APPLIED KALMAN FILTERING: AN OVERVIEW

R. Grover Brown
Electrical Engineering Department
Iowa State University

ABSTRACT

A brief resume of the evolution of Kalman filtering from classical filter theory is presented. The required format of the discrete filter model is discussed. The recursive equations for the discrete Kalman filter filter are then presented, but not derived. Two scalar examples are given to illustrate the use of the recursive equations. The first deals with estimation of a random constant; the second illustrates the Wiener process.

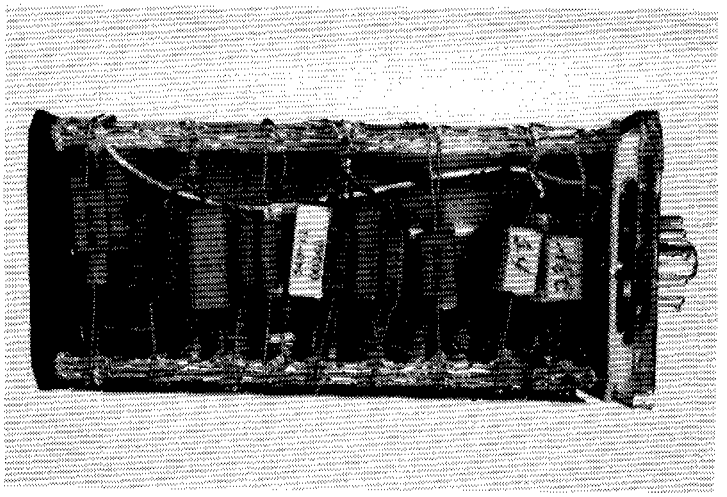
1. INTRODUCTION

It has been about 20 years since R. E. Kalman published his classic paper on recursive least-squares filtering [1]. In addition to its theoretical elegance, Kalman filtering has proved to be eminently practical. This, no doubt, accounts for its continued durability. The first applications in the early 1960's were in position and velocity determination [2,3,4], both in space and terrestial settings. Since then, applications outside the navigation field have become more common. For example, a recent issue of the Bell System Technical Journal was devoted entirely to applications of Kalman filtering to load forecasting [5]. Also, a recent issue of the IEEE Transactions on Automatic Control was dedicated to new applications of Kalman filtering, many of which were outside the traditional application area of navigation [6]. Thus, Kalman filtering is alive and well, and the list of applications continues to grow!

Of necessity, this paper must be brief. Thus, it will be largely an overview or, if you like, a guided tour of the bare essentials of Kalman filtering. With this thought in mind, it is appropriate that we begin with a brief historical perspective.

2. HISTORICAL PERSPECTIVE

Figure 1 shows a conventional telephone bandpass filter side-by-side with a Kalman filter. At first glance, it looks ridiculous even to try to compare the two. On one hand, we have a circuit consisting of resistors, capacitors, etc.; on the other, we have just a set of mathematical equations. One might logically ask, "How in the world did that set of equations ever become known as a 'filter'?" The answer lies in the historical evolution of modern statistical filter theory from classical theory.



BAND-PASS

TELEPHONE FILTER

$$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + R_k]^{-1}$$

$$\hat{x}_k = \hat{x}_k^- + K_k (z_k - H_k \hat{x}_k^-)$$

$$P_k = [I - K_k H_k] P_k^-$$

$$\hat{x}_{k+1}^- = \Phi_k \hat{x}_k$$

$$P_{k+1}^- = \Phi_k P_k \Phi_k^T + Q_k$$

KALMAN FILTER

RECURSIVE EQUATIONS

Fig. 1 (a) Conventional analog filter and (b) Kalman filter

Figure 2 shows a simplified chronology of the three major branches of filter theory as we know it today. Classical filter theory began in the early days of telegraphy, and it continues today as an active discipline within electrical

