Semi-unsupervised Bayesian convex image restoration with location mixture of Gaussian

François Orieux, Raphael Chinchilla

Semi-Unsupervised Bayesian Convex Image Restoration with Location Mixture of Gaussian

François Orieux
Laboratoire des Signaux et Systèmes – Univ. Paris-Sud 11, CNRS, CentraleSupélec, Université Paris-Saclay – Gif-sur-Yvette, France.
Email: orieux@l2s.centralesupelec.fr

Raphael Chinchilla
Laboratoire des Signaux et Systèmes – Univ. Paris-Sud 11, CNRS, CentraleSupélec, Université Paris-Saclay – Gif-sur-Yvette, France.
Email: raphael@chinchillas.net

***Abstract*—Convex image restoration is a major field in inverse problems. The problem is often addressed by hand-tuning hyper-parameters. We propose an incremental contribution about a Bayesian approach where a convex field is constructed via Location Mixture of Gaussian and the estimator computed with a fast MCMC algorithm. Main contributions are a new field with several operator avoiding crosslike artifacts and a fallback sampling algorithm to prevent numerical errors. Results, in comparison to standard supervised results, have equivalent quality in a quasi-unsupervised approach and go with uncertainty quantification.**

I. Introduction

Image restoration, or deconvolution, is a major problem with an abundant literature and applications found in, for instance, optics, radio astronomy, microscopy [1]–[6]. Deconvolution is also related to the resolution of ill-posed inverse problems where the likelihood of data presents defects like missing data or instability [7].

Optimization methods, based on the minimization of mixed criteria, are common [8], [9] and popular in high-dimension, thanks to efficient algorithms, especially in the convex case [10]. However, the hyper-parameters that determine the balance between the data and prior adequacy are usually hand-tuned (supervised) and uncertainty about the optimum estimate are not well defined or studied.

Bayesian methods, on the contrary, present a natural way to estimate the hyper-parameters (unsupervised) by considering them as nuisance parameter and marginalization of the joint *a posteriori* law [7], [11]. Moreover, *a posteriori* law analysis brings confidence interval analysis and quantification of uncertainty about the estimation.

Within the Bayesian methods, variationnal approaches had interest by providing apparently fast algorithms [12]–[14]. However, variationnal approaches build an approximation of the *a posteriori* law and the uncertainty is known to be underestimated, induced by the separability and the non-correlation model.

The approach of this work is the use of Monte Carlo Markov Chains (MCMC) algorithms for exact *a posteriori* law exploration and computation of the Expectation *a posteriori* estimator (EAP). However, results on unsupervised method are sparser with some existing work [15], for instance based on marginal likelihood [16], quadratic prior [17], [18], per variables Metropolis-Hastings [19] or more recently Moreau approximation [20].

In [21], authors present a fast MCMC deconvolution method limited to quadratic prior. Work of J.-F. Giovannelli [22] presents an unsupervised convex deconvolution approach. However, the use of multiple regularization operator, like horizontal and vertical gradient, is missing. In addition, some numerical instabilities are present due to special functions. Finally, divergence of the MCMC chains may be observed in practice when the full set of hyper-parameters is estimated.

We propose an incremental work based on the method of [22]. We show that multiple operators can be used, providing better results and no crosslike effect of the Laplacian around the edges [23]. To prevent numerical instability of the simulation we present a fallback solution based on a simple and efficient Metropolis-Hastings algorithm. Finally, by just considering a known or estimated noise level, the method becomes robust and estimate the *a priori* shape and precision. The proposed algorithm does not pretend to achieve the best possible results but rather to complement ongoing work for unsupervised convex image restoration.

The paper is organized as follows : Sec. II presents the notation and the methodological context. Sec. III is devoted to the convex potential, the unsupervised approach and the MCMC algorithm. Finally, Sec. IV shows some results and comparison to standard algorithm of the same class.

