Synthesis of Optimal Nonlinear Observers

Scott A. Bortoff **Alan F. Lynch** *
Department of Electrical and Computer Engineering
University of Toronto
Toronto, Ontario M5S 1A4 Canada
bortoff@control.toronto.edu alanl@control.toronto.edu

September 7, 1995

Abstract

After briefly reviewing the well-known necessary and sufficient existence conditions for a nonlinear observer with linear error dynamics (NOLED), we show how spline functions can be used to solve the generalized characteristic equation (GCE). The splines transform the GCE, which is a high-order partial differential equation, into a well-conditioned algebraic one. This equation can be solved even if the original GCE does not posses a solution, i.e., when the non-generic involutivity condition is not satisfied. In this case, we compute an observer with approximately linear error dynamics. Moreover, our approximation is optimal in a well-defined sense. Two examples, one of which is not integrable, illustrate observer construction and performance.

1 Introduction

The problem of constructing an approximate feedback linearizing controller for a nonlinear system has attracted the attention of a number of authors in recent years, e.g., [1]-[3]. The results are important because they do not require a non-generic involutivity condition to be satisfied. Thus, approximate feedback linearization can be applied to a large class of nonlinear systems. We emphasize that these results usually provide a means of *computing* the required coordinate transformation and nonlinear control, sometimes in an optimal sense, e.g., [4]; they are more than existence conditions. However, dual results, which would weaken the involutivity condition required for the nonlinear observer with linear error dynamics (denoted NOLED for the remainder of this paper), and would provide a means of computing an observer with approximately linear error dynamics for a generic class of nonlinear systems, have lagged the controller results, with some exceptions, e.g., [5].

In this paper, we represent the NOLED's output injection vector and part of its coordinate transformation using *spline functions*. This transforms a high-order partial differential equation known as the generalized characteristic equation (GCE) [6] into an algebraic equation, whose solution is easily computed regardless of plant complexity. But our result is more than a numerical method of solution. It can be generalized to non-integrable systems, i.e., those for which the necessary involutivity condition fails. In this case, we compute an observer with approximately linear error dynamics. This observer exists *and can be computed* for a generic class of nonlinear systems. Further, because our synthesis allows for the minimization of the high-order terms that perturb the exponentially stable error dynamics, we call it "optimal." In short, we have generalized the well-know Extended Luenberger Observer, and "dualized" the results of [1].

This paper is organized as follows. In Section 2, we provide some background in NOLED synthesis for single-input single-output systems. Much of the development follows [6] and [7]. In Section 3, we present our new synthesis method for the NOLED, for both the integrable and non-integrable cases. Section 4 provides two numerical examples that illustrate observer performance. Conclusions are drawn in Section 5, along with some indications of future research.

2 Background

We consider the synthesis of nonlinear observers for single-input, single-output systems of the form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u) \tag{1}$$
$$y = h(\boldsymbol{x}), \tag{2}$$

where $\boldsymbol{x} \in \mathcal{R}^n$ is the state, $u \in \mathcal{R}$ is the control input, $y \in \mathcal{R}$ is the measured output, $\boldsymbol{f}$ is a C^∞-vector

*This research supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) under grant OGP0138423.

field parameterized by u, h is a C^∞-output function, $\boldsymbol{f}(0,0)=0$, and $h(0)=0$. We begin with

Assumption A1. The input $u(t)$ is piecewise constant, i.e., $u(t)=\bar{u}_k$ for $t \in [kT,\,(k+1)T)$, where k is a positive integer, and T is the fixed sample period, and $\bar{u}_k \in [a,b] \subset \mathcal{R}$.

This assumption allows us to extend existing results for systems without inputs: we simply consider $\boldsymbol{f}$ to be parameterized by a constant control on each time interval. Our synthesis will not depend upon time derivatives of u, unlike [6], [8] and [9]. We justify this assumption by the fact that u is piecewise constant when implemented using digital hardware, and remark that the sample intervals need not be uniform; this merely simplifies our notation.

Now, if we denote $\bar{\boldsymbol{f}} := \boldsymbol{f}(\boldsymbol{x},\,\bar{u})$, where $\bar{u} \in [a,b]$ is constant, then necessary and sufficient conditions for the existence of the NOLED are well-known and can be found in [7], and [10]-[13]:

Condition C1 (Observability). The codistribution

$$\Omega := \text{span}\,\{dh,\, d\mathcal{L}_{\bar{f}}h, \ldots,\, d\mathcal{L}_{\bar{f}}^{n-1}h\} \qquad (3)$$

(where d is the exterior derivative) must be full-rank at $\boldsymbol{x}=0$, for every $\bar{u} \in [a,b]$; and

Condition C2 (Integrability). The n vector fields $\boldsymbol{\pi}_i(\boldsymbol{x},\,\bar{u})$, defined by

$$\begin{aligned}
\langle d\mathcal{L}_{\bar{f}}^{i}h\,;\,\boldsymbol{\pi}_1\rangle &= 0 \quad \text{for} \quad i=0,1,\ldots,n-2, && (4)\\
\langle d\mathcal{L}_{\bar{f}}^{n-1}h\,;\,\boldsymbol{\pi}_1\rangle &= 1 && (5)\\
\boldsymbol{\pi}_j &= -\text{ad}_{\bar{f}}^{j-1}\boldsymbol{\pi}_1 \quad \text{for} \quad j=2,\ldots,n && (6)
\end{aligned}$$

must satisfy $[\boldsymbol{\pi}_r\,,\,\boldsymbol{\pi}_s]=0$ for all $\bar{u}\in[a,b]$ and for all $1 \le r,\, s \le n$, in a neighborhood $\boldsymbol{\Gamma}$ of $\boldsymbol{x}=0$.

