ℓ_1-NORM MINIMIZATION AND REDUCED ORDER MODELS IN MULTIVARIABLE CONTROL

J. Chau-Chavez

A THESIS
IN THE
FACULTY OF ENGINEERING

Presented in partial fulfillment of the requirements
for the degree of MASTER OF ENGINEERING at
Concordia University, Montreal, CANADA

December 1980

ℓ_1-NORM MINIMIZATION AND REDUCED ORDER MODELS IN MULTIVARIABLE CONTROL

J. Chau-Chavez

ABSTRACT

The ℓ_1-norm minimization problem is studied together with some of its applications to approximation and sub-optimal multi-input-output control systems design. The nondifferentiable unconstrained ℓ_1-problem is transformed to a sequence of differentiable problems with a dynamic scaling factor allowing a reduction in the number of iterations. The existing gradient methods can be used to solve each of the new minimizations. It is shown that, under mild conditions every limit point of the sequence of minimums is a solution of the ℓ_1-problem. The numerical study conducted has shown that the proposed method is numerically stable and robust.

A new method for obtaining reduced order models for multi-input-output strictly proper and proper linear time invariant systems using the ℓ_1-norm minimization is thoroughly studied. This procedure for obtaining order reduction ensures stable meaningful reduced order models for stable high-order systems. It is shown that using the proposed method for order reduction of linear time invariant systems together with a dissagregation scheme yields new procedures to obtain a sub-optimal Wiener-Kalman Filter and the order reduction of a class of linear time variant systems.

ACKNOWLEDGEMENTS

I would like to thank Dr. M. Vidyasagar for proposing the topic of this thesis and for his financial support.

Also I want to express my thanks to two special people, Mr. Aaron Pila and Miss Betty Liebermann. Aaron Pila assisted with proofreading and gave helpful comments. Without the moral support and untiring encouragement given to me by Betty and Aaron, this work could not have been finished. I am indebted to them for their support.

I also express my thanks to Mrs. Marie Berryman for her diligent typing of this thesis.

LIST OF TABLES

LIST OF FIGURES

TABLE OF CONTENTS

DEDICO ESTA TESIS A MI QUERIDA MADRE

Sra. JUANA CHAVEZ de CHAU

CHAPTER 1
INTRODUCTION
1.0 MOTIVATION

In this thesis we are concerned with the study of the ℓ_1-norm minimization problem, specifically in context of ℓ_1-approximations and some of its applications to multivariable control system design. The raison d'etre of the minimization of cost functions of the ℓ_1-norm type is based on the well known [9] fact that the best ℓ_1-approximations are often superior to best ℓ_p-approximations where the observations contain wild points.

It is also well known (Sinha and Titli [46]) that the computational methods of optimal control theory, ~~for example~~, for linear dynamic systems which use quadratic type cost functions run into numerical difficulties when their order is greater than about ten. This fact makes it attractive to consider instead, the solution of the sub-optimal problem, i.e. where the high dimensional system is replaced by one of lower dimensionality and the optimal control policies calculated for the lower order system are used by the higher order one. However since lower dimensional systems cannot be arbitrarily chosen, a systematic procedure for their selection can be achieved through approximation theory, specifically "reduced order modelling", which is a branch of approximation analysis.

Since, as previously mentioned, ℓ_1-norm approximations yield better results than their ℓ_p-norm counterparts, the reduced order model problem is therefore studied from the ℓ_1-norm minimization point of view. Due to its importance in control theory a major part of this thesis is devoted to a study of the R.O.M. and its applications to control system design and estimation, specifically the linear output regulator problem, filtering problem and the modelling of linear time varying systems. The R.O.M. problem can be formulated as a nonlinear ℓ_1-norm minimization problem and therefore part of this thesis is dedicated to the study of an efficient computational method for dealing with its solution.

1.1 ORGANIZATION OF THE THESIS

Chapter 1 contains the motivating factors for this research and the outline of the thesis. In Chapter 2, an analysis of the ℓ_1-norm minimization and nonlinear ℓ_1-norm minimization problems are presented. A brief review of some existing techniques for the determination of the best ℓ_1-approximation is given together with some remarks and comments. The transformation of the unconstrained ℓ_1-minimization problem into a sequence of problems, each involving the optimization of a continuous differentiable function, due to El Attar et al. [21], is reviewed in some detail. Based on the abovementioned approach, an algorithm which deals with the ℓ_1-minimization problem, is presented, and an iterative procedure which implements the proposed algorithm is thoroughly discussed. Also it is shown that, under mild conditions, this algorithm converges to the solution of the unconstrained ℓ_1 problem. The efficiency of the method is illustrated through several numerical examples. A comparison between the S.U.M. algorithm and the proposed one, is exhibited when both are used in solving some ℓ_1-approximation problems.

In Chapter 3, the reduced order modelling problem is formulated and studied as a minimization of the induced norm of the input-output map for the error system. Due to the fact that, in general, it is not possible to minimize the induced norm of this mapping, an alternative procedure is proposed for obtaining R.O.M.'s for multi-input, multi-output strictly proper, as well as proper systems. It is shown that this method yields stable reduced order models, whenever the original system is stable. Several numerical examples illustrate the proposed technique.

Chapter 4 deals with the application of the reduced order model in obtaining sub-optimal control policies for the output regulator problem. With the use of the disaggregation scheme [2], a method is proposed for obtaining a sub-optimal Weiner Kalman Filter. An example is given which illustrates the performance of this method, and it is apparent that its numerical behaviour is comparable to that of the optimal W.K.F. Finally, a procedure is presented whereby with the use of Wu [52] and Rao [37] transformations and the methods presented in the previous chapter for obtaining R.O.M.'s for linear time invariant systems, a reduced order model for a class of linear time varying systems is obtained. The iterative procedure for this technique is given in detail.

The last chapter contains concluding remarks and possible areas for further investigation. Throughout this investigation, the digital computer used to solve all the numerical examples is a CDC Cyber 173, and all programs were written in Fortran IV.

CHAPTER 2

ℓ_1-NORM MINIMIZATION AND NONLINEAR ℓ_1-APPROXIMATION

2.0 INTRODUCTION

Although the problem of minimizing an ℓ_1-norm type of objective function is not new, efficient solution techniques are available only in the linear case. On the other hand, little has been done for the corresponding nonlinear problem. This chapter is devoted to the study of the nonlinear problem.

The first two sections state the ℓ_1-norm minimization problem and its well known equivalent nonlinear programming problem. In sections 2.4 and 2.5 the nonlinear ℓ_1-approximation problem is stated and a brief review of some of the existing algorithms used to tackle it are presented. Following this the efficient algorithm proposed by El-Attar and associates [22], is presented in some detail. In sections 2.7 the new algorithm termed "Family of Unconstrained Minimizations" based on the sequential unconstrained minimization approach is proposed. Section 2.8 contains the iterative procedure of the algorithm prposed in the preceding section. Finally section 2.9 and 2.10 contain the numerical examples that illustrate the computation performance of the proposed method and the conclusions of this chapter respectively. Futhermore for the sake of illustration the results obtained by using S.U.M. method [22] are compared with the results obtained using the proposed approach for the same examples.

2.1 THE ℓ_p-PROBLEM

Consider real valued functions f_i, $i=1,\ldots,m$, continuously differentiable, with domain D in R^n. The ℓ_p-norm of the vector $\underset{\sim}{f} = \{f_1,\ldots,f_m\}$ is defined by

$$F_p(\underset{\sim}{x}) = \|f(\underset{\sim}{x})\|_p = \{\sum_{i=1}^{m} |f_i(\underset{\sim}{x})|^p\}^{\frac{1}{p}} \qquad (2.1)$$

where $1 \leq p \leq \infty$ and $\underset{\sim}{x} \in D$. Then the ℓ_p-norm minimization problem can be stated as follows:

Problem 2.1 With the f_i as defined above, for some p in $[1,\infty]$ minimize

$$F_p(\underset{\sim}{x}) = \|f(\underset{\sim}{x})\|_p . \qquad (2.2)$$

First consider this problem for $p > 1$; then the vector gradient of $F_p(\underset{\sim}{x})$ is

$$\nabla_{\underset{\sim}{x}} F_p(\underset{\sim}{x}) = \sum_{i=1}^{m} \left\{\frac{|f_i(\underset{\sim}{x})|}{\|f(\underset{\sim}{x})\|_p}\right\}^{p-1} \cdot \nabla_{\underset{\sim}{x}} f_i(\underset{\sim}{x}) \cdot \text{sign}\,(f_i(\underset{\sim}{x})), \qquad (2.3)$$

or in alternative form for $p \geq 2$

$$\nabla_{\underset{\sim}{x}} F_p(\underset{\sim}{x}) = \sum_{i=1}^{m} \frac{\{|f_i(\underset{\sim}{x})|\}^{p-2}}{\{\|f_i(\underset{\sim}{x})\|_p\}^{p-1}} \cdot f_i(\underset{\sim}{x}) \cdot \nabla_{\underset{\sim}{x}} f_i(\underset{\sim}{x}) \cdot \qquad (2.4)$$

Clearly the $\nabla_{\underset{\sim}{x}} F_p(\underset{\sim}{x})$ exists $\forall\, \underset{\sim}{x} \in D$. Now as p approaches 1, $F_p(\underset{\sim}{x})$ becomes nondifferentiable in general. Specifically, if $f_i(\underset{\sim}{x}') = 0$, for some i and some $\underset{\sim}{x}' \in D$ then $\nabla_{\underset{\sim}{x}} F_p(\underset{\sim}{x})$ does not exist. In general the

f_i's can be linear or nonlinear functions of $\underset{\sim}{x}$. The first case can be treated by linear programming, the second, by nonlinear programming but for $p = 1$, methods that do not use derivatives are the suitable ones. However it is a well known fact that methods using gradient techniques have better computational performance (see e.g. Avriel [7]).

2.2 THE ℓ_1-PROBLEM

The ℓ_1-problem can be stated as

$$\text{minimize} \quad F_1(\underset{\sim}{x}) = \sum_{i=1}^{m} |f_i(\underset{\sim}{x})| \, . \tag{2.5}$$

In order to overcome the nondifferentiability of $F_1(\underset{\sim}{x})$ it is known [38], that (2.5) can be reformulated as a general nonlinear programming problem (NPP) as follows:

Problem 2.2 Minimize $\psi : R^{n+m} \to R$

$$\psi = \sum_{i=1}^{m} \phi_i \, , \tag{2.6}$$

subject to

$$C_1: \begin{cases} c_i^1 : -f_i(\underset{\sim}{x}) + \phi_i \geq 0 \quad \forall i, & (2.7) \\ c_i^2 : f_i(\underset{\sim}{x}) + \phi_i \geq 0 \quad \forall i \, . & (2.8) \end{cases}$$

If $\underset{\sim}{\bar{x}} \in D$ and $\bar{\phi}$ solves problem 2.2 then we have $\bar{\phi}_i = |f_i(\bar{x})| \quad \forall i$.

2.3 OPTIMALITY CONDITIONS

The optimality condition for problem 2.2 can be easily derived from the Karush-Kuhn-Tucker conditions, [31], [32]. The necessary conditions are enunciated in the following lemma:

Lemma 2.1 (First order optimality conditions). If $\bar{x} \in D$ is a local minimizer of problem 2.2 and the following holds

$$z^T \nabla c_1(\bar{x}) \geq 0, \quad \forall i \in \bar{B}, \quad \forall z \in R^n, \quad z \neq 0, \tag{2.9}$$

where, $\bar{B} = \{i: \; C_1(\bar{x}) = 0, \forall i\}$, then there exists constants $\alpha_i \in [-1,1]$, $\forall i \in K(\bar{x})$ such that

$$\sum_{i \notin K(\bar{x})} \text{sign } f_i(\bar{x}) \cdot \nabla f_i(\bar{x}) + \sum_{i \in K(\bar{x})} \alpha_i \nabla f_i(\bar{x}) = 0, \tag{2.10}$$

Where $K(\bar{x}) = \{i: \; f_i(\bar{x}) = 0\}$.

2.4 NONLINEAR ℓ_1-APPROXIMATION

Consider the ℓ_1-problem and define the f_i's as follows:

$$f_i \triangleq f(x_i) - K(A, x_i), \quad \forall i: \; i=1 \ldots t_s \tag{2.11}$$

where $f(x_i)$ are real valued functions defined on a discrete set $x: \; \{x_1 \ldots x_{t_s}\}$ and $A:\{a_1 \ldots a_r\}$ is a set of real parameters. Then the ℓ_1-approximation problem can be stated as follows:

Problem 2.3

$$\underset{A \in R^n}{\text{Minimize}} \quad F(A,\underset{\sim}{x}) = \sum_{i=1}^{t_s} |f(\underset{\sim}{x}_i) - K(A,\underset{\sim}{x}_i)| \;, \tag{2.12}$$

for a chosen function $K(A,\underset{\sim}{x})$, where $\underset{\sim}{x}_i \in \underset{\sim}{x}$. Let A^* minimize $F(\cdot)$ and $A^* \in A$, then

$$\sum_{i=1}^{t_s} |f(\underset{\sim}{x}_i) - K(A^*,\underset{\sim}{x}_i)| \leq \sum_{i=1}^{t_s} |f(\underset{\sim}{x}_i) - K(A,\underset{\sim}{x}_i)| \tag{2.13}$$

$\forall A \in R^n$ so that $K(A^*,\underset{\sim}{x})$ is the best ℓ_1-approximation of $f(\underset{\sim}{x})$; moreover if $K(A,\underset{\sim}{x})$ is a nonlinear function of the parameters A then $K(A^*,\underset{\sim}{x})$ is the best nonlinear ℓ_1-approximation.

Remark 1.1 The case of best linear ℓ_1-approximation, where $K(A,x)$ is a linear function of A is well documented in the literature where, for example, the existence of the best approximation is guaranteed (Rice [40]). On the uniqueness, Jackson [5] gives a proof, for the case of Chebyschev sets, but in general the uniqueness cannot be guaranteed. However, several algorithms are available to solve this problem, e.g. [47], [41], [8] among others. On the existence and uniqueness of best nonlinear approximations the literature contains virtually no reference to these problems. Rice [39] presented a proof of the existence for a particular case of $F(A,\underset{\sim}{x})$, by assuming convexity of $F(A,\underset{\sim}{x})$ and proceeding in a similar manner as for the linear case. However if $K(A,\underset{\sim}{x})$ is nonlinear, it cannot be established that, $F(A,x)$ is convex for all $f(\underset{\sim}{x})$. Therefore these problems remain open. However by using lemma 2.1 we can derive the necessary conditions for $K(A^*,\underset{\sim}{x})$ to be the best nonlinear ℓ_1-approximation. From (2.11) if $f_i=0$, then

$f(x_i) = K(A^*,x_i)$. Defining the set $T(A^*) = \{i: f_i = 0\}$ and taking the gradient of (2.11), we have that $\nabla_{A^*} f_i = \nabla_{A^*} K(A^*;x)$. Then if A^* minimizes (2.12), the necessary conditions for $K(A^*,x)$ to be the best nonlinear ℓ_1-approximation are:

$$\sum_{i \notin T(A^*)} \operatorname{sign}\,(K(A^*;x_i) - f_i(x_i)) \cdot \nabla_{A^*} K(Z^*_A,x_i) +$$

$$\sum_{i \in T(A^*)} \alpha_i \cdot \nabla_{A^*} K(A^* x_i) = 0 , \qquad (2.14)$$

where

$$\alpha_i \in [-1,1].$$

2.5 ALGORITHMS FOR NONLINEAR ℓ_1-APPROXIMATION

In spite of the importance of this problem very few algorithms are available in the literature. However, in one of the existing methods due to Barrodale-Robert and Hunt [9] , the best ℓ_1-approximation is computed by functions nonlinear in one parameter in the following way:

Step I Search over a grid of values of nonlinear parameters to find some interval containing the minimum.

Step II Locate the minimum in this interval by using the Fibonacci search.

Clearly, Step I, is the well known separable programming problem e.g. [15], [11] among others, and Step II is the standard linear search over a closed bounded interval. While this algorithm is highly efficient, the class of practical problems with which it deals is restricted to a very small number.

Another algorithm covering a wider range of functions, is due to Osborne and Watson [35]. In this approach the best ℓ_1-nonlinear approximation is computed by linearization around some point followed by the solution of the linear ℓ_1-approximation problem by linear programming. An analysis of this algorithm can be found in [21], where it is shown that its numerical performance belongs to the steepest descent type and furthermore, [21] reported some examples where the algorithm failed to converge. In order to overcome the deficiencies in the previously described algorithms and to accelerate the convergence of the nonlinear ℓ_1-minimization, El Attar [20] presents an algorithm which converts the nonlinear ℓ_1-minimization problem into a sequence of unconstrained minimizations. The advantage of this approach is that gradient techniques such as quasi-Newton methods, may be used, thereby providing superlinear convergence for any hypersphere minimization. This technique is described in the next section.

2.6 SEQUENTIAL UNCONSTRAINED MINIMIZATION (S.U.M.) [1]

Consider the following function

$$P(\underset{\sim}{x},\varepsilon) = \sum_{i=1}^{m} [f_i^2(\underset{\sim}{x}) + \varepsilon]^{\frac{1}{2}}, \quad \varepsilon > 0 \qquad (2.15)$$

where $f_i(x)$ are continuous differentiable functions, for $i=1,\ldots,m$. The gradient vector and hessian matrix of $F(x,\varepsilon)$ are:

$$\nabla P(\underset{\sim}{x},\varepsilon) = \sum_{i=1}^{m} \frac{f_i(\underset{\sim}{x})}{(f_i^2(\underset{\sim}{x})+\varepsilon)^{\frac{1}{2}}} \cdot \nabla f_i(\underset{\sim}{x}) \qquad (2.16)$$

$$\nabla^2 P(\underset{\sim}{x},\varepsilon) = \sum_{i=1}^{m} \{(f_i^2(\underset{\sim}{x}) + \varepsilon)^{\frac{1}{2}} \cdot (f_i(\underset{\sim}{x}) \cdot \nabla^2 f_i(\underset{\sim}{x}) +$$

$$\nabla f_i(\underset{\sim}{x}) \cdot \nabla^T f_i(\underset{\sim}{x})) - (f_i^2(\underset{\sim}{x}) + \varepsilon)^{\frac{3}{2}} \cdot$$

$$(f_i^2(\underset{\sim}{x}) \cdot \nabla f_i(\underset{\sim}{x}) \cdot \nabla^T f_i(\underset{\sim}{x}))\} \cdot \qquad (2.17)$$

The S.U.M. approach can be restated in the form of the following problem:

Problem 2.4

$$\text{Minimize } P(\underset{\sim}{x},\varepsilon) = \sum_{i=1}^{m} (f_i^2(\underset{\sim}{x}) + \varepsilon)^{\frac{1}{2}}, \ \varepsilon > 0, \qquad (2.18)$$

for decreasing values of ε.

Suppose $\underset{\sim}{x}^*$ minimizes $P(x,\varepsilon)$, then $\underset{\sim}{x}^*$ is also a solution of the ℓ_1-problem (2.5). In order to justify this claim, the following two lemmas, taken from [21], are provided.

Lemma 2.2 For every $\underset{\sim}{x}_j \in R^n$ the following is true,

$$\lim_{\varepsilon \to 0} \sum_{i=1}^{m} (f_i^2(\underset{\sim}{x}_j) + \varepsilon)^{\frac{1}{2}} \to \sum_{i=j}^{m} |f_i(\underset{\sim}{x}_j)| \cdot \qquad (2.19)$$

Lemma 2.3 Let $\underset{\sim}{x}^*_{\varepsilon}$ minimize $P(\underset{\sim}{x},\varepsilon)$ and $\{\varepsilon_i\}$ be any sequence converging to zero. Then every limit point of the sequence $\{\underset{\sim}{x}^*_{\varepsilon_1}\}$ is a solution of the ℓ_1-problem.

Proof: Suppose $\underset{\sim}{x}^*$ is a limit point of the sequence $\{\underset{\sim}{x}^*_{\varepsilon_i}\}$, then there exists a subsequence which we can renumber as $\{\varepsilon_i\}$ such that $\varepsilon_i \to 0$ and $\underset{\sim}{x}^*_{\varepsilon_i} \to \underset{\sim}{x}^*$ as $i \to \infty$. Let $\underset{\sim}{x}_i$ be any element of R^n then

$$P(\underset{\sim}{x}^*_{\varepsilon_i}, \varepsilon_i) \leq P(\underset{\sim}{x}_i, \varepsilon_i), \quad \forall i \tag{2.20}$$

Letting $i \to \infty$ gives

$$F_1(\underset{\sim}{x}^*) \leq F_1(\underset{\sim}{x}) \quad \forall\, x \in R^n . \tag{2.21}$$

The iterative procedure of the S.U.M. approach is given by the following steps;

Step I   Pick $\hat{x}_1 \in R^n$, set $\varepsilon_1 > 0$ and set $K=1$, select L where $L \in R > 1$, where ε_1 can be computed as follows:

$$\varepsilon_1 = \frac{1}{10} \max_{i \in [1,m]} |f_i(\hat{\underset{\sim}{x}}_i)| \tag{2.22}$$

or

$$\varepsilon_1 = \frac{1}{10m} \left[\sum_{i=1}^{m} |f_i(\hat{\underset{\sim}{x}}_i)| \right]. \tag{2.23}$$

Step II   Minimize $P(\hat{x}, \varepsilon_K)$, denote the solution by x^*_k.

Step III   Set $\varepsilon_{k+1} = \varepsilon_k / L$.

Step IV   If $\varepsilon_{k+1} \leq \sigma$ and/or $|\bar{\underset{\sim}{x}}_k - \bar{x}_{k-1}| \leq \beta$, where σ, β are small numbers which determine the accuracy desired, stop.

Step V   If $K=1$, set $\hat{\underset{\sim}{x}}_{k+1} = \bar{\underset{\sim}{x}}_K$. If $K \geq 2$ find and estimate $\hat{\underset{\sim}{x}}_{k+1}$ to $\underset{\sim}{x}_{k+1}$ by Fiacco and McCormick extrapolation technique [23].

Step VI   Set $K = K+1$, go to Step II.

The justification of Step V can be found in [23], [21] and it should be noticed that the use of extrapolation in the S.U.M. algorithm can be associated with the S.U.M.T. [23] where it is used in order to accelerate the convergence. S.U.M. uses Step V for the same purpose and in addition it is used in order to improve its numerical stability, i.e. as $\varepsilon_i \to 0$ $P(x,\varepsilon_i)$ becomes ill conditioned, then a good estimate of the minimum can be obtained (under mild conditions) through extrapolations of the previous minima.

Remark 2.2 This algorithm is the most efficient and practical of all the algorithms available in the literature. However, the approach presents two basic drawbacks, i) a high number of iterations is required in order to reach the minimum and consequently it is unsuitable for minimizing functions of the penalty type [7] for the constrained problems, ii) there is no clear way to choose the parameter L, i.e. in examples 1,2, and 3, [21] uses L=10 and in example 4, L=16. An attempt was made however to solve example 4 with L=10 but the minimum was not reached to full accuracy and the number of function evaluations at the end of the computation was greater than that required for L=16. Moreover, [21] used the algorithm to obtain reduced order models for the single input-output case where the typical values of L were 10,16,22. A new method for ℓ_1-norm minimization and nonlinear ℓ_1-approximation based on the sequential unconstrained minimization approach is presented in the following sections.

