Non-linear least squares estimation with an iteratively updated data set
by Marc Leonard Keppler

A thesis submitted in partial fulfillment of the requirements for the Degree of Master of Science in Electrical Engineering
Montana State University

Abstract:
This thesis presents a nonlinear least squares approach that iteratively updates the parameter estimate while stepping through the data set. The approach presented proceeds by first computing the parameter estimate for a reduced portion of the data set and then adding consecutive data samples and recomputing the parameter estimate for the new data set. This process repeats until a final parameter estimate for the entire data set is obtained. This approach was developed to estimate the parameters of systems that can be modeled by differential equations. Two routines, referred to collectively as the continuation methods, that utilize this approach are examined. One method steps through the data space using an arbitrary schedule supplied by the user. The other routine computes the number of consecutive data samples to add on each iteration. The continuation methods were validated by comparing their performance to an implementation of Levenburg-Marquardt. Experimental results obtained from a wide variety of functions and dynamic data obtained from a practical fuel cell power system demonstrated that the continuation methods have significantly larger regions of convergence than the traditional Levenburg-Marquardt approach.

NON-LINEAR LEAST SQUARES ESTIMATION WITH AN ITERATIVELY UPDATED DATA SET

by

Marc Leonard Keppler

A thesis submitted in partial fulfillment
of the requirements for the Degree

of

Master of Science

in

Electrical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

September 2003

APPROVAL

of a thesis submitted by

Marc Leonard Keppler

This thesis has been read by each member of the thesis committee and has been found to be satisfactory regarding content, English usage, format, citations, bibliographic style, and consistency, and is ready for submission to the College of Graduate Studies.

Dr. Steven R. Shaw _____ 9-16-2003
(Signature) Date

Approved for the Department of Electrical and Computer Engineering

Dr. Jim Peterson _____ 9/18/2003
(Signature) Date

Approved for the College of Graduate Studies

Dr. Bruce McLeod _____ 9-23-03
(Signature) Date

STATEMENT OF PERMISSION OF USE

In presenting this thesis in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library.

If I have indicated my intention to copyright this thesis by including a copyright notice page, copying is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for permission for extended quotation from or reproduction of this thesis in whole or in parts may be granted only by the copyright holder.

Signature Marie L Kepple

Date 9/12/2003

ACKNOWLEDGMENTS

I would like to thank Professor Steven Shaw for his support, contributions, and patience. Professor Shaw's input and advice over the last few years are greatly appreciated. Professor Shaw was a motivating and inspiring advisor.

TABLE OF CONTENTS

TABLE OF CONTENTS – CONTINUED

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES – CONTINUED

LIST OF FIGURES – CONTINUED

GLOSSARY

local minimum	Minimum of a function on a closed interval
global minimum	Minimum of the entire function
residual	Difference between the observed data and estimated data.
region of convergence	Range of initial parameter estimates for which the routine converges to the actual value.
iteration	One complete cycle through the continuation methods with a fixed number of data points.
loss function	A mathematical expression that computes how the parameter estimate changes the residual.
convergence	Percentage of total tests for which the routine converged to the actual parameters.
test problem	The problem of estimating the parameters of a single data set using a single initial guess.
trust region	Region of the loss function that is not expected to contain a local minima.
sensible heat	Heat energy stored in a substance as a result of an increase in its temperature.
eps	Machine precision
$\vec{N}$	A vector that completely defines the steps taken in data space by the simple continuation method.
Δ_N	Size of the step taken in the data space
N_k	Number of data samples evaluated by the continuation methods on the k'th iteration
N_{max}	Maximum number of samples in a measured data set
$\|\| * \|\|$	The l_2 vector norm of the quantity $*$

