SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER

2. GOVT ACCESSION NO.

3. RECIPIENT'S CATALOG NUMBER

4. TITLE *(and Subtitle)*
Maximum Likelihood Parameter Estimation for Acoustic Transducer Calibration

5. TYPE OF REPORT & PERIOD COVERED
Final Report

6. PERFORMING ORG. REPORT NUMBER
CSP - NRL - 3

7. AUTHOR(s)
P. L. Ainsleigh,
J. D. George
V. K. Jain

8. CONTRACT OR GRANT NUMBER(s)
N00014-88-K-2012

9. PERFORMING ORGANIZATION NAME AND ADDRESS
Center for Communications and Signal Proc. (EE)
University of South Florida
Tampa, Florida 33620

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS
Naval Research Laboratory
Underwater Sound Reference Detachment
Orlando, Florida 32856

12. REPORT DATE
August 1988

13. NUMBER OF PAGES
94

14. MONITORING AGENCY NAME & ADDRESS*(if different from Controlling Office)*
Office of Naval Research
Georgia Institute of Technology
Atlanta, Georgia 30332

15. SECURITY CLASS. *(of this report)*
Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES
NRL Project Engineer: J.D. George

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*
Signal modeling
Maximum likelihood parameter estimation
Nonlinear least-squares
Variable projection functional
Orthogonal projection operator
Generalized inverse
Minimum norm solution
QR Factorization

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

This report examines maximum likelihood parameter estimation for signal models characteristic of the stepped sinusoid response of underwater acoustic transducers. The estimation problem is found to be particularly difficult when the stepped sinusoid excitation is at or near a resonance and the observation time is short compared to the model transient. A variable projection implementation of a maximum likelihood estimator is used to study parameter estimation when the excitation is near resonance.

MAXIMUM LIKELIHOOD PARAMETER ESTIMATION FOR

ACOUSTIC TRANSDUCER CALIBRATION

P.L. Ainsleigh, J.D. George, and V.K. Jain

CENTER FOR COMMUNICATIONS AND SIGNAL PROCESSING
UNIVERSITY OF SOUTH FLORIDA

Prepared for

Naval Research Laboratory
Underwater Sound reference Detachment

August 1988

ABSTRACT

This report examines maximum likelihood parameter estimation for signal models characteristic of the stepped sinusoid response of underwater acoustic transducers. This includes a review of the principal component linear prediction mthod, an exposition of the variable projection nonlinear least-squares method, a review of linear least-squares theory with special emphasis on generalized inverses and projection operators, and a discussion of iterative techniques for nonlinear least squares algorithms. The estimation problem is found to to be particularly difficult when the stepped sinusoid excitation is at or near a resonance and the observation time is short compared to the model transient. Characteristic least-squares error surfaces and contours are obtained for a two pole high-pass transducer model. A variable projection implementation of a maximum likelihood estimator is used to study parameter estimation when the excitation is near resonance.

KEYWORDS - signal modeling, maximum likelihood parameter estimation, nonlinear least-squares, variable projection functional, orthogonal projection operator, generalized inverse, minimum norm solution, QR Factorization, Gauss-Newton method.

TABLE OF CONTENTS

1. INTRODUCTION

Calibration of an underwater acoustic transducers entails, in part, estimating the steady-state response of the transducer to stepped sinusoid excitations. Of particular interest is the steady-state amplitude, which is used to characterize the transducer radiation pattern. An inherent problem arises, though, in making the response measurements, since reflections from measurement volume boundaries can corrupt the signal limiting the length of available data. The desired signal data is thus confined to a finite observation window occurring between the arrival of the wave at the hydrophone via the direct path from the projector and via the reflected paths (see Figure 1).

The problems caused by these reflections become critical at low frequencies (particularly for complex high power devices) because the decaying component of the transducer's transient response may not settle to a negligible level at any time during the available observation window, thus making direct measurement of the steady-state amplitude and phase impossible. This, therefore, necessitates estimating the steady-state information from the transient portion of the response.

In research previously carried out at the Naval Research Laboratory, Underwater Sound Reference Detachment, and at the Center for Communications and Signal Processing, University of South Florida [1], a signal parameter estimation algorithm utilizing principal component linear prediction as described by Kumerasean and Tufts [2], was used for estimating these parameters. This method was found to yield acceptable results (as assessed by the Cramer-Rao bounds for unbiased estimates) for excitation frequencies away from the resonant frequency of the transducer. For excitation frequencies near the resonance, however, the mean square errors of the estimates were unacceptably larger than the Cramer-Rao bound, thus suggesting the need for a maximum likelihood estimator. This report addresses this problem.

The maximum likelihood algorithm presented is based upon the variable projection nonlinear least-squares method described by Golub and Pereyra [3]. This method essentially reduces the number of parameters which must be optimized iteratively by defining a new cost functional, the variable projection functional, which is a function only of the observation vector and those parameters which occur nonlinearly in the signal model. For example, the two-pole high-pass model of a transducer used in our simulations results in a variable projection functional which can be mapped solely in terms of the damping factor and frequency of the decaying component of the transient. A contour plot of this error function for a particular signal parameter model is shown in Figure 2.

The goals of this document are to provide an exposition of the theory for obtaining maximum likelihood parameter estimators for stepped sine response signal models and to report results of simulation studies of the effectiveness of an ML algorithm. This will include (1) a reveiw of principal component linear prediction, (2) an exposition of the variable projection nonlinear least-squares method, (3) a review of linear least-squares theory with special emphasis on the construction of generalized inverses and projection operators, and (4) a discussion of iterative techniques necessary for the implementation of nonlinear least-squares algorithms.

In addition to presenting the results of the computer simulations using the algorithm outlined, contour and surface plots of the variable projection functional are provided. The parameter estimation problem is seen to be particularly difficult for stepped sinusoid excitations near a resonance. Maximum likelihood performance is achieved by the computer implementation of the variable projection algorithm described herein down to a threshold signal-to-noise ratio which depends on the quality factor (or Q) of the transducer model.

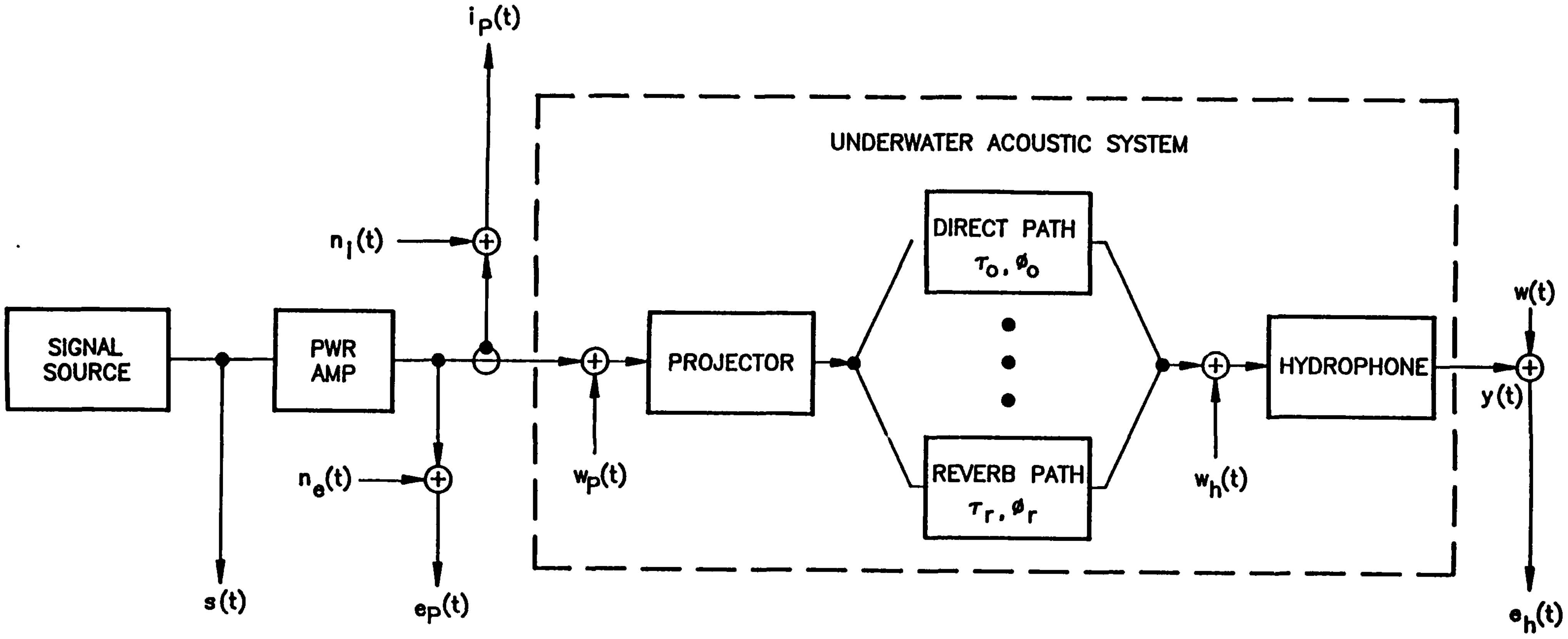


Figure 1: Transducer Calibration Environment

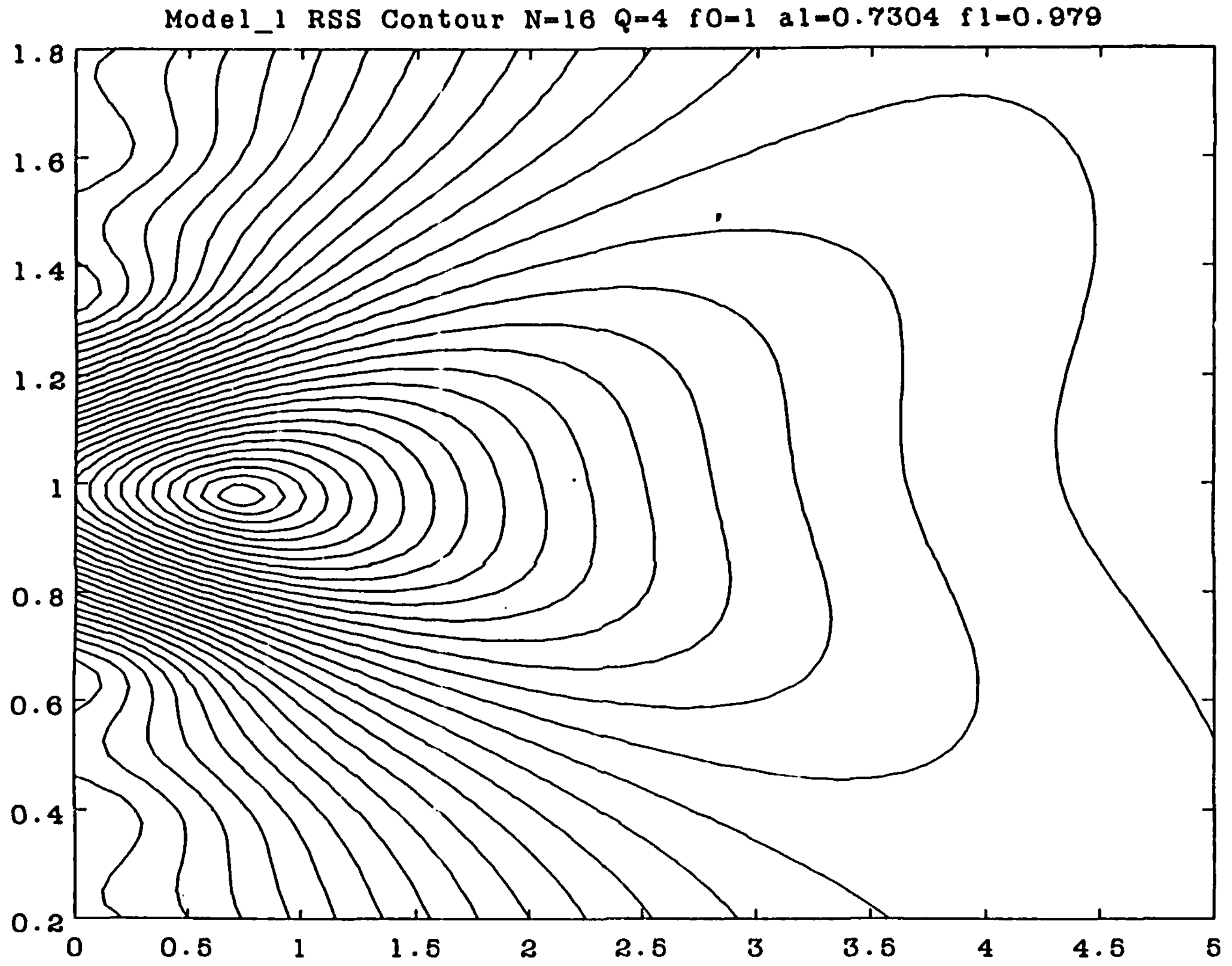


Figure 2: Variable Projection Least-Squares Functional Contour

2. BACKGROUND

This chapter provides a background of a parameter estimation problem that arises in acoustic transducer calibration. A model will be defined and the parameter set which uniquely defines the signal will be chosen. It will then be shown that, if the reflection free signal is corrupted with white gaussian noise, then nonlinear least-squares and the maximum likelihood estimators are the same. A review of the principal component linear prediction method will then be provided.

2.1 Acoustic Transducer Response to a Stepped Sinusoid Excitation

A suitable model for a transducer's response to a stepped sinusoid excitation is a steady-state sinusoid (of the same frequency as the excitation) and a sum of damped sinusoids (corresponding to the system poles) [4]. For real signals, this can be written as

$$x(t) = A_0 \cos(2\pi f_0 t+\phi_0) + \sum_{j=1}^{K} A_j \exp(-\alpha_j t) \cos(2\pi f_j t+\phi_j). \tag{1}$$

where K is the number of real system poles plus the number of complex conjugate system pole pairs. The parameters which uniquely define the signal are

$$\left\{ A_0, \phi_0, f_0, A_1, \phi_1, f_1, \alpha_1, \ldots, A_K, \phi_K, f_K, \alpha_K \right\}.$$

Each sinusoid in this model may be further decomposed into its Cartesian components so that the signal poles are the only parameters which enter into the model nonlinearly. Thus the signal model becomes

$$x(t) = A_{0_C}\cos(2\pi f_0 t) + A_{0_S}\sin(2\pi f_0 t)$$
$$+ \sum_{j=1}^{K} \exp(-\alpha_j t) \left\{ A_{j_C}\cos(2\pi f_j t) + A_{j_S}\sin(2\pi f_j t) \right\}, \tag{2}$$

from which we obtain the new parameter set

$$\left\{ A_{0_C},\ A_{0_C},\ f_0,\ A_{1_C},\ A_{1_S},\ f_1,\ \alpha_1,\ \ldots,\ A_{K_C},\ A_{K_S},\ f_K,\ \alpha_K \right\}. \tag{3}$$

Having estimated this parameter set, we may calculate the desired amplitudes and phases from

$$A_i = \left[A_{i_C}^{\ 2} + A_{i_S}^{\ 2} \right]^{1/2} \quad \text{and} \quad \phi_i = -\tan^{-1}\left[\frac{A_{i_S}}{A_{i_C}} \right]. \tag{4}$$

2.2 Nonlinear Least-Squares Maximum Likelihood Parameter Estimation

The parameter estimation approach to system identification has proven to be a powerful tool in problems where the system of interest is known, a priori, to have a particular model structure, M [5]. In this case, the model can be parameterized as $M(\underline{\theta})$ using the parameter vector $\underline{\theta} \subset D_M$, where D_M is an appropriate domain. Thus the family of models is

$$\{M(\underline{\theta}) \mid \underline{\theta} \subset D_M\} \tag{5}$$

and the search for the model which best decribes the system becomes a search for the best parameter vector, $\underline{\theta}$. In determining the best model parameters, we invoke some penalty function, or cost functional, which quantifies in some way the error of estimation.

Since the data used to estimate the model parameters are generally corrupted by additive noise, the observations are themselves random variables. The Maximum Likelihood Estimator (MLE) has been shown [6] to be the best possible estimator for the model parameters given the uniform prior probability distribution. In maximum likelihood estimation, we wish to find the parameter vector which maximizes the probability of the observed data. This is

$$\underline{\theta}_{ML} = \operatorname{Arg} \max_{\underline{\theta}} \; p(y_1, y_2, \ldots, y_N | \underline{\theta}) \tag{6}$$

where $p(y_1, y_2, \ldots, y_N | \underline{\theta})$ is the likelihood function.

Let us assume the the observations are, in fact, the sum of an ideal signal x_i and zero mean gaussian white noise, ϵ_i. Thus

$$y_n = x_n(\underline{\theta}) + \epsilon_n. \tag{7}$$

The estimation error is then

$$e_n = y_n - \hat{x}_n(\underline{\theta}) \tag{8}$$

and the likelihood function for estimating $\underline{\theta}$ becomes

$$p(y_1, y_2, \ldots, y_N | \underline{\theta}) = \frac{1}{\left[2\pi\sigma_\epsilon^{\,2}\right]^{N/2}} \exp\left\{ - \frac{1}{2\sigma_\epsilon^{\,2}} \sum_{n=1}^{N} \left[y_n - \hat{x}_n(\underline{\theta}) \right]^2 \right\}. \tag{9}$$

Now since the logarithm is a monotonic function, maximizing the logarithm of the likelihood function yields the same result as maximizing the likelihood function itself. Thus we may define the log likelihood function as

$$\begin{aligned} L(\underline{\theta}) &= \ln \left\{ p(y_1, y_2, \ldots, y_N | \underline{\theta}) \right\} \\ &= - \frac{N}{2} \ln(2\pi) - N \ln(\sigma_\epsilon) - \frac{1}{2\sigma_\epsilon^{\,2}} \sum_{n=1}^{N} \left[y_n - \hat{x}_n(\underline{\theta}) \right]^2 \end{aligned} \tag{10}$$

Of the terms in (10), only the last is dependent upon $\underline{\theta}$ and this term appears in the expression with a negative sign. Therefore, the parameter vector which maximizes the log likelihood function, and thus the likelihood function, is the one which minimizes

$$\phi(\underline{\theta}) = \sum_{n=1}^{N} \left[y_n - \hat{x}_n(\underline{\theta}) \right]^2 \tag{11}$$

which is simply the least-squares functional. Thus we see that for the case of an ideal signal in gaussian white noise, nonlinear least-squares estimation provides the maximum likelihood estimator.

Before moving on to the direct nonlinear least-squares approach to the transducer signal parameter estimation problem, an indirect approach will be described.

2.3 The Principal Component Linear Prediction Method

Linear prediction is a method of difference equation modelling used to estimate the poles of exponential signals. Linear prediction simplifies a typically nonlinear problem by solving a related linear problem, namely estimating the coefficients in the difference equation. From these coefficients, the exponential poles can be obtained by forming and solving for the roots of the prediction polynomial (the z-plane representation of the difference equation).

Least squares linear prediction dictates the solution of an overdetermined system of equations to obtain the prediction coefficients. Principal component linear prediction takes this a step further and dictates the use of an overmodelled prediction error filter, i.e. a difference equation of order larger than the expected signal order, and the use of a rank reduced approximation to the pseudoinverse in the least squares solution for the coefficients. Consequently, principal component linear prediction neccessitates a selection process to separate the signal poles from the remaining estimates.

The signal parameter estimation algorithm can be summarized in the following three steps:

(1) solution for the prediction coefficients using the principal component method,

(2) signal pole selection, and

(3) linear least-squares solution of the signal amplitudes.

Step 1: Solution for Prediction Coefficients

It is well known that an q'th order discrete time linear system may be described by the q'th order forward difference equation

$$y(n) = a_1 y(n-1) + a_2 y(n-2) + \cdots + a_q y(n-q).$$

We may similarly describe the system using the backward difference equation

$$y(n) = b_1 y(n+1) + b_2 y(n+2) + \cdots + b_q y(n+q).$$

Moving all terms to the left hand side and taking the z-transform of the backward difference equation yields

$$Y(z)\ [1 - b_1\ z - b_2\ z^2 - \cdots - b_q\ z^q] = 0. \qquad (12)$$

By determining the coefficients $b_1, \ldots, b_q$ and equating the polynomial in brackets to 0, the system's z-plane poles can be found as the reciprocals of the backward prediction polynomial roots.

In linear prediction [11], an arbitrary linear system is modeled by a difference equation whose order (say, order L) is not necessarally equal to the system order; the prediction equation is said to be overmodelled or undermodelled, depending on if the order L is chosen greater or less than the system order. Also, depending on whether the coefficients are determined for the forward or backward difference equation, the technique is called forward or backward prediction, respectively.

Given data samples y_n, n=0,1,...,N-1, the system of backward backward predictions is

$$A\,\underline{b} = -\,\underline{h}, \tag{13}$$

where $\underline{b} = [b_1,\ldots,b_L]^T$ is the unknown vector of backward prediction coefficients, $\underline{h} = [y_0,\ldots,y_{N-L-1}]^T$, and A is the Hankel data matrix

$$A = \begin{bmatrix} y_1 & y_2 & \cdots & y_L \\ y_2 & y_3 & \cdots & y_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-L} & y_{N-L+1} & \cdots & y_{N-1} \end{bmatrix} \tag{14}$$

A primary difference which distinguishes the principal component method of linear prediction from that used in typical least squares linear prediction is first the use of a prediction order L much larger than the known or estimated system order (overmodelling), and then the use of a rank reduced approximant to the Hankel matrix using the singular value decomposition.

Briefly stated, the singular value decomposition factors an arbitrary N×M (usually N > M) matrix, A, as

$$A = U\,\Sigma\,V^T, \tag{15}$$

where U is the N×N orthogonal matrix of left singular vectors, V is the M×M orthogonal matrix of right singular vectors, and Σ is an N×M matrix whose only nonzero elements lie along the diagonal of the first M rows. These diagonal elements of Σ, called the singular values of A, are non-negative and arranged in non-increasing order. If A is a numerically rank deficient matrix with rank $r < M$, then only the first r singular values will be nonzero. If the true rank of the system is $r < M$, yet noise in the data causes the numerical rank of the matrix to equal M, then the first r singular values are called the principal singular values and rank reduction can be performed on the matrix A by setting to zero all but the these principal singular values.

Having performed the singular value decomposition and rank reduction of the Hankel data matrix A, the Moore-Penrose generalized inverse of A can be defined as

$$A^{\#} = V \Sigma^{\#} U^{T}. \tag{16}$$

Here, $\Sigma^{\#}$ is an M×N matrix whose only nonzero elements are the first r elements along the diagonal of the first M columns. These nonzero elements are the reciprocals of the r principal singular values. Given this pseudoinverse, the minimum norm linear least-squares solution for the backward prediction coefficient vector is

$$\underline{b} = - A^{\#} \underline{h}. \tag{17}$$

<u>Step 2</u> Signal Pole Selection

With the prediction coefficients in hand, a pool of z-plane pole estimates can be obtained by solving for and taking the reciprocal of the roots of the backward prediction polynomial

$$1 + b_1 z + b_2 z^2 + \cdots + b_L z^L = 0, \tag{18}$$

Due to the overmodeling used in the principal component method, these roots will be far more numerous than the actual signal poles, so that a number of these are 'extraneous' roots, from which the actual signal poles must be distinguished. In the transducer signal parameter estimation problem, these signal poles will include a steady-state pole which lies on the unit circle in the z-plane and the system poles which will fall outside of the unit circle. Kumerasean has observed [12] that while these signal poles fall outside of the unit circle, the extraneous roots will fall within the unit circle. Thus the location at which the backward prediction polynomial roots fall within the z-plane provides a method of pole selection.

While Kumerasean's method of selection works well at high signal-to-noise ratios, simulations have indicated that a signal-to-noise ratio threshold is reached, below which this method of pole selection cannot distinguish between the extraneous roots and the signal poles. Because of this, a more general method utilizing subset selection [10] has been adopted. This signal pole selection technique consists of three parts:

(1) reflection of roots into the unit circle and transformation to the s-plane

(2) replacement of the excitation pole, and

(3) subset selection of the remaining poles.

In part 1, the roots are selectively reflected about the unit circle so that all roots fall within the unit circle in the z-plane (so that all pole estimates will be stable). The roots are then transformed to the s-plane using the following equations:

$$s_i = -\alpha_i \pm j2\pi f_i$$

where

$$\alpha_i = -\frac{1}{T}\ln|z_i|$$

and

$$f_i = \frac{1}{2\pi T}\tan^{-1}\left[\frac{\text{imag}(z_i)}{\text{real}(z_i)}\right].$$

where T is the sampling interval, s_i are the s-plane poles, and z_i are the z-plane poles.

In part 2, all roots are examined one at a time and the root which lies closest to the known excitation pole in the s-plane is marked as the steady-state pole and removed from the pool of roots. The known theoretical pole value $(\alpha_0=0, 2\pi f_0)$ is then used throughout the remainder of the estimation process (in the subset selection of the other poles and the amplitude solution.