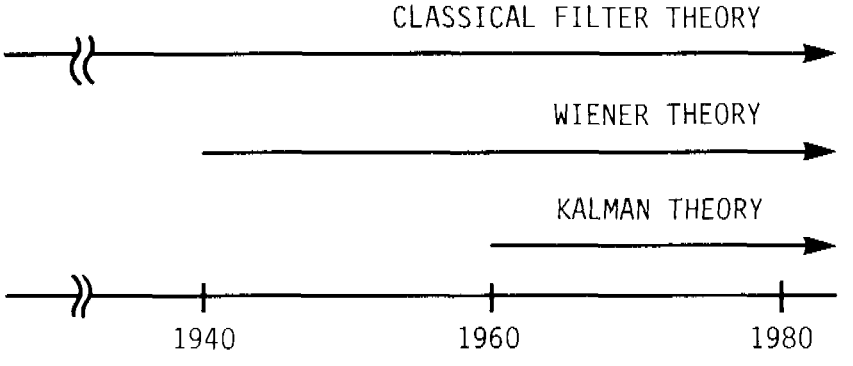


Fig. 2 Chronology of filter theory

engineering. The basic problem of classical theory is shown in Figure 3. It is basically one of synthesizing a given frequency response with R, L, C and (nowadays) active elements. The word given is underscored to emphasize the fact that the designer is assumed to know a priori what response is desirable in the application at hand. The only problem remaining then is that of practical implementation of the desired response (to some degree of approximation). Obviously, this problem is just as fresh and important today as it was in the 1800's.

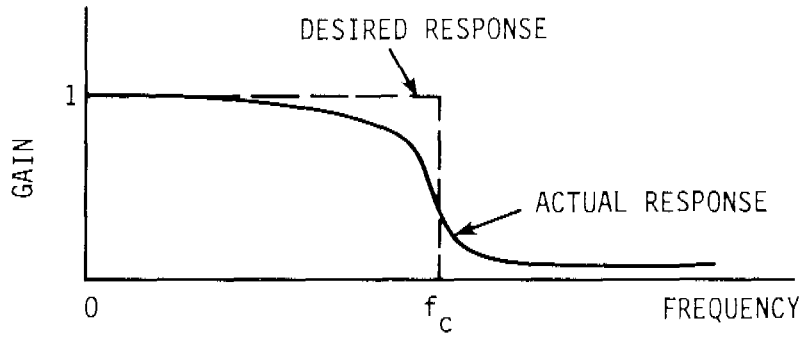
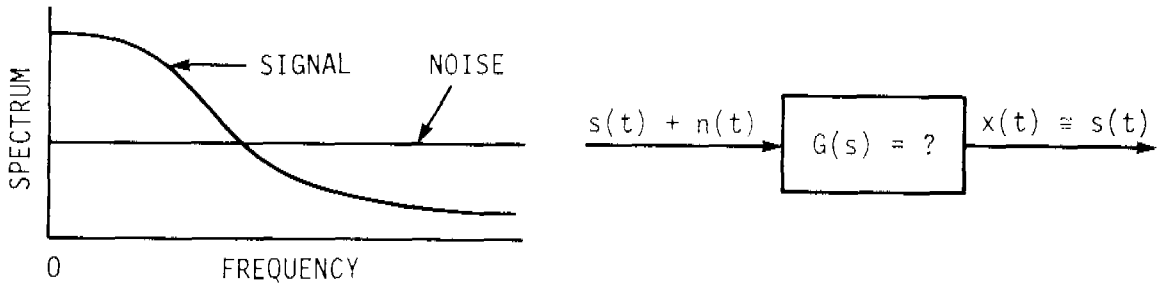


Fig. 3 The classical filter problem (low-pass example)

In the early 1940's, Norbert Wiener considered a different type of filtering problem [7]. Suppose, as shown in Fig. 4, that one has an additive combination of signal and noise with overlapping frequency spectra. We wish to remove the noise from the signal. However, it should be apparent that no filter in the usual sense will do a perfect job of removing noise without destroying the signal. Furthermore, it is not at all obvious what sort of compromise filter response will get rid of most of the noise with minimal damage to the signal. This is the problem that Wiener addressed during the World War II period.



