2
COLORADO STATE
UNIVERSITY
FORT COLLINS, COLORADO
80523

department of electrical engineering

UCRL 15424

#6

FINAL REPORT

by

Thomas A. Brubaker

April 7, 1981

LAWRENCE LIVERMORE LABORATORY

MASTER

UCRL--15424

DE82 006857

COMPUTER AUTOMATION FOR CHEMISTRY INSTRUMENTS

Time Period: February 1, 1980 - January 31, 1981

Contract 7864309

between

THE REGENTS OF THE
UNIVERSITY OF CALIFORNIA

and

COLORADO STATE UNIVERSITY

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

NOTICE

Work performed under the auspices of the U.S. Department of Energy by the Lawrence Livermore Laboratory under contract number W-7405-ENG-48.

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

INTRODUCTION

This report covers the time frame from February 1, 1980 through January 31, 1981. The work performed consisted of the following:

1. Development of computer aided design software for polynomial filters and Butterworth-Chebyshev filter design.
2. Development of algorithms for resolving peaks in spectroscopic data.
3. Interaction with LLL personnel to develop a structure for automation of surface chemistry spectroscopy.

Each work phase is now discussed in detail.

1. Computer Aided Design Software

The polynomial filters use a polynomial model over a finite data window. The programs were written to allow the user to specify the order of polynomial, the point in the window where the output is to occur and the window size. Coefficients are then automatically calculated. Since the design is based on state space concepts, the result is a set of filters that estimate the data and the data derivatives. The number of estimators is determined by the order of the model. For example, if a second order model is chosen, the program generates the coefficients for estimating the data and the first two derivatives of the data.

The package also permits the design of a Kalman filter using a polynomial data model. Here, the user specifies the fading factor and the program then generates estimates of the data and derivatives.

The Butterworth-Chebyshev filters operate in the same way except that different parameters are specified. These filters are designed using frequency domain principles so the user specifies center frequency

and bandwidth as well as the filter order. The program then generates the filter coefficients for the digital filter representation.

The programs both operate on data files, however, no real time capability exists. To accomplish this one would link to the real time data acquisition routines after the filters are designed. Reports on the programs were included when the software was delivered

2. Peak Resolution

An M.S. Thesis was written on this topic. Marquardt nonlinear parameter estimation was employed first and then a combined Marquardt-Kalman parameter estimation approach was investigated and shown to work. A copy of the thesis including all results is included in Appendix A.

3. Automation

The automation project has evolved and the software was developed by LLL personnel.

Appendix A

THESIS

SPECTRAL PEAK SEPARATION OF SUMMED
GAUSSIAN MODELS

Submitted by

Celeste H. Johnson

Electrical Engineering Department

In partial fulfillment of the requirements

for the degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall, 1980

COLORADO STATE UNIVERSITY

Fall 1980

WE HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER OUR SUPERVISION
BY CELESTE H. JOHNSON
ENTITLED SPECTRAL PEAK SEPARATION OF SUMMED GAUSSIAN MODELS
BE ACCEPTED AS FULFILLING IN PART REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

Committee on Graduate Work

C. Budak

R. K. Hopwood

1
C. Budak

Adviser

ABSTRACT OF THESIS

SPECTRAL PEAK SEPARATION OF SUMMED GAUSSIAN MODELS

Spectroscopy measurements of chemical compounds yield information which can be correlated to the elements forming the compound's composition. The information from the data that is required to identify the components of the compound is normally not directly visible in the observed spectrum. Methods for separating each element's contribution to the measured spectrum are of interest to aid in the analysis of spectral data.

This thesis examines a recursive algorithm for spectral peak splitting. The recursion is established using a nonlinear representation of a Kalman filter. The nonlinear parameters required by the filter are obtained by a normalized windowed Marquardt or a "stochastic" method. These nonlinear parameters describe a gaussian model which has physical significance in the representation of an element's spectral response. The amplitude, a linear term, of each gaussian model is separated from the nonlinear parameters and is considered after the gaussian separation has been achieved. Summed gaussian data sets with additive white noise are used in this analysis.

The recursion algorithm in conjunction with the separation of variables requires less computer time and memory than batch Marquardt, a commonly used nonlinear parameter estimation technique. Batch Marquardt is examined in terms of the accuracy of the estimates of the

linear and nonlinear parameters describing the gaussian models. The accuracy and flexibility of batch Marquardt is compared with the recursive algorithm described within this thesis.

Celeste H. Johnson
Electrical Engineering Department
Colorado State University
Fort Collins, Colorado
Fall, 1980

TABLE OF CONTENTS

<u>Chapter</u>	<u>Page</u>
I. INTRODUCTION	1
II. BACKGROUND	6
A. Linear Parameter Estimation--Method of Least Squares	10
B. Recursive State Space Filtering--A Kalman Filter	13
C. Nonlinear-Parameter Estimation--Marquardt	20
III. MARQUARDT NONLINEAR PARAMETER ESTIMATION	26
IV. RECURSIVE NONLINEAR KALMAN FORMULATION	79
V. CONCLUSIONS	111
REFERENCES	114
APPENDIX A--LIST OF SYMBOLS	116

CHAPTER I

INTRODUCTION

The measured spectral response of a compound is often described in terms of a sequence of summed gaussian models. This modelling is based on a well defined physical interpretation of the chemistry and state of the compound, and the measurement procedure. Each element in the compound can be identified by its own characteristic gaussian model. The data set represents a summation of these gaussian models, making the direct measurement of a single element's contribution impossible. Consequently, the ability to separate each gaussian in the data set is desirable to obtain the spectral contribution from each element.

The ability to identify these individual components deteriorates due to several factors. First, as the number of components per data set increases, the number of parameters required to identify the spectrum's elements multiplies, increasing the complexity of the problem. Second, the contributions from an element often overlap those of another element, biasing each representative gaussian model. In the data set this biasing induces peak shifting of the element's actual peak location and a broadening of the dispersion about the peak. Finally, as the signal to noise ratio decreases, the spectral information is lost in random noise.

Traditionally, digital algorithms used to identify these gaussian models have been based on locating the peaks and shoulders

associated with each component in the data set. This is typically attained by searching for zero crossings in various orders of derivatives [1]. Smoothed data sets are required to make these derivatives a viable means of sensing zero crossings of the signal rather than the noise. Savitsky and Golay's method of least squares filtering by convoluting data points with chosen integers accomplishes this smoothing [2].

An algorithm which depends solely on peak and shoulder detection to quantify an element's contribution to the data set can lead to erroneous results. The peak locations derived from the zero crossings of the derivatives may be biased from their actual locations by supplemental values added from adjacent components. The need for optimization is clearly shown by the difference between the fit obtained in Figure 1-1 to Figure 1-2 [3]. Peak and shoulder detection of each component can then be utilized in the selection of an accurate model for a batch mode optimization routine.

Marquardt's nonlinear parameter estimation technique is one example of a batch mode optimization algorithm. Under most conditions, this method accurately parameterizes each gaussian model summed in a noisy data set; however, the computer time and memory requirements are high. The batch Marquardt algorithm requires that the entire data set and each partial with respect to the parameters of the model be located in memory throughout processing. As the size of the parameter and data set increases the rate of convergence decreases, while the memory allocation necessary to store this information expands.

These limitations motivate a search for an algorithm which can obtain results comparable to batch Marquardt while consuming less time

Figure 1-1. A mass spectrum separated by peak and shoulder detection [3].

Figure 1-2. A mass spectrum separated by optimization of information supplied by peak and shoulder detection [3].

and storage in its implementation. The algorithm subsequently derived depends on separable variables and a recursive formulation to meet these constraints. In Chapter II the necessary background information and derivations are presented which form the basis of this algorithm. Basic stochastic definitions, linear parameter estimation, recursive Kalman state space estimation, and Marquardt's nonlinear parameter estimation are reviewed.

Chapter III concentrates on Marquardt's nonlinear parameter estimation algorithm. The accuracy of the parameter estimates returned by the batch mode technique are studied using specific summed gaussian models. This algorithm is then used over limited portions, windows, of the data set which have been normalized. By selectively windowing the data, this technique results in the ability to obtain the nonlinear parameters of each gaussian model contained in the measured spectrum. The normalization of the windowed data removes the linear term describing the amplitude of the gaussian model. Thus, reduction in the size of the data and parameter set which is used by the algorithm at any one time is accomplished.

The nonlinear parameters, the mean and the standard deviation, of the gaussian models have been estimated, but each gaussian has not yet been separated from the measured spectrum. Chapter IV redevelops the Kalman filter as a nonlinear estimator of a gaussian. This filter depends only on the nonlinear parameters previously estimated for each gaussian in the data set. The filter is capable of removing each gaussian from the summed gaussian expression. Limitations of the normalized window Marquardt algorithm necessitate an alternative "stochastic" method to be developed for obtaining the nonlinear

parameters of each gaussian model. A recursive formulation between the "stochastic" parameter estimation, Kalman filtering, and normalized window Marquardt is implemented to obtain close approximations of the ideal summed gaussians in a noisy measurement set.

Chapter V summarizes the general conclusions. A method of obtaining the amplitude of each of the gaussians now separated from the data set and other suggestions are given for future work.

CHAPTER II

BACKGROUND

This chapter will briefly review various classic estimation techniques along with some probability theory. The estimation techniques include linear parameter estimation by least squares, recursive state space filtering by a Kalman filter, and nonlinear parameter estimation using the Marquardt algorithm. In later chapters these techniques are used in an iterative fashion to obtain both linear and nonlinear parameters for a gaussian model. Consequently, a basic understanding of each technique is necessary before this iteration process can be developed.

Often, the notation attached to these techniques is confusing, for example, differentiating between the elements of a simple linear equation and a matrix representation of simultaneous equations. To avoid this confusion, the standard form of vectors and matrices will be capital letters underlined, i.e., $\underline{A}$. Matrix dimensions will be denoted in parenthesis by the number of rows and then columns, (r,c) and scalar quantities, which form a matrix, will be lower case letters whose subscript will show their respective location in the matrix.

$$\underline{A}(n,p) = \begin{bmatrix} a_{11} & \dots & a_{1j} & \dots & a_{1p} \\ a_{i1} & \dots & a_{ij} & \dots & a_{ip} \\ a_{n1} & \dots & a_{nj} & \dots & a_{np} \end{bmatrix}$$

The matrix is $\underline{A}$ of dimension $n \times p$. Each a_{ij} is a scalar quantity in the i^{th} row and j^{th} column of the matrix. In the case of vectors, only one subscript is necessary to define the location of a vector element.

$$\underline{A} = \begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix}$$

In this case, column vector $\underline{A}$ has n elements, each designated by a_i . Lower case letters will also signify scalar multiples of matrices and vectors. The reader may find it convenient to refer to Appendix A for a specific definition of terms at some point in this or succeeding chapters.

In each of the following sections, discussing estimation techniques, we will assume a model plus additive noise to describe the actual measurement made from the physical process. It is also assumed that this additive noise has a gaussian density distribution. Stochastically, this defines the results of an infinite number of observations and describes the way in which these observations are distributed. The probability of observing any noise value, n_j an element of the set N , is represented by

$$P[N=n_j] = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} (n_j - \mu)^2/\sigma^2\right]$$

Here μ is the mean of the parent distribution, $P[N]$, and is equivalent to the average value of an infinite number of observations.

$$\mu = \lim_{p \rightarrow \infty} \frac{1}{p} \sum_{i=1}^p y_i$$

where y_i is the measured observation.

The standard deviation, δ , is a measure of the dispersion of the observations about the mean. The variance, δ^2 , is defined by

$$\delta^2 = \lim_{p \rightarrow \infty} \left[\frac{1}{p} \sum_{i=1}^p (y_i - \mu)^2 \right]$$

The standard deviation and the variance characterize the uncertainties which are associated with the measurements [4].

Figure 2-1 shows that the gaussian density distribution is symmetric. This means that the probability of observing a noise value less than or equal to the mean, is the same as the probability of observing a value greater than or equal to the mean.

Figure 2-1. Gaussian density distribution.

It is assumed that the random variables are stationary and ergodic. This means that the probability densities are independent of time, and the mean and the variance are constants over all time [5]. Furthermore, the mean of the noise will be defined to be zero.

It is further assumed that there will be no correlation between noise samples. This means that the expected value between two noise

samples at different times will be zero, while the expected value of the noise at one time squared is equal to the variance. In symbolic notation

$$E[n_i, n_j] = \begin{cases} 0 & i \neq j \\ \sigma^2 & i = j \end{cases}$$

To relate this notation to the mean and variance previously defined, the mean and the variance must be redefined in terms of the parent distribution and the possible discrete values of n_j , the noise.

It is realistically impossible to measure continuous values of n , therefore a discrete distribution is assumed, n_j . It is also impossible to measure the infinite possibilities of n_j . Consequently, a finite set of points must be taken to obtain the best estimate for the mean and standard deviation. The mean can then be redefined as

$$\mu = \sum_{j=1}^p n_j P(N=n_j) \approx E[N]$$

and similarly,

$$\sigma^2 = \sum_{j=1}^p (n_j - \mu)^2 P(N = n_j) \approx E[(N - E[N])^2]$$

where N represents the finite set of p discrete values of n .

With these definitions in mind, the covariance matrix can be defined as:

$$\underline{S} = E \begin{bmatrix} n_1 n_1 & n_1 n_j & \dots & n_1 n_p \\ n_j n_1 & n_j n_j & \dots & n_j n_p \\ \vdots & \vdots & \ddots & \vdots \\ n_p n_1 & n_p n_j & \dots & n_p n_p \end{bmatrix}$$

Remember that the mean, $E[N]$, is assumed zero. With the expectation taken inside the matrix the covariance matrix becomes:

$$\underline{S} = \begin{bmatrix} \delta^2 & 0 & 0 & \dots & 0 \\ 0 & \delta^2 & 0 & \dots & 0 \\ 0 & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \delta^2 \end{bmatrix} = \delta^2 \underline{I}$$

where $\underline{I}$ is the identity matrix. The zeros in the off-diagonal terms result from the assumption of uncorrelated noise samples, while the diagonal terms are derived from the ergodic process of constant variance δ^2 [1], [5].

This stationary, ergodic, zero mean, and uncorrelated noise process is termed white gaussian noise.

A. Linear Parameter Estimation--Method of Least Squares

Measurements from a physical system may be represented by a model of the process summed with white gaussian noise. Given n measurements from this system the i^{th} measurement is defined as

$$y_i = m_i + n_i$$

where y_i represents the measurement, n_i represents the additive noise, and m_i is the model for the process at this observation. A model is defined to represent the essential aspects of the process under study [6]. For a process linear with respect to the parameters, the model for the i^{th} observation is chosen in the form

$$m_i = \sum_{j=1}^p h_{ij} b_j$$

where h_{ij} represents independent variables which relate the p parameters ($b_1, b_2, \dots, b_j, \dots, b_p$) to the model, m_i .

