Electronic supplement to the paper: Solving the Couette inverse problem using a wavelet-vaguelette decomposition

Christophe Ancey

École Polytechnique Fédérale de Lausanne,

Ecublens, 1015 Lausanne, Switzerland

(Dated: November 1, 2004)

I. INTRODUCTION

The objective is to recover the shear-rate function $\dot{\gamma}$ from the rotational velocity ω in a coaxial cylinder (Couette) rheometer. The shear-rate function is solution to the Fredholm equation

$$\omega(\tau) = \frac{1}{2}\int_{\beta\tau}^{\tau} \frac{\dot{\gamma}(S)}{S} dS, \tag{1}$$

where $\beta = (R_1/R_2)^2$, $M(\omega)$, where τ is the shear stress, $\dot{\gamma}$ denotes the shear rate, and ω is the rotational velocity of the inner cylinder (Coleman *et al.*, 1966). The shear stress τ exerted on the inner cylinder of radius R_1 can be directly related to the measured torque M by $\tau = \alpha_1 M$, with $\alpha_1 = 1/(2\pi R_1^2)$, independently of the form of the constitutive equation.

Equation (1) can be represented in a generic form: $\omega(\tau) = (K\dot{\gamma})(\tau)$, where K is the integral operator

$$(Kf)(z) = \int_{\beta z}^{z} \frac{f(x)}{x} dx, \tag{2}$$

with β a constant parameter ($\beta < 1$). The operator K maps a function space $\mathcal{H}$ to another function space $\mathcal{H}'$; a typical example for $\mathcal{H}$ is the space $L_2(\mathbb{R})$ including all the functions f that are square-integrable: $\int_{\mathbb{R}} |f(x)|^2 dx < \infty$.

II. PROPERTIES OF THE INTEGRAL OPERATOR K

The integral operator K introduced in Eq. (2) is linear and homogeneous of order 0, i.e., $K[f(ax)] = a^0 K[f](ax)$ for any scalar a; the operator is scale invariant. As a consequence, according to Donoho (1995), an operator of this kind admits a wavelet-vaguelette decomposition provided that the wavelet basis is sufficiently regular; here the regularity conditions given by Donoho imply that Daubechies wavelets of at least order 3 must be used. In order to proceed with this approach, we need to define the adjoint operator.

The adjoint operator K^* is defined as $\langle Kf,\ g\rangle = \langle f,\ K^*g\rangle$. After basic integral computations (integration by parts), it can be shown that the adjoint operator is

$$(K^*g)(x) = \frac{1}{2x}\int_{x/\beta}^{x} g(y) dy. \tag{3}$$

A few words must be said about the domain over which operator K is defined and its action on the smoothness of the image of f. In order to characterize the regularity of functions, we can introduce *Sobolev spaces* H_s, indexed by the real s and defined as follows

$H_s = \{f \in L_2(\mathbb{R}) \mid ||f||_s < \infty\}$, with the Sobolev norm: $||f||_s = (\int_{\mathbb{R}} (1+|\xi|^2)^s |\hat{f}(\xi)|^2 d\xi)^{1/2}$ where $\hat{f}$ denotes the Fourier transform of f: $\hat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-i\xi x} dx$. The larger the index s is, the faster the Fourier transform $\hat{f}$ must decay; this implies that the highest frequencies in function f must not play a significant role and that function f must be smooth. The particular case $s = 0$ corresponds to $L_2(\mathbb{R})$; discontinuous functions such as the Dirac function belong to spaces H_s with $s < -1/2$, whereas for $\mathcal{C}^k(\mathbb{R})$ functions, we have the following embedding: $H_s \subset \mathcal{C}^k(\mathbb{R})$ for $s > k + 1/2$ ($k \in \mathbb{N}$). In order to characterize the action of K, we need to define it in the frequency domain. By using basic properties of the Fourier transformation, we can show that the Fourier transform of Kf is

$$\widehat{Kf}(\xi) = \frac{1}{\xi}\left(\hat{F}(\xi) - \frac{1}{\beta}\hat{F}\left(\frac{\xi}{\beta}\right)\right),$$

with $\hat{F}$ the integral of $\hat{f}$: $\hat{F}(\xi) = \int_{-\infty}^{\xi} \hat{f}(x) dx$. It follows that, if f is an H_s function, then Kf is also an H_s function since an integration by parts shows that $||Kf||_s \propto ||f||_s$. Therefore the important result is that the integral operator does not alter the regularity of function f (this result is also a consequence of the scale invariance of the operator); if we have information on Kf (e.g. experimental data), we can deduce the regularity of f.