If **A1**, **C1** and **C2** are satisfied, then a coordinate transformation $\boldsymbol{z} := \boldsymbol{\Phi}(\boldsymbol{x},\,\bar{u}_k)$, where $\boldsymbol{\Phi} : \boldsymbol{\Gamma}\times[a,b] \to \mathcal{R}^n$, exists such that (1)-(2) can be expressed in observer form

$$\begin{aligned}
\dot{\boldsymbol{z}} &= \boldsymbol{A}\boldsymbol{z} + \boldsymbol{a}(y,\,\bar{u}_k) && (7)\\
y &= \boldsymbol{C}\boldsymbol{z}, && (8)
\end{aligned}$$

for $t \in \big[kT,\,(k+1)T\big)$, where the pair $(\boldsymbol{A},\,\boldsymbol{C})$ is in observer canonical form, and $\boldsymbol{a}$ is the C^∞ output injection vector field. Synthesis of an observer for (7)-(8) follows naturally. Denoting $\hat{\boldsymbol{z}}$ as the estimate of the state $\boldsymbol{z}$, we have the observer with linear error dynamics, defined for $t \in \big[kT,\,(k+1)T\big)$:

$$\begin{aligned}
\dot{\hat{\boldsymbol{z}}} &= \boldsymbol{A}\hat{\boldsymbol{z}} + \boldsymbol{a}(y,\,\bar{u}_k) + \boldsymbol{G}(y-\hat{y}) && (9)\\
\hat{y} &= \boldsymbol{C}\hat{\boldsymbol{z}}. && (10)
\end{aligned}$$

Note that, because the input is piecewise constant, the right-hand side of (9)-(10) is only piecewise C^∞ ($\boldsymbol{a}$ is discontinuous at $t=kT$). However, the estimation error $\tilde{\boldsymbol{z}} := \boldsymbol{z}-\hat{\boldsymbol{z}}$ is governed by the C^∞ system

$$\dot{\tilde{\boldsymbol{z}}} = (\boldsymbol{A}-\boldsymbol{G}\boldsymbol{C})\tilde{\boldsymbol{z}}, \qquad (11)$$

whose spectrum can be arbitrarily assigned.

Of course, the purpose of the observer is to estimate $\boldsymbol{x}$, not $\boldsymbol{z}$. Although estimates can be generated via $\hat{\boldsymbol{x}} = \boldsymbol{T}(\hat{\boldsymbol{z}},\bar{u}_k) := \boldsymbol{\Phi}^{-1}(\hat{\boldsymbol{z}},\bar{u}_k)$, it is advantageous to implement (9)-(10) explicitly in the $\boldsymbol{x}$-coordinates:

$$\begin{aligned}
\dot{\hat{\boldsymbol{x}}} = \;& \boldsymbol{f}(\hat{\boldsymbol{x}},\,\bar{u}_k) + D\boldsymbol{T}(\boldsymbol{\Phi}(\hat{x},\bar{u}_k),\,\bar{u}_k)\,\cdot\\
& \big(\boldsymbol{a}(y,\,\bar{u}_k) - \boldsymbol{a}(\hat{y}\,,\bar{u}_k) + \boldsymbol{G}\cdot(y-\hat{y})\big). \qquad (12)
\end{aligned}$$

Here, $D\boldsymbol{T}(\hat{\boldsymbol{z}},\bar{u}_k)$ denotes the Jacobian of $\boldsymbol{T}(\hat{\boldsymbol{z}},\bar{u}_k)$ with respect to $\hat{\boldsymbol{z}}$, parameterized by $\bar{u}_k$. Representation (12) has the advantage that neither $\boldsymbol{\Phi}$ nor $\boldsymbol{\Phi}^{-1}$ appear. Only the Jacobian of $\boldsymbol{T}$, expressed in the $\boldsymbol{x}$-coordinates, and the vector field $\boldsymbol{a}$ need be known.

There are basically two approaches to synthesizing (12). The first involves computing $D\boldsymbol{T}$ and then integrating and inverting to produce $\boldsymbol{\Phi}$. Following [7], we would define the set

$$\boldsymbol{\Pi} := \{\boldsymbol{\pi}_1,\,\boldsymbol{\pi}_2,\ldots,\,\boldsymbol{\pi}_n\}, \qquad (13)$$

which can be written as a matrix, and where $\boldsymbol{\pi}_i$ are defined in (4)-(6). Then the following partial differential equation must be satisfied [7]:

$$D\boldsymbol{T}(\boldsymbol{z},\bar{u})|_{\boldsymbol{z}=\boldsymbol{\Phi}(\boldsymbol{x},\bar{u})} = \frac{\partial \boldsymbol{T}(\boldsymbol{z},\bar{u})}{\partial \boldsymbol{z}}|_{\boldsymbol{z}=\boldsymbol{\Phi}(\boldsymbol{x},\bar{u})} = \boldsymbol{\Pi}(\boldsymbol{x},\bar{u}), \qquad (14)$$

where the subscript k has been dropped. (The piecewise constant input will be denoted $\bar{u}$ for the rest of the paper.) Integrating (14) yields $\boldsymbol{T}$, which is then inverted to give $\boldsymbol{\Phi}$, and differentiated to produce $\boldsymbol{a}$.

