21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

On numerical differentiation algorithms for observation problems

S. Diop[1] and H. Mounier[1]

***Abstract*— By numerical differentiation it is meant here approximation of the derivative at current time t of a real valued function of a single real variable (the time, t) which is available only through finitely many of its past samples which are possibly imprecisely known. Numerical differentiation is a major ill-posed inverse problem. It has not yet received a definite solution. But it is encountered in many engineering problems. Efficient numerical differentiation algorithms thus are key-enablers of engineering designs such as in control system theory, specifically estimation, control synthesis, and other practical issues. This communication presents theoretical and implementation details on several numerical differentiation algorithms for potential use in observation problems. In particular, these algorithms may be used as ingredients for alternative solutions to the longstanding problem of observer design for nonlinear systems.**

Keywords. Observer design; Numerical differentiation algorithms; Ill-posed inverse problems

2010 Mathematics Subject Classification: 65D25; 65F22; 93B50

I. Introduction

By numerical differentiation it is meant here approximation $\widehat{\dot{y}}(t)$ of $\dot{y}(t)$, the derivative at current time t of $y : \tau \mapsto y(\tau)$, a real valued function of a single real variable (the time) which is available only through finitely many of its past samples which are possibly imprecisely known, $\overline{y}(t_W), \overline{y}(t_{W-1}), \cdots, \overline{y}(t_1)$. Very often the sampling instants $t_W < t_{W-1} < \cdots < t_1 = t$ ($W \in \mathbb{N}$, $W \geq 2$) are fixed by experimental conditions, and the samples $\overline{y}(t_i)$ are approximates of exact values $y(t_i)$: $y(t_i) - \overline{y}(t_i) = \widetilde{\overline{y}}(t_i)$. A numerical differentiation algorithm thus is an approximation of the differentiation operator from, say the space of differentiable real valued functions defined on the time interval $[t_W, t]$, into the space of real valued functions defined on the same time interval. It is well-known that the differentiation operator is *discontinuous* with respect to norms which are usually of interest in practice. For instance, for the uniform norm,

$$||z||_\infty = \max_{\tau \in [t_W, t]} |z(\tau)|$$

on the space $\mathcal{C}^1(t_W, t)$ of continuously differentiable real valued functions defined on the time interval $[t_W, t]$, and on the space $\mathcal{C}^0(t_W, t)$ of continuous real valued functions on $[t_W, t]$, there are functions with small norms while with high norms for their derivatives. This is the case for

$$\widetilde{\overline{y}}(\tau) = \varepsilon \sin \frac{2\pi\tau}{\varepsilon^2}\,, \quad \varepsilon > 0.$$

In this case the norm of the error in the derivative approximate of y

$$\dot{y}(\tau) - \dot{\overline{y}}(\tau) = \dot{\widetilde{\overline{y}}}(\tau) = \frac{2\pi}{\varepsilon} \cos \frac{2\pi\tau}{\varepsilon^2}$$

grows indefinitely while the norm of the uncertainty on y gets smaller and smaller. Numerical differentiation algorithm by means of the popular backward difference,

$$\widehat{\dot{y}}(t) = \frac{\overline{y}(t) - \overline{y}(t-T)}{T}\,,$$

therefore is a bad approximation of the differentiation operator in function spaces containing $\mathcal{C}^1(t_W, t)$. Actually the backward difference algorithm consists of the differentiation of the *polynomial interpolant* $\widehat{y}(\tau) = \alpha_0 + \alpha_1\tau$ on the grid consisting of the 2 instants $t - T$ and t, so that $\widehat{\dot{y}}(t) = \dot{\widehat{y}}(t)$, where $t - T$ is a past instant where $\overline{y}$ is available. It follows that

$$\begin{aligned} y(\tau) - \widehat{y}(\tau) &= (\overline{y}(\tau) - \widehat{y}(\tau)) + \widetilde{\overline{y}}(\tau) \\ &= \dot{\widetilde{y}}(\widetilde{\tau})(t - \tau) + \widetilde{\overline{y}}(\tau) \end{aligned}$$

for all $\tau \in [t-T, t]$, where $\widetilde{\tau}$ is known, by the mean value theorem, to lie in $[t - T, t]$. At $\tau = t$ it is seen that the error in the interpolation of y equals the uncertainty on y,

$$y(t) - \widehat{y}(t) = \widetilde{\overline{y}}(\tau)\,,$$

hence the error in the derivative approximate is the derivative of the uncertainty on y,

$$\dot{y}(t) - \widehat{\dot{y}}(t) = \dot{\widetilde{\overline{y}}}(\tau)\,.$$

For use in practice, a *stable* numerical differentiation algorithm thus is a *continuous* operator which approximates the discontinuous differentiation operator. Describing efficient stable numerical differentiation algorithms has been, and still is, a longstanding numerical analysis problem, see for instance [1], [2], [3], [4], [5], [6], [7].