Because of the presence of $1/x$ in the integral operator (see Eq. 2), problems may arise in the vicinity of zero. When $z \to 0$, a constant c in $[\beta z,\ z]$ exists such that $Kf(z) = (1-\beta) z f(c) c^{-1} \approx (1-\beta) f(z)$; thus, if $f(z)$ is finite when $z \to 0$, then the integral is properly defined. Note that the condition $|f(0)| < \infty$ gives $|\int_{\mathbb{R}} \hat{f}(\xi) d\xi| < \infty$ in the frequency domain. Schwarz's inequality implies that f must satisfy $\int_{\mathbb{R}} |\hat{f}(\xi)|^2 d\xi < \infty$, i.e., f must belong to H_0. This proves that Kf is defined for any function f of $L_2(\mathbb{R})$.

In short, we can summarize its properties as follows: K is an integral linear, homogeneous (scale-invariant) operator that maps the space $\mathcal{H} = L_2(\mathbb{R})$ to $\mathcal{H}' = \{g \in L_2(\mathbb{R}) \mid g(0) = 0\}$.

III. WAVELET FUNCTIONS

To define a wavelet basis, we need two basic ingredients:

- the wavelet function ψ, usually referred to as the *mother wavelet* ;
- the *scaling function* ϕ, also called the *father wavelet*.

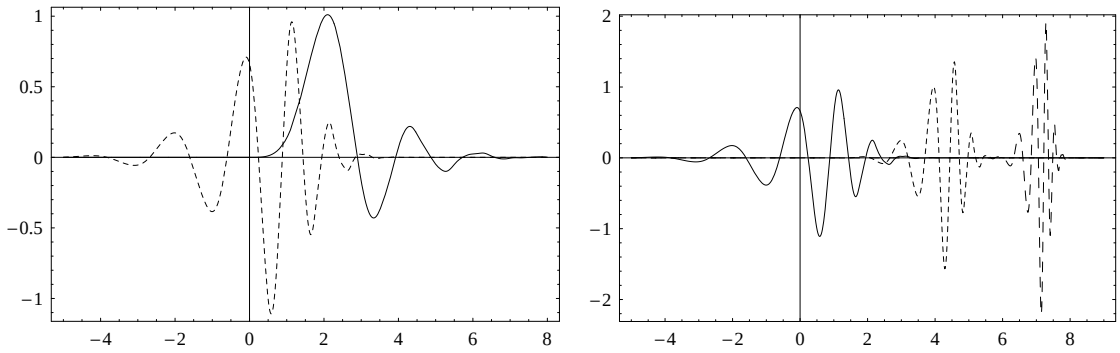

FIG. 1: Left: Scaling function ϕ (solid line) and wavelet function (dashed line) ψ for the D8 Daubechies wavelet. Right: mother wavelet $\psi_{00} = \psi$ (solid line), wavelets ψ_{18} (dashed line), and ψ_{28} (long-dashed line) for D8.

Their supports are denoted by suppψ and suppϕ, respectively. They can be finite or infinite, but here, for numerical reasons, we will give preference to finite-support wavelets. We can then define a family of orthogonal functions by dilation and translation (Louis *et al.*, 1997; Mallat, 1998)

$$\psi_{ij}(x) = 2^{i/2}\psi(2^i x - j) \text{ and } \phi_{ij}(x) = 2^{i/2}\phi(2^i x - j). \tag{4}$$

The index i controls the magnitude and the shape of the functions: the larger i is, the sharper the function ϕ_{ij} or ψ_{ij} is (see Fig. 1). The index j controls the position of the function along the abscissa axis.

