Strongly Convex Optimization for Joint Fractal Feature Estimation and Texture Segmentation

Barbara Pascal, Nelly Pustelnik, Patrice Abry

Strongly Convex Optimization for Joint Fractal Feature Estimation and Texture Segmentation

Barbara Pascal, Nelly Pustelnik and Patrice Abry[♯]

April 3, 2021

Abstract

The present work investigates the segmentation of textures by formulating it as a strongly convex optimization problem, aiming to favor piecewise constancy of fractal features (local variance and local regularity) widely used to model real-world textures in numerous applications very different in nature. Two objective functions combining these two features are compared, referred to as *joint* and *coupled*, promoting either independent or co-localized changes in local variance and regularity. To solve the resulting convex nonsmooth optimization problems, because the processing of large size images and databases are targeted, two categories of proximal algorithms (dual forward-backward and primal-dual), are devised and compared. An in-depth study of the objective functions, notably of their strong convexity, memory and computational costs, permits to propose significantly accelerated algorithms. A class of synthetic models of piecewise fractal texture is constructed and studied. They enable, by means of large-scale Monte-Carlo simulations, to quantify the benefits in texture segmentation of combining local regularity and local variance (as opposed to regularity only) while using strong-convexity accelerated primal-dual algorithms. Achieved results also permit to discuss the gains/costs in imposing co-localizations of changes in local regularity and local variance in the problem formulation. Finally, the potential of the proposed approaches is illustrated on real-world textures taken from a publicly available and documented database.

1 Introduction

Context: unsupervised texture segmentation. Textured images appear as *natural* models for a large variety of real-world applications very different in nature. Often, fractal attributes are used to relevantly characterize such real-world textures. This is the case with biological tissues [30], tomography-based pathology diagnostics [20, 29], art painting expertise [1], and microfluidics [35], to name a few examples.

[♯]B. Pascal, N. Pustelnik and P. Abry are with the Université de Lyon, ENS de Lyon, CNRS, Laboratoire de Physique, Lyon, France, e-mail: `firstname.lastname@ens-lyon.fr`

[†]This work was supported by Defi Imag'in SIROCCO and by ANR-16-CE33-0020 MultiFracs, France.

Often in these applications, texture segmentation (i.e., partitioning images into regions with homogeneous features) remains an on-going and major challenge. In computer vision or scene analysis, there exist numerous well-established and efficient methods to partition images, mostly relying on their geometrical properties (cf. e.g., [19, 23, 31, 39]). For textured images, segmentation remains challenging, as geometry is more difficult to capture, relying essentially on the statistics of the texture features.

Most classical texture segmentation approaches rely on two-step procedures, consisting of first computing *a priori* chosen texture features and, second, grouping features into regions with homogeneous feature statistics. Attempts to improve these traditional approaches address their two main limitations: arbitrary *a priori* feature selection and suboptimality of the two-step procedure. For instance, deep learning and neural network approaches have strongly contributed to renew image segmentation, avoiding notably the *a priori* selection of specific features, and combining feature estimation and segmentation (cf. e.g., [3]). However, their use remains designed mostly for the specific context of supervised segmentation: Typically, they require a very large labeled dataset, together with massive computing capabilities. Often, the required databases are not available, as expert annotations may be too expensive (in time and money) or technically beyond reach. Besides technical and financial issues, assembling such databases may raise non trivial ethical problems (cf. e.g., [13]). Along that line, a severe drawback of neural network approaches lies in the lack of *interpretability* or even of *identification* of the features decisions are based on: Doctors for instance might legitimately remain reluctant to base diagnostics on undocumented features. Therefore, despite the potential of deep learning, in contexts of absence of labeled database, of requirements for accurately estimating boundaries, or when there is a need of understandability, unsupervised segmentation strategies remain of critical importance.

Similarly to supervised neural network approaches, the present contribution aims to outperform traditional two-step texture segmentation procedures, while remaining in an unsupervised framework. A major advantage of our procedure, compared to neural networks or standard two-steps approaches, lies in its strongly convex formulation. It guarantees the existence and the uniqueness of the solution of the minimization problem involved in the segmentation procedure. Further, it provides computational efficiency.

Related work. Amongst classical features commonly enrolled in texture segmentation, one can list e.g., Gabor or short-term Fourier transform coefficients [18, 22], fractal dimension [11], Amplitude/Frequency Modulation models [24]. More recently, fractal (or scale-free) features were also involved in texture segmentation (cf. e.g., [47]). Notably, *local regularity* was used to quantify the fluctuations of roughness along the texture. Local regularity is quantified as an *optimal* local power-law behavior across scales for some multiscale quantities [47]. Modulus of wavelet coefficients [27] were initially used as multiscale quantities, followed by continuous wavelet transform modulus maxima [28, 32]. More recently, *wavelet leaders* (local suprema of discrete wavelet coefficients), used here, were shown to permit a theoretically accurate and practically robust estimation of local regularity [47], and successfully involved in texture characterization in e.g., [1, 29, 33]. Wavelet leaders are used to estimate local regularity at each pixel, mostly by linear regressions in log-log coordinates.

For the segmentation step, local Gabor coefficient histograms were grouped using matrix factorization [48] ; textons were combined with brightness and color features to yield a multiscale contour detection procedure [4]. Further, with fractal features, pixels sharing similar estimates of local regularity are grouped together *via* the minimization of the Rudin-Osher-Fatemi (ROF) functional [42], which amounts to performing *total variation* (TV) denoising of the local regularity estimate [38].

Further, in [38], it was also attempted to combine both steps into a single one by incorporating the regression weight estimation into the optimization procedure. The high computational burden implied by iteratively solving optimization problems as well as by tuning the regularization parameter has tentatively been addressed by *block splitting approaches* as suggested in [40] and explored in [34]. Strong convexity [10] constitutes another proposed track to address iterative optimization acceleration. However, the proposed joint features estimation and segmentation procedures, despite showing satisfactory and state-of-the-art segmentation performance suffered from major limitations: Their high computing cost prevents their use to process large images and databases ; While based on a key fractal feature, local regularity, they neglect changes in local power, potentially relevant information for texture segmentation, notably to extract accurate region boundaries.

Goals, contributions and outline. Aiming to address the above limitations, the overall goals of the present contribution are to investigate the potential benefits for texture segmentation brought by

- the use of joint estimates of fractal features (local variance and local regularity) ;
- the formulation of feature estimation and segmentation as a single step taking the form of a convex minimization ;
- the derivation of fast and scalable iterative algorithms to solve the optimization problem, permitted by explicitly proving (and measuring) strong convexity of the data fidelity term in the objective function ;
- the comparisons of several optimization formulations, favoring changes in features that are either co-localized in space or independent ;
- the derivation of an explicit stochastic process, modeling piecewise fractal textures, thus permitting to conduct large-size Monte-Carlo simulation for performance benchmarking and comparisons.

To that end, Section 2 recalls the key concepts and definitions related to local regularity, wavelet leaders and corresponding state-of-the-art linear regression followed by TV-based estimation procedures. The drawbacks of this approach are discussed in detail, and illustrated on a texture segmentation example. Section 3 describes the methodology of the first key contribution, a one-step estimation/segmentation procedure aiming to overcome the limitations of the two-step procedure. Thus, Section 3 proposes two alternative objective functions for the combined estimation/segmentation of local variance and regularity, respectively referred to as *joint* and *coupled*, based on the same data fidelity term but on two different total-variation based regularization strategies. Studying strong

convexity and duality gaps, two classes of fast iterative algorithms (*Primal-Dual* and *Dual Forward-backward*), are devised. Section 4 proposes the second contribution of this work. It describes a model of piecewise homogeneous fractal textures, providing well-controlled synthetic textures. Section 5 describes the third contribution. The performance of the estimation/segmentation procedures of Section 3 are assessed from Monte-Carlo simulations conducted on synthetic texture samples generated from the model of Section 4. Conducting such simulations requires to address issues related to regularization parameter selection, iterative algorithm stopping criterion, and performance metrics. It permits relevant answers to the final aim of assessing the actual benefits of using both local variance and regularity in texture segmentation with respect to issues such as sensitivity to fractal parameter changes, computational costs, impact of the different optimization formulations. Finally, the potential of the proposed segmentation approaches is illustrated at work on real-world piecewise homogeneous textures taken from a publicly available and documented texture database. Performance is also compared against a state-of-the-art segmentation strategy.

A MATLAB toolbox implementing the analysis and synthesis procedures devised here will be made freely and publicly available at the time of publication.

2 Local regularity

2.1 Local regularity and Hölder exponent

Local regularity can classically be assessed by means of the Hölder exponent [21, 26], defined as follows:

Definition 1 *Let $f\colon \Omega \to \mathbb{R}$ denote a 2D real field defined on an open set $\Omega \subset \mathbb{R}^2$. The Hölder exponent $h(\underline{z}_0)$ at location $\underline{z}_0 \in \Omega$ is defined as the largest $\alpha > 0$ such that there exists a constant $\chi > 0$, a polynomial $\mathcal{P}_{\underline{z}_0}$ of degree lower than α and a neighborhood $\mathcal{V}(\underline{z}_0)$ satisfying:*

$$\big(\forall \underline{z} \in \mathcal{V}(\underline{z}_0)\big) \quad \big|f(\underline{z}) - \mathcal{P}_{\underline{z}_0}(\underline{z})\big| \leq \chi \, \|\underline{z} - \underline{z}_0\|^{\alpha}$$

where $\|\cdot\|$ denotes the Euclidian norm.

Definition 1 does not provide a practical way to estimate local regularity. Thus, the practical assessment of $h(\underline{z}_0)$ usually relies on the use of multiscale quantities, such as wavelet coefficients, or wavelet leaders [21, 26, 47].

2.2 Local regularity and wavelet leaders

Because the practical aim is to analyse digitized images, definitions are given in a discrete setting. Let $X = (X_{\underline{n}})_{\underline{n} \in \Upsilon} \in \mathbb{R}^{|\Upsilon|}$ denote the digitized version of the 2D real field f on a finite grid $\Upsilon = \{1, \ldots, N\}^2$. Let $d_j = \mathbf{W}_j X$ denote the discrete wavelet transform (DWT) coefficients of X, at resolution $j \in \{j_1, \ldots, j_2\}$, with $\mathbf{W}_j\colon \mathbb{R}^{N \times N} \to \mathbb{R}^{M_j \times M_j}$ the operator formulation of the DWT.

Further, let the wavelet leader $\mathcal{L}_{j,\underline{k}}$, at scale 2^j and location $\underline{n} = 2^j \underline{k}$, be defined as the local supremum of modulus of wavelet coefficients in a small

neighborhood across all finer scales, [21, 26, 47] $\mathcal{L}_{j,\underline{k}} = \sup_{\lambda_{j',\underline{k}'} \subset 3\lambda_{j,\underline{k}}} |2^j d_{j',\underline{k}'}|$, with $\lambda_{j,\underline{k}} = [\underline{k}2^j, (\underline{k}+1)2^j)$ and $3\lambda_{j,\underline{k}} = \bigcup_{\underline{p}\in\{-1,0,1\}^2} \lambda_{j,\underline{k}+\underline{p}}$.

It was proven in [21,26,47] that wavelet leaders provide multiscale quantities intrinsically tied to Hölder exponents, insofar as when the underlying field f has Hölder exponent $h(\underline{x})$ at location $\underline{x}$, then the wavelet leaders $\mathcal{L}_{j,\underline{k}}$ satisfy:

$$\mathcal{L}_{j,\underline{k}} \simeq \eta(\underline{x}) 2^{j h(\underline{x})} \text{ as } 2^j \to 0, \tag{1}$$

for $\underline{x} \in \lambda_{j,\underline{k}}$. The quantity $\eta(\underline{x})$, referred to as the local *power*, is proportional to the local variance of f at location $\underline{x}$. The discretized field X inherits from the properties of f, and we denote by $h_{\underline{n}}$ (resp. $\eta_{\underline{n}}$) its local regularity (resp. power) at pixel $\underline{n}$.

