On numerical differentiation algorithms for nonlinear estimation

S. Diop, J. W. Grizzle, F. Chaplais

Abstract. Practical implementations of the operation which consists of differentiating a signal, a real valued function of the time, known only through its on line samples is much wanted, given the numerous areas in control theory where it is encountered. We present theoretical as well as implementation details on some numerical differentiation algorithms which may be useful in the area of nonlinear estimation, where, in control theory in particular, these algorithms may be used as ingredients for alternative solutions to the longstanding problem of asymptotic observer design. Numerical differentiation is known as an ill-posed inverse problem. We call on regularization concepts and techniques to investigate potentially practical algorithms for numerical differentiation. Specific details that we provide here include three differentiation schemes. All of them being almost equally attractive for the balance they make between online computation burden, speed and convergence behavior. The first one is a polynomial differentiation scheme. The second one is an averaged finite differences scheme. And the third option uses wavelets.

Keywords. Numerical differentiation; Regularization; Mollification; Nonlinear estimation; Observers

1 Introduction

While this work mainly targets applications in control theory nonlinear estimation issues we shall content ourselves to referring to [11, 12, 10, 7] for motivations found in control theory to invoke numerical differentiation, and focus our attention on the theory and implementation of some numerical differentiation algorithms.

The difficulty with the implementation of differentiation is that it is an operator which is *not continuous* with respect to norms we usually tend to consider. The following standard example makes this clear. Assume $y = y(t) \in \mathcal{C}^1(a,b), a < b \in \mathbb{R}$, so that its derivative x is in $\mathcal{C}^0(a,b)$, and endow $\mathcal{C}^0(a,b)$ and $\mathcal{C}^1(a,b)$ with the uniform norm,

$$||z||_\infty = \max_{\tau\in[a,b]} |z(\tau)| \,.$$

If y is known through some experimental data

$$\overline{y} = y + \varepsilon \,,$$

for instance,

$$\overline{y}(t) = y(t) + \sigma \sin \frac{2\pi t}{\sigma^2} \,,$$

then

$$\dot{\overline{y}}(t) = \dot{y}(t) + \frac{2\pi}{\sigma} \cos \frac{2\pi t}{\sigma^2} \,.$$

Therefore,

$$||\overline{y} - y||_\infty = \sigma,$$

while

$$||\dot{y} - \dot{\overline{y}}||_\infty = \frac{2\pi}{\sigma} \,.$$

In other words, the smaller the uncertainty on the data the larger the error in the derivative of y. This abstract example hardly exaggerates what happens in real applications. One of the consequences of this non continuity of the differentiation operator is that we cannot use simple schemes such as the backward difference

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

to implement differentiation operators. The backward difference scheme is easily seen as equivalent to the differentiation scheme obtained through the *polynomial interpolant*

$$\widehat{y}(t) = \alpha_0 + \alpha_1 t \,,$$

on the grid consisting of the 2 instants $t - T, t$, so that

$$\widehat{\dot{y}}(t) = \dot{\widehat{y}}(t) \,,$$

where $t - T$ is a past instant where $\overline{y}$ is available. Assuming $\overline{y}$ to be twice differentiable we then have

$$\begin{aligned} y(\tau) - \widehat{y}(\tau) &= (\overline{y}(\tau) - \widehat{y}(\tau)) - \varepsilon(\tau) \\ &= \frac{1}{2}\ddot{\overline{y}}(\widetilde{\tau})(\tau - t + T)(\tau - t) - \varepsilon(\tau) \end{aligned}$$

S. Diop <diop@lss.supelec.fr>: Laboratoire des Signaux & Systèmes; Supélec; 91192 Gif sur Yvette cedex, France.
J. W. Grizzle <grizzle@umich.edu>: Control Systems Laboratory; EECS Dept.; Univ of Michigan; Ann Arbor, MI 48109-2122, USA.
F. Chaplais <chaplais@cas.ensmp.fr>: Centre Automatique et Systèmes; ENSM Paris; 77305 Fontainebleau cedex, France.

for all $\tau \in [t-T, t]$, where $\widetilde{\tau}$ is known, by Rolle's Theorem, to lie in $[t-T, t]$. Applying this remainder equation at $\tau = t$ yields

$$y(t) - \widehat{y}(t) = -\varepsilon(t)\,,$$

hence

$$\dot{y}(t) - \widehat{\dot{y}}(t) = -\dot{\varepsilon}(t)\,,$$

that is, the derivative estimation error through the backward difference scheme is the derivative of the uncertain signal on the data. The backward difference scheme thus has no immunity against the higher frequency content of this uncertain signal which tend to drown the derivative of the true signal.

Differentiation is well-known as one of the most important so-called *ill-posed inverse problems*. Describing efficient algorithms which implement differentiation has been and still is a longstanding numerical analysis problem. But, as for many other ubiquitous ill-posed inverse problems arising in physics and engineering science, the formulation of regularization concepts by A. N. Tikhonov, about four decades ago, has led to a substantial breakthrough.

This work is an attempt to collect details on some of the various numerical differentiation algorithms which are most relevant to the control theory issues in online estimation.

The communication is organized as follows. We first present differentiation as the inverse of the integration operator on normed spaces of real-valued functions. This throws some light on the ill-posedness of differentiation, and, at the same time, relates differentiation to ill-posed inverse problems and their regularization strategies. Among regularization techniques mollification is close to the control theory notion of filtering. We then present an overview of this notion. We discuss its use as an ingredient in building blocks of numerical differentiation nonlinear estimation. Then we discuss some implementation issues, mainly the discrete nature of the data which will be available for digital processing. Finally three differentiation schemes are presented in more details. One is the so-called Savitzky-Golay differentiation scheme. The second one is an averaged finite difference differentiation scheme. And the last one is in the class of wavelet differentiation scheme.

2 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

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

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.,

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

The operator H_r

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

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 [14]):

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

where the norm is the operator norm.

The operator K^{-1} is thus *not continuous* 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 (ii) and (iii) above, differentiating 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.

We call *differentiation scheme* a regularization strategy for the integral equation (1). The regularization error on uncertain data,

$$\overline{y} = y + \varepsilon\,,$$

then takes the form

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

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 on 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* ones for the differentiation operator. In other words, when we try to improve the accuracy of differentiation schemes on exact data by choosing λ small, at the same time we are most likely amplifying data uncertainties by an increased factor of $||R_\lambda||$. Therefore, a differentiation scheme should include, in the choice of λ, a *compromise* between accuracy on exact data and faculty to smooth out data uncertainties.

How to choose a strategy $\lambda(\overline{y},\varepsilon)$, for λ? This actually is a matter of *quantity of information available* on y and ε.

A rather mild requirement is that we know ε to be a bounded function of t and that we know the value of its bound, σ. A more involved information that we also may have at our disposal is a *model* for y. This is the case in control theory where a model is usually taken as a dynamic system with a known input and with output y where the need for derivatives of y comes from model state estimation needs. All such an information is used in the design of the strategy R_λ and $\lambda(\overline{y},\varepsilon)$ for the choice of the regularization parameter.

The use of a supposedly known dynamic model of y in control theory estimation problems has been magnificently done in the Kalman-Bucy filter for linear models. This discussion is pursued in a separate work [9]. In this work we assume simpler candidates as models for y which reveal useful in case the given models for y are too complex, or with low information content for the differentiation scheme design.

Let us assume that we know σ. Then we let $\lambda(\overline{y},\sigma) = \lambda(\overline{y},\varepsilon)$. A regularization strategy is said to be *admissible* [18, 14] if

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

for all y.