The total measurement vector can be represented as

$$\underline{Y} = \underline{H} \underline{B} + \underline{N}$$

or

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1p} \\ h_{21} & h_{22} & \cdots & h_{2p} \\ \vdots & \vdots & & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{np} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_n \end{bmatrix}$$

It should be noted that the term linear refers only to the relationship of the parameters to the model and does not comment on the system's linearity [7]. Consequently, the $\underline{H}$ matrix components, h_{ij} , may be any type of function, not necessarily linear.

It is clear that the parameters $\underline{b}$ are not directly observable. An estimated parameter vector, $\hat{\underline{b}}$, must be found such that the difference between the actual measurement and the model response is minimized. It is expected that the difference will not be exactly zero, thus an error vector, $\underline{r}$, with elements r_i and length n , will represent this difference.

$$\underline{r} = \underline{y} - \hat{\underline{m}}$$

where

$$\hat{\underline{m}} = \underline{H} \hat{\underline{b}}$$

The i^{th} error element can be expressed as

$$r_i = y_i - \sum_{j=1}^p h_{ij} \hat{b}_j$$

To minimize this observed difference requires some criteria for optimizing the estimates of the parameters. Generally, the least-squares method is used for its simplicity and its experimental validity.

The method of least-squares is built on the hypothesis that the optimum description of the measurements is one which minimizes the sum

of the squares of the deviations of the data from the estimated model [4]. The error term of the sum squared errors will then be

$$e = \sum_{i=1}^n r_i^2 = \sum_{i=1}^n [y_i - \hat{m}_i]^2$$

and in vector form

$$e = \underline{R}^T \underline{R} = (\underline{Y} - \hat{\underline{M}})^T (\underline{Y} - \hat{\underline{M}}) \quad (2-1)$$

Notice that e will always be positive, from the fact that r_i^2 is always positive. The sum squared error is in even or quadratic form due to these squared terms. This means that the error term appears as a parabola with respect to any one of the estimated parameters of the linear model [8]. This guarantees the existence of p parameter estimates at the minimum error value of e .

To minimize the error term, differentiate e with respect to $\hat{\underline{B}}$. The resulting gradient is set equal to zero and solved for $\hat{\underline{B}}$. The gradient can be shown to be

$$\frac{\partial e}{\partial \underline{B}} = -2\underline{H}^T \underline{Y} + 2\underline{H}^T \underline{H} \hat{\underline{B}}$$

When this gradient is set to zero the optimum estimate is found.

$$\hat{\underline{B}} = (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{Y}$$

$\hat{\underline{B}}$ is the least squares estimate of the parameters and depends upon the measurements and independent variables.

These estimates are unbiased, meaning the estimates $\hat{\underline{B}}$ are random variables with mean $\underline{B}$. To observe this, look at the expected value of $\hat{\underline{B}}$.

$$\begin{aligned}
E[\hat{\underline{B}}] &= E[(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{Y}] \\
&= E[(\underline{H}^T \underline{H})^{-1} \underline{H}^T (\underline{H} \underline{B} + \underline{N})] \\
&= E[\underline{B} + (\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{N}] \\
&= E[\underline{B}] = \underline{B}
\end{aligned}$$

where $E[(\underline{H}^T \underline{H})^{-1} \underline{H}^T \underline{N}] = \emptyset$ because $E[\underline{N}] = \emptyset$ for white gaussian noise. In a similar fashion, the covariance of $\hat{\underline{B}}$ can be found to be

$$\underline{S} = \text{Cov } \hat{\underline{B}} = (\underline{H}^T \underline{H})^{-1} \delta^2$$

where δ^2 is the variance of the white gaussian noise.

If the noise was assumed correlated rather than uncorrelated, with the expected value still zero, the covariance matrix of the noise would no longer be a scalar multiple of the identity matrix. This change would effect the estimation procedure just outlined.

The noise covariance matrix becomes

$$E[\underline{N} \underline{N}^T] = \underline{W}$$

for $E[\underline{N}]$ is still assumed zero. The optimum estimate of $\hat{\underline{B}}$ would now become a minimum variance estimate [5].

$$\hat{\underline{B}} = [\underline{H}^T \underline{W}^{-1} \underline{H}]^{-1} \underline{H}^T \underline{W}^{-1} \underline{Y}.$$

For our purposes the weighted least squares will not be used since the weights are not known.

B. Recursive State Space Filtering-- A Kalman Filter

Often the process measured in the laboratory has an ordinary or partial differential equation as a model. Here a linear dynamic system is assumed and described as

$$\frac{d\underline{B}_k}{dt} = \underline{F}(t)\underline{B}_k + \underline{D}(t) \underline{U}(t)$$

and

$$y_k = \underline{O}(t) \underline{B}_k + n_k \quad (2-2)$$

$\underline{B}_k$ is a p-element state vector at time t_k representing the p parameters necessary to define the state of the system. $\underline{U}(t)$ is an m-element vector representing the inputs of the system, while $\underline{F}(t)$ and $\underline{D}(t)$ are (p,p) and (p,m) matrices, respectively. Again, a stationary system is assumed; hence, $\underline{F}(t)$, $\underline{D}(t)$, $\underline{U}(t)$ and $\underline{O}(t)$ reduce to $\underline{F}$, $\underline{D}$, $\underline{U}$, and $\underline{O}$, constants over time. Finally, y_k represents the measurement of the system at time t_k , and n_k is an additive noise element, whose stochastic description is defined as white gaussian noise. $\underline{O}$ is a row vector of p-elements relating the state vector to the measurement.

The dependent variables, b_{ik} , of the model are called state variables. The differential equation defines these state variables to be linear with respect to the model. Kalman filtering redefines a least squares formulation in a recursive form. This recursion effectively utilizes the digital computer's computational ability. The complex matrix manipulations shown in section A are minimized in achieving the estimates. Iterative summations and multiplications of, ideally, scalar quantities, or at worst, small matrices are used to obtain these estimates. It is computationally faster and easier to add and multiply on a computer than to invert large matrices.

Wiener and his successors, Kalman and Bucy, used this technique to solve problems in the prediction of random signals and the separation of these random signals from random noise [9]. In this development, the technique will be applied to the estimation of signals of a known form in the presence of white noise. This implies that no errors are associated with the state vector $\underline{B}_k$; it is the true state vector.

Equation 2-2 denotes a Markov process. The future behavior of the system depends only on its state at a particular time and the statistical nature of the additive noise [10]. This results in the definition of a state vector at time t_{k+1} as

$$\underline{B}_{k+1} = \underline{\Phi}_k(1) \underline{B}_k \quad (2-3)$$

$\underline{\Phi}_k$ is (p,p) and represents the known dynamics of the system. It is also called a state transition matrix. $\underline{\Phi}_k(1)$ indicates the change from a state at time t_k to a state at time t_{k+1} . Similarly, a change from t_k to t_{k-1} is signified by $\underline{\Phi}_k(-1)$. In general terms, the state transition matrix at time t has the following properties:

- 1] $\underline{\Phi}_\ell(0) = \underline{I}$
- 2] $\underline{\Phi}_\ell(n)^{-1} = \underline{\Phi}_{\ell+n}(-n)$
- 3] $\underline{\Phi}_\ell(n+r) = \underline{\Phi}_\ell(n)\underline{\Phi}_{\ell+n}(r)$

From these properties it can be shown that any state may be represented in terms of another state.

$$\underline{B}_{k+n} = \underline{\Phi}_k(n)\underline{B}_k$$

Driving noise could be added to this equation just as the previous section could have used weighted least squares; however, the noise statistics and weighting are again unknown.

The nonaccented $\underline{B}_k$ represents the true state values. The measurement process at each time is then described by the following equations:

$$\begin{aligned} y_k &= \underline{0} \underline{B}_k + n_k \\ y_{k-1} &= \underline{0} \underline{B}_{k-1} + n_{k-1} \\ &\vdots \\ y_1 &= \underline{0} \underline{B}_1 + n_1 \end{aligned}$$

Because each of the states are related through a $\underline{\Phi}_k$ matrix these equations can be reformulated to:

$$y_k = \underline{0} \underline{\Phi}_k (0) \underline{B}_k + n_k$$

$$y_{k-1} = \underline{0} \underline{\Phi}_k (-1) \underline{B}_k + n_{k-1}$$

$$y_1 = \underline{0} \underline{\Phi}_k (-k+1) \underline{B}_k + n_1$$

and in matrix form

$$\underline{Y}_k = \underline{H}_k \underline{B}_k + \underline{N}_k$$

where

$$\underline{H}_k = \begin{bmatrix} \underline{0} \underline{\Phi}_k (0) \\ \underline{0} \underline{\Phi}_k (-1) \\ \vdots \\ \underline{0} \underline{\Phi}_k (-k+1) \end{bmatrix} \quad \text{and} \quad \underline{Y}_k = \begin{bmatrix} y_1 \\ y_{k-1} \\ \vdots \\ y_1 \end{bmatrix}$$

This is now in the same format as the least squares formulation.

The optimal estimate for the true state value is

$$\hat{\underline{B}}_k = (\underline{H}_k^T \underline{H}_k)^{-1} \underline{H}_k^T \underline{Y}_k$$

From the definition of the $\underline{\Phi}$ matrix, which relates states in a linear way, the forecast of the optimal estimate can be obtained without the related data point present. Consider,

$$\underline{Y}_k = \underline{H}_{k+1}' \underline{B}_{k+1}' + \underline{N}_k$$

where

$$\underline{H}_{k+1}' = \begin{bmatrix} \underline{0} \underline{\Phi}_{k+1} (-1) \\ \underline{0} \underline{\Phi}_{k+1} (-2) \\ \vdots \\ \underline{0} \underline{\Phi}_{k+1} (-k) \end{bmatrix} = \underline{H}_k \underline{\Phi}_{k+1} (-1)$$

and $\underline{\hat{B}}_{k+1}$ is the true state vector at time t_{k+1} which is being predicted.

The optimum value of the estimate forecast is

$$\begin{aligned}
 \underline{\hat{B}}_{k+1} &= (\underline{H}_{k+1}^T \underline{H}_{k+1})^{-1} \underline{H}_{k+1}^T \underline{Y}_k \\
 &= [\underline{\Phi}_{k+1}^T (-1) \underline{H}_{k-k}^T \underline{H}_{k+1} \underline{\Phi}_{k+1} (-1)]^{-1} \underline{\Phi}_{k+1}^T (-1) \underline{H}_{k-k}^T \underline{Y}_k \\
 &= \underline{\Phi}_{k+1}^{-1} (-1) \underline{H}_{k-k}^{-1} [\underline{H}_{k-k}^T]^{-1} [\underline{\Phi}_{k+1}^T (-1)]^{-1} \underline{\Phi}_{k+1}^T (-1) \underline{H}_{k-k}^T \underline{Y}_k \\
 &= \underline{\Phi}_k (1) (\underline{H}_{k-k}^T \underline{H}_{k-k})^{-1} \underline{H}_{k-k}^T \underline{Y}_k \\
 &= \underline{\Phi}_k (1) \underline{\hat{B}}_k
 \end{aligned} \tag{2-5}$$

The forecasted estimate is found to be the last state's optimum estimate multiplied by the respective state transition matrix.

The covariance of this forecasted estimate is defined as

$$\begin{aligned}
 \underline{S}_{k+1} &= (\underline{H}_{k+1}^T \underline{H}_{k+1})^{-1} \delta^2 \\
 &= \underline{\Phi}_k (1) \underline{S}_k \underline{\Phi}_k^T (1)
 \end{aligned} \tag{2-6}$$

where $\underline{S}_k$ is the covariance of $\underline{\hat{B}}_k$.

$$\underline{S}_k = (\underline{H}_k^T \underline{H}_k)^{-1} \delta^2$$

The variance, δ^2 , is derived from the additive white gaussian noise.

To formulate the recursion, a new data point is added such that the measurement vector becomes

$$\underline{Y}_{k+1} = \underline{H}_{k+1} \underline{B}_{k+1} + \underline{N}_{k+1}$$

where

$$\underline{H}_{k+1} = \begin{bmatrix} 0 & \underline{\Phi}_{k+1}(0) \\ 0 & \underline{\Phi}_{k+1}(-1) \\ \vdots & \vdots \\ 0 & \underline{\Phi}_{k+1}(-k) \end{bmatrix} = \begin{bmatrix} 0 & \underline{\Phi}_k(1) \\ \vdots & \vdots \\ \underline{H}_k \end{bmatrix} \underline{\Phi}_{k+1}(-1)$$

and

$$\underline{Y}_{k+1} = \begin{bmatrix} y_{k+1} \\ \vdots \\ y_k \end{bmatrix}$$

The covariance of $\hat{\underline{B}}_{k+1}$ can be rewritten in terms of the forecasted covariance.

$$\begin{aligned} \underline{S}_{k+1} &= (\underline{H}_{k+1}^T \underline{H}_{k+1})^{-1} \delta^2 \\ &= (\underline{\Phi}_{k+1}^T (-1) [\underline{\Phi}_k^T (1) \underline{0}^T : \underline{H}_k^T] \begin{bmatrix} \underline{0} & \underline{\Phi}_k^T (1) \\ \vdots & \vdots \\ \underline{H}_k \end{bmatrix} \underline{\Phi}_{k+1} (-1))^{-1} \delta^2 \\ &= (\underline{0}^T \underline{0} + \underline{\Phi}_{k+1}^T (-1) \underline{H}_k^T \underline{H}_k \underline{\Phi}_{k+1} (-1))^{-1} \delta^2 \\ &= \left(\frac{\underline{0}^T \underline{0}}{\delta^2} + \frac{\underline{\Phi}_{k+1}^T (-1) \underline{H}_k^T \underline{H}_k \underline{\Phi}_{k+1} (-1)}{\delta^2} \right)^{-1} \end{aligned}$$

and since

$$(\underline{S}_{k+1}^-)^{-1} = \frac{\underline{\Phi}_{k+1}^T (-1) \underline{H}_k^T \underline{H}_k \underline{\Phi}_{k+1} (-1)}{\delta^2}$$

from (2-6) and (2-4) then

$$\underline{S}_{k+1} = \left(\frac{\underline{0}^T \underline{0}}{\delta^2} + (\underline{S}_{k+1}^-)^{-1} \right)^{-1} \quad (2-7)$$

