LESSON 24

Iterated Least Squares and Extended Kalman Filtering

SUMMARY

This lesson is devoted primarily to the extended Kalman filter (EKF), which is a form of the Kalman filter "extended" to nonlinear dynamical systems of the form described in Lesson 23. We show that the EKF is related to the method of iterated least squares (ILS), the major difference being that the EKF is for dynamical systems, whereas ILS is not.

We explain ILS for the model $z(k) = f(\theta, k) + \nu(k)$, $k = 1, 2, \ldots, N$, where the objective is to estimate θ from the measurements. A four-step procedure is given for ILS, from which we observe that in each complete cycle of this procedure *we use both the nonlinear and linearized models* and that ILS uses the estimate obtained from the linearized model to generate the nominal value of θ about which the nonlinear model is *relinearized*.

The notions of relinearizing about a filter output and using both the nonlinear and linearized models are also at the very heart of the EKF. The EKF is developed in predictor–corrector format. Its prediction equation is obtained by integrating the nominal differential equation associated with $\mathbf{x}^*(t)$, from t_k to t_{k+1}. To do this, we learn that $\mathbf{x}^*(t_k) = \hat{\mathbf{x}}(k|k)$, whereas $\mathbf{x}^*(t_{k+1}) = \hat{\mathbf{x}}(k+1|k)$. The corrector equation is obtained from the Kalman filter associated with the discretized perturbation state-variable model derived in Lesson 23. The gain and covariance matrices that appear in the corrector equation depend on the nominal $\mathbf{x}^*(t)$ that results from prediction, $\hat{\mathbf{x}}(k+1|k)$.

An *iterated EKF* is an EKF in which the correction equation is iterated a fixed number of times. This improves the convergence properties of the EKF, because convergence is related to how close the nominal value of the state vector is to its actual value.

at the ith and $(i+1)$st iterations, respectively. Convergence of the ILS method occurs when

$$|\hat{\theta}^{i+1}_{\mathrm{WLS}}(N) - \hat{\theta}^{i}_{\mathrm{WLS}}(N)| < \epsilon \tag{24-8}$$

where ϵ is a prespecified small positive number.

We observe, from this four-step procedure, that ILS uses the estimate obtained from the linearized model to generate the nominal value of θ about which the nonlinear model is *relinearized*. Additionally, in each complete cycle of this procedure, *we use both the nonlinear and linearized models*. The nonlinear model is used to compute $z^*(k)$ and subsequently $\delta z(k)$, using (24-3).

The notions of relinearizing about a filter output and using both the nonlinear and linearized models are also at the very heart of the EKF.

EXTENDED KALMAN FILTER

The nonlinear dynamical system of interest to us is the one described in Lesson 23. For convenience to the reader, we summarize aspects of that system next. The nonlinear state-variable model is

$$\dot{\mathbf{x}}(t) = \mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t] + \mathbf{G}(t)\mathbf{w}(t) \tag{24-9}$$

$$\mathbf{z}(t) = \mathbf{h}[\mathbf{x}(t), \mathbf{u}(t), t] + \mathbf{v}(t), \qquad t = t_i, \quad i = 1, 2, \ldots \tag{24-10}$$

Given a nominal input $\mathbf{u}^*(t)$ and assuming that a nominal trajectory $\mathbf{x}^*(t)$ exists, $\mathbf{x}^*(t)$ and its associated nominal measurement satisfy the following nominal system model:

$$\dot{\mathbf{x}}^*(t) = \mathbf{f}[\mathbf{x}^*(t), \mathbf{u}^*(t), t] \tag{24-11}$$

$$\mathbf{z}^*(t) = \mathbf{h}[\mathbf{x}^*(t), \mathbf{u}^*(t), t], \qquad t = t_i, \quad i = 1, 2, \ldots \tag{24-12}$$

