International Journal of
Georesources and Environment
Journal home: http://ijge.camdemia.ca. Articles at http://ojs.library.dal.ca/ijge
Geo-Resource Research
IJGE 2021 **6:** 83

The Composite Covariance of a Kalman Filter Estimation

Winston C Chow
Clermont, FL

Abstract: A Kalman filter estimation of the state of a system is merely a random vector that has a normal, also called Gaussian, distribution. Elementary statistics teaches any Gaussian distribution is completely and uniquely characterized by its mean and covariance (variance if univariate). Such characterization is required for statistical inference problems on a Gaussian random vector. This mean and composite covariance of a Kalman filter estimate of a system state will be derived here. The derived covariance is in recursive form. One must not confuse it with the "error covariance" output of a Kalman filter. Potential applications, including geological ones, of the derivation are described and illustrated with a simple example.

Keywords: Statistical inference, Kalman filter, composite covariance, mean, estimation, Gaussian, process noise, measurement noise, probabilistic distribution, prediction, unbiased

1. Introduction

Analysts often use the Kalman filter (KF) to estimate the state of a system from measurements of observations of this state that are perturbed by Gaussian (normally distributed) white process noise. Since the estimated state of a system, based on measurements, is, itself, a random variable or vector, as well as a stochastic process over time, analysts may wish to also characterize the probabilistic distribution (that is, the statistics) of the estimate, itself, in order to understand its random behavior. Since noise is assumed to be Gaussian (normally distributed), the estimated state, a linear function of Gaussian noises, is also Gaussian (Hogg and Craig 1978). Being Gaussian, it has the convenient mathematical property that its statistics is completely determined by the mean vector and covariance matrix (variance if univariate) (Hogg and Craig 1978). We derive here the covariance and mean of the KF state estimate, as a Gaussian distributed random vector/variable. A recursive formula for the covariance will be derived as well as the formula for the mean.

Given initial conditions, a formula for a type of covariance is already in the KF. It is recursive and represents the estimation error covariance, that is, the covariance of the difference between the true state of the system and the KF estimated state (Meditch 1969). It is a priori (does not depend on measurement data). This covariance is

$$P(k|k) \equiv cov\left(\hat{X}(k|k) - X(k)\right),$$

where $\hat{X}(k|k)$ is the KF state estimate at time $k = 0,1,2,\cdots$, given k measurements, initializing at

$$\hat{X}(0|0) = E[X(0)] \equiv x_0.$$

Note E is the symbol for expected value. The $X(k)$ represents the actual state at time k. Knowledge of this covariance is used to determine the statistical behavior of estimation error. Thus, it gives information on the KF state estimation's reliability.

Although the error covariance of the KF measures the random variation of the state estimator from the actual random state, it does not give the covariance around the deterministic mean function of the estimator. The mean can be interpreted to be the deterministic state model or the prediction model prior to observations, as we will show later that the mean of the estimate equals the mean of the state. This "composite" covariance, not given in the KF, of the estimator, itself, along with its mean, according to statistical theory, are required to determine the true statistics of the estimator, being Gaussian. We shall derive this composite covariance here, as it is not found in textbooks. We also derive here the formula for the mean vector.

Notice this new covariance is completely different from the KF output error covariance $P(k|k)$. Recalling error covariance is

$$P(k|k) \equiv cov\left(\hat{X}(k|k) - X(k)\right)$$

by the definition of covariance, this must be

$$E\left[\left(\hat{X}(k|k) - X(k) - 0\right)\left(\left(\hat{X}(k|k) - X(k) - 0\right)\right)^T\right]$$

where the superscript T represents matrix inverse and with the 0's unnecessarily inserted to show the mean of the estimate error, $\hat{X}(k|k) - X(k)$, is hypothesize as 0. Yet our

* Corresponding Author: Winston C Chow, Mathematician/Statistician, Retiree, Clermont, FL,

Received Oct. 6, 2020; revised March 15, 2021; accepted Sept. 10, 2021.

DOI: 10.15273/ijge.2021.06.083

new KF estimate covariance we will derive here is $cov\left(\hat{X}(k|k)\right)$, which, in contrast, must be

$$E\left[\left(\hat{X}(k|k) - E\left[\hat{X}(k|k)\right]\right)\left(\hat{X}(k|k) - E\left[\hat{X}(k|k)\right]\right)^T\right]$$

Notice how this very much differs from $P(k|k)$. These formulas will be restated later when they are used in explanations. We also show later that

$$E\left[\hat{X}(k|k)\right] = E[X(k)],$$

that is, the estimate is unbiased.

2. Proposed Applications

Along with the application to mathematical statistics that knowing the exact distribution of a KF estimate gives mathematical completeness from a statistical perspective, we now describe how this knowledge is useful in the analysis of estimations of physical systems. Such systems surely include geology or geological engineering analysis such as performing estimations on seismic measurement data or observation data from radioactivity coming from the ground (or elsewhere). We briefly describe below.

