

ROBUST IDENTIFICATION OF DYNAMIC SYSTEMS

By

SARATCHANDRAN C. PUTHENPURA

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University



December 1985

ROBUST IDENTIFICATION OF DYNAMIC SYSTEMS

DEDICATED

to

my parents

and to the loving memory of

my grandfather

DOCTOR OF PHILOSOPHY (1985)

McMASTER UNIVERSITY

(Electrical Engineering)

Hamilton, Ontario

TITLE: Robust Identification of Dynamic Systems

AUTHOR: Saratchandran C. Puthenpura, B.Tech (Elec. Eng.,
Indian Institute of Technology, Madras.)

SUPERVISOR: Naresh K. Sinha, Professor of Electrical
and Computer Engineering
B.Sc. Eng. (Banaras)
Ph.D. (Manchester)

NUMBER OF PAGES: xix, 220

ABSTRACT

The problem of identification (estimation of system parameters and characteristics) is considered. The aspect emphasized is the robustness in identification. By saying robustness, two points are meant. One is the robustness with respect to bad pieces of data and the other is the numerical robustness with respect to truncation and round-off errors in computation.

A thorough study has been made on the robust statistical principles and their applicability in system identification is critically evaluated. Off-line and on-line identification algorithms are proposed, which are resilient to undetectable spurious errors in the data and at the same time computationally simple and efficient. The convergence of these algorithms is theoretically established. Besides, a robust recursive algorithm is also proposed which jointly estimates the states and parameters of a linear system in a bootstrap manner. This algorithm is also proven to be converging with probability one. In addition to this, a very general method is developed for evaluating the asymptotic efficiency of robust identification methods. The superiority of the proposed approaches in contrast with the conventional identification methods (least squares and its generalizations) is illustrated with the help of several

simulated as well as real-life examples.

The numerical instability caused by improper choice of sampling rates is also subjected to considerable study in the context of identification of continuous-time systems from samples of input-output data. Methods are suggested to overcome this problem. Also, numerically robust schemes are introduced for transforming discrete-time models to their continuous-time equivalents and their performance is compared with other existing methods.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude to Professor N.K. Sinha whose expert guidance, constant encouragement and overwhelming enthusiasm have been most invaluable to me throughout this work. I also thank Professor K.M. Wong and Professor S.G. Mohanty, members of my Ph.D. supervisory committee for their continuing interest and support.

I am very grateful to Dr. O. Vidal, Mr. F.K. Omani and Dr. N. Balakrishnan for several stimulating discussions which have been highly educational for me. I would also extend my appreciation to Prof. J.F. MacGregor for his very helpful suggestions and for providing the packed-bed reactor data. In addition, it is my pleasure to acknowledge the reviewers of Automatica and IEE Proceedings for their constructive criticisms.

I am indebted to my colleagues, Mr. L. Liu, Mr. J.R. Rao, Mr. J. Ching, Mr. A.J. Nagatakiya, Dr. A.M. Karnik, Dr. M.S. El-Sobki and Mrs. Z. Abu-El-Magd for participating in my several presentations and subsequent discussions.

I am also thankful to Mr. P.G. Madhavan, Mr. B.W. Dahanayake, Mr. B.P. Mohanty and Miss. S. Mini for their encouragement throughout this work.

TABLE OF CONTENTS

	PAGE
ABSTRACT	v
ACKNOWLEDGEMENTS	vii
LIST OF FIGURES	xiv
LIST OF TABLES	xviii
CHAPTER 1: INTRODUCTION	1
1.1 Introduction to the Problem	1
1.2 Organization of the Thesis and Original Contributions	4
CHAPTER 2: A SURVEY OF ROBUST STATISTICAL METHODS	8
2.1 Overview of the Chapter	8
2.2 History and Development of Robust Statistical Methods.	8
2.3 Robust Methods of Estimation of a Location Parameter.	10
2.3.1 Maximum Likelihood Type Estimates (M-Estimates)	11
2.3.2 Linear Combination of Order Statistics (L-Estimates)	12
2.3.3 Modified Maximum Likelihood Method Based on Censored Samples (MML Estimates)	12
2.3.4 Estimates Using Rank Tests (R-Estimates)	13
2.3.5 Trimmed and Winsorized Means	14
2.3.6 Other Estimates	15
2.4 Numerical Studies on Robustness	15

2.5	Application of Robust Methods in Engineering	16
2.5.1	Filtering and Signal Processing	16
2.5.2	System Identification	17
2.6	Concluding Remarks	18
CHAPTER 3:	ROBUST OFF-LINE IDENTIFICATION USING MODIFIED MAXIMUM LIKELIHOOD (MML) METHOD	20
3.1	Overview of the Chapter	20
3.2	The Modified Maximum Likelihood (MML) Method	20
3.3	Application of MML Method in System Identification	23
3.3.1	Introduction	
3.3.2	Formulation of the Off-line Algorithm	25
3.4	Important Properties of MML Estimation	30
3.5	Results of Simulation	30
3.6	Concluding Remarks	32
CHAPTER 4:	APPLICATION OF M-ESTIMATION IN ROBUST RECURSIVE IDENTIFICATION	36
4.1	Overview of the Chapter	36
4.2	A Review of M-Estimation	36
4.3	Introduction to Recursive Identification	39
4.4	A Robust, Weighted Least Squares Type Algorithm.	42
4.5	Convergence Analysis	45
4.6	Practical Considerations	51

4.6.1	How to Decide on M-Estimation?	52
4.6.2	Comments on the Choice of Influence Functions	53
4.7	Efficiency of the Robust Algorithm	54
4.7.1	The Scalar Fisher Information	55
4.7.2	Evaluation of Efficiency of the Robust Algorithm	57
4.8	Examples of Application of the Robust Algorithm	61
4.8.1	Simulation Studies	62
4.8.2	Application of the Robust Algorithm to a Real-life Problem (Identification of a Packed-bed Chemical Reactor)	63
4.9	Extension of M-Estimation to Instrumental Variables Method	65
4.9.1	Introduction	65
4.9.2	IV Method as Approximate Gradient Method	66
4.9.3	Robust Version of the IV Method	68
4.9.4	Convergence Analysis	70
4.9.5	Simulation Results	71
4.9.6	Identification of the Packed-bed Reactor	72
4.10	Extension of M-Estimation to Correlation Method	73
4.10.1	Introduction	73
4.10.2	A Robust Correlation Method	74
4.10.3	Discussion on the Convergence	76
4.10.4	Simulation Results	77
4.11	Extension to Impulse and Step Response Methods	77
4.11.1	Introduction	77
4.11.2	Robustness Aspects	78
4.11.3	Notes on Convergence	79
4.11.4	Simulation Results	79
4.12	Concluding Remarks	80

CHAPTER 5: ROBUST RECURSIVE IDENTIFICATION - A BAYESIAN APPROACH	93
5.1 Overview of the Chapter	
5.2 Introduction to Contaminated Density Functions and Huber's Minimax Approach	94
5.3 Robust Recursive Estimation - Bayesian Approach	95
5.3.1 Introduction	95
5.3.2 Derivation of Recursive Identification Algorithms	96
5.3.3 Tracking Time-varying Systems	100
5.4 Convergence and Efficiency of the Algorithm	101
5.5 Examples of Application of the Proposed Method	103
5.5.1 Simulation Results	
5.5.2 The Packed-bed Reactor Identification Problem	103
5.6 Concluding Remarks	104
CHAPTER 6: ROBUST BOOTSTRAP METHOD FOR THE JOINT ESTIMATION OF PARAMETERS AND STATES	111
6.1 Overview of the Chapter	111
6.2 The Problem of Joint Parameters-States Estimation	111
6.3 The Bootstrap Method	112
6.3.1 Introduction	112
6.3.2 Special Canonical Forms	113
6.3.3 Formulation of the Bootstrap Method as a Pseudo Linear Regression (PLR)	116

6.3.4	The Robust Bootstrap Algorithm	118
6.4	Convergence and Efficiency of the Robust Bootstrap Algorithm	120
6.5	Extension to Multivariable Systems	122
6.6	Simulation Results	124
6.7	Concluding Remarks	125
CHAPTER 7:	NUMERICAL ROBUSTNESS ISSUES IN THE IDENTIFICATION OF CONTINUOUS-TIME SYSTEMS	132
7.1	Overview of the Chapter	132
7.2	Identification of Continuous-time Models	132
7.3	Effects of Choosing Very Small Sampling Intervals	136
7.3.1	Introduction	136
7.3.2	Illustrations of the Effect of Very Small Sampling Intervals	138
7.4	Procedures to Obtain Proper Values of Sampling Intervals	141
7.4.1	Introduction	141
7.4.2	The W-Plane Method	142
7.4.3	Critical Evaluation of the W-Plane Method	146
7.4.4	The P-Plane Method	147
7.4.5	Critical Evaluation of the P-Plane Method	153
7.5	Transformation from Discrete-time Models to Continuous-time Models	154
7.5.1	Introduction	154
7.5.2	A Numerically Robust Transformation Scheme	154
7.5.3	Reduction in Computation Using Chebyshev Approximation	158
7.5.4	Examples and Comparison of the	

Proposed Method with Other Methods	160
7.6. Concluding Remarks	162
CHAPTER 8. CONCLUSIONS AND RECOMMENDATIONS	173
8.1. Conclusions	173
8.2. Suggestions for Further Research	178
APPENDIX I. OUTLINE OF DERIVATION OF THE MML ESTIMATION ALGORITHM	181
APPENDIX II. PROOF OF THEOREM 4.1	185
APPENDIX III. PROOF OF THEOREM 4.6	189
APPENDIX IV. PROOF OF THEOREM 6.1	192
APPENDIX V.A. FIXED POINT ITERATION SCHEME OF SINHA AND LASTMAN	199
APPENDIX V.B. THE DIRECT APPROACH OF HUNG ET. AL.	200
REFERENCES	201

LIST OF FIGURES

FIGURE		PAGE
3.1	Variation of a and B for Different Degrees of Censoring	35
4.1	Huber's Influence Function	84
4.2a	Geometric Interpretation of Theorem 4.5	84
4.2b	$I(f_2)/I(f_1)$ for Different a/σ	85
4.3	Parameter Error Norm: Least Squares	85
4.4	Parameter Error Norm: Robust Weighted Least Squares Type Algorithm	86
4.5	Convergence of Parameters: Robust Weighted Least Squares Type Algorithm	86
4.6	Variation of Reactor Wall Temperature Over 5 Hours	87
4.7	Variation of Percentage Butane Conversion Over 5 Hours	87
4.8	Residuals of the Least Squares Model	88
4.9	Residuals of the Robust Model	88

4.10	Parameter Error Norm of the IV Method in Presence of Outliers	89
4.11	Parameter Error Norm of the Robust IV Method in Presence of Outliers	89
4.12	Residuals of the IV Model	90
4.13	Residuals of the Robust IV Model	90
4.14	Comparison of Linear and Robust Versions of the Recursive Correlation Method	91
4.15	Parameter Error Norm: Identification from Impulse Response, Least Squares Method (in presence of no outliers)	91
4.16	Parameter Error Norm: Identification from Impulse Response, Least Squares Method (in presence of outliers)	92
4.17	Parameter Error Norm: Identification from Impulse Response, Robust Method (in presence of outliers)	92
5.1	The Mixed Gaussian pdf	107
5.2	Parameter Error Norm: Least Squares with no Outliers in the Data	107

5.3	Parameter Error Norm: Least Squares with Outliers in the Data	108
5.4	Parameter Error Norm: Robust Algorithm (Bayesian Approach) with Outliers in the Data	108
5.5	Variation of Reactor Wall Temperature Over 8 Hours	109
5.6	Variation of Percentage Butane Conversion Over 8 Hours	109
5.7	Residuals of the Least Squares Model	110
5.8	Residuals of the Robust Model	110
6.1	Parameter Error Norm: Robust Bootstrap (in Presence of Outliers)	128
6.2	State Error Norm: Robust Bootstrap (in Presence of Outliers)	128
6.3	Parameter Error Norm: Robust Bootstrap (in Presence of no Outliers)	129
6.4	State Error Norm: Robust Bootstrap (in Presence of no Outliers)	129
6.5	Parameter Error Norm: Non Robust Bootstrap (in Presence of no Outliers)	130

6.6	State Error Norm: Non Robust Bootstrap (in Presence of no Outliers)	130
6.7	Parameter Error Norm: Non Robust Bootstrap (in Presence of Outliers)	131
6.8	State Error Norm: Non Robust Bootstrap (in Presence of Outliers)	131
7.1	Regions in the Z-Plane where the Eigenvalues of the Discrete-time Model of a Stable System Must Occur for Different Sampling Intervals	169
7.2	Parameter Error Norm for Different Sampling intervals (Discrete-time Identification)	170
7.3	Parameter Error Norm for Different Sampling intervals (Direct Method)	170
7.4	Mapping from the Z-Plane to W-Plane	171
7.5	Mapping from the Z-Plane to P-Plane for Different Contraction Factors	172
7.6	Number of Terms Required in the Series Expansion in the P Domain with Contraction Factor 0.95 for 5 Decimal Place Accuracy	172

LIST OF TABLES

TABLE	PAGE
3.1 Results of MML Estimation	34
3.2 Results of Identification Using Ordinary Maximum Likelihood and Generalized Least Squares in Presence of Outliers	34
4.1 Result of Identification of the Packed-bed Reactor Using Least Squares Method	82
4.2 Result of Identification of the Packed-bed Reactor Using Robust (Weighted Least Squares Type) Method	82
4.3 Result of Identification of the Packed-bed Reactor Using IV Method	83
4.4 Result of Identification of the Packed-bed Reactor Using Robust IV Method	83
5.1 Result of Identification of the Packed-bed Reactor Using the Least Squares Method and the Robust Algorithm Based on Mixed Gaussian.pdf.	106
6.1 Comparison of Robust and Non Robust Bootstrap Method at SNR = 10 dB. Number of Iterations = 500	127

7.1	Results of Parameter Estimation for Different Sampling Intervals (Discrete - time Model)	164
7.2	Results of Parameter Estimation for Different Sampling Intervals (Continuous - time Model)	165
7.3	Estimated Poles in the P-Plane with $T=0.01$ Secs. and Contraction Factor = 0.93	166
7.4	The Optimum Value of the Sampling Interval in Different Cases	166
7.5	Estimated Poles in the Z-Plane Model with $T=0.01$ Secs.	167
7.6	CPU Time (Seconds) Requirements for Different Procedures of Identification	167
7.7	Comparison of Various Transformation Algorithms in Terms of Accuracy and Computational Requirements	168

CHAPTER 1

INTRODUCTION

1.1 Introduction to the Problem

Identification, which is the estimation of system parameters and characteristics, is a very important field in modern control theory. Besides the vital role in control and systems theory, identification finds application in a series of other areas like econometrics, social sciences, physical sciences and bio-engineering. Due to this reason, during the past three decades there has been a flurry of research activity on the subject from different perspectives. This can be substantiated by the numerous papers (eg. the survey papers by Cuenod and Sage, 1968; Eykhoff, 1968; Astrom and Eykhoff, 1971; Nieman et al., 1971; Gustavsson, 1972; Sinha and Sen, 1973; Iserman et al., 1974; Saridis, 1974(a), 1974(b); Rajbman, 1976; Rajbman and Sinha, 1977; Sinha, 1977; El-Sherief and Sinha, 1979; Young, 1981) and books (eg. Desai and Lalwani, 1972; Graupe, 1972; Sage and Melsa, 1972; Mendel, 1973; Eykhoff, 1974; Mehra and Lainiotis, 1976; Sorenson, 1980; Sinha and Kuszta, 1983; Ljung and Soderstrom, 1983; Goodwin and Sin, 1984; Young, 1984). In spite of the vast literature available on the subject, it is interesting to note that many people still consider the area of identification to be

a "bag of tricks" or a "fiddler's paradise" rather than a theory. This stems down from the fact that no identification method can be applied to a variety of situations without different approximations or modifications to improve the performance.

Most of the identification schemes employ samples of input-output data of the system to be identified. This is because of the current availability of digital computers. Basically, system identification consists of three steps: structure determination, parameter estimation and model validation (Box and Jenkins, 1970; Sinha and Kuszta, 1983). In the first step one tries to find the order of the model (discrete or continuous) by which the system can be closely represented. The second step is the application of a suitable algorithm to estimate the parameters of the model as accurately as possible. The last step is the application of some criterion (eg. the Akaike information criterion) to check how closely the fitted model represents the system under consideration.

From the point of view of computation, identification schemes can be classified into two categories: off-line and on-line algorithms. During recent times, on-line algorithms have assumed greater importance because of their extensive applications in adaptive control and due to the availability of inexpensive microprocessors.

Normally one would use least squares or its

generalizations for obtaining optimal parameter estimates. But these methods are very sensitive to deviations from the probability distribution laws based on which they are derived. In the presence of bad data, failures, gross errors or 'outliers' as they are called in statistics, these methods cease to work. In engineering, real-life data quite often contain outliers due to transducer failures, analog to digital conversion errors, data transmission problems and large disturbances. Hence it is very important to modify the existing algorithms so that they are insensitive to outliers and at the same time efficient with respect to a class of probability distribution laws. Such algorithms are called 'robust identification algorithms'. From the point of various aspects explained above, robust algorithms are more realistic for practical applications than their non-robust counterparts.

Another important factor to be considered is the issue of numerical robustness in identification. By this, one means the sensitivity to truncation and round-off errors in computers due to their finite word length. This aspect is especially serious in the identification of continuous-time systems from the samples of input-output data where improper selection of sampling intervals can create severe numerical problems.

The various problems elucidated above require considerable study, both from theoretical as well as practical considerations. This thesis is an endeavour to

illustrate the gravity of these problems and some possible solutions to them in the context of linear systems.

1.2 Organization of the thesis and original contributions

The thesis deals with robust off-line and on-line algorithms applicable to a variety of cases including joint parameter-state estimation. Besides, numerical problems associated with the identification of continuous-time systems are also considered from the point of view of the choice of the sampling interval and schemes for transformations from discrete to continuous-time models.

Chapter 2 is a survey of robust statistical methods and their applications to engineering problems. All the important robust statistical procedures have been covered and critically evaluated.

In chapter 3, a robust off - line identification method, called the 'modified maximum likelihood' method, is introduced. This method is based on censored samples of order statistics. The important properties of this powerful method have been exposed along with simulation studies.

Chapter 4 contains a thorough discussion of a class of robust recursive algorithms based on a well known statistical procedure called M-Estimation. These algorithms are the robust versions of recursive least squares, recursive instrumental variables, recursive correlation and impulse response methods. The convergence and efficiency of the proposed algorithms have been theoretically established

and their practical usefulness is illustrated through several simulation studies. In addition to that, a real-life identification problem is also considered.

Chapter 5 deals with a new robust recursive algorithm derived following the Bayesian approach using mixed Gaussian probability density functions. The efficiency and convergence of the algorithm are quantitatively investigated. Simulation studies and a practical application of the algorithm are also included.

In chapter 6, a 'robust bootstrap algorithm' is introduced for the joint parameter-state estimation. This is basically an extension of the scheme introduced in the previous chapter. The necessary and sufficient conditions for the convergence of the robust bootstrap method are obtained. Once again, the powerful features of the proposed method are shown through simulation studies.

Chapter 7 is about numerical robustness issues in the identification of 'continuous-time' systems. Numerical difficulties due to the choice of very small sampling intervals are illustrated and possible solutions are suggested for overcoming this problem. Numerically robust algorithms for the transformation of discrete-time models to continuous-time models are also introduced. These methods require fewer arithmetic operations and avoid matrix inversions in contrast with other existing techniques.

Conclusions and suggestions for future investigations

are delineated in chapter 8.

The significant contributions made in this thesis are enumerated below.

1. A thorough survey has been made on robust statistical principles and their applicability in system identification is critically analyzed.

2. A powerful robust off-line identification method is derived using censored samples of order statistics in conjunction with the maximum likelihood method.

3. Several robust on-line algorithms (the choice among them depends on the context of application, which will be clear from subsequent chapters) have been formulated and their convergence and efficiency have been analyzed. These include

(a). Derivation of a weighted least squares type algorithm based on M-Estimation and obtaining necessary and sufficient conditions for its convergence in quadratic mean. Also, evaluation of the efficiency of this type of algorithms based on 'scalar Fisher information' in a very general framework.

(b). Extension of the above idea to the recursive instrumental variables method and proving the necessary and sufficient conditions for the convergence of the algorithm in probability.

(c). Extension of the approach depicted in (a) to recursive correlation and impulse response methods of identification and evaluation of their convergence.

(d). Derivation of a robust recursive algorithm following the Bayesian approach using mixed Gaussian probability density functions and obtaining necessary and sufficient conditions for its convergence in quadratic mean. In addition, evaluation of its efficiency.

4. The proposed on-line identification methods have been successfully applied for the identification of a packed-bed catalytical chemical reactor where the conventional methods were ineffective due to the presence of undetectable outliers. Thus the practical usefulness of robust algorithms is demonstrated.

5. A robust pseudo linear regression algorithm called 'robust bootstrap algorithm' is derived for the joint estimation of parameters and states. Necessary and sufficient conditions are obtained for the convergence of the algorithm with probability one, using martingale theory.

6. Numerical difficulties associated with the selection of very small sampling intervals for the identification of continuous-time systems have been studied. Methods have been suggested to solve this problem.

7. Numerically robust and computationally simple methods have been proposed for the transformation of discrete-time models to their continuous-time equivalents. These methods are compared with other existing algorithms.

CHAPTER 2

A SURVEY OF ROBUST STATISTICAL METHODS

2.1 Overview of the chapter.

A survey of different robust statistical procedures is presented. Their applicability in the areas of signal processing and system identification is investigated.

2.2 History and development of robust statistical methods

There have been a lot of investigations in the field of robust statistical procedures for the past two decades and even now it is going on very vigorously. However, the history of robust statistical methods is about 180 years old and inseparable from classical statistical approaches.

To the best of the author's knowledge, it was Legendre (1805) who first pointed out that bad pieces of data have very detrimental effects on least squares estimates. Later, Laplace (1818) showed that least absolute deviation regression models are less susceptible to spurious data compared to least squares regression models. In fact, the probability density function (pdf) originated from this concept is now named after him, in contrast with the Gaussian pdf which gives least squares estimates. Similar results were obtained by Edgeworth (1886), who showed that sample median (derived from least absolute deviations) is a

more robust estimate of location than sample mean (derived from least squares).

The normality assumption in statistical inference was first attacked vehemently by Pearson (1931). Eden and Yates (1933) and Tiku (1971) give real-life examples which question the validity of Fisher's approach of maximum likelihood estimation, assuming normality on data.

It seems that Poincaré (1912) is the one who first proposed the idea of censoring extreme observations before estimating the parameters. Apparently, a similar idea was used by Mendeleev in 1895 indirectly, as pointed out by Eisenhart (1971) but the evidences are not concrete. The use and asymptotic normality of linear combinations of order statistics were first introduced by Daniel and later expounded by Godwin (1948), Lloyd (1952), Tiku (1967), Stigler (1974) and several other authors.

The Princeton studies led by Tukey gave a significant momentum to robust statistics (see Tukey, 1967). Other interesting advances during this period are due to Hodges and Lehman (1963) and the estimates proposed by them are now known after their names. These estimates are based on rank tests.

One of the most popular methods is that proposed by Huber (1964) using a minimax approach. Most of the applications of robust methods in signal processing are based on this method. This trend can be easily understood

from a very recent survey by Kassam and Poor (1985).

Systematic methods to detect outliers were first proposed by Pierce (1892). Some of the notable results in this aspect are due to Dixon (1953), Anscombe (1960), Daniel (1960) and Cook and Prescott (1981).

Robustness aspects in time series analysis have received a lot of attention after the work by Abraham and Box (1975).

The discussions above give a rough idea of the important milestones in the history of robust statistics. In the next section we briefly consider some important robust statistical methods, currently in use.

2.3 Robust methods of estimation of a location parameter

Here we attempt to estimate the location parameter μ described by the equation

$$y(t) = \mu + e(t) \quad (2.3.1)$$

where $t=1,2,\dots,n$. The sequence $\{y(t)\}$ represents the samples of the measured data and $\{e(t)\}$ is the error sequence of independent and identically distributed random variables, often assumed to be of zero mean and having a symmetric unimodal probability density function. If we assume normality on $\{e(t)\}$, the maximum likelihood estimate of μ coincides with the least squares estimate, which is nothing but the sample mean.

$$\text{ie., } \hat{\mu}_{LS} = (1/n) \sum_{t=1}^n y(t) \quad (2.3.2)$$

This estimate of μ is highly non robust since it is affected very much by extreme observations. Therefore one has to seek for other estimates, like those given below, to ensure a good degree of protection against outliers.

2.3.1 Maximum likelihood type estimates (M-Estimates)

Any estimate originated from the minimization problem of the form

$$\sum_{t=1}^n \rho(y(t)-\mu) = \min! \quad (2.3.3)$$

or by the solution of an implicit equation

$$\sum_{t=1}^n \psi(y(t)-\mu) = 0 \quad (2.3.4)$$

where $\rho(y,\mu)$ is an arbitrary cost function and

$$\psi(y,\mu) = (\partial/\partial\mu)[\rho(y,\mu)]$$

is called an M-Estimate. Note that the choice $\rho(y,\mu) = -\log[f(y,\mu)]$ where $f(\cdot)$ is the probability density function of $e(t)$ gives the ordinary maximum-likelihood estimate of μ .

M-Estimates are further discussed in chapter 4. More details on this subject in a compiled form can be obtained from the book by Huber (1981).

2.3.2 Estimates by linear combinations of order statistics
(L-Estimates)

Let $y(t)$ represents the t^{th} order statistics of observations; $t=1,2,\dots,n$. Define $h(z)$ such that $h(z) > 0$

$$h(z) = h(1-z) \text{ and}$$

$$\int_0^1 h(z) dz = 1; z \in [0,1].$$

Also define the weights $w(t) = h[t/(1+n)]$, $t=1,2,\dots,n$.

Then, the L - Estimate is obtained as the weighted sum of order statistics:

ie.,
$$u_L = (1/n) \sum_{t=1}^n w(t)y(t) \tag{2.3.5}$$

A very similar method to construct best linear unbiased estimates (BLUE) using order statistics, applying Gauss Markov theorem can be found in Balakrishnan and Puthenpura (1985a, 1985b). Further discussion on L-Estimates can be seen in a number of papers. Some of them are Chernoff et al. (1967), Stigler (1969, 1974) and Huber (1981).

2.3.3 Modified maximum likelihood estimates based on censored samples (MML Estimates)

This method, introduced by Tiku (1967), is similar to the estimation procedure explained in the previous section. Here r_1 small and r_2 large observations are censored among n order statistics of observations. The probability density

function of the censored samples is used as the likelihood function to estimate the location parameter. Computations are made simple by approximating $f(x)/F(x) = \alpha + \beta x$ where $f(\cdot)$ and $F(\cdot)$ are the probability density function and the cumulative probability function of x , respectively. The coefficients α and β depend on the degree of censoring and can be easily obtained (Tiku, 1967). In the next chapter MML estimation will be considered in detail.

2.3.4 Estimates using rank tests (R-Estimates)

Define $J(z)$ which is arbitrary monotone increasing and such that $J(z) = -J(1-z)$. Generate $2n$ numbers, $\{-y(t)+u\}$ and $\{y(t)-u\}$ where $t=1,2,\dots,n$, for an arbitrary 'slippage parameter' u and order them. Set $w(t)=1$ if the t^{th} element in the ordered sequence is of the form $\{y(t)-u\}$ and set $w(t)=0$ if the t^{th} element has the form $\{-y(t)+u\}$. The R-estimate is the solution of the equation;

$$R(u) = \sum_{t=1}^n w(t)J[t/1+n] = 0 \tag{2.3.6}$$

ie., the estimate which makes a rank test statistic vanish. A detailed and comprehensive discussion on R-estimates can be found in Chernoff and Savage (1958), Hajek (1968), Huber (1981) and Puri and Sen (1985).

It is important to note that the estimate given by equation (2.3.6) is asymptotically equivalent to the famous estimates of Hodges and Lehman (1963) which is given by

$$\mu_{HL} = \text{med}[y(t)+y(s)]/2. \quad (2.3.7)$$

$t > s$ (med represents 'median')

Jurečková (1977) established the asymptotic equivalence of M and R-estimates.

The advantage of this method is that it does not demand any distribution assumptions. But the heavy computational requirements make it unattractive for practical implementation.

2.3.5 Trimmed and Winsorized means

Let $y_{(t)}$ be the t^{th} order statistics of n observations. To obtain a trimmed mean, we discard qn ($0 < q < 1$) largest and smallest observations and get the mean of the remaining samples.

$$\text{i.e., } \mu_{TM} = \frac{1}{n-2qn} \sum_{t=qn+1}^{n-qn} y_{(t)} \quad (2.3.8)$$

Suggested references in this topic are Dixon (1960), Tukey (1967) and Rey (1983).

In the case of Winsorized means (named after C.P. Winsor), the extreme values at each ends of the ordered samples are not discarded, but projected on to the remaining part of the ordered sample. The heuristic idea behind this is that we are not completely disregarding the information contained in the samples which are censored off.

Thus the expression for the Winsorized mean is

$$\hat{\mu}_W = [1/n] \left[\sum_{t=q_n+2}^{n-q_n-1} y_{(t)} + q_n(y_{(q_n+1)} + y_{(n-q_n)}) \right] \quad (2.3.9)$$

Bickel (1965), Tukey (1967) and Hogg (1979) give sufficient details on this topic.

2.3.6 Other estimates

There are several other robust methods, which have been introduced during the past fifteen years, but they are not as well established as the methods explained in the previous sections. However, for the sake of completeness, some of these methods are very briefly mentioned below.

Hogg (1974) introduced a class of estimates called 'adaptive estimates'. These are estimates which adapt to the underlying probability distribution. At this time, the connection between adaptivity and robustness is far from clear and further research is required in this area.

Other two existing robust schemes are the bootstrap (Effron, 1979, 1980) and jack-knife (Quenouille, 1956; Effron, 1980) methods. A critical evaluation of these methods is presented in the book by Rey (1983).

2.4 Numerical studies on robustness

A classic work in this topic is the book by Andrews et al. (1972). They consider about 72 different robust estimates over 20 distributions. Numerical comparison of R-estimates and M-estimates are given by Leon et al. (1967). Similar results can be found in Tukey and McLaughlin (1963).

and Hodges (1967). Computational results on M-estimation using the simplex algorithm are presented in the book by Bierens (1981). Another important paper in the context of numerical robustness is the one by Gastwirth and Cohen (1970) where a very good investigation is made of the various estimates explained in the previous sections, from the point of view of their numerical properties. Numerical aspects on multivariate populations have been discussed by Box (1953), Olson (1974) and Pillai and Sudjana (1975).

2.5 Applications of robust method in engineering

Probably the first reported application of robust methods in engineering is due to Newcomb (1886), who pointed out the effects of anomalous measurements in astronomical data and suggested possible remedies to overcome this difficulty. His approaches were conceptually and computationally very similar to modern methods. The landmark paper of Stigler (1977) illustrates the practical usefulness of robust methods. In this section we concentrate on only two potential cases of application of robust procedures - (1) filtering and signal processing and (2) system identification.

2.5.1 Filtering and signal processing

Considerable amount of work has been done in this area as suggested by a very recent survey by Kassam and Poor (1985). This survey is very thorough and the reader's

attention is directed towards it, rather than making an attempt to conduct another survey on this subject. However, some of the important referenced are mentioned below.

Robust estimation of power spectra is discussed by Klefner et al. (1979), Martin and Thomson (1982) and Vastola and Poor (1984).

Construction of robust Weiner filters is explored in a number of papers. Some important ones are Kassam and Lim (1979), Poor (1980), Cimini and Kassam (1981), Moustakides and Kassam (1983) and Vastola and Poor (1983, 1984).

Robustification of the Kalman filter has also been subjected to considerable study. Some notable results in this area are due to Masreliez and Martin (1977), Morris (1976), Boncelet and Dickinson (1983) and Tsai and Kurz (1983).

Robust estimation of signal amplitude is discussed in the papers by Martin (1972), Martin and Masreliez (1976) and Price and Vandelinde (1979).

A very interesting application of order statistic filters in sonar data processing (which is basically a robust noise power estimation problem) is well explained in a recent work by Wong (1986). The most attractive feature of this work is the introduction of a robust generalized residual filter which is a substantial improvement over its counter parts.

2.5.2 System Identification

Unlike filtering and signal processing, reported applications of robust principles to system identification are relatively few.

The most notable work in this area is due to Poljak and Tsytkin (1980) where they consider minimax recursive estimators. Basu and Vandelinde (1977) discuss a robust recursive correlation technique for system identification. Bierens (1981) demonstrates the application of robust parameter estimation methods in some real-life cases but through a laborious solution of non-linear equations. Some theoretical approaches to deal with contaminated normal probability distributions are discussed by Papantoni-Kazakos (1977). Robust identification of multivariable systems is considered by Kovacevic and Stancovic (1979, 1982 and 1984) using iterative gradient methods (like Gauss-Newton and Newton-Raphson methods) and robust estimation of time series autoregressions is analyzed by Martin (1979). Agee and Turner (1979) illustrate the use of robust regression in deducing inference from the measurements of a cinetheodolite.

Robust estimation via stochastic approximation is well explained in the paper by Martin and Masreliez (1975). In the book by Ljung and Soderstrom (1983) some aspects are mentioned about the choice of cost functions for robust identification, along with several simulation results. A few theoretical guidelines on the robust estimation of time varying systems are mentioned in a very recent paper by

Kovacević and Filipovic (1985).

2.6 Concluding remarks

Even though robust methods are about two centuries old, most of the important advancements in this field have been made during the past two decades. They find new horizons in signal processing and system identification. Actual experimental data often contain outliers, hence robust procedures are essential in these contexts. The most widely used method in engineering applications is the M-estimation, others are less popular due to their computational complexity. But with the present trends in computers, one can expect that other robust methods will also become computationally affordable and hence very popular in the near future.

CHAPTER 3
ROBUST OFF-LINE IDENTIFICATION USING
MODIFIED MAXIMUM LIKELIHOOD (MML) METHOD

3.1 Overview of the chapter

A robust off-line identification method is presented. This is based on censoring ordered data and employing the maximum likelihood method. The proposed method is compared with the ordinary maximum likelihood and generalized least squares methods using a simulated example.

3.2 The modified maximum likelihood (MML) method

The MML estimation method introduced by Tiku (1967) works as follows.

Suppose we have n samples of an observation arranged in the ascending order;

$$y_1, y_2, y_3, \dots, y_n \quad (3.2.1)$$

so that

$$y_1 \leq y_2 \leq y_3 \leq \dots \leq y_n \quad (3.2.2)$$

We assume that these n samples are from the Gaussian distribution given by

$$f(y) = [1/\sqrt{2\pi}\sigma] \exp[-(y-\mu)^2/(2\sigma^2)] \quad (3.2.3)$$

Now out of these n samples $r=n_1$ smallest and $r=n_2$ largest

are censored (symmetric censoring) where $0 < q < 1$. The remaining observations in order of magnitude are

$$y_{r+1}, y_{r+2}, y_{r+3}, \dots, y_{n-r} \quad (3.2.4)$$

The probability density function of this symmetrically censored samples of size $n-2r$ can be written as (Tiku, 1967)

$$[n!/(r!)^2] [1/\sqrt{2\pi}\sigma]^{(-n+2r)} \exp[-(1/2) \sum_{i=r+1}^{n-r} z_i^2] G_1 G_2 \quad (3.2.5)$$

where

$$G_1 = [1 - F(z_{r+1})]^r, \quad G_2 = [F(z_{n-r})]^r, \quad z_i = (y_i - \mu)/\sigma \quad (3.2.6)$$

and

$$F(z) = \int_{-\infty}^z \{ [1/\sqrt{2\pi}\sigma] \exp[-(y-\mu)^2/(2\sigma^2)] \} dt \quad (3.2.7)$$

which is the cumulative density function of Gaussian. In the MML estimation procedure, (3.2.5) is taken as the likelihood function.

Let us recall the way we obtain maximum likelihood estimators (see for example Sorenson, 1980). Suppose $y_1, y_2, y_3, \dots, y_M$ are M independent samples from a Gaussian distribution $N(\mu, \sigma)$ where μ and σ are unknown. The likelihood function is

$$f(y_1, y_2, y_3, \dots, y_M | \mu, \sigma)$$

which is equal to

$$[1/2\pi\sigma^2]^{-M/2} \exp[-(1/2\sigma^2) \sum_{i=1}^M (y_i - \mu)^2]. \quad (3.2.8)$$

Now we take the logarithm of (3.2.8) on both sides and set

$$[\partial/\partial\mu] \ln[f(y, \mu, \sigma)] = (1/\sigma^2) \sum_{i=1}^M (y_i - \mu) = 0 \quad (3.2.9)$$

which gives

$$\mu_{ML} = (1/M) \sum_{i=1}^M y_i \quad (3.2.10)$$

Similarly

$$[\partial/\partial\sigma] \ln[f(y, \mu, \sigma)] = 0 \quad (3.2.11)$$

gives

$$\sigma_{ML}^2 = (1/M) \sum_{i=1}^M (y_i - \mu_{ML})^2 \quad (3.2.12)$$

In a similar fashion, Tiku (1967) derives the MML estimates of the location and scale parameters (μ and σ) of the sample as,

$$\mu_{MML} = (1/d) [\sum_{i=r+1}^{n-r} y_i + rS(y_{r+1} + y_{n-r})] \quad (3.2.13)$$

and

$$\sigma_{MML}^2 = \{B + \sqrt{B^2 + 4AC}\} / \{2\sqrt{A(A-1)}\} \quad (3.2.14)$$

where

$$d=n-2r+2r\beta, \quad A=n-2r, \quad B=rc[y_{n-r} - y_{r+1}]$$

and

$$C = \sum_{i=r+1}^{n-r} y_i^2 + r\beta(y_{r+1}^2 + y_{n-r}^2) - du_{MML}^2 \quad (3.2.15)$$

by making an approximation, $f(z)/F(z) = a + \beta z$. The coefficients a and β are given by Tikku (1967) for different degrees of censoring. However, for large values ($n > 10$) of n these coefficients can be obtained in a simple manner as given below.

$$\begin{aligned} \beta &= -f(t)[t-f(t)/q]/q \\ a &= [f(t)/q] - \beta t \end{aligned} \quad (3.2.16)$$

where

$$F(t) = \int_0^t f(z) dz = 1 - q$$

and

$$[1/\sqrt{2\pi}] \exp(-z^2/2), \quad -\infty < z < \infty \quad (3.2.17)$$

Also recall that $q=r/n$

Note that $0 < a < 1$ and $0 < \beta < 1$. As an example, for $n=10$ the variation of a and β for different values of q is given in fig. 3.1.

3.3 Application of MML method in system identification

3.3.1 Introduction

The MML method can be applied very successfully for the identification of linear time invariant systems

(Puthenpura and Sinha, 1985a).

Assume that we have N samples of input and output of a linear, time invariant single input single output system described by,

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx + e(t) \end{aligned} \tag{3.3.1}$$

where $x \in R^n$ and $y, u \in R$. A, B and C are $n \times n$, $n \times 1$ and $1 \times n$ matrices respectively. The quantity $e(t)$ is the contaminating noise, assumed to be normally distributed with zero mean and variance σ^2 . It is important to note that the system should be completely observable, the input $u(t)$ should persistently excite the system and the sampling interval is selected properly (see chapter 7).

Let the discrete-time transfer function model of the system (3.3.1) be

$$\frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + a_3 z^{-3} + \dots + a_n z^{-n}} \tag{3.3.2}$$

where z^{-1} is the backward shift operator. Now let there be $K+1$ samples of input and $(K+1)p$ samples of output are available, so that p readings of each of $y(k)$, $k=0,1,2,\dots,K$ are taken. Hence the data available are $\{u(k)\}$ and $\{y_i(k)\}$; $k=0,1,2,\dots,K$ and $i=1,2,\dots,p$. Thus we can write (Ljung and Soderstrom, 1983 - also see section 4.3)

$$y_i(k+1) = \phi_i^T(k) \theta + e_i(k) \tag{3.3.3}$$

where

$$\phi_i^T(k) = [u(k), u(k-1), \dots, u(k-m), -y_i(k-1), -y_i(k-2), \dots, -y_i(k-n)]$$

$$\text{for } k=0, 1, 2, \dots, K; \quad i=1, 2, \dots, p. \quad (3.3.4)$$

and

$$\theta^T = [b_0, b_1, \dots, b_m, a_1, a_2, \dots, a_m] \quad (3.3.5)$$

We apply the MML estimation to the linear model given by (3.3.3). Our aim is to get a robust estimate of θ ; i.e., the best estimate of θ for which $y_i(k)$ are most likely to occur even for considerably large deviations of $y_i(k)$ from the assumed probability density function.

3.3.2 Formulation of the off-line algorithm

Here we take the likelihood function $L[y_i(k); \theta]$ to be as in (3.2.5) and perform a maximum likelihood estimation.

Thus

$$[\partial/\partial \mu]\{L[y_i(k); \theta]\} = 0; \quad [\partial/\partial \sigma^2]\{L[y_i(k); \theta]\} = 0$$

and

$$[\partial/\partial \theta]\{L[y_i(k); \theta]\} = 0 \quad (3.3.6)$$

will give the MML estimates of μ , σ^2 and θ respectively. In particular we are interested in the solution for θ . This is given by the set of equations below. The derivation of these equations is algebraically lengthy and may not be particularly interesting to the reader. Therefore, a simplified solution (of reduced dimension) is given in appendix I, which is very straightforward and easily

comprehensible.

