UC-32

LBL-21160

Preprint c.1

Lawrence Berkeley Laboratory

UNIVERSITY OF CALIFORNIA

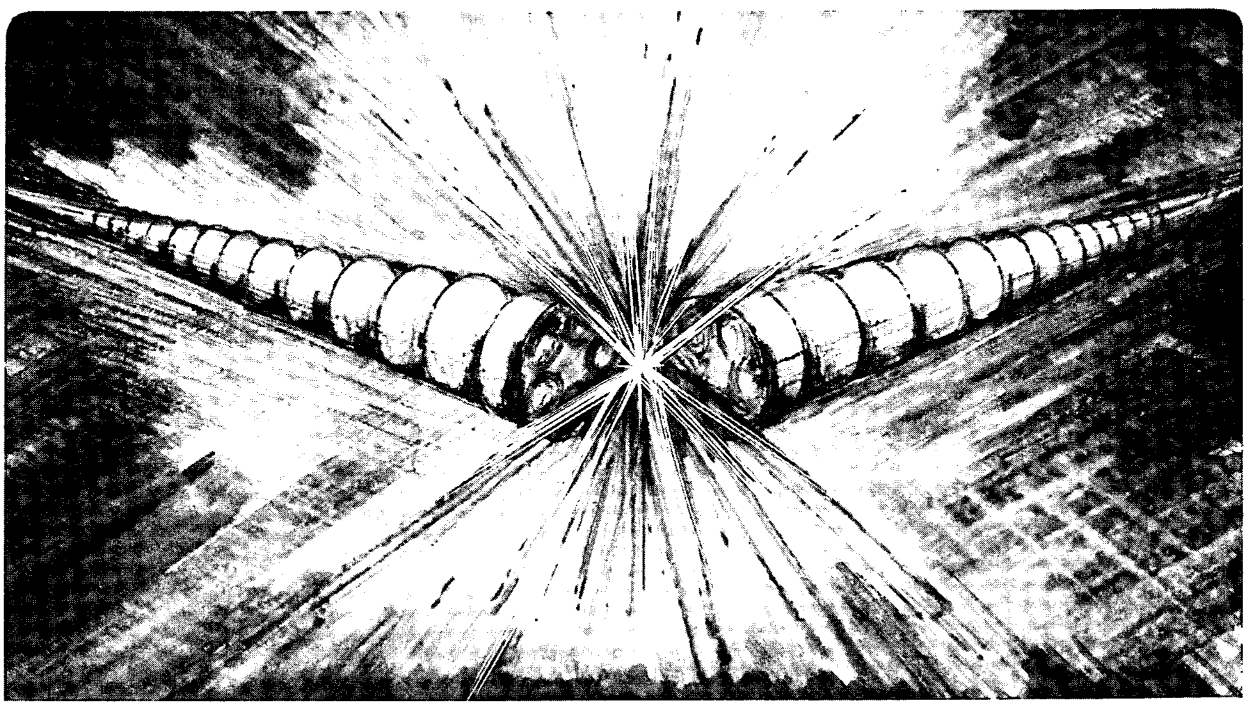
Accelerator & Fusion Research Division

Submitted to American Journal of Physics

USE OF OPTIMAL ESTIMATION THEORY, IN PARTICULAR THE KALMAN FILTER, IN DATA ANALYSIS AND SIGNAL PROCESSING

W.S. Cooper

January 1986



Prepared for the U.S. Department of Energy under Contract DE-AC03-76SF00098

DISCLAIMER

LBL-21160

USE OF OPTIMAL ESTIMATION THEORY, IN PARTICULAR THE KALMAN FILTER, IN DATA ANALYSIS AND SIGNAL PROCESSING*

William S. Cooper

Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

*This work was supported by the U.S. DOE under contract No. DE-AC03-76SF00098.

LBL-21160

USE OF OPTIMAL ESTIMATION THEORY, IN PARTICULAR THE KALMAN FILTER, IN DATA ANALYSIS AND SIGNAL PROCESSING

William S. Cooper

Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

14 January 1985

Abstract

A powerful and useful optimal estimation technique, the Kalman filter, does not seem to be widely known among physicists. In this article we outline the derivation of the algorithm, and give three examples of its use: a) in estimating the value of a constant, with both system and measurement noise, b) in numerical differentiation of noisy data, and c) in optimally estimating the amplitude of a signal with arbitrary but known time dependence superimposed on a noisy background.

Submitted to American Journal of Physics

I. Introduction

Sometimes it is desirable to make the best estimate possible of a set of parameters that describe the state of a system, under circumstances in which information about the state of the system can be gleaned from periodic measurements, and under conditions in which both the state of the system and the measurements are affected by the presence of noise. This is a much more common problem than might be thought from the description just given. Examples of systems in this category include the estimation of the first derivative of noisy data, the amplitude of a coded signal buried in noise, the position, velocity, and acceleration of a moving object being tracked by radar, the location of an aircraft traveling under inertial navigation with periodic updates from another navigational system, the state of a chemical plant or other complex system being monitored and controlled, and the age distribution of the French beef cattle herd as a function of time.[1]

An extensive body of theory has been developed to deal with problems such as these--the general field is called "optimal estimation theory." It deals with the development of optimal estimators, or algorithms, for utilizing new input (measurements) to provide an updated optimal estimate of the state of the system being observed. Pioneering work in the development of efficient, recursive algorithms for estimation was done in the early 1960's by Kalman[2] and Kalman and Bucy.[3] Optimal estimation theory is extensively used in the fields of real-time measurement and control, but does not seem to be widely known among physicists. It can by applied with advantage in many circumstances commonly occurring in physics, however; we shall give several examples later in this article.

In this article we deal with the type of estimator developed by Kalman and Bucy, the least-squares optimal recursive estimator for linear systems, which has come to be known as the "Kalman filter." The use of the word "filter" reflects the electrical engineering origin of much of the original work with these algorithms. The approach is to incorporate the known or anticipated dynamic behavior of the system being observed into a mathematical model, then to obtain the optimal estimate, in a least-squares sense, of a set of parameters, comprising the "state vector," that describes the state of the system. In this way, the system dynamics, as well as the measurement noise, can be taken into account. As an example, which we will consider in more detail later, suppose we have to determine as accurately as possible the length of a metal rod in a room within which the temperature is subject to random fluctuations, and we have as a measuring tool a micrometer with limited resolution. The system in this case is the length of the rod, which we model as a constant plus additive random noise; the length of the rod changes randomly between measurements. The measurement process also introduces (different) random noise. We will see later in this article how to make the optimal estimate of the length of the rod under these conditions.

