KALMAN FILTER MODELING

R. Grover Brown
Electrical and Computer Engineering Department
Iowa State University
Ames, Iowa 50010

ABSTRACT

The main emphasis of this tutorial paper is on the formulation of appropriate state-space models for Kalman filtering applications. The so-called "model" is completely specified by four matrix parameters and the initial conditions of the recursive equations. Once these are determined, the die is cast, and the way in which the measurements are weighted is determined foreverafter. Thus, finding a model that fits the physical situation at hand is all important. Also, it is often the most difficult aspect of designing a Kalman filter. Formulation of discrete state models from the spectral density and ARMA random process descriptions is discussed. Finally, it is pointed out that many common processes encountered in applied work (such as band-limited white noise) simply do not lend themselves very well to Kalman filter modeling.

INTRODUCTION

Kalman filtering is now well known, and tutorial discussions of the technique are given in a number of standard references [1,2,3]. The filter recursive equations are summarized in Figure 1 for reference purposes here. It should be noted that once the initial conditions and the ϕ_k, H_k, R_k, Q_k, parameters are specified, the die is cast and the way in which the measurement sequence is processed is completely determined. Thus, the specification of these parameters is especially important -- they are, in effect, the filter "model". The emphasis in this tutorial paper will be on the modeling aspect of Kalman filtering. To see where these parameters come from, we will now review the basic process and measurement equations.

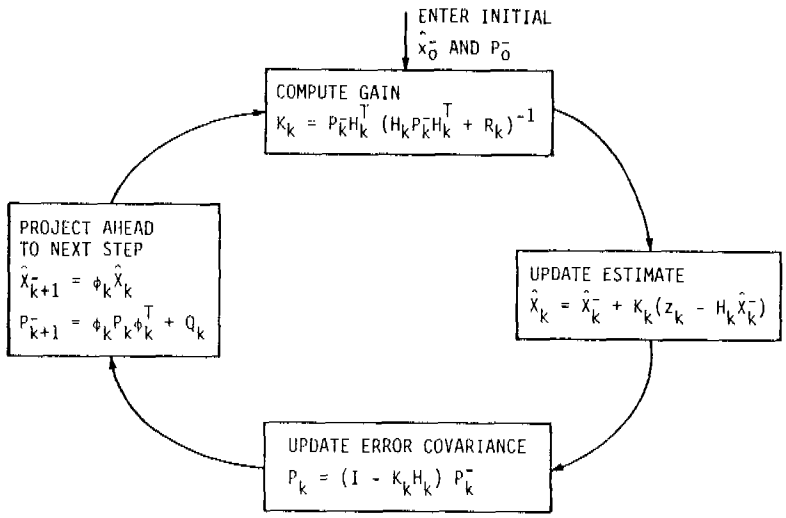

Figure 1 Kalman filter loop

THE DISCRETE PROCESS AND MEASUREMENT EQUATIONS

The starting point for discrete Kalman filter theory begins with the process and measurement equations. The random process under consideration is assumed to satisfy the following recursive equation

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

where k refers to the k-th step in time, x_k is a vector random process, ϕ_k is the transition matrix, and w_k is a Gaussian white sequence with a covariance structure given by

$$E[x_k x_k^T] = Q_k \tag{2}$$

The measurement relationship is assumed to be of the form

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

where v_k is also a Gaussian white sequence, uncorrelated with w_k, and described by the covariance

$$E[v_k v_k^T] = R_k \tag{4}$$

In words, then, the key parameters of a Kalman filter model can be described as follows:

(1) ϕ_k is the transition matrix that describes the natural dynamics of the process in going from step k to k+1.

(2) H_k is the linear connection matrix that gives the ideal (noiseless) relationship between the measurement z_k and the process to be estimated x_k.

(3) Q_k describes the additional noise that comes into the x_k process in the Δt interval between step k and k+1.

(4) R_k describes additive measurement noise.

It is important to note that the discrete model described by Eqs. (1) through (4) stands in its own right. It is not an approximation of some continuous system, nor does it have to be related to another continuous linear dynamical system in any way. Once the discrete model is assumed, the recursive estimation process given in Fig. 1 follows directly.