A second method of synthesis involves computing $\boldsymbol{a}$ first, and then calculating $\boldsymbol{\Phi}$. Following [6] and [9], we rewrite the observer form (7) for convenience:

$$\begin{aligned}
\dot{z}_1 &= a_1(y,\bar{u}) && (15)\\
\dot{z}_2 &= z_1 + a_2(y,\bar{u}) && (16)\\
&\vdots\\
\dot{z}_n &= z_{n-1} + a_n(y,\bar{u}) && (17)\\
y &= z_n. && (18)
\end{aligned}$$

Differentiating $y=h(x)$ a total of n times gives:

$$y^{(n)} = \mathcal{L}_{\bar{f}}^{n}h = a_1 + \mathcal{L}_{\bar{f}}a_2 + \cdots + \mathcal{L}_{\bar{f}}^{n-1}a_n. \qquad (19)$$

Inspection of (19) reveals that the observer (15)-(18) exists if and only if the equation

$$\mathcal{L}_{\bar{f}}^{n}h - \mathcal{L}_{\bar{f}}^{n-1}a_n - \cdots - \mathcal{L}_{\bar{f}}^{1}a_2 - \mathcal{L}_{\bar{f}}^{0}a_1 = 0, \qquad (20)$$

called the **generalized characteristic equation** [6], has a solution for $a_1,\ldots,a_n$. Clearly, a solution to

(20) exists if the nonlinear observer (15)-(18) exists. Conversely, if the n functions $a_1, \ldots, a_n$ can be computed from (20), then the coordinate transformation $\boldsymbol{\Phi} = [\phi_1 \ \cdots \ \phi_n]^T$ follows:

$$\begin{aligned} \phi_{n-1}(\boldsymbol{x}, \bar{u}) &:= \mathcal{L}_{\bar{f}}\phi_n(\boldsymbol{x}, \bar{u}) - a_n(h(\boldsymbol{x}), \bar{u}) & (21) \\ &\vdots \\ \phi_1(\boldsymbol{x}, \bar{u}) &:= \mathcal{L}_{\bar{f}}\phi_2(\boldsymbol{x}, \bar{u}) - a_2(h(\boldsymbol{x}), \bar{u}). & (22) \end{aligned}$$

Note that (20) is a single $(n-1)^{th}$-order partial differential equation in the n unknowns a_k. It may appear to be more difficult to solve than the first-order PDE (14). However, by eliminating $\boldsymbol{\Phi}$ from appearing explicitly in (20), we have simplified the computations, because each of the n functions $a_k(y, \bar{u})$ depends on only *two* variables. Of course, closed-form solutions to (20) often will be difficult to find, depending on n and the complexity of $\boldsymbol{f}$ and h.

3 Observer Synthesis

Motivated by the computational challenges outlined in the previous section, we now develop a method to compute spline-function approximations of $\boldsymbol{a}$ and $\boldsymbol{\Phi}$. We begin by computing the Lie derivatives in (20):

$$\begin{aligned} \mathcal{L}^0_{\bar{f}} a_1 &= a_1 & (23) \\ \mathcal{L}^1_{\bar{f}} a_2 &= a_2' \cdot \mathcal{L}_{\bar{f}} h & (24) \\ \mathcal{L}^2_{\bar{f}} a_3 &= a_3' \cdot \mathcal{L}^2_{\bar{f}} h + a_3'' \cdot (\mathcal{L}_{\bar{f}} h)^2 & (25) \\ &\vdots \\ \mathcal{L}^{n-1}_{\bar{f}} a_n &= a_n' \cdot \mathcal{L}^{n-1}_{\bar{f}} h + \cdots + a_n^{(n-1)} \cdot (\mathcal{L}_{\bar{f}} h)^{n-1}, & (26) \end{aligned}$$

where $'$ denotes the derivative with respect to the first argument, e.g., $a_2'(y, \bar{u}) = \frac{\partial a_2(y, \bar{u})}{\partial y}$. From these expressions, we see that (20) is not sufficient (in itself) to compute $a_2, a_3, \ldots, a_n$ because only the first- and higher-order derivatives of these terms appear. However, as we shall show, (20) can be used to compute a_1, and $a_2', \ldots, a_n'$, and we may integrate the result along y, computing the "integration constant" from (21)-(22). Thus, we begin by defining $\alpha_k := a_k'$ for $k = 2, \ldots, n$, and $\alpha_1 := a_1$.

