# Chapter 6

# Nonlinear State Estimation

KENNETH R. MUSKE
Computational Science Methods Group
Los Alamos National Laboratory
Los Alamos, NM 87545

THOMAS F. EDGAR
Department of Chemical Engineering
The University of Texas at Austin
Austin, TX 78712-1062

## 6.1 Introduction

The implementation of the nonlinear model predictive control techniques discussed in the previous chapter requires knowledge of the current state of the nonlinear system in order to compute the solution to the open-loop optimal control problem formulated at each control interval. Feedback in these controllers comes from the update of the current state. Since the full state of the nonlinear system is not directly measurable in most applications, some method of reconstructing the current state of the system from the measured outputs must be employed. In this chapter, several methods of state estimation for nonlinear systems are presented.

The estimation of the state of a linear system is performed using techniques based on well-established optimal linear estimation theory. Due to the mathematical complexity introduced by a nonlinear model, nonlinear optimal state estimation is much less established in practice. However, a

number of sub-optimal techniques that are computationally tractable have been proposed for solving the nonlinear problem. Each of these techniques is based on some simplification of or approximation to the underlying nonlinear stochastic system. Consequently, they have their associated advantages and limitations. It is these techniques, which can reasonably be implemented in real-time applications, that are discussed in this chapter.

We begin with a review of linear state estimation to set the stage for the nonlinear approaches. Linearized filters for the nonlinear system are the next topic and include extended Kalman filtering and statistical approximation methods. Nonlinear observers are then presented followed by a nonlinear moving horizon state estimator. We close with a brief discussion of combined state and parameter estimation. The goal of this chapter is to summarize the methods currently available for nonlinear state estimation from an applications-oriented viewpoint. Since these techniques will be implemented using a process control computer, we emphasize techniques in which the measured outputs are available at discrete sampling times.

Before proceeding with our discussion of state estimation, let us define the state, inputs, parameters, and outputs of a nonlinear dynamic system. The *state* is the minimum amount of information necessary at the current time to uniquely determine the dynamic behavior of the system at all future times given the inputs and parameters. An *input* is an independent variable of the system analogous to the forcing function of a dynamic differential equation. Chemical process examples include the feed rate and composition to a process unit and the inlet cooling medium temperature to a heat exchanger. *Parameters* consist of the physical properties used in the description of the system. For chemical processes, parameters include thermodynamic properties such as density and heat capacity, physical dimensions such as volume and area, and rate terms such as heat transfer coefficients and reaction rate constants. In the discussion that follows, we assume that the inputs and parameters of the system are known. An *output* of the system is a variable that can be measured directly. Outputs may be either a state variable or some function of the state. State estimation is the process of determining the state from the output measurements given a dynamic model of the system. We illustrate this terminology with the following example.

*Example 6.1*

Consider an exothermic reaction carried out in a perfectly mixed batch reactor with a constant temperature cooling coil. The reaction is second-order

and irreversible in which chemical species A dimerizes to form species B.

$$\mathrm{A} + \mathrm{A} \longrightarrow \mathrm{B}$$

The rate of reaction of species A is expressed as follows.

$$r = -kC_{\mathrm{A}}{}^2, \qquad k = k_0 \exp(-E_a/RT)$$

The rate coefficient, $k$, is described by an Arrhenius relation in which $k_0$ is the frequency factor, $E_a$ is the activation energy, $T$ is the reactor temperature, and $R$ is the gas constant. Heat is generated due to the heat of reaction and removed via heat transfer through the cooling coil in which $\Delta H$ is the heat of reaction, $V$ is the volume of the reactor, and $U$, $A$, and $T_c$ are the overall heat transfer coefficient, heat transfer area, and temperature of the cooling coil.

$$h = -\Delta H V k C_{\mathrm{A}}{}^2 + UA(T_c - T)$$

This expression assumes that the heat capacity of the cooling fluid is large and the resistance through the coil wall is negligible such that the coil temperature remains constant at the cooling fluid temperature.

A dynamic model of the batch reactor can be obtained from mass and energy balances in which $\rho$ and $C$ are the density and heat capacity of the fluid in the reactor, respectively.

$$\frac{dC_{\mathrm{A}}}{dt} = -k_0 \exp(-E_a/RT) C_{\mathrm{A}}{}^2 \tag{6.1}$$

$$\frac{dT}{dt} = -\frac{\Delta H}{\rho C} k_0 \exp(-E_a/RT) C_{\mathrm{A}}{}^2 + \frac{UA}{V\rho C}(T_c - T) \tag{6.2}$$

This model assumes that the heat of reaction, density, and heat capacity are independent of temperature.

For this batch reactor model, the concentration of species A, $C_{\mathrm{A}}$, and the reactor fluid temperature, $T$, are the state variables, the cooling coil temperature, $T_c$, is the input, and the remaining items are parameters. The dynamic model consists of the two differential equations in Eqs. 6.1 and 6.2 that determine the behavior of the state variables. In a typical application, the temperature is measured and used to estimate the concentration.

## 6.2 Linear Systems

In this section, the basic concepts of state estimation are discussed using linear models. These concepts are extended to nonlinear models in Section 6.4.

We adopt this approach in order to introduce the topic within a simpler framework. In addition, several nonlinear state estimation techniques are based on the application of linear state estimation on a linearization of the nonlinear model. More information on linear system theory and state estimation can be found in standard linear system theory reference texts such as Kwakernaak and Sivan [34] and Kailath [26].

### 6.2.1 Linear System Models

A continuous-time linear system model can be constructed using the following system of linear differential equations.

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) \end{aligned} \tag{6.3}$$

This representation is referred to as a linear *state-space* model in which $y \in \Re^p$ is the output, $u \in \Re^m$ is the input, $x \in \Re^n$ is the state of the system, and $A \in \Re^{n \times n}$, $B \in \Re^{n \times m}$, and $C \in \Re^{p \times n}$ are the system matrices. We restrict the input to bounded, integrable functions in the following discussion.

If the output measurements are available only at equally spaced time intervals or *sampling periods*, referred to as $\Delta t$, a discrete linear state-space model can be used to describe the dynamic behavior of the system. In this chapter, we represent discrete-time models with the discrete time index $k$ specified as a subscript and the system matrices denoted by an overbar.

$$\begin{aligned} x_{k+1} &= x\left((k+1)\Delta t\right) = \bar{A}x_k + \bar{B}u_k \\ y_k &= y\left(k\Delta t\right) = \bar{C}x_k \end{aligned} \tag{6.4}$$

The discrete-time system matrices are obtained from the continuous-time matrices in Eq. 6.3 as follows.

$$\bar{A} = e^{A\Delta t}, \qquad \bar{B} = \left( \int_0^{\Delta t} e^{A\tau} d\tau \right) B, \qquad \bar{C} = C$$

The difference equation in Eq. 6.4 is valid when the input remains constant over the sampling period and the output is sampled at the same time the input is injected [34]. Due to the long process response times for most chemical processes, sampling periods much larger than the computational time required by the controller are used. Therefore, the validity of Eq. 6.4 typically can be assumed for chemical process computer control applications.

The linear system models presented in Eqs. 6.3 and 6.4 are referred to as *time-invariant* since the system matrices are independent of time. For continuous-time systems, *time-varying* models, in which the system matrices are functions of time, are represented as follows.

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) \\ y(t) &= C(t)x(t) \end{aligned} \tag{6.5}$$

For the discrete-time system in Eq. 6.4, the time dependence of the system matrices is represented by a subscript indicating the discrete time index.

$$\begin{aligned} x_{k+1} &= \bar{A}_k x_k + \bar{B}_k u_k \\ y_k &= \bar{C}_k x_k \end{aligned} \tag{6.6}$$

### 6.2.2 Observability of Linear Systems

Before proceeding with a discussion of state estimation, we first consider whether it is possible to estimate the state from the output measurements for a given linear system model. If $C = 0$, for example, state estimation is clearly not possible because the output measurements are always zero and contain no information about the state. A system in which the state can be determined uniquely from the output measurements is referred to as *observable*. Observability for linear systems is a global property of the system matrices that does not depend on the specific state, input, or output values. A characterization of observable linear systems is developed as follows.

Consider the output of the discrete linear system in Eq. 6.4 at successive sample times up to sample time $k = n - 1$. The outputs are related to the initial state of the system, $x_0$, and the previous inputs, $\{u_0, u_1, \ldots, u_{n-2}\}$.

$$\begin{aligned} y_0 &= \bar{C}x_0 \\ y_1 &= \bar{C}\bar{A}x_0 + \bar{C}\bar{B}u_0 \\ y_2 &= \bar{C}\bar{A}^2x_0 + \bar{C}\bar{A}\bar{B}u_0 + \bar{C}\bar{B}u_1 \\ &\vdots \qquad\qquad \vdots \\ y_{n-1} &= \bar{C}\bar{A}^{n-1}x_0 + \bar{C}\bar{A}^{n-2}\bar{B}u_0 + \ldots + \bar{C}\bar{B}u_{n-2} \end{aligned}$$

These relations can be combined into a linear system of equations, $\bar{\mathcal{O}}x_0 = \mathcal{Y}$,

in which the $pn \times n$ matrix $\bar{\mathcal{O}}$ is called the *observability matrix.*

$$\bar{\mathcal{O}} = \begin{bmatrix} \bar{C} \\ \bar{C}\bar{A} \\ \vdots \\ \bar{C}\bar{A}^{n-1} \end{bmatrix}, \quad \mathcal{Y} = \begin{bmatrix} y_0 \\ y_1 - \bar{C}\bar{B}u_0 \\ \vdots \\ y_{n-1} - \bar{C}\bar{A}^{n-2}\bar{B}u_0 - \ldots - \bar{C}\bar{B}u_{n-1} \end{bmatrix} \tag{6.7}$$

The initial state $x_0$ can be determined uniquely if $\bar{\mathcal{O}}$ is full rank. In this case, the linear system in Eq. 6.4 is observable or reconstructable. Since $\bar{\mathcal{O}}$ is constructed from only $\bar{A}$ and $\bar{C}$, $(\bar{C}, \bar{A})$ is referred to as observable. If $\bar{\mathcal{O}}$ is not full rank, any initial state with a component in the null space of $\bar{\mathcal{O}}$ cannot be uniquely determined from the output measurements. In this case, the linear system in Eq. 6.4 is not observable and the null space of $\bar{\mathcal{O}}$ is called the *unobservable subspace* of the linear system. If the unobservable subspace does not contain unstable modes of $A$, the system is *detectable.*

The linear system model in Eq. 6.4 can be transformed into *observer canonical* or modal form using the nonsingular transformation matrix $V$. This transformation matrix is partitioned into the observable and unobservable subspaces of the linear system. The observable subspace, $V_o$, is composed of the basis vectors of the range of $\bar{\mathcal{O}}$. The unobservable subspace, $V_u$, is composed of the basis vectors of the null space of $\bar{\mathcal{O}}$.

$$z_k = Vx_k, \qquad V = \begin{bmatrix} V_o \\ V_u \end{bmatrix}, \qquad \begin{aligned} z_{k+1} &= V\bar{A}V^{-1}z_k + V\bar{B}u_k \\ y_k &= \bar{C}V^{-1}z_k \end{aligned}$$

The advantage of observer canonical form is that the observable modes, $z^o$, and the unobservable modes, $z^u$, are apparent by inspection of the partitioned system.

$$\begin{bmatrix} z^o_{k+1} \\ z^u_{k+1} \end{bmatrix} = \begin{bmatrix} \tilde{A}_o & 0 \\ \tilde{A} & \tilde{A}_u \end{bmatrix} \begin{bmatrix} z^o_k \\ z^u_k \end{bmatrix} + \begin{bmatrix} \tilde{B}_o \\ \tilde{B}_u \end{bmatrix} u_k$$

$$y_k = \begin{bmatrix} \tilde{C}_o & 0 \end{bmatrix} \begin{bmatrix} z^o_k \\ z^u_k \end{bmatrix}$$

Since the canonical form is generated by a similarity transform that results in a lower triangular matrix, the characteristic values of the system are composed of the eigenvalues of $\tilde{A}_o$ and $\tilde{A}_u$. The linear system is detectable if and

only if the eigenvalues of $\tilde{A}_u$, which are the characteristic values of the unobservable modes, have moduli strictly less than one. Detectability ensures that the modes of the system that cannot be observed or reconstructed from the output measurements are stable.

The continuous-time linear system in Eq. 6.3 is observable if there exists some $\delta$, $0 < \delta < \infty$, such that the state $x(t-\delta)$ can be uniquely determined from $y(\tau)$ and $u(\tau)$ for all $u(\tau)$ and $t - \delta \leq \tau \leq t$. It can be shown that when the observability matrix in Eq. 6.7 is constructed from the continuous-time system matrices $A$ and $C$, the same conditions for observability and detectability apply for continuous-time systems. A proof using the derivatives of the output and following an argument similar to that presented here for discrete-time systems is available in [26]. The continuous-time observer canonical form is constructed in the same manner as that for discrete-time systems. In this case, the system is detectable if and only if the eigenvalues of $\tilde{A}_u$ are strictly negative.

For time-varying linear systems, the conditions for observability are more complicated and involve the *state transition matrix*, $\Phi$, and the *observability Gramian matrix*, $M$. The time-varying, continuous-time system in Eq. 6.5 is observable if and only if for all $t > 0$ there exists a $\sigma > 0$ such that the observability Gramian matrix $M(t-\sigma, t)$ is nonsingular. The continuous-time observability Gramian is a symmetric, nonnegative definite matrix computed from the state transition matrix $\Phi(t, t_o)$.

$$M(t-\sigma, t) = \int_{t-\sigma}^{t} \Phi^T(\tau, t-\sigma) C^T(\tau) C(\tau) \Phi(\tau, t-\sigma) d\tau \qquad (6.8)$$

The continuous-time state transition matrix is the solution to the matrix differential equation

$$\frac{d\Phi(t, t_o)}{dt} = A(t)\Phi(t, t_o), \qquad \Phi(t_o, t_o) = I \qquad (6.9)$$

that describes the dynamic behavior of the following unforced or *autonomous* linear system.

$$\dot{x}(t) = A(t)x(t), \qquad x(t) = \Phi(t, t_o)x(t_o)$$

Note that when $A$ is a time-invariant matrix, the state transition matrix can be determined from the matrix exponential for $\bar{A}$ in Eq. 6.4.

The time-varying discrete linear system in Eq. 6.6 is observable if and only if for every $k$ there exists a $j \leq k-1$ such that the discrete observability

Gramian matrix

$$\bar{M}(j,k) = \sum_{i=j+1}^{k} \bar{\Phi}^T(i,j+1)\bar{C}_i^T\bar{C}_i\bar{\Phi}(i,j+1) \tag{6.10}$$

is nonsingular. The observability Gramian matrix is computed using the discrete state transition matrix, $\bar{\Phi}(k,j)$, which is the solution to the following difference equation.

$$\bar{\Phi}(k+1,j) = \bar{A}_k\bar{\Phi}(k,j), \qquad \bar{\Phi}(j,j) = I \tag{6.11}$$

This matrix describes the dynamic behavior of the autonomous discrete linear system.

$$x_{k+1} = \bar{A}_k x_k, \qquad x_k = \bar{\Phi}(k,0)x_0$$

A more detailed discussion of observability for linear systems can be found in Kwakernaak and Sivan [34] and Kailath [26].

## 6.3 Linear State Estimation

We denote estimated variables with the symbol "^". Since these variables are determined using the output measurements, we will also indicate the time index of the output measurements along with the time index of the estimate. For the discrete-time system in Eq. 6.4, $\hat{x}_{j|k}$ represents the estimate of the state at sample time $j$ given output measurements up to sample time $k$. For continuous-time systems, $\hat{x}\,(\tau \mid k)$ represents the estimate of the state at time $\tau$ given discrete output measurements up to sample time $k$.

The estimate $\hat{x}_{j|k}$ is a *smoothed* state estimate when $j < k$, a *filtered* state estimate when $j = k$, and a *predicted* state estimate when $j > k$. The filtered estimate is the current estimate based on all of the available output measurements. It is this estimate of the state that is normally used for control applications. Predicted estimates are obtained by using the model in Eq. 6.4 to extrapolate the filtered estimate into the future. These estimates typically are used to compute the objective or cost function in model predictive control formulations. Smoothed estimates are based on both past and future output measurements and provide a more accurate estimate than the filtered estimate. Smoothed estimates usually are computed off-line from previously collected process data and used in process analysis and diagnosis. For continuous-time systems, the estimate $\hat{x}\,(\tau \mid k)$ is a smoothed estimate

when $\tau < k\Delta t$, a filtered estimate when $\tau = k\Delta t$, and a predicted estimate when $\tau > k\Delta t$.

A very simple conceptual approach to discrete linear state estimation is to solve Eq. 6.7 for the unknown initial state $x_0$. The filtered state at sample period $k$ is then easily computed from the known inputs. However, in any practical application there are unmeasured disturbances present such that the solution to Eq. 6.7 may not exist or may result in an unreliable estimate. Therefore, an approach that considers disturbances in the determination of the state estimate is more useful.

### 6.3.1 Least Squares Estimator

We assume that disturbances to the state and measurement can be represented by additive terms to the state-space model. The discrete linear system in Eq. 6.4 with disturbances is expressed as

$$\begin{aligned} x_{k+1} &= \bar{A}x_k + \bar{B}u_k + w_k \\ y_k &= \bar{C}x_k + v_k \end{aligned} \tag{6.12}$$

in which $w \in \Re^n$ is the process or state disturbance vector and $v \in \Re^p$ is the measurement disturbance vector. The process disturbance vector models unmeasured disturbances to the process that effect the state. The measurement disturbance vector represents the error in the measuring device caused by instrument noise.

