Variational algorithms to remove stationary noise. Application to microscopy imaging.

Jérôme Fehrenbach, Pierre Weiss, Corinne Lorenzo

Variational algorithms to remove stationary noise. Applications to microscopy imaging.

Jérôme Fehrenbach, Pierre Weiss, and Corinne Lorenzo

***Abstract*—A framework and an algorithm are presented in order to remove stationary noise from images. This algorithm is called VSNR (Variational Stationary Noise Remover). It can be interpreted both as a restoration method in a Bayesian framework and as a cartoon+texture decomposition method. In numerous denoising applications the white noise assumption fails: structured patterns (e.g. stripes) appear in the images. The model described here addresses these cases. Applications are presented with images acquired using different modalities: scanning electron microscope, FIB-nanotomography, and an emerging fluorescence microscopy technique called SPIM (Selective Plane Illumination Microscope).**

***Index Terms*—Stripe removal, non linear filtering, fast algorithms, total variation, Gabor filter, SPIM microscope, scanning electron microscope, nanotomography, texture-geometry decompositions.**

I. Introduction

Many imaging modalities tend to impair images with structured noise. For instance, stripes appear in images produced by recent microscopes such as Single Plane Illumination Microscope (SPIM) [13], Atomic Force Microscope (AFM), scanning electron microscopy (SEM) [8], [15], Synchrotron X-Ray Tomographic images, Focused Ion Beam nanotomography [18], or in remote sensing imaging such as MODerate-resolution Imaging Spectroradiometer (MODIS) images or digital elevation models [6], [7]. Figure 1 shows examples of stripes in images obtained using different modalities.

The main purpose of this work is to propose a simple noise model which describes accurately the kind of distortions appearing in Figure 1. This statistical modeling allows to formulate the restoration problem in a Bayesian framework. It leads to the resolution of a large scale convex programming problem which is solved using recent advances in numerical optimization. We call our algorithm VSNR which stands for Variational Stationary Noise Remover.

As far as we know, few works address this restoration problem. Recently, [18] proposed a combined wavelet and Fourier method which allows to interpolate and restore wavelet coefficients impaired by stripes. In [8], a Fourier filtering is proposed, where noisy Fourier coefficients are first detected using statistical tests. Contrarily to our work, these approaches do not rely on a precise statistical modeling of the noise and incorporate prior information on the images only implicitly.

The general image formation model we consider can be stated as follows:

$$u_0 = Hu + b$$

Jérôme Fehrenbach and Pierre Weiss are with Université de Toulouse, the IMT-UMR5219 and with ITAV-UMS3039, France.

Corinne Lorenzo is with Université de Toulouse and ITAV-UMS3039, France.

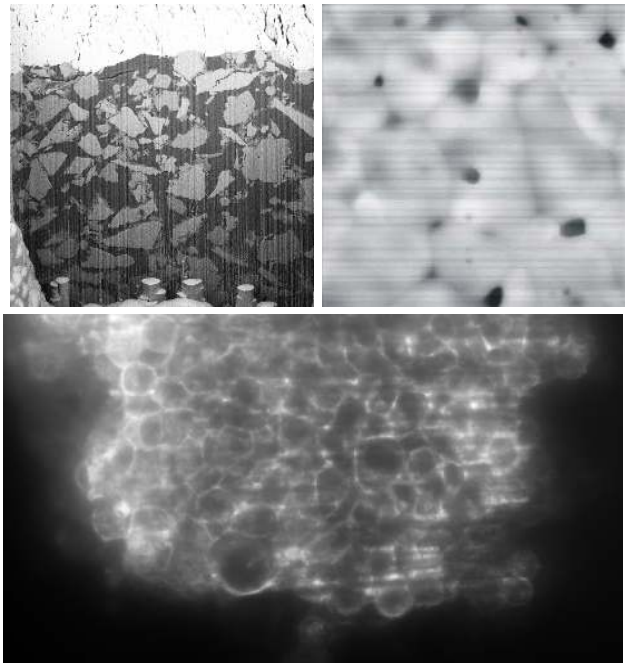


Fig. 1. Stripes for various imaging modalities. Top-left: cement paste image using Ion beam nanotomography. Top-right: sintered specimen of CeO2 image using Scanning Electron Microscope (SEM). Bottom: medaka fish embryo image using Selective Plane Illumination Microscope (SPIM)

where u_0 is the observed digital image, H is a deterministic linear operator that models the acquisition process (sampling, convolution, indirect measurements, ...) and b is a noise or a texture. In restoration applications, the objective is to retrieve u knowing the observation u_0 and the observation operator H. In texture+cartoon decomposition, H is the identity operator, and the objective is to retrieve both the cartoon component u and the texture component b.

A standard assumption in the literature is that the noise is white, meaning that the different components of b are independent random variables of finite variance. Though this noise assumption models accurately some real applications, it fails to do so in many scenarii. In the case of image restoration/denoising problems, it often leads to unsatisfactory results.

In this article we propose a noise model that suits more complex frameworks and that can be handled numerically in reasonable computing times. We replace the white noise assumption by a *stationary noise* assumption. A stationary noise b can be defined as a stochastic process having the following translation invariance property:

$$\forall \tau \in \mathbb{Z}^2, \quad \mathbf{p}(b) = \mathbf{p}(T_\tau(b)) \tag{1}$$

where $\mathbf{p}$ denotes the probability density function and T_τ is the translation of vector τ with periodic boundary conditions on the image domain.

This hypothesis is natural in many applications, since it amounts to assume that there is no a-priori knowledge about the location of any feature or noise pattern in the image. However, the sole hypothesis of stationarity appears to be hardly numerically tractable. In this work, we restrict to a subclass of stationary stochastic processes, that can be easily described in the frequency domain: these processes are generated by convolving white noise with a given kernel. The noise thus appears as "structured" in the sense that some pattern might be visible, see Figure 3.

The proposed model shares many similarities with the negative norm models of Y. Meyer [17], its successors [2] and the decomposition algorithms proposed in [11]. Meyer's idea is to decompose an image into a piecewise smooth component and an oscillatory component. We refer to [2] for a review of the principles, algorithms and results using this approach. An alternative way of decomposing images was proposed in [11]. The idea is to seek for components that are sparse in given dictionaries. Different choices for the elementary atoms composing the dictionary allow to recover different kind of textures.

The main contributions of the present paper are:

1) We propose a simple class of random processes that describes real-life noises and textures.
2) Similarly to [11], the texture is described through a dictionary. In our work each dictionary is composed of a single pattern shifted in space, ensuring translation invariance and fast computation in Fourier domain.
3) The proposed formalism and algorithms can be used when dealing with inverse problems such as deblurring or sampling with indirect measurements.
4) A Bayesian approach is provided in order to take into account the statistical nature of textures precisely.
5) The decomposition problem is recast into a convex optimization problem that is solved with a recent algorithm of A. Chambolle and T. Pock [4] allowing to obtain results in an interactive time.
6) We propose a C and Matlab implementation on our webpage in the spriit of reproducible research http://www.math.univ-toulouse.fr/~weiss/PageCodes.html.

The outline of this paper is as follows. In section II, we introduce the notation. Section III contains detailed definitions and some elementary properties of the noise model. Section IV presents a restoration or decomposition model based on the maximum a posteriori (MAP) estimator. Section V details an efficient numerical scheme based on [4] to solve the resulting convex programming problems. In section VI we present some applications and results on synthetic and real images.

II. Notation

In all the paper $n = n_x \times n_y$ will refer to the number of pixels of the degraded image. m stands for the number of filters used to describe the noise. Let $u \in \mathbb{R}^{n_x \times n_y}$ be an image. $u(\mathbf{x})$ denotes the pixel of coordinates $\mathbf{x} \in \{1, 2, \cdots, n_x\} \times \{1, 2, \cdots, n_y\}$.

$\|u\|_p$ refers to the standard ℓ^p-norm of u. Let $(u_1, u_2) \in \mathbb{R}^n \times \mathbb{R}^n$ the dot product is denoted $\langle u_1, u_2 \rangle = u_1^T u_2$.

Let $\Lambda = \mathbb{R}^{n \times m}$ and $Q = \mathbb{R}^{n \times 2}$. These spaces are endowed with inner products $\langle \cdot, \cdot \rangle_\Lambda$ and $\langle \cdot, \cdot \rangle_Q$. We set $\|q\|_Q = \sqrt{\langle q, q \rangle_Q}$ and $\|\lambda\|_\Lambda = \sqrt{\langle \lambda, \lambda \rangle_\Lambda}$.

