In the Proceedings of the IEEE Conference on Decision and Control, Sydney, Australia, 2000.

On Numerical Differentiation Algorithms for Nonlinear Estimation

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

Abstract. Practical methods of differentiating a signal known only through its on-line samples are much needed, given the numerous areas in control theory and practice where differentiation is encountered. This communication presents theoretical as well as implementation details on several numerical differentiation algorithms which may be useful in the area of nonlinear estimation. In particular, these algorithms may be used as ingredients for alternative solutions to the long-standing problem of observer design for nonlinear systems.

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

1 Introduction

A ripe application area for numerical differentiation is observer design for nonlinear systems. The basis for this has already been described in [6, 7, 5, 3]. This communication focuses its attention on theoretical and implementation aspects of several numerical differentiation algorithms.

The difficulty with the implementation of differentiation is that it is an operator which is *discontinuous* 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 both $\mathcal{C}^0(a,b)$ and $\mathcal{C}^1(a,b)$ with the uniform norm,

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

Suppose that y is known through some experimental data and hence is uncertain. Write y as

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

where,

$$\varepsilon(t) = \sigma \sin \frac{2\pi t}{\sigma^2} \,,$$

and hence

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

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 discontinuity 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 $t - T$ and 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

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

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 uncertainty on the data. The backward difference scheme thus has no immunity against the higher frequency content of this uncertain signal which tends to drown out the derivative of the true signal.

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

This communication is an attempt to collect details on some of the various numerical differentiation algorithms

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.

which are most relevant to control theoretic issues in online estimation. The paper is organized as follows. We first present differentiation as a typical ill-posed inverse problem and relate its study to regularization techniques. Among the latter, mollification is close to the control theory notion of filtering. For a brief presentation see [5]. Here, three differentiation schemes are presented in more detail. One is the so-called Savitzky-Golay differentiation scheme. The second is an averaged finite difference differentiation scheme. And the last uses wavelet transforms.

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 [9]):

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

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 (ii) and (iii) 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* [9] if

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

for all y.

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

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

consisting of positive singular values $\tau_i \geq \tau_{i+1}$, (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 bases $(x_i)_{i \in \mathbb{N}}$, $(y_i)_{i \in \mathbb{N}}$ 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 of $K(x) = y$ 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 a differentiation scheme since, with experimental data, the high frequency content of ε would tend to drown out 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 sequence $(y_i)_{i \in \mathbb{N}}$.

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

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

The norm of the estimation error

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

is

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

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 over smoothing the derivative. One possible trade-off consists of choosing S and requiring

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

Then from the above expression of $||\widetilde{\dot{y}}(t)||^2_{L^2}$ 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 a bound on ε, a bound on its spectral power density.

3 The Savitzky-Golay differentiation scheme

This is perhaps the simplest differentiation scheme. It is polynomial in the sense that the basis functions are polynomials in t. It does not employ orthogonal polynomial basis functions, but rather uses the basis of monomials. We have followed [10] in naming this differentiation scheme after A. Savitzky and J. E. Golay. 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 and for all) and also for its quality which may be sufficient in many cases.

In what follows, we assume that the data are sampled at some given frequency, f with $T = 1/f$. 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

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

As may be suggested by the previous spectral regularization methods, we may truncate formula (4) by considering finite dimensional subspaces X_N of $\mathcal{C}^r_0\,(a,b)$ generated by subsequences $(x_i)_{1\leq i\leq N}$ of a singular system of $\mathcal{C}^r_0\,(a,b)$. In practice, we extend this viewpoint by choosing an arbitrary finite dimensional subspace $\mathbf{F}$ of $\mathcal{C}^r_0\,(a,b)$ with basis $(\mathfrak{b}_j(t))_{1\leq j\leq N}$. 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. Practical choices for $\mathbf{F}$ include 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, $\alpha = (\alpha_1, \ldots, \alpha_N)'$, and β is the $(W \times N)$-matrix with coefficients $\beta(i,j) = \mathfrak{b}_j(t_i)$. 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 is then $\mathfrak{b}_j(t_i) = (i - W + \delta)^{j-1}$. 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 denotes matrix transpose. Let $\lambda > 0$ be a positive constant. Then a better estimate of α in $\widehat{y}(t) = \sum_{j=1}^{N} \alpha_j \mathfrak{b}_j(t)$ 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 greater than or equal to λ.