Here are the main lines of the so-called spectral design of regularization strategies. We recall that the *singular system* for a linear bounded compact operator K over the Hilbert spaces X and Y is

$$(\tau_i, x_i, y_i)_{i\in\mathbb{N}}$$

with positive singular values $\tau_i \ge \tau_{i+1}$, (the numbers $\left(\tau_i^2\right)_{i\in\mathbb{N}}$ being the eigenvalues of K^*K where K^* is the

adjoint of K), orthonormal systems $(x_i)_{i\in\mathbb{N}}$, $(y_i)_{i\in\mathbb{N}}$ of X and Y, respectively, such that

$$\left\{\begin{array}{rcl} Kx_i &=& \tau_i y_i\,, \\ K^*y_i &=& \tau_i x_i \end{array}\right.$$

for all $i\in\mathbb{N}$. If K is one-to-one then its singular values are all positive, and for the equation

$$Kx = y\,,$$

we have the following so-called Picard criterion

$$\sum_{i=0}^{\infty}\frac{|\,(y,y_i)\,|^2}{\tau_i^2}<\infty \tag{3}$$

and the solution is given by

$$x=\sum_{i=0}^{\infty}\frac{(y,y_i)}{\tau_i}x_i\,, \tag{4}$$

where (y,y_i) denotes the inner product on Y of y and y_i. The Picard criterion reflects the necessary (and sufficient) convergence condition for a series $\sum_{i=0}^{\infty}\lambda_i e_i$ in a Hilbert space where (e_i) is an orthonormal sequence of vectors. It is closely related to the invertibility of the operator K. The unboundedness of K^{-1} is related to the fact that the singular values of K accumulate at 0. It is apparent that we cannot use equation (4) as is to implement differentiation schemes since, with experimental data, the high frequency content of ε would tend to drown the derivative of the true signal y. The notion of high frequency content of ε refers here to the fact that (ε,y_i) may be seen as the Fourier coefficients of ε in the orthonormal system $(y_i)_{i\in\mathbb{N}}$.

Spectral regularization strategies then assume the following general form:

$$R_\lambda\overline{y}=\sum_{i=0}^{\infty}\omega(\lambda,\tau_i)\,(\overline{y},y_i)\,x_i$$

where the coefficient $1/\tau_i$ in equation (4) has been replaced by the function $\omega(\lambda,\tau_i)$ for the sake of better behavior for low values of τ_i. For instance the well-known Tikhonov regularization corresponds to the choice

$$\omega(\lambda,\mu)=\frac{\mu}{\mu^2+\lambda}\;,$$

while the TSVD (truncated singular value decomposition) strategy corresponds to

$$\omega(\lambda,\mu)=\left\{\begin{array}{ll}\dfrac{1}{\mu}\,, & \mu\geq\lambda\;,\\ 0, & \mu<\lambda\;.\end{array}\right.$$

Details on the TSVD for the differentiation operator H_r of first order ($r=1$) are as follows. A singular system for H_1 is

$$\left\{\begin{array}{rcl} \tau_i &=& \dfrac{2\,(b-a)}{(2i+1)\,\pi}\,, \\ x_i(t) &=& \sqrt{\dfrac{2}{b-a}}\cos\left((2i+1)\,\dfrac{\pi}{2}\,\dfrac{t-a}{b-a}\right), \\ y_i(t) &=& \sqrt{\dfrac{2}{b-a}}\sin\left((2i+1)\,\dfrac{\pi}{2}\,\dfrac{t-a}{b-a}\right). \end{array}\right. \tag{5}$$

If $N=N(\lambda)$ denotes the largest integer which is lower than or equal to

$$\frac{1}{2}\left(\frac{2\,(b-a)}{\lambda\pi}-1\right)$$

then the TSVD strategy yields

$$\widehat{\dot{y}}(t)=\sum_{i=0}^{N}\frac{(2i+1)\,\pi}{2\,(b-a)}\,(\overline{y},y_i)\,x_i\,.$$

The norm of the estimation error

$$\begin{array}{rcl}\widetilde{\dot{y}}(t)=\dot{y}(t)-\widehat{\dot{y}}(t) &=& \displaystyle\sum_{i=N+1}^{\infty}\frac{(2i+1)\,\pi}{2\,(b-a)}\,(y,y_i)\,x_i- \\ && \displaystyle\sum_{i=0}^{N}\frac{(2i+1)\,\pi}{2\,(b-a)}\,(\varepsilon,y_i)\,x_i\,,\end{array}$$

is

$$\begin{array}{rcl}||\widetilde{\dot{y}}(t)||_{L^2}^2 &=& \displaystyle\sum_{i=N+1}^{\infty}\frac{(2i+1)^2\,\pi^2}{4\,(b-a)^2}\,(y,y_i)^2+ \\ && \displaystyle\sum_{i=0}^{N}\frac{(2i+1)^2\,\pi^2}{4\,(b-a)^2}\,(\varepsilon,y_i)^2\;.\end{array}$$

For a fixed $\lambda>0$ the estimation error may be large but is definitely *bounded*. When λ is made small the first term converges to 0 while the second term, depending on the nature of the uncertainty ε, will most likely go to ∞. Conversely, by making λ large, we reduce the contribution of the uncertainties at the expense of oversmoothing the derivative. One possible trade-off consists of choosing S and requiring

$$||\widetilde{\dot{y}}(t)||_{L^2}^2\leq S\,.$$

Then from the above expression of $||\widetilde{\dot{y}}(t)||_{L^2}^2$ we see that

$$\sum_{i=0}^{N(\lambda)}\frac{(2i+1)^2\,\pi^2}{4\,(b-a)^2}\,(\varepsilon,y_i)^2\leq S$$

which provides an inequality to solve in λ in order to obtain a potentially admissible first order differentiation scheme. The extra assumptions needed in practical implementations of such a regularization strategy are loosely indicated in the latter inequality. We may, for instance, assume σ to be, instead of the bound on ε, the bound on its spectral power density.

3 Mollification

A regularization strategy for the differentiation operator (1) may also be obtained by mollification as follows. A *mollifier* [16] is a nonnegative function φ of a single variable with integral 1 over the reals. An example is the so-called Gaussian kernel

$$\varphi(t) = \frac{1}{\sqrt{\pi}} e^{-t^2} \quad (t \in \mathbb{R}) .$$

Given a mollifier φ, we constructs the family $\varphi_\lambda(t) = \frac{1}{\lambda}\varphi(\frac{t}{\lambda}) \quad (t \in \mathbb{R})$, where $\lambda \in \mathbb{R}$. Mollified numerical differentiation is then defined as

$$\widehat{\dot{y}}(t) = R_\lambda y(t) = (\dot{\varphi}_\lambda * y)(t) ,$$

where,

$$\dot{\varphi}_\lambda(\tau) = \frac{\mathrm{d}}{\mathrm{d}\tau}\varphi_\lambda(\tau) ,$$

and $*$ denotes convolution of functions, and where y assumes a compact support.

The main reason for using mollification is the following set of results, which may be found in [1, 18], for instance. The filtered data $\varphi_\lambda * y$ is *infinitely differentiable*. Moreover, for the Gaussian mollifier

$$\begin{aligned} ||R_\lambda y||_{L^2} &\leq \frac{4}{\lambda\sqrt{\pi}}||y||_{L^2} \\ ||R_\lambda K x||_{L^2} &\leq 2||x||_{L^2}; \end{aligned}$$

i.e., the family of $(R_\lambda)_{\lambda>0}$ is *uniformly bounded*. Now the regularization error with uncertain data is bounded by

$$\begin{aligned} ||\dot{y} - \widehat{\dot{y}}||_{L^2} &= ||x - R_\lambda \overline{y}||_{L^2} \\ &\leq 2\sqrt{2}\lambda||\dot{x}||_{L^2} + \frac{4\sigma}{\lambda\sqrt{\pi}} \\ &= 2\sqrt{2}\lambda||\ddot{y}||_{L^2} + \frac{4\sigma}{\lambda\sqrt{\pi}} . \end{aligned}$$

Under the assumption $||\ddot{y}||_{L^2} \leq E$, this bound is minimized (optimal) for

$$\lambda = \lambda(\sigma) = \sqrt{\frac{2}{\sqrt{2\pi}}}\sqrt{\frac{\sigma}{E}}$$

which yields

$$||\dot{y} - \widehat{\dot{y}}||_{L^2} \leq \frac{8}{\sqrt[4]{2\pi}}\sqrt{\sigma E}$$

that is an admissible differentiation scheme for the first order differentiation operator.