Let $A : \Lambda \to Q$ be a linear operator. The adjoint A^* of A is defined by:

$$\langle A\lambda, q \rangle_Q = \langle A^* q, \lambda \rangle_\Lambda, \quad \forall (\lambda, q) \in \Lambda \times Q. \tag{2}$$

The norm of the operator A is defined by:

$$|||A||| = \max_{\|\lambda\|_\Lambda \leq 1} \|A\lambda\|_Q \tag{3}$$

$*$ is the convolution operator with periodic boundary conditions. $\mathcal{F}$ and $\mathcal{F}^{-1}$ respectively denote the discrete Fourier and inverse Fourier transforms. We will also use the notation $\hat{u} = \mathcal{F}u$. Finally, ∇ denotes the discrete gradient operator. In this article we use the following discretization:

$$\nabla u = (\partial_x u, \partial_y u)$$

where

$$(\partial_x u)(i,j) = \begin{cases} u(i+1,j) - u(i,j) & \text{if } \ i < n_x \\ 0 & \text{if } \ i = n_x \end{cases} \tag{4}$$

$$(\partial_y u)(i,j) = \begin{cases} u(i,j+1) - u(i,j) & \text{if } \ j < n_y \\ 0 & \text{if } \ j = n_y. \end{cases} \tag{5}$$

Let $\phi : \Lambda \to \mathbb{R} \cup \{+\infty\}$ be a convex, closed function with non-empty domain. The domain of ϕ is defined by $\mathrm{dom}(\phi) = \{\lambda \in \Lambda, \phi(\lambda) < +\infty\}$. ϕ^* refers to the Fenchel conjugate of ϕ defined by:

$$\phi^*(\eta) = \sup_{\lambda \in \Lambda} \langle \eta, \lambda \rangle_\Lambda - \phi(\lambda). \tag{6}$$

The sub-differential of ϕ at λ_1 is the set defined by:

$$\partial\phi(\lambda_1) = \{\eta \in \Lambda, \phi(\lambda_2) \geq \phi(\lambda_1) + \langle \eta, \lambda_2 - \lambda_1 \rangle_\Lambda, \ \forall \lambda_2 \in \Lambda\}.$$

A function ϕ is said to be strongly convex of parameter $\gamma > 0$, if the following inequality holds for all $\lambda_1, \lambda_2 \in \mathrm{dom}(\phi)$ and for all $\eta \in \partial\phi(\lambda_1)$:

$$\phi(\lambda_2) \geq \phi(\lambda_1) + \langle \eta, \lambda_2 - \lambda_1 \rangle_\Lambda + \frac{\gamma}{2} \|\lambda_2 - \lambda_1\|_\Lambda^2,$$

Let $\Xi \subseteq \Lambda$ be a set. The indicator function of Ξ is denoted χ_Ξ and defined by:

$$\chi_\Xi(x) = \begin{cases} 0 & \text{if } x \in \Xi \\ +\infty & \text{otherwise} \end{cases}.$$

The resolvent or proximal operator of ϕ at point λ is defined by:

$$(\mathrm{Id} + \partial\phi)^{-1}(\lambda) = \arg\min_{\lambda' \in \Lambda} \phi(\lambda') + \frac{1}{2} \|\lambda' - \lambda\|_\Lambda^2.$$

It can be seen as a generalization of the projection operator since $(\mathrm{Id} + \partial\chi_\Xi)^{-1}(\lambda)$ is the projection of λ on the set Ξ. We refer to [24] for a detailed presentation of the above tools.

III. Noise model

A. A class of stationary noises

In this paper, we address the class of stationary noises that satisfy the following hypothesis.

Hypothesis 3.1: We assume that the noise b reads:

$$b = \sum_{i=1}^{m} \lambda_i * \psi_i, \tag{7}$$

where:

- each $\psi_i \in \mathbb{R}^n$ is a known elementary pattern,
- The λ_i's are independent realizations of white noise processes with known probability density functions.

Proposition 3.1: This class of noises is *stationary* - in the sense that it satisfies property (1).
The proof of this proposition can be found in [25], page 404.

B. Rationale for the noise model

The images presented in Figure 1 are corrupted by parallel stripes. A noise removal algorithm should be able both to locate their position and to estimate their intensity. Such a perturbation b can be modeled by the following equation

$$b = \lambda * \psi, \tag{8}$$

where ψ is an elementary pattern, similar to a stripe, and λ describes the pattern positions and intensities.

In realistic applications, the noise is seldom describable simply as in Equation (8). A more realistic noise description would be the sum of processes of type (8). For instance, in SPIM imaging in addition to stripes, Poisson or Gaussian white noise appears due to the imperfections of the imaging system. The Gaussian white noise is a special case of (8) where ψ is a Dirac function and λ is a sample of a Gaussian white noise process.

The common feature of the stripes is their orientation θ, as well as their dimensions (width and length) that belong to a restricted range. In this paper, we simply model a stripe as an anisotropic Gaussian function defined by :

$$\psi(x, y) = \exp\left(-\frac{x'^2}{\sigma_x^2} - \frac{y'^2}{\sigma_y^2}\right),$$

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$. In order to describe stripes of different width or length, several filters $\psi_1, \ldots, \psi_m$ with different geometrical parameters can be used.

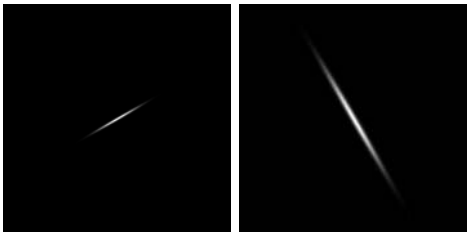


Fig. 2. Left: a Gaussian function in space domain – Right: the same Gaussian function in Fourier domain.

IV. Restoration/decomposition model

Throughout this paper, we assume the following image formation model:

$$u_0 = Hu + b \tag{9}$$

where H is a linear operator and b is independent of u and satisfies hypothesis 3.1. In the numerical results, we will concentrate on the case $H = \mathrm{Id}$, however all the proposed algorithms extend to an arbitrary operator $H \neq \mathrm{Id}$.

A. A MAP reconstruction approach

In this work we assume that the samples $\lambda = \{\lambda_i\}_{i=1..m}$ are drawn from stochastic processes with known probability density functions. The prior probability on the space of images is denoted $\mathbf{p}(u)$. The maximum a posteriori (MAP) approach leads to maximize:

$$\begin{aligned}\mathbf{p}(\lambda, u|u_0) &= \frac{\mathbf{p}(u_0|\lambda, u)\mathbf{p}(\lambda, u)}{\mathbf{p}(u_0)} \\ &= \frac{\mathbf{p}(u_0|\lambda, u)\mathbf{p}(\lambda)\mathbf{p}(u)}{\mathbf{p}(u_0)}.\end{aligned}$$

Taking the log of the conditional probability leads to the following optimization problem:

$$\text{Find } \lambda^* \in \operatorname*{Arg\,min}_{\lambda\in\Lambda} -\log \mathbf{p}(u_0|\lambda, u) - \log \mathbf{p}(\lambda) - \log \mathbf{p}(u). \tag{10}$$

As we assumed independence of the λ_is,

$$-\log \mathbf{p}(\lambda) = \sum_{i=1}^{m} -\log \mathbf{p}(\lambda_i)$$

and

$$\begin{aligned}\mathbf{p}(u_0|\lambda, u) &= \mathbf{p}(Hu + \sum_{i=1}^{m} \lambda_i * \psi_i|\lambda, u) \\ &= \delta\left(Hu + \sum_{i=1}^{m} \lambda_i \star \psi_i - u_0\right),\end{aligned}$$

where δ denotes the Dirac mass.

In order to specify problem (10), we need to define the densities $\mathbf{p}(u)$ (image prior) and $\mathbf{p}(\lambda)$ (noise prior).

Remark 4.1: In some applications the independence of u and λ is questionable and a multiplicative model could be considered. However our numerical experiments using multiplicative models did not improve the results sensibly compared to the additive models and required heavier computations.

B. Image prior

In this work, we use the standard assumption that images are smooth or piecewise smooth. This can be promoted by a p.d.f. of the form

$$\mathbf{p}(u) \propto \exp(-\|\nabla u\|_{1,\epsilon}). \tag{11}$$

where

$$\begin{aligned}\|\cdot\|_{1,\epsilon}: \quad &\mathbb{R}^{n\times 2} \to \mathbb{R} \\ &q \mapsto \|q\|_{1,\epsilon} = \textstyle\sum_{\mathbf{x}} \phi_\epsilon(\|q(\mathbf{x})\|_2)\end{aligned}$$

and

$$\phi_\epsilon(t) = \begin{cases} \frac{t^2}{2\epsilon} & \text{if } |t| \leq \epsilon \\ |t| - \frac{\epsilon}{2} & \text{otherwise.} \end{cases}$$

This function can be rewritten using duality as [27]:

$$||z||_{1,\epsilon} := \max_{\|q\|_\infty \leq 1} \langle z, q \rangle - \frac{\epsilon}{2}\|q\|_2^2.$$

The interest of this prior is twofold:

- by choosing $\epsilon = 0$, $\|\cdot\|_{1,0}$ corresponds to the standard isotropic ℓ^1-norm $\mathbb{R}^{n\times 2}$ and $\lim_{\epsilon\to+\infty} \epsilon\|\cdot\|_{1,\epsilon} = \|\cdot\|_2^2$. This formalism thus captures the standard ℓ^1 (total variation) and ℓ^2 (Tikhonov) regularization.
- by setting $\epsilon \neq 0$, the numerical scheme designed to solve (10) converges faster than for $\epsilon = 0$ and may provide better results in some applications.