One approach that can be considered to estimate the state of the system in Eq. 6.12 is to minimize the estimated process and measurement disturbances in a least squares sense. The following least squares minimization problem is solved to obtain an estimate of the initial state for the discrete linear system with disturbances.

$$\min_{\{\hat{w}_{-1|k},\ldots,\hat{w}_{k-1|k}\}} \hat{w}_{-1|k}^T Q_0^{-1} \hat{w}_{-1|k} + \sum_{j=0}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=0}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \tag{6.13}$$

$$\text{Subject to:} \quad \begin{aligned} \hat{x}_{0|k} &= \bar{x}_0 + \hat{w}_{-1|k} \\ \hat{x}_{j+1|k} &= \bar{A}\hat{x}_{j|k} + \bar{B}u_j + \hat{w}_{j|k} \\ \hat{v}_{j|k} &= y_j - \bar{C}\hat{x}_{j|k} \end{aligned}$$

In this approach, $\hat{x}_{0|k}$ is the estimate of the initial state given $k$ output measurements, $\bar{x}_0$ is an *a priori* estimate of the initial state, $\hat{w}_{j|k}$ are the

estimated state disturbances in which $\hat{w}_{-1|k}$ is a correction to the *a priori* estimate, and $\hat{v}_{j|k}$ are the estimated measurement disturbances. The weighting matrices, $Q_0^{-1}$, $Q^{-1}$, and $R^{-1}$, specify the relative contribution of each of the terms in the quadratic objective. The inverse is used for notational consistency in the succeeding sections.

The optimal solution to this least squares problem at time $k$, $\hat{w}^*_{j|k}$, is used to compute the state estimate at time $j$ given the $k$ output measurements, $\hat{x}_{j|k}$, as follows.

$$\hat{x}_{j|k} = A^j \bar{x}_0 + \sum_{i=0}^{j} \bar{A}^{j-i} \hat{w}^*_{i-1|k} + \sum_{i=1}^{j} \bar{A}^{j-i} \bar{B} u_{i-1} \tag{6.14}$$

This expression computes a smoothed state estimate when $j < k$, a filtered state estimate when $j = k$, and a predicted state estimate when $j > k$.

The choice of the coefficients in the weighting matrices is a compromise between minimizing the estimated process disturbances versus minimizing the estimated measurement disturbances. This choice is based on the expected magnitudes of each of these disturbances. If the output measurements are reliable, then $R^{-1}$ is chosen to be large relative to $Q^{-1}$. This choice tends to reduce the deviations between the measured output and the predicted output by increasing the estimated process disturbances. In this case, process disturbances are estimated to bring the model into agreement with the output measurements. On the other hand, if the output measurements are poor, then $R^{-1}$ is chosen to be small relative to $Q^{-1}$. This choice prevents uncertain measurements from causing large estimated state disturbances. In this case, the model is assumed to be a more reliable indication of the state than the output measurements.

### 6.3.2 Recursive Estimation

The least squares approach presented in the previous section is referred to as non-sequential or *batch* estimation. It requires the solution of a least squares problem using all $k$ previous output measurements to obtain a filtered state estimate at each sample time. As time progresses, the number of decision variables in this optimization problem quickly becomes prohibitive, making it a rather cumbersome procedure that cannot be implemented in practice. However, a sequential or *recursive* solution to the batch estimation problem posed in Eq. 6.13 that requires information only from the previous sample

time can be constructed.

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k(y_k - \bar{C}\hat{x}_{k|k-1}) \tag{6.15}$$

$$\hat{x}_{k+1|k} = \bar{A}\hat{x}_{k|k} + \bar{B}u_k \tag{6.16}$$

$$\hat{x}_{0|0} = \bar{x}_0$$

The filtered state estimate is computed from the predicted state estimate and the output measurement at each sample time. A gain matrix, $L_k$, multiplies the difference between the current measured output, $y_k$, and the predicted output, $\bar{C}\hat{x}_{k|k-1}$, to form a correction to the predicted state estimate. The gain matrix is a function of the system and weighting matrices used in the batch least squares estimator as shown in the following section.

### 6.3.3 Kalman Filter

The least squares recursive estimator in Eq. 6.15 can also be derived by minimization of the mean square reconstruction error [2]. The reconstruction error is a measure of the deviation between the actual and estimated state.

$$e_{k|k-1} = x_k - \hat{x}_{k|k-1} \tag{6.17}$$

When $w_j$ and $v_j$ are independent, zero mean, normally distributed random variables with covariances $Q$ and $R$, respectively, and $\bar{x}_0$ is an independent, normally distributed random variable with covariance $Q_0$, the recursive estimator in Eq. 6.15 produces the minimum variance estimate of the state. For linear Gaussian systems this is also the most probable or *maximum likelihood* estimate. This recursive estimator is referred to as the discrete *Kalman filter*. A detailed probabilistic derivation of the Kalman filter can be found in Jazwinski [25] and Lewis [36].

In the probabilistic formulation, the covariances specify the expected magnitudes of the disturbances to the output measurement and the state. If the covariance of the measurement noise, $R$, is small relative to the covariance of the process noise, $Q$, the measurements are relatively noise free and the deviations between the measured output and the predicted output should be small. If the measurement noise covariance is large relative to the process noise covariance, the measurements are relatively uncertain and the feedback correction to the model prediction should be small.

The probabilistic approach provides a rigorous way to determine the tuning parameters for the estimator provided that the assumptions concerning

the stochastic process are valid and the covariances are known or can be determined. The filter gain is computed at each sampling time using the covariance of the state estimate in Eq. 6.16, which we denote as $P_k$.

$$L_k = P_k\bar{C}^T(\bar{C}P_k\bar{C}^T + R)^{-1} \tag{6.18}$$

This covariance is propagated using the discrete filtering Riccati equation.

$$\begin{aligned} P_{k+1} &= \bar{A}P_k\bar{A}^T + Q - \bar{A}P_k\bar{C}^T(\bar{C}P_k\bar{C}^T + R)^{-1}\bar{C}P_k\bar{A}^T \\ P_0 &= Q_0 \end{aligned} \tag{6.19}$$

The Riccati equation contains a term, $\bar{A}P_k\bar{A}^T + Q$, that increases the covariance of the state estimate at each sampling time due to the presence of state or process noise. The remaining term represents the contribution due to the output measurement and generally decreases the covariance. The recursion in Eq. 6.19 tends to a constant matrix at large $k$ when $(\bar{C}, \bar{A})$ is detectable, $Q$ and $R$ are positive definite, and $Q_0$ is non-negative definite.

$$P_\infty = \bar{A}(P_\infty - P_\infty\bar{C}^T(\bar{C}P_\infty\bar{C}^T + R)^{-1}\bar{C}P_\infty)\bar{A}^T + Q \tag{6.20}$$

This matrix is called the steady-state discrete filtering Riccati matrix. The steady-state Kalman filter gain is computed by using $P_\infty$ in Eq. 6.18.

For the time-varying discrete linear system in Eq. 6.6, the optimal filter gain is computed from the solution of the discrete filtering Riccati equation at each sampling time using the following expressions.

$$L_k = P_k\bar{C}_k^T(\bar{C}_kP_k\bar{C}_k^T + R)^{-1} \tag{6.21}$$

$$\begin{aligned} P_{k+1} &= \bar{A}_kP_k\bar{A}_k^T + Q - \bar{A}_kP_k\bar{C}_k^T(\bar{C}_kP_k\bar{C}_k^T + R)^{-1}\bar{C}_kP_k\bar{A}_k^T \\ P_0 &= Q_0 \end{aligned} \tag{6.22}$$

There is no general steady-state solution to this Riccati equation or steady-state filter gain for time-varying systems.

A stochastic model for the time-varying, continuous-time linear system in Eq. 6.5 can also be constructed.

$$\begin{aligned} \dot{x}(t) &= A(t)x(t) + B(t)u(t) + w(t) \\ y(t) &= C(t)x(t) + v(t) \end{aligned} \tag{6.23}$$

When $v(t) \in \Re^p$ and $w(t) \in \Re^n$ are independent, zero mean, white noise processes with spectral densities $Q$ and $R$, respectively, a continuous-time Kalman filter similar to that for discrete systems is obtained [21].

$$\dot{\hat{x}}(t \mid t) = A(t)\hat{x}(t \mid t) + B(t)u(t) + L(t)(y(t) - C(t)\hat{x}(t \mid t)) \quad (6.24)$$

$$L(t) = P(t)C(t)^T R^{-1} \quad (6.25)$$

$$\dot{P}(t) = A(t)P(t) + P(t)A(t)^T + Q - P(t)C(t)^T R^{-1} C(t)P(t) \quad (6.26)$$

$$P(0) = Q_0$$

The result in Eq. 6.24 is the minimum variance filtered estimate in which the filter gain, $L(t)$, is computed via the differential equation in Eq. 6.26 and $P(t)$ is the covariance of the state estimate $\hat{x}(t \mid t)$. There is again a term for the increase in the covariance due to the presence of state noise, $A(t)P(t) + P(t)A(t)^T + Q$, and a remaining term, $P(t)C(t)^T R^{-1} C(t)P(t)$, that generally decreases the covariance due to the contribution of the measurement.

The Kalman filter in Eq. 6.15 for discrete-time systems or Eq. 6.24 for continuous-time systems produces the minimum variance and maximum likelihood state estimate provided that the assumptions concerning the stochastic process are valid. In this case, the Kalman filter is the optimal linear state estimator. It is nominally stable provided $R$ is positive definite, $Q$ and $Q_0$ are positive semidefinite, and the corresponding linear system is detectable. *Nominal stability* ensures that the filtered state will converge to the true state in the absence of disturbances for any initial reconstruction error given an exact model of the process.

In addition to the standard linear system references previously cited, optimal linear state estimation is discussed in Åström [2] and Bryson and Ho [8]. A discrete-time presentation is provided in Catlin [10]. A discussion of square-root algorithms, which result in more accurate numerical computations, is presented in Park and Kailath [50]. An excellent practical review of linear filtering and smoothing is contained in Gelb [21].

For the linear discrete-time optimal estimator, the gain in Eq. 6.15 must be computed at each sampling time. A common approximation to the optimal estimator that avoids this computation is to use the steady-state solution of the discrete filtering Riccati equation in Eq. 6.20 to determine the steady-state Kalman filter gain matrix. The steady-state gain is then used at each sampling time. Another common approximation is to discontinue the update of the gain matrix at a given sample time and use this gain at each successive sampling time. This technique avoids the computation of

the steady-state gain matrix, but does not guarantee nominal stability when the gain calculation is stopped at an arbitrary sample time [15].

### 6.3.4 Luenberger Observer

Another method of constructing a recursive estimator is to use a constant gain matrix in Eq. 6.15 that achieves some desired performance criteria.

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L(y_k - \bar{C}\hat{x}_{k|k-1}) \tag{6.27}$$

The performance criterion is based on the dynamic behavior of the reconstruction error and ignores the stochastic properties of the disturbances $w_k$ and $v_k$. This recursive estimator is referred to as a deterministic or Luenberger *observer* [40]. For discrete-time systems, the observer is nominally stable if and only if the eigenvalues of the $n \times n$ matrix $\bar{A} - \bar{A}L\bar{C}$ have moduli strictly less than one. This result comes from the dynamic response of the reconstruction error from an initial nonzero value with no disturbances.

$$e_{k+1} = (\bar{A} - \bar{A}L\bar{C})e_k \tag{6.28}$$

The gain matrix can be determined by choosing the eigenvalues of $\bar{A} - \bar{A}L\bar{C}$, which are referred to as the *closed-loop observer poles.* This technique is referred to as *pole placement* and requires that the system be observable. If the gain is chosen such that all of the eigenvalues are zero, the result is a *deadbeat* observer. The choice of the observer poles is a compromise between rapid decay of the reconstruction error, which requires that the poles be placed close to the origin, and sensitivity to measurement noise and modeling error, which increases as the poles are moved toward the origin.

An observer for continuous-time systems can be constructed analogous to that for discrete-time systems. The reconstruction error for the estimate in Eq. 6.24 behaves as follows.

$$\begin{aligned} \dot{e}(t) &= (A - LC)e(t) \\ e(t) &= x(t) - \hat{x}(t \mid t) \end{aligned} \tag{6.29}$$

In this case, stability of the observer requires that the gain be selected such that the eigenvalues of $A - LC$, which are the closed-loop observer poles, are strictly negative. As the poles are moved further into the left half plane, the rate of convergence of the reconstruction error increases. For the continuous-time observer, there is also a compromise between the rapid decay of the reconstruction error and sensitivity to measurement noise and modeling error.

The Luenberger observer is a deterministic estimator that is employed when the stochastic model of the process is unknown or if the optimal filter exhibits poor performance. It can also be employed to reduce the computational requirements of the estimator. For modern chemical process control applications, however, computational time is generally not an issue with linear systems except in large-scale applications. Further discussion of observers for linear systems can be found in Kwakernaak and Sivan [34] and Kailath [26].

### 6.3.5 Moving Horizon Estimator

A recursive form of the batch state estimation problem in Eq. 6.13 can be constructed using a moving horizon approach. In this approach, the state estimate at time $k$ is determined recursively from the solution of the following least squares problem that uses the predicted estimate at time $k-N-1$, $\hat{x}_{k-N|k-N-1}$, and the most recent $N+1$ output measurements.

$$\min_{\{\hat{w}_{k-N-1|k},\ldots,\hat{w}_{k-1|k}\}} \quad \hat{w}_{k-N-1|k}^T P_{k-N}^{-1} \hat{w}_{k-N-1|k} + \sum_{j=k-N}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \tag{6.30}$$

$$\text{Subject to:} \quad \begin{aligned} \hat{x}_{k-N|k} &= \hat{x}_{k-N|k-N-1} + \hat{w}_{k-N-1|k} \\ \hat{x}_{j+1|k} &= \bar{A}\hat{x}_{j|k} + \bar{B}u_j + \hat{w}_{j|k} \\ \hat{v}_{j|k} &= y_j - \bar{C}\hat{x}_{j|k} \end{aligned}$$

The moving horizon of output measurements allows for a finite number of decision variables at each sampling time. This procedure is initialized by computing the first $N$ predicted estimates using the batch estimator in Eq. 6.13. The state estimate at time $n+j$ given $k$ output measurements, in which $n = k-N$, can then be computed from the predicted state estimate at sample time $n$ and the solution of the least squares problem in a manner similar to the batch estimator.

$$\hat{x}_{n+j|k} = A^j \hat{x}_{n|n-1} + \sum_{i=0}^{j} A^{j-i} \hat{w}_{n-1+i|k}^* + \sum_{i=1}^{j} A^{j-i} B u_{n-1+i} \tag{6.31}$$

This expression computes a smoothed state estimate when $j < N$, a filtered state estimate when $j = N$, and a predicted state estimate when $j > N$. It

can be shown that the filtered estimate from the moving horizon estimator is the optimal filtered estimate for the discrete stochastic linear system in Eq. 6.12 in which $v_k$, $w_k$ and $\bar{x}_0$ follow the same assumptions made in the Kalman filter, the weighting matrices $Q^{-1}$ and $R^{-1}$ are the inverses of the covariance matrices for $w_k$ and $v_k$, respectively, and the weighting matrix $P_{k-N}^{-1}$ is the inverse of the discrete filtering Riccati matrix at sampling time $k-N$ [47]. This matrix is the covariance of the state estimate $\hat{x}_{k-N|k-N-1}$ and is computed using the recursion in Eq. 6.19.

### Constraints

Since the moving horizon estimator produces the same estimate as the Kalman filter, there is no incentive to implement this approach due to the increased computational effort required to solve the least squares problem in Eq. 6.30. The motivation for employing the moving horizon formulation is the addition of constraints on the estimated states and state disturbances.

$$h_{\min} \leq H\hat{x}_{j|k} \leq h_{\max}, \qquad j = k-N+1, \ldots, k \tag{6.32}$$

$$w_{\min} \leq \hat{w}_{j|k} \leq w_{\max}, \qquad j = k-N, \ldots, k-1 \tag{6.33}$$

The estimated state constraint in Eq. 6.32 specifies maximum and minimum limits on the estimate of the state. These constraints are applied to prevent physically unrealistic state estimates, such as negative concentrations, that can be due to spurious output measurements. The estimated state disturbance constraint in Eq. 6.33 specifies limits on the estimated state disturbances. These limits can be viewed as altering the probability distribution function of the state disturbances such that the probability of any state disturbance outside of the constraints is zero. This constraint prevents estimated state disturbances that cannot realistically occur in the process.

The constraints are chosen to define a convex region in the $\hat{w}$ space containing the origin. For systems forced by an input, the state constraints are shifted by the nominal trajectory due to the input. A feasible constraint set can always be achieved since $\hat{w}_{k-N-1|k}$ is unconstrained at each sample time. Due to these constraints, this formulation is a nonlinear estimator for the linear system that requires the solution of the quadratic program consisting of Eqs. 6.30 through 6.33. Muske *et al.* [47] construct this constrained moving horizon observer quadratic program. Meadows *et al.* [43] present an example of moving horizon constrained state estimation in a linear model predictive control application.

## 6.4 Nonlinear Systems

Our presentation of nonlinear state estimation parallels the discussion for linear systems in the previous section. We begin with a definition of observability for nonlinear systems. Nonlinear recursive filters and observers analogous to the linear estimators follow. We then present nonlinear moving horizon state estimation. Combined state and parameter estimation, a topic not addressed in our presentation of linear systems, followed by a summary of the nonlinear state estimation techniques closes our presentation.

Nonlinearity introduces additional complexity to the state estimation problem. This complexity applies to the derivation as well as the implementation of the nonlinear state estimator. Since the optimal estimator generally is not available for the nonlinear problem, nonlinear techniques are based on sub-optimal approaches. These approaches lead to estimators that are quite different from each other. There are also restrictions on the nonlinear systems that each approach can adequately address.