Thus, following exactly in the same way as described in the appendix F, we obtain the MML estimate of θ as the solution of

$$\begin{bmatrix} S_{11} & S_{12} & \dots & S_{1h} \\ S_{21} & S_{22} & \dots & S_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ S_{h1} & S_{h2} & \dots & S_{hh} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_m \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \\ \vdots \\ Q_h \end{bmatrix} \quad (3.3.7)$$

where $h=m+n+1$. Also,

$$Q_j = \sum_{k=j}^K [u(k-j+1) - u_j^*] W(k) \quad \text{for } j=1, 2, \dots, m+1$$

$$Q_{j+m+1} = \sum_{k=j}^K [-y_c(k-j) - y_j^*] W(k) \quad \text{for } j=1, 2, \dots, n \quad (3.3.8)$$

$$S_{jj} = \sum_{k=j}^K [u(k-j+1) - u_j^*]^2 \quad \text{for } j=1, 2, \dots, m+1$$

$$S_{(j+m+1)(j+m+1)} = \sum_{k=j}^K [-y_c(k-j) - y_j^*]^2 \quad \text{for } j=1, 2, \dots, n \quad (3.3.9)$$

$$S_{j\ell} = \sum_{k=\max(j,\ell)}^K [u(k-j+1) - u_j^*] u(k+1-\ell) \quad \text{for } j=1,2,\dots,m+1 \\ \ell=1,2,\dots,m+1 \\ j \neq \ell$$

$$S_{(j+m+1)(\ell+m+1)} = \sum_{k=\max(j+m+1,\ell+m+1)}^K [-y_c^*(k-j) - y_j^*] y_c(k-1) \\ \text{for } j=1,2,\dots,n \\ \ell=1,2,\dots,n \\ j \neq \ell$$

(3.3.10)

where

$$u_j^* = [1/(K-j+1)] \sum_{k=j}^K u(k-j+1) \quad j=1,2,\dots,m+1$$

$$y_j^* = [-1/(K-j)] \sum_{k=j}^K \tilde{y}_c(k-j) \quad j=1,2,\dots,n \quad (3.3.11)$$

and

$$W(k) = [\sum_{i=r+1}^{p-r} y_i(k) + r\beta(y_{r+1}(k) + y_{p-r}(k))] \quad (3.3.12)$$

The variance of θ can be computed as,

$$V(\theta) = \sigma_c^2 / \{ (p-2r+2\beta) [\sum_{j=1}^h S_{jj}] \} \quad (3.3.13)$$

where σ_c^2 is computed as

$$\sigma_c = \{2\sqrt{A_1(A_1 - h - 1)}\} \sigma^*$$

$$\sigma^* = \{B_1 + \sqrt{(B_1^2 + 4A_1 C_1)}\} / (2A_1) \quad (3.3.14)$$

with

$$A_1 = (K+1)(p-2r)$$

$$B_1 = \sum_{k=0}^K r\alpha [y_{p-r}(k) - y_{r+1}(k)]$$

$$C_1 = \sum_{k=0}^K \left\{ \sum_{i=r+1}^{p-r} y_i^2(k) + r\beta (y_{r+1}^2(k) + y_{p-r}^2(k)) \right\}$$

$$- [1/(K+1)] \left\{ \sum_{k=0}^K W(k) \right\}^2 (p-2r+2r\beta)$$

$$- \sum_{j=1}^h \theta_j Q_j \quad (3.3.15)$$

Note that

$$\theta_j = b_{j-1} \quad ; \quad j=1, 2, \dots, m+1$$

$$\theta_{j+m+1} = -a_j \quad ; \quad j=1, 2, \dots, n \quad (3.3.16)$$

Note:

To start with, we take $y_c(k)$ as the mean of the censored samples.

$$i.e., \quad y_c(k) = [1/(p-2r)] \sum_{i=r+1}^{p-r} y_i(k)$$

for $k=0,1,2,\dots,K$ (3.3.17)

After getting the first estimates of θ by MML method,

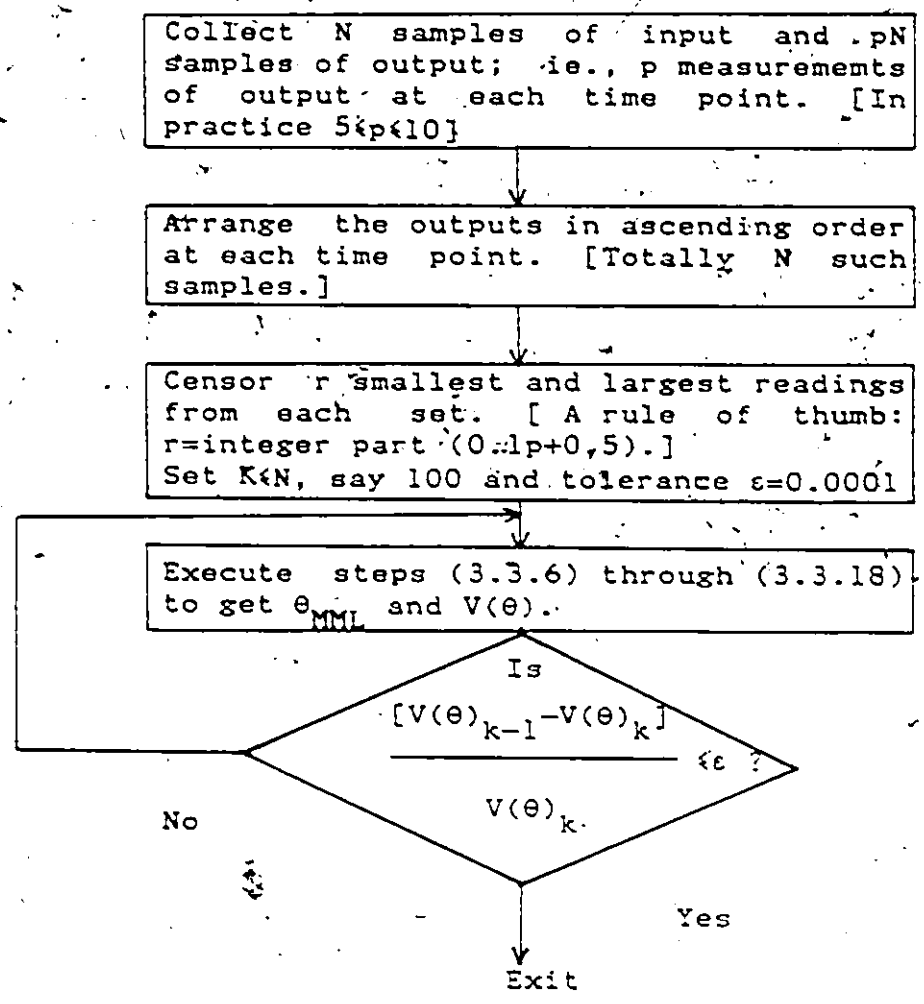
we compute

$y_c(k) = \Phi_c^T(k) \theta_{MML}$

where

$\Phi_c^T(k) = [u(k), u(k-1), \dots, u(k-m), -y_c(k-1), -y_c(k-2), \dots, -y_c(k-n)]$
for $k=0,1,2,\dots,K$ (3.3.18)

The whole procedure can be summarized as follows.



3.4 Important properties of MML Estimation

The important properties of MML estimates are discussed by Tiku (1978, 80, 81 and 82). Also Tan (1985) investigates these properties in a Bayesian approach. However, some important properties of MML method are given below without proof. For the convergence properties of MML estimation refer Tiku (1981).

1. MML method can be applied to any 'location and scale family' of probability distribution. By saying location and scale family it is meant that, if $\{x_i\}$ has a probability density function $f(x; \mu, \sigma)$ where μ and σ are the location and scale parameters, by making the transformation $z = (x - \mu) / \sigma$ the new density function $f(z)$ should be independent of μ and σ . Examples of location and scale family are (David, 1981) Gaussian, Double exponential, logistic, Gamma and Cauchy.

2. The estimators θ_{MML} and μ_{MML} are unbiased and asymptotically normal.

3. The estimators θ_{MML} , μ_{MML} and σ are considerably more efficient than other robust estimates like Hampel, wave and bisquare estimates.

4. For $p < 4$, MML estimators lose their efficiency. Hence it is desirable to have $p > 5$.

3.5 Results of simulation

Consider the third order system

$$H(s) = \frac{40(s+0.1)}{(s^2+2s+5)(s+5)} \quad (3.5.1)$$

The impulse invariant equivalent of discrete-time model of the above continuous-time domain representation can be written as,

$$H(z) = \frac{b_0 + b_1 z + b_2 z^2}{z^3 + a_1 z^2 + a_2 z + a_3} \quad (3.5.2)$$

where

$$b_0 = 0.0$$

$$b_1 = 4.7576563 \exp(-T) \cos(2T + .204185) - 4.6588235 \exp(-10T)$$

$$b_2 = 4.7576563 \exp(-11T) \cos(2T - .204185) - 4.6588235 \exp(-2T)$$

$$a_1 = -[\exp(-10T) - 2 \exp(-T) \cos(2T)]$$

$$a_2 = \exp(-2T) + 2 \exp(-11T) \cos(2T)$$

$$a_3 = -\exp(-12T)$$

and T is the sampling interval.

The above system was perturbed with a sequence of random inputs and the corresponding output sequence was calculated choosing $T=0.04$ seconds. To simulate large measurement errors, arbitrary large positive and negative numbers were also added to some of the readings and 5 values of each $y(k)$ are taken. Before using the MML method, the mean of each group of $y(k)$ were taken and usual identification methods like generalized least squares and ordinary maximum likelihood were tried. They were either

diverging or gave highly biased estimates.

The results of MML method are given in table 3.1. It is interesting to note that the estimated parameters are more or less the same for all the sets of simulated data, showing the high robustness of the method.

Note:

Here $p=5$. The value of $r = \text{int.part}[0.1 \times 5 + 0.5] = 1$. Hence $q = 1/5 = 0.2$. For this value of p and q , $a = 0.7540$ and $\beta = 0.8108$.

For the sake of comparison, the estimated parameters using the usual maximum likelihood method and the generalized least squares method (Sinha and Kuszta, 1983) are given in table 3.2. Note that even for a higher signal to noise ratio poor results are obtained due to the presence of outliers.

3.5 Concluding remarks

The method described in this chapter is purely an off-line method. The algorithm described in section 3.3 requires several samples of output at each time point of observation. This is not a major draw back, since one can employ several transducers to obtain a particular measurement (eg. thermocouples for a chemical reactor) or it may possible to perform the experiment several times for the same input. This of course, increases the the cost of identification, but definitely paid off in terms of the accuracy of estimates, as suggested by several simulated

examples.

TABLE 3.1
Results of MML estimation

Signal to Noise Ratio in dB	Estimated Parameters		Actual Parameters
	Trial 1	Trial 2	
100	1.263912 -1.261374 -2.585194 2.210001 -0.619100	1.264013 -1.264013 -2.585109 2.210002 -0.619101	
10	1.303101 -1.246713 -2.590000 2.219374 -0.619107	1.302991 -1.245910 -2.587412 2.218997 -0.619101	1.264858 -1.260129 -2.585753 2.207070 -0.618783
3	0.943721 1.703142 0.023416 -4.573121 3.671128	0.941738 1.701741 0.022999 -4.569123 3.673172	

TABLE 3.2
Result of identification using ordinary maximum likelihood
method and generalized least squares in presence of
outliers

Signal to Noise Ratio in dB	Estimated Parameters	
	Max. Likelihood	Gen. Least squares
100	0.936374 1.813437 0.099673 -3.997431 4.137384	-1.516166 0.298989 0.176134 1.228888 0.398164
10	0.999375 -4.131276 -0.001315 0.914111 -3.000134	-3.143716 1.111163 -17.319870 -0.011315 4.448187
3	17.998137 -17.183162 3.395516 -1.881376 -0.085554	-2.385167 1.287303 0.010133 -12.566371 16.001301

Note: The order in which the parameters are given above is
 b_1, b_2, a_1, a_2 and a_3 .

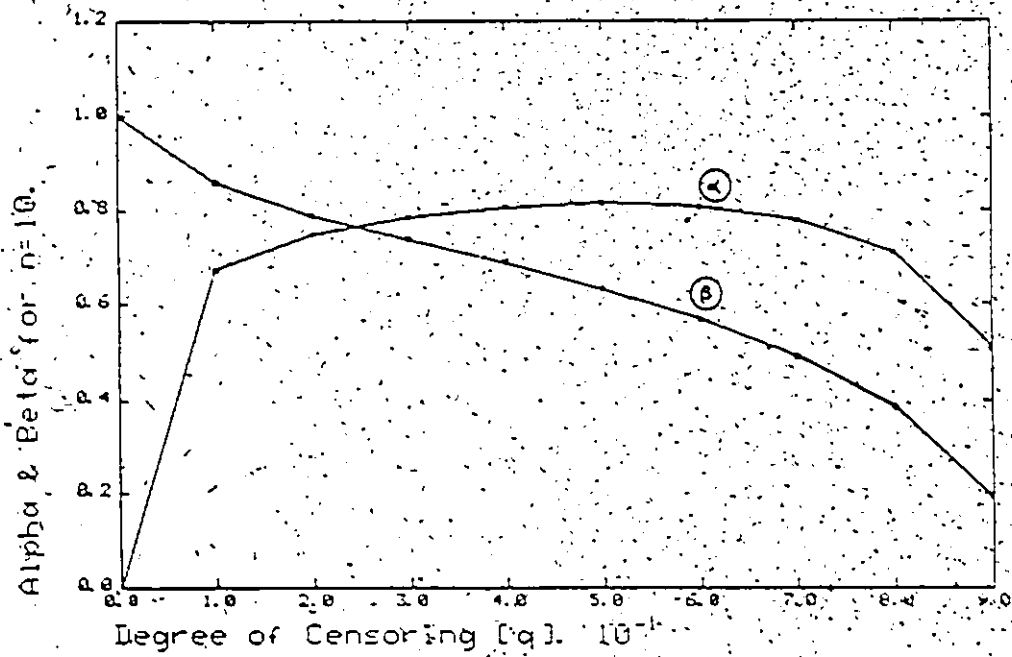


Fig: 3.1 Variation of α and β for Different Degrees of Censoring

CHAPTER 4

APPLICATION OF M-ESTIMATION IN ROBUST RECURSIVE IDENTIFICATION

4.1 Overview of the chapter

This chapter is devoted to the development and analysis of a class of robust recursive identification algorithms, based on the M-Estimation method. To begin with, a weighted least squares type of algorithm is introduced (Puthenpura et.al., 1985) which is extended to the recursive instrumental variables (Puthenpura and Sinha, 1985b), recursive correlation (Puthenpura et.al; 1985) and impulse response methods. The real-life identification problem clearly points out the relevant features of robust system identification.

4.2 A review of M-Estimation (of location)

As mentioned earlier in section 2.3.1, the concept of M-estimation is explained in a somewhat detailed manner here.

Let X_1, X_2, \dots, X_n be a random sample, independently and identically distributed, drawn from a distribution with density $f(x-\theta)$ of the continuous type, where θ is the location parameter. Recall that any location and scale probability density function can be expressed as

$$[1/\sigma]f[(x-\theta)/\sigma],$$

where σ is called the scale parameter. Here, for simplicity we consider the scale known case and without losing generality, the scale parameter can be assumed to be unity. The problem is to estimate the location parameter, θ . We proceed as follows. The logarithm of the likelihood function is

$$\ln[L(\theta)] = \sum_{i=1}^n \ln[f(x_i - \theta)] \quad (4.2.1)$$

Denoting

$$\ln[f(x)] = -\rho(x) \quad (4.2.2)$$

We get

$$\ln[L(\theta)] = - \sum_{i=1}^n \rho(x_i - \theta) \quad (4.2.3)$$

In the method of maximum likelihood, we wish to maximize $L(\theta)$ or $\ln[L(\theta)]$, equivalently minimize the cost function

$$J(\theta) = \sum_{i=1}^n \rho(x_i - \theta) \quad (4.2.4)$$

Suppose that this minimization can be achieved by setting

$$[\partial/\partial\theta]J(\theta) = 0 \quad (4.2.5)$$

We get the estimate of θ as the solution of

$$\sum_{i=1}^n \psi(x_i - \theta) = 0 \quad (4.2.6)$$

where

$$\psi(x) = \rho'(x) = -f'(x)/f(x) \quad (4.2.7)$$

The estimate of θ obtained by solving eqn. (4.2.6) in general (for an arbitrary ψ function) is called the M-estimator (maximum likelihood type estimator) as depicted in section 2.3.1 and the function ψ is called the 'influence function'. Note that the particular choice (4.2.7) of influence function reduces the M-estimator to be the conventional maximum likelihood estimator.

Examples

(a) $f(x)$ is normal with unit scale

This gives us

$$\rho(x) = (1/2)x^2 + c, \text{ where } c \text{ is some constant and } \psi(x) = x.$$

$$\text{i.e., } \sum_{i=1}^n \psi(x_i - \theta) = 0 \Rightarrow \hat{\theta} = \bar{x} \text{ (sample mean)} \quad (4.2.8)$$

[$\hat{\theta}$ indicates the estimate]

(b) $f(x)$ is Laplacian (double exponential) with unit scale

Here

$$\rho(x) = |x| + c, \text{ where } c \text{ is some constant and}$$

$$\psi(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ +1, & x > 0 \end{cases} \text{ (signum function).}$$

$$\text{i.e., } \sum_{i=1}^n \psi(x_i - \theta) = 0 \Rightarrow \hat{\theta} = \text{sample median.} \quad (4.2.9)$$

From these examples one can conclude that the ψ function of example (b) is bounded and the corresponding estimate of θ is not influenced by outliers. Therefore we say that this estimator of θ is robust. On the other hand, in example (a), the ψ function is unbounded and as a result the estimate of θ is greatly influenced by extreme observations. In other terms, the least squares estimate of θ , obtained by assuming normality on $\{x_i\}$ is not robust.

Thus in robust M-estimation we look for a ψ function such that the resulting estimator is protected from outliers. Besides, the efficiency of the resulting procedure should be reasonably good when the data actually enjoys normality. It will be shown later that the premium we pay for the robustness is a little loss of efficiency in the normal case. However, when the amount of data available is not small, a little loss of efficiency is not a major draw back.

4.3 Introduction to recursive identification

Let us rewrite eqn. (3.3.3) as,

$$\frac{y(t)}{u(t)} = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \dots + a_n z^{-n}} \quad (4.3.1)$$

where $\{u(t)\}$ and $\{y(t)\}$ are the input and output sequence respectively. Here t represents the discrete-time index and z^{-1} is the backward shift operator. In time series analysis the model represented by eqn. (4.3.1) is referred as an autoregressive moving average (ARMA) model. In a recursive identification scheme, the parameters of (4.3.1) are sequentially estimated. The major reason for the popularity of recursive identification schemes is their extensive applicability in adaptive control, adaptive filtering, adaptive prediction and adaptive signal processing problems. In addition to that, recursive identification schemes are computationally simple and demand less computer memory, in contrast with their off-line counterparts. There are two disadvantages of recursive identification in comparison with off-line identification. One is that the decision of model structure should be made a priori, before starting the recursive procedure. The second disadvantage is the lesser accuracy of estimates compared to off-line methods.

The most widely used recursive identification method is the 'equation error approach'. Note that the linear model (4.3.1) can be put as

$$y(t) + a_1 y(t-1) + \dots + a_n y(t-n) = b_0 u(t) + b_1 u(t-1) + \dots + b_m u(t-m) + e(t) \quad (4.3.2)$$

where $e(t)$ is some disturbance of unspecified character. Often it is convenient to express (4.3.2) in the form of a regression equation

$$y(t) = \phi^T(t)\theta + e(t) \quad (4.3.3)$$

$$t=1,2,\dots,N$$

where

$$\phi^T(t) = [-y(t-1), \dots, -y(t-n), u(t), \dots, u(t-m)] \quad (4.3.4)$$

called the observation vector

and

$$\theta^T = [a_1, \dots, a_n, b_0, \dots, b_m] \quad (4.3.4)$$

called the parameter vector, which is to be estimated. Throughout our discussions, it will be assumed that (i) correct model order has been selected, (ii) the input is of persistent excitation and (iii) and the system is completely observable.

Since the noise term $e(t)$ in (4.3.3) corresponds to an 'equation error' in the difference equation (4.3.2), methods to estimate θ in (4.3.3) are called 'equation error methods'. Usually one would use least squares or its generalizations for this purpose (Ljung and Soderstrom, 1983) assuming $\{e(t)\}$ to be a Gaussian white noise sequence or it can be derived by passing a Gaussian white noise sequence through some ARMA filter. The least squares method is optimal in the mean square sense in the case when $\{e(t)\}$ is normal. In the presence of outliers, the least squares method can be far from satisfactory. There is no doubt that real-life data often contain outliers. Therefore, robust estimation procedures are quite useful to handle such problems.

4.4 A robust, weighted least squares type algorithm

It is well known that the cost function which is minimized in the least squares approach is

$$J(\theta) = (1/N) \sum_{t=1}^N [y(t) - \phi^T(t)\theta]^2 \quad (4.4.1)$$

which implies that the estimate of θ can be obtained by solving

$$(1/N) \sum_{t=1}^N \phi(t)[y(t) - \phi^T(t)\theta] = 0 \quad (4.4.2)$$

The estimate of θ from the above equation is

$$\hat{\theta}(N) = \left[\sum_{t=1}^N \phi(t)\phi^T(t) \right]^{-1} \left[\sum_{t=1}^N \phi(t)y(t) \right], \quad (4.4.3)$$

whenever the inverse exists. Now we re-formulate the problem in a different perspective as explained by eqns. (4.2.1) through (4.2.7). This provides us

$$J_1(\theta) = (1/N) \sum_{t=1}^N \rho[y(t) - \phi^T(t)\theta] \quad (4.4.4)$$

which is the new cost function and the estimate of θ can be obtained by solving

$$(1/N) \sum_{t=1}^N \phi(t)\psi[y(t) - \phi^T(t)\theta] = 0 \quad (4.4.5)$$

In general $\psi(\cdot)$ is nonlinear and the solution of (4.4.5) becomes cumbersome. Note that under normality assumption, $\psi(x) = x$ and (4.4.5) reduces to (4.4.2).

Note that (4.4.5) can also be written as

$$(1/N) \sum_{t=1}^N \frac{\psi[y(t) - \phi^T(t)\theta]}{[y(t) - \phi^T(t)\theta]} \phi(t) [y(t) - \phi^T(t)\theta] = 0 \quad (4.4.6)$$

since $[y(t) - \phi^T(t)\theta]$ is a scalar.

Now notice that (4.4.6) is similar to the solution of the weighted least squares approach (eg. Ljung and Soderstrom, 1983) where the weight $\alpha(t)$ is chosen as (which is the necessary condition for optimality),

$$\alpha(t) = \frac{\psi[y(t) - \phi^T(t)\theta]}{[y(t) - \phi^T(t)\theta]} \quad (4.4.7)$$

and $\alpha(t) := 1$ when $[y(t) - \phi^T(t)\theta] = 0$.

($:=$ means 'by definition')

so that the new estimate of θ can be computed as

$$\hat{\theta}(N) = \left[\sum_{t=1}^N \alpha(t) \phi(t) \phi^T(t) \right]^{-1} \left[\sum_{t=1}^N \alpha(t) \phi(t) y(t) \right] \quad (4.4.8)$$

whenever the inverse exists. Using the matrix inversion lemma (eg. Ljung and Soderstrom, 1983), the estimates of θ can be obtained recursively from the following

relationships.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-1)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{\{1/\alpha(t)\} + \phi^T(t)P(t-1)\phi(t)} \quad (4.4.9a)$$

and

$$P(t) = P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{\{1/\alpha(t)\} + \phi^T(t)P(t-1)\phi(t)} \quad (4.4.9b)$$

where the value of $\alpha(t)$ is described by eqn. (4.4.7), θ being taken as $\hat{\theta}(t-1)$. The starting values can be taken as,

$$\hat{\theta}(0) = \underline{0}; \quad P(0) = kI, \quad k \gg 0. \quad (4.4.9c)$$

Thus the nonlinear estimation problem (4.4.5) is reduced approximately to a weighted least squares type problem.

Note:

Kovacevic and Stancovic (1979) pointed out that, in general non-linear (robust) estimation algorithms can be sensitive to $P(0)$. In that paper, the sensitivity of $P(0)$ on a class of algorithms is illustrated using simulated examples. It is interesting to note that the algorithm described by (4.4.9) is not very sensitive to $P(0)$. However, too small or large values of k seem to affect the convergence rate of the algorithm (4.4.9). From a large number of simulated examples, the author came to the conclusion that values of k in between 10 and 100 give very good results.

4.5 Convergence analysis

In this section we look into the convergence properties of the proposed algorithm, given by eqns. (4.4.9a) and (4.4.9b). To begin with some lemmas are introduced, which will be useful in proving the main convergence result of the algorithm.

Lemma 4.1 (Ruthenpura et. al., 1985)

For a given influence function ψ , which is (i) odd, (ii) Riemann summable over any finite interval and such that (iii) $\exists M, m: 0 < M < \infty, m > 1$ so that $\forall x \geq M, \psi(x) \geq m/x$, the function

$$f(x) = K \exp\left[-\int_0^x \psi(t) dt\right], \text{ with } 0 < K < \infty \text{ such that}$$

$$\int f(x) dx = 1, \text{ is a probability density function (pdf).}$$

Furthermore $f(\cdot)$ is even.

[Note: Condition (i) is necessary and sufficient for the symmetry of $f(\cdot)$.

Condition (ii) is sufficient for the construction for an absolute continuous $f(\cdot)$.

Condition (iii) is sufficient for $f(\cdot)$ being a pdf.]

Proof:

$$\text{Define } \rho(x) = \int_0^x \psi(t) dt,$$

the existence of an absolute continuous $\rho(x)$ over any finite interval is guaranteed due to (ii).

Now

$$\rho(-x) = \int_0^{-x} \psi(t) dt,$$

$$= - \int_0^x \psi(-u) du \text{ where } u = -t$$

$$= - \int_0^x \psi(u) du = -\rho(x), \text{ since } \psi(\cdot) \text{ is odd.}$$

$$\Rightarrow \rho(\cdot) \text{ is even.}$$

Now define $f_1(x) = \exp(-\rho(x))$.

Therefore,

$$f_1(-x) = \exp[-\rho(-x)] = f_1(x) \text{ since } \rho(\cdot) \text{ is even.}$$

$$\Rightarrow f_1(\cdot) \text{ is even.}$$

Therefore $f(x) = Kf_1(x)$ is even; $K \in \mathbb{R}$.

Now from the assumption (iii) and $\forall x > M$

$$\rho(x) = \int_0^x \psi(t) dt = \int_0^M \psi(t) dt + \int_M^x \psi(t) dt.$$

Since (ii), for some $0 < \eta < \infty$,

$$p(x) \geq \eta + \int_M^x [m/t] dt$$

$$\text{i.e., } p(x) \geq \eta + \ln[x/M]^m > 0$$

Therefore, $f_1(x) \leq \exp(-\eta) \exp[-\ln(x/M)^m], \forall x \geq M.$

$$\text{i.e., } \int_0^M f_1(x) dx \leq \int_0^M f_1(x) dx + \exp(-\eta) \int_M^{\infty} [M/x]^m dx$$

$$\Rightarrow \int_0^M f_1(x) dx < \infty, \text{ if } m > 1.$$

Since $f_1(\cdot)$ is even then $\exists K, 0 < K < \infty$

such that

$$K \int_0^{\infty} f_1(x) dx = 1.$$

which means $f(x) = Kf_1(x)$ is a pdf.

Comment

The condition (iii) put forward by the above lemma suggests that the ψ function should not re-descend fast. Collins (1976) also showed that the ψ function should not re-descend 'steeply' from the point of view of the convexity of the resulting cost function, p . The above lemma puts this argument in a convenient quantitative fashion. Also see Hogg (1978) for some discussions on this aspect.

Now assume that x is a random variable,

$x: (\Omega, \mathcal{F}, P) \rightarrow (R, \mathcal{B}_R)$ [where (Ω, \mathcal{F}, P) and (R, \mathcal{B}_R) represent the probability space and the Borel field, respectively] such that the induced probability measure P_x on (R, \mathcal{B}_R) is defined by

$$\forall A \in \mathcal{B}_R, P_x(A) = \int_A f(x), \text{ where } f(x) \text{ is defined by lemma 4.1}$$

With this background we go to the following lemma.

Lemma 4.2 (Puthenpura et. al., 1985)

In addition to the assumption of lemma 4.1, assume that $x\psi(x) > 0, \forall x \in R$. Then $E[\psi(k+x)]$ is the same sign as k , where k is any real constant and E denotes the expectation, with respect to the measure $P_x(A)$.

Proof

$$E[\psi(k+x)] = \lim_{u \rightarrow -u} \int_{-u}^u \psi(k+x)f(x)dx$$

which exists; $\psi(\cdot)$ is summable (hence Borel measurable) and $f(\cdot)$ is a pdf. (Loe'Ve, 1977):

Now

$$\int_{-u}^u \psi(k+x)f(x)dx = \int_{-u}^0 \rho(k+x)f(x)dx + \int_0^u \rho(k+x)\psi(x)f(x)dx \quad (4.5.1)$$

But it is obvious that

$$\int_{-u}^u \rho(k+x)\psi(x)f(x)dx = \int_{-u}^0 \rho(k+x)\psi(x)f(x)dx + \int_0^u \rho(k+x)\psi(x)f(x)dx$$

and

$$\int_{-u}^0 \rho(k+x)\psi(x)f(x)dx = \int_0^u \rho(k-v)\psi(v)f(v)dv, \quad v = -x$$

$$= - \int_0^u \rho(k-v)\psi(v)f(v)dv.$$

Therefore (4.5.1) becomes

$$f(u)[\rho(k+u)-\rho(k-u)] + \int_0^u [\rho(k+x)-\rho(k-x)]\psi(x)f(x)dx.$$

Now observe that $[\rho(k+x)-\rho(k-x)]$ is of the same sign as k since $\rho(x)$ is positive, even and monotonically non-decreasing for $x \geq 0$, since $x\psi(x) \geq 0$. Also $f(\cdot)$ and $\psi(\cdot)$ are positive, $\forall x \geq 0$ and the lemma is proved.

Lemma 4.3 (Puthenpura et. al., 1985)

Let $r(t) = \phi^T(t)[P(t-1)]^2 \phi(t)$ where $P(t)$ and $\phi(t)$ are described by the algorithm (4.4(9)).

Then $\sum_{t=1}^{\infty} r(t) < \infty$ if $\sup_{t \in [0, \infty)} \{ \varepsilon(t) / \psi[\varepsilon(t)] \} < \infty$

where $\varepsilon(t) = y(t) - \phi^T(t)\hat{\theta}(t-1)$, the prediction error.

Proof

$$r(t) = \phi^T(t)[P(t-1)]^2\phi(t)$$

$< \phi^T(t)P(t-1)P(t-2)\phi(t)$, using matrix inversion lemma (eg. Ljung and Soderstrom, 1983)

$$= \text{trace}[P(t-1)\phi(t)\phi^T(t)P(t-2)]$$

$$= [\text{trace}P(t-2) - \text{trace}P(t-1)]/[1/\alpha(t)]$$

Since $\alpha(t) > 0 \Rightarrow \text{trace} P(t) < \text{trace} P(t-1)$, $\forall t \in [0, \infty)$.

Therefore,

$$\sum_{t=1}^{\infty} r(t) \leq \sup_{t \in [0, \infty)} \{1/\alpha(t)\} [\text{trace} P(0) - \text{trace} P(\infty)]$$

$$< \infty \text{ if } \sup_{t \in [0, \infty)} \{1/\alpha(t)\} < \infty \quad (4.5.2)$$

Note:

For asymptotic efficiency, in the neighbourhood of origin $[(-a, -a)]$, $\psi(x)$ is often taken to be linear. See Huber (1981) and Poljak and Tsybakin (1980). Let $\psi(x) = x$ be the choice when $|\varepsilon(t)| \leq \Delta$, Δ being some positive real number, often taken as σ , 1.5σ or 3σ , where σ is the standard deviation of the measurement noise. Assume that $|\varepsilon(t)| \leq \lambda < \infty$, $\forall t \in [0, \infty)$. The the following condition is sufficient to ensure (4.5.2).

$$\exists \beta; |\psi(\varepsilon(t))| \geq \beta > 0 \quad \forall \varepsilon(t) : 0 < |\varepsilon(t)| \leq \lambda \quad (4.5.3)$$

Now we are in a position to state the main convergence result, in the form of the following theorem.

Theorem 4.1 (Puthenpurath et al., 1985)

Let the assumptions of lemmas 4.1 through 4.3 hold. Then for the algorithm (4.4.9)

$$\lim_{t \rightarrow \infty} \theta(t) = \theta_0 \text{ in quadratic mean (qm)}$$

where θ_0 is the true parameter, provided

(i) $\phi(t)$ and $e(t)$ are statistically independent $\forall t$
 $e[0, \infty)$ and

(ii)

$$\limsup_{N \rightarrow \infty} (1/N) \sum_{t=1}^N |\phi(t)|^2 < \infty$$

The proof is along the same way as that of Robbins and Monro (1951) and it is given in appendix II.

The discussions presented above give a good insight into the various theoretical aspects of the convergence of the proposed algorithm. In the next section, some practical considerations on the algorithm are presented.

4.6 Practical considerations

To use the algorithm described by (4.4.9), one should look into two main points. First of all, the user should make sure that the employment of the algorithm can improve the quality of the estimates in comparison with the ordinary least squares estimates, before he actually applies it.

Secondly, he should be able to decide upon the influence function to be used. These two issues are stressed in the following subsections.

4.6.1 How to decide on M-estimation?

From a number of simulation studies and experience with real data, it is concluded that the 'residual kurtosis' (the peakedness of the pdf of residuals) is a good source of knowledge on the quality of the data used for identification. Thus, the following theorem is quite appropriate in this context. [In section 4.8.2, the usefulness of the theorem is illustrated with the help of a real-life identification problem.]

Theorem 4.2 (Bierens)

Let $\{s(t)\}$ represents the residual sequence of a least squares regression. Then the regression coefficients may be improved (in the sense of less standard error) if the residual kurtosis

$$(1/n) \sum_{t=1}^n [s(t)]^4$$

$$- 3 > 0$$

$$\left\{ (1/n) \sum_{t=1}^n [s(t)]^2 \right\}^2$$

(Note that this condition is sufficient, not necessary.)

Proof

Bierens (1981), provides an elegant proof for this result:

4.6.2 Comments on the choice of the influence function

The proposed algorithm permits the use of any influence function provided it satisfies the conditions put forward by lemmas 4.1 through 4.3. One simple ψ function proposed by Huber (1981) seems to be quite adequate for most of the practical situations, from a number of studies made by the author on simulated as well as real data. The ψ function of Huber is given in fig. 4.1. The choice of a is often taken to be σ , 1.5σ or 3σ , σ being some estimate (or even guess) of the standard deviation of the measurement noise. Also in accordance with the central limit theorem, the distribution of noise, which is the sum of a large number of uncorrelated terms, can be assumed to be normal; hence give further justification for choosing ψ function to be linear around the neighbourhood of origin. It will be shown in section 4.7.2 that in the case of normality on data, the efficiency of the algorithm will become poorer and poorer as a approaches the origin. This is the price we pay for robustness. But if we have sufficient number of samples, this is not a major drawback. Also in practice, experimental data seldom enjoys normality. These aspects are very clear from the examples discussed in section 4.8.

4.7 Efficiency of the robust algorithm

An important factor to be considered about recursive identification is the efficiency, since the rate of convergence of such algorithms is dependent on their efficiency. One successful method of efficiency evaluation of recursive algorithms is via Cramér-Rao inequality (Poljak and Tsytkin, 1980). In this section, the same method is adopted to determine the efficiency of the robust algorithm described in section 4.4.

In general the estimation problem is of the form

$$y(t) = p(\phi(t), \theta) + e(t) \quad (4.7.1)$$

where $p(\cdot)$ is some arbitrary function. [Compare eqn. (4.7.1) with the linear estimation problem expressed by eqn. (4.3.3)]. Let $\phi(t) \in R^r$, $\theta \in R^n$, $p: R^r \times R^n \rightarrow R$. Also assume that $e(t)$'s are independent and identically distributed random variables with an absolute continuous pdf, $f(x)$. Now we can state the following theorem.

Theorem 4.3 (Generalized Cramér-Rao inequality)

If $p(\phi(t), \theta)$ is differentiable with respect to θ for all t and $f(\cdot)$ [the pdf of $e(t)$] is absolutely continuous and the matrix

$$B = E\left\{ \begin{matrix} [\text{Grad } p(\cdot)] \\ \theta \end{matrix} [\text{Grad } p(\cdot)]^T \right\} \quad (4.7.2)$$

is positive definite

then $\forall t \in [0, \infty)$

$$E\{[\hat{\theta}(t) - \theta_0][\hat{\theta}(t) - \theta_0]^T\} \geq [I(f)]^{-1} B^{-1} \quad (4.7.3)$$

where $\hat{\theta}(t)$ is the estimate of the parameter vector θ at time t , $\theta_0 = E[\hat{\theta}(t)]$ which is the the actual value of θ and the quantity

$$I(f) = \int [f'(x)/f(x)]^2 f(x) dx \quad (4.7.4)$$

where $'$ stands for the derivative with respect to x , is called the 'scalar Fisher information'.

Proof

Exactly in the same way as given in Zacks (1971).

As can be easily seen in the above theorem, the crucial point in the efficiency of recursive algorithms is the scalar Fisher information (SFI). Further study on SFI (in connection with the influence function used for robust algorithms) is made in the next section.

4.7.1 The scalar Fisher information (SFI)

The classical definition of SFI is given by eqn. (4.7.4). For any absolute continuous pdf $f(x)$, it can be shown that $I(f) < \infty$. [For proof, see Huber (1981).] But the form of the classical definition of SFI is quite inadequate for comparing the efficiencies of a set of robust algorithms over a class of influence functions. Therefore, a transformation of SFI is made to suit this requirement. This is given below as a theorem.

Theorem 4.4 (Transformation of SFI)(Puthenpura et. al., 1985)

Let the assumption of lemmas 4.1 and 4.2 hold good. Also assume that $\psi(\cdot)$ is continuous. Then,

$$I(f) = \left[\int \exp[-\rho(t)] dt \right]^{-1} \int \{ [\exp(-t)] / [\rho^{-1}(t)]' \} dt$$

Proof:

From (4.7.4) and the of $f(\cdot)$, we get

$$I(f) = \int \psi^2(x) f(x) dx = 2K \int [\rho'(x)]^2 \exp[-\rho(x)] dx$$

with

$$K^{-1} = \int \exp[-\rho(x)] dx.$$

(4.7.5)

Now for any closed set $[0, u]$, from (4.7.5)

$$I_u(f) = 2K \int_0^{\rho(u)} \rho'[\rho^{-1}(y)] [\exp(-y)] dy, \text{ where } \rho(t) = y.$$

(4.7.6)

Assume that $\psi(\cdot)$ is continuous and $x\psi(x) > 0 \forall x \neq 0$, then $\rho(x)$ is strictly increasing and $\rho^{-1}(\cdot)$ exists $\forall x > 0$. Noting that for any non zero function $g'[g^{-1}]$, $g(\cdot)$ being strictly increasing,

$$[g^{-1}(y)]' = 1/[g'[g^{-1}(y)]] \quad (4.7.7)$$

we obtain

$$I(f) = \lim_{\alpha \rightarrow \infty} I_{\alpha}(f) =$$

$$\left[\int \exp[-\rho(t)] dt \right]^{-1} \int \left\{ \frac{\exp(-t)}{[\rho^{-1}(t)]'} \right\} dt$$

With the help of the above result, now we can proceed further to evaluate the efficiency of the robust algorithm introduced in section 4.4.

4.7.2 Evaluation of efficiency of the robust algorithm

From equation (4.3.3) and theorem 4.3, we obtain,

$$E\{[\hat{\theta}(t) - \theta_0][\hat{\theta}(t) - \theta_0]^T\} = [I(f)]^{-1} B^{-1} \quad (4.7.8a)$$

where

$$B = E\{\dot{\phi}(t)\dot{\phi}^T(t)\} \quad (4.7.8b)$$

and $I(f)$ is as given by (4.7.4).

Now the problem is to compare the efficiency of a set of robust algorithms, examining the influence function they employ, when $\{e(t)\}$ actually enjoys normality. Since the matrix B is one and the same for all the algorithms, it is sufficient to compare the SFI for different for different influence functions. However, direct evaluation of SFI may be very difficult, in general. Therefore the following result is very valuable in those situations.

Theorem 4.5 (Puthenpura et. al., 1985)

Let $\psi_1(x)$ and $\psi_2(x)$ be two different influence

functions satisfying the conditions of theorem 4.3. Then if

(a) $\psi_1(x) \leq \psi_2(x)$ and (b) $\rho_2^{-1}(x) - \rho_1^{-1}(x)$ is monotonically increasing, where

$$g(x) = \int_0^x \psi(x)dx, \text{ it follows that } I(f_2) < I(f_1).$$

Proof:

Given

$$\psi_2(x) \leq \psi_1(x), \forall x \in [0, \infty)$$

$$\Rightarrow \rho_2(x) \leq \rho_1(x), \forall x \in [0, \infty)$$

$$\Rightarrow \left[\int_0^{\infty} \exp(-\rho_2(x))dx \right]^{-1} < \left[\int_0^{\infty} \exp(-\rho_1(x))dx \right]^{-1}$$

(4.7.9)

Now if $\rho_2^{-1}(x) - \rho_1^{-1}(x)$ is monotonically increasing then

$$[\rho_2^{-1}(x)]' \geq [\rho_1^{-1}(x)]' \tag{4.7.10}$$

This finishes the proof of the theorem by applying inequalities (4.7.9) and (4.7.10) to the result of theorem 4.4.

The geometric interpretation of conditions (a) and (b) of the above theorem is illustrated in fig.4.2a.

Example:(Efficiency when the measurement noise is normal)

Assume that the measurement noise $\{e(t)\}$ is normally distributed. Let,

$$\psi_1(x) = x/(\sigma^2) \quad (4.7.11)$$

which is the influence function associated with a Gaussian pdf

$$f_1(x) = [1/\sqrt{2\pi}\sigma] \exp[-x^2/2\sigma^2]. \quad (4.7.12)$$

It can be easily worked out that

$$[I(f_1)]^{-1} = \sigma^2 \quad (4.7.13)$$

Now consider a Huber type ψ function with a slope of $1/\sigma^2$ over $[-a, a]$.

ie.,

$$\psi_2(x) = \begin{cases} x/\sigma^2, & |x| \leq a \\ a/\sigma^2 \text{ sign}(x), & |x| > a. \end{cases} \quad (4.7.14)$$

After a few cumbersome calculations (which are omitted here for brevity) we get

$$[I(f_2)]^{-1} = \sigma^2 + [\sigma^3/a][\sqrt{(2/\pi)}][\exp[-a^2/2\sigma^2][\text{erf}(a/(\sqrt{2}\sigma))]^{-1} \quad (4.7.15)$$

Note that

$$\text{erf}(x) = [2/\sqrt{\pi}] \int_0^x \exp[-t^2] dt > 0, \quad \text{called the error function.}$$

From (4.7.13) and (4.7.15) it follows that

$$[I(f_2)]^{-1} > [I(f_1)]^{-1}. \quad (4.7.16)$$

Moreover,

$$\lim_{x \rightarrow \infty} \text{erf}(x) = 1$$

implies that $\lim_{a \rightarrow 0} [I(f_2)]^{-1} = [I(f_1)]^{-1}$.