Using this result in the estimate of state $\underline{B}_{k+1}$

$$\hat{\underline{B}}_{k+1} = (\underline{H}_{k+1}^T \underline{H}_{k+1})^{-1} \underline{H}_{k+1}^T \underline{Y}_{k+1}$$

becomes

$$\begin{aligned} &= \frac{\underline{S}_{k+1}}{\delta^2} \left\{ \underline{\Phi}_{k+1}^T (-1) [\underline{\Phi}_k^T (1) \underline{0}^T : \underline{H}_k^T] \begin{bmatrix} y_{k+1} \\ \vdots \\ y_k \end{bmatrix} \right\} \\ &= \frac{\underline{S}_{k+1}}{\delta^2} \left\{ \underline{0}^T y_{k+1} + \underline{\Phi}_{k+1}^T (-1) \underline{H}_k^T y_k \right\} \\ &= \frac{\underline{S}_{k+1}}{\delta^2} \left\{ \underline{0}^T y_{k+1} + (\underline{S}_{k+1}^-)^{-1} \delta^2 \hat{\underline{B}}_{k+1} \right\} \end{aligned}$$

$$\begin{aligned}
&= \frac{S_{k+1}}{\delta^2} \{ \delta^2 (S_{k+1}')^{-1} \hat{B}_{k+1}' + 0^T 0 \hat{B}_{k+1}' - 0^T 0 \hat{B}_{k+1}' + 0^T y_{k+1} \} \\
&= S_{k+1} \left\{ [(S_{k+1}')^{-1} + \frac{0^T 0}{\delta^2}] \hat{B}_{k+1}' - \frac{0^T 0}{\delta^2} \hat{B}_{k+1}' + \frac{0^T}{\delta^2} y_{k+1} \right\} \\
&= S_{k+1} \left\{ (S_{k+1}')^{-1} \hat{B}_{k+1}' - \frac{0^T 0}{\delta^2} \hat{B}_{k+1}' + \frac{0^T}{\delta^2} y_{k+1} \right\}
\end{aligned}$$

and finally

$$\hat{B}_{k+1} = \hat{B}_{k+1}' + \frac{S_{k+1}}{\delta^2} 0 \{ y_{k+1} - 0 \hat{B}_{k+1}' \} \quad (2-8)$$

To define each term for clarity

- 1] $\hat{B}_{k+1}' = \Phi_k(1) \hat{B}_k$ forecast estimate from (2-5)
- 2] $S_{k+1}' = \Phi_k(1) S_k \Phi_k^T(1)$ forecast covariance from (2-6)
- 3] $S_{k+1} = \left[\frac{0^T 0}{\delta^2} + (S_{k+1}')^{-1} \right]^{-1}$ covariance at time t_{k+1} from (2-7)
- 4] $P_{k+1} = \frac{S_{k+1} 0^T}{\delta^2}$
- 5] $\hat{B}_{k+1} = \hat{B}_{k+1}' + P_{k+1} \{ y_{k+1} - 0 \hat{B}_{k+1}' \}$ recursive equation from (2-8)

Notice that the recursion equation represents a growing-memory window that is easily implemented as each new data point is introduced. Also notice that the covariance matrix plays an important role in weighting the new data point. The covariance matrix gives some indication of the reliability of the model's prediction.

These equations can be reformulated into the classic Kalman form [5], [10], through some matrix manipulation. They become

- 1] $\hat{B}_{k+1}' = \Phi_k(1) \hat{B}_k$
- 2] $S_{k+1}' = \Phi_k(1) S_k \Phi_k^T(1)$

$$3] \quad \underline{P}_{k+1} = \underline{S}_{k+1}^{-1} \underline{O}^T [\delta^2 + \underline{O} \underline{S}_{k+1}^{-1} \underline{O}^T]^{-1}$$

$$4] \quad \underline{S}_{k+1} = [\underline{I} - \underline{P}_{k+1} \underline{O}] \underline{S}_{k+1}^{-1}$$

$$5] \quad \hat{\underline{B}}_{k+1} = \hat{\underline{B}}_{k+1}^{-1} + \underline{P}_{k+1} [y_{k+1} - \underline{O} \hat{\underline{B}}_{k+1}^{-1}]$$

C. Nonlinear-Parameter Estimation--Marquardt

Until now the functions used to model the system have been linear with respect to the parameters and states estimated. Methods have been extrapolated from linear least-squares to fit nonlinear system models.

The optimum estimates from least squares were obtained by calculating the first derivative of the sum squared error term with respect to each of the parameters and equating this result to zero. In the case of a nonlinear function, an analytical expression for the parameters may not be trivial to obtain. Instead of a closed form solution, the error term may be expressed as a continuous function of n parameters describing a hyperspace. A two dimensional hyperspace is illustrated in Fig. 3-14. This hypersurface can then be searched for a minimum of the n -parameters. This minimum may not necessarily be the global minimum; therefore, it is necessary that the search for this minimum be started within a reasonable distance, so that convergence is ensured.

Marquardt in 1963 developed a "maximum neighborhood" method which performs an optimum interpolation between the Newton-Raphson method and gradient method search techniques.

The Newton-Raphson method expresses the error term hypersurface in terms of a second-order Taylor series expansion. It can be interpreted as finding the minimum of a hyperparaboloid.

$$e(\hat{\underline{B}}) = e(\hat{\underline{B}}_i) + \underline{\nabla}_{\underline{B}}^{\wedge}(e) \big|_{\underline{B}=\hat{\underline{B}}_i} (\hat{\underline{B}} - \hat{\underline{B}}_i) + \frac{1}{2} (\hat{\underline{B}} - \hat{\underline{B}}_i)^T \underline{\nabla}_{\underline{B}}^2 \hat{\underline{B}}^{\wedge}(e) \big|_{\underline{B}=\hat{\underline{B}}_i} (\hat{\underline{B}} - \hat{\underline{B}}_i)$$

The error term, e , is now a nonlinear function of the estimated parameters. The $\underline{\nabla}_{\underline{B}}^{\wedge}(e)$ represents the gradient of e or the first partial of e with respect to $\underline{B}$ and is evaluated at the i^{th} iteration's value of $\hat{\underline{B}}$, $\hat{\underline{B}}_i$. The hessian, or the second partial of e with respect to $\underline{B}$ is denoted by $\underline{\nabla}_{\underline{B}}^2 \hat{\underline{B}}^{\wedge}(e)$.

The minimum of the error term is found by differentiating $e(\hat{\underline{B}})$ with respect to $\underline{B}$, evaluating the result at the next iteration's estimate, $\hat{\underline{B}}_{i+1}$ and setting the result to zero. This recursion equation is

$$\hat{\underline{B}}_{i+1} = \hat{\underline{B}}_i - (\underline{\nabla}_{\underline{B}}^2 \hat{\underline{B}}^{\wedge}(e) \big|_{\underline{B}=\hat{\underline{B}}_i})^{-1} \underline{\nabla}_{\underline{B}}^{\wedge}(e) \big|_{\underline{B}=\hat{\underline{B}}_i}$$

The hessian expresses the step size, while the gradient denotes the direction of the step. Subsequently, an error hyperspace is searched until a minimum is found. A one parameter search is shown in Figure 2-2 [6], [8].

Figure 2-2. Example of a Newton-Raphson search of a one parameter error space.

The error term, a nonlinear function of the estimated variables, can still be expressed as in (2-1) of the linear least squares case.

$$c(\underline{\hat{B}}) = (\underline{Y} - \underline{\hat{M}})^T (\underline{Y} - \underline{\hat{M}})$$

where the estimated model, $\underline{\hat{M}}$, now is nonlinear with respect to the parameters. It can be shown that

$$\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(e) = \underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)(\underline{\hat{M}} - \underline{Y})$$

and

$$\underline{\nabla}_{\underline{\hat{B}}}^2(e) = 2\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)^T + 2\underline{\nabla}_{\underline{\hat{B}}}^2(\underline{\hat{M}}^T)(\underline{\hat{M}} - \underline{Y})$$

The hessian, now expressed in terms of the jacobian and the difference between the model and the measurements, must be positive definite throughout the search for a minimum. If not, the hessian will interfere with the direction that the gradient supplies for convergence.

When it is assumed that the estimate is close to the minimum then the hessian reduces to

$$\underline{\nabla}_{\underline{\hat{B}}}^2(e) = 2\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)^T$$

and the Newton-Raphson method has reduced to the Gauss-Newton method.

$$\underline{\hat{B}}_{i+1} = \underline{\hat{B}}_i - [2\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)\underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)^T]^{-1} \underline{\nabla}_{\underline{\hat{B}}}^{\wedge}(\underline{\hat{M}}^T)(\underline{\hat{M}} - \underline{Y})$$

Both of these methods use the inverse of the hessian, or its approximation, to define the step size to converge on the minimum estimate. As the minimum is approached this quantity goes to zero causing the step sizes to become very large. The parameter estimates become oscillatory and may even step out of range of the computer.

In the case where one of the parameters does not effect the error surface significantly, a small sensitivity coefficient, a near singularity or singularity in the hessian would exist. A near singularity in the matrix would cause a very large step size to be produced

and a singularity would cause the inverse to become impossible to compute. From these limitations, it is clear that convergence is not guaranteed.

The various methods of steepest descent also search the error hypersurface for a minimum. Their general form is described by

$$\hat{\underline{B}}_{i+1} = \hat{\underline{B}}_i - g \nabla_{\underline{B}}^{\wedge}(e) \Big|_{\hat{\underline{B}}=\hat{\underline{B}}_i}$$

In this recursion g is a constant which varies the step size. This method works well far from the minimum, but as the minimum is approached the gradient decreases rapidly, leading to a very slow convergence rate. Even with careful control of the step size, slow convergence is the typical outcome.

Marquardt's algorithm interpolates between these two methods as seen in the following equation.

$$\hat{\underline{B}}_{i+1} = \hat{\underline{B}}_i - \underbrace{[\nabla_{\underline{B}}^{\wedge}(\hat{\underline{M}}^T) \nabla_{\underline{B}}^{\wedge}(\hat{\underline{M}}^T)^T + g \underline{L}]}_{\text{step size}}^{-1} \underbrace{\nabla_{\underline{B}}^{\wedge}(\hat{\underline{M}}^T)(\hat{\underline{M}} - \underline{Y})}_{\text{direction}} \Big|_{\hat{\underline{B}}=\hat{\underline{B}}_i}$$

As the minimum is approached the step size will not become unstable. The product $g\underline{L}$ is chosen such that it is large enough to prevent instability, yet small enough to provide step sizes large enough for rapid convergence. This eliminates the major drawbacks encountered with each of the preceding methods.

Marquardt utilized Levenberg's notion of "damped least squares," for the choice of the $\underline{L}$ matrix. Although, the weighting supplied by g is different from Levenberg's original suggestion [11].

$$\underline{L} = \text{diagonal } (\nabla_{\underline{B}}^{\wedge}(\hat{\underline{M}}^T) \nabla_{\underline{B}}^{\wedge}(\hat{\underline{M}}^T)^T) \Big|_{\hat{\underline{B}}=\hat{\underline{B}}_i}$$

The value chosen for g attempts to keep the product $g\underline{L}$ as small as possible without causing instability. The total sum squared error, e , is monitored to ensure this result.

Marquardt's strategy is as follows [12]:

1. Find $e\{\underline{B}_i\}$
2. Let $v > 1$, actually some factor of 2
3. Let g_i denote the value of g when estimating $\hat{\underline{B}}_i$, initially set $g_0 = .01$
4. Compute the error term using estimates of $\hat{\underline{B}}_{i+1}$ derived from g_i and g_i/v . Notationally, find $e\{\hat{\underline{B}}_{i+1}(g_i)\}$ and $e\{\hat{\underline{B}}_{i+1}(g_i/v)\}$.
 - a) If $e\{\hat{\underline{B}}_{i+1}(g_i/v)\} \leq e\{\hat{\underline{B}}_i\}$
 then $g_{i+1} = g_i/v$
 - b) If $e\{\hat{\underline{B}}_{i+1}(g_i/v)\} > e\{\hat{\underline{B}}_i\}$
 and $e\{\hat{\underline{B}}_{i+1}(g_i)\} \leq e\{\hat{\underline{B}}_i\}$
 then $g_{i+1} = g_i$
 - c) If $e\{\hat{\underline{B}}_{i+1}(g_i/v)\} > e\{\hat{\underline{B}}_i\}$
 and $e\{\hat{\underline{B}}_{i+1}(g_i)\} > e\{\hat{\underline{B}}_i\}$
 increase g_i by successive multiples of v until for some smallest w
 $e\{\hat{\underline{B}}_{i+1}(g_i v^w)\} \leq e\{\hat{\underline{B}}_i\}$
 and then $g_{i+1} = g_i v^w$

This guarantees that a feasible neighborhood is obtained and a continuous progression toward the minimum is achieved.

Although Marquardt's algorithm has improved in many respects on the Gauss-Newton and steepest descent techniques, there are still some inherent difficulties. If the initial values of the parameters are not

"close" to the global minimum, then local minimums may be encountered first or no minimum at all is found. A parameter's insensitivity to the sum squared error term may result in a minimum whose estimate of all the parameters save this insensitive parameter is accurate. There is little indication by which this problem could be detected except a priori information or relative sum squared error results for different convergence values. Additive noise to the model changes the error hypersurface. The resulting estimates are unbiased; however, the associated variance of this value defines this estimate only to within a certain distance of the ideal parameter.

The following chapter shall study the accuracy of this technique with respect to a specific model.

CHAPTER III

MARQUARDT NONLINEAR PARAMETER ESTIMATION

The application of Marquardt's nonlinear parameter estimation algorithm has spread to many disciplines since its introduction. In particular, the parameterization of peaks, often modeled as gaussian-shaped signals, in spectroscopy is a perfect candidate for this technique. Marquardt can closely estimate the amplitude, mean, and standard deviation of each gaussian in a summed gaussian expression. These estimates completely define the entire noiseless data set. In this discussion, the summed gaussian data set will consist of two gaussian signals dependent on a sampled abscissa. The independent variable may be atomic mass units, electron volts, frequency, or time depending on the measurement procedure [1]. Consequently, the model reviewed will not explicitly define the units used; rather, it will deal with incremental step sizes that can be defined by the user at a later time.

The stochastic processes discussed in the previous chapter, relating a gaussian shaped distribution to noise values dependent on time are still applicable. The exception in this case is that the modeled gaussian signal now represents the dispersion of incremental steps whose dependence on units is determined by the measurement process.

With this in mind, the model of the process becomes

$$m_i = a_1 \exp \left[-\frac{(n_i T - \mu_1)^2}{2\delta_1^2} \right] + a_2 \exp \left[-\frac{(n_i T - \mu_2)^2}{2\delta_2^2} \right]$$

where a_1 and a_2 are the respective amplitudes of the summed gaussians, μ_1 and μ_2 are their respective means, and δ_1 and δ_2 are their respective standard deviations. These gaussians are sampled at integer multiples, n_i , of the step size, T . The batch Marquardt technique will use a rich data set of five hundred points to describe the measurements, and the step size of these measurements will be set at .01 for each increment from .01 to 5.0.