Numerical differentiation is invoked in many engineering problems, see for instance [8], [9], [10] for applications in frequency estimation in signal processing, [11] in pattern recognition, [12] in biomedical engineering, [13] in mechatronics, [14] in flatness path tracking.

[1]Laboratoire des Signaux et Systèmes, CNRS – Université Paris-Sud Orsay – Supélec, Plateau de Moulon, 91192 Gif sur Yvette cedex, France, diop@lss.supelec.fr, Hugues.Mounier@lss.supelec.fr

ISBN: 978-90-367-6321-9

It has been invoked in observation problems in [15], [16], [17], [18], and in [19], [20] a new observer structure has been proposed with stable numerical differentiation algorithms as ingredients.

This communication is an attempt to collect details on some of the various numerical differentiation algorithms which are most relevant to control theoretic issues in online estimation. Some of such algorithms have been reported in previous publications of the authors (this is the case for the popular Savitzky-Golay differentiation scheme) and will not be addressed here. The paper is organized as follows. We first recall the formal explanation of differentiation as a typical ill-posed inverse problem and relate its study to regularization techniques. Among the latter, the so-called linear filtering technique widely used in practice as a numerical differentiation algorithm is presented. Next, another popular numerical differentiation scheme, namely the one based upon polynomial splines is presented. We also included in this section an averaged finite difference differentiation scheme. Finally, numerical differentiation algorithms based upon operational calculus is included.

II. On regularizations of differentiation

The rth derivative, $r \in \mathbb{N}, r \geq 1$, of $y = y(t) \in \mathcal{C}^r(a,b)$ is a solution, $x(t)$, of the integral equation

$$\int_a^t \frac{1}{(r-1)!}(t-\tau)^{r-1} x(\tau)\mathrm{d}\tau = y(t)$$

for x. We may always *detrend* y by replacing it in the latter equation by

$$\breve{y}(t) = y(t) - \sum_{i=0}^{r-1} \frac{1}{i!} y^{(i)}(a)(t-a)^i$$

and then assume that y satisfies the initial conditions

$$y(a) = 0, \dot{y}(a) = 0, \ldots, y^{(r-1)}(a) = 0.$$

The subspace of $\mathcal{C}^r(a,b)$ consisting of functions satisfying the previous initial conditions is denoted by $\mathcal{C}_0^r(a,b)$. Let

$$h_r(t,\tau) = \frac{1}{(r-1)!}(t-\tau)^{r-1}\Upsilon(t-\tau)$$

where Υ is the Heaviside function, i.e.,

$$\Upsilon(\tau) = 1 \text{ if } \tau > 0 \text{ and } \Upsilon(\tau) = 0 \text{ if } \tau \leq 0\,.$$

The operator H_r

$$\begin{array}{ccc} \mathcal{C}^0(a,b) = X & \to & Y = \mathcal{C}_0^r(a,b) \\ x & \mapsto & H_r x \end{array}$$

with

$$(H_r x)(t) = \int_a^b h_r(t,\tau)x(\tau)\mathrm{d}\tau\,.$$

is thus one-to-one, and the rth derivative of $y \in \mathcal{C}_0^r(a,b)$ is the *unique* solution of the Fredholm integral equation of the first kind

$$H_r x = y\,. \tag{1}$$

Integral operators

$$x \mapsto Kx, \quad t \mapsto \int_a^b k(t,\tau)x(\tau)\mathrm{d}\tau$$

where $a, b \in \mathbb{R}$, are *bounded* and *compact* (they map bounded sets into relatively compact sets) whenever the kernel k satisfies one of the following conditions:

(i) k is continuous on $[a,b]\times[a,b]$, in which case

$$||K||_\infty = \max_{a\leq t\leq b}\int_a^b |k(t,\tau)|\mathrm{d}\tau\,.$$

(ii) k is square integrable over $[a,b]\times[a,b]$

$$||k||_2 = \sqrt{\int_a^b\int_a^b |k(t,\tau)|^2\mathrm{d}\tau\mathrm{d}t} \;<\; \infty\,,$$

in which case

$$||K||_2 \leq ||k||_2\,.$$

(iii) k is *weakly singular*, i.e., it is continuous on the subset of points (t,τ) of $[a,b]\times[a,b]$ such that $t \neq \tau$, and there are reals $\gamma > 0$ and ι such that

$$|k(t,\tau)| \leq \frac{\gamma}{|t-\tau|^\iota}\,.$$

Linear, bounded, one-to-one, compact operators with infinite dimensional range are known to have *unbounded inverses* (see, e.g., section 2.2 of [21]):

$$||K^{-1}|| = \infty\,,$$

where the norm is the operator norm.

The operator K^{-1} is thus *discontinuous* implying that the effect of uncertainties in y may be *indefinitely amplified* in the solution of the equation $Kx = y$. Such equations are said to be *ill-posed*. More precisely, an equation (or problem) $Kx = y$ resulting from an operator $K : X \longrightarrow Y$ on normed spaces is said to be *well-posed* if it has one, and only one, solution x for each given y, and if the solution depends continuously on the data in the sense that, given a sequence $(y_n)_{n\in\mathbb{N}}$, $\lim_{n\to\infty} y_n = y$ implies that the corresponding sequence of solutions, $(x_n)_{n\in\mathbb{N}}$, verifies $\lim_{n\to\infty} x_n = x$.

Given that the kernel of equation (1) satisfies conditions (i) and (i) above, *differentiation is an ill-posed problem.*

A *regularization strategy* for an *ill-posed* problem $Kx = y$ is a family $(R_\lambda)_{\lambda>0}$ of *linear* and *bounded* operators $R_\lambda : Y \longrightarrow X$ such that $\lim_{\lambda\to 0} R_\lambda y = K^{-1}y$ for all y.

A regularization strategy for the integral equation (1) is called a *differentiation scheme* . The *regularization error* on uncertain data, $\overline{y} = y + \varepsilon$, takes the form

$$K^{-1}y - R_\lambda\overline{y} = \left(K^{-1} - R_\lambda\right)y - R_\lambda\varepsilon\,. \tag{2}$$

It is the sum of two error terms, the first one being contributed by the inaccuracy of the differentiation scheme on exact data and the second one being the result of the action of the differentiation scheme on the uncertainty

of the data. It is thus apparent that differentiation schemes should care about two main features: *accuracy* (of the differentiation approximation) on exact data, and capability of *smoothing* out uncertainties on the non exact data. These two tasks are *conflicting* for the differentiation operator. In other words, when we try to improve the accuracy of a differentiation scheme on exact data by choosing λ small, at the same time we are most likely amplifying the effect of uncertainties in the data by increasing the factor $||R_\lambda||$. Therefore, a differentiation scheme should include, in the choice of λ, a *compromise* between accuracy on exact data and the ability to smooth out data uncertainties.

How to choose a strategy $\lambda(\overline{y}, \varepsilon)$, for λ? This depends on what is known about y and ε. The use of a known stochastic, dynamic model for y in linear estimation theory problems has been magnificently illustrated in the Kalman-Bucy filter. In this communication, we assume simpler candidate models for y which are useful in cases where the given models for y are too complex, or have low 'information content' for the design of a differentiation scheme.

Let us assume that we know that the uncertainty in y, ε, is bounded and that we know the value of its bound, σ. Then we let $\lambda(\overline{y}, \sigma) = \lambda(\overline{y}, \varepsilon)$. A regularization strategy is said to be *admissible* [21] if

$$\begin{cases} \lambda(\overline{y}, \sigma) & \xrightarrow[\sigma \to 0]{} \quad 0\,, \\ \sup\limits_{\overline{y}, ||\overline{y}-y|| \leq \sigma} ||K^{-1} y - R_{\lambda(\overline{y},\sigma)} \overline{y}|| & \xrightarrow[\sigma \to 0]{} \quad 0\,, \end{cases}$$

for all y.

