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

Jérôme Fehrenbach, Pierre Weiss, Corinne Lorenzo

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

Jérôme Fehrenbach, Pierre Weiss, and Corinne Lorenzo

***Abstract*—In the present paper, a framework and an algorithm are presented in order to remove stationary noise from images. This algorithm can be seen 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 cases where the white noise assumption is replaced by a stationary noise assumption. An application to an emerging fluorescence microscopy technique (SPIM: Selective Plane Illumination Microscope) is presented, where the adequate noise modeling allows to neatly improve the image quality and provides much better results than methods designed for white noise.**

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

I. Introduction

The main purpose of this work is to propose a general noise or texture model that is both numerically tractable, and describes complex real-life situations.

Many image formation models can be stated as follows:

$$u_0 = Hu + b$$

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, 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:

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

Jérôme Fehrenbach and Pierre Weiss are with the Institut de Mathématiques de Toulouse, CNRS and University of Toulouse, France.

Corinne Lorenzo is with ITAV-UMS3039 (CNRS).

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 saying that there is no a-priori knowledge about location of any feature or noise pattern in the image. This can be seen as an axiomatic approach to noise modelling as in [1]. However, the general 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: the 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 Figures 2, 3, 4.

This work was initially motivated by the recent development of a microscope called Selective Plane Illumination Microscope (SPIM). The SPIM is a fluorescence microscope which performs optical sectioning of a specimen, see [12] for more details. It allows to do long term 3D imaging of thick specimens and seems to have no rival for many biological applications [2], [13], [23]. One of the differences with conventional microscopy is that the fluorescence light is detected at an angle of 90 degrees with the illumination axis. This procedure leads to the generation of stripe artifacts that can be seen as irregular lines aligned with the illumination axis see Figures 8 (a,d), 10 (a,c) and 11 (a). They strongly impair image interpretation. One of the applications of our formalism is to model and remove these stripes using variational principles.

Even though this work was initially devoted to the SPIM microscope, it can be useful in a variety of applications, where stationary patterns must be removed or detected. For instance, stripes appear in many recent imaging modalities as Atomic Force Microscope (AFM) [8], [14], electron microscopy and Synchrotron X-Ray Tomographic images [16], or in remote sensing imaging such as Moderate-resolution Imaging Spectroradiometer (MODIS) images or digital elevation models [6], [7].

The proposed model shares many similarities with the negative norm models of Y. Meyer [15], its successors [3], [11] and the decomposition algorithms proposed in [10], [26]. Meyer's idea is to decompose an image into a piecewise smooth component and an oscillatory component. We refer to [3] for a review of the principles, algorithms and results using this approach. An alternative way of decomposing images was proposed in [26]. The idea is to seek for components that are sparse in given dictionaries. Different choices for the elementary atoms composing the dictionary will allow to recover different kind of textures. See [10] for a review of these methods and a generalization to the decomposition into

an arbitrary number of components.

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 [10], 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. It sheds a new light on the negative norm models and allows to interpret them as maximum a posteriori approaches.
5) The decomposition problem is recast into a convex optimization problem that is solved with a recent algorithm of A. Chambolle and T. Pock [5] allowing to obtain results in an interactive time.
6) We propose a C and Matlab implementation on our webpage in the spririt 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 and tools helpful for the paper understanding. 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 likelihood estimator. Section V details an efficient numerical scheme based on [5] 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 $\Lambda = \mathbb{R}^{n \times n}$ and $Q = \mathbb{R}^{n \times 2}$. These Euclidean 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 make use of the notation $\hat{u} = \mathcal{F}u$. Finally, ∇ denotes the discrete gradient operator. In this article we use the discretization proposed in [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:

$$\phi^*(\lambda_2) = \sup_{\lambda_1 \in \Lambda} \langle \lambda_1, \lambda_2 \rangle_\Lambda - \phi(\lambda_1).$$

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 γ, if it is convex and that the following inequality holds for all $(\lambda_1, \lambda_2) \in \mathrm{dom}(\phi)$ and for all $\eta \in \partial\phi(\lambda_1)$:

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

The resolvent of ϕ at point λ is defined by:

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

We refer to [24] for a complete introduction of the above tools.

III. Noise model

In most denoising methods, the noise is supposed to be white. However this assumption does not hold in some applications. For instance, Figures 5 (b,e) and 8 (a,d) are perturbed by a directional noise, which is clearly not pixelwise independent.

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 can be written as:

$$b = \sum_{i=1}^{m} \lambda_i * \psi_i, \tag{4}$$

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.

This class of noises is *stationary* - in the sense that it satisfies Equation (1) - as shown below.

Proposition 3.1: Let L be a random vector in $\mathbb{R}^n$, the components of which are i.i.d. Let $B = h * L$ be a random vector where h is a convolution kernel. Then B is stationary. The proof of this proposition can be found in [25], page 404.

Proposition 3.2: The class of noises described by Equation (4) is stationary.

Proof: It suffices to use proposition 3.1 and the fact that the sum of two independent stationary processes is stationary. ■

B. Justification of the noise model in SPIM imaging

In light-sheet microscopy, images 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

$$b = \lambda * \psi, \tag{5}$$

where ψ is an elementary pattern, similar to a stripe, and λ describes the positions and intensities of the pattern. We assume that if ψ is chosen adequately, then λ can be modeled as a sample of a white noise process.

In realistic applications, the noise is seldom describable simply as in Equation (5). A more realistic noise description would be the sum of processes of type (5). 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 (5) where ψ is a Dirac function and λ is a sample of a Gaussian white noise process. In this work, we will make experiments with Gaussian noise, but the extension to Poisson noise is quite straightforward.