2.7 FAMILY OF SEQUENTIAL UNCONSTRAINED MINIMIZATION

Given m continuous differentiable functions $f_i(x)$ with domain D in R^n, define the following cost function

$$\Gamma(\underset{\sim}{x},\phi,\beta) = \sum_{i=1}^{m} [f_i^2(\underset{\sim}{x}) + \beta_{i_K}\phi(f_i(\underset{\sim}{x}))]^{\frac{1}{2}}, \tag{2.24}$$

where

$$\beta_{i_k} \in R, > 0, \quad k = \{1,2,3,\ldots\}.$$

Let $s \subset R^n$, then define $\phi(\cdot)$ as follows:

1) $\phi(f_i(\underset{\sim}{x}))$, is a continuous differentiable function $\forall i$ and $\forall\, x \in S$. (2.25)

2) $\phi(f_i(\underset{\sim}{x})) \leq f_i^2(\underset{\sim}{x})$, $\forall i$, $\forall\, \underset{\sim}{x} \in s$. (2.26)

3) $\phi(f_i(\underset{\sim}{x})) \geq 0 \quad \dfrac{d\phi(f_i(\underset{\sim}{x}))}{df_i(\underset{\sim}{x})} < \infty$, $\forall i, \forall\, \underset{\sim}{x} \in D$. (2.27)

4) If $\downarrow\uparrow f_i^2(\underset{\sim}{x})$ then $\downarrow\uparrow \phi(f_i(\underset{\sim}{x}))$, $\forall i$, $\forall\, \underset{\sim}{x} \in s$. (2.28)

Now we can state the approach in the following problem.

Problem 2.5 Minimize

$$\Gamma(\underset{\sim}{x},\phi,\beta) = \sum_{i=1}^{m} [f_i^2(\underset{\sim}{x}) + \beta_{i_k}\phi(f_i(\underset{\sim}{x}))]^{\frac{1}{2}}, \tag{2.29}$$

for decreasing values of β_{i_k}.

Suppose that $\bar{\underset{\sim}{x}}$ solves problem 2.5, then it is claimed that $\bar{\underset{\sim}{x}}$ solves the ℓ_1-problem. The following lemmas justify this claim.

Lemma 2.4 The following is true

$$\sum_{i=1}^{m} [f_i^2(x) + \beta_{i_k}\phi(f_i(x))]^{\frac{1}{2}} \to \sum_{i=1}^{m} |f_i(x)|, \tag{2.30}$$

as $\beta_{i_k} \to 0$ and furthermore independently of $\phi(f_i(x))$.

Lemma 2.5 Let $\bar{x}$ be an interior point of D , then a necessary condition for $\bar{x}$ to solve problem 2.5 is that there exist constants $\alpha_i \in [-1,1]$ $\forall i \in (\bar{x})$ such that

$$\sum_{i \notin C(\bar{x})}^{m} \text{sign } f_i(x) \cdot \nabla f_i(\bar{x}) + \sum_{i \in C(\bar{x})} \alpha_i \cdot \nabla f_i(\bar{x}) = 0 \tag{2.31}$$

where $C(\bar{x}) = \{i: \ f_i(\bar{x}) = 0\}$. (2.31)

Proof: Taking the gradient of (2.29) we have

$$\nabla\Gamma(x,\phi,\beta) = \sum_{i=1}^{m} \frac{(f_i(x)+\frac{1}{2}\cdot\beta_{i_k}\cdot\frac{d\phi(f_i(x))}{df_i(x)})}{(f_i^2(x)+\beta_{i_k}\phi(f_i(x)))^{\frac{1}{2}}} \cdot \nabla f_i(x) \ . \tag{2.32}$$

Let $\{x_k\}$ and $\{\beta_{i_k}\}$ be two sequences converging to $\bar{x}$ and zero respectively $\forall i$. Then letting $k\to\infty$ in (2.32) and $i \notin C(\bar{x})$ we have

$$\nabla\Gamma(\cdot) \to \sum_{i=1}^{m} \text{sign } f_i(\bar{x}) \cdot \nabla f_i(\bar{x}). \tag{2.33}$$

Furthermore if $\bar{x}$ minimizes problem 2.5 then

$$\sum_{i=1}^{m} \text{sign } f_i(\bar{x}) \cdot \nabla f_i(\bar{x}) = 0.. \tag{2.34}$$

Now consider that $i \in C(\bar{x})$. We see that the sequence

$$\frac{f_i(x) + \frac{1}{2} \cdot \beta_{i_k} \frac{d\phi f_i(x))}{df_i(x)}}{(f_i^2(x) + \beta_{i_k}\phi(f_i(x)))^{\frac{1}{2}}} \tag{2.35}$$

does not have a definite limit in general. However, it is a bounded sequence with values between -1, +1 then (2.33) becomes

$$\sum_{i=1}^{m} \alpha_i \cdot \nabla f_i(\bar{x}) = 0 \tag{2.36}$$

and from (2.33) and (2.36) we get (2.31) where $\alpha i \in [-1,1]$.

<u>Lemma 2.6</u> Let $\bar{x}$ satisfy (2.31) and let $\{\beta_{i_k}\}$ be any sequence converging to zero. Then every limit point of the sequence $\{\bar{x}_{\beta_{i_k}}\}$ is a solution of Problem 2.5.

<u>Proof</u> If $\bar{x}$ minimizes $\Gamma(\cdot)$ we have

$$\Gamma(\bar{x}_{\beta_{i_k}}, \phi, \beta_{i_k}) \leq \Gamma(x, \phi, \beta_{i_k}), \forall i. \tag{2.37}$$

Letting $k \to \infty$ gives

$$\{\beta_{i_k}\} \to 0 \quad \forall i, \tag{2.38}$$

$$\bar{x}_{\beta_{i_k}} \to \bar{x} \quad \forall i, \tag{2.39}$$

$$F_1(\bar{x}) \leq F_1(x) \quad \forall x \in D, \tag{2.40}$$

and then the lemma holds.

Remark 2.3 Until now nothing has been said about the role played by $\phi(\cdot)$. The convergence proof as demonstrated previously is independent of $\phi(\cdot)$. Therefore the justification for its inclusion is given in this remark. In (2.24), $\phi(\cdot)$ can be regarded as a dynamic scaling factor. Let $\underset{\sim}{x} \in s$ and if $f_i^2(\underset{\sim}{x})$ decreases for some i, then $\phi(f_i(\underset{\sim}{x}))$ decreases. If $\phi(\cdot)$ decreases sufficiently, it is reasonable to expect that β_{i_k} does not have to become very small in order to converge to $\bar{\underset{\sim}{x}}$, due to the fact that $\beta_{i_k}\phi(\cdot) \to 0$ as $k \to \infty$. Consequently it is expected that the use of the dynamic factor ϕ, under suitable conditions, can lead to a reduction in the overall number of iterations. Furthermore, the parameter L (i.e. $\beta_{i_{k+1}} = \beta_{i_k}/L$, $L > 1$) becomes less critical in this algorithm than in S.U.M.

One way of decreasing the overall number of iterations, may be to decrease β_{i_k} by a large amount $\forall i$ at each iteration. Then clearly there exists one drawback (also shared by S.U.M.) in that, if the decrease is too drastic, it will lead to numerical difficulties. Suppose there exist L, $L' > 1$ where $L' \gg L$, and $\bar{\underset{\sim}{x}}$ is the vector point where $\Gamma(\cdot)$ attains its minimum after k_m iterations when β_{i_k} is reduced by a factor L. However, if the decreasing factor is chosen to be L' with the intention of obtaining $\bar{\underset{\sim}{x}}$ in K_r iterations such that $K_r < K_m$, then the following two cases may appear. I) If f_i $\forall i$ are not close to zero, the minimum might not be reached because $\Gamma(\cdot)$ becomes nondifferentiable before $\bar{\underset{\sim}{x}}$ is found. II) If any f_i becomes zero then $[\nabla^2\Gamma(\cdot)]^{-1}$ becomes singular, where $\nabla^2\Gamma(\cdot)$ is given by

$$\nabla\Gamma(\cdot) = \sum_{i=1}^{m} \{[f_i^2(\underset{\sim}{x}) + \beta_{i_k}\phi(f_i(\underset{\sim}{x}))]^{-\frac{1}{2}} \cdot [f_i(\underset{\sim}{x}) \cdot \nabla^2 f_i(\underset{\sim}{x}) +$$

$$\nabla f_i(\underset{\sim}{x}) \cdot \nabla^T f_i(\underset{\sim}{x})] - [f_i^2(\underset{\sim}{x}) + \beta_{ik}\phi(f_i(\underset{\sim}{x}))]^{-\frac{3}{2}} \cdot$$

$$[f_i(x) \cdot \nabla f(\underset{\sim}{x}) \cdot \nabla^T \Gamma_i(\underset{\sim}{x})]\} +$$

$$\{[f_i^2(\underset{\sim}{x}) + \beta_{i_k}\phi(f_i(\underset{\sim}{x}))]^{-\frac{3}{2}} \cdot [\tfrac{1}{2}\, \beta_{i_k} \frac{d\phi(f_i(\underset{\sim}{x}))}{df_i(\underset{\sim}{x})} \cdot \nabla^2 f_i(\underset{\sim}{x}) +$$

$$\tfrac{1}{2}\, \beta_{i_k} \cdot \frac{d^2\phi(f_i(x))}{df_i^2(\underset{\sim}{x})} \cdot \nabla f_i(\underset{\sim}{x})\, \nabla^T f_i(\underset{\sim}{x})] -$$

$$[\tfrac{1}{2}\, \beta_{i_k} \cdot \nabla\Gamma(\cdot)\ \ \nabla^T f_i(\underset{\sim}{x})]\}. \tag{2.41}$$

In order to avoid decreasing β_{j_k} too quickly, the following ideas are considered:

Suppose that $\beta_{jk}\ \forall j$ is computed as $\beta_{i_k} = \hat{\beta}_{i_k}/L$ (where β_{i_k}, L depend on the particular choice of $\phi(\cdot)$) for the iteration K and for the next iteration $\beta_{i_{k+1}} = \beta_{i_{k+1}}/L$,) then the following three cases are considered.

i) If the ratio $\gamma = \beta_{i_k}/\beta_{i_{k+1}}$ for any i is equal to 1, then $\beta_{i_{k+1}}$ is recalculated as

$$\beta_{i_{k+1}} = \frac{\hat{\beta}_{i_k}}{c}\ \ \forall i, \tag{2.42}$$

where $c \in R$, $1 < c \ll L$.

ii) If $\beta_{i_{k+1}} < \beta_{i_k}$ then $\beta_{i_{k+1}}$ is retained.

iii) If $\beta_{i_{k+1}} \ll \beta_{i_k}$ i.e. $r \geq r_m$ for some i then $\beta_{i_{k+1}}$ is recalculated in the following form

$$\beta_{i_{k+1}} = \beta_{i_{k+1}} \cdot t, \quad t > 1 \tag{2.43}$$

such that (ii) is satisfied.

Based on the above material, an algorithm for a particular choice of $\phi(\cdot)$ is presented in the next section.

2.8 AN ALGORITHM FOR ℓ_1-NORM MINIMIZATION AND NONLINEAR ℓ_1-APPROXIMATION

First define $\phi(f_i(\underset{\sim}{x}))$ as follows:

$$\phi(f_i(\underset{\sim}{x})) \triangleq \phi_L(f_i(\underset{\sim}{x})) = \text{Ln}(f_i^2(\underset{\sim}{x}) + \sigma), \quad \forall i, \tag{2.44}$$

where $\sigma = 1 + \zeta, \quad 0 < \zeta < 1.$

Then $\Gamma(\cdot)$ becomes

$$\Gamma(\underset{\sim}{x},\phi,\beta) = \sum_{i=1}^{m} (f_i^2(\underset{\sim}{x}) + \beta_{i_k} \cdot \text{Ln}(f_i^2(\underset{\sim}{x}) + \sigma))^{\frac{1}{2}} \tag{2.45}$$

and

$$\nabla\Gamma(\cdot) = \sum_{i=1}^{m} \frac{[f_i^3(\underset{\sim}{x}) + (\sigma_k + \beta_{i_k})\, f_i(\underset{\sim}{x})]}{(f_i^2(\underset{\sim}{x}) + \beta_{i_k} \cdot \text{Ln}[f_i^2(\underset{\sim}{x}) + \sigma])^{\frac{1}{2}}} \cdot \nabla f_i(\underset{\sim}{x}) \tag{2.46}$$

$$\nabla^2\Gamma(\cdot) = \sum_{i=1}^{m} \{(3.\, f_i^2(\underset{\sim}{x}) + \sigma_k + \beta_{i_k}) \cdot \nabla f(\underset{\sim}{x}) \cdot \nabla^T f(\underset{\sim}{x}) +$$

$$(f_i^3(\underset{\sim}{x}) + (\sigma_k + \beta_{i_k}) \cdot f_i(\underset{\sim}{x}) \cdot \nabla^2 f_i(\underset{\sim}{x})\} \cdot \Gamma_{2_i}^{-1} -$$

$$\{(f_i^3(\underset{\sim}{x}) + (\sigma_k + \beta_{i_k}) \cdot f_i(\underset{\sim}{x})) \cdot \nabla f_i(\underset{\sim}{x}) \cdot \nabla^T f_i(x)\} \cdot \Gamma_{2_i}^{-3} \cdot \Gamma_{1_i} \tag{2.47}$$

where

$$\Gamma_{1_i} = f_i^3(\underset{\sim}{x}) + (\sigma_k + \beta_{i_k}) \cdot f_i(\underset{\sim}{x}) \tag{2.48}$$

$$\Gamma_{2_i} = f_i^2(\underset{\sim}{x}) + \beta_{i_k} \cdot L\ [f_i^2(\underset{\sim}{x}) + \sigma_k] \cdot \tag{2.49}$$

The parameter σ can be seen as a to avoid numerical problems due to the $\log_e$ function (i.e. due to a fast decrease of $f_i(\underset{\sim}{x})$ in the early iterations). Clearly $\phi_L(\cdot)$ satisfies the definitions (2.25-2.28), i.e. if $\zeta_k = 0.1$, the set of all $\underset{\sim}{x}$ such that $f_i^2(\underset{\sim}{x}) \geq 1$, belongs to the set s. Then, problem 2.5 for $\phi(\cdot)$ as defined above can be restated as follows:

Problem 2.6 Minimize

$$\Gamma(\cdot) = \sum_{i=1}^{m} (f_i^2(\underset{\sim}{x}) + \beta_{i_k} \cdot L\ (f_i^2(\underset{\sim}{x}) + \sigma_k)]^{\frac{1}{2}} \tag{2.50}$$

for decreasing values of β_{i_k} and σ_k.

The choice of the appropriate values of the parameters β_{i_k}, σ_k, c,t,L are related to the scaling of the particular problem under consideration. In general, numer experience with the algorithm shows that the following are good choices of these parameters.

1) $\zeta_1 = 10^{-1}$, $\zeta_k = \{10^{-2}, 10^{-4} \ldots 10^{-10}\}$ for $1 \leq K \leq 6$, for $K > 6$ $\zeta_k = 10^{-12}$.

2) $L = 100$, $c = \frac{1}{2}$, $t = 10$, $r_m = 100$.

3) $b_{i_k} = \max |f_i(x)| \quad \forall i$, and let

z_{i_k} be an integer variable different than zero $\forall i,k$.

IF $b_{i_k} \geq 1$ then

$$\beta_{i_k} = \frac{z_{i_k}}{L}, \quad z_{i_k} \leftarrow b_{i_k}.$$

IF $b_{i_k} < 1$ then

$$\beta_{i_k} = \frac{z_{i_k}}{L \cdot 10^{\alpha}}, \quad z_{i_k} \leftarrow b_{i_k} \cdot 10^{\alpha}$$

$\alpha = \{1,2,3,\ldots\}$.

4) β_{i_k} can be expressed as $\beta_{i_k} = J \cdot 10^{\alpha_{i_k}}$ where J and α_{i_k} are positive or negative integer numbers and r can be computed as

$$r_i = \frac{10^{\alpha_{i_{k-1}}}}{10^{\alpha_{i_k}}}, \text{ where } r_1 \leq r_m \quad \forall i \tag{2.51}$$

Several efficient techniques based on gradient methods are available in the literature and among them is a quasi Newton algorithm that uses inexact linear search due to Fletcher [24]. The Fletcher method is the technique used in this thesis for the minimization of $\Gamma(\cdot)$ due to its good computational characteristic. The following steps describe the proposed algorithm for the particular choice of $\phi(\cdot) = \phi_L(\cdot)$:

Step I Pick the starting point $\underset{\sim}{x}_0$ and set $k=1$

Step II Set c, t, ζ_1, L, r_m

Step III Compute $b_{i_k} = \max |f_i(\underset{\sim}{x})| \quad \forall i,$

if $b_{i_k} \geq 1$ then $\beta_{i_k} = z_{i_k}/L,$

if $b_{i_k} < 1$ then $\beta_{i_k} = z_{i_k}/L \cdot 10^{\alpha}$

Step IV Using [24] minimize

$$\Gamma(\cdot) = \sum_{i=1} [f_i^2(\underset{\sim}{x}) + \beta_{i_k} \cdot \operatorname{Ln}(f_i^2(\underset{\sim}{x}) + \sigma_k)]^{\frac{1}{2}}$$

and denote the solution by $\overline{\underset{\sim}{x}}_k$. If $k=1$ set $\sigma_k=1$

Step V If $\|\underset{\sim}{x}_{k-1} - \underset{\sim}{x}_k\| \leq q_1$ or $|\Gamma(\cdot) - F_1(\cdot)| \leq q_2$, where q_1 and q_2 prespecified small numbers, depending on the desired accuracy, stop.

Step VI Set $k = k+1$ and compute β_{i_k} as in Step III.

Step VII Compute $\sigma_k = \sigma_k \cdot t$, if $\zeta_k \geq 10^{-12}$ then $\zeta_k = 10^{-12}$.

Step VIII Compute $r_i = \dfrac{10^{\alpha_{i_{k-1}}}}{10^{\alpha_{i_k}}}$.

If for any i $r_i = 1$ and $r_p = 0$, then $\beta_{i_k} = \beta_{i_{k-1}}/C \quad \forall i$, $r_p = 1$, go to Step IV. If for any i, $r_i = 1$ and $r_p = 1$, Step XI. Else, next step.

Step IX If $r_i < r_m \; \forall i$, then $\beta_{i_k} \leftarrow \beta_{i_k} \; \forall i$, $r_p = 0$, Step IV. Else next step.

Step X If $r_i \geq r_m$ then $\beta_{i_k} \leftarrow \beta_{i_k} \cdot t$, Compute r_i, go to Step IV.

Step XI Set $c \leftarrow c/t$, compute $\beta_{i_k} = \beta_{i_{k-1}}/c \; \forall i$, to Step IV.

In order to illustrate the performance of this algorithm the same examples used by [20] to test S.U.M. are solved with the proposed algorithm, and a comparison with S.U.M. is made in the respective tables. These results are presented in the following section.

2.9 Numerical Examples

Example 2.1: Given the set of nonlinear Equations

$$f_1(\underset{\sim}{x}) = x_1^2 + x_2 - 10$$

$$f_2(\underset{\sim}{x}) = x_1 + x_2 - 7$$

$$f_3(\underset{\sim}{x}) = x_1 - x_2^3 - 1$$

$$\text{minimize} \quad F(x) = \sum_{i=1}^{3} |f_i(x)|,$$

where $x_0' = [1,1]$.

This problem was solved using the proposed algorithm. After 5 iterations the computation ended with the following values:

$$F(\underset{\sim}{\overline{x}}) = .470424$$

$$f_1(\underset{\sim}{x}) = 0.$$

$f_2(\underset{\sim}{\bar{x}}) = -.4704$

$f_3(\underset{\sim}{\bar{x}}) = 0.$

The progress of the computation is shown in Table 2.1. From Table 2.3 we can see that the value of $F(\underset{\sim}{x})$ coincides with the one reached by the S.U.M. algorithm. But requires 7 iterations and 3 function evaluations less respectively. Furthermore at the point $\underset{\sim}{\bar{x}}$ the function $F(\underset{\sim}{\bar{x}})$ is not differentiable. Also it should be noticed that the minimium $\{\bar{\beta}_{i_5}\} = 2\times10^{-6}$ and for S.U.M. $\bar{\varepsilon}_{14} = 5.8 \times 10^{-13}$. Clearly β_{i_k} does not have to become very small in order to reach the minimum.

<u>Example 2.2</u>: Given the following set of nonlinear equations

$$f_1(\underset{\sim}{x}) = x_1^2 + x_2^2 + x_3^2 - 1$$

$$f_2(\underset{\sim}{x}) = x_1^2 + x_2^2 + (x_3-2)^2$$

$$f_4(\underset{\sim}{x}) = x_1 + x_2 - x_3 + 1$$

$$f_5(\underset{\sim}{x}) = 2x_1^3 + 6x_2^2 + 2\ (5x_3 - x_1 + 1)^2$$

$$f_6(\underset{\sim}{x}) = x_1^2 - 9x_3$$

minimize $F(\underset{\sim}{x}) = \sum_{i=1}^{6} |f_i(\underset{\sim}{x})|$, where $\underset{\sim}{x}^T = [1,1,1]$.

The minimum $F(\underset{\sim}{\bar{x}})$ found by the algorithm is $F(\underset{\sim}{\bar{x}}) = 7.89422$ and the progress of the computation is shown in Table 2.2. From Table 2.3 we see that the minimum found by the proposed algorithm is slightly lower than that obtained by S.U.M. Moreover 7 iterations and 17 function evaluations less were needed.

K	X_K		Γ	L_1-norms error.	No. function evaluations
	x_1	x_2			
1	2.849503	1.924619	.591692	.500051	28
2	2.842554	1.920228	.472734	.470814	24
3	2.842501	1.920178	.470952	.470451	13
4	2.842503	1.920175	.470469	.470425	13
5	2.842503	1.920176	.470429	.470424	7
					85

K	β_1	β_2	β_3
1	0.080000	0.050000	0.050000
2	0.000400	0.004000	0.000900
3	0.000200	0.002000	0.000450
4	0.000020	0.000200	0.000045
5	$2x10^{-5}$	$2x10^{05}$	$4.5x10^{-6}$

TABLE 2.1

K	X_K		X_3	Γ	L_1-norms error.	No. function evaluation
	X_1	X_2				
1	.522176	.000142	.021178	8.18656	7.92120	22
2	.536166	.000089	.031838	7.90620	7.89448	31
3	.536111	.000045	.031927	7.89994	7.89424	15
4	.535985	.000004	.031919	7.89479	7.89423	24
5	.535985	0.0	.031920	7.89428	7.89422	12
						104

K	β_1	β_2	β_3	β_4	β_5	β_6
1	.02	.03	.02	.02	.58	.08
2	.007	.004	.004	.01	.009	.0008
3	.0035	.002	.002	.005	.0045	.0004
4	.00035	.0002	.0002	.0005	.00045	.00004
5	3.5×10^{-5}	$2.\times10^{-5}$	2×10^{-5}	5×10^{-5}	4.5×10^{-5}	4×10^{-5}

TABLE 2.2

Example.	K	Optimal Point X_K			P,Γ	L_1 - norm of error *	L_1 - norm of error **	No. of function evaluation. *	No. of function evaluation. **
5.1	12*	2.842504*	1.920176*		.470429*	.470424	.470424	88	85
	5**	2.842503**	1.920176**		.470429**				
5.3	14*	.535971*	0.0*	.31918*	7.89423*	7.89423	7.89422	121	104
	5**	.535985**	0.0**	.031920**	7.89428**				
	11*	2.240826*	1.857637*	7.769832*					
6.1		-1.644823*	.165725*	.740422*	.559818*	.559817	.559815	166	148
	4**	2.240758**	1.857692**	6.770026**	.559827*				
		-1.644891**	.165873**	.742099**					

No. of Iterations = K ; * S.U.M. Algorithm ; ** Proposed algorithm

TABLE 2.3

<u>Examples 2.3</u>: Given the following functions

i) $\sqrt{x}$ where $x \in [0,1]$

ii) $e^x \cos x$ where $x \in [0,2]$

iii) $\sin x$ where $x \in [0,2\pi]$,

find a rational approximation of the form

$$K(A,x) = \frac{a_0 + a_1 x + a_2 \cdot x^2}{1+b_1 \cdot x + b_2 \cdot x^2}$$

such that $\| f(\underset{\sim}{x}) - K(A,\underset{\sim}{x}) \|_1$ is minimized. These problems were solved by discretizing every function into 51 uniformly spaced samples on the respective interval. The results are presented in Table 2.4. For Part i) a slightly lower minimum was obtained than for S.U.M. and 6 iterations and 71 function evaluations less were used. For ii) the minimum reached was slightly lower than that reached by S.U.M. and 6 iterations less were required while 5 more functions evaluations were necessary. For case iii) the minimum found was slightly higher than that found by S.U.M. but 3 iterations and 28 functions evaluations less were necessary.