### 6.4.1 Nonlinear System Models

Continuous-time nonlinear systems are described by the following nonlinear system of differential equations in which $y(t) \in \Re^p$ is the output, $u(t) \in \Re^m$ is the input, $x(t) \in \Re^n$ is the state of the system, $\mathbf{f} : \Re^n \times \Re^m \times \Re^+ \rightarrow \Re^n$ is the system function, and $\mathbf{g} : \Re^n \times \Re^+ \rightarrow \Re^m$ is the measurement function.

$$\begin{aligned} \dot{x}(t) &= \mathbf{f}(x, u, t) \\ y(t) &= \mathbf{g}(x, t) \end{aligned} \tag{6.34}$$

We assume that the model parameters are known and do not include them in the argument list for the system and measurement functions. We also restrict the input to bounded, integrable functions. Discrete-time nonlinear systems are described by the following difference equation in which the discrete system function is $\bar{\mathbf{f}} : \Re^n \times \Re^m \times \mathcal{Z}^+ \rightarrow \Re^n$ and the discrete measurement function is $\bar{\mathbf{g}} : \Re^n \times \mathcal{Z}^+ \rightarrow \Re^m$.

$$\begin{aligned} x_{k+1} &= \bar{\mathbf{f}}(x_k, u_k, k) \\ y_k &= \bar{\mathbf{g}}(x_k, k) \end{aligned} \tag{6.35}$$

In general, there exists no functional representation for the discrete sampling of a continuous-time nonlinear system such as Eq. 6.4 for linear systems. In this case, the continuous-time nonlinear model in Eq. 6.34 with

a discrete sampling period is used to model the system as a discrete process. The output at each sample time is determined from the solution of the nonlinear system of differential equations.

$$\begin{aligned} \dot{x}(\tau) &= \mathbf{f}(x(\tau), u(\tau), \tau), \qquad x(k\Delta t) = x_k \\ x_{k+1} &= x\left((k+1)\Delta t\right) \\ y_k &= \mathbf{g}(x_k, k\Delta t) \end{aligned} \tag{6.36}$$

When the input remains constant over the sampling period, $u(\tau)$ can be represented as a discrete input in which $k$ is the discrete sampling time.

$$u(\tau) = u_k, \qquad k\Delta t \leq \tau < (k+1)\Delta t \tag{6.37}$$

Although we assume a uniform sampling period in the following discussion, we are not restricted to this assumption.

### 6.4.2 Observability of Nonlinear Systems

Observability of a linear system is a global property that can be determined either from the rank of the observability matrix for time-invariant systems or from the rank of the observability Gramian matrix for time-varying systems. In contrast, observability of a nonlinear system is determined locally about a given state or equilibrium point. In this section, we present a brief description of the requirements for local observability of nonlinear systems. The reader is referred to Hermann and Krener [22], Sontag [57], and Krener and Respondek [33] for a more detailed discussion on continuous-time systems. A further discussion on discrete systems is contained in Nijmeijer [49], Keerthi and Gilbert [28], and Lee and Nam [35].

#### Continuous-time Systems

Let $y(t, t_0, x_0, u(t))$ denote the output trajectory from an initial state $x_0$, initial time $t_0$, and input $u(t)$ for the continuous-time nonlinear system in Eq. 6.34. Two states $x_0$ and $z_0$ are defined to be *indistinguishable* if $y(t, t_0, x_0, u(t)) = y(t, t_0, z_0, u(t))$ for $t_0 < t < T$, $T < \infty$, and all admissible input trajectories $u(t)$. The nonlinear system is observable at $x_0$ if the set of states indistinguishable from $x_0$ contains only $x_0$ [22]. A nonlinear system is observable if it is observable at $x$ for all $x \in \Re^n$. A weaker form of observability is obtained if we require that it only be necessary to distinguish

$x_0$ from its neighbors. In this case, a nonlinear system is *weakly observable* at $x_0$ if there exists a neighborhood $\mathcal{S}^n$ of $x_0$ such that the set of states in $\mathcal{S}^n$ that are indistinguishable from $x_0$ contains only $x_0$ [22]. A nonlinear system is weakly observable if it is weakly observable at $x$ for all $x \in \Re^n$. These are global concepts of observability in which it might be necessary to compute a trajectory that is very far from $x_0$ or in which $T$ is very large. By also restricting the state trajectories to a neighborhood of $x_0$, *local observability* can be defined. A nonlinear system is *locally weakly observable* at $x_0$ if for every neighborhood $\mathcal{X}^n$ of $x_0$ there exists another neighborhood $\mathcal{S}^n$ contained in $\mathcal{X}^n$ such that the set of states in $\mathcal{S}^n$ that are indistinguishable from $x_0$ for all input trajectories in which $x(t, t_0, z_0, u(t)) \in \mathcal{X}^n$ and $z_0 \in \mathcal{S}^n$ contains only $x_0$ [57]. This definition is slightly different from that presented in [22]. A nonlinear system is locally weakly observable if it is locally weakly observable at $x$ for all $x \in \Re^n$.

Local weak observability implies that states close to $x_0$ are distinguishable from $x_0$ without large excursions. The advantage to this concept is that a linearized algebraic test similar to that for linear systems can be developed. The nonlinear system in Eq. 6.34 is observable at $x_0$ if there exists a neighborhood of $x_0$ and a $p$-tuple of integers $(k_1, k_2, \ldots, k_p)$, referred to as *observability indices*, such that

i) $k_1 \geq k_2 \geq \ldots \geq k_p \geq 0, \quad \sum_{i=1}^{p} k_i = n$

ii) The $n$ row vectors of $\{L_f^{j-1}(dg_i) \; : \; i = 1, \ldots, p; \; j = 1, \ldots, k_i\}$ are linearly independent.

in which $L_f^j(dg_i)$ is the $j$th Lie derivative of the gradient of $\mathbf{g}_i$ by the vector field $\mathbf{f}$ [33]. An observability matrix can be constructed using the row vectors $L_f^{j-1}(dg_i)$ in which full rank of this matrix implies local weak observability.

$$\mathcal{O} \; = \; \left[L_f^{j-1}(dg_i)\right], \qquad i = 1, \ldots, p; \;\; j = 1, \ldots, k_i \tag{6.38}$$

Note that when $\mathbf{f}$ and $\mathbf{g}$ are linear functions, this matrix is equivalent to the observability matrix for continuous-time linear systems. If the continuous-time system in Eq. 6.34 is observable, the continuous-time system with discrete measurements in Eq. 6.36 is also observable for sufficiently small sampling periods [57]. Certain unobservable systems can be decomposed into local observable and unobservable subspaces by a state transformation [62].

A simpler observability test that is commonly applied to nonlinear systems can be constructed by linearizing the system about $x_0$ using the first

order terms of the following Taylor series expansion [52].

$$\begin{aligned} \dot{x} &= \mathbf{f}(x_0, u, t) + \left.\frac{\partial \mathbf{f}}{\partial x}\right|_{x=x_0} (x - x_0) + \ldots \\ y &= \mathbf{g}(x_0, t) + \left.\frac{\partial \mathbf{g}}{\partial x}\right|_{x=x_0} (x - x_0) + \ldots \end{aligned} \tag{6.39}$$

Observability is determined using the linear system techniques discussed in Section 6.2.2 with the linearized system matrices $A(t) = \partial \mathbf{f}/\partial x$ and $C(t) = \partial \mathbf{g}/\partial x$. Note that this is an approximation to the nonlinear observability matrix presented in Eq. 6.38. Observability of this linearized system may not imply local weak observability of the original nonlinear system.

### Discrete-time Systems

For the discrete nonlinear system in Eq. 6.35, consider any admissible input sequence $\{u_0, u_1, \ldots, u_{k-1}\}$ and an initial state $x_0$. Let $x_j$, $j = 1, \ldots, k$, denote the state sequence generated by the model, $x_j = \bar{\mathbf{f}}(x_{j-1}, u_{j-1}, j-1)$, and $y_j(x_0)$ denote the output trajectory corresponding to $x_0$, $y_j(x_0) = \bar{\mathbf{g}}(x_j, j)$. Two states $x_0$ and $z_0$ are indistinguishable if $y_j(x_0) = y_j(z_0)$ for all $j$ and all admissible input sequences. A nonlinear system is observable at $x_0$ if the set of states indistinguishable from $x_0$ contains only $x_0$ [49]. This definition of observability requires an infinite test and, therefore, is not useful in practice. A finite time test can be constructed in which the discrete nonlinear system is *finite time observable* or *strongly observable* at $x_0$ if for all $z_0 \in \Re^n$ and any admissible input sequence $\{u_0, u_1, \ldots u_{n-1}\}$, $y_j(z_0) = y_j(x_0)$, $j = 1, \ldots, n-1$ implies $z_0 = x_0$. The system is strongly observable if it is strongly observable at $x$ for all $x \in \Re^n$. A local concept of observability for discrete systems can also be defined if we require that it only be necessary to distinguish $x_0$ from its neighbors. In this case, the discrete nonlinear system is *strongly locally observable* at $x_0$ if there exists a neighborhood $\mathcal{S}^n$ of $x_0$ such that for all $z_0 \in \mathcal{S}^n$ and all admissible input sequences, $y_j(z_0) = y_j(x_0)$, $j = 1, \ldots, n-1$ implies $z_0 = x_0$ [49]. The system is strongly locally observable if it is strongly locally observable at $x$ for all $x \in \Re^n$.

A stronger observability condition than finite time observability is *uniform observability.* The discrete nonlinear system in Eq. 6.35 is uniformly observable at $x_0$ if for all $z_0 \in \Re^n$ and all admissible input sequences, there

exists an $N \in [n-1, \infty)$ and a function $\alpha : \Re^+ \to \Re^+$ such that

$$\sum_{j=0}^{N} \|y_j(x_0) - y_j(z_0)\| \geq \alpha(\|x_0 - z_0\|) \tag{6.40}$$

in which $\alpha$ is a continuous, increasing function with $\alpha(0) = 0$ [46]. The system is uniformly observable if it is uniformly observable at $x$ for all $x \in \Re^n$. This condition is similar to the uniform observability property in [28] and is a weaker discrete-time version of that presented in [44].

A linearized algebraic test for strong local observability of discrete nonlinear systems is developed as follows. The discrete nonlinear system in Eq. 6.35 is observable at $x_0$ if there exists a neighborhood of $x_0$ and a $p$-tuple of integers $k_1 \geq k_2 \geq \ldots \geq k_p \geq 0$, $\sum_{i=1}^{p} k_i = n$ such that the observability matrix $\bar{\mathcal{O}}$ for the discrete nonlinear system is full rank in which $\bar{\mathbf{F}}^i$ is the $i$th composite of the system function $\bar{\mathbf{f}}$ [35].

$$\bar{\mathcal{O}} = \begin{bmatrix} \bar{\mathcal{O}}_1 \\ \vdots \\ \bar{\mathcal{O}}_p \end{bmatrix}, \qquad \bar{\mathcal{O}}_i = \frac{\partial}{\partial x} \begin{bmatrix} \bar{\mathbf{g}}_i(x_0) \\ \bar{\mathbf{g}}_i(\bar{\mathbf{F}}^1(x_0)) \\ \vdots \\ \bar{\mathbf{g}}_i(\bar{\mathbf{F}}^{k_i-1}(x_0)) \end{bmatrix} \tag{6.41}$$

When $\bar{\mathbf{f}}$ and $\bar{\mathbf{g}}$ are linear functions, this matrix is equivalent to the discrete linear system observability matrix in Eq. 6.7.

### Summary of Nonlinear Observability

Observability of nonlinear systems is determined by a linearized test that is only guaranteed within a neighborhood of $x_0$. This neighborhood may be arbitrarily small or may extend over the entire operating range of interest. Ray [52] states that observability of nonlinear systems is often determined by the structure of the system and is not dependent on the state in a complex manner. Consequently, linearized observability tests are usually adequate. An additional complexity with nonlinear systems is that the observability matrix can be a function of the input. In these cases, observability must be verified for all admissible inputs since there may exist input values for which the nonlinear system looses observability. For a number of practical cases, however, the system is observable independent of the inputs. These systems are referred to as *observable for any input.* Gauthier *et al.* [20] present a characterization of these systems using single input–single output models.

*Example 6.2*

Consider the batch reactor model presented in Example 6.1. The nonlinear observability matrix in Eq. 6.38 is constructed as follows.

$$\mathcal{O} = \begin{bmatrix} 0 & 1 \\ -2\frac{\Delta H}{\rho C} k C_{\mathrm{A}} & -\frac{E}{RT^2}\frac{\Delta H}{\rho C} k {C_{\mathrm{A}}}^2 - \frac{UA}{V\rho C} \end{bmatrix}$$

When $C_{\mathrm{A}} \neq 0$, the nonlinear observability matrix is full rank and the system is observable. In this example, the system is observable for any input since the observability matrix is not a function of the input. Note that the same observability matrix is obtained from the linearized model using Eq. 6.39.

## 6.5 Nonlinear Recursive Filters

The optimal linear filter in Eq. 6.15 produces the minimum variance and maximum likelihood estimate of the state of the linear system in Eq. 6.4. To develop a nonlinear analogy, one must first specify which estimate is desired since they may not be the same for a nonlinear stochastic system. The minimum variance estimate is the conditional mean of the state. The maximum likelihood or Bayesian estimate is the conditional mode or most probable estimate. These estimates are determined from the conditional probability density of the state given the output measurements. The conditional probability density of the state of a linear system with Gaussian noise is Gaussian and completely determined by the mean and covariance. For a nonlinear system, the conditional probability density of the state is not Gaussian even when the state and measurement disturbances, $w(t)$ and $v(t)$, are Gaussian. The determination of the conditional probability requires that Eq. 6.34 be represented as a vector nonlinear stochastic differential equation.

$$\begin{aligned} dx(t) &= \mathbf{f}(x,u,t)dt + d\beta \\ \frac{d\beta}{dt} &= w(t) \end{aligned}$$

The evolution of the conditional probability is determined by the use of stochastic calculus and, in general, requires some approximations in order to obtain an implementable filter.

These techniques are beyond the scope and intent of the discussion here. We refer the interested reader to the following references for more information. A Bayesian approach to stochastic state estimation for both linear and

nonlinear systems is discussed in Ho and Lee [23]. A detailed discussion of the probabilistic approach to nonlinear state estimation is presented in Jazwinski [25]. An overview is provided in Bryson and Ho [8]. A brief review of the stochastic calculus concepts necessary to understand these techniques is provided in Marcus [42].

### 6.5.1 Extended Kalman Filter

A straight-forward approximation to optimal nonlinear state estimation is to linearize the nonlinear model about a given operating point and apply optimal linear state estimation to the linearized system. The extended Kalman filter computes a state estimate at each sampling time by the use of Kalman filtering on a linearized model of the nonlinear system. This technique is justified if there exists a sufficiently large neighborhood in which the linearized model is a good representation of the nonlinear system. If, in addition, the disturbances are well represented by zero mean Gaussian state and measurement noise, the optimal estimate for the linearized system should be a reasonable approximation to the optimal estimate for the nonlinear system. In this case, extended Kalman filtering is expected to provide an accurate estimate of the state of the nonlinear system.

A linearized model of the continuous-time system in Eq. 6.34 about the state $x^*$ can be developed from the Taylor series expansion in Eq. 6.39. Depending on the selection of $x^*$, several variations of the extended Kalman filter can be developed. The most common approach is the first order filter in which the nonlinear system is linearized about the current state estimate at each sampling time using the first order terms in Eq. 6.39. More complex approaches that attempt to compensate for the inaccuracy caused by linearization, such as iterative and higher order filters, have also been implemented. These extended Kalman filter techniques are presented in the following sections. Further discussion of extended Kalman filtering is contained in Jazwinski [25], Gelb [21], Bryson and Ho [8], Lewis [36], and Stengel [58].

#### First Order Extended Kalman Filter

For the continuous-time nonlinear system with discrete output measurements in Eq. 6.36, a linearized approximation to the nonlinear model can be obtained by truncating Eq. 6.39 after the first order terms. This approximate model, linearized about the current state, is used to construct a

time-varying Kalman filter at each sampling time. In order to ensure that the partial derivatives exist, the system and measurement functions are restricted to continuous functions in $x$ such that $\mathbf{f}, \mathbf{g} \in \mathcal{C}^1$.

The filtered state estimate is determined from the current output measurement in the same manner as the linear recursive filter in Eq. 6.15.

