Interpolation and Numerical Differentiation for Observer Design

S. Diop,* J. W. Grizzle†P. E. Moraal†and A. Stefanopoulou†

September 26, 1993

Abstract

This paper explores interpolation and numerical differentiation as a basis for constructing a new approach to the design of nonlinear observers. Numerical differentiation is well known to be an ill-conditioned problem in the sense that small perturbations on the function to be differentiated may induce large changes in the derivatives. We illustrate through examples that some specific methods may be able to overcome this difficulty, turning numerical differentiation techniques into a possible challenger to existing observer design methods.

Keywords: Observer design; Interpolation; Numerical differentiation; System theory.

1. Introduction.

Observer design is a fundamental problem in system theory and control practice. Indeed, many feedback design techniques, especially for nonlinear system models, start with the assumption that the full state is available for feedback and that the *user* will design separately a suitable observer for the actual implementation. This in turn has motivated a great deal of research into observer design.

The specific problem addressed herein involves the situation where one seeks to "estimate" the states of a continuous-time model from observations collected at discrete instants in time. To date, this problem has been addressed only in the context of sampled data systems [23, 25, 24, 30], where the procedure has been to first compute a discrete-time model from the continuous-time model, and then to design an observer on the basis of the discretized system model. Of course, the discretization process sometimes can be quite non-trivial, from a numerical or computational point of view; also, it is usually quite important to assume that the observed data has been collected at regular intervals of time, and thus dealing with problems where event driven sampling is necessary can be difficult.

This paper comes out of the observation that on the one hand, the numerical analysis literature contains a substantial body of results on numerical differentiation, while on the other hand, one of the very basic problems in system theory turns out to be that of obtaining the time derivatives of a continuous system variable, which is usually known only through the differential equations of the system, and time samples of this variable. We present a first attempt in applying results on numerical differentiation to the observer design problem, mainly by working out some specific examples.

To see how numerical differentiation and observer design may tie together, let us consider the following simple observer design problem. Let a system be given as

$$
\begin{cases}
\dot{x}_1 & = & x_2 \\
\dot{x}_2 & = & -ux_1^3 \\
y & = & x_1
\end{cases}
\tag{1.1}
$$

and let the question be to build an observer for (1.1); that is, to reconstruct the states of the system on the basis of the available measurements. It is clear that the variable x_1 need not be computed since it is directly given by the measurement: $x_1 = y$. From the equations of (1.1), we see that $x_2 = \dot{y}$. If, as is usually the case, we have no access to $\dot{y}$ then we need to *compute* $\dot{y}$ from the measurement y. This is where numerical differentiation comes into play: we explore the feasibility of *numerically differentiating* y in order to obtain x_2. If y is known only through its time samples then numerical differentiation yields an approximation of $\dot{y}$ at the current sampling time t_k by some number $\hat{\dot{y}}(t_k)$. An algorithm for the computation of $\hat{\dot{y}}(t_k)$ from the samples of y, and *estimates of the error bounds* should thus be worked out. Usually, not only $\dot{y}$, but, also, finitely many higher derivatives of y should be approximated at time t_k in order to be able to compute all of the desired state components.

The intent of this simple example is to illustrate our main objectives in applying numerical differentiation techniques to the observer design problem. If efficient numerical differentiation algorithms are available, then they should be applicable to many other system theoretic problems that require the computation of the derivatives of continuous variables which are known only through their time samples and the differential equations they satisfy. The inversion problem is among other potential applications that come to mind.

Our investigation led us directly to interpolation theory, and more generally, to approximation theory. As is well known, there is a considerable number of different ways for approximating the value of an unknown function at some point. One commonly chooses among polynomial, rational, periodic, or exponential functions as basis elements. (See also methods stemming from the Shannon sampling theorem [6, 7, 45, 46, 5, 4, 36, 40].) The optimality of any approximation process is, of course, relative to the assumed basis functions. Approximation methods based on polynomials are the simplest ones, but spline approximations seem to have some interesting fundamental properties.

The paper is organized as follows. Section 2. briefly reviews the notion of observability we are using. Section 3. provides a considerable review of the numerical analysis literature on numerical differentiation. Section 4. illustrates on specific examples how numerical differentiation may be integrated into an overall observer design strategy.

2

2.   An observability property.

Given a system described by the state equations

$$\begin{cases} \dot{x}_i & = g_i(x, u) \quad (1 \le i \le n) \\ y_j & = h_j(x, u) \quad (1 \le j \le p) \end{cases} \tag{2.1}$$

we will adopt the following specific observability property (see [8, 23, 24, 25, 30, 44, 31, 32]): system (2.1) is **observable** if there is an integer N such that the map $H\left(u, \dot{u}, \cdots, u^{(N-2)}, \cdot\right)$ of the state space into some space of output values, defined by

$$H\left(u, \dot{u}, \cdots, u^{(N-1)}, x\right) = \begin{pmatrix} y \\ \dot{y} \\ \vdots \\ y^{(N-1)} \end{pmatrix}, \tag{2.2}$$

is *injective* for any fixed *universal input*. It is easiest at this stage to assume that there is actually no input in (2.1), or that all inputs are universal in the sense that the above map is injective for any input.

For an observable system (2.1), we then may write

$$x = L\left(u, \dot{u}, \cdots, u^{(N-2)}, y, \dot{y}, \cdots, y^{(N-1)}\right).$$

The existence of the map L is guaranteed by our definition of observability, but an explicit expression of L may not be easy to obtain so that numerical equation solving methods such as Newton's algorithm may be required in order to obtain L. When (2.1) is a rational system in the sense that the g_i's and the h_j's are rational functions of their arguments, then the type of observability we just defined corresponds to *rational observability* of [19, 18], and then the explicit expression of L may be constructively derived through the use of certain differential algebraic techniques.

For these observable systems, the observer design problem may thus be seen as a problem of numerical differentiation: once estimates of the derivatives of y and u can be determined from the available measurements, then x can be determined from L.

3.   On numerical differentiation.

The basic problem we are facing may be roughly stated as follows:

Problem P. A real valued function f of one real variable (called the *time, t*) defined on an interval I is known only through the differential equations it satisfies and through the values it takes at some instants $t_0, t_1, \cdots, t_k \in I$, with $t_0 < t_1 < \cdots < t_k$. The question is to provide estimates of the first q derivatives $f^{(1)}(t_k), f^{(2)}(t_k), \cdots, f^{(q)}(t_k)$ of f at time t_j, for some $0 \le j \le k$. The difficulty of the problem may be increased by the presence of noise in the samples of f, in which case, the approximation technique should provide *smoothing*, and even the values of f would need to be estimated.

Remarks:

- The more general problem we should end up facing in system theory is actually a multi-variable one in that f is a vector valued function instead of scalar valued as in the latter statement.

- If r is the order of the differential equations defining f then only at most the first $q = r - 1$ derivatives of f need be estimated, the succeeding derivatives being computable from the previous ones by means of these differential equations defining f.

- Problem P is a particular case of the general problem of approximation of functions on the basis of partial knowledge of these functions. Its seems that there is not yet a complete solution with expressions of the error bounds in the literature. One partial solution consists in taking the derivatives of the *interpolant* $\hat{f}$ for those of f. On this approach, there is a substantial amount of literature. Another possible direction consists in trying to overcome the ill-conditionedness of the numerical differentiation problem by using some regularization techniques. The only known works are the one by Cullum [13], and those in the Russian literature (cited in the latter paper) to which we do not yet have access.

3.1. Generalities on approximation theory.

The basic problem that approximation theory addresses is the following: Assume that the unknown object to be approximated is an element x of a given metric space E, and find a *best estimate* $\hat{x}$ of x in an (also given) subspace F of E subject to

$$d(x, \hat{x}) = d(x, F) = \min_{y \in F} d(x, y),$$

where d denotes the distance on E. The word *best* is of course relative to the distance used in E and to the choice of F. For example, for a fixed distance, the larger F is, the finer the approximation will be. This problem has at least one solution if the subspace F is compact, or if E is a normed vector space and F is a linear subspace of finite dimension, or if E is a Hilbert space and F is a closed subspace.

In our case, the object to be approximated is the set of derivatives $f^{(1)}, f^{(2)}, \cdots, f^{(q)}$ of a real valued function f defined on, say a compact interval I of $\mathbb{R}$, so that the space E may be taken as the set of real valued functions which are $q - 1$ times differentiable with $f^{(q-1)}$ absolutely continuous and $f^{(q)}$ square integrable on I, or which are infinitely differentiable, or, even, which are analytic on I. E is thus endowed with its usual $\mathbb{R}$-algebra structure and may be converted into a real normed vector space by one of the following

$$
\begin{aligned}
\|g\|_p &= \left(\int_I |g(t)|^p dt \right)^{1/p}, \, g \in E, p \in \mathbb{R}, \, p \geq 1, \\
\|g\|_\infty &= \sup_{t \in I} |g(t)|, \, g \in E, \\
\|g\|_{(q)} &= \|g\|_2 + \|g^{(1)}\|_2 + \cdots + \|g^{(q)}\|_2, \, g \in E, \\
&\quad \text{etc.}
\end{aligned}
$$

The norm $\| \|_2$ corresponds to the Hilbert space structure on E, the corresponding inner product being

$$(g, h) = \int_I g(t)h(t)dt, \, g, h \in E.$$

The norm $\| \|_\infty$ is usually called the *infinity* or *uniform* norm.

