1967

Variance of the state estimation error of an adaptive Kalman filter

Allan Eugene Baker
Iowa State University

VARIANCE OF THE STATE ESTIMATION ERROR OF AN ADAPTIVE KALMAN FILTER

by

Allan Eugene Baker

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

Major Subject: Electrical Engineering

Approved:

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

Head of Major Department

Signature was redacted for privacy.

Dean of Graduate College

Iowa State University
Of Science and Technology
Ames, Iowa

1967

TABLE OF CONTENTS

INTRODUCTION

The estimation problem in control or communication theory consists of estimating the value of a signal, or message process, in the presence of noise, or measurement error. The signal and the noise are assumed to be random processes.

The optimum estimator in our case is an operation on the combined signal and noise which yields an estimate of the signal such that the mean-squared error is minimized over the period of time in question. The error is defined as the difference between the estimate of the signal and the signal itself.

One solution to this problem was originated by Wiener (16). The Wiener filter is designed to operate on an infinite amount of past data and is usually realized as a physically realizable network of lumped circuit elements. Wiener's original work has been extended and clarified by many later authors.

A somewhat different approach to the solution of the estimation problem is that taken by Kalman (6, 7). This filtering method was also derived independently by Battin (1). The Kalman filter is realized as a sequential method of operating on each reading or measurement as it is taken. This is done by means of a digital computer program. The Kalman filter is designed to operate on a finite amount of past data. Kalman's work has also been extended and restated, both by Kalman himself and by other authors.

Both the Wiener and Kalman filters require the knowledge of the means and variances of the signal and noise in order for the optimal filter to be specified. In cases where they are not known, they must be either estimated by statistical methods, or guessed at, or an alternative filtering method must be used.

Since the Kalman filter is a sequential technique capable of being easily modified as the statistics of the process change in time, it lends itself much more easily to the case where the statistics are not known and must be estimated from the available measurements.

The Kalman filter is the optimal filter for Gaussian random processes. In the case where the processes are not Gaussian, the Kalman filter is not the optimal filter, but only the optimal linear filter; and even then it is only optimal in the least squares sense.

When the mean and variance are not known exactly and must be estimated, the filtering technique used is suboptimal when compared to the case where the process statistics are known; and therefore the mean-squared error in the estimate of the state is larger.

If the statistical properties of the signal or noise are not known completely and must be estimated statistically, what is the actual variance of the state estimation error? How is this error variance improved by statistical estimation

techniques? It is desired to know this information both when the statistical parameter in question is time-stationary and when it is changing, either abruptly or slowly, with time. Certain estimation techniques will give better results when the parameter is time-stationary, and others will give better results when it is time-changing. This is especially the case when the filter is to be optimized over a finite time interval after a rapid change in the parameter.

Information on the actual state estimation error is useful also because some solutions to the adaptive filtering problem require more involved computations than others, but give greater accuracy. Ultimately some compromise between accuracy and complexity will probably have to be made.

In the problem considered here, it is assumed that the statistical properties of the signal, or message process, are known. The noise, or measurement error, is assumed to be uncorrelated, with zero mean and unknown variance.

REVIEW OF LITERATURE

Much literature has been written on various aspects of Kalman filtering techniques, but very little on the particular problem considered here.

Bucy and Follin (2), in a 1962 paper, analyze the steady-state and transient behavior of an adaptive filter for one particular simple two-state system. Their analysis is done using the continuous form of the Kalman filter equations, rather than the discrete-time form. The discrete-time equations are of more general use, because of their compatibility with digital computing techniques. Bucy and Follin include results obtained by a simulation of the system on an analog computer.

The two main formulations of adaptive Kalman filters which apply directly to the discrete-time case are those of Shellenbarger (13) and Magill (9). Magill, by separating the possible values of the unknown statistical parameter into a finite set of numbers and assigning a priori probabilities to each of these possibilities, arrives at an adaptive technique which is claimed to be optimal. This technique will work especially well in the case of a change in time of the parameter from one discrete value to another, although it will involve more computations and assumes some a priori knowledge of the parameter involved.

Shellenbarger has given solutions to the adaptive estimation problem when the measurement error covariance matrix, or the signal response covariance matrix, or both, are unknown. These solutions involve a maximum likelihood (10) or "maximum-probability" type of statistical estimator for the unknown parameter. Since Shellenbarger's adaptive filter requires fewer computations than other known methods, it will be used as the basis for the work considered here.

Some work has been published which deals with the effect of erroneous parameters in the signal and noise processes on the actual error in the estimation of the signal. Nishimura (11) has considered the effect of incorrect initial values of the estimation error covariance matrix, P, on the solutions of the variance equation. He has identified three quantities, P_c (calculated error variance), P_a (actual error variance), and P_o (variance of the error due to an optimal filter). He has derived difference equations for $E_{ca} = P_c - P_a$, $E_{ao} = P_a - P_o$, and $E_{co} = P_c - P_o$. These equations assume that the models of the signal and noise functions, with the exception of the initial error covariance matrix, are known. Nishimura also simulates an example system, which includes a phase-locked loop, and derives a relative "time of convergence" of the state estimation error variance to its final value.

Heffes (5), in a short paper, extends Nishimura's work to the case where there is an error in the noise model. He works

a simple example and gives graphical results for P_c, P_a, and P_o.

Fagin (4) has also derived some equations similar to those of Heffes and Nishimura.

Questions not treated by Nishimura, Heffes, or Fagin are those of obtaining additional information about the model and of how any additional information obtained would affect the actual estimation error.

Swerling (14) also mentions the existence of problems quite similar to these, and gives a few hints as to their possible solution.

REVIEW OF KALMAN FILTERING

In order to identify and standardize the notation used in the following sections, a review of the Kalman filter equations seems to be in order. Since they have been derived many times before, no attempt will be made to derive them here. A summary of some of the different methods of derivation is made by Lee (8).

The signal and noise processes are defined as in Figure 1. The equations of the system are:

$$\dot{\underline{x}} = A\underline{x} + B\underline{u} \tag{1}$$

$$\underline{y} = M\underline{x} + \underline{v} \tag{2}$$

Since the output is being sampled at discrete instants of time t_k, the equivalent difference equations of the system are:

$$\underline{x}(k+1) = \Phi(k+1;k)\ \underline{x}(k) + \underline{g}(k) \tag{3}$$

$$\underline{y}(k) = M\underline{x}(k) + \underline{v}(k), \tag{4}$$

where Φ is the state transition matrix,

$$\underline{g} = \int_{t_k}^{t_{k+1}} \Phi(t,\tau)\ B\underline{u}(\tau)d\tau, \tag{5}$$

and M is the measurement matrix.

Since the measurement of the state is contaminated by additive noise, one can not determine the exact value of the state at a particular time, but can only estimate it.

In the Kalman filter, the optimal estimate is given by

$$\hat{\underline{x}}_k = \hat{\underline{x}}'_k + b_k(\underline{y}_k - \hat{\underline{y}}'_k) \tag{6}$$

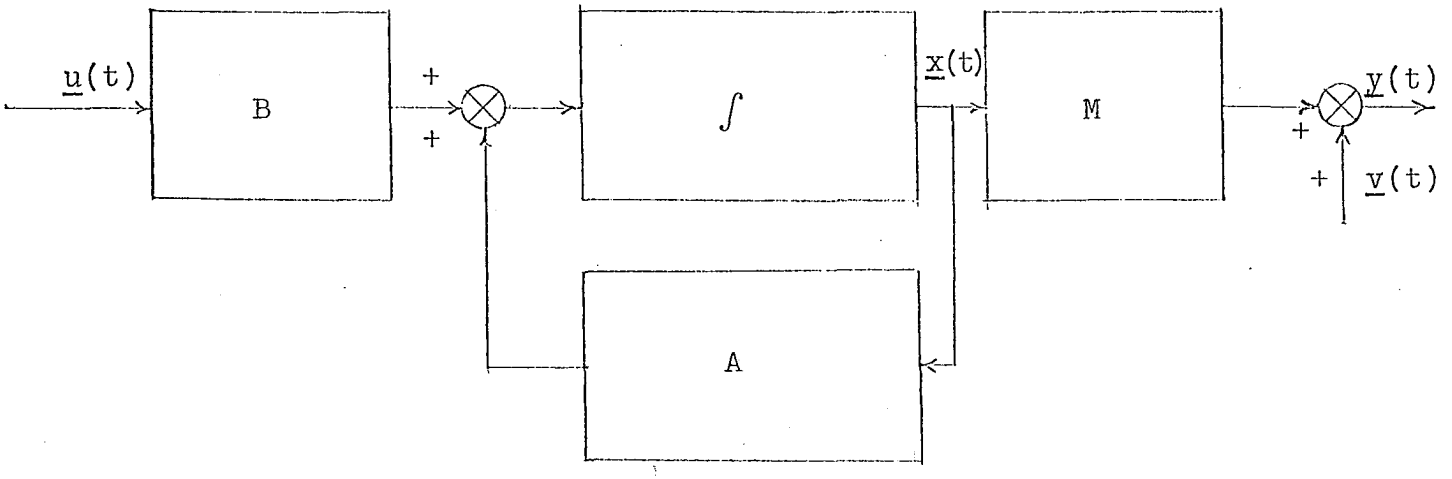


Figure 1. Signal and noise processes as used in Kalman filter formulation

where $$\hat{x}_k' = \Phi\, \hat{x}_{k-1} \text{ and } \hat{y}_k' = M\hat{x}_k', \tag{7}$$

and $$b_k = P_k^* M_k^T (M_k P_k^* M_k^T + V)^{-1}$$

$$P_k^* = E[(\hat{x}_k' - x_k)(\hat{x}_k' - x_k)^T] \tag{8}$$

is the covariance matrix of the error of the estimate of x due to all measurements up to and including y_{k-1}.

$V_k = E[v_k v_k^T]$ is the covariance matrix of the measurement error.

P_k is the covariance matrix of the error of the estimate of x by all measurements up to and including y_k, and is equal to $E[(\hat{x}_k - x_k)(\hat{x}_k - x_k)^T]$.

P_k and P_k^* are related as follows:

$$P_k^* = \Phi P_{k-1} \Phi^T + H \tag{9}$$

$$P_k = P_k^* - P_k^* M_k^T (M_k P_k^* M_k^T + V_k)^{-1} M_k P_k^*$$

$$= P_k^* - b_k (M_k P_k^* M_k^T + V_k) b_k^T. \tag{10}$$

$H = E(g_k g_k^T)$ is the covariance matrix of the response of the system due to the white-noise inputs between times t_k and t_{k+1}.

PARAMETER ESTIMATION METHODS

If nothing about the variance of the measurement errors is known, the most attractive way of attacking the estimation problem seems to be that of using the framework of the Kalman filter equations, as Shellenbarger has done. The variance is estimated from the measured data and inserted into the Kalman filter equations to form the weighting factor.

From one viewpoint, there are basically two aspects of this problem. The first is that of fixing some sort of bound on the mean-squared error in the estimation of V, and the second is to see how this error in the estimation of V propagates and affects the actual mean-squared error in the estimate of the state variable. The latter aspect is the more important, because it is the estimate of the state that we are interested in. We are interested in the characteristics of the filter both in the case where the noise properties are time-stationary and in the case where they are time-changing.

Two adaptive estimation techniques are given by Shellenbarger. They work well, but they do have some shortcomings, as will be described below.

Maximum-Probability Estimation

First we shall consider the "maximum-probability" estimate of Shellenbarger. This estimate is given by

$$\hat{V}_i = y_i y_i^T - M_i X_i M_i^T, \tag{11}$$

where $\hat{V}_i$ is the estimate of the measurement error variance matrix at time t_i, y_i is the measurement vector at time t_i, X_i is the variance of the message or signal process at t_i, and M_i is the measurement matrix at t_i.

The above estimate, taken alone, gives an estimate of V based on one reading. A more accurate estimate is that based on several readings taken over a period of time and is given by a simple average $\bar{V} = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i$, or by an exponentially weighted average. In order to fix some bounds on the range of the estimate $\bar{V}$ and hence to have a basis for comparison of various estimation methods and to gain some insight into the behavior of the state estimation error, the properties of this estimate will be considered here.

Since $\hat{V}_i = y_i y_i^T - M_i X_i M_i^T$,

$$\bar{V} = \frac{1}{n} \sum_{i=1}^{n} y_i y_i^T - \frac{1}{n} \sum_{i=1}^{n} M_i X_i M_i^T. \qquad (12)$$

It is assumed that the X_i's are known. X_0 is the covariance matrix of the random process at time t_0, and $X_{i+1} = \Phi_i X_i \Phi_i^T = H_i$. The latter term in Equation 2 is then deterministic and the first term is statistical. Since the mean of the process y is assumed to be zero, the first term of $\bar{V}_i$ would be the maximum likelihood estimator for the variance of y if the individual readings y_i were independent of each other. The expression

$$\frac{1}{n} \sum_{i=1}^{n} y_i^2,$$

which applies to the one-dimensional case, would then have a chi-square distribution with n degrees of freedom (9). However, since the signal process is correlated from one instant of time to the next, the measurements are not independent, but are correlated normally distributed random variables. The quadratic form of this distribution, $y^T Y^{-1} y$, where Y^{-1} is the inverse of the covariance matrix, is distributed as a chi-square variate with n degrees of freedom for n readings. Although the distribution of $\frac{1}{n} \sum_{i=1}^{n} y_i y_i^T$ is difficult to determine, the variance of the random variable $\overline{V} = \frac{1}{n} \sum_{i=1}^{n} y_i y_i^T$ can be determined and compared with the variance of an average of independent observations.

Shellenbarger defines the variance of $\hat{V}$ as a matrix whose elements are the variances of the corresponding elements of the estimate of V. When the measurement is a scalar quantity, the variance of $\hat{V}$ is a one-by-one matrix, or scalar.

The variance of the estimate $\hat{\theta}$ of a parameter θ is defined as the expectation $E(\hat{\theta} - \theta)^2$. For an estimate based on a single measurement, Shellenbarger finds that the elements of the matrix which represents the variance of $\hat{V}$ are given by

$$E[(\hat{V}-V)*(\hat{V}-V)]_{ij} = E[(yy^T - Y)*(yy^T - Y)]_{ij}$$
$$= (Y)_{ij}^2 + (Y)_{ii}(Y)_{jj}. \qquad (13)$$

Y is the variance of the process $y = Mx + v$ and is equal to $M_i X_i M_i^T + V_i$. When y is a scalar quantity the variance of $\hat{V}$ reduces to $2Y^2$.

Let us now look at an estimate $\bar{V}$ formed from two readings. Since, in most of the systems of interest here, y is a scalar quantity, we will consider that case first.

$$E[\tfrac{1}{2}(y_1y_1 + y_2y_2) - Y]^2 = E[\tfrac{1}{4}(y_1y_1 + y_2y_2)^2 - Y(y_1y_1 + y_2y_2) + Y^2] \quad (14)$$

$$= E[\tfrac{1}{4}(y_1y_1 + y_2y_2)^2] - 2Y^2 + Y^2.$$

$$\tfrac{1}{4}(y_1y_1 + y_2y_2)^2 = \tfrac{1}{4}(y_1y_1y_1y_1 + 2y_1y_1y_2y_2 + y_2y_2y_2y_2). \quad (15)$$

We know

$$E(y_iy_iy_iy_i) = 3[E(y_i^2)]^2 = 3Y^2 \quad (16)$$

where $Y = E(y^2)$,

and also

$$E(y_iy_iy_jy_j) = Y^2 + 2[E(y_iy_j)]^2. \quad (17)$$

By using this information, we find that the right side of Equation 14 is equal to

$$\tfrac{1}{4}[3Y^2 + 2Y^2 + 4E(y_1y_2)^2 + 3Y^2] - Y^2$$

$$= Y^2 + [E(y_1y_2)]^2. \quad (18)$$

Now $E(y_1y_2) = E[(M_1x_1 + v_1)(M_2x_2 + v_2)]$

$$= E[(M_1x_1x_2M_2 + M_1x_1v_2 + M_2x_2v_1 + v_1v_2)]$$

$$= M_1\Phi(t_2-t_1)M_2, \quad (19)$$

where $\Phi(t_2-t_1)$ is the autocorrelation function of the process x evaluated between instants t_1 and t_2.

So finally Equation 14 is equal to

$$Y^2 + M_1^2 M_2^2 [\Phi(t_2 - t_1)]^2. \qquad (20)$$

For our purposes here we will assume M is constant; that is $M_1 = M_2 = \ldots = M_n$.

