Parameter-free and fast nonlinear piecewise filtering.

Application to experimental physics

**Barbara Pascal · Nelly Pustelnik ·
Patrice Abry · Jean-
Christophe Géminard ·
Valérie Vidal**

Abstract Numerous fields of nonlinear physics, very different in nature, produce signals and images, that share the common feature of being essentially constituted of piecewise homogeneous phases. Analyzing signals and images from corresponding experiments to construct relevant physical interpretations thus often requires detecting such phases and estimating accurately their characteristics (borders, feature differences, . . .). However, situations of physical relevance often comes with low to very low signal to noise ratio precluding the standard use of classical linear filtering for analysis and denoising and thus calling for the design of advanced nonlinear signal/image filtering techniques. Additionally, when dealing with experimental physics signals/images, a second limitation is the large amount of data that need to be analyzed to yield accurate and relevant conclusions requiring the design of fast algorithms. The present work proposes a unified signal/image nonlinear filtering procedure, with fast algorithms and a data-driven automated hyperparameter tuning, based on proximal algorithms and Stein unbiased estimator principles. The interest and potential of these tools are illustrated at work on low-confinement solid friction signals and porous media multiphase flows.

All Authors
Univ Lyon, ENS de Lyon, Univ Claude Bernard, CNRS, Laboratoire de Physique, Lyon, France
Tel.: +33 47272 8493
E-mail: firstname.lastname@ens-lyon.fr

1 Introduction

Signals or images collected from numerous experiments in physics can be, at least as a first order approximation, described as piecewise homogeneous (piecewise constant, piecewise linear,...). Detecting and estimating such piecewise homogeneous regions thus constitute a crucial goal to extract the physically relevant information conveyed in such data. This remains, however, often challenging, as signals or images are usually altered by superimosed noises, possibly with low signal-to-noise ratio, which may hinder the interpretation of the corresponding experiments. The joint need to denoise data while preserving edges and discontinuities pertaining phase changes or region boundaries often preclude the use of classical linear filtering and call for the use of advanced nonlinear signal and image processing techniques.

Solid friction provides us with a first representative example. Indeed, when two, nominally flat, solid surfaces in contact are forced to slide against one another, the shear force at the contact surface exhibits generally a characteristic tooth-shape signal (Fig. 1a): the force signal thus consists of ab alternation of slow linear rises, corresponding to the loading of elastic energy in the driving system while the surfaces in contact do not move with respect to each another, followed by sudden drops, corresponding to fast energy releases when surfaces slide [1]. When solids are strongly pressed one against the other, these two *phases* (rest and sliding at the contact surface) can easily be identified. However, high confinement pressures tend to damage surfaces, a major limitations in the study of the microscopic mechanisms at play. Therefore, probing effectively and accurately frictional material properties required that experiments are performed at low confinement pressure. This, however, induces that collected signals have low to very low signal-to-noise-ratios [2] thus requiring advanced non linear filtering signal processing techniques to detect and analyse the piecewise linear shape of data.

Multiphase flows in porous media constitute another rich example. Hydrodynamics in porous media, notably mass transfer, is of prominent practical relevance in industry, e.g., for catalytic process studies. Hydrodynamics and mass transfer studies of multiphase flows in porous media traditionally involve packed beds and are well characterized. However, recent experiments [3, 4] consider innovative materials such as open-cell solid foams which, due to high porosity and the resulting low pressure loss, are promising for industrial applications. In such experiments, a liquid and a gas are forced to flow simultaneously through the foam and the characterization of such multiphase spatiotemporal dynamics stems from image analysis (Fig. 1b). The challenge is here to identify liquid flows from gas bubbles. The rationale is that each phase can be associated to homogeneous textures in images and a crucial stake consists in identifying precisely gas bubble contours so as to measure their lengths. Liquid and gas are both transparent and the foam is itself introducing a scrambled background, yielding low-contrast and blurred images, thus requiring advanced nonlinear image processing techniques form texture segmentation and contour estimation.
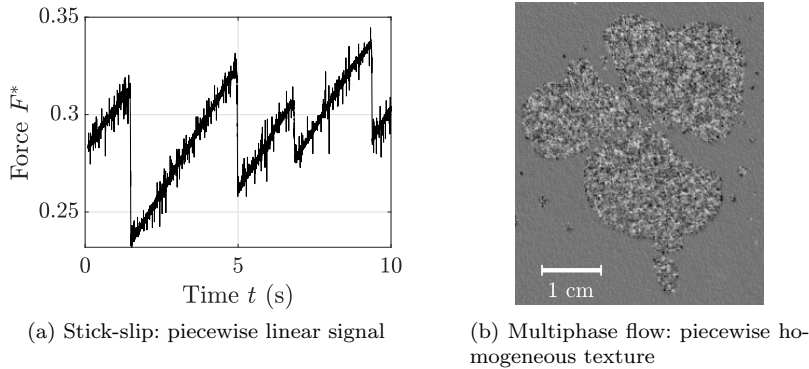
(a) Stick-slip: piecewise linear signal

(b) Multiphase flow: piecewise homogeneous texture

Fig. 1: **Experimental data in non linear physics.** (a): Normalized driving force $F^* = k\Delta x/(mg)$ [see Section 4.1] as function of time t in a solid friction experiment. (b): Direct image of a gas bubble in porous media multiphase flows (gas corresponds to the most scrambled region whereas liquid corresponds to smoother regions).

These two emblematic examples share in common that the key aspects of the physics to be understood are driven by piecewise homogeneous phases. On one hand, studying friction requires identifying the stick and the slip phases, each associated with a piecewise linear signal. On other hand, studying multiphase flows implies detecting fluid phases, each associated with a piecewise homogeneous texture. Piecewise-homogeneous signals or images are very common in numerous fields of nonlinear physics, very different in nature, such as time reversal of the magnetic field in turbulent dynamo [5], on-off intermittency in creeping granular matter [6], DNA detection during translocation through nuclear pores [7],...

The present work proposes a generic nonlinear signal/image filtering unified framework for the analysis of piecewise homogeneous (piecewise-constant, piecewise-linear) experimental datasets. The major challenges here are both to obtain fast algorithms so as to handle the large amount of data that need to be analyzed to yield accurate and relevant conclusions (e.g., in producing a phase diagram or in analyzing video frames of large size images) and to be able to perform an automated and data-driven tuning of hyperparameters, unavoidably entering any nonlinear filtering procedure, and whose arbitrary selection (by expert visual inspection) might have drastic impacts on achieved outcomes and thus on a posteriori drawn physical interpretations.

The unified signal/image nonlinear filtering framework proposed here is based on proximal schemes [8,9] to obtain fast algorithms, and on the Stein unbiased estimator framework to design an automated data-driven hyperparameter tuning procedure.

Section 2 is dedicated to the formulation of this framework as an inverse problem, and recalls state-of-the-art strategies with focus on piecewise constant or linear estimation both in signal or images. Section 3 details the pro-

posed algorithmic framework. Section 4 illustrates the performance on the two examples discussed above, solid friction and multiphase flows.

A documented toolbox (in MATLAB), for the implementation of this signal/image processing nonlinear filtering and data-driven hyperparameter tuning, is freely available at
`https://github.com/bpascal-fr/stein-piecewise-filtering`.

2 Non linear filtering formulated as inverse problems

2.1 Direct models

Let $S = \{\underline{n} = (n_1, n_2) : 1 \leq n_1 \leq N_1, 1 \leq n_2 \leq N_2\}$ denote a lattice, supporting $\overline{x} = (\overline{x}_{\underline{n}})_{\underline{n} \in S}$, the unknown signal/image of size $N = N_1 \times N_2$ ($N_1 = 1$ for univariate 1D signal analysis and $N_1 = K$ for multivariate 1D signal analysis with K components). Observation $z = (z_{m_1, m_2})_{1 \leq m_1 \leq M_1, 1 \leq m_2 \leq M_2}$ is of size $M = M_1 \times M_2$ consists of a degraded version of $\overline{x}$, which stems both from a linear degradation (e.g. filtering), denoted $A \in \mathbb{R}^{M \times N}$, and additive random noise, denoted $\mathcal{B}$.

Handling an inverse problem relies first on an accurate design of the observation (or direct) model, which here takes the following form:

$$z = \mathcal{B}(A\overline{x}).$$

In this work, S corresponds to an homogeneous neighborhood system. For instance, when considering a 1D signal, a site $n_2 \in \{2, \ldots, N_1 - 1\}$ has two nearest neighbors $\mathcal{N}_{n_2} = \{n_2 - 1, n_2 + 1\}$. In a general regular rectangular lattice S and for a 4-neighborhood system, every interior point has four neighbors that yields to $\mathcal{N}_{\underline{n}} = \{(n_1 - 1, n_2), (n_1 + 1, n_2), (n_1, n_2 - 1), (n_1, n_2 + 1)\}$. The pair $(S, \mathbb{E})$ constitutes a graph where S contains the nodes and $\mathbb{E}$ determines the links between the nodes according to the neighboring relationship.

We detail this direct model on the two nonlinear physics problems described in Introduction (low confinement solid friction and porous media multiphase flows) and illustrated in Figure 1. For solid friction, the challenging question consists in denoising obserbation z (Figure 1(a)), with $A = \text{Id}$ and the presence of additive impulsive noise. For multiphase flows, the challenging issues pertains to segmentation of textures such as the one in Figure 1(b). In such a case, information $\overline{x}$ refers to piecewise constant scale-free texture features, and observation z is obtained from a nonlinear multiscale transform (cf. Section 4.2). Noise is assumed additive and Gaussian, with spatial and multiscale correlations.

2.2 State-of-the-art

Solving an inverse problem consists in providing an estimator $\widehat{x}$, as close as can be from information $\overline{x}$. This has been largely addressed in the literature and

we propose first a brief overview of the main inverse problem solving streams (see also [10,11]), before focusing, second, on the specific assumptions on the model required to design parameter-free and fast nonlinear piecewise filtering.

Bayesian arguments and most standard models – On the first hand, Markov Random Fields (MRF) have been introduced in visual labelling to establish probabilistic distributions of interacting labels, aiming to analyze dependencies of a physical phenomena [12]. In such a formalism $\overline{x}$ and z are considered as realizations of random vectors $\overline{X}$ and Z defined on the set S. X is said to be a MRF on S with respect to a neighborhood system $\mathbb{E}$ if and only if positivity (i.e. $P(X = x) > 0$) and Markovianity $P(x_{\underline{n}}|x_{\{S\}-\underline{n}}) = P(x_{\underline{n}}|x_{\mathcal{N}_{\underline{n}}})$, which models the local characteristics of X, are satisfied. Other properties such as homogeneity and isotropy can be depicted.