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + L_k \left( y_k - \mathbf{g}(\hat{x}_{k|k-1}, k\Delta t) \right) \\ \hat{x}_{0|0} &= \bar{x}_0 \end{aligned} \tag{6.42}$$

The Kalman filter gain, $L_k$, is computed using the discrete, time-varying Kalman filter formulation in Eq. 6.21 in which the linear system matrix $\bar{C}_k$ is replaced by the linearized measurement function $G_k$.

$$\begin{aligned} L_k &= \hat{P}_{k|k-1} G_k^T \left( G_k \hat{P}_{k|k-1} G_k^T + R \right)^{-1} \\ G_k &= \left. \frac{\partial \mathbf{g}(x,t)}{\partial x} \right|_{x=\hat{x}_{k|k-1},\, t=k\Delta t} \end{aligned} \tag{6.43}$$

The estimated covariance of the state is updated at each sample time due to the contribution of the discrete measurement as follows.

$$\begin{aligned} \hat{P}_{k|k} &= \left( I - L_k G_k \right) \hat{P}_{k|k-1} \\ \hat{P}_{0|0} &= Q_0 \end{aligned} \tag{6.44}$$

Between sampling times, the state estimate is propagated using the nonlinear system model in Eq. 6.36.

$$\begin{aligned} \dot{\hat{x}}\left(\tau \mid k\right) &= \mathbf{f}(\hat{x}\left(\tau \mid k\right), u(\tau), \tau), \qquad \hat{x}\left(k\Delta t \mid k\right) = \hat{x}_{k|k} \\ \hat{x}_{k+1|k} &= \hat{x}\left((k+1)\Delta t \mid k\right) \end{aligned} \tag{6.45}$$

The covariance of the state estimate in Eq. 6.45 is propagated between sampling times using the linear differential equation in Eq. 6.26 in which the linear system matrix $A(t)$ is replaced by the linearized system function $F\left(\tau \mid k\right)$. Since there are no output measurements available between sampling times, the contribution to the covariance from the output measurement is removed.

$$\begin{aligned} \dot{\hat{P}}\left(\tau \mid k\right) &= F\left(\tau \mid k\right) \hat{P}\left(\tau \mid k\right) + \hat{P}\left(\tau \mid k\right) F\left(\tau \mid k\right)^T, \quad \hat{P}\left(k\Delta t \mid k\right) = \hat{P}_{k|k} \\ \hat{P}_{k+1|k} &= \hat{P}\left((k+1)\Delta t \mid k\right) + Q \\ F\left(\tau \mid k\right) &= \left. \frac{\partial \mathbf{f}(x,u,t)}{\partial x} \right|_{x=\hat{x}(\tau|k),\, u=u(k\Delta t),\, t=\tau} \end{aligned} \tag{6.46}$$

This continuous-time linear approximation of the estimated state covariance is used in Eq. 6.42 to compute the Kalman filter gain. A discrete linear approximation to the covariance can also be used in the Kalman filter gain calculation. It is obtained from the time-varying filtering Riccati equation in Eq. 6.21 with the contribution from the output measurement removed.

$$\hat{P}_{k+1|k} = \Phi\left((k+1)\Delta t, k\Delta t\right)\hat{P}_{k|k}\Phi^T\left((k+1)\Delta t, k\Delta t\right) + Q \tag{6.47}$$

$$\frac{d\Phi\left(t, k\Delta t\right)}{dt} = F\left(t \mid k\right)\Phi\left(t, k\Delta t\right), \qquad \Phi\left(k\Delta t, k\Delta t\right) = I \tag{6.48}$$

In this expression, $\Phi\left((k+1)\Delta t, k\Delta t\right)$ is the transition matrix of the time-varying linear system obtained by linearization of the nonlinear system. This discrete covariance approximation does not require the solution of a differential equation to determine the covariance, but the state transition matrix must be computed. A further approximation can be made by linearizing about the state $\hat{x}_{k|k}$ and assuming a time-invariant linear system for the transition matrix calculation in Eq. 6.48. Wishner *et al.* [64] compared this method of discrete covariance propagation to the continuous-time method and found agreement within 0.1 percent for the example presented in the reference.

The extended Kalman filter for the discrete nonlinear system in Eq. 6.35 is similar to that presented for continuous-time systems with discrete measurements.

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k\left(y_k - \bar{\mathbf{g}}(\hat{x}_{k|k-1}, k)\right) \tag{6.49}$$

$$L_k = \hat{P}_{k|k-1}\bar{G}_k^T\left(\bar{G}_k\hat{P}_{k|k-1}\bar{G}_k^T + R\right)^{-1} \tag{6.50}$$

$$\hat{P}_{k|k} = \left(I - L_k\bar{G}_k\right)\hat{P}_{k|k-1} \tag{6.51}$$

$$\hat{x}_{k+1|k} = \bar{\mathbf{f}}(\hat{x}_{k|k}, u_k, k) \tag{6.52}$$

$$\hat{P}_{k+1|k} = \bar{F}_k\hat{P}_{k|k}\bar{F}_k^T + Q \tag{6.53}$$

The discrete linear system matrices are formed from the partial derivatives of the discrete nonlinear measurement and system functions.

$$\bar{G}_k = \left.\frac{\partial\bar{\mathbf{g}}(x,k)}{\partial x}\right|_{x=\hat{x}_{k|k-1}} \tag{6.54}$$

$$\bar{F}_k = \left.\frac{\partial\bar{\mathbf{f}}(x,u,k)}{\partial x}\right|_{x=\hat{x}_{k|k},\, u=u_k} \tag{6.55}$$

The functions $\bar{\mathbf{f}}$ and $\bar{\mathbf{g}}$ are also restricted to continuous functions in $x$ so that the partial derivatives in Eqs. 6.54 and 6.55 exist.

A first order extended Kalman filter was implemented on an experimental batch reactor to estimate heat generation by DeVallière and Bonvin [16] and on an experimental batch styrene reactor to estimate conversion and chain length distribution by Schuler and Suzhen [55]. Kozub and MacGregor [31] implemented an extended Kalman filter on an semi-batch polymerization reactor to estimate monomer concentration and a reactive impurity disturbance. Dimitratos *et al.* [19] estimated the monomer concentration in a copolymerization reactor. Myers and Luecke [48] discuss an efficient numerical integration technique for the system and covariance equations and used this technique to estimate the specific growth rate in a fed-batch fermenter.

### Linearized Extended Kalman Filter

The linearized extended Kalman filter determines the Kalman filter gains from a nominal state and input trajectory [21]. The nominal state, $\tilde{x}(t)$, and input, $\tilde{u}(t)$, are specified *a priori* based on an assumed behavior of the system. The Kalman filter gain for continuous-time systems with discrete measurements is then computed as shown in Eqs. 6.43 through 6.47 in which the measurement and system functions are linearized about the nominal state and input trajectories.

$$G_k = \left.\frac{\partial \mathbf{g}(x,t)}{\partial x}\right|_{x=\tilde{x}(k\Delta t),\, t=k\Delta t} \tag{6.56}$$

$$F(\tau \mid k) = \left.\frac{\partial \mathbf{f}(x,u,t)}{\partial x}\right|_{x=\tilde{x}(\tau),\, u=\tilde{u}(k\Delta t),\, t=\tau} \tag{6.57}$$

The Kalman filter gain for discrete systems is computed as shown in Eqs. 6.50 through 6.53 in which the discrete measurement and system functions are linearized about the nominal trajectories.

$$\bar{G}_k = \left.\frac{\partial \bar{\mathbf{g}}(x,k)}{\partial x}\right|_{x=\tilde{x}_k} \tag{6.58}$$

$$\bar{F}_k = \left.\frac{\partial \bar{\mathbf{f}}(x,u,k)}{\partial x}\right|_{x=\tilde{x}_k,\, u=\tilde{u}_k} \tag{6.59}$$

The advantage to this approach is that the filter gains can be computed off-line, resulting in a considerable reduction in the real-time computational

requirements of the filter. The disadvantage is the lack of feedback from the process in the determination of the filter gain. Since the estimated state is normally a better representation of the true state than the nominal state, the linearized filter is less accurate. Systems that have large process disturbances or are controlled between multiple operating points generally are not well suited to this approach. This filter is applicable for systems with a well-defined state and input trajectory. It is normally used when the improvement in the performance of the first order filter does not justify the real-time computations necessary to implement it.

Because of the approximations made in the propagation of the covariance matrix in extended Kalman filtering, this covariance matrix can become a poor estimate in some applications. In these cases, updating the filter gain at each sampling period may result in computational effort that does not improve the performance of the filter. Safonov and Athans [54] present a linearized extended Kalman filter formulation that uses a constant filter gain and present sufficient conditions for convergence of this filter. This approach consists of a steady-state Kalman filter applied to the nonlinear process. The nominal state trajectory is composed of a single operating point. The robustness of the Kalman filter to model mismatch is sufficient to obtain acceptable performance with this technique in some applications.

### Iterative Extended Kalman Filter

In the first order extended Kalman filter approaches already discussed, the nonlinear system equations are linearized about some estimate of the state. If this estimate is not close to the true state, the linearized model may be a poor approximation to the nonlinear system. Iterative schemes have been developed in which a linearized model is determined from an updated state estimate at each iteration. These schemes attempt to reduce the estimation error by improving the approximation to the nonlinear system that is used in the determination of the filtered state.

One iterative scheme that can be employed is to repeat the calculation of $\hat{x}_{k|k}$ in Eq. 6.42 [25]. This method is useful when compensating for nonlinearity in the measurement function. Letting $\hat{x}_{k|k,i}$ represent the $i$th iterate of the filtered state, the next iteration is determined as follows.

$$\begin{aligned} \hat{x}_{k|k,i+1} &= \hat{x}_{k|k-1} + L_{k,i}\left(y_k - \mathbf{g}(\hat{x}_{k|k,i}, k\Delta t) - G_{k,i}(\hat{x}_{k|k-1} - \hat{x}_{k|k,i})\right) \quad (6.60) \\ \hat{x}_{k|k,0} &= \hat{x}_{k|k-1} \end{aligned}$$

The filter gain and linearized measurement function are recomputed based on the current iterate of the filtered state.

$$L_{k,i} = \hat{P}_{k|k-1} G_{k,i}^T \left( G_{k,i} \hat{P}_{k|k-1} G_{k,i}^T + R \right)^{-1} \tag{6.61}$$

$$G_{k,i} = \left. \frac{\partial \mathbf{g}(x,t)}{\partial x} \right|_{x=\hat{x}_{k|k,i},\, t=k\Delta t} \tag{6.62}$$

Eqs. 6.60 through 6.62 are repeated until there is no significant difference between the iterated filtered states. The estimated state covariance is then updated in the same manner as Eq. 6.44 using the converged gain and linearized measurement function.

$$\hat{P}_{k|k} = \left( I - L_{k,i} G_{k,i} \right) \hat{P}_{k|k-1} \tag{6.63}$$

The state estimate and covariance are propagated between sampling times as shown in Eqs. 6.45 and 6.46 or Eq. 6.47 for continuous-time systems. For discrete systems, Eqs. 6.52 and 6.53 are used. Note that a single iteration results in the first order extended Kalman filter.

Nonlinearity in the system function is not addressed by the preceding method. A second iterative method that takes system function nonlinearity into account can be developed by also updating the estimated state $\hat{x}_{k|k-1}$ in the iteration scheme [64]. The filtered state is iterated in the same manner as the first method in which the filter gain, $L_{k,i}$, and linearized measurement function, $G_{k,i}$, are computed as shown in Eqs. 6.61 and 6.62.

$$\hat{x}_{k|k,i+1} = \hat{z}_{k|k-1,i} + L_{k,i} \left( y_k - \mathbf{g}(\hat{x}_{k|k,i}, k\Delta t) - G_{k,i}(\hat{z}_{k|k-1,i} - \hat{x}_{k|k,i}) \right) \tag{6.64}$$

$$\hat{x}_{k|k,0} = \hat{x}_{k|k-1}$$

The current value of the estimated state, $\hat{z}_{k|k-1,i}$, used in the filtered state calculation is computed using the following procedure.

$$\hat{z}_{k|k-1,i} = \hat{z}_{k|k,i} + \Phi\left(k\Delta t, (k-1)\Delta t, i\right) \left( \hat{x}_{k-1|k-1} - \hat{z}_{k-1|k,i} \right) \tag{6.65}$$

$$\hat{z}_{k|k-1,0} = \hat{x}_{k|k-1}$$

The estimated state update requires a smoothed estimate at sample time $k-1$ and a filtered state at sample time $k$. The smoothed state estimate, $\hat{z}_{k-1|k,i}$, is computed iteratively as follows.

$$\hat{z}_{k-1|k,i+1} = \hat{x}_{k|k} + S_{k,i} \left( \hat{x}_{k|k,i+1} - \hat{z}_{k|k-1,i} \right) \tag{6.66}$$

$$\hat{z}_{k-1|k,0} = \hat{x}_{k-1|k-1}$$

See Jazwinski [25] or Gelb [21] for a review of the nonlinear smoothing techniques used to generate this expression. The filtered state estimate $\hat{z}_{k|k,i}$ used in the estimated state calculation in Eq. 6.65 is determined by propagating the current smoothed estimate forward one sample period using the system equation.

$$\begin{aligned} \dot{\hat{z}}\left(\tau \mid k,i\right) &= \mathbf{f}(\hat{z}\left(\tau \mid k,i\right), u(\tau), \tau), \qquad \hat{z}\left((k-1)\Delta t \mid k,i\right) = \hat{z}_{k-1|k,i} \quad (6.67) \\ \hat{z}_{k|k,i} &= \hat{z}\left(k\Delta t \mid k,i\right), \qquad \hat{z}_{k|k,0} = \hat{x}_{k|k-1} \end{aligned}$$

The transition matrix used in the estimated state update in Eq. 6.65, $\Phi\left(k\Delta t,(k-1)\Delta t,i\right)$, is determined by linearizing about the current smoothed state estimate $\hat{z}_{k-1|k,i}$.

$$\begin{aligned} \frac{d\Phi\left(t,(k-1)\Delta t,i\right)}{dt} &= F\left(t \mid k,i\right)\Phi\left(t,(k-1)\Delta t,i\right), \qquad (6.68) \\ \Phi\left((k-1)\Delta t,(k-1)\Delta t,i\right) &= I \\ F\left(\tau \mid k,i\right) &= \left.\frac{\partial \mathbf{f}(x,u,t)}{\partial x}\right|_{x=\hat{z}_{k-1|k,i},\, u=u((k-1)\Delta t),\, t=\tau} \end{aligned}$$

The gain matrix used in the smoothed state iteration, $S_{k,i}$, is computed from this transition matrix as follows.

$$S_{k,i} = \hat{P}_{k-1|k-1}\Phi^T\left(k\Delta t,(k-1)\Delta t,i\right)\hat{P}^{-1}_{k|k-1,i} \qquad (6.69)$$

$$\hat{P}_{k|k-1,i} = \Phi\left(k\Delta t,(k-1)\Delta t,i\right)\hat{P}_{k-1|k-1}\Phi^T(k\Delta t,(k-1)\Delta t,i) + Q \qquad (6.70)$$

In this iterative method, Eqs. 6.64 through 6.70 are repeated until there is no significant difference between the iterated filtered states. The estimated state covariance is then updated as shown in Eq. 6.63. The state estimate and covariance are propagated between sampling times as in the previous iterative method. Wishner *et al.* [64] state that 95 percent of filter performance is achieved in two iterations for the example presented in the reference. A single iteration of this method results in the first order extended Kalman filter.

This iterative method uses a smoothed estimate from the previous sample time to determine the iterated state estimates. A *globally iterated* extended Kalman filter can be obtained by determining a smoothed estimate of the initial state, $\hat{z}_{0|k}$, and then reprocessing all of the output measurements starting from this smoothed initial state estimate [25]. Since the *a*

*priori* estimate of the initial state is generally no better than an educated guess, this procedure can significantly improve the performance of the filter. However, the increased computational burden may become substantial. This technique was implemented on a semi-batch polymerization reactor by Kozub and MacGregor [31]. It was found to improve the convergence of the estimated states to the true state by providing a better estimate of the initial state of the reactor.

These iterative approaches are used to compensate for the errors introduced by linearizing a nonlinear system. However, multiple iterations of the filtered state calculation require additional computation time which must be taken into account when using these methods. In addition, there is no guarantee that the iterations will converge since these methods are approximations generated by successive linearization.

### Second Order Extended Kalman Filter

If we restrict the system and measurement functions to twice differentiable functions in $x$ such that $\mathbf{f}, \mathbf{g} \in \mathcal{C}^2$, the second order Taylor series expansion for the continuous-time nonlinear system in Eq. 6.34 is as follows.

$$\begin{aligned} \dot{x} &= \mathbf{f}(x^*, u, t) + F(x - x^*) + \frac{1}{2}\sum_{j=1}^{n} e^j (x - x^*)^T \mathcal{F}^j (x - x^*) \quad (6.71) \\ y &= \mathbf{g}(x^*, t) + G(x - x^*) + \frac{1}{2}\sum_{j=1}^{m} e^j (x - x^*)^T \mathcal{G}^j (x - x^*) \\ e^j &= [0, \ldots, 0, \underbrace{1}_{j\text{th location}}, 0, \ldots, 0]^T \end{aligned}$$

In this expansion, $\mathcal{F}^j$ and $\mathcal{G}^j$ are the second order Hessian matrices of the $j$th component of the system and measurement functions and $e^j$ are the standard unit basis vectors. The use of this expansion in the approximate filter design results in the second order extended Kalman filter [3].

The filtered state is determined using the discrete output measurements in the same manner as the first order filter in Eq. 6.42 with the addition of a bias correction term $\Pi$.

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + L_k \left( y_k - \mathbf{g}(\hat{x}_{k|k-1}, k\Delta t) - \Pi_k \right) \quad (6.72)$$

The bias correction term is calculated as follows.

$$\Pi_k = \frac{1}{2}\sum_{j=1}^{m} e^j \operatorname{tr}\left(\mathcal{G}_k^j \hat{P}_{k|k-1}\right) \tag{6.73}$$

The Kalman filter gain, $L_k$, is computed as shown in Eq. 6.43 for the first order filter with an additional correction term $\Gamma$.

$$L_k = \hat{P}_{k|k-1} G_k^T \left(G_k \hat{P}_{k|k-1} G_k^T + R + \Gamma_k\right)^{-1} \tag{6.74}$$

$$\Gamma_k = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m} e^i e^{jT} \operatorname{tr}\left(\mathcal{G}_k^i \hat{P}_{k|k-1} \mathcal{G}_k^j \hat{P}_{k|k-1}\right) \tag{6.75}$$

$$G_k = \left.\frac{\partial \mathbf{g}(x,t)}{\partial x}\right|_{x=\hat{x}_{k|k-1},\, t=k\Delta t} \qquad \mathcal{G}_k^j = \left.\frac{\partial^2 \mathbf{g}^j(x,t)}{\partial x^2}\right|_{x=\hat{x}_{k|k-1},\, t=k\Delta t}$$

The covariance of the filtered state estimate, $\hat{P}_{k|k}$, is calculated from $\hat{P}_{k|k-1}$ in the same manner as the first order filter using Eq. 6.44.