In the case where we have three correlated readings we may derive, in a manner similar to that above,

$$E[(\hat{V}-V)(\hat{V}-V)] = \frac{2Y^2}{3} + \frac{8M^4\Phi^2(t_2-t_1)}{9} + \frac{4}{9}M^4\Phi^2(t_3-t_1) \qquad (21)$$

In general, by following derivations similar to the above, one can find that the variance of an MP estimate of V based on n observations is

$$\frac{1}{n^2}\left[2nY^2 + \sum_{k=1}^{n-1} 4KM^4\Phi^2(n-k)\right] \qquad (22)$$

Maximum-Conditional-Probability Estimation

The "maximum-conditional-probability" (MCP) estimator of Shellenbarger is given by

$$\hat{V}_i = (y_i - \hat{y}_i')(y_i - \hat{y}_i')^T - M_i P_i^*(c) M_i^T \qquad (23)$$

where $\hat{V}_i$ = estimate of measurement error covariance matrix at time t_i

y_i = measurement vector at time t_i

$P_i^*(c)$ = calculated a priori covariance matrix of estimation error at time t_i

M_i = measurement matrix

Again, as with the MP estimate, averages of the form

$$\bar{V}_i = \frac{1}{n}\sum_{i=1}^{n} \hat{V}_i \qquad (24)$$

or $$\bar{V}_i = (1-e^{-\frac{1}{p}})\,\hat{V}_i + e^{-\frac{1}{p}}\,\bar{V}_{i-1} \tag{25}$$

may be formed for more accurate estimation.

The mean of the estimate formed from a single measurement is as follows:

$$E[\hat{V}_i] = E[(y_i-\hat{y}_i')(y_i-\hat{y}_i')^T - M_i P_i^*(c)M_i^T] \tag{26}$$

$$E(y_i-\hat{y}_i')(y_i-\hat{y}_i')^T = MP_i^*(a)M^T + V_i \tag{27}$$

$$\begin{aligned} E[\hat{V}] &= [MP^*(a)M^T + V - MP^*(c)M^T] \\ &= V - ME_{ca}^*M^T \end{aligned} \tag{28}$$

where $E_{ca}^* = P^*(c) - P^*(a)$

In general, the average value of E_{ca}^* is equal to zero.

The variance of the estimate formed by one measurement is given below for the case where Y is one-dimensional.

$$\begin{aligned} &E[(y-\hat{y}')^2 - MP^*(c)M^T - V]^2 \\ &\quad = E[(y-\hat{y}')^4 - 2(y-\hat{y}')^2(MP(c)^*M^T + V) + \\ &\qquad (MP(c)^*M^T + V)^2]. \end{aligned} \tag{29}$$

Since $$E[(y-\hat{y}')]^4 = 3[E(y-\hat{y}')^2]^2 = 3(MP^*(a)M^T + V)^2, \tag{30}$$

we have $$\begin{aligned} E[(\hat{V}-V)^2] = {} & 3(MP^*(a)M^T + V)^2 - 2(MP^*(a)M^T + V) \\ & (MP^*(c)M^T + V) + (MP^*(c)M^T + V)^2 \end{aligned} \tag{31}$$

Let us compare this with the variance of a single MP estimate of V, which is $2Y^2 = 2(MXM^T + V)^2$.

Since P*(c) and P*(a) should both be less than X, this MCP estimate has a smaller variance, as is also shown by Shellenbarger (13).

Let us now look at the variance of two averaged estimates, for the one-dimensional case.

$$E[\hat{V}-V]^2 = E\left\{\frac{1}{2}[(y_1-\hat{y}_1')^2 + (y_2-\hat{y}_2')^2 - MP_1^*(c)M^T - MP_2^*(c)M^T] - V\right\}^2 \qquad (32)$$

The quantity inside the braces is equal to

$$\frac{1}{4}[(y_1-\hat{y}_1')^2 + (y_2-\hat{y}_2')^2 - MP_1^*(c)M^T - MP_2^*(c)M^T]^2$$

$$- [(y_1-\hat{y}_1')^2 + (y_2-\hat{y}_2')^2 - MP_1^*(c)M^T - MP_2^*(c)M^T]V + V^2$$

$$= \frac{1}{4}[(y_1-\hat{y}_1')^4 + 2(y_1-\hat{y}_1')^2(y_2-\hat{y}_2')^2 + (y_2-\hat{y}_2')^4$$

$$- 2(y_1-\hat{y}_1')^2(MP_1^*(c)M^T) - 2(y_1-\hat{y}_1')^2(MP_2^*(c)M^T)$$

$$- 2(y_2-\hat{y}_2')^2(MP_1^*(c)M^T) - 2(y_2-\hat{y}_2')^2(MP_2^*(c)M^T)$$

$$+ (MP_1^*(c)M^T)^2 + 2(MP_1^*(c)M^T)(MP_2^*(c)M^T) + (MP_2^*(c)M^T)^2]$$

$$- [(y_1-\hat{y}_1')^2 + (y_2-\hat{y}_2')^2 - MP_1^*(c)M^T - MP_2^*(c)M^T]V + V^2. \qquad (33)$$

The above expression contains the term

$$E[(y_1-\hat{y}_1')^2(y_2-\hat{y}_2')^2].$$

We know that $E[y_1y_1y_2y_2] = E(y_1y_1)E(y_2y_2) + E(y_1y_2)E(y_2y_1)$

$$+ E(y_1y_2)E(y_1y_2) \qquad (34)$$

Therefore $E[(y_1-\hat{y}_1')^2(y_2-\hat{y}_2')^2]$

$$= E(y_1-\hat{y}_1')^2E(y_2-\hat{y}_2')^2 + 2E[(y_1-\hat{y}_1')(y_2-\hat{y}_2')]. \quad (35)$$

In order to evaluate these terms, we must digress still further and evaluate

$$E[(y_1-\hat{y}_1')(y_2-\hat{y}_2')].$$

We know that

$$\begin{aligned}(y_i-\hat{y}_i') &= Mx_i + v_i - M\Phi x_{i-1} \\ &= M(x_i-\Phi x_{i-1}) + v_i \\ &= M(x_i-\Phi x_{i-1}) - M\Phi(x_{i-1}-x_{i-1}) + v_i \\ &= Mg_{i-1} - M\Phi e_{i-1} + v_i \quad (36)\end{aligned}$$

Similarly $(y_{i+1} - y_{i+1}') = Mg_i - \Phi Me_i + v_{i+1}$ (37)

$$e_n' = \Phi e_{n-1} - g_{n-1} \quad (38)$$

$$\begin{aligned}e_n &= \hat{x}_n' - x_n + b_n(y_n-\hat{y}_n') \\ &= e_n' + b_n(Mg_{n-1} - \Phi Me_{n-1} + v_n) \quad (39)\end{aligned}$$

therefore $e_n = \Phi_{n-1}e_{n-1} - g_{n-1} + b_n(Mg_{n-1} - \Phi Me_{n-1} + v_n)$ (40)

and $(y_{i+1} - \hat{y}_{i+1}') = Mg_i - M\Phi^2e_{i-1} + M\Phi g_{i-1}$

$$- M\Phi b_i(Mg_{i-1} - \Phi Me_{i-1} + v_i) + v_{i+1} \quad (41)$$

$$E[(y_i-\hat{y}_i')(y_{i+1} - \hat{y}_{i+1}')] = \Phi^3M^2P_{i-1}(a)(1-b_iM)$$

$$+ M^2\Phi(1-b_iM)H - b_iV_iM\Phi \quad (42)$$

By carrying the above derivation one step further, we have

$$E[(y_i-\hat{y}_i')(y_{i+2}-\hat{y}_{i+2}')]$$

$$= \Phi^4M^2(1-b_iM)(1-b_{i+1}M)P_{i+1}(a) + \Phi^2M^2(1-b_iM)(1-b_{i+1}M)H$$

$$- \Phi^2b_iM(1-b_{i+1}M)V. \tag{43}$$

In general, the relationship is

$$E[(y_n-\hat{y}_n')(y_{n+k}-\hat{y}_{n+k}')] = M^2\Phi^k\left[\prod_{i=n}^{n+k-1}(1-b_iM)\right]H$$

$$+ M^2\Phi^{k+2}\left[\prod_{i=n}^{n+k-1}(1-b_iM)\right]P_n(a)$$

$$- \Phi^kb_nM\left[\prod_{i=n+1}^{n+k-1}(1-b_iM)\right]V. \tag{44}$$

From Equation 42, we have

$$E[(y_1-\hat{y}_1')(y_2-\hat{y}_2')] = M^2\Phi(1-b_1M)H + M^2\Phi^3(1-b_1M)P_o - \Phi b_1MV \tag{45}$$

We now return to Equation 32 and evaluate the variance of the estimate $\bar{V}$. When n=2 we have

$$E[(\hat{V}-V)(\hat{V}-V)] = E\left\{\frac{1}{4}[(y_1-\hat{y}_1')^4 + 2(y_1-\hat{y}_1')^2(y_2-\hat{y}_2')^2 + (y_2-\hat{y}_2')^4\right.$$

$$- 2(y_1-\hat{y}_1')^2(MP_1^*(c)M^T) - 2(y_1-\hat{y}_1')^2(MP_2^*(c)M^T)$$

$$- 2(y_2-\hat{y}_2')^2(MP_1^*(c)M^T) - 2(y_2-\hat{y}_2')^2(MP_2^*(c)M^T) + (MP_1^*(c)M^T)^2$$

$$+ 2(MP_1^*(c)M^T)(MP_2^*(c)M^T) + (MP_2^*(c)M^T)^2] - [(y_1-\hat{y}_1')^2$$

$$\left.+ (y_2-\hat{y}_2')^2 - MP_1^*(c)M^T - MP_2^*(c)M^T]V + V^2\right\} \tag{46}$$

$$= \frac{1}{4}\{3(MP_1^*(a)M^T + V)^2 + 2(MP_1^*(a)M^T + V)(MP_2^*(a)M^T + V)$$

$$+ 4[M^2\Phi^3(1-b_1M)P_o + M^2\Phi(1-b_1M)H - \Phi b_1 MV]^2$$

$$+ 3(MP_2^*(a)M^T + V)^2 - 2(MP_1^*(a)M^T + V)(MP_1^*(c)M^T)$$

$$- 2(MP_1^*(a)M^T + V)(MP_2^*(c)M^T) - 2(MP_2^*(a)M^T + V)(MP_1^*(c)M^T)$$

$$- 2(MP_2^*(a)M^T + V)(MP_2^*(c)M^T) + (MP_1^*(c)M^T)^2$$

$$+ 2(MP_1^*(c)M^T)(MP_2^*(c)M^T) + (MP_2^*(c)M^T)^2\}$$

$$- [(MP_1^*(a)M^T + V) + (MP_2^*(a)M^T + V) - MP_1^*(c)M^T - MP_2^*(c)M^T]V$$

$$+ V^2. \qquad (47)$$

In order to gain more insight into the nature of the above expression, let us assume that $P_1^*(a) = P_2^*(a) = P_1^*(c) = P_2^*(c) = P^*$. Then $E[(\hat{V}-V)(\hat{V}-V)] =$

$$2(MP^*M^T + V)^2 - 2(MP^*M^T + V)(MP^*M^T) + (MP^*M^T)^2$$

$$+ [M^2\Phi^3(1-b_1M)P_o + M^2\Phi(1-b_1M)H - \Phi b_1 MV]^2 + V^2$$

$$= (MP^*M^T + V)^2 + [M^2\Phi^3(1-b_1M)P_o + M^2\Phi(1-b_1M)H - \Phi b_1 MV]^2 \qquad (48)$$

This is comparable to Equation 20 for the Maximum-Probability Estimator.

Let us look at the second term of Equation 48 when weighting is optimal; that is, when $b_1 = \frac{P_1^*}{P_1^* + V}$ and $M = 1.0$, in the one-dimensional case.

$$E[(y_1-\hat{y}_1')(y_2-\hat{y}_2')] = \Phi^3(1 - \frac{P^*}{P^* + V})P_o + \Phi(1 - \frac{P^*}{P^* + V})H$$

$$- \frac{\Phi P^* V}{P^* + V} = \frac{\Phi^3 V P_o}{P^* + V} + \frac{\Phi V H}{P^* + V} - \frac{\Phi P^* V}{P^* + V} \tag{49}$$

$$P_1^* = \Phi P_o \Phi^T + H.$$

Substitution of this term into the above expression gives:

$$\frac{\Phi P^* V}{P^* + V} - \frac{\Phi P^* V}{P^* + V} = 0.$$

Therefore, in the optimal case, $E[(y_1-\hat{y}_1')(y_2-\hat{y}_2')] = 0$, and when the weighting is nearly optimal, the value of this term is small. The same relation can be shown to hold for $E[(y_n-\hat{y}_n')(y_{n+k}-\hat{y}_{n+k}')]$.

By following derivations similar to those performed in the MP case, one can find that the variance of an MCP estimate of V based on three observations is

$$\frac{2}{3}(MP^*M^T + V)^2 + \frac{8}{9}\left\{E[(y_n-\hat{y}_n')(y_{n+1}-\hat{y}_{n+1}')]\right\}^2$$
$$+ \frac{4}{9}\left\{E[(y_n-\hat{y}_n')(y_{n+2}-\hat{y}_{n+2}')]\right\}^2 \tag{50}$$

and in the general case

$$E[(\hat{V}-V)(\hat{V}-V)] = \frac{1}{n^2}\left\{2n(MP^*M^T + V)^2\right.$$
$$\left. + \sum_{k=1}^{n-1} 4k[E(y_k-\hat{y}_k')(y_n - \hat{y}_n')]^2\right\} \tag{51}$$

where the terms of the summation are given by Equation 44.

Similar derivations can be carried through for the variance of the exponentially weighted estimate.

Since $(MP^*M^T + V)$ is less than $(MXM^T + V)$ and since, when the system is near optimal, $E[(y_n - \hat{y}'_n)(y_{n+k} - \hat{y}'_{n+k})]$ is less than $\Phi(\Delta t)$, the variance of an averaged MCP estimator is somewhat less than that of an averaged MP estimator.

The MP estimation technique is more difficult to apply than the MCP technique because a larger number of estimates of V must be made in order to get a usable estimate. When V is relatively small and the variance of the signal process is large, the chances of obtaining a negative value for V, and hence an unusable estimate, are larger. For these reasons, the MCP estimate is used exclusively in the adaptive filter computer simulations.

The result of both these estimation techniques is that the weighting factor is not optimum in the sense of a situation where all the statistics are known. The calculated variance of the estimation error (P-matrix) is neither the optimum value nor the true value. The actual value of the estimation error is inaccessible in a practical situation. However, for purposes of comparing different estimation methods, it can be assumed that the actual value of V is known. Of course, this value is not used in the state estimation, and the weighting of the data is still suboptimal.

The actual variance of the state estimation error is given by the relationships

$$P_a(k) = (I - bM)\, P^*_a(k)\, (I - bM)^T + bVb^T \qquad (52)$$

$$P_a^*(k+1) = \Phi P_a(k)\Phi^T + H, \tag{53}$$

where V is the actual value of V, b is the weighting factor used (not necessarily optimum), and $P_a^*(0)$ is the $P^*(0)$ actually used, which is usually X_o, the variance of the signal process.

SIMULATIONS

In order to obtain information on the behavior of the covariance of the estimation error, several systems were simulated on a digital computer and runs were made for various conditions.

One-State System

White noise with spectral amplitude 0.604 is driving a system with a rational transfer function, as shown in Figure 2. The output y is corrupted by additive uncorrelated measurement errors with unknown variance. The sampling interval is one second. The differential equation for the system is

$$\dot{x} = -0.2x + f(t) \tag{54}$$

$$y = x + v \tag{55}$$

The state transition matrix is given by

$$\Phi = \mathcal{L}^{-1}[sI - A]^{-1} = \mathcal{L}^{-1}\left[\frac{1}{s + 0.2}\right] \tag{56}$$

$$= e^{-0.2t}$$

The symbol $\mathcal{L}^{-1}$ stands for the inverse Laplace transformation. The variance of the system response is given by

$$H = E[g_n g_n^T] = \int_0^1 \int_0^1 e^{-0.2u}\, e^{-0.2v} (0.604)\delta(u-v)\, du\, dv$$

$$= 0.604 \int_0^1 e^{-0.4v}\, dv$$

$$= 0.500 \tag{57}$$

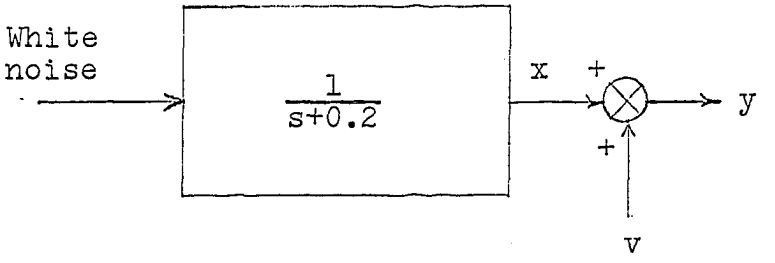


Figure 2. One-state system

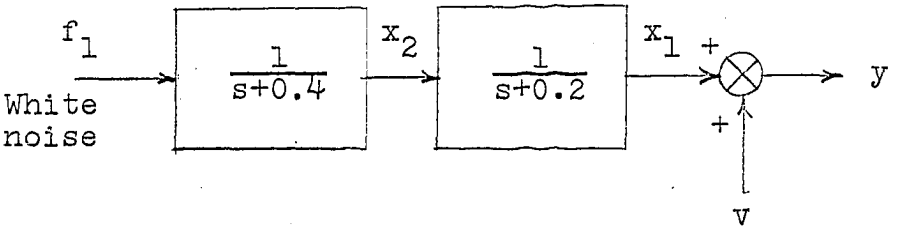


Figure 3. Two-state system

$$\Phi(\Delta t) = e^{-0.2\Delta t} = e^{-0.2} = 0.819 \tag{58}$$

$$P_o^* = \sigma^2 = 1.510 \tag{59}$$

The above values of H, Φ, and P_o^* were used in the Kalman filter equations. V was assumed to be unknown but estimable.