<u>Example 2.4</u> Given the following impulse response corresponding to a seventh order single input-output system

$$f(t) = \frac{1}{2} e^{-t} - e^{-2t} + \frac{1}{2} e^{-2t} + \frac{1}{2} e^{-3t} + \frac{3}{2} e^{-\frac{3}{2}t} \sin (7t) + e^{-\frac{5}{2}} \sin (5t) ,$$

Function	Starting point a_0	a_1	a_2	b_1	b_2	Optimal point $\bar{a}_0$	$\bar{a}_1$	$\bar{a}_2$	$\bar{b}_1$	$\bar{b}_2$	No. of iterations *	No. of iterations **	No. of function evaluation *	No. of function evaluation **	l_1-norm of error *	l_1-norm of error **
$\sqrt{x}$	.1706	1.758	0.0	.9537	0.0	0.0*	8.563022*	29.314374*	24.738527*	12.229623*	11	5	301	230	.707195	.707182
sa						0.0**	8.562872**	29.312342**	24.737448**	18.22847**						
$e^x \cos x$	1.0	1.0	1.0	1.0	.1,0	.982518*	.567515*	- .759189*	-.570984*	.110569*	10	[illegible]	129	134	.170838	.170837
sa						.982515**	.567515**	- .759187**	-.570984**	.110569**						
Sinx	0.0	1.0	1.0	1.0	1.0	.641239*	-.204113*	0.0 *	-.529810*	.084322*	10	7	172	144	7.373005	7.373006
sa						.641289**	-.204113**	0.0 **	-.529817**	.084322**						

* S.U.M. Algorithm
** Proposed Algorithm
No. Samplings sa = 51.

TABLE 2.4

K	X_K			Γ	L_1-norms Error	No. of function evaluation
1	2.260052	1.892936	6.813806	2.856335	.601162	45
	-1.651138	.164252	.712679			
2	2.243663	1.863993	6.770822	.567899	.561006	27
	-1.658208	.165820	.752047			
3	2.241031	1.857987	6.769864	.560165	.559871	30
	-1.644800	.1657077	.740921			
4	2.240758	1.857692	6.770026	.559827	.559815	46
	1.644891	.165873	.942099			
						148

K	β_1	β_2	β_3	β_4	β_5	β_6
1	.02	.03	.02	.02	.58	.08
2	.007	.004	.004	.01	.009	.0008
3	.0035	.002	.002	.005	.0045	.0004
4	.00035	.0002	.0002	.0005	.00045	.00004

TABLE 2.5

find a third order system with impulse response of the form

$$K(a_1 t) = a_1 e^{-a_2 t} \cos(a_3 t + a_4) + a_5 e^{-a_6 t}$$

such that it is the best approximation in the sense of the ℓ_1-norm i.e.

$$\text{minimize} \quad \|f(t) - K(A,t)\| , \quad t \in [0,5].$$

With starting point $[2,2,7,0,-2,1]'$ the problem was discretized into 51 uniform samples in the respective interval and the progress of the computation is shown in Table 2.5. From Table 2.3 we can see that 7 iterations and 18 function evaluations less than S.U.M. were necessary to reach the minimum. Moreover this minimum was slightly lower than the one found by S.U.M.

2.10 CONCLUSIONS

A new and efficient algorithm for ℓ_1-norm minimization has been presented. One interesting feature of this approach is that the parameters are robust in the sense that L, σ, β, etc. are unchanged for all the examples presented in this thesis.

This algorithm was tested extensively on other problems and the same numerical performance was found. In addition, the algorithm is both numerically stable and converges to the ℓ_1-norm minimum for all examples tried. The efficiency of the algorithm lies in the reduced number of iterations and function evaluations required, due to the incorporation of the dynamic scaling factor $\phi(\cdot)$. Although the algorithm tends to become ill-conditioned when $\beta\phi(\cdot)$ tends to zero, the algorithm still converges to its minimum value.

CHAPTER 3
ORDER REDUCTION

3.0 INTRODUCTION

The main interest of this chapter is to study the reduced order model problem for multi-input multi-ouput (M.I.M.O) linear time invariant systems, i.e. given an m-input ℓ-output system S, find another system S_r m-input, ℓ-output, of lower dimension such that S_r approximates S in some sense. In section 3.1 this problem is enunciated and some of the important approaches, which deal with the solution of this problem, are reviewed. An analysis of the optimal order reduction for the M.I.M.O. case is presented in section 3.2 by considering the minimization of the input-output mapping, namely the impulse matrix response which characterizes the system. In section 3.3 a method is proposed for obtaining reduced order models for the M.I.M.O. case, based on the theory of the previous section. Several numerical examples are given in section 3.4 and the conclusions of this chapter are contained in section 3.5.

3.1 REDUCED ORDER MODELS (R.O.M.)

Let us consider a linear time invariant system (L.T.I.) represented as follows

System S:

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{X}(t) + B\underset{\sim}{U}(t), \quad \underset{\sim}{x}(0) = 0$$

$$\underset{\sim}{y}(t) = C\underset{\sim}{X}(a) + D\underset{\sim}{U}(t) \qquad (3.1)$$

Where A,B,C,D are constant matrices of dimension nxn, nxm, ℓxn, ℓxm respectively and $\underset{\sim}{x},\underset{\sim}{y},\underset{\sim}{u}$, vectors of corresponding dimensions n, ℓ, and m with McMillan degree ρ[42] (e.g. $\rho \leq n$) and with transfer matrix H(s), impulse response matrix H(t). The R.O.M. Problem can be stated in the following way:

Problem 3.1 Given the system S as above, find a system S_r with transfer matrix $H_r(s)$ and impulse response matrix $H_r(t)$ such that

System S_r in represented as:

$$\dot{x}_r(t) = A_r x_r(t) + B_r U(t), \quad x_r(0) = 0$$

$$y_r(t) = C_r x_r(t) + D_r U(t) \qquad (3.2)$$

where y_r and x_r are vectors of dimensions ℓ, r and A_r, B_r, C_r, D_r are constant matrices of proper dimensions $r<n$ and $H_r(t)$ approximates $H(t)$ in some sense.

The following remark is appropriate at this point.

Remark 3.1 The R.O.M. problem is sometimes stated as above but without the condition $\rho=n$. However, this condition is relevant in the sense that if $\rho<n$ for S and an S_r is found such that it is a good approximation to S then $r<n$. But if $\rho<r$ then the question to be answered is whether or not it is more practical to find the minimal realization which is an exact representation of system S or just an approximated system S_r. However it is not assumed here that system S is controllable and observable, rather, it is assumed that its Macmillan degree is known. Several methods are presented in the literature which deal with the R.O.M. problem. In general, one can classify these methods into the following categories: i) Singular perturbation ii) Continued fraction expansion iii) Power series Expansion of $H(s)$ iv) Error minimization. We present here a brief review of these methods along with some comments.

i) Singular Perturbation Consider the system S where x is expressed as

$$\dot{x} = \begin{bmatrix} \dot{x}_r \\ \mu z \end{bmatrix}, \quad \mu \text{ is a scalar } > 0. \qquad (3.3)$$

Then by partitioning the triple (A,B,C,) we have

$$\dot{x}_r(t) = A_{11}x_r(t) + A_{12}z(t) + B_1U(t), \qquad (3.4)$$

$$\mu\dot{z} = A_{21}x_r(t) + A_{22}z(t) + B_2U(t). \qquad (3.5)$$

Now if we set $\mu=0$ and solve (3.5) we have

$$z(t) = -A_{22}^{-1}A_{21}x_r(t) - A_{22}^{-1}B_2U(t). \qquad (3.6)$$

Then the reduced model S_r is clearly

$$\dot{x}_r(t) = (A_{11}X_r(t) + A_{12}A_{22}^{-1}A_{21})\,X_r(t) + (B_1 - A_{12}A_{22}^{-1}B_2)\,U(t)$$

$$y_r(t) = C_{11}x_r(t). \qquad (3.7)$$

Two method based on the singular perturbation approach for obtaining R.O.M.'s are due to Davison [15] and Hutton [4]. Davison's methods, known as the "Dominant mode", neglects the high frequency modes and retains only the low frequency components. Two variations of these methods are due to [13] [17]. While these techniques preserve the stability and the states of the reduced systems are physically meaningful, their chief drawbacks are:

a) When the system poles are not clearly distinguished, i.e. when the poles are too close to each other numerically, the method fails, b) For systems where $n \geq 10$ this technique becomes both computationally costly and numerically inaccurate.

ii) Continued Fraction Expansion. Consider the system S as defined before and H(s) as its transfer matrix. Assume that $\rho = n$, then H(s) can be expressed as follows:

$$H(s) = \frac{C\Gamma(s)B}{|SI-A|} \tag{3.8}$$

where $\Gamma(s) = \text{adj}\,(SI-A)$,

then (3.8) can be rewritten as

$$H(s) = \frac{B_0 + B_1 S + \ldots B_{p-1} S^{p-1}}{a_0 + a_1 S + \ldots + a_p S^p} \tag{3.9}$$

where $B_0 \in R^{\ell \times m}$ and a_i are scalar $\forall i = (1 \ldots p)$

or

$$H(s) = \{H_1 + s\{H_2 + s\{H_3 + \ldots s\, H_j\}^{-1} \ldots\}^{-1}\}^{-1}\}^{-1}. \tag{3.10}$$

Now the R.O.M can be obtained from (3.10) for some specific j. This method orginally developed for single input-output systems by Chen and Shieh [13] was later extended by Chen [14] (as shown above) to cover the M.I.M.O. case. This technique guarantees the stability of the R.O.M. for the S.I.S.O case but not for the M.I.M.O. case even though the original system might be stable. Some of its drawbacks are as follows: a) a zero eigenvalue often causes numerical failure b) the requirement that B_0 be non-singular cannot be met in general. A more detailed analysis and critique of this method can be found in Calfe [12] where he concludes that this method is totally unsuitable for the order reduction problem in the M.I.M.O. case.

iii) Power Expansion of $H(s)$ Again consider S with transfer matrix $H(s)$ expanding $H(s)$ around $s=0$ we have

$$H(s) = \sum_{i=1}^{\infty} CA^{i-1}BS^{i-1} \qquad (2.11)$$

then R.O.M.'s can be obtained by applying an algorithm to $H(s)$ which finds the minimal realization using the Hankel matrix approach. Shamash (44) used the Silverman [45] algorithm on $H(s)$ in order to obtain R.O.M. for M.I.M.O. systems. However methods based on those ideas give R.O.M.'s that are not optimal in any sense. Moreover, they fail to reproduce the steady state response.

iv) Error Minimization The underlying idea of these methods is to minimize the error function of the quadratic type, of the form:

$$J = \int_0^{\infty} \| \underset{\sim}{y}(t) - \underset{\sim}{y}_r(t) \|_Q^2 \qquad (3.12)$$

for suitable input $\underset{\sim}{U} \in R^m$ and where Q is a constant matrix $\in R^{\ell x \ell}$. One variation proposed by Galiana [25] is to minimize J^I of the form

$$J^I = \int_0^{\infty} w^{\frac{1}{2}} (H(t) - H_r(t)) Q (H(t)-H_r(t))' \, w^{\frac{1}{2}} \, dt \qquad (3.13)$$

for $\ell \geq m$ and J^{II} for $m<\ell$ as follows

$$J^{II} = \int_0^{\infty} Q^{\frac{1}{2}} (H(t) - H_r(t))' \, w (H(t)-H_r(t)) \, Q^{\frac{1}{2}} \, dt \qquad (3.14)$$

where Q and w are constant diagonal positive definite matrices and where impulse functions are used as inputs. R.O.M. were obtained by Wilson [50], where he used the expectation operator E for stochastic processes while considering deterministic impulses as inputs. The cost function to be minimized was then

$$J^{III} = \lim_{t\to\infty} E\{\| \underset{\sim}{y}(t)-\underset{\sim}{y}_r(t)\|_Q^2\}, \text{ or} \tag{3.15}$$

$$J^{III} = \text{trace } [PS], \text{ where} \tag{3.16}$$

$$E[\underset{\sim}{U}(t)] = 0, \quad E[\underset{\sim}{U}(t)\ \underset{\sim}{U}^t(\tau)] = N\sigma\ (t-\tau), \tag{3.17}$$

N is a positive definite symetric matrix. It can be shown that solution of (3.16) implies the solution of

$$F'P + PF = -B'Q\ B, \tag{3.18}$$

$$FR + RF' = -S, \tag{3.19}$$

where Q is as before, $F = \text{diag}[A, A_r]$, σ is the Dirac function

$G = (C, -C_r)$ and

$$S = \begin{bmatrix} BNB' & BNB_r' \\ B_rNB' & B_rNB_r' \end{bmatrix}. \tag{3.20}$$

Furthermore the reduced system is assumed to be an aggregation form (see section 4.1). Therefore the aggregation matrix and the R.O.M. has to be found through the direct minimization of J^{III}, and consequently the large numbers of variables to be minimized imply an increase in computational effort. The Hirzinger and Kreisselmeier [27] method minimizes

$$J_i = \| \underset{\sim}{x}_0 \|_P^2 \qquad (3.21)$$

where $\underset{\sim}{x}(0)$ is the inital state and P is the solution of the Liapunov Eq.(3.18) for every input $U(t) \in \underset{\sim}{U}(t)$. Then the R.O.M. can be obtained by minimizing

$$J^{(IV)} = \sum_{i=1}^{m} J_i . \qquad (3.22)$$

The main disadvantages of these methods are the selection of the weighting matrices W,Q,N [~~note~~ that an optimum reduced model can be obtained for a specific choise of W,Q,N but it may not necessarily be the global optimum in the strict sense, due to the fact that optimum weighting matrices have to be found first] and the fact that the numerical effort is considerably high even for small sized problems (i.e. $n \leq 10$) and these methods in general can not reproduce the steady-state response. Recently Wilson and Mishra [51] presented an improved version of Wilson's method, by decomposing the output responseinto transient and steady state portions. The transient portion was then reduced using [50] and the steady state part matched exactly. However, the drawbacks mentioned above still applies to this method with the exception that now the steady state of the original system is reproduced by the R.O.M.

3.2 OPTIMAL ORDER REDUCTION

Consider the multi-input , multi-output system (3.1) described by the convolution integral

$$\underset{\sim}{y}(t) = \int_0^t H(t-\tau)\,\underset{\sim}{U}(\tau)\, d \qquad (3.23)$$

where $\underset{\sim}{y} \in R^L$, $\underset{\sim}{U} \in R^m$ and $H \in R^{\ell xm} \cdot H(\cdot)$ is known as the matrix impulse response of the form

$$H(t) = [h_{ij}(t)] \quad \forall i \in I, \quad \forall J \in \theta \tag{3.24}$$

where from now on $I = \{1 \ldots \ell\}$, $\theta = \{1 \ldots m\}$. We assume that $h_{iJ}(t)$ is of the form

$$h_{iJ}(t) = h^{I}_{iJ} \ \sigma(t) + h^{II}_{iJ}(t) \tag{3.25}$$

where $h^{II}_{iJ}(t)$ is a measurable absolutely integrable function i.e.

$$\int_0^t |h^{II}_{iJ}(t)| dt < \infty \qquad \forall t < \infty \ \forall i \in I, \quad \forall J \in \theta \tag{3.26}$$

and $\sigma(t)$ is the unit impulse distribution. Now consider the system (3.2) described by its convolution integral

$$\underset{\sim}{y_r}(t) = \int_0^t H_r(t-\tau) \ \underset{\sim}{U}(t) \, d\tau \tag{3.27}$$

where $y_r \in R^{\ell}$, $H_r \in R^{\ell xm}$ and $\underset{\sim}{U}$ is the same as before. $H_r(t)$ and $h_{r_{iJ}}(t)$ are of the form

$$H_r(t) = [h_{r_{iJ}}(t)] \,, \tag{3.28}$$

$$h_{r_{iJ}}(t) = h^{I}_{r_{iJ}} \sigma(t) + h^{II}_{r_{iJ}}(t) \quad \forall i \in I, \quad \forall_J \in \theta \tag{3.29}$$

and $h^{II}_{r_{iJ}}(t)$ is a measurable, and absolutely integrable function. It is desired to apprixmate system S by system S_r, in other words, to find the quadruple $[A_r, B_r, C_r, D_r]$ such that S_r is a good approximation to S. If the same input vector is applied to both systems the output error is given by

$$\underset{\sim}{e}(t) = \underset{\sim}{y}(t) - \underset{\sim}{y}_r(t), \tag{3.30}$$

and the system error by

$$\underset{\sim}{e}(t) = \int_0^t H_e(t-\tau)\, \underset{\sim}{U}(\tau)\, d\tau, \tag{3.31}$$

where

$$H_e(t) = [h_{e_{ij}}(t)]. \tag{3.32}$$

and

$$h_{e_{ij}}(t) = h_{e_{ij}}(t) - h_{e_{ij}}(t), \quad \forall i \in I, \quad \forall j \in \Theta. \tag{3.33}$$

Equation (3.31) can be rewritten in operator form as $(H_e U)(t)$ where $H_e(\cdot)$ is the associated impulse response matrix of the operator H. Clearly $H_e(\cdot)$ maps U onto e for all inputs U i.e. $U \to e = H_e * U$. Let $U \in L_1^m$ and define

$$\|h_{e_{ij}}(t)\| = |h^{I}_{e_{ij}}| + \|h^{II}_{e_{ij}}\|_1 . \tag{3.34}$$

The induced L_1 norm of $H_e(t)$ is

$$I_N = \max_{J\in\theta} \sum_{i\in I} \| h_{e_{iJ}}(t)\| \cdot \qquad (3.35)$$

The following facts concerning the systems S, S_r, S_e are straight forward consequences of known results [53], [19], [48]:

<u>Fact 1</u> The following four conditions are equivalent:

i) the system S, (S_r, S_e) is BIBO stable or L_∞ stable

ii) the system S, (S_r, S_e) is L_p stable for all $p\in[1,\infty]$

iii) the system S, (S_r, S_e) is L_1 stable.

iv) $h^{II}_{iJ}(\cdot) \in L_1$ (Resp. $h^{II}_{r_{e_{ij}}}(\cdot)$, $h^{II}_{e_{iJ}}(\cdot)$).

Basically this fact states, that an m-input ℓ-ouput system with impulse response matrix $H(\cdot)$ is I, II, III, if, and only if, each component of $H(\cdot)$ namely, $h^{II}_{iJ}(\cdot) \in L_1$.

<u>Fact 2</u> Let $H(\cdot)$ be a stable impulse matrix then, whenever $U(\cdot)$ L^m_1, we have

$$\|HU\| \leq \alpha_1 . \| U\|_1 \qquad (3.36)$$

$$\text{where} \quad \alpha_1 = \|\hat{H}\|_{i1} \qquad (3.37)$$

and where $\|\hat{H}\|_{i1}$ represents the ℓ_1-induced matrix norm R^{Lxm} and $\hat{H}$ is the matrix whose iJ^{th} entry is as in (3.34). Then

$$\sup_{U\in L_1} \frac{\|HU\|}{\|U\|_1} = \alpha_1 \qquad (3.38)$$

In the next section an approach to the R.O.M. problem is developed based on the material presented above.

3.3 M.I.M.O. REDUCED ORDER MODEL

Consider the systems S and S_r where it is desired that the system S_r provide a uniformly good approximation over all inputs $U(t)$. Clearly the induced operator norm I_N as defined in the previous section is the cost funtion to be minimized but unfortunately in practice it is almost impossible to do so. However an alternative procedure is as follows:

First let $L(He(t))$ be a strictly proper matrix then (3.4) becomes

$$\| h_{e_{iJ}}(t) \| = h^{II}_{e_{iJ}}(t) \tag{3.39}$$

For BIBO stability we have

$$\int_{t_o}^{t} \| H_e(t) \| \, dt = K < \infty \tag{3.40}$$

where

$$\|U(t)\| \leq K\mu < \infty \tag{3.41}$$

Defining the $\|\cdot\|$ as

$$\|H_e(t)\| = \sum_{i \in I} \sum_{J \in \theta} |h_{e_{iJ}}(t)| \tag{3.42}$$

then from (3.40) we have

$$J_1 = \int_{t0}^{t} \sum_{i \in I} \sum_{J \in \theta} |h_{e_{iJ}}(t)| \, dt = K . \tag{3.43}$$

In the case that $[H_e(t)]$ is a proper matrix, then $h_{e_{ij}}(t)$ is as in (3.34). Define the matrix

$$z(t) = \{z_{ij}(t)\} \tag{3.44}$$

whose entries are of the form

$$z_{ij}(t) = |h^{I}_{e_{ij}}| + \|h^{II}_{e_{ij}}(t)\|_1, \ \forall i\in I, \ \forall j\in\theta \tag{3.45}$$

Let $\bar{J}_2$ be a cost function of the form

$$\bar{J}_2 = \sum_{i\in I} \sum_{j\in\theta} z_{ij}(t) , \tag{3.46}$$

then $\bar{J}_2$ becomes

$$\bar{J}_2 = \bar{J}_1 + \sum_{i\in I} \sum_{j\in\theta} \|h^{I}_{e_{ij}}\| . \tag{3.47}$$

In view of the above, the reduced order model can be obtained by minimizing $\bar{J}_1$ and $\bar{J}_2$. However for computational simplicity we are going to consider the discretized form of $\bar{J}_1$ and $\bar{J}_2$. First let $\hat{A}$ be a set of real parameters $A : \{a_1,\dots a_q\}$ and T be the discrete set $T: \{t_0,\dots t_s\}$ where $t\in R_+$. Defining the impulse matrix of the system S_r as follows

$$H_e(A,t) = \{h_{e_{ij}}(A,t)\} , \tag{3.48}$$

The discretized form of the L_1 norm of $h_{e_{ij}}(t)$ for the strictly proper case is denoted by

$$\sum_{k\in T} |h_{e_{ij}}(A,t_k)| \cdot \tag{3.49}$$

Clearly, $\bar{J}_1$ becomes

$$\bar{J}_1 = \sum_{k\in T} \sum_{i\in I} \sum_{j\in\theta} |h_{e_{ij}}(A, t_k)| \tag{3.50}$$

and it follows that reduced order models can be obtained by minimizing $\bar{J}_1$ for the strictly proper case. Moreover, by minimizing $\bar{J}_1$, we can also obtain reduced order models for the proper case and, the following is a justification of this claim. Let $H(s)$ be a proper matrix in s. It is known that

$$H_1(s) = H(s) - \Big|_{s\to\infty} H(s) \tag{3.51}$$

where $H_1(s)$ is the strictly proper matrix associated with the triple $[A,B,C]$ and $H(s)$ is the matrix associated with the quadruple $[A,B,C,D]$. Then, finding a R.O.M. for the proper case implies finding the triple $[A_r,B_r,C_r]$ where $D=D_r$. Now clearly by minimizing J_1, we can obtain reduced order models S_r that are a good approximation to S for all inputs $\underset{\sim}{U}(t)$. However the minimization of $\bar{J}_1$ is accomplished by using the algorithm for ℓ_1-norm minimization proposed in Chapter II. At this point the following remarks are appropptiate:

<u>Remark 3.2</u> Choosing the form of the triple (A_r,B_r,C_r) is a difficult task especially for the continuous case. The choice of the structure clearly affects the computational effort (i.e. introducing more unknown parameters). However the difficulty is due to the fact that the structure of

(A_r, B_r, C_r) is not known a priori. Some forms to circumvent this problem are presented in the literature [27], [6], i.e. [6] the triple (A_r, B_r, C_r) is chosen to have the pair (A_r, B_r) in controllable canonical form. However, in this thesis what has been chosen is $(\hat{A}_r, B_r, C_r)$ where B_r, C_r are full matrices and $\hat{A}_r$ is in Jordan canonical form. This structure presents the advantage that the closed form solution of the transition matrix can be easily found by prespecifying the nature of its eigenvalues namely real, complex or imaginary.