In part 3, the remaining roots are taken p at a time (where p is the number of system poles), and along with the excitation pole, used to form the basis function matrix for the linear least-squares equation for the amplitudes. The observation vector is then projected onto the column space of each of these basis function matrices and the group of p roots which result in the lowest residual sum of squares is chosen as the remaining pole estimates. In this way, the pole subset which best fits the data in the least-squares sense is determined.

Step 3: Amplitude Solution

The final step in the signal parameter estimation algorithm is the linear least-squares estimation of the signal amplitudes. This is performed by constructing the basis function matrix, F, consisting of complex exponentials corresponding to each of the estimates poles, for which the desired amplitudes are simply linear weighting factors. Defining $\underline{a}$ to be the vector of unknown amplitudes

$$\underline{a} = \left[A_{0_C}, A_{0_S}, A_{1_C}, A_{1_S}, \cdots , A_{K_C}, A_{K_S} \right]^T .$$

the problem is to solve for $\underline{a}$ in the linear least-squaresd problem

$$F \, \underline{a} \simeq \underline{y} . \qquad (19)$$

3. VARIABLE PROJECTION NONLINEAR LEAST-SQUARES THEORY

This chapter provides an exposition of the theory of the variable projection approach to solving nonlinear least-squares problems, as developed by Golub and Pereyra [3]. This approach partitions the unknown model parameters into two groups: those that occur linearly in the model and those that occur nonlinearly in the model. By performing this separation of variables, one is provided the opportunity to solve the nonlinear least-squares problem as a sequence of two computationally simpler problems.

Sections 3.1 through 3.4 introduce the notation and review vector-matrix calculus material necessary for the subsequent derivations. The remainder of the chapter derives three principal results: (1) the relationship between the linear and nonlinear parameters within the least-squares framework which allows the separation of variables, (2) an expression for the derivative of the projection operator, and (3) an expression for the gradient of the variable projection functional. With these results, a wide range of standard solution techniques are available for minimizing the least-squares functional in the two-step procedure described.

3.1 Parameter Model Definition

The variable projection method is applicable to parameter estimation problems in which the signal model can be decomposed as a set of nonlinear basis functions weighted by a set of coefficients. These nonlinear basis functions are a function only of the independent variable (say, time) and a set of parameters which we will call collectively the nonlinear parameter vector. Symbolically, we may write this as

$$x_n = \sum_{j=1}^{M} a_j \, f_j(\underline{\theta}, t_n) \quad , \tag{20}$$

alternatively written as

$$x_n = \underline{f}^T(\underline{\theta}, t_n) \, \underline{a} \tag{21}$$

where

$$\underline{\theta} = \left[\, \theta_1, \, \theta_2, \, \ldots \, , \, \theta_K \, \right]^T \tag{22}$$

$$= \text{nonlinear parameter vector,}$$

$$\underline{a} = \left[\, a_1, \, a_2, \, \ldots \, , \, a_M \, \right]^T \tag{23}$$

$$= \text{linear parameter vector,}$$

and

$$\underline{f}^T(\underline{\theta}, t_n) = \left[\, f_1(\underline{\theta}, t_n), \, f_2(\underline{\theta}, t_n), \, \ldots \, , \, f_M(\underline{\theta}, t_n) \, \right] \tag{24}$$

$$= \text{basis function vector.}$$

3.2 The Least-Squares Functional

The parameter estimation problem attempts to form an estimate, $\hat{x}_n$, of an ideal signal, x_n, from observed data, presumably of the ideal signal in noise. This is written symbolically as

$$y_n = x_n + w_n, \qquad n = 0,1,\ldots,N\text{-}1 \tag{25}$$

where w_n are independently distributed random variables and x_n are as described in the previous section.

In forming this estimate of x_n, we seek to minimize, by appropriate choice of $\underline{a}$ and $\underline{\theta}$, the Least-Squares Functional (LSF), defined as

$$\phi(\underline{a},\underline{\theta}) = \left\| \, \underline{e}(\underline{a},\underline{\theta}) \, \right\|_2^2 \tag{26}$$

$$= \sum_{n=0}^{N-1} \left[\, y_n - \hat{x}(\underline{a},\underline{\theta},t_n) \, \right]^2 \tag{27}$$

$$= \sum_{n=0}^{N-1} \left[\, y_n - \sum_{j=1}^{M} a_j \; f_j(\underline{\theta},t_n) \, \right]^2 \tag{28}$$

$$= \sum_{n=0}^{N-1} \left[\, y_n - \underline{f}^T(\underline{\theta},t_n) \; \underline{a} \, \right]^2 \tag{29}$$

Here, $\underline{e}(\underline{a},\underline{\theta})$ is the signal estimation error vector. Since the Euclidean norm (2-norm) $||\bullet||_2$ is used throughout this development, the subscript will subsequently be dropped for convenience.

Let us now define the independent variable vector

$$\underline{t} = \left[\, t_0, \, t_1, \, \cdots \, , \, t_{N-1} \, \right]^T,$$

and write the observed data in vector notation as

$$\underline{y} = \left[\, y_0, \, y_1, \, \cdots \, , \, y_{N-1} \, \right]^T. \tag{30}$$

Let us also define the basis function matrix

$$F(\underline{\theta},\underline{t}) \;=\; \begin{bmatrix} \underline{f}^T(\underline{\theta},t_0) \\ \underline{f}^T(\underline{\theta},t_1) \\ \vdots \\ \underline{f}^T(\underline{\theta},t_{N-1}) \end{bmatrix} \;=\; \begin{bmatrix} f_1(\underline{\theta},t_0) & \cdots & f_M(\underline{\theta},t_0) \\ f_1(\underline{\theta},t_1) & \cdots & f_M(\underline{\theta},t_1) \\ \vdots & \ddots & \vdots \\ f_1(\underline{\theta},t_{N-1}) & \cdots & f_M(\underline{\theta},t_{N-1}) \end{bmatrix} \tag{31}$$

whose elements are independent of the linear parameters and whose rows each correspond to an element of the observation vector.

We may now define the LSF in terms of the basis function matrix by writing

$$\phi(\underline{a},\underline{\theta}) = \left\| \, \underline{y} - F(\underline{\theta}) \; \underline{a} \, \right\|^2, \tag{32}$$

where we have dropped the explicit time dependence for convenience.

$\phi(\underline{a},\underline{\theta})$ is minimized when $\nabla\phi(\underline{a},\underline{\theta})=0$, where ∇ is the gradient operator. While developing an expression for this gradient (of the LSF), the linear and nonlinear parts of the model will be separated and a new cost functional, the variable projection functional, will be defined.

3.3 Differentiation with Respect to a Vector

Consider the vector

$$\underline{x} = \left[\, x_1,\, x_2,\, \ldots\, ,\, x_K \,\right]^T .$$

Differentiation of a scalar function $\Psi(\underline{x})$ with respect to $\underline{x}$ yields

$$\frac{\partial\Psi(\underline{x})}{\partial\underline{x}} = \left[\, \frac{\partial\Psi(\underline{x})}{\partial x_1} ,\, \frac{\partial\Psi(\underline{x})}{\partial x_2} ,\, \ldots\, ,\, \frac{\partial\Psi(\underline{x})}{\partial x_K} \,\right]^T . \tag{33}$$

A particularly interesting case of this is the quadratic form,

$$\Psi(\underline{x}) = \underline{x}^T A\, \underline{x} \tag{34}$$

where A is a K × K constant matrix. For this case, we obtain simply

$$\frac{\partial\Psi(\underline{x})}{\partial\underline{x}} = 2\, A\, \underline{x} . \tag{35}$$

Now consider the (row) vector function $\underline{f}\ (\underline{x})$ defined by

$$\underline{f}^T(\underline{x}) = \left[\, f_1(\underline{x}),\, f_2(\underline{x}),\, \ldots\, ,\, f_M(\underline{x}) \,\right] .$$

Differentiation of $\underline{f}\ (\underline{x})$ with respect to $\underline{x}$ yields the K × M Jacobian matrix

$$\frac{\partial \underline{f}^T(\underline{x})}{\partial \underline{x}} = \begin{bmatrix} \frac{\partial f_1(\underline{x})}{\partial x_1} & \cdots & \frac{\partial f_M(\underline{x})}{\partial x_1} \\ \frac{\partial f_1(\underline{x})}{\partial x_2} & \cdots & \frac{\partial f_M(\underline{x})}{\partial x_2} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_1(\underline{x})}{\partial x_K} & \cdots & \frac{\partial f_M(\underline{x})}{\partial x_K} \end{bmatrix} . \tag{36}$$

Finally, consider the N x M matrix function $F(\underline{x},\underline{t})$ defined by

$$\left\{ F(\underline{x},\underline{t}) \right\}_{ij} = f_j(\underline{x},t_i) \ , \quad i = 1,2,\ldots,N \ , \quad j = 1,2,\ldots,M.$$

Defining $D(F) = \frac{\partial F(\underline{x},\underline{t})}{\partial \underline{x}}$, the resulting derivative will be an NxMxK tensor (three-dimensional array) defined by

$$\left\{ D(F) \right\}_{ijk} = \frac{f_j(\underline{x},t_i)}{\partial x_k} \ , \quad i = 1,2,\ldots,N \ , \quad j = 1,2,\ldots,M \ , \quad k = 1,2,\ldots,K. \tag{37}$$

Note that this derivative tensor can be viewed as a series of partial derivative matrices, or 'slabs', each corresponding to differentiation with respect to one of the variables in the vector. This type of differentiation is described in [3] and is called the Frechet derivative of a mapping.

3.4 Differentiation of the Squared Norm

The squared norm of a vector function $\underline{f}(\underline{x})$ is a scalar function $\gamma(\underline{x})$ defined by

$$\gamma(\underline{x}) = \underline{f}^T(\underline{x}) \ \underline{f}(\underline{x}) \tag{38}$$

$$= \sum_{j=1}^{M} \left[\ f_j(\underline{x}) \ \right]^2 \tag{39}$$

Differentiation of this with respect to $\underline{x}$ thus yields

$$\frac{\partial\gamma(\underline{x})}{\partial\underline{x}} = \left[\frac{\partial\gamma(\underline{x})}{\partial x_1} , \frac{\partial\gamma(\underline{x})}{\partial x_2} , \ldots , \frac{\partial\gamma(\underline{x})}{\partial x_K} \right]^T \tag{40}$$

where

$$\frac{\partial\gamma(\underline{x})}{\partial x_k} = 2 \sum_{j=1}^{M} f_j(\underline{x}) \frac{\partial f_j(\underline{x})}{\partial x_k} . \tag{41}$$

Substituting (41) into (40) and simplifying, we obtain

$$\frac{\partial\gamma(\underline{x})}{\partial\underline{x}} = \left[\begin{array}{c} 2 \sum_{j=1}^{M} f_j(\underline{x}) \frac{\partial f_j(\underline{x})}{\partial x_1} \\ 2 \sum_{j=1}^{M} f_j(\underline{x}) \frac{\partial f_j(\underline{x})}{\partial x_2} \\ \bullet \\ \bullet \\ \bullet \\ 2 \sum_{j=1}^{M} f_j(\underline{x}) \frac{\partial f_j(\underline{x})}{\partial x_K} \end{array} \right] = 2 \frac{\partial \underline{f}^T(\underline{x})}{\partial\underline{x}} \underline{f}(\underline{x}) . \tag{42}$$

3.5 Gradient of the Least-Squares Functional

Recall the Least-Squares Functional (LSF) given by

$$\begin{aligned} \phi(\underline{a},\underline{\theta}) &= \left|\left| \underline{e}(\underline{a},\underline{\theta}) \right|\right|^2 \\ &= \left|\left| \underline{y} - F(\underline{\theta}) \underline{a} \right|\right|^2 \end{aligned}$$

where $\underline{y}$ is the N-vector of observations,

$\underline{a}$ is the M-vector of linear parameters,

$\underline{\theta}$ is the K-vector of nonlinear parameters, and

$F(\underline{\theta})$ is the N × M basis function matrix.

By partitioning the overall parameter vector as $[\underline{a}^T, \underline{\theta}^T]^T$, we may write the gradient function in partitioned form as

$$\nabla\phi(\underline{a},\underline{\theta}) = \begin{bmatrix} \dfrac{\partial\phi(\underline{a},\underline{\theta})}{\partial\underline{a}} \\[2ex] \dfrac{\partial\phi(\underline{a},\underline{\theta})}{\partial\underline{\theta}} \end{bmatrix} \tag{43}$$

A critical point of $\phi(\underline{a},\underline{\theta})$ is found by evaluating $\nabla\phi(\underline{a},\underline{\theta})=0$, which, in general, requires the simultaneous satisfaction of

$$\frac{\partial\phi(\underline{a},\underline{\theta})}{\partial\underline{a}} = 0 \tag{44}$$

and

$$\frac{\partial\phi(\underline{a},\underline{\theta})}{\partial\underline{\theta}} = 0. \tag{45}$$

Let us focus, for the moment, on the evaluation of (44). Applying (42) to (32), we obtain

$$\frac{\partial\phi(\underline{a},\underline{\theta})}{\partial\underline{a}} = 2\left\{ \frac{\partial}{\partial\underline{a}} \left[\underline{y} - F(\underline{\theta})\ \underline{a} \right]^T \right\} \left[\underline{y} - F(\underline{\theta})\ \underline{a} \right] \tag{46}$$

$$= -2\left\{ \frac{\partial}{\partial\underline{a}} \left[F(\underline{\theta})\ \underline{a} \right]^T \right\} \left[\underline{y} - F(\underline{\theta})\ \underline{a} \right]$$

$$= -2\ F^T(\underline{\theta}) \left[\underline{y} - F(\underline{\theta})\ \underline{a} \right] \tag{47}$$

where it has been noted that

$$\frac{\partial F^T(\underline{\theta})}{\partial\underline{a}} = 0 \quad \text{and} \quad \frac{\partial\underline{a}^T}{\partial\underline{a}} = I.$$

Equating (47) to zero and rearranging yields

$$F^T(\underline{\theta})\ F(\underline{\theta})\ \underline{a} = F^T(\underline{\theta})\ \underline{y} \tag{48}$$

which, for a given $\underline{\theta}$, represent the linear least-squares normal equations in which the observation vector $\underline{y}$ is projected onto the range of the basis function matrix to obtain the vector $\underline{a}$. This result will be utilized in the next section in developing a cost functional for the nonlinear parameters independent of the linear parameters.

3.6 The Variable Projection Functional

In the previous section, it was shown that differentiating the least-squares functional with respect to the linear parameter vector and evaluating at zero led to the linear least-squares normal equations for $\underline{a}$. The Gauss-Markov Theorem [6] demonstrates that the Moore-Penrose generalized inverse, or pseudoinverse, provides the minimum variance solution to the linear least-squares problem. Thus, given a maximum likelihood estimator for the nonlinear parameter vector, we may write the maximum likelihood estimator for the linear parameters as

$$\underline{\hat{a}}_{ML} = F^{\#}(\underline{\hat{\theta}}_{ML})\ \underline{y} \qquad (49)$$

where $F^{\#}$ denotes the Moore-Penrose generalized inverse.

By substituting $\underline{\hat{a}}_{ML}$ back into the LSF, we are able to transform the minimization problem into one in which we first minimize with respect to the nonlinear parameters, and then solve for the linear parameters as a linear least-squares problem. This technique leads to the Variable Projection Functional (VPF) which is defined as

$$\phi_2(\underline{\theta}) = \phi(\underline{a},\underline{\theta}\,|\,\underline{\hat{a}}_{ML}) \qquad (50)$$

$$= \left|\left|\ \underline{y} - F(\underline{\theta})\ F^{\#}(\underline{\theta})\ \underline{y}\ \right|\right|^2 \qquad (51)$$

$$= \left|\left|\ (I - P_F)\ \underline{y}\ \right|\right|^2 \qquad (52)$$

$$= \left|\left|\ P_F^{\perp}\ \underline{y}\ \right|\right|^2 \qquad (53)$$

Here $P_F = F(\underline{\theta})\ F(\underline{\theta})^{\#}$ is the projection operator onto the column space of the basis function matrix and $P_F^{\perp} = I - P_F$ is the projection operator onto the orthogonal complement of the column space of the basis function matrix.

The argument just laid out is the key to the variable projection method. Golub and Pereyra [3] provide a proof that minimization using the variable projection method leads to the same critical point as would the traditional least-squares solution technique in which the LSF is minimized with respect to all parameters simultaneously.

3.7 The Derivative of the Projection Operator

In developing the gradient for the Variable Projection Functional (VPF), an expression for the derivative of the projection operator with respect to the nonlinear parameter vector, $D(P_F)$, will be needed. It is also useful in itself when minimizing the VPF using the Gauss-Newton iterative scheme.

It is important to note prior to this development that although the generalized inverse (g-inverse) arose in our application from the linear least-squares normal equations, and thus implied the Moore-Penrose pseudoinverse, the formation of the VPF and its derivative require a g-inverse suitable for forming the projection operator only, thus this g-inverse need satisfy only (54) - (56) below. A heuristic argument for this is that projecting a vector onto the column space of a matrix is a simpler task than finding the minimum norm linear least-squares solution, thus the requirements upon the g-inverse can be less stringent. A more formal treatment of this matter will be provided in the next chapter, while the present discussion will proceed with (54) - (58) as assertions. The first three assertions are properties of g-inverses, (57) is an expression for the projection operator, and (58) is the product rule of differentiation. The symbol F^+ will denote the g-inverse of the matrix F throughout the discussion.

$$F\,F^+\,F = F \tag{54}$$

$$\left[\,F\,F^+\,\right]^T = F\,F^+ \tag{55}$$

$$F^+\,F\,F^+ = F^+ \tag{56}$$

$$P_F = F\,F^+ \tag{57}$$

$$D(AB) = D(A)\,B + A\,D(B) \tag{58}$$

Combining (54) and (57) and then applying eq. (58), we obtain

$$\begin{aligned} D(F) &= D(P_F F) \\ &= D(P_F)\,F + P_F\,D(F). \end{aligned}$$

Rearranging yields

$$D(P_F)\,F = D(F) - P_F\,D(F),$$

and recalling that

$$P_F^{\perp} = I - P_F,$$

we see that

$$D(P_F)\,F = P_F^{\perp}\,D(F). \tag{59}$$

Postmultiplying by F^+ then yields

$$\begin{aligned} D(P_F)\,P_F &= D(P_F)\,F\,F^+ \\ &= P_F^{\perp}\,D(F)\,F^+. \end{aligned} \tag{60}$$

Transposing the left-hand side of this equation yields

$$\left[\,D(P_F)\,P_F\,\right]^T = P_F^{\,T}\left[\,D(P_F)\,\right]^T.$$

If we now partition $D(P_F)$ as

$$D(P_F) = \left[\,\frac{\partial P_F}{\partial\theta_1}\,\middle|\,\frac{\partial P_F}{\partial\theta_2}\,\middle|\,\cdots\,\middle|\,\frac{\partial P_F}{\partial\theta_K}\,\right],$$

then transposition within the derivative tensor is equivalent to transposition within each of the partial derivative 'slabs' (see section 3.3) shown in the above partition. Thus

$$\left[\ D(P_F)\ \right]^T = \left[\ \left(\frac{\partial P_F}{\partial \theta_1}\right)^T \middle|\ \left(\frac{\partial P_F}{\partial \theta_2}\right)^T \middle|\ \cdots\ \middle|\ \left(\frac{\partial P_F}{\partial \theta_K}\right)^T\ \right].$$

Now noting the symmetry of the projection operator and its partial derivative (the projection operator is, by definition, symmetric and idempotent) we have shown that

$$\left[\ D(P_F)\ \right]^T = D(P_F) \quad \text{and} \quad \left[\ D(P_F)\ P_F\ \right]^T = P_F\ D(P_F). \tag{61}$$

Now noting that the projection operator is idempotent, we write

$$(P_F)^2 = P_F.$$

Substituting this into (58), we obtain

$$D(P_F) = D(P_F\ P_F) = D(P_F)\ P_F + P_F\ D(P_F)$$

which, after applying (60) and (61), becomes

$$D(P_F) \ = \ P_F^{\perp}\ D(F)\ F^+ + \left[\ D(P_F)\ P_F\ \right]^T. \tag{62}$$

Again using (60) in the rightmost term, we obutain an expression for the derivative of the projection operator which can be evaluated,

$$D(P_F) \ = \ P_F^{\perp}\ D(F)\ F^+ + \left[\ P_F^{\perp}\ D(F)\ F^+\ \right]^T. \tag{63}$$

3.8 The Gradient of the Variable Projection Functional

Armed with the derivative of the projection operator, we may now form an expression for the gradient of the Variable Projection Functional (VPF).

Recall the definition of the VPF, given as

$$\phi_2(\underline{\theta}) = \left|\left|\ P_F^{\perp}\ \underline{y}\ \right|\right|^2.$$

Applying (42), i.e. taking the derivative of the VPF with respect to the nonlinear parameter vector, we obtain the gradient of the VPF as follows:

$$\begin{aligned}\nabla\phi_2(\underline{\theta}) &= 2\,\frac{\partial}{\partial\underline{\theta}}\left\{\left[\,P_F^{\perp}\;\underline{y}\,\right]^T\right\}P_F^{\perp}\,\underline{y}\\ &= 2\,\frac{\partial}{\partial\underline{\theta}}\left\{\underline{y}^T\left[\,P_F^{\perp}\,\right]^T\right\}P_F^{\perp}\,\underline{y}\\ &= 2\,\underline{y}^T\,D\left\{\left[\,P_F^{\perp}\,\right]^T\right\}P_F^{\perp}\,\underline{y}\;.\end{aligned}$$

Noting that

$$D\left\{\left[\,P_F^{\perp}\,\right]^T\right\} = D\left\{\,P_F^{\perp}\,\right\} = -D(P_F),$$

we obtain

$$\frac{1}{2}\,\nabla\phi_2(\underline{\theta}) = -\,\underline{y}^T\,D(P_F)\,P_F^{\perp}\,\underline{y}\;. \tag{64}$$

If we now substitute (63) for the derivative of the projector, we get

$$\begin{aligned}\frac{1}{2}\,\nabla\phi_2(\underline{\theta}) &= -\,\underline{y}^T\left\{\,P_F^{\perp}\,D(F)\,F^{+} + \left[\,P_F^{\perp}\,D(F)\,F^{+}\,\right]^T\right\}P_F^{\perp}\,\underline{y}\\ &= -\,\underline{y}^T\,P_F^{\perp}\,D(F)\,F^{+}\,P_F^{\perp}\,\underline{y}\;-\,\underline{y}^T\left[\,P_F^{\perp}\,D(F)\,F^{+}\right]^T P_F^{\perp}\,\underline{y}\\ &= -\,\underline{y}^T\,P_F^{\perp}\,D(F)\,F^{+}\,P_F^{\perp}\,\underline{y}\;-\,\underline{y}^T\,(F^{+})^T\,D(F^T)\,P_F^{\perp}\,\underline{y}\;.\end{aligned} \tag{65}$$

In arriving at (65), we recognized that $P_F^{\perp}$ is symmetric and idempotent, therefore

$$\left[\,P_F^{\perp}\,\right]^T P_F^{\perp} = P_F^{\perp}\;.$$

Now noting that

$$F^{+}\,P_F^{\perp} = F^{+}\left[\,I - F\,F^{+}\,\right] = F^{+} - \;F^{+}\,F\,F^{+} = 0,$$

we see that the first term in the gradient becomes zero, leaving

$$\begin{aligned}\frac{1}{2}\,\nabla\phi_2(\underline{\theta}) &= -\,\underline{y}^T\,(F^{+})^T\,D(F^T)\,P_F^{\perp}\,\underline{y}\\ &= -\left[\,\underline{y}^T\,P_F^{\perp}\,D(F)\,F^{+}\,\underline{y}\,\right]^T.\end{aligned} \tag{66}$$

Equations (63) and (66) provide the ability to use any of the gradient minimization techniques, which will be summarized as part of Chapter 5, as well as the variable metric techniques, in solving the variable projection nonlinear least-squares problem.

4. LINEAR LEAST-SQUARES THEORY

As is clear from the preceeding chapter, linear least-squares theory, particularly the concepts of the generalized inverse and the orthogonal projection operator, are fundamental to general least-squares theory.

The purpose of this chapter is two-fold. The first objective is to review in some detail the characteristics of the projection operator and the generalized inverse. The second objective is to summarize what works, what doesn't, and when, as the popular QR factorizations are applied to forming projection operators and solving linear least-squares problems.

Section 4.1 focuses on the orthogonal projection operator. Here, following the the work of Halmos [13], the properties of the projection operator, namely that it is idempotent and symmetric, will be discussed. The eigenstructure of this operator is also examined.