When the scaling and wavelet functions are properly defined, these functions provide helpful properties:

1. Pairwise orthogonality: $\langle \psi_{ij},\ \psi_{kl} \rangle = \delta_{ik}\delta_{jl}$ and $\langle \phi_{ij},\ \phi_{ik} \rangle = \delta_{jk}$.
2. Normality: $\int_{\mathbb{R}} \psi^2(x)dx = 1$ and $\int_{\mathbb{R}} \phi^2(x)dx = 1$.
3. Orthogonality $\langle \phi_{ij},\ \psi_{kl} \rangle = 0$ for $k \geq i$.
4. 0-moments: $\int_{\mathbb{R}} \psi(x)dx = 0$ and $\int_{\mathbb{R}} \phi(x)dx = 1$.
5. Scaling properties: there is a usually finite set of coefficients h_k such that:
$$\phi(x) = \sum_{k\in\mathbb{Z}} h_k \sqrt{2}\phi(2x - k). \tag{5}$$
6. If the mother wavelet has m vanishing moments $M_i = \langle x^i,\ \psi(x) \rangle = \int_{\mathbb{R}} \psi(x)x^i dx = 0$, then for any $0 \leq k < p$, the function $\sum_{n\in\mathbb{Z}} n^k \phi(t - n)$ is a polynomial of degree k.

Properties (1) and (2) are fundamental since they allow expression of a function f as $f = \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} \psi_{ij}(x)$ or $f = \sum_{j\in\mathbb{Z}} \phi_{i_0 j}(x)$. Properties (3–5) are the cornerstone of the *multiresolution analysis* (Mallat, 1998): the key point is that it is possible to relate $\phi(x)$ to $\phi(2x-k)$, i.e. by averaging what happens on a fine scale, we can obtain the trend (see Sec. IV). The last property states that if the mother wavelet has m vanishing moments, then any polynomial of degree $m-1$ is fully reproduced by the scaling function ϕ. This has important implications when finding sparse representations of functions, i.e., approximations with a few non-zero coefficients. Indeed, if ψ has m vanishing moments, then ψ is orthogonal to any polynomial of degree $m-1$. Recall that if a function is locally $\mathcal{C}^k$ (i.e., its derivatives $f^{(i)}$ exist and are continuous for $1 \leq i \leq k$), then over a small interval it is approximated by a polynomial of degree k. Therefore, we have $\langle f,\ \psi_{ij} \rangle \approx 0$ for $0 \leq i < m$.

Now a large number of wavelet functions exist that present various properties in terms of support, symmetry, vanishing moments, etc. Here we will use Daubechies wavelets because they are optimal in terms of the support width and number of vanishing moments: indeed, the wider the support of ψ compared to the support of a function f, the more numerous the shift indexes required to span $\mathrm{supp} f$. Furthermore, if f has few isolated singularities and is very regular, it is more convenient to select a wavelet with a large support and many vanishing moments to provide a sparse representation of f. Otherwise, it may be better to decrease the support width of ψ to single out singularities and avoid high-amplitude coefficients. Daubechies wavelets form a group of wavelet functions, usually referred to generically as Dq where q is an integer ($q \geq 1$) representing the order of the wavelet function. The support of ϕ and ψ together with the number of vanishing moments depend on q: $\mathrm{supp}\psi = [1-q, q]$, $\mathrm{supp}\phi = [0, 2q-1]$, and $m = q$. In Figure 1, the left panel represents the D8 scaling and wavelet functions, whereas the right panel shows shifted and dilated wavelets. In the remainder of the paper, we will only consider Dq Daubechies wavelets.

IV. WAVELET APPROXIMATION AND MULTIRESOLUTION REPRESENTATION

We can use wavelets to represent and/or approximate functions in different ways. The most classic is the approximation in terms of scaling functions. Property (1) in Sec. III shows that by translating the functions $\phi_{i_0 j}$ at a given scale i_0, we obtain orthonormal functions.

