FORECASTING:
STATE-SPACE MODELS AND KALMAN FILTER ESTIMATION

Kemal Gürsoy[1] and Melike Baykal-Gürsoy[2]

[1]Department of Mathematics, Bogazici University
Anderson Hall, Bebek, Istanbul, Turkey
e-mail: gursoyk@boun.edu.tr

[2]Laboratory for Stochastic Systems
Department of Industrial&Systems Engineering
Rutgers, The State University of New Jersey
e-mail: gursoy@rci.rutgers.edu

Abstract

We present continuous and discrete time linear state-space models and explain the Kalman filter algorithm that is used to obtain the one-step-ahead state estimations recursively. Adaptive and nonlinear state estimations, as well as parameter estimation, with a Kalman filter are also briefly described.

Introduction

Humans observe their environment, learn its properties, and by forecasting the future events, they make plans to influence future outcomes to their benefit. There are risks that those plans, for which actions are based, are not good at all. Hence, it is desirable to set up expectations, based on the detailed analysis of empirical evidence and sound knowledge in order to reduce the risk of future regrets.

Forecasting methods to set up reasonable expectations for the future events are mainly based on the properties of relevant historical data (time-series) assuming that the environment will not change significantly. Forecasting models express relationships between what the past evidence shows and what could possibly take place in the future.

System models are useful concepts to represent relations, interactions and to build up tools for predicting the collective behaviour of entities. An input-output model expresses the behaviour of an open system that interacts with its environment, in a causal manner (the past and the present of the system shape its future), which takes from its environment (input, or stimulus) and gives to its environment (output, or response). If one is interested in the behaviour of an observable (measurable) open system, then one should observe the system for some time for stimulus-response relations and construct an input-output model that suitably fits to these observations.

A similar causal model (local in time) is the state-space representation of an open system, where the underlying system's behaviour is represented by its state and its response to an input at that state. The collection of all possible (admissible) states of a system is known as the state-space [1, 15, 7]. In order to predict the future states of a dynamic system, with minimum mean square error, we can design a recursive predictor based on the current estimate and the current filtered estimation error. The best estimate of the state of a linear stochastic system from partial observations is known as Kalman filter [11, 12] (a.k.a. Kalman-Bucy filter).

Previous work on this subject has been done by Kolmogorov [14], Wiener [24], Kalman [11], Bucy [13] and many others in the last century. We can also mention closely related work by Wald [22], and Robbins and Monro [20], where the *stochastic approximation* algorithm developed in [20] may be utilized for parameter estimation.

Stationary and nonstationary time series models such as autoregressive (AR), moving average (MA), mixed AR and MA (ARMA), autoregressive integrated moving average (ARIMA), ARMA with external input (ARMAX), can be translated into state-space models [3, 4, 7, 15, 10], thus benefit from the recursive prediction algorithm. Multivariate time series are especially suitable to state-space representation. For example, Aviv [2] uses state-space models and Kalman filter in demand estimation and inventory control. Bayesian estimation can also exploit the Kalman filtering methodology, as discussed in [19, 23].

We introduce the general state-space model in the next section, and then focus on linear Gaussian model in Section 2. We describe Kalman filter iterations in Section 3. The adaptive and nonlinear generalizations follow as Sections 4 and 5, respectively. The parameter estimation can also be approached by utilizing a Kalman filter, as presented in Section 6.

1 State-Space Structure

A mathematical model of a dynamic system with multiple inputs and outputs can be constructed as follows. Let x_t be the multidimensional (vector) state of the system, $\mathcal{X}$ be the state-space, u_t be an observable multidimensional input to the system and y_t be an observable multidimensional output of the system, at time t. Also, let v_t and w_t denote multidimensional random noise processes, at time t. Assume $x_{t+1} = f_t(x_t, u_t, v_t)$, $y_t = g_t(x_t, w_t)$, where f, g are measurable functions at all time. By using this simple model, we try to capture and express the relationships between the state of the system, the input that drives it, and the output of the system. The k-step ahead state, x_{t+k}, depends upon the present state, x_t, as well as the future inputs, such that, $x_{t+k} = f_{t+k-1}(f_{t+k-2}(...(f_t(x_t, u_t, v_t))...), u_{t+k-1}, v_{t+k-1})$. This recursion is the basis for state transitions from period t to k-step ahead period $t + k$.