2.3 Local regularity estimation

2.3.1 Linear regression

Eq. (1) naturally leads to estimate $\boldsymbol{h}$ and $\boldsymbol{v} = \log_2 \boldsymbol{\eta}$ by means of linear regression in log-log coordinates [21, 26, 47], denoted $\left(\widehat{\boldsymbol{v}}_{\mathrm{LR}}, \widehat{\boldsymbol{h}}_{\mathrm{LR}}\right)$:

$$(\forall \underline{n} \in \Upsilon) \qquad \begin{pmatrix} \widehat{v}_{\mathrm{LR},\underline{n}} \\ \widehat{h}_{\mathrm{LR},\underline{n}} \end{pmatrix} = \mathbf{J}^{-1} \begin{pmatrix} \mathcal{S}_{\underline{n}} \\ \mathcal{T}_{\underline{n}} \end{pmatrix} \tag{2}$$

$$\text{with} \quad \mathbf{J} = \begin{pmatrix} R_0 & R_1 \\ R_1 & R_2 \end{pmatrix} \quad \text{and} \quad R_m = \sum_j j^m, \tag{3}$$

$$\text{and} \quad \mathcal{S}_{\underline{n}} = \sum_j \log_2 \mathcal{L}_{j,\underline{n}}, \quad \mathcal{T}_{\underline{n}} = \sum_j j \log_2 \mathcal{L}_{j,\underline{n}}. \tag{4}$$

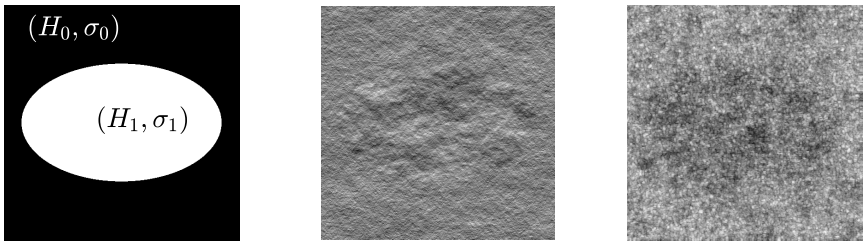
The sum $\sum_j$ implicitly stands for $\sum_{j=j_1}^{j_2}$ with (j_1, j_2) the range of octaves involved in the estimation. Though linear regressions provide estimates both for $\boldsymbol{h}$ and $\log_2 \boldsymbol{\eta}$, the later remains to date rarely used. Estimates of $\boldsymbol{h}$ were for instance used in [34,38]. Bayesian based extension of linear regressions were also and alternatively proposed to estimate η and h [46].

A sample of piecewise homogeneous fractal texture is displayed in Fig. 1b, synthesized using the mask sketched in Fig. 1a. The corresponding linear regression based estimates $\widehat{\boldsymbol{\eta}}_{\mathrm{LR}} \equiv 2^{\widehat{\boldsymbol{v}}_{\mathrm{LR}}}$, and $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$ (Fig. 1(c-d)) show too poor performance (high variability) to permit to detect the two regions and the corresponding boundary.

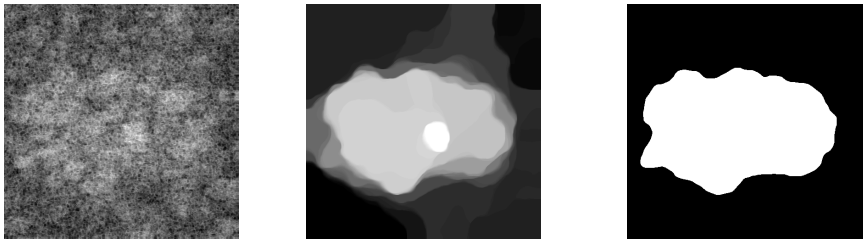
2.3.2 Total variation based estimates

To address the poor estimation performance achieved by linear regressions for local estimation, it has been proposed to *denoise* $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$ by using an optimisation framework involving TV regularization, as a mean to enforce piecewise constant estimates [33,38]:

$$\widehat{\boldsymbol{h}}_{\mathrm{ROF}} = \operatorname*{argmin}_{\boldsymbol{h}\in\mathbb{R}^{|\Upsilon|}} \frac{1}{2} \|\boldsymbol{h} - \widehat{\boldsymbol{h}}_{\mathrm{LR}}\|_2^2 + \lambda \, \mathrm{TV}(\boldsymbol{h}). \tag{5}$$



(a) Mask (b) Texture X (c) $\widehat{\boldsymbol{v}}_{\mathrm{LR}}$



(d) $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$ (e) $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$ (f) $\widehat{\boldsymbol{M}}_{\mathrm{ROF}}$

Figure 1: **Piecewise homogeneous fractal texture and local variance and regularity estimates.** (a) Synthesis mask $(H_0, \Sigma_0) = (0.5, 0.6)$ and $(H_1, \Sigma_1) = (0.8, 0.65)$; (b) sample texture (see Section 4), (c) and (d) linear regression based estimates of local variance and regularity $\widehat{\boldsymbol{v}}_{\mathrm{LR}}$ and $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$; (e) and (f) Total variation based estimates of local regularity $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$ and segmentation $\widehat{\boldsymbol{M}}_{\mathrm{ROF}}$ obtained with Alg. 1.

The above functional balances a data fidelity term against a total variation penalization defined as:

$$\mathrm{TV}(\boldsymbol{h}) = \|\mathbf{D}\boldsymbol{h}\|_{2,1}, \tag{6}$$

where the operator $\mathbf{D} : \mathbb{R}^{|\Upsilon|} \to \mathbb{R}^{2\times|\Upsilon|}$, consists of horizontal and vertical increments:

$$(\mathbf{D}\boldsymbol{h})_{n_1,n_2} = \begin{pmatrix} h_{n_1,n_2+1} - h_{n_1,n_2} \\ h_{n_1+1,n_2} - h_{n_1,n_2} \end{pmatrix} \tag{7}$$

and where the mixed norm, $\|\cdot\|_{2,1}$, defined for $\boldsymbol{z} = [\boldsymbol{z}_1; \ldots; \boldsymbol{z}_I]^\top \in \mathbb{R}^{I\times|\Upsilon|}$ as:

$$\|\boldsymbol{z}\|_{2,1} = \sum_{n_1=1}^{N_1-1}\sum_{n_2=1}^{N_2-1}\sqrt{\sum_{i=1}^{I}(z_i)_{n_1,n_2}^2} \tag{8}$$

which fosters an isotropic and piecewise constant behaviour of the estimates [38].

While showing substantial benefits compared to linear regression based estimates and although algorithmically efficient, the TV-based procedure above, proposed in [38], suffers from several significant limitations: i) It is restricted to the estimation of local regularity only and neglects local variance ; ii) It consists of a two-step process (first, apply linear regression to obtain $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$; second, apply TV to $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$ to obtain $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$) and is hence potentially not optimal. Section 3 below will propose several solutions that address these two limitations by

performing the estimate of both local variance and regularity in a single pass at similar computational cost as the two-step procedure.

2.3.3 A posteriori segmentation and local regularity global estimation

When the number of regions is known *a priori*, it has been shown in [7,8] that a fast iterative thresholding post-processing procedure can be applied to $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$ to obtain a segmentation of the TV denoised estimate. Papers [7,8] provided theoretical guarantees linking this Threshold-ROF (T-ROF) strategy to Mumford-Shah-like segmentation. Alg. 1 explicitly describes the procedure applied to $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$ to obtain a two-region T-ROF segmentation $\widehat{\boldsymbol{M}}_{\mathrm{ROF}}$ with $\widehat{M}_{\mathrm{ROF},\underline{n}} = 0$ if $\underline{n} \in \Upsilon_0$ and 1 otherwise. Figs. 1e, 1f display examples of TV denoised estimate $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$ and T-ROF segmentation $\widehat{\boldsymbol{M}}_{\mathrm{ROF}}$.

Algorithm 1: T-ROF: iterative thresholding of $\widehat{\boldsymbol{h}}_{\mathrm{ROF}}$

Input: $\boldsymbol{h}$

Initialization: $\mathrm{m}_0^{[0]} = \min_{\underline{n}\in\Upsilon} h_{\underline{n}}, \quad \mathrm{m}_1^{[0]} = \max_{\underline{n}\in\Upsilon} h_{\underline{n}}.$

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 $\boldsymbol{h}$*:*

$$\Upsilon_0^{[t]} = \{\underline{n} \,|\, h_{\underline{n}} \leq \mathrm{T}^{[t]}\}, \quad \Upsilon_1^{[t]} = \{\underline{n} \,|\, h_{\underline{n}} > \mathrm{T}^{[t]}\}$$

Update region mean:

$$\mathrm{m}_0^{[t]} = 1/|\Upsilon_0| \sum_{\underline{n}\in\Upsilon_0} h_{\underline{n}}, \quad \mathrm{m}_1^{[t]} = 1/|\Upsilon_1| \sum_{\underline{n}\in\Upsilon_1} h_{\underline{n}}.$$

Output: $\Upsilon_0 = \Upsilon_0^{[\infty]}, \quad \Upsilon_1 = \Upsilon_1^{[\infty]}$

From T-ROF segmentation, Texture X can be interpreted as the concatenation of two fractal textures X_0 and X_1, each with its own uniform regularity and variance. Therefore, *a posteriori* global estimates, $\widehat{H}_{0,\mathrm{ROF}}$ and $\widehat{H}_{1,\mathrm{ROF}}$, for the local regularity of each region, Υ_0 and Υ_1, can be obtained by performing linear regressions applied to the logarithm of multiscale quantities $\boldsymbol{\mathcal{L}}_j$ averaged over Υ_0 (resp. Υ_1) (cf. [47]).

3 Total variation based estimation of local variance and regularity

3.1 Design of the objective function

3.1.1 Linear regression as functional minimization

Linear regression can (of course) be viewed as an optimization scheme. Indeed, setting $\boldsymbol{v} = \log_2 \boldsymbol{\eta}$ the following strictly convex functional (in variables $(\boldsymbol{v}, \boldsymbol{h})$):

$$\Phi(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{\mathcal{L}}) = \frac{1}{2} \sum_j \|\boldsymbol{v} + j\boldsymbol{h} - \log_2 \boldsymbol{\mathcal{L}}_j\|_{\mathrm{Fro}}^2 , \tag{9}$$

has a unique minimum corresponding to the linear regression estimates in 2.