Letting $\delta\mathbf{x}(t) = \mathbf{x}(t) - \mathbf{x}^*(t)$, $\delta\mathbf{u}(t) = \mathbf{u}(t) - \mathbf{u}^*(t)$, and $\delta\mathbf{z}(t) = \mathbf{z}(t) - \mathbf{z}^*(t)$, we also have the following *discretized perturbation state-variable model* that is associated with a linearized version of the original nonlinear state-variable model:

$$\delta\mathbf{x}(k+1) = \mathbf{\Phi}(k+1, k;^*)\delta\mathbf{x}(k) + \mathbf{\Psi}(k+1, k;^*)\delta\mathbf{u}(k) + \mathbf{w}_d(k) \tag{24-13}$$

$$\begin{aligned}\delta\mathbf{z}(k+1) &= \mathbf{H}_{\mathbf{x}}(k+1;^*)\delta\mathbf{x}(k+1) \\ &\quad + \mathbf{H}_{\mathbf{u}}(k+1;^*)\delta\mathbf{u}(k+1) + \mathbf{v}(k+1)\end{aligned} \tag{24-14}$$

In deriving (24-13) and (24-14), we made the important assumption that higher-order terms in the Taylor series expansions of $\mathbf{f}[\mathbf{x}(t), \mathbf{u}(t), t]$ and $\mathbf{h}[\mathbf{x}(t), \mathbf{u}(t), t]$ could be neglected. Of course, this is only correct as long as $\mathbf{x}(t)$ is "close" to $\mathbf{x}^*(t)$ and $\mathbf{u}(t)$ is "close" to $\mathbf{u}^*(t)$.

As discussed in Lesson 23, if $\mathbf{u}(t)$ is an input derived from a feedback control law, so that $\mathbf{u}(t) = \mathbf{u}[\mathbf{x}(t), t]$, then $\mathbf{u}(t)$ can differ from $\mathbf{u}^*(t)$, because $\mathbf{x}(t)$ will differ from $\mathbf{x}^*(t)$. On the other hand, if $\mathbf{u}(t)$ does not depend on $\mathbf{x}(t)$, then usually $\mathbf{u}(t)$ is the same as $\mathbf{u}^*(t)$, in which case $\delta\mathbf{u}(t) = \mathbf{0}$. We see, therefore, that

$\mathbf{x}^*(t)$ is the critical quantity in the calculation of our discretized perturbation state-variable model.

Suppose $\mathbf{x}^*(t)$ is given a priori; then we can compute predicted, filtered, or smoothed estimates of $\delta\mathbf{x}(k)$ by applying all our previously derived estimators to the discretized perturbation state-variable model in (24-13) and (24-14). We can precompute $\mathbf{x}^*(t)$ by solving the nominal differential equation (24-11). The Kalman filter associated with using a precomputed $\mathbf{x}^*(t)$ is known as *a relinearized KF*.

A relinearized KF usually gives poor results, because it relies on an open-loop strategy for choosing $\mathbf{x}^*(t)$. When $\mathbf{x}^*(t)$ is precomputed, there is no way of forcing $\mathbf{x}^*(t)$ to remain close to $\mathbf{x}(t)$, and this must be done or else the perturbation state-variable model is invalid. Divergence of the relinearized KF often occurs; hence, we do not recommend the relinearized KF.

The relinearized KF is based only on the discretized perturbation state-variable model. It does not use the nonlinear nature of the original system in an active manner. The extended Kalman filter relinearizes the nonlinear system about each new estimate as it becomes available; i.e., at $k = 0$, the system is linearized about $\hat{\mathbf{x}}(0|0)$. Once $\mathbf{z}(1)$ is processed by the EKF, so that $\hat{\mathbf{x}}(1|1)$ is obtained, the system is linearized about $\hat{\mathbf{x}}(1|1)$. By "linearize about $\hat{\mathbf{x}}(1|1)$," we mean $\hat{\mathbf{x}}(1|1)$ is used to calculate all the quantities needed to make the transition from $\hat{\mathbf{x}}(1|1)$ to $\hat{\mathbf{x}}(2|1)$ and subsequently $\hat{\mathbf{x}}(2|2)$. This phrase will become clear later. The purpose of relinearizing about the filter's output is to use a better reference trajectory for $\mathbf{x}^*(t)$. Doing this, $\delta\mathbf{x} = \mathbf{x} - \hat{\mathbf{x}}$ will be held as small as possible, so that our linearization assumptions are less likely to be violated than in the case of the relinearized KF.