2 Linear-Gaussian Model

In this section, we will consider linear state-space models operating in continuous or discrete time. Let the initial (starting) state be denoted as $x(t_0)$ for the continous and $x(k_0)$ for the discrete time process.

Continuous-Time Model: Assume that the state dynamics, or evolution is as follows,

$$\frac{d}{dt}x(t) = A_t x(t) + B_t u(t) + G_t v(t).$$

Also the observed output, or response is given by

$$y(t) = C_t x(t) + w(t),$$

where $x(t), u(t), y(t)$ are finite real vectors, A_t, B_t, C_t, G_t are real matrices with proper dimensions and $t \geq t_0$ is a real number.

The noise processes $\{v(t)\}, \{w(t)\}$ are independent Gaussian random vectors, with zero means and finite variances, conditionally independent of $x(t)$ and $u(t)$. Let $\{v(t)\}$ be a white Gaussian stochastic process with covariance $E[v(t)v^T(t-\tau)] = Q(t)\delta(t-\tau)$, and $\{w(t)\}$ be a white Gaussian stochastic process with covariance $E[w(t)w^T(t-\tau)] = R(t)\delta(t-\tau)$, where $\delta(.)$ is the Kronecker delta function, i.e. its value is one when its argument is zero, its value is zero otherwise. Since these stochastic processes are uncorrelated, we have $E[v(t)w^T(t)] = E[x(0)v^T(t)] = E[x(0)w^T(t)] = E[u(t)v^T(t)] = E[u(t)w^T(t)] = 0$.

Discrete-Time Model: Assume that the state dynamics is given by

$$x(k+1) = A_k x(k) + B_k u(k) + G_k v(k),$$

and the observation equation is

$$y(k) = C_k x(k) + w(k),$$

where $x(k), u(k), y(k)$ are finite real vectors, A_k, B_k, C_k, G_k are real matrices with proper dimensions and $k \geq k_0$ is an integer.

The noise processes $\{v(k)\}, \{w(k)\}$ are independent Gaussian random vectors, with zero means and finite variances, conditionally independent of $x(k)$ and $u(k)$, and temporally uncorrelated, i.e. $E[v(k)v^T(j)] = Q(k)\delta_{kj}$, $E[w(k)w^T(j)] = R(k)\delta_{kj}$, where δ_{kj} is the Kronecker delta, i.e. its value is one when $k = j$, and zero otherwise.

In these models, the system's evolution can be represented by state transition,

$$x(t|t_0) = \Phi(t, t_0)x_{t_0} + \int_{t_0}^{t} \Phi(t, s)B(s)u(s)ds,$$

where $\Phi(.,.)$ is the state-transition function (a matrix, or a dynamic semi-group operator on the state-space), $x(t_0)$ is the initial state of the system and integral is a summation operator in the Lebesgue sense [17] that operates both in the continuous-time and in the discrete-time, componentwise.

Note that for linear systems, the initial state is a critical point for the future trajectory of the system. On the other hand, if the system is not linear, then the initial state of the system is not sufficient for the system's behaviour, system's trajectories would depend upon other factors as well. Many trajectories (perhaps infinitely many) from the same initial state are possible. There are two essential questions:

Q1: Is the system *observable*, i.e. is it possible to determine the states of the system $\{x(s)|t_0 \leq s < t\}$, from the observations of $y(s)$ and $u(s)$, for $s \in [t_0, t)$?

Q2: Is the system *controllable*, i.e. is it possible to drive the system into a desired state $x(t_f)$, from an initial state $x(t_0)$, by selecting a stream of feasible inputs (stimuli) $\{u(s)|t_0 \leq s \leq t_f\}$, in a finite interval of time $[t_0, t_f]$?

For the discrete-time system dynamics, we replace t, t_0, t_f by k, k_0, k_f, respectively.

The duality of these questions, as well as the state-space formulation of linear systems for predicting and controlling the system's behavior were explained by many researchers, for one particular approach see [12].

