

A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise

Yosra Marnissi, Yuling Zheng, Emilie Chouzenoux, Jean-Christophe Pesquet

▶ **To cite this version:**

Yosra Marnissi, Yuling Zheng, Emilie Chouzenoux, Jean-Christophe Pesquet. A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise. [Research Report] Laboratoire Informatique Gaspard Monge. 2016. hal-01418399

HAL Id: hal-01418399
https://hal.science/hal-01418399

Submitted on 16 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Public Domain

A Variational Bayesian Approach for Restoring Data Corrupted with Non-Gaussian Noise

Yosra Marnissi, Yuling Zheng, Emilie Chouzenoux and Jean-Christophe Pesquet *

October 24, 2016

Abstract

In this paper, a methodology is investigated for signal recovery in the presence of non-Gaussian noise. In contrast with regularized minimization approaches often adopted in the literature, in our algorithm the regularization parameter is reliably estimated from the observations. As the posterior density of the unknown parameters is analytically intractable, the estimation problem is derived in a variational Bayesian framework where the goal is to provide a good approximation to the posterior distribution in order to compute posterior mean estimates. Moreover, a majorization technique is employed to circumvent the difficulties raised by the intricate forms of the non-Gaussian likelihood and of the prior density. We demonstrate the potential of the proposed approach through comparisons with state-of-the-art techniques that are specifically tailored to signal recovery in the presence of mixed Poisson-Gaussian noise. Results show that the proposed approach is efficient and achieves performance comparable with other methods where the regularization parameter is manually tuned from an available ground truth.

Keywords: inverse problems, restoration, Variational Bayesian methods, parameter estimation, non-Gaussian noise, majorization, minimization.

1 Introduction

One of the most challenging tasks in signal processing is restoration, where one aims at providing an accurate estimate of the original signal from degraded observations. These degradations may arise due to various phenomena which are often unavoidable in practical situations. Undesirable blurring may be introduced by the atmosphere or may also stem from the intrinsic limitations of the acquisition system characterized by its point spread function. Furthermore, data can be perturbed by noise which can be viewed as parasite signal added to the information of interest, hence altering the extraction of this information. Noise may originate from various sources. On the one hand, sensors generally suffer from internal fluctuations referred to electrical or thermal noise. This type of noise is additive, independent of the signal of interest, and it can be modeled by a Gaussian distribution. On the other hand, it has been experimentally proven that in many situations, the signal of interest may suffer from noise with more complex

*Y. Marnissi and E. Chouzenoux are with University Paris-Est, LIGM, UMR CNRS 8049, 77454 Marne-la-Vallée, France (e-mail: first.last@u-pem.fr). Y. Zheng is with IBM Research China. Building 19, Zhongguancun Software Park, 8 Dongbeiwang Western Road, Haidian District, Beijing, 11 100193, CN, China (e-mail: ylzh@cn.ibm.com). J.-C. Pesquet is with Center for Visual Computing, CentraleSupelec, University Paris-Saclay, Grande Voie des Vignes, 92295 Châtenay-Malabry, France (e-mail: jean-christophe@pesquet.eu). This work was supported by the CNRS Imag'in project under grant 2015OPTIMISME, by the CNRS MASTODONS project under grant 2016TABASCO, and by the Agence Nationale de la Recherche under grant ANR-14-CE27-0001 GRAPHSIP.

characteristics. In fact, many devices lead to measurements distorted by heteroscedastic noise i.e. the characteristics of the noise depend on the characteristics of the unknown signal [1–7]. For example, to better reflect the physical properties of optical communication, the noise remains additive Gaussian but its variance is assumed to be dependent on the unknown signal [8]. Signals can also be corrupted with multiplicative noise [9–12] such as the speckle noise which commonly affects synthetic aperture radar (SAR), medical ultrasound and optical coherence tomography images [13], as well as with impulsive noise [14]. A mixture of Gaussian and impulsive noise has also been studied in [15,16]. Furthermore, in applications such as astronomy, medicine, and fluorescence microscopy where signals are acquired via photon counting devices, like CMOS and CCD cameras, the number of collected photons is related to some non-additive counting errors resulting in a shot noise [1–3, 17, 18]. The latter is non-additive, signal-dependent and it can be modeled by a Poisson distribution [19–28]. In this case, when the noise is assumed to be Poisson distributed, the implicit assumption is that Poisson noise dominates over all other noise kinds. Otherwise, the involved noise is a combination of Poisson and Gaussian (PG) components [29–41]. Most of the existing denoising methods only consider the noise as independent Gaussian, mainly because of the difficulties raised in handling other noise sources than the Gaussian one. In this paper, we focus on signal recovery beyond the standard additive independent Gaussian noise assumption.

1.1 State-of-the-art

Most of the existing strategies for such problems define the estimate as a minimizer of an appropriate cost function. The latter is composed of two terms: the so-called data fidelity term whose role is to make the solution consistent with the observation and the regularization term that incorporates prior information about the target signal so as to ensure the stability of the solution [42]. Several algorithms have been proposed to tackle the problem of restoration for signals corrupted with non-Gaussian noise by using minimization approaches. For example, in [43], a method is proposed to restore signals degraded by a linear operator and corrupted with an additive Gaussian noise having a signal-dependent variance. An early work in [44] and more recent developments in [29, 34–36, 39, 45, 46] have proposed to restore signals corrupted with mixed PG noise using different approximations of the PG data fidelity term. In all these approaches, the regularization parameter allows a tradeoff to be performed between fidelity to the observations and the prior information. Too small values of this parameter may lead to noisy estimates while too large values yield oversmoothed solutions. Consequently, the problem of setting a proper value of the regularization parameter should be addressed carefully and may depend on both the properties of the observations and the statistics of the target signal. When ground truth is available, one can choose the value of the regularization parameter that gives the minimal residual error evaluated through some suitable metric. However, in real applications where no ground truth is available, the problem of selecting the regularization parameter remains an open issue especially in situations where the images are acquired under poor conditions i.e. when the noise level is very high. Among existing approaches dealing with regularization parameter estimation, the works in [47–53] have to be mentioned. However, most of the mentioned methods were developed under the assumption of a Gaussian noise and their extension to the context of non-Gaussian noise is not easy. One can however mention the works in [31, 33] proposing efficient estimators in the context of denoising i.e. problems that do not involve linear degradation. Other approaches can be found in [54, 55] proposing efficient estimates in the specific case of a Poisson likelihood.

To address the shortcomings of these methods, one can adopt the Bayesian framework. In particular, Bayesian estimation methods based on Markov Chain Monte Carlo (MCMC)

sampling algorithms have been recently extended to inverse problems involving non-Gaussian noise [56–59]. However, despite good estimation performance that has been obtained, such methods remain computationally expensive for large scale problems. Another alternative approach is to rely on variational Bayesian approximation (VBA) [60–64]. Instead of simulating from the true posterior distribution, VBA approaches aim at approximating the intractable true posterior distribution with a tractable one from which the posterior mean can be easily computed. These methods can lead generally to a relatively low computational complexity when compared with sampling based algorithms.

1.2 Contributions

In this paper, we propose a VBA estimation approach for signals degraded by an arbitrary linear operator and corrupted with non-Gaussian noise. One of the main advantages of the proposed method is that it allows us to jointly estimate the original signal and the required regularization parameter from the observed data by providing good approximations of the Minimum Mean Square Estimator (MMSE) for the problem of interest. While using VBA, the main difficulty arising in the non-Gaussian case is that the involved likelihood and the prior density may have a complicated form and are not necessarily conjugate. To address this problem, a majorization technique is adopted providing a tractable VBA solution for non-conjugate distributions. Our approach allows us to employ a wide class of a priori distributions accounting for the possible sparsity of the target signal after some appropriate linear transformation. Moreover, it can be easily applied to several non Gaussian likelihoods that have been widely used [29, 39, 65, 66]. In particular, experiments in the case of images corrupted by PG noise showcase the good performance of our approach compared with methods using the discrepancy principle for estimating the regularization parameter [67]. Moreover, we propose variants of our method leading to a significant reduction of the computational cost while maintaining a satisfactory restoration quality.

1.3 Outline

This paper is organized as follows. In Section 2, we formulate the considered signal recovery problem in the Bayesian framework and we present a short overview on the variational Bayesian principle. In Section 3, we present our proposed estimation method based on VBA. In Section 5, we provide simulation results together with comparisons with state-of-the-art methods in terms of image restoration performance and computation time. Finally, some conclusions are drawn in Section 6.

2 Problem statement

2.1 Bayesian formulation

In this paper, we consider a wide range of applications where the degradation model can be formulated as an inverse possibly ill-posed problem as follows:

$$(\forall i \in \{1, \ldots, M\}), \quad y_i = \mathcal{D}([\mathbf{H}\overline{\mathbf{x}}]_i), \tag{1}$$

where $\overline{\mathbf{x}} \in \mathbb{R}^N$ is the signal of interest, $\mathbf{H} \in \mathbb{R}^{M \times N}$ is the linear operator typically modeling a blur, a projection, or a combination of both degradations, $[\mathbf{H}\overline{\mathbf{x}}]_i$ denotes the i-th component of $\mathbf{H}\overline{\mathbf{x}}$, $\mathbf{y} = (y_i)_{1 \leqslant i \leqslant M} \in \mathbb{R}^M$ is the measured data and $\mathcal{D}$ is the noise model that may depend

on the target data. The objective is to find an estimator $\hat{\mathbf{x}}$ of $\overline{\mathbf{x}}$ from $\mathbf{H}$ and $\mathbf{y}$. The neg-log-likelihood ϕ of the observations reads

$$(\forall \mathbf{x} \in \mathbb{R}^N) \quad \phi(\mathbf{x}; \mathbf{y}) = -\ln \mathsf{p}(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^{M} \phi_i([\mathbf{H}\mathbf{x}]_i \,; y_i). \tag{2}$$

Depending on the noise statistical model $\mathcal{D}$, ϕ_i may take various forms [3,14,16,31]. In particular, it reduces to a least squares function for additive Gaussian noise. In the Bayesian framework, we apply regularization by assigning a prior distribution to the data $\mathbf{x}$ to be recovered. In this paper, we adopt the following flexible expression of the prior density of $\overline{\mathbf{x}}$:

$$\mathsf{p}(\mathbf{x} \mid \gamma) = \tau \gamma^{\frac{N}{2\kappa}} \exp\Big(-\gamma \sum_{j=1}^{J} \|\mathbf{D}_j \mathbf{x}\|^{2\kappa} \Big), \tag{3}$$

where κ is a constant in $(0, 1]$, $\|\cdot\|$ denotes the ℓ_2-norm and $(\mathbf{D}_j)_{1 \leqslant j \leqslant J} \in (\mathbb{R}^{S \times N})^J$ where $\mathbf{D} = [\mathbf{D}_1^\top, \ldots, \mathbf{D}_J^\top]^\top$ is an operator that may be a matrix computing the discrete difference between neighboring pixels or a frame-analysis operator [68]. The constant $\gamma \in (0, +\infty)$ can be viewed as a regularization parameter that plays a prominent role in the restoration process and $\tau \in (0, +\infty)$ is a normalization constant independent of γ. Note that the form of the partition function for such a prior distribution follows from the fact that the associated potential is 2κ-homogeneous [69].

In this paper, the noise shape parameter κ is chosen to be fixed through empirical methods and we aim at estimating parameter γ together with $\mathbf{x}$. To this end, we choose a Gamma prior for γ, i.e. $\mathsf{p}(\gamma) \propto \gamma^{\alpha-1} \exp(-\beta\gamma)$ where α and β are positive constants (set in practice to small values to ensure a weakly informative prior).

Using the Bayes' rule, we can obtain the posterior distribution of the set of unknown variables $\boldsymbol{\Theta} = (\mathbf{x}, \gamma)$ given the vector of observations $\mathbf{y}$:

$$\mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y}) \propto \mathsf{p}(\mathbf{y} \mid \mathbf{x}) \mathsf{p}(\mathbf{x} \mid \gamma) \mathsf{p}(\gamma). \tag{4}$$

However, this distribution has an intricate form. In particular, its normalization constant does not have a closed form expression. To cope with this problem, we resort to the variational Bayesian framework. The rationale of this work is to find a simple approximation to the true posterior distortion, leading to a tractable computation of the posterior mean estimate.