The common feature of the stripes in SPIM is their orientation θ, as well as their dimensions (width and length) that belong to a restricted range. Some stripes could also have a frequency that characterizes their spatial oscillations. One given stripe will thus be modeled by a Gabor function, which is a 2-dimensional Gaussian modulated by a plane sinusoidal wave. Its expression in space domain is

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

where $x' = x\cos\theta + y\sin\theta$ and $y' = -x\sin\theta + y\cos\theta$.

In Fourier domain, the expression of this Gabor function is the convolution of a Gaussian and the Fourier transform of the plane wave (two Dirac masses).

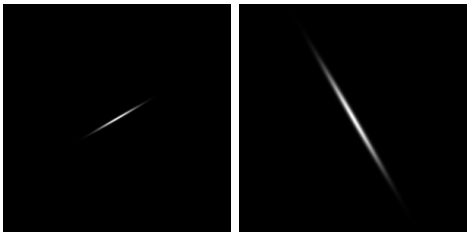


Fig. 1. Left: a Gabor function in space domain – Right: the same Gabor function in Fourier domain.

In order to describe stripes of different width or length, several Gabor filters $\psi_1, \ldots, \psi_m$ with different geometrical parameters can be used. As a summary we propose a description of the stripes as in Equation (4) where the λ_i are white noise realizations that will be described later and the ψ_i are Gabor filters with the same orientation θ and different geometrical parameters.

IV. Restoration/decomposition model

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

$$u_0 = Hu + b \tag{6}$$

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 $H = \mathrm{Id}$ for simplicity, however all the proposed algorithms extend in a straightforward manner to $H \neq \mathrm{Id}$.

A. A MAP reconstruction approach

The maximum a posteriori (MAP) approach in a Bayesian framework usually leads to retrieve the image u that maximizes the likelihood probability

$$\mathbf{p}(u|u_0) = \frac{\mathbf{p}(u_0|u)\mathbf{p}(u)}{\mathbf{p}(u_0)}.$$

In our setting, it is more natural to estimate the weights $\lambda = \{\lambda_i\}_{i=1..m}$ since we assume that these processes have a known probability density functions. The problem we propose is to maximize the following probability:

$$\mathbf{p}(\lambda|u_0) = \frac{\mathbf{p}(u_0|\lambda)\mathbf{p}(\lambda)}{\mathbf{p}(u_0)}.$$

Maximizing $\mathbf{p}(\lambda|u_0)$ amounts to minimize $-\log \mathbf{p}(\lambda|u_0)$. The restoration problem is thus expressed as:

$$\text{Find } \lambda^* \in \operatorname*{Arg\,min}_{\lambda\in\Lambda} -\log \mathbf{p}(u_0|\lambda) - \log \mathbf{p}(\lambda). \tag{7}$$

As we assumed independence of the λ_is,

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

and

$$\begin{aligned}\mathbf{p}(u_0|\lambda) &= \mathbf{p}(Hu + \sum_{i=1}^{m} \lambda_i * \psi_i|\lambda) \\ &= \mathbf{p}(Hu)\end{aligned}$$

since we assume independence of λ and u.

Remark 4.1: In some applications the independence hypothesis is questionable and a multiplicative model could be considered. However our numerical experiments using such multiplicative models did not improve the results sensibly compared to the additive models and made the computations much more involved.

B. Image prior

We need to devise a prior $\mathbf{p}(Hu)$ on the space of observations or a p.d.f. $\mathbf{p}(u)$ on the space of images (see remark 4.4) in order to specify problem (7).

In this work, we use the standard assumption that images are smooth or piecewise smooth. This can be promoted by standard p.d.f. of the form $\mathbf{p}(u) \propto \exp(-\alpha J_R(u))$ with:

$$J_R(u) = \varphi(Bu).$$

In this equation, B is a linear operator and φ is a positive convex function. Typically $B = \nabla$ (the discrete gradient operator) or a discrete wavelet transform and φ is a norm or a regularized norm.

In our experiments we use a quantity $J_R(u)$ that quantifies the spatial regularity of the image:

$$J_R(u) = \|\nabla u\|_{1,\epsilon}$$

with:

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

and

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

Note that this function can be rewritten using duality as:

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

The interest of setting $\varphi(\cdot) = \|\cdot\|_{1,\epsilon}$ 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 schemes designed to solve (7) converge faster than for $\epsilon = 0$ and give similar practical results.

Remark 4.2: it is straightforward to use sparsity priors in the proposed formalism and we give the resulting optimization problems for the sake of completeness in equation (11). In the numerical experiments, we restricted our attention to total variation like priors for the sake of simplicity.

C. Noise priors

The p.d.f. $\mathbf{p}(\lambda_i)$ in equation (7) 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 this paper we considered three cases:

- *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. A typical example of such noise is given in Figure 2.

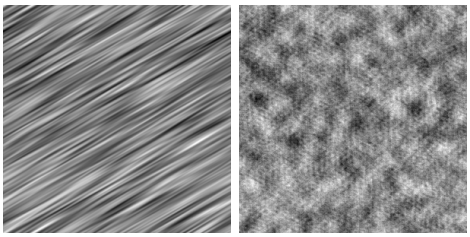


Fig. 2. Examples of Gaussian colored noises. Left: noise generated by convolving a Gabor filter with a Gaussian white noise. Right: noise generated by convolving a fish pattern with the same Gaussian white noise.

- *Laplace noise*. This hypothesis consists in setting $\mathbf{p}(\lambda_i) \propto \exp(-\alpha_i\|\lambda_i\|_1)$. This distribution is know to be a good convex approximation of Bernoulli processes. The corresponding noise component $b_i = \lambda_i * \psi_i$ is thus close to a sparse signal in the dictionary composed of pattern ψ_i translated in space. An example of such a process is given in Figure 3.