III. Linear filtering

In control systems literature and practice numerical differentiation is often implemented as linear filters which approximate the Laplace representation s of the derivation operator by rational fractions of the variable s. How to choose parameters of such filters to make them regularized numerical differentiators? This is shown below.

The discussion will be limited to the following class of such linear approximation of the derivation operator:

$$F_{\lambda,n}(s) = \frac{s}{(\lambda s + 1)^n}, \tag{3}$$

where $\lambda > 0$ et $n \in \mathbb{N}^*$ and derivative is estimated as

$$\widehat{\dot{y}}(s) = F_{\lambda,n}(s) \cdot \overline{y}(s)\,.$$

Theorem 1: The estimation error by means of (3) of the derivative of the signal $\overline{y} = y - \widetilde{\overline{y}}$ is bounded as

$$\left\| \dot{y} - \widehat{\dot{y}} \right\|_2 \leq n\,\lambda\,\|\ddot{y}\|_2 + \frac{\sqrt{\dfrac{(n-1)^{n-1}}{n^n}}}{\lambda}\,\|\widetilde{\overline{y}}\|_2 \tag{4}$$

for $n \geq 1$ where $(n-1)^{n-1}$ is 1 when $n = 1$. This bound is minimal and is

$$\left\| \dot{y} - \widehat{\dot{y}} \right\|_2 \leq 2\,n \sqrt[4]{\frac{(n-1)^{n-1}}{n^{n+2}}}\,\sqrt{\|\ddot{y}\|_2}\,\sqrt{\|\widetilde{\overline{y}}\|_2} \tag{5}$$

for the following choice

$$\lambda = \sqrt[4]{\frac{(n-1)^{n-1}}{n^{n+2}}}\,\sqrt{\frac{\|\widetilde{\overline{y}}\|_2}{\|\ddot{y}\|_2}} \tag{6}$$

of the regularization parameter λ.

Corollary 2: The operator (3) is a stable differentiation scheme for

$$\lambda(y, \overline{y}) = \sqrt[4]{\frac{(n-1)^{n-1}}{n^{n+2}}}\,\sqrt{\frac{\|\widetilde{\overline{y}}\|_2}{\|\ddot{y}\|_2}}\,.$$

Remark 3: The error bound of the differentiator given by (5) is a decreasing function of n for $n \geq 2$. It is important to notice that it takes the same value for $n = 1$ and $n = 2$, that is

$$2\,\sqrt{\|\ddot{y}\|_2} \cdot \sqrt{\|\widetilde{\overline{y}}\|_2}\,.$$

It is better to take $n = 2$ instead of $n = 1$ if the resulting dimension increase does not matter.

It is also important to notice that this error bound is minimal for a value of n close to $\frac{5}{4}$. That is to say that the "best" differentiator in terms of linear filters of the form (3) is given by a transform function of rational order, therefore of infinite dimension. The difference between the error bound of this fractional linear filter and the one for $n = 2$ is about 8 %. The qualifier "best" is not to be taken seriously here since the minimization which is invoked previously is not done by minimizing with respect to the couple (λ, n) but first with respect to λ and next with respect to n.

IV. Spline differentiation scheme

The main idea of the numerical spline differentiation scheme may be summarized as follows.

Referring to [22], *the j-th B-spline of order ℓ for $\mathbf{t}$ is defined to be*

$$B_{j,\ell,\mathbf{t}}(t) = (t_{j+\ell} - t_j)\,[t_j, \cdots, t_{j+\ell}]\,(\cdot - t)_+^{\ell-1}\,,$$

where $t \in \mathbb{R}$, and $(x)_+ = \max(x, 0)$ denotes the *truncation* function, and the symbol $[\tau_i, \tau_{i+1}, \cdots, \tau_{i+k}]\,g$ designates the k-th *divided difference of a function g at the points* $\tau_i, \tau_{i+1}, \cdots, \tau_{i+k}$. The j-th B-spline of order ℓ for $\mathbf{t}$ verifies

$$B_{j,\ell,\mathbf{t}}(t) = 0 \quad (t \notin [t_j, t_{j+\ell}]),$$

$$B_{j,\ell,\mathbf{t}}(t) > 0 \quad (t_j < t < t_{j+\ell}),$$

and

$$\sum_{j=k-\ell+1}^{l-1} B_{j,\ell,\mathbf{t}}(t) = 1 \quad (t_k < t < t_l).$$

A spline function of order ℓ for $\mathbf{t}$ is a linear combination

$$s = \sum_{j \in \mathbb{N}} \alpha_i B_{j,\ell,\mathbf{t}}$$

over $\mathbb{R}$ of B-splines of order ℓ for $\mathbf{t}$.