2.2 Variational Bayes principle

The variational Bayes approach has been first introduced in physics [70]. The idea behind it is to approximate the posterior distribution $\mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y})$ with another distribution denoted by $q(\boldsymbol{\Theta})$ which is as close as possible to $\mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y})$, by minimizing the Kullback-Leibler divergence between them [60, 68, 71]:

$$q^{\text{opt}} = \underset{q}{\operatorname{argmin}} \ \mathcal{KL}\big(q(\boldsymbol{\Theta}) \| \mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y})\big), \tag{5}$$

where

$$\mathcal{KL}\big(q(\boldsymbol{\Theta}) \| \mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y})\big) = \int q(\boldsymbol{\Theta}) \ln \frac{q(\boldsymbol{\Theta})}{\mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y})} \mathrm{d}\boldsymbol{\Theta}. \tag{6}$$

This minimization becomes tractable if a suitable factorization structure of $q(\boldsymbol{\Theta})$ is assumed. In particular, we assume that $q(\boldsymbol{\Theta}) = \prod_{r=1}^{R} q(\Theta_r)$. Hence, the optimal density approximation $q^{\text{opt}}(\Theta_r)$ for each variable Θ_r, is obtained by minimizing the $\mathcal{KL}$ divergence while holding the remaining densities for the rest of variables fixed. In this case, there exists an optimal solution

to the optimization problem (5) for each density $q(\Theta_r)_{1\leqslant r\leqslant R}$, given by the exponential of the expectation of the joint density with respect to the distribution of all the unknown parameters except the one of interest i.e.

$$(\forall r \in \{1,\ldots,R\}) \quad q^{\text{opt}}(\Theta_r) \propto \exp\left(\langle \ln \mathsf{p}(\mathbf{y},\Theta)\rangle_{\prod_{i\neq r} q^{\text{opt}}(\Theta_i)}\right) \tag{7}$$

where $\langle\cdot\rangle_{\prod_{i\neq r} q(\Theta_i)} = \int \cdot \prod_{i\neq r} q(\Theta_i)\mathrm{d}\Theta_i$. Due to the implicit relations existing between $\left(q^{\text{opt}}(\Theta_r)\right)_{1\leqslant r\leqslant R}$, an analytical expression of $q^{\text{opt}}(\Theta)$ generally does not exist. Usually, these distributions are determined in an iterative way, by updating one of the separable components $(q(\Theta_r))_{1\leqslant r\leqslant R}$ while fixing the others [60]. Applications of classical VBA approaches can be found in [62–64, 72, 73] while improved VBA algorithms have been proposed in [61, 74]. Once the approximate distributions are computed, the unknown parameters are then estimated by the means of the obtained distributions.

3 Proposed approach

Name	$\phi_i(v;y_i)$	$\phi_i'(v;y_i)$	$\beta_i(y_i)$	Domain of validity	Noise model
Gaussian	$\frac{1}{2\sigma^2}(v-y_i)^2$	$\frac{1}{\sigma^2}(v-y_i)$	$\frac{1}{\sigma^2}$	$y_i\in\mathbb{R}$	Gaussian
Cauchy	$\ln\left(1+\frac{(v-y_i)^2}{\sigma^2}\right)$	$\frac{2(v-y_i)}{\sigma^2+(v-y_i)^2}$	$\frac{2}{\sigma^2}$	$y_i\in\mathbb{R}$	Cauchy
Anscombe Transform	$2\left(\sqrt{y_i+\frac{3}{8}}-\sqrt{v+\frac{3}{8}}\right)^2$	$2-\frac{2\sqrt{y_i+\frac{3}{8}}}{\sqrt{v+\frac{3}{8}}}$	$\left(\frac{3}{8}\right)^{-3/2}\sqrt{y_i+\frac{3}{8}}$	$y_i\geqslant -\frac{3}{8}$	Poisson
Generalized Anscombe Transform	$2\left(\sqrt{y_i+\sigma^2+\frac{3}{8}}-\sqrt{v+\sigma^2+\frac{3}{8}}\right)^2$	$2-\frac{2\sqrt{y_i+\frac{3}{8}+\sigma^2}}{\sqrt{v+\frac{3}{8}+\sigma^2}}$	$(\frac{3}{8}+\sigma^2)^{-3/2}\sqrt{y_i+\frac{3}{8}+\sigma^2}$	$y_i\geqslant -\frac{3}{8}-\sigma^2$	Poisson-Gaussian
Shifted Poisson	$(v+\sigma^2)-(y_i+\sigma^2)\ln(v+\sigma^2)$	$1-\frac{y_i+\sigma^2}{v+\sigma^2}$	$\frac{y_i+\sigma^2}{\sigma^4}$	$y_i\geqslant -\sigma^2$	Poisson-Gaussian
Weighted least squares	$\frac{(y_i-v)^2}{2(\sigma^2+v)}+\frac{1}{2}\ln(\sigma^2+v)$	$\frac{1}{2}-\frac{(y_i+\sigma^2)^2}{2(v+\sigma^2)^2}+\frac{1}{2(\sigma^2+v)}$	$\max\left\{\frac{(y_i+\sigma^2)^2}{\sigma^6}-\frac{1}{2\sigma^4},\frac{1}{54(y_i+\sigma^2)^4}\right\}$	$y_i\in\mathbb{R}\backslash\{-\sigma^2\}$	Poisson-Gaussian

Table 1: Examples of differentiable likelihoods satisfying Assumption 3.1. ϕ_i' denotes the first derivative of function ϕ_i and $\beta_i(y_i)$ is the Lipschitz constant of ϕ_i' (for functions in lines 3-6, we assume that ϕ_i is replaced on $\mathbb{R}_-$ by its quadratic extension (9).) The expression for the Lipschitz constant of the gradient of the weighted least squares likelihood was established in [75, Chap. IV].

In this work, we assume the following separable form for q:

$$q(\boldsymbol{\Theta}) = q(\mathbf{x})q(\gamma). \tag{8}$$

Unfortunately, by using directly (7), we cannot obtain an explicit expression of $q(\mathbf{x})$ due to the intricate form of both the prior distribution and the likelihood when the statistics of the noise are no longer Gaussian. In this paper, we propose to use deterministic methods to construct quadratic upper bounds for the negative logarithms of both the likelihood and the prior density [76]. This allows us to derive an upper bound of the desired cost function in (5) as will be described in the following.

3.1 Construction of the majorizing approximation

3.1.1 Likelihood

One popular approach in signal recovery is the half-quadratic formulation [77]. Under some mild assumptions and by introducing some auxiliary variables, a complicated criterion can be written

as the infimum of a surrogate half-quadratic function i.e. the latter is quadratic with respect to the original variables and the auxiliary variables appear decoupled. This half-quadratic criterion can be then efficiently minimized using classical optimization algorithms. This formulation has been widely used in energy-minimization approaches [78–80] where the initial optimization problem is replaced by the minimization of the constructed surrogate function. Furthermore, this technique has been recently extended to sampling algorithms [81]. The initial intractable posterior distribution to sample from is replaced by the conditional distribution of the target signal given the auxiliary variables. The obtained distribution has been shown to be much simpler to explore by using standard sampling algorithms. In this paper, we propose to use half-quadratic approaches to construct an upper bound for the objective function in (5).

We assume that the likelihood satisfies the following property:

Assumption 3.1. *For every $i \in \{1, \ldots, M\}$, ϕ_i is differentiable on $\mathbb{R}$ and there exists $\mu_i(y_i) > 0$ such that the function defined by $v \mapsto \frac{v^2}{2} - \frac{\phi_i(v;y_i)}{\mu_i(y_i)}$ is convex on $\mathbb{R}$.*

In particular, this assumption is satisfied when, for every $i \in \{1, \ldots, M\}$, ϕ_i is $\beta_i(y_i)$-Lipschitz differentiable on $\mathbb{R}$ for every $\mu_i(y_i) \geqslant \beta_i(y_i)$. Table 1 shows some examples of useful functions satisfying the desired property (up to an additive constant). Note that, since the functions in lines 3-6 of Table 1 are $\beta_i(y_i)$-Lipschitz differentiable only on $\mathbb{R}_+$, we propose to use on $\mathbb{R}_-$ a quadratic extension of them defined as follows:

$$(\forall v \in \mathbb{R}_-) \quad \phi_i(v; y_i) = \phi_i(0; y_i) + \phi_i^{'}(0; y_i)v + \frac{1}{2}\phi_i^{''}(0; y_i)v^2, \tag{9}$$

so that the extended version of $\phi_i(.; y_i)$ is now differentiable on $\mathbb{R}$ with $\beta_i(y_i)$-Lipschitzian gradient.

For every $i \in \{1, \ldots, M\}$ and $v \in \mathbb{R}$, let us define the following function:

$$(\forall v \in \mathbb{R}) \quad \varsigma_i(v; y_i) = \sup_{t \in \mathbb{R}} \left(-\frac{1}{2}(v - t)^2 + \frac{\phi_i(t; y_i)}{\mu_i(y_i)} \right). \tag{10}$$

Then, the following property holds:

Proposition 3.1. *For every $i \in \{1, \ldots, M\}$,*

$$(\forall v \in \mathbb{R}) \quad \phi_i(v; y_i) = \inf_{w_i \in \mathbb{R}} \quad T_i(v, w_i; y_i). \tag{11}$$

where, for every $v \in \mathbb{R}$,

$$T_i(v, w_i; y_i) = \mu_i(y_i) \left(\frac{1}{2}(v - w_i)^2 + \varsigma_i(w_i; y_i) \right). \tag{12}$$

Moreover, the unique minimizer of $w_i \mapsto T_i(v, w_i; y_i)$ reads

$$\widehat{w}_i(v) = v - \frac{1}{\mu_i(y_i)} \phi_i'(v; y_i). \tag{13}$$

Proof. See Appendix 6. □

It follows from this result that

$$\left(\forall \mathbf{x} \in \mathbb{R}^N\right) \quad \phi(\mathbf{x}; \mathbf{y}) = \inf_{\mathbf{w} \in \mathbb{R}^M} T(\mathbf{x}, \mathbf{w}; \mathbf{y}), \tag{14}$$

where $T(\mathbf{x}, \mathbf{w}; \mathbf{y}) = \sum_{i=1}^{M} T_i([\mathbf{H}\mathbf{x}]_i, w_i; y_i)$.

Note that (11) shows that, for every $i \in \{1, \ldots, M\}$, $\phi_i(\cdot; y_i)$ is a so-called Moreau envelope of the function $\varsigma_i(\cdot; y_i)$. A more direct proof of Proposition 3.1 can thus be derived from the properties of the proximity operator [82] when the functions $(\phi_i)_{1 \leq i \leq M}$ are convex. The proof we provide in the appendix however does not make such a restrictive assumption.

