2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

New insights into derivative estimation via least squares approximation - theory and application

Philipp Mai and Claus Hillermeier

Abstract— **In this article, we revise a well-known derivative estimation scheme which is based on a least squared error polynomial approximation of a noisy measurement signal. Our contribution is to determine the influence of the estimation parameters onto the covariance matrix and the temporal delay of the estimation result. Also, it is shown that the least squares estimator is statistically optimal in the presence of white gaussian measurement noise. Our ideas are applied to the estimation of derivatives of the first state of Chen's Chaotic Oscillator and to the fault tolerant swing up of the Inverted Pendulum on a Cart.**

I. INTRODUCTION

The need for a fast and robust time derivative estimation regarding noisy measurement signals arises in a variety of disciplines, as in control engineering, fault diagnosis, signal processing and transmission. Various ideas for derivative estimation have been suggested in the past. In the continuous time case, popular approaches to derivative estimation are based on high gain observers, sliding mode observers or the so called 'algebraic' identification of the parameters of a spline approximation, see [1],[2],[9],[10] and references therein, for instance. In the discrete time case, the problem of derivative estimation was commonly viewed in the context of FIR filters, see [3], [8], [11] and the vast literature collected in [4], for instance.

In this paper, we revise the well established derivative estimation scheme that is based on a least squares polynomial approximation of the measurement signal, see [5], [6] for an overview. Our contribution is to determine the influence of the estimation parameters such as approximation interval length, sampling rate and degree of the approximating polynomial onto the variance and the delay of the estimation result. These questions have not been adressed so far, to the best of our knowledge, but appear as fundamental problems from our own practical experience. We show that in the presence of white gaussian measurement noise, the least squares estimator is optimal in the following sense: The variance of the estimated derivative lies exactly on the theoretical lower bound of any unbiased estimator, the so called Cramér Rao Bound (CRB) [13]. For large sampling numbers, a very good approximation of the covariance matrix of the polynomial's estimated parameter vector is then calculated, that gives insight into the effect of the estimation scheme parameters. In order to determine the temporal delay of the estimation result, it is assumed that on each approximation interval the clean measurement signal corresponds to a polynomial of one degree higher than the derivative to be estimated. For true signals not meeting this requirement, the taken assumption means that for analytical purposes the true signal is replaced by its truncated Taylor expansion, whose derivative (of the degree to be estimated) is a linear approximation of the true derivative. A general result for the delay is then derived. As a demonstration of our ideas, Chen's Chaotic Oscillator is simulated and derivatives of the first state component are estimated. Furthermore, our ideas are applied in the laboratory to the fault tolerant swing up of the Inverted Pendulum on a Cart. While the results in Section II are developed for the discrete time setting, in Section III the continuous time case is considered in order to facilitate the mathematical treatment. This does not represent a major restriction of the area of application, since the properties of the discrete time and continuous time estimators nearly do not differ if the sampling time is sufficiently small, which is assumed in the remainder of this work.

Our contribution is structured as follows: In Section II, we first recall the least squares estimation of polynomial coefficients before showing the optimality of the estimator. Afterwards, an approximate analytical expression of the covariance matrix is derived. The temporal delay of the estimation scheme is tackled in Section III. Section IV discusses the trade-off between parameters. Simulation results of the derivative estimation applied to Chen's Chaotic Oscillator are provided in Section V-A, and the fault tolerant swing up of the Inverted Pendulum on a Cart is described in Section V-B. We draw our conclusions in Section VI.

II. STATISTICAL PROPERTIES OF THE LEAST SQUARES ESTIMATION

In this section we start with recalling the derivative estimation scheme based on least squares in a discrete time setting. Afterwards, we show that, under the condition of white gaussian noise, the least squares derivative estimator is statistically optimal in that the variance of the estimated derivative lies exactly on the Cramér Rao Bound. Finally, by calculating an explicit expression for the covariance matrix of the unknown parameter vector of the signal's polynomial model, the effect of each estimation scheme parameter on the result is enlighted.

A. Basics of the estimation scheme

Let $y(t)$ be a noisy measurement of the clean signal $x(t)$,

$$y(t) = x(t) + \eta(t), \tag{1}$$

This work was supported by the CoTeSys project.

Philipp Mai and Claus Hillermeier are with the Faculty of Electrical Engineering, Institute of Automation and Control, University der Bundeswehr München, D-85579 Neubiberg, Germany. (pmai@ieee.org)

where $\eta(t)$ is assumed to be white, gaussian noise. The task is to estimate the j-th derivative of $x(t)$, $x^{(j)}(t)$. The result of the estimation be denoted by $\hat{x}^{(j)}(t)$. The estimation scheme is based on a local polynomial model of $x(t)$. Let M be its order, so that the polynomial model $x_M(t)$ is given by

$$x_M(t) = \sum_{i=0}^{M} a_i t^i. \tag{2}$$

The parameters of $x_M(t)$ are summarized in the vector

$$\theta = (a_0, a_1, \ldots, a_M)^{\mathrm{T}} \tag{3}$$

and have to be estimated on the basis of the measurement signal $y(t)$. Let $\hat{\theta}$ be the estimated parameter vector (with elements $\hat{a}_0, \hat{a}_1$ etc.). Suppose that $n+1$ samples have been measured on the time interval $[0, T]$ at equidistant instants $t = i \cdot T_s$, with T_s denoting the sampling time. The time interval may be shifted arbitrarily later on to allow for an online estimation scheme. Let y be the measurement vector, which is given by

$$y = (y(0), y(T_{\mathrm{s}}), y(2T_{\mathrm{s}}), \ldots, y(nT_{\mathrm{s}}))^{\mathrm{T}}, \tag{4}$$

with $T = nT_s$. Once $\hat{\theta}$ has been estimated, the sought derivative $\hat{x}^{(j)}(t)$ at time instant $t = T$ can be calculated by

$$\hat{x}^{(j)}(T) = \sum_{i=j}^{M} \frac{i!}{(i-j)!} \hat{a}_i T^{i-j} = c^T \hat{\theta} \tag{5}$$

with the row vector

$$c^T = (0,\ 0, ...,\ j!,\ (j+1)!T,\ ...,\ \frac{M!}{(M-j)!} T^{M-j}). \tag{6}$$

It is straightforward to show that in view of (5), the variance of $\hat{x}^{(j)}(T)$ is given by

$$\mathrm{Var}(\hat{x}^{(j)}(T)) = c^T C_{\hat{\theta}}\, c, \tag{7}$$

where $C_{\hat{\theta}}$ denotes the covariance matrix of $\hat{\theta}$. In view of (7), the performance of the derivative estimation depends on the quality of estimating θ, i.e. on the covariance matrix $C_{\hat{\theta}}$.