Sections 4.2 through 4.4 examine the properties of the generalized inverse for finding solutions to consistent equations and linear least-squares problems and finding minimum norm solutions to consistent equations. The Moore-Penrose pseudoinverse, which leads to the minimum norm least-squares solution, will then be examined.

Sections 4.6 through 4.9 look at the QR factorization, which is seen in Section 4.6 to lead to a generalized inverse which is adequate for solving full rank linear least-squares problems. Sections 4.7 and 4.8 look at the rank deficient case, and give results for what we will call the truncated QR factorization and for the complete orthogonal factorization as outlined by Hanson and Lawson [14], Golub and Pereyra [15], and Golub and Van Loan [16]. Here it is seen that while the g-inverse formed using the truncated QR factorization does not lead to the Moore-Penrose pseudoinverse, the complete orthogonal factorization does achieve the minimum norm solution, and thus provides an alternative to the singular value decomposition for performing the necessary rank reduction. Finally, the application of the complete orthogonal factorization to nearly rank deficient matrices is discussed breifly in section 4.9.

4.1 The Projection Operator

Consider the N-dimensional space $\mathbf{R}^N$ and a linear subspace $S \subset \mathbf{R}^N$. There exists a linear subspace $S^\perp$, called the orthogonal complement of S in $\mathbf{R}^N$, such that $\mathbf{R}^N$ is the direct sum of S and $S^\perp$.

Drawing from the work of Halmos [13], the following definitions and equations (67) through (73) characterize the projection operators.

Definition There exists an operator, P_S, called the projector onto S, which maps every vector in $\mathbf{R}^N$ onto S.

Definition There exists an operator, $P_S^\perp$, called the projector onto the complement of S^Nin $\mathbf{R}$, which maps every vector in $\mathbf{R}^N$ onto $S^\perp$.

Furthermore,

$$P_S^\perp = I - P_S. \qquad (67)$$

Now consider the vectors

$$\underline{x} \in S \qquad \underline{y} \in S^\perp \qquad \underline{z} \in \mathbf{R}^N,$$

where

$$\underline{z} = \underline{z}_1 + \underline{z}_2, \qquad \underline{z}_1 \in S \text{ and } \underline{z}_2 \in S^\perp.$$

The projection operators defined above satisfy the following six relationships:

$$P_S \underline{x} = \underline{x} \qquad (68) \qquad\qquad P_S^{\perp} \underline{x} = 0 \qquad (69)$$

$$P_S \underline{y} = 0 \qquad (70) \qquad\qquad P_S^{\perp} \underline{y} = \underline{y} \qquad (71)$$

$$P_S \underline{z} = \underline{z}_1 \qquad (72) \qquad\qquad P_S^{\perp} \underline{z} = \underline{z}_2 \qquad (73)$$

From (68) and (71), the projection operator is easily shown to be idempotent.

$$P_S^2 \underline{x} = P_S (P_S \underline{x}) = P_S \underline{x},$$

therefore

$$P_S^2 = P_S. \tag{74}$$

Also, we see from (70) and (73) that

$$P_S (P_S^{\perp} \underline{z}) = 0 \qquad \text{so that} \qquad P_S P_S^{\perp} = 0. \tag{75}$$

Similarly, from (69) and (72), we see that

$$P_S^{\perp} (P_S \underline{z}) = 0 \qquad \text{so that} \qquad P_S^{\perp} P_S = 0.$$

Finally, from (72) and (73), and from the definition of orthogonality

$$(P_S \underline{z})^T (P_S^{\perp} \underline{z}) = 0.$$

Substituting (67) and transposing within the parantheses

$$\underline{z}^T P_S^T (I - P_S) \underline{z} = 0.$$

Since this is true for all $\underline{z}$ in $\mathbf{R}^N$,

$$P_S^T = P_S^T P_S. \tag{76}$$

The right hand side is seen, by inspection, to be symmetric. Therefore the projection operator must also satisfy

$$P_S = P_S^T. \tag{77}$$

In summary, the projection operators P_S and $P_S^{\perp}$ are idempotent, symmetric, and mutually annihilating.

Further insight can be gained by examining the eigenstructure of the projector. Since the subspace S is invariant under the transformation P_S, it is known that S is spanned by some set of the eigenvectors of P_S which correspond to a multiple unity eigenvalue [13]. The remaining eigenvectors span the complement of S in R^N and correspond to eigenvalues of zero. The eigenvectors of P_S will also form a basis for R^N.

To see that this eigenstructure exemplifies the operation of the projector, consider the arbitrary N-vector $\underline{z}$ in the space R^N, which has a set of basis vectors $\underline{u}_i$, i=1,...,N. We may choose this set of basis vectors to be the eigenvectors of the N x N matrix P_S, which is the projector onto the (say, M-dimensional) subspace S. We may now write

$$P_S \underline{u}_i = \lambda_i \underline{u}_i . \tag{78}$$

and

$$\underline{z} = a_1 \underline{u}_1 + a_2 \underline{u}_2 + \cdots + a_N \underline{u}_N \tag{79}$$

thus

$$P_S \underline{z} = \lambda_1 a_1 \underline{u}_1 + \lambda_2 a_2 \underline{u}_2 + \cdots + \lambda_N a_N \underline{u}_N . \tag{80}$$

M of the N eigenvalues, those corresponding to the eigenvectors which span the subspace S, will have value unity, while the remaining eigenvalues will have a value of zero. We therefore have the result

$$P_S \underline{z} = \sum_{i=1}^{M} a_{j_i} \underline{u}_{j_i} \tag{81}$$

where the indices j_i denote those eigenvectors which span S.

4.2 The Generalized Inverse

In general, an N x M matrix A is a linear transformation which maps an arbitrary vector $\underline{x}$ from an M-dimensional space to an N-dimensional space which is the range (column space) of the mapping (matrix). We desire an inverse transformation which will map an N-vector $\underline{y}$ lying in the range of A back into an M-dimensional space. If the vector $\underline{y}$ does not lie in the range of A, then the inverse mapping must first approximate $\underline{y}$ with a suitable vector which is in the range of the mapping. Let us first consider the case where $\underline{y}$ does lie in the column space of A (consistent equations).

For the N × M matrix A, the M × N matrix A^+ is a generalized inverse (g-inverse) of A if

$$\underline{x} = A^+ \underline{y}$$

is a solution to the equation

$$A \underline{x} = \underline{y},$$

for any $\underline{y}$ which makes the system consistent [17]. Clearly, then, for consistent equations,

$$A A^+ \underline{y} = \underline{y}.$$

Suppose

$$\underline{y} = A \underline{z}$$

for some arbitrary M-vector $\underline{z}$ (which is obviously in the range of A), then

$$A A^+ A \underline{z} = A \underline{z}.$$

In general, this requires that

$$A A^+ A = A. \tag{82}$$

In the most unrestricted sense, this is all that is required of a g-inverse. If we wish, however, to consider the case of inconsistent equations, then we must impose further restrictions which determine how we wish to approximate $\underline{y}$ before transforming.

4.3 g-inverse for Linear Least-Squares Solution

Geometrically, the best approximant to $\underline{y}$ which makes the system consistent is the projection of $\underline{y}$ onto the column space of A. Thus for the arbitrary vector $\underline{z}$, the inconsistent equation

$$A\ \underline{x} \simeq \underline{z}$$

can be made consistent by premultiplying both sides by the projection operator for the column space of A, yielding

$$P_A\ A\ \underline{x} = P_A\ \underline{z}.$$

But the projection of the columns of A onto themselves leaves them unaffected, so this reduces to

$$A\ \underline{x} = P_A\ \underline{z}. \tag{83}$$

The g-inverse solution for $\underline{x}$ is then

$$\underline{x} = A^{+}\ P_A\ \underline{z}.$$

If we substitute this into (83) and note, on the right hand side of (83), that the projection operator is idempotent, this becomes

$$A\ A^{+}\ P_A\ \underline{z} = P_A\ P_A\ \underline{z}.$$

From this, we see that the generalized inverse for solving linear least-squares problems must be such that

$$P_A = A\ A^{+}$$

is, in fact, the projector onto the columns space of A. This then requires that the product A A^{+} be idempotent and symmetric. That this is idempotent follows from (82), so that no new restriction is imposed. We do have the further restriction, though, that A^{+} must satisfy

$$[A\ A^{+}]^{T} = A\ A^{+}. \tag{84}$$

4.4 g-inverse for Minimum Norm Solution

We know from the Section 4.2 that the g-inverse which solves the consistent equation

$$A \underline{x} = \underline{y} \tag{85}$$

must satisfy

$$A A^{+} A = A.$$

From this, it follows that

$$A - A A^{+} A = 0,$$

thus

$$A [I - A^{+} A] = 0.$$

We can therefore state that for any $\underline{z}$,

$$A [I - A^{+} A] \underline{z} \tag{86}$$

is a solution to the homogeneous equation

$$A \underline{x} = 0.$$

Analogously to the linear differential equatian, the general solution for the set of simultaneous linear equations in (85) is the sum of the homogeneous solution and a particular solution, and can thus be written

$$\underline{x} = A^{+}\underline{y} + [I - A^{+} A] \underline{z}.$$

Denote the g-inverse which leads to the minimum norm solution as A^{+}_{m}. We desire then that

$$\left|\left| A^{+}_{m} \underline{y} \right|\right|^2 \leq \left|\left| A^{+} \underline{y} + \left[I - A^{+} A \right] \underline{z} \right|\right|^2 \tag{87}$$

for all $\underline{y}$ and $\underline{z}$.

Note that

$$\left|\left| A^{+} \underline{y} + \left[I - A^{+} A \right] \underline{z} \right|\right|^2$$

$$= \left\{ A^{+} \underline{y} + \left[I - A^{+} A \right] \underline{z} \right\}^{T} \left\{ A^{+} \underline{y} + \left[I - A^{+} A \right] \underline{z} \right\}.$$

Thus

$$\left|\left| A_m^+ \underline{y} \right|\right|^2 \leq \left|\left| A^+ \underline{y} \right|\right|^2 + \underline{y}^T \left[A^+ \right]^T \left[I - A^+ A \right] \underline{z}$$
$$+ \underline{z}^T \left[I - A^+ A \right]^T A^+ \underline{y} + \left|\left| \left[I - A^+ A \right] \underline{z} \right|\right|^2.$$

This is a minimum when the two middle terms are zero, which occurs when the particular solution is orthogonal to the homogeneous solution. In general, this requires that

$$[A^+]^T \, [I - A^+ A] = 0.$$

Thus

$$[A^+]^T = [A^+]^T A^+ A.$$

For this to be true, it is necessary and sufficient [17] that

$$A^+ A A^+ = A^+ \tag{88}$$

and

$$[A^+ A]^T = A^+ A. \tag{89}$$

From (88) and (89), we see that the product $A^+ A$ is both idempotent and symmetric and is thus a projection operator. We will now show that it is, in fact, the projector onto the row space of A.

$$\begin{aligned} A^+ A &= [A^+ A]^T \\ &= A^T [A^+]^T \\ &= A^T [A^T]^+, \end{aligned}$$

which is the projection operator onto the columns of A^T, which are the rows of A.

In summary, the g-inverse for obtaining a minimum norm solution to

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

must be such that

$${}_A P = P_{A^T} = A^+ A \tag{90}$$

is the projection operator onto the row space of A.

Let us now re-examine the general solution, now written

$$\underline{x} = A^{+} \underline{y} + [I - {}_{A}P] \ \underline{z}$$

where we have substituted (90) into the homogeneous solution. Recall from Section 4.1 that [I - P] is the projector onto the complement of the subspace for which P is the projector; thus we see that the homogeneous solution is confined to the null space of A. Since the minimum norm solution must be orthogonal to this homogeneous solution, what we are really striving for in the minimum norm solution is that solution which lies in the row space of A.

4.5 g-inverse for Minimum Norm Least-squares Solution

Combining the results of the last two sections, we see that the g-inverse for obtaining the minimum norm least-squares solution, i.e. the Moore-Penrose generalized inverse (pseudoinverse), must be such that

$$P_A = A \ A^{+}$$

and

$${}_{A}P = A^{+} \ A$$

are, respectively, the orthogonal projectors onto the column space and the row space of OA. This is equivalent to the following:

$$A \ A^{+} \ A = A$$

$$[A \ A^{+}]^{T} = A \ A^{+}$$

$$A^{+} \ A \ A^{+} = A^{+}$$

$$[A^{+} \ A]^{T} = A^{+} \ A.$$

It is interesting to note that in forming the minimum norm linear least-squares solution, we are actually performing a three-stage process. Starting with the least-squares problem

$$A \, \underline{x} \simeq \underline{y} \tag{91}$$

where A is not necessarily full rank, we obtain the minimum norm solution

$$\underline{x} = A^{+} \, \underline{y}. \tag{92}$$

Stage I: Projection of $\underline{y}$ onto the column space of A to obtain a consistent set of equations. This can be shown explicitly by substituting (92) into (91) above to yield

$$\begin{aligned} \underline{\hat{y}} &= A \, A^{+} \, \underline{y} \\ &= P_{A} \, \underline{y} \end{aligned}$$

Stage II: Solution to the consistent set of equations

$$A \, \underline{\hat{x}} = \underline{\hat{y}}$$

to yield the general solution

$$\underline{\hat{x}} = A^{+} \, \underline{\hat{y}} + [I - A^{+} \, A] \, \underline{z},$$

where, $\underline{z}$ is an arbitrary vector in R^{N}.

Stage III: Projection of $\underline{\hat{x}}$ onto the row space of A to obtain the minimum norm solution (eliminate the homogeneous part of the solution). This can be shown explicitly by substiting (91) into (92) above to yield

$$\begin{aligned} \underline{x} &= A^{+} \, A \, \underline{\hat{x}} \\ &= {}_{A}P \, \underline{\hat{x}} \\ &= {}_{A}P \, A^{+} \, P_{A} \, \underline{y}. \end{aligned}$$

With this, we conclude our formal discussion of linear least-squares theory. The remainder of this chapter will examine the QR factorization family as they are used for forming projection operators and solving linear least-squares problems. In particular, we will look at how well the g-inverses constructed with these factorizations conform to to the equations outlined in this section.

4.6 QR Factorization of Full Rank Matrices

Consider the $N \times M$ matrix A of rank $r = M \leq N$. There exists an $N \times N$ orthogonal matrix Q, such that

$$Q\,A = R \equiv \left[\begin{array}{c} R_1 \\ \hline 0 \end{array}\right] \tag{93}$$

where

$$R_1 = \begin{bmatrix} \diagdown & /\!/ \\ 0 & \diagdown \end{bmatrix}_{M\times M}$$

is square, upper triangular, and nonsingular.

With this, we may write

$$A = Q^T R,$$

and define a g-inverse of A as

$$A^+ = \left[\begin{array}{c|c} R_1^{-1} & 0 \end{array}\right] Q. \tag{94}$$

Recalling that $P_A = A\,A^+$, the projection operator becomeas

$$P_A = Q^T \left[\begin{array}{c} R_1 \\ \hline 0 \end{array}\right] \left[\begin{array}{c|c} R_1^{-1} & 0 \end{array}\right] Q$$

$$= Q^T \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array}\right] Q \tag{95}$$

where I_M is the $M \times M$ identity matrix. We see by inspection that this is symmetric, and by squaring we see that it is idempotent,

$$P_A{}^2 = Q^T \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array}\right] Q\,Q^T \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array}\right] Q$$

$$= Q^T \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array}\right] Q = P_A,$$

where we have used the orthogonality of matrix Q. Thus we see that the g-inverse defined in (94) is adequate for forming the projection operator onto the column space of A.

Now recall the projection operator onto the row space of A,

$$_{A}P = A^{+} A$$

Substituting (94) into this equation yields

$$_{A}P = \left[R_1^{-1} \;\middle|\; 0 \right] Q \, Q^T \left[\begin{array}{c} R_1 \\ \hline 0 \end{array} \right]$$

$$= \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array} \right].$$

Thus this factorization satisfies the requirements for the Moore-Penrose g-inverse. The least-squares functional for the linear least-squares problem $A \, \underline{x} \simeq \underline{b}$ then becomes

$$\min_{\underline{x}} \phi(\underline{x}) = \min_{\underline{x}} \left|\left| \, A \, \underline{x} - \underline{b} \, \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| \, Q \, A \, \underline{x} - Q \, \underline{b} \, \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| \, R \, \underline{x} - Q \, \underline{b} \, \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| \left[\begin{array}{c} R_1 \\ \hline 0 \end{array} \right] \underline{x} - \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array} \right] \underline{b} \, \right|\right|^2. \qquad (96)$$

Here, Q has been partitioned as $Q = \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array} \right] \begin{array}{l} \} \, M \\ \} \, N\text{-}M \end{array}$. The solution for $\underline{x}$ in

in this case is determined uniquely as $\underline{x}_{LS} = R_1^{-1} \, Q_1 \, \underline{b}$, (97)

leaving a residual sum of squared error

$$\phi(\underline{x}_{LS}) = \left|\left| \, Q_2 \, \underline{b} \, \right|\right|^2. \qquad (98)$$

4.7 QR Factorization of Rank Deficient Matrices

In the case of rank deficient matrices, the QR factorization does not lead to a g-inverse which satisfies the Moore-Penrose conditions. In this section, it is shown that a truncated version of the QR factorization with column pivotting can, however, be used to construct a g-inverse suitable for forming projection operators. In the next section, the complete orthogonal factorization will be presented, which solves the problem of rank degeneracy and does lead to the Moore-Penrose pseudoinverse.

Consider the $N \times M$ matrix A of rank $r < M \leq N$. There exists an $N \times N$ orthogonal matrix Q, and an $M \times M$ permutation matrix S, such that

$$Q\ A\ S = R \equiv \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & 0 \end{array}\right] \tag{99}$$

where

$$R_{11} = \begin{bmatrix} \diagdown & /\!/ \\ 0 & \diagdown \end{bmatrix}_{r \times r} .$$

By truncating R, i.e. replacing R_{12} by a zero matrix, a g-inverse of A is

$$A^{+} = S \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q . \tag{100}$$

For this factorization, the projector onto the range of A becomes

$$P_A = Q^T \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & 0 \end{array}\right] S^T\ S \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q$$

$$= Q^T \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array}\right] Q \tag{101}$$

which, as in the full rank case, conforms to the requirements of a projection operator, thus this g-inverse is suitable if the formation of the projector for the column space is all that is required. As we shall now see, however, the product $A^+ A$ does not form a suitable projection operator with this g-inverse.

$$A^+ A = S \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q\, Q^T \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & 0 \end{array}\right] S^T$$

$$= S \left[\begin{array}{c|c} I_r & R_{11}^{-1} R_{12} \\ \hline 0 & 0 \end{array}\right] S^T \qquad (102)$$

While this is not symmetric and therefore cannot be used as a projector, it is interesting to note that this factorization does satisfy the third requirement of the Moore-Penrose pseudoinverse, namely that

$$A^+ A A^+ = A.$$

$$A^+ A A^+ = {}_A P\, A^+ = S \left[\begin{array}{c|c} I_r & R_{11}^{-1} R_{12} \\ \hline 0 & 0 \end{array}\right] S^T S \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q$$

$$= S \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q = A^+.$$

Note also that this g-inverse satisfied all of the conditions for forming the derivative of the projection operator in Chapter 2, thus it is suitable for use in minimizing the variable projection functional even when the basis function matrix is rank deficient.

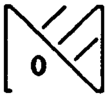
4.8 Complete Orthogonal Factorization of Rank Deficient Matrices

In this section, it is shown that an extension of the QR factorization, the complete orthogonal factorization, is a suitable alternative to the singular value decomposition for performing the rank reduction necessary to obtain the minimum norm linear least-squares solution.

Again consider the N × M matrix A with rank $r < M \leq N$ and the orthogonal factorization

$$Q\ A\ S = R \equiv \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & 0 \end{array}\right]$$

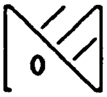
where

$R_{11} =$  $_{r \times r}$.

There exists an M × M orthogonal matrix V such that

$$R\ V = Q\ A\ S\ V = R' \equiv \left[\begin{array}{c|c} R_{11}' & 0 \\ \hline 0 & 0 \end{array}\right] \tag{103}$$

where

$R_{11}' =$  $_{r \times r}$.

From this, we may write $A = Q^T R' V^T S^T$, and define the g-inverse

$$A^+ = S\ V \left[\begin{array}{c|c} R_{11}'^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q \tag{104}$$

Now forming the projection operator P_A, we obtain

$$P_A = Q^T \left[\begin{array}{c|c} R_{11}' & 0 \\ \hline 0 & 0 \end{array}\right] V^T S^T\ S\ V \left[\begin{array}{c|c} R_{11}'^{-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q$$

$$= Q^T \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] Q \tag{105}$$

which is, once again, seen to be symmetric and idempotent. If we now attempt to form the projector onto the row space of A, we get

$$A^P = S\ V \left[\begin{array}{c|c} R_{11}^{'\,-1} & 0 \\ \hline 0 & 0 \end{array}\right] Q\ Q^T \left[\begin{array}{c|c} R_{11}^{'} & 0 \\ \hline 0 & 0 \end{array}\right] V^T\ S^T$$

$$= S\ V \left[\begin{array}{c|c} I_r & 0 \\ \hline 0 & 0 \end{array}\right] V^T S^T \qquad (106)$$

which is symmetric and idempotent. Thus the complete orthogonal factorization leads to a g-inverse which is suitable for forming both projection operators and minimum norm least-squares solutions.

With this g-inverse, the least-squares functional becomes

$$\min_{\underline{x}} \phi(\underline{x}) = \min_{\underline{x}} \left|\left| A\ \underline{x} - \underline{b} \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| Q\ A\ \underline{x} - Q\ \underline{b} \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| Q\ A\ S\ S^T\ \underline{x} - Q\ \underline{b} \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| Q\ A\ S\ V\ V^T S^T\ \underline{x} - Q\ \underline{b} \right|\right|^2$$

$$= \min_{\underline{x}} \left|\left| R' V^T S^T\ \underline{x} - Q\ \underline{b} \right|\right|^2.$$

If we let $\underline{y} = S^T\ \underline{x}$, partition Q as $Q = \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array}\right]\begin{array}{l} \} \ r \\ \} \ N\text{-}r \end{array}$, and partition V as $V = [V_1 \mid V_2\ ]$, we then obtain

$$\min_{\underline{y}} \phi(\underline{y}) = \min_{\underline{y}} \left|\left| \left[\begin{array}{c|c} R_{11}^{'} & 0 \\ \hline 0 & 0 \end{array}\right] \left[\begin{array}{c} V_1^{\ T} \\ \hline V_2^{\ T} \end{array}\right] \underline{y} - \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array}\right] \underline{b} \right|\right|^2$$

$$= \min_{\underline{y}} \left|\left| \left[\begin{array}{c} R_{11}^{'}\ V_1^{\ T} \\ \hline 0 \end{array}\right] \underline{y} - \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array}\right] \underline{b} \right|\right|^2$$

from which, if we partition $\underline{y}$ as $\left[\ \underline{y}_1^{\ T} \mid \underline{y}_2^{\ T}\ \right]^T$, we obtain the solution

$$\underline{y}_1 = V_1\ R_{11}^{'\,-1}\ Q_1\ \underline{b}. \qquad (107)$$

We can now obtain a minimum norm solution by letting $\underline{y}_2 = 0$, yielding

$$\underline{x}_{LS} = S \left[\begin{array}{c} V_1 \, R_{11}'^{\,-1} \, Q_1 \, \underline{b} \\ \hline 0 \end{array} \right], \tag{108}$$

and, as in the full rank case, leaving a residual sum of squares

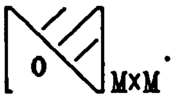
$$\phi(\underline{x}_{LS}) = \left|\left| \, Q_2 \, \underline{b} \, \right|\right|^2. \tag{109}$$

4.9 Near Rank Deficient Matricies

Consider the N × M matrix A with numerical rank $\rho = M \leq N$, but whose expected (ideal rank) is $r \leq M$. There exists an N × N orthogonal matrix, Q, and an M × M permutation matrix, S, such that

$$Q \; A \; S = R \equiv \left[\begin{array}{c} R_1 \\ \hline 0 \end{array} \right] \tag{110}$$

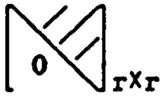
where

$R_1 =$ 

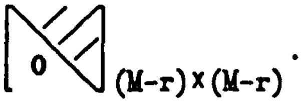
Column pivotting at each stage of the factorization will result in a matrix which can be further partitioned as

$$R_1 = \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & R_{22} \end{array} \right],$$

where

$R_{11} =$ 

and

$R_{22} =$ 

If A were truly rank deficient, R_{22} would consist of zeros. But because of perturbations in A, R_{22} will have non-zero elements. If the perturbations are small, however, then the elements of R_{22} should also be small, so that the rank deficiency can be uncovered when $||R_{22}||$ becomes much smaller than $||A||$. Then rank reduction can be achieved by setting R_{22} to zero and solving the remainder of the problem as a truly rank deficient case.