<u>Remark 3.3</u> Some care is required in treating the R.O.M. problem for the M.I.M.O. case. Consider the matrices $H_r(s)$ and $H_r(t)$ associated with the triple $(\hat{A}_r, B_r, C_r)$. Then we have that

$$h_{r_{iJ}}(t) = \sum_{k=1}^{n} \phi_{ik} \left(\sum_{n=1} b_{kh} c_{hJ} \right) \tag{3.52}$$

where ϕ_{ik} is the transition matrix, $A_{iJ} \in A_r$, $b_{kn} \in B_r$, $C_{hJ} \in C_r$ and $a_{iJ}, b_{kh}, C_{hJ} \in \hat{A}_r$. However $H_r(t)$ can be rewritten in the form:

$$h_{r_{iJ}}(t) = \sum_{q=1}^{p} \gamma_q^{iJ} \, t^{\beta_q} e^{h_q t} \tag{3.53}$$

where $h_q \in R$, $\forall q$, where β is a positive integer $\forall q$. Let us consider the case where $\hat{A}_r$ has distinguishable eigenvalues, then $\beta_q = 0$ $\forall q$. Moreover, $H_r(s)$ can be rewritten in the following form:

$$H_r(s) = \sum_{q=1}^{z} \frac{M_q}{(s+\lambda_q)} \tag{3.54}$$

$$\text{or} \qquad H_r(t) = \sum_{q=1}^{z} M_i \, e^{\lambda_i t} \tag{3.55}$$

$$M_q = \{\gamma_q^{ij}\} \qquad \gamma_q^{ij} \in \hat{A}$$

$$M_q = \lim_{S \to \lambda_q} H_r(s)\,(s+\lambda_q) \quad \forall q. \tag{3.56}$$

Suppose that the method proposed in this section is applied to system S and S_r where $H_r(t)$ is as in (3.53), then we obtain

$$\gamma_q^{*ij}, \quad \lambda_q^* \in A^* \text{ and } M_q^* = \{\lambda_q^{*ij}\} . \tag{3.57}$$

By Gilbert [26] we know that if R_q is the rank of M_q^* then the Macmillan degree ρ is given by

$$\rho = \sum_{q=1}^{z} R_q. \tag{3.58}$$

It is clear though that the relationship $n>\rho$ need not hold. Then (3.53) is an unsuitable expression for the reduced system. However from (3.52) it can easily be shown that $n>r>\rho$ is always true.

In the next section we present several numercial examples which illustrate the method proposed in this chapter.

3.4 NUMERICAL EXAMPLES

Example 3.1. Given the transfer matrix

$$H(s) = \begin{bmatrix} \frac{1}{(s+1)} & \frac{1}{(s+1)(s+2)} \\ \frac{1}{(s+1)(s+3)} & \frac{1}{(s+3)} \end{bmatrix}$$

with Macmillan degree $\rho=4$ and eigenvalues λ_i: {-1, -1, -2, -3}, find a reduced order model of dimension 2. a) The A_r was choosen to be a diagonal matrix and this problem was discretize in the interval [0,5] for 51 uniform samples. At the minimum the error was $J_1^* = 4.138910$, and the reduced model is given in Table 3.1. b) Now the A_r was chosen to have complex eigenvalues namely $\lambda_1 = a + b_J$, $\lambda_2 = \bar{a} - b_J$. At the minimum the error was $J_1^* = 5.1775260$ and the reduced model is given in Table 3.2. As an illustration, the plots of $h_{iJ}(t)$, $h_{r_{iJ}}(t)$ are shown in Figs. 3.1-2 and 3.3-4 for part a,b respectively.

Example 3.2. Given the transfer matrix

$$H(s) = \begin{bmatrix} (\frac{s_0}{s+1} - \frac{100}{s+1,1} + \frac{s_0}{s+1,2}) & (\frac{s_0}{s+0,4} \; \frac{100}{s+1} + \frac{s_0}{s+1,1}) \\ (\frac{s_0}{s+0,a} \; \frac{100}{s+1} + \frac{s_0}{s+1,1}) & \frac{1}{s+1} \end{bmatrix}$$

with $\rho=7$. Find a reduced order model of dimension $r=5$, where A_r is a diagonal matrix. The minimum obtained by the proposed method is $J_1^* = .482986$ and the reduced order model is given in Table 3.3 and in Figures 3.5 and 3.6 $h_{iJ}(t)$ and $h_{r_{iJ}}(t)$ are plotted vs time $\forall i,J$.

Example 3.3 given $H(s)$, transfer matrix corresponding to a M.I.M.O. system with 4 inputs and 3 outputs, $H(s)$ is given in Table3.4. This well known transfer matrix [29], has $\rho=9$ and the eigenvalues are λ_i: {-1,-1,-1,-2,-2,-3,-3,-4,-5}. A reduced order model of dimension 7 was found, using the proposed method, and at the minimum the error was $J_1=5.144338$, where the matrix A_r was choosen to be diagonal. The R.O.M. is given in Table 3.5 and the respective plots are given in figures 3.7-3.12. The following

$$A_r = \begin{bmatrix} -.89157712 & 0. \\ 0. & -.2735897 \end{bmatrix}$$

$$B_r = \begin{bmatrix} .42067 & .6140634 \\ .1111591 & .1343549E+01 \end{bmatrix}$$

$$C_r = \begin{bmatrix} .2119433E+01 & .9965355 \\ .2876781 & .5158638 \end{bmatrix}$$

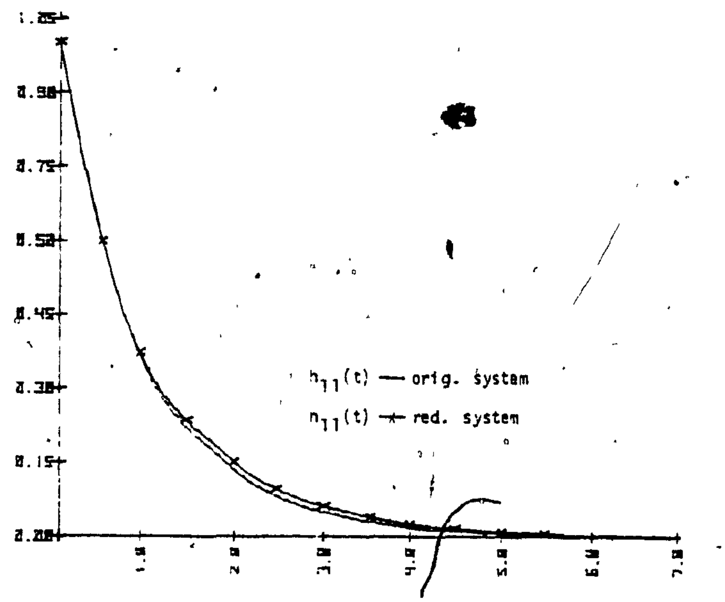
TABLE 3.1

$$A_r = \begin{bmatrix} -1.405113E+01 & .6301060E-03 \\ -.6301060E-03 & -.1405113E+0.1 \end{bmatrix}$$

$$B_r = \begin{bmatrix} .3787430E-1 & .3158332E-01 \\ .2883521E+02 & .9006512E+0.2 \end{bmatrix}$$

$$C_r = \begin{bmatrix} .3390468E+02 & -.1184214E-01 \\ -.3316715E+01 & .8652492E-02 \end{bmatrix}$$

TABLE 3.2


h11(t) — orig. system
h11(t) —x— red. system

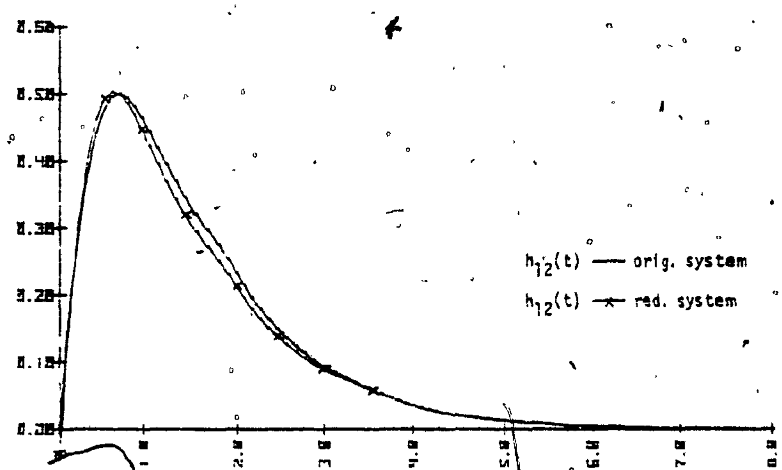
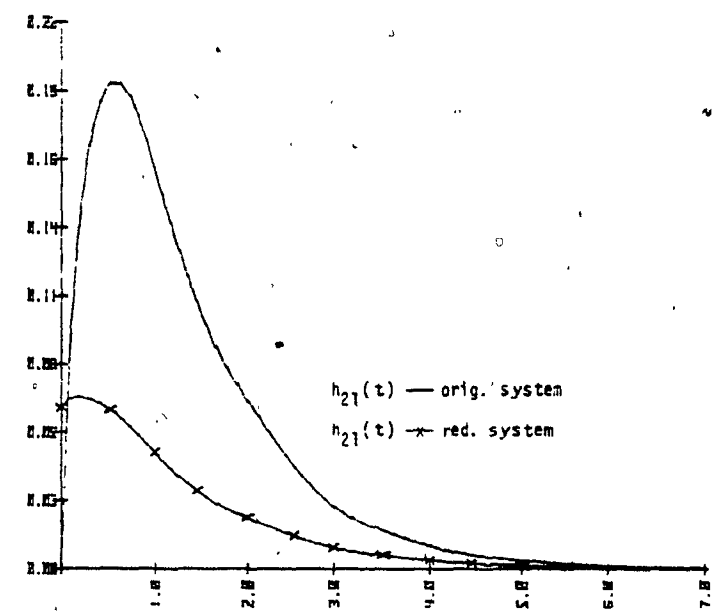
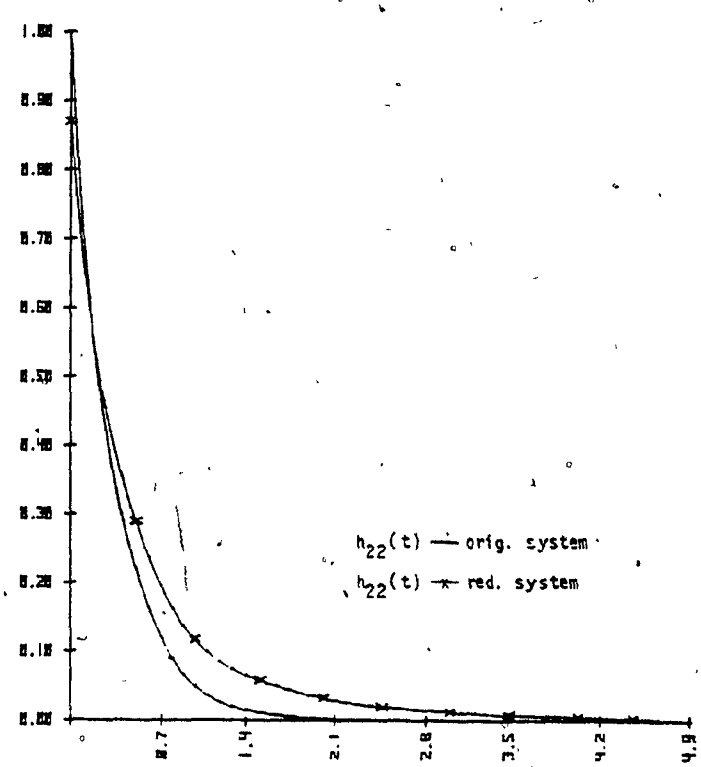

h12(t) — orig. system
h12(t) —x— red. system

FIG. 3.1

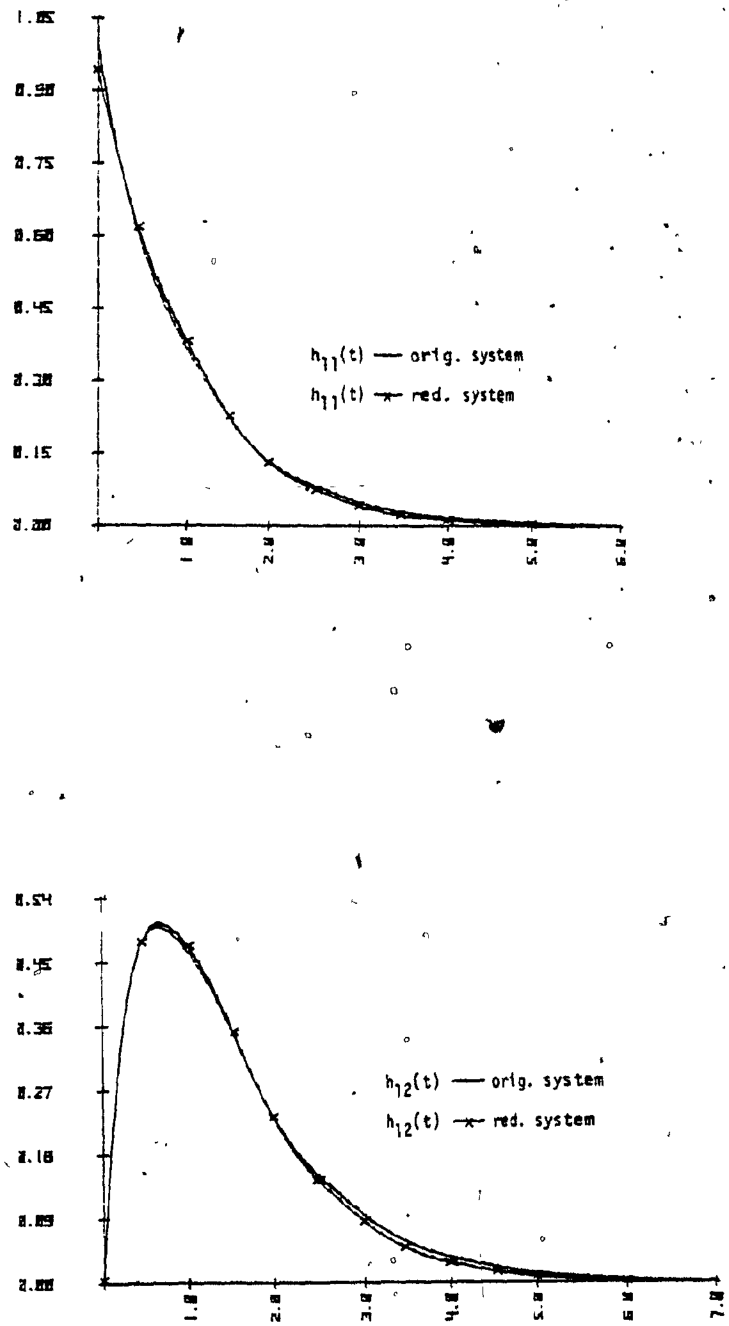


$h_{21}(t)$ — orig. system
$h_{21}(t)$ -x- red. system



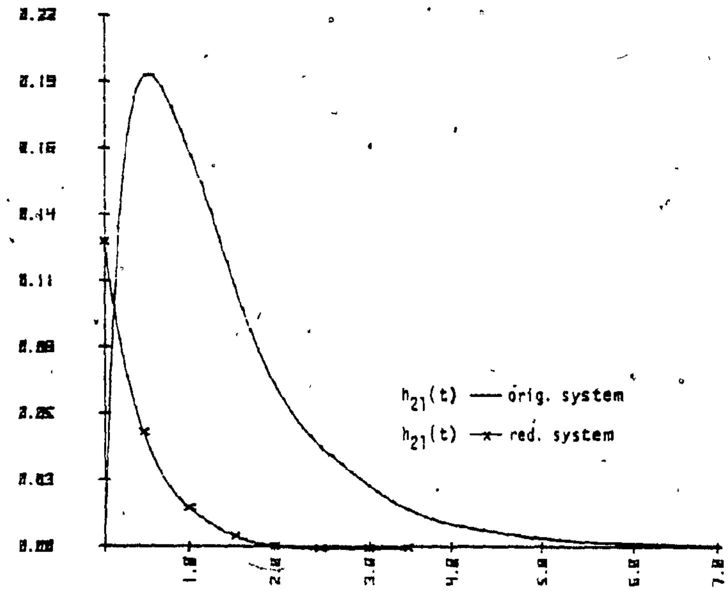
$h_{22}(t)$ — orig. system
$h_{22}(t)$ -x- red. system

FIG. 3.2

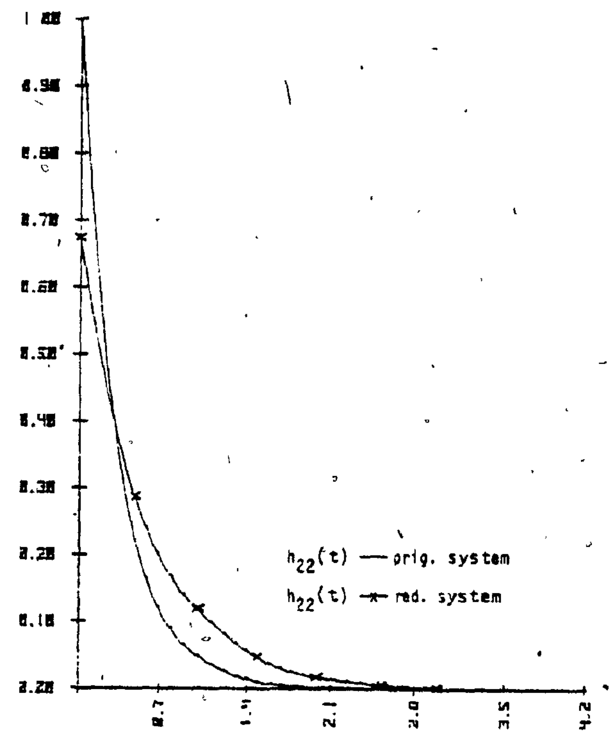


1.05
0.90
0.75
0.60
0.45
0.30
0.15
0.00
1.0
2.0
3.0
4.0
5.0
6.0
h11(t) — orig. system
h11(t) -x- red. system
0.54
0.45
0.36
0.27
0.18
0.09
0.00
1.0
2.0
3.0
4.0
5.0
6.0
7.0
h12(t) — orig. system
h12(t) -x- red. system

FIG. 3.3



$h_{21}(t)$ — orig. system
$h_{21}(t)$ —x— red. system



$h_{22}(t)$ — orig. system
$h_{22}(t)$ —x— red. system

FIG. 3.4

$$A_r = \begin{bmatrix} -.5348654 & 0. & 0. & 0. & 0. \\ 0. & -.8878999 & 0. & 0. & 0. \\ 0. & 0. & -.1045389E+01 & 0. & 0. \\ 0. & 0. & 0. & -.7297035 & 0. \\ 0. & 0. & 0. & 0. & -.1348871E+01 \end{bmatrix}$$

$$B_r = \begin{bmatrix} -.5933413E+01 & .7664345E+01 \\ -.8605920E+01 & .8542592E+01 \\ .7170082E+01 & -.4327124E+01 \\ .1915893E+01 & -.2743977E+01 \\ .2593567E+02 & -.4105723E+01 \end{bmatrix}$$

$$C_r = \begin{bmatrix} -.7398051 & -.8221810E+01 & -.8238091E+01 & -.1544721E+02 & .5215084 \\ .8179787 & -.5298589E+01 & -.5314049E+01 & -.1131270E+02 & .3594887 \end{bmatrix}$$

TABLE 3.3

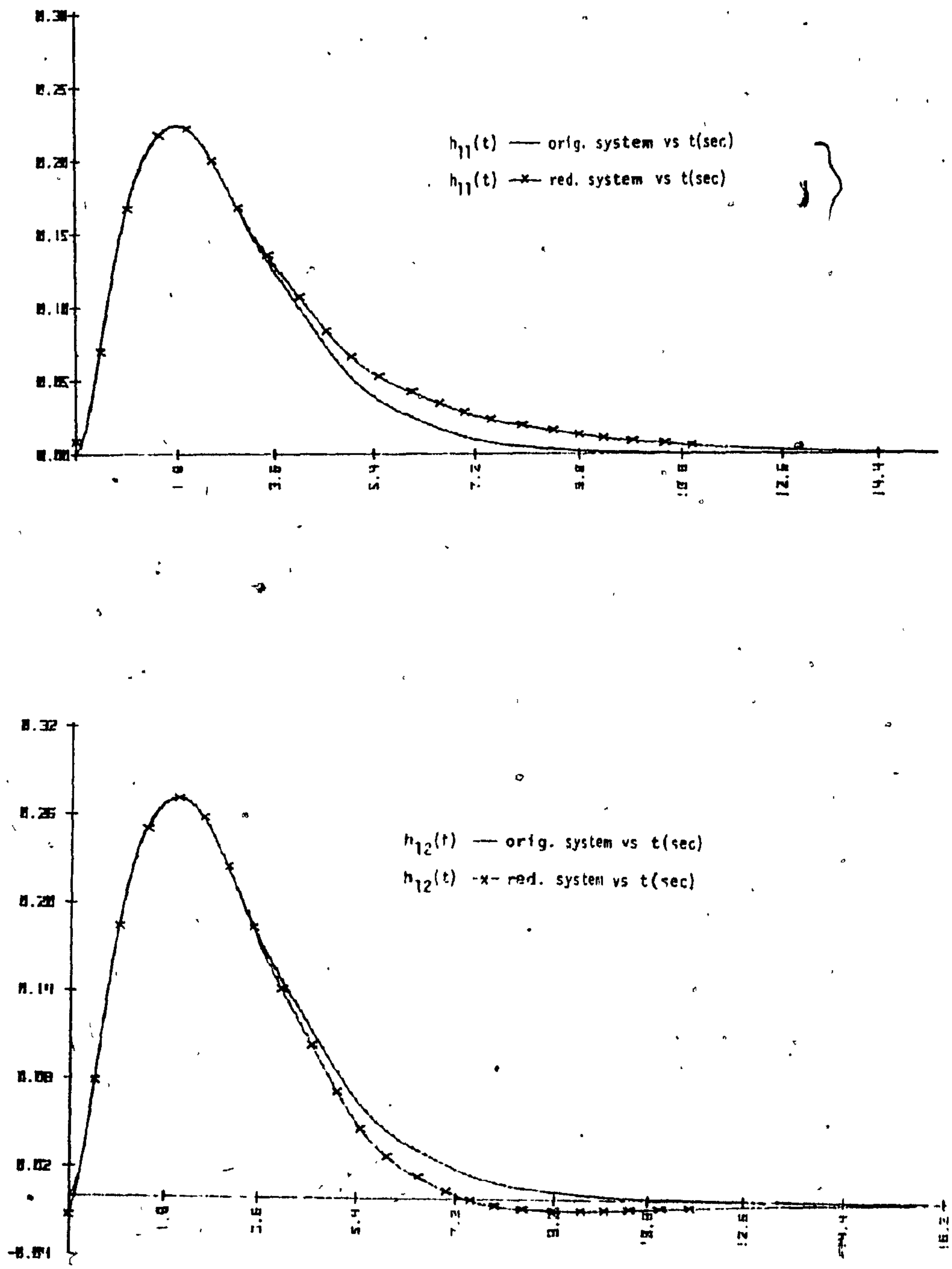

$h_{11}(t)$ — orig. system vs t(sec)
$h_{11}(t)$ -x- red. system vs t(sec)
$h_{12}(t)$ — orig. system vs t(sec)
$h_{12}(t)$ -x- red. system vs t(sec)

