

1972

Near-optimal state estimation for interconnected systems

Terry Bernard Cline
Iowa State University

Follow this and additional works at: <https://lib.dr.iastate.edu/rtd>

 Part of the [Electrical and Electronics Commons](#)

Recommended Citation

Cline, Terry Bernard, "Near-optimal state estimation for interconnected systems " (1972). *Retrospective Theses and Dissertations*. 4725.
<https://lib.dr.iastate.edu/rtd/4725>

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.

INFORMATION TO USERS

This dissertation was produced from a microfilm copy of the original document. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the original submitted.

The following explanation of techniques is provided to help you understand markings or patterns which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting thru an image and duplicating adjacent pages to insure you complete continuity.
2. When an image on the film is obliterated with a large round black mark, it is an indication that the photographer suspected that the copy may have moved during exposure and thus cause a blurred image. You will find a good image of the page in the adjacent frame.
3. When a map, drawing or chart, etc., was part of the material being photographed the photographer followed a definite method in "sectioning" the material. It is customary to begin photoing at the upper left hand corner of a large sheet and to continue photoing from left to right in equal sections with a small overlap. If necessary, sectioning is continued again — beginning below the first row and continuing on until complete.
4. The majority of users indicate that the textual content is of greatest value, however, a somewhat higher quality reproduction could be made from "photographs" if essential to the understanding of the dissertation. Silver prints of "photographs" may be ordered at additional charge by writing the Order Department, giving the catalog number, title, author and specific pages you wish reproduced.

University Microfilms

300 North Zeeb Road
Ann Arbor, Michigan 48106
A Xerox Education Company

73-9432

CLINE, Terry Bernard, 1945-
NEAR-OPTIMAL STATE ESTIMATION FOR INTERCONNECTED
SYSTEMS.

Iowa State University, Ph.D., 1972
Engineering, electrical

University Microfilms, A XEROX Company, Ann Arbor, Michigan

**Near-optimal state estimation
for interconnected systems**

by

Terry Bernard Cline

**A Dissertation Submitted to the
Graduate Faculty in Partial Fulfillment of
The Requirements for the Degree of
DOCTOR OF PHILOSOPHY**

Major: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

**Iowa State University
Ames, Iowa**

1972

PLEASE NOTE:

Some pages may have
indistinct print.

Filmed as received.

University Microfilms, A Xerox Education Company

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
II. DESCRIPTION OF AN INTERCONNECTED SYSTEM	7
III. MODAL-TRAJECTORY STATE ESTIMATION FOR NONLINEAR INTERCONNECTED SYSTEMS	17
A. Derivation of the Algorithms	18
B. Performance Analysis	36
C. Example	37
IV. STATE ESTIMATION FOR LINEAR INTERCONNECTED SYSTEMS	51
A. Two Subsystem Case	52
B. s-Subsystem Case	65
C. Performance Analysis	67
D. Examples	71
V. CONCLUSION	98
VI. REFERENCES	102
VII. ACKNOWLEDGMENTS	105
VIII. APPENDIX	106

I. INTRODUCTION

In applied science and technology an important problem is that of estimating the behavior of a physical process subject to random disturbances and measurement errors in some manner which is "best" in the context of some performance criterion. Frequently the description of the process is in terms of a nonlinear stochastic differential equation and a nonlinear measurement process.

The state estimation problem for such a system is a formidable one both theoretically and practically. Two principal criteria have been used to judge the performance of a particular procedure: conditional mean and conditional maximum-likelihood estimation. In conditional mean estimation the object is to design an estimator that produces an estimate which is the mean of the state conditioned on the observations. Such an estimator minimizes the variance of the state conditioned on the observations. On the other hand, the goal of conditional maximum-likelihood estimation is to maximize the probability density function of the state conditioned on the observations.

Kushner [1] has determined the equations which are satisfied by the conditional mean and conditional covariance matrix. However, the conditional density function, which

is usually unknown, is required to implement the equations for a nonlinear system.

In the maximum-likelihood approach two philosophies have been adopted. One is to compute the mode of the conditional density function. Kushner [2] has derived the equation for the mode, but as in the conditional mean approach, the unknown conditional density function is needed to implement it. The second approach, taken by Cox [3] and Mortensen [4], is to compute the most probable trajectory to be followed by the system. This is called the modal-trajectory estimator. The advantage of this approach is that optimal control techniques can be used to characterize the optimal estimator. However, problems similar to the ones above arise when it comes to implementing the estimator.

As one may have concluded by now, the best we can hope for in an application is to be able to find an adequate approximation for the optimal estimator. Two trends have appeared in the efforts to find good approximations.

Since the linear estimation problem has been solved by Kalman [5] and Kalman and Bucy [6], it is natural to attempt to approximate the nonlinear problem in such a way that the solution to the linear problem can be utilized. A brief statement of the linear problem and a summary of the Kalman-Bucy equations are provided for reference in the

Appendix. The algorithms which have resulted are the linearized, extended, and iterated extended Kalman filters, among others. In the linearized Kalman filter the system is linearized about a nominal trajectory and the Kalman filter is applied to this linearized model. In the extended Kalman filter the same linearization is used, but the linearization is performed about the most recent estimate obtained from the filter. The iterated extended Kalman filter is the same as the extended Kalman filter with the exception that an iterative procedure is employed at each step to improve the nominal trajectory by reducing the effects of measurement nonlinearities.

The other approach is to approximate the nonlinear problem directly by assuming a density function or by attempting to parameterize the density function. Some of the results have been the second-order gaussian estimator and the truncated second-order estimator. The truncated second-order filter is obtained by assuming that third and higher order moments of the conditional distribution may be neglected. The second-order gaussian filter is obtained by assuming the conditional distribution to be gaussian and account for moments up to fourth-order by computing them in terms of the second-order moments. System nonlinearities are approximated to second-order.

Surveys of the above mentioned algorithms and others

may be found in Jazwinski [7] and Sage and Melsa [8].

One of the problems with these approximate methods is that it is often difficult to justify the assumptions required by a particular algorithm. Also it is very difficult to ascertain the quality of performance which will result from the use of the algorithm.

Another problem, which also plagues the Kalman filter, is that in many applications the dimension of the system may be large. Because of this, the computational and high-speed memory requirements of the algorithm may be excessive, particularly if real-time implementation via a small on-board computer is desired.

In attempting to deal with these problems, we take a new approach to the approximation of the modal-trajectory estimator. By adopting the modal-trajectory approach we are able to utilize approximation methods from such areas as large-scale control (see Kokotović and Singh [9]) and optimally sensitive control systems (see Werner and Cruz [10]) to synthesize a new approximate modal-trajectory estimator.

Our approach to the problem is formulated in the context of large interconnected systems. We will regard the system as a collection of interconnected subsystems. Our description of this composite system in Chapter II is a generalization of that used by Bailey [11].

Using the modal-trajectory approach Mortensen [4] was able to obtain a nonlinear two-point boundary value problem (TPBVP) whose solution is the optimal modal-trajectory estimator. In Chapter III we derive an approximate algorithm for solving this TPBVP using a method used by Werner and Cruz [10] and others (see Kokotović and Singh [9], Kokotović and Sannuti [12], and Cruz [13]) to design optimally sensitive control systems. We will call this method the ϵ -coupling technique. The ϵ -coupling technique together with the interconnected system formulation allows us to obtain an estimator with the following properties:

- (1) computations are carried out at the lower-dimensional subsystem level,
- (2) the smoothing as well as the filtering solution is an integral component of the method,
- (3) a qualitative estimate of the performance of the algorithm can be obtained.

In Chapter IV we re-derive in our formulation the ϵ -coupling solution of Haddad and Cruz [14] to the linear estimation problem. We also specialize our results in Chapter III to the linear case to obtain another ϵ -coupling solution to the linear problem and state a theorem on the performance of the algorithms.

At appropriate points in the exposition examples are presented to illustrate the employment of the algorithms

and to demonstrate that the algorithms do work. A full computational study, however, is beyond the scope of this work.

II. DESCRIPTION OF AN INTERCONNECTED SYSTEM

In this chapter we describe a class of interconnected systems which retains enough structural properties to enable us to obtain meaningful computational and theoretical results while being reasonably general.

The class of systems which we describe is basically a generalization of that described by Bailey [11]. To Bailey's model we add plant and measurement noise and allow a nonlinear interconnection structure. In later chapters we shall actually deal with just two subclasses of the type of system described here.

The basic unit of our composite system; i.e., interconnected system, is the subsystem. The subsystem is described by specifying its external inputs, its internal structure, and its outputs. The composite system may be described by specifying its external inputs, its subsystems and their interconnections, and its outputs.

Suppose we have a composite system composed of s subsystems, S_i , $i = 1, 2, \dots, s$. Then each S_i may be described in the following manner. The state $x_i(t)$ of S_i is determined at any time $t \in (T_1, T_2)$ by

$$\dot{x}_i(t) = f_i(x_i(t), t) + u_i(t) + w_i(t) \quad (2.1)$$

where $f_i(\cdot, \cdot)$ maps $E^{n_i} \times E^1$ into E^{n_i} , $u_i(t) \in E^{n_i}$ is the input into S_i at t , and $w_i(t) \in E^{n_i}$ is a white noise disturbance process. E^{n_i} denotes the n_i -dimensional Euclidean vector space, where n_i is the dimension of S_i . Equation (2.1) is to be interpreted in the sense of Stratonovich (see e.g. Jazwinski [7]).

The output of S_i is given by

$$z_i(t) = g_i(x_i(t), t) \quad (2.2)$$

where $g_i(\cdot)$ maps $E^{n_i} \times E^1$ into E^{q_i} .

The relationships among the subsystems are given by

$$u_i(t) = \sum_{j=1}^S b_{ij}(z_j(t), t) + K_i(t)u(t) \quad (2.3)$$

where $b_{ij}(\cdot, \cdot)$ maps $E^{q_i} \times E^1$ into E^{n_i} , $K_i(\cdot)$ maps E^1 into the space of $n_i \times \ell$ matrices, and $u(t) \in E^\ell$ is the external input of the composite system.

The assumption is made that the internal structure of the individual subsystems is not affected by the interconnection with other subsystems. In other words, we make the usual "no loading" assumption. It may appear that the assumption of noise-free connections between subsystems has been made. However, the effects of noisy connections can be accounted for in the w_i terms because surely in practical systems the connections between subsystems are

not noise-free.

The plant description of the composite system can be obtained by combining equations (2.1) - (2.3). Denote by $x(t) \in E^{n_1} \times E^{n_2} \times \dots \times E^{n_s}$ the composite system state at time t ; i.e.,

$$x' = (x_1, x_2, \dots, x_s)'$$

where the prime denotes transpose. Then

$$\dot{x}(t) = f(x, t) + c(x, t) + K(t)u(t) + w(t) \quad (2.4)$$

where

$$f'(x, t) = (f_1(x_1, t), f_2(x_2, t), \dots, f_s(x_s, t))'$$

and

$$c'(x, t) = (c_1(x, t), c_2(x, t), \dots, c_s(x, t))',$$

such that

$$c_i(x, t) = \sum_{j=1}^s b_{ij}(z_j(x_j), t), \quad i = 1, 2, \dots, s;$$

and

$$K(t) = \begin{bmatrix} K_1(t) & & & & \\ & K_2(t) & & & \\ & & \cdot & & \\ & \emptyset & & \cdot & \\ & & & & K_s(t) \end{bmatrix},$$

and

$$w'(t) = (w_1(t), w_2(t), \dots, w_s(t))'.$$

We call $c(\cdot, \cdot)$ the interconnection term, $u(\cdot)$ the composite system (external) input, and $w(\cdot)$ the composite system plant disturbance. It is assumed that the function $f + c + Ku$ is continuous and continuously differentiable up to some order $N + 1$ in all its arguments in some domain D of the (x, u, t) space.

Suppose now that we can make measurements on the subsystems. Then for S_i we have $(y_i(t) \in E^{m_i})$

$$y_i(t) = h_i(x_i(t), t) + d_i(x(t), t) + v_i \quad (2.5)$$

and for the composite system measurement $(y(t) \in E^{m_1} \times E^{m_2} \times \dots \times E^{m_s})$

$$y(t) = h(x(t), t) + d(x(t), t) + v \quad (2.6)$$

where the definitions of y , h , d , and v without subscripts refer to the whole composite system.

The $v_i(\cdot)$ term in (2.5) is a white noise process which expresses the uncertainty in each subsystem measurement. The possibility of other subsystem states coupling into the measurement is the motivation for the $d(\cdot, \cdot)$ term.

Observe also that external self-feedback (possibly nonlinear) around a subsystem is allowed in this formula-

tion. Similarly, self (possibly nonlinear) coupling is allowed in the subsystem measurements. These two features will be important in our approach to the nonlinear estimation problem.

Thus far we have not completely specified the noise processes in the model. This can be done by specifying their mean and covariances (assuming gaussian distributions). Thus, let

$$\text{mean}\{w(t)\} = 0, \text{ for all } t \in (T_1, T_2)$$

and

$$\text{cov}\{w(t_1), w(t_2)\} = Q(t_1) \delta(t_1 - t_2), \text{ for all } t_1, t_2 \in (T_1, T_2)$$

where $\delta(\cdot)$ denotes the Dirac delta function. We will assume $w(\cdot)$ is independent of $v(\cdot)$ and that

$$\text{mean}\{v(t)\} = 0 \text{ for all } t \in (T_1, T_2)$$

and

$$\text{cov}\{v(t_1), v(t_2)\} = R(t_1) \delta(t_1 - t_2)$$

$$= \begin{bmatrix} R_1(t_1) & & & & \\ & R_2(t_1) & & & \\ & & \cdot & & \\ & \emptyset & & \cdot & \\ & & & & R_s(t_1) \end{bmatrix} \delta(t_1 - t_2)$$

for all $t_1, t_2 \in (T_1, T_2)$.

Note that Q is not necessarily block diagonal so that the effects of noisy connections can be included. R is block diagonal because it is assumed that the subsystem measurements are made independently.

Before proceeding to the estimation problem we give three examples of composite systems whose subsystems are interconnected in different ways.

EXAMPLE 2.1. A simple cascade system is obtained by letting K_2, \dots, K_s be zero and $u_k = z_{k-1}$, $k = 2, \dots, s$. The block diagram of Figure 2.1 indicates the structure of such a system. The fact that no measurements are made on S_1, \dots, S_{s-1} can be accounted for by letting h_i and d_i , $i = 1, \dots, s-1$, be identically zero, causing y_1, \dots, y_{s-1} to be meaningless.

EXAMPLE 2.2. The block diagram of Figure 2.2 indicates the structure of a simple parallel connection of subsystems in which the only measurement is the sum of the subsystem outputs. This measurement may be arbitrarily assigned to S_1 with y_2, \dots, y_s discarded as in Example 2.1. Of course, in many practical situations measurements can be performed at other points in the system and y determined accordingly. The form of y in these illustrative examples is not meant to give the impression that other possibilities are precluded.

EXAMPLE 2.3. A simple loop connection of subsystems is depicted in Figure 2.3. The only measurement available is the output of S_{k-1} . In this case, only K_1 is nonzero.

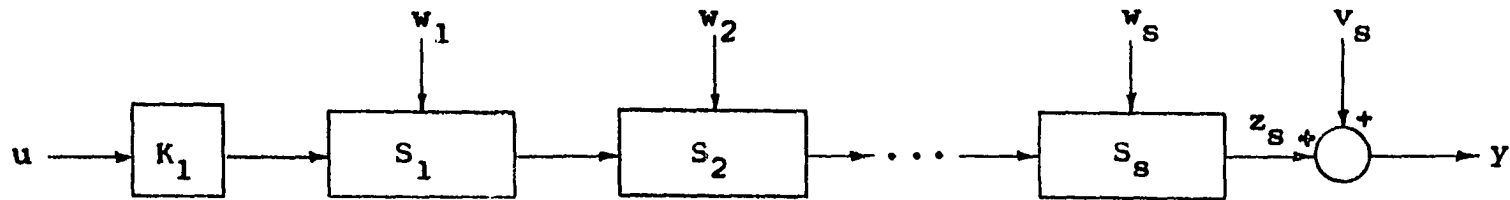


Figure 2.1. A simple cascade system

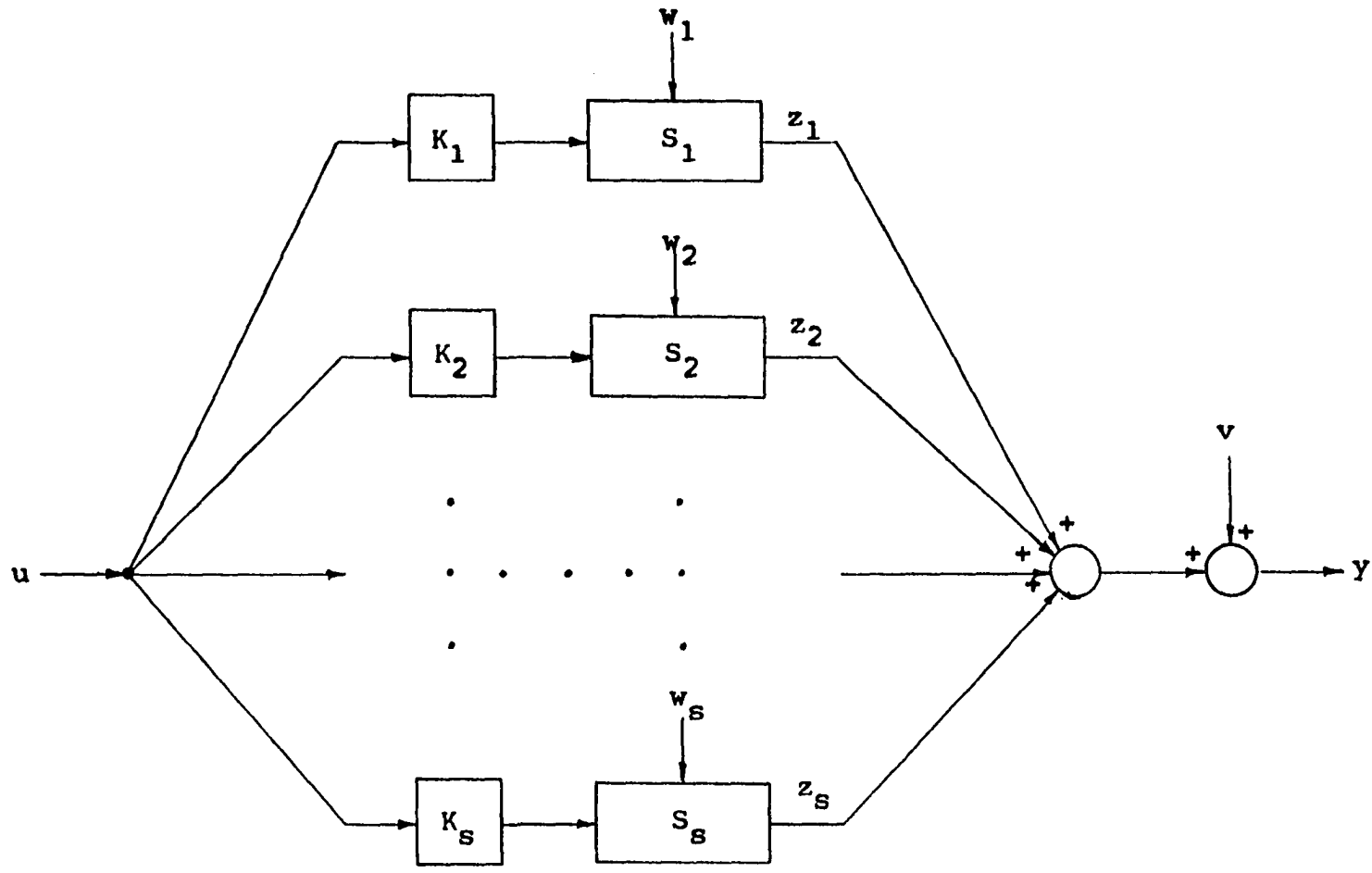


Figure 2.2. A simple parallel connected system

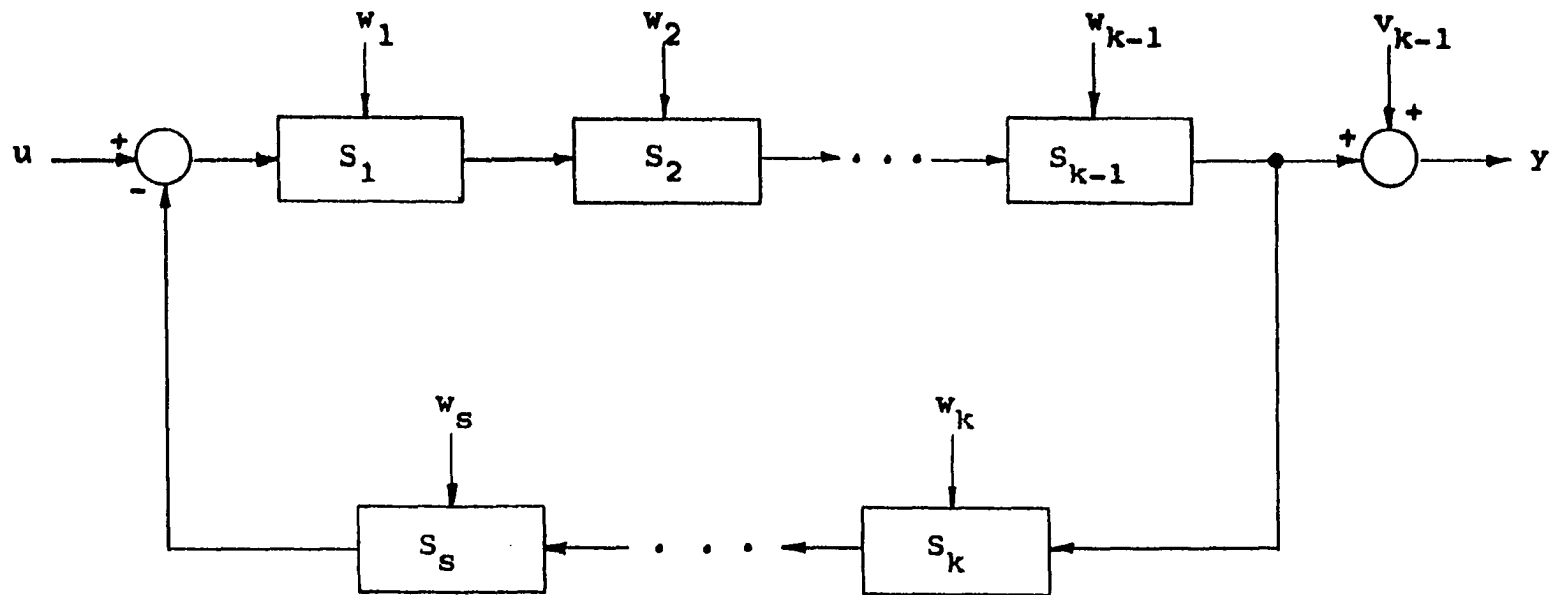


Figure 2.3. A simple closed-loop system

III. MODAL-TRAJECTORY STATE ESTIMATION FOR NONLINEAR INTERCONNECTED SYSTEMS

In the modal-trajectory approach the conditional density function is not required. Instead, the most probable trajectory in function space is computed to obtain the smoothed estimates, with the filtered estimate being the final state of this trajectory. The appeal of the method is that the estimator equations can be obtained via control theoretic techniques. Hence, one might expect that approximation techniques which have been used successfully on control problems might be utilized to obtain an approximate modal-trajectory estimator. As we shall see, this is precisely the case, and we shall also see that this fact will enable us to evaluate the quality of our approximate algorithm.

The ϵ -coupling method which we will use to obtain our approximate estimator originated in sensitivity analysis and the theory of optimally sensitive control systems; see e.g. Cruz [13]. The method may be outlined in the following way. Suppose we have a control system with a cost functional which contains a parameter. The procedure then is to expand the state (as well as the costate and the control) in a series about some nominal value of the parameter and to solve for the coefficients

(in function space) in the series. The necessary conditions for optimality can be used to evaluate the coefficients. The advantage of this method in our problem is that the resulting equations which must be solved are decoupled by subsystem.

In practice one uses only a truncated version of the series. The degree of approximation is determined by how many terms one is willing to compute. The natural question which arises at this point is why is this procedure any better than computing a truncated series about a nominal state trajectory or the latest estimate (linearized and extended Kalman filters). As we shall see, if the problem is formulated appropriately, considerable reductions in computation can be obtained so that more terms in the series can be carried. The procedure also gives one a systematic method for computing the terms in the series and a way of ascertaining the quality of the resulting algorithm.

