NONLINEAR OBSERVERS

A. J. Krener

University of California, *Davis*, *CA* , *USA*

Keywords: nonlinear observer, state estimation, nonlinear filtering, observability, high gain observers, minimum energy estimation, H^∞ estimation, extended Kalman filter

Contents

Summary

This paper is a review of the existing methods for designing an observer for a system modeled by nonlinear equations. We focus our attention on autonomous, finite dimensional systems described by ordinary differential equations. The current condition of such a system is described by its state variables about which we just have partial and possibly noisy measurements. The goal of the observer is to process these measurements and any information regarding the initial state of the system and to obtain an estimate of the current state of the system. This estimate should improve with additional measurements and, ideally, converge to the true value in the absence of noise. The observer does this by taking advantage of our *a priori* knowledge of the dynamics of the system.

1. Introduction

Systems are sets of components, physical or otherwise, which are connected in such a manner as to form and act as entire units. A nonlinear system is described by a mathematical model consisting of inputs, states, and outputs whose dynamics is given by nonlinear equations. Such models are used to represent a wide variety of dynamic processes in the real world. The inputs are the way the external world affects the system, the states are the internal memory of the system and the outputs are the way the system affects the external world. An example of such a system is

$$\dot{x}(t) = f\left(t, x(t), u(t)\right) \quad (1)$$

$$y(t) = h\left(t, x(t), u(t)\right) \quad (2)$$

$$x(0) \approx \hat{x}^0 . \tag{3}$$

The input is the m vector u, the state is the n vector x and the output is the p vector y. The state of the system at the initial time $t = 0$ is not known exactly but is approximately $\hat{x}^0$. Typically, the dimensions of the input and output are less than that of the state.

A particular case is an autonomous linear system

$$\dot{x} = Ax + Bu \tag{4}$$

$$y = Cx + Du \tag{5}$$

$$x(0) \approx \hat{x}^0 . \tag{6}$$

Other examples include systems described by difference equations

$$x(t+1) = f\left(t, x(t), u(t)\right) \tag{7}$$

$$y(t) = h\left(t, x(t), u(t)\right) \tag{8}$$

and infinite dimensional systems described by partial differential and/or difference equations, delay differential equations or integro-differential equations. This review will focus on finite dimensional systems described by ordinary differential equations.

An observer is a method of estimating the state of the system from partial and possibly noisy measurements of the inputs and outputs and inexact knowledge of the initial condition. More precisely an observer is a causal mapping from any prior information about the initial state x^0 and from the past inputs and outputs

$$\left\{\left(u(\tau), y(\tau)\right) : t^0 \leq \tau \leq t\right\} \tag{9}$$

to an estimate $\hat{x}(t)$ of the current state $x(t)$ or an estimate $\hat{z}(t)$ of some function $z(t) = k(x(t))$ of the current state. Causality means that the estimate at time t does not depend on any information about the inputs and outputs after time t. This restriction reflects the need to use the estimate in real time to control the system. The essential requirement of an observer is that the estimate converges to the true value as t gets large.

Sometimes it is not necessary to estimate the full state but only some function of it, say $\kappa(t, x)$. For example, if one wishes to use the feedback control $u = \kappa(t, x)$. This article will focus on observers of the full state.

The prototype of an observer is that of an autonomous linear system Eqs. (4) - (6). The system

$$\dot{\hat{x}} = A\hat{x} + Bu + L\left(y - \hat{y}\right) \tag{10}$$

$$\hat{y} = C\hat{x} + Du \tag{11}$$

$$\hat{x}(0) = \hat{x}^0 \tag{12}$$

is an observer where L is an $n \times p$ matrix to be chosen by the designer. The dynamics of the error $\tilde{x} = x - \hat{x}$ is given by

$$\dot{\tilde{x}} = \left(A - LC\right)\tilde{x} \tag{13}$$

$$\tilde{x}(0) = x^0 - \hat{x}^0 \tag{14}$$

If the spectrum of the matrix $A - LC$ lies in the open left half plane, then the error decays to zero exponentially fast. In this way, the problem of designing an observer for an autonomous linear system is reduced to the following problem. Given A, C, find L so that $A - LC$ is Hurwitz, i. e., the spectrum of $A - LC$ is in the open left half plane. We discuss when L can be so chosen in the next section (see *Design Techniques for Time Varying Systems* for further details.)

For nonlinear systems the distinction between nonautonomous Eqs. (1) - (3) and autonomous systems

$$\dot{x} = f(x,u) \tag{15}$$

$$y = h(x,u) \tag{16}$$

$$x(0) = \hat{x}^0 \tag{17}$$

is frequently not important as one can add time as an extra state $x_{n+1} = t - t^0$ and thereby reduce the former to the latter. Since an observer operates in real time, time is usually observable and so can be added as an extra output also. Frequently models depend on parameters θ as in $\dot{x} = f(x,u,\theta)$. But in a nonlinear system the distinction between states and parameters is not always clearcut. Parameters can always be treated as additional states by adding the differential equation $\dot{\theta} = 0$. Therefore, the problem of real time parameter estimation reduces to the problem of real time state estimation and may be solvable by an observer. If the state estimate is not going to be used in real time, then one can collect data after time t to estimate $x(t)$. This problem is sometimes called nonlinear smoothing and is related to the identification of nonlinear systems (see 6.43.10) .

Another example of an observer is the extended Kalman filter described in more detail in (see *State Reconstruction by Extended Kalman Filter*) and in the following statements. This is an observer for a nonlinear, nonautonomous system Eqs. (1) - (3) which is derived using stochastic arguments. Two quantities $\hat{x}(t)$ and $P(t)$ are computed by the extended Kalman filter. The stochastic interpretation is that the distribution of the true state $x(t)$ is approximately Gaussian with mean $\hat{x}(t)$ and covariance $P(t)$.

Most observers are described recursively as a dynamical system whose input is the measured variables $\begin{bmatrix} u \\ y \end{bmatrix}$ and whose output is the state estimate $\hat{x}$ such as

$$\dot{z} = \hat{f}(t,z,u,y) \tag{18}$$

$$\hat{x} = \hat{h}(t,z,u,y) \tag{19}$$

If the state of the observer, z, is of the same dimension as the state of the system, then it is called a full order observer; if it is of greater dimension then it is called an expanded order observer, and if it is of lesser dimension, then it is called a reduced order observer.

