A STUDY OF THREE ALGORITHMS FOR NONLINEAR LEAST SQUARES PARAMETER ESTIMATION

by

MICKEY LINN STILSON

B.S., Kansas State University, 1971

A MASTER'S REPORT

submitted in partial fulfillment of the

requirements for the degree

MASTER OF SCIENCE

Department of Statistics

KANSAS STATE UNIVERSITY

1976

Approved by:

George A. Milliken

Major Professor

Table of Contents

1. Introduction.

Scientists are frequently interested in investigating the relationship between some response or dependent variable, denoted by y, and a vector of independent variables, denoted by $\underline{x}$. Assume the relationship is postulated to be some function f involving a vector of parameters $\underline{\theta}$, as

$$y=f(\underline{x},\underline{\theta}).$$

For an experimental situation where y is a random variable, the assumed relationship is

$$E(y)=f(\underline{x},\underline{\theta})$$

and the model with additive error structure can be expressed as

$$y=f(\underline{x},\underline{\theta})+\varepsilon.$$

Thus over the course of n observations, we construct the model

$$\underline{y}=\underline{f}(\underline{x},\underline{\theta})+\underline{\varepsilon},$$

where $\underline{y}$, $\underline{f}(\underline{x},\underline{\theta})$, and $\underline{\varepsilon}$ are $n\times 1$ vectors.

In order to more accurately classify the type of models to be analyzed, we present the following definitions ([9]):

Definition 1.1 A model is $y=f(\underline{x};\underline{\theta};\underline{\varepsilon})$, where y is the value of an observed random variable which is to measure the phenomena under study (dependent variable), $\underline{x}$ is a vector of constants or other observed random variables (independent variables), $\underline{\theta}$ is a vector of unknown parameters, $\underline{\varepsilon}$ is a vector of unobserved random variables with some assumed distribution, and $f(\cdot,\cdot,\cdot)$ is some known mathematical function of $\underline{x}$, $\underline{\theta}$, and $\underline{\varepsilon}$.

Definition 1.2 A model is defined to be a linear model if $y=f(\underline{x},\underline{\theta})+h(\underline{\varepsilon})$, where $f(\underline{x},\underline{\theta})$ is a linear function of the elements of $\underline{\theta}$.

Note the additive error structure implied in the above definition of a linear model. An example is $y=\theta_0+\theta_1x+\theta_2x^2+\varepsilon$.

Definition 1.3 A model is defined to be essentially linear if there exists a transformation $\ell(y)=\ell(f(\underline{x},\underline{\theta},\underline{\varepsilon}))$ such that $\ell(y)=g(\underline{x},\underline{\theta})+h(\underline{\varepsilon})$, where $g(\underline{x},\underline{\theta})$ is a linear function of $\underline{\theta}$, and $h(\underline{\varepsilon})$ is a function of $\underline{\varepsilon}$ only. An example of an essentially linear model is

$$y=(e^{\theta_0}e^{\theta_1x})(\varepsilon)$$

as $ln(y)=\theta_0+\theta_1x+ln(\varepsilon)$. The model $y=\exp(\theta_0+\theta_1x)+\varepsilon$ is nonlinear because we cannot make a transformation to a linear model retaining an additive error structure. The additive error structure is important in order to apply least squares to making interval estimates about $\underline{\theta}$ and examining various distributional properties.

Definition 1.4 Any model $y=f(\underline{x},\underline{\theta},\underline{\varepsilon})$ that is not linear or essentially linear is defined to be nonlinear.

Statistical theory offers many techniques for obtaining estimators of $\underline{\theta}$ from the model $y=f(\underline{x},\underline{\theta})+\underline{\varepsilon}$, including maximum likelihood, Bayesian, and least squares. For an account of general methods of obtaining estimators see [10](Chapter VII). For most techniques some objective function of $\underline{\theta}$, say $\Phi(\underline{\theta})$, is to be optimized. Examples of Φ include risk functions (Bayesian estimation), likelihood functions (maximum likelihood estimation), and sums of squares (least squares estimation). In this paper we restrict ourselves to obtaining least squares estimators for $\underline{\theta}$ from the model $y=f(\underline{x},\underline{\theta})+\underline{\varepsilon}$, where f is generally a nonlinear function in $\underline{\theta}$. These estimators are also maximum likelihood estimators when normality is assumed. Since linear estimation is a special case of nonlinear estimation, all results discussed will apply equally to linear and nonlinear estimation.

In our study we will begin with a general formulation of a technique of finding a least squares estimator $\hat{\underline{\theta}}$ for $\underline{\theta}$ in $y=f(\underline{x},\underline{\theta})+\underline{\varepsilon}$. The technique utilizes a Taylor series linear approximation to f and develops an iterative scheme to approach $\hat{\underline{\theta}}$. The scheme is generally referred to as the Gauss-Newton or Taylor series method. We then study three modifications to the general technique of the Gauss-Newton method. These include the modified Gauss-Newton [6], the Marquardt [8], and the Spiral [7] algorithms. These modifications are based on the premise that a procedure that converges in fewer iterations and/or with less computational effort is an improvement.

Section 3 is devoted to several examples and the problem of parameter estimation under constraints is considered in Section 4. The appendix documents a computer program developed by the author incorporating the algorithms of the modified Gauss-Newton, Marquardt, and Spiral techniques. The results of Section 4 are included in the program so that constrained estimation is possible using either the modified Gauss-Newton or Spiral algorithms. The appendix includes a user's guide to the program along with sample output.

In this study all theorems are quoted without proof and often with less than complete rigor. The reader is referred to the references for detail.