One useful physical application follows, as the KF is an "unbiased" state estimator, which we show later. Thus, a KF estimation mean equals the actual state mean, which can also be interpreted, given the initial state, as the predicted state before measurement. A covariance shows how much a random variable or process varies from its mean. Since, being unbiased, the mean of the KF is the same as the state mean, this composite covariance is useful in determining how closely a KF estimation follows the state mean, that is, the predicted state. It shows how much unpredictability due to randomness is inherit to the chosen application of the KF for each given time. If the covariance matrix shows large variances for some components of the state vector at any given time, one concludes the estimated state will probably vary a lot from the deterministic predicted model at these state components. As random disturbances of the state, modeled by process noise, is cumulative, the deviation of the estimate from the prediction will, generally speaking. grow with time. For example, if a state is disturbed by air, this disturbance always exists and causes the state to veer off the predicted more and more with time. Our covariance gives user ideas of when to expect the estimate to no longer resemble the prediction. Some analysts may like some anticipations of when may estimates no longer resemble predictions. As our covariance as well as the error covariance do not use actual measurement data, this analysis can be done a priori giving ideas of what to expect before actual observations. This extra piece of information previously unavailable may prove useful to some analysts.

In state estimation, including geological ones, this covariance can be used whenever Kalman filters are used. Kalman filters are useful in situations when the user has a state model and a measuring device. Given the state model and the measurement, a Kalman filter estimates the state. Our composite covariance, as mentioned, gives a priori ideas of how much and when actual estimates will deviate from predictions (which actually is also the state model without the random process noise) up to a point where it no longer resembles the prediction.

An article by Kitagawa, Takanani, and Matsumoto in the book Time Series Analysis and Applications to Geophysical Systems (Brillinger et al. 2004) describes how seismological data can use state space models and measurement devices such as seismometers. The article stated seismic signals are contaminated by process noise from natural (and human) background disturbances such as the air and numerous other sources. With the existence of state models and measuring devices, the KF is useful to estimate the state. According to the article, the states to be estimated by Kalman filters may actually be time varying parameters in a time series model. Parameters of states, to be analyzed as states themselves, for Kalman filters may include seismic signals or "coseismic" effects from groundwater. This must be extracted in order to observe signals from earthquakes that the seismic device is measuring. Given KFs are utilized here, we can use our composite covariance to conjecture how much estimates will deviate from the predicted model and when will deviations grow so large it no longer resembles predictions.

We will give a simple numerical example illustrating our proposed application immediately following the derivations of our estimation covariance and mean. This will help clarify the above description.

One also may wish to know whether estimates of components of a state vector statistically affect each other. For example, if a component of the state estimate is geological ground vibrations in several directions, one may like to know whether an increase in one direction greatly affects that of another direction. Also, the measurement of displacements from rock deformation or landslides can have several locations as well as perhaps directions forming a multivariate random vector. The analyst may not be concerned only whether a component of the state, itself, affects that of another but may also wish to know whether estimated components are affected by the estimation of other components of the estimated state vector. We use correlations to analyze this. The covariance matrix of any random process gives cross covariance between any 2 component elements. Correlations between these elements can be computed by dividing the product of the square roots of the variances of the 2 elements (standard deviations) (Meditch 1969). Since the composite covariance that we will derive is the actual covariance of the KF, it can show the amount of correlation between estimated state components and therefore how much one estimated component affects estimation of another statistically. As before, being a priori, this is the conjectured correlation, which can be obtained before the actual observation.

Being a priori, insights of how closely estimates follow predictions and how estimates of components are affected by other components give knowledge that can be archived as reference on similar future runs. One can thus have better ideas of what to expect in similar scenes.

If we are given actual observations and measurements, another use of composite covariance occurs if observations

in the field or lab experiment can be replicated. In that case this composite covariance can be used to perform a test of hypothesis on the correctness of the state mean model, that is, how correct is the predicted state. Again, this is true since for Kalman filters, being unbiased, its estimate mean equals that of the predicted state. Since Gaussian processes are fully determined by the covariance and mean, a statistical test of hypothesis on the mean can be constructed. We must replicate and measure an experimental run a given *m* times for a chosen time *k*. Then one takes the vector sample mean of the estimate at time *k*. The state mean, if the model is correct, is as we will derive. We do a multivariate test of hypothesis to test the correctness of this mean (prediction) by using the probability distribution of the sample mean computed from the distribution of the KF estimate, whose distribution is determined by our composite covariance. We construct a multivariate rejection ("critical") region for the hypothesized mean (prediction) model at any desired confidence level using the *n* dimensional joint probability density of the sample mean, rejecting the assumed prediction at time *k* if it falls in this region. We can also test the hypotheses on any component of the mean vector by using the element's variance within the covariance matrix and the mean to be tested of that element, constructing an univariate sample mean critical region. See (Hogg and Craig 1978) on a univariate test of hypothesis, for which the *n* dimensional version in this article generalizes. In geology, for example, one may have some samples taken from the earth for analysis. One may perform some tests such as testing for how much radioactive minerals it contains or other attributes that can be modeled a priori. Then measurements can be taken and estimates computed from Kalman filters. Perhaps this experiment can be replicated with the lab sample multiple times. Then such a test of hypothesis can be performed to accept or reject the predicted model.

In our presentation and derivations, we use mathematical symbols that in large part followed the textbook by (Meditch 1969), as these symbols very clearly described the Kalman filter (KF). However, in contrast to the book, we do not assume the deterministic state model to be 0 in the absence of process noise.

3. Elementary Well-Known Results

In order to prove the covariance formula, some elementary well-known results from probability and KF are required. These are presented here.