A. Derivation of the Algorithms

In this section we derive the approximate modal-trajectory smoothing and recursive filtering equations for a sub-class of nonlinear composite systems of those described in Chapter II. We consider only systems which have linear subsystems and no external inputs. Our

$$\dot{x}(t) = F(t)x(t) + \epsilon c(x(t), t) + w(t) \quad (3.3)$$

$$y(t) = H(t)x(t) + \epsilon d(x(t), t) + v(t) \quad (3.4)$$

$$Q = \begin{bmatrix} Q_1 & \epsilon Q_{12} & \epsilon Q_{13} & \cdot & \cdot & \cdot & \epsilon Q_{1s} \\ \epsilon Q_{21} & Q_2 & \epsilon Q_{23} & \cdot & \cdot & \cdot & \\ \cdot & & \cdot & & & & \\ \cdot & & \cdot & & & & \\ \cdot & & \cdot & & & & \\ \epsilon Q_{s-1,1} & \cdot & \cdot & \cdot & Q_{s-1} & \epsilon Q_{s-1,s} & \\ \epsilon Q_{s,1} & \cdot & \cdot & \cdot & \epsilon Q_{s,s-1} & Q_s & \end{bmatrix}, \quad (3.5)$$

$$P(t_0) = \begin{bmatrix} P_1(t_0) & \epsilon P_{12}(t_0) & \cdot & \cdot & \cdot & \epsilon P_{1s}(t_0) \\ \epsilon P_{21}(t_0) & P_2(t_0) & \cdot & \cdot & \cdot & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & P_s(t_0) \end{bmatrix}. \quad (3.6)$$

ϵ is allowed to vary on the interval $[0,1]$. When $\epsilon = 0$, we have a set of decoupled linear subsystems and when $\epsilon = 1$, we have the original composite system.

Now, the likelihood function for this composite system is (see Mortensen [4])

$$\begin{aligned}
J_t &= 0.5[x(t_0) - \mu] \cdot P^{-1}(t_0)[x(t_0) - \mu] \\
&+ 0.5 \int_{t_0}^t \{ [\dot{x}(\tau) - F(\tau)x(\tau) \\
&\quad - \epsilon c(x(\tau), \tau)] \cdot Q^{-1}(\tau, \epsilon) [\dot{x}(\tau) \\
&\quad - F(\tau)x(\tau) - \epsilon c(x(\tau), \tau)] \} d\tau \\
&+ 0.5 \int_{t_0}^t \{ [y(\tau) - H(\tau)x(\tau) \\
&\quad - \epsilon d(x(\tau), \tau)] \cdot R^{-1}(\tau) [y(\tau) \\
&\quad - H(\tau)x(\tau) - \epsilon d(x(\tau), \tau)] \} d\tau \quad (3.7)
\end{aligned}$$

Equation (3.7) can be derived by use of the theory of Feynman integrals in function space, Mortensen [15].

For a given realization of $y(\tau)$, $t_0 \leq \tau \leq t$, J_t is a functional only on $x(\cdot)$ (and ϵ). The most probable trajectory is computed by minimizing J_t with the end conditions $x(t_0)$ and $x(t)$ free. The minimizing trajectory $x(\tau)$, $t_0 \leq \tau \leq t$, is the modal-trajectory smoothed estimate and the final state $x(t)$ is the modal-trajectory filtered estimate.

In order to perform sequential filtering it is necessary to continually update the optimal estimate as t varies. As discussed in Mortensen, this requires us to

perform the minimization for a new modal-trajectory for each new t ; i.e., perform an infinite sequence of minimizations. Clearly, this is an undesirable property of the modal-trajectory estimator. However, we shall see that application of the ϵ -coupling technique to the problem eliminates this prohibitive computational burden.

To put our problem into more convenient form, substitute (3.3) into (3.7) and treat $w(\tau)$, $t_0 \leq \tau \leq t$, as a control. The problem now is to choose the control which minimizes

$$\begin{aligned}
 J_t = & 0.5[x(t_0) - \mu]^T P^{-1}(t_0)[x(t_0) - \mu] \\
 & + 0.5 \int_{t_0}^t \{w^T(\tau) Q^{-1}(\tau, \epsilon) w(\tau) + [y(\tau) - H(\tau)x(\tau) \\
 & \quad - \epsilon d(x(\tau), \tau)]^T R^{-1}(\tau)[y(\tau) - H(\tau)x(\tau) \\
 & \quad - \epsilon d(x(\tau), \tau)]\} d\tau \tag{3.8}
 \end{aligned}$$

subject to

$$\dot{x}(\tau) = F(\tau)x(\tau) + \epsilon c(x(\tau), \tau) + w(\tau), \quad t_0 \leq \tau \leq t \tag{3.9}$$

with $x(t)$ and $x(t_0)$ free. In this form, we have the tracking problem of optimal control theory.

Denoting the costate by λ , we define the Hamiltonian as

$$\mathcal{H}(x, \lambda, w, \epsilon, \tau) = 0.5 w^T(\tau) Q^{-1}(\tau, \epsilon) w(\tau) + \lambda^T(\tau)[F(\tau)x(\tau)$$

$$\begin{aligned}
& + \epsilon c(x(\tau), \tau) + w(\tau)] + 0.5 [y(\tau) - H(\tau)x(\tau) \\
& - \epsilon d(x(\tau), \tau)] \cdot R^{-1}(\tau) [y(\tau) - H(\tau)x(\tau) \\
& - \epsilon d(x(\tau), \tau)]. \tag{3.10}
\end{aligned}$$

Considering $y(\cdot)$ to be a known function of τ , we apply the minimum principle (Pontryagin, et al. [16]). The resulting necessary conditions are

$$\dot{x}(\tau) = \mathcal{H}_{\lambda} = F(\tau)x(\tau) + \epsilon c(x(\tau), \tau) + w(\tau) \tag{3.11}$$

$$\begin{aligned}
\dot{\lambda}(\tau) = -\mathcal{H}_{\lambda} = & -[F'(\tau) + \epsilon c'_x(x(\tau), \tau)]\lambda(\tau) \\
& + [H'(\tau) + \epsilon d'_x(x(\tau), \tau)]R^{-1}(\tau)[y(\tau) \\
& - H(\tau)x(\tau) - \epsilon d(x(\tau), \tau)] \tag{3.12}
\end{aligned}$$

$$\mathcal{H}_w = Q^{-1}(\tau, \epsilon)w(\tau) + \lambda(\tau) = 0 \tag{3.13}$$

$$\lambda(t_0) = P^{-1}(t_0)[x(t_0) - \mu],$$

$$\lambda(t) = 0 \tag{3.14}$$

where the subscripts indicate partial differentiation with respect to the indicated argument.

Solving (3.13) for $w(\tau)$,

$$w(\tau) = -Q(\tau, \epsilon)\lambda(\tau). \tag{3.15}$$

Substituting (3.15) into (3.11), we obtain the following

two-point boundary value problem (TPBVP)

$$\dot{x}(\tau) = F(\tau)x(\tau) + \epsilon c(x(\tau), \tau) - Q(\tau, \epsilon)\lambda(\tau), \quad (3.16)$$

$$\begin{aligned} \dot{\lambda}(\tau) = & -[F(\tau) + \epsilon c_x(x(\tau), \tau)] \cdot \lambda(\tau) \\ & + [H(\tau) + \epsilon d_x(x(\tau), \tau)] \cdot R^{-1}(\tau)[y(\tau) \\ & - H(\tau)x(\tau) - \epsilon d(x(\tau), \tau)], \end{aligned} \quad (3.17)$$

$$x(t_0) = \mu + P(t_0)\lambda(t_0),$$

$$\lambda(t) = 0. \quad (3.18)$$

For each fixed t , this TPBVP must be solved for $x(\tau)$ and $\lambda(\tau)$, $t_0 \leq \tau \leq t$ since the boundary condition $\lambda(t) = 0$ must be re-enforced at each new t . To emphasize that x , λ , Q , and $P(t_0)$ are also functions of ϵ , we shall often write $x(\tau, \epsilon)$, $\lambda(\tau, \epsilon)$, etc.

Because of this and the regularity conditions which were imposed on the system model in Chapter II, we know that solutions to (3.16) - (3.17) are continuously differentiable up to arbitrary order $N + 1$ (see Coddington and Levinson [17]).

Thus, the solution of (3.16) - (3.18) can be written in the form

$$x(\tau, \epsilon) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!} x^k(\tau, 0) \quad (3.19)$$

and

$$\lambda(\tau, \epsilon) = \sum_{k=0}^{\infty} \frac{\epsilon^k}{k!} \lambda^k(\tau, 0), \quad (3.20)$$

where

$$x^k(\tau, 0) = \left. \frac{\partial^k}{\partial \epsilon^k} x(\tau, \epsilon) \right|_{\epsilon=0} \quad \text{and}$$

$$\lambda^k(\tau, 0) = \left. \frac{\partial^k}{\partial \epsilon^k} \lambda(\tau, \epsilon) \right|_{\epsilon=0}$$

The solution (3.19) is the optimal trajectory. In practice, we must be content with only a finite series approximation to (3.19) and (3.20) which, of course, is suboptimal. Denote this suboptimal solution by x^* and λ^* . Then

$$x^*(\tau, \epsilon) = \sum_{k=0}^{N-1} \frac{\epsilon^k}{k!} x^k(\tau, 0) \quad (3.21)$$

and

$$\lambda^*(\tau, \epsilon) = \sum_{k=0}^{N-1} \frac{\epsilon^k}{k!} \lambda^k(\tau, 0). \quad (3.22)$$

We call x^* the Nth-order near-optimal estimate.

Our task now is to solve for the x^k and λ^k terms in (3.21) and (3.22).

Step 0: (x^0, λ^0) is the zeroth-order solution which results from solving the TPBVP which is obtained by setting $\epsilon = 0$ in (3.16) - (3.18). That TPBVP is (where the arguments τ and $\epsilon = 0$ are to be understood unless otherwise noted)

$$\dot{x}^0 = Fx^0 - Q^0\lambda^0 \quad (3.23)$$

$$\dot{\lambda}^0 = -F^*\lambda^0 + H^*R^{-1}[y(\tau) - Hx^0] \quad (3.24)$$

$$x^0(t_0) = \mu + P^0(t_0)\lambda^0(t_0),$$

$$\lambda^0(t) = 0 \quad (3.25)$$

Observe that if we now write out (3.23) - (3.25) in partitioned form by subsystem we obtain a set of s linear decoupled TPBVP's. For $i = 1, 2, \dots, s$

$$\dot{x}_i^0 = F_i x_i^0 - Q_i \lambda_i^0 \quad (3.26)$$

$$\dot{\lambda}_i^0 = -F_i^* \lambda_i^0 + H_i^* R_i^{-1} [y_i(\tau) - H_i x_i^0] \quad (3.27)$$

$$x_i^0(t_0) = \mu_i + P_i(t_0)\lambda_i^0(t_0),$$

$$\lambda_i^0(t) = 0 \quad (3.28)$$

This linear TPBVP is easily solved (see Bryson and Frazier [18] or Jazwinski [7]). Suppose α_i^0 and β_i^0 are solutions to (3.26) and (3.27) respectively with $\alpha_i^0(t_0) = \mu_i$ and

$\beta_i^0(t_0) = 0$ and that x_i^0 and y_i^0 are matrices which solve the initial value problem

$$\dot{x}_i^0 = F_i x_i^0 + Q_i y_i^0, \quad x_i^0(t_0) = P_i(t_0); \quad (3.29)$$

$$\dot{y}_i^0 = -F_i^* y_i^0 + H_i^* R_i^{-1} H_i x_i^0, \quad y_i^0(t_0) = I; \quad (3.30)$$

where I is the identity matrix. Then it can easily be verified that the solution of (3.26) - (3.28) is

$$x_i^0(\tau) = \alpha_i^0(\tau) + x_i^0(\tau) [y_i^0(\tau)]^{-1} \beta_i^0(\tau) \quad (3.31)$$

$$\lambda_i^0(\tau) = \beta_i^0(\tau) - y_i^0(\tau) [y_i^0(\tau)]^{-1} \beta_i^0(\tau) \quad (3.32)$$

assuming that the inverse exists.

To demonstrate that this is so, substitute (3.31) into (3.26),

$$\begin{aligned} \dot{\alpha}_i^0 + \dot{x}_i^0 [y_i^0(\tau)]^{-1} \beta_i^0(\tau) &= F_i \alpha_i^0 - Q_i \beta_i^0 + F_i x_i^0 [y_i^0(\tau)]^{-1} \beta_i^0(\tau) \\ &\quad + Q_i y_i^0 [y_i^0(\tau)]^{-1} \beta_i^0(\tau) \end{aligned} \quad (3.33)$$

$$\dot{\alpha}_i^0 = F_i \alpha_i^0 - Q_i \beta_i^0 + [F_i x_i^0 + Q_i y_i^0 - \dot{x}_i^0] [y_i^0(\tau)]^{-1} \beta_i^0(\tau) \quad (3.34)$$

Since α_i^0 and β_i^0 must satisfy (3.26), we have, for $\tau \in [t_0, t]$,

$$\dot{x}_i^0 = F_i x_i^0 + Q_i y_i^0. \quad (3.35)$$

Similarly, it can be shown that (3.30) holds.

Step 1: Having solved for x^0 and λ^0 in Step 0, we

differentiate (3.16) - (3.18) with respect to ϵ and set $\epsilon = 0$ to obtain the set of s linear decoupled TPBVP's

$$\dot{x}_i^1 = F_i x_i^1 - Q_i \lambda_i^1 + A_i^0(x^0, \lambda^0, \tau) \quad (3.36)$$

$$\dot{\lambda}_i^1 = -F_i^1 \lambda_i^1 - H_i^1 R_i^{-1} H_i x_i^1 + B_i^0(x^0, \lambda^0, y, \tau) \quad (3.37)$$

$$x_i^1(t_0) = P_i(t_0) \lambda_i^1(t_0) + \sum_{j \neq 1}^s P_{ij}(t_0) \lambda_j^0(t_0),$$

$$\lambda_i^1(t) = 0 \quad (3.38)$$

where

$$A_i^0(x^0, \lambda^0, \tau) = C_i(x^0, \tau) - \sum_{j \neq 1}^s Q_{ij} \lambda_j^0 \quad (3.39)$$

$$B_i^0(x^0, \lambda^0, y, \tau) = [d_x^*(x^0, \tau) R^{-1} (y - Hx^0) - c_x^*(x^0, \tau) \lambda^0 - H^* R^{-1} d(x^0, \tau)]_i \quad (3.40)$$

do not depend on x^1 or λ^1 . The $[\cdot]_i$ in (3.40) is used to denote the i th-partitioned block of the enclosed vector quantity.

Now, (3.36) - (3.38) is a linear TPBVP and can be solved in the same manner as in Step 0. Let α_i^1 and β_i^1 be solutions of (3.36) and (3.37) respectively with $\alpha_i^1(t_0) = \beta_i^1(t_0) = 0$, then x_i^1 and λ_i^1 are given by

$$x_i^1(\tau) = \alpha_i^1(\tau) + x_i^1(\tau) [Y_i^1(\tau)]^{-1} \beta_i^1(\tau) \quad (3.41)$$

$$\lambda_i^1(\tau) = \beta_i^1(\tau) - \gamma_i^1(\tau)[Y_i^1(t)]^{-1}\beta_i^1(t) \quad (3.42)$$

where the matrices X_i^1 and Y_i^1 solve the initial value problem

$$\dot{X}_i^1 = F_i X_i^1 + Q_i Y_i^1, \quad X_i^1(t_0) = \emptyset \quad (3.43)$$

$$\dot{Y}_i^1 = -F_i Y_i^1 + H_i R_i^{-1} H_i X_i^1, \quad Y_i^1(t_0) = I \quad (3.44)$$

Step k: Having solved for x^j and λ^j in step j ,

$j = 0, 1, 2, \dots, k-1$, differentiate (3.16) - (3.18) k times and set $\epsilon = 0$ to obtain the set of s linear decoupled TPBVP's

$$\dot{x}_i^k = F_i x_i^k - Q_i \lambda_i^k + A_i^{k-1} \quad (3.45)$$

$$\dot{\lambda}_i^k = -F_i \lambda_i^k - H_i R_i^{-1} H_i x_i^k + B_i^{k-1} \quad (3.46)$$

$$x_i^k(t_0) = P_i(t_0)\lambda_i^k(t_0) + \kappa \sum_{j=1}^s P_{ij}(t_0)\lambda_j^{k-1}(t_0),$$

$$\lambda_i^k(t) = 0 \quad (3.47)$$

where A_i^{k-1} and B_i^{k-1} do not depend on x^k and λ^k .

The TPBVP can be solved in the same manner as those of the previous steps. Let α_i^k and β_i^k be solutions of (3.45) and (3.46) respectively with $\alpha_i^k(t_0) = \beta_i^k(t_0) = 0$, then

$$x_i^k(\tau) = \alpha_i^k(\tau) + x_i^1(\tau)[Y_i^1(t)]^{-1}\beta_i^k(t) \quad (3.48)$$

$$\lambda_i^k(\tau) = \beta_i^k(\tau) - \gamma_i^1(\tau)[Y_i^1(t)]^{-1}\beta_i^k(t) \quad (3.49)$$

where the matrices X_i^1 and Y_i^1 solve (3.43) - (3.44).

Let us summarize briefly the computational procedure for determining the Nth-order near-optimal estimator.

Stage 0: For $i = 1, 2, \dots, s$ and $t_0 \leq \tau \leq t$, for fixed t , perform the following three steps.

Step 1: Solve the initial-value problem

$$\dot{\alpha}_i^0 = F_i \alpha_i^0 - Q_i \beta_i^0, \quad \alpha_i^0(t_0) = \mu_i \quad (3.50)$$

$$\dot{\beta}_i^0 = -F_i \beta_i^0 + H_i^T R_i^{-1} [y_i(\tau) - H_i \alpha_i^0], \quad \beta_i^0(t_0) = 0 \quad (3.51)$$

Step 2: Solve the matrix initial-value problem

$$\dot{X}_i^0 = F_i X_i^0 + Q_i Y_i^0, \quad X_i^0(t_0) = P_i(t_0) \quad (3.52)$$

$$\dot{Y}_i^0 = -F_i Y_i^0 + H_i^T R_i^{-1} H_i X_i^0, \quad Y_i^0(t_0) = I \quad (3.53)$$

Step 3: Compute the zeroth-order term in the estimator via equations (3.31) and (3.32).

Stage k: ($k = 1, 2, \dots, N$) for $i = 1, 2, \dots, s$ and $t_0 \leq \tau \leq t$, for fixed t , perform three steps.

Step 1: Solve the initial-value problem

$$\dot{\alpha}_i^k = F_i \alpha_i^k - Q_i \beta_i^k + A_i^{k-1}, \quad \alpha_i^k(t_0) = 0 \quad (3.54)$$

$$\dot{\beta}_i^k = -F_i \beta_i^k - H_i^T R_i^{-1} H_i \alpha_i^k + B_i^{k-1}, \quad \beta_i^k(t_0) = 0 \quad (3.55)$$

Step 2: Perform this step if $k = 1$. If $k > 1$, skip and go on to Step 3. Solve the matrix initial-value problem

$$\dot{X}_i^1 = F_i X_i^1 + Q_i Y_i^1, \quad X_i^1(t_0) = \emptyset \quad (3.56)$$

$$\dot{Y}_i^1 = -F_i^* Y_i^1 + H_i^* R_i^{-1} H_i X_i^1, \quad Y_i^1(t_0) = I \quad (3.57)$$

Step 3: Compute the kth-order term in the estimator via equations (3.48) and (3.49).

Stage N: For the Nth-order near-optimal estimator,

$$x^*(\tau, 1) = \sum_{k=0}^{N-1} \frac{1}{k!} x^k(\tau, 0), \quad t_0 \leq \tau \leq t. \quad (3.58)$$

If an estimate is desired for some $x(t_1)$, $t_1 > t$, set $t = t_1$, return to Stage 0, and repeat the procedure.

Several observations can be made about the above algorithm. First, all computations are carried out at the subsystem level. Hence, for large composite systems we would expect a considerable savings in computation over that required by other methods which attempt to solve the original $2n$ -dimensional TPBVP, if only a small number of terms in the series are required.

Secondly, λ^k must be carried along at each stage so that A^k and B^k can be computed in the next stage.

In the present form, the algorithm obtains the smoothed estimates whether or not they are of interest. It may happen that only the filtered estimates are required.

Finally, recursive filtering via this algorithm

requires iterating through all of the stages for each new value of t and suffers from a "growing memory" problem since the differential equations must be solved from t_0 for each new t .

In computing smoothed estimates of the state of a system, we would expect to encounter computational difficulties such as those above. Rarely would we expect to perform smoothing in real-time. As an off-line procedure, the above algorithm becomes practical with the use of a general-purpose computer and the utilization of "slow" memory where required.

However, if the task is to perform recursive filtering, then the above algorithm is clearly unsuitable for real-time operation and the memory requirements make off-line applications possible only for small problems. For the purpose of recursive filtering, we can reduce the algorithm to a more suitable one which does not have the growing memory problem and which does not require computation of the costate.

For purposes of identification, we designate the first algorithm as the smoothing algorithm and the filtering algorithm which we will now derive as the recursive filtering, or just filtering, algorithm.

Recall that the zeroth-order term in the filtered estimate is obtained by evaluating equation (3.31) at

$\tau = t$. Thus,

$$x_i^0(t) = \alpha_i^0(t) + P_i^0(t)\beta_i^0(t) \quad (3.59)$$

where we have set

$$x_i^0(t)[Y_i^0(t)]^{-1} = P_i^0(t). \quad (3.60)$$

Suppose now that we differentiate (3.59) with respect to t and recall that α_i^0 and β_i^0 must satisfy (3.50) and (3.51) even at $\tau = t$. Then

$$\begin{aligned} dx_i^0/dt &= d\alpha_i^0/dt + [dP_i^0/dt]\beta_i^0 + P_i^0[d\beta_i^0/dt] \\ &= F_i x_i^0 + P_i^0 H_i^* R_i^{-1} [y_i(t) - H_i x_i^0] \\ &\quad + [dP_i^0/dt - F_i P_i^0 - P_i^0 F_i^* - Q_i \\ &\quad + P_i^0 H_i^* R_i^{-1} H_i P_i^0] \beta_i^0. \end{aligned} \quad (3.61)$$

If we require that $P_i^0(t)$ satisfy

$$dP_i^0/dt = F_i P_i^0 + P_i^0 F_i^* + Q_i - P_i^0 H_i^* R_i^{-1} H_i P_i^0 \quad (3.62)$$

then

$$dx_i^0/dt = F_i x_i^0 + P_i^0 H_i^* R_i^{-1} [y_i(t) - H_i x_i^0]. \quad (3.63)$$

To determine the initial conditions on (3.62) and (3.63), let $t = t_0$ in (3.28) and recall the definition of P_i^0 , equation (3.60), and the initial conditions on X_i^0 and Y_i^0 ,

$$\begin{aligned}
 x_i^0(t_0) &= \mu_i, \\
 P_i^0(t_0) &= P_i(t_0)
 \end{aligned}
 \tag{3.64}$$

Thus, the zeroth-order term in the near-optimal recursive filter is a set of s decoupled Kalman filters.

Following the same procedure, similar results can be obtained for the higher order terms in the series. Therefore, let us summarize briefly the N th-order near-optimal recursive filtering algorithm.

Stage 0: For $i = 1, 2, \dots, s$ and $t \geq t_0$, the zeroth-order term is given by

$$\begin{aligned}
 \dot{P}_i^0 &= F_i P_i^0 + P_i^0 F_i^T + Q_i - P_i^0 H_i^T R_i^{-1} H_i P_i^0, \\
 P_i^0(t_0) &= P_i(t_0)
 \end{aligned}
 \tag{3.65}$$

$$\begin{aligned}
 \dot{x}_i^0 &= F_i x_i^0 + P_i^0 H_i^T R_i^{-1} [y_i(t) - H_i x_i^0], \\
 x_i^0(t_0) &= \mu_i
 \end{aligned}
 \tag{3.66}$$

where the dot now denotes differentiation with respect to t .

Stage k : ($k = 1, 2, \dots, N$) for $i = 1, 2, \dots, s$ and $t \geq t_0$, the k th-order term is given by

$$\begin{aligned}
 \dot{P}_i^k &= F_i P_i^k + P_i^k F_i^T + Q_i - P_i^k H_i^T R_i^{-1} H_i P_i^k, \\
 P_i^k(t_0) &= \emptyset
 \end{aligned}
 \tag{3.67}$$

$$\begin{aligned} \dot{x}_i^k &= [F_i - P_i^1 H_i^T R_i^{-1} H_i] x_i^k + P_i^1 B_i^{k-1} + A_i^{k-1}, \\ x_i^k(t_0) &= 0 \end{aligned} \quad (3.68)$$

where P_i^1 is computed only at Stage 1.