C. Noise priors

The p.d.f. $\mathbf{p}(\lambda_i)$ in equation (10) still needs to be defined in order to specify the optimization problem completely. The p.d.f. should be chosen depending on the noise nature. In the experiments of this paper we consider three cases:

- *Uniform noise*. It is obtained by setting

$$\mathbf{p}(\lambda_i) \propto \begin{cases} 1 & \text{if } \|\lambda_i\|_\infty \leq \alpha_i \\ 0 & \text{otherwise.} \end{cases}.$$

  This assumption allows to model bounded noises. See right images in Figure 3.
- *Gaussian noise*. This hypothesis consists in setting $\mathbf{p}(\lambda_i) \propto \exp(-\alpha_i\|\lambda_i\|_2^2)$. The corresponding noise component $b_i = \lambda_i * \psi_i$ is then a colored Gaussian noise. Its power spectral density is given by the Fourier transform of ψ_i. See middle images in Figure 3.
- *Laplace noise*. It is obtained by setting $\mathbf{p}(\lambda_i) \propto \exp(-\alpha_i\|\lambda_i\|_1)$. This distribution is known to be a convex approximation of Bernoulli processes [9], [16]. See left images in Figure 3.

All these noise priors are generalized Gaussian distributions and their distribution function writes:

$$\mathbf{p}(\lambda_i) \propto \exp(-\phi_i(\lambda_i)). \tag{12}$$

Remark 4.2: The middle and right columns in Figure 3 look very similar. This can be explained using Lindeberg-Feller theorem (an extension of the central limit theorem). It asserts that under suitable hypotheses, a weighted sum of independent variables tends to a normal distribution. This remark has an important practical consequence: in most applications, it is convenient to approximate λ by a Gaussian process. Indeed, the numerical algorithms we develop have a convergence rate of order $O(1/k^2)$ - where k is the iteration counter - for Gaussian priors, while it is only of order $O(1/k)$ for the others (see proposition 5.3).

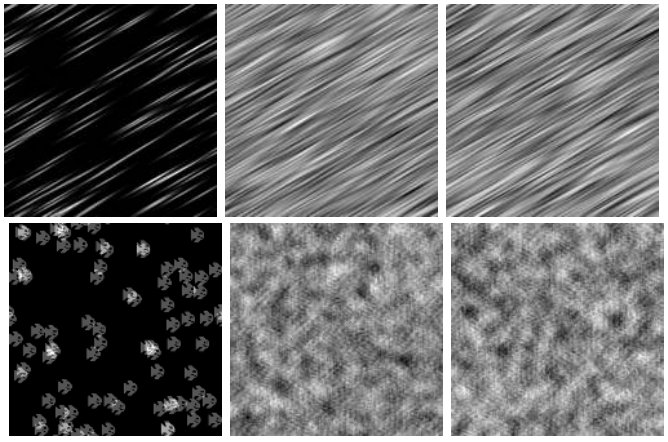


Fig. 3. Examples of stationary processes obtained by convolving white noises with different marginals with a filter. Top row : convolution with a Gabor filter. Bottom row: convolution with a fish pattern. From left to right : convolution with a white Bernoulli process, a Gaussian white noise and a uniform white noise.

D. Resulting optimization problems

By replacing formulas (11) and (12) in (10) and denoting $\lambda = \{\lambda_i\}_{i=1}^m$, we obtain the following optimization problem:

$$\begin{array}{lll} \text{Find } (\lambda, u) \in & \arg\min & \|\nabla u\|_{1,\epsilon} + \sum_{i=1}^m \phi_i(\lambda_i) \\ & \text{subject to} & u \in \mathbb{R}^n, \lambda \in \Lambda \\ & & Hu + \sum_{i=1}^m \lambda_i * \psi_i = u_0 \end{array} \tag{13}$$

In the case $H = \text{Id}$ it simplifies to:

$$\text{Find} \quad \lambda \in \operatorname*{Arg\,min}_{\lambda\in\Lambda} P(\lambda), \tag{14}$$

where

$$P(\lambda) = \left|\left|\nabla\left(u_0 - \sum_{i=1}^m \lambda_i * \psi_i\right)\right|\right|_{1,\epsilon} + \sum_{i=1}^m \phi_i(\lambda_i). \tag{15}$$

Remark 4.3: We show in [12] that the proposed model provides a Bayesian interpretation and generalizes the negative norm texture+cartoon decomposition models [2], [17] in the discrete setting.

V. Numerical algorithms

In the case of Tikhonov regularization (which corresponds to $\phi_i(\lambda) = \frac{\alpha_i}{2}\|\lambda_i\|_2^2$ and $J_R(u) = \|\nabla u\|_2^2$), Problem (14) can be solved exactly in $O(m^3 n \log n)$ operations using Fourier transforms and inversion of small $m \times m$ linear systems. In the more general case, it is impossible to get an exact solution. The objective of this section is to design iterative methods that lead to approximate solutions of (14).

A. Problem relaxation

Since problem (14) cannot be solved exactly, we are interested in finding an ϵ-solution defined by:

$$\text{Find } \lambda_\epsilon \in \Lambda \text{ such that } P(\lambda_\epsilon) - P(\lambda^*) \leq \epsilon. \tag{16}$$

An iterative optimization algorithm defines a sequence $(\lambda_k)_{k\in\mathbb{N}}$ that converges to a solution of (14). The resolution of

(16) amounts to define an appropriate stopping criterion. Since $P(\lambda^*)$ is unknown, $P(\lambda_k) - P(\lambda^*)$ cannot be evaluated, but this quantity can be bounded by the duality gap defined in equation (23) [26]. In order to prevent the duality gap to be infinite (corresponding to an infeasible dual variable), we add a box constraint to problem (16). This extra constraint also allows to obtain convergence rates, see Proposition 5.3.

To summarize, the constrained problem we consider reads :

$$\text{Find } \lambda_\epsilon \in \Xi \text{ such that } P(\lambda_\epsilon) - P(\lambda^*) \leq \epsilon. \tag{17}$$

where

$$\Xi = \left\{\lambda \in \mathbb{R}^{n\times m}, \|\lambda\|_\infty \leq C\right\}$$

and where λ^* is one exact solution of (14) and C is large enough (e.g. $C \geq \|\lambda^*\|_\infty$). Under the latter condition, the extra box-constraint is inactive and the solutions of (17) are the same as those of (16).

B. Reformulation as a saddle-point problem

This section is devoted to the resolution of (17). The extension to problems (13) is straightforward. In order to apply the ideas presented in [4], we reformulate (17) as a saddle-point problem.

To simplify the reading, let us first introduce some notation. A is the following linear operator:

$$\begin{array}{rccc} A: & \Lambda & \rightarrow & Q \\ & \lambda & \mapsto & \nabla\left(\sum_{i=1}^m \lambda_i * \psi_i\right) \end{array} \tag{18}$$

We detail a procedure to compute the highest singular value L of A in appendix B. By denoting

$$F(g) = \|\nabla u_0 - g\|_{1,\epsilon} \tag{19}$$

and

$$G(\lambda) = \sum_{i=1}^m \phi_i(\lambda_i) + \chi_\Xi(\lambda), \tag{20}$$

problem (17) can be recast as the following convex-concave saddle-point problem:

$$\min_{\lambda\in\Xi}\max_{q\in Q} \langle A\lambda, q\rangle_Q - F^*(q) + G(\lambda). \tag{21}$$

The interest of this saddle-point reformulation is twofold:

- It allows the use of primal-dual algorithms which are known as being robust and efficient.
- It will allow to define a duality gap, which will provide a reliable stopping criterion for the iterative algorithm.

C. Elementary properties of the problem

By inverting the minimum and the maximum in (21) we obtain the following dual problem :

$$\max_{q\in Q}\min_{\lambda\in\Xi} \langle \lambda, A^*q\rangle_Q - F^*(q) + G(\lambda). \tag{22}$$

Let us denote:

$$\begin{aligned} P(\lambda) &= \max_{q\in Q} \langle \lambda, A^*q\rangle_\Lambda - F^*(q) + G(\lambda) \\ &= F(A\lambda) + G(\lambda), \end{aligned}$$

$$\begin{aligned} D(q) &= \min_{\lambda\in\Lambda} \langle A\lambda, q\rangle_Q - F^*(q) + G(\lambda) \\ &= -F^*(q) - G^*(-A^*q) \end{aligned}$$

and

$$\Delta(\lambda, q) = P(\lambda) - D(q). \tag{23}$$

The letter P stands for Primal and D stands for Dual. By using standard results of convex programming, we prove that:

Theorem 5.1 (Characterization of problem (17) solutions):

1) Problem (17) admits a convex, non-empty set of solutions.
2) Problem (21) admits a non-empty set of saddle points.
3) Let (λ^*, q^*) be a saddle point of (21). It satisfies $\Delta(\lambda^*, q^*) = 0$ and λ^* is a solution of problem (17).
4) Finally, for any $(\lambda, q) \in \Xi \times \mathbb{R}^{n\times 2}$,

$$\Delta(\lambda, q) \geq P(\lambda) - P(\lambda^*).$$

The last item in this theorem indicates that the duality gap might be used as a stopping criterion for an iterative algorithm: it is an upper bound of the difference between the objective function and the minimum.