Theorem 1: Given random vector X, Y, U, V, and matrices A, B, C, D,

$$cov[AX + BY, CU + DV] = A\,cov(X, U)C^T + B\,cov(Y, U)\,C^T + A\,cov(X, V)\,D^T + B\,cov(Y, V)\,D^T \quad (1)$$

Note, for all random variable R, we define

$$cov(R) \equiv E\{[R - E(R)][R - E(R)]^T\} \equiv cov(R, R) \quad (2)$$

Corollary 1: Given random vectors X, Y

$$cov(AX + BY) = A\,cov(X)\,A^T + B\,cov(Y)\,B^T + A\,cov(X, Y)\,B^T + B\,cov(Y, X)A^T \quad (3)$$

Theorem 2: Given random vectors X and Y,

$$cov(X, Y) = cov^T(Y, X) \quad (4)$$

The state and measurement equations will be stated respectively as follows (Meditch 1969), letting$X(j)$ &$Z(j)$ be state & measurement vectors respectively at time t_j. Let $X(0) = X \quad (5)$, X a random variable.

$$X(k + 1) = \Phi(k + 1, k)X(k) + \Gamma(k + 1, k)W(k) \quad (6)$$

$$Z(k + 1) = H(k + 1)X(k + 1) + V(k + 1) \quad (7)$$

$\Phi(k + 1, k)$ is called the state transition matrix from time $t_k$to t_{k+1}. Equation (6) is actually the solution of the differential equation model of the state at time t_{k+1}, initialized at t_k, where the transition matrix in (6) is actually the state transition matrix of this solution. The differential equation is equation (8) below, which we will describe very soon. W and V represent Gaussian white process and measurement noises respectively, which arises from the original state differential equation motion. H and Γ are matrix functions of k. Γ has $n \times p$ dimension, as we define the state X to be n dimensional for n components and process noise W to have p dimensions. H has $q \times n$ dimensions, as we define the measurement $Z(K)$ to have q dimensions. The differential equation representing the state is

$$dX(t) = F(t)X(t)dt + G(t)dB(t); t \geq 0 \quad (8)$$

where B is Brownian motion, supposedly the integral of Gaussian white noise. Yet, Gaussian white noise mathematically does not exist, as Brownian motion is with probability 1 ("almost surely") not differentiable. See (Mandelbrot and Van Ness 1968), which states this by way of processes called Fractional Brownian Motion, which Brownian motion is a specific case. We say non-differentiable with probability 1 because there is a set in the probability space where Brownian motion is differentiable, but it has probability measure 0. Noting a probability space is a measure space where its measure is defined as its probability, the reader is referred to (Royden 1969) to learn about measure spaces and its measures. Its differential, $dB(t)$, is merely symbolized as $W(t)dt$, such that W is continuous Gaussian white noise, which, as explained, is mathematically non-existent. This is why we prefer using differentials in our differential equation to avoid the derivative definition of Gaussian white noise,

$$W(t) \equiv \frac{dB(t)}{dt} \quad (9)$$

which with probability 1, cannot exist. [Wong and Hajek 1985] discusses such differential equations as well as its generalization. Symbol k represent time $t_k > 0; k = 0,1, \ldots$. F and G are matrices involved in describing the differential equation model, and Γ is the coefficient matrix of the discrete time noise that arises as the solution of the

differential equation. Note in solving the differential equation we assume that the process noise is constant between observations. So actually

$$\Gamma(k+1,k) \equiv \int_{t_k}^{t_{k+1}} \Phi(k+1,k)G(\alpha)\,d\alpha \qquad (10)$$

from solving the differential equation with process noise constant between observation (Meditch 1969). (This avoids the non-differentiability of Brownian motion and also simplifies and modifies the Gaussian white process noise to a Gaussian process that is independent for different $k = 0,1,\cdots$.).

The KF equations for the state estimator, written in reference (Meditch 1969) form, which is used in our derivations is:

Theorem 3: Kalman Filter (Meditch 1969): Given the state model X the KF estimator is:

$$\hat{X}(k+1|k+1) = [I - K(k+1)H(k+1)] \times \Phi(k+1,k)\hat{X}(k|k) + K(k+1)Z(k+1) \qquad (11)$$

where

$$K(k+1) = P(k+1|k)H^T(k+1)[H(k+1) \times P(k+1|k)H^T(k+1) + R(k+1)]^{-1} \qquad (12)$$

$$P(k+1|k) = \Phi(k+1,k)P(k|k)\Phi^T(k+1,k) + \Gamma(k+1,k)Q(k)\Gamma^T(k+1,k) \qquad (13)$$

$$P(k+1|k+1) = [I - K(k+1)H(k+1)]P(k+1|k);$$
$$I \text{ an } n \times n \text{ identity matrix} \qquad (14)$$

$$\hat{X}(0|0) = E[X(0)] \equiv x_0 \qquad (15)$$

is the initial condition vector and $\hat{X}(k|k)$ represents the KF estimate of $X(k)$, given observed measurements $\{Z(1),\dots,Z(k)\}$. $Z(j)$ has a given r dimension $j = 1,\dots,p$ for a total of p observed measurements.

$$P(k|k) \equiv cov\{[\hat{X}(k|k) - X(k)]\}$$
$$\equiv E\left\{[\hat{X}(k|k) - X(k) - 0][\hat{X}(k|k) - X(k) - 0]^T\right\}$$
$$= E\left\{[\hat{X}(k|k) - X(k)][\hat{X}(k|k) - X(k)]^T\right\} \qquad (16)$$

is the filter error covariance given k measurements, which assumes initial condition $P(0|0)$ as a given P_0, and k represents $t_k \geq 0$. $K(k)$ is a matrix with associated dimensions from the KF known as the "Kalman gain". Matrix$P(k+1|k)$ is the predicted error covariance at $k+1$ given k measurements (Meditch 1969). $Q(k); k = 0,1,\cdots$ represents process noise covariance, and $R(k+1); k = 0,1,\dots$ represents measurement noise covariance. The covariance Q definition is modified to the discrete time noise that is constant for all k, as explained above. In the identical way, the covariance R is also now made discrete.