The EKF is developed in predictor–corrector format (Jazwinski, 1970). Its prediction equation is obtained by integrating the nominal differential equation for $\mathbf{x}^*(t)$, from t_k to t_{k+1}. To do this, we need to know how to choose $\mathbf{x}^*(t)$ for the entire interval of time $t \in [t_k, t_{k+1}]$. Thus far, we have only mentioned how $\mathbf{x}^*(t)$ is chosen at t_k, i.e., as $\hat{\mathbf{x}}(k|k)$.

Theorem 24-1. *As a consequence of relinearizing about* $\hat{\mathbf{x}}(k|k)(k = 0, 1, \ldots)$,

$$\delta\hat{\mathbf{x}}(t|t_k) = \mathbf{0}, \qquad \text{for all } t \in [t_k, t_{k+1}] \tag{24-15}$$

This means that

$$\mathbf{x}^*(t) = \hat{\mathbf{x}}(t|t_k) \qquad \text{for all } t \in [t_k, t_{k+1}] \tag{24-16}$$

Before proving this important result, we observe that it provides us with a choice of $\mathbf{x}^*(t)$ over the entire interval of time $t \in [t_k, t_{k+1}]$, and it states that at the left-hand side of this time interval $\mathbf{x}^*(t_k) = \hat{\mathbf{x}}(k|k)$, whereas at the right-hand side of this time interval $\mathbf{x}^*(t_{k+1}) = \hat{\mathbf{x}}(k+1|k)$. The transition from $\hat{\mathbf{x}}(k+1|k)$ to $\hat{\mathbf{x}}(k+1|k+1)$ will be made using the EKF's correction equation.

Proof. Let t_l be an arbitrary value of t lying in the interval between t_k and t_{k+1} (see Figure 24-1). For the purposes of this derivation, we can assume that $\delta\mathbf{u}(k) = \mathbf{0}$ [i.e., perturbation input $\delta\mathbf{u}(k)$ takes on no new values in the interval from t_k to t_{k+1}; recall the piecewise-constant assumption made about $\mathbf{u}(t)$ in the derivation of (23-38)]; i.e.,

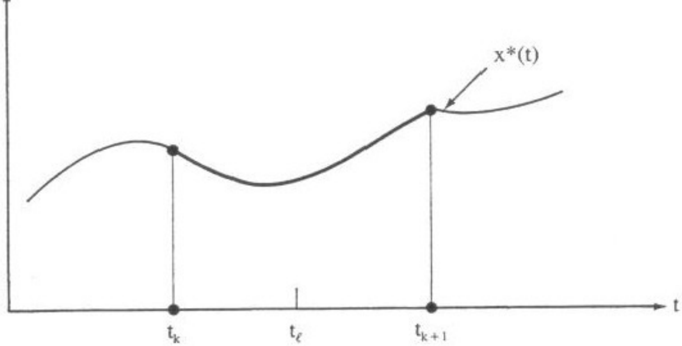

Figure 24-1 Nominal state trajectory $x^*(t)$.

$$\delta\mathbf{x}(k+1) = \mathbf{\Phi}(k+1, k;^*)\delta\mathbf{x}(k) + \mathbf{w}_d(k) \tag{24-17}$$

Using our general state-predictor results given in (16-14), we see that (remember that k is short for t_k, and that $t_{k+1} - t_k$ does not have to be a constant; this is true in all of our predictor, filter, and smoother formulas)

$$\begin{aligned} \delta\hat{\mathbf{x}}(t_l|t_k) &= \mathbf{\Phi}(t_l, t_k;^*)\delta\hat{\mathbf{x}}(t_k|t_k) \\ &= \mathbf{\Phi}(t_l, t_k;^*)[\hat{\mathbf{x}}(k|k) - \mathbf{x}^*(k)] \end{aligned} \tag{24-18}$$