3.1.2 Penalization

To improve on the poor performance of linear regression, a generic optimization framework can be proposed as $(\widehat{\boldsymbol{v}}, \widehat{\boldsymbol{h}}) = \underset{\boldsymbol{v}\in\mathbb{R}^{|\Upsilon|},\boldsymbol{h}\in\mathbb{R}^{|\Upsilon|}}{\arg\min}\ \Phi(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\mathcal{L}}) + \Psi(\boldsymbol{v},\boldsymbol{h})$, where the data fidelity term Φ is that of the linear regression (cf.(9)), and where the penalization term Ψ favors piecewise constancy of $\boldsymbol{v}$ and $\boldsymbol{h}$. Two different strategies are proposed:

(i) *Joint estimation*:

$$(\widehat{\boldsymbol{v}}_{\mathrm{J}}, \widehat{\boldsymbol{h}}_{\mathrm{J}}) = \underset{\boldsymbol{v},\boldsymbol{h}\in\mathbb{R}^{|\Upsilon|}\times\mathbb{R}^{|\Upsilon|}}{\arg\min}\ \Phi(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\mathcal{L}}) + \Psi_J(\boldsymbol{v},\boldsymbol{h}), \tag{10}$$

$$\text{with} \quad \Psi_{\mathrm{J}}(\boldsymbol{v},\boldsymbol{h}) = \lambda\left(\mathrm{TV}(\boldsymbol{v}) + \alpha\mathrm{TV}(\boldsymbol{h})\right), \tag{11}$$

that does not favor changes in $\boldsymbol{v}$ and $\boldsymbol{h}$ occurring at the same location;

(ii) *Coupled estimation*:

$$(\widehat{\boldsymbol{v}}_{\mathrm{C}}, \widehat{\boldsymbol{h}}_{\mathrm{C}}) = \underset{\boldsymbol{v},\boldsymbol{h}\in\mathbb{R}^{|\Upsilon|}\times\mathbb{R}^{|\Upsilon|}}{\arg\min}\ \Phi(\boldsymbol{v},\boldsymbol{h};\boldsymbol{\mathcal{L}}) + \Psi_{\mathrm{C}}(\boldsymbol{v},\boldsymbol{h}), \tag{12}$$

$$\begin{aligned}\text{with} \quad \Psi_{\mathrm{C}}(\boldsymbol{v},\boldsymbol{h}) &= \lambda\,\mathrm{TV}_\alpha(\boldsymbol{v},\boldsymbol{h}),\\ &= \lambda\left\|\left[\mathbf{D}\boldsymbol{v};\alpha\mathbf{D}\boldsymbol{h}\right]^\top\right\|_{2,1}\end{aligned} \tag{13}$$

where TV_α couples spatial variations of $v_{\underline{n}}$ and $h_{\underline{n}}$ and thus favor their occurrences at same location.

For both constructions, $\lambda > 0$ and $\alpha > 0$ constitute regularization hyperparameters that need to be selected.

Further, the thresholding procedure in [7, 8] is generalized to "*joint*" and "*coupled*" estimation strategies: Alg. 1, applied to $\widehat{\boldsymbol{h}}_{\mathrm{J}}$ (resp. $\widehat{\boldsymbol{h}}_{\mathrm{C}}$) provides "*joint*" (resp. "*coupled*") segmentation $\widehat{\boldsymbol{M}}_{\mathrm{J}}$ (resp. $\widehat{\boldsymbol{M}}_{\mathrm{C}}$).

Finally, from the two-region segmentation $\widehat{\boldsymbol{M}}_{\mathrm{J}}$ (resp. $\widehat{\boldsymbol{M}}_{\mathrm{C}}$) obtained with Alg. 1, one can use global techniques [47] to obtain *a posteriori* estimates of the Hölder exponents of each region $\widehat{H}_{0,\mathrm{J}}$ and $\widehat{H}_{1,\mathrm{J}}$ (resp. $\widehat{H}_{0,\mathrm{C}}$ and $\widehat{H}_{1,\mathrm{C}}$) (cf. Sec. 2.3.2).

3.2 Minimization algorithms

The data fidelity function Φ is Lipschitz differentiable. This is the case neither for Ψ_{J} nor for Ψ_{C}. Therefore, gradient descent methods may not be appropriate to solve (10) or (12). In contrast, proximal algorithms are often better adapted to the minimization of such nonsmooth functions [15]. While the nondifferentiability stems from the mixed norm $\|\cdot\|_{2,1}$, appearing both in Ψ_{J} and in Ψ_{C}, the linear operator $\mathbf{D}$ makes the computation of the proximal operator of total variation difficult [12, 37]. Below, two algorithms are devised: the dual FISTA and the primal-dual.

3.2.1 Dual FISTA

The well-known Fast Iterative Shrinkage Thresholding Algorihtm (FISTA) algorithm [9] is here customized to Problems (10) and (12), to achieve faster convergence than with the basic dual forward-backward. Corresponding iterations are detailed in Algorithms 2 and 3. Convergence guarantees are specified in Theorems 1 and 2.

Theorem 1 (Convergence of FISTA_J**)** *The sequence* $(\boldsymbol{v}^{[t]}, \boldsymbol{h}^{[t]})$ *generated by Algorithm 2 converges towards a solution of the joint estimation problem* (10)*.*

Theorem 2 (Convergence of FISTA_C**)** *The sequence* $(\boldsymbol{v}^{[t]}, \boldsymbol{h}^{[t]})$ *generated by Algorithm 3 converges towards a solution of coupled estimation problem* (12)*.*

Proof 1 *The proofs of Theorems 1 and 2 stem directly from the choice of descent parameters proposed in Algs. 2 and 3 following the reasoning in [9].*

Algorithm 2: FISTA_J: Joint estimation (Pb. (10))

Initialization: Set $\boldsymbol{u}^{[0]} \in \mathbb{R}^{2\times|\Upsilon|}$, $\bar{\boldsymbol{u}}^{[0]} = \boldsymbol{u}^{[0]}$;
Set $\boldsymbol{\ell}^{[0]} \in \mathbb{R}^{2\times|\Upsilon|}$, $\bar{\boldsymbol{\ell}}^{[0]} = \boldsymbol{\ell}^{[0]}$;
Let $\left(\mathcal{S}_{\underline{n}}, \mathcal{T}_{\underline{n}}\right)$ defined in (4);
Let $\mathbf{J}$ defined in (3);
Set $(\forall \underline{n})$ $\left(v_{\underline{n}}^{[0]}, h_{\underline{n}}^{[0]}\right)^\top = \mathbf{J}^{-1}\left(\mathcal{S}_{\underline{n}}, \mathcal{T}_{\underline{n}}\right)^\top$;
Set $b > 2$ and $\tau_0 = 1$;
Set $\alpha > 0$ and $\lambda > 0$;
Set $\gamma > 0$ such that $\gamma \|\mathbf{J}^{-1}\|\|\mathbf{D}\|^2 < 1$;

for $t \in \mathbb{N}$ **do**

Dual variable update:

$$\boldsymbol{u}^{[t+1]} = \text{prox}_{\gamma(\lambda\|\cdot\|_{2,1})^*}\left(\bar{\boldsymbol{u}}^{[t]} + \gamma\mathbf{D}\boldsymbol{v}^{[t]}\right)$$

$$\boldsymbol{\ell}^{[t+1]} = \text{prox}_{\gamma(\lambda\alpha\|\cdot\|_{2,1})^*}\left(\bar{\boldsymbol{\ell}}^{[t]} + \gamma\mathbf{D}\boldsymbol{h}^{[t]}\right)$$

FISTA parameter update

$$\tau_{t+1} = \frac{t+b}{b}$$

Auxiliary variable update

$$\bar{\boldsymbol{u}}^{[t+1]} = \boldsymbol{u}^{[t+1]} + \frac{\tau_t - 1}{\tau_{t+1}}\left(\boldsymbol{u}^{[t+1]} - \boldsymbol{u}^{[t]}\right)$$

$$\bar{\boldsymbol{\ell}}^{[t+1]} = \boldsymbol{\ell}^{[t+1]} + \frac{\tau_t - 1}{\tau_{t+1}}\left(\boldsymbol{\ell}^{[t+1]} - \boldsymbol{\ell}^{[t]}\right)$$

Primal variable update

$$\begin{pmatrix}\boldsymbol{v}^{[t+1]} \\ \boldsymbol{h}^{[t+1]}\end{pmatrix} = \begin{pmatrix}\boldsymbol{v}^{[t]} \\ \boldsymbol{h}^{[t]}\end{pmatrix} - \mathbf{J}^{-1}\begin{pmatrix}\mathbf{D}^*(\boldsymbol{u}^{[t+1]} - \boldsymbol{u}^{[t]}) \\ \mathbf{D}^*(\boldsymbol{\ell}^{[t+1]} - \boldsymbol{\ell}^{[t]})\end{pmatrix}$$

3.2.2 Primal-dual

Primal-dual algorithms [10, 16, 25, 44] can also be customized to solve Problems (10) and (12).

This requires the derivation of a closed-form expression for the proximal operator associated with the quadratic data fidelity term Φ, provided in Proposition 1.

Algorithm 3: FISTA$_{\mathrm{C}}$: Coupled estimation (Pb. (12))

Initialization:
Set $\boldsymbol{u}^{[0]} \in \mathbb{R}^{2\times|\Upsilon|}$, $\bar{\boldsymbol{u}}^{[0]} = \boldsymbol{u}^{[0]}$;
Set $\boldsymbol{\ell}^{[0]} \in \mathbb{R}^{2\times|\Upsilon|}$, $\bar{\boldsymbol{\ell}}^{[0]} = \boldsymbol{\ell}^{[0]}$;
Let $\left(\mathcal{S}_{\underline{n}}, \mathcal{T}_{\underline{n}}\right)$ defined in (4);
Let $\mathbf{J}$ defined in (3);
Set $(\forall \underline{n})$ $\left(v_{\underline{n}}^{[0]}, h_{\underline{n}}^{[0]}\right)^{\top} = \mathbf{J}^{-1}\left(\mathcal{S}_{\underline{n}}, \mathcal{T}_{\underline{n}}\right)^{\top}$;
Set $b > 2$ and $\tau_0 = 1$;
Set $\alpha > 0$ and $\lambda > 0$;
Set $\gamma > 0$ s. t. $\gamma \max(1,\alpha)\,\|\mathbf{J}^{-1}\|\|\mathbf{D}\|^2 < 1$;

for $t \in \mathbb{N}$ **do**

Dual variable update:

$$\begin{pmatrix}\boldsymbol{u}^{[t+1]}\\ \boldsymbol{\ell}^{[t+1]}\end{pmatrix} = \mathrm{prox}_{\gamma(\lambda\|\cdot\|_{2,1})^*}\begin{pmatrix}\bar{\boldsymbol{u}}^{[t]} + \gamma\mathbf{D}\boldsymbol{v}^{[t]}\\ \bar{\boldsymbol{\ell}}^{[t]} + \gamma\alpha\mathbf{D}\boldsymbol{h}^{[t]}\end{pmatrix}$$

FISTA parameter update

$$\tau_{t+1} = \tfrac{t+b}{b}$$

Auxiliary variable update

$$\bar{\boldsymbol{u}}^{[t+1]} = \boldsymbol{u}^{[t+1]} + \tfrac{\tau_t - 1}{\tau_{t+1}}\left(\boldsymbol{u}^{[t+1]} - \boldsymbol{u}^{[t]}\right)$$

$$\bar{\boldsymbol{\ell}}^{[t+1]} = \boldsymbol{\ell}^{[t+1]} + \tfrac{\tau_t - 1}{\tau_{t+1}}\left(\boldsymbol{\ell}^{[t+1]} - \boldsymbol{\ell}^{[t]}\right)$$

Primal variable update

$$\begin{pmatrix}\boldsymbol{v}^{[t+1]}\\ \boldsymbol{h}^{[t+1]}\end{pmatrix} = \begin{pmatrix}\boldsymbol{v}^{[t]}\\ \boldsymbol{h}^{[t]}\end{pmatrix} - \mathbf{J}^{-1}\begin{pmatrix}\mathbf{D}^*(\boldsymbol{u}^{[t+1]} - \boldsymbol{u}^{[t]})\\ \alpha\mathbf{D}^*(\boldsymbol{\ell}^{[t+1]} - \boldsymbol{\ell}^{[t]})\end{pmatrix}$$

Proposition 1 (Computation of prox_{Φ}) *For every $(\boldsymbol{v}, \boldsymbol{h}) \in \mathbb{R}^{|\Upsilon|} \times \mathbb{R}^{|\Upsilon|}$, denoting $(\boldsymbol{p}, \boldsymbol{q}) = \mathrm{prox}_{\Phi}(\boldsymbol{v}, \boldsymbol{h}) \in \mathbb{R}^{|\Upsilon|} \times \mathbb{R}^{|\Upsilon|}$ one has*

$$\begin{cases} \boldsymbol{p} = \frac{(1+R_2)(\boldsymbol{\mathcal{S}}+\boldsymbol{v})-R_1(\boldsymbol{\mathcal{T}}+\boldsymbol{h})}{(1+R_0)(1+R_2)-R_1^2}, \\ \boldsymbol{q} = \frac{(1+R_0)(\boldsymbol{\mathcal{T}}+\boldsymbol{h})-R_1(\boldsymbol{\mathcal{S}}+\boldsymbol{v})}{(1+R_0)(1+R_2)-R_1^2}. \end{cases}$$

with $\boldsymbol{\mathcal{S}}$ and $\boldsymbol{\mathcal{T}}$ defined in (4) *and R_0, R_1, R_2 defined in* (3).