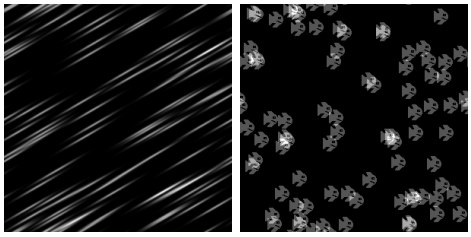


Fig. 3. Left: noise generated by convolving a Gabor filter with a Bernoulli process. Right: noise generated by convolving a fish pattern with the same Bernoulli process.

- *Uniform noise*. This hypothesis consists in 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. An example of such process is given in Figure 4.

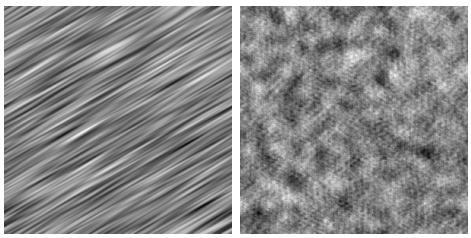


Fig. 4. Left: noise generated by convolving a Gabor filter with a uniform white noise. Right: noise generated by convolving a fish pattern with the same uniform white noise.

Remark 4.3: Notice that the images in Figure 2 and 4 look very similar. This can be explained using an extension of the central limit theorem and has important practical consequences. At each pixel, the result of a convolution product between an elementary pattern and a white noise can be interpreted as a weighted sum of independent random variables. Lindeberg-Feller theorem's [22] states that if the random variables have finite variance, the pixelwise p.d.f. of b in equation (5) tends to a Normal distribution as the sampling step goes to 0, whatever the p.d.f. of the white noise. Berry-Esseen's theorem [22] states that the convergence rate in infinite-norm to a normal distribution is of order $O\left(\frac{C}{\sqrt{n}}\right)$ where n is the convolution kernel's support size and C depends on the second and third order moments of the process marginals. As a consequence: if the convolution kernel has sufficiently slow spatial decrease, the p.d.f. of b in equation (5) will be approximately the same whatever the marginal distribution of λ. In practice it is much more convenient to assume that λ is a Gaussian process for numerical purposes. Indeed, Gaussian priors lead to the minimization of strongly convex functions which can be achieved much more efficiently using the iterative schemes we propose in this paper (see proposition 5.3).

D. Resulting optimization problems

We stated every necessary details to state the restoration problem completely. In the case $H = \mathrm{Id}$, which is the one used in our experiments, it reads:

$$\text{Find} \quad \lambda \in \arg\min_{\lambda \in \Lambda} P(\lambda). \tag{8}$$

where

$$P(\lambda) = \left|\left| \nabla \left(u_0 - \sum_{i=1}^{m} \lambda_i * \psi_i \right) \right|\right|_{1,\epsilon} + \sum_{i=1}^{m} \alpha_i \phi_i(\lambda_i), \tag{9}$$

and

$$\lambda = \{\lambda_i\}_{i=1}^m.$$

Remark 4.4: The model can be extended to cases where H is not the identity. In order to avoid using the inverse operator H^{-1}, it suffices to add an extra variable u and solve the linearly constrained problem:

$$\begin{array}{rl} \min & \|\nabla u\|_{1,\epsilon} + \sum_{i=1}^m \alpha_i \phi_i(\lambda_i) \\ \text{subject to} & u \in \mathbb{R}^n \\ & \lambda \in \mathbb{R}^{n \times m} \\ & Hu = u_0 - \sum_{i=1}^m \lambda_i * \psi_i \end{array} \tag{10}$$

The resolution of (10) can be achieved using the algorithm proposed later.

Remark 4.5: Sparsity priors would lead to problems of kind:

$$\begin{array}{rl} \min & \|\gamma\|_1 + \sum_{i=1}^m \alpha_i \phi_i(\lambda_i) \\ \text{subject to} & \gamma \in \mathbb{R}^p \\ & \lambda \in \mathbb{R}^{n \times m} \\ & u_0 - \sum_{i=1}^m \lambda_i * \psi_i = HS\gamma \end{array} \tag{11}$$

where $S : \mathbb{R}^p \to \mathbb{R}^n$ is a dictionnary or a synthesis operator and γ are the coefficients of u in this dictionnary.

Remark 4.6: The proposed model turns out to be equivalent to the one proposed in [3] in the case of $m = 1$ component and $\phi_1(\lambda_1) = \frac{\alpha}{2}\|\lambda_1\|_2^2$. In this work, the authors propose to solve:

$$\min_{u \in \mathbb{R}^n} J_R(u) + \frac{\alpha}{2}\|u - u_0\|_K^2$$

where $\|u\|_K^2 = \langle u, Ku \rangle$ is a Hilbert norm defined with a symmetric positive definite (SPD) convolution operator K. Thus, there exists h such that $\sqrt{K}u = h * u$. Let us denote $\psi = \mathcal{F}^{-1}\frac{1}{\mathcal{F}h}$ and $\lambda = h * (u - u_0)$ then:

$$\begin{aligned} &\min_{u \in \mathbb{R}^n} J_R(u) + \frac{\alpha}{2}\|u - u_0\|_K^2 \\ =&\min_{u \in \mathbb{R}^n} J_R(u) + \frac{\alpha}{2}\|h * (u - u_0)\|_2^2 \\ =&\min_{\lambda \in \mathbb{R}^n} J_R(u_0 - \psi * \lambda) + \frac{\alpha}{2}\|\lambda\|_2^2. \end{aligned}$$

This last line is exactly equation (9). In this particular case our models are equivalent, however our framework allows ψ to vanish in the Fourier domain.

As a matter of fact, in the discrete setting, the model proposed here can be seen as generalization of the negative norm texture+cartoon decomposition models [15]. This will be shown it in a forthcoming research report.