It is then possible to use $\phi_{i_0 j}$ as a basis of the subspace $\mathcal{H}_{i_0} = \text{span}(\phi_{i_0 j})_{j\in\mathbb{Z}}$ of $L_2(\mathbb{R})$. The projection $\mathcal{P}_{i_0}$ of a function f of $L_2(\mathbb{R})$ onto this subspace is then

$$\mathcal{P}_{i_0} f = \sum_{j\in\mathbb{Z}} a_j \phi_{i_0 j}, \tag{6}$$

where $a_j = \langle f,\ \phi_{i_0 j}\rangle = \int_{\mathbb{R}} f(x)\phi_{i_0 j}(x)dx$. In this projection, the scaling index i controls the accuracy of the representation of f as a series $a_j\phi_{i_0 j}$. The larger i is, the finer the representation of f is, but also the less sparse it is. Indeed, let us consider a function f of $\mathcal{C}^k([a,\ b])$; then, at scale i_0, for the coefficient a_j not to be zero, the shift index must lie within the range J_0: $2^{i_0}a - 2q - 1 < j < 2^{i_0}b$. In the meantime, it can be shown that there a positive constant C exists such that: $||f - \mathcal{P}_{i_0} f|| \leq C2^{-ki_0}||f^{(k)}||$ (Louis *et al.*, 1997; Strang and Nguyen, 1997). Thus, the accuracy varies as 2^{-ki_0} with i_0 and the number of non-zero terms increases approximately as $2^{i_0}(b-a)$. The projection operator $\mathcal{P}_{i_0}$ can also be interpreted as an operator controlling the loss of information across different levels [low-pass filter in signal processing (Mallat, 1998)]: indeed, the approximation of a function f at a resolution level 2^{i_0} is equivalent to a local average of f over neighborhoods of size 2^{i_0}.

An alternative way of representing f is to use the wavelet functions ψ_{ij}, which are pairwise orthogonal. A direct consequence is that $L_2(\mathbb{R}) = \text{span}(\psi_{ij})_{i,j\in\mathbb{Z}}$. Thus for any function f of $L_2(\mathbb{R})$, we have the infinite decomposition

$$f = \sum_{i\in\mathbb{Z}}\sum_{j\in\mathbb{Z}} b_{ij}\psi_{ij},$$

with $b_{ij} = \langle f,\ \psi_{ij}\rangle$, which is not very helpful in this form since it involves an infinite number of terms. However, it is possible to truncate this series by noting that the function ψ_{ij} and ϕ_{kl} are also orthogonal. So we can replace the previous decomposition by

$$f(x) = \sum_{j\in J_0} \alpha_{i_0 j}\phi_{i_0 j}(x) + \sum_{i\geq i_0}^{\infty}\sum_{j\in J_i} \beta_{ij}\psi_{ij}(x), \tag{7}$$

referred to as the *multiresolution decomposition* (Mallat, 1998). J_i denotes the set of j-indexes needed to describe f at scale i. The first term on the right-hand side of the equation is called the *trend*, i.e., it represents the mean or filtered behavior of a function f at a scale i_0. The second term represents the deviation from this trend or the *details* (from a geometrical viewpoint, it is the complement to the orthogonal projection of f onto $\mathcal{H}_{i_0}$); the summation is made over a different scale i.

Up to this point we have considered functions and wavelets in space $L_2(\mathbb{R})$, but in many applications, computations are done on a compact support, e.g., on $L_2([0,\ 1])$. For the projection operator [see Eq. (6)], this has no implication, whereas for the multiresolution decomposition, this involves finding wavelet functions that span space $L_2([0,\ 1])$. A practical way of doing this is to periodize the wavelets by introducing

$$\phi_{ij}^{\rm per}(x) = \sum_{k\in\mathbb{Z}} \phi_{ij}(x+k) \text{ and } \psi_{ij}^{\rm per}(x) = \sum_{k\in\mathbb{Z}} \psi_{ij}(x+k). \tag{8}$$

Most of the properties of the nonperiodic wavelets are kept by the periodized functions; moreover, although their definition involves an infinite sum of terms, a finite number is really needed in practice because of the finite width of the wavelet support. The price to pay for this fairly simple treatment is that spurious effects (wide oscillations) at the boundaries may be introduced because property (6) no longer holds. Another consequence is that $||f - \mathcal{P}_{i_0} f|| \leq C||f||O(2^{-i_0/2})$, instead of $||f^{(k)}||O(2^{-ki_0})$, regardless of the function smoothness. These disadvantages can be alleviated or removed by using other constructions, but at the price of more complex computations. Since the spurious effects affect only the boundaries and their extent can be controlled by altering the scaling level, here we will use the periodic construction given in Eq. (8).