In the EKF we set $\mathbf{x}^*(k) = \hat{\mathbf{x}}(k|k)$; thus, when this is done,

$$\delta\hat{\mathbf{x}}(t_l|t_k) = \mathbf{0} \tag{24-19}$$

and, because $t_l \in [t_k, t_{k+1}]$,

$$\delta\hat{\mathbf{x}}(t_l|t_k) = \mathbf{0}, \qquad \text{for } \textit{all } t_l \in [t_k, t_{k+1}] \tag{24-20}$$

which is (24-15). Equation (24-16) follows from (24-20) and the fact that $\delta\hat{\mathbf{x}}(t_l|t_k) = \hat{\mathbf{x}}(t_l|t_k) - \mathbf{x}^*(t_l)$. ☐

We are now able to derive the EKF. As mentioned previously, the EKF must be obtained in predictor–corrector format. We begin the derivation by obtaining the predictor equation for $\hat{\mathbf{x}}(k+1|k)$.

Recall that $\mathbf{x}^*(t)$ is the solution of the nominal state equation (24-11). Using (24-16) in (24-11), we find that

$$\frac{d}{dt}\hat{\mathbf{x}}(t|t_k) = \mathbf{f}[\hat{\mathbf{x}}(t|t_k), \mathbf{u}^*(t), t] \tag{24-21}$$

Integrating this equation from $t = t_k$ to $t = t_{k+1}$, we obtain

$$\hat{\mathbf{x}}(k+1|k) = \hat{\mathbf{x}}(k|k) + \int_{t_k}^{t_{k+1}} \mathbf{f}[\hat{\mathbf{x}}(t|t_k), \mathbf{u}^*(t), t]dt \tag{24-22}$$

which is the *EKF prediction equation*. Observe that the nonlinear nature of the system's state equation is used to determine $\hat{\mathbf{x}}(k+1|k)$. The integral in (24-22) is evaluated by means of numerical integration formulas that are initialized by $\mathbf{f}[\hat{\mathbf{x}}(t_k|t_k), \mathbf{u}^*(t_k), t_k]$.

The corrector equation for $\hat{\mathbf{x}}(k+1|k+1)$ is obtained from the Kalman filter associated with the discretized perturbation state-variable model in (24-13) and (24-14) and is

$$\begin{aligned}\delta\hat{\mathbf{x}}(k+1|k+1) = \delta\hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1;^*)[\delta\mathbf{z}(k+1) \\ - \mathbf{H}_{\mathbf{x}}(k+1;^*)\delta\hat{\mathbf{x}}(k+1|k) - \mathbf{H}_{\mathbf{u}}(k+1;^*)\delta\mathbf{u}(k+1)]\end{aligned} \tag{24-23}$$

As a consequence of relinearizing about $\hat{\mathbf{x}}(k|k)$, we know that

$$\delta\hat{\mathbf{x}}(k+1|k) = \mathbf{0} \tag{24-24}$$

$$\begin{aligned}\delta\hat{\mathbf{x}}(k+1|k+1) &= \hat{\mathbf{x}}(k+1|k+1) - \mathbf{x}^*(k+1) \\ &= \hat{\mathbf{x}}(k+1|k+1) - \hat{\mathbf{x}}(k+1|k)\end{aligned} \tag{24-25}$$

and

$$\begin{aligned}\delta\mathbf{z}(k+1) &= \mathbf{z}(k+1) - \mathbf{z}^*(k+1) \\ &= \mathbf{z}(k+1) - \mathbf{h}[\mathbf{x}^*(k+1), \mathbf{u}^*(k+1), k+1] \\ &= \mathbf{z}(k+1) - \mathbf{h}[\hat{\mathbf{x}}(k+1|k), \mathbf{u}^*(k+1), k+1]\end{aligned} \tag{24-26}$$