4. Derivation of the Covariance and the Mean Forming the Complete Gaussian Statistics

Finally, what follows is the formula, stated in the form of a mathematical theorem, for the composite covariance of the estimated state. The mathematical proof is also provided. As already explained, it characterizes the total variation of the state estimation. It is based upon the given process noise, measurement noise covariance, and state model. Advantages of this algorithm is that it is a relatively simple formula that can be programmed on the computer in a straightforward manner. It is a sequential (or recursive) formula and based on the model and 2 input covariance parameter matrices. An advantage of this methodology is that it is computationally very fast, requiring no computationally intensive, CPU time consuming. nor memory space intensive processes, being a set of recursive equations. Since the measurement data are not used, the accuracy of this methodology relies upon the reliability of the state equation model and the accuracy of the input processes noise and measurement noise covariance matrices.

Let $\hat{X}(k|k)$ represent the KF estimate for state $X(k)$ given the first k measurement data where data at k represents measurement at time t_k. Then the sequential composite covariance algorithm is given in the next theorem.

Recall that we define this covariance as follows:

$$cov[\hat{X}(k|k)] \equiv E\left\{\left[\hat{X}(k|k) - E\left(\hat{X}(k|k)\right)\right]\left[\hat{X}(k|k) - E\left(\hat{X}(k|k)\right)\right]^T\right\} \qquad (17)$$

Theorem 4: The equations for $cov[\hat{X}(k+1|k+1)]$ is as follows:

$$cov[\hat{X}(k+1|k+1)] \equiv E\left\{\left[\hat{X}(k+1|k+1) - E\left(\hat{X}(k+1|k+1)\right)\right] \times \left[\hat{X}(k+1|k+1) - E\left(\hat{X}(k+1|k+1)\right)\right]^T\right\}$$
$$= [I - K(k+1)H(k+1)]\Phi(k+1,k)\,cov\left(\hat{X}(k|k)\right) \times \Phi^T(k+1,k)[I - K(k+1)H(k+1)]^T$$
$$+K(k+1)H(k+1)\,cov\big(X(k+1)\big)H^T(k+1) \times K^T(k+1) + K(k+1)\,cov\big(V(k+1)\big)K^T(k+1)$$
$$+\big(I - K(k+1)H(k+1)\big)\Phi(k+1,k) \times cov[\hat{X}(k|k), X(k+1)]H^T(k+1) \times K^T(k+1)$$
$$+K(k+1)H(k+1)\,cov^T[\hat{X}(k|k), X(k+1)] \times \Phi^T(k+1,k)[I - K(k+1)H(k+1)]^T \qquad (18)$$

$$cov[\hat{X}(k|k), X(k+1)] \equiv E\left\{\left[\hat{X}(k|k) - E\left(\hat{X}(k|k)\right)\right]\left[X(k+1) - E\big(X(k+1)\big)\right]^T\right\}$$
$$= [I - K(k)H(k)]\Phi(k,k-1) \times cov[\hat{X}(k-1|k-1), X(k)]\,\Phi^T(k+1,k)$$
$$+K(k)H(k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k) \qquad (19)$$

$$cov\big(X(k+1)\big) \equiv$$
$$E\left\{\big[X(k+1) - E\big(X(k+1)\big)\big] \times \big[X(k+1) - E\big(X(k+1)\big)\big]^T\right\}$$
$$= cov[\Phi(k+1,k)X(k) + \Gamma(k+1,k)W(k)] =$$
$$\Phi(k+1,k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k) +$$
$$\Gamma(k+1,k)\,cov\big(W(k)\big)\,\Gamma^T(k+1,k) \quad (20)$$

Initialize as follows:

a) Given
$$\hat{X}(0|0) \equiv E[X(0)] = x_0,$$
$$cov\big[\hat{X}(0|0)\big] = cov(x_0) \quad (21)$$

But this is zero since x_0 is a constant.

b) $cov\big[\hat{X}(0|0), X(1)\big] = 0$ since $\hat{X}(0|0) = X_0$, a constant (22)

c) Since $cov\big(W(k)\big) = Q(k)$, let $cov\big(X(0)\big) = Q_0$ (23)

d) $cov\big[X(1), \hat{X}(0|0)\big] = 0$ (24)

for the same reason as in b)

Assume that W, V are the discrete versions derived from Gaussian white noises, as previously described. Also assume $W(k)$ is independent of $V(m)$, and $V(m)$ is independent of $X(p)$, for any integer *k,m,p*, and $W(k)$ is independent of $X(p)$ for all $p \le k$.

