

## Processing Stationary Noise: Model and Parameter Selection in Variational Methods\*

Jérôme Fehrenbach<sup>†</sup> and Pierre Weiss<sup>‡</sup>

**Abstract.** Additive or multiplicative stationary noise recently became an important issue in applied fields such as microscopy or satellite imaging. Relatively few works address the design of dedicated denoising methods compared to the usual white noise setting. We recently proposed a variational algorithm to tackle this issue. In this paper, we analyze this problem from a statistical point of view and provide deterministic properties of the solutions of the associated variational problems. In the first part of this work, we demonstrate that in many practical problems, the noise can be assimilated to a colored Gaussian noise. We provide a quantitative measure of the distance between a stationary process and the corresponding Gaussian process. In the second part, we focus on the Gaussian setting and analyze denoising methods which consist in minimizing the sum of a total variation term and an  $l^2$  data fidelity term. While the constrained formulation of this problem allows us to easily tune the parameters, the Lagrangian formulation can be solved more efficiently since the problem is strongly convex. Our second contribution consists in providing analytical values of the regularization parameter in order to approximately reach a given noise level.

**Key words.** stationary noise, Berry–Esseen theorem, Morozov principle, negative norm models, destriping, convex analysis and optimization

**AMS subject classifications.** 94A08, 65K10, 49M29, 60F05

**DOI.** 10.1137/130929424

**1. Introduction.** In a recent paper [9], a variational method that decomposes an image into the sum of a piecewise smooth component and a set of stationary processes was proposed. This algorithm has a large number of applications, such as deconvolution or denoising when *structured patterns* degrade the image contents. A typical example of an application that received considerable attention lately is destriping [4, 6, 9, 14, 18]. It was also shown to generalize the negative norm models [1, 16, 21, 26] in the discrete setting [10]. Figures 1, 2, and 6 show examples of applications of this algorithm in an additive noise setting, and Figure 3 shows an example with a multiplicative noise model.

This algorithm is based on the hypothesis that the observed image  $u_0$  can be written as

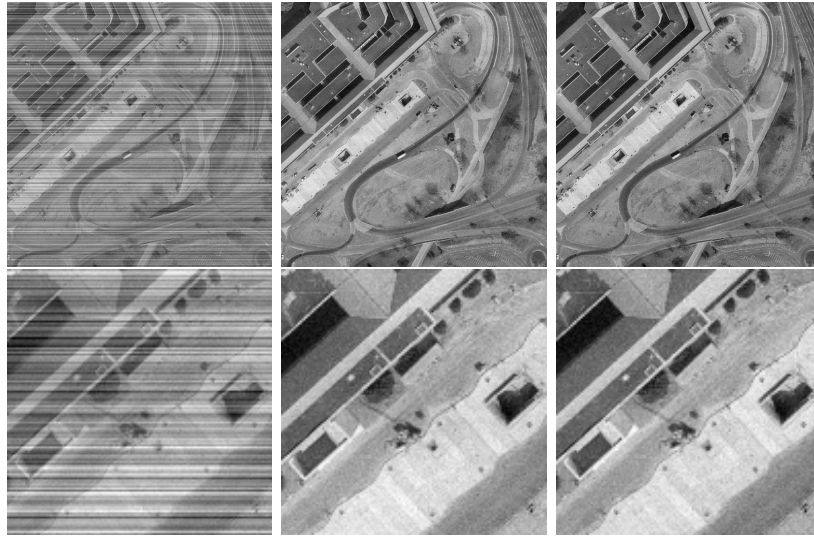
$$u_0 = u + \sum_{i=1}^m b_i,$$

\*Received by the editors July 16, 2013; accepted for publication (in revised form) January 22, 2014; published electronically April 17, 2014. This work was partially supported by ANR SPH-IM-3D (ANR-12-BSV5-0008) and mission pour l'interdisciplinarité from CNRS.

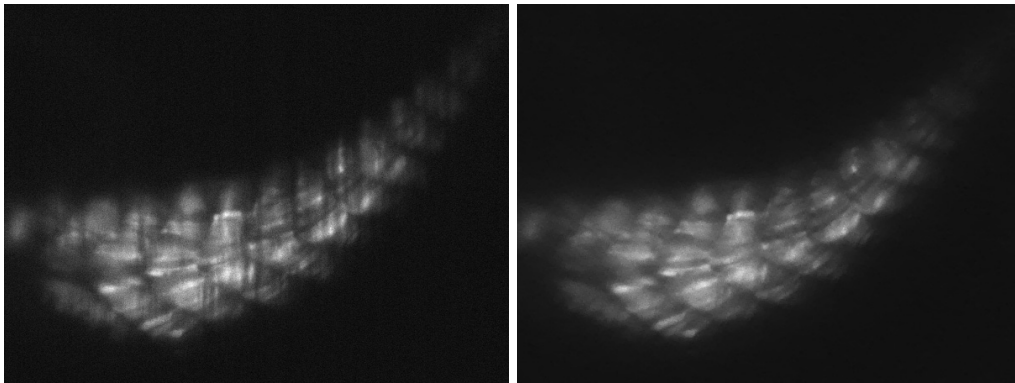
<http://www.siam.org/journals/siims/7-2/92942.html>

<sup>†</sup>IMT-UMR5219, Université de Toulouse, CNRS, 31062 Toulouse, France ([jerome.fehrenbach@math.univ-toulouse.fr](mailto:jerome.fehrenbach@math.univ-toulouse.fr)).

<sup>‡</sup>ITAV-USR3505, Université de Toulouse, CNRS, 31106 Toulouse, France ([pierre.weiss@itav-recherche.fr](mailto:pierre.weiss@itav-recherche.fr)).



**Figure 1.** Top: full size images. Bottom: zoom on a small part. From left to right: Noisy image (17dB), denoised using the method proposed in [9] (PSNR = 34.3dB), original image.



**Figure 2.** Left: SPIM image of a slow muscle fibers labeled zebrafish embryo using the *Tg(smγhcl1:EGFP)i104* line. Right: denoised image using variational stationary noise removal (VSNR). (Image credit: Julie Batut, CNRS and University of Toulouse; Centre de Biologie du Développement; F-31062 Toulouse, France.)

where  $u$  denotes the original image and  $(b_i)_{i \in \{1, \dots, m\}}$  denotes a set of realizations of independent stochastic processes  $B_i$ . These processes are further assumed to be stationary and read as  $B_i = \psi_i \star \Lambda_i$ , where  $\psi_i$  denotes a known kernel and each  $\Lambda_i$  is a random vector with independent and identically distributed (i.i.d.) entries. The decomposition algorithm can then be deduced from a Bayesian approach, leading to large scale convex optimization problems of size  $m \times n$ , where  $n$  is the number of pixels/voxels in the image.

This method is now used routinely in the context of microscopy imaging. Its main weakness for a broader use lies in the difficulty in setting its parameters adequately. One basically needs to input the filters  $\psi_i$  and the marginals of each random vector  $\Lambda_i$ , which is uneasy even for



**Figure 3.** An example involving a multiplicative noise model. Top-left: original image. Top-right: noisy image. It is obtained by multiplying each line of the original image by a uniform random variable in  $[0.1, 1]$ .  $SNR=10.6dB$  – Bottom-left: denoised using the method proposed in [9] on the logarithm of the noisy image.  $SNR=29.1dB$  – Bottom-right: ratio between the original image and the denoised image. The multiplicative factor is retrieved accurately.

imaging specialists. Our aim in this paper is to provide a set of mathematically founded rules to simplify the parameter selection and reduce computing times. We do not tackle the problem of finding the filters  $\psi_i$  (which is a problem similar to blind deconvolution), but focus on the choice of the marginals of  $\Lambda_i$ . Our main contribution is to propose an analytic expression of the regularization parameters in order to approximately reach a given noise level. Contrarily to other approaches, our method is one shot in the sense that no iterative process is used. This is similar in spirit to the recent work [2].

The outline of the paper is as follows. Notation is described in section 2.1. In section 2.2, we review the main principles motivating the decomposition algorithm. In section 3, we show that, from a statistical point of view and for many applications, assuming that  $\Lambda_i$  is a Gaussian process is nearly equivalent to selecting other marginals. This has the double advantage of simplifying the analysis of the model properties and reducing the computational complexity. In section 4, we show that when  $b_i$  are drawn from Gaussian processes, parameter selection can be performed in a *deterministic* way, by analyzing the primal-dual optimality conditions. We also show that the proposed ideas allow us to reduce the problem dimension from  $m \times n$  to  $n$  variables, thus dividing the storage cost and computing times by a factor roughly equal

to  $m$ . Section 5, the appendix, contains the proofs of the results stated in section 4.

## 2. Notation and context.

**2.1. Notation.** We consider discrete  $d$ -dimensional images  $u \in \mathbb{R}^n$ , where  $n = n_1 \cdot n_2 \cdots n_d$  denotes the pixel number. The pixel indices belong to the set  $\Omega = \{1, \dots, n_1\} \times \cdots \times \{1, \dots, n_d\}$ . The pixel value of  $u$  at location  $\mathbf{x} \in \Omega$  is denoted by  $u(\mathbf{x}) = u(x_1, \dots, x_d)$ . Let  $u \in \mathbb{R}^n$  denote an image. The image  $u^{mean} \in \mathbb{R}^n$  has all its components equal to the mean of  $u$ . The standard  $l^p$ -norms on  $\mathbb{R}^n$  are denoted by  $\|\cdot\|_p$ . Discrete vector fields

$$\mathbf{q} = \begin{pmatrix} q_1 \\ \vdots \\ q_d \end{pmatrix} \in \mathbb{R}^{n \cdot d}$$

are denoted by bold symbols. The isotropic  $l^p$ -norms on vector fields are denoted by  $\|\cdot\|_{\mathbf{p}}$  and defined by

$$\|\mathbf{q}\|_{\mathbf{p}} = \|\sqrt{q_1^2 + \cdots + q_d^2}\|_p,$$

where  $q_i^2$  should be understood componentwise.

The discrete partial derivative in direction  $k$  is defined by

$$\partial_k u(\cdot, x_k, \cdot) = \begin{cases} u(\cdot, x_k + 1, \cdot) - u(\cdot, x_k, \cdot) & \text{if } 1 \leq x_k < n_k, \\ u(\cdot, 1, \cdot) - u(\cdot, n_k, \cdot) & \text{if } x_k = n_k. \end{cases}$$

Using periodic boundary conditions allows one to rewrite partial derivatives as circular convolutions:  $\partial_k u = d_k \star u$ , where  $d_k$  is a finite difference filter. The discrete gradient operator in the  $d$ -dimension is defined by

$$\nabla = \begin{pmatrix} \partial_1 \\ \partial_2 \\ \vdots \\ \partial_d \end{pmatrix}.$$

The discrete isotropic total variation of  $u \in \mathbb{R}^n$  is defined by  $TV(u) = \|\nabla u\|_{\mathbf{1}}$ . Let  $\|\cdot\|_{\alpha}$  and  $\|\cdot\|_{\beta}$  denote norms on  $\mathbb{R}^n$  and  $\mathbb{R}^m$ , respectively, and let  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$  denote a linear operator. The subordinate operator norm  $\|A\|_{\alpha \rightarrow \beta}$  is defined as follows:

$$(2.1) \quad \|A\|_{\alpha \rightarrow \beta} = \max_{\|x\|_{\alpha} \leq 1} \|Ax\|_{\beta}.$$

Let  $u$  and  $v$  be two  $d$ -dimensional images. The pointwise product between  $u$  and  $v$  is denoted by  $u \odot v$ , and the pointwise division is denoted by  $u \oslash v$ . The conjugate of a number or a vector  $a$  is denoted by  $\bar{a}$ . The transconjugate of a matrix  $\mathcal{M} \in \mathbb{C}^{m \times n}$  is denoted by  $\mathcal{M}^*$ . The canonical basis of  $\mathbb{R}^n$  is denoted by  $(e_i)_{i \in \{1, \dots, n\}}$ . The discrete Fourier basis of  $\mathbb{C}^n$  is denoted by  $(f_i)_{i \in \{1, \dots, n\}}$ . We use the convention that  $\|f_i\|_{\infty} = 1$  for all  $i$  so that  $\|f_i\|_2 = \sqrt{n}$  (see, e.g., [15]). The discrete  $d$ -dimensional Fourier transform of  $u$  is denoted by  $\mathcal{F}u$  or  $\hat{u}$ .