(a) SIGNAL AND NOISE SPECTRA (b) WIENER PROBLEM

Fig. 4 Wiener filter problem

Wiener began by assuming both the signal and noise to be noiselike in character with known spectral characteristics. He then chose minimum-mean-square-error as the performance criterion, and he proceeded to develop a solution for the weighting function (inverse transform of the frequency response) of the filter. The "solution" which is in the form of an integral equation can be solved (with some difficulty) in the stationary (time invariant) case, and Wiener theory is still used in some applications. However, its extension to discrete, multiple-input multiple-output situations is awkward, to say the least. Thus, in a large class of problems the older Wiener methods have been replaced by those of Kalman. Note that the final solution of the Wiener problem is just a transfer function. The classical problem of synthesizing the transfer function still remains after the Wiener problem is solved.

In 1960, R. E. Kalman published a new solution of the least-squares filtering problem [1]. Even though Kalman's methods were much different than those of Wiener, the underlying assumptions were the same. In Kalman's formulation of the problem, the noisy measurements were assumed to be discrete rather than continuous in time, and the signal and noise were modeled in vector rather than scalar form. Kalman then proceeded to develop a solution for the conditional mean of the vector process (i.e., the "signal"), conditioned on all available measurements. In the Gaussian case, this conditional mean also minimizes the mean square error, so the end result is the same as that obtained with the Wiener theory. Kalman's solution is recursive in form, though, which makes it readily ammeanable to programming on a digital computer, either in real time or off line.

In summary, we might characterize the Wiener approach as a scalar weighting function method, whereby the desired estimate of the signal is computed as a weighted sum of past measurements. This procedure has been modified somewhat in recent years to ease the computational burden, but it still is basically a weighting function approach. On the other hand, Kalman's solution was discrete from the outset, and it is characterized by vector modeling of the noise and signal processes and recursive processing of the measurement data. Over the past 20 years, this approach has proved to be remarkably versatile in its ability to accommodate a wide variety of practical filtering problems. There is every reason to expect this activity will continue in the future.

The evolution of linear filter theory from frequency-selective RLC circuits to a set of mathematical equations should now be clear. Once one makes the transition from continuous to discrete measurement, the filter becomes just a prescribed set of arithmetic operations on the sequence of samples. The prescribed operation in the Kalman filter case is developed using the methods of probability and statistics because of the minimum-mean-square-error performance criterion. Thus, at times, the theory looks more like statistics than electrical engineering. Yet, its roots go back to the early days of telegraphy.

3. A SIMPLE AVERAGING EXAMPLE

The recursive philosophy which is essential in Kalman filter theory can be illustrated with a simple example. Suppose we have a sequence of noisy measurements of some unknown constant. We wish to simply average the available measurements and use this as a measure of the unknown constant. Assume that the measurements come to us sequentially in time, and denote them as $z_0, z_1, \ldots, z_k, .$

The last measurement, z_k, is the measurement at current time t_k. We can now compute the average at each point in time in either of two ways:

Batch Processing		Recursive Processing	
Time	Averaging Formula	Time	Averaging Formula
t_o	$(\text{Ave})_o = z_o$	t_o	$(\text{Ave})_o = z_o$
t_1	$(\text{Ave})_1 = \frac{z_o + z_1}{2}$	t_1	$(\text{Ave})_1 = \frac{1}{2}(\text{Ave})_o + \frac{1}{2} z_1$
t_2	$(\text{Ave})_2 = \frac{z_o + z_1 + z_2}{3}$	t_3	$(\text{Ave})_2 = \frac{2}{3}(\text{Ave})_1 + \frac{1}{3} z_3$
	. . .		. . .
	etc.		etc.

Clearly, both methods lead to the same sequence of sample averages. However, as the process progresses, the recursive computation has two distinct advantages over the batch method: (1) The measurements $z_o, z_1, \ldots, z_k$ do not have to be stored individually, and (2) the number of arithmetic operations remains the same with each step. Thus, the recursive approach avoids an escalation of the computational problem as the amount of measurement data increases. This is certainly important in problems involving a large number of measurements. In effect, with recursive processing the new measurement is simply used to update or refine the old estimate. The updated estimate is then projected ahead to the next step, and the process is repeated. This is the basic computational philosophy of the Kalman filter; it is to be contrasted with the batch philosophy, whereby all available measurements are stored and then summed with appropriate weight factors to yield the desired estimate.