Whenever F is a finite dimensional subspace of E and the norm $\| \|_2$ is used, the *least squares* approximation process results in the existence and uniqueness of the best estimate of any given f in E.

The subset F of E is usually taken as a subspace of one of the following linear subspaces:

4

- $\mathbb{R}[t]$, the subspace of polynomial functions,
- $\mathbb{R}(t)$, the subspace of rational functions,
- $\mathbb{R}[\cos t, \sin t]$, the subspace of $\mathbb{R}$-linear combinations of $\cos(it)$ and $\sin(it)$, $i \in \mathbb{N}$.

Assume that F is finite dimensional, with dimension $N' + 1$, and norm $\|\ \|$. We then know that, given any f_0 in E, there is at least one best estimate $\hat{f}$ of f_0 in the sense of minimizing $\|f - f_0\|$ over $f \in F$. When $\|\ \| = \|\ \|_\infty$, $\hat{f}$ is called a best uniform estimate. A base $(\zeta_0, \zeta_1, \ldots, \zeta_{N'})$ of F is said to satisfy the Haar condition if any real linear combination

$$c_1 \zeta_1(t) + \cdots + c_N \zeta_{N'}(t)$$

vanishes at less than $N' + 1$ distinct points in I.

Theorem 3.1 ((Haar), [17]) . *A necessary and sufficient condition for the best uniform estimate to be unique is that the base $(\zeta_0, \zeta_1, \ldots, \zeta_{N'})$ of F satisfies the Haar condition.*

3.2. Differentiating the interpolant.

Since the interpolation problem has been given a large number of elegant and practical solutions, the easiest way to approach Problem P is to take the derivatives of the interpolant $\hat{f}$ for those of f. Let us make the following remark at this stage even if we have not yet introduced the material which makes it clear: There might be, due the reality of a particular physical problem, no freedom in the choice of the sampling instants t_i, but the error bounds for the estimated derivatives are reasonably expected to depend on both the number of samples and their distribution on the time interval I as apparent from the Tchebychev theorem. Our study should address this issue, as well, but for now, we shall suppose that the current time is t_N; later on, we shall discuss how we should deal with the succeeding sampling instants $t_{N+1}, \ldots$ as time evolves.

Let $F_{N'}$ be a linear subspace of E of finite dimension $N' + 1$, and with base $(\zeta_0, \zeta_1, \ldots, \zeta_{N'})$. An element

$$\hat{\zeta}_{N'}(t) = c_0 \zeta_0(t) + \cdots + c_{N'} \zeta_{N'}(t)$$

of $F_{N'}$ *interpolates* f if

$$\hat{\zeta}_{N'}(t_i) = f(t_i)\,(0 \leq i \leq N),$$

that is, if

$$\sum_{j=1}^{N'} c_j \zeta_j(t_i) = f(t_i)\,(0 \leq i \leq N).$$

Therefore,

Theorem 3.2 *There is one and only one function in $F_{N'}$ interpolating f if and only if the previous set of equations in the c_j's has one and only one solution, that is, the existence and uniqueness in $F_{N'}$ of an interpolant of f is equivalent to $N' = N$ and the determinant of the following matrix*

$$\begin{pmatrix} \zeta_0(t_0) & \cdots & \zeta_N(t_0) \\ \vdots & & \vdots \\ \zeta_0(t_N) & \cdots & \zeta_N(t_N) \end{pmatrix}$$

is nonsingular; in this case, we may use the well-known Cramer formula to compute the inter-
polant, and write it in the form

$$\hat{\zeta}_N(t) = \sum_{i=1}^{N} f(t_i)s_i(t) \, (t \in I),$$

with $s_i(t_j) = \delta_{ij}$, where δ_{ij} is the Kronecker symbol.

Note that the Haar condition guarantees that the latter square matrix is nonsingular. Note also that this previous result shows that, when interpolating linearly a function at N points, the subspace of projection, F, should be of dimension precisely N.

Among the well-established results on interpolation theory, we will now discuss those due to Lagrange, those based on splines, etc.

3.2.1. Differentiating the Lagrange interpolant.

It is clear that there are infinitely many polynomials of arbitrary degree *interpolating f*, but, according to the previous general result, there is one and only one of degree at most N which interpolates f. This unique polynomial L_N is known as the Lagrange (or polynomial) interpolant of f.

Assuming

$$\ell_i(t) = \prod_{\substack{j=0 \\ j \neq i}}^{N} \frac{t - t_j}{t_i - t_j},$$

we have

$$L_N(t) = \sum_{i=0}^{N} f(t_i)\ell_i(t) \, (t \in I),$$

with $\ell_i(t_j) = \delta_{ij}$. This is not the most interesting form for the computation of the Lagrange interpolant. The Newton formulae lead to more efficient computations, see [47, 26, 34] for example.

If the only knowledge we have on f is the set of its samples, then being able to estimate the error bounds is hopeless. Given more information, the following result is known: If f is N times continuously differentiable in I, then for any t in I, there is a point ξ_t in $[t_0, t_N]$ such that

$$\varepsilon(t) = f(t) - \hat{\zeta}_N(t) = \frac{1}{(N+1)!}\ell(t)f^{(N+1)}(\xi_t) \, (t \in I),$$

where $\ell(t) = \prod_{i=0}^{N}(t - t_i)$.

This follows mainly from the Rolle theorem. In general, given t, we do not know ξ_t nor $f^{(N+1)}(\xi_t)$. This is the reason why we shall need to know a bound for the higher derivatives of f:

$$\|\varepsilon\|_\infty \leq \frac{\Delta^N}{(N+1)!}\|f^{(N+1)}\|_\infty,$$

where Δ is $\max_{1 \leq i \leq N}(t_i - t_{i-1})$. This bound depends both on the number $N + 1$ of samples and on their distribution. If it is possible to choose freely the sampling instants, then we would be able to decrease the latter error bound by sampling according to the Tchebychev theorem, see [34]. Now, the error when taking $\hat{\zeta}_N^{(j)}(t)$ for $f^{(j)}(t)$ is given by

$$\varepsilon^{(j)}(t) = f^{(j)}(t) - \hat{\zeta}_N^{(j)}(t) = \sum_{\iota=0}^{j} \frac{1}{(N+\iota+1)!}\binom{j}{\iota}\ell^{(j-\iota)}(t)f^{(N+\iota+1)}(\xi_\iota) \, (t \in I),$$

where the ξ_ι's are t dependent and in $[t_0, t_N]$. Therefore, we have

$$\left\|\varepsilon^{(j)}\right\|_\infty \leq c\,\Delta^{N+1-j} \max_{0 \leq \iota \leq j} \left\|f^{(N+1+\iota)}\right\|_\infty,$$

where c is an absolute constant (it does not depend on f nor the samples), see [38].

Remark: Though simple and elegant, Lagrangian interpolation suffers from the Runge phenomenon which is the fact that a small change to the data points in the middle of the data set may produce a very large excursion in the end points, t_0, t_N, see §3.2.3 of [34]. According to the latter book, Lagrangian interpolation is thus not recommended for large set of equispaced data. Nevertheless, there are some improvements of the Lagrangian interpolation in the literature: [38] addresses the problem of numerical differentiation with application to computer-aided-design (the basic assumption that we know the range of f made in that paper makes it unrealistic in control theory, but still deserves some attention), [11, 12, 9, 10] propose an *osculatory* alternative by assuming available some extra data points which represent values of the higher derivatives of f at the sampling instants.

3.2.2. Differentiating the spline interpolant.

Excellent introductions to spline theory may be found in [22, 34]; §3 of the latter book is particularly recommended for a motivated step by step introduction of splines in the theory of interpolation.

A *natural spline* of *degree* $2l - 1$, $l \in \mathbb{N}$, $l \geq 1$ with *knots* $t_0, t_1, \cdots, t_N$ is a real valued function, s, defined on $\mathbb{R}$, and such that:

(i) s has $2l - 2$ first derivatives defined and continuous on $\mathbb{R}$,

(ii) In each interval $[t_i, t_{i+1}]$, $i = 0, 1, \cdots, N - 1$, s is polynomial and with degree at most $2l - 1$,

(iii) In each interval $]\infty, t_0]$, $[t_N, \infty[$ s is polynomial with degree at most $l - 1$.

The set of natural splines of degree $2l - 1$ with knots $t_0, t_1, \cdots, t_N$ is classically denoted by $\mathcal{N}_{2l-1}(t_0, t_1, \ldots, t_N)$. It is clear that $\mathcal{N}_{2l-1}(t_0, t_1, \ldots, t_N)$ contains the set of real coefficient polynomials in t with degree less than l.

The fine points of spline interpolation theory are the following fundamental properties, see [39, 16, 22], for example, for more properties and proofs.

Interpolation property. Let f be in E. For any natural integer l such that $l \leq N + 1$ there is one and only one natural spline $\hat{s}$ in $\mathcal{N}_{2l-1}(t_0, t_1, \ldots t_N)$ which interpolates f. When $l = N + 1$ then, of course, $\hat{s}$ is the Lagrange polynomial which is the unique polynomial of degree less than $N + 1$ interpolating f. For $l \geq N + 1$, there are infinitely many natural splines which interpolate f.