Golub and Van Loan [16] point out that there are cases in which at no step during the orthogonalization procefss is the norm of R_{22} very small, even though the oriuginal matrix is rank deficient. But they also go on to say that this method of rank deternmination "works well in practice." The reader is referred to Section 6.4 of Golub and Van Loan [16], and to Golub, Klema, and Stewart [18].

5. ALGORITHM

The purpose of this chapter is to develop an algorithm for the maximum likelihood parameter estimation technique which utilizes the variable projection method. Following Bard [5], Sections 5.1 through 5.4 review the gradient methods for iterative minimization. This review will culminate in a discussion of the Gauss-Newton method applied to minimization of a squared norm function (inclusive of the least-squares and the variable projection functionals). Section 5.5 will then utilize the results of Chapter 3 to formulate the Gauss-Newton step for the Variable Projection Nonlinear Least-squares method. Following this, a simplification to the algorithm noted by Kaufman [19] will be reviewed. Finally, Marquardt's modification to the Gauss-Newton step will be discussed.

5.1 Iterative Minimization Techniques

Given an objective functional $\phi(\underline{\theta})$ of the vector of parameters

$$\underline{\theta} = \left[\, \theta_1, \, \theta_2, \, \ldots, \, \theta_K \, \right]^T,$$

we wish to determine the values for $\underline{\theta}$ such $\phi(\underline{\theta})$ is minimized. Iterative minimization techniques [5] generate a sequences of vectors $\underline{\theta}_i$, i=1,2,... which hopefully converge to the true minimum of the objective function. The vector $\underline{\theta}_i$ is called the i'th iterate.

Let us define

$$\begin{aligned} \underline{\Delta}_i &= \underline{\theta}_{i+1} - \underline{\theta}_i \\ &= \text{the i'th update step} \end{aligned} \qquad (111)$$

and

$$\phi_i = \phi(\underline{\theta}_i).$$

<u>Definition</u>: The i'th iterate is <u>acceptable</u> if $\phi_{i+1} < \phi_i$, that is, if the addition of the i'th update step to the i'th iterate causes a decrease in the value of the objective function.

Each iteration consists of determining

(1) a vector $\underline{d}_i$ in the direction of the i'th update step, and

(2) a scalr ρ_i, such that the step $\underline{\Delta}_i = \rho_i\ \underline{d}_i$ produces an acceptable iteration. Thus we require that $\phi(\underline{\theta}_i+\underline{\Delta}_i) < \phi(\underline{\theta}_i)$.

5.2 Gradient Methods for Determining Step Direction

During the i'th iteration, we strike out from $\underline{\theta}_i$ along a direction $\underline{d}$ generating the ray

$$\underline{\theta}(\rho) = \underline{\theta}_i+\rho\underline{d}. \tag{112}$$

Here we have noted that, when confined to this ray, $\underline{\theta}$, and hence $\phi(\underline{\theta})$, are functions of ρ alone. We may now define the confined objective function as

$$\Psi_{i\underline{d}}(\rho) = \phi\Big[\ \underline{\theta}(\rho)\ \Big] = \phi(\underline{\theta}_i+\ \rho\ \underline{d}). \tag{113}$$

Differentiating this with respect to ρ yields

$$\frac{\partial\Psi_{i\underline{d}}}{\partial\rho} = \left[\frac{\partial\phi}{\partial\underline{\theta}}\right]^T \left[\frac{\partial\underline{\theta}}{\theta\rho}\right] = \left[\frac{\partial\phi}{\partial\underline{\theta}}\right]^T \underline{d}, \tag{114}$$

and evaluating at ρ=0 yields the directional derivative of ϕ relative to $\underline{d}$ at $\underline{\theta}_i$, defined as

$$\Psi_{i\underline{d}}' = \frac{\partial\Psi_{i\underline{d}}}{\partial\rho}\Bigg|_{\rho=0} = g_i^{\ T}\ \underline{d}. \tag{115}$$

Here, g_i is the gradient of ϕ evaluated at $\underline{\theta}_i$,

$$g_i = \frac{\partial\phi(\underline{\theta})}{\partial\underline{\theta}}\bigg|_{\underline{\theta}=\underline{\theta}_i} . \tag{116}$$

A small positive value of ρ is guaranteed to produce a step which decreases the value of the objective function if the directional derivative at $\underline{\theta}_i$ is negative. Thus we may define $\underline{d}$ as an acceptable direction if $\Psi' < 0$. This simply states that $\underline{d}$ is a downhill direction on the contour of Ψ if it forms a greater than 90° angle with the gradient at $\underline{\theta}_i$.

One obvious choice of direction for the i'th iterate is simply

$$\underline{d} = -g \tag{117}$$

This is the direction used for all iterations in the steepest descent method, named for the fact that this is the direction in which the objective function initially decreases most rapidly. Unfortunately, this often produces steps which zigzag back and forth down the contour, leading to extremely slow convergence.

As an alternative, we may find an acceptable direction by finding a suitable positive definite matrix R, and defining

$$\underline{d}_i = - R\, g_i . \tag{118}$$

The acceptability of this direction follows from the definition of positive definiteness as follows:

$$\Psi_{i\underline{d}}{}' = g_i^T\, \underline{d}_i = - g_i^T R\, g_i < 0 \tag{119}$$

Minimization techniques in which directions are obtained in this way are called gradient methods. If the positive definiteness of R is strictly adhered to, then the method is called an acceptable gradient method.

5.3 Newton's Method

It is well known from single variable calculus that a zero in a function $g(x)$ can be found iteratively using the first order Taylor approximation around a point x_i, given by

$$g(x) \simeq g(x_i) + [g'(x_i)](x-x_i).$$

This approximation can be extended to finding a local minimum of a function f(x) by letting $g(x) = f'(x)$,

$$f'(x) \simeq f'(x_i) + [f''(x_i)](x-x_i).$$

Equating this to zero and rearranging, the i'th iteration becomes

$$x_{i+1} = x_i - [f''(x_i)]^{-1} f'(x_i)$$

provided $f''(x_i) \neq 0$.

Extending this to multivariable calculus, we obtain the update relation

$$\underline{\theta}_{i+1} = \underline{\theta}_i - H_i^{-1} g_i \qquad (120)$$

where $\underline{\theta}$ and $\underline{g}$ are as defined in the previous section and $H_i = H(\underline{\theta}_i)$ is the Hessian matrix evaluated at $\underline{\theta}_i$. The Hessian matrix is defined by

$$\left\{ H(\underline{\theta}) \right\}_{mn} = \frac{\partial\phi(\underline{\theta})}{\partial\theta_m \, \partial\theta_n} . \qquad (121)$$

Note here that H_i must be nonsingular.

Newton's method may be alternatively viewed as a second order Taylor series approximation to the original function, given by

$$\hat{\phi}(\underline{\theta}) = \phi(\underline{\theta}_i) + g_i^T \left[\underline{\theta}-\underline{\theta}_i \right] + \frac{1}{2} \left[\underline{\theta}-\underline{\theta}_i \right]^T H_i \left[\underline{\theta}-\underline{\theta}_i \right] \qquad (122)$$

which is the best second order approximation to the original function. Differentiating, we obtain

$$\frac{\partial\hat{\phi}(\underline{\theta})}{\partial\underline{\theta}} = \underline{g}_i + H_i \left[\underline{\theta}-\underline{\theta}_i \right] \qquad (123)$$

which, when set to zero yields the recursion

$$\underline{\theta}_{i+1} = \underline{\theta}_i - H^{-1}\, \underline{g}_i . \qquad (124)$$

This relation satisfies the general formula for a gradient method iteration with ρ_i=1 and R_i=H^{-1}. If H is positive definite, then H^{-1} will also be positive definite, and Newton's method will produce acceptable iterations. Furthermore, if the objective function is quadratic, then $\hat{\phi}(\underline{\theta}) = \phi(\underline{\theta})$ and Newton's method will converge in a single iteration.

In order to avoid calculating second derivatives, one may use the Gauss approximation to Newton's method, or the Gauss-Newton method, which requires only the evaluation of first derivatives. This will be brought to light in the next section.

5.4 Gauss-Newton Method

Consider an objective function of the form

$$\phi(\underline{\theta}) = \left|\left| \; \underline{e}(\underline{\theta}) \; \right|\right|^2 \qquad (125)$$

$$= \sum_{j=1}^{N} e_j^{\,2} \qquad (126)$$

Among others, this form includes the least-squares and variable projection functionals.

Differentiating with respect to the m'th component of the parameter vector (and dropping the iteration index for convenience) yields the m'th component of the gradient vector

$$q_m = \frac{\partial\phi(\underline{\theta})}{\partial\theta_m} = \sum_{j=1}^{N} e_j \left[\frac{\partial e_j}{\partial\theta_m} \right] \qquad (127)$$

Now differentiating this with respect to the n'th component of the parameter vector yields the typical component of the Hessian matrix

$$\left\{ H(\underline{\theta}) \right\}_{mn} = 2 \sum_{j=1}^{N} \left[\frac{\partial e_j}{\partial \theta_m} \right] \left[\frac{\partial e_j}{\partial \theta_n} \right] + 2 \sum_{j=1}^{N} e_j \left[\frac{\partial^2 e_j}{\partial \theta_m \, \partial \theta_n} \right] \qquad (128)$$

Near a minimum of $\phi(\underline{\theta})$, the error e_j will be small and will make the second term above negligible compared to the first. It is by neglecting this term that we obtain the Gauss approximation to the Hessian matrix, given as

$$\left\{ N(\underline{\theta}) \right\}_{mn} = 2 \sum_{j=1}^{N} \left[\frac{\partial e_j}{\partial \theta_m} \right] \left[\frac{\partial e_j}{\partial \theta_n} \right]. \qquad (129)$$

Let us now define the cost function derivative matrix

$$B = \left[\frac{\partial \underline{e}}{\partial \theta_1} \;\middle|\; \frac{\partial \underline{e}}{\partial \theta_1} \;\middle|\; \cdots \;\middle|\; \frac{\partial \underline{e}}{\partial \theta_K} \right]. \qquad (130)$$

Having thus defined this derivative matrix, we may now rewrite both the gradient vector and the Gauss approximation to the Hessian matrix in terms of the derivative matrix as

$$\underline{g} = 2 \, B^T \, \underline{e} \qquad (131)$$

and

$$N = 2 \, B^T \, B \qquad (132)$$

If we now substitute the Gauss approximation to the Hessian matrix into the gradient method equation for the step direction, i.e. $R = N^{-1}$, we obtain

$$\underline{d} = - \, N^{-1} \, \underline{g},$$

or

$$N \, \underline{d} = - \, \underline{g}. \qquad (133)$$

Now substituting in (131) and (132) above, we get

$$B^T \, B \, \underline{d} = - \, B^T \, \underline{e}. \qquad (134)$$

But these are just the normal equations for the linear least-squares problem in which the error vector is projected onto the range of the derivative matrix to obtain the vector $\underline{d}$. Thus, the solution for $\underline{d}$ at each iteration is simply

$$\underline{d} = - B^{\#} \, \underline{e}, \tag{135}$$

where $B^{\#}$ is a Moore-Penrose pseudoinverse of B.

5.5 Variable Projection Nonlinear Least-Squares Gauss-Newton Iteration

This section will combine the results of the present chapter and those of the previous two chapters to devise an algorithm for the Variable Projection method. Recall that the Gauss-Newton step is obtained from

$$\underline{d} = - B^{\#} \, \underline{e},$$

where

$$B = \left[\frac{\partial \underline{e}}{\partial \theta_1} \;\middle|\; \frac{\partial \underline{e}}{\partial \theta_1} \;\middle|\; \cdots \;\middle|\; \frac{\partial \underline{e}}{\partial \theta_K} \right] = \frac{\partial \underline{e}}{\partial \underline{\theta}^T}$$

From Chapter 3, equation (53), we have

$$\underline{e} = P_F^{\perp} \, \underline{y}.$$

Thus

$$\begin{aligned} B = \frac{\partial \left(P_F^{\perp} \underline{y} \right)}{\partial \underline{\theta}^T} &= \left[\frac{\partial P_F^{\perp}}{\partial \underline{\theta}^T} \right] \underline{y} \\ &= D\left(P_F^{\perp} \right) \underline{y} \\ &= - D\left(P_F \right) \underline{y} \end{aligned} \tag{136}$$

Substituting (63) for the Frechet derivative of the projection operator yields

$$B = - P_F^{\perp} D(F) \, F^{+} \underline{y} - \left[P_F^{\perp} D(F) \, F^{+} \right]^T \underline{y}. \tag{137}$$

Now recall, from Section 4.7, the factorization of the N × M matrix F given by

$$Q \, F \, S = \left[\begin{array}{c|c} R_{11} & R_{12} \\ \hline 0 & 0 \end{array} \right]$$

where Q is an N × N orthogonal matrix and S is an M × M permutation matrix. From this we had defined the g-inverse of F

$$F^{+} = S \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array} \right] Q$$

The projection operator onto the column space of F is then

$$P_F = Q^T \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array} \right] Q,$$

from which we can define the projector onto the orthogonal complement of F in $\mathbb{R}^N$

$$P_F^{\perp} = Q^T \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I_{N-M} \end{array} \right] Q.$$

To simplify notation, let us define

$$I_1 = \left[\begin{array}{c|c} I_M & 0 \\ \hline 0 & 0 \end{array} \right], \quad I_2 = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & I_{N-M} \end{array} \right], \text{ and } \quad \tilde{R}_1^{-1} = \left[\begin{array}{c|c} R_{11}^{-1} & 0 \\ \hline 0 & 0 \end{array} \right].$$

With these definitions, we can write the g-inverse and the projection operators, respectively, as

$$F^{+} = S \, \tilde{R}_1^{-1} \, Q, \tag{139}$$

$$P_F = Q^T \, I_1 \, Q, \tag{140}$$

and

$$P_F^{\perp} = Q^T \, I_2 \, Q. \tag{141}$$

Substituting these definitions into (137) yields

$$B = - Q^T I_2 Q D(F) S \tilde{R}_1^{-1} Q \underline{y} - \left[Q^T I_2 Q D(F) S \tilde{R}_1^{-1} Q \right]^T \underline{y}. \qquad (142)$$

This equation can be regrouped as follows to demonstrate how one might implement the formation of the matrix. First, taking the transpose of the second term yields

$$\left[Q^T I_2 Q D(F) S \tilde{R}_1^{-1} Q \right]^T \underline{y} = Q^T \left[\tilde{R}_1^{-1} \right]^T S^T \left[Q D(F) \right]^T I_2 Q \underline{y}$$

If we now let $\underline{v} = Q \underline{y}$ and $C = Q D(F)$, and $\underline{x} = F^+ \underline{y}$ we obtain

$$B = - Q^T \left\{ I_2 C \underline{x} + \left[\tilde{R}_1^{-1} \right]^T S^T C^T I_2 \underline{v} \right\}.$$

5.6 Kaufman's Variable Projection Algorithm

A much simpler version of the projection operator derivative, and thus a simpler version of (142), was derived by Kaufman [19]. By exploiting the structure of the projection operator and the isometric properties of the orthogonal matrix Q, Kaufman has shown that the second term on the right hand side of (142) can effectively be ignored.

Noting that Q has orthonormal columns, we know that

$$\begin{aligned} \left\| P_F^{\perp} \underline{y} \right\| &= \left\| Q P_F^{\perp} \underline{y} \right\| \qquad (143) \\ &= \left\| Q Q^T I_2 Q \underline{y} \right\| \\ &= \left\| I_2 Q \underline{y} \right\|. \end{aligned}$$

Following Kaufman, we can define the new objective function

$$\phi_3(\underline{\theta}) = \left\| Q_2 \underline{y} \right\|^2 \qquad (144)$$

where we have partitioned Q as

$$Q = \left[\begin{array}{c} Q_1 \\ \hline Q_2 \end{array}\right] \begin{array}{l} \}\ M \\ \}\ N\text{-}M \end{array}$$

While the derivative of Q_2 is dependent upon the orthogonalization process in which the matrix Q is determined, and is therefore not unique, Kaufman derives the following general formula whose results, though nonunique, are similar "within an orthogonal transformation":

$$D(Q_2) = -\ Q_2\ D(F)\ S_1\ \tilde{R}_1^{-1}\ Q\ +\ Z\ Q \tag{145}$$

where

$$Z^T + Z = 0. \tag{146}$$

Since the matrix Z is not unique, neither is $D(Q_2)$. We can, however, choose $Z = 0$, which certainly satisfies (146), leaving

$$B = D(Q_2)\ \underline{y} = -\ Q_2\ D(F)\ S_1\ \tilde{R}_1^{-1}\ Q\ \underline{y}\ , \tag{147}$$

which is the same result derived by Golub and Pereyra with the modified projection operator and the second term in (142) disregarded.

With this definition for the derivative, the Gauss-Newton step direction becomes

$$\underline{d} = \left[\ Q_2\ D(F)\ S_1\ \tilde{R}_1^{-1}\ Q\ \underline{y}\ \right]^+\ Q_2\ \underline{y}. \tag{148}$$

where we have partitioned S as

$$S = \left[\ S_1\ \middle|\ S_2\ \right].$$

5.7 Marquardt's Modification

Recall that for a gradient method to produce acceptable steps, the matrix R_i has to be positive definite. Neither Newton's method nor the Gauss approximation to it ensure that R_i would be positive definite, so that they cannot be considered, in their original form, acceptable gradient methods. Noting an observation by Marquardt [20], we may force the R_i matrix in both cases to be positive definite, thus making the methods acceptable.

For some positive definite matrix, P, any matrix A can be made positive definite by adding λP, provided that the positive scalar λ is large enough. Thus we can ensure that an iteration produces an acceptable direction by letting

$$R_i = \left[A_i + \lambda_i P_i \right]^{-1} \tag{149}$$

where A_i is H_i, N_i, or some other appropriate matrix.

Several choices are available for the matrix P_i. In partricular, suppose that P_i is diagonal. We may then define the diagonal matrix G with elements

$$g_{jj} = \left\{ P_i \right\}_{jj}^{1/2}. \tag{150}$$

With this choice of G, we may write R_i as

$$R_i = \left[A_i + \lambda_i G_i^T G_i \right]^{-1}. \tag{151}$$

If we focus specifically on the Gauss method, then the equation for the step direction with Marquardt's modification is

$$\left[B_i^T B_i + \lambda_i G_i^T G_i \right] \underline{d} = - B_i^T \underline{e}_i . \tag{152}$$

This can be solved using the Cholesky factorization for symmetric matrices, or we can note that these are simply the normal equations for the linear least-squares problem

$$B_i^T B_i \underline{d} + \lambda_i G_i^T G_i \ \underline{d} = - B_i^T \underline{e}_i + \lambda_i^{1/2} G_i^T \underline{0}$$

where we have added zero to the right hand side. The i'th step direction can be calculated using the QR factorization for the linear least-squares problem

$$\begin{bmatrix} B_i \\ \hline \lambda_i^{1/2}\, G_i \end{bmatrix} \underline{d} \simeq - \begin{bmatrix} \underline{e}_i \\ \hline \underline{0} \end{bmatrix} \tag{153}$$

whose solution is

$$\underline{d} = - \begin{bmatrix} B_i \\ \hline \lambda_i^{1/2}\, G_i \end{bmatrix}^{\#} \begin{bmatrix} \underline{e}_i \\ \hline \underline{0} \end{bmatrix}. \tag{154}$$

5.8 Step-Size Determination

Upon determination of the step direction, the optimum, or near optimum, step-size is determined using a line search along the given direction. Essentially the step-size which yields the minimum residual sum of squares is found by increasing the step-size (with large increments) until the steps cease to cause a decrease in the residual, then small decrements are taken until the actual optimum value is found (when the steps again cease to yield a decrease in residual sum of squares).

6. RESULTS AND CONCLUSIONS

Monte Carlo simulations were run to test the effectiveness of the algoritihm outlined in this report. This chapter will present the transducer model used in the computer simulations, the results obtained, and conclusions based on those results.

6.1 Transducer Model

The transducer model used was a two-pole high-pass filter excited by stepped sinusoid (Model 1). The Laplace transform of the theoretical response is

$$Y_1(s) = \frac{\omega_0}{s^2+\omega_0^2} H_1(s) \qquad , \omega_0 = 2\pi f_0$$

where

$$H_1(s) = \frac{s^2}{s^2+2\varsigma_1\omega_1 s+\omega_{C1}^2} \qquad , \omega_{C1} = 2\pi f_{C1}$$

The peak of the transfer function magnitude $|H(j\omega)|$ occurs at

$$f_{m1} = f_{C1}(1-2\varsigma_1^2)^{-1/2}$$

The Q (quality factor) is the ratio of the peak response frequency to the 3 dB bandwidth and is approximately

$$Q_1 \simeq 1/2\varsigma_1 - 2\varsigma_1 - 3\varsigma_1^3 - 9\varsigma_1^5.$$

This model can be seen as either the acoustic signal from a projector modeled as a 2-pole high-pass or as the electrical signal seen at the output of a hydrophone when the acoustic signal is an ideal stepped sinusoid.

The s-plane pole parameters for the decaying component of the transient are

$$\alpha_1 = 2\pi f_{C1}\, \zeta_1$$

and

$$f_1 = f_{C1}\, (1-\zeta_1^2)^{1/2}$$

The exact time function, y(t), is

$$y_1(t) = A_{0_C} \cos(2\pi f_0 t) + A_{0_S} \sin(2\pi f_0 t)$$

$$+ A_{1_C}\, e^{-\alpha t} \cos(2\pi f_1 t) + A_{1_S}\, e^{-\alpha t} \sin(2\pi f_1 t)$$

where

$$A_{0_C} = \frac{2\zeta_1 (f_0/f_{C1})^3}{\{(2\zeta_1 f_0/f_{C1})^2 + [(f_0/f_{C1})^2 - 1]^2\}} \,,$$

$$A_{0_S} = \frac{(f_0/f_{C1})^2\, [(f_0/f_{C1})^2 - 1]}{\{(2\zeta_1 f_0/f_{C1})^2 + [(f_0/f_{C1})^2 - 1]^2\}} \,,$$

$$A_{1_C} = \frac{2\zeta_1 (f_0/f_{C1})^3}{\{4\zeta_1^{\,2}(1-\zeta_1^{\,2}) + [(f_0/f_{C1})^2 + 2\zeta_1^{\,2} - 1]^2\}} \,,$$

and

$$A_{1_S} = \frac{(f_0/f_{C1})[(f_0/f_{C1})^2(1-2\zeta_1^{\,2})-1]}{\sqrt{1-\zeta_1^{\,2}}\ \{4\zeta_1^{\,2}(1-\zeta_1^{\,2}) + [(f_0/f_{C1})^2 + 2\zeta_1^{\,2} - 1]^2\}}$$

6.2 Simulation Results

All results were calculated based on 100 Monte Carlo trials. The signal-to-noise ratio used was defined in terms of the steady state amplitude as follows:

$$\mathrm{SNR(dB)} = 10 \log \left[\frac{A_0^{\,2}}{2\sigma^2} \right]$$

Tables 1, 2, and 3 give a numerical representation of the results for Q's of 4, 8, and 12, respectively. In each table, the bias, standard deviation, root mean square error, and Cramer-Rao bound are given for the steady-state amplitude (A_0) and the transient damping factor (α_1) and frequency (f_1) at each of the signal-to-noise ratios simulated.

Figures 3-5 give a graphical comparison of the results for this computer implementation of the variable projection nonlinear least-squares method to those of the principal component linear prediction method described in Chapter 2. For each parameter (A_0, α_1, and f_1), a plot of normalized mean square error vs. signal-to-noise ratio (MSE's are normalized to the C-R bound) for each Q (4, 8, and 12).