3.1 Integrable Case

We compute approximate solutions to (20) by representing α_k as a multivariable spline function, for $k = 1, 2, \ldots, n$:

$$\alpha_k(y, \bar{u}) := \sum_{i=1}^{N} \sum_{j=1}^{M} c_{ijk} \cdot B_i(y) \cdot B_j(\bar{u}), \qquad (27)$$

where $\{B_j(\bar{u}) : \bar{u} \in [a, b] \subset \mathcal{R}, \ j = 1, \ldots, M\}$ is a spline basis for $\bar{u}$,

$$\{B_i(y) : y \in [c, d] \subset \mathcal{R}, \ i = 1, 2, \ldots, N\} \qquad (28)$$

is a spline basis for y, and $c_{ijk} \in \mathcal{R}$ is one of $N \cdot M \cdot n$ spline coefficients. Note that we assume $y \in [c, d]$, which is the practical case of interest.

Substituting (27) into (20) and using the definition of α_k, we obtain a sum of terms, each of the form

$$\begin{aligned} \mathcal{L}^0_{\bar{f}} a_1 &= \sum_{i,j}^{N,M} c_{ij1} B_j(\bar{u}) \cdot \{B_i(y)\} & (29) \\ \mathcal{L}^1_{\bar{f}} a_2 &= \sum_{i,j}^{N,M} c_{ij2} B_j(\bar{u}) \cdot \{B_i(y) \cdot \mathcal{L}_{\bar{f}} h\} & (30) \\ &\vdots \\ \mathcal{L}^{n-1}_{\bar{f}} a_n &= \sum_{i,j}^{N,M} c_{ijn} B_j(\bar{u}) \cdot \Big\{B_i(y) \cdot \mathcal{L}^{n-1}_{\bar{f}} h \\ &\quad + \cdots + B_i^{(n-2)}(y) \cdot (\mathcal{L}_{\bar{f}} h)^n\Big\}. & (31) \end{aligned}$$

Define w_{ki} to be the term inside the $\{\cdot\}$ in $\mathcal{L}^{k-1}_{\bar{f}}$, e.g., $w_{3i}(\boldsymbol{x}, \bar{u}) = B_i(y) \cdot \mathcal{L}^2_{\bar{f}} h(\boldsymbol{x}) + B_i'(y) \cdot (\mathcal{L}_{\bar{f}} h(\boldsymbol{x}))^2$. Then, we may rewrite the GCE (20) compactly as:

$$\mathcal{L}^n_{\bar{f}} h(\boldsymbol{x}) - \sum_{i=1}^{N} \sum_{j=1}^{M} B_j(\bar{u}) \cdot \big(c_{ij1} w_{1\,i}(\boldsymbol{x}, \bar{u}) + c_{ij2} w_{2\,i}(\boldsymbol{x}, \bar{u}) + \cdots + c_{ijn} w_{n\,i}(\boldsymbol{x}, \bar{u})\big) = \epsilon(\boldsymbol{x}, \bar{u}), \quad (32)$$

where ϵ represents the spline-function approximation error. In general, this is nonzero, unless a_k is itself in the spline space. Further discussion of ϵ is beyond the scope of this paper, and we refer the reader to [14]. Suffice it to say that $\| \epsilon \|$ can be made small by increasing N or M, without significantly increasing the computational complexity of what follows. To simplify our presentation, we ignore ϵ.

Lemma 3.1 If Assumption **A1** and Condition **C1** are satisfied, then $\{w_{ki} : \ k = 1, \ldots, n, \ i = 1, \ldots, N\}$ is a linearly independent (LI) set of C^∞-functions in the neighborhood $\boldsymbol{\Gamma}$ of $\boldsymbol{x} = 0$, for all $u \in [a, b]$.

Proof 3.1 We only outline the proof. The functions w_{ik} are LI iff their gradients dw_{ki} are LI in the entire cotangent bundle $\mathrm{T}^*\boldsymbol{\Gamma} \times \boldsymbol{\Gamma}$. Using (29)-(31), dw_{ki} can be expressed as the sum of terms which include $B_i(y) \cdot d\mathcal{L}^k_f h$. One can form an $n \cdot N \times n$ matrix $D\boldsymbol{w}$ with dw_{ki} as the $k \cdot i^{th}$ row. Then, $D\boldsymbol{w}$ can be decomposed the product of a lower-triangular matrix $\boldsymbol{B}$ which has the spline basis vectors $B_i(y)$ on its (block) diagonal, and the observability matrix $\boldsymbol{\Omega}$, i.e., $D\boldsymbol{w} = \boldsymbol{B} \cdot \boldsymbol{\Omega}$. Since **C1** implies $\boldsymbol{\Omega}$ has rank n, and since the rows of $\boldsymbol{B}$ are LI, the rows of $D\boldsymbol{w}$ are also LI on $\mathrm{T}^*\boldsymbol{\Gamma} \times \boldsymbol{\Gamma}$. It follows that w_{ki} are all linearly independent. □

Defining the $N \cdot M \cdot n$–dimensional row vector

$$\boldsymbol{c} := [c_{111} \cdots c_{11n} \ c_{121} \cdots c_{12n} \ \cdots\cdots \ c_{NM1} \cdots c_{NMn}]$$

and the $N \cdot M \cdot n$–dimensional column vector

$$\boldsymbol{w}(\boldsymbol{x}, \bar{u}) := [B_1(\bar{u})w_{11}(\boldsymbol{x}, \bar{u}) \cdots B_1(\bar{u})w_{n1}(\boldsymbol{x}, \bar{u}) \cdots \\ \cdots B_M(\bar{u})w_{1N}(\boldsymbol{x}, \bar{u}) \cdots B_M(\bar{u})w_{nN}(\boldsymbol{x}, \bar{u})]^T,$$