3.1.2 Prior

Similarly, we construct a surrogate function for the prior distribution. More precisely, we follow the same idea as in [63] and we use the following convexity inequality to derive a majorant for the ℓ_κ-norm :

$$(\forall \nu \geq 0)(\forall \upsilon > 0) \quad \upsilon^\kappa \leq (1-\kappa)\nu^\kappa + \kappa \nu^{\kappa-1}\upsilon.$$

Hence, we obtain the following majorant function for the negative logarithm of the prior distribution:

$$\gamma \sum_{j=1}^{J} \|\mathbf{D}_j \mathbf{x}\|^{2\kappa} \leqslant \gamma \sum_{j=1}^{J} \frac{\kappa \|\mathbf{D}_j \mathbf{x}\|^2 + (1-\kappa)\lambda_j}{\lambda_j^{1-\kappa}}. \tag{15}$$

where $(\lambda_j)_{1 \leqslant j \leqslant J}$ are positive variables. In the following, we will denote by $Q(\mathbf{x}, \boldsymbol{\lambda}; \gamma) = \sum_{j=1}^{J} Q_j(\mathbf{D}_j \mathbf{x}, \lambda_j; \gamma)$, the function in the right-hand side of the above inequality where, for every $j \in \{1, \ldots, J\}$,

$$Q_j(\mathbf{D}_j \mathbf{x}, \lambda_j; \gamma) = \gamma \frac{\kappa \|\mathbf{D}_j \mathbf{x}\|^2 + (1-\kappa)\lambda_j}{\lambda_j^{1-\kappa}}. \tag{16}$$

3.1.3 Proposed majorant

Thus, we can derive the following lower bound for the posterior distribution:

$$\mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y}) \geqslant L(\boldsymbol{\Theta} | \mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}), \tag{17}$$

where function L is defined as

$$L(\boldsymbol{\Theta} | \mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}) = C(\mathbf{y}) \exp\left[-T(\mathbf{x}, \mathbf{w}; \mathbf{y}) - Q(\mathbf{x}, \boldsymbol{\lambda}; \gamma)\right] \mathsf{p}(\gamma)$$

with $C(\mathbf{y}) = \mathsf{p}(\mathbf{y})^{-1}(2\pi)^{-M/2}\tau\gamma^{\frac{N}{2\kappa}}$. The minorization of the distribution leads to an upper bound for the $\mathcal{KL}$ divergence:

$$\mathcal{KL}(q(\boldsymbol{\Theta}) \| \mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y})) \leqslant \mathcal{KL}(q(\boldsymbol{\Theta}) \| L(\boldsymbol{\Theta} | \mathbf{y}; \mathbf{w}, \boldsymbol{\lambda})). \tag{18}$$

Note that, although the constructed lower bound in (17) is tangent to the posterior distribution i.e.

$$\mathsf{p}(\boldsymbol{\Theta} \mid \mathbf{y}) = \sup_{\mathbf{w} \in \mathbb{R}^M, \boldsymbol{\lambda} \in \mathbb{R}^J} L(\boldsymbol{\Theta} | \mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}),$$

the tangency property may not be generally satisfied in (18). Thus, the tightness of the constructed majorant of the $\mathcal{KL}$ divergence may have a significant impact on the accuracy of the method. By minimizing the constructed bound (18) with respect to $\mathbf{w}$ and $\boldsymbol{\lambda}$, we make this bound as tight as possible. Note that, for every $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, J\}$, $\lambda_j \mapsto \mathcal{KL}(q(\boldsymbol{\Theta}) \| L(\boldsymbol{\Theta} | \mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}))$ and $w_i \mapsto \mathcal{KL}(q(\boldsymbol{\Theta}) \| L(\boldsymbol{\Theta} | \mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}))$ can be minimized separately. Hence, Problem (5) can be solved by the following four-step alternating optimization scheme:

- Minimizing the upper bound in (18) w.r.t. $q(\mathbf{x})$;
- Updating the auxiliary variables w_i in order to minimize $\mathcal{KL}(q(\boldsymbol{\Theta})\|L(\boldsymbol{\Theta}|\mathbf{y};\mathbf{w},\boldsymbol{\lambda}))$, for every $i \in \{1,\ldots,M\}$;
- Updating the auxiliary variable λ_j in order to minimize $\mathcal{KL}(q(\boldsymbol{\Theta})\|L(\boldsymbol{\Theta}|\mathbf{y};\mathbf{w},\boldsymbol{\lambda}))$, for every $j \in \{1,\ldots,J\}$;
- Mimimizing the upper bound in (18) w.r.t. $q(\gamma)$.

The main benefit of this majorization strategy is to guarantee that the optimal approximate posterior distribution for $\mathbf{x}$ belongs to the Gaussian family and the optimal approximate posterior distribution for γ belongs to the Gamma one, i.e.

$$q(\mathbf{x}) \equiv \mathcal{N}(\mathbf{m},\boldsymbol{\Sigma}), \quad q(\gamma) \equiv \mathcal{G}(a,b).$$

Therefore, the distribution updates can be performed by updating their parameters, namely $\mathbf{m}$, $\boldsymbol{\Sigma}$, a, and b.

4 Iterative algorithm

Subsequently, at a given iteration k of the proposed algorithm, the corresponding estimated variables will be indexed by k.

4.1 Updating $q(\mathbf{x})$

Because of the majorization step, we need to minimize the upper bound on the $\mathcal{KL}$ divergence. The standard solution (7) can still be used by replacing the joint distribution by a lower bound $L(\boldsymbol{\Theta},\mathbf{y};\mathbf{w},\boldsymbol{\lambda})$ chosen proportional to $L(\boldsymbol{\Theta}|\mathbf{y};\mathbf{w},\boldsymbol{\lambda})$:

$$\begin{aligned} q^{k+1}(\mathbf{x}) &\propto \exp\left(\left\langle \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^k,\boldsymbol{\lambda}^k)\right\rangle_{q^k(\gamma)}\right) \\ &\propto \exp\left(\int \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^k,\boldsymbol{\lambda}^k) q^k(\gamma)\mathrm{d}\gamma\right) \\ &\propto \exp\left(-\sum_{i=1}^{M}\frac{1}{2}\mu_i(y_i)\left([\mathbf{H}\mathbf{x}]_i - w_i^k\right)^2 - \frac{a_k}{b_k}\sum_{j=1}^{J}\frac{\kappa\|\mathbf{D}_j\mathbf{x}\|^2 + (1-\kappa)\lambda_j^k}{(\lambda_j^k)^{1-\kappa}}\right). \end{aligned} \tag{19}$$

The above distribution can be identified as a multivariate Gaussian distribution whose covariance matrix and mean parameter are given by

$$\boldsymbol{\Sigma}_{k+1}^{-1} = \mathbf{H}^\top \mathrm{Diag}(\boldsymbol{\mu}(\mathbf{y}))\mathbf{H} + 2\frac{a_k}{b_k}\mathbf{D}^\top\boldsymbol{\Lambda}^k\mathbf{D}, \tag{20}$$

$$\mathbf{m}_{k+1} = \boldsymbol{\Sigma}_{k+1}\mathbf{H}^\top\mathbf{u}, \tag{21}$$

where $\boldsymbol{\mu}(\mathbf{y}) = [\mu_1(y_1),\ldots,\mu_M(y_M)]^\top$, $\mathbf{u}$ is a $M \times 1$ vector whose i-th component is given by $u_i = \mu_i(y_i)w_i^k$ and $\boldsymbol{\Lambda}$ is the diagonal matrix whose diagonal elements are $\left(\kappa(\lambda_j^k)^{\kappa-1}\mathbf{I}_S\right)_{1\leq j\leq J}$.

4.2 Updating w

The auxiliary variable $\mathbf{w}$ is determined by minimizing the upper bound of $\mathcal{KL}$ divergence with respect to this variable:

$$\begin{aligned}
\mathbf{w}^{k+1} &= \underset{\mathbf{w}}{\operatorname{argmin}} \int q^{k+1}(\mathbf{x}) q^k(\gamma) \ln \frac{q^{k+1}(\mathbf{x}) q^k(\gamma)}{L(\boldsymbol{\Theta}|\mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}^k)} \mathrm{d}\mathbf{x}\mathrm{d}\gamma \\
&= \underset{\mathbf{w}}{\operatorname{argmin}} \int q^{k+1}(\mathbf{x}) q^k(\gamma) \left(- \ln L(\boldsymbol{\Theta}|\mathbf{y}; \mathbf{w}, \boldsymbol{\lambda}^k) \right) \mathrm{d}\mathbf{x}\mathrm{d}\gamma \\
&= \underset{\mathbf{w}}{\operatorname{argmin}} \int q^{k+1}(\mathbf{x}) \sum_{i=1}^{M} T_i([\mathbf{H}\mathbf{x}]_i, w_i; y_i) \mathrm{d}\mathbf{x} \qquad (22) \\
&= \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{M} T_i([\mathbf{H}\mathbf{m}_{k+1}]_i, w_i; y_i), \qquad (23)
\end{aligned}$$

where the equality in (22) follows from the expression in (12). Interestingly, it follows from Property 3.1 that

$$\begin{aligned}
w_i^{k+1} &= \underset{w_i}{\operatorname{argmin}} \; T_i([\mathbf{H}\mathbf{m}_{k+1}]_i, w_i; y_i) \\
&= [\mathbf{H}\mathbf{m}_{k+1}]_i - \frac{1}{\mu_i(y_i)} \phi_i^{'}([\mathbf{H}\mathbf{m}_{k+1}]_i; y_i). \qquad (24)
\end{aligned}$$

4.3 Updating $\boldsymbol{\lambda}$

The variable $\boldsymbol{\lambda}$ is determined in a similar way: for every $j \in \{1, \ldots, J\}$,

$$\begin{aligned}
\lambda_j^{k+1} &= \underset{\lambda_j \in [0, +\infty)}{\operatorname{argmin}} \; \mathcal{KL}(q^{k+1}(\mathbf{x}) q^k(\gamma) \| L(\boldsymbol{\Theta}|\mathbf{y}; \mathbf{w}^{k+1}, \boldsymbol{\lambda})) \\
&= \underset{\lambda_j \in [0, +\infty)}{\operatorname{argmin}} \; \sum_{i=1}^{Q} \int q^{k+1}(\mathbf{x}) q^k(\gamma) Q_i(\mathbf{D}_i\mathbf{x}, \lambda_i; \gamma) \mathrm{d}\mathbf{x}\mathrm{d}\gamma \\
&= \underset{\lambda_j \in [0, +\infty)}{\operatorname{argmin}} \; \int q^{k+1}(\mathbf{x}) q^k(\gamma) Q_j(\mathbf{D}_j\mathbf{x}, \lambda_j; \gamma) \mathrm{d}\mathbf{x}\mathrm{d}\gamma \\
&= \underset{\lambda_j \in [0, +\infty)}{\operatorname{argmin}} \; \int q^{k+1}(\mathbf{x}) q^k(\gamma) \gamma \frac{\kappa \|\mathbf{D}_j\mathbf{x}\|^2 + (1-\kappa)\lambda_j}{\lambda_j^{1-\kappa}} \mathrm{d}\mathbf{x}\mathrm{d}\gamma \\
&= \underset{\lambda_j \in [0, +\infty)}{\operatorname{argmin}} \; \frac{\kappa \, \mathbb{E}_{q^{k+1}(\mathbf{x})} \left[\|\mathbf{D}_j\mathbf{x}\|^2 \right] + (1-\kappa)\lambda_j}{\lambda_j^{1-\kappa}}. \qquad (25)
\end{aligned}$$

The minimum is attained at

$$\begin{aligned}
\lambda_j^{k+1} &= \mathbb{E}_{q^{k+1}(\mathbf{x})} \left[\|\mathbf{D}_j\mathbf{x}\|^2 \right] \\
&= \|\mathbf{D}_j\mathbf{m}_{k+1}\|^2 + \operatorname{trace} \left[\mathbf{D}_j^\top \mathbf{D}_j \boldsymbol{\Sigma}_{k+1} \right]. \qquad (26)
\end{aligned}$$

4.4 Updating $q(\gamma)$

Using (7) where the joint distribution is replaced by its lower bound function, we obtain

$$\begin{aligned} q^{k+1}(\gamma) \propto & \exp\left(\left\langle \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^{k+1},\boldsymbol{\lambda}^{k+1})\right\rangle_{q^{k+1}(\mathbf{x})}\right) \\ \propto & \exp\left(\int \ln L(\mathbf{x},\gamma,\mathbf{y};\mathbf{w}^{k+1},\boldsymbol{\lambda}^{k+1}) q^{k+1}(\mathbf{x}) d\mathbf{x}\right) \\ \propto & \gamma^{\frac{N}{2\kappa}+\alpha-1}\exp(-\beta\gamma)\exp\left(-\gamma\sum_{j=1}^{J}\frac{\kappa\mathbb{E}_{q^{k+1}(\mathbf{x})}\left[\|\mathbf{D}_j\mathbf{x}\|^2\right]+(1-\kappa)\lambda_j^{k+1}}{(\lambda_j^{k+1})^{1-\kappa}}\right) \\ \equiv & \mathcal{G}(a_{k+1},b_{k+1}). \end{aligned} \tag{27}$$

Using (26), one can recognize that the above distribution is a Gamma one with parameters

$$a_{k+1} = \frac{N}{2\kappa}+\alpha = a, \qquad b_{k+1} = \sum_{j=1}^{J}(\lambda_j^{k+1})^{\kappa}+\beta. \tag{28}$$

4.5 Resulting algorithm

The proposed method is outlined in Algorithm 2. It alternates between the update of the auxiliary variables and the distribution of the unknown parameters.

Algorithm 1 VBA approach for recovery of signals corrupted with non-Gaussian noise.

1. Set initial values: $\mathbf{w}^0, \boldsymbol{\lambda}^0, b_0$. Compute a with (28).
2. For $k = 0, 1, \ldots$
   (a) Update parameters $\boldsymbol{\Sigma}_{k+1}$ and $\mathbf{m}_{k+1}$ of $q^{k+1}(\mathbf{x})$ using (20) and (21).
   (b) Update $\mathbf{w}^{k+1}$ using (24).
   (c) Update $\boldsymbol{\lambda}^{k+1}$ using (26).
   (d) Update parameter b_{k+1} of $q^{k+1}(\gamma)$ using (28).