FIG. 3.5

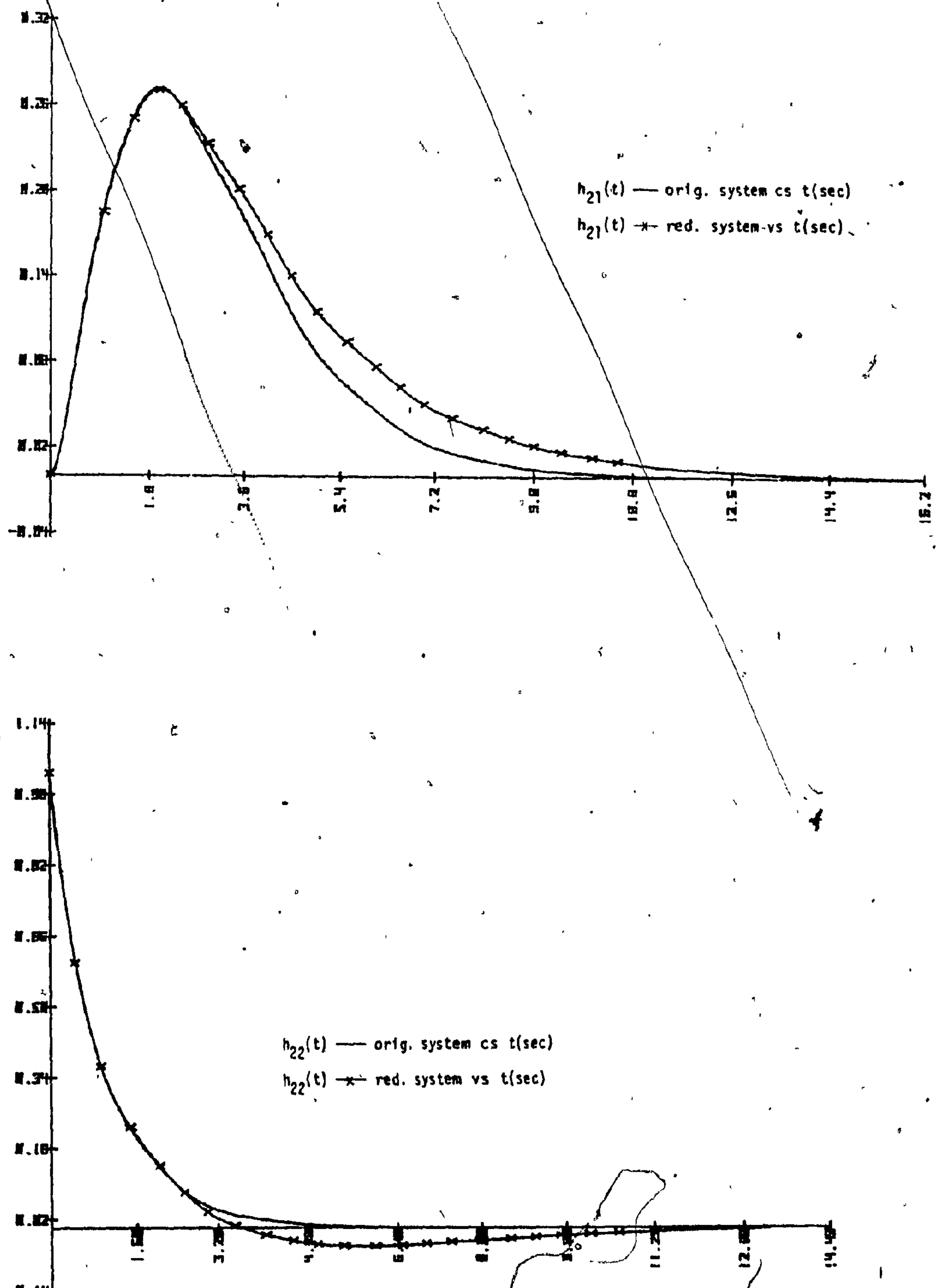

$h_{21}(t)$ — orig. system cs t(sec)
$h_{21}(t)$ -x- red. system-vs t(sec)
1.8
3.6
5.4
7.2
9.0
10.8
12.6
14.4
16.2
$h_{22}(t)$ — orig. system cs t(sec)
$h_{22}(t)$ -x- red. system vs t(sec)

FIG. 3.6

$$\begin{bmatrix} \frac{3(s+3)(s+5)}{(s+1)(s+2)(s+4)} & \frac{6(s+1)}{(s+2)\ (s+4)} & \frac{2s+7}{(s+3)(s+4)} & \frac{2s+5}{(s+2)(s+3)} \\ \frac{2}{(s+3)(s+5)} & \frac{1}{(s+3)} & \frac{2(s-5)}{(s+1)(s+2)(s+3)} & \frac{8(s+2)}{(s+1)(s+3)(s+5)} \\ \frac{2(s^2+7s+18)}{(s+1)(s+3)(s+5)} & \frac{2s}{(s+1)\ (s+3)} & \frac{1}{(s+3)} & \frac{2(5s^2+27s+34)}{(s+1)(s+3)(s+5)} \end{bmatrix}$$

TABLE 3.4

- 59 -

A_r =

-.9549936	0.	0.	0.	0.	0.	0.
0.	-.10[illegible]449E+01	0.	0.	0.	0.	0.
0.	0.	-.9188060	0	0.	0.	0.
0.	0.	0.	-.6[illegible]898	0.	0.	0
0.	0.	0.	0.	-.4398299E+01	0.	0.
0.	0.	0.	0.	0.	-.4891157E+01	0.
0.	0.	0.	0.	0.	0.	-.4506368E+01

B_r =

.73[illegible]8964E+01	-.2687[illegible]1E+01	-.1522361	-.1449836E+01
.3107887E+01	-.1784647E.01	-.1101949E+01	-.1164601E+01
.1607332E+01	.4437802	.3025[illegible]48E+01	.1021676E+01
.1069377E+01	.6471428	-.6641115	.4444452
.3476769	.3957664	-.1459830	-.2504654E+0[illegible]
-.3835976	.1090012E+01	-.9572694	-.1163615E+01
-.7477381	.2092929E+01	.6773049	.6856783

C_r =

.4496871	.113609[illegible]	.4196122	.1008619E+01	-.6698834E+01	.5468542	.3119614E+01
-.2294246	.15[illegible]382[illegible]	.6222282	-.[illegible]8952	-.233746[illegible]	.1496924E+01	-.309463[illegible]
-.3747013	.1144729E+01	.5981412[illegible]	.[illegible]30184	-.2618372E-01	-.635314	-.1626[illegible]

TABLE 3.5

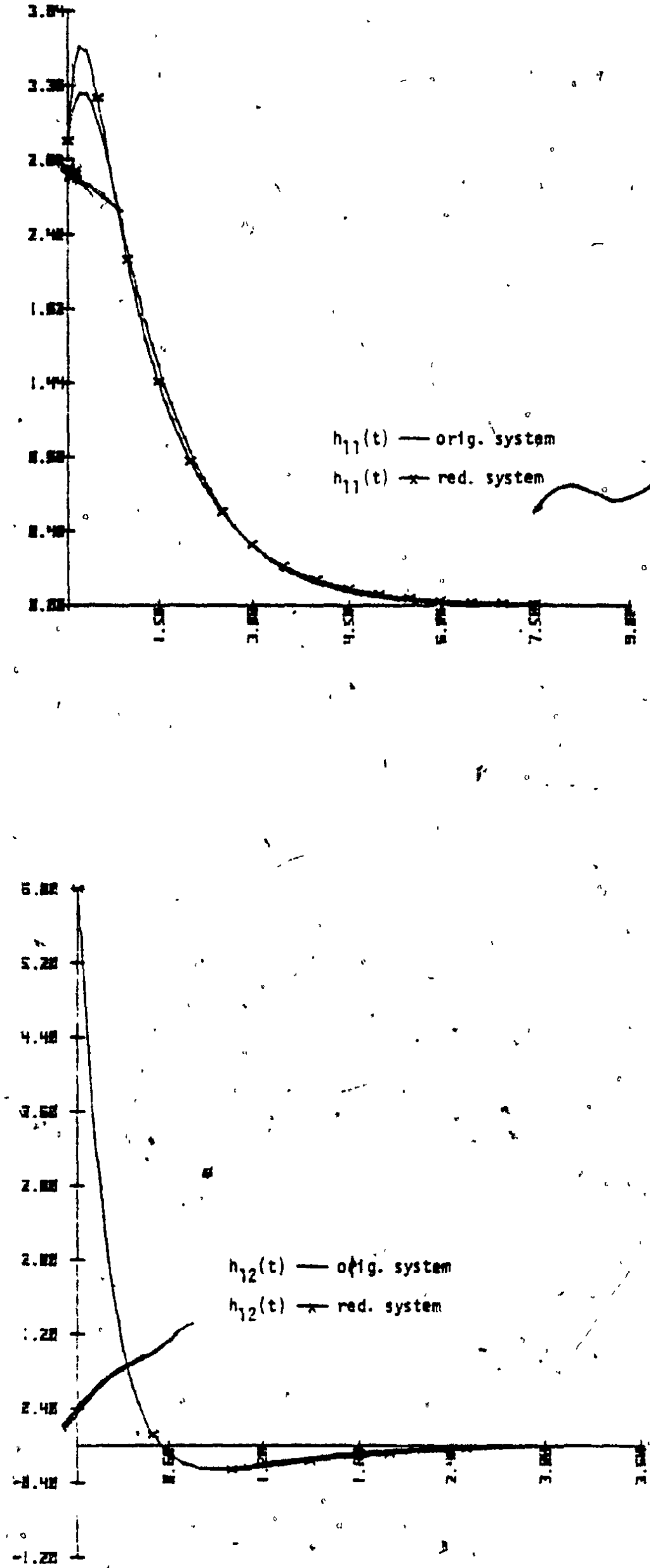

h11(t) — orig. system
h11(t) —x— red. system
h12(t) — orig. system
h12(t) —x— red. system

FIG. 3.7

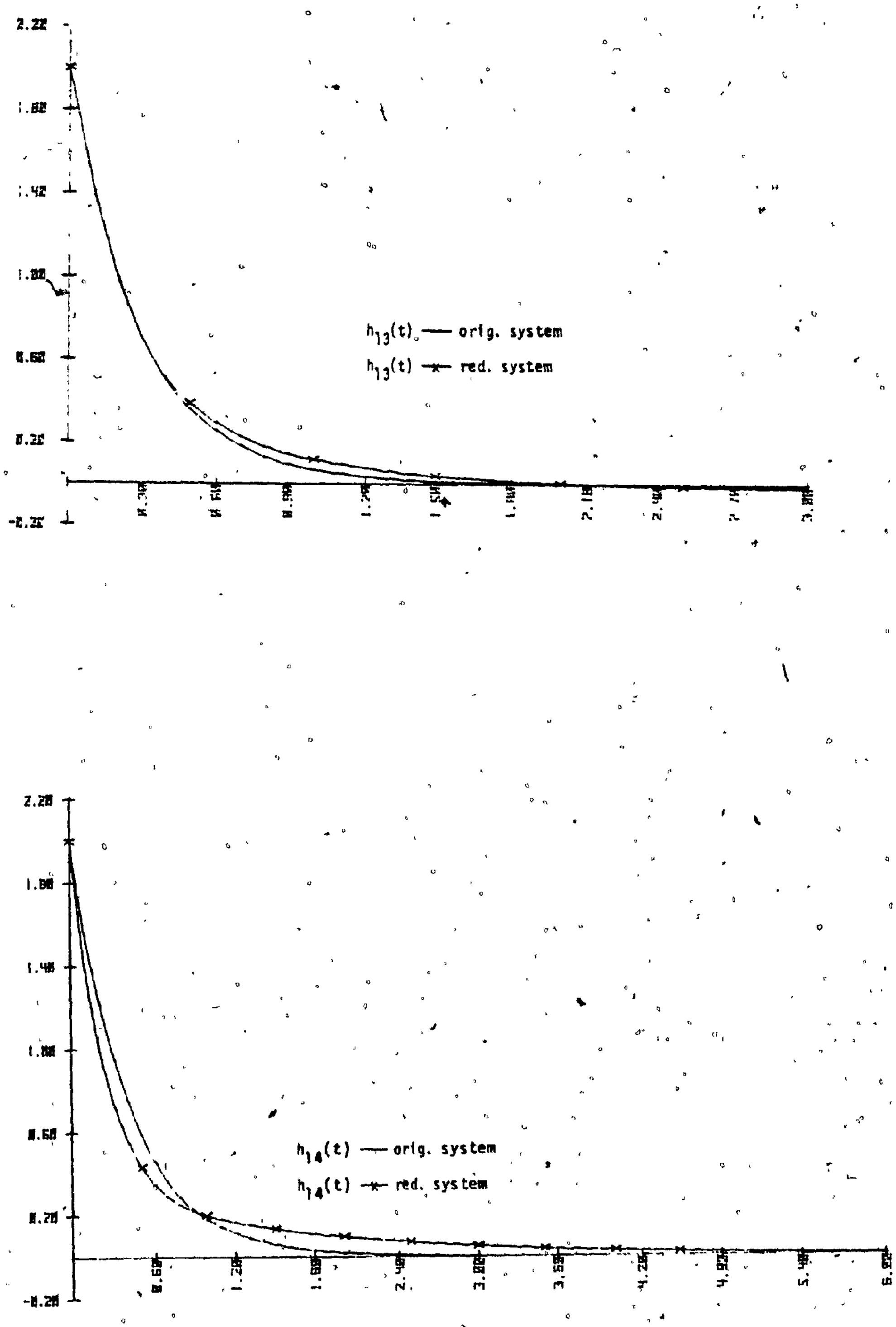
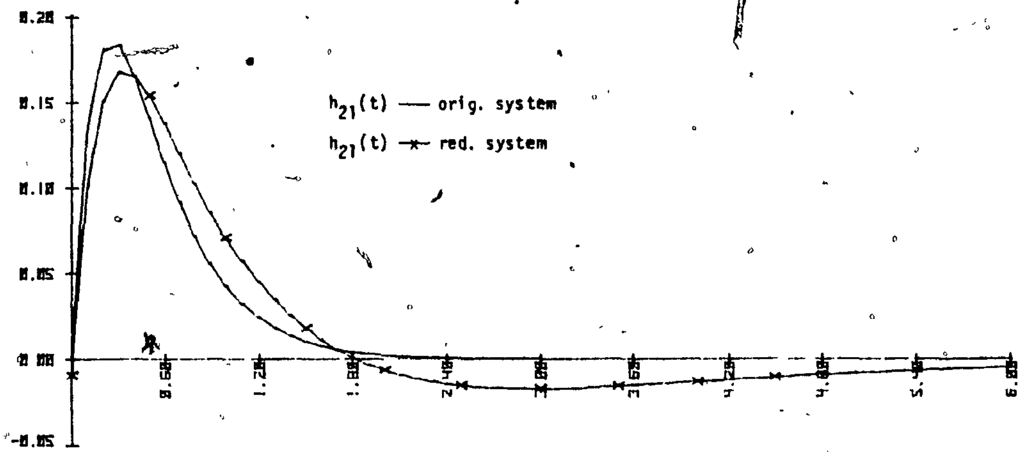

h13(t) —— orig. system
h13(t) —x— red. system
h14(t) —— orig. system
h14(t) —x— red. system

FIG. 3.8


0.20
0.15
0.10
0.05
0.00
-0.05
$h_{21}(t)$ —— orig. system
$h_{21}(t)$ —x— red. system

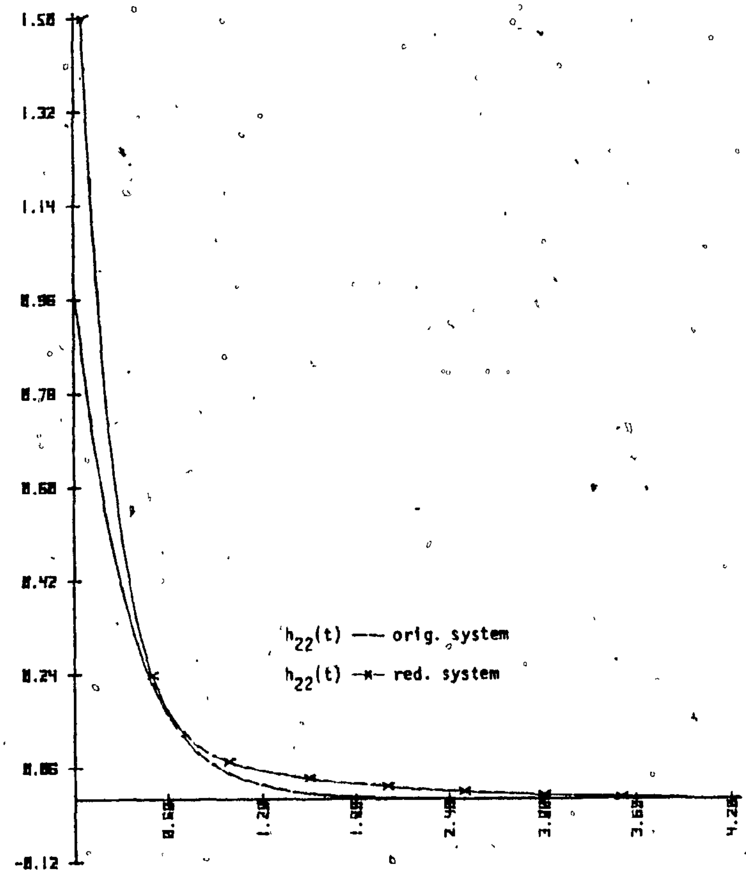

1.50
1.32
1.14
0.96
0.78
0.60
0.42
0.24
0.06
-0.12
0.60
1.20
1.80
2.40
3.00
3.60
4.20
$h_{22}(t)$ —— orig. system
$h_{22}(t)$ —x— red. system

FIG. 3.9

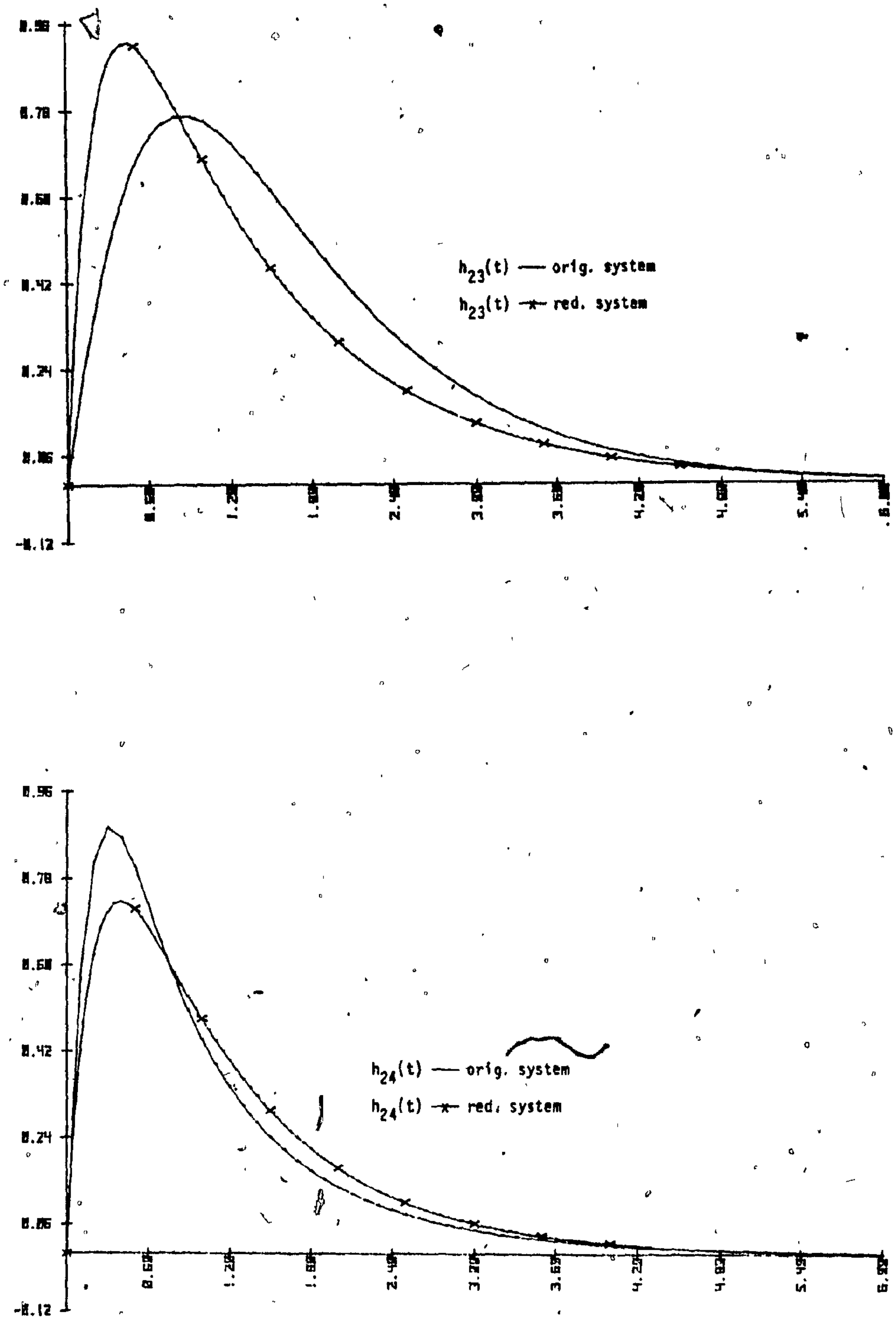

h23(t) —— orig. system
h23(t) —x— red. system
h24(t) —— orig. system
h24(t) —x— red. system