4 Implementation issues

When we turn to the implementation of differentiation schemes as depicted in the previous sections then we face two main issues. One is that data are available as discrete samples of the underlying continuous time variables. Moreover, calculations are most likely done on digital computers which require discrete samples of the data. The second issue we face in practice is that we shall have only finite room to store the data. The finiteness of computer storage memory may be overcome by using iterative versions of algorithms at the expense of slower convergence rate and potential stability issues. In view of *online* differentiation schemes there is a major third issue: the amount of computation should be moderate.

In what follows, we assume that the data are sampled at some given frequency, f with $T = 1/f$ the sampling time interval. We discuss noniterative implementations only. And we denote by W the width of the window of data we use at the current time, t. Let the sampling instants in the window be

$$a = t_1, t_2, \ldots, t_W = b ,$$

and the sampled data be denoted by

$$\overline{Y} = \begin{pmatrix} \overline{y}_1 = \overline{y}(t_1) \\ \vdots \\ \overline{y}_W = \overline{y}(t_W) \end{pmatrix} .$$

Implementation of mollification will be by means of fast Fourier transforms (FFT). Due to a much larger data window which would be needed we do not recommend implementing the mollification differentiation scheme in the previous section as is. Instead, mollification is used as a filtering device whose output is passed to to a much simpler differentiation schemes such as the Savitzky-Golay or the averaged finite differences differentiation schemes as will be detailed in the next sections. Precisely, we choose a mollifier such as the Gaussian. The mollification parameter, λ, controls the bandwidth of the signals. It should be chosen by optimizing the trade-off between smoothing and accuracy as was discussed in the previous section. However, in many cases the value of λ may be set up in an *a priori* basis given some knowledge on the bandwidth of the signal being differentiated. One advantage of using mollification is a clear separation between the tasks of smoothing and differentiating.

Here we shall content ourselves with two main ideas for differentiation scheme implementations. The first one is reminiscent of C. F. Gauss' least squares technique while the second one is of statistical flavor.

As may be suggested by our presentation of spectral regularization methods, we may consider truncating formula (4) by considering finite dimensional subspaces X_N of $\mathcal{C}_0^r(a,b)$ generated by subsequences $(x_i)_{1\leq i\leq N}$ of a singular system of $\mathcal{C}_0^r(a,b)$. In practice we extend this viewpoint by choosing an arbitrary finite dimensional subspace $\mathbf{F}$ of $\mathcal{C}_0^r(a,b)$ with base

$$(\mathfrak{b}_j(t))_{1\leq j\leq N} .$$

Next we search for a *linear* estimate

$$\widehat{y}(t) = \sum_{j=1}^{N} \alpha_j \mathfrak{b}_j(t)$$

in the sense of a *semi-norm* on $\mathbf{F}$ and then we take

$$\widehat{y^{(r)}}(t) = \widehat{y}^{(r)}(t)$$

as estimates of the derivatives of y.

The practical choices for $\mathbf{F}$ being subspaces of polynomials, polynomial splines, trigonometric functions (as in the TSVD), etc.

The semi-norm on $\mathbf{F}$ is chosen to be

$$\|\overline{y} - \widehat{y}\|^2 = \sum_{i=1}^{W} (\overline{y}_i - \widehat{y}(t_i))^2 = \left\|\overline{Y} - \beta\alpha\right\|^2$$

where the last symbol $||\cdot||$ is the Euclidean norm,

$$\alpha = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{pmatrix},$$

and β is the $(W \times N)$-matrix with coefficients

$$\beta(i,j) = \mathfrak{b}_j(t_i).$$

The choice of W and N are actually part of the regularization strategy. Smoothing behaviors are obtained if we choose W larger than N, while quite accurate reproduction of $\overline{y}$ is possible if W is chosen to be equal to N. In any case we should have $W \geq N$. As we shall see in the next section, a more precise regularization parameter may be defined which makes this method cope with the general regularization theory depicted earlier in this communication. For the least squares estimate to be uniquely defined we need a base for $\mathbf{F}$ such that the resulting matrix β is of rank N. The following section provides details of a polynomial implementation of this general differentiation scheme.

5 The Savitzky-Golay differentiation scheme

This is perhaps the simplest differentiation scheme. It is polynomial in the sense that the base functions are polynomials in t. It does not have recourse to orthogonal polynomial base functions, but rather uses the base of monomials. We have followed [15] in naming this differentiation scheme after A. Savitzky and J. E. Golay [25]. This differentiation scheme is singled out here for its speed (it reduces to a multiplication of the data vector by a constant matrix which is computed offline once for all) and also for its quality which may be sufficient in many cases.

Let δ be a nonnegative integer at most equal to W. We have

$$t_i = t_{W-\delta} + (i - W + \delta)T$$

The matrix β then is

$$\mathfrak{b}_j(t_i) = (i - W + \delta)^{j-1}.$$

Though of rank N, the matrix β may be *poorly conditioned* for large values of N and some choices of δ. We use the TSVD to regularize this least squares problem. Let

$$\beta = USV'$$

be the singular value decomposition of β, where the prime is for the transposition of matrices. Let $\lambda > 0$ be a positive constant. Then a better estimate of α is

$$\alpha = V\beta^{\dagger}U'$$

where $\beta^{\dagger} = \mathrm{diag}(1/\sigma_1, \ldots, 1/\sigma_s, 0, \ldots, 0)$ and $\sigma_1 \geq \sigma_2, \ldots, \geq \sigma_s \geq \lambda$ are the singular values of β which are at least equal to λ.

From

$$\widehat{y}(t) = \sum_{j=1}^{N} \alpha_j \mathfrak{b}_j(t)$$

we obtain the kth derivative of $\widehat{y}(t)$:

$$\widehat{y}^{(k)}(t) = \sum_{j=r+1}^{N} \alpha_j (j-1)(j-2)\cdots(j-k)\mathfrak{b}_{j-k}(t),$$

so that

$$\underline{\widehat{y}}^{(r)}(t_{W-\delta}) = J\,\alpha$$

where

$$\underline{\widehat{y}}^{(r)} = \begin{pmatrix} \widehat{y} \\ \widehat{\dot{y}} \\ \vdots \\ \widehat{y^{(r)}} \end{pmatrix},$$

and J is the $(r+1) \times N$-matrix with main diagonal the vector

$$\begin{pmatrix} 1 \\ f \\ \vdots \\ r!f^r \end{pmatrix}.$$

In summary, the Savitzky-Golay differentiation scheme reads as the the filter

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

where $C = JV\beta^{\dagger}U'$. The explicit regularization parameter may be chosen in very heuristic way by requiring the norm of C to be no higher than some prespecified value. This provides guidelines for choosing values for λ, δ, N and W as well. A formal proof of the convergence of this differentiation scheme in the lines of the

general regularization theory earlier presented would be rather involved. We merely skip it.

Finally let us note that δ is a quite accurate measure of the delay introduced by the differentiation scheme. Since we know its value actions may be taken in an overall observer design to compensate for this delay, see [9].

6 The averaged finite differences differentiation scheme

We refer the reader to the original works by R. S. Anderssen, F. de Hoog, and M. Hegland in 1984 [5] and 1998 [4] for more details.