Two-State Systems

Several two-state systems were used. The first had a rational transfer function with two real roots, as shown in Figure 3.

The equations for the system are as follows.

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -0.2 & 1 \\ 0 & -0.4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ f_1 \end{bmatrix} \tag{60}$$

$$y = [1 \quad 0] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v \tag{61}$$

The quantities for the Kalman filter equations are calculated as follows.

$$\Phi(t) = \mathcal{L}^{-1}[sI - A]^{-1}$$

$$= \mathcal{L}^{-1} \begin{bmatrix} \frac{1}{s+0.2} & \frac{1}{(s+0.2)(s+0.4)} \\ 0 & \frac{1}{s+0.4} \end{bmatrix} \tag{62}$$

$$= \begin{bmatrix} e^{-0.2t} & 5(e^{-0.2t} - e^{-0.4t}) \\ 0 & e^{-0.4t} \end{bmatrix} \tag{63}$$

If $\Delta t = 1$ second,

$$\Phi(\Delta t) = \begin{bmatrix} 0.819 & 0.745 \\ 0 & 0.670 \end{bmatrix} \tag{64}$$

$$H_{11} = E[g_1 g_1^T]$$

$$= \int_0^1 \int_0^1 \Phi_{12}(u)\Phi_{12}(v)\, \widetilde{f(t-v)f(t-u)}\, du\, dv \tag{65}$$

$$\widetilde{f(t-u)f(t-v)} = \varphi(t-v-t+u)$$

$$= \varphi(u-v) = \delta(u-v) \text{ for white noise.}$$

So
$$H_{11} = \int_0^1 (25)(e^{-0.2v} - e^{-0.4v})^2\, dv = \underline{\underline{0.500}} \tag{66}$$

$$H_{22} = E[g_2 g_2^T]$$

$$= \int_0^1 \int_0^1 \Phi_{22}(u)\Phi_{22}(v)\delta(u-v)\, du\, dv$$

$$= \int_0^1 (e^{-0.4v})^2\, dv = 0.690 \tag{67}$$

$$H_{12} = H_{21} = E[g_1 g_2^T] = E[g_2 g_1^T]$$

$$= \int_0^1 \int_0^1 e^{-0.4u}(5)(e^{-0.2v} - e^{-0.4v})\delta(u-v)\, du\, dv$$

$$= 5\int_0^1 e^{-0.4v}(e^{-0.2v} - e^{-0.4v})dv = 0.310 \tag{68}$$

Therefore
$$H = \begin{bmatrix} 0.500 & 0.310 \\ 0.310 & 0.690 \end{bmatrix} \tag{69}$$

Another two-state system which was tried was the one shown in Figure 4, in which two independent Markov processes are added together, with independent measurement noise added to the output. The quantities used in the Kalman filter formulation are given below.

$$\Phi(\Delta t=1) = \begin{bmatrix} 0.819 & 0 \\ 0 & 0.670 \end{bmatrix} \qquad (70)$$

$$M = [1 \quad 1] \qquad (71)$$

$$\sigma_1^2 = 1.510 \qquad \sigma_2^2 = 0.909$$

$$H_{11} = \int_0^1 \int_0^1 \sigma_1^2 \, e^{-0.2u} \, e^{-0.2v} \, \delta(u-v) \, du \, dv$$

$$= \int_0^1 \sigma_1^2 (e^{-0.4v}) \, dv = 0.500 \qquad (72)$$

$$H_{12} = H_{21} = 0 \qquad (73)$$

$$H_{22} = \int_0^1 \int_0^1 (0.909)(e^{-0.4u})(e^{-0.4v}) \delta(u-v) \, du \, dv \qquad (74)$$

$$= 0.500$$

More interesting results were obtained from systems whose transfer functions are rational, with imaginary roots.

Let the system in Figure 5 be driven by white noise with unity spectral amplitude. The measurement y is composed of the output of this system plus additive uncorrelated measurement error. The system is set up with phase variables chosen

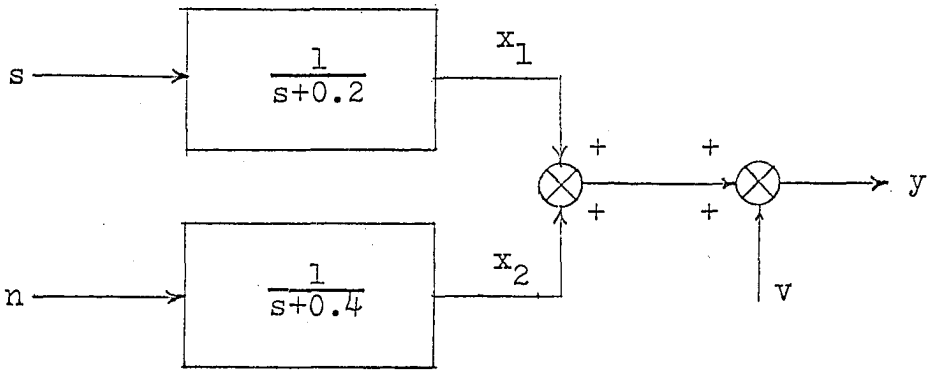


Figure 4. Two-state system with independent states

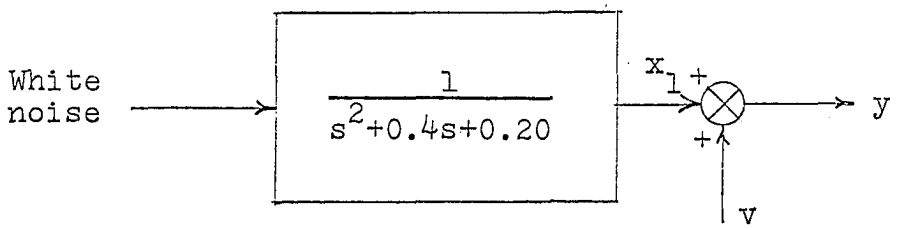


Figure 5. Two-state system with imaginary roots

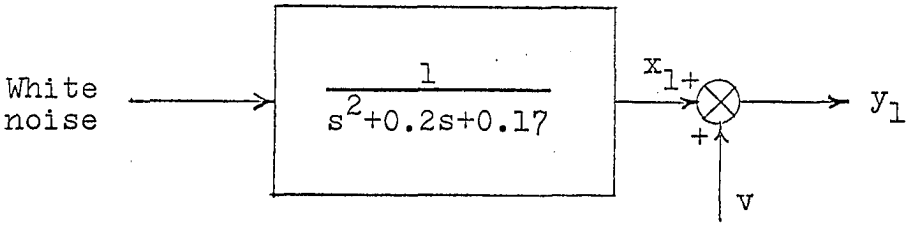


Figure 6. Less heavily damped two-state system

as state variables. The state variable description of the system is given below.

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ -0.20 & -0.40\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix} u(t) \qquad (75)$$

$$y = [1 \quad 0]\begin{bmatrix}x_1\\ x_2\end{bmatrix} + v_1 \qquad (76)$$

The transition matrix $\Phi(t)$ is equal, as before, to the inverse Laplace transform of $[sI-A]^{-1}$.

$$[sI-A]^{-1} = \frac{1}{s^2+0.45+0.20}\begin{bmatrix}s+0.4 & 1\\ -0.20 & s\end{bmatrix} \qquad (77)$$

Therefore

$$\Phi(t) = \begin{bmatrix}e^{-0.2t}(\cos 0.4t+0.5\sin 0.4t) & 2.5e^{-0.2t}\sin 0.4t\\ -0.5e^{-0.2t}\sin 0.4t & e^{-0.2t}(\cos 0.4t-0.5\sin 0.4t)\end{bmatrix} \qquad (78)$$

If we let Δt, the interval between measurements, equal one second, we have

$$\Phi(\Delta t) = \begin{bmatrix}0.914 & 0.795\\ -0.159 & 0.595\end{bmatrix} \qquad (79)$$

As before, we have

$$\underline{g}(t) = \int_0^t \Phi(t,\tau)\, B\, u(\tau)\, d\tau. \qquad (80)$$

Therefore $$g_1(t) = \int_0^t \Phi_{12}(t,\tau)\, u(\tau)\, d\tau \qquad (81)$$

and $$g_2(t) = \int_0^t \Phi_{22}(t,\tau)\, u(\tau)\, d\tau. \tag{82}$$

The elements of the H matrix are again given by the formula

$$H_{ij} = E[g_i g_j^T]. \tag{83}$$

$$H_{11} = \int_0^1 \int_0^1 6.25\, e^{-0.2u} e^{-0.2v} \sin 0.4u \sin 0.4v\, \delta(u-v)\, du\, dv$$

$$= 0.243 \tag{84}$$

$$H_{22} = E[g_2 g_2^T]$$

$$= \int_0^1 \int_0^1 e^{-0.2u} e^{-0.2v} (\cos 0.4u - 0.5 \sin 0.5u)$$
$$(\cos 0.4v - 0.5 \sin 0.4v)\, \delta(u-v)\, du\, dv$$

$$= 0.650 \tag{85}$$

$$H_{12} = E[g_1 g_2^T]$$

$$= \int_0^1 \int_0^1 e^{-0.2u} (\cos 0.4u - 0.5 \sin 0.4u) e^{-0.2v} (2.5 \sin 0.4v)$$
$$\delta(u-v)\, du\, dv$$

By the use of suitable trigonometric identities and integration, we have

$$H_{12} = 0.316$$

and $$\tag{86}$$

$$H_{21} = H_{12}$$

since H is symmetric.

Therefore $$H = \begin{bmatrix} 0.243 & 0.316 \\ 0.316 & 0.650 \end{bmatrix} \tag{87}$$

Another system which has been simulated is shown in Figure 6. This system is similar to the one in Figure 5, but is less heavily damped. The formulation of the system parameters is given below.

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ -0.17 & -0.20\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix} u(t) \tag{88}$$

$$\Phi = \mathcal{L}^{-1}[sI-A]^{-1}$$

$$\text{where } [sI-A]^{-1} = \frac{1}{(s+0.1)^2 + (0.160)}\begin{bmatrix}s+0.2 & 1\\ -0.170 & s\end{bmatrix} \tag{89}$$

$$= \begin{bmatrix}e^{-0.1t}(\cos 0.4t+0.25\sin 0.4t & 2.5e^{-0.1t}\sin 0.4t\\ -0.425e^{-0.1t}\sin 0.4t & e^{-0.1t}(\cos 0.4t-0.25\sin 0.4t\end{bmatrix} \tag{90}$$

$$g_1 = \int_0^1 2.5e^{-0.1u}\sin 0.4u\ u(t,u)\ du \tag{91}$$

$$g_2 = \int_0^1 e^{-0.1u}(\cos 0.4u-0.25\sin 0.4u)\ u(t,u)\ du \tag{92}$$

$$H_{11} = E[g_1 g_1^T]$$

$$= \int_0^1\int_0^1 (2.5)^2 e^{-0.1u}e^{-0.1v}\sin 0.4u\sin 0.4v\ \delta(u-v)\ du\ dv$$

$$= 0.278 \tag{93}$$

$$H_{12} = E[g_1 g_2^T]$$

$$= \int_0^1\int_0^1 e^{-0.1u}e^{-0.1v}(\cos 0.4u-0.25\sin 0.4u)(2.5)$$
$$(\sin 0.4v)\ \delta(u-v)\ du\ dv$$

By using appropriate trigonometric identities and integrating we have

$$H_{12} = 0.392 \tag{94}$$

and

$$H_{21} = H_{12} \tag{95}$$

$$H_{22} = E[g_2 g_2^T]$$

$$= \int_0^1 \int_0^1 e^{-0.2u} e^{-0.2v} (\cos 0.4u - 0.25 \sin 0.4u)$$

$$(\cos 0.4v - 0.25 \sin 0.4v)\ \delta(u-v)\ du\ dv$$

$$= 0.781 \tag{96}$$

Therefore H, for $\Delta t = 1$ second, is equal to

$$\begin{bmatrix} 0.278 & 0.392 \\ 0.392 & 0.781 \end{bmatrix} \tag{97}$$

and

$$\Phi(\Delta t) = \begin{bmatrix} 0.920 & 0.880 \\ -0.150 & 0.745 \end{bmatrix} \tag{98}$$

Simulations of Random Processes

In order to implement different filtering techniques by means of a digital computer, we must simulate random variables which correspond to samples taken from this random process at points Δt seconds apart. The process x is defined as the output of a first-order system with a specified transfer function. This system is being driven by white noise of a specified spectral amplitude. If the process is sampled at discrete intervals, the values x_k satisfy the difference equation

$$x_{k+1} = \Phi(t_{k+1};t_k)\, x_k + \int_{t_k}^{t_{k+1}} \Phi(t,\tau)Bu(\tau)\, d\tau. \qquad (99)$$

For a constant sampling interval Δt we have

$$x_{k+1} = \Phi(\Delta t)x_k + g_k. \qquad (100)$$

g_k is the response of the system to the white noise input between the times t_k and t_{k+1}. Also, $E[g_k g_{k+1}] = 0$. Therefore the g_k's are random variables which are independent, with mean zero if the mean of the input is zero. They are normally distributed with covariance matrix H.

In order to simulate values of x_k for the one-dimensional case, one random number is chosen from a population with variance X. For a second reading, this first value x_1 is multiplied by $\Phi(\Delta t)$ and a second random variable g_k, chosen from a population with variance H, is added to it. The same procedure is applied to x_2, and so on.

For values of uncorrelated measurement noise v, independent random variables are chosen from a population with the desired variance. The measurement y is then formed by the following operation: $y = Mx + v$. (101)

The process of forming two-dimensional vector random variables with correlated elements is somewhat more involved. The vector g_k is composed of two random variables with variances and covariances equal to the elements of the matrix H. To form the vector g_k, one must choose two independent random

variables from an available population and make a linear transformation on them.

It is known (10) that if a p x 1 vector y is distributed as a p-variate normal with mean μ and covariance V, and A is a q x p matrix of rank $q \leq p$, then the q x 1 vector $\underline{x}^* = A\underline{y}$ is a q-variate normal with mean $A\mu$ and covariance matrix AVA^T.

Let us assume we have available two sets of independent normally distributed random variables, y_1 and y_2, each with a desired variance. Some linear transformation T can be applied to the vector $\underline{y}$ to give correlated random variables with variance $TAT^T = H$. The matrix H as used here is a symmetric matrix. We can diagonalize this matrix by finding the eigenvalues and the corresponding eigenvectors. The transformation T in the above expression can be formed by these eigenvectors $\underline{x}_1$ and $\underline{x}_2$ as follows:

$$T = \begin{bmatrix} x_{11} & x_{21} \\ x_{12} & x_{22} \end{bmatrix} \tag{102}$$

This T matrix can then be used as the linear transformation which operates on the vector of independent random variables to form random variables with variance H. These independent random variables themselves must have variances equal to the eigenvalues λ_1 and λ_2 of the H matrix. If, however, we have two independent normally distributed random variables with unity variance, we must transform these variables by

$$W = \begin{bmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{bmatrix} \tag{103}$$

in order to obtain random variables with the desired variance $TWIW^T T^T = H$. Therefore the total transformation on the independent random variables is the product of the transformations

$$TW = \begin{bmatrix} x_{11}\sqrt{\lambda_1} & x_{21}\sqrt{\lambda_2} \\ x_{12}\sqrt{\lambda_1} & x_{22}\sqrt{\lambda_2} \end{bmatrix} \tag{104}$$

This same derivation can also be generalized to the case of a vector of n dependent random variables.