Table 1: Normalized Estimator Results for Parameter A_0

			Standard		
Q	SNR	Bias	Deviation	RMS Error	C-R Bound
		$\frac{\bar{A}_0 - A_0}{A_0}$	$\frac{\hat{\sigma}_{A_0}}{A_0}$	$\frac{\sqrt{MSE_{A_0}}}{A_0}$	$\frac{\sigma_{CR_{A_0}}}{A_0}$
4	25	3.1397E-02	7.8014E-02	7.7686E-02	5.7052E-02
4	27	3.8494E-03	5.5601E-02	5.5456E-02	4.5318E-02
4	30	3.3125E-03	3.8165E-02	3.8118E-02	3.2083E-02
4	33	2.4347E-03	2.6829E-02	2.6805E-02	2.2713E-02
4	37	1.7988E-03	1.7345E-02	1.7352E-02	1.4331E-02
4	40	1.0275E-03	1.1992E-02	1.1976E-02	1.0145E-02
4	50	2.6404E-04	3.7056E-03	3.6965e-03	3.2083E-03
4	60	8.3761E-05	1.1663E-03	1.1634E-03	1.0145E-03
8	25	-9.3672E-02	2.6597E-01	2.8072E-01	1.7476E-01
8	27	-1.1479E-03	2.2148E-01	2.2038E-01	1.3882E-01
8	30	1.5533E-02	1.1790E-01	1.1834E-01	9.8276E-02
8	33	1.0685E-02	7.8684E-02	7.9015E-02	6.9575E-02
8	37	6.8030E-03	4.6857E-02	4.7116E-02	4.3899E-02
8	40	5.2801E-03	3.2613E-02	3.2877E-02	3.1078E-02
8	50	1.7833E-03	9.7773E-03	9.8904E-03	9.8276E-03
8	60	4.8134E-04	3.2679E-03	3.2869E-03	3.1078E-03
12	25	-3.0256E-01	4.0855E-01	5.0674E-01	3.7311E-01
12	27	1.5185E-01	4.0974E-01	4.3505E-01	2.9637E-01
12	30	7.8057E-02	6.0740E-01	6.0938E-01	2.0981E-01
12	33	3.8631E-02	2.0769E-01	2.1023E-01	1.4854E-01
12	37	1.8663E-02	1.0343E-01	1.0459E-01	9.3721E-02
12	40	1.4054E-02	6.9019E-02	7.0096E-02	6.6349E-02
12	50	4.0782E-03	2.1160E-02	2.1445E-02	2.0981E-02
12	60	1.2375E-03	6.5496E-03	6.6333E-03	6.6349E-03

Table 2: Estimator Results for Parameter α_1

			Standard		
Q	SNR	Bias	Deviation	RMS Error	C-R Bound
		$\bar{\alpha}_1 - \alpha_1$	$\hat{\sigma}_{\alpha_1}$	$\sqrt{MSE_{\alpha_1}}$	$\sigma_{CR_{\alpha_1}}$
4	25	-1.0488E-02	2.9812E-01	2.9681E-02	1.1262E-01
4	27	1.2867E-02	9.4643E-02	9.5044E-02	8.9461E-02
4	30	1.0796E-02	6.5190E-02	6.5755E-02	6.3334E-02
4	33	7.9022E-03	4.7570E-02	4.7987E-02	4.4836E-02
4	37	5.4118E-03	3.1725E-02	3.2026E-02	2.8290E-02
4	40	3.3593E-03	2.1771E-02	2.1921E-02	2.0028E-02
4	50	9.3497E-04	6.9761E-03	7.0038E-03	6.3334E-03
4	60	2.9473E-04	2.2050E-03	2.2246E-03	2.0028E-03
8	25	-2.2919E-01	5.3391E-01	5.7857E-01	1.2295E-01
8	27	-4.5185E-02	2.5344E-01	2.5619E-01	9.7661E-02
8	30	9.5538E-03	6.7355E-02	6.7695E-02	6.9138E-02
8	33	7.4555E-03	4.8321E-02	4.8654E-02	4.8946E-02
8	37	5.0330E-03	3.0256E-02	3.0522E-02	3.0883E-02
8	40	4.0504E-03	2.1383E-02	2.1658E-02	2.1863E-02
8	50	1.4110E-03	6.5945E-03	6.7114E-03	6.9138E-03
8	60	3.8313E-04	2.2454E-03	2.2667E-03	2.1863E-03
12	25	-5.6826E-01	7.4937E-01	9.3747E-01	1.4781E-01
12	27	-3.3185E-01	6.0141E-01	6.8425E-01	1.1741E-01
12	30	-2.7766E-02	2.2919E-01	2.2972E-01	8.3121E-02
12	33	9.7796E-03	5.7614E-02	5.8153E-02	5.8845E-02
12	37	6.5181E-03	3.5428E-02	3.5848E-02	3.7129E-02
12	40	5.2898E-03	2.4906e-02	2.5339e-02	2.6285E-02
12	50	1.6507e-03	8.1051e-03	8.2316e-03	8.3121E-03
12	60	5.1621e-04	2.5330e-03	2.5727e-03	2.6285E-03

Table 3: Estimator Results for Parameter f_1

Q	SNR	Bias	Standard Deviation	RMS Error	C-R Bound
		$\bar{f}_1 - f_1$	$\hat{\sigma}_{f_1}$	$\sqrt{MSE_{f_1}}$	$\sigma_{CR_{f_1}}$
4	25	-1.0416E-02	8.0899E-02	8.1165E-02	2.1509E-02
4	27	-2.6716E-03	1.7504E-02	1.7620E-02	1.7085E-02
4	30	-1.3759E-03	1.2783E-02	1.2793E-02	1.2095E-02
4	33	-9.6422E-04	9.0174E-03	9.0239E-03	8.5629E-03
4	37	-6.0650E-04	5.3263E-03	5.3342E-03	5.4028E-03
4	40	-4.2107E-04	3.7593E-03	3.7641E-03	3.8249E-03
4	50	-1.0745E-04	1.2411E-03	1.2396E-03	1.2095E-03
4	60	-3.1262E-05	3.8806E-04	3.8738E-04	3.8249E-04
8	25	-9.8704E-02	1.9825E-01	2.2057E-01	2.1502E-02
8	27	-2.9563E-02	1.1967E-01	1.5051E-01	1.7080E-02
8	30	-2.2884E-03	1.1726E-02	1.1890E-02	1.2091E-02
8	33	-1.4406E-03	8.4651E-03	8.5450E-03	8.5600E-03
8	37	-7.6443E-04	5.2576E-03	5.2868E-03	5.4010E-03
8	40	-6.2716E-04	3.6980E-03	3.7325E-03	3.8236E-03
8	50	-1.5717E-04	1.2063E-03	1.2105E-03	1.2091E-03
8	60	-4.1869E-05	3.8314E-04	3.3869E-04	3.8236E-04
12	25	-2.3453E-01	2.6917E-01	3.5599E-01	2.5071E-02
12	27	-1.3253E-01	2.2330E-01	2.5870E-01	1.9920E-02
12	30	-2.4989E-02	1.1681E-01	1.1888E-01	1.4102E-02
12	33	-1.7707E-03	9.5334E-03	9.6495E-03	9.9836E-03
12	37	-9.7337E-04	6.1725E-03	6.2183E-03	6.2992E-03
12	40	-7.3483E-04	4.3372E-03	4.3776E-03	4.4595E-03
12	50	-2.3559E-04	1.3686E-03	1.3820E-03	1.4102E-03
12	60	-7.4384E-05	4.1825E-04	4.2275E-04	4.4595R-04

Figure 3. MSE(alpha$_1$) rel Cramer-Rao Lower Bound

Excitation at Resonance of 2-pole high-pass

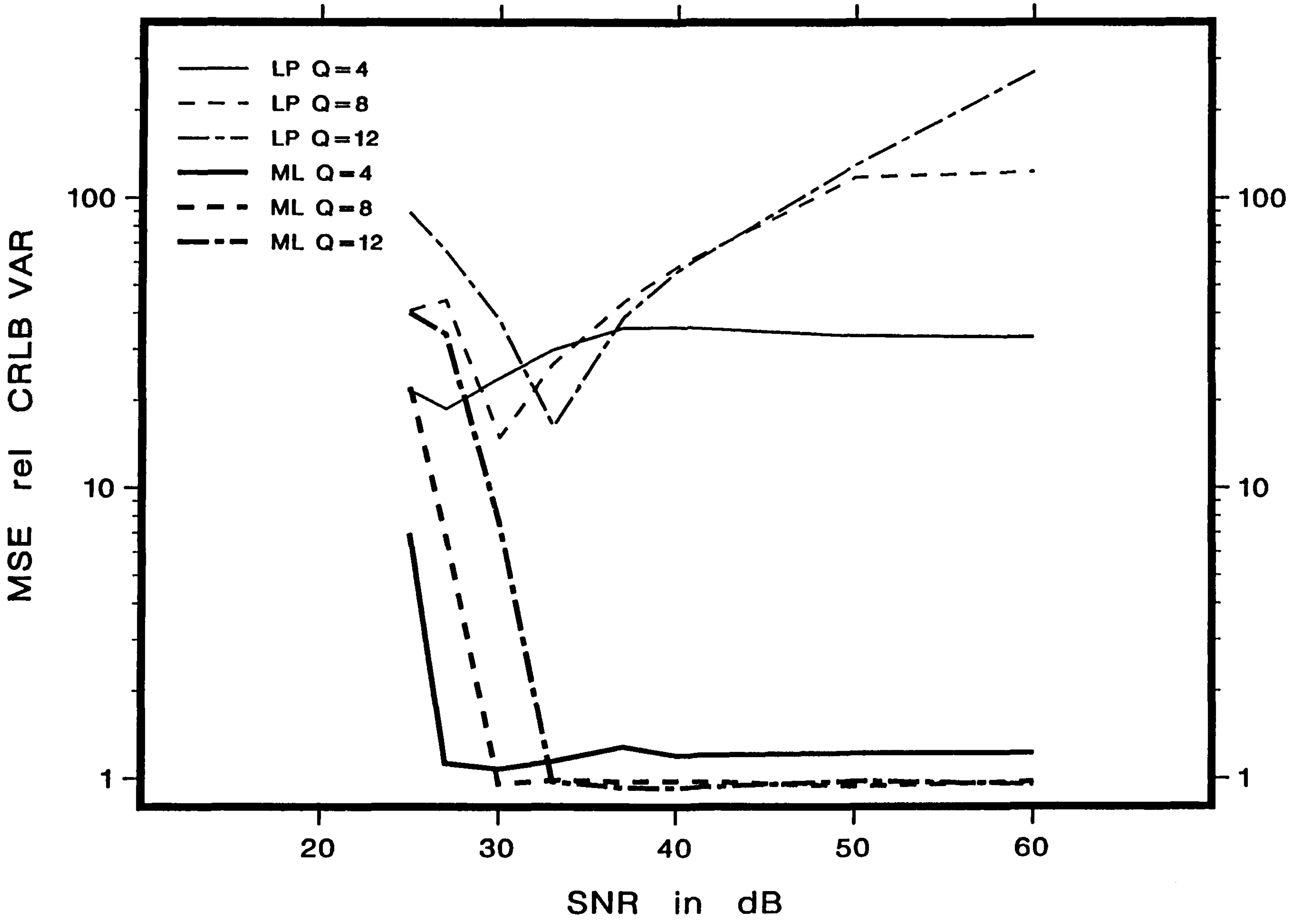


Figure 4. MSE(f_1) rel Cramer-Rao Lower Bound

Excitation at Resonance of 2-pole high-pass

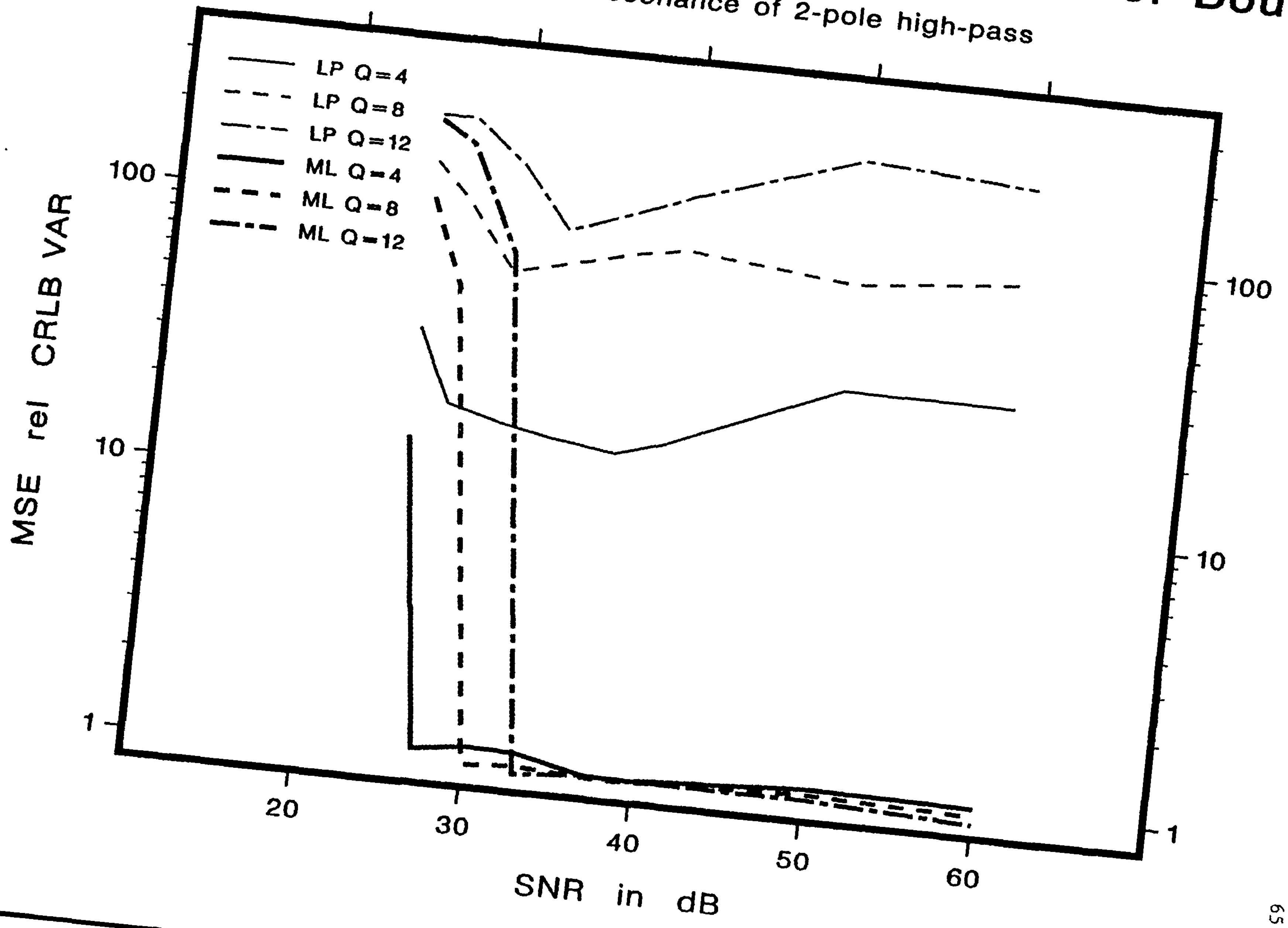
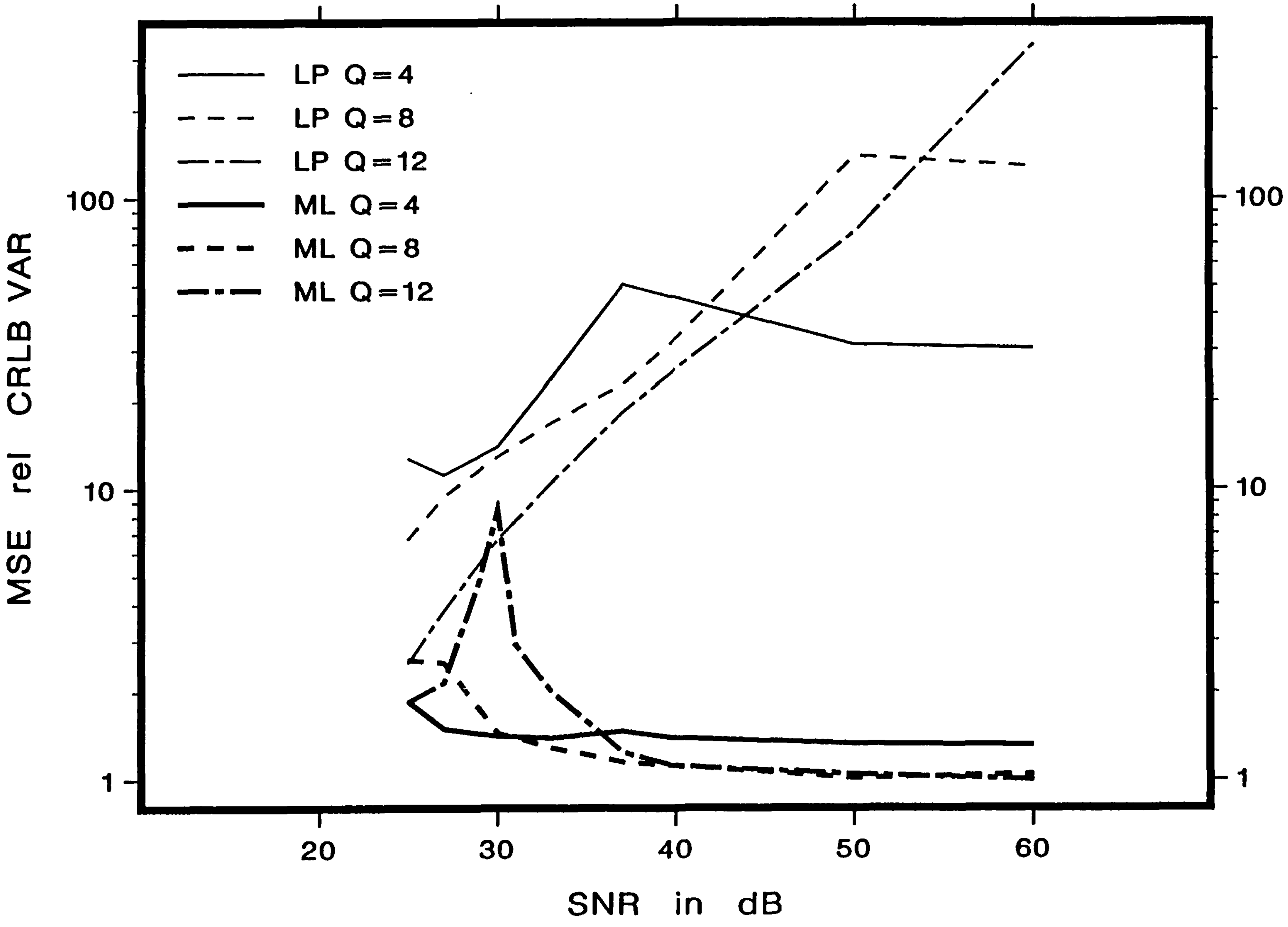


Figure 5. MSE(A_0) rel Cramer-Rao Lower Bound

Excitation at Resonance of 2-pole high-pass



6.3 Conclusions

The results given demonstrate that this computer implementation of the variable projection nonlinear least-squares algorithm exhibits maximum likelihood performance above a threshold signal-to-noise ratio which depends upon the Q of the model. Below the threshold SNR, this implementation departs from maximum likelihood performance. Judging from the behavior of the results and observations of the convergence behavior below the threshold SNR, it appears that the current implementation of the variable projection method is unable to converge when the initial estimate from the principal component linear prediction method is far removed from the true minimum of the variable projection functional (see Appendix B). Future work will examine alternative step-direction and step-size search methods which should improve the performance at these low signal-to-noise ratios.

REFERENCES

[1] J.D. George, V.K. Jain, and P.L. Ainsleigh, "Estimating Steady-State Response of a Resonant Transducer in a Reverberant Underwater Environment", IEEE ICASSP88, pp 2737-2740 (1988).

[2] R. Kumaresean and D.W. Tufts, "Estimating the Parameters of Exponentially Damped Sinusoids and Pole-zero Modeling in Noise", IEEE ASSP-30, pp 833-840 (1982).

[3] G.H. Golub and V. Pereyra, "The Differentiation of Pseudoinverses and Nonlinear Least Squares Problems whose Variables Separate", SIAM Journal Numerical Analysis 10, (1973).

[4] L.G. Beatty, J.D. George, and A.Z. Robinson, "Use of Complex Exponential Expansion as a Signal Representation for Underwater Acoustic Calibration", Journal of the Acoustical Society of America Vol 63(6), pp. 1782-1794 (1978).

[5] Y. Bard, Nonlinear Parameter Estimation, Academic Press, New York NY (1974).

[6] M.K. Kendall and A. Stuart, The Advanced Theory of Statistics, Charles Griffin and Company, Ltd., High Wycombe, England (1979).

[7] S.L. Marple, Jr., Digital Spectral Analysis with Applications, Prentice Hall, Englewood Cliffs NJ (1987).

[8] A.C. Kot, S. Parthasarathy, D.W. Tufts, and R.J. Vaccaro, "The Statistical Performance of State-Variable Balancing and Prony's Method in Parameter Estimation", ICASSP88, pp.1549-1552.

[9] G.W. Stewart, Introduction to Matrix Computations, Academic Press, New York NY (1973).

[10] D.C. Montgomery and E.A. Peck, Introduction to Linear Regression Analysis, John Wiley and Sons, New York NY (1982).

[11] S.J. Orfanidis, Optimum Signal Processing, Macmillan Publishing Co., New York NY (1985)

[12] R. Kumaresean, "On the Zeroes of the Linear Prediction Error Filter for Deterministic Signals", IEEE ASSP-31 pp 217-220 (1983).

[13] P.R. Halmos, Introduction to Hilbert Space, Chelsea Publishing Co., New York NY (1957).

[14] C.L. Hanson and R.J. Lawson, Solving Least Squares Problems, Prentice Hall, Englewood Cliffs, NJ (1974).

[15] G.H. Golub and V. Pereyra, "The Differentiation of Pseudoinverses, Separable Nonlinear Least Squares Problems, and Other Tales", Generalized Inverses and Applicatications, ed. M.Z. Nashed, Academic Press, New York NY, pp. 303-324 (1976).

[16] G.H. Golub and C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore MD (1983)

[17] R.C. Rao and S.K. Mitra, The Generalized Inverse of a Matrix and its Applications, John Wiley, New York NY (1971).

[18] G.H. Golub, V. Klema, and G.W. Stewart, "Rank Degeneracy and Least Squares Problems", Stan-CS-76-559 (AD-A032 348), Computer Science Department, Stanford University (August 1976).

[19] L. Kaufman, "A Variable Projection Method for Solving Separable Nonlinear Least Squares Problems", BIT 15, pp. 49-57 (1975).

[20] D. Marquardt, "An Algorithm for Least Squares Estimation of Nonlinear Parameters", Journal of the Society of Industrial and Applied Mathemetics 11 (1963).

APPENDICES

A. QR Factorisation by Successive Householder Transformation

The Householder reflector is defined [9] as

$$U = I - \beta^{-1} \underline{u}\, \underline{u}^T,$$

such that

$$U\, \underline{x} = -\sigma\, \underline{e}_1,$$

where

$$\sigma = \mathrm{sgn}(x_1)\ ||\underline{x}||,$$

$$u_1 = x_1 + \sigma,$$

$$u_i = x_i\ ,\ i=1,2,\ldots,n$$

and

$$\beta = \sigma\, u_1.$$

When triangularizing the matrix F using successive Householder reflectors, each column of F is transformed by reflectors formed from each of the preceeding columns; i.e. denoting as U_i the Householder reflector which zeros the elements below the i'th diagonal, then we define

$$H_i = \left[\begin{array}{c|c} I_{i-1} & 0 \\ \hline 0 & U_i \end{array} \right],$$

such that H_1 zeros the elements below the first diagonal and transforms columns 2 through M, H_2 zeros the elements below the second diagonal and transforms columns 3 through M, and so on. Thus the i'th reflector effects only the rightmost M-i+1 columns and the lower N-i+1 rows. After the reflector has been constructed for all M columns, the orthogonal matrix Q is defined as

$$Q = H_M\, H_M{-1} \cdots H_2\, H_1.$$

Furthermore, since only the vector $\underline{u}_i$ and the scalar β_i are necessary for forming H_i at each stage, all information concerning the construction of Q can be saved by storing the last n-i elements of $\underline{u}_i$ below below the i'th diagonal element of F and storing the pre-transformation value of the diagonal element in an auxillary vector (note that the post-transformation value of the diagonal is $-\sigma_i$, so that π_i is indirectly available).