ABSTRACT

This thesis presents a nonlinear least squares approach that iteratively updates the parameter estimate while stepping through the data set. The approach presented proceeds by first computing the parameter estimate for a reduced portion of the data set and then adding consecutive data samples and recomputing the parameter estimate for the new data set. This process repeats until a final parameter estimate for the entire data set is obtained. This approach was developed to estimate the parameters of systems that can be modeled by differential equations. Two routines, referred to collectively as the continuation methods, that utilize this approach are examined. One method steps through the data space using an arbitrary schedule supplied by the user. The other routine computes the number of consecutive data samples to add on each iteration. The continuation methods were validated by comparing their performance to an implementation of Levenburg-Marquardt. Experimental results obtained from a wide variety of functions and dynamic data obtained from a practical fuel cell power system demonstrated that the continuation methods have significantly larger regions of convergence than the traditional Levenburg-Marquardt approach.

INTRODUCTION

Least squares estimation is a common solution to the problem of determining unknown system parameters from a measured data set. Least squares fitting estimates the actual parameters of a modeled system by minimizing the sum of the squared errors between the estimated and measured data. Two major categories of least squares fitting are linear least squares and nonlinear least squares (NLS) estimation. Linear least squares can be used when the function relating the system model is linear from the parameters to the output. Otherwise, NLS is used [1].

NLS estimation routines generally update an initial parameter estimate $\hat{\theta}_0$ iteratively to obtain the final estimate $\hat{\theta}$. These routines generally converge only for a limited range of initial guesses. The continuation methods presented in this thesis were developed to increase the region of convergence (ROC) provided by least squares estimation. Development of the continuation methods was motivated by the desire to estimate the parameters of an induction motor from transient stator current measurements. A general approach for estimating the parameters of these motors based on increasing amounts of data was proposed in [2] and the continuation methods presented in this thesis are an extension of this research. The continuation methods are based on the NLS approach and were developed to estimate the parameters of systems that can be modeled by differential equations.

Systems that can be modeled by differential equations represent a broad class of useful problems. The continuation methods developed in this thesis are beneficial whenever an accurate initial guess for a system that can be modeled by differential equations is not available or easy to obtain. Two interesting specializations of this problem class are noninvasive system diagnostics and identification of fuel cell model parameters. One example of noninvasive system diagnostics is the Nonintrusive Load Monitor (NILM) [3, 4, 5, 6, 7, 8]. This device is installed at a centralized location in a

power distribution network and monitors electrical transients to determine the condition of the individual loads in the network. This requires the NILM to identify a wide range of systems without an accurate initial guess for every transient. Another interesting specialization of this problem class is the identification of fuel cell parameters. Much research has focused on the development of steady-state [9, 10, 11, 12, 13] and dynamic [14, 15, 16, 17, 18] models for fuel cells. System identification using these models can improve the design of individual fuel cells and fuel cell power systems.

This thesis begins with an overview of least squares estimation in Chapter 2. The description in Chapter 2 includes general least squares estimation along with the Gauss-Newton and Levenberg-Marquardt methods. Chapter 3 introduces the continuation methods and examines the computations involved with each method. The continuation methods are compared to a traditional Levenberg-Marquardt routine using simulated data sets in Chapter 4 and using dynamic data obtained from a practical fuel cell power system in Chapter 5. Chapter 6 summarizes the experimental results and provides suggestions for future research.

LEAST SQUARES ESTIMATION

This chapter summarizes the mathematical computations involved with least squares estimation and the Gauss-Newton and Levenberg-Marquardt iterations. The overview provided is intended to facilitate a description of the continuation methods and the problem of local minima from a mathematical point of view. This review is not exhaustive. More rigorous treatments of these topics are given in [19, 20, 21].

The primary purpose of least squares is to determine an m dimensional parameter vector $\hat{\theta}$ that minimizes the error between an n dimensional measurement vector y and an n dimensional estimate vector $\hat{y}$ given an p dimensional input vector u and a function relating these quantities

$$\hat{y} = f(\hat{\theta}, u). \tag{2.1}$$

The error measure for least squares is the loss function

$$V(\theta) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \frac{1}{2}r'r, \tag{2.2}$$

where r is the n dimensional residual vector defined by

$$r = y - \hat{y}. \tag{2.3}$$

The loss function (2.2) produces a maximum likelihood estimate when the data has normal, identically distributed, and independent errors [20].