Smoothing properties. Given f in E, and $l \leq N + 1$ then

$$\int_I \left(s^{(l)}(t)\right)^2 dt \leq \int_I \left(f^{(l)}(t)\right)^2 dt$$

for any s in $\mathcal{N}_{2l-1}(t_0, t_1, \ldots, t_N)$, and equality holds only if f is the unique natural spline $\hat{s}$ interpolating f.

The reason why this last property is called *smoothing* is explained very clearly in §3.4 of [34]. It refers to *cubic splines* where the above integrals may be interpreted in terms of energy. Note that this minimum property may be generalized a bit farther as is done in Sard's linear approximation theory [37], see also [22], and [27, 29, 28]. Roughly speaking, the generalization

7

consists in substituting a general linear functional for $f^{(l)}(t)$ in the above integrals. We need the theory of Peano kernels, or more generally, of reproducing kernels in Hilbert spaces in order to understand and use these generalizations of spline interpolation. One may ask if an ultimate generalization to the nonlinear case would not be a definite solution to Problem P;see [28] and the references therein, for what seems to be a step towards this idea.

We postpone to the next draft of this paper the summary of the results on error bounds of spline based, numerical differentiation. Let us just mention some papers containing partial results: [21, 48, 3, 41, 37, 22, 20]. We have not yet consulted [15] which is reported to be an excellent source for spline interpolation. This last paper contains an innovative idea which consists in improving the spline interpolation by iterating the basic process of spline interpolation as many times as desired. It is claimed that arbitrary precision may thus be obtained.

Spline interpolation certainly provides better estimation, and constitutes the most elaborate method of interpolation when one restricts oneself to linear approximation with polynomial base functions. Beyond that method, one has to enter the theory of *nonlinear approximation* (see [35]) with its simplest form being *rational approximation.* But then, caution is recommended in the literature, and, anyway, results are far less simple and clear.

The error bounds, say ε_0, ε_1,..., on the respective derivatives $f^{(0)}$, $f^{(1)}$,... of a function f, no matter what the interpolation process is, are reasonably expected to be such that $\varepsilon_0 < \varepsilon_1 < \ldots$, see [14].

Finally, let us note that we do yet not have at hand the references [33, 2] which seem to be very interesting[1].

3.3. Numerical differentiation by regularization.

As we said earlier, the only reference on this approach that we yet have at hand is the paper by Cullum [13]. This paper mentions some other works by Russian numerical analysts. Further bibliographic research is also due since there might be other works which have been completed since 1971.

4. Illustrative examples.

In order to better understand some of the issues involved in the use of numerical differentiation and interpolation as a basis for an observer design method, three academic examples are currently being pursued. The first example involves the estimation of a time function and a few of its derivatives from noisy samples collected at discrete instants of time. The second example builds upon the first one by showing that if the signal is generated by a known, continuous-time linear model, then the model's dynamcis can be incorporated into the estimation process with advantage. This will bring us into contact with more classical observer theory. The third example will illustrate that the techniques sketched in the first two examples are also applicable to nonlinear system models, which is of course, the main point.

In each case, polynomials are used as the interpolating funtions, total least squares at the interpolating points is the metric to be minimized, and the discrete data will be gathered at

---

[1]It is remarkable that these researchers were with engineering departments such as the IBM Research Center at Yorktown Heights, NY; some other contributors on numerical differentiation used to be with Ford Mo. Co., Dearborn, MI.

a uniform rate. These simplifications are made so that the discussion can be focused on other issues.

It is emphasized that no theorems are formulated and thus no proofs of validity are offered for any of the algorithms proposed herein. This paper is meant to be an exploration of a set of ideas. Formalization of these ideas into a rigorous design methodology will be pursued when sufficient numerical evidence exists suggesting that it is useful to do so.

4.1.   Sum of two sinusoids.

The purpose of this example is to illustrate the use of numerical differentiation, alone and in combination with standard system theoretic notions, to recover or estimate several derivatives of a signal in noise. The key parameters of interest are: N, the order of the interpolating polynomial; Δt, the time interval between data points; $W * \Delta t$, the length in time of the moving window used for data interpolation (note that $W + 1$ is the number of data points in the window); K, the data node or "knot" within the moving window where the derivative is to be estimated (counted from the left, with the first node in the window numbered zero); and σ, the "intensity" of the noise process corrupting the measurements.

Let an analog signal be given by the sum of two sinusoids of frequencies 1 Hz. and 5 Hz., respectively:
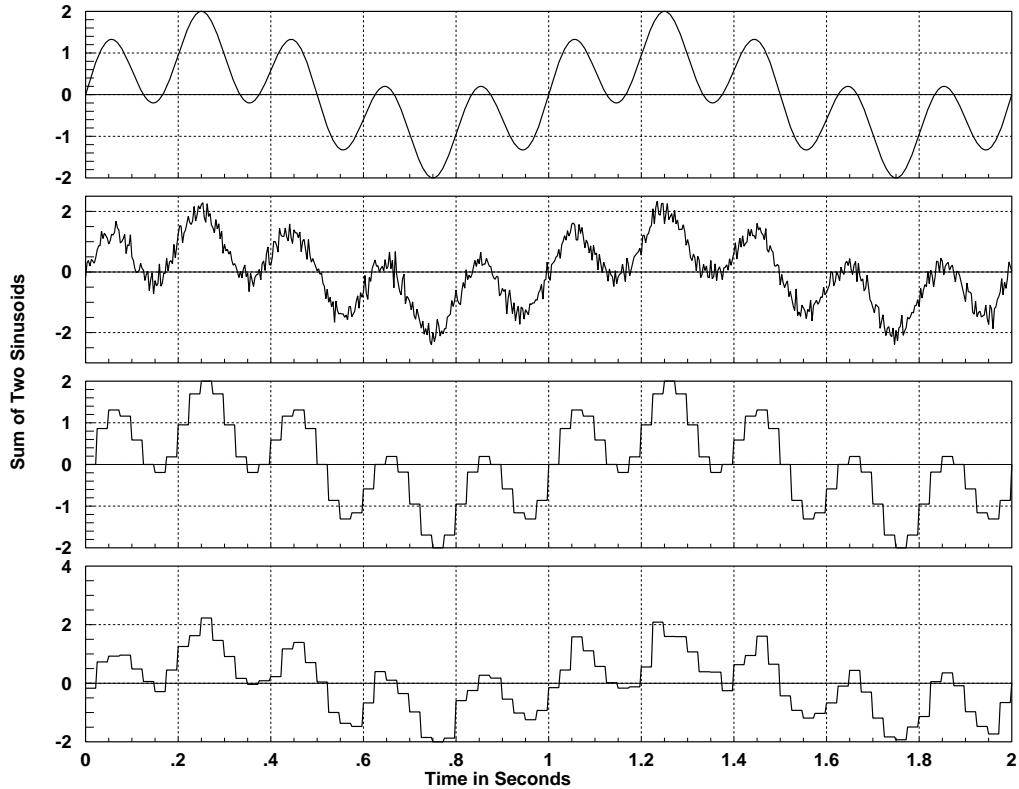
$$y(t) = sin(2\pi t) + sin(10\pi t). \tag{4.1}$$

Figure 1: Time plots of $y(t)$, $y(t) + w(t)$, y_k^m and $y_k^m + w_k$, respectively.

Assume that the measured signal, y^m, is the sum of y and w, where w is a Gaussian white noise process, with standard deviation σ. It is further supposed that y^m is to be sampled at

discrete instants of time, $k\Delta t$, $k = 0, 1, \ldots$. In the absence of noise, the minimum sampling rate to reconstruct y from sampled data would be the Nyquist rate, 10 Hz.; with noise, at least four times this rate is warranted (an anti-aliasing filter is not being used though it would be standard practice to do so). Discrete samples of $y + w$ will be collected therefore at 40 Hz.:

$$
\begin{aligned}
y^m(t) &:= y(t) + w(t) &\qquad (4.2)\\
y_k^m &:= y^m(k\Delta t).
\end{aligned}
$$

The time signals discussed are depicted in Figure 1. The noise intensity was selected to provide a (numerically computed) signal to noise ratio of 20/1, and corresponded to $\sigma = 0.22$. This should provide a feel for the sample rate and noise intensity being used in all later simulations.

Let the interpolating polynomial for the window of data $\{y_{k-W}^m, \ldots, y_k^m\}$ be denoted by

$$
\hat{y}_k(t) = a_0 + a_1(t - t_{k-W}) + \ldots + a_N(t - t_{k-W})^N, \qquad (4.3)
$$

where $t_k := k\Delta t$. The coefficients $\{a_0, a_1, \ldots, a_N\}$ are determined from the least squares solution of

$$
\begin{bmatrix}
1 & 0 & \ldots & 0 \\
1 & \Delta t & \ldots & (\Delta t)^N \\
\vdots & \vdots & \vdots & \vdots \\
1 & W\Delta t & \ldots & (W\Delta t)^N
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
\vdots \\
a_N
\end{bmatrix}
=
\begin{bmatrix}
y_{k-W}^m \\
y_{k-W+1}^m \\
\vdots \\
y_k^m
\end{bmatrix}, \qquad (4.4)
$$

with respect to the Euclidean norm. The estimates of the derivatives of y at time $\bar{t}$ are determined by

$$
\widehat{\frac{d^j}{dt^j} y_k}(\bar{t}) := \frac{d^j}{dt^j} \hat{y}_k(t)_{|t=\bar{t}} \ ; \qquad (4.5)
$$

for simplicity of notation, this is written as $\frac{d^j}{dt^j} \hat{y}_k(\bar{t})$.