Further, in [10], it was described how to take advantage of the strong convexity[1] of the objective function to obtain fast implementation for primal-dual algorithms, with linear convergence rates. In Proposition 2 below, we prove that the data fidelity term Φ is strongly convex and derive the closed form expression of the strong convexity coefficient.

Proposition 2 *Function $\Phi(\boldsymbol{v}, \boldsymbol{h}; \mathcal{L})$ is μ-strongly convex w.r.t the variables $(\boldsymbol{v}, \boldsymbol{h})$, with $\mu = \chi$ where $\chi > 0$ is the lowest eigenvalue of the symmetric and positive definite matrix $\boldsymbol{J}$ defined in Eq.* (3).

Indeed, since $\nabla\Phi(\boldsymbol{v}, \boldsymbol{h}; \mathcal{L}) = \mathbf{J}(\boldsymbol{v}, \boldsymbol{h})^{\top}$ is linear, the condition for Φ to be $\mu-$strongly convex can be recast as:

$$\begin{aligned} \langle \nabla\Phi(\boldsymbol{v}, \boldsymbol{h}; \mathcal{L}), (\boldsymbol{v}, \boldsymbol{h}) \rangle &= \left\langle \mathbf{J}(\boldsymbol{v}, \boldsymbol{h})^{\top}, (\boldsymbol{v}, \boldsymbol{h})^{\top} \right\rangle \\ &\geq \mu \left\| (\boldsymbol{v}, \boldsymbol{h}) \right\|^2 . \end{aligned} \tag{14}$$

Intuitively, a function with a large strong-convexity constant μ is very peaked around its minimum, hence yielding a good theoretical convergence rate toward the minimizer. The strong-convexity parameter is represented in Fig. 2 as a function of the range of scales involved in the estimation procedure (cf. (9)).

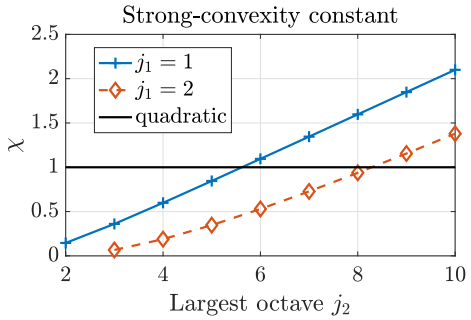


Figure 2: **Strong convexity constant** as a function of the range of scales involved in the estimation.

The iterations of the devised fast primal-dual algorithms are detailed in Algorithms 4 and 5. Convergence guarantees are specified in Theorem 3 (resp. Theorem 4).

[1] A function $\varphi : \mathcal{H} \to \mathbb{R}$, defined on an Hilbert space $\mathcal{H}$, is said to be $\mu-$strongly convex, for a given $\mu > 0$, if the function $y \mapsto \varphi(y) - \frac{\mu}{2}\|y\|_2^2$ is convex. When the function φ is differentiable, φ is $\mu-$strongly convex if and only if $(\forall (y, z) \in \mathcal{H} \times \mathcal{H}) \quad \langle \nabla\varphi(y) - \nabla\varphi(z), y - z \rangle \geq \mu \|y - z\|^2$ where $\langle \cdot, \cdot \rangle$ denotes the Hilbert scalar product and $\|\cdot\|$ the associated scalar product.

Theorem 3 (Convergence of $\mathrm{PD_J}$) *The sequence $(\boldsymbol{v}^{[\ell]}, \boldsymbol{h}^{[\ell]})$ generated by Algorithm 4 converges towards a solution of the joint estimation problem* (10).

Theorem 4 (Convergence of $\mathrm{PD_C}$) *The sequence $(\boldsymbol{v}^{[\ell]}, \boldsymbol{h}^{[\ell]})$ generated by Algorithm 5 converges towards a solution of the joint estimation problem* (12).

Proof 2 *The proofs of Theorems 3 and 4 come directly from the choice of descent parameters proposed in Algs. 4 and 5 following the reasoning of [10].*

Algorithm 4: $\mathrm{PD_J}$: Joint estimation (Pb. (10))

Initialization: Set $\boldsymbol{v}^{[0]} \in \mathbb{R}^{|\Upsilon|}$, $\boldsymbol{u}^{[0]} = \mathbf{D}\boldsymbol{v}^{[0]}$, $\bar{\boldsymbol{u}}^{[0]} = \boldsymbol{u}^{[0]}$;
Set $\boldsymbol{h}^{[0]} \in \mathbb{R}^{|\Upsilon|}$, $\boldsymbol{\ell}^{[0]} = \mathbf{D}\boldsymbol{h}^{[0]}$, $\bar{\boldsymbol{\ell}}^{[0]} = \boldsymbol{\ell}^{[0]}$;
Set $\alpha > 0$ and $\lambda > 0$;
Set (δ_0, ν_0) such that $\delta_0 \nu_0 \|\mathbf{D}\|^2 < 1$;

for $t \in \mathbb{N}^*$ **do**

Primal variable update:

$$\begin{pmatrix} \boldsymbol{v}^{[t+1]} \\ \boldsymbol{h}^{[t+1]} \end{pmatrix} = \mathrm{prox}_{\delta_t \Phi}\left(\begin{pmatrix} \boldsymbol{v}^{[t]} \\ \boldsymbol{h}^{[t]} \end{pmatrix} - \delta_t \begin{pmatrix} \mathbf{D}^* \bar{\boldsymbol{u}}^{[t]} \\ \mathbf{D}^* \bar{\boldsymbol{\ell}}^{[t]} \end{pmatrix}\right)$$

Dual variable update:

$$\boldsymbol{u}^{[t+1]} = \mathrm{prox}_{\nu_t (\lambda \|\cdot\|_{2,1})^*}\left(\boldsymbol{u}^{[t]} + \nu_t \mathbf{D}\boldsymbol{v}^{[t]}\right)$$

$$\boldsymbol{\ell}^{[t+1]} = \mathrm{prox}_{\nu_t (\lambda\alpha \|\cdot\|_{2,1})^*}\left(\boldsymbol{\ell}^{[t]} + \nu_t \mathbf{D}\boldsymbol{h}^{[t]}\right)$$

Descent steps update:

$\vartheta_t = (1 + 2\mu\delta_t)^{-1/2}$, $\delta_{t+1} = \vartheta_t \delta_t$, $\nu_{t+1} = \nu_t / \vartheta_t$

Auxiliary variable update:

$$\begin{pmatrix} \bar{\boldsymbol{u}}^{[t+1]} \\ \bar{\boldsymbol{\ell}}^{[t+1]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} + \vartheta_t \left(\begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} - \begin{pmatrix} \boldsymbol{u}^{[t]} \\ \boldsymbol{\ell}^{[t]} \end{pmatrix}\right)$$

3.3 Duality gap

To ensure fair comparisons between the different algorithms, we construct a stopping criterion based on the *duality gap*. Let $\mathcal{H}, \mathcal{G}$ Hilbert spaces, $\Theta : \mathcal{H} \to]-\infty, +\infty]$, $\Xi : \mathcal{G} \to]-\infty, +\infty]$ and $\mathbf{L} : \mathcal{H} \to \mathcal{G}$ a bounded linear operator. From a primal optimization problem of the general form $\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x} \in \mathcal{H}} \Theta(\boldsymbol{x}) + \Xi(\mathbf{L}\boldsymbol{x})$, and its associated dual problem $\widehat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y} \in \mathcal{G}} \Theta^*(-\mathbf{L}^*\boldsymbol{y}) + \Xi^*(\boldsymbol{y})$, we define:

$$\Gamma(\boldsymbol{x}; \boldsymbol{y}) := \Theta(\boldsymbol{x}) + \Xi(\mathbf{L}\boldsymbol{x}) + \Theta^*(-\mathbf{L}^*\boldsymbol{y}) + \Xi^*(\boldsymbol{y}), \tag{15}$$

Θ^* (resp. Ξ^*), being the Fenchel conjugate of Θ (resp. Ξ).

Definition 2 *The* duality gap $\delta\Gamma$ *associated to primal and dual optimization problems is defined as the infimum:*

$$\delta\Gamma := \inf_{(\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{H} \times \mathcal{G}} \Gamma(\boldsymbol{x}; \boldsymbol{y}). \tag{16}$$

Algorithm 5: $\mathrm{PD_C}$: Coupled estimation (Pb. (12))

Initialization:

Set $\boldsymbol{v}^{[0]} \in \mathbb{R}^{|\Upsilon|}$, $\boldsymbol{u}^{[0]} = \mathbf{D}\boldsymbol{v}^{[0]}$, $\bar{\boldsymbol{u}}^{[0]} = \boldsymbol{u}^{[0]}$;
Set $\boldsymbol{h}^{[0]} \in \mathbb{R}^{|\Upsilon|}$, $\boldsymbol{\ell}^{[0]} = \alpha\mathbf{D}\boldsymbol{h}^{[0]}$, $\bar{\boldsymbol{\ell}}^{[0]} = \boldsymbol{\ell}^{[0]}$;
Set $\alpha > 0$ and $\lambda > 0$.
Set (δ_0, ν_0) such that $\delta_0\nu_0 \max(1, \alpha)\|\mathbf{D}\|^2 < 1$;

for $t \in \mathbb{N}^*$ **do**

Primal variable update:

$$\begin{pmatrix} \boldsymbol{v}^{[t+1]} \\ \boldsymbol{h}^{[t+1]} \end{pmatrix} = \mathrm{prox}_{\delta_t \Phi}\left(\begin{pmatrix} \boldsymbol{v}^{[t]} \\ \boldsymbol{h}^{[t]} \end{pmatrix} - \delta_t \begin{pmatrix} \mathbf{D}^*\bar{\boldsymbol{u}}^{[t]} \\ \alpha\mathbf{D}^*\bar{\boldsymbol{\ell}}^{[t]} \end{pmatrix}\right)$$

Dual variable update:

$$\begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} = \mathrm{prox}_{\nu_t(\lambda\|.\|_{2,1})^*}\begin{pmatrix} \boldsymbol{u}^{[t]} + \nu_t\mathbf{D}\boldsymbol{v}^{[t]} \\ \boldsymbol{\ell}^{[t]} + \nu_t\alpha\mathbf{D}\boldsymbol{h}^{[t]} \end{pmatrix}$$

Descent steps update:

$\vartheta_t = (1 + 2\mu\delta_t)^{-1/2}$, $\delta_{t+1} = \vartheta_t\delta_t$, $\nu_{t+1} = \nu_t/\vartheta_t$

Auxiliary variable update:

$$\begin{pmatrix} \bar{\boldsymbol{u}}^{[t+1]} \\ \bar{\boldsymbol{\ell}}^{[t+1]} \end{pmatrix} = \begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} + \vartheta_t\left(\begin{pmatrix} \boldsymbol{u}^{[t+1]} \\ \boldsymbol{\ell}^{[t+1]} \end{pmatrix} - \begin{pmatrix} \boldsymbol{u}^{[t]} \\ \boldsymbol{\ell}^{[t]} \end{pmatrix}\right)$$

Under some loose assumptions on Θ, Ξ and $\mathbf{L}$, referred as *strong duality* in [5], the duality gap $\delta\Gamma$ is zero and the respective solutions $\widehat{\boldsymbol{x}}$ and $\widehat{\boldsymbol{y}}$ of primal and dual problems are characterized by $\Gamma(\widehat{\boldsymbol{x}}; \widehat{\boldsymbol{y}}) = \delta\Gamma = 0$.