[The value of $[I(f_2)]/[I(f_1)]$ is plotted for different a/σ ratios in fig. 4.2b., as a matter of interest.]

In the case of normality of measurement noise, it can be shown that the minimum value of the error covariance matrix of the parameter estimates is obtained when $c(t) = 1/\sigma^2$, σ^2 being the variance of the noise (see eg. Deutch, 1965; Young, 1984). This is obtained by selecting ψ as given by (4.7.11). If it is desired to implement the algorithm in section 4.4 using the ψ given by (4.7.14), the premium paid will be the loss of efficiency (only in the normal case), due to (4.7.16). Also see that this loss of efficiency will be more and more prominent as 'a' approaches the origin, which is evident from (4.7.15).

A similar result can be obtained by a straightforward application of theorem 4.5 involving two ψ_2 functions when the measurement noise is normal. Notice that the application of theorem 4.5 does not necessitate the direct calculation of SFI, which is often difficult.

Two important aspects to be remembered in this context are that (i) the above discussions are useful only for the cases where the true pdf of the measurement noise is normal but we use influence functions other than (4.7.11) and (ii) the deduction that using $\psi_1(\cdot)$ is more efficient than using $\psi_2(\cdot)$ is valid only when the strict inequality $[I(f_2)]^{-1} > [I(f_1)]^{-1}$ holds. This is because of the fact

that (4.7.8a) is not a strict equality. Putting in other terms, the problems (i) and (ii) arise due to the fact that we have used only a bound (right hand side of the Cramér-Rao inequality) for comparing the efficiency. This can be illustrated as follows. Assume that the underlying distribution is Laplacian [i.e., $f_3(x) = (1/2)\exp(-|x|/\sigma)$, $\psi_3(x) = (1/\sigma)\text{sign}(x)$]. Instead of using this influence function if we use $\psi_1(x) = x/\sigma^2$, it does not mean that we do not loose efficiency in spite of the fact that $[I(f_3)]^{-1} = [I(f_1)]^{-1} = \sigma^2$. Similarly, assume that the true pdf is Huber's pdf but we use (4.7.11) as our influence function. In this case also, it is not meant that we gain efficiency, eventhough the strict innequality (4.7.16) holds.

In the absence of any information about the measurement noise, one usually selects $a(t) = 1 \forall t$, while using recursive least squares algorithm. This corresponds to a ψ function with unit slope in the interval $(-,+)$. So in practical situations, where not much information is available on noise [see examples given in the next section], one can use the ψ functions given in fig 4.1 with a unity slope in the neighbourhood of origin.

4.8 Examples of application of the robust algorithm

In this section simulation studies and results of robust identification of a packed-bed chemical reactor, using the proposed approach, are presented.

4.8.1 Simulation studies

Consider the simple discrete-time system given by

$$\frac{(1.0)z^{-1} + (0.5)z^{-2}}{1 - (1.5)z^{-1} + (0.7)z^{-2}} \quad (4.8.1)$$

The system is perturbed with a sequence of random inputs and the corresponding outputs are collected. To this noise-free output data, random numbers are added with proper variance to simulate different signal-to-noise ratios. These random numbers are drawn from a Gaussian distribution. To simulate outliers, another set of random numbers (drawn from a Laplacian distribution) are substituted to some output data points instead of the normal measurement noise. These outliers are simulated with different probabilities of occurrence. For example, the probability of occurrence is 0.1 means that one out of 10 readings is an outlier.

To begin with, ordinary least squares method is used to estimate parameters. The parameter error norm

$$E(t) = \frac{\|\hat{\theta}(t) - \theta_0\|}{\|\theta_0\|} \quad (4.8.2)$$

is computed and this is shown in fig. 4.3. [For the sake of brevity, the results are given only for the case when 80% of the noise is from a normal distribution, $N(0, 0.3)$ and 20% is from a Laplacian, $L(0, 20)$]. As shown in fig. 4.3, the

result is unsatisfactory.

Now the algorithm 4.4.9 is applied using a simple ψ function as shown in fig 4.1 with $a=0.6$. [Note that this value is near $1.5\sigma = 0.45$]. The result is given in fig.4.4 where there is a substantial improvement in the error norm. The convergence of different parameters of the model (4.8.1) is also shown in fig.4.5.

4.8.2 Application of the robust algorithm to a real-life problem

(Identification of a packed-bed chemical reactor)

The study of packed-bed catalytical reactors is one of the most interesting problems in the modern chemical industry. Basically, the reaction taking place in this type of reactors is the hydrogenolysis of n-butane over a nickel on silica gel catalyst. The percentage conversion of n-butane can be controlled by the reactor wall temperature. [For details on the reactor refer to Jutan et al., (1977)]. In this context the problem is to model the reactor in such a manner that the percentage conversion of n-butane is obtained as a function of the reactor wall temperature (temperature of the hot point).

The reactor wall temperature is varied randomly as shown in fig.4.7 and it is measured using thermocouples.

The resulting percentage conversion of n-butane is obtained by using a gas chromatograph and it is shown in fig.4.7.

The sampling period is taken to be 3 minutes.

By correlation analysis (see Box and Jenkins, 1970), an auto regressive moving average (ARMA) model of order (2,2) with no delay seems to be adequate to represent the model. So the model is of the form,

$$\frac{\theta_1 + \theta_2 z^{-1} + \theta_3 z^{-2}}{1 + \theta_4 z^{-1} + \theta_5 z^{-2}} \quad (4.8.3)$$

Now ordinary recursive least squares method is used on five hours of data to identify the parameters of (4.8.3). The results are shown in table 4.1.

This model passes all model validation tests (eg. the chi-square test), which are commonly used. But it can be seen that the residual kurtosis is very high. This shows that the residual pdf is highly leptokurtic and the results may be improved by robust methods, by virtue of theorem 4.2. Moreover, the standard error of the parameters are very unsatisfactory.

Now we apply the proposed method using the ψ function of fig.4.1, choosing $a=0.05$, which is nearly 1.5σ . The results are given in table 4.2.

Analysis of results

Comparing tables 4.1 and 4.2, the parameters are not dramatically different (except θ_1). But it can be claimed easily that the parameters obtained in table 4.2 are robust because of the following reasons.

(i) From table 4.1 and table 4.2 it is clear that the standard error (see Box and Jenkins, 1970) on each parameters is drastically reduced. This improvement is substantial.

(ii) Figures 4.8 and 4.9 are the residuals of the least squares and the robust fit, respectively. It is clear from fig.4.9 that the robust method could pick out all the outliers present in the data, so that they do not affect the quality of estimates. For example, data point 34 does not appear to be an outlier (from fig.4.7) but it is an outlier as suggested by fig.4.9. Similarly data point 89 looks like an outlier (see fig.4.7) but from fig.4.9 we understand that actually it is not an outlier. These features are not revealed by the residuals of the least squares method (fig.4.8).

Thus, this example vividly illustrates the power of the proposed method in practical situations.

4.9 Extension of M-estimation to the instrumental variables method

4.9.1 Introduction

One immediate extension of the above mentioned theory of the weighted least squares type M-estimation algorithm is to the so called 'recursive instrumental variables method'.

Instrumental variables (IV) method is quite popular

in obtaining unbiased estimates. Identification of linear discrete-time systems using instrumental variables has been discussed extensively in literature (eg. Soderstrom and Stoica, 1981, 1983; Young, 1970, 1984). It is to be noted that, in literature, lot of attention has been given to the choice of instrumental variables and different schemes of using them. This is logical due to the heavy dependency of the accuracy of IV estimates on the choice of instrumental variables and the set up in which one uses this method. [See for example: Puthenpura and Sinha, 1984a, 1985c]. However, it is important to note that some observations with large errors have very detrimental effects on the IV scheme, just like they affect the least squares estimation. This necessitates a robust version of the IV algorithm to suite practical situations (Puthenpura and Sinha, 1985b).

It is important to realize that IV methods are only an improvement of the least squares approach. The basic motivation behind both the methods is one and the same. This is explained in the next section.

4.9.2 IV method as approximate gradient method

In the least squares estimation case we are interested in finding the value of θ such that the prediction error criterion

$$J = E\left\{\frac{1}{2}[\epsilon(t, \theta)]^2\right\} \quad (4.9.1a)$$

is minimized, where

$$\epsilon(t, \theta) = y(t) - \phi^T \theta, \quad (4.9.1b)$$

which is the prediction error and E denotes the expectation. The desired value of θ can be obtained by solving

$$[\partial J / \partial \theta] = E\{[\partial \varepsilon(t, \theta) / \partial \theta][\varepsilon(t, \theta)]\} = 0 \quad (4.9.2a)$$

Thus the problem reduces to the evaluation of the gradient,

$$\text{grad}'[\varepsilon(t, \theta)] = [\partial \varepsilon(t, \theta) / \partial \theta] \quad (4.9.2b)$$

If we ignore the implicit dependence of θ on $\phi(t)$ we can get the approximate gradient as

$$\partial \varepsilon(t, \theta) / \partial \theta = \phi(t) \quad (4.9.2c)$$

Substituting (4.9.2c) in (4.9.2a) we get

$$E\{\phi(t)[y(t) - \phi^T \theta]\} = 0 \quad (4.9.2d)$$

which is the ordinary least squares solution.

Suppose we consider the actual output as

$$y(t) = \phi^T \theta_0 + e(t) \quad (4.9.3)$$

for some θ_0 and stochastic disturbance $\{e(t)\}$. We may refer θ_0 as the true value of θ . Now we see that θ_0 minimizes J , i.e., θ_0 is a solution of (4.9.2d) only if

$$E\{\phi(t)e(t)\} = 0 \quad (4.9.4)$$

which means that $e(t)$ must be zero mean and uncorrelated to $\phi(t)$. In other situations (which is often the case) the approximation (4.9.2c) will be inadequate and the solution of (4.9.2d) will give biased estimates. Therefore, under these circumstances, it is necessary to improve the gradient given by (4.9.2c). This leads to the concept of instrumental variables.

Let us consider the vector $\zeta(t)$ such that

$$E\{\zeta(t)e(t)\} = 0 \quad (4.9.5a)$$

and

$$E\{\zeta(t)\zeta^T(t)\} \quad (4.9.5b)$$

positive definite or at least non-singular.

Then we can solve

$$E\{\zeta(t)[y(t) - \zeta^T(t)\theta]\} = 0 \quad (4.9.5c)$$

instead of (4.9.2c); by taking the gradient

$$\partial \zeta(t, \theta) / \partial \theta = \zeta(t) \quad (4.9.5d)$$

which is a better approximation than (4.9.2c) from that it gives unbiased estimates. The vector $\zeta(t)$ is called the instrumental variable vector. The solution of (4.9.5c) is called the IV estimate, which will converge to θ_0 as t approaches infinity (Ljung and Soderstrom, 1983).

4.9.3 Robust version of the IV method

In previous sections, it was substantiated that extreme observations have very bad effects on the least squares method. Since IV method is basically a least squares method (as pointed out in the previous section) one can expect that this method is also susceptible to bad data. However, since the IV method is based on the fact that it provides a better approximation for the gradient via condition (4.9.5a) one can also conjecture that the effect of outliers on IV approach is not as much as that on least squares. Unfortunately, this improvement is not substantial (in fact negligible in most of the cases) and we have to depend on extra techniques to robustify the conventional IV approach.

As shown in previous sections one effective way to achieve the above goal is to modify the criterion function, instead of adopting a quadratic criterion (Ljung and Soderstrom, 1983; Puthenpura et. al., 1985; Puthenpura and Sinha, 1985c).

So, for the general case, consider the criterion function

$$J_1 = -E\{\rho[\varepsilon(t, \theta)]\} \quad (4.9.6a)$$

instead of (4.9.1a). Proceeding exactly as before

$$[\partial J_1 / \partial \theta] = E\{[\partial \varepsilon(t, \theta) / \partial \theta] \psi(\varepsilon(t, \theta))\} = 0 \quad (4.9.6b)$$

$\psi(\cdot)$ being the derivative of $\rho(\cdot)$ with respect to the argument.

Now proceeding exactly as described in section 4.4, we arrive at the following recursive relationships.

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-1)\zeta(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{\{1/\alpha(t)\} + \phi^T(t)P(t-1)\zeta(t)} \quad (4.9.7a)$$

and

$$P(t) = P(t-1) - \frac{P(t-1)\phi(t)\zeta^T(t)P(t-1)}{\{1/\alpha(t)\} + \phi^T(t)P(t-1)\zeta(t)} \quad (4.9.7b)$$

with

$$\alpha(t) = \frac{\psi[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{[y(t) - \phi^T(t)\hat{\theta}(t-1)]} \quad (4.9.7b)$$

and $\alpha(t) = 1$ when $[y(t) - \phi^T(t)\hat{\theta}(t-1)] = 0$.

The starting values can be taken as,

$$\hat{\theta}(0) = \underline{0}; \quad P(0) = kI, \quad k \gg 0. \quad (4.9.7c)$$

4.9.4 Convergence analysis

The result on convergence of algorithm (4.9.7) is stated as the following theorem (Puthenpura and Sinha, 1985c).

Theorem 4.6 (Puthenpura and Sinha, 1985b)

Let the $\psi(\cdot)$ function satisfy the condition of theorem 4.1. Also if

- (i) $e(t)$ has zero mean,
- (ii) $\zeta(t)$ and $e(t)$ are statistically independent,
- (iii) $\langle \psi(t), \zeta(t) \rangle > 0 \quad \forall t \in [0, \infty)$,

[where $\langle \dots \rangle$ represents the inner product]

- (iv)

$$\lim_{N \rightarrow \infty} \sup_{t=1}^N \langle \psi(t), \zeta(t) \rangle < \infty \quad \text{and}$$

- (v) $\sup_{t \in [0, \infty)} \{1/\alpha(t)\} < \infty$,

then the algorithm (4.9.7)

$$\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta_0 \quad \text{in probability.}$$

The proof of this theorem is given in appendix III.

It is interesting to note that condition (ii) implies (4.9.5a) and the condition (4.9.5b) is satisfied if

condition (iii) and (iv), are met. Assuming that

$$|\varepsilon(t, \theta)| < \lambda < \infty \quad \forall t \in [0, \infty).$$

Then condition (4.5.3) is sufficient to ensure (v):

Another important thing to notice here is that the convergence property of the robust IV method is slightly weaker than that of the weighted least squares type algorithm introduced in section 4.4. [Note that convergence in quadratic mean always implies convergence in probability, but the converse is not true (Lukacs, 1975)]. However, for practical purposes this type of convergence is strong enough.

4.9.5 Simulation results

Let us consider the same system given by (4.8.1). The data containing outliers is generated exactly in the same way as described in section (4.8.1). Besides, the same error criterion (4.8.2) is used to compare the performance of the conventional IV algorithm and its robust version. Note that the conventional IV algorithm is obtained by putting $c(t)=1$, $\forall t \in [0, \infty)$, in (4.9.7).

For the sake of brevity, only a typical case is given, where the signal-to-noise ratio is 10dB with 0.1 probability of occurring outliers with variance equal to 20. Figures 4.10 and 4.11 show the results which are self explanatory. In both the cases, instrumental variables are produced by passing the same input through an auxiliary model and updating the model with the latest estimates of system

parameters (Sinha and Kuszta, 1983). The influence function used for the robust algorithm is that shown in fig.4.1 with $a=0.5$. The comparison of fig.4.10 and fig.4.11 shows the substantial improvement of the error norm when the robust algorithm is used for identification.

Since we have taken 'a' close to the origin, from fig.4.11 it follows that the rate of convergence is slow. This is the price we pay for robustness. However, if one doesn't know the actual variance of the measurement noise (in which case one can choose 'a' to be 1.5σ , 2.0σ or 3σ), it is safer to choose 'a' to be closer to the origin, paying a premium of convergence rate.

4.9.6 Identification of the packed-bed reactor

Let us now consider the same identification problem discussed in section 4.8.2.

Table 4.3 shows the results of identification using the conventional IV method.

From the residual kurtosis, by virtue of theorem 4.2, robust procedures can improve this result. So, we apply algorithm (4.9.7), using the ψ function of fig.4.1 with $a=0.05 = 1.5\sigma$. Table 4.4 shows the results.

Note:

In both the cases, the instrumental variables are formed exactly in the same way as explained in the previous section, i.e., updating the auxiliary model with the latest estimates of system parameters.

Analysis of results

Comparing table 4.1 and 4.3, one can see that there is no significant improvement in the standard error of parameters by employing the instrumental variables method instead of the least squares method. This is clear from the residual plots also (fig.4.12), which is not much different from fig.4.8.

From table 4.4 and the residual plot of the robust IV algorithm (fig.4.13), the robustness of identification is quite clear. Also note that there is no substantial improvement in the standard error of parameters given by the robust IV algorithm over the weighted least squares type algorithm of section 4.4 in this particular case. This can be understood by comparing tables 4.2 and 4.4. This is because of the fact that the measurements are not very noisy. Instrumental variables method gives better estimates compared to the least squares approach in cases where the noise level is quite substantial.

4.10 Extension of M-estimation to correlation method

4.10.1 Introduction

In correlation methods (eg. Isermann et.al., 1974; Sinha and Kuszta, 1983) no assumption is made on the noise. The only requirements are, (a) stationarity of the input and noise sequences and (b) no correlation between the input and the noise sequences. Therefore one might expect that

outliers may not affect correlation methods. But Basu and Vandelinde (1977) show that this is not true. They pointed out that recursive estimation schemes such as the one suggested by Isermann et.al. (1974), though always consistent, are affected by the distribution of the noise in terms of efficiency, speed of convergence and variance of estimates. That is, unmodelled noise considerably affects correlation methods. In this section we apply the M-estimation method to reduce the sensitivity of the correlation method to the actual distribution of noise.

4.10.2 A robust correlation method

Let the $\{u(t)\}$ and $\{y(t)\}$ be the input and output sequences of the system. The auto and cross correlation functions are

$$Y_{uu}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{t=0}^N u(t)u(t-\tau) \quad (4.10.1a)$$

and

$$Y_{yu}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N+1} \sum_{t=0}^N y(t)u(t-\tau) \quad (4.10.1b)$$

respectively. Now (4.10.1b) is often recursively calculated as

$$\hat{Y}_{yu}(\tau, t) = \hat{Y}_{yu}(\tau, t-1) + (1/t+1)[y(t)u(t-\tau) - \hat{Y}_{yu}(\tau, t-1)] \quad (4.10.2)$$

When $\{u(t)\}$ is white or a pseudo random binary sequence (PRBS),

$$\begin{aligned} \gamma_{uu}(\tau) &= \gamma_{uu}(0)\delta(\tau); & \delta(\tau) &= 1 \text{ when } \tau=0 \\ & & \delta(\tau) &= 0 \text{ when } \tau \neq 0. \end{aligned} \quad (4.10.3)$$

Then the estimated impulse weights $\{\hat{h}(\tau)\}$ are

$$\hat{h}(\tau) = [\hat{\gamma}_{yu}(\tau) / \hat{\gamma}_{uu}(0)] \quad (4.10.4a)$$

and $\hat{\gamma}_{yu}(\tau)$ are recursively calculated by (4.10.2)

The corresponding outputs are

$$\hat{y}(t) = \sum_{i=0} \hat{h}(t-i)u(i) \quad (4.10.4b)$$

Now the parameter vector θ can be estimated from the following equation.

$$\hat{Y}(l) = C(l)\theta \quad (4.10.5)$$

where

$$\hat{Y}(l) = [\hat{y}(1), \hat{y}(2), \hat{y}(3), \dots, \hat{y}(l)]^T$$

$$C(l) = \begin{bmatrix} 0 & 0 & \dots & 0 & u(0) & 0 & 0 \\ -\hat{y}(1) & 0 & & 0 & u(1) & u(0) & 0 \\ -\hat{y}(2) & -\hat{y}(1) & & 0 & u(2) & u(1) & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ -\hat{y}(l-1) & -\hat{y}(l-2) & \dots & -\hat{y}(l-n) & u(l-1) & u(l-2) & u(l-m) \end{bmatrix}$$

and

$$\theta^T = [a_1, a_2, \dots, a_n, b_1, b_2, \dots, b_m]^T$$

The application of M-estimation method changes equation

(4.10.2) to

$$Y_{yu}(\tau, t) = Y_{yu}(\tau, t-1) + (1/t+1)\psi[y(t)u(t-1) - Y_{yu}(\tau, t-1)] \quad (4.10.6)$$

The rest of the steps are not changed.

4.10.3 Discussion on the convergence

Proceeding exactly in the similar way as shown in appendix II, for the algorithm (4.10.6), the following result can be obtained.

Corollary 4.1 [of theorem 4.1]

Let the ψ function of (4.10.6) satisfy the conditions proposed by lemmas 4.1 and 4.2. Also assume that

(i) the input sequence $\{u(t)\}$ and the noise sequence $\{e(t)\}$ are statistically independent and

(ii) $\sup_{t \in [0, \infty)} [y(t)u(t-1) - Y_{yu}(\tau, t-1)]$. Then,

$$\lim_{t \rightarrow \infty} Y_{yu}(\tau, t) = Y_{yu}(\tau, t) \text{ in quadratic mean.} \quad (4.10.7)$$

Note that here, the scalar multiplier sequence of the recursive algorithm (4.10.6) is $\{1/(t+1)\}$, which is independent of $\psi(\cdot)$ and such that

$$\sum_{t=1}^{\infty} [1/(t+1)] = \infty$$

and

$$\sum_{t=1}^{\infty} [1/(t+1)]^2 < \infty. \quad (4.10.8)$$

Hence, lemma 4.3 is not necessary in this case. In fact Basu and Vandelinde (1977) have shown the same result (4.10.7) for the particular choice of $\psi(\cdot)$ function shown in fig.4.1, more or less in the same method adopted here to prove the theorem 4.1.

4.10.4. Simulation results

Consider the same system given by equation (4.8.1). The input and output sequences are generated exactly as explained in section 4.8.1. The parameter error norm is compared for the two situations, the linear case (equation 4.10.2) and the robust case (equation 4.10.6). The results are shown in fig.4.14, which are self explanatory.

4.11. Extension to impulse and step response methods

To make the discussions complete, another two special cases of identification are also covered in the following subsections.

4.11.1 Introduction

The impulse response method (Sinha and Kuszta, 1983) is as follows. Let $\{h(t)\}$ be the impulse response of the model given by equation (4.3.1). It then follows that

$$b_0 + b_1 z^{-1} + \dots + b_m z^{-m} = h_0 + (h_1 + a_1 h_0) z^{-1} + \dots \quad (4.11.1)$$

Equating coefficients of z^{-j} , one arrives at

$$h_0 = b_0$$

$$[h_j + \sum_{i=1}^j a_i h_{j-i}] = b_j \quad j=1, 2, \dots, m \quad (4.11.2a)$$

$$[h_j + \sum_{i=1}^n a_i h_{j-i}] = 0 \quad j=m+1, m+2, \dots \quad (4.11.2b)$$

The impulse response method is to estimate the values of a_i ($i=1, 2, \dots, n$) from (4.11.2b) and to substitute these values in (4.11.2a) to obtain b_i ($i=1, 2, \dots, m$).

Instead of impulse responses, if we have step responses $\{w(t)\}$, we can obtain the equivalent impulse responses as follows, before applying the above method.

$$h(0) = w(0)$$

$$h(t) = w(t) - w(t-1), \quad t=1, 2, \dots \quad (4.11.3)$$

4.11.2 Robustness aspects

Because of measurement noise, the right hand side of equation (4.11.2b) seldom be zero and it is often practised to minimize the cost function

$$J = [-h_j - \sum_{i=1}^n a_i h_{j-i}]^2 \quad j=m+1, m+2, \dots \quad (4.11.2b)$$

This is a regression problem and recursive least squares method can be used to solve this (Sinha and Kuszta, 1983).

Obviously, the resulting estimates are sensitive to outliers and the algorithm described in section 4.4 can be used here to get robust estimates (Puthenpura and Sinha, 1985d). Thus for this purpose the algorithm (4.4.9) will be modified as

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-1)\phi(t)[-h(t) - \phi^T(t)\hat{\theta}(t-1)]}{\{1/\alpha(t)\} + \phi^T(t)P(t-1)\phi(t)} \quad (4.11.3a)$$

and

$$P(t) = P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{\{1/\alpha(t)\} + \phi^T(t)P(t-1)\phi(t)} \quad (4.11.3b)$$

with

$$\alpha(t) = \frac{\phi[-h(t) - \phi^T(t)\hat{\theta}(t-1)]}{[-h(t) - \phi^T(t)\hat{\theta}(t-1)]} \quad (4.11.3c)$$

and $\alpha(t) := 1$ when $[y(t) - \phi^T(t)\hat{\theta}(t-1)] = 0$,

with starting values as before.

Note that

$$\phi^T(t) = [h(t-1), h(t-2), \dots, h(t-n)]$$

and

$$\theta^T = [a_1, a_2, \dots, a_n]$$

After getting these autoregressive parameters, the moving average parameters can be obtained by using (4.11.2a).

4.11.3 Notes on convergence

In the line light of theorem 4.1, the convergence result of the above algorithm can be stated as follows.

Corollary 4.2 [of theorem 4.1]

Let the $\psi(\cdot)$ function of algorithm (4.11.3) satisfy the conditions put forward by theorem 4.1. Then

$$\lim_{t \rightarrow \infty} \hat{\theta} = \theta_0 \text{ in quadratic mean}$$

provided

$$(i) \limsup_{N \rightarrow \infty} (1/N) \sum_{t=1}^N [h(t)]^2 < \infty \quad (\text{strictly stable system})$$

and (ii) $\{h(t)\}$ and the noise sequence $\{e(t)\}$ are statistically independent.

4.11.4 Simulation results

We again consider the system given by (5.8.1). The input is now a pulse of height 10 and duration one sampling interval. To the noise free output, measurement noise and outliers are added exactly as before. The parameter error norms are calculated and plotted (figures 4.15, 4.16 and 4.17) for a typical case where outliers are from a Laplacian pdf. These plots are self explanatory.

4.12 Concluding remarks

The main motivation behind the discussions of this chapter was to present a new class of recursive identification algorithms, which can be considered as a wide extension of the well known weighted least squares algorithm, based on a robust statistical method called M-estimation. This result is obtained by introducing the 'influence function' involved in the determination of the

maximum likelihood estimator in the weight of the the recursive least squares type of algorithms.

The convergence of the new class of algorithms is proved for a general class of influence functions for different cases. It is notable that the lemmas used in the proofs do not require the continuity of the influence function; thus the cost function to be optimized can be non differentiable. For example, one can minimize the total absolute prediction error instead of the total squared prediction error.

The results of identification of the packed-bed reactor using actual experimental data suggest that the performance of the robust algorithm is generally better in the context of real-life where the data is seldom normal.

In addition, the implementation of these robust algorithms is inexpensive since they demand only minor modifications of the existing software, developed for conventional identification methods.

Table 4.1

Result of identification of the packed bed reactor using recursive least squares method

	θ_1	θ_2	θ_3	θ_4	θ_5
Parameter	-0.0643	1.4879	-0.3774	-1.0867	0.4522
Standard error	0.1661	0.6489	0.7396	0.0576	0.0152
Statistics	Residual variance = 1.429×10^{-3}				
	Residual Kurtosis = 14.96				
	Chi-square of residuals (20 degrees of freedom) = 30.10				

Table 4.2

Result of identification of the packed bed reactor using the robust (weighted least squares type) method

	θ_1	θ_2	θ_3	θ_4	θ_5
Parameter	-0.3390	1.6193	-0.4762	-1.1867	0.5521
Standard error	0.0108	0.0467	0.0552	0.0056	0.0016
Statistics	Residual variance = 6.238×10^{-4}				
	Chi-square of residuals (20 degrees of freedom) = 24.47				

Table 4.3
Result of identification of the packed bed reactor using
the conventional IV method

	θ_1	θ_2	θ_3	θ_4	θ_5
Parameter	-0.0668	1.5030	-0.3805	-1.0989	0.5004
Standard error	0.1537	0.3071	0.3947	0.0481	0.0142
Statistics	Residual variance = 1.314×10^{-3}				
	Residual Kurtosis = 11.37				
	Chi-square of residuals (20 degrees of freedom) = 28.19				

Table 4.3
Result of identification of the packed bed reactor using
the robust IV method

	θ_1	θ_2	θ_3	θ_4	θ_5
Parameter	-0.3391	1.6193	-0.4760	-1.1866	0.5212
Standard error	0.0107	0.0466	0.0569	0.0055	0.0015
Statistics	Residual variance = 6.199×10^{-4}				
	Chi-square of residuals (20 degrees of freedom) = 23.99				