(32) can be written compactly as

$$\mathcal{L}_f^n h = \boldsymbol{c} \cdot \boldsymbol{w}. \tag{33}$$

Lemma 3.1 then implies that the $n \cdot N \cdot M$ rows of $\boldsymbol{w}$ are linearly independent functions in $\boldsymbol{\Gamma}$, which allows us to solve (33) for $\boldsymbol{c}$. One way to do this is by evaluating $\boldsymbol{w}(\boldsymbol{x}, \bar{u})$ at $L = L_x \cdot L_u$ points that are properly selected from state and control space, where $L_x \geq n \cdot N$ and $L_u \geq M$. Let us denote this set of points, usually called a *mesh* in spline-interpolation parlance, as

$$\mathcal{S}_x := \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{L_x}\} \tag{34}$$
$$\mathcal{S}_u := \{\bar{u}_1, \ldots, \bar{u}_{L_u}\}. \tag{35}$$

"Proper selection" must ensure that the L vector samples of $\boldsymbol{w}$ span $\mathcal{R}^{n \cdot N \cdot M}$. But, because $\boldsymbol{w}$ is full row-rank, such a mesh is guaranteed to exist. The mesh should include points from state and control space near which $\boldsymbol{x}$ and u are expected to be.

With $\mathcal{S}_x$ and $\mathcal{S}_u$, we define the $(N \cdot M \cdot n) \times L$ matrix $\boldsymbol{W}$ as the "sampled" $\boldsymbol{w}$:

$$\boldsymbol{W} := [\boldsymbol{w}(\boldsymbol{x}_1, \bar{u}_1) \cdots \boldsymbol{w}(\boldsymbol{x}_{nL_x}, \bar{u}_1) \cdots \\ \cdots \boldsymbol{w}(\boldsymbol{x}_1, \bar{u}_{L_u}) \cdots \boldsymbol{w}(\boldsymbol{x}_{nL_x}, \bar{u}_{L_u})]. \tag{36}$$

Similarly, we define the L-dimensional row vector $\boldsymbol{\nu}$ as the left-hand side of (33) sampled at points in $\mathcal{S}_x$ and $\mathcal{S}_u$:

$$\boldsymbol{\nu} := [\mathcal{L}^n_{f(x,\bar{u}_1)} h(\boldsymbol{x}_1) \ \cdots \ \mathcal{L}^n_{f(x,\bar{u}_1)} h(\boldsymbol{x}_{nL_x}) \ \cdots \\ \cdots \ \mathcal{L}^n_{f(x,\bar{u}_{L_u})} h(\boldsymbol{x}_1) \ \cdots \ \mathcal{L}^n_{f(x,\bar{u}_{L_u})} h(\boldsymbol{x}_{nL_x})]. \tag{37}$$

Then we can express (33) as

$$\boldsymbol{\nu} = \boldsymbol{c} \cdot \boldsymbol{W} \tag{38}$$

which can be solved for the coefficient vector $\boldsymbol{c}$ by using the pseudoinverse of $\boldsymbol{W}$:

$$\boldsymbol{c} = \boldsymbol{\nu} \cdot \boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{W}^T)^{-1}. \tag{39}$$

With $\boldsymbol{c}$, the functions α_k are defined for $k = 1, \ldots, n$. The last $(n-1)$ of these must be integrated to give $\boldsymbol{a}$:

$$a_k(y, \bar{u}) = \int_{r=0}^{r=y} \alpha_k(r, \bar{u})\, dr + \gamma_k(\bar{u}) \tag{40}$$

for $k = 2, \ldots, n$, where $\gamma_k(\bar{u})$ is a "constant" of integration. Because α_k is a B-spline, the first term in (40) is easy to compute [14]; the only nontrivial step is determining γ_k. To do this, we first compute $\boldsymbol{z} = \boldsymbol{\Phi}(\boldsymbol{x}, \bar{u})$ using (21)-(22), noting ϕ_k will depend upon γ_{k+1}, for $k = 1, \ldots, n-1$. Differentiating ϕ_k along $\boldsymbol{f}$, and evaluating at the $\boldsymbol{x} = 0$, we obtain

$$\boldsymbol{a}_k(0, \bar{u}) = \gamma_k(\bar{u}) = \langle d\phi_k(0, \bar{u}) \ ; \ \boldsymbol{f}(0, \bar{u}) \rangle, \tag{41}$$

where $\langle \cdot \, ; \, \cdot \rangle$ denotes contraction between gradient and vector field. Now, because $d\phi_k$ does *not* depend upon γ_{k+1}, the right-hand side of (41) is a function of only $\bar{u}$, yielding γ_k.

Remark 3.1 We must consider the degree of differentiability when selecting a B-spline basis (28). Equation (20) contains derivatives up to order $n-1$. Thus, if B_i is C^{n-1}, then $a_n(y, \bar{u})$ is (at least) C^0. In this case, (21)-(22) ensure that $D\boldsymbol{\Phi}$ is also (at least) C^0, which is sufficient for (12) to possess a solution.