B. Optimality of the Least Squares estimator

In the absence of noise, the measurement would yield the data vector [1]

$$x = (x_M(0), x_M(T_{\mathrm{s}}), x_M(2T_{\mathrm{s}}), \ldots, x_M(nT_{\mathrm{s}}))^{\mathrm{T}}. \tag{8}$$

Estimating the parameter vector θ corresponds to adapting the linear model

$$x = V\,\theta \tag{9}$$

to the noisy measurement vector. The matrix V is given by [5]

$$V = \begin{pmatrix} 1 & 0 & \ldots & \ldots & 0 \\ 1 & T_{\mathrm{s}} & T_{\mathrm{s}}^2 & \ldots & T_{\mathrm{s}}^M \\ 1 & 2T_{\mathrm{s}} & (2T_{\mathrm{s}})^2 & \ldots & (2T_{\mathrm{s}})^M \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & nT_{\mathrm{s}} & (nT_{\mathrm{s}})^2 & \ldots & (nT_{\mathrm{s}})^M \end{pmatrix}. \tag{10}$$

[1]We consider a polynomial signal $x(t) = x_M(t)$ for the subsequent statistical analysis of the estimator.

The least squares estimate $\hat{\theta}$ minimizes the mean squared error between model and measurement vector y. In other words, it is the solution of the minimization problem

$$\hat{\theta} = \min_{\theta} \parallel y - V\theta \parallel^2. \tag{11}$$

Clearly, $\hat{\theta}$ is given by the pseudoinverse of V:

$$\hat{\theta} = (V^{\mathrm{T}}V)^{-1}V^{\mathrm{T}}\, y, \tag{12}$$

which is only defined for $n \geq M$.
It is a known result in the theory of estimation that linear estimation problems under white gaussian noise can be solved in a *statistically optimal* fashion by the least squares estimator. This is based on the fact that the covariance matrix of the least squares estimator lies exactly on the lower theoretical bound, the so called Cramér Rao Bound (CRB) [13], [12]. The CRB states that the following inequality holds for the covariance matrix $C_{\hat{\rho}}$ of any unbiased estimator of the parameter vector ρ:

$$C_{\hat{\rho}} - \mathcal{I}^{-1}(\rho) \geq 0, \tag{13}$$

where $\mathcal{I}(\rho)$ denotes the Fisher Information matrix and where the symbol '≥ 0' means that the symmetric matrix on the left hand side of (13) is positive semidefinite. Eq. (13) becomes an equality for the least squares estimator, and its covariance matrix $C_{\hat{\theta}}$ is given by

$$C_{\hat{\theta}} = \sigma^2 \left(V^{\mathrm{T}}V\right)^{-1} = \mathcal{I}^{-1}(\theta), \tag{14}$$

where σ^2 is the noise power.
We now show that the optimal estimation of the parameter vector θ transforms into an optimal estimation of the sought derivative $x^{(j)}(T)$. This is done by considering the general form of the CRB, which states that the estimation of a value α, that depends on the parameter vector θ through a known function

$$\alpha = g(\theta), \tag{15}$$

is reigned by the following inequality [13]:

$$\mathrm{Var}(\hat{\alpha}) \geq \frac{\partial g(\theta)}{\partial \theta} \mathcal{I}^{-1}(\theta) \left(\frac{\partial g(\theta)}{\partial \theta}\right)^T. \tag{16}$$

In our case, α corresponds to $x^{(j)}(T)$, and according to (5), the function $g(\theta)$ reduces to a dot product $x^{(j)}(T) = c^T\theta$. In light of this, the variance of $\hat{x}^{(j)}(T)$ is bounded by

$$\mathrm{Var}(\hat{x}^{(j)}(T)) \geq c^T \mathcal{I}^{-1}(\theta) c = c^T C_{\hat{\theta}}\, c. \tag{17}$$

The comparison of (17) und (7) shows that the least squares estimation scheme makes the variance of $\hat{x}^{(j)}(T)$ lie on the CRB and, therefore, estimates the derivative $x^{(j)}(T)$ in a statistically optimal way.

C. Covariance matrix

In order to gain more insight into the effect that the parameters of the estimation scheme (i.e. sampling time T_s, window length T, polynomial order M) have onto the variance of the estimation result, an explicit formula for the covariance matrix $C_{\hat{\theta}}$ is needed. For that purpose, the matrix