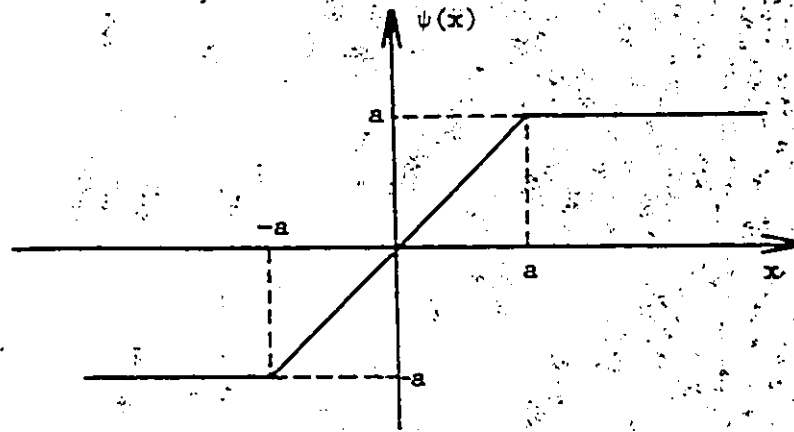


Fig. 4.1 Huber's Influence Function

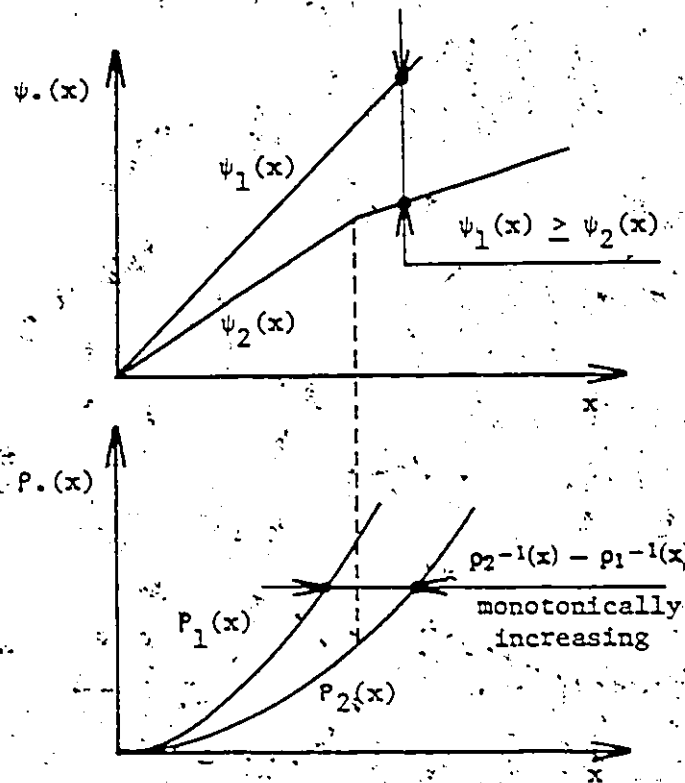


Fig. 4.2a Geometric Interpretation of Theorem 4.5

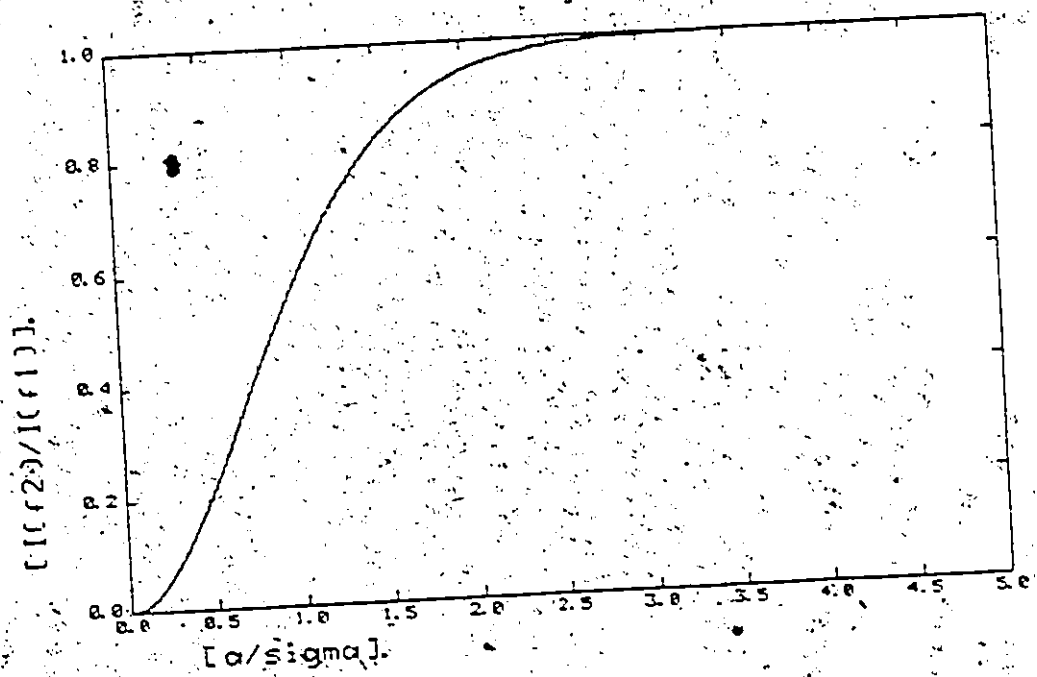


Fig. 4.2b $I(f_2)/I(f_1)$ for Different a/σ

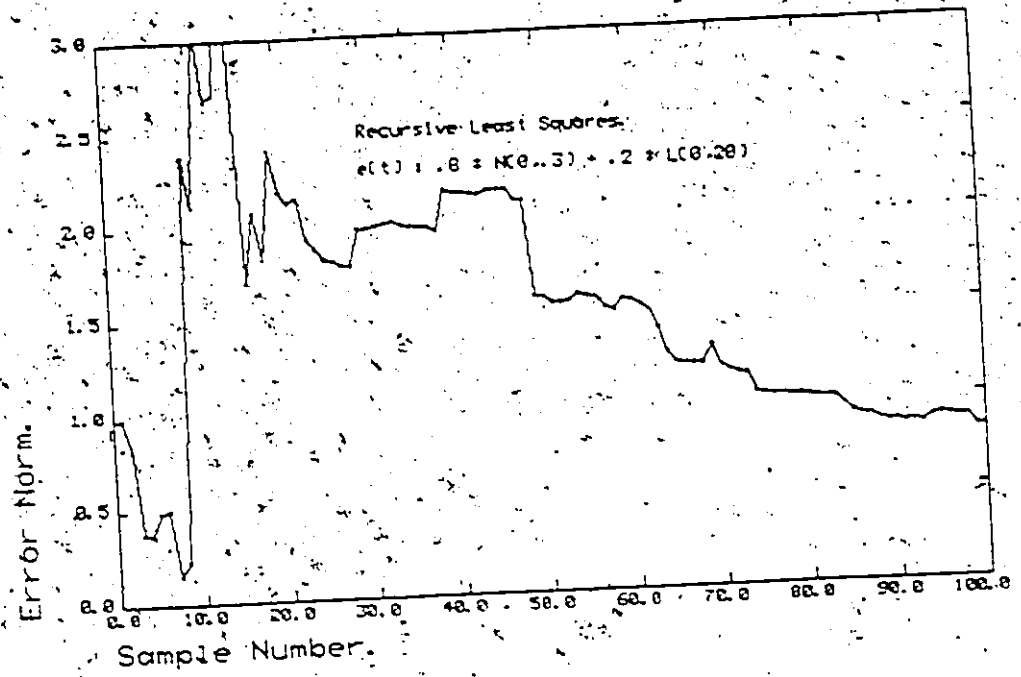


Fig. 4.3 Parameter Error Norm: Least Squares

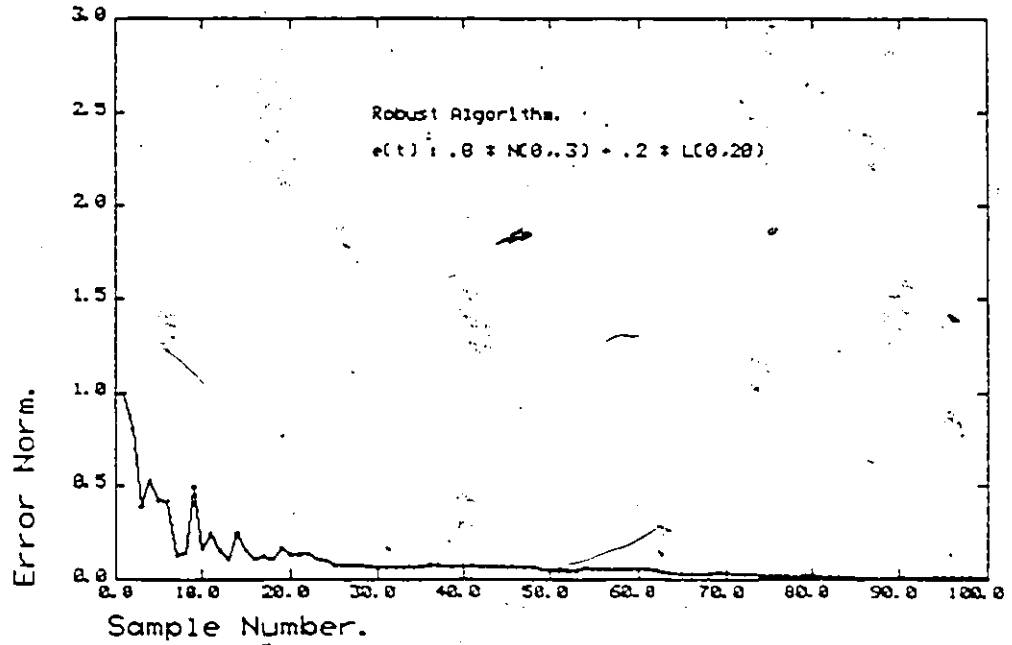


Fig. 4.4 Parameter Error Norm: Robust Weighted Least Squares Type Algorithm

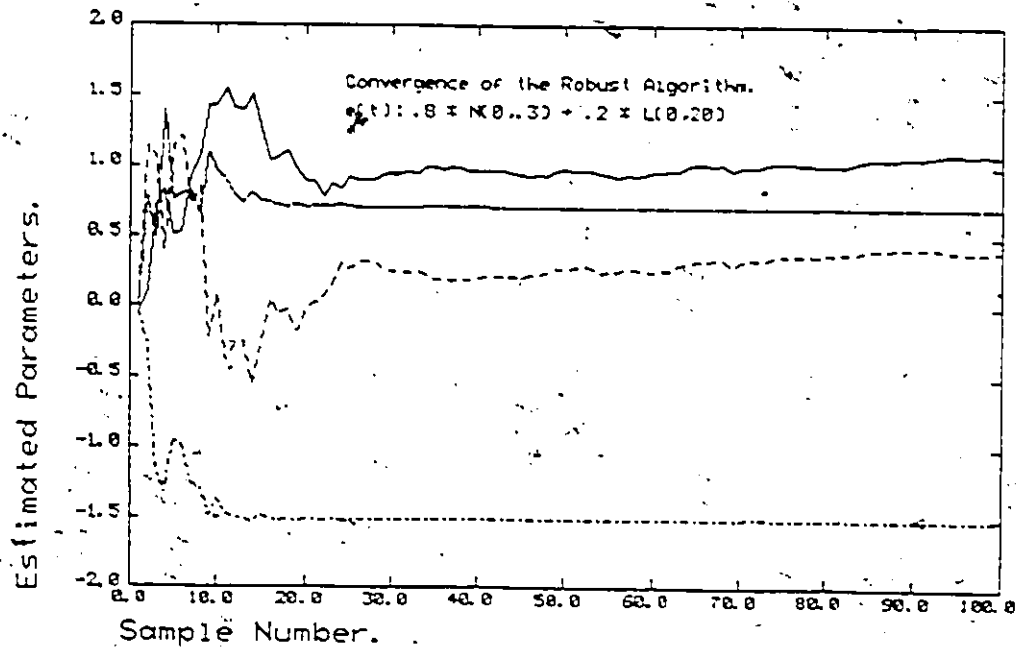


Fig. 4.5 Convergence of Parameters: Robust Weighted Least Squares Type Algorithm

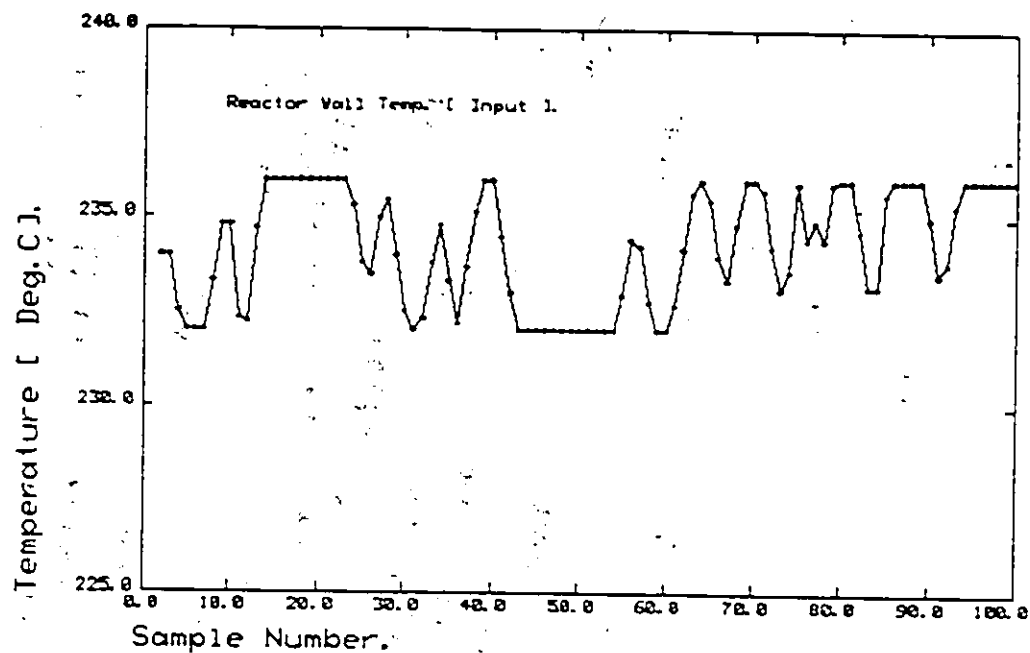


Fig. 4.6 Variation of Reactor Wall Temperature Over 5 Hours

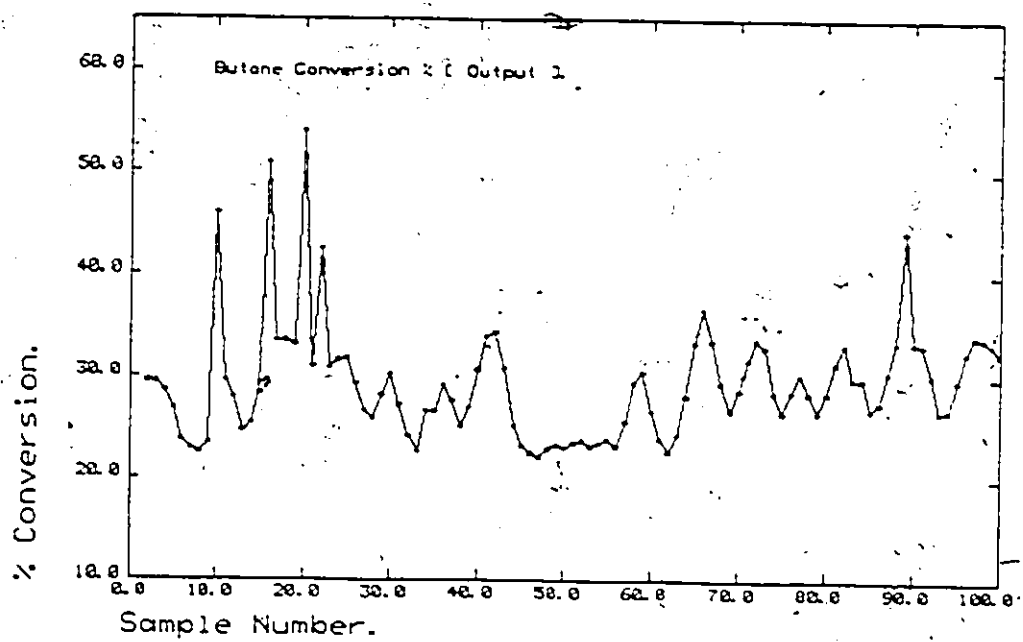


Fig. 4.7 Variation of Percentage Butane Conversion Over 5 Hours

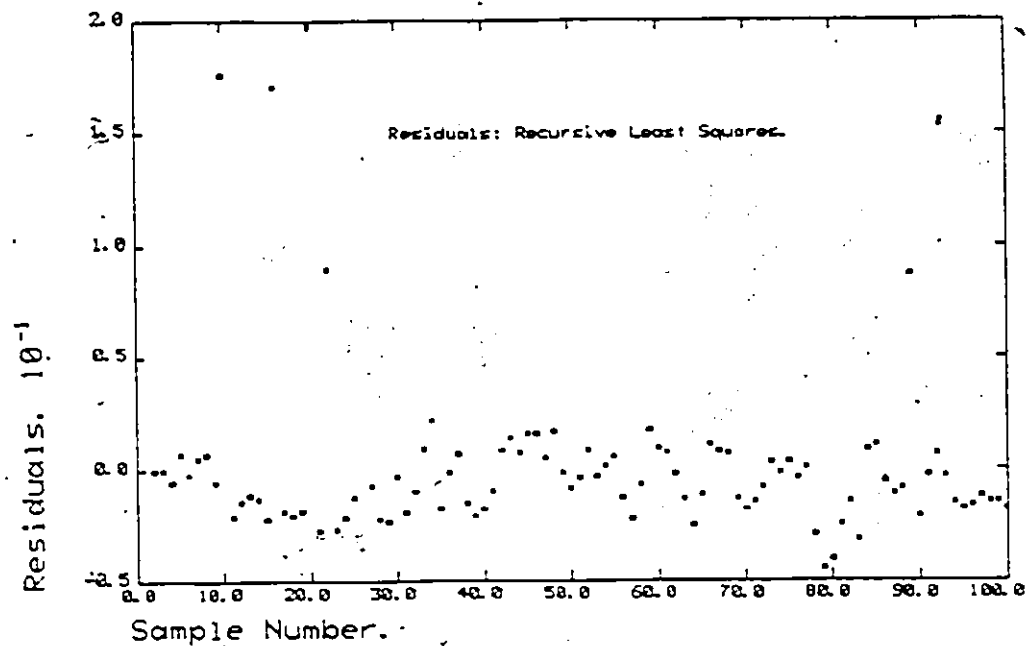


Fig. 4.8 Residuals of the Least Squares Model

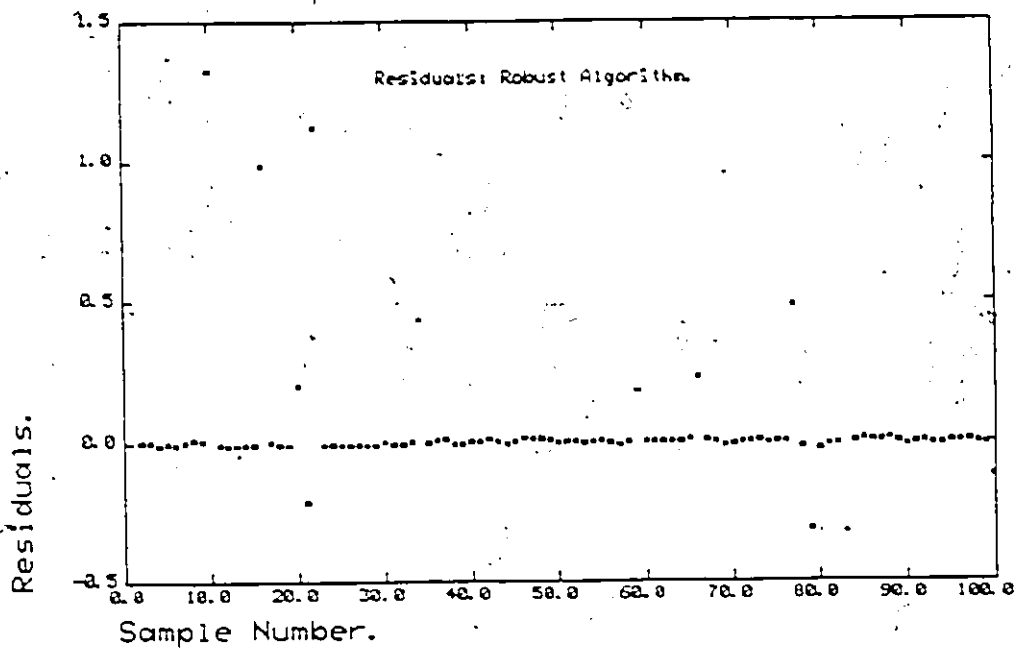


Fig. 4.9 Residuals of the Robust Model

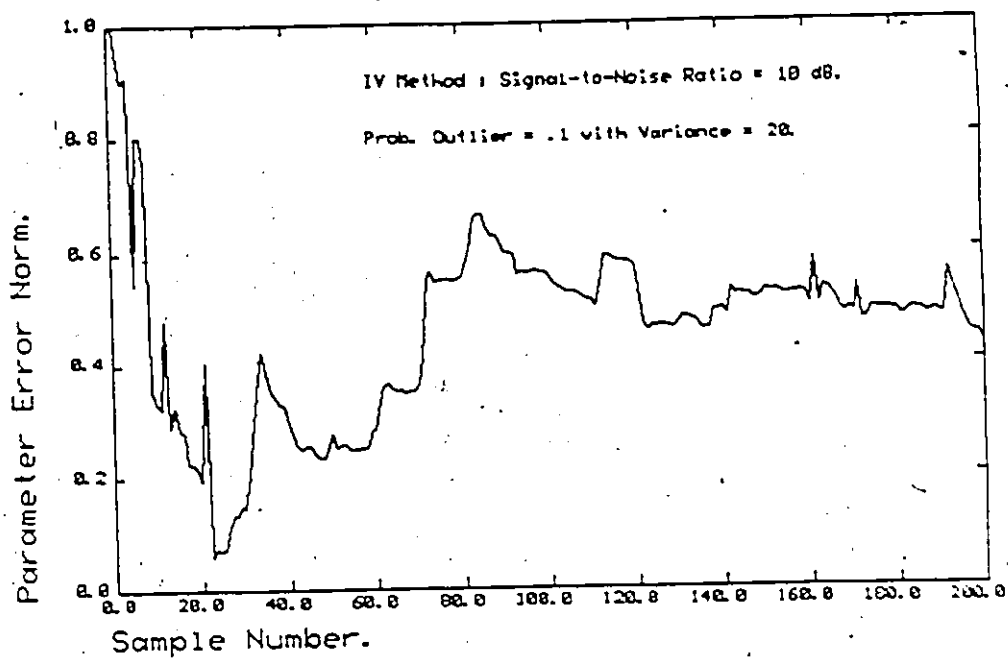


Fig. 4.10 Parameter Error Norm of the IV Method in Presence of Outliers

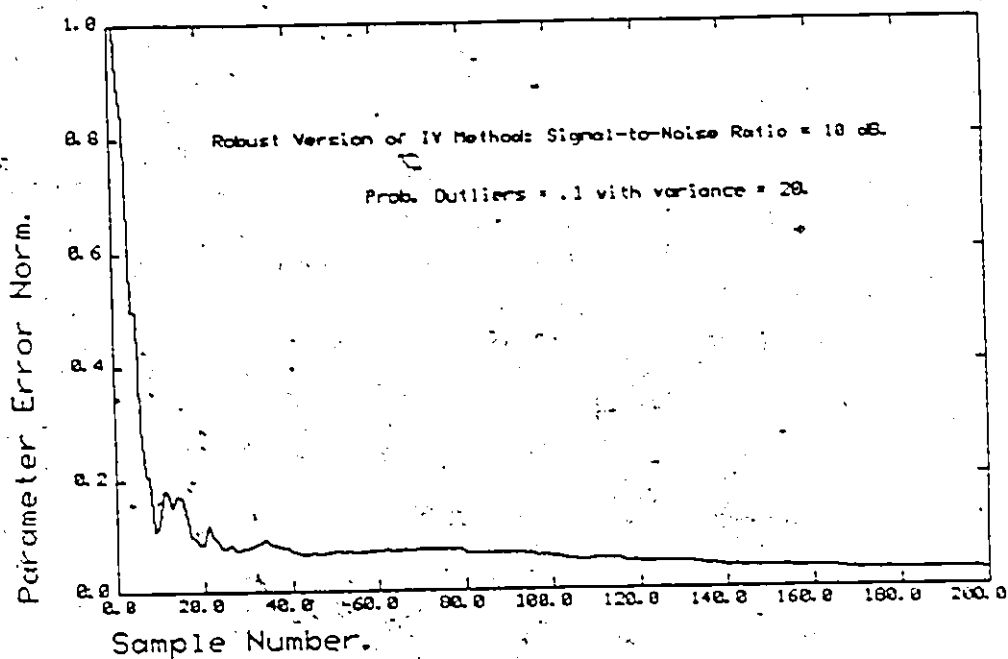


Fig. 4.11 Parameter Error Norm of the Robust IV Method in Presence of Outliers

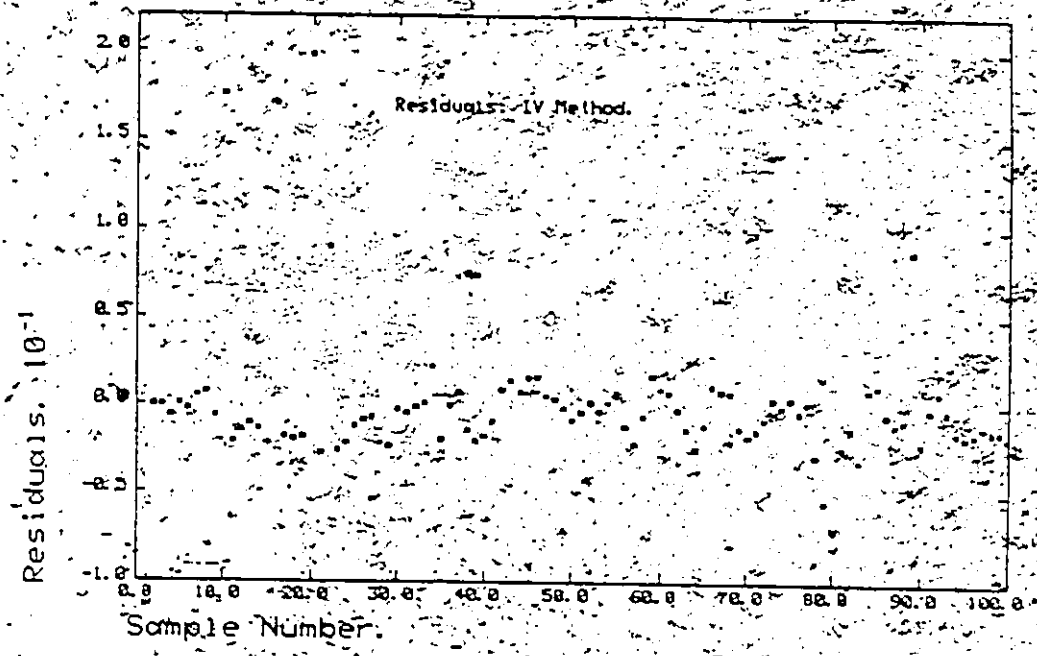


Fig. 4.12 Residuals of the IV Model.

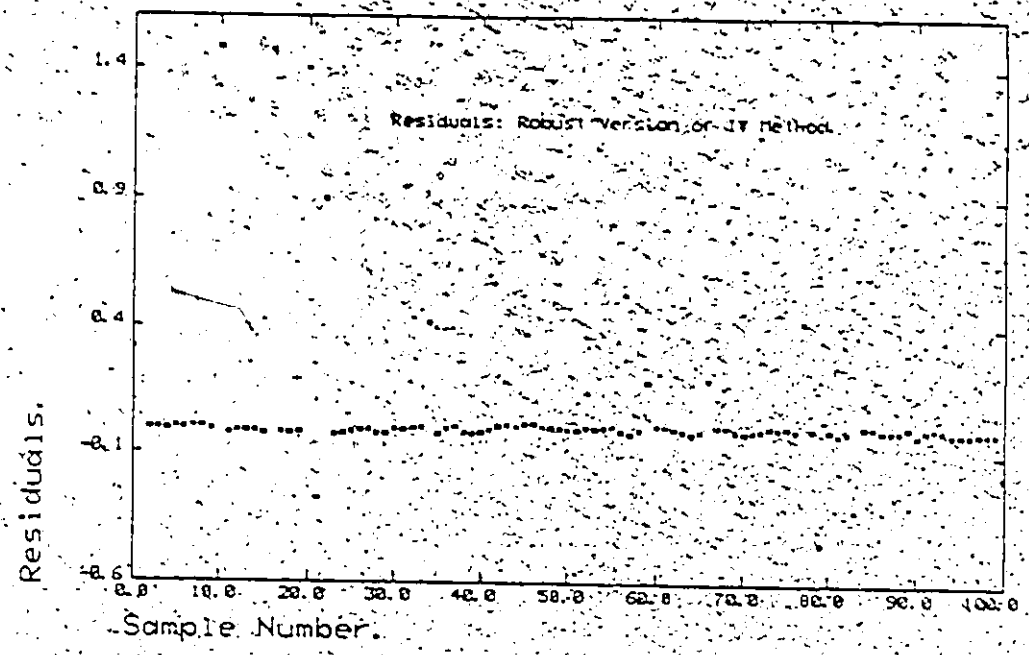


Fig. 4.13 Residuals of the Robust IV Model

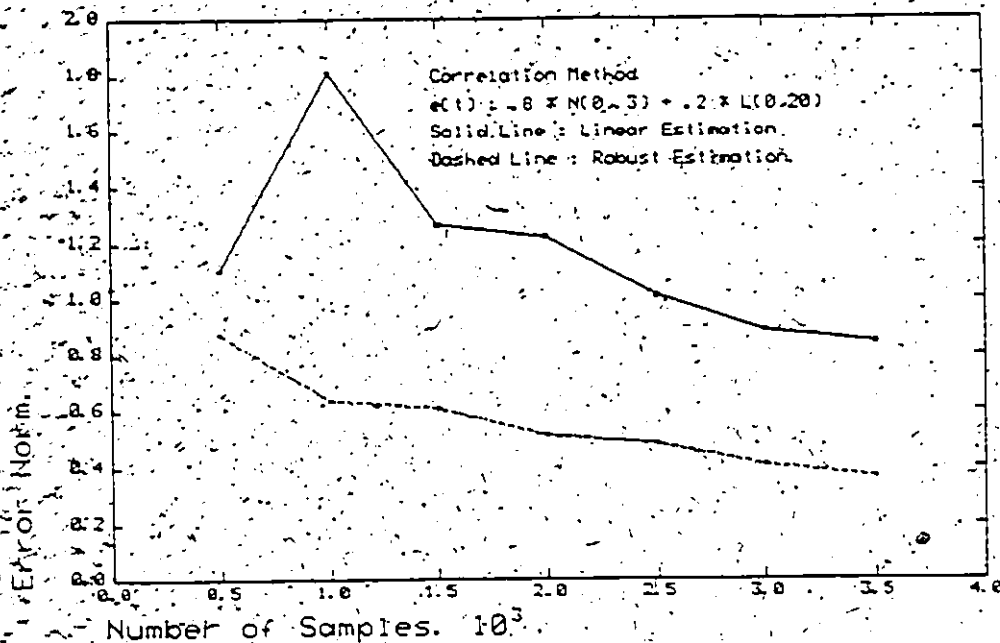


Fig. 4.14 Comparison of Linear and Robust Versions of the Recursive Correlation Method

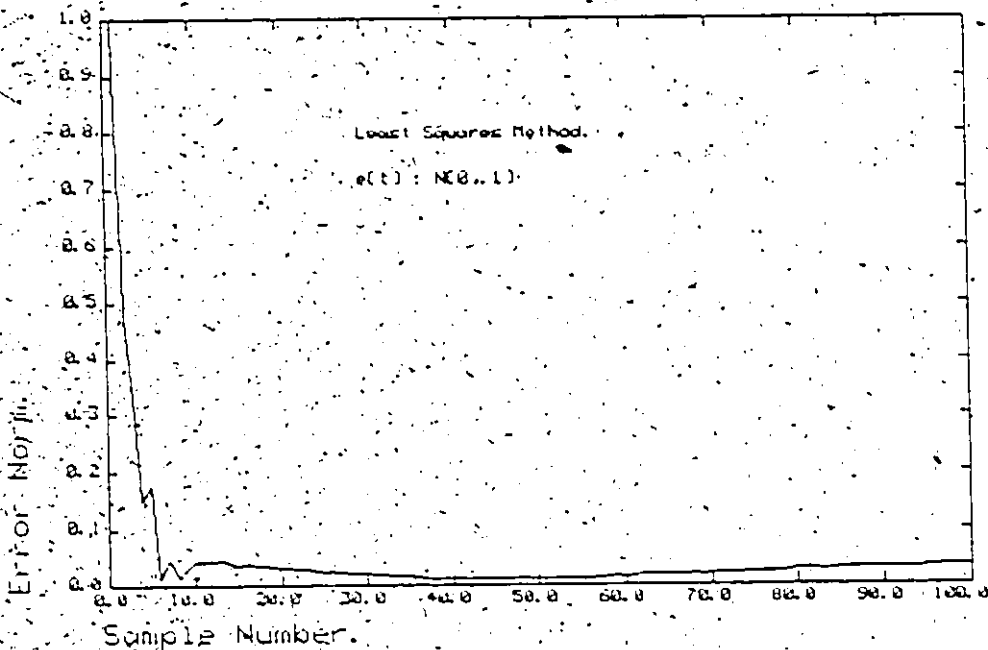


Fig. 4.15 Parameter Error Norm: Identification from Impulse Response, Least Squares Method (in presence of no outliers)

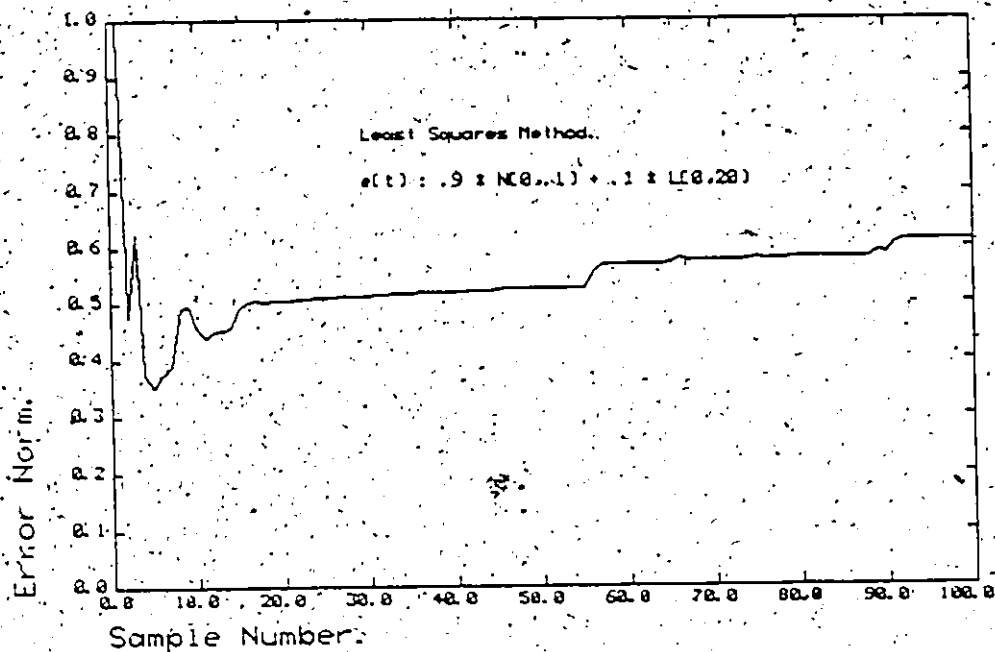


Fig. 4.16 Parameter Error Norm: Identification from Impulse Response, Least Squares Method (in presence of outliers)

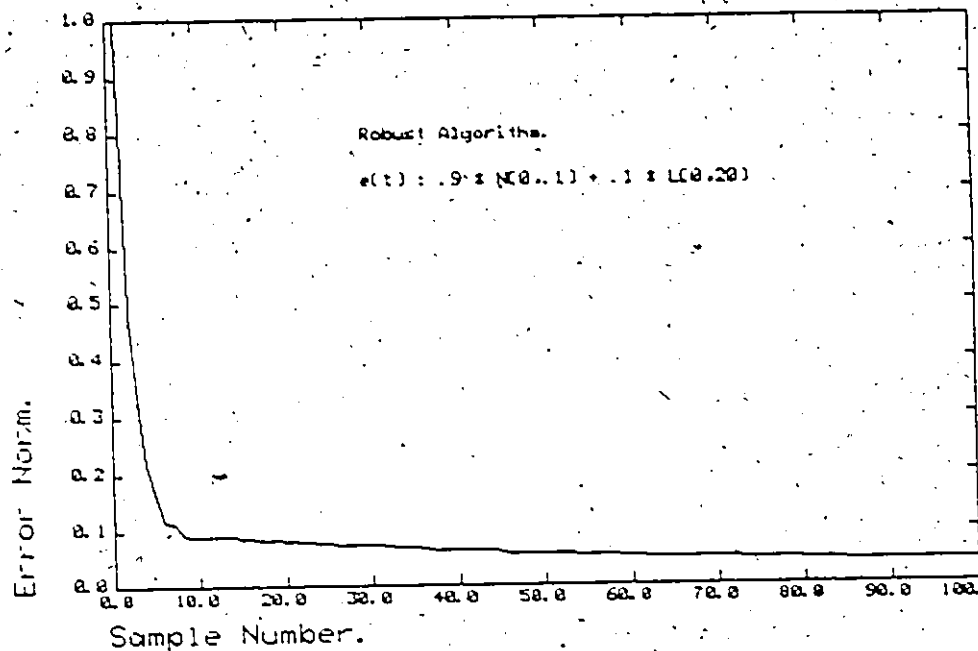


Fig. 4.17 Parameter Error Norm: Identification from Impulse Response, Robust Method (in presence of outliers)

CHAPTER 5
ROBUST RECURSIVE IDENTIFICATION - A BAYESIAN
APPROACH

5.1 Overview of the chapter

A robust recursive parameter estimation algorithm is derived in a Bayesian frame work. The resulting algorithm is very similar to the weighted least squares type algorithm presented in section 4.4. Here also, the powerful features of the suggested algorithm are illustrated through simulated as well as real-life examples.

5.2 Introduction to contaminated density functions and
Huber's minimax approach

One way to characterize the outlier problem is to consider outliers as another class of observation errors superimposed over the regular observation noise. Based on this idea, Huber (1964) developed the following procedure, called minimax approach.

Consider a class of estimates T and a class of distribution F . Define $V(T, F)$ as the asymptotic variance of T when the distribution is F . If we choose our estimate T while the data belong to the distribution is F (where F is unknown), the pay off will be $V(T, F)$. Therefore we have to look for the saddle point solution pair (T_0, F_0) such that

$$\min_{T \in T} \max_{F \in F} V(T, F) = V(T_0, F_0) = \max_{F \in F} \min_{T \in T} V(T, F) \quad (5.2.1)$$

where T_0 is called the 'minimax robust estimator' and F_0 is called the 'least informative or least favourable' distribution (Huber, 1964). Normally we assume F to be Gaussian but it is too idealistic. In our application, we consider the 'u contaminated family' F_u such that (Huber, 1964)

$$F_u = \{F | F = (1-u)G + uH, 0 \leq u \leq 1\} \quad (5.2.2)$$

where G represents Gaussian distribution, H is some symmetrical distribution and u is the probability of occurring outliers. We simplify the matter further by assuming

$$F_u = F_u(\cdot | 0, \sigma_1^2, \sigma_2^2) = (1-u)N(\cdot | 0, \sigma_1^2) + uN(\cdot | 0, \sigma_2^2) \quad (5.2.3)$$

where $N(\cdot | \mu, \sigma^2)$ denotes normal density with mean μ and variance σ^2 . Here the assumption is that the occasional large errors are also from a Gaussian distribution with larger variance (ie., $\sigma_1^2 < \sigma_2^2$) or with a longer tail. We take (5.2.3) as the least favourable distribution and look for an estimate of system parameters which minimizes the asymptotic variance under this assumption. Based on this, we can obtain a recursive weighted least squares type algorithm in a Bayesian approach (Puthenpura and Sinha, 1985e). Note that, under these circumstances, the estimates obtained by Bayesian approach are minimum variance estimates (thus

satisfying the saddle point condition 5.2.1) since F_0 given by (5.2.3) satisfies all the requirements for the variance of the estimates to reach the minimum variance bound, asymptotically (Huber 1964, 1981).

5.3 Robust recursive estimation, Bayesian approach

5.3.1 Introduction

In Bayesian approach, the parameter itself is considered as a random variable. Based on observations on another random variable which is correlated with the parameter, we try to estimate the parameter. The celebrated Kalman filter was developed in such a set up. The only difference is that the Kalman filter estimates the states instead of parameters. Here we assume that the parameter vector θ described by (4.3.3) is a random vector with certain prior distribution. From the fact that the input and output of the system are correlated to θ , we aim to determine the posterior probability density of θ , i.e., $p[\theta|y(t), u(t)]$ so that the estimate of θ can be found by maximizing the posterior probability density function with respect to θ . This is called the maximum posteriority (MAP) estimation method. Assuming that the input sequence $\{u(t)\}$ is purely deterministic, we deduce that the posterior probability density functions for θ can be written as $p[\theta|y(t)]$. Also note that the MAP estimator coincides with the maximum likelihood estimator in the case of a uniform prior distribution (often called a 'non-informatory prior').

5.3.2 Derivation of recursive identification algorithms

Now we assume that the errors in data are also observation errors superimposed over the usual noise contaminating the output. Therefore, the error term $e(t)$ of (4.3.3) is assumed to belong to the v contaminated distribution (5.2.3). Hence we can write

$$E[e(t)] = 0$$

and

$$E[e(t) \cdot e(t)] = (1-u)\sigma_1^2 + u\sigma_2^2 \quad (5.3.1)$$

Thus

$$p[y(t)|\theta, y^{t-1}] = \frac{\exp\{-[y(t) - \Phi^T(t)\hat{\theta}(t-1)]^2 / [2[(1-u)\sigma_1^2 + u\sigma_2^2]]\}}{\sqrt{2\pi[(1-u)\sigma_1^2 + u\sigma_2^2]}}$$

(5.3.2)

Applying the famous Bayes's rule

$$p[\theta|y^t] = \frac{p[\theta|y(t), y^{t-1}] p[y(t)|\theta, y^{t-1}] p[\theta|y^{t-1}]}{p[y(t)|y^{t-1}]}$$

(5.3.3)

where the superscript $[y^{t-1}]$ indicates the observations upto the $(t-1)^{th}$ instant.

Also assuming that the prior distribution of θ is Gaussian with mean $\theta(0)$ and covariance matrix $P(0)$, the posterior distribution is also Gaussian with mean $\hat{\theta}(t)$ and

covariance matrix $P(t)$ which implies

$$p[\theta|y^0] = \frac{(2\pi)^{-\dim\theta/2} \exp[-(1/2)\{\theta - \hat{\theta}(0)\}^T P^{-1}(0)\{\theta - \hat{\theta}(0)\}]}{\sqrt{\det P(0)}} \quad (5.3.4a)$$

and

$$p[\theta|y^{t-1}] = \frac{(2\pi)^{-\dim\theta/2} \exp[-(1/2)\{\theta - \hat{\theta}(t-1)\}^T P^{-1}(t-1)\{\theta - \hat{\theta}(t-1)\}]}{\sqrt{\det P(t-1)}} \quad (5.3.4b)$$

Now, maximizing (5.3.4b) with the help of (5.3.2), (5.3.3) and (5.3.4a) we get the following recursive relationships (Ljung and Soderstrom, 1983).

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-1)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{[(1-u)\sigma_1^2 + u\sigma_2^2] + \phi^T(t)P(t-1)\phi(t)} \quad (5.3.5a)$$

and

$$P(t) = P(t-1) - \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{[(1-u)\sigma_1^2 + u\sigma_2^2] + \phi^T(t)P(t-1)\phi(t)} \quad (5.3.5b)$$

The starting values can be taken as,

$$\hat{\theta}(0) = \underline{0}; \quad P(0) = kI, \quad k \gg 0. \quad (5.3.5c)$$

If one knows the values of u , σ_1^2 and σ_2^2 , the algorithm (5.3.5) can be used as a robust estimation

procedure to obtain the parameter vector θ . But in practice we do not know the actual values of u , σ_1^2 and σ_2^2 and wild guess on these parameters may not be effective. Therefore, it is necessary to modify (5.3.5) so that it will be suitable for practical applications.

One obvious solution to this problem is to draw information from the prediction error,

$$e(t) = [y(t) - \phi^T(t)\hat{\theta}(t-1)]. \quad (5.3.6)$$

Define $\delta(t)$ such that

$$\begin{aligned} \delta(t) &= 0 \text{ if } |e(t)| < M \\ \delta(t) &= 1 \text{ if } |e(t)| > M \end{aligned} \quad (5.3.7)$$

where M is some pre-assigned bound, which can be taken as $3\sigma_1$. The idea is to make $\text{prob}[\delta(t)=1] = u$. Now modify (5.2.3) as

$$F_{\delta(t)} = F_{\delta(t)}(.|0, \sigma_1^2; \sigma_2^2) = [1-\delta(t)]N(.|0, \sigma_1^2) + \delta(t)N(.|0, \sigma_2^2) \quad (5.3.8)$$

The mixed pdf is shown in fig.5.1.

As a result of this modification, (5.3.5) will be changed to

$$\begin{aligned} \hat{\theta}(t) = & \\ & \hat{\theta}(t-1) + \frac{P(t-1)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{[[1-\delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2] + \phi^T(t)P(t-1)\phi(t)} \end{aligned} \quad (5.3.9a)$$

and

$$P(t) = \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{[[1-\delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2] + \phi^T(t)P(t-1)\phi(t)} \quad (5.3.9b)$$

where $\delta(t)$ is defined as (5.3.7).

Moreover, if necessary, one can update σ_1^2 and σ_2^2 for large sample size as,

$$\hat{\sigma}_1^2(t) = \hat{\sigma}_1^2(t-1) + [1/\tau][\varepsilon^2(t) - \hat{\sigma}_1^2(t-1)] \quad \text{for } |\varepsilon(t)| < 3\hat{\sigma}_1(t-1)$$

and

$$\hat{\sigma}_2^2(t) = \hat{\sigma}_2^2(t-1) + [1/\tau][\varepsilon^2(t) - \hat{\sigma}_2^2(t-1)] \quad \text{for } |\varepsilon(t)| > 3\hat{\sigma}_1(t-1) \quad (5.3.9c)$$

where τ represents the number of times that $|\varepsilon(t)|$ exceeded the $3\hat{\sigma}_1$ bound until time t . $\hat{\sigma}_1^2(0)$ and $\hat{\sigma}_2^2(0)$ are the initial guess on variances. Note that the estimates of variances given by (5.3.9c) are unbiased for sufficiently large values of t and τ , hence the formula is effective only for large sample sizes. A good modification of (5.3.9c) could be that proposed by Young (1984). In that method, poor and early values of $\varepsilon(t)$ are exponentially forgotten. for example, the estimate of σ_1^2 is modified as

$$\hat{\sigma}_1^2(t) = \hat{\sigma}_1^2(t-1) + [\gamma(t)][\varepsilon^2(t) - \hat{\sigma}_1^2(t-1)]$$

where

$$\gamma(t) = \gamma(t-1) - [1/\lambda(t)][\gamma^2(t-1)/\{\gamma(t-1) + \lambda(t)\}]$$

and

$$\lambda(t) = \lambda_0 \lambda(t-1) + (1-\lambda_0) \quad (5.3.9d)$$

typically $\lambda(0) = 0.95$ and $\lambda_0 = 0.99$.

Similar expression holds good for σ_2^2 also.

Thus the algorithm (5.3.9) is a very practicable and easy to use one. It is also noteworthy that this algorithm is a weighted least squares method with weights taken as $\{[1-\delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2\}^{-1}$.

5.3.3 Tracking time-varying systems

One possible extension of algorithm (5.3.9) is the tracking of time-varying parameter of a system (Ljung and Soderstrom, 1983). Here, instead of assuming the true parameter θ_0 to be constant, we consider

$$\theta_0(t+1) = \theta_0(t) + w(t)$$

and

$$y(t) = \phi^T(t)\theta_0(t) + e(t) \quad (5.3.10)$$

where $e(t)$ is assumed to of a mixed pdf given by (5.3.8) and statistically independent to $w(t)$.

ie.,

$$E[w(t)w^T(t)] = \Omega(t)I$$

and $E[w(t)e(s)] = 0 \quad \forall t \text{ and } s.$

Under these circumstances we get the following updating equations.

$$\hat{\theta}(t) =$$

$$\hat{\theta}(t-1) + \frac{P(t-1)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{\{[1-\delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2\} + \phi^T(t)P(t-1)\phi(t)} \quad (5.3.11a)$$

and

$$P(t) = \Omega(t)I + \frac{P(t-1)\phi(t)\phi^T(t)P(t-1)}{\{[1-\delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2\} + \phi^T(t)P(t-1)\phi(t)} \quad (5.3.11b)$$

5.4 Convergence and efficiency of the algorithm

The algorithm introduced in the previous section can be considered as a special case of the algorithm (4.4.9) with the influence function

$$\psi(x) = \begin{cases} x/\sigma_1^2 & ; |x| \leq M \\ x/\sigma_1^2 & ; |x| > M \end{cases} \quad (5.4.1)$$

With this information, the following result can be stated.

Corollary 5.1 [of theorem 4.1]

Let the following assumptions hold good.

- (i) $e(t)$ has a pdf given by (5.3.8)
- (ii) $\phi(t)$ and $e(t)$ are statistically independent,
- (iii)

$$\limsup_{N \rightarrow \infty} (1/N) \sum_{t=1}^N |\phi(t)|^2 < \infty \quad \text{and}$$

$$(iv) \quad 0 < \sigma_1^2, \sigma_2^2 < \infty$$

Then for the algorithm (5.3.9);

$\lim_{t \rightarrow \infty} \theta(t) = \theta_0$ in quadratic mean.

Proof:

Note that the influence function (5.4.1) meets all the requirements put forward by theorem 4.1, subjected to the conditions of the corollary and the proof is immediate.

Note: Sometimes it is found that discarding an 'outlier' completely and proceeding to the next observation gives biased estimates. This is equivalent to setting $\sigma_z^2 = -$ in (5.3.9) and (5.3.11) which is against condition (iv) of corollary 5.1 and the result follows.

The evaluation of efficiency of the above algorithms is also in the same way as described in section 4.7. But here the treatment is quite simple and it is presented below.

Proposition 5.1

When $\{e(t)\}$ actually enjoys normality the robust algorithms presented above are less efficient than their non-robust counter parts.

Proof:

The non-robust algorithm has a weight equal to $1/\sigma_1^2$, inverse of the variance of the measurement noise (eg. Ljung and Soderstrom, 1983).

If the probability of occurring outliers is u then the scalar Fisher information $I(f)$ corresponding to the

underlying pdf of the above algorithms is

$$1/[(1-u)\sigma_1^2 + u\sigma_2^2] < 1/\sigma_1^2$$

when $u > 0$ and the result follows. •

5.5 Examples of application of the proposed method

5.5.1 Simulation results

Let us consider the system given by (4.8.1) once again. The data containing outliers is generated in the same way as described in section (4.8.1), but this time the outliers are from another Gaussian pdf with a larger variance. The same error criterion (4.8.2) is used to compare the performance of the conventional optimal weighted least squares algorithm and the robust algorithm (5.3.9).

Note that the conventional algorithm has a weight $1/\sigma_1^2$.

Figures 5.2, 5.3 and 5.4 show typical results, when for a standard signal to noise ratio of 10 dB., which do not require much explanations. Fig.5.2 shows the error norm of the estimates in case of no outliers, which 'explodes' in the presence of outliers as suggested by fig.5.3. But the application of the proposed algorithm (5.3.9) with starting values of variances 0.2 and 10. improves the results, as obvious from fig.5.4. substantial improvement of the error norm when the robust algorithm is used for identification.

5.5.2 The packed-bed reactor identification problem

Let us now consider the packed bed reactor identification problem introduced in section 4.8.2.

A difficulty one encounters here is the length of the

data. It is obvious that the estimates of σ_1^2 and σ_2^2 by the adaptation algorithm (5.3.9c) or (5.3.9d) improve as the length of the data increases. Here it is found that 100 data points (data over 5 hours) does not give good results because of the above reason. Therefore 60 more points are collected (data over 8 hours) to use the algorithm (5.3.9). Refer fig.5.5 and 5.6 to see the input output data. This gave satisfactory results, which are given in table 5.1. The initial estimates of the variances are taken as 0.1 and 4.0, respectively.

The residuals of the least squares fit and the robust fit are given in fig.5.7 and 5.8 respectively.

Comparing the results of table 5.1 and the previous results, it can be seen that this algorithm gave poorer results in contrast with the algorithms presented in the previous chapter. This is further emphasised by the residual plots too. However, the present algorithm is very simple and seems to be adequate for practical applications.

5.6 Concluding remarks

A natural question arises at the end of the discussions of this chapter is this. Since the algorithm presented in this chapter is a special case of the M-estimation algorithm (which is fairly general), what is the use of introducing the same algorithm in a different framework? In fact the real-life example showed that the results are somewhat inferior compared to the original M-estimation

algorithms! The answer to this question is this. The power of the above algorithm lies in its simplicity and versatility. It can be easily and directly extended to a variety of problems like identification of multivariable systems, joint parameter-state estimation (see the next chapter) and filtering. In fact, a straightforward extension of the above algorithm to the robust estimation of satellite orbits (which is basically a filtering problem) can be found in a very recent work by Raol (1985).