The link between the MRF, characterized by its local properties, and another standard random field, the Gibbs random field, characterized by global properties, has been provided by Hammersley and Clifford [13, Theorem 1]. We recall that a Gibbs distribution relative to the graph $\{S, \mathbb{E}\}$ is a probability measure and it has the following representation :

$$P(\omega) = \frac{1}{C} e^{-U(\omega)/T} \tag{1}$$

where C is the normalizing constant called the partition function such that $C = \sum_{\omega} e^{-U(\omega)/T}$ and T stands for temperature, which controls the sharpness of the distribution. High temperature leads to all configurations equally distributed, while a temperature close to 0 concentrates the distribution around the global energy minima. $U(\omega)$ denotes the energy function and $P(\omega)$ measures the probability of the occurence of a specific configuration ω. The more probable configurations are those with the lower energies. The terminology comes from statistical physics where such measures are equilibrium states for physical systems (e.g. ferromagnets). $U(\omega)$ can be formulated with contributions from external fields (i.e. x_{n_1,n_2}) and pair interactions (e.g. $x_{n_1,n_2} x_{n_1+1,n_2}$). For instance, the Ising model reads

$$U(x) = -\alpha \sum x_{n_1,n_2} - \beta \left(\sum x_{n_1,n_2} x_{n_1+1,n_2} + \sum x_{n_1,n_2} x_{n_1,n_2+1} \right) \tag{2}$$

considering $\omega = x$ and for some parameters $\alpha \in \mathbb{R}$ and $\beta > 0$, which measure, the external magnetic moment and bonding strengths.

Geman and Geman [14] handle the maximization of the conditional probability distribution of $(x, e) \in \{S, \mathbb{E}\}$ given the data z (i.e. find the mode of the posterior distribution), which is known as the maximum a posteriori estimation or penalized maximum likelihood. In [15,14,16], the authors prove that the posterior is a Gibbs distribution over $\{S, \mathbb{E}\}$ with energy function

$$U(x, e) = \frac{1}{2\sigma^2} \|Ax - z\|^2 + \beta \sum_{\underline{n},\underline{n}' \in \mathcal{N}_{\underline{n}}} \varphi(x_{\underline{n}} - x_{\underline{n}'})(1 - e_{\underline{n},\underline{n}'}) + \alpha \psi(e) \tag{3}$$

so that $\omega = (x, e)$ when $\mathcal{B}$ designates an additive zero-mean Gaussian noise with a variance σ^2. $e \in \mathbb{E}$ denotes the coded line states and $\varphi(\eta) = -1$ if $\eta = 0$ and 1 if $\eta \neq 0$. The first term acts as a data fidelity term and forces the approximation x to be close to z, the second term allows small variations of x except at the locations where $e_{\underline{n},\underline{n}'} = 1$, and ψ is constructed to organize the line process. Finally, $\alpha > 0$ and $\beta > 0$ denote regularization parameters controlling the smoothness of the solution and the length of the interfaces. This model can be interpreted as a coupled MRF dealing jointly with image restoration and edges detection: one MRF for the pixel values and one for the edges values that are described respectively in the image lattice or in its dual lattice. This model has strong link with the continuous Mumford-Shah setting [17].

For specific choices of φ and ψ [16, 18], an alternative equivalent formulation is the Blake-Zisserman model formulated as

$$U(x) = \frac{1}{2\sigma^2}\|Ax - z\|^2 + \lambda \sum_{\underline{n},\underline{n}' \in \mathcal{N}_{\underline{n}}} \min\left((x_{\underline{n}} - x_{\underline{n}'})^2, \eta\right) \tag{4}$$

where $\lambda, \eta > 0$, leading to the so-called truncated ℓ_2 and whose interest is to favor piecewise smooth solution. Another model very close is the Potts model that can be interpreted as a ℓ_0-penalization over $x_{\underline{n}} - x_{\underline{n}'}$ that is designed to provide piecewise-constant estimates [19]. For numerical reasons detailed below, the most standard convex relation is the anisotropic total-variation penalization which reads [20, 21]:

$$U(x) = \frac{1}{2\sigma^2}\|Ax - z\|^2 + \lambda \sum_{\underline{n},\underline{n}' \in \mathcal{N}_{\underline{n}}} |x_{\underline{n}} - x_{\underline{n}'}|. \tag{5}$$

Solving inverse problems – Once an energy (or functional) has been designed to fit the considered problem, numerical strategies have to be designed to implement both the hyperparameter selection and the computation of the minimizing solution x, also corresponding to the most probable ω or moments of P.

On one hand, Markov Chain Monte Carlo algorithms address simulations from a probability distribution P. The function P can be written in a closed-form expression but the objective is generally to access the moments of P, which are not computable analytically. The two main techniques used in MCMC are Metropolis-Hastings, which relies on accept/reject mechanism and Gibbs sampler, which simplifies the high dimensional problem by successively simulating from different smaller dimensional components. The main limitation of these techniques is to be computationally intensive for solving large size inverse problems (see a contrario [22, 23]). We should also refer to some specific configurations where a closed form expression is available such as for the Ising model in 1D and 2D but which is not adapted for general inverse problem solving considered in this work.

When one wants to estimate jointly the maximum of a posteriori and its hyperparameters, Bayesian hierarchical inference frameworks are particularly

adapted and received considerable interest for addressing change-point detection or piecewise denoising problems [24–26] or texture segmentation [23]. However, to the best of our knowledge, for the proposed unified 1D-2D framework considered in this work, such a general efficient strategy has not yet been designed.

On other hand, during the last twenty years, important research efforts have been dedicated to convex but non-smooth energy generally formulated as a sum of two or three terms: a data-fidelity term, a penalization and a constraint [27,9,28]. This framework is thus especially adapted when dealing with an energy such as (5). These algorithmic strategies are particularly efficient when dealing with A full-rank which is rarely the case in standard restoration/reconstruction problems but which is more encountered in experimental physics processing when dealing either with denoising i.e. $A = \mathrm{Id}$ such as in friction experiments or when dealing with texture reconstruction especially adapted to study multiphase flow dynamics as it will be described later. However, when one handles such optimization strategy to find the minimizer of the energy U, the selection of the automated regularization parameter(s) is not addressed. For automated selection, one could consider either an empirical rule that consists in setting $\lambda \sim N^{1/2}\sigma/4$, with N the signal size and σ the noise standard deviation, estimated e.g., from the median value of the absolute value of the wavelet coefficients [29], or an hybrid Bayesian hierarchical inference framework combined with ℓ_0-minimization startegy [30] in the specific case of piecewise-constant denoising, or the recourse to Stein Unbiased Risk Estimator (SURE) which provides an unbiased estimator of the mean square error [31–33]. Our contribution focuses on such a strategy, its implementability, and its applicability on real physics experiments.

3 Proximal operator based nonlinear filtering: fast algorithms and automated data-driven hyperparameter tuning

3.1 Nonlinear filtering formulation

In this work, we consider an estimator $\widehat{x}(z; \Lambda)$ of the quantity of interest $\bar{x}$, from a corrupted observation z, parametrized by Λ. The estimate is obtained from the minimization of an energy, inspired from (5), and defined as:

$$\widehat{x}(z; \Lambda) \in \operatorname*{Argmin}_{x \in \mathbb{R}^N} \|Ax - z\|_2^2 + \|D_\Lambda x\|, \tag{6}$$

where the matrix D_Λ models a weighted discrete differentiation operator, so that the penalization $\|D_\Lambda x\|$ enforces some regularity of the estimate $\widehat{x}(z; \Lambda)$. Specific instances of (6) are provided :

– To favor joint piecewise *constancy* of K multivariate signals, the regularization parameters are stored in a vector $\Lambda = (\lambda_1, \dots, \lambda_K) \in \left(\mathbb{R}_+^*\right)^K$, and the operator D_Λ is a first order differentiation operator, acting componentwise, also called discrete gradient, writing for $k \in \{1, \dots, K\}$, $n_2 \in$

$\{1, \ldots, N_2 - 1\}$,

$$(D_\Lambda x)_{k,n_2} = \lambda_k \left(x_{k,n_2+1} - x_{k,n_2} \right) \tag{7}$$

and where $\|\cdot\|$ is the mixed $\ell_{1,2}$-norm

$$\|D_\Lambda x\|_{1,2} = \sum_{n_2=1}^{N_2-1} \sqrt{\sum_{k=1}^{K} (D_\Lambda x)_{k,n_2}^2}. \tag{8}$$

– Enforcing joint piecewise *linearity* requires a second order differentiation operator, named discrete Laplacian, defined for $k \in \{1, \ldots, K\}$ and $n_1 \in \{2, \ldots, N_1 - 1\}$,

$$(D_\Lambda x)_{n_1,k} = \lambda_k \left(x_{n_1+1,k} - 2x_{n_1,k} + x_{n_1-1,k} \right) \tag{9}$$

composed with the $\ell_{1,2}$-norm defined in (8).

– Image segmentation is performed imposing piecewise constancy prior, using a two dimensional difference operator. For an image $x \in \mathbb{R}^{N_1 \times N_2}$ the horizontal and vertical gradients are computed for each pixel with $1 \leq n_1 \leq N_1 - 1$ and $1 \leq n_2 \leq N_2 - 1$

$$(D_\Lambda x)_{n_1,n_2} = \lambda \begin{pmatrix} (D_1 x)_{n_1,n_2} \\ (D_2 x)_{n_1,n_2} \end{pmatrix} = \lambda \begin{pmatrix} x_{n_1,n_2+1} - x_{n_1,n_2} \\ x_{n_1+1,n_2} - x_{n_1,n_2} \end{pmatrix} \tag{10}$$

and coupled into an $\ell_{1,2}$-norm

$$\|D_\Lambda x\| = \lambda \sum_{n_1=1}^{N_1-1} \sum_{n_2=1}^{N_2-1} \sqrt{(D_1 x)_{n_1,n_2}^2 + (D_2 x)_{n_1,n_2}^2} := \lambda \mathrm{TV}(x). \tag{11}$$

The above penalization is known as the isotropic Total Variation [20].

The estimate $\widehat{x}(z; \Lambda)$ is the result of a trade-off between the fidelity to the observation model and some regularity priors and the balance is tuned by the hyperparameter Λ. Hence, our purpose is twofold. First, solving the minimization Problem (6), that is, for fixed hyperparameter Λ, given an observation z, compute $\widehat{x}(z; \Lambda)$ the minimizer of (6). Second, finding the *best* hyperparameter $\Lambda^\dagger$ minimizing the quadratic error $\mathbb{E}\{\|\widehat{x}(z; \Lambda) - \overline{x}\|^2\}$, i.e.

Problem 1 Find

$$\Lambda^\dagger = \underset{\Lambda \in \left(\mathbb{R}_+^* \right)^K}{\arg\min} \ \mathbb{E}\{\|\widehat{x}(z; \Lambda) - \overline{x}\|^2\} \tag{12}$$

where $\widehat{x}(z; \Lambda)$ is defined by (6) and the expectation is taken over all realizations of the noise $\mathcal{B}$ corrupting the observation $z = \mathcal{B}(A\overline{x})$.

In the next sections, we specify the assumptions over A and $\mathcal{B}$ allowing us to derive a fast algorithmic scheme to estimate $\Lambda^\dagger$.