Substituting these three equations into (24-23), we obtain

$$\begin{aligned}\hat{\mathbf{x}}(k+1|k+1) = \hat{\mathbf{x}}(k+1|k) + \mathbf{K}(k+1;^*)\{\mathbf{z}(k+1) \\ - \mathbf{h}[\hat{\mathbf{x}}(k+1|k), \mathbf{u}^*(k+1), k+1] - \mathbf{H}_{\mathbf{u}}(k+1;^*)\delta\mathbf{u}(k+1)\}\end{aligned} \tag{24-27}$$

which is the *EKF correction equation*. Observe that the nonlinear nature of the system's measurement equation is used to determine $\hat{\mathbf{x}}(k+1|k+1)$. One usually sees this equation for the case when $\delta\mathbf{u} = \mathbf{0}$, in which case the last term on the right-hand side of (24-27) is not present.

To compute $\hat{\mathbf{x}}(k+1|k+1)$, we must compute the EKF gain matrix $\mathbf{K}(k+1;^*)$. This matrix, as well as its associated $\mathbf{P}(k+1|k;^*)$ and $\mathbf{P}(k+1|k+1;^*)$ matrices, depend on the nominal $\mathbf{x}^*(t)$ that results from prediction, $\hat{\mathbf{x}}(k+1|k)$. Observe, from (24-16), that $\mathbf{x}^*(k+1) = \hat{\mathbf{x}}(k+1|k)$ and that the argument of $\mathbf{K}$ in the correction equation is $k+1$; hence, we are indeed justified to use $\hat{\mathbf{x}}(k+1|k)$ as the nominal value of $\mathbf{x}^*$ during the calculations of $\mathbf{K}(k+1;^*)$, $\mathbf{P}(k+1|k;^*)$, and $\mathbf{P}(k+1|k+1;^*)$. These three quantities are computed from

$$\begin{aligned}\mathbf{K}(k+1;^*) &= \mathbf{P}(k+1|k;^*)\mathbf{H}'_{\mathbf{x}}(k+1;^*)[\mathbf{H}_{\mathbf{x}}(k+1;^*) \\ &\quad \mathbf{P}(k+1|k;^*)\mathbf{H}'_{\mathbf{x}}(k+1;^*) + \mathbf{R}(k+1)]^{-1}\end{aligned} \tag{24-28}$$

$$\begin{aligned}\mathbf{P}(k+1|k;^*) &= \mathbf{\Phi}(k+1,k;^*)\mathbf{P}(k|k;^*)\mathbf{\Phi}'(k+1,k;^*) \\ &\quad + \mathbf{Q}_d(k+1,k;^*)\end{aligned} \tag{24-29}$$

$$\mathbf{P}(k+1|k+1;^*) = [\mathbf{I} - \mathbf{K}(k+1;^*)\mathbf{H}_{\mathbf{x}}(k+1;^*)]\mathbf{P}(k+1|k;^*) \tag{24-30}$$

Remember that in these three equations * denotes the use of $\hat{\mathbf{x}}(k+1|k)$.

The EKF is very widely used, especially in the aerospace industry; however, it does not provide an optimal estimate of $\mathbf{x}(k)$. The optimal estimate of $\mathbf{x}(k)$ is still $\mathbf{E}\{\mathbf{x}(k)|\mathbf{Z}(k)\}$, regardless of the linear or nonlinear nature of the system's model. The EKF is a first-order approximation of $\mathbf{E}\{\mathbf{x}(k)|\mathbf{Z}(k)\}$ that sometimes works quite well, but cannot be guaranteed always to work well. No convergence results are known for the EKF; hence, the EKF must be viewed as an ad hoc filter. Alternatives to the EKF, which are based on nonlinear filtering, are quite complicated and are rarely used.

A flow chart for implementation of the EKF is depicted in Figure 24-2.