It is the intent of this article to outline the derivation of the Kalman filter algorithm and give several examples of its use, in sufficient detail that physicists and others can adapt this powerful tool to their particular applications. The development and notation used in this article generally follows that of the excellent book on optimal estimation edited by Gelb.[4]

II. The Recursive Estimator

A recursive estimator is an algorithm that uses new information derived from measurements to update the previous estimate of the parameters describing the state of a system. Only the most recent estimate is needed; the previous information that led to the estimate is not required. A recursive estimator is well suited, therefore, for analyzing data arriving in real time, but can also be applied to "batch" data.

A. The Linear Recursive Estimator

In this article we deal only with linear systems. One might expect a recursive estimator for a simple linear scalar system to be of the form

$$\hat{x}_k = \hat{x}_{k-1} + K_k (z_k - \hat{x}_{k-1}). \tag{1}$$

Here $\hat{x}_k$ is the estimate of the quantity x after the k^{th} measurement,

$\hat{x}_{k-1}$ is the estimate of x after the $(k-1)^{th}$ measurement,

K_k is the "gain" of the filter, and is a function of k, and

z_k is the k^{th} measurement of x.

This equation simply states that the new estimate is the previous estimate plus a weight factor times the difference between the new measurement and the previous estimate. This weight factor K_k is called the gain of the filter.

It is a characteristic of recursive estimators that the gain K decreases with the number of measurements (possibly after some initial transient response). Initially, the filter weights new measurements heavily, then, as the estimate $\hat{x}$ becomes more and more accurate, the weighting factor K either continually decreases or decreases and asymptotically approaches a constant value, so that new measurements have a relatively small effect on the estimate.

B. Response Time of the Filter

We can deduce the response time τ of the filter from Eq. (1). If the change over a single interval of measurement ($\Delta k = 1$) is small, and K_k is constant or very slowly varying, the discrete form given in Eq. (1) is an approximation to the continuous form given by

$$d\hat{x}_{k-1}/dk = -K_k\ (\hat{x}_{k-1} - z_k). \tag{2}$$

Suppose x changes discontinuously at k = 0 from x = 0 to $x = x_0$, and z_k is a good approximation to $\hat{x}_k$, so that $z = x_0$. Then with the substitution $y = (\hat{x}_{k-1} - x_0)$, it is obvious that the time constant for response to change is

$$\tau = 1/K_k. \tag{3}$$

C. An Example of a Recursive Estimator

Let us consider a simple example of a recursive estimator. Suppose we have made a series of measurements $z_1, z_2, \ldots, z_{k-1}$ of a scalar quantity x, and we wish an estimate for x. The least-square optimal estimate of x is just the mean of the set of measurements. Therefore, the $(k-1)^{th}$ estimate of x is

$$\hat{x}_{k-1} = 1/(k-1) \sum_{i=1}^{k-1} z_i. \tag{4}$$

If we make one more measurement, we obtain the k^{th} estimate of x,

$$\hat{x}_k = (1/k) \sum_{i=1}^{k-1} z_i. \tag{5}$$

This expression can be rewritten as

$$\hat{x}_k = (1/k) \sum_{i=1}^{k-1} z_i + (1/k)z_k = ((k-1)/k)\hat{x}_{k-1} + (1/k)z_k, \tag{6}$$

and finally, as

$$\hat{x}_k = \hat{x}_{k-1} + (1/k)(z_k - \hat{x}_{k-1}). \quad (7)$$

The estimator is now in the anticipated form, and we see that the gain is

$$K_k = 1/k. \quad (8)$$

Since we have no prior knowlege of x, we take the initial estimate of x to be zero; in that case the first estimate for x, $\hat{x}_1$, is just equal to the first measurement z_1. With increasing number of measurements the gain K_k ultimately becomes vanishingly small in this example, and new measurements have no effect on the estimate.

III. The Optimal Recursive Estimator (Kalman Filter)

We now broaden the concept of the recursive estimator and seek one that will give the optimal estimate of the state of a system in the presence of noise, both in the measurement and in the parameters characterizing the state of the system. These n parameters, which are assumed to be either constant or varying slowly compared with the response time of the filter, form the n components of the state vector $\underline{x}_k$ at the time of the k^{th} measurement. Our approach is to define models for the time evolution of both the system and the measurement, and then to find the filter gain K_k that minimizes in a least-squares sense the errors in estimating the state of the system after each new measurement.

A. The System Model

We assume that the system is examined periodically, and that the state at time t_k is linearly related to the state at time t_{k-1} by a matrix Φ_{k-1}, with the addition of a random noise vector $\underline{w}_{k-1}$. We further assume that the vector

$\underline{w}$ is gaussian, with zero mean, and that the values at different measurement times are uncorrelated. The system model is then

$$x_k = \Phi_{k-1} \underline{x}_{k-1} + \underline{w}_{k-1}. \tag{9}$$

This model is general enough to describe any system that can be represented by an n^{th} order linear ordinary differential equation, or by a set of such equations (e.g., the equations of motion of an object, or the behavior of an electrical circuit). The vector $\underline{w}$ represents the system noise, that is, random uncertainties in the state of the system from one measurement time to the next. Examples include the noise introduced by measuring the length of a metal rod in a room with poor temperature control (the length of the rod actually varies from one measurement time to the next), the position of an aircraft with erratic engine performance, or subject to variable winds, or the price of a stock or bond subject to random market pressures. If the system noise $\underline{w}$ is zero, the system evolves in a completely deterministic fashion.

The system model given in Eq. (9) generates a random walk in the set of system parameters that make up the state vector.

B. The Measurement Model

Now suppose that we make a set of ℓ measurements at time k, forming the ℓ components of the measurement vector $\underline{z}_k$, and that these measurements are linearly related to the state vector $\underline{x}_k$ by the measurement matrix H_k. The measurements are also corrupted by noise $\underline{v}_k$, which we assume to be gaussian with zero mean. This gives us the measurement model

$$\underline{z}_k = H_k \underline{x}_k + \underline{v}_k. \tag{10}$$

The measurement matrix H_k can be an arbitrary function of time or k, but must, of course, be known or measurable at the time of each measurement. We will use this to advantage in an example of signal processing later in this article.

C. Derivation of the Equations for the Optimal Linear Estimator (Kalman Filter)

In this section we outline, without going into detail, the arguments followed in deriving the optimal recursive estimator. For the omitted details, see Gelb,[4] whose derivation and notation we generally follow.

We seek a new estimate for the state of the system immediately following the kth measurement, denoted by $\hat{\underline{x}}_k(+)$, that is a linear combination of the estimate immediately preceeding the k^{th} measurement, denoted by $\hat{\underline{x}}_k(-)$, and the new measurement $\underline{z}_k$.

$$\hat{\underline{x}}_k(+) = K_k' \hat{\underline{x}}_k(-) + K_k \underline{z}_k \; . \tag{11}$$

Here K_k' and K_k are time-varying matrices to be determined.