In a *real time* process a *finite* moving window, $\mathbf{t}_W = (t_{k-W}, t_{k-W-1}, \ldots, t_k)$, of measurement sampling instants is considered, where t_k stands for the current instant. In other words, $\mathbf{t}$ is finite, and, for the sake of simplicity of the notations, $\mathbf{t}_W$ will be shortened as $\mathbf{t}$, implicitly assuming the current instant to be W, and $B_{j,\ell,\mathbf{t}}$ will be designated by B_j. At the current instant, t_W, the derivatives must, therefore, be approximated by using the measurements samples contained in the window, $\mathbf{t}$. In the time interval $[t_0, t_W]$ only the B-splines $B_0, \ldots, B_W$ might be nonzero. The splines defined on the latter interval form a linear $\mathbb{R}$-vector space of dimension at most $W+1$. Again, in order to simplify the notations, one component, y_i, of the measurements vector will be considered and denoted by y.

One basic existence theorem in spline approximation states that *among all splines of order $\ell = 2\mathfrak{l}$ for the window $\mathbf{t}$, there is one, and only one, which minimizes the following performance index*

$$J = \lambda \sum_{k=0}^{W} \left(\frac{\overline{y}(t_k) - \hat{y}(t_k)}{\delta_k} \right)^2 + (1-\lambda) \int_{t_0}^{t_W} {\hat{y}^{(\mathfrak{l})}}^2 \mathrm{d}t$$

where $\overline{y}(t_k) = y(t_k) + \varepsilon(t_k)$ are the measurements of the time function $y(t)$ which are corrupted by a white noise $\varepsilon(t)$, and where λ, $0 \leq \lambda \leq 1$, and δ_k are the regularizing *parameters.* This spline will be called *smoothing*. It realizes a compromise between the fidelity to the data and the smoothness of the spline. The order $\ell = 2\mathfrak{l}$ has to be chosen in such a way as it let the spline be sufficiently smooth.

Noting that

$$\sum_{k=0}^{W} \left(\frac{\overline{y}(t_k) - \hat{y}(t_k)}{\delta_k} \right)^2 = (\overline{y} - B\alpha)' D (\overline{y} - B\alpha)$$

where $B = [b_{i,j}]$, $b_{i,j} = B_i(t_j)$, and $\alpha = (\alpha_0, \ldots, \alpha_W)'$, and $\overline{y} = (\overline{y}_0, \ldots, \overline{y}_W)'$, and

$$\int_{t_0}^{t_W} {\hat{y}^{(\mathfrak{l})}}^2 (t) dt = \alpha' R \alpha$$

where $R = [r_{i,j}]$,

$$r_{i,j} = \int_{t_0}^{t_W} B_i^{(\mathfrak{l})}(t) B_j^{(\mathfrak{l})}(t) \mathrm{d}t,$$

the above criterion may be rewritten in the form

$$J = \lambda \, (\overline{y} - B\alpha)' D \, (\overline{y} - B\alpha) + (1-\lambda) \, \alpha' \, R \, \alpha$$

where

$$D = \mathrm{diag} \left(\delta_1^2, \delta_2^2, \ldots, \delta_W^2 \right) .$$

In other words, the problem of finding the coefficients α of the spline is one of least squares. The solution is readily

$$\alpha = \left(\lambda \, B \, D \, B' + (1-\lambda) R \right)^{-1} \lambda \, B \, D \, \overline{y}.$$