For example, the prototype autonomous linear observer Eqs. (10) - (12) can be written as

$$\dot{z} = (A - LC)z + \begin{bmatrix} B - LD & L \end{bmatrix} \begin{bmatrix} u \\ y \end{bmatrix} \tag{20}$$

$$\hat{x} = z \tag{21}$$

$$z(0) = \hat{x}^0 \tag{22}$$

and hence is a full order observer. The state of extended Kalman filter discussed as follows is the pair $z=(\hat{x},P)$, so it is an expanded order observer. We briefly discuss the Luenberger observer, a reduced order observer for a linear autonomous system in the form

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} u \qquad (23)$$

$$y = x_1 + Du \qquad (24)$$

$$x(0) = x^0 \qquad (25)$$

The reduced order observer is given by

$$\dot{z} = (A_{22} - LA_{12})z + \left[(A_{21} - LA_{11}) + (A_{22} - LA_{12})L\right](y - Du) \qquad (26)$$

$$\hat{x}_1 = y - Du \qquad (27)$$

$$\hat{x}_2 = z + L(y - Du) \qquad (28)$$

where L is a design parameter. If the model is exact then $\hat{x}_1 = x_1$ and

$$\dot{\tilde{x}}_2 = (A_{22} - LA_{12})\tilde{x}_2 \qquad (29)$$

so if the spectrum of the matrix $A_{22} - LA_{12}$ lies in the open left half plane then the error decays to zero exponentially fast. We discuss when L can be so chosen in the next section. For more on reduced order linear observers, (see (6.43.5.3)) .

The state z of the observer is some measure of the likely distribution of the state of the original system given the past observations. If the observer is derived using stochastic arguments, the state of the observer is typically the conditional density of the state of the system given the past observations and the initial information. In the extended Kalman filter, the state $z=(\hat{x},P)$ is the mean and the covariance of the approximately Gaussian distribution of the true state. For the full and reduced order linear observers described previously, which were derived by nonstochastic arguments, one can view the conditional density as being singular and concentrated at a single point, $\hat{x}(t)$.

2. Observability

The question of whether an observer converges is of paramount importance. A more immediate question is when a nonlinear system Eqs. (15) - (17) admits a convergent observer. This leads to the concepts of observability and detectability which are discussed in (see 6.43.21.7) . Briefly two states x^{01}, x^{02} are said to be distinguishable by an input $u(t)$ if the outputs $y^1(t)$, $y^2(t)$ of Eqs. (15) - (17) satisfying the initial conditions $x^0 = x^{01}, x^0 = x^{02}$ differ at some time $t \geq 0$. The system is said to be observable if every pair x^{01}, x^{02} can be distinguished by some input $u(t)$. An input $u(t)$ which distinguishes every pair x^{01}, x^{02} is said to be universal. A system where every input is universal is said to be uniformly observable.

Consider a smooth autonomous nonlinear system without inputs

$$\dot{x} = f(x) \qquad (30)$$

$$y = h(x) \qquad (31)$$

$$x(0) = x^0 \tag{32}$$

At time $t = 0$ the output and its time derivatives are given by the iterated Lie derivatives

$$y(0) = h(x^0) \tag{33}$$

$$\dot{y}(0) = L_f(h)(x^0) = \frac{\partial h}{\partial x}(x^0) f(x^0) \tag{34}$$

$$\ddot{y}(0) = L_f^2(h)(x^0) = \frac{\partial L_f(h)}{\partial x}(x^0) f(x^0) \tag{35}$$

and so on. If the p-vector-valued functions h, $L_f(h)$, $L_f^2(h)$, . . . distinguish points then clearly the system is observable. For a real analytic system, this is a necessary and sufficient condition for observability. This suggests a way of reconstructing the state of a system, differentiate the output numerous times, and find the state which generates such values. One does not proceed in this fashion because differentiation greatly accentuates the effect of the almost inevitable noise that is present in the observations, and multiple differentiations greatly increase this problem. That is why observers are usually dynamic systems driven by measurements. When such systems are integrated, the effect of the noise is mitigated not enhanced.

For simplicity of exposition, suppose that $n = kp$. If the matrix

$$\begin{bmatrix} \dfrac{\partial(h)}{\partial x}(x^0) \\ \dfrac{\partial L_f(h)}{\partial x}(x^0) \\ \vdots \\ \dfrac{\partial L_f^{k-1}(h)}{\partial x}(x^0) \end{bmatrix} \tag{36}$$

is invertible then the p-vector-valued functions

$$\xi_1 = h(x), \tag{37}$$

$$\xi_2 = L_f(h)(x),\ldots, \tag{38}$$

$$\xi_k = L_f^{k-1}(h)(x) \tag{39}$$

are local coordinates around x^0 and in these coordinates the system Eqs. (30) - (32) becomes

$$y = \xi_1 \tag{40}$$

$$\dot{\xi}_1 = \xi_2 \tag{41}$$

$$\begin{gathered} \dot{\xi}_2 = \xi_3 \\ \vdots \end{gathered} \tag{42}$$

$$\dot{\xi}_k = \bar{f}_k(\xi) \tag{43}$$

Each ξ_i is a p-vector. Such a system is said to be in observable form, since it is clearly observable. Many algorithms for constructing observers start with the assumption that the system is in observable form. The observable form of a $n = kp$ system with inputs is

$$y = \xi_1 + g_0(\xi, u) \tag{44}$$

$$\dot{\xi}_1 = \xi_2 + g_1(\xi, u) \tag{45}$$

$$\begin{aligned} \dot{\xi}_2 &= \xi_3 + g_2(\xi, u) \\ &\vdots \end{aligned} \tag{46}$$

$$\dot{\xi}_k = \bar{f}_k(\xi) + g_k(\xi, u). \tag{47}$$

where $g_i(\xi, 0) = 0$. Such a system is clearly observable as the input $u(t) = 0$ distinguishes every pair of points, but it may not be uniformly observable. A system

$$y = \xi_1 + g_0(u) \tag{48}$$

$$\begin{aligned} \dot{\xi}_1 &= \xi_2 + g_1(\xi_1, u) \\ &\vdots \end{aligned} \tag{49}$$

$$\begin{aligned} \dot{\xi}_i &= \xi_{i+1} + g_2(\xi_1, \ldots, \xi_i, u) \\ &\vdots \end{aligned} \tag{50}$$

$$\dot{\xi}_k = \bar{f}_k(\xi) + g_k(\xi_1, \ldots, \xi_k, u). \tag{51}$$

is said to be in uniformly observable form for it is clearly uniformly observable. From the knowledge of $u(t)$, $y(t)$ we can determine $\xi_1(t)$, from the knowledge of $u(t)$, $y(t)$, $\dot{\xi}_1(t)$ we can determine $\xi_2(t)$, etc.

An autonomous linear system is observable if, and only if, the matrix

$$O = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \tag{52}$$

is of full column rank in which case C, A is said to be an observable pair. Moreover, for such systems the spectrum of $A - LC$ can be set up arbitrarily to complex conjugation by choice of L. (As a real matrix the spectrum of $A - LC$ is invariant with respect to complex conjugation.) (See (6.43.5.3)).

A system Eqs. (15) - (17) is detectable, if whenever the outputs are equal $y^1(t) = y^2(t)$ from the initial states x^{01}, x^{02} using the same control $u(t)$, then the state trajectories converge $x^1(t) - x^2(t) \to 0$.

For an autonomous linear system, the kernel of the matrix Eq. (52) is the largest invariant subspace of the matrix A contained in the kernel of C. It is not hard to show that the system is detectable if, and only if, the spectrum of A restricted to the kernel of Eq. (52) is in the open left half plane. Clearly, the spectrum of $A - LC$ on the kernel of Eq. (52) does not depend on L. The rest of the spectrum of $A - LC$ can be set up arbitrarily to complex conjugation by choice of L.