$V^{\mathrm{T}}V$ has to be inverted. An approximate general inverse of $V^{\mathrm{T}}V$ can be found whenever the number of samples $n+1$ is sufficiently large, e.g. $n > 100$, which is quite in accordance with practical settings.
To this end, note that the element V_{ij} of matrix V is

$$V_{ij} = ((i-1)T_{\mathrm{s}})^{j-1}, \tag{18}$$

and, therefore,

$$\begin{aligned}\left[V^{\mathrm{T}}V\right]_{ij} &= \sum_{\nu=1}^{n+1} (T_{\mathrm{s}}(\nu-1))^{i+j-2} \\ &= T_{\mathrm{s}}^{i+j-2}\left(\delta_{i1}\delta_{j1} + \sum_{\nu=1}^{n} \nu^{i+j-2}\right),\end{aligned} \tag{19}$$

where $\delta_{i1}\delta_{j1} = 1$ only if $i = j = 1$, elsewhere it is zero.
To the knowledge of the authors, there is no result for a general inverse of $V^{\mathrm{T}}V$ yet.
From geometric reasoning, it is obvious that

$$\int_0^n x^k \mathrm{d}x = \tfrac{n^{k+1}}{k+1} < \sum_{i=1}^{n} i^k < \int_1^{n+1} x^k \mathrm{d}x = \tfrac{(n+1)^{k+1}-1}{k+1} < \tfrac{(n+1)^{k+1}}{k+1}, \tag{20}$$

since x^k is strictly monotonous for $k > 1, x \geq 0$. This results in the inclusion

$$1 < \frac{k+1}{n^{k+1}} \sum_{i=1}^{n} i^k < \frac{(n+1)^{k+1}}{n^{k+1}} = (1+\tfrac{1}{n})^{k+1}. \tag{21}$$

Hence, it is is clear that for large n

$$\sum_{i=1}^{n} i^k \approx \frac{n^{k+1}}{k+1} \tag{22}$$

is a valid approximation. Therefore, the matrix elements $\left[V^{\mathrm{T}}V\right]_{ij}$ may be approximated by

$$\left[V^{\mathrm{T}}V\right]_{ij} \approx [\widetilde{V^{\mathrm{T}}V}]_{ij} = T_{\mathrm{s}}^{i+j-2}\frac{n^{i+j-1}}{i+j-1}. \tag{23}$$

Note that $\left[V^{\mathrm{T}}V\right]_{11} = n+1$ whereas $[\widetilde{V^{\mathrm{T}}V}]_{11} = n$, which is also a reasonable approximation for large n.
With (23) the factorization of the matrix $\widetilde{V^{\mathrm{T}}V}$ is clear:

$$\widetilde{V^{\mathrm{T}}V} = \begin{pmatrix} n & \frac{n^2}{2}T_{\mathrm{s}} & \frac{n^3}{3}T_{\mathrm{s}}^2 & \iddots & \frac{n^{M+1}}{M+1}T_{\mathrm{s}}^M \\ \frac{n^2}{2}T_{\mathrm{s}} & \frac{n^3}{3}T_{\mathrm{s}}^2 & \frac{n^4}{4}T_{\mathrm{s}}^3 & \iddots & \frac{n^{M+2}}{M+2}T_{\mathrm{s}}^{M+1} \\ \frac{n^3}{3}T_{\mathrm{s}}^2 & \frac{n^4}{4}T_{\mathrm{s}}^3 & \frac{n^5}{5}T_{\mathrm{s}}^4 & \iddots & \frac{n^{M+3}}{M+3}T_{\mathrm{s}}^{M+2} \\ \iddots & \iddots & \iddots & \iddots & \iddots \\ \frac{n^{M+1}}{M+1}T_{\mathrm{s}}^M & \frac{n^{M+2}}{M+2}T_{\mathrm{s}}^{M+1} & \frac{n^{M+3}}{M+3}T_{\mathrm{s}}^{M+2} & \iddots & \frac{n^{2M+1}}{2M+1}T_{\mathrm{s}}^{2M} \end{pmatrix}$$

$$= \frac{1}{nT_{\mathrm{s}}^2} D \underbrace{\begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \iddots & \frac{1}{M+1} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \iddots & \frac{1}{M+2} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \iddots & \frac{1}{M+3} \\ \iddots & \iddots & \iddots & \iddots & \iddots \\ \frac{1}{M+1} & \frac{1}{M+2} & \frac{1}{M+3} & \iddots & \frac{1}{2M+1} \end{pmatrix}}_{=:\, H_{M+1}} D = \frac{D\,H_{M+1}D}{nT_{\mathrm{s}}^2} \tag{24}$$

with the diagonal matrix

$$D = \mathrm{diag}\left(nT_{\mathrm{s}}, (nT_{\mathrm{s}})^2, \ldots, (nT_{\mathrm{s}})^{M+1}\right) \tag{25}$$

and the Hilbert-matrix H_{M+1} of dimension $M+1$.
A general inverse of H_{M+1} is given by (see [7])

$$[H_{M+1}^{-1}]_{ij} = (-1)^{i+j}(i+j-1)\tbinom{M+i}{M+1-j}\tbinom{M+j}{M+1-i}\tbinom{i+j-2}{i-1}^2. \tag{26}$$

Due to (24), the sought inverse reads

$$\left(\widetilde{V^TV}\right)^{-1} = D^{-1}H_{M+1}^{-1}D^{-1}(nT_{\mathrm{s}}^2), \tag{27}$$

and its components are given by

$$\boxed{\begin{aligned}\left[\left(\widetilde{V^TV}\right)^{-1}\right]_{ij} &= (-1)^{i+j}(i+j-1)\frac{1}{n^{i+j-1}}\frac{1}{T_{\mathrm{s}}^{i+j-2}}\times \\ &\binom{M+i}{M+1-j}\binom{M+j}{M+1-i}\binom{i+j-2}{i-1}^2.\end{aligned}} \tag{28}$$