IMPORTANCE OF THE GAUSSIAN ASSUMPTION

We will digress for a moment and look at the Gaussian assumption used in Eqs. (1) through (4). If w_k and v_k are Gaussian white sequences, then x_k and z_k will be Gaussian processes. Even though the Gaussian assumption is often omitted in discussions of least-squares filtering, we make here with no apology. The reason for this is that minimizing the mean square error really does not make very good sense for non-Gaussian processes. To illustrate this, consider the two processes shown in Fig. 2. The first is a scalar Gauss-Markov process which has the general appearance of typical noise. The second process is the random telegraph wave which switches between +1 and -1 at random points in time. If the parameters of the two processes are adjusted appropriately, they can be made to have identical power spectral density functions. Yet, they are radically different processes! The least-squares prediction far out into the future is zero for both cases. This makes good sense in the Gauss-Markov case because zero is the mean and most likely value. On the other hand, it is ridiculous to predict zero in the random telegraph wave case. We know a priori that this waveform is never zero. We would be better off to predict either +1 or -1 and be correct half the time than to predict zero and be wrong all the time! Thus, the Gaussian assumption is a reasonable one in the least squares estimation theory, and to stray from it leads us into dangerous territory.

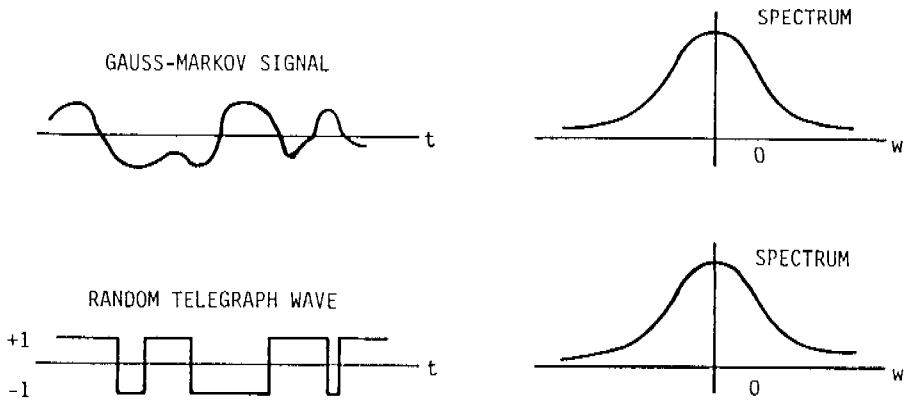

Figure 2 Gauss-Markov and random telegraph waves

TRANSITION FROM A SPECTRAL DESCRIPTION TO A DISCRETE STATE MODEL

In Kalman filter applications, we frequently begin with a spectral description of the various random processes involved. The problem then is to convert this information to a model of the form specified by Eqs. (1) through (4). The general procedure for making the transition to the discrete model is as follows:

(1) Look for a continuous dynamical system that yields the desired process when driven by white noise. (The white noise input assures that w_k will be a white sequence.)

(2) Then write the dynamical equations in state-space form:

$$\dot{x} = Ax + Bu \tag{5}$$

(3) Solve the state equations for step size Δt and obtain

$$x_{k+1} = \phi_k x_k + w_k \tag{6}$$

(4) Determine the measurement equation from the particular situation at hand.

To illustrate the procedure further, suppose the y process power spectral density function $S_y(s)$ can be written as a ratio of polynomials in s^2 (or ω^2, where $\omega^2 = -s^2$). The spectral function can then always be factored into two symmetric parts, one with its poles and zeros in the left-half s plane, the other with mirror-image poles and zeros in the right-half plane. This is called spectral factorization and is represented mathematically as

$$S_y(s) = S_y^+(s) \cdot S_y^-(s) \tag{7}$$

where S_y^+ and S_y^- are the left- and right-half plane parts respectively. $S_y^+(s)$ then becomes the shaping filter that will shape unity white noise into a process y(t) with a spectral function $S_y(s)$. (See Ref. [1] for further details.)