Primal-dual algorithms presented above built a minimizing sequence $(\boldsymbol{x}^{[t]}, \boldsymbol{y}^{[t]})_{t\in\mathbb{N}}$ for Γ, therefore converging towards the unique point achieving the infimum of Eq. (16): $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{y}})$. Then the convergence of $\boldsymbol{x}^{[t]}$ (resp. $\boldsymbol{y}^{[t]}$) towards $\widehat{\boldsymbol{x}}$ (resp. $\widehat{\boldsymbol{y}}$) can be measured evaluating the primal-dual functional Γ, as:

$$\Gamma(\boldsymbol{x}^{[t]}; \boldsymbol{y}^{[t]}) \underset{t\to\infty}{\longrightarrow} 0. \tag{17}$$

The "*joint*" (10) and "*coupled*" (12) optimization problems, with primal variable $\boldsymbol{x} = (\boldsymbol{v}, \boldsymbol{h}) \in \mathbb{R}^{|\Upsilon|} \times \mathbb{R}^{|\Upsilon|}$, share the same data fidelity term $\Theta(\boldsymbol{x}) = \Phi(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{\mathcal{L}})$ and a penalization based on the mixed norm $\Xi = \lambda\|\cdot\|_{2,1}$. Yet, the linear operator $\mathbf{L}$ differs between *joint* and *coupled* formulations, as:
i) $\mathbf{L}\boldsymbol{x} = [\mathbf{D}\boldsymbol{v}, \alpha\mathbf{D}\boldsymbol{h}] \in \mathbb{R}^{2\times 2|\Upsilon|}$, for "*joint*" (10),
ii) $\mathbf{L}\boldsymbol{x} = [\mathbf{D}\boldsymbol{v}; \alpha\mathbf{D}\boldsymbol{h}] \in \mathbb{R}^{4\times|\Upsilon|}$, for "*coupled*" (12).
For Pb. (10) (resp. (12)), denoting $\boldsymbol{y} = (\boldsymbol{u}, \boldsymbol{\ell})$ the dual variable, the evaluation of $\Gamma_{\mathrm{J\,(resp.\,C)}}(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{u}, \boldsymbol{\ell})$ requires computing:
i) $\mathbf{L}^*$ (straightforward from $\mathbf{D}^*$),
ii) $\Xi^* = (\lambda\|\cdot\|_{2,1})^* = \iota_{\|\cdot\|_{2,+\infty}\leq\lambda}$ (direct computation),
iii) $\Theta^* = \Phi^*$ which is devised in Proposition 3.

Proposition 3 *Let $\boldsymbol{J}$ and $(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}})$ defined in* (3) *and* (4)*,*

$$\Phi^*(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{\mathcal{L}}) = \frac{1}{2}\langle (\boldsymbol{v}, \boldsymbol{h})^\top, \boldsymbol{J}^{-1}(\boldsymbol{v}, \boldsymbol{h})^\top \rangle + \langle (\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}})^\top, \boldsymbol{J}^{-1}(\boldsymbol{v}, \boldsymbol{h})^\top \rangle + \boldsymbol{\mathcal{C}}, \tag{18}$$
$$\boldsymbol{\mathcal{C}} = \frac{1}{2}\langle (\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}})^\top, \boldsymbol{J}^{-1}(\boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{T}})^\top \rangle - \frac{1}{2}\sum_j (\log_2 \boldsymbol{\mathcal{L}}_j)^2.$$

Proof 3 *By definition of the Fenchel conjugate,*

$$F^*(\boldsymbol{v}, \boldsymbol{h}; \boldsymbol{\mathcal{L}}) = \sup_{\widetilde{\boldsymbol{v}} \in \mathbb{R}^{|\Upsilon|}, \widetilde{\boldsymbol{h}} \in \mathbb{R}^{|\Upsilon|}} \langle \widetilde{\boldsymbol{v}}, \boldsymbol{v} \rangle + \langle \widetilde{\boldsymbol{h}}, \boldsymbol{h} \rangle - F(\widetilde{\boldsymbol{v}}, \widetilde{\boldsymbol{h}}; \boldsymbol{\mathcal{L}}). \tag{19}$$

The supremum is obtained at $(\bar{\boldsymbol{v}}, \bar{\boldsymbol{h}})$ such that, for every $\underline{n} \in \Omega$,

$$\begin{cases} v_{\underline{n}} - \sum_j \left(\bar{v}_{\underline{n}} + j\bar{h}_{\underline{n}} - \log_2 \mathcal{L}_{j,\underline{n}} \right) = 0 \\ h_{\underline{n}} - \sum_j j \left(\bar{v}_{\underline{n}} + j\bar{h}_{\underline{n}} - \log_2 \mathcal{L}_{j,\underline{n}} \right) = 0. \end{cases} \tag{20}$$

or equivalently,

$$\begin{cases} R_0 \bar{v}_{\underline{n}} + R_1 \bar{h}_{\underline{n}} = v_{\underline{n}} + \mathcal{S}_{\underline{n}} \\ R_1 \bar{v}_{\underline{n}} + R_2 \bar{h}_{\underline{n}} = h_{\underline{n}} + \mathcal{T}_{\underline{n}} \end{cases} \tag{21}$$

that yields

$$\begin{pmatrix} \bar{v}_{\underline{n}} \\ \bar{h}_{\underline{n}} \end{pmatrix} = \boldsymbol{J}^{-1} \begin{pmatrix} v_{\underline{n}} + \mathcal{S}_{\underline{n}} \\ h_{\underline{n}} + \mathcal{T}_{\underline{n}} \end{pmatrix} \tag{22}$$

and it is then necessary to substitute this expression into the explicit expression of Ψ.

4 Piecewise homogeneous fractal textures

Homogeneous fractal textures. Numerous models of homogeneous textures have been proposed in the literature, mostly consisting of 2D extensions of fractional Brownian motion, the unique univariate Gaussian exactly selfsimilar process with stationary increments [14]. These extensions are often referred to as fractional Brownian field fBf. However, in most applications, real-world textures are better modeled by stationary processes, hence by increments of fBf, referred to as fractional Gaussian fields. Further, to construct a formally relevant definition for piecewise homogeneous fractal textures, as originally proposed here, it is easier to *manipulate* stationary processes.

For Gaussian processes, space or frequency domain definitions of fBf are theoretically equivalent and several variations were proposed in either domain, mostly consisting in different tuning of the 2D extension of the fractional integration kernel underlying selfsimilar textures (cf. e.g., [2, 14, 36]). Yet, in practice, numerical issues need to be accounted for, such as discrete sampling from a continuous process, or computation of integrals defined from a theoretically infinite support kernel and related border effects. Further, fast circulant embedding matrix algorithms were developed ; but they do not permit to reach the full range $0 < H < 1$ for the selfsimilarity parameter [43]. Therefore, we will make use of a self-customized construction that combines an effective implementation scheme with excellent theoretical and practical control of the local

variance and regularity, while permitting direct extension to piecewise homogeneous construction.

Following e.g., [6, 17, 41], we start from an harmonizable representation of fBf $B_{\underline{n}}$, $\underline{n} \in \mathbb{R}^{|\Upsilon|}$, $\Upsilon = \{1, \ldots, N\}^2$:

$$B_{\underline{n}} = \frac{\Sigma}{C(H)} \int \frac{\mathrm{e}^{-\mathrm{i}\underline{f}\cdot\underline{n}} - 1}{\|\underline{f}\|^{H+1}} \mathrm{d}\widetilde{G}(\underline{f}), \tag{23}$$

with $\mathrm{d}\widetilde{G}(\underline{f})$ the Fourier transform of a white Gaussian noise and

$$C(H) = \frac{\pi^{1/2}\Gamma(H+1/2)}{2^{d/2}H\Gamma(2H)\sin(\pi H)\Gamma(H+d/2)}.$$

The covariance function of the zero mean process $B_{\underline{n}}$ reads:

$$\mathbb{E}\left[B_{\underline{n}}B_{\underline{m}}\right] = \frac{\Sigma^2}{2}\left(\|\underline{n}\|^{2H} + \|\underline{m}\|^{2H} - \|\underline{n} - \underline{m}\|^{2H}\right). \tag{24}$$

Texture $Y_{\underline{n}}$ is constructed from increments of $B_{\underline{n}}$ as (with $\mathbf{e}_1$, $\mathbf{e}_2$, unitary vectors in horizontal and vertical directions):

$$Y_{\underline{n}} = \frac{\Sigma}{2\delta^H\sqrt{1-2^{H-2}}}\Big(\underbrace{B_{\underline{n}+\delta\mathbf{e}_1} - B_{\underline{n}}}_{\text{horizontal increment}} + \underbrace{B_{\underline{n}+\delta\mathbf{e}_2} - B_{\underline{n}}}_{\text{vertical increment}} \Big), \tag{25}$$

Proposition 4 *The field $Y = (Y_{\underline{n}})_{\underline{n}\in\Upsilon}$ is a zero-mean Gaussian process with variance $\mathbb{E}\left[Y_{\underline{n}}^2\right] = \Sigma^2$ and covariance:*

$$\begin{aligned}
&\mathbb{E}\left[Y_{\underline{n}+\Delta\underline{n}}Y_{\underline{n}}\right] = \frac{\Sigma^2\delta^{-2H}}{4-2^H}(\|\Delta\underline{n} + \delta\boldsymbol{e}_1\|^{2H} \\
&+ \|\Delta\underline{n} - \delta\boldsymbol{e}_1\|^{2H} + \|\Delta\underline{n} + \delta\boldsymbol{e}_2\|^{2H} + \|\Delta\underline{n} - \delta\boldsymbol{e}_2\|^{2H} - 3\|\Delta\underline{n}\|^{2H} \\
&- \frac{1}{2}\|\Delta\underline{n} + \delta\boldsymbol{e}_1 - \delta\boldsymbol{e}_2\|^{2H} - \frac{1}{2}\|\Delta\underline{n} - \delta\boldsymbol{e}_1 + \delta\boldsymbol{e}_2\|^{2H}).
\end{aligned}$$

Piecewise homogeneous fractal textures. Piecewise homogeneous fractal textures $X = (X_{\underline{n}})_{\underline{n}\in\Upsilon}$ are defined as the concatenation of M (pairwise disjoint) regions, denoted $(\Upsilon_m)_{0\le m\le M-1}$, such that $\Upsilon = \cup_m \Upsilon_m$, with textures in each region consisting of an homogeneous fractal $Y_{\underline{n}}^{(m)}$ $(0 \le m \le M-1)$, with variance and regularity (Σ_m^2, H_m):

$$X_{\underline{n}} = Y_{\underline{n}}^{(m)}, \text{ when } \underline{n} \in \Upsilon_m.$$

Examples of piecewise fractal textures are shown in Fig. 3.