4.6 Implementation issues

An additional difficulty arising in the implementation of Algorithm 2 is that the determination of $\boldsymbol{\Sigma}_{k+1}$ requires inverting the matrix given by (20), which is computationally expensive in high dimension. To bypass this operation, we propose to compare two approaches. The first one follows the idea in [62]: we make use of the linear conjugate gradient method to approximate $\mathbf{m}_{k+1}$ iteratively and in (26), where an explicit form of $\boldsymbol{\Sigma}_{k+1}$ cannot be sidestepped, this matrix is approximated by a diagonal one whose diagonal entries are equal to the inverse of the diagonal elements of $\boldsymbol{\Sigma}_{k+1}^{-1}$. The second technique uses Monte-Carlo sample averaging to approximate $\mathbf{m}_{k+1}$ and λ_j^{k+1}: specifically, we generate samples $(\mathbf{n}_s)_{1\leq s\leq N_s}$ from Gaussian distribution with mean $\mathbf{m}_{k+1}$ and covariance matrix $\boldsymbol{\Sigma}_{k+1}$ using [83], as summarized in Algorithm 2. This estimator has two desirable properties. First, its accuracy is independent of the problem size, its relative error only depends on the number of samples and it decays as $\sqrt{2/N_s}$ (only $N_s = 2/\rho^2$ samples are required to reach a desired relative error ρ) [84]. Second, for the simulation of N_s independent Gaussian samples, one can take advantage of a multiprocessor architecture by resorting to parallel implementation allowing us to reduce the computation time.

Algorithm 2 Stochastic approach for computing the parameters of $q(\mathbf{x})$.

1. For $s = 1, 2, \ldots, N_s$
   (a) **Perturbation** : Generate

$$\boldsymbol{\nu}_s \sim \mathcal{N}\left(\mathbf{u}, \left(\mathrm{Diag}\big(\boldsymbol{\mu}(\mathbf{y})\big)\right)^{1/2}\right)$$
$$\boldsymbol{\eta}_s \sim \mathcal{N}\left(0, \sqrt{2\gamma^k}\boldsymbol{\Lambda}_k^{1/2}\right)$$

   with $\gamma^k = a_k/b_k$.
   (b) **Optimization**: Compute $\mathbf{n}_s$ as the minimizer of $\mathcal{J}(\mathbf{v}) = \|\boldsymbol{\nu}_s - \mathrm{Diag}\big(\boldsymbol{\mu}(\mathbf{y})\big)\mathbf{H}\mathbf{v}\|^2_{(\mathrm{Diag}(\boldsymbol{\mu}(\mathbf{y})))^{-1}} + \frac{1}{2\gamma^k}\|\boldsymbol{\eta}_s - 2\gamma^k\boldsymbol{\Lambda}_k\mathbf{D}\mathbf{v}\|^2_{\boldsymbol{\Lambda}_k^{-1}}$, which is equivalent to minimize $\tilde{\mathcal{J}}(\mathbf{v}) = \mathbf{v}^\top\boldsymbol{\Sigma}_{k+1}^{-1}\mathbf{v} - 2\mathbf{v}^\top\mathbf{z}_s$ where $\mathbf{z}_s = \mathbf{H}^\top\boldsymbol{\nu}_s + \mathbf{D}^\top\boldsymbol{\eta}_s$. The minimizer is computed using the conjugate gradient algorithm.
2. Update

$$\mathbf{m}_{k+1} = \frac{1}{N_s}\sum_{s=1}^{N_s}\mathbf{n}_s$$
$$(\forall j \in \{1, \ldots, J\}) \quad \lambda_j^{k+1} = \frac{1}{N_s}\sum_{s=1}^{N_s}\|\mathbf{D}_j\mathbf{n}_s\|^2.$$

5 Application to Poisson-Gaussian image restoration

5.1 Problem formulation

In this section, we propose to illustrate the usefulness of our algorithm via experiments on PG noise models in the context of image restoration. More specifically, the vector of observations $\mathbf{y} = (y_i)_{1\leq i\leq M} \in \mathbb{R}^M$ is related to the original image $\overline{\mathbf{x}}$ through

$$\mathbf{y} = \mathbf{z} + \mathbf{b}, \tag{29}$$

where $\mathbf{z}$ and $\mathbf{b}$ are assumed to be mutually independent random vectors and

$$\mathbf{z} \mid \overline{\mathbf{x}} \sim \mathcal{P}(\mathbf{H}\overline{\mathbf{x}}), \quad \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_M),$$

$\mathcal{P}$ denoting the independent Poisson distribution. Hence, we can deduce the following expression of the likelihood function [39]:

$$\mathsf{p}(\mathbf{y} \mid \mathbf{x}) = \prod_{i=1}^{M}\left(\sum_{n=1}^{+\infty}\frac{e^{-[\mathbf{H}\mathbf{x}]_i}\left([\mathbf{H}\mathbf{x}]_i\right)^n}{n!}\frac{e^{-\frac{1}{2\sigma^2}(y_i-n)^2}}{\sqrt{2\pi\sigma^2}}\right). \tag{30}$$

The exact expression of the PG likelihood (30) involves an infinite sum which makes its exact computation impossible in practice. In [39], the infinite sum was replaced by a finite summation with bounds depending on the current estimate of $\bar{\mathbf{x}}$. However, this method implies a higher computational burden in the reconstruction process when compared with other likelihoods proposed in the literature as accurate approximations of (30). In [72], VBA inference techniques have been successfully applied to the restoration of data corrupted with PG noise using the generalized Anscombe transform (GAST) likelihood [31, 32, 65, 66]. In this paper, we will employ

the GAST likelihood , as well as the shifted Poisson likelihood (SPoiss) [6] and the weighted least squares approximation (WL2) [34, 43, 45] (see Table 1).

5.2 Numerical results

We evaluate the performance of the proposed approach for the restoration of images degraded by both blur and PG noise. We consider six test images, displayed in Figure 1, whose intensities have been rescaled so that pixel values belong to a chosen interval $[0, x^+]$. Images $\overline{\mathbf{x}}_1$ and $\overline{\mathbf{x}}_6$ are HST astronomical images while images $\overline{\mathbf{x}}_2$, $\overline{\mathbf{x}}_3$, $\overline{\mathbf{x}}_4$ and $\overline{\mathbf{x}}_5$ correspond to the set of confocal microscopy images considered in [39]. These images are then artificially degraded by an operator $\mathbf{H}$ modeling spatially invariant blur with point spread function h and by PG noise with variance σ^2. In our experiments, we choose a total variation prior, i.e. $\kappa = 1/2$ and for every pixel $j \in \{1, \ldots, N\}$, $\mathbf{D}_j\mathbf{x} = \left[[\nabla^h\mathbf{x}]_j, [\nabla^v\mathbf{x}]_j\right]^\top \in \mathbb{R}^2$ where ∇^h and ∇^v are the discrete gradients computed in the horizontal and vertical directions. As a result, $J = N$. It can be noted that we can apply the proposed algorithm to all approximations of the PG fidelity term presented in Table 1. For the GAST and the SPoiss approximations, we use $\mu_i(y_i) \equiv \max\{\beta_i(y_i), \varepsilon\}$ where $\varepsilon > 0$. For the WL2 approximation, we use $\mu_i(y_i) = \max\left\{(y_i + \sigma^2)^2/\sigma^6, \varepsilon\right\}$, so as to satisfy Assumption 3.1. Note that a data truncation is performed as a pre-processing step on the observed image $\mathbf{y}$ in order to satisfy the domain condition given in the fifth column of Table 1.

The goal of our experiments is twofold. First, for each likelihood, we compare the accuracy of the two proposed approximations of the covariance matrix described in Section 4.6 namely the diagonal approximation (denoted as approximation 1) and the Monte Carlo averaging strategy (denoted as approximation 2) with different number of samples N_s, namely $N_s = 160$ or 640. Second, the proposed method is compared with state-of-the-art algorithms that compute the MAP estimate. More specifically, as GAST and SPoiss likelihoods are convex and Lipschitz differentiable, we use the method presented in [39] where a primal-dual splitting algorithm is adopted to minimize the associated penalized criterion. For the WL2 approximation, we consider the variable metric forward-backward algorithm presented in [43]. For all these approaches, we test two variants. First, we propose to estimate the regularization parameter using the discrepancy principle [54, 55, 67]. In the second variant, the regularization parameter γ is adjusted empirically to achieve the maximum Signal-to-Noise Ratio (SNR) value defined as $\text{SNR} = 20\log_{10}\left(\frac{\|\overline{\mathbf{x}}\|}{\|\overline{\mathbf{x}}-\hat{\mathbf{x}}\|}\right)$, which requires the availability of the true image. Tables 2-7 report the results obtained with the different images in terms of SNR, SSIM [85], and approximate computation time needed for convergence. For each likelihood, we emphasize in bold the approximation of the covariance matrix that achieves the best quantitative result in the shortest computational time. Simulations are performed on an Intel(R) Xeon(R) CPU E5-2630, @ 2.40 GHz, using a Matlab7 implementation. All tested methods are initialized with the degraded image. The Monte Carlo averaging approximation is computed using parallel implementation with 16 cores by means of the command *PARFOR* of the Matlab® Parallel Computing Toolbox™.

One can observe that in most studied situations (see Tables 4-7), the diagonal approximation of the covariance matrix appears to give satisfactory qualitative results in a small computation time. However, in few other situations (see Tables 2 and 3), it fails to capture the real qualitative structures of the covariance matrix leading to a poorer performance. The latter issue is well alleviated by using the Monte Carlo approximation where good results, in terms of image quality, are achieved within $N_s = 160$ samples which is equivalent to a relative approximation error equal to 11%. A few improvements are observed by decreasing the approximation error to 5% using $N_s = 640$ samples.

We also notice that the GAST approximation does not seem to be suitable for very low

count images (see Tables 2 and 3), whereas, the other likelihoods lead to competitive results in all the experiments. The best tradeoff between restoration quality and small computational time seems to be achieved by the WL2 approximation.

Finally, it can be observed that, in Tables 2 and 4, our VBA method yields comparable performance in terms of SNR to the MAP estimate when the latter is computed with the optimal regularization parameter, while our approach requires less time to converge. In the other experiments, our approach leads to the best qualitative results. For instance, in Table 3, the gain in terms of SNR reaches up to 0.20 dB compared with the MAP estimator using the best regularization parameter, but our approach needs more time to converge. In Tables 5, 6 and 7, we achieve both the best quantitative results and the smallest computational time. It should be noted that for most tested scenarii, discrepancy based approaches perform relatively poorly compared with the other methods, especially in the case of low count images (see Table 2).

In Figures 2 - 7, we show some examples of visual results obtained with the different approaches, when the best approximation strategy for the covariance matrix is retained in the VBA method. It can be noticed that, unlike the other methods, the reconstructed images with the proposed VBA algorithm exhibit very few artifacts, without over-smoothed regions.

It should be emphasized that the problem of setting the regularization parameter for these algorithms must be carefully addressed as it highly impacts the quality of the restored image. The main advantage of our approach is that this parameter is tuned automatically without the need of the ground truth, while also often being the most competitive in terms of computation time. Furthermore, the performance of the proposed method could be further improved by using parallel implementation with more than 16 cores for the Monte Carlo approximation of the covariance matrix allowing either generating a higher number of samples (i.e. an improved estimation error) or a reduction of the computation time.

6 Conclusion

In this paper, we have proposed a variational Bayesian approach for solving signal recovery problems in the presence of non-Gaussian noise. Our approach has two main advantages. First, the regularization parameter is tuned automatically during the recovery process. Second, the designed method is applicable to a wide range of prior distributions and data fidelity terms. Simulations carried out on various images corrupted with mixed Poisson-Gaussian noise have shown that the proposed strategy constitutes a competitive solution for low computational and high-quality restoration of images compared with state-of-the art methods.