Stage N: The Nth-order near-optimal recursive filter is given by

$$x^*(t,1) = \sum_{k=0}^{N-1} \frac{1}{k!} x^k(t,0), \quad t \geq t_0. \quad (3.69)$$

The filter is recursive in the following sense. We have in mind, of course, the implementation of the algorithm on a digital machine. The filtered estimate is specified by the solutions of (3.65) - (3.68) evaluated at t . For the filtered estimate at $t + \Delta t$, one merely integrates (3.65) - (3.68) forward to $t + \Delta t$. The only quantities which must be stored are the values of the solutions at t . Once the values at $t + \Delta t$ have been obtained, the values at t may be discarded. Thus, the filtering process proceeds from step to step, the present step depending only on the previous one.

We have derived approximate algorithms for the performance of nonlinear smoothing and recursive filtering which require us to solve only linear equations. The

computations are performed at the lower dimensional subsystem level, and we expect that it will be most useful in applications to large system problems. Note also that the "gains" can be precomputed if desired, whereas in the extended Kalman filter this is not true because of the need to re-nominalize about each new estimate.

The degree to which the solutions resulting from the algorithms approximate the optimal solutions depends only on the number of terms in the series that are computed. The natural question then asks how many terms are enough. A partial answer to this question will be given in the next section.

B. Performance Analysis

The degree of approximation achieved by the Nth-order near-optimal estimator can be determined at least qualitatively via a theorem first stated and proved by Werner and Cruz [10] and later generalized by Cruz [13]. It was used to assess the suboptimality of a design technique for feedback control systems subject to parameter variations. Although stated for a control problem, the result translates directly to the modal-trajectory estimation problem.

Suppose that we evaluate J_t for both $x(\cdot, \epsilon)$ and

$x^*(\cdot, \epsilon)$. Denote these values by $\hat{J}_t(\epsilon)$ and $J_t^*(\epsilon)$ respectively. If we now expand both $\hat{J}_t(\epsilon)$ and $J_t^*(\epsilon)$ in a series about $\epsilon = 0$, we get the following result, which is applicable to both smoothing and filtering.

THEOREM 3.1. For an Nth-order near-optimal estimator, the first $2N$ terms of the series for J_t^* are equal to the corresponding $2N$ terms of the series for \hat{J}_t ,

$$\left. \frac{d^k}{d\epsilon^k} J_t^*(\epsilon) \right|_{\epsilon=0} = \left. \frac{d^k}{d\epsilon^k} \hat{J}_t(\epsilon) \right|_{\epsilon=0}, \quad k = 0, 1, 2, \dots, 2N-1 \quad (3.70)$$

Proof. See Chapter 6 of Cruz [13].

Although the above theorem tells us that the terms in the series match one another, for up to the $2N$ th-order, no estimate of the difference between the two series is available. Thus, it fails to answer quantitatively the question of how good the performance of our estimator is. We might speculate that the answer can be found by determining bounds on the remaining terms in the series for \hat{J}_t and J_t^* . Unfortunately, a procedure for determining such bounds has not yet been discovered.

C. Example

In this section we present a simple application of the near-optimal recursive filtering algorithm to demon-

strate that the method does work for at least some problems. A full computational study of the algorithm, although desirable, is beyond the scope of this work. It should also be noted that the computational advantages of the algorithm cannot be fully realized on small problems. The advantage of the subsystem level computations become more pronounced with larger composite systems.

For our example we consider the following system

$$\dot{x}_1 = ax_1 + bx_1^2 + cx_2 + w_1 \quad (3.71)$$

$$\dot{x}_2 = dx_2 + w_2 \quad (3.72)$$

$$\dot{x}_3 = ex_3 + w_3 \quad (3.73)$$

$$y = x_1 + x_1^2 + fx_3 + v \quad (3.74)$$

$$x_1(0) \sim N(\mu_1, p_1(0)), x_2(0) \sim N(\mu_2, p_2(0)), \quad (3.75)$$

$$x_3(0) \sim N(\mu_3, p_3(0)) \quad (3.75)$$

This system may have originated as a scalar system with a Riccati type plant description and with exponentially correlated plant and measurement noise components. The states x_2 and x_3 result from augmenting the state with the correlated noise processes. Although small, this problem presents the algorithm with several elements which provide a fair test: nonlinear self-feedback in one subsystem,

a nonlinear measurement, and coupling between subsystems in both the plant dynamics and the measurement process.

Choosing to view the system as an interconnection of three scalar subsystems with states x_1 , x_2 , and x_3 respectively, the composite system has the structure depicted in Figure 3.1. The third-order near-optimal recursive filter is given by

$$\dot{p}_1^0 = 2ap_1^0 + q_1 = (p_1^0)^2/r, \quad p_1^0(0) = p_1(0) \quad (3.76)$$

$$\dot{p}_2^0 = 2dp_2^0 + q_2, \quad p_2^0(0) = p_2(0) \quad (3.77)$$

$$\dot{p}_3^0 = 2ep_3^0 + q_3, \quad p_3^0(0) = p_3(0) \quad (3.78)$$

$$\dot{x}_1^0 = ax_1^0 + p_1^0(y - x_1^0)/r, \quad x_1^0(0) = \mu_1 \quad (3.79)$$

$$\dot{x}_2^0 = dx_2^0, \quad x_2^0(0) = \mu_2 \quad (3.80)$$

$$\dot{x}_3^0 = ex_3^0, \quad x_3^0(0) = \mu_3 \quad (3.81)$$

$$\dot{p}_1^1 = 2ap_1^0 + q_1 - (p_1^1)^2/r, \quad p_1^1(0) = 0 \quad (3.82)$$

$$\dot{p}_2^1 = 2dp_2^1 + q_2, \quad p_2^1(0) = 0 \quad (3.83)$$

$$\dot{p}_3^1 = 2ep_3^1 + q_3, \quad p_3^1(0) = 0 \quad (3.84)$$

$$\dot{x}_1^1 = (a - p_1^1/r)x_1^1 + p_1^1[2x_1^0(y - x_1^0)/r] + b(x_1^0)^2 + cx_2^0, \quad x_1^1(0) = 0 \quad (3.85)$$

$$\dot{x}_2^1 = dx_2^1, \quad x_2^1(0) = 0 \quad (3.86)$$

$$\dot{x}_3^1 = ex_3^1 + p_3^1 f(y - x_1^0)/r, \quad x_3^1(0) = 0 \quad (3.87)$$

$$\begin{aligned} \dot{x}_1^2 = & (a - p_1^1/r)x_1^2 + 4p_1^1[x_1^1(y - x_1^0) - x_1^0 x_1^1 - x_1^1(x_1^0)^2 \\ & - fx_1^1 x_3^0 - x_1^0(2x_1^0 x_1^1 + fx_3^1)]/r + 2(2bx_1^1 x_1^0 + cx_2^1), \\ & x_1^2(0) = 0 \end{aligned} \quad (3.88)$$

$$\dot{x}_2^2 = dx_2^2, \quad x_2^2(0) = 0 \quad (3.89)$$

$$\begin{aligned} \dot{x}_3^2 = & ex_3^2 - 2p_3^1 f[x_1^1 + (x_1^0)^2 + fx_3^0]/r, \\ & x_3^2(0) = 0 \end{aligned} \quad (3.90)$$

$$x_i^*(t) = x_i^0(t) + x_i^1(t) + 0.5x_i^2(t), \quad i = 1, 2, 3 \quad (3.91)$$

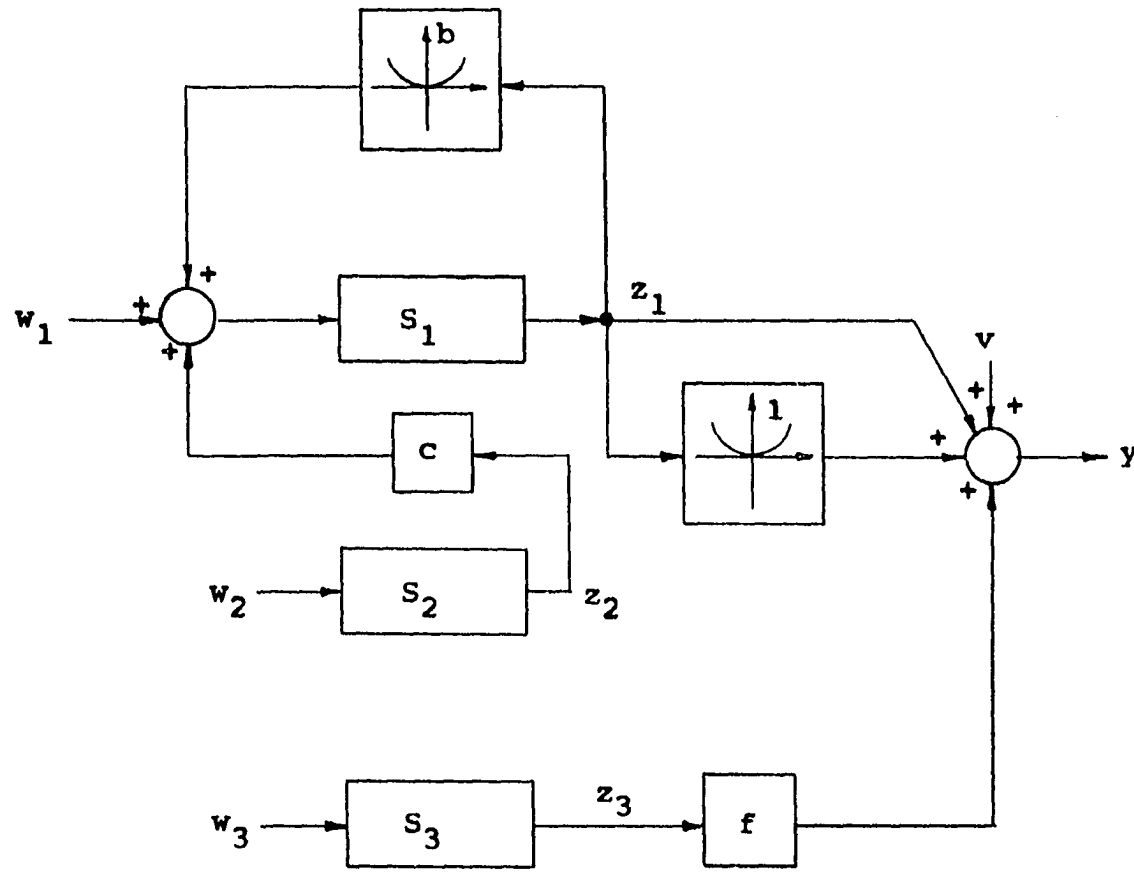


Figure 3.1. A simple nonlinear system

Notice that p_2^0 , p_3^0 , and p_2^1 are not actually needed anywhere in the computations so that equations (3.77), (3.78), and (3.83) may be deleted. Also, further simplification can be obtained by recognizing that the solutions of (3.80), (3.81), (3.86), and (3.89) can be found analytically, eliminating the necessity of integrating them numerically. Hence,

$$x_2^0(t) = \mu_2 \exp\{dt\} \quad (3.91)$$

$$x_3^0(t) = \mu_3 \exp\{et\} \quad (3.92)$$

$$x_2^1(t) = x_2^2(t) = 0 \quad (3.93)$$

for all $t \geq 0$.

The system (3.71) - (3.74) was simulated for several sets of parameter values and sample functions using rectangular integration with a step size of 0.001. Simultaneously, the above near-optimal modal-trajectory filter, and for comparative purposes, the extended Kalman filter for this system were simulated.

In all cases, the a , d , and c parameters were set as follows:

$$a = -0.5 \quad \text{cov}\{v(t_1), v(t_2)\} = r\delta(t_1 - t_2) = 0.1\delta(t_1 - t_2)$$

$$d = -100.0 \quad \text{cov}\{w(t_1), w(t_2)\} = \text{diag}\{q_1, q_2, q_3\}\delta(t_1 - t_2)$$

$$e = -100.0 \quad q_1 = q_2 = q_3 = 0.1$$

The values for d and e were chosen to make the response of the correlated noise states appear nearly white relative to the response of x_1 .

In Figures 3.2 - 3.7, for four sets of parameters, a typical sample function and the estimation error resulting from the extended Kalman filter, the second-order near-optimal filter, and the third-order near-optimal filter were plotted on semi-log axes. Because of the wide range of values encountered, this type of scaling was required. The results for states x_2 and x_3 are given for only one set of parameters as they did not change appreciably from one set to the next.

In Figure 3.2 note that after an initial transient period the third-order near-optimal filter produces a smaller estimation error than the extended Kalman filter. The second-order filter does better than the extended Kalman filter at times, but not consistently so. Thus, it appears that the improvement in performance warrants the addition of the third term in the series for this example.

In Figures 3.3 and 3.4 we see that all three filters do an equally poor job of tracking the correlated noise states.

Figure 3.5 demonstrates the result of increasing the effect of the plant nonlinearity. The extended Kalman

filter performed about the same. However, the near-optimal filters did well until near the end of the run, when the estimation error began to increase.

Figure 3.6 illustrates the effect of increasing the coupling in the plant dynamics. Very little change in performance resulted.

In Figure 3.7 we find, as expected, that the effect of decreasing the coupling in the measurement results in even better performance by the near-optimal filters.

From this example, we can draw several preliminary conclusions regarding the performance of the near-optimal modal-trajectory filter. The near-optimal filter should perform well under conditions of light coupling and small nonlinearities. One might reply that this is true also of the extended Kalman filter, which is less tedious to design. However, it should be remembered that the near-optimal filter is intended for application to large-dimensional problems where its decoupling properties provide significant computational savings over methods such as the extended Kalman filter. In addition, unlike other approximate filters, performance can be improved by computing more terms in the series, with Theorem 3.1 indicating how much the performance is to be improved.