Now suppose that the shaping filter is of the form shown in Fig. 3. We seek a state-space model for that dynamical system. One way of achieving this is

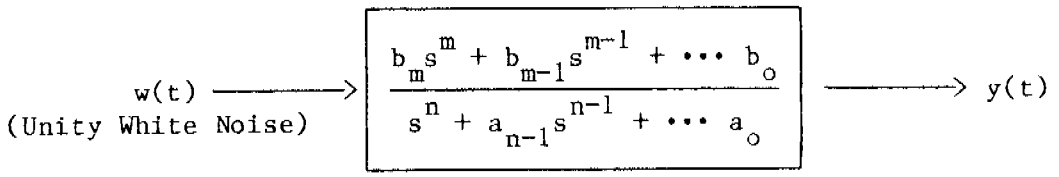

Figure 3 Shaping filter

shown in block diagram form in Fig. 4. The state-space equations are then

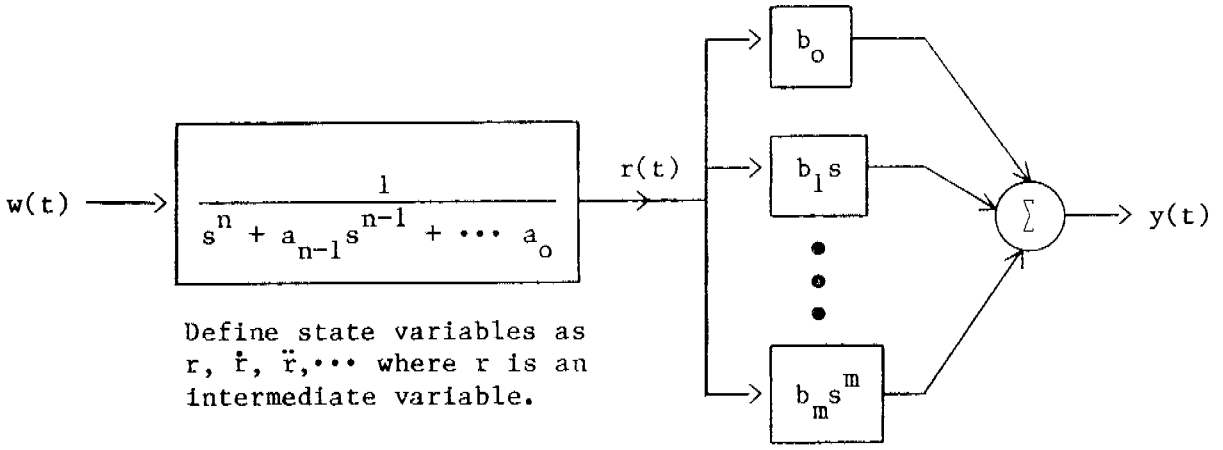

Figure 4 Shaping filter redrawn

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \cdot \\ \cdot \\ \cdot \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 1 & \cdot & \cdot & \cdot \\ & & \cdot & & & & \\ & & \cdot & & & & \\ -a_o & -a_1 & \cdot & \cdot & \cdot & & -a_{n-1} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} w(t) \tag{8}$$

$$y = [b_o \quad b_1 \quad \cdots \quad b_{n-1}] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} \tag{9}$$

Control system engineers refer to this as the controllable canonical form, and it can always be achieved for the dynamical system as shown in Fig. 3. If y is the process that is actually measured, then the H matrix is just the row matrix of b's given in Eq. (9).