The two-dimensional random variables for the simulations here were then formed according to the following relationship:

$$\underline{x}(k + 1) = \Phi(\Delta t)\ \underline{x}(k) + TW\underline{y}. \tag{105}$$

Here $\underline{y}$ is a two-vector of independent normally distributed random variables with unity variance and zero mean.

In all the simulations made, the random variables were formed from random normal numbers with mean 0 and variance one selected from tables originally published by the Rand Corporation (12) and reprinted in Dixon and Massey (3).

A computer subroutine for generating normal random numbers with a specified mean and variance was also available. This subroutine utilized a second subroutine which generated uniformly distributed random numbers. A normally distributed random number was formed by adding a finite number of

uniformly distributed numbers together, taking the mean, and scaling. These subroutines resulted in the use of somewhat more computer time than was needed to read in numbers selected from tables and store them. Therefore this method was not used. If a large amount of statistical data is needed, however, the use of these or similar subroutines is recommended in order to save the labor involved in selecting random numbers from tables and punching these onto cards.

By using random numbers punched onto cards, it was possible to make a large number of runs using the same set of data, but with modifications of the system parameters. If more than one set of data was desired for the same set of conditions, the cards could be reordered, or shuffled, for a different random process.

It was desired to make comparisons of different systems and adaptive techniques to obtain information on their relative merits, rather than to obtain statistical data on the variance of each individual case which could be relied upon to within a certain confidence interval. Such data would be valuable, but would require a large number of computations for each case considered.

ADAPTIVE FILTERS WITH TIME-CHANGING MEASUREMENT ERROR

In the cases considered here we will assume that all statistical properties of the system are known except for the variance of the measurement error. It is noted that if this variance is known completely, even though it is time-changing, we have all the information necessary for the optimal weighting of estimates in the Kalman filter.

If, at $t = 0$, no information about the variance V of the measurement error is known, the procedure used here combines a statistical estimation of the variance parameter with the Kalman filter equations to estimate the state at $t = t_k$.

If V is time-stationary, obviously the best way to estimate the state is to carry along all the estimates of V in an average. As $t_k - t_o$ grows large, an unbiased estimate of V will converge to V; and P_a, the actual variance of the state estimate, will converge to the final value of P(optimal). If, however, V is apt to change with time, this method may not detect this change in V at all.

If V changes with time, the parameter estimation technique must be such that the Kalman filter is adaptive to these changes in V. Two techniques are used here. One involves carrying along a finite number of estimates of V and forming an averaged estimate V. The second involves a form of exponential weighting of the past data.

The truncated average of V is formed as follows:

$$\overline{V}_i = \frac{1}{n} \sum_{i=1}^{n} \hat{V}_i \qquad (106)$$

The exponentially weighted average chosen is the simplest of the possible forms of the exponentially weighted averages which may be used:

$$\overline{V}_i = e^{-\frac{1}{p}}(\overline{V}_{i-1}) + (1 - e^{-\frac{1}{p}})\hat{V}_i. \qquad (107)$$

For purposes of comparison, p was chosen such that $(1 - e^{-\frac{1}{p}})$ was equal to 1/n in the truncated estimate. The two estimates then seemed to be comparable in performance.

In general, the truncated estimate seemed to converge faster and to be more adaptable to changes in V. The exponential average seemed to give better estimates in the steady-state case for a particular n. One particular advantage of the exponential average is that it is less susceptible to sudden irregularities in the noise which sometimes cause truncated estimates of V to be highly erratic and inaccurate.

The truncated average was modified by using different values of n; that is, by using a large or small amount of past data in forming an estimate for V. The exponentially weighted average was modified by changing the value of $e^{-\frac{1}{p}}$.

As stated before, it is desired to estimate the state of a system with a message process whose statistical properties are known. The measurements of the output, however, are

contaminated by independent measurement errors with zero mean and unknown variance. The variance of this measurement error will change rapidly at some unknown instant of time. It may be desired to "design" a filter which will give the minimum mean squared estimation error over a finite time interval from the instant of this change in V to the present time t_f. This design consists of choosing the optimum number of measurements to weight in estimating V.

In the general case, a compromise must be made here. If V were time-stationary, the optimal filter would involve averaging all available estimates together to estimate V. However, after V makes a step change in value, there is an interval of time during which the mean of the averaged estimate of V is not equal to V. This interval is n seconds long and is described by n-1 estimates $\overline{\hat{V}}$, where n is the number of estimates $\hat{V}$ used to form the average. This leads to the estimation-error variance during this transient interval being larger than the steady-state value and larger yet than the optimal value of P_a. Therefore a compromise must be made between two extremes. A large value of n causes a small steady-state error variance, although values are similar for different large values of n. At the same time we have the disadvantages of a longer interval of adaptation and a larger peak error variance during this interval. A small n leads to a relatively large and uneven steady-state error with the

advantages of a short interval of adaptation and usually a smaller error during this interval.

We may also think of V as periodically switching between two values, as shown in Figure 7. With the adaptive filters described here, P_a for this situation will appear as in Figure 8.

In the periodic case, one possible problem is that of finding the value of n to be used to form $\overline{V}$ such that the mean squared error of the estimate over one cycle is a minimum.

In order to obtain some idea of the optimal value of n for a certain length of time between switching, or a certain period, several computer simulations were made. These would enable one to determine average values of P, both during the time in which V is stationary, and during the transient period. The length of the transient period is taken to be n time intervals, described by (n - 1) estimates.

One-State System

For an example situation, V was postulated to switch from 0.5 to 4.0 and back to 0.5. Four computer runs of 80 pieces of data each were made using values of n of 5, 8, 10, and 15. The averaged results of these runs are given in Figures 9 and 10. Values of 7, 9, 12, 20, 25, and 30 were also used for n on some computer runs, as were different values of V.

It would seem that larger numbers of n should be used for longer time periods, because with the large n the estimate is

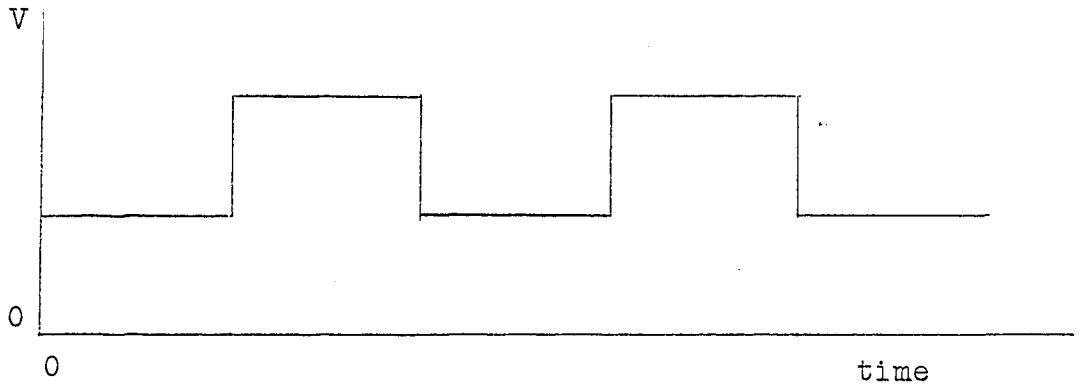


Figure 7. V periodically switching between two values

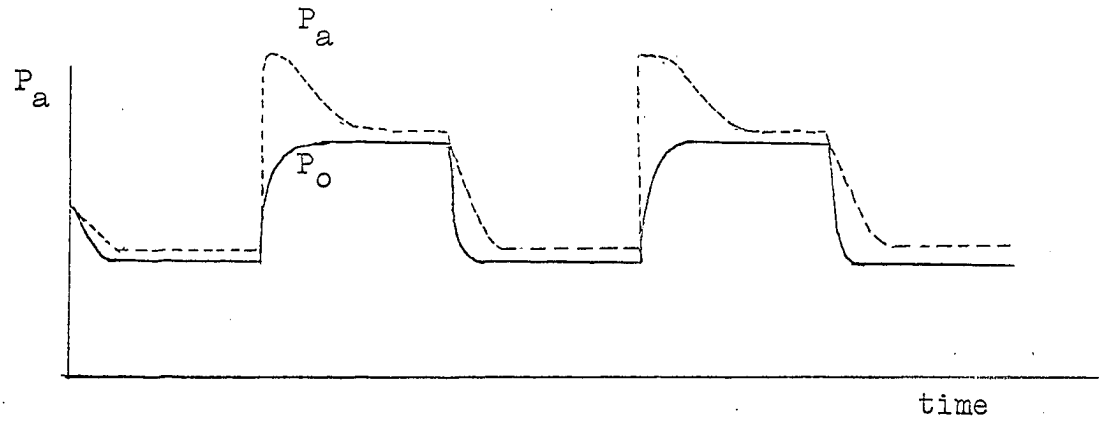


Figure 8. P_a and P_o obtained with an adaptive filter applied to the situation shown in Figure 7

slow to adapt but has slightly lower and more even steady-state error.

Two-State Systems

Several two-state systems were also simulated. Here, if one is interested in the total state estimation error variance, he must consider the trace of the P-matrix as well as the individual elements. The trace is defined as the sum of the elements along the principal diagonal; it is $P_{11} + P_{22}$ in the two-state case.

In the two-state systems considered here, a step change in V causes a large change in P_{11} and a relatively small steady-state change in P_{22}. However, pronounced transient characteristics are discernible in both P_{11} and P_{22}. In the two-state case, a larger number of estimates must be averaged together in order for the estimate of V to converge. This was especially noticeable in the system described in Figure 6, where the damping ratio was small.

The behavior of the different systems which were simulated is best described by the graphs which are included and described in the following section.

Description of Graphs

Graphs of the results of some of the computer runs are presented here for illustrative purposes. Figures 9 through 14 pertain to the one-state system.

Figures 9 and 10 give the averaged results of four similar computer runs. In Figure 9 P_a for n equal to 5 and 8 are compared with the variance of the error of the optimal estimate. Similar results are given in Figure 10 for n = 10 and n = 15. V is equal to 0.5 for the first 30 state estimates, and it then makes a step change to 4.0 for the remaining 50 estimates. Note that the values of P for each n are markedly different during the time interval immediately after the change in V. The steady-state variance appears to be erratic for n = 5, but is not markedly different in value for n = 8, 10, and 15.

Figures 11 and 12 compare a 5-measurement truncated average and an equivalent exponentially weighted average. In Figure 11, V goes from 0.5 to 1.0 and back to 0.5. In Figure 12, V goes from 0.5 to 4.0 and back to 0.5. Here it can be seen that the truncated estimator adapts slightly faster, but the exponentially averaged estimator is "smoother" in that peak values of P are not as high when V is time-stationary.

Figure 13 shows the convergence of P_a for the one-state system. The estimate of V is formed by averaging all available estimates together up to time t_k. Results are shown for V = 0.5 and V = 1.0.

Figure 14 shows the results of truncated averaged estimators of V where n = 5 and n = 8. V goes from 0.5 to 1.0 and back to 0.5.

Figures 15 through 25 apply to various two-state systems.

Figure 15 shows P_{11} for one run for the two-state system shown in Figure 3, for n equal to 8, 15, and 25. Notice that the estimator for n = 8 is much more erratic than the other two, when V is time-stationary. A "spike" also occurs here after the change in V.

Figure 16 shows P_{22} for the same run for the same system. P_{22} is reasonably close to the optimum value, except for a period of approximately 5 seconds after the change in V. If the sum of the elements along the major diagonal of the covariance matrix is used as a measure of the quality of the total state estimate, that estimate is especially poor during this time interval.

Figure 17 shows P_{11} for the two-state system shown in Figure 5. Again V is estimated by a truncated average with n equal to 8, 15, and 25. V switches from 0.5 to 4.0 at t = 30.

Figure 18 shows P_{22} for the same situation. Note that the results obtained with this system, whose transfer function has imaginary roots, do not appear to be markedly different from those obtained with the system of Figures 15 and 16, whose transfer function has real roots.

Figure 19 again shows P_{11} for the two-state system of Figure 5. Again n is equal to 8, 15, and 25. V goes from 4.0 to 0.5.

Figure 20 shows P_{11} and P_{22} for the same two-state system for V going from 0.5 to 1.0. The averaged estimate is formed exponentially. The estimate formed is approximately equivalent to 8 averaged estimates.

Figure 21 shows P_{11} and P_{22} when a truncated estimate, with n equal to 20, is used.

Figures 22 through 25 pertain to the system of Figure 6.

Figure 22 shows P_{11} for a truncated and an exponentially weighted estimate when V goes from 0.5 to 4.0.

Figure 23 shows the convergence of P_{11} and P_{22} when all available estimates are averaged together.

Figure 24 shows the results of a truncated and an exponentially weighted estimate when V switches from 0.5 to 1.0. Note that the estimation-error variance of this system is higher and more irregular than that obtained with the other two-state systems used.

Figure 25 compares values of y for similar runs made with the systems of Figures 5 and 6. Note that the peak values of y are higher for the system of Figure 6. This would appear to help explain the fact that the adaptive filter used here is more difficult to apply to this system.

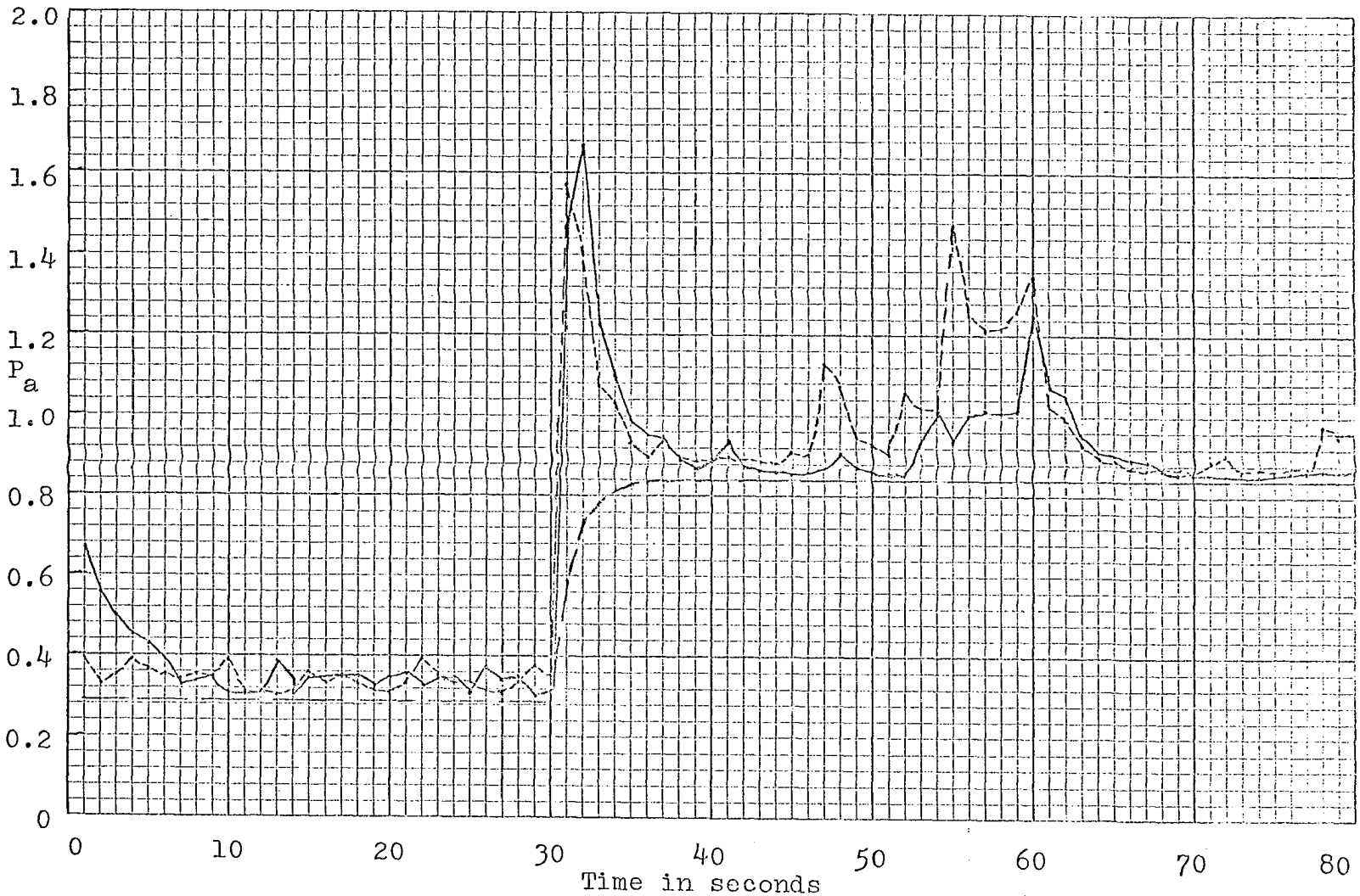


Figure 9. Averaged results of four computer runs for one-state system
--- n=5 —— n=8 — — optimal