Hence a linear system admits a convergent observer if, and only if, it is detectable. It is not hard to show that the system Eq. (23) - (25) is detectable if, and only if, the reduced system is.

$$\dot{x}_2 = A_{22} x_2 \tag{53}$$

$$\bar{y} = A_{12} x_2 \tag{54}$$

Hence a linear system admits a convergent reduced order observer if and only if it is detectable.

3. Construction of Observers by Linear Approximation

Consider an autonomous nonlinear system without inputs Eqs. (30) - (32) . If the system is known to operate in a neighborhood of some fixed state, say $x = 0$ where $f(0) = 0$, $h(0) = 0$, then the simplest approach to constructing an observer is to approximate the dynamics around this operating condition by the linear autonomous system

$$\dot{x} = Ax \tag{55}$$

$$y = Cx \tag{56}$$

$$x(0) = x^0 \approx 0 \tag{57}$$

where

$$A = \frac{\partial f}{\partial x}(0) \tag{58}$$

$$C = \frac{\partial h}{\partial x}(0) \tag{59}$$

and use an observer for the latter,

$$\dot{\hat{x}} = A\hat{x} + L(y - \hat{y}) \tag{60}$$

$$\hat{y} = C\hat{x} \tag{61}$$

$$\hat{x}(0) = \hat{x}^0 . \tag{62}$$

The error $\tilde{x} = x - \hat{x}$ dynamics is

$$\dot{\tilde{x}} = (A - LC)\tilde{x} + \bar{f}(x) - L\bar{h}(x) \tag{63}$$

where

$$\bar{f}(x) = f(x) - Ax , \tag{64}$$

$$\bar{h}(x) = h(x) - Cx . \tag{65}$$

If the linear system is detectable, then there are choices of L such that the spectrum of $A - LC$ lies in the open left half plane. But the original system must be locally asymptotically stable to 0 for the observer to converge. And, if it is locally asymptotically stable to 0, an observer may not be needed as the estimate $\hat{x} = 0$ is asymptotically correct.

A slightly more sophisticated approach is preferable. Define the observer to be

$$\dot{\hat{x}} = f(\hat{x}) + L(y - \hat{y}) \tag{66}$$

$$\hat{y} = h(\hat{x}) \tag{67}$$

$$\hat{x}(0)=\hat{x}^0 \quad (68)$$

Suppose the system Eqs. (30) - (32) is stable around $x = 0$ in the sense of Lyapunov; i. e., if the system starts in a sufficiently small neighborhood of 0 it stays close to 0. If the original state x^0 and state estimate error $\tilde{x}^0 = x^0 - \hat{x}^0$ are sufficiently small then the observer error converges to zero. Hence, this is a local observer where local is meant in two senses. Both the original state x^0 and the original state estimate error $\tilde{x}^0 = x^0 - \hat{x}^0$ must be close to 0 for guaranteed convergence of the error.

To see this, suppose the spectrum of $A - LC$ lies in the open left half plane, then there exists a positive definite solution P^{-1} to the Lyapunov equation

$$(A-LC)' P^{-1} + P^{-1}(A-LC) = -I \quad (69)$$

So, if $x(t)$ satisfies Eqs. (30) - (32) and $\hat{x}(t)$ satisfies Eqs. (66) - (68), then

$$\frac{d}{dt}\left(\tilde{x}'P^{-1}\tilde{x}\right) = -\tilde{x}'\tilde{x} + \left(\tilde{f} - L\tilde{h}\right)' P^{-1}\tilde{x} + \tilde{x}'P^{-1}\left(\tilde{f} - L\tilde{h}\right) \quad (70)$$

where

$$\tilde{f} = \tilde{f}(x,\tilde{x}) = f(x) - f(x-\tilde{x}) - A\tilde{x}, \quad (71)$$

$$\tilde{h} = \tilde{h}(x,\tilde{x}) = h(x) - h(x-\tilde{x}) - C\tilde{x}. \quad (72)$$

Assuming that the system is sufficiently smooth, the last two terms on the right side are $O(x)O(\tilde{x})^2$ and so are dominated by $\tilde{x}'\tilde{x}$ for small $x, \tilde{x}$. Hence, the right side is negative and the error converges to zero.

If the system has an input Eqs. (15) - (17) and the input is measurable, then

$$\dot{\hat{x}} = f(\hat{x},u) + L(y-\hat{y}) \quad (73)$$

$$\hat{y} = h(\hat{x},u) \quad (74)$$

$$\hat{x}(0) = \hat{x}^0 \quad (75)$$

is an local observer. If the controlled system Eqs. (15) - (17) stays in a sufficiently small neighborhood of the origin then the observer converges as before based on the analysis of Eq. (70) but with

$$\tilde{f} = f(x,u) - f(x-\tilde{x},u) - A\tilde{x}, \quad (76)$$

$$\tilde{h} = h(x,u) - h(x-\tilde{x},u) - C\tilde{x}. \quad (77)$$

Frequently, the state estimate $\hat{x}$ is used in a feedback control law $u = \kappa(\hat{x})$. For the observer Eqs. (73) - (75) to converge with the open loop system Eqs. (15) - (17) need not be stable, but the closed loop system should be. If the spectrum of $A - LC$ lies in the open left half plane, if the state feedback $u = \kappa(x)$ locally exponential stabilizes the system Eqs. (15) - (17), and if x^0, $\tilde{x}^0$ are sufficiently small then the state estimate feedback $u = \kappa(\hat{x})$ will also be locally exponential stabilizing the system and the observer will converge locally.

The previously shown techniques require choosing L so that $A - LC$ is Hurwitz. Of course, if there is one such L, there are many and the question is which one to choose. One reasonable way of choosing L is via an approximating Kalman filter (see *Kalman Filters*). Assume that the linear approximating system Eqs. (55) - (57) is corrupted by noise,

$$\dot{x} = Ax + Gw \tag{78}$$

$$y = Cx + Jv \tag{79}$$

$$x(0) = \hat{x}^0 + \tilde{x}^0 \tag{80}$$

where w, v are standard independent white Gaussian noises, and $\tilde{x}^0$ is an independent Gaussian initial condition. Let the system be detectable and $Q = GG'$, $R = JJ'$. If R is invertible, then the long time, stationary Kalman filter for this system is

$$\dot{\hat{x}} = Ax + L(y - \hat{y}) \tag{81}$$

$$\hat{y} = C\hat{x} \tag{82}$$

$$x(0) = \hat{x}^0 \tag{83}$$

The observer gain is

$$L = PC'R^{-1} \tag{84}$$

where P is the unique positive definite solution to the algebraic Riccati equation

$$0 = AP + PA' + Q - PC'R^{-1}CP \tag{85}$$

This observer gain L is used in the observer Eqs. (73) - (75).

The Kalman filtering approach in effect replaces the design parameter L by a pair of design parameters Q, R. The tradeoff between these two parameters is roughly as follows. The smaller that R is as compared to Q, the more weight the observer puts on the most recent observations in arriving at its estimate. Making R smaller while holding Q constant tends to move the spectrum of $A-LC$ further left. At first this might seem an unmitigated benefit, but the further left that the spectrum is the more errors in the observations increase the errors in the estimate. An observer with the spectrum far to the left is severely compromised by observation noise and even by driving noise although to a lesser extent. The Kalman filter finds the optimal place to put the spectrum given the relative magnitudes (covariances) of the noise.