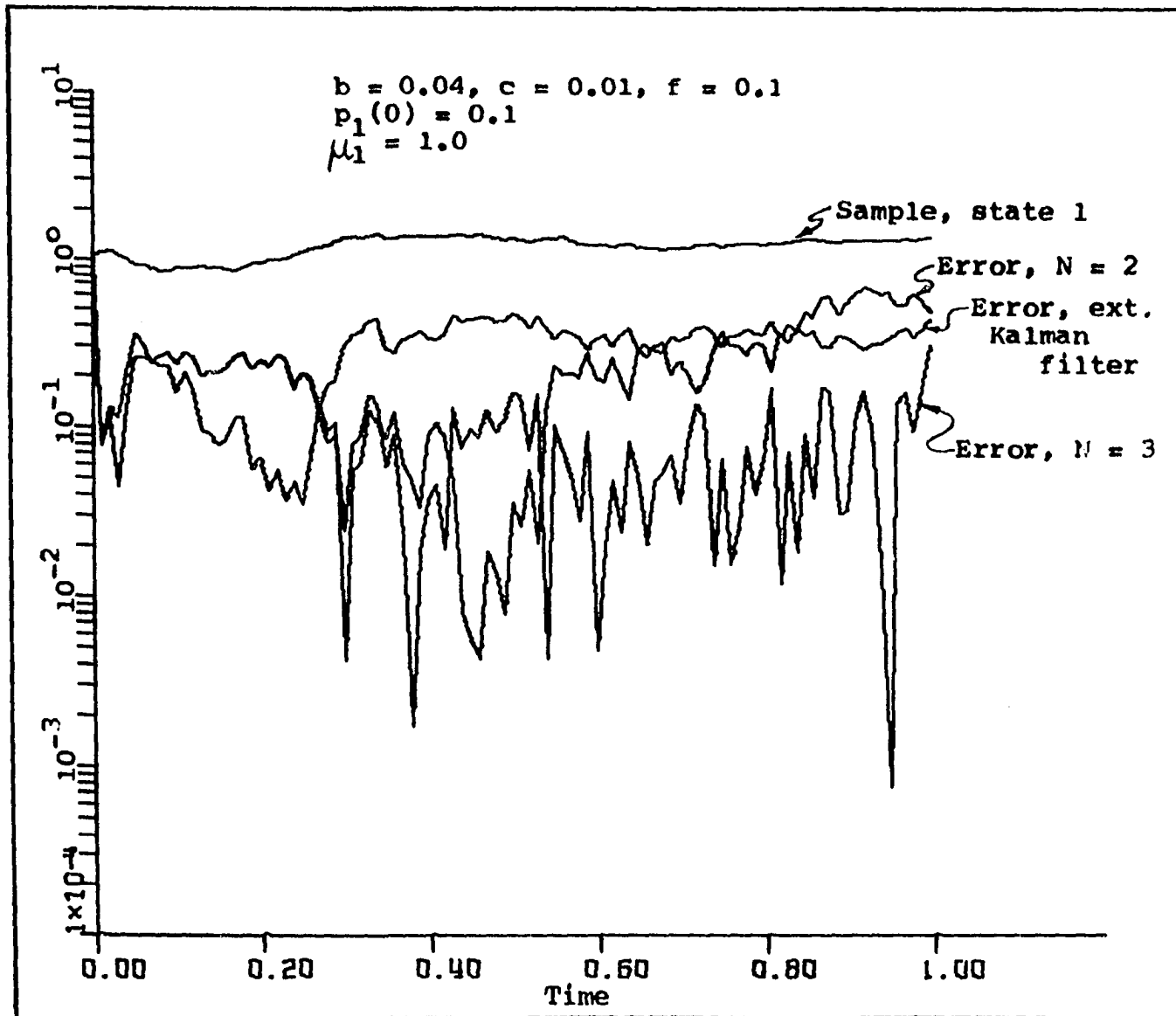


Figure 3.2. Filter performance for state 1

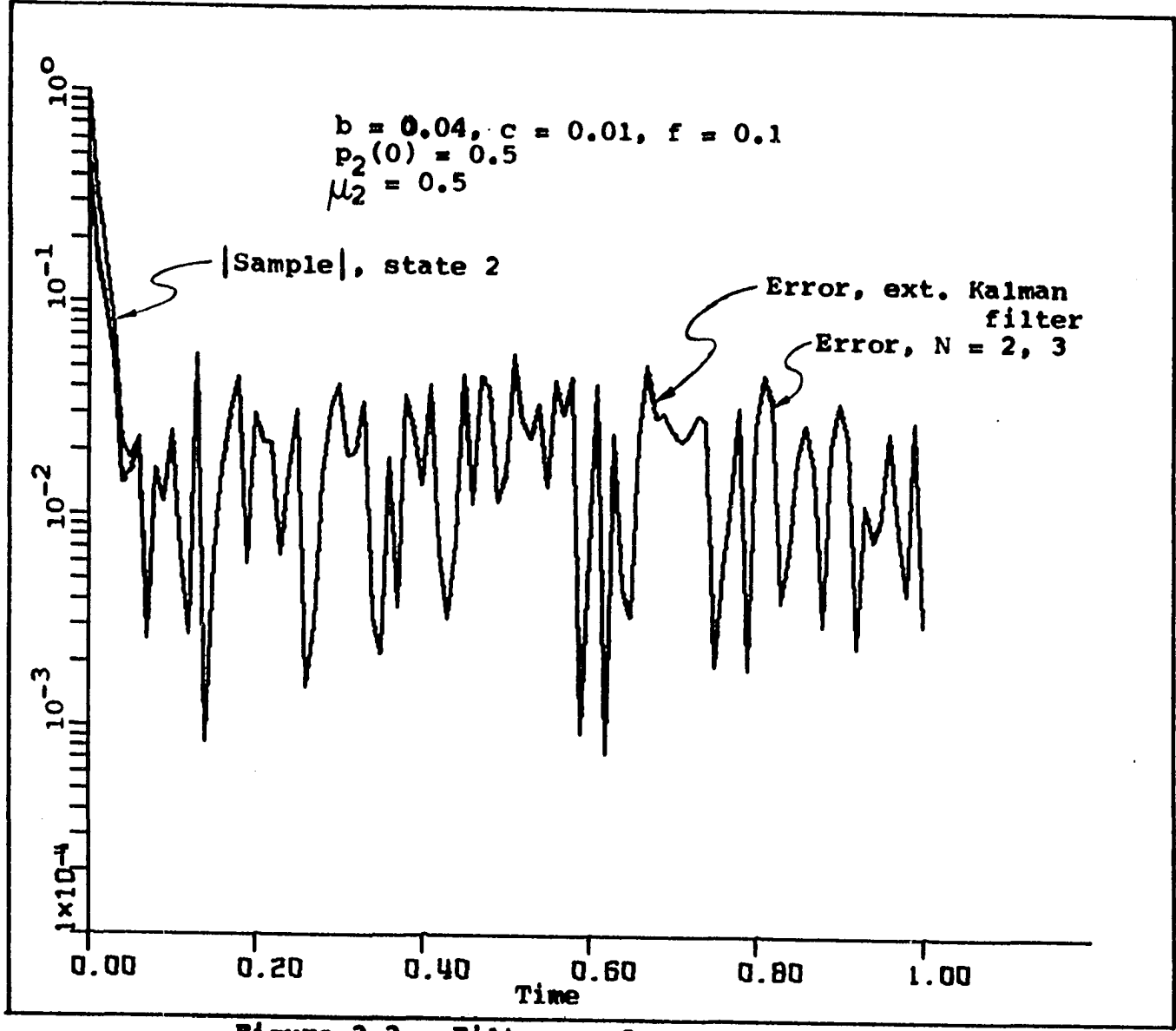


Figure 3.3. Filter performance for state 2

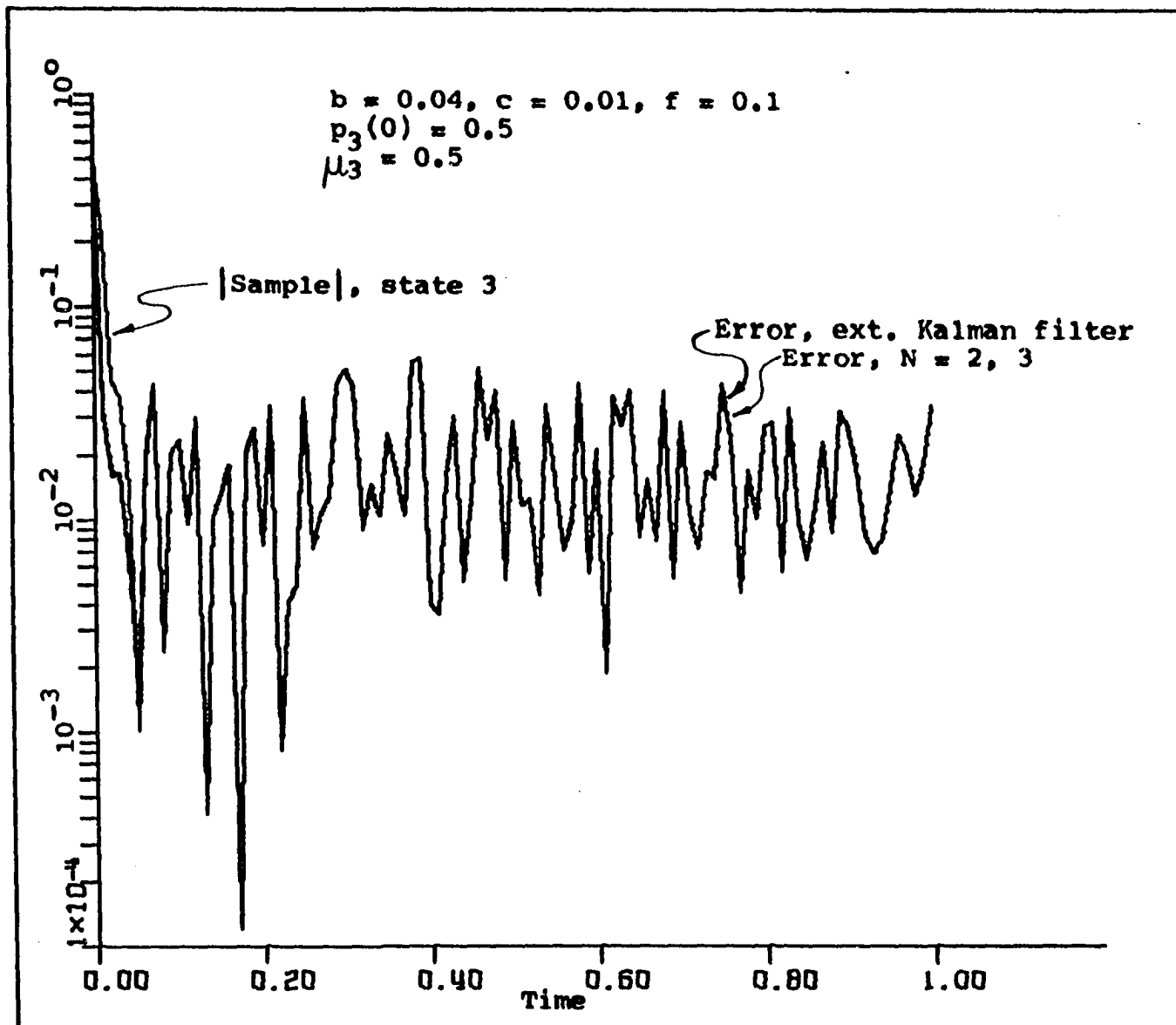


Figure 3.4. Filter performance for state 3

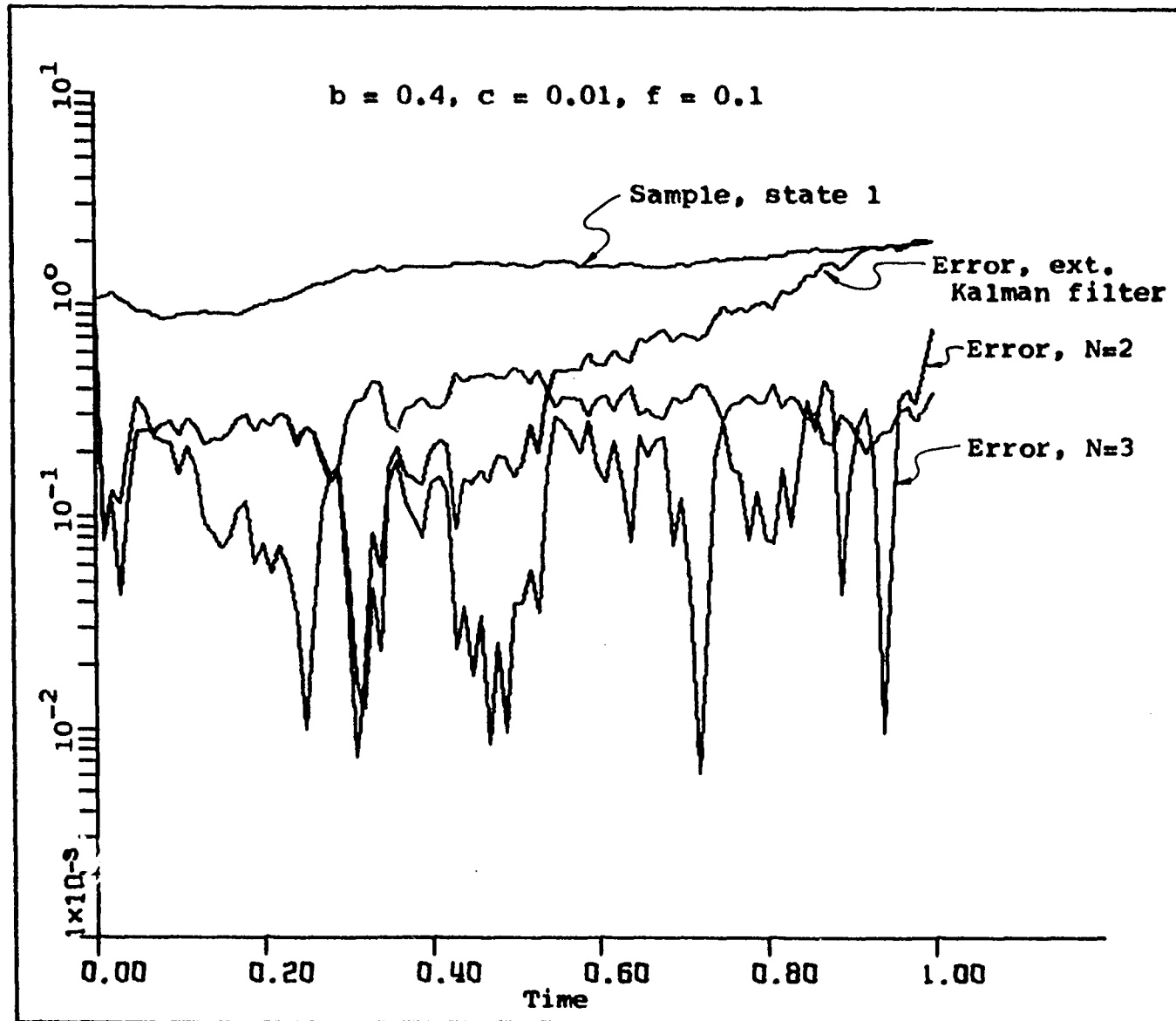


Figure 3.5. Filter performance for increased plant nonlinearity

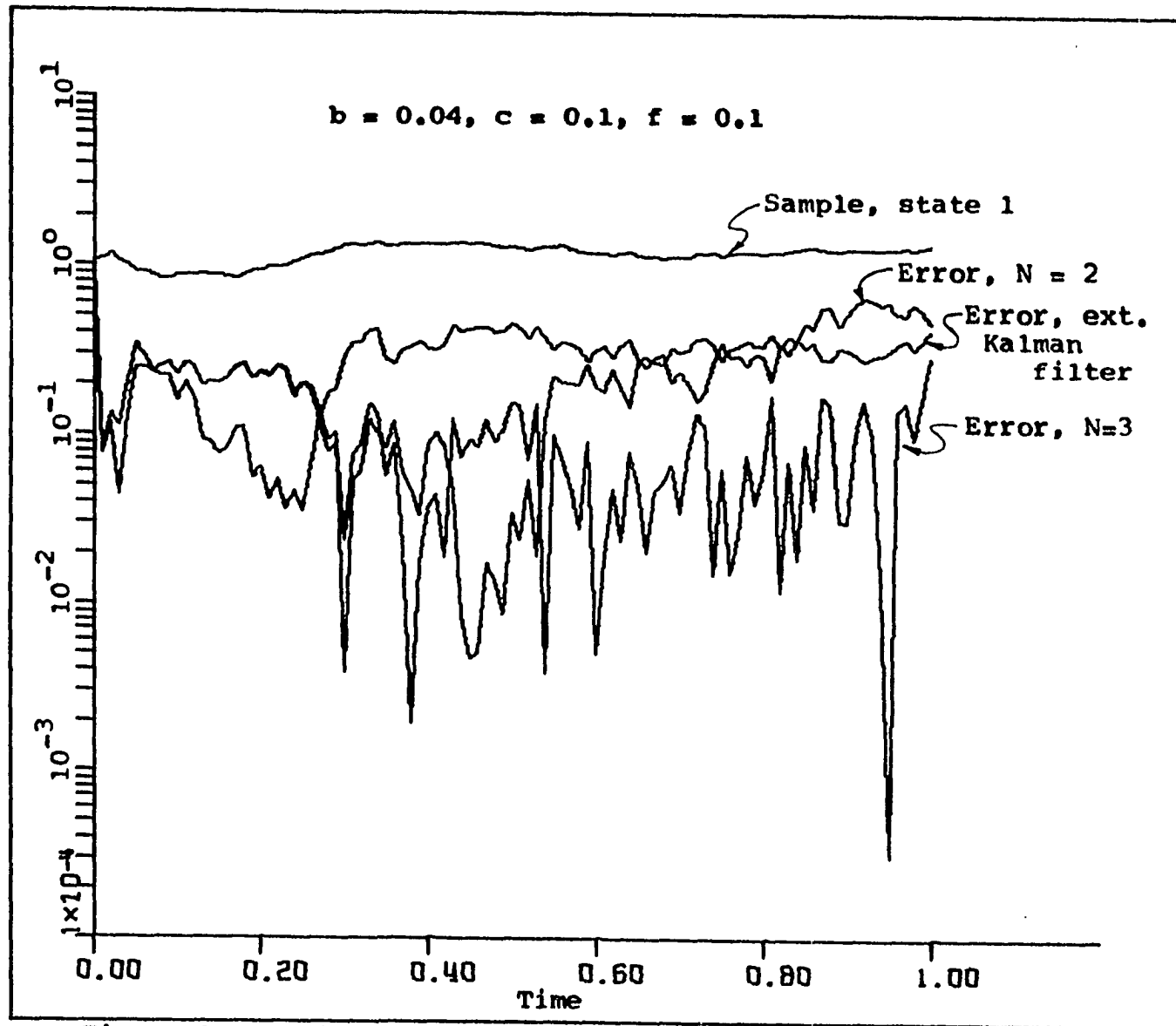


Figure 3.6. Filter performance for increased plant coupling

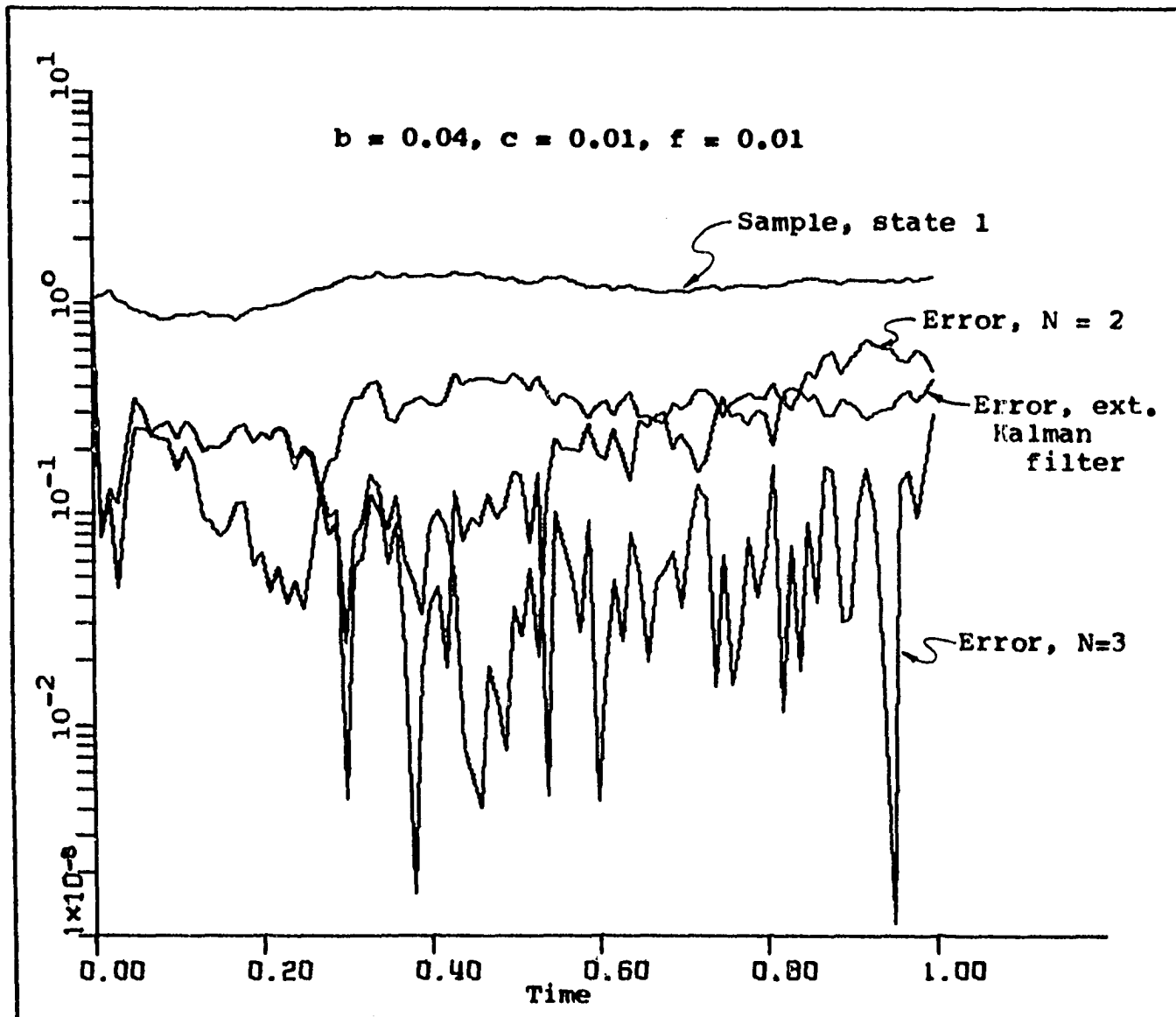


Figure 3.7. Filter performance for decreased measurement coupling

IV. STATE ESTIMATION FOR LINEAR INTERCONNECTED SYSTEMS

In this chapter we present two approaches to filtering for linear interconnected systems. Both approaches are based on the ϵ -coupling technique. Unlike the nonlinear problem, the linear problem has a known closed-form solution, the Kalman filter. See the Appendix for a statement of the linear estimation problem and a summary of the Kalman filter equation. For large systems, computational problems arise, however, which may render implementation of the Kalman filter impractical in a particular application because of the severe computational limitations which may be imposed by the capability of the on-board computer. Over the years a number of ad hoc methods have been proposed to remedy the problem; see e.g. Pentecost [19], Aoki and Huddle [20], and Bucy and Joseph [21]. The ϵ -coupling method provides two systematic approaches to this problem. The computational load is relieved by performing computations at the lower-dimensional subsystem level.

Given the system model, the Kalman filter is uniquely specified by the covariance matrix of the estimation error. This matrix satisfies a matrix Riccati equation. In Section A we apply the ϵ -coupling method to this Riccati equation to obtain an approximate error covariance matrix. This

approach is restricted to two-subsystem composite systems. This approach is based on the approach of Kokotović, et al. [22] to the linear regulator problem, which was later extended by Haddad and Cruz [14] to the linear estimation problem. Although formulated within the framework of interconnected systems, our results are substantially the same as Haddad and Cruz.

In Section B we specialize the results of Chapter III to the linear case to obtain a near-optimal linear filter which is not restricted in the number of subsystems.

In Section C we give some performance results which indicate the degree to which our filters are suboptimal.

We present some examples in Section D to illustrate the use of the algorithms.

A. Two Subsystem Case

We now apply the method of Kokotović, et al. [22] to the covariance matrix, P , of the estimation error of the Kalman filter to obtain the results of Haddad and Cruz [14].

Let our interconnected system be composed of two linear subsystems which are linearly connected. Then

$$\dot{x}(t) = F(t)x(t) + \epsilon C(t)x(t) + w(t) \quad (4.1)$$

$$y(t) = H(t)x(t) + \epsilon D(t)x(t) + v(t) \quad (4.2)$$

where we have introduced the coupling parameter ϵ and

$$\mathbf{x}^* = [x_1, x_2]^T,$$

$$\mathbf{w}^* = [w_1, w_2]^T,$$

$$\mathbf{v}^* = [v_1, v_2]^T,$$

$$\mathbf{F} = \begin{bmatrix} F_1 & | & \emptyset \\ \hline \emptyset & | & F_2 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} H_1 & | & \emptyset \\ \hline \emptyset & | & H_2 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} \emptyset & | & C_{12} \\ \hline C_{21} & | & \emptyset \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} \emptyset & | & D_{12} \\ \hline D_{21} & | & \emptyset \end{bmatrix},$$

\mathbf{w} and \mathbf{v} are independent zero-mean gaussian white noise processes with

$$\text{cov}\{\mathbf{w}(t_1), \mathbf{w}(t_2)\} = \begin{bmatrix} Q_1(t_1) & | & \epsilon Q_{12}(t_1) \\ \hline \epsilon Q_{12}^*(t_1) & | & Q_2(t_1) \end{bmatrix} \delta(t_1 - t_2)$$

and

$$\text{cov}\{\mathbf{v}(t_1), \mathbf{v}(t_2)\} = \begin{bmatrix} R_1(t_1) & | & \emptyset \\ \hline \emptyset & | & R_2(t_1) \end{bmatrix} \delta(t_1 - t_2).$$

Q_1 , Q_2 , R_1 , and R_2 are assumed to be positive definite.

Given the above model, the optimal minimum variance, unbiased, a posteriori maximum-likelihood (modal-trajectory) recursive estimator is specified by the solution of the matrix Riccati equation

$$\begin{aligned} \dot{P}(t) = & [F(t) + \epsilon C(t)]P(t) + P(t)[F(t) + \epsilon C(t)]' + Q(t, \epsilon) \\ & - P(t)[H(t) + \epsilon D(t)]' R^{-1}(t)[H(t) + \epsilon D(t)]P(t), \\ P(t_0, \epsilon) & \text{ given,} \end{aligned} \quad (4.3)$$

where $P(t_0, \epsilon)$ is of the same form as $Q(t, \epsilon)$.

Since P actually depends on ϵ as well as t , we will often denote it as $P(t, \epsilon)$. $P(t, \epsilon)$ is an analytic function of ϵ on $[0, 1]$ from a theorem on the differentiability of solutions of ordinary differential equations with respect to a parameter (Coddington and Levinson [17]). Hence, the solution $P(t, \epsilon)$ can be expanded in a series about $\epsilon = 0$,

$$P(t, \epsilon) = \sum_{k=0}^{\infty} \epsilon^k p^k(t, 0) \quad (4.4)$$

where $p^k(t, 0)$ denotes $\partial^k P(t, \epsilon) / \partial \epsilon^k$ evaluated at $\epsilon = 0$ for every $t \geq t_0$.

Of course, in practice we would not expect to use (4.4), but an approximate truncated version

$$P_S(t, \epsilon) = \sum_{k=0}^{N-1} \frac{\epsilon^k}{k!} P^k(t, 0). \quad (4.5)$$

We now proceed to find the terms in the series (4.5) via the same procedure as in Chapter III. Denote

$$P^k(t, 0) = \begin{bmatrix} P_1^k(t) & | & P_{12}^k(t) \\ \hline P_{12}^k(t) & | & P_2^k(t) \end{bmatrix}.$$

Stage 0: Let $\epsilon = 0$ in (4.3) and write out the resultant equation in partitioned form

$$\begin{aligned} \dot{P}_i^0 &= F_i P_i^0 + P_i^0 F_i^* + Q_i - P_i^0 H_i^* R_i^{-1} H_i P_i^0, \\ P_i^0(t_0) &= P_i(t_0), \quad i = 1, 2 \end{aligned} \quad (4.6)$$

and

$$\dot{P}_{12}^0 = 0, \quad P_{12}^0(t_0) = \emptyset. \quad (4.7)$$

Thus, in the zeroth-order term, the off-diagonal block is zero for all t and the diagonal blocks satisfy decoupled matrix Riccati equations. Therefore, the filter corresponding to the zeroth-order term is a set of two decoupled Kalman filters. The dimensions of the filters are that of the subsystems.

Stage 1: Having solved for P^0 , differentiate (4.3)

with respect to ϵ and set $\epsilon = 0$.

$$\begin{aligned} \dot{P}^1 &= [F - P^0 H^* R^{-1} H] P^1 + P^1 [F - P^0 H^* R^{-1} H]^* + Q^1 \\ &\quad + C P^0 + P^0 C^* - P^0 [D^* R^{-1} H + H^* R^{-1} D] P^0, \\ P^1(t_0) &= \left. \frac{\partial}{\partial \epsilon} P(t_0, \epsilon) \right|_{\epsilon = 0} \end{aligned} \quad (4.8)$$

Writing out (4.8) in partitioned form yields

$$\begin{aligned} \dot{P}_1^1 &= [F_1 - P_1^0 H_1^* R_1^{-1} H_1] P_1^1 + P_1^1 [F_1 - P_1^0 H_1^* R_1^{-1} H_1]^*, \\ P_1^1(t_0) &= 0 \end{aligned} \quad (4.9)$$

$$\begin{aligned} \dot{P}_2^1 &= [F_2 - P_2^0 H_2^* R_2^{-1} H_2] P_2^1 + P_2^1 [F_2 - P_2^0 H_2^* R_2^{-1} H_2]^*, \\ P_2^1(t_0) &= 0 \end{aligned} \quad (4.10)$$

$$\begin{aligned} \dot{P}_{12}^1 &= [F_1 - P_1^0 H_1^* R_1^{-1} H_1] P_{12}^1 + P_{12}^1 [F_2 - P_2^0 H_2^* R_2^{-1} H_2]^* + Q_{12} \\ &\quad + C_{12} P_2^0 + P_1^0 C_{21}^* - P_1^0 [D_{21}^* R_2^{-1} H_2 + H_1^* R_1^{-1} D_{12}] P_2^0 \\ &\quad - P_2^0 [H_2^* R_2^{-1} D_{21} + D_{12}^* R_1^{-1} H_1] P_1^0, \quad P_{12}^1(t_0) = P_{12}(t_0) \end{aligned} \quad (4.11)$$

Since equations (4.9) and (4.10) are linear and homogeneous with zero initial conditions, $P_1^1(t)$ and $P_2^1(t)$ equal zero for all $t \geq t_0$. Thus, for Stage 1 we have

$$P_1^1(t) = \emptyset \quad (4.12)$$

$$P_2^1(t) = \emptyset \quad (4.13)$$

for all $t \geq t_0$. The off-diagonal block is obtained from

$$\dot{P}_{12}^1 = G_1 P_{12}^1 + P_{12}^1 G_2^* + Q_{12} + A_{12}^0, \quad P_{12}^1(t_0) = P_{12}(t) \quad (4.14)$$

where

$$G_1 = F_1 - P_1^0 H_1^* R_1^{-1} H_1 \quad (4.15)$$

$$G_2 = F_2 - P_2^0 H_2^* R_2^{-1} H_2 \quad (4.16)$$

$$\begin{aligned} A_{12}^0 = & C_{12} P_2^0 + P_1^0 C_{21}^* - P_1^0 [D_{21}^* R_2^{-1} H_2 + H_1^* R_1^{-1} D_{12}] P_2^0 \\ & - P_2^0 [H_2^* R_2^{-1} D_{21} + D_{12}^* R_1^{-1} H_1] P_1^0 \end{aligned} \quad (4.17)$$

Note that for this step the equation for the nonzero off-diagonal block P_{12}^1 is linear.

Stage k: ($k \geq 1$) Having solved for P^0, P^1, \dots, P^{k-1} we differentiate (4.3) with respect to ϵ and set $\epsilon = 0$. There are two cases.

Case 1: k even.

$$\dot{P}_i^k = G_i P_i^k + P_i^k G_i^* + A_i^{k-1}, \quad P_i^k(t_0) = \emptyset, \quad i = 1, 2 \quad (4.18)$$

where A_i^{k-1} does not depend on P^k .

$$P_{12}^k(t) = \emptyset \text{ for all } t \geq t_0. \quad (4.19)$$

Case 2: k odd.

$$\dot{P}_{12}^k = G_1 P_{12}^k + P_{12}^k G_2^* + A_{12}^{k-1}, \quad P_{12}^k(t_0) = \emptyset \quad (4.20)$$

where A_{12}^{k-1} does not depend on P^k .

$$P_1^k(t) = \emptyset$$

$$P_2^k(t) = \emptyset$$

for all $t \geq t_0$.

Thus, we see the terms in the filter gain which represent the cross-coupling between subsystems correspond to the odd-numbered terms in the series.

Recapitulating, the procedure for computing an approximation to the P matrix, and hence the filter gain, is as follows.

For an N-term approximation, the even-numbered terms are of the form

$$P^{2k}(t) = \left[\begin{array}{c|c} P_1^{2k}(t) & \emptyset \\ \hline \emptyset & P_2^{2k}(t) \end{array} \right] \quad k = 0, 1, 2, \dots \quad (4.21)$$

where for $k = 0$,

$$\begin{aligned} \dot{P}_i^{2k}(t) = & F_i(t)P_i^{2k}(t) + P_i^{2k}(t)F_i^*(t) + Q_i(t) \\ & - P_i^{2k}(t)H_i^*(t)R_i^{-1}(t)H_i(t)P_i^{2k}(t), \end{aligned}$$

$$P_i^{2k}(t_0) = P_i(t_0), \quad i = 1, 2 \quad (4.22)$$

and for $k > 0$,

$$\begin{aligned} \dot{P}_i^{2k}(t) &= G_i(t)P_i^{2k}(t) + P_i^{2k}(t)G_i^*(t) \\ &\quad + A_i^{2k-1}(t) \\ P_i^{2k}(t_0) &= \emptyset, \quad i = 1, 2 \end{aligned} \quad (4.23)$$

The odd-numbered terms are of the form

$$P^{2k+1}(t) = \left[\begin{array}{c|c} \emptyset & P_{12}^{2k+1}(t) \\ \hline \text{---} & \text{---} \\ \hline [P_{12}^{2k+1}(t)]^* & \emptyset \end{array} \right] \quad (4.24)$$

where for $k = 0$,

$$\begin{aligned} \dot{P}_{12}^{2k+1}(t) &= G_1(t)P_{12}^{2k+1}(t) + P_{12}^{2k+1}(t)G_2^*(t) \\ &\quad + Q_{12}(t) + A_{12}^{2k}(t), \\ P_{12}^{2k+1}(t_0) &= P_{12}(t_0) \end{aligned} \quad (4.25)$$

and for $k > 0$,

$$\begin{aligned} \dot{P}_{12}^{2k+1} &= G_1(t)P_{12}^{2k+1}(t) + P_{12}^{2k+1}(t)G_2^*(t) \\ &\quad + A_{12}^{2k}(t), \\ P_{12}^{2k+1}(t_0) &= \emptyset \end{aligned} \quad (4.26)$$

The suboptimal covariance matrix is given by

$$P_s(t,1) = \sum_{k=0}^{N-1} \frac{1}{K} P^k(t,1). \quad (4.27)$$

The principal appeal of this algorithm is that computation of the estimation error covariance can be accomplished by solving a sequence of problems whose dimensionality is lower than that of the original problem. The advantage of this is particularly apparent for rather large-dimensional composite systems. Suppose that $n = 16$ and $n_1 = n_2 = 8$. Then instead of solving $n(n+1)/2 = 136$ coupled nonlinear equations, one solves, say for $N = 3$, two decoupled sets of $n_1(n_1+1)/2 = 36$ coupled nonlinear equations and three decoupled sets of 36 coupled linear equations. It is well known that for large n , convergence difficulties plague the solution of the Riccati equation, see e.g. Meditch [23]. Also, for certain classes of problems additional savings can be obtained in computation due to the special structure of the composite system. This will be dealt with in more detail in the examples.