4. THE DISCRETE KALMAN FILTER

The theory and details of Kalman filtering can be found in a number of textbooks [8,9,10], so we will only touch on the high points here. In discrete Kalman filter theory, the process to be estimated must be modeled in the following vector form:

$$x_{k+1} = \phi_k x_k + w_k \tag{1}$$

where

x_k = State vector at time t_k

ϕ_k = State transition matrix

w_k = Input white noise sequence characterized by a covariance matrix Q_k

which is defined as

$$E[w_k w_i^T] = \begin{cases} Q_k, & i=k \\ 0, & i \neq k \end{cases} \tag{2}$$

Equation (1) may result from sampling a continuous system driven by white noise. This does not have to be the case, though. Some physical problems are inherently discrete at the outset and there need be no continuous counterpart. The discrete process equation, Eq. (1), stands in its own right.

The measurement relationship connecting the noisy measurement z_k to the state vector x_k must be of the form

$$z_k = H_k x_k + v_k \tag{3}$$

where

z_k = Measurement at time t_k

H_k = Linear connection matrix

v_k = White measurement noise sequence characterized by a covariance matrix R_k

which is defined by

$$E[v_k v_i^T] = \begin{cases} R_k, & i=k \\ 0, & i \neq k \end{cases} \tag{4}$$

and

$$E[w_k v_i^T] = 0, \quad \text{for all } i \text{ and } k. \tag{5}$$

Equations (3), (4) and (5) simply state that there must be a linear connection between the measurement and the process to be estimated, and that the process (which is driven by w_k) and measurement noise must be uncorrelated.

Equations (1) and (3) are sometimes loosely referred to as the Kalman filter "model", and they may seem unduly restrictive at first glance. However, experience of the past 20 years has shown that the model is remarkably versatile in its ability to accommodate a wide variety of physical applications. Formulating the model is, without a doubt, the most difficult part of any applied Kalman filter problem, and there is no one simple rule for doing this. In the case of continuous processes, one must ask the question, "What set of linear differential equations relates the various random processes under consideration to white noise inputs?" Or, saying the same thing another way, "What linear dynamical operations will shape a set of white noise inputs into the processes being considered?" If the appropriate linear dynamical connection can be found, then the problem can be put into state space form, and the discrete form specified by Eq. (1) can be found. Of course, the linear connection between the measurement sequence and the process must also be formulated. This, though, is usually the easier half of the modeling problem.

5. THE RECURSIVE EQUATIONS

We begin by assuming that the estimation problem at hand fits the form given by Eqs. (1) and (3). We then pose the question: What sequence of estimates of the state vector will minimize the mean square error? The recursive solution of this problem is summarized in Fig. 5. Its derivation is given in the tutorial references previously cited, so it will not be included here. However, each major step in the recursive loop deserves further comment.