Convergence to an estimate will be determined when any one of the following three criteria has been satisfied. The first test of convergence is satisfied, if on any two successive iterations, the parameter estimates agree component-wise for 5-significant figures. The second criterion tests the summed squared error on successive iterations; convergence is satisfied if the difference is zero. The final test checks the euclidean norm of the gradient. When this norm is zero the third criterion has been satisfied [13].

Consider first a noiseless data set illustrated in Figure 3-1. The summed gaussians are quite distinct and well separated. Their ideal amplitudes have a magnitude of 1.0 and their standard deviations are defined to be .5. Their respective means are at 1.0 and 4.0, separated by six standard deviations. This wide separation signifies that each gaussian hardly effects the other, or that their respective magnitudes are relatively unbiased. By initializing the parameters in Marquardt as far as $\pm 50\%$ of the true amplitudes' values, $\pm 15\%$ of the

Figure 3-1. Gaussians of equal dispersion and widely separated are summed to form a noiseless data set.

true means' values and $\pm 40\%$ of the true standard deviations' values individually and combinatorially, it is found that convergence is obtained in all cases. The estimated parameters agree with the ideal parameters to five significant digits and the sum of the squares of residuals, e , is of order 10^{-13} . This error hyperspace of six dimensions does not appear to have any local minima closely surrounding the global minimum, assuring the user of convergence even when the initialized parameters are fairly "far" from their true values.

Similar results were found for the noiseless measurements of Figure 3-2 and 3-3. Again, the peaks are well separated and each gaussian contributes little to biasing the other gaussian. The true parameters of Figure 3-2 are the same as in Figure 3-1, except the amplitudes are now different. Convergence to five significant figures of the true values was obtained with an e of magnitude 10^{-13} . These results were obtained with initialization values of amplitude, mean and standard deviation varying by $\pm 50\%$, $\pm 15\%$, and $\pm 40\%$, respectively.

In Figure 3-3, the ideal parameters are once more the same as in Figure 3-1, with the exception of the standard deviation. This change causes the peaks to be not as clearly separated in relation to the preceeding two figures; however, the effect of biasing is not seen significantly in the resulting estimates. The estimates again agree to five significant digits with the ideal parameters and e is of 10^{-13} in magnitude. Initialization values varied in the same proportion as previously noted.

It should be evident from just these 3 cases that this process of initiating each parameter progressively further from its true value, independently and in some kind of combination with the other five

Figure 3-2. A noiseless summed gaussian data set for batch Marquardt estimation.

Figure 3-3. A noiseless summed gaussian data set for batch Marquardt estimation.

parameters, to test for local and global minimums, becomes tedious quickly. Especially, when it is noted that the cases just considered are under noiseless conditions and have well separated peaks which are easily detected by the operator. These are trivial cases to estimate, using little a priori knowledge of the process while reasonable initial estimates can be chosen by the user with little effort.

Consider now, a noiseless measurement set, in which the peaks are within half a standard deviation of each other. To the user without information about the process, the measurements almost appear as a single gaussian rather than the sum of two, Figure 3-4. If a single gaussian, a three parameter model, was used to fit these measurements, Marquardt returns an amplitude of about 2.0, a mean of approximately 2.5, and a standard deviation of around 1.0. The error for this fit is of 10^{-4} order, which when compared to previous error terms is extremely high, reflecting a poor model choice. In cases of noise imposed on the model, this error value will not appear high. In fact, with noisy data sets, estimates returned may or may not closely coincide with the ideal parameter values, while the sum squared error will remain approximately the same.

It is interesting to notice, that when the correct model of two summed gaussians was used, the amplitude estimates appeared rather insensitive about their minimum values. This can be seen by reviewing some of the results obtained from various initialization points (Table 3.1). Initial estimates of the ideal parameters with a $\pm 10\%$ deviation about the mean's true value were given to Marquardt. The results agreed with the ideal parameters to three significant figures, while the error term was in the $10^{-10} - 10^{-11}$ area. Next, just the standard

Figure 3-4. Gaussians of equal dispersion and in close proximity are summed to form a noiseless data set for batch Marquardt estimation (see Table 3.1).

Table 3.1. Batch Marquardt results from the data set in Figure 3-4.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.25	1.0				
Second gaussian	1.0	2.75	1.0				

	Experimental Results						
	Initialization			Results			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.1	2.25	1.0	1.082	2.271	1.002	.8734E-8
Second gaussian	.9	2.75	1.0	.918	2.771	.998	
First gaussian	1.2	2.25	.8	1.120	2.280	1.003	.9071E-8
Second gaussian	.8	2.75	.8	.880	2.780	.997	
First gaussian	1.1	2.25	.8	1.093	2.274	1.002	.5020E-8
Second gaussian	.9	2.75	.8	.906	2.774	.998	
First gaussian	1.0	2.25	.9	.999	2.250	1.000	.1082E-11
Second gaussian	1.0	2.75	.9	1.000	2.750	1.000	
First gaussian	1.0	2.25	.8	.999	2.250	1.000	.1167E-11
Second gaussian	1.0	2.75	.8	1.000	2.750	1.000	
First gaussian	1.0	2.0	1.0	.999	2.250	1.000	.1147E-11
Second gaussian	1.0	3.0	1.0	1.000	2.750	1.000	
First gaussian	1.1	2.0	.8	1.076	2.269	1.002	.3269E-8
Second gaussian	.9	3.0	.8	.924	2.770	.998	
First gaussian	1.0	2.25	.6	.984	2.246	.999	.5381E-10
Second gaussian	1.0	2.75	.6	1.016	2.746	1.000	
Only one gaussian for model	1.5	2.2	2.0	1.939	2.500	1.032	.7077E-4

deviation initialization was varied by $\pm 40\%$ of its true value, while the other parameters were again ideally initialized. The results were similar to the $\pm 10\%$ variation initialization of the amplitude. Finally, the amplitude was the only varying initial estimate by $\pm 20\%$, with the results favoring the initialization rather than the true value of the parameter. The error sum did reflect this insensitivity with terms in the neighborhood of 10^{-8} . However, if various initial estimates were not compared with respect to their resulting error term, it is entirely feasible that the final estimates of the amplitudes could be $\pm 20\%$ of their actual values. This amount would be significant if the resulting parameter estimates were used in the calculation of the concentration of an element in a material.

In Figure 3-5, the amplitudes of the summed gaussians just considered are changed to concentrate on this amplitude insensitivity. The remaining parameters are identical to the previous example, Figure 3-4. In this example the results become confusing as many local minima appear to have moved closer to the global minimum (review Table 3-2). Even when the operator has prior knowledge of four of the six parameters and initializes all but two to their true values; all six may diverge away from the global minimum. The error term will reflect this mismatch, but it can only be noticed in relation to a true fit. The true fit can only be found by guessing the "correct" initial values.

Four more noiseless models are considered using this batch Marquardt approach. Each of these summed gaussian models' means are mediumly separated with respect to the previously considered examples. The basic results are summarized in adjoining tables to the Figures

Figure 3-5. A noiseless summed gaussian data set whose components are unequal in amplitude and in close proximity.

Table 3.2. Batch Marquardt results from the data set in Figure 3-5.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.25	1.0				
Second gaussian	.4	2.75	1.0				

	Experimental Results						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	.7	2.25	1.0	.7474	2.194	.9957	.1875E-7
Second gaussian	.7	2.75	1.0	.6486	2.619	1.011	
First gaussian	.8	2.25	1.0	.8073	2.206	.9966	.1217E-7
Second gaussian	.6	2.75	1.0	.5903	2.646	1.009	
First gaussian	.9	2.25	1.0	.9114	2.228	.9983	.3286E-8
Second gaussian	.5	2.75	1.0	.4881	2.698	1.004	
First gaussian	1.1	2.25	1.0	.9994	2.250	1.000	.2043E-11
Second gaussian	.5	2.75	1.0	.4006	2.750	1.000	
First gaussian	1.5	2.25	1.0	.9458	2.237	.9989	.1367E-8
Second gaussian	1.0	2.75	1.0	.4540	2.717	1.003	
First gaussian	1.0	2.25	.5	.7153	2.187	.9952	.2254E-7
Second gaussian	.4	2.75	.5	.6798	2.605	1.012	
First gaussian	1.0	2.25	1.2	1.020	2.255	1.000	.2507E-9
Second gaussian	.4	2.75	.8	.3797	2.763	.9989	
First gaussian	1.0	2.25	1.2	1.002	2.250	1.000	.2971E-11
Second gaussian	.4	2.75	1.2	.3982	2.751	.9999	
First gaussian	1.0	2.0	1.0	.9961	2.249	.9999	.9454E-11
Second gaussian	.4	3.0	1.0	.4039	2.748	1.000	
First gaussian	1.0	1.5	1.0	.9133	2.517	1.020	.4358E-7
Second gaussian	.4	3.5	1.0	.4730	2.146	.9900	
First gaussian	1.0	1.75	1.0	.9722	2.243	.9994	.3961E-9
Second gaussian	.4	3.25	1.0	.4278	2.733	1.001	
First gaussian	1.0	2.1	1.0	.9883	2.247	.9997	.7523E-10
Second gaussian	.4	2.8	1.0	.4117	2.743	1.001	

Table 3.2 (cont.).

	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.0	2.25	2.0	.5424	2.157	.9920	.4110E-7
Second gaussian	1.0	2.75	2.0	.8466	2.540	1.018	
First gaussian	1.1	2.25	1.5	.9925	2.248	.9998	.3207E-10
Second gaussian	.5	2.75	1.5	.4075	2.745	1.000	
First gaussian	1.1	2.25	.5	.9718	2.243	.9994	.4044E-9
Second gaussian	.5	2.75	.5	.4281	2.733	1.001	
First gaussian	.8	2.25	.5	.8196	2.208	.9967	.1093E-7
Second gaussian	.6	2.75	.5	.5782	2.652	1.008	
First gaussian	.8	2.25	1.2	.8255	2.209	.8968	.1035E-7
Second gaussian	.6	2.75	.8	.5724	2.655	1.008	
First gaussian	1.5	2.25	2.0	.9517	2.238	.9999	.1109E-8
Second gaussian	1.0	2.75	2.0	.4482	2.721	1.002	
First gaussian	1.1	2.1	.9	1.068	2.268	1.002	.3597E-8
Second gaussian	.3	2.8	1.1	.3314	2.797	.9959	
First gaussian	1.4	2.5	1.1	1.155	2.447	1.025	.6552E-8
Second gaussian	.5	2.5	.9	.2224	2.094	.9714	
Only one gaussian for model	2.0	2.5	1.0	1.365	2.390	1.026	.7307E-3

Figure 3-6. Closely located, unequally dispersed gaussians are summed to form a noiseless data set.

Table 3.3. Batch Marquardt estimates of the data set in Figure 3-6.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.25	1.0				
Second gaussian	1.0	2.75	.5				

	<u>Experimental Results</u>						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.5	2.25	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	1.5	2.75	.5	1.0	2.75	.5	
First gaussian	1.5	2.25	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	.5	2.75	.5	1.0	2.75	.5	
First gaussian	2.0	2.25	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	2.0	2.75	.5	1.0	2.75	.5	
First gaussian	3.0	2.25	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	.5	2.75	.5	1.0	2.75	.5	
First gaussian	.5	2.25	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	3.0	2.75	.5	1.0	2.75	.5	
First gaussian	1.0	2.2	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	2.8	.5	1.0	2.75	.5	
First gaussian	1.0	2.0	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	3.0	.5	1.0	2.75	.5	
First gaussian	1.0	1.6	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	3.4	.5	1.0	2.75	.5	
First gaussian	1.0	1.4	1.0	1.841	2.667	.6528	.2923
Second gaussian	1.0	3.6	.5	.4383	1.262	.6260	
First gaussian	1.0	2.25	1.2	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	2.75	.6	1.0	2.75	.5	
First gaussian	1.0	2.25	1.4	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	2.75	.8	1.0	2.75	.5	
First gaussian	1.0	2.25	1.5	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	2.75	1.0	1.0	2.75	.5	

Table 3.3 (cont.).

	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.0	2.25	2.0	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	2.75	1.5	1.0	2.75	.5	
First gaussian	1.0	2.25	.5	1.0	2.25	1.0	.3275E-12
Second gaussian	1.0	2.75	.2	1.0	2.75	.5	
First gaussian	.8	2.25	1.2	1.0	2.25	1.0	.3275E-12
Second gaussian	1.2	2.75	.6	1.0	2.75	.5	
First gaussian	1.5	2.25	1.4	1.0	2.25	1.0	.3275E-12
Second gaussian	1.5	2.75	.8	1.0	2.75	.5	
First gaussian	1.5	2.25	1.4	1.0	2.25	1.0	.3275E-12
Second gaussian	.5	2.75	.8	1.0	2.75	.5	
First gaussian	2.0	2.25	2.0	1.0	2.25	1.0	.3275E-12
Second gaussian	2.0	2.75	1.5	1.0	2.75	.5	
First gaussian	2.0	2.25	1.0	1.0	2.25	1.0	.3275E-12
Second gaussian	2.0	2.75	1.0	1.0	2.75	.5	
First gaussian	1.5	2.2	1.4	1.0	2.25	1.0	.3275E-12
Second gaussian	1.5	2.8	.8	1.0	2.75	.5	
First gaussian	1.5	2.0	1.5	1.0	2.25	1.0	.3350E-12
Second gaussian	.5	3.0	1.0	1.0	2.75	.5	
First gaussian	3.0	1.4	2.0	1.841	2.667	.6528	.2923
Second gaussian	.5	3.6	1.5	.4383	1.262	.6260	
First gaussian	4.0	2.25	1.0	1.732	2.538	.7353	4.257
Second gaussian	.1	2.75	.5	.1065	.2907E7	.244E10	
First gaussian	1.0	1.4	1.0	1.841	2.667	.6528	.2923
Second gaussian	1.0	3.6	.5	.4383	1.262	.6260

Figure 3-7. Gaussians of equal dispersion and mediumly separated are summed to form a noiseless data set.