It satisfies  $\|\hat{u}\|_2 = \sqrt{n}\|u\|_2$ . The inverse Fourier transform is denoted by  $\mathcal{F}^{-1}$  and satisfies  $\mathcal{F}^{-1} = \frac{\mathcal{F}^*}{n}$ . The discrete symmetric of  $u$  is denoted by  $\tilde{u}$  and defined by  $\tilde{u} = \mathcal{F}^{-1}\hat{\tilde{u}}$ . The convolution product between  $u$  and  $\psi$  is denoted by  $u \star \psi$  and defined for any  $\mathbf{x} \in \Omega$  by

$$u \star \psi(\mathbf{x}) = \sum_{\mathbf{y} \in \Omega} u(\mathbf{y})\psi(\mathbf{x} - \mathbf{y}),$$

where periodic boundary conditions are used. It satisfies

$$u \star \psi = \mathcal{F}^{-1} \left( \hat{u} \odot \hat{\psi} \right).$$

Since the discrete convolution is a linear operator, it can be represented by a matrix. The convolution matrix associated with a kernel  $\psi$  is denoted in capital letters  $\Psi$ :

$$\Psi u = u \star \psi.$$

The transpose of a convolution operator with  $\psi$  is a convolution operator with the symmetrized kernel:  $\Psi^T u = \tilde{\psi} \star u$ .

**2.2. Decomposition algorithm.** The VSNR algorithm is described in [9]. The starting point of our algorithm is the following image formation model:

$$u_0 = u + \sum_{i=1}^m \lambda_i \star \psi_i,$$

where  $u_0 \in \mathbb{R}^n$  is the observed image and  $u \in \mathbb{R}^n$  is the image to recover. Each  $\psi_i \in \mathbb{R}^n$  is a known filter, and each  $\lambda_i \in \mathbb{R}^n$  is the realization of a random vector with i.i.d. entries. We assume that its density reads as  $\mathcal{P}(\lambda_i) \propto \exp(-\phi_i(\lambda_i))$ , where  $\phi_i$  is a *separable* function of kind

$$(2.2) \quad \phi_i(\lambda_i) = \sum_{\mathbf{x} \in \Omega} \varphi_i(\lambda_i(\mathbf{x}))$$

with  $\varphi_i : \mathbb{R} \rightarrow \mathbb{R} \cup \{+\infty\}$  (typical examples are  $l^p$  to the  $p$ -norms). Note that hypothesis (2.2) is a simple consequence of the i.i.d. hypothesis.

Our aim is to recover both the stationary components  $b_i = \lambda_i \star \psi_i$  and the image  $u$ . Assuming that the noise  $b = \sum_{i=1}^m b_i$  and the image are drawn from independent random vectors, the maximum a posteriori (MAP) approach leads to the following optimization problem:

$$\text{Find } (\boldsymbol{\lambda}^\dagger, u^\dagger) \in \underset{\boldsymbol{\lambda} \in \mathbb{R}^{n \times m}, u \in \mathbb{R}^n}{\text{Arg max}} \mathcal{P}(\boldsymbol{\lambda}, u | u_0).$$

Bayes' rule allows us to reformulate this problem as (see [9] for details)

$$\text{Find } \boldsymbol{\lambda}^\dagger \in \underset{\boldsymbol{\lambda} \in \mathbb{R}^{n \times m}, u \in \mathbb{R}^n \text{ such that (s.t.) } u = u_0 - \sum_{i=1}^m \lambda_i \star \psi_i}{\text{Arg min}} -\log \mathcal{P}(u_0 | \boldsymbol{\lambda}, u) - \log \mathcal{P}(\boldsymbol{\lambda}) - \log \mathcal{P}(u).$$

Since we assumed independence of the  $\lambda_i$ 's,

$$-\log \mathcal{P}(\boldsymbol{\lambda}) = \sum_{i=1}^m -\log \mathcal{P}(\lambda_i).$$

If we further assume that  $\mathcal{P}(u) \propto \exp(-\|\nabla u\|_1)$ , the optimization problem we aim at solving finally writes as

$$(2.3) \quad \text{Find } \lambda \in \text{Arg min}_{\lambda \in \mathbb{R}^{n \times m}} \left\| \nabla \left( u_0 - \sum_{i=1}^m \lambda_i \star \psi_i \right) \right\|_1 + \sum_{i=1}^m \phi_i(\lambda_i).$$

This problem is convex whenever all the  $\phi_i$  are convex and can be solved efficiently using first order algorithms such as proximal splitting methods [5, 7, 10]. The filters  $\psi_i$  and the functions  $\phi_i$  are user defined and should be selected using prior knowledge about the noise properties. Unfortunately, the choice of  $\phi_i$  proved to be very complicated in applications. Even for the special case  $\phi_i(\cdot) = \frac{\alpha_i}{2} \|\cdot\|_2^2$ ,  $\alpha_i$  is currently obtained by trial and error, and interesting values vary in the range  $[10^{-8}, 10^{10}]$  depending on the filters  $\psi_i$  and the noise level. It is thus essential to restrict the range of these parameters in order to ease the task of end users.

Problem (2.3) is a very large scale problem since typical three-dimensional images contain from  $10^8$  to  $10^9$  voxels. Most automatized parameter selection methods, such as generalized cross validation [12] or generalized SURE [25], require solving several instances of (2.3). This leads to excessive computational times in our setting. In this paper, we propose estimating *analytically* the parameters  $\alpha_i$ . The principle underlying our approach is a variant of the Morozov discrepancy principle: we propose finding parameters  $\alpha_i$  such that the  $i$ th noise component approximately has a given norm. Let us note that the idea of fixing the noise level to denoise an image was successfully applied in various works, such as [19, 24, 27].

**3. Effectiveness of the Gaussian model in the non-Gaussian setting.** In this section we analyze the statistical properties of random processes that can be written as  $\Lambda \star \psi$ , where  $\Lambda$  is a noise process with independent components. Our main result is that the stationary noise  $b_i = \lambda_i \star \psi_i$  can be assimilated to a *Gaussian* colored noise for many applications of interest even if  $\Lambda$  is non-Gaussian.

**3.1. Distance of stationary processes to the Gaussian distribution.** Our results are simple consequences of the Berry–Esseen theorem [3] that quantifies the distance between a sum of independent random variables and a Gaussian.

**Theorem 3.1 (Berry–Esseen).** *Let  $X_1, X_2, \dots, X_n$  be independent centered random variables of finite variances  $\sigma_i^2$  and finite third order moments  $\rho_i = \mathbb{E}(|X_i|^3)$ . Let*

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{\sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2}}.$$

*Let  $F_n$  denote the cumulative distribution function (cdf) of  $S_n$ . Let  $\Phi$  denote the cdf of the standard normal distribution. Then*

$$(3.1) \quad \|F_n - \Phi\|_\infty \leq C_0 \frac{\sum_{i=1}^n \rho_i}{(\sum_{i=1}^n \sigma_i^2)^{3/2}},$$

where  $C_0 \leq 0.56$ .

In our problem, we consider random vectors of kind

$$(3.2) \quad B = \psi \star \Lambda = \Psi \Lambda,$$

so that

$$B(\mathbf{x}) = \sum_{\mathbf{y} \in \Omega} \Lambda(\mathbf{x} - \mathbf{y})\psi(\mathbf{y}),$$

where  $(\Lambda(\mathbf{x}))_{\mathbf{x} \in \Omega}$  are centered i.i.d. random variables. Let us further assume that they are of finite second and third order moments.<sup>1</sup> Denote  $\sigma^2 = \mathbb{E}(\Lambda(\mathbf{y})^2) < +\infty$  and  $\rho = \mathbb{E}(|\Lambda(\mathbf{y})|^3) < +\infty$ . The mean of  $B$  is  $\mathbb{E}(B) = 0$  since convolution operators preserve the set of vectors with zero mean. Moreover, its covariance matrix is  $\text{Cov}(B) = \sigma^2 \Psi^T \Psi$  with  $\Psi^T \Psi = \mathcal{F}^{-1} \text{diag}(|\hat{\psi}|^2) \mathcal{F}$  whatever the distribution of  $\Lambda$ . The following results state that every component  $B(\mathbf{x})$  is close to a Gaussian whatever the law of  $\Lambda$  if the filter  $\psi$  satisfies a geometrical criterion discussed later.

**Proposition 3.2.** *Let  $G$  denote the cdf of  $\frac{B(\mathbf{x})}{s\sigma}$ , where  $s = \|\psi\|_2$ . This cdf is independent of  $\mathbf{x}$ ; moreover,*

$$(3.3) \quad \|G - \Phi\|_\infty \leq 0.56 \frac{\rho}{\sigma^3} \frac{\|\psi\|_3^3}{\|\psi\|_2^3}.$$

*Proof.* The independence w.r.t.  $\mathbf{x}$  is a direct consequence of the stationarity of  $B$ . Bound (3.3) is a direct consequence of the Berry–Esseen theorem, Theorem 3.1. It suffices to notice that  $\mathbb{E}(|\Lambda(\mathbf{x} - \mathbf{y})\psi(\mathbf{y})|^2) = \psi(\mathbf{y})^2 \sigma^2$ ,  $\mathbb{E}(|\Lambda(\mathbf{x} - \mathbf{y})\psi(\mathbf{y})|^3) = |\psi(\mathbf{y})|^3 \rho$  for any  $(\mathbf{x}, \mathbf{y}) \in \Omega^2$  and to apply Theorem 3.1. ■

Thus, if  $\frac{\|\psi\|_3^3}{\|\psi\|_2^3}$  is sufficiently small, the distribution of  $B(\mathbf{x})$  will be near Gaussian. The following result clarifies this condition in an asymptotic regime.

**Proposition 3.3.** *Let  $\psi : \mathbb{R}_+^d \rightarrow \mathbb{R}$  denote a function. Let  $\Omega_n = [1, n]^d \cap \mathbb{Z}^d$  denote a Euclidean grid. Let  $s_n = \sqrt{\sum_{\mathbf{x} \in \Omega_n} \psi^2(\mathbf{x})}$ . If  $\Lambda(\mathbf{x})$  is of finite second and third order moments and the sequence  $(\psi(\mathbf{x}))_{\mathbf{x} \in \mathbb{Z}^d}$  is uniformly bounded  $|\psi(\mathbf{x})| \leq M < +\infty$  for all  $\mathbf{x} \in \mathbb{Z}^d$  and has infinite variance  $\lim_{n \rightarrow +\infty} s_n = +\infty$ , then for all  $\mathbf{x} \in \Omega_n$ ,*

$$\frac{B(\mathbf{x})}{s_n} \xrightarrow{(\mathcal{D})} \mathcal{N}(0, \sigma^2).$$

*Proof.* Let us denote

$$f(n) = \frac{\sum_{\mathbf{x} \in \Omega_n} |\psi(\mathbf{x})|^3}{(\sum_{\mathbf{x} \in \Omega_n} \psi(\mathbf{x})^2)^{3/2}}.$$

We have

$$\begin{aligned} \sum_{\mathbf{x} \in \Omega_n} |\psi(\mathbf{x})|^3 &\leq \sum_{\mathbf{x} \in \Omega_n} \|\psi\|_\infty \psi(\mathbf{x})^2 \\ &\leq M s_n^2. \end{aligned}$$

Thus

$$f(n) \leq \frac{M s_n^2}{s_n^3} = \frac{M}{s_n}.$$

---

<sup>1</sup>This hypothesis is not completely necessary, but it simplifies the exposition.

The right-hand side in (3.3) is  $f(n)$  and goes to 0 as  $n \rightarrow +\infty$ . The Lindeberg–Feller theorem could also be used in this context and allow us to avoid moment conditions. ■

**3.2. Examples.** We present different examples of kernels where Theorem 3.1 applies.

*Example 1.* We first consider a kernel that is the indicator function of a “large” set, namely  $\psi(\mathbf{x}) = 1$  if  $\mathbf{x} \in I$ , and  $\#I = N$ . Then the upper bound in (3.1) is  $C_0/\sqrt{N}$ . It becomes small when  $N$  becomes large.

*Example 2.* Let us study the case of kernels with a (slow enough) power decay:  $\psi(\mathbf{x}) = |\mathbf{x}|^\alpha$  for some  $-d/2 < \alpha < 0$ . In this case, the quantity  $s_n$  tends to infinity since it is asymptotic to

$$\int_{[1,n]^d} |\mathbf{x}|^{2\alpha} d\mathbf{x} \sim K \int_{r=1}^n r^{d-1} r^{2\alpha} dr \sim K n^{d+2\alpha}$$

for some constant  $K$ . Therefore Proposition 3.3 applies. This result is still valid for  $\alpha \geq 0$ .