A number of numerical experiments or simulations were conducted to ascertain the effects of the parameters N, Δt, W, K and σ on the ability to estimate the derivatives of y. The general trends of these experiments are summarized here; quantifying these observations analytically would be a worthy goal:

- The order of the interpolating polynomial should be selected as low as possible in order to smooth the noise in the signal. On the other hand, for the estimation of the 3rd derivative of a signal, for example, at least a 3rd order polynomial is required.

- The window length $W\Delta t$ must be chosen large enough to capture the variations of the signal so that the interpolating polymomial can reproduce its derivatives. However, larger $W\Delta t$ require higher order polynomials to be chosen.

- Estimates of the derivatives made at the end-points of the window $W\Delta t$ are considerably less accurate than those made in the interior. However, selecting $K < W$ introduces a time delay of $(W - K)\Delta t$ in the estimates of the signal and its derivatives.

- For fixed N and $W\Delta t$, decreasing the time between samples Δt or increasing the number of points in the window W, allows a higher noise intensity σ to be tolerated.

- For a fixed Δt, though increasing W can produce more accurate results, it must be considered that it requires more on line compuation and a larger delay must be introduced (i.e., $(W - K)\Delta t$ increased) in order to retain the accuracy of the estimates.
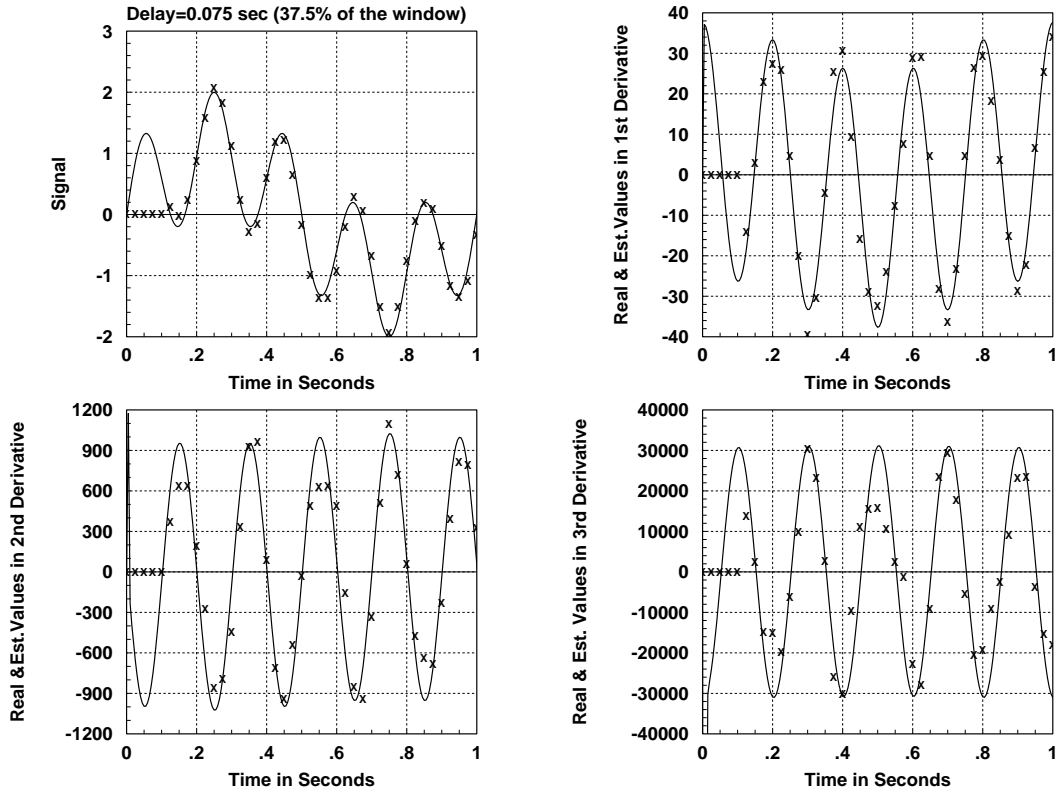
10

Figure 2: Comparison of the true and estimated ($\times$) values of y, $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$, and $\frac{d^3y}{dt^3}$, respectively using interpolation/numerical differentiation.

Figure 2 compares the true and estimated values of y, $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$, and $\frac{d^3y}{dt^3}$ when $N = 4$, $\Delta t = .025$, $W = 8$ and $K = 5$. Figure 3 displays the normalized errors in the estimates: the errors in the estimates for $(y, \ldots, \frac{d^3}{dt^3}y)$ are normalized by

$$lim_{T\to\infty} \frac{1}{T} \int_0^T |\frac{d^j}{dt^j} y(t)| dt. \tag{4.6}$$

Thus, the error in y is divided by 0.796, $\frac{dy}{dt}$ by 19.3, $\frac{d^2y}{dt^2}$ by 650, and $\frac{d^3y}{dt^3}$ by 22,700. The wide range in the magnitudes makes the estimation problem quite challenging.

Estimating the signal and its derivatives on the basis of the $W + 1$ data points in the moving window corresponds to using an FIR (finite impulse response) filter. By introducing a simple modification to the above, a sort of IIR (infinite impulse response) filter can be achieved. The modification is somewhat analogous to what is done in spline approximations, but to a systems person, it would be called a state.

The idea is to append to (4.4), the estimates of y and its derivatives at node K from the previous window, that is, one appends the relations

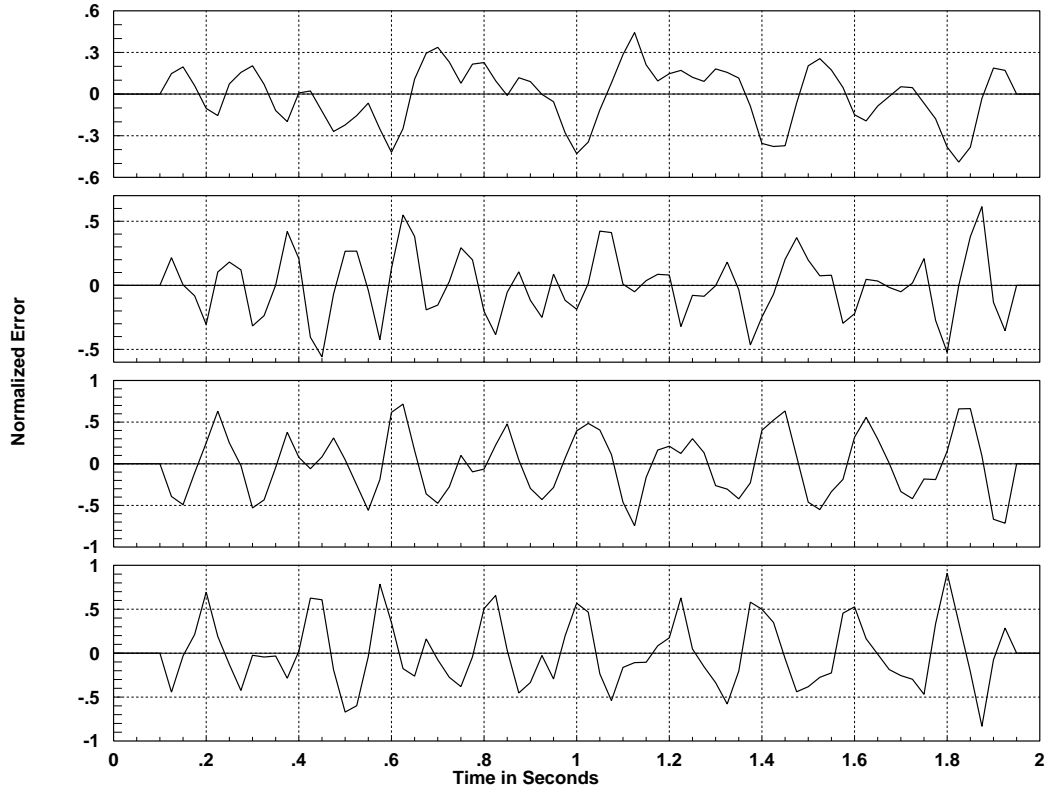$$\frac{d^j}{dt^j} \hat{y}_k(t_{k-1-W+K}) = \frac{d^j}{dt^j} \hat{y}_{k-1}(t_{k-1-W+K}), \tag{4.7}$$

11

Figure 3: Normalized estimation errors in y, $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$, and $\frac{d^3y}{dt^3}$, respectively using interpolation/numerical differentiation.

for $j = 0, \ldots, J$, for some $0 \le J \le N$. This is equivalent to

$$
\begin{bmatrix}
1 & (K-1)\Delta t & ((K-1)\Delta t)^2 & \ldots & ((K-1)\Delta t)^J & \ldots & ((K-1)\Delta t)^N \\
0 & 1 & 2(K-1)\Delta t & \ldots & J((K-1)\Delta t)^{J-1} & \ldots & N((K-1)\Delta t)^{N-1} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\
0 & 0 & \ldots & \ldots & J! & \ldots & N(N-1)\ldots(N-J+1)((K-1)\Delta t)^{N-J}
\end{bmatrix} \times
$$

$$
\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_N \end{bmatrix} =
\begin{bmatrix} \hat{y}_{k-1}(t_{k-1-W+K}) \\ \frac{d}{dt}\hat{y}_{k-1}(t_{k-1-W+K}) \\ \vdots \\ \frac{d^J}{dt^J}\hat{y}_{k-1}(t_{k-1-W+K}) \end{bmatrix} \qquad (4.8)
$$

The right-hand side of (4.7) must be initialized to start the interpolation process, and thus becomes a state. This state allows derivative information to be passed from one window of data to the next. Weights can be added to trade-off errors in interpolating the measured data points versus errors in interpolating the derivatives in (4.7).