Linear Least Squares

Linear least squares is applied to systems that are linear in the parameters. These systems can be expressed

$$y(u) = \sum_{i=1}^{m}\theta_i f_i(u), \tag{2.4}$$

where f_i represents the i'th component of f in (2.1) except f is no longer a function of θ. Linear systems can also be expressed using the linear matrix equation

$$y = A\theta, \tag{2.5}$$

where y is linearly related to the parameters θ. If the system is linear in the parameters, it is possible to obtain the final least squares estimate in a single iteration. If $m = n$ then the solution can be found by inverting A as follows

$$\hat{\theta} = A^{-1}y \tag{2.6}$$

If $n > m$ then the pseudo-inverse must be utilized. When $n > m$ the linear least squares problem is often solved using the singular value decomposition (SVD) [20, 22].

Non-Linear Least Squares

If a system is not linear in the parameters nonlinear least squares must be applied. Nonlinear least squares is applied to systems that are expressed

$$y(u) = f(\theta, u), \tag{2.7}$$

where the parameters θ are inside the nonlinear function f. These systems require an iterative approach to minimize the loss function in (2.2). Many methods exist that solve the NLS problem. However, the Levenberg-Marquardt method [23, 1, 24] and its variations [25, 26] are most commonly used [20].

The Levenberg-Marquardt method is an adaptation of the Gauss-Newton method. The Gauss-Newton iteration is

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \delta_{GN}^{(k)}, \tag{2.8}$$

$$\delta_{GN} = (J^T J)^{-1} J^T r, \tag{2.9}$$

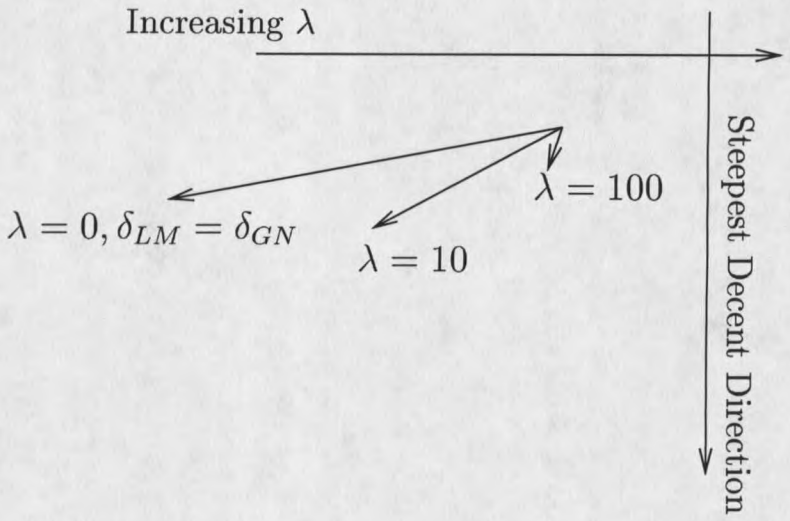

Fig. 2.1. An illustration of the relationship between δ_{GN} and δ_{LM}.

where J is the Jacobian matrix [19] with elements

$$J_{i,j} = -\frac{\partial r_i}{\partial \theta_j}\bigg|_{\hat{\theta}}. \tag{2.10}$$

The Levenberg-Marquardt step δ_{LM} follows δ_{GN} with an additional weighting term that allows the step to interpolate between the steepest decent direction and δ_{GN}. The Levenberg-Marquardt step is given by

$$\delta_{LM} = (J^T J + \lambda D^T D)^{-1} J^T r, \tag{2.11}$$

where λ and sometimes D are changed on each iteration [27]. The term λ is often referred to as the Levenberg-Marquardt parameter and D is a weighting matrix that accounts for the scaling of the problem. Implementations of Levenberg-Marquardt vary primarily in how D and λ are computed [24]. Notice that δ_{LM} approaches δ_{GN} as $\lambda \to 0$ and δ_{LM} reduces to zero and approaches the steepest-descent direction as λ increases with $D = I$ [20]. The relationship between δ_{GN} and δ_{LM} is illustrated in Fig. 2.1.