V. Numerical algorithms

We completely specified problem (7) and we can now describe a numerical scheme to solve it.

A. Problem relaxation

In the case of Tikhonov regularization (which corresponds to $\phi_i(\lambda) = \frac{\alpha_i}{2}\|\lambda_i\|_2^2$ and $J_R(u) = \|\nabla u\|_2^2$), the above problem 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 and we need to design iterative methods that will lead to approximate solutions.

The objective of this section is to provide a numerical scheme to solve the following natural relaxation of problem (8):

$$\text{Find } \lambda_\epsilon \in \Xi \text{ such that } P(\lambda_\epsilon) - P(\lambda^*) \leq \epsilon \tag{12}$$

where

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

This constrained problem is convex. We will show that adding the constraint $\lambda \in \Xi$ allows to design reliable stopping criteria to ensure $P(\lambda_\epsilon) - P(\lambda^*) \leq \epsilon$. Moreover if C is large enough, the extra box-constraint is inactive and the solutions of (12) can be shown to be the same as those of:

$$\text{Find } \lambda_\epsilon \in \Lambda \text{ such that } P(\lambda_\epsilon) - P(\lambda^*) \leq \epsilon. \tag{13}$$

Problem (12) is a generic convex-concave saddle-point problem. Many algorithms are proposed in the literature to solve these kinds of problems (see e.g. [9] for an overview of existing methods). Recently, Nesterov paved the way to the development of new efficient algorithms ([4], [17], [20], [21]). 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, [5] is probably the most versatile and we decided to present and implement this strategy.

B. Reformulation as a saddle-point problem

This section is devoted to the resolution of (12). It is straightforward to extend the algorithms to problems (10) and (11). In order to apply the ideas presented in [5], we first need to reformulate (12) as a saddle-point problem.

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

$$\begin{array}{rccc} A: & \mathbb{R}^{m \times n} & \to & \mathbb{R}^{n \times 2} \\ & \lambda & \mapsto & \nabla \left(\sum_{i=1}^m \lambda_i * \psi_i \right) \end{array} \tag{14}$$

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

$$F(q) = \|\nabla u_0 + q\|_{1,\epsilon}$$

and

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

problem (12) can be recast as the following convex-concave saddle-point problem:

$$\min_{\lambda\in\Xi}\max_{q\in Q}\langle A\lambda, q\rangle_Q - F^*(q) + G(\lambda). \tag{15}$$

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 (15) we obtain the following dual problem :

$$\max_{q\in Q}\min_{\lambda\in\Lambda}\langle \lambda, A^*q\rangle_Q - F^*(q) + G(\lambda). \tag{16}$$

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

$$\Delta(\lambda, q) = P(\lambda) - D(q).$$

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

Theorem 5.1 (Characterization of solutions of problem (12)):

1) Problem (12) admits a convex, non-empty set of solutions.
2) Problem (15) admits a non-empty set of saddle points.
3) Let (λ^*, q^*) be a saddle point of (15). It satisfies $\Delta(\lambda^*, q^*) = 0$ and λ^* is a solution of problem (12).
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:

$$\forall (q, q') \in Q^2,\ F^*(q') \geq F^*(q) + \langle \partial F^*(q), q' - q\rangle_Q + \frac{\epsilon}{2}\|q' - q\|_Q^2$$

Proof: Function F is differentiable and its gradient is $1/\varepsilon$-Lipschitz. The result follows from the fact that the conjugate of a convex function with Lipschitz gradient is strongly convex (see e.g. [27]). ∎

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:

$$\gamma = \min_{i\in\{1..m\}} \alpha_i \tag{17}$$

Proof: Since $G(\lambda) = \sum_{i=1}^{m} \phi_i(\lambda_i) + \chi_\Xi(\lambda)$, $\mathrm{dom}(G) = \Xi$. Let $\lambda = (\lambda_i)$ and $\lambda' = (\lambda_i') \in \Xi$. Then:

$$\begin{aligned} G(\lambda') - G(\lambda) &= \sum_{i=1}^{m} \frac{\alpha_i}{2}\langle \lambda_i' - \lambda_i, \lambda_i' + \lambda_i\rangle \\ &= \sum_{i=1}^{m} \frac{\alpha_i}{2}\left(\langle \lambda_i' - \lambda_i, 2\lambda_i\rangle + \|\lambda_i' - \lambda_i\|^2\right) \\ &= DG(\lambda).(\lambda' - \lambda) + \sum_{i=1}^{m} \frac{\alpha_i}{2}\|\lambda_i' - \lambda_i\|^2 \\ &\geq DG(\lambda).(\lambda' - \lambda) + \frac{\gamma}{2}\|\lambda' - \lambda\|^2 \end{aligned}$$

∎

Note that these constants change if we change the inner products.

We have provided all necessary preliminaries to describe the algorithm.

D. A primal-dual algorithm

The algorithm proposed by Chambolle and Pock in [5] applied to problem (12), can be written as follows:

Algorithm 1: Primal-Dual algorithm [5]

Input:
ϵ: the desired precision;
(λ_0, q_0): a starting point;
σ_0, τ_0 such that $\sigma_0\tau_0 = L^2$;
Output:
λ_ϵ: a solution to problem (12).
begin
  k=0;
  while $\Delta(\lambda, q) \leq \epsilon\Delta(\lambda_0, q_0)$ **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. The Update function corresponds to the following updating rules:

Algorithm 2: Step size update rules

Input:
ε: parameter of strong convexity of F^*
γ: parameter of strong convexity of G (see eq. (17)
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 (15). 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 $\omega = \frac{1}{1+\frac{\sqrt{\gamma\epsilon}}{L}}$.

These convergence rates can be shown to be optimal for first order methods [17], [19].

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 paragraph VI-C for more details.