We obtain the vector of the $k+1$ first derivatives:

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

where C is computed offline. The explicit regularization parameter may be chosen in a 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 and is skipped.

Finally let us note that δ is a good measure of the delay introduced by the differentiation scheme. Since we know its value, corrective action may be taken in an overall observer design to compensate for this delay.

4 The averaged finite differences differentiation scheme

The idea (see [1] for more details) is to take quite standard finite difference operators

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

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

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

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 we define the estimate of the rth derivative of y as

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

More explicitly, let $2\ell+1$ be the number of data points needed in the finite difference operator $\Delta^{(r)}_{f,\ell,k,q}$. 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 $\Delta^{(r)}_{f,\ell,k,q}$ 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 [1]. 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 [1] 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.

5 The wavelet differentiation scheme

We refer the reader to [2, 13] for more details on wavelets. Functions considered here are assumed to be in the space, $L^2(\mathbb{R})$, of real-valued functions of the real variable t which are square integrable. This space $L^2(\mathbb{R})$ is equipped with the usual inner product $\langle f, g\rangle = \int_{\mathbb{R}} f(t)g(t)\mathrm{d}t$. A scaling function is a function with unit average which satisfies a scaling equation

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

where $(h_k)_{\mathbb{Z}}$ is a sequence of real (or perhaps complex) numbers. Two scaling functions ϕ and ϕ^* form a pair of conjugate scaling functions if they satisfy $\langle \phi(t-i), \phi^*(t-j)\rangle = \delta_{i,j}$. A scaling function is orthogonal if it is its own conjugate. In practice, if the g_k's are the coefficients for ϕ^*, we have $g_k = (-1)^{k-1} h^*_{1-k}$ and $g^*_k = (-1)^{k-1} h_{1-k}$. Given a function x, we denote by $x_{j,n}$ the function $x_{j,n}(t) = \sqrt{2^{-j}}x(2^{-j}t-n)$. For suitable h and h^*, the families $(\psi_{j,n})_{j,n\in\mathbb{Z}}$ and $(\psi^*_{j,n})_{j,n\in\mathbb{Z}}$ define dual Riesz bases of $L^2(\mathbb{R})$ [4, 13]. If h and h^* are compactly supported, then any signal $y(t)$ with locally finite energy can be decomposed as

$$y = \sum_{n\in\mathbb{Z}} c_{J,n}\phi_{J,n} + \sum_{j\in\mathbb{Z}, j\le J} d_{j,n}\psi_{j,n} \tag{6}$$

with $c_{j,n} = \langle y, \phi^*_{j,n}\rangle$ and $d_{j,n} = \langle y, \psi^*_{j,n}\rangle$. The coefficients $c_{j,n}$ and $d_{j,n}$ can be computed recursively across the scales j using finite impulse response filter banks (see Mallat [13]): for the direct wavelet transform

$$c_{j+1,p} = \sum_{n=-\infty}^{+\infty} h^*_{n-2p} c_{j,n} \tag{7}$$

$$d_{j+1,p} = \sum_{n=-\infty}^{+\infty} g^*_{n-2p} c_{j,n}, \tag{8}$$

while for the inverse wavelet transform

$$\begin{aligned} c_{j,p} \quad = \quad & \sum_{n=-\infty}^{+\infty} h_{p-2n}\, c_{j+1,n} + \\ & \sum_{n=-\infty}^{+\infty} g_{p-2n}\, d_{j+1,n}. \end{aligned} \tag{9}$$

In practice, j is lower bounded and the coefficients $c_{j,n}$ at the finest scale are taken to be the signal samples.