EXAMPLE

Suppose we have a scalar Gauss-Markov process y(t) whose power spectral density function is

$$S_y(s) = \frac{2\sigma^2\beta}{-s^2 + \beta^2} \quad \left(\text{or } \frac{2\sigma^2\beta}{\omega^2 + \beta^2}\right) \tag{10}$$

We first factor S_y as follows:

$$S_y(s) = \frac{\sqrt{2\sigma^2\beta}}{s + \beta} \cdot \frac{\sqrt{2\sigma^2\beta}}{-s + \beta} \tag{11}$$

The shaping filter is then $\sqrt{2\sigma^2\beta}/(s+\beta)$ which corresponds to the dynamical equation

$$\dot{y} + \beta y = \sqrt{2\sigma^2\beta}\ w(t) \tag{12}$$

This is a simple first order differential equation, so we only have one state variable. Call it x_1. Our state equation is then

$$\dot{x}_1 = -\beta x_1 + \sqrt{2\sigma^2\beta}\ w(t) \tag{13}$$

The solution of this equation for a step size Δt is

$$x_{k+1} = e^{-\beta\Delta t} x_k + w_k \tag{14}$$

and $e^{-\beta\Delta t}$ can be seen to be the transition matrix ϕ_k. The mean square value of w_k can be determined from random process theory [1], and it works out to be

$$Q_k = E[w_k^2] = \sigma^2(1-e^{-2\beta\Delta t}) \tag{15}$$

The process model is now complete.

UNIQUENESS

We might pose a question at this point:

Are Kalman filter models unique?

The answer is an emphatic NO. We know from linear system theory that any nonsingular linear transformation on the state vector leads to another equally legitimate state vector. The choice of coordinate frame for performing the estimation process is purely a matter of convenience. Optimal estimates can be transformed freely from one coordinate frame to another (through a linear transformation) and still remain optimal estimates in the new frame of reference.

ARMA MODEL

Sometimes the random process model comes to us in the form of a difference equation rather than a continuous differential equation. For example, consider the auto-regressive moving average (ARMA) model that relates a discrete process y(k) to an input white sequence u(k).

$$y(k+n) + \alpha_1 y(k+n-1) + \cdots \alpha_n y(k) = \beta_1 u(k+n-1) + \cdots \beta_n u(k) \tag{16}$$

There is a close analogy between difference and differential equations, and it works out that this nth-order difference equation can be converted to vector form in much the same manner as for a differential equation. If we define an intermediate variable $y'(k)$ as the solution to Eq. (16) with just u(k) as the driving function, and then define our state variables as

$$x_1(k) = y'(k), \quad x_2(k) = y'(k+1), \quad \text{etc.} \tag{17}$$

then the system of Eq. (16) translates into state-space form as

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ \cdot \\ \cdot \\ \cdot \\ x_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdot & \cdot & \cdot \\ & & \cdot & & & & \\ & & \cdot & & & & \\ & & \cdot & & & & \\ -\alpha_n & -\alpha_{n-1} & \cdot & \cdot & \cdot & & -\alpha_1 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \\ \cdot \\ \cdot \\ \cdot \\ x_n(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} u(k) \tag{18}$$

$$y(k) = [\beta_n \quad \beta_{n-1} \quad \cdots \quad \beta_1] \begin{bmatrix} x_1(k) \\ x_2(k) \\ \cdot \\ \cdot \\ \cdot \\ x_n(k) \end{bmatrix} \tag{19}$$

Note that our choice of state variables leads to the controllable canonical form, just as in the continuous dynamical case. Of course, we could have defined our state variables differently and arrived at a form different from Eqs. (18) and (19). We will not pursue this further other than to say the choice of state variables is (within limits) a matter of convenience for the situation at hand.

PROCESSES DERIVED FROM IRRATIONAL SHAPING FILTERS

The random process modeling procedures discussed thus far have been straightforward. They may be tedious for higher-order processes, but they do not call for much imagination. There exists, however, a whole class of processes where this is not the case. These are the processes that cannot be thought of as the result of passing vector white noise through a linear dynamical system of finite order. Such processes are commonplace in engineering literature. For example, bandlimited Gaussian white noise is a very useful abstraction in communication theory. It is Gaussian noise that has a flat spectrum in the baseband and then is zero out beyond the cutoff frequency. It can be thought of as the result of passing pure white noise through an idealized lowpass filter, but no such filter can be represented as a ratio of polynomials in s of finite order. (Note that a Butterworth filter can be made to approximate the ideal case, but not equal it.) The

idealizations of bandlimited white noise are often a convenience in communication theory; however, they are an obstruction in Kalman filter theory.