Proof:

The preceding theorems and corollary are used to prove these results. Each of the 3 equations of this theorem are proven separately. The third, equation (20) is simplest and proven first.

$$cov\big(X(k+1)\big) =$$
$$cov[\Phi(k+1,k)X(k) + \Gamma(k+1,k)W(k)] =$$
$$\Phi(k+1,k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k) +$$
$$\Gamma(k+1,k)\,cov\big(W(k)\big)\,\Gamma^T(k+1,k) \quad (20)$$

Proof:

By definition given as equation (6),

$$X(k+1) = \Phi(k+1,k)X(k) + \Gamma(k+1,k)W(k).$$

Therefore,

$$cov\big(X(k+1)\big) = cov[\Phi(k+1,k)X(k) + \Gamma(k+1,k)W(k)]$$
$$= \Phi(k+1,k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k)$$
$$+\Phi(k+1,k)\,cov[X(k),W(k)]\,\Gamma^T(k+1,k) +$$
$$\Gamma(k+1,k)\,cov[W(k),X(k)]\,\Phi^T(k+1,k)$$
$$+\Gamma(k+1,k)\,cov\big(W(k)\big)\,\Gamma^T(k+1.k) =$$
$$\Phi(k+1,k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k) +$$
$$\Gamma(k+1,k)\,cov\big(W(k)\big)\,\Gamma^T(k+1,k) \quad (25)$$

Now we will prove our first relationship, equation (18).

1) $cov\big[\hat{X}(k+1|k+1)\big]$
$$= [I - K(k+1)H(k+1)]\Phi(k+1,k)$$
$$cov\left(\hat{X}(k|k)\right)\Phi^T(k+1,k)[I - K(k+1)H(k+1)]^T$$
$$+K(k+1)H(k+1)\,cov\big(X(k+1)\big)\,H^T(k+1)$$
$$\times K^T(k+1) + K(k+1)\,cov\big(V(k+1)\big)\,K^T(k+1)$$
$$+\big(I - K(k+1)H(k+1)\big)\Phi(k+1,k) \times$$
$$cov\big[\hat{X}(k|k), X(k+1)\big]\,H^T(k+1)K^T(k+1)$$
$$+K(k+1)H(k+1)\,cov^T\big[\hat{X}(k|k), X(k+1)\big] \times$$
$$\Phi^T(k+1,k)[I - K(k+1)H(k+1)]^T \quad (18)$$

Proof: Let

$$A = [I - K(k+1)H(k+1)]\Phi(k+1,k);$$
$$K = K(k+1); H = H(k+1) \quad (26)$$

Then by the Kalman filter formula previously given,

$$cov\left(\hat{X}(k+1|k+1)\right) = cov\big[A\hat{X}(k|k) + KZ(k+1)\big]$$
$$= cov\big[A\hat{X}(k|k) + K\big(HX(k+1) + V(k+1)\big)\big] \quad (27)$$

As measurement model, equation (7) implies

$$Z(k+1) = HX(k+1) + V(k+1) \quad (28)$$

$$cov\big[A\hat{X}(k|k) + K\big(HX(k+1) + V(k+1)\big)\big]$$
$$= A\,cov\left(\hat{X}(k|k)\right)A^T + K\,cov[HX(k+1) + V(k+1)]\,K^T$$
$$+A\,cov\big[\hat{X}(k|k), HX(k+1) + V(k+1)\big]\,K^T$$
$$+K\,cov\big[HX(k+1) + V(k+1), \hat{X}(k|k)\big]\,A^T$$
$$= A\,cov\left(\hat{X}(k|k)\right)A^T + KH\,cov\big(X(k+1)\big)\,H^TK^T + K\,cov\big(V(k+1)\big)\,K^T$$
$$+A\,cov\big[\hat{X}(k|k), HX(k+1) + V(k+1)\big]\,K^T$$
$$+K\,cov\big[HX(k+1) + V(k+1), \hat{X}(k|k)\big]\,A^T$$
$$= A\,cov\left(\hat{X}(k|k)\right)A^T + KH\,cov\big(X(k+1)\big)\,H^TK^T$$
$$+K\,cov\big(V(k+1)\big)\,K^T + A\,cov\big[\hat{X}(k|k), X(k+1)\big]\,H^TK^T$$
$$+KH\,cov\big[X(k+1), \hat{X}(k|k)\big]\,A^T$$
$$= A\,cov\left(\hat{X}(k|k)\right)A^T + KH\,cov\big(X(k+1)\big)\,H^TK^T +$$
$$K\,cov\big(V(k+1)\big)\,K^T + A\,cov\big[\hat{X}(k|k), X(k+1)\big]\,H^TK^T +$$
$$KHcov^T\big[\hat{X}(k|k), X(k+1)\big]A^T \quad (29)$$

By our definitions of *K, H,* and *A,* proof is now complete.

Note that the independence of $V(k)$ and $X(j)$ for all $j, k = 0,1,2..$ is utilized in this proof. Also, notice that this implies that $V(k)$ must also be independent of $\hat{X}(j|j)$ for all $j < k = 0,1,2,...$, and this fact was also needed in the proof of this result. Also note that Theorem 2 was used.