Due to the second order terms, the state and covariance are coupled between sampling times by a second bias correction term, $\mu$, and are propagated using the following nonlinear system of differential equations.

$$\dot{\hat{x}}\left(\tau \mid k\right) = \mathbf{f}(\hat{x}\left(\tau \mid k\right), u(\tau), \tau) + \mu\left(\tau \mid k\right), \qquad \hat{x}\left(k\Delta t \mid k\right) = \hat{x}_{k|k} \tag{6.76}$$

$$\dot{\hat{P}}\left(\tau \mid k\right) = F\left(\tau \mid k\right)\hat{P}\left(\tau \mid k\right) + \hat{P}\left(\tau \mid k\right)F\left(\tau \mid k\right)^T, \quad \hat{P}\left(k\Delta t \mid k\right) = \hat{P}_{k|k} \tag{6.77}$$

$$\mu\left(\tau \mid k\right) = \frac{1}{2}\sum_{j=1}^{n} e^j \operatorname{tr}\left(\mathcal{F}^j\left(\tau \mid k\right)\hat{P}\left(\tau \mid k\right)\right) \tag{6.78}$$

$$F\left(\tau \mid k\right) = \left.\frac{\partial \mathbf{f}(x,u,t)}{\partial x}\right|_{x=\hat{x}(\tau|k),\, u=u(k\Delta t),\, t=\tau}$$

$$\mathcal{F}^j\left(\tau \mid k\right) = \left.\frac{\partial^2 \mathbf{f}^j(x,u,t)}{\partial x^2}\right|_{x=\hat{x}(\tau|k),\, u=u(k\Delta t),\, t=\tau}$$

The covariance of the state estimate at the next sample time is determined as follows.

$$\hat{P}_{k+1|k} = \hat{P}\left((k+1)\Delta t \mid k\right) + Q \tag{6.79}$$

If $\Pi_k$, $\Gamma_k$, and $\mu\left(\tau \mid k\right)$ are removed, the first order filter is recovered. When $\mathbf{g}$ is a linear function, the only difference between the second and first

order filters is the term $\mu(\tau \mid k)$ that couples the covariance and estimated state differential equations. Athans *et al.* [3] state that the major improvement over the first order filter is due to this term for the example presented in the reference. A detailed derivation of this filter is presented in [3].

### 6.5.2 Statistical Approximation

In the statistical approximation approach, the system and measurement functions are represented by a truncated polynomial series expansion. The coefficients of these expansions are determined by minimizing the error between the function and the truncated polynomial based on the statistics of the state. Consider a scalar system in which the system function error is represented as follows.

$$e = \mathbf{f}(x,u,t) - \left(a_0(u,t) + a_1(u,t)x + a_2(u,t)x^2 + \ldots + a_j(u,t)x^j\right)$$

This approach seeks coefficients of the polynomial expansion that minimize the mean square error between the system function and polynomial, $E\left[e^2\right]$, in which $E$ is the expectation operator. The coefficients are determined by taking the partial derivative of the mean square error with respect to each coefficient, equating these derivatives to zero, and interchanging the derivative and expectation operations. The result is a linear system of equations in the coefficients containing moments and cross-moments of the state and system function. The conditional probability density of the state is required to compute these moments which are generated by the expectation operation. Since the exact probability density is normally not available, an approximation must be used.

If the polynomial expansion is truncated after the linear term, the approach is referred to as *statistical linearization* [21] or *quasilinearization* [58]. The coefficients for the statistically linearized expansion of the system function, $\tilde{\mathbf{f}}(x,u,t) = f + Fx$, are determined as follows.

$$f = E\left[\mathbf{f}(x,u,t)\right], \qquad F = \frac{E\left[\mathbf{f}(x,u,t)x^T\right] - E\left[\mathbf{f}(x,u,t)\right]E[x]^T}{E\left[(x - E[x])^2\right]}$$

An approximation for the measurement function is developed in the same manner as the system function approximation. These expressions are referred to as *describing functions* for the nonlinear system.

For continuous-time systems with discrete measurements, the filter equations using the statistically linearized filter are similar to those of the extended Kalman filter. The filtered state is determined recursively from the

current output measurement as shown in Eq. 6.42. The filter gain is computed as shown in Eq. 6.43 and the covariance of the filtered state estimate is updated at each sample time as shown in Eq. 6.44 using the statistically linearized measurement function $\tilde{\mathbf{g}}_k$.

$$L_k = \hat{P}_{k|k-1}\tilde{\mathbf{g}}_k^T \left(\tilde{\mathbf{g}}_k \hat{P}_{k|k-1}\tilde{\mathbf{g}}_k^T + R\right)^{-1} \tag{6.80}$$

$$\hat{P}_{k|k} = \left(I - L_k\tilde{\mathbf{g}}_k\right)\hat{P}_{k|k-1} \tag{6.81}$$

Between sampling times, the state estimate is propagated using the nonlinear system function as shown in Eq. 6.45. The covariance of the state estimate is propagated as shown in Eq. 6.46 using the statistically linearized system function $\tilde{\mathbf{f}}_k$.

$$\dot{\hat{P}}\left(\tau \mid k\right) = \tilde{\mathbf{f}}_k \hat{P}\left(\tau \mid k\right) + \hat{P}\left(\tau \mid k\right)\tilde{\mathbf{f}}_k^T, \qquad \hat{P}\left(k\Delta t \mid k\right) = \hat{P}_{k|k} \tag{6.82}$$

$$\hat{P}_{k+1|k} = \hat{P}\left((k+1)\Delta t \mid k\right) + Q$$

The statistically linearized measurement and system functions used in this approach are determined at each sampling time as follows.

$$\tilde{\mathbf{g}}_k = \left(E\left[\mathbf{g}(\hat{x}_{k|k-1}, k\Delta t)\hat{x}_{k|k-1}^T\right] - E\left[\mathbf{g}(\hat{x}_{k|k-1}, k)\right]E\left[\hat{x}_{k|k-1}^T\right]\right)\hat{P}_{k|k-1}^{-1} \tag{6.83}$$

$$\tilde{\mathbf{f}}_k = \left(E\left[\mathbf{f}(\hat{x}_{k|k}, u_k, k\Delta t)\hat{x}_{k|k}^T\right] - E\left[\mathbf{f}(\hat{x}_{k|k}, u_k, k\Delta t)\right]E\left[\hat{x}_{k|k}^T\right]\right)\hat{P}_{k|k}^{-1} \tag{6.84}$$

The expectation operations required to determine these statistically linearized functions are computed using an assumed probability density function $p(x)$ in which $e^j$ and $e^i$ are the standard unit vectors.

$$E\left[f(x,u,t)\right] = \sum_{j=1}^{n} e^j \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^j(x,u,t)p(x)\, dx^1 \ldots dx^n$$

$$E\left[f(x,u,t)x^T\right] = \sum_{j=1}^{n}\sum_{i=1}^{n} e^j e^{i^T} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f^j(x,u,t)x^i p(x)\, dx^1 \ldots dx^n$$

A Gaussian distribution function typically is assumed in this approach. Significant computational effort may be required to perform these integrations. These computational aspects are discussed in Gelb [21] and Stengel [58].

Higher order polynomial expansions of the system and measurement functions can be used in the development of an approximate nonlinear estimator. For example, Mahalanabis and Farooq [41] present a second order

formulation for scalar systems. The higher order expansions are easily generated and can improve the performance of the estimator. However, the corresponding filter equations require the determination of higher order moments and cross-moments of the state and model functions. Determination of these higher order moments can greatly increase the computational requirements of the filter.

### 6.5.3 Summary of Nonlinear Recursive Filters

Extended Kalman filtering is typically the first nonlinear state estimation technique that would be attempted in most situations. It has been found to work well for a number of chemical process control applications and is relatively easy to implement. However, the performance of extended Kalman filtering is directly related to the quality of the linear approximations made in the state and covariance estimates. If the state disturbances are large, the initial state estimate is poor, or the system is highly nonlinear, the linearized equations may not be a good approximation to the nonlinear stochastic process. In these cases, the extended Kalman filter estimates may not be close to the optimal nonlinear estimates and may not converge to the true state. Extended Kalman filtering can produce either biased or divergent estimates under these conditions. It can also be difficult to tune the extended Kalman filter to achieve acceptable performance. Poor choices of the covariance matrices $Q$, $Q_0$, and $R$ can also lead to biased or divergent state estimates. Agarwal and Bonvin [1] discuss this issue in the context of state estimation for a batch reactor. Dimitratos *et al.* [19] discuss adaptive schemes that estimate the process noise covariance along with the state to improve the filter tuning. Kozub and MacGregor [31] discuss the importance of including integrating disturbance states in the model to prevent bias in the state estimates due to unmeasured disturbances and model error.

Deciding which extended Kalman filtering technique to implement is based on a trade-off between performance and complexity. The linearized formulation is the easiest to implement but may not work well for many applications. The second order and iterative approaches have been shown to produce improved estimates [64], but they are more difficult to implement and require more computational time. These approaches should normally not be considered as an initial choice unless the first order filter is known not to perform well for the given application or it is necessary to achieve higher performance. For some applications, none of the extended Kalman filtering approaches may perform well and another technique will be required.

Statistical approximation techniques have generally been found to be more accurate than extended Kalman filtering [21]. Another advantage of these techniques is that derivatives of the system and measurement functions are not required. Therefore, a much broader class of nonlinear functions can be considered since there are no continuity restrictions that were necessary for the extended Kalman filter. However, the computational requirements may be greater due to the calculation of the expectation operations. These requirements will depend on the system and measurement functions and the assumed probability distribution. Although it is convenient, the choice of a Gaussian distribution can be a gross approximation that may result in poor performance for some applications. Gelb [21] presents a detailed example of the implementation of this approach.

*Example 6.3*

We again consider the batch reactor presented in Example 6.1 and compare the performance of the linearized and first order extended Kalman filters in estimating the concentration of component A from the temperature measurements. In this example, perfect temperature measurements are available at a discrete sampling period of 30 seconds and there is no model error. The nominal initial state of the reactor is a component A concentration of 1M and a temperature of 20 °C. The filters are tuned to indicate that disturbances are more likely to occur in the composition due to possible polymerization side reactions not accounted for in the model. The covariance of the initial state is chosen to indicate that the initial temperature estimate is much more reliable than the estimate of the initial composition. The model parameters and filter covariances are listed in Table 6.1.

The steady-state Kalman filter gain used in the linearized filter is determined by linearization of the model equations about the nominal initial state. The continuous-time linear approximation of the estimated state covariance is used in the first order filter. We simulate three different initial temperatures using an estimated initial concentration of 1M in which the actual concentration is 0.9M. Figure 6.1 shows the excellent performance of both filters when the initial batch temperature is at the nominal initial condition of 20 °C. When the initial temperature is lowered to 10 °C, both filters demonstrate slower convergence to the true concentration. However, the degradation in performance of the linearized filter is considerably greater as shown in Figure 6.2. When the initial temperature is increased to 30 °C, there is little effect on the first order filter performance when compared to

| Model Parameters | | Filter Covariances | |
|---|---|---|---|
| $\frac{\Delta H}{\rho C}$ | $-30.0\ \frac{\text{liter}-\ ^\circ\text{C}}{\text{gmole}}$ | $Q$ | $\begin{bmatrix} 10 & 0 \\ 0 & 1 \end{bmatrix}$ |
| $\frac{UA}{V\rho C}$ | $1.00 \times 10^{-3}\ \text{sec}^{-1}$ | | |
| $T_c$ | $20.0\ ^\circ\text{C}$ | $R$ | $1$ |
| $k_0$ | $1.00 \times 10^{8}\ \frac{\text{liter}}{\text{gmole-sec}}$ | | |
| $E_a/R$ | $7.50 \times 10^{3}\ ^\circ\text{C}$ | $Q_0$ | $\begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}$ |

Table 6.1: Batch Reactor Model Parameters and Filter Covariances

20 °C. As shown in Figure 6.3, however, the linearized filter estimates oscillate around the true state and do not converge. This example demonstrates the good performance of the first order filter and the limited applicability of the linearized filter over this 20 °C initial temperature range. It also suggests that an iterated or higher order extended Kalman filter may not be necessary for this application.

## 6.6 Nonlinear Observers

The nonlinear recursive filter techniques presented in the previous section are intended to produce an approximate optimal estimate of the state. Similar to the linear observer discussed in Section 6.3.4, nonlinear observers consider the deterministic dynamic behavior of the reconstruction error. In this section, we briefly discuss observer error linearization and extended Luenberger-type observers. Additional approaches are mentioned in the summary.

### 6.6.1 Observer Error Linearization

Analogous to feedback linearization for nonlinear control presented in Chapter 5, observer error linearization attempts to find a local coordinate transformation such that the corresponding transformed system is linear in the reconstruction error dynamics. An estimate of the state of the transformed system is then obtained using a linear observer. The advantage to this ap-

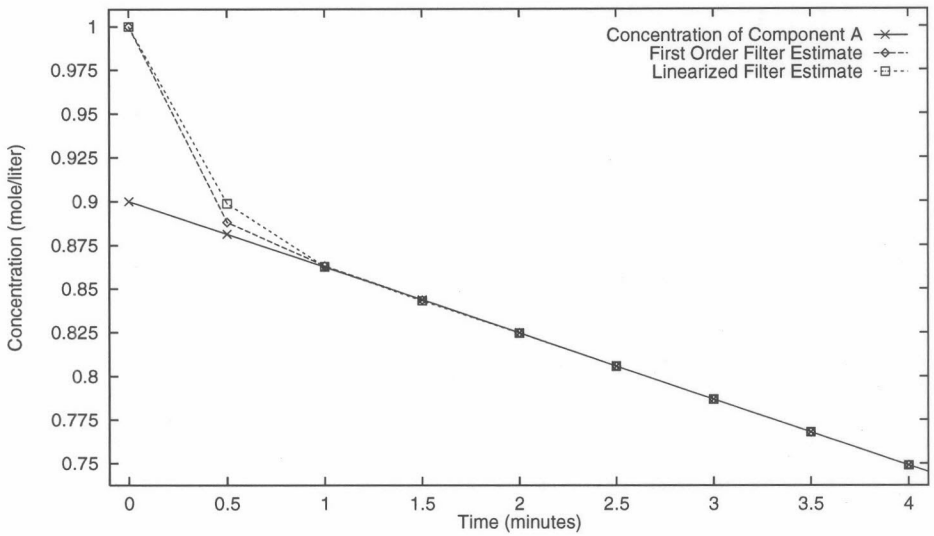


Figure 6.1: First order and linearized extended Kalman filter estimated concentrations at the nominal initial temperature of 20 °C.

proach is that the linearized model is an exact representation of the nonlinear system within a neighborhood of a given nominal state. Therefore, this technique does not introduce the linearization errors present in the recursive filter approaches. The disadvantage is that a coordinate transformation must exist which imposes a restriction on the nonlinear systems that can be considered. This technique was independently introduced using scalar systems by Bestle and Zeitz [6] and Krener and Isidori [32]. It was extended to multivariable systems by Krener and Respondek [33] and Xia and Gao [66].

Consider a nonlinear system described by Eq. 6.34 that is the result of a nonlinear change of coordinates, $x(t) = \mathcal{F}(z(t))$, on the reference system

$$\begin{aligned} \dot{z}(t) &= Az(t) + \gamma(u(t), y(t)) = h(z(t), u(t)) \\ y(t) &= Cz(t) \end{aligned} \tag{6.85}$$

in which the matrices $A$ and $C$ are in *Brunovsky canonical form.* This canonical form results in the following relation for each output $y_i$ in which $k_i$ the corresponding observability index. The observability index is discussed in Section 6.4.2.

$$\dot{z}_2^i = z_1^i, \;\; \dot{z}_3^i = z_2^i, \;\; \ldots, \;\; \dot{z}_{k_i}^i = z_{k_i-1}^i; \;\; y_i = z_{k_i}^i \tag{6.86}$$

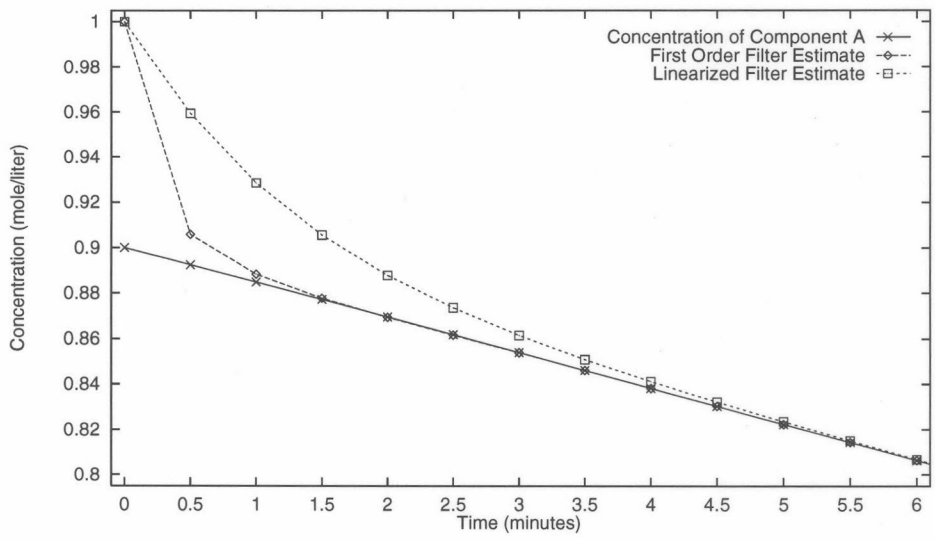


Figure 6.2: First order and linearized extended Kalman filter estimated concentrations at an initial temperature of 10 °C.