3.2 Convex non-smooth minimization

The objective function appearing in Problem (6) is convex, since the composition of a linear operator and a norm is convex. Yet, because of the presence of the norm $\|\cdot\|$, it is non-smooth. Consequently, the minimization of (6) requires proximal algorithms [9,27,34], which in general suffer from low convergence rate. Nevertheless, provided that the operator A is injective, it is possible to design accelerated primal-dual schemes [21] and to obtain linear convergence rate toward the minimizer of (6). Such algorithms relies on proximity operators [34], whose definition is recalled in Definition 1. Further, disposing closed-form expressions of the proximity operators of the data fidelity term and the penalization function is a key element to design fast implementations of primal-dual algorithms.

Definition 1 For a convex lower semi-continuous function $\varphi : \mathbb{R}^N \to \mathbb{R} \cup \{+\infty\}$ and $\tau > 0$, the proximity operator is defined as

$$(\forall x \in \mathbb{R}^N) \qquad \mathrm{prox}_{\tau\varphi}(x) = \arg\min_{\widetilde{x}} \frac{1}{2}\|\widetilde{x} - x\|^2 + \tau\varphi(\widetilde{x}) \qquad (13)$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^N$.

Few examples of well-established closed-form expression for proximity operator of interest in this work are recalled.

Example 1 The proximity operator of the data fidelity term $\|A \cdot - z\|_2^2$ as a closed form expression that is, for every $\tau > 0$,

$$(\forall x \in \mathbb{R}^N) \qquad \mathrm{prox}_{\tau\|A\cdot - z\|^2}(x) = \left(\mathrm{Id} + 2\tau A^\top A\right)^{-1}\left(x + 2\tau A^\top z\right). \qquad (14)$$

Example 2 The proximity operator of the multivariate 1D $\ell_{1,2}$-norm defined in (8) is, for every $y \in \mathbb{R}^{K \times N_2}$,

$$\left(\mathrm{prox}_{\tau\|\cdot\|1,2}(y)\right)_{k,n_2} = \begin{cases} \left(1 - \frac{\tau}{\|y_{\cdot,n_2}\|_2}\right) y_{k,n_2} & \text{if } \|y_{\cdot,n_2}\|_2 > \tau, \\ 0 & \text{otherwise,} \end{cases} \qquad (15)$$

where $\|y_{\cdot,n_2}\|_2 := \sqrt{\sum_{k=1}^{K} y_{k,n_2}^2}$.

Example 3 The proximity operator of the 2D $\ell_{1,2}$-norm of Equation (11), for $y = \left(y^{(H)}, y^{(V)}\right) \in \mathbb{R}^{2 \times N_1 \times N_2}$,

$$\left(\mathrm{prox}_{\tau\|\cdot\|1,2}(y)\right)_{n_1,n_2} = \begin{cases} \left(1 - \frac{\tau}{\|y_{n_1,n_2}\|_2}\right) y_{n_1,n_2} & \text{if } \|y_{n_1,n_2}\|_2 > \tau, \\ 0 & \text{otherwise,} \end{cases} \qquad (16)$$

where $\|y_{n_1,n_2}\|_2 := \sqrt{(y_{n_1,n_2}^{(H)})^2 + (y_{n_1,n_2}^{(V)})^2}$.

Theorem 1 *Assuming that the deformation operator A is injective and denoting by $\mu > 0$ the smallest eigenvalue of $2A^\top A$, the sequence $\left(x^{[t]}\right)_{t\in\mathbb{N}}$ defined in Algorithm 1 converges toward the solution $\widehat{x}(z; \Lambda)$ of*

$$\underset{x\in\mathbb{R}^N}{minimize} \|Ax - z\|_2^2 + \|D_\Lambda x\|. \tag{17}$$

Further, it has been proven in [21] that, for any $\epsilon > 0$, there exists t_0 such that

$$(\forall t \geq t_0) \quad \left\|\widehat{x}(z; \Lambda) - x^{[t]}\right\| \leq \frac{1+\epsilon}{t^2}\left(\frac{\|\widehat{x}(z; \Lambda) - x^{[0]}\|^2}{\mu^2\tau_0^2} + \frac{\|A\|^2\|\widehat{y}(z; \Lambda) - y^{[0]}\|^2}{\mu^2}\right) \tag{18}$$

where $\widehat{y}(z; \Lambda)$ denotes the solution of the dual problem of Problem (17). Hence, the convergence rate of the iterates $\left(x^{[t]}\right)_{t\in\mathbb{N}}$ scales like $\mathcal{O}(1/t^2)$.

Proof This theorem is a direct application of [21, Theorem 2], stated and demonstrated for the minimization of objective functionals defined as the sum of a μ-strongly convex data fidelity and convex, proper, lower semi-continuous penalization, which is the case here.

Indeed, thanks to the assumption that A is full-rank, the considered data fidelity term $\|Ax - z\|_2^2$ is μ-strongly convex with modulus $\mu = 2\min\operatorname{Sp}(A^\top A) > 0$. Further, the penalization being the composition of a linear operator and a norm is satisfies the aforementioned conditions.

Then, the primal-dual updates of Algorithm 1 corresponds to the customization of the Algorithm 2 of [21] to the problem of finding $\widehat{x}(z, \Lambda)$ solution of (6), hence corresponding to $G(x) = \|Ax - z\|^2$, linear operator $K = D_\Lambda$ and $F = \|\cdot\|_{1,2}$.

We have to note that, because of the operation $\left(\mathrm{Id} + 2\tau A^\top A\right)^{-1}$, the proximity operator of the data-term might be uneasy to evaluate. However, for numerous configuration, this expression as a closed form expression. First when $A = \mathrm{Id}$. Second, when A is diagonalizable in a specific basis such as Fourier for circulant matrices (leading to $\mathcal{O}(MN)$ operations). Another specific choice of A will be discussed in Section 4.2.

3.3 Stein Unbiased Risk Estimate

Once an efficient algorithmic strategy has been identified to estimate $\widehat{x}(z; \Lambda)$, the second major difficulty raised by Problem 1 is that, in practice, one does not have access to the true signal/image $\bar{x}$ and hence cannot compute $\mathbb{E}\left\{\|\widehat{x}(z; \Lambda) - \bar{x}\|^2\right\}$. To handle this limitation, Stein proposed an Unbiased Risk Estimator [35], denoted $\mathrm{SURE}(\Lambda)$, providing an usable approximation of the quadratic risk in the case of i.i.d. Gaussian noise. This estimator was then widely extended to more general noise models [36,37].

Theorem 2 (Stein Unbiased Risk Estimate) *We denote $\widehat{x}(z; \Lambda)$ the parametric estimator defined in (17) of the ground truth $\bar{x}$ from observation $z = \mathcal{B}(A\bar{x})$ corrupted by a full-rank deformation operator A and additive (possibly correlated) Gaussian noise $\mathcal{B}$, with covariance matrix $\mathcal{S}$. The Stein Unbiased Risk Estimate, defined as*

$$\mathrm{SURE}(\Lambda) := \|\Phi\left(A\widehat{x}(z;\Lambda) - z\right)\|^2 + 2\mathrm{Tr}\left(\mathcal{S}\Phi^\top \frac{\partial \widehat{x}(z;\Lambda)}{\partial z}\right) - \mathrm{Tr}(\Phi\mathcal{S}\Phi^\top), \quad (19)$$

satisfies the following unbiasedness property

$$\mathbb{E}\left\{\mathrm{SURE}(\Lambda)\right\} = \mathbb{E}\{\|\widehat{x}(z;\Lambda) - \overline{x}\|^2\}. \tag{20}$$

where $\Phi := \left(A^\top A\right)^{-1} A^\top$ and $\partial\widehat{x}(z;\Lambda)/\partial z$ denotes the Jacobian of $\widehat{x}(z;\Lambda)$ w.r.t. observation z.

Proof A complete and detailed proof was proposed in [37].

Definition 2 (Degrees of freedom) The second term in the definition of SURE(Λ), in Equation (19),

$$2\mathrm{Tr}\left(\mathcal{S}\Phi^\top \frac{\partial \widehat{x}(z;\Lambda)}{\partial z}\right) \tag{21}$$

is called the *degrees of freedom*.

Since $\Phi \in \mathbb{R}^{M \times N}$ and $\partial\widehat{x}(z;\Lambda)/\partial z \in \mathbb{R}^{N \times M}$, are large size matrices computing the trace of $\Phi^\top \partial\widehat{x}(z;\Lambda)/\partial z$ is very expansive and hence the evaluation of the degrees of freedom requires additional tools. This difficulty is overcome using, a Monte Carlo (MC) strategy, only requiring the evaluation of the Jacobian on a random vector $\delta \in \mathbb{R}^M$. Hence, it is only necessary to store a vector of size N, instead of manipulating a matrix of size $N \times M$, which decreases drastically the computational and memory costs. Moreover, when $\widehat{x}(z;\Lambda)$ is obtained from a minimization scheme, there is often no closed-form expression of the Jacobian, hence we will approximate $\partial\widehat{x}(z;\Lambda)/\partial z[\delta]$ using Finite Difference (FD) approximation of the Jacobian. Altogether, Monte Carlo and Finite Difference strategies lead to the following FDMC Stein Unbiased Risk Estimate.

Theorem 3 (Finite Difference Monte Carlo SURE) *Let $\widehat{x}(z; \Lambda)$ denote a parametric estimator of ground truth $\bar{x}$ from observation $z = \mathcal{B}(A\bar{x})$ corrupted by a full-rank deformation operator A and additive (possibly correlated) Gaussian noise $\mathcal{B}$, with covariance matrix $\mathcal{S}$ and $\Phi = \left(A^\top A\right)^{-1} A^\top$. The FDMC Stein Unbiased Risk Estimate is defined as*

$$\mathrm{SURE}_{\varepsilon,\delta}(\Lambda) := \|\Phi\left(A\widehat{x}(z;\Lambda) - z\right)\|^2$$
$$+ \frac{2}{\varepsilon}\left\langle \Phi^\top\left(\widehat{x}(z + \varepsilon\delta;\Lambda) - \widehat{x}(z;\Lambda)\right), \mathcal{S}\delta\right\rangle - \mathrm{Tr}(\Phi\mathcal{S}\Phi^\top). \tag{22}$$

Provided that $\widehat{x}(z;\Lambda)$ is uniformly Lipschitz w.r.t. observation z and integrable against Gaussian density, $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$ is an asymptotically unbiased estimator of the quadratic risk, i.e.

$$\lim_{\varepsilon\to 0}\mathbb{E}\left\{\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)\right\} = \mathbb{E}\{\|\widehat{x}(z;\Lambda) - \overline{x}\|^2\}, \qquad (23)$$

where the expectation in the left hand side is taken over both the noise $\mathcal{B}$ and the Monte Carlo vector $\delta \sim \mathcal{N}(0,\mathrm{Id})$.

Proof The proof directly follows from Theorem 2 in [37].

Thus, Problem 1 is replaced by

Problem 2 Find $\widehat{\Lambda}^\dagger = \arg\min_{\Lambda\in\left(\mathbb{R}_+^*\right)^\kappa}\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$.