The disadvantage of most NLS estimation algorithms, and Gauss-Newton and Levenberg-Marquardt in particular, is that they evaluate the loss function using local

information. Thus, they converge to the minimum closest to the initial parameter estimate. This is a significant problem because the loss function of a nonlinear system can have many local minima. Routines that evaluate the loss function using local information are susceptible to local minima because the derivative of the loss function is zero at both global and local extrema. The derivative of the loss function at any extrema is

$$\nabla V(\theta) = \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial}{\partial\theta_1} \quad \frac{\partial}{\partial\theta_2} \quad \cdots \quad \frac{\partial}{\partial\theta_m}\right) r_i^2 = 0. \tag{2.12}$$

Expanding this derivative yields

$$\nabla V(\theta) = \sum_{i=1}^{n}\left(\frac{\partial r_i}{\partial\theta_1} \quad \frac{\partial r_i}{\partial\theta_2} \quad \cdots \quad \frac{\partial r_i}{\partial\theta_m}\right) r_i. \tag{2.13}$$

Noticing that the term

$$\left(\frac{\partial r_i}{\partial\theta_1} \quad \frac{\partial r_i}{\partial\theta_2} \quad \cdots \quad \frac{\partial r_i}{\partial\theta_m}\right) \tag{2.14}$$

is a row of the Jacobian matrix defined in (2.10), it follows that [24]

$$\nabla V(\theta) = \sum_{i=1}^{n}\left(\frac{\partial r_i}{\partial\theta_1} \quad \frac{\partial r_i}{\partial\theta_2} \quad \cdots \quad \frac{\partial r_i}{\partial\theta_m}\right) r_i = -(J^T r)^T = 0, \tag{2.15}$$

$$\delta_{LM} = (J^T J + \lambda D^T D)^{-1} J^T r = 0. \tag{2.16}$$

The Levenberg-Marquardt step δ_{LM} is zero at a local minimum and the iteration does not improve the parameter estimate.

An Example of Local Minima

Estimating the frequency of a sine wave illustrates the problem of local minima. Traditional NLS routines that utilize local information generally require an accurate initial guess to estimate the frequency of a sine wave. Suppose the observed data is given by the function

$$y = \sin(6t), \tag{2.17}$$

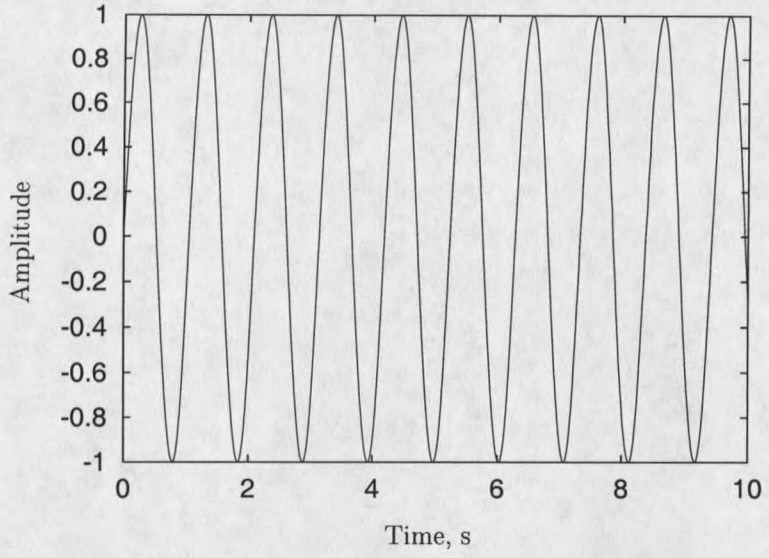

Fig. 2.2. Observation set for the single parameter sine wave (2.17) used to illustrate the problem of local minima.