We define the error $\underline{x}(+)$ in the estimate after the k^{th} measurement to be

$$\tilde{\underline{x}}_k(+) = \hat{\underline{x}}_k(+) - \underline{x}_k , \tag{12}$$

with a similar definition for $\tilde{\underline{x}}_k(-)$. Here $\underline{x}_k$ is the true value of the state vector at the time of the k^{th} measurement. When we substitute these error definitions into Eq. (11), and require that the new error estimate $\tilde{\underline{x}}_k(+)$ be unbiased (expectation value equal to zero), we find that

$$K_k' = I - K_k H_k , \tag{13}$$

where I is the identity matrix. This expression, when substituted into Eq. (11), gives an estimator of the form

$$\hat{\underline{x}}_k(+) = \hat{\underline{x}}_k(-) + K_k\ (\underline{z}_k - H_k\ \hat{\underline{x}}_k(-)), \tag{14}$$

which shows how the estimate of the state of the system $\hat{\underline{x}}_k(-)$ just before the measurement is modified by the new measurement. The corresponding change in the error estimate is given by

$$\tilde{\underline{x}}_k(+) = (I - K_k H_k)\tilde{\underline{x}}_k(-) + K_k \underline{v}_k. \tag{15}$$

We are unlikely ever to know the actual error $\tilde{\underline{x}}_k$. Statistically speaking, we can only determine the root mean square error in the components of the state vector. As the first step in calculating these quantities, we define the <u>covariance matrix</u> associated with any vector $\underline{a}$ to be the matrix $E[\underline{a}\ \underline{a}^T]$, where the symbol E denotes the <u>expectation value</u>, or mean, of the elements of the matrix, and the $\underline{a}^T$ denotes the transpose of the vector $\underline{a}$. We then define the <u>error covariance matrix</u> after the k^{th} measurement to be

$$P_k = E\ [\tilde{\underline{x}}_k(+)\ \tilde{\underline{x}}_k(+)^T]. \tag{16}$$

The diagonal elements of this matrix are the expectation values of the squares of the errors in the estimate of the components of $\underline{x}$ after the k^{th} measurement. Shortly we will choose K_k to minimize the "square of the length of the error vector", the sum of the diagonal elements (the trace) of P_k.

Next we substitute the expression for $\tilde{\underline{x}}_k(+)$ given by Eq. (15) into Eq. (16). We define

$$P_k' = P_k(-) = E[\tilde{\underline{x}}_k(-)\ \tilde{\underline{x}}_k(-)\ ]\ , \tag{17}$$

$$R_k = E\ [\underline{v}_k\ \underline{v}_k\ ^T], \tag{18}$$

assume that the measurement errors and the errors in the state vector are uncorrelated, and find that

$$P_k = (I - K_k H_k)(P_k') (I - K_k H_k)^T + K_k R_k K_k^T . \quad (19)$$

This expression shows how the error covariance matrix just before the k^{th} measurement, P_k', is modified by the new measurement.

We are now ready to minimize the error estimates. We need to differentiate the trace of P_k with respect to the matrix K_k (this is not trivial), set the result equal to zero, and solve for K_k. This provides us with an expression for the filter gain K_k that minimizes in a least-squares sense the estimate of the error in the state vector $\underline{x}_k$ after the k^{th} measurement; the result is

$$K_k = P_k' H_k^T [H_k P_k' H_k^T + R_k]^{-1} . \quad (20)$$

This choice of the filter gain K_k gives the <u>optimal linear recursive estimator</u>. From this point on, K_k will be assumed to have this form.

The Kalman filter is an optimal least-squares estimator, and gives exactly the same answer as other least-squares estimators that use the same information.

So far we have only concerned ourselves with how various quantities change instantaneously when new information becomes available--the change from just before the new measurement , (-), to just after the measurement, (+). To complete the picture, we need to know how the estimate of the state vector and how the error covariance matrix propagate between measurements (from just after the last measurement to just before the new measurement). One can easily show that these propagate as

$$\hat{\underline{x}}_k(-) = \Phi_{k-1}\hat{\underline{x}}_{k-1}(+) \text{ and} \quad (21)$$

$$P_k' = \Phi_{k-1} P_{k-1} \Phi_{k-1}{}^T + Q_{k-1}, \tag{22}$$

where we have defined

$$Q_k = E[\underline{w}_k \underline{w}_k{}^T] . \tag{23}$$

Q_k is the covariance matrix for the system noise, as defined in Eq. (9).

D. Recursion Relations for the Optimal Linear Estimator (Kalman filter)

We can now combine the above calculations into four matrix equations that tell us how to calculate the error covariance just prior to the new measurement, and how to use the new information provided by the k^{th} measurement to calculate the optimal filter gain, the new estimate of the system state vector, and the error covariance of the state vector just after the new measurement. These four equations comprise a complete algorithm for updating the estimate of the system state vector in a stepwise, recursive, fashion; they constitute the optimal linear recursive estimator, or Kalman filter. The equations are

Error covariance matrix prior to measurement:

$$P_k' = \Phi_{k-1} P_{k-1} \Phi_{k-1}^T + Q_{k-1} ; \qquad (24a)$$

Optimal (Kalman) filter gain:

$$K_k = P_k' H_k^T [H_k P_k' H_k^T + R_k]^{-1} ; \qquad (24b)$$

New estimate of the system state vector:

$$\hat{\underline{x}}_k = \Phi_{k-1} \hat{x}_{k-1} + K_k[\underline{z}_k - H_k \Phi_{k-1} \hat{\underline{x}}_{k-1}]; \qquad (24c)$$

Updated error covariance matrix:

$$P_k = [I - K_k H_k] P_k' . \qquad (24d)$$

By definition,

$$Q_k = E [\underline{w}_k \underline{w}_k^T] \text{ (system noise covariance) , and} \qquad (24e)$$

$$R_k = E [\underline{v}_k \underline{v}_k^T] \text{ (measurement noise covariance) .} \qquad (24f)$$

E. Initial Conditions

To start off the filter, we need to know the system and measurement error covariance matrices R and Q, and we need intial guesses for the state vector $\underline{x}$ and for the error covariance matrix P. After each measurement, the filter can then be advanced one step by using the Kalman filter algorithm summarized in Eqs. (24a) through (24d).

F. General Observations on P_k

The heart of the Kalman filter is in the error covariance matrix P_k, which provides the estimate of the errors in the state vector at each step.