As we indicated earlier, the results just presented

are not new. However, Haddad and Cruz [14] presented only the algorithm without indicating for what classes of estimation problems this approach might be applicable and without demonstrating that the algorithm performs computationally in a well-behaved manner. It is the purpose of our examples to make some preliminary exploration in that direction.

For a large class of linear estimation problems the corrupting plant noise is not white. Often this correlated noise can be modeled as the output of a linear shaping filter operating on white noise. Thus, to obtain the proper form for the model required by the Kalman filter, the state vector is augmented with the states due to the correlated noise processes of the system. This model can naturally be viewed as an interconnection of two subsystems where the matrices F and C are of the form

$$F = \left[\begin{array}{c|c} F_1 & \emptyset \\ \hline \emptyset & F_2 \end{array} \right], \quad C = \left[\begin{array}{c|c} \emptyset & C_{12} \\ \hline \emptyset & \emptyset \end{array} \right] \quad (4.28)$$

where F_2 is often a diagonal matrix. The matrices H and D are of the form

$$H = \left[\begin{array}{c|c} H_1 & \emptyset \\ \hline \emptyset & \emptyset \end{array} \right], \quad D = \emptyset \quad (4.29)$$

In other words, interconnections occur only in the dynamics of the composite system.

In a similar manner, if the observations are corrupted by the sum of a correlated noise process and a white noise process, the state can be augmented to include the effects of the correlated measurement noise in the model of a form required by the Kalman algorithm. For this class of problems, the system can be naturally viewed as an interconnection of two subsystems with

$$F = \left[\begin{array}{c|c} F_1 & \emptyset \\ \hline \emptyset & F_2 \end{array} \right], \quad C = \emptyset \quad (4.30)$$

and

$$H = \left[\begin{array}{c|c} H_1 & \emptyset \\ \hline \emptyset & \emptyset \end{array} \right], \quad D = \left[\begin{array}{c|c} \emptyset & D_{12} \\ \hline \emptyset & \emptyset \end{array} \right] \quad (4.31)$$

where F_2 is often a diagonal matrix.

A large number of estimation problems in automatic navigation and control fall into the above classes of problems or their combination. Another approach to this class of problems, which utilizes lower-dimensional computations, is that of Bryson and Henrikson [24], but it is restricted to the case of correlated measurement noise only.

The next natural question is whether the above

algorithm can be generalized to include systems consisting of more than two subsystems. The answer to that question is no. The reason is thus.

In deriving the equations for the succeeding terms in the series, the following facts about the multiplication of certain classes of partitioned matrices were used.

Suppose a is any matrix in the class of block-diagonal matrices (two blocks on diagonal) and b is any matrix in the class of zero block-diagonal matrices (two blocks on diagonal); i.e. a and b are of the form

$$a = \begin{bmatrix} a_1 & | & \emptyset \\ \hline \emptyset & | & a_2 \end{bmatrix}, \quad b = \begin{bmatrix} \emptyset & | & b_1 \\ \hline b_2 & | & \emptyset \end{bmatrix}.$$

Then ab , ba , and bb belong to the class of zero block-diagonal matrices and aa belongs to the class of block-diagonal matrices. Because of this property, the equations for the diagonal blocks in the odd-numbered terms are homogeneous with zero initial conditions, resulting in zero blocks on the diagonal. Similarly, the same result is true for the off-diagonal blocks in the even-numbered terms.

Now, suppose we generalize to s subsystems, then F , C , H , and D are of the form

$$F = \begin{bmatrix} F_1 & & & & \\ & F_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ \emptyset & & & & \cdot \\ & & & & & F_s \end{bmatrix}, \quad C = \begin{bmatrix} \emptyset & C_{12} & \cdot & \cdot & \cdot & C_{1s} \\ C_{21} & \emptyset & & C_{23} & \cdot & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ C_{s1} & \cdot & \cdot & \cdot & \cdot & \emptyset \end{bmatrix}$$

and

$$H = \begin{bmatrix} H_1 & & & & \\ & H_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ \emptyset & & & & \cdot \\ & & & & & H_s \end{bmatrix}, \quad D = \begin{bmatrix} \emptyset & D_{12} & \cdot & \cdot & \cdot & D_{1s} \\ D_{21} & \emptyset & & D_{23} & \cdot & \cdot \\ \cdot & \cdot & & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ D_{s1} & \cdot & \cdot & \cdot & \cdot & \emptyset \end{bmatrix}$$

In deriving the equations to be satisfied by the terms in the series, we must consider products of matrices of the form

$$a = \begin{bmatrix} a_1 & & & & \\ & a_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ \emptyset & & & & \cdot \\ & & & & & a_s \end{bmatrix}, \quad b = \begin{bmatrix} \emptyset & b_{12} & \cdot & \cdot & \cdot & b_{1s} \\ b_{21} & \emptyset & & \cdot & \cdot & b_{2s} \\ \cdot & \cdot & & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ b_{s1} & \cdot & \cdot & \cdot & \cdot & \emptyset \end{bmatrix}.$$

Then aa is of the same form as a and ab and ba are of the same form as b . But bb in general is a full matrix, not of the form of either a or b . Hence, for the s subsystem

case, multiplication within the classes of partitioned matrices encountered is not "closed."

Because of this lack of "closure" property, terms of the form $CP^{k-1} + P^{k-1}C'$, for $k \geq 2$, appear which result in coupled nonhomogeneous equations for the subsystem partitions. Therefore, the series no longer alternates with matrices of the form a and b and P^k can no longer be computed via uncoupled sets of equations, which is the chief justification for using the method.

How then do we generalize to the s subsystem case? Clearly, we can specialize the results of Chapter III to the linear system, linear connections case. This will be done in Section B. In exchange for the ability to deal with arbitrary numbers of subsystems, we lose some of the simple intuitive interpretation associated with the above method and our ability to quantitatively assess the performance of the resulting filter.

B. s -Subsystem Case

In this section we specialize the results of Chapter III to the case of linear subsystems with linear interconnections and observations. Thus,

$$C = \begin{bmatrix} \emptyset & C_{12} & \dots & C_{1s} \\ C_{21} & \emptyset & C_{23} & \dots & C_{2s} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ C_{s1} & \cdot & \cdot & \cdot & \emptyset \end{bmatrix}, \quad D = \begin{bmatrix} \emptyset & D_{12} & \dots & D_{1s} \\ D_{21} & \emptyset & D_{23} & \dots & D_{2s} \\ \cdot & \cdot & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \cdot \\ D_{s1} & \cdot & \cdot & \cdot & \emptyset \end{bmatrix}. \quad (4.32)$$

so that the Nth-order near-optimal modal-trajectory linear filter is given by

Stage 0: For $i = 1, 2, \dots, s$ and $t \geq t_0$, the zeroth-order term is given by

$$\begin{aligned} \dot{P}_i^0 &= F_i P_i^0 + P_i^0 F_i^0 + Q_i - P_i^0 H_i^0 R_i^{-1} H_i^0 P_i^0, \\ P_i^0(t_0) &= P_i(t_0) \end{aligned} \quad (4.33)$$

$$\begin{aligned} \dot{x}_i^0 &= F_i x_i^0 + P_i^0 H_i^0 R_i^{-1} [y_i(t) - H_i x_i^0], \\ x_i^0(t_0) &= \mu_i \end{aligned} \quad (4.34)$$

Stage k: ($k = 1, 2, \dots, N-1$) for $i = 1, 2, \dots, s$ and $t \geq t_0$, the kth-order term is given by

$$\begin{aligned} \dot{P}_i^k &= F_i P_i^k + P_i^k F_i^0 + Q_i - P_i^k H_i^0 R_i^{-1} H_i^0 P_i^k, \\ P_i^k(t_0) &= \emptyset \end{aligned} \quad (4.35)$$

$$\begin{aligned} \dot{x}_i^k &= [F_i - P_i^k H_i^0 R_i^{-1} H_i^0] x_i^k + P_i^k B_i^{k-1} + A_i^{k-1}, \\ x_i^k(t_0) &= 0 \end{aligned} \quad (4.36)$$

where P_i^1 is computed only at Stage 1 and B_i^{j-1} and A_i^{j-1} do not depend on x^k , $j = 1, 2, \dots, k$.

Stage N: The Nth-order near-optimal recursive linear filter is given by

$$x^*(t,1) = \sum_{k=0}^{N-1} \frac{1}{k!} x^k(t,0), \quad t \geq t_0. \quad (4.37)$$

Note that the expense we pay for the flexibility of allowing any number of subsystems, attended by possibly greater computational savings, is that no intuitively appealing interpretation in terms of filter gain can be made as could for the method of Section A. Perhaps the greatest drawback is that this fact prevents us from making a complete assessment of the performance of the filter.

C. Performance Analysis

In this section we attempt to answer the questions of how suboptimal the above two filters are. In the latter case, a partial answer can be given. The theorem of Chapter III is applicable and for an Nth-order approximation the optimal cost functional is approximated up to the $2N$ -order.

However, a more complete accounting can be given for

the filter of Section A. If the optimal estimation error covariance matrix is used, the filter gain is given by

$$K(t) = P(t)[H(t) + D(t)]'R^{-1}(t) \quad (4.38)$$

where $P(t)$ satisfies (4.3) with $\epsilon = 1$. The value of the optimal cost functional is given by

$$\hat{J}_t = \text{tr}P(t). \quad (4.39)$$

If the N th-order approximation P_S to P is used, the filter gain is given by

$$K_S(t) = P_S(t,1)[H(t) + D(t)]'R^{-1}(t) \quad (4.40)$$

and the value of the cost functional is given by

$$J_t^* = \text{tr}P_a(t) \quad (4.41)$$

where P_a is the covariance matrix of the actual estimation error resulting from the use of K_S . According to Friedland [25], P_a satisfies

$$\begin{aligned} \dot{P}_a &= [F - P_S(H + D)'R^{-1}(H + D)]P_a \\ &\quad + P_a[F - P_S(H + D)'R^{-1}(H + D)]' \\ &\quad + Q + P_S(H + D)'R^{-1}(H + D)P_S, \\ P_a(t_0) &= P(t_0). \end{aligned} \quad (4.42)$$

Now, the degradation of the filter resulting from using P_s instead of P in the gain computation is the difference between the value of the suboptimal cost functional and the optimal cost functional; i.e.

$$\begin{aligned}\Delta J &= J_t^* - \hat{J}_t \\ &= \text{tr}[P_a - P].\end{aligned}\quad (4.43)$$

Denote $P_a - P$ by Γ , then subtracting (4.3) evaluated at $\epsilon = 1$ from (4.42) we obtain

$$\begin{aligned}\dot{\Gamma} &= [F - P(H + D) \cdot R^{-1}(H + D)]\Gamma \\ &\quad + \Gamma[F - P(H + D) \cdot R^{-1}(H + D)]' \\ &\quad + \epsilon^N [\Gamma(H + D) \cdot R^{-1}(H + D)\Lambda \\ &\quad + \Lambda(H + D) \cdot R^{-1}(H + D)\Gamma] \\ &\quad + \epsilon^{2N} \Lambda(H + D) \cdot R^{-1}(H + D)\Lambda,\end{aligned}$$

$$\Gamma(t_0) = 0 \quad (4.44)$$

where

$$\Lambda = \sum_{k=N}^{\infty} \frac{\epsilon^{k-N}}{k!} P^k(t, \epsilon) \Big|_{\epsilon=0}. \quad (4.45)$$

A more convenient form for Λ can be found by taking the

difference between P and P_s to find that

$$P - P_s = \epsilon^N \Lambda. \quad (4.46)$$

By solving (4.44) and computing the trace of the solution we have a direct method of evaluating the performance of the N th-order near-optimal linear filter. Moreover, using (4.44) we can prove a theorem like that of Chapter III by the method of Kokotović and Cruz [26] for the corresponding result for the linear regulator problem.

THEOREM 4.1. For an N th-order near-optimal linear filter,

$$\frac{\partial^k}{\partial \epsilon^k} P(t, \epsilon) \Big|_{\epsilon=0} = \frac{\partial^k}{\partial \epsilon^k} P_a(t, \epsilon) \Big|_{\epsilon=0}, \quad k = 0, 1, 2, \dots, 2N-1 \quad (4.47)$$

Proof. To prove the theorem, it is sufficient to show that the first $2N$ terms in the series expansion of Γ are zero. Setting $\epsilon = 0$ in (4.44) yields a linear homogeneous equation with zero initial condition for Γ^0 . Hence, $\Gamma^0(t) = 0$ for all $t \geq t_0$. Differentiating (4.44) with respect to ϵ and setting $\epsilon = 0$, we find a similar result for Γ^1 . Proceeding inductively, one can show that this is true for terms up to the $2N$ th-order. Beyond the $2N$ th-order, the resulting linear equation is no longer homogeneous. For more details, see the analogous proof of the corresponding theorem for the linear regulator

problem by Kokotović and Cruz [26].

Comments on the interpretation of the theorem of Chapter III are applicable here also.

D. Examples

In this section we present three examples which represent, on a lesser scale, classes of estimation problems for which the ϵ -coupling method might be expected to yield successful results. We point out that the purpose of these simple examples is to demonstrate the use of the algorithms and some of their properties. However, the principal attraction of the method, computational savings, can only be realized on applications to fairly large problems.

EXAMPLE 4.1. Suppose we are given a scalar plant described by

$$\dot{x}_1 = -ax_1 + bx_2 + w_1, \quad a > 0$$

with measurements

$$y = dx_1 + v$$

where w_1 and v are independent gaussian white noise processes and x_2 is an exponentially correlated plant disturbance.

Since x_2 can be modeled by the scalar system

$$\dot{x}_2 = -cx_2 + w_2, \quad c > 0$$

where w_2 is a gaussian white noise process independent of v , we have the equivalent two-dimensional linear system

$$\dot{x} = \begin{bmatrix} -a & b \\ 0 & -c \end{bmatrix} x + w$$

$$y = [d, 0] x + v$$

with

$$\text{cov}\{w(t_1), w(t_2)\} = Q\delta(t_1 - t_2)$$

and

$$\text{cov}\{v(t_1), v(t_2)\} = r\delta(t_1 - t_2).$$

Viewing this system as an interconnection of two scalar subsystems with states x_1 and x_2 respectively, the third-order approximation to P is given by

$$\dot{p}_1^0 = -2ap_1^0 + q_1 - (dp_1^0)^2/r, \quad p_1^0(0) = p_1(0)$$

$$\dot{p}_2^0 = -2cp_2^0 + q_2, \quad p_2^0(0) = p_2(0)$$

$$\dot{p}_{12}^1 = -[a + c + d^2p_1^0/r]p_{12}^1 + bp_2^0 + q_{12},$$

$$p_{12}^1(0) = p_{12}(0)$$

$$\dot{p}_1^2 = -2[a + d^2 p_1^0 / r] p_1^2 + 2b p_{12}^1, \quad p_1^2(0) = 0$$

$$\dot{p}_2^2 = -2[cp_2^2 + (dp_{12}^1)^2 / r], \quad p_2^2(0) = 0$$

$$P_s = \begin{bmatrix} p_1^0 + 0.5p_1^2 & p_{12}^1 \\ p_{12}^1 & p_2^0 + 0.5p_2^2 \end{bmatrix}$$

The equations for P_s , P , and Γ were solved for both the second and third-order approximations using a fourth-order Runge-Kutta algorithm in the CSMP language. For all runs, $a = c = d = r = q_1 = q_2 = p_1(0) = p_2(0) = 1.0$ and $q_{12} = p_{12}(0) = 0$ were used.

In Figures 4.1 - 4.3 the elements of Γ for $b = 1.0$ and for $N = 2$ and $N = 3$ are plotted. In order to infer the meaning of the magnitude of the numbers, the steady-state value of the corresponding element of the P matrix is printed on the plot also. The elements of the P matrix were found to behave exponentially and decayed quickly to their steady-state values.

From these plots we can see that the addition of the third term in the series does not produce a filter which is appreciably less suboptimal. For example, the steady-state value of Γ_1 for $N = 2$ is about 0.79% of the steady-state value of p_1 and the steady-state value of Γ_1 for $N = 3$ is about 0.1% of the value of p_1 . In most cases

this marginal improvement would not warrant the computation of the extra term.

In Figures 4.4 - 4.6 similar plots are given for $b = 0.1$. As expected, by reducing the amount of coupling, the degree of suboptimality is markedly decreased.

EXAMPLE 4.2. In this example we consider a scalar system with measurements which are corrupted by the sum of exponentially correlated noise and gaussian white noise; i.e.,

$$\dot{x}_1 = -ax_1 + w_1, \quad a > 0$$

$$y = cx_1 + dx_2 + v$$

where w_1 and v are independent gaussian white noise and x_2 is exponentially correlated noise which can be modeled by

$$\dot{x}_2 = bx_2 + w_2, \quad b > 0$$

where w_2 is gaussian white noise independent of v .

Augmenting the state, we have the equivalent two-dimensional linear system

$$\dot{\mathbf{x}} = \begin{bmatrix} -a & 0 \\ 0 & -b \end{bmatrix} \mathbf{x} + \mathbf{w}$$

$$y = [c, d]x + v$$

with

$$\text{cov}\{w(t_1), w(t_2)\} = Q\delta(t_1 - t_2)$$

and

$$\text{cov}\{v(t_1), v(t_2)\} = r\delta(t_1 - t_2)$$

If we view this system as an interconnection of two scalar subsystems with states x_1 and x_2 respectively, the third-order approximation to P is given by

$$\dot{p}_1^0 = -2ap_1^0 + q_1 - (cp_1^0)^2/r, \quad p_1^0(0) = p_1(0)$$

$$\dot{p}_2^0 = -2bp_2^0 + q_2, \quad p_2^0(0) = p_2(0)$$

$$\dot{p}_{12}^1 = -(a + b + c^2 p_1^0/r)p_{12}^1 - cdp_1^0 p_2^0/r + q_{12},$$

$$p_{12}^1(0) = p_{12}(0)$$

$$\dot{p}_1^2 = -2(a + c^2 p_1^0/r)p_1^2 - rcdp_1^0 p_{12}^1/r,$$

$$p_1^2(0) = 0$$

$$\dot{p}_2^2 = -2[bp_2^2 + (dp_2^0)^2/r + 2cdp_2^0 p_{12}^1/r + (cp_{12}^1)^2/r],$$

$$p_2^2(0) = 0$$

$$P_S = \begin{bmatrix} p_1^0 + 0.5p_1^2 & p_{12}^1 \\ p_{12}^1 & p_2^0 + 0.5p_2^2 \end{bmatrix}$$

The equations for P_S , P , and Γ were solved as in Example 4.1 for both second and third-order approximations. For all runs, the values $a = b = c = r = q_1 = q_2 = p_1(0) = p_2(0) = 1.0$ and $q_{12} = p_{12}(0) = 0$ were used.

The results are plotted in Figures 4.7 - 4.12 for $d = 1.0$ and $d = 0.1$. From a qualitative point of view, the results were the same as in Example 4.1.

EXAMPLE 4.3. In this example we consider the error propagation in a slow-moving vehicle pure inertial navigation system. We will assume that the vehicle is a slow-moving surface vehicle such as a ship and that the navigation system is operating in a locally level, latitude-longitude coordinate system with x-axis north, y-axis west, and z-axis up. Since we consider only surface vehicles, the vertical channel will not be implemented. We assume also that the "platform errors" can be neglected.

Given these assumptions, we have the following error model for the position and velocity errors (Pitman [27]).

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & | & 0 & 0 \\ -\omega_0^2 & 0 & | & 0 & 2\Omega_z \\ \hline 0 & 0 & | & 0 & 1 \\ 0 & -2\Omega_z & | & -\omega_0^2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ w_{1,2} \\ 0 \\ w_{2,2} \end{bmatrix}$$

where x_1 is the x-channel (north error) and x_2 is the y-channel (west error) and ω_0 is the Schuler frequency and Ω_z is the vertical component of the earth angular rate. The second subscript denotes the component of the subsystem vector. For example, $x_{1,1}$ is the position error in the x-channel, $x_{1,2}$ is the velocity error in the x-channel, and $w_{1,2}$ is the accelerometer error in the x-channel. The fact that some components of the w vector are identically zero can be accounted for by allowing Q to be semi-definite.

If we view the error model as an interconnection of the north error model and the west error model, then

$$F_1 = F_2 = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & 0 \end{bmatrix},$$

and

$$C_{12} = \begin{bmatrix} 0 & 0 \\ 0 & 2\Omega_z \end{bmatrix}, \quad C_{21} = \begin{bmatrix} 0 & 0 \\ 0 & -2\Omega_z \end{bmatrix}.$$

Suppose that we have only position measurements available, then

$$y = \begin{bmatrix} 1 & 0 & | & 0 & 0 \\ \hline & & | & & \\ 0 & 0 & | & 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v$$

and

$$H_1 = H_2 = [1, 0], \quad D_{12} = D_{21} = \emptyset$$

The equations for P , P_s , and Γ for $N = 1, 2$, and 3 were solved using rectangular integration and a step size of 0.2 seconds for a nominal latitude of 45 degrees. The initial P matrix was chosen to be diagonal with

$$p_1(0) = p_3(0) = 10^6 \text{ ft.}^2,$$

$$p_2(0) = p_4(0) = 10^6 \times \omega_0^2 \text{ (ft/sec)}^2.$$

The Q matrix was chosen to be diagonal with $q_1 = q_3 = 0$ and $q_2 = q_4 = 3$ chosen to correspond to an rms position error of one foot after one second. The value of r_1 and r_2 (10^6 ft.^2) was chosen to correspond to an rms position error of approximately 450 ft. for one second between sampling times.

In Figure 4.13 a portion of the results are listed. The values of some of the quantities of interest for comparison were too nearly equal for meaningful plots to

be made. After the first column in which the time is listed, the next six columns list the diagonal elements followed by the trace of P_S for $N = 1$ and 3 , P , and $\Gamma(G)$ for $N = 1, 2$, and 3 . The last two columns list the elements of the off-diagonal block of P_S for $N = 2$ and 3 and for P in the order p_{13} , p_{14} , p_{23} , and p_{24} . This was done to get some idea of the effect of the coupling between the two subsystems, which we can see from Figure 4.13 is fairly small.

As we can see from Figure 4.13, the diagonal terms for the three suboptimal covariance matrices are very close in value to those of the optimal covariance matrix. The fact that these approximations are not very suboptimal is further verified by the magnitude of the elements of Γ for the three cases.

It is apparent that round-off errors have effected the computation of Γ . This is to be expected since the difference of two large, but nearly equal, quantities, Λ , plays a major role in the computation of Γ . If this were a significant problem, and it is not in this example, programming precautions could be taken to minimize its effect.

EXAMPLE 4.4. For our final example we chose to compute the near-optimal filter of Section B and compare it with the Kalman filter for the navigation system in

Example 4.3. Because the near-optimal filter does not have a simple interpretation in terms of a suboptimal gain, simulation of the error model of the navigation system, as in the example of Chapter III, was required. In other words, a covariance analysis was not possible.

For $N = 3$, the near-optimal filter is given by ($i = 1, 2$)

$$\dot{P}_i^0 = F_i P_i^0 + P_i^0 F_i^* + Q_i - P_i^0 H_i^* H_i P_i^0 / r_i, \quad P_i^0(0) = P_i(0)$$

$$\dot{x}_i^0 = F_i x_i^0 + P_i^0 H_i^* (y_i - H_i x_i^0) / r_i, \quad x_i^0(0) = \mu_i$$

$$\dot{P}_i^1 = F_i P_i^1 + P_i^1 F_i^* + Q_i - P_i^1 H_i^* H_i P_i^1 / r_i, \quad P_i^1(0) = \emptyset$$

$$\dot{x}_i^1 = (F_i - P_i^1 H_i^* H_i / r_i) x_i^1 + C_{ij} x_j^0, \quad x_i^1(0) = 0$$

$$\dot{x}_i^2 = (F_i - P_i^1 H_i^* H_i / r_i) x_i^2 + 2C_{ij} x_j^1, \quad x_i^2(0) = 0$$

where $j = 1$ when $i = 2$ and $j = 2$ when $i = 1$.