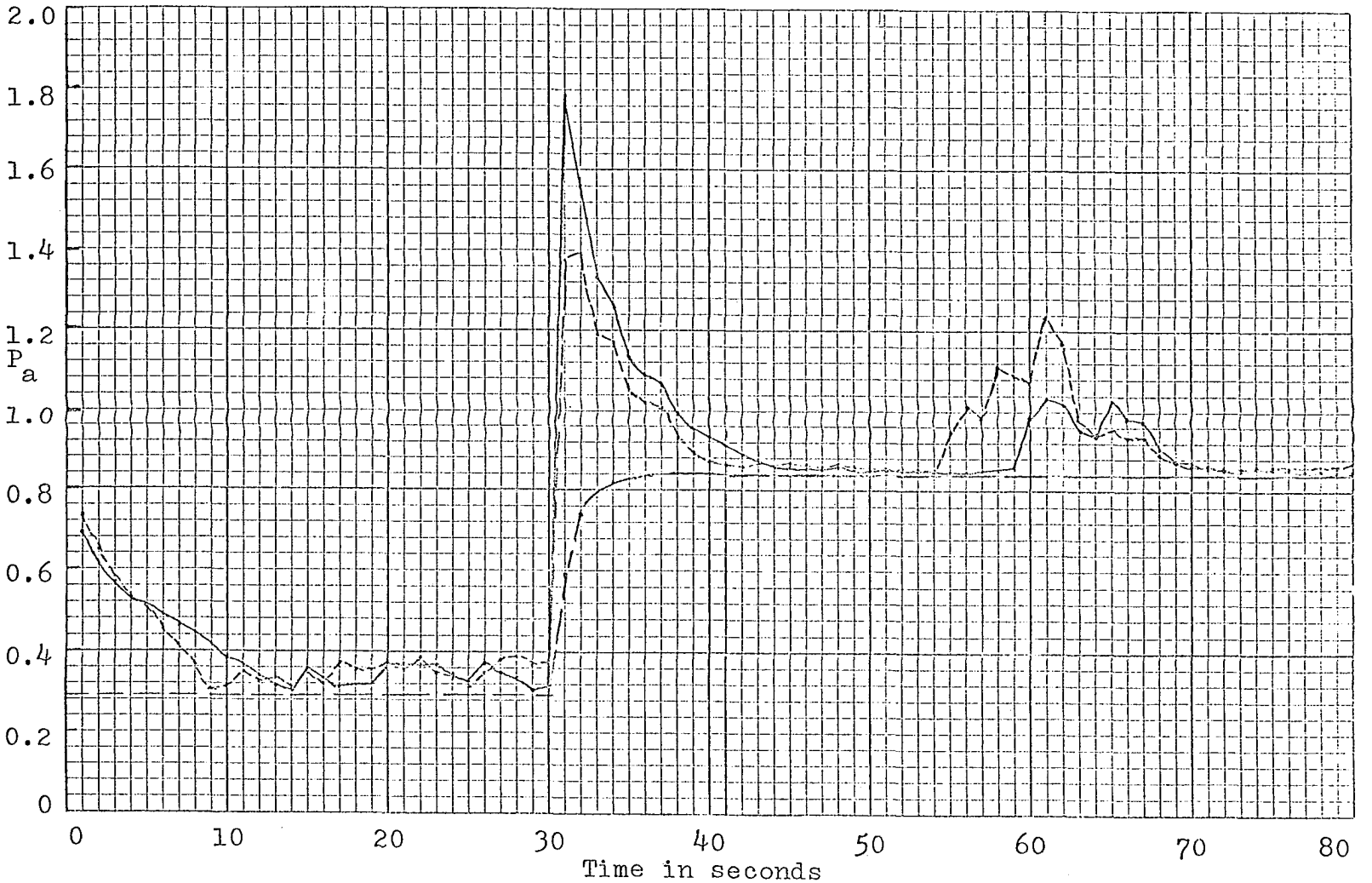


Figure 10. Averaged results of four computer runs for one-state system
--- n=10 ——— n=15 — — optimal

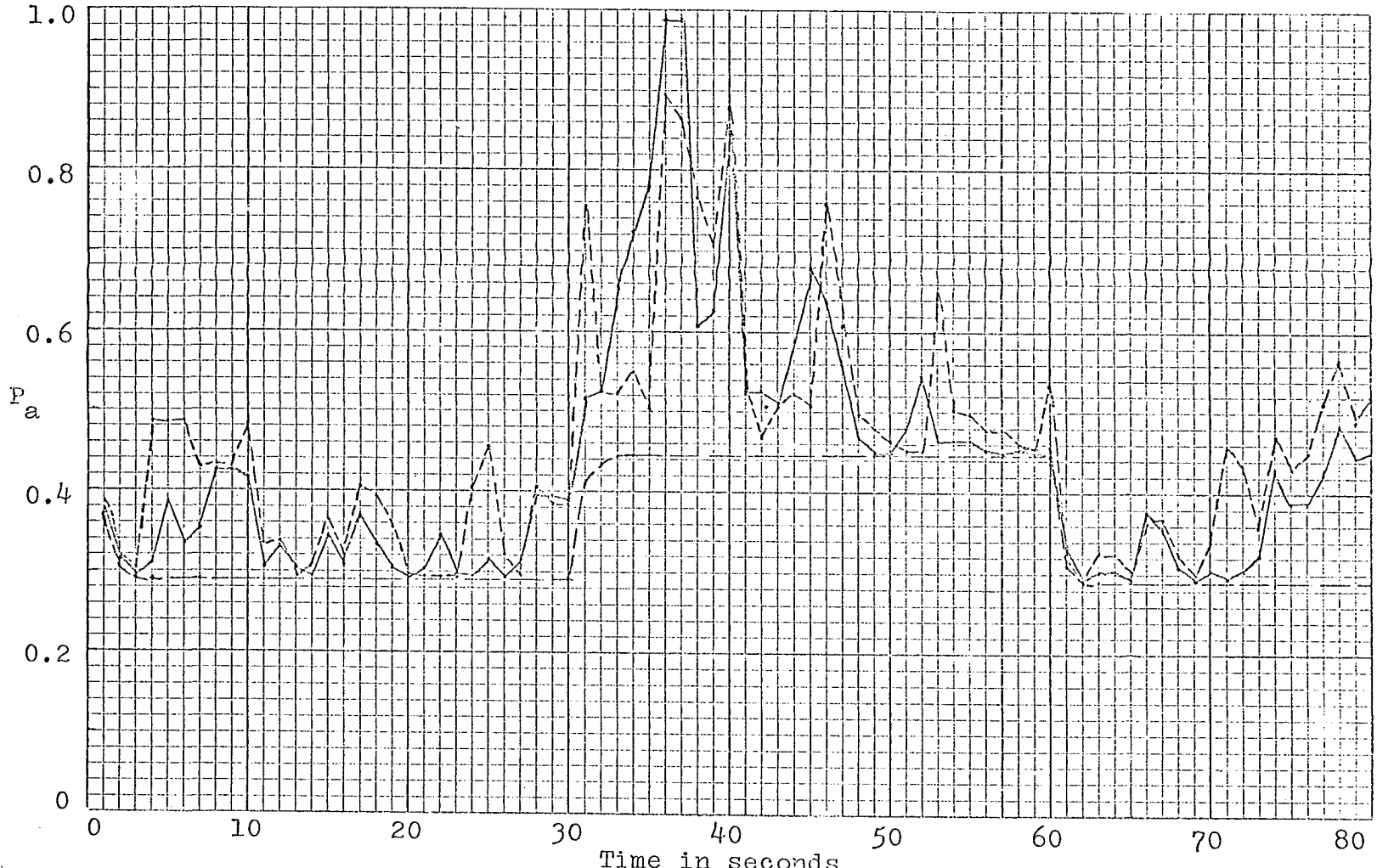
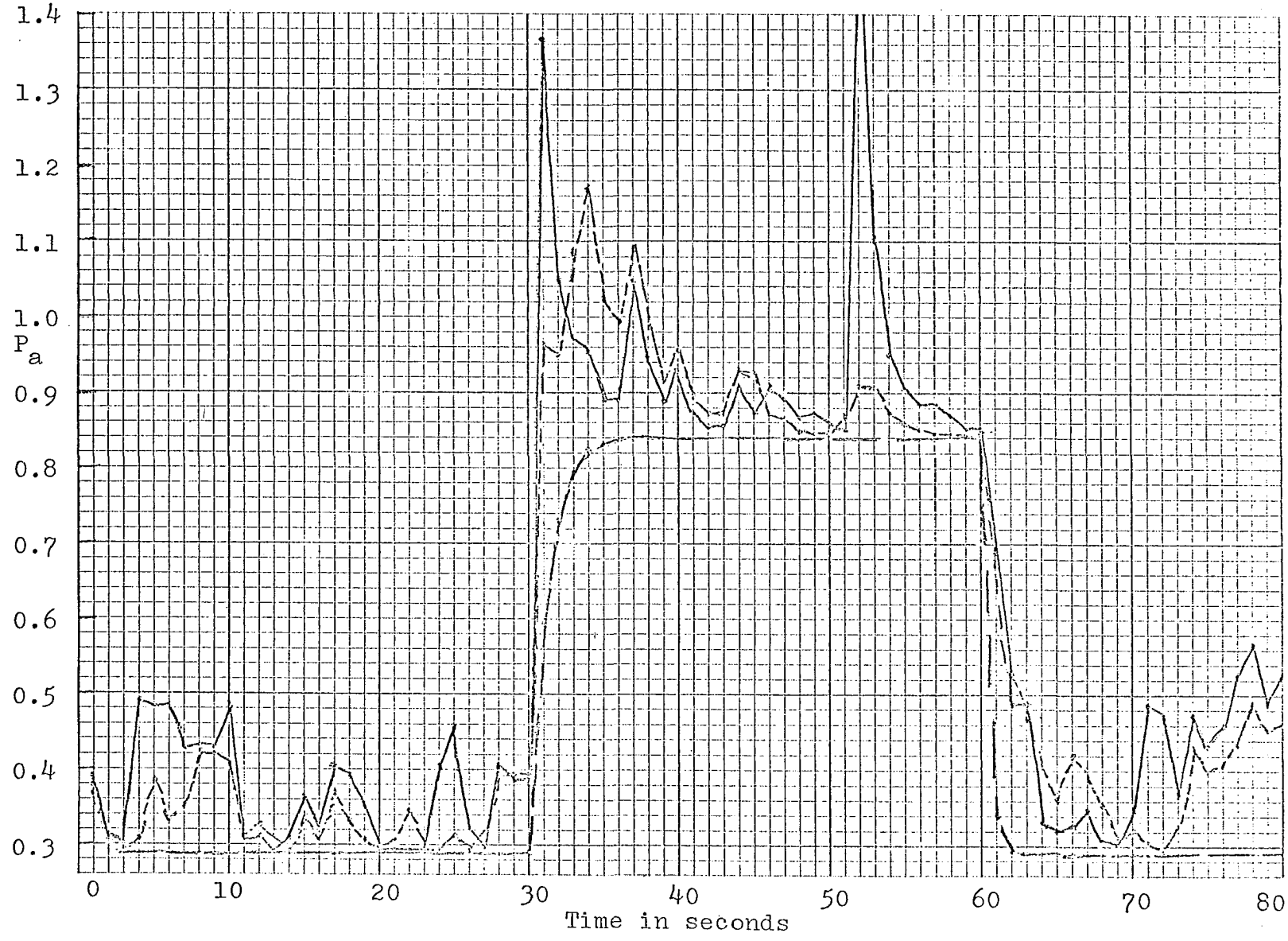


Figure 11. Comparison of truncated and exponentially weighted averages
--- truncated average —— exponential average —— optimal case

Figure 12. Comparison of truncated and exponentially weighted averages

——— truncated --- exponentially weighted
— — optimal


1.4
1.3
1.2
1.1
1.0
P_a
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
10
20
30
40
50
60
70
80
Time in seconds