3.2 Non-integrable case

The approach introduced in Section 3.1 is easily generalized to the generic class of nonlinear systems that fail the integrability condition **C2**, but satisfy **A1** and **C1**. In this case, the GCE (20) does not have a solution; every choice of $a_k(y, \bar{u})$, $k = 1, 2, \ldots, n$ gives a nonzero function $\eta(\boldsymbol{x}, \bar{u})$:

$$\mathcal{L}_f^n h - \mathcal{L}_f^{n-1} a_n - \cdots - \mathcal{L}_f^1 a_2 - \mathcal{L}_f^0 a_1 = \eta. \tag{42}$$

We consider the minimization of $\| \ \eta(\boldsymbol{x}, \bar{u}) \ \|$ in the neighborhood $\boldsymbol{\Gamma}$ of $\boldsymbol{x} = 0$. Again, we use spline functions to represent $a_k(y, \bar{u})$. Because the development is exactly parallel to that in Section 3.1, we reuse all of the established notation.

In the non-integrable case, equation (33) becomes

$$\mathcal{L}_f^n h = \boldsymbol{c} \cdot \boldsymbol{w} + \eta, \tag{43}$$

and η is not negligible. To carry out the calculation of the spline coefficient vector $\boldsymbol{c}$, we again "sample" $\boldsymbol{w}$ and the left-hand side of (43) at a finite number of states and controls. These samples should be drawn from the region of state space and control space in which the observer is expected to function, because η is to be minimized over this finite set. Furthermore, the sampled vectors $\boldsymbol{w}$ must span $\mathcal{R}^{N \cdot M \cdot n}$, as in the exact case. The development is identical up to equation (38), which in the inexact case is now

$$\boldsymbol{\nu} = \boldsymbol{c} \cdot \boldsymbol{W} + \boldsymbol{\eta}. \tag{44}$$

Of course, there is no solution to (44), but an approximate one can be found which minimizes the Euclidean norm $\| \ \boldsymbol{\eta} \ \|$ on the data set $\mathcal{S}_x \times \mathcal{S}_u$:

$$\boldsymbol{c}^* = \boldsymbol{\nu} \cdot \boldsymbol{W}^T(\boldsymbol{W}\boldsymbol{W}^T)^{-1}. \tag{45}$$

Once the coefficient vector $\boldsymbol{c}^*$ is computed, the coordinate transformation $\boldsymbol{z} = \boldsymbol{\Phi}(\boldsymbol{x}, \bar{u})$ is computed using (21)-(22), just as in the exact case. This involves only differentiation. Next, we compute $D\boldsymbol{\Phi}$, and evaluate the right-hand side of (41), giving the constant of integration γ_k for $\boldsymbol{a}_k$, for $k = 2, \ldots, n$. Finally, we have to invert $D\boldsymbol{\Phi}$ for the observer realization (12).

It is of interest to express the original system (1)-(2) in the $\boldsymbol{z}$-coordinates when the exact observer does not exist. Using (21)-(22), we have

$$\begin{aligned} \dot{z}_1 &= a_1(y, \bar{u}) + \eta(\boldsymbol{x}, \bar{u}) && (46) \\ \dot{z}_2 &= z_1 + a_2(y, \bar{u}) && (47) \\ &\vdots \\ \dot{z}_n &= z_{n-1} + a_n(y, \bar{u}) && (48) \\ y &= z_n. && (49) \end{aligned}$$

That is, the disturbance term η, which cannot be canceled by a function of only y and $\bar{u}$, appears only in (46). Since we implement the observer in the $\boldsymbol{x}$-coordinates, $\eta(\boldsymbol{x}, \bar{u})$ is implicitly canceled by its estimate, $\eta(\hat{\boldsymbol{x}}, \bar{u})$, so it does not "bias" the estimates. However, η does appear in the error dynamics, which, in the $\boldsymbol{z}$-coordinates are

$$\dot{\tilde{\boldsymbol{z}}} = (\boldsymbol{A} - \boldsymbol{G}\boldsymbol{C})\tilde{\boldsymbol{z}} + [\eta(\boldsymbol{x}, \bar{u}) - \eta(\hat{\boldsymbol{x}}, \bar{u}) \;\; 0 \;\; \cdots \;\; 0]^T. \quad (50)$$

We formally state the synthesis as a theorem, whose proof follows directly from the above development.

Theorem 3.1 If (1)-(2) satisfies Assumption **A1** and Condition **C1**, then an observer of the form (12), which minimizes $\| \eta \|$ on a mesh of states and controls, can be constructed using spline functions to represent $\boldsymbol{a}$.

4 Examples

This section presents two examples that illustrate the performance of the spline-based observer design. Both are modified versions of those in [15].

4.1 Integrable Example

An armature-controlled DC motor of negligible mass with its shaft connected to a pendulum is described by $\boldsymbol{f}(\boldsymbol{x}, \bar{u}) = (x_2, \sin x_1 + x_3, -x_2 - x_3 + 10\bar{u})^T$. With the output measuring the pendulum's angle, $h(\boldsymbol{x}) = x_1$, there exists a linear change of coordinates to put the system in observer form:

$$\left.\begin{aligned} \dot{\boldsymbol{z}} &= \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \boldsymbol{z} + \begin{bmatrix} \sin z_3 + 10\bar{u} \\ \sin z_3 - z_3 \\ -z_3 \end{bmatrix} \\ y &= h(\boldsymbol{\Phi}^{-1}(\boldsymbol{z}, \bar{u})) = z_3 \end{aligned}\right\}. \quad (51)$$

Conditions **A1**, **C1**, and **C2** hold, allowing us to compute $D\boldsymbol{T}$ from $\boldsymbol{f}$ and h using (14).