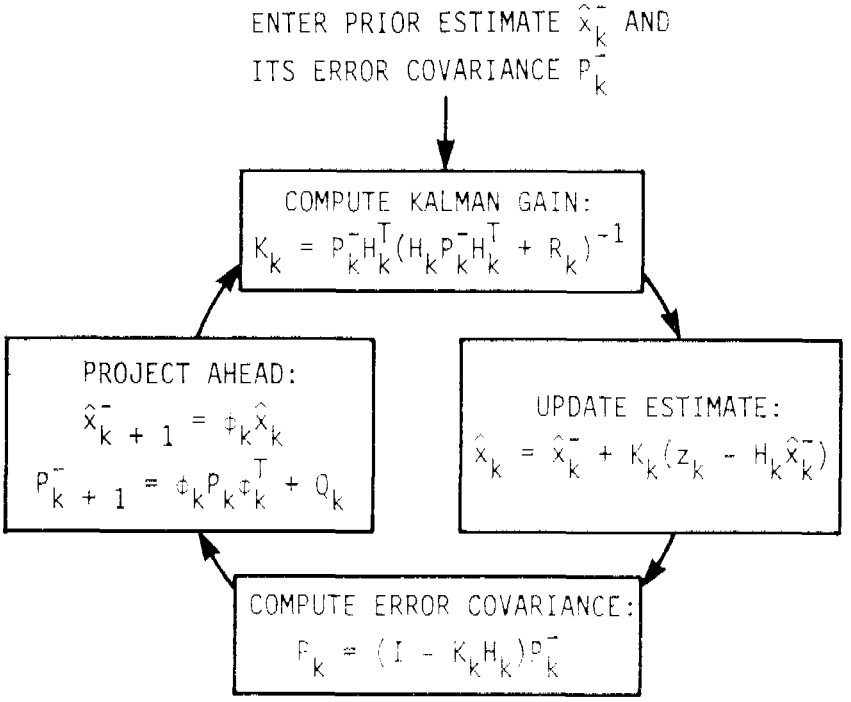


Fig. 5 Kalman filter recursive loop (super minus denotes a priori)

We enter the loop with an initial estimate and its error covariance. The starting point in time occurs with the first measurement, and we usually let this be t=0 (k=0). The initial estimate $\hat{x}_0^-$ is based on our prior knowledge of the process, and this is zero for processes where we have no prior measurements and we know this process mean to be zero. In this case, the initial error covariance P_0^- is just the covariance of the process itself, which is assumed to be known a priori.

The first step in the loop is to compute a matrix known as the Kalman gain. Note from the first block of the loop in Fig. 5 that the Kalman gain depends on the known model parameters H_0, R_0 and the initial P_0^-, but it does not depend on the actual measurement z_0.

The next step in the recursive process is to use the measurement z_0 to update the prior estimate $\hat{x}_0^-$. The equation for doing this is shown in the second block of the loop, and note that it is similar in form with that of the simple averaging example considered in Section 3. That is, the updated estimate $\hat{x}_0$ is formed as the sum of the prior estimate $\hat{x}_0^-$ plus a correction term which is the measurement residual weighted by the Kalman gain.

The third step is to update the prior error covariance P_0^- and obtain the error covariance associated with updated estimate. One can think of P_0 as a measure of the "fuzziness" associated with the updated estimate $\hat{x}_0$. The terms along the major diagonal of P are the variances of the estimation errors for the respective components of the state vector. The off-diagonal terms are the corresponding covariances.

The final step in the recursive loop is to project $\hat{x}_0$ and P_0 ahead to time t_1. This is done via the equations given in the fourth block of the loop shown in Fig. 5. In effect, the state estimate is projected ahead through the natural dynamics of the system as determined by the state transition matrix. Additional uncertainty is added to the projected estimate because of the process noise w_k, and this is accounted for with the Q_k matrix in the error covariance projection.

After the projection step, the estimator is ready to repeat the recursive loop and assimilate the next measurement z_1 at time t_1. If the time between measurements is sufficient to permit all the required computations, the recursive estimation can be done on line. If not, it must be done off line. Two simple scalar examples will now illustrate the use of the recursive equations.

6. KALMAN FILTER EXAMPLES

(a) Estimating an Unknown Random Constant

Suppose we wish to estimate an unknown constant based on a sequence of noisy samples of the constant. For example, this might be the bias on a particular instrument that had just come off the assembly line and is now ready for calibration. Let us say that our past experience with similar instruments tells us that the bias is as likely to be positive as negative, and most of the previous instruments tested were found to have biases in the range of ± 2 units. Based on this crude prior information, it would be reasonable to model the unknown constant (i.e., the bias of the untested instrument) to be a zero-mean Gaussian random variable with a variance of 2^2 units.

A typical sequence of measurements for this examples is shown in Fig. 6.