Proof: Points (1) and (2) result from boundedness of the set $\Xi \times \mathrm{dom}(F^*)$ and convexity of the functionals. Points (3) and (4) are standard results of convex analysis [24]. ■

The following propositions will be useful in order to determine the algorithm step sizes.

Proposition 5.1: In the standard Euclidean metric, if $\epsilon > 0$, then F^* is strongly convex with parameter ϵ.

Proposition 5.2: In the standard Euclidean metric, if $\phi_i(\lambda_i) = \frac{\alpha_i}{2}\|\lambda_i\|_2^2$ for all $i \in \{1..m\}$, then G is strongly convex with parameter:

$$\gamma = \min_{i\in\{1..m\}} \alpha_i \tag{24}$$

D. A primal-dual algorithm

Indeed, Nesterov paved the way to the development of new efficient algorithms in ([19], [22]). From a theoretical point of view, these algorithms are shown to be optimal in the class of first order methods and they outperform second order algorithms - like interior point methods - for large scale problems and moderate accuracy. Among all the papers published recently on this topic, [4] is probably the most versatile and we decided to present and implement this strategy.

The algorithm proposed by Chambolle and Pock in [4] applied to problem (17), can be written as follows:

Algorithm 1: Primal-Dual algorithm [4]

Input:
ϵ: the desired precision;
$(\bar{\lambda}_0, q_0)$: a starting point;
σ_0, τ_0 such that $\sigma_0\tau_0 L^2 < 1$;
Output:
λ_ϵ: a solution to problem (17).
begin
  k=0;
  while $\Delta(\lambda_k, q_k) > \epsilon\Delta(\lambda_0, q_0)$ **do**
    $q_{k+1} = (\mathrm{Id} + \sigma_k \partial F^*)^{-1}(q_k + \sigma_k A\bar{\lambda}_k)$
    $\lambda_{k+1} = (\mathrm{Id} + \tau_k \partial G)^{-1}(\lambda_k - \tau_k A^* q_{k+1})$
    $(\tau_{k+1}, \sigma_{k+1}, \theta_k) =$Update$(\tau_k, \sigma_k, \epsilon, \gamma, L)$
    $\bar{\lambda}_{k+1} = \lambda_{k+1} + \theta_k(\lambda_{k+1} - \lambda_k)$
    $k = k + 1$;
  end
end

The calculation of the resolvents of F^* and G are described in the appendix. Following [4], the Update function should be implemented according to the following rules:

Algorithm 2: Step size update rules

Input:
ε: parameter of strong convexity of F^*
γ: parameter of strong convexity of G (see eq. (24))
L: highest singular value of G
σ_k and τ_k: step sizes at iteration k
Output: θ_k, σ_{k+1} and τ_{k+1}.
begin
  if $\phi_i(\cdot) = \frac{\alpha_i}{2}\|\cdot\|_2^2$ *for all* $i \in \{1..m\}$ **then**
    if $\varepsilon > 0$ **then**
      The sequence $((\tau_k, \sigma_k))_{k\in\mathbb{N}}$ is constant:
      $\mu = 2\sqrt{\gamma\varepsilon}/L$
      $\tau_{k+1} = \frac{\mu}{2*\gamma}$
      $\sigma_{k+1} = \frac{\mu}{2*\varepsilon}$
      $\theta_k = \frac{1}{1+\mu}$
    else
      $\theta_k = \frac{1}{\sqrt{1+2\gamma\tau}}$
      $\tau_{k+1} = \theta_k\tau_k$
      $\sigma_{k+1} = \frac{\sigma_k}{\theta_k}$
    end
  else
    The sequence $((\tau_k, \sigma_k))_{k\in\mathbb{N}}$ is constant:
    $\theta_k = 1$
    $\tau_{k+1} = \tau_k$
    $\sigma_{k+1} = \sigma_k$
  end
end

The convergence properties of this algorithm are summarized below.

Proposition 5.3: The sequence $(\lambda_k, q_k)_{k\in\mathbb{N}}$ converges to a saddle-point of (21). Moreover it guarantees the following convergence rates:

- If $\epsilon = 0$ and a function ϕ_i is non quadratic then:
$$\Delta(\lambda_k, q_k) = O\left(\frac{L}{k}\right)$$
- If $\epsilon = 0$ and all functions ϕ_i are squared ℓ^2-norms:
$$\Delta(\lambda_k, q_k) = O\left(\frac{L}{k^2}\right)$$
and
$$\|\lambda_k - \lambda^*\|_2^2 = O\left(\frac{L^2}{k^2}\right)$$
- If $\epsilon > 0$ and all functions ϕ_i are ℓ^2-norms, the convergence is at least linear:
$$\Delta(\lambda_k, q_k) = O\left(\omega^{k/2}\right)$$
and
$$\|\lambda_k - \lambda^*\|_2^2 = O\left(\omega^{k/2}\right)$$
with $\omega = \frac{1}{1+\frac{\sqrt{\gamma\epsilon}}{L}}$.

The proof of this proposition is a straightforward application of results in [4]. These convergence rates can be shown to be optimal for first order methods [19], [21].

Remark 5.1: As Chambolle-Pock's method is first-order, the algorithm depends on the choice of the inner products on the primal and dual spaces. It is possible to define different inner products $\langle\cdot,\cdot\rangle_\Lambda$ and $\langle\cdot,\cdot\rangle_Q$, leading to different iterative methods. This remark can be used to precondition the problem and obtain faster convergence rates, see [5] and paragraph VI-C for more details.

VI. Results

In this section we first present experimental results on images impaired by synthetic noise. We perform comparisons between the proposed algorithm and two recent works dedicated to stripes removal [8], [18]. We also present an example where the noise is synthetized as the sum of $m = 2$ stationary processes. This illustrates our algorithm's ability to cope with more complicated noises. We then present results on real images acquired using different modalities: FIB-nanotomography, SEM and SPIM. We finish the section by showing the experimental convergence behavior of the algorithm.

A. Synthetic noise examples

In order to evaluate the performance of the proposed VSNR algorithm, we synthetise noise according to model (7) and add it to Figure 4. In the first experiment, we consider $m = 1$ noise component, the filter ψ_1 is a straight vertical line running over the entire vertical direction and λ_1 is the realization of a Gaussian white noise process with different variances. The noisy images are then restored using three algorithms: VSNR, the DeStripe algorithm available on http://biodev.cea.fr/destripe/ [8], and the wavelet-FFT algorithm [18]. The wavelet-FFT algorithm implementation was kindly provided by Dr. Beat Münch as a Fiji plugin.

We compare the results of the different restoration algorithms using a rescaled SNR measure (denoted SNRr) defined as follows:

$$SNRr(u_0, u) = SNR(\widetilde{u}_0, u) = 10\log_{10}\left(\frac{\|u\|_2^2}{\|\widetilde{u}_0 - u\|_2^2}\right)$$

where $\widetilde{u}_0 = au_0 - b$ and $(a,b) = \underset{(a,b)\in\mathbb{R}^2}{\arg\min} \|au_0 - b - u\|_2^2$. This measure provides the best SNR among all possible affine transforms of the grey levels of u_0. We chose this measure since the wavelet-FFT algorithm tends to decrease the contrasts.

The results are shown in Table I and in Figure 5. VSNR outperforms the two other methods by a few dBs. Note that when the noise level is low (first line of Table I), both the Destripe and wavelet-FFT methods tend to degrade images. This is due to the fact that the noise description is too rough and this affects signal recovery.

Fig. 4. Original pirate image.

In the second experiment, we consider $m = 1$ noise component, the filter ψ_1 is an elongated Gaussian and λ_1 is the realization of a Gaussian white noise process with different variances. The results are shown in table II and in Figure 6. Once again, VSNR clearly outperforms the two other methods. Figure 6 shows that the algorithm recovers most of the signal even when the noise level is higher than the signal level.

TABLE I
SNRr OF THE IMAGES DENOISED USING DIFFERENT ALGORITHMS. THE NOISE IS OBTAINED BY CONVOLVING WHITE GAUSSIAN NOISE WITH A STRAIGHT LINE. THE THIRD LINE CORRESPONDS TO FIGURE 5.