B. Variable Projection Functional Contour and Surface Plots

This appendix contains contour and surface plots for the variable projection functional of the two-pole high-pass filter transducer model (see Section 6.1). For a given parameter set, the noiseles observation vector, $\underline{y}$, and the known excitation frequency will be fixed and this functional can be written solely in terms of the parameters α and f as

$$\mathrm{RSS}(\alpha,f) = \left|\left| \underline{y} - P_{F(\alpha,f)}\, \underline{y} \right|\right|^2 .$$

For each of these signal parameters sets

{N=16, T=0.25, Q=4, fm1=1.0, f0=1.0}

{N=16, T=0.25, Q=8, fm1=1.0, f0=1.0}

{N=16, T=0.25, Q=12, fm1=1.0, f0=1.0}

{N=16, T=0.25, Q=4, fm1=1.0, f0=0.5}

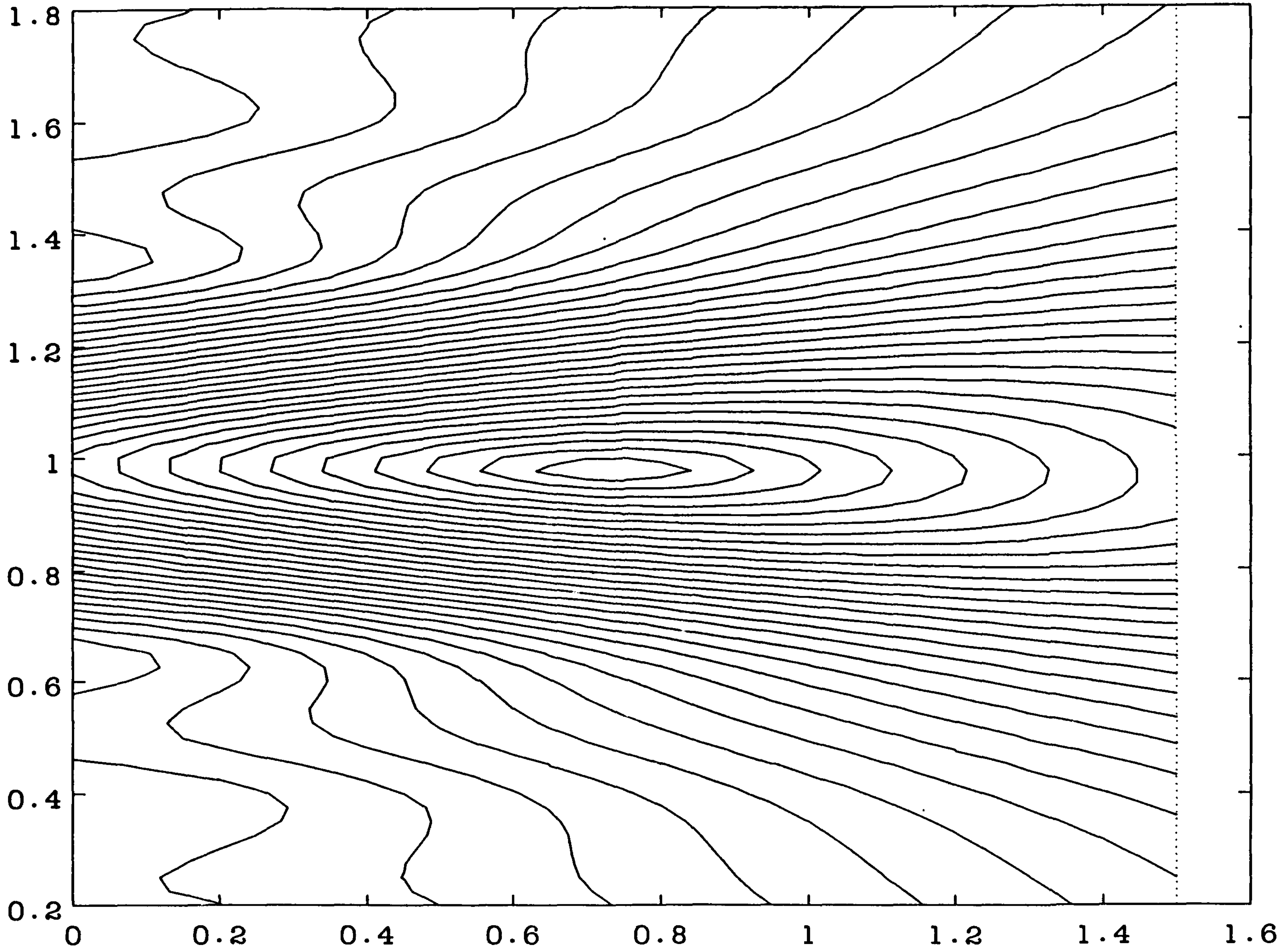
contour plots are given with α in Nepers given along the horizontal axis and f in Hertz given along the vertical axis. Following each contour plot, the following views of each error surface (RSS vs. α and f) are provided:

1. Surface viewed from large f and small α
2. Surface viewed from small f and large α
3. Surface viewed from the α=0 plane (front side)
4. Surface viewed from the α=1.5 plane (back side).

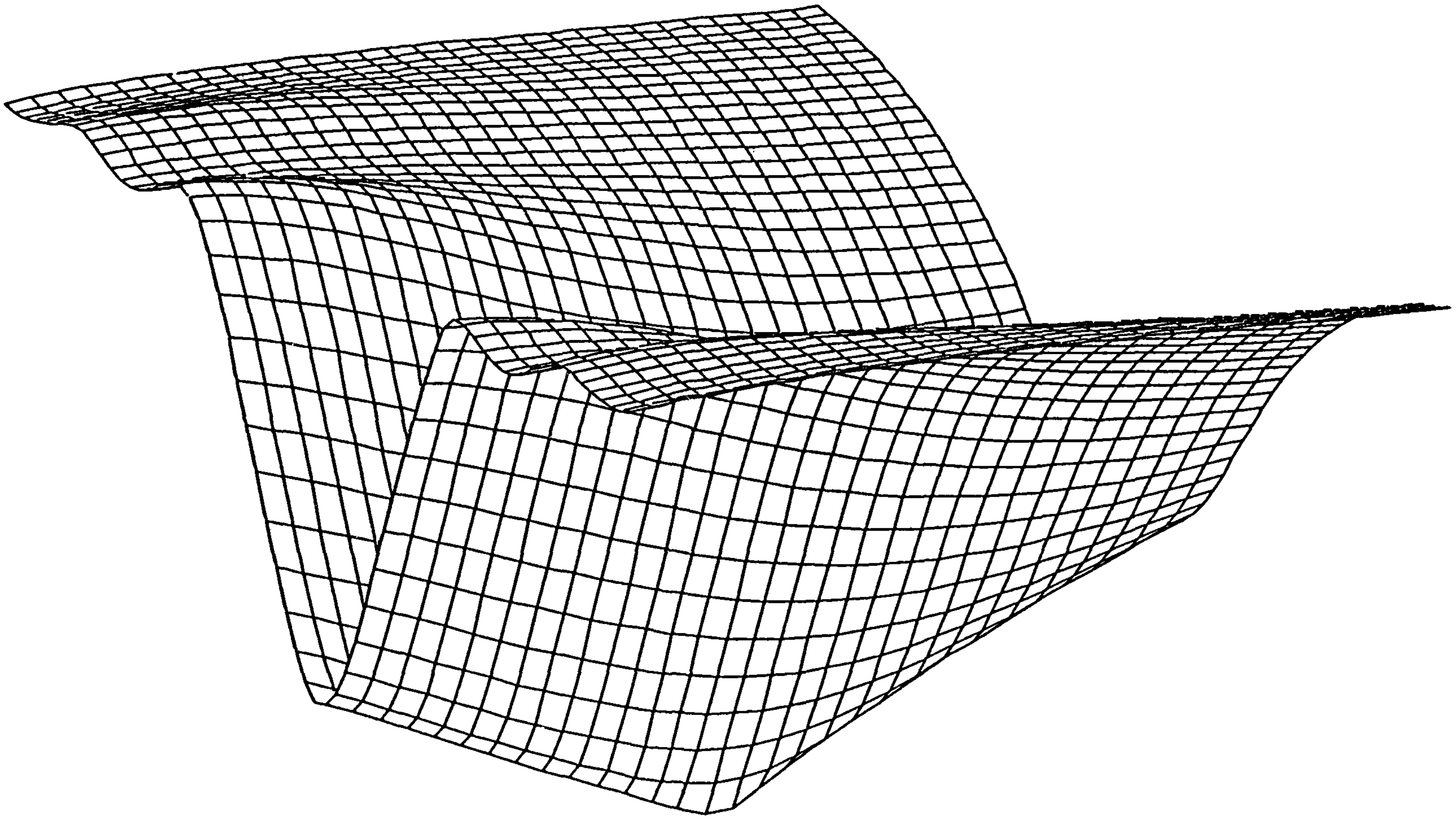
These plots exhibit the flat nature of the functional's surface as the frequency estimate (f) becomes far removed from the true minimum of the functional. It is in these areas particularly that the problems described in Section 6.3 occur.

The non-resonance excitation case (f0=0.5) was included above for comparison. This contour suggests a smoother surface than the excitation-at-resonance case.

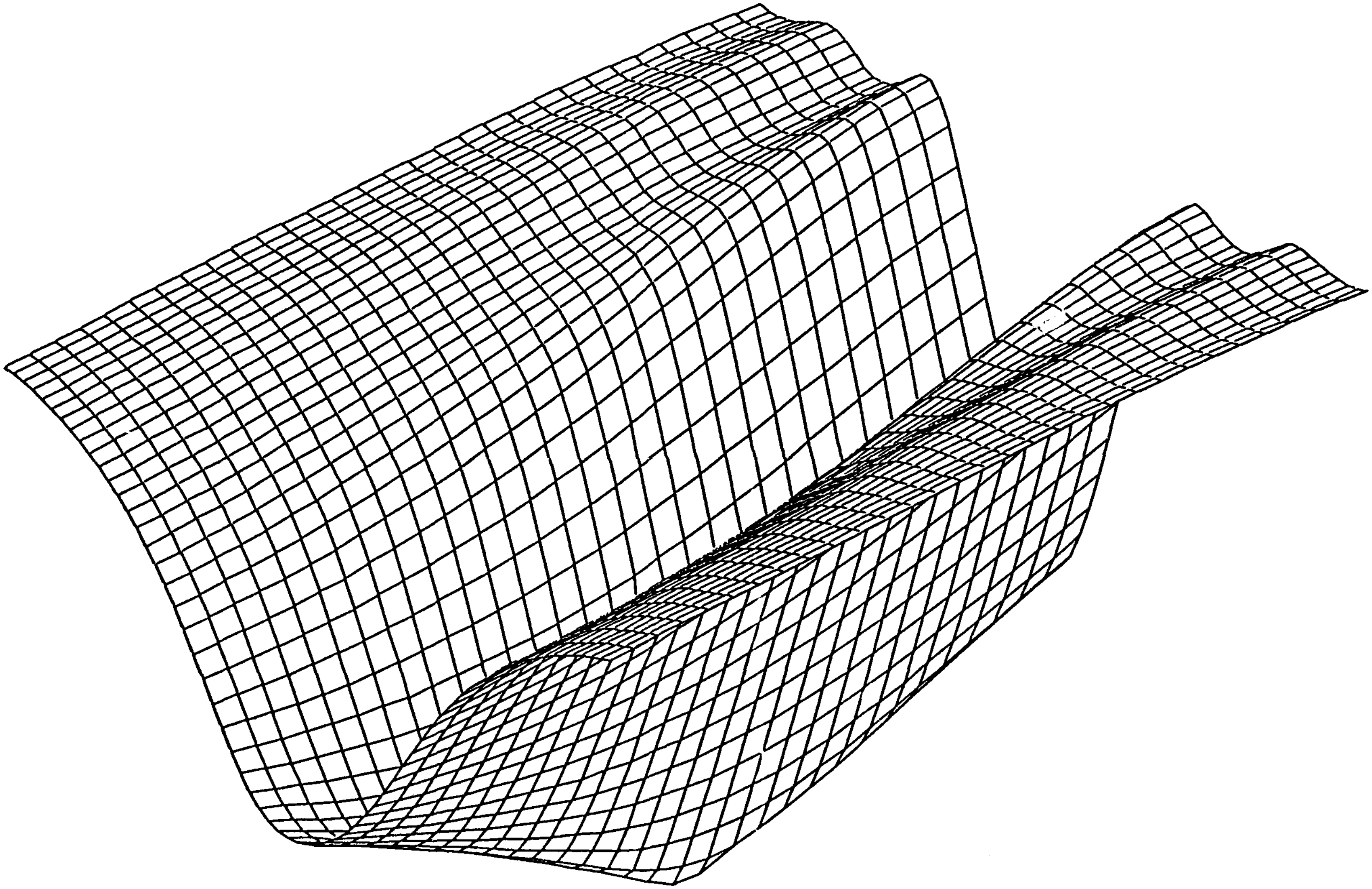
Model_1 RSS Contour N=16 Q=4 f0=1 a1=0.7304 f1=0.979



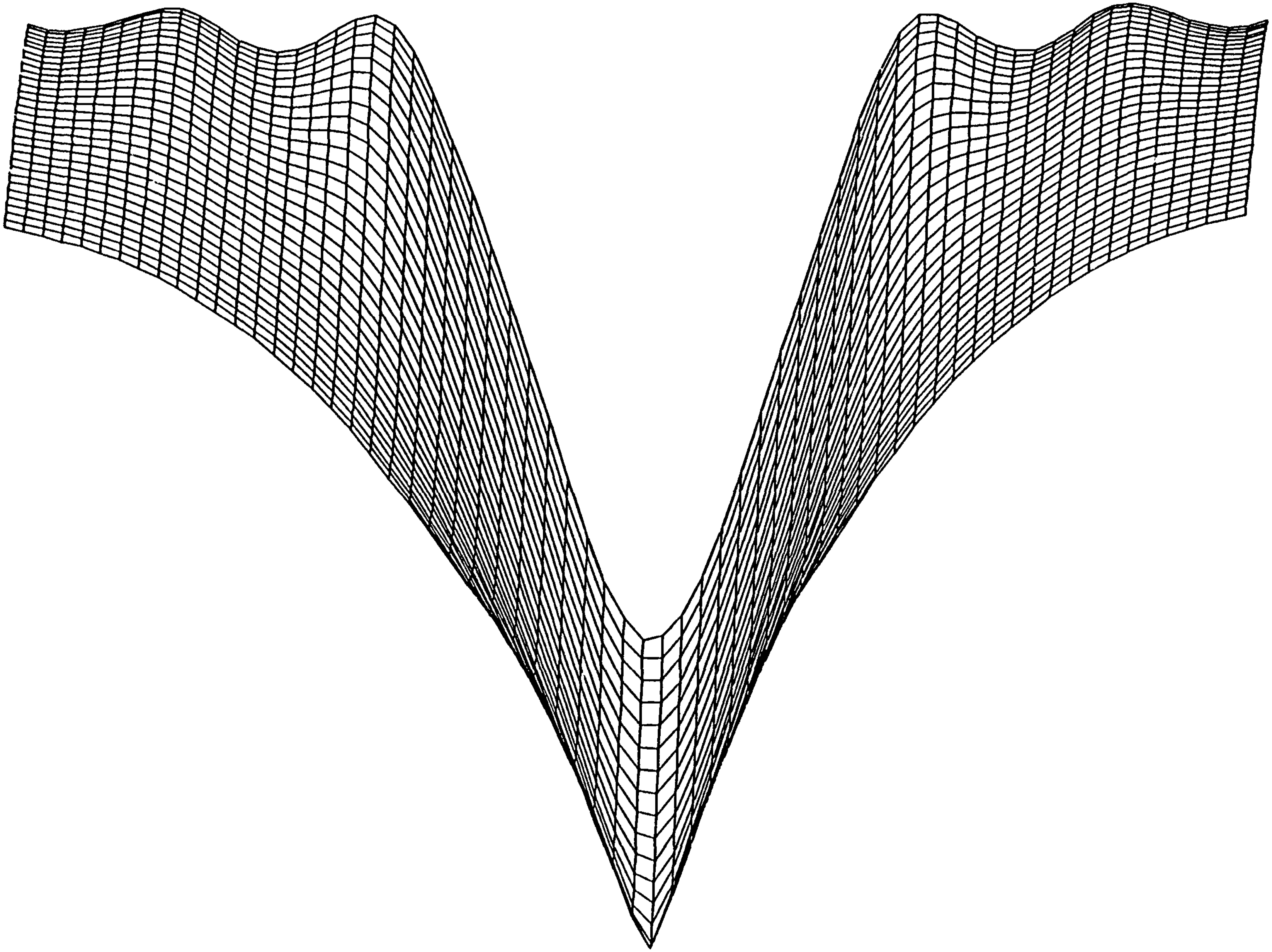
Model_1 RSS Surface N=16 Q=4 f0=1 a1=0.7304 f1=0.979 (-38.8,30)



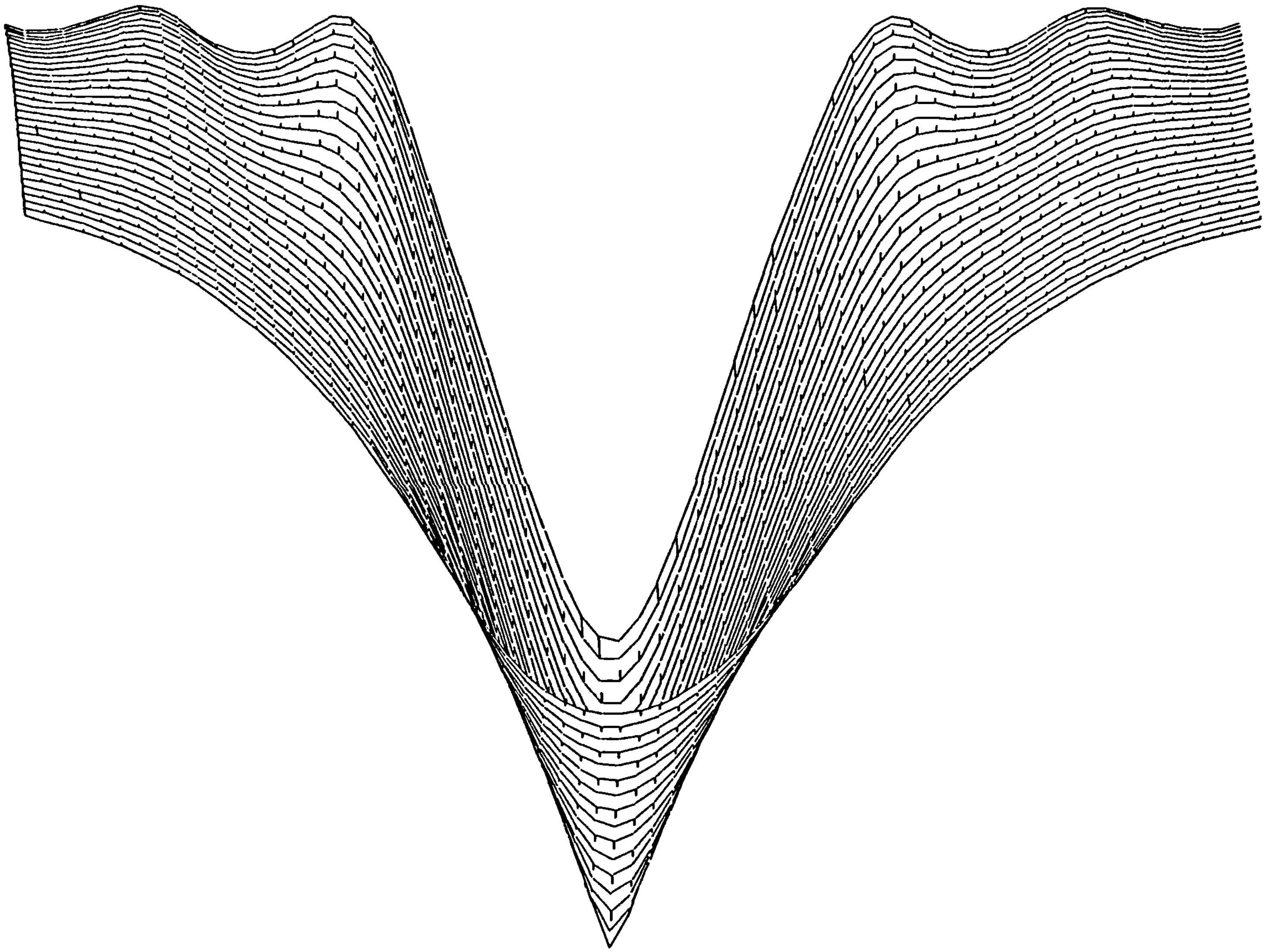
Model_1 RSS Surface N=16 Q=4 f0=1 a1=0.7304 f1=0.979 (142.2,30)



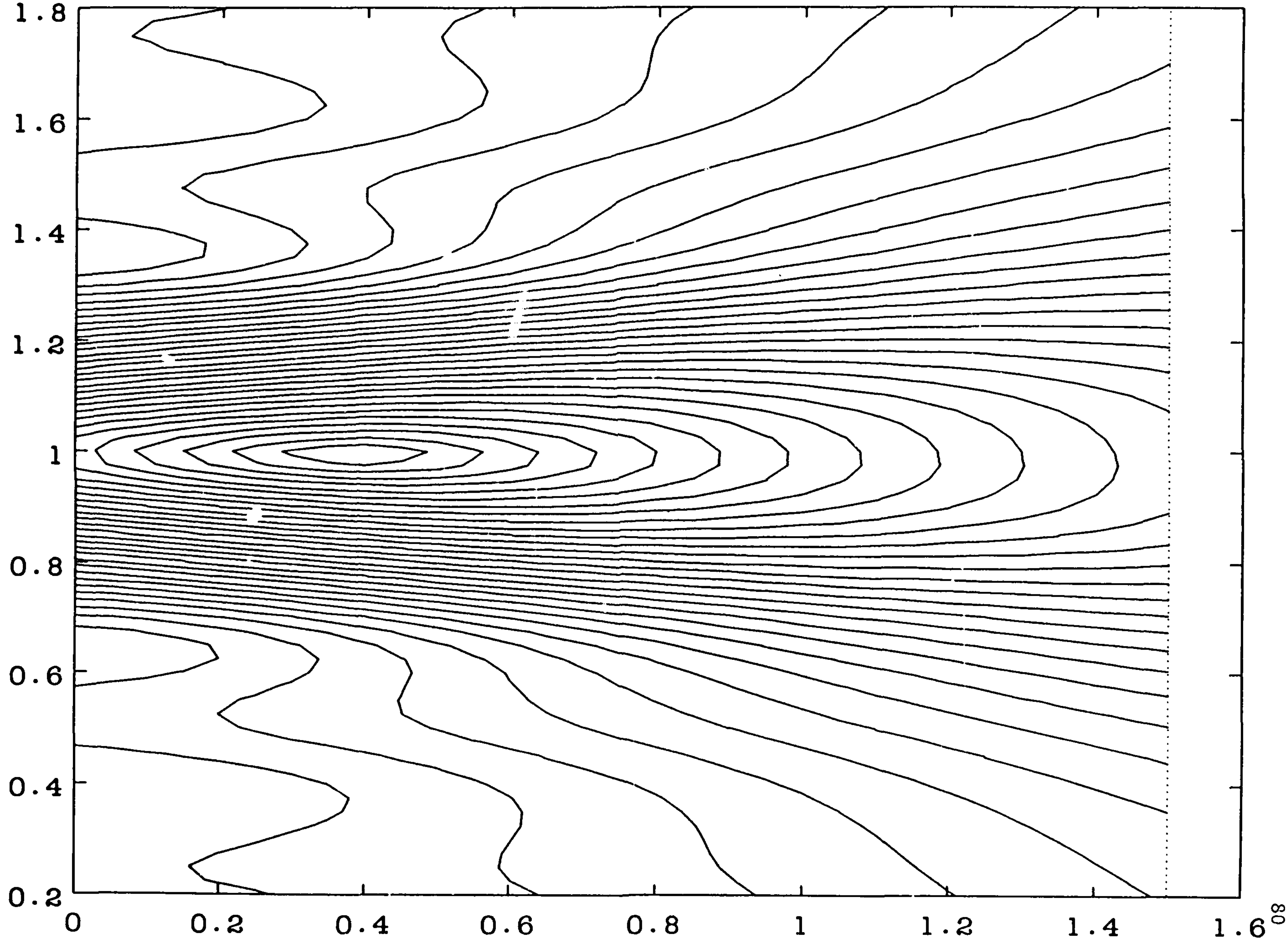
Model_1 RSS Surface N=16 Q=4 f0=1 a1=0.7304 f1=0.979 (270,0)