The first question is relevant to forecasting, or estimating the states of a linear dynamical system, and can be approached by the Wiener filter [24]. Wiener filter provides the best estimates (in the sense of minimum mean square error [5]) of the Markovian signals $x(t)$, by using a set of observations $\{y(s),\ s \in [t_0, t]\}$. State estimates, $\hat{x}(s|t)$ may be obtained for prediction, or, $\hat{x}(t|t) := \hat{x}(t)$, for filtering when $s > t > t_0$. The Wiener filter is based on a function, $h(t)$, such that its convolution with the past observations, $h \star y$, would generate the state estimator,

$$\hat{x}(t) = \int_0^\infty h(s)y(t-s)ds.$$

The function, $h(t)$ is selected to minimize the expected squared estimation error,

$$E[(x(t) - \hat{x}(t))^2] = E[x^2(t)] - 2E[\hat{x}(t)x(t)] + E[\hat{x}^2(t)].$$

Since $E[\hat{x}^2(t)] = \int_0^\infty h(s)ds \int_0^\infty h(\tau)R_{yy}(t,\tau)d\tau$ and $E[\hat{x}(t)x(t)] = \int_0^\infty h(s)R_{xy}(t,s)ds$, where R_{yy} is the autocorrelation function of the observations $y(t)$ and R_{xy} is the correlation function of $x(t)$ and $y(t)$, then one can introduce an optimal $h(t)$, say $h^*(t)$, such that $h^*(t) = h(t) + \epsilon w(t)$, where $w(t)$ is a perturbation and ϵ is a scaling parameter. Consequently, $\frac{\partial E[(x(t)-\hat{x}(t))^2]}{\partial \epsilon} = 0$ is required for the expected minimum error square. Hence, the expected minimum mean square error will be achieved when h^* satisfies,

$$\int_0^\infty h^*(\tau)R_{yy}(t,\tau)d\tau = R_{xy}(t),$$

which is known as the Wiener-Hopf equation, and its solution for h^* yields the best linear estimator for the state $x(t)$ [24].

3 Kalman Filter for State Estimation

The Kalman filter, a.k.a. Kalman-Bucy filter, is a sequential estimator for the states of a stationary linear dynamical system and it is based on the Wiener filter, subject to a white Gaussian noise. Without loss of generality, we can set the initial time t_0 to be 0. The usual Kalman filter formulation, such as given in [21], for a known distribution of the initial state, $x(0)$, is carried on the following simplified state-space equations. Note that, the control component is not included in this model, thus one is only concerned with the state estimation problem.

Continuous Time: $\frac{d}{dt}x(t) = A_t x(t) + G_t v(t)$ and $y(t) = C_t x(t) + w(t)$, for $t \geq 0$.
Discrete Time: $x(k+1) = A_k x(k) + G_k v(k)$ and $y(k) = C_k x(k) + w(k)$, for $k \geq 0$.

The Kalman filter first estimates the state of the system, then updates the estimation process, according to the estimation error that manifests itself in the difference of actual observation and the estimated observation which is based on the state estimation. Hence, these two stages of estimation-correction process repeats itself for the entire forecasting horizon. In our model, noises and states are independent Gaussian random processes with finite second moments. Also let $E[x(0)] = \bar{x}(0)$ and $Var(x(0)) = P_0$.

For continuous time case: Let $\hat{x}(s|t)$ be the estimate of $x(s)$, given all $y(\tau)$, for $0 \leq \tau \leq t$ and $s > t \geq 0$. Also, let $e(s|t) = x(s) - \hat{x}(s|t)$ be the estimation error, at time s.

Therefore, one must find the linear filter $K(t, \tau)$ such that $\hat{x}(s|t) = \int_0^t K(s, \tau) y(\tau) d\tau$, where $\hat{x}(0|0) = \bar{x}(0)$. Also, the mean squared error, $E[||e(s|t)||^2]$ is minimized by this filter, where $||e(s|t)||$ denotes the Euclidean norm of the error vector.