Initial SNRr	DeStripe	Wavelet - FFT	VSNR
33.85	33.70	22.61	43.00
19.73	20.52	22.03	37.54
8.63	16.60	15.15	25.32

TABLE II
SNRr OF THE IMAGES DENOISED USING DIFFERENT ALGORITHMS. THE NOISE IS OBTAINED BY CONVOLVING WHITE GAUSSIAN NOISE WITH AN ELONGATED GAUSSIAN. THE THIRD LINE CORRESPONDS TO FIGURE 6.

Initial SNRr	DeStripe	Wavelet-FFT	VSNR
32.68	30.64	23.63	34.42
21.80	22.53	18.98	26.31
8.14	8.15	8.29	13.31

In the third experiment, we consider $m = 2$ noise components, the first filter ψ_1 is a radial $sinc$ function, the second filter ψ_2 is an elongated Gaussian in the vertical direction, λ_1 is a sample of a white Gaussian process and λ_2 is a sample of a white Bernoulli process with low probability. The noisy image is shown in Figure 7 and the decomposition result is shown in Figure 8.

Fig. 5. Top-left: noisy pirate. Top-right: denoised using VSNR. Bottom-left: denoised using DeStripe [8]. Bottom-right: denoised using Wavelet-FFT [18]. This experiment corresponds to the last line of Table I. The noise pattern is a vertical line. The Destripe and Wavelet-FFT algorithms do not completely remove low frequencies while VSNR leads to near perfect results.

B. Real images

In order to assess the efficiency of VSNR algorithm, we perform tests on images coming from different modalities:

- FIB-nanotomography. This image was used in [18]. Results are shown in Figure 9.
- SEM imaging. This image was used in [8]. Results are shown in Figure 10.
- SPIM imaging. Results are shown in Figure 11 and 12. The images come from a microscope prototype developed at ITAV-UMS3039 (CNRS).

In all experiments, algorithm 1 is applied with $m = 2$ filters ψ_1 and ψ_2. The first filter ψ_1 is a Dirac (which allows the recovery of white noise), and the second filter ψ_2 is an anisotropic Gabor filter with principal axis directed by the stripes (this orientation is determined by the user). Different parameters were tested for filter ψ_2 and led to similar results, outlining the robustness of this approach with respect to the filter choice. Since no ground truth is available for these images, only visual inspection can help assessing the algorithm's performance. Let us discuss the results:

- Figure 9 and 10 show that VSNR is able to preserve the small image details and to remove the stripes and most of the white noise. The output of the wavelet-FFT and DeStripe algorithms are still impaired with low frequency striping and white noise.
- In Figure 11 only the cells contour is stained. The images are significantly denoised and small details are preserved. This is a particularly hard example as the images are not piecewise constant (i.e. the perfect BV images).

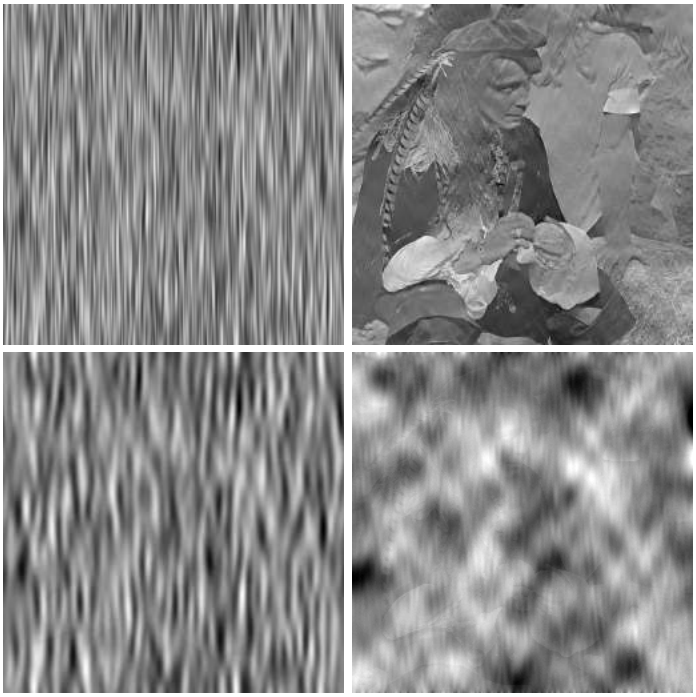


Fig. 6. Top-left: noisy pirate. Top-right: denoised using VSNR. Bottom-left: denoised using DeStripe [8]. Bottom-right: denoised using Wavelet-FFT [18]. This experiment illustrates the third line of Table II. In that example, the noise pattern is an elongated Gaussian in the vertical direction. The noise level is higher that the signal level. The VSNR algorithm retrieves most of the signal while the other two algorithms do not.

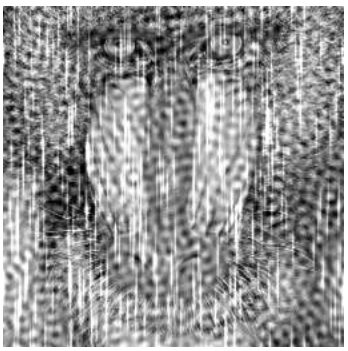


Fig. 7. Synthetic image used for the toy example. PSNR = 21.5dB.

This shows that the proposed noise model is a good approximation of reality.

- Figure 12 is a 3D visualisation of a stack, corresponding to a region of a multicellular tumor spheroid expressing a nuclear fluorescent protein, the H2B-HcRed. The denoising obtained on the stack is significant. Moreover, the 3D reconstruction of some objects (nuclei) performed by extraction of the isosurfaces of fluorescence intensity (in red, Imaris software) is substantially improved after processing.

C. Numerical behavior

In this section, we briefly describe some aspects of the algorithm and of its numerical performance.

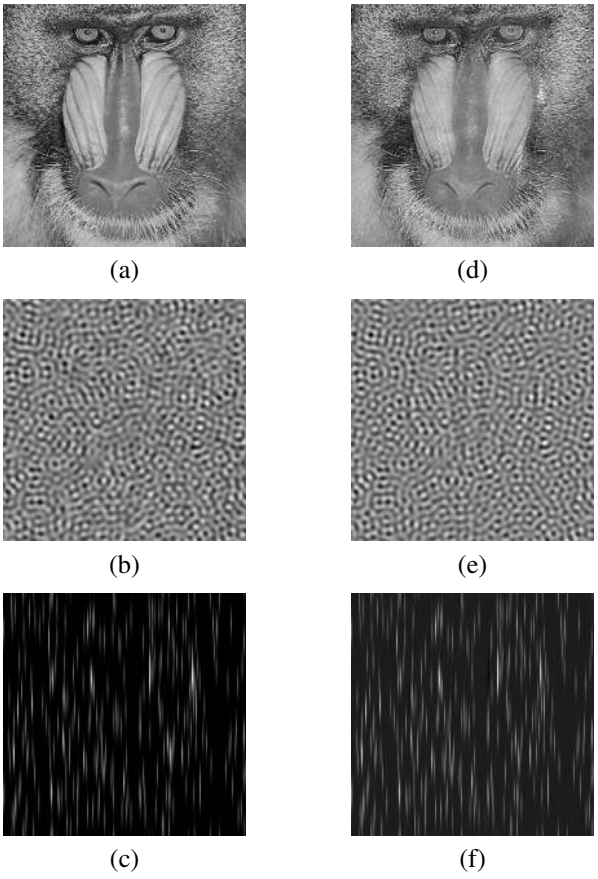


Fig. 8. Toy example. Left column: real components; right column: estimated components using our algorithm. (a,d): Baboon image. PSNR = 27dB. - (b,e): Colored Gaussian noise - (c,f): Impulse like noise

a) Parameter tuning: The convergence rates depend on the parameters in algorithm 1. There are at least 4 parameters that should be tuned:

1) First, it is important to have a tight estimate of L, the highest singular value of A, in order to choose σ and τ as large as possible such that $\sigma\tau L^2 < 1$. An overestimation of L will slow down the scheme and an underestimation will make the scheme diverge. In the case of total variation, the calculation of L can be done explicitly, see appendix B.
2) Second, the relationship $\sigma\tau L^2 < 1$ still leaves a degree of freedom in the choice of σ and τ. It is uneasy to choose these step sizes and it seems that only experimental tuning is available today.
3) Third, as stated in remark 5.1, choosing different inner products in the primal space Λ and dual space Q will lead to different algorithms and can change the numerical performances of algorithm 1 drastically. This question is related to the previous one and we are currently conducting theoretical investigations in order to provide analytical solutions for these choices. A first answer was provided in [5].
4) Finally, it is important to have a reliable stopping criterion in order to automatize the algorithm. In all our experiments, choosing $\epsilon = 10^{-3}$ as a stopping criterion

in algorithm 1 led to solutions that were accurate enough for the visual system (no visible difference with the true solution). We believe that this is a very nice property of the proposed scheme as no user input is necessary.

b) Analytical and empirical complexity: The scheme's analytical complexity $\mathcal{A}$ is given by:

$$\begin{aligned}\mathcal{A} = \quad & (\text{iterations number}) \times 2 \times m \times (\text{FFT}(n) + O(n) \text{ op.}) \\ & + m \times \text{FFT}(n)\end{aligned} \tag{25}$$

where n is the pixels number, m is the number of filters used in the model, FFT(n) indicates an FFT applied to a size n vector and op. means operations. This theoretical convergence rate indicates that the scheme is adapted to large scale problems: the dependence in n is just $O(n \log(n))$.