There is a theorem from linear systems theory that is useful at this point. Chen [4] gives us the following criterion for the realization of linear dynamical models.

A linear dynamical model of the form

$$\dot{x} = Ax + Bu$$
$$y = Cx + Du \tag{20}$$

will exist for a system with an input-output impulsive response $G(t,\tau)$, if and only if, $G(t,\tau)$ is factorable in the form

$$G(t,\tau) = M(t)N(\tau) \tag{21}$$

M and N are finite-order matrices, so if $G(t,\tau)$ is scalar (i.e., single-input, single-output), M(t) is a row vector and N(t) is a column vector. This theorem can then be used as a test to see if a dynamical system will exist for a corresponding impulsive response function. Furthermore, the factorization provides the necessary information for realization of the model. (See Chen [4] for further details.) We will use flicker noise to illustrate the use of Chen's theorem. Flicker noise is of special interest to the PTTI community because of its presence in precision frequency standards. It is characterized by a power spectral density function of the form of $1/f$ at the frequency level, or $1/f^3$ when referred to the phase level [5,6]. A block diagram showing the relationship between flicker noise and white noise is given in Fig. 5.

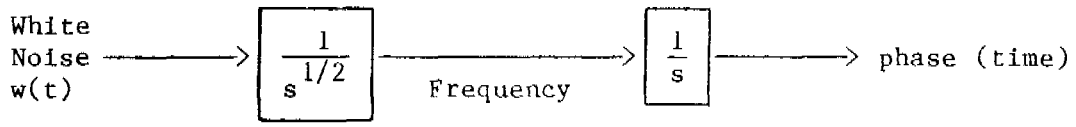

Figure 5 Block diagrams relating flicker noise to white noise

Clearly, the transfer function relating input white noise to the output phase x(t) is $1/s^{3/2}$. The inverse transform of $1/s^{3/2}$ gives the impulsive response to an impulse applied at t=0. This is $2\sqrt{t}/\sqrt{\pi}$. Thus, for an impulse applied at $t=\tau$, we have (in Chen's notation)

$$G(t,\tau) = \frac{2}{\sqrt{\pi}}\sqrt{t-\tau}\ , \qquad t > \tau \tag{22}$$

The question is, "Is $G(t,\tau)$ factorable in the form $M(t)N(\tau)$?" It appears that it is not, although this is difficult to show in a rigorous sense. This being the case, Chen's theorem says that no linear dynamical system will exist that corresponds to the $G(t,\tau)$ of Eq. (22). This is to say that no finite-order state model will exactly represent flicker noise! Of course, the state model is essential for Kalman filtering, so this leads to a dilema when one attempts to include flicker noise in a Kalman filter clock model. This is the subject of a companion paper in these Proceedings [6], so we will not pursue this further here.

SUMMARY

Various aspects of Kalman filtering modeling have been discussed briefly in this paper. Perhaps the most important thing to remember is that the random processes under consideration must be modeled in vector state-space form. This can often be done with exact methods. If the exact methods discussed here cannot be used, as in the case of flicker noise, then one must seek approximate finite-order vector models in order to form a workable Kalman filter. The measurement model usually does not cause difficulty, because it simply depends on what state variables are being observed.

REFERENCES

[1] R. Grover Brown, Introduction to Random Signal Analysis and Kalman Filtering, John Wiley, Inc., 1983.

[2] P. S. Maybeck, Stochastic Models, Estimation and Control, (Vol. 1), Academic Press, 1979.

[3] A. Gelb (Ed.), Applied Optimal Estimation, MIT Press, 1974.

[4] C. T. Chen, Introduction to Linear System Theory, Holt, Rinehart and Winston, 1970.

[5] J. A. Barnes and D. W. Allan, "A Statistical Model of Flicker Noise", Proc. of the IEEE, Vol. 54, No. 2, February 1966.

[6] A. J. Van Dierendanck, J. B. McGraw and R. Grover Brown, "Relationship Between Allan Variances and Kalman Filter Parameters", Proceedings of this Conference (1984).