With the approximate representation of the inverse according to (28), it is clear that by increasing the number of samples, that is by increasing n, the entries of the covariance matrix can be reduced arbitrarily close to zero, corresponding to an infinite noise damping. The identity

$$\frac{1}{n^{i+j-1}}\frac{1}{T_{\mathrm{s}}^{i+j-2}} = \frac{T_s}{T^{i+j-1}}, \tag{29}$$

shows explicitly that for a fixed window length T the sampling time T_s should be chosen as small as technically possible, whereas for a fixed sampling time increasing the window length T steadily reduces the noise. In those cases where the number of samples $n+1$ is fixed due to limited storage capacity, a larger sampling time T_s is preferable. One should additionally bear in mind that the polynomial approximation of the original signal $x(t)$ is only locally valid, so that T should not be chosen too large. It is furthermore not recommendable to increase the polynomial order M of the signal model $x_M(t)$ far beyond the sought derivative order j, because the binomial coefficients in (28) will tend to be very large. This effect means that increasing M will cause the approximating polynomial to actually *model* the noise.
Equation (28) also explains why estimating higher derivatives is highly sensitive to noise. Let us suppose $M = j$, which minimizes the variance of the estimation result. In this case, the vector c relating $C_{\hat{\theta}}$ and $\mathrm{Var}(\hat{x}^{(j)}(T))$ according to (7) reads $c^T = (0, .., 0, M!)$. Therefore, the element $(M+1, M+1)$ of $\left(\widetilde{V^TV}\right)^{-1}$ times $(M!)^2$ corresponds to the approximate noise gain when the M-th derivative is estimated by approximation with minimal order M. For $n = 250$ and $T_s = 0.001$, which are typical values, the approximate noise gain is equal to 0.77 for the estimation of the first derivative ($M = 1$), equal to 737.3 for the second derivative ($M = 2$) and equal to $1.65 \cdot 10^6$ for the third derivative ($M = 3$). It is for that reason that estimates of the

third or even higher derivatives are rarely seen in practice [2].

III. DELAY OF THE LEAST SQUARES ESTIMATOR

Estimating a signal's derivative by polynomial approximation in an online fashion induces a temporal delay due to the fact that the true signal contains in its Taylor expansion terms of higher degree than the approximating polynomial. Since this delay is not caused by noise, we discuss it for the noiseless case. While we are considering the time continuous approximation here, the time discrete approximation was discussed in section II. This does not affect the validity of the following analysis, since for small sampling times T_s the time discrete case converges to the time continuous case.

For describing the online scenario, the present time can - without loss of generality - be denoted by $t = T$, in accordance with the previous chapter. The estimation result available at $t = T$ is based on the signal information gathered during the past time interval $[0, T]$. Let us assume that the j-th derivative at $t = T$ of the true signal $x(t)$ is to be estimated; the approximation polynomial is denoted by $\hat{x}(t)$. If there is a time $\tau \in [0, T]$ for which $\hat{x}^{(j)}(T) = x^{(j)}(T - \tau)$ holds, then the estimation result obtained for time T reproduces the derivative of the *real* signal at time $T - \tau$. Therefore, we call this time interval τ the *temporal delay* of the estimation scheme.

In order to make analytical calculations possible, certain assumptions about the true signal have to be made. We assume that the true noiseless signal $x(t)$ is a polynomial of degree $N = j + 1$ in the time interval $[0, T]$. As mentioned earlier, this assumption means that for analytical purposes the original true signal is replaced by its truncated Taylor expansion, whose j-th derivative is a linear approximation of the j-th derivative of the true original signal. For small T and sufficiently smooth signals, the analytical results based on that assumption will approximately hold for the original signal as well.

In the following, by 'true signal $x(t)$' we will denote a polynomial of degree N. By least squares estimation, $x(t)$ is approximated by a polynomial $\hat{x}(t)$ of degree M. Two choices of M make sense: Recalling that the $j-th$ derivative is to be estimated, the choice $M = j = N - 1$ is the minimal possible order. According to section II-C, this choice minimizes the noise of the estimation result but, as will be shown below, induces a temporal delay. Alternatively, we might opt for $M = j + 1 = N$ at the cost of getting more noise into the estimation result. For that choice, no delay appears in the estimation for small T, since the true signal $x(t)$ would be perfectly approximated in the noiseless case.

We now consider $M = j$. Let L^2 be the Hilbert space of quadratically integrable real functions on $[0, T]$. For each $n \in \mathbb{N}_0$ let V_n be the subspace of L^2 of real-valued polynomials of degree equal to or less then n. Let P_n be the Legendre polynomial of degree n on the interval $[-1, 1]$, and let Q_n be defined by

$$Q_n(t) = P_n((\frac{2t}{T} - 1)),\ \forall t \in [0, T]. \tag{30}$$

Clearly, the set of $\{Q_i \mid i = 0, .., n\}$ defines an orthogonal basis of V_n. Thus $x(t)$ can be written as

$$x(t) = \sum_{i=0}^{M+1} a_i Q_i(t), \ \ a_i = \frac{< x, Q_i >}{\parallel Q_i \parallel^2}, \tag{31}$$

where

$$< x(t), y(t) >= \int_0^T x(\tau) y(\tau) \mathrm{d}\tau, \tag{32}$$

$$\parallel x(t) \parallel = \sqrt{\int_0^T x(\tau)^2 \mathrm{d}\tau}. \tag{33}$$

As $\hat{x}(t)$ is the least squares approximation of $x(t)$, it corresponds to the element in V_M that minimizes $\parallel x(t) - y(t) \parallel$, $y(t) \in V_M$. Therefore, $\hat{x}(t)$ is the orthogonal projection of $x(t)$ onto V_M, which is given by

$$\hat{x}(t) = \sum_{i=0}^{M} a_i Q_i(t). \tag{34}$$

Theorem 1 *Let $x(t)$ be a polynomial of degree $M + 1$ on the time interval $[0, T]$, and let $\hat{x}(t) \in V_M$ be its orthogonal projection onto V_M. The M-th derivative of $\hat{x}(t)$ at $t = T$, $\hat{x}^{(M)}(T)$ will be subject to a delay of $\frac{T}{2}$ compared to the M-th derivative of the true signal, $x^{(M)}(t)$. In other words, $\hat{x}^{(M)}(T) = x^{(M)}(\frac{T}{2})$ holds.*