*Example 3.* We treat the case of an anisotropic Gaussian filter  $\psi$  with axes aligned with the coordinate axes. In this case the variance is finite and Proposition 3.3 does not apply. However, we can give an explicit value of the upper bound in (3.3), which ensures that the process is close to a Gaussian. Let us assume that

$$\psi(\mathbf{x}) = K e^{-\sum_{i=1}^d x_i^2/2\sigma_i^2},$$

where  $K$  is a normalizing constant and  $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathbb{Z}^d$ . We provide in this case an upper bound for  $f(n)$  in terms of  $(\sigma_i)$ . For the sake of simplicity we assume that  $K = 1$ .

$$\begin{aligned} \sum_{\mathbf{x} \in \mathbb{Z}^d} |\psi(\mathbf{x})|^3 &= \sum_{(n_1, \dots, n_d) \in \mathbb{Z}^d} e^{-3\sum_{1 \leq i \leq d} n_i^2/2\sigma_i^2} \\ &= \prod_{i=1}^d \left( \sum_{n \in \mathbb{Z}} e^{-3n^2/2\sigma_i^2} \right) \\ &= \prod_{i=1}^d \left( 1 + 2 \sum_{n>0} e^{-3n^2/2\sigma_i^2} \right) \end{aligned}$$

and, similarly,

$$\sum_{\mathbf{x} \in \mathbb{Z}^d} |\psi(\mathbf{x})|^2 = \prod_{i=1}^d \left( 1 + 2 \sum_{n>0} e^{-n^2/\sigma_i^2} \right).$$

We use the inequalities

$$\frac{1}{2} \sqrt{\frac{\pi}{\alpha}} - 1 \leq \int_1^{+\infty} e^{-\alpha t^2} dt \leq \sum_{n>0} e^{-\alpha n^2} \leq \int_0^{+\infty} e^{-\alpha t^2} dt = \frac{1}{2} \sqrt{\frac{\pi}{\alpha}}$$

to obtain

$$\max \left( 1, \sqrt{\frac{\pi}{\alpha}} - 1 \right) \leq 1 + 2 \sum_{n>0} e^{-\alpha n^2} \leq 1 + \sqrt{\frac{\pi}{\alpha}}.$$



It follows that

$$\lim_{n \rightarrow \infty} f(n) \leq \prod_{i=1}^d \frac{(1 + \sigma_i \sqrt{2\pi/3})}{\max(1, \sigma_i \sqrt{\pi} - 1)^{3/2}} = \prod_{i=1}^d g(\sigma_i).$$

Note that  $g(\sigma) = \mathcal{O}_{+\infty}(\frac{1}{\sqrt{\sigma}})$ . In other words, if the Gaussian kernel has sufficiently large variances, the constant in the upper bound of (3.1) is small.

*Example 4.* In this example, we illustrate Theorem 3.1 in a practical setting. Let us assume that  $\Lambda(\mathbf{x})$  is a Bernoulli-uniform random variable in order to model sparse processes. With this model  $\Lambda(\mathbf{x}) = 0$  with probability  $1 - \gamma$  and takes a random value distributed uniformly in  $[-1, 1]$  with probability  $\gamma$ . Simple calculation leads to  $\sigma^2 = \frac{\gamma}{3}$  and  $\rho = \frac{\gamma}{4}$  so that (3.3) gives

$$\|G - \Phi\|_{\infty} \leq \frac{0.73 \|\psi\|_3^3}{\sqrt{\gamma} \|\psi\|_2^3}.$$

Let us define a two-dimensional anisotropic Gaussian filter as

$$\psi(x_1, x_2) = C \exp\left(-\frac{x_1^2}{2\sigma_1^2} - \frac{x_2^2}{2\sigma_2^2}\right),$$

where  $C$  is a normalization constant. This filter is used frequently in the microscopy experiments we perform and is thus of particular interest. Figure 4 shows practical realizations of stationary processes defined as  $\Lambda \star \psi$ . Note that as  $\sigma_1$  or  $\gamma$  increases, the texture becomes similar to the Gaussian process in the last row. Table 1 quantifies the proximity of the non-Gaussian process to the Gaussian one using Proposition 3.2. The processes can hardly be distinguished from a perceptual point of view when the right hand-side in (3.3) is less than 0.4.

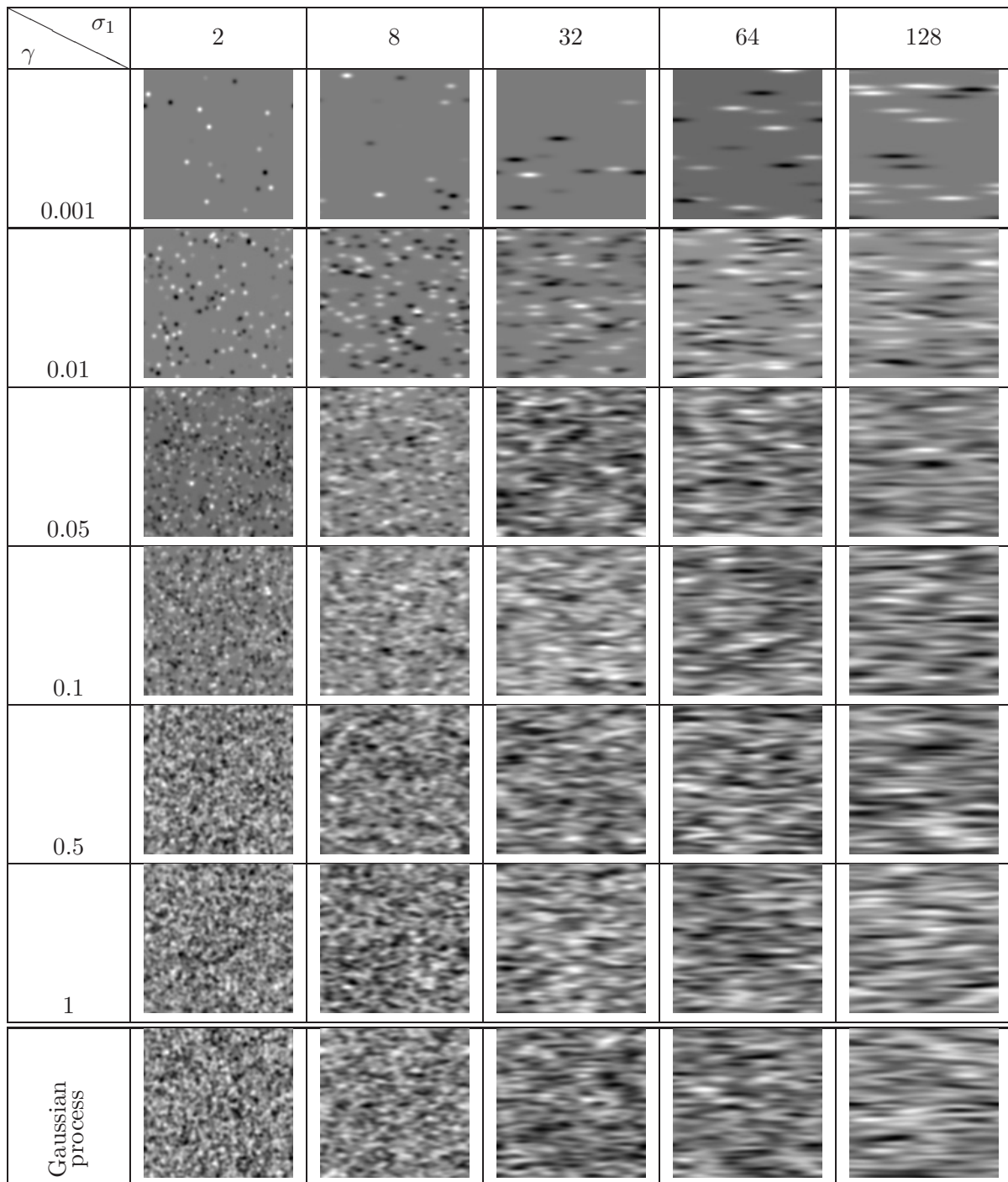
**3.3. Numerical validation.** In the previous sections we showed that in many situations, stationary random processes  $B$  of kind

$$B = \Lambda \star \psi,$$

where  $\Lambda$  denotes a noise process with independent components, can be assimilated to a colored Gaussian noise. A Bayesian approach thus indicates that problem (2.3) can be replaced by the following approximation:

$$(3.4) \quad \text{Find } \boldsymbol{\lambda}(\alpha) = \text{Arg min}_{\boldsymbol{\lambda} \in \mathbb{R}^{n \cdot m}} \left\| \nabla \left( u_0 - \sum_{i=1}^m \psi_i \star \lambda_i \right) \right\|_{\mathbf{1}} + \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2$$

for a particular choice of  $\alpha_i$  discussed later. This new problem has an attractive feature compared to (2.3): it is strongly convex in  $\boldsymbol{\lambda}$ , which implies uniqueness of the minimizer and the existence of efficient minimization algorithms. Unfortunately, it is well known that Bayesian approaches can substantially deviate from the prior models that underlie the MAP estimator [20]. The aim of this section is to validate the proposed approximation experimentally. We consider a problem of stationary noise removal.



**Figure 4.** The first six rows show stationary processes obtained by convolving an anisotropic Gaussian filter with Bernoulli-uniform processes for  $\sigma_2 = 2$ , different values of  $\gamma$ , and different values of  $\sigma_1$ . The last row shows a Gaussian process obtained by convolving Gaussian white noise with the same Gaussian filter.

**Table 1**  
 Values of bound (3.3) w.r.t.  $\gamma$  and  $\sigma_1$ .

$\gamma \backslash \sigma_1$	2	8	32	64	128
0.001	1.00	1.00	1.00	1.00	1.00
0.01	1.00	1.00	0.98	0.82	0.69
0.05	0.88	0.62	0.44	0.37	0.31
0.1	0.62	0.44	0.31	0.26	0.22
0.5	0.28	0.20	0.14	0.12	0.10
1	0.20	0.14	0.10	0.08	0.07

We generate stationary processes from the models described in Example 4 and Figure 4 for different values of  $\gamma$ . Bernoulli-uniform processes are generated from functions  $\phi_i$  that are nonconvex ( $l^0$ -norms), and in this case, problem (2.3) is a hard combinatorial problem. We denoise the image using either a standard  $l^1$ -norm relaxation,

$$\text{Find } \lambda \in \text{Arg min}_{\lambda \in \mathbb{R}^n} \|\nabla(u_0 - \psi \star \lambda)\|_1 + \alpha \|\lambda\|_1,$$

or the  $l^2$ -norm approximation suggested by the previous results:

$$(3.5) \quad \text{Find } \lambda \in \text{Arg min}_{\lambda \in \mathbb{R}^n} \|\nabla(u_0 - \psi \star \lambda)\|_1 + \frac{\alpha}{2} \|\lambda\|_2^2.$$




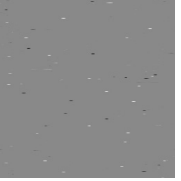
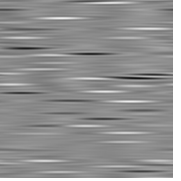



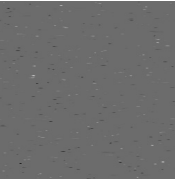
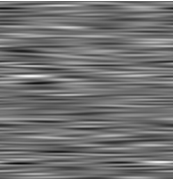



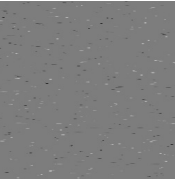
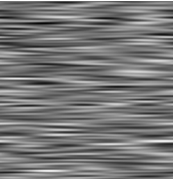



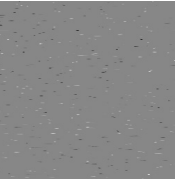
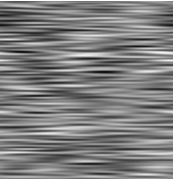



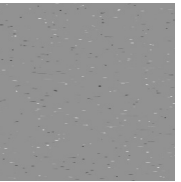
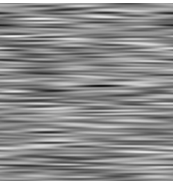



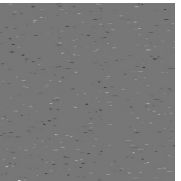
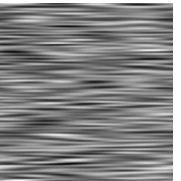
The optimal parameter  $\alpha$  is estimated by dichotomy in order to maximize the SNR of the denoised image. As can be seen in Figure 5, the  $l^1$ -norm approximation provides better results for very sparse Bernoulli processes and the  $l^2$  approximation provides similar or better results when the Bernoulli process gets denser. This confirms the results presented in section 3.1.

The fourth and fifth columns in Figure 5 show an important fact: even though  $b(\alpha)$  is correctly estimated by using an  $l^2$  minimization, it is not necessarily the case for  $\lambda$ . If the user is interested not only in denoising the images but also in finding  $\lambda$ , then precise statistical modeling is necessary: the  $l^2$  estimated component is far from being sparse, while the  $l^1$  estimated one is sparse. Finally, note that the  $l^2$  estimated component does not look like a white noise process. This is another practical example of the distortions that exist between the statistical modeling of processes and the MAP estimation results.