VICTOR REINHARDT, HUGHES AIRCRAFT COMPANY: I think you are right about that not being able to be factored, and I think that I have a reason for that. You can show that flicker noise can be mathematically generated by the sum of an infinite number of gaussian processes where the beta term goes from zero to infinity. Therefore, there are infinite time constants in the process. So, you can't give a state vector at any one time, because the beta term goes from zero to infinity.

MR. BROWN: I agree with what you say. I think that it fits my intuition to think the same thing, and I have read that paper that you wrote on it. I think that it's a very nice paper, and a nice way to look at it.

Other people have also approximated flicker noise with a cascaded sequence of what we, in control system engineering, call lead or lag networks, which gives kind of a staircase sort of frequency response function, which, to a certain degree of approximation, drops off at ten dB per decade rather than twenty dB.

If you take any rational transfer function, or one that is written out in integer powers, and look at the Bode plot, the slopes go in multiples of twenty dB per decade, There are no thirty dB per decade, or fifty dB per decade slopes.

In the case of flicker noise, and consider the filter that shapes white noise into flicker noise, it requires an s to the negative one-half power in the transfer function. That would give a Bode plot that drops off at ten dB per decade instead of twenty. What you would do is approximate that ten dB per decade slope with a whole sequence of filters with alternating zeros and poles. You then end up with a staircase shape response which, on the average, has a ten dB per decade slope.

Incidentally, I think that this is a very good way to model flicker noise. The difficulty is that every time you put a new pole in the system you have a new state model. If you want get a reasonably accurate approximation of flicker noise that way, it does involve escalating the order of the Kalman filter considerably. There is nothing wrong with doing it off-line for analysis purposes. I think that there are some on-line cases where it would not be accepted.

MR. REINHARDT: I think that some people have reported on a similar method where they used a finite number of filters and it worked very well in an operational case. If you try to limit that process though, what happens is that all the poles run together, and you end up with a branch line.

MR. BROWN: I guess my answer to that would be that, in any of these processes, in the case of flicker noise for example, at zero frequency and out at infinity, there are singular conditions for either case. If it drops off as one over f, the area under the curve out at infinity is not finite. You are talking about a process with infinite variance, which is physically ridiculous.

The same thing happens at the other end of the spectrum, the

area under the curve doesn't converge there, either. Physically it makes sense, if you want to be careful and talk about processes of finite variance, that you have to bound the power spectral density at the low frequency end and at the high frequency end. It has to roll off at least twenty dB per decade in order to have a process of finite variance.

It doesn't bother me to think of putting in a filter at the origin which will bound the frequency content at zero frequency, and also put one in at the high end and make it roll off at least twenty dB per decade.

Incidentally, that impulse response function is not original with me. Other people have written about that before, including yourself, I think.

JIM BARNES, AUSTRON, INC.: I have done a fair amount of simulation of flicker noise with polynomials, the lead-lag networks you mentioned, and have one comment in their defense: Three or four stages can do an amazing amount. You can cover as much as three to four decades of frequency with only three or four stages.

MR. BROWN: Oh, is that right? It isn't as bad as it might appear at first glance then. I haven't used it, but would have imagined that you would need a fairly large number.

MR. REINHARDT: As another comment, even a single filter, which generates a random telegraph, will generate a flat Allan variance of about two orders of magnitude in tau, right around the peak. Then you really have to put a pole every order of magnitude or even every two orders of magnitude.

MR. BROWN: All of these are, of course, approximate models for the reasons which I just cited.

MR. ALLAN: I think, in practice, the problem with flicker noise is not a serious one, because it's only at the extremes, as you pointed out, at zero and at infinity that you have difficulties with one over f integration. In practice, that's not where the Fourier frequencies are. In reality, a few stages of the filter will work very nicely in describing, predicting or simulating a flicker process.

MR. BROWN: You need something like that though as far as the Kalman filter is concerned. You can't afford to have these fractional powers of s is you are going to do the state model. You have to have something where you only need to worry about integer powers of s, and if you can do that by only adding two or three poles, that would be a very feasible way to approximate it certainly.