Proof: We have

$$\begin{aligned} x^{(M)}(\frac{T}{2}) &= \hat{x}^{(M)}(\frac{T}{2}) + a_{M+1} Q_{M+1}^{(M)}(\frac{T}{2}) \\ &= \hat{x}^{(M)}(T) + a_{M+1} Q_{M+1}^{(M)}(\frac{T}{2}) \\ &= \hat{x}^{(M)}(T) + a_{M+1} \left(\frac{2}{T}\right)^M P_{M+1}^{(M)}(0), \end{aligned}$$

[2]For given $n = 250$, $T_s = 0.001$ and M, the exact noise gains can be calculated instead of the approximate ones, since in this case the inverse of $(V^T V)$ is known. The exact noise gains are found as 0.76 for $M = 1$, 722.8 for $M = 2$ and $1.61 \cdot 10^6$ for $M = 3$. The small difference between exact and approximate noise gains underlines the correctness of the approximation of $(V^T V)^{-1}$.

where the second equals sign is due to fact that $\hat{x}(t)$ is a polynomial of degree M and, therefore, $\hat{x}^M(t) = \text{constant}$. It remains to show that $P_{M+1}^{(M)}(0) = 0$. The Legendre Polynomials are defined by

$$P_{M+1}(x) = c_{M+1}\frac{d^{M+1}}{dx^{M+1}}[(x^2-1)^{M+1}], c_{M+1} = \frac{2^{-(M+1)}}{(M+1)!}.$$

It follows

$$\begin{aligned} P_{M+1}^{(M)}(x) = & \quad c_{M+1}\tfrac{d^{2M+1}}{dx^{2M+1}}[(x^2-1)^{M+1}] \\ = & \ c_{M+1}\tfrac{d^{2M+1}}{dx^{2M+1}}\textstyle\sum_{i=0}^{M+1}\binom{M+1}{i}x^{2i}(-1)^{M+1-i} \\ = & \ c_{M+1}\tfrac{d^{2M+1}}{dx^{2M+1}}[x^{2M+2}-(M+1)x^{2M}+..] \\ = & \quad c_{M+1}(2M+2)!\,x, \end{aligned}$$

from which $P_{M+1}^{(M)}(0) = 0$ follows. That finishes the proof. ■

For the applications, we restate this result in the following corollary [3]:

Corollary 1 *If the $j-th$ derivative of a polynomial of degree $j+1$ is estimated by approximation with a polynomial of degree j on a time interval of length T by minimizing the mean squared error, the resulting derivative estimation is subject to a delay of length $\frac{T}{2}$.*

IV. TRADE-OFF BETWEEN PARAMETERS

The results of sections II-C and III should be taken into account when choosing the parameters of the derivative estimator. Note that in general a discrete time implementation will be used. The parameters to be designed are the degree M of the approximating signal, the sampling time T_s and the window length T. While no generally applicable optimal tuning rule for the parameters exist, since their effects are cross-coupled, we suggest the following procedure: By reducing T_s, the continuous-time case is approached, and the noise can be eliminated arbitrarily, see (28) in combination with (29). Therefore, T_s should always be reduced to the minimum feasible value, depending on the application, unless the sampling number $n+1$ is fixed due to memory capacity reasons - in that special case, a large sampling time is preferable. Secondly, T and M have to be designed. The choice of M depends on the fact whether the estimation result is used in an online application and whether it is fed back in the closed control loop. By choosing $M = j$, with j being the derivative order to be estimated, the noise is further minimized, but a temporal delay in the estimation result will appear that is approximately equal to $\frac{T}{2}$, since for small T the true signal can be well approximated by a polynomial of degree $M+1$ - see the next sections for application examples. Therefore, the tolerable delay of the application has to be identified, and T chosen accordingly. If no delay can be tolerated due to stability reasons, $M = j+1$ has to be chosen. In this case, the estimation will be subject to more noise, but almost no delay will appear. Note that for $M = j+1$, T cannot be raised arbitrarily in order to further reduce the noise, since for too large T, the true signal can not be modeled as a polynomial of degree $j+1$ anymore, which will in turn create a delay. Increasing M further beyond $j+1$ leads to very noisy results, unless the sampling time T_s is extremely small, which is usually not the case. If the estimation result is used for an offline application, the temporal delay does not represent a problem, since the estimation result can be shifted by the amount of the delay. Therefore, in this case the noise reduction is the main concern, and M should be chosen as $M = j$. Here, the choice of T depends on the noise level: on the one hand, increasing T reduces the noise arbitrarily, but on the other hand, only for a limited time interval the M-th order Taylor approximation coincides well with the measured signal. Therefore, T should slowly be increased starting from small values, until the estimation result is sufficiently smooth.

[3]Since $\hat{x}^M(t) = \text{constant}$, the preceding proof equally shows that $x^{(M)}(\frac{T}{2}) = \hat{x}^{(M)}(\frac{T}{2})$. Therefore, one might conclude that the discussed derivative estimation scheme is delayless. This would indeed be a false conclusion, since the approximation is carried out on the time interval $[0, T]$ in an online fashion, and the result of the approximation will only be available at $t = T$.

V. APPLICATION EXAMPLES

In this section, the preceding results are applied to two benchmark systems of the control community. First, Chen's Chaotic Oscillator serves as a system to demonstrate that the results concerning the delay apply even when the examined signal is of chaotic nature - noise is omitted during the simulation, since the focus is on the estimation's delay. Second, we apply our results in a laboratory experiment to the Inverted Pendulum on a cart. The pendulum performs a swing up under faulty conditions. The faults are identified in an offline fashion during a diagnostic sidestep before the swing up, making use of the derivative estimation and the presented insights about its delay.