VI. Results

In this section we first present experimental results of the proposed algorithm on synthetic images and then turn to the application on SPIM imaging. We finish the section by showing the experimental convergence behavior of the algorithm.

A. Synthetic images

The method is first validated on two synthetic examples, where ground truths are available.

In the first example, we use a real color image. We added straigth lines (stripes) on the images with random independent Gaussian amplitudes. The image is denoised by setting $m = 1$ (one noise component is sought after), the filter ψ_1 is a horizontal line of width 1 pixel and $\phi_1(\lambda_1) = \frac{\alpha_1}{2}\|\lambda_1\|_2^2$ which corresponds to a Gaussian noise assumption. Every channels are treated separately. The extension to color total variation is straightforward. Removing this kind of noise is probably a simpler task than standard white noise, as it is incoherent with the image content.

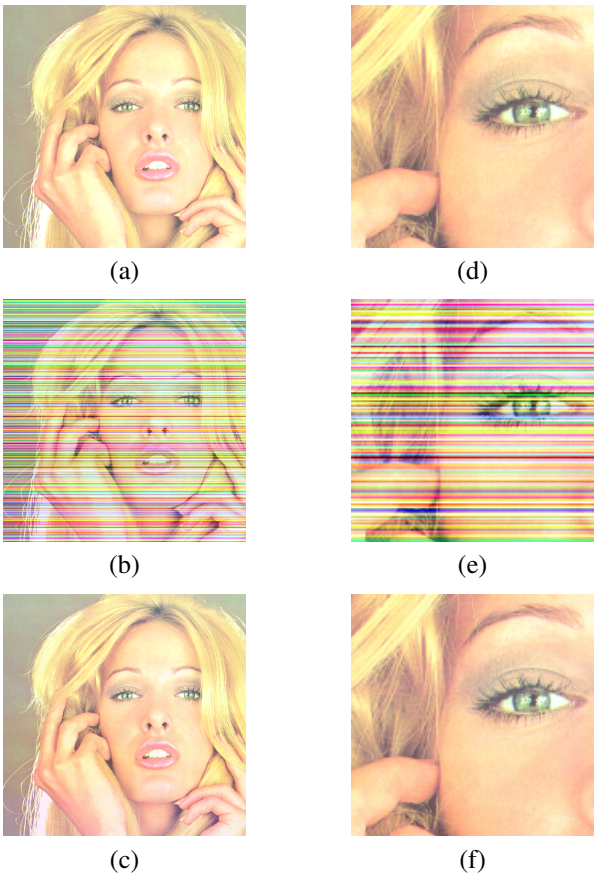


Fig. 5. (a),(d) Original image. (b),(e) Noisy image. (c),(f) Denoised image using the proposed algorithm.

In a second example, we added the sum of two stationnary noises to the Baboon image. The first one is the convolution product of a Gaussian white noise with an isotropic sinc. The second one is the convolution product of a Bernoulli process with a Gabor function.

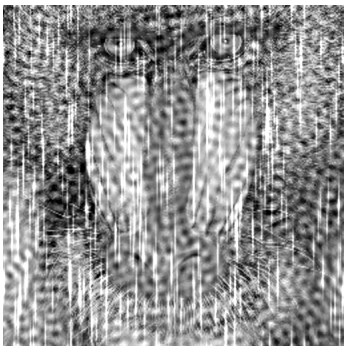


Fig. 6. Synthetic image used for the toy example. PSNR = 21.5dB.

The results are shown below:

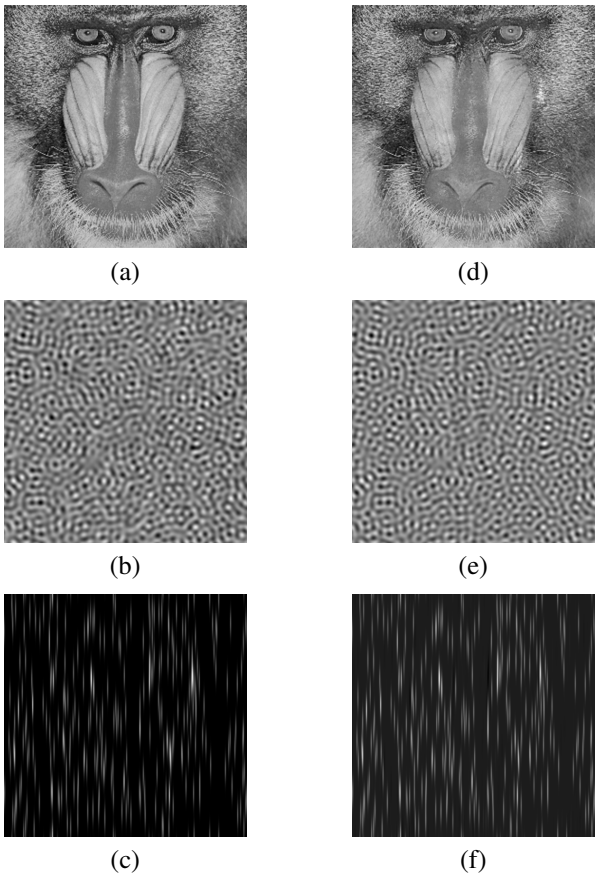


Fig. 7. Toy example. Left column: real components; right column: estimated components using our algorithm. (a,d): Baboon image. PSNR = 27dB. - (b,e): Colored Gaussian noise - (c,f): Impulse like noise

B. Real images

The proposed algorithm is tested on 3 SPIM images with different contents. The images all come from the SPIM microscope prototype developed at ITAV-UMS3039 (CNRS). Algorithm 1 is applied with two filters ψ_1 and ψ_2. The first filter ψ_1 is a Dirac (which allows the recovery of Gaussian 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. This outlines the robustness of this approach with respect to the filter choice. A comparison with a TV-L^2 denoising algorithm is presented in Figure 8.