If this filter, $K(t, \tau)$, satisfies the Wiener-Hopf equation, then $\hat{x}(s|t)$ will be an unbiased minimum variance estimator of $x(s)$, for the proof see [13]. Hence, this filter will satisfy,

$$\frac{d}{dt}\hat{x}(t|t) = \int_0^t \frac{\partial}{\partial t} K(t, \tau) y(\tau) d\tau + K(t, t) y(t),$$

and

$$\frac{d}{dt}\hat{x}(t|t) = A_t \hat{x}(t|t) + K(t, t)[y(t) - C_t \hat{x}(t|t)],$$

where $\hat{x}(0|0) = \bar{x}(0)$. The estimation error, $e(t|t) = x(t) - \hat{x}(t|t)$, will be subject to the following evolution,

$$\frac{d}{dt}e(t|t) = [A_t - K(t, t) C_t] e(t|t) - G_t v(t) - K(t, t) w(t).$$

Let the covariance of the estimation error be $E[e(t|0) e^T(t|0)] = P(t)$, and $P(0) = P_0$.

In order to obtain the least square error estimator, the estimation error, $e(t|t)$, must be orthogonal to the state estimate, $\hat{x}(t|t)$, by the Pythagoras theorem. Hence, the orthogonal projection of the error vector on the state-space would give, $E[e(t|t)\hat{x}(t|\tau)] = 0$, for $t \geq \tau \geq 0$.

Thus, by solving for the above estimation error differential equation, we obtain the evalution of the error function as follows;

$$e(t|t) = \Phi(t,0)e(0|0) + \int_0^t \Phi(t,\tau)[K(\tau,\tau)w(\tau) - G_\tau v(\tau)]d\tau,$$

where $\Phi(.,.)$ denotes the transition function for e. Therefore,

$$\begin{aligned} P(t) &= \Phi(t,0)E[e(0|0)e^T(0|0)]\Phi^T(t,0) \\ &+ \int_0^t \int_0^t \Phi(t,\tau)[K(\tau,\tau)E[w(\tau)w^T(s)]]\Phi^T(t,s)dsd\tau \\ &+ \int_0^t ds \int_0^t \Phi(t,\tau)G_\tau E[v(\tau)v^T(s)]G_s^T\Phi^T(t,\tau)d\tau. \end{aligned}$$

By taking the time derivative of the above function, we have the following Riccati equation for the covariance of the estimation error,

$$\frac{d}{dt}P(t) = G_tQ(t)G_t^T + A_tP(t) + P(t)A_t^T - P(t)C_t^TR^{-1}(t)C_tP(t),$$

where, $P(t)C_t^T = E[e(t|t)e^T(t|t)]C_t^T = K(t,t)R(t)$. Hence, this yields $K(t,t) = P(t)C_t^TR^{-1}(t)$.

In summary, the Kalman filter is defined by the following equations, for $t \geq 0$:

$\frac{d}{dt}\hat{x}(t|t) = A_t\hat{x}(t|t) + K(t,t)[y(t) - C_t\hat{x}(t|t)]$, where $\hat{x}(0|0) = \bar{x}(0)$,
$K(t,t) = P(t)C_t^TR^{-1}(t)$,
$\frac{d}{dt}P(t) = A_tP(t) + P(t)A_t^T + G_tQ(t)G_t^T - P(t)C_t^TR^{-1}(t)C_tP(t)$.

The forecasting of future states can be done by $\hat{x}(s|t) = \Phi(s,t)\hat{x}(t|t)$, for $s > t > 0$ and by using the state transition function $\Phi(.,.)$.

For discrete time case: Let $\hat{x}(k|m) = E[x(k)|y(0), y(1), \ldots, y(m)]$ be the state estimator and $e(k|m) = x(k) - \hat{x}(k|m)$ be the estimation error, for $k \geq m$. The best estimator of states has the minimum mean square estimation error property, which is obtained by the orthogonal projection of the state estimator on the observations, with the following Kalman recursions.

$$\hat{x}(k|k) = \hat{x}(k|k-1) + P(k-1)C_k^T[C_kP(k-1)C_k^T + R(k)]^{-1}(y(k) - C_k\hat{x}(k|k-1)),$$

$$P(k|k) = P(k-1) - P(k-1)C_k^T[C_kP(k-1)C_k^T + R(k)]^{-1}C_kP(k-1),$$

where $P(k|k) = E[e(k|k)e^T(k|k)]$. Also, $\hat{x}(k+1|k) = A_k\hat{x}(k|k)$, i.e.,

$\hat{x}(k+1|k) = A_k\hat{x}(k|k-1) + A_kP(k-1)C_k^T[C_kP(k-1)C_k^T + R(k)]^{-1}(y(k) - C_k\hat{x}(k|k-1))$, and

$e(k+1|k) = A_ke(k|k-1) + A_kP(k-1)C_k^T[C_kP(k-1)C_k^T + R(k)]^{-1}(y(k) - C_k\hat{x}(k|k-1) - G_kv(k))$.

