A

CONTRIBUTION

BUT
CE&M=16.

to

THE SOLVING OF NONLINEAR ESTIMATION PROBLEMS

by

RENÉ GONIN

Submitted
in fulfilment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematical Statistics
University of Cape Town

Promotor: Professor A H Money

August 1983

to Joyce

and

my parents who rest in peace

ACKNOWLEDGEMENTS

I wish to express my gratitude to my promotor, Prof A H Money, for his enthusiasm and encouragement throughout this study. Arthur, thank you for your support. This has been the most rewarding period of my academic career.

My heartfelt gratitude to my wife, Joyce for her patience, support and encouragement throughout. Joyce, thank you for the thorough editing and proofreading of the manuscript at the cost of preparation time for your own final medical examinations.

My gratitude to Dr Steve du Toit for his interest in my research endeavours and for suggesting the simulation study reported in Chapter 5.

Many thanks to Dr Debbie Bradshaw for her perusal of the manuscript and suggestions for improving the manuscript.

My sincere gratitude to Miss Antoinette van Zyl for her accurate and imaginative typing and to Mrs Marie Kotze for her efficient co-ordination of the typing. Also to Dave Gargan and Dominic Rooney for the smooth running of the computer.

Finally I wish to thank Dr Stephen Fellingham for the use of the facilities of the Institute of Biostatistics at the Medical Research Council and my colleagues in the Institute for their interest in my research.

R GONIN

CAPE TOWN
AUGUST 1983

TABLE OF CONTENTS

Chapter 1 : Introduction: The nonlinear curve fitting problem.

The first curve fitting models were encountered in Astronomy where the prediction of the orbit of a celestial body was of interest. This precipitated the need for estimating the parameters of the underlying mathematical model. In essence all the early techniques utilised a process of averaging the empirical data or functions of the data.

Galileo Galilei (1632) interested in determining the distance of a new star from the earth remarked that "it will be appropriate for us to apply the minimum amendments and smallest corrections that we can - just enough to remove the observations from impossibility and restore them to possibility". This statement heralded the beginning of the theory of errors in which the true model is derived from a number of inconsistent observations.

A brief resume of historical events will be enlightening and appropriate.

1. The origin and history of curve fitting problems.

One of the first methods for smoothing random errors was based on averages and was known as the Principle of the Arithmetic mean: Suppose we wish to estimate the model $y = bx + a$ from the observations (x_i, y_i) $i=1,\ldots n$.

Slopes between all possible pairs of points,

$$b_{ij} = \frac{y_j - y_i}{x_j - x_i} \qquad (\text{with } x_j \neq x_i \text{ for } i=1,\ldots n-1;\ j=i+1,\ldots n)$$

are first calculated and the corresponding a_{ij} in each case is then calculated by substitution. The averages of the slopes and intercepts $\bar{b}$ and $\bar{a}$ respectively are then taken as estimates of b and a.

Cotes (1722) noted in certain models that only the dependent variable (observations y) is subject to measurement errors and suggested a procedure based on weighted arithmetic means with weights proportional to $|x|$. In the model $y = bx + e$, b is estimated by $\hat{b}=\bar{y}/\bar{x}$, the ratio of the two means, which in turn is equivalent to the zero sum residuals condition:

$$\sum_{i=1}^{n} (y_i - bx_i) = 0,$$

which is the same as stipulating that the line must pass through the centroid $(\bar{x},\bar{y})$ of the data.

Euler (1749) and Mayer (1750) independently derived the so-called Method of Averages for fitting a straight line to observed data. The observations are subdivided into as many subsets as there are coefficients. The grouping is made according to the value of one of the independent variables. Those with the largest values of this variable are grouped together and so on. The condition of zero sum residuals is then applied to each observation of the subset. The formation of subsets is, however, subjective and arbitrary (consult Nyquist (1980) for more detail).

Boscovich (1757) considered the model $y = bx + a + e$ and proposed two criteria for fitting the best straight line:

1) the sum of the positive and negative residuals in the y-variable must be equal or $\sum_{i=1}^{n} (y_i - a - bx_i) = 0$ and

2) $\sum_{i=1}^{n} |y_i - a - bx_i|$ must be a minimum.

Condition 2) is still used in the minimum absolute deviations (MAD), least absolute errors (LAE) or L_1-norm estimation procedures. Boscovich's solution procedure is based on geometric principles. Laplace (1786) also used the Boscovich principle to test the adequacy of the relationship $y = a + bx$ for the data (x_i, y_i). His procedure for determining the coefficients a and b was analytical in nature.

Gauss developed the method of minimizing the squared observation errors in his works on celestial mechanics which subsequently became known as the method of least squares (L_2-norm estimation). Although Gauss (1806) had used least squares since 1795, Legendre (1805) was the first to publish the method. He derived the normal equations algebraically without using calculus. He claimed that least squares was superior to other existing methods but gave no proof. In 1809 Gauss derived the normal (Gaussian) law of error which states that the arithmetic mean of the observations of an unknown variable x will be the most probable. Gauss (1820) succinctly writes: "Determining a magnitude by observation can justly be compared to a game in which there is a danger of loss but no hope of gain... Evidently the loss in the game cannot be compared directly to the error which has

been committed, for then a positive error would represent loss and a negative error a gain. The magnitude of the loss must on the contrary be evaluated by a function of the error of which the value is always positive... it seems natural to choose the simplest (function), which is, beyond contradiction, the square of the error." Nyquist op. cit.

Laplace (1818) examined the distributional properties of the parameter b in the simple regression model y=bx+e when L_1-norm estimation is used. He assumed that all the errors had the same symmetric distribution about zero and derived the density function f of the errors e. He also showed that the slope b was normally distributed with mean zero

and variance $\{4f(0)^2 \sum_{i=1}^{n} x_i^2\}^{-1}$ for large sample sizes(n).

This is a well known result for the sample median in the location model y=b+e.

Cauchy (1824) examined the fitting of a straight line y=a+bx to data and proposed minimizing the maximum absolute residual. This he achieved by means of an iterative procedure. Chebychev (1854) in the approximation of functions also proposed the estimation of parameters by means of minimizing the maximum absolute difference between the observed function and the estimated function. This minimax procedure later became known as Chebychev approximation or L_∞-norm approximation.

Edgeworth (1883) questioned the universal use of the normal law of error and also examined the problem of outliers. This problem was taken further by Doolittle (1884). Edgeworth (1887a) and (1887b) used the Boscovich

principle and abandoned the zero sum residual condition which forces the line through the centroid of the data. He considered the case where least squares is inappropriate, i.e. where the error distributions are unknown or contaminated (normal) distributions.

In his monumental survey of the history of curve fitting, Harter (1974a), (1974b), (1975a), (1975b), (1975c), (1975d), (1976) summarises the work of the 20th century. Since these contemporary developments are well known they will not be discussed here.

2. The linear L_p-norm estimation problem.

The three main L_p-norm estimation procedures in this century have been L_1, L_2 and L_∞-norm estimation. In recent years, however, statisticians and mathematicians have shown an interest in L_p-norm estimation where p is any value in the range $1 < p < \infty$. The linear L_p-norm estimation problem is then defined as:

Find the parameters $\underline{b} = (b_1, b_2, \ldots b_k)'$ which minimize

$$\sum_{i=1}^{n} |y_i - \underline{b}'\underline{x}_i|^p \tag{1.1}$$

with y_i the response (dependent) and $\underline{x}_i = (x_{1i}, \ldots x_{mi})'$ the independent variables.

Extensive research in linear L_1-norm estimation has been reported in the statistical and mathematical journals. A comprehensive survey of references was undertaken by Narula and Wellington (1982) see also Gentle (1977). In addition to their list we can add: Anderson and Osborne (1976) who considered linear approximation problems in polyhedral norms (L_1- and L_∞-norm approximations are special cases); Harvey (1977) who compared the L_1-norm estimator to two other well known robust estimators (Hinnich and Talwar (1975) and Andrews (1974)) and concluded that the L_1-norm estimator is asymptotically more efficient than the other two.

The linear L_1- and L_∞-norm estimation problems are solved by means of linear programming. Two very efficient algorithms for solving L_1-norm problems are those by Barrodale & Roberts (1973) and algorithm 79-01 by Armstrong, Frome and Kung (1979). Algorithms for solving the L_∞-norm (minimax) problem are those given by Armstrong and Kung (1979) and (1980) as well as Barrodale and Phillips (1975). The interested reader is also referred to the unifying text by Arthanari and Dodge (1981). Sadovski (1974) followed an alternative approach which is based on the the original procedure proposed by Edgeworth (1888).

Sklar & Armstrong (1982) used the linear least squares solution as an initial basis in the linear programming formulation of the L_1- and L_∞-norm estimation problems. They showed that a significant saving in computation is achieved on a number of the well known algorithms. See also McCormick and Sposito (1976) and Hand and Sposito (1980).

It is interesting to note that stepwise selection of variables has been incorporated into L_1-norm algorithms. See for example Gentle & Hanson (1977) or Roodman (1974).

The more general linear L_p-norm estimation problem ($p \neq 1,2$ or ∞) has also received a considerable amount of attention. Descloux (1963) and Fletcher, Grant and Hebden (1974a) showed that the L_∞-norm estimator is the limiting L_p-norm estimator as $p \to \infty$. Fletcher, Grant and Hebden (1974b) demonstrated that the parameters of the continuous L_p-norm approximation problem for $p \geq 2$ are both continuous and differentiable functions of p. This result, however, does not hold for the discrete L_p-norm problem (1.1).

Barrodale and Roberts (1970) showed that the linear L_p-norm estimation problem can be formulated as a nonlinear programming (NLP) problem in which the objective function is concave for $0 < p < 1$ and convex for $1 < p < \infty$ and the constraints linear. They suggested the use of the convex simplex method for solving the latter problem. (A homogeneous unconstrained algorithm for linear constraints by Breytenbach (1978) may also be used). They proposed a modification of the simplex method for linear programming to solve the problem when $0 < p < 1$.

Fletcher, Grant and Hebden (1971) derived a method which takes the structure of problem (1.1) into account. Their method is analagous to Newton's method for solving algebraic equations. They proved that provided certain integrals exist, the method converges for all values of $p \geq 2$ and that the convergence rate is quadratic when $p \geq 3$. Prior to this paper the Davidon-Fletcher-Powell (DFP) quasi-Newton method was the method of choice. This is a standard method for unconstrained optimization problems. The method by Fletcher (1970) is, however, more efficient than the DFP method (see Himmelblau (1972) and Himmelblau and Lindsay (1980) for numerical comparisons). Kahng (1972) and Rey (1975) have also presented algorithms based on Newton's method.

Merle and Späth (1973) remarked that the errors in the response variables are often non-normally distributed with unequal variances. They suggested two algorithms; the first for problems where $1 \leq p \leq 2$ and the second for problems where $p > 2$. In the first algorithm, also known as iteratively reweighted least squares, they set zero residuals equal to a small positive constant. The second algorithm is also based on Newton's method as mentioned above. They found that the first algorithm converges on numerical examples (although the convergence is not proved) whilst the second Newton-type algorithm is, of course, known to converge.

Schlossmacher (1973) derived a method for solving linear L_p-norm problems which uses iteratively reweighted least squares. The method temporarily deletes observations which give rise to zero residuals and then reinstates them in subsequent iterations if their residuals become larger. The method will not always yield a solution nor has a convergence proof been derived (Kennedy and Gentle (1980) p 532).

Ekblom (1973) reformulated problem (1.1) as the perturbed problem:

Find parameters $\underline{b}$ which minimize

$$(1.2) \qquad \sum_{i=1}^{n} \left|(y_i - \underline{b}'\underline{x}_i)^2 + e^2\right|^{p/2} \qquad \text{(where e is finite)}$$

The author used the modified (damped) Newton method to solve problem (1.2). The advantage is that the Hessian of the perturbed problem remains positive definite as long as $e \neq 0$, hence a decrease is assured at every iteration. Ekblom then showed that the limiting solution as $e \to 0$ is the solution to the original problem (1.1).

Schlossmacher's method was extended by Sposito, Kennedy and Gentle (1977) to the case $1 \leq p \leq 2$ for the simple linear model $y = a + bx$. Kennedy and Gentle (1978) showed that a conventional quasi-Newton method should be used when $1 < p < 2$ and that a modified Newton method works well on problems when $p > 2$. Barr et al. (1980) extended the method of Sposito et al. (1977) to the multiple regression case. They concluded that their method is useful for solving linear L_p-norm regression problems for p values in the range [1,2.6]. However, they demonstrated that for $p > 2.6$ the DFP method is superior and that when $p \geq 3$ their method will not converge.

Wolfe (1979) examined the first algorithm of Merle and Späth and went on to prove convergence of their algorithm when $1 < p < 2$. He showed that the rate of convergence is geometric with an asymptotic convergence constant of 2-p. A similar result holds for p=1 if the best approximation is unique. This paper supports the empirical findings of Merle and Späth op.cit.

Fischer (1981) considered problem (1.1) for the case $1 < p < 2$. He transformed it to the following linearly constrained problem by setting $r_i = \underline{b}'\underline{x}_i - y_i$ for i=1,...n. Find the parameters $\underline{\theta}$ which minimize

$$\sum_{i=1}^{n} |r_i|^p \text{ subject to } r_i = 0.$$

This problem, known as the primal (see Chapter 2 of Zangwill (1969)) can be formulated as:

$$\min_{(\underline{x},\underline{r})} \max_{\underline{u}} L(\underline{x},\underline{r},\underline{u}) = \sum_{i=1}^{n} (|r_i|^p + u_i(\underline{b}'\underline{x}_i - r_i - y_i))$$

convergent for all $p > 1$ are presented. The third method is a descent method and the fourth, a Newton-based method. Numerical difficulties were experienced when p was close to 1. In view of this problem, Watson is undertaking further research with regard to the solution of the orthogonal L_1-norm problem.

Forsythe (1972), in estimating the parameters a and b in the simple regression model $y = a + bx$, proposed the use of L_p-norm estimation with $1 < p \leq 2$. He argued that since the mean is sensitive to deviations from normality the L_p-norm estimator will be more robust than least squares in estimating the mean. This will be the case when outliers are present. He suggested the compromise use of p=1.5 when contaminated or skewly distributed error distributions are encountered and the DFP method as minimization technique.

In a simulation study Ekblom (1974) compared the L_p-norm estimators with the Huber M-estimators. He also considered the case when $p < 1$. He concluded that the Huber estimator is superior to the L_p-norm estimates when the errors are contaminated normally distributed. For other error distributions (Laplace, Cauchy) he suggested that p=1.25 should be used. The proposal that $p \leq 1$ should be used for skewly distributed (Chi-square) errors is interesting and shows that the remark by Rice (1964) that problems where $p < 1$ are not of interest, is not justified. Ekblom warns against the use of least squares when the errors are non-normally distributed.

Harter (1977) suggested an adaptive scheme which relies on the kurtosis of the regression error distribution. He suggested L_1-norm estimation if the kurtosis $\beta_2 > 3.8$, least squares if $2.2 \leq \beta_2 \leq 3.8$ and Chebychev or L_∞-norm estimation if $\beta_2 < 2.2$. This scheme has been extended by Barr (1980), Money et al. (1982) and Sposito et al. (1983) and will be discussed in Chapters 4 and 5 of this thesis.

Using the asymmetrical estimator concept introduced by Sielken and Hartley (1973), Harvey (1978) showed that the linear L_p-norm estimator will always be unbiased for $1 < p < \infty$ given the assumptions that the regression errors are symmetrically distributed, the first moment exists (for Cauchy distributed errors the estimator will be biased) and that the model is of full column rank. When p=1 the estimator may not be unique and the estimator may be biased. Sielken and Hartley op. cit. showed how an unbiased estimator may be obtained by means of linear programming. Sposito (1982) extended this result to the general case where $p \geq 1$. This he achieved by formulating the linear L_p-norm problem as a convex programming problem which can be solved by the convex-simplex method of Zangwill (1969). The procedure by Sielken and Hartley is then applied. Hence unbiased L_p-norm estimates for all $p \geq 1$ can be obtained.

In his thesis Nyquist (1980) considered the statistical properties of linear L_p-norm estimators. He derived the asymptotic distribution of linear L_p-norm estimators and showed it to be normal for sufficiently small values of p. It is not stated, however, how small p should be. The multicollinear, stochastic regressor and linearly dependent residual cases were also examined. A procedure for selecting the optimal value of p based on the asymptotic variance is proposed which validates the empirical

studies by Barr (1980) and Money et al. (1982). These results have been submitted for publication by Nyquist (1982). Asymptotic properties of the L_1-norm estimator were derived by Bassett and Koenker (1978). They also showed that the relative efficiency of the L_1-norm estimator to the least squares estimator is the same as the relative efficiency of the sample median to the sample mean.

By means of a simulation study, Barr (1980) studied the properties of linear L_p-norm estimation and constructed an adaptive L_p-norm estimation procedure. Some of the results may be found in Money et al. (1982) who derived an empirical relationship between the optimal p-value and the kurtosis of the error distribution.

Sposito et al. (1983) derived a different empirical relationship which also relates the optimal p-value to the kurtosis of the error distribution. Both these predictor formulae will be the object of study in Chapter 5. These authors also showed that the Money et al. formula yields a reasonable value of p for error distributions with a finite range and suggested the use of their own formula for large sample sizes ($n \geq 200$) when it is known that $1 \leq p \leq 2$. The following modification of Harter's rule was suggested: Use p=1.5 (Forsythe) if $3 < \beta_2 < 6$, least squares if $2.2 \leq \beta_2 \leq 3$ and L_∞-norm estimation if $\beta_2 < 2.2$.

We therefore conclude that active research currently concentrates on deriving more efficient algorithms for solving linear L_p-norm estimation problems as well as studying the distributional properties of these estimates. This research also includes the related problems of heteroscedasticity of errors and problems in which both the dependent and

independent variables are subject to error. In conjunction with paragraph 3, this brief survey is intended to highlight the similarities, differences and areas yet to be explored in the research into linear and nonlinear L_p-norm estimation.

3. Survey of recent research in nonlinear L_p-norm estimation.

The literature on nonlinear least squares is fairly extensive. Surveys of algorithms may be found in Dennis and Welsch (1978), the excellent text by Kennedy and Gentle (1980) and for large residual problems, Nazareth (1980). A numerical comparison of various nonlinear least squares algorithms was undertaken by Hiebert (1979).

The main results of research into the statistical aspects of nonlinear least squares will now be discussed briefly.

Jennrich (1969) derived conditions for consistency of the estimators $\hat{\underline{\theta}}$ in nonlinear regression and showed that these estimators are asymptotically normally distributed with mean $\underline{\theta}$ (optimal) and variance $\sigma^2(J'J)^{-1}$. His work was supported by that of Malinvaud (1970). Jennrich's results may be used to construct confidence intervals for $\hat{\underline{\theta}}$. Since these intervals are based on first-order derivative information only they are approximate and may therefore be quite inaccurate. Clarke (1980) derived an expression for the variance-covariance matrix of $\hat{\underline{\theta}}$ which takes into account third- and fourth-order derivative information.

Hypothesis testing in the nonlinear case is approached in basically the same way as in linear regression and may be carried out using 1) the likelihood ratio test, 2) a test based on the asymptotic normality of $\underline{\theta}$ or 3) Hartley's (1964) test. Gallant (1975) compared these tests in a simulation study and suggested that the likelihood ratio test be used. Milliken and DeBruin (1978) also derived a procedure for testing hypotheses about $\underline{\theta}$. Khorasani and Milliken (1982) used the confidence regions of $\underline{\theta}$ in conjunction with optimization procedures to determine a conservative simulataneous confidence band around the nonlinear model. Their procedure depends on the finding of an appropriate confidence region for $\underline{\theta}$. If this is not possible a meaningless confidence band will be constructed. Johnson and Milliken (1983) discussed a procedure which tests linear hypotheses about the parameters $\underline{\theta}$. For example, if we wish to know whether a particular nonlinear relationship holds for different populations, then the parameters of the model for each population group may be compared by testing a linear hypothesis about the $\underline{\theta}$. Their procedure for hypothesis testing was be carried out using weighted least squares.

Clarke (1982) summarised the properties of $\hat{\underline{\theta}}$ as follows:

1) The least squares estimate $\hat{\underline{\theta}}$ that minimizes $S_2(\underline{\theta})$ also maximizes the likelihood function.
2) Contours of constant likelihood for given σ^2(the variance of the error distribution) may be defined by $S_2(\underline{\theta})=c$, where c is some constant.
3) The estimates $\hat{\underline{\theta}}$ are not sufficient statistics for $\underline{\theta}$ (an exception occurs when all of the parameters enter linearly).

4) The maximum likelihood estimator $\hat{\underline{\theta}}$ is consistent and asymptotically normally distributed with mean $\underline{\theta}$ (true parameter value) and variance-covariance matrix $\tilde{J}$ ($\tilde{J}$ is Fisher's information matrix).

5) $s^2 = S_2(\hat{\underline{\theta}})/n$ is a consistent estimator of σ^2.

6) If the likelihood ratio statistic is defined as

$$T = \frac{S_2(\tilde{\underline{\theta}}) - S_2(\hat{\underline{\theta}})}{S_2(\hat{\underline{\theta}})} \quad x \quad \frac{n-k}{k}$$

where $\tilde{\underline{\theta}}$ is the value of $\underline{\theta}$ hypothesised under the test. If this hypothesis is true, then as a large sample approximation T follows the F distribution with k and n-k degrees of freedom.

Most of the results in the literature on nonlinear L_p-norm estimation relate to computational considerations. Relatively little appears to have been published with regard to the statistical aspects of nonlinear L_p-norm estimation. In Chapters 4 and 5 we shall consider some of the statistical properties of the parameters and the optimal p-value.

Wagner (1962) proposed that mathematical programming procedures should be used to solve nonlinear regression problems. However, the model he considered while nonlinear in the independent variables is linear in the parameters. It is stated that the exact functional form of the model is not of importance but that the functions must be monotonically increasing in the independent variable(s). The purpose is then to minimise the sum of the absolute deviations. Given additional assumptions, including the so-called weak curvature constraint, the minimum absolute deviation problem is reformulated as a linear programming problem.

Barrodale, Roberts and Hunt (1970) considered the L_p-norm estimation problem (for $p=1,2,\infty$) for a class of approximating functions which are nonlinear in only one of several parameters (the others obviously enter linearly). Their method proceeds as follows: For a given value (not necessarily optimal) of the single nonlinear parameter the best L_p-approximation can be accomplished using one of the available linear approximation algorithms (e.g. linear programming or least squares). The resulting error of approximation will then be a function of the one nonlinear parameter only. This univariate function can then be minimized by univariate search methods.

Two useful algorithms for multiple nonlinear L_1- and L_∞-norm estimation problems were proposed by Osborne and Watson (1971) and (1969) respectively. The L_1-norm algorithm will be discussed in Chapter 4 where it is used as part of the simulation study. In their method the nonlinear problem is reduced to a series of linear L_1-problems which can be solved efficiently by linear programming methods. The authors proved convergence of their algorithm given the following two conditions: 1) The model function f is once continuously differentiable and 2) The Jacobian of the regression functions is of full column rank. The order of convergence is linear (convergence rates and ranks of matrices are discussed in Appendix B, Chapter 3).

Anderson and Osborne (1977a) considered the discrete nonlinear approximation problem in a polyhedral norm. The L_1- and L_∞-norms are special cases of a polyhedral norm. Problems in these latter two norms may be formulated as linear programming problems. They then generalise these problems to polyhedral norm approximation problems which are expressed in terms of linear inequalities. The polyhedral norm is defined as follows:

Let B , $\underline{x}$ and $\underline{e}$ be mxn, nx1 and mx1 matrices with m > n and $\underline{e}$ the vector with unit components. Consider the set of inequalities $B\underline{x} \leq \underline{e}$ given the following conditions: (i) The feasible region $F = \{ \underline{x} | B\underline{x} \leq \underline{e} \}$ is bounded with a proper interior ($F \neq \phi$). (ii) $x \in F$ if and only if $-x \in F$. Then the polyhedral norm of $\underline{x}$ is written as

$$|| \underline{x} ||_B = \min\{ a | B\underline{x} \leq a\underline{e} \}, \text{ (where a is any scalar number).}$$

A thorough discussion of these norms may be found in Chapter 14 of Fletcher (1981).

Anderson and Osborne then considered the class of problems:

Find parameters $\underline{\theta}$ which minimize $||f(\underline{\theta})||_B$ where $f:R^k - R^1$.

The following two conditions were set on the problem; firstly a solution $\underline{\theta}^*$ to the problem exists and secondly $f(\underline{\theta})$ is sufficiently smooth (e.g. $f(\underline{\theta})$ is twice continuously differentiable). The authors derived an algorithm in which each step involves the solution of a linear polyhedral norm approximation problem. The method is a modification of Newton's method for locating the solution of a system of nonlinear equations. The authors proved convergence of their algorithm and show that the rate of convergence is at least quadratic. The authors conclude their paper with a number of numerical examples of L_1- and L_∞-norm approximation problems. No numerical examples of other polyhedral norms are given. The advantage of this algorithm over the original Osborne-Watson (1971) algorithm seems to be that convergence is faster i.e. quadratic as opposed to linear.

unconstrained minimization problems requiring solution. This is known as the penalty function method in nonlinear programming (see e.g. Fiacco and McCormick (1968)). These penalty functions are differentiable. The quasi-Newton method of Fletcher (1970) (also known as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method) is then used in the unconstrained minimization steps. The authors accomplished accelerated convergence of the algorithm by means of an extrapolation procedure. This procedure is well known in numerical analysis, (Henrici (1964) chapter 12). The authors concluded from their numerical studies that the Osborne-Watson algorithm converges slowly and that their algorithm is more efficient than the Osborne-Watson algorithm.

These authors also derived necessary first-order optimality conditions for the nonlinear L_1-approximation problem. The first set of conditions is essentially the Kuhn-Tucker conditions and the second set is based on directional derivatives. Consider the following L_1-problem:

(1.5) Find parameters $\underline{\theta}$ which minimize $F(\underline{\theta}) = \sum_{i=1}^{n} |f_i(\underline{\theta})|$,
$f: R^k \to R^1$ are continuously differentiable functions.

It is equivalent to the following nonlinear programming (NLP) problem:

$$\text{(1.6)} \qquad \text{minimize} \quad z = \sum_{i=1}^{n} g_i$$

$$\text{subject to} \quad -g_i + f_i(\underline{\theta}) \leq 0$$

$$i=1,\ldots n$$

$$-g_i - f_i(\underline{\theta}) \leq 0$$

with $z: R^{k+n} \to R^1$.

From the Kuhn-Tucker (1951) conditions the following first-order necessary optimality conditions result:

Lemma 1.1: A necessary condition for $\underline{\theta}^*$ to be a local solution to problem (1.6) is that there exists multipliers $v_i \in [-1,1]$ for all $i \in K(\underline{\theta}^*) = \{i \,|\, f_i(\underline{\theta}^*)=0\}$ such that

$$(1.7) \qquad \sum_{i \notin K(\underline{\theta}^*)} \text{sign } f_i(\underline{\theta}^*) \nabla f_i(\underline{\theta}^*) + \sum_{i \in K(\underline{\theta}^*)} v_i \nabla f_i(\underline{\theta}^*) = \underline{0}$$

In terms of directional derivative (see Appendix A, Chapter 2) we obtain:

Lemma 1.2: A necessary condition for $\underline{\theta}^*$ to be a local solution to problem (1.5) is that

$$(1.8) \qquad \sum_{i \notin K(\underline{\theta}^*)} \underline{h}' \nabla f_i(\underline{\theta}^*) \text{sign } f_i(\underline{\theta}^*) + \sum_{i \in K(\underline{\theta}^*)} |\underline{h}' \nabla f_i(\underline{\theta}^*)| \geq 0 \text{ for all } \underline{h} \in R^k.$$

These two sets of optimality conditions are equivalent.

Shrager and Hill (1980) also considered a Marquardt-Levenberg algorithm for solving the nonlinear L_1- and L_∞-norm estimation problems. Since linear programming is involved in calculating succesive estimates of the parameters, they suggested that a good starting estimate would reduce the overall computational burden. See also McCormick and Sposito (1976); Sklar & Armstrong (1982) and Hand and Sposito (1980) for similar suggestions. A second difficulty, the nonuniqueness of the parameters, is also overcome by using the linear programming procedure to select a single solution. The algorithm by Barrodale and Roberts (1973) as well as others have this

capability. The authors indicated that in the L_1-norm case the parameters are discontinuous with respect to the steplength in the line search procedure. They propose that this difficulty could be overcome by means of linear interpolation between two discontinuous points.

Charalambous (1979) and Ben-Tal and Zowe (1982) derived optimality conditions for the L_1-norm problem (1.5). Optimality conditions for more general non-differentiable problems may be found in Hiriart-Urruty (1978) and Fletcher and Watson (1980). These conditions involve the directional derivatives and curvature of the objective function in contrast to the gradient and Hessian of differentiable functions. The first-order necessary conditions are those given by (1.8). The following second-order sufficiency conditions were derived:

Lemma 1.3: Suppose the functions f_i i=1,... n are twice continuously differentiable. Then $\underline{\theta}^*$ is an isolated (strong) local minimum of $F(\underline{\theta})$ if there exist multipliers $-1 \leq v_i \leq 1$, $i \in K(\underline{\theta}^*)$, such that

$$\sum_{i \notin K(\underline{\theta}^*)} \text{sign } f_i(\underline{\theta}^*) \nabla f_i(\underline{\theta}^*) + \sum_{i \in K(\underline{\theta}^*)} v_i \nabla f_i(\underline{\theta}^*) = 0.$$

and for every $i \in K(\underline{\theta}^*)$ and $\underline{d} \neq 0$ satisfying:

$\underline{d}' \nabla f_i(\underline{\theta}^*) = 0$ for all i such that $|v_i| \neq 1$,

$\underline{d}' \nabla f_i(\underline{\theta}^*) \geq 0$ for all i such that $v_i = 1$ and

$\underline{d}' \nabla f_i(\underline{\theta}^*) \leq 0$ for all i such that $v_i = -1$ it follows that

$$\underline{d}' \Big[\sum_{i \notin K(\underline{\theta}^*)} \text{sign } f_i(\underline{\theta}^*) \nabla^2 f_i(\underline{\theta}^*) + \sum_{i \in K(\underline{\theta}^*)} v_i \nabla^2 f_i(\underline{\theta}^*) \Big] \underline{d} > 0.$$

Oberhofer (1982) considered the nonlinear L_1-norm estimation problem and established conditions that will not only guarantee the existence but also the consistency of the parameter estimates $\underline{\hat{\theta}}$. Given that the parameter space K is compact, that $\underline{\theta}$ is an inner point of K and that the problem functions $\underline{f}$ are continuous in $\underline{\theta}$ over K, the existence of the parameters is proved. Under 5 additional assumptions it is shown that the L_1-norm estimator of $\underline{\theta}$ is weakly consistent. His first three assumptions were first used by Jennrich (1969) and Malinvaud (1970) to prove consistency of nonlinear least squares estimates.

We conclude this section by observing that:

1) The research into the statistical aspects of nonlinear L_p-norm estimation is still in its infancy.

2) As far as the solving of nonlinear L_p-norm estimation problems is concerned the following remark by Fletcher (1981), Chapter 14, is relevant: " Algorithms for basic NDO (nondifferentiable optimization) have not progressed as far because of the difficulties caused by the limited availability of information. ...In fact there is currently much research interest in NDO algorithms of all kinds and further developments can be expected."

4. Scope and contribution of this thesis to nonlinear estimation.

The following nonlinear estimation problem will be considered : Given the data :

$$(y_i, x_{1i}, \ldots x_{mi}), \quad i=1, \ldots n$$

where y_i is the dependent and $\underline{x}_i = (x_{1i}, \ldots x_{mi})$ the independent variables, the problem is to estimate the k parameters $\underline{\theta} = (\theta_1, \ldots \theta_k)'$ from the nonlinear model

$$y_i = f(\underline{x}_i; \underline{\theta}) + e_i \qquad i=1,\ldots n$$

where $n > k$ in general, $f_i = f(\underline{x}_i; \underline{\theta})$ is the response function, $\underline{\theta}$ the vector of unknown parameters and the e_i unobserved error variates. Least squares (L_2-norm estimation) is the appropriate method for solving this problem when the error variates are normally distributed with expected value and variance: $E(e_i) = 0$, $var(e_i) = \sigma^2$ respectively. Alternatively, the solution may be obtained by means of the more general L_p-norm estimation problem:

Find the parameters $\underline{\theta}$ which minimize $S_p(\underline{\theta}) = \sum_{i=1}^{n} |y_i - f_i|^p$ for $1 \leq p < \infty$.

The following contibutions will be made in this thesis:

1) In Chapter 2 expressions for the first- and second-order partial derivatives of the objective function $S_p(\underline{\theta})$ with respect to $\underline{\theta}$ will be derived. A new compact matrix notation for these derivatives will be introduced and at the same time it will be shown that the nonlinear least squares problem is imbedded in the general L_p-norm estimation problem.

2) In addition, an algorithm will be derived which takes the structure of the L_p-norm estimation problem into account. The Gauss-Newton method for nonlinear least squares is imbedded in this gradient method. The gradient method uses only first-derivative information whilst second-order derivative information is ignored. It was designed to solve fairly mild small residual problems. At each iteration a direction of descent is needed. It is shown that the gradient direction is a descent direction if the matrix of second-order partial derivatives, the Hessian, is positive definite. The Hessian need not be positive definite and hence a modified Choleski factorization which transforms insufficiently positive definite and even indefinite matrices into positive definite matrices will be used. Graphical displays of the numerical examples will also be provided.

L_p-norm estimators will be examined and related properties for nonlinear L_p-norm estimators postulated. It will be shown that these theoretical proposals are in complete agreement with the simulation results. The generation of random numbers from the uniform, parabolic, triangular, normal, contaminated normal (kurtoses 4 and 5) and Laplace distributions will be considered. In addition a new formula which uses the analytical roots of cubic equations will be derived for generating random numbers from a parabolic distribution.

5) In Chapter 5 an adaptive procedure will be derived to calculate systematically the estimates of the optimal p-values for a given error distribution. This procedure will be used in a simulation study to derive the empirical distribution of these estimates . It will be shown that the estimates are asymptotically normal. Some examples of nonlinear models in medical research will also be discussed. The value of this adaptive procedure in identifying outlying observations will be illustrated by means of graphical displays. These examples will show that in the event where least squares is appropriate, the alternative estimation procedures are equally efficient.

Chapter 2 : A numerical algorithm for the small residual nonlinear L_p-norm estimation problem.

In this chapter we shall define the nonlinear estimation problem and show how it is formulated in the L_p-norm approximation context. The solution to the L_p-norm problem can either be obtained directly by numerical minimization or by transformation of the original problem into a nonlinear programming problem (NLP) which is then solved by means of standard nonlinearly constrained minimization procedures. The former approach will be employed as the latter is both cumbersome and numerically intractable.

We shall derive a new algorithm to solve the L_p-norm approximation problem. The new algorithm requires the first- and second-order partial derivatives of $S_p(\underline{\theta})$ with respect to $\underline{\theta}$. Expressions for these derivatives will be derived and at the same time a compact matrix notation will be introduced. These expressions are entirely new as far as the author is aware. The analogy between these results and those of least squares will also be drawn (see Corollary 2.2).

The algorithm makes use of the structure of the L_p-norm estimation problem and is an an extension of the classical Gauss-Newton method which was designed to solve nonlinear least squares problems. It will be shown that the Gauss-Newton method is imbedded within this new algorithm and therefore that only first-order partial derivatives need be considered. The use of second-order derivative information, which is costly from a computational point of view, is therefore ignored.

It will be shown that the algorithm is efficient in solving fairly well-behaved small residual nonlinear L_p-norm estimation problems where p is finite and $p > 1$. By means of numerical examples it will be indicated that the objection against the use of gradient methods on this type of non-differentiable problem may be ignored. To facilitate the understanding of this complex algorithm a numerical example with intermediate numerical calculations will be presented. Graphical displays of $S_p(\underline{\theta})$ for various examples are given in Figures 2.1 to 2.4.

1. The nonlinear estimation problem.

Suppose that the following data were collected on n occasions :

$$(y_i, x_{1i}, \ldots x_{mi}) \qquad i=1, \ldots n$$

where y_i is the response (or dependent) variable and $\underline{x}_i = (x_{1i}, \ldots x_{mi})$ the independent variables. Our problem is then to estimate the k parameters $\underline{\theta} = (\theta_1, \ldots \theta_k)'$ from the nonlinear model

$$y_i = f(\underline{x}_i; \underline{\theta}) + e_i \qquad i=1, \ldots n \tag{2.1}$$

where $n > k$ in general, $f_i = f(\underline{x}_i; \underline{\theta})$ is the response function, $\underline{\theta}$ the unknown parameters and e_i the unobserved error variates.

The method of least squares (L_2-norm estimation) is the appropriate method for solving this problem when the error variates are normally distributed with expected value and variance:

$$E(e_i) = 0 \quad , \; var(e_i) = \sigma^2$$

respectively. The error variates are frequently not normally distributed in which case least squares estimation is inappropriate. An alternative such as L_p-norm estimation then has to be considered.

The L_p-norm estimation problem for (2.1) is defined as:

Find the parameters $\underline{\theta}$ which minimize

$$(2.2) \quad S_p(\underline{\theta}) = \sum_{i=1}^{n} |y_i - f(\underline{x}_i;\underline{\theta})|^p$$

where $1 \leq p < \infty$.

Problem (2.2) can be reformulated as the nonlinear programming(NLP) problem:

$$\text{Minimize } z = \sum_{i=1}^{n} u_i$$

subject to

$$
\begin{aligned}
(2.3) \qquad & -u_i^{1/p} + y_i - f(\underline{x}_i;\underline{\theta}) \leq 0 \\
& -u_i^{1/p} - y_i + f(\underline{x}_i;\underline{\theta}) \leq 0 \\
& u_i \geq 0 \qquad i=1,\ldots n \\
& \underline{\theta} \text{ unconstrained in sign.}
\end{aligned}
$$

Note that the constraints are equivalent to stating that

$$|y_i - f(\underline{x}_i;\underline{\theta})|^p \leq u_i \quad \text{for all } i=1,\ldots n$$

The original problem in k unknowns has now been defined as an NLP problem in 2n nonlinear inequality constraints, n non-negative variables (u_i) and k unconstrained variables ($\underline{\theta}$). Although there are efficient numerical methods for solving problems in nonlinear constraints see for example Fiacco and McCormick (1968), Buys (1972), Bertsekas (1976) as well as El-Attar et al. (1979), this is a cumbersome way of solving the original estimation problem in view of the number of nonlinear constraints and additional number of constrained variables.

We shall, however, follow an approach which solves the L_p-norm estimation problem by means of a numerical minimization technique. The first- and second-order partial derivatives of $S_p(\underline{\theta})$ with respect to $\underline{\theta}$ will be required and hence expressions for these will be derived in Lemma 2.1.

We shall now introduce a compact matrix notation which will simplify the algebraic expressions for the first- and second-order partial derivatives of $S_p(\underline{\theta})$ with respect to $\underline{\theta}$.