**4. Primal-dual estimation in the  $l^2$ -case.** Motivated by the results presented in the previous section, we focus on the  $l^1$ - $l^2$  problem (3.4). Since the mapping  $\lambda \mapsto \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2$  is strictly convex, this problem admits a unique minimizer. We aim at proposing an automatic estimation of an adequate value of  $\alpha = (\alpha_1, \dots, \alpha_m)$ . In practice the manual tuning of the parameter  $\alpha$  appears to be very difficult since it heavily depends on the filters' shape. A natural choice for the regularization parameter  $\alpha$  is to ensure that

$$(4.1) \quad \|\psi_i \star \lambda_i(\alpha)\| = \|b_i\|$$

for a given norm  $\|\cdot\|$ . This idea is similar but different from Morozov's discrepancy principle [17], which consists in fixing  $\|\lambda_i\|$  and not  $\|\lambda_i \star \psi_i\|$ . In practice,  $\|b_i\|$  is usually unknown, but

0.001	6.02dB 	27.09dB 	16.87dB 		
0.01	6.02dB 	16.41dB 	15.87dB 		
0.05	6.02dB 	17.65dB 	17.53dB 		
0.1	6.02dB 	18.61dB 	18.24dB 		
0.5	6.02dB 	14.29dB 	15.41dB 		
1	6.02dB 	17.67dB 	18.15dB 		

**Figure 5.** Denoising results with the resolution of a  $TV-l^1$  or  $TV-l^2$  problem. From top to bottom: increasing value of  $\gamma$ . From left to right: noisy images, denoised using an  $l^1$  prior, denoised using an  $l^2$  prior, retrieved  $\lambda$  for the  $l^1$ -norm minimization, and retrieved component for the  $l^2$ -norm minimization.

the user usually has an idea of the noise level and can set  $\|b_i\| \simeq \eta_i \|u_0\|$ , where  $\eta_i \in ]0, 1[$  denotes a noise fraction. Our aim in this section is thus to find values of  $\alpha$  to approximately satisfy (4.1).

In what follows, we denote  $b(\alpha) = \psi \star \lambda(\alpha)$  in the case  $m = 1$  where  $\lambda(\alpha)$  is the solution of (3.5), and  $b_i(\alpha) = \psi_i \star \lambda_i(\alpha)$  in the case  $m \geq 2$ .

In order to satisfy (4.1) approximately, our strategy is to find explicit functions  $\ell^- = (\ell_i^-)_{i \in \{1, \dots, m\}}$  and  $\ell^+ = (\ell_i^+)_{i \in \{1, \dots, m\}} : \mathbb{R}_+^m \rightarrow \mathbb{R}_+^m$  such that for every  $\alpha \in \mathbb{R}_+^m$  and every  $i \in \{1, \dots, m\}$ ,

$$(4.2) \quad \ell_i^-(\alpha) \leq \|b_i(\alpha)\| \leq \ell_i^+(\alpha).$$

The interest of such functions appears in the following proposition.

**Proposition 4.1.** *Let  $\alpha$  denote a parameter such that  $\|b_i(\alpha)\| = \gamma_i$  for all  $i$ . Then it necessarily belongs to the set*

$$(\ell^+)^{-1}([\gamma_1, +\infty) \times \dots \times [\gamma_m, +\infty)) \cap (\ell^-)^{-1}((-\infty, \gamma_1] \times \dots \times (-\infty, \gamma_m]).$$

*Proof.* This is a direct consequence of inequality (4.2). ■

The rest of this section is organized as follows. In section 4.1, we provide lower and upper bounds  $\ell^-$  and  $\ell^+$  for  $\|b(\alpha)\|_2$  in the case  $m = 1$  filter. Then we derive upper bounds  $\ell^+$  for  $\|b_i(\alpha)\|_2$  in the general case in section 4.2. When the parameters  $(\alpha_i)_{i \in \{1, \dots, m\}}$  are given, the problem with  $m$  filters is equivalent to a related problem with 1 filter. The link is detailed in section 4.3. Finally, section 4.4 shows how the proposed results can be used in a practical algorithm. The proofs are provided in the appendix.

**4.1. Results for the case  $m = 1$  filter.** We first state our results in the particular case of  $m = 1$  filter in order to clarify the exposition. We give an upper bound  $\ell^+$  in Theorem 4.2 and a lower bound  $\ell^-$  in Proposition 4.3.

**Theorem 4.2.** *Let  $\alpha > 0$ , and denote  $h_k = \psi \star \tilde{\psi} \star \tilde{d}_k$  for  $k \in \{1, \dots, d\}$ , where  $d_k$  denotes the finite difference filter in the direction of  $x_k$ ; see section 2.1. Then*

$$(4.3) \quad \|b(\alpha)\|_2 \leq \frac{\sqrt{n}}{\alpha} \max_{k \in \{1, \dots, d\}} \|\hat{h}_k\|_\infty.$$

*If we further assume that  $\hat{\psi}$  does not vanish, there exists a value  $\bar{\alpha} > 0$  such that for all  $\alpha \in (0, \bar{\alpha}]$ ,  $b(\alpha) = u_0 - u_0^{mean}$ .*

Note that  $\alpha \mapsto \|b(\alpha)\|_2$  is not necessarily monotonically decreasing. Indeed,  $\|b(\alpha)\|_2 = \|\Psi \lambda(\alpha)\|_2$  can be interpreted as the norm of  $\lambda(\alpha)$  in the metric induced by  $\Psi$  and there is no reason that it decreases, although  $\|\lambda(\alpha)\|_2$  decreases [11]. As a consequence, the quantity  $\|u_0 - u_0^{mean}\|_2$  which is an upper bound in a neighborhood of 0 is not necessarily an upper bound for all  $\alpha > 0$ . In our numerical tests, we never encountered a situation where  $\|b(\alpha)\|_2 > \|u_0 - u_0^{mean}\|_2$ . In the following, we make the abuse to refer to  $\min(\frac{\sqrt{n}}{\alpha} \max_{k \in \{1, \dots, d\}} \|\hat{h}_k\|_\infty, \|u_0 - u_0^{mean}\|_2)$  as an ‘‘upper bound.’’

The following proposition provides a lower bound with the same asymptotic decay rate in  $\frac{1}{\alpha}$  for  $\|b(\alpha)\|_2$ .

**Proposition 4.3.** *Assume that all components of  $\hat{\psi}$  are different from 0. Let  $b(\alpha) = \psi \star \lambda(\alpha)$ , where  $\lambda(\alpha)$  is the solution of (3.5). Let  $P_1$  denote the orthogonal projector on the range of*

the operator  $\Psi^T \nabla^T$  and  $b_1 = P_1(\Psi^{-1}u_0)$ . Then if  $\alpha$  is sufficiently large,

$$\|b(\alpha)\|_2 \geq \frac{1}{\alpha} \frac{1}{\|\Psi^{-1}\|_{2 \rightarrow 2}} \frac{\|b_1\|_2}{\|A^+ b_1\|_\infty}.$$

Theorem 4.2 and Proposition 4.3 provide upper and lower bounds that can be used in Proposition 4.1 to find a suitable regularization parameter. Unfortunately the lower bound in Proposition 4.3 is too coarse to be of practical interest for parameter selection. It is still interesting from a theoretical point of view since it provides an asymptotic on  $\|b(\alpha)\|_2$  when  $\alpha \rightarrow +\infty$ . On the other hand, the upper bound provided in Theorem 4.2 turns out to be remarkably tight in practice, as will be demonstrated in the numerical experiments (section 4.5). It can thus be used in practice to get a crude idea of the correct regularization parameter. By setting  $\alpha = \frac{\sqrt{n}}{\gamma} \max_{k \in \{1, \dots, d\}} \|\hat{h}_k\|_\infty$ , we ensure that  $\|b(\alpha)\|_2 \leq \gamma$  and, more importantly, we will see that  $\frac{\|b(\alpha)\|_2}{\gamma}$  does not exceed values from 2 to 15 depending on the filter  $\psi$ .

**4.2. Results for the general case  $m \geq 1$  filters.** In this section, we state results that generalize Theorem 4.2 to the case of  $m \geq 2$  filters.

**Theorem 4.4.** Let  $\alpha = (\alpha_1, \dots, \alpha_m)$  denote positive weights. Let  $h_{i,k} = \psi_i \star \tilde{\psi}_i \star \tilde{d}_k$  for  $k \in \{1, \dots, d\}$ . Then

$$(4.4) \quad \|b_i(\alpha)\|_2 \leq \frac{\sqrt{n}}{\alpha_i} \max_{k \in \{1, \dots, d\}} \|\hat{h}_{i,k}\|_\infty.$$

**Theorem 4.5.** Denote  $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_m) \in R^{n \times nm}$ , and assume that  $\Psi \Psi^T$  has full rank (this is equivalent to the fact that for all  $\xi$ , there exists  $i \in \{1, \dots, m\}$ ,  $\hat{\psi}_i(\xi) \neq 0$ ). Let  $\hat{\lambda}^0(\alpha) = (\hat{\lambda}_1^0, \dots, \hat{\lambda}_m^0)$  be defined by

$$\hat{\lambda}_i^0(\alpha)(\xi) = \begin{cases} 0 & \text{if } \xi = 0, \\ \frac{\tilde{\psi}_i(\xi) \hat{u}_0(\xi)}{\alpha_i \sum_{j=1}^m \frac{|\tilde{\psi}_j(\xi)|^2}{\alpha_j}} & \text{otherwise.} \end{cases}$$

Then there exists a value  $\bar{\alpha} > 0$  such that for all  $\alpha \in ]0, \bar{\alpha}]^m$  the solution  $\lambda(\alpha)$  of problem (3.4) is

$$\lambda(\alpha) = \lambda^0(\alpha).$$

In other words, this theorem states that if the regularization parameter is small enough, then all the data is interpreted as noise.

Theorems 4.4 and 4.5 generalize Theorem 4.2. The algorithm presented in section 4.4 shows how to use this theorem to provide the user warm start regularization parameters.

**4.3. Equivalence with a single filter model.** In section 3, we showed that the following image formation model is rich enough for many applications of interest:

$$u_0 = u + \sum_{i=1}^m \lambda_i \star \psi_i,$$

where  $\lambda_i$  is the realization of a *Gaussian* random vector of distribution  $\mathcal{N}(0, \sigma_i^2 I)$ . Since the sum of independent Gaussian stationary processes is still a Gaussian stationary process, the random vector  $B = \sum_{i=1}^m \Lambda_i \star \psi_i$  can be rewritten simply as  $B = \Lambda \star \psi$ , where  $\Lambda$  is a Gaussian random vector  $\mathcal{N}(0, I)$  and  $\psi$  satisfies

$$(4.5) \quad |\hat{\psi}(\xi)|^2 = \sum_{i=1}^m \sigma_i^2 |\hat{\psi}_i(\xi)|^2 \quad \forall \xi.$$

This condition ensures that both processes have the same covariance matrix.

The above remark has a pleasant consequence: problems (4.6) and (4.7) below are equivalent from a Bayesian point of view if only the noise component  $b = \sum_{i=1}^m \lambda_i \star \psi_i$  and the denoised image  $u$  are sought:

$$(4.6) \quad \min_{\lambda \in \mathbb{R}^{n \times m}} \left\| \nabla \left( u_0 - \sum_{i=1}^m \lambda_i \star \psi_i \right) \right\|_1 + \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2,$$

$$(4.7) \quad \min_{\lambda \in \mathbb{R}^n} \|\nabla(u_0 - \lambda \star \psi)\|_1 + \frac{1}{2} \|\lambda\|_2^2.$$

Hence the optimization can be performed on  $\mathbb{R}^n$  instead of  $\mathbb{R}^{n \cdot m}$ . The following result states that this simplification is also justified from a deterministic point of view.

**Theorem 4.6.** *Let  $\lambda_i(\alpha)$  denote the minimizer of (4.6) and  $\lambda(\alpha)$  denote the minimizer of (4.7). Let  $b_i(\alpha) = \lambda_i(\alpha) \star \psi_i$  and  $b(\alpha) = \lambda(\alpha) \star \psi$ . If condition (4.5) is satisfied, the following equality holds:*

$$(4.8) \quad \sum_{i=1}^m b_i(\alpha) = b(\alpha).$$

Moreover, the noise components  $b_i(\alpha)$  can be retrieved from  $b(\alpha)$  using the following formula:

$$(4.9) \quad \hat{\lambda}_i(\xi) = \begin{cases} \frac{\tilde{\psi}_i(\xi) \hat{b}(\xi)}{\alpha_i \sum_{j=1}^m \frac{|\hat{\psi}_j(\xi)|^2}{\alpha_j}} & \text{if } \sum_{j=1}^m |\hat{\psi}_j(\xi)|^2 \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

In practice, this theorem allows one to divide the computing times and memory requirements by a factor approximately equal to  $m$ .

**4.4. Algorithm.** The following algorithm summarizes how the results presented in this paper allow us to design an effective supervised parameter estimation.

Note that in the computation of  $\hat{\psi}$  (second step) we consider a filter  $\psi$  that has a real Fourier transform. In fact the results of the denoising algorithm do not depend on the phase of each Fourier coefficient of the filter, and for simplicity we restrict ourselves to real coefficients. Moreover, we have chosen  $\sigma = 1$ ; another choice for  $\sigma$  is possible as long as the weight in front of the regularization term is modified adequately.

**4.5. Numerical experiments.** The objective of this section is to validate Theorems 4.2 and 4.4 experimentally and to check that the upper bound in the right-hand side of (4.4) is not too coarse. We also show how the proposed ideas can be used to find reasonable denoising parameters in practice.