The idea is to take quite standard finite difference operators

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

which approximates $y^{(r)}$ 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 integers. Then we define the estimate of the rth derivative of y 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 $\Delta_{f,\ell,k,q}^{(r)}$. We consider a window of

$$W = 2\left((\ell k+1)\, q+1\right)$$

data points. For instance

$$\Delta_{f,\ell,k,q}^{(1)} y(t) = \frac{y(t+(kq+1)T) - y(t-(kq+1)T)}{2(kq+1)T} \ ,$$

$$\Delta_{f,\ell,k,q}^{(2)} y(t) = \frac{y(t+(kq+1)T) - 2y(t) + y(t-(kq+1)T)}{\left((kq+1)T\right)^2} \ .$$

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 $\Delta_{f,\ell,k,q}^{(r)}$ which are used to form the average and is computed offline and once for all. To be a regularization strategy the parameters of the averaged finite difference should satisfy conditions such as [4]

$$\begin{cases} \frac{kq}{f} \ll 1 \\ \\ \frac{q}{f} \ll 1 \\ \\ \sqrt{q}(kq/f)^2 \approx 1 \end{cases}$$

As for the Savitzky-Golay differentiation scheme, we may define a formal explicit regularization parameter in terms of the singular values of the matrix C. It is proved in [4] that we actually have 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

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

it introduces is generally much larger.

7 The wavelet differentiation scheme

7.1 Decomposition in wavelet bases

A scaling function is a function with unit average which satisfies a scaling equation

$$\phi\left(\frac{t}{2}\right) = \sqrt{2} \sum_{k\in\mathbb{Z}} h_k \phi(t-k) \qquad (6)$$

A scale recursion on the Fourier transform of (6), together with the condition $\hat{\phi}(0) = 1$, entirely determines ϕ. If ϕ is well defined, then it has a compact support if and only if a finite number of coefficients h_k is non zero. If the Fourier series of h satisfies a positivity condition [21, 24], then a scaling function defines a multiresolution analysis. Multiresolution approximations are not presented here; for details, see [8, 23, 24].
Two scaling functions ϕ and ϕ^* form a pair of conjugate scaling functions if they satisfy $\langle \phi(t-i), \phi(t-j)\rangle = \delta_{i,j}$. The scaling function is orthogonal if it is its own conjugate.
Dyadic wavelets are also defined by a scaling equation

$$\psi\left(\frac{t}{2}\right) = \sqrt{2} \sum_{k\in\mathbb{Z}} g_k \phi(t-k) \qquad (7)$$

which relates them to the scaling functions. If ϕ and ϕ^* are conjugate scaling functions, and $g_k = (-1)^{k-1} h^*_{1-k}$ and $g^*_k = (-1)^{k-1} h_{1-k}$, then the families

$$\psi_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t-2^j n}{2^j}\right) \ , \ \psi^*_{j,n}(t) = \frac{1}{\sqrt{2^j}} \psi^*\left(\frac{t-2^j n}{2^j}\right) \ , \ j,n \in$$

define dual Riesz bases of $L^2(\mathbb{R})$ [21, 24, 6, 8]. If h and h^* are compactly supported, then any signal x with a local finite energy can be decomposed as

$$\begin{aligned} x(t) &= \frac{1}{\delta} \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} \phi\left(\frac{t}{\delta} - k\right) \phi^*\left(\frac{s}{\delta} - k\right) x(s) ds \qquad (8) \\ &+ \sum_{j=0}^{j=-\infty} \frac{1}{2^j\delta} \sum_{k\in\mathbb{Z}} \int_{\mathbb{R}} \psi\left(\frac{t}{2^j\delta} - k\right) \psi^*\left(\frac{s}{2^j\delta} - k\right) x(s) ds \qquad (9) \end{aligned}$$

where δ defines the finest time scale.
Denote by P_δ the operator defined by (8), e.g. $P_\delta x(t) = 1/\delta \sum_k \int \phi(t/\delta - k)\phi^*(s/\delta - k)x(s)ds$. Section 7.4.1 shows that P_δ is a regularizing operator which accurately approximates regular signals. While ϕ^* is qualified as a mollifier since it has a unit average, the regularized signal is obtained by a discrete reconstruction using the fine elements $\phi(t/\delta - k)$, instead of using a continuous time convolution.

7.2 Differentiation and wavelet bases

Lemarié [19] has related the wavelet decomposition of a signal to the wavelet decomposition of its derivative in a related wavelet basis:

Theorem 1 (Lemarié) *Let ϕ and ϕ^* two conjugate scaling functions such that ϕ is $C^{1+\epsilon}$ for some $\epsilon > 0$. Then there exist two conjugate scaling functions $\tilde{\phi}$ and $\tilde{\phi}^*$ such that:*

$$\phi'(t) = \tilde{\phi}(t) - \tilde{\phi}(t-1) \text{ and } \phi^*(t+1) - \phi^*(t) = \tilde{\phi}^{*\prime}(t) \tag{10}$$

Moreover, the trigonometric polynomials $\tilde{h}$ and $\tilde{h}^*$ and the biorthogonal wavelets $\tilde{\psi}$ and $\tilde{\psi}^*$ related to $\tilde{\phi}$ and $\tilde{\phi}^*$, the projectors $\tilde{P}_\delta$ defined by the pair $(\tilde{\phi}, \tilde{\phi}^*)$ and the projector $\tilde{Q}_\delta = Id - \tilde{P}_\delta$ satisfy

$$\widehat{\tilde{h}}(\omega) = \frac{2}{1+e^{-i\omega}}\hat{h}(\omega) \tag{11}$$

$$\widehat{\tilde{h}^*}(\omega) = \frac{1+e^{i\omega}}{2}\widehat{h^*}(\omega) \tag{12}$$

$$\tilde{\psi}(t) = \frac{1}{4}\psi'(t) \text{ and } \tilde{\psi}^{*\prime}(t) = -4\psi^*(t) \tag{13}$$

$$\frac{d}{dt} \circ P_\delta = \tilde{P}_\delta \circ \frac{d}{dt} \text{ (commutation formula)} \tag{14}$$

$$\frac{d}{dt} \circ Q_\delta = \tilde{Q}_\delta \circ \frac{d}{dt}. \tag{15}$$

7.3 Implementations with filter banks

7.3.1 Discrete wavelet computations

The scaling equations (6,7) make it possible to compute the decomposition and reconstruction of a signal in a wavelet basis using multirate filter banks.
Define

$$c_j[n] = \frac{1}{2^j\delta}\int_{\mathbb{R}} x(s)\phi^*\left(\frac{s}{2^j\delta} - n\right)ds \;,\; d_j[n] = \frac{1}{2^j\delta}\int_{\mathbb{R}} x(s)\psi^*\left(\frac{s}{2^j\delta} - n\right)ds$$

Then

Theorem 2 (Mallat [20, 22]) *At the decomposition*

$$c_{j+1}[p] = \sum_{n=-\infty}^{+\infty} h^*[n-2p] \text{ and } d_{j+1}[p] = \sum_{n=-\infty}^{+\infty} g^*[n-2p]. \tag{16}$$

At the reconstruction,

$$c_j[p] = \sum_{n=-\infty}^{+\infty} h[p-2n]\, c_{j+1}[n] + \sum_{n=-\infty}^{+\infty} g[p-2n]\, d_{j+1}[n]. \tag{17}$$

If δ is "small", then the scaling coefficients $c_0[n]$ can be assimilated to a sampled data because at fine scales P_δ can be assimilated to a Dirac comb. Hence theorem 2 provides a fast algorithm for the "wavelet decomposition" of discrete signals. Moreover, the algorithm is directly related to purely discrete signal processing objects known as perfect reconstruction filter banks [2, 3, 28, 29, 30].
The decomposition and reconstruction algorithm is summarized in Figure 1.
To compute the derivative of the decomposed signal, a suitable transformations must be performed on the coefficients and the right reconstruction filters must be used.

7.3.2 Computing the coefficients of the derivative

Proposition 3 *A differentiable signal x decomposed as in (8,9) satisfies*

$$\frac{dx}{dt} = \frac{1}{\delta}\sum_{k\in\mathbb{Z}} \frac{c_k - c_{k-1}}{\delta}\tilde{\phi}\left(\frac{t}{\delta} - k\right) + \sum_{j=0}^{j=-\infty} \frac{1}{2^j\delta}\sum_{k\in\mathbb{Z}} \frac{4d_{j,k}}{2^j\delta}\tilde{\psi}\left(\frac{t}{2^j\delta} - k\right) \tag{18}$$

The proof is quite straightforward using equations (10) and (13).