How do we estimate the derivatives of a signal through its wavelet transforms? This is answered by the following result. Let ϕ and ϕ^* be two conjugate scaling functions such that ϕ is $\mathcal{C}^{1+\epsilon}$ for some $\epsilon > 0$. Then (see Lemarié-Rieusset [12]) there exist two conjugate scaling functions $\tilde{\phi}$ and $\tilde{\phi}^*$ such that:

$$\begin{cases} \dfrac{\mathrm{d}\phi}{\mathrm{d}t}(t) = \tilde{\phi}(t) - \tilde{\phi}(t-1) \\ \phi^*(t+1) - \phi^*(t) = \dfrac{\mathrm{d}\tilde{\phi}^*}{\mathrm{d}t}(t) \end{cases}$$

The related wavelets are defined by

$$\tilde{\psi}(t) = \frac{1}{4}\frac{\mathrm{d}\psi}{\mathrm{d}t}(t) \text{ and } \frac{\mathrm{d}\tilde{\psi}^*}{\mathrm{d}t}(t) = -4\psi^*(t)\,.$$

If y is differentiable then the decomposition of $\dot{y}$ in the basis defined by $(\tilde{\phi}, \tilde{\phi}^*)$ is

$$\dot{y} = \sum_{n\in\mathbb{Z}} \frac{c_{J,n} - c_{j,n-1}}{2^J}\tilde{\phi}_{J,n} + \sum_{j\leq J}\sum_{n\in\mathbb{Z}} \frac{4d_{j,n}}{2^j}\tilde{\psi}_{j,n}\ .$$

This equation computes the coefficients of the derivative from the coefficients of the signal. The derivative is computed using suitable 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

$$\tilde{h}(z) = \frac{2}{1+z^{-1}}h(z) \text{ and } \tilde{h}^*(z) = \frac{1+z}{2}h^*(z) \qquad (10)$$

The high pass filters $\tilde{g}$ and $\tilde{g}^*$ satisfy

$$\tilde{g}(z) = \frac{1-z^{-1}}{2}g(z) \text{ and } \tilde{g}^*(z) = -\frac{2}{z-1}g^*(z) \qquad (11)$$

Formulae (10, 11) and (7–9) are then used to compute the derivative from its coefficients.

Specific properties of the wavelet analysis can be used to regularize the differentiation operator. For instance, Jaffard [11, 4] has related the pointwise Lipschitz regularity of a signal to the decay of its wavelet coefficients at the fine scales. Roughly speaking, a signal is Lipschitz α at t iff its wavelet coefficients decay like $2^{-j(\alpha+\frac{1}{2})}$ in the neighborhood of t when $j \to +\infty$. This is to say that the pointwise regularity of a signal can be analyzed *and* controlled through its wavelet coefficients. For example, setting to 0 all the wavelet coefficients which are beyond a certain scale yields the maximum regularity allowed by the wavelet. The regularized signal is then computed by considering $d_{j,n} = 0$ in the reconstruction procedure. Moreover, this procedure accurately approximates regular signals (Strang & Fix [14]).

When facing noisy signals, Donoho and Johnstone [8] (see also [13]) have proposed wavelet (*hard* and *soft*) *thresholding*. Hard thresholding sets to 0 all coefficients which are below a given threshold. The latter is theoretically related to the noise variance and to the number of data samples; it can also be determined based on practical issues. Zeroing small wavelet coefficients removes most of the noise contribution while preserving the large coefficients which represent the sharp transients of the signal.

References

[1] 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.

[2] C. S. Burrus, R. A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms. A Primer*, Prentice-Hall, Upper Saddle River, NJ, 1998.

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

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

[5] 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.

[6] 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.

[7] 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.

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

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

[10] 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.

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

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

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

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