4.2. Including a system model in the interpolation process.

The goal of this example is to demonstrate two methods for incorporating a dynamic model of a system into the interpolation/numerical differentiation process, whenever the signal in question is generated by a finite set of ordinary differential equations. This will bring us into more direct contact with observer design because we will tightly connect estimating the derivatives of a system's output with estimating the states of the model.

The signal (4.1) can obviously be generated by an initialized, uncontrolled, linear model with four states. Let

$$
\begin{aligned}
\frac{dx}{dt} &= Ax \\
y &= Cx
\end{aligned}
\tag{4.9}
$$

be such a model; it will be observable. This means that x can be recovered from estimates of y and its derivatives through

$$
\begin{bmatrix} y \\ \frac{d}{dt}y \\ \vdots \\ \frac{d^n}{dt^n}y \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^\eta \end{bmatrix} x
\tag{4.10}
$$

for η large enough, which in the particular case of (4.9) is $\eta = 3$.

The first way to include the model in the estimation process is now explained. Suppose that in general the model (4.9) has dimension n, and for simplicity, assume that it has a single output. As before, let

$$
\begin{bmatrix} \hat{y}_{k-1}(t_{k-1-W+K}) \\ \frac{d}{dt}\hat{y}_{k-1}(t_{k-1-W+K}) \\ \vdots \\ \frac{d^{n-1}}{dt^{n-1}}\hat{y}_{k-1}(t_{k-1-W+K}) \end{bmatrix}
\tag{4.11}
$$

be estimates of y and its first $n-1$ derivatives from the previous window. By substituting (4.11) for the right-hand side of (4.10) with $\eta = n - 1$, one can determine $\hat{x}_{k-1}(t_{k-1-W+K})$. If the resulting estimate of x is "accurate", then it satisfies (4.10) for any η; if not, choosing $\eta \geq n$ introduces constraints whose errors can be minimized in order for x to be more closely compatible with the dynamics (4.9). Thus, in place of (4.7), one augments (4.4) with

$$
\frac{d^j}{dt^j}\hat{y}_k(t_{k-1-W+K}) = CA^j\hat{x}_{k-1}(t_{k-1-W+K}),
\tag{4.12}
$$

For $j = 0, \ldots, \eta$, and proceeds as indicated previously. Note that taking $\eta = J \leq n - 1$ reduces (4.12) to (4.7).

Figure 4 shows the effect of this modification on the estimation process for $N = 6$, $\Delta t = .025$, $W = 8$ and $K = 5$. The weight on the constraints was chosen as

$$
Q = \begin{bmatrix}
1.26 & 0 & 0 & 0 & 0 & 0 \\
0 & 5.19\ 10^{-2} & 0 & 0 & 0 & 0 \\
0 & 0 & 1.54\ 10^{-3} & 0 & 0 & 0 \\
0 & 0 & 0 & 4.4\ 10^{-5} & 0 & 0 \\
0 & 0 & 0 & 0 & 1.3\ 10^{-6} & 0 \\
0 & 0 & 0 & 0 & 0 & 4.4\ 10^{-8}
\end{bmatrix}.
\tag{4.13}
$$

The rationale for selecting the entries of the weight Q is that the magnitude of each successive derivative grows by a approximatley 30 (recall (4.6), or see (4.1)), and thus Q corresponds to equal "relative weighting" on y and its derivatives; this same rationale will be applied to subsequent examples. Including the model interpolation constraints allowed the same window size and delay as in Figure 3 to be used with an increased polynomial order, without "following" the noise.

The method just outlined for estimating x from y and its derivatives is akin to a "dead-beat" observer for x, though, as explained earlier, there is smoothing in the interpolation of y, which
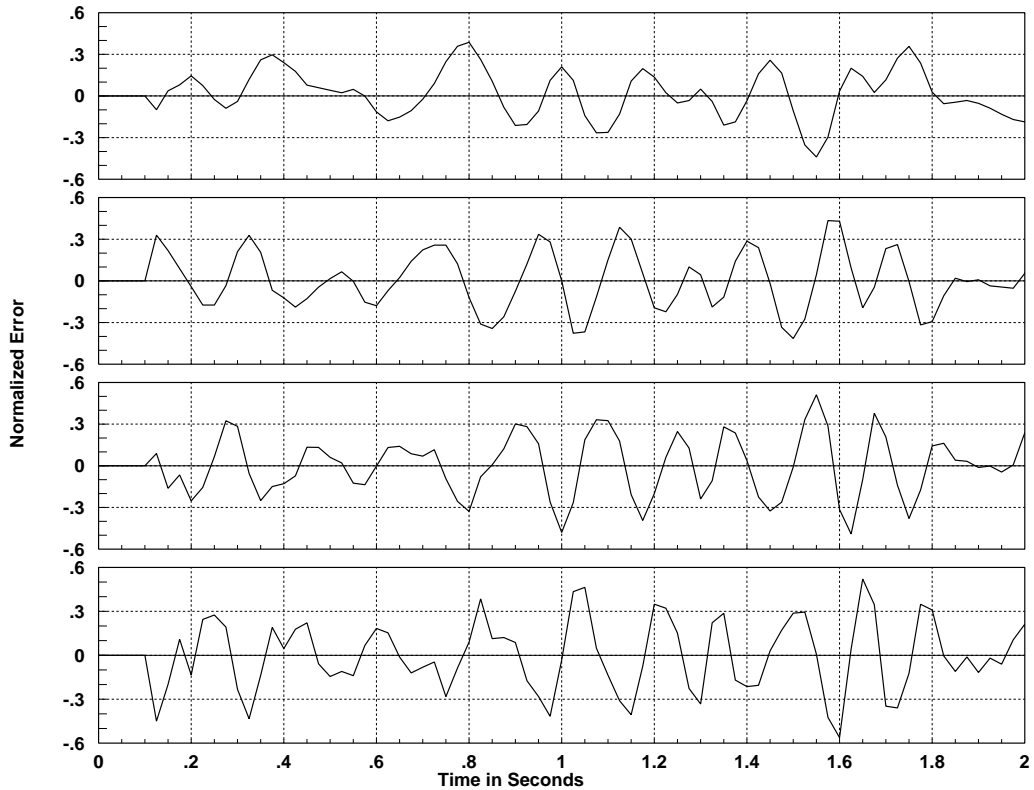
Figure 4: Normalized estimation errors in y, $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$, and $\frac{d^3y}{dt^3}$, respectively using interpolation/numerical differentiation and $\eta = 5$ in (4.12). Note the reduced error with respect to Figure 3.

is "IIR". Some additional smoothing from the model can be obtained, and this constitutes the second method for introducing the model into the interpolation/numerical differentiation process. This technique, which is reminiscent of [23, 25, 24, 30], will be directly applicable to nonlinear systems.

Let A_d denote the discretization of the linear model, (4.9), and let $\hat{x}_{k-1}^+(t_{k-1-W+K})$ be the previous estimate of x. Update x based on the latest estimate of y and its derivatives through a damped Newton method:

$$
\begin{aligned}
\hat{x}_k^-(t_{k-W+K}) &= A_d \hat{x}_{k-1}^+(t_{k-1-W+K}) \\
\hat{x}_k^+(t_{k-W+K}) &= \hat{x}_k^-(t_{k-W+K}) + \\
\epsilon \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix}^{-1} & \left(\begin{bmatrix} \hat{y}_k(t_{k-W+K}) \\ \frac{d}{dt}\hat{y}_k(t_{k-W+K}) \\ \vdots \\ \frac{d^{n-1}}{dt^{n-1}}\hat{y}_k(t_{k-W+K}) \end{bmatrix} - \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{bmatrix} \hat{x}_k^-(t_{k-W+K}) \right),
\end{aligned}
\tag{4.14}
$$

where $0 < \epsilon < 1$; using $\epsilon = 1$ reduces the above to a "dead-beat" estimator for x. The damped Newton update can be used with (4.4) alone or with any of the additions to (4.4) discussed previously.