FIG. 3.10

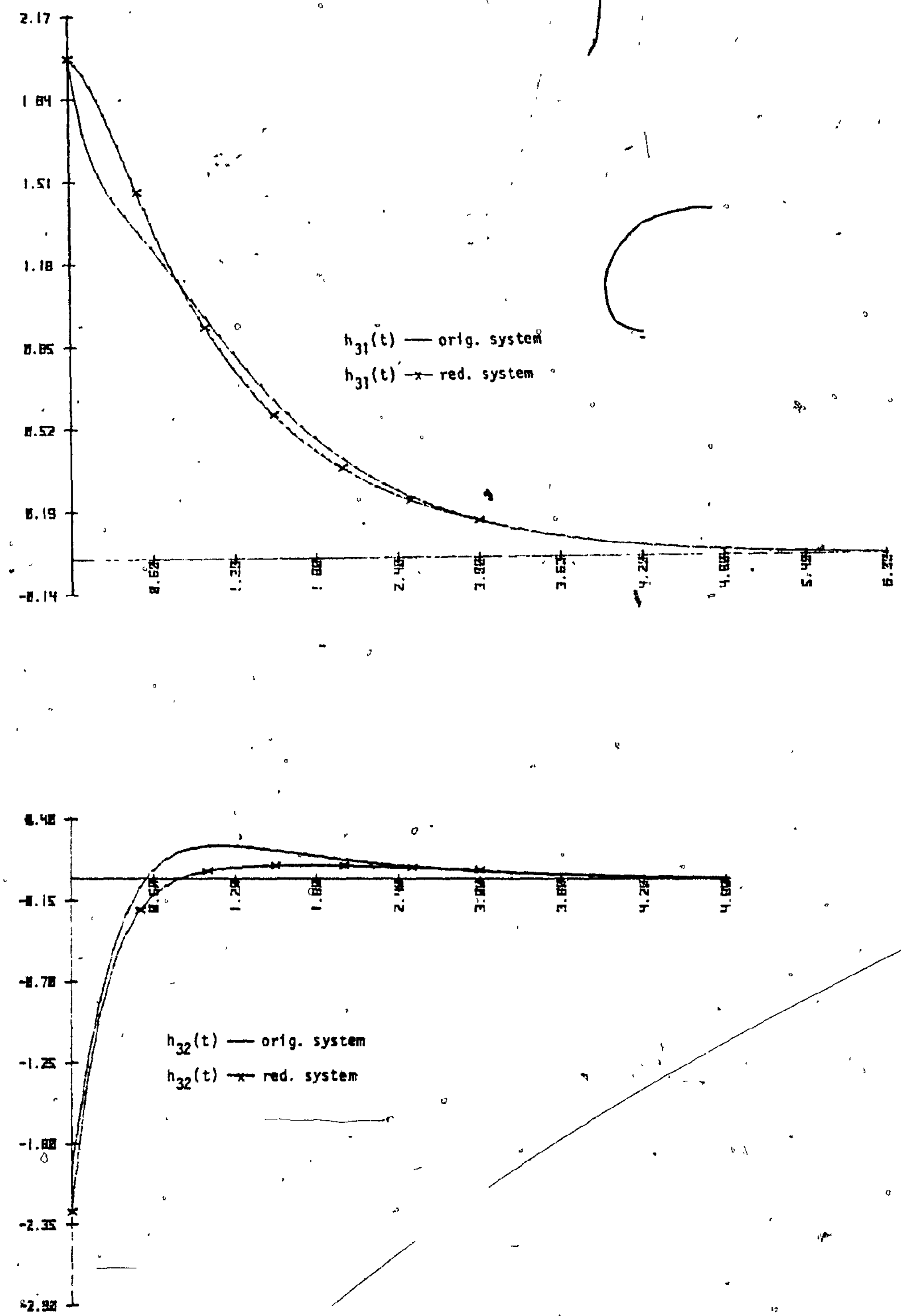
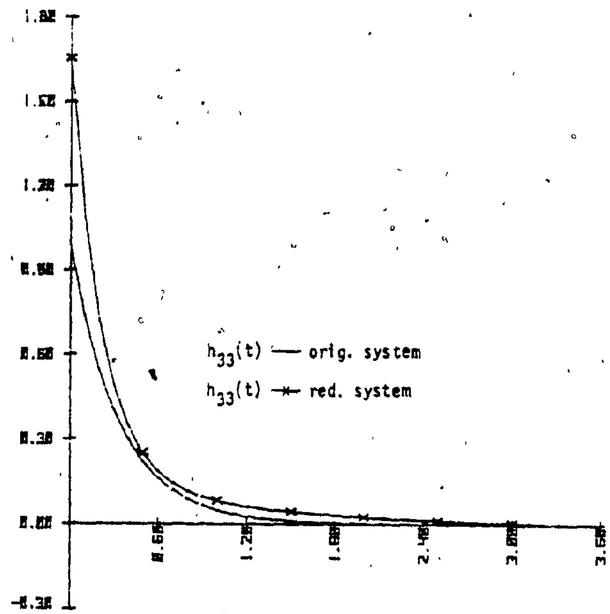

h31(t) — orig. system
h31(t) -x- red. system
h32(t) — orig. system
h32(t) -x- red. system

FIG. 3.11


h33(t) — orig. system
h33(t) —x— red. system

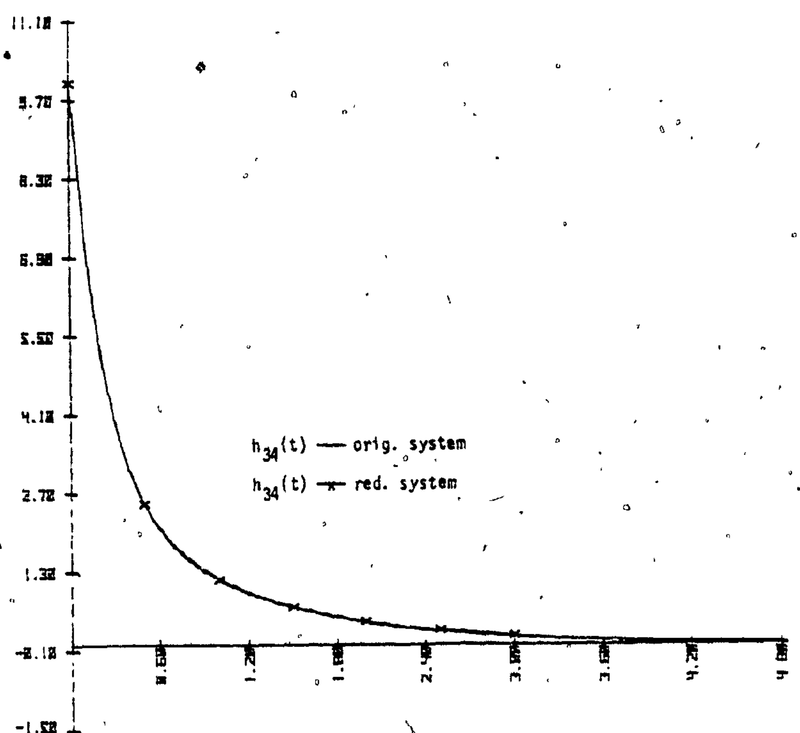

h34(t) — orig. system
h34(t) —x— red. system

FIG. 3.12

comment is appropriate at this point:

It is clear that the quality of the approximation depends on the dimension of the reduced model. However several of models for different values of r were tried and the ones shown in the above examples were chosen for their satisfactory approximations. From the computational point of view the selection of the starting parameters (A_r^0, B_r^0, C_r^0) are critical (e.g. increasing the computes timer used). For S.I.S.O. [21] uses λ_i (eigenvalues of A_r) $\{\forall i = 1 \ldots r\}$ to be $\lambda_i \subset \tilde{\lambda}_J$ where $j = \{1 \ldots n\}$ and $\tilde{\lambda}_J$ are the dominant eigenvalues of A. We found that for the M.I.M.O. case the choice of $\lambda_i < \tilde{\lambda}_J$ $\forall i$ gives better numerical performance and a substantial saving in CPU time.

3.5 CONCLUSIONS

The procedure proposed in this chapter has the following advantages over some of the existing methods:

1. R O.M. is optimal, in the sense that J_1 is minimized.
2. R.O.M.'s can be obtained without depending on a priori knowledge of the nature of the inputs to the system.
3. For a stable system, it always yields a stable R.O.M.
4. The R.O.M. can have real, or complex eigenvalues or a combination of both.
5. The starting parameters (A_r^0, B_r^0, C_r^0) need not be controllable e.g. $(A_r^0 B_r^0)$, to ensure an optimal R.O.M. at the end of the computation. The choice of Real $\{\lambda^0_{r_i}\} < 0$, where $\lambda^0_{r_i}$ are the eigenvalues or A_r^0 is the only requirement to obtain stable meaningful R.O.M.'s.

CHAPTER 4

APPLICATIONS OF ORDER REDUCTION

4.0 INTRODUCTION

In this chapter the method for obtaining sub-optimal control policies for the state linear regulator problem using the aggregation scheme of Aoki [1] is reviewed. In general the aggregation scheme cannot yield an exact solution and therefore the obvious approach is to find an approximate aggregation scheme. With the results of Aoki and the approximate aggregation scheme, a sub-optimal control policy, for the M.I.M.O. case, using reduced order models, is presented. By the use of the concept of disaggregation [2], in conjunction with the proposed method for R.O.M.'s of Chapter III, a procedure for obtaining a sub-optimal Wiener Kalman Filter for M.I.M.O. stationary systems is proposed and the technique is illustrated with an example. In addition the degradation in performance or loss factor for the sub-optimal filter, as compared to the optimal Wiener-Kalman estimator, is derived. Finally in section 4.4, using the WU [52] and Rao [46] transformations, the disaggregation scheme presented in the previous section and the method of Chapter III for obtaining R.O.M's of L.T.I. systems, a procedure is proposed in order to obtain R.O.M.'s for a class of linear time varying M.I.M.O. systems.

4.1 SUB-OPTIMAL STATE AND OUTPUT LINEAR REGULATORS

Problem 4.1 (State Regulator)

Consider a linear time invariant system described by

$$\dot{x}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{U}(t) \qquad (4.1)$$

where $\underset{\sim}{x}$, is a vector of dimension n and A,B are constant matrices of appropriate dimensions. It is desired to find a control $\underset{\sim}{\overline{U}}(t)$ that transfers the original state $\underset{\sim}{x}(0)$ to the final state $\underset{\sim}{x}(\infty) = 0$, such that the following cost function is minimized

$$J = \frac{1}{2} \int_0^\infty \{\|\underset{\sim}{x}(t)\|_Q^2 + \|\underset{\sim}{U}(t)\|_R^2\} \, dt \tag{4.2}$$

where Q,R are positive semi definite and positive definite constant matrices. The solution of this problem is well known [48] and involves the solution of the matrix Riccati equation of the form

$$A'P + PA - PBR^{-1}B'P = -Q \tag{4.3}$$

and the optimal control is given by

$$\underset{\sim}{\overline{U}}(t) = -R^{-1}B^TB'Px(t). \tag{4.4}$$

A sub-optimal control can be obtained using aggregation methods. Following Aoki's procedure, consider the system described by

$$\dot{\underset{\sim}{x}}_r(t) = A_r\underset{\sim}{x}_r(t) + B_r\underset{\sim}{U}(t) \tag{4.5}$$

where $\underset{\sim}{x}_r$ is a vector of dimensions r, $A_r \in R^{rxr}$ $B_r \in R^{rxm}$ and $r < n$, and moreover $\underset{\sim}{x}_r$ is an approximation to $\underset{\sim}{x}$. Suppose that it is desired to design a state regulator for (4.5). Then the cost function to be minimized is

$$J_r = \tfrac{1}{2} \int_0^{\infty} \{\| \underset{\sim}{x}_r(t)\|^2_{Q_r} + \| \underset{\sim}{U}(t)\|^2_R\} \, dt \qquad (4.6)$$

and the Riccati Eq. and the optimal control are respectively

$$A_r'P_r + P_rA_r - P_rB_rR^{-1}B_rP_r = -Q_r, \qquad (4.7)$$

$$\overline{\underset{\sim}{U}}_r(t) = R^{-1}B_r'P_r\underset{\sim}{x}_r(t) \,. \qquad (4.8)$$

Now suppose that a matrix $M \in R^{rxn}$ exists such that

$$\underset{\sim}{x}_r(t) = M\underset{\sim}{x}(t) \,. \qquad (4.9)$$

Then clearly from (4.1) and (4.5) we have

$$A_rM - MA = 0 \qquad (4.10)$$

$$B_r - MB = 0 \,. \qquad (4.11)$$

If an M does not exist such that (4.10) and (4.11) are satisfied namely $\underset{\sim}{x}_r(t) \cong M\underset{\sim}{x}(t)$, then clearly $\overline{M}$ can be obtained by finding an approximate solution of (4.10) and (4.11). El Attar [21] proposed that it can be accomplished with the use of the ℓ_1-norm algorithm for solving over determined set of linear equations due to Barrodale and Roberts [11].

Choosing Q_r as

$$Q_r = (\overline{M}\overline{M}')^{-1} \, \overline{M}Q\overline{M} \, (\overline{M}\overline{M}')^{-1}, \qquad (4.12)$$

minimizing the cost function (4.6) and using (4.10-4.11) we can obtain a sub-optimum control of the form

$$\underset{\sim}{\overline{U}}_s(t) = -R^{-1}B_r'P_r, \widetilde{M}\underset{\sim}{x}_r(t) \cdot \tag{4.13}$$

These results shall be directly extended to the output regulator.

Problem 3.2 (Output Regulator)

Consider the L.T.I. M.I.M.O. system described by

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{u}(t),$$

$$\underset{\sim}{y}(t) = C\underset{\sim}{x}(t) \tag{4.14}$$

where $\underset{\sim}{x}, \underset{\sim}{y}$, are vectors of dimension n, ℓ respectively and $A \in R^{nxn}$, $B \in R^{nxm}$, $C \in R^{\ell xr}$. It is desired to find a control $\underset{\sim}{\overline{U}}(t)$ that transfers the original state $\underset{\sim}{x}(0)$ to the final state $\underset{\sim}{x}(\infty) = 0$ such that the following cost function is minimized

$$J = \tfrac{1}{2} \int_0^\infty \{\| \underset{\sim}{y}(t)\|_Q^2 + \| \underset{\sim}{U}(t)\|_R^2\} \, dt \cdot \tag{4.15}$$

This involves solving the following Ricatti Eq.

$$TA + A'T + TBR^{-1}B'T - C'RQC = 0 \cdot \tag{4.16}$$

Then the optimal control $\underset{\sim}{\overline{U}}$ is

$$\underset{\sim}{\overline{U}}(t) = -R^{-1}B'T\underset{\sim}{x}(t) \cdot \tag{4.17}$$

Proceeding in a manner similar to problem 4.1, consider the lower order system described by

$$\dot{x}_r(t) = A_r x_r(t) + B_r U(t),$$
$$y_r(t) = C_r x_r(t), \qquad (4.18)$$

where x_r and y_r are vectors of dimension r and ℓ respectively and $A_r \in R^{rxn}$, $B_r \in R^{rxm}$, $C_r \in R^{\ell xr}$, $r < n$. It is desired to design an output regulator for (4.18). The solution of the following equations yield the optimal control $\overline{U}_r(t)$.

$$J_r = \int_0^\infty \{\| y(t)\|^2_{Q_r} + \| U(t)\|^2_R\} \, dt, \qquad (4.19)$$

$$T_r A_r + A_r' T + T_r B_r R^{-1} B_r' T_r - C_r' Q_r C_r = 0, \qquad (4.20)$$

$$\overline{U}_r(t) = -R^{-1} B_r' T_r x_r(t) . \qquad (4.21)$$

Now suppose that a matrix $M \in R^{rxn}$ exists such that

$$x_r(t) = Mx(t), \qquad (4.22)$$

then from (4.22), (4.14) and (4.18) we have

$$A_r M - MA = 0, \qquad (4.23)$$

$$B_r - MB = 0, \qquad (4.24)$$

$$C - C_r M = 0. \qquad (4.25)$$

Solving this overdetermined set of equations with the Barrodale [10] algorithm we get $\bar{M} \simeq M$. Substituting (4.23 - 4.25) in (4.16) and (4.20) and comparing, then (4.20) becomes

$$T_r A_r + A_r' T_r + T_r B_r R^{-1} B' T_t - C_r' Q\, C = 0, \tag{4.26}$$

and a sub-optimal control $\bar{U}_s(t)$ can be obtained as

$$\bar{\underset{\sim}{U}}_s(t) = -R^{-1} B_r' P_r \bar{M} \underset{\sim}{x}_r(t) \tag{4.27}$$

Based on the above results, we propose a procedure to obtain a sub-optimal control law for the linear output regulator problem as follows:

I. Find a M.I.M.O. reduced order Model (4.18) of (4.14) by using the method proposed in Chapter 2.

II Solve the overdetermined set of equations (4.23 - 4.25) using [10], call the solution $\bar{M}$.

III. Solve the lower dimensional Ricatti equation (4.26).

IV. Calculate the sub-optimal control $\bar{\underset{\sim}{U}}_s(t)$ (4.27) using the results of II, III.

4.2 SUB-OPTIMAL ESTIMATOR

Consider the following two systems

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{U}(t), \tag{4.28}$$

$$\dot{\underset{\sim}{x}}_r(t) = A_r \underset{\sim}{x}_r(t) + B_r \underset{\sim}{U}(t), \tag{4.29}$$

defined as in problem 4.1. Assume that a matrix $M \in R^{rxn}$ exists such that $\underset{\sim}{x}_r = M\underset{\sim}{x}$ as in the previous section, then

$$MA = A_r M, \tag{4.30}$$

$$MB = B_r \cdot \tag{4.31}$$

A_r can be uniquely defined (Aoki [2]) by

$$A_r = MAM' (MM')^{-1} \cdot \tag{4.32}$$

that is (4.29) can be obtained by perfect aggregation of (4.28). Consider now the inverse problem. Knowing the state of the aggregate systems, $\underset{\sim}{x}_r(t)$, can we estimate or reconstruct the state $\underset{\sim}{x}(t)$ of the system (4.27)? This problem is known as disaggregation and was studied by Aoki [1], [3] and other researchers. Aoki shows that perfect disaggregation can be achieved if the matrix A of (4.28) has a special structure namely

$$A = MDE \cdot \tag{4.33}$$

Let $\hat{M}$ be a disaggregation matrix such that

$$\underset{\sim}{x}(t) = \hat{M}\underset{\sim}{x}_r(t) \cdot \tag{4.34}$$

If $\hat{M}$ is defined to be

$$\hat{M} = D(MD)^{-1}, \tag{4.35}$$

where

$$A_r M = MA = MDEM, \tag{4.36}$$

and the matrix E is the unknown to be determined, then $\underset{\sim}{x}(t)$ can be reconstructed from $\underset{\sim}{x}_r(t)$ as follows:

$$\underset{\sim}{x}(t) = \hat{M}\underset{\sim}{x}_r(t) + [I_n - \hat{M}M]\, B\underset{\sim}{U}(t) \cdot \tag{4.37}$$

The natural application of the disaggregation scheme is in the filtering problem specifically in the Wiener-Kalman estimator. The computational burden of the Wiener-Kalman filter is well known and several methods are proposed in the literature to alleviate the computational effort. Aoki and Hudle [4] propose a procedure to obtain the Weiner-Kalman estimator from a low order aggregated model and use the disaggregation scheme to obtain higher order estimates, in particular for discrete time systems. Furthermore, the high dimensional system is assumed to have a special structure (4.33). The continuous version of the Aoki-Hudle filter was presented by Newman [34] where he concluded that the solution of this continuous filter implied a formidable effort, even for the simplest problem. In the next section we propose a simple but efficient procedure to obtain a sub-optimal continuous Wiener-Kalman estimator for the stationary case.

4.3 SUB-OPTIMAL WIENER-KALMAN FILTER

Consider the following process model

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{U}(t), \tag{4.38}$$

with measurements

$$y(t) = Cx(t) + V, \tag{4.39}$$

$x \in R^n$, $y \in R^\ell$, $A \in R^{nxn}$, $B \in R^{nxm}$, $C \in R^{\ell xr}$.

with noise characteristics

$$E[U(t)U'(t)] = Q\sigma(t-\tau) , \tag{4.40}$$

$$E[V(t) V'(\tau)] = R\sigma(t-\tau), \tag{4.41}$$

$$E[U(t) V'(\tau)] = 0 , \tag{4.42}$$

where E is the expectation operator, Q, R are positive semi-definite and positive definite symmetric matrices respectively, V is a white noise process and σ is the Dirac function.

First suppose that A,B,C,Q,R are time varying matrices. Then it is well known (Kalman [30]) that the optimal estimator $\hat{x}(t)$ of $x(t)$ is given by

$$\dot{\hat{x}}(t) = A\hat{x}(t) + K(t) [y(t) - C\hat{x}(t)] . \tag{4.43}$$

The error covariance propagation $P(t)$ is the solution of the following Ricatti equation

$$\dot{P}(t) = AP(t) + P(t) A' + BQB' - P(t) C'R^{-1} CP(t), \tag{4.44}$$

and the Kalman gain $K(t)$ is

$$K(t) = P(t)\, C'R^{-1} \,. \tag{4.45}$$

Now consider the stationary case namely where A,B,C,Q,R, are constant matrices. The filtering may reach a steady state whenever P is constant, $\dot{P} = 0$ and (4.44) becomes

$$AP + PA' + BQB' - PC'RCP = 0 \,. \tag{4.46}$$

The estimator and the Kalman gain are respectively

$$\dot{\hat{\underset{\sim}{x}}}(t) = A\hat{\underset{\sim}{x}}(t) + K[\underset{\sim}{y}(t) - C\hat{\underset{\sim}{x}}(t)], \tag{4.47}$$

$$K = PC'R^{-1} \,. \tag{4.48}$$

Consider the following system

$$\dot{\underset{\sim}{x}}_r(t) = A_r \underset{\sim}{x}_r(t) + B_r \underset{\sim}{U}(t), \tag{4.49}$$

with measurements

$$\underset{\sim}{y}_r(t) = C_r \underset{\sim}{x}_r(t) + V, \tag{4.50}$$

where $A_r \in R^{rxn}$, $B_r \in R^{rxm}$, $C_r^{\ell xr}$, $r < n$, and with noise characteristics given by equations (4.40, 4.42). Let us set $V = 0$ in (4.50) and suppose that there exists a constant matrix (disaggregation matrix) S such that

$$S\underset{\sim}{x}_r(t) = \underset{\sim}{x}(t) \cdot \tag{4.51}$$

Then by (4.38), (4.39) and (4.49), (4.50) we have

$$SA_r = AS = 0, \tag{4.52}$$

$$SB_r - B = 0, \tag{4.53}$$

$$CS - C_r = 0. \tag{4.54}$$

Solving these equations approximately, we arrive at an S such that $S\underset{\sim}{x}_r \cong \underset{\sim}{x}$. Now let $V \neq 0$, then the error covariance matrix for system (4.49), (4.50) is

$$A_rP_r + P_rA_r' + B_rQ_rB_r' - P_rC_r'R^{-1}C_rP_r = 0 \cdot \tag{4.55}$$

Premultiplying (4.55) by S and post multiplying by S' we have

$$SA_rP_rS' + SP_rA_r'S' + \hat{Q}_r - SP_rC_r'R^{-1}CP_rS' = 0 \tag{4.56}$$

(4.56)

where $\hat{Q}_r = SB_rQ_rB_r'S' \cdot$ (4.57)

By comparison of (4.56) with (4.46) and using the relations (4.52-4.54) we find that solving (4.55) for an appropriate choice of $\hat{Q}_r$, in particular choosing $\hat{Q}_r$ to be

$$\hat{Q}_r = (S'S)^{-1} BQB'S(S'S)^{-1}, \tag{4.58}$$

we can obtain a approximate solution to the high order Ricatti equation (4.46), where the relationship between the lower and the high order Ricatti equations is

$$S P_r S' = \hat{P} \cong P \cdot \qquad (4.59)$$

Therefore the sub-optimal Wiener-Kalman filter is

$$\dot{\tilde{x}} = A\tilde{x}(t) + \hat{K}[y(x) - C\tilde{x}(t)] \cdot \qquad (4.60)$$

From the above equations we can see that if $\hat{P}$ is a good approximation to P then $\hat{K}$ approximates K and in turn $\tilde{x}(t)$ approximates $\hat{x}(t)$.