The estimation error covariance, $P(k) = E[e(k+1|k)e^T(k+1|k)]$, i.e.

$P(k) = A_kP(k-1)A_k^T - A_kP(k-1)C_k^T[C_kP(k-1)C_k^T + R(k)]^{-1}C_kP(k-1)A_k^T + G_kQ(k)G_k^T$,

where, $\hat{x}(0|0) = E[x(0)]$ and $P(0) = Var(x(0))$.

This discrete time Kalman filter is a minimum variance and minimum square error sequential estimator of the state.

We can construct a predictor, based on this discrete Kalman filter, to forecast a distant future, $\hat{x}(n|k) = \prod_{j=k,\ldots,n} A_j\hat{x}(k|k)$, for $n > k > 0$.

In this approach, predictor-corrector structure of the Kalman filter is generating an *innovation* sequence, $y(k) - \hat{y}(k)$, which is a zero-mean and independent stochastic process (based on the maximum likelihood estimation, or, orthogonal projection method). There are other approaches for optimal state estimators, such as recursive Bayesian estimators, but when the system is linear and the noise structure is independent Gaussian, these methods converge to the Kalman filter [13].

4 Adaptive State Estimation

When there are uncertainties in the system parameters, Kalman filter cannot provide the state estimation. We can approach to those cases with a Bayesian construct [16, 23, 19]. Let the parameter vector θ is subject to uncertainty, where the linear system dynamics are as before.

For continuous time:

$\frac{d}{dt}x(t) = A(t,\theta)x(t) + G(t,\theta)v(t)$ and $y(t) = C(t,\theta)x(t) + w(t)$, where $v(t)$ and $w(t)$ are independent zero mean white Gaussian random vectors with $Q(t)\delta(t)$ and $R(t)\delta(t)$ covariances, respectively. The initial state is an independent Gaussian random vector with mean $\hat{x}(0|0,\theta)$ and variance $P(0|0,\theta)$.

The system evolution is subject to its unknown parameter vector, $\theta \in \Omega$, with its prior probability density $f(\theta)$. Based on the observation data, $Y_t = \{y(s)|0 \leq s \leq t\}$, the minimum mean square estimate of the state, $\hat{x}(t|t) = \int_\Omega \hat{x}(t|t,\theta)f(\theta|Y_t)d\theta$, where $\hat{x}(t|t) = E[x(t)|Y_t]$, $\hat{x}(t|t,\theta) = E[x(t)|Y_t,\theta]$, over the sample space of the parameter, Ω.

The posterior probability density of the parameter, based on the observations, would be as follows.

$f(\theta|Y_t) = f(\theta)exp(-\int_0^t \hat{x}(\tau|\tau,\theta)C^T(\tau,\theta)R^{-1}(\tau)y(\tau)d\tau$
$-\frac{1}{2}\int_0^t ||C(\tau,\theta)\hat{x}(\tau|\tau,\theta)||^2R^{-1}(\tau)d\tau)exp(-\int_0^t \hat{y}^T(\tau|\tau)R^{-1}(\tau)y(\tau)d\tau$
$-\frac{1}{2}\int_0^t ||\hat{y}(\tau|\tau)||^2R^{-1}(\tau)d\tau)$,

where $\hat{y}(\tau|\tau) = \int_\Omega C(\tau,\theta)\hat{x}(\tau|\tau)f(\theta|Y_t)d\tau$.

The conditional covariance of the state estimation error would be,

$$P(t) = \int_{\Omega} P(t|t,\theta)[\hat{x}(t|t,\theta) - \hat{x}(t|t)][\hat{x}(t|t,\theta) - \hat{x}(t,t)]^T f(\theta|Y_t) d\theta,$$

where $P(t|t,\theta) = E[[x(t) - \hat{x}(t|t,\theta)][x(t) - \hat{x}(t|t,\theta)]^T|Y_t,\theta]$, for all $\theta \in \Omega$.