Inspection of the Kalman filter algorithm shows that the error covariance matrix P_k depends on the intial conditions, the description of the system (Φ), the description of the measurements (H), the system noise (Q), and the measurement noise (R), but not at all on the measurements z_k! This means that the estimated actual (not relative) errors in the components of the state vector (the square roots of the diagonal elements of P) can be calculated as a function of k in advance of any measurements, provided that Φ, H, Q, and R are either constant or have known dependence on time or k. In fact, to save time in real-time data processing, once the dependence of P_k on k has been determined, P_k can be approximated by an algebraic function of k or by a step-wise varying function. Except in very simple cases, the functional dependences of P_k on k cannot be determined.

IV. Advantages of the Kalman Filter

Before giving examples of the use of the Kalman filter, we list some of the advantages its use affords, although some of these are not yet obvious from the discussion given so far:

1. The Kalman filter is recursive, which minimizes memory requirements and makes the filter well suited to deal with data in real time.
2. It provides after each measurement the optimal estimate of the state of a linear system--it can be shown that the Kalman filter is the optimal estimator for linear systems.
3. It provides after each measurement an estimate of the errors in the parameters characterizing the state of the system.
4. In many cases the time response of the filter is fast enough to follow the evolution of physical systems in real time.

5. The filter responds gracefully to discontinuous changes (steps) in the measurements, without erratic transients.
6. The filter is flexible, and can be used with many physical systems, such as those that can be represented by a set of linear ordinary differential equations, and so is broadly applicable to problems arising in physics.
7. The filter uses knowledge of the system dynamics in the estimation process, which acts as a constraint against estimates based on occasional implausible measurements.
8. While the Kalman filter is equivalent to other least squares estimators that use the same information, it is often easier to implement in complex situations.

V. Examples

In this section we discuss three examples of the use of the Kalman filter in the optimal estimation of

1. A constant, with both system and measurement noise,
2. The first derivative of a set of noisy data, and
3. The amplitude of a time-dependent signal superimposed on a constant background, with both signal and measurement noise.

We begin with

A. A constant, with both system and measurement noise

We now return to the problem of optimally estimating the length of the metal rod in the room with poor temperature control. This example illustrates the use of a Kalman filter to optimally estimate the value of a constant (the length of the metal rod) subject to system noise (the room temperature fluctuates randomly, and so does the length of the rod, correspondingly) and to measurement noise (the micrometer has limited resolution).

In this example, the state vector is just the length of the rod.

State Vector:

$\underline{x} = x$, a scalar.

System Model:

$x_k = x_{k-1} + w_{k-1}$ (the new length is the previous length plus a random component).

therefore $\Phi = 1$, $Q = E\,[w^2] = \sigma_s^2 = q$.

Measurement Model:

$z_k = x_k + v_k$,

therefore $H = 1$, $R = E\,[v^2] = \sigma_m^2 = r$.

We determine the recursion relations for the Kalman filter by working our way through one step of the filter algorithm:

$$P_k' = P_{k-1} + Q_{k-1} = P_{k-1} + q.$$

$$K_k = (P_{k-1} + q)/(P_{k-1} + q + r), \text{ and}$$

$$P_k = [1 - (P_{k-1} + q)/(P_{k-1} + q + r)]\,(P_{k-1} + q)$$
$$= r(P_{k-1} + q)/(P_{k-1} + q + r)\ .$$

After running a number of cycles, the filter will reach a steady state, with constant values for K_k and P_k. The steady-state value for P, P_∞, can in the case of this simple example be found by setting $P_k = P_{k-1} = P_\infty$ in the above expression and solving for P_∞; the solution is

$$P_\infty = 1/2\ [q^2 + 4qr - q].$$

The square root of this expression is the minimal error in estimating x that can be achieved even with an unlimited number of measurements.

Next we consider two special cases: first, systems with no system noise and second, those with no measurement noise.

1. No System Noise

If there is no system noise ($q = 0$), then from the system model given by Eq. (9), the value of x cannot change ($\Phi = 1$). The only way that x can change is through the random variable w. This is an important point--if the physics of the system dictates that the system state vector can change randomly with time, then there must be system noise ($Q \neq 0$). In the alternative case ($Q = 0$), the system can only evolve deterministically.

In this case, for $q = 0$, we can find P_k in closed form as a function of k. The only uncertainty in the problem is that of the measurement noise. Therefore, we set $P_1 = r$. By successive invocations of the filter algorithm we find that

$$P_2 = r/2,\ P_3 = r/3, \ldots, P_k = r/k.$$

This means that the error in the estimate of x after k measurements is $\sigma_m/\sqrt{k}$, in agreement with the usual result for the error in the mean of a set of measurements. Since the value of x is truly constant, we can reduce the error in the estimate of x to any desired value by making a sufficient number of measurements.

2. No Measurement Noise

If there is no measurement noise ($r = 0$), then $P_k = 0$ for all k. The state of the system can be determined exactly at each measurement. In this case the Kalman gain is $K_k = 1$, so the estimate after the k^{th} measurement, from Eq. (24c), is

$$\hat{x}_k = \hat{x}_{k-1} + [z_k - \hat{x}_{k-1}] = z_k \; ;$$

the new measurement is the optimal estimate of x which is, in fact, determined <u>exactly</u> at the instant of the measurement.

B. <u>The First Derivative of Noisy Data</u>

Occasionally we have to practice numerical differentiation of noisy numerical data. We can use the Kalman filter to generate an optimal estimate of the first derivative and the amplitude of a measured signal.

State Vector:

Let $\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, where

x_1 = the amplitude of the signal, and

$x_2 = dx_1/dk$, the change in x_1 between measurements.

System Model:

$$x_{1,k} = x_{1,k-1} + x_{2,k-1} + w_{1,k-1}$$
$$x_{2,k} = \qquad\qquad x_{2,k-1} + w_{2,k-1} \; .$$

The system model generates a straight line with x_1 changing by an amount x_2 at every step, plus noise in both the amplitude and the slope. We see that

$$\Phi = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \text{a constant matrix.}$$

We assume the noises associated with x_1 and x_2 are uncorrelated, so

$$Q = E\,[\underline{w}\;\underline{w}^T] = \begin{pmatrix} \sigma_{s1}^{\;2} & 0 \\ 0 & \sigma_{s2}^{\;2} \end{pmatrix} = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix} .$$

Measurement Model:

We can only measure the amplitude of the signal, a scalar:

$z_k = x_{1,k} + v_k$, so $H = (1 \quad 0)$. Assume

$R = E\,[\underline{v}_k\;\underline{v}_k^{\;T}] = r$.