Define

$$F_i = |y_i - f(\underline{x}_i, \underline{\theta})|^p ,$$

the p-Jacobian matrix J_p with (i,j)-th element

$$|y_i - f_i|^{\frac{1}{2}p-1} \frac{\partial f_i}{\partial \theta_j} \qquad i=1,\ldots n \; j=1,\ldots k$$

and the p-residual vector as

$$(\underline{y} - \underline{f})_p = [|y_i - f_i|^{\frac{1}{2}p-1} (y_i - f_i)] \quad i=1,\ldots n.$$

Let $\nabla^2 f_i$ be the Hessian matrix of f_i with respect to $\underline{\theta}$ and define the matrix

$$B_p(\underline{\theta}) = \sum_{i=1}^{n} |y_i - f_i|^{p-2} (f_i - y_i) \nabla^2 f_i .$$

Lemma 2.1: The first- and second-order partial derivatives of $S_p(\underline{\theta})$ with respect to $\underline{\theta}$ are given by:

$$(2.4) \quad \nabla S_p(\underline{\theta}) = -pJ_p'(\underline{y}-\underline{f})_p$$

$$(2.5) \quad \nabla^2 S_p(\underline{\theta}) = p[(p-1)J_p'J_p + B_p(\underline{\theta})].$$

Proof: We have $F_i^{1/p} = |y_i - f_i|$ for i=1,...n. By means of implicit differentiation we find that

$$\frac{1}{p} F_i^{\frac{1}{p}-1} \frac{\partial F_i}{\partial \theta_\ell} = -\text{sign}(y_i - f_i) \frac{\partial f_i}{\partial \theta_\ell} \quad \text{for some index } \ell \text{ when } y_i \neq f_i.$$

$$\text{Thus } \frac{\partial F_i}{\partial \theta_\ell} = -p|y_i - f_i|^{p-1} \text{ sign}(y_i - f_i) \frac{\partial f_i}{\partial \theta_\ell}$$

$$= -p|y_i - f_i|^{p-2} (y_i - f_i) \frac{\partial f_i}{\partial \theta_\ell}$$

By summation over i=1,...n, we find

$$\frac{\partial S_p}{\partial \theta_\ell} = \sum_{i=1}^{n} -p|y_i - f_i|^{p-2} (y_i - f_i) \frac{\partial f_i}{\partial \theta_\ell} \quad \text{for } \ell=1,...k.$$

By inspection we can see that the gradient

$$\nabla S_p(\underline{\theta}) = -pJ_p'(\underline{y}-\underline{f})_p$$

Similarly

$$\frac{\partial^2 F_i}{\partial \theta_\ell \partial \theta_s} = -p|y_i - f_i|^{p-2} (y_i - f_i) \frac{\partial^2 f_i}{\partial \theta_\ell \partial \theta_s} + p|y_i - f_i|^{p-2} \frac{\partial f_i}{\partial \theta_\ell} \cdot \frac{\partial f_i}{\partial \theta_s}$$

$$+ p(p-2)|y_i - f_i|^{p-4} (y_i - f_i) \frac{\partial f_i}{\partial \theta_\ell} \cdot \frac{\partial f_i}{\partial \theta_s}$$

$$= p(p-1)|y_i - f_i|^{p-2} \frac{\partial f_i}{\partial \theta_\ell} \cdot \frac{\partial f_i}{\partial \theta_s} - p|y_i - f_i|^{p-2} (y_i - f_i) \frac{\partial^2 f_i}{\partial \theta_\ell \partial \theta_s} .$$

By summation over i=1,...n, we find the desired expression for

$$\frac{\partial^2 S_p}{\partial\theta_\ell \partial\theta_s} = p \sum_{i=1}^{n} |y_i - f_i|^{p-2} \left\{ (p-1) \frac{\partial f_i}{\partial\theta_\ell} \cdot \frac{\partial f_i}{\partial\theta_s} + (f_i - y_i) \frac{\partial^2 f_i}{\partial\theta_\ell \partial\theta_s} \right\} .$$

By inspection we can see that the Hessian matrix

$\nabla^2 S_p(\underline{\theta}) = p[(p-1)J_p'J_p + B_p(\underline{\theta})]$ and the Lemma is proved.

Remark: Since $(\underline{f}-\underline{y})_p = [|f_i - y_i|^{\frac{1}{2}p-1} (f_i - y_i)]_{i=1,\ldots n}$

$= -[|y_i - f_i|^{\frac{1}{2}p-1} (y_i - f_i)]$

$= -(\underline{y}-\underline{f})_p$,

we can conveniently rewrite

$$\nabla S_p(\underline{\theta}) = -pJ_p'(\underline{y}-\underline{f})_p = pJ_p'(\underline{f}-\underline{y})_p .$$

The next two corollaries follow immediately:

Corollary 2.2: In L_2-norm (least squares) estimation the first- and second-order partial derivatives of $S_2(\underline{\theta})$ with respect to $\underline{\theta}$ are given by:

(2.6) $$\nabla S_2(\underline{\theta}) = -2J'(\underline{y}-\underline{f})$$

(2.7) $$\nabla^2 S_2(\underline{\theta}) = 2(J'J + B_2(\underline{\theta})),$$

where J is the usual Jacobian matrix with (i,j)-th element $\frac{\partial f_i}{\partial \theta_j}$ and vector $(\underline{y}-\underline{f})$ the nx1 vector $[(y_i - f_i)]$ whilst matrix

$$B_2(\underline{\theta}) = \sum_{i=1}^{n} (f_i - y_i)\, \nabla^2 f_i .$$

Corollary 2.3: In L_1-norm (least absolute value) estimation the first- and second-order partial derivatives of $S_1(\underline{\theta})$ with respect to $\underline{\theta}$ take the form:

$$\nabla S_1(\underline{\theta}) = -pJ_1'(\underline{y}-\underline{f})_1 = [-\sum_{i=1}^{n} \operatorname{sign}(y_i - f_i) \frac{\partial f_i}{\partial \theta_\ell}] \qquad \ell=1,\ldots k,$$

$$\nabla^2 S_1(\underline{\Theta}) = B_1(\underline{\theta}) = [\sum_{i=1}^{n} \operatorname{sign}(f_i - y_i) \frac{\partial^2 f_i}{\partial \theta_\ell \partial \theta_s}] \qquad \ell,s=1,\ldots k,$$

where J_1 has the (i,j)-th element $|y_i - f_i|^{-\frac{1}{2}} \frac{\partial f_i}{\partial \theta_j}$,

$$(\underline{y} - \underline{f})_1 = [|y_i - f_i|^{-\frac{1}{2}} (y_i - f_i)] \quad i=1,\ldots n$$

and $$B_1(\underline{\theta}) = \sum_{i=1}^{n} (f_i - y_i)/|y_i - f_i|)\, \nabla^2 f_i .$$

If a parameter θ_ℓ is included as a constant term in f_i, then its corresponding second-order partial derivatives $\frac{\partial^2 f_i}{\partial\theta_\ell \partial\theta_s} = 0$ for all s=1,...k. For example, suppose we wish to fit the model

$$f = \theta_3 + \theta_2 \exp(\theta_1 x)$$

by means of L_1-norm estimation. Then the corresponding Hessian matrix

$$\nabla^2 S_1(\underline{\theta}) = \begin{bmatrix} S_{11} & S_{12} & 0 \\ S_{12} & S_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{where} \quad S_{12} = -\sum_{i=1}^{n} \operatorname{sign}(f_i - y_i) \frac{\partial^2 f_i}{\partial\theta_1 \partial\theta_2} \quad \text{etc.}$$

We can see that this Hessian $\nabla^2 S_1(\underline{\theta})$ will be singular. Consequently second-order gradient methods cannot be used in the case of parameters entering as constants in L_1-norm estimation problems. Our algorithm will therefore only apply to such problems when p>1. When p=1, the Osborne and Watson (1971) algorithm may be used.

We are now in a position to state an algorithm for solving the nonlinear estimation problem (2.2).

2. Rationale of the L_p-norm first-order gradient method.

In our algorithm we shall use the newly derived expressions for the partial derivatives of $S_p(\underline{\theta})$ with respect to $\underline{\theta}$. The classical Gauss- Newton method will be extended to solve the more general nonlinear L_p-norm estimation problem.

In order to guarantee that the objective function $S_p(\underline{\theta})$ decreases at each succeeding iteration we must use a descent direction (a definition may be found in Lemma A1.1 in the Appendix). A necessary condition for this is that the Hessian matrix $\nabla^2 S_p(\underline{\theta})$ should be positive definite at each iteration (Lemma A1.2). It is well known that a function will be (strictly) convex if and only if its corresponding Hessian matrix is positive (semi) definite (see Theorem A5.1). It is also reasonable to assume that $S_p(\underline{\theta})$ behaves like a quadratic function in the vicinity of a local minimum $\underline{\theta}^*$. (a proof of this statement can be found in paragraph A5 in the Appendix). Note that a quadratic function which has a minimum (maximum) is convex (concave). It may however, happen that the Hessian will not be positive definite at each iteration, especially at points $\underline{\theta}$ away from the optimum. In this event we shall use the modified Choleski factorisation procedure due to Gill and Murray (1974) which transforms a given (Hessian) matrix into one which is positive definite (see Appendix paragraph A3).

The following notation will be used throughout.

$$(2.8) \quad \underline{g}^j = \nabla S_p(\underline{\theta}^j) = pJ_p'(\underline{f}-\underline{y})_p = pJ_p(\underline{\theta}^j)'J_p(\underline{\theta}^j)'(\underline{f}-\underline{y})_p$$

$$(2.9) \quad G^j = p(p-1)J_p'J_p = p(p-1)J_p(\underline{\theta}^j)'J_p(\underline{\theta}^j) \ .$$

In the classical Gauss-Newton method for nonlinear least squares a sequence of iterates $\{\underline{\theta}^j\}$ is constructed so that

$$\underline{\theta}^{j+1} = \underline{\theta}^j + \gamma_j \underline{d}^j$$

where γ_j is the steplength and $\underline{d}^j$ a direction of search satisfying the equation:

$$(2.10) \quad J'J\underline{d}^j = -J'(\underline{y}-\underline{f})$$

Note that only first derivative information (given by the Jacobian) is taken into account since for small residual problems the norm $||B_2(\underline{\theta})||$ is small compared to the norm $||J'J||$. The second-order terms can be ignored since $\underline{f} \to \underline{y}$ as $\underline{\theta}^j \to \underline{\theta}^*$ (the optimal point).

If we examine the Hessian, $\nabla^2 S_p(\underline{\theta})$, which arises in the general L_p-norm estimation problem we see that for $p>2$ we may ignore the second derivative term $B_p(\underline{\theta})$ in expression (2.5) when $\underline{f} \simeq \underline{y}$. If $p<2$ we may experience non-convergence of the algorithm especially when $\underline{y} \simeq \underline{f}$. This case will be discussed further in the next chapter.

When it is feasible to ignore the second derivative term $B_p(\underline{\theta})$ the Hessian can be approximated by

$$p(p-1)J_p'J_p.$$

We can therefore construct the first-order gradient iteration,

$$p(p-1)J_p'J_p\underline{d}^j = -pJ_p'(\underline{f}-\underline{y})_p \text{ or}$$

$$(p-1)J_p'J_p\underline{d}^j = -J_p'(\underline{f}-\underline{y})_p \tag{2.11}$$

for solving nonlinear L_p-norm estimation problems.

It is clear that the classical Gauss-Newton method is imbedded in the new method. This is so since iteration step (2.11) is expressed in terms of p, the p-Jacobian and the p-residual vector as opposed to the Jacobian and residual vector used in the Gauss-Newton iteration (2.10).

A provisional algorithm

Step 0 : Given an initial estimate $\underline{\theta}^\circ$ of $\underline{\theta}^*$, set j=0.

Step 1 : Calculate $\underline{g}^j = \nabla S_p(\underline{\theta}^j)$, J_p and solve the set of linear equations

$$(p-1)J_p'J_p\underline{d}^j = J_p'(\underline{y}-\underline{f})_p = -\underline{g}^j$$

Step 2 : Calculate γ_j by line search and set

$$\underline{\theta}^{j+1} = \underline{\theta}^{j} + \gamma_j \underline{d}^{j}$$

Step 3 : Continue until certain convergence criteria are met, otherwise set j:=j+1 and return to Step 1.

However, the above algorithm does not take into account the fact that the approximation of the Hessian given by matrix $J_p'J_p$ may not be positive definite at each iteration. The modified Choleski factorisation procedure will be used to overcome this difficulty. Individual steps in the algorithm are discussed in more detail in the Appendix.

THE ALGORITHM

Let $\varepsilon_1, \varepsilon_2$ and ε_3 $(= 10^{-9})$ be prescribed tolerances.

Step 0 : Select an initial estimate $\underline{\theta}^{\circ}$ of $\underline{\theta}^{*}$, set j=0.

Step 1 : Calculate (i) $J_p(\underline{\theta}^{j})$, $(\underline{y}-\underline{f})_p$,
(ii) $\underline{g}^{j}$ and G^{j} using (2.8) and (2.9).

Step 2 : Compute the modified Choleski factorization of G which yields
$L^{j}D^{j}(L^{j})' = G^{j} + E^{j}$ (Appendix paragraphs A2 and A3).

Step 3 : (a) If $||\underline{g}^j|| \leq \varepsilon_1$ and $||E^j|| = 0$ then $\underline{\theta}^j$ is optimal and STOP.

(b) If $||\underline{g}^j|| > \varepsilon_1$ determine the search direction by solving the set of linear equations for $\underline{d}^j$.

$$G(\underline{\theta}^j)\underline{d}^j = L^j D^j (L^j)' \underline{d}^j = -\underline{g}^j$$

(c) If $||\underline{g}^j|| \leq \varepsilon_1$ and $||E^j|| \neq 0$ determine $\underline{d}^j$ by the following search procedure: Solve

$$(L^j)'\underline{u} = e_s$$

where $d_s^j - E_{ss}^j = \min \{d_i^j - E_{ii}^j : 1 \leq i \leq k\}$.

$$\text{Set } \underline{d}^j = \begin{cases} -\text{sign}(\underline{u}'\underline{g}^j)\underline{u} & \text{if } ||\underline{g}^j|| > 0 \\ \underline{u} & \text{if } ||\underline{g}^j|| = 0. \end{cases}$$

Step 4 : Calculate γ_j by means of Fletcher's rule (Appendix paragraph A4) and set

$$\underline{\theta}^{j+1} = \underline{\theta}^j + \gamma_j \underline{d}^j .$$

Step 5 : Continue until the following convergence criteria are met :

a) $|g_i^j| < \varepsilon_2 \qquad i=1, \ldots k$

b) $|S_p(\underline{\theta}^{j+1}) - S_p(\underline{\theta}^j)|/|S_p(\underline{\theta}^j) < \varepsilon_3$

in for example 4 consecutive iterations. Otherwise return to Step 1.

$$\underline{g}^\circ = \begin{bmatrix} -215.1514 \\ -88.00 \end{bmatrix} \qquad G^\circ = \begin{bmatrix} 1153.9996 & 480.0 \\ 480.0 & 200.0 \end{bmatrix}$$

$$||\, \underline{g}^\circ \,|| = 232.87 \qquad ||\, E^\circ \,|| = 0.0$$

Step 3(b) $\qquad \underline{d}^\circ = \begin{bmatrix} 2.2 \\ -4.84 \end{bmatrix}$

Step 4: $\gamma_o = 0.072$ and $\underline{\theta}^1 = \begin{bmatrix} -1.042 \\ 0.652 \end{bmatrix}$ and $S_2(\underline{\theta}^1) = 22.95$

and so on.

If we started with the steepest descent step initially i.e. $\underline{d}^\circ = - \nabla S_2(\underline{\theta}^\circ)$ then

$$\underline{\theta}^2 = \begin{bmatrix} -1.0076 \\ 1.0785 \end{bmatrix} \text{ and } S_2(\underline{\theta}^1) = 4.4316 .$$

The algorithm performed 46 function evaluations to reach the optimal solution. The Fletcher (1970) and Jacobson and Oksman (1972) algorithms are established methods for general optimization problems. These two methods performed 47 and 69 function evaluations respectively. A method specifically designed for least squares problems due to Jones (1970) performed 17 function evaluations.

It is not the purpose to illustrate that our method is better than other unconstrained minimization techniques but rather to show that it converges in the same order of number of function evaluations as the established methods. This observation indicates that the algorithm is numerically efficient. The steepest descent step was taken initially and then every 5 iterations to enhance convergence.

It is of interest to note the solution for other values of p. The problem then becomes:

Find parameters θ_1 and θ_2 so that $S_p(\underline{\theta}) = |10(\theta_2 - \theta_1^2)|^p + |1-\theta_1|^p$ is a minimum.

Table 2.1 Solution to Rosenbrock example for differing values of p

	p					
	1.5	1.75	2.0	2.5	2.75	3.0
θ_1	1.0000	1.0000	1.0000	0.9999	0.9996	0.9994
θ_2	1.0000	1.0000	1.0000	0.9998	0.9992	0.9989
$S_p(\underline{\theta})$	0.0	0.0	0.0000	0.0000	0.0000	0.0000
No. evaluations	>100	>100	46	62	60	60

Example 2

This is an example due to Beale (1958) (see also Betts (1976) example 8.4). It is a data-fitting problem.

Find parameters θ_1 and θ_2 so that

$$S_2(\underline{\theta}) = \sum_{i=1}^{3} [y_i - \theta_1(1-\theta_2^i)]^2$$

$$(y_1 = 1.5 \qquad y_2 = 2.25 \qquad y_3 = 2.625)$$

is a minimum.

The usual starting value is $\theta_1^\circ = \theta_2^\circ = 0.1$. The optimal point is located at $\theta_1^* = 3$ and $\theta_2^* = 0.5$ and $S_2(\underline{\theta}^*)=0$. In Figure 2.3 the Beale function is displayed in 3 dimensions. Note that the surface is plotted over the following region $-2 \leq \theta_1 \leq 2$ and $-2 \leq \theta_2 \leq 2$. The surface was also artificially flattened out at $S_2(\underline{\theta}) = 300$. In Figure 2.4 various contours of $S_2(\underline{\theta})$ are given. From this we can see that the optimum is located in the region of the point (3,0.5).

The algorithm performed 6 function evaluations to reach the optimal solution. Betts (1976) reported 10 function evaluations. The best reported up to that time was 20 function and gradient evaluations. Our algorithm shows a marginal improvement in convergence. In Table 2.2 we show the solution for differing values of p (no steepest descent steps were taken).

Table 2.2 Solution to Beale example for differing values of p

	p					
	1.5	1.75	2.0	2.5	2.75	3.0
θ_1	3.0000	3.0000	3	2.9996	2.9995	2.9987
θ_2	0.5000	0.5000	0.5	0.4999	0.4999	0.4998
$S_p(\underline{\theta})$	0.0000	0.0000	0.0	0.0000	0.0000	0.0000
No. evaluations	13	18	6	11	13	15

Example 3

This is a data-fitting example due to Bard (1970) (see also Betts (1976) example 8.7). The problem is to find parameters θ_1, θ_2 and θ_3 so that

$$S_2(\underline{\theta}) = \sum_{i=1}^{15} [y_i - (\theta_1 + \frac{u_i}{\theta_2 v_i + \theta_3 w_i})]^2$$

is a minimum. The data are provided in Table 2.3.

Table 2.3 Bard example data

y_i	u_i	w_i	v_i
0.14	1	1	15
0.18	2	2	14
0.22	3	3	13
0.25	4	4	12
0.29	5	5	11
0.32	6	6	10
0.35	7	7	9
0.39	8	8	8
0.37	9	7	7
0.58	10	6	6
0.73	11	5	5
0.96	12	4	4
1.34	13	3	3
2.10	14	2	2
4.39	15	1	1

The starting value is $\theta_1^\circ = \theta_2^\circ = \theta_3^\circ = 1$. The optimal point is located at $\underline{\theta}^* = (0.08241, 1.1330, 2.3437)'$ and $S_2(\underline{\theta}^*) = 8.214877 \times 10^{-3}$.

Our algorithm performed 6 function evaluations to reach the optimal solution. Betts (1976) reported 7 function evaluations. The best reported up to that time was 36 function and gradient evaluations. Gill and Murray (1978) also reported 6 function and gradient evaluations. In Table 2.4 the solution for differing values of p are shown.

Table 2.4 Solution to Bard example for differing values of p

	p					
	1.5	1.75	2.0	2.5	2.75	3.0
θ_1	0.09617	0.08977	0.08241	0.07115	0.06731	0.06432
θ_2	1.41707	1.2756	1.1330	0.93479	0.87233	0.82516
θ_3	2.07603	2.2098	2.3437	2.5282	2.5858	2.62913
$S_p(\underline{\theta})$	0.031598	0.01632	8.21488^{-3}	1.9470^{-3}	9.3118^{-4}	4.4275^{-4}
No. evaluations	10	8	6	13	12	13

APPENDIX A

A1. Directions of search.

We shall now characterise descent directions by means of the following useful Lemmas which may be found in standard texts in optimization (see e.g Zangwill (1969)).

Lemma A1.1 Suppose the function $S_p(\underline{\theta})$, $\underline{\theta} \in R^k$ is differentiable (i.e. its first order partial derivatives exist) at $\underline{\theta}$ and there exists a direction $\underline{u}$ such that the directional derivative

$$\nabla S_p(\underline{\theta})'\underline{u} < 0,$$

then there exists a sufficiently small constant $a>0$ so that for all $0 < b < a$

$$S_p(\underline{\theta}+b\underline{u}) < S_p(\underline{\theta}) .$$

Proof The DIRECTIONAL DERIVATIVE of $S_p(\underline{\theta})$ at $\underline{\theta}$ is defined as:

$$\lim_{b\to 0} (S_p(\underline{\theta}+b\underline{u}) - S_p(\underline{\theta}))/b = \nabla S_p(\underline{\theta})'\underline{u}$$

which is negative by assumption. By definition of the limit there exist a constant $a > 0$ such that for all $b\neq 0$ and $-a < b < a$

$$(S_p(\underline{\theta}+b\underline{u}) - S_p(\underline{\theta}))/b < 0$$

Choose b to preserve this inequality and the Lemma is proved.

Remark: A direction $\underline{u}$ such that the directional derivative $\nabla S_p(\underline{\theta})'\underline{u} < 0$ is termed a descent direction. Thus a decrease in $S_p(\underline{\theta})$ can be obtained by taking a sufficiently small step in the direction $\underline{u}$.

Lemma A1.2 Direction $\underline{d} = -\nabla^2 S_p(\underline{\theta})^{-1} \nabla S_p(\underline{\theta})$ is a descent direction if $\nabla^2 S_p(\underline{\theta})$ is a positive definite matrix.

Proof If $\nabla^2 S_p(\underline{\theta})$ is positive definite its inverse will also be positive definite. Thus

$$\nabla S_p(\underline{\theta})'\underline{d} = -\nabla S_p(\underline{\theta})'\nabla^2 S_p(\underline{\theta})^{-1} \nabla S_p(\underline{\theta}) < 0$$

since $\nabla^2 S_p(\underline{\theta})$ is positive definite. From Lemma A1.1 it follows that $\underline{d}$ is a descent direction and the Lemma is proved.

Lemma A1.1 shows that a decrease can be obtained by taking a sufficiently small step in the direction $\underline{d}$, provided that $\nabla S_p(\underline{\theta})'\underline{d}<0$. Lemma A1.2 shows that the direction $\underline{d} = -\nabla^2 S_p(\underline{\theta})^{-1} \nabla S_p(\underline{\theta})$ is a descent direction if $\nabla^2 S_p(\underline{\theta})$ is a positive definite matrix.

A2. The Choleski factorisation of a symmetric positive definite matrix.

Consider the system of equations

(A2.1) $A\underline{x} = \underline{b}$,

where A, (mxm) is a positive definite matrix and $\underline{b}$ an mx1 vector of known constants. A numerically stable (i.e. minimizing the effect of rounding errors) method for calculating $\underline{x}$ is to factorise A by the method of Choleski into the form

(A2.2) $A = LDL'$

where L is a unit lower-triangular matrix and D a diagonal matrix. The vector $\underline{x}$ is then computed using forward and backward substitution. Let A_{ij}, ℓ_{ij} denote the ij-th element of A and L respectively and d_j the jj-th element of D (d_j not to be confused with the direction $\underline{d}^j$).

The j-th step of Choleski's method is then given by

(A2.3) $$d_j = a_{jj} - \sum_{k=1}^{j-1} \ell_{jk}^2 d_k,$$

(A2.4) $\ell_{jk} = C_{jk}/d_k$ with the auxiliary quantities C_{ij} given by

(A2.5) $$C_{ij} = a_{ij} - \sum_{k=1}^{j-1} \ell_{ij}\ell_{ik}d_k \quad , \quad i=j+1,\ldots m.$$

Note that $d_1 = a_{11}$ and $C_{i1} = a_{i1}$, i=2,...m. Set T = LD then (A2.2) may be written as $TL'\underline{x} = \underline{b}$. If we set $\underline{y} = L'\underline{x}$ then the m unknowns $\underline{y}$ in the system of linear equations $T\underline{y} = \underline{b}$ can be computed by forward substitution:

$$(A2.6) \qquad y_i = \frac{b_i - \sum_{k=1}^{i-1} t_{ik}\, y_k}{t_{ii}} \quad , \qquad i = 1,2,\ldots,m \ .$$

Finally, the m unknowns $\underline{x}$ in the system

$$L\underline{x} = \underline{y}$$

can be computed by backward substitution:

$$(A2.7) \qquad x_i = \frac{y_i - \sum_{k=i+1}^{m} \ell_{ij}\, x_k}{\ell_{ii}} \quad , \qquad i=m,m-1,\ldots,1$$

$$= y_i - \sum_{k=i+1}^{m} \ell_{ik}\, x_k \quad ,$$

since L is a unit lower triangular matrix.

A3. The modified Choleski factorisation method for insufficiently positive definite matrices.

In this paragraph we shall discuss the modification of the Choleski factorization method which constructs a positive definite matrix from a given matrix which may be indefinite, negative definite or insufficiently positive definite. This method was derived by Gill and Murray (1974).

Equations (A2.4) and (A2.5) show that $\ell_{ij} = C_{ij}/d_j$ may be too large whenever d_j is too small. This will happen when A is insufficiently positive definite. We know that $A = (LD^{\frac{1}{2}})(LD^{\frac{1}{2}})'$ since A is positive definite. The diagonal and off-diagonal elements of $LD^{\frac{1}{2}}$ are therefore $d_j^{\frac{1}{2}}$ and $\ell_{ij}d_j^{\frac{1}{2}}$ respectively. From equation (A2.3)

$$d_j^{\frac{1}{2}} = [a_{jj} - \sum_{k=1}^{j-1} \ell_{jk}^2 d_k]^{\frac{1}{2}} .$$

Now, since d_j is positive it follows that each individual term in the summation above will be smaller than a_{jj}, i.e. $\ell_{ij}d_k^{\frac{1}{2}} < a_{jj}^{\frac{1}{2}}$. This means that no element of $LD^{\frac{1}{2}}$ can exceed $a_{jj}^{\frac{1}{2}}$. Thus, a large element(s) in $LD^{\frac{1}{2}}$ can only occur if A has a large diagonal element(s). A Choleski decomposition will now be described with the following properties:

a) All elements of $LD^{\frac{1}{2}}$ are bounded above by a value β.

b) All elements of D are bounded below by a value δ.

This modified Choleski factorisation will only be carried out if A is insufficiently positive definite. The diagonal elements d_j of D are either retained or modified according to certain rule which will now be discussed, we shall also show how the values of β and δ are calculated. First the following quantities have to be defined:

Let ζ = maximum in modulus of the off-diagonal elements of A

$$= \max \{|a_{ij}| : i=j+1, \ldots m\}.$$

η = maximum in modulus of the diagonal elements of A

$$= \max \{|a_{jj}| : j=1, \ldots m\}.$$

θ_j = $\max \{|C_{ij}| : i=j+1, \ldots m\}$.

EPS = relative machine precision (i.e. the smallest positive machine representable floating point number).

Gill and Murray (1974) show that the choice

$\beta^2 = \max \{\zeta/m, \eta, \text{EPS}\}$

δ = $\max \{\text{EPS}.||A||, \text{EPS}\}$ and

d_j^* = $\max \{\delta, |d_j|, \theta_j^2/\beta^2\}$

automatically ensures that $|\ell_{ij}\, \tilde{d}_j^{\frac{1}{2}}| \leq \beta$ where $\tilde{d}_j = \max\{\delta, |d_j|\}$. This simply states that the terms of $LD^{\frac{1}{2}}$ are bounded above by β. We see that the choice of β^2 and d_j^* prevents the off-diagonal elements of A from becoming too large and the diagonal elements of A from becoming too small. We shall see that if the matrix A is sufficiently positive definite (i.e. when $d_j > \delta$) no modification to d_j is made, i.e. $d_j^* = d_j = |d_j| > 0$. Hence the stability of the factorisation is assured. We therefore select:

$$\text{(A3.1)} \qquad d_j^* = \begin{cases} \delta \text{ if } \delta \geq \max\{|d_j|, \theta_j^2/\beta^2\} \\ |d_j| \text{ if } |d_j| \geq \max\{\theta_j^2/\beta^2, \delta\} \\ \theta_j^2/\beta^2 \text{ if } \theta_j^2/\beta^2 \geq \max\{\delta, |d_j|\}. \end{cases}$$

The factors obtained by the modified procedure are (cf. Gill and Murray, 1974) identical to those obtained by applying Choleski's method to the matrix

$$A^* = A + E$$

where E is a diagonal matrix with a typical element

$$\text{(A3.2)} \qquad E_{ii} = \begin{cases} \delta - d_j \text{ if } \delta \geq \max\{|d_j|, \theta_j^2/\beta^2\} \\ |d_j| - d_j \text{ if } |d_j| \geq \max\{\theta_j^2/\beta^2, \delta\} \\ \theta_j^2/\beta^2 - d_j \text{ if } \theta_j^2/\beta^2 \geq \max\{\delta, |d_j|\}. \end{cases}$$

If A is sufficiently positive definite then E=0. This can be seen if d_j is the largest of the quantities

$$\theta_j^2/\beta^2 \quad \text{and} \quad \delta$$

then $d_j = |d_j|$ whence $E_{ii} = 0$

and no modification is therefore made.

In the case where matrix $G = \nabla^2 S_p(\underline{\theta})$ is indefinite, further examination is required. We observe that if G is negative definite, i.e. when $d_j < 0$, then d_j is replaced by $|d_j|$.

The direction

$$\underline{d}^j = -G^*(\underline{\theta}^j)^{-1}\underline{g}^j \quad \text{with } G^* = LDL^T = G + E$$

will be zero when the gradient $\underline{g}^j$ is zero. If G, however, is indefinite then $\underline{\theta}^j$ will not be local (isolated) minimum. An alternative direction or so-called direction of negative curvature has to be used whenever the norm, $||\underline{g}^j|| > \varepsilon$ where ε (=0.1) is some prescribed tolerance. This procedure will therefore prevent convergence to saddlepoints, i.e. when G is indefinite and $||\underline{g}^j|| = 0$. The information regarding the indefiniteness of G is already available in the modified Choleski factorization of G.

Definition: A direction $\underline{u}$ is a DIRECTION OF NEGATIVE CURVATURE with respect to the indefinite matrix G if

$$\underline{u}'G\underline{u} < 0 \ .$$

The following postulates are stated without proof (see Gill and Murray op. cit.).

1. Let s be any integer such that $d_s^* \leq \min \{d_i^* : 1 \leq i \leq m\}$. If G is indefinite then $d_s^* < 0$.

2. If G is indefinite and $d_s^* = 0$, then we can set $\ell_{sj} = 0$ for $j=1,\ldots q-1,q+1,\ldots s-1$ and $\ell_{sq}^2 \, d_q^* = \beta^2$ for some q such that $1 \leq q \leq m$.

3. Suppose $\underline{\theta}^j$ is such that $||\underline{g}^j|| = 0$ and G is indefinite. Let $\underline{u}^*$ be a solution to the equation $L'\underline{u}^* = e_s$ (e_s is the s-th unit co-ordinate vector) with $d_s^* \leq \min \{d_i^* : 1 \leq i \leq m\}$, then $\underline{u}^*$ is a direction of negative curvature.

Remark: The following is not explicitly stated in Gill and Murray op. cit. In postulate 3 it is assumed that $||\underline{g}^j|| = 0$. If, however, $||\underline{g}^j|| > 0$ and $(\underline{u}^*)'\underline{g}^j > 0$ (indicating an increase in $S_p(\underline{\theta})$ in the direction $\underline{u}^*$) then $-\underline{u}^*$ will be a descent direction. The following decision rule will therefore be used:

$$(A3.3) \quad \underline{u}^* = \begin{cases} -\text{sign}((\underline{u}^*)'\underline{g}^j)\underline{u}^* & \text{if } 0<||\underline{g}^j||<\varepsilon \\ \\ \underline{u}^* & \text{if } ||\underline{g}^j||= 0 \end{cases}$$

where $\varepsilon>0$ is a small positive number.

A4. One-dimensional line search algorithms

These methods are generally known as steplength algorithms. Recall that in Step 4 of our algorithm we set

$$\underline{\theta}^{j+1} = \underline{\theta}^j + \gamma_j \underline{d}^j$$

where the vectors $\underline{\theta}^j$ and $\underline{d}^j$ are known. All we need to determine is the value of the scalar $\gamma_j > 0$.

a) Minimization procedure

Minimization (exact line search) is perhaps the most well known. In this case the maximum possible decrease in $S_p(\underline{\theta})$ occurs for a given $\underline{d}^j$ if we perform the one-dimensional minimization:

$$(A4.1) \quad S_p(\underline{\theta}^j + \gamma_j \underline{d}^j) = \min \{S_p(\underline{\theta}^j + \gamma \underline{d}^j): \gamma > 0\}$$

where $\underline{d}^j$ is a descent direction. The main drawback of this procedure is the number of function evaluations required in the minimization process.

Through the years various line search (inexact as well as exact) procedures have been proposed. e.g Fibonacci search, quadratic and cubic interpolation methods etc. A discussion of these methods may be found in standard nonlinear programming texts (see e.g. Luenberger (1973), Avriel (1976) and Fletcher (1980)).

b) Armijo-Goldstein algorithm

This method (Armijo (1966)) is based on the Goldstein-Armijo principle:

A sufficient decrease in $S_p(\underline{\theta}^j)$ is obtained if γ_j satisfies

$$\mu_2\gamma_j(\underline{g}^j)'\underline{d}^j \leq S_p(\underline{\theta}^j + \gamma_j\underline{d}^j) - S_p(\underline{\theta}^j) \leq \mu_1\gamma_j(\underline{g}^j)'\underline{d}^j \tag{A4.2}$$

where μ_1 and μ_2 are scalars and $0 < \mu_1 \leq \mu_2 < 1$. The bounds ensure that γ_j is neither too small or too large.

Algorithm:

Step 0 : Select a trial value γ_j (=1).

Step 1 : If $S_p(\underline{\theta}^j + \gamma_j \underline{d}^j) - S_p(\underline{\theta}^j) \leq \mu_1 \gamma_j (\underline{g}^j)' \underline{d}$, STOP.

Otherwise proceed.

Step 3 : Set $\gamma_j := w\gamma_j$ and return to Step 1.

Typical choices for the parameters are μ_1=0.0001 and w=0.1. This rule is feasible if we examine Lemma A1.1.

c) Fletcher's line search algorithm

The line search algorithm due to Fletcher (1970) makes use of gradient information. The following condition is required to ensure that the magnitude of the gradient is sufficiently increased away from $\underline{\theta}^j$, i.e.

(A4.3) $$|\underline{g}^j(\underline{\theta}^j + \gamma_j \underline{d}^j)' \underline{d}^j| > -\rho(\underline{g}^j)' \underline{d}^j \text{ for } 0 < \rho < 1.$$

A value of ρ = 0.9 gives a weak line search and ρ=0.1 gives an accurate line search.

Algorithm :

Step 0 : Let ΔS_p be a user supplied estimate of the likely reduction in $S_p(\underline{\theta})$. Set $S_1 = S_p(\underline{\theta}^j)$.

Step 1 : Calculate $\gamma = \min (1,-2\ \Delta S_p/(\underline{g}^j)'\underline{d}^j)$.

Step 2 : Calculate $\underline{\theta}_N = \underline{\theta}^j + \gamma\underline{d}^j$, $S_2 = S_p(\underline{\theta}_N)$ and

$$\underline{g}_2'\underline{d}^j = \underline{g}(\underline{\theta}_N)'\underline{d}^j .$$

Step 3 : If $S_2 \geq S_1$ go to Step 7, otherwise go to Step 4.

Step 4 : If $|\underline{g}_2\ '\underline{d}^j/(\underline{g}^j)'\underline{d}^j| < \rho$ STOP. The optimal value of γ has been found. Otherwise proceed.

Step 5 : If $\underline{g}_2'\underline{d}^j > 0$ go to Step 7, otherwise proceed.

Step 6 : If $(\underline{g}^j)'\underline{d}^j < \underline{g}_2'\underline{d}^j$ set

$$\gamma = \gamma \min(10,\ \underline{g}_2'\underline{d}^j/((\underline{g}^j)'\underline{d}^j - \underline{g}_2'\underline{d}^j)$$

otherwise set $\gamma := 10\gamma$, return to Step 2.

Step 7: Cubic interpolation.

Calculate $$z = \frac{3}{\gamma}(S_1 - S_2) + \underline{g}_2'\underline{d}^j + (\underline{g}^j)'\underline{d}^j$$

$$w = (z^2 - \underline{g}_2'\underline{d}^j(\underline{g}^j)'\underline{d}^j)$$

$$\gamma := \gamma \left(1 - \frac{w - z + \underline{g}_2'\ \underline{d}^j}{\underline{g}_2'\ \underline{d}^j - (\underline{g}^j)'\underline{d}^j} \right)$$

Return to Step 2.

A5. Quadratic behaviour of a nonlinear function in a small neighbourhood of a local minimum.

Consider the quadratic function $f(x) = \frac{1}{2}\, \underline{x}'Q\underline{x} + \underline{b}'\underline{x} + c$ where Q is an nxn matrix, $\underline{x}$ and $\underline{b}$ are nx1 vectors and c a scalar (1x1). Given two points $\underline{x}^1$ and $\underline{x}^2$ then

$$(A5.1) \qquad g(\underline{x}^1) - g(\underline{x}^2) = Q(\underline{x}^1 - \underline{x}^2) .$$

Consider a general function $f(\underline{x})$, then by a first-order Taylor series expansion:

$$f(\underline{x}^* + \underline{h}) \simeq \nabla f(\underline{x}^*) + \nabla^2 f(\underline{x}^*)\underline{h}$$

$$\text{i.e} \qquad \nabla^2 f(\underline{x}^*)\underline{h} \simeq \nabla f(\underline{x}^* + \underline{h}) - \nabla f(\underline{x}^*) .$$

Relation (A5.1) is also known as the quasi-Newton or secant relation for nonlinear functions where $\underline{g} = \nabla f$ and $Q = \nabla^2 f$. It illustrates the quadratic behaviour of the nonlinear function in the neighbourhood of an optimal point $\underline{x}^*$.

Definition A5.1: A set $F \subset R^k$ is convex if $\underline{x},\underline{y} \in F \Rightarrow \underline{w} = \lambda\underline{x} + (1-\lambda)\underline{y} \in F$ whenever $0 \leq \lambda \leq 1$.

Definition A5.2: A function $f(\underline{x})$ is CONVEX on a convex set F if $\underline{x},\underline{y} \in F$ $\Rightarrow f(\lambda\underline{x} + (1-\lambda)\underline{y}) \leq \lambda f(\underline{x}) + (1-\lambda)f(\underline{y})$ for $0 \leq \lambda \leq 1$. A function will be strictly convex if the strict inequality holds for $0 < \lambda < 1$.

The following theorems characterise convexity. Proofs may be found in Mangasarian (1969).

Theorem A5.1: Let $f(\underline{x})$ be twice continuously differentiable then the Hessian matrix of $f(\underline{x})$ is positive semi-definite if and only if $f(\underline{x})$ is convex. It will be positive definite iff $f(\underline{x})$ is strictly convex.