Finally, we show the second part of theorem, equation (19).

$$cov\big[\hat{X}(k|k), X(k+1)\big] =$$
$$[I - K(k)H(k)]\Phi(k,k-1) \times$$

$$cov\left[\hat{X}(k-1|k-1), X(k)\right]\Phi^T(k+1,k) + K(k)H(k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k) \quad (19)$$

Proof:

$$cov\big[\hat{X}(k|k), x(k+1)\big] = cov\big\{[I-K(k)H(k)]\Phi(k,k-1)\hat{X}(k-1|k-1) + K(k)Z(k), \Phi(k+1,k)X(k) + \Gamma(k+1.k)W(k)\big\} =$$
$$[I-K(k)H(k)]\Phi(k,k-1)\,cov\big[\hat{X}(k-1|k-1), X(k)\big] \times \Phi^T(k+1,k) + K(k)\,cov[Z(k), X(k)]\,\Phi^T(k+1,k) +$$
$$[I-K(k)H(k)]\Phi(k,k-1)\,cov\left(\hat{X}(k-1|k-1), W(k)\right) \times \Gamma^T(k+1,k) + K(k)\,cov\big(Z(k), W(k)\big)\,\Gamma^T(k+1,k) =$$
$$[I-K(k)H(k)]\Phi(k,k-1)\,cov\big[\hat{X}(k-1|k-1), X(k)\big] \times \Phi^T(k+1,k) + K(k)\,cov[Z(k), X(k)]\,\Phi^T(k+1,k)$$
$$+K(k)\,cov[H(k)X(k)+V(k), W(k)]\,\Gamma^T(k+1,k) =$$
$$[I-K(k)H(k)]\Phi(k,k-1)\,cov\big[\hat{X}(k-1|k-1), X(k)\big] \times \Phi^T(k+1,k) + K(k)\,cov[Z(k), X(k)]\,\Phi^T(k+1,k) =$$
$$[I-K(k)H(k)]\Phi(k,k-1)\,cov\big[\hat{X}(k-1|k-1), X(k)\big]\Phi^T(k+1,k)$$
$$+K(k)\,cov[H(k)X(k)+V(k), X(k)]\,\Phi^T(k+1,k) =$$
$$[I-K(k)H(k)]\Phi(k,k-1)\,cov\big[\hat{X}(k-1|k-1), X(k)\big] \times \Phi^T(k+1,k) + K(k)H(k)\,cov[X(k), X(k)]\,\Phi^T(k+1,k)$$
$$\overset{def}{\equiv} [I-K(k)H(k)]\Phi(k,k-1) \times cov\big[\hat{X}(k-1|k-1), X(k)\big]\Phi^T(k+1,k) + K(k)H(k)\,cov\big(X(k)\big)\,\Phi^T(k+1,k) \quad (30)$$

This completes the proof of our theorem that gives an algorithm for finding the estimator's covariance.

Since the estimator is a linear function of normally distributed process noise and measurement vectors. we said the estimator, itself, is also Gaussian, describable by its mean vector and covariance matrix. Now that the covariance matrix is derived, we now give and prove a formula for the mean vector.

Theorem 5:

$$E\big[\hat{X}(k+1|k+1)\big] = \prod_{j=0}^{k}\Phi(j+1,j)\,E[X(0)] \quad (31)$$

Proof (via mathematical induction):

$$E\big[\hat{X}(1|1)\big] = (I-KH)\Phi(1,0)E\big[\hat{X}(0|0)\big] + KHE[X(1)]$$
$$= (I-KH)\Phi(1,0)E\big[E[X(0)]\big] + KH\Phi(1,0)E[X(0)]$$
$$= (I-KH)\Phi(1,0)E[X(0)] + KH\Phi(1,0)E[X(0)]$$
$$= \Phi(1,0)E[X(0)] \quad (32)$$

where *K* and *H* are as previously defined in proof of equation 1. Note that the above relation assumed that the random noises are of zero mean.

$$E\big[\hat{X}(2|2)\big] = (I-KH)\Phi(2,1)E\big[\hat{X}(1|1)\big] + KHE[X(2)]$$
$$= (I-KH)\Phi(2,1)\Phi(1,0)E[X(0)] + KH\Phi(2,1)E[X(1)]$$
$$= (I-KH)\Phi(2,1)\Phi(1,0)E[X(0)] + KH\Phi(2,1)\Phi(1,0)E[X(1,0)]$$
$$= \Phi(2,1)\Phi(1,0)E[X(0)] \quad (33)$$

For the induction step of the proof, let us assume that

$$E\left[\hat{X}[k|k]\right] = \Phi(k,k-1)\Phi(k-1,k-2)\ldots\Phi(1,0)E[X(0)] \equiv \prod_{j=0}^{k-1}\Phi(j+1,j)\,E[X(0)] \quad (34)$$

Then we have

$$E\big[\hat{X}(k+1|k+1)\big] = (I-KH)\Phi(k+1,k)E\big[\hat{X}(k|k)\big] + KHE[X(k+1)] \quad (35)$$

Using the fact that that

$$E[X(k+1)] = E[\Phi(k+1,k)X(k) + \Gamma(k+1,k)W(k)] = \Phi(k+1,k)E[X(k)] = \prod_{j=0}^{k}\Phi(j+1,j)\,E[X(0)] \quad (36)$$

with the process noise W being of zero mean and also using the induction step assumption given above,

$$E\big[\hat{X}(k+1|k+1)\big] = (I-KH)\prod_{j=0}^{k}\Phi(j+1,j) \times E[X(0)] + KH\prod_{j=0}^{k}\Phi(j+1,j)\,E[X(0)] \quad (37)$$

$$\therefore E\big[\hat{X}(k+1|k+1)\big] = \prod_{j=0}^{k}\Phi(j+1,j)\,E[X(0)] \quad (31)$$

and the result is proved. (Note we iterated the definition of $X(k)$ and took its expectation to get the last expression on equation (36).) As this formula is clearly also the state mean, we give can a corollary.