Define $P_{k-1} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}$; after substitution into Eqs 24a

through 24d, we find that the recursion relations are

$$P_k{}' = \begin{pmatrix} P_{11}+P_{12}+P_{21}+P_{22}+q_1 & P_{12}+P_{22} \\ P_{21}+P_{22} & P_{22}+q_2 \end{pmatrix} = \begin{pmatrix} P_{11}{}' & P_{12}{}' \\ P_{21}{}' & P_{22}{}' \end{pmatrix} ,$$

$K_k = 1/D \begin{pmatrix} P_{11}{}' \\ P_{21}{}' \end{pmatrix} = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}$, where $D = (P_{11}{}' + r)$, and

$$P_k = \begin{pmatrix} (1-K_1)\,P_{11}{}' & (1-K_1)\,P_{12}{}' \\ -K_2\,P_{11}{}' + P_{21}{}' & -K_2\,P_{12}{}' + P_{22}{}' \end{pmatrix} .$$

The estimator is

$$\hat{x}_{1,k} = (1-K_1)(\hat{x}_{1,k-1} + \hat{x}_{2,k-1}) + K_1 z_k \quad ,$$

$$\hat{x}_{2,k} = (1-K_2)\hat{x}_{2,k-1} + K_2(z_k - \hat{x}_{1,k-1}) \quad .$$

This simple system is already too complicated for us to proceed very far algebraically, so we have left the recursion relations in forms convenient for numerical evaluation.

The performance of this Kalman differentiator on simulated data is shown in Figure 1. We have chosen a signal consisting of a constant portion, then a linear ramp with slope $dx_1/dk = +1$, a second constant portion, then an abrupt decrease to the original level. Noise variances used in this example are $q_1 = 1.0$, $q_2 = 0.05$, and $r = 10.0$.

We used as initial conditions $x_1 = 0.0$, $x_2 = 0.0$. Since our initial estimate for x_1 cannot be better than the measurement noise, we have set $P_{11} = r$, and have derived our initial estimate of the variance in x_2 from a hypothetical measurement of the slope based on two consecutive amplitude measurements: $P_{22} = 2\, r$. P_{12} and P_{21} were initially set equal to zero. With the noise variances used, the filter reaches steady-state in about 20 measurements.

Gaussian noise with variance (q_1+r) has been added to the signal to approximate the effects of the system noise and the measurement noise; this is the top trace in Figure 1. The second trace shows the Kalman filter's estimate of the amplitude, $\hat{x}_1$-- it duplicates the behavior of the noisy signal, but the noise is substantially less, as expected. The bottom trace is the estimated slope, $\hat{x}_2$, with the ideal response (no noise) superimposed.

The estimated error in $\hat{x}_2$ is about 60% of the amplitude of the rectangular pulse. Differentiation is easier in hindsight than in real time. If we knew that the slope was "supposed" to remain constant during the ramp portion, we could, of course make a more accurate estimate of the slope. The filter does not know this, and neither would we in a situation in which we had no prior knowledge of the shape of the signal. The filter output shown is the <u>optimal estimate of the amplitude and slope after each measurement, and cannot be improved upon</u>.

C. <u>The Amplitude of a Time-dependent Signal Superimposed on a Constant Background, With Both Signal and Measurement Noise</u>

The Kalman filter can also be used for optimally estimating the amplitude of a time-dependent signal buried in noise; the following example gives an illustration of this application.

1. <u>The Models</u>

Suppose we have a time-dependent signal $x_1 g(t)$ superimposed on a constant background x_2 and we wish to optimally estimate the signal amplitude x_1.

State Vector: We take

$$\underline{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} .$$

System Model:

$$x_k = x_{k-1} + w_{k-1}, \text{ so } \Phi = I, \text{ the identity matrix.}$$

Assume the system noise (components of $\underline{w}$) to be uncorrelated. Then

$$Q = \begin{pmatrix} \sigma_{s1}^2 & 0 \\ 0 & \sigma_{s2}^2 \end{pmatrix} = \begin{pmatrix} q_1 & 0 \\ 0 & q_2 \end{pmatrix} = \text{a constant matrix.}$$

Measurement Model:

We will incorporate the time dependence of the signal into the measurement model. At time t_k we measure the scalar quantity

$$z_k = x_{1,k}\, g(t_k) + x_{2,k} + v_k \;.$$

From the information available to us from the measurement, we cannot tell whether the signal is actually time-varying or whether the time dependence is built into the measurement model--the two cases are indistinguishable.

From the form of the measurement, we identify

$$H = (g \quad 1),$$

where we have written g for g(t). We need also to define

$$R = E\,[\underline{v}\ \underline{v}^T] = \sigma_m^{\ 2} = r \text{ , a scalar.}$$

2. <u>The Recursion Relations</u>

The first equation of the filter algorithm, Eq.(24a), gives

$P_k' = P_{k-1} + Q$. Let

$$P_{k-1} = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{21} \end{pmatrix} \text{ . Then}$$

$$P_k' = \begin{pmatrix} P_{11} + q_1 & P_{12} \\ P_{21} & P_{22} + q_2 \end{pmatrix} = \begin{pmatrix} P_{11}' & P_{12}' \\ P_{21}' & P_{22}' \end{pmatrix}, \text{ and}$$

$$K_k = 1/D \begin{pmatrix} g(P_{11} + q_1) + P_{12} \\ g\ P_{21} + P_{22} + q_2 \end{pmatrix} = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix}, \text{ where}$$

$$D = g^2(P_{11} + q_1) + g\ (P_{12} + P_{21}) + P_{22} + q_2 + r\ .$$

The estimate of the state vector is

$$\hat{x}_{1,k} = \hat{x}_{1,k-1}\ (1-g\ K_1) + K_1\ (z_k - \hat{x}_{2,k-1})\ ,$$

$$\hat{x}_{2,k} = \hat{x}_{2,k-1}\ (1 - K_2) + K_2(z_k - g\ \hat{x}_{1,k-1}).$$

The updated error covariance follows from Eq. (24d):

$$P_k = \begin{pmatrix} (1 - g\ K_1)\ P_{11}' - K_1\ P_{21}' & (1 - g\ K_1)\ P_{12}' - K_1\ P_{22}' \\ (1 - g\ K_1)\ P_{21}' - K_1\ P_{22}' & -g\ K_2\ P_{12}' + (1 - K_2)\ P_{22}' \end{pmatrix}.$$

So far we have not had to specify the functional form of the time dependence of the signal, g(t). Aside from some practical constraints (g(t) should not change appreciably during the measurement time, and g(t) cannot have the same value for all measurements), the only restriction on g(t) is that its value must be known at the time of each measurement. The function g(t) could be periodic in time, a random sequence of pulses, or even (if one wanted to make a clandestine measurement) derived from a random noise generator! If the signal were generated by an unpredictable event, such as an earthquake, an explosion, or a solar flare, it should be possible to use the measured time dependence of the disturbance, together with a variable time delay and any modifications known to have occurred in the signal during its transit, to search for the signal. In some cases when one has control of g(t), it might be possible to choose its functional form in a way that minimizes the error estimates. We will give an example of this signal optimization later.