Theorem A5.2: Let $f(\underline{x})$ be a convex function on a convex set $F \subset R^k$ then any local minimum of $f(\underline{x})$ in F will also be a global minimum of $f(\underline{x})$ over F. If $f(\underline{x})$ is strictly convex on F then there exists at the most one global minimum point of $f(\underline{x})$ over F.

Fig. 2.1: Plot of the Rosenbrock function (p=2)

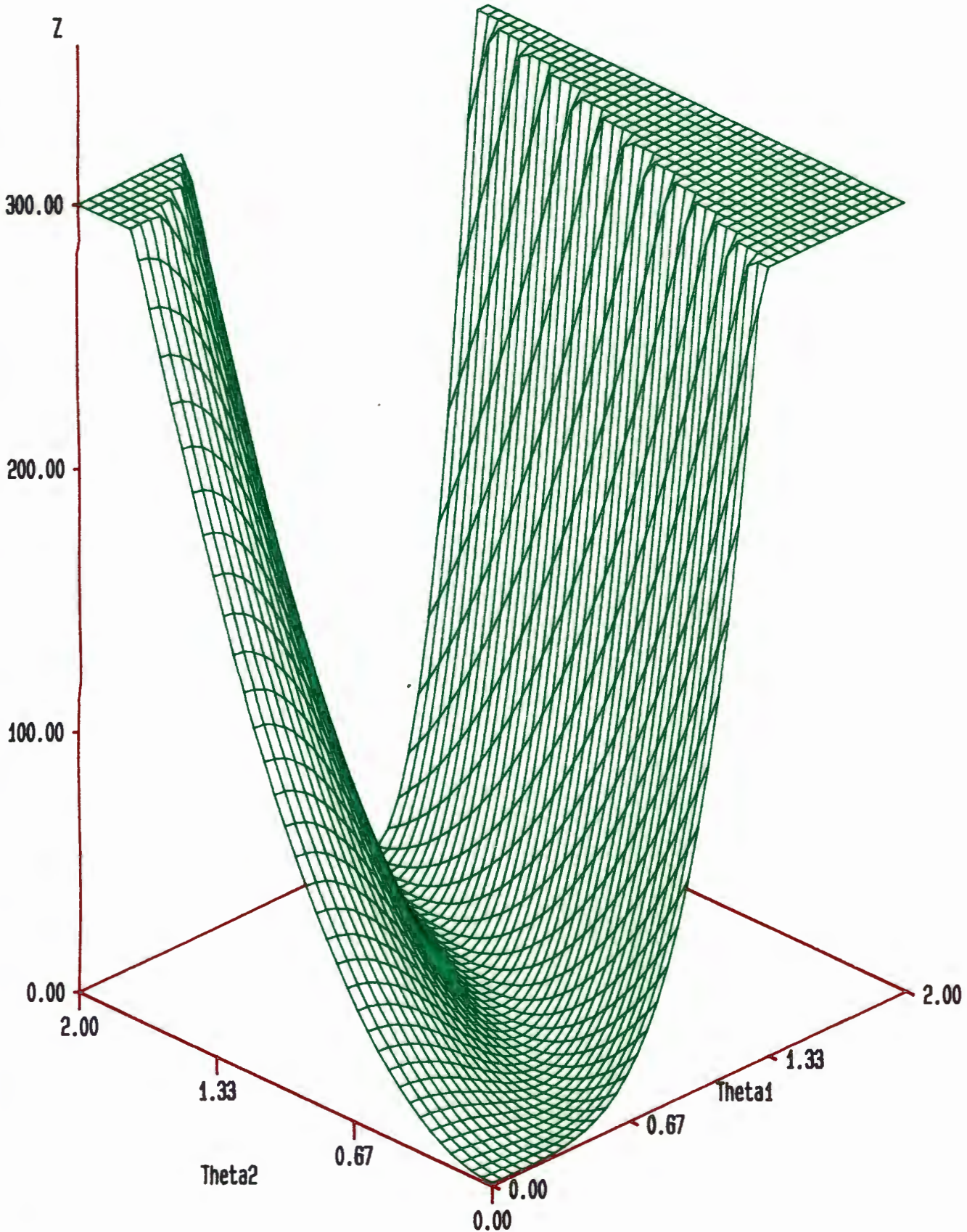


Fig. 2.2: Contours of the Rosenbrock function (p=2)

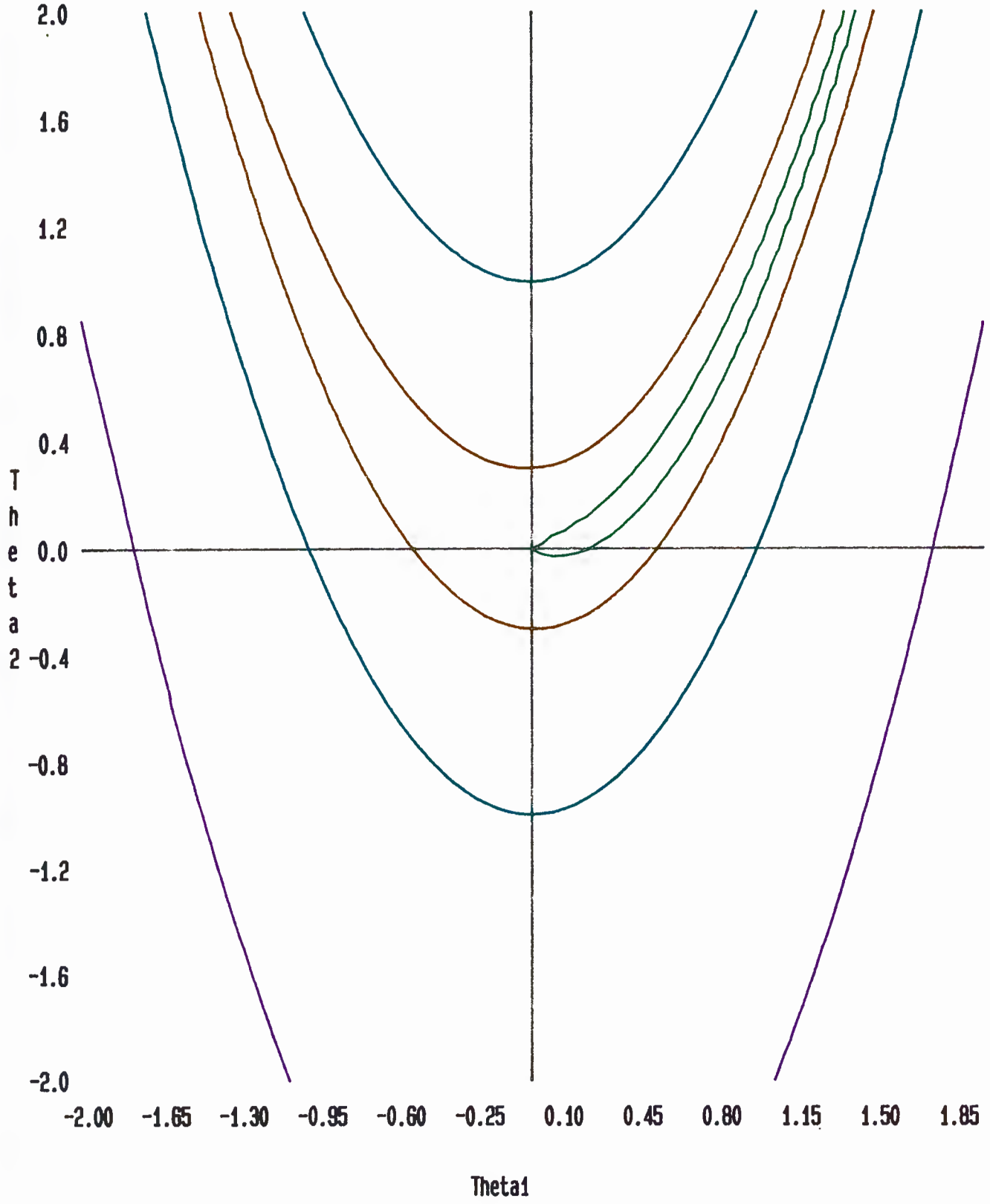


Fig. 2.3: Plot of the Beale example (p=2)

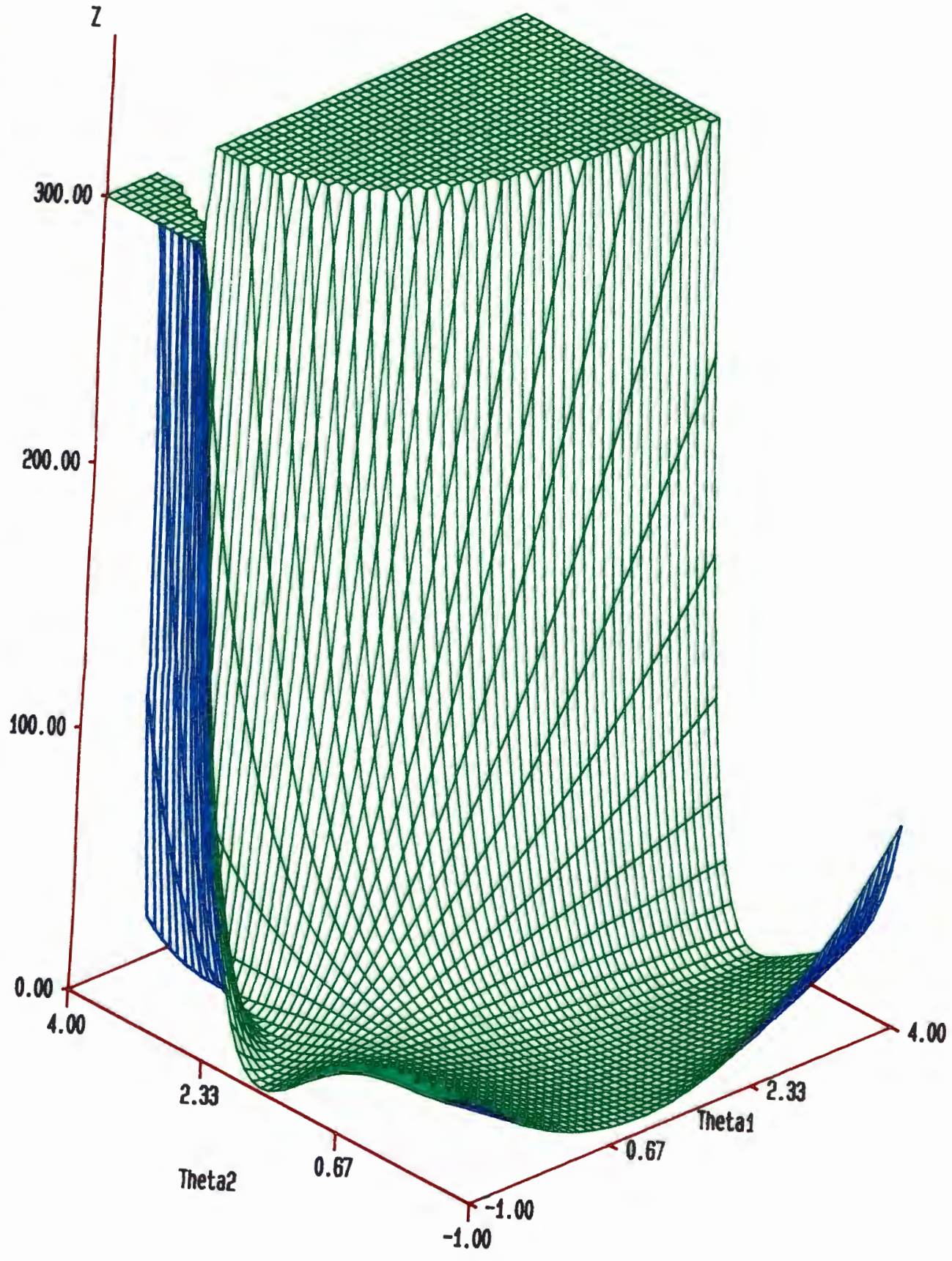
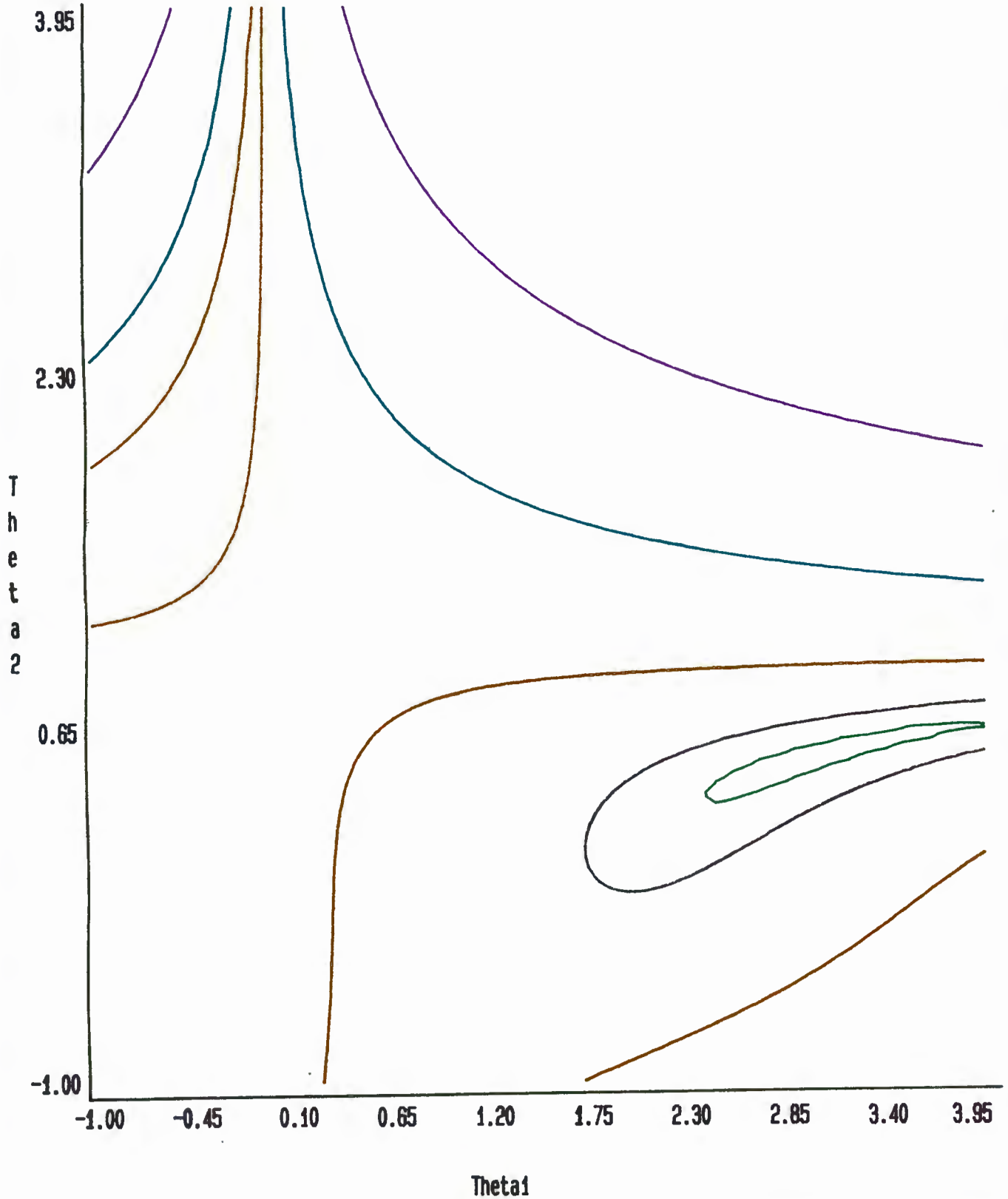


Fig. 2.4: Contours of the Beale example (p=2)



Chapter 3 : A numerical algorithm for solving the large residual nonlinear L_p-norm estimation problem.

In the previous chapter an algorithm was derived for solving small residual problems and examples were given to show that the convergence to the optimal solution is satisfactory. However, poor convergence was experienced on large residual problems and problems that are ill-conditioned by nature (e.g. models involving sums of exponential terms).

In both large residual and in ill-conditioned problems second-derivative information has to be taken into account. Gauss-Newton methods as well as our first-order gradient method ignore second-order derivative information hence poor convergence behaviour occurred in both these cases.

In this chapter we shall derive a second nonlinear L_p-norm estimation algorithm which uses a mixture of first-order gradient (Gauss-Newton) and Newton search directions. The algorithm employs numerical differentiation and utilises second-order derivative information as sparingly as possible since it is expensive in terms of computer calculations. It will be shown that the large residual nonlinear least squares algorithm of Gill and Murray (1978) is imbedded in the new algorithm.

Singular-value decomposition of matrix J_p will be used to stabilize rounding error (note that this is in addition to Choleski factorization which was introduced for the same reason). In the procedure the singular values of matrix J_p will be examined. A first-order gradient (Gauss-Newton) direction $\underline{d}_1$ corresponding to the dominant singular values will then be computed.

This is perfectly valid since the dominant singular values correspond to that part of the p-Jacobian (and indirectly the Hessian) which is well-conditioned. The direction $\underline{d}_1$ will therefore be based on first-order derivative information only. For the remaining non-dominant singular values the second-order derivative information is used and a direction $\underline{d}_2$ based on first- and second-order information is calculated. We therefore only calculate second-order information where it is absolutely necessary.

In geometrical terms, $\underline{d}_1$ may be regarded as the first-order gradient (Gauss-Newton) direction in the subspace spanned by the dominant singular vectors of the p-Jacobian and $\underline{d}_2$ the Newton direction in the subspace spanned by the non-dominant singular vectors of the p-Jacobian. The arithmetic sum of the directions, $\underline{d}_1 + \underline{d}_2 = \underline{d}$ is then used as a descent direction by the algorithm. The choice involving $\underline{d}_1$ and $\underline{d}_2$ constitutes the mixture algorithm.

Numerical examples will be used to illustrate the feasibility and efficiency of the algorithm. The steps of the algorithm will be illustrated by intermediate calculations. Graphical displays of specific examples will also be provided.

To promote the smooth flow of thought, certain mathematical concepts to be used in the remainder of this chapter will be reserved for Appendix B. The following concepts will be outlined: rank of matrix, vector and matrix norms, eigenvalue - eigenvector decomposition of a symmetric matrix, singular-value decomposition of a matrix, Gram-Schmidt orthogonalization, Householder orthogonal transformations and convergence rates of algorithms.

1. Nonlinear least squares

We shall digress slightly to discuss nonlinear least squares problems. In general nonlinear least squares problems may be classified as follows:

(a) The small residual problem with rank (J)=k (i.e. the Jacobian matrix is of full column rank).

(b) The large residual problem with or without rank deficiency of the Jacobian matrix J.

(c) The small residual problem with an ill-conditioned Jacobian matrix in which the eigenvalues of the second-order derivative matrix B_2 may differ in sign from the eigenvalues of J.

In large residual problems notably the example due to Jennrich and Sampson (1968) (see also Betts (1976) example 8.8) the conventional Gauss-Newton methods often fail to converge. This problem can be explained as follows:

Recall that in Chapter 2 the gradient and Hessian of $S_2(\underline{\theta})$ in nonlinear least squares were given by expressions (2.6) and (2.7) respectively.

$$\nabla S_2(\underline{\theta}) = -2J'(\underline{y}-\underline{f})$$
$$\nabla^2 S_2(\underline{\theta}) = 2[J'J + B_2(\underline{\theta})]$$

where $J = J(\underline{\theta}) = \left[\frac{\partial f_i}{\partial \theta_j} \right]$ $\quad i=1,\ldots n \; j=1,\ldots k$

and $B_2(\underline{\theta}) = \sum_{i=1}^{n} (f_i - y_i)\nabla^2 f_i$, $\nabla^2 f_i$ the Hessian of f_i with respect to $\underline{\theta}$.

2. Derivation of the large residual mixture method

In order to derive the descent direction, the derivatives will be examined for the following conditions:

(a) Matrix $(p-1)J_p'J_p + B_p(\underline{\theta})$ is positive definite.

(b) Matrix $J_p'J_p$ is ill-conditioned.

(c) Matrix $(p-1)J_p'J_p + B_p(\underline{\theta})$ is indefinite.

(d) Matrix $(p-1)J_p'J_p + B_p(\underline{\theta})$ is negative definite.

In case (a) the singular value decomposition will be followed by a Choleski factorization and the usual first-order method will result. In case (b) the mixture method will be used. This depends on the number of dominant singular values of J_p. In case (c) the modified Choleski factorisation will be used to provide a direction of descent whilst in case (d) the search direction will be reversed (i.e. the direction will be replaced by minus its original value) with a special safeguard when the grade of J_p (number of dominant singular values of J_p) is positive.

Reconsider the L_p-norm estimation problem (2.2) and the gradient and Hessian expressions:

$$\nabla S_p(\underline{\theta}) = pJ_p'(\underline{f}-\underline{y})_p \tag{3.1}$$

$$\nabla^2 S_p(\underline{\theta}) = p(p-1)J_p'J_p + pB_p(\underline{\theta}). \tag{3.2}$$

A second-order Taylor series expansion of $S_p(\underline{\theta})$ about $\underline{\theta}^j$ yields:

$$(3.3) \quad S_p(\underline{\theta}) \simeq S_p(\underline{\theta}^j) + (\underline{\theta} - \underline{\theta}^j)'\nabla S_p(\underline{\theta}^j) + \tfrac{1}{2}(\underline{\theta} - \underline{\theta}^j)' \nabla^2 S_p(\underline{\theta}^j)(\underline{\theta} - \underline{\theta}^j)$$

The same argument may be used as that employed in the derivation of the classical Newton-Raphson method.

A necessary condition (first-order necessary condition) for $\underline{\theta}$ to be a local minimum of $S_p(\theta)$ is that:

$$(3.4) \quad \nabla S_p(\underline{\theta}) = \underline{0} .$$

Differentiation of (3.3) with respect to $\underline{\theta}$ yields:

$$\nabla S_p(\underline{\theta}^j) + \nabla^2 S_p(\underline{\theta}^j)(\underline{\theta} - \underline{\theta}^j) \simeq \underline{0} .$$

If we define our Newton search direction as

$$\underline{d}^j = \underline{\theta} - \underline{\theta}^j \quad \text{then}$$

$$(3.5) \quad \underline{d}^j = -[\nabla^2 S_p(\underline{\theta}^j)]^{-1} \nabla S_p(\underline{\theta}^j) .$$

In the previous chapter we discussed a Gauss-Newton type of method in which the Hessian matrix $\nabla^2 S_p(\underline{\theta}^j)$ was approximated by first-order derivative terms only i.e. $p(p-1)J_p'J_p$. In our experience with large residual problems this approximation is inadequate and hence second-order derivative information will be incorporated into our algorithm.

The direction $\underline{d}^j$ can therefore be calculated from:

$$\nabla^2 S_p(\underline{\theta}^j)\ \underline{d}^j = -\nabla S_p(\underline{\theta}^j)$$

which is equivalent to

$$[\,p(p-1)J_p'J_p + pB_p(\underline{\theta})]\underline{d}^j = pJ_p'(\underline{y}-\underline{f})_p = -pJ_p'(\underline{f}-\underline{y})_p$$

or

$$(3.6)\qquad [\,(p-1)J_p'J_p + B_p(\underline{\theta}^j)]\,\underline{d}^j = J_p'(\underline{y}-\underline{f})_p = -J_p'(\underline{f}-\underline{y})_p\ .$$

Note the correspondence between direction $\underline{d}^j$ in (3.6) and the direction in nonlinear least squares

$$[\,J'J + B_2(\underline{\theta}^j)\,]\,\underline{d}^j = -J'(\underline{f}-\underline{y})\ .$$

We shall see that because of this correspondence, the nonlinear least squares algorithm will be imbedded in the mixture algorithm.

A numerically stable (minimizing the rounding error) method for calculating $\underline{d}^j$ from the system of equations (3.5) involves the singular-value decomposition process. The singular-value decomposition of matrix J_p enables us to write matrix J_p as the product of three matrices:

$$(3.7) \qquad J_p = U \begin{bmatrix} D(s) & 0 \\ 0 & 0 \end{bmatrix} V'$$

where matrix U is nxn, D is nxk and V' is kxk. A more detailed discussion of singular-value decomposition is provided in Appendix B.

Let $B_p = B_p(\underline{\theta})$ and substitute (3.7) into (3.6) thus yielding:

$$[(p-1)\ VD'U'UDV' + B_p]\underline{d}^j = -VD'U'(\underline{f}-\underline{y})_p$$

which becomes

$$[(p-1)\ VD^2(s)V' + B_p]\underline{d}^j = -V[D(s):0]U'(\underline{f}-\underline{y})_p \ .$$

Premultiplication by V' yields

$$(3.8) \quad [(p-1)\ D^2(s)V' + V'B_p]\underline{d}^j = -[D(s) : 0]\ U'(\underline{f}-\underline{y})_p \ .$$

Define $\underline{d}^j = V\underline{z}^j$, then (3.8) may be rewritten as:

$$(3.9) \quad [(p-1)\ D^2(s) + V'B_pV]\underline{z}^j = -[D(s) : 0]\ U'(\underline{f}-\underline{y})_p \ .$$

The following distinct cases have to be considered at this stage:

(a) $(p-1)J_p'J_p + B_p$ is positive definite

In this event the matrix

$$(3.10) \qquad (p-1)D^2(s) + V'B_pV$$

in expression (3.9) will be positive definite and hence the Choleski factorization (Appendix A, paragraph A3)

$$LDL' = (p-1)D^2(s) + V'B_pV$$

can be used to calculate $\underline{z}^j$ from (3.9). In the event that the matrix in (3.10) is insufficiently positive definite the modified Choleski method may be used.

(b) Matrix $J_p'J_p$ is ill-conditioned

In the event that $J_p'J_p$ is ill-conditioned; this ill-conditioning will be reflected in the matrix

$$D^2(s) + V'B_pV \quad \text{especially when } ||B_p|| \ll ||J_p'J_p|| .$$

This ill-conditioning often occurs in data-fitting problems with small residuals. The approach is based on an idea by Gill and Murray and has been adapted for the case $p \neq 2$.

A procedure to determine the grade of J_p will be described subsequently. Although at present the process of determining r has not altogether been resolved, we shall describe a procedure which has worked well in practice.

Partition matrix D(s) into two submatrices:

D_1 : rxr = diag $(s_1, s_2, \ldots s_r)$

D_2 : (k-r)x(k-r) = diag $(s_{r+1}, \ldots s_n)$.

Similarly partition V into

$V = [V_1 : kxr; \quad V_2 : kx(k-r)]$ and

$$\underline{z} = \begin{bmatrix} \underline{z}_1 : & rx1 \\ \\ \underline{z}_2 : & (k-r)x1 \end{bmatrix}$$ and

$$U'(\underline{y}-\underline{f})_p = \begin{bmatrix} \underline{f}_1 : & rx1 \\ \underline{f}_2 : & (k-r)x1 \\ \underline{f}_3 : & (n-k)x1 \end{bmatrix}$$

where $\underline{f}_3$ is a dummy component not to be used any further.

Note that the direction $\underline{d}^j$ will then be the sum of two components:

$$\underline{d}^j = V_1\underline{z}_1 + V_2\underline{z}_2 = \underline{d}_1 + \underline{d}_2 .$$

Substitution of these partitions into (3.8) yields the following systems of linear equations:

$$\left((p-1)\begin{bmatrix} D_1^2 & 0 \\ 0 & D_2^2 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix} + \begin{bmatrix} V_1' \\ V_2' \end{bmatrix} B_p\right)\left(V_1\underline{z}_1 + V_2\underline{z}_2\right) = \begin{bmatrix} -D_1\underline{f}_1 \\ -D_2\underline{f}_2 \end{bmatrix}$$

or

$$(3.11)\quad (p-1)D_1^2\underline{z}_1 + V_1'B_p(V_1\underline{z}_1 + V_2\underline{z}_2) = -D_1\underline{f}_1$$

$$(3.12)\quad [(p-1)D_2^2 + V_2'B_pV_2]\,\underline{z}_2 + V_2'B_pV_1\underline{z}_1 = D_2\underline{f}_2 .$$

Recall that our partitioning proceeded according to the number of dominant singular values r and that $||B_p|| << ||J_p'J_p||$ or to be more precise $||B_p||$ is small in comparison to $||(p-1)J_p'J_p||$. Hence the contribution of the term

$$V_1'B_p(V_1\underline{z}_1 + V_2\underline{z}_2) = V_1'B_p(\underline{d}_1 + \underline{d}_2) = V_1'B\underline{d}^j$$

in expression (3.11) will be small and may be neglected. We may therefore solve $\underline{z}_1$ from the approximate system of equations

$$(p-1)\, D_1^2\underline{z}_1 \simeq -D_1\underline{f}_1$$

i.e. $$\underline{z}_1 = -D_1^{-1}\underline{f}_1/(p-1)$$

and calculate

$$(3.13) \qquad \underline{d}_1 = V_1\underline{z}_1 = -V_1D_1^{-1}\underline{f}_1/(p-1) \ .$$

We observe that $\underline{d}_1$ is the first-order gradient (Gauss-Newton) direction in the subspace spanned by V_1. Substitution of (3.13) into (3.12) yields

$$(3.14) \qquad [(p-1)D_2^2 + V_2'B_pV_2]\underline{z}_2 = -D_2\underline{f}_2 - V_2'B_p\,\underline{d}_1$$

$$= -D_2\underline{f}_2 + V_2'B_pV_1D_1^{-1}\underline{f}_1/(p-1)$$

We can therefore solve for $\underline{z}_2$ from (3.14) once $\underline{d}_1$ is known and calculate $\underline{d}_2 = V_2\underline{z}_2$. Note that $\underline{d}_2$ is the Newton direction in the subspace spanned by V_2. Our direction is therefore $\underline{d}^j = \underline{d}_1 + \underline{d}_2$ a sum of a first-order gradient (Gauss-Newton) and full Newton directions respectively. This direction constitutes the mixture algorithm to be stated subsequently.

(c) Matrix $(p-1)J_p'J_p + B_p$ is indefinite

In this case $\underline{d}^j$ will no longer be satisfactory as a search direction since it may not be a descent direction. The modified Choleski method may again be used since it will provide a direction of negative curvature (Appendix A paragraph A3). Recall that information regarding the indefiniteness of matrix $(p-1)J_p'J_p + B_p$ or $(p-1)D_2^2 + V_2'B_pV_2$ will be available in its modified Choleski factorization. The computation of such a direction of negative curvature was described in step 3(c) of the algorithm in the paragraph 2 of the previous chapter.

(d) Matrix $(p-1)J_p'J_p + B_p$ is negative definite

In this event the modified Choleski factorization of $(p-1)J_p'J_p + B_p$ will simply reverse the direction to one of descent. Recall from the previous chapter, notably expressions (2.8) and (2.9); that if $(p-1)J_p'J_p + B_p$ is negative definite then the diagonal elements of D in LDL' will be negative and in the modified Choleski method d_j will be replaced by $|d_j|$. This is equivalent to replacing matrix $(p-1)J_p'J_p + B_p$ by $-(p-1)J_p'J_p - B_p$. Hence our search direction will be a descent direction. Note that d_j refers to the diagonal elements of D in LDL' not to be confused with D(s) or $\underline{d}^j$. The rest of the procedure will follow as described in Chapter 2.

Another important point has to be considered. Suppose the matrix $(p-1)J_p'J_p + B_p$ has grade r and that it is negative definite (i.e. k negative eigenvalues). The modified Choleski method is only applied to the portion

$(p-1)D_2^2 + V_2'B_pV_2$ hence only k-r eigenvalues will be made positive and not the remaining r eigenvalues. As a safeguard Gill and Murray op.cit. suggest the use of the projected gradient $(\underline{g}^j)'\underline{d}^j$. If the quantity

$$-(\underline{g}^j)'\underline{d}^j/(||\underline{g}^j||\ ||\underline{d}^j||) < \eta < ||D_1^{-2}||$$

with η some small positive constant, then the direction $\underline{d}^j$ is recomputed by setting r=0 and proceeding as before. In this event B_p need not be recomputed and $\underline{d}_1$ will be set equal to zero.

We are now in a position to state a numerical algorithm for solving the nonlinear estimation problem (2.2). Let tol, gtol, Ftol, ε, η be prescribed tolerances (for example tol=0.1*, gtol=Ftol=10^{-9}, $\varepsilon = 10^{-9}$, $\eta = 0.0001$). Denote the ii-th element of B_p by B_{ii}.

The ALGORITHM

Step 0 : Set j:=0 and select the initial estimate $\underline{\theta}^\circ$ of $\underline{\theta}$*.

Step 1 : Calculate (i) the p-Jacobian $J_p(\underline{\theta}^j)$

(ii) the p-residual vector $(\underline{y}-\underline{f})_p$

(iii) the gradient of $S_p(\underline{\theta}^j)$;

$\underline{g}^j = -pJ_p(\underline{\theta}^j)'(\underline{y}-\underline{f})_p$ using (2.8).

* In the Fortran program tol may be defined according to values taken on by ρ_1. This definition is purely heuristic and by no means the best.

Step 2 : Compute the singular-value decomposition of $J_p(\underline{\theta}^j)$ viz: $U \begin{bmatrix} D(s) & 0 \\ 0 & 0 \end{bmatrix} V'$ (Appendix B, paragraph B1).

Step 3 : Calculate the grade r as follows:

$$\rho_1 = ||B_{ii}||/||(p-1)J_p'J_p||$$

$$\rho_2 = \min \{s_{i+1}/s_i : s_i > 0,\ i=1,\ldots k\}$$

Set $r = i_{min}$

If $\rho_2 >$ tol set r = k (hence a full first-order gradient step is taken).

Step 4 : If grade r=0 then $\underline{d}_1 = 0$. Take a full Newton step. Go to Step 5.

For r > 0. Calculate the first-order gradient direction

$$\underline{d}_1 = V_1 D_1^{-1} \underline{f}_1/(p-1)$$

If grade r=k set $\underline{d}^j = \underline{d}_1$, $\underline{d}_2 = \underline{0}$ and take a full first-order gradient step. Go to Step 8.

If grade $0 < r < k$ go to Step 5.

Step 5 : (a) Compute the second derivative matrix $B_p(\underline{\theta}^j)$ numerically.

(b) Then compute the modified Choleski factorization of

$$(p-1)D_2^{\ 2} + V_2'B_p(\underline{\theta}^j)V_2 + E^j = L^j\tilde{D}^j(L^j)'$$

(Appendix A paragraphs A2 and A3. $\tilde{D}$ not to be confused with D(s)). Go to step 6.

Step 6 : (a) If $||\underline{g}^j|| \leq \varepsilon$ and $||E^j|| = 0$ then $\underline{\theta}^j$ is optimal and STOP.

(b) If $||\underline{g}^j|| > \varepsilon$ determine the search direction by solving the set of linear equations for $\underline{z}_2$:

$$L^j\tilde{D}^j(L^j)\underline{z}_2 = -D_2\underline{f}_2 - V_2'B_p\underline{d}_1$$

(c) If $||\underline{g}^j|| \leq \varepsilon$ and $||E^j|| \neq 0$ determine $\underline{z}_2$ by means of the following search procedure:

Let q be the subscript for which

$$\tilde{d}_q^j - E_{qq}^j = \min\{\tilde{d}_i^j - E_{ii}^j : r+1 \leq i \leq k\}$$

Solve the system by back substitution for $\underline{u}$

$$(L^j)'\underline{u} = e_q$$

$$\text{Set } \underline{z}_2 = \begin{cases} -\text{sign}(\underline{u}'\underline{g}^j)\underline{u} & \text{if } ||\underline{g}^j|| > 0 \\ \underline{u} & \text{if } ||\underline{g}^j|| = 0 \end{cases}$$

Go to step 7.

Step 7 : Compute the direction

$$\underline{d}_2 = V_2\underline{z}_2$$

and define the direction

$$\underline{d}^j = \underline{d}_1 + \underline{d}_2$$

Go to step 8.

Step 8 : If the negative of the normalised gradient

$$-\underline{g}^j\underline{d}^j/(||\underline{g}^j|| \; ||\underline{d}^j||) < \eta < ||D_1^{-2}||$$

return to Step 5(b) with grade r=0. Otherwise proceed to Step 9.

Step 9 : Compute a steplength γ_j using Fletcher's line search procedure (Appendix A, paragraph A4) and set

$$\underline{\theta}^{j+1} = \underline{\theta}^{j} + \gamma_j \underline{d}^{j}$$

Go to Step 10.

Step 10 : Continue until certain convergence criteria are met:

(a) $|g_i^j| < \text{gtol}$ for i=1,...k

(b) $|S_p(\underline{\theta}^{j+1}) - S_p(\underline{\theta}^{j})| / S_p(\underline{\theta}^{j}) < \text{Ftol}$
in 4 consecutive iterations.

Otherwise set j:=j+1 and return to Step 1.

Remark: As an alternative to test 10(a) the generalised gradient may be used.

The generalised gradient $\underline{g}^*$ is defined as

$$g_i^* = g_i^j / ((p-1)(J_p'J_p)_{ii} . S_p(\underline{\theta}^{j}))^{\frac{1}{2}}$$

and is due to Dennis (1977).

3. Numerical considerations and programme implementation

In the execution of the various steps in the numerical algorithm the following procedures will be used:

(a) Singular value decomposition: Subroutine SVDRS by courtesy of Lawson and Hanson (1974).

The singular value decomposition (SVD) of an nxk matrix A (rank (A)=k) is calculated in two stages:

(i) Construct a sequence of Householder transformations (paragraph B1).

P_i, $i=1,\ldots,k$, $\bar{P}_i$, $i=1,\ldots,k-1$ so that

$$P_k P_{k-1} \cdots P_1 A \bar{P}_1 \bar{P}_2 \cdots \bar{P}_{k-1} \equiv P' A \bar{P} = Q$$

where Q is an nxk bidiagonal matrix:

$$Q = \begin{bmatrix} a_1 & b_1 & & 0 \\ & a_2 & b_2 & \\ & & \ddots & \\ & & & b_{k-1} \\ 0 & & & a_k \\ \hline & & 0 & \end{bmatrix}$$

and where Q has the same singular values as A.

(ii) The singular values of Q are then computed with the aid of a special QR (Gram-Schmidt orthogonalization, paragraph B2) algorithm for computing singular values. For further information see Lawson and Hanson op.cit.

Subroutine SVDRS uses the following subroutines:

H12 : Constructs a Householder transformation and applies the transformation to a given vector.

QRBD : Computes the singular value decomposition of a bidiagonal matrix

G1, G2 : Construction and application of rotation matrices.

DIFF : Termination criterion.

A more detailed discussion of singular value decomposition is given in Appendix B paragraph B1.

(b) Modified Choleski factorization: Subroutines LDLT, LDLSOL and TEST were programmed to carry out this procedure.

(c) Numerical derivatives: Subroutines FGRAD and SBHESS calculate the first- and second-order partial derivatives numerically.

(d) Matrix multiplications: Subroutine MLPY by courtesy of Browne and du Toit (1977) is used in the multiplication of matrices in Steps 4, 5, 6 and 7 of the algorithm.

(e) Line search: Subroutine LINE was programmed to carry out Fletcher's line search algorithm.

A FORTRAN code of the full programme is given in Appendix E.

4. Numerical examples

The following numerical examples will be considered:

Example 1: Betts example 8.7 (see Chapter 2, example 3).

When tol:=0.04 the grade of J_p was found to equal 3. In the remaining iterations it remained at 3. This implies that full first-order gradient steps were taken throughout. The optimal solution was reached in 7 function evaluations as compared to the six previously. This may be due to the fact that numerical first-order derivatives were computed.