If the system Eqs. (1) - (3) is not operating in the neighborhood of some fixed state, then the extended Kalman filtering approach can be used to construct an observer Eqs. (100)- (101). In effect, the nonlinear system Eqs. (1) - (3) is approximated by a time varying linear system along the estimate of the state trajectory with standard independent white Gaussian noises w, v,

$$\dot{x} = A(t)x + B(t)u + G(t)w \tag{86}$$

$$y = C(t)x + D(t)u + J(t)v \tag{87}$$

where

$$A(t) = \frac{\partial f}{\partial x}(t, \hat{x}(t), u(t)), \tag{88}$$

$$B(t) = \frac{\partial f}{\partial u}(t, \hat{x}(t), u(t)), \tag{89}$$

$$C(t)=\frac{\partial h}{\partial x}(t,\hat{x}(t),u(t)), \quad (90)$$

$$D(t)=\frac{\partial h}{\partial u}(t,\hat{x}(t),u(t)). \quad (91)$$

A Kalman filter for this linear system is

$$\dot{\hat{x}}=A(t)\hat{x}+B(t)u+L(t)(y-\hat{y}) \quad (92)$$

$$\hat{y}=C(t)\hat{x}+D(t)u \quad (93)$$

$$\dot{P}(t)=A(t)P(t)+P(t)A'(t)+Q(t)-P(t)C'(t)R^{-1}(t)C(t)P(t) \quad (94)$$

$$\hat{x}(t^0)=\hat{x}^0 \quad (95)$$

$$P(t^0)=P^0 \quad (96)$$

$$Q(t)=G(t)G'(t) \quad (97)$$

$$R(t)=J(t)J'(t) \quad (98)$$

$$L(t)=P(t)C'(t)R^{-1}(t) \quad (99)$$

The form of an extended Kalman filter is slightly different and obtained by changing the first two equations as before

$$\dot{\hat{x}}=f(t,\hat{x},u)+L(t)(y-\hat{y}) \quad (100)$$

$$\hat{y}=h(t,\hat{x},u). \quad (101)$$

The matrices $Q(t)$, $R(t)$ are design parameters, the former represents the uncertainty in the system dynamics (the driving noise covariance) and must be chosen to be nonnegative definite. The latter represents the uncertainty in the system measurements (the measurement noise covariance) and must be chosen to be positive definite. The initial state estimate $\hat{x}^0$ and its covariance P^0 describe the prior knowledge of the true state at the beginning of the process.

The extended Kalman filter is the most widely used nonlinear observer. Its virtues are its relative simplicity and its frequently good performance. Unfortunately, though it is not guaranteed to converge, here is a simple example where it fails.

$$\dot{x}=f(x)=x(1-x^2) \quad (102)$$

$$y=h(x)=x^2-x/2. \quad (103)$$

The system is observable as h, $L_f(h)$, $L_f^2(h)$ separate points. The dynamics has stable equilibria at $x=\pm 1$ and an unstable equilibrium at $x=0$. Under certain conditions, the extended Kalman filter fails to converge. Suppose the $x^0=1$ so $x(t)=1$ and $y(t)=1/2$ for all $t\geq 0$. But $h(-1/2)=1/2$ so if $\hat{x}^0\leq -1/2$ the extended Kalman filter will not converge. To see this notice that when $\hat{x}(t)=-1/2$, the term $y(t)-\hat{y}(t)=0$ so $\dot{\hat{x}}=f(\hat{x}(t))=f(-1/2)=-3/8$. Therefore $\hat{x}(t)\leq -1/2$ for all $t\geq 0$.

It is not hard to see that any one dimensional observer will encounter the same difficulties as the extended Kalman filter. One way around this difficulty might be to embed the system into a higher dimensional system in observer form.

4. Construction of Observers by Error Linearization

There are several approaches that rely on finding a change of state coordinates that makes the problem of constructing an observer easier. Perhaps, the simplest way is to try to find a change of state and output coordinates

$$z=\theta(x) \tag{104}$$

$$w=\gamma(y) \tag{105}$$

that transforms the nonlinear autonomous system Eqs. (15) - (17) into a linear autonomous system with input output injection,

$$\dot{z}=Az+\alpha(u,y) \tag{106}$$

$$w=Cz+\beta(u,y). \tag{107}$$

One would like the transformations to be global diffeomorphisms, but one may have to settle for local diffeomorphisms around x^0, y^0 which map to $z^0 = 0$, $w^0 = 0$.

It is easy to construct an observer for the latter system,

$$\dot{\hat{z}}=A\hat{z}+\alpha(u,y)+L(w-\hat{w}) \tag{108}$$

$$\hat{w}=C\hat{z}+\beta(u,y) \tag{109}$$

$$\hat{z}(0)=\hat{z}^0=\theta(\hat{x}^0) \tag{110}$$

$$\hat{x}=\theta^{-1}(\hat{z}) \tag{111}$$

with linear error dynamics

$$\dot{\tilde{z}}=(A-LC)\tilde{z} \tag{112}$$

$$\tilde{z}(0)=z^0-\hat{z}^0. \tag{113}$$

where $\tilde{z}=z-\hat{z}$. If the nonlinear autonomous system Eqs. (15) - (17) is linearly observable at x^0, then C, A is an observable pair, so one can set the spectrum of $A-LC$ in the open left half plane and the error will go to zero exponentially.

One can leave the observer in $\hat{z}$, w coordinates or transform the observer Eqs. (108) - (111) back into $\hat{x}$, y coordinates,

$$\dot{\hat{x}}=f(\hat{x},u)+\left(\frac{\partial\theta}{\partial x}(\hat{x})\right)^{-1}\left(\alpha(u,y)-\alpha(u,h(\hat{x}))+L\left(\gamma(y)-C\theta(\hat{x})-\beta(u,y)\right)\right) \tag{114}$$

$$\hat{x}(0)=\hat{x}^0 \tag{115}$$

The advantage of $\hat{x}$, y coordinates is that they may be natural to the system. The advantage of $\hat{z}$, w coordinates is that the observer is a stable linear system driven by a signal that depends on the input and the output. If the signal is bounded, then the estimate remains bounded; even if the coordinate transformations are only approximate.

The problem with this approach is that there are very few systems Eqs. (15) - (17) that can be transformed into Eqs. (106) - (107). The functions Eqs. (104) - (105) must satisfy a first order system of partial differential equations and must be at least local diffeomorphisms. To be solvable, the system of partial differential equations must satisfy integrability conditions that are quite restrictive. Also, the system Eqs. (15) - (17) needs to be linearly observable at x^0.

This last condition can sometimes be avoided by allowing Eqs. (104) - (105) to be semi-diffeomorphisms: that is, smooth functions with continuous inverses.