3. Initial Conditions

To aid in understanding the working of a Kalman filter, we will continue with a numerical example of signal processing. Suppose our signal consists of a periodic series of pulses, so that at the time of any measurement g is either zero or 1. The signal is a periodic rectangular (not necessarily square) wave.

Our first problem is in choosing suitable initial conditions for starting the filter. This is not as difficult a problem as it might appear; except for singularly unfortunate choices of the initial error covariance matrix, the filter will reach the same steady state (provided, or course, that Q and R are the same in all cases). Still, we should make reasonable choices, based on our understanding of the problem, so that the error estimates will be reasonable as the filter approaches steady-state.

In this case, let us assume that we make the first measurement when there is no signal present; g = 0. The measurement cannot affect $\hat{x}_1$ since it contains no information about the signal amplitude. We see, then, from the equation for the estimate of the state vector, that for $\hat{x}_{1,1}$ to be unaffected by the measurement, K_1 (the first component of the vector K) must be zero. With g = 0, this leads to

$K_1 = P_{12}/(P_{22} + q_2 + r)$, so we require

$P_{12} = P_{21} = 0$ initially.

What can we use for initial estimates of P_{11} and P_{12}? We are measuring a quantity (forget the time dependence for the moment)

$$z = x_1 + x_2 + v,$$

and we know the noise covariances for all three quantities on the right hand side of this equation. We therefore expect z to have a noise covariance n given by

$$n = q_1 + q_2 + r.$$

In the absence of other information, it is reasonable to assign half the initial error to x_1 and half to x_2, so we choose for initial conditions

$$P_{11} = P_{22} = (1/2)\ n\ .$$

Since we measure the sum of $\tilde{x}_1$ and $\tilde{x}_2$, if the actual error $\tilde{x}_1$ is positive, $\tilde{x}_2$ must necessarily be negative, so the expectation value of their product (P_{12} and P_{21}) when the filter reaches steady-state must be negative; the errors are anti-correlated.

In practice, the noise characteristics must be either estimated from the physics of the problem or measured. Remarkably, it is possible to determine all the noise characteristics (the matrices Q and R) from measurements alone.[5] This is a very useful fact, as in many cases (in economics, for example), we may not have a physical basis for estimating the system noise. The components of Q must be large enough to give the expected variation of $\underline{x}$ during the observation time by the random walk process implicit in the system model. This fact can also be used as a basis for estimating the system noise.

4. Results for a Square Wave

In this example, we will express the noise covariances in terms of the measurement noise covariance, r, which we take to be 1. We choose

$$q_1 = 10^{-4}, \quad q_2 = 10^{-2}, \text{ and } r = 1\ .$$

The characteristic noise amplitude of the background ($\sqrt{q_2} = \sigma_{s2}$) is then 1/10 of the measurement noise amplitude, and that of the signal ($\sqrt{q_1} = \sigma_{s1}$) is 1/100 of the measurement noise. The sum of the noise covariances (n) is very nearly 1, so we take for the initial error covariance matrix

$$P = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix} .$$

The dependence of the components of P_k on k are shown for the case of a square wave in Figure 2. P has reached steady state within 1% by the 500th measurement; in steady-state P is

$$P = \begin{pmatrix} 1.995 \times 10^{-2} & -1.002 \times 10^{-2} \\ -1.002 \times 10^{-2} & 1.003 \times 10^{-1} \end{pmatrix} \text{ for k even, and}$$

$$P = \begin{pmatrix} 1.996 \times 10^{-2} & -9.030 \times 10^{-3} \\ -9.030 \times 10^{-3} & 9.931 \times 10^{-2} \end{pmatrix} \quad \text{for k odd.}$$

Within 200 measurements (100 signal periods), the error covariances are nearly as small, and give estimated rms errors of 0.16 for $\hat{x}_1$ and 0.32 for $\hat{x}_2$. If we require that the estimate of the signal amplitude $\hat{x}_1$ be determined to 10%, then $\hat{x}_1$ must be 1.6 or greater. To meet the same requirement with a single measurement, as one might make from an oscilloscope picture, with rms noise of unit amplitude (n) in both signal and background, $\hat{x}_1$ must be $10\sqrt{2}$, or 8.8 times larger. In this case, the use of the Kalman filter provides an improvement of a factor of 8.8 (when the filter reaches steady-state, the improvement is a factor of 10) in our ability to measure the amplitude of a small signal buried in noise.

The values of P_{11} and P_{22} in Figure 2 oscillate from one measurment to the next, depending on whether the signal is present or absent. The dashed portions of the curves approximate the mean values.

5. Optimization of g(t)

The function g(t) in this example can be a periodic rectangular wave, not necessarily square, as was the case just studied. At each measurement the signal is either present or absent. Is there an optimal choice of signal duty factor, the fraction of the period that the signal is present? Figure 3 shows the estimated errors in $\hat{x}_1$ and $\hat{x}_2$ as a function of signal duty factor. The error in $\hat{x}_1$ is a weak function of signal duty factor over most of the range, with the minimum occurring at 0.5. A square wave is therefore the best choice for estimating the signal amplitude. For a signal duty factor of 1, g = 1 always, and both error estimates become infinite--we always measure the sum of x_1 and x_2, and have no means to determine either one separately. For a signal duty factor of zero, we have no information about x_1, so the error is again infinite, but we minimize the error in x_2.

VI. Related Topics

Straightforward extensions of this technique permit us to do optimal smoothing of data and optimal prediction. Nonlinear problems can also be handled. It is possible to make an adaptive filter--one that modifies the original estimates of Q and R to further optimize the performance of the filter. An extensive body of literature concerns model identification. These and other related topics are discussed by Gelb.[4]

VII. Conclusions

We have reviewed the origins of optimal estimation theory, the notion of a recursive estimator, and the derivation of the optimal linear recursive estimator, or Kalman filter, and have given several examples of its use. The Kalman filter offers many advantages in data analysis and signal processing, and is capable of efficient and optimal estimation of parameters of interest in many situations occurring in physics. I hope that this article will stimulate others to use and enjoy this powerful tool.

VIII. Acknowledgements:

I wish to thank Ed Theil for several useful suggestions and discussions, and Oscar Anderson and Wulf Kunkel for their comments on the manuscript. This work was supported by the U.S. DOE under Contract No. DE-AC03-76SF00098.