If we choose tol=0.1 then the grade of J_p is selected as 2 initially and 3 in the remaining steps. In this case the total number of function evaluations was equal to 10. This indicates the superiority of the first-order gradient algorithm over the mixture method for this small residual example. Similarly for tol=10, the algorithm took 13 function evaluations and in one step the grade of J_p was found to equal 1.

	p					
	1.5	1.75	2.0	2.50	2.75	3.0
θ_1	.09618	.08976	.08241	.07115	.06732	.06433
θ_2	1.41701	1.27552	1.1330	.93479	.87294	.826496
θ_3	2.07608	2.20989	2.3437	2.5282	2.5852	2.67812
$S_p(\underline{\theta})$	.031598	.01632	8.214877^{-3}	1.9470^{-3}	9.3118^{-4}	4.4275^{-4}
Evaluations	11	9	7	17	18	21

Example 2: This example is due to Jennrich and Sampson (1968) (see also Betts example 8.8).

$$\text{Min } S_2(\theta) = \sum_{i=1}^{10} \{y_i - (e^{i\theta_1} + e^{i\theta_2})\}^2$$

where $y_i = (2+2i)$ note that i is not a complex number.

The optimal solution is

$$\underline{\theta}^* = (0.25783,\ 0.25783)$$

$$S_2(\underline{\theta}^*) = 124.36$$

The usual starting value

$$\underline{\theta}^o = (0.3,\ 0.4)$$

was used.

Betts op.cit. reports 19 function and gradient evaluations whilst the best reported up to that time by Jennrich and Sampson was 100. Gill and Murray op.cit. reported 32, our algorithm took only 7 function and gradient evaluations.

p

	1.50	1.75	2.0	2.50	2.75	3.0
θ_1	0.25752	0.25784	0.25783	0.25754	0.257398	0.25729
θ_2	0.25752	0.25784	0.25783	0.25754	0.257398	0.25729
$S_p(\underline{\theta})$	62.6425	88.0693	124.362	250.537	357.026	509.883
Evaluations	6	6	7	8	8	9

We shall use this example to illustrate the intermediate steps of the algorithm. We have chosen p=2.

Step 0 : $\underline{\theta}^\circ = (0.5,1.5,-1.0,0.01,0.02)'$

$S_p(\underline{\theta}^\circ) = 0.879026$

Step 1 : The matrices $J_p(\underline{\theta}^\circ)$ and $(\underline{y}-\underline{f})_p$ will be 33x5 and 33x1 matrices respectively and will therefore not be reported. The gradient

$\underline{g}^\circ = (10.71000,3.06465,1.58106,-411.62400,76.25780)'$

$||\underline{g}^\circ|| = 418.7790000.$

Step 2: Singular values $\underline{s} = (237.000,28.530,2.514,0.7968,0.007084)'$.

Step 3: (a) $\rho_1 = 2886.0$

(b) $\rho_2 = 0.008891 < \text{tol} = 0.1$

grade $r = i_{min} = 4$

Step 4: $\underline{d}_1 = (-0.1187,-0.0158,-0.0184,0.0019,0.0047)'$

Step 5 : Matrix $B_p(\underline{\theta}^\circ)$ will be a 5x5 matrix and will not be reported here

$||E^\circ|| = 1.1986$

Step 6(b): $\underline{z}_2 = (-0.2643,-0.0058,0.1042,-0.0616,0.0000)'$

Step 7: $\underline{d}_2 = (-0.0047,-0.1848,0.1889,-0.0005,0.0012)'$

$\underline{d}^\circ = (-0.1234,-0.2006,0.1704,0.0013,0.0059)'$

Step 8: $-(\underline{g}^\circ)'\underline{d}^\circ/(||g^\circ||.||\underline{d}^\circ||) = .01444$

Step 9: $\gamma_o = 1.0$ and $\underline{\theta}^1 = (0.376621, 1.29940, -0.829575, 0.011313, 0.025889)'$

$S_p(\underline{\theta}^1) = 0.00403722$

The algorithm proceeds in this manner to the optimal solution.

The best reported algorithm before 1976 took 34 function and gradient evaluations. Betts reports 10 function and gradient evaluations. Gill and Murray report 13 function and gradient evaluations, ours took 34 function evaluations.

p

	1.50	1.75	2.0	2.50	2.75	3.0
θ_1	0.3736	0.3751	0.3754	0.3743	.3727	.3701
θ_2	1.7349	1.8995	1.9358	1.8224	1.6991	1.5416
θ_3	-1.2631	-1.4284	-1.4647	-1.3500	-1.2252	-1.0653
θ_4	0.0124	0.0128	0.01290	0.0126	0.0123	0.0118
θ_5	0.0231	0.0223	0.0221	0.0226	0.0233	0.0243
$S_p(\theta)$	0.1286^{-2}	0.2542^{-3}	0.5465^{-4}	0.2718^{-5}	0.6279^{-6}	0.1597^{-6}
Evaluations	75	52	34	35	33	32

APPENDIX B

B1. LINEAR ALGEBRA

A matrix J, nxk is RANK-DEFICIENT if its rank(J) < min(n,k). It is of FULL RANK if rank(J)=r=min(n,k). Matrix J will be termed of FULL COLUMN RANK if rank(J)=k. A square nxn matrix B is NON-SINGULAR if rank(B)=n and singular if rank(B) < n. Matrix B will be termed ORTHOGONAL if B'B=BB'=I. Matrix B is termed ILL-CONDITIONED if B is nearly singular. In this case its CONDITION NUMBER (defined as the ratio of the largest to the smallest eigenvalue) will be very large.

The Euclidean norm of a vector $\underline{x} \in R^k$, denoted by $||\underline{x}||$, is defined as

$$||\underline{x}|| = [\sum_{i=1}^{n} x_i^2]^{\frac{1}{2}} .$$

The Euclidean (Frobenius) norm of a square nxn matrix B, denoted by $||B||$ is defined as

$$||B|| = [\sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij}^2]^{\frac{1}{2}} .$$

Eigenvalue-eigenvector decomposition of a symmetric matrix

Let A_{nxn} be a given real symmetric matrix. The eigenvalue-eigenvector decomposition of A is of the form.

(B1.1) $A = VDV'$

where the columns of V are the eigenvectors of A and D is a diagonal matrix with diagonal elements the eigenvalues of A; ordered as follows:

$$\lambda_1 \geq \lambda_2 \geq \cdots \lambda_n .$$

By definition of the eigenvectors, V is an orthogonal matrix i.e. $VV'=V'V=I$. The relationship (B1.1) is derived as follows:

Per definition of the eigenvalue-eigenvector problem

$$Av_i = \lambda_i v_i \qquad i=1,..n$$

or $$AV = VD$$

hence $A=AVV' = VDV'$ since V is orthogonal.

The singular value decomposition (SVD) of a matrix

The singular value decomposition of matrix A_{nxk} hinges on the following theorem:

Theorem B1.1

Suppose the rank(A)=r < k, then there exists an orthogonal matrix U_{nxn} partitioned as

$$U = \{ U_1 : nxr \; ; \; U_2 : nx(n-r) \}$$

and an orthogonal matrix V_{kxk} partitioned as

$$V = \{V_1 : kxr \; ; \; V_2 : kx(k-r)\}$$

such that $A = UDV'$

where $$D_{nxk} = \begin{bmatrix} \text{diag}(s_1,\ldots,s_r) & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} D(s) & 0 \\ 0 & 0 \end{bmatrix}$$

with $s_1 \geq s_2 \geq \ldots s_r > 0$.

Proof: The matrix A'A will either be positive or positive semidefinite of rank r. It will therefore possess an eigenvalue-eigenvector decomposition of the form:

$$A'A = V \begin{bmatrix} D^2(s) & 0 \\ 0 & 0 \end{bmatrix} V' = VD^2V'. \quad \text{Thus } V'A'AV = D^2$$

where $s_1 \geq s_2 \geq \ldots s_r > 0$.

Denote the first r columns of V by V_1 and the remaining (k-r) columns by V_2. Then

$$I=VV'=V_1V_1' + V_2V_2' \quad \text{or}$$

$$V_1V_1' = I - V_2V_2' .$$

Since V'A'AV=D it follows that

$$V_1'A'AV_1 = D^2(s) \quad \text{and} \quad V_2'A'AV_2 = \underline{0}$$

Thus it follows that

$$AV_2 = \underline{0} .$$

We see that

$$A = A - (AV_2)V_2' \qquad \text{since } AV_2 = \underline{0}$$

$$= A\,(I - V_2V_2')$$

$$= AV_1V_1'$$

$$= AV_1(D(s)^{-1}D(s))\,V_1'$$

$$= (AV_1D(s)^{-1})D(s)V_1'$$

$$= U_1D(s)V_1'.$$

By the Gram-Schmidt orthogonalization process it is possible to construct a matrix U_2 such that $U = [U_1, U_2]$ is orthogonal.

(Note that U_1 is columnwise orthogonal i.e. $U_1'U_1=D(s)^{-1}V_1'A'AV_1D(s)^{-1}=I$).

Observe that

$$UDV' = [U_1 \;\; U_2]\begin{bmatrix} D(s) & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

$$= U_1D(s)V_1'$$

This completes the proof.

Remarks: 1. Matrices U and V consist of the orthonormalized eigenvectors of AA' and A'A respectively and the singular values s_i are the non-negative square roots of the eigenvalues of A'A.

2. The grade of matrix J_p, denoted by r, is the number of dominant singular values of J_p.

Gram-Schmidt orthogonalization

The following two theorems describe the orthogonalization process and may be found in standard texts on linear algebra. We shall state the proofs for completeness.

Theorem B1.2

Given a set of m linearly independent vectors $\underline{a}_1, \underline{a}_2, \ldots \underline{a}_m$. Then an orthonormal set of vectors $\underline{b}_1, \ldots \underline{b}_m$ can be constructed where each $\underline{b}_i$ is a linear combination of the vectors $\underline{a}_i$, i=1,...m.

Proof: The proof is by construction and is also known as the Gram-Schmidt orthogonalization method:

Choose $\underline{c}_1 = \underline{a}_1$ and set $\underline{b}_1 = \underline{c}_1/||\underline{c}_1||$. Set $\underline{c}_2 = \underline{a}_2 - \alpha_1\underline{b}_1$. Choose α_1 such that $\underline{b}_1$ and $\underline{c}_2$ are orthogonal:

$$\underline{b}_1'\underline{c}_2 = \underline{b}_1'\underline{a}_2 - \alpha_1\underline{b}_1'\underline{b}_1 = 0$$

$$\Rightarrow \quad \alpha_1 = \underline{b}_1'\underline{a}_2/\underline{b}_1'\underline{b}_1$$

$$= \underline{b}_1'\underline{a}_2 \qquad \text{since } \underline{b}_1'\underline{b}_1 = 1$$

$$\Rightarrow \quad \underline{c}_2 = \underline{a}_2 - (\underline{b}_1'\underline{a}_2)\underline{b}_1 \ .$$

Select $\underline{b}_2 = \underline{c}_2/||\underline{c}_2||$.

The only way in which this process would fail is when $\underline{c}_2 \equiv \underline{0}$. This would imply $\underline{a}_2$ and $\underline{a}_1$ are linearly dependent thus contradicting the assumption of the theorem. Next we form

$$\underline{c}_3 = \underline{a}_3 - \alpha_2\underline{b}_2 - \beta_1\underline{b}_1.$$

Choose α_2 and β_1, such that $\underline{c}_3$ is orthogonal to $\underline{b}_1$ and $\underline{b}_2$:

$\underline{c}_3'\underline{b}_1 = \underline{0} \Rightarrow \underline{a}_3'\underline{b}_1 - \beta_1 = 0$ since $\underline{b}_1$ and $\underline{b}_2$ are orthogonal.

$\underline{c}_3'\underline{b}_2 = \underline{0} \Rightarrow \underline{a}_3'\underline{b}_2 - \alpha_2 = 0$ for the same reason.

Hence we choose

$$\underline{c}_3 = \underline{a}_3 - (\underline{a}_3'\underline{b}_2)\underline{b}_2 - (\underline{a}_3'\underline{b}_1)\underline{b}_1$$

and $\underline{b}_3 = \underline{c}_3/||\underline{c}_3||$.

In general

$$\underline{c}_j = \underline{a}_j - \sum_{i=1}^{j-1} (\underline{a}_j'\underline{b}_i)\underline{b}_i$$

and $\underline{b}_j = \underline{c}_j/||\underline{c}_j||$.

Again the only way in which this process would fail is when $\underline{c}_j = 0$ for some j. This would imply that the vectors $\underline{a}_1,\ldots\underline{a}_j$ are linearly dependent since $\underline{c}_j$ is a linear combination of $\underline{a}_1$, ... $\underline{a}_j$ thus contradicting our assumption. This completes the proof.

Theorem B1.3

Let $\underline{a}_1$, $\underline{a}_2,\ldots\underline{a}_m$ be a set of nx1 orthonormal vectors ($m < n$) then there exist vectors $\underline{b}_1,\ldots\underline{b}_{n-m}$ such that the matrix

$$Q = [\underline{a}_1,\ldots\underline{a}_m, \underline{b}_1,\ldots\underline{b}_{n-m}]$$

is orthogonal.

Proof: Let $\underline{w}_1,\ldots\underline{w}_n$ be any set of linearly independent nxl vectors.

Consider the set of n+m vectors

$$\underline{a}_1,\ldots\underline{a}_m,\ \underline{w}_1,\ldots\underline{w}_n\ .$$

This set of vectors can be made linearly independent. If $\underline{a}_1,\ldots\underline{w}_n$ span a vector space we can always select from these a linearly independent set that spans the same space. The vectors $\underline{a}_1,\ldots\underline{a}_m$ are linearly independent. If $\underline{w}_1$ is linearly independent of the preceeding vectors $\underline{a}_1,\ldots\underline{a}_m$ we include it in the set as $\underline{c}_1$, if not we exclude it. Proceeding this way we obtain a set of n vectors

$$\underline{a}_1,\ldots\underline{a}_m,\ \underline{c}_1,\ldots\underline{c}_{n-m}$$

where any vector is linearly independent of the preceeding vectors in the set. By the Gram-Schmidt procedure we can replace $\underline{c}_1,\ldots\underline{c}_{n-m}$ by $\underline{b}_1,\ldots\underline{b}_{n-m}$ such that $\underline{a}_1,\ldots\underline{a}_m,\ \underline{b}_1,\ldots\underline{b}_{n-m}$ is an orthonormal set of vectors. Hence matrix Q is orthogonal. This completes the proof.

In matrix notation this process would imply

$$A = QR$$

where Q has columns $\underline{q}_j$ and R is an upper triangular matrix with

$$r_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i > j \\ \underline{a}_j'\underline{q}_i/(\underline{q}_i'\underline{q}_i) & \text{otherwise.} \end{cases}$$

In this instance it can be shown that

$$\underline{q}_j = \underline{a}_j - \sum_{i=1}^{j-1} \{\underline{a}_j'\underline{q}_i/\underline{q}_i'\underline{q}_i\}\,\underline{q}_i \quad .$$

Householder orthogonal transformations

A Householder transformation is an orthogonal matrix

$$P = I - 2\,\underline{w}\,\underline{w}'$$

where P and I are nxn and vector $\underline{w}$ is nx1 such that $||\,w\,|| = 1$.

The idea is to transform a real arbitrary vector $\underline{u}$ into a real vector $\underline{v}$ of the same length that is

$$\underline{v} = P\underline{u} \quad .$$

Define $\underline{w} = \underline{v} - \underline{u} \,/\, ||\underline{v} - \underline{u}||$

then $P\underline{u} = (I - 2\underline{w}\,\underline{w}')\underline{u}$

$$= \underline{u} - \frac{2(\underline{v} - \underline{u})(\underline{v} - \underline{u})'\underline{u}}{||\underline{v} - \underline{u}||^2} \quad .$$

Since $||\underline{v}|| = ||\underline{u}||$ we have

$$||\underline{v} - \underline{u}||^2 = \underline{v}'\underline{v} + \underline{u}'\underline{u} - 2\underline{u}'\underline{v}$$

$$= 2\underline{u}'\underline{u} - 2\underline{u}'\underline{v}$$

$$= -2(\underline{v}-\underline{u})'\underline{u} \quad .$$

Thus $P\underline{u} = \underline{u} + \underline{v} - \underline{u} = \underline{v}$.

Observe that P represents a rotation from $\underline{u}$ to $\underline{v}$ around an axis through the origin perpendicular to the plane containing $\underline{u}$ and $\underline{v}$.

Finally

$$P'P = PP' = (I - 2\underline{w}\,\underline{w}')'(I - 2\underline{w}\,\underline{w}')$$

$$= I - 4\,\underline{w}\,\underline{w}' + 4\,\underline{w}\,\underline{w}'\,\underline{w}\,\underline{w}'$$

$$= I \text{ by definition of } \underline{w}.$$

B2. CONVERGENCE RATES

Let $\underline{\theta}^0, \underline{\theta}^1, \ldots \in R^k$ be a sequence of vectors that converges to $\underline{\theta}^*$. If there exists numbers q and $a \neq 0$ such that

$$\lim_{n\to\infty}\{||\underline{\theta}^{n+1} - \underline{\theta}^*||/||\underline{\theta}^n - \underline{\theta}^*||^q\} = a$$

then q is the order of convergence of the sequence and a the root convergence factor. If q=1 we say the convergence rate is LINEAR. If q=1 and a=0 then we say the convergence rate is SUPERLINEAR. If q=2 then the convergence rate is QUADRATIC.

It can be shown that the steepest descent method converges at a linear rate and Newton's method is quadratic see e.g. Luenberger (1973). Gill and Murray op. cit. remark that in the case of small residual problems "Gauss-Newton methods will ultimately converge at the same rate as Newton's method despite the fact that only first derivatives are required". Obviously the higher the order of convergence the quicker convergence of $\underline{\theta}^k \to \underline{\theta}^*$ occurs. Powell (1971) shows that if the objective function to be minimized is convex and if exact line searches are used then a class of quasi-Newton or variable metric methods for unconstrained minimization is superlinearly convergent.

Fig. 3.1: Plot of the Jennrich and Sampson example

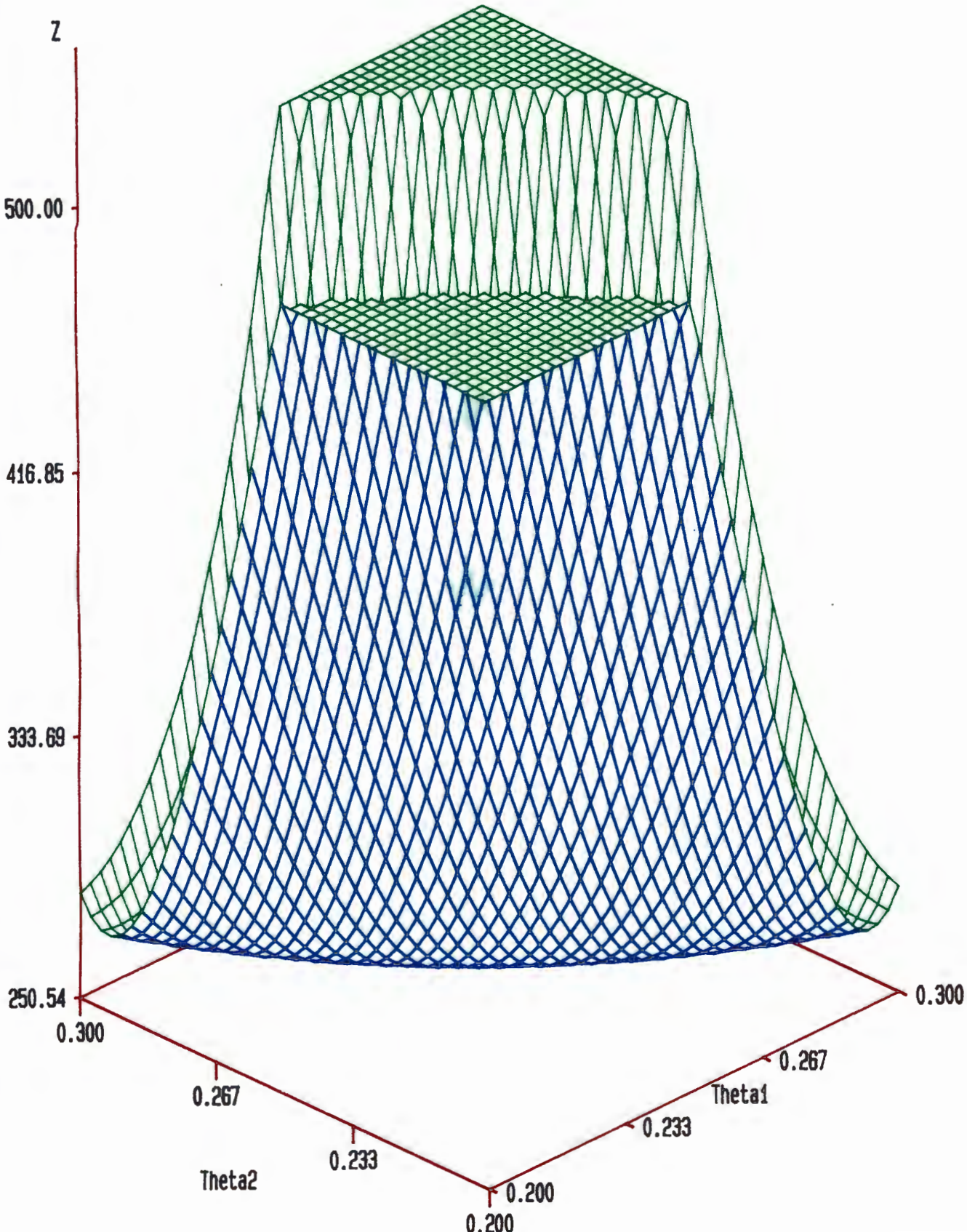
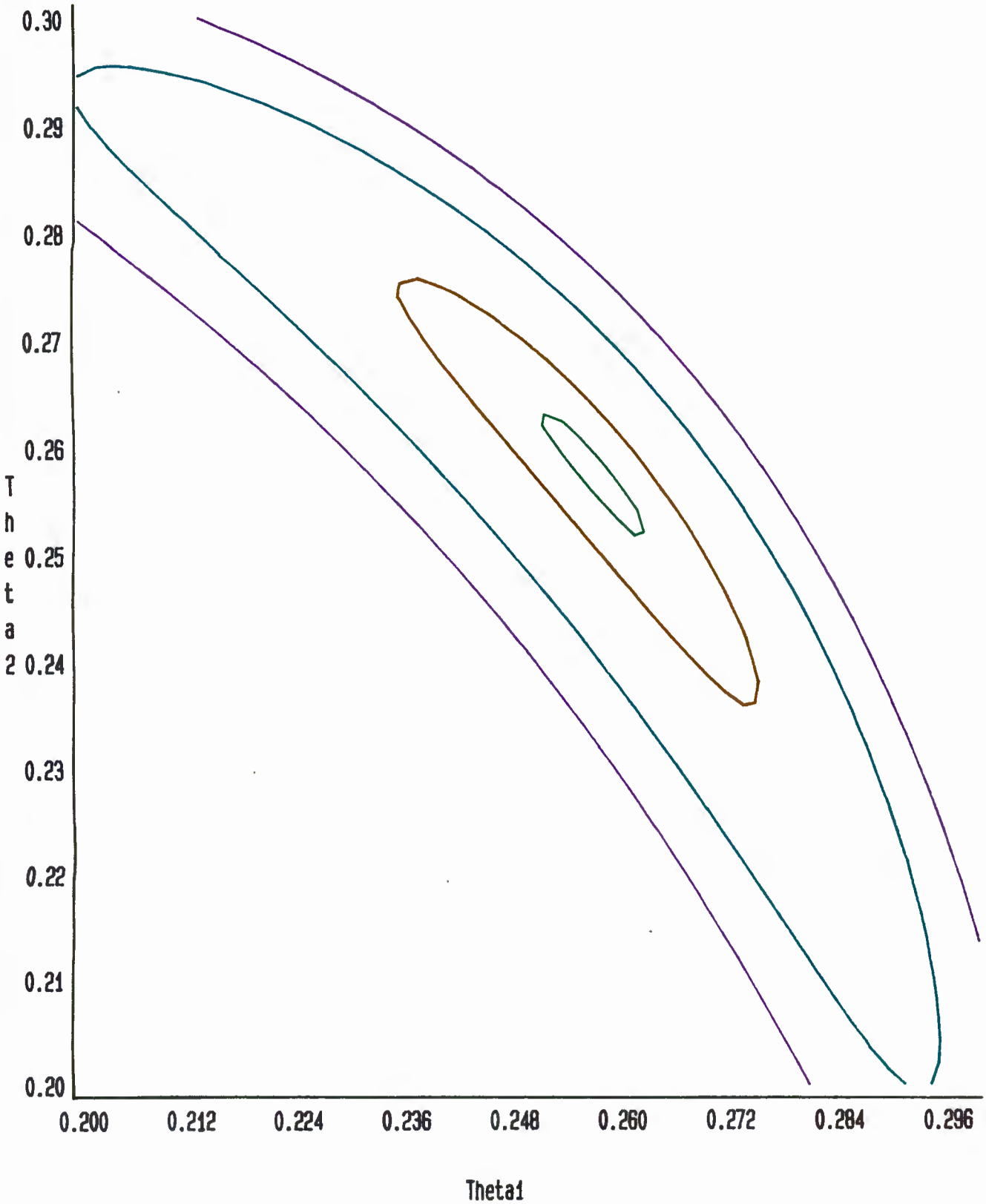


Fig. 3.2: Contours of the Jennrich and Sampson example



Chapter 4: A simulation study to establish the best value of p in L_p-norm estimation of a class of nonlinear models.

1. Introduction

Recall that in Chapter 2 we considered the nonlinear estimation problem: Given the data $(y_i, x_{1i},\ldots x_{mi})$ for i=1, ... n where $\underline{y}$ is the response or (dependent) and $\underline{x}_i = (\underline{x}_{1i},\ldots \underline{x}_{mi})$ the independent variables. Estimate the k parameters $\underline{\theta} = (\theta_1, \theta_2,\ldots, \theta_k)'$ from the nonlinear model:

$$y_i = f_i(x_{1i},\ldots x_{mi}; \underline{\theta}) + e_i \qquad i=1, \ldots n \tag{4.1}$$

where $n>k$ in general, $f(.)$ is the response function, $\underline{\theta}$ the unknown parameters and e_i unobserved error variates.

Then we can define the L_p- norm estimation problem as:

Find the parameters $\underline{\theta}$ which minimize

$$S_p(\underline{\theta}) = \sum_{i=1}^{n} |y_i - f_i(\underline{x}_i, \underline{\theta})|^p \tag{4.2}$$

where $1 < p < \infty$ in general.

In this Chapter we shall be concerned with the errors e_i in the model (4.1) as opposed to the actual estimation of the parameters $\underline{\theta}$, the subject of Chapters 2 and 3.

Laplace (1818) examined the asymptotic distribution of the least squares estimator as well the L_1-norm estimator for the linear regression model $y_i = \beta x_i + e_i$. He concluded that the L_1-norm estimation should be used instead of least squares when the residual distribution has a long tail (e.g. the Laplace and Cauchy distributions; in the former the kurtosis equals 6 whilst in the latter the kurtosis is not defined). It therefore seems logical to relate the p-value in the L_p-norm estimation to the kurtosis of the error distribution when it is symmetrically distributed.

The method of least squares (L_2-norm estimation) is the appropriate method for solving this problem when the error variates are normally distributed with expected value and variance:

$$E(e_i) = 0 \quad , \; var(e_i) = \sigma^2 > 0$$

respectively. The error variates are more often than not non-normally distributed hence an alternative to least squares such as L_p-norm estimation has to be considered.

Harter (1977) related the kurtosis of the error distribution to the value of p in linear L_p-norm estimation. He suggested that $p=\infty$ should be used when the error distribution kurtosis $\beta_2 > 3.8$ and $p=1$ when $\beta_2 < 2.2$. Money et al. (1982) used Monte Carlo simulation to derive an empirical relationship between the kurtosis of the error distribution and the value of p in linear L_p-norm estimation.

We shall, however, consider the nonlinear L_p-norm estimation problem. In this chapter a simulation study will be carried out to determine a relationship between the "optimal" value of p and the kurtosis of a given symmetric error distribution. The applicability of the empirical relationship of Money et al. to the nonlinear case with additive errors will be highlighted by this simulation study and will be supported even further by the recently established asymptotic properties of linear L_p-norm estimators which we shall apply to the nonlinear case.

The generation of random numbers from the uniform, parabolic, triangular, normal, contaminated normal (kurtoses 4 and 5) and Laplace distributions will be considered. We shall derive a formula for generating random numbers from a parabolic distribution which uses the analytical roots of a cubic equation. As far as the author is aware this is a new result.

2. The asymptotic variance of L_p-norm estimators for additive errors.

The mathematical regression model (4.1) is nonlinear in its parameters but the errors are additive (i.e. the errors are additive in the response variables).

Nyquist (1982) considers the L_p-norm estimation of the linear model:

$$\underline{y} = X\underline{\theta} + \underline{e} \tag{4.3}$$

where $\underline{y}$, X, $\underline{\theta}$ and $\underline{e}$ are nx1, nxm, mx1 and nx1 matrices respectively.

The following assumptions are then made:

A1 : The errors e_i are i.i.d. with common distribution F (e.g. F=uniform, Laplace etc.).

A2 : The L_1 and L_∞-norm estimators are unique. In general the L_p-norm estimators are uniquely defined for $1 < p < \infty$.

A3 : Matrix $Q = \lim_{n\to\infty} X'X/n$ is positive definite with rank(Q) = m.

A4a: F is continuous with a continuous positive derivative at the median i.e. $F'(e_i) > 0$ at $e_i=0$ when p=1.

A4b: When $1 < p < \infty$ the following expectations exist:

1) $E\{|e_i|^{p-1}\}$
2) $E\{|e_i|^{p-2}\}$
3) $E\{|e_i|^{2p-2}\}$ and
4) $E\{|e_i|^{p-1} .\text{sign}(e_i)\} = 0.$

If we substitute p = 2 in assumption 4, we find that

$$E\{|e_i| .\text{sign}(e_i)\} = E\{e_i\} = 0 .$$

This is simply the well known condition that the mean of the least squares errors is zero. Analogously, in terms of the residuals $\hat{e}_i$, this condition can be written as:

$$\sum_{i=1}^{n} |\hat{e}_i|^{p-1} \text{sign}(\hat{e}_i) = 0; \quad \text{for p=2 we have } \sum_{i=1}^{n} \hat{e}_i = 0.$$

Let $\hat{\underline{\theta}}$ be the estimate of $\underline{\theta}$ when L_p-norm approximation is used. Nyquist then shows that given these four assumptions, $\sqrt{n}\ (\hat{\underline{\theta}} - \underline{\theta})$ is asymptotically normally distributed with mean $\underline{0}$ and variance $w_p' Q^{-1}$

where

$$(4.4)\quad w_p^2 = \begin{cases} 1/4F'(0)^2 & \text{if } p = 1 \\ E(|e_i|^{2p-2})/\{(p-1)E(|e_i|^{p-2})\}^2 & \text{if } 1 < p < \infty. \end{cases}$$

We shall now postulate similar asymptotic results for the nonlinear model with additive error terms. Jennrich (1969) (see also Goldfeldt and Quandt (1972) and Gallant (1975)) shows that the least squares estimators $\sqrt{n}\ (\hat{\underline{\theta}}-\underline{\theta})$ are asymptotically normally distributed with mean zero and variance $w_2^2 Q^{-1}$ where

$$(4.5)\qquad Q = \lim_{n\to\infty} J'(\underline{\theta})J(\underline{\theta})/n$$

and

$$w_2^2 = S_2(\hat{\underline{\theta}})/(n-k) \text{ is an estimate of } \operatorname{var}(e_i) = \sigma^2$$

where $J(\underline{\theta})$ is the Jacobian matrix defined as the nxk matrix of first derivatives. It has element (i,j):

$$\frac{\partial f_i}{\partial \theta_j} \qquad i=1,\ldots n;\ j=1,\ldots k.$$

Gallant (1975) states two conditions on the limiting behaviour of the response function and its derivatives which correspond to the concept of estimability in the linear model:

1) $\lim \frac{1}{n} \sum_{i=1}^{n} [f_i(x_{1i},\ldots,x_{mi},\underline{\theta}) - f_i(x_{1i},\ldots x_{mi},\underline{\theta})]^2$ has a unique minimum at $\underline{\theta}$.

2) Q is nonsingular.

He goes on to say that "These two conditions are tedious to verify in applications and few would bother to do so." Clarke (1980) derived asymptotic results based on higher order derivatives (up to order 4). The derivation of these expressions is extremely complex.

In view of the results of Jennrich and Nyquist as well as the fact that the error terms are additive, we propose that the L_p-norm estimators for model (4.1) are asymptotically normally distributed with mean zero and variance $w_p^2 Q^{-1}$ where Q is given by expression (4.5) and w_p^2 by (4.4). Support for this proposal will be presented in the subsequent paragraphs and will be based on a Monte Carlo simulation study.

3. The choice of model and the error distribution

In physiological problems exponential growth models occur naturally. Such a model describes the relationship between the oxygen saturation of the haemoglobin of arterial blood and the partial pressure of oxygen in the pulmonary capillary. An exponential growth model which seeks to define this relationship is described in Du Toit and Gonin (1983). The model is

used by anaesthetists and respiratory physicians to determine the actual oxygen content of the blood. Moreover close familiarity with the model is essential for an understanding of many aspects of gas exchange in the body.

Exponential models are also used in the modelling of drug formulations. In this case the model describes the serum concentration of the drug over time and consists of a sum of decaying exponential terms. These examples will be considered in more detail in Chapter 5. It is important to note that both the models discussed have a definite physical interpretation.

In view of the importance of exponential models in medical research it was decided to use an exponential model in this simulation study. Other models may prove equally constructive. In this simulation study we used a two parameter nonlinear model of the form:

$$(4.6) \qquad y = 5 + 4\exp(\theta_1 x_1) + 3\exp(\theta_2 x_2) + e.$$

The coefficients 5,4 and 3 were arbitrarily chosen. Following Money et al. (1982) and Sposito et.al (1983) various SYMMETRIC error distributions were considered. Properties of the distributions may be found in Johnson and Kotz (1970). Some of these properties will only be stated briefly in the text, whilst the generation of random numbers from these distributions will be discussed in a fair amount of detail.

The Rand Corporation's 1 million digits are available on magnetic tape and since these numbers are now considered to be "truly" random and free of autocorrelation and periodicity it was decided to use them rather than a pseudo random number generator. In all the error distributions that were considered we assumed that

$$E(e_i) = 0 \quad \text{and} \quad Var(e_i) = \sigma^2 = 25.$$

The distributions vary from shortailed to mediumtailed (kurtosis = 3) to longtailed.

In previous studies of a similar nature, the random numbers were generated from various distributions without ensuring that the variances of these distributions were the same. This is important when the p value is calculated from the sample generalised variances.

A FORTRAN subroutine SIMUL which generates random numbers from the uniform, parabolic, triangular, normal and Laplace distributions is given in Appendix C.

The uniform distribution ($\beta_2 = 1.8$)

$$f_X(x) = \begin{cases} 1/(b - a) & , \quad a \leq x \leq b \\ 0 & , \quad \text{otherwise.} \end{cases}$$

It has moments:

$$\mu_r' = \frac{(b^{r+1} - a^{r+1})}{(b-a)(r+1)} = \mu_r \text{ (since } \mu = 0)$$

$$\mu_1 = \mu_1' = (b + a)/2 = 0 \Rightarrow b = -a$$

$$\mu_2 = \sigma^2 = b^2/3$$

$$\mu_4 = b^4/5 \Rightarrow \beta_2 = \mu_4/(\mu_2)^2 = 1.8.$$

To generate a random number from this distribution we simply use uniform [0,1] random numbers r, and set

$$x = a + (b-a)r \quad \text{then } x \sim \text{uniform } [a,b] .$$

For mean $\mu = 0$ and variance $\sigma^2 = 25$ we have $b = 8.6603$, $a = -b$.

$$x = 8.6603\ (2r-1). \tag{4.7}$$

The parabolic distribution ($\beta_2 = \frac{15}{7}$)

$$f_X(x) = \begin{cases} a(b-x^2) & , \ -\sqrt{b} \leq x \leq \sqrt{b} \\ 0 & , \text{ otherwise.} \end{cases}$$

It has mean $\mu = 0$ and distribution function

$$F(x) = -\frac{1}{3}ax^3 + abx + \frac{2}{3}ab^{1.5} \quad .$$

The moments are:

$$\mu_r = \mu'_r = \begin{cases} \dfrac{4ab^{(r+3)/2}}{(r+1)(r+3)} & , \ r = \text{even} \\ 0 & , \ r = \text{odd} \end{cases}$$

$$\therefore \quad \mu_2 = \frac{4ab^{2.5}}{15} = \sigma^2$$

$$\mu_4 = \frac{4ab^{3.5}}{35}$$

Since $\mu_0 = \dfrac{4ab^{1.5}}{3} = 1$ it follows that

$$\mu_2 = \frac{b}{5} = \sigma^2 \qquad \text{or } b = 5\sigma^2$$

$$\mu_4 = \frac{3b^2}{35} \Rightarrow \beta_2 = 15/7 \quad .$$

We shall use the inverse transform method to calculate a random number from this distribution.

Set

$$r = -\frac{1}{3}ax^3 + abx + \frac{2}{3}ab^{1.5} \quad .$$

The analytical roots may be calculated (Spiegel (1968) formula 9.4):

$$\theta = \arccos\left(+Q/\sqrt{-R^3}\right)$$

$$= \arccos(1-2r) \qquad \text{since } Q = b^{1.5}(1-2r) \text{ and } R = -b \quad . \tag{4.8}$$

The three <u>real</u> roots of this cubic are:

$$x_1 = 2\sqrt{b}\cos(\theta/3)$$

$$x_2 = 2\sqrt{b}\cos(\theta/3 + 120^o)$$

$$x_3 = 2\sqrt{b}\cos(\theta/3 + 240^o)$$

By inspection we see that $-\sqrt{b} \leq x_3 \leq \sqrt{b}$ for $0 \leq r \leq 1$:

r	θ	x_1	x_2	x_3
0.0	0^o	$2\sqrt{b}$	$-\sqrt{b}$	$-\sqrt{b}$
0.5	90^o	$\sqrt{3b}$	$-\sqrt{3b}$	0
1.0	180^o	$\sqrt{b}$	$-2\sqrt{b}$	$\sqrt{b}$

(b) Using the convolution theorem it can be shown that the sum of two uniform random numbers:

$$x_1 \sim [-2b, -b]$$
$$x_2 \sim [b, 2b]$$

i.e. $x = x_1 + x_2$ will have the required triangular distribution.

With mean $\mu = 0$ and $\sigma^2 = 25$

$b = 12.2474$ and in

(a)

$$(4.10) \qquad x = \begin{cases} 12.2472\ (\sqrt{2r} - 1) & \text{if } 0 \leq r \leq .5 \\ 12.2474\ (1 - \sqrt{2(1-r)}\ ) & \text{if } .5 \leq r \leq 1.0 \end{cases} .$$

(b) $x_1 = 12.2474\ (r-2)$, $x_2 = 12.2474\ (r+1)$

or

$$(4.11) \qquad x = 12.2474\ (2r-1) .$$

Computationally, however, (4.11) is preferred to (4.10), since it does not involve the taking of square roots nor is selection based on the value of r as in (4.10).

The normal distribution ($\beta_2 = 3$)

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad -\infty < x < \infty \quad .$$

To generate random normal deviates, we use the Odeh and Evans (1974) approximation to the inverse error function $F^{-1}(r)$ which is accurate to 7 decimals (see also Kennedy and Gentle (1980) p 95). It is a rational fraction approximation of the inverse error function (see subroutine SIMUL in the Appendix).