For the simulations, the mean values chosen for the initial states were

$$\mu_1 = \mu_2 = \begin{bmatrix} 100 \\ 0.125 \end{bmatrix}.$$

In Figure 4.14 is a portion of the results from a typical sample run. For each of the states at a particular time, the sample value is listed followed by the absolute

value of the estimation error obtained from the Kalman filter and the near-optimal filter for $N = 1, 2,$ and $3.$

It is clear that none of the filters tracked the position states very well and the velocity states only slightly better. The important thing to note, in the context of our work here, is that the near-optimal filters did as well as the Kalman filter. This is an indication that the method has some potential, particularly in light of the computational savings which result for large-dimensional problems.

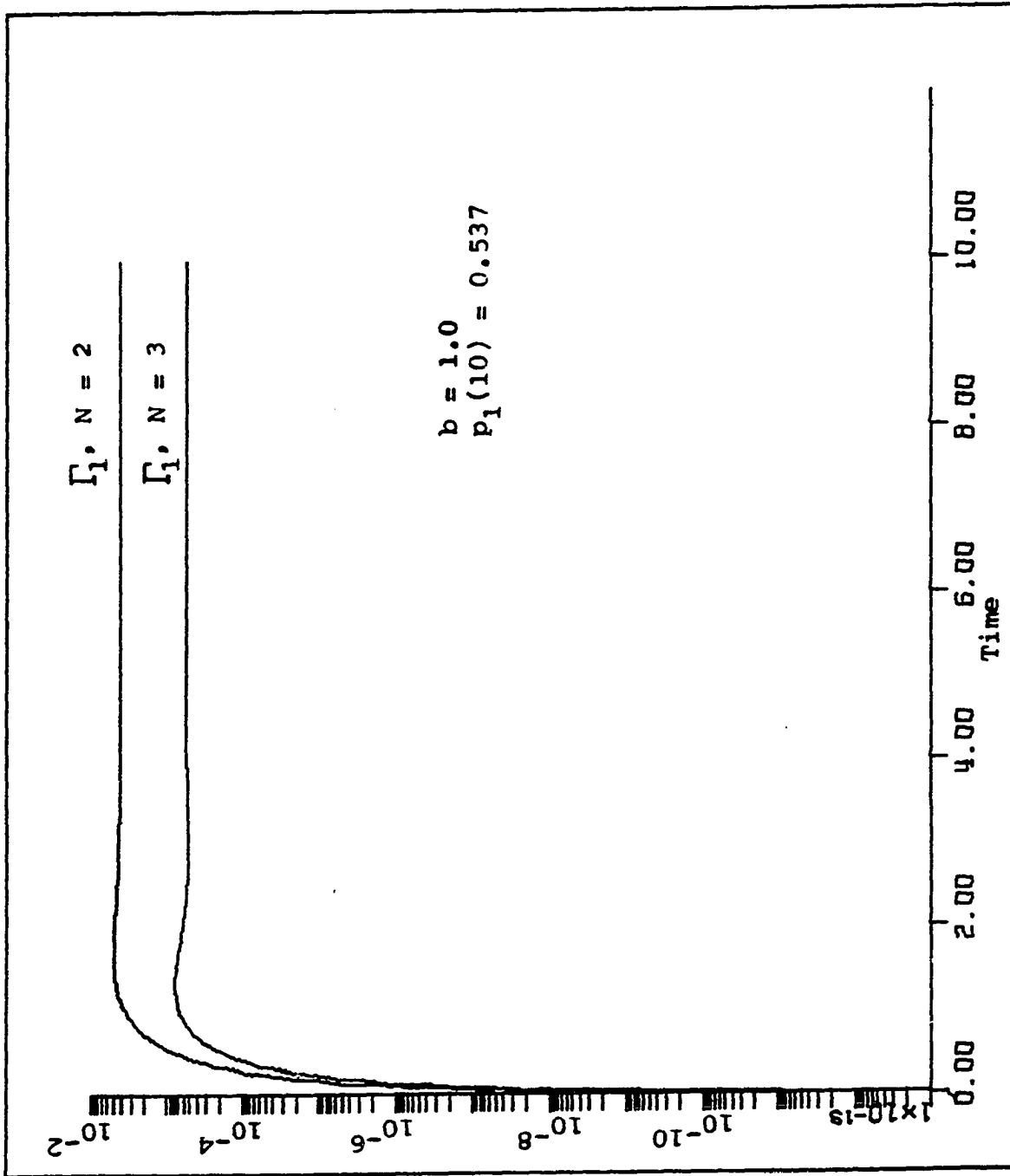


Figure 4.1. Suboptimality of P_g for Example 4.1

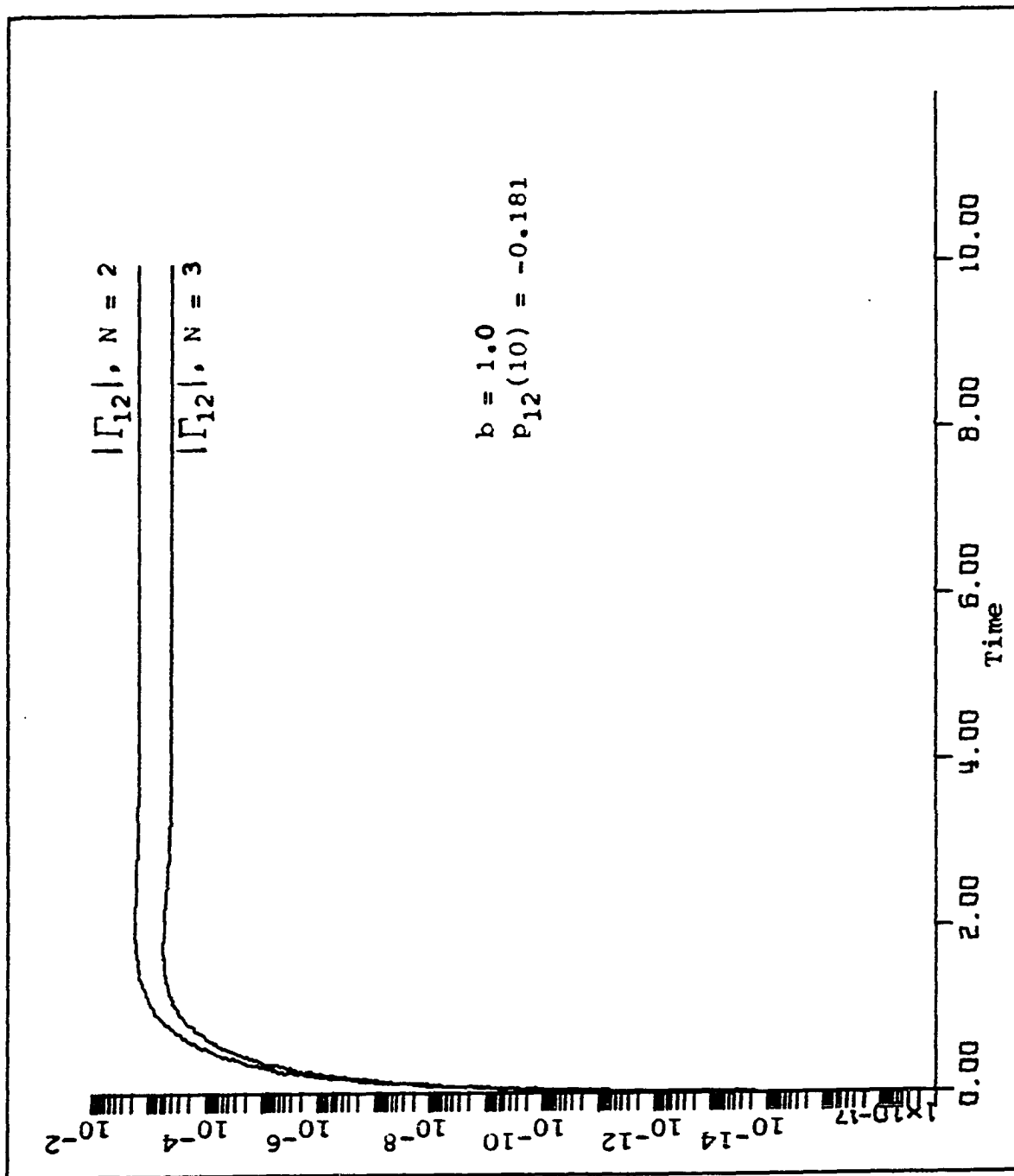


Figure 4.2. Suboptimality of P_g for Example 4.1

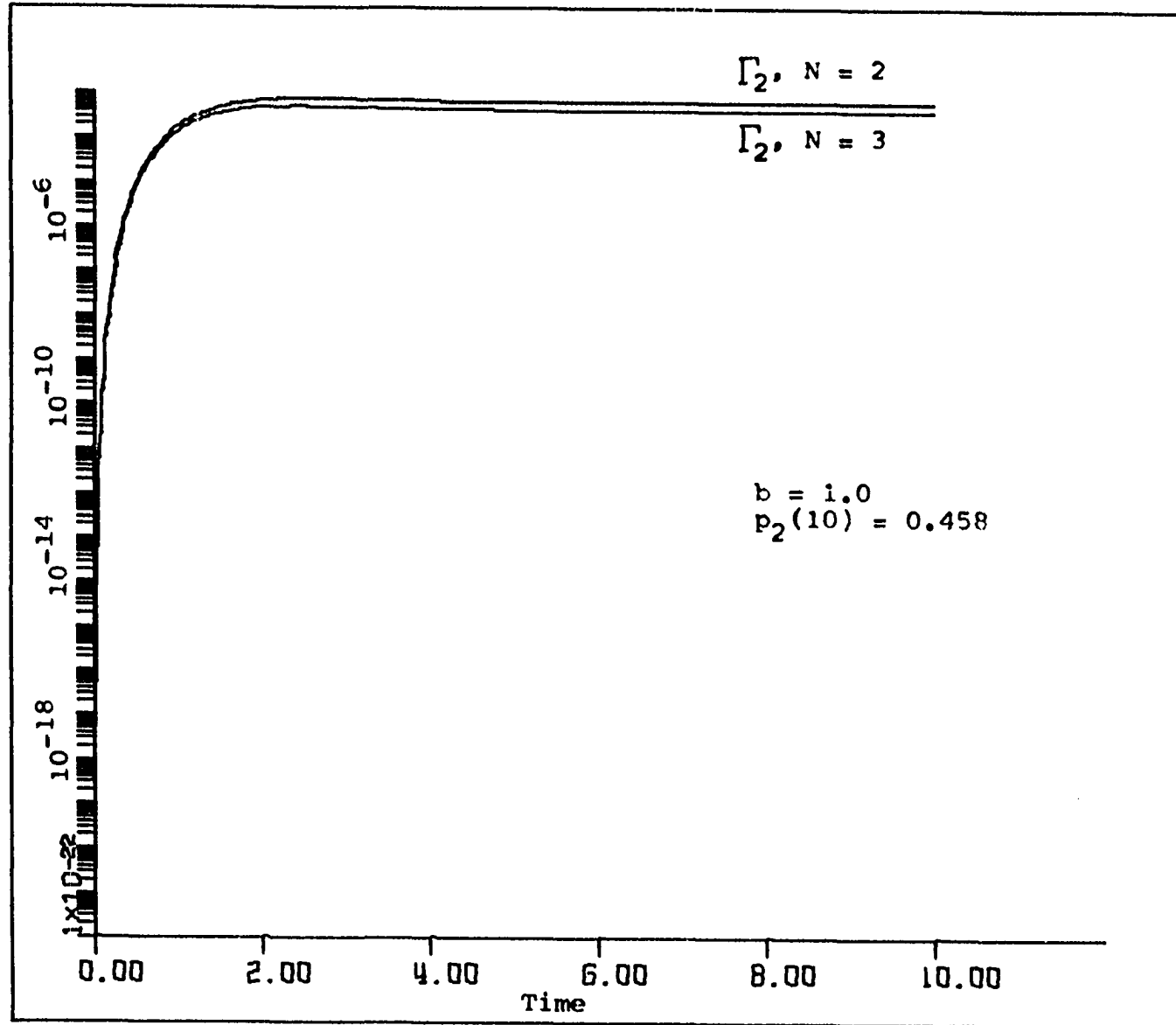


Figure 4.3. Suboptimality of P_g for Example 4.1

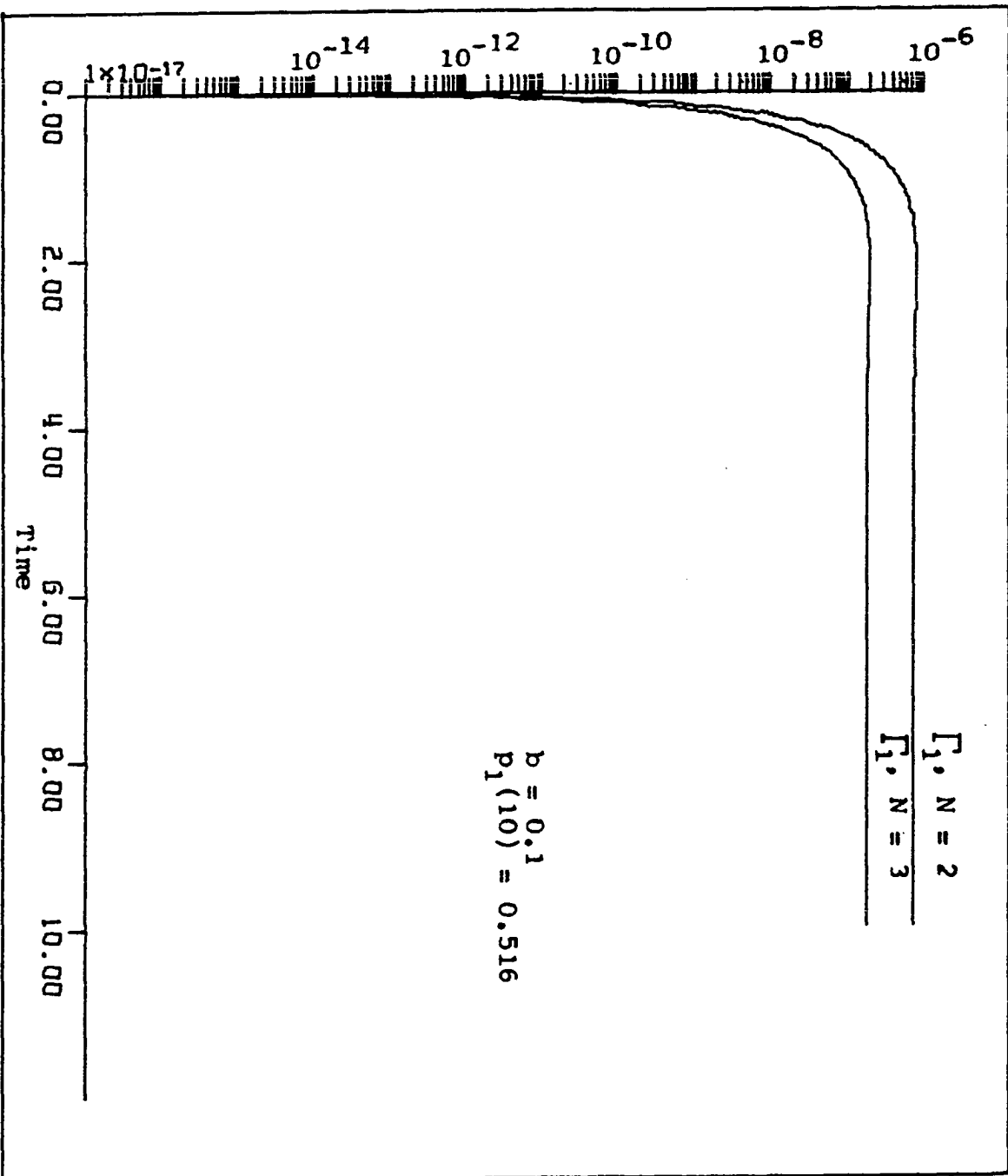


Figure 4.4. Suboptimality of P_s for Example 4.1 with decreased coupling

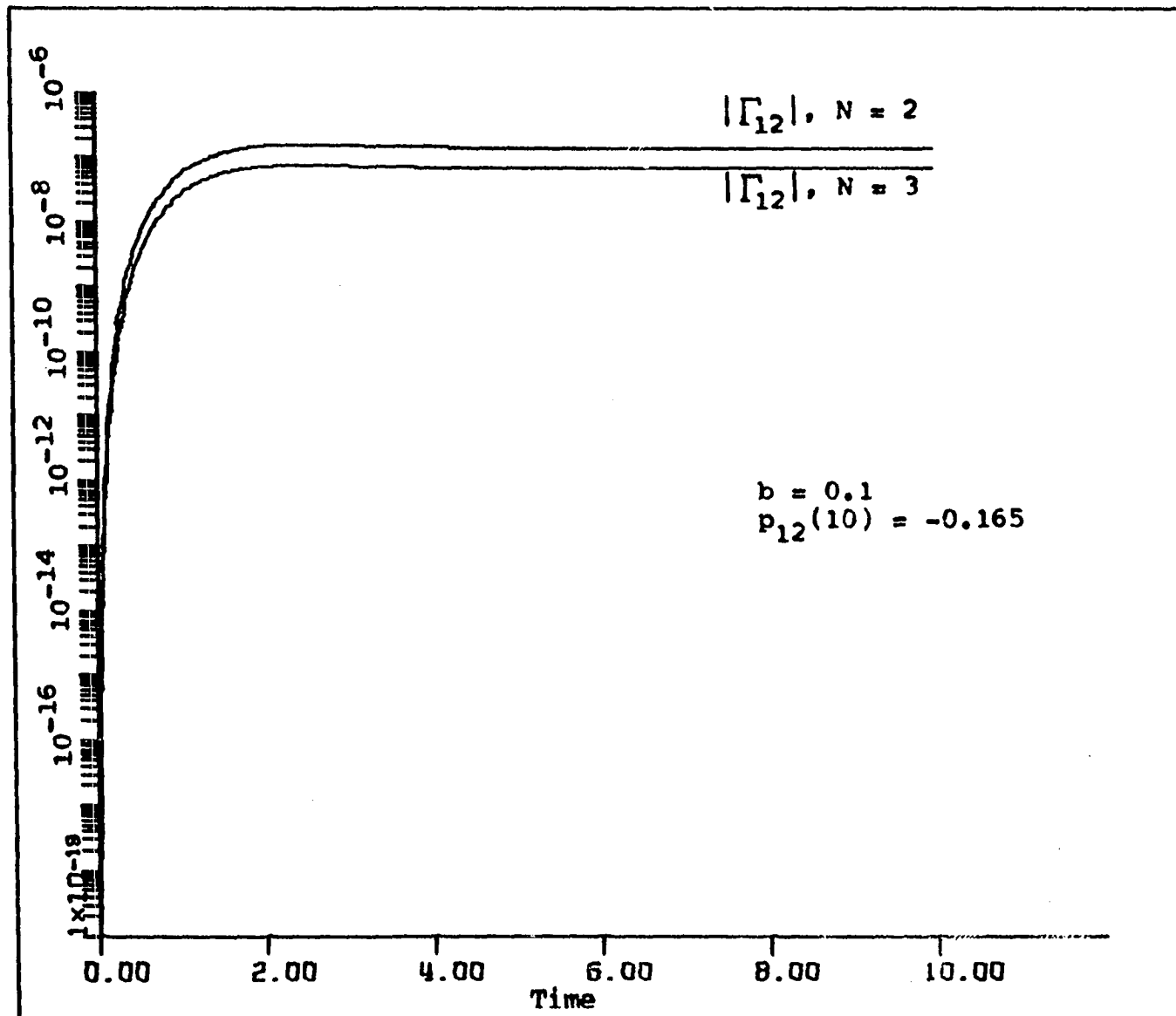


Figure 4.5. Suboptimality of P_g for Example 4.1 with decreased coupling

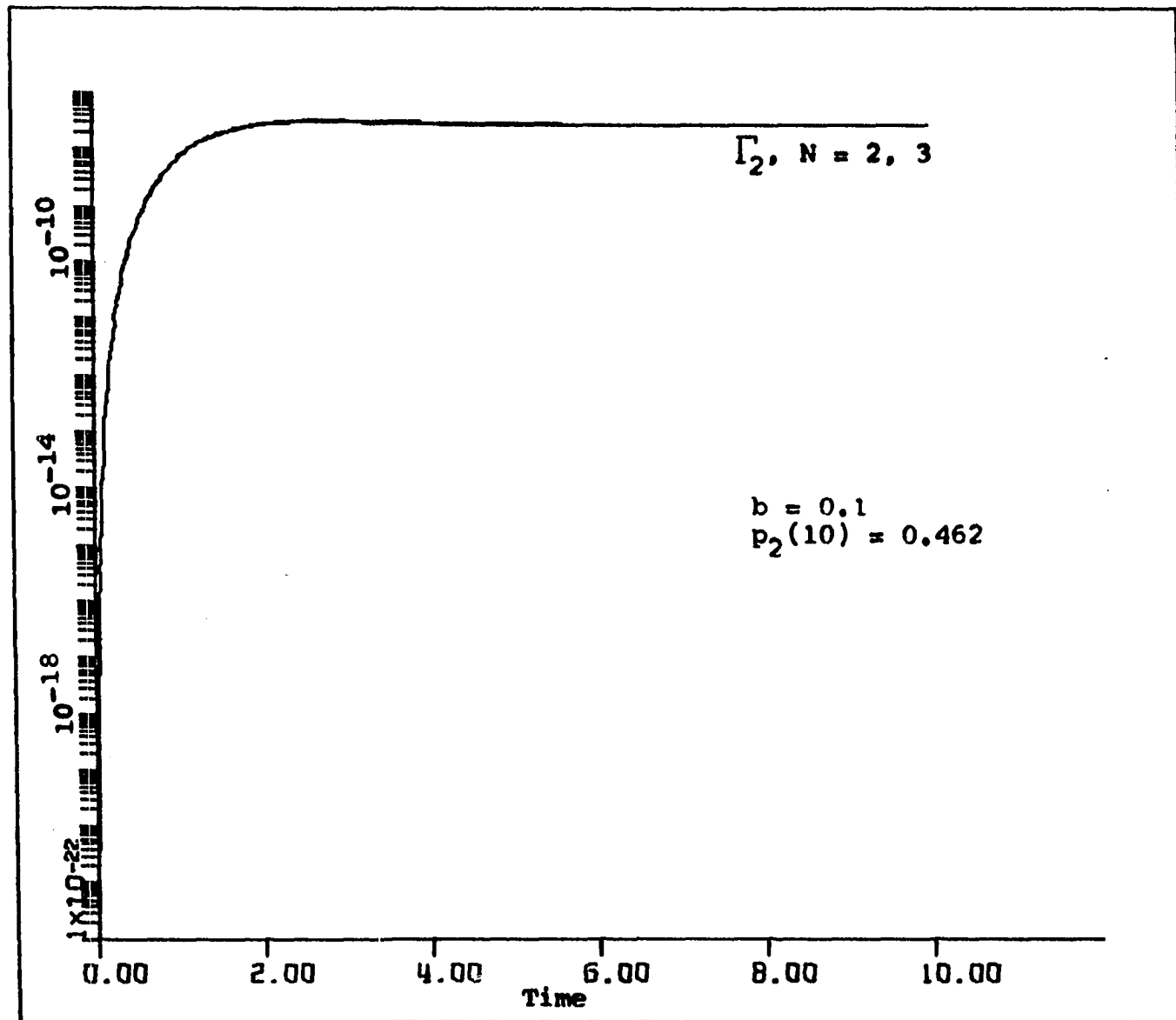


Figure 4.6. Suboptimality of P_g for Example 4.1 with decreased coupling

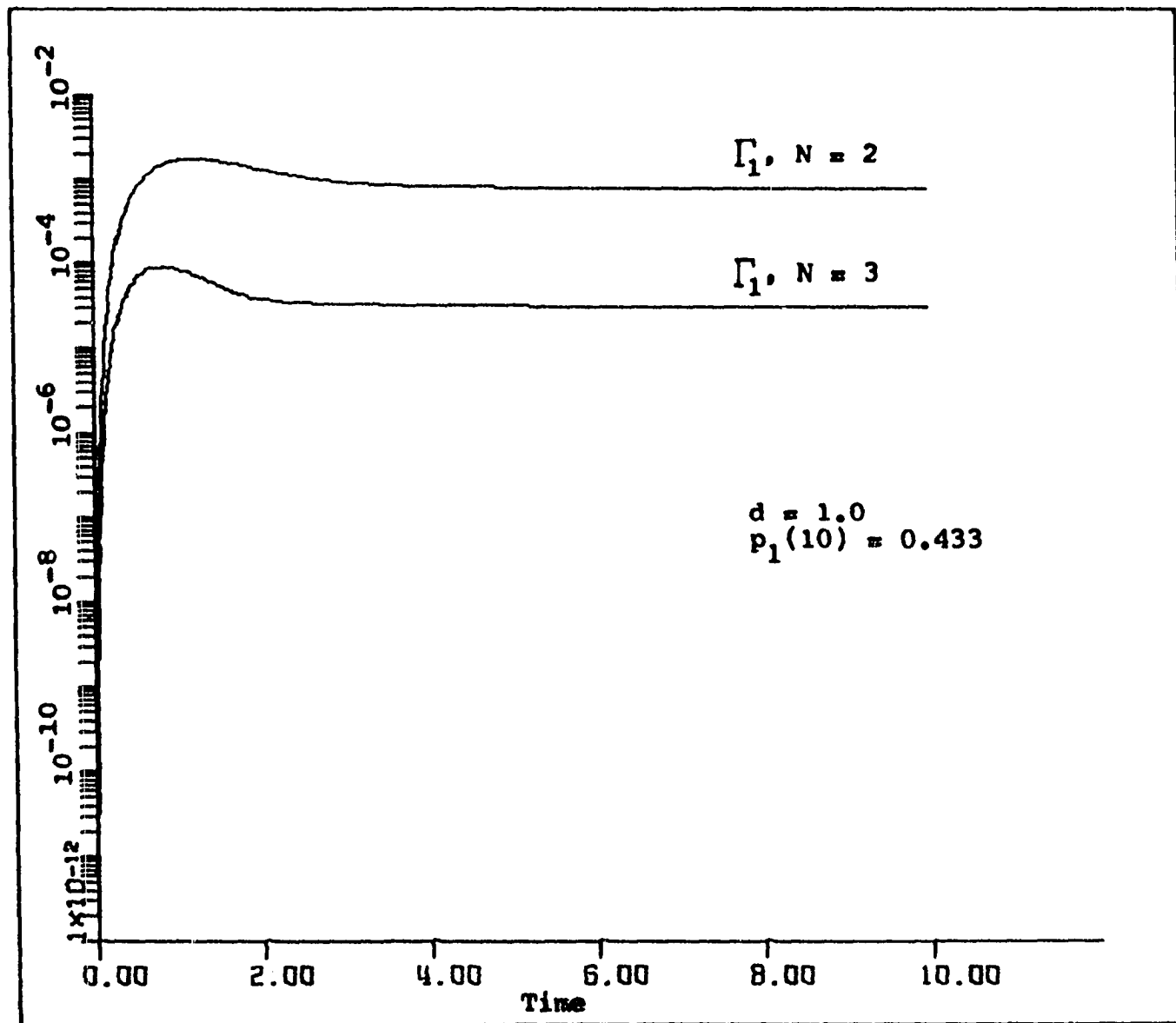


Figure 4.7. Suboptimality of P_g for Example 4.2

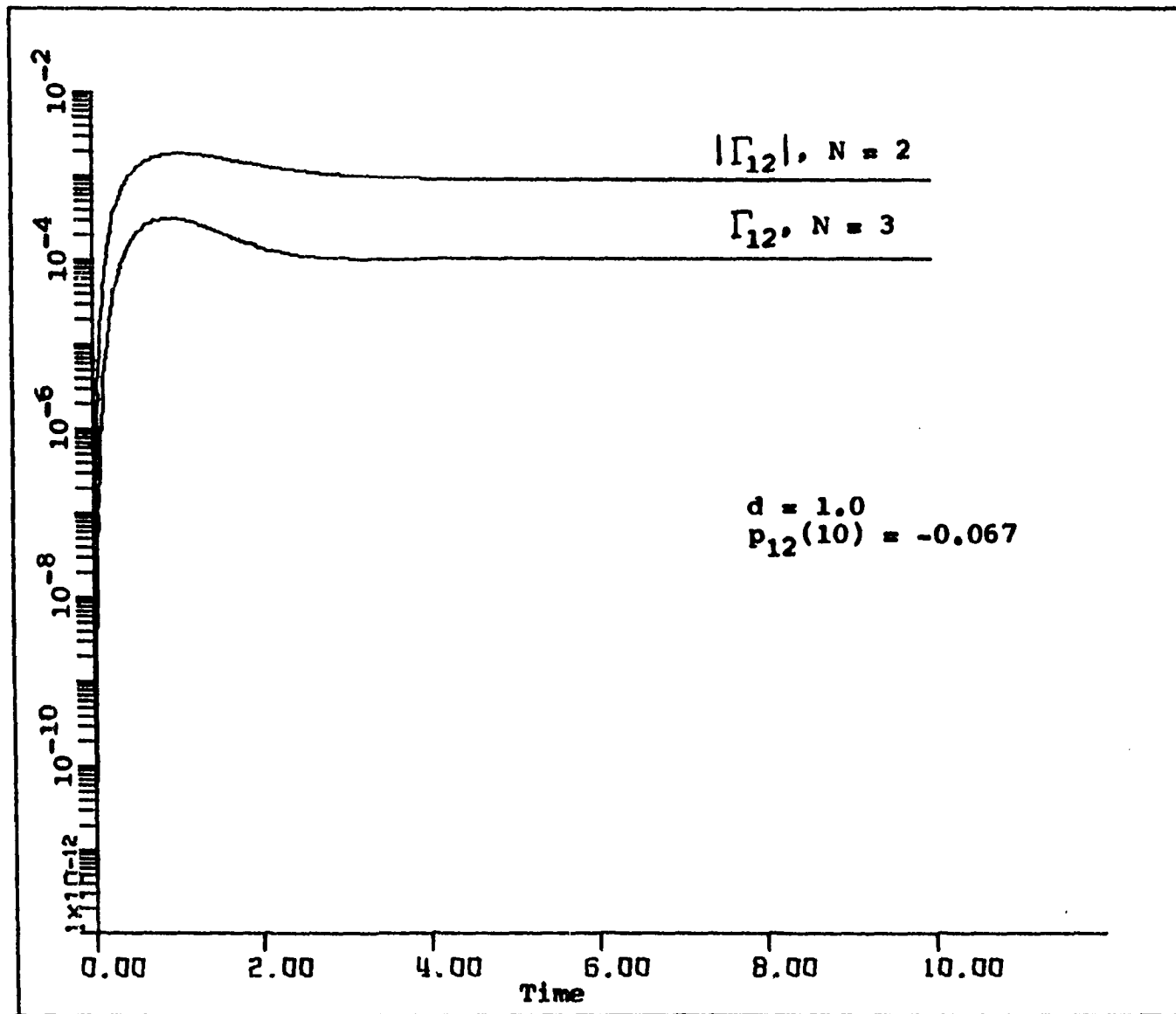


Figure 4.8. Suboptimality of P_B for Example 4.2

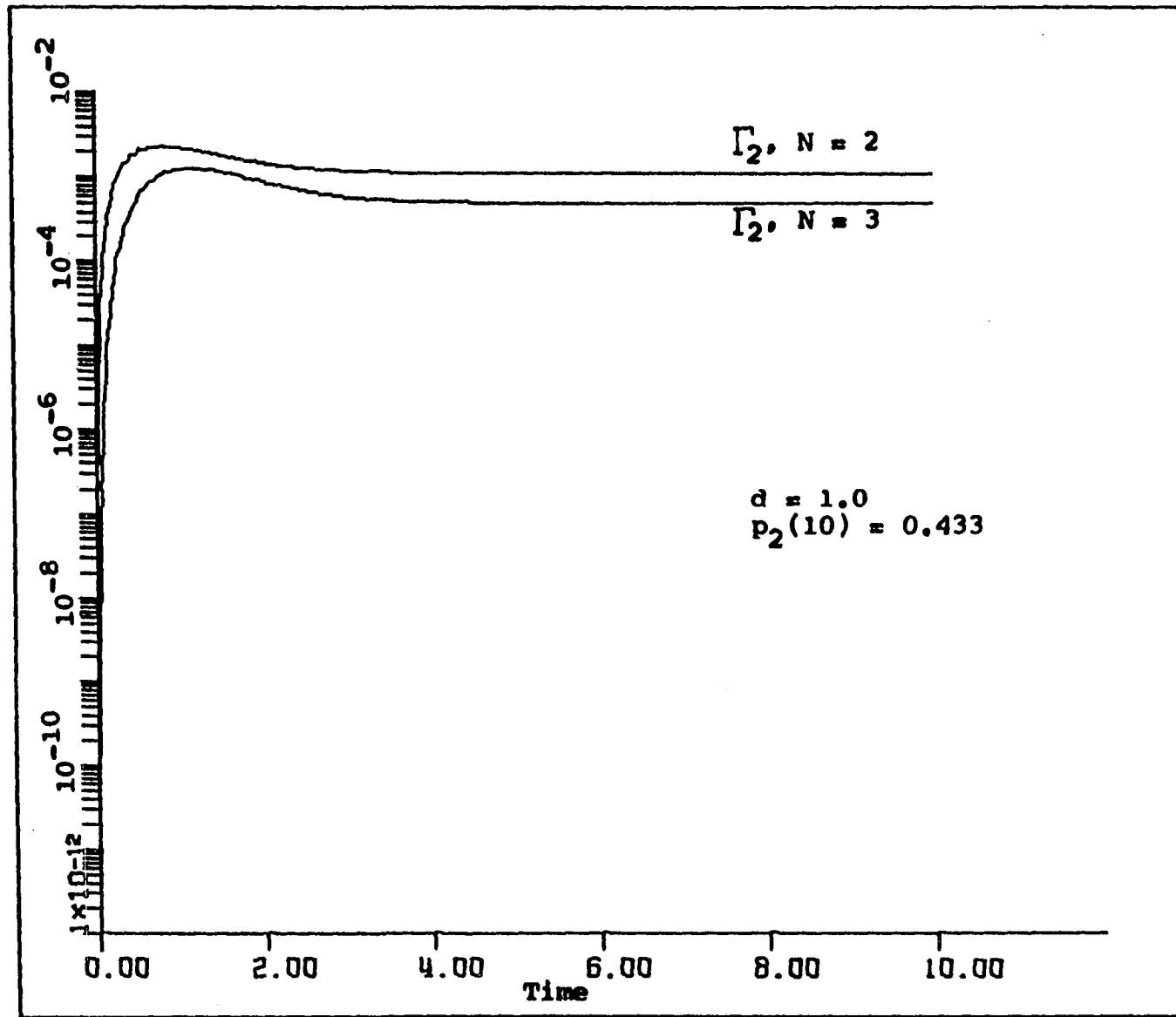


Figure 4.9. Suboptimality of P_g for Example 4.2

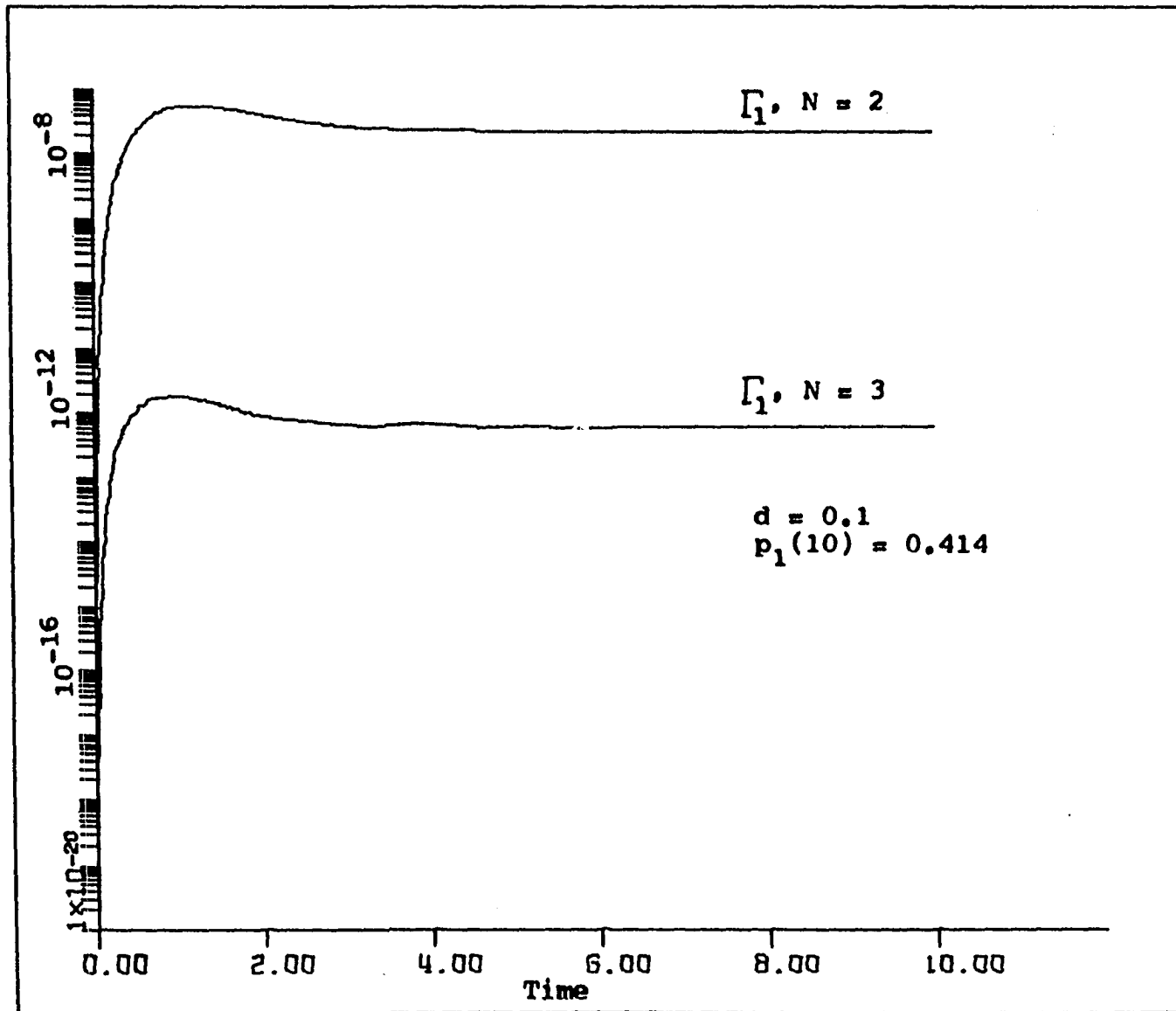


Figure 4.10. Suboptimality of P_g for Example 4.2 with decreased coupling

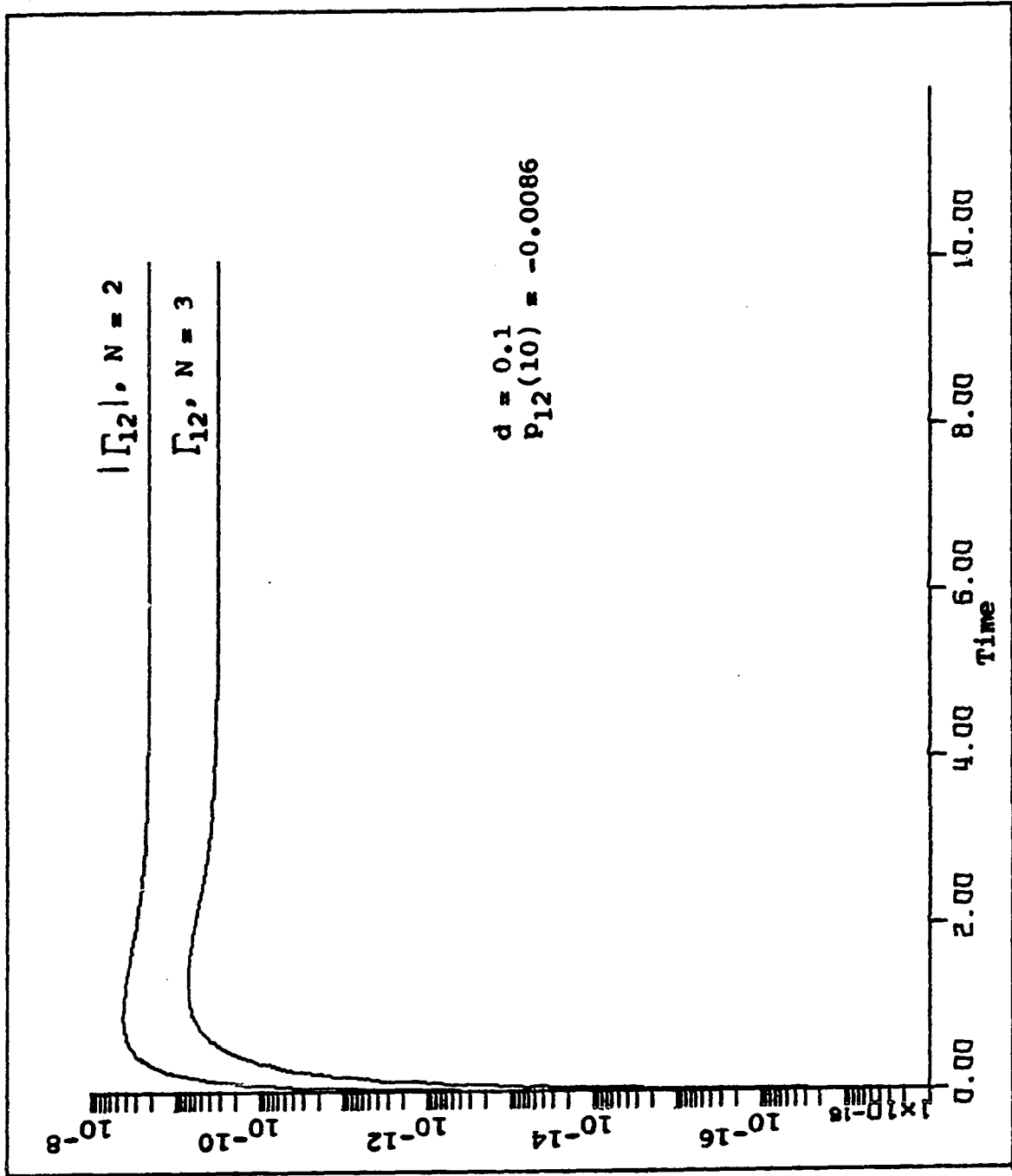


Figure 4.11. Suboptimality of P_s for Example 4.2 with decreased coupling

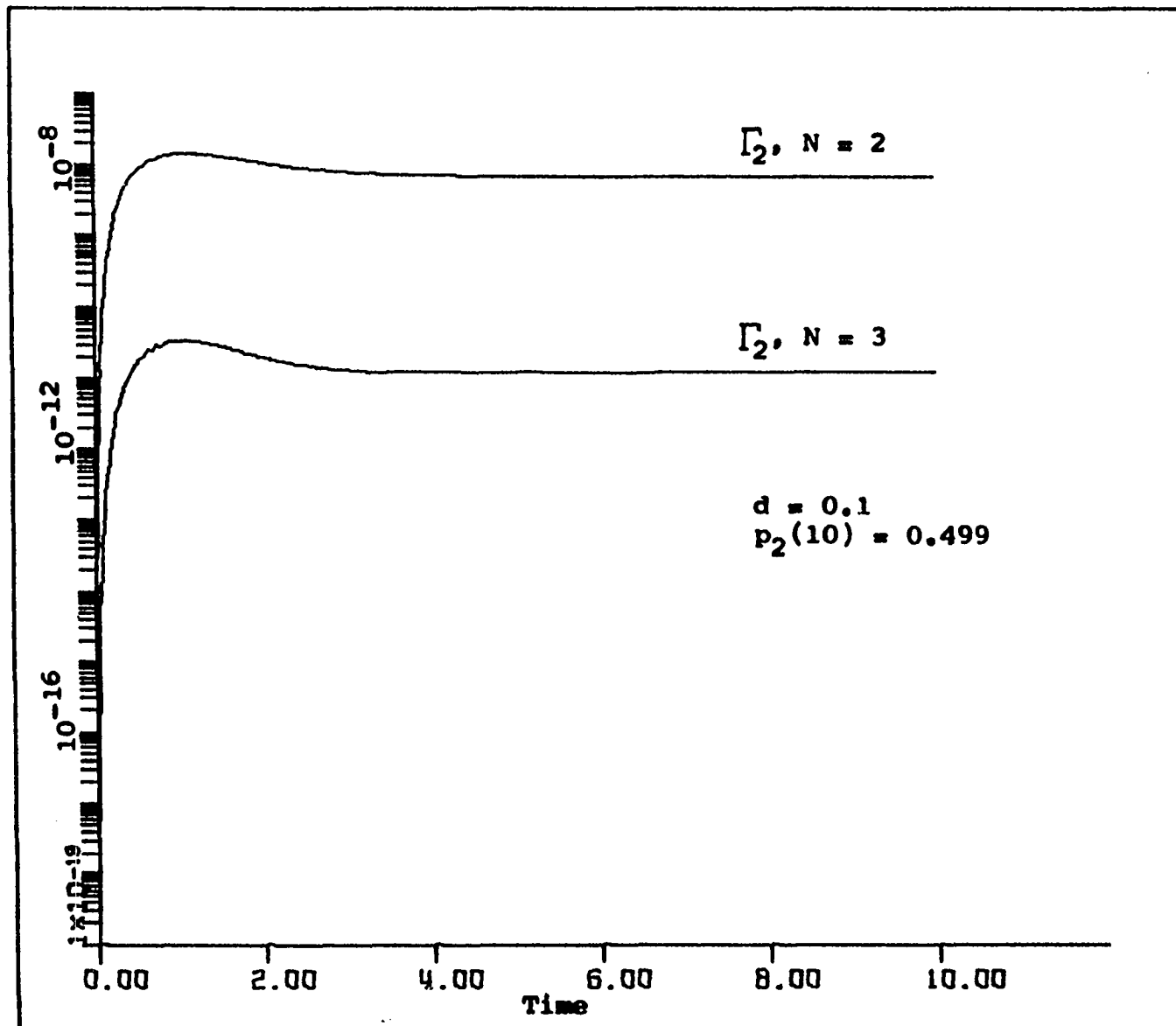


Figure 4.12. Suboptimality of P_8 for Example 4.2 with decreased coupling

Figure 4.13. Filter performance for Example 4.3