References:

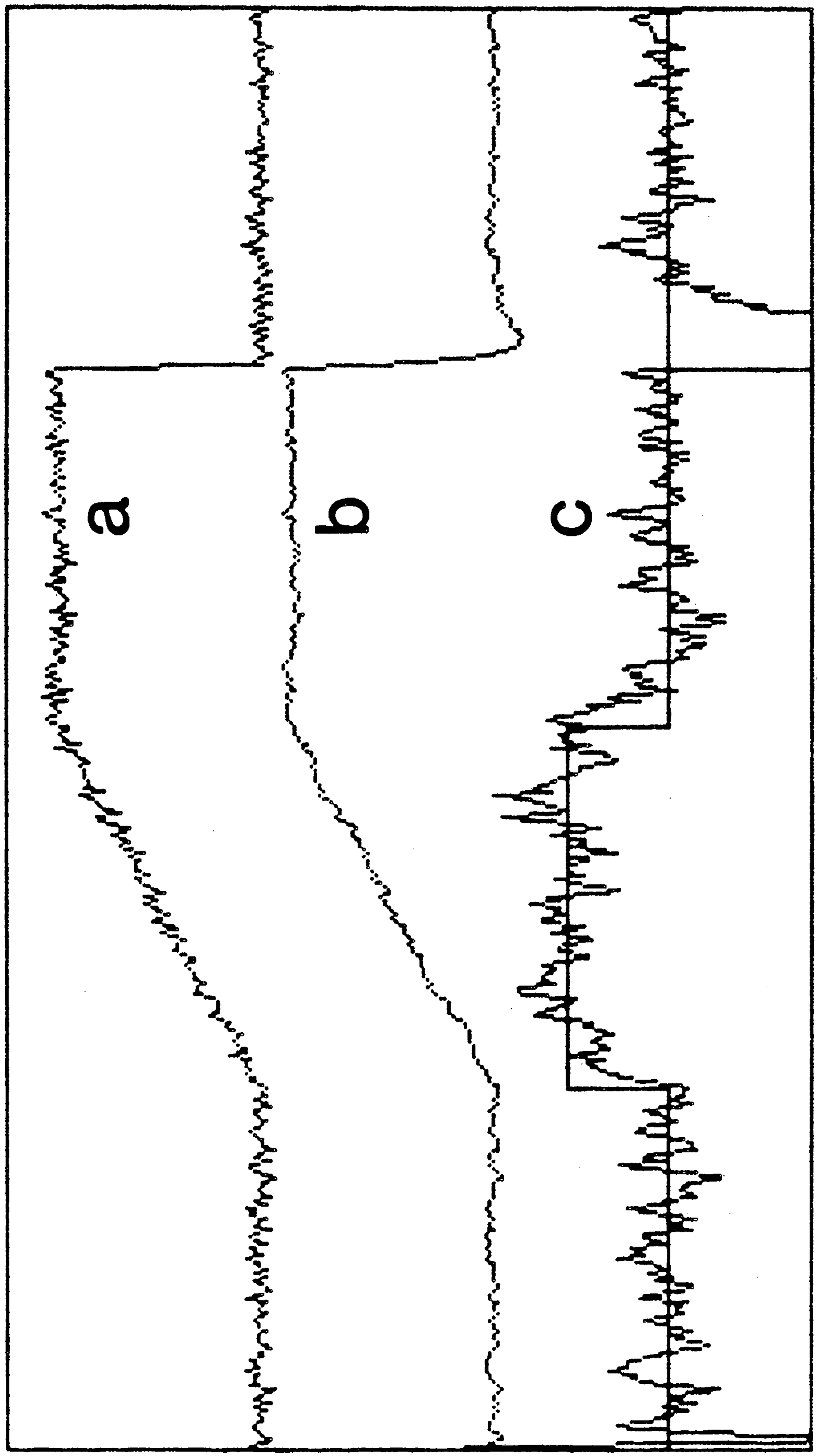
1. For this and other applications, see IEEE Trans. on Automatic Control, AC-28, March, 1983. This is the Special Issue on Applications of Kalman Filtering.
2. R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems," ASME J. Basic Eng., March, 1960, pp. 35-46.
3. R. E. Kalman and R. S. Bucy, "New Results in Linear Filtering and Prediction Theory," AMSE J. Basic Eng., March, 1961, pp. 95-108.
4. A. Gelb, Editor, Applied Optimal Estimation, The M. I. T. Press, Cambridge, Mass., 1974.
5. T. T. Lee, "A Direct Approach to Identify the Noise Covariances of Kalman Filtering," IEEE Trans. on Automatic Control, AC-25, August, 1980, pp.841-842.

Figure Legends

Fig. 1 Use of the Kalman filter in numerical differentiation of noisy data. Trace (a) is the simulated noisy data, (b) is the filter's estimate of the amplitude of the data, and (b) is the estimate of the slope, with the ideal (noise free) response superimposed.