Let us discuss the results:

- Figure 8: in the image (d), cell groups are visible (top left and bottom right). The raw image is seriously interfered with stripes artifacts which are very marked within the cells. In the result (e), the stripes disappeared and the contours of cells are preserved and they are still well resolved. The image is restored without adding others artefacts. The results of TV-L^2 optimization is also presented in (c) and (f).
- In Figure 10 only the contour of the cells 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). This shows that the proposed noise model is a good approximation of reality.
- Figure 11 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 algorithms and its numerical performances.

a) Parameter tuning: These performances depend quite heavily on the parameters choice 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 such that $\sigma\tau = L^2$. An overestimation of L will slow down the scheme and an underestimation will make the scheme diverge. In the case of total variation, this computation can be done explicitly as shown in appendix B.
2) Second, the relationship $\sigma\tau = L^2$ still leaves a degree of freedom in the choice of σ and τ. It makes 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

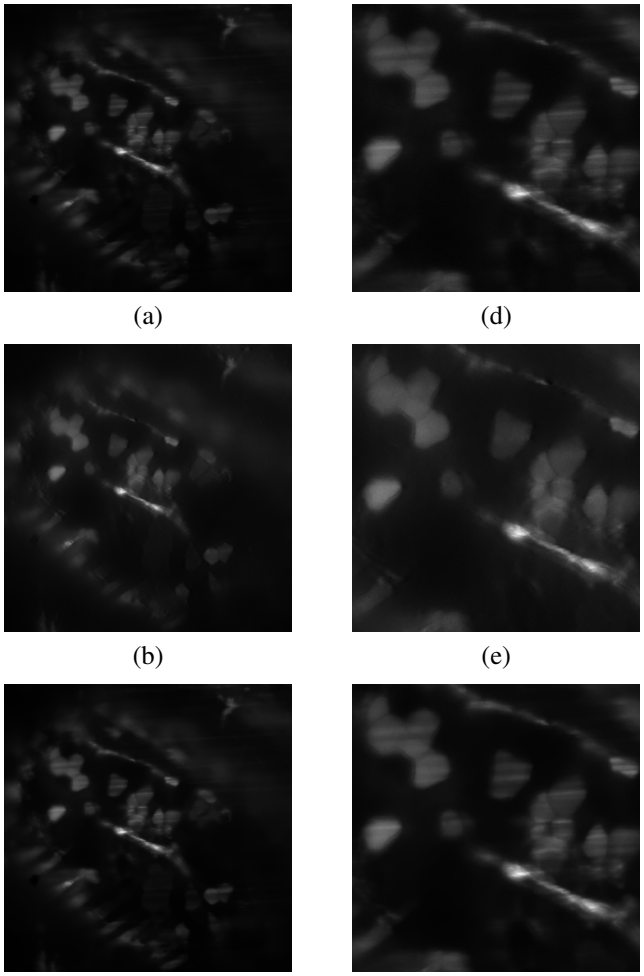


Fig. 8. SPIM images of branchial arcs of Xenopus leavis's late taibud acquired with an objective 40X NA 0.8, an excitation wavelength of 473 nm and an emission wavelength of 510 nm. The voxel size is 0.16*0.16*1 μm. Left column, from top to bottom: (a) Single plane of a 3D stack. (b) Restored image using our algorithm. (c) Restored image using TV-L^2 optimization. Right column: details.

to provide analytical solutions for these choices. We propose a heuristic choice described later in this section.

4) Finally, it is important to have a reliable stopping criterion in order to automatize the algorithm and simplify its use for people not specialized in computing. 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) [1]. 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:

$$\begin{aligned}\mathcal{A} = \quad & (\text{iterations number}) \times 2 \times m \times (\text{FFT}(n) + O(n) \text{ op.}) \\ & + m \times \text{FFT}(n)\end{aligned} \tag{18}$$

[1]For most problems, we experienced that $\epsilon = 10^{-2}$ was enough, so that $\epsilon = 10^{-3}$ is a very cautious choice.

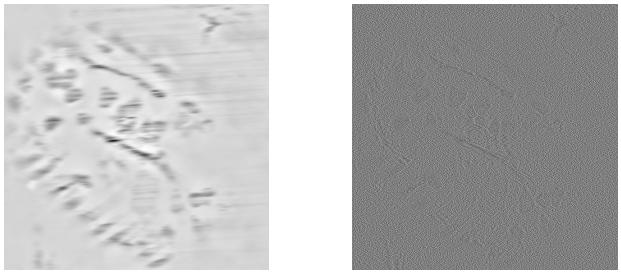


Fig. 9. Left: the stripes identified by our algorithm ($\lambda_2 * \psi_2$). Right: the gaussian noise identified by our algorithm ($\lambda_1 * \psi_1$)

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 $O(n \log(n))$.

Overall, if all 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 $\epsilon = 10^{-3}$ [2]. The overall cost is thus around $100 \times m$ FFT computations. This takes around 5 seconds for a 1024×1024 image, on a 1.2GHz laptop, using $m = 1$ filter. This is quite fast and it can be accelerated using parallel computing. For instance, the scheme can be implemented easily on a GPU. This would decrease the computing times to less than 1 second for a 1024×1024 image.