The same procedure is also suitable for the discrete time estimations, as follows.

For discrete time:

$x(k+1) = A(k,\theta)x(k) + v(k)$ and $y(k) = C(k,\theta)x(k) + w(k)$, where $v(k)$, $w(k)$ are independent zero mean Gaussian vectors with covariances $Q(k)\delta_{kj}$ and $R(k)\delta_{kj}$ respectively.

The initial state is an independent random vector with mean $\hat{x}(0|0,\theta)$ and covariance $P(0|0,\theta)$. Let the unknown parameter has a prior probability density $f(\theta)$. Hence, based on a realization of past observations, $Y_k = \{y(j)|j = 1,2,\ldots,k\}$, the minimum mean square error state estimate would be $\hat{x}(k|k) = \int_{\Omega} \hat{x}(k|k,\theta) f(\theta|Y_k) d\theta$, where $\hat{x}(k|k) = E[x(k)|Y_k]$ and $\hat{x}(k|k,\theta) = E[x(k)|Y_k,\theta]$.

The posterior probability of the parameter, based on the observation data,

$f(\theta|Y_k) = [|P_z(k|k-1,\theta)|^{-\frac{1}{2}} exp(-\frac{1}{2}||y(k) - C(k,\theta)\hat{x}(k|k,\theta)||^2 P_z^{-1}(k|k-1,\theta)) f(\theta|Y_{k-1})]/[\int_{\Omega} f(\theta|Y_{k-1})|P_z(k|k-1,\theta)|^{-\frac{1}{2}}||y(k) - C(k,\theta)\hat{x}(k|k,\theta)||^2 P_z^{-1}(k|k-1,\theta)d\theta]$,

where $y(k) - C(k,\theta)\hat{x}(k|k,\theta)$ is a conditional white noise process with covariance $P_z(k|k-1,\theta) = C(k,\theta)P(k|k,\theta)C^T(k,\theta) + R(k)$.

Moreover, the state estimation error covariance would be

$P(k|Y_k) = \int_{\Omega}(P(k|k,\theta) + [\hat{x}(k|k,\theta) - \hat{x}(k|k)][\hat{x}(k|k,\theta) - \hat{x}(k|k)]^T) f(\theta|Y_k) d\theta$.

5 Nonlinear State Estimation

The state estimation for nonlinear systems is very difficult, to say the least, unlike the linear systems. Under suitable conditions, a linearization could be an aproximation of a nonlinear system's representation, then we may employ the Kalman filter approach, but always be mindful of the divergence of the approximation from the actual system.

Let us modify our state-space model for the nonlinear structure, as follow.

1. $\frac{d}{dt}x(t) = g(x,t) + v(t)$ and $y(t) = h(x,t) + w(t)$, for continuous time,
2. $x(t+1) = g(x,k) + v(k)$ and $y(k) = h(x,k) + w(k)$, for discrete time.

where x, y are state and output (observation) vectors, also let g, h are differentiable vector functions, representing the system dynamics (evolution), and v, w are zero-mean independent Gaussian random vectors with time-varying covariances, Q and R, respectively.

1) Continuous time: At time t, the first-order Taylor series expansion for g and h, around the estimated state, $\hat{x}$, would be as follows.

$$g(x,t) = g(\hat{x},t) + A(\hat{x},t)(x - \hat{x}) + \epsilon$$

and

$$h(x,t) = h(\hat{x},t) + C(\hat{x},t)(x-\hat{x}) + \varepsilon,$$

where ϵ, ε are the approximation errors, such that

$$A(\hat{x},t) = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} \cdots \frac{\partial g_1}{\partial x_n} \\ \vdots \\ \frac{\partial g_n}{\partial x_1} \cdots \frac{\partial g_n}{\partial x_n} \end{pmatrix}, \; C(\hat{x},t) = \begin{pmatrix} \frac{\partial h_1}{\partial x_1} \cdots \frac{\partial h_1}{\partial x_n} \\ \vdots \\ \frac{\partial h_m}{\partial x_1} \cdots \frac{\partial h_m}{\partial x_n} \end{pmatrix},$$

$$and\ H_i(\hat{x},t) = \begin{pmatrix} \frac{\partial^2 h_i}{\partial x_1^2} \cdots \frac{\partial^2 h_i}{\partial x_1 \partial x_n} \\ \vdots \\ \frac{\partial^2 h_i}{\partial x_n \partial x_1} \cdots \frac{\partial^2 h_i}{\partial x_n^2} \end{pmatrix}, \; for\ i = 1, \ldots, m.$$

Here, the functions g and h are evaluated at $\hat{x}(t)$.