3.4 Automated data-driven hyperparameter tuning

In order to solve Problem 2, two strategies can be considered. First a grid search, computing $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$ over a large range of hyperparameter values, corresponding to the discrete set $\boldsymbol{\Lambda} = (\Lambda_i)_{i=1}^I$ and selecting *a posteriori* the hyperparameters of the grid for which $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda_i)$ is minimal, denoted $\widehat{\Lambda}_{\mathrm{G}}$, as described in Algorithm 2. The major drawback of this approach is its computational cost, increasing algebraically with the dimension of the hyperparameters Λ.

Algorithm 1 Accelerated primal dual scheme for minimization of (6).

Require: Set $\varepsilon > 0$, $\delta \in \mathbb{R}^M$, $\tau_0 > 0$, $\sigma_0 > 0$, such that $\tau_0\sigma_0\|D_\Lambda\|^2 < 1$. .
  for $\widetilde{z} = \{z, z+\varepsilon\delta\}$ **do**
    Choose $\widetilde{x}^{[0]} \in \mathbb{R}^N$, $x^{[0]} \in \mathbb{R}^N$, $y^{[0]} = D_\Lambda x^{[0]}$
    $\partial_\Lambda\widetilde{x}^{[0]} \leftarrow 0_N$
    $\partial_\Lambda x^{[0]} \leftarrow 0_N$
    $\partial_\Lambda y^{[0]} \leftarrow D_\Lambda\partial_\Lambda x^{[0]}$
    for $t = 0$ **to** $T_{\max} - 1$ **do**
      {Primal-dual updates}
      $y^{[t+1]} = \mathrm{prox}_{\sigma_t(\|\cdot\|)^*}\left(y^{[t]} + \sigma_t D_\Lambda\widetilde{x}^{[t]}\right)$
      $x^{[t+1]} = \mathrm{prox}_{\tau_t\|A\cdot-\widetilde{z}\|^2}\left(x^{[t]} - \tau_t D_\Lambda y^{[t+1]}\right)$
      $\vartheta_t = \sqrt{1+2\mu\tau_t}$, $\tau_{t+1} = \tau_t/\vartheta_t$ and $\sigma_{t+1} = \vartheta_t\sigma_t$
      $\widetilde{x}^{[t+1]} = x^{[t+1]} + \vartheta_t\left(x^{[t+1]} - x^{[t]}\right)$
    end for
    $\widehat{x}(\widetilde{z};\Lambda) \leftarrow x^{[T_{\max}]}$
  end for
  Compute $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$ injecting $\widehat{x}(z;\Lambda)$ and $\widehat{x}(z+\varepsilon\delta;\Lambda)$ in Formula (22)
  return $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$

In order to provide faster implementations, we consider automated selection of hyperparameters. To that aim, the number of hyperparameters is assumed to be $K = \mathcal{O}(1)$ and hence quasi-Newton algorithms are appropriate

---

Algorithm 2 Grid search for SURE minimization.

Require: Grid $\Lambda = (\Lambda_i)_{i=1}^I$, $\varepsilon > 0$, $\delta \sim \mathcal{N}(0, \mathrm{Id}) \in \mathbb{R}^M$,
  for $i = 1$ **to** I **do**
    ERROR$(i) \leftarrow \mathrm{SURE}_{\varepsilon,\delta}(\Lambda_i)$, computed from Algorithm 1
  end for
  $\widehat{i}_{\mathrm{G}} \leftarrow \underset{1 \leq i \leq I}{\arg\max} \, \mathrm{ERROR}(i)$
  return $\widehat{\Lambda}_{\mathrm{G}} = \Lambda_{\widehat{i}_{\mathrm{G}}}$

---

since they can handle very efficiently minimization in low dimension. It requires to compute the gradient of $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$ w.r.t. the hyperparameter Λ. For this purpose, it was proposed a Stein Unbiased GrAdient Risk estimate, denoted $\mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda)$ [33,37], which, under the conditions of Theorem 3, writes

$$
\mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda) := 2 \left(\Phi A \frac{\partial \widehat{x}(z; \Lambda)}{\partial \Lambda} \right)^\top (\Phi (A \widehat{x}(z; \Lambda) - z))
$$
$$
+ \frac{2}{\varepsilon} \left\langle \Phi^\top \left(\frac{\partial \widehat{x}(z + \varepsilon\delta; \Lambda)}{\partial \Lambda} - \frac{\partial \widehat{x}(z; \Lambda)}{\partial \Lambda} \right), \mathcal{S}\delta \right\rangle. \qquad (24)
$$

A sketch of quasi-Newton descent [38], particularized to Problem 2, is detailed in Algorithm 4. It generates a sequence $\left(\Lambda^{[j]} \right)_{j \in \mathbb{N}}$ converging toward a minimizer of $\mathrm{SURE}_{\varepsilon,\boldsymbol{\delta}}(\Lambda)$, denoted $\widehat{\Lambda}_{\mathrm{BFGS}}$.

This algorithm relies on a gradient descent step involving a descent direction $d^{[j]}$ obtained from the product of BFGS approximated inverse Hessian matrix $H^{[j]}$ and the gradient $\mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda)$ obtained from Algorithm 3. The descent step size $\alpha^{[j]}$ is obtained from a line search which stops when Wolfe conditions are fulfilled [38,39]. Finally, the approximated inverse Hessian matrix $H^{[j]}$ is updated according to a BFGS strategy.

Remark 1 The line search is the most time consuming. Indeed, the routines SURE and SUGAR are called for several hyperparameters of the form $\Lambda^{[j]} + \alpha d^{[j]}$, each call requiring to run differentiated primal-dual scheme twice.

Algorithm 3 Accelerated primal dual scheme for minimization of (6) with iterative forward differentiation.

Require: Set $\varepsilon > 0$, $\delta \in \mathbb{R}^M$, $\tau_0 > 0$, $\sigma_0 > 0$, such that $\tau_0 \sigma_0 \|D_\Lambda\|^2 < 1$

   for $\widetilde{z} = \{z, z + \varepsilon\delta\}$ **do**

      Choose $\widetilde{x}^{[0]} \in \mathbb{R}^N$, $x^{[0]} \in \mathbb{R}^N$, $y^{[0]} = D_\Lambda x^{[0]}$

      $\partial_\Lambda \widetilde{x}^{[0]} \leftarrow 0_N$

      $\partial_\Lambda x^{[0]} \leftarrow 0_N$

      $\partial_\Lambda y^{[0]} \leftarrow D_\Lambda \partial_\Lambda x^{[0]}$

      for $t = 0$ **to** $T_{\max} - 1$ **do**

         {Primal-dual updates}

         $y^{[t+1]} = \mathrm{prox}_{\sigma_t(\|\cdot\|)^*}\left(y^{[t]} + \sigma_t D_\Lambda \widetilde{x}^{[t]}\right)$

         $x^{[t+1]} = \mathrm{prox}_{\tau_t \|A\cdot -\widetilde{z}\|^2}\left(x^{[t]} - \tau_t D_\Lambda y^{[t+1]}\right)$

         $\vartheta_t = \sqrt{1 + 2\mu\tau_t}$, $\tau_{t+1} = \tau_t/\vartheta_t$ and $\sigma_{t+1} = \vartheta_t \sigma_t$

         $\widetilde{x}^{[t+1]} = x^{[t+1]} + \vartheta_t\left(x^{[t+1]} - x^{[t]}\right)$

         {Forward iterative differentiation}

         $\partial_\Lambda y^{[t+1]} = \partial_y \mathrm{prox}_{\sigma_t(\|\cdot\|)^*}\left(\partial_\Lambda y^{[t]} + \sigma_t D_\Lambda \partial_\Lambda \widetilde{x}^{[t]} + \sigma_t\left(\partial_\Lambda D_\Lambda\right)\widetilde{x}^{[t]}\right)$

         $\partial_\Lambda x^{[t+1]} = \partial_x \mathrm{prox}_{\tau_t \|A\cdot -\widetilde{z}\|}\left(\partial_\Lambda x^{[t]} - \tau_t D_\Lambda \partial_\Lambda y^{[t+1]} - \tau_t\left(\partial\Lambda D_\Lambda\right)y^{[t+1]}\right)$

         $\partial_\Lambda \widetilde{x}^{[t+1]} = \partial_\Lambda x^{[t+1]} + \vartheta_t\left(\partial_\Lambda x^{[t+1]} - \partial_\Lambda x^{[t]}\right)$

      end for

      $\widehat{x}(\widetilde{z}; \Lambda) \leftarrow x^{[T_{\max}]}$

      $\partial_\Lambda \widehat{x}(\widetilde{z}; \Lambda) \leftarrow \partial_\Lambda x^{[T_{\max}]}$

   end for

   Compute $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$ injecting $\widehat{x}(z; \Lambda)$ and $\widehat{x}(z + \varepsilon\delta; \Lambda)$ in Formula (22)

   Compute $\mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda)$ injecting $\widehat{x}(z; \Lambda)$, $\widehat{x}(z + \varepsilon\delta; \Lambda)$, $\partial_\Lambda \widehat{x}(z; \Lambda)$ and $\partial_\Lambda \widehat{x}(z + \varepsilon\delta; \Lambda)$ in Formula (24)

   return $\mathrm{SURE}_{\varepsilon,\delta}(\Lambda)$ and $\mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda)$

Algorithm 4 Automated selection of hyperparameters minimizing quadratic risk.

Require: $\varepsilon > 0$, $\delta \sim \mathcal{N}(0, \mathrm{Id}) \in \mathbb{R}^M$

Initialization: $\Lambda^{[0]} \in (\mathbb{R}_+)^K$, $H^{[0]} \in \mathbb{R}^{K \times K}$

   $\mathrm{SUGAR}^{[0]} \leftarrow \mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda^{[0]})$ computed from Algorithm 3

   for $j = 0$ **to** $J_{\max} - 1$ **do**

      $d^{[j]} = -H^{[j]} \mathrm{SUGAR}^{[j]}$

      $\alpha^{[j]} \in \underset{\alpha \in \mathbb{R}}{\mathrm{Argmin}}\, \mathrm{SURE}_{\varepsilon,\delta}(\Lambda^{[j]} + \alpha d^{[j]})$, SURE computed from Algorithm 3

      $\Lambda^{[j+1]} = \Lambda^{[j]} + \alpha^{[j]} d^{[j]}$

      $\mathrm{SUGAR}^{[j+1]} \leftarrow \mathrm{SUGAR}_{\varepsilon,\delta}(\Lambda^{[j+1]})$ computed from Algorithm 3

      $u^{[j]} = \mathrm{SUGAR}^{[j+1]} - \mathrm{SUGAR}^{[j]}$

      $H^{[j+1]} = \left(\mathrm{Id} - \dfrac{d^{[j]}\left(u^{[j]}\right)^\top}{\left(u^{[j]}\right)^\top d^{[j]}}\right) H^{[j]} \left(\mathrm{Id} - \dfrac{u^{[j]}\left(d^{[j]}\right)^\top}{\left(u^{[j]}\right)^\top d^{[j]}}\right) + \alpha^{[j]} \dfrac{d^{[j]}\left(d^{[j]}\right)^\top}{\left(u^{[j]}\right)^\top d^{[j]}}.$

   end for

   return $\widehat{\Lambda}_{\mathrm{BFGS}} = \Lambda^{[T_{\max}]}$

4 Nonlinear denoising in non linear physics: low confinement solid friction and porous media multiphase flows

4.1 Low confinement solid friction: Piecewise linear denoising.

Context – Friction experiments aim at probing not only the characteristics of materials, but also the dynamics of systems involving surfaces in contact. In particular, they are paradigms for modeling and attempting to predict earthquake dynamics [40]. The classical solid friction experiment consists of towing a mass m (so-called *slider*) over a substrate via a spring of stiffness k pulled at velocity V (see for instance Figure 2 in [2]). The signal representative of the slider dynamics is the force measured at the contact point between the spring and the slider. Among the different regimes described in solid friction, we can distinguish the stick-slip, characterized by a tooth-shaped signal alternating slow, linear rise and fast drops, the inertial regime, in which the signal becomes periodic and resembles a sine curve, and the continuous sliding regime, characterized by an almost constant signal superimposed with noise [1]. The appearance of creep, slow forward motion of the slider previous to a slip event, may also modify the signal shape. The challenge in such studies is to establish a *regime diagram* describing (and therefore, predicting) the system dynamics depending on the parameters (m, k, V). If the recognition of the different regimes is easy for large mass m, experiments with low confinement pressure, necessary to avoid surface wear, are challenging as they add noise to the experimental signals [2]. In this context, new signal processing tools are required. Here we focus in particular on signal denoising by approximating, at first order, the stick-slip signals to piecewise linear signals.