A last point concerns the computation of the coefficients a_{ij} and b_{ij} defined by using inner products. The equivalent of the fast Fourier transform exists for wavelets; so in most practical situations, it is not necessary to compute integrals numerically, but algebraic computations can used. For a sufficiently high scaling level i, we start with $b_{ij} = \langle f,\ \psi_{ij}\rangle = 2^{-i/2} f(j/2^i) + O(2^{-i})$. If we can sample a function $f(x)$ at the dyadic points $j2^{-i}$ over $[0,\ 1]$, then we have an estimate of b_{ij}; using a *cascade* algorithm (Mallat, 1998), we can iteratively compute the coefficients b_{kj} and a_{kj} at coarser levels ($0 \leq k < i$) without computing any integral.

V. ESTIMATE OF THE ERROR

An error estimate is given by the mean square error $\mathbb{E}[||\tilde{f} - f||^2]$, where the average is made over $\mathcal{H}$. We assume that the function g is known within a certain small tolerance ζ satisfying $||\zeta|| \ll 1$, but $\zeta \notin L_2(\mathbb{R})$ (it does not make sense to state that the error is a

well-behaved function); so we have to solve the problem $K\tilde{f} = g + \zeta$. Here we find

$$\mathbb{E}[||\tilde{f} - f||^2] = \sum_{i=p}^{\infty} \langle g,\ u_i \rangle^2 + \sum_{i=0}^{p-1} \mathbb{E}[\langle \zeta,\ u_i \rangle^2].$$

If noise is modeled as a white noise of strength ε, then we can write $\zeta = \varepsilon W$, with $W = \sum_{i \in \mathbb{Z}} \nu_i \Psi_i$ and ν_i a random Gaussian variable of mean 0 and variance 1, since the base Ψ_i is orthonormal. In this case and using the fact that $\langle g,\ u_i \rangle = \langle f,\ \Psi_i \rangle$, the previous equation simplifies into

$$\mathbb{E}[||\tilde{f} - f||^2] = \sum_{i=p}^{\infty} \langle f,\ \Psi_i \rangle^2 + \varepsilon^2 \sum_{i=0}^{p-1} \langle \Psi_i,\ u_i \rangle^2. \tag{9}$$

The former term on the right-hand side of the equation represents the bias, i.e., the distance between the true solution and the approximation, whereas the latter term is the variance, i.e., the error resulting from noisy data.

As is typical for inverse problems (Bertero *et al.*, 1985), increasing the approximation order p leads to decreasing the bias, but also to increasing the variance. Thus, an optimal order p exists that optimizes the accuracy gained by increasing p and the variance growth induced by this increase. To go further in this direction, let us consider a function of $\mathcal{C}^k([a,\ b])$. Since the support of f is compact, the number of wavelet functions involved in its projection onto $\mathcal{H}_{i_0} = \text{span}(\phi_{i_0 j})_{j \in J_0}$ is finite and the length of J_0 is $p = 2^{i_0}(b-a)+2q-1 \propto (b-a)2^{i_0}$ (see Appendix IV). The bias in Eq. (9) is bounded at level i_0 by $C||f^{(k)}||2^{-ki_0}$, where C is a constant. The inner products $\langle \Psi_i,\ u_i \rangle$ are bounded by a constant D, whatever the scale level i_0, because of the scale-invariance of the operator. This implies that the variance is bounded by $pD^2\varepsilon^2$. The mean-square error is then bounded by

$$\mathbb{E}[||\tilde{f} - f||^2] \leq C||f^{(k)}||2^{-ki_0} + pD^2\varepsilon^2 \propto C||f^{(k)}||2^{-ki_0} + (b-a)2^{i_0}D^2\varepsilon^2.$$

The optimal scale level i_0 is the scale index that minimizes the right-hand side of the inequality, i.e., the integer closest to $((1+k)\ln 2)^{-1} \ln[kC||f^{(k)}||/D^2/(b-a)/\varepsilon^2]$, thus $2^{i_0} \approx \varepsilon^{-2/(k+1)}$. The optimal rate (within a multiplicative factor) of the mean-square error is then $\mathbb{E}[||\tilde{f} - f||^2] \leq \varepsilon^{2k/(k+1)}$ when $\varepsilon \to 0$. This shows that for irregular functions ($k = 1$), the mean-square error varies asymptotically as $\varepsilon^{2/3}$, whereas for smooth functions (large k values), the mean-square error decreases much quicker as ε^2. This optimal rate can be substantially increased by using shrinkage techniques (Cai, 1999; Donoho and Johnstone, 1995).