Table 5.1
Result of identification of the packed bed reactor using
the least squares method and the robust algorithm
based on mixed Gaussian pdf

Parameter	θ_1	θ_2	θ_3	θ_4	θ_5
Least squares	0.2214	0.9001	0.0432	-1.0821	0.5346
Robust method	-0.3972	1.5137	-0.4301	-1.1347	0.4914
Standard error	θ_1	θ_2	θ_3	θ_4	θ_5
Least squares	0.1837	0.6198	0.7198	0.3171	0.1159
Robust method	0.1001	0.0537	0.1039	0.1170	0.1037

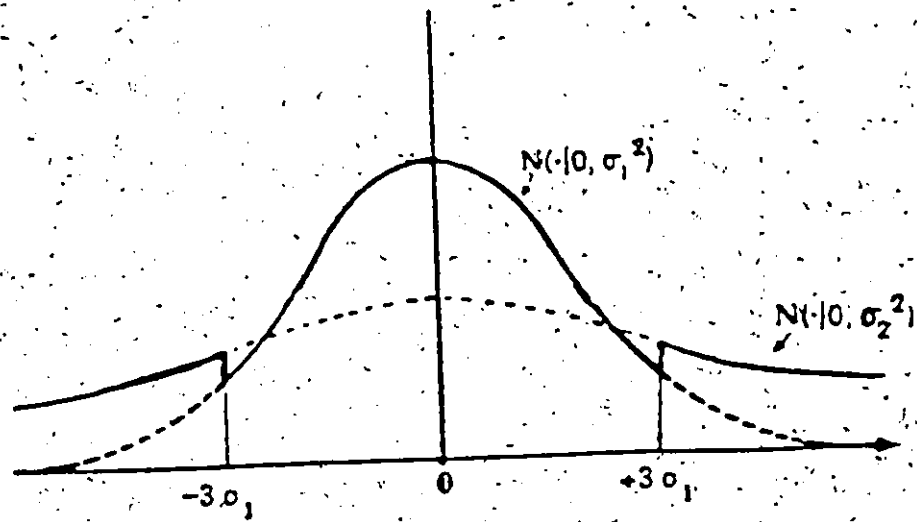


Fig. 5.1 The Mixed Gaussian pdf

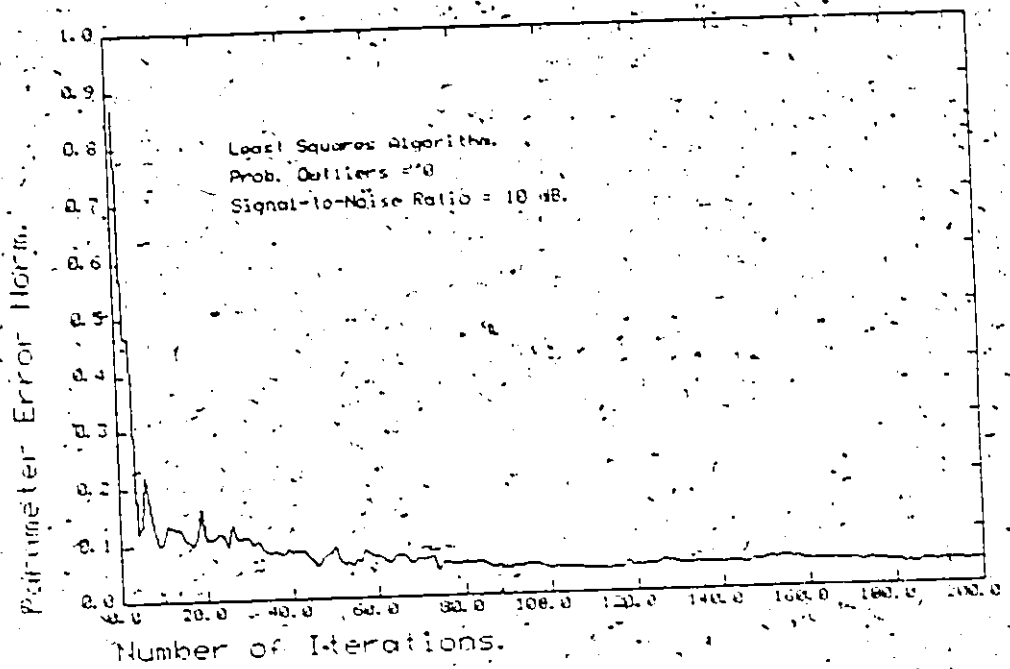


Fig. 5.2 Parameter Error Norm: Least Squares with no Outliers in the Data

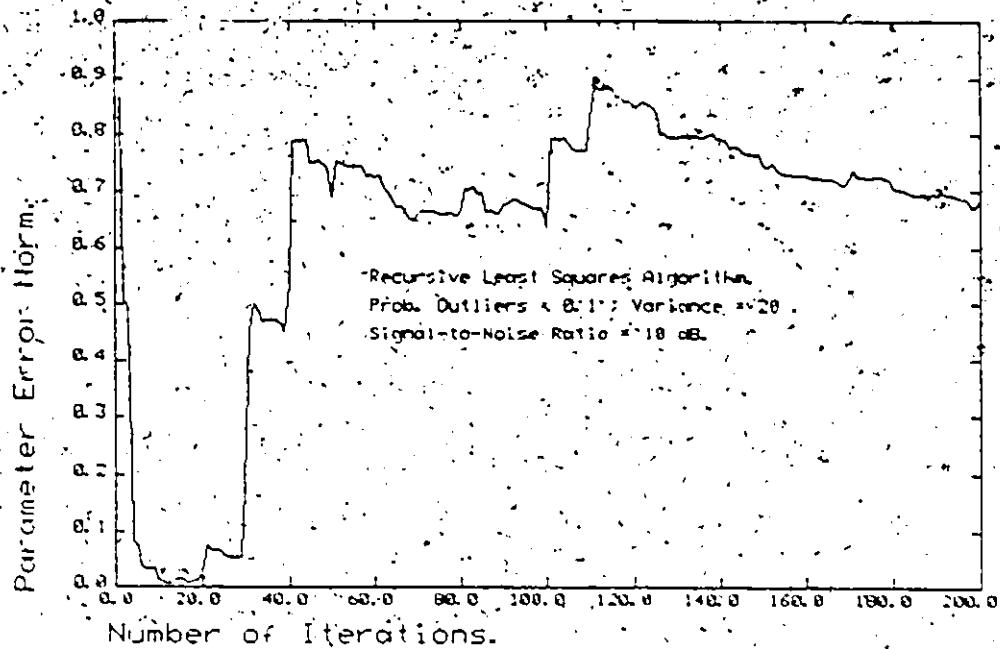


Fig. 5.3 Parameter Error Norm: Least Squares with Outliers
 in the Data

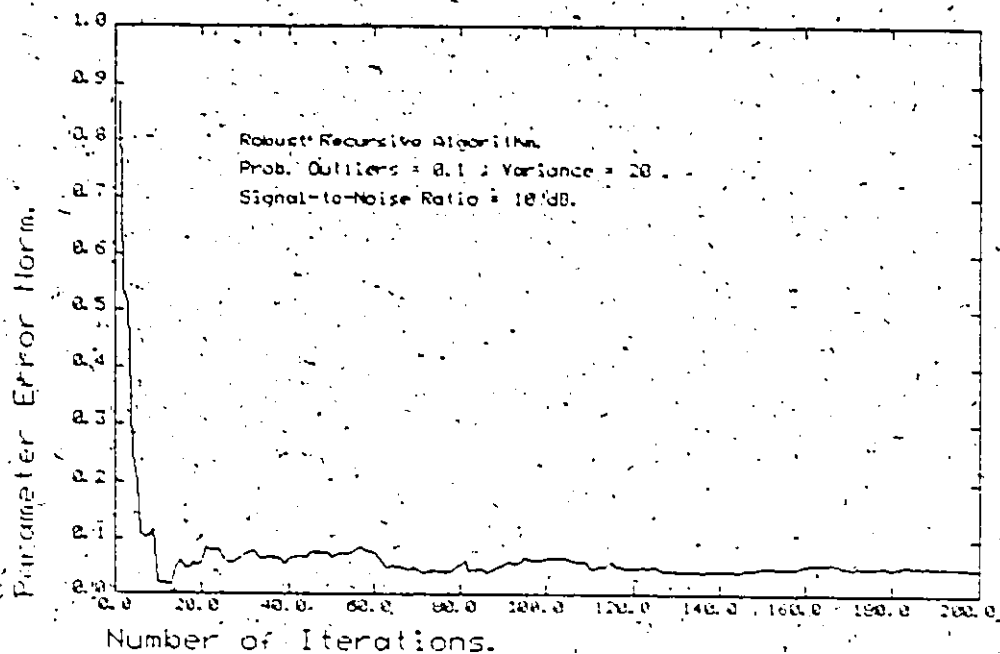


Fig. 5.4 Parameter Error Norm: Robust Algorithm (Bayesian
 Approach) with Outliers in the Data

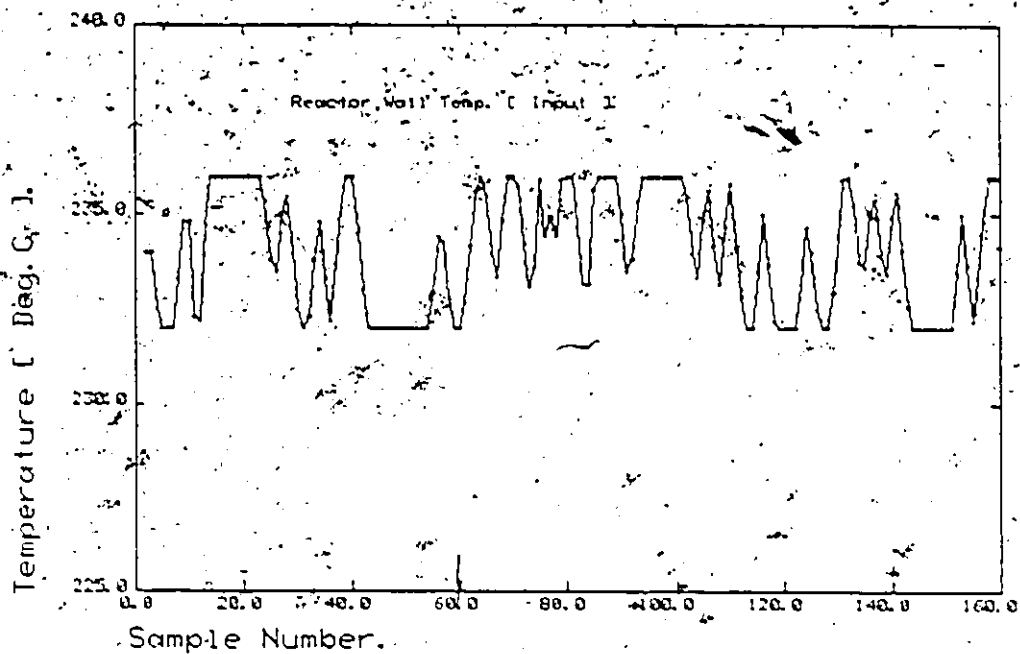


Fig. 5.5 Variation of Reactor Wall Temperature Over 8 Hours

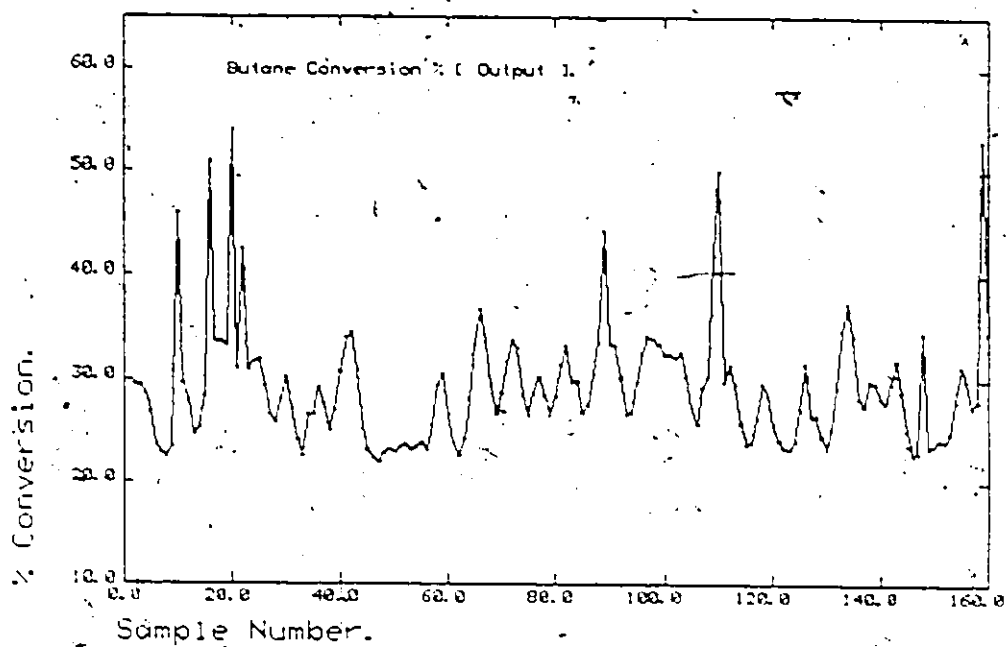


Fig. 5.6 Variation of Percentage Butane Conversion Over 8 Hours

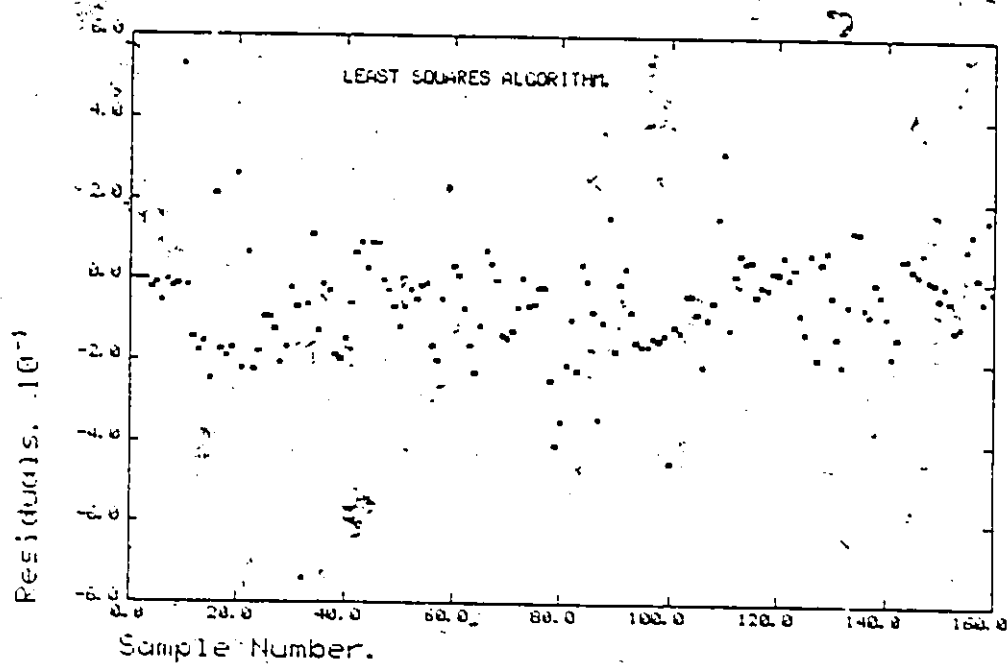


Fig. 5.7 Residuals of the Least Squares Model

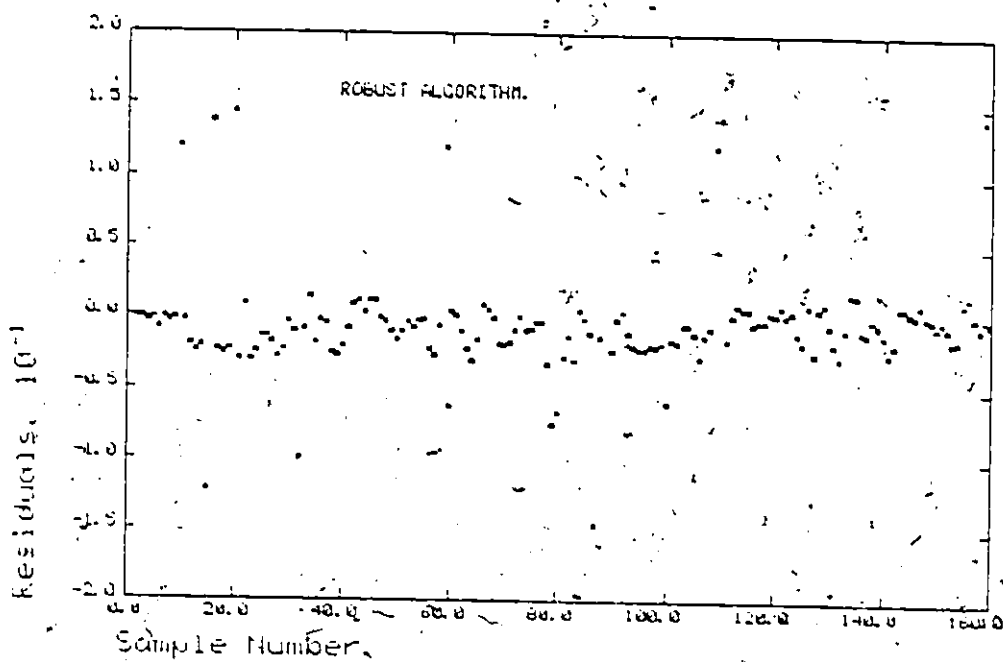


Fig. 5.8 Residuals of the Robust Model

CHAPTER 6

ROBUST BOOTSTRAP METHOD FOR THE JOINT ESTIMATION OF PARAMETERS AND STATES

6.1 Overview of the chapter

The problem of robust joint estimation of parameters and states of linear systems is considered. This is done by employing the algorithm introduced in the previous chapter along with special canonical structures of state space representation of systems.

6.2 The problem of joint parameters-states estimation

The necessity of joint parameters-states estimation became very important during the past few years because of its tremendous use in adaptive control. A typical application of this is well elucidated by Omani and Sinha (1985) in the context of adaptive pole placement using state variable feed back. The problem of joint parameters-states estimation is quite a challenging problem in the field of system identification because of its structure. That is to say, the parameter estimates will be correct if the estimates of states are correct, but the state estimates will be correct only if the the estimates of parameters are correct! This elusive nature of the problem gave rise to

considerable reaserch during the past few years.

The joint parameters-states estimation problem was originally posed as a nonlinear problem and extended Kalman filter was used to solve this problem (eg. Jazwinski, 1970). Later, a two stage algorithm was proposed in which the parameters and states are estimated in a bootstrap manner (Prasad and Sinha, 1977; El-Sherief and Sinha, 1979). But the convergence properties of the bootstrap algorithm were not well established in the presence of driving and significant observation noise and this led to several discussions (Sprevak and Newman, 1980; Ahmed, 1983). However, very recently, asymptotic convergence of the bootstrap algorithm has been assured (for a variety of cases of system noise) by incorporating modifications in the origininal algorithm (Goodwin and Sin, 1984; Omani and Sinha, 1985; Puthenpura and Sinha, 1985f).

6.3 The bootstrap method

6.3.1 Introduction

For simplicity, let us concentrate our attention to single input single output systems. It will be shown later that similar results hold good for multivariable systems.

Basically, the bootstrap algorithm can be divided into three steps as follows.

(i) Realization of the given system in special canonical form(s)

(ii) Estimation of parameters from samples of input.

output and from the nominal values of states

(iii) Use the above parameters to estimates and go to step (ii)

Thus steps (ii) and (iii) are coupled in a bootstrap manner.

The above three steps are explained in detail in subsequent discussions

6.3.2 Special canonical forms

For simplicity let us first consider a single input single output system represented by the ARMAX model

$$A'(z^{-1})y(t) = B'(z^{-1})u(t-f) + C'(z^{-1})e(t) \quad (6.3.1)$$

where z^{-1} represents the backward shift operator and

$$A'(z^{-1}) = 1 + a'_1 z^{-1} + a'_2 z^{-2} + \dots + a'_{n1} z^{-n1}$$

$$B'(z^{-1}) = b'_1 z^{-1} + b'_2 z^{-2} + \dots + b'_{n2} z^{-n2} \quad (6.3.2)$$

$$C'(z^{-1}) = 1 + c'_1 z^{-1} + c'_2 z^{-2} + \dots + c'_{n3} z^{-n3}$$

and f represents the integral periods of system delay. Also, $y(t)$ and $u(t)$ represent the output and input, respectively at the t^{th} instant of observation. The quantity $e(t)$ is the t^{th} element of a sequence of random numbers $\{e(t)\}$ such that

$$E[e(t)] = 0$$

$$\text{and } E[e(t)e(t)] = \sigma^2 \quad (6.3.3)$$

If we denote $n = \max(n1, n2+f, n3)$, (6.3.1) can be expressed as

$$A(z^{-1})y(t) = B(z^{-1})u(t-f) + C(z^{-1})e(t) \quad (6.3.4)$$

such that

$$A(z^{-1}) = 1 + a_1z^{-1} + a_2z^{-2} + \dots + a_nz^{-n}$$

$$B(z^{-1}) = b_1z^{-1} + b_2z^{-2} + \dots + b_nz^{-n} \quad (6.3.5)$$

$$C(z^{-1}) = 1 + c_1z^{-1} + c_2z^{-2} + \dots + c_nz^{-n}$$

It is possible to write (6.3.4) in the following state space form called the 'innovation model' as,

$$x(t+1) = F_0x(t) + G_0u(t) + K_0e(t)$$

$$y(t) = H_0x(t) + e(t) \quad (6.3.6)$$

where

F_0 - $n \times n$ matrix

G_0 , K_0 and $x(t) \in R^n$ (K_0 is often called the 'steady state Kalman gain vector')

$u(t)$, $y(t) \in R$ and $H_0^T \in R$.

The structure of (6.3.6) depends upon the choice of the canonical form. Here, we are particularly interested in two canonical forms.

1. The observer canonical form

In this particular realization,

$$F_0^1 = \begin{bmatrix} -a_1 & 1 & 0 & \dots & 0 \\ -a_2 & 0 & 1 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ -a_{n-1} & 0 & 0 & \dots & 1 \\ -a_n & 0 & 0 & \dots & 0 \end{bmatrix}$$

$$G_0^1 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix}, \quad H_0^1 = [1 \ 0 \ 0 \ \dots \ 0]$$

and

$$K_0^1 = \begin{bmatrix} c_1 - a_1 \\ c_2 - a_2 \\ \vdots \\ c_{n-1} - a_{n-1} \\ c_n - a_n \end{bmatrix} = \begin{bmatrix} k_1 \\ k_2 \\ \vdots \\ k_{n-1} \\ k_n \end{bmatrix}$$

(6.3.8)

2. The observability canonical form

Here,

$$F_0^2 = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \\ -a_n & -a_{n-1} & \dots & \dots & -a_1 \end{bmatrix}$$

$$G_0^2 = F_0^{-1} G_0^1 = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_{n-1} \\ B_n \end{bmatrix}, \quad H_0^2 = [1 \ 0 \ 0 \ \dots \ 0]$$

and

$$K_0^2 = T^{-1} K_0^1 = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_{n-1} \\ t_n \end{bmatrix} \quad (6.3.9)$$

where the transformation matrix T is the Toeplitz matrix given by

$$T = \begin{bmatrix} a_1 & 0 & \dots & 0 & 0 \\ a_2 & 1 & \dots & 0 & 0 \\ & a_1 & \dots & & \\ & & \dots & & \\ a_{n-1} & a_{n-2} & \dots & 1 & 0 \\ a_n & a_{n-1} & \dots & a_1 & 1 \end{bmatrix} \quad (6.3.10)$$

6.3.3 Formulation of bootstrap method as a pseudo linear regression (PLR)

The attractive feature of the above mentioned canonical forms is that that the parameters and states can be concatenated and put in an equation error form (section 4.3).

$$y(t) = \phi^T(t)\theta_0 + v(t) \quad (6.3.11)$$

where $v(t)$ is the measurement noise (coloured), $\phi(t)$ is the observation vector and θ_0 is the true parameter vector. It can be easily worked out that

$$\phi^T(t) = [z^T(t) \mid u(t-1), \dots, u(t-n) \mid e(t-1), \dots, e(t-n)]$$

where

$$z^T(t) = [x_1(t-1), x_1(t-2), \dots, x_1(t-n)]$$

and

$$\theta_o^T = [-a_1 \dots -a_n \mid b_1 \dots b_n \mid k_1 \dots k_n] \quad (6.3.12)$$

for the observer canonical form and

$$D^T(t) = [x^T(t-n) \mid u(t-1), \dots, u(t-n) \mid e(t-1), \dots, e(t-n)]$$

where

$$x^T(t-n) = [x_1(t-n), x_2(t-n), \dots, x_n(t-n)]$$

and

$$\theta_o^T = [-a_1 \dots -a_n \mid \beta_1 \dots \beta_n \mid \lambda_1 \dots \lambda_n] \quad (6.3.13)$$

for the observability canonical form. Note that $e(t)$ is given by eqn. (6.3.3).

Thus it is possible to estimate θ_o in a recursive extended least squares fashion. Let $\hat{\theta}(t)$ represents the estimate of θ_o at time t . Let us denote $F[\hat{\theta}(t)] = F$, $G[\hat{\theta}(t)] = G$, $H[\hat{\theta}(t)] = H$ and $K[\hat{\theta}(t)] = K$. Then using (6.3.6), the estimates of various states can be obtained as

$$x(t+1) = Fx(t) + Gu(t) + Ke(t) \quad (6.3.14)$$

where $e(t)$ is the residual error. The new estimates of states can be used in the observation vector and the whole procedure can be repeated several times in a bootstrap manner.

In practice, the $e(t)$ terms of $\phi(t)$ in eqns. (9.3.12) and (9.3.13) are replaced by their estimates, $\hat{e}(t)$. Moreover, the $x(t)$ terms in these equations are also

replaced by their estimates from (6.3.14). Thus the regression vector $\hat{\phi}(t)$ is not a true regression vector (which should contain the actual values of regressors), but a pseudo regression vector. Hence the bootstrap algorithm can be categorized as a pseudo linear regression (PLR) and the analysis can be carried out in that direction. The name PLR was first suggested by Solo (1979) and a general treatment of PLR's can be seen in this reference.

6.3.4 The robust bootstrap algorithm

The method described above is intuitively very appealing and works quite well when the data is devoid of any errors. Unfortunately, as pointed out in early chapters, this is not the case when one deals with real-life data. Therefore, we need robust procedures for the parameter estimation part of the bootstrap algorithm and the convergence of the whole method should be re-assessed.

The algorithm presented in the previous chapter can be applied very effectively for this purpose. [We particularly choose the estimation algorithm of the previous chapter, not the general M-estimation algorithm of chapter 4, because of the simplicity in proving the convergence. If we use the general M-estimation algorithm, the regression vector, the stochastic weight $\alpha(t)$ and the updating equation (6.3.14) will be coupled together via $\varepsilon(t)$'s and the analysis will become extremely complicated. Thus convergence proof of such an algorithm still remains as an open problem. Note that

even with a simplified estimation algorithm such as that of chapter 5, the convergence analysis is not very easy (appendix IV).]

Incorporating the modification using the approach of chapter 5, the robust bootstrap algorithm can be expressed as follows.

(a) Choose a suitable canonical form, observer or observability.

(b) Construct the observation vector

$$\Phi^T(t) = [\hat{S}^T(t) \mid U(t) \mid \hat{E}(t)] \quad (6.3.15a)$$

where

$$\begin{aligned} \hat{S}^T(t) &= [\hat{x}_1(t-1), \hat{x}_1(t-2) \dots \hat{x}_1(t-n)] \\ &= [\hat{y}(t-1), \hat{y}(t-2) \dots \hat{y}(t-n)] \end{aligned}$$

and

$$\theta_o^T = [-a_1 \dots -a_n \mid b_1 \dots b_n \mid k_1 \dots k_n] \quad (6.3.15b)$$

for the observer canonical form and

$$\hat{S}^T(t) = [\hat{x}_1(t-n), \hat{x}_2(t-n) \dots \hat{x}_n(t-n)]$$

and

$$\theta_o^T = [-a_1 \dots -a_n \mid \beta_1 \dots \beta_n \mid \ell_1 \dots \ell_n] \quad (6.3.15c)$$

Also,

$$U(t) = [u(t-1), u(t-2), \dots, u(t-n)] \quad (6.3.15d)$$

$$E(t) = [\varepsilon(t-1), \varepsilon(t-2), \dots, \varepsilon(t-n)] \quad (6.3.15e)$$

$$\text{Note that } \varepsilon(t) = y(t) - \hat{y}(t) = \hat{y}(t) - \Phi^T(t)\hat{\theta}(t) \quad (6.3.15f)$$

(c) Choose $\hat{\theta}(0) = \underline{0}$ and $P(0) = kI$, $k \gg 0$,

(d) Use algorithm (5.3.9) or (5.3.11) to obtain

parameter estimate $\hat{\theta}$

- (e) Update the estimates of states using (6.3.14)
- (f) $t \rightarrow t+1$ and go back to (b).

6.4 Convergence and efficiency of the robust bootstrap algorithm

Since the bootstrap algorithm is a typical PLR algorithm, the convergence analysis can be carried out exactly in the same way that of the extended least squares algorithm, which is the most celebrated example of a PLR (Ljung and Soderstrom, 1983).

The following theorem gives the convergence conditions of the robust bootstrap algorithm.

Theorem 6.1 (Puthenpura and Sinha, 1985f)

Let the following assumptions hold good.

(i) The correct model order has been selected.

$$\begin{aligned} \text{(ii)} \quad E[e^2(t)|F(t-1)] &= \sigma_1^2(1-u) + \sigma_2^2u \\ &= E[\sigma_1^2[1-\delta(t)] + \sigma_2^2\delta(t)|F(t-1)] \end{aligned}$$

so that $\text{prob}[\delta(t)=1|F(t-1)] = u$

$e(t)$ being the observation noise, $F(t-1)$ is the sigma algebra of observations upto time $t-1$.

Then the parameter estimate sequence $\{\hat{\theta}(t)\}$ provided by the robust bootstrap algorithm will converge to the true parameter (θ_0) with probability 1 only if

$$(a) \quad 0 < \sigma_1^2, \sigma_2^2 < \infty$$

$$(b) E[\phi^T(t)\phi(t)] = 0.$$

$$(c) \limsup_{t \rightarrow \infty} [1/t] \sum_{n=1}^t |\phi(n)|^2 < \infty$$

and if

$$(d) \text{Real} [H(z^{-1}) - (1/2)I] > 0$$

where

$$H(z^{-1}) = [1/C(z^{-1})]D(z^{-1})$$

where $D(z^{-1}) = I$ for the observer canonical form

and

$$D(z^{-1}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & D_1(z^{-1}) & 0 \\ 0 & 0 & D_1(z^{-1}) \end{bmatrix}$$

$$D_1(z^{-1}) = \begin{bmatrix} \lambda_1(z^{-1}) & & & 0 \\ 0 & \lambda_1(z^{-1}) & & 0 \\ & \ddots & \ddots & \vdots \\ & & & \lambda_{n-1}(z^{-1}) \\ 0 & & & & 1 \end{bmatrix}$$

such that

$$\lambda_i(z^{-1}) = \sum_{j=1}^i a_j z^{-j} \quad \text{for the observability canonical form.}$$

Note that $C(z^{-1})$ and a_j 's are described by (6.3.5).

The proof of this theorem is long and technical, and

it is given in appendix IV.

Important note:

It can be verified that when we use the observability canonical form, if we replace the observation vector $\phi(t)$ by a filtered observation vector $\phi_f(t)$ where

$$\phi_f(t) = [\hat{x}(t-n) | \hat{\lambda}_1(z^{-1})u(t-1) \dots \hat{\lambda}_1(z^{-1})u(t-i) \dots u(t-n) | \hat{\lambda}_1(z^{-1})\varepsilon(t-1) \dots \hat{\lambda}_1(z^{-1})\varepsilon(t-i) \dots s(t-n)]$$

where $\hat{\lambda}_1$ is the estimate of λ_1 , then the matrix $D(z^{-1})$ reduces to I for the observability canonical form too. Under these circumstances, the sufficient condition (d) of theorem 6.1 is one and the same for both the canonical form.

The efficiency of the robust bootstrap method can be evaluated exactly in the same way as explained in section 5.4 (see proposition 5.1).

Thus it can be said that in general the robust bootstrap method is less efficient when the data enjoys normality.

6.5 Extension to Multivariable Systems

Here, instead of (6.3.6) we get (El-Sherief, 1979)

$$\begin{aligned} x(t+1) &= F_0^* x(t) + G_0^* u(t) + K_0^* e(t) \\ y(t) &= H_0^* x(t) + e(t) \end{aligned} \quad (6.5.1)$$

where

$$x(t) \in R^n, \quad u(t) \in R^p, \quad y(t) \in R^m \text{ and } e(t) \in R^m.$$

Then partition F_0^* in the form

$$F_0^* = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1n} \\ F_{21} & F_{22} & \dots & F_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ F_{n1} & F_{n2} & \dots & F_{nn} \end{bmatrix} \quad (6.5.2)$$

where the elements have the form

$$F_{ii} = \begin{bmatrix} 0 & & & \\ & & I_{n_i-1} & \\ & & & \\ a_{ii}(1) & a_{ii}(2) & \dots & a_{ii}(n_i) \end{bmatrix} \quad (6.5.3a)$$

$$F_{ij} = \begin{bmatrix} 0 & & & 0 \\ & & & \\ & & & \\ a_{ij}(1) & a_{ij}(2) & \dots & a_{ij}(n_j) \end{bmatrix} \quad (6.5.3b)$$

for the observability form and

$$F_{ii} = \begin{bmatrix} a_{ii}(1) & & & \\ a_{ii}(2) & & & \\ & & I_{n_i-1} & \\ a_{ii}(n_i) & 0 & \dots & 0 \end{bmatrix} \quad (6.5.4a)$$

$$F_{ij} = \begin{bmatrix} a_{ij}(1) & & & 0 \\ a_{ij}(2) & & & \\ & & & \\ a_{ij}(n_j) & & & 0 \end{bmatrix} \quad (6.5.4b)$$

for the observer canonical form. Note that G_0^* and K_0^* will also be changed accordingly. Rest of the procedure is exactly similar to what was presented in section 6.3.4 except for the difference that we have to consider each subsystem $[F_{ij}; i, j = 1, 2, \dots, n]$ separately. That is, we will be estimating $\hat{\theta}_j(t)$ and $P_j(t)$ using algorithms 5.3.9 or 5.3.11 for $j=1, 2, \dots, n$ recursively.

6.6 Simulation Results

To illustrate the robust bootstrap method the following example is provided. For conciseness only one canonical form is considered, the other one follows exactly in the same way.

Consider the ARMAX model of a system

$$[1 - 1.5z^{-1} + 0.7z^{-2}]y(t) = [1 + 0.5z^{-1}]u(t) + [1 - 0.3z^{-1}]e(t) \quad (6.6.1a)$$

which can be put in the observability canonical form

$$\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -0.7 & 1.5 \end{bmatrix} \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + \begin{bmatrix} 1.0 \\ 2.0 \end{bmatrix} u(t) + \begin{bmatrix} 1.2 \\ 1.1 \end{bmatrix} e(t)$$

$$y(t) = [1 \quad 0] \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} + e(t) \quad (6.6.1b)$$

which gives

$$\phi^T(t) = [x_1(t-2) \quad x_2(t-2) \quad | \quad u(t-1) \quad u(t-2) \quad | \quad \varepsilon(t-1) \quad \varepsilon(t-2)]$$

and

$$\theta_0 = [-0.7 \quad 1.5 \quad | \quad 1.0 \quad 2.0 \quad | \quad 1.2 \quad 1.1] \quad (6.6.1c)$$

The noisy output containing outliers are generated exactly in the way as described in section 5.5.1.

Table 6.1 shows some results when the SNR = 10 dB, probability of occurring outliers is 0.1 with variance 50. The robust bootstrap method (6.3.15) is compared with the non robust version [ie., by setting $\{1-\delta(t)\}\sigma_1^2 + \delta(t)\sigma_2^2 = 1$ V t] under these circumstances. The starting values of the variances are taken to be 1 and 30. To make the comparison complete, the robust and non-robust bootstrap algorithms are compared when no outliers are present in the data. Also refer fig.6.1 through fig.6.8, which are self explanatory. In particular, fig.6.7 and fig.6.8 clearly indicate the failure of the non-robust bootstrap method in presence of outliers in the data. Fig.6.1 and fig.6.2 illustrate the powerfulness of the robust bootstrap algorithm in tackling outliers. But notice the slower convergence rate of the robust bootstrap algorithm (fig.6.1 and 6.3). Hence this example clearly illustrates the various theoretical aspects discussed in the previous sections.

6.7 Concluding remarks

The robust bootstrap method described and analyzed in previous sections is an immediate extension of the discussions presented in chapter 5. The approach used in the convergence analysis is a fairly general and unified treatment, so that it can be applied to a variety of

situations. In fact, the results obtained by Goodwin and Sin (1984) and Omani and Sinha (1985) are only special cases of the results obtained by this approach, where the analyses are carried out using different techniques.

The choice of canonical form (between the two discussed) depends upon the context of application of the bootstrap algorithm. For example, in adaptive pole placement techniques (eg. Warwick, 1981) the robust bootstrap method with the observability canonical form is preferred from the fact that the observability matrix in that case reduces to an identity matrix.

The poor convergence rate of the robust bootstrap method may be a criticism against it, at this time. However, the author is quite optimistic that some means may be found in the near future to accelerate its convergence so that the only draw back of this simple and attractive tool for joint parameters-states estimation is eliminated.

Table 6.1.
Comparison of robust and non robust bootstrap method
at SNR = 10dB. Number of iterations = 500.

	<u>Estimated Parameters</u>		<u>Actual parameters</u>
	<u>Non-robust bootstrap</u>	<u>Robust bootstrap</u>	
$u = 0.1$ $V = 50$	failed to converge	-0.722258 1.522595 1.060739 2.081306 1.254283 1.187507	-0.7 1.5 1.0 2.0
$u = 0$	-0.716047 1.516693 1.054972 2.073849 1.243374 1.169769	-0.717760 1.518319 1.066371 2.080849 1.249132 1.178820	1.2 1.1

Note : u means the probability of occurring outliers and
 V means the the variance of outliers.