Our experiments showed that when the parameters in algorithm 1 are chosen correctly (σ, τ and the metrics), the scheme requires less than 50 iterations in order to decrease the initial duality gap by a factor $\epsilon = 10^{-3}$. The overall cost is thus around $100 \times m$ FFT computations. The computational time is around 5 seconds for a 1024×1024 image, on a 1.2GHz laptop, using $m = 1$ filter. This can be accelerated using parallel computing.

A more formal complexity result could be obtained using the worst case convergence rates of the algorithm, but this rate seems to be always outperformed for the class of problems considered in this paper.

c) A heuristic choice of the metrics: All the tools used in algorithm 1 (Adjoints, singular values, Fenchel transforms, subgradients, proximal operators,...) depend on the choice of the inner products $\langle \cdot, \cdot \rangle_\Lambda$ and $\langle \cdot, \cdot \rangle_Q$. In [5] some choices of the metrics are proposed in order to improve the convergence rate. In the present work, we simply rescaled the filters ψ_i and weights α_i. This normalization is equivalent to a block-wise constant diagonal preconditionner.

VII. Conclusion

We proposed a variational approach to denoise images impaired by a class of stationary noises. It takes advantage of recent advances in convex optimization, leading to interactive computing times. Applications to synthetic images, and to images issued from different microscopy imaging modalities were presented, leading to clear improvements of the image quality compared to existing methods.

Appendix A
Explicit expressions of the resolvents

In this section, we assume that $\langle \cdot, \cdot \rangle_\Lambda$ and $\langle \cdot, \cdot \rangle_Q$ are the standard dot products. The Fenchel conjugates F^* and G^* can be computed using Equation (6):

$$F^*(q) = \begin{cases} \frac{\epsilon}{2}\|q\|_2^2 - \langle \nabla u_0, q \rangle_Q & \text{if } \|q\|_\infty \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

and

$$G^*(\eta) = \sum_{i=1}^{m} (\phi_i + \chi_{[-C,C]^n})^* (\eta_i).$$

The adjoint of A is $A^* = A^T$, where:

$$\begin{aligned} A^T : \quad \mathbb{R}^{n \times 2} &\to \mathbb{R}^{m \times n} \\ q &\mapsto \left(\nabla^T q * \check{\psi}_1, \cdots, \nabla^T q * \check{\psi}_m\right) \end{aligned}$$

and $\check{\psi}_i = \mathcal{F}^{-1}\psi_i$.

The different Fenchel transforms useful for the computation of G^* are:

- If $\phi_i(\lambda) = \alpha_i \|\lambda\|_1$:
$$(\phi_i + \chi_{[-C,C]^n})^*(\eta) = C\| \max(0, |\eta| - \alpha_i)\|_1.$$
- If $\phi_i(\lambda) = \frac{\alpha_i}{2}\|\lambda\|_2^2$:
$$\begin{aligned}&(\phi_i + \chi_{[-C,C]^n})^*(\eta) \\ &= \alpha \left\| \min\left(\left|\frac{\eta}{\alpha}\right|, C\right) \left|\frac{\eta}{\alpha}\right| - \frac{1}{2}\min\left(\left|\frac{\lambda}{\alpha}\right|, C\right)^2 \right\|_1\end{aligned}$$
- If $\phi_i(\lambda) = \chi_{[-\alpha_i, \alpha_i]}(\lambda)$:
$$(\phi_i + \chi_{[-C,C]^n})^*(\eta) = \min(\alpha_i, C)\|\eta\|_1$$

1) Resolvent of F^*: Let us detail the calculation of:

$$\begin{aligned} q^* &= (\text{Id} + \sigma \partial F^*)^{-1}(q_n) \\ &= \underset{q \in Q, \|q\|_\infty \leq 1}{\arg\min} \frac{\sigma\epsilon}{2}\|q\|_2^2 - \langle \sigma \nabla u_0, q \rangle \\ &\quad + \frac{1}{2}\|q - q_n\|_Q^2. \end{aligned} \tag{26}$$

Let

$$\tilde{q}_n = q_n + \sigma \nabla u_0$$

By writing the Karush-Kuhn-Tucker optimality conditions, we obtain the following solution to problem (26):

$$q^*(\mathbf{x}) = \frac{\tilde{q}_n(\mathbf{x})}{\max(|\tilde{q}_n(\mathbf{x})|, \sigma\epsilon + 1)}$$

2) Resolvent of G: In the second step of the algorithm, we must compute

$$\begin{aligned} \lambda^* &= (I + \tau \partial G)^{-1}(\lambda_n) \\ &= \underset{\lambda \in \Lambda}{\arg\min}\, \tau G(\lambda) + \frac{1}{2}\|\lambda - \lambda_n\|_\Lambda^2 \\ &= \underset{\lambda \in \Xi}{\arg\min} \sum_{i=1}^{m} \phi_i(\lambda_i) + \frac{1}{2\tau}\|\lambda_i - \lambda_{n,i}\|_{\Lambda_i}^2 \end{aligned}$$

As all the functions ϕ_i are separable, this problem reduces to $m \times n$ unidimensional problems of form:

$$\underset{\lambda \in \mathbb{R}, |\lambda| \leq C}{\arg\min}\, \tau f(\lambda) + \frac{1}{2}|\lambda - \lambda_n|^2 \tag{27}$$

where f is a convex function. In this work we focus on the cases:

- $f(\lambda) = \alpha|\lambda|$.

The solution of (27) is given by:

$$\lambda^* = \frac{\max(|\lambda_n| - \tau\alpha, 0) \cdot \text{sign}(\lambda_n)}{\max(1, \max(|\lambda_n| - \tau\alpha, 0)/C)}.$$

- $f(\lambda) = \alpha\frac{\lambda^2}{2}$.

  The solution of (27) is given by:

$$\lambda^* = \frac{\frac{\lambda_n}{\tau\alpha+1}}{\max(1, |\frac{\lambda_n}{\tau\alpha+1}|/C)}.$$

- $f(\lambda) = \begin{cases} 0 & \text{if } |\lambda| \leq \alpha \\ +\infty & \text{otherwise} \end{cases}$

  Let $\delta = \min(C, \alpha)$. The solution of (27) is given by:

$$\lambda^* = \frac{\lambda_n}{\max(1, |\lambda_n|/\delta)}.$$

APPENDIX B
COMPUTATION OF THE OPERATOR NORM L

In order to Algorithm 1 as fast as possible, it is important to use a tight estimation of L, the highest singular value of A defined in equation (18). In the case of total variation, L can be computed exactly if periodic boundary conditions are used to define the discrete gradient operator.

With these boundary conditions, the discrete gradient operator $\nabla = \begin{pmatrix} \partial_1 \\ \partial_2 \end{pmatrix}$ can be rewritten using convolution products and is thus diagonalized by the discrete Fourier tranform (DFT):

$$\partial_1 u = d_1 * u = \mathcal{F}^{-1}\mathrm{diag}(\hat{d_1})\mathcal{F}u$$

and

$$\partial_2 u = d_2 * u = \mathcal{F}^{-1}\mathrm{diag}(\hat{d_2})\mathcal{F}u$$

where d_1 and d_2 are finite difference filters in the horizontal and vertical directions.

Let H_i be the matrix associated to the convolution product with ψ_i. It satisfies $H_i\lambda_i = \psi_i * \lambda_i$ and is thus diagonalized by the DFT:

$$H_i = \mathcal{F}^{-1}\mathrm{diag}\left(\hat{\psi_i}\right)\mathcal{F}. \tag{28}$$

Let us denote $D_i = \mathrm{diag}\left(\hat{d_i}\right)$, $|D_i|^2 = \bar{D}_i D_i$ where $\bar{D}_i$ is the complex conjugate of D_i and $\Sigma = \sum_{i=1}^m \mathrm{diag}\left(|\hat{\psi_i}|^2\right)$. Elementary calculus then leads to:

$$AA^T = \begin{bmatrix} \mathcal{F}^{-1} & 0 \\ 0 & \mathcal{F}^{-1} \end{bmatrix} \cdot \begin{bmatrix} |D_1|^2\Sigma & D_1\Sigma D_2^* \\ D_2\Sigma D_1^* & |D_2|^2\Sigma \end{bmatrix} \cdot \begin{bmatrix} \mathcal{F} & 0 \\ 0 & \mathcal{F} \end{bmatrix}.$$

The eigenvalues of AA^T are the same as those of

$$\begin{bmatrix} |D_1|^2\Sigma & D_1\Sigma D_2^* \\ D_2\Sigma D_1^* & |D_2|^2\Sigma \end{bmatrix}.$$

This matrix is symmetric, positive, semi-definite and is constituted of four diagonal blocks. L is thus given by the maximum largest eigenvalue of the following n, 2×2 matrices:

$$M(\mathbf{k}) = \begin{bmatrix} |\hat{d_1}(\mathbf{k})|^2\Sigma(\mathbf{k}) & \hat{d_1}(\mathbf{k})\overline{\hat{d_2}(\mathbf{k})}\Sigma(\mathbf{k}) \\ \overline{\hat{d_1}(\mathbf{k})}\hat{d_2}(\mathbf{k})\Sigma(\mathbf{k}) & |\hat{d_2}(\mathbf{k})|^2\Sigma(\mathbf{k}) \end{bmatrix},$$

where $\mathbf{k}$ belongs to the frequency domain and $\Sigma(\mathbf{k}) = \sum_{i=1}^m \mathrm{diag}\left(|\hat{\psi_i}(\mathbf{k})|^2\right)$. This computation is achieved in $O(n \log n)$ arithmetic operations.