[Proof of Proposition 3.1]

Let $i \in \{1, \ldots, M\}$. Let us define $g_i : \mathbb{R} \to \mathbb{R}$ such that

$$(\forall v \in \mathbb{R}) \quad g_i(v) = \frac{v^2}{2} - \frac{\phi_i(v; y_i)}{\mu_i(y_i)}. \tag{31}$$

According to Assumption 3.1, g_i is convex. Its conjugate function [86, Chapter 13] reads:

$$(\forall w \in \mathbb{R}) \quad g_i^*(w) = \sup_{v \in \mathbb{R}} \left(vw - g_i(v)\right) \tag{32}$$

$$= \sup_{v \in \mathbb{R}} \left(vw + \frac{\phi_i(v; y_i)}{\mu_i(y_i)} - \frac{v^2}{2} \right) \tag{33}$$

$$= \sup_{v \in \mathbb{R}} \left(-\frac{1}{2}(v - w)^2 + \frac{\phi_i(v; y_i)}{\mu_i(y_i)} \right) + \frac{w^2}{2}. \tag{34}$$

According to Definition (10),

$$(\forall w \in \mathbb{R}) \quad g_i^*(w) = \varsigma_i(w; y_i) + \frac{w^2}{2}. \tag{35}$$

The conjugate of g_i^* is

$$\begin{aligned}(\forall v \in \mathbb{R}) \quad g_i^{**}(v) &= \sup_{w \in \mathbb{R}} \big(vw - g_i^*(w) \big) && (36)\\ &= \sup_{w \in \mathbb{R}} \left(vw - \frac{w^2}{2} - \varsigma_i(w; y_i) \right) \\ &= \sup_{w \in \mathbb{R}} \left(-\frac{1}{2}(v-w)^2 - \varsigma_i(w; y_i) \right) + \frac{v^2}{2} \\ &= - \inf_{w \in \mathbb{R}} \left(\frac{1}{2}(v-w)^2 + \varsigma_i(w; y_i) \right) + \frac{v^2}{2}. && (37)\end{aligned}$$

Since g_i is convex, $g_i = g_i^{**}$ so that

$$(\forall v \in \mathbb{R}) \quad - \frac{\phi_i(v; y_i)}{\mu_i(y_i)} = - \inf_{w \in \mathbb{R}} \left(\frac{1}{2}(v-w)^2 + \varsigma_i(w; y_i) \right) \tag{38}$$

which is equivalent to

$$(\forall v \in \mathbb{R}) \quad \phi_i(v; y_i) = \mu_i(y_i) \inf_{w \in \mathbb{R}} \left(\frac{1}{2}(v-w)^2 + \varsigma_i(w; y_i) \right), \tag{39}$$

so that (11) holds.

For every $v \in \mathbb{R}$, let

$$\widehat{w}_i(v) = g_i'(v). \tag{40}$$

By using standard properties of the subdifferential of convex functions [86, Corollary 16.24], the latter equality is equivalent to

$$v \in \partial g_i^* \big(\widehat{w}_i(v) \big), \tag{41}$$

which is a necessary and sufficient condition for $\widehat{w}_i(v)$ to be a minimizer of the convex function $w \mapsto g_i^*(w) - vw$. This minimizer is unique since $\widehat{w}_i(v)$ is uniquely defined by (40). We have therefore established that

$$\widehat{w}_i(v) = \operatorname*{argmax}_{w \in \mathbb{R}} \ \big(vw - g_i^*(w) \big). \tag{42}$$

The definition of g_i in (31) shows that (40) also reads

$$\widehat{w}_i(v) = v - \frac{1}{\mu_i(y_i)} \phi_i'(v; y_i). \tag{43}$$

According to (36), it is straightforward that $\widehat{w}_i(v)$ also reaches the infimum in (39). Hence the result by using (43) and (12).

References

[1] G. E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 16, no. 3, pp. 267–276, 1994.

[2] H. Tian, B. Fowler, and El. A. Gamal, "Analysis of temporal noise in CMOS photodiode active pixel sensor," *IEEE J. Solid-State Circuits*, vol. 36, no. 1, pp. 92–101, 2001.

[3] James R Janesick, *Photon transfer*, SPIE press San Jose, 2007.

[4] L. Azzari and A. Foi, "Gaussian-Cauchy mixture modeling for robust signal-dependent noise estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2014)*, Florance, Italy, 4-9 May 2014, pp. 5357–5361.

[5] X. Liu, M. Tanaka, and M. Okutomi, "Practical signal-dependent noise parameter estimation from a single noisy image," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4361–4371, 2014.

[6] A. Chakrabarti and T. E. Zickler, "Image restoration with signal-dependent camera noise," Research report, 2012.

[7] L. Boubchir, S. Al-Maadeed, and A. Bouridane, "Undecimated wavelet-based Bayesian denoising in mixed Poisson-Gaussian noise with application on medical and biological images," in *Proc. Int. Conf. Image Process. Theory, Tools and Applicat. (IPTA 2014)*, Paris, France, 14-17 Oct. 2014, pp. 1–5.

[8] Stefan M Moser, "Capacity results of an optical intensity channel with input-dependent Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 207–223, 2012.

[9] G. Aubert and J-F Aujol, "A variational approach to removing multiplicative noise," *SIAM J. Appl. Math.*, vol. 68, no. 4, pp. 925–946, 2008.

[10] A. Buades, B. Coll, and J-M Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Comput. Soc. Conf. Comput. Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, USA, 20-25 Jun. 2005, vol. 2, pp. 60–65.

[11] Y. Dong and T. Zeng, "A convex variational model for restoring blurred images with multiplicative noise," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1598–1625, 2013.

[12] Y-M. Huang, M. K. Ng, and Y-W Wen, "A new total variation method for multiplicative noise removal," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 20–40, 2009.

[13] S. Parrilli, M. Poderico, Cesario V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens*, vol. 50, no. 2, pp. 606–616, 2012.

[14] J-F. Cai, R. H. Chan, and M. Nikolova, "Fast two-phase image deblurring under impulse noise," *J. Math. Imaging Vision*, vol. 36, no. 1, pp. 46–53, 2010.

[15] Y. Xiao, T. Zeng, J. Yu, and M. K. Ng, "Restoration of images corrupted by mixed gaussian-impulse noise via $\ell_1 - \ell_0$ minimization," *Pattern Recogn.*, vol. 44, no. 8, pp. 1708–1720, 2011.

[16] M. Yan, "Restoration of images corrupted by impulse noise and mixed gaussian impulse noise using blind inpainting," *SIAM J. Imaging Sci.*, vol. 6, no. 3, pp. 1227–1245, 2013.

[17] J. Boulanger, L. Condat, T. Piolot, L. Sengmanivong, and N. Pustelnik, "Nonsmooth convex optimization for structured illumination microscopy image reconstruction," Research report, GIPSA-lab, 2015.

[18] A. P. Petropulu, J. C. Pesquet, and Xueshi Yang, "Power-law shot noise and its relationship to long-memory α-stable processes," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 1883–1892, Jul 2000.

[19] J. Salmon, Z. Harmany, C-A. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *J. Math. Imaging Vis.*, vol. 48, no. 2, pp. 279–294, 2014.

[20] S. Setzer, G. Steidl, and T. Teuber, "Deblurring Poissonian images by split Bregman techniques," *J. Vis. Comm. Image Repr.,*, vol. 21, no. 3, pp. 193–199, 2010.

[21] T. Jeong, H. Woo, and S. Yun, "Frame-based Poisson image restoration using a proximal linearized alternating direction method," *Inverse Probl.*, vol. 29, no. 7, pp. 075007, 2013.

[22] S. Harizanov, J-C Pesquet, and G. Steidl, "Epigraphical projection for solving least squares Anscombe transformed constrained optimization problems," in *Proc. Int. Conf. Scale Space and Variational Methods Comput. Vision (SSVM 2013)*, Schloss Seggau, Austria, 2-6 Jun. 2013, pp. 125–136.

[23] F-X. Dupé, M. J. Fadili, and J-L Starck, "Deconvolution under Poisson noise using exact data fidelity and synthesis or analysis sparsity priors," *Stat. Methodol.*, vol. 9, no. 1, pp. 4–18, 2012.

[24] C. Chaux, J-C Pesquet, and N. Pustelnik, "Nested iterative algorithms for convex constrained image recovery problems," *SIAM J. Imaging Sci.*, vol. 2, no. 2, pp. 730–762, 2009.

[25] S. Bonettini and M. Prato, "New convergence results for the scaled gradient projection method," *Inverse Probl.*, vol. 31, no. 9, pp. 095008, 2015.

[26] S. Anthoine, J. F. Aujol, Y. Boursier, and C. Mélot, "Some proximal methods for Poisson intensity CBCT and PET," *Inverse Probl. Imaging*, vol. 6, no. 4, pp. 565–598, 2012.

[27] S Bonettini and V Ruggiero, "An alternating extragradient method for total variation-based image restoration from Poisson data," *Inverse Probl.*, vol. 27, no. 9, pp. 095001, 2011.

[28] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 1935–1946, May 2016.

[29] T. D. Roberts and N. G. Kingsbury, "Bayesian denoising/deblurring of Poisson-Gaussian corrupted data using complex wavelets," in *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI 2014)*, Beijing, China, 29 Apr. - 2 May 2014, pp. x–x+2.

[30] M. Mäkitalo and A. Foi, "Noise parameter mismatch in variance stabilization, with an application to Poisson–Gaussian noise estimation," *IEEE Trans. Image Process.*, vol. 23, no. 12, pp. 5348–5359, 2014.

[31] M. Mäkitalo and A. Foi, "Poisson-Gaussian denoising using the exact unbiased inverse of the generalized Anscombe transformation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2012)*, Kyoto, Japan, 25-30 Mar. 2012, pp. 1081–1084.

[32] M. Mäkitalo and A. Foi, "Optimal inversion of the generalized Anscombe transformation for Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 91–103, 2013.

[33] F. Luisier, T. Blu, and M. Unser, "Image denoising in mixed Poisson-Gaussian noise," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 696–708, 2011.

[34] J. Li, Z. Shen, R. Jin, and X. Zhang, "A reweighted ℓ_2 method for image restoration with Poisson and mixed Poisson-Gaussian noise," *Inverse Probl. Imaging*, vol. 9, no. 3, pp. 875–894, 2015.

[35] H. Lantéri and C. Theys, "Restoration of astrophysical images - the case of Poisson data with additive Gaussian noise," *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 15, pp. 2500–2513, 2005.

[36] A. Jezierska, E. Chouzenoux, J.-C. Pesquet, and H. Talbot, "A primal-dual proximal splitting approach for restoring data corrupted with Poisson-Gaussian noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2012)*, Kyoto, Japan, 25-30 Mar. 2012, pp. 1085–1088.

[37] A. Jezierska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler, "An EM approach for time-variant Poisson-Gaussian model parameter estimation," *IEEE Trans. Signal Process.*, vol. 62, no. 1, pp. 17–30, 2014.

[38] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, pp. 1737–1754, 2008.

[39] E. Chouzenoux, A. Jezierska, J.-C. Pesquet, and H. Talbot, "A convex approach for image restoration with exact Poisson-Gaussian likelihood," *SIAM J. Imag. Sci.*, vol. 8, no. 4, pp. 2662–2682, 2015.

[40] P. Kittisuwan, "Medical image denoising using simple form of MMSE estimation in Poisson–Gaussian noise model," *Int. J. Biomath.*, vol. 9, no. 02, pp. 1650020, 2016.

[41] L. Calatroni, C. Cao, J. C. De Los Reyes, C.-B. Schönlieb, and T. Valkonen, "Bilevel approaches for learning of variational imaging models," 2016, to appear in RADON book series.

[42] Guy Demoment, "Image reconstruction and restoration: Overview of common estimation structures and problems," *IEEE Trans. Acous., Speech Signal Process*, vol. 37, no. 12, pp. 2024–2036, 1989.

[43] A. Repetti, E. Chouzenoux, and J.-C. Pesquet, "A penalized weighted least squares approach for restoring data corrupted with signal-dependent noise," in *Proc. Eur. Sig. Proc. Conf. (EUSIPCO 2012)*, Bucharest, Roumania, 27-31 Aug. 2012, pp. 1553–1557.

[44] D. L. Snyder, A. M. Hammoud, and R. L. White, "Image recovery from data acquired with a charge-coupled-device camera," *J. Opt. Soc. Am. A*, vol. 10, no. 5, pp. 1014–1023, 1993.

[45] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lantéri, and M. Bertero, "The study of an iterative method for the reconstruction of images corrupted by Poisson and Gaussian noise," *Inverse Probl.*, vol. 24, no. 3, pp. 035016, 2008.