Remark 4.1 Kalman [30] showed that complete observability is a sufficient condition for the existence of a steady state solution of (4.44) and furthermore that complete controllability will assure the uniqueness of the steady state solution. As a consequence the following question is proposed: Given a system as in (4.38-4.39) with $V = 0$, let Ψ be the set of reduced models which approximate x in some sense, e.g. all possible structures of the triple (A_r, B_r, C_r), with a state vector of dimension r. Suppose that $\exists S^i \in \Phi$, $I \triangleq \{i\}$, such that $S^i: x \to x_r$ $\forall x_r^i \in \Psi$. Then, letting the noise $V \neq 0$ and $\{\hat{K}_i\}$ be the set of Kalman gains corresponding to every S^i, if $\{\hat{K}_i\}$ in (4.47) satisfies the complete observability criterion $\forall i$, is $\hat{K}^J$, where $J \in I$, the best approximation to K (the optimal Kalman gain) such that (4.47) is absolutely controllable?

It is well known that optimality of the Kalman filter does not guarantee its stability, i.e. the solution of (4.44) leads to the optimum filter in the sense of minimum variance, but this filter can be unstable.

The impulse response matrices of (4.47), (4.60) are clearly,

$$H_1(t-\tau) = e^{(A-KC)(t-\tau)} K$$

$$H_2(t-\tau) = e^{(A-\hat{K}C)(T-\tau)} \hat{K} .$$

Then $H_2(\cdot)$ will be stable if $R_e\{\lambda_{B_2}\} \leq 0$, where $\{\lambda_{B_2}\}$ is the set of eigenvalues of $(A-\hat{K}C)$. Therefore this introduces a constraint in the determination of $\hat{K}$ and consequently in the selection of the reduced model. To obtain the optimal gain K, (4.44) has to be solved which implies solving $\frac{n x(n+1)}{2}$ nonlinear coupled equations. But for the sub-optimal gains only $\frac{r x(r+1)}{2}$ equations have to be solved, i.e. if $n=9$, $r=7$ then 28 nonlinear equations must be solved instead of 45. This clearly demonstrates the usefulness of the sub-optimal filter. The iterative procedure of the proposed method is summarized in the following steps:

Step I Disregard the noise i.e. set $V = 0$ in (4.39)(4.50).

Step II Find a reduced order model of the form

$\dot{\underset{\sim}{x}}_r = A_r \underset{\sim}{X}_r + B_r \underset{\sim}{U}$, $\underset{\sim}{y}_r = C_r \underset{\sim}{X}_r$ by using the proposed method of Chapter 3.

Step III Compute the disaggregation matrix S by solving Eqs. (4.52-4.54) Using [10].

Step IV Reintroduce the noise into the reduced model and solve the lower dimensionality Ricatti equation ($\frac{1}{2}r(r+1)$ simultaneous quadratic equations) for $\hat{Q}_r$ given by (4.58) and call it $\hat{P}_r$.

Step V. With P_r and (4.59) compute the approximate solution of the high order Ricatti equation and label it $\hat{P}$.

Step VI With $\hat{P}$ and (4.48), (4.60) compute the sub-optimal Kalman gain and the sub-optimal estimator.

4.4 DEGRADATION IN PERFORMANCE

Consider the state $\hat{x}(t)$ corresponding to the Wiener Kalman estimator (4.47) and $x(t)$ the state of the process to be estimated (4.38). As mentioned in the preceeding section, by solving (4.46) we can obtain the optimal estimator. However it is well known that in practice this is not true due to the fact that errors can arise from modelling or measurements in the dynamics of the system, i.e. an incorrect estimate of Q,R. Therefore the Wiener-Kalman filter is no longer optimal. Of the several sources of error, we are interested in the error due to the use of a sub-optimal Kalman gain. Let's define this error as follows;

$$\underset{\sim}{e}(t) = \underset{\sim}{x}(t) - \hat{\underset{\sim}{x}}(t) . \tag{4.61}$$

It is known (Meditch [33]) that the covariance matrix of this error $P_e = \text{cov}\,[\underset{\sim}{e}(t), \underset{\sim}{e}'(t)]$ is the solution of the following Ricatti equation

$$\dot{P}_e = A_e P_e + P_e A_e' + K_e \tag{4.62}$$

where

$$A_e = A - KC \tag{4.63}$$

$$K_e = KRK' + BQB' \tag{4.64}$$

Using (4.62) the degradation in performance may be determined in the following way; it is known [43] that if K_o is the optimum Kalman gain then K_o minimizes the following cost function, for every time t, resulting in the minimum variance filter

$$J(t) = tr[var\ \underset{\sim}{e}(t)] = tr\ [P_e(t)] \ . \tag{4.65}$$

This can be accomplished by minimizing

$$J'(t) = \frac{dJ(t)}{dt} = t_r\ [\dot{P}_e(t)] \ . \tag{4.66}$$

Now it is clear that for the stationary case, namely $t \to \infty$, (4.66) becomes $\dot{J}(\infty) = 0$. Then the index of degradation in performance for the stationary filter J_D can be calculated as follows

$$J_D = T_r[P_e],\ \dot{P}_e = 0, \tag{4.67}$$

where $J_D \geq 0$ and P_e is the solution of (4.62). Suppose that a gain $\hat{K}$ is used instead of K in order to determine the minimum variance filter, and $\hat{K}$ is the approximate Kalman gain obtained by the procedure proposed in Section (4.3), then the degradation in performance is

$$J_{D_r} = t_r\ [P_{e_r}], \tag{4.68}$$

where $J_{D_r} \geq J_D$ and P_{e_r} is the solution of the following linear matrix equations.

$$A_{e_r} P_{e_r} + P_{e_r} A'_{e_r} + K_{e_r} = 0 \qquad (4.69)$$

where

$$A_{e_r} = A - \hat{K}C \qquad (4.70)$$

$$K_{e_r} = \hat{K}R\hat{K}' + BQB' \qquad (4.71)$$

Now the degradation in performance J_{D_S} of the sub-optimal Wiener-Kalman filter using $\hat{K}$ compared to the optimal estimator using K can therefore be computed as follows

$$J_{D_S} = \frac{J_{D_r} - J_D}{J_D} \cdot 100\% \qquad (4.72)$$

In order to illustrate the proposed procedure, a numerical example is presented in the next section.

4.5 NUMERICAL EXAMPLE

Given the following M.I.M.O. system

$$\dot{\underset{\sim}{x}}(t) = A\underset{\sim}{x}(t) + B\underset{\sim}{U}(t), \qquad (4.73)$$

$$\underset{\sim}{y}(t) = C\,\underset{\sim}{x}(t) + V, \qquad (4.74)$$

with noise characteristics Q,R, and V is a white noise process, where these parameters and system are given in Table 4.1.

First, we find a reduced order model of the form

$$\dot{\underset{\sim}{x}}_r(t) = A_r\underset{\sim}{x}_r(t) + B_r\underset{\sim}{U}(t), \qquad (4.75)$$

$$\underset{\sim}{y}_r(t) = C_r\underset{\sim}{x}_r(t), \qquad (4.76)$$

Using the procedure presented in Chapter II by setting $v=0$ in (4.50). The parameters (A_r, B_r, C_r) of this reduced model are given in Table 4.2, where the error $\bar{J}_1^* = 4.139$. The disaggregation matrix S and the solution of the lower order Ricatti eq. P_r, are show in Table 4.3. With P_r and S we compute the approximate solution of (4.59) $\hat{\tilde{P}}$. In Table 4.4 it shows the matrix $\hat{P}$ and for the sake of clarity the solution of (4.55) the matrix P is given in Table 4.5. Now from (4.42) and (4.60) we can compute the optimal Kalman filter and the sub-optimal Kalman filter. Moreover, (4.47), (4.60) can be rewriting as follows:

$$\dot{\hat{\underset{\sim}{x}}}(t) = F\hat{\underset{\sim}{x}}(t) + K\underset{\sim}{U}(t) \tag{4.77}$$

and

$$\dot{\tilde{\underset{\sim}{x}}}(t) = \hat{F}\tilde{\underset{\sim}{x}}(t) + \hat{K}\underset{\sim}{y}(t), \tag{4.78}$$

where $F = A-KC$, $\hat{F}=A-\hat{K}C$ respectively. These results are shown in Tables 4.6 and 4.7. The degradation in performance due to the use of the sub-optimal Kalman gains respect to the optimal filter is

$$J_{D_3} = 1.705\% \tag{4.79}$$

In order to show graphically the performance of the sub-optimal estimator the following steps will apply.

Step I Set $U = \{1\cdot\sigma(t), 1\cdot\sigma(t) \ldots\}$ in (4.38)

Step II $\underset{\sim}{y}(t)$ obtaining from (4.39) wit $\underset{\sim}{v}=0$ was applied to the optimal and sub-optimal estimator.

$$A = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -3 \end{bmatrix}$$

$$B = \begin{bmatrix} 1 & 2 \\ 0.5 & 0 \\ 0 & -2 \\ -0.5 & 1 \end{bmatrix}$$

$$C = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}$$

$$Q = \begin{bmatrix} 10 & 0 \\ 0 & 10 \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

System Process Parameters

TABLE 4.1

$$A_r = \begin{bmatrix} -.89157712 & 0. \\ 0. & -.2735897 \end{bmatrix}$$

$$B_r = \begin{bmatrix} .42067 & .6140634 \\ .1111591 & .1343549E+01 \end{bmatrix}$$

$$C_r = \begin{bmatrix} .2119433E+01 & .9965355 \\ .2876781 & .5158636 \end{bmatrix}$$

TABLE 4.2

$$S = \begin{bmatrix} .2465401E+01 & -.3339363 \\ .2876781 & .1314821 \\ -.3459676 & .1330472E+01 \\ 0. & -.7442974 \end{bmatrix}$$

$$P_r = \begin{bmatrix} .1386112E+01 & -.1363597E+01 \\ -.1362597E+01 & .2809949E+01 \end{bmatrix}$$

TABLE 4.3

.1098203E + 02	.5489186	-.7057663E + 01	.3198761E + 01
.5489186	.6021077E-01	-.1059507	.1677000E - 01
-.7057663E + 01	-.1059507	.6394363E + 01	-.3133472E + 01
.3198761E + 01	.1677000E-01	-.3133472E + 01	.1556652E + 01

MATRIX $\hat{P}$

TABLE 4.4

$$
\begin{bmatrix}
.1237228E + 02 & .4439207 & -.8731095E + 01 & .3019777E + 01 \\
.4439207 & .7755319 & .5132501 & -.5945343 \\
-.8731095E + 01 & .5132501 & .7664106E + 01 & -.3377707E + 01 \\
.3019777E + 01 & -.5945343 & -.3377707E + 01 & .1814094E + 01
\end{bmatrix}
$$

MATRIX P

TABLE 4.5

$$K = \begin{bmatrix} .3641185E+01 & .3463698E+01 \\ .9571708 & .1809976 \\ -.1066989E+01 & -.2864457E+01 \\ -.3579300 & .1219560E+01 \end{bmatrix}$$

$$F = \begin{bmatrix} -.4641185E+01 & -.3463698E+01 & -.3641185E+01 & -.3463698E+01 \\ -.9571708 & -.1180998E+01 & -.9571708 & -.1809976 \\ .1066989E+01 & .2864457E+01 & -.9330110 & .2864457E+01 \\ .3579300 & -.1219560E+01 & .3579300 & -.4[illegible]19560E+01 \end{bmatrix}$$

Parameters of the optional Kalman estimator

TABLE 4.6

$$\hat{K} = \begin{bmatrix} .3924367E+01 & .3747679E+01 \\ .4429678 & .7698085E-01 \\ -.6633003 & -.3239422E+01 \\ .6528911E-01 & .1573422E+01 \end{bmatrix}$$

$$\hat{F} = \begin{bmatrix} -.4924367E+01 & -.3747679E+01 & -.3924367E+01 & -.3747679E+01 \\ -.4429678 & -.1076981E+01 & -.4429678E+01 & -.7698085E-01 \\ .6633003 & .3239422E+01 & -.1336700E+01 & .3229422E+01 \\ -.6528911E-01 & -.1573422E+01 & -.6528911E-01 & -.4573422E+01 \end{bmatrix}$$

Parameters of Sub-Optimal Kalman Estimator

TABLE 4.7

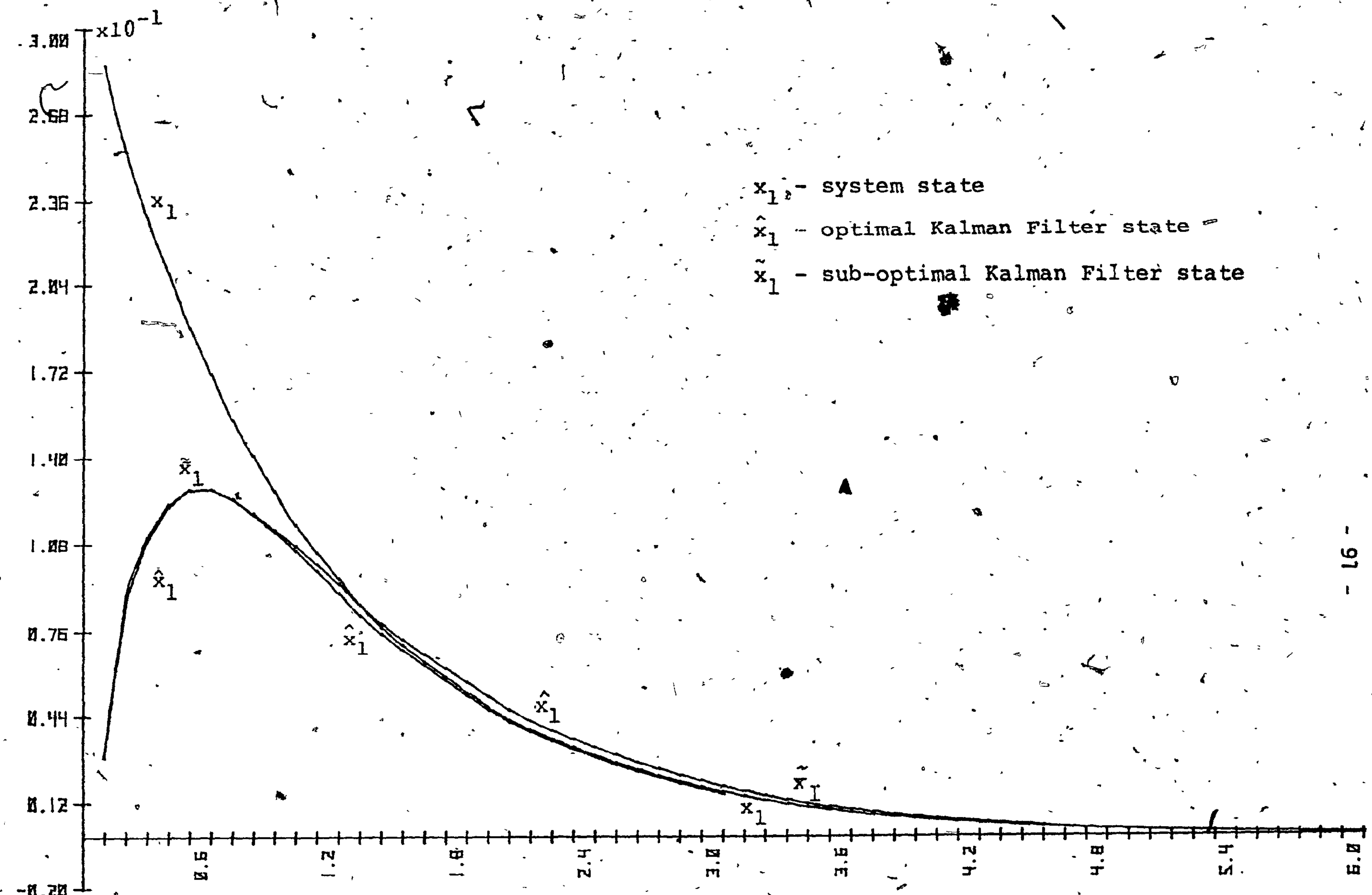

x10^-1
3.00
2.68
2.36
2.04
1.72
1.40
1.08
0.76
0.44
0.12
-0.20
0.6
1.2
1.8
2.4
3.0
3.6
4.2
4.8
5.4
6.0
x_1 - system state
$\hat{x}_1$ - optimal Kalman Filter state
$\tilde{x}_1$ - sub-optimal Kalman Filter state

Fig. 4.1

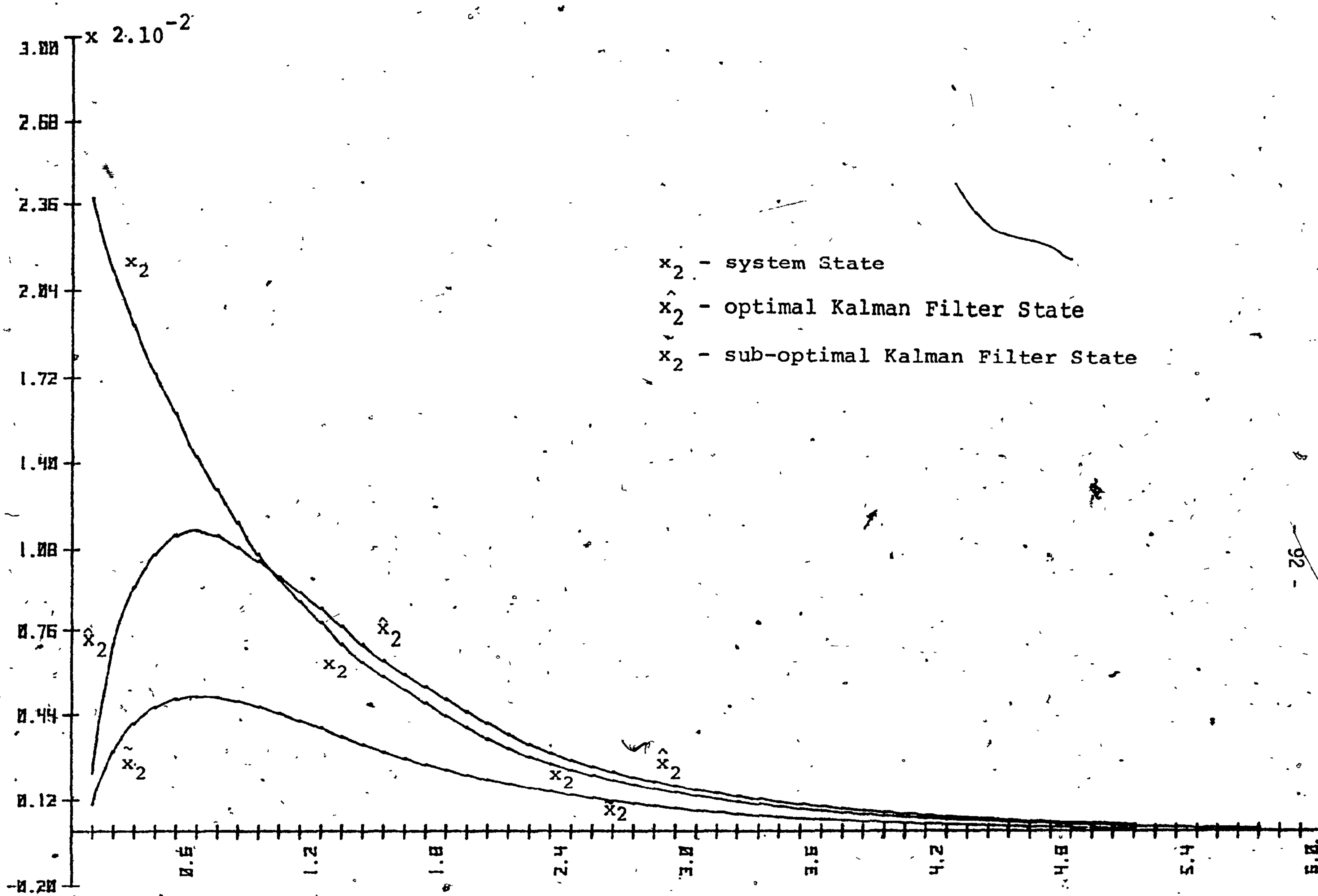

x 2.10^{-2}
3.00
2.68
2.36
2.04
1.72
1.40
1.08
0.76
0.44
0.12
-0.20
0.6
1.2
1.8
2.4
3.0
3.6
4.2
4.8
5.4
6.0
x_2 - system State
$\hat{x}_2$ - optimal Kalman Filter State
$\tilde{x}_2$ - sub-optimal Kalman Filter State
x_2
$\hat{x}_2$
$\tilde{x}_2$

FIG. 4.2

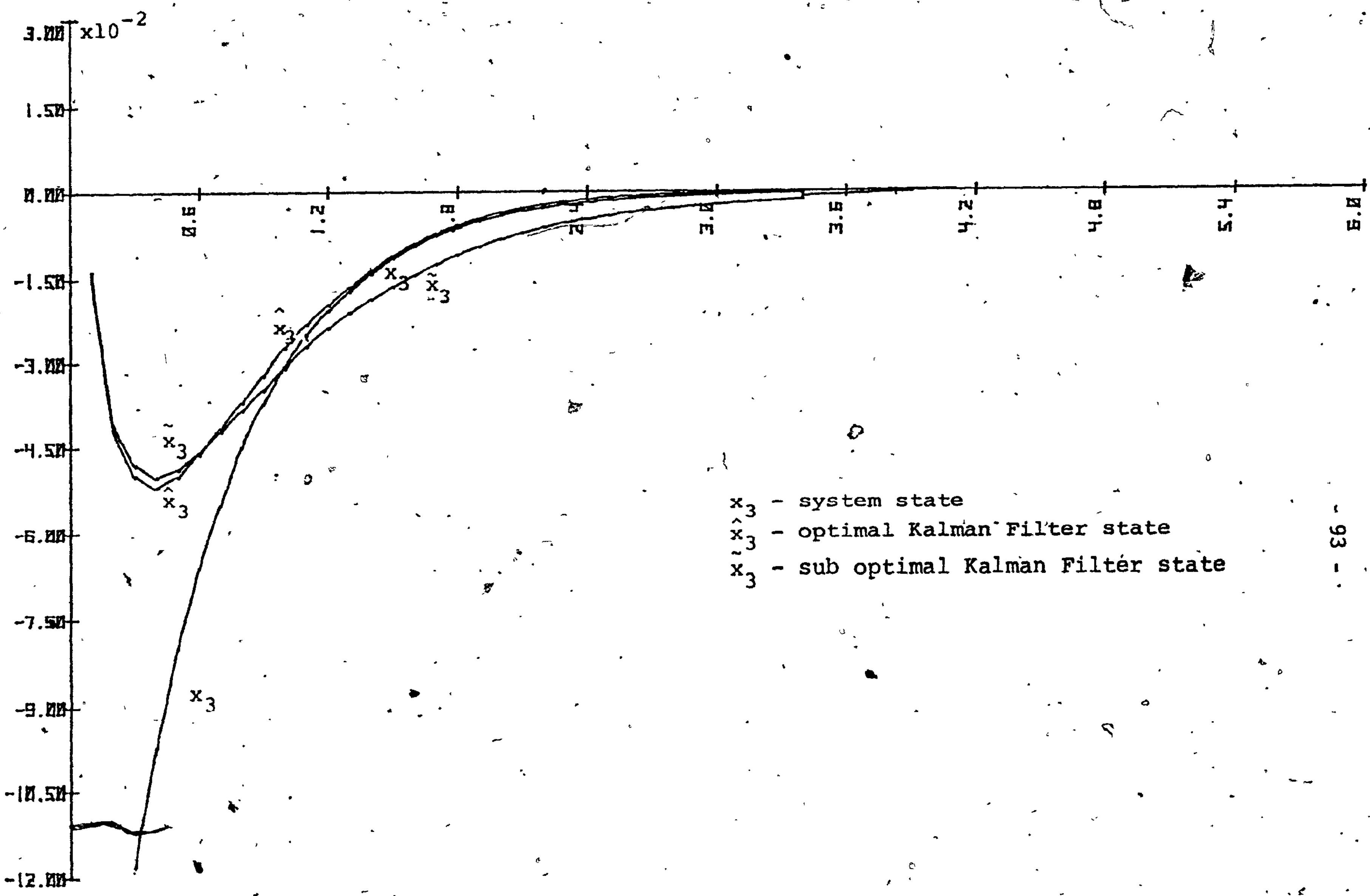

x10^-2
3.00
1.50
0.00
-1.50
-3.00
-4.50
-6.00
-7.50
-9.00
-10.50
-12.00
x_3 - system state
$\hat{x}_3$ - optimal Kalman Filter state
$\tilde{x}_3$ - sub optimal Kalman Filter state