Selected simulations of the above estimation methods are now presented. In what follows,

14

the state in (4.9) was chosen as

$$x = \begin{bmatrix} y \\ \frac{d}{dt}y \\ \frac{d^2}{dt^2}y \\ \frac{d^3}{dt^3}y \end{bmatrix}.$$

It was properly initialized so that its output is precisely given by (4.1). The same noise sequence used in Figure 3 was added to the output of (4.9).
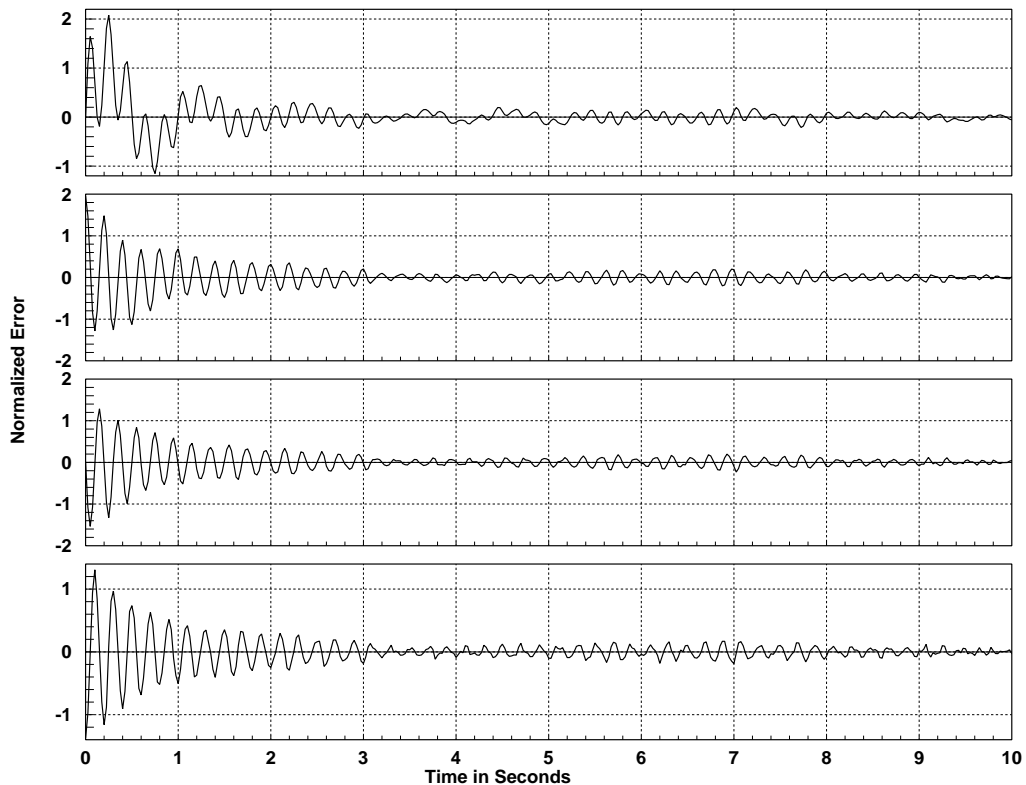

Figure 5: Normalized estimation errors in y, $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$ and $\frac{d^3y}{dt^3}$, respectively, incorporating "spline method" and Newton update in interpolation/numerical differentiation.

Figure 5 shows the normalized estimation error resulting from the application of (4.4) in combination with (4.7) and (4.14). The parameter values were chosen as: $N = 5$, $\Delta t = 0.025$ sec., $W = 6$ and $K = 3$ in (4.4), $\eta = 4$ in (4.12) with a weight on (4.7) selected as

$$Q = 0.1 \times \begin{bmatrix} 1.26 \; 10^{-1} & 0 & 0 & 0 \\ 0 & 5.19 \; 10^{-3} & 0 & 0 \\ 0 & 0 & 1.54 \; 10^{-4} & 0 \\ 0 & 0 & 0 & 4.4 \; 10^{-5} \end{bmatrix}, \qquad (4.15)$$

and $\epsilon = 0.03$ in (4.14). For comparison purposes, a steady state, discrete-time, Kalman filter was designed with state noise covariance equal to $0.01 \; \sigma^2 \; Q^{-1}$ and output noise variance equal to σ^2; even though (4.9) does not have any process noise per se, the pair consisting of the matrix A and the state noise convariance matrix had to be made stabilizable for a stable
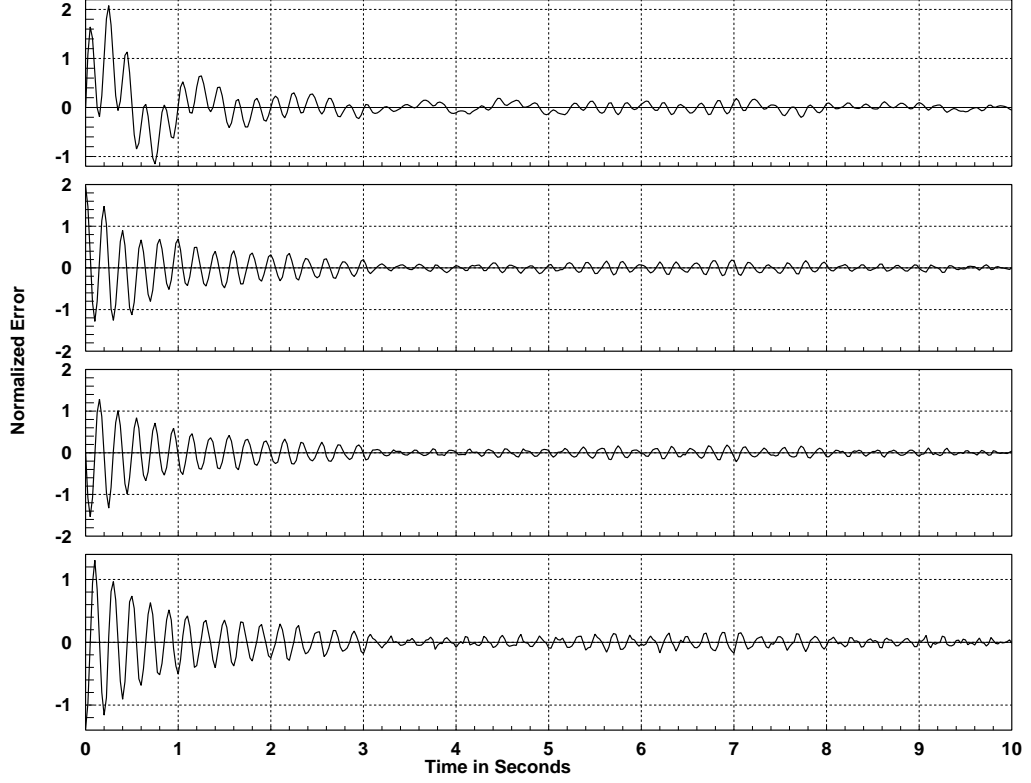
15

Figure 6: Normalized estimation errors in y, $\frac{dy}{dt}$, $\frac{d^2y}{dt^2}$ and $\frac{d^3y}{dt^3}$, respectively, incorporating model derivative constraints and damped Newton updates with interpolation/numerical differentiation.

Kalman filter to exist. The magnitudes of the steady state errors were essentially identical to those in Figure 5.

The normalized estimation errors resulting from the application of (4.4) in combination with (4.12) and (4.14) are shown in Figure 6. The parameter values were chosen as: $N = 5$, $\Delta t = 0.025$ sec., $W = 6$ and $K = 3$ in (4.4), $\eta = 4$ in (4.12) with a weight selected as

$$
Q = \begin{bmatrix}
1.26\ 10^{-1} & 0 & 0 & 0 & 0 \\
0 & 5.19\ 10^{-3} & 0 & 0 & 0 \\
0 & 0 & 1.54\ 10^{-4} & 0 & 0 \\
0 & 0 & 0 & 4.4\ 10^{-5} & 0 \\
0 & 0 & 0 & 0 & 1.33\ 10^{-6}
\end{bmatrix} ; \qquad (4.16)
$$

ϵ in (4.14) was set at 0.03.

4.3.   Third order, nonlinear dynamical system.

The goal of this example is to show that the method illustrated on the linear example above is also applicable to nonlinear models. For this purpose, consider the following system which is a "convex combination" of an unstable linear dynamics and a stable linear dynamics:

$$
\begin{bmatrix} \frac{dx_1}{dt} \\ \frac{dx_2}{dt} \\ \frac{dx_3}{dt} \end{bmatrix} = f(x) = \begin{bmatrix} x_2 \\ x_3 \\ f_3(x) \end{bmatrix} \qquad (4.17)
$$

16

$$y = h(x) = x_1,$$

where,

$$
\begin{aligned}
f_3(x) &= \alpha(x)g_s(x) + (1 - \alpha(x))g_u(x), \\
\alpha(x) &= \beta\frac{x'x}{1 + x'x} \\
g_s(x) &= -54x_1 - 36x_2 - 9x_3 \\
g_u(x) &= 54x_1 - 36x_2 + 9x_3.
\end{aligned}
$$