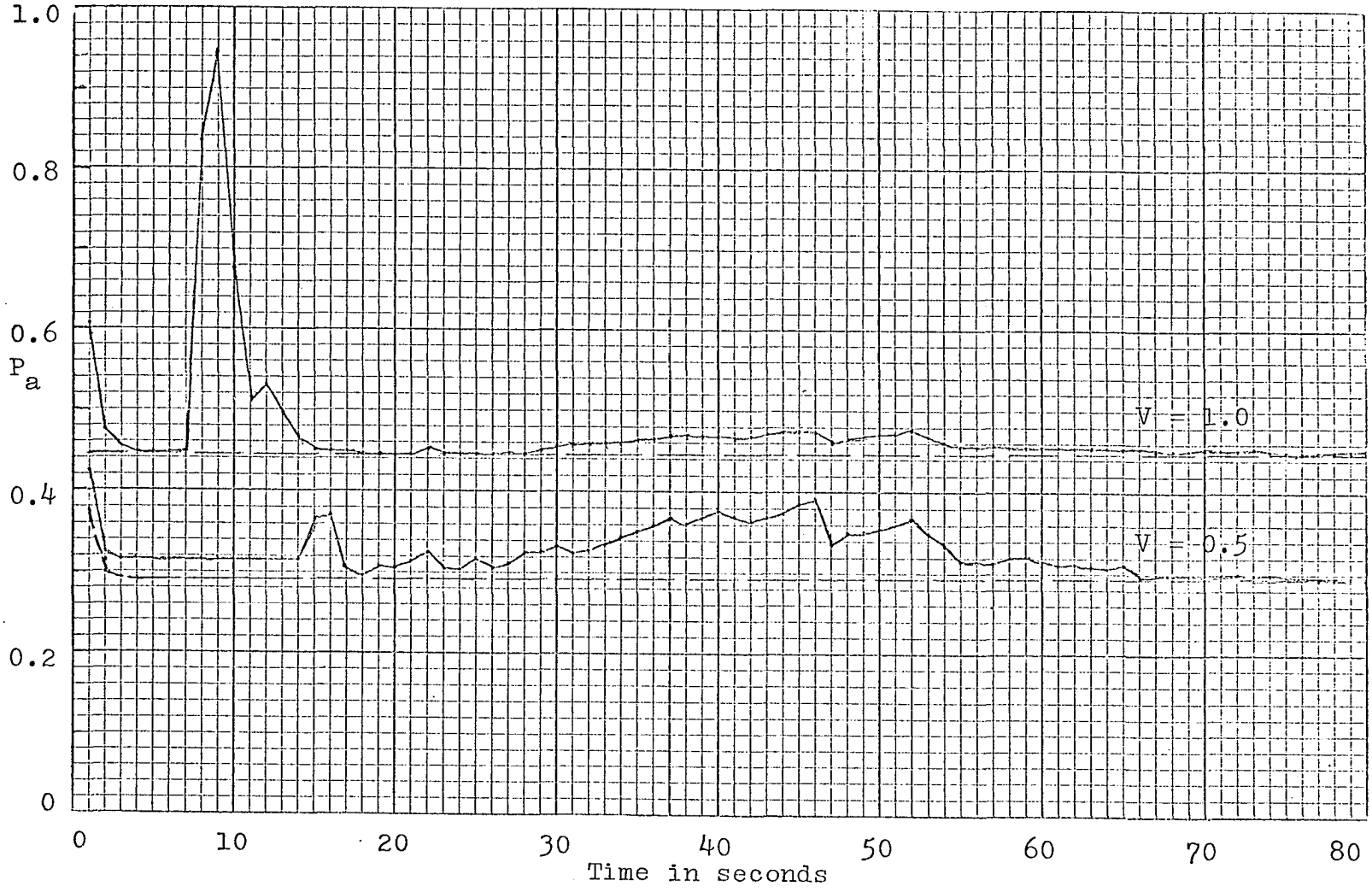


Figure 13. Convergence of P_a to P_o for a one-state system

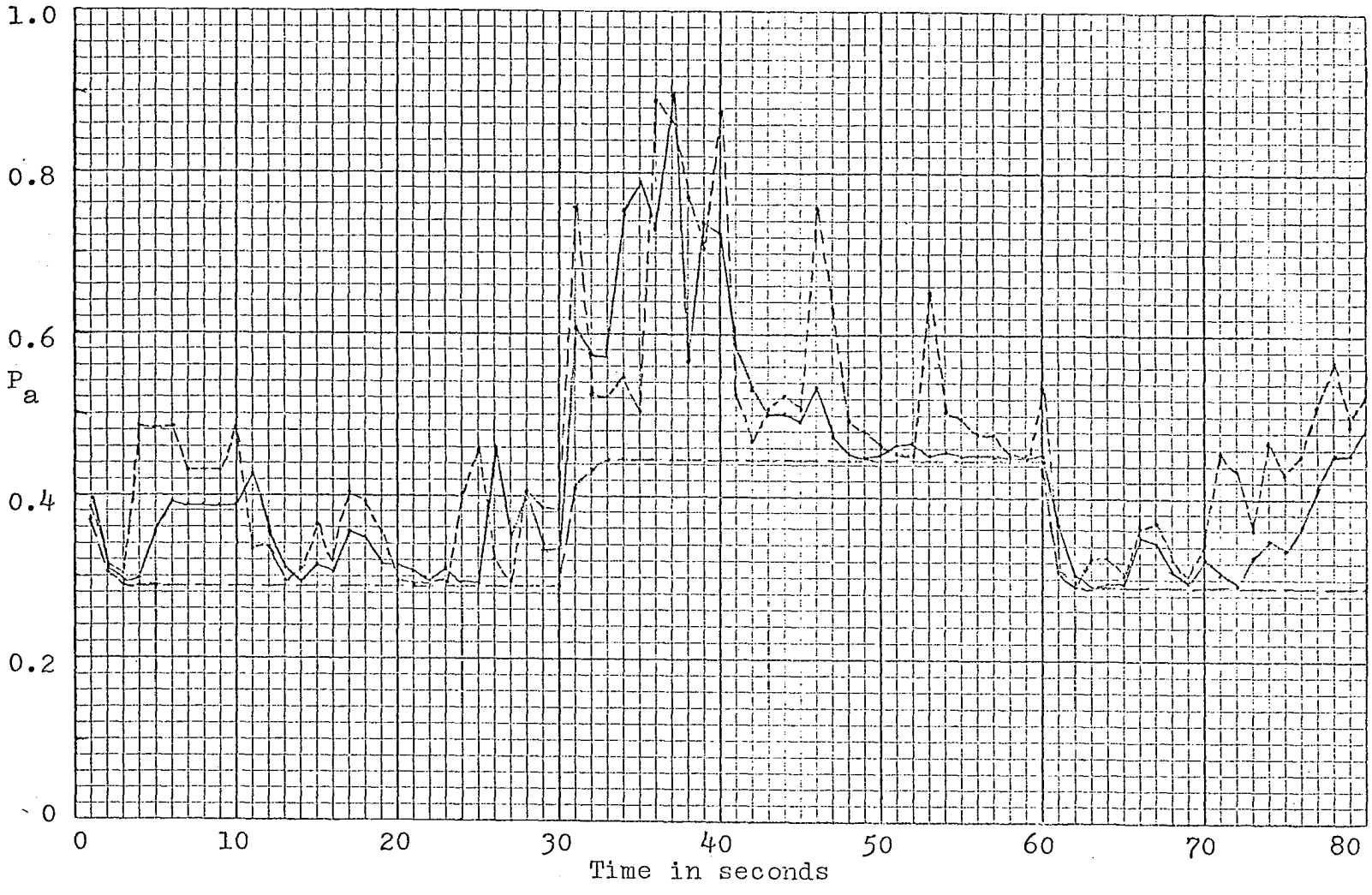


Figure 14. Results of two truncated estimators
--- n=5 ——— n=8 — — optimal

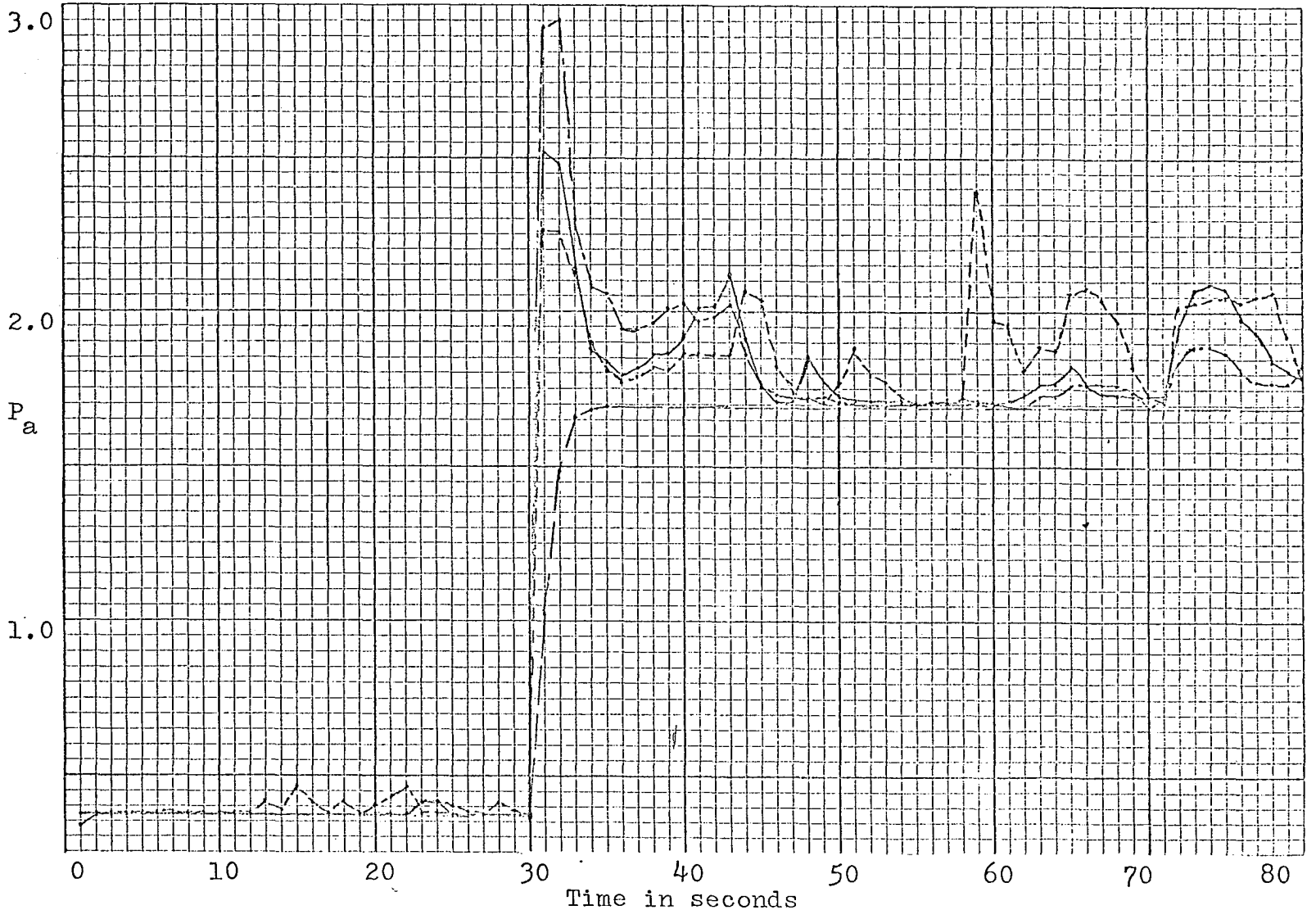


Figure 15. P_{11} for two-state system of Figure 3
--- n=8 ——— n=15 — - — - n=25 — — optimal

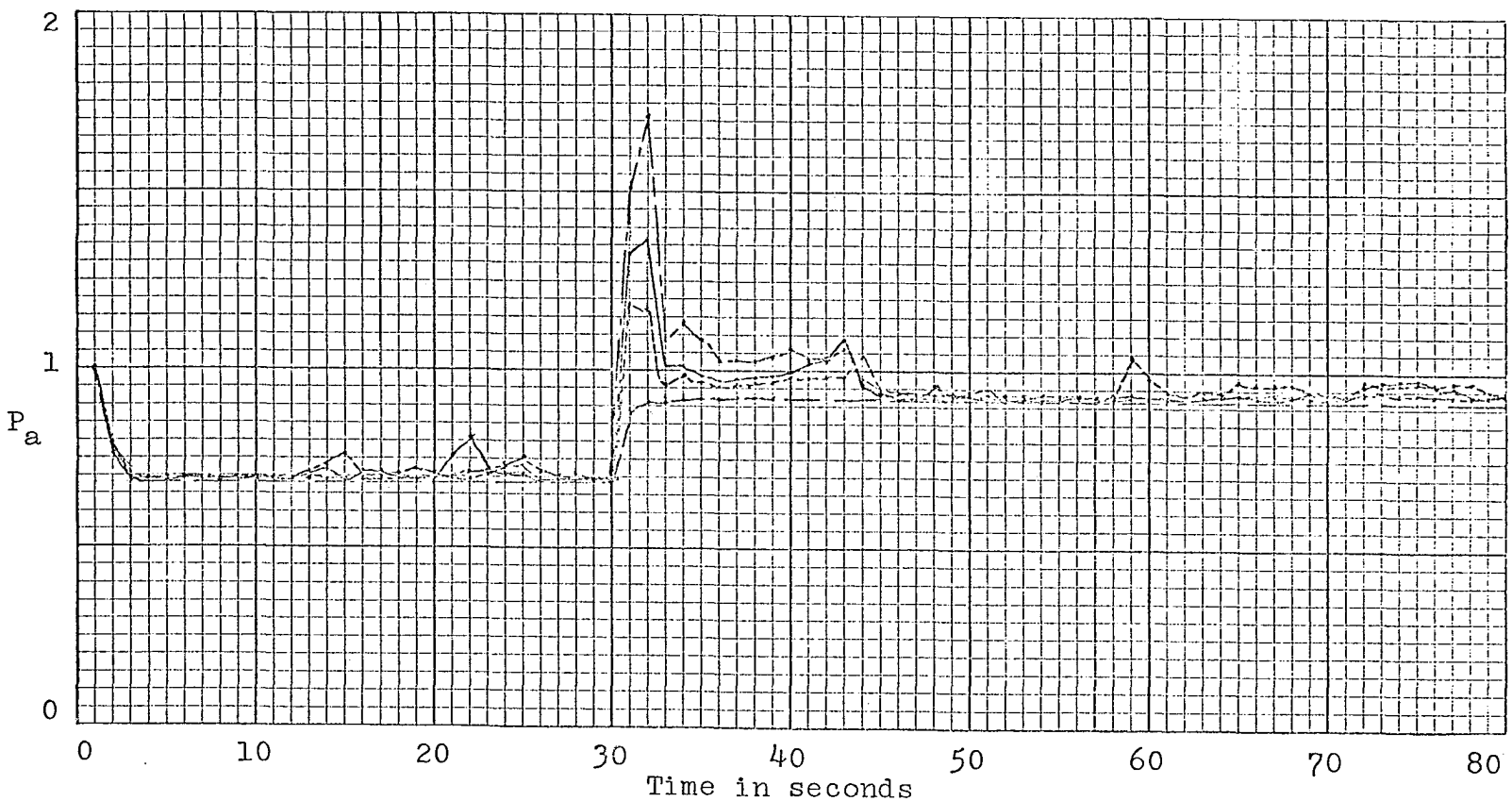


Figure 16. P_{22} for two-state system of Figure 3
--- n=8 ——— n=15 — - — n=25 — — optimal

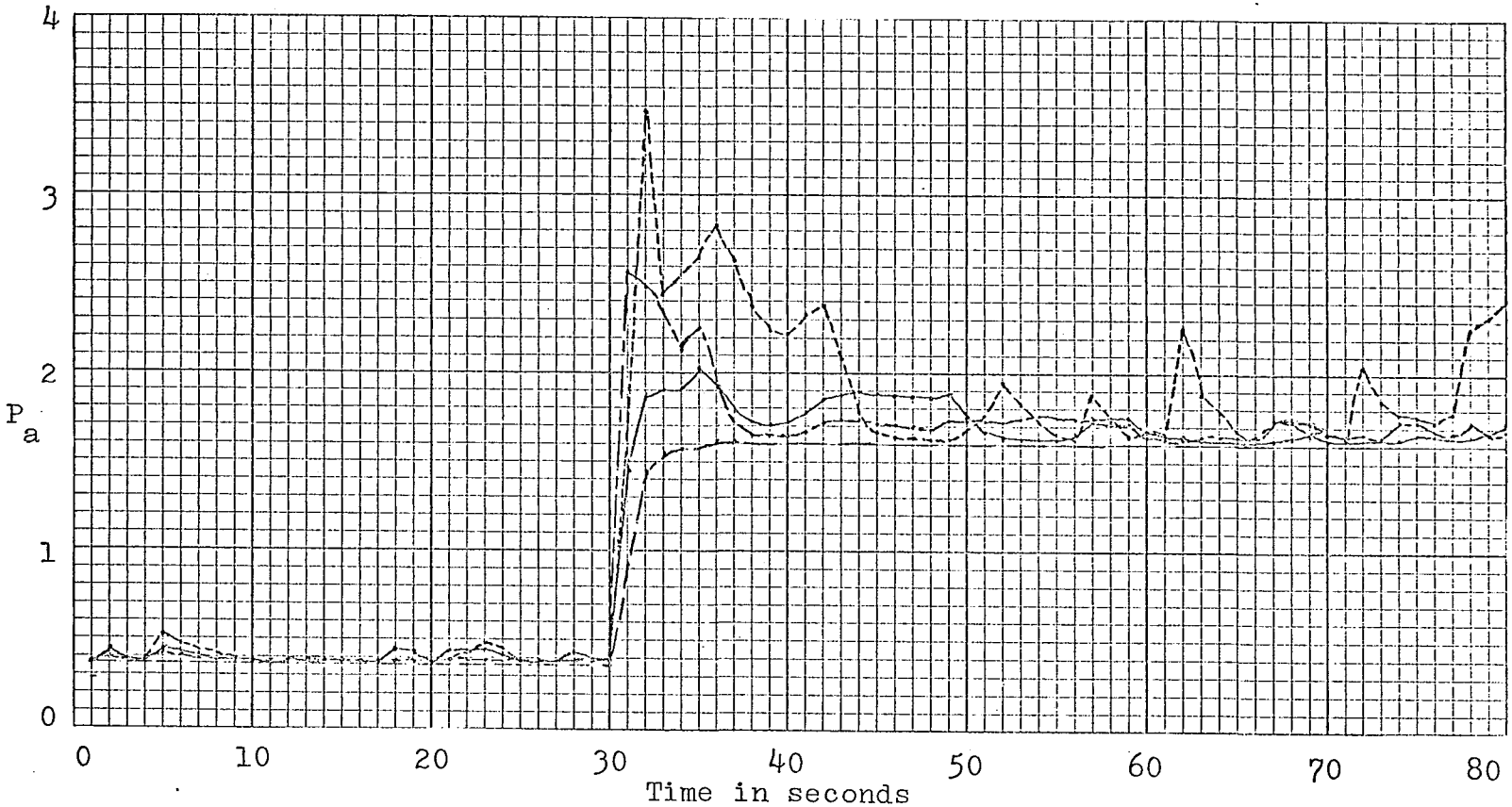


Figure 17. P_{11} for the two-state system of Figure 5
--- n=8 ——— n=15 —·— n=25 — — optimal