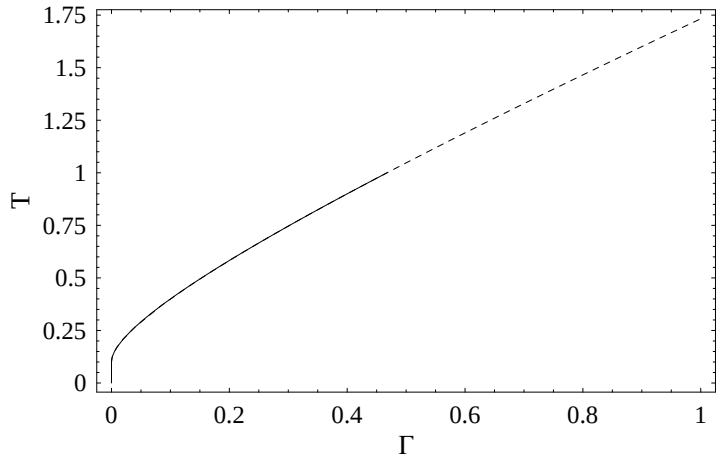

FIG. 2: Flow curve for the (dimensionless) Casson model. The solid line represents the numerical approximation obtained by using the WVD approach, whereas the dashed line represents the exact solution Eq. (10).

VI. NUMERICAL EXAMPLE

For the numerical test, let us consider the following Casson constitutive equation in a dimensionless form

$$\sqrt{T} = \sqrt{T_c} + \sqrt{\Gamma}, \tag{10}$$

where T_c is the dimensionless yield stress. For numerical applications, let us set $T_c = 0.1$. Computations are made over the interval $[0,\ 1]$. The corresponding dimensionless torque per unit height is

$$\Omega(T) = \begin{cases} 0 & \text{for } T < T_c, \\ T - T_c - \sqrt{TT_c} + 4T_c + T_c \ln(T/T_c) & \text{for } T_c \leq T < \beta T_c, \\ T_c - \beta T + 4\sqrt{TT_c}(\sqrt{\beta} - 1) - T_c \ln \beta & \text{for } T \geq \beta T_c. \end{cases} \tag{11}$$

First we consider the projection method. The scale level is set to $i_0 = 6$ such that the details (yield stress and nonlinear behavior) can be properly reproduced (the scaling level must satisfy $2^{-i_0} < (4q-2)^{-1}$). We use the Daubechies D8 wavelets ($q = 8$). To represent functions whose support lies within $[0,\ 1]$, $p = 78$ wavelets are needed; we use $(\Psi_k)_{1\leq k\leq 78} = \{\phi_{6,-14},\ \phi_{6,-13}, \cdots,\ \phi_{6,63}\}$. We denote $\tilde{\Gamma}$ the approximate dimensionless shear rate: $\tilde{\Gamma} = \mathcal{P}_{78}\Gamma$. The recovered shear rate is reported in Fig. 2. Note the perfect reconstruction of the shear rate over the entire interval. The mean-square error is approximately 1.21965×10^{-9}