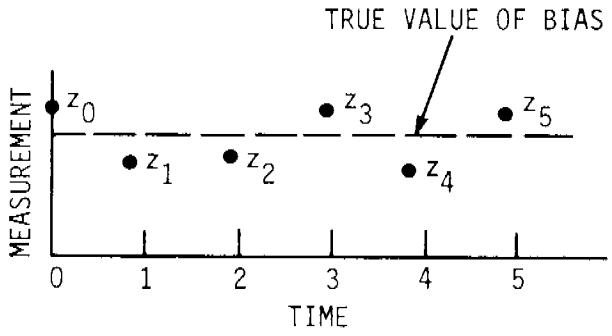


Fig. 6 Constant bias example

Suppose the master instrument used for measurement has a random error of 0.5 units rms, and that this error uncorrelated from step to step in the measurement process. The model parameters can now be specified as follows:

(1) The process satisfies the discrete state equation $x_{k+1} = 1 \cdot x_k + 0$, because x is assumed to be constant. Thus

$$\phi_k = 1$$

$$Q_k = 0$$

(2) We assumed a direct one-to-one noisy measurement of x. Thus

$$H_k = 1$$

(3) We assumed the measurement error to be 0.5 units rms, or a variance of 0.25. Therefore, R_k is

$$R_k = .25$$

Note that ϕ_k, Q_k, H_k, and R_k do not depend on time, so the subscript k could be omitted in this simple example.

As stated previously, prior knowledge of the process led us to assume x to be a zero-mean Gaussian random variable with a variance of 4 units. Thus, the initial conditions are

$$\hat{x}_o^- = 0$$

$$P_o^- = 4$$

We now enter the loop at t=0 and compute the Kalman gain (see Fig. 5).

$$K_o = 4 \cdot 1(1 \cdot 4 \cdot 1 + .25)^{-1} = \frac{16}{17}$$

We update the estimate next.

$$\hat{x}_o = 0 + \frac{16}{17}(z_o - 1 \cdot 0)$$

$$= \frac{1}{17} \cdot 0 + \frac{16}{17} \cdot z_o$$

The P matrix associated with $\hat{x}_o$ is then computed as

$$P_o = (1 - \frac{16}{17} \cdot 1)4 = \frac{4}{17}$$

Finally, we project $\hat{x}_o$ and P_o ahead to the next measurement.

$$P_1^- = 1 \cdot \frac{4}{17} \cdot 1 + 0 = \frac{4}{17}$$

$$\hat{x}_1^- = 1 \cdot \hat{x}_o$$

We are now ready to repeat the process at t_1 and assimilate z_1 in a similar manner, and so forth.

Before leaving this example, it should be noted that the Kalman filter result is not the same as that obtained by simply averaging the measurements (Section 3). The difference arises because the Kalman filter gives the initial a priori estimate nonzero weight when blending it with the first measurement z_o. This will, of course, cause the a priori estimate to propagate indirectly into the subsequent estimates as the recursive process proceeds. This distinguishes the Kalman filter from maximum likelihood estimation, where we usually assume that no a priori information is available. To make a Kalman filter artificially look like maximum likelihood estimation, all one has to do is make the initial P matrix very large. Then the initial estimate is given zero weight on the first step, and all subsequent estimates will depend only on the measurement sequence.

(b) Brownian Motion (Wiener) Process

Figure 7 shows a noise process which is being generated as the output of an integrator driven by Gaussian white noise. This experiment can be readily demonstrated in the laboratory just by connecting the output of a wideband noise source to an analog integrator with the initial condition being set at zero. Such a process is known as a Brownian-motion or Wiener process. Clearly, x(t) satisfies the first order differential equation

$$\dot{x} = f(t) \qquad (6)$$

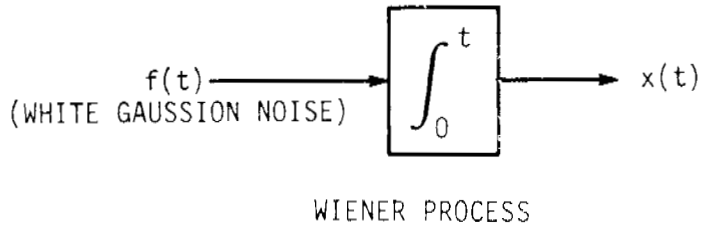


Fig. 7 Generation of a Wiener process

Thus, discrete samples of x(t) are related by the recursive equation

$$x_{k+1} = x_k + w_k \tag{7}$$

where

$$w_k = \int_{t_k}^{t_{k+1}} f(t)dt \tag{8}$$

Clearly, if f(t) is white noise, the w_k sequence will be white and the process model fits the required format.