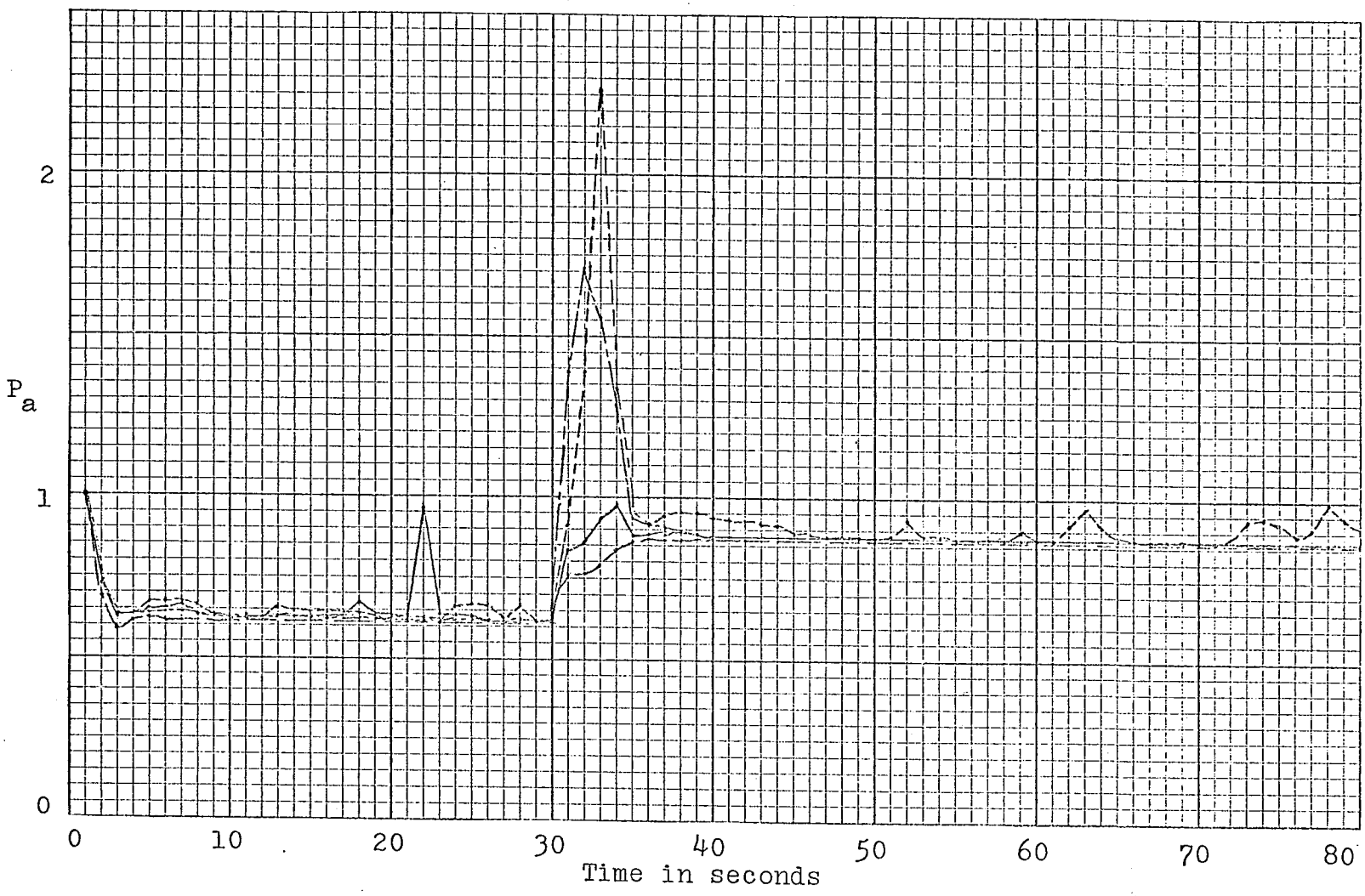


Figure 18. P_{22} for two-state system of Figure 5

--- n=8 —— n=15 —·— n=25 — — optimal

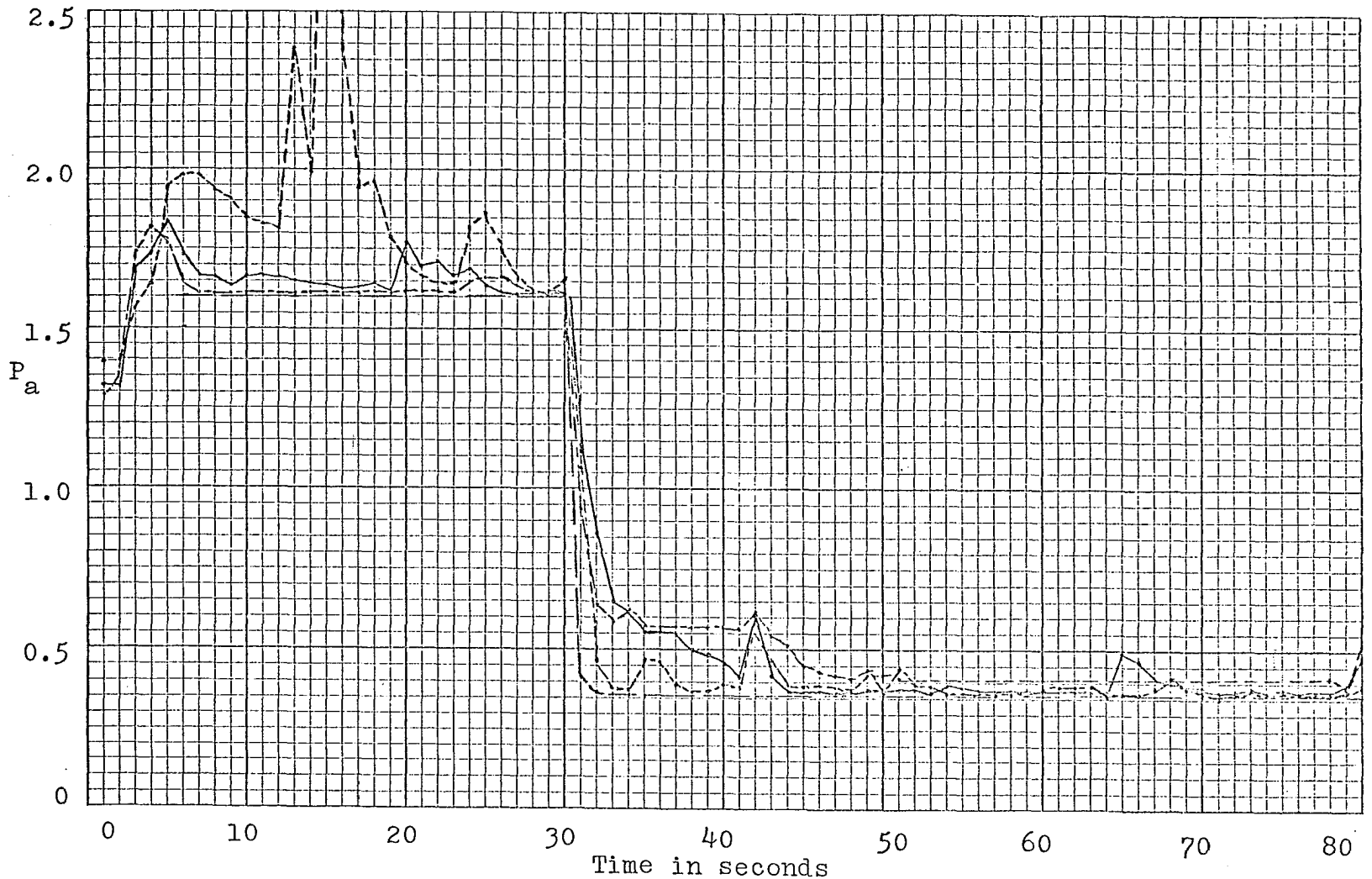


Figure 19. P_{11} for two-state system of Figure 5

--- n=8 ——— n=15 — - — n=25 — — optimal

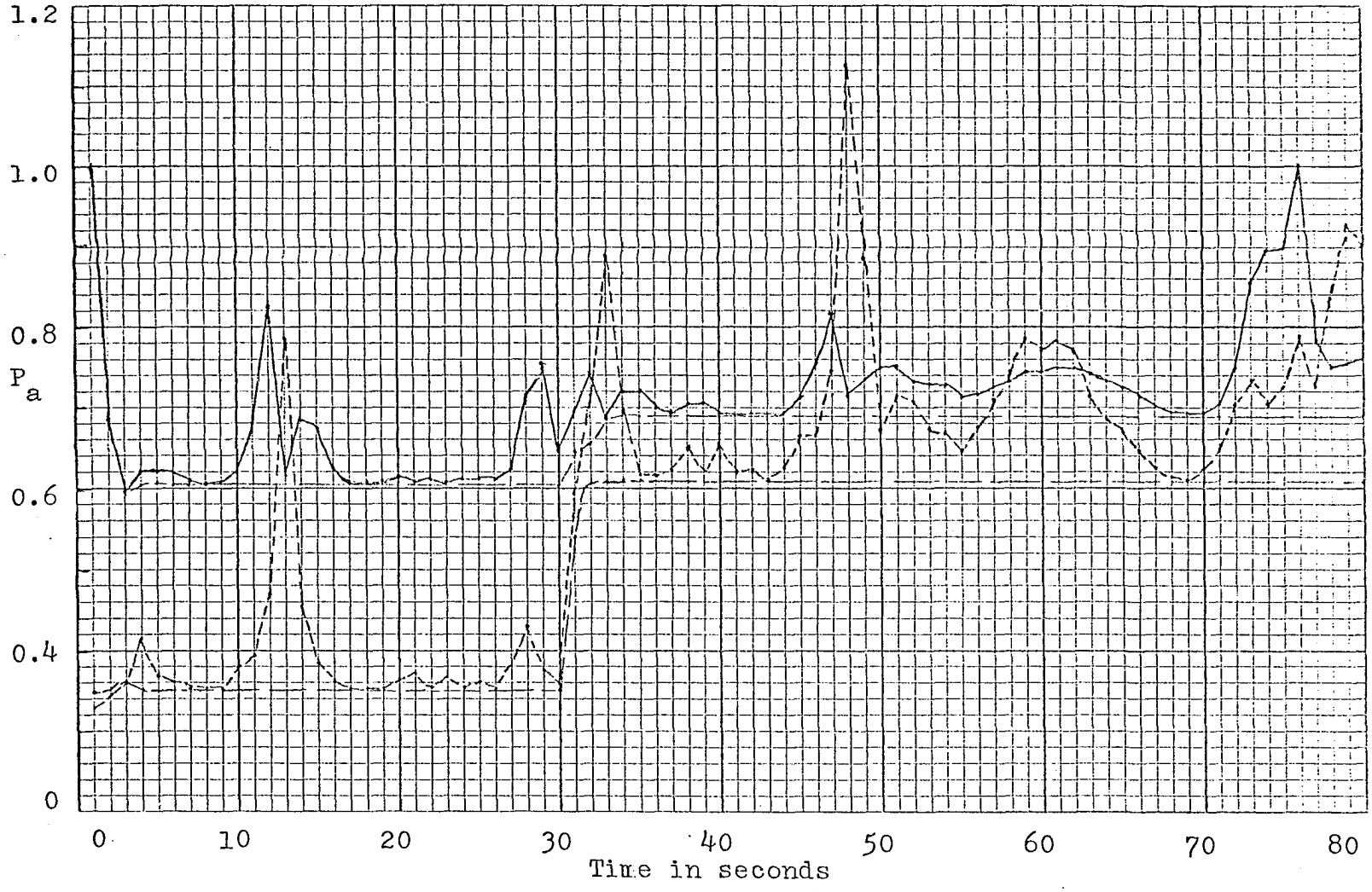


Figure 20. P_{11} and P_{22} for system of Figure 5

--- P_{11} —— P_{22} — — optimal values

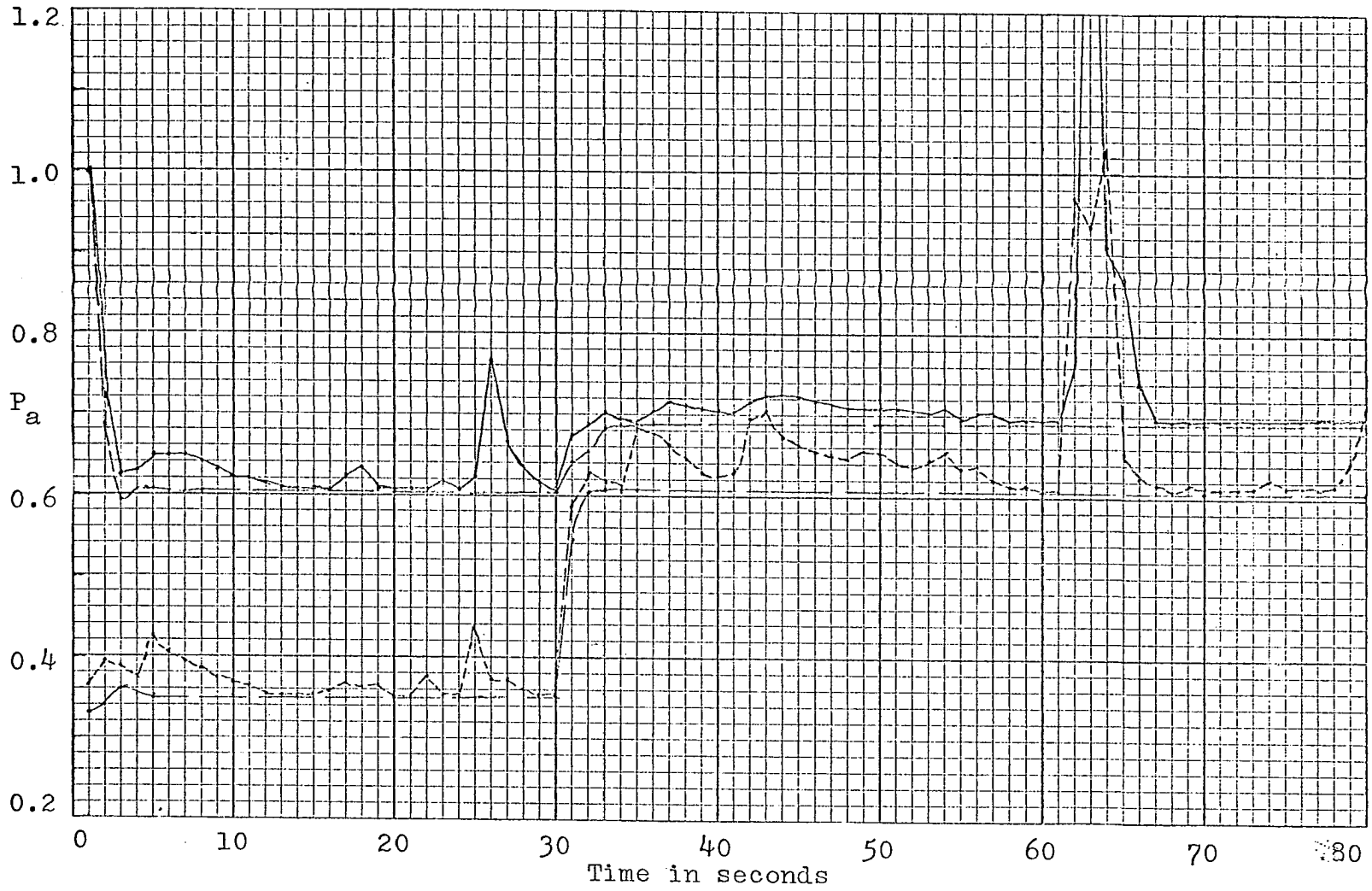


Figure 21. P_{11} and P_{22} for system of Figure 5
--- P_{11} —— P_{22} — — optimal values