Table 3.4. Parameters returned by batch Marquardt from Figure 3-7.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.0	1.0				
Second gaussian	1.0	3.0	1.0				

Experimental Results							
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.2	2.0	1.0	1.0	2.0	1.0	.2362E-11
Second gaussian	1.2	3.0	1.0	1.0	3.0	1.0	
First gaussian	1.2	2.0	1.0	1.0	2.0	1.0	.9350E-12
Second gaussian	.8	3.0	1.0	1.0	3.0	1.0	
First gaussian	1.4	2.0	1.0	1.0	2.0	1.0	.1016E-11
Second gaussian	.6	3.0	1.0	1.0	3.0	1.0	
First gaussian	1.6	2.0	1.0	1.0	2.0	1.0	.1074E-11
Second gaussian	.4	3.0	1.0	1.0	3.0	1.0	
First gaussian	.4	2.0	1.0	1.0	2.0	1.0	.1465E-11
Second gaussian	1.6	3.0	1.0	1.0	3.0	1.0	
First gaussian	1.0	2.2	1.0	1.0	2.0	1.0	.3787E-11
Second gaussian	1.0	2.8	1.0	1.0	3.0	1.0	
First gaussian	1.0	2.4	1.0	1.0	2.0	1.0	.2097E-11
Second gaussian	1.0	2.6	1.0	1.0	3.0	1.0	
First gaussian	1.0	1.8	1.0	1.0	2.0	1.0	.2395E-11
Second gaussian	1.0	3.2	1.0	1.0	3.0	1.0	
First gaussian	1.0	1.5	1.0	1.0	2.0	1.0	.1217E-11
Second gaussian	1.0	3.5	1.0	1.0	3.0	1.0	
First gaussian	1.0	2.0	1.2	1.0	2.0	1.0	.9875E-12
Second gaussian	1.0	3.0	.8	1.0	3.0	1.0	
First gaussian	1.0	2.0	.8	1.0	2.0	1.0	.1137E-11
Second gaussian	1.0	3.0	1.2	1.0	3.0	1.0	
First gaussian	1.0	2.0	1.6	1.0	2.0	1.0	.1151E-11
Second gaussian	1.0	3.0	.5	1.0	3.0	1.0	

Table 3.4 (cont.).

	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.4	2.0	.8	1.0	2.0	1.0	.1033E-11
Second gaussian	.6	3.0	1.2	1.0	3.0	1.0	
First gaussian	1.4	2.0	1.6	1.0	2.0	1.0	.1813E-11
Second gaussian	.6	3.0	.5	1.0	3.0	1.0	
First gaussian	1.6	2.0	1.6	1.0	2.0	1.0	.4584E-12
Second gaussian	.4	3.0	.5	1.0	3.0	1.0	
First gaussian	1.4	2.2	.8	1.0	2.0	1.0	.1057E-11
Second gaussian	.6	2.8	1.2	1.0	3.0	1.0	
First gaussian	1.6	1.5	1.6	1.0	2.0	1.0	.8925E-12
Second gaussian	.4	3.5	.5	1.0	3.0	1.0	
First gaussian	.4	1.5	1.6	1.0	2.0	1.0	.1849E-11
Second gaussian	1.6	3.5	.5	1.0	3.0	1.0	
First gaussian	.4	2.2	.5	1.0	2.0	1.0	.1137E-11
Second gaussian	1.6	3.2	1.6	1.0	3.0	1.0	
Only one gaussian for model	1.8	2.5	1.0	1.772	2.5	1.134	.1467E-1

Figure 3-8. Unequally dispersed mediumly separated gaussians form a noiseless data set.

Table 3.5. Batch Marquardt estimations describing Figure 3-8.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.0	1.0				
Second gaussian	1.0	3.0	.5				

	Experimental Results						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.2	2.0	1.0	1.0	2.0	1.0	.3542E-12
Second gaussian	.8	3.0	.5	1.0	3.0	.5	
First gaussian	1.2	2.0	1.0	1.0	2.0	1.0	.3542E-12
Second gaussian	1.2	3.0	.5	1.0	3.0	.5	
First gaussian	2.0	2.0	1.0	1.0	2.0	1.0	.3542E-12
Second gaussian	2.0	3.0	.5	1.0	3.0	.5	
First gaussian	1.0	2.1	1.0	1.0	2.0	1.0	.3542E-12
Second gaussian	1.0	3.1	.5	1.0	3.0	.5	
First gaussian	1.0	2.5	1.0	1.0	2.0	1.0	.3542E-12
Second gaussian	1.0	3.5	.5	1.0	3.0	.5	
First gaussian	1.0	1.5	1.0	1.0	2.0	1.0	.3542E-12
Second gaussian	1.0	2.5	.5	1.0	3.0	.5	
First gaussian	1.0	2.0	1.2	1.0	2.0	1.0	.9219E-12
Second gaussian	1.0	3.0	.3	1.0	3.0	.5	
First gaussian	1.0	2.0	.8	1.0	2.0	1.0	.3542E-12
Second gaussian	1.0	3.0	.7	1.0	3.0	.5	
First gaussian	2.0	2.0	1.2	1.0	2.0	1.0	.3542E-12
Second gaussian	2.0	3.0	.3	1.0	3.0	.5	
First gaussian	1.4	2.0	.8	1.0	2.0	1.0	.3542E-12
Second gaussian	.6	3.0	.7	1.0	3.0	.5	
First gaussian	2.0	2.0	.8	1.0	2.0	1.0	.9001E-12
Second gaussian	2.0	3.0	.7	1.0	3.0	.5	
First gaussian	2.0	1.5	.8	1.0	2.0	1.0	.3542E-12
Second gaussian	2.0	2.5	.7	1.0	3.0	.5

Figure 3-9. Gaussians unequal in amplitude and mediumly separated form a noiseless data set.

Table 3.6. Parameter estimates from batch Marquardt for Figure 3-9.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.0	1.0				
Second gaussian	.4	3.0	1.0				

	Experimental Results						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.2	2.0	1.0	1.0	2.0	1.0	.6792E-13
Second gaussian	.7	3.0	1.0	.4	3.0	1.0	
First gaussian	1.4	2.0	1.0	1.0	2.0	1.0	.8200E-12
Second gaussian	.2	3.0	1.0	.4	3.0	1.0	
First gaussian	.8	2.0	1.0	1.0	2.0	1.0	.8224E-12
Second gaussian	.7	3.0	1.0	.4	3.0	1.0	
First gaussian	1.6	2.0	1.0	1.0	2.0	1.0	.7780E-12
Second gaussian	.2	3.0	1.0	.4	3.0	1.0	
First gaussian	1.8	2.0	1.0	1.0	2.0	1.0	.2490E-12
Second gaussian	1.0	3.0	1.0	.4	3.0	1.0	
First gaussian	1.0	2.2	1.0	1.0	2.0	1.0	.5212E-12
Second gaussian	.4	2.8	1.0	.4	3.0	1.0	
First gaussian	1.0	2.4	1.0	1.0	2.0	1.0	.6028E-12
Second gaussian	.4	2.6	1.0	.4	3.0	1.0	
First gaussian	1.0	1.5	1.0	1.0	2.0	1.0	.1071E-11
Second gaussian	.4	3.5	1.0	.4	3.0	1.0	
First gaussian	1.0	2.0	1.2	1.0	2.0	1.0	.5793E-12
Second gaussian	.4	3.0	.8	.4	3.0	1.0	
First gaussian	1.0	2.0	.8	1.0	2.0	1.0	.6473E-12
Second gaussian	.4	3.0	1.2	.4	3.0	1.0	
First gaussian	1.0	2.0	1.4	2.532	2.399	1.050	No
Second gaussian	.4	3.0	.6	-1.286	2.527	.9800	Convergence
First gaussian	1.2	2.0	1.2	1.0	2.0	1.0	.7290E-12
Second gaussian	.7	3.0	.8	.4	3.0	1.0	

Table 3.6 (cont.).

	Initilization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.8	2.0	1.2	1.0	2.0	1.0	.7184E-12
Second gaussian	.2	3.0	.8	.4	3.0	1.0	
First gaussian	1.2	2.2	1.2	1.0	2.0	1.0	.6537E-12
Second gaussian	.7	2.8	.8	.4	3.0	1.0	
First gaussian	1.8	2.2	1.2	2.384	2.395	1.053	No
Second gaussian	.2	2.8	.8	-1.138	2.534	.9752	Convergence
First gaussian	1.8	2.4	.8	1.0	2.0	1.0	.6894E-12
Second gaussian	.2	2.6	1.2	.4	3.0	1.0	
First gaussian	1.8	1.0	.8	.01825	12.86	-.9723	.04287
Second gaussian	1.0	4.0	1.2	1.271	2.263	1.104	
Only one gaussian for model	1.4	2.4	1.0	1.271	2.263	1.104	.04287

(Figures 3-6, 3-7, 3-8, and 3-9 with Tables 3.3, 3.4, 3.5, and 3.6, respectively).

Consider now white gaussian noise added to the first three cases considered (Figures 3-1, 3-2 and 3-3) of widely separated gaussians. The standard deviation of the noise shall be chosen approximately 20dB down from the minimum amplitude; consequently, additive white gaussian noise of standard deviation .05 is chosen to be added to each of these three cases. Figures 3-10, 3-11, and 3-12 show the summed gaussians with the noise added. From the same initial estimates used in the noiseless cases, the final estimated parameters are again very close to the ideal values. Their worst deviation from the ideal parameters was less than 1%. The sum of the squares of residuals returned was 1.146, which actually agrees with what one would expect. To show this, assume worst case that each data point had noise of value $\pm .05$ added to the original value. If the ideal parameters are estimated, the error term now represents the sum of the squares of the noise values, $\sum (.05)^2$ over the entire data set. For five hundred points, this sum is 1.157. If noise of about 90dB down from the minimum amplitude is considered, the sum of the squares of the noise would still be on the order of 10^{-5} . In recalling the error values obtained for the gaussians separated by no more than one-half of a standard deviation (Figure 3-4 and 3-5), their error magnitudes were around 10^{-7} to 10^{-8} when the final estimates were poor. Now considering noise levels as low as 90dB down from the minimum amplitude, giving the appearance of an almost ideal gaussian, the sum squared error will immediately be as high as 10^{-5} and the estimates may or may not closely approximate the true parameters. This additive noise can easily mask

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

Figure 3-10. Data set from Figure 3-1 with additive noise of $\mu = 0$ and $\delta = 0.05$.

Figure 3-11. Data set from Figure 3-2 with additive noise of $\mu = 0$ and $\delta = -0.05$. ---

Figure 3-12. Data set from Figure 3-3 with additive noise of $\mu = 0$ and $\delta = 0.05$.

out the differences seen before between an ideal fit with an associated error term of 10^{-13} and a poor fit with its respective error term of 10^{-7} .

Under additive noise conditions the results may appear good, although as has been noted, the amplitudes may be as far off as 20%, or the location of the means far from the true location.

When white gaussian noise of .05 standard deviation was added to the summed gaussians, which lay in close proximity, (Figures 3-13 and 3-14 from Figures 3-4 and 3-5, respectively) Marquardt could not discern the existence of two equal amplitude, equal dispersion gaussians. Instead, for Figure 3-10, with all converging initial estimates, it found one gaussian specified by an amplitude, mean, and standard deviation of approximately 1.9, 2.5, and 1.0. The other gaussian was identified by an amplitude of 0.01, a varying mean, and a standard deviation of 2.0. Similar results were obtained for Figure 3-11. The final estimates for one gaussian's amplitude, mean, and standard deviation were approximately 1.4, 2.4, and 1.0, while the other gaussian was shifted about several means with an amplitude of about .01 and a standard deviation of 2.5. Marquardt attributed almost the entire contribution to the data set from only one gaussian, instead of the two from which the data was actually comprised.

In reviewing the final cases of Figures 3-6 to 3-9 with additive noise (Figures 3-15 to 3-18), it is found that the noise in two out of the three cases has interfered with the good results obtained in the noiseless cases. These examples show the limitations of Marquardt. To obtain proper convergence under these noise conditions, separations of greater than one standard deviation must exist between gaussians.

Figure 13. Data set of Figure 3-4 with additive noise of $\mu = 0$ and $\delta = 0.05$.

Figure 3-14. Data set of Figure 3-5 with additive noise of $\mu = 0$ and $\delta = 0.05$.

Figure 3-15. Data set of Figure 3-6 with additive noise of $\mu = 0$ and $\sigma = 0.05$.

Table 3.7. Batch Marquardt estimates of noisy* data set; Figure 3-15.

	Ideal Parameters						
	Amplitude			Mean		Standard Deviation	
First gaussian	1.0			2.25		1.0	
Second gaussian	1.0			2.75		.5	

	Experimental Results						
	Initilization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.2	2.25	1.0	.9443	2.226	1.014	1.143
Second gaussian	1.2	2.75	1.0	1.045	2.744	.5201	
First gaussian	1.5	2.25	1.0	.9443	2.226	1.014	1.143
Second gaussian	1.5	2.75	.5	1.045	2.744	.5201	
First gaussian	.5	2.25	1.0	.9444	2.227	1.014	1.014
Second gaussian	3.0	2.75	.5	1.045	2.744	.5201	
First gaussian	4.0	2.25	1.0	1.823	2.671	.6562	1.386
Second gaussian	.1	2.75	.5	.4284	1.271	.6616	
First gaussian	1.0	2.2	1.0	.9444	2.227	1.014	1.143
Second gaussian	1.0	2.8	.5	1.045	1.014	.5201	
First gaussian	1.0	1.8	1.0	.9444	2.227	1.014	1.143
Second gaussian	1.0	3.2	.5	1.045	2.744	.5201	
First gaussian	1.0	1.4	1.0	1.823	2.671	.6562	1.386
Second gaussian	1.0	3.6	.5	.4284	1.271	.6615	
First gaussian	1.0	2.25	1.2	.9444	2.227	1.014	1.143
Second gaussian	1.0	2.75	.6	1.045	2.744	.5201	
First gaussian	1.0	2.25	.5	.9443	2.226	1.014	1.143
Second gaussian	1.0	2.75	.2	1.045	2.744	.5201	
First gaussian	1.0	2.25	.05	.9444	2.227	1.014	1.143
Second gaussian	1.0	2.75	.01	1.045	2.744	.5201	
First gaussian	.8	2.25	1.2	.9444	2.227	1.014	1.143
Second gaussian	1.2	2.75	.6	1.045	2.744	.5201	
First gaussian	2.0	2.25	.2	.9445	2.227	1.014	1.143
Second gaussian	2.0	2.75	.05	1.045	2.744	.5201	

Table 3.7 (cont.).

	Initilization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.5	2.25	1.0	.9445	2.227	1.014	1.143
Second gaussian	1.5	2.75	1.0	1.045	2.744	.5201	
First gaussian	1.5	2.2	1.4	.9443	2.226	1.014	1.143
Second gaussian	1.5	2.8	.8	1.045	2.744	.5201	
First gaussian	3.0	1.8	2.0	1.823	2.671	.6562	1.386
Second gaussian	.5	3.2	1.5	.4285	1.271	.6610	

*Additive noise mean 0.0 standard deviation 0.05.

Figure 3-16. Noise of $\mu = 0$ and $\delta = 0.05$ added to data set of Figure 3-7.

Table 3.8. Batch Marquardt estimates of noise* data sets, Figure 3-16.