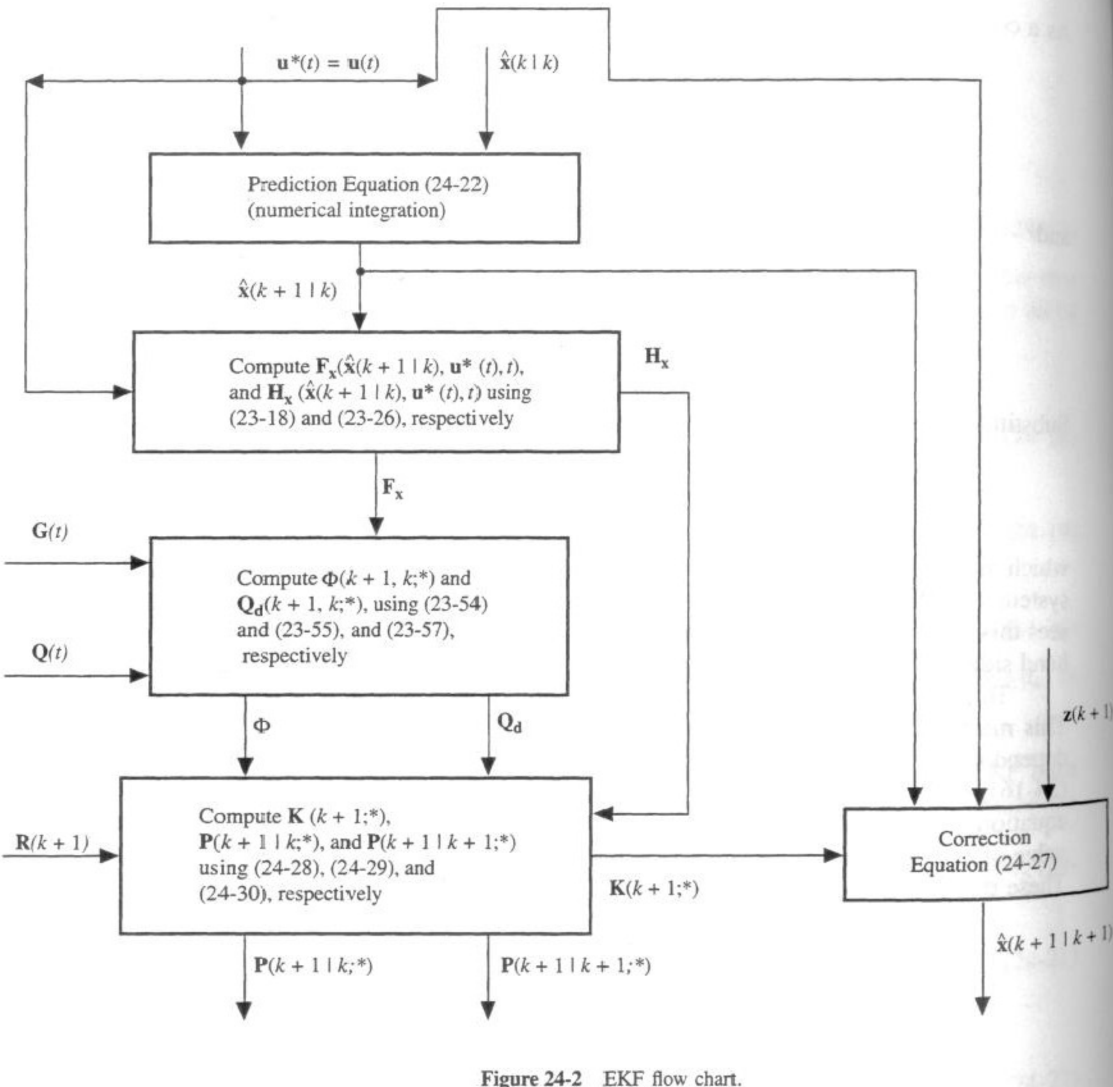

Figure 24-2 EKF flow chart.