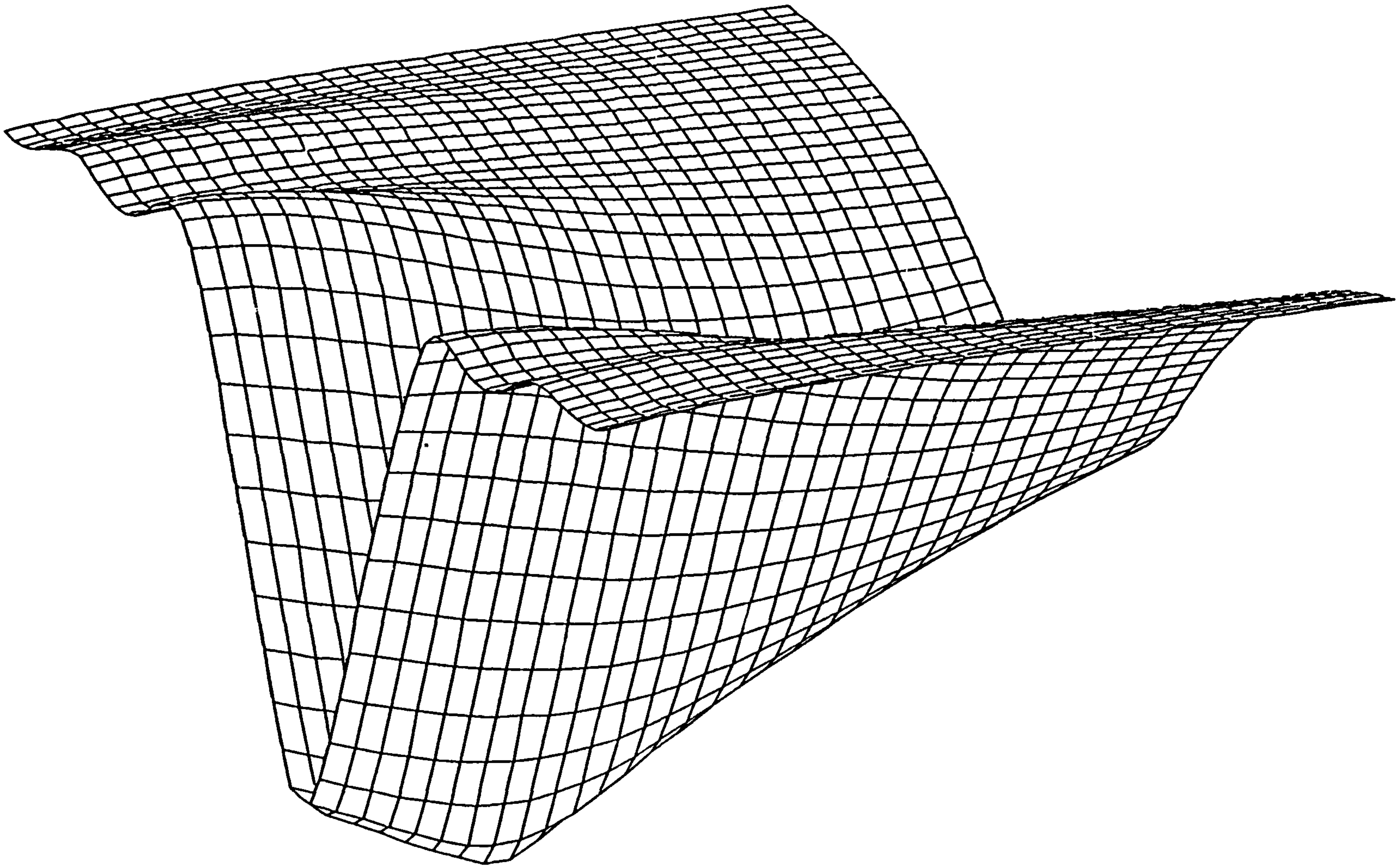
Model_1 RSS Surface N=16 Q=4 f0=1 a1=0.7304 f1=0.979 (90,0)



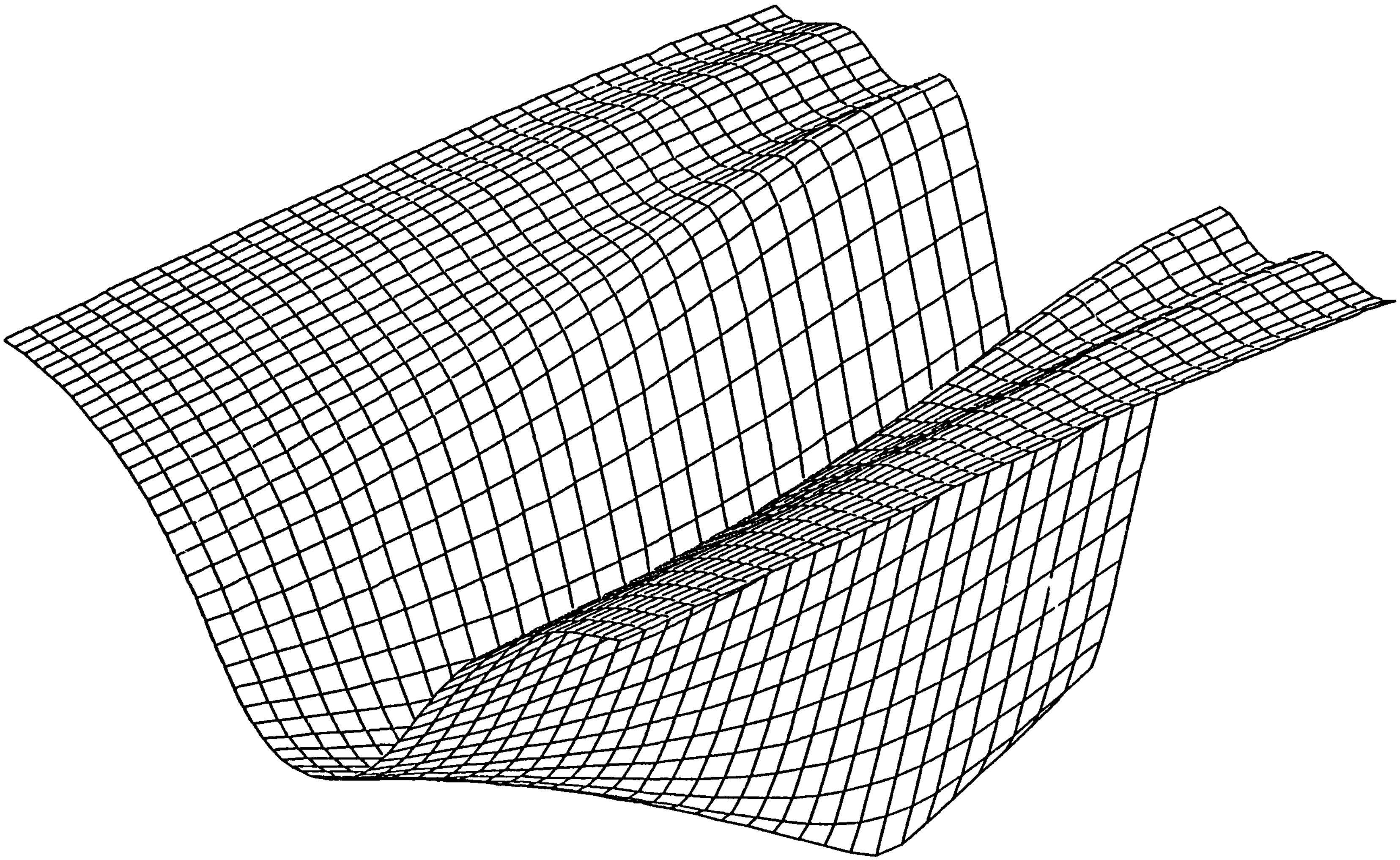
Model_1 RSS Contour N=16 Q=8 f0=1 a1=0.06155 f1=0.9962



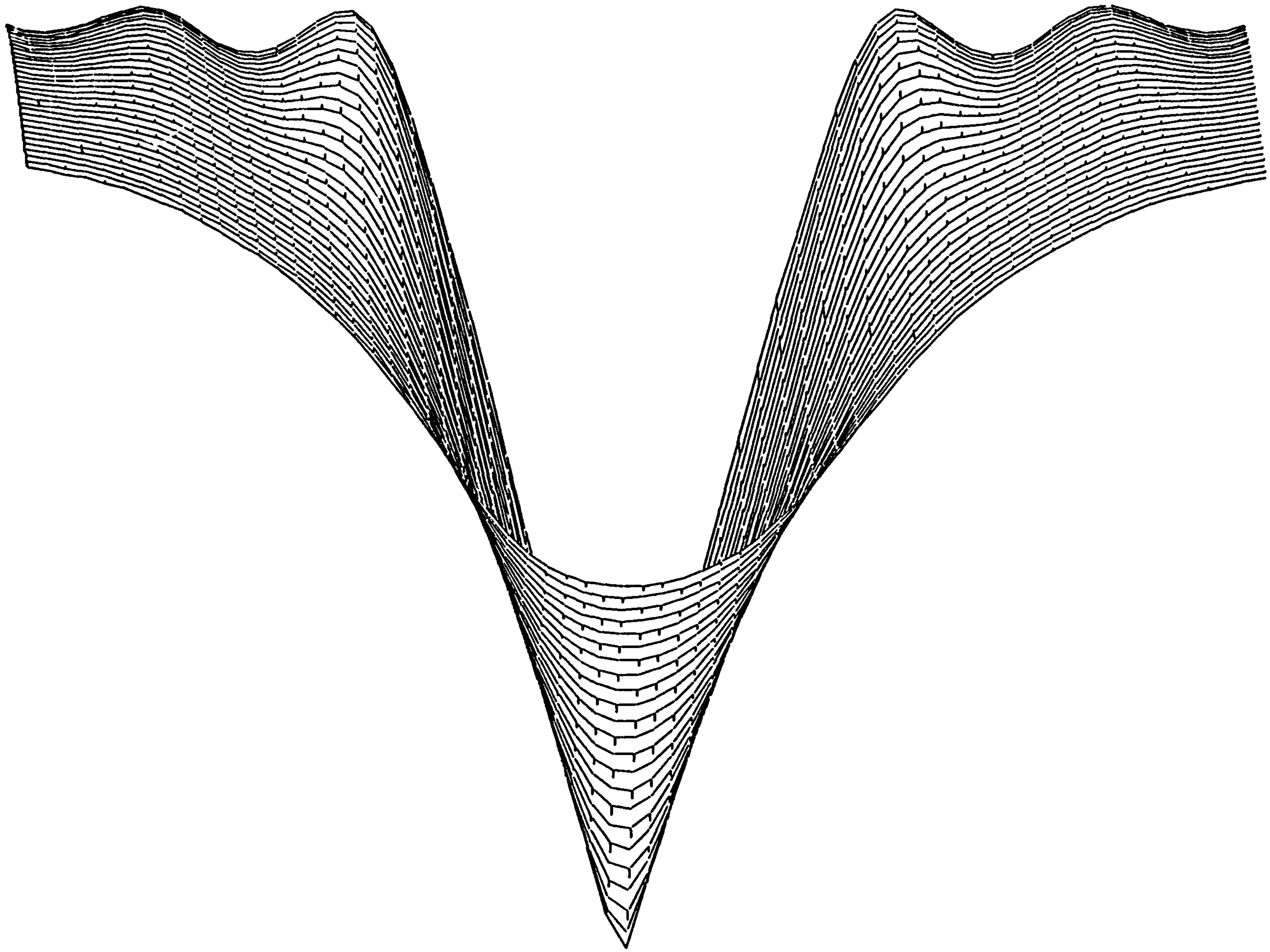
Model_1 RSS Surface N=16 Q=8 f0=1 a1=0.06155 f1=0.9962 (-38.8,30)



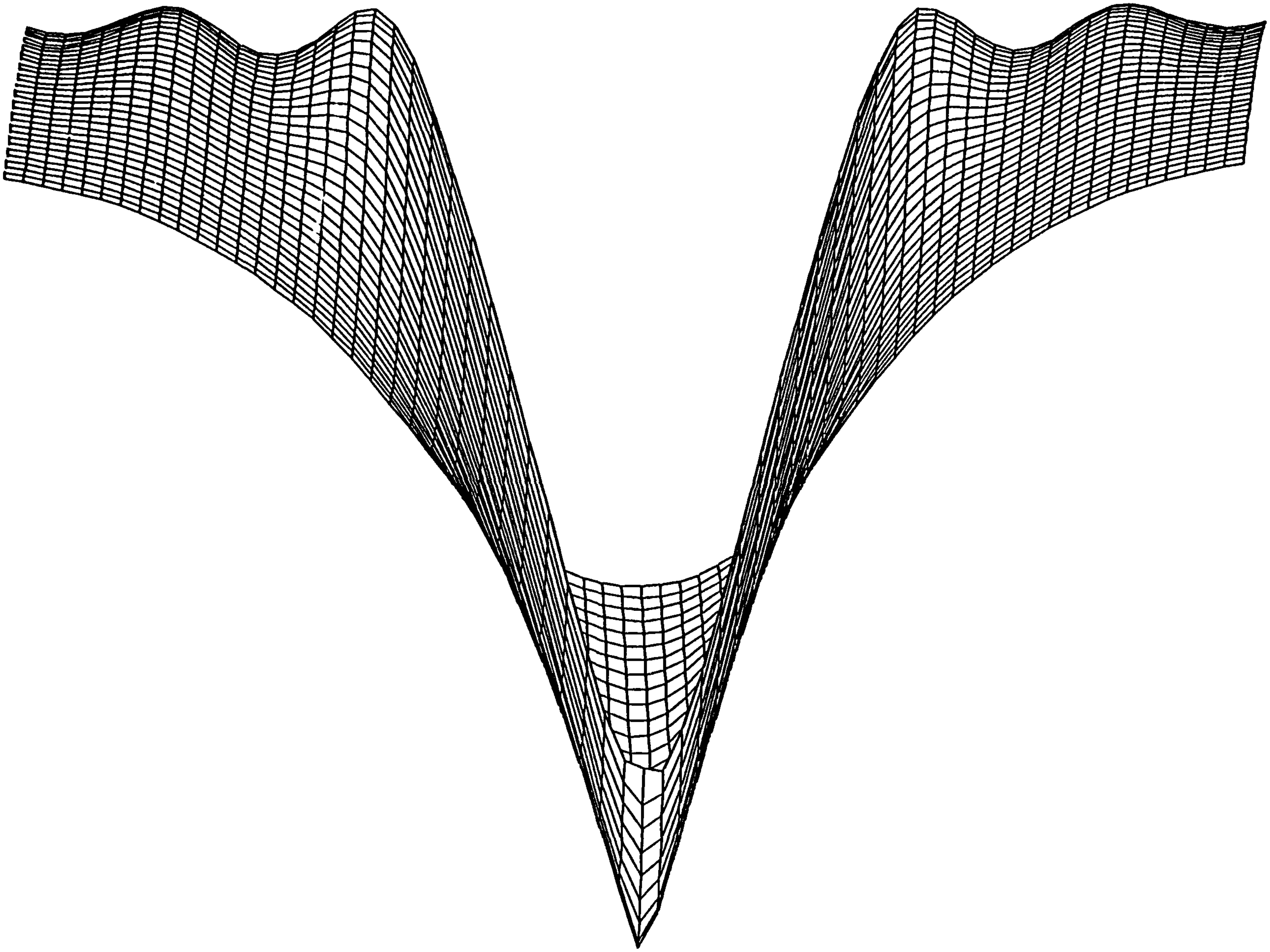
Model_1 RSS Surface N=16 Q=8 f0=1 a1=0.06155 f1=0.9962 (142.2,30)



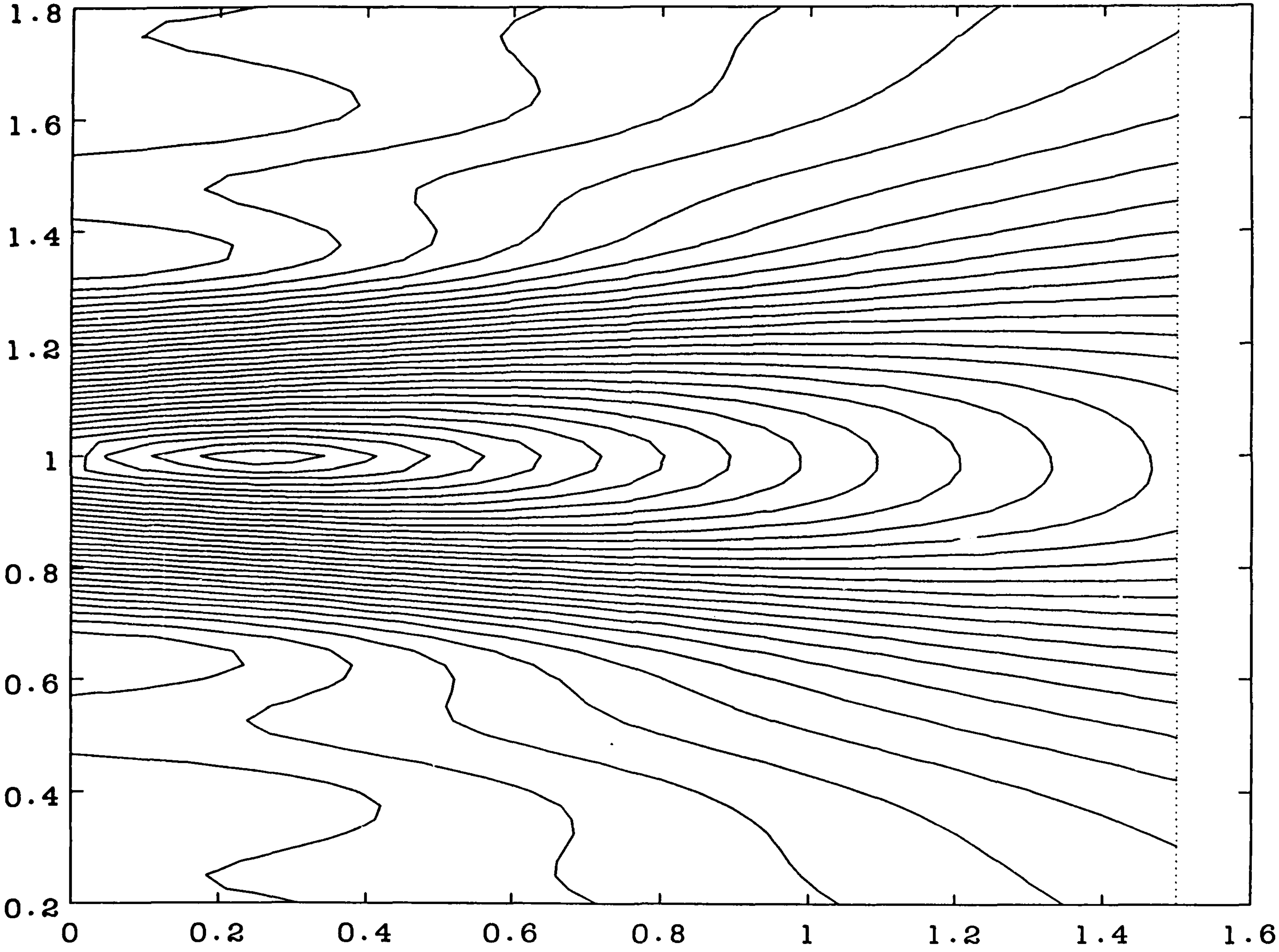
Model_1 RSS Surface N=16 Q=8 f0=1 a1=0.06155 f1=0.9962 (90,0)



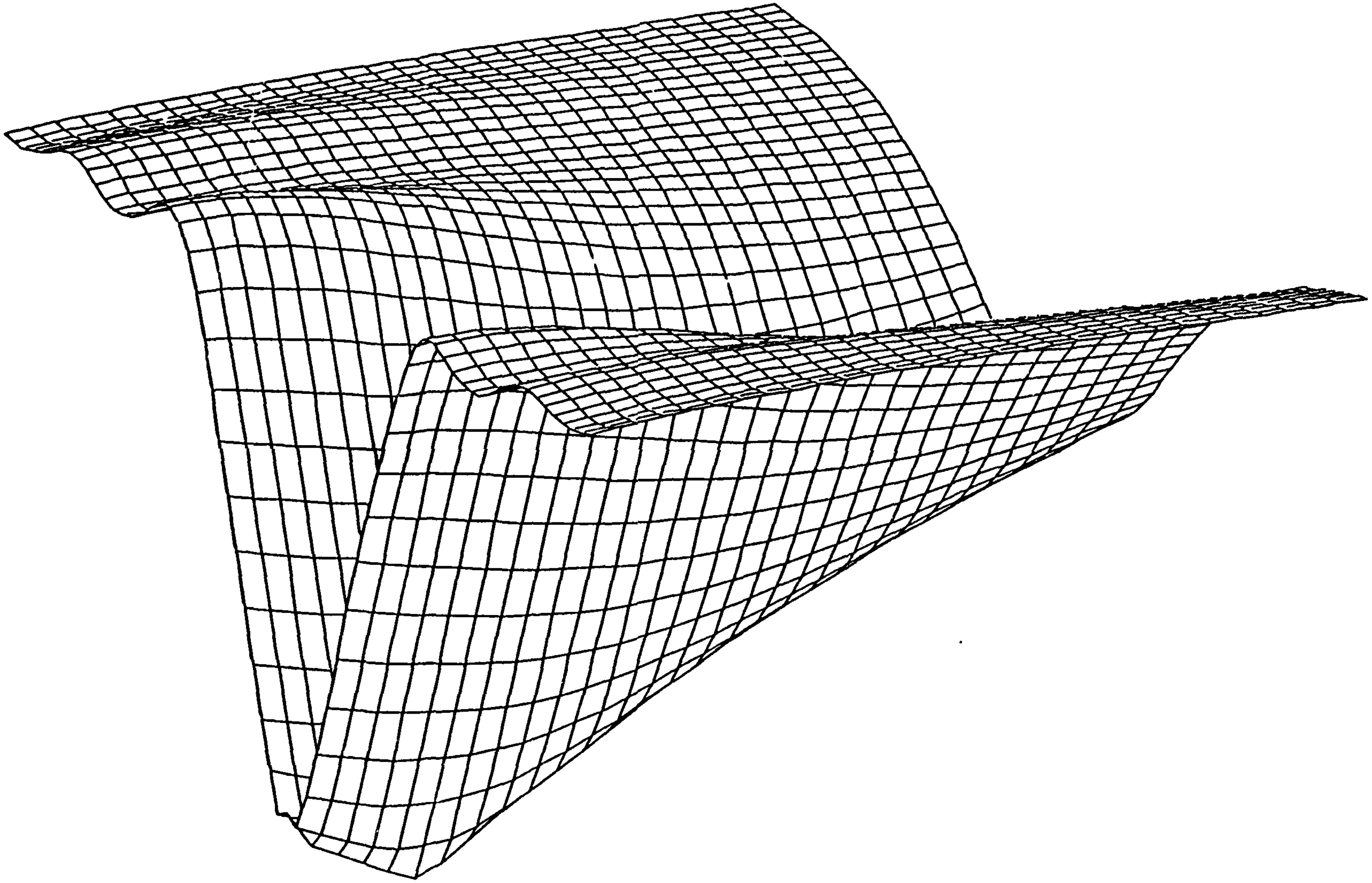
Model_1 RSS Surface N=16 Q=8 f0=1 a1=0.06155 f1=0.9962 (270,0)



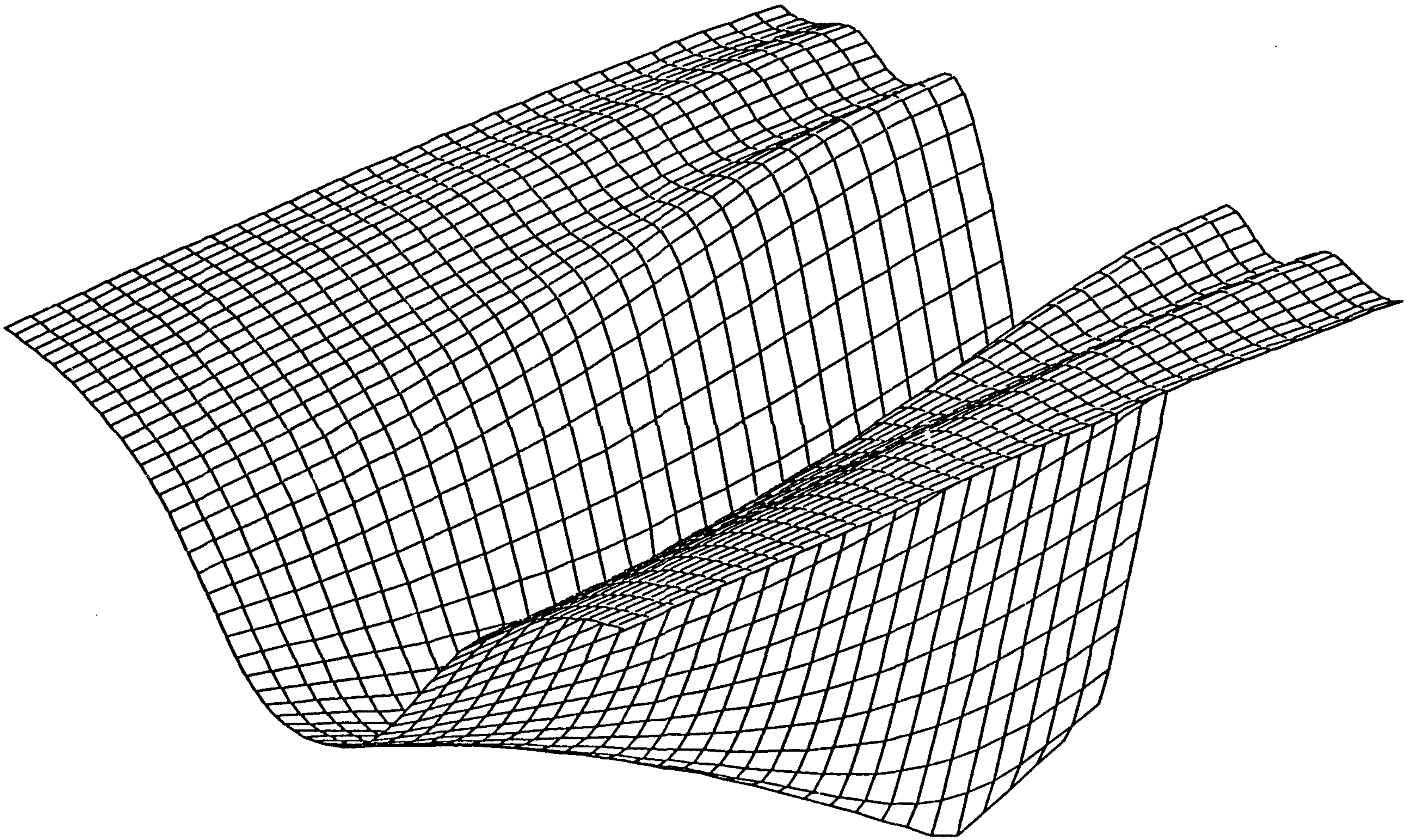
Model_1 RSS Contour N=16 Q=12 f0=1 a1=0.04139 f1=0.9983