To get around this problem, one can broaden the class of systems Eqs. (106) - (107) to those in so-called state affine form

$$\dot{z} = A(u,y)z + \alpha(u,y) \tag{116}$$

$$w = C(u,y)z + \beta(u,y). \tag{117}$$

For these systems one can use a Kalman filtering approach,

$$\dot{\hat{z}} = A(u,y)\hat{z} + \alpha(u,y) + L(t)(w - \hat{w}) \tag{118}$$

$$\hat{w} = C(u,y)\hat{x} + \beta(u,y) \tag{119}$$

$$\dot{P}(t) = A(t)P(t) + P(t)A'(t) + Q(t) - P(t)C'(t)R^{-1}(t)C(t)P(t) \tag{120}$$

$$Q(t) = G(t)G'(t) \tag{121}$$

$$R(t) = J(t)J'(t) \tag{122}$$

$$L(t) = P(t)C'(t)R^{-1}(t). \tag{123}$$

The partial differential equations for these state affine transformations are more complicated, but the integrability conditions are less stringent. These techniques have been successful employed for low dimensional problems but the calculations grow in complexity as the dimensions increase.

If one assumes that the input $u(t)$ and/or the output $y(t)$ is differentiable, then one can allow A, C, α, β to depend on their derivatives. The partial differential equations for θ, γ get even more complicated, but the integrability conditions become even less stringent. Of course, this approach is not advisable if there is noise present.

Recently, a simpler approach has been introduced for real analytic systems without inputs,

$$\dot{x} = f(x) = Fx + f^{[2]}(x) + f^{[3]}(x) + \ldots \tag{124}$$

$$y = h(x) = Hx + h^{[2]}(x) + h^{[3]}(x) + \ldots \tag{125}$$

$$x(0) \approx x^0 = 0 \tag{126}$$

where $f^{[d]}(x), h^{[d]}(x)$ denote the degree d terms in the Taylor series expansion of $f(x)$, $h(x)$. One seeks a local diffeomorphism $z = \theta(x)$ and an output injection $\beta(y)$ that transforms Eqs. (30) - (32) into

$$\dot{z} = Az - \beta(y) \tag{127}$$

where A is Hurwitz. If these can be found, then the observer

$$\dot{\hat{z}} = A\hat{z} - \beta(y) \tag{128}$$

$$\hat{x} = \theta^{-1}(\hat{z}) \tag{129}$$

has exponentially stable linear error dynamics, with linear error dynamics

$$\dot{\tilde{z}} = A\tilde{z}. \tag{130}$$

Necessary conditions for the existence of such a local diffeomorphism and output injection are that H, F be an observable pair, and that there are no resonances between the eigenvalues of F and those of A. Suppose the spectrum of F is $(\lambda_1, \ldots, \lambda_n)$ and the spectrum of A is $(\mu_1, \ldots, \mu_n)$. A resonance occurs when there is a nonzero vector $k = (k_1, \ldots, k_n)$ of nonnegative integers and some j, $1 \le j \le n$, such that

$$\sum_{i=1}^{n} k_i \lambda_i = \mu_j. \tag{131}$$

As originally proposed, a sufficient condition for the existence of such a local diffeomorphism and output injection was that the spectrum of F be either in the open left half plane or the open right half plane. This ruled out many interesting cases. The former implies that the system is exponentially stable, and so $\hat{x}(t) = 0$ is a convergent observer. The latter implies that the system is exponentially unstable, and so a local observer is not of much use. But, recently, a much weaker sufficient condition has been found, and the spectrum of F can be arbitrary. This method has not been extended to systems with inputs as yet.

All of the above approaches lend themselves to power series methods for finding the desired transformations term by term up to any degree of accuracy. For brevity, we illustrate this for only the last method. The Hurwitz matrix A, the local diffeomorphism $z = \theta(x)$ and the output injection $\beta(y)$ must satisfy the first order partial differential equation

$$\frac{\partial\theta}{\partial x}(x) f(x) = A\theta(x) - \beta(y). \tag{132}$$

We expand in a power series assuming without loss of generality that the linear part of $\theta(x)$ is the identity,

$$\theta(x) = x + \theta^{[2]}(x) + \theta^{[3]}(x) + \ldots \tag{133}$$

$$\beta(y) = BHx + \beta^{[2]}(y) + \beta^{[3]}(y) + \ldots \tag{134}$$

The linear part of Eq. (132) is

$$A = F + BH. \tag{135}$$

and if H, F is an observable pair the spectrum of A can be set arbitrarily by choice of B.

The quadratic part of Eq. (132) is

$$\frac{\partial\theta^{[2]}}{\partial x}(x) Fx - A\theta^{[2]}(x) = -f^{[2]}(x) - \beta^{[2]}(y). \tag{136}$$

If there is no resonance, then this equation has a unique solution, $\theta^{[2]}(x)$, for any right side equation. The unknown $\beta^{[2]}(y)$ can be chosen to keep $\theta^{[2]}(x)$ close to 0 so $\theta(x)$ remains a diffeomorphism over a wide region.

The degree d part of Eq. (132) is

$$\frac{\partial\theta^{[d]}}{\partial x}(x)Fx - A\theta^{[d]}(x) = -\sum_{i=1}^{d-1}\frac{\partial\theta^{[i]}}{\partial x}(x)f^{[d-i]}(x) - \beta^{[d]}(y). \tag{137}$$

If there is no resonance, then this equation has a unique solution, $\theta^{[d]}(x)$, for any right side. Again the unknown $\beta^{[d]}(y)$ can be chosen to keep $\theta^{[d]}(x)$ close to 0 so $\theta(x)$ remains a diffeomorphism over a wide region.

If we stop at degree d

$$\theta(x) = x + \theta^{[2]}(x) + \ldots + \theta^{[d]}(x) \tag{138}$$

$$\beta(y) = BHx + \beta^{[2]}(y) + \ldots + \beta^{[d]}(y) \tag{139}$$

the observer Eq. (128) - (129) has approximately linear error dynamics

$$\dot{\tilde{z}} = A\tilde{z} + O(z)^{d+1}. \tag{140}$$

5. High Gain Observers

A high gain observer can be constructed for a system in uniformly observable form Eqs. (48) - (51) which satisfies certain Lipschitz conditions. For scalar output systems, it takes the form

$$\hat{y} = \hat{\xi}_1 + g_0(u) \tag{141}$$

$$\begin{aligned}\dot{\hat{\xi}}_1 &= \hat{\xi}_2 + g_1(\hat{\xi}_1, u) + L_1(y - \hat{y})\\ &\vdots\end{aligned} \tag{142}$$

$$\begin{aligned}\dot{\hat{\xi}}_i &= \hat{\xi}_{i+1} + g_2(\hat{\xi}_1, \ldots, \hat{\xi}_i, u) + L_i(y - \hat{y})\\ &\vdots\end{aligned} \tag{143}$$

$$\dot{\hat{\xi}}_n = \bar{f}(\hat{\xi}) + g_n(\hat{\xi}_1, \ldots, \hat{\xi}_n, u) + L_n(y - \hat{y}). \tag{144}$$

The gain is chosen as

$$L_i = \binom{n}{i}\theta^i \tag{145}$$

where θ is a constant that depends on the Lipschitz constant of the system. It has been proven that if θ is sufficiently large, then the observer estimate is globally asymptotically convergent to the system state. Of course, this assumes that there is no noise in the dynamics nor the observations. High gain observers converge to differentiators as the gain increases and therefore become very sensitive to noise. If one uses a high gain observer, then even small noise can lead to substantial degradation in the performance of the observer. In most of the examples in the literature, the gain parameter θ is chosen small without regard to global convergence. The resulting observers seem to perform well in simulations, but this is probably due to linear rather than nonlinear effects. There is no theoretical explanation of why this should happen. When the gain is set too low, there is no guarantee of convergence even if there is no noise.