Corollary 2: The expectation of a KF is the same as that of the state, making the KF an unbiased estimator.

Note for the special case where the transition matrix is constant, our result gives

$$E\big[\hat{X}(k+1|k+1)\big] = \Phi^{k+1}(1,0)E[X(0)] \quad (38)$$

Recall that in this article, as in [Meditch 1969], we defined the transition matrix symbol $\Phi(j+1,j)$ as the transition matrix from time t_j to time t_{j+1}.

One property of transition matrices is the fact that

$$\Phi(k)\Phi(k-1)\ldots\Phi(0) = \Phi(k+1,0) \quad (39)$$

[Meditch 1969]

Thus, this theorem, concerning the mean of the estimator, can be stated as

$$E\big[\hat{X}(k+1|k+1)\big] = \Phi(k+1,0)E[X(0)] \quad (40)$$

Thus, given that $E[X(0)] = x_0$ (41) is assumed known, this theorem gives the mean of the estimator.

5. Simple Example

We now give a simple numerical example of how to use this covariance in showing how close an estimate follows the

prediction, that is, mean. We also show that traditional error covariance that Kalman filters outputs for comparison.

Suppose we have a predicted model of phenomena, such as seismic data [Brillinger et al. 2004] or numerous other applications. We give numbers here that are chosen and simplified to illustrate how the new covariance is used in applications. For simplicity, assume the model is 2 dimensions only and at time, say $k = 3$, the predicted state is $E[X(3)] = \begin{bmatrix}10\\30\end{bmatrix}$. Now suppose our Kalman filter estimation of the state is $\hat{X}(3) = \begin{bmatrix}12\\34\end{bmatrix}$. To compare with our new composite covariance, suppose the error covariance is $P(3|3) = \begin{bmatrix}1 & .1\\ .1 & 2\end{bmatrix}$.

Suppose our newly derived composite covariance is $cov\left[\hat{X}(3|3)\right] = \begin{bmatrix}1.44 & 1\\ 1 & 2.25\end{bmatrix}$. Note the variance of the 1st and 2nd components of the estimation error is 1 and 2 respectively from the error covariance, while the variance of the 1st and 2nd components of the estimation covariance is 1.44 and 2.25 respectively from the composite covariance. We conclude that the state estimate at time $k = 3$ follows the actual state, $X(3)$ well via the traditional error covariance matrix, and the estimate also varies little from the predicted state $E[X(3)] = \begin{bmatrix}10\\30\end{bmatrix}$ via our composite covariance matrix, which is computed by equations (18)-(20) with assumed initial condition by equations (21)-(24).

Now suppose at $k = 60$, we have the predicted state $E[X(60)] = \begin{bmatrix}20\\40\end{bmatrix}$ and suppose our Kalman filter estimation of the state is $\hat{X}(60) = \begin{bmatrix}80\\91\end{bmatrix}$. Suppose the error covariance is $P(60|60) = \begin{bmatrix}1.4 & .5\\ .5 & 2.4\end{bmatrix}$ but the new composite covariance is now $cov\left[\hat{X}(60|60)\right] = \begin{bmatrix}49 & 1\\ 1 & 64\end{bmatrix}$. We conclude from this that at $k = 60$, the estimate still follows that actual state quite well via variances in the traditional error covariance matrix, but this same estimate may deviate greatly from the prediction and most likely will not resemble the predicted path due to the larger variances of 49 and 64 in the composite covariance matrix.

As previously explained, since process noise contribute cumulatively with time, deviations from the prediction will not lessen, but, instead, in most applications will continue to grow with time, increasing our new covariance. Yet, the timing of when such deviations may get large is one of the new applications we explained, as it gives novel conjecture information, which may prove helpful to some analysts.

We also stated that another application of our composite covariance is to determine how each component of the estimate is correlated to another component. By the definition of correlation, we have in our example, at time $k = 60$, $correlation\left[\hat{X}_1, \hat{X}_2\right] = \frac{1}{\sqrt{49}\sqrt{64}} = \frac{1}{56}$. Since the maximum magnitude of a correlation is 1.0, we conclude that the 2 components of the estimate show very little correlation at time $k = 60$ in our example.

6. Conclusions

In this article, we derived the "composite" covariance and mean of the Kalman filter state estimate. Since the Kalman filter estimate is a normally distributed (Gaussian) random process, this covariance, which differs from the error covariance, and the mean are required to exactly describe its probabilistic distribution. We described how this covariance is useful in statistical analysis of attributes of a Kalman filter state estimation and explained why such analysis is useful in various physical problems. We concluded with a simple illustrative numerical example.

References

Brillinger, D, Robinson, E, Schoenberg, F (Editors), *Time Series Analysis and Applications to Geophysical Systems,* Springer-Verlag, 2004, pages 11-39 (actual authors here: Kitagawa, Takanani, and Matsumoto).