For mean $\mu = 0$ and variance $\sigma^2 = 25$ we set

(4.12) $\quad x = 5F^{-1}(r).$

The contaminated normal distribution ($\beta_2 = 4$ and 5)

$$f_X(x) = \frac{w_1}{\sqrt{2\pi\sigma_1^2}} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\} + \frac{w_2}{\sqrt{2\pi\sigma_2^2}} \exp\left\{-\frac{(x-\mu_2)^2}{2\sigma_2^2}\right\}$$

and $w_1 + w_2 = 1$ $\qquad -\infty < x < \infty \quad .$

We shall choose $w_1 = w_2 = \frac{1}{2}$, $\mu_1 = \mu_2 = 0$.

The central moments are:

$$\tilde{\mu}_1 = \tilde{\mu}_3 = 0$$

$$\tilde{\mu}_2 = \tilde{\mu}_2' = (\sigma_1^2 + \sigma_2^2)/2$$

$$\tilde{\mu}_4 = \frac{3}{2}(\sigma_1^4 + \sigma_2^4)$$

$$\therefore \quad \beta_2 = \frac{6(\sigma_1^4 + \sigma_2^4)}{(\sigma_1^2 + \sigma_2^2)^2} \quad .$$

In order to generate a contaminated normal with variance σ^2 and kurtosis β_2 we need to solve the two equations for σ_1 and σ_2 :

$$\sigma_1^2 + \sigma_2^2 = 2\sigma^2$$

$$\sigma_1^4 + \sigma_2^4 = \frac{2}{3}\sigma^4 \beta_2 \ .$$

This can be solved algebraically to yield:

$$\sigma_1^2 = \sigma^2 \ (1 + (\beta_2/3-1)^{\frac{1}{2}})$$

$$\sigma_2^2 = \sigma^2 \ (1 - (\beta_2/3-1)^{\frac{1}{2}}) \quad .$$

With mean $\mu = 0$ and $\sigma^2 = 25$ set:

$$x_1 = 6.2796F^{-1}(r)$$

(4.13) for $\beta_2 = 4$

$$x_2 = 3.2506F^{-1}(s)$$

$$x_1 = 6.7389F^{-1}(r)$$

(4.14) for $\beta_2 = 5$

$$x_2 = 2.1419F^{-1}(s)$$

where r,s ~ uniform [0,1].

and $F^{-1}(.)$ is calculated as in the normal distribution. We then lump the random numbers from the two differing normal distributions together. The combination will then be random numbers from a contaminated normal distribution with mean $\mu = 0$, $\sigma^2 = 25$.

The Laplace distribution ($\beta_2 = 6$)

$$f_X(x) = \frac{1}{2b} \exp\left[-|x - a|/b\right] \qquad -\infty < x < \infty$$

and distribution function:

$$F(x) = \begin{cases} .5 \exp((x-a)/b) & x \leq a \\ 1 - .5 \exp(-(x-a)/b) & x > a \end{cases}$$

We must have

$$\sigma^2 = 2b^2 = 25 \Rightarrow b = 3.5355$$

$$\mu = a = 0 \quad .$$

By the inverse transform method we choose

$$(4.15) \qquad x = \begin{cases} 3.5355 \ln 2r & 0 < r \leq .5 \\ -3.5355 \ln 2(1-r) & .5 < r < 1 \quad . \end{cases}$$

4. Numerical considerations and the design of the simulation study

In the simulation study two algorithms were used. The first algorithm was outlined in Chapter 2. This algorithm appears to be numerically efficient for values of $p > 1$. To solve the nonlinear L_1-problem an algorithm due to Osborne and Watson (1971) was used, it is as follows:

Given an initial estimate say $\underline{\theta}^0$ of the optimal $\underline{\theta}^*$, set j=0.

Step 1 : Set j:=j+1. Calculate $\delta\underline{\theta}^j$ so that

$$\hat{S}_j = \min_{\delta\underline{\theta}^j} \sum_{i=1}^{n} |y_i - f_i(\underline{x};\ \underline{\theta}^j) - \nabla_{\underline{\theta}} f_i(\underline{x},\ \underline{\theta}^j)'\ \delta\underline{\theta}^j| \quad .$$

Step 2 : Calculate γ^j so that $\bar{S}^{j+1} = \min_{\gamma>0} \sum_{i=1}^{n} |y_i - f_i(\underline{x},\ \underline{\theta}^j + \gamma\delta\underline{\theta}^j)|$.

Step 3 : Set $\underline{\theta}^{j+1} = \underline{\theta}^j + \gamma^j\ \delta\ \underline{\theta}^j$ and go to Step 1. Repeat until certain convergence criteria are met.

In Step 1 use is made of the first-order Taylor series expansion of $f_i(\underline{x},\underline{\theta})$ about $\underline{\theta}$. The original nonlinear problem is therefore reduced to a linear L_1-norm problem with unknown variables $\underline{\theta}$. This linear L_1-norm problem can be solved efficiently by a linear programming algorithm such as the Barrodale and Roberts (1973) procedure.

In the simulation study it was decided to arbitrarily fix the values of θ_1 and θ_2 in model (4.4) at 1.0 and 1.5 respectively. Thirty values for $(x_{1i},\ x_{2i})$ i=1,... 30 were selected from a $[0.5,1.5]$ uniform distribution and held fixed in all 500 samples for all error distributions and L_p-norm approximations. The values for (x_{1i},x_{2i}) are given in Table 4.1.

In the subsequent simulation runs it was found that the number of estimates lying above or below the true values was about the same. For example when p=2.75 was used, 263 (247) of the 500 estimates lay above (below) $\theta_1 = 1$ and 248 (252) lay above (below) $\theta_2 = 1.5$ in the simulation run for the uniform error distribution. Note that the number of values lying above (or below) the true values fell in the 95% confidence limits of the binomial distribution.

Since the normal distribution is a good approximation of the binomial distribution for large sample sizes (n = 500), we derive the 95% confidence limits of the number of estimates m as:

$$|m| < 1.96(nP(1-P))^{\frac{1}{2}} + 250$$

or $$228 < m < 272$$

where m is the number of estimates, probability P = 0.5, n = 500 and the normal z-value, $z_{.05} = 1.96$.

We found the actual numbers lying above or below the true values to be closer to 250 i.e. a range of 240 to 260 as compared to the above range of 228 to 272. Tables 4.2 and 4.3 illustrate for each error distribution the sample means and sample variances (based on 500 simulated experiments) of the parameters θ_1 and θ_2 for varying values of p.

5. The generalised variance of $\hat{\theta}_1$ and $\hat{\theta}_1$ and the choice of p

As Money et al. observe: "the empirical generalised variance... (of $\hat{\theta}_1$ and $\hat{\theta}_2$) ... defined as the determinant of the empirical covariance matrix... ($\text{cov}(\hat{\theta}_1, \hat{\theta}_2)$) ... can be considered as an univariate summary of the information present in the sample covariance matrix." The choice of the estimate of the optimal value of p is therefore based on the minimum sample generalised variance. We shall not be in a position to find this smallest value for that would entail an exhaustive search over all

values of p. Instead, we shall perform the simulations for selected values of p (i.e. 1.0, 1.25. 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, and 3.0). We shall shortly show how this minimum value can be found with the aid of our asymptotic results (Paragraph 2).

The generalised variances of the 500 sample estimates are given in Table 4.4 for various values of p as well as the 7 different symmetric error distributions. In Figure 4.1 a graphical picture of the optimum p value versus the kurtosis of the error distribution is given.

We shall now show that these numerical results are in complete agreement with the theoretical results outlined in paragraph 2. Recall that the asymptotic variance of $\sqrt{n}\,(\underline{\hat{\theta}} - \underline{\theta})$ was postulated to be $w_p^2 Q^{-1}$ where w_p^2 is given by (4.4) and Q by (4.5). Since Q^{-1} is independent of p, the expression will yield the optimal value of p for a given error distribution. Consider therefore the normal, uniform and Laplace distributions with kurtoses of 3, 1.8 and 6 respectively. In order to compare the analytical results with our simulation results, we shall assume that the error distributions all have the same variance, $\text{var}(e_i) = 25 = \sigma^2$.

We shall now calculate the theoretical expressions for the uniform, normal and Laplace distributions respectively:

a) Uniform $U \sim [-a,a]$: $E(|u|)^{2p-2}) = 2\int_0^a \frac{|u|^{2p-2}}{2a}\,du = a^{2p-2}/(2p-1)$

$$E(|u|)^{p-2}) = 2\int_0^a \frac{|u|^{p-2}}{2a}\,du = a^{p-2}/(p-1)$$

$$\therefore \quad w_p^2 = 3\sigma^2/(2p-1) \qquad 1 < p < \infty \ .$$

b) Normal $N(0,\sigma^2)$: $E(|u|)^{2p-2}) = \int_{-\infty}^{\infty} \frac{|u|^{2p-2} \exp(-u^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} du$

Set $t = u^2/2\sigma^2$ and integrate over it:

$$\therefore \quad E(|u|^{2p-2}) = (2\sigma^2)^{p-1} \Gamma(p-\tfrac{1}{2})/\sqrt{\pi} \quad .$$

Similarly

$$E(|u|)^{p-2}) = \int_{-\infty}^{\infty} \frac{|u|^{p-2} \exp(-u^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} du$$

$$= (2\sigma^2)^{\frac{p-2}{2}} \Gamma(\tfrac{p-1}{2})/\sqrt{\pi}$$

$$\therefore \quad w_p^2 = \frac{E(|u|^{2p-2})}{\{(p-1)E(|u|^{p-2})\}^2}$$

$$= \frac{2\sqrt{\pi}\sigma^2 \Gamma(p-\frac{1}{2})}{\{(p-1)\Gamma(\frac{p-1}{2})\}^2} \qquad 1 < p < \infty \quad .$$

Laplace: $E(|u|^{2p-2}) = \frac{1}{2b} \int_{-\infty}^{\infty} |u|^{2p-2} \exp(-|u|/b) du$

$$= b^{2p-2} \Gamma(2p-1) \quad .$$

Similarly

$$E(|u|^{p-2}) = b^{p-2}\Gamma(p-1)$$

$$\therefore \quad w_p^2 = \frac{b^2\Gamma(2p-1)}{\{(p-1)\Gamma(p-1)\}^2} \qquad 1 < p < \infty$$

Figure 4.2 displays the w_p^2 curve for each of the distributions. All that we need to do now is to read off the minimum value of the w_p^2 with respect to p for each distribution. A numerical procedure may also be used to calculate this minimum value. For the normal distribution we find that p=2, a well known result. For the Laplace distibution we find that p=1 and for the uniform $p=\infty$. This is in agreement with our simulation results. It is also interesting to observe the agreement with the results found by Money et al. in the linear case.

In Figure 4.2 we observe that the w_p^2 curves all intersect at p=2 when the variances are equal. This must be so since:

$w_2^2 = 2\sqrt{\pi}\ \sigma^2\Gamma(3/2)/\pi = \sigma^2$ for the normal distribution

$w_2^2 = \sigma^2$ for the uniform distribution

$w_2^2 = \sigma^2\Gamma(3)/2 = \sigma^2$ for the Laplace distribution

$w_2^2 = \sigma^2$ is a well known result in least squares theory. In Figure 4.3 we have plotted w_p^2 with differing error distribution variances. See also Nyquist op cit.

Our simulation results indicate a certain relationship between the optimal value of p to be used and the kurtosis of the underlying error distribution (assuming the errors are additive in the model). Money et al. suggested the following empirical relationship in linear L_p-norm estimation.

$$(4.16) \qquad p = 9/(\text{kurtosis})^2 + 1 \,.$$

The theoretical analysis supports this relationship. The following alternative formula

$$(4.17) \qquad p = 6/\text{kurtosis} \,.$$

has been suggested by Sposito et.al. (1983). We shall use both formulae in Chapter 5 when we discuss practical applications.

6. The relative efficiency of the L_p - norm estimates for varying values of p.

In this paragraph we shall consider the ratio

$$R = \frac{\text{generalised variance } (|\text{cov}(\hat{\theta}_1, \hat{\theta}_2)|) \text{ using the optimal p}}{\text{generalised variance } (|\text{cov}(\hat{\theta}_1, \hat{\theta}_2)|) \text{ with another p value}}$$

where the generalised variances are given in Table 4.4. These ratios are displayed in Table 4.5. We observe, using the generalised variance as a basis, that these ratios indicate the efficiency of the L_p-norm estimation.

This ratio also indicates the behaviour of the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ when p values other than the optimal one are used in the estimation procedure. For example, if the errors are Laplace distributed and least squares estimation is used, then we would expect the estimates to be about 56% as efficient as those obtained using the optimal p = 1.25. Similarly if the errors are uniformly distributed and least squares estimation is used, then we would expect the estimates to be 20% as efficient as those obtained using the optimal p value.

To summarise then: The empirical results indicate a relationship between the kurtosis of the error distribution and the optimal p value based on the sample generalised variance. This relationship and the efficiency of the estimates highlight the inherent danger of using, for example, least squares estimation when the data is symmetrically though non-normally distributed. This observation will be taken up further in Chapter 5.

APPENDIX C

```
      SUBROUTINE SIMUL(MU,A1,A2,B1,B2,R,S,Y1,Y2,IOPT)
C
C     IOPT=1 SELECTS THE LAPLACE DISTRIBUTION
C     IOPT=2 SELECTS THE NORMAL  DISTRIBUTION (ODEH AND EVANS (1974)
C            W.J.KENNEDY,J.R. AND J.E.GENTLE,STATISTICAL COMPUTING
C            (MARCEL DEKKER, NEW YORK, 1980),P 95)
C     IOPT=3 SELECTS THE PARABOLIC DISTRIBUTION
C     IOPT=4 SELECTS THE TRIANGULAR DISTRIBUTION (INVERSE TRANSFORM)
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DOUBLE PRECISION MU,LIM
C
C     IN THIS PROGRAM RANDOM DEVIATES ARE GENERATED ACCORDING
C     TO THE INVERSE METHOD
C     MU = SHIFT PARAMETER
C     B = SHAPE PARAMETER
C
      GO TO (10,20,30,40),IOPT
C
C     IN THIS SECTION RANDOM LAPLACE DEVIATES ARE GENERATED
C
10    IF(R.GT.0.5) GO TO 1
      Y1=A1+B1*DLOG(2.D0*R)
      RETURN
 1    Y1=A1-B1*DLOG(2.D0-2.D0*R)
      RETURN
C
C     IN THIS SECTION RANDOM NORMAL DEVIATES ARE GENERATED ACCORDING
C     TO THE RATIONAL APPROXIMATION FORMULA FOR THE INVERSE NORMAL DIS=
C     TRIBUTION  (ODEH AND EVANS (1974))
C     A = MEAN OF THE NORMAL DISTRIBUTION
C     B = STANDARD DEVIATION OF THE NORMAL DISTRIBUTION
C
20    LIM=1.D-20
      P0=-0.322232431088D0
      P1=-1D0
      P2=-0.342242088547D0
      P3=-0.0204231210245D0
      P4=-0.453642210148*1.D-4
      Q0=0.099348462606D0
      Q1=0.588581570495D0
      Q2=0.531103462366D0
      Q3=0.103537752850D0
      Q4=0.38560700634*1D-2
      IERROR=1
      Y1=0D0
      IF (R.GT.0.5D0) GO TO 15
      IC=1
      GO TO 16
15    R=1D0-R
      IC=0
16    IF (R.LT.LIM) GO TO 31
      IF (R.EQ.0.5D0) GO TO 11
```

```
      Y=DSQRT(DLOG(1D0/R**2))
      Y1=Y+((((Y*P4+P3)*Y+P2)*Y+P1)*Y+P0)/((((Y*Q4+Q3)*Y+Q2)*Y+Q1)*
     *Y+Q0)
14    IF(IC.EQ.1)Y1=-Y1
      GO TO 31
11    IERROR=0
31    CONTINUE
      Y1=A1+B1*Y1
      RETURN
C
C     THIS SECTION COMPUTES A RANDOM PARABOLIC DEVIATE OVER THE INTERVAL
C     (-SQR(B1),SQRT(B1)) WHERE
C     R IS A UNIFORM (0,1) RANDOM NUMBER
C
30    FOURPI=4D0*3.141592654D0
      THETA3=DARCOS(1-2D0*R)/3D0
      Y1 = 2D0*DSQRT(B1)*DCOS(THETA3 + FOURPI/3D0)
      RETURN
C
C     THIS SECTION COMPUTES A RANDOM TRIANGULAR DEVIATE OVER THE INTERVAL
C     (A1 + A2, B1 + B2) WHERE
C     R IS A UNIFORM (0,1) RANDOM NUMBER
C
40    X1=A1 +(B1-A1)*R
      X2=A2 +(B2-A2)*S
      Y1=X1+X2
      RETURN
      END
```

TABLE 4.1

TABLE OF UNIFORM [0.5,1.5] RANDOM NUMBERS (X_{1i}, X_{2i}) i=1,30

X_{1i}	X_{2i}
0.5973	1.0202
1.2652	0.6491
1.3635	0.5722
0.9877	0.5559
1.4091	0.5934
0.7927	1.0007
0.9227	0.9530
0.6965	0.5629
0.5323	1.3314
0.7560	0.5078
0.8348	1.0406
0.5803	1.2118
1.3080	1.0947
1.3016	1.3752
0.9764	1.1992
1.1654	1.3929
0.6169	1.0614
0.6718	0.9783
0.5601	0.6244
0.9557	0.6773
0.5635	0.6545
0.7615	0.7374
1.4074	1.0783
0.9031	0.8855
1.0733	1.4163
0.5533	1.1625
0.9891	0.9837
1.2548	1.2204
1.3287	0.6140
0.9913	0.9006

TABLE 4.2

COMPARISON OF MEAN VALUES OF THE ESTIMATED REGRESSION COEFFICIENTS $\hat{\theta}_1$ AND $\hat{\theta}_2$ WITH POPULATION VALUES

$\theta_1 = 1.0$, $\theta_2 = 1.5$ (n=30, $\sigma^2 = 25$)

Distrbn. kurt.	p = 1.0*	1.25	1.50	1.75	2.0	2.25	2.50	2.75	3.0
Uniform 1.8	.998 1.489	.992 1.494	.995 1.495	.996 1.496	.997 1.497	.997 1.498	.997 1.498	.997 1.498	.997 1.499
Parabolic 2.14	.991 1.492	.994 1.495	.995 1.496	.996 1.497	.997 1.498	.997 1.498	.997 1.498	.997 1.498	.996 1.499
Triangular 2.4	.993 1.493	.995 1.496	.996 1.497	.996 1.497	.997 1.498	.996 1.498	.996 1.498	.996 1.498	.996 1.499
Normal 3.0	.993 1.494	.995 1.496	.996 1.497	.997 1.498	.997 1.498	.997 1.498	.996 1.498	.996 1.498	.995 1.498
C.Normal 4.0	.997 1.494	.997 1.496	.997 1.497	.997 1.497	.997 1.497	.997 1.496	.996 1.496	.996 1.496	.995 1.495
C.Normal 5.0	.999 1.496	.998 1.497	.998 1.497	.998 1.497	.998 1.496	.997 1.496	.996 1.495	.995 1.494	.993 1.493
Laplace 6.0	.997 1.497	.998 1.499	.998 1.499	.998 1.499	.997 1.498	.996 1.498	.994 1.497	.993 1.497	.990 1.496

*

TABLE 4.3

EMPIRICAL VARIANCES OF $\hat{\theta}_1$ AND $\hat{\theta}_2$ (n=30, σ^2=25)

Distrbn. kurt.	p = 1.0*	1.25	1.50	1.75	2.0	2.25	2.50	2.75	3.0
Uniform	.030	.020	.015	.012	.011	.009	.009	.008	.008
1.8	.016	.011	.008	.007	.006	.005	.005	.005	.004
Parabolic	.024	.017	.013	.011	.011	.010	.009	.009	.009
2.14	.013	.009	.007	.007	.006	.006	.005	.005	.005
Triangular	.020	.015	.012	.011	.010	.010	.010	.010	.010
2.4	.011	.008	.007	.006	.006	.005	.006	.006	.006
Normal	.018	.014	.011	.011	.010	.011	.011	.012	.012
3.0	.010	.007	.006	.006	.006	.006	.006	.007	.007
C.Normal	.016	.013	.012	.012	.013	.014	.016	.017	.019
4.0	.008	.007	.006	.006	.007	.007	.008	.009	.010
C.Normal	.012	.011	.011	.013	.014	.017	.019	.021	.024
5.0	.005	.005	.005	.006	.007	.008	.010	.011	.012
Laplace	.010	.008	.008	.009	.010	.012	.014	.017	.020
6.0	.005	.004	.004	.005	.006	.008	.009	.010	.013

*(p=1 : Osborne-Watson algorithm)

TABLE 4.4

GENERALISED VARIANCE OF $\hat{\theta}_1$ AND $\hat{\theta}_2$: $|\mathrm{cov}(\hat{\theta}_1, \hat{\theta}_2)|$ $(n=30, \sigma^2=25)$

p

Distrbn. kurt.	1.0*	1.25	1.50	1.75	2.0	2.25	2.50	2.75	3.0
Uniform 1.8	2.275	1.058	.5911	.3956	.2909	.2277	.1869	.1590	.1390
Parabolic 2.14	1.484	.7537	.4678	.3488	.2866	.2503	.2279	.2140	.2055
Triangular 2.4	1.007	.5694	.3861	.3165	.2843	.2692	.2633	.2635	.2677
Normal 3.0	.8103	.4773	.3355	.2907	.2807	.2891	.3104	.3423	.3836
C.Normal 4.0	.5143	.3545	.3104	.3228	.3686	.4424	.5443	.6740	.8312
C.Normal 5.0	.2199	.1947	.2253	.2994	.4195	.5890	.8120	1.090	1.422
Laplace 6.0	.2086	.1547	.1556	.1966	.2755	.4047	.5996	.8753	1.248

*(actual values are x 10^{-4})

TABLE 4.5

EFFICIENCY BASED ON THE EMPIRICAL GENERALISED VARIANCE OF $\hat{\theta}_1$ AND $\hat{\theta}_2$
($n=30$, $\sigma^2=25$)

Distrbn. kurt.	p = 1.0	1.25	1.50	1.75	2.0	2.25	2.50	2.75	3.0
Uniform* 1.8	.03	.05	.10	.15	.20	.26	.31	.37	.42
Parabolic† 2.14	.13	.26	.42	.57	.69	.79	.87	.92	.96
Triangular 2.4	.33	.46	.68	.83	.93	.98	1.00	.999	.98
Normal 3.0	.35	.59	.84	.97	1.00	.97	.90	.82	.73
C.Normal 4.0	.60	.88	1.00	.96	.84	.70	.57	.46	.37
C.Normal 5.0	.88	1.00	.86	.65	.46	.33	.24	.18	.14
Laplace 6.0	.74	1.00	.99	.79	.56	.38	.26	.17	.12

* The optimum p value is 11.5 where the generalised variance is equal to 0.05817.

† The optimum p value is 3.75 where the generalised variance is equal to 0.1977.

Fig. 4.1 : p value vs Kurtosis

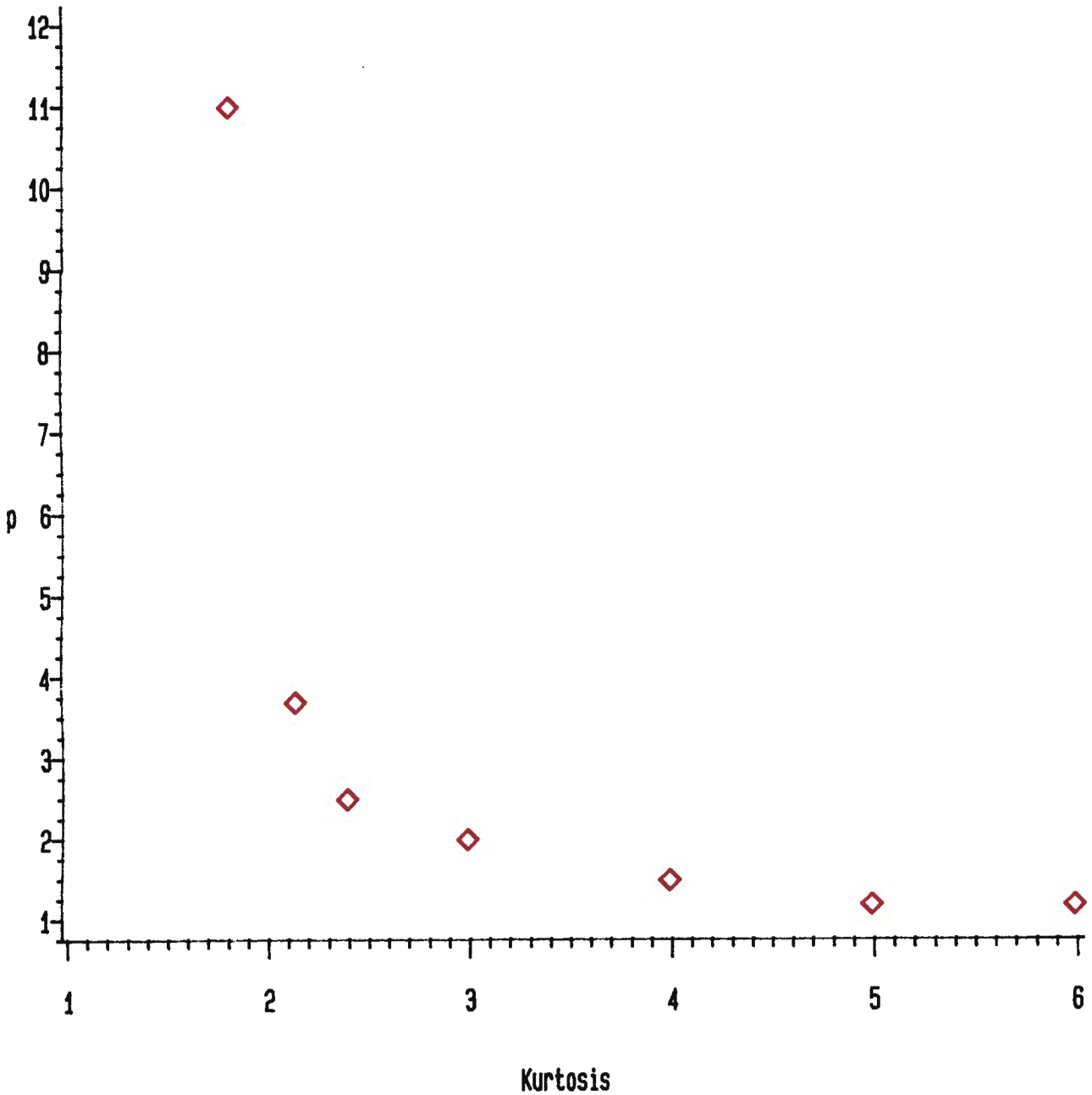


*Fig. 4.2: (Wp)**2 for 3 power distributions vs p*

error variances = 25

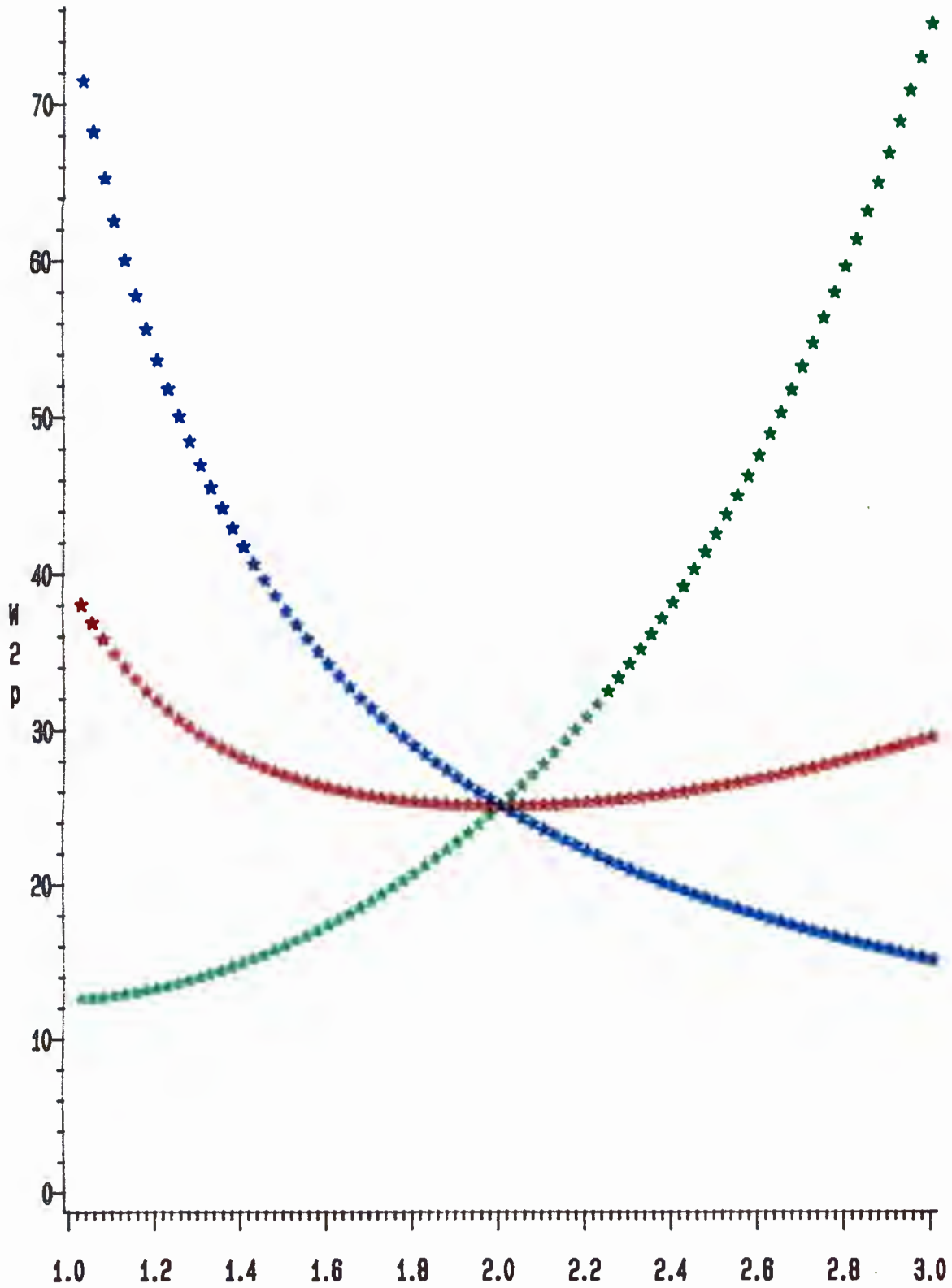
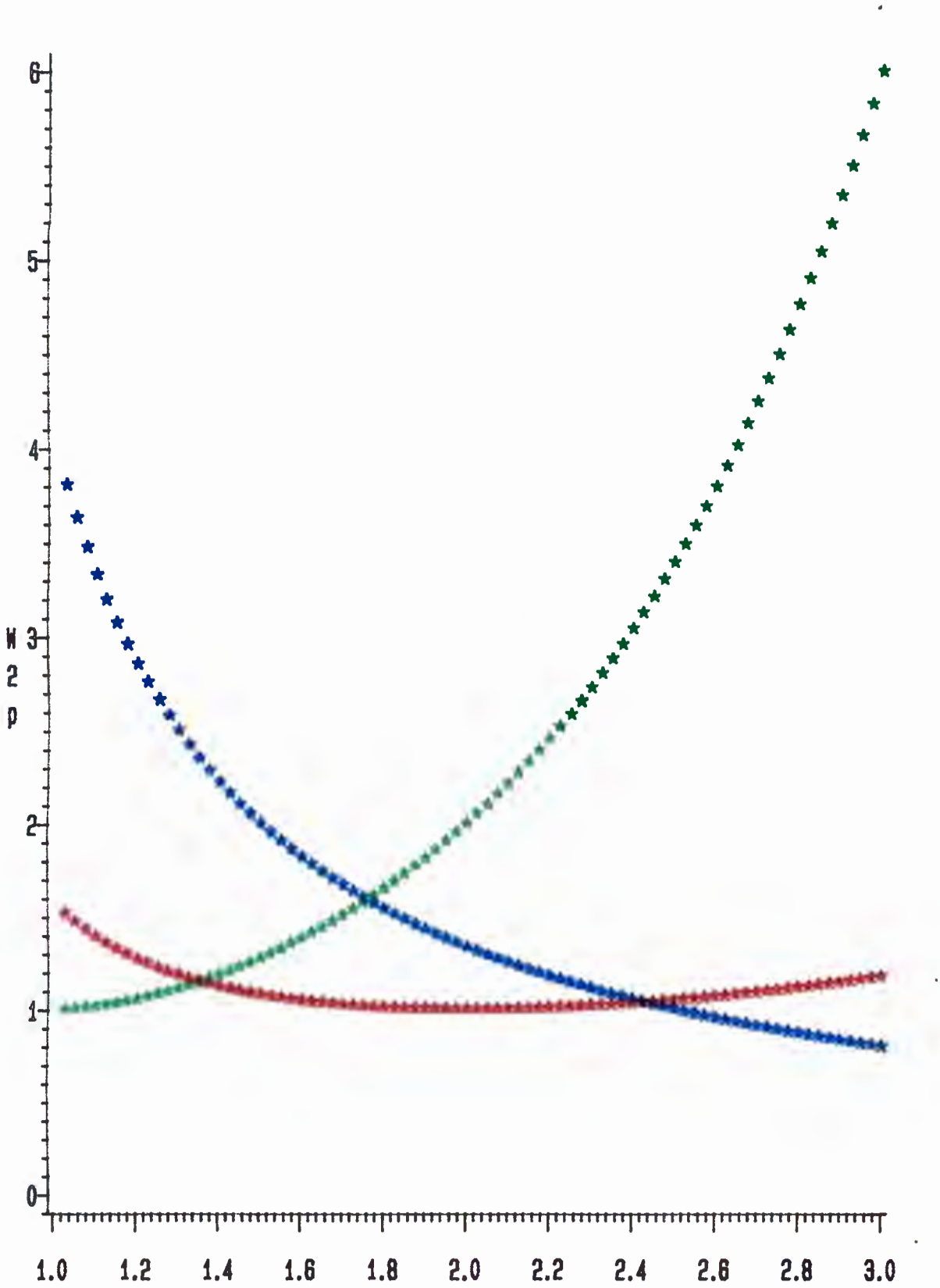


*Fig. 4.3: (Wp)**2 for 3 power distributions vs p*



Chapter 5 : The practical application of adaptive nonlinear L_p-norm estimation.

In Chapters 2 and 3 the actual solving of nonlinear L_p-norm problems was considered and in Chapter 4 the relationship between the true kurtosis of the regression error distribution and the value of p was examined. In this chapter the determination of the optimal p-value for the solution of a given L_p-norm estimation problem will be the object of scrutiny. An adaptive procedure for determining the optimal p-value for a given set of data will also be derived. A Monte Carlo simulation study, utilizing the adaptive procedure, will be performed to determine the empirical distribution of the optimal p-value for an unknown error distribution. This adaptive procedure will then be applied to the estimation of the model parameters of two physiological processes. Its value in identifying outlying observations will also be illustrated.

1. The adaptive algorithm for L_p-norm estimation.

Two formulae relating p to the kurtosis of the error distribution have been proposed in linear L_p-norm estimation. These are:

(5.1) $\quad p = 9/(\beta_2)^2 + 1$

by Money et al. (1982) and

(5.2) $\quad p = 6/\beta_2$

by Sposito et al. (1983).

2. A simulation study to determine the empirical distribution of the estimate of the optimal p-value.

In order to gain some insight into the behaviour of the optimal p a restricted simulation study was undertaken. The model employed in Chapter 4 will be used here:

$$y = 5 + 4\exp(\Theta_1 x_1) + 3\exp(\Theta_2 x_2) + e \,, \tag{5.3}$$

where the coefficients 5,4 and 3 were arbitrarily chosen. The following symmetric error distributions will be considered: normal(β_2=3), uniform (β_2=1.8), parabolic (β_2=2.14) and Laplace (β_2=6). It was assumed that $E(e_i) = 0$ and $Var(e_i) = \sigma^2 = 9$ for all the error distributions.

In the simulation study we shall again fix the values of Θ_1 and Θ_2 in model (5.3) at 1.0 and 1.5 respectively. A sample size of n values for x_{1i} and x_{2i} from a [0.5,1.5] uniform distribution will be selected and held fixed. In each case 500 experiments were simulated such that the error term was a random variable from a specified distribution (see Appendix C for the subroutine). Various sample sizes n were used in order to study the asymptotic properties of the optimal p-value.

The adaptive procedure described in the previous paragraph was used to determine the optimal p-value for each simulation experiment. The optimal p-value, the mean, variance, skewness and kurtosis of the resulting residual distribution were recorded for each of the 500 experiments. This process was then repeated for each of the four regression error distributions for differing sample sizes (n = 30,50,100,200,400). The distribution of the optimal p-value for a given error distribution was then determined from the results of the 500 simulation experiments.

Goodness-of-fit tests were then used to determine whether the optimal p-values are normally distributed. These tests use statistics based on the observed empirical distribution function (EDF) of the data. EDF statistics are discussed by Stephens (1974). The Cramer-von Mises test statistic $W^*(W^2)$ was used in this simulation study as it is appropriate when testing for normality where the two unknown parameters (μ and σ^2) are estimated by the sample mean and variance. A discussion of this test may be found in Appendix D.

The simulation results based on additive normally distributed errors and using the prediction formulae (5.1) and (5.2) for p, are given in Tables 5.1a and 5.1b respectively.

TABLE 5.1a MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (normal errors, Money et al. formula).

	Moments of p				W^*	
Sample size n	Mean	Variance	Skewness	Kurtosis		
30	2.373	0.345	0.491	3.043	0.287	(P<0.01)
50	2.261	0.203	0.489	3.155	0.215	(P<0.01)
100	2.117	0.101	0.236	2.901	0.129	(P<0.05)
200	2.070	0.048	0.185	2.776	0.074	(P>0.15)
400	2.040	0.028	0.186	2.728	0.075	(P>0.15)

TABLE 5.1b MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (normal errors, Sposito et al. formula).

	Moments of p				W^*	
Sample size n	Mean	Variance	Skewness	Kurtosis		
30	2.242	0.260	-0.157	2.725	0.110	(P>0.05)
50	2.219	0.164	-0.175	2.973	0.093	(P>0.10)
100	2.097	0.093	-0.146	2.761	0.095	(P>0.10)
200	2.056	0.047	-0.206	3.358	0.069	(P>0.15)
400	2.019	0.028	-0.132	3.279	0.097	(P>0.10)

The unexpected result is that the optimal p-values are ASYMPTOTICALLY NORMAL. Note that the Cramer-von Mises statistic indicates that the p-values are normally distributed for sample sizes $n \geq 200$ (formula (5.1)) and smaller samples $n \geq 30$ (formula (5.2)). Sposito et al. suggest that that their formula (5.2) is suitable for large samples ($n \geq 200$) and that formula (5.1) may be applied to small samples where the error distribution has a finite range. (The reverse seems to apply in the above situation). The mean value of p approaches 2 as the sample size increases. The variance decreases with increasing sample size. The skewness appears to oscillate around 0 and the kurtosis around 3. The variation in kurtosis is not unexpected since it is known that the variance of the kurtosis is large (see Kendall and Stuart (1963) volume 1, p 243).