**Algorithm 1:** Effective supervised algorithm.**Input:**  $u_0 \in \mathbb{R}^n$ : noisy image. $(\psi_i)_{i \in \{1, \dots, m\}} \in \mathbb{R}^{n \times m}$ : a set of filters. $(\eta_1, \dots, \eta_m) \in [0, 1]^m$ : noise levels.**Output:**  $u \in \mathbb{R}^n$ : denoised image $(b_i)_{i \in \{1, \dots, m\}} \in \mathbb{R}^{n \times m}$ : noise components (satisfying  $\|b_i\|_2 \simeq \eta_i \|u_0\|_2$ ).**begin**

Compute $\alpha_i = \frac{\sqrt{n} \max_{k \in \{1, \dots, d\}} \ \hat{h}_{i,k}\ _\infty}{\ u_0\ _2 \eta_i}$ (see Theorem 4.4).
---

Compute $\hat{\psi} = \sqrt{\sum_{i=1}^m \frac{\ \hat{\psi}_i\ ^2}{\alpha_i}}$ .
--

Find $\lambda \in \text{Arg min}_{\lambda \in \mathbb{R}^n} \ \nabla(u_0 - \lambda \star \psi)\ _1 + \frac{1}{2} \ \lambda\ _2^2$ (see [9]).
--

Compute $u = u_0 - \lambda \star \psi$ .
--

Compute $b = \lambda \star \psi$ .
------------------------------------

Compute $b_i = \lambda_i \star \psi_i$ using Theorem 4.6.
---

**4.5.1. Experiment for  $m = 1$  filter.** In this first experiment, we aim at illustrating Theorem 4.2. We compute the minimizers of (3.4) using an iterative algorithm for various filters, various images, and various values of  $\alpha$ . In all experiments, we assume that  $\Omega = \{-\frac{n_1}{2} + 1, \dots, \frac{n_1}{2}\} \times \{-\frac{n_2}{2} + 1, \dots, \frac{n_2}{2}\}$ . All the images have a size  $n_1 \times n_2 = 512 \times 512$ . The different filters are as follows:

- filter 1 is an isotropic sinc function:  $\psi(x_1, x_2) \propto \frac{\sin(\sigma \sqrt{x_1^2 + x_2^2})}{\sigma \sqrt{x_1^2 + x_2^2}}$  with  $\sigma = 2.56$ ;
- filter 2 is an anisotropic Gaussian function:

$$(4.10) \quad \psi(x_1, x_2) \propto \exp\left(-\frac{x_1^2}{2\sigma_1^2} - \frac{x_2^2}{2\sigma_2^2}\right)$$

with  $\sigma_1 = 1.5$  and  $\sigma_2 = 3$ ; and

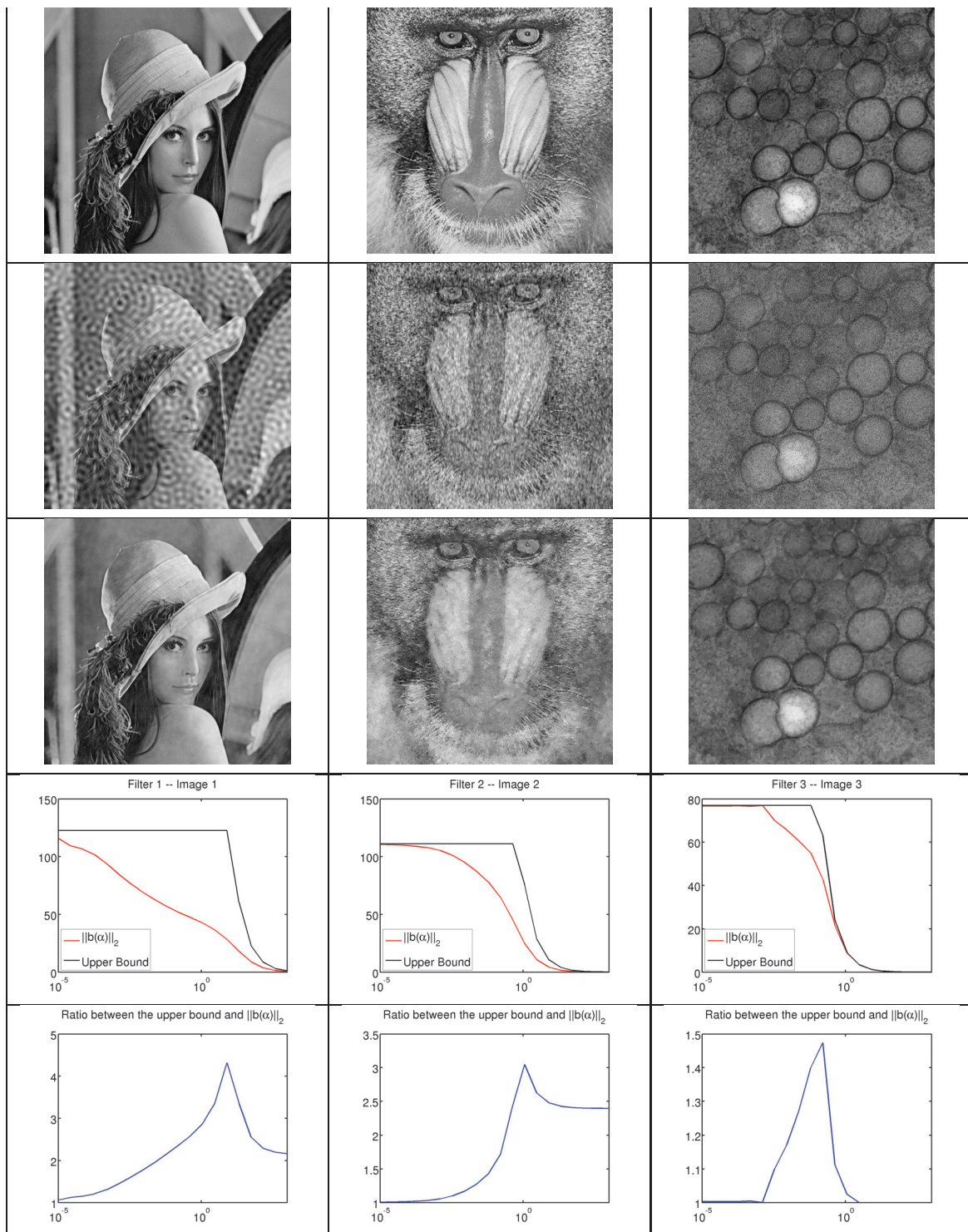
- filter 3 is a Dirac.

The filters were used to produce stationary noises with different covariance matrices. The noisy images  $u_0$  corrupted with each stationary noise can be seen in the second row of Figure 6. The denoised image using a manually selected regularization parameter  $\alpha$  can be seen in the third row of Figure 6.

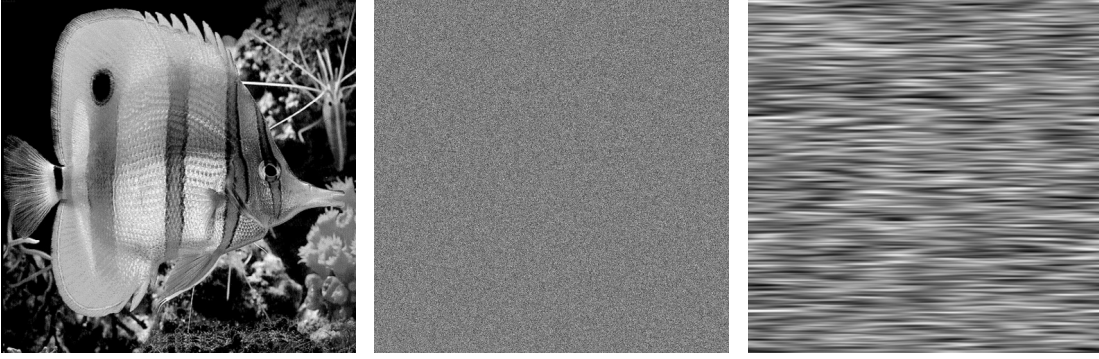
In the fourth and fifth rows, we aim at illustrating Theorem 4.2. We solve problem (3.5) with different values of  $\alpha$  and set  $u_0$  to be the image in the second row. In the fourth row, we plot the values  $\|b(\alpha)\|_2$  in red and  $\min(\frac{\sqrt{n}\|\hat{h}\|_\infty}{\alpha}, \|u_0 - u_0^{mean}\|_2)$  in black. We observe that  $\|b(\alpha)\|_2 \leq \frac{\sqrt{n}\|\hat{h}\|_\infty}{\alpha}$ , which confirms Theorem 4.2. As stated in section 4.1, the quantity  $\|u_0 - u_0^{mean}\|_2$  is not strictly speaking an upper bound, but we could not find examples of practical interest where  $\|b(\alpha)\|_2 \geq \min(\frac{\sqrt{n}\|\hat{h}\|_\infty}{\alpha}, \|u_0 - u_0^{mean}\|_2)$ .

In the fifth row, we plot the ratio between  $\min(\frac{\sqrt{n}\|\hat{h}\|_\infty}{\alpha}, \|u_0 - u_0^{mean}\|_2)$  and  $\|b(\alpha)\|_2$ . As observed in the fourth row, this ratio is always greater than 1. More importantly, it can





**Figure 6.** Comparison of the upper bound in (4.4) with  $\|b(\alpha)\|_2$ . First row: original image. 2nd row: noisy image. 3rd row: denoised using the proposed algorithm. 4th row: comparison of the upper bound (4.4) with  $\|b(\alpha)\|_2$ . Last row: ratio between the upper bound and the true value of  $\|b(\alpha)\|_2$ .



**Figure 7.** From left to right: original image, a realization of  $b_1$ , and a realization of  $b_2$ .

be seen that the upper bound deviates at most from a factor 4.5 from  $\|b(\alpha)\|_2$ . This shows that the upper bound (4.4) can provide a good hint on how to choose a correct value of the regularization parameter. The user can then refine this bound manually to get a visually satisfactory result.

#### 4.5.2. Experiment for $m = 2$ filters.

- In this experiment, we aim to
- validate Theorem 4.4 experimentally and
  - show how it can be used to assist a user with parameter selection.

We consider the problem of denoising the left image in Figure 7 degraded with the sum of two stationary noise components depicted in the middle and right images in Figure 7 with various noise levels. The filters used for this experiment are as follows:

- $\psi_1$  is a Dirac, leading to a standard i.i.d. Gaussian noise; and
- $\psi_2$  is an anisotropic Gaussian defined using (4.10) with  $\sigma_1 = 20$  and  $\sigma_2 = 1$ .

In Figure 8, we show noisy images corresponding to various noise levels. We then apply Algorithm 1 twice to denoise the images. We use two different values for  $(\eta_1, \eta_2)$ :

1. we set  $\eta_i$  to be the exact ratio between the  $l^2$ -norms of the noise components and the image:  $\eta_i = \frac{\|b_i\|_2}{\|u_0\|_2}$ ; and
2. we overestimate the noise level by setting  $\eta_i = \gamma_i \frac{\|b_i\|_2}{\|u_0\|_2}$  with  $\gamma_i > 1$ .

Tables 2 and 3 present the ratio between the estimated noise component norm and the true noise component norm  $\frac{\|b_i(\alpha)\|_2}{\|b_i\|_2}$  for both experiments. We can observe in Table 2 that the ratios do not exceed one order of magnitude. This means that if the true noise level is provided to the algorithm, the estimated noise components will have the right order of magnitude. In Table 3, we set  $\gamma_1 = 1.5$  and  $\gamma_2 = 20$ . We observe that the ratios are close to 1, meaning that the parameters  $\alpha_1$  and  $\alpha_2$  are chosen appropriately. This experiment suggests that the overestimation parameters  $\gamma_i$  are valid over a wide range of noise levels and depend mostly on the filter type. The denoised images and the SNRs are displayed in Figure 9.

In practice, the true noise levels are unknown and this experiment shows that if the user provides overestimated noise levels to Algorithm 1, reasonable denoising results are expected.