7.3.3 Reconstruction of the derivative

Equation (18) computes the coefficients of the derivative from the coefficients of the signal. The bases, and hence, the filters, are not the same. This section computes the "derivative" filters from the original ones.

Proposition 4 (Low pass filters) *Let z denote the advance operator. Then the low pass filters $\tilde{h} = \sum \tilde{h}_k z^{-k}$ and $\tilde{h}^* = \sum \tilde{h}^*_k z^{-k}$ satisfy*

$$\tilde{h}(z) = \frac{2}{1+z^{-1}}h(z) \text{ and } \tilde{h}^*(z) = \frac{1+z}{2}h^*(z) = z\frac{1+z^{-1}}{2}h^*(z) \tag{19}$$

Taking $z = e^{i\omega}$ in (11,12) proves the proposition.
Observe that $\tilde{h}$ is computed from h by a euclidian division (this how the computation is actually implemented), and $\tilde{h}^*$ is obtained by a polynomial product.

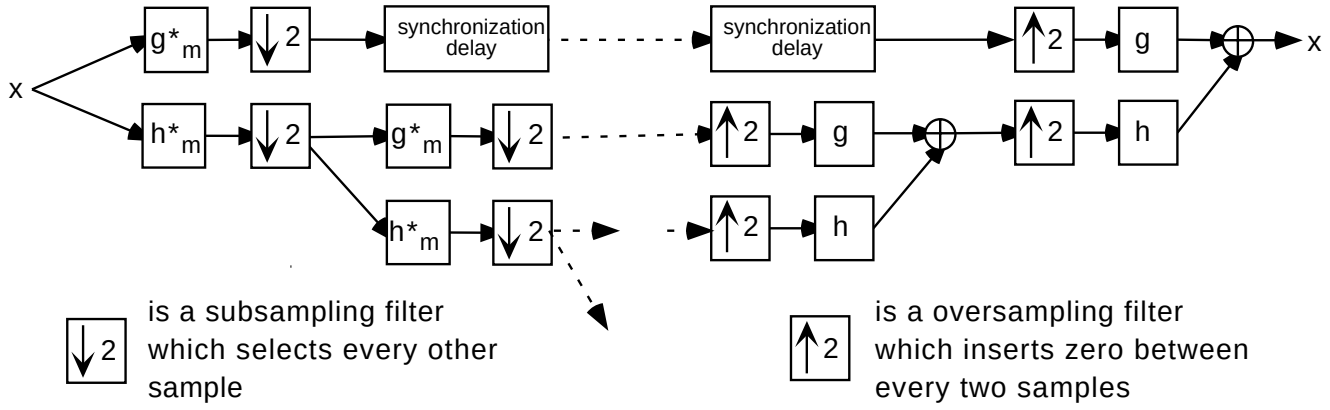

Figure 1: The decomposition and reconstruction in a wavelet basis is implemented by a cascade of filter banks. The wavelet and scale coefficients are computed in the left part of the figure. The reconstruction of the signal from these coefficients is performed in the right part. For a given filter f, f_m denotes its mirror filter defined by $f_m[n] = f[-n]$. Synchronization delays may be required on the non processed channels to compensate for the non causality of some filters.

Proposition 5 (High pass filters) *The high pass filters $\tilde{g}$ and $\tilde{g}^*$ satisfy*

$$\tilde{g}(z) = \frac{1-z^{-1}}{2} g(z) \text{ and } \tilde{g}^*(z) = -\frac{2}{z-1} g^*(z) = -z^{-1} \frac{2}{1-z^{-1}} g^*(z) \quad (20)$$

The coefficients $\tilde{g}_k$ are defined by the scaling equation:

$$\tilde{\psi}(t) = \sqrt{2} \sum \tilde{g}_k \tilde{\phi}(2t-k)$$

Differentiating the previous equation and using (13) gives the expression of $\tilde{g}$. Using the same technique on the scaling equation of $\tilde{\psi}^*$ gives $\tilde{g}^*$.
As noticed before, the mirror filters of $\tilde{h}^*$ and $\tilde{g}^*$ are used at the decomposition. While they are not directly used in the differentiation process (only the the reconstruction filters are used), it is important to study them in order to characterize the various delays created by possible non causalities of the filter banks. Ignoring these delays leads to a "reconstructed" signal which is completely distorted.
Equations (19,20) imply

$$\tilde{h}^*_m(z) = \frac{1+z^{-1}}{2} h^*_m(z) \ , \ \tilde{g}^*_m(z) = \frac{2}{1-z^{-1}} g^*_m(z) \quad (21)$$

Now we can pay some attention to the causality of the filters. In general, not all of the filters h^*_m, g^*_m, h and g are causal; for instance, if the wavelet basis is orthogonal, then $h^*_m[n] = h[-n]$. To preserve synchronicity of the various signal channels, a delay d is applied to the non processed signals at the decomposition, and a delay rd is applied at the reconstruction. This is equivalent to putting the filters under the form

$$h(z) = z^{rd} h_r(z^{-1}) \ , \ g(z) = z^{rd} g_r(z^{-1}) \ , \ h^*_m(z) = z^d h^*_r(z^{-1}) \ , \ g^*_m(z) = z^d g^*_r(z^{-1})$$

where h_r, g_r, h^*_r and g^*_r are polynomials. The previous factorization is used in (19,21) to get the decomposition of the "derivative filters":

$$\tilde{h}(z) = z^{rd} \tilde{h}_r(z^{-1}) \quad \text{with} \quad \tilde{h}_r(z^{-1}) = \frac{2}{1+z^{-1}} h_r(z^{-1}) \quad (22)$$

$$\tilde{g}(z) = z^{rd} \tilde{g}_r(z^{-1}) \quad \text{with} \quad \tilde{g}_r(z^{-1}) = \frac{1-z^{-1}}{2} g_r(z^{-1}) \quad (23)$$

$$\tilde{h}^*_m(z) = z^d \tilde{h}^*_r(z^{-1}) \quad \text{with} \quad \tilde{h}^*_r(z^{-1}) = \frac{1+z^{-1}}{2} h^*_r(z^{-1}) \quad (24)$$

$$\tilde{g}^*_m(z) = z^d \tilde{g}^*_r(z^{-1}) \quad \text{with} \quad \tilde{g}^*_r(z^{-1}) = \frac{2}{1-z^{-1}} g^*_r(z^{-1}) \quad (25)$$

As a consequence, during the reconstruction of the derivative, the same delays are used as in the reconstruction of the original signal. Figure 2 illustrates the differentiation process. Observe that the computations on the coefficients are very simple:

- the wavelet coefficients are multiplied by a scale factor which is due to the fact that $d/dt(x(\lambda t)) = \lambda dx/dt(\lambda t)$
- the scale coefficients are computed by the classical first order finite difference scheme

7.4 Using wavelets theory to improve the differentiation operator

Specific properties of the wavelet analysis can be used to improve the differentiation process.

7.4.1 Regularization

Jaffard [17] and Daubechies [8] have related the pointwise Lipschitz regularity of a signal to the decay of its wavelet coefficients at the fine scales. It is assumed that the wavelet ψ is C^1 and compactly supported, and that the related scaling function is orthogonal.

Theorem 6 *If x is Lipschitz-α at a with $0 < \alpha < 1$, then its wavelet coefficients $d_j[k]$ satisfy*

$$\max_k \left[|d_j[k]|\mathrm{dist}(a, \mathrm{supp}(\psi_{-j}(t-k)))\right] = O\left(2^{-j(\frac{1}{2}+\alpha)}\right) \tag{26}$$

as $j \to +\infty$. Conversely, define, for $\epsilon > 0$,

$$S(a, j, \epsilon) = \{k \in \mathbb{Z}; \mathrm{supp}(\psi_{-j}(t-k)) \cap \ ]a-\epsilon, a+\epsilon[\ \neq \emptyset\}$$

If , for some $\epsilon > 0$ and $0 < \alpha < 1$,

$$\max_{k \in S(a,j,\epsilon)} |d_j[k]| = O\left(2^{-j(\frac{1}{2}+\alpha)}\right) \tag{27}$$

then x is Lipschitz-α at a.