We take $[a, b] = [-20, 20]$, and $[c, d] = [-3\pi, 3\pi]$. The B-spline basis is fifth-degree for y, first-degree for u, $N = 12$, and $M = 2$, giving $n \cdot N \cdot M = 72$ spline coefficients. When evaluating $\boldsymbol{w}$ to construct $\boldsymbol{W}$ in (36) a mesh of 480 points ($L_x = 240, L_u = 2$) on $\{(\boldsymbol{x}, \bar{u}) \in [-3\pi, 3\pi] \times [-\pi, \pi] \times [-\pi, \pi] \times [-20, 20]\}$ is used. To stabilize the pendulum about its inverted position $x_1 = 0$, linear feedback of the *estimated* state is applied to place the poles of the linearized plant at -1. The observer gain $\boldsymbol{G}$ in (11) is chosen to place the poles at -1. Figure 1 shows the state error $\boldsymbol{e} := \boldsymbol{x} - \hat{\boldsymbol{x}}$ for the nonlinear observer (12), along with that for for a linear observer, constructed using the Jacobian linearization of (51). The initial conditions are $\hat{\boldsymbol{x}}(0) = (0, 0, 0)^T$ and $\boldsymbol{x}(0) = (\pi, 0, 0)^T$. The error dynamics for the nonlinear observer correspond exactly to those of an exact linear system with poles at -1, modulo the small spline function approximation error ϵ.

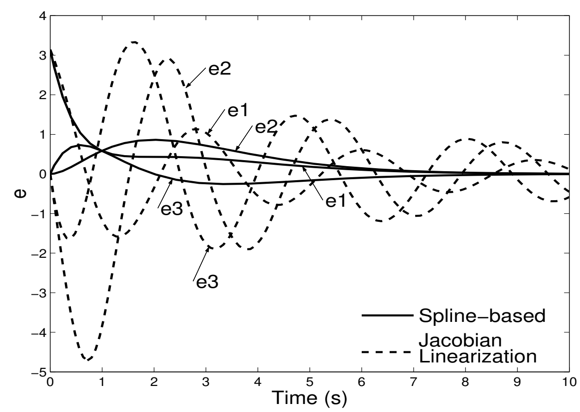

Figure 1: State error for the nonlinear and Jacobian linearization observers in Example 1.

4.2 Non-Integrable Example

A controlled undamped pendulum with output function $h(\boldsymbol{x}) = 2x_1 + x_2$ is described by $f(\boldsymbol{x}, \bar{u}) = (x_2, \sin x_1 + \bar{u})^T$. Although **A1** and **C1** are satisfied, **C2** is not. From (46)-(49), $z_1 = \mathcal{L}_{\bar{f}} h(\boldsymbol{x}) - a_2(y, \bar{u})$ and $z_2 = h(\boldsymbol{x}) = 2x_1 + x_2$, giving

$$D\boldsymbol{T}(\boldsymbol{\Phi}(\boldsymbol{x}, \bar{u}), \bar{u}) = \begin{bmatrix} \frac{1}{\cos x_1 - 4} & \frac{a_2'(y,\bar{u}) - 2}{\cos x_1 - 4} \\ \frac{-2}{\cos x_1 - 4} & \frac{\cos x_1 - 2a_2'(y,\bar{u})}{\cos x_1 - 4} \end{bmatrix}.$$

We compare the performance of our proposed spline-based observer with an Extended Linearized observer described in [16], which is governed by $\dot{\hat{\boldsymbol{x}}} = \boldsymbol{f}(\hat{\boldsymbol{x}}, \bar{u}) + \boldsymbol{g}(y) - \boldsymbol{g}(y)$, where $\boldsymbol{g}$ is designed so that the spectrum of the linearized error equation is

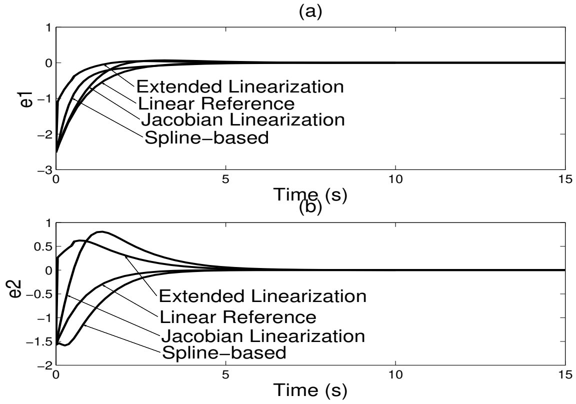

Figure 2: The first component of $\boldsymbol{e}$ for Example 2 is shown in (a) and (b) shows its second component.