Finally, Figure 10 shows the restoration results for a given noise level when  $(\gamma_1, \gamma_2) = (1, 1)$  (middle) and  $(\gamma_1, \gamma_2) = (1.5, 20)$  (left). This experiment demonstrates that our estimation procedure is only rough: if the true noise level is provided to the algorithm, then the images

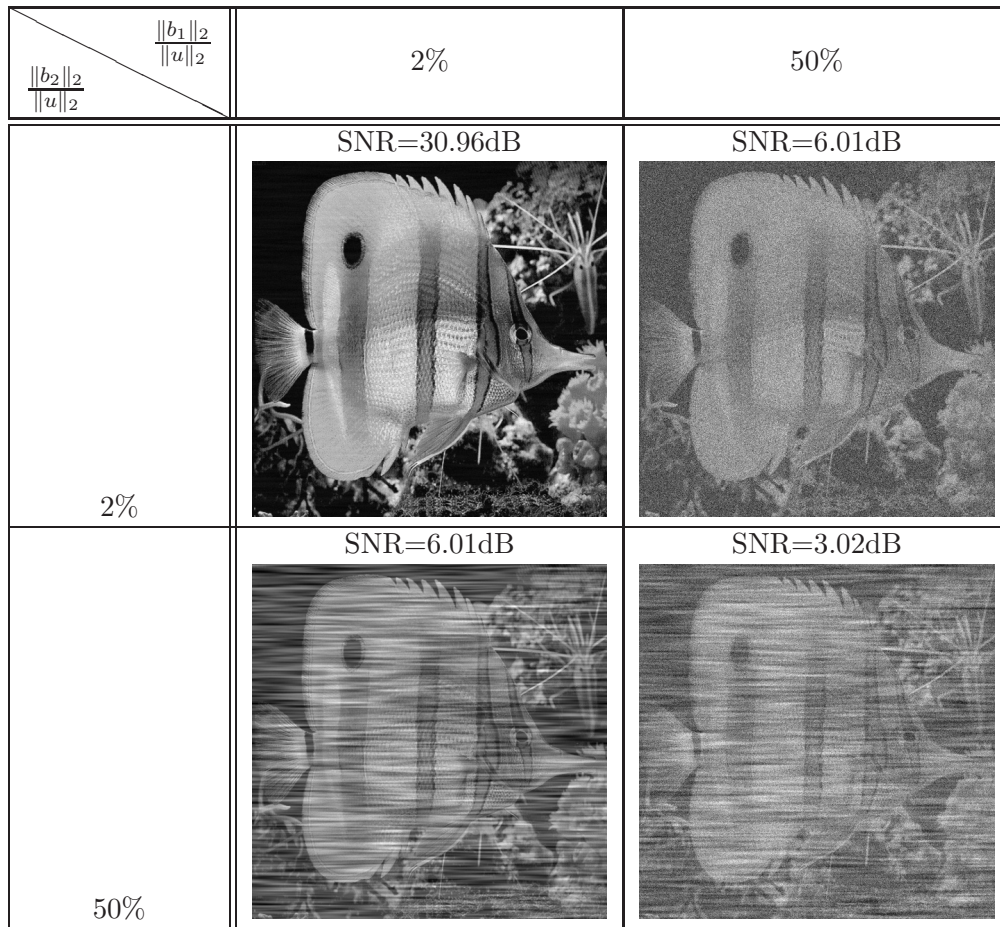


Figure 8. Noisy image with different noise levels for  $b_1$  and  $b_2$ .

Table 2

Values of the ratios  $\frac{\|b_i(\alpha)\|_2}{\|b_i\|_2}$  for various noise levels by setting  $\gamma_1 = 1$  and  $\gamma_2 = 1$ .

$\frac{\ b_2\ _2}{\ u\ _2}$	$\frac{\ b_1\ _2}{\ u\ _2}$	2%	5%	20%	50%	2%	5%	20%	50%
2%		1.28	1.30	1.38	1.46	9.01	6.39	4.17	3.29
5%		1.33	1.34	1.42	1.49	10.39	7.20	4.16	3.25
20%		1.35	1.35	1.39	1.45	13.66	10.13	5.11	3.68
50%		1.32	1.32	1.33	1.37	15.60	13.61	7.16	4.61
		ratio $\frac{\ b_1(\alpha)\ _2}{\ b_1\ _2}$				ratio $\frac{\ b_2(\alpha)\ _2}{\ b_2\ _2}$			

are not sufficiently denoised. Let us recall that our aim in this paper is just to assist the user with parameter selection, so that this problem is no too serious. In practice, if the user observes that the retrieved noise component level is around 10 times too small, he can just launch the algorithm once more by indicating a 10 times higher noise level.

Table 3

Values of the ratios  $\frac{\|b_i(\alpha)\|_2}{\|b_i\|_2}$  for various noise levels by setting  $\gamma_1 = 1.5$  and  $\gamma_2 = 20$ .

$\frac{\ b_2\ _2}{\ u\ _2}$ \diagdown $\frac{\ b_1\ _2}{\ u\ _2}$	2%	5%	20%	50%	2%	5%	20%	50%
2%	0.91	0.91	0.92	0.96	1.12	1.20	1.19	1.13
5%	0.99	0.98	0.99	1.01	1.05	1.17	1.17	1.13
20%	1.08	1.08	1.08	1.08	0.97	1.11	1.19	1.14
50%	1.09	1.09	1.09	1.09	0.90	1.10	1.23	1.19
	ratio $\frac{\ b_1(\alpha)\ _2}{\ b_1\ _2}$				ratio $\frac{\ b_2(\alpha)\ _2}{\ b_2\ _2}$			

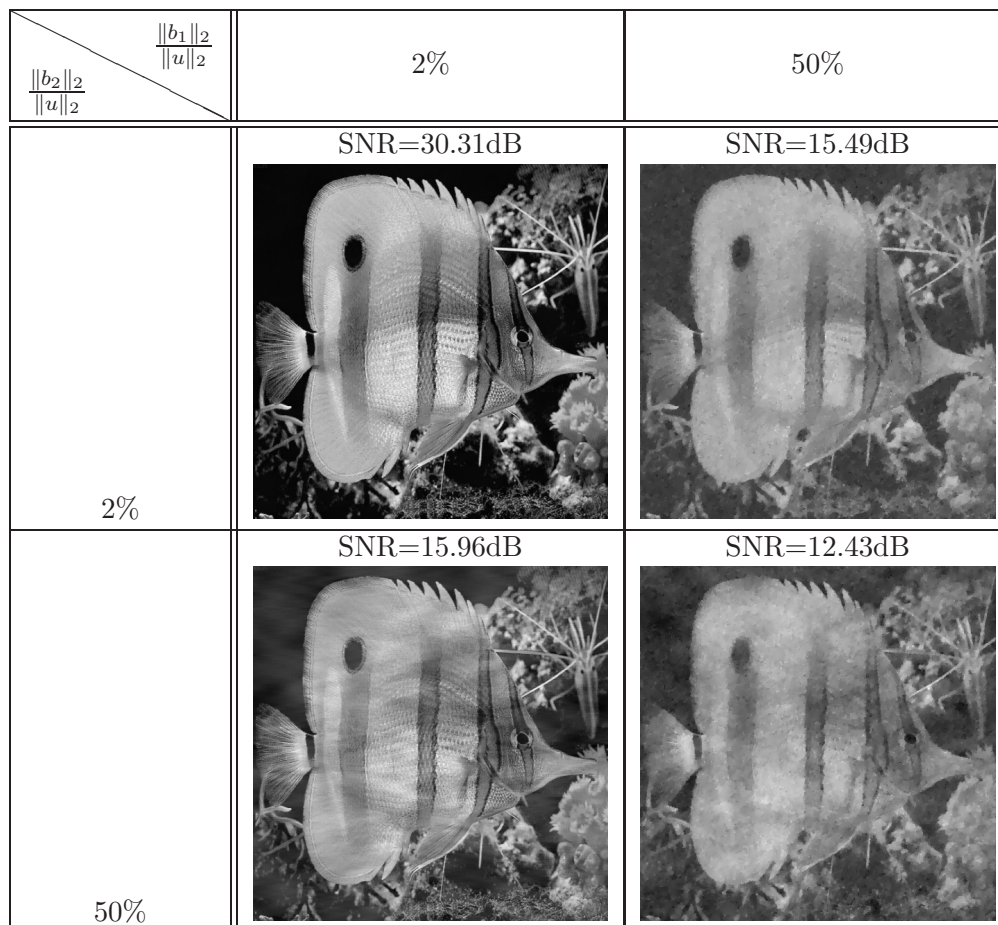
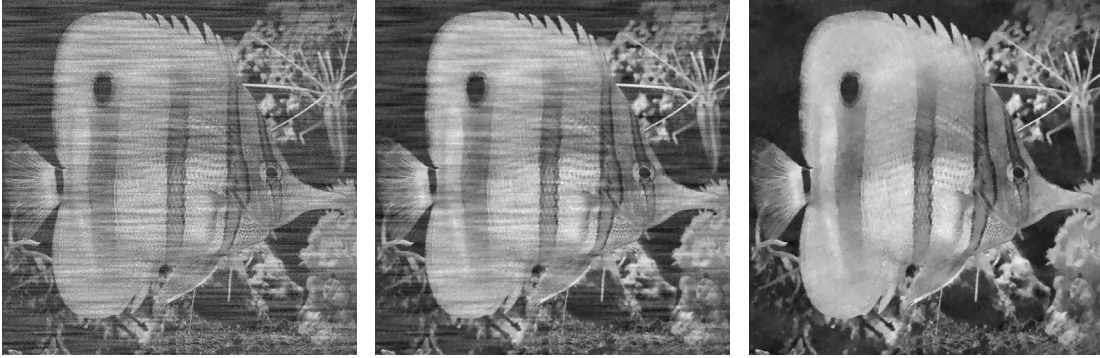


Figure 9. Restored images with different noise levels with  $\gamma_1 = 1.5$  and  $\gamma_2 = 20$ .

**Conclusion.** This paper focused on the problem of stationary noise removal using variational methods. In the first part, we showed that assuming the noise to be Gaussian is reasonable under conditions that are met in many applications, such as destriping. In the second part we thus concentrated on variational problems that consist in minimizing  $l^1$ - $l^2$  functionals. We derived upper and lower bounds on the  $l^2$ -norm of the solutions of these



**Figure 10.** From left to right: noisy image (PSNR=10.97dB) with  $\frac{\|b_1\|_2}{\|u\|_2} = 20\%$  and  $\frac{\|b_2\|_2}{\|u\|_2} = 20\%$ , denoised image (PSNR=14.51dB) by setting  $\gamma_1 = 1$  and  $\gamma_2 = 1$ , and denoised image (PSNR=17.01dB) by setting  $\gamma_1 = 1.5$  and  $\gamma_2 = 20$ .

functionals and showed that they can be used for simplifying the task of parameter selection. We also provided a numerical trick that allows us to drastically reduce the computing times for cases where the noise is described as a sum of stationary processes. Overall this work allows us to strongly reduce the computing times, to ease the parameter selection, and to make our algorithms robust to different conditions. Note that the proofs' structures proposed here can be used to tackle more general classes of problems, where the total variation is replaced by general regularizing priors and the convolution operator is replaced by operators diagonal in other bases (e.g., discrete cosine transform, wavelet).

As a perspective, let us notice that the lower bound proposed in Proposition 4.3 is coarse, and it would be interesting to obtain tighter results highlighting why the upper bound is near tight in practice. We also plan to study the problem of deterministic parameter selection in a more general setting such as  $l^p$ - $l^q$  functionals.

The results implemented in this paper were implemented both on MATLAB and on a Fiji plugin and are available from <http://www.math.univ-toulouse.fr/~weiss/PageCodes.html>.

**5. Appendix.** In this section we provide detailed proofs of the results presented in section 4.

**5.1. Proof of Theorem 4.4.** Theorem 4.4 is a direct consequence of Lemma 5.1 and Proposition 5.2 below.

**Lemma 5.1.** Let  $\|\cdot\|_N$  denote a norm on  $\mathbb{R}^n$ , and let  $\lambda = (\lambda_i)_{1 \leq i \leq m}$  be the solution of problem (3.4). The following inequality holds:

$$(5.1) \quad \|\psi_i \star \lambda_i(\alpha)\|_N \leq \frac{1}{\alpha_i} \|\Psi_i \Psi_i^T \nabla^T\|_{\infty \rightarrow N}.$$

*Proof.* Problem (3.4) can be recast as the following saddle-point problem:

$$\min_{\lambda \in \mathbb{R}^{n \times m}} \max_{\mathbf{q} \in \mathbb{R}^{n \times d}, \|\mathbf{q}\|_{\infty} \leq 1} \left\langle \nabla \left( u_0 - \sum_{i=1}^m \lambda_i \star \psi_i \right), \mathbf{q} \right\rangle + \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2.$$

The dual problem obtained using the Fenchel–Rockafellar duality [8, p. 161] or [22, p. 393] reads as

$$(5.2) \quad \max_{\mathbf{q} \in \mathbb{R}^{n \times d}, \|\mathbf{q}\|_\infty \leq 1} \min_{\lambda \in \mathbb{R}^{n \times m}} \left\langle \nabla \left( u_0 - \sum_{i=1}^m \lambda_i \star \psi_i \right), \mathbf{q} \right\rangle + \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2.$$

Let  $\mathbf{q}(\alpha)$  denote the solution of the dual problem (5.2). The primal-dual optimality conditions imply that

$$(5.3) \quad \lambda_i(\alpha) = -\frac{\Psi_i^T \nabla^T \mathbf{q}(\alpha)}{\alpha_i}.$$

Using (5.3), we obtain  $\psi_i \star \lambda_i(\alpha) = -\frac{1}{\alpha_i} \Psi_i \Psi_i^T \nabla^T \mathbf{q}(\alpha)$ . Moreover,  $\|\mathbf{q}(\alpha)\|_\infty \leq 1$ . It then suffices to use the operator norm definition (2.1) to obtain inequality (5.1). ■

In order to use inequality (5.1) for practical purposes, one needs to estimate upper bounds for  $\|\cdot\|_{\infty \rightarrow N}$ . Unfortunately, it is known to be a hard mathematical problem, as shown in [13, 23]. The special case  $N = 2$ , which corresponds to a Gaussian noise assumption, can be treated analytically.