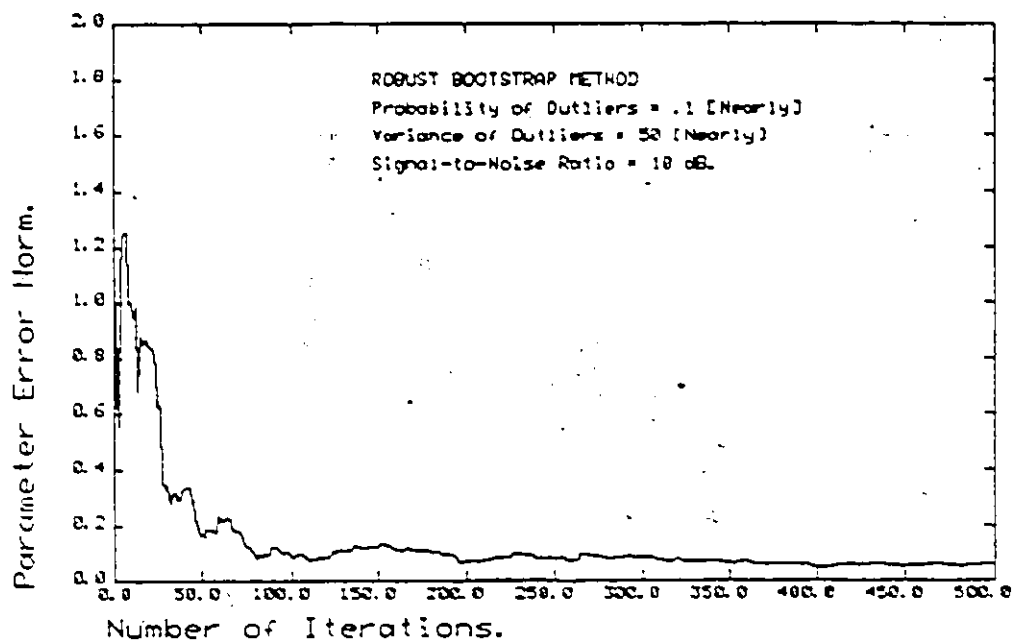


Fig. 6.1 Parameter Error Norm: Robust Bootstrap (in Presence of Outliers)

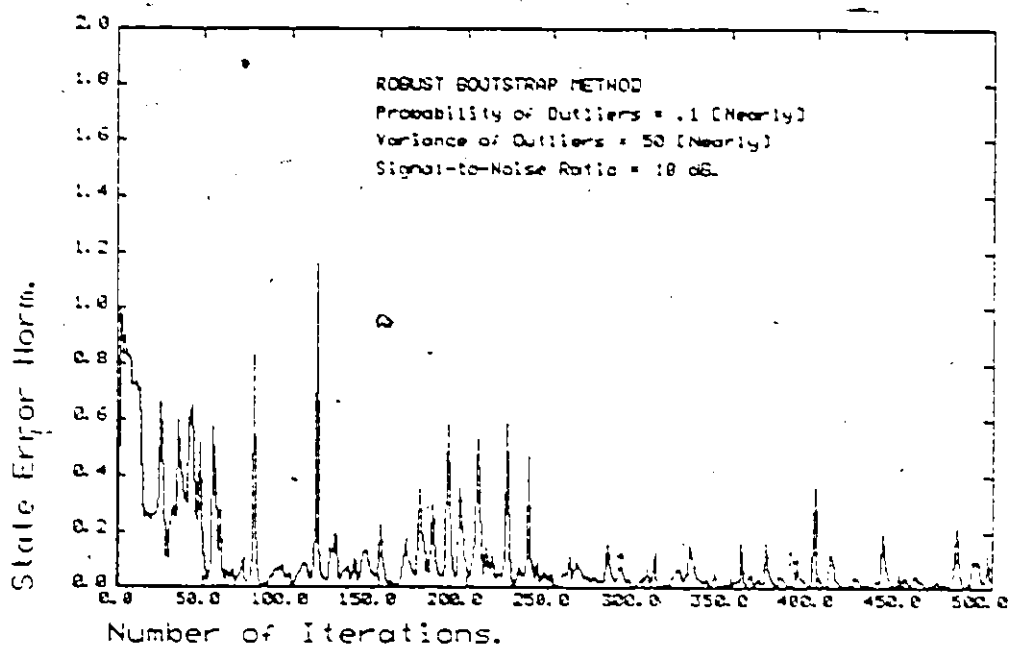


Fig. 6.2 State Error Norm: Robust Bootstrap (in Presence of Outliers)

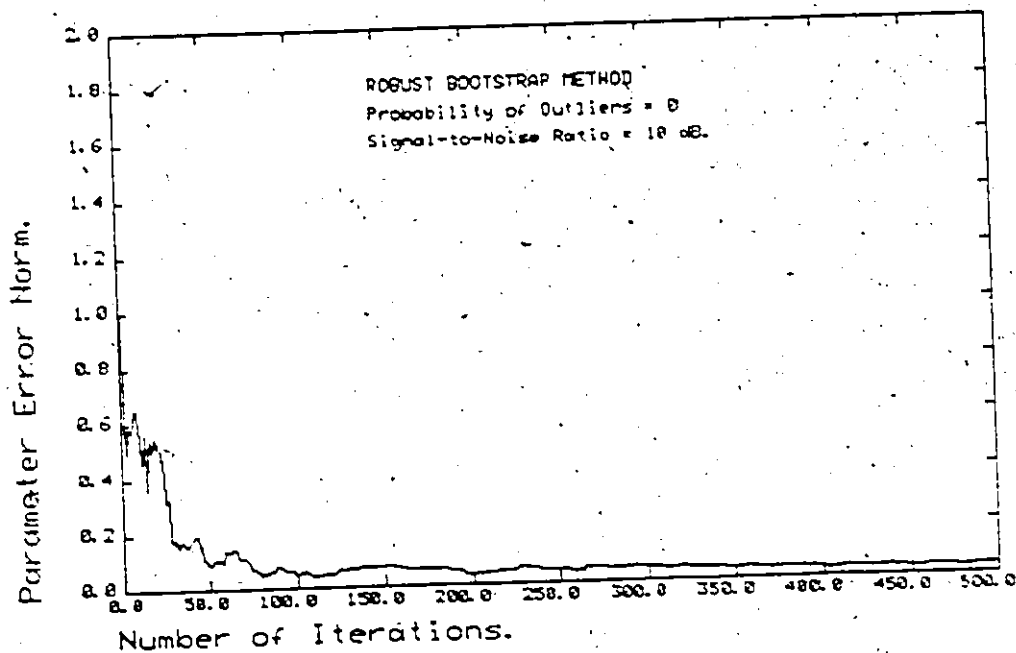


Fig. 6.3 Parameter Error Norm: Robust Bootstrap (in Presence of no Outliers)

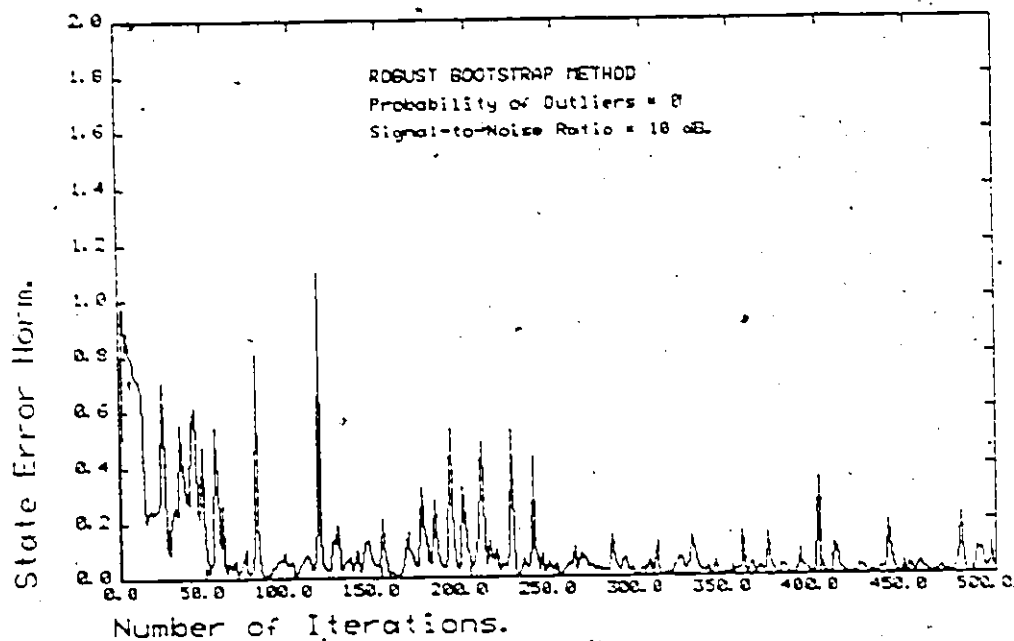


Fig. 6.4 State Error Norm: Robust Bootstrap (in Presence of no Outliers)

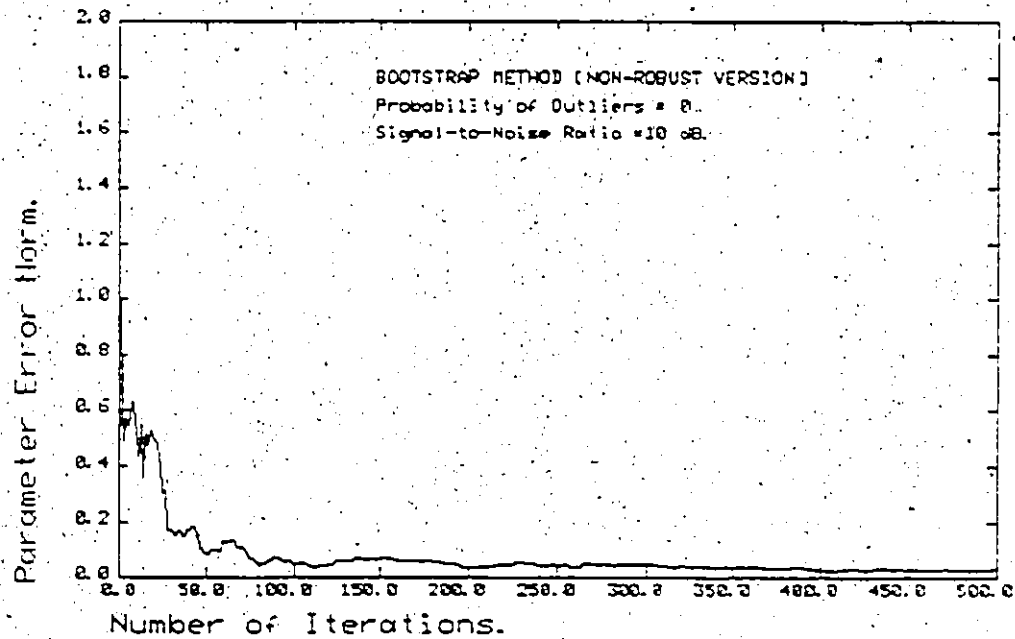


Fig. 6.5 Parameter Error Norm: Non Robust Bootstrap (in Presence of no Outliers)

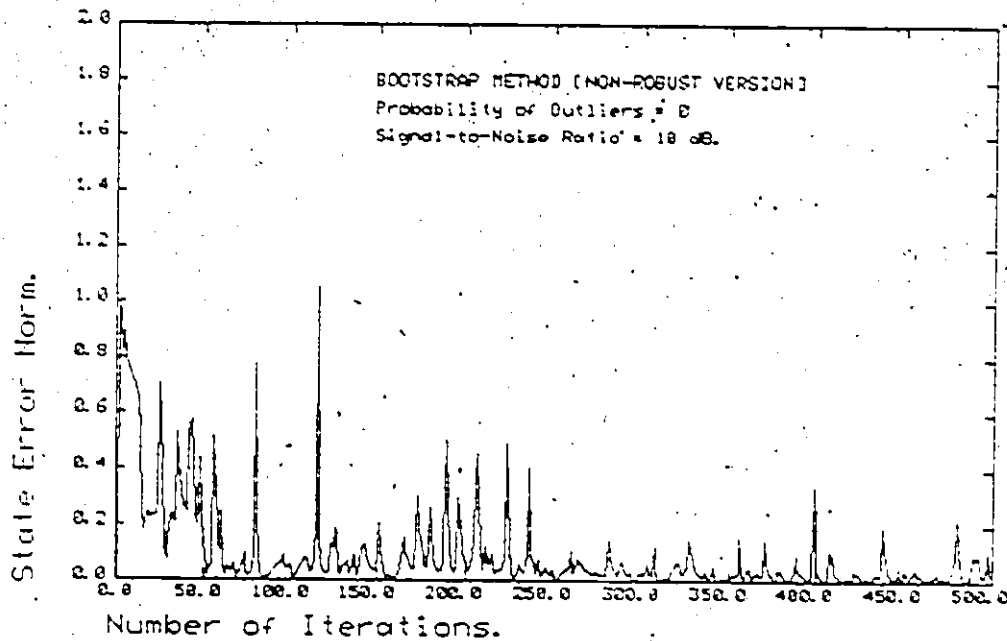


Fig. 6.6 State Error Norm: Non Robust Bootstrap (in Presence of no Outliers)

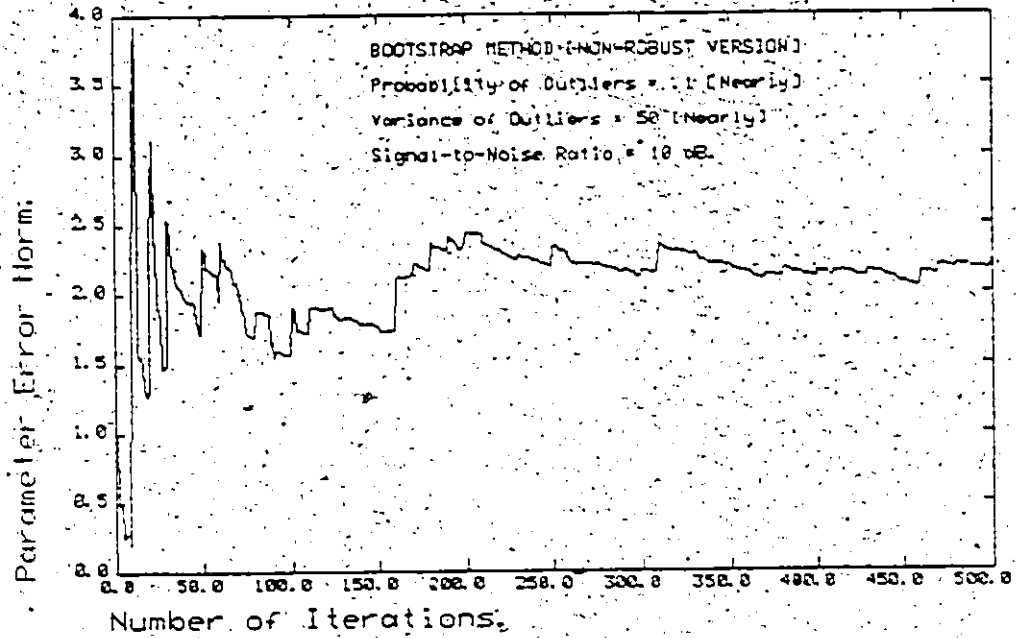


Fig. 6.7 Parameter Error Norm: Non Robust Bootstrap (in Presence of Outliers)

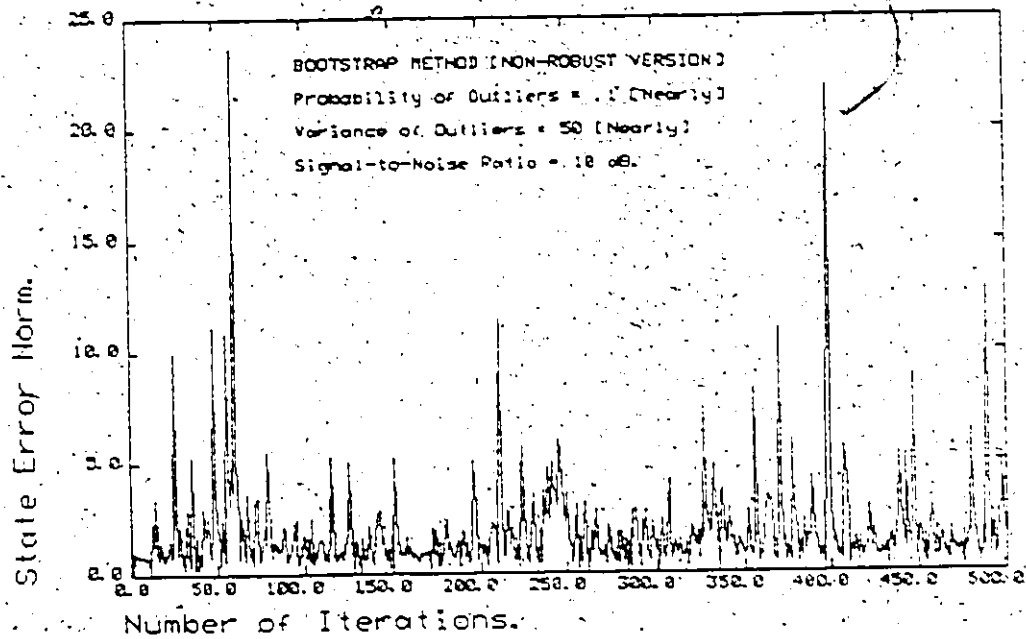


Fig. 6.8 State Error Norm; Non Robust Bootstrap (in Presence of Outliers)

CHAPTER 7

NUMERICAL ROBUSTNESS ISSUES IN THE IDENTIFICATION OF CONTINUOUS-TIME SYSTEMS

7.1 Overview of the chapter

This chapter is about continuous-time systems. The attention has been focussed on two aspects; (i) the choice of sampling interval so that the identification schemes are less sensitive to truncation and round off errors in computers, due to their finite word length and (ii) numerically robust algorithms for the transformation of discrete-time models to their continuous-time equivalents.

7.2 Identification of continuous-time models

Sometimes it is required to obtain a continuous-time model of the system under consideration for control applications. The continuous-time model of the system can be obtained basically by two methods. The first one is the so called "indirect method" which can be subdivided further into two parts. (1). Obtaining the discrete-time model of the system from the samples of input-output data and (2). the determination of a continuous-time model corresponding to the estimated discrete-time model (Sinha and Kuszta, 1983). The other approach is the "direct method", which is based on obtaining piecewise constant solutions of linear

differential equations in between sampling intervals. This can be done by using state variable filters (Young, 1984) or by interpolating the data in between sampling intervals (Sinha and Zhou Qi-Jie, 1984). The notable interpolation methods are block pulse, trapezoidal pulse and cubical splines. The main advantage of the indirect method is the availability of considerable literature and relatively less computation compared to the direct method. A careful analysis of the direct and indirect methods can be found in Sinha and Lastman (1982) by which the above mentioned points can be substantiated. It is also noteworthy that the main disadvantage of the indirect method is that the input is not always maintained constant in between sampling interval. These aspects are explained quantitatively below.

Consider the discrete-time model described by eqn. (4.3.1). An equivalent continuous-time model can be written as

$$H(s) = \frac{p_r s^r + p_{r-1} s^{r-1} + \dots + p_1 s^1 + p_0}{s^n + q_{n-1} s^{n-1} + \dots + q_1 s^1 + q_0} \quad (7.2.1)$$

Alternatively, let the corresponding state-space forms are

$$\dot{x}(kT+T) = Fx(kT) + Gu(kT)$$

$$y(kT) = Cx(kT), \quad T \text{ being the sampling interval}$$

(7.2.2a)

and

$$\dot{x}(t) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t) \quad (7.2.2b)$$

respectively.

The identification procedure is to estimate the parameters of the discrete-time or the continuous-time transfer function model from the samples of input output data. After this, the state space models can be easily obtained with the help of canonical forms.

In the indirect method of identification of continuous-time systems, first we estimate the parameters of the discrete-time model by any standard method of identification and utilize the following relationships to get (7.2.2b) and hence (7.2.1).

$$F = \exp(AT) \quad \text{and}$$

$$G = \int_0^T \exp(At)Bdt \quad (7.2.3)$$

As mentioned earlier, (7.2.2a) is not very accurate if the input $u(t)$ is not held constant in between sampling intervals. It may be noted that the exact solution of the state equation during the interval $kT < t \leq (k+1)T$ is given by

$$x(kT+T) = [\exp(AT)]x(kT) + \int_{kT}^{kT+T} \exp[A(kT+T-t)]Bu(t)dt \quad (7.2.4)$$

Equation (7.2.4) is a special case when

$$u(t) = u(kT) \quad \text{for } kT < t \leq kT+T \quad (7.2.5)$$

The actual value of the the integral in eqn. (7.2.4)

depends upon the nature of variation of $u(t)$ in between sampling instants. But we know the value of $u(t)$ only at the sampling instants. Therefore, an intuitively more appealing approximation is to use the average value of u from the sampling instant kT to $kT+T$. Hence, eqn. (7.2.2a) is replaced by

$$x(kT+T) = Fx(kT) + (1/2)G[u(kT) + u(kT+T)] \quad (7.2.6)$$

in cases where $u(t)$ is not piecewise constant.

However, in simulation studies presented in this chapter, the state space form is not directly used. Instead, the equation error approach is used to estimate the parameters of the discrete or continuous-time transfer function model of the system. If the indirect method is followed, first the discrete-time transfer function model of the the system is estimated, which can be put in the form (7.2.6). Later, an appropriate transformation (see section 7.5) to obtain the continuous-time equivalent model, given by (7.2.2b). In the direct approach, one attempts to estimate the parameters of the continuous-time transfer function model (7.2.1), directly using methods described before.

It is very important to note that all these procedures of identification are based on the fact that the sampling interval has been selected "properly", in addition to the assumption that the system is completely observable and the input is of persistantly exciting. A rule of thumb, which is commonly used, is that the sampling interval (T) is

chosen in such a way that (Haykin, 1972)

$$\lambda_m T \leq 0.5 \quad (7.2.7)$$

where λ_m is the magnitude of the magnitude of the largest eigenvalue of the continuous-time model. In practice, however, one does not know the value of λ_m a priori. All that can be done is to make an intelligent guess based on the expected time constants of the system under consideration. It may appear that this problem can be easily solved by making the sampling interval too small. This is not always true and may cause two other severe problems. First of all, we may have too many samples as a result the computational burden is increased. This is obviously undesirable. Secondly and most importantly, the identification problem gets ill conditioned; as will be shown in the next section with the help of examples; both in direct and indirect methods.

7.3 Effects of choosing very small sampling intervals

7.3.1 Introduction

As mentioned earlier, the choice of very small sampling interval makes the identification problem ill conditioned which may lead to severe computational problems. This can be explained as follows.

If λ is an eigenvalue of A , then the corresponding eigenvalue of F will be $\exp(\lambda T)$, T being the sampling interval. If we denote

$$\lambda T = c \pm j\beta. \quad (7.3.1)$$

Then, T should be chosen such that

$$|\alpha \pm j\beta| \leq 0.5 \quad (7.3.2)$$

to satisfy condition (7.2.7). Therefore,

$$\alpha^2 + \beta^2 \leq 0.25 \quad (7.3.3)$$

The eigenvalue of F corresponding to $\alpha \pm j\beta$ is

$$\exp(\alpha \pm j\beta) = \exp(\alpha)[\cos\beta \pm j\sin\beta] \quad (7.3.4)$$

For stable poles in the z -plane and still satisfying (7.3.3), the values of α and β should be such that

$$-0.5 \leq \alpha \leq 0$$

$$\text{and } -0.5 \leq \beta \leq 0.5 \quad (7.3.5)$$

Hence, the region in the z -plane subjected to the above condition will be as shown in fig.7.1. Thus it is obvious that when T becomes smaller and smaller the eigenvalues will be constrained into narrower regions in the z -plane, very close to the boundary of the unit circle, even though the corresponding continuous-time model may have its eigenvalues scattered widely over the left half of the s -plane. This produces a large sensitivity to numerical errors caused by truncation and round-off in computation due to the finite word length of computers used for identification. Moreover, eigenvalues close to the boundary of the unit circle will lead to considerable numerical instability in identification. This problem will be serious if the measurements are contaminated with noise, as usual in practical situations.

7.3.2 Illustrations of the effect of very small sampling intervals

Here an example is presented to ill conditioning of the identification problem due to improper selection of sampling intervals.

Consider the system

$$H(s) = \frac{2s + 1}{s^2 + 5s + 6} \quad (7.3.6)$$

The above system is perturbed with an input

$$u(t) = \cos(0.1t) - \cos(1.5t) \quad (7.3.7)$$

and the corresponding output is obtained as

$$y(t) = 0.536\exp(-2t) - 0.332\exp(-3t) + 0.169\cos(0.1t) - 0.019\sin(1.5t) - 0.373\cos(1.5t) + 0.053\sin(1.5t) \quad (7.3.8)$$

The above input and output are sampled at an interval of 0.001 seconds and 15000 samples are taken. To produce coloured noise, random numbers (of zero mean Gaussian) are passed through a moving average filter of the form

$$C(z^{-1}) = 1 - 0.3z^{-1} \quad (7.3.8)$$

and the variance of the input random numbers are adjusted to obtain different signal-to-noise ratios. Noise produced in this manner is added to the entire 15000 samples of data. This whole set of data can be used for identifying system models for different sampling intervals. For example, one out of ten data points is taken corresponding to a sampling interval (T) of 0.01 seconds, one out of 100 is taken for T = 0.1 and so on. Note that in each case the total length of

data (hence the total information content) is the same, data collected over 15 seconds. This makes the comparison of different sampling intervals fair. Now we can proceed as follows.

Indirect method

Here first we identify the discrete-time model of the system in the form (Box and Jenkins, 1971)

$$y(t) = \frac{B(z^{-1})}{A(z^{-1})} u(t) + \frac{C(z^{-1})}{D(z^{-1})} e(t) \quad (7.3.8)$$

for different sampling intervals. The identification method used is the off-line maximum likelihood. It is important to note that the noise model is not same for all the sampling intervals. It varies for different sampling intervals along with the system transfer function.

If we assume a zero order hold which keeps the input constant between sampling intervals, we get a step invariant z-domain transfer function model of (7.3.6) as

$$H(z^{-1}) = \frac{a_1 z^{-1} + a_2 z^{-2}}{1 + b_1 z^{-1} + b_2 z^{-2}}$$

where $a_1 = (1/6)[1 + 9\exp(-2T) - 10\exp(-3T)]$

$a_2 = (1/6)[-10\exp(-2T) + 9\exp(-3T) + \exp(-5T)]$

and $b_1 = -[\exp(-3T) + \exp(-2T)]$

$b_2 = \exp(-5T)$ (7.3.9)

T being the sampling interval.

The noise model is quite difficult to be calculated analytically for each sampling interval and its parameters are implicitly sampling time dependent. The only information about the noise model is that it reduces to that given by (7.3.9) at $T=0.001$ seconds. As a result, we obtain different noise models for different sampling intervals.

Table 7.1 shows some of the results obtained. The parameters of the noise model is not provided as they cannot be compared to the actual parameters as they are not available.

It is noteworthy that the identification scheme failed due to severe numerical problems at sampling intervals smaller than 0.01 seconds.

These results clearly indicate the ill conditioning of the identification problem due to improper selection of sampling intervals. It is also interesting to note that best numerical results are obtained when T is such that $\lambda_m T$ is around 0.5.

Fig.7.2 shows the parameter error norm for different sampling intervals.

The transformation from the discrete-time domain to continuous-time domain is omitted here for simplicity. It is quite obvious that if the discrete-time model is not accurate, the corresponding continuous-time model will also be inaccurate.

Direct method

Consider the same system described by (7.3.6), the same input output data has been used here also. But here we use the the recurrent trapezoidal rule of integration along with the instrumental variables method (eg. Prasad and Sinha, 1983; Puthenpura and Sinha, 1985c) to identify the continuous-time model directly. Note that this method avoids the use of conventional state variable filters or the noisy derivative evaluation (eg. using modulating functions).

Just like the direct method, too small or too large sampling intervals reduce the accuracy of identification as illustrated by table 7.2 and fig.7.3. As quite well known, the direct method will be not accurate at poor signal-to-noise ratios. Hence unlike the indirect method it is not very meaningful to go below a SNR of 20dB.

7.4 Procedures to obtain proper values of sampling intervals

7.4.1 Introduction

In spite of the vitality of the problem, it is surprising to note that the amount of research that has been put into it is negligible. The only reported work in this topic are due to Sinha et.al. (1982), Puthenpura and Sinha (1985g) and Sinha and Puthenpura (1985). (But as per the comment by one of the reviewers of the last reference, prof. K.J. Astrom of Lund Institute of Technology, Sweden was also involved with the problem for sometime).

The procedures which will be described in subsequent

discussions, are not fully based on theoretical foundations. On the other hand, they are derived on a number of simulation studies, hence somewhat heuristic. These methods have their own drawbacks, but seem to be adequate to tackle most of the practical situations.

7.4.2 The w-plane method

This method can be considered as one of the very first attempts to tackle the sampling time problem and it is due to Sinha and Puthenpura (1985).

One of the obvious solutions for obtaining the proper sampling interval(T) is to adjust the value of T so that the eigenvalues of the discrete-time system fall closer to the boundary ABC of fig.7.1 than the boundary of the unit circle.

Since the region in which the eigenvalues of the discrete-time model of the system are located is very small (fig.7.1), any direct adjustment of the position of eigenvalues in the z-plane can produce serious numerical problems, especially when the measurements are noisy (Sinha et. al., 1982).

By introducing a bilinear transformation given by

$$w = (z+1)/(z-1) \quad (7.4.1)$$

the area inside the unit circle can be spread out to the entire left half of the w-plane.

If $z = \exp(c)[\cos\beta \pm j\sin\beta]$, by applying (7.4.1) the corresponding point in the w-plane is given by

$$w = w_1 + jw_2 \quad \text{where}$$

$$w_1 = [\exp(2\alpha) - 1] / [\exp(2\alpha) - 2\exp(\alpha)\cos\beta + 1] \quad \text{and}$$

$$w_2 = [2\exp(2\alpha)\sin\beta] / [\exp(2\alpha) - 2\exp(\alpha)\cos\beta + 1] \quad (7.4.2)$$

Hence the region in the z -plane shown in fig.7.1 will be mapped into the region shown in fig.7.4 in the w -plane. Also it can be seen that the boundary ABC in the z -plane is mapped as \overline{ABC} in the w -plane.

To make the identification procedure accurate, poles near to the boundary of the unit circle should be pulled inside but within the area shown in fig.7.1, by adjusting the sampling interval (T). We focus our attention on poles only, since the success of identification procedure as well as the transformation from discrete-time to continuous-time domain depends on the poles of the system. On the transformation from the discrete to continuous-time domain, we map the poles first and evaluate the residues at poles (e.g. Franklin and Powell, 1980). Hence the poles play an important part of our problem, not the zeros.

As mentioned before, it can be easily seen that the small region in the z -plane is expanded to a larger region in the w -plane. As a result, placing the poles "properly" in the w -plane gives better numerical stability. Hence a criterion for choosing the sampling interval can be developed as follows.

Our aim is to place the poles near to the boundary ABC in the shaded region of fig.7.1. Equivalently, the poles are to be placed near to the boundary \overline{ABC} of fig.7.4

within the shaded area. The point \bar{E} is at a distance of

$$[\exp(-0.5) + 1]/[\exp(-0.5) - 1] = 4.0830 \quad (7.4.3)$$

from the origin. Hence a criterion can be proposed to place the poles at a distance R at a distance of 5.0000 from the origin in the w -plane. We choose this value of R instead of 4.0830 exactly, because it is desirable to have a "tolerance" in numerical algorithms. This was proved to be quite effective from a large number of numerical examples tried by the author.

The value of R , the distance from the origin to the pole $w_1 = jw_2$ is given by

$$\begin{aligned} R &= [w_1^2 + w_2^2]^{1/2} \\ &= \frac{[\rho^4 - 2\rho^2 \cos(2\phi) + 1]^{1/2}}{[\rho^2 - 2\rho \cos(2\phi) + 1]} \end{aligned} \quad (7.4.4)$$

where $\rho \exp(\mp j\phi)$ is the corresponding pole in the z -plane. The attempt is to find a correction factor K , such that $T_{\text{new}} = T_{\text{old}} \times K$ so that the value of R corresponding to the smallest magnitude in the z -domain is about 5.0 units; that is, placing the equivalent w -plane pole within a semi-circle of radius 5.0 units.

Thus we can find the value of K as the solution of the equation,

$$\frac{[\rho_1^{4K} - \rho_1^{2K} \cos(2K\phi_1) + 1]^{1/2}}{[\rho_1^{2K} - \rho_1^K \cos(K\phi_1) + 1]} = 5.0 \quad (7.4.5)$$

$\rho_1 \exp(\mp j\phi_1)$ is the pole of smallest magnitude in the z-domain.

Note: Equation 7.4.5 is nonlinear, but can be solved by using one of the common numerical techniques, as for example, the Newton-Raphson method. The starting value of K can be taken as 1.

The whole procedure can be carried out as follows.

(i) Assume a value of T on the basis of system time constants. It could be about one fourth of the smallest time constant.

(ii) For this value of T, estimate the parameters of the discrete-time model of the system and then the eigenvalues of F.

(iii) Make the transformation, $w = (z+1)/(z-1)$; where z is the the eigenvalue of F which is of smallest magnitude.

(iv) If the magnitude of w is around 5.0, then the choice of T is good. Otherwise change T to KT, K being calculated by eqn. (7.4.5).

(v) Go back to step (ii).

After obtaining the proper sampling interval one can use the direct or indirect method for identifying the continuous-time model of the system. The example given below illustrates the algorithm.

Example

Consider the same system as before described by (7.3.6). To begin with T is chosen as 0.1 seconds. For this value, the eigenvalues of F are found to be 0.7381 and 0.7389 (refer table 7.1). The value of R in the w -plane corresponding to the pole of smallest magnitude in the z -plane (i.e., 0.7381) is 6.64 which indicates that T is too small. Hence the proposed method is applied to get the correction factor K which is found to be 1.4. Therefore, the corrected sampling interval is 0.14 seconds.

At this sampling interval the eigenvalues of the discrete-time model are found to be 0.6670 and 0.6913. The value of R corresponding to the smallest eigenvalue (i.e., 0.6670) is about 5.0. Therefore, further corrections on T is not necessary. Hence the computationally optimum value of the sampling interval is accepted as 0.14 seconds.

7.4.3 Critical evaluation of the w -plane method

The method described above is simple and works satisfactorily under normal circumstances. However, the method fails in the case of very stiff systems and does not permit the user to start with an arbitrary small sampling interval. Putting it in equivalent terms, some a priori knowledge about system time constants is absolutely necessary. Note that, if one starts off with an arbitrary small sampling interval, the discrete-time model itself is not accurate enough to apply the proposed method which is

evident from table 7.1. Therefore one has to explore other means to obtain an improved performance in these cases. The method explained in the next section is a possible answer to this problem.

7.4.4 The p-plane method

Here we introduce a transformation from the z-plane to the so called p-plane which is given by

$$p = 1/[1-\epsilon z] \tag{7.4.6}$$

where ϵ is defined as the "contraction factor", lies between 0 and 1 (Puthenpura and Sinha, 1985g). The mapping of the lense shaped region of the z-plane (fig.7.1) to the plane is shown in fig.7.5(a) and fig.7.5(b).

Hence, instead of identifying the parameters of the discrete-time model the parameters of the equivalent p-plane model which is given by

$$H(p) = \frac{A_m p^m + A_{m-1} p^{m-1} + \dots + A_0}{B_n p^n + A_{n-1} p^{n-1} + \dots + B_1 p + 1} \tag{7.4.7}$$

are estimated. In the p-plane instead of the unit delay operator (z^{-1}), the p-operator $1/[1-\epsilon z]$ will operate on $y(t)$ and $u(t)$ where t represents the discrete-time index. Therefore the observation vector in the p-plane can be written as,

$$\phi_1(t) = [-y_1(t), -y_1(t), \dots, -y_n(t), u(t), u_1(t), \dots, u_m(t)] \tag{7.4.8}$$

where

$$y_r(t) = y(t) + [r/1!]\epsilon y(t+1) + [r(r+1)/2!]\epsilon^2 y(t+2) + \dots \quad (7.4.9)$$

and

$$u_r(t) = u(t) + [r/1!]\epsilon u(t+1) + [r(r+1)/2!]\epsilon^2 u(t+2) + \dots \quad (7.4.10)$$

$$\text{so that } p^r y(p) = [1-\epsilon z]^{-1} y(z), \quad (7.4.11)$$

Thus the key point behind this technique is to transform the data vector such that the eigenvalues of the resulting model (p-plane model), which are now situated, in an expanded region, can be estimated directly without any data interpolation.

Since $|z| < 1$ and $\epsilon < 1$, (7.4.9) and (7.4.10) can be truncated after a specified number of terms (depends upon the accuracy needed). It is obvious that as ϵ approaches unity, eventhough the region in which the p-plane poles are situated expands, one needs a larger number of terms of the above series as a result the computational effort and time are increased. Therefore a good choice of ϵ is about 0.95 so that one can sufficiently enlarge the region in which the p-plane poles are located without a very large increase in computation. Also note that the number of terms required increases with the order of the model. For five decimal point accuracy in computation of the series (7.4.9) and (7.4.10), the number of terms required is shown in fig.7.6 for models of different orders at $\epsilon = 0.95$. For the example considered below, ϵ is taken to be 0.93. It may also

recalled that at $\epsilon = 1$, the p operator will turn out to be an integrator, causing nonstationarity in the process.

After the transformation of the observation vector, any usual identification methods can be used for estimating the parameters of the p -plane model of the system.

Obtaining the proper sampling interval

If $\alpha \pm j\beta$ is an eigenvalue of the system under consideration in the sT plane. The corresponding eigenvalue in the p -plane with a contraction factor ϵ will be

$$1/(1 - \epsilon[\exp(\alpha)(\cos\beta \pm j\sin\beta)]) \quad (7.4.11)$$

which on simplification gives

$$P_1 \pm P_2$$

where

$$P_1 = \frac{[1 - \epsilon \exp(\alpha) \cos\beta]}{[1 - 2\epsilon \exp(\alpha) \cos\beta + \epsilon^2 \exp(2\alpha)]} \quad \text{and}$$

$$P_2 = \frac{[\epsilon \exp(\alpha) \sin\beta]}{[1 - 2\epsilon \exp(\alpha) \cos\beta + \epsilon^2 \exp(2\alpha)]} \quad (7.4.12)$$

It can be verified that the points A, B, C and D of fig.7.1 are mapped as points P, Q, R and S respectively, in fig.7.5. Note that when $\epsilon = 1$ (fig.7.5(a)) the point S is at infinity. Thus the boundary ABC of the lense shaped region of z -plane (fig.7.1) is mapped as the boundary PQR in the p -plane (figs. 7.5(a) and 7.5(b)). In other words, the

eigenvalue of the smallest magnitude in the z -plane (which is the eigenvalue of the largest magnitude in the s -plane) will be mapped as the eigenvalue of smallest magnitude in the p -plane.

Let $p_1^* \pm jp_2^*$ be a pole of smallest magnitude in the p -plane for a sampling interval T_1 . The corresponding z -plane pole can be easily obtained as

$$\frac{[(p_1^*)^2 + (p_2^*)^2 - (p_1^* \pm jp_2^*)]}{[\epsilon \{(p_1^*)^2 + (p_2^*)^2\}]}$$

$$= p \exp(\pm j\theta) \text{ say} \quad (7.4.13)$$

Proceeding exactly in the lines as explained in section 7.4.2, we can get the computationally optimum sampling interval T_2 as

$$T_2 = T_1 \times K \quad (7.4.14)$$

$$\text{where } K = 0.5 / [(\ln p)^2 + (\theta)^2]^{1/2} \quad (7.4.15)$$

By this way we can make sure that the pole of smallest magnitude in the z -plane lies on the boundary ABC and the rest of the poles lie within the lense shaped region of fig.7.1.

After getting the computationally optimum sampling interval, one can go back to usual direct or indirect methods of identifying continuous-time systems. To illustrate the procedure explained above, the following example is considered.

Example

Consider the fourth order system

$$H(s) = \frac{30(s+5)(s^2 + 4s + 8)}{(s+0.1)(s+0.2)(s^2 + 20s + 350)} \quad (7.4.16)$$

which is a stiff system. [Note that λ_m is 18.7082869 hence the optimum value of T is 0.02678 secs.]

The impulse invariant discrete-time model can be written as

$$H(z) = \frac{b_0 + b_1z + b_2z^2 + b_3z^3}{z^3 + a_1z^2 + a_2z + a_4} \quad (7.4.17)$$

with

$$b_0 = 30.0$$

$$b_1 = 2.144776\exp(-0.1T) - 60.12308\exp(-0.2T) \\ - 32.25550\exp(-10T)\sin(wT + 0.1218206)$$

$$b_2 = 2.016457\exp(-20T) + 27.983543\exp(-0.3T) \\ + 64.408908\exp(-10.2T)\sin(wT + 0.0608937) \\ + 60.383963\exp(-10.1T)\sin(wT + 1.5058382)$$

$$b_3 = 2.016457\exp(-20T) \\ + \exp(-10.3T)\sin(wT + 0.1391651)$$

$$\text{and } a_1 = -(\exp(-0.1T) + \exp(-0.2T) + 2\exp(-10T)\cos wT)$$

$$a_2 = \exp(-0.3T) + \exp(-20T) \\ + 2\exp(-10T)[\exp(-0.1T) + \exp(-0.2T)]\cos wT$$

$$a_3 = -(\exp(-20.1T) + \exp(-20.2T) + 2\exp(-10.3T)\cos wT)$$

$$a_4 = \exp(-20.3T)$$

where $w = 15.811388$ rad/sec.

The above system is perturbed with a PRBS and the output is obtained in the same way as explained in section 7.3.2. The sampling interval is arbitrarily chosen as 0.01 seconds. After this, the p-plane transformation is applied with a contraction factor 0.93. The parameters of the resulting p-plane model is then obtained by using the maximum likelihood method and the generalized least squares method. The series (7.4.9) and (7.4.10) are truncated after taking 300 terms, which gave sufficient accuracy. Some of the results are shown in table 7.3.

After this, steps (7.4.12) through (7.4.15) are performed to get the optimum sampling interval. The results are summarized in table 7.4

In order to demonstrate the superiority of the proposed method over the discrete-time model identification, the results of the z-domain model identification is given in table 7.5.

The simulation has been done on a VAX 11/780 computer in double precision (64 bit accuracy). Table 7.6 provides the CPU time taken to solve the example with 200 iterations and 1000 data points.

Note that this is one of the worst cases of system identification where the eigenvalues are scattered widely over the left half of the s-plane. Even in this case the proposed method gave satisfactory performance.

7.4.5 Critical evaluation of the p-plane method

Since the eigenvalues of the p-plane model are located in a wide region, one can afford to start with a very small arbitrary sampling interval by making ϵ near to 1 at the expense of a substantial increase in computation. Therefore, strictly speaking, priori knowledge of system time constants is not necessary, if one is ready to accept a large amount of computation. However, care must be taken to choose the initial sampling time so that no mode of the system is missed. Note that computational effort becomes larger when

(a) eigenvalues of the continuous-time model of the system are spread out in the left half of the s-plane (stiff system).

(b) the initial guess of T is very small,

i.e., $\lambda_m T \ll 0.5$ and ϵ is near to 1.0 and

(c) the order of the system is large.

Thus the accuracy in estimation is obtained at the expense of an increase in computational time and effort. But since the present day computers are powerful enough, one may consider that this is not a big drawback.

In the light of the discussions presented in the preceding sections, one arrives at a conclusion that if the system is stiff and no priory knowledge about the system time constants are available, the p-plane method is preferred. [A practical example of this can be found in the Ph.D. thesis by Karnik (1985) where he started with a

sampling interval of 0.001 seconds which is corrected to 0.008 seconds, while modelling the arm of an industrial robot (UNIMATE 2000)]. In other cases the w-plane method seems to sufficient, which is computationally much simpler.

7.5 Transformation from discrete-time models to continuous-time models

7.5.1 Introduction

As mentioned in section 7.2.1, if an indirect method is used to identify the continuous-time model of a system, it is necessary that the the discrete-time model should be transformed to its continuous-time equivalent. The commonly used methods (e.g. Sinha and Lastman, 1981; Hung et.al., 1980) are not computationally simple and involve matrix inversions. Because of this reason, these algorithms are not numerically robust. In the following discussions, numerically simple and robust (without any matrix ineversions) transformation algorithms are presented. At length, they are compared with other existing algorithms with the help of a numerical example.

7.5.2 A numerically robust transformation scheme

The transformation problem is to obtain the matrices A and B of eqn. (7.2.2b) from their discrete-time equivalents F and G of eqn. (7.2.2a).

Now from equation (7.2.3) the following relationships are obvious.

$$AT = \ln(F) \quad (7.5.1a)$$

$$\text{and } BT = (AT)(F-I)^{-1}G \quad (7.5.1b)$$

One of the popular methods is the fixed point iteration of Sinha and Lastman(1981) which is given in appendix V(A). Note that this method needs good starting values and involves many matrix inversions which is quite undesirable especially when the system under consideration is stiff. Another algebraic method proposed by Hung et.al. (1980) utilizes the states of the system and demands large computation. This algorithm is given in appendix V(B). Another draw back of these methods is that none of them provides error bounds for A and B which is quite important.

The following methods are simple, straightforward and provide compact error bounds for A and B.

From (7.5.1a) we obtain

$$AT = \ln[I+L] \text{ where } L = F-I \quad (7.5.2a)$$

$$= \sum_{k=1}^{\infty} (1/k)(-1)^{(k+1)} L^k \quad (7.5.2b)$$

The above series converges uniformly in all the cases of its applications in the present context of system identification by virtue of the following lemma.

Lemma 7.1 (Raol, Puthenpura and Sinha, 1985)

Let A be a nxn matrix describing a stable system and the spectral radius of AT = r(AT) < 0.5 [see equation (7.2.7)] so that T is the usual choice of the sampling

interval. Then the series (7.5.2b) converges uniformly.

Proof

$$r(AT) < 0.5 \Rightarrow \exp(-0.5) < \gamma_i < 1.0 ; i=1,2,\dots,n. \quad (7.5.3a)$$

where γ_i is the i -th root of the characteristic equation

$$\det[zI - F] = 0 \quad (7.5.3b)$$

Now the spectral radius of L ,

$r(L) = \max [|\lambda_i| ; i=1,2,\dots,n]$ where λ_i is the i -th root of the characteristic equation

$$\det[zI - (F-I)] = \det[(z+1)I - F] = 0. \quad (7.5.3c)$$

From (7.5.3b) and (7.5.3c) it follows that

$$\lambda_i + 1 = \gamma_i ; i=1,2,\dots,n \quad (7.5.3d)$$

which together with (7.5.3a) implies

$$|\lambda_i| < 1 ; i=1,2,\dots,n$$

i.e., $r(L) < 1$ and the proof is immediate.