Then the extended Kalman filter for the state estimation would be as follows [6].

$\frac{d}{dt}\hat{x}(t) = g(\hat{x},t) + \hat{P}(t)C(\hat{x},t)R^{-1}(t)[y(t) - h(\hat{x},t)]$,
$\frac{d}{dt}\hat{P}(t) = \hat{P}(t)A^T(\hat{x},t)\hat{P}(t) + Q(t) - \hat{P}(t)C^T(\hat{x},t)R^{-1}(t)C(\hat{x},t)\hat{P}(t) + \sum_{i=1,\ldots,m} \hat{P}(t)H_i(\hat{x},t)I[i]R^{-1}(t)[y(t) - h(\hat{x},t)]$, where $I[i]$ is the ith row of the identity matrix I.

2) Discrete time:

The extended Kalman filter construct can be carried out as follows,

$\hat{x}(k+1|k+1) = g(\hat{x}(k|k)) + \hat{P}(k+1|k+1)C^T(\hat{x},k+1)R^{-1}(k+1)[y(k+1) - C(\hat{x},k+1)g(\hat{x}(k|k))]$,
$\hat{P}(k+1|k+1) = [R(k) + C^T(\hat{x},k+1)P(k+1|k)C(\hat{x},k+1)]^{-1}\hat{P}(k+1|k)$,
$\hat{P}(k+1|k) = A(\hat{x},k)\hat{P}(k|k)A^T(\hat{x},k) + Q(k)$.

There are methods to improve the convergency of extended Kalman filter [21], [18] and apply these improved extended Kalman filters to adaptive controller design [8], [9].

The linearization could also be made around a reference (nominal) trajectory, $\bar{x}(t)$, with a given $\bar{x}(0)$, such that $\frac{d}{dt}\bar{x}(t) = g(\bar{x}(t),t)$, for $t \geq 0$.

Also, let $\Delta x(t) = x(t) - \bar{x}(t)$ be a Gaussian perturbation from the reference trajectory. Hence, $\frac{d\Delta x(t)}{dt} = g(x(t),t) - g(\bar{x}(t),t) + v(t)$. Therefore, the first-order Taylor series approximation would be, $g(x(t),t) - g(\bar{x}(t),t) = A(\bar{x}(0),t)\Delta x(t)$, where $A(\bar{x}(0),t)$ is evaluated along the reference trajectory, $\frac{d\Delta x(t)}{dt} = A(\bar{x}(0),t)\Delta x(t) + v(t)$. Hence, the Kalman filter approach could be used with $\hat{x}(0) = \bar{x}(0)$.

6 Parameter Estimation

We can estimate the parameter, together with the state of a system, by using the above constructed extended Kalman filter structure. Let the state vector x be augmented with the parameter vector θ, such that,

$$z(t) = \begin{pmatrix} x(t) \\ \theta(t) \end{pmatrix} \; and \; V(t) = \begin{pmatrix} v(t) \\ 0 \end{pmatrix}.$$

1. $\frac{d}{dt}z(t) = g(z,t) + V(t)$ and $y(t) = h(z,t) + w(t)$, for continuous time,
2. $z(t+1) = g(z,k) + V(k)$ and $y(k) = h(z,k) + w(k)$, for discrete time.