**Proposition 5.2.** *Let*

$$\mathbf{h}_i = \begin{pmatrix} h_{i,1} \\ \vdots \\ h_{i,d} \end{pmatrix}$$

with  $h_{i,k} = \psi_i \star \tilde{\psi}_i \star \tilde{d}_k$ . Then

$$\|\Psi_i \Psi_i^T \nabla^T\|_{\infty \rightarrow 2} = \sqrt{n} \max_{k \in \{1, \dots, d\}} \|\hat{h}_{i,k}\|_\infty.$$

*Proof.* First, remark that

$$\begin{aligned} \|\Psi_i \Psi_i^T \nabla^T\|_{\infty \rightarrow 2} &= \max_{\|\mathbf{q}\|_\infty \leq 1} \left\| \sum_{k=1}^d h_{i,k} \star q_k \right\|_2 \\ &\leq \max_{\|\mathbf{q}\|_2 \leq \sqrt{n}} \left\| \sum_{k=1}^d h_{i,k} \star q_k \right\|_2 \\ &= \sqrt{n} \max_{\sum_{k=1}^d \|\hat{q}_k\|_2^2 \leq 1} \left\| \sum_{k=1}^d \hat{h}_{i,k} \odot \hat{q}_k \right\|_2 \\ &= \sqrt{n} \max_{k \in \{1, \dots, d\}} \|\hat{h}_{i,k}\|_\infty. \end{aligned}$$

In order to obtain the reverse inequality, let us define

$$\mathcal{Q}_k = \{\mathbf{q} \in \mathbb{R}^{n \times d}, q_k \in \{f_1, \dots, f_n\} \text{ and } q_i = 0, i \in \{1, \dots, d\} \setminus \{k\}\}$$

and the Fourier transform of this set, which is

$$\hat{\mathcal{Q}}_k = \{\hat{\mathbf{q}} \in \mathbb{C}^{n \times d}, \hat{q}_k \in \{ne_1, \dots, ne_n\} \text{ and } \hat{q}_i = 0, i \in \{1, \dots, d\} \setminus \{k\}\}.$$

Let us denote  $\mathcal{Q} = \cup_{k=1}^d Q_k$  and  $\hat{\mathcal{Q}} = \cup_{k=1}^d \hat{Q}_k$ . Thus we obtain

$$\begin{aligned} \|\Psi_i \Psi_i^T \nabla^T\|_{\infty \rightarrow 2} &= \max_{\|\mathbf{q}\|_{\infty} \leq 1} \left\| \sum_{k=1}^d h_{i,k} \star q_k \right\|_2 \\ &\geq \max_{\mathbf{q} \in \mathcal{Q}} \left\| \sum_{k=1}^d h_{i,k} \star q_k \right\|_2 \\ &= \max_{\mathbf{q} \in \hat{\mathcal{Q}}} \frac{\left\| \sum_{k=1}^d \hat{h}_{i,k} \odot \hat{q}_k \right\|_2}{\sqrt{n}} \\ &= \sqrt{n} \max_{k \in \{1, \dots, d\}} \|\hat{h}_{i,k}\|_{\infty}, \end{aligned}$$

which ends the proof.  $\blacksquare$

**5.2. Proof of Theorem 4.5.** Denote  $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_m) \in R^{n \times nm}$ , and assume that  $\Psi \Psi^T$  has full rank. This condition ensures the existence of  $\lambda$  satisfying  $\sum_{i=1}^m \lambda_i \star \psi_i = u_0 - u_0^{\text{mean}}$ , where  $u_0^{\text{mean}}$  denotes the mean of  $u_0$ .

**Proposition 5.3.** Let  $\lambda^0(\alpha)$  denote the solution of the following problem:

$$\lambda^0(\alpha) = \underset{\lambda \in \mathbb{R}^{n \times m} \text{ s.t. } \sum_{i=1}^m \lambda_i \star \psi_i = u_0 - u_0^{\text{mean}}}{\text{Arg min}} \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2.$$

Then the vector  $\hat{\lambda}^0(\alpha) = (\hat{\lambda}_1^0, \dots, \hat{\lambda}_m^0)$  is equal to

$$(5.4) \quad \hat{\lambda}_i^0(\xi) = \begin{cases} 0 & \text{if } \xi = 0, \\ \frac{\bar{\psi}_i(\xi) \hat{u}_0(\xi)}{\alpha_i \sum_{j=1}^m \frac{|\psi_j(\xi)|^2}{\alpha_j}} & \text{otherwise.} \end{cases}$$

*Proof.* First, notice that the full rank hypothesis on  $\Psi \Psi^T$  is equivalent to assuming that for all  $\xi$ , there exists  $i \in \{1, \dots, m\}$ ,  $\hat{\psi}_i(\xi) \neq 0$ , since  $\Psi_i = \mathcal{F}^{-1} \text{diag}(\hat{\psi}_i) \mathcal{F}$ . Then

$$\begin{aligned} \lambda^0(\alpha) &= \underset{\lambda \in \mathbb{R}^{n \times m} \text{ s.t. } \sum_{i=1}^m \lambda_i \star \psi_i = u_0 - u_0^{\text{mean}}}{\text{Arg min}} \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2 \\ &= \underset{\lambda \in \mathbb{R}^{n \times m} \text{ s.t. } \sum_{i=1}^m \hat{\lambda}_i \odot \hat{\psi}_i = u_0 - \widehat{u_0^{\text{mean}}}}{\text{Arg min}} \sum_{i=1}^m \frac{\alpha_i}{2} \|\hat{\lambda}_i\|_2^2. \end{aligned}$$

This problem can be decomposed as  $n$  independent optimization problems of size  $m$ . If  $\xi = 0$ , it remains to observe that  $u_0 - \widehat{u_0^{\text{mean}}}(0) = 0$  since  $u_0 - u_0^{\text{mean}}$  has zero mean. For  $\xi \neq 0$ , this amounts to solving the  $m$ -dimensional quadratic problem:

$$\underset{\hat{\lambda}(\xi) \in \mathbb{C}^m}{\text{Arg min}} \sum_{i=1}^m \frac{\alpha_i}{2} |\hat{\lambda}_i(\xi)|_2^2 \quad \text{s.t.} \quad \sum_{i=1}^m \hat{\psi}_i(\xi) \hat{\lambda}_i(\xi) = \hat{u}_0(\xi).$$

It is straightforward to derive the solution (5.4) analytically.  $\blacksquare$

**Corollary 5.4.** *If  $\hat{\psi}_i(\xi) = 0$ , then  $\hat{\lambda}_i^0(\alpha)(\xi) = 0$ , and if  $\hat{\psi}_i(\xi) \neq 0$ , then  $|\hat{\lambda}_i^0(\alpha)(\xi)| \leq \frac{|\hat{u}_0(\xi)|}{|\hat{\psi}_i(\xi)|}$ .*

Therefore, for every  $\alpha$ ,  $\|\hat{\lambda}_i^0(\alpha)\|_2 \leq \|\hat{u}_0 \odot \hat{\psi}_i\|_2$  (with the convention of replacing by 0 the terms where the denominator vanishes).

*Proof.* It is a direct consequence of (5.4). ■

*Proof of Theorem 4.5.* Let  $F_\alpha(\boldsymbol{\lambda}) = G(\boldsymbol{\lambda}) + \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2$  with  $G(\boldsymbol{\lambda}) = \|(\nabla(\Psi\boldsymbol{\lambda} - u_0))\|_1$ . The objective is to prove that  $\partial F_\alpha(\boldsymbol{\lambda}^0(\alpha)) \ni 0$  for sufficiently small  $\alpha$ . Denote  $C = \{\beta \mathbf{1}_{\mathbb{R}^n}, \beta \in \mathbb{R}\}$  the space of constant images. Since  $\text{Ker}(\nabla) = C$  and  $\Psi\boldsymbol{\lambda}^0(\alpha) - u_0 \in C$ ,  $\nabla(\Psi\boldsymbol{\lambda}^0 - u_0) = 0$ . Standard results of convex analysis lead to

$$\begin{aligned} \partial G(\boldsymbol{\lambda}^0(\alpha)) &= \Psi^T \nabla^T \partial_{\|\cdot\|_1}(0) \\ &= \Psi^T \nabla^T Q, \end{aligned}$$

where  $Q = \{\mathbf{q} \in \mathbb{R}^{n \cdot d} \text{ s.t. } \|\mathbf{q}(\mathbf{x})\|_2 \leq 1 \ \forall \mathbf{x} \in \Omega\}$  is the unit ball associated with the dual norm  $\|\cdot\|_1^*$ . Since  $\text{Ran}(\nabla^T) = \text{Ker}(\nabla)^\perp$ , we deduce that  $\text{Ran}(\nabla^T) = C^\perp$  is the set of images with zero mean. Therefore, since  $Q$  has a nonempty interior, there exists  $\gamma > 0$  such that  $\nabla^T Q \supset B(0, \gamma) \cap C^\perp$ , where  $B(0, \gamma)$  denotes a Euclidean ball of radius  $\gamma$ . Therefore

$$\begin{aligned} (\partial F_\alpha(\boldsymbol{\lambda}^0(\alpha)))_i &= (\partial G(\boldsymbol{\lambda}^0(\alpha)))_i + \alpha_i \lambda_i^0(\alpha) \\ &= (\Psi^T \nabla^T Q)_i + \alpha_i \lambda_i^0(\alpha) \\ &\supset (\Psi^T (B(0, \gamma) \cap C^\perp))_i + \alpha_i \lambda_i^0(\alpha). \end{aligned}$$

Note that

$$\Psi^T (B(0, \gamma) \cap C^\perp) = \Psi_1^T (B(0, \gamma) \cap C^\perp) \times \dots \times \Psi_m^T (B(0, \gamma) \cap C^\perp).$$

Since convolution operators preserve  $C^\perp$ , we obtain

$$\Psi_i^T (B(0, \gamma) \cap C^\perp) \supset B(0, \gamma_i) \cap \text{Ran}(\Psi_i) \cap C^\perp \quad \text{for some } \gamma_i > 0.$$

Now it remains to remark that Proposition 5.3 ensures

$$\boldsymbol{\lambda}^0(\alpha) \in (\text{Ran}(\Psi_1) \cap C^\perp) \times \dots \times (\text{Ran}(\Psi_m) \cap C^\perp).$$

Therefore, for  $\alpha_i \leq \frac{\gamma_i}{\|\hat{u}_0 \odot \hat{\psi}_i\|_2}$ ,

$$(\partial F_\alpha(\boldsymbol{\lambda}^0))_i \supset (B(0, \gamma_i) \cap \text{Ran}(\Psi_i) \cap C^\perp) + \alpha_i \lambda_i^0(\alpha) \ni 0.$$

In view of Corollary 5.4 it suffices to set  $\bar{\alpha} = \min_{i \in \{1, \dots, m\}} \frac{\gamma_i}{\|\hat{u}_0 \odot \hat{\psi}_i\|_2}$  to conclude the proof. ■

**5.3. Proof of Proposition 4.3.** We now concentrate on problem (4.7) in the case of  $m = 1$  filter and provide a lower bound on  $\|b(\alpha)\|_2$ . We assume that  $\Psi$  is invertible, meaning that every component of  $\hat{\psi}$  is nonzero.