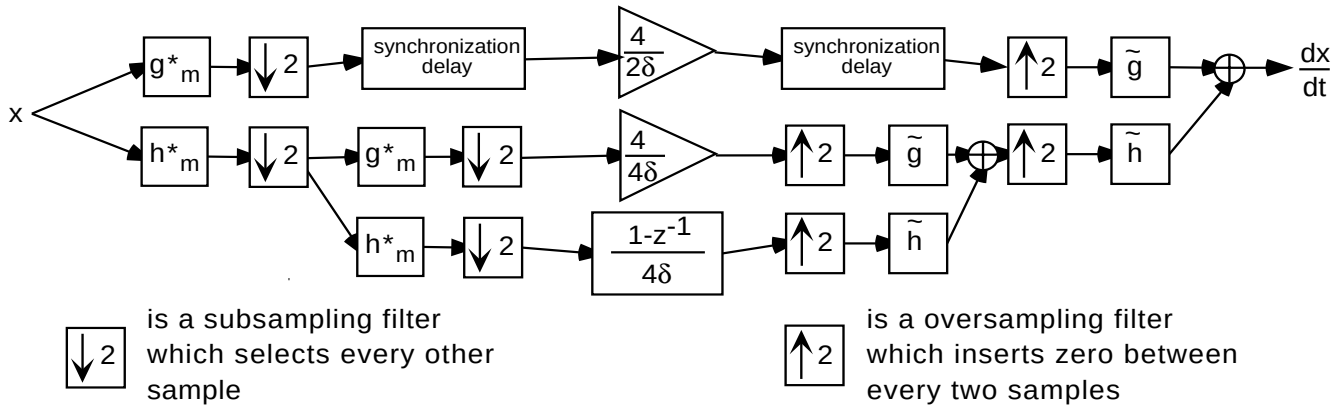

Figure 2: The derivative is reconstructed from finite differences on the scaling coefficients and scaled wavelet coefficients using the synthesis filters given by propositions 4 and 5. This example uses two cascades.

This result can be extended to Lipschitz-α regularity with $p-1 < \alpha < p$ by using a wavelet of class C^p with p vanishing moments. It proves that the pointwise regularity of a signal can be analyzed *and* controlled through its wavelet coefficients. For instance, setting 0 all the wavelet coefficients beyond a certain scale guarantees that the resulting signal has the maximum regularity that is allowed buy the wavelet. Hence the scaling operator P_δ is a regularization operator. Moreover, it accurately approximates regular signals:

Theorem 7 (Fix-Strang [27]) *Let $p \in \mathbb{N}$. The following three conditions are equivalent*

- *for any $0 \leq k \leq N$, there exists a polynomial θ_k of degree k such that*

$$\sum_{n=-\infty}^{+\infty} \theta_k(n)\, \phi(t-n) = t^k \tag{28}$$

- *ψ^* has $N+1$ vanishing moments*
- *$\exists C \left(\forall x \in H^{N+1}(\mathbb{R})\right) (\forall j \leq 0)$*

$$\|P_\delta x - x\|_{L^2} \leq C\delta^{N+1} \|x^{(N+1)}\|_{L^2}$$

where H^{N+1} is the Sobolev space of functions with $N+1$ derivatives in $L^2(\mathbb{R})$.

Putting together theorems 6 and 7 indicates that a regularization by wavelet coefficient control will be localized to the neighborhood of singularities, and that the regular part will be efficiently represented by their coarse scale approximation P_δ.

Here is an application example. The control objective is to make the oscillatory system

$$y = \frac{1}{10^{-4}s^2 + 10^{-4}s + 1} u \tag{29}$$

follow a piecewise constant setpoint c. The abstract answer is to take

$$u = 10^{-4}\frac{d^2c}{dt^2} + 10^{-4}\frac{dc}{dt} + c, \tag{30}$$

which is impossible since c is not differentiable. Tracking c is an badly conditioned problem.

Since the static gain is 1, the output may follow very slow setpoints. Figure 3 features the piecewise constant setpoint and the response of the systems to it. The system is not damped enough to mask the response to the steps. To improve the response, the setpoint input is locally regularized by the scale operator P_δ, e.g., the wavelet coefficients are set to zero beyond a given scale. Figure 4 illustrate the response to the regularized input obtained by setting the coefficients of the six first wavelet scales to 0. The advantage of this method is that it is model independent. The response is, however, a little slack. If there is some confidence in the model, then a regularized setpoint can be used in the inversion formula (30). Figure 5 shows the resulting input and output.

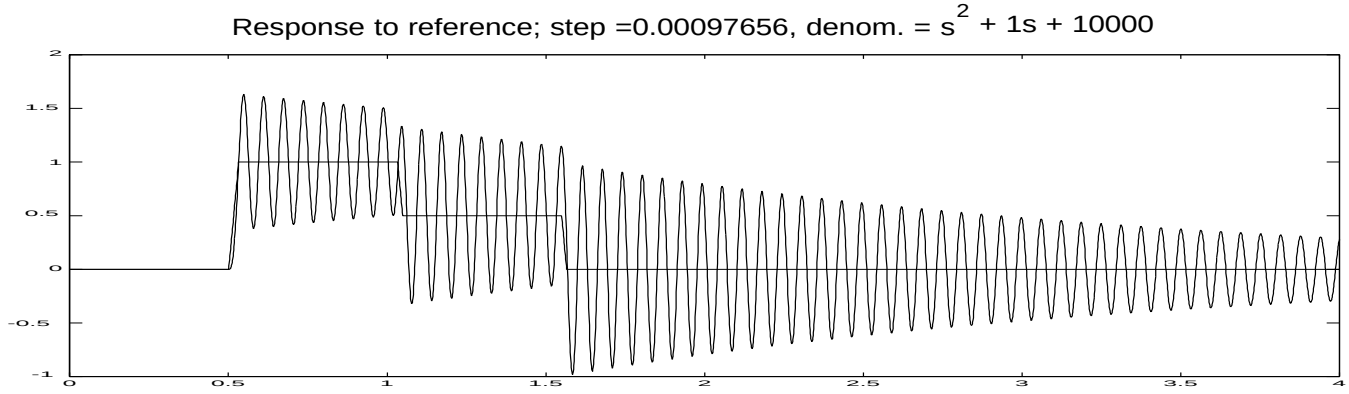

Figure 3: Piecewise constant setpoint and system response to it.

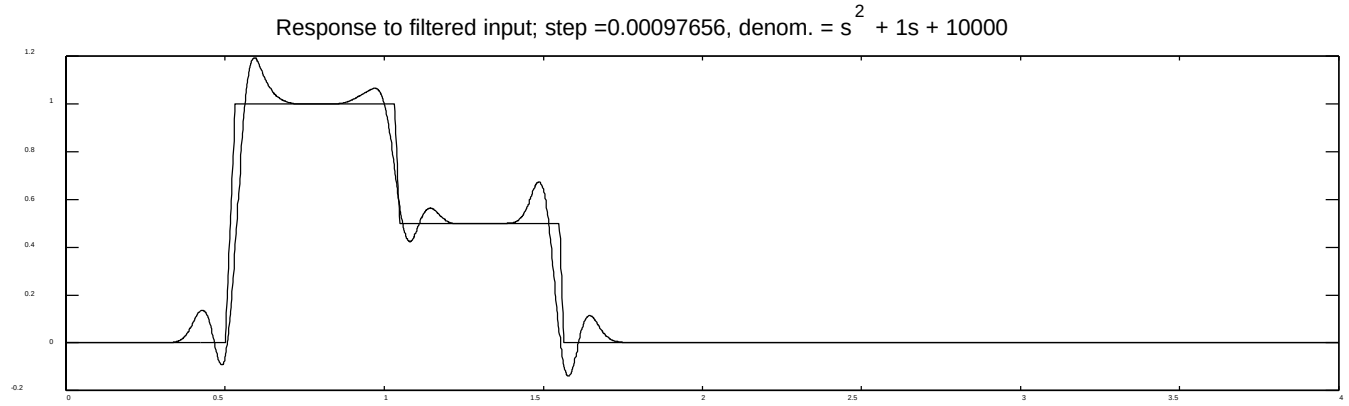

Figure 4: Response to a filtered setpoint computed by setting the wavelet coefficients to zero at the first 6 scales. The wavelets are a biorthogonal spline wavelet systems with 4 and 2 vanishing moments.