6. Nonlinear Filtering

The stochastic approach to nonlinear estimation is to assume that the nonlinear system is described by Ito's stochastic differential equations

$$dx = f(x,u)\,dt + g(x)\,dw \tag{146}$$

$$dy = h(x,u)\,dt + dv \tag{147}$$

$$x(0) = x^0 \tag{148}$$

corrupted by observation and driving noises. Here $w(t)$ and $v(t)$ are standard Wiener processes and the initial condition x^0 is assumed to have density $p^0(x)$. It can be shown that the unnormalized density $q(x, t)$ of $x(t)$ conditioned on the past controls and observations satisfies a stochastic partial differential equation called the Zakai equation,

$$dq(x,t) = -\sum_{i=1}^{n} \frac{\partial}{\partial x_i}\left(f_i(x,u)\,q(x,t)\right) + \sum_{i,j=1}^{n} \frac{\partial^2}{\partial x_i \partial x_j}\left(a_{ij}(x)\,q(x,t)\right) + q(x,t)\sum_{i=1}^{p} h_i(x,u)\,dy_i \tag{149}$$

$$q(x,0) = p^0(x) \tag{150}$$

where

$$a_{ij}(x) = \sum_{l=1}^{n} g_{il}(x)\,g_{lj}(x). \tag{151}$$

The conditional density $p(x, t)$ is obtained by normalizing $q(x, t)$

$$p(x,t) = \frac{q(x,t)}{\int q(\xi,t)\,d\xi} \tag{152}$$

and the conditional mean is

$$\hat{x}(t) = \frac{\int \xi q(\xi,t)\,d\xi}{\int q(\xi,t)\,d\xi}. \tag{153}$$

One can also take the conditional mode as the estimate,

$$\hat{x}(t) = \arg\max_x q(x,t). \tag{154}$$

The Zakai equation is a stochastic parabolic partial differential equation in the Ito sense and its numerical integration is quite delicate. It is theoretically quite important but of limited practical use. It is driven by the observation process y(t), so it must be solved in real time. This is generally not possible when the state dimension is greater than 1 and is difficult even when it is 1 because the accuracy of the solution is dictated by the step sizes in x and t. Hence, the numerical integration of the Zakai equation is generally not a practical approach to estimating the state of a nonlinear system.

Notice when thought of as an observer, the Zakai Eq. (149) and state estimate Eqs. (153) or (154) is an expanded order observer Eqs. (18) - (19) with the state $z(t)$ being $q(\cdot, t)$. The state of the observer is infinite dimensional.

7. Minimum Energy and H^∞ Estimation

An alternative to the stochastic approach is to assume that the noises $w(t)$, $v(t)$ are not stochastic but unknown L_2 functions corrupting the system,

$$\dot{x} = f(x,u) + g(x)w \tag{155}$$

$$y = h(x,u) + v \tag{156}$$

$$x(0) = \hat{x}^0 + \tilde{x}^0 \tag{157}$$

where $\tilde{x}^0$ is an unknown error in the initial condition. We seek the initial state error $\tilde{x}^0$ and noises

$$\{(w(\tau), v(\tau)) : 0 \le \tau \le t\} \tag{158}$$

of “minimum energy”

$$\frac{1}{2}|\tilde{x}^0|^2 + \frac{1}{2}\int_0^t |w(\tau)|^2 + |v(\tau)|^2 \, d\tau \tag{159}$$

which are consistent with the initial estimate $\hat{x}^0$ and the past controls and observations

$$\{(u(\tau), y(\tau)) : 0 \le \tau \le t\} \tag{160}$$

and the system Eqs. (155) - (157). This is an optimal control problem. We consider the optimal cost

$$Q(x,t) = \inf_{\tilde{x}^0, w(\cdot)} \left(\frac{1}{2}|\tilde{x}^0|^2 + \frac{1}{2}\int_0^t |w(\tau)|^2 + |y(\tau) - h(x(\tau), u(\tau))|^2 \, d\tau \right) \tag{161}$$

subject to Eqs. (155) - (157) and $x(t) = x$. The optimal estimate is then given by

$$\hat{x}(t) = \arg\min_x Q(x,t). \tag{162}$$

The dynamic programming approach yields a partial differential equation for $Q(x, t)$,

$$0 = \frac{\partial Q}{\partial t}(x,t) + \sum_{i=1}^{n} \frac{\partial Q}{\partial x_i}(x,t) f_i(x,u) + \frac{1}{2}\sum_{i,j=1}^{n} \frac{\partial Q}{\partial x_i}(x,t) a_{ij}(x) \frac{\partial Q}{\partial x_j}(x,t) - \frac{1}{2}|y(t) - h(x,u)|^2 \tag{163}$$

It is of the Hamilton-Jacobian type, first order, nonlinear and again driven by the observations. As with the Zakai equation, it is very difficult to compute an accurate solution in real time. Moreover, it may not admit a smooth solution so the Eq. (163) must be interpreted in the viscosity sense. This is an infinite dimensional observer with state $Q(\cdot,t)$ evolving according to Eq. (163) with the state estimate given by Eq. (162). Hence it is of limited practical use. In some sense, it is the larger deviations limit of the Zakai observer.

The H^∞ approach to nonlinear estimation is an extension of this. If there were no noise and the initial conditions were known exactly, then the estimation of $x(t)$ would be relatively easy, just integrate the differential equation. If there are disturbances, i.e., driving and observation noises and an unknown initial condition, we would like the gain from these to the estimation error to be as small as possible. Hence, our estimate $\hat{z}(\tau)$ should satisfy the

$$\int_0^t |x(\tau) - \hat{x}(\tau)|^2 \le \gamma^2 \left(|\tilde{x}^0|^2 + \int_0^t |w(\tau)|^2 + |v(\tau)|^2 \, d\tau \right) \tag{164}$$

for as small γ as possible. Finding the minimal γ is a difficult problem to solve directly, so instead, we choose a gain level γ, and see if we can construct an observer that achieves it. If this is possible, then we try to do it for a smaller γ, if not we try a larger γ, etc.