DT = 0.010

RI = 0.1000E-05

Q = 0.3000D 01

TIME	DIAG PSO	DIAG PS2	DIAG P	DIAG
1.250	0.444816D 06	0.442838D 06	0.442838D 06	0.62186
	0.526271D 01	0.529271D 01	0.529271D 01	0.16581
	0.444816D 06	0.442838D 06	0.442838D 06	0.62186
	0.526271D 01	0.529271D 01	0.529271D 01	0.16581
TRACE	0.889643D 06	0.885686D 06	0.885686D 06	0.34407
2.500	0.285515D 06	0.284700D 06	0.284700D 06	0.22885
	0.901266D 01	0.904265D 01	0.904265D 01	0.15234
	0.285515D 06	0.284700D 06	0.284700D 06	0.22885
	0.901266D 01	0.904265D 01	0.904265D 01	0.15234
TRACE	0.571048D 06	0.569418D 06	0.569418D 06	0.35045
3.750	0.210308D 06	0.209866D 06	0.209866D 06	0.70905
	0.127624D 02	0.127924D 02	0.127924D 02	0.21349
	0.210308D 06	0.209866D 06	0.209866D 06	0.70905
	0.127624D 02	0.127924D 02	0.127924D 02	0.21349
TRACE	0.420641D 06	0.419757D 06	0.419757D 06	0.56879
5.000	0.166514D 06	0.166237D 06	0.166237D 06	0.81473
	0.165116D 02	0.165416D 02	0.165416D 02	0.14034
	0.166514D 06	0.166237D 06	0.166237D 06	0.81473
	0.165116D 02	0.165416D 02	0.165416D 02	0.14034
TRACE	0.333060D 06	0.332507D 06	0.332507D 06	0.44363
6.250	0.137870D 06	0.137681D 06	0.137681D 06	0.54537
	0.202596D 02	0.202896D 02	0.202896D 02	0.60996
	0.137870D 06	0.137681D 06	0.137681D 06	0.54537
	0.202596D 02	0.202896D 02	0.202896D 02	0.60996
TRACE	0.275781D 06	0.275403D 06	0.275403D 06	0.23106
7.500	0.117700D 06	0.117563D 06	0.117563D 06	0.259425
	0.240056D 02	0.240356D 02	0.240356D 02	0.203931
	0.117700D 06	0.117563D 06	0.117563D 06	0.259425
	0.240056D 02	0.240356D 02	0.240356D 02	0.203931
TRACE	0.235448D 06	0.235173D 06	0.235173D 06	0.926714
8.750	0.102756D 06	0.102652D 06	0.102652D 06	0.974415
	0.277483D 02	0.277782D 02	0.277782D 02	0.568464
	0.102756D 06	0.102652D 06	0.102652D 06	0.974415
	0.277483D 02	0.277782D 02	0.277782D 02	0.568464
TRACE	0.205567D 06	0.205359D 06	0.205359D 06	0.308576
10.000	0.912699D 05	0.911888D 05	0.911888D 05	0.307714
	0.314857D 02	0.315156D 02	0.315156D 02	0.138624
	0.912699D 05	0.911888D 05	0.911888D 05	0.307714
	0.314857D 02	0.315156D 02	0.315156D 02	0.138624
TRACE	0.182603D 06	0.182441D 06	0.182441D 06	0.892675