A. On error bounds

A complete answer to the fundamental question of error bounds estimate of smoothing splines is not known to the authors. Partial results are, however available. One concerns the *interpolation* case ($\lambda = 1$) in the smoothing spline: *the derivative $\hat{y}^{(\mathfrak{l})}$ of $\hat{y}$ is the least squares approximation to $y^{(\mathfrak{l})}$ from some space of splines of order $\ell = 2\mathfrak{l}$. For cubic splines and four times continuously differentiable outputs,* best *error bounds are known:*

$$\begin{cases} ||y - \hat{y}|| & \leq & \dfrac{5}{384} |\mathbf{t}|^4 ||y^{(4)}||, \\ ||\dot{y} - \dot{\hat{y}}|| & \leq & \dfrac{1}{24} |\mathbf{t}|^3 ||y^{(4)}||, \\ ||\ddot{y} - \ddot{\hat{y}}|| & \leq & \dfrac{3}{8} |\mathbf{t}|^2 ||y^{(4)}||, \end{cases}$$

where $||\cdot||$ denotes the infinity norm over the window, and

$$|\mathbf{t}| = \max_{0 \leq k \leq W-1} (t_{k+1} - t_k)$$

is the so called *mesh size*. When the sampling instants can freely be chosen then better error bounds may be obtained.

B. The averaged differences differentiation operator

The idea (see [23] for more details) is to take quite standard finite difference operators

$$\Delta_{f,\ell,k,q}^{(r)}$$

which approximate $y^{(r)}$ in the mean sense

$$\Delta_{f,\ell,k,q}^{(r)} y(t) = y^{(r)}(t) + c \left((kq+1)T \right)^2 y^{(r+2)}(\zeta)$$

for some constant c and a mean value ζ. Here f is the sampling frequency, $T = 1/f$, and ℓ, k, q are natural numbers. Then the estimate of the rth derivative of y is defined as

$$\widehat{y^{(r)}}(t) = \frac{1}{2q+1} \sum_{j=-q}^{q} \Delta_{f,\ell,k,q}^{(r)} y(t + jT) \ .$$

More explicitly, let $2\ell + 1$ be the number of data points needed in the finite difference operator $\Delta_{f,\ell,k,q}^{(r)}$. The window length is $W = 2 \left((\ell k + 1) \, q + 1 \right)$. The differentiation scheme takes the form

$$\widehat{\underline{y}}^{(r)}(t_{W-\delta}) = C \overline{Y}$$

where the matrix C depends on the differentiation operators $\Delta_{f,\ell,k,q}^{(r)}$ that are used to form the average ; C is computed offline, once and for all. To be a regularization strategy the parameters of the averaged finite difference should satisfy some conditions [23]. As for the Savitzky-Golay differentiation scheme, a formal explicit regularization parameter in terms of the singular values of the matrix C may be defined. It is proved in [23] that this yields a differentiation scheme according to the definition of this notion.

In practice, this scheme is quite flexible given the number of parameters which control its behavior. It is more involved than the Savitzky-Golay scheme, and the time delay

$$\delta = ((\ell k + 1)q + \ell)\, T$$

it introduces is generally much larger.

V. Algebraic differentiators

New algebraic numerical differentiation methods appeared in 2004 (see, e.g., [24], [25], [17]) and have been successfully used in signal and image processing, control theory (see [26], [27], [28], [25], [29], [30], [31]). These differentiators are based on the truncated taylor approximation of a signal. One translates the estimation, always kept short, and operational calculus is used to estimate the desired coefficients. The results are expressed as integrals over the estimation window $[t; t + T_e]$ (t being the current time and T_e the window size) of the signal $f(t)$, multiplied by a certain temporal polynomial, characterizing the differentiator. This integral can, for instance, be computed using the trapezoidal method. The obtained iterative formula may be interpreted as the output of a numerical finite filter, whose inputs are the successive integration points $f(t_k)$.

As an illustration, consider the signal $y(t)$ of which we want to estimate the first derivative. Approximate $y(t)$ around $t = 0$ by a truncated Taylor polynomial until the desired differentiation order, here 1: $p(t) = a_0 + a_1 t$. Thus, the coefficients identifiifation is straightforward: a_0 and a_1 are estimators in $t = 0$ of the signal and of its first derivative. With classical operational calculus notations (see [32]), one gets, for $t \geq 0$:

$$P(s) = \frac{a_0}{s} + \frac{a_1}{s^2} \tag{7}$$

We want to eliminate a_0 since we are looking for an estimate of a_1. To this end, multiply (7) by s:

$$sP(s) = a_0 + \frac{a_1}{s}$$

and differentiate the obtained expression with respect to s to suppress a_0:

$$P(s) + s\frac{\mathrm{d}P(s)}{\mathrm{d}s} = -\frac{a_1}{s^2}.$$

Before coming back to the temporal domain, a multiplication by s^{-N}, with $N > 1$, $N = 2$ for instance, is necessary to yield only integrals:

$$s^{-2}P(s) + s^{-1}\frac{\mathrm{d}P(s)}{\mathrm{d}s} = -s^{-4}a_1.$$

One comes back to the temporal domain by recalling (see [32]) that $\frac{d}{ds}$ corresponds to the multiplication by $-t$:

$$\begin{aligned} a_1 &= \frac{3!\left(\int_{t_0}^{t} \tau y(\tau)\mathrm{d}\tau - \int_{t_0}^{t}\int_{t_0}^{\tau} y(\kappa)\mathrm{d}\kappa\mathrm{d}\tau\right)}{t^3} \\ &= \frac{3!\left(\int_{t_0}^{t} \tau y(\tau)\mathrm{d}\tau - \int_{t_0}^{t} (t-\tau) y(\tau)\mathrm{d}\tau\right)}{t^3} \\ &= \frac{3!\int_{t_0}^{t} \big(\tau y(\tau) - (t-\tau)y(\tau)\big)\mathrm{d}\tau}{t^3} \end{aligned}$$

Therefore the derivative of $y(t)$ at $t = 0$ is approximated by the the above integral:

$$\widehat{\dot{y}(0)} = \frac{3!}{t^3}\int_{t_0}^{t} \big(\tau y(\tau) - (t-\tau)y(\tau)\big)\mathrm{d}\tau$$

For more details see [18] where the latter rough introduction is lenghtly expanded and justified in more conventional mathematical arguments.

References

[1] T. F. Dolgopolova and V. K. Ivanov, "Numerical differentiation," *USSR Comput. Math. Math. Phys.*, vol. 6, pp. 223–232, 1966.

[2] J. Cullum, "Numerical differentiation and regularization," *SIAM J. Numer. Anal.*, vol. 8, pp. 254–265, 1971. [Online]. Available: http://www.jstor.org/stable/2949474

[3] R. S. Anderssen and P. Bloomfield, "Numerical differentiation procedures for non-exact data," *Numer. Math.*, vol. 22, pp. 157–182, 1974.

[4] A. G. Ramm and A. B. Smirnova, "On stable numerical differentiation," *Math. Comp.*, vol. 70, pp. 1131–1153, 2001.

[5] O. Bruno and D. Hoch, "Numerical differentiation of approximated functions with limited order-of-accuracy deterioration," *SIAM J. Numer. Anal.*, vol. 50, pp. 1581–1603, 2012.

[6] V. I. Dmitriev and J. G. Ingtem, "Numerical differentiation using spline functions," *Comput. Math. Modeling*, vol. 23, pp. 312–318, 2012.

[7] W. Y. Kong and V. Rokhlin, "A new class of highly accurate differentiation schemes based on the prolate spheroidal wave functions," *Appl. Comput. Harmonic Analysis*, vol. 33, pp. 226—260, 2012.

[8] T. E. Duncan, P. Mandl, and B. Pasik-Duncan, "Numerical differentiation and parameter estimation in higher-order linear stochastic systems," *IEEE Trans. Automat. Control*, vol. 41, pp. 522–532, 1996.

[9] J. K. Wu, J. Long, Y. Liang, and J. X. Wang, "Numerical differentiation based algorithms for frequency estimation of multiple signals," in *Proceedings of the IEEE SoutheastCon 2004*, 2004, pp. 251–254.

[10] J. K. Wu, J. Long, and J. X. Wang, "High-accuracy, wide-range frequency estimation methods for power system signals under nonsinusoidal conditions," *IEEE Trans. Power Delivery*, vol. 20, pp. 366–374, 2005.

[11] A. M. Taratorin and S. Sideman, "Constrained regularized differentiation," *IEEE Trans. Pattern Anal. Mach. Intelligence*, vol. 16, pp. 88–92, 1994.

[12] M. P. Wachowiak, G. S. Rash, P. M. Quesada, and A. H. Desoky, "Wavelet-based noise removal for biomechanical signals: a comparative study," *IEEE Trans. Biomedical Engrg.*, vol. 47, pp. 360–368, 2000.