FIG. 4.3

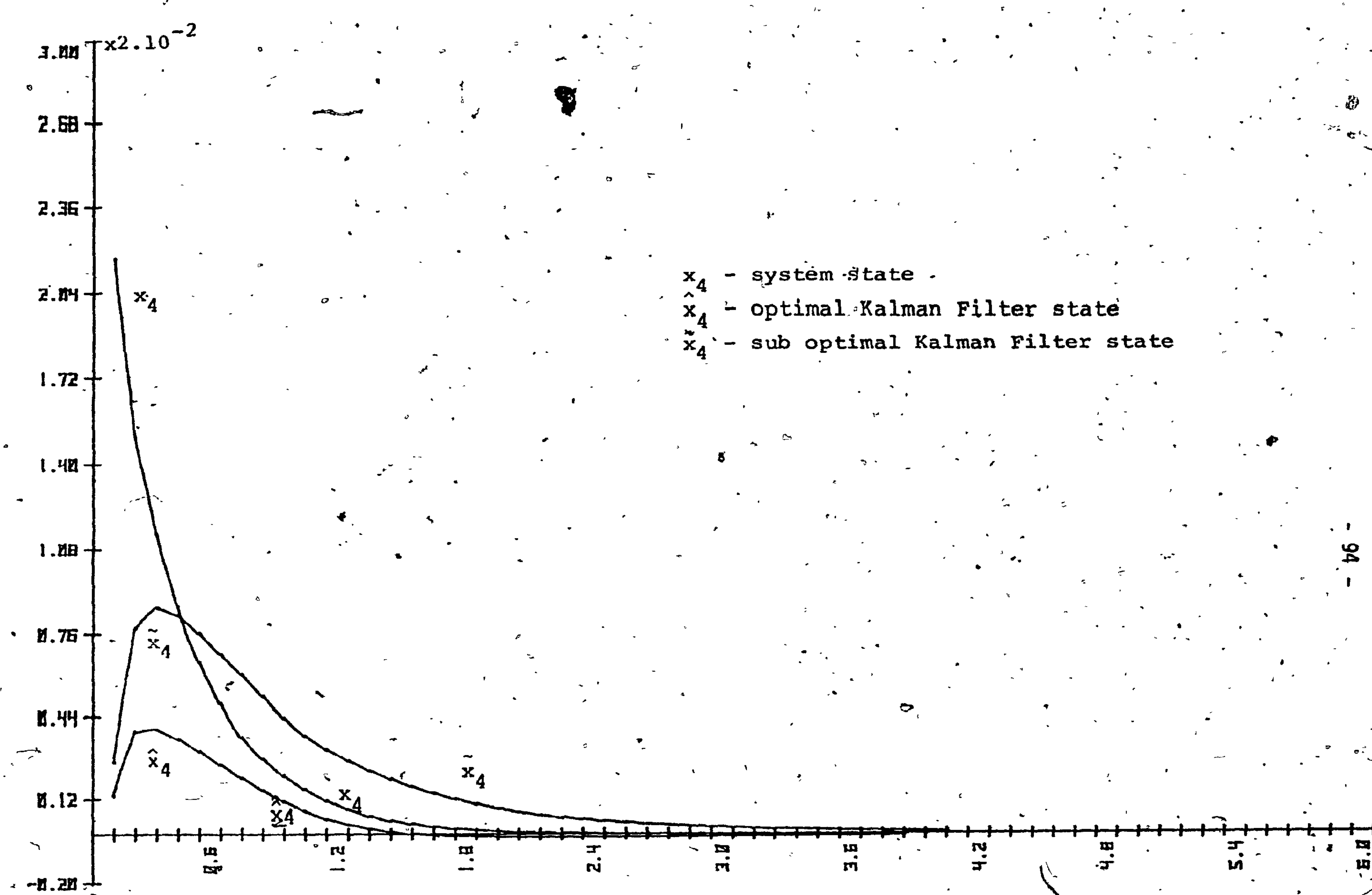

x2.10^-2
3.00
2.68
2.36
2.04
1.72
1.40
1.08
0.76
0.44
0.12
-0.20
0.6
1.2
1.8
2.4
3.0
3.6
4.2
4.8
5.4
6.0
x4 - system state
x̂4 - optimal Kalman Filter state
x̃4 - sub optimal Kalman Filter state
x4
x̃4
x̂4

FIG. 4.4

Step III (4.77), (4.78) were solved for $\hat{\underset{\sim}{x}}(t)$ and $\tilde{\underset{\sim}{x}}(t)$.

Similar procedure was used by Wilson [51] for a S.I.S.O.L. Problems in other context. In Figure 4.1, 4.2, 4.3, 4.4 are show $x_i(t)$, $\tilde{x}_i(t)$, $\hat{x}_i(t)$, for i=1 to 4 respectively. From the value of J_{D_S} we can see that the performance of the sub-optimal estimator is comparable to the optimal estimator and Figs. 4.1 to 4.4 shows that the approximate is satisfactory. Moreover for this example and several others where this method was tested, yield always an stable estimator $\forall$ r.

4.6 REDUCED ORDER MODEL OF A CLASS OF TIME VARYING SYSTEMS

Consider the following multi input-output linear time varying system of the form

$$\dot{\underset{\sim}{x}}(t) = A(t)\,\underset{\sim}{x}(t) + B(t)\underset{\sim}{U}(t) \tag{4.80}$$

$$\underset{\sim}{y}(t) = C(t)\underset{\sim}{x}(t) \tag{4.81}$$

where the state vector is of dimension , A(t), B(t), C(t) are matrices of dimensions nxn, nxm, Lxn respectively. Recently Wu[52] showed that l.t.v. autonomous sytems which are algebraically invariable or $\tilde{\tau}$ algebraically invariable can be explicity transformed into time invariant systems without using the full information contained in the transition matrix of the l.T.v. systems. Based on Wu's theory Rao [37] extended the results for nonautonomous time varying systems (4.80 - 4.81) and also presented a method for obtaining a reduced model for single -input-output systems by using a Routh approximation method due to Rao [36] on the transformed time varying system. With the use of the results from Wu and Rao and the disaggregation matrix S presented in section 4.30. We shall present a procedure for obtaining a reduced order model for a class of time varying systems of the multi input-output type. First lets consider the Wu transformation of the form

$$\underset{\sim}{x}(t) = e^{A_1 t}\, \underset{\sim}{z}(t), \quad A_1 \in e^{n x n} \tag{4.82}$$

Applying (4.82) to Eqs. (4.80-81) we have

$$\dot{\underset{\sim}{z}}(t) = \hat{A}\underset{\sim}{z}(t) + \hat{B}(t)\, \underset{\sim}{U}(t) \tag{4.83}$$

$$h(t) = \hat{C}(t)\underset{\sim}{z}(t) \tag{4.84}$$

Wehre $\hat{A} \in R^{n x m}$ and the following eq. are satisfied

$$A_1 A(t) - A(t)\, A_1 = A(t) \quad \forall t \tag{4.85}$$

$$\hat{A} = A(0) = A_1 \tag{4.86}$$

$$\hat{B}(t) = e^{A_1 t} B(t) \tag{4.87}$$

$$\hat{C}(t) = C(t)\, e^{A_1 t} \tag{4.88}$$

Now consider the Rao transformation as follows

$$\underset{\sim}{z}(t) = \underset{\sim}{x}(t) + \underset{\sim}{\psi}(t) \tag{4.89}$$

Then Eqn. (4.83) becomes

$$\dot{\underset{\sim}{x}}(t) = \hat{A}\underset{\sim}{x}(t) + \hat{A}\underset{\sim}{\psi}(t) + \hat{B}(t)\underset{\sim}{U}(t) - \underset{\sim}{\psi}(t) \tag{4.90}$$

Defining

$$\dot{\psi}(t) = \hat{A}\psi(t) + [\hat{B}(t) - B]\, U(t), \tag{4.91}$$

where B is any matrix such that the pair (A,B) is controlable, $B \in R^{n\times m}$ then from (3.77), (3.82-3.84) (4.83-84) (4.90-91) we have the L.T.I. system

$$\dot{\hat{x}}(t) = \hat{A}x(t) + BU(t), \tag{4.92}$$

$$\hat{h}(t) = \hat{C}(t)\, [\hat{X}(t) + \psi(t)]\,. \tag{4.93}$$

Consider now the following system

$$\dot{x}_1(t) = A_r X_r(t) + B_r U(t) \tag{4.94}$$

$$y_r(t) = C_r(t)\, [X_r(t) + \psi_r(t)] \tag{4.95}$$

Suppose the relationship between (4.92) and (4.94) is given by

$$S x_r = \hat{x} \tag{4.96}$$

Where S is the disaggregation matrix rxn. From (4.96), (4.92), (4.94) we have

$$SA_r = AS \tag{4.97}$$

$$SB_r = B \tag{4.98}$$

and the output equations (4.93) (4.95) are

$$\hat{\underset{\sim}{h}}(t) = \hat{C}(t)S\underset{\sim}{X}_r + \hat{C}(t)\,\underset{\sim}{\psi}(t), \tag{4.99}$$

$$\underset{\sim}{y}_r(t) = \hat{C}_r(t)\underset{\sim}{X}_r + \hat{C}_r(t)\underset{\sim}{\psi}_r(t). \tag{4.100}$$

Then

$$\hat{C}_r(t) = \hat{C}(t)S, \tag{4.101}$$

$$\hat{C}(t)\psi(t) = \hat{C}_r(t)\,\psi_r(t), \tag{4.102}$$

where

$\underset{\sim}{\psi}_r(t)$ is defined as in (4.91), we then have

$$\dot{\underset{\sim}{\psi}}_r(t) = A_r\underset{\sim}{\psi}_r(t) + (\hat{B}_r(t) - B_r)\underset{\sim}{U}(t)m \tag{4.103}$$

premultiplying (4.103) by S and comparing with (4.91) we arrive at

$$\hat{B}_r(t) = (S'S)^{-1}S'\hat{B}(t) \tag{4.104}$$

Clearly $\underset{\sim}{\psi}_r(t)$ can be obtained by solving (4.103) where $\hat{B}_r(t)$ is given by (4.104).

REMARKS

Remark 4.2 Consider the systems given by (4.92),(4.94) and $H(t)$, $H_r(t)$ are the respective impulse matrix. Suppose that the Reduced system (4.94) was obtained by use of the algorithm presented in Chapter 3, then $H_r(t)$ is a good approximation to $H(t)$ and therefore (4.94) is a good approximation to (4.92) for all inputs $\underset{\sim}{U}(t)$. However, this is not true for $\underset{\sim}{\psi}_r(t)$ and $\underset{\sim}{\psi}(t)$ where (4.103) has to be solved for every input $\underset{\sim}{U}(t)$. But the computation becomes easy due to the fact that matrices A_r, B_r, $\hat{B}_r$ were obtained before the output equations were computed and therefore need not be recalculated.

The following steps summarize the proposed method.

Step I. Using Wu and Rao transformations, transforms the L.T.V. systems (3.74-5) (4.90) (4.81) in to an L.T.I. system (4.92), (4.93).

Step II Find a reduced order model namely $\dot{\underset{\sim}{x}}_r(t) = A_r\underset{\sim}{X}(t) + B_r\underset{\sim}{U}(t)$ that approximates the system. $\dot{\hat{\underset{\sim}{x}}}(t) = A\hat{\underset{\sim}{x}}(t) + B\underset{\sim}{U}(t)$.

Step III Compute the disaggregation matrix S by solving equations (3.90-3.91) (4.97-4.98) using [10]

Step IV With S of Step III and (4.88) find $\underset{\sim}{\psi}C(t)$.

Step V Solve (4.103) for $\underset{\sim}{\psi}_r(t)$ where $\hat{B}_r(t)$ is given by $\hat{B}_r(t) = (S'S)^{-1}S'\hat{B}(t)$, for every input $\underset{\sim}{U}(t)$.

4.7 CONCLUSIONS

Several methods for obtaining R.O.M.'s from the error minimization approach exist but from applications and specific numerical examples are available in the literature. In this chapter therefore several applications and computational procedures are proposed for multi input-output systems that include, sub-optimal control policies for the output linear regulator problem, sub-optimal Wiener-Kalman filter for the stationary case. Furthermore, by using Wu an Rao Transformations a method for order reduction of a class of linear time varying systems is proposed based on the R.O.M.'s techniques for L.T.I. systems of Chapter 3.

CHAPTER 5

CONCLUDING REMARKS

In this thesis we were concerned mainly with two related topics, namely, the ℓ_1-norm minimization and the system order reduction problems. We established the following:

1. A new procedure for the unconstrained ℓ_1-norm minimization problem, which enables one to solve it efficiently using gradient techniques.

2. A new procedure for optimal order reduction for M.I.M.O. systems (proper or strictly proper H(s)) which ensures meaningful stable reduced-order models for stable higher order systems.

3. A new procedure for obtaining sub-optimal Kalman filters whose performance is comparable to the optimal K.W.F., by using reduced order models.

4. A new procedure for obtaining reduced order models of linear time varying systems using the procedure proposed for L.T.I. systems.

For the constrained ℓ_1-norm minimization problem, irregardless of the fact that the number of iterations is reduced by fifty percent over that of the previously available algorithm, further study must be done in order to render this technique suitable, i.e. make the number of function evaluations reasonable, for machine implementation. For the order reduction of multivariable systems, further effort must be made in order to clarify the structural relationship between the system and its reduced models, for example, increasing the number of parameters in the matrix A_r for the continuous case.

It is important also to remark that the techniques presented in the previous chapter can be applied to the following problem:

Let S_r be a system L.T.I. associated with a triple (A_r, B_r, C_r) where the pairs (A_r, B_r), (A_r, C_r) are weakly controllable or uncontrollable and weakly observable or unobservable, respectively. This system can be controllable and observable by increasing the dimension of the system. Namely, find a triple (A_n, B_n, C_n) associated to the system S_n where $n > r$, such as, S_n is a good approximation of S_r. Then, clearly, the techniques proposed to find reduced order models for L.T.I. in Chapter 2 can be used in a similar manner to find a system of higher dimension that is a good approximation of the lower order system.

Finally, an area in which investigation can be initiated is the combination of the ℓ_1-norm minimization with the disaggregation scheme in order to tackle the two-boundary value problem in optimal control, i.e., the sub-optimal time and sub-optimal fuel control policies, etc.

REFERENCES

[1] Aoki, M., "Control of Large Scale Dynamic Systems by Aggregation", IEE Trans. Automat. Contr., Vol. AC-13, No. 3, June 1968.

[2] Aoki, M., "Some Approximations Methods for Estimation and Control of Large Scale Systems", IEEE Trans. Automat. Contr., Vol. AC-23, No. 2, April 1978.

[3] Aoki, M. and M.T. Li, "Partial Reconstruction of State in Decentralized Dynamic Systems", IEEE Trans. Automat. Contr., Vol. AC-18, pp. 289-292, 1973.

[4] Aoki, M. and J.R. Huddle, "Estimation of State Vector of a Linear Stochastic System with a Constrained Estimation", IEEE Trans. Automat., Contr., Vol. AC-12, pp. 432-433, August 1967.

[5] Achieser, N.I. Theory of Approximation, Unger, [1,p.67], 1956.

[6] Arbel, A. and E. Tse, "Reduced Order Models, Canonical Forms and Observers", Int. J. Control, Vol. 30, No. 3, pp. 513-531, 1979.

[7] Avriel, M., Nonlinear Programming Analysis and Methods, Prentice-Hall, 1976.

[8] Barrodale, I. and Young A., "Algorithms for Best L_1 and L_∞ Linears Approximation on Discrete Set", Numer. Math. 8, pp. 295-306, 1966.

[9] Barrodale, I, and F.D.K. Roberts and C.R. Hunt, "Computing Best L_p Approximation by Functions Nonlinear in One Parameter", Comp. J., Vol. 13, pp. 382-386, 1970.

[10] Barrodale, I. and F.D.K. Roberts, "An Improved Algorithm for Discrete L_1 Linear Approximation", SIAM J. Numer. Anal., Vol. 10, pp. 838-843, 1973.

[11] Beale E.M.L., "Numerical Methods in Nonlinear Programming", J. Abadie (Ed.), North-Holland Publishing Co. Amsterdam, 1967.

[12] Calfe, H.R., "Continued-Fraction Model Reduction Technique for Multivariable Systems", Proc. IEE, Vol. 121, No. 5, May 1974.

[13] Chen, C.F., and L.S. Shieh, "A Novel Approach to Linear Model Amplification", Int. J. Contr., Vol. 8, pp. 561-570, 1968.

[14] Chen, C.F., "Model Reduction of Multivariable Control Systems by Means of Matrix Continued Fraction", Int. J. Control, Vol. 20, pp. 225-236, 1974.

[15] Dautzig, G.B., Linear Programming and Extensions, Princeton University Press, Princeton, N.J. 1963.

[16] Davison, E.J., A Method for Simplifying Linear Dynamic Systems, IEEE Trans. Automat. Contr., AC-11, pp. 93-101, 1966.

[17] Davison, E.J. and M.R. Chidambara "On a Method for Simplifying Linear Dynamic Systems", IEEE Trans. Automat. Contr., Vol. AC-12, pp. 119-121, Feb. 1967.

[18] Davison, E.J., "A New Method for Simplfying Large Linear Dynamic Systems", IEEE Trans. Automat. Contr., Vol. AC-13, pp. 214-215, 1968.

[19] Desoer, C.A. and Vidyasagar, M., Feedback Systems: Input-output Properties, Acadmeic Press, 1975.

[20] El-Attar, R.A., M. Vidyasagar, R.R. Dutta, An Algorithm for L_1-norm Minimization with Application to Nonlinear L_1-approximation, SIAM, J. Numer. Anal. Vol. 16, No. 1, February 1979.

[21] El-Attar, R.A. "New Algorithms for Nonlinear L_1-norm Minimization With Applications to Control System Design", Phd. Thesis, Concordia University, 1978.

[22] El-Attar, R.A., M. Vidyasagar, S.R.K. Dutta,"Optimality Conditions for L_1-norm Minimization", 19th Midwest Symposium on Circuits and Systems, pp. 272-275, 1976.

[23] Fiacco, A.V., G.P. McCormick, Nonlinear Programming: Sequential Unconstrained Minimization Techniques, John Wiley and Sons, New York, 1968.

[24] Fletcher, R., "A New Approach to Variable Metric Algorithms", Comput. J., Vol. 13, pp. 317-322, Aug. 1970.

[25] Galiania, F.D., "On the Approximation of Multiple Input-Multiple Output Constant Linear Systems, Int. J. Control, Vol. 7, No. 6, pp. 1313-1324, 1973.

[26] Gilbert, E.G., "Controlability and Observability in Multivariable Control Systems", SIAM J. of Control Series A, Vol. 1, No. 2, pp. 128-151, 1963.

[27] Hirzuiger, G. and Kreisselmeier, G., "On Optimal Approximation of High Order Linear System by Low Order Models", Int. J. Contr., Vol. 22, No. 3, pp. 399-408, 1975.

[28] Hutton, M.F., B. Friedlard. Routh Approximation for Reducing Order of Linear Time-Invariant Systems, IEE Trans. Automat. Contr., Vol. AC-20, No. 3, June 1975.

[29] Kalman. R.E., "Mathematical Description a Linear Dynamic Systems", Journ. Soc. Ind. Appl. Math. Contr. Series, pp. 51-60, March 1964.

[30] Kalman, R.E. and R.S. Bucy, New Results in Linear Filtering and Prediction Theory Journal of Basic Engineering, ASME, pp. 95-108, March 1961.

[31] Kuhn, H.W., "Nonlinear Programming: A Historical View", SIAM-AMS Proceedings, Vol. IX, Nonlinear Programing 1976.

[32] Kuhn, H.W., A.W. Tucker, "Nonlinear Programming", in Jerzy Neyman (ed), Proceedings of the second Berkeley Symposium on Mathematical Statistics and Probability (Berkeley University of Califor. Press), pp. 481-492, 1950.

[33] Meditch, J.S."Suboptimal Linear Filtering for Continuous Dynamic Process", Aerospace Corp. Tech. Dept., July 1964.

[34] Newmann, M.M., "A Continuous-Time Reduced-Order Filter for S the State Vector of a Linear Stochastic System", Int. J. Control, Vol. 11, No. 2, pp. 229-239, 1970.

[35] Osborne, M.R. and G.A. Watsons, "On an Algorithm for Discrete Nonlinear L_1 Approximations", Comp. J. Vol. 14, pp. 184-188, 1971.

[36] Rao, A.S., Lamba, S.S., Rao, S.V. Rao., Routh Approximate time Domain Reduced Order Models for S.I.S.O. Systems" Proc., IEE, pp. 1059-1063, 1978.

[37] Rao, A.S., Lamba, S.S. and Vittal Rao S., "Application of Routh Approximant Method for Reducing the Order of a Class of Time Varying Systems", 22nd Midwest Symposium Circuits on Systems, 1979.

[38] Rabinowitz, R., "Mathematical Programming and Approximation", in A. Talbot (ed.), Approximation Theory Academic Press, London, pp. 217-231, 1970.

[39] Rice, J.R., "On Nonlinear L_1 Approximation", Arch. Rat. Mech. Anal., 17, pp. 61-66, 1964.

[40] Rice, J.R., "The Approximation of Functions", Vol. 1, Addison Wesley, 1964.

[41] Robers, P.D., Ben-Israel, A., "An Interval Programming Algorithm for Discrete Linear L_1 Approximations Problems", System Research Memorandum, No. 223, Northwestern University, 1969.

[42] Rosenbrock, H.H., State-Space and Multivariate Theory, Nelson, 1970.

[43] Sage, A.P., C.C. White III, Optimum Systems Control, Englewood Cliffs, New Jersey, 1977.

[44] Shamrash, Y., "Model Reduction Using Minimal Realization Algorithm", Electronics Letters, Vol. 11, No. 16, pp. 385-387, 1975.

[45] Silverman, L.M., "Realization of Linear Dynamical Systems", IEEE Trans., AC-16, pp. 554-567, 1971.

[46] Singh, M.G. and Titli, Systems: Decomposition Optimization and Control, Bergarnon Press, 1978.

[47] Usow, K.H., "On L_1 Approximation, II: Computation for Discrete Functions and Discretizations Effects", SIAM J. Numer. Anal. 4, pp. 233-244, 1967.

[48] Vidyasagar, M., Nonlinear System Analysis, Prentice-Hall, Englewood Cliffs, N.J., 1978.

[49] Wilson, D.A. and Mishra, R.N., "Optimal Reduction of Multivariable Systems", Int. J. Control, Vol. 29, No. 2, pp. 267-278, 1977.

[50] Wilson, D.A., "Model Reduction for Multivariable Systems", Int. J. Contr. Vol. 20, No. 1, pp. 57-64, 1974.

[51] Wilson, D.A. and Mishra, "Design of Low Order Estimators Using Reduced Models", Int. J. Control, Vol. 29, No. 3, pp. 447-456, 1979.

[52] Wu, M.Y., "Transformation of a Linear Time Varying System into a Linear Time Invariant System", Int. J. Control, Vol. V7, No. 4, pp. 589-602,

[53] Zadeh, L.A., and C.A. Desoer, "Linear System Theory: the State Space Approach", McGraw-Hill Book Company, Inc., 1963.