A. Chen's Chaotic Oscillator

As a demonstration system for our delay considerations, we consider Chen's Chaotic oscillator, which is given by

$$\begin{aligned} \dot{x}_1 &= a(x_2 - x_1) \\ \dot{x}_2 &= (c-a)x_1 + cx_2 - x_1x_3 \\ \dot{x}_3 &= x_1x_2 - bx_3. \end{aligned} \tag{35}$$

The first and second derivative of x_1 are estimated and plotted against their true values. Clearly, $\ddot{x}_1$ is given by

$$\ddot{x}_1 = a(\dot{x}_2 - \dot{x}_1) = a(c(x_1 + x_2) - ax_2 - x_1 x_3). \tag{36}$$

We used the following set of parameters: $a = 35, b = 3, c = 28, x_1(0) = -10, x_2(0) = 0, x_3(0) = 20$. The sampling time was chosen as $T_s = 0.001$, which is a typical value in control applications. In figure 1, $\dot{x}_1$ is plotted versus its estimation for $T = 0.05$, while a linear and a quadratic approximation was used, e.g. $M = 1, 2$. The appearing delay corresponds to $T/2 = 0.025$ for the linear approximation, as predicted. Furthermore, for $M = 2$, we observe almost no delay, but the approximation shows little overshooting at the extremums of $\dot{x}_1$. This can be explained by the fact that the extremums of $\dot{x}_1$ correspond to turning points of $x_1(t)$. Near those points the cubic term of the Taylor series expansion is essential for characterizing the signal $x_1(t)$.

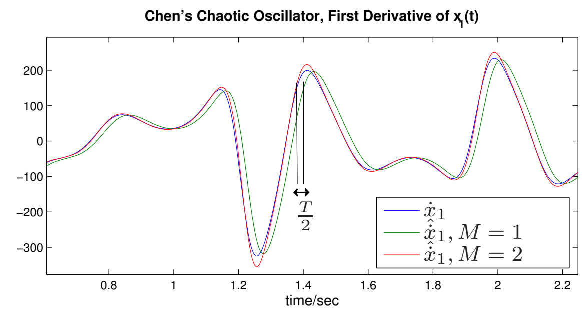

Fig. 1. True $\dot{x}_1$ and its estimations for $M = 1$ and $M = 2$.

In figure 2, the second derivative of $x_1(t)$ is estimated, by quadratic and cubic approximation. The same observations are made as in the case of the first derivative estimation. Here, the sampling time was reduced to $T_s = 5e - 4$, and the approximation length was chosen as $T = 0.03$. Again, the predicted delay appears when the approximating polynomial is of the same order as the derivative to be estimated.

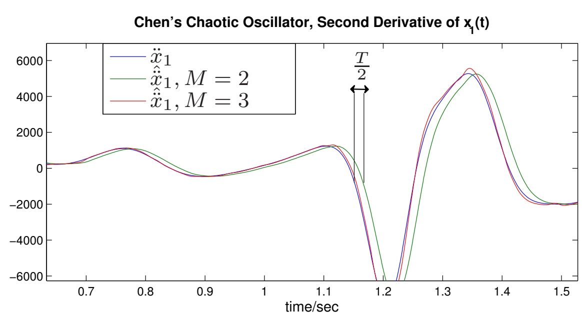

Fig. 2. True $\ddot{x}_1$ and its estimations for $M = 1$ and $M = 2$.

B. Experimental Validation - Fault Tolerant Swing Up of the Inverted Pendulum on a Cart via offline Diagnosis

As a laboratory application of the preceding analysis, the Inverted Pendulum on a Cart is swung up in the presence of a fault. For this purpose, a diagnostic sidestep is performed before the swing up maneuver, allowing for an offline analysis of possible faults which could be an actuator performance loss, a cart position sensor offset, or an angle sensor offset. The faults were implemented as software faults and were chosen to be constant, but unkown to the diagnosis unit. After the side step, the pendulum is swung up in open loop and afterwards stabilized by a linear controller in the upward position. During side step and stabilization, the control signals were modified in such a way that the effect of the estimated fault was neutralized.
The system is governed by the following set of equations [14]:

$$\ddot{\phi} = \frac{m}{J}(gl \sin\phi - l\dot{u} f_a \cos\phi) - d\dot{\phi} \tag{37}$$
$$\dot{x} = u f_a \tag{38}$$
$$x_m = x + f_x \tag{39}$$
$$\phi_m = \phi + f_\phi, \tag{40}$$

where ϕ is the angle of the pole to the vertical position, x is the cart's position, u is the control input, which corresponds to the cart's speed, x_m and ϕ_m denote the measured cart position and pole angle, J is the moment of inertia of the pendulum with respect to the axis of rotation, m is its mass, l is the distance of the pivot point of the pole to the center of its mass, g is the acceleration due to gravity and d is a friction constant. The constant faults are called f_a (multiplicative actuator fault), f_x and f_ϕ (additive sensor faults). Clearly, in the faultfree case, $f_a = 1, f_x = 0, f_\phi = 0$. The parameters were identified as (all values in SI-units) $J/m = 0.294, l = 0.43, d = 0.064$.
First, a reference trajectory $x^\star_{\text{up}}(t), t \in [0, T_{\text{up}}]$ corresponding to the swing-up maneuver was designed by numerical solution of (37) with the help of MATLAB's *bvp4c.m*, see [15] for details - $T_{\text{up}} = 2.1\,\text{sec}$ was chosen. The swing up is preceded by a sidestep of 0.75m, which brings the pendulum from the downward equilibrium state at $x = 0.6\text{m}$ to the downward equilibrium state at $x = -0.15\text{m}$ during a transfer time of $T_{\text{side}} = 2.0\,\text{sec}$. The swing up starts at $x = -0.15\text{m}$, accordingly. The corresponding reference trajectory of $x(t)$ during side step is denoted by $x^\star_{\text{side}}(t), t \in [-T_{\text{side}}, 0]$, and was designed in the same way as $x^\star_{\text{up}}(t)$. Clearly, the corresponding nominal control inputs are given by $u^\star_{\text{up}}(t) = \dot{x}^\star_{\text{up}}(t)$ and $u^\star_{\text{side}}(t) = \dot{x}^\star_{\text{side}}(t)$ according to (38). During the side step, two residuals were generated in the