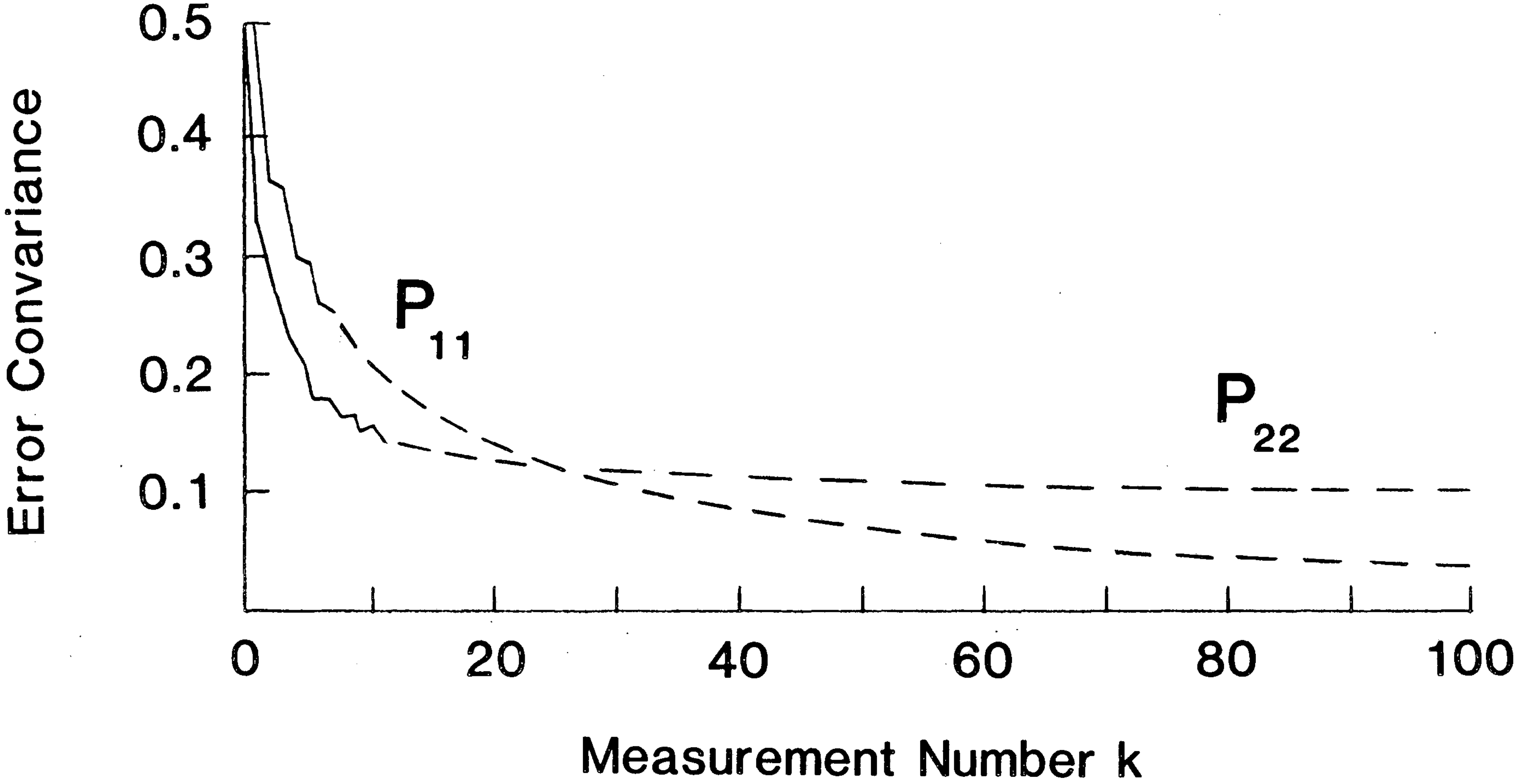
Fig. 2 The dependence of P_{11}, the square of the error in the estimate of the amplitude of the square wave, and P_{22}, the square of the error in the estimate of the background, on measurement number k, for the square wave signal processor.

Fig. 3. Estimated errors in the signal and background amplitudes as functions of the signal duty factor for a rectangular wave signal processor.



XBL 863-765

Fig. 1



XBL 863-767

Fig. 2

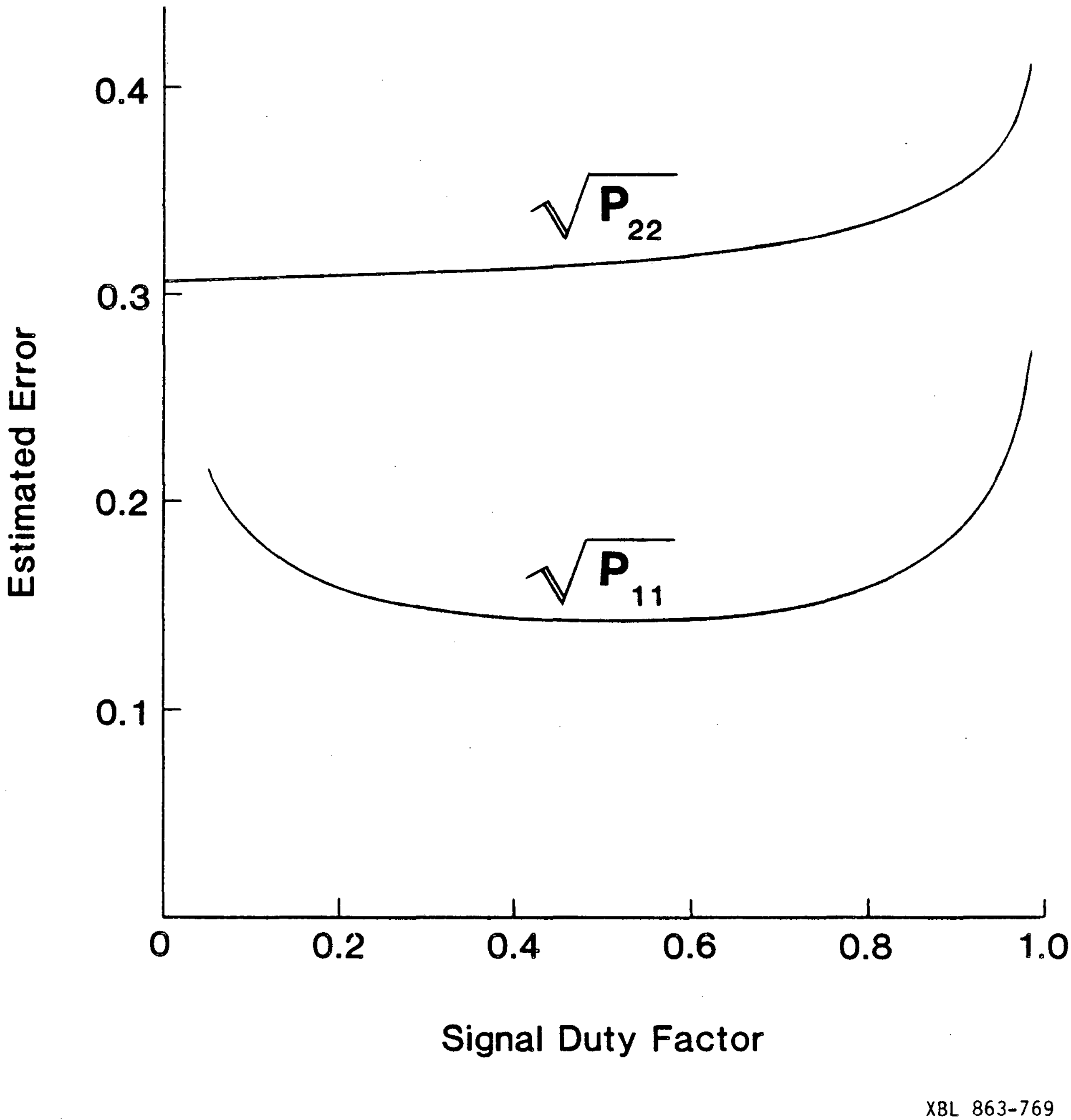

0.4
0.3
0.2
0.1
Estimated Error
$\sqrt{P_{22}}$
$\sqrt{P_{11}}$
0
0.2
0.4
0.6
0.8
1.0
Signal Duty Factor
XBL 863-769

Fig. 3

LAWRENCE BERKELEY LABORATORY
TECHNICAL INFORMATION DEPARTMENT
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720