The mean of the optimal p-value as predicted by both formulae tends to 2 which is expected since least squares is the appropriate method to use in the case of normally distributed errors. Hence, in the case of normal additive regression errors either formula may be used since they are both efficient in predicting the optimal p-value.

The additional knowledge we now have is that the optimal p-values are asymptotically normally distributed. We shall subsequently see that this result is true for other model error distributions as well.

The simulation results based on additive uniformly distributed errors and using the prediction formulae (5.1) and (5.2) for p, are given in Tables 5.2a and 5.2b respectively.

TABLE 5.2a MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (uniform errors, Money et al. formula).

	Moments of p				W^*	
Sample size n	Mean	Variance	Skewness	Kurtosis		
30	3.632	0.532	0.423	3.284	0.160	(P<0.01)
50	3.697	0.309	0.270	3.220	0.067	(P>0.15)
100	3.742	0.135	0.210	3.099	0.041	(P>0.15)
200	3.745	0.063	-0.070	2.963	0.055	(P>0.15)
400	3.753	0.033	-0.055	3.141	0.036	(P>0.15)

TABLE 5.2b MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (uniform errors, Sposito et al. formula).

	Moments of p				W^*	
Sample size n	Mean	Variance	Skewness	Kurtosis		
30	3.200	0.188	-0.269	3.131	0.106	(P>0.05)
50	3.259	0.102	-0.243	3.427	0.096	(P>0.10)
100	3.300	0.052	0.042	3.046	0.095	(P>0.10)
200	3.307	0.023	-0.205	3.037	0.072	(P>0.15)
400	3.313	0.012	0.034	2.641	0.082	(P>0.15)

It is known (Johnson and Kotz (1970)) that in the location model if we minimize $\max_i |x_i - \theta|$ then $\hat{\theta}$ = midrange (x_i) will also be the maximum likelihood estimate of θ if the x_i's follow a uniform distribution. This is also known as Chebychev estimation and hence $p \to \infty$ should be used. In the case of uniformly distributed errors, the optimal p-value is predicted as 3.79 by formula (5.1). However, formula (5.2) predicts it as 3.33. Observe that for large sample sizes ($n \geq 400$) the mean value of p (using either formula for predicting p) approaches this corresponding value. Observe therefore that although both formulae will yield optimal p-values that are asymptotically normal, the means of the optimal p-values will differ markedly.

Next consider the case where the errors are additive and parabolically distributed. The results of this simulation using prediction formulae (5.1) and (5.2) for p, are given in Tables 5.3a and 5.3b respectively.

TABLE 5.3a MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (parabolic errors, Money et al. formula).

	Moments of p				W^*	
Sample size n	Mean	Variance	Skewness	Kurtosis		
30	3.042	0.388	0.454	3.100	0.219	(P<0.01)
50	2.995	0.221	0.285	3.180	0.136	(P<0.05)
100	2.985	0.101	0.137	2.876	0.082	(P>0.15)
200	2.955	0.045	0.283	3.228	0.087	(P>0.15)
400	2.970	0.022	-0.023	2.844	0.109	(P>0.05)

TABLE 5.3b MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (parabolic errors, Sposito et al. formula).

	Moments of p				W^*	
Sample size n	Mean	Variance	Skewness	Kurtosis		
30	2.815	0.187	-0.069	2.914	0.036	(P>0.15)
50	2.820	0.107	0.050	3.315	0.050	(P>0.15)
100	2.785	0.050	-0.071	3.153	0.034	(P>0.15)
200	2.803	0.023	0.085	3.072	0.028	(P>0.15)
400	2.803	0.014	0.068	2.500	0.098	(P>0.10)

It is interesting that the optimal p-values are normally distributed even when the sample sizes are small ($n \geq 30$ using formula (5.2) and $n \geq 50$ using formula (5.1)). In the case of parabolic errors, the true optimal value is predicted as 2.96 by formula (5.1) whilst formula (5.2) predicts it as 2.8. Observe that for large sample sizes ($n \geq 400$) the mean value of p (using either formula for predicting p) approaches this corresponding value. Observe therefore that although both formulae will yield optimal p-values that are asymptotically normal, the means of the optimal p-values will again differ markedly.

Finally consider the case where the errors are additive and Laplace distributed. The simulation results are given in Tables 5.4a and 5.4b.

TABLE 5.4a MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (Laplace errors, Money et al. formula).

Sample size n	Moments of p Mean	Variance	Skewness	Kurtosis	W^2	
30	1.685	0.218	1.390	5.639	1.902	(P<0.01)
50	1.536	0.106	1.104	4.965	1.391	(P<0.01)
100	1.406	0.042	0.653	3.252	0.451	(P<0.01)
200	1.334	0.021	0.382	2.768	0.201	(P<0.01)
400	1.291	0.012	0.329	2.535	0.342	(P<0.01)

TABLE 5.4b MOMENTS OF OPTIMAL p FOR VARYING SAMPLE SIZES n BASED ON 500 EXPERIMENTS (Laplace errors, Sposito et al. formula).

Sample size n	Moments of p Mean	Variance	Skewness	Kurtosis	W^2	
30	1.569	0.297	0.336	2.747	0.161	(P<0.05)
50	1.419	0.192	0.252	2.923	0.060	(P>0.15)
100	1.281	0.115	0.060	2.765	0.033	(P>0.15)
200	1.134	0.065	0.030	2.576	0.047	(P>0.15)
400	1.059	0.041	-0.082	2.478	0.105	(P>0.05)

The following interesting situation occurs: Recall that formula (5.2) predicts the optimal p-values for Laplace distributed errors to be equal to 1. Observe that these optimal p-values are asymptotically normal with mean value approximately equal to 1. It is known (Johnson and Kotz (1970)) that in the location model if we minimize

$$\sum_{i=1}^{n} |x_i-\theta| \text{ then } \hat{\theta} = \text{median } (x_i) \text{ will also be the maximum likelihood}$$

estimate of θ if the x_i's follow a Laplace distribution.

Money et al. op. cit. found in a simulation study that linear L_1-norm estimation is only 60% as efficient as linear $L_{1.25}$-norm estimation when the errors follow a Laplace distribution. Their formula therefore predicts this optimal p-value to be 1.25. In this case the optimal p-values are no longer asymptotically normal.

The fact that the p-values are normally distributed enables us to construct a 95% confidence interval for the mean of the optimal p-values. These intervals are, however, narrow. They are nevertheless shown in Table 5.5 for the various error distributions (except for the Money et al. formula with Laplace errors). A sample size of 400 was used.

TABLE 5.5 95% CONFIDENCE INTERVALS FOR THE MEAN OF THE OPTIMAL p-VALUES (n=400) .

Error distribution	95% confidence interval	
	Formula (5.1)	Formula (5.2)
Uniform	[3.73,3.76]	[3.30,3.32]
Parabolic	[2.96,2.98]	[2.80,2.81]
Normal	[2.03,2.06]	[2.02,2.04]
Laplace	-	[1.05,1.07]

In conclusion we can say that both formulae predict optimal p-values that are asymptotically normal (except where formula (5.1) is applied to Laplace errors). The means of the optimal p-values differ for the two formulae (except for normal errors). Throughout we observe that the variance of the optimal p-values is small and decreases with an increase in sample size. The skewness oscillates around 0 with diminishing amplititude as the sample size increases. The kurtosis, though quite variable, oscillates around 3.

3. The application of the adaptive L_p-norm estimation procedure.

In the remainder of this chapter we shall consider the application of L_p-norm estimation in practical problems. We shall, for example, consider the modelling of certain physiological processes of interest in medical research. We shall also see how useful the adaptive procedure is in identifying outlying observations.

3.1 Oxygen saturation in respiratory physiology.

The following problem has been considered by Du Toit and Gonin (1982) in the least squares context.

Pulmonary disorders such as asthma, chronic bronchitis and emphysema produce specific patterns when lung function tests are performed on the patient. An important lung function test involves the relationship between oxygen saturation in arterial blood (So_2) and the partial pressure of oxygen in the pulmonary capillary (Po_2). So_2 is expressed as a percentage (%) and Po_2 is measured in mm of mercury (mm Hg). The relationship So_2/Po_2 is important because although Po_2 can be measured directly in arterial blood; this requires catheterisation, an invasive procedure carrying the risk of infection as well as one which causes discomfort to the severely ill patient. Oxygen saturation, however, can be measured non-invasively using a fibre-optic ear oximeter. Once the saturation is known, the Po_2 can be calculated by means of the oxygen dissociation curve (So_2/Po_2 relationship). This calculation is only possible and valid if the So_2/Po_2 relationship is mathematically invertible. Du Toit and Gonin op. cit. discuss the derivation of the model and its inverse.

The data which have previously been used to describe this relationship (see Severinghaus (1979)) are given in Table 5.6 and depicted in Figure 5.1.

TABLE 5.6 : OXYGEN SATURATION DATA: So_2 vs Po_2 .

Po_2(mm Hg)	So_2(%)	Po_2(mm Hg)	So_2(%)	Po_2(mm Hg)	So_2(%)
4.00	2.56	36.00	68.63	80.00	95.84
6.00	4.37	38.00	71.94	85.00	96.42
8.00	6.68	40.00	74.69	90.00	96.88
10.00	9.58	42.00	77.29	95.00	97.25
12.00	12.96	44.00	79.55	100.00	97.49
14.00	16.89	46.00	81.71	110.00	97.91
16.00	21.40	48.00	83.52	120.00	98.21
18.00	26.50	50.00	85.08	130.00	98.44
20.00	32.12	52.00	86.59	140.00	98.62
22.00	37.60	54.00	87.70	150.00	98.77
24.00	43.14	56.00	88.93	175.00	99.03
26.00	48.27	58.00	89.95	200.00	99.20
28.00	53.16	60.00	90.85	225.00	99.32
30.00	57.54	65.00	92.73	250.00	99.41
32.00	61.69	70.00	94.06		
34.00	65.16	75.00	95.10		

The following Gompertz model will be used:

$$(5.4) \quad So_2 = \theta_1 \exp(-\theta_2(\theta_3)^{Po_2}) \quad , \; 0 < \theta_3 < 1.$$

The initial values are $\underline{\theta}^\circ = (98.0, 4.6, 0.93)$ $S_2(\underline{\theta}^\circ) = 62.4101$. We shall use the adaptive procedure of paragraph 1 to calculate the optimal value of p for this data using both formulae (5.1) and (5.2). To initiate the estimation procedure we shall use least squares.

The optimal solutions for the respective optimal p-values using formulae (5.1) and (5.2) are given in Table 5.7

TABLE 5.7 OPTIMAL SOLUTIONS FOR DIFFERING OPTIMAL p-VALUES.

	p	$\underline{\theta}^*$			$S_p(\underline{\theta}^*)$
Least squares	2.00	98.0012	4.6059	0.9316	23.9549
Formula (5.1)	3.49	98.1407	4.5753	0.9319	21.2424
Formula (5.2)	3.15	98.1194	4.5799	0.9318	21.6380

We see that the estimated parameter values only differ in the first decimal for the three L_p-norm estimations. It is a large residual problem. The results of the adaptive procedure and the moments of the resulting residual distribution are given in Table 5.8a and 5.8b.

TABLE 5.8a CALCULATION OF THE OPTIMAL p-VALUE OF OXYGEN SATURATION DATA (Money et al. formula).

Step	p_i	Moments of residual distribution				Predicted p_{i+1}
i		Mean	Variance	Skewness	Kurtosis	
1	2.000	-0.047	0.519	0.539	1.944	3.381
2	3.381	-0.124	0.513	0.518	1.903	3.487
3	3.487	-0.127	0.513	0.518	1.901	3.491
4	3.487	-0.127	0.513	0.518	1.901	3.491

TABLE 5.8b CALCULATION OF THE OPTIMAL p-VALUE OF OXYGEN SATURATION DATA (Sposito et al. formula) .

Step	p_i	Moments of residual distribution				Predicted p_{i+1}
i		Mean	Variance	Skewness	Kurtosis	
1	2.000	-0.047	0.519	0.539	1.944	3.086
2	3.086	-0.113	0.514	0.521	1.908	3.145
3	3.145	-0.115	0.514	0.521	1.907	3.147
4	3.145	-0.115	0.514	0.521	1.907	3.147

Convergence occurred in 4 iterations. In the first and subsequent steps we notice that the kurtosis of the residuals is less than 2, hence the normality of the data is in question. Note how the mean value of the residuals increases with increasing values of p. The residuals have a kurtosis markedly less than 3. The Cramer-von Mises test for normality of the errors yielded $W^{*}=0.2397$, $P < 0.01$ in the case of least squares. Observe that the residuals are slightly positively skew hence care should be taken when using either formula (5.1) or (5.2) where symmetry has been assumed. We observe a marked difference between the final values of p as calculated by formulae (5.1) and (5.2) respectively. Note, however, that the moments of the residual distributions for the two formulae are in agreement to one decimal place. It is encouraging to note that the discrepancy in the predicted p-values does not seem to affect the residual distribution.

The differences in the predicted optimal p-values are due to inherent differences in the formulae and are exacerbated if the residual distribution is significantly skew. Further research should therefore concentrate on the derivation of an empirical relationship of the form:

$$p = f(\text{skewness}, \text{kurtosis})$$

which will hopefully alleviate the inherent differences encountered in formulae (5.1) and (5.2).

The L_p-norm estimation can also be used to show its efficacy in predicting a future observation. We shall then compare it with the predictive ability of least squares. The Severinghaus (1979) data contain an additional point viz: $Po_2 = 500$ and $So_2 = 99.72$ %. The following predictions can be made:

p	$\hat{So}_2$	$So_2 - \hat{So}_2$
2.00	98.00	1.72
3.49	98.14	1.58
3.15	98.12	1.60

We see that a marginal improvement over least squares was found when the alternative p-values were used. In general we should expect some improvement in the prediction of future obsevations when least squares is no longer adequate as an estimation procedure.

A significantly better fit to the data is found if the following extended Richards growth curve (Du Toit and Gonin op. cit.) is fitted:

$$So_2 = \theta_5\theta_1^{\theta_2^{(1-\theta_3^{Po_2})^{\theta_4}}} \qquad (5.5)$$

with $0 < \theta_i \leq 1$ for i=1,2,3

$1 \leq \theta_4$ and $\theta_5 \leq 100$.

The starting values that were used are:

$\underline{\theta}^\circ = (0.0148, 0.0014, 0.9820, 1.3465, 100.0)$ $\quad S_2(\underline{\theta}^\circ) = 8.51353.$

The optimal solutions for the respective optimal p-values using formulae (5.1) and (5.2) are given in Table 5.9:

TABLE 5.9 OPTIMAL SOLUTIONS FOR DIFFERING OPTIMAL p-VALUES.

	p	$\underline{\theta}^*$					$S_p(\underline{\theta}^*)$
Least squares	2.00	0.0125	0.0007	0.9843	1.2808	99.6080	1.31436
Formula (5.1)	2.56	0.0126	0.0008	0.9840	1.2887	99.6576	0.60014
Formula (5.2)	2.48	0.0126	0.0008	0.9840	1.2876	99.6515	0.66308

We see that the estimated parameter values only differ in the second and third decimal for the three L_p-norm estimations. Note that the residual sum of squares is now markedly smaller than in the case of the Gompertz fit. The results of the adaptive procedure along with the moments of the residual distribution are given in the Tables 5.10a and 5.10b.

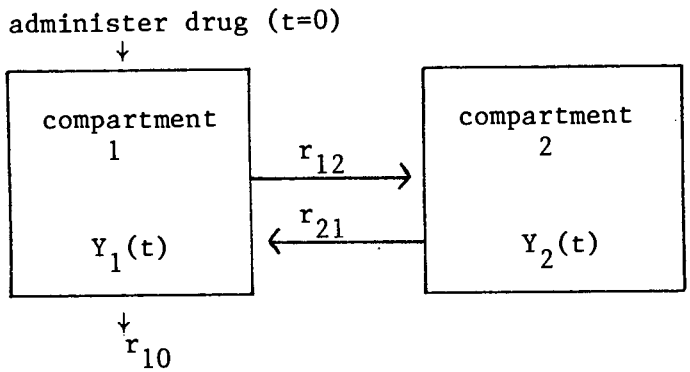
3.2 Mathematical models of drug bioavailability.

Metzler (1974) states that

> "the object of a bioavailability study is to quantify the relative amount and rate of absorption of the administered drug which reaches the general circulation intact."
> comparative bioavailability is defined as the the comparison of the bioavailability of the new drug to a standard or reference drug.

In bioavailability studies compartmental models are used. One such compartment, the vascular compartment contains the drug and carries it to another compartment the tissue. These compartments form a system characterised by the transfer rate of the drug between compartments.

Consider a two compartment system where the one is open to the environment (for example oral administration or intravenous injection). Schematically:



Legend: r_{ij} = constant transfer rate from compartment i to j.

$i=1,2$ $j=0,1,2.$

Y_i = drug concentration in compartment i at time t.

Assuming that the rate of change in concentration is proportional to the amount present, we can derive the following system of differential equations:

$$Y_1'(t) = -(r_{10} + r_{12})Y_1(t) + r_{21}Y_2(t)$$
$$Y_2'(t) = \quad r_{12}Y_1(t) - r_{21}Y_2(t).$$

This can be solved by the standard analytical techniques of linear differential equations to yield:

$$Y_1(t) = P_1\exp(-\lambda_1 t) + P_2\exp(-\lambda_2 t) \tag{5.6}$$

where the roots λ_i are given by:

$$-(r_{10} + r_{12} + r_{21}) \pm ((r_{10} + r_{12} + r_{21})^2 - 4r_{10}r_{21})^{\frac{1}{2}}$$

and P_1 and P_2 are arbitrary constants.

Since the concentration of the drug is measured in the vascular compartment we are only interested in $Y_1(t)$ (the transfer rates cannot be obtained directly from the differential equations). We shall therefore fit a model which resembles the analytical solution (5.6). Denote the observed concentrations $Y_1(t)$ by $y(t)$ and use the model

(5.7) $$y(t) = \theta_1 \exp(-\theta_3 t) + \theta_2 \exp(-\theta_4 t).$$

Metronidazole is a drug used to combat anaerobic infections. Two standard formulations of the drug were compared in a bioavailability study. Ten patients were used in a two period cross-over experimental design. Further information on the experimental design may be found in Juritz, Gonin and Bridle (1983). The drug formulation was administered and the concentration measured at nineteen different times for each patient on each given treatment. We shall only consider the actual fitting of the model to the data.

The data of one such patient are given in Table 5.11 and depicted in Figure 5.2.

TABLE 5.11 METRONIDAZOLE CONCENTRATION OVER TIME.

Time (Hrs)	Concentration (μg/ml)	Time (Hrs)	Concentration (μg/ml)	Time (Hrs)	Concentration (μg/ml)
0.000	0.000	4.000	6.866	10.000	3.424
1.000	0.795	4.500	8.516	12.000	3.206
1.500	3.527	5.000	8.042	18.000	1.558
2.000	4.066	5.500	8.159	24.000	0.701
2.500	6.246	6.000	5.276	37.500	0.087
3.000	6.972	7.000	5.126		
3.500	7.637	8.000	4.462		

We shall now show that model (5.7) adequately describes the data. Function y(t) has a maximum where

$$y'(t) = 0 \text{ , i.e. where } t = t_{max} = \ln(-\theta_1\theta_3 / \theta_2\theta_4)/(\theta_3 - \theta_4).$$

$(y''(t) < 0$ at $t = t_{max}$ provided that $0 < \theta_3 < \theta_4)$.

Note that $y \to 0$ as $t \to 0$ i.e. $\theta_1 + \theta_2 \simeq 0 \Rightarrow \theta_1 \simeq -\theta_2$; and that $y \to 0$ as $t \to \infty$.

Once the model describing the drug concentrations is fitted the following bioavailability parameters can be calculated:

1) The area under the curve (AUC):

$$AUC = \int_0^\infty y(t)dt = \theta_1/\theta_3 + \theta_2/\theta_4$$ which will have units mass.time e.g. μg-hours.

2) Time till maximum concentration:

$$t_{max} = \ln(-\theta_1\theta_3/\theta_2\theta_4)/(\theta_3-\theta_4).$$

3) Maximum concentration $C_{max} = y(t_{max})$.

Similarly a 3-compartment (6-parameter) model could be used in which case the concentration is modelled by

$$(5.8) \qquad y(t) = \theta_1 \exp(-\theta_4 t) + \theta_2 \exp(-\theta_5 t) + \theta_3 \exp(-\theta_6 t) \ .$$

We shall fit this model to the data of Table 5.11.

The initial values are:

$$\underline{\theta}^\circ = (50,-200,200,0.1,0.3,0.5) \qquad S_2(\underline{\theta}^\circ) = 3281.95 \ .$$

The optimal solutions for the respective optimal p-values using formulae (5.1) and (5.2) are given in Table 5.12:

TABLE 5.12 OPTIMAL SOLUTIONS FOR DIFFERING OPTIMAL p-VALUES .

	p	$\underline{\theta}^*$						$S_p(\underline{\theta}^*)$	AUC
Least squares	2.00	17.65	-315.04	297.41	0.16	0.98	1.05	7.593	74.61
Formula (5.1)	2.48	18.19	-315.37	297.25	0.16	0.96	1.03	7.267	73.35
Formula (5.2)	2.43	18.14	-315.20	307.12	0.16	0.96	1.03	7.290	74.25

We see that the estimated parameter values and AUC's agree to one significant digit for the three L_p-norm estimations. The results of the adaptive procedure and the moments of the resultant residual distribution are given in the Tables 5.13a and 5.13b.

TABLE 5.13a CALCULATION OF THE OPTIMAL p-VALUE OF METRONIDAZOLE DATA (Money et al. formula).

Step	p_i	moments of residual distribution				Predicted p_{i+1}
i		Mean	Variance	Skewness	Kurtosis	
1	2.000	0.029	0.399	-0.133	2.483	2.459
2	2.459	0.033	0.399	-0.184	2.467	2.479
3	2.479	0.033	0.399	-0.186	2.467	2.479

TABLE 5.13b CALCULATION OF THE OPTIMAL p-VALUE OF METRONIDAZOLE DATA (Sposito et al. formula).

Step	p_i	moments of residual distribution				Predicted p_{i+1}
i		Mean	Variance	Skewness	Kurtosis	
1	2.000	0.029	0.399	-0.133	2.483	2.416
2	2.416	0.032	0.399	-0.181	2.469	2.430
3	2.430	0.032	0.399	-0.182	2.468	2.431

We see also that the optimal values of p as predicted by the two formulae do not differ much (2.48 and 2.43). Note also that the moments of the resultant residual distributions are in agreement to two decimal places. These moments do not differ markedly from the moments of the least squares residuals. Note that the variance of the residuals remains constant at 0.399. We therefore conclude that both formulae appear to predict a p-value that results in a residual distribution with similar properties. It is also reasonable to conclude that the L_p-norm estimations with p=2.48 and p=2.43 will be as efficient as least squares estimation in this example.

3.3 Outlying observations.

Frome and Yakatan (1980) consider the one compartment model which can be formulated as:

$$y(t) = \frac{\theta_3\theta_1}{\theta_1-\theta_2}(\exp(-\theta_2 t) - \exp(-\theta_1 t)). \tag{5.9}$$

The parameter θ_1 is the absorption rate constant, θ_2 the elimination rate constant whilst $\theta_3 = fD/V$ where D is the initial amount of drug administered, V the volume of distribution and f the fraction of the drug absorbed from the gastro-intestinal tract. Note that θ_3 can therefore be interpreted as physical density (mass/volume).

The bioavailability parameter AUC can be calculated by:

$$AUC = \int_0^\infty y(t)dt = \theta_3/\theta_2$$ (there is a misprint in the article which gives it as θ_3/θ_1).

The authors claim that gradient methods for parameter estimation often fail on real data and suggest the Nelder-Mead derivative-free method for solving nonlinear least squares problems. This method exhibits fair convergence behaviour (see for example Himmelblau (1972), Himmelblau and Lindsay (1980) who conducted a survey of unconstrained methods). It was found that the gradient method proposed in Chapter 3 converges well for nonlinear least squares problems and is considerably more efficient than the Nelder-Mead algorithm (see also program MINPACK by Moré et al. (1980)). Both these methods are numerically stable.

The authors also suggest L_1-norm estimation for outliers since it is known to be resistant to outlying observation. They use the Nelder-Mead algorithm to solve the L_1-norm problem and do not refer to the work by Osborne and Watson (1971) on nonlinear L_1-norm estimation problems (Chapter 4).

Frome and Yakatan op. cit. suggest the consideration of two factors in the parameter estimation: the first is the sampling distribution of the response variable and the second the resistance of the estimation procedure to outlying observations. The first aspect was considered in Chapter 4. The second aspect will be dealt with here. The adaptive procedure will predict the optimal p-value close to 1 with the result that very little weight is attached to the outlying observations.

Rodda, Sampson and Smith (1975) also considered the effect of outliers on least squares estimation in the one compartment model. Frome and Yakatan adjust their data to contain certain patterns of outliers. We shall consider two of these 8 data sets (Patterns 0 and 1) and suggest one of our own (Pattern 8). (There appears to be a misprint in their Pattern 7 since it is identical to Pattern 0).

Pattern 0

In Pattern 0 we found the residual sum of squares using both our large residual method (p=2) and MINPACK to be 0.0104. The starting values $\underline{\theta}^{\circ} = (25,1,10)$ used by Frome and Yakatan op.cit. were also used here.

In view of the small sample size and the fact that the data are fairly well-behaved the optimal p-value as predicted by formulae (5.1) or (5.2) will not improve much on the least squares result. This shows that least squares may be used and that even if the predicted p-values are used the parameter values and the bioavailibility parameter AUC remain unaffected. The results are given in Table 5.14.

TABLE 5.14 PATTERN 0 : OPTIMAL SOLUTIONS FOR DIFFERING OPTIMAL p-VALUES.

	p	$\underline{\theta}^{*}$			$S_p(\underline{\theta}^{*})$	No. function evaluations	AUC
Least squares	2.00	2.995	0.300	50.014	0.0104	19	166.5
Formula (5.1)	3.67	2.993	0.300	50.011	0.0000	19+2+1+1=23	166.6

Note that although the optimal p-values differ markedly the optimal parameter values and AUC are the same. (True values are $\underline{\theta}^* = (3, 0.3, 50)$ and AUC=166.7).

Pattern 1

In Pattern 1 we can see the effect of one outlier and the value of our adaptive procedure in identifying such an outlier. This is achieved by L_p-norm approximation when p is close to 1. The results are given in Table 5.15.

TABLE 5.15 PATTERN 1 : OPTIMAL SOLUTIONS FOR DIFFERING OPTIMAL p-VALUES.

	p	$\underline{\theta}^*$			$S_p(\underline{\theta}^*)$	No. function evaluations	AUC
Least squares	2.00	2.144	0.306	49.490	304.682	21	161.6
Formula (5.1)	1.06	2.994	0.301	50.032	24.826	40+22+24=86	166.4
Formula (5.2)	0.50	3.003	0.300	49.971	6.116	19+111+8=138	166.8

Observe the marked deviation of the parameter values and value of AUC from the true values in the case of least squares. The adaptive L_p-norm procedure estimates the parameters and AUC value close to the true values. The slow convergence when formula (5.2) is used is due to the fact that p-values ≤ 1 are predicted and our algorithm is not designed to solve L_p-norm problems for p values ≤ 1.

The data, fitted values and observed residuals are given in Table 5.16.

TABLE 5.16 PATTERN 1 : DATA, FITTED VALUES AND RESIDUALS (p=2, 0.5, 1.06) .

		p=2		p=0.50		p=1.06	
t_i	y_i	$\hat{y}_i$	$\hat{y}_i - y_i$	$\hat{y}_i$	$\hat{y}_i - y_i$	$\hat{y}_i$	$\hat{y}_i - y_i$
0.083	10.9	7.96	-2.94	10.88	-0.02	10.87	-0.03
0.167	19.1	14.50	-4.60	19.18	0.08	19.16	0.06
0.250	25.3	19.70	-5.60	25.30	-0.00	25.28	-0.02
0.500	15.0"	29.78	14.78	35.42	20.42 ←	35.41	20.41 ←
0.750	38.5	34.33	-4.17	38.50	-0.00	38.50	-0.00
1.000	38.4	35.74	-2.66	38.38	-0.02	38.39	-0.01
1.500	34.8	34.15	-0.65	34.80	0.00	34.80	0.00
2.250	28.2	28.52	0.32	28.23	0.03	28.20	0.00
3.000	22.6	22.94	0.34	22.59	-0.01	22.55	-0.05
4.000	16.7	16.95	0.25	16.75	0.05	16.70	0.00
6.000	9.2	9.19	-0.01	9.20	0.00	9.15	-0.05
8.000	5.0	4.98	-0.02	5.05	0.05	5.01	0.01
10.000	2.8	2.70	-0.10	2.78	-0.02	2.75	-0.05
12.000	1.5	1.46	-0.04	1.52	0.02	1.51	0.01

("true value is 35.4).

With the optimal p=1.06 (or 0.5) the outlier at t=0.5 with a residual of 20.4 is clearly identified while all the remaining residuals are small. When we examine the residuals in the least squares case we find that the first 6 residuals are all larger than 1.0 whilst in the former case all the remaining residuals are less than 0.1. L_p-norm estimation is therefore useful in not only identifying an outlier but is also sufficiently robust to cope with such an outlier.

Pattern 8

In this example there are two outliers present. It differs from the other patterns of Frome and Yakatan where outliers occur as adjacent values. In this example non-adjacent values were chosen.

The results are given in Table 5.17.

TABLE 5.17 PATTERN 8 : OPTIMAL SOLUTIONS FOR DIFERING OPTIMAL p-VALUES .

	p	$\underline{\theta}^*$			$S_p(\underline{\theta}^*)$	No.function evaluations	AUC
Least squares	2.00	2.566	0.245	44.582	474.511	32	182.3
Formula (5.1)	1.15	2.987	0.300	49.962	52.405	32+13+35=80	166.8
Formula (5.2)	0.78	2.987	0.300	49.962	18.731	32+42+22=96	166.7

Again observe the marked deviation of the parameter values from the true parameter values when least squares is used. The adaptive L_p-norm estimation procedure correctly estimates the parameters and AUC value.

The data, fitted values and observed residuals are given in Table 5.18.

TABLE 5.18 PATTERN 8 : DATA, FITTED VALUES AND RESIDUALS (p=2, 0.78, 1.15).

		p=2		p=0.78		p=1.15	
t_i	y_i	$\hat{y}_i$	$\hat{y}_i-y_i$	$\hat{y}_i$	$\hat{y}_i-y_i$	$\hat{y}_i$	$\hat{y}_i-y_i$
0.083	10.9	8.46	-2.44	10.83	-0.07	10.83	-0.07
0.167	19.1	15.20	-3.90	19.10	0.00	19.10	0.00
0.250	25.3	20.41	-4.89	25.21	-0.09	25.21	-0.09
0.500	15.0"	29.95	14.95	35.33	20.33←	35.33	20.33←
0.750	38.5	33.83	-4.67	38.45	-0.05	38.45	-0.05
1.000	38.4	34.80	-3.60	38.35	-0.05	38.35	-0.05
1.500	34.8	33.10	-1.70	34.80	-0.00	34.80	0.00
2.250	28.2	28.27	0.07	28.23	0.03	28.23	0.03
3.000	22.6	23.63	1.03	22.60	-0.00	22.60	-0.00
4.000	16.7	18.52	1.82	16.75	0.05	16.75	0.05
6.000	9.2	11.35	2.15	9.20	-0.00	9.20	0.00
8.000	5.0	6.96	1.96	5.05	0.05	5.05	0.05
10.000	2.8	4.26	1.46	2.78	-0.02	2.78	-0.02
12.000	15.0"	2.61	-12.39	1.52	-13.48←	1.52	-13.48←

(" true values are 35.4 and 1.5).

With the optimal p=1.15 (or 0.78) we can clearly identify the outliers at t=0.5 and t=12. The observed values and fitted values for p=2 and p=1.5 are depicted in Figures 5.3 to 5.5. It may be argued that these outliers come from a skew distribution. Ekblom (1974) suggested in linear estimation that p-values $\leq$ 1 should be used for skewly distributed errors. Our results suggest that his argument may be valid for nonlinear estimation problems with skewly distributed errors.

Note that in all the examples the optimal p-value as predicted by formula (5.1) was larger than the value predicted by (5.2).

We conclude that our algorithm is efficient in solving L_p-norm estimation problems. The suggested adaptive procedure also deals efficiently with outliers and identifies them in a systematic way. In well behaved problems, where least squares is appropriate, the procedure predicts p-values which yield equally efficient results.

Appendix D: The Cramer-von Mises goodness-of-fit test.

Suppose the observations have been arranged in ascending order.

$$r_1 < r_2 \ldots < r_n .$$

We shall assume that the population parameters μ and σ^2 are unknown but estimated by $\bar{r}$ and $s^2 = \sum_{i=1}^{n} (r_i - \bar{r})^2 /(n-1)$.

Let the standardised normal distribution function be given by $F(w)$.

The test proceeds as follows:

Step 1: Calculate the quantities $w_i = (r_i - \bar{r})/s$.

Step 2: Calculate the standardised normal values $z_i = F(w_i)$.

Step 3: Calculate the Cramer-von Mises statistic

$$W^2 = \sum_{i=1}^{n} [z_i - (2i-1)/2n]^2 + 1/(12n)$$ and the modified statistic

$$W^* = W^2(1 + 1/2n).$$

Step 4: Check for significance of W^* in the following table:

Critical values of W^* for the α-level of significance

α	Critical value
.150	.091
.100	.104
.050	.126
.025	.148
.001	.178

If W^* is greater than the critical value at level α then we conclude, at the α-level of significance, that the data do not follow a normal distribution.

Fig. 5.1: Saturation So2 (%) vs Po2 (mm Hg)