Suppose we have a sequence of noisy measurements of this process as shown by the dots in Fig. 8. Further, suppose that the rms measurement error is 0.5, and that

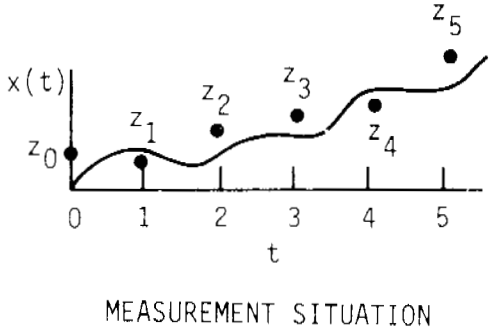


Fig. 8 Typical measurement situation for random walk example

w_k for a sampling interval of 1 sec is normal with zero mean and unit variance. The model parameters are then

$$\phi_k = 1$$

$$Q_k = 1$$

$$H_k = 1$$

$$R_k = (.5)^2 = .25$$

The initial conditions in this case are

$$\hat{x}_o^- = 0$$

$$P_o^- = 0$$

Note that the P_o^- condition is unusual. It says that we know the initial process state at t=0 perfectly! This is due to initial zero condition on the integrator; we know its output cannot change instantaneously. To see how this unusual P_o^- affects the estimator, we cycle through the recursive loop at t=0 (see Fig. 5):

Calculate gain: $K_o = 0 \cdot 1(1 \cdot 0 \cdot 1 + .25)^{-1} = 0$

Update estimate: $\hat{x}_o = 0 + 0 \cdot (z_o - 1 \cdot 0) = 0$

Update P: $P_o = (1 - 0 \cdot 1) \cdot 0 = 0$

Project ahead: $\hat{x}_1^- = 1 \cdot 0 = 0$

$$P_1^- = 1 \cdot 0 \cdot 1 + 1 = 1$$

Note that the Kalman filter gives the noisy measurement z_o zero weight. This is just as it should be; it is worthless relative to our perfect knowledge of the state at t=0.

It is instructive now to cycle through one more step of the recursive process. At t=1:

Calculate gain: $K_1 = 1 \cdot 1(1 \cdot 1 \cdot 1 + .25)^{-1} = \frac{4}{5}$

Update estimate: $\hat{x}_1 = 0 + \frac{4}{5}(z_1 - 1 \cdot 0) = \frac{4}{5} z_1$

Update P: $$P_1 = \left(1 - \frac{4}{5} \cdot 1\right) = \frac{1}{5}$$

Project ahead: $$\hat{x}_2^- = 1 \cdot \hat{x}_1$$

$$P_2^- = 1 \cdot \frac{1}{5} \cdot 1 + 1 = \frac{6}{5}$$

Notice now that the a priori estimate at t=1 is only given a weight of 1/5, and the measurement receives a weight of 4/5. This is due to the process "random walk" that takes place in the interval from t=0 to t=1. The a priori estimate is thus quite uncertain at t=1, and the measurement, even though noisy, contains valuable new information about the process.

The recursive data processing can now be continued on ad infinitum. It might be mentioned that this least-squares estimation problem works out to be elementary when viewed from the Kalman viewpoint. It is somewhat elusive, though, when viewed from the Wiener viewpoint because of the nonstationary character of the Wiener process.

7. CLOSING COMMENT

We will close with these two simple examples, knowing that there are a number of application papers to follow in this session of the PTTI meeting. Hopefully, this tutorial overview will be of help in understanding the subsequent papers.

References

[1] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", Trans. of the ASMEE--Jour. of Basic Engr., pp. 35-45, March 1960.

[2] S. F. Schmidt, "Application of State-Space Methods to Navigation Problems, in Advances in Control Systems, Vol. 3, C. T. Leondes (Ed.), New York, Academic Press, 1966.

[3] R. G. Brown and D. T. Friest, "Optimization of a Hybrid Inertial Solar-Tracker System", IEEE International Convention Record, Part 7, pp. 121-135, 1964.