ACKNOWLEDGMENT

The authors would like to thank Bernard Ducommun, Raphaël Jorand and Valérie Lobjois from the IP3D team in Toulouse cancéropole for their tireless support during this work and for all SPIM related questions. They also thank François De Vieilleville for fruitful discussions. They thank Beat Münch for providing a Fiji implementation of the Wavelet-FFT algorithm and test images and Jean-Luc Pellequer for providing us images. Jérôme Fehrenbach was partly supported by a grant from CNRS.

REFERENCES

[1] A. Arrenberg AB, D. Stainier, H. Baier H, and J.Huisken, "Optogenetic control of cardiac function," *Science*, vol. 330(6006), pp. 971–974, 2010.
[2] J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher, "Structure-Texture Image Decomposition - Modeling, Algorithms, and Parameter Selection," International Journal of Computer Vision, vol. 67, no. 1, pp. 111–136, April 2006.
[3] H. Carfantan and J. Idier, "Statistical linear destriping of satellite-based pushbroom-type images," IEEE Transactions on Geoscience and Remote Sensing, 48,1860–1871, 2010.
[4] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Mathematical Imaging and Vision*, vol. 40, no.1, pp. 120–145, 2011.
[5] T. Pock and A. Chambolle, "Diagonal preconditioning for first order primal-dual algorithms in convex optimization," *ICCV*, 2011.
[6] J. Chen, Y. Shao, H. Guo, W. Wang, B. Zhu, "Destriping CMODIS Data by Power Filtering," *IEEE Trans. Geosci. Remote Sens.* vol. 41, pp. 2119–2124, 2003.
[7] J. Chen, H. Lin, Y. Shao, L. Yang, "Oblique Striping Removal in Remote Sensing Imagery Based on Wavelet Transform," *Int. J. Remote Sens.* vol. 27, pp. 1717–1723, 2006.
[8] S. Chen and J.L. Pellequer, "DeStripe: frequency-based algorithm for removing stripe noises from AFM images" *BMC Structural Biology*, vol. 11, no.7, 2011.
[9] S.S. Chen, D.L. Donoho and M.A. Saunders, "Atomic decomposition by basis pursuit", SIAM review, 129–159, 2001.
[10] F. Facchinei and J. Pang, "Finite-dimensional variational inequalities and complementarity problems", Springer Verlag, 2003.
[11] M.J. Fadili, J.-L. Starck, J. Bobin, Y. Moudden, "Image decomposition and separation using sparse representations: an overview," *Proc. of the IEEE, Special Issue: Applications of Sparse Representation*, vol. 98, no. 6, pp. 983–994, 2010.
[12] J. Fehrenbach, P. Weiss and C. Lorenzo, "Variational algorithms to remove stripes: a generalization of the negative norm models," *Proc. ICPRAM*, 2012.
[13] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. Stelzer, "Optical Sectioning Deep Inside Live Embryos by Selective Plane Illumination Microscopy," *Science*, vol. 305, no. 5686, pp. 1007–1009, 2004.
[14] P. Keller, A. Schmidt, A. Santella, K. Khairy, Z. Bao, J. Wittbrodt, and E. Stelzer, "Fast, high-contrast imaging of animal development with scanned light sheet-based structured-illumination microscopy," *Nat. Methods*, vol. 7, no. 8, pp. 637–642, 2010.
[15] F. Kienberger, VP Pastushenko, G. Kada, T. Puntheeranurak, L. Chtcheglova, C. Riethmueller, C. Rankl, A. Ebner, P. Hinterdorfer I "Improving the contrast of topographical AFM images by a simple averaging filter." *Ultramicroscopy*, vol. 106, pp. 822–828, 2006.
[16] S. Mallat, "A Wavelet Tour of Signal Processing: The Sparse Way." 2008, Academic Press.
[17] Y. Meyer, "Oscillating Patterns in Image Processing and Nonlinear Evolution Equations," The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures, American Mathematical Society, Boston, MA, 2001
[18] B. Münch, P. Trtik, F. Marone, M. Stampanoni "Stripe and ring artifact removal with combined wavelet-Fourier filtering." *Opt. Express*,vol. 17, pp. 8568–8591, 2009.
[19] A. Nemirovski, "Information-based complexity of linear operator equations," *Journal of Complexity*, vol. 8, pp. 153–175, 1992.
[20] A. Nemirovski, "Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM J. Optimization*, vol. 15, no. 1, pp. 229–251, 2005.

[21] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004.
[22] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–152, 2005.
[23] S. Preibisch, S. Saalfeld, J. Schindelin, and P. Tomancak, "Software for bead-based registration of selective plane illumination microscopy data," Nat. Methods, vol. 7, no. 6, pp. 418–419, 2010.
[24] R.T. Rockafellar, *Convex Analysis*, Princeton University Press 1970.
[25] A. Shiryaev, *Probability Second Edition*, Springer, Graduate in Mathematics 95.
[26] J.-B. Hiriart-Urruty and C. Lemarechal, "Convex Analysis and Minimization Algorithms", Vol. 2, Springer Verlag, Berlin, Germany, 1993
[27] P. Weiss, L. Blanc-Féraud and G. Aubert, *Efficient schemes for total variation minimization under constraints in image processing*, SIAM Journal on Scientific Computing, 31, 3, 2047–2080, 2009.

Jérôme Fehrenbach received the PhD degree in mathematics from University of Nice-Sophia Antipolis in 1998. He is with Institut de Mathématiques de Toulouse in France. His current research interest are inverse problems and image processing.

Pierre Weiss graduated from INSA de Toulouse in 2004. He received a PhD degree in signal processing from the University of Nice-Sophia Antipolis in 2008 and did a post-doc at Hong Kong Baptist University in 2009. Since 2010, he serves as an associate professor at INSA de Toulouse and is a member of Institut de Mathématiques de Toulouse. His research interests are in the field of imaging and applied mathematics. It includes inverse problems, convex optimization and variational methods applied to topics as image restoration, change detection, MRI or SPIM.

Corinne Lorenzo , PhD, graduated from the university of Toulouse in 2004. She is a member of IP3D group, ITAV-UMS3039 (Toulouse, France) and she is in charge of the SPIM imaging developement.