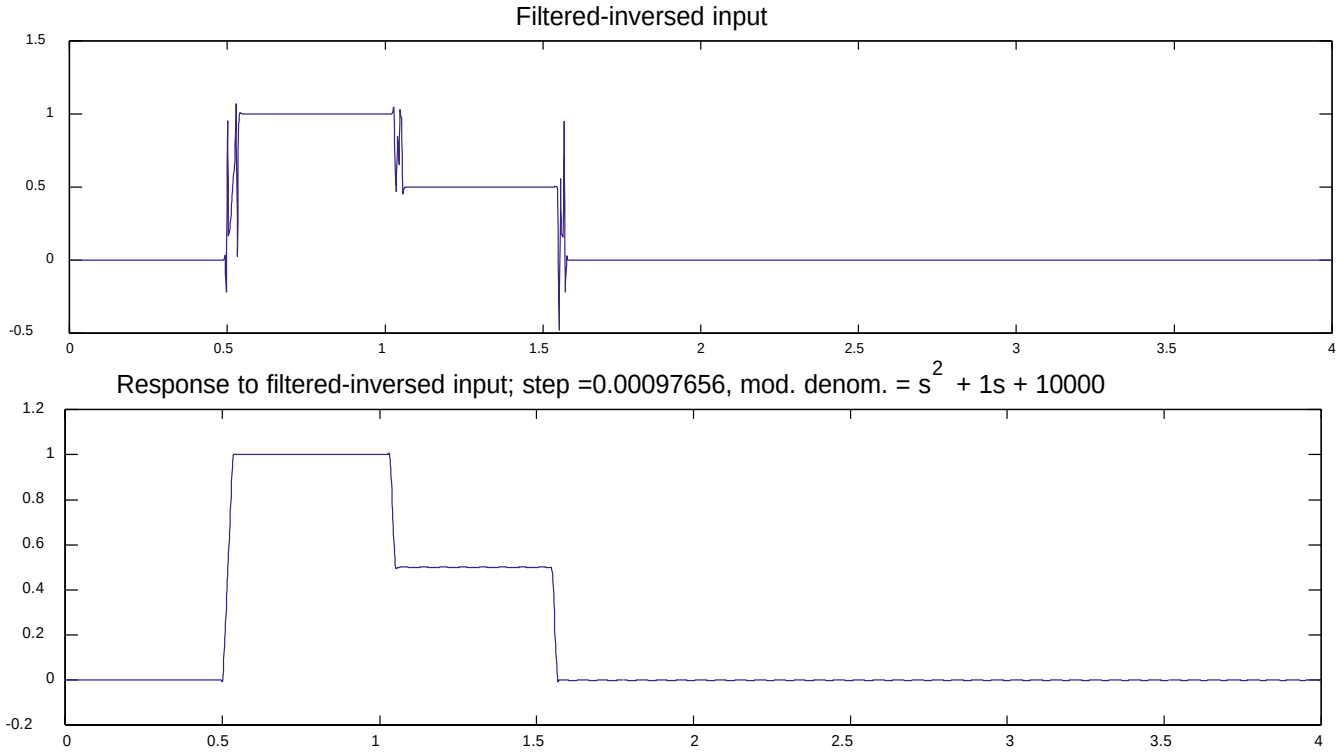

Figure 5: Response to an input computed by applying the inverse system to a filtered setpoint obtained by setting to zero the first two wavelet coefficients. The wavelets are the same as in figure 4.

7.4.2 Denoising

Donoho and Johnstone [13] have proved that wavelet thresholding is almost as efficient for estimating piecewise regular signals as an ideal statistic nonlinear estimator. Hard wavelet thresholding sets to 0 all wavelet coefficients which are below a given threshold. The threshold is

$$T = \sigma\sqrt{2\log_e N} \tag{31}$$

where σ is the noise variance and N the number of parameters. Removing small wavelet coefficients removes most of the noise contribution while preserving the large coefficients which represent the sharp transients of the signal.

Wavelet thresholding can be applied before differentiation to denoise the signal. Since the differentiation operator is diagonal with respect to the wavelet coefficients, thresholding before the differentiation is equivalent to a threshold after the differentiation (but with a different basis). Figure 6 shows a noisy signal and figure 7 compares the derivative of an ideal Wiener linear estimator (which uses information on the signal which generally not available) with the derivative of a wavelet hard threshold estimator, which uses only a covariance information.

Wavelet thresholding can be performed on-line like the wavelet decomposition and reconstruction.

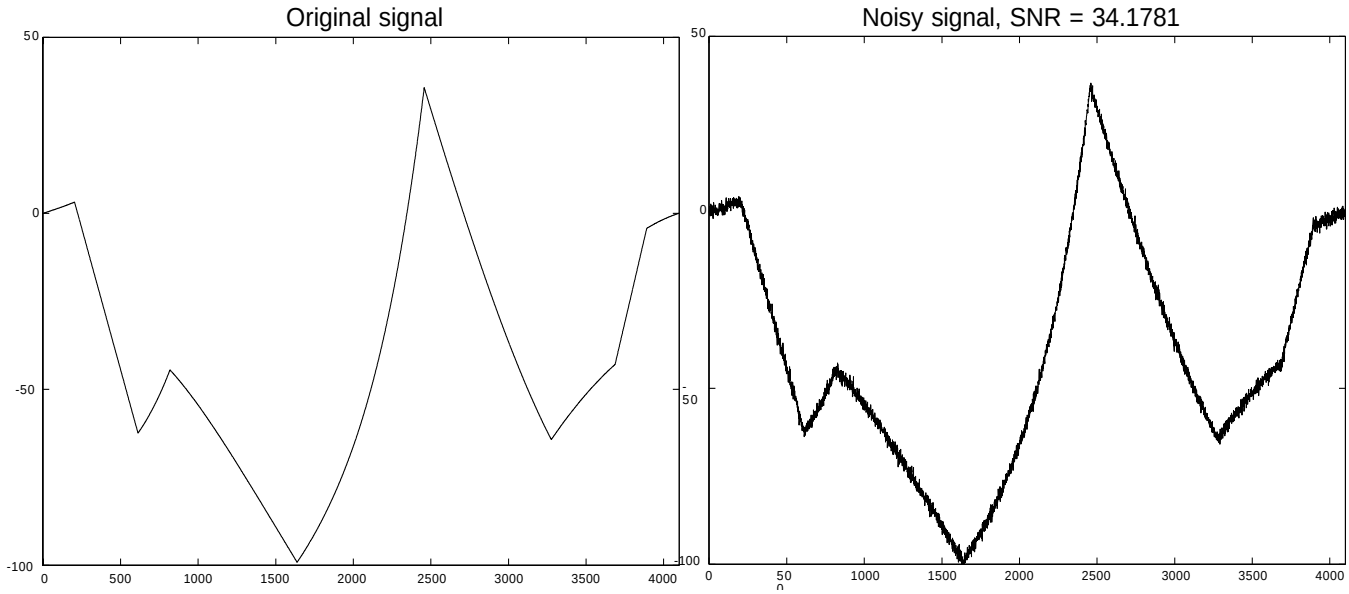

Figure 6: A piecewise polynomial signal (left) with 4096 samples at intervals of 0.0025 is perturbed by a white noise. Signal to Noise Ratio is 34.2 dB. The derivative of the left signal has jumps.