We seek an estimator such that if

$$Q(x,t)=\inf_{\tilde{x}^0,w(\cdot)}\left(\frac{\gamma^2}{2}\left|\tilde{x}^0\right|^2+\frac{\gamma^2}{2}\int_0^t\left|w(\tau)\right|^2+\left|y(\tau)-h(x(\tau),u(\tau))\right|^2d\tau-\frac{1}{2}\int_0^t\left|x(\tau)-\hat{x}(\tau)\right|^2d\tau\right) \tag{165}$$

subject to Eqs. (155) - (157) and $x(t) = x$ then $Q(x, t) \geq 0$. If such a $Q(x, t)$ exists then our estimate is given by Eq. (162) and Q satisfies in the viscosity sense

$$\begin{aligned}0=&\frac{\partial Q}{\partial t}(x,t)+\sum_{i=1}^{n}\frac{\partial Q}{\partial x_i}(x,t)f_i(x,u)+\frac{1}{2\gamma^2}\sum_{i,j=1}^{n}\frac{\partial Q}{\partial x_i}(x,t)a_{ij}(x)\frac{\partial Q}{\partial x_j}(x,t)\\&-\frac{\gamma^2}{2}\left|y(t)-h(x,u)\right|^2+\frac{1}{2}\left|x-\hat{z}(\tau)\right|^2.\end{aligned} \tag{166}$$

Again this is an infinite dimensional observer and hence of limited practical use. There is a stochastic version of the H^∞ observer, but it too is infinite dimensional.

From a theoretical point of view there are two additional problems with the minimum energy and H^∞ observers. The first is that the criteria weights the distant past as much as the present, and the other is that $Q(x, t)$ tends to grow. These don't arise in the Zakai estimator because the second order terms in the Zakai partial differential equation tend to diffuse away the past, and the solution can always be renormalized Eq. (152), and the partial differential equation restarted. One way around these difficulties is to introduce a forgetting factor $\alpha \geq 0$. For the H^∞ observer it takes the form

$$Q(x,t)=\inf_{\tilde{x}^0,w(\cdot)}\left(\begin{aligned}&\frac{\gamma^2}{2}e^{-\alpha t}\left|\tilde{x}^0\right|^2+\frac{\gamma^2}{2}\int_0^t e^{-\alpha(t-\tau)}\left(\left|w(\tau)\right|^2+\left|y(\tau)-h(x(\tau),u(\tau))\right|^2\right)d\tau\\&-\frac{1}{2}\int_0^t e^{-\alpha(t-\tau)}\left|x(\tau)-\hat{x}(\tau)\right|^2d\tau\end{aligned}\right) \tag{167}$$

and the partial differential equation becomes

$$\begin{aligned}0=&\alpha Q(x,t)+\frac{\partial Q}{\partial t}(x,t)+\sum_{i=1}^{n}\frac{\partial Q}{\partial x_i}(x,t)f_i(x,u)+\frac{1}{2\gamma^2}\sum_{i,j=1}^{n}\frac{\partial Q}{\partial x_i}(x,t)a_{ij}(x)\frac{\partial Q}{\partial x_j}(x,t)\\&-\frac{\gamma^2}{2}\left|y(t)-h(x,u)\right|^2+\frac{1}{2}\left|x-\hat{z}(\tau)\right|^2.\end{aligned} \tag{168}$$

In H^∞ estimation one tries to find an observer that minimizes the induced L_2 gain from disturbances to estimation error. An alternate is to use a different L_p norm on the disturbances and errors. Probably the most useful norm from a practical point view is the L_∞ norm, so that bounded disturbances produce bounded estimation errors. Unfortunately, the mathematics is not so nice and the resulting partial differential equation is even more difficult to state and solve. And it is still an infinite dimensional observer.

As the reader may surmise, it is not hard, theoretically to construct infinite dimensional observers for finite dimensional nonlinear systems. In fact every nonlinear observer has an infinite dimensional realization as we shall show in a moment. The hard part is finding a finite dimensional realization. Considerable effort has been expended on this topic for the Zakai observer with only limited success.

Recall that an observer is a causal functional

$$\begin{bmatrix} \hat{x}^0 \\ u(\tau) \\ y(\tau) \end{bmatrix} \mapsto \hat{x}(t), \quad 0 \le \tau \le t \tag{169}$$

from the initial state estimate, $\hat{x}^0$ and the past control and observation, $u(\tau), y(\tau), 0 \le \tau \le t$, to the current state estimate, $\hat{x}(t)$ such that $\left|x(t) - \hat{x}(t)\right| \to 0$ as $t \to \infty$.

For such an observer $\hat{x}(t)$, there exists a function $Q(x, t)$ which is a causal functional of the initial state estimate $\hat{x}^0$ and the past controls and observations, $\{u(\tau), y(\tau), 0 \le \tau \le t\}$ such that $Q(x, t)$ has a minimum at $x = \hat{x}(t)$ and along any state trajectory $x(t)$, input $u(t)$, observation $y(t)$ and noises $w(t)$, $v(t)$ are consistent with the system

$$Q(x(t_2), t_2) \le Q(x(t_1), t_1) + \int_{t_1}^{t_2} \left|w(\tau)\right|^2 + \left|v(\tau)\right|^2 d\tau \tag{170}$$

Define $Q(x, t)$ as

$$Q(x,t) = \inf\left\{ \left|z(\tau_1) - \hat{x}(\tau_1)\right|^2 + \int_{\tau_1}^{t} \left|w(\tau)\right|^2 + \left|v(\tau)\right|^2 d\tau : 0 \le \tau_1 \le t \right\} \tag{171}$$

where the infimum is over all $z(\tau)$, $w(\tau)$, $v(\tau)$ satisfying

$$\frac{d}{d\tau} z(\tau) = f(z(\tau), u(\tau)) + g(z(\tau)) w(\tau) \tag{172}$$

$$y(\tau) = h(z(\tau)) + v(\tau) \tag{173}$$

$$z(t) = x. \tag{174}$$

and $u(\tau)$, $y(\tau)$ are the control and observation. This can be viewed as an infinite dimensional observer with state $Q(\cdot, t)$ and output Eq. (162).

8. Multiple Extended Kalman Filters

There is a hybrid approach to the nonlinear estimation. One can compute the solution of one of the partial differential equations described previously on a very coarse spatial and temporal grid. Of course, this will not lead to a very accurate estimate. But at each local maximum of the coarse solution to the Zakai equation one can initiate an extended Kalman filter, and this should quickly evolve to the true mode if it is nearby. For the minimum energy and H^∞ partial differential equations, one initiates the extended Kalman filter at the local minima.

The relative accuracy of the different Kalman filters can be assessed by how well they explain the most recent observations. If the observation is $y(t)$ and an estimate is $\hat{x}(t)$, then a measure of its recent accuracy is $q(t)$ where

$$\dot{q}(t) = -\alpha q(t) + \frac{1}{2}\left|y(t) - h(\hat{x}(t), u(t))\right|^2 \tag{175}$$

where $\alpha \ge 0$ is a forgetting factor. The extended Kalman filter estimate $\hat{x}(t)$ with the smallest current value of $q(t)$ is taken as true. This leads to extremely fast transitions between estimates as different $q(t)$ cross. The bookkeeping of creating and merging extended Kalman filters is substantial, but the computation of their updates is relatively trivial.

One can eliminate the step of solving the partial differential equation and instead continuously initiate extended Kalman filters at likely spots in the state space. The practicality of these approaches has yet to be demonstrated.