Data – Experiments of solid friction (taken from [2]) were performed by pulling a mass $m = 30.7$ g (slider area 9×6 cm^2) over a solid substrate. Both surfaces in contact consist of paper samples (Canson$^®$, characterized by its roughness). A cantilever spring (metallic blade of stiffness k between 168 and 3337 N/m) is pulled at constant velocity V (between 42 and 7200 μm/s) and is in contact with the slider by a steel ball glued onto this latter, ensuring a punctual contact and the free motion of the contact point. The slider dynamics is quantified though the measurement of the blade deflection, Δx, by an inductive sensor (Baumer, IPRM 12I9505/S14). In all experiments, the mass m is kept constant. We vary the parameters (k, V) and, for each experiment, record the normalized force signal F^* from the blade deflection, $F^* = k\Delta x/(mg)$, where $g = 9.81$ m/s^{-2} is the gravitational acceleration. This signal is recorded with a sampling frequency of 2 kHz, and its size varies from about 4.5×10^3 to 7.5×10^5.

Piecewise linear denoising – In [2], stick-slip signals were processed using an optimization formalism, falling under formulation (6), in order to enforce piecewise linear behavior. The observation z corresponds to the measured force signals, $A = $ Id, the linear operator D_Λ is chosen to be the discrete Laplacian

described in Equation (9) (for an univariate signal, i.e. $K = 1$), and $\|\cdot\|$ is the $\ell_{1,2}$-norm defined in (8), which reduces to the ℓ_1-norm in the context of univariate signals.

The tedious task of tuning the regularization parameter λ was performed by expert visual inspection and led to a choice $\lambda_{\mathrm{expert}} = 0.8$ uniformly applied to all signals, irrespective of the different experiment settings. Examples of noisy observations are shown in Figure 2(gray), with associated piecewise linear estimates obtained with $\lambda_{\mathrm{expert}}$ displayed in red. The regularized signals appear to capture well the transition between the stick and slip regimes.

Here, we propose to illustrate the use of the regularization parameter automated tuning strategy presented in Section 3.4 for piecewise linear denoising on stick-slip signals. $\mathrm{SURE}_{\varepsilon,\delta}(\lambda)$ is used as the quality criterion and minimized over λ. Therefore, both grid search and automated tuning are implemented and compared.

Automated data-driven hyperparameter tuning – The Finite Difference step ε, involved in $\mathrm{SURE}_{\varepsilon,\delta}$ and $\mathrm{SUGAR}_{\varepsilon,\delta}$ computation [see Equations (22) and (24)] is set to

$$\varepsilon = \frac{2\sigma^2}{N_1^{0.3}} \tag{25}$$

with N_1 the length of the considered stick-slip signal and σ^2 the estimated variance of the additive noise corrupting the signal. Since no additional information about the noise is available, σ^2 is estimated using the sample variance estimator applied to observations.

$\mathrm{SURE}_{\varepsilon,\delta}$ (black curve in Figure 3) is first computed over a grid of 15 logarithmically spaced values of the regularization parameter λ, using Algorithm 2. Then, λ_{grid} is defined as the minimum of $\mathrm{SURE}_{\varepsilon,\delta}(\lambda)$ over the grid and indicated by the '+' symbol. Finally, the quasi-Newton Algorithm 4 for automated tuning of regularization parameter is run, providing λ_{BFGS}, represented by the '∗' symbol. The regularization parameter chosen by the expert is displayed for comparison purpose, an indicated by the '×' marker. For each experimental setting (k, V), the grid search optimal regularization parameter λ_{grid} and the automatically tuned regularization parameter $\lambda_{\mathrm{BFGS}}(k, V)$ obtained respectively from Algorithms 2 and 4 are compared in Table 1, showing satisfactory agreement.

Denoised data analysis – Table 1 shows first that λ_{BFGS} is within the same order of magnitude as $\lambda_{\mathrm{expert}}$. This is consistent with Figure 2, that further shows that denoised experimental signals obtained from nonlinear filtering enforcing piecewise linear behavior, with automated hyperparameter tuning (blue) and expert-selected hyperparameter (red) display similar shapes and behaviors. This is a very satisfactory outcome as the proposed data-driven and automated hyperparameter tuning yields outcomes very consistent with those obtained from expert choices, without making use of any a priori information, and relying on data only instead.
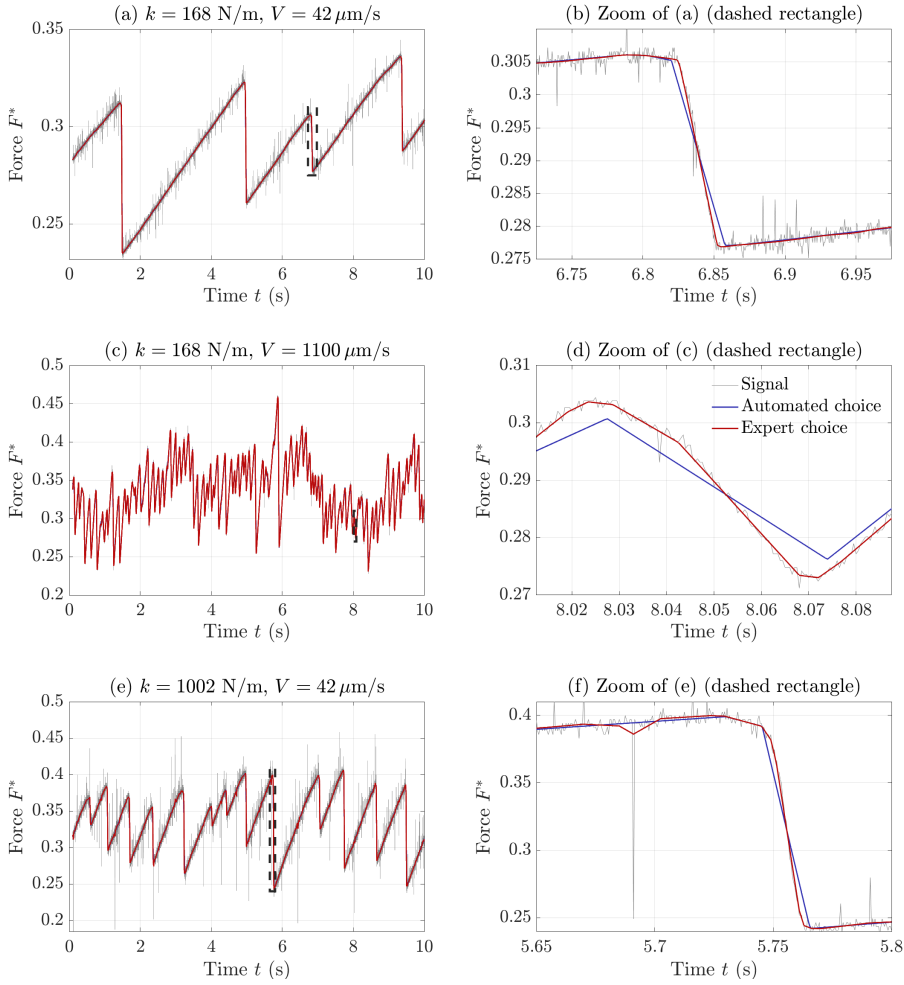
Fig. 2: **Stick-slip force signals.** Experimental data ((in grey) for three different experimental settings and nonlinear filtering enforcing piecewise linear behavior, with automated hyperparameter tuning (blue) and expert-selected hyperparameter (red).

Table 1 also shows that the automated procedure yields different regularization parameters for the different (k, V) configurations, illustrating an ability to finely adapt to data, which would not be possible - or would be too much time-consuming - for an expert. Table 1 further reveals that the automatically selected regularization parameters, λ_{BFGS}, are, for almost all (k, V) configurations, slightly larger than the expert-selected ones, λ_{expert}, hence yielding overall more regular signals. Figure 2(f) shows that the red signal, obtained with λ_{expert}, displays discontinuities (e.g, around $t = 5.7$ s) which are due to noise rather than to the physical mechanisms of interest, that are sat-

$\lambda_{\mathrm{grid}}(k,V)$		k (N/m)			$\lambda_{\mathrm{BFGS}}(k,V)$		k (N/m)		
		168	**1002**	**2254**			**168**	**1002**	**2254**
V (μm/s)	**42**	21.6	12.7	23.1	V (μm/s)	**42**	7.9	10.0	0.2
	1100	16.6	3.7	76.6		**1100**	16.5	3.4	2.2
	4300	8.8	6.2	2.6		**4300**	9.1	4.9	3.2

Table 1: Grid search v.s. automated tuning of regularization parameter in piecewise linear denoising of stick-slip signals for different stiffness k and velocity V.
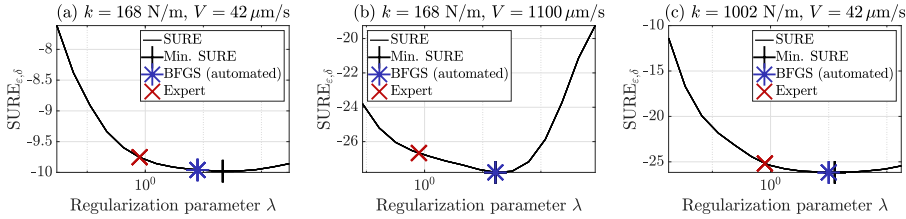

Fig. 3: $\mathrm{SURE}_{\varepsilon,\delta}(\lambda)$. Grid search v.s. automated tuning of regularization parameter.

isfactorily properly discarded on blue signal, obtained with the automated selection λ_{BFGS}, hence showing the interest of tuning the hyperparameter to each specific signal. However, Figure 2(b) and (d) also shows small yet visible differences during the slip-phase (fast decrease) between the red signals, obtained with $\lambda_{\mathrm{expert}}$, and the blue signals, obtained with the automated selection λ_{BFGS}. To decide which one is the most relevant requires returning to a detailed analysis of solid friction: The stick phases actually produce force signals that are exactly linearly increasing ; For the slip phase, while they can be described in first approximation as an abrupt linear decrease, detailed analysis indicates that they actually consist of arches of sinusoidal functions that connect the stick phases. Therefore, it can be considered that the expert-driven signals (red) better fit the experimental data, at the price though of concatenating a series of short linear segments that are irrelevant with respect to the underlying physics, whereas the data-driven signals (blue) yield more stylized piecewise linear approximations of the data, that may however better capture the times of transitions between stick and slip phases, an information of premier importance to analyze solid friction regimes.