RI = 0.1000E-05

Q = 0.3000D 01

PSO	DIAG PS2	DIAG P	DIAG GO	DIAG G1	DIAG G2	PS
6D 06	0.442838D 06	0.442838D 06	0.621862D-16	0.693312D-24	0.107416D-24	0.11
1D 01	0.529271D 01	0.529271D 01	0.165818D-14	0.112217D-23	0.150356D-27	-0.92
6D 06	0.442838D 06	0.442838D 06	0.621862D-16	0.693312D-24	0.107416D-24	0.92
1D 01	0.529271D 01	0.529271D 01	0.165818D-14	0.112216D-23	0.149241D-27	0.18
6D 06	0.885686D 06	0.885686D 06	0.344073E-14	0.363096D-23	0.215132D-24	
5D 06	0.284700D 06	0.284700D 06	0.228851D-13	0.371872D-21	0.966700D-24	0.12
6D 01	0.904265D 01	0.904265D 01	0.152340D-12	0.397052D-21	0.928116D-24	-0.62
5D 06	0.284700D 06	0.284700D 06	0.228851D-13	0.371871D-21	0.965983D-24	0.62
6D 01	0.904265D 01	0.904265D 01	0.152340D-12	0.397046D-21	0.924858D-24	0.59
8D 06	0.569418D 06	0.569418D 06	0.350451E-12	0.153764D-20	0.378566D-23	
8D 06	0.209866D 06	0.209866D 06	0.709055D-12	0.231230D-19	0.207374D-21	0.22
4D 02	0.127924D 02	0.127924D 02	0.213490D-11	0.122655D-19	0.203685D-21	-0.19
8D 06	0.209866D 06	0.209866D 06	0.709055D-12	0.231227D-19	0.207246D-21	0.19
4D 02	0.127924D 02	0.127924D 02	0.213490D-11	0.122647D-19	0.203291D-21	0.51
1D 06	0.419757D 06	0.419757D 06	0.568790E-11	0.707758D-19	0.821597D-21	
4D 06	0.166237D 06	0.166237D 06	0.814730D-11	0.460866D-18	0.177573D-19	0.18
6D 02	0.165416D 02	0.165416D 02	0.140342D-10	0.147663D-18	0.103789D-19	-0.42
6D 06	0.166237D 06	0.166237D 06	0.814730D-11	0.460851D-18	0.177505D-19	0.42
6D 02	0.165416D 02	0.165416D 02	0.140342D-10	0.147638D-18	0.103664D-19	0.24
0D 06	0.332507D 06	0.332507D 06	0.443631E-10	0.121702D-17	0.562532D-19	
0D 06	0.137681D 06	0.137681D 06	0.545373D-10	0.507635D-17	0.596125D-18	0.96
6D 02	0.202896D 02	0.202896D 02	0.609964D-10	0.114368D-17	0.230150D-18	-0.80
0D 06	0.137681D 06	0.137681D 06	0.545373D-10	0.507601D-17	0.595971D-18	0.80
6D 02	0.202896D 02	0.202896D 02	0.609964D-10	0.114330D-17	0.229962D-18	0.84
1D 06	0.275403D 06	0.275403D 06	0.231067E-09	0.124393D-16	0.165221D-17	
0D 06	0.117563D 06	0.117563D 06	0.259425D-09	0.408119D-16	0.108570D-16	0.37
6D 02	0.240356D 02	0.240356D 02	0.203931D-09	0.730420D-17	0.297627D-17	-0.13
0D 06	0.117563D 06	0.117563D 06	0.259425D-09	0.408075D-16	0.108550D-16	0.13
6D 02	0.240356D 02	0.240356D 02	0.203931D-09	0.730070D-17	0.297451D-17	0.23
8D 06	0.235173D 06	0.235173D 06	0.926714E-09	0.962243D-16	0.276628D-16	
6D 06	0.102652D 06	0.102652D 06	0.974415D-09	0.279043D-15	0.128720D-15	0.12
3D 02	0.277782D 02	0.277782D 02	0.568464D-09	0.425682D-16	0.263605D-16	-0.20
6D 06	0.102652D 06	0.102652D 06	0.974415D-09	0.279003D-15	0.128702D-15	0.20
3D 02	0.277782D 02	0.277782D 02	0.568464D-09	0.425448D-16	0.263487D-16	0.50
7D 06	0.205359D 06	0.205359D 06	0.308576E-08	0.643160D-15	0.310132D-15	
9D 05	0.911888D 05	0.911888D 05	0.307714D-08	0.172096D-14	0.111012D-14	0.3
7D 02	0.315156D 02	0.315156D 02	0.138624D-08	0.227411D-15	0.176346D-15	-0.3
9D 05	0.911888D 05	0.911888D 05	0.307714D-08	0.172069D-14	0.111000D-14	0.3
7D 02	0.315156D 02	0.315156D 02	0.138624D-08	0.227289D-15	0.176285D-15	0.1
3D 06	0.182441D 06	0.182441D 06	0.892675E-08	0.389634D-14	0.257275D-14	

Q = 0.3000 01

DIAG P	DIAG GO	DIAG G1	DIAG G2	PS1 OFF D	P OFF D
0.442838D 06	0.621862D-16	0.693312D-24	0.107416D-24	0.119794D-10	0.185942D-
0.529271D 01	0.165818D-14	0.112217D-23	0.150356D-27	-.922985D-04	-.943650D-
0.442838D 06	0.621862D-16	0.693312D-24	0.107416D-24	0.922986D-04	0.943650D-
0.529271D 01	0.165818D-14	0.112216D-23	0.149241D-27	0.188168D-09	0.512624D-
0.885686D 06	0.344073E-14	0.363096D-23	0.215132D-24		
0.284700D 06	0.228851D-13	0.371872D-21	0.966700D-24	0.128089D-08	0.992916D-
0.904265D 01	0.152340D-12	0.397052D-21	0.928116D-24	-.625981D-03	-.632912D-
0.284700D 06	0.228851D-13	0.371871D-21	0.965983D-24	0.625986D-03	0.632912D-
0.904265D 01	0.152340D-12	0.397046D-21	0.924858D-24	0.596660D-08	0.110893D-
0.569418D 06	0.350451E-12	0.153784D-20	0.378566D-23		
0.209866D 06	0.709055D-12	0.231230D-19	0.207374D-21	0.227580D-07	0.401042D-
0.127924D 02	0.213490D-11	0.122655D-19	0.203685D-21	-.192058D-02	-.193486D-
0.209866D 06	0.709055D-12	0.231227D-19	0.207246D-21	0.192064D-02	0.193486D-
0.127924D 02	0.213490D-11	0.122647D-19	0.203291D-21	0.511716D-07	0.174855D-
0.419757D 06	0.568790E-11	0.707758D-19	0.821597D-21		
0.166237D 06	0.814730D-11	0.460866D-18	0.177573D-19	0.185301D-06	0.130659D-
0.165416D 02	0.140342D-10	0.147663D-18	0.103789D-19	-.428099D-02	-.430516D-
0.166237D 06	0.814730D-11	0.460851D-18	0.177505D-19	0.428133D-02	0.430516D-
0.165416D 02	0.140342D-10	0.147638D-18	0.103664D-19	0.245568D-06	0.238782D-
0.332507D 06	0.443631E-10	0.121702D-17	0.562532D-19		
0.137681D 06	0.545373D-10	0.507635D-17	0.596125D-18	0.967764D-06	0.214122D-
0.202896D 02	0.609964D-10	0.114368D-17	0.230150D-18	-.800947D-02	-.804634D-
0.137681D 06	0.545373D-10	0.507601D-17	0.595971D-18	0.801089D-02	0.804634D-
0.202896D 02	0.609964D-10	0.114330D-17	0.229962D-18	0.846012D-06	0.304308D-
0.275403D 06	0.231067E-09	0.124393D-16	0.165221D-17		
0.117563D 06	0.259425D-09	0.408119D-16	0.108570D-16	0.379053D-05	0.333461D-
0.240356D 02	0.203931D-09	0.730420D-17	0.297627D-17	-.134064D-01	-.134595D-
0.117563D 06	0.259425D-09	0.408075D-16	0.108550D-16	0.134111D-01	0.134595D-
0.240356D 02	0.203931D-09	0.730070D-17	0.297451D-17	0.235094D-05	0.367064D-
0.235173D 06	0.926714E-09	0.962243D-16	0.276628D-16		
0.102652D 06	0.974415D-09	0.279043D-15	0.128720D-15	0.121318D-04	0.597706D-
0.277782D 02	0.568464D-09	0.425682D-16	0.263605D-16	-.207696D-01	-.208438D-
0.102652D 06	0.974415D-09	0.279003D-15	0.128702D-15	0.207826D-01	0.208438D-
0.277782D 02	0.568464D-09	0.425448D-16	0.263487D-16	0.561770D-05	0.430105D-
0.205359D 06	0.308576E-08	0.643160D-15	0.310132D-15		
0.911888D 05	0.307714D-08	0.172096D-14	0.111012D-14	0.334290D-04	0.101675D-
0.315156D 02	0.138624D-08	0.227411D-15	0.176346D-15	-.303928D-01	-.304956D-
0.911888D 05	0.307714D-08	0.172069D-14	0.111000D-14	0.304244D-01	0.304956D-
0.315156D 02	0.138624D-08	0.227289D-15	0.176285D-15	0.120020D-04	0.493907D-
0.182441D 06	0.892675E-08	0.389634D-14	0.257275D-14		

Figure 4.14. Filter performance for Example 4.4

	STATE 1	STATE 2	STATE 3	STATE 4
TIME = 0.0				
SAMPLE	-0.1955E 04	0.1118E 01	-0.7547E 03	-0.1379E 01
TIME = 1.250				
SAMPLE	-0.1955E 04	0.9449E 00	-0.7568E 03	-0.2531E 01
ERRK	0.8622E 03	0.8187E 00	0.5155E 02	0.2653E 01
ERRCO	0.8623E 03	0.8187E 00	0.5154E 02	0.2653E 01
ERRC1	0.8623E 03	0.8187E 00	0.5154E 02	0.2653E 01
ERRC2	0.8623E 03	0.8187E 00	0.5154E 02	0.2653E 01
TIME = 2.500				
SAMPLE	-0.1955E 04	-0.2159E 01	-0.7587E 03	-0.1522E 01
ERRK	0.9839E 03	0.2287E 01	0.4382E 03	0.1638E 01
ERRCO	0.9839E 03	0.2287E 01	0.4382E 03	0.1638E 01
ERRC1	0.9839E 03	0.2287E 01	0.4382E 03	0.1638E 01
ERRC2	0.9839E 03	0.2287E 01	0.4382E 03	0.1638E 01
TIME = 3.750				
SAMPLE	-0.1959E 04	-0.3617E 01	-0.7613E 03	-0.1479E 01
ERRK	0.4123E 03	0.3723E 01	0.5585E 03	0.1592E 01
ERRCO	0.4123E 03	0.3723E 01	0.5585E 03	0.1592E 01
ERRC1	0.4123E 03	0.3723E 01	0.5585E 03	0.1592E 01
ERRC2	0.4123E 03	0.3723E 01	0.5585E 03	0.1592E 01
TIME = 5.000				
SAMPLE	-0.1962E 04	-0.1202E 01	-0.7614E 03	0.1628E 01
ERRK	0.3455E 03	0.1298E 01	0.5805E 03	0.1514E 01
ERRCO	0.3455E 03	0.1298E 01	0.5805E 03	0.1514E 01
ERRC1	0.3455E 03	0.1298E 01	0.5805E 03	0.1514E 01
ERRC2	0.3455E 03	0.1298E 01	0.5805E 03	0.1514E 01
TIME = 6.250				
SAMPLE	-0.1965E 04	-0.2649E 01	-0.7576E 03	0.4526E 01
ERRK	0.4183E 03	0.2769E 01	0.6121E 03	0.4415E 01
ERRCO	0.4183E 03	0.2769E 01	0.6121E 03	0.4415E 01
ERRC1	0.4183E 03	0.2769E 01	0.6121E 03	0.4415E 01
ERRC2	0.4183E 03	0.2769E 01	0.6121E 03	0.4415E 01
TIME = 7.500				
SAMPLE	-0.1967E 04	-0.1965E 01	-0.7516E 03	0.5530E 01
ERRK	0.7099E 03	0.2206E 01	0.7508E 03	0.5468E 01
ERRCO	0.7099E 03	0.2206E 01	0.7508E 03	0.5468E 01
ERRC1	0.7099E 03	0.2206E 01	0.7508E 03	0.5468E 01
ERRC2	0.7099E 03	0.2206E 01	0.7508E 03	0.5468E 01
TIME = 8.749				
SAMPLE	-0.1969E 04	-0.8146E 00	-0.7447E 03	0.5701E 01
ERRK	0.5977E 03	0.9799E 00	0.8384E 03	0.5691E 01
ERRCO	0.5977E 03	0.9799E 00	0.8384E 03	0.5691E 01
ERRC1	0.5977E 03	0.9799E 00	0.8384E 03	0.5691E 01
ERRC2	0.5977E 03	0.9799E 00	0.8384E 03	0.5691E 01
TIME = 9.999				
SAMPLE	-0.1970E 04	-0.1047E 01	-0.7371E 03	0.4986E 01
ERRK	0.5509E 03	0.1134E 01	0.8521E 03	0.4976E 01
ERRCO	0.5509E 03	0.1134E 01	0.8521E 03	0.4976E 01
ERRC1	0.5509E 03	0.1134E 01	0.8521E 03	0.4976E 01
ERRC2	0.5509E 03	0.1134E 01	0.8521E 03	0.4976E 01

V. CONCLUSION

In the preceding chapters we have proposed new approximate nonlinear filtering and smoothing algorithms based upon a synthesis of ideas in optimal control theory, large-scale systems analysis, and Bayesian estimation theory.

The principal advantage of the method is that for large-scale systems, significant computational savings can be obtained over the computational requirements of other approximate algorithms. From the preliminary computational studies made for the examples presented, we might expect that the method will perform best when the composite system has light coupling between the subsystems and/or when the nonlinearities in the system are small.

Another advantage of the method is that the smoothing solution is a natural by-product in the derivation. The nonlinear smoothing problem is a difficult one and most previous methods resorted to merely linearizing the system and applying the linear smoothing algorithm. The smoothing problem in general requires a great deal of computation, hence the decoupling property of our method makes it particularly attractive.

Other attractions of the method are that only linear

and Riccati type equations need be solved and that a qualitative estimate can be obtained for the performance of the algorithm.

As in the nonlinear problem, the computational savings can be significant for the linear case when dealing with large-dimensional systems. The approach of Haddad and Cruz [14] is attractive because many systems may be naturally viewed as consisting of two interconnected subsystems, the results have a simple interpretation in terms of a suboptimal gain, and a quantitative estimate of the suboptimality of the algorithm can be obtained.

The modal-trajectory approach to the linear problem, on the other hand, enables one to break the problem apart into more than two subsystems. This results in perhaps additional computational savings as well as a decoupled solution to the smoothing problem.

On the debit side for the ϵ -coupling approach, the algorithm has no advantage over the Kalman filter for small linear problems. For small nonlinear problems, the only justification for its use is that if a sufficient number of terms in the series are computed, better performance can be obtained than that achieved with other methods. It is expected that usually only two or three terms will be needed.

The filter is tedious to design. Somewhat more hand

calculation is needed than in, say, the extended Kalman filter to specify the equations in the algorithm for a particular application.

Perhaps a more serious drawback is that no intuitive interpretation can be made in terms of a suboptimal gain or error covariance matrix. This perhaps hinders the evaluation of the performance of the filter. However, to some degree this is a problem common to the other approximate methods as well.

Several investigations appear promising and interesting related to this study. From a practical viewpoint, perhaps the first requirement is a full computational study of properties of these algorithms on a large-scale system problem. This should be done coincident with a computational study of the other approximate nonlinear algorithms for comparative purposes.

Clearly, a quantitative measure of the quality of performance of the algorithms is desirable. One might conjecture that the way to do this is to establish a bound on the remainder term for the series expansion for the cost functional. That does not appear to be a trivial problem.

Along these same lines, an analysis would seem desirable of the convergence properties of the solution of the modal-trajectory TPBVP for ϵ on the interval $[0,1]$.

We have used the terms light coupling and small nonlinearity. An investigation establishing a quantitative definition of those terms within the context of the problems studied here would be of great practical value.

In Joseph [28] and Pentecost [19] a method was proposed for linear filtering which utilized a partitioning of a system into subsystems. A study of the relationship of that method with the present one would be of interest.

Throughout this work we have indicated that the estimation procedure would be implemented digitally. In many, perhaps most, applications it would be more desirable to convert the continuous problem to a discrete one and perform discrete estimation rather than convert the continuous estimation procedure to a discrete form. Given the foundation presented here, it should be a simple matter to derive the discrete analogy to the algorithms given here.

VI. REFERENCES

- [1] H. J. Kushner, "Dynamical equations for optimal nonlinear filtering," J. Differential Equations, vol. 3, pp. 179-190, 1967.
- [2] _____, "Nonlinear filtering: the exact dynamical equations satisfied by the conditional mode," IEEE Trans. Automat. Contr., vol. AC-12, pp. 262-267, Jun. 1967.
- [3] H. Cox, "Estimation of state variables via dynamic programming," Proc. 1964 Joint Automatic Control Conf., 1964, pp. 376-381.
- [4] R. E. Mortensen, "Maximum-likelihood recursive nonlinear filtering," J. Optimization Theory Appl., vol. 2, no. 6, pp. 386-394, 1968.
- [5] R. E. Kalman, "A new approach to linear filtering and prediction problems," Trans. ASME, Ser. D: J. Basic Eng., vol. 82, pp. 394-400, 1960.
- [6] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," Trans. ASME, Ser. D: J. Basic Eng., vol. 83, pp. 95-108, 1961.
- [7] A. H. Jazwinski, Stochastic Processes and Filtering Theory. New York: Academic Press, 1970.
- [8] A. P. Sage and J. L. Melsa, Estimation Theory with Applications to Communications and Control. New York: McGraw-Hill, 1971.
- [9] P. V. Kokotovic and G. Singh, "Optimization of coupled non-linear systems," Int. J. Contr., vol. 14, no. 1, pp. 51-64, 1971.
- [10] R. A. Werner and J. B. Cruz, Jr., "Feedback control which preserves optimality for systems with unknown parameters," IEEE Trans. Automat. Contr., vol. AC-13, pp. 621-629, Dec. 1968.
- [11] F. N. Bailey, "The application of Lyapunov's second method to interconnected systems," J. SIAM Contr., Ser. A, vol. 3, no. 3, pp. 443-462, 1966.

- [12] P. V. Kokotović and P. Sannuti, "Singular perturbation method for reducing the model order in optimal control design," IEEE Trans. Automat. Contr., vol. AC-13, pp. 377-384, Aug. 1968.
- [13] J. B. Cruz, Jr., Feedback Systems. New York: McGraw-Hill, 1972.
- [14] A. H. Haddad and J. B. Cruz, " ϵ -coupling method for near-optimum design of large-scale linear systems," Proc. IEE (correspondence), vol. 117, no. 1, pp. 223-224, 1970.
- [15] R. E. Mortensen, "Optimal control of continuous-time stochastic systems," Ph.D. dissertation, Univ. of California, Berkeley, California, 1966.
- [16] L. S. Pontryagin, V. Boltyanskii, R. Gamkrelidze, and E. Mishchenko, The Mathematical Theory of Optimal Processes. New York: Wiley-Interscience, 1962.
- [17] E. A. Coddington and N. Levinson, Theory of Ordinary Differential Equations. New York: McGraw-Hill, 1955.
- [18] A. E. Bryson and M. Frazier, "Smoothing for linear and nonlinear dynamic systems," U.S. Air Force Tech. Rept. ASD-TDR-063-119, 1963.
- [19] E. E. Pentecost, "Synthesis of computationally efficient sequential linear estimators," Ph.D. dissertation, Univ. of California, Los Angeles, California, 1965.
- [20] M. Aoki and J. R. Huddle, "Estimation of the state vector of a linear stochastic system with a constrained estimator," IEEE Trans. Automat. Contr., vol. AC-12, pp. 432-433, Aug. 1967.
- [21] R. S. Bucy and P. D. Joseph, Filtering for Stochastic Processes with Applications to Guidance. New York: Wiley-Interscience, 1968.
- [22] P. V. Kokotović, W. R. Perkins, J. B. Cruz, Jr., and G. D'Ans, " ϵ -coupling method for near-optimum design of large-scale linear systems," Proc. IEE, vol. 116, no. 5, pp. 49-52, 1969.

- [23] J. S. Meditch, "A class of suboptimal linear controls," IEEE Trans. Automat. Contr., vol. AC-11, pp. 433-439, Jul. 1966.
- [24] A. E. Bryson and L. J. Henrikson, "Estimation using sampled data containing sequentially correlated noise," J. Spacecraft and Rockets, vol. 5, no. 6, pp. 662-665, Jun. 1968.
- [25] B. Friedland, "On the effect of incorrect gain in the Kalman filter," IEEE Trans. Automat. Contr., vol. AC-12, no. 5, p. 610, Oct. 1967.
- [26] P. V. Kokotovic and J. B. Cruz, Jr., "An approximation theorem for linear optimal regulators," J. Math. Anal. Appl., vol. 27, pp. 249-252, 1969.
- [27] G. R. Pitman, ed., Inertial Guidance. New York: Wiley-Interscience, 1962.
- [28] P. D. Joseph, "Sub-optimal linear filtering," Interoffice correspondence 9321.4-653, Space Technology Laboratories, Inc., Redondo Beach, California, December 5, 1963.

VII. ACKNOWLEDGMENTS

The author wishes to thank his major professor, Dr. R. G. Brown, for his encouragement and counsel during the course of this study and throughout his graduate career.

He also would like to thank Dr. A. N. Michel who first stimulated his interest in the study of interconnected systems.

Additionally, he wishes to acknowledge the financial support of a National Defense Education Act fellowship and the partial support of the Themis Project, Office of Naval Research, contract N00014-68-A-0162.

To his wife, Linda, no acknowledgment can possibly convey the appreciation for her expressions of encouragement and confidence and for her excellent typing of the manuscript under trying circumstances.

VIII. APPENDIX

The purpose of this Appendix is to state for reference purposes the linear estimation problem and summarize its solution, the Kalman filter.

Suppose that a representation of a dynamical system is given by the (formal) stochastic differential equation

$$\dot{x}(t) = F(t)x(t) + w(t), \quad t \geq t_0 \quad (\text{A.1})$$

where $x(t)$ is the n -vector state and $F(\cdot)$ is an $n \times n$ known continuous matrix time-function. Continuous measurements are taken via

$$y(t) = H(t)x(t) + v(t), \quad t \geq t_0 \quad (\text{A.2})$$

where $y(t)$ is the m -vector measurement and $H(\cdot)$ is a known $m \times n$ continuous matrix time-function. $\{w(t), t \geq t_0\}$ and $\{v(t), t \geq t_0\}$ are zero-mean gaussian white noise processes with

$$\begin{aligned} \text{cov}\{w(t_1), w(t_2)\} &= Q(t_1)\delta(t_1 - t_2), \\ \text{cov}\{v(t_1), v(t_2)\} &= R(t_1)\delta(t_1 - t_2). \end{aligned} \quad (\text{A.3})$$

The initial state $x(t_0)$ is normally distributed, $x(t_0) \sim N(\mu, P(t_0))$ with $x(t_0)$, $\{w(t)\}$, and $\{v(t)\}$ independent.

The problem then is, given the above model of the

dynamical system, determine the minimum variance estimate of the state at any time $t \geq t_0$ given the measurement data $\{y(\tau), t_0 \leq \tau \leq t\}$. The solution to this problem is given by the well-known Kalman-Bucy filter which was first derived (in the continuous form) by Kalman and Bucy in 1961 [6].

The Kalman-Bucy filter for the continuous system (A.1) - (A.2) is given by the differential equations

$$\begin{aligned} \dot{x}^*(t) &= F(t)x^*(t) + P(t)H'(t)R^{-1}(t)[y(t) - H(t)x^*(t)], \\ x^*(t_0) &= \mu, \quad t \geq t_0 \end{aligned} \quad (\text{A.4})$$

where $P(t)$ satisfies the matrix Riccati equation

$$\begin{aligned} \dot{P}(t) &= F(t)P(t) + P(t)F'(t) + Q(t) - P(t)H'(t)R^{-1}H(t)P(t), \\ t &\geq t_0 \end{aligned} \quad (\text{A.5})$$

with the initial condition the covariance matrix of the initial state, $P(t_0)$. $x^*(t)$ is the minimum variance estimate of the state at time t and $P(t)$ is the covariance matrix of the estimation error resulting from $x^*(t)$.

Solution of the linear equation (A.4) with the Riccati equation (A.5) as measurements are obtained produces the minimum variance estimate of the state as a linear operation on the measurement data.