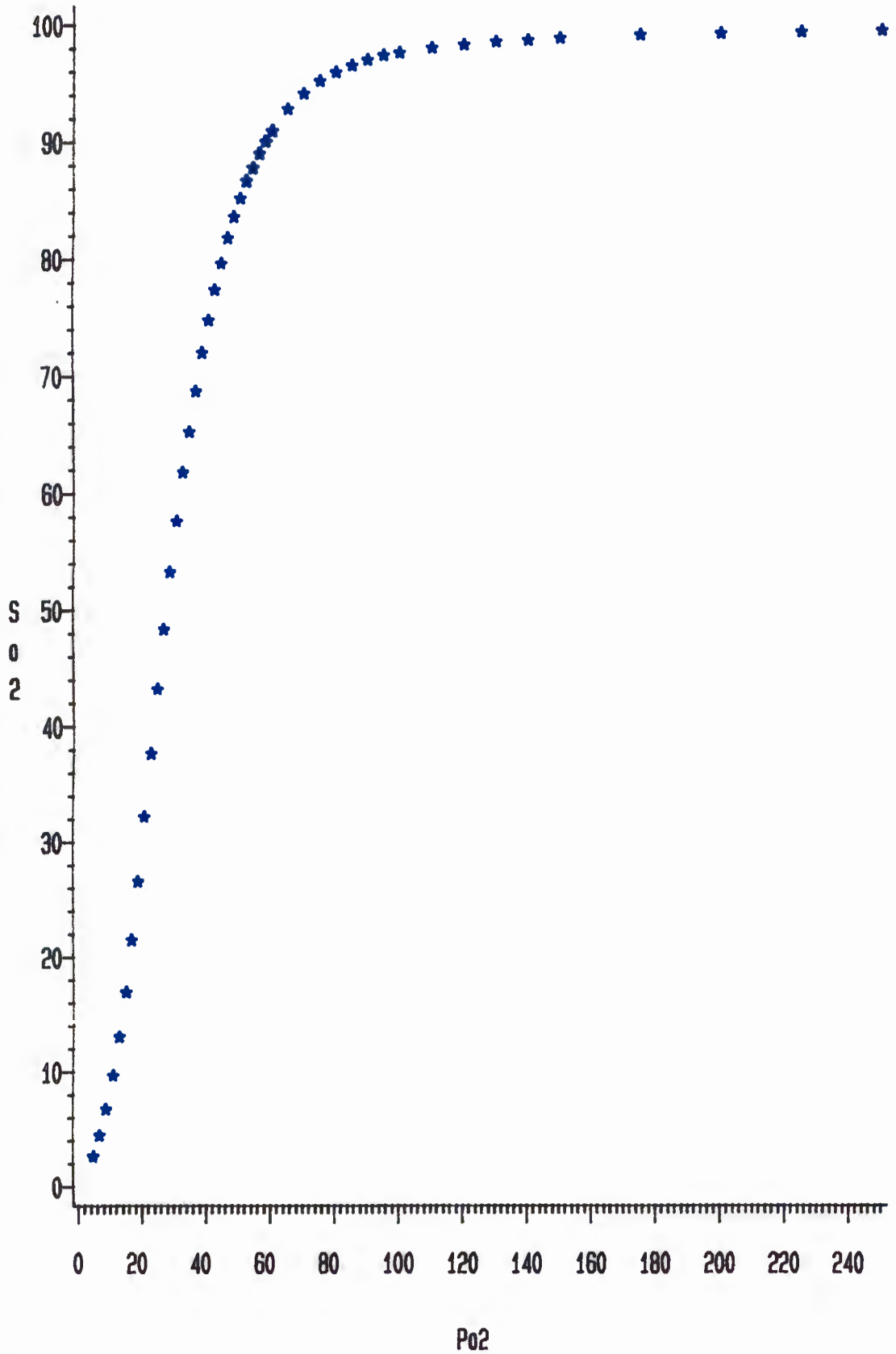


Fig. 5.2 : Metronidazole concentration vs time

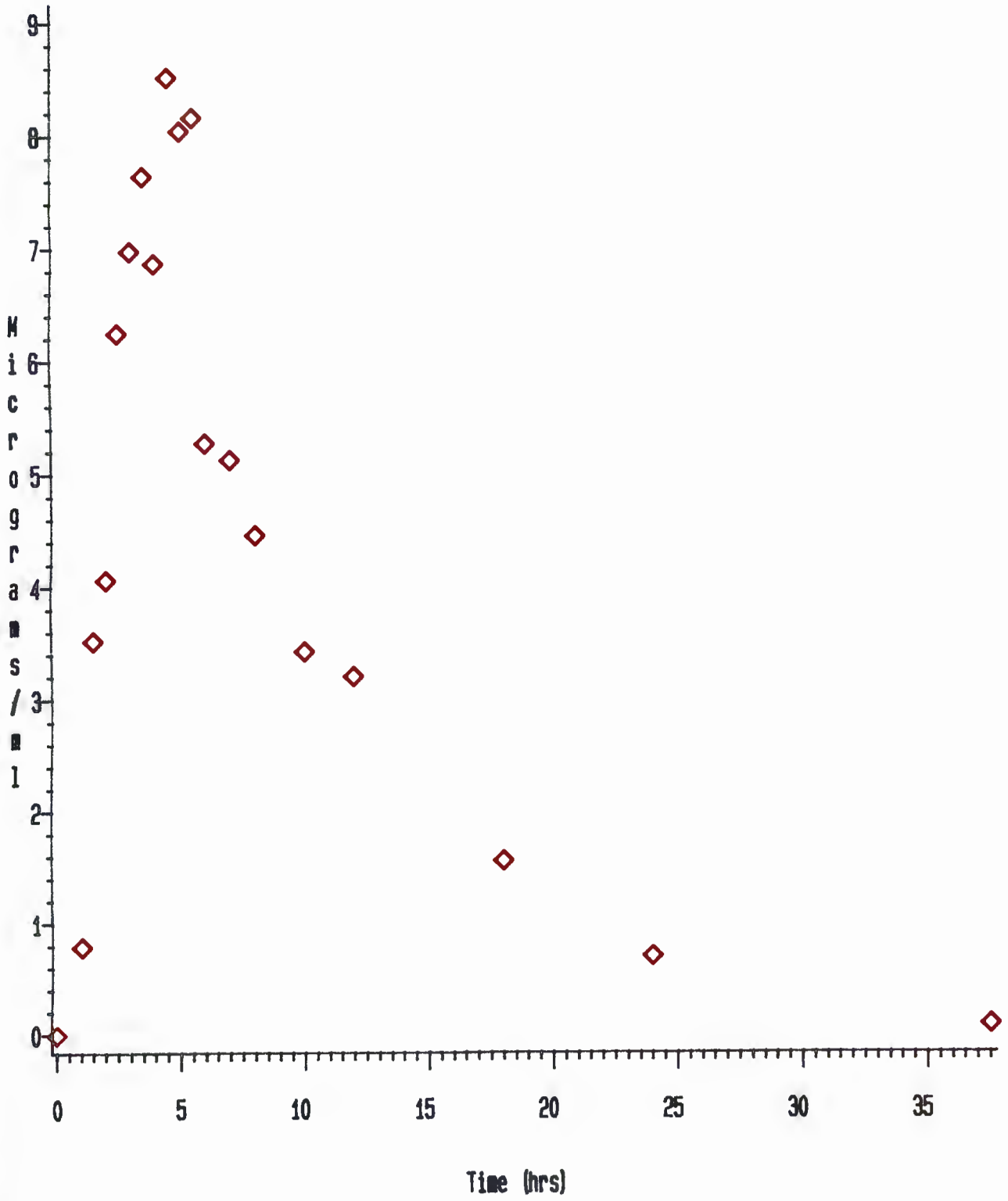


Fig. 5.3 : Pattern 8 data

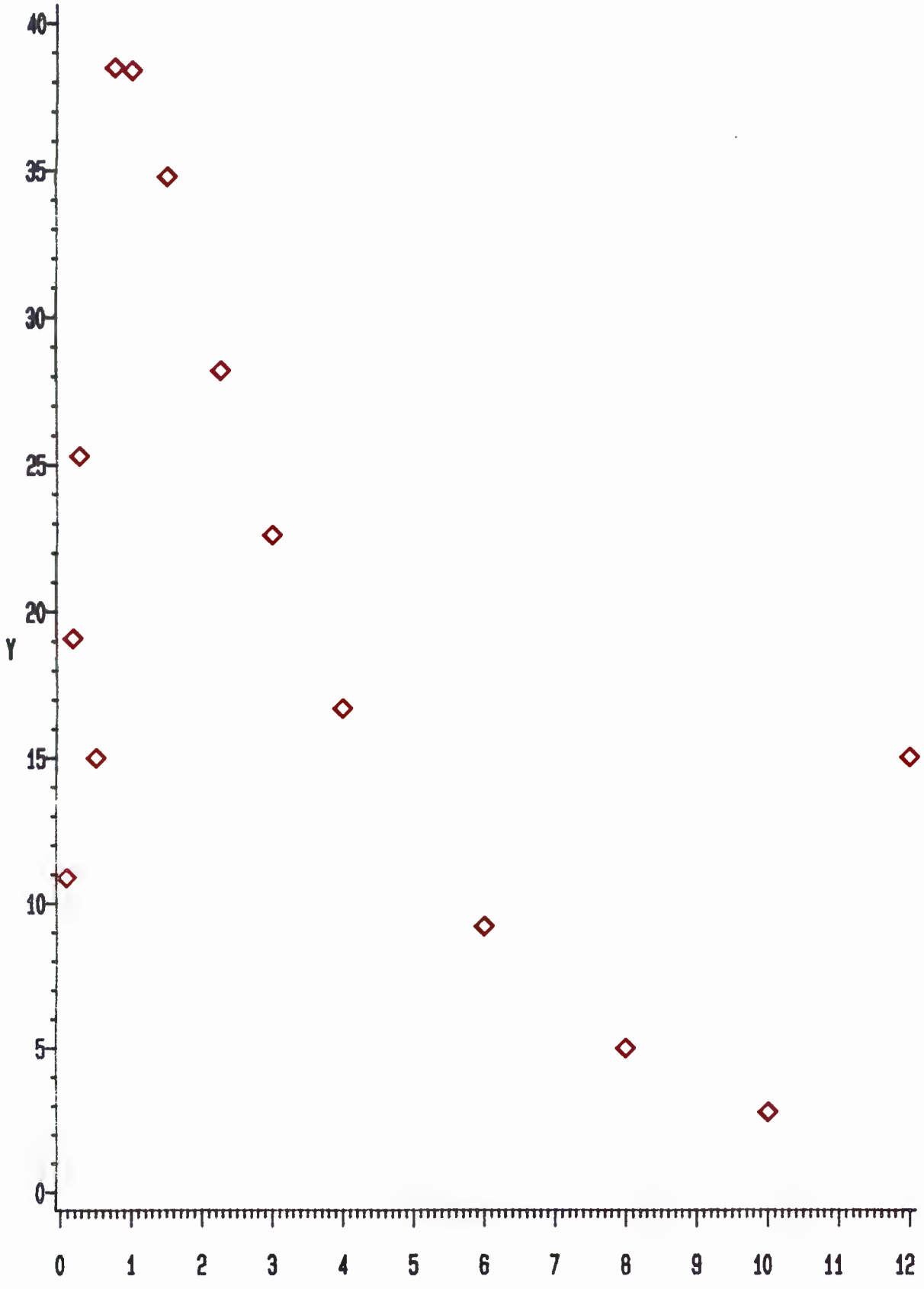


Fig. 5.4: *Pattern* 8 *data with fitted values for* p=1.15

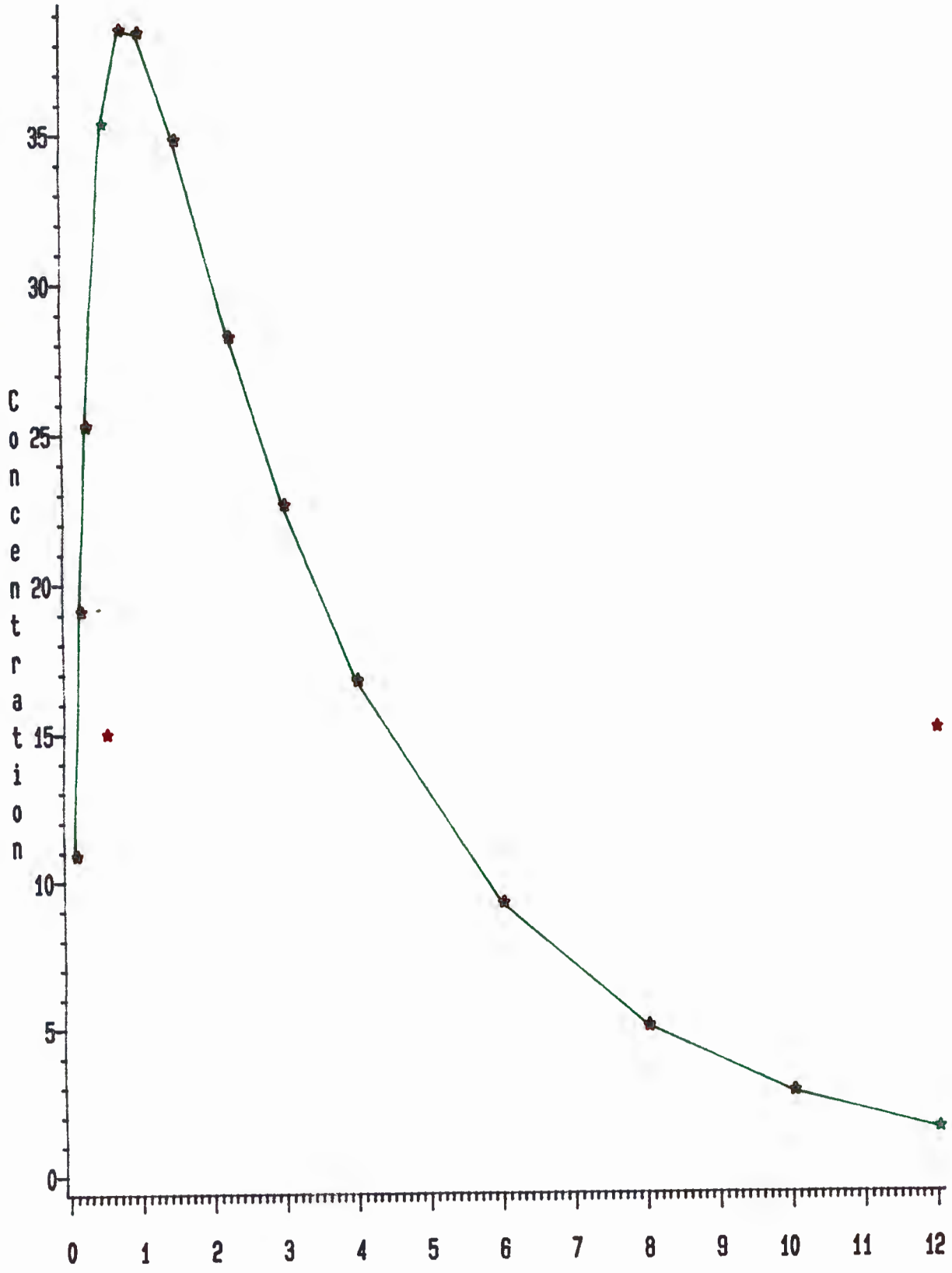
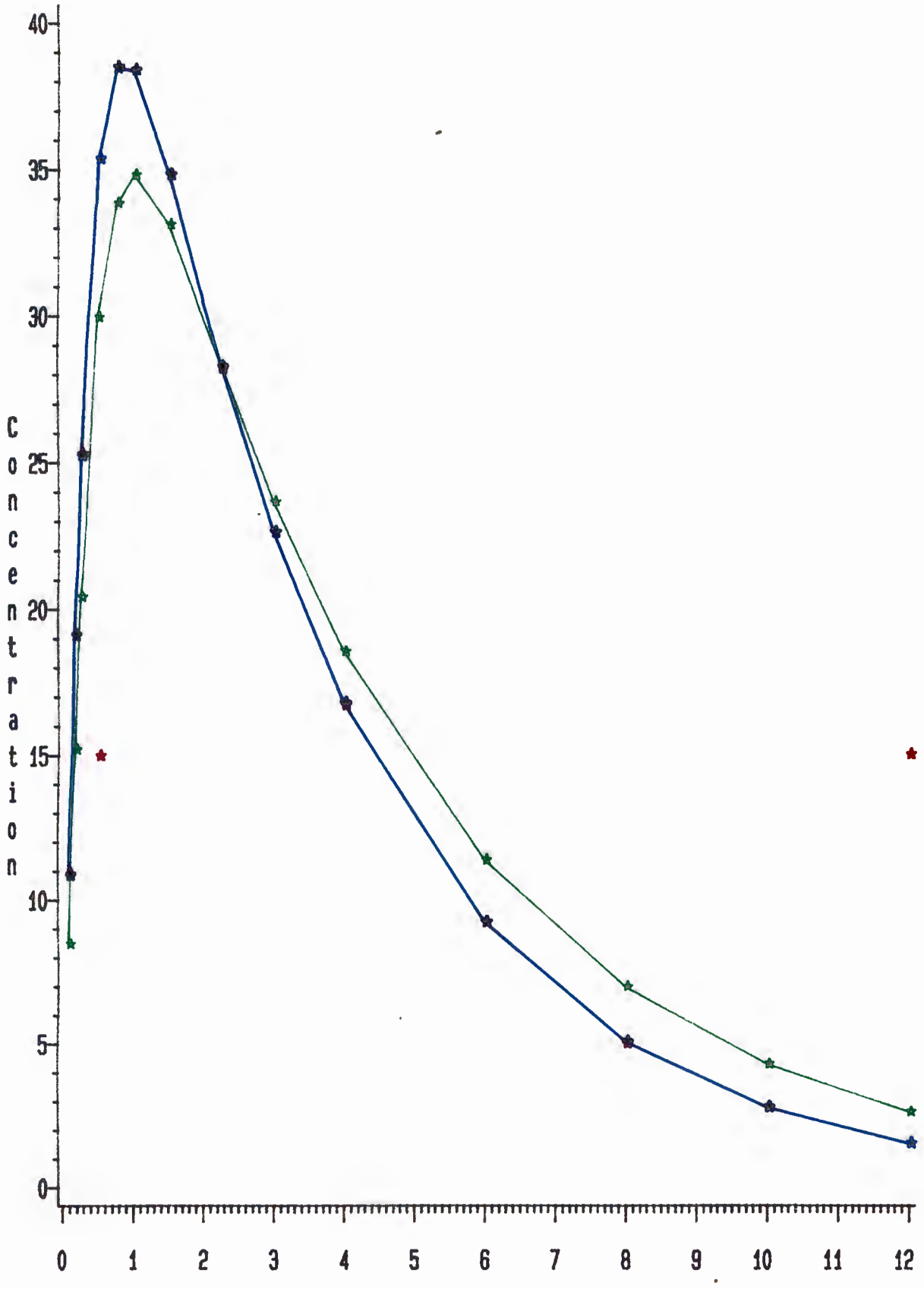


Fig. 5.5: *Pattern* 8 *data with fitted values for* p=2 *and* p=1.15



CONCLUSION

In this thesis two numerically efficient algorithms were derived for solving nonlinear L_p-norm estimation problems. The first- and second- order partial derivatives of the objective function $S_p(\underline{\theta})$ were expressed in terms of a new and compact matrix notation. At the same time it was shown that the nonlinear least squares problem is imbedded in the general L_p-norm estimation problem. Graphical displays were also provided to illustrate numerical examples.

A simulation study was carried out to establish the best p-value to use for a given additive symmetric error distibution. Similar results to that of linear L_p-norm estimation were found. Established empirical relationships between the value p and the kurtosis of symmetrical error distributions were confirmed for the nonlinear case. In addition recently derived theoretical sampling (asymptotic) properties of linear L_p-norm estimators were examined and related properties for nonlinear L_p-norm estimators were postulated. It was shown that these theoretical proposals are in complete agreement with the simulation results.

An adaptive procedure was derived to calculate systematically the optimal p-value for a given error distribution. This procedure was used in a simulation study to derive the empirical distribution of the optimal p-values. It was shown that the optimal p-values are asymptotically normally distributed. The value of this adaptive procedure in identifying outlying observations was also illustrated.

This research has demonstrated the value of L_p-norm estimation as an alternative to nonlinear least squares. In the case of non-normally distributed observations (especially long-tailed distributions where outliers will be prevalent) its use represents a substantial improvement on classical least squares. This result is, of course, also true in linear estimation. Moreover in the event where least squares is appropriate, the alternative estimation procedure is equally efficient.

SUGGESTIONS FOR FUTURE RESEARCH

Various aspects of nonlinear L_p-norm estimation still remain unresolved. Future research may be undertaken on a number of fronts.

With regard to numerical analysis the following problems need to be studied:

a) The derivation of optimality conditions for general nonlinear L_p-norm estimation problems where the objective function is non-differentiable. The implementation of these conditions in usable and practical algorithms. The algorithms of Chapters 2 and 3 in essence solve the usual first-order necessary conditions for optimality.

b) A convergence proof of the algorithms in Chapters 2 and 3 has not been worked out. The convergence proof of Gill and Murray (1978) for nonlinear least squares may be used. The convergence rate of the algorithm should also be derived. It is expected to be linear at worst and superlinear at best.

c) The author of this thesis found a direct correspondence between the value of p and the degree of homogeneity γ of the Jacobson-Oksman (1972) homogeneous algorithm when it was applied to nonlinear L_p-norm estimation problems. An investigation of this phenomenon would not only contribute to the theory of estimation but also to the theory of homogeneous algorithms.

On the statistical side the following problems are still unresolved:

d) Asymptotic properties of the nonlinear L_p-norm estimators have yet to be determined. An intuitive proposal was made in Chapter 4 based on related properties derived by Nyquist (1980) for linear L_p-norm estimators.

e) Confidence regions for the parameters $\underline{\theta}$ are still nonexistent. This is likely to prove to be a difficult problem in view of the difficulty experienced in nonlinear least squares (see for example Clarke (1980) and Jennrich (1969)).

f) The simulation results suggest that the optimal p-values are asymptotically normal. An analytical proof is lacking.

g) Further research relating skewed distributions to the optimal p-value needs to be done to establish the appropriate relationship. The work by Ekblom (1974), who suggested that p-values less than 1 should be used in linear estimation, will be relevant. We feel that formulae (5.1) and (5.2) for p should be extended to take into account the skewness (in addition to the kurtosis) of the resulting residual distributions.

h) In the nonlinear estimation problem we have only considered additive errors (errors in the dependent variable). The effect of errors in the independent variables (multiplicative) would also be a fruitful area of research. This would be especially useful in the area of biological modelling. The idea by Watson (1982) of using orthogonal deviations

could be used to derive an algorithm for the nonlinear estimation problem in which both the dependent and independent variables are subject to error. The effect of heteroscedastic errors should also be examined. The work by Nyquist op. cit. will again be relevant.

BIBLIOGRAPHY

Anderson, D.H. and Osborne, M.R. (1976). Discrete, linear approximation problems in Polyhedral norms. Numer. Math. 26, pp 179-189.

Anderson, D.H. and Osborne, M.R. (1977a). Discrete, nonlinear approximation problems in polyhedral norms. Numer. Math. 28, pp 143-156.

Anderson, D.H. and Osborne, M.R. (1977b). Discrete, nonlinear approximation problems in polyhedral norms. A Levenberg-like algorithm. Numer. Math. 28, pp 157-170.

Andrews, D.F. (1974). A robust method for multiple linear regression. Technometrics 16, pp 523-531.

Armijo, L. (1966). Minimization of functions having Lipschitz continuous partial derivatives. Pacific J. Math. 16, pp 1-3.

Armstrong, R.D., Frome, E.L. and Kung, D.S. (1979). A revised simplex algorithm for the absolute deviation curve fitting problem. Commun. Statist.-Simula. Computa., B8(2), pp 175-190.

Armstrong, R.D. and Kung, D.S. (1979). Min-max estimates for a linear multiple regression problem. Appl.Stat. 28, pp 93-100.

Armstrong, R.D. and Kung, D.S. (1980). A dual method for discrete Chebychev curve fitting. Math. Programming 19, pp 186-199.

Arthanari, T.S. and Dodge, Y. (1981). Mathematical programming in Statistics, Wiley, New York.

Avriel, M. (1976). Nonlinear programming, analysis and methods. Prentice-Hall. New Jersey.

Bard, Y. (1970). Comparison of gradient methods for the solution of nonlinear parameter estimation problems. SIAM J. Numer. Anal. 7, pp 157-186.

Barr, G.D.I. (1980). A contribution to adaptive robust estimation. PhD Thesis. University of Cape Town.

Barr, G.D.I., Affleck-Graves, J.F., Money, A.H. and Hart, M.L. (1980). Performance of a generalized algorithm for L_p-norm regression estimates. Technical Report No. ALS-4. Dept. Math. Stats. University of Cape Town.

Barrodale, I. and Phillips, C. (1975). Solution of an overdetermined system of linear equations in the Chebychev norm. ACM Trans. on Math. Software 1, pp 264-270.

Barrodale, I. and Roberts, F.D.K. (1970). Applications of Mathematical Programming to Lp approximation. In: Nonlinear programming (eds.) J.B. Rosen, O.L. Mangasarian & K. Ritter, Academic Press. New York.

Barrodale, I., Roberts, F.D.K. and Hunt, C.R. (1970). Computing best L_p approximations by functions nonlinear in one parameter. Computer J. 13, 4 pp 382-386.

Barrodale, I. and Roberts, F.D.K. (1973). An improved algorithm for discrete L_1 linear approximation. SIAM J. Numer. Anal. 10, pp 839-848.

Bassett, G. and Koenker, R. (1978). Asymptotic theory of least absolute error regression. JASA 73, pp 618-620.

Beale, E.M.L. (1958). On an iterative method for finding a local minimum of a function of more than one variable. Technical report no. 25. Statistical research group. Princeton University.

Ben-Tal, A and Zowe, J. (1982). Necessary and sufficient optimality conditions for a class of nonsmooth minimization problems. Math. Programming 24, pp 70-88.

Bertsekas, D.P. (1976). Multiplier methods: A survey. Automatica 12, pp 133-145.

Betts, J.T. (1976). Solving the nonlinear least squares problem: Application of a general method. JOTA 18, pp 469-483.

Boscovich, R.J. (1757). De litteraria expeditione per pontificiam ditionem et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa, Bononiensi Scientiarum et Artum Instituto atque Academia Commentarii 4, pp 353-396.

Breytenbach, N.J. (1978). Homogeneous programs in optimization. PhD Thesis. University of South Africa, Pretoria.

Browne, M.B. and Du Toit, S.H.C. (1977). Suroutine MLPY.

Buys, J.D. (1972). Dual algorithms for constrained optimization problems. PhD Thesis. Leiden University.

Cauchy, Augustin-Louis (1824). Sur le système des valeurs qu'il faut attribeur a deux éléments détermines par un grand nombre d'observations, pour que la plus grande de toutes les erreurs, abstraction faite du signe, devienne un minimum. Bulletin de la Societe Philomatique (Paris), pp 92-99.

Charalambous, C. (1979). On conditions for optimality of the nonlinear L_1 problem. Math. Programming 17, pp 123-135.

Chebychev, P.L. (1854). Théorie des mécanismes connus sous le nom de parallélogrammes. (Reprinted in Ouvres de P.L. Chebychev (ed. A. Markoff & N. Sonin) Vol I, (1899) pp 111-143. Impremerie de l'Académie Impériale des Sciences St. Petersbourg.

Clarke, G.P.Y. (1980). Moments of the least squares estimators in a non-linear regression model. J.R. Statist. Soc. B, pp 227-237.

Clarke, G.P.Y. (1982). The analysis of nonlinear regression models. PhD Thesis. University of London.

Cotes, Roger (1722). Aestimatio errorum im mixta mathesi, per variationes partium trianguli plani et sphaerici. Opera Miscellanea, (Appended to Cotes, Harmonia Mensur, Cantabrigiae. pp 1-22.

Dennis, J.E. (1977). Non-linear least squares and equations. In: The state of the art in numerical analysis (ed.) D. Jacobs, Academic Press. London.

Dennis, J.E. and Welsch, R.E. (1978). Techniques for nonlinear least squares and robust regression. Commun. Statist. Computa, B7(4), pp 345-359.

Descloux, J. (1963). Approximations in L^p and Chebyshev approximations. J. Soc. Ind. Appl. Math 11, pp 1017-1026.

Doolittle, H.M. (1884). The rejection of doubtful observations (abstract). Bulletin of the Philosophical Society of Washington (Mathematical Section) 6, pp 153-156.

Du Toit, S.H.C. and Gonin, R. (1982). An extension of the Richards family of growth curves. Lecture/paper PR-77, Human Sciences Research Council, Pretoria.

Edgeworth, F.Y. (1883). The law of error. Phil. Mag. 16, pp 360-375.

Edgeworth, F.Y. (1887a). On discordant observations. Phil. Mag. 23, pp 364-375.

Edgeworth, F.Y. (1887b). On observations relating to several quantities. Hermathena, 6 (13), pp 279-285.

Edgeworth, F.Y. (1888). On a new method of reducing observations relating to several quantities. Phil. Mag. 25, pp 184-191.

Ekblom, H. (1973). Calculation of linear best L_p-approximations. BIT 13, pp 292-300.

Ekblom, H. (1974). L_p-methods for robust regression. BIT 14, pp 22-32.

El-Attar, R.A. Vidyasagar, M. and Dutta, S.R.K. (1979). An algorithm for L_1-norm minimization with application to nonlinear L_1-approximation . SIAM J Numer. Anal. 16, pp 70-86.

Euler, Leonhard (1749). Pièce qui a Remporté le Prix de l'Academie Royale des Sciences en 1748 , sur les Inegalites du Movement de Saturn et de Jupiter, Paris.

Fiacco, A.V. and McCormick, G.P. (1968). Nonlinear programming: Sequential unconstrained minimization techniques, Wiley, New York.

Fischer, J. (1981). An algorithm for discrete linear L_p approximation. Numer. Math. 38, pp 129-139.

Fletcher, R. (1970). A new approach to variable metric algorithms. Comput. J. 13, pp 317-322.

Fletcher, R. (1980). Practical methods of optimization, Volume 1 unconstrained optimization, Wiley, Chichester.

Fletcher, R. (1981). Practical methods of optimization, Volume 2 constrained optimization, Wiley, Chichester.

Fletcher, R., Grant, J.A. and Hebden, M.D. (1971). The calculation of linear best L_p approximations. Comput. J. 14, pp 276-279.

Fletcher, R., Grant, J.A. and Hebden, M.D. (1974a). Linear minimax approximation as the limit of best L_p approximation. SIAM J. Numer. Anal. 11, pp 123-136.

Fletcher, R., Grant, J.A. and Hebden, M.D. (1974b). The continuity and differentiabilty of the parameters of best linear L_p approximations. J. Approx. Theory 10, pp 69-73.

Fletcher, R. and Watson, G.A. (1980). First and second order conditions for a class of nondifferentiable optimization problems. Math. Programming 18, pp 291-307.

Forsythe, A.B. (1972). Robust estimation of straight line regression coefficients by minimizing pth power deviations. Technometrics 14, pp 159-166.

Frome, E.L. and Yakatan, G.J. (1980). Statistical estimation of the pharmokinetic parameters in the one compartment open model. Commun. Statist.-Simula. Computa., B9(3), pp 201-222.

Galilei, Galileo (1632). (English translation, Dialogue concerning the two chief world systems, Ptolemaic and Copernican, by Stillman Drake (foreword Albert Einstein) UCLA Press 1953).

Gallant, A.R. (1975). Nonlinear regression. The American Statistician, 29, pp 73-81.

Gauss, C.F. (1806). II Comet von Jahr 1805. Monatliche Correspondenz zur Beförderung der Erd- und Himmelskunde 14, pp 181-186.

Gauss, C.F. (1820). Theoria combinationis observationum erroribus minimus obnoxiae, pars prior, printed in Werke, (Göttingen 1880) IV pp 6-7.

Gentle, J.E. (1977). Least absolute values estimation: An introduction. Commun. Statist.-Simula. Computa., B6(4), pp 313-328.

Gentle, J.E. and Hanson, T.A. (1977). Variable selection under L_1. Proc. Statist. Comp. Section A.S.A., pp 228-230.

Gill, P.E. and Murray, W. (1974). Newton-type methods for unconstrained and linearly constrained optimization. Math. Programming 7, pp 311-350.

Gill, P.E. and Murray, W. (1978). Algorithms for the solution of the nonlinear least-squares problem. Siam J. Numer. Anal. 15, pp 977-992.

Goldfeldt, S.M. and Quandt, R.E. (1972). Nonlinear methods in Econometrics, North-Holland, Amsterdam.

Hand, M. and Sposito, V.A. (1980). Using the least squares estimator in Chebychev estimation. Commun. Statist.-Simula. Computa., B9 pp 43-49.

Harter, H.L. (1974a). The method of least squares and some alternatives I. Int. Stat. Rev. 42, pp 147-174.

Harter, H.L. (1974b). The method of least squares and some alternatives II. Int. Stat. Rev. 42, pp 235-264.

Harter, H.L. (1975a). The method of least squares and some alternatives III. Int. Stat. Rev. 43, pp 1-44.

Harter, H.L. (1975b). The method of least squares and some alternatives IV. Int. Stat. Rev. 43, pp 125-190.

Harter, H.L. (1975c). The method of least squares and some alternatives V. Int. Stat. Rev. 43, pp 269-272.

Harter, H.L. (1975d). The method of least squares and some alternatives - Addendum to part IV. Int. Stat. Rev. 43, pp 273-278.

Harter, H.L. (1976). The method of least squares and some alternatives VI. Int. Stat. Rev. 44, pp 113-159.

Harter, H.L. (1977). The nonuniqueness of absolute values regression. Commun. Statist.-Simula. Computa. A6, pp 829-838.

Hartley, H.O. (1964). Exact confidence regions for the parameters in non-linear regression laws. Biometrika 51, pp 347-353.

Harvey, A.C. (1977). A comparison of preliminary estimators for robust regression. JASA 72, pp 910-913.

Harvey, A.C. (1978). On the unbiasedness of robust regression estimators. Commun. Statist.-Theor. Meth., A7, pp 779-783.

Henrici, P. (1964). Elements of numerical analysis, Wiley, New York.

Hiebert, K.L. (1979). A comparison of nonlinear least squares software. Sandia laboratories report SAND 79-0483, Albuquerque, New Mexico.

Himmelblau, D.M. (1972) Applied nonlinear programming, McGraw-Hill, New York.

Himmelblau, D.M. and Lindsay, J.V. (1980). An evaluation of substitute methods for derivatives in unconstrained optimization. Operations Res. 28, pp 668-686.

Hinnich, M.J. and Talwar, P.P. (1975). A simple method for robust regression. JASA 70, pp 113-119.

Hiriart-Urruty, J.B. (1978). On optimality conditions in nondifferentiable programming. Math. Programming 14, pp 73-86.

Jacobson, D.H. and Oksman W. (1972). An algorithm that minimizes homogeneous functions of N variables in N+2 iterations and rapidly minimizes general functions. J. Math. Anal. Appl. 38, pp 535-552.

Jennrich, R.I. and Sampson, P.F. (1968). Application of stepwise regression to non-linear estimation. Technometrics 10, pp 63-71.

Jennrich , R.I. (1969). Asymptotic properties of non-linear least squares estimators. Ann. Math. Statist. 40, pp 633-643.

Johnson, N.L. and Kotz, S. (1970). Distributions in statistics: Continuous univariate distributions-1. Houghton Mifflin, New York.

Johnson, N.L. and Kotz, S. (1970). Distributions in statistics: Continuous univariate distributions-2. Houghton Mifflin, New York.

Johnson, P. and Milliken, G.A. (1983). A simple procedure for testing linear hypotheses about the parameters of a nonlinear model using weigted least squares. Commun. Statist.-Simula. Computa., 12(2), pp 135-145.

Jones, A.P. (1970). SPIRAL - A new algorithm for nonlinear parameter estimation using least squares. Comput. J. 13, pp 301-308.

Juritz, J.M., Gonin, R. and Bridle, T. (1983). The comparison of immediate release Theophylline drug preparations in comparative bioavailability studies. Institute for Biostatistics working paper, Medical Research Council, Cape Town.

Kahng, S.W. (1972). Best L_p approximation. Math. Comput. 26, pp 505-508.

Kendall, M.G. and Stuart, A. (1963). The advanced theory of statistics, Volume 1, Distribution theory. Charles Griffin, London.

Kennedy, W.J. and Gentle, J.E. (1978). Comparisons of algorithms for minimum L_p norm linear regression., Proceedings of Computer Science and Statistics: 10th Annual Symposium on the Interface. (Ed.) D. Hogben, pp 373-378.

Kennedy, W.J. and Gentle, J.E. (1980). Statistical Computing. Marcel Dekker. New York.

Khorasani, F. and Milliken, G.A. (1982). Simultaneous confidence bands for nonlinear models. Commun. Statist.-Theor. Meth., A11(11), pp 1241-1253.

Kuhn, H.W. and Tucker, A.W. (1951) Nonlinear programming. In: Proceedings of the second Berkeley symposium on mathematical statistics and probability, (ed.) J. Neyman, UCLA Press, Berkeley, California, pp 481-492.

Laplace, Pierre S. (1786). Exposition du systéme du monde, Paris.

Laplace, Pierre S. (1818). Deuxiéme supplément a la théorie analytique des probabilités, Paris, Courcier. Reprinted (1887) in Ouvres Complétes de Laplace 7, pp 531-580 Paris, Gouthier-Villars.

Lawson, C.L. and Hanson R.J. (1974). Solving least squares problems. Prentice-Hall, New Jersey.

Legendre, A.M. (1805). Nouvelles methodes pour la détermination des orbites des cometes. Courcier Paris. (Appendice sur la methode des moindres quarrés, pp 72-80).

Luenberger, D.G. (1973). Introduction to linear and nonlinear programming, Addison-Wesley, Reading, Mass.

Malinvaud, E. (1970). The consistency of nonlinear regressions. Ann. Math. Statist. 41, pp 956-969.

Mangasarian, O.L. (1969). Nonlinear programming. McGraw-Hill, New York.

Mayer, Johann T. (1750). Abhandlung über die Umwälzung des Mondes um seine Axe. Kosmographische Nahrichten und Sammlungen fur 1748, 1, pp 52-183.

McCormick, G.F. and Sposito, V.A. (1976). Using the L_2-estimator in L_1-estimation. SIAM J. Numer. Anal. 13, pp 337-343.

Merle, G. and Späth, H. (1973). Computational experiences with discrete L_p-approximation. Computing 12, pp 315-321.

Metzler, C.M. (1974). Bioavailability - A problem in equivalence. Biometrics 30, pp 309-317.

Milliken, G.A. and DeBruin , R.L. (1978). A procedure to test hypotheses for nonlinear models. Commun. Statist.-Theor. Meth., A7(1), pp 65-79.

Money, A.H., Affleck-Graves, J.F., Hart, M.L. and Barr, G.D.I. (1982). The linear regression model: L_p norm estimation and the choice of p. Commun. Statist.-Simula. Computa. 11, pp 89-109.

Moré, J.J. , Garbow, B. and Hillstrom, K.E. (1980). User guide for MINPACK-1. Argonne National Laboratory, Argonne, Illinois.

Narula, S.C. and Wellington, J.F. (1982). The minimum sum of absolute errors regression : A state of the art survey. Int. Stat. Rev. 50, pp 317-326.

Nazareth, L. (1980). Some recent approaches to solving large residual nonlinear least squares problems. Siam Rev. 22, pp 1-11.

Nyquist, H. (1980). Recent studies on L_p-norm estimation. PhD Thesis. University of Umeå, Sweden.

Nyquist, H. (1982). The optimal L_p norm estimator in linear regression models. Submitted for publication to Comm. Stat.

Oberhofer, W. (1982). The consistency of nonlinear regression minimizing the L_1-norm . Ann. Stat. 10, pp 316-319.

Odeh, R.E. and Evans, J.O. (1974). Algorithm AS 70: Percentage points of the normal distribution Appl. Stat. 23, pp 96-97.

Osborne, M.R. and Watson, G.A. (1969). An algorithm for minimax approximation in the nonlinear case. Comput. J. 12, pp 63-68.

Osborne, M.R. and Watson, G.A. (1971). On an algorithm for discrete nonlinear L_1 approximation. Comput. J. 14, pp 184-188.

Osborne, M.R. (1972). Some aspects of non-linear least squares calculations. In: Numerical methods for optimization, (ed.) F. Lootsma, Academic Press pp 171-189.

Powell, M.J.D. (1971). On the convergence of the variable metric algorithm. J. Inst. Math. Appl. 7, pp 21-36.

Rey, W. (1975). On least p-th power methods in multiple regressions and location estimations. BIT 15, pp 174-185.

Rice, J.R. (1964). The approximation of functions, vol 1: Linear Theory. Addison-Wesley.

Rodda, B.E., Sampson, C.B. and Smith, D.W. (1975). The one-compartment open model: Some statistical aspects of parameter estimation. Appl. Stat. 24, pp 309-318.

Roodman, G. (1974). A procedure for optimal stepwise MSAE regression analysis. Operations Res. 22, pp 393-399.

Rosenbrock, H.H. (1960). An automatic method for finding the greatest or the least value of a function. Comp. J. 3. pp 175-184.

Sadovski, A.N. (1974). L_1-norm fit of a straight line. Appl. Stat. 23, pp 244-248.

Schlossmacher, E.J. (1973). An iterative technique for absolute deviation curve fitting. JASA 68, pp 857-859.

Severinghaus, J.W. (1979). Simple, accurate equations for human blood O_2 dissociation computations. J. Appl. Physiol. 46, pp 599-602.

Shrager, R.I. and Hill, E. (1980). Nonlinear curve-fitting in the L_1 and L_∞ norms. Math. Comput., 34, pp 529-541.

Sielken, R.L. and Hartley, H.O. (1973). Two linear programming algorithms for unbiased estimation of linear models. JASA 68, pp 639-641.

Sklar, M.G. and Armstrong, R.D. (1982). Least absolute value and Chebychev estimation utilizing least squares results. Math. Programming 24, pp 346-352.

Späth, H. (1982). On discrete linear orthogonal L_p approximation. Z. Angew. Math. Mech. 62, pp 354-355.

Spiegel, M.R. (1968). Mathematical handbook of formulas and tables. Schaum's Outline Series, McGraw-Hill, New York.

Sposito, V.A. Kennedy, W.J. and Gentle, J.E. (1977). L_p norm fit of a straight line. Appl. Statist. 26, pp 114-118.

Sposito, V.A. (1982). On unbiased L_p regression estimators. JASA 77, pp 652-654.

Sposito, V.A. Hand, M.L. and Skarpness, B. (1983). On the efficiency of using the sample kurtosis in selecting optimal L_p estimators. Commun. Statist.-Simula. Computa. 12, pp 265-272.

Stephens, M.A. (1974). EDF Statistics for goodness of fit and some comparisons. JASA 69, pp 730-737.

Wagner H.M. (1962). Non-linear regression with minimal assumptions. JASA 57, pp 572-578.

Watson G.A. (1979). Dual methods for nonlinear best approximation problems. J. Approx. Theory 26, pp 142-150.

Watson G.A. (1982). Numerical methods for linear orthogonal L_p Approximation. IMA J. Numer. Anal. 2, pp 275-287.

Wolfe J.M. (1979). On the convergence of an algorithm for discrete L_p approximation. Numer. Math. 32, pp 439-459.

Zangwill, W.I. (1969). Nonlinear Programming: A unified approach. Prentice-Hall.