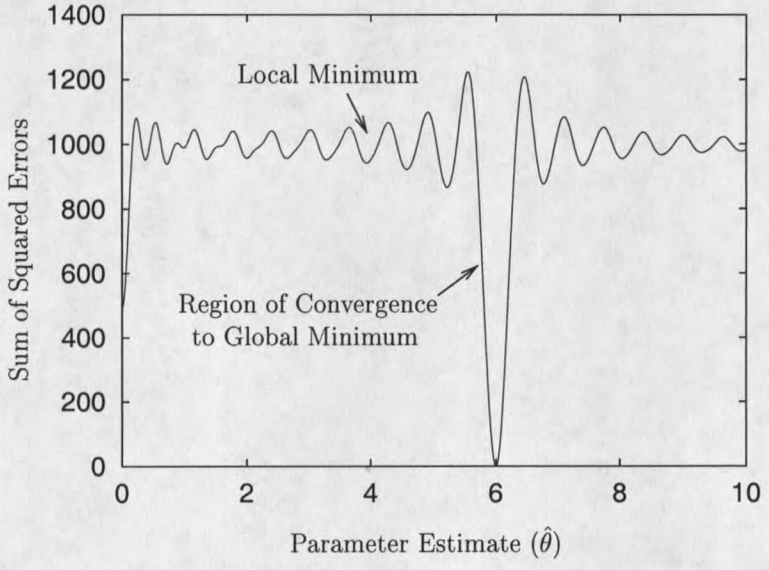

Fig. 2.3. Loss function for (2.17).

where t is a time vector on the interval $[0, 10]$ seconds with a sample rate of 100 samples/sec (see Fig. 2.2). The loss function for this observation set is shown in Fig. 2.3. Many local minima exist in this loss function. Traditional least squares approaches will only converge to the actual frequency if the initial guess is within the concave-up region directly surrounding the global minimum. For this example, traditional least squares routines will converge to the actual parameter estimate only if the initial guess is within approximately 7.5% of the actual value over the interval $[0, 10]$. This is a very small region of convergence. The disadvantage of NLS routines that use local information is that their convergence completely depends on the initial guess and the shape of the loss function.

THE CONTINUATION METHODS

The Approach

Before the specific computations of the continuation methods are presented it is necessary to understand the fundamental approach of the routines. The continuation methods avoid local minima and increase the ROC by stepping iteratively through the data space and solving the NLS problem on each iteration. One iteration through the continuation methods corresponds to one minimization of the loss function (2.2) for a particular number of data points N_k. The continuation methods update the data set by adding consecutive data samples. That is, they begin with the first data point d_1 and add data points following the scheme

$$(d_1 \quad d_2 \quad d_3 \quad \cdots \quad d_{N_k} \quad \cdots \quad d_{N_{k+1}} \quad \cdots \quad d_{N_{max}})$$

where d_i represents the i'th data point, N_k is the number of data points evaluated on the k'th iteration and N_{max} is the maximum number of data points in the observation set. The general approach taken by the continuation methods is

1) Obtain the initial data points to evaluate $N_{k=0}$
2) Minimize the loss function over N_k
3) Increase the data set to N_{k+1}
4) Repeat 2) and 3) until $N_k = N_{max}$ and
   the convergence criteria is met.

Both continuation methods follow the general approach outlined above. However, they differ in how they obtain the initial number of data points $N_{k=0}$, how the data space is updated on each iteration, and how the parameter estimate is computed. The specific differences between these routines are examined in the following sections. For