Now from (7.5.1b) and (7.5.2b) it follows that

$$\begin{aligned} BT &= \left[\sum_{k=1}^{\infty} (1/k)(-1)^{(k+1)} L^k \right] L^{-1} G \\ &= \left[\sum_{k=1}^{\infty} (1/k)(-1)^{(k+1)} L^{k-1} \right] G \end{aligned} \quad (7.5.4)$$

More over, if one denotes

$$D = \left[\sum_{k=1}^{\infty} (1/k)(-1)^{(k+1)} L^{k-1} \right] \quad (7.5.5a)$$

then $AT = DL$ (7.5.5b)

and $BT = DG$ (7.5.5c)

Based on the above results a simple algorithm can be formulated as follows.

Select a positive integer N and (or) a nonnegative number ϵ .

Set $D^{(1)} = I$; $M^{(1)} = I$; $k = 1$

Repeat

$$\begin{aligned} M^{(k+1)} &\leftarrow -[k/(k+1)]LM^{(k)} \\ D^{(k+1)} &\leftarrow D^{(k)} + M^{(k+1)} \\ k &\leftarrow k+1 \end{aligned} \quad (7.5.6)$$

until $k = N$ or the metric $d(M^{(k+1)}, M^{(k)}) < \epsilon$. One useful metric (Lastman et. al., 1984) is found to be

$$\begin{aligned} \max_{i,j} \frac{|M_{i,j}^{(k+1)} - M_{i,j}^{(k)}|}{\max_{i,j} |M_{i,j}^{(k+1)}|} &= d(M^{(k+1)}, M^{(k)}) \end{aligned} \quad (7.5.7)$$

Error analysis

If the series (7.5.5a) is approximated by the first n terms,

$$\text{i.e., } D = P_n(L) + E_n(L) \quad (7.5.6a)$$

where

$$P_n(L) = \left[\sum_{k=1}^n (1/k)(-1)^{(k+1)} L^{k-1} \right] \quad (7.5.6b)$$

and $E_n(L)$ is the error in approximation. It can be easily

worked out that the error in the evaluation of AT is

$$\begin{aligned} \|E_n^{AT}(L)\| &= \|AT - P_n(L)\| \leq \\ &\sum_{k=n+1}^{\infty} (1/k) \|L^{k-1}\| \|L\| \\ &\leq [1/(n+1)] \|L^n\| \|L\| / [1 - \|L\|] \end{aligned} \quad (7.5.7a)$$

If the spectral radius of L is known it follows that

$$\|E_n^{AT}(L)\| \leq [1/(n+1)] [r(L)]^{n+1} / [1 - r(L)] \quad (7.5.7b)$$

Similarly the error in approximation of BT is computed as

$$\begin{aligned} \|E_n^{BT}(L)\| &\leq \sum_{k=n+1}^{\infty} (1/k) \|L^{k-1}\| \|G\| \\ &\leq [1/(n+1)] \|L^n\| \|G\| / [1 - \|L\|] \end{aligned} \quad (7.5.8a)$$

Hence

$$\|E_n^{BT}(L)\| \leq [1/(n+1)] [r(L)]^n [r(GG^T)] / [1 - r(L)] \quad (7.5.8b)$$

7.5.3 Reduction in computation using Chebyshev approximation

The Chebyshev approximation theory can be stated as follows.

Theorem 7.1 (Chebyshev)

Given a polynomial $P_n(x)$ of degree n with leading term $a_0 x^n$, the best approximation (in the least squares sense) to $P_n(x)$ by another polynomial $Q_m(x)$ of lower degree, at most $n-1$ in the range $(-1,1)$, is such that

$P_n(x) - Q_m(x) = kT_n(x)$, $k = a_0 2^{1-n}$, where $T_n(x)$ is the Chebyshev polynomial of degree n .

Proof

Refer any standard book on Chebyshev polynomials or approximation theory (e.g. Rivlin, 1974).

Now the first few Chebyshev polynomials and their reversals can be written as follows, for $x \in (-1, 1)$.

$$\begin{aligned} T_0(x) &= 1 & 1 &= T_0(x) \\ T_1(x) &= x & x &= T_1(x) \\ T_2(x) &= 2x^2 - 1 & x^2 &= (1/2)[T_0(x) + T_2(x)] \end{aligned}$$

and so on.

These will satisfy a recurrent relationship

$$T_{r+1}(x) = 2xT_r(x) - T_{r-1}(x) \quad (7.5.9a)$$

For $x \in (0, 1)$ the so called shifted Chebyshev polynomials can be written as,

$$\begin{aligned} T_0^*(x) &= 1 & 1 &= T_0^*(x) \\ T_1^*(x) &= 2x - 1 & x &= (1/2)[T_0^*(x) + T_1^*(x)] \end{aligned}$$

and so on with a recurrent formula

$$T_{r+1}^*(x) = 2(2x-1)T_r^*(x) - T_{r-1}^*(x) \quad (7.5.9b)$$

It is also important to note that $|T_r^*(x)|$ and $|T_r(x)| \leq 1$ in the respective ranges of x .

Example

Let $x \in (-1, 1)$.

We seek to approximate $p_3(x) = x^3 + x^2$ by a lower order polynomial.

From (7.5.9a) we get $p_3(x) = x^3 + x^2 = x^2 + (3/4)x + (1/4)T_3(x)$. Introducing the notation $q_r[p_n(x)]$ which means the r -th degree Chebyshev approximation of a polynomial $p_n(x)$ of degree n , where $r < n$,

$$q_2[p_3(x)] = x^2 + (3/4)x + n \text{ with } |h| < 1/4$$

which is the best lower order approximation of $p_3(x)$ in the least squares sense.

Now consider equation (7.5.6a) which may be written as

$$D = -Q_m(L) + H_n(L) \quad (7.5.10)$$

where $Q_m(L)$ is a lower order (i.e. $m < n$) Chebyshev approximation of $P_n(L)$. Note that $Q_m(L)$ is obtained by replacing $L^{m+1}, L^{m+2}, \dots, L^n$ of $P_n(L)$ by their expansions in terms of the shifted Chebyshev polynomials of degree at most degree m (Puthenpura and Sinha, 1984b). This, of course, introduces a little more error in the approximation of (7.5.2b) by a finite series but with a good amount of reduction in computation in terms of matrix multiplication.

7.4.5 Examples and comparison of the proposed method with other methods

The example given below is taken from Raol et.al. (1985).

Let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -1 & -2 & -2 \end{bmatrix} \text{ and } B = [1 \quad 3 \quad -2]^T$$

Also take $T = 0.5$

The matrix F and vector G are calculated and they are as follows.

$$F = \begin{bmatrix} 0.983876 & 0.465627 & 0.088281 \\ -0.088281 & -0.807313 & 0.289064 \\ -0.289064 & -0.666409 & 0.229185 \end{bmatrix}$$

and

$$G = [0.827214 \quad 1.204191 \quad -1.244470]^T$$

The results of various transformation algorithms are given below.

1. Direct method (Hung et.al., 1980)

$$\hat{A} = \begin{bmatrix} 0.001932 & 0.901357 & -0.002002 \\ -0.002121 & 0.001003 & 0.910123 \\ -0.919813 & -1.933714 & -1.934999 \end{bmatrix}$$

$$\hat{B} = [0.919735 \quad 2.909918 \quad -1.964137]^T$$

2. Fixed point iteration (Sinha and Lastman, 1981)

$$\hat{A} = \begin{bmatrix} 0.000000 & 1.000004 & 0.000004 \\ 0.000008 & 0.000002 & 1.000000 \\ -1.000000 & -2.000000 & -2.000002 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0.999997 & 2.999995 & -2.000005 \end{bmatrix}^T$$

3. The proposed algorithm

(a) With direct expansion of 13 terms

$$\hat{A} = \begin{bmatrix} 0.000000 & 0.999999 & -0.000002 \\ 0.000002 & 0.000001 & 1.000001 \\ -1.000001 & -2.000001 & -2.000002 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0.999995 & 2.999995 & -2.000002 \end{bmatrix}^T$$

(b) With Chebyshev approximation using 8 terms

$$\hat{A} = \begin{bmatrix} 0.000012 & 1.000023 & 0.000010 \\ -0.000011 & 0.000009 & 0.999999 \\ -0.999999 & -2.000000 & -2.000001 \end{bmatrix}$$

$$\hat{B} = \begin{bmatrix} 0.999979 & 3.000005 & -2.000005 \end{bmatrix}^T$$

The actual errors in the approximation of AT and BT , the respective error bounds and the computational requirements are also evaluated for these algorithms and are presented in table 7.8.

7.6 Concluding Remarks

Two important aspects of numerical robustness in the identification of continuous-time systems have been considered. The first problem, which is the choice of

sampling interval for identification, is utmost important and the possible solutions suggested in this chapter are only adhoc. However, they seem to work in a number of simulated examples and a real-life problem (see Karnik, 1985). It is also noteworthy that the only existing solutions of the sampling time problem (from the numerical robustness point of view) are those given in this chapter, to the best of the knowledge of the author.

The second problem of transformation of discrete-time models to their continuous-time equivalents is solved in simple and attractive way. This method has been gathered much popularity in recent times due to its efficiency, accuracy, knowledge about the error bounds and above all the computational simplicity.

Table 7.1
Results of parameter estimation for different sampling intervals (discrete-time model)

T in secs.	$\lambda_m T$	No. of samples used	Actual sys. parameters	Estim. sys. parameters SNR = 10dB	Error Norm
0.01	0.03	15000	0.01956 -0.01946 -1.95064 0.95123	0.33470 -0.33400 -1.10665 0.11055	0.5859
0.10	0.3	1500	0.16006 -0.15223 -1.55955 0.60653	0.33470 -0.32924 -1.47701 0.54538	0.2153
0.15	0.45	100	0.21518 -0.19953 -1.37845 0.47237	0.22262 -0.21457 -1.28353 0.33768	0.1114
0.20	0.60	75	0.25746 -0.23267 -1.23267 0.36789	0.42486 -0.38489 -0.81855 0.05908	0.4198

Note: The parameter error norm is the same as that given by (4.8.2). The order in which the parameters are given is a_1 , a_2 , b_1 and b_2 . SNR means signal-to-noise ratio.

Table 7.2
Results of parameter estimation for different sampling
intervals (continuous-time model)

T in secs.	$\lambda_m T$	No. of samples used	Actual sys. parameters	Estim. sys. parameters SNR = 25dB	Error Norm
0.01	0.03	1500		2.3091	0.5609
				0.1173	
				2.4385	
				2.3367	
0.10	0.30	150	2.0000	2.1119	0.1162
			1.0000	1.2731	
			5.0000	5.1487	
			6.0000	6.8843	
0.15	0.45	100		1.8934	0.0559
				1.1001	
				5.1991	
0.20	0.60	75		6.3718	0.6358
				0.6843	
				0.3978	
				1.6103	
				2.3817	

TABLE 7.3
Estimated poles of the p-plane model with T=0.01 secs.
and contraction factor = 0.93

SNR in dB.	Estimated p-plane poles		Actual Values.
	Max. likelihood Method.	Gen. least sq. Method .	
30	14.100836	14.29327	14.098499 13.916307 3.664604 ± j2.873147
	13.901899	13.818845	
	3.669893 ±	3.7000001 ±	
	j2.881432	j2.791838	
15	13.814432	12.999415	
	13.001337	13.007719	
	3.441837 ±	3.404149 ±	
	j2.778108	j2.616337	

TABLE 7.4
The optimum value of the sampling interval in different
cases

SNR in dB	Max. likelihood Method.	Gen. least sq. Method .
30	0.02141	0.02122
15	0.02022	0.01953

TABLE 7.5
Estimated poles of the z-plane model with T=0.01 secs.

SNR in dB.	Estimated z-plane poles		Actual Values.
	Max. likelihood Method.	Gen. least sq. Method .	
30	1.118442	1.347816	
	0.810018	0.743533	
	0.483744 ±	0.412439 ±	0.999001
	j0.637551	j0.574483	0.998002
			0.893551 ±
			j0.142470
15	procedures failed due to numerical instability		

TABLE 7.6
CPU time (seconds) requirements for different procedures
of identification
(SNR = 30 dB, T = 0.01 sec)

	Max. likelihood	Gen. least squares
z-plane	3.24	2.11
p-plane ε = 0.93	16.29	15.15

TABLE 7.7
Comparison of various transformation algorithms
in terms of accuracies and computational requirements

	Hung et.al.	Sinha and Lastman	Proposed Algorithm	
Comp. method	Least Squares	Fixed point iteration	Direct Expansion 13 terms	Chebyshev Approx. 13 terms
$r(\hat{A}T-AT)$	4.23×10^{-3}	1.38×10^{-7}	1.01×10^{-6}	7.33×10^{-6}
$\ E_n^{AT}(L)\ $	-	-	1.26×10^{-5}	8.89×10^{-5}
$r(\hat{B}T-BT)$	7.61×10^{-5}	2.15×10^{-11}	2.43×10^{-11}	5.34×10^{-10}
$\ E_n^{BT}(L)\ $	-	-	5.49×10^{-6}	8.97×10^{-6}
Number of Matrix Mult.	-	22	14	9
Number of Matrix Inv.	-	3	0	0
<u>Note:</u> $r(\cdot)$ represents the spectral radius				

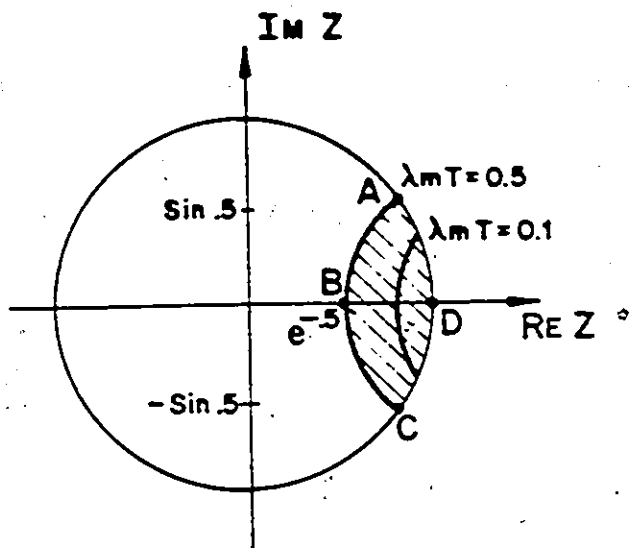


Fig. 7.1 Regions in the z-Plane where the Eigenvalues of the Discrete-time Model of a Stable System Must Occur for Different Sampling Intervals

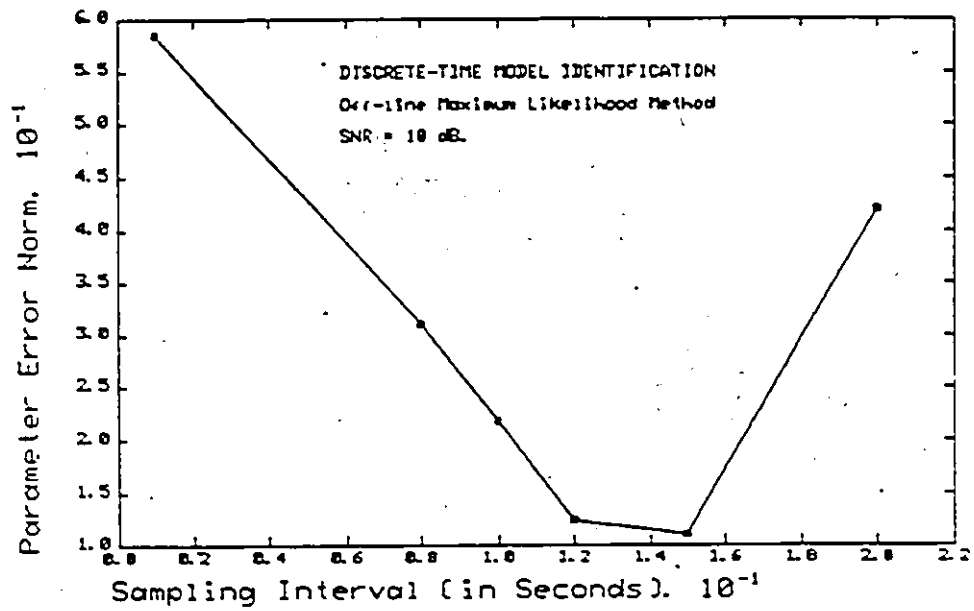


Fig. 7.2 Parameter Error Norm for Different Sampling intervals (Discrete-time Identification)

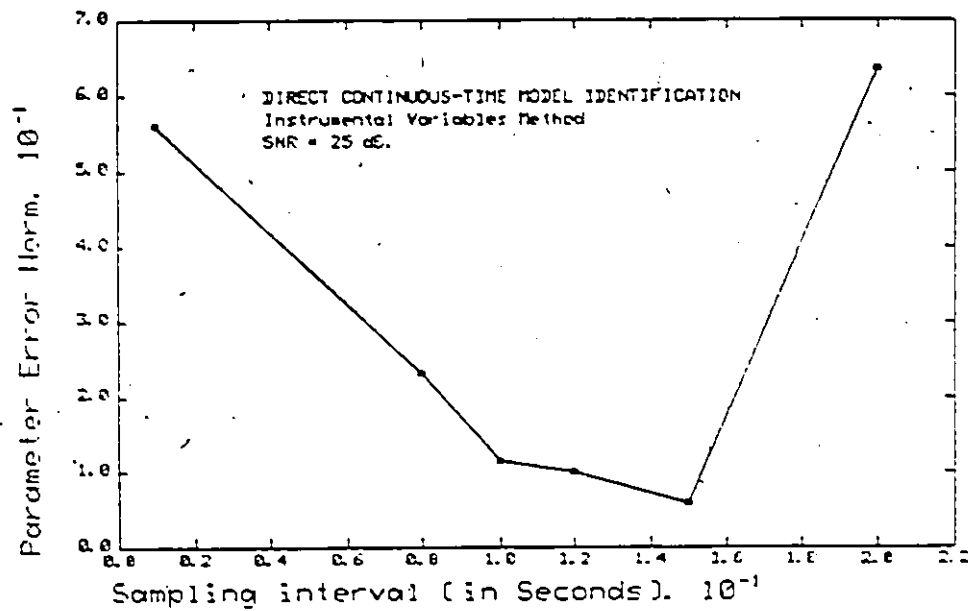


Fig. 7.3 Parameter Error Norm for Different Sampling intervals (Direct Method)

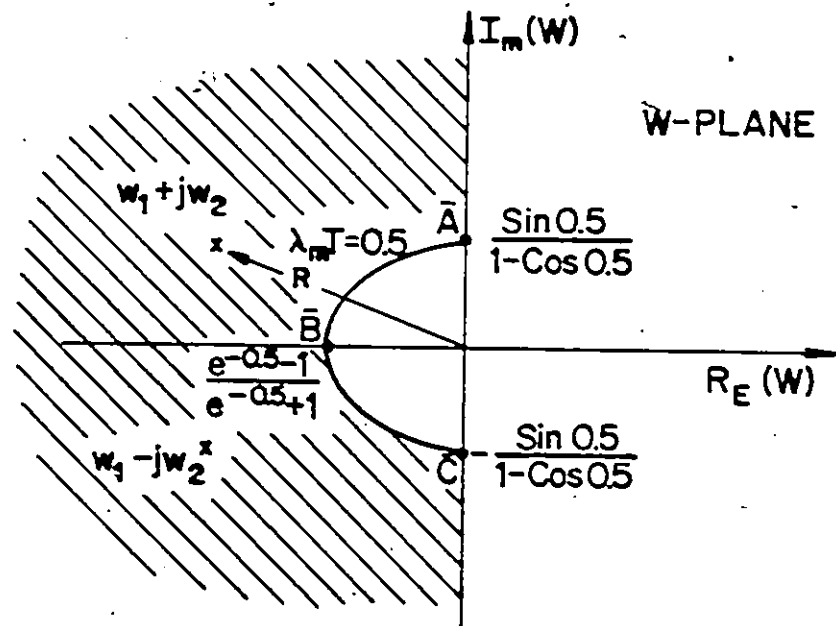


Fig. 7.4 Mapping from the z -Plane to w -Plane

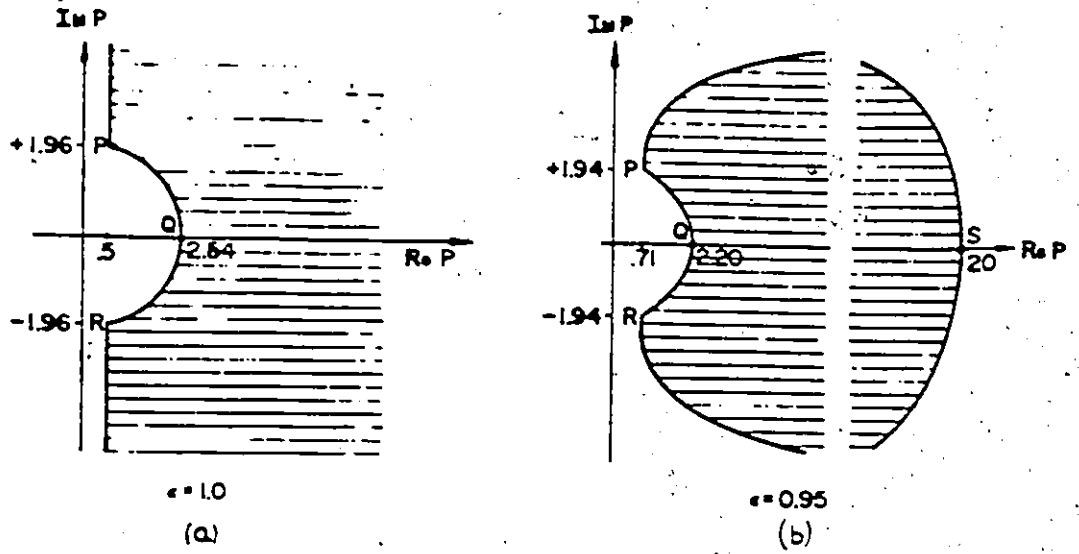


Fig. 7.5 Mapping from the z-Plane to p-Plane for Different Contraction Factors

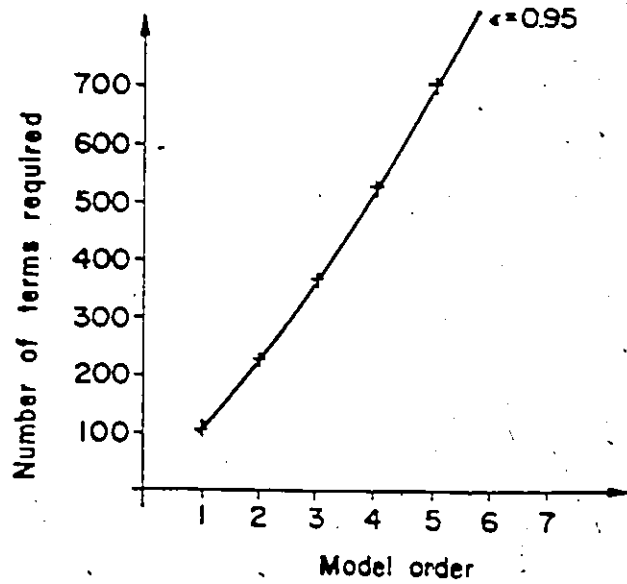


Fig. 7.6 Number of Terms Required in the Series Expansion in the p Domain with Contraction Factor 0.95 for 5 Decimal Place Accuracy

CHAPTER 8

CONCLUSIONS AND RECOMMENDATIONS

8.1 Conclusions

In this thesis, an effort has been made towards obtaining the robust solutions of various problems in identification and control of practical systems. The major emphasis has been on studying the effects of non normality of the data on identification and the introduction of new on-line and off-line identification schemes which are protected from spurious data (outliers).

The discussions presented in chapter 2 give a brief idea of the existing robust statistical methods which may be applied to system identification and signal processing problems. However, it is to be pointed out that some of these methods are complex from the computational point of view. As a result one may think that the computational requirements not worth the extra accuracy obtained by using robust methods. This is not quite true since one can reduce the computations substantially by introducing some extra techniques, approximations or simplifications. For example, see Tiku's linear approximation in chapter 3 and the weighted least squares type formulation of the non linear M-estimation problem in chapter 4. Thus it is clear that in most of the situations robust methods can replace

conventional least squares methods effortlessly with the help of proper substitutions or approximations.

The modified maximum likelihood (MML) method, which is based on censored samples of order statistics, is used for the first time in the context of system identification in this thesis (chapter 3). This method takes only almost same CPU time as the conventional maximum likelihood method, at least in the case of single input single output systems. The draw back of this method is the requirement of several samples of observations at each time point, which is the general draw back of any methods based on order statistics. However, situations exist where one uses several transducers to get the same data, or data over several experiments on the system. The MML method is well suited for such situations.

Several robust on-line parameter identification algorithms are introduced for the first time in chapter 4 and 5. It is obvious that the choice of a particular algorithm is entirely dependant on the context of application. For example, if the measurement noise is highly interfering with the data, the robust instrumental variables method is preferred to the robust weighted least squares type method but with an extra effort in generating the instrumental variables. But the key point in all these algorithms is one and the same; that is to say, posing the non linear estimation problem as a weighted least squares

type problem. Here, the weights are chosen in such a way that the necessary condition of optimality is satisfied so that the resulting estimates are asymptotically equivalent to the M-estimates.

The convergence of the recursive algorithms are established theoretically. The tools used for convergence analysis are quite well known and standard (eg. the method of Robbins and Monro). The weighted least squares type robust algorithm convergence in the mean squares sense under relatively weak assumptions. However, the convergence of the robust instrumental variables method is not proven to be in quadratic mean but slightly weak (i.e., convergence in probability) under not so weak assumptions. At this point further knowledge about the convergence properties of the robust instrumental variables method is far from obvious.

The simplified robust algorithm (which is derived in a Bayesian approach) presented in chapter 5 looks very promising. In fact, this may be the simplest and the most straightforward robust identification algorithm existing. The advantage of this method is its versatility. An immediate extension of this robust estimation algorithm is to robust Kalman filtering and joint parameter-state estimation.

The efficiency of these robust on-line algorithms is also discussed in a very general set up. The generalized Cramer-Rao inequality is used for this purpose. This method is quite simple but has a draw back from the point of

view that only a 'bound' is used to evaluate the efficiency. The theorems presented in chapter 4 (theorems 4.4 and 4.5) and several illustrative examples give good insight into the advantages and disadvantages of this approach.

Another important achievement presented in this thesis is the introduction of a robust bootstrap algorithm for the joint estimation of parameters and states of a linear system, which may be multivariable in structure. As far as the author is aware, this is the first robust algorithm for this purpose. The treatment of this algorithm is quite wide and general so that other existing bootstrap algorithms are only special cases of this method. This algorithm is extremely useful in adaptive control with state variable feed back. The proof of convergence is via martingale theory and this is the most accepted method for analyzing pseudo linear regressions (PLR). The positive real (causality) condition for the convergence of the robust bootstrap method is a typical sufficient condition for the convergence of any PLR's. Attempts to weaken this strong condition are yet to be succeeded.

The common disadvantage of all these robust recursive method is the relatively slow convergence rate in comparison with their non robust counterpart when the data actually enjoy nonnormality. However this is not a major drawback if we have enough data points, which often the case. Also it is to be remembered that in practice the data available are seldom

normal.

The real life example of identifying a packed bed chemical reactor gives a good physical feel of the application and performance of the various robust recursive algorithms.

It is also important to note that the robust identification algorithms presented in this thesis require only minor modifications of the existing software developed for the conventional (non robust) identification algorithms.

Numerical robustness aspects in the identification of continuous-time systems described in chapter 7 is one of the very few studies made on this topic. The solutions presented to the sampling time choice problem, eventhough ad-hoc, provide reasonably good tools for the practising people to tackle this difficult problem. To the best of the knowledge of the author, these are the only reported solutions to the the problem of optimal choice of sampling interval.

Numerically attractive methods for transforming discrete-time models to their continuous-time equivalents are also presented in chapter 7. They are simple to use and devoid of any complex matrix computations like eigenvalue determination and inversion. The compact error bounds on the transformed matrices can also be evaluated easily. It is also noteworthy that these are the only methods where error bounds on transformed matrices are available.

8.2 Suggestions for further research

The various discussions presented in this thesis open several challenging problems. Some of the interesting ones are enumerated concisely below.

(1). Application of other robust methods (see chapter 2) to the identification problem. This may require some approximations or substitutions which are applicable in the context of system identification for computational simplicity.

(2). Multivariable version of the MML method introduced in chapter 3, considering the properties of multivariate distributions. One may adopt the approach of Abraham and Box (1975) for this purpose.

(3). The multivariable version of the weighted least squares type method introduced in chapter 4. Note that in multivariable case there will be 'coupling of noise' which adds extra complexity to the robustness problem. Thus, a careful look at this problem shows that the extension of the single input single output (SISO) algorithm to the multivariable case is not conceptually straightforward. One possible paper that may help the interested is due to Rajbman and Sinha (1977).

(4). Extension of the idea presented in chapter 4 to the Kalman filtering problem. Note that in the case of Kalman filter, the weight matrix $\alpha(t)$ is a function of the prediction error. Hence one really runs into difficulty here and the solution to this problem becomes quite

challenging. [Note that the algorithm in chapter 5 can be easily extended to the Kalman filtering problem unlike the algorithm presented in chapter 4]. A reference that may give some idea for handling this problem is Boncelet and Dickinson (1983).

(5). Derivation and analysis of SISO and multivariable PLR's based on the algorithm in chapter 4. For example, the analysis of the bootstrap algorithm utilizing the estimation algorithm in chapter 4 is very complicated due the fact that the the regression vector $\phi(t)$ and the weight $\alpha(t)$ are coupled unlike the bootstrap algorithm in chapter 6. Needless to say, a multivariable version of this algorithm will be a much more complicated one.

(6). A general robust method for identifying time varying systems. [See the algorithm in chapter 5 for tracking time varying parameters for the simplified case]. Also the recent paper of Kovacevic and Filipovic (1985) is a good reference in this topic.

(7). Application of the robust weighted least squares type method to ladder (lattice) structures (eg. Goodwin and Sin, 1984) in the context of adaptive filtering. This will be a good achievement for biomedical signal processing.

(8). A more mathematically rigorous treatment of the sampling time problem. That is, one can link the word length of the digital computer used for identification to the sampling rate and get a more elegant solution to the problem

of the computationally optimum sampling interval for identification. For further guidelines see the recent paper by Goodwin (1985).

These are roughly some of the possible and useful extensions of the various results presented in the thesis. Probably a good way to conclude this thesis is by quoting John W. Tukey who contributed a lot to the field of robust estimation.

"NORMALITY IS A MYTH. IT HAS NEVER HAPPENED, AND WILL NEVER HAPPEN !"

APPENDIX I

OUTLINE OF DERIVATION OF THE MML ESTIMATION

ALGORITHM

Consider a simple regression problem

$$y_i(k) = \theta x(k) + e_i(k) \quad (I.1)$$

$$k = 0, 1, \dots, N$$

$$i = 0, 1, \dots, p$$

$e_i(k)$ is the noise assumed to be normally distributed with zero mean and standard deviation σ . Also note that y , θ , x and e are scalars. Let at the k -th instant the p observations $y_i(k)$'s are arranged in ascending order and the extreme r observations on each side are censored. Therefore, we have

$$y_{r+1}(k) \leq y_{r+2}(k) \leq \dots \leq y_{p-r}(k) \quad (I.2)$$

taking the likelihood function as the one described by (3.2.5), we get $\hat{\theta}$ (estimate of θ) as the solution of the equation (Harter and Moore, 1966)

$$[\partial L / \partial \theta] = (1/\sigma) \sum_{k=0}^N x(k) \left\{ \sum_{i=r+1}^{p-r} z_i(k) - r g_1(z_{r+1}(k)) + r g_2(z_{p-r}(k)) \right\} = 0 \quad (I.3)$$

where $z_i(k) = [y_i(k) - \mu] / \sigma$; μ and σ are the location and scale parameters, respectively. The functions $g_1(\cdot)$ and

$g_2(\cdot)$ are defined as

$$g_1(z) = f(z)/F(z)$$

and

$$g_2(z) = f(z)/[1 - F(z)] \quad (\text{I.4})$$

where $f(z)$ is the pdf defined by (3.2.5) and $F(z)$ is its cumulative distribution function.

Similarly the the estimate of scale (σ) is obtained from

$$[\partial L / \partial \sigma] = (1/\sigma) \sum_{k=0}^N \{ -(p-2r) + \sum_{i=r+1}^{p-r} z_i^2(k) - r z_{r+1}(k) g_1(z_{r+1}(k)) + r z_{p-r}(k) g_2(z_{p-r}(k)) \} = 0 \quad (\text{I.5})$$

Now make the approximations (Tiku, 1967)

$$g_1(z_{r+1}(k)) = \alpha - \beta z_{r+1}(k)$$

and

$$g_2(z_{p-r}(k)) = \alpha + \beta z_{p-r}(k) \quad (\text{I.6})$$

where

$$\beta = -f(t)[t - f(t)/q]/q ; \alpha = f(t)/q - \beta t$$

t is determined by $1 - F(t) = q = r/p$.

Thus α and β are functions of the degree of censoring (q) and are tabulated by Tiku (1967).

Now (I.3) and (I.5) will be transformed as

$$[\partial L / \partial \theta] = (1/\sigma) \sum_{k=0}^N x(k) \{ \sum_{i=r+1}^{p-r} z_i(k) + r\beta [z_{r+1}(k) + z_{p-r}(k)] \} = 0 \quad (\text{I.7a})$$

and

$$[\partial L / \partial \sigma] = (1/\sigma) \left\{ \sum_{k=0}^N \left[-(p-2r) + \sum_{i=r+1}^{p-r} z_i^2(k) - r\alpha [z_{r+1}(k) - z_{p-r}(k)] + r\beta [z_{r+1}^2(k) + z_{p-r}^2(k)] \right] \right\} = 0 \tag{I.7b}$$

From these equations, the MML estimates of θ and σ are immediately obtained as follows (Tiku, 1981).

$$\hat{\theta} = \frac{\sum_{k=0}^N [x(k) - \bar{x}] W(k)}{\sum_{k=0}^N [x(k) - \bar{x}]^2} \tag{I.8}$$

where

$$W(k) = \sum_{i=r+1}^{p-r} y_i(k) + r\beta [y_{r+1}(k) + y_{p-r}(k)] \tag{I.9}$$

and $\bar{x} = [1/(N+1)] \sum_{k=0}^N \hat{x}(k)$

The estimate of σ ($\hat{\sigma}$) is obtained as the positive root of the quadratic equation;

$$A\hat{\sigma}^2 - B\hat{\sigma} - C = 0 \tag{I.10a}$$

$$A = (N+1)(p-2r)$$

$$B = \sum_{k=0}^N r\alpha [y_{p-r}(k) - y_{r+1}(k)] \quad \text{and}$$

$$C = \sum_{k=0}^N \left\{ \sum_{i=r+1}^{p-r} y_i^2(k) + r\beta [y_{r+1}^2(k) + y_{p-r}^2(k)] \right\}$$

$$+ \left[\frac{1}{N+1} \right] \left[\sum_{k=0}^N W(k) \right]^2 - \hat{\theta} \left[\sum_{k=0}^N x(k) - \bar{x} \right]^2 \quad (I.10b)$$

Finally the variance of $\hat{\theta}$ is obtained as (Sarhan and Greenberg, 1962)

$$V(\hat{\theta}) = \hat{\sigma} / H(x) \quad \text{where}$$

$$H(x) = \left(\left[\sum_{k=0}^N x(k) - \bar{x} \right]^2 \right) (p - 2r + 2r\beta) \quad (I.11)$$

By changing $x(k)$ and θ to $\phi(k)$ and $\underline{\theta}$ (vectors) respectively, one gets the results depicted in (3.3.7) through (3.3.17)

APPENDIX II

PROOF OF THEOREM 4.1

We have from (4.4.9a) that

$$\hat{\theta}(t) = \hat{\theta}(t-1) + \frac{P(t-1)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)]}{[1/\alpha(t)] + \phi^T(t)P(t-1)\phi(t)} \quad (\text{II.1})$$

Denote $\varepsilon_1(t) = \phi^T(t)\hat{\theta}(t-1) - y(t) = -\varepsilon(t)$

(see lemma 4.3)

Then (II.1) can be written as

$$\hat{\theta}(t) = \hat{\theta}(t-1) - \frac{P(t-1)\phi(t)\varepsilon_1(t)}{[\varepsilon_1(t)/\psi(\varepsilon_1(t))] + \phi^T(t)P(t-1)\phi(t)} \quad (\text{II.2})$$

Since $\psi(\cdot)$ is odd $\alpha(t) = [\varepsilon(t)/\psi(\varepsilon(t))] = [\varepsilon_1(t)/\psi(\varepsilon_1(t))]$.

Let $\tilde{\theta}(t) = \hat{\theta}(t) - \theta_0$. Then

$$\tilde{\theta}(t) = \tilde{\theta}(t-1) - \frac{P(t-1)\phi(t)\psi(\varepsilon_1(t))}{1 + \alpha(t)\phi^T(t)P(t-1)\phi(t)} \quad (\text{II.3})$$

Let $d(t) = 1 + \alpha(t)\phi^T(t)P(t-1)\phi(t) > 0 \quad \forall t \in [0, \infty)$.

From (II.3) it follows that

$$\begin{aligned} E[\tilde{\theta}^T(t)\tilde{\theta}(t)|\hat{\theta}(t-1)] &= \tilde{\theta}^T(t-1)\tilde{\theta}(t-1) \\ &+ \phi^T(t)[P(t-1)]^2\phi(t)E\{[\psi_1(\varepsilon_1(t))]^2|\hat{\theta}(t-1)\} \\ &- 2E[\psi_1(\varepsilon_1(t))|\hat{\theta}(t-1)]\phi^T(t)P(t-1)\tilde{\theta}(t-1) \end{aligned} \quad (\text{II.4})$$

where $\psi_1(\varepsilon_1(t)) = [1/d(t)]\psi(\varepsilon_1(t))$.

Let

$E[\hat{\theta}^T(t)\hat{\theta}(t)] = V(t)$. Then from (II.4) we get (by taking expectation again)

$$V(t) = V(t-1) + p(t) - 2q(t) \quad (\text{II.5})$$

where

$$p(t) = E\{\phi^T(t)[P(t-1)]^2\phi(t)E\{[\psi_1(\varepsilon_1(t))]^2|\hat{\theta}(t-1)\}$$

and

$$q(t) = E\{E[\psi_1(\varepsilon_1(t))|\hat{\theta}(t-1)]\phi^T(t)P(t-1)\hat{\theta}(t-1)\} \quad (\text{II.6})$$

Obviously, $p(t) \geq 0$, $d(t) > 0$ and $V(t) \geq 0 \forall t \in [0, \infty)$.

Now,

$$\begin{aligned} & \text{sign of } \{E[\psi_1(\varepsilon_1(t))|\hat{\theta}(t-1)]\} \\ & = \text{sign of } \{E[\psi(\phi^T(t)\hat{\theta}(t-1) - e(t))]\}, \quad (\text{II.7}) \end{aligned}$$

assuming $\phi(t)$ and $e(t)$ are independent and $E(\cdot)$ is the expectation with respect to the measure defined in section 4.5. Applying lemma 4.2 on (II.7) it can be easily verified that $q(t) \geq 0 \forall t \in [0, \infty)$.

From (II.5) we get

$$V(t) = V(1) + \sum_{n=2}^t p(n) - 2 \sum_{n=2}^t q(n) \geq 0 \quad (\text{II.8})$$

from which it follows that

$$\sum_{n=2}^t q(n) \leq (1/2)[V(1) + \sum_{n=2}^t p(n)]$$

Let $\varepsilon(t)$ be bounded. From lemma 4.3 it follows that $\{p(n)\}$

converges and hence $\{q(n)\}$ converges.

Therefore, $\lim_{t \rightarrow \infty} V(t) = V > 0$ exists. (II.9)

Now the scalar multiplier sequence of the algorithm (II.1) $\{g(t)\}$ is such that (see Ljung and Soderstrom, 1983; Poljak and Tsypkin, 1979)

$$g(t) = \frac{1}{[1/\alpha(t)] + \phi^T(t)P(t-1)\phi(t)}$$

$$= \frac{\alpha(t)}{1 + \alpha(t)\phi^T(t)P(t-1)\phi(t)}$$

is non negative $\forall t \in [0, \infty)$.

From lemma 4.3, $\text{trace}P(t) < \text{trace}P(t-1)$.

Therefore if

$$\limsup_{N \rightarrow \infty} (1/N) \sum_{t=1}^N |\phi(t)|^2 < \infty$$

we see that $\lim_{t \rightarrow \infty} \phi^T(t)P(t-1)\phi(t) < \infty$ (e.g., Ljung and Soderstrom, 1983)

So, it follows that $\lim_{t \rightarrow \infty} g(t) \neq 0$

which implies

$$\sum_{t=1}^{\infty} g(t) = \infty \quad (\text{II.10})$$

Using (II.8), (II.9) and (II.10), from Robbins and Monro (1951) it follows that

$$\lim_{t \rightarrow \infty} V(t) = 0$$

t →

which means

$$\lim_{t \rightarrow \infty} \theta(t) = \theta_0 \text{ in quadratic mean.}$$

t →

APPENDIX III

PROOF OF THEOREM 4.6

We have from (4.9.3) and (4.9.7a)

$$\begin{aligned} \hat{\theta}(t) &= \hat{\theta}(t-1) \\ &+ \frac{P(t-1)\zeta(t)[\phi^T(t)\theta_0 - \phi^T(t)\hat{\theta}(t-1) + e(t)]}{[1/\alpha(t)] + \phi^T(t)P(t-1)\zeta(t)} \end{aligned} \quad (\text{III.1})$$

Let $\tilde{\theta}(t) = \hat{\theta}(t) - \theta_0$ and

$$d(t) = [1/\alpha(t)] + \phi^T(t)P(t-1)\zeta(t).$$

Therefore from (III.1)

$$\begin{aligned} \tilde{\theta}(t) &= \tilde{\theta}(t-1) - \frac{P(t-1)\zeta(t)\phi^T\tilde{\theta}(t-1)}{d(t)} \\ &+ \frac{P(t-1)\zeta(t)e(t)}{d(t)} \end{aligned} \quad (\text{III.2})$$

Now we construct the error function, also popular as the 'Mahalanobis distance' (cf. Vinod and Ullah, 1981)

$$V(t) = \tilde{\theta}^T(t)[P(t)]^{-1}\tilde{\theta}(t) = \tilde{\theta}^T(t)R(t)\tilde{\theta}(t) \geq 0$$

and define $\Delta V(t) = V(t) - V(t-1)$.

i.e.,

$$\Delta V(t) = \tilde{\theta}^T(t)[P(t)]^{-1}\tilde{\theta}(t) - \tilde{\theta}^T(t-1)[P(t-1)]^{-1}\tilde{\theta}(t-1) \quad (\text{III.3})$$

Now applying matrix inversion lemma on eqns. (4.9.7a) and (4.9.7b) (see Goodwin and Sin, 1984)

$$\tilde{\theta}^T(t) = P(t)[P(t-1)]^{-1}\tilde{\theta}(t-1) \quad (\text{III.4})$$

Therefore from (III.3) and (III.4)

$$\Delta V(t) = [\tilde{\theta}^T(t) - \tilde{\theta}^T(t-1)]^T [P(t-1)]^{-1} \tilde{\theta}(t-1) \quad (\text{III.5})$$

Using (III.2) and (III.5)-

$$\Delta V(t) = - \frac{\tilde{\theta}^T(t-1)\phi(t)\zeta^T(t)\tilde{\theta}(t-1)}{d(t)} + \frac{e(t)\zeta^T(t)\tilde{\theta}(t-1)}{d(t)} \quad (\text{III.6})$$

Let us denote $\Delta V_1(t) = \Delta V(t)d(t)$.