	Ideal Parameters						
	Amplitude		Mean		Standard Deviation		
First gaussian	1.0		2.0		1.0		
Second gaussian	1.0		3.0		1.0		

	Experimental Results						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.2	2.0	1.0	1.761	2.487	1.129	1.144
Second gaussian	1.2	3.0	1.0	.03968	3.622	.4267	
First gaussian	1.2	2.0	1.0	1.761	2.487	1.129	1.144
Second gaussian	.8	3.0	1.0	.03967	3.622	.4266	
First gaussian	1.1	2.0	1.0	1.761	2.487	1.129	1.144
Second gaussian	1.1	3.0	1.0	.03964	3.622	.4264	
First gaussian	1.0	2.2	1.0	1.761	2.487	1.129	1.144
Second gaussian	1.0	2.8	1.0	.03964	3.622	.4267	
First gaussian	1.0	1.5	1.2	1.761	2.487	1.129	1.144
Second gaussian	1.0	3.5	.8	.03968	3.622	.4264	
First gaussian	1.0	2.0	1.2	1.761	2.487	1.129	1.144
Second gaussian	1.0	3.0	.8	.03968	3.622	.4264	
First gaussian	1.4	2.0	.8	1.761	2.487	1.129	1.144
Second gaussian	.6	3.0	1.2	.03968	3.622	.4266	
First gaussian	1.1	2.1	1.1	1.761	2.489	1.129	1.144
Second gaussian	1.1	3.1	.9	.03968	3.622	.4266	
First gaussian	.4	2.2	.5	1.761	2.487	1.129	1.144
Second gaussian	1.6	3.2	1.6	.03968	3.623	.4266	
Only one gaussian for model	1.8	2.5	1.0	1.769	2.502	1.136	1.167

*Additive noise mean 0.0 standard deviation 0.05.

Figure 3-17. Noise of $\mu = 0$ and $\delta = 0.05$ added to data set of Figure 3-8.

Table 3.9. Batch Marquardt estimates of noisy* data set, Figure 3-17.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.0	1.0				
Second gaussian	1.0	3.0	.5				

	Experimental Results						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.2	2.0	1.0	.9851	1.982	.9976	1.1493
Second gaussian	.8	3.0	.5	1.016	3.002	.5091	
First gaussian	2.0	2.0	1.0	.9851	1.982	.9976	1.149
Second gaussian	2.0	3.0	.5	1.016	3.002	.5094	
First gaussian	1.0	1.5	1.0	.9851	1.982	.9976	1.149
Second gaussian	1.0	2.5	.5	1.016	3.002	.5094	
First gaussian	1.0	2.0	.8	.9851	1.982	.9976	1.149
Second gaussian	1.0	3.0	.7	1.016	3.002	.5094	
First gaussian	2.0	2.0	.8	.9851	1.982	.9975	1.149
Second gaussian	2.0	3.0	.7	1.016	3.002	.5095	
First gaussian	2.0	1.5	.8	.9851	1.982	.9976	1.149
Second gaussian	2.0	2.5	.7	1.016	3.002	.5094	

*Additive noise mean 0.0 standard deviation 0.05.

Figure 3-18. Data set of Figure 3-9 with additive noise of $\delta = 0$ and $\delta = 0.05$.

Table 3.10. Batch Marquardt estimates of noisy* data set, Figure 3-18.

	Ideal Parameters						
	Amplitude	Mean	Standard Deviation				
First gaussian	1.0	2.0	1.0				
Second gaussian	.4	3.0	1.0				

	Experimental Results						
	Initialization			Results Returned			
	Amplitude	Mean	Standard Deviation	Amplitude	Mean	Standard Deviation	Error (e)
First gaussian	1.0	2.0	1.0	1.263	2.218	1.068	1.148
Second gaussian	.4	3.0	1.0	.07246	3.785	.6982	
First gaussian	1.2	2.0	1.0	1.263	2.218	1.068	1.148
Second gaussian	.7	3.0	1.0	.07260	3.785	.6985	
First gaussian	1.0	2.2	1.0	1.263	2.218	1.068	1.148
Second gaussian	.4	2.8	1.0	.07281	3.784	.6992	
First gaussian	1.0	2.0	1.2	1.263	2.218	1.068	1.148
Second gaussian	.4	3.0	.8	.07271	3.784	.6990	
First gaussian	1.2	2.0	1.2	1.263	2.218	1.068	1.148
Second gaussian	.7	3.0	.8	.07279	3.784	.6991	
First gaussian	1.1	2.0	1.1	1.263	2.218	1.068	1.148
Second gaussian	.5	3.0	.9	.07279	3.784	.6991	
First gaussian	1.8	1.0	.8	1.263	2.218	1.068	1.148
Second gaussian	1.0	4.0	1.2	.07298	3.783	.7003	

*Additive noise mean 0.0 standard deviation 0.05.

Marquardt's convergence ability depends almost exclusively on the hypersurface evolved from the sum of the squares of the errors for each iteration's estimate. It is clear from Figure 3-19, that local minima surround the global minimum. Therefore, the initializing point is crucial to true convergence. Figure 3-19 represents a two parameter problem with the minimum defined by a noiseless gaussian of amplitude one, a mean of two and a standard deviation of one. The varying parameters are the mean and the standard deviation whose range is .8 to 6.6 by .2 and .4 to 6.4 by .2, respectively. With the addition of -20dB white gaussian noise, zero mean and standard deviation of .10, to this noiseless gaussian, the contours of the noisy error space look identical to the noiseless case. Looking at the difference between these two cases, Figure 3-20, it is found that the actual values of the error surface have changed. On closer inspection, a point by point review of the error term, it is found that in this case the minimum of the noisy error space is at a mean of 2.0 with a standard deviation of 1.01, and not at 2.0 and 1.0 as in the noiseless case. The question arises, has the noise biased the minimum parameter values of the error hypersurface?

To discuss this possibility allow m_i to represent the noiseless gaussian model of true mean, μ_T , and of true dispersion, δ_T , where i represents the i^{th} point in a data set of n points.

$$m_i = \exp \left[-\frac{(i_T - \mu_T)^2}{2\delta_T^2} \right]$$

To define an error surface let $\hat{m}_i$ signify the changing gaussian whose parameters, μ_k and δ_j , will vary over a specified range at

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

Figure 3-19. Two parameter error surface of gaussian model.

Figure 3-20. Difference error surface between noisy and noiseless gaussian models.

discrete intervals, and k and j designate values of μ and δ surrounding and containing μ_T and δ_T .

$$\hat{m}_i = \exp \left[-\frac{(\mu_T - \mu_k)^2}{2\delta_j^2} \right]$$

The noiseless error space can then be defined as

$$e_I = \sum_{i=1}^n (m_i - \hat{m}_i)^2$$

Consider noise added to the ideal gaussian, changing the hypersurface to

$$e_N = \sum_{i=1}^n (m_i + n_i - \hat{m}_i)^2$$

which can be reformulated as

$$e_N = e_I + 2 \sum_{i=1}^n (m_i - \hat{m}_i)n_i + \sum_{i=1}^n n_i^2$$

Then

$$E[e_N] = e_I + n\delta^2$$

since

$$E[n_i] = 0 \text{ and } E[n_i^2] = \delta^2$$

The expected values of the noisy error surface will all be shifted by a constant in a positive direction from the noiseless error surface. The expected minimum point is maintained at the true parameter values, μ_T and δ_T . This unbiased estimate will have an associated variance of complex form, which on any one measurement set may result in a minimum not associated with the ideal parameter values. This occurred in the previous example.

From these results, it has been established that the noisy error surface's expected minimum occurs at the unbiased values of the true parameters while each point on this hypersurface has a variance describing the maximum deviation of that value from its expected value. To ensure that the technique by which this error surface is searched does not induce bias on the estimated parameters, review the expected values of these estimated parameters obtained by the Marquardt algorithm.

Redefine the iteration expression for Marquardt given in Chapter II to:

$$\hat{\underline{B}}_{k+1} = \hat{\underline{B}}_k - [\underline{H}_k + g\underline{L}_k]^{-1} \underline{J}_k^T (\hat{\underline{M}}_k - \underline{Y})$$

where $\underline{H}$ is the approximated hessian

$$\underline{H}_k = \underline{V}_B^{\wedge}(\hat{\underline{M}}_k^T) \underline{V}_B^{\wedge}(\hat{\underline{M}}_k^T)^T$$

$\underline{J}$ is the jacobian of the estimated model

$$\underline{J}_k = \underline{V}_B^{\wedge}(\hat{\underline{M}}_k)$$

$\underline{L}_k$ is the diagonal of $\underline{H}_k$, and g is the Marquardt parameter. Let $\underline{W}_k$ be an (n,m) matrix represented by

$$\underline{W}_k = (\underline{H}_k + g\underline{L}_k)^{-1} \underline{J}_k^T$$

The (n,m) is defined from the n parameters describing the model and the m measurements taken from the process.

Now

$$E[\hat{\underline{B}}_{k+1}] = E[\hat{\underline{B}}_k] - E[\underline{W}_k(\hat{\underline{M}}_k - \underline{Y})]$$

Since the data can be represented as the sum of an ideal model and additive white gaussian noise, substitute

$$\underline{Y} = \underline{M} + \underline{N}$$

to obtain

$$\begin{aligned} E[\hat{\underline{B}}_{k+1}] &= E[\hat{\underline{B}}_k] - E[\underline{W}_k(\hat{\underline{M}}_k - \underline{M} - N)] \\ &= E[\hat{\underline{B}}_k] - E[\underline{W}_k(\hat{\underline{M}}_k - \underline{M})] + E[\underline{W}_k N] \end{aligned}$$

At this point, it is necessary to know the stochastics of $\underline{W}_k$. In Marquardt, elements of the jacobian are found by the following equations:

$$\begin{aligned} j_{ik} &= \frac{\text{Residuals}(\hat{\underline{B}}_k + 10^{-3.5}\hat{\underline{B}}_k) - \text{Residuals}(\hat{\underline{B}}_k)}{10^{-3.5}\hat{\underline{B}}_k} \\ &= \frac{[y_i - \hat{m}_i(\hat{\underline{B}}_k + 10^{-3.5}\hat{\underline{B}}_k)] - [y_i - \hat{m}_i(\hat{\underline{B}}_k)]}{10^{-3.5}\hat{\underline{B}}_k} \end{aligned}$$

where i denotes the i^{th} measurement and k is the k^{th} iteration. The elements of the jacobian depend only on a numerical estimate of the first partial of the model. The noisy data cancels itself.

The approximated hessian, $\underline{H}_k$, is the product of $\underline{J}_k^T \underline{J}_k$ making it independent of noise, as is $\underline{L}_k$. Consequently, the expected value of the parameter estimates is

$$E[\hat{\underline{B}}_{k+1}] = E[\hat{\underline{B}}_k] - \underline{W}_k(\hat{\underline{M}}_k - \underline{M}) + \underline{W}_k E[N]$$

The noise vector has a mean of zero, reducing this equation to

$$E[\hat{\underline{B}}_{k+1}] = E[\hat{\underline{B}}_k] - \underline{W}_k(\hat{\underline{M}}_k - \underline{M})$$

or from the recursion

$$E[\hat{\underline{B}}_{k+1}] = \hat{\underline{B}}_0 - \underline{W}_0(\hat{\underline{M}}_0 - \underline{M}) - \underline{W}_1(\hat{\underline{M}}_1 - \underline{M}) - \dots - \underline{W}_k(\hat{\underline{M}}_k - \underline{M})$$

which is an unbiased estimate, not dependent on noise. Marquardt proves this series converges [12].

$$\text{Var}[\hat{\underline{B}}_{k+1}] = -E[\underline{N}^T \underline{W}_k^T \underline{W}_k \underline{N}]$$

Both the search for the true parameters and the error space, over which the search is conducted, provide random variables of unbiased means with variances related to the noise variance.

The preceding noisy examples accentuated Marquardt's limitations. Under less severe noise, or a separation of greater than one standard deviation the estimates returned are normally close to the ideal parameters even with noise variability introduced. However, when the parameters returned are far from their actual values, there are only special cases where transformations can apply constraints to resolve the convergence problem [15]. To gain confidence in the returned parameters, many runs from different initialization points must be made to ensure that the global minimum is returned. This is time consuming, at an average of twenty minutes a run for six parameters on a PDP-11:60.

In addition, problems with computer memory size can arise. Not considering the storage required to contain the basic Marquardt algorithm, the entire data set that is being estimated must be held in memory (m points). In addition, the first partial of the model with respect to each of the n parameters leads to an $m \times n$ storage array. Symmetric storage conserves some space in holding the hessian in triangular form. Each parameter estimated adds n new elements to the hessian matrix and $5 + n$ elements to a basic work array. The work array starts with $2m + 6$ elements for a one parameter model. The larger the data set and the greater the number of parameters to be estimated in a model creates rapidly increasing storage and time difficulties.

To minimize these problems, as well as constrain the degrees of freedom in initialization of multiple parameters, consider separating

the gaussians from one another so that a three parameter problem may be isolated. In many instances, as have been shown, the data set containing two gaussians may not represent any one gaussian over a long period of incremental steps. To isolate one of the gaussians, limited windowing of the data set is necessary. The window of data may not contain the entire gaussian sought, but it may hopefully contain enough information to describe the entire curve.

In limiting the number of parameters and the size of the window to a fractional amount of the initial problem, necessary storage area and calculation time is drastically reduced.

With minimal noise levels present on the data set, this three parameter model may be reduced to a two parameter problem through normalization. The linear amplitude term may be separated from the nonlinear mean and standard deviation terms. This concept of separable nonlinear least squares problems is illustrated by an algorithm originally developed by Golub and Pereyra [16], and later examined by Corradi and Stefanini [17]. This algorithm employs a batch mode method and requires all of the data points in the measurement set. Since windowed data rather than batch mode measurements were considered, this particular algorithm was not implemented for nonlinear least squares estimation.

The window of data supplied to Marquardt is chosen by the user to correctly represent a portion of one gaussian curve. The length of this data set will be defined as m , where m is much less than the total number of measurements in the data set. The window will begin at the 1^{th} increment. Each measurement in this window can then be represented as:

$$y_{\ell} = a_T \exp[-\frac{1}{2}(\ell T - \mu_T)^2/\delta_T^2]$$

$$y_{\ell+1} = a_T \exp[-\frac{1}{2}\{(\ell + 1)T - \mu_T\}^2/\delta_T^2]$$

$$y_{\ell+2} = a_T \exp[-\frac{1}{2}\{(\ell + 2)T - \mu_T\}^2/\delta_T^2]$$

$$\vdots$$

$$y_{\ell+m-1} = a_T \exp[-\frac{1}{2}\{(\ell + m - 1)T - \mu_T\}^2/\delta_T^2]$$

where there is little to no additive noise.