[4] B. E. Bona and R. J. Smay, "Optimum Reset of Ship's Inertial Navigation System", IEEE Trans. on Aerospace and Electronic Systems, AES-2, No. 4, pp. 409-414, July 1966.

[5] Bell System Technical Journal, January 1982.

[6] IEEE Trans. on Automatic Control, Vol. AC-28, No. 3, March 1983.

[7] N. Wiener, Extrapolation, Interpolation and Smoothing of Stationary Time Series, New York, John Wiley, Inc., 1949.

[8] R. G. Brown, Introduction to Random Signal Analysis and Kalman Filtering, New York, John Wiley, Inc. 1983.

[9] A. Gelb (Ed.), Applied Optimal Estimation, Cambridge, MA, The M.I.T. Press, 1974.

[10] P. S. Maybeck, Stochastic Models, Estimation and Control, Vol. 1, New York, Academic Press, 1979.

QUESTIONS AND ANSWERS

MR. BROWN:

What you always wanted to know about Kalman Filtering but were afraid to ask, or something like that.

MR. McCONAHY:

Mac McConahy, JHU/Applied Physics Lab. Grover would you like to say a few words about one of the things that I think rather puzzles many people about Kalman Filters, and that is, 'choosing' the process noise co-variance matrix in a practical application?

MR. BROWN:

Okay. The process noise co-variance matrix, that is the Q-matrix you're talking about. Well, I got to go clear back here, I'm afraid. Can I get the projector turned on again, please?

Going back to this particular model here, you have to have the characterization of W_K and, of course, the Q-matrix is W_K times W_K transposed and then the expectation of that; so it is a covariance matrix, and what you have to do is describe this in terms of the dynamics of the system. Usually, you end up getting this equation by starting out with some kind of continuous dynamics that relate the state vector to white noise inputs. Now, if you have that, if that's the starting point, then you can write a set of dynamical--you know, you can write it out explicitly what this W_K is; usually write it out as a convolution integral. Now, you take W_K transpose times W_K, take the expectation of that, and you end up having to evaluate a whole bunch of double integrals, is what it amounts to; and I guess I can't say more than that, except that is one of the most difficult parts of the problem. It's a very doggy job.

One of the sneaky approaches is the risk of adding something to that. There is a sneaky approach sometimes. If the step size is relatively large, sometimes the effective way of getting the Q-matrix is to subdivide the large integral into infinitesimal integrals, approximate the Q_K for the infinitesimal integral as being a diagonal matrix. No, not a diagonal matrix, but a matrix consisting of only first order terms of delta t.

Then you cycle through a whole bunch of steps to find the Q for the whole integral, just using the projection on the P-matrix form.

Use $Q\emptyset Q^+$, that extra term, and you keep cycling through that. That is sort of a vague description, but the Q-matrix is kind of hard to find.

MR. WEISS:

Mark Weiss, N.B.S. To what extent does using a Kalman Filter assume a certain form of a noise, such as excluding flicker noise?

MR. BROWN:

I'll simply say this, if this is a bit of a cop-out that will apply to any situation if you can make it fit the model that is up there now. Now, there are certain cases where you can always fit it into this particular form. If you start out with the processes of--if you start out with stationary processes that are describable in terms of--well, where the spectral characteristics are rational. Then you can always choose the state vector to describe that which is such that it's the result of putting input white noise into some kind of linear dynamics to give that spectral characteristic. So any time the spectral characteristics are describable by rational spectral functions, then your in business as far as the Kalman Filter is concerned--. But that is not the only case. You can also handle non-stationary cases.

You can model any case where the process that you are talking about can be thought of as the result of--let's see what I want to say--putting white noise inputs into a system of linear dynamics.

MR. WEISS:

Would you handle flicker noise?

MR. BROWN:

No, I don't think so. You see, that's something you people know a lot more about than I do, but it appears to be that that's a case where you have a spectral function, which is not rational. It's fractional powered, and I don't know how to do that.

DR. BARNES:

You can approximate it with an empirical approach.

MR. BROWN:

If it's truly a spectral function where it involves fractional powers of Omega, then I don't know any way to model it exactly, using the Kalman Filtering methods.