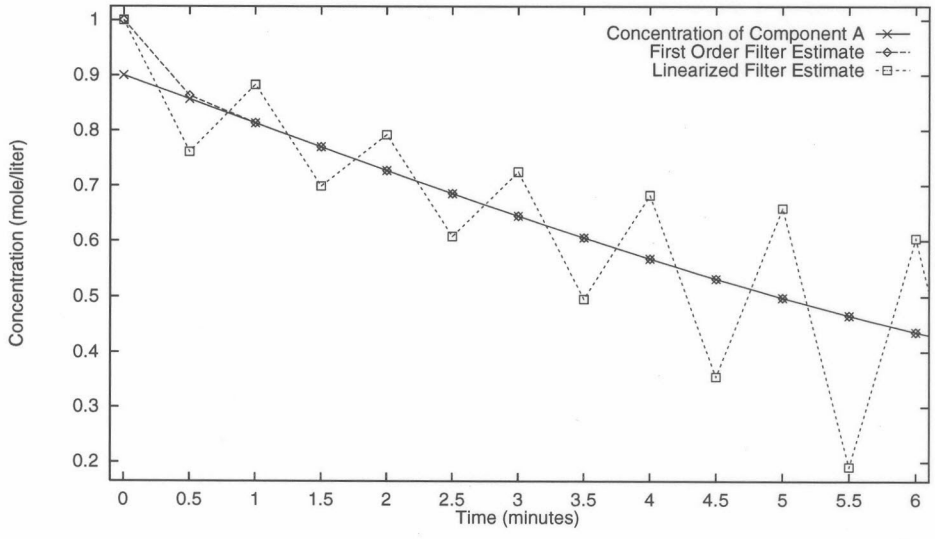


Figure 6.3: First order and linearized extended Kalman filter estimated concentrations at an initial temperature of 30 °C.

The reference system, which is in *observer form*, is linear in the transformed state $z$ with the addition of a nonlinear function of the input and output. An estimate of the state of this system can be obtained using a linear observer as discussed in Section 6.3.4.

$$\dot{\hat{z}}(t) = A\hat{z}(t) + L(y(t) - C\hat{z}(t)) + \gamma(u, y) \tag{6.87}$$

Note that the reconstruction error of the reference system, $z - \hat{z}$, satisfies the linear differential equation in Eq. 6.29. The state estimate for the original nonlinear system is obtained using the nonlinear coordinate transformation.

$$\hat{x}(t) = \mathcal{F}(\hat{z}(t)) \tag{6.88}$$

Let us consider the autonomous system. If a coordinate transformation $\mathcal{F}$ that transforms Eq. 6.85 into Eq. 6.34 exists, the following expressions must hold in which $C_j$ is the $\sum_{i=1}^{j} k_i$ row of $C$.

$$f(x) = \frac{\partial \mathcal{F}}{\partial z} h(z) \tag{6.89}$$

$$h_j(\mathcal{F}(z)) = C_j z, \quad j = 1, \ldots, p \tag{6.90}$$

In this method, the observer problem consists of determining whether a solution to these equations exists and, if so, finding this solution. Performing the indicated differentiation and rearranging Eq. 6.89 results in a system of differential equations for $i = 1, \ldots, p$ that must be solved.

$$\frac{\partial \mathcal{F}}{\partial z_{j+1}^i} = \frac{\partial f}{\partial x}\frac{\partial \mathcal{F}}{\partial z_j^i} - \frac{\partial}{\partial x}\left(\frac{\partial \mathcal{F}}{\partial z_j^i}\right) f, \qquad j = 1, \ldots, k_i - 1 \tag{6.91}$$

$$\frac{\partial \mathcal{F}}{\partial z}\frac{\partial \gamma}{\partial z_{k_i}^i} = \frac{\partial f}{\partial x}\frac{\partial \mathcal{F}}{\partial z_{k_i}^i} - \frac{\partial}{\partial x}\left(\frac{\partial \mathcal{F}}{\partial z_{k_i}^i}\right) f \tag{6.92}$$

Eq. 6.90 results in a system of linear algebraic equations for $\partial \mathcal{F}/\partial z_1^i$ with $i = 1, \ldots, p$.

$$L_f^{j-1}(dg_k)\frac{\partial \mathcal{F}}{\partial z_1^i} = \delta_{i,k}\delta_{k_i,j} \qquad j = 1, \ldots, k_i;\ k = 1, \ldots, p \tag{6.93}$$

The original nonlinear system in Eq. 6.34 can be expressed in observer form if and only if there is a $\mathcal{F}$ and $\gamma$ that satisfy Eqs. 6.91 through 6.93. Since the nonlinear system must be observable for a transformation to exist, the

relationship in Eq. 6.93 can be expressed using the observability matrix in Eq. 6.38 as follows.

$$\left[\frac{\partial \mathcal{F}}{\partial z_1^1} \; \frac{\partial \mathcal{F}}{\partial z_1^2} \; \cdots \; \frac{\partial \mathcal{F}}{\partial z_1^p}\right] \;=\; \mathcal{O}^{-1}C \tag{6.94}$$

The necessary and sufficient conditions for the existence of a linearizing transformation along with a detailed computational procedure to determine the transformation is provided in Xia and Gao [66]. For systems with inputs, it can be shown that the conditions are similar to those presented for autonomous systems and must hold for all admissible inputs [66]. In practice, analytical solutions for $\mathcal{F}$ and $\gamma$ are not obtainable and numerical solutions are used. The formulation presented here considers a continuous output measurement for the estimated state calculation in Eq. 6.87. A corresponding observer for autonomous discrete-time nonlinear systems is discussed in Lee and Nam [35] and Chung and Grizzle [11]. The existence of a continuous-time linearizing transformation for a given nonlinear system does not imply the existence of a discrete linearizing transformation [11]. An excellent presentation of the application of this technique to a continuous-time stirred tank reactor is provided in Kantor [27].

### 6.6.2 Extended Luenberger Observer

Application of observer error linearization involves the determination of a coordinate transformation. This transformation exists only for a limited class of nonlinear systems and requires the solution of a system of partial differential equations. The observer in Birk and Zeitz [7] does not require the integration of these equations. This formulation is referred to as an *extended Luenberger observer* because the nonlinear observer gain is determined by linearizing these partial differential equations. Consider the following nonlinear observer.

$$\dot{\hat{x}}(t) \;=\; \mathbf{f}(\hat{x}, u, t) + L(\hat{x}, u)\,(y(t) - \mathbf{g}(\hat{x}, t)) \tag{6.95}$$

In this approach, an extended linearization of the partial differential equations in Eqs. 6.91 and 6.92 about the estimated state is used to obtain the nonlinear observer gain $L(\hat{x}, u)$. After specifying the characteristic equation or observer poles for the transformed system in Eq. 6.86, the observer gain in Eq. 6.95 becomes

$$L(\hat{x}, u) \;=\; \left[(l_{11} + l_{12}L_f + \ldots + l_{1k_1}L_f^{k_1-1} + L_f^{k_1})\mathbf{s}_1, \ldots, \right. \tag{6.96}$$

$$(l_{p1} + l_{p2}L_f + \ldots + l_{pk_p}L_f^{k_p-1} + L_f^{k_p})\mathbf{s}_p\Big] \left[\frac{\partial \mathbf{g}(\hat{x})}{\partial \hat{x}}\right]^{-1}$$

in which $l_{ij}$ are the observer poles for the linearized reconstruction error, $k_j$ are the observability indices, and $L_f^j$ is the Lie derivative operator. The vectors $\mathbf{s}_j$ are determined from the following relationship in which $e^j$ are the standard unit basis vectors and $\partial g_j^*/\partial z_{k_j}^j \neq 0$ can be freely specified such that the calculation of the Lie derivatives in Eq. 6.96 is made as simple as possible.

$$\mathbf{s}_j = \frac{\partial g_j^*(z)}{\partial z_{k_j}^j} \mathcal{O}^{-1} e^j \tag{6.97}$$

A complete derivation of this observer can be found in Birk and Zeitz [7]. The approach is restricted to systems observable for every input with $\mathbf{f}$ and $\mathbf{g}$ smooth analytic functions.

Ciccarella *et al.* [13] present a discrete-time nonlinear Luenberger-type observer that considers single output systems. Let $U_k^n$ represent the series of $n$ past inputs at sample time $k$ and $\Phi(x, U_k^{n-1})$ represent the series of $n$ predicted outputs generated by the discrete nonlinear system in Eq. 6.35 from an initial state $x$ and input sequence $U_k^n$. We define $\bar{\mathbf{F}}^j(x, U^j)$ as the $j$th composite of the system function $\bar{\mathbf{f}}$ with input sequence $U^j$ and $n$ is the dimension of the state vector.

$$U_k^n = \begin{bmatrix} u_{k-n+1} \\ u_{k-n+2} \\ \vdots \\ u_k \end{bmatrix}, \quad \Phi(x, U_k^{n-1}) = \begin{bmatrix} \bar{\mathbf{g}}(\bar{\mathbf{F}}^{n-1}(x, U_{k-1}^{n-1})) \\ \bar{\mathbf{g}}(\bar{\mathbf{F}}^{n-2}(x, U_{k-2}^{n-2})) \\ \vdots \\ \bar{\mathbf{g}}(x) \end{bmatrix}$$

The filtered state estimate is computed from the following discrete system in which $L$ is the observer gain and $z_{k-n+j|k} = \bar{\mathbf{F}}^{j-1}(z_{k-n+1|k}, U_{k-n+j-1}^{j-1})$.

$$\begin{aligned} z_{k-n+2|k+1} &= z_{k-n+2|k} + \bar{\mathcal{O}}^{-1}(z_{k-n+1|k}, U_{k-1}^{n-1}) \\ &\quad \times \left[\bar{B}\left(y_{k+1} - \bar{\mathbf{g}}(z_{k+1|k})\right) + L\left(y_{k-n+1} - \bar{\mathbf{g}}(z_{k-n+1|k})\right)\right] \quad (6.98) \\ \hat{x}_{k|k} &= \bar{\mathbf{F}}^{n-1}(z_{k-n+1|k}, U_{k-1}^{n-1}) \quad (6.99) \end{aligned}$$

In this system, $\bar{\mathcal{O}}$ is the discrete nonlinear observability matrix at sample time $k$ computed from the partial derivative of $\Phi(x, U_k^{n-1})$.

$$\bar{\mathcal{O}}(x, U_k^{n-1}) = \frac{\partial \Phi(x, U_k^{n-1})}{\partial x} \tag{6.100}$$

The observer gain is selected such that the eigenvalues of the perturbed linear system for the reconstruction error in $z$ is locally stable.

$$e_{k-n+2|k+1} = (\bar{A} - L\bar{C})e_{k-n+1|k} + O\left(\left\|e_{k-n+1|k}\right\|^2\right) \quad (6.101)$$

This linear system, which is in Brunovsky canonical form, is determined from a nonlinear transformation on $z$. The approach is restricted to systems observable for every input with $\bar{\mathbf{g}}$ and $\bar{\mathbf{F}}^n(\Phi^{-1}(x, U^{n-1}))$ uniformly Lipschitz continuous functions of the state. Convergence is guaranteed only for a neighborhood of initial state estimates around the true initial state. This neighborhood may be small. A detailed derivation of this observer is contained in [13].

### 6.6.3 Summary of Nonlinear Observers

In addition to the techniques discussed in this section, there have been a number of other nonlinear observer formulations proposed in the literature. A review of several of these observers for continuous-time systems can be found in Walcott *et al.* [63]. Sliding mode observers for continuous-time systems are discussed in Slotine *et al.* [56]. The conditions under which exponentially convergent nonlinear observers exist for continuous-time systems are presented in Xia and Gao [65]. A Lyapunov function based approach for continuous-time systems is discussed in Tsinias [60]. A Luenberger-type observer that considers single-output nonlinear systems of the form

$$\begin{aligned} \dot{x} &= \mathbf{f}(x) + \mathbf{h}(x)u \\ y &= \mathbf{g}(x) \end{aligned} \quad (6.102)$$

is presented in Gauthier *et al.* [20], Ciccarella *et al.* [12], and Deza *et al.* [18]. Deza *et al.* [18] also consider discrete output measurements. Implementation of this observer on a free radical polymerization CSTR is discussed in Van Dootingh *et al.* [61].

The advantage of nonlinear observers is that the calculation of the state estimate generally involves a gain multiplying the difference between the estimated and measured outputs. After the initial construction of the observer, this calculation can be performed quickly and relatively easily for most of these techniques. The disadvantage is the limitations on the systems that can be considered. Many of these approaches have further restrictions on

the system and measurement functions in addition to the observability requirement. The continuous-time formulations are also not appropriate for applications in which the output measurements are available at discrete sampling periods. Another consideration with observer error linearization is the calculations required to construct the observer.

## 6.7 Nonlinear Moving Horizon Estimator

In this section, the moving horizon state estimator presented previously for linear systems is extended to nonlinear systems. The objective is a deterministic least squares estimator for the discrete nonlinear system in Eq. 6.35 with state and measurement disturbances. In this presentation, we assume that these disturbances are additive in which $v \in \Re^p$ is the measurement disturbance vector and $w \in \Re^n$ is the process or state disturbance vector.

$$\begin{aligned} x_{k+1} &= \bar{\mathbf{f}}(x_k, u_k, k) + w_k \\ y_k &= \bar{\mathbf{g}}(x_k, k) + v_k \end{aligned} \tag{6.103}$$

We also require uniform observability of the nonlinear system as discussed in Section 6.4.2.

### 6.7.1 Estimated State Disturbance Approach

Because it is not possible to obtain a general closed-form recursive solution to the nonlinear batch state estimation problem, it is difficult to develop a moving horizon estimator equivalent to the batch problem in the nonlinear case. Therefore, a moving horizon estimator similar to that for linear systems is usually constructed.

$$\min_{\{\hat{x}_{k-N|k}, \hat{w}_{k-N|k}, \ldots, \hat{w}_{k-1|k}\}} \quad \sum_{j=k-N}^{k-1} \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \tag{6.104}$$

$$\text{Subject to:} \qquad \begin{aligned} \hat{x}_{j+1|k} &= \bar{\mathbf{f}}(\hat{x}_{j|k}, u_j, j) + \hat{w}_{j|k} \\ \hat{v}_{j|k} &= y_j - \bar{\mathbf{g}}(\hat{x}_{j|k}, j) \end{aligned} \tag{6.105}$$

This nonlinear moving horizon estimator does not penalize the initial state disturbance in the horizon allowing the initial state estimate in the horizon, $\hat{x}_{k-N|k}$, to be chosen freely. The advantage of this approach is that nominal

convergence of the estimated state to the true state is achieved [46]. However, this choice ignores information from the output measurements prior to sample time $k-N$. Robertson *et al.* [53] suggest penalizing the initial state disturbance by the inverse of an approximate covariance matrix of the state estimate at time $k-N$. This approach attempts to take prior information into account, but does not have a convergence guarantee. As the horizon length is increased, these two approaches yield similar estimates.

We consider estimated process disturbance and state constraints in which $\bar{\mathbf{h}}$ is the estimated state constraint function.

$$h_{\min} \leq \bar{\mathbf{h}}(\hat{x}_{j|k}) \leq h_{\max}, \qquad j = k-N+1, \ldots, k \tag{6.106}$$

$$w_{\min} \leq \hat{w}_{j|k} \leq w_{\max}, \qquad j = k-N, \ldots, k-1 \tag{6.107}$$

The constraints are chosen to define a convex region in the $\hat{w}$ space containing the origin. For systems forced by an input, the state constraints are shifted by the nominal trajectory due to the input. With the first state estimate unconstrained, these constraints form a feasible set at each sampling time.

Since the process model is nonlinear, this estimator requires the solution to a general nonlinear optimization problem. The techniques required to solve this optimization problem are discussed in Section 6.7.4. The state estimate at sample time $n+j$, in which $n = k-N$, is computed from the optimal solution using the following recursion.

$$\begin{aligned} \hat{x}_{n+j+1|k} &= \bar{\mathbf{f}}(\hat{x}_{n+j|k}, u_{n+j}, n+j) + \hat{w}^*_{n+j} \\ \hat{x}_{n|k} &= \hat{x}^*_{n|k} \end{aligned} \tag{6.108}$$

A moving horizon estimator for the continuous-time nonlinear system with discrete output measurements presented in Eq. 6.36 is constructed by assuming discrete state and measurement disturbances.

$$\dot{x}(\tau) = \mathbf{f}(x(\tau), u(\tau), \tau), \qquad x(k\Delta t) = x_k \tag{6.109}$$

$$\begin{aligned} x_{k+1} &= x\left((k+1)\Delta t\right) + w_k \\ y_k &= \mathbf{g}(x_k, k) + v_k \end{aligned} \tag{6.110}$$

This model is used for the equality constraint of the moving horizon estimator in Eqs. 6.104 and 6.105. The state estimates are computed from the optimal solution using the recursion in Eqs. 6.109 and 6.110.

### 6.7.2 Initial State Estimate Approach

A simpler moving horizon estimator can be constructed by not considering estimated state disturbances in the optimization problem and estimating only the initial state in the horizon. This approach is well motivated if the state disturbances are negligible and the measurement is corrupted by zero mean noise [21]. The objective is to determine an initial state estimate that minimizes the difference between the measured and predicted outputs in a least squares sense.

$$\min_{\hat{x}_{k-N-1|k}} \quad \sum_{j=k-N}^{k} \hat{v}_{j|k}^T \hat{v}_{j|k} \tag{6.111}$$

$$\text{Subject to:} \quad \begin{aligned} \hat{x}_{j+1|k} &= \bar{\mathbf{f}}(\hat{x}_{j|k}, u_j, j) \\ \hat{v}_{j|k} &= y_j - \bar{\mathbf{g}}(\hat{x}_{j|k}, j) \end{aligned} \tag{6.112}$$

Nominal convergence of this approach for continuous-time systems is shown by Michalska and Mayne [44]. Nominal convergence for discrete systems follows in the same manner as the moving horizon estimator presented in the previous section. Continuous-time systems with discrete output measurements can also be considered with this approach. In this case, the state disturbance in Eq. 6.110 is removed and this model is used for the equality constraint in Eq. 6.112.

The implementation of this estimator on continuous-time chemical process systems with discrete output measurements is discussed in Jang *et al.* [24], Bequette [5], Kim *et al.* [30], Liebman *et al.* [37], and Ramamurthi *et al.* [51]. The implementation in Jang *et al.* [24], however, is not a moving horizon approach. The observer horizon is completely replaced with $N$ new output measurements before an estimate is made. Kozub and McGregor [31] compare this approach with extended Kalman filtering techniques. The implementation in Ramamurthi *et al.* [51] linearizes the nonlinear model equations about a nominal trajectory within the observer horizon to estimate the initial state. The filtered state estimate is then obtained using the nonlinear model from this initial state. This approach reduces the computational requirements since a quadratic program is solved at each sample time. If the nominal trajectory is close to the estimated trajectory and the system is not highly nonlinear, this implementation can be a reasonable approximation.