time interval $t \in [-1.7, -0.3]$, making use of the derivative estimation:

$$r_1(t) = \hat{\dot{x}}(t) - u^{\star}_{\text{side}}(t) \tag{41}$$

$$r_2(t) = \frac{J}{m}\hat{\ddot{\phi}}(t + \frac{T}{2}) + \frac{J}{m}d\hat{\dot{\phi}}(t) - \left(gl\sin\phi_m(t) - l\dot{u}^{\star}_{\text{side}}(t)\cos\phi_m(t)\right), \tag{42}$$

where $\hat{\dot{x}}(t)$, $\hat{\dot{\phi}}(t)$ and $\hat{\ddot{\phi}}(t)$ denote the respective derivative estimates of $x(t)$ and $\phi(t)$ on the basis of the measured signals $x_m(t)$ and $\phi_m(t)$ [4]. The approximation window T was chosen as $T = 0.16$. Since $\hat{\ddot{\phi}}(t)$ was estimated by quadratic approximation, i.e. $M = 2$, the occuring delay of $\frac{T}{2}$ had to be taken into account by shifting $\hat{\ddot{\phi}}(t)$ by $-\frac{T}{2}$ in (42). Naturally, this could only be done in an offline fashion. The estimates of the first derivative were also calculated via quadratic approximation ($M = 2$), and therefore no delay appears. A digital implemention was used with a sampling time of $T_{\text{s}} = 0.001$. At $t = -0.3$, the mean and variance of $r_1(t)$ and $r_2(t)$ were evaluated. Under the condition that only a single fault appears at once, a simple logic, based on checking whether the mean and variance of $r_1(t)$ and $r_2(t)$ exceed certain thresholds, determines if a fault is present and isolates the fault. Fig. 3 plots $r_2(t)$ for the fault-free case and the case where the actuator has a performance loss of 20%. Taking the variance of $r_2(t)$, we can state a clear increase from $6.4 \cdot 10^{-4}$ for the fault-free case to 0.024 for the faulty case. Therefore, taking a threshold of 0.001 allows to detect whether an actuator fault is present or not[5]. In order to show the occuring delay, Fig. 4 displays the second derivative of the reference angle trajectory, $\ddot{\phi}^{\star}_{\text{side}}$ and the estimation $\hat{\ddot{\phi}}$ for different time windows T in the absence of faults. $\phi^{\star}_{\text{side}}(t)$ is obtained by the numerical solution of (37).

After a possible fault has been isolated, its magnitude is then estimated by

$$\hat{f}_a = \frac{1}{1.4}\int_{-1.7}^{-0.3} \frac{\hat{\dot{x}}(\tau)}{u^{\star}_{\text{side}}(\tau)}\mathrm{d}\tau \tag{43}$$

$$\hat{f}_x = \frac{1}{1.4}\int_{-1.7}^{-0.3} \left(x_m(\tau) - x^{\star}_{\text{side}}(\tau)\right)\mathrm{d}\tau \tag{44}$$

$$\hat{f}_\phi = \frac{1}{1.4}\int_{-1.7}^{-0.3} \left(\phi_m(\tau) - \phi^{\star}(\tau)\right)\mathrm{d}\tau. \tag{45}$$

Having estimated the respective fault, the open loop swing up of the pole and its closed loop control in the upward position can be adapted by replacing the control signal $u(t)$ by $\frac{u(t)}{\hat{f}_a}$ and the measurement signals $\phi_m(t), x_m(t)$ by $\phi_m(t) - \hat{f}_\phi, x_m(t) - \hat{f}_x$.

We'd like to point out that the calculation of $r_2(t)$ is not necessary in the present set-up. Instead, one could have directly estimated $\hat{f}_a$ in any case. Our intention in doing so is to provide a laboratory demonstration of a successful estimation of the second derivative, including a shift of the magnitude of the occuring delay.

The swing up was sucessfully performed for all kinds of faults, and only one representative result is depicted in Fig. 5, where the actuator was subject to a performance loss of 20%.

[4]In this case, the measurement noise characteristics were unkown, since the angle sensor was not modelled in detail. To some extent, a high proportion of quantization noise can be assumed. Still, the properties of the derivative estimation of the previous chapters could be confirmed.

[5]To be more precise, in the case of an actuator fault, both residuals $r_1(t)$ and $r_2(t)$ deviate from zero, which is checked by the diagnostic logic.

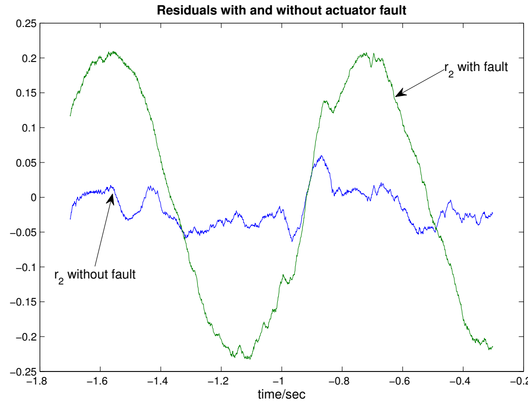

Fig. 3. Residuals $r_2(t)$ in fault-free case and for an actuator loss of 20%.

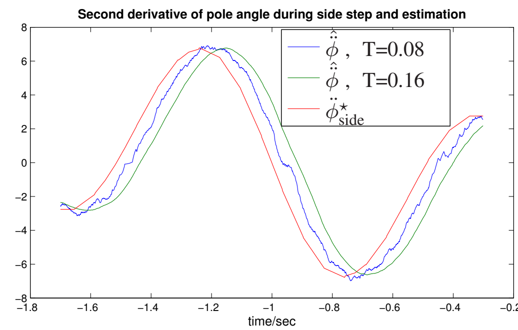

Fig. 4. Second derivative $\ddot{\phi}^{\star}_{\text{side}}(t)$ and its estimation for $M = 2$ and two choices for T. Clearly, increasing T raises the predicted delay of $\frac{T}{2}$, while the noise attenuation improves, see (28) together with (29).