Let $F(t)$ be the increasing sigma algebra comprising of the instrumental variables and observations upto time t . Assume that $\tilde{\theta}(t)$ is adapted to $F(t)$, i.e., $F(t)$ measurable.

Also remembering that for any random variable $x: (\Omega, F, P) \rightarrow (R, B_R)$ [where (Ω, F, P) is the probability space and (R, B_R) the Borel field] such that the induced probability measure P_x defined by

$$\forall A \in B_R, P_x(A) = \int_A f(x)d(x) \text{ exists by virtue of lemma 4.1 (Loe'Ve, 1977).}$$

Under the above circumstances

$E[\Delta V_1(t)|F(t)]$ exists and its value will be equal to

$$- \tilde{\theta}^T(t-1)\phi(t)\zeta^T(t)\tilde{\theta}(t-1) \quad (\text{III.7})$$

if $\zeta(t)$ and $e(t)$ are independent and $e(t)$ has zero mean.

Note that (III.7) is always negative definite $\forall t \in [0, \infty)$ if

$$\langle \phi(t), \zeta(t) \rangle > 0 \quad \forall t \in [0, \infty)$$

and

$$\limsup_{N \rightarrow \infty} (1/N) \sum_{t=1}^N \langle \phi(t), \zeta(t) \rangle < -$$

The latter along with

$$\sup_{t \in [0, \infty)} \{1/\alpha(t)\} < -$$

ensures that $0 < d(t) < - \quad \forall t \in [0, \infty)$.

Therefore the error function (Mahalanobis distance) is expected to decrease with respect to the increasing sigma algebra of observations and this ensures the convergence of the robust IV algorithm in probability (Mahalanobis, 1936; Lukacs, 1975) which means

$\exists N: \forall t > N, E[V(t)|F(t)] < \delta; \delta > 0$ and arbitrary small

i.e., $\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta_0$ in probability.

APPENDIX IV

PROOF OF THEOREM 6.1

To begin with consider the following theorem (Ne'veu, 1975).

Theorem A-IV.1

Let $\{V(t)\}$, $\{s_1\}$ and $\{s_2\}$ be sequences of random variables adapted to a sequence of increasing sigma algebras $F(t)$ [i.e., $V(t), s_1, s_2 \in F(t)$] and $E[V(t)|F(t-1)] \in V(t-1) + s_1 - s_2$ and

$$\sum_{t=1}^{\infty} s_1 < \infty \text{ with probability 1 (w.p.1)}$$

then $\lim_{t \rightarrow \infty} V(t) = V$ and $\sum_{t=1}^{\infty} s_1 < \infty$ with probability 1 (w.p.1).

Now we proceed as follows:

Denote $\Lambda(t) = [1 - \delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2 + \phi^T(t)P(t-1)\phi(t)$.

Then from (5.3.11) we get

$$\hat{\theta}(t) = \hat{\theta}(t-1) + [1/\Lambda(t)]P(t-1)\phi(t)[y(t) - \phi^T(t)\hat{\theta}(t-1)] \tag{IV.1}$$

$$\text{and } P^{-1}(t) = P^{-1}(t-1) + [1/\Lambda(t)]\phi(t)\phi^T(t) \tag{IV.2}$$

Let θ_0 be the actual parameter vector and define

$$\bar{\theta}(t) = \hat{\theta}(t) - \theta_0 \text{ and construct}$$

$$V(t) = \bar{\theta}^T(t)P^{-1}(t)\bar{\theta}(t) > 0 \quad \forall t \in [0, \infty)$$

Subtracting θ_0 from both sides of (IV.1), pre-multiplying by $P^{-1}(t-1)$ and denoting $[y(t) - \phi^T(t)\hat{\theta}(t-1)] = \varepsilon(t)$ we get

$$P^{-1}(t-1)\tilde{\theta}(t) = P^{-1}(t-1)\tilde{\theta}(t-1) + [1/\Lambda(t)]\phi(t)\varepsilon(t)$$

Adding $[1/\Lambda(t)]\phi(t)\phi^T(t)\tilde{\theta}(t)$ on both sides, pre-multiplying by $\tilde{\theta}^T(t)$ and using the matrix inversion lemma

$$V^*(t) = V^*(t-1) + \phi^T(t)\varepsilon(t)\tilde{\theta}(t-1) + \tilde{\theta}^T(t)\phi(t)\varepsilon(t) + [\tilde{\theta}^T(t)\phi(t)]^2 - \phi^T(t)P(t-1)\phi(t)\varepsilon^2(t) \quad (IV.3)$$

where $V^*(t) = V(t)\Lambda(t) > 0 \quad \forall t \in [0, \infty)$.

Making a substitution in (IV.3) that

$$\begin{aligned} & \tilde{\theta}^T(t)\phi(t)\varepsilon(t) + [\tilde{\theta}^T(t)\phi(t)]^2 \\ &= \phi^T(t)\tilde{\theta}(t)\{\phi^T(t)\tilde{\theta}(t) + [\varepsilon(t) - e(t)]\} + e(t)\phi^T(t)\tilde{\theta}(t) \end{aligned} \quad (IV.4)$$

and noting that (a few steps are omitted here for brevity)

$$E[e(t)\phi^T(t)\tilde{\theta}(t)|F(t-1)] = \phi^T(t)P(t)\phi(t)[(1-u)\sigma_1^2 + u\sigma_2^2]$$

$$\text{where } u = \text{prob}[\delta(t) = 1] \quad (IV.5)$$

we get the following expression after some simplifications.

$$\begin{aligned} E[V^*(t)|F(t-1)] &= V^*(t-1) - 2E[p(t)q(t)|F(t-1)] \\ &+ 2\phi^T(t)P(t)\phi(t)[(1-u)\sigma_1^2 + u\sigma_2^2] \\ &- \phi^T(t)P(t-1)\phi(t)E[\varepsilon^2(t)|F(t-1)] \end{aligned} \quad (IV.6)$$

$$\text{where } p(t) = -\phi^T(t)\tilde{\theta}(t) \text{ and} \quad (IV.7)$$

$$q(t) = \varepsilon(t) - e(t) + (1/2)\phi^T(t)\tilde{\theta}(t) \quad (IV.8)$$

Introducing

$$\Delta V(t) = V(t) + [2/\Lambda(t)] \sum_{k=1}^t p(k)q(k) \quad (IV.9)$$

and noting that $V^*(t) = \Lambda(t)V(t)$,

$$E[\Delta V(t)|F(t-1)] \leq \Delta V(t-1) - [1/\lambda(t)]\Delta V(t-1) + \gamma(t) \quad (\text{IV.10})$$

where

$$\gamma(t) = 2\phi^T(t)P(t)\phi(t)[(1-u)\sigma_1^2 + u\sigma_2^2] \leq 0 \quad (\text{IV.11})$$

Now we claim

$$\sum_{t=1}^{\infty} \gamma(t) < \infty \text{ only if } 0 < \sigma_1^2, \sigma_2^2 < \infty$$

The proof can be as follows (similar to the proof of Lemma 4.3):

$$\begin{aligned} & \phi^T(t)[P(t)]^2\phi(t) \\ & \leq \phi^T(t)P(t)P(t-1)\phi(t) \\ & = \text{trace}[P(t)\phi(t)\phi^T(t)P(t-1)] \\ & = [\text{trace}P(t-1) - \text{trace}P(t)]d(t) \end{aligned} \quad (\text{IV.13})$$

$$\text{where } d(t) = [1 - \delta(t)]\sigma_1^2 + \delta(t)\sigma_2^2.$$

Now from the above steps it follows that

$$\sum_{t=1}^{\infty} \phi^T(t)[P(t)]^2\phi(t) \leq \max(\sigma_1^2, \sigma_2^2)[\text{trace } P(0) - P(\infty)]$$

$$< \infty \text{ only if } 0 < \sigma_1^2, \sigma_2^2 < \infty.$$

Now

$$\sum_{t=1}^{\infty} \gamma(t) \leq 2\max(\sigma_1^2, \sigma_2^2) \sum_{t=1}^{\infty} \phi^T(t)P(t)\phi(t)$$

But

$$\phi^T(t)P(t)\phi(t) \leq \|P^{-1}(t)\| \phi^T(t)[P(t)]^2\phi(t)$$

(where $\|\cdot\|$ is the operator norm)

which shows that

$$\sum_{t=1}^{\infty} \phi^T(t) P(t-1) \phi(t) = \sum_{t=1}^{\infty} \text{trace} [P^{-1}(t)] \phi^T(t) [P(t)]^2 \phi(t)$$

and the result is immediate.

Our next aim is to show the conditions in which $\Delta V(t)$ will be non negative. We consider two distinct cases.

Case (i) : The observability canonical form

Let $U(t-1)$ and $\hat{E}(t-1)$ be as defined by (15.3.15d) and (15.3.15e) respectively. In addition, define

$$E(t-1) = [e(t-1), e(t-2), \dots, e(t-n)] \quad (\text{IV.14})$$

we have $\varepsilon(t) = \phi^T(t) [\theta_0 - \theta] + [\hat{\phi}_0^T(t) - \phi_0^T(t)] \theta_0 + e(t)$

(the subscript 0 means 'true value')

$$\begin{aligned} &= \hat{x}^T(t-n) [\bar{a}_0 - \bar{a}] + U(t-1) [\bar{\beta}_0 - \bar{\beta}] \\ &+ \hat{E}(t-1) [\bar{\xi}_0 - \bar{\xi}] + [x^T(t-n) - \hat{x}^T(t-n)] \bar{a}_0 \\ &\quad + [\hat{E}(t-1) - E(t-1)] \bar{\xi}_0 + e(t) \quad (\text{IV.15}) \end{aligned}$$

$[\bar{a}_0, \bar{a}, \bar{\beta}_0, \bar{\beta}, \bar{\xi}_0$ and $\bar{\xi}]$ are vectors and from the canonical structure (6.3.9)]

After some straightforward simplifications we get the following relationships.

$$C(z^{-1}) [\varepsilon(t) - e(t)] = \hat{x}^T(t-n) [\bar{a}_0 - \bar{a}]$$

$$\begin{aligned} &= \sum_{i=1}^n [\beta_{0,i} - \beta_i] \lambda_i(z^{-1}) u(t-i) + \sum_{i=1}^n [\xi_{0,i} - \xi_i] \lambda_i(z^{-1}) \varepsilon(t-i) \quad (\text{IV.16}) \end{aligned}$$

where $C(z^{-1})$ is as in (6.3.5) and

$$\lambda_1(z^{-1}) = 1 - \sum_{j=1}^{n-1} a_j z^{-j} \quad (\text{IV.17})$$

a_j 's are given by (6.3.5).

$$\text{Thus we can write } \varepsilon(t) = -\phi^T(t)H(z^{-1})(\theta - \theta_0) + e(t)$$

where

$$H(z^{-1}) = [1/C(z^{-1})] \begin{bmatrix} 1 & 0 & 0 \\ 0 & D & 0 \\ 0 & 0 & D \end{bmatrix} \quad (\text{IV.18})$$

where

$$D = \begin{bmatrix} \lambda_1(z^{-1}) & & & \\ & \lambda_2(z^{-1}) & & \\ & & & 0 \\ 0 & & & \lambda_{n-1}(z^{-1}) \\ & & & & 1 \end{bmatrix}$$

From (IV.17) and (IV.18)

$$q(t) = [H(z^{-1}) - (1/2)I]p(t).$$

Now from (IV.8) it follows immediately that

$$\Delta V(t) > 0 \text{ if } \text{Real}[H(z^{-1}) - (1/2)I] > 0.$$

Case(ii) : The observer canonical form

Following exactly as before (but simplifications are easier here!), we get

$$C(z^{-1})[\varepsilon(t) - e(t)] = \sum_{i=1}^n [\bar{a}_{0,i} - a_i]y(t-i)$$

$$+ \sum_{i=1}^n [b_{0,i} - b_i] u(t-i) + \sum_{i=1}^n [k_{0,i} - k_i] \varepsilon(t-i) \quad (\text{IV.19})$$

Hence $H(z^{-1}) = [1/C(z^{-1})]$

i.e., $\Delta V(t) > 0$, if $\text{Real}[1/C(z^{-1}) - (1/2)] > 0$.

which is the same causality condition for the convergence of the extended least squares algorithm.

Thus we have shown that in (IV.10)

$$\sum_{t=1}^{\infty} \Delta V(t) < \infty \text{ and } \Delta V(t) > 0 \quad \forall t \in [0, \infty)$$

subjected to certain conditions derived. By applying Neveu's theorem in (IV.10) we get

$$\sum_{t=1}^{\infty} [1/\Lambda(t)] \Delta V(t-1) < \infty \text{ w.p.1}$$

hence

$$\lim_{t \rightarrow \infty} [1/\Lambda(t)] \Delta V(t-1) = 0 \text{ w.p.1.}$$

As $\Lambda(t) < \infty \quad \forall t \in [0, \infty)$ if $0 < \sigma_1^2, \sigma_2^2 < \infty$ and

$$\limsup_{t \rightarrow \infty} [1/t] \sum_{n=1}^t |\phi(n)|^2 < \infty.$$

Hence it follows that

$$\lim_{t \rightarrow \infty} \Delta V(t) = 0 \text{ w.p.1.}$$

Therefore it immediately follows from (IV.8) that

$$\lim_{t \rightarrow \infty} V(t) = 0 \quad \text{w.p.1.}$$

from which it is clear that

$$\lim_{t \rightarrow \infty} \hat{\theta}(t) = \theta_0 \quad \text{w.p.1.}$$

and the proof is complete.

APPENDIX V A

FIXED POINT ITERATION SCHEME OF SINHA AND LASTMAN

The fixed point iteration of Sinha and Lastman (1981) can be briefly put as follows.

$$\begin{aligned} (AT)^{(k+1)} &= (AT)^{(k)} + F^{-1}[F - F^{(k)}] \\ &= (AT)^{(k)} + I + F^{-1}F^{(k)} \end{aligned} \tag{V.A.1}$$

where $(AT)^{(k)}$ is the value of AT at the k-th iteration and

$$F^{(k)} = \text{Exp}((AT)^{(k)}) = I + (AT)^{(k)} + (1/2!)[(AT)^{(k)}]^2 + \dots \tag{V.A.2}$$

This will work well provided

(a) if T is chosen such that the spectral radius of $AT < 1$, which will be fulfilled if T satisfies the condition given by eqn. (7.2.7) and

(b) the initial value $(AT)^{(0)}$ is good. Sinha and Lastman suggest $(AT)^{(0)} = (1/2)[F - F^{-1}]$ based on truncated Taylor series expansion of $\text{Exp}(AT)$ and $\text{Exp}(-AT)$ after three terms or $(AT)^{(0)} = 2(F + I)^{-1}(F + I)$ based on the Pade' approximation of $\text{Exp}(AT)$. Once AT is obtained by using the above algorithm, B can be calculated as

$$B = R^{-1}G$$

where $R = [I + (1/2!)(AT) + (1/3!)(AT)^2 + \dots]T$.

APPENDIX V B

THE DIRECT APPROACH OF HUNG ET.AL.

An algebraic approach using the trapezoidal rule of numerical integration is given by Hung et.al. (1980). It can be considered as a combination of the block pulse technique and the inverse-trapezoidal rule (Sinha and Lastman, 1982).

From the discrete-time state equations we obtain

$$x(k+1) = x(k) + A \int_{kT}^{kT+T} x(t)dt + B \int_{kT}^{kT+T} u(t)dt \quad (V.B.1)$$

using trapezoidal rule of integration

$$x(k+1) = x(k) + (1/2)AT[x(k+1) + x(k)] + (1/2)BT[u(k+1) + u(k)] \quad (V.B.2)$$

which after some rearrangements give

$$A[x(k+1) + x(k)] + B[u(k+1) + u(k)] = (2/T)[x(k+1) - x(k)] \quad (V.B.2)$$

Using samples of the input and the states we can obtain least squares estimates of the parameters of the matrix A and B. However it should be emphasized that this method requires the states of the system and is not very simple computationally .

REFERENCES

Abraham and Box (1975), "Outliers in Time Series", Technical Report No.440, Dept. of Statistics, University of Wisconsin.

W.S.Agee and R.H.Turner (1979), "Application of Robust Regression to Trajectory Data Reduction", in Robustness in Statistics, Edited by R.L.Launer and G.N.Wilkinson, Academic Press, New York.

M.S.Ahmed (1983), "On Bootstrap Estimation of Parameters And States", IEEE Trans. Aut. Cont. Vol. 28, pp. 805-806.

D.F.Andrews, P.J.Bickel, F.R.Hampel, P.J.Huber, W.H.Rogers and J.W.Tukey (1972), "Robust Estimates of Location. Survey and Advances", Princeton University Press, Princeton.

F.J.Anscombe (1960), "Rejection of Outliers", Technometrics, Vol.. 2, pp. 123-147.

K.J.Astrom (1980), "Maximum Likelihood and Prediction Error Method", Automatica, vol.16, pp. 551-574.

K.J.Astrom and P.Eykhoff (1971), "System Identification - A Survey", Automatica, Vol. 7, pp. 123-162.

N.Balakrishnan and S.Puthenpura (1985a) "Best Linear Unbiased Estimators of Location and Scale Parameters of the

"Half Logistic Distribution", to appear in the Journal of Statistical Computations and Simulations.

N.Balakrishnan and S.Puthenpura (1985b). "BLUE of Location and Scale Parameters of Log Logistic Distribution from Order Statistics", (under preparation).

S.Basu and D.Vandelinde (1977), "Robust Identification of Parameters of a Linear System", Proc. 15th Ann. Allerton Conf. on Comm. Comp. and Cont. (Monticello), pp. 221-230.

P.J.Bickel (1965), "On Some Robust Estimation of Location", Ann. Math. Stat., Vol. 36, pp. 847-858.

H.J.Bierens (1981), "Robust Methods and Asymptotic Theory in Non Linear Econometrics", Springer Verlag, New York.

C.C.Bonchelet, Jr. and B.W.Dickinson (1983), "An Approach to Robust Kalman Filtering", Proc. 22nd IEEE Conf. Decision Control (San Antonio, Tx.), pp. 304-305.

G.E.P.Box (1953), "Non Normality and Test on Variances", Biometrika, Vol. 40, pp. 318 - 335.

G.E.P.Box and G.M.Jenkins (1970), "Time Series Analysis Forecasting and Control", Holden-Day, Sanfrancisco.

H.Chernoff and I.R.Savage (1958), "Asymptotic Normality and Efficiency of Certain Non Parametric Test Statistics", Ann. Math. Stat., Vol. 29, pp. 872-914.

H.Chernoff, J.L.Gastwirth and W.V.John (1967), "Asymptotic Distributions of Order Statistics with Applications to Estimation", Ann. Math. Stat., Vol: 40, pp. 52-72.

L.J.Cimini and S.A.Kassam (1981), "Optimum Piecewise Constant Weiner Filtering", J. Opt. Soc. Amer., Vol. 77, pp. 1162-1171.

J.R.Collins (1976); "Robust Estimation of a Location Parameter in Presence of Asymmetry", Ann. Stat., Vol.4, pp. 68-85.

R.D.Cook and C.Prescott (1981), "On the Application of Bonferroni Significance Levels for Detecting Outliers in Linear Models", Technometrics, Vol. 22, pp. 59-63.

M.Cuenod and A.P.Sage (1968), "Comparison of Some Method Used for System Identification", Automatica, Vol.4, pp. 235-269.

C.Daniel (1920), "Observation Weighted According to Order", American Journal of Mathematics, Vol. 42, pp. 222-236.

C.Daniel (1976), "Application of Statistics to Industrial Experiments", John Wiley, New York.

H.A.David, (1981), "Order Statistics" (second edition), John Wiley, New York.

R.C.Desai and C.S.Lalwani (1972). "Identification Techniques", Tata McGraw-Hill, Bombay.

R.Deutsch (1965), "Estimation Theory", Prentice Hall, New Jersey.

W.J.Dixon (1953), "Processing Data for Outliers", Biometrika Vol. 9, pp. 74-89.

W.J.Dixon (1960), "Simplified Mean for Censored Order Samples", Ann. Math. Stat., Vol. 3, pp. 385-391.

T.Eden and F.Yates (1933), "On the Validity of Fisher's Z-Test when Applied to an Actual Example of Non Normal Data", Biometrika, J. Agricultural Sciences, Vol. 23, pp. 6-16.

F.Y.Edgeworth (1886), "Problems in Probabilities", Philosophical Magazine, Vol. 22, pp. 371-384.

B.Effron (1979), "Bootstrap Methods, Another Look at the Jack knife", Ann. Stat., Vol. 7, pp. 1-26.

B.Effron (1980), "The Jack knife, the Bootstrap and the Resampling Plans", Tech. report No. 163, Stanford University.

C.Eisenhart (1971), "Mendeleev's expressed preference of the $33 \frac{1}{3} \%$ trimmed mean", National Bureau of Standards, Washington D.C.

H.El-Sherief and N.K.Sinha (1979a), "Identification and Modeling of Linear Multivariable Systems - A Survey", J. of Cybernetics, Vol. 9, pp. 43-71.

H.El-Sherief and N.K.Sinha (1979b) "Bootstrap Estimation of Parameters and States of a Linear System", IEEE Trans. Aut. Control, Vol. 24, pp. 340-343.

H.El-Sherief (1979), "Identification of Linear Multivariable Discrete-time Systems", Ph.D. thesis, McMaster University.

P.Eykhoff (1968), "Process Parameter and State Estimation", Automatica, Vol. 4, pp. 205-233.

P.Eykhoff (1974), "System Identification", John Wiley, London.

G.F.Franklin and J.D.Powell (1980), "Digital Control of Dynamic Systems", Addison Wesley Inc.

J.L.Gastwirth and M.L.Cohen (1970), "Small Sample behaviour of Some Robust Linear Estimates of Location", J. of Am. Stat. Assn., Vol. 65, pp. 946-973.

H.J.Godwin (1949), "On the Estimation of Dispersion by Linear Systematic Statistics.", Biometrika, Vol. 39, pp. 88-95.

G.C.Goodwin and K.S.Sin (1984), "Adaptive Filtering Prediction and Control", Prentice Hall, New Jersey.

D.Graupe (1972). "Identification of Systems", Van Nostrand Reinhold, New York.

I.Gustavsson (1972), "Comparison of Different Methods for Identification of Industrial Processes", Automatica, Vol. 8, pp. 127-142.

Ha'jeck (1968), "Asymptotic Normality of Simple Linear Rank Statistics Under Alternatives", Ann. Math. Stat., Vol. 34, pp. 325-346.

F.R.Hampel (1971), "A General Qualitative Definition for Robustness", Ann. Math. Stat., vol.42, pp. 1887-1896.

H.L.Harter and A.H.Moore (1966), "Iterative Maximum Likelihood Estimation of the Parameters of Normal Populations from Singly or Doubly Censored Samples", Biometrika, Vol. 53, pp. 205-213.

S.S.Haykin (1972), "A Unified Treatment in Recursive Digital Filtering", IEEE Trans. Aut. Cont., Vol. 17, pp. 113-116.

J.L.Hodges and E.L.Lehman (1963), "Estimates of Location Based on Rank Tests", Ann. Math. Stat., Vol 34, pp. 598-611.

J.L.Hodges (1967), "Efficiency of Normal Samples and Tolerance of Extreme Values for the Estimation of Location", Proc. 5th Berkeley Symp. on Math. and Stat. Problems, University of California (Berkeley) Press, pp. 163-186.

R.V.Hogg (1974), "Adaptive Robust Procedures", J. Am. Stat. Assn., Vol 69, pp. 909-927.

R.V.Hogg (1979), "An Introduction to Robust Estimation", in

"Robustness in Statistics", edited by R.L.Launer and G.Wilkinson Academic press, New York.

P.J.Huber (1981), "Robust Statistics", John Wiley, New York.

P.J.Huber (1964), "Robust Estimation of a Location Parameter", Ann. Math. Stat., vol.35, pp. 73-101.

J.C.Hung, C.C.Liu and P.Y.Chan, (1980), "An Algebraic Method for System Parameter Identification", Proc. 14th Asilomar Conf. on Cir. Sys. and Comp. (Houghton, Michigan), pp. 325-328.

R.Iserman, U.Baur, W.Bamberger, P.Kneppo and H.Siebert (1974), "Comparison of Six On-line Identification and Parameter Estimation Methods", Automatica, Vol. 10, pp. 81-103.

A.H.Jaswinski (1970), "Stochastic Processes and Filtering Theory", Academic Press, New York.

J.Jureckova (1977), "Asymptotic Relations of M-Estimates and R-Estimates in Linear Regression Models", Ann. Stat., Vol. 5, pp. 464-472.

A.Jutan, J.P.Tremblay, J.F.MacGregor and J.D.Wright (1977), "Multivariable Control of a Butane Hydrogenolysis Reactor", AI. Che.J., vol.23, pp. 732-742.

A.M.Karnik (1985), "Adaptive Control of an Industrial

Robot", Ph.D. thesis, McMaster University.

S.A.Kassam and T.L.Lim (1977), "Robust Weiner Filters", J. Frank. Inst. Vol. 304, pp. 171-185.

S.A.Kassam and V.H.Poor (1985), "Robust Techniques in Signal Processing, a Survey", Proc. IEEE, vol.73, No.3, pp. 433-484.

B.Kleiner, R.D.Martin and D.J.Thomson (1979), "Robust Estimation of Power Spectra", J. Roy. Stat. Soc., Vol. 41, pp. 313-351.

B.Kovacevic and S.Stancovic (1979), "Comparative Analysis of a Class of Robust Real-time Identification Methods", Proc. 5th IFAC Symp. Ident. Param. Estim., (Darmstadt), pp. 763-770.

B.Kovacevic and S.Stancovic (1982), "Robust Real-time Identification of Multivariable Methods", Proc. 6th IFAC Symp. Ident. Par. Est. (Washington), pp. 935-940.

B.Kovacevic and S.Stancovic (1984), "Approaches to Robust Real-time Identification of Stochastic Systems", Proc. 9th IFAC Cong. Ident. Sys. Par. Est. (Budapest), Vol. X, pp. 84-89.

B.Kovacevic and V.Z.Filipovic (1985), "Robust Real-time Identification of Linear Time Varying Dynamic Systems", Proc. 7th IFAC Symp. Ident. Par. Est. (York), pp. 1467-1471.

P.S.Laplace (1818), "Deuxieme Supplement a la Theorie

Analytique des Probabilités", in Oeuvres de Laplace, 7, Gauthier-Villars, Paris (published in 1886).

G.J.Lastman, S.Puthenpura and N.K.Sinha (1984). "Algorithm for the Identification of Continuous-time Models from their Discrete-time Models", Electronics Letters, 1984, Vol.22, No.22, pp. 918-919.

R.L.Launer and G.Wilkinson (editors) (1978), "Robustness in Statistics", Academic press, New York.

A.M.Legendre (1805), "On the Method of Least Squares", in "A Source Book in Mathematics" (edited by D.E.Smith) (1959), Dover, New York. (Original article in French.)

F.C.Leone, T.Jayachandran and Eisenstat (1967), "A Study of Robust Estimation", Technometrics, Vol. 9, pp. 652-660.

L.Ljung and T.Soderstrom (1983), "Theory and Practice of Recursive Identification", MIT Press, Massachusetts.

E.H.Lloyd (1952), "Least Squares Estimation of Location and Scale Parameters Using Linear Combinations of Order Statistics", Biometrika, Vol. 39, pp. 88-95.

M.Loe've (1977), "Probability Theory" (4th edition), Springer Verlag, New York.

E.Lukacs (1975), "Stochastic convergence" (second edition), Academic press, New York.

P.C.Mahalanobis (1936)., "On the Generalized Distance in Statistics", Proc. National Institute of Sciences, India, pp. 49-55.

R.D.Martin (1972), "Robust Estimation of Signal Amplitude", IEEE Trans. Inf. Theory, Vol. 18, pp. 596-602.

R.D.Martin (1979), "Robust Estimation for Time Series Autoregressions", in Robustness in Statistics, Edited by R.L.Launer and G.N.Wilkinson, Academic Press, New York.

R.D.Martin and C.J.Masreliez (1975), "Robust Estimation via stochastic Approximation", IEEE Trans. Inf. Theory, Vol. 21, pp. 263-271.

R.D.Martin and D.J.Thomson (1982), "Robust-resistant Spectrum Estimation", Proc.IEEE, Vol. 70, pp. 1097-1115.

C.J. Masreliez and R.D.Martin (1978), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter", IEEE Trans. Aut. Control, vol.22, pp. 361-371.

R.K.Mehra and D.G.Lainiotis (1976), "System Identification - Advances and Case Studies", Academic Press, New York.

J.M.Mendel (1973), "Discrete Techniques for Parameter Estimation: the Equation Error Approach". Marcell Dekker Inc. New York.

J.M.Morris (1976), "The Kalman Filter: A Robust Estimator

for Some Quadratic and Estimation and Control Problems", IEEE Trans. Inf. Theory, Vol. 22, pp. 526-534.

G.Moustakides and S.A.Kassam (1983), "Robust Weiner Filters for Random Signals and Correlated Noise", IEEE Trans. Inf. Theory, Vol. 29, pp. 614-619.

J.Neveu (1975), "Discrete Parameter Martingales", North Holland Company, Amsterdam.

S.Newcomb (1886), "A generalized Theory of the Combination of Observations so as to Obtain the Best Results", American J. of Math., Vol. 8, pp. 343-366.

R.E.Neiman, D.G.Fisher and D.E.Seborg (1971), "A Review of Process Identification and Parameter Estimation Techniques", Int. J. Control, pp. 209-264.

C.L.Olson (1974), "Comparative Robustness of Six Tests in Multivariate Analysis of Variance", J. Am. Stat. Assn., Vol. 69, pp. 894-908.

F.Omani and N.K.Sinha (1985), "A Modified Approach to State Space Self-tuning Control with Pole Assignment", accepted for publication in the IEE Proceedings.

P.Papapantoni-Kazakoz (1977), "Robustness in Parameter Estimation", IEEE Trans. Inf. Theory, Vol. 23, pp. 223-231.

E.S.Pearson (1931), "The Analysis of Variance in the Case of Non Normal Variations", Biometrika, Vol. 20, pp. 114-133.

C.S.Pillai and C.Sudjana (1975), "Exact Robustness Studies of Tests of Two Multivariate Hypothesis Based on Four Criteria and Their Distribution Problems Under Violations", Ann. Stat., Vol. 3, pp. 617-636.

B.Pierce (1852), "Criteria for Rejecting Doubtful Observations" Astronomical Journal, Vol. 2, pp.161-163.

H.Poincare' (1912), "Calculus des Probabilites", Gauthier-Villars, Paris.

B.T.Poljak and Ja.Z. Tsytkin (1980), "Robust Identification", Automatica, vol.16, pp. 53-63.

H.V.Poor (1980)," On Robust. Weiner Filtering", IEEE Trans. Aut. Control, Vol. 25, pp. 531-536.

R.M.Prasad and A.K.Sinha (1977), "On Bootstrap Identification Using Stochastic Approximation", IEEE. Trans. Aut. Cont. Vol. 23, pp. 671-672.

T.Prasad and N.K.Sinha (1983), "Modeling of Continuous-time Systems from Sampled Data Using Trapezoidal Pulse Functions", Proc. IEEE Conf. Sys. Man and Cyb. (Bombay, India), pp. 427-430.

E.L.Price and V.D.Vandelinde (1979), "Robust Estimation Using Robbin Monro Stochastic Approximation Algorithm", IEEE Trans. Inf. Theory, Vol. 25, pp. 698-704.

M.L.Puri and P.K.Sen (1985), "Non Parametric Methods in General Linear Models", John Wiley, New York.

S.Puthenpura and N.K.Sinha (1984a). "On the Instrumental Variables Method", Proc. Twenty-second Annual Allerton Conference on Communication, Control and Computing, Univ. of Illinois, Urbana Champaign, , pp. 613-614

S.Puthenpura and N.K.Sinha (1984b). "Transformation of Continuous-time Model of a Linear Multivariable System from its Discrete-time model", Electronics Letters, , Vol.20, No.18, pp. 737-738.

S.Puthenpura and N.K.Sinha (1985a) "Modified Maximum Likelihood Method for the Robust Estimation of System Parameters from Very Noisy Data", accepted for publication in Automatica.

S.Puthenpura and N.K.Sinha (1985b). "Robust Instrumental Variables Method for System Identification", accepted for publication in Control Theory and Advanced Technology.

S.Puthenpura and N.K.Sinha (1985c) "Identification of Continuous-time Systems Using Instrumental Variables with Application to an Industrial Robot", accepted for publication in IEEE Transactions on Industrial Electronics.

S.Puthenpura and N.K.Sinha (1985d), "Robust Identification from Impulse and Step Responses" (under preparation).

S.Puthenpura and N.K.Sinha (1985e). "A Robust Recursive Identification Algorithm", under consideration for publication.

S.Puthenpura and N.K.Sinha (1985f) "Robust Bootstrap Method for the Joint Estimation of States and Parameters of a Linear System", Proc. of the annual winter meeting (Florida, November) of ASME on Dynamic Systems Modelling and Control, pp. 175-184. (Also under consideration for publication in the ASME Journal of Dynamic Systems Measurement and Control).

S.Puthenpura and N.K.Sinha (1985g). "A Procedure for Determining the Optimal Sampling Interval for System Identification Using a Digital Computer", Canadian Electrical Engineering Journal, Vol.10, No.4, pp. 152-157.

S.Puthenpura, N.K.Sinha and O.P.Vidal (1985) "Application of M-Estimation in Robust Recursive System Identification", accepted for presentation at the IFAC symp. on stochastic control (Vilnius, U.S.S.R., May, 1986). (Also under consideration for publication in Automatica).

U.V.Quinouille (1956), "Notes on Bias in Estimation", Biometrika, Vol. 43, pp. 353-360.

N.S.Rajbman (1976), "Application of Identification Methods in U.S.S.R. - a Survey", Automatica, Vol. 12, pp. 73-95.

N.S.Rajbman and N.K.Sinha (1977), "Identification of Multivariable Systems: A Critical Review", Int. J. Sys. Sc. Vol. 8, pp. 1415-1427.

J.R.Raol (1985), "Stochastic State Estimation with Application to Satellite Orbit Determination", Ph.D. thesis, McMaster University (under preparation).

J.R.Raol, S.Puthenpura and N.K.Sinha (1985). "An Efficient Algorithm for Transformation of Multivariable Discrete-time Models to Continuous-time Models", "accepted for the AMSE International Conference on Modelling and Simulation, Gorakhpur(Dec. 22-24, India).]

W.J.J.Rey (1983), "Robust and Quasi-Robust Statistical Methods", Springer Verlag, New York.

T.J.Rivlin (1974), "The Chebyshev Polynomials", John Wiley, New York.

H.Robbins and S.Monro (1951), "A Stochastic Approximation Method", Ann. Math. Stat., Vol. 20, pp. 400-407.

A.P.Sage and J.L.Melsa (1972). "System Identification", Academic Press, New York.

A.E.Sarhan and B.G.Greenberg (1962), "Contribution to Order Statistics", John Wiley, New York.

G.N.Saridis (1974a), "A Comparison of Six On-line Identification Algorithms", Automatica, Vol. 10, pp. 69-79.

G.N.Saridis (1974b), "Stochastic Approximation Methods for Identification and Control - A Survey", IEEE Trans. Aut. Cont., Vol. 19, pp. 798-809.

N.K.Sinha and B.Kusztá (1983), "Modeling and Identification of Dynamic Systems", Van Nostrand Reinhold Company, New York.

N.K.Sinha and G.J.Lastman (1981), "Transformation Algorithm for Identification of Continuous-time Multivariable Systems from Discrete Data", Electronic Letters, Vol. 21, pp. 779-780

N.K.Sinha and G.J.Lastman (1982), "Identification of Continuous-time Systems from Sampled Data", Int. J. Control, vol.35, No.1, pp. 117-126.

N.K.Sinha, G.J.Lastman and Zhou Qi-Jie (1982), "On the Choice of Sampling Interval for the Identification from Sampled Data", Proc. 25th Midwest. Symp. on Cir. and Sys., pp. 329-332.

N.K.Sinha and A.Sen (1975), "A Critical Evaluation of On-line Identification Methods", Proc. IEE, Vol. 122, pp. 1153-1158.

N.K.Sinha and S.Puthenpura (1985). "Choice of the Sampling Interval for the Identification of Continuous-time Systems from Samples of Input-Output Data", accepted for publication in IEE Proceedings on Control Theory and Applications.

N.K.Sinha (1977), "On-line Identification of Linear Multivariable Systems - A Survey and New Results", Proc. Int. Conf. on Inf. Sc. and Sys., (Patras, Greece), pp. 301-305.

N.K.Sinha and Zhou Qi-Jie (1984), "Identification of Continuous-time systems from Samples of Input-Output Data", Control Theory and Applications (Publ. Academia Sinica, China), Vol. 1, pp. 110-122.

T.Soderstrom and P.Stoica (1981), "Comparison of Some Instrumental Variables Methods - Consistency and Accuracy Aspects", Automatica, Vol. 17, pp. 101-116.

T.Soderstrom and P.Stoica (1983), "Instrumental Variables Method for System Identification", Springer Verlag, Berlin.

V.Solo (1979), "The Convergence of AML", IEEE Trans. Aut. Cont., Vol. 24, pp. 395-410.

H.W.Sorenson (1980), "Parameter Estimation - Principles and Problems", Marcel Dekker, New York.

D.Sprevak and M.M.Newman (1980), "Comments on Bootstrap Estimates of Parameters and States of Linear Multivariable Systems", IEEE Trans. Aut. Cont. Vol. 25, pp. 138-139.

S.M.Stigler (1974), "Linear Combination of Order statistics with Smooth Weight Functions", Annals of Stat., Vol. 2, pp.

676-693. [Correction: Vol. 7, page 466].

S.M. Stigler (1969), "Linear Combinations of Order Statistics", Ann. Math. Stat., Vol. 40, pp. 770-788.

W.Y. Tan (1985), "On Tiku's Robust Procedure - A Bayesian Insight", J. Stat. Planning and Inference, Vol. 11, pp. 329-340.

M.L. Tiku (1967), "Estimating the Mean and Standard Deviation from Censored Normal Samples", Biometrika, Vol. 54, pp. 155-165.

M.L. Tiku (1971), "Student's t distribution Under Non Normal Situations", Australian J. Statistics, Vol. 13, pp. 142-148.

M.L. Tiku (1978), "Linear Regression Model with Censored Observations", Commun. in Stat. Theory. Meth. A7(13), pp. 1219-1232.

M.L. Tiku (1980), "Robustness of MML Estimators Based on Censored Samples and Robust Test Statistics", J. Stat. Plann. Inf., pp. 123-143.

M.L. Tiku (1981), "Robust Two Sample Test and Robust Regression and Analysis of Variance via MML Estimates" in Statistics and Related Topics, published by North Holland Company, pp. 297-314.

M.L. Tiku (1982), "Robust Statistics for Testing the Equality of Means or Variances", Comm. Stat. Th. and Methods, Vol. 11

(Special issue edited by M.L.Tiku), pp. 2543-2558.

C.Tsai and L.Kurz (1983), "Adaptive Robustizing Approach to Kalman Filtering", Automatica, Vol 18, No.3, pp. 279-288.

J.W.Tukey (1960), "A Survey of Sampling from Contaminated Distributions", in "Contributions to Probability and Statistics", edited by I.Olkin et.al., Stanford University Press, pp. 433-438.

J.W.Tukey and D.H.McLaughlin (1963), "Less Vulnerable Confidence Intervals and Significance Procedures Based on a Single Sample", Sankhya, Vol 25, pp. 331-352.

J.W.Tukey (1967), "The Future of Data Analysis", Ann. Math. Stat., Vol 33, pp. 1-67.

K.V.Vastola and H.V.Poor (1984), "Robust Wiener-Kolmogorov Theory", IEEE Trans. Inf. Theory, Vol 30, pp. 316-327.

K.V.Vastola and H.V.Poor (1983), "An Analysis of the Effect of Spectral Uncertainty on Wiener Filtering", Automatica, Vol 28, pp. 289-293.

H.D.Vinod and A.Ullah (1981), "Recent Advances in Regression Models", Marcel Dekker, New York.

P.C. Young (1970), "An Instrumental Variables Method for Real-time Identification of a Noisy Process", Automatica, Vol. 6, pp. 271-287.

P.C. Young (1984), "Recursive Estimation and Time-Series Analysis", Springer-Verlag, New York.

K.M. Wong (1986), "Sonar Data Normalization Using Order Statistical Filters", Report No.CRL 149, Communications Research Laboratory, McMaster University.

S. Zacks (1971), "The Theory of Statistical Inference", John Wiley, New York.