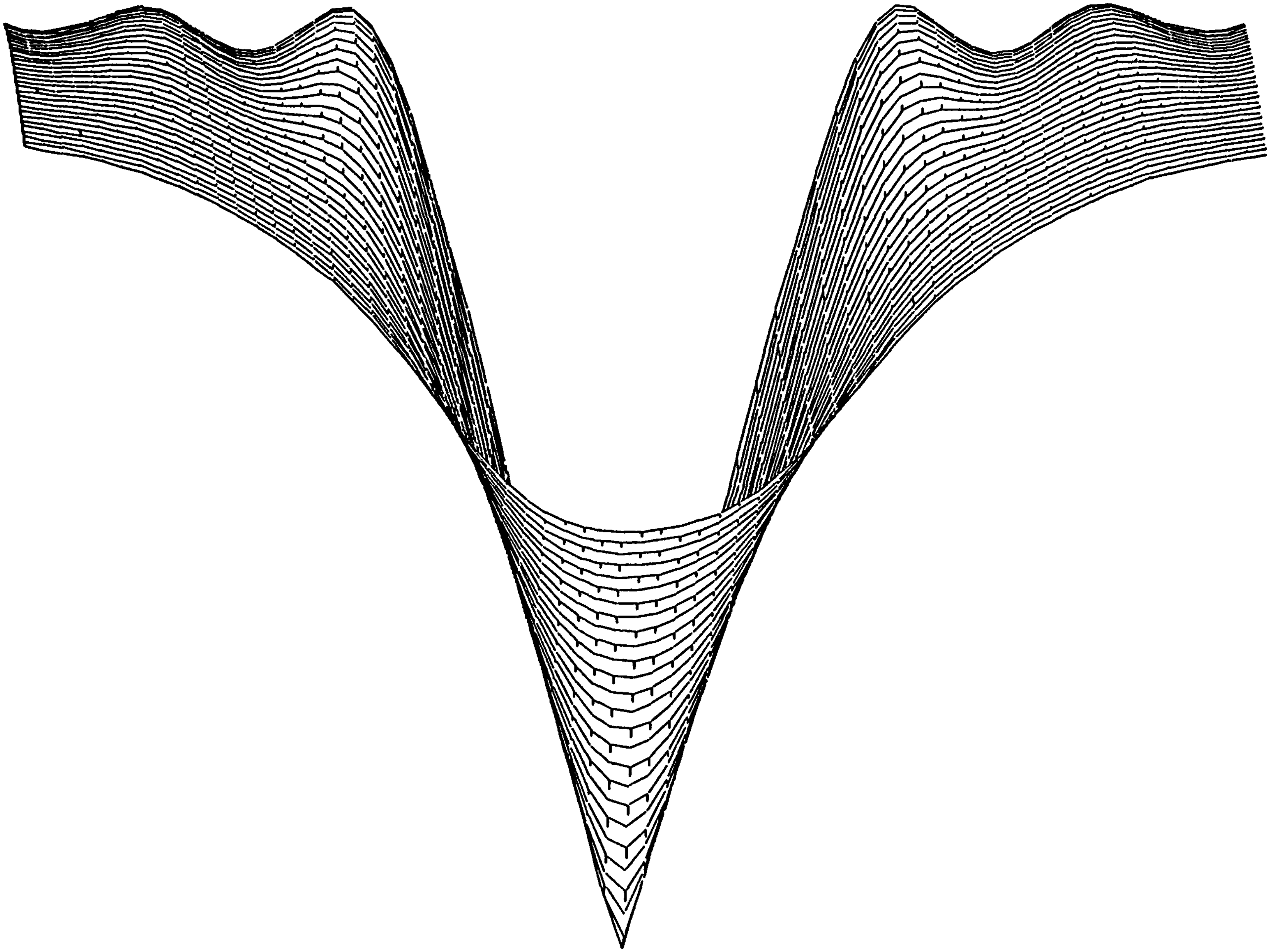
Model_1 RSS Surface N=16 Q=12 f0=1 a1=0.04139 f1=0.9983 (-38.8,30)



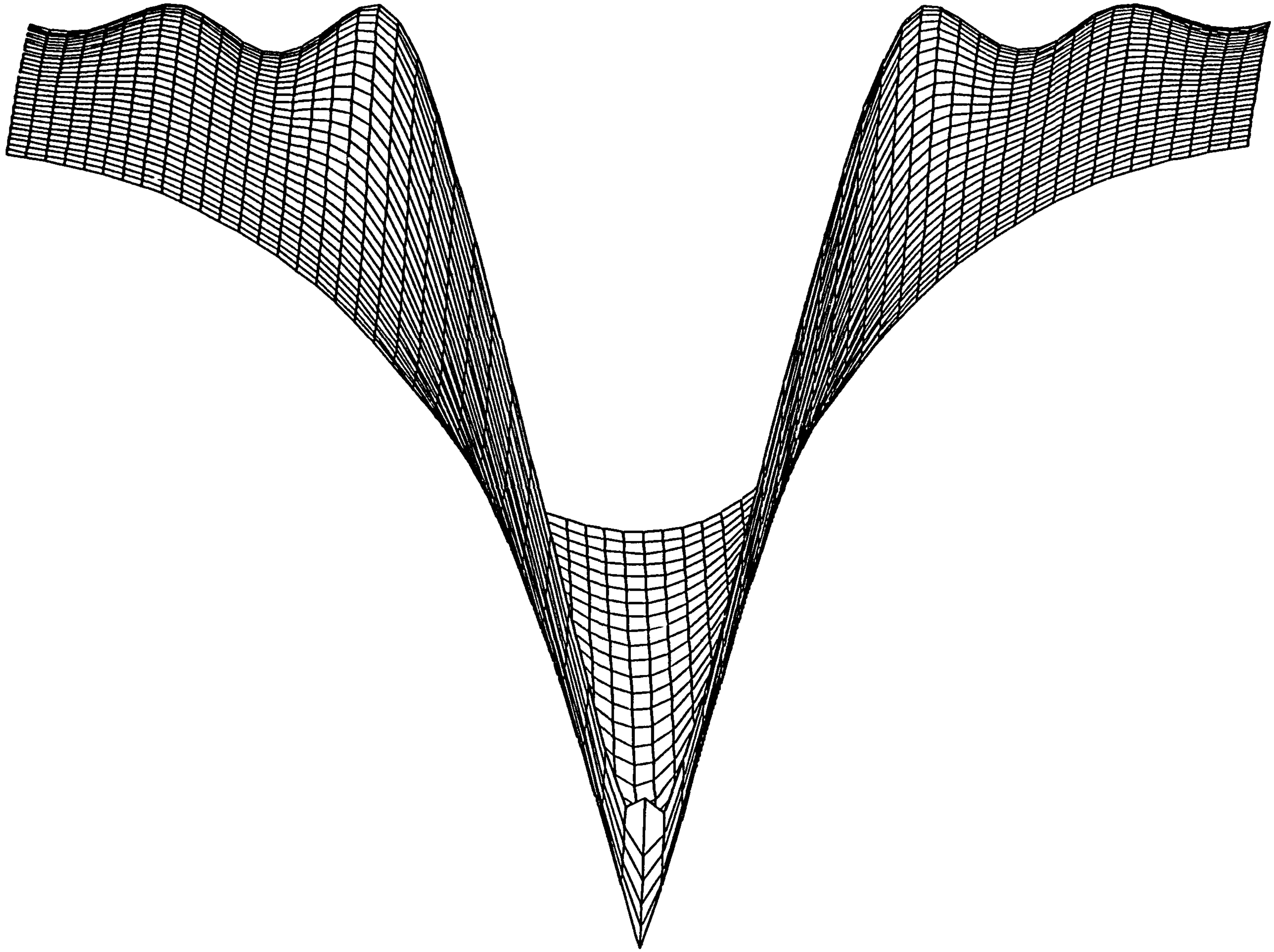
Model_1 RSS Surface N=16 Q=12 f0=1 a1=0.04139 f1=0.9983 (142.2,30)



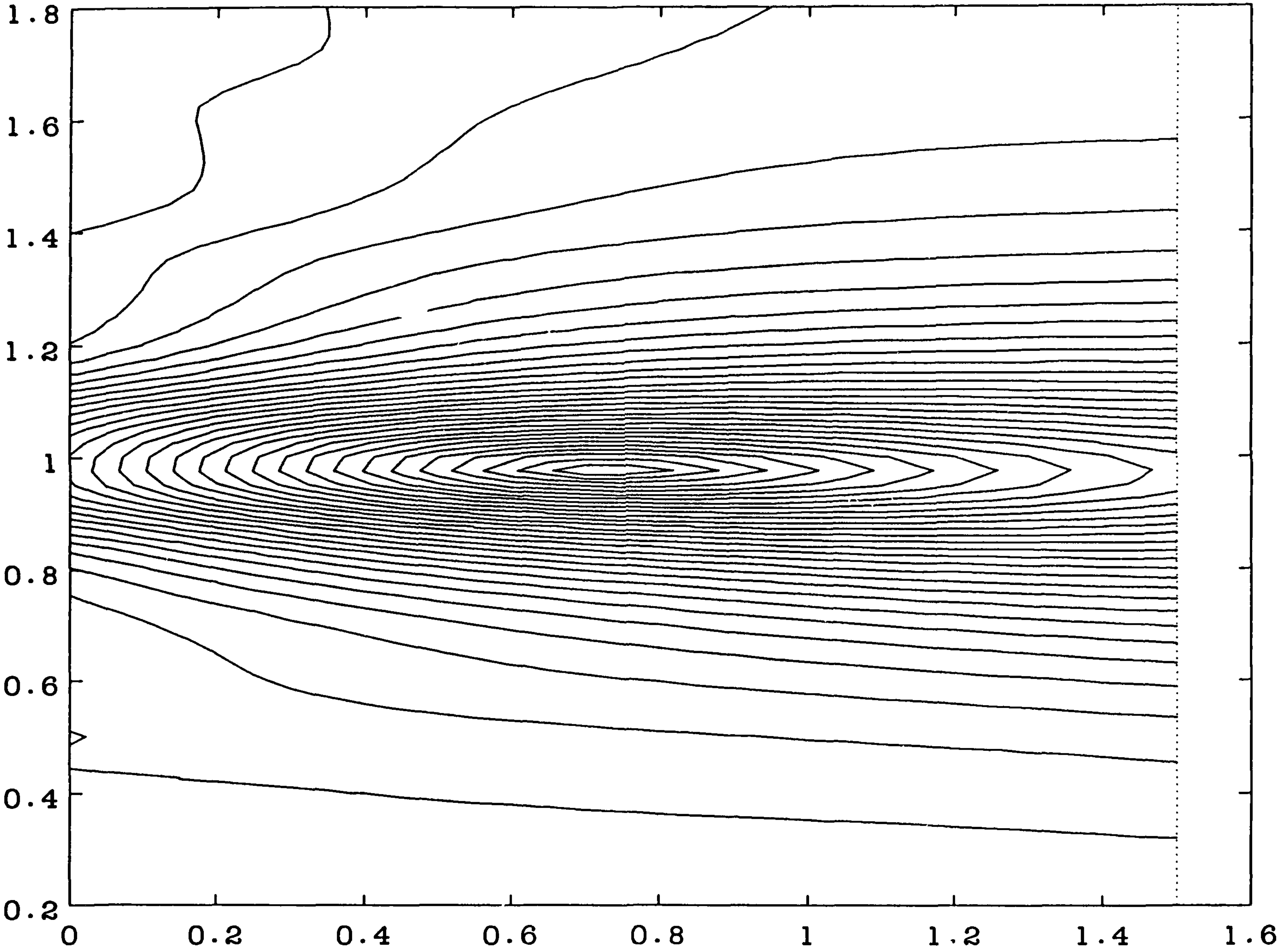
Model_1 RSS Surface N=16 Q=12 f0=1 a1=0.04139 f1=0.9983 (90,0)



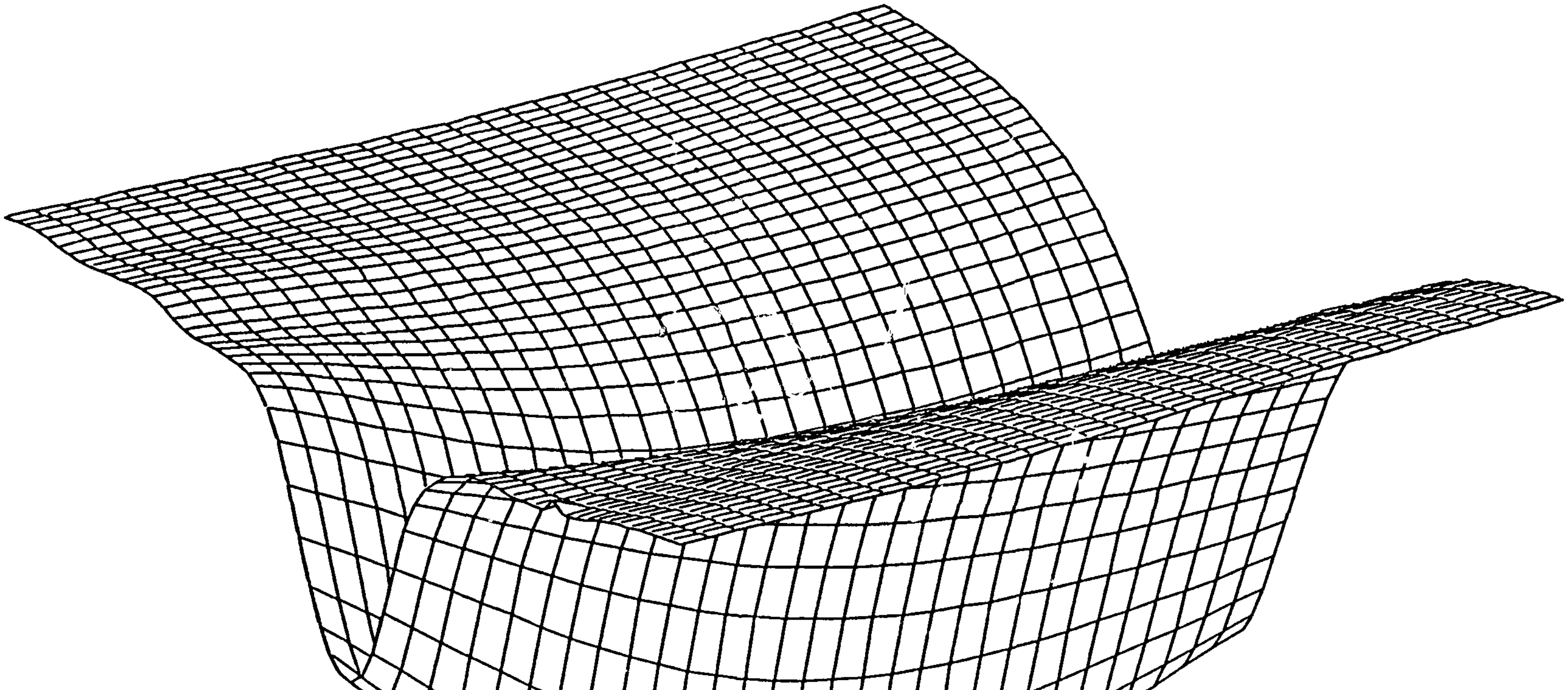
Model_1 RSS Surface N=16 Q=12 f0=1 a1=0.04139 f1=0.9983 (270,0)



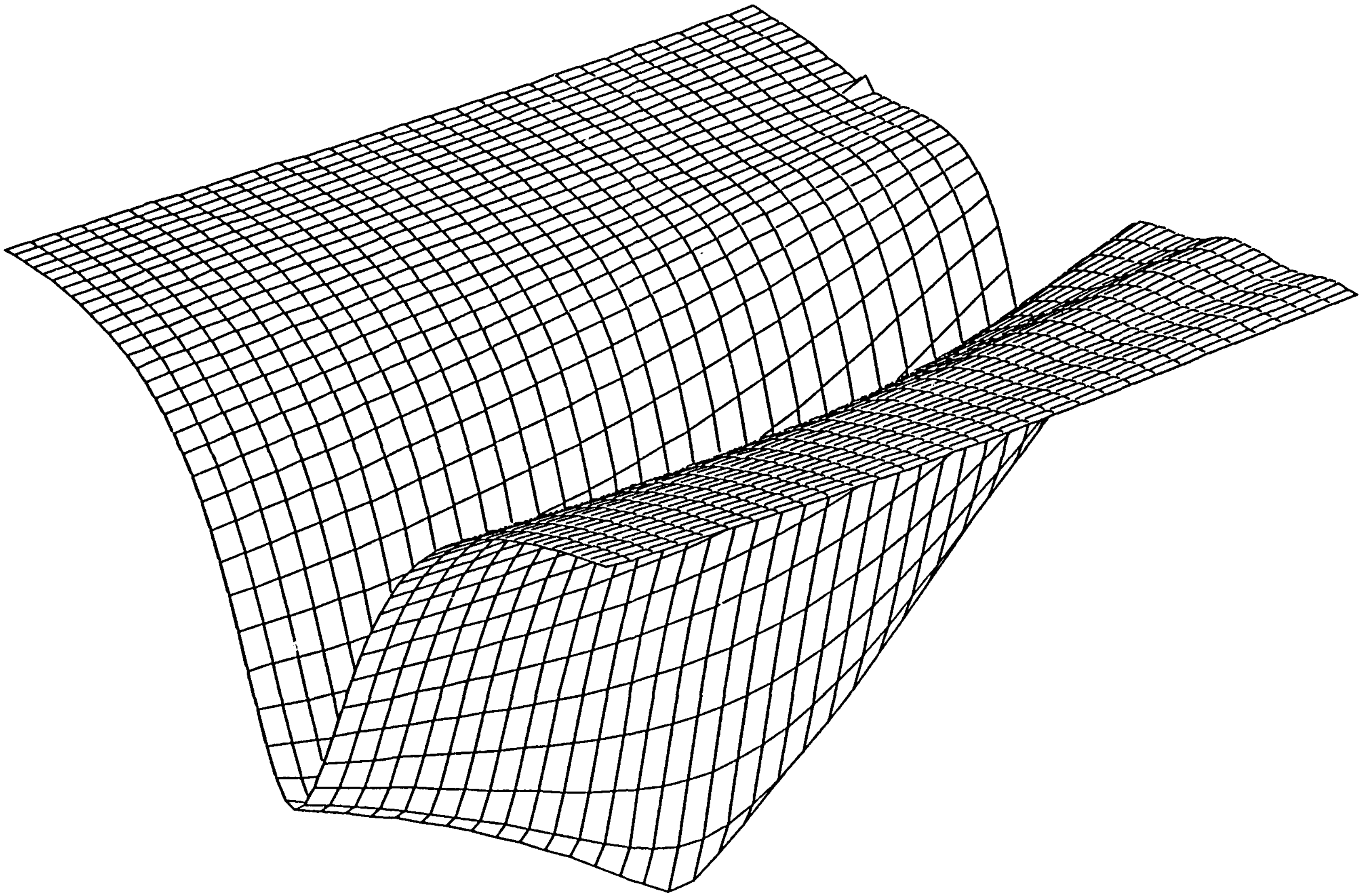
Model_1 RSS Contour N=16 Q=4 f0=0.5 a1=0.7304 f1=0.979



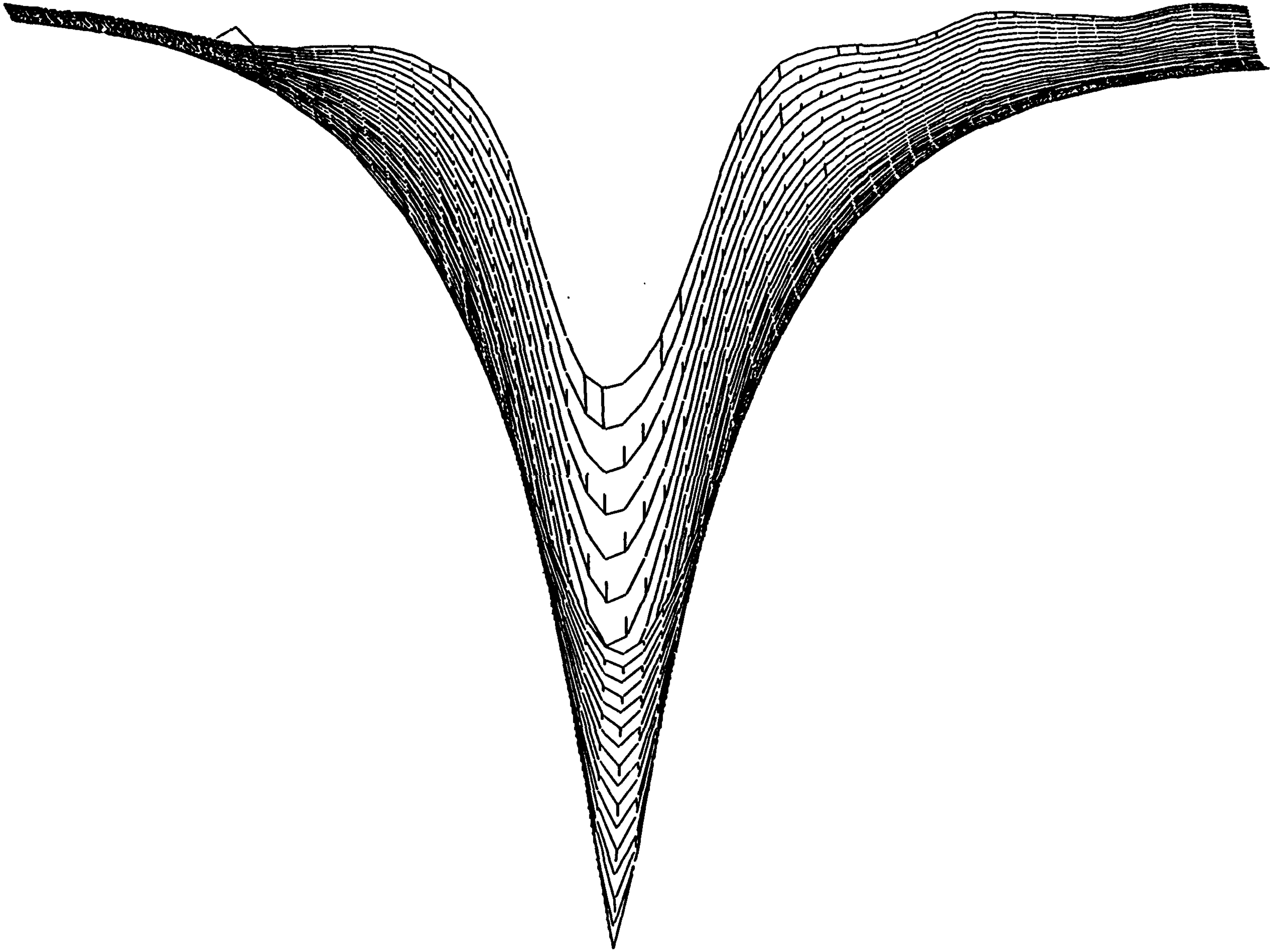
Model_1 RSS Surface N=16 Q=4 f0=0.5 al=0.7304 fl=0.979 (-38.8,30)



Model_1 RSS Surface N=16 Q=4 f0=0.5 a1=0.7304 f1=0.979 (142.2,30)



Model_1 RSS Surface N=16 Q=4 f0=0.5 a1=0.7304 f1=0.979 (90,0)



Model_1 RSS Surface N=16 Q=4 f0=0.5 a1=0.7304 f1=0.979 (270,0)