Hogg, R.V. and Craig, A.T., *Introduction to Mathematical Statistics*, MacMillan Publishing Co., Inc, New York, 1978, Sec. 7.1.

Mandelbrot, B and Van Ness, J.W. "Fractional Brownian Motions, Fractional Noises and Application", *SIAM Review*, Vol. 10, 1968, page 427.

Meditch, J.S. *Stochastic Optimal Linear Estimation and Control*, McGraw-Hill, New York, 1969, Sec. 2.3, 2.4, 3.5, 3.6, 5.3, 5.4.

Royden, H.L., *Real Analysis Second Edition*, Macmillian Publishing Co., Inc. New York, 1968, Ch. 11

Tapley, B.D., Schutz, B.E., Born, G.H., *Statistical Orbit Determination,* Elsevier. Academic Press, Burlington, MA, San Diego, London, 2004, Sec 4.9, Appendix F.

Wong, E, Hajek, B., *Stochastic Processes in Engineering Systems*, Springer Verlag, New York, 1985, Sec. 3.8, Ch. 4.

Appendix

Along with our estimator covariance just derived, some may like to know how the Kalman filter's error covariance is affected by the process noise alone. Our composite covariance of this article and the Kalman filter's error covariance are modeled a priori (without empirical data), and we can actually derive still another new covariance in addition to the one in this article by merely setting the measurement noise covariance to be 0. That is, let

$$R(k+1) = 0 \forall k = 0,1,\cdots \quad (42)$$

This measurement noise covariance is used to compute equation (12), the Kalman gain matrix, K. Because this Kalman gain is used in computing the error covariance, which is done recursively over all time, we zeroed out the measurement noise error covariance for all measured time, not just the current one or the ones we are analyzing. Thus equation (12) now becomes

$$K(k+1) = P(k+1|k)H^T(k+1) \times [H(k+1)P(k+1|k)H^T(k+1)]^{-1} \forall k = 0,1,\cdots (43)$$

Our modified error covariance can be defined as

$$P(k+1|k+1|V(j); j = 1,2,\cdots,k+1) \equiv E\{[X\,(k+1|k+1) - X(k+1)] \times [X\,(k+1|k+1) - X(k+1)]^T\,|V(j)\;; j = 1,2,\cdots k+1\} \quad (44)$$

This indicates that we assume knowledge of the value of $V(j); j = 1,2,\cdots k+1$, and so it is not random. The formula for this modified error covariance is the same as equation for $P(k+1|k+1)$, that is equation (14), except that we use equation (43) for the Kalman gain.

Analogously some analyst may also wish to know the behavior of the error covariance in the absence of random process noise. In this case we actually know the state but may still wish to check out the error covariance, perhaps in a laboratory to test measurement devices. Again, due to the fact that the error covariance formula is recursive, we must assume the process noise to be 0 for all time in order to derive error covariances conditioned upon knowledge of the states. That is, let

$$Q(k) = 0 \forall k = 0,1,\cdots \quad (45)$$

The new error covariance is defined to be

$$P(k+1|k+1|X(j); j = 0,1,\cdots k+1) \equiv E\left\{\left[\hat{X}(k+1|k+1) - X(k+1)\right] \times \left[\hat{X}(k+1|k+1) - X(k+1)\right]^T |X(j); j = 0,1,\cdots k+1\right\} \quad (46)$$

However, these new error covariances may be meaningless, since Kalman filters often tend to diverge when process noises have 0 covariance over all time. This divergence is because state models generally are not perfectly accurate and so the inaccuracy falls into process noise. For example, linear models are often used to model phenomena that is not truly linear [Tapley et al 2004]. Perhaps to a smaller extent, numerical errors may also contribute to divergence without process noise too. Therefore, we will not proceed with this.

Recall since our model is discrete here, process noise covariance is discrete, and we represented it by

$Q(k) \equiv cov\big(W(k)\big)$,

where $W(k)$ represents process noise at time $k = 0,1,2\cdots$. Empirically finding a good, hypothesized model for Q may be difficult, but discussion of it is beyond the scope of this article. [Tapley et al 2004] discusses this somewhat, regarding their application of Kalman filters.

Finally, note that any covariance that is derived from the Kalman filter is also applicable to the Bayesian (generalized batch least squares) estimator. This is due to the following theorem.

Theorem: The Kalman filter is equivalent to the Bayesian optimal estimator, given any admissible loss function [Meditch 1969].

Proof:

For Kalman filters, the estimator is derived from the conditional expectation $E[X(k)|Y(1),\cdots,Y(k)]$ where the random variables Y 's can be interpreted as the observations for the Kalman Filter [Meditch 1969]. Note that both the measurement noise and process noise are Gaussian. This implies the Bayesian optimal estimate must be the conditional mean of the state conditioned on the observations for any admissible loss function [Meditch 1969]. Since any admissible loss function will work, it must surely work for the "mean square loss function" [Meditch 1969]. Thus, Kalman filters are equivalent to Bayesian optimal estimators [Tapley et al 2004].

Note that a Bayesian estimator is the generalized version of the batch least squares. It is generalized in the sense that the parameters (or states) themselves are random instead of deterministic as in the least squares estimator in the traditional sense of the term. Furthermore, a Bayesian estimator can generalize a Kalman filter because the Kalman filter assumes white noises, but white is not required in a Bayesian estimator