APPENDIX E:

The FORTRAN programme

```
C
C   THIS PROGRAM SOLVES LARGE RESIDUAL NONLINEAR LP-NORM PROBLEMS.
C   IN: GONIN, R. (1983). A CONTRIBUTION TO SOLVING NONLINEAR ESTIMATION
C            PROBLEMS. THESIS SUBMITTED FOR THE DEGREE OF PHD.
C            UNIVERSITY OF CAPE TOWN.
C
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION THETA(30),S(30),G(30),D1(30),D2(30),F2(100),DELF(500)
     *,BHESS(1000),YHAT(100)
      COMMON/NVAL/NITER,MAXITR,NEVAL,IFREQ
      COMMON/PRECIS/GTOL,FTOL,XTOL,TOLER
      COMMON/NEGCUR/EJ,PHIMIN,IMIN,NEGOPT
      COMMON/DATA/Y(100),T(100),P,N,M
      COMMON/RESUL/YRES(100)
C
C            MAXITR = MAXIMUM NO OF ITERATIONS REQUIRED
C            NEVAL= NUMBER OF STEPS IN LINE SEARCH
C            N  = NO OF PARAMETERS
C            M  = NO OF OBSERVATIONS
C            P VALUE TO BE USED IN LP-NORM ESTIMATION
C            THETA = PARAMETER VECTOR
C            RHO IS THE LINE SEARCH PARAMETER IN STEP 9
C            XTOL,GTOL AND FTOL CONVERGENCE TOLERANCES USED IN NONLPS
C            YRES(.) = VECTOR OF RESIDUALS : Y(FIT)-Y(OBS)
C            IFREQ: A STEEPEST DESCENT STEPS EVERY IFREQ-TH ITERATION
C
C            BETTS EXAMPLE 8.8 JENNRICH & SAMPSON (1968) EXAMPLE
C
      M=10
      N=2
      DO 1 I=1,M
      T(I)=I
1     Y(I)=2.0 + 2.0*T(I)
      MAXITR=50
      P=2.0D0
      RHO=0.4
      TOLER=0.1
      IFREQ=65
C
C                  STEP 0 : INITIALISATION
C
999   NEVAL=0
      NITER=0
      THETA(1)=0.3
      THETA(2)=0.4
      WRITE(6,10) P
10    FORMAT('  P = ',F6.2)
      GTOL=1.D-9
      FTOL=GTOL
      XTOL=GTOL
      WRITE(6,20)
20    FORMAT(25X,' SP(THETA)',T50,'   THETA1',T65,'  THETA2')
      CALL NONLPS(RHO,THETA,F2,DELF,G,BHESS,FOBJ,D1,D2,S)
      WRITE(6,30)
```

```
30    FORMAT(T10,'T(I)        Y(I)      YHAT(I)')
      DO 2 I=1,M
      CALL FUNC(I,THETA,YH)
      YHAT(I)=YH
      WRITE(6,3)T(I),Y(I),YHAT(I)
3     FORMAT(1X,3F12.2)
2     CONTINUE
500   STOP
      END
C
C     SUBROUTINE FUNC(I,THETA,YHAT)
C
      SUBROUTINE FUNC(I,THETA,YHAT)
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION THETA(1)
      COMMON/DATA/Y(100),T(100),P,N,M
          FA=DEXP(THETA(1)*T(I))
          FB=DEXP(THETA(2)*T(I))
          YHAT=FA+FB
      RETURN
      END
C
C     SUBROUTINE NONLPS(RHO,THETA,F2,DELF,G,BHESS,FOBJ,D1,D2,S)
C
C
C     SUBROUTINE NONLPS SOLVES THE NONLINEAR LP-NORM PROBLEM BY MEANS
C     OF THE ALGORITHM BY GONIN (1983) (CHAPTER3 OF THE THESIS).
C

      SUBROUTINE NONLPS(RHO,THETA,F2,DELF,G,BHESS,FOBJ,D1,D2,S)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION THETA(1),DELF(1),G(1),BHESS(1),D1(1),D2(1),S(1),F2(1)
     *,GG(30)
      COMMON/NVAL/NITER,MAXITR,NEVAL,IFREQ
      COMMON/PRECIS/GTOL,FTOL,XTOL,TOLER
      COMMON/DATA/Y(100),T(100),P,N,M
      COMMON/RESUL/YRES(100)
      COMMON/NEGCUR/EJ,PHIMIN,IMIN,NEGOPT
      COMMON/SPACE/WSPA(2000),WSPB(2000),WSPC(2000),WSPD(2000)
      COMMON/OPTION/IOPDER
C
C     IOPDER = 1 OPTION CHOOSES NUMERICAL DERIVATIVES
C     IOPDER = 2 OPTION CHOOSES ANALYTICAL DERIVATIVES
C
      IOPDER=1
      NPROD=N*M
      ZERO = 0D0
      PHALF=0.5D0*P-1D0
      PONE=P-1D0
C
C     NITER =      ITERATION NUMBER
C
      IND=-4
      INDIC = -N
      N4=MIN0(4,N)
      NITER=0
      CALL FGRAD(N,M,THETA,DELF,G,FOBJ)
      DF=FOBJ
```

```
      NEVAL=0
    1 NITER=NITER+1
      WRITE(6,100) NITER,NEVAL,FOBJ,(THETA(I),I=1,N4)
  100 FORMAT(4X,I5,1X,'*',4X,I5,2X,'*',2X,G12.6,1X,'*',5X,4F14.6)
      IF(N.GT.4) WRITE(6,150) (THETA(I),I=5,N)
  150 FORMAT(1H0,32X,'*',5X,4F14.6)
      IK=MOD(NITER,IFREQ)
      IF(IK.EQ.0) GO TO 500
      NEWTON=1
      ITEST = 0
C
C                  STEP 1
C
C            COMPUTE  GRADIENT g(.) AND  ||g(.)||
C            DNORM:=  ||JP(TRANSP)JP||  WHERE JP IS THE P-JACOBIAN
C            GG(I) IS GENERALISED GRADIENT DUE TO DENNIS (1977)
C
      DNORM=ZERO
      GNORM  = ZERO
      DO 3 I = 1,N
      DELFII=ZERO
      DO 2 J = 1,N
      IJ = N*(I-1)+J
      DELFII = DELFII + DELF(IJ)**2
    2 CONTINUE
      DNORM=DNORM+DELFII*DELFII
      GG(I)=G(I)/DSQRT(FOBJ*DELFII)
      GNORM= GNORM +G(I)**2
    3 CONTINUE
      DNORM=DSQRT(DNORM)
      GNORM=DSQRT(GNORM)
C
C     f2 = ABS(YRES)**(P/2-1)*YRES     p-RESIDUAL mx1 vector
C
      DO 4 I=1,NPROD
    4 WSPA(I)=DELF(I)
      DO 5  I=1,M
      IF(DABS(YRES(I)).NE.0D0) GO TO 777
      ABYRES=ZERO
      GO TO 5
  777 ABYRES=DABS(YRES(I))**PHALF
    5 F2(I)=YRES(I)*ABYRES
C
C                  STEP 2
C
C     SINGULAR-VALUE DECOMPOSITION OF Jp=DELF (m x n) IS COMPUTED
C
      CALL SVDRS(DELF,M,M,N,F2,M,1,S)
C
C     VECTOR f2 = U(TRANSP)f2 = U(TRANSP).(yhat-y)P IN (2.6)
C     VECTOR DELF = MATRIX V (n x n)
C
      IF(NEWTON.EQ.0)GOTO 6
C
C                    STEP 5
C
C     BHESS:= HESSIAN OF SECOND DERIVATIVE TERMS (NUMERICAL)
C     BNORM=  ||BHESS(I,I)||
```

```
C
      CALL SBHESS(N,M,THETA,BHESS)
      BNORM=ZERO
      DO 13 I=1,N
   13 BNORM=BNORM+BHESS(I*(I+1)/2)**2
      BNORM=DSQRT(BNORM)
C
C     RATIO = || BHESS(I,I) || /(p-1)|| Jp(TRANSP)Jp) ||
C
      DNORM=PONE*DNORM
      RATIO=BNORM/DNORM
C
C                    STEP 3
C
      CALL GRADE(N,IR,RATIO,S)
      IF(IR.EQ.N) NEWTON=0
    6 SIGMA=0D0
      DO 21 I=1,IR
      IF(S(I).LT.1D0)GO TO 21
      SIGMA=SIGMA+1D0/S(I)**2
   21 CONTINUE
      SIGMA=DSQRT(SIGMA)
      SIGMA=DMIN1(SIGMA,0.0001D0)
    7 ITEST=ITEST + 1
      IF(IR.GT.0) GO TO 9
C
C     IR = 0 IMPLIES A FULL NEWTON STEP THUS DIRECTION d1 = 0
C
      DO 8 I=1,N
    8 D1(I) = ZERO
      GO TO 14
    9 DO 10 I=1,IR
   10 WSPB(I)=-F2(I)/S(I)
C
C     EXPRESSION (D1)**(-1)*F1 IS STORED IN WSPB
C
      L=0
      DO 11 I=1,N
      DO 11 J=1,N
        L=L+1
        JI= M*(I-1)+J
        DELF(L)=DELF(JI)
   11 CONTINUE
C
C     SORTING V INTO FIRST n x n ELEMENTS OF DELF
C
C                STEP 4
C
      CALL MLPY(DELF,N,N,IR,WSPB,IR,IR,1,D1,N)
      DO 70 I=1,N
   70 D1(I)=D1(I)/PONE
C
C     CALCULATION OF d1 = -V1.(D1)**(-1).f1/(p-1)
C     IF GRADE r = n TAKE FULL GAUSS-NEWTON STEP
C
   12 IF(IR.NE.N) GO TO 14
      DO 17 I=1,N
   17 D2(I)=ZERO
```

```
      GO TO 25
   14 L=0
      NMR = N - IR
      IR1 = IR + 1
      DO 15 I =IR1,N
      DO 15 J = 1,N
      L=L+1
      IJ = N*(I-1) +  J
      WSPD(L) = DELF(IJ)
C
C     V2 : (n x nmr) STORED COLUMNWISE IN FIRST (n-r)
C     COLUMNS OF WSPD
C
   15 CONTINUE
      CALL MLPY(WSPD,-N,N,NMR,BHESS,0,N,N,WSPC,NMR)
C
C     MATRIX PRODUCT  V2(TRANSP.) X BHESS=Y ( STORED IN WSPC )
C
      CALL MLPY(WSPC,NMR,NMR,N,WSPD, N, N,NMR,WSPA,0)
C
C     MATRIX PRODUCT  V2(TRANSP.) X BHESS X V2=Q STORED IN WSPA
C     IN H.S. FORM
C
      CALL MLPY(WSPC,NMR,NMR,N,D1, N, N,1,WSPB,NMR)
C
C     MATRIX PRODUCT V2(TRANSP).BHESS.d1 STORED IN WSPB
C
      DO 27 I = 1,NMR
      II =I*(I+1)/2
      WSPA(II) = WSPA(II)+ PONE*S(IR+I)**2
      WSPB(I) = -F2(IR+I)* S(IR+I) - WSPB(I)
   27 CONTINUE
C
C     (P-1).D2**2 + V2(TRANSP).BHESS.V2 STORED IN WSPA (ORDER = NMR)
C     VECTOR( D2.F2)- V2(TRANSP).BHESS.VECT(d1)  STORED IN WSPB
C
C                  STEP 5
C
C       CHOLESKI LDL(TRANSP) FACTORIZATION OF WSPA
C
      CALL LDLT(NMR,WSPA)
C     WRITE(6,991)IMIN,PHIMIN,NEGOPT,GNORM,EJ
  991 FORMAT(1X,'PHIMIN(',I2,')=',F12.4,' NEGOPT=',I2,' GNORM=',F12.4
     *,' EJ=',F12.4)
C
C     STEP 6(C) OPTIMAL SOLUTION
C
      IF(NEGOPT.EQ.0 .AND. GNORM.LE.GTOL) RETURN
      IF(NEGOPT.EQ.-1 .AND. GNORM.LE.GTOL) GO TO 36
      GO TO 19
C
C     STEP 6 (C)
C
C     IF NEGOPT=-1 CALCULATE THE DIRECTION OF NEGATIVE CURVATURE
C     ONLY THE SYTEM (L-TRANSP)*Y = C(SUB(S)) IS SOLVED IN THE SECOND
C     PART OF LDLSOL, WHERE C(SUB(S)) = VECTOR WITH A 1 IN POSITION S
C     AND ZERO'S ELSEWHERE
C
```

```
   36 DO 18 I=1,NMR
   18 WSPC(I)=0.0D0
      WSPC(IMIN)=1.0D0
      CALL LDLSOL(NMR,WSPA,WSPB,WSPC)
C
C     CALCULATION OF PROJECTED GRADIENT GSUBK*DSUBK =GS
C
      UG=0.
      DO 16 I=1,N
16    UG=UG+G(I)*WSPB(I)
      IF( GNORM.GT.0D0) GO TO 22
      GO TO 52
C
C     DSIGN(X,X)=SIGN(X)*ABS(X)
C
22    DO 51 I=1,N
        WSPB(I)=-DSIGN(UG,UG)*WSPB(I)/DABS(UG)
51    CONTINUE
      GO TO 21
C
C     STEP 6 (B)
C
  19  CALL LDLSOL(NMR,WSPA,WSPB,WSPC)
52    CONTINUE
C
C                 STEP 7
C
C     CALCULATION OF DIRECTION d2 = V2.Z2
C
      CALL MLPY(WSPD,N, N,NMR,WSPB,NMR,NMR,1,D2, N)
C
C     VECTOR dk = d1 + d2 AND IS PLACED IN d1
C
25    GS=ZERO
      D1NORM=ZERO
      DO 20 I=1,N
        D1(I)= D1(I)+D2(I)
        D1NORM= D1NORM +D1(I)**2
        GS = GS +G(I)*D1(I)
   20 CONTINUE
      D1NORM =DSQRT(D1NORM)
C
C     STEP 8
C
      SG= -GS/(GNORM*D1NORM)
C     WRITE(6,23)GS,SG,SIGMA,ITEST
23    FORMAT(' PROJEC. GRAD. ',G14.4/' NEG. STAND. PROJ. GRAD.',G14.4,
     *' SIGMA ',G14.4/ ' ITEST',I4)
      IF( SG.LT.SIGMA .AND. ITEST.EQ.1) GO TO 40
      GO TO 505
C
C             STEP 9
C
C     THE LINE SEARCH STEP - LINE IS A CUBIC INTERPOLATION METHOD
C     SEE FLETCHER (1970).
C
500   GS=0D0
      DO 502 I=1,N
```

```
      D1(I)=-G(I)
  502 GS=GS+G(I)*D1(I)
C     WRITE(6,23)GS
  505 CONTINUE
      FOLD=FOBJ
      CALL LINE (DF,GTOL,RHO,GS,THETA,DELF,G,FOBJ,D1)
      IF(GS.GE.0D0) RETURN
C
C     CALCULATE DECREASE IN OBJECTIVE FUNCTION
C
      DF=FOBJ
C
C                STEP 10
C
C      THIS STEP CHECKS TO SEE IF THE CONVERGENCE CRITERIA HAVE BEEN MET
C      INDIC = 0   INDICATES |G(I)| < GTOL FOR ALL I=1,...N
C      THE ALGORITHM HAS CONVERGED TO A (LOCAL) MINIMUM
C      IND =       CONVERGENCE PARAMETER INDICATING THE NO OF CONSECUTIVE
C                  ITERATIONS IN WHICH NO CHANGE WAS FOUND (STEP 1)
C
      DO 75 I=1,N
   75 IF (DABS(G(I)).LT.GTOL) INDIC = INDIC + 1
      IF (INDIC.EQ.0) RETURN
      IF(DABS((FOLD-FOBJ)/FOLD).GT.GTOL) IND=-3
      IF(DABS((FOLD-FOBJ)/FOLD).LE.GTOL) IND=IND+1
      IF(IND.EQ.0) RETURN
      IF(NITER.GT.MAXITR) RETURN
      GO TO 1
C
C     STEP 8 OF ALGORITHM
C
   40 IR=0
C
C     RETURN FROM STEP 8 TO STEP 6
C     ALGORITHM CONTINUES FROM STEP 6 (WITH d1 = ZERO AND IR = 0 )
C     I.E. WITH A FULL NEWTON STEP
C
      GO TO 7
      END
C
C     SUBROUTINE FGRAD(N,M,X,DELF,G,SSQ)
C     CALCULATES FIRST ORDER DERIVATIVES
C
      SUBROUTINE FGRAD(N,M,X,DELF,G,SSQ)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION X(1),DELF(1),G(1),BHESS(1)
      COMMON/NVAL/NITER,MAXITR,NEVAL
      COMMON/OPTION/IOPDER
      COMMON/RESUL/YRES(100)
      COMMON/DATA/Y(100),T(100),P
      COMMON/SPACE/WSPA(2000),WSPB(2000),WSPC(2000),WSPD(2000)
      COMMON/JAC/ DELFJ(500)
      ZERO = 0D0
      NEVAL=NEVAL +1
      PHALF=0.5D0*P-1D0
      PTWO=P-2D0
      SSQ  = ZERO
      DO 1 I=1,N
```

```
      G(I) = ZERO
    1 CONTINUE
    3 DO 10 K = 1, M
      CALL FUNC (K,X,FTK)
      YRES(K) = FTK-Y(K)
      BB=DABS(YRES(K))
      IF(BB.EQ.0D0) GO TO 910
      BBTWO=BB**PTWO
      BBHALF=BB**PHALF
      BBP=BB**P
      GO TO 900
  910 WRITE (6,999)
  999 FORMAT(' WARNING A RESIDUAL IS ZERO')
      BBHALF=0D0
      BBTWO=0D0
      BBP=0D0
  900 CONTINUE
      SSQ  = SSQ + BB**P
C
C     SSQ = OBJECTIVE FUNCTION FOBJ
C
      GO TO (4,5),IOPDER
C
C     NUMERICAL DERIVATIVES OF FTK) W.R.T. X(I),I=1,N STORED IN
C     WSPD (N X 1)
C
    4 CALL NDERIV(N,K,FTK,X,WSPD)
      GO TO 6
C
C     ANALYTICAL DERIVATIVES
C
    5 CALL ADERIV(N,K,FTK,X,WSPD)
C
C     DELF (M X N): IS MATRIX JP
C     G (N X 1) GRADIENT VECTOR OF SP SEE (2.1)
C     DELFJ (M X N): IS ORDINARY JACOBIAN MATRIX
C
    6 IJ=0
      DO 8 I=1,N
      KI = M*(I-1) + K
      DFKXI = WSPD(I)
C
C     FORMULA FOR G(I) AS IN (1.4)
C
      G(I) = G(I)+P*(BBTWO)*YRES(K)*DFKXI
      DELFJ(KI)=DFKXI
      DELF(KI) = DFKXI*BBHALF
    8 CONTINUE
   10 CONTINUE
      RETURN
      END
C
C     SUBROUTINE NDERIV(N,K,FT,X,DFKDX)
C
      SUBROUTINE NDERIV(N,K,FT,X,DFKDX)
      IMPLICIT REAL*8(A-H,P-Z)
      DIMENSION X(1),DFKDX(1)
      H1=1D-6
```

```
      DO 1  J =1,N
        X(J)=X(J)+H1
      CALL FUNC(K,X,FTD)
      DFKDX(J) = (FTD-FT)/H1
        X(J)=X(J)-H1
    1 CONTINUE
      RETURN
      END
C
C     SUBROUTINE SBHESS(N,M,X,BHESS)
C     CALCULATES THE HESSIAN P.B(THETA) = BHESS NUMERICALLY.
C
      SUBROUTINE SBHESS(N,M,X,BHESS)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION X(1),DELF(1),BHESS(1)
      COMMON/RESUL/YRES(100)
      COMMON/DATA/Y(100),T(100),P
      COMMON/JAC/ DELFJ(500)
C
C     DELFJ (M X N): IS ORDINARY JACOBIAN MATRIX
C     BHESS (M*N*(N+1)/2 X 1) VECTOR :  HESSIAN IN HALF-SYMMETRIC FORM
C
C
      PTWO=P-2D0
      NELV=N*(N+1)/2
      DO 20 I=1,NELV
   20 BHESS(I)=0.0D0
      H=1D-6
      HSQ=H*H
      DO 10 K=1,M
      FTK=Y(K)+YRES(K)
      IF(DABS(YRES(K)).NE.0D0) GO TO 777
      ABYRES=ZERO
      GO TO 776
  777 ABYRES=DABS(YRES(K))**PTWO
  776 BBP=ABYRES*YRES(K)
      IJ=0
      DO 8 I = 1,N
      KI = M*(I-1) + K
      DFKXI=DELFJ(KI)
      X(I)=X(I)+H
      DO 7 J=1,I
      IJ = IJ + 1
      X(J)=X(J)+H
      CALL FUNC (K,X,FKIJ)
      KJ = M*(J-1) + K
      BHESS(IJ)=BHESS(IJ) + BBP*((FKIJ-FTK)/HSQ-(DFKXI+DELFJ(KJ))/H)
      X(J)=X(J)-H
    7 CONTINUE
      X(I)=X(I)-H
    8 CONTINUE
   10 CONTINUE
      RETURN
      END
```

```
C
C     SUBROUTINE ADERIV(N,K,FT,X,DFKDX)
C
      SUBROUTINE ADERIV(N,K,FT,X,DFKDX)
      IMPLICIT REAL*8(A-H,P-Z)
      DIMENSION X(1),DFKDX(1)
      COMMON/DATA/Y(100),T(100)
C
C     ANALYTICAL DERIVATIVES
C
      RETURN
      END
C
C     SUBROUTINE LINE(DF,GTOL,RHO,GS,THETA,DELF,G,FOBJ,D)
C
      SUBROUTINE LINE(DF,GTOL,RHO,GS,THETA,DELF,G,FOBJ,D)
      IMPLICIT REAL*8 (A-H,O-Z)
      DIMENSION THETA(1),DELF(1),G(1),D(1)
      COMMON/NVAL/NITER,MAXITR
      COMMON/DATA/Y(100),T(100),P,N,M
      COMMON/RESUL/YRES(100)
C
C     DF=LIKELY REDUCTION IN THE OBJECTIVE FUNCTION
C
      NIVAL=0
      IEXIT=2
      IF(GS.GE.0D0) GO TO 90
      GSO=GS
C
C     INITIAL STEPLENGTH SETTING OF GAM = ETA IN FLETCHER POWELL(1963)
C
      GAM1=-2D0*DF/GS
      GAM = DMIN1(1D0,GAM1)
30    CONTINUE
      IEXIT=3
      IF(NIVAL.GE.20) GO TO 80
      ICON=0
      IEXIT=1
113   DO  31 I=1,N
      Z =  GAM*D(I)
C
C     CALCULATION OF NEW THETA
C
      IF(DABS(Z).GE.GTOL) ICON=1
31    THETA(I)=THETA(I)+Z
111   NIVAL=NIVAL+1
      CALL FGRAD(N,M,THETA,DELF,G,FNEW)
C
C     CALCULATION OF PROJECTED GRADIENT GYS= GSUB(K+1)*DSUBK
C
      GYS =0D0
      DO 32 I=1,N
32    GYS=GYS +G(I)*D(I)
      IF(FNEW.GE.FOBJ) GO TO 40
      IF(DABS(GYS/GSO).LE.RHO)GO TO 50
C
C     THIS IS A PROJECTED GRADIENT TEST WITH RHO=.9 GIVING A WEAK
C     LINE SEARCH. RHO=.1 GIVES A FAIRLY ACCURATE LINE SEARCH.
```

```
C
      IF(GYS.GT.0D0)GO TO 40
      Z = 10D0
      IF(GS.LT.GYS)Z = GYS/(GS-GYS)
      Z = DMIN1(10D0,Z)
      GAM = GAM*Z
      FOBJ=FNEW
      GS=GYS
      GO TO 30
40    CONTINUE
C
C     NO DECREASE IN THE OBJECTIVE FUNCTION WAS FOUND RESET TO OLD THETA
C
      DO 41 I=1,N
41    THETA(I)=THETA(I)-GAM*D(I)
C
C     CHANGE IN THETA IS LT GTOL - EXIT
C
      IF(ICON.EQ.0) GO TO 78
C
C     CUBIC INTERPOLATION
C
      Z = 3D0*(FOBJ-FNEW)/GAM + GYS+GS
      ZZ= DSQRT(Z**2-GS*GYS)
      Z = 1D0 - (GYS+ZZ-Z)/(2D0*ZZ+GYS-GS)
      GAM = GAM*Z
      GO TO 30
50    CONTINUE
      FOBJ =FNEW
C
C     LINE SEARCH IS COMPLETE
C
78    WRITE(6,79)
79    FORMAT(' NORMAL EXIT')
      GO TO 100
80    WRITE(6,81)NIVAL
81    FORMAT(' MORE THAN ',I3,'FUNCTION EVALUATIONS IN THE LINE SEARCH')
      GO TO 100
90    WRITE(6,91)
91    FORMAT(' PROJECTED GRADIENT DENOTES AN INCREASE')
 100  RETURN
      END
C
C     SUBROUTINE GRADE(N,IR,RHO1,S) SELECTS THE GRADE (IR) OF THE
C     THE P-JACOBIAN JP
C
      SUBROUTINE GRADE(N,IR,RHO1,S)
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION S(1)
      COMMON/PRECIS/GTOL,FTOL,XTOL,TOLER
C
C     HEURISTIC RULE TO DEFINE TOLER
C
      IF(RHO1.LT.5D-2) TOLER=2D-1
      IF(RHO1.LT.5D-3) TOLER=1D-2
      IF(RHO1.LT.5D-4) TOLER=1D-3
      IF(RHO1.LT.5D-5) TOLER=1D-4
```

```
C
C     CALCULATION OF THE MINIMUM SINGULAR VALUE (S(.)) RHO1
C
      IR=0
      RHO2=1D0
      NM1=N-1
      DO 10 I=1,NM1
      SRMIN=S(I+1)/S(I)
      IF(SRMIN.LT.RHO2) IR=I
      RHO2=DMIN1(RHO2,SRMIN)
   10 CONTINUE
C
C     IF RHO2 > TOLER COMPUTE FULL GAUSS-NEWTON STEP
C
      IF(RHO2.GT.TOLER) IR=N
      WRITE(6,20)IR,TOLER,RHO2,(S(I),I=1,N)
   20 FORMAT(1H0,' GRADE OF JP =',I3,' TOLER ',G11.4,' RHO2 ',G11.4,
     */' SINGULAR VALUES',1X,5G11.4)
      RETURN
      END
C
C     INSERT:  SUBROUTINE SVDRS (A,MDA,MM,NN,B,MDB,NB,S)
C              SUBROUTINE QRBD (IPASS,Q,O,NN,V,MDV,NRV,C,MDC,NCC)
C              SUBROUTINE H12 (MODE,LPIVOT,L1,M,U,IUE,UP,C,ICE,ICV,NCV)
C              SUBROUTINE G1 (A,B,COS,SIN,SIG)
C              SUBROUTINE G2(COS,SIN,X,Y)
C              DOUBLE PRECISION  FUNCTION DIFF(X,Y)
C     C.L.LAWSON AND R.J.HANSON (1974). SOLVING LEAST SQUARES PROBLEMS.
C          PRENTICE-HALL.
C
      SUBROUTINE MLPY(A,IDA1,NROWA,NCOLA,B,IDB1,NROWB,NCOLB,C,IDC)
C     SUBROUTINE MLPY
C
C     PURPOSE
C        MATRIX MULTIPLICATION
C
C     PROGRAMMER
C        M.W. BROWNE N.I.P.R. NSM34 - MODIFIED BY WEBB - 7035/4366
C        MODIFIED BY S.H.C. DU TOIT: UNIV. OF PRET. 1977
C
C     USAGE
C        CALL MLPY(A,IDA1,NROWA,NCOLA,B,IDB1,NROWB,NCOLB,C,IDC)
C
C     DESCRIPTION OF PARAMETERS
C        A    - 1ST INPUT MATRIX
C        IDA1 - STORAGE MODE INDICATOR OF A
C        NROWA- NUMBER OF ROWS OF A
C        NCOLA- NUMBER OF COLUMNS OF A
C        B    - 2ND INPUT MATRIX
C        IDB1 - STORAGE MODE INDICATOR OF B
C        NROWB- NUMBER OF ROWS OF B
C        NCOLB- NUMBER OF COLUMNS OF B
C        C    - RESULTANT OUTPUT MATRIX
C        IDC  - STORAGE MODE INDICATOR OF C
C
C     REMARKS
C        X.Y=C
C        IF IDA1=NROWA   THEN X=A
```

```
C        IF IDA1=0        THEN X IS IN H.S. FORM
C        IF IDA1=-NROWA   THEN X=A(TR)
C
C        SIMILARLY FOR B
C
C        IF IDC=NROWA     THEN C IS GENERAL
C        IF IDC=0         THEN C IS GIVEN IN H.S. FORM
C
C     REMARKS
C        IDC=0 ONLY ALLOWABLE IF RESULTANT PRODUCT IS SYMM.
C
C     SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED
      IMPLICIT REAL*8(A-H,P-Z)
      DIMENSION A(1),B(1),C(1)
      IDC1=NROWA
      IF(IDA1.LE.0) IDC1=NCOLA
      IDA=IABS(IDA1)
      IDB=IABS(IDB1)
      IF(IDA1)3,3,4
    4 INCIA=1
      NRW=NROWA
      NAB=NCOLA
      INCJA=IDA
      GO TO 5
    3 INCJA=1
      NRW=NCOLA
      NAB=NROWA
      INCIA=IDA
    5 IF(IDB1)6,6,7
    7 INCIB=1
      NCL=NCOLB
      NAB2=NROWB
      INCJB=IDB
      GO TO 8
    6 INCJB=1
      NCL=NROWB
      NAB2=NCOLB
      INCIB=IDB
    8 IF(NAB-NAB2)10,9,10
   10 STOP
    9 IJJ=0
      KJB=-INCJB+1
      DO 23 J=1,NCL
      IJC=IJJ
      KJB=KJB+INCJB
      KIA=-INCIA+1
      IF(IDC.EQ.0) NRW=J
      DO 14 I=1,NRW
      LIDC=J*(J-1)/2+I
   22 KIA=KIA+INCIA
      IJC=IJC+1
      X=0.0D0
      IL=KIA
      LJ=KJB
      DO 15 L=1,NAB
      IF(IDA)17,18,17
   18 IL=ISYM(I,L)
   17 IF(IDB)19,20,19
```

```
   20 LJ=ISYM(L,J)
   19 CONTINUE
      X=X+A(IL)*B(LJ)
      IL=IL+INCJA
   15 LJ=LJ+INCIB
      C(IJC)=X
      IF(IDC.EQ.0) C(LIDC)=X
   14 CONTINUE
      IJJ=IJJ+IDC1
   23 CONTINUE
      RETURN
      END
C
      SUBROUTINE LDLSOL(NVAR,A,B,C)
C
C     PURPOSE:
C     SOLUTION OF THE EQUATIONS A.VECT(B-STAR) = VECT(B)
C     METHOD. A ON INPUT IS IN L.D.L-TRANSPOSE FORM
C     FIRST COMPUTE THE SOLUTION OF (L.D).VECT(C)= VECT(B) USING
C     FORWARD SUBSTITUTION
C     THEN COMPUTE THE SOLUTION OF  (L-TRANSP).VECT(B-STAR) = VECT(C)
C     USING BACKWARD SUBSTITUTION
C
C     PROGRAMMERS:
C     S.H.C.DU TOIT: UNIV.OF PRETORIA  AND
C     R. GONIN : INSTITUTE FOR BIOSTATISTICS MRC CAPE TOWN 1980
C
C     USAGE:
C     CALL LDLSOL(NVAR,A,B,C)
C
C     DESCRIPTION OF PARAMETERS:
C     NVAR.INPUT : DIMENSION OF A
C     A.INPUT :(NVARXNVAR) STORED IN H.S. FORM,FOR FURTHER DESCRIPTION
C     REFER TO SUBROUTINE LDLT
C     B.INPUT :(NVARX1) VECTOR OF KNOWN COEFFICIENTS
C     B.OUTPUT :(NVARX1) SOLUTION VECT B-STAR)
C     C.INPUT : (NVARX1) WORKSPACE
C
C
C
      IMPLICIT REAL*8(A-H,P-Z)
      DIMENSION A(1),B(1),C(1)
      COMMON/NEGCUR/EJ,PHIMIN,IMIN,NEGOPT
      ZERO=0D0
C
C   FORWARD SUBSTITUTION
C
      C(1) = B(1)/A(1)
      IF(NVAR.EQ.1)GO TO 3
      IF(NEGOPT.EQ.-1 )GO TO 3
C
C     DOES NOT CONVERGE
C     IF(NEGOPT.EQ.-1 .OR. PHIMIN.LT. 0D0)GO TO 3
C
      DO 2 I=2,NVAR
        IMIN1=I-1
        C(I)= B(I)
        DO 1 K=1,IMIN1
```

```
      IK=ISYM(I,K)
      KK=ISYM(K,K)
    1 C(I)=C(I)- A(IK)*A(KK)*C(K)
      II=ISYM(I,I)
      C(I)=C(I)/A(II)
    2 CONTINUE
C
C     BACKWARD SUBSTITUTION
C
    3 B(NVAR) =C(NVAR)
      IF(NVAR.EQ.1) RETURN
      DO 5 I1=2,NVAR
        I = NVAR -I1+1
        B(I)=C(I)
        IPLUS1= I+1
        DO 4 K=IPLUS1,NVAR
        IK=ISYM(I,K)
    4   B(I)=B(I)-A(IK)*B(K)
    5 CONTINUE
      RETURN
      END
         SUBROUTINE LDLT(N,A)
C
C
C     SUBROUTINE LDLT
C
C     PURPOSE:
C      FACTORIZE A SYMMETRIC MATRIX A INTO THE  FORM A=L*D*L(TRANSPOSE)
C      WHERE L IS A UNIT LOWER-TRIANGULAR MATRIX AND D A DIAGONAL MATRIX
C      A MODIFIED CHOLESKY FACTORIZATION IS PERFORMED AS DESCRIBED BY
C      GILL,P.E. AND MURRAY,W (1974).
C
C     PROGRAMMERS:
C       S.H.C. DU TOIT: UNIV.OF PRETORIA  AND
C       R. GONIN : INSTITUTE FOR BIOSTATISTICS MRC CAPE TOWN   1980
C
C       SEPTEMBER 1981:
C       SUBROUTINES TEST AND LDLT WERE MODIFIED BY R. GONIN TO AVOID
C       NUMERICAL DIFFICULTIES EXPERIENCED IN NEARLY SINGULAR MATRICES
C       E.G. THE HILBERT MATRIX.
C
C     USAGE:
C       CALL LDLT(N,A)
C
C     DESCRIPTION OF PARAMETERS:
C     A   -INPUT.MATRIX STORED IN HALF-SYMMETRIC(HS) FORM
C     N   -INPUT.ORDER OF MATRIX A
C     A   -OUTPUT. DIAGONAL ELEMENTS OF A WILL BE REPLACED BY DIAGONAL
C         -ELEMENTS OF D.
C         OFF-DIAGONAL ELEMENTS A(I,J) WILL BE REPLACED BY THE ELEMENTS
C         L(I,J) OF L WHERE  J=1,...,N ; I= J+1,...,N
C
C     REMARKS:
C      P.E GILL AND W.MURRAY (1974) NEWTON-TYPE METHODS FOR UNCON-
C      STRAINED AND LINEARLY CONSTRAINED OPTIMIZATION, MATH. PROGRAMMING
C      VOL. 7, PP. 311-350.
C
C     SUBROUTINES AND FUNCTION SUBPROGRAMS REQUIRED:
```

```
C       FUNCTION ISYM  LOCATES ELEMENT(I,J) OF MATRIX STORED IN HS
C       FORM
C       SUBROUTINE TEST MODIFIES ELEMENTS OF DIAGONAL MATRIX D WHEN
C       WHEN A IS INSUFFICIENTLY POSITIVE DEFINITE
C
C
C
      IMPLICIT REAL*8(A-H,O-Z)
      DIMENSION A(1)
      COMMON/NEGCUR/EJ,PHIMIN,IMIN,NEGOPT
      EPS=1.0D-30
      ZERO=0.0D0
      ANORM=ZERO
C
C     ANORM IS THE EUCLIDEAN NORM OF A:(NXN)
C     ZI IS THE LARGEST IN MODULUS OF THE OFF-DIAGONAL ELEMENTS
C     GAMMA IS THE LARGEST IN MODULUS OF THE DIAGONAL ELEMENTS OF A
C
      IMIN=1
      IJ=0
      ZI=ZERO
      GAMMA=ZERO
      DO 2 I=1,N
      DO 1 J=1,I
      IJ=IJ+1
      D=DABS(A(IJ))
      C=D*D
      ANORM=ANORM+2.D0*C
      IF(I.NE.J) ZI=DMAX1(D,ZI)
    1 CONTINUE
      ANORM=ANORM-C
      GAMMA=DMAX1(D,GAMMA)
    2 CONTINUE
      ANORM=DSQRT(ANORM)
      XN=N
      ZIN=ZI/XN
      BETA =DMAX1(GAMMA,ZIN,EPS)
C
C     BETA IS EQUAL TO MAX(GAMMA,ZI/N,EPS)
C
      ANORM=ANORM*EPS
      DELTA=DMAX1(ANORM,EPS)
C
C     DELTA IS EQUAL TO MAX(ANORM*EPS,EPS)
C
      IF(N.GT.1) GO TO 3
      CALL TEST(A(1),THETA,BETA,DELTA,1)
C
C     MODIFY A(1) IF A:(1X1) IS INSUFFICIENTLY POSITIVE DEFINITE
C
      GO TO 55
    3 THETA=EPS
      DO 4 I=2,N
      I1= ISYM(I,1)
      D=DABS(A(I1))
      THETA=DMAX1(D,THETA)
    4 CONTINUE
C
C     THETA IS THE LARGEST IN MODULUS OF C(I,1)=L(I,1)*D(I),I=2,...N
```

```
C
      CALL TEST(A(1),THETA,BETA,DELTA,2)
C
C     MODIFY A(1,1) IF NECESSARY
C
      PHIMIN=A(1)
      DO 5 I=2,N
      I1= ISYM(I,1)
      A(I1)=A(I1)/A(1)
  5   CONTINUE
C
C     ELEMENTS L(I,1),I=2,...,N  OVERWRITE THE CORRESPONDING A(I,1)
C
      DO 50 J=2,N
        THETA=EPS
        JP1=J+1
        JM1=J-1
        JJ =JP1*J/2
        DO 30 I=J,N
          C = ZERO
          DO 25 K =1,JM1
            KK=ISYM(K,K)
            IK=ISYM(I,K)
            JK=ISYM(J,K)
            C = C + A(JK)*A(IK)*A(KK)
   25     CONTINUE
          IJ = ISYM(I,J)
          A(IJ)= A(IJ)-C
          IF(I.NE.J) THETA=DMAX1(DABS(A(IJ)),THETA)
   30   CONTINUE
C     IMIN DETERMINES THE VARIABLE INDEX FOR NEGATIVE CURVATURE,
C          THEOREM 2.3.1 IN GILL & MURRAY (1974)
C          I.E. IMIN = SUBSCRIPT (S) AND PHIMIN =  PHI(S)
C
      IF(A(JJ).LT.PHIMIN) IMIN=J
      PHIMIN=DMIN1(A(JJ),PHIMIN)
      CALL TEST(A(JJ),THETA,BETA,DELTA,3)
      IF(J.EQ.N) GO TO 55
        DO 40 I= JP1,N
          IJ =ISYM(I,J)
          A(IJ)= A(IJ)/A(JJ)
   40   CONTINUE
   50 CONTINUE
   55 RETURN
      END
C
C     FUNCTION ISYM(I,J)
C
      FUNCTION ISYM(I,J)
      IF(J-I) 10,20,30
   10 ISYM=(I*(I-1))/2+J
      RETURN
   20 ISYM=(I*(I+1))/2
      RETURN
   30 ISYM=(J*(J-1))/2+I
      RETURN
       END
```

```
C
C     SUBROUTINE TEST(AJJ,THETA,BETA,DELTA,IOPT)
C
      SUBROUTINE TEST(AJJ,THETA,BETA,DELTA,IOPT)
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON/NEGCUR/EJ,PHIMIN,IMIN,NEGOPT
C
C     CALLED BY SUBROUTINE LDLT
C     MODIFY ELEMENTS OF DIAGONAL MATRIX D IN FACTORIZATION OF A=LXDXL(T
C     IF A IS INSUFFICIENTLY POSITIVE DEFINITE
C
      NEGOPT=1
      D=DABS(AJJ)
      AJJ3=DMAX1(DELTA,D)
      IF(IOPT.EQ.1) GO TO 4
      TB=THETA**2/BETA
      AJJ1=DMAX1(D,TB)
      AJJ2=DMAX1(TB,DELTA)
      IF(DELTA.GE.AJJ1)GO TO 1
      IF(D.GE.AJJ2) GO TO 2
      IF(TB.GE.AJJ3) GO TO 3
      GO TO 5
    1 EJ=DELTA-AJJ
       AJJ=DELTA
      GO TO 5
    2 EJ = D -AJJ
      AJJ=D
      NEGOPT=0
      RETURN
    3 EJ=TB-AJJ
       AJJ=TB
      GO TO 5
    4 EJ=AJJ3-AJJ
       AJJ=AJJ3
    5 IF(EJ.NE.0.0D0) NEGOPT=-1
      RETURN
      END
```