It follows that if each element in this window is divided by the 1th increment, y_{ℓ} , then amplitude will no longer be a parameter. This operation transforms the previous measurements to:

$$y_{\ell} = 1$$

$$y_{\ell+1} = \exp[-\frac{1}{2}\{2\ell T^2 + T^2 - 2T\mu_T\}/\delta_T^2]$$

$$y_{\ell+2} = \exp[-\frac{1}{2}\{4\ell T^2 + 4T^2 - 4T\mu_T\}/\delta_T^2]$$

$$\vdots$$

$$y_{\ell+m-1} = \exp[-\frac{1}{2}\{2(m-1)\ell T^2 + (m-1)^2 T^2 - 2(m-1)T\mu_T\}/\delta_T^2]$$

From this normalized window of data Marquardt can estimate for the mean and the standard deviation without considering amplitude. This normalization may only be used on extremely clean signals. The advantage of this separation is that each time linear, nonlinear, and summed terms can be considered independently.

To gain confidence in Marquardt's ability to characterize a gaussian's nonlinear terms from a normalized window, review the following noiseless cases. In the first example, Figure 3-21, the mean is 2.0, and the standard deviation is 1.0. In the second case, Figure 3-22, the

Figure 3-21. Percent deviation from ideal parameters in initialization of normalized ten point window Marquardt. Convergence to five significant digits to each ideal parameter describing the noiseless data is consistently achieved.

Figure 3-22. Another example of the maximum percent deviation from the ideal parameters in initializing a normalized ten point window Marquardt algorithm to obtain close estimates of the ideal parameters.

mean is 3.0, and the standard deviation is .5. In both figures, windows of ten points were taken throughout the data set. Convergence to five significant places was obtained from each window with error terms returned of magnitude 10^{-13} . The percentage figures seen in each window exhibit how far each initialization guess could deviate from the actual parameters and still obtain the global minimum. As the mean is approached from the left with small windows, the negative percent variance from the true mean to initialize the search becomes more stringent. Similarly, this phenomena is repeated for the positive percent deviation when approaching the mean from the right. The standard deviation does not experience this limitation.

Figures 3-23 and 3-24 are additive noise compliments of the noiseless cases just considered. In these examples, ideal convergence was not always achieved, thus the percentage figures denote Marquardt's returned estimates' deviation from the ideal parameters. Various window sizes were used to search for the best size.

It is obvious that this normalized window approach gives excellent results using a model of a single noiseless gaussian, while limited success is encountered when small noise is added to these models. Under "large noise" conditions, approximately 20dB down rather than 50dB down from the peak amplitude, the complete normalization technique breaks down. The results returned from Marquardt under these conditions do not resemble the ideal parameters. The resolution is not improved with larger windows as this inaccuracy stems from the normalization rather than the randomness of the additive noise throughout the window.

Figure 3-23. Percent deviation from ideal parameters in final estimates returned by various sizes of normalized windows by Marquardt.

Figure 3-24. Another example of the estimation ability of the normalized window Marquardt.

It is readily apparent that incorrect estimates will be returned by this technique if the windowed data set does not closely represent a portion of a single gaussian. In other words, referring to Figures 3-4, 3-5, and 3-7, any window taken from these summed data sets will not adequately represent either one of the gaussians. The bias added from the close proximity of the other gaussian makes it impossible to isolate just one of the gaussians. The normalized window Marquardt will return the same parameters found when giving batch Marquardt a one gaussian model for these summed gaussian data sets. Figure 3-25 explicitly demonstrates how close two summed gaussians can appear as a single gaussian.

The success of this normalized windowed Marquardt estimation of the nonlinear terms of a gaussian is a summed gaussian data set is constrained by noise and a representative window.

Noting these shortcomings, this technique still enables the nonlinear terms to be separated from the linear parameters. With the nonlinear terms defined, one can proceed to filter out each gaussian in a summed gaussian data set without knowledge of the amplitude of either gaussian. This filtering procedure motivates the redefinition of the Kalman filter previously defined.

Figure 3-25. Summed versus single gaussian data set.

CHAPTER IV

RECURSIVE NONLINEAR KALMAN FORMULATION

In the previous discussion of Kalman, it was assumed that the process modeled was in the form of an ordinary or partial differential equation, which is inherently linear with respect to the states. A state transition matrix was defined to relate one state to the next, making it possible to forecast estimates and their covariances. The recursion formula used these forecasts in conjunction with incoming data to estimate the state in a recursive least squares formulation.

To apply the Kalman filter to the nonlinear function of a gaussian whose independent variable is not necessarily time, a "transition" matrix must be found to define the change from one increment to another increment of the independent variable. This transition matrix must still satisfy the same properties established by the state transition matrix (equation 2-4) to fit the Kalman formulation already developed in Chapter II.

To find this transition matrix, look at a window of m points taken from a noiseless gaussian. Begin at the ℓ^{th} increment.

$$y[\ell T] = a \exp\left[-\frac{1}{2}(\ell T - \mu)^2/\delta^2\right]$$

$$\begin{aligned} y[(\ell + 1)T] &= a \exp\left[-\frac{1}{2}[(\ell + 1)T - \mu]^2/\delta^2\right] \\ &= y[\ell T] \exp\left[-\frac{1}{2}(2\ell T^2 + \ell^2 T^2 - 2T\mu)/\delta^2\right] \end{aligned}$$

$$\begin{aligned}
y[(\ell + 2)T] &= a \exp\left[-\frac{1}{2}\{(\ell + 2)T - \mu\}^2/\delta^2\right] \\
&= y[(\ell + 1)T] \exp\left[-\frac{1}{2}(2\ell T^2 + \ell^2 T^2 - 2T\mu)/\delta^2\right] \\
&= y[\ell T] \exp\left[-\frac{1}{2}(4\ell T^2 + 4T^2 - 4T\mu)/\delta^2\right] \\
&\vdots \\
y[(\ell + m - 1)T] &= a \exp\left[-\frac{1}{2}\{(\ell + m - 1)T - \mu\}^2/\delta^2\right] \\
&= y[(\ell + m - 2)T] \exp\left[-\frac{1}{2}(2\ell T^2 + \ell^2 T^2 - 2T\mu)/\delta^2\right] \\
&= y[\ell T] \exp\left[-\frac{1}{2}\frac{2\ell(m-1)T^2 + (m-1)^2 T^2 - 2(m-1)T\mu}{\delta^2}\right]
\end{aligned}$$

It is obvious that each data point may be expressed in terms of a previous data point. Consequently, the transition matrix, in this case (1,1), a single element, will be defined as [18]

$$\Phi_{\ell}[k] = \exp\left[-\frac{1}{2}(2\ell kT^2 + k^2 T^2 - 2kT\mu)/\delta^2\right]$$

If k is positive the transition matrix is placing measurements at $\ell + k$ in terms of the ℓ^{th} increment data value. If k is negative then the $\ell + k$ increment, which is less than ℓ , is placed in terms of the ℓ^{th} increments value. Thus, k is an element of positive or negative integers, while ℓ must be a positive integer.

For the recursion formula, the transition matrix will relate each point in the window to the ℓ^{th} point. The Φ matrix is dependent on where in the entire data set the window is taken, ℓ , and how far away the point is from ℓ .

Check now, the state transition properties with respect to this transition matrix just defined.

$$1) \quad \Phi_{\ell}[0] = 1 \quad \text{which is the identity matrix of one element}$$

$$\begin{aligned}
2) \quad \Phi_{\ell}[k]^{-1} &= \exp[-\frac{1}{2}(2\ell kT^2 + k^2T^2 - 2kT_{\mu})/\delta^2]^{-1} \\
&= \exp[-\frac{1}{2}(-2\ell kT^2 - k^2T^2 + 2kT_{\mu})/\delta^2] \\
&= \exp[-\frac{1}{2}\{2(\ell + k)(-k)T^2 + k^2T^2 + 2kT_{\mu}\}/\delta^2] \\
&= \Phi_{\ell+k}[-k]
\end{aligned}$$

$$\begin{aligned}
3) \quad \Phi_{\ell}[n + r] &= \exp[-\frac{1}{2}\{2\ell(n + r)T^2 + (n + r)^2T^2 - 2(n+r)T_{\mu}\}/\delta^2] \\
&= \exp[-\frac{1}{2}\{2\ell nT^2 + n^2T^2 - 2nT_{\mu}\}/\delta^2] \times \\
&\quad \exp[-\frac{1}{2}\{2(\ell + n)rT^2 + r^2T^2 - 2rT_{\mu}\}/\delta^2] \\
&= \Phi_{\ell}[n]\Phi_{\ell+n}[r]
\end{aligned}$$

With the transition matrix defined, a Kalman filter can be developed for this nonlinear case as was established in Chapter II for the linear case. Many of the equations follow from the traditional state space Kalman filter. Redeveloping the formulation for this nonlinear function would only be repetitious. The necessary changes of definitions are restricted to the transition matrix, already defined, and the $\underline{H}$ matrix which becomes

$$\underline{H}_k = \begin{bmatrix} \underline{\Phi}_k[0] \\ \underline{\Phi}_k[-1] \\ \vdots \\ \underline{\Phi}_k[-(k-1)] \end{bmatrix}$$

This $\underline{H}_k$ matrix is identical to that previously defined with the exception of the $\underline{Q}(t)$ matrix which becomes the identity matrix (for a (1,1) matrix, $\underline{Q}(t)$ would be just 1).

To initialize the recursion formula, both $\hat{\underline{B}}_k$ and $\underline{S}_k$ must be defined. Both of these vectors contain only one element and the matrix notation shall be dropped.

Using the linear least squares formula, the best estimate and covariance for the initialization of the recursion can be found.

$$\hat{b}_k = (H_k^T H_k)^{-1} H_k^T Y_k \quad \text{and} \quad s_k = (H_k^T H_k)^{-1} \sigma^2$$

The number of points, k , to initialize the recursion are user defined. However, it has been found that three points are sufficient. It is best to minimize the number of points used for initialization as the filter does not affect the magnitudes of these increments. In terms of the gaussian split from the data set, these initialization points do not define the gaussian at those particular incremental points.

It is important to observe that the only parameters required by the Kalman filter to split out a specific gaussian are the mean and the standard deviation. The amplitude information is stored in the data points and is recovered from the initialization and the weighting of the difference between the measurement and the forecasted value.

A data set comprised of a single noisy gaussian is illustrated in Figure 4-1. Given the correct mean and standard deviation to define the transition matrix, the Kalman filter traces out almost an exact copy of the original noiseless gaussian. The initialization could begin from the left or right hand side of the data set requiring the transition matrix to forecast increasing increment values or decreasing increment values, respectively.

In considering two summed gaussians, it becomes important to associate the left or right side in initialization and filtering with the mean and standard deviation which correctly defines the nearest gaussian (Figure 4-2). Mismatching the sides and nonlinear parameters result in the obvious poor approximation of the original summed gaussians (Figure 4-3).

Figure 4-1. Kalman filter using ideal nonlinear parameters.

Figure 4-2. Nonlinear parameters correctly matched for Kalman filter.

Figure 4-3. Nonlinear parameters mismatched using Kalman filter.

Kalman's ability to ideally split these summed gaussians apart is dependent on the correct values of μ and δ specified for each gaussian, and on the proximity of the gaussians. It is apparent from the difference between the Kalman filtered estimates and the original gaussians used in the summation seen in Figure 4-4, that even when the ideal mean and standard deviation of each gaussian in the data set is known, the Kalman approximation of each gaussian are biased by the close proximity of the other gaussian. A recursion between approximations is required to ultimately split out the original gaussians.

The recursion is defined in the following manner. The Kalman filter initially splits out a biased approximation of one of the gaussians from the original data set. Either the left or the right gaussian may be chosen in this step. For clarity in this discussion, choose the left gaussian for this initial cut. Once the filtered form of the left gaussian is obtained, subtract the estimate from the data set. The remaining data contains a biased form of the right gaussian. Kalman is used on this data set to obtain an approximation for the right gaussian. Subtract this filtered approximation from this present data set to obtain a noise residual. This residual should appear as white gaussian noise, random about a zero mean. If it is not, then the two filtered approximations are not close estimates of the ideally summed gaussians. In this case return to the approximation of the left gaussian and add the noise residual. Use Kalman again to filter this data set to obtain a second approximation of the left gaussian. Subtract this approximation from the data set and add the resulting noise residual to the approximation of the right gaussian. This process continues, adding remains, filtering approximations, and subtracting

Figure 4-4. First iteration of a Kalman recursion using ideal parameters.

estimates of each gaussian, until the change seen in approximations between iterations becomes small and the noise residual appears random about a zero mean. An easily obtained criterion to check is the sum of the squares of the noise residual.

Ultimately, the ideal gaussians are closely approximated (Figure 4-5). The initial choice of which gaussian to filter must be made carefully, and on the basis of which trailing edge has the least interference [18].

In the preceeding discussion it was assumed that the ideal mean and standard deviation were known and used for each iteration's filtered approximation of the left and right gaussian. This is rarely the case. Chapter III showed how Marquardt could be used over small windows under certain limiting conditions to obtain close approximations of these parameters. These approximations of the nonlinear parameter can be updated at each iteration and returned to Kalman to split out the approximated gaussian. The window used by Marquardt to estimate the μ and δ must be chosen at the rising edge of the gaussian which is least biased by the other approximated gaussian. These updated approximations will always be close in value but will improve as the recursion eliminates bias, induced by the proximity of the other gaussian. As this bias is reduced, larger windows can be taken to estimate the parameters. This reduces the noise influence and improves the accuracy of the estimates. At the conclusion of the iterations Marquardt can estimate the nonlinear parameters from the filtered approximations made by Kalman rather than from a data set containing the approximation plus residual noise. The results are extremely close to the ideal values, due to the noise filtering ability of Kalman. Using these

THIS PAGE
WAS INTENTIONALLY
LEFT BLANK

Figure 4-5(a). Complete Kalman recursion using ideal parameters.

Figure 4-5(b). Noise residual after last iteration.

approximations Marquardt is essentially estimating the parameters from a noiseless data set. Results from Chapter III verify that these approximations should be extremely close to the ideal values. Figures 4-6, 4-7, and 4-8 show several examples of Kalman recursion with normalized windowed Marquardt reestimation of the parameters at each iteration.

As good as these results look, it is important to remember that this technique is restricted to limited cases. Limited windowed Marquardt requires that the additive noise level is low and that there must exist at least a small portion of the data set which closely tracks the rising edge of the ideal gaussian. The user must not only find this window but recognize it. These are limiting conditions, but if they exist, good results are quickly obtained with limited computing facilities.