The dual problem of

$$(5.5) \quad \min_{\lambda \in \mathbb{R}^n} \|\nabla(u_0 - \lambda \star \psi)\|_1 + \frac{\alpha}{2} \|\lambda\|_2^2$$

is

$$(5.6) \quad \max_{\mathbf{q} \in \mathbb{R}^{n \times d}, \|\mathbf{q}\|_\infty \leq 1} \langle \nabla u_0, \mathbf{q} \rangle - \frac{1}{2\alpha} \|\Psi^T \nabla^T \mathbf{q}\|_2^2.$$

The solution  $\lambda(\alpha)$  of (5.5) can be deduced from the solution  $\mathbf{q}(\alpha)$  of (5.6) by using the primal-dual relationship

$$\lambda(\alpha) = -\frac{1}{\alpha} \Psi^T \nabla^T \mathbf{q}(\alpha).$$

We can write

$$(5.7) \quad \begin{aligned} & \text{Arg max}_{\mathbf{q} \in \mathbb{R}^{n \times d}, \|\mathbf{q}\|_\infty \leq 1} \langle \nabla u_0, \mathbf{q} \rangle - \frac{1}{2\alpha} \|\Psi^T \nabla^T \mathbf{q}\|_2^2 \\ &= \text{Arg min}_{\mathbf{q} \in \mathbb{R}^{n \times d}, \|\mathbf{q}\|_\infty \leq 1} \frac{1}{2} \|\Psi^T \nabla^T \mathbf{q} - \alpha \Psi^{-1} u_0\|_2^2. \end{aligned}$$

Let  $P_1$  denote the orthogonal projector on  $\text{Ran}(\Psi^T \nabla^T)$  and  $P_2$  denote the orthogonal projector on  $\text{Ran}(\Psi^T \nabla^T)^\perp$ . Using these operators, we can write  $\alpha \Psi^{-1} u_0 = \alpha b_1 + \alpha b_2$ , where  $b_1 = P_1 \Psi^{-1} u_0$  and  $b_2 = P_2 \Psi^{-1} u_0$ . Problem (5.7) becomes

$$\mathbf{q}(\alpha) = \text{Arg min}_{\mathbf{q} \in \mathbb{R}^{n \times d}, \|\mathbf{q}\|_\infty \leq 1} \frac{1}{2} \|\Psi^T \nabla^T \mathbf{q} - \alpha b_1\|_2^2.$$

Let us denote  $A = \Psi^T \nabla^T$  and  $\mathbf{q}'(\alpha) = \frac{A^+ b_1}{\|A^+ b_1\|_\infty}$ . Since  $b_1 \in \text{Ran}(A)$ ,  $A \mathbf{q}'(\alpha) = \frac{b_1}{\|A^+ b_1\|_\infty}$ . Thus, as long as  $\|A^+ b_1\|_\infty \geq \frac{1}{\alpha}$ ,

$$\begin{aligned} \|\|A \mathbf{q}(\alpha)\|_2 - \alpha \|b_1\|_2\| &\leq \|A \mathbf{q}(\alpha) - \alpha b_1\|_2 \\ &= \min_{\|\mathbf{q}\|_\infty \leq 1} \|A \mathbf{q} - \alpha b_1\|_2 \\ &\leq \|A \mathbf{q}'(\alpha) - \alpha b_1\|_2 \\ &= \left\| \frac{b_1}{\|A^+ b_1\|_\infty} - \alpha b_1 \right\|_2 \\ &= \left( \alpha - \frac{1}{\|A^+ b_1\|_\infty} \right) \|b_1\|_2. \end{aligned}$$

Since  $A \mathbf{q}(\alpha)$  is a projection of  $\alpha b_1$  on a convex set that contains the origin,  $\alpha \|b_1\|_2 \geq \|A \mathbf{q}(\alpha)\|_2$  and we get

$$\alpha \|b_1\|_2 - \|A \mathbf{q}(\alpha)\|_2 \leq \left( \alpha - \frac{1}{\|A^+ b_1\|_\infty} \right) \|b_1\|_2,$$

which is equivalent to

$$\|A \mathbf{q}(\alpha)\|_2 \geq \frac{b_1}{\|A^+ b_1\|_\infty}.$$

Since  $\lambda(\alpha) = -\frac{1}{\alpha}A\mathbf{q}(\alpha)$ , we get

$$\|\lambda(\alpha)\|_2 \geq \frac{1}{\alpha} \frac{\|b_1\|_2}{\|A^+b_1\|_\infty},$$

and since  $\lambda = \Psi^{-1}b$ , we obtain

$$\|b(\alpha)\|_2 \geq \frac{1}{\alpha} \|\Psi^{-1}\|_{2 \rightarrow 2}^{-1} \frac{\|b_1\|_2}{\|A^+b_1\|_\infty}.$$

**5.4. Proof of Theorem 4.6.** Theorem 4.6 is a simple consequence of a more general result described below.

Let  $F$  be a convex lower semicontinuous (l.s.c.) function, and let  $\Psi : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\Psi_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $i = 1 \dots m$ , denote linear operators. Define

$$(P_1) \quad (\lambda_i^\dagger)_{1 \leq i \leq m} = \underset{(\lambda_i)_{1 \leq i \leq m} \in (\mathbb{R}^n)^m}{\text{Arg min}} F \left( \sum_{i=1}^m \Psi_i \lambda_i \right) + \frac{1}{2} \sum_{i=1}^m \|\lambda_i\|_2^2$$

and

$$(P_2) \quad \lambda^\dagger = \underset{\lambda \in \mathbb{R}^n}{\text{Arg min}} F(\Psi\lambda) + \frac{1}{2} \|\lambda\|_2^2.$$

**Proposition 5.5.** *If the operators  $\Psi$  and  $(\Psi_i)_{i=1, \dots, m}$  satisfy the relation*

$$\Psi\Psi^* = \sum_{i=1}^m \Psi_i\Psi_i^*,$$

then the solutions  $(\lambda_i^\dagger)_{1 \leq i \leq m}$  of  $(P_1)$  and  $\lambda^\dagger$  of  $(P_2)$  are related by

$$\Psi\lambda^\dagger = \sum_{i=1}^m \Psi_i\lambda_i^\dagger.$$

*Proof.* We define  $\Psi : \mathbb{R}^{n \times m} \rightarrow \mathbb{R}^n$  by  $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_m)$  so that for

$$\lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_m \end{pmatrix},$$

$\Psi\lambda = \sum_{i=1}^m \Psi_i\lambda_i$ . The optimality condition of  $(P_1)$  reads as

$$\Psi^* \partial F(\Psi\lambda^\dagger) + \sum_{i=1}^m \lambda_i^\dagger \ni 0,$$

and the optimality condition of  $(P_2)$  reads as

$$\Psi^* \partial F(\Psi\lambda^\dagger) + \lambda^\dagger \ni 0.$$

The minimization problem  $(P_2)$  admits a unique minimizer denoted by  $\lambda^\dagger$ . By hypothesis,  $\Psi\Psi^* = \Psi\Psi^*$ , and hence  $\text{Ran}\Psi = \text{Ran}\Psi$  and there exists  $\lambda_0$  such that

$$\Psi\lambda^\dagger = \Psi\lambda_0.$$

The optimality condition of  $(P_2)$  implies that

$$0 \in \Psi\Psi^*\partial F(\Psi\lambda^\dagger) + \Psi\lambda^\dagger,$$

and hence

$$0 \in \Psi\Psi^*\partial F(\Psi\lambda_0) + \Psi\lambda_0 = \Psi(\Psi^*\partial F(\Psi\lambda_0) + \lambda_0).$$

This proves that there exists  $\lambda' \in (\Psi^*\partial F(\Psi\lambda_0) + \lambda_0) \cap (\text{Ker}\Psi)$ . If we choose such a  $\lambda'$  and set  $\lambda = \lambda_0 - \lambda'$ , we have

$$\Psi^*\partial F(\Psi\lambda) + \lambda = \Psi^*\partial F(\Psi\lambda_0) + \lambda_0 - \lambda' \ni 0.$$

This implies that  $\lambda$  is the minimizer of  $(P_1)$  and we have

$$\Psi\lambda = \Psi(\lambda_0 - \lambda') = \Psi\lambda_0 = \Psi\lambda^\dagger,$$

which ends the proof. ■

Let us now turn to the proof of Theorem 4.6.

*Proof.* To obtain (4.8), it suffices to make the change of variables  $\lambda'_i = \frac{\lambda_i}{\sqrt{\alpha_i}}$  in problem  $(P_1)$  and to apply Proposition 5.5 together with condition (4.5). To obtain (4.9), it remains to observe that since  $\sum_{k=1}^m b_i(\alpha) = b(\alpha)$ , the determination of  $\lambda_i$  boils down to the following quadratic problem:

$$\begin{aligned} (\lambda_i(\alpha))_{i \in \{1, \dots, m\}} &= \underset{\sum_{i=1}^m \lambda_i \star \psi_i = b(\alpha)}{\text{Arg min}} \sum_{i=1}^m \frac{\alpha_i}{2} \|\lambda_i\|_2^2 \\ &= \underset{\sum_{i=1}^m \hat{\lambda}_i \odot \hat{\psi}_i = \hat{b}(\alpha)}{\text{Arg min}} \sum_{i=1}^m \frac{\alpha_i}{2} \|\hat{\lambda}_i\|_2^2. \end{aligned}$$

The solution of this problem can be obtained analytically by deriving its optimality conditions. It leads to (4.9). ■

**Acknowledgments.** The authors wish to thank Gabriele Steidl and Mila Nikolova for their constructive remarks on the paper. They thank Fabrice Gamboa and Frank Barthe for stimulating discussions on stationary processes. They also wish to thank the anonymous referees for their suggestions, which helped a lot in clarifying the paper.

REFERENCES

[1] J. F. AUJOL AND A. CHAMBOLLE, *Dual norms and image decomposition models*, Int. J. Comput. Vis., 63 (2005), pp. 85–104.  
 [2] F. BAUSS, M. NIKOLOVA, AND G. STEIDL, *Fully smoothed l1-TV models: Bounds for the minimizers and parameter choice*, J. Math. Imaging Vision, 48 (2014), pp. 295–307.

- [3] P. BILLINGSLEY, *Convergence of Probability Measures*, Wiley Ser. Probab. Stat. 493, Wiley-Interscience, New York, 2009.
- [4] H. CARFANTAN AND J. IDIER, *Statistical linear destriping of satellite-based pushbroom-type images*, IEEE Trans. Geosci. Remote Sensing, 48 (2010), pp. 1860–1871.
- [5] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
- [6] S. CHEN AND J.-L. PELLEQUER, *DeStripe: Frequency-based algorithm for removing stripe noises from AFM images*, BMC Struct. Biol., 11 (2011), 7.
- [7] P. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing*, in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer, New York, 2011, pp. 185–212.
- [8] I. EKELAND AND R. TEMAM, *Analyse convexe et problèmes variationnels*, Dunod, Paris, 1974.
- [9] J. FEHRENBACH, P. WEISS, AND C. LORENZO, *Variational algorithms to remove stationary noise. Application to microscopy imaging*, IEEE Trans. Image Process., 21 (2012), pp. 4420–4430.
- [10] J. FEHRENBACH, P. WEISS, AND C. LORENZO, *Variational algorithms to remove stripes: A generalization of the negative norm models*, in Proceedings of the 1st International Conference on Pattern Recognition Applications and Methods (ICPRAM), 2012.
- [11] A. V. FIACCO, *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*, Academic Press, New York, 1983.
- [12] G. GOLUB, M. HEATH, AND G. WAHBA, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223.
- [13] J. M. HENDRICKX AND A. OLSHEVSKY, *Matrix  $p$ -norms are NP-hard to approximate if  $p \neq 1, 2, \infty$* , SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2802–2812.
- [14] U. LEISCHNER, A. SCHIERLOH, W. ZIEGLGÄNSBERGER, AND H.-U. DODT, *Formalin-induced fluorescence reveals cell shape and morphology in biological tissue samples*, PloS One, 5 (2010), e10391.
- [15] S. MALLAT, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, 1999.
- [16] Y. MEYER, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures*, American Mathematical Society, Providence, RI, 2001.
- [17] V. MOROZOV, *On the solution of functional equations by the method of regularization*, Soviet Math. Dokl., 7 (1966), pp. 414–417.
- [18] B. MÜNCH, P. TRTIK, F. MARONE, AND M. STAMPANONI, *Stripe and ring artifact removal with combined wavelet-Fourier filtering*, Opt. Express, 17 (2009), pp. 8567–8591.
- [19] M. K. NG, P. WEISS, AND X. YUAN, *Solving constrained total-variation image restoration and reconstruction problems via alternating direction methods*, SIAM J. Sci. Comput., 32 (2010), pp. 2710–2736.
- [20] M. NIKOLOVA, *Model distortions in Bayesian MAP reconstruction*, Inverse Probl. Imaging, 1 (2007), pp. 399–422.
- [21] S. OSHER, A. SOLÉ, AND L. VESE, *Image decomposition and restoration using total variation minimization and the  $H^{-1}$  norm*, Multiscale Model. Simul., 1 (2003), pp. 349–370.
- [22] R. T. ROCKAFELLAR, *Convex Analysis*, Vol. 28, Princeton University Press, Princeton, NJ, 1996.
- [23] J. ROHN, *Computing the norm  $\|A\|_{\infty \rightarrow 1}$  norm is NP-hard*, Linear Multilinear Algebra, 47 (2000), pp. 195–204.
- [24] T. TEUBER, G. STEIDL, AND R. CHAN, *Minimization and parameter estimation for seminorm regularization models with  $I$ -divergence constraints*, Inverse Problems, 29 (2013), 035007.
- [25] S. VAITER, C. DELEDALLE, G. PEYRÉ, J. FADILI, AND C. DOSSAL, *Local behavior of sparse analysis regularization: Applications to risk estimation*, Appl. Comput. Harmon. Anal., 35 (2012), pp. 433–451.
- [26] L. VESE AND S. OSHER, *Modeling textures with total variation minimization and oscillating patterns in image processing*, J. Sci. Comput., 19 (2003), pp. 553–572.
- [27] Y. WEN AND R. CHAN, *Parameter selection for total variation based image restoration using discrepancy principle*, IEEE Trans. Image Process., 21 (2012), pp. 1770–1781.