The EKF is designed to work well as long as $\delta\mathbf{x}(k)$ is "small". The *iterated EKF* (Jazwinski, 1970), depicted in Figure 24-3, is designed to keep $\delta\mathbf{x}(k)$ as small as possible. The iterated EKF differs from the EKF in that it iterates the correction equation L times until $\|\hat{\mathbf{x}}_L(k+1|k+1) - \hat{\mathbf{x}}_{L-1}(k+1|k+1)\| \le \epsilon$. Corrector 1 computes $\mathbf{K}(k+1;^*)$, $\mathbf{P}(k+1|k;^*)$, and $\mathbf{P}(k+1|k+1;^*)$ using $\mathbf{x}^* = \hat{\mathbf{x}}(k+1|k)$; corrector 2 computes these quantities using $\mathbf{x}^* = \hat{\mathbf{x}}_1(k+1|k+1)$; corrector 3 computes these quantities using $\mathbf{x}^* = \hat{\mathbf{x}}_2(k+1|k+1)$; etc.

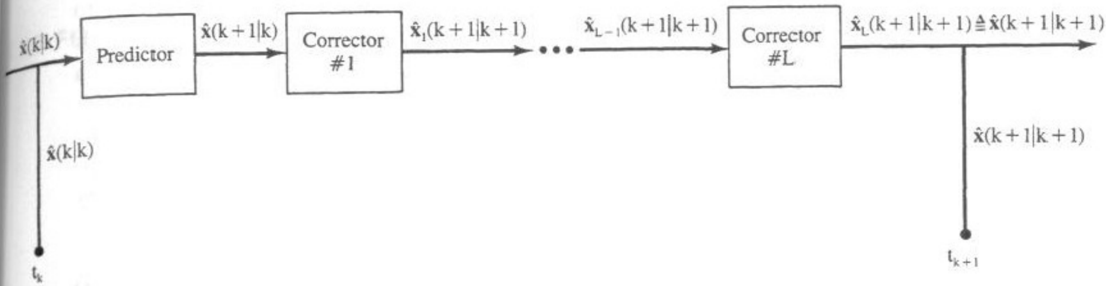

Figure 24-3 Iterated EKF. All the calculations provide us with a refined estimate of $\mathbf{x}(k+1)$, $\hat{\mathbf{x}}(k+1|k+1)$, starting with $\hat{\mathbf{x}}(k|k)$.

Often, just adding one additional corrector (i.e., $L = 2$) leads to substantially better results for $\hat{\mathbf{x}}(k+1|k+1)$ than are obtained using the EKF.

APPLICATION TO PARAMETER ESTIMATION

One of the earliest applications of the extended Kalman filter was to parameter estimation (Kopp and Orford, 1963). Consider the continuous-time linear system

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{w}(t) \tag{24-31a}$$

$$\mathbf{z}(t) = \mathbf{H}\mathbf{x}(t) + \nu(t), \quad t = t_i, \quad i = 1, 2, \ldots \tag{24-31b}$$

Matrices $\mathbf{A}$ and $\mathbf{H}$ contain some unknown parameters, and our objective is to estimate these parameters from the measurements $\mathbf{z}(t_i)$ as they become available.

To begin, we assume differential equation models for the unknown parameters, i.e., either

$$\dot{a}_l(t) = 0, \qquad l = 1, 2, \ldots, l^* \tag{24-32a}$$

$$\dot{h}_j(t) = 0, \qquad j = 1, 2, \ldots, j^* \tag{24-32b}$$

or

$$\dot{a}_l(t) = c_l a_l(t) + n_l(t), \qquad l = 1, 2, \ldots, l^* \tag{24-33a}$$

$$\dot{h}_j(t) = d_j h_j(t) + \eta_j(t), \qquad j = 1, 2, \ldots, j^* \tag{24-33b}$$

In the latter models $n_l(t)$ and $\eta_j(t)$ are white noise processes, and we often choose $c_l = 0$ and $d_j = 0$. The noises $n_l(t)$ and $\eta_j(t)$ introduce uncertainty about the "constancy" of the a_l and h_j parameters.

Next, we augment the parameter differential equations to (24-31a) and (24-31b). The resulting system is nonlinear, because it contains products of states [e.g.,