[46] B. Bajić, J. Lindblad, and N. Sladoje, "Blind restoration of images degraded with mixed Poisson-Gaussian noise with application in transmission electron microscopy," in *Proc. IEEE Int. Symp. Biomedical Imaging (ISBI 2016)*, Prague, Czech Republic, 13-16 Apr. 2016, pp. 123–127.

[47] S. Ramani, T. Blu, and M. Unser, "Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithms," *IEEE Trans. Image Process.*, vol. 17, no. 9, pp. 1540–1554, 2008.

[48] Y. C. Eldar, "Generalized SURE for exponential families: Applications to regularization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 471–481, 2009.

[49] J.-C. Pesquet, A. Benazza-Benyahia, and C. Chaux, "A SURE approach for digital signal/image deconvolution problems," *IEEE Trans. Signal Process.*, vol. 57, no. 12, pp. 4616–4632, 2009.

[50] C.-A. Deledalle, S. Vaiter, J. Fadili, and G. Peyré, "Stein unbiased gradient estimator of the risk (SUGAR) for multiple parameter selection," *SIAM J. Imaging Sci.*, vol. 7, no. 4, pp. 2448–2487, 2014.

[51] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comput. Harmon. Anal.*, vol. 30, no. 3, pp. 407–422, 2011.

[52] M. S. C. Almeida and M. Figueiredo, "Parameter estimation for blind and non-blind deblurring using residual whiteness measures," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2751–2763, 2013.

[53] P. C. Hansen, M. E. Kilmer, and R. H. Kjeldsen, "Exploiting residual information in the parameter choice for discrete ill-posed problems," *BIT Numer. Math.*, vol. 46, no. 1, pp. 41–59, 2006.

[54] M. Bertero, P. Boccacci, G. Talenti, R. Zanella, and L. Zanni, "A discrepancy principle for Poisson data," *Inverse Probl.*, vol. 26, no. 10, pp. 105004, 2010.

[55] L. Zanni, A. Benfenati, M. Bertero, and V. Ruggiero, "Numerical methods for parameter estimation in Poisson data inversion," *J. Math. Imaging Vis.*, vol. 52, no. 3, pp. 397–413, 2015.

[56] W. Ying, Y. Jiang, and C. Fan, "A blind receiver with multiple antennas in impulsive noise with Gaussian mixtures through MCMC approaches," in *Proc. IEEE Int. Conf. Com. Technol. (ICCT 2012)*, Chengdu, China, 9-11 Nov. 2012, pp. 961–965.

[57] Y. Altmann, A. Wallace, and S. McLaughlin, "Spectral unmixing of multispectral Lidar signals," *IEEE Trans. Signal Process.*, vol. 63, no. 20, pp. 5525–5533, 2015.

[58] J. Murphy and S. Godsill, "Joint Bayesian removal of impulse and background noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2011)*, Prague, Czech Republic, 22-27 May 2011, pp. 261–264.

[59] L. Chaâri, T. Vincent, F. Forbes, M. Dojat, and P. Ciuciu, "Fast joint detection-estimation of evoked brain activity in event-related FMRI using a variational approach," *IEEE Trans. Med. Imaging*, vol. 32, no. 5, pp. 821–837, May 2013.

[60] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*, Springer-Verlag, New York, USA, 2005.

[61] Y. Zheng, A. Fraysse, and T. Rodet, "Efficient variational Bayesian approximation method based on subspace optimization," *IEEE Trans. Image Process.*, vol. 24, no. 2, pp. 681–693, 2015.

[62] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian super resolution," *IEEE Trans. Image Process.*, vol. 20, no. 4, pp. 984–999, 2011.

[63] Z. Chen, S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian methods for multimedia problems," *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 1000–1017, 2014.

[64] A. Drémeau, C. Herzet, and L. Daudet, "Boltzmann machine and mean-field approximation for structured sparse decompositions," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3425–3438, 2012.

[65] F. Murtagh, J.-L. Starck, and A. Bijaoui, "Image restoration with noise suppression using a multiresolution support," *Astron. Astrophys. Suppl.*, vol. 112, pp. 179–189, 1995.

[66] J.-L. Starck and F. Murtagh, "Automatic noise estimation from the multiresolution support," *Publ. Astron. Soc. Pac.*, vol. 110, no. 744, pp. 193–199, 1998.

[67] J. M. Bardsley and J. Goldes, "Regularization parameter selection methods for ill-posed Poisson maximum likelihood estimation," *Inverse Probl.*, vol. 25, no. 9, pp. 095005, 2009.

[68] N. Pustelnik, A. Benazza-Benhayia, Y. Zheng, and J.-C. Pesquet, "Wavelet-based image deconvolution and reconstruction," 2016, to appear in Wiley Encyclopedia of Electrical and Electronics Engineering.

[69] M. Pereyra, J. Bioucas-Dias, and M. A. T. Figueiredo, "Maximum-a-posteriori estimation with unknown regularisation parameters," in *Proc. Eur. Sig. Proc. Conf. (EUSIPCO 2015)*, Nice, France, 31 Aug. - 4 Sept. 2015, pp. 2244–2248.

[70] G. Parisi, *Statistical Field Theory*, Addison Wesley, Redwood City, CA, 1988.

[71] M. Pereyra, P. Schniter, E. Chouzenoux, J.-C. Pesquet, J.-Y. Tourneret, A. O Hero, and S. McLaughlin, "A survey of stochastic simulation and optimization methods in signal processing," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 2, pp. 224–241, Mar. 2016.

[72] Y. Marnissi, Y. Zheng, and J-C Pesquet, "Fast variational Bayesian signal recovery in the presence of Poisson-Gaussian noise," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP 2016)*, Shanghai, China, 20-25 Mar. 2016, pp. 3964–3968.

[73] E. W. Tramel, A. Drémeau, and F. Krzakala, "Approximate message passing with restricted Boltzmann machine priors," *J. Stat. Mech.: Theory and Experiment*, vol. 2016, no. 7, pp. 073401, 2016.

[74] A. Fraysse and T. Rodet, "A measure-theoretic variational Bayesian algorithm for large dimensional problems," *SIAM J. Imaging Sci.*, vol. 7, no. 4, pp. 2591–2622, 2014.

[75] A. Repetti, *Algorithmes d'optimisation en grande dimension: applications à la résolution de problèmes inverses*, Ph.D. thesis, Université Paris-Est Marne la Vallée, 2015.

[76] M. Seeger and G. Bouchard, "Fast variational Bayesian inference for non-conjugate matrix factorization models," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS 2012)*, La Palma, Canary Islands, 21-23 Apr. 2012, pp. 1012–1018.

[77] D. Geman and C. Yang, "Nonlinear image recovery with half-quadratic regularization," *IEEE Trans. Image Process.*, vol. 4, no. 7, pp. 932–946, 1995.

[78] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.

[79] M. Nikolova and M.-K. Ng, "Analysis of half-quadratic minimization methods for signal and image recovery," *SIAM J. Sci. Comput.*, vol. 27, pp. 937–966, 2005.

[80] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and EM algorithms," *IEEE Signal Process. Lett.*, vol. 11, no. 9, pp. 709–712, 2004.

[81] Y. Marnissi, E. Chouzenoux, J.-C. Pesquet, and A. Benazza-Benyahia, "An auxiliary variable method for langevin based MCMC algorithms," in *Proc. IEEE Stat. Signal Process. Workshop (SSP 2016)*, Palma de Mallorca, Spain, 26-29 Jun. 2016, pp. 297–301.

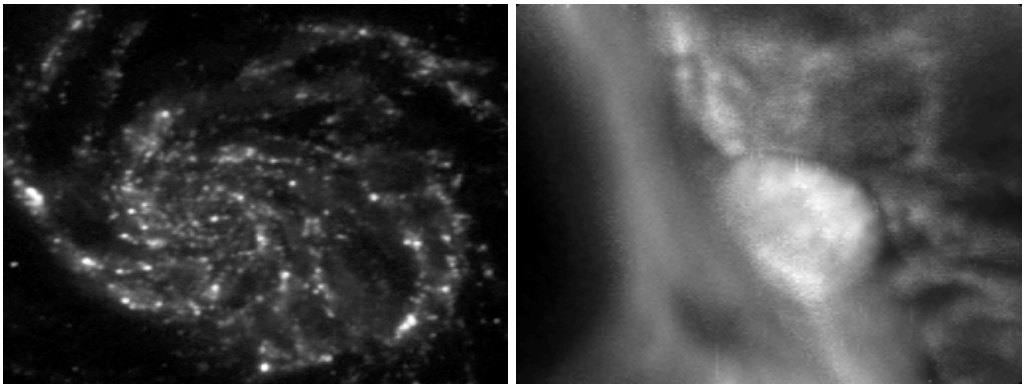
[82] P. L. Combettes and J.-C. Pesquet, "Proximal splitting methods in signal processing," in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, H. H. Bauschke, R. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, Eds., pp. 185–212. Springer-Verlag, New York, 2010.

[83] F. Orieux, O. Féron, and J. F. Giovannelli, "Sampling high-dimensional Gaussian distributions for general linear inverse problems," *IEEE Signal Process. Lett.*, vol. 19, no. 5, pp. 251–254, 2012.

[84] G. Papandreou and A. L. Yuille, "Gaussian sampling by local perturbations," in *Proc. Neural Inform. Process. Syst. Conf. 23 (NIPS 2010)*, J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, and A. Culotta, Eds., Vancouver, Canada, 06-11 Dec. 2010, pp. 1858–1866.

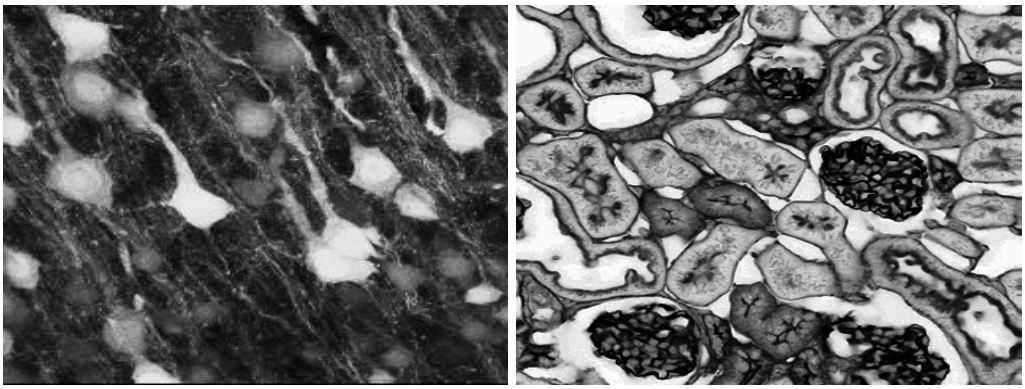
[85] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, 2004.

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



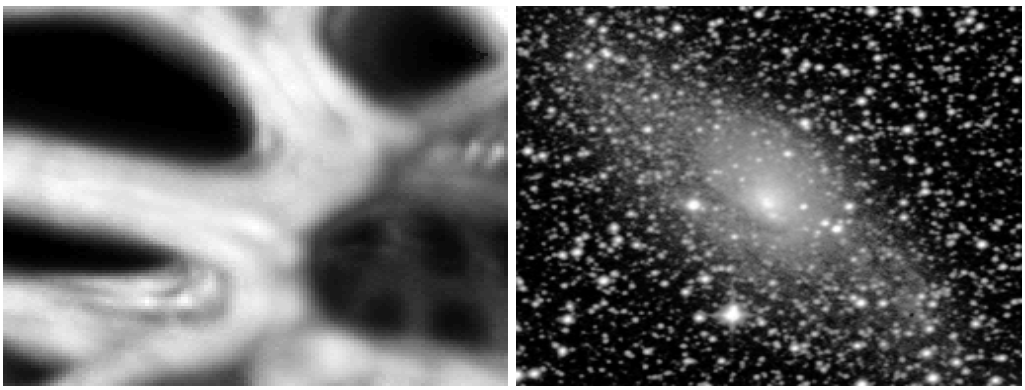
(a) Image $\overline{\mathbf{x}}_1$
(256×256, $x^+ = 10$)

(b) Image $\overline{\mathbf{x}}_2$
(190×190, $x^+ = 12$)



(c) Image $\overline{\mathbf{x}}_3$
(257×256, $x^+ = 15$)

(d) Image $\overline{\mathbf{x}}_4$
(350×350, $x^+ = 20$)



(e) Image $\overline{\mathbf{x}}_5$
(128×128, $x^+ = 20$)

(f) Image $\overline{\mathbf{x}}_6$
(256×256, $x^+ = 100$)

Figure 1: Original images.