In sums, deciding between the use of expert versus automated tuning of the hyperparameters combines several issues ranging from feasibility (expert tuning is time consuming, prone to errors and may lack reproducibility) to relevance (denoised signals must permit relevant access to quantity of interest for the physics).

4.2 Porous media multiphase flow: Piecewise homogeneous texture segmentation with weighted isotropic TV.

Context – Understanding and predicting the dynamics of multiphase flows is a major issue in geosciences (soil decontamination, CO_2 sequestration) and in the industry (enhanced oil recovery, heterogeneous catalysis) [41–44]. Among these processes, many involve a joint gas and liquid flow through a porous medium. Quantifying the contact areas between the different phases, where chemical reactions take place, is of tremendous importance for analyzing and predicting the efficiency of such processes [45]. However, even when direct visualization is possible, the porous medium generates a global, multiscale texture on images which makes it difficult to extract the gas-liquid interfaces. Segmentation techniques based on morphological tools used so far to differentiate phases in multiphase flows [46] present severe limitations: arbitrary threshold setting, non-physical irregular bubble contour, non detection of small bubbles. In addition, recent developments in high-resolution and high-speed imaging yield large-size images and large data sets, thus bringing forward issues in memory and computational costs. Here, we focus on the identification of the different phases (liquid and gas) in textured images. As a first approximation, the liquid and the gas appear as homogeneous fractal textures. Hence, discriminating phases requires to solve a texture segmentation problem.

Data – Experiments of joint gas and liquid flow through a porous medium were performed in a quasi-2D vertical Hele-Shaw cell of width 210 mm, height 410 mm and gap 1.75 mm (see Figure 1 in [47]). The porous medium is an open cell solid foam of NiCrFeAl alloy (Alantum), with a typical pore diameter of 580 μm. Constant gas and liquid flow rates are injected at the bottom of the cell through nine injectors (air) and a homogeneous slit (water). Images of the multiphase flow are acquired by a high-resolution camera (Basler A2040-90um, 2048×2048 pixels + 16 mm lens) at 100 Hz [48,47]. After cropping the region of interest, the size of the images to analyze is 1626×1160. An example is provided in Figure 4(a), showing that the gas phase (dark gray or white structures) is textured because of the presence of the foam struts which are not captured by the camera resolution. The liquid phase (in gray) is also textured though at smaller scales, as can be observed in Figure 5(a). For all experimental data sets, 50 to 3000 images are recorded. Similarly to the friction experiment, a large number of data sets associated with different parameters (here the gas and liquid flow rate) are investigated, to analyze the different hydrodynamic regimes.

Fractal features – We consider *fractal*, or *scale-free*, features, consisting of the local behavior as functions of scales of the wavelet *leader* coefficients $\mathcal{L}_{j,\underline{n}}$ and scale $j \in \{1, \ldots, J\}$, built as a local supremum of wavelet coefficients [49, 50]. For each pixel $\underline{n} \in \Omega = \{1, \ldots, N_1\} \times \{1, \ldots, N_2\}$, the *leader* coefficients of the image X to analyze, denoted $\mathcal{L}_{j,\underline{n}}$, evidence the following local scaling

property [51]

$$\mathcal{L}_{j,\underline{n}} \sim \eta_{\underline{n}} 2^{jh_{\underline{n}}}, \quad \text{as } 2^j \to 0 \tag{26}$$

where 2^j denotes the scale of the multiscale transform. The quantity $h_{\underline{n}}$ measures the *local regularity* of the texture at pixel $\underline{n}$. In log-log coordinates, Equation (26) corresponds to a linear behavior through octaves j

$$\log_2(\mathcal{L}_{j,\underline{n}}) \simeq \log_2(\eta_{\underline{n}}) + jh_{\underline{n}}, \quad \text{as } 2^j \to 0. \tag{27}$$

Setting $v_{\underline{n}} := \log_2(\eta_{\underline{n}})$, which will be called in the following the *local power* of the texture, a texture X is characterized by $\left(h_{\underline{n}}, v_{\underline{n}}\right)_{\underline{n} \in \Omega}$.

Definition 3 An *homogeneous* texture is characterized by a uniform local regularity $h_{\underline{n}} \equiv H$ and local power $v_{\underline{n}} \equiv V$.

Then, texture segmentation consists in identifying a partition of the image domain

$$\Omega = \Omega_1 \cup \cdots \cup \Omega_Q, \quad \Omega_q \cap \Omega_{q'} = \emptyset \text{ for } q \neq q', \tag{28}$$

for which both $h_{\underline{n}}$ and $v_{\underline{n}}$ are uniform on each Ω_q. In other words, it consists in obtaining piecewise constant maps of local regularity and local power.

Regularized estimates – Linear regression on log-*leaders* (27) can be formulated as the minimization of the following least-squares

$$\Phi(h, v; \mathcal{L}) = \frac{1}{2} \sum_{j=j_1}^{j_2} \|jh + v - \log_2 \mathcal{L}_j\|^2, \tag{29}$$

and provides estimates $\left(\widehat{h}_{\text{LR}}, \widehat{v}_{\text{LR}}\right)$ of fractal features. As an example, the linear regression estimate of the local regularity of the (zoomed) flow image of Figure 4(a) (Figure 5(a)) is presented in Figure 4(b) (Figure 5(b)). These estimates turn out to suffer from large variances precluding their use of actual segmentation, thus calling for nonlinear estimation tools.

To favor piecewise homogeneous segmentation, we enforce piecewise constancy in estimated features via two different Total Variation-based penalizations, leading to the minimization of the *Joint* and the *Coupled* functionals

$$\left(\widehat{h}^{\text{J/C}}, \widehat{v}^{\text{J/C}}\right) = \arg\min_{h,v} \Phi(h, v; \mathcal{L}) + \lambda \Psi_{\text{J/C}}(h, v; \alpha). \tag{30}$$

The *Joint* and *Coupled* penalizations are defined as

$$\Psi_{\text{J}}(h, v) := \lambda \left(\alpha \text{TV}(h) + \text{TV}(v)\right), \tag{31}$$

$$\Psi_{\text{C}}(h, v) := \lambda \sum_{n_1=1}^{N_1-1} \sum_{n_2=1}^{N_2-1} \sqrt{\alpha^2 (Hh)_{n_1,n_2}^2 + \alpha^2 (Vh)_{n_1,n_2}^2 + (Hv)_{n_1,n_2}^2 + (Vv)_{n_1,n_2}^2}. \tag{32}$$

where the total variation (TV) is defined in Equation (11) and the horizontal and vertical discrete gradients, H and V, are defined at Equation (10). While the *Joint* penalization imposes independently piecewise constancy of local regularity h and local power v, the *Coupled* penalization is more restrictive and favors co-localized changes in h and v. The trade-off between fidelity to the mathematical model (27) and piecewise constancy of h and v is controlled by the regularization parameter $\lambda > 0$ and $\alpha > 0$.

Iterated thresholding – From the regularized estimates, $\widehat{h}_{\mathrm{J/C}}$, one can obtain a segmentation by applying a post-processing thresholding. The iterated thresholding procedure, proposed in [52,53], benefiting from theoretical assessment, is customized to the gas/liquid segmentation problem in Algorithm 5. It is used systematically in the following, leading to the proposed T-*Joint* and T-*Coupled* segmentation procedures introduced in [54].

Algorithm 5 T-ROF: iterative thresholding of $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$

Require: $\widehat{h}$
Initialization: $\mathrm{m}_0^{[0]} = \min\limits_{\underline{n}\in\Omega} \widehat{h}_{\underline{n}}, \quad \mathrm{m}_1^{[0]} = \max\limits_{\underline{n}\in\Omega} \widehat{h}_{\underline{n}}.$

    for $t \in \mathbb{N}^*$ **do**
        {Compute the threshold:}
        $\mathrm{T}^{[t-1]} = \left(\mathrm{m}_0^{[t-1]} + \mathrm{m}_1^{[t-1]} \right)/2$
        {Threshold $\widehat{h}$:}
        $\Omega_0^{[t]} = \{\underline{n} \,|\, \widehat{h}_{\underline{n}} \leq \mathrm{T}^{[t]}\}, \quad \Omega_1^{[t]} = \{\underline{n} \,|\, \widehat{h}_{\underline{n}} > \mathrm{T}^{[t]}\}$
        {Update region mean:}
        $\mathrm{m}_0^{[t]} = 1/|\Omega_0| \sum\limits_{\underline{n}\in\Omega_0} \widehat{h}_{\underline{n}}, \quad \mathrm{m}_1^{[t]} = 1/|\Omega_1| \sum\limits_{\underline{n}\in\Omega_1} \widehat{h}_{\underline{n}}.$
    end for
    return $\Omega_0 = \Omega_0^{[\infty]}$ (liquid), $\quad \Omega_1 = \Omega_1^{[\infty]}$ (gas)

Compared texture segmentation procedures – Four procedures falling under Model (6) and satisfying assumptions of Theorem 1 and Theorem 2 will be compared for texture segmentation, summarized in Table 2. Note that they differ both by the functional minimized and the noise model, which is of crucial importance in Stein procedures.

The first one, denoted ROF-Id, is a state-of-the-art piecewise constant denoising method, applied on $\widehat{h}_{\mathrm{LR}}$ seen as an observation of $\bar{h}$ corrupted by additive i.i.d. zero-mean Gaussian noise of variance σ^2, hence with scalar covariance matrix $\mathcal{S} = \sigma^2\mathrm{Id}$.

The three procedures ROF-$\mathcal{S}$, *Joint* and *Coupled* take into account the covariance structure of the log-*leaders* coefficients, evidencing both inter-scale and spatial correlations encapsulated in a non-diagonal covariance matrix $\mathcal{S}$.

The linear operator intervening in the data fidelity term of *Joint* and *Coupled* procedures, denoted J, acts on the double variable (h,v) as $J(h,v) := (jh+v)_{j=J_1}^{j_2}$. We showed in a previous work [54] that it is full-rank. Hence,

Theorem 1 applies. Moreover, the strong-convexity modulus $\mu = 2\min \mathrm{Sp}(J^\top J)$, where Sp denotes the spectrum of a linear operator, only depends on the octave range $\{j_1, \ldots, j_2\}$ and its numerical values are provided for fixed $j_1 = 1$ and varying j_2 in Table 3 [54].