The parameter β will be fixed at 2 for now; in the final version of the paper, it will be assumed unknown.
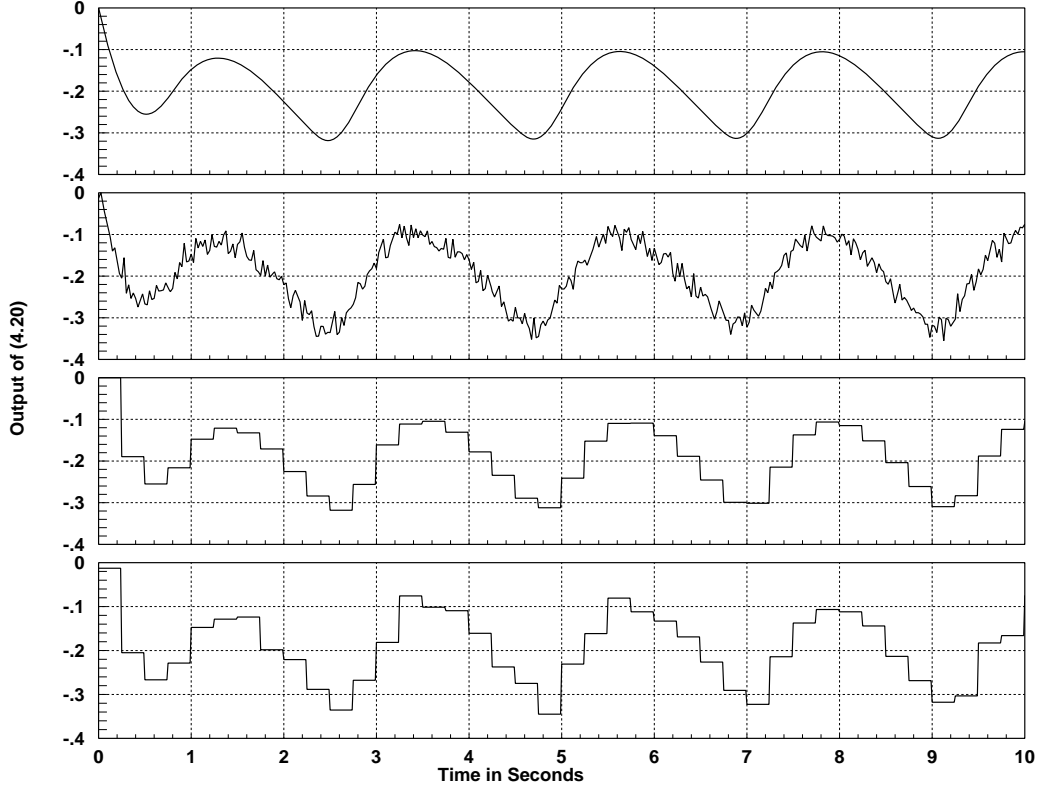

Figure 7: Time plots of $y(t)$, $y(t) + w(t)$, y_k^m and $y_k^m + w_k$, respectively.

The output of system (4.18) has a fundamental frequency of about 1 Hz., so the sample rate for the system was selected as 4 Hz.; this is probably a little too slow, but will challenge the observers to be designed. Figure 7 shows the output of the system (4.18), y, $y + w$, y_k^m and $y_k^m + w_k$, respectively, for $\Delta t = 0.25$ sec. and $\sigma = 0.0168$ (which corresponds once again to a 20/1 signal to noise ratio). Figure 8 shows the effectiveness of estimating the states on the basis of (4.4) with $N = 6$, $W = 8$, $K = 4$. The performance seems quite good.

A state estimator was also constructed using the interpolation/numerical differentiation method of (4.4), with (4.14) replaced by

$$\hat{x}_k^-(t_{k-W+K}) = F_{\Delta t}(\hat{x}_{k-1}^+(t_{k-1-W+K}))$$
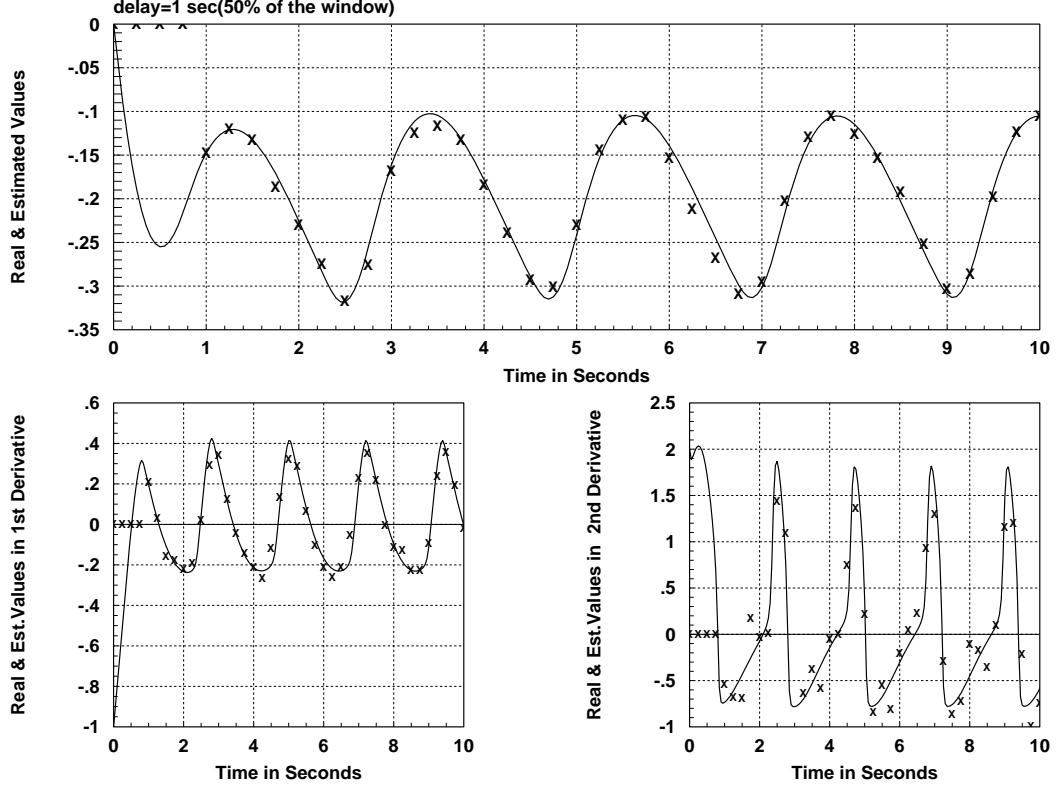
17

Figure 8: Comparison of estimated signals ($\times$)'s and true signals using interpolation/numerical differentiation.

$$\hat{x}_k^+(t_{k-W+K}) \quad = \quad \hat{x}_k^-(t_{k-W+K}) + \tag{4.18}$$

$$\epsilon \, (\frac{\partial}{\partial x} \begin{bmatrix} h \\ L_f h \\ L_f^2 h \end{bmatrix})_{|\hat{x}_k^-(t_{k-W+K})}^{-1} \qquad (\begin{bmatrix} \hat{y}_k(t_{k-W+K}) \\ \frac{d}{dt}\hat{y}_k(t_{k-W+K}) \\ \frac{d^2}{dt^2}\hat{y}_k(t_{k-W+K}) \end{bmatrix} - \begin{bmatrix} h \\ L_f h \\ L_f^2 h \end{bmatrix}_{|\hat{x}_k^-(t_{k-W+K})} \quad).$$

In the above, $F_{\Delta t}(x)$ is the sampled-data representation of (4.17) for $t = \Delta t$; for the estimator results which follow, it was computed by Euler integration with step size 0.0125 (i.e., 80 Hz. updates). The parameter ϵ was selected as

$$\epsilon = \begin{bmatrix} 0.6 & 0 & \\ 0 & 0.12 & 0 \\ 0 & 0 & 0.03 \end{bmatrix},$$

to reflect scaling in the magnitudes of the estimates of the derivatives and uncertainty due to noise. Figure 9 compares the estimated values of x_1, x_2 and x_3 to the true values.

One could explore replacing the model constraints (4.12) by

$$\frac{d^j}{dt^j}\hat{y}_k(t_{k-1-W+K}) = L_f^j h(\hat{x}_{k-1}(t_{k-1-W+K})), \tag{4.19}$$
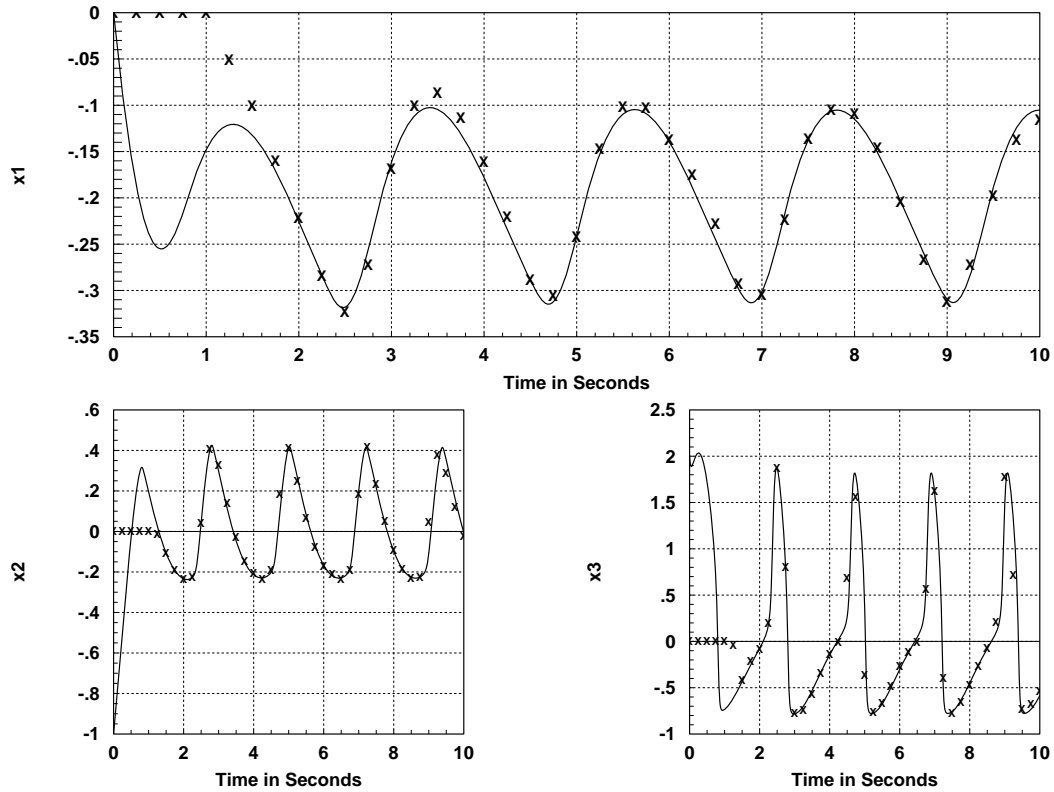
for $j = 0, \ldots, J$, but this was not done.