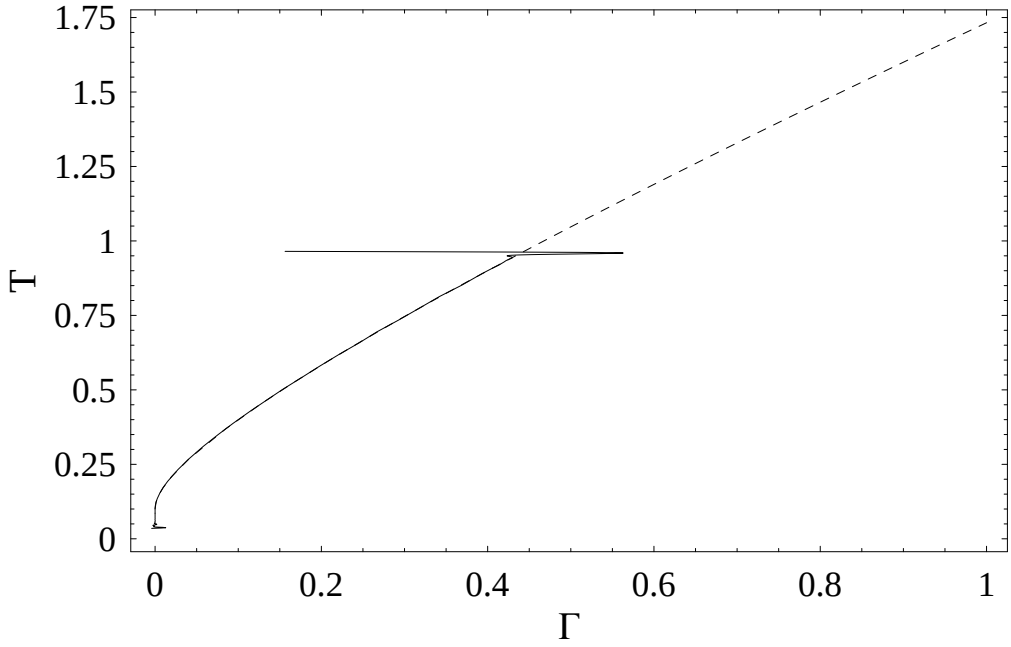

FIG. 3: Flow curve for the (dimensionless) Casson model. The solid line represents the numerical approximation obtained using a multiresolution decomposition and the WVD approach, whereas the dashed line represents the exact solution [Eq. (10)].

versus 1.21675×10^{-9} when $\tilde{\Gamma} = \mathcal{P}_{78}\Gamma$ is computed from the exact function.

Similar tests were carried out by replacing the exact dimensionless torque with an interpolating polynomial (piecewise polynomial of order 3) fitting $n_d = 100$ data computed from $\Omega(i\delta)$, for $i = 0, \cdots, n_d - 1$ and $\delta = (n_d - 1)^{-1}$. A very good agreement was found, with no difference to the naked eye with the exact flow curve (mean-square error 1.296932×10^{-9}).

Second we consider a multiresolution analysis. The wavelet series in Eq. **??**eq:multiresolution) was truncated for $i > 6$. As previously, we use the periodized Daubechies D8 wavelets. To represent functions whose support lies within $[0,\ 1]$, $p = 128$ wavelets are needed; we use $(\Psi_k)_{1 \le k \le 128} = \{\phi_{0,0},\ \psi_{0,0},\ \psi_{1,0},\ \psi_{1,1},\ \psi_{2,0}, \cdots\ \psi_{6,63}\}$. The recovered shear rate is reported in Fig. 3. Note the fairly good reconstruction of the shear rate over the entire interval except for the boundaries, where large fluctuations are observed (because we used periodic wavelets). The mean square error is approximately 4.0508×10^{-4} versus 1.30239×10^{-8} when $\tilde{\Gamma} = \mathcal{P}_{78}\Gamma$ is computed from the exact function.

Bertero, M., C. De Mol and E. Pike, "Linear inverse problems with discrete data. I: General formulation and singular system analysis," Inv. Prob. **1**, 301–330 (1985).

Cai, T., "Adaptative wavelet estimation: a block thresholding and oracle inequality approach," Ann. Stat. **27**, 898–924 (1999).

Coleman, B., H. Markowitz and W. Noll, *Viscometric flows of non-Newtonian fluids*, vol. 5 of *Springer Tracts in natural philosophy*, Springer-Verlag, Berlin (1966).

Donoho, D., "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," Appl. Comp. Harmonic Anal. **2**, 101–126 (1995).

Donoho, D. and I. Johnstone, "Wavelet shrinkage: asymptotia?" J. Royal. Stat. Soc. B **57**, 301–369 (1995).

Louis, A., P. Maass and A. Rieder, *Wavelets, Theory and Applications*, Pure & Applied Mathematics, John Wiley & Sons, Chichester (1997).

Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, 2nd edition edn. (1998).

Strang, G. and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge, Wellesley (1997).