VI. CONCLUSIONS

The well-known least squares polynomial approximation scheme for estimating the derivative(s) of a measured noisy signal has been analysed with respect to its statistical proper-

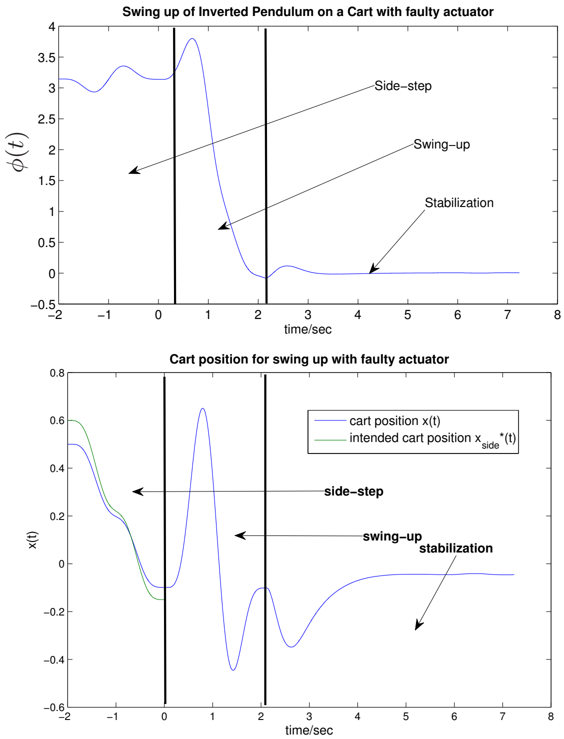

Fig. 5. Side-step, swing-up and stabilization of the pole under faulty conditions: pole angle (top) and cart position (bottom). The actuator loss of 20% was successfully identified during the side-step and could be compensated for later on. The initial cart position is shifted by $0.1m$ in order to avoid an exceeding of the available rail length.

ties and its temporal delay. The analysis gives important hints for the optimal tuning of the parameters of the estimator, such as sampling time, approximation window length and order of the polynomial approximation. While no generally applicable optimal set of parameters exists, a systematic way to design the estimator is given in Section IV, which takes into account the requirements of the given application. The results have been validated by two sample problems: The simulated estimation of the first and second derivative of the first state component of Chen's Chaotic Oscillator and the experimental swing up of the Inverted Pendulum on a Cart in the presence of actuator and sensor faults, where the fault identification and isolation scheme was based on the described derivative estimation method. The results in Sections II and III have been derived under the assumption that the true, noiseless measurement signal is of polynomial type of known order on each time interval T. Even more, the results concerning the estimation's delay (section III) are based on a polynomial model of order $j+1$ when the j-th derivative is to be estimated. The effect that might be caused by neglecting the higher order terms of the signal's Taylor expansion has not been considered in this work. Therefore, the preceding results are only valid for signals with sufficiently fast converging Taylor series on each interval T.

VII. ACKNOWLEDGMENTS

The authors gratefully acknowledge the very helpful contributions of Rainer v. Chossy, Georg Rappl and Berthold Lankl and the support within the DFG excellence initiative research cluster "Cognition for Technical Systems" - CoTeSys. Also, we would like to thank the reviewers for their helpful comments.

REFERENCES

[1] L.K. Vasiljevic and H. Khalil, Differentiation with High-Gain Observers the Presence of Measurement Noise, Proceedings of the 45th IEEE Conference on Decision&Control, 2006.

[2] N. Orani, A. Pisano and E. Usai, On a new sliding-mode differentiation scheme, IEEE International Conference on Industrial Technology, 2006.

[3] M.A. Al-Alaoui, Novel Approach to Designing Digital Differentiators, *Electronics Letters*, Vol. 28, No. 15, 1992, pp. 1376–1378.

[4] Dr. Yang Quan Chen's Digital Differentiator (DD) Webpage: http://www.ece.usu.edu/csois/people/yqchen/dd/index.html

[5] S. Diop, J.W. Grizzle, P.E. Moraal, and A. Stefanopoulou, "Interpolation and Numerical Differentiation for Observer Design", *1994 American Control Conference*, Baltimore, 1994, pp. 1329–1333.

[6] S. Diop, J.W. Grizzle and F. Chaplais, "On Numerical Differentiation Algorithms for Nonlinear Estimation", *Proc. of the 2000 IEEE Conference on Decision and Control*, Australia, December 2000.

[7] Gene H. Golub, Charles F. Van Loan, *Matrix computations*, Johns Hopkins University Press, 3rd edition, 1996.

[8] B. Kumar, S.C. Dutta Roy, Design of efficient FIR digital differentiators for low frequencies, *Proc. of the IEEE*, Vol. 76, No. 3, 1988, pp. 287–289.

[9] M. Mboup, C. Join, M. Fliess, "A revised look at numerical differentiation with an application to nonlinear feedback control", MED'07 Workshop, Athens, 2007.

[10] J. Reger, H. Sira-Ramírez, M. Fliess, On non-asymptotic observation of nonlinear systems, Proc. of the 44th IEEE Conference on Decision&Control, 2005.

[11] G. Mollova, "Compact formulas for least-squares design of digital differentiators", *Electronics Letters*, Vol. 35, No. 20, 1999, pp. 1695–1697.

[12] S.M. Kay, Fundamentals of Statistical Signal Processing: Estimation Theory, Prentice Hall 1993.

[13] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I*, Wiley, 1968, pp.74-85.

[14] I. Fantoni and R. Lozano, Non-linear Control for Underactuated Mechanical Systems, 2002, Springer Verlag.

[15] K. Graichen, M. Treuer and M. Zeitz, Swing-up of the double pendulum on a cart by feedforward and feedback control with experimental validation, *Automatica* 43(1), 2007.