18

Figure 9: Comparison of estimated signals ($\times$)'s and true signals using (4.18). Note the reduced error with respect to Figure 8.

References

[1] *Proceedings of the Americal Control Conference*, 1992.

[2] R. S. Anderssen and P. Bloomfield. On the numerical differentiation of data. Technical report, Princeton University, No. 13, Ser. 2, 1972.

[3] G. Birkhoff and C. R. de Boor. Error bounds for spline interpolation. *J. Math. Mech.*, 13:827–835, 1964.

[4] P. L. Butzer. The Shannon sampling theorem and some of its generalizations: An overview. In Sendov et al. [42], pages 258–274.

[5] P. L. Butzer. A survey of the Whittaker-Shannon sampling theorem and some of its extensions. *J. Math. Res. Exposition*, 3:185–212, 1983.

[6] P. L. Butzer and R. J. Nessel, editors. *Fourrier Analysis and Approximation, I.* Birkhäuser, Basel, 1971.

[7] P. L. Butzer and W. Splettstösser. A sampling theorem for duration-limited functions with error estimates. *Inform. and Control*, 34:55–65, 1977.

[8] S. T. Chung and J. W. Grizzle. Sampled-data observer error linearization. *Automatica*, 26:997–1000, 1990.

[9] G. Criscuolo, B. Della Vecchia, and G. Mastroianni. Hermite-Fejér and Hermite interpolation. In Singh et al. [43].

[10] G. Criscuolo and G. Mastroianni. New results on Lagrange interpolation. In Singh et al. [43], pages 333–340.

[11] G. Criscuolo, G. Mastroianni, and D. Occorsio. Convergence of extended Lagrange interpolation. *Math. Comput.*, 55(191):197–212, July 1990.

[12] G. Criscuolo, G. Mastroianni, and D. Occorsio. Uniform convergence of derivatives of extended Lagrange interpolation. *Numer. Math.*, 60:195–218, 1990.

[13] J. Cullum. Numerical differentiation and regularization. *SIAM J. Numer. Anal.*, 8:254–265, 1971.

[14] J. Czipszer and G. Freud. Sur l'approximation d'une fonction périodique et ses dérivées successives par un polynome trigonométrique et par ses dérivées successives. *Acta Math.*, 99:33–51, 1958.

[15] C. de Boor. *A Practical Guide to Splines.* Springer-Verlag, New York, 1978.

[16] C. R. de Boor and R. E. Lynch. On splines and their minimum properties. *J. Math. Mech.*, 15:953–969, 1966.

[17] C. de la Vallée Poussin. *L'Approximation.* Chelsea, New York, 1970.

[18] S. Diop. Rational system equivalence, and generalized realization theory. In *Proceedings of the IFAC-Symposium NOLCOS'92*, pages 402–407, Bordeaux, 1992. IFAC.

[19] S. Diop and M. Fliess. On nonlinear observability. In C. Commault, D. Normand-Cyrot, J. M. Dion, L. Dugard, M. Fliess, A. Titli, G. Cohen, A. Benveniste, and I. D. Landau, editors, *Proceedings of the First European Control Conference*, pages 152–157, Paris, 1991. Hermès.

[20] V. Dolezal and R. Tewarson. Error bounds for spline–on–spline interpolation. *J. Approx. Theory*, 36:213–225, 1982.

[21] M. Golomb and H. F. Weinberger. Optimal approximation and error bounds. In R. E. Langer, editor, *On Numerical Approximation*, pages 117–190, Madison, 1959. University of Wisconsin Press.

[22] T. N. E. Greville. Introduction to spline functions. In T. N. E. Greville, editor, *Theory and Applications of Spline Functions*, pages 1–35, New York, 1969. Academic Press.

[23] J. W. Grizzle and P. E. Moraal. Newton observers and nonlinear discrete-time control. In *Proceedings of the 29th IEEE Conference on Decision and Control*, pages 760–767, New York, 1990. IEEE Press.

[24] J. W. Grizzle and P. E. Moraal. Observer based control of nonlinear discrete-time systems. Research note, University of Michigan, Ann Arbor, Michigan, College of Engineering, Control Group, February 1990.

[25] J. W. Grizzle and P. E. Moraal. On observers for smooth nonlinear digital systems. In A. Bensoussan and J. L. Lions, editors, *Analysis and Optimization of Systems*, Lect. Notes Control Inform. Sci., pages 401–410, Berlin, 1990. Springer-Verlag.

[26] G. I. Marchuk. *Methods of Numerical Mathematics.* Number 2 in Appl. Math. Springer-Verlag, New York, second edition, 1982.

[27] C. A. Micchelli. Cardinal *l*-splines. In S. Karlin, C. A. Micchelli, A. Pinkus, and I. J. Schoenberg, editors, *Studies in Splines and Approximation Theory*, New York, 1976. Academic Press.

[28] C. A. Micchelli. On an optimal method for the numerical differentiation of smooth funcions. *J. Approx. Theory*, 18:189–204, 1976.

[29] C. A. Micchelli, T. J. Rivlin, and S. Winograd. The optimal recovery of smooth functions. *Numer. Math.*, 26:191–200, 1976.

[30] P. E. Moraal and J. W. Grizzle. Nonlinear discrete-time observers using Newton's and Broyden's methods. In acc1992 [1], pages 3086–3090.

[31] P. E. Moraal and J. W. Grizzle. Observer design for nonlinear systems with discrete-time measurements. Research note, University of Michigan, Ann Arbor, Michigan, College of Engineering, Control Group, 1993.

[32] P. E. Moraal, J. W. Grizzle, and J. A. Cook. An observer design for single-sensor individual cylinder pressure control. In *Proceedings of the 32nd IEEE Conference on Decision and Control*, New York, 1993. IEEE Press.

[33] D. J. Newman. Numerical differentiation of smooth data. Technical report.

[34] T. R. F. Nonweiler. *Computational Mathematics: An Introduction to Numerical Approximation*. Ellis Horwood Ser. Math. Appl. John Wiley & Sons, New York, 1984.

[35] J. R. Rice. Nonlinear approximation. In H. L. Garabedian, editor, *Approximation of Functions*, pages 111–133, Amsterdam, 1965. Elsevier.

[36] S. Ries and R. L. Stens. Approximation by generalized sampling series. In Sendov et al. [42], pages 746–756.

[37] A. Sard. Optimal approximation. *J. Funct. Anal.*, 1:222–244, 1967.

[38] R. Schaback. Geometrical differentiation and high-accuracy curve interpolation. In Singh et al. [43], pages 445–462.

[39] I. J. Schoenberg. On interpolation by spline functions and its minimal properties. In *Proc. Conf. on Approximation Theory, Oberwolfach, Germany*, number 5 in Internat. Ser. Numer. Math., pages 109–129, New York, 1964. Academic Press.

[40] H. W. Schüssler, editor. *EUSIPCO-83, Signal Processing: Theories and Applications: Proceedings of the Second European Signal Processing Conference*, Amsterdam, 1983. North-Holland.

[41] D. Secrest. Error bounds for interpolation and differentiation by the use of spline funcions. *SIAM J. Numer. Anal.*, 2:440–447, 1965.

[42] B. Sendov et al., editors. *Constructive Function Theory '81: Proceedings of Conference, Varna, 1981*, Sofia, 1983. Publ. House Bulgarian Acad. Sci.

[43] S. P. Singh, A. Carbone, R. Charron, and B. Watson, editors. *Approximation Theory, Spline Functions and Applications*, number 356 in NATO ASI Ser. C Math. Phys. Sci., Boston, 1992. Kluwer Academic Publishers.

[44] Y. Song and J. W. Grizzle. The extended Kalman filter as a local asymptotic observer for nonlinear discrete-time systems. In acc1992 [1], pages 3365–3369.

[45] W. Splettstösser. Error estimates for sampling approximation of non-band limited funcions. *Math. Meth. Appl. Sci.*, 1:127–137, 1979.

[46] W. Splettstösser. 75 years aliasing error in the sampling theorem. In Schüssler [40].

[47] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*. Springer-Verlag, New York, 1980.

[48] A. K. Varma and K. L. Katsifarakis. Optimal error bounds for cubic spline interpolation. *, pages 214–229, 0000.