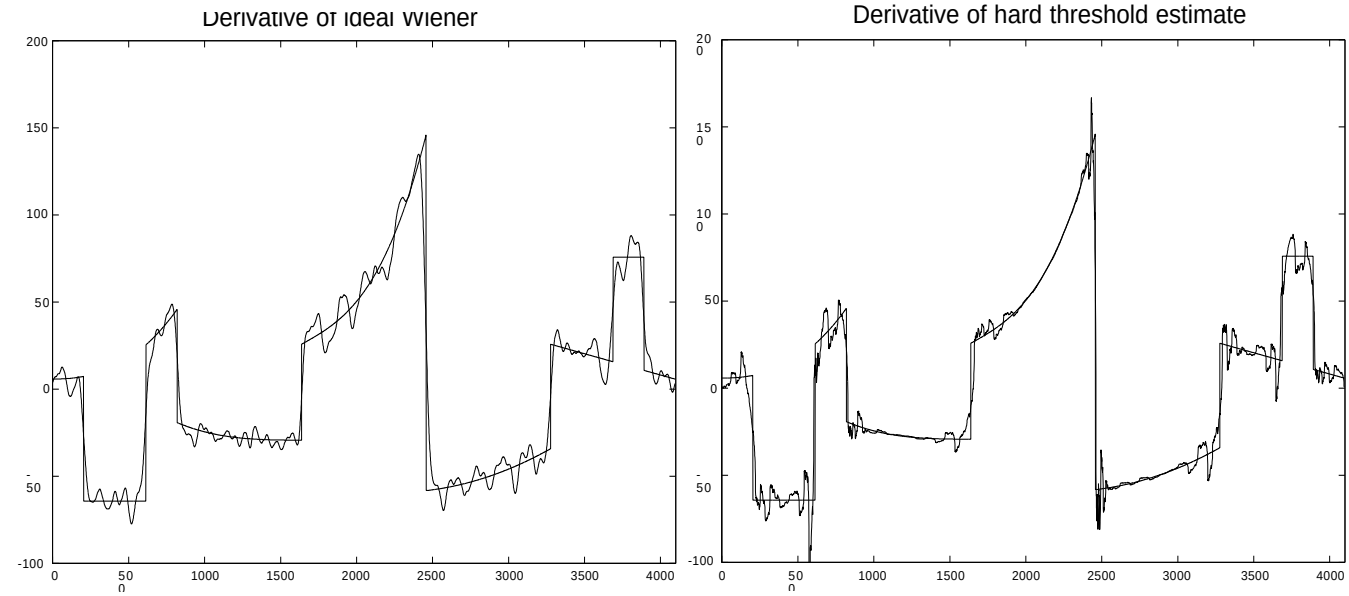

Figure 7: The left plot compares the derivative of an ideal Wiener estimate to the derivative of the unperturbed signal. The right plot compares the derivative of a hard wavelet threshold estimate to the derivative of the unperturbed signal. The wavelets used are Symlets with length 8.

7.5 Wavelet summary

The differentiation process can be decomposed in four steps:

- the wavelet transform (decomposition) is performed using filter banks
- suitable signal processing is performed on the wavelet transform, using its specific abilities
- the wavelet transform of the derivative is computed from the previous filtered transform
- the derivative signal (or an approximation of it) is reconstructed from its wavelet transform.

Since step 3 is particularly straightforward, all necessary signal processing can be done before it. For instance, wavelet thresholds are equivalent before and af-

ter the differentiation (up to a scaling factor), since the differentiation operator is diagonal in this representation.
Gains to be expected from using the wavelet transform by contrast to a classic numerical scheme on a mollified signal are in the ability to fine tune the wavelet coefficients in order to locally smooth the signal or denoise it while preserving transients.

References

[1] R. A. Adams, *Sobolev Spaces*, Academic Press, New York, 1975.

[2] A. Akansu, R. Haddad, *Multiresolution Signal Decomposition*, Academic Press, 1993.

[3] A. Akansu, M. J. Smith (eds.), *Subband and Wavelet Transforms*, Kluwer, 1995.

[4] R. S. Anderssen, F. de Hoog, M. Hegland, A stable finite difference Ansatz for higher order differentiation of non-exact data, *Bull. Austral. Math. Soc.*, **58**(1998), 223–232.

[5] R. S. Anderssen, F. R. de Hoog, Finite difference methods for the numerical differentiation of non-exact data, *Computing*, **33**(1984), 259–267.

[6] A. Cohen, I. Daubechies, J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, *Commun. Pure Appl. Math.*, **45**(1992), 485–560.

[7] A. M. Dabroom, H. K. Khalil, Discrete-time implementation of high-gain observers for numerical differentiation, *Internat. J. Control*, **72**(1999), 1523–1537.

[8] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

[9] S. Diop, V. Fromion, J. W. Grizzle, A resettable extended Kalman filter based on numerical differentiation, Research Note, Laboratoire des Signaux & Systèmes, Gif sur Yvette, France, 2000.

[10] S. Diop, J. W. Grizzle, S. Ibrir, On regularized numerical observers, *in Proceedings of the IEEE Conference on Decision and Control*, IEEE Press, New York, 1999, p. Paper no. 0526.

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

[12] S. Diop, S. Ibrir, On numerical observers. Application to the control of simple academic adaptive control example, *in Proceedings of the European Control Conference*, European Union Control Association, Brussels, Belgium, 1997, p. Paper no. 429.

[13] D. Donoho, I. Johnstone, Ideal spatial adaptation via wavelet shrinkage, *Biometrika*, **81**(1994), 425–455.

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

[15] W. Gander, U. von Matt, Smoothing filters, *in* W. Gander, J. Hřebíček (eds.), *Solving Problems in Scientific Computating Using Maple and Matlab. 3rd Edition*, Springer, Berlin, 1997, pp. 121–139.

[16] M. Hegland, R. S. Anderssen, A mollification framework for improperly posed problems, *Numer. Math.*, **78**(1998), 549–575.

[17] S. Jaffard, Exposants de Hölder en des points donnés et coefficients d'ondelettes, *C. R. Acad. Sci. Paris Sér. I*, **308**(1989), 79–81.

[18] A. Kirsch, *An Introduction to the Mathematical Theory of Inverse Problems*, New York, Springer-Verlag, 1996.

[19] P. G. Lemarie-Rieusset, Analyses multi-résolutions non orthogonales, commutation entre projecteurs et dérivation et ondelettes vecteurs à divergence nulle, *Revista Mat. Iberoamericana*, **8**(1992), 221–237.

[20] S. Mallat, An efficient image representation for multiscale analysis, *in Proc. Machine Vision Conference*, Lake Taho, February 1987.

[21] S. Mallat, Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$, *Trans. Amer. Math. Soc.*, **315**(1989), 69–87.

[22] S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Recognition Mach. Intell.*, **11**(1989), 674–693.

[23] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1998.

[24] Y. Meyer, *Wavelets and Operators*, Cambridge University Press, Cambridge, MA, 1992.

[25] A. Savitzky, M. J. E. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.*, **36**(1964), 1627–1639, see also [26].

[26] J. Steinier, Y. Termonia, J. Deltour, Comments on smoothing and differentiation of data by simplified least square procedure, *Anal. Chem.*, **44**(1972), 1906–1909.

[27] G. Strang, G. Fix, A Fourier analysis of the finite element variational method, *Constructive Aspects Funct. Anal.*, 1971, 796–830.

[28] G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Boston, MA, 1996.

[29] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice-Hall, Englewood Cliffs, NJ, 1993.

[30] M. Vetterli, Kovacevic.J., *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.