5 Estimation/Segmentation performance

5.1 Monte-Carlo simulation set-up

The proposed *joint* and *coupled* segmentations are now compared in terms of performance and computational costs to the state-of-the-art T-ROF procedure, in the context of a two-region segmentation, by means of Monte-Carlo simulations.

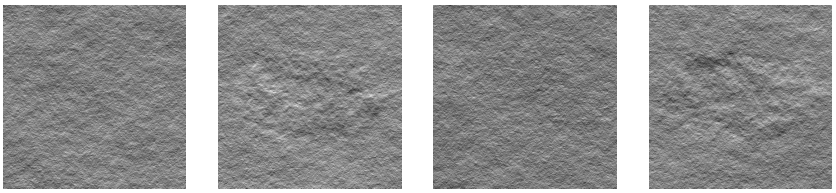


Figure 3: **Piecewise homogeneous fractal textures** X generated using the mask displayed in Fig. 1a, with parameters $(H_0, \Sigma_0) = (0.5, 0.6)$ and $(H_1, \Sigma_1))$ chosen as: (a) $(0.5, 0.6)$, no change ; (b) $(0.5, 0.65)$, change in Σ ; (c) $(0.8, 0.6)$ change in H ; (d) $(0.8, 0.65)$, change in Σ and H.

5.1.1 Synthetic textures

Piecewise homogeneous fractal textures are generated as defined in Section 4, using the mask shown in Fig. 1a, with $N = 512$. They consist of two regions: i) a background, with variance and local regularity $\left(\Sigma_0^2, H_0\right) = (0.6, 0.5)$ kept fixed, ii) a central ellipse, for which variance and local regularity $\left(\Sigma_1^2, H_1\right)$ are varied, as illustrated in Fig. 5. For each configuration, 5 realizations of piecewise homogeneous fractal textures are generated and analyzed ; performance is reported as averages over realizations.

5.1.2 Wavelet transform

2D-Wavelet decompositions are performed using tensor product wavelets constructed from 1D-Daubechies orthonormal least asymmetric wavelets with $N_\psi = 3$ vanishing moments [27]. Wavelet leaders are computed as in Section 2.2. Local estimate $\widehat{\boldsymbol{h}}_{\mathrm{LR}}$ in T-ROF is computed as in Sec. 2.3. Estimates from all procedures involve octaves $(j_1, j_2) = (2, 5)$. Octave $j_1 = 1$ is a priori excluded as leaders are biased [45, 47].

5.1.3 Optimization algorithm parameters

To achieve best convergence of the optimization schemes, descent steps are chosen as large as permitted:

- Alg. 2 ($\mathrm{FISTA_J}$): $\gamma = 0.99/\left(\|\mathbf{J}^{-1}\|\|\mathbf{D}\|^2\right)$,
- Alg. 3 ($\mathrm{FISTA_C}$): $\gamma = 0.99/\left(\max(1,\alpha)\,\|\mathbf{J}^{-1}\|\|\mathbf{D}\|^2\right)$,
- Alg. 4 ($\mathrm{PD_J}$): $\delta_0 = \nu_0 = 0.99/\|\mathbf{D}\|$.
- Alg. 5 ($\mathrm{PD_C}$): $\delta_0 = \nu_0 = 0.99/\left(\max(1,\sqrt{\alpha})\|\mathbf{D}\|\right)$,

where $\|\mathbf{D}\| = 2\sqrt{2}$, with $\mathbf{D}$ defined in Eq. (7), and where $\|\mathbf{J}^{-1}\|$ depends on the octaves involved in estimation ($\mathbf{J}$ defined in Eq. (3)): With $(j_1, j_2) = (2, 5)$, $\|\mathbf{J}^{-1}\| \simeq 2.88$. For FISTA, the inertia parameter is set to $b = 4$.

5.2 Issues in performance and algorithm comparison

5.2.1 Stopping criteria for proximal algorithms

To ensure fair comparisons either between algorithms or between *joint* and *coupled* formulations (Pb. (10) vs. Pb. (12)), an effective stopping criterion is needed. Fig. 4 illustrates the convergence toward zero of two potential candidates, for the particular case, chosen as representative example, of the *joint* estimation using primal-dual Alg. 4, and for several choices of hyperparameters.

First, the normalized increments of the objective function, as a function of the number of iterations, are presented in Fig. 4a ; they decrease with an extremely irregular behavior, which makes them ill-suited to serve as effective stopping criterion.

Second, the *normalized* primal-dual functional (with notations as in Sec. 3.3)

$$\widetilde{\Gamma}^{[t]} := \frac{\Gamma^{[t]}}{\left|\Theta(\boldsymbol{x}^{[t]}) + \Xi(\mathbf{L}\boldsymbol{x}^{[t]})\right| + \left|\Theta^*(-\mathbf{L}^*\boldsymbol{y}^{[t]}) + \Xi^*(\boldsymbol{y}^{[t]})\right|},$$

is presented in Fig. 4b, where it is observed to decrease smoothly. Systematic inspections of such decreases together with that of the corresponding achieved solutions lead us to devise a stopping criterion, which is effective both for primal-dual and forward-backward algorithms and for *joint* and *coupled* estimations, as,

$$\widetilde{\Gamma}_{\bullet}^{[t]} < 5.10^{-3}, \text{ for } \bullet \in \{\text{ROF, J}\}, \text{ and } \widetilde{\Gamma}_{\text{C}}^{[t]} < 10^{-4}. \tag{26}$$

The use of the normalized quantities, $\widetilde{\Gamma}^{[t]}$, makes this stopping criterion robust to variations of the hyperparameters. For practical purposes, we also impose an upper limit on the number of iterations: $t < 2.5\,10^5$.

5.2.2 Choice of regularization hyperparameters

The choice of regularization parameters (λ, α) appearing in Pb (5), (10), (12) is of prime importance as λ tunes the trade-off between fidelity to the fractal model (1) and expected piecewise constancy, while α controls the relative weight given to local wavelet log-variance $\boldsymbol{v}$ compared to local regularity $\boldsymbol{h}$, in the (*joint* (11) and *coupled* (13)) total variation penalization. The automated choice of the regularization parameters is a difficult issue, beyond the scope of the present work. In this study, a grid search strategy is used to find the parameters λ and α achieving the best segmentation. In practice, logarithmically spaced ranges are used, from 10^{-1} to 10^3 for λ and from 10^{-2} to 10^3 for α.

5.2.3 Performance assessment

A natural performance criterion consists in comparing the achieved classification, denoted $\widehat{\boldsymbol{M}}_{\text{ROF}}$, $\widehat{\boldsymbol{M}}_{\text{J}}$ and $\widehat{\boldsymbol{M}}_{\text{C}}$ respectively for the three segmentation procedures compared here, (ROF, *joint* and *coupled*), to the mask in Fig 1a, regarded as ground truth. It leads to define the *classification score* as the percentage of correctly labeled pixels. Classification scores for $\widehat{\boldsymbol{M}}_{\text{ROF}}$, $\widehat{\boldsymbol{M}}_{\text{J}}$ and $\widehat{\boldsymbol{M}}_{\text{C}}$, applied to different configurations of piecewise fractal textures are reported in Table 1, together with the difference between the a posteriori global

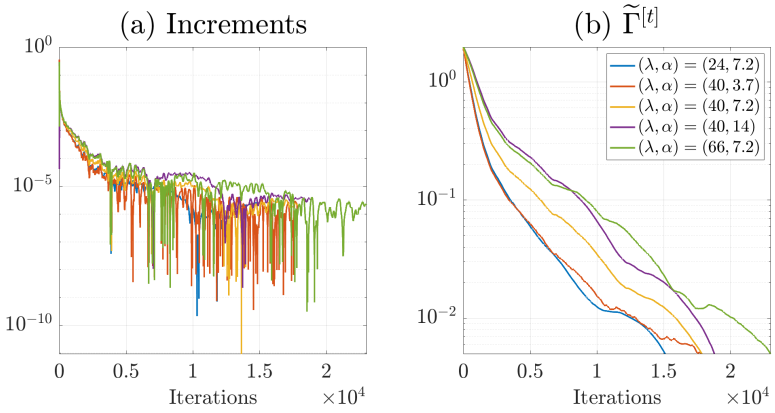


Figure 4: **Stopping criteria for proximal algorithms.** Convergence of Alg. 4 solving *joint* problem (10) for five pairs of hyperparameters (λ, α) evaluated with: (a) (Normalized) increments of objective function, (b) Normalized primal-dual functional $\widetilde{\Gamma}_{\mathrm{J}}^{[t]}$.

estimates obtained for each segmented regions Υ_0 and Υ_1 (cf. Section 2.3.3):

$$\widehat{\Delta H}_\bullet := \widehat{H}_{1,\bullet} - \widehat{H}_{0,\bullet}, \text{ for } \bullet \in \{\mathrm{ROF}, \mathrm{J}, \mathrm{C}\}.$$

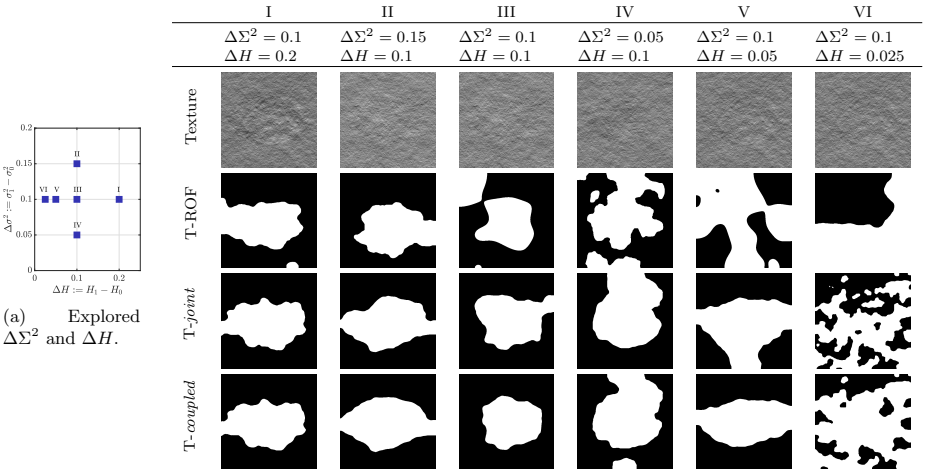
5.3 Performance comparisons

5.3.1 Segmentation and estimation performance

Fig. 5 and Table 1 report segmentation and estimation performance for 7 different configurations and for the optimal set of hyperparameters (i.e., those that maximize the classification scores).

Configurations I, III, V, VI correspond to a decrease in the difference between the regularity of each region of the piecewise fractal texture: $\Delta H = H_1 - H_0$, (hence to an increase in difficulty) for a fixed $\Delta\Sigma^2 = 0.1$. While the segmentation performance of the three procedures (T-ROF, T-*joint* and T-*coupled*) are comparable for *easy* configuration, those of T-ROF decrease drastically when ΔH decreases while those of T-*joint* and T-*coupled* decrease significantly less. Along the same line, the estimation of ΔH remains more satisfactory at small ΔH for T-*joint* and T-*coupled* than for T-ROF. It can also be observed that the performance of T-*coupled* degrade slightly less than those T-*joint*.