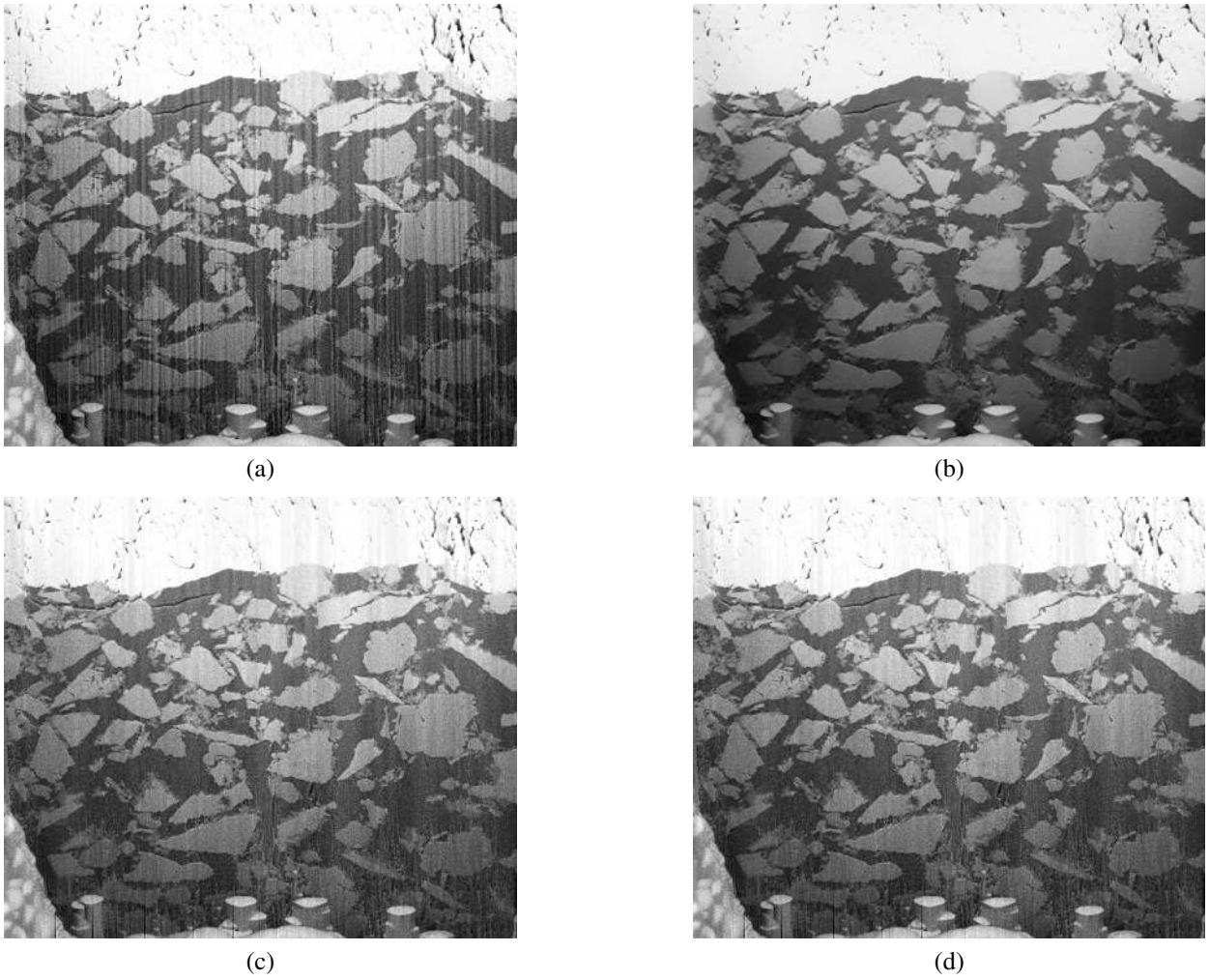


Fig. 9. Destriping of a single section from a 3D volume of unhydrated particles of cement paste accessed by FIB-nanotomography. (a) original image is severely affected by the waterfall effect. (b) restored image using VSNR. (c) restored image using wavelet-FFT. (d) restored image using DeStripe.

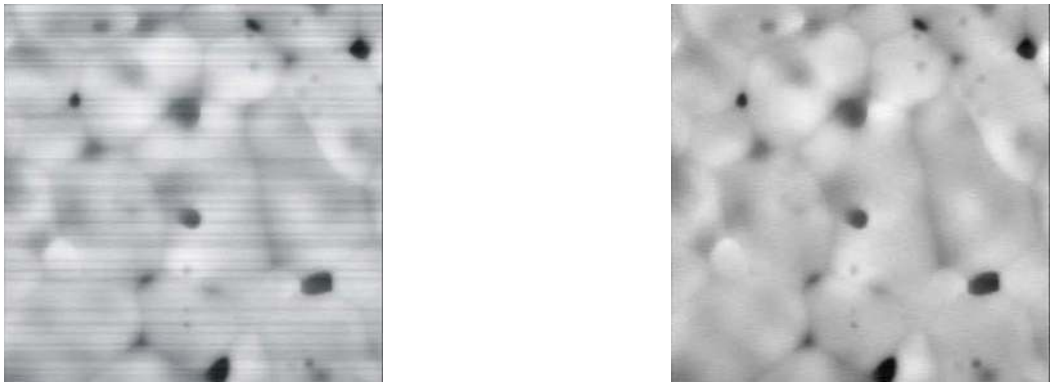
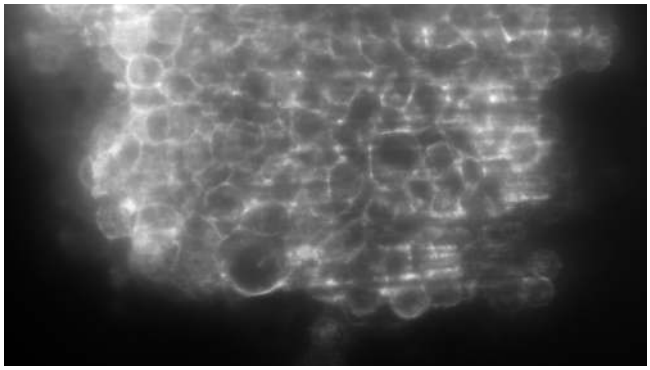
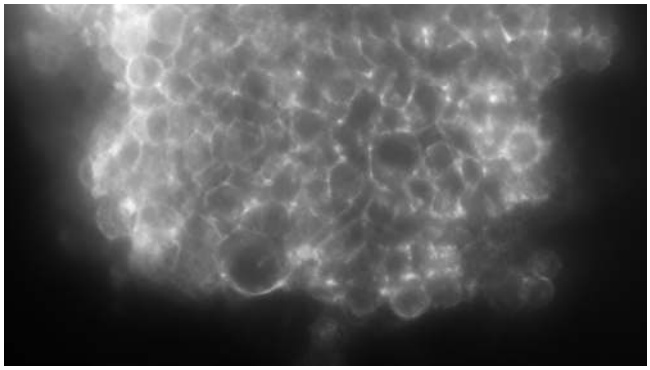


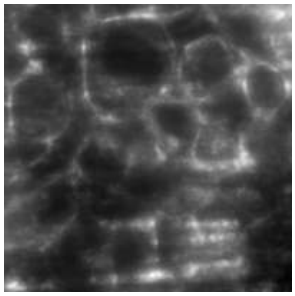
Fig. 10. SEM imaging on a sintered specimen of CeO2 (cerium oxide) at a sintering temperature of 1400 C for 2 hours and 50 minutes [8]. The image is composed of 512x512 pixels with each side measuring 70.55 μm; the intensity unit is 0.1 μm.



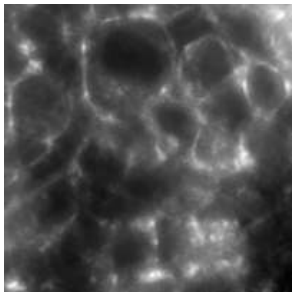
(a)



(b)

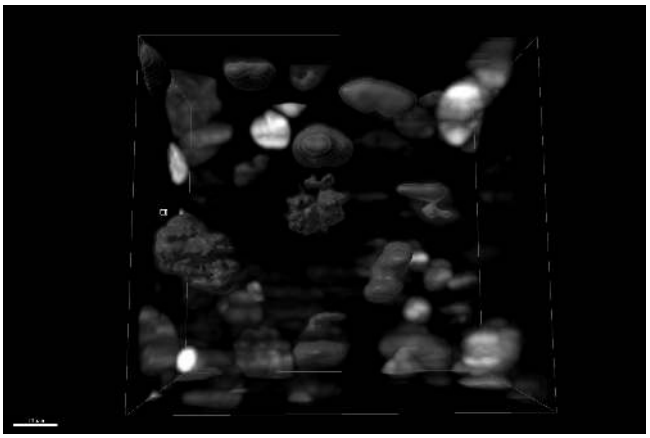


(c)

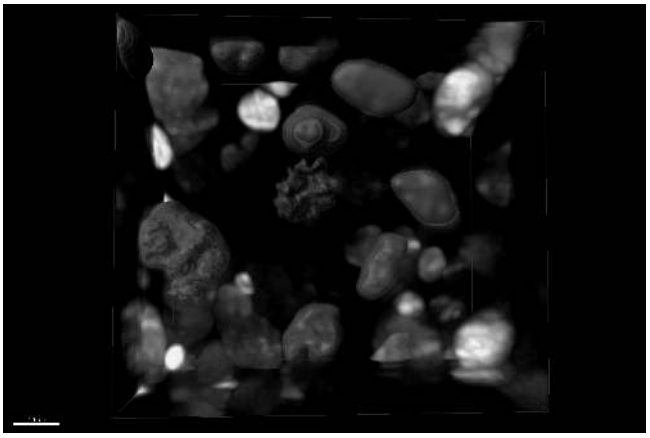


(d)

Fig. 11. SPIM images of a Multicellular Tumor Spheroid stained with DiI acquired with an objective 20X NA 0.5, an excitation wavelength of 532 nm and an emission wavelength of 593 nm. The voxel size is 0.32*0.32*1 μm. (a) Single plane of a 3D stack; (c) A magnified view of a region of (a); (b) Denoised image; (d) A magnified view of the same region of (b).



(a)



(b)

Fig. 12. Isosurface rendering (in red) of a three- dimensional stack of 42 planes of a Multicellular Tumor Spheroid expressing a fluorescent nuclear protein, H2B-HcRed. SPIM images were acquired with an objective 10X NA 0.25, an excitation wavelength of 595 nm and a detection using a 593 nm long pass filter (a) raw data (b) denoised data. The voxel size is 0.645*0.645*1μm.