An alternative to Marquardt's estimation of these nonlinear parameters may be taken almost directly from statistics [19]. In Chapter II, the best estimate for the mean was defined as

$$\mu = \frac{\sum_{j=1}^P n_j P(N = n_j)}{P}$$

Instead of the probability density function, $P(N = n_j)$, describing the distribution of the noise, let the independent variable be nT . The probability density function which describes this distribution of nT is just the measurements $y[nT]$. These measurements, under no additive noise, conveniently follow the gaussian curve, although not normalized, for a single gaussian model. Since additive noise is present consider the measurements squared to represent $P(N)$. This alleviates the possibility of this analagous probability density

Figure 4-6(a). Kalman recursion, parameter estimates from normalized window Marquardt

Figure 4-6(b). Noise residual.

Figure 4-7(a). Kalman recursion, parameter estimates from normalized window Marquardt.

Figure 4-7(b). Noise residual.

Figure 4-8(a). Kalman recursion, parameter estimates from normalized window Marquardt.

Figure 4-8(b). Noise residual.

function to become negative. With these changes, the estimated mean from the observations can be found from

$$\mu = \frac{\sum_{n=\ell}^{\ell+m} nT y^2[nT]}{\sum_{n=\ell}^{\ell+m} y^2[nT]}$$

where the denominator term is required for normalization.

The standard deviation of the gaussian must now be found. Since the measurements have been squared, the resulting signal is still gaussian, although the standard deviation has changed. In effect, the squaring has divided the variance by two; consequently, the standard deviation we seek is

$$\delta = \sqrt{2} T_x$$

where T_x is the standard deviation of the squared measurements. It can be obtained from

$$T_x^2 = \frac{\sum_{n=\ell}^{\ell+m} [nT - \mu]^2 y^2[nT]}{\sum_{n=\ell}^{\ell+m} y^2[nT]}$$

This directly corresponds to the definition of variance given in Chapter II.

These "stochastic" approximations of the mean and standard deviation assume that the data set consists of only one gaussian. In the case of well separated summed gaussians, Figure 4-9, limited windows of the data set can be used to estimate these nonlinear parameters. Windows approximately symmetric about the apparent mean of the gaussian give the closest approximation to the ideal gaussian's parameters. Notice that the additive noise on this example is much

Figure 4-9(a). Kalman recursion parameter estimates from "stochastic" approximations.

Figure 4-9(b). Noise residual.

higher than that considered for windowed Marquardt and the results obtained are good.

For overlapping gaussian signals, this technique of limiting the window over which the parameter can be estimated is not viable on the first iteration. Instead, one of the gaussians must be manually removed [18] from the data set before these "stochastic" approximations can be made. A priori knowledge must be utilized to supply the mean and standard deviation to Kalman to separate this first approximation. Again the choice of which gaussian to filter out must be based on which side has the minimum bias from the other gaussian. If ideal parameters are used in the first iteration, Figure 4-10 shows the speed of convergence. Using Marquardt approximations, the parameters were reestimated at each iteration. This process was found to be disastrous when using the "stochastic" method. Convergence was never achieved. Instead of iterative reestimation, initial values of the parameters were kept throughout the recursion process until the sum squared of the noise residuals was at a minimum. This minimum can be quickly identified, for the sum squared noise residual will remain relatively unchanged between iteration's evaluations.

Naturally the ideal parameter values will not always be known and thus used on one of the gaussians in the sum. Hopefully the parameters initially guessed will be close to the ideal values. Figure 4-11 shows that with nonideal parameters used initially, the convergence takes longer but the results obtained are as accurate as those found in Figure 4-10. Figure 4-11 is divided into three major sections, in which the method of evaluating the parameters for the filtering is different. In the first section, Figure 4-11a, the

Figure 4-10(a). Kalman recursion, ideal and "stochastic" estimates of parameters.

Figure 4-10(b). Noise residual.

initial guess parameters and the "stochastically" estimated parameters are retained throughout the recursion until the minimum in the sum squared noise is obtained. The noise residual is then reviewed over the data set to verify its randomness about a zero mean, Figure 4-11b.

Although the sum of the squares of the noise residuals had been reached, the dispersion was not satisfactory. As a result, the parameters were then "stochastically" reestimated before each filtering process occurred. This continued until another minimum of the sum squared noise residual was achieved. This constituted the second major section shown in Figure 4-11c with a noise residual of Figure 4-11d. Finally, the last filtered approximations from section two were handed to the normalized window Marquardt and the parameters were reestimated from these smoothed data sets. The parameters estimated by Marquardt were then only used by the Kalman filter to separate the least biased gaussian in the set. The other gaussian's parameters were again "stochastically" estimated at each iteration. This process continued until the last sum squared noise residual minimum occurred. The normalized window Marquardt again approximated the nonlinear parameters from the filtered estimates. This was section three. The results were extremely close to the ideal parameters. When looking back at the batch mode Marquardt estimates of these same parameters, Figure 3-8, it appears that the results presently obtained are closer approximations to the ideal parameters than those obtained by batch Marquardt. This is predictable as both the error space and the parameter estimates of the batch Marquardt have an attached variance to their unbiased estimations; hence, the apparent inaccuracy can occur due to additive noise. In the recursive form outlined here, this noise is ultimately

Figure 4-11(a). Kalman recursion, section I

Figure 4-11(b). Noise residual after section I.

Figure 4-11(c). Kalman recursion, section II.

Figure 4-11(d). Noise residual after section II.

Figure 4-11(e). Kalman recursion, section III.

Figure 4-11(f). Noise residual after section III.

filtered out so this same variability about an unbiased estimate does not occur.

Figures 4-12 and 4-13 are other examples of this three sectional filtering technique. Figure 4-13 is somewhat different than Figures 4-11 and 4-12. In this case, the summed gaussian models are equally distributed and biased. After filtering this data set with each sectional method (Figure 4-13a, b, c, d, e, f) the resulting noise residual was not adequately centered about the zero mean, specifically, the first half of the spectrum. To find better estimates, the third section (Figure 4-13g) used the estimated Marquardt parameters for only the first iteration, the remaining iterations in this section were "stochastically" reestimated before filtering. Again, a minimum in the sum squared noise residual denoted the end of this section. The third sectional filtering techniques were then applied to these new estimations (Figure 4-13i). The final results achieved were good. To summarize the processing of this data set:

- 1] Section I techniques were used (Figure 4-13a).
- 2] Section II techniques were applied (Figure 4-13c).
- 3] A Marquardt evaluation of the nonlinear parameters was made on 2]'s filtered estimates.
- 4] One iteration was made using 3]'s parameter estimations (Figure 4-13g).
- 5] Section II's techniques were used (Figure 4-13g).
- 6] Section III techniques were applied (Figure 4-13i).

Notice the additive noise levels on each of these examples. This method has the capability of handling much higher noise than Marquardt's normalized window estimation.

Figure 4-12(a). Kalman recursion, section I.

Figure 4-12(b). Noise residual.

Figure 4-12(c). Kalman recursion, section II.

Figure 4-12(d). Noise residual.

Figure 4-12(e). Kalman recursion, section III.

Figure 4-12(f). Noise residual.

Figure 4-13(a). Kalman recursion, section I.

Figure 4-13(b). Noise residual.

Figure 4-13(c). Kalman recursion, section II.

Figure 4-13(d). Noise residual.

Figure 4-13(e). Kalman recursion, section III.

Figure 4-13(f). Noise residual.

Figure 4-13(g). Revision, add after section II another section II Kalman recursion.

Figure 4-13(h). Noise residual.

Figure 4-13(i). Kalman recursion, new section III.

Figure 4-13(j). Noise residual.

CHAPTER V

CONCLUSIONS

In considering both batch and recursive nonlinear estimators, the dependence on the operator is high. The batch mode Marquardt can only estimate parameters as well as the operator can supply an appropriate model and close initial guesses. Similarly, the recursive method presented bases its ability to separate gaussians on the operator's choice in window size and location.

The actual time saved using the recursive method rather than the batch Marquardt is difficult to quantize. All of the step sizes in the search for a minimum are automatically chosen using Marquardt. This recursive technique has not yet been automated. However, the results are quickly returned from the Kalman filter and the estimation of the nonlinear parameters by either the "stochastic" or windowed Marquardt are immediately available. The operator can see the trend of convergence within, at most, ten iterations. All of the convergence factors in the examples shown here occurred in less than thirty iterations.

This recursion method works well on most noisy gaussians only mediumly separated. For closely located gaussians, batch Marquardt could provide a better fit if an exact model and "proper" initial guesses were chosen. This means that both techniques have difficulty separating gaussians which heavily bias each other. Special conditions

must exist for each technique before an accurate estimation of each gaussian can be achieved under biasing from the other gaussian.

The separation of linear from nonlinear terms, and the consideration of only one gaussian at a time, conceptually makes this division of components easier to visualize. The storage required is also broken into smaller pieces. Because each part can be considered separately, the memory allocated to process each portion of the problem is smaller. Each part can be brought into memory in an overlay configuration.

The recursive algorithm defined has parameterized the nonlinear terms and removed the actual gaussians from a summed gaussian expression. The linear term, amplitude, remains to be found for the gaussians. Analytically the amplitude may be determined by generating a unity amplitude gaussian from the nonlinear terms previously determined. This unity amplitude gaussian can be used for the $\underline{H}$ matrix in a least squares formulation, since the amplitude is linearly related to this exponential data. The model, the unity gaussian, can be nonlinear in terms of the independent variable [20], the incremental steps.

With this result, the total parameterization of each gaussian is known and the individual curves are available. These results have been derived from rich data sets of only two summed gaussians. The next consideration are data sets composed of more than two gaussians with only sparse data points defining the set. Interpolation between points may be required to achieve the same information contained in the data sets considered here. Recursive techniques will require some

modification in convergence criteria and method when several rather than two gaussians are closely associated. This will require further investigation.

REFERENCES

1. Yule, H.P., "Computer studies of complex full energy peaks using 2nd and 3rd derivatives," Modern Trends in Activation Analysis, U.S. Department of Commerce, National Bureau of Standards, Special Publication 312, Vol. II, pp. 1256-1259.
2. Savitsky, A. and Golay, M.V.E., "Smoothing and differentiation of data by simplified least squares procedures," Analytical Chemistry, Vol. 36, #8, July 1964, pp. 1627.
3. Edwards, R.T., Settle, L.G., and Knight, R.D., "A digital algorithm for spectral deconvolution with noise filtering and peak picking--NOFIPP-DECON," NASA Technical Memorandum, NASA TMX-64927, February 1975.
4. Bevington, P.R., Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, 1969.
5. Brubaker, T.A., Colorado State University, Fort Collins, Colorado, unpublished notes for EE581 parameter estimation, Summer 1979.
6. Loseke, T., Nonlinear Parameter Estimation with Gaussian Models, Colorado State University, M.S., 1978.
7. Draper, N.R. and Smith, H., Applied Regression Analysis, John Wiley and Sons Inc., 1966.
8. Eykhoff, P., System Identification, John Wiley and Sons Inc., 1974.
9. Kalman, R.E., "A new approach to linear filtering and prediction problems," Journal of Basic Engineering, March 1960, pp. 35-45.
10. Barham, P.M. and Humphries, D.E., "Derivation of the Kalman filtering equations from elementary statistical principles," in "Theory and applications of Kalman filtering," AGARDograph, No. 139, C.T. Leondes, ed., 1970.
11. Levenberg, K., "A method for the solution of certain nonlinear problems in least squares," Quarterly of Applied Mathematics, Vol. 2, 1977, pp. 164-168.
12. Marquardt, D.W., "An algorithm for least-squares estimation of nonlinear parameters," Journal of the Society for Industrial and Applied Mathematics, Vol. 11, No. 2, June 1963, pp. 431-441.

13. Balaban, D., Lawrence Livermore Laboratory, Livermore, California, unpublished notes on the Levenberg-Marquardt method, 1978.
14. Beck, J.V. and Arnold, K.J., Parameter Estimation in Engineering and Science, John Wiley and Sons Inc., 1977.
15. Box, M.J., "A comparison of several current optimization methods and the use of transformations in constrained problems," Computer Journal, Vol. 9, 1966, pp. 67-77.
16. Golub, G.H. and Pereyra, V., "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," SIAM Journal of Numerical Analysis, Vol. 10, 1973, pp. 413-432.
17. Corradi, C. and Stefanini, L., "Computational experience with algorithms for separable nonlinear least squares problems," CNR-GNIM meeting in Florence, Sept. 1976.
18. Brubaker, T.A., Cornett, F.N., and Pomernachi, C.L., "Linear digital filtering for laboratory automation," Proceedings of the IEEE, Vol. 63, No. 10, Oct. 1975, pp. 1475-1486.
19. Franks, L.E., Signal Theory, New Jersey: Prentice-Hall Inc., 1969.
20. Brubaker, T.A., Tracy, R. and Pomernachi, C.L., "Linear Parameter Estimation," Analytical Chemistry, Vol. 50, No. 11, pp. 1017A-1024A.

APPENDIX A

LIST OF SYMBOLS

Vector and Matrix Notation

$\underline{M}$	(n,1)	Model Vector (dependent variable)
$\underline{B}$	(p,1)	True Parameter or state vector
$\underline{H}$	(n,p)	Independent Variable Matrix
$\underline{N}$	(n,1)	White gaussian noise vector
$\hat{\underline{B}}$	(p,1)	Estimated parameter or state vector
$\underline{Y}$	(n,1)	Measurement vector
$\underline{R}$	(n,1)	Error difference vector ($\underline{Y} - \underline{M}$)
$\hat{\underline{M}}$	(n,1)	Estimated model vector
$\underline{W}$	(n,n)	Weight matrix
$\underline{O}(t)$	(1,p)	Vector relating states of system to measurements
$\underline{\phi}_k$	(p,p)	Transition matrix at k^{th} point
$\hat{\underline{B}}_{k+1}$	(p,1)	Forecast of estimated state or parameter vector
$\underline{S}$	(p,p)	Covariance matrix
$\underline{S}^-$	(p,p)	Forecasted covariance matrix
$\underline{\nabla}_{\underline{B}}^{\hat{\underline{M}}^T}$	(p,n)	Gradient of model, jacobian
$\underline{\nabla}_{\underline{B}}^{\hat{\underline{M}}^2}$	(p,p)	Hessian of model
$\underline{\nabla}_{\underline{B}}^{\hat{\underline{e}}}$	(p,1)	Gradient of error
$\underline{\nabla}_{\underline{B}}^{\hat{\underline{e}}^2}$	(p,p)	Hessian of error
$\underline{L}$	(p,p)	Adjustment matrix
$\underline{I}$	(p,p) or (n,n)	Identity matrix

Scalar Notation

e	total squared error ($\underline{R}^T \underline{R}$)
n, m	number of observations
p	number of parameters
i, j, k, l	index numbers
T	step size
a_i	amplitude
μ_i	mean
δ_i	standard deviation
g	Marquardt parameter