Method	Figures 4, 5	Observation z	Operator A	Variable x	Penalization	Covariance
ROF-Id	(c), (d)	$\widehat{h}_{\mathrm{LR}}$	Id	h	TV	$\sigma^2\mathrm{Id}$
ROF-$\mathcal{S}$	(e), (f)	$\widehat{h}_{\mathrm{LR}}$	Id	h	TV	$\mathcal{S}$
Joint	(g), (h)	$\log_2(\mathcal{L})$	J	(h, v)	Ψ_{J}	$\mathcal{S}$
Coupled	(i), (j)	$\log_2(\mathcal{L})$	J	(h, v)	Ψ_{C}	$\mathcal{S}$

Table 2: Four different settings considered in the experiments of local regularity-based texture segmentation with automated choice of hyperparameters procedures. $\widehat{h}_{\mathrm{LR}}$ stands for the minimizer of (29), $\mathcal{L}$ denotes the wavelet leaders of the image to analyze, TV stands for total-variation penalization as defined in (11), and with h, v, Ψ_{J} and Ψ_{C}, $\mathcal{S}$ are defined in this section denote respectively the local regularity, the local variance, the Joint penalization, the Coupled penalization and the covariance matrix.

Automated hyperparameter tuning – Stein based formalism, described in Section 3.4, is used, first, to obtain an estimation of the quadratic risk from $\mathrm{SURE}_{\varepsilon,\delta}(\lambda, \alpha)$, second, for automated tuning of regularization parameters thanks to $\mathrm{SUGAR}_{\varepsilon,\delta}(\lambda, \alpha)$ estimate.

For this purpose, it is necessary to provide an estimate of the covariance matrix of the noise. The estimated noise variance σ^2 involved in ROF-Id is obtained from the variance of $\widehat{h}_{\mathrm{LR}}$, while the covariance matrix $\mathcal{S}$ is assimilated to the covariance of the log-*leaders* of the textured image X to be segmented.

The Finite Difference step ε, involved in $\mathrm{SURE}_{\varepsilon,\delta}$ and $\mathrm{SUGAR}_{\varepsilon,\delta}$ computation (see Equations (22) and (24)) is set to

$$\varepsilon = \frac{2\sqrt{\max \mathcal{S}}}{M^{0.3}} \tag{33}$$

where M is the size of the observation vector and the maximum is taken over all coefficients of the covariance matrix and $M = N_1 \times N_2$ in the case of ROF-Id and ROF-$\mathcal{S}$, $M = (j_2 - j_1 + 1) \times N_1 \times N_2$ in the case of *Joint* and *Coupled* procedures.

Accuracy of the automated tuning – Grid search minimization of $\mathrm{SURE}_{\varepsilon,\delta}(\lambda, \alpha)$ (Algorithm 2) being costly, due to the large number of Algorithm 1 runs required, it is performed on a zoomed image of 281×231 pixels, presented in Figure 5(a). Then, automated tuning of λ and α from Algorithm 4, based on $\mathrm{SUGAR}_{\varepsilon,\delta}(\lambda, \alpha)$, is performed on the same zoomed image.

	$j_2 = 2$	$j_2 = 3$	$j_2 = 4$	$j_2 = 5$	$j_2 = 6$
μ	**0.29**	0.72	1.20	1.69	2.20

Table 3: Strong-convexity modulus μ of data-fidelity term of (6), for fixed $j_1 = 1$ and varied j_2. The bold entry correspond to the range of scales used in the experiments.

In practice, $\mathrm{SURE}_{\varepsilon,\delta}(\lambda)$ is computed on 15 values of the hyperparameter λ for ROF-Id (Figure 6(a)) and ROF-$\mathcal{S}$ (Figure 6(b)) procedures, and over a 15×15 grid of hyperparameters (λ, α) for *Joint* (Figure 6(c)) and *Coupled* (Figure 6(d)) methods. The grid search minimum, Λ_{grid}, indicated by the '+' symbol, is compared to the optimal regularization parameters found applying Algorithm 4, Λ_{BFGS}, indicated by the '∗' symbol. The optimal parameters Λ_{grid} and Λ_{BFGS} appear to coincide perfectly for ROF-Id and *Joint* procedures. As for ROF-$\mathcal{S}$ and *Coupled* strategies, even though they are different, they are consistent with $\mathrm{SURE}_{\varepsilon,\delta}$ profile, in the sense that they correspond to similar values of $\mathrm{SURE}_{\varepsilon,\delta}$. We observed that, while grid search minimization (Algorithm 2) required 225 runs of Algorithm 1 for *Joint* and *Coupled* methods, the automated tuning via BFGS quasi-Newton minimization (Algorithm 4) needed no more than 50 runs of Algorithm 3. Hence, when several parameters are involved, an automated strategy (Algorithm 4) is significantly faster than a grid search (Algorithm 2).

Segmentation results – Figure 3 indicates that the automated selection of regularization parameters is consistent with $\mathrm{SURE}_{\varepsilon,\delta}$ minimization. Hence, the complete images of 1626×1160 pixels will be analyzed only with Algorithm 4. The corresponding segmentation results are presented in Figures 4 and 5. State-of-the-art ROF-Id and ROF-$\mathcal{S}$ procedures yield regularized $\widehat{h}_{\mathrm{TV/ROF}}$ presenting artifacts, as observed in Figures 4(d) and 4(f), and hence lead to inaccurate segmentation, cf. Figures 4(c) and 4(e). In addition, a key point in such experiments is to estimate precisely the contact surface between the liquid and gas. Both T-ROF-Id and T-ROF-$\mathcal{S}$ (Figures 4(c) and (e)) present irregular contours, which are not representative of the real contours and strongly overestimate bubble perimeters. *Joint* and *Coupled* procedures, taking into account both the local regularity and the local variance yield more regular contours. In addition, the *Joint* and *Coupled* methods detect less artifacts (see Figures 4(g) and 4(i)). However, the *Joint* estimate $\widehat{h}_{\mathrm{J}}$ appears to be over-regularized, leading to non-detection of small bubbles in the segmentation of Figure 4(g). The *Coupled* procedure turns out to perform a satisfactory compromise, avoiding artifacts, yet, detecting small gas bubbles, as illustrated in Figures 4(i) and 4(j).

5 Conclusion and perspectives

The present work has described a unified framework for signal/image non linear filtering, formulated as an inverse problem, that can actually be affiliated to several functional minimization problems encountered in statistical (nonlinear) physics. This inverse problem formulation aims at favoring piecewise homogeneous signal and images, that naturally correspond to solutions on numerous problems in nonlinear physics, often very different in nature. Piecewise homogeneity assessment entails non smooth convex optimization, here handled via proximal operators. In addition to yielding relevant piecewise homogeneous estimates, the proposed framework also achieves an automated and data-driven tuning of hyperparameters inherently present in inverse problems and nonlinear filtering, thus avoiding the burden of conducting a prone to error and sometimes lacking reproductibily expert inspection. The potential and interest of nonlinear filtering has been illustrated at work on two, different in nature, real nonlinear physics experiments (low confinement solid friction and porous media multiphase flow). However, the approach has a fairly general level of applicability and a documented MAT-LAB toolbox both for multivariate signals and images, implementing both the nonlinear filtering favoring piecewise homogeneity and the automated data-driven hyperparameter selection, has been made publicly available at `https://github.com/bpascal-fr/stein-piecewise-filtering`.

References

1. T. Baumberger, C. Caroli, Adv. Phys. **55**(3-4), 279 (2006)
2. J. Colas, N. Pustelnik, C. Oliver, P. Abry, J.C. Géminard, V. Vidal, Phys. Rev. E **100**, 032803 (2019)
3. M. Serres, T. Maison, R. Philippe, V. Vidal, Int. J. Multiph. Flow **105**, 134 (2018)
4. B. Pascal, N. Pustelnik, P. Abry, M. Serres, V. Vidal, in *Proc. Int. Conf. Image Process.* (IEEE, Athens, Greece, 2018), pp. 2092–2096
5. M. Berhanu, R. Monchaux, S. Fauve, N. Mordant, F. Petrelis, A. Chiffaudel, F. Daviaud, B. Dubrulle, L. Marie, F. Ravelet, M. Bourgoin, P. Odier, J.F. Pinton, R. Volk, Eur. Phys. Lett. **77**, 59001 (2007)
6. T. Divoux, H. Gayvallet, J.C. Geminard, Phys. Rev. Lett. **101**, 148303 (2008)
7. T. Auger, J. Mathe, V. Viasnoff, G. Charron, J.M. Di Meglio, L. Auvray, F. Montel, Phys. Rev. Lett. **113**, 028302 (2014)
8. P.L. Combettes, V.R. Wajs, Multiscale Model. and Simul. **4**(4), 1168 (2005)
9. H.H. Bauschke, P.L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, 2nd edn. (Springer, New York, 2017)
10. N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, J.C. Pesquet, Wiley Encyclopedia of Electrical and Electronics Engineering (2016)
11. J.F. Cai, B. Dong, S. Osher, Z. Shen, J. Amer. Math. Soc. **25**, 1033 (2012)
12. S.Z. Li, *Markov Random Field modeling in image analysis* (Springer, 2009)
13. J. Møller, *Spatial statistics and computational methods* (Lecture Notes in Statistics, Springer, 2003)
14. S. Geman, D. Geman, in *Readings in Computer Vision* (Elsevier, 1987), pp. 564–584
15. D. Geman, S. Geman, in *Disordered systems and biological organization* (Springer, 1986), pp. 301–319
16. D. Geman, G. Reynolds, IEEE Trans. Pattern Anal. Match. Int. **14**(3), 367 (1992)
17. A. Chambolle, SIAM J. Appl Math. **55**, 827 (1995)