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 $\langle \cdot, \cdot \rangle_Q$. Let $M : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times n}$ and $N : \mathbb{R}^{n \times 2} \to \mathbb{R}^{n \times 2}$ be two symmetric, positive definite matrices. They allow to specify the inner products on Λ and Q by:

$$\begin{aligned}\langle \lambda_1, \lambda_2 \rangle_\Lambda &= \langle M\lambda_1, \lambda_2 \rangle \\ \langle q_1, q_2 \rangle_N &= \langle Nq_1, q_2 \rangle_Q.\end{aligned} \tag{19}$$

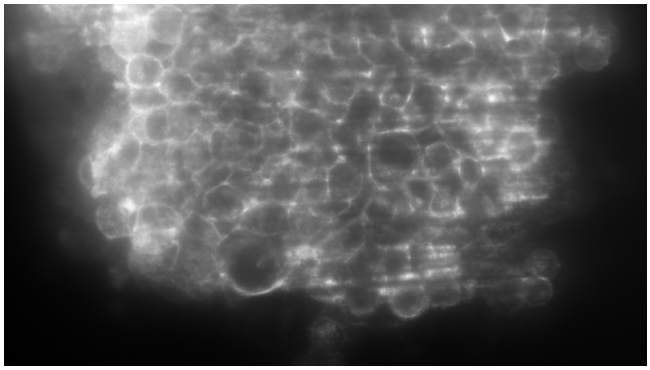
It is not possible to choose N and M arbitrarily if we wish to use explicit expressions for the resolvents of F^* and G. In this work, we thus simply use diagonal matrices which allow explicit computations (see appendix A).

Usually, metric change or preconditionning are used in linear systems in order to improve the condition numbers of matrices. The case of convex programming is more involved: the Lipschitz constant of the gradient and the modulus of strong convexity - which generalize the notion of smallest and highest singular value - can be infinite or null. A potential goal for choosing σ, τ, N and M could thus be to minimize:

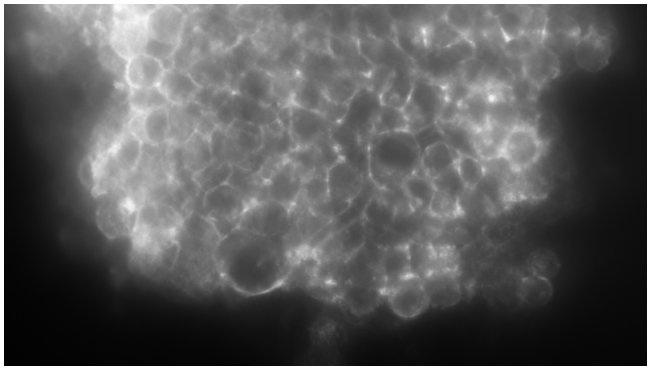
$$\frac{\|\lambda^* - \lambda_0\|_\Lambda^2}{2\tau} + \frac{\|q^* - q_0\|_Q^2}{2\sigma} \tag{20}$$

such that $\sigma\tau = L^2$. This quantity defines the scheme's theoretical convergence rate [5]. However our tries using this strategy did not provide good enough results.

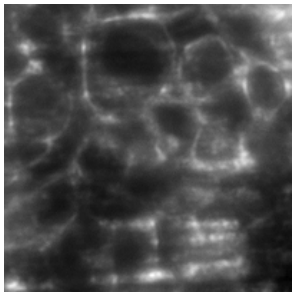
[2]for easy problems, even 10 iterations are enough.



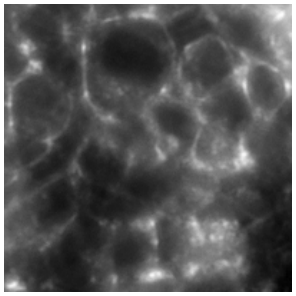
(a)



(b)



(c)



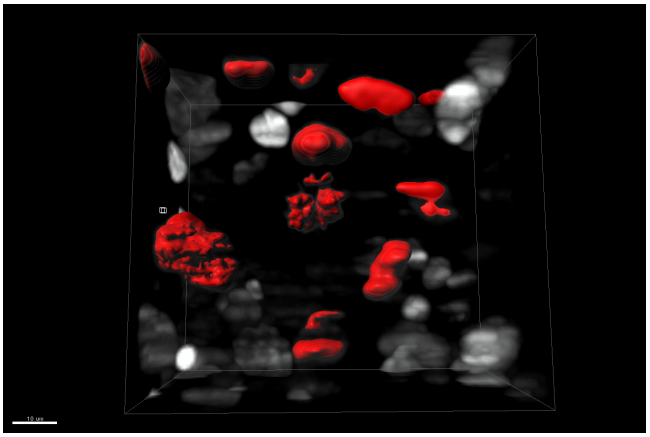
(d)

Fig. 10. 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).

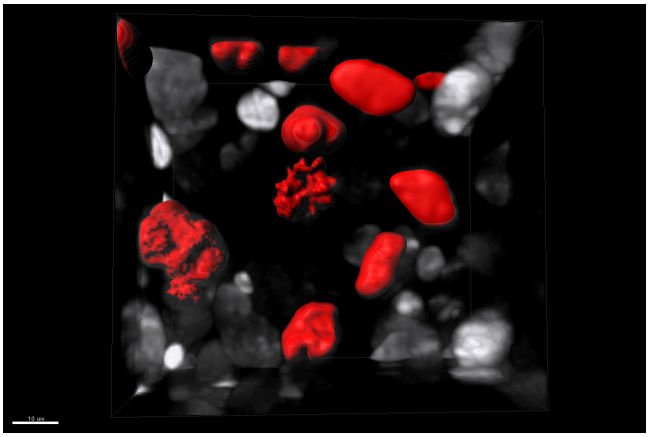
That is why we finally proceed by trial and error using very simple matrices. The matrix N is used the dual space which represents the level lines orientation. There is a priori no reason to promote different pixels so that we simply use $N = \mathrm{Id}$. The matrix M is used in the primal space and allows to balance the different filters. We define it as a diagonal matrix of form:

$$M = \begin{pmatrix} m_1 \mathrm{Id}_n & 0_n & 0_n \\ 0_n & \ddots & 0_n \\ 0_n & 0_n & m_m \mathrm{Id}_n \end{pmatrix} \tag{21}$$

where m_m are positive real numbers allowing to weight the different filters. The coefficients m_m are obtained by trial and error. This strategy is satisfactory when dealing with a single imaging modality (here we use the same SPIM device), and the parameters can be tuned once for all.