The advantage to the initial state estimate approach is a smaller number of decision variables for a given horizon length. This reduction in decision

variables results in less computational time to solve the optimization problem. In addition, the horizon length is the only tuning parameter. However, the assumption of negligible process disturbances is not realistic for many applications. Issues such as unmeasured or unmodeled disturbances, process modeling error, and variation in the model parameters cannot be addressed adequately when only the initial state in the horizon is estimated.

### 6.7.3 Input Observation Error

We have assumed that the input variables of the process are known exactly. However, there is uncertainty in these values in many applications. For example, a control valve may stick and not attain the position indicated by the signal sent to the valve. Errors in the observation of the input values can be taken into account by assuming that there are disturbances to the input. Kim *et al.* [29] and Tjoa and Biegler [59] refer to this technique as an *error-in-variables* method.

A moving horizon estimator for input observation error is obtained by solving a modification of the least squares problem in Eq. 6.104 in which $\mu \in \Re^m$ is the assumed input disturbance.

$$\min_{\{\hat{x}_{k-N|k}, \hat{w}_{j|k}, \hat{\mu}_{j|k}\}} \sum_{j=k-N}^{k-1} \left( \hat{w}_{j|k}^T Q^{-1} \hat{w}_{j|k} + \hat{\mu}_{j|k}^T S^{-1} \hat{\mu}_{j|k} \right) + \sum_{j=k-N}^{k} \hat{v}_{j|k}^T R^{-1} \hat{v}_{j|k} \tag{6.113}$$

$$\text{Subject to:} \quad \begin{aligned} \hat{x}_{j+1|k} &= \bar{\mathbf{f}}(\hat{x}_{j|k}, \hat{u}_{j|k}, j) + \hat{w}_{j|k} \\ \hat{u}_{j|k} &= u_j + \hat{\mu}_{j|k} \\ \hat{v}_{j|k} &= y_j - \bar{\mathbf{g}}(\hat{x}_{j|k}, j) \end{aligned} \tag{6.114}$$

A corresponding moving horizon estimator for continuous-time systems with discrete output measurements is constructed by using the model in Eq. 6.109 for the equality constraint in Eq. 6.114. Removing the estimated state disturbances, as in the initial state estimate approach, results in an estimator that assumes the process disturbances enter the system only through the input.

### 6.7.4 Solution Techniques

The nonlinear moving horizon estimator problem formulation requires the minimization of a least squares objective function subject to nonlinear differential and algebraic constraints. There are two basic solution strategies

that can be employed. The first is referred to as *sequential solution* [30]. In this method, a nonlinear optimization technique is used to determine the estimated state disturbances that are the decision variables of the least squares problem. At each iteration of the optimizer, the objective function is computed by either integrating the nonlinear modeling equations for continuous-time systems or iterating for discrete systems. For gradient based optimizers used with continuous-time systems, the gradient of the objective function can be computed by integration of the sensitivity equations at the same time the model equations are solved [9]. The sequential solution has the advantage of being relatively straightforward to implement and having a small number of decision variables in the optimization problem. However, the computation time required by this method for continuous-time systems may be excessive since the model equations must be integrated at each iteration of the optimizer. This method also does not readily admit algebraic inequality constraints on the states of the system.

The second method is referred to as *simultaneous solution* [37]. In this method, the state variables at each sample time are included as decision variables in an optimization problem that is solved only once. For continuous-time systems, the model equations are discretized in time to convert the differential equation equality constraints into algebraic equality constraints. The discretization is normally performed using orthogonal collocation on finite elements. This strategy easily handles constraints on the states of the system at each sample time. The major drawback is the large number of decision variables and constraint functions that result in a very large order optimization problem that must be solved. The solution normally requires a large-scale, sparse nonlinear optimization technique [37]. A reduced space nonlinear optimization strategy in which the collocation equations are solved inexactly is presented in Logsdon and Biegler [39].

### 6.7.5 Moving Horizon Algebraic Estimator

Instead of using nonlinear optimization techniques to determine the initial state in the horizon as discussed in Section 6.7.2, Moraal and Grizzle [45] employ Newton's method to solve the system of nonlinear equations generated by the state estimation problem. Let $Y_k^N$ represent the series of $N$ past measured outputs, $U_k^N$ the series of $N$ past inputs, and $\Phi(x, U_k^N)$ the series of $N$ predicted outputs generated by the discrete nonlinear system in

Eq. 6.35 from an initial state $x$ and input sequence $U_k^N$.

$$Y_k^N = \begin{bmatrix} y_{i+1} \\ y_{i+2} \\ \vdots \\ y_k \end{bmatrix}, \quad U_k^N = \begin{bmatrix} u_{i+1} \\ u_{i+2} \\ \vdots \\ u_k \end{bmatrix}, \quad \Phi(x, U_k^N) = \begin{bmatrix} \bar{\mathbf{g}}(x) \\ \bar{\mathbf{g}}(\bar{\mathbf{F}}^1(x, U_{i+1}^1)) \\ \vdots \\ \bar{\mathbf{g}}(\bar{\mathbf{F}}^{N-1}(x, U_{k-1}^{N-1})) \end{bmatrix}$$

In this expression, $\bar{\mathbf{F}}^j(x, U^j)$ is the $j$th composite of the system function $\bar{\mathbf{f}}$ with input sequence $U^j$ and $i = k-N$. Similar to the linear algebraic system in Eq. 6.7, a nonlinear system of equations in $\hat{x}_{k-N+1}$ can be constructed using the known inputs and output measurements.

$$Y_k^N - \Phi(\hat{x}_{k-N+1}, U_k^N) = 0 \tag{6.115}$$

The state estimate $\hat{x}_{k-N+1|k}$ is then determined iteratively using Newton's method in which it is assumed that the Jacobian is square.

$$\hat{x}_{k-N+1|k,i+1} = \hat{x}_{k-N+1|k,i} + \left[\frac{\partial \Phi(\hat{x}_{k-N+1|k,i}, U_k^N)}{\partial x}\right]^{-1} \times \left(Y_k^N - \Phi(\hat{x}_{k-N+1|k,i}, U_k^N)\right) \tag{6.116}$$

If there are more equations than states, the inverse is replaced by a pseudo inverse. An initial estimate is obtained from the solution at the previous sample time. This approach does not consider constraints.

The observability requirement for this technique is the existence of a bounded $N$ such that the nonlinear system of equations in Eq. 6.115 has a unique solution for all admissible $U^N$. Sufficient conditions for nominal convergence are provided in [45]. This formulation may fail to converge if not properly initialized and the region of convergence may be small. Modified Newton methods, in which the Jacobian is not explicitly evaluated at each iteration, can be used to reduce the computational requirements. However, these modified methods can effect the rate of convergence. It can be shown that a single iteration of this estimator is equivalent to the extended Kalman filter with $R = \epsilon I$ in the limit as $\epsilon \to 0$. Application of this observer on a continuous stirred tank bioreactor is presented in [45].

### 6.7.6 Summary of Nonlinear Moving Horizon Estimation

Moving horizon estimation can be applied to a very general class of problems since there are few restrictions on the systems that can be considered.

Nominal convergence of the estimator can also be guaranteed. This approach easily incorporates constraints on the estimates and very general disturbance modeling. It is not restricted to the additive process and measurement disturbances presented in this discussion. Since the nonlinear system is solved exactly, the performance does not suffer from the use of an approximate model. However, this approach requires more computational effort than most of the other nonlinear filters and observers. The additional computational effort can become substantial and prohibit the use of this technique for large-scale systems or systems with short sampling periods.

The solution methods required for an efficient implementation are also more complex. The choice of solution strategy to be employed depends on the application. The sequential approach is normally considered for simple problems due to its relative ease of implementation. This technique is also considered for systems with a large number of states since a prohibitive number of decision variables may be required by the simultaneous approach. However, the sequential approach will normally not be as efficient as the simultaneous approach. The sequential approach may also fail in applications in which there are constraints on the estimated states.

*Example 6.4*

We demonstrate the performance of the moving horizon estimators presented in Sections 6.7.1 and 6.7.2 on a simple linear integrating process with $A = 1$ and $C = 1$ and no constraints on the estimator. We consider the case of zero mean measurement noise with no process disturbances and the case of random process disturbances with no measurement noise. When only measurement noise is present in the system, the true state is stationary at zero. When only process disturbances are present, the measurement is the true state. The performance of the estimator with state disturbances in Section 6.7.1 is shown in Figures 6.4 and 6.5. For the measurement noise case, this estimator is tuned to reflect that the measurement is unreliable using a horizon length $N = 20$, $Q = 1$, and $R = 100$. The deviations between the estimated output and measured output are not heavily penalized in comparison to the estimated state disturbances which results in state estimates that are very close to the true zero state. For the process disturbance case, the estimator is tuned with $Q = 1$ and $R = 0.1$ to reflect that the measurements are relatively reliable. With a horizon length $N = 20$, the estimated state essentially tracks the true state.

The performance of the initial state estimator in Section 6.7.2 with the

same horizon length, $N = 20$, is shown in Figures 6.6 and 6.7. This estimator is essentially identical to the previous approach when only measurement noise is present. However, when process disturbances are present, this estimator does a very poor job of estimating the true state. The estimate of an initial state at the beginning of the horizon cannot take into account the state disturbances that enter the system within the observer horizon.

When the horizon length is shortened to $N = 2$, the performance with process disturbances present improves, as shown in Figure 6.8, but the estimate still lags the true state. With the shorter horizon, however, the performance of this estimator with measurement noise degrades considerably as shown in Figure 6.9. This example clearly shows that initial state estimate approaches are not suitable when significant process disturbances are present.

## 6.8 Combined State and Parameter Estimation

The assumption of constant model parameters made throughout this discussion may not be realistic in many practical applications. Parameters, such as heat transfer coefficients, can change during the operation of chemical processes. In these cases, a more accurate determination of the state can be made by allowing one or more model parameters to vary. The parameter values are then estimated along with the state. This technique is referred to as combined state and parameter estimation. A brief summary of this topic is presented in this section. Further information on parameter estimation for nonlinear systems is available in Bard [4].

Combined state and parameter estimation typically is performed by augmenting the state of the system to include those parameters to be estimated. The most common assumption made to describe the dynamic behavior of these parameters is an integrated white noise process. Using this assumption, the continuous-time nonlinear model in Eq. 6.34 and the discrete nonlinear model in Eq. 6.35 are augmented as follows in which $\Theta$ is the vector of parameters to be estimated.

$$\begin{bmatrix} \dot{x} \\ \dot{\Theta} \end{bmatrix} = \begin{bmatrix} \mathbf{f}(x, u, \Theta, t) \\ 0 \end{bmatrix}; \qquad \begin{bmatrix} x_{k+1} \\ \Theta_{k+1} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{f}}(x_k, u_k, \Theta_k, k) \\ \Theta_k \end{bmatrix} \tag{6.117}$$

$$y = \mathbf{g}(x, \Theta, t); \qquad y_k = \bar{\mathbf{g}}(x_k, \Theta_k, k)$$

Observability of the augmented system is not implied by observability of the original system and must be verified using the techniques discussed

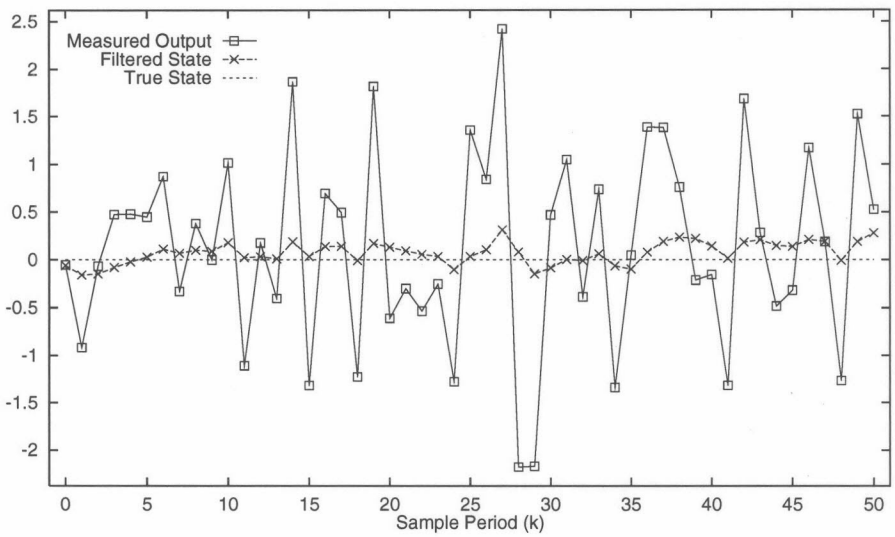


Figure 6.4: Filtered state estimate for the estimated state disturbance moving horizon estimator with measurement noise, $N = 20$.

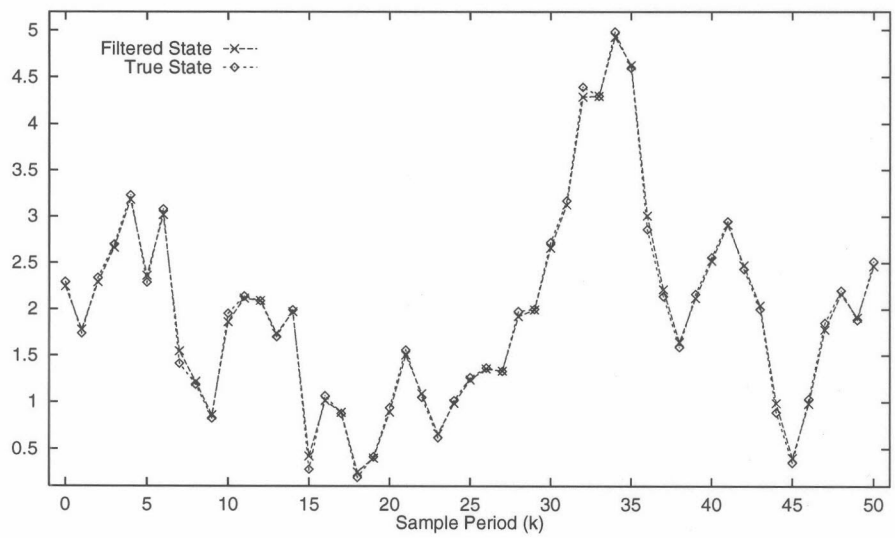


Figure 6.5: Filtered state estimate for the estimated state disturbance moving horizon estimator with process disturbances, $N = 20$.

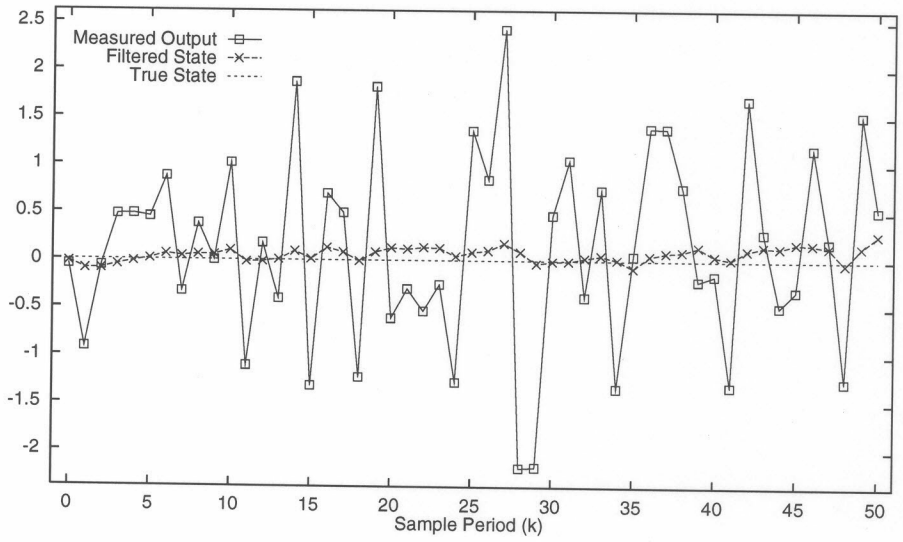


Figure 6.6: Filtered state estimate for the initial state estimate moving horizon estimator with measurement noise, $N = 20$.

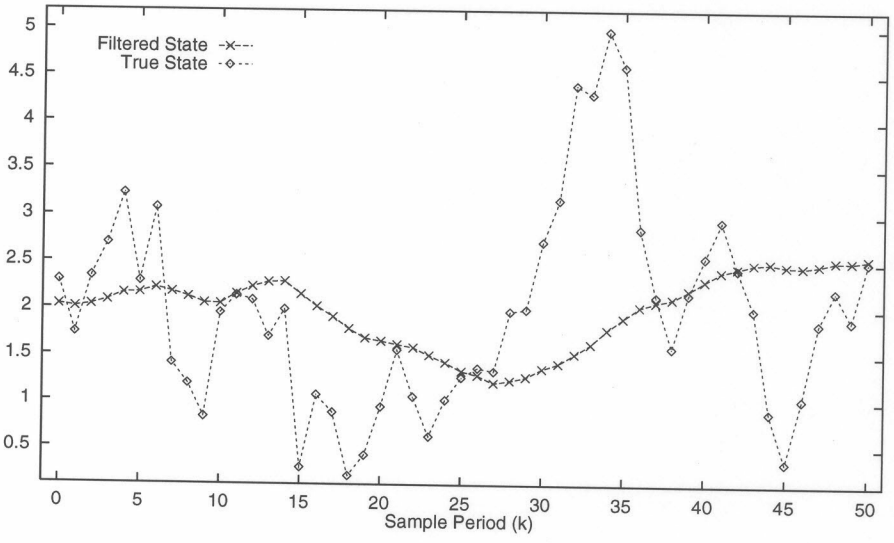


Figure 6.7: Filtered state estimate for the initial state estimate moving horizon estimator with process disturbances, $N = 20$.