A linearization is obtained by the first-order Taylor series expansion, around the augmented state estimate $\hat{z}$. Without loss of generality, in discrete time a linearized Kalman filter that generates state and parameter estimation would be as follows,

$\hat{z}(k+1) = \hat{z}(k+1|k) + K(k+1)[y(k+1) - h(\hat{z}(k+1|k), k)]$, with

$\hat{z}(0|0) = z(0)$,

$h(\hat{z}(k+1|k), k) = C(k, \hat{\theta}_k)\hat{x}(k+1|k)$,

$\hat{z}(k+1|k) = F(k)\hat{z}(k)$,

$K(k+1) = P(k+1|k+1)G(k)R^{-1}(k)$,

$P(k+1|k+1) = P(k+1|k) - P(k+1|k)G(k)[G^T(k)P(k+1|k)G(k) + R(k)]^{-1}G^T(k)P(k+1|k)$,

$P(k+1|k) = F(k)P(k|k)F^T(k) + Q(k)G(k)G^T(k)$, where F, G, R, Q are proper matrices for the augmented state-parameter vector z, and θ_0 is a priori information about the parameter.

These methods provide sequential estimation of the state and the parameter for stationary and linear, or linearized, systems yet they suffer immensely in real time computations. Divergency is a common problem for linearized models [9], that is the success of estimation depends upon a very good initial point for the nominal trajectory and persistent corrections for very well behaved systems.

References

[1] K. J. Astrom. *Introduction to Stochastic Control*. Academic Press, New York, NY, 1970.

[2] Y. Aviv. A time-series framework for supply-chain inventory management. 51(2):210–227, 2003.

[3] G.E.P. Box, G.M. Jenkins, and G.C. Reinsel. *Time Series Analysis: Forecasting and Control*. Prentice Hall, New Jersey, 1994.

[4] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting*. Springer, New York, 2002.

[5] C.F. Gauss. *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae: Translated by G.W. Steward as Theory of the Combination of Observations Least Subject to Error*. SIAM, Philadelphia, 1995.

[6] A. Gelb. *Applied Optimal Estimation*. MIT press, Massachusets, 1974.

[7] G. C. Goodwin and K. S. Sin. *Adaptive Filtering Prediction and Control*. Prentice Hall, Englewood Cliffs, New Jersey, 1984. Information and System Science Series, T. Kailath Series Editor.

[8] H.Ö. Gülcür, G. Akman, and K. Gürsoy. Extended Kalman filter design for time-varying system identification: An adaptive controller. *IFAC Transactions*, 1:452–473, 1982.

[9] K. Gürsoy. Time varying parameter estimation with extended Kalman filters for adaptive control applications. Master's thesis, METU, Turkey, 1981.

[10] J. D. Hamilton. *Time Series Analysis*. Princeton University Press, Princeton, NJ, 1994.

[11] R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82:35–45, 1960.

[12] R. Kalman. Canonical structure of linear dynamical systems. *Proceedings of the National Academy of Sciences*, 3:596–600, 1962.

[13] R. Kalman and R. Bucy. New results in linear filtering and prediction theory. *Trans. ASME J. Basic Eng.*, 83:95–108, 1961.

[14] A. Kolmogoroff. Interpolation und extrapolation von stationaren zufalligen folgen. *Bulletin de l'Academie des Sciences de USSR*, Ser. Math. 5:3–14, 1941.

[15] P. R. Kumar and P. Varaiya. *Stochastic Systems: Estimation, Identification and Adaptive Control*. Prentice Hall, 1986.

[16] D. Lainiotis. Optimal adaptive estimation: Structure and parameter adaptation. *IEEE Trans. Automatic Control*, AC-16:160–170, 1971.

[17] H. Lebesgue. *Measure and the Integral.* Holden-Day, San Fransisco, 1966.

[18] L. Ljung. Asymptotic behaviour of the EKF as a parameter estimator. *IEEE Trans. Automatic Control*, AC-24:36–50, 1979.

[19] R. J. Meinhold and N. D. Singpurwalla. Understanding the Kalman filter. *The American Statistician*, 37(2):123–127, 1983.

[20] H. Robbins and S. Monro. A stochastic approximation method. *Annals of Mathematical Statistics*, 22:400–425, 1951.

[21] G. Saridis. *Self-Organizing Control of Stochastic Systems.* Dekker, New York, 1977.

[22] A. Wald. *Sequential Analysis.* Wiley, New York, 1947.

[23] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models.* Springer, NY, NY, 2nd edition, 1999.

[24] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series with Engineering Applications.* MIT and Wiley, New York, 1949.