$\{-1, -1\}$, independent of constant operating points. We define $[a, b] = [-10, 10]$, $[c, d] = [-\pi, \pi]$, use fifth-degree spline basis functions for y, first-degree spline basis functions for $\bar{u}$, and set $N = 10$ and $M = 4$, giving $n \cdot N \cdot M = 80$ spline coefficients. We use a mesh of 400 points ($L_x = 100, L_u = 4$) on the set $\{(\boldsymbol{x}, \bar{u}) \in [-\pi, \pi] \times [-\pi, \pi] \times [-10, 10]\}$. Estimated state feedback regulates the pendulum about its inverted position, with the poles of the linearized model at -1, while the poles of (11) are placed at -1. For the initial conditions $\boldsymbol{x}(0) = (-4\pi/5, 0)^T$ and $\hat{\boldsymbol{x}}(0) = (0, \pi/2)^T$, Figures 2 (a) and (b) show the components of the state error $\boldsymbol{e}$ for the spline-based observer, a design based on a Jacobian linearization of the example (with poles at -1), and the Extended Linearization observer. As well, the response of a second-order linear system with poles at -1 is plotted for comparison. Since the spline-based design minimizes the nonlinear disturbance term η, its response is the closest to the "reference" system. Similar results are obtained for different initial conditions.

5 Conclusions

In this paper, we have presented a method of synthesis for nonlinear observers. When the system satisfies **A1**, **C1**, and the non-generic **C2**, the observer has linear, exponentially stable error dynamics. We employ spline functions to reduce the integration of the generalized characteristic equation to computationally efficient algebra, which can be solved regardless of the complexity of the original nonlinear system. The approach is also applied to systems that satisfy **A1** and **C1**, but fail **C2**. In this case, there is some nonlinearity that cannot be compensated by output injection. However, we minimize the Euclidean norm of this "disturbance," and in this sense, the observer synthesis is optimal.

The work can be extended in several directions. For example, in the case the NOLED does not exist, $\boldsymbol{\Pi}$ and hence "$D\boldsymbol{T}$," which we write in quotes because it is not a Jacobian, still can be computed via (13). It is tempting to use the non-integrable $\boldsymbol{\Pi}$ in an observer configuration such as (12), although in this case it is not obvious how to compute $\boldsymbol{a}(y, \bar{u})$. The spline-function synthesis of $\boldsymbol{a}$ presented in this paper may offer a solution.

References

[1] J. Hauser, "Nonlinear control via uniform system approximation," *Systems and Control Letters*, vol. 17, pp. 145–154, 1991.

[2] J. Hauser, S. Sastry, and P. Kokotović, "Nonlinear control via approximate input-output linearization: The ball and beam example," *IEEE Transactions on Automatic Control*, vol. 37, pp. 392–398, Mar. 1992.

[3] S. A. Bortoff, "Advanced nonlinear control using digital signal processing," *IEEE Transactions on Industrial Electronics*, vol. 41, pp. 32–39, Feb. 1994.

[4] A. Banaszuk and J. Hauser, "Approximate feedback linearization: Homotopy operator approach," in *Proceedings of the 1994 American Control Conference*, pp. 1690–1694, June 1994.

[5] S. A. Bortoff and M. W. Spong, "Observer-based pseudo-linearization using splines: The rolling acrobot example," in *Proceedings of the ASME-WAM 1992*, (Anaheim, CA), ASME, Nov. 1992.

[6] H. Keller, "Nonlinear observer design by transformation into a generalized observer canonical form," *International Journal of Control*, vol. 46, no. 6, pp. 1915–1930, 1987.

[7] A. Isidori, *Nonlinear Control Systems, Second Edition*. Berlin: Springer-Verlag, 1989.

[8] D. Bestle and M. Zeitz, "Canonical form observer design for nonlinear time-variable systems," *International Journal of Control*, vol. 38, no. 2, pp. 419–431, 1983.

[9] T. P. Proychev and R. L. Mishkov, "Transformation of nonlinear systems in observer canonical form with reduced dependency on derivatives of the input," *Automatica*, vol. 29, no. 2, pp. 495–498, 1993.

[10] A. J. Krener and A. Isidori, "Linearization by output injection and nonlinear observers," *Systems and Control Letters*, vol. 3, no. 1, pp. 47–52, 1983.

[11] A. J. Krener and W. Respondek, "Nonlinear observers with linearizable error dynamics," *SIAM Journal of Control and Optimization*, vol. 23, pp. 197–216, Mar. 1985.

[12] X. H. Xia and W. B. Gao, "Nonlinear observer design by observer error linearization," *SIAM Journal of Control and Optimization*, vol. 27, pp. 199–216, Jan. 1989.

[13] A. R. Phelps, *A simplification of nonlinear observer theory*. PhD thesis, University of California at Berkeley, University Microfilms International, 1987. Order Number 8726336.

[14] L. L. Schumaker, *Spline Functions: Basic Theory*. New York: Wiley, 1981.

[15] B. Walcott, M. Corless, and S. Zak, "Comparative study of non-linear state-observation techniques," *International Journal of Control*, vol. 45, no. 6, pp. 2109–2132, 1987.

[16] W. T. Baumann and W. J. Rugh, "Feedback control of nonlinear systems by extended linearization," *IEEE Transactions on Automatic Control*, vol. AC-31, pp. 40–46, Jan. 1986.