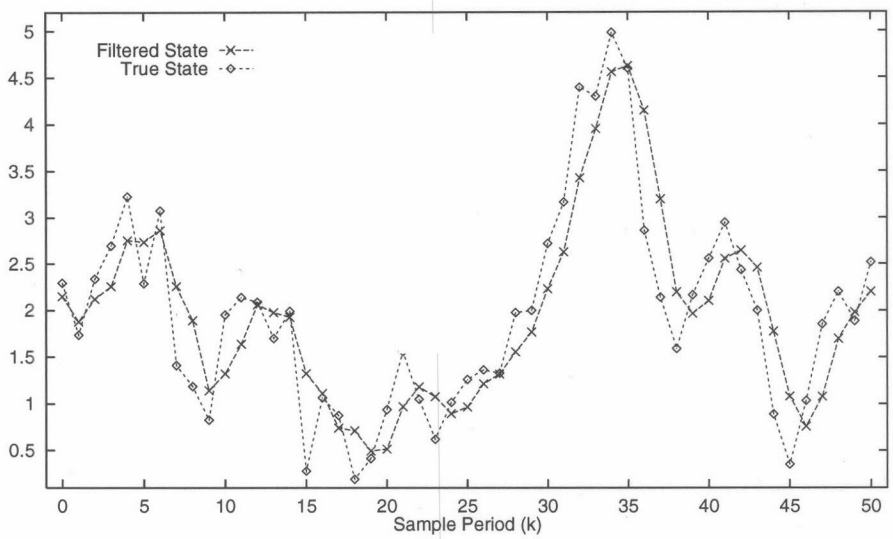


Figure 6.8: Filtered state estimate for the initial state estimate moving horizon estimator with process disturbances, $N = 2$.

in Section 6.4.2. Observability may impose restrictions on the parameters that can be estimated for a given system. Application of the nonlinear state estimation techniques discussed in this chapter on the augmented system is used to provide an estimate of both the state and parameters. In practice, only a small number of the model parameters are estimated.

Parameter estimation using the extended Kalman filter is discussed in Cox [14], Ljung [38], and DeVallière and Bonvin [16]. The extended Schmidt–Kalman filter is discussed in Jazwinski [25]. This filter accounts for the effect of parameter variation on the state, but does not compute an estimate of the parameter values. A discussion of modeling techniques that can improve the estimation of parameter values is presented in DeVallière and Bonvin [17]. Agarwal and Bonvin [1] recommend a decoupled estimator in which the states and parameters are computed separately to reduce the nonlinearity due to parameter-state interactions.

Much of the research in the area of moving horizon estimation has been related to data reconciliation in which parameter estimation is routinely performed. In the moving horizon data reconciliation problem, the model parameters are also included as decision variables in the least squares objective. The optimization problem can be set up to either directly compute

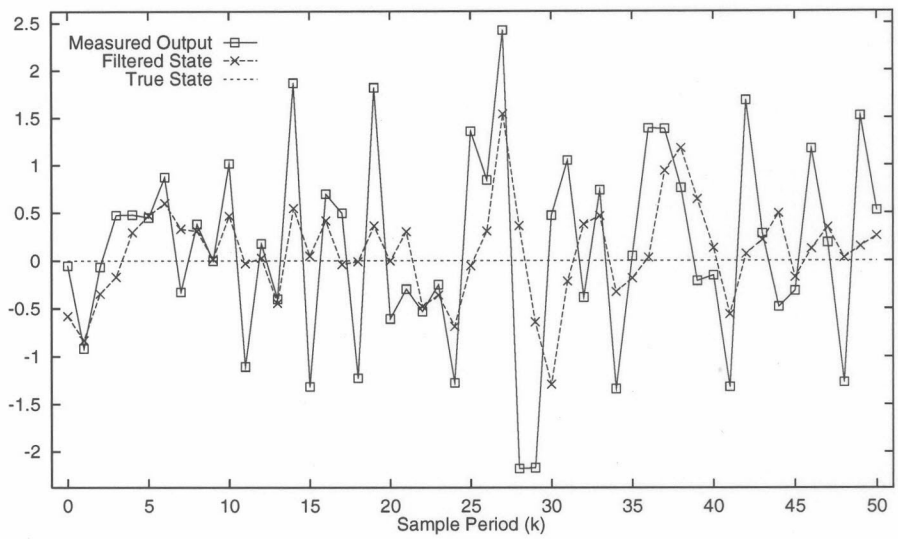


Figure 6.9: Filtered state estimate for the initial state estimate moving horizon estimator with measurement noise, $N = 2$.

the parameter values or compute deviations from nominal values. Combined state and parameter estimation using a moving horizon approach is discussed in Kim *et al.* [30], Liebman *et al.* [37], Tjoa and Biegler [59], and Ramamurthi *et al.* [51].

## 6.9 Nonlinear State Estimation Summary

The selection of an appropriate nonlinear state estimation technique depends on several factors. A significant consideration is the computational requirement. In order to implement an estimator, it must be able to determine the state estimate within the sampling period of the application. Computationally intensive techniques, such as moving horizon estimation and globally iterated extended Kalman filtering, may not be appropriate for applications with short sampling periods because of the computational time required. Large-scale systems can also restrict the techniques that can be considered since the computational requirement generally increases with the number of states that must be estimated.

Modeling error and unmeasured disturbances are other considerations in the selection of a nonlinear estimator. In essentially every application,

there will be a difference between the process and the process model. This difference is referred to as model error or mismatch. There will also be disturbances entering the system that are not accounted for in the model. In most cases, it is not possible to completely eliminate these sources of error that can impose limitations on the performance of the estimator. With a poor or uncertain model, it may not be worthwhile to implement a complex estimation technique since the achievable performance may be no better than simpler approaches. Unfortunately, no general guidelines exist regarding the effect of model error on the selection of a nonlinear state estimation technique.

Simulation studies in which a given technique performs very well for an application may not be a reliable indication of how that technique will perform in practice on similar applications. Slight structural errors in the model, mismatch in the parameter values, and unmeasured disturbances can seriously impact the performance of these techniques. A comprehensive simulation study that includes variation in the model parameters and disturbances to the process and measurements should be carried out before a technique is implemented. The selection of these perturbations is based on engineering judgment of what reasonably can be expected to occur in the process.

The proper nonlinear state estimation technique depends on the application being considered and the computational constraints. For some applications, several attempts at selecting and tuning an estimator may be required before a suitable approach is found. In many cases, however, more than one technique will be appropriate and the choice generally will be based on ease of implementation. The intent here is to present sufficient background information on the selection of nonlinear state estimation techniques for the process control engineer and to provide references for those requiring more detail.

## References

[1] M. Agarwal and D. Bonvin. Limitations of the extended Kalman filter for batch reactors. In J. Rinsdorp, editor, *Proceedings of the 1989 IFAC DYCORD Symposium*, pages 287–294, Oxford, 1989. Pergamon Press.

[2] K. J. Åström. *Introduction to Stochastic Control Theory.* Academic Press, San Diego, California, 1970.

[3] M. Athans, R. P. Wishner, and A. Bertolini. Suboptimal state estimation for continuous–time nonlinear systems from discrete noisy measurements. *IEEE Trans. Auto. Cont.*, 13(5):504–514, 1968.

[4] Y. Bard. *Nonlinear Parameter Estimation.* Academic Press, New York, 1974.

[5] B. W. Bequette. Nonlinear predictive control using multi-rate sampling. *Can. J. Chem. Eng.*, 69:136–143, 1991.

[6] D. Bestle and M. Zeitz. Canonical form observer design for non-linear time-variable systems. *Int. J. Control*, 38(2):419–431, 1983.

[7] J. Birk and M. Zeitz. Extended Luenberger observer for non–linear multivariable systems. *Int. J. Control*, 47(6):1823–1836, 1988.

[8] A. E. Bryson and Y. Ho. *Applied Optimal Control.* Hemisphere Publishing, New York, 1975.

[9] M. Caracotsios and W. E. Stewart. Sensitivity analysis of initial value problems with mixed ODEs and algebraic equations. *Comput. Chem. Eng.*, 9(4):359–365, 1985.

[10] D. E. Catlin. *Estimation, Control, and the Discrete Kalman Filter.* Springer-Verlag, New York, 1989.

[11] S. T. Chung and J. W. Grizzle. Observer error linearization for sampled–data systems. *Automatica*, 26:997–1007, 1990.

[12] G. Ciccarella, M. Dalla Mora, and A. Germani. A Luenberger-like observer for nonlinear systems. *Int. J. Control*, 57(3):537–556, 1993.

[13] G. Ciccarella, M. Dalla Mora, and A. Germani. Observers for discrete-time nonlinear systems. *Sys. Cont. Let.*, 20:373–382, 1993.

[14] H. Cox. On the estimation of state variables and parameters for noisy dynamic systems. *IEEE Trans. Auto. Cont.*, 9(1):5–12, 1964.

[15] C. E. De Souza. On stabilizing properties of solutions of the Riccati difference equation. *IEEE Trans. Auto. Cont.*, 34(12):1313–1316, 1989.

[16] P. DeVallière and D. Bonvin. Application of estimation techniques to batch reactors–II. Experimental studies in state and parameter estimation. *Comput. Chem. Eng.*, 13(1):11–20, 1989.

[17] P. DeVallière and D. Bonvin. Application of estimation techniques to batch reactors–III. Modeling refinements which improve the quality of state and parameter estimation. *Comput. Chem. Eng.*, 14(7):799–808, 1990.

[18] F. Deza, E. Busvelle, J. P. Gauthier, and D. Rakotopara. High gain estimation for nonlinear systems. *Sys. Cont. Let.*, 18:295–299, 1992.

[19] J. Dimitratos, C. Georgakis, M. El-Aasser, and A. Klein. Dynamic modeling and state estimation for an emulsion copolymerization reactor. *Comput. Chem. Eng.*, 13(1):21–33, 1989.

[20] J. P. Gauthier, H. Hammouri, and S. Othman. A simple observer for nonlinear systems – Applications to bioreactors. *IEEE Trans. Auto. Cont.*, 37(6):875–880, 1992.

[21] A. Gelb, editor. *Applied Optimal Estimation.* The M.I.T. Press, Cambridge, Massachusetts, 1974.

[22] R. Hermann and A. J. Krener. Nonlinear controllability and observability. *IEEE Trans. Auto. Cont.*, 22(5):728–740, 1977.

[23] Y. C. Ho and R. C. K. Lee. A Bayesian approach to problems in stochastic estimation and control. *IEEE Trans. Auto. Cont.*, 9(5):333–339, 1964.

[24] S.-S. Jang, B. Joseph, and H. Mukai. Comparison of two approaches to on-line prarameter and state estimation of nonlinear systems. *Ind. Eng. Chem. Proc. Des. Dev.*, 25:809–814, 1986.

[25] A. H. Jazwinski. *Stochastic Proceses and Filtering Theory.* Academic Press, New York, 1970.

[26] T. Kailath. *Linear Systems.* Prentice-Hall, Englewood Cliffs, New Jersey, 1980.

[27] J. C. Kantor. A finite dimensional nonlinear observer for an exothermic stirred-tank reactor. *Chem. Eng. Sci.*, 44(7):1503–1510, 1989.

[28] S. S. Keerthi and E. G. Gilbert. Optimal infinite-horizon feedback laws for a general class of constrained discrete-time systems: Stability and moving-horizon approximations. *J. Optim. Theory Appl.*, 57(2):265–293, 1988.

[29] I. Kim, M. Liebman, and T. Edgar. Robust error-in-variables estimation using nonlinear programming techniques. *AIChE J.*, 36(7):985–993, 1990.

[30] I. Kim, M. Liebman, and T. Edgar. A sequential error-in-variables method for nonlinear dynamic systems. *Comput. Chem. Eng.*, 15(9):663–670, 1991.

[31] D. J. Kozub and J. F. MacGregor. State estimation for semi–batch polymerization reactors. *Chem. Eng. Sci.*, 47(5):1047–1062, 1992.

[32] A. J. Krener and A. Isidori. Linearization by ouput injection and nonlinear observers. *Sys. Cont. Let.*, 3(1):47–51, 1983.

[33] A. J. Krener and W. Respondek. Nonlinear observers with linearizable error dynamics. *SIAM J. Cont. Opt.*, 23(2):197–216, 1985.

[34] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* John Wiley and Sons, New York, 1972.

[35] W. Lee and K. Nam. Observer design for autonomous discrete–time nonlinear systems. *Sys. Cont. Let.*, 17(1):49–58, 1991.

[36] F. L. Lewis. *Optimal Estimation.* John Wiley and Sons, New York, 1986.

[37] M. Liebman, T. Edgar, and L. Lasdon. Efficient data reconciliation and estimation for dynamic processes using nonlinear programming techniques. *Comput. Chem. Eng.*, 16(10/11):963–986, 1992.

[38] L. Ljung. Asymptotic behaviour of the extended Kalman filter as a parameter estimator for linear systems. *IEEE Trans. Auto. Cont.*, 24(1):36–50, 1979.

[39] J. Logsdon and L. T. Biegler. A relaxed reduced space SQP strategy for dynamic optimization problems. *Comput. Chem. Eng.*, 17(4):367–372, 1993.

[40] D. G. Luenberger. An introduction to observers. *IEEE Trans. Auto. Cont.*, 16(2):596–602, 1971.

[41] A. K. Mahalanabis and M. Farooq. A second–order method for state estimation of non–linear dynamical systems. *Int. J. Control*, 14(4):631–639, 1971.

[42] S. I. Marcus. Stochastic integrals and stochastic calculus. In M. G. Singh, editor, *Systems and Control Encyclopedia: Theory, Technology, Applications*, pages 4631–4638, Oxford, 1987. Pergamon Press.

[43] E. S. Meadows, K. R. Muske, and J. B. Rawlings. Constrained state estimation and discontinuous feedback in model predictive control. In *Proceedings of the 1993 European Control Conference*, pages 2308–2312. European Automatic Control Council, June 1993.

[44] H. Michalska and D. Q. Mayne. Moving horizon observers and observer-based control. *IEEE Trans. Auto. Cont.*, 40(6):995–1006, 1995.

[45] P. E. Moraal and J. W. Grizzle. Observer design for nonlinear systems with discrete-time measurements. *IEEE Trans. Auto. Cont.*, 40(3):395–404, 1995.

[46] K. R. Muske and J. B. Rawlings. Nonlinear moving horizon state estimation. In R. Berber, editor, *NATO Advanced Study Institute on Methods of Model Based Control Proceedings*, Antalya, Turkey, July 1994. Kluwer Academic Publishers.

[47] K. R. Muske, J. B. Rawlings, and J. H. Lee. Receding horizon recursive state estimation. In *Proceedings of the 1993 American Control Conference*, pages 900–904, June 1993.

[48] M. Myers and R. H. Luecke. Process control applications of an extended Kalman filter algorithm. *Comput. Chem. Eng.*, 15(12):853–857, 1991.

[49] H. Nijmeijer. Observability of autonomous discrete time nonlinear systems: A geometric approach. *Int. J. Control*, 36(5):867–874, 1982.

[50] P. Park and T. Kailath. New square-root algorithms for Kalman filtering. *IEEE Trans. Auto. Cont.*, 40(5):895–899, 1995.

[51] Y. Ramamurthi, P. Sistu, and B. Bequette. Control-relevant dynamic data reconciliation and parameter estimation. *Comput. Chem. Eng.*, 17(1):41–59, 1993.

[52] W. H. Ray. *Advanced Process Control.* McGraw-Hill, New York, 1981.

[53] D. Robertson, J. H. Lee, and J. B. Rawlings. A moving horizon based approach for least squares state estimation. Submitted to *AIChE J.*, 1995.

[54] M. G. Safonov and M. Athans. Robustness and computational aspects of nonlinear stochastic estimators and regulators. *IEEE Trans. Auto. Cont.*, 23(4):717–725, 1978.

[55] H. Schuler and Z. Suzhen. Real-time estimation of the chain length distribution in a polymerization reactor. *Chem. Eng. Sci.*, 40(10):1891–1904, 1985.

[56] J. J. E. Slotine, J. K. Hedrick, and E. A. Misawa. On sliding observers for nonlinear systems. *J. Dynam. Sys. Meas. Cont.*, 109(3):245–252, 1987.

[57] E. D. Sontag. A concept of local observability. *Sys. Cont. Let.*, 5:41–47, 1984.

[58] R. F. Stengel. *Stochastic Optimal Control: Theory and Application.* John Wiley and Sons, New York, 1986.

[59] I.-B. Tjoa and L. T. Biegler. Reduced sucessive quadratic programming strategy for errors–in–variables estimation. *Comput. Chem. Eng.*, 16(6):523–533, 1992.

[60] J. Tsinias. Further results on the observer design problem. *Sys. Cont. Let.*, 14:411–418, 1990.

[61] M. Van Dootingh, F. Viel, D. Rakotopara, J. P. Gauthier, and P. Hobbes. Nonlinear deterministic observer for state estimation: Application to a continuous free radical polymerization reactor. *Comput. Chem. Eng.*, 16(8):777–791, 1992.

[62] M. Vidyasagar. *Nonlinear Systems Analysis.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 2nd edition, 1993.

[63] B. Walcott, M. Corless, and S. Żak. Comparative study of non–linear state–observation techniques. *Int. J. Control*, 45(6):2109–2132, 1987.

[64] R. P. Wishner, J. A. Tabaczynski, and M. Athans. A comparison of three nonlinear filters. *Automatica*, 5:478–495, 1969.

[65] X.-H. Xia and W.-B. Gao. On exponential observers for nonlinear systems. *Sys. Cont. Let.*, 11:319–325, 1988.

[66] X.-H. Xia and W.-B. Gao. Nonlinear observer design by observer error linearization. *SIAM J. Cont. Opt.*, 27(1):199–216, 1989.