Configurations II, III, IV correspond to a decrease in variance, $\Delta\Sigma^2$, (hence to an increase in difficulty) for a fixed $\Delta H = 0.1$ which can already be regarded as a difficult case. As expected, T-ROF is not helped by the increase of variance between IV and II as estimation of local regularity does not depend on



(a) Explored $\Delta\Sigma^2$ and ΔH.

Figure 5: **Compared optimal segmentation.** Piecewise fractal textures are characterized by $\Sigma_0^2 = 0.6$, $H_0 = 0.5$ and different $\Delta\Sigma^2$, ΔH as sketched in Fig. 1a. First row: T-ROF segmentation $\widehat{\boldsymbol{M}}_{\mathrm{ROF}}$. Second row: T-*joint* segmentation $\widehat{\boldsymbol{M}}_{\mathrm{J}}$. Third row: T-*coupled* segmentation $\widehat{\boldsymbol{M}}_{\mathrm{C}}$.

variance [38, 47], and T-ROF segmentation results are not satisfactory. It can also be observed that performance of T-*joint* and T-*coupled* improve when $\Delta\Sigma^2$ increases, and, again, that the improvement is slightly larger for T-*coupled* than for T-*joint*.

These results permit to draw two clear conclusions. First, there are quantifiable benefits in using the side information brought by $\Delta\Sigma^2$, notably when the changes in regularity become small (low ΔH): T-*joint* and T-*coupled* outperform T-ROF. Second, T-*coupled* –which in principle favors co-localized changes in regularity and variance– shows overall better performance than T-*joint* – which does not favor co-localized changes. This is a satisfactory outcome, as all the configurations chosen follow the *a priori* intuition, relevant for real-world applications, that changes of textures naturally imply co-localized changes in local variance and local regularity.

5.3.2 Computational costs

Comparisons in terms of computational costs both between the three approaches, and between the two classes of proximal algorithms, dual forward-backward, standard and accelerated (FISTA), vs. primal-dual, standard and accelerated by strong convexity (cf. Sec. 3), are reported in Table 2, for configurations I and III (regarded as *easy* and *difficult*) considered as representative. Computational costs are reported in number of iterations actually used to reach the stopping criterion and in real time, for the optimal set of hyperparameters and averaged over 5 realizations.

Table 2 shows first that, as expected, accelerated algorithms always require fewer iterations than non-accelerated ones, thus generally leading to lower computational times (though this is not always the case with FISTA whose complexity per iteration is larger). Also, T-ROF shows always lower computational

		I	II	III	IV	V	VI
		$\Delta\Sigma^2 = 0.1$ $\Delta H = 0.2$	$\Delta\Sigma^2 = 0.15$ $\Delta H = 0.1$	$\Delta\Sigma^2 = 0.1$ $\Delta H = 0.1$	$\Delta\Sigma^2 = 0.05$ $\Delta H = 0.1$	$\Delta\Sigma^2 = 0.1$ $\Delta H = 0.05$	$\Delta\Sigma^2 = 0.1$ $\Delta H = 0.025$
T-ROF	Score	$86.7 \pm 2.1\%$	$79.5 \pm 1.2\%$	$78.5 \pm 1.1\%$	$77.5 \pm 2.9\%$	$69.9 \pm 7.1\%$	$59.5 \pm 2.4\%$
	$\widehat{\Delta H}$	0.21 ± 0.07	0.05 ± 0.02	0.05 ± 0.06	0.07 ± 0.04	0.01 ± 0.06	0.05 ± 0.07
T-*joint*	Score	$91.6 \pm 1.7\%$	$91.5 \pm 2.0\%$	$90.2 \pm 1.9\%$	$84.2 \pm 4.5\%$	$84.3 \pm 3.2\%$	$74.7 \pm 8.2\%$
	$\widehat{\Delta H}$	0.21 ± 0.06	0.07 ± 0.03	0.10 ± 0.02	0.04 ± 0.07	0.05 ± 0.02	0.11 ± 0.28
T-*coupled*	Score	$91.7 \pm 1.7\%$	$91.9 \pm 4.0\%$	$91.1 \pm 1.5\%$	$85.5 \pm 3.8\%$	$86.1 \pm 4.3\%$	$74.3 \pm 8.2\%$
	$\widehat{\Delta H}$	0.20 ± 0.05	0.06 ± 0.04	0.10 ± 0.02	0.08 ± 0.04	0.05 ± 0.02	0.06 ± 0.04

Table 1: **Optimal segmentation performance for different configurations of fractal textures, averaged over** 5 **realizations.** Piecewise fractal textures are characterized by $(\Sigma_0^2, H_0) = (0.6, 0.5)$ and different $(\Delta\Sigma^2, \Delta H)$ as sketched in Fig. 1a. First row: T-ROF segmentation. Second row: T-*joint* segmentation. Third row: T-*coupled* segmentation.

costs compared to T-*joint* and T-*coupled*. This is expected, as T-ROF only works with the regularity and does not use variance.

FISTA vs. Accelerated primal-dual. For T-ROF, FISTA is overall preferable to the accelerated primal-dual algorithm, as both show equivalent computational costs for Configuration I but FISTA is ten times faster (both in number of iterations and computation time) for Configuration III. For T-*joint* and T-*coupled*, for both configurations, accelerated primal-dual is faster than FISTA. For T-*coupled*, in configuration III, FISTA has actually not converged when meeting the the upper limit of iterations. Therefore, FISTA is to be preferred for T-ROF, while accelerated primal-dual algorithms are more relevant for T-*joint* and T-*coupled*.

T-joint vs. T-coupled. Focusing on T-*joint* and T-*coupled* and thus on the accelerated primal-dual algorithm that is faster for these two methods, Table 2 shows that T-*joint* is solved 3 to 4 times faster (both in number of iterations and computational cost) than T-*coupled*.

		Configuration I			Configuration III		
		T-ROF	T-*joint*	T-*coupled*	T-ROF	T-*joint*	T-*coupled*
Iterations (10^3 it.)	DFB	96 ± 48	> 250	> 250	241 ± 18	> 250	> 250
	FISTA	1.7 ± 0.4	50.2 ± 21.0	231 ± 37	$\mathbf{3.7 \pm 0.7}$	48.1 ± 3.4	> 250
	PD	31.8 ± 17.0	> 250	> 250	201 ± 69	> 250	> 250
	AcPD	$\mathbf{1.5 \pm 0.4}$	$\mathbf{31.4 \pm 4.6}$	$\mathbf{125 \pm 67}$	45.2 ± 43	$\mathbf{40.5 \pm 2.8}$	$\mathbf{121 \pm 42}$
Time (s)	DFB	$1,090 \pm 520$	$4,840 \pm 15$	$4,210 \pm 76$	$2,010 \pm 73$	$4,810 \pm 215$	$4,200 \pm 76$
	FISTA	16 ± 4	$1,030 \pm 410$	$4,800 \pm 560$	$\mathbf{30 \pm 5}$	989 ± 64	$5,110 \pm 340$
	PD	297 ± 150	$4,180 \pm 69$	$4,110 \pm 43$	$1,580 \pm 490$	$4,150 \pm 18$	$4,100 \pm 15$
	AcPD	$\mathbf{15 \pm 4}$	$\mathbf{619 \pm 96}$	$\mathbf{2,420 \pm 1,300}$	349 ± 330	$\mathbf{785 \pm 59}$	$\mathbf{2,320 \pm 790}$

Table 2: Number of iterations and computational time necessary to reach Condition (26) for the different proximal algorithms investigated, illustrated on two configurations I ($\Delta H = 0.2$, $\Delta\Sigma^2 = 0.1$) and III ($\Delta H = 0.1$, $\Delta\Sigma^2 = 0.1$). **DFB**: Dual Forward-Backward, **FISTA**: inertial acceleration of DFB, **PD**: primal-dual, **AcPD**: strong-convexity based acceleration of PD.

5.3.3 Overall comparison

As an overall conclusion, results reported above show that there are benefits to use together local regularity and variance, compared to regularity only, when the changes in regularity become small. This implies switching from accelerated dual forward-backward algorithms (FISTA, for T-ROF) to accelerated primal-dual algorithms (for T-*joint* and T-*coupled*).

For difficult configurations, T-*coupled* (slightly) outperforms T-*joint* in terms of segmentation performance, at the price of non negligible increases of computational costs. In that sense, T-*joint* can be considered a reasonable trade-off between relatively poor segmentation performance (as those of T-ROF) and relatively large computational costs (as those of T-*coupled*).

5.4 Real-world textures

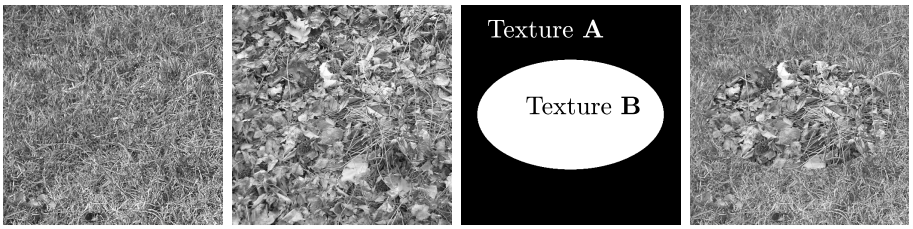
The actual potential of the proposed segmentation strategies is further illustrated on real-world textures, arbitrarily chosen from the commonly used large, documented and publicly available *University of Maryland, HighResolution (UMD HR)*[2] texture dataset. This dataset consists of 50 classes of homogeneous textures, each consisting of 40 images of specific texture classes under different imaging conditions (light, angle,...). From arbitrarily chosen pairs of images, piecewise homogeneous textures are constructed by including textures one into the other using the ellipse-shaped mask used above, as sketched and illustrated in Fig. 6. Homogeneous textures are centered and normalized in variance independently and prior to inclusion, to avoid detections be based only on change of means or variances.

The T-ROF, T-*Joint* and T-*Coupled* segmentation approaches proposed here are applied to and compared on these real-world piecewise homogeneous textures. Further, they are compared against one of the state-of-the-art texture segmentation procedures, based on matrix factorization and local spectral histogram clustering, proposed in [48] ; this joint feature selection and segmentation procedure is hereafter referred to as *LSHC-MF*.

Fig. 7 reports and compares segmentation performance, in terms of percentage of pixels correctly labelled. For the T-ROF, T-*Joint* and T-*Coupled* approaches, optimal performance is reported, after a grid search for the selection of the hyperparameters λ and α. For *LSHC-MF*, optimal performance is also reported, after a grid search on the window size, within which local histogram statistics are computed.

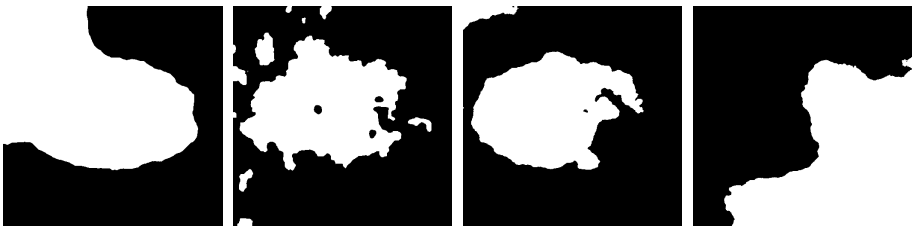
Fig. 7 clearly shows that the *LSHC-MF* approach fails to segment correctly the piecewise homogeneous texture and is outperformed by the T-ROF, T-*Joint* and T-*Coupled* strategies. It also shows that T-ROF, while not being inconsistent, does not yield satisfactory segmentations, whereas T-*Joint* and T-*Coupled* do. Further, it is worth noting that not only does the *Coupled* strategy reach (slightly) better segmentation performance compared to the *Joint* one, but that it also yields smoother boundaries, an outcome of utmost relevance for decision/diagnosis in numerous real-world applications, cf. e.g. [35]. Such significant improvement in segmentation performance is at the cost of significantly higher computational cost: When *LSHC-MF* requires of the order of 1s to perform its most accurate segmentation, as in Fig. 7(d), the T-ROF, T-*Joint* and T-*Coupled*

[2] `http://legacydirs.umiacs.umd.edu/ fer/website-texture/texture.htm`



(a) Texture **A** (b) Texture **B** (c) Mask (d) Mixture **A**-**B**

Figure 6: **Real world textures.** two samples of real-world textures taken from the UMD HR texture dataset ((a) and (b)), piecewise constant mask (c) and piecewise homogeneous texture (d).