[13] R. J. E. Merry, M. J. G. van de Molengraft, and M. Steinbuch, "Velocity and acceleration estimation for optical incremental encoders," *Mechatronics*, vol. 20, pp. 20—26, 2010.

[14] L. Menhour, B. d'Andréa Novel, M. Fliess, and H. Mounier, "Coupled nonlinear vehicle control: Flatness-based setting with algebraic estimation techniques," *Control Engrg. Practice*, vol. 22, pp. 135—146, 2014.

[15] S. Diop, J. W. Grizzle, P. E. Moraal, and A. G. Stefanopoulou, "Interpolation and numerical differentiation for observer design," in *Proceedings of the American Control Conference*. Evanston, IL: American Automatic Control Council, 1994, pp. 1329–1333.

[16] E. Cruz-Zavala, J. A. Moreno, and L. M. Fridman, "Uniform robust exact differentiator," in *Proceedings of the IEEE Conference on Decision and Control*. New York: IEEE Press, 2010, pp. 1189–1194.

[17] M. Mboup, C. Join, and M. Fliess, "Numerical differentiation with annihilators in noisy environment," *Numerical Algorithms*, vol. 50, pp. 439–467, 2009.

[18] D. Y. Liu, "Analyse d'erreurs d'estimateurs des dérivées de signaux bruités et applications," PhD. Thesis, Université des Sciences et Technologies de Lille, Villeneuve d'Ascq, France, 2011.

[19] S. Diop and S. Ibrir, "A model based numerical differentiation observer for linear systems," in *Proceedings of the IFAC World Congress*. Amsterdam: Elsevier, 1999, pp. Paper D–2b–08–5.

[20] S. Diop, V. Fromion, and J. W. Grizzle, "A global exponential observer based on numerical differentiation," in *Proceedings of the IEEE Conference on Decision and Control*, vol. 4. New York: IEEE Press, 2001, pp. 3344–3349.

[21] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of Inverse Problems*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 1996.

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

[23] R. S. Anderssen, F. de Hoog, and M. Hegland, "A stable finite difference ansatz for higher order differentiation of non-exact data," *Bull. Austral. Math. Soc.*, vol. 58, pp. 223–232, 1998.

[24] M. Fliess and H. Sira-Ramírez, "Compression différentielle de transitoires bruités," *C. R. Acad. Sci. Paris Sér. I*, vol. 339, pp. 821–826, 2004.

[25] M. Fliess, C. Join, and H. Sira-Ramírez, "Non-linear estimation is easy," *Internat. J. Modelling Identification Control*, vol. 4, pp. 12–27, 2008.

[26] G. Fedele, F. Chiaravalloti, and C. Join, "An algebraic derivative-based approach for the zero-crossings estimation," in *17th European Signal Processing Conference 2009 (EUSIPCO 2009)*. European Association for Signal Processing, 2009, pp. 2455–2459.

[27] M. Fliess and C. Join, "Model-free control," *Internat. J. Control*, vol. 86, pp. 2228–2252, 2013.

[28] ——, "Towards new technical indicators for trading systems and risk management," in *Proceedings of the IFAC Symposium on* System Identification. Oxford, UK: International Federation of Automatic Control, 2009, pp. 1435–1440.

[29] C. Join and S. Tabbone, "Robust curvature extrema detection based on new numerical derivation," in *Advanced Concepts for Intelligent Vision Systems. Proceedings of the 10th International Conference, ACIVS 2008, Juan-les-Pins, France, October 20-24, 2008*, ser. Lect. Notes Comput. Sci., J. Blanc-Talon, S. Bourennane, W. Philips, D. Popescu, and P. Scheunders, Eds., vol. 5259. Springer-Verlag, 2008, pp. 485–493.

[30] J. Villagra, B. d'Andrea Novel, M. Fliess, and H. Mounier, "Estimation of longitudinal and lateral vehicle velocities: An algebraic approach," in *Proceedings of the American Control Conference*. Evanston, IL: American Automatic Control Council, 2008, pp. 3941–3946.

[31] ——, "Synthèse algébrique d'estimateurs de vitesses longitudinale et latérale d'une automobile," *e-STA*, vol. 5, no. 3, pp. 12–27, 2008.

[32] K. Yosida, *Operational Calculus. A Theory of Hyperfunctions*. New York: Springer-Verlag, 1984.