(a)



(b)

Fig. 11. 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.

VII. Conclusion

We proposed a variational approach to denoise images impaired by a large class of stationary noise. It takes advantage of recent advances in convex optimization, leading to interactive computing times. Applications to synthetic images, and to images issued from an emerging microscope (SPIM) were presented, leading to very clear improvements of the image quality.

Appendix A
Explicit expressions of the resolvents

The resolvent of F^* and G depend on the choices of the inner products. By using definition (19), we obtain - using standard rules of conjugacy - that:

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

and

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

The adjoint of A also depends on M and N. Elementary calculus leads to:

$$A^* = M^{-1} A^T N$$

where the transpose is:

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

The different Fenchel transforms of ϕ_i are:

- If $\phi_i(\lambda) = \alpha_i \|\lambda\|_1$:

$$(\phi_i + \chi_{[-C,C]^n})^*(\lambda) = C\| \max(0, |\lambda| - \alpha_i)\|_1.$$

- If $\phi_i(\lambda) = \frac{\alpha_i}{2}\|\lambda\|_2^2$:

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

- If $\phi_i(\lambda) = \chi_{[-\alpha_i,\alpha_i]}(\lambda)$:

$$(\phi_i + \chi_{[-C,C]^n})^*(\lambda) = \min(\alpha_i, C)\|\lambda\|_1$$

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

$$\begin{aligned} q^* &= (\mathrm{Id} + \sigma\partial F^*)^{-1}(q_n) \\ &= \underset{q\in Q, \|Nq\|_\infty \le 1}{\arg\min} \frac{\sigma\epsilon}{2}\|Nq\|_2^2 - \langle \sigma\nabla u_0, Nq\rangle \\ &\quad + \frac{1}{2}\|q - q_n\|_Q^2 \end{aligned} \tag{22}$$

We assume that N is a diagonal matrix and $N(\mathbf{x})$ denotes the diagonal element of N associated to pixel $\mathbf{x}$. 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 (22):

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

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 and M is diagonal, this problem reduces to $m \times n$ unidimensional problems of form:

$$\underset{\lambda\in\mathbb{R}, |\lambda|\le C}{\arg\min}\, \tau f(\lambda) + \frac{1}{2}m|\lambda - \lambda_n|^2 \tag{23}$$

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

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

  The solution of (23) is given by:

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

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

  The solution of (23) is given by:

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

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

  Let $\delta = \min(C, \alpha)$. The solution of (23) is given by:

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

APPENDIX B
COMPUTATION OF THE OPERATOR NORM L

In order to make the iterative methods as fast as possible, it is important to use a tight estimation of L, the highest singular value of A defined in equation (14). 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:

$$H_i = \mathcal{F}^{-1}\mathrm{diag}\left(\hat{\psi}_i\right)\mathcal{F}. \tag{24}$$

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, $\hat{d}_1$ and $\hat{d}_2$ where defined above and $\Sigma(\mathbf{k}) = \sum_{i=1}^{m} \text{diag}\left(|\hat{\psi}_i(\mathbf{k})|^2\right)$. This computation is achieved in $O(n \log n)$ arithmetic operations [3].

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. Jérôme Fehrenbach was partly supported by a grant from CNRS.

REFERENCES

[1] L. Alvarez, F. Guichard, P.-L. Lions, and J.-M. Morel "Axioms and Fundamental Equations of Image Processing," *Arch. Rational Mech. Anal.*, vol. 123, pp. 199–257, 1993.
[2] A. Arrenberg AB, D. Stainier, H. Baier H, and J.Huisken, "Optogenetic control of cardiac function," *Science*, vol. 330(6006), pp. 971–974, 2010.
[3] 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.
[4] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no.1, pp. 183–202, 2009.
[5] 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.
[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] F. Facchinei and J. Pang, "Finite-dimensional variational inequalities and complementarity problems", Springer Verlag, 2003.
[10] 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.
[11] J. Garnett, T. Le, Y. Meyer, and L. Vese, "Image decompositions using bounded variation and generalized homogeneous Besov spaces," *Appl. Comput. Harmon. Anal.*, vol. 23, pp. 25–56, 2007.
[12] 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.
[13] 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.
[14] 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.
[15] 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
[16] B. Munch, P. Trtik, F. Marone, M. Stampanoni "Stripe and ring artifact removal with combined wavelet-Fourier filtering." *Opt. Express*,vol. 17, pp. 8568–8591, 2009.
[17] A. Nemirovski, "Information-based complexity of linear operator equations," *Journal of Complexity*, vol. 8, pp. 153–175, 1992.
[18] 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.
[19] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, Kluwer Academic Publishers, 2004.
[20] Y. Nesterov, "Smooth minimization of non-smooth functions," *Mathematical Programming*, vol. 103, no. 1, pp. 127–152, 2005.
[21] Y. Nesterov, *Gradient methods for minimizing composite objective function*, CORE Discussion paper 2007.
[22] V. Petrov and A. Brown, "Sums of independent random variables", Springer-Verlag, 1975.
[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.-L. Starck, M. Elad, and D. Donoho, "Image decomposition via the combination of sparse representations and a variational approach," *IEEE Trans. Image Processing*, vol. 14, no. 10, 2005.
[27] J.-B. Hiriart-Urruty and C. Lemarechal, "Convex Analysis and Minimization Algorithms", Vol. 2, Springer Verlag, Berlin, Germany, 1993

[3]The computation with a different diagonal metric can still be achieved explicitely.

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.