82.6% 86.1% 87.9% 55.2%
(a) T-ROF (b) Joint (c) Coupled (d) LSHC-MF

Figure 7: **Segmentation performance on real-world textures.** For the piecewise homogeneous texture shown in Fig. 6(d), performance are computed in term of the percentage of well-classified pixels for T-ROF (a), *Joint* (b), *Coupled* (c) and LSHC-MF (d).

segmentation approaches require 10s, 600s 1800s respectively, in agreement with computational cost evaluation reported in Table 2. Similar conclusions can be drawn for other choices of classes of textures in the UMD HR dataset.

6 Conclusion and future work

The present article has aimed to significantly advanced the state-of-the-art in the segmentation of piecewise fractal textures.

First, it has been proposed to base the segmentation of fractal textures not only on the estimation of the sole local regularity parameter, but to use an additional local parameter, the log-wavelet variance, tightly related to the local variance of the textures. Two variations were investigated, *coupled* and *joint*, that respectively enforce or not co-localized changes in regularity and variance. It has been shown, using large-scale Monte-Carlo simulations, that the use of this additional feature improves drastically segmentation performance when the difference in regularity becomes negligible. This, however, comes at the price of a non-negligible increase in computational costs.

Therefore, a second contribution has been to construct accelerated primal-dual algorithms, requiring the explicit calculation of the strong convexity constant underlying the data fidelity term form. The achieved substantial reduction

in computational costs has turned critical both to be able to conduct large-scale Monte-Carlo simulations and to perform the greedy search for an optimal set of hyperparameters. This low computational cost is also crucial for application on real-world data.

The investigations reported here have permitted to show that accelerated primal-dual algorithms outperform accelerated dual forward-backward (FISTA-type) algorithms for piecewise fractal texture segmentation as soon as the joint use of regularity and variance is required. Further, they showed that the *coupled* formulation, that favors co-localized changes in regularity and variance, performs better than the *joint* formulation, yet at the price of a significantly larger computational cost. Thus, depending on budget constraints on time and requested quality of the solution, the *joint* formulation can be regarded as an effective trade-off.

The proposed theoretical formulations for piecewise fractal texture segmentation and the corresponding accelerated algorithms are mature enough for applications on real-world data, and competitive with respect to state-of-the-art as shown in Sec. 5.4. Application to the segmentation of multiphasic flows is under current investigations. The automation of the tuning of the hyperparameters, which is a crucial point for real-world applications, as underlined in Sec. 5.4, is also being investigated. Extensions to piecewise multifractal textures are further targeted.

A MATLAB toolbox implementing the analysis and synthesis procedures devised here will be made freely and publicly available at the time of publication.

References

[1] P. Abry, S. Jaffard, and H. Wendt. When Van Gogh meets Mandelbrot: Multifractal classification of painting's texture. *Signal Process.*, 93(3):554–572, 2013.

[2] P. Abry and F. Sellan. The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation, 1996.

[3] V. Andrearczyk and P. F. Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recogn. Lett.*, 84:63–69, 2016.

[4] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Match. Int.*, 33(5):898–916, 2011.

[5] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.

[6] H. Biermé, M. Meerschaert, and H. Scheffle. Operator scaling stable random fields. *Stochastic Processes and their Applications*, 117(3):312–332, 2007.

[7] X. Cai, R. Chan, C.-B. Schonlieb, and T. Steidl, G.and Zeng. Linkage between Piecewise Constant Mumford-Shah model and ROF model and its virtue in image segmentation. *arXiv preprint arXiv:1807.10194*, 2018.

[8] X. Cai and G. Steidl. Multiclass segmentation by iterated ROF thresholding. In *International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 237–250. Springer, 2013.

[9] A. Chambolle and C. Dossal. On the convergence of the iterates of "FISTA". *J. Optim. Theory Appl.*, 166(3):25, August 2015.

[10] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1):120–145, 2011.

[11] B. B. Chaudhuri and N. Sarkar. Texture segmentation using fractal dimension. *IEEE Trans. Pattern Anal. Match. Int.*, 17(1):72–77, 1995.

[12] C. Chaux, P.L. Combettes, and V.R. Pesquet, J.-C. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4):1495–1518, Jun. 2007.

[13] M. Cimpoi, S. Maji, I. Kokkinos, and A. Vedaldi. Deep filter banks for texture recognition, description, and segmentation. *Int. J. Comp. Vis.*, 118(1):65–94, 2016.

[14] S. Cohen and J. Istas. *Fractional fields and applications*. Springer, 2013.

[15] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York, 2011.

[16] L. Condat. A primal-dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2):460–479, 2013.

[17] G. Didier, M. M. Meerschaert, and V. Pipiras. Domain and range symmetries of operator fractional Brownian fields. *Stochastic Processes and their Applications*, 128(1):39–78, 2018.

[18] D. Dunn, W. E. Higgins, and J. Wakeley. Texture segmentation using 2-D Gabor elementary functions. *IEEE Trans. Pattern Anal. Match. Int.*, 16(2):130–149, 1994.

[19] P. Getreuer. Chan-Vese segmentation. *Image Processing On Line*, 2:214–224, 2012.

[20] M. A. Ibrahim, O. A. Ojo, and P. A. Oluwafisoye. Identification of emphysema patterns in high resolution computed tomography images. *Journal of Biomedical Engineering and Informatics*, 4(1):16, 2017.

[21] S. Jaffard. Wavelet techniques in multifractal analysis. *Fractal Geometry and Applications: A Jubilee of Benoît Mandelbrot, M. Lapidus and M. van Frankenhuijsen Eds., Proceedings of Symposia in Pure Mathematics*, 72(2):91–152, 2004.

[22] A. K. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recogn.*, 24(12):1167–1186, 1991.

[23] M. Jung, G. Peyré, and L. D. Cohen. Non-local Active Contours. In *Proc. SSVM*, volume 6667, pages 255–266, Ein-Gedi, Israel, 2011. Springer.

[24] I. Kokkinos, G. Evangelopoulos, and P. Maragos. Texture analysis and segmentation using modulation features, generative models, and weighted curve evolution. *IEEE Trans. Pattern Anal. Match. Int.*, 31(1):142–157, 2009.

[25] N. Komodakis and J.-C. Pesquet. Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Process. Mag.*, 32(6):31–54, Nov. 2015.

[26] R. Leonarduzzi, H. Wendt, P. Abry, S. Jaffard, C. Melot, S. G. Roux, and M. E. Torres. p-exponent and p-leaders, Part II: Multifractal Analysis. Relations to Detrended Fluctuation Analysis. *Physica A*, 448:319–339, 2016.

[27] S. Mallat. *A wavelet tour of signal processing.* Academic Press, San Diego, USA, 1997.

[28] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. Pattern Anal. Match. Int.*, 14(7):710–732, 1992.

[29] Z. Marin, K. A. Batchelder, B. C. Toner, L. Guimond, E. Gerasimova-Chechkina, A. R. Harrow, A. Arneodo, and A. Khalil. Mammographic evidence of microenvironment changes in tumorous breasts. *Medical Physics*, 44(4):1324–1336, 2017.

[30] M. T. McCann, D. G. Mixon, M. C. Fickus, C. A. Castro, J. A. Ozolek, and J. Kovacevic. Images as occlusions of textures: A framework for segmentation. *IEEE Trans. Image Process.*, 23(5):2033–2046, 2014.

[31] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied Math.*, 42(5):577–685, 1989.

[32] J.F. Muzy, E. Bacry, and A. Arneodo. Wavelets and multifractal formalism for singular signals: Application to turbulence data. *Phys. Rev. Lett.*, 67(25):3515–3518, 1991.

[33] J.D.B. Nelson, C. Nafornita, and A. Isar. Semi-local scaling exponent estimation with box-penalty constraints and total-variation regularization. *IEEE Trans. Image Process.*, 25(7):3167–3181, 2016.

[34] B. Pascal, N. Pustelnik, P. Abry, and J.-C. Pesquet. Block-coordinate proximal algorithms for scale-free texture segmentation. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, Calgary, Alberta, Canada, Apr. 15-20 2018.

[35] B. Pascal, N. Pustelnik, P. Abry, M. Serres, and V. Vidal. Joint Estimation of Local Variance and Local Regularity for Texture Segmentation. Application to Multiphase Flow Characterization. In *Proc. Int. Conf. Image Process.*, pages 2092–2096, Athens, Greece, 2018. IEEE.

[36] R. M. Pereira, C. Garban, and L. Chevillard. A dissipative random velocity field for fully developed fluid turbulence. *Journal of Fluid Mechanics*, 794:369–408, 2016.

[37] N. Pustelnik, C. Chaux, and J.-C. Pesquet. Parallel ProXimal Algorithm for image restoration using hybrid regularization. *IEEE Trans. Image Process.*, 20(9):2450–2462, 2011.

[38] N. Pustelnik, H. Wendt, P. Abry, and N. Dobigeon. Combining local regularity estimation and total variation optimization for scale-free texture segmentation. *IEEE Trans. Computational Imaging*, 2(4):468–479, 2016.

[39] S. Ray and R. H. Turi. Determination of number of clusters in K-means clustering and application in colour image segmentation. In *Proc. Int. Conf. Adv. Pattern Recognition and Digital Techniques*, pages 137–143. Calcutta, India, 1999.

[40] A. Repetti, E. Chouzenoux, and J.-C. Pesquet. A parallel block-coordinate approach for primal-dual splitting with arbitrary random block selection. In *Proc. Eur. Sig. Proc. Conference*, pages 235–239, Nice, France, Aug. 31-Sept. 4, 2015.

[41] S.G. Roux, M. Clausel, B. Vedel, S. Jaffard, and P. Abry. Self-Similar Anisotropic Texture Analysis: the Hyperbolic Wavelet transform contribution. *IEEE Trans. Image Process.*, 22(11):4353 – 4363, 2013.

[42] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.

[43] M. L. Stein. Fast and exact simulation of fractional Brownian surfaces. *Journal of Computational and Graphical Statistics*, 11(3):587–599, 2002.

[44] B.C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38(3):667–681, Apr. 2013.

[45] H. Wendt, P. Abry, and S. Jaffard. Bootstrap for Empirical Multifractal Analysis. *IEEE Signal Process. Mag.*, 24(4):38–48, 2007.

[46] H. Wendt, S. Combrexelle, Y. Altmann, J.-Y. Tourneret, S. McLaughlin, and P.Abry. Multifractal analysis of multivariate images using gamma Markov random field priors. *SIAM J. Imaging Sci.*, 11(2):1294–1316, 2018.

[47] H. Wendt, S. G. Roux, P. Abry, and S. Jaffard. Wavelet leaders and bootstrap for multifractal analysis of images. *Signal Process.*, 89(6):1100–1114, 2009.

[48] J. Yuan, D. Wang, and A. M. Cheriyadat. Factorization-based texture segmentation. *IEEE Trans. Image Process.*, 24(11):3488–3497, November 2015.