			GAST	SPoiss	WL2
VBA	Approx. 1	SNR	8.13	9.36	9.90
		SSIM	0.3987	0.4790	0.5140
		Time (s.)	55	62	67
	Approx. 2 $N_s = 160$	SNR	9.57	10.17	10.22
		SSIM	0.5260	0.6017	0.6058
		Time (s.)	688	601	1011
	Approx. 2 $N_s = 640$	SNR	**9.61**	**10.20**	**10.27**
		SSIM	**0.5308**	**0.6088**	**0.6112**
		Time (s.)	**3606**	**3507**	**3510**
MAP	Discrepancy principle	SNR	-1.13	5.24	10.17
		SSIM	0.0980	0.2961	0.6131
		Time (s.)	3326	2215	3053
	Best parameter	SNR	9.46	10.40	10.39
		SSIM	0.5078	0.6029	0.5920
		Time (s.)	4380	2560	13740

Table 2: Restoration results for image $\overline{\mathbf{x}}_1$ with $x^+ = 10$ and $\sigma^2 = 4$. Uniform kernel with size 5×5. Initial SNR= -2.55 dB.

			GAST	SPoiss	WL2
VBA	Approx. 1	SNR	13.97	15.19	16.41
		SSIM	0.3544	0.4167	0.4959
		Time (s.)	58	64	70
	Approx. 2 $N_s = 160$	SNR	18.05	19.07	19.11
		SSIM	0.6664	0.6930	0.7066
		Time (s.)	524	498	491
	Approx. 2 $N_s = 640$	SNR	**18.11**	**19.12**	**19.13**
		SSIM	**0.6778**	**0.7034**	**0.7152**
		Time (s.)	**2048**	**1828**	**1735**
MAP	Discrepancy principle	SNR	16.52	17.41	18.09
		SSIM	0.5484	0.7570	0.6732
		Time (s.)	594	583	2286
	Best parameter	SNR	17.83	18.73	19.09
		SSIM	0.6519	0.6646	0.6702
		Time (s.)	674	705	4164

Table 3: Restoration results for the image $\overline{\mathbf{x}}_2$, $x^+ = 12$ and $\sigma^2 = 9$. Gaussian kernel with size 25×25, std 1.6. Initial SNR= 2.21 dB.

			GAST	SPoiss	WL2
VBA	Approx. 1	SNR	11.42	11.94	12.25
		SSIM	0.4184	0.4403	0.4588
		Time (s.)	45	47	53
	Approx. 2 $N_s = 160$	SNR	12.04	12.31	12.28
		SSIM	0.4555	0.4624	0.4627
		Time (s.)	328	332	396
	Approx.2 $N_s = 640$	SNR	**12.09**	**12.36**	**12.33**
		SSIM	**0.4617**	**0.4684**	**0.4683**
		Time (s.)	**1965**	**2051**	**2019**
MAP	Discrepancy principle	SNR	12.08	12.38	12.08
		SSIM	0.4523	0.4582	0.4314
		Time (s.)	6252	3865	1929
	Best parameter	SNR	12.17	12.45	12.37
		SSIM	0.4531	0.4576	0.4565
		Time (s.)	3348	2441	2525

Table 4: Restoration results for the image $\overline{\mathbf{x}}_3$ with $x^+ = 15$ and $\sigma^2 = 9$. Uniform kernel with size 5×5. Initial SNR= 3.14 dB.

			GAST	SPoiss	WL2
VBA	Approx. 1	SNR	**13.80**	**13.90**	**13.66**
		SSIM	**0.5752**	**0.5769**	**0.5582**
		Time (s.)	**29**	**34**	**88**
	Approx. 2 $N_s = 160$	SNR	13.72	13.76	13.56
		SSIM	0.5667	0.5641	0.5491
		Time (s.)	555	580	757
	Approx. 2 $N_s = 640$	SNR	13.78	13.81	13.61
		SSIM	0.5715	0.5687	0.5534
		Time (s.)	1897	2170	2719
MAP	Discrepancy principle	SNR	13.48	13.60	13.39
		SSIM	0.5348	0.5393	0.5103
		Time (s.)	3049	769	2644
	Best parameter	SNR	13.60	13.71	13.75
		SSIM	0.5568	0.5605	0.5602
		Time (s.)	8390	8477	2397

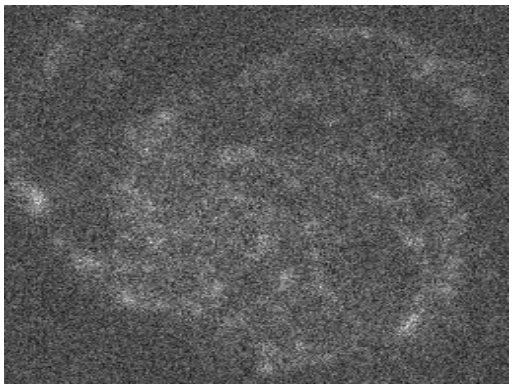
Table 5: Restoration results for the image $\overline{\mathbf{x}}_4$ with $x^+ = 20$ and $\sigma^2 = 9$. Uniform kernel with size 5×5. Initial SNR= 7.64 dB.

			GAST	SPoiss	WL2
VBA	Approx. 1	SNR	19.5	20.23	**20.71**
		SSIM	0.6649	0.7135	**0.7793**
		Time (s.)	5	6	**14**
	Approx. 2 $N_s = 160$	SNR	20.27	20.59	20.56
		SSIM	0.7473	0.7660	0.7877
		Time (s.)	49	56	62
	Approx.2 $N_s = 640$	SNR	**20.35**	**20.67**	20.64
		SSIM	**0.7563**	**0.7798**	0.7989
		Time (s.)	**178**	**184**	220
MAP	Discrepancy principle	SNR	19.39	19.50	18.70
		SSIM	0.7458	0.7550	0.7448
		Time (s.)	717	1201	1087
	Best parameter	SNR	20.15	20.41	20.44
		SSIM	0.7535	0.7594	0.7628
		Time (s.)	559	125	253

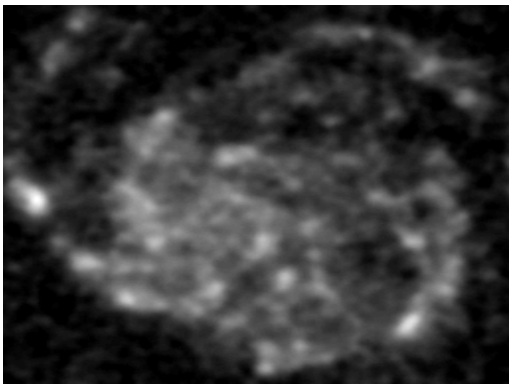
Table 6: Restoration results for image $\overline{\mathbf{x}}_5$ with $x^+ = 20$ and $\sigma^2 = 9$. Gaussian kernel with size 7×7, std 1. Initial SNR= 8.55 dB.

			GAST	SPoiss	WL2
VBA	Approx. 1	SNR	**14.17**	14.13	13.90
		SSIM	**0.7655**	0.7647	0.7569
		Time (s.)	**9**	8	26
	Approx.2 $N_s = 160$	SNR	14.1	14.13	14.09
		SSIM	0.7605	0.7619	0.7620
		Time (s.)	104	148	246
	Approx. 2 $N_s = 640$	SNR	14.16	**14.19**	**14.16**
		SSIM	0.7639	**0.7650**	**0.7658**
		Time (s.)	332	**479**	**913**
MAP	Discrepancy principle	SNR	13.23	13.29	13.32
		SSIM	0.7104	0.7126	0.7117
		Time (s.)	2796	4900	1045
	Best parameter	SNR	13.77	13.79	13.84
		SSIM	0.7565	0.7570	0.7591
		Time (s.)	10084	10005	821

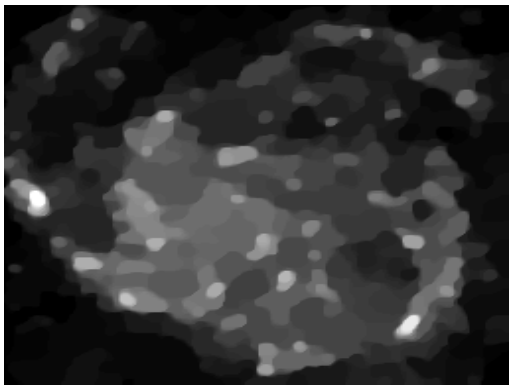
Table 7: Restoration results for the image $\overline{\mathbf{x}}_6$ with $x^+ = 100$ and $\sigma^2 = 36$. Uniform kernel with size 3×3. Initial SNR= 10.68 dB.



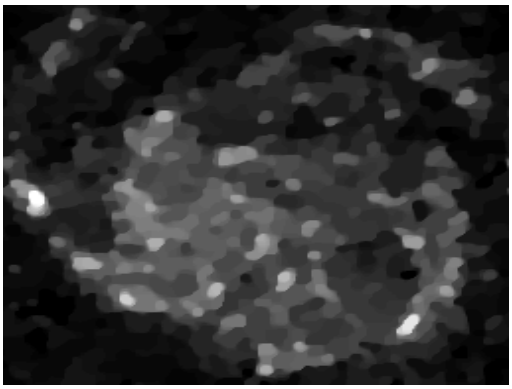
(a) Degraded image with SNR= -2.55 dB (Uniform kernel 5×5, $x^+ = 10$ and $\sigma^2 = 4$).



(b) Restored image with VBA approach using the Monte Carlo approximation with 640 samples: SNR= 10.27 dB

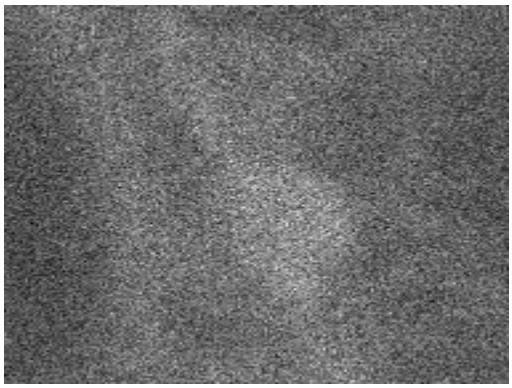


(c) Restored image with discrepancy principle: SNR= 10.17 dB

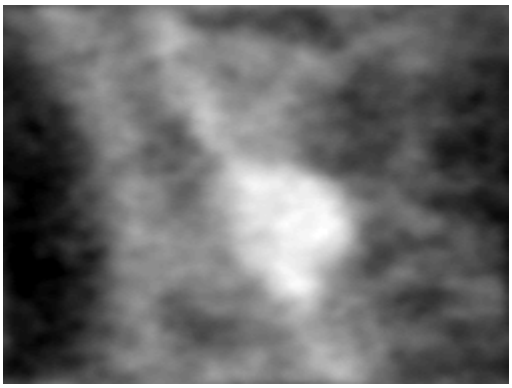


(d) Restored image with best parameter: SNR= 10.39 dB

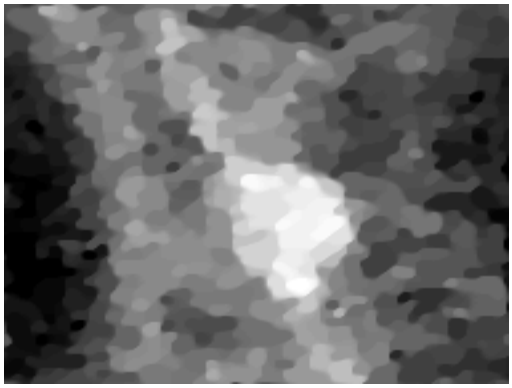
Figure 2: Restoration results for image $\bar{\mathbf{x}}_1$ using WL2 approximation.



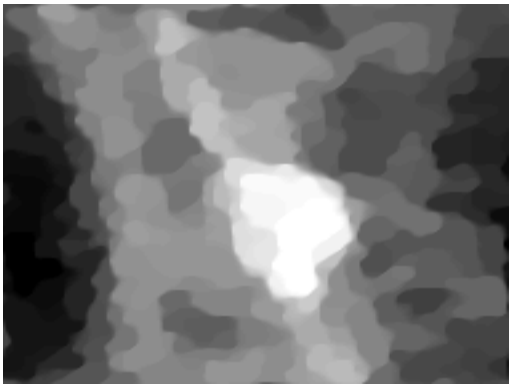
(a) Degraded image with SNR= 2.21 dB (Gaussian kernel 25×25, std 1.6, $x^+ = 12$ and $\sigma^2 = 9$).