9. Conclusion

We have surveyed some but not all of the ways of constructing an observer for a nonlinear system. The high gain observer is a theoretical finite dimensional solution to a broad class of noise free problems, but performs poorly when noise is present. The Zakai, minimum energy, and H^∞ observers are theoretically infinite dimensional solutions to broad classes of noisy problems. But none of these are practically implementable. The linearization techniques give local and sometimes only approximate solutions for narrower classes of problems, and sometimes they are very hard to implement.

The extended Kalman filter is probably still the most robust and practical approach for most problems. If there are substantial nonlinearities, e.g., multiple stable equilibria and/or stable limit cycles, then the use of multiple extended Kalman filters is probably the preferred approach.

All these methods rely on the linear observability of the system to insure convergence. Other than the method of semi-diffeomorphisms, little is known about constructing observers around an equilibrium state where the system is not linearly observable. In summary, much is known about state reconstruction by observers but much remains to be done to find implementable solutions for broad classes of nonlinear systems.

Glossary

System: A system is a set of components, physical or otherwise, which are connected in such a manner as to form and act as an entire unit. Frequently, we use the term to describe a mathematical model of a physical system. A system has state variables which are its internal memory, input variables by which the external world affects the system, and output variables by which it affects the external world. The input variables that can be chosen or measured by the operator are called controls. Other input variables are frequently viewed as noise. The output variables that can be measured are called measurements.

Observer: A system which accepts as inputs the measured inputs and outputs of another system and returns an estimate of the other system's state. The observer is said to be of full, reduced, or expanded order when its state is of the same, smaller, or larger dimension than the other system.

Autonomous system: A system whose behavior is independent of time.

Causality: A property of systems, future inputs cannot affect past outputs.

Controls: Those inputs that can be chosen by the operator.

Detectable system: A system where the measurements eventually determine the state.

Error linearization: A technique for constructing an observer which results in linear error dynamics in some coordinate system

Extended Kalman filter: An extension of the Kalman filter to nonlinear systems.

***H^∞* estimation:** An approach to estimation that attempts to minimize the gain between the size of noise and the size of the estimation error.

Hurwitz matrix: A square matrix with all its eigenvalues in the open left half plane.

Kalman filter: A linear observer constructed using stochastic methods.

Linear approximating system: The linear part of a nonlinear system around an equilibrium point.

Linear system: A system described by linear equations.

Linear observer: A linear system that is an observer for another system.

Luenberger observer: A reduced order linear observer for a linear system.

Minimum energy estimation: An approach to estimation that minimizes the size of the noise necessary to produce the observations

Nonautonomous system: A system whose behavior varies with time.

Nonlinear filter: An infinite dimensional, nonlinear observer constructed using stochastic methods described by a stochastic partial differential equation, called the Zakai equation.

Nonlinear system: A system described by nonlinear equations.

Nonlinear observer: A nonlinear system that is an observer for another nonlinear system.

Observable form: A mathematical way of representing a system using the measurement function and its derivatives as state coordinates.

Observable system: A system where the measurements and their time derivatives determine the state.

Bibliography

Bestle D. and Zeitz M. (1983). Canonical form observer design for non-linear time-variable systems. *Internat. J. Control* **38**, 419-431. [One of the first papers to treat nonlinear observer design by error linearization.]

Davis M.H.A. and Marcus S.I. (1981). An introduction to nonlinear filtering. In M. Hazewinkel, J.C. Willems, eds., *Stochastic Systems: The Mathematics of Filtering and Identification and Applications*, pp. 53-76. Dordrecht: D. Reidel Publishing. [This paper contains a rigorous derivation of the Zakai equation for the evolution of the conditional density of a nonlinear filter.]

Gauthier J.P., Hammouri H. and Othman S. (1992). A simple observer for nonlinear systems with applications to bioreactors. *IEEE Trans. Auto. Contr.* **37**, 875-880. [This paper introduces the high gain observer and proves its convergence if the gain is high enough and there is no noise.]

Gelb A. (1989). *Applied Optimal Estimation*. Cambridge, MA: MIT Press. [An excellent introduction to Kalman filtering, extended Kalman filtering and their extensions.]

Isidori A. (1985). *Nonlinear Control Systems*. New York: Springer-Verlag. [A basic reference on nonlinear control systems.]

Kalman R.E. (1960). A new approach to linear filtering and prediction problems. *Trans. of ASME*, *Part D*, *J. of Basic Engineering* **82**, 35-45. [The introduction of the Kalman filter.]

Kalman R.E. and Bucy R.S. (1961). New results in linear filtering and prediction theory. *Trans. of ASME*, *Part D*, *J. of Basic Engineering* **83**, 95-108. [More on the Kalman filter.]

Kazantzis N. and Kravaris C. (1998). Nonlinear observer design using Lyapunov's auxiliary theorem. *Systems and Control Letters* **34**, 241-247. [This paper introduces an improved method of nonlinear observer design via error linearization that is applicable to a wider class of nonlinear systems.]

Krener A.J. (1997). Necessary and sufficient conditions for nonlinear worst case (H-infinity) control and estimation. *Journal of Mathematical Systems*, *Estimation*, *and Control* **7**, 81-106. [An introduction to H-infinity observer design.]

Krener A.J. and Duarte A. (1996). A hybrid computational approach to nonlinear estimation. In *Proc. of 35th Conference on Decision and Control*, pp. 1815-1819. Kobe, Japan. [This paper presents the multiple extended Kalman filtering method.]

Krener A.J. and Isidori A. (1983). Linearization by output injection and nonlinear observers. *Systems and Control Letters* **3**, 47-52. [One of the first papers to treat nonlinear observer design by error linearization.]

Krener A.J. and Respondek W. (1985). Nonlinear observers with linearizable error dynamics. *SIAM J. Control and Optimization* 23, 197-216. [This paper extends the error linearization method to systems with vector observations].

Luenberger D.G. (1964). Observing the state of a linear system. *IEEE Trans. on Military Electronics* **8**, 74-80. [This paper presents a reduced order linear observer for a linear system.]

Misawa E.A. and Hedrick J.K. (1989). Nonlinear observers, a state of the art survey. *Trans. of ASME*, *J. of Dynamic Systems*, *Measurement and Control* **111**, 344-352. [A survey of observer design methods.]

Muske K.R. and Edgar T.F. (1997). Nonlinear state estimation. In M.A. Henson, D.E. Seborg, eds., *Nonlinear Process Control*, Upper Saddle River, NJ: Prentice Hall. [Another survey of observer design methods.]

Nijmeijer H. and Fossen T. (1999). *New Directions in Nonlinear Observer Design*. New York: Springer-Verlag. [The proceedings of a conference on nonlinear observer design.]

Nijmeijer H. and van der Schaft A. (1990). *Nonlinear Dynamical Control Systems*, 467pp. New York: Springer-Verlag. [Another basic reference on nonlinear control systems.]

Xia X. and Zeitz M. (1997). On nonlinear continuous observers. *Internat. J. Control* **66**, 943-954. [Introduces the use of semi-diffeomorphisms to observer design.]