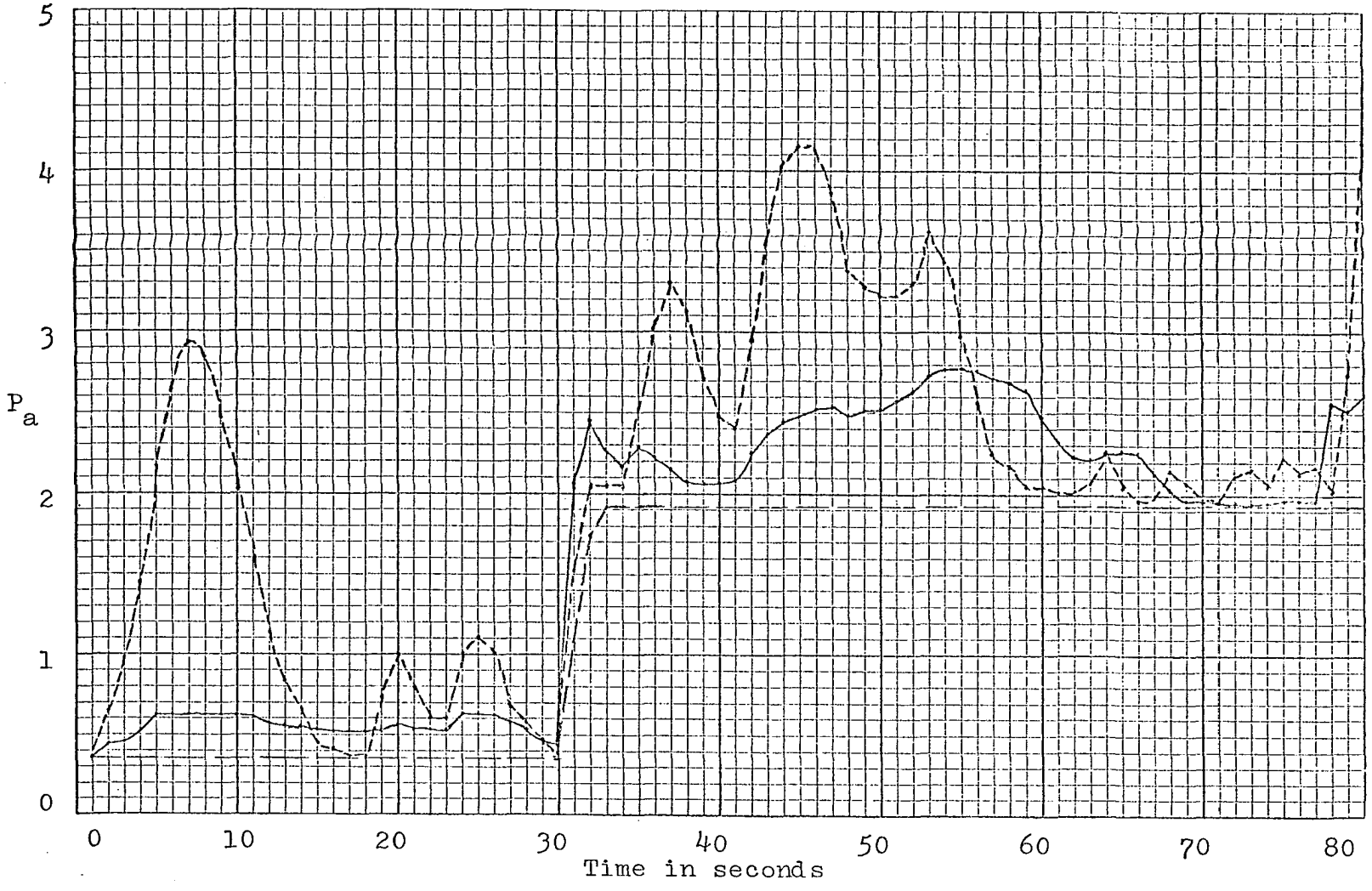


Figure 22. P_{11} for system of Figure 6

--- 5 exponentially weighted estimates — — optimum
—— 25 truncated estimates

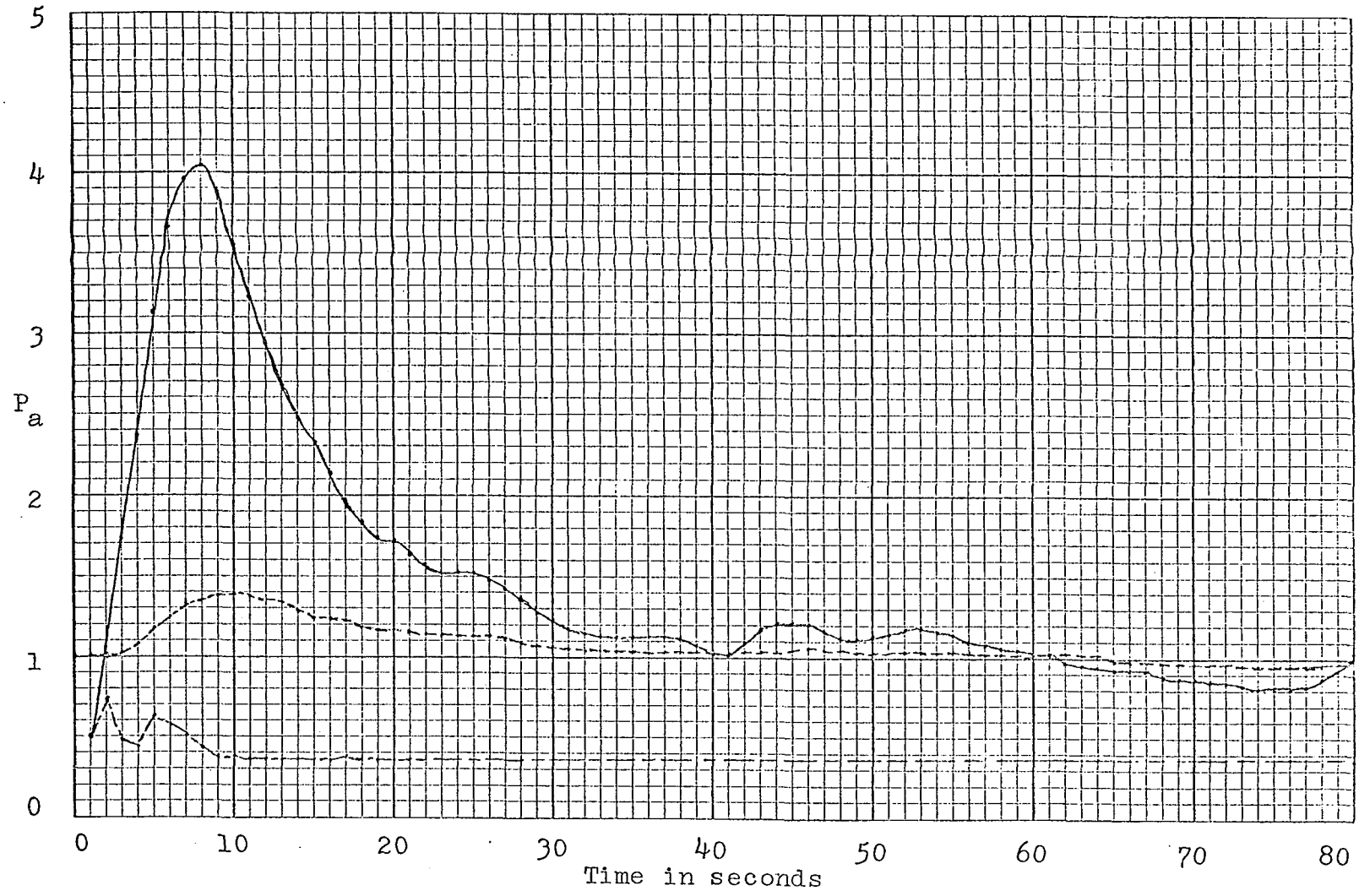


Figure 23. Convergence of P_{11} and P_{22} for systems of Figures 5 and 6
—— P_{11} for system of Figure 6 --- P_{22} for system of Figure 6
— - — P_{11} for system of Figure 5

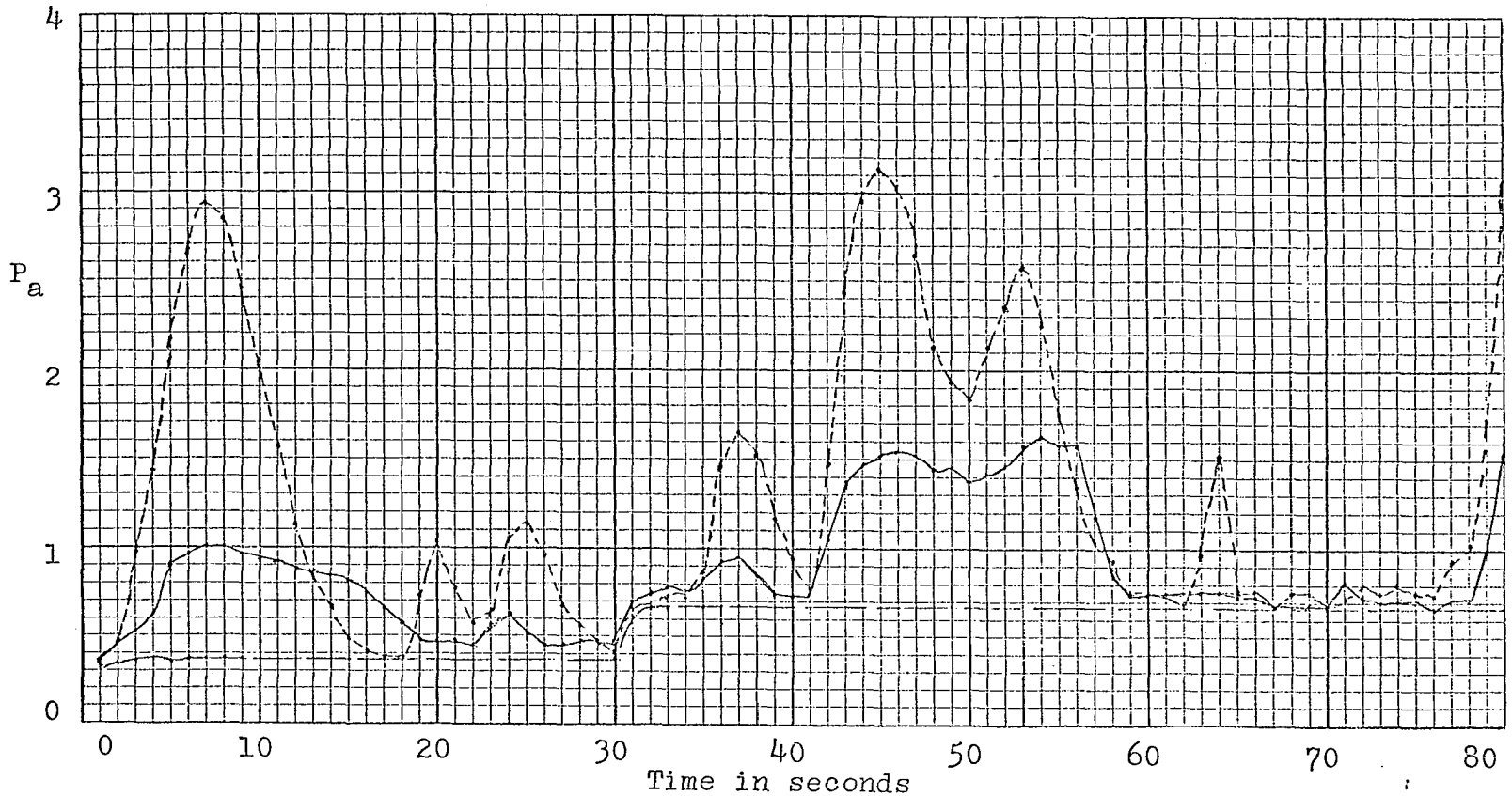


Figure 24. Results of two estimators for system of Figure 6
--- n=5 exp. wtd. —— n=15 truncated — — optimal

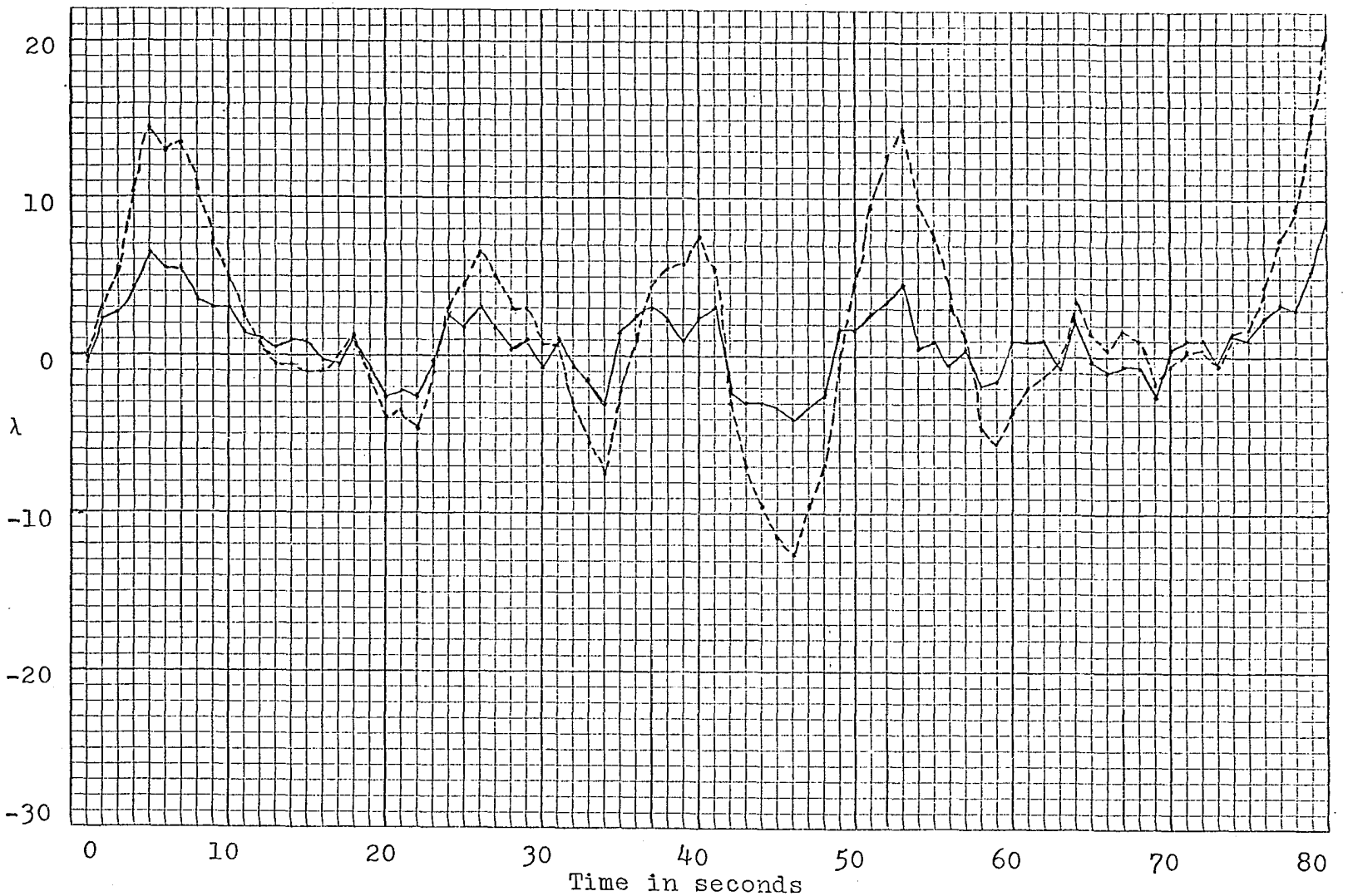


Figure 25. Comparative outputs of systems of Figures 5 and 6
—— system of Figure 5 --- system of Figure 6

CONCLUSIONS

It is concluded that, of the two possibilities considered, the MCP estimator is the most useful estimator of V. With a suitable choice of n it enables one to make state estimates with an error variance consistently within 20 per cent or less of those obtained when all statistical properties of the system are known.

Some computer simulations of actual system situations are carried out. The data from these simulations illustrate some of the properties of the adaptive filter when the variance of the measurement error is time-changing. Simple one- and two-state systems are used. However, the methods used here can be extended to the study of more complex systems.

With two-state systems, more problems exist in using an adaptive filter. The state estimate is sometimes erratic, with high peaks in the estimation error. This seems to be especially evident when the damping ratio is small and the system transfer function has complex roots. However, it appears that satisfactory results can be achieved in some cases.

LITERATURE CITED

1. Battin, R. H. A statistical optimizing navigation procedure for space flight. American Rocket Society Journal 32: 1681-1696. 1962.

2. Bucy, R. S. and Follin, J. W., Jr. Adaptive finite time filtering. Institute of Radio Engineers Transactions on Automatic Control 7: 10-20. 1962.

3. Dixon, Wilfrid J. and Massey, Frank J., Jr. Introduction to statistical analysis. New York, N.Y., McGraw-Hill Book Company, Inc. 1957.

4. Fagin, S. L. Recursive linear regression theory, optimal filter theory, and error analyses of optimal systems. Institute of Electrical and Electronics Engineers Convention Record 12, Part 1: 216-240. 1964.

5. Heffes, H. The effect of erroneous models on the Kalman filter response. Institute of Electrical and Electronics Engineers Transactions on Automatic Control 11: 541-542. 1966.

6. Kalman, R. E. A new approach to linear filtering and prediction problems. American Society of Mechanical Engineers Transactions, Series D, Journal of Basic Engineering 82: 35-45. 1960.

7. Kalman, R. E. New methods in Wiener filtering theory. In Bogdanoff, John L. and Kozin, Frank, eds. Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability. pp. 270-388. New York, N.Y., John Wiley and Sons, Inc. 1963.

8. Lee, Robert C. K. Optimal estimation, identification, and control. Massachusetts Institute of Technology Research Monograph 28. 1964.

9. Magill, D. T. Optimal adaptive estimation of sampled stochastic processes. Institute of Electrical and Electronics Engineers Transactions on Automatic Control 10: 434-439. 1965.

10. Mood, A. M. and Graybill, F. A. Introduction to the theory of statistics. New York, N.Y., McGraw-Hill Book Company, Inc. 1963.

11. Nishimura, T. On the a priori information in sequential estimation problems. Institute of Electrical and Electronics Engineers Transactions on Automatic Control 11: 197-204. 1966.

12. Rand Corporation. A million random digits with 100,000 normal deviates. Glencoe, Ill., Free Press. 1955.

13. Shellenbarger, J. C. Estimation of covariance parameters for an adaptive Kalman filter. Unpublished Ph.D. thesis. Ames, Iowa, Library, Iowa State University of Science and Technology. 1966.

14. Swerling, Peter. Topics in generalized least-squares signal estimation. Society for Industrial and Applied Mathematics Journal on Applied Mathematics 14: 998-1032. 1966.

15. Tou, J. T. Modern control theory. New York, N.Y., McGraw-Hill Book Company, Inc. 1964.

16. Wiener, Norbert. Extrapolation, interpolation, and smoothing of stationary time series. Cambridge, Mass., The Massachusetts Institute of Technology Press. 1949.

ACKNOWLEDGMENTS

The author gratefully acknowledges the suggestions and guidance of his major professor, Dr. R. G. Brown, in both the developmental and writing phases of this dissertation. The author also wishes to thank the National Science Foundation for their help in the form of a fellowship.