(b) Restored image with VBA approach using the Monte Carlo approximation with 640 samples: SNR= 19.12 dB

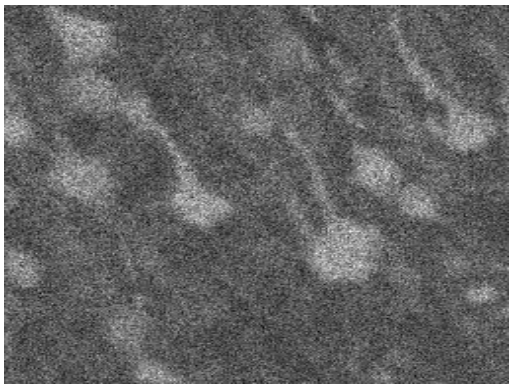


(c) Restored image with discrepancy principle: SNR= 17.41 dB

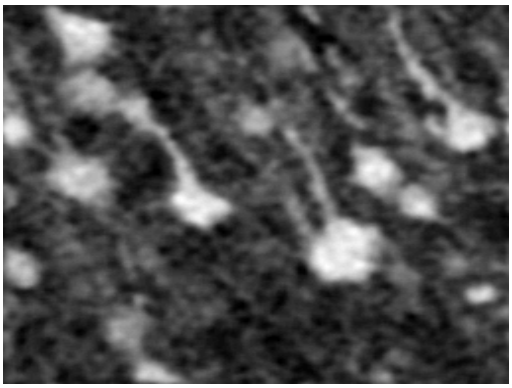


(d) Restored image with best parameter: SNR= 18.73 dB

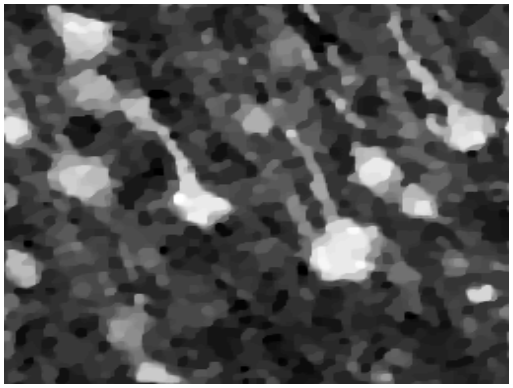
Figure 3: Restoration results for image $\bar{\mathbf{x}}_2$ using SPoiss approximation.



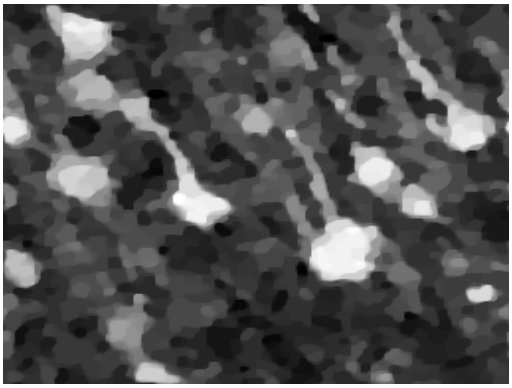
(a) Degraded image with SNR= 3.14 dB (Uniform kernel 5×5, $x^+ = 15$ and $\sigma^2 = 9$).



(b) Restored image with VBA approach using the Monte Carlo approximation with 640 samples: SNR= 12.36 dB

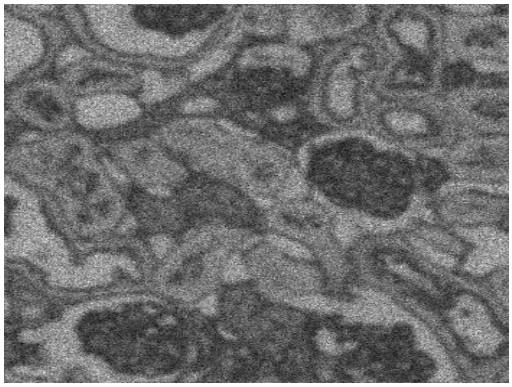


(c) Restored image with discrepancy principle: SNR= 12.38 dB

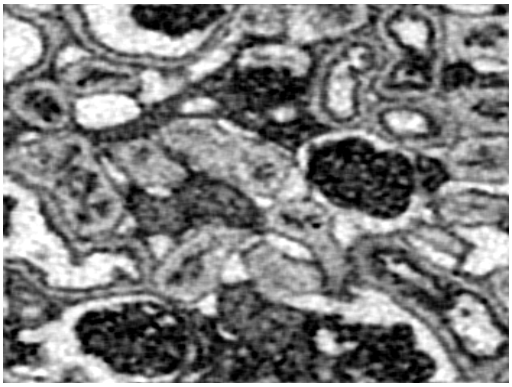


(d) Restored image with best parameter: SNR= 12.45 dB

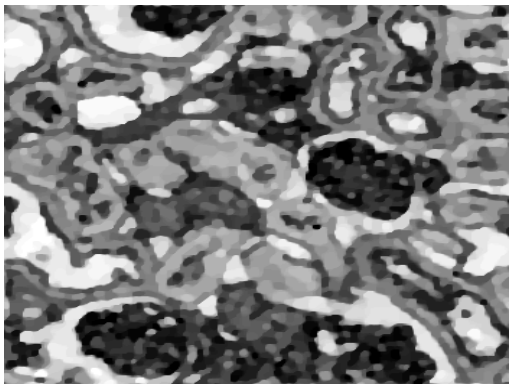
Figure 4: Restoration results for image $\bar{\mathbf{x}}_3$ using SPoiss approximation.



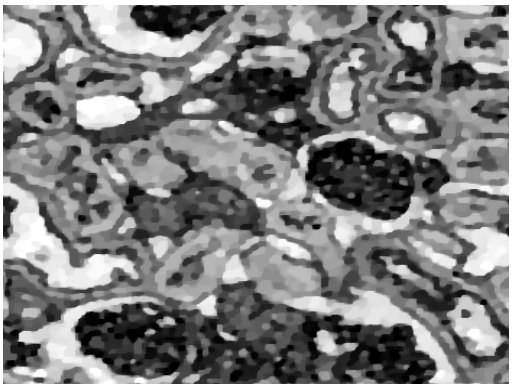
(a) Degraded image with SNR=7.64 dB (Uniform kernel 5×5, $x^+ = 20$ and $\sigma^2 = 9$).



(b) Restored image with VBA approach using the diagonal approximation: SNR= 13.80 dB

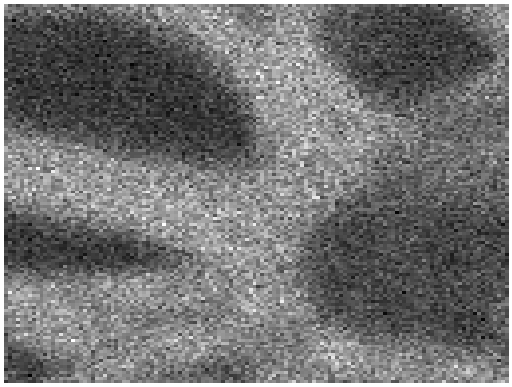


(c) Restored image with discrepancy principle: SNR= 13.48 dB

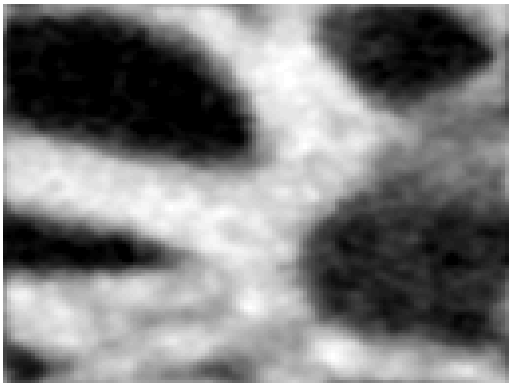


(d) Restored image with best parameter: SNR= 13.60 dB

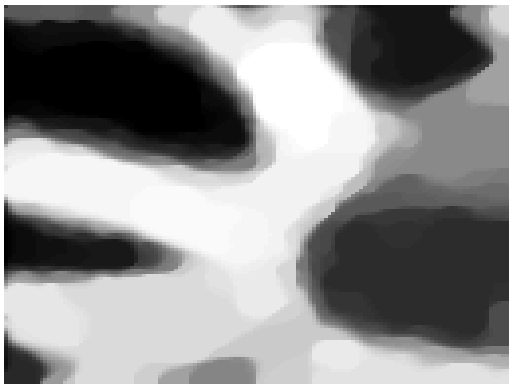
Figure 5: Restoration results for image $\bar{\mathbf{x}}_4$ using GAST approximation.



(a) Degraded image with SNR= 8.55 dB (Gaussian kernel 7×7, std 1. $x^+ = 20$ and $\sigma^2 = 9$).



(b) Restored image with VBA approach using the diagonal approximation: SNR= 20.71 dB

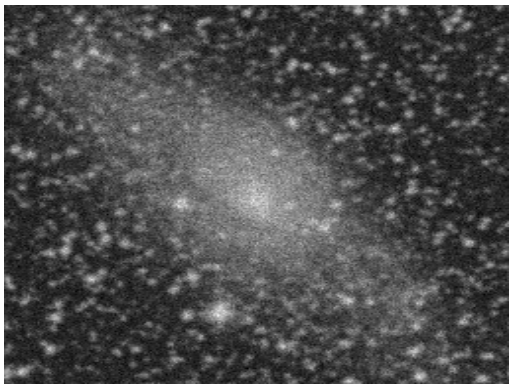


(c) Restored image with discrepancy principle: SNR= 18.70 dB

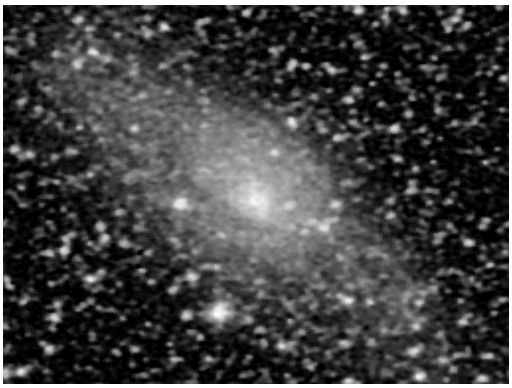


(d) Restored image with best parameter: SNR= 20.44 dB

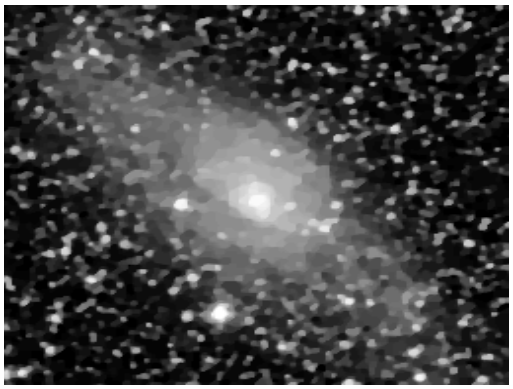
Figure 6: Restoration results for image $\bar{\mathbf{x}}_5$ using WL2 approximation.



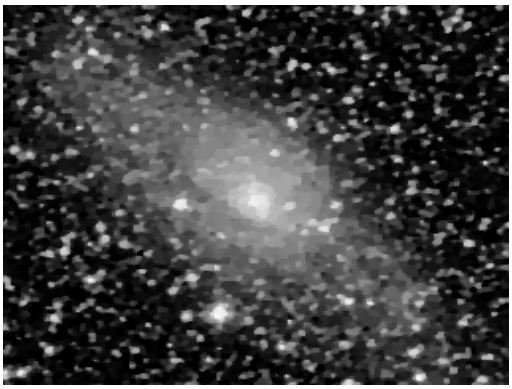
(a) Degraded image with SNR= 10.68 dB (Uniform kernel 3×3, $x^+ = 100$ and $\sigma^2 = 36$).



(b) Restored image with VBA approach using the Monte Carlo approximation with 640 samples: SNR= 14.16 dB



(c) Restored image with discrepancy principle: SNR= 13.32 dB



(d) Restored image with best parameter: SNR= 13.84 dB

Figure 7: Restoration results for image $\bar{\mathbf{x}}_6$ using WL2 approximation.