18. P. Lobel, C. Pichot, L. Blanc-Féraud, M. Barlaud, IEEE Antennas Propag. Mag. **39**(2), 12 (1997)
19. M. Storath, A. Weinmann, J. Frikel, M. Unser, Inverse Problems **31**(2), 025003 (2015)
20. L.I. Rudin, S. Osher, E. Fatemi, Physica D: Nonlinear Phenomena **60**(1-4), 259 (1992)
21. A. Chambolle, T. Pock, J. Math. Imag. Vis. **40**(1), 120 (2011)
22. Y. Marnissi, E. Chouzenoux, A. Benazza-Benyahia, J.C. Pesquet, Entropy **20**(110), 35p. (2018)
23. C. Vacar, J.F. Giovannelli, EURASIP J. Adv. Signal Process., special issue on Advanced Computational Methods for Bayesian Signal Processing (17) (2019)
24. N. Dobigeon, J.Y. Tourneret, M. Davy, IEEE Trans. Signal Process. **55**(4), 1251 (2007)
25. N. Dobigeon, J.Y. Tourneret, Comput. Stat. Data Anal. **51**(12), 5603 (2007)
26. M. Pereyra, N. Dobigeon, H. Batatia, J.Y. Tourneret, IEEE Trans. Image Process. **22**(6), 2385 (2013)
27. P.L. Combettes, J.C. Pesquet, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, ed. by H.H. Bauschke, R.S. Burachik, P.L. Combettes, V. Elser, D.R. Luke, H. Wolkowicz (Springer-Verlag, New York, 2011), pp. 185–212
28. L. Condat, J. Optim. Theory Appl. **158**(2), 460 (2013)
29. D.L. Donoho, J.M. Johnstone, Biometrika **81**(3), 425 (1994)
30. J. Frecon, N. Pustelnik, N. Dobigeon, H. Wendt, P. Abry, IEEE Trans. Signal Process. **65**(19), 5215 (2017)
31. A. Benazza-Benyahia, J.C. Pesquet, IEEE Trans. Image Process. **14**(11), 1814 (2005)
32. S. Ramani, T. Blu, M. Unser, IEEE Trans. Image Process. **17**(9), 1540 (2008)
33. C.A. Deledalle, S. Vaiter, J. Fadili, G. Peyré, SIAM J. Imaging Sci. **7**(4), 2448 (2014)
34. N. Parikh, S. Boyd, Foundations and Trends® in Optimization **1**(3), 127 (2014)
35. C.M. Stein, Ann. Stat. pp. 1135–1151 (1981)
36. Y.C. Eldar, IEEE Trans. Signal Process. **57**(2), 471 (2008)
37. B. Pascal, S. Vaiter, N. Pustelnik, P. Abry, Preprint arXiv:2004.09434 (2020)
38. J. Nocedal, S. Wright, *Numerical optimization* (Springer Science & Business Media, 2006)
39. F.E. Curtis, T. Mitchell, M.L. Overton, Optim. Methods Softw. **32**(1), 148 (2017)
40. C. Marone, Ann. Rev. Earth Planet. Sci. **26**, 643 (1998)
41. K.R. Reddy, J.A. Adams, J. Geotech. Geoenviron. Eng. **127**(3), 234 (2001)
42. Q. Kang, I.N. Tsimpanogiannis, D. Zhang, P.C. Lichtner, Fuel Process. Technol. **86**, 1647 (2005)
43. V. Hessel, P. Angeli, A. Gavriilidis, H. Löwe, Industrial & engineering chemistry research **44**, 9750 (2005)
44. A. Babchin, I. Brailovsky, P. Gordon, G. Sivashinsky, Phys. Rev. E **77**, 026301 (2008)
45. M.T. Kreutzer, F. Kapteijn, J.A. Moulijn, J.J. Heiszwolf, Chemical Engineering Science **60**, 5895 (2005)
46. M. Serres, M.L. Zanota, R. Philippe, V. Vidal, Int. J. Multiph. Flow **85**, 157 (2016)
47. T. Busser, M. Serres, R. Philippe, V. Vidal, in revision at Chem. Eng. Sci. (2020)
48. M. Serres, PhDThesis, École Normale Supérieure de Lyon, Université de Lyon p. 203 (2017)
49. H. Wendt, P. Abry, S. Jaffard, IEEE Signal Process. Mag. **24**(4), 38 (2007)
50. H. Wendt, S.G. Roux, P. Abry, S. Jaffard, Signal Process. **89**(6), 1100 (2009)
51. S. Jaffard, Fractal geometry and applications: A jubilee of Benoit Mandelbrot, M. Lapidus et M. van Frankenhuysen Eds, Proceedings of Symposia in Pure Mathematics (AMS) **72**(2), 91 (2004)
52. X. Cai, G. Steidl, in *Int. Workshop on Energy Minimization Methods in Comp. Vis. and Pat. Rec.* (Springer, 2013), pp. 237–250
53. X. Cai, R. Chan, C.B. Schonlieb, T. Steidl, G.and Zeng, Preprint arXiv:1807.10194 (2018)
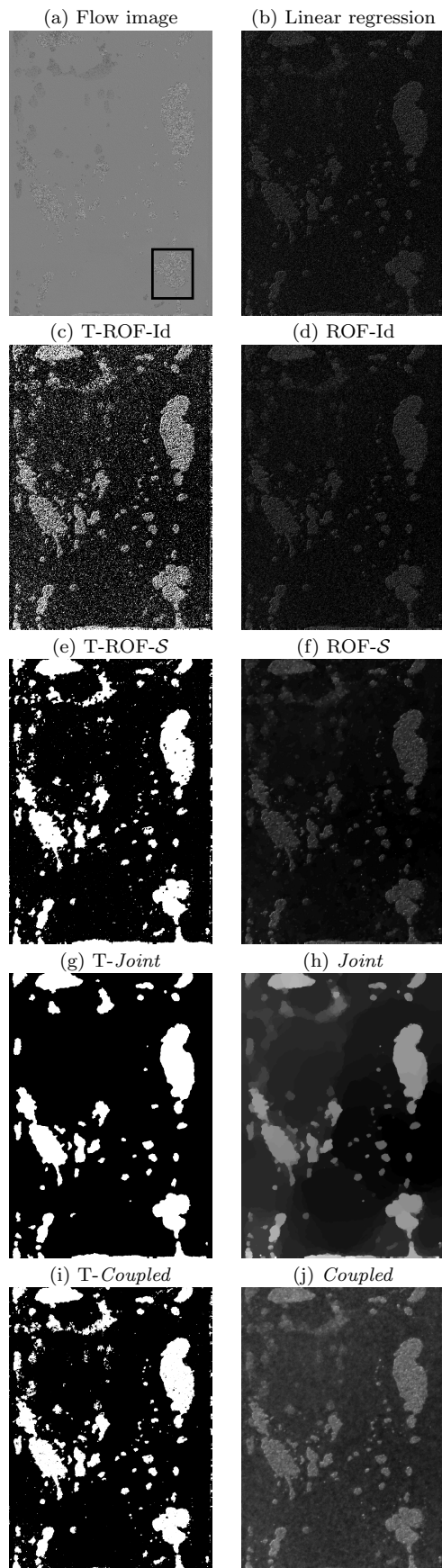54. B. Pascal, N. Pustelnik, P. Abry, Preprint arXiv:1910.05246 (2019)

Fig. 4: **Porous media multiphase flow texture segmentation based on fractal features.** Comparisons between different approaches as summarized in Table 2.
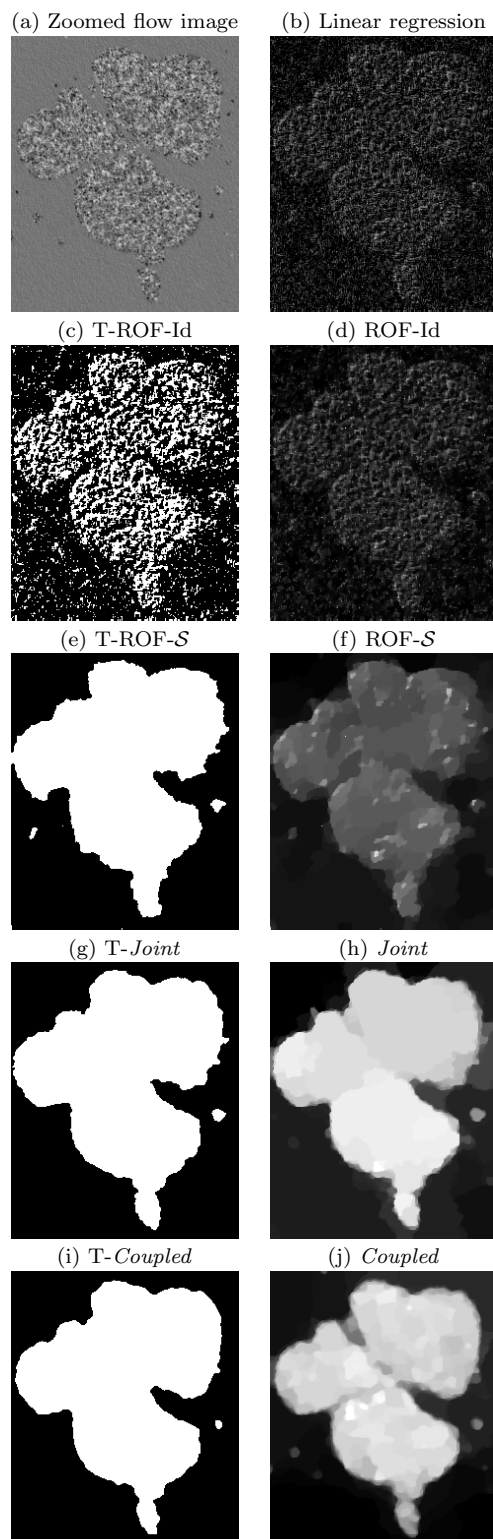
(a) Zoomed flow image (b) Linear regression

(c) T-ROF-Id (d) ROF-Id

(e) T-ROF-$\mathcal{S}$ (f) ROF-$\mathcal{S}$

(g) T-*Joint* (h) *Joint*

(i) T-*Coupled* (j) *Coupled*

Fig. 5: **Porous media multiphase flow texture segmentation based on fractal features.** Comparisons between different approaches summarized in Table 2. *Zoom* on the area marked by the black rectangle in Figure 4(a).
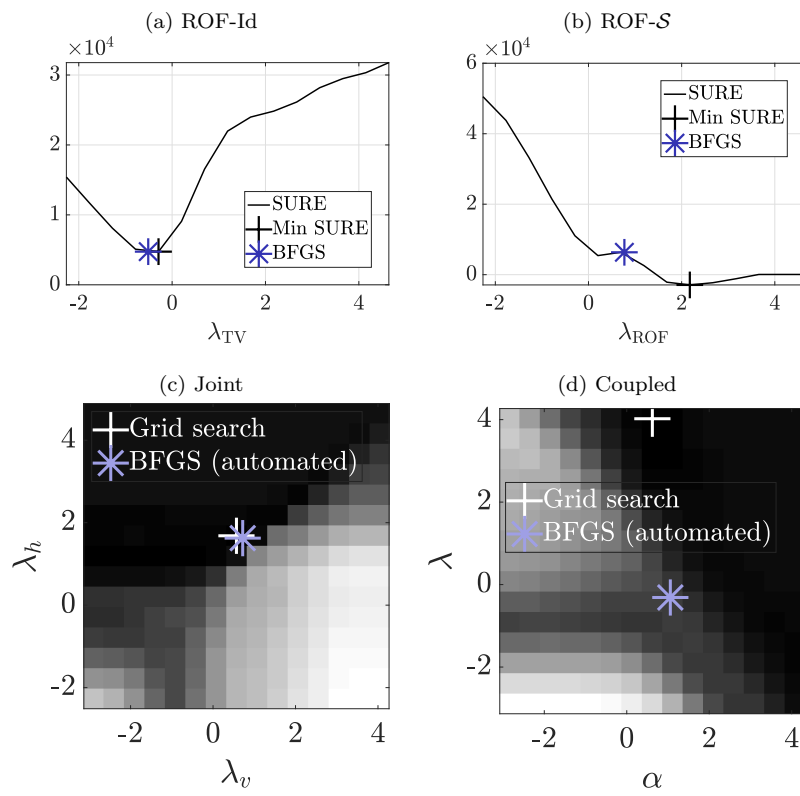
Fig. 6: Grid search stategy to minimize SURE for the segmentation of a *zoomed* multiphase flow image.