II. Notations and Problems

We consider N pixels images $\boldsymbol{x} \in \mathbb{R}^N$. The data model is $\boldsymbol{y} = \boldsymbol{H}\boldsymbol{x} + \boldsymbol{n}$ where $\boldsymbol{y} \in \mathbb{R}^N$ collects the data, $\boldsymbol{H} \in \mathbb{R}^{N \times N}$ is the convolution operator and $\boldsymbol{n} \in \mathbb{R}^N$ an unknown random noise. The convolution are circulant and computed via Discrete Fourier Transform $\boldsymbol{F}$ with a diagonalizable operator $\boldsymbol{H} = \boldsymbol{F}^\dagger \boldsymbol{\Lambda}_{\boldsymbol{H}} \boldsymbol{F}$, where $\boldsymbol{\Lambda}_{\boldsymbol{H}} = \operatorname{diag}(\mathring{h}_0, \dots, \mathring{h}_{N-1})$.

The *a priori* noise law is supposed white Gaussian with a *known* noise level $\sigma_{\boldsymbol{n}} = \gamma_{\boldsymbol{n}}^{-1/2} > 0$ leading to the likelihood

$$\mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}; \gamma_{\boldsymbol{n}}) = (2\pi)^{-\frac{N}{2}} \gamma_{\boldsymbol{n}}^{\frac{N}{2}} \exp\left(-\frac{\gamma_{\boldsymbol{n}}}{2} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2\right). \quad (1)$$

We also consider *a priori* image law using Markov Random Field

$$\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{\theta}) = K(\boldsymbol{\theta})^{-1} \exp\left(-\frac{1}{2} E_{\boldsymbol{\theta}}(\boldsymbol{x})\right) \quad (2)$$

where $K(\boldsymbol{\theta}) = \int \exp(-E_{\boldsymbol{\theta}}(\boldsymbol{x})/2)\,\mathrm{d}\boldsymbol{x}$ is the partition function. In the general case, K is dependant wrt. regularization parameters $\boldsymbol{\theta}$ which is a major difficulty that blocks $\boldsymbol{\theta}$ estimation [16], [21], [22].

For the well-posedness of the inverse problem we consider the energy $E_{\boldsymbol{\theta}}(\boldsymbol{x}) = \gamma_{\boldsymbol{x}} \sum_{c \in \mathcal{C}} \phi\left(\boldsymbol{d}_c^t \boldsymbol{x}; \boldsymbol{\theta}\right)$ with the set $\mathcal{C}$ of cliques c and neighborhoods $\boldsymbol{d}_c$ [7] and $\gamma_{\boldsymbol{x}} > 0$. For edge preserving consideration and uniqueness of the supervised solution, the potential ϕ is chosen as convex.

A common estimator in image processing community is the maximum *a posteriori* (MAP) obtained by minimization of the anti logarithm of the *a posteriori* law

$$\boldsymbol{x}_{\text{MAP}} = \arg\min_{\boldsymbol{x}} - \ln(\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{y}, \gamma_{\boldsymbol{n}}, \boldsymbol{\theta})) \tag{3}$$

$$= \arg\min_{\boldsymbol{x}} \|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2 + \frac{\gamma_{\boldsymbol{x}}}{\gamma_{\boldsymbol{n}}} E_{\boldsymbol{\theta}}(\boldsymbol{x}). \tag{4}$$

In that case of convex optimization, an abundant literature is available to compute it efficiently [10], [24].

III. Bayesian Framework and Algorithm

A. LogErf potential and LMG

The LogErf potential ϕ [22], [25] is the convex function

$$\phi(u) = -2\ln(\chi(u) + \chi(-u)) \tag{5}$$

with

$$\chi(u) = \exp\Big(\frac{\gamma_{\boldsymbol{b}}}{2} u\Big) \operatorname{erfc}\Bigg(\sqrt{\frac{\gamma_{\boldsymbol{x}}}{2}} \left(\frac{\gamma_{\boldsymbol{b}}}{2\gamma_{\boldsymbol{x}}} + u\right)\Bigg) \tag{6}$$

with $\gamma_{\boldsymbol{x}} > 0$ and $\gamma_{\boldsymbol{b}} > 0$. This potential, liken to Huber potential, brings several advantages when used for Gibbs field:

1) ϕ is convex and, therefore, $E_{\boldsymbol{\theta}}(\boldsymbol{x})$ is convex wrt. $\boldsymbol{x}$.
2) The distribution $p(u)$ can be expressed as a Location Mixture of Gaussian (LMG)

$$\mathrm{p}(u) \propto \exp\Big(-\frac{1}{2}\phi(u)\Big) \propto \int_{\mathbb{R}} \mathrm{p}(u \mid b)\,\mathrm{p}(b)\,\mathrm{d}b \tag{7}$$

$$= P^{-1} \int_{\mathbb{R}} \exp\Big(-\frac{\gamma_{\boldsymbol{x}}}{2}(u - b)^2 - \frac{\gamma_{\boldsymbol{b}}}{2}|b|\Big)\,\mathrm{d}b \tag{8}$$

with a Laplace distribution $\mathrm{p}(b)$ on the mean of $\mathrm{p}(u \mid b)$.

3) The partition function is tractable and does not depend on b. For the LogErf potential, we have

$$P = \sqrt{2\pi}\gamma_{\boldsymbol{x}}^{-\frac{1}{2}} \left(\frac{\gamma_{\boldsymbol{b}}}{4}\right)^{-1}. \tag{9}$$

A Gibbs field on the image $\boldsymbol{x}$ can be constructed as

$$\mathrm{p}(\boldsymbol{x} \mid \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}) = K^{-1} \int \exp\Big(-\frac{\gamma_{\boldsymbol{x}}}{2}\|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{b}\|^2\Big) \prod_{c \in \mathcal{C}} \exp\Big(\frac{\gamma_{\boldsymbol{b}}}{2}|b_c|\Big)\,\mathrm{d}\boldsymbol{b} \tag{10}$$

that is, an LMG for each clique $\boldsymbol{d}_c^t \boldsymbol{x}$ associated with an auxiliary variable b_c. Initial work [22] states the necessity that the number of clique equals the number of pixels. The operator was circulant convolution with second order difference (Laplacian), leading to *a priori* independent clique in image and Fourier space

$$\mathrm{p}(\boldsymbol{x}) \propto \prod_{c=1}^{N} \int_{\mathbb{R}} \exp\Big(-\gamma_{\boldsymbol{x}} |\bar{x}_c - b_c|^2\Big) \exp\Big(\frac{\gamma_{\boldsymbol{b}}}{2}|b_c|\Big)\,\mathrm{d}b_c \tag{11}$$

where $\bar{\boldsymbol{x}} = \boldsymbol{D}\boldsymbol{x}$, and $\bar{x}_c$ the c-element of $\bar{\boldsymbol{x}}$. However, this model introduces crosslike artifacts around the edges [23]. Moreover, this limitation is not necessary and a field build with more clique and auxiliary variables than pixels is possible, at the price of no longer orthogonal cliques.

For image, we consider two group of cliques (horizontal and vertical gradient for instance) $\boldsymbol{D}_h$ and $\boldsymbol{D}_v$ with conditional mean $\boldsymbol{b}_h \in \mathbb{R}^N$ and $\boldsymbol{b}_v \in \mathbb{R}^N$

$$\mathrm{p}(\boldsymbol{x} \mid \boldsymbol{b}) \propto \exp\Big(-\frac{\gamma_{\boldsymbol{x}}}{2}\|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{b}\|^2\Big) \tag{12}$$

$$= (2\pi)^{-\frac{N}{2}} |\boldsymbol{Q}|^{\frac{1}{2}} \exp\Bigg(-\frac{(\boldsymbol{x} - \boldsymbol{\mu})^t \boldsymbol{Q} (\boldsymbol{x} - \boldsymbol{\mu})}{2}\Bigg) \tag{13}$$

with $\boldsymbol{D}^t = [\boldsymbol{D}_h^t, \boldsymbol{D}_v^t]$, $\boldsymbol{b}^t = [\boldsymbol{b}_h^t, \boldsymbol{b}_v^t]$, $\boldsymbol{Q} = \gamma_{\boldsymbol{x}}(\boldsymbol{D}_h^t \boldsymbol{D}_h + \boldsymbol{D}_v^t \boldsymbol{D}_v)$ and $\boldsymbol{\mu} = \boldsymbol{Q}^{-1}\boldsymbol{D}^t\boldsymbol{b}$, if $\boldsymbol{Q}$ is not singular. Contrary to [22], the distribution on $\boldsymbol{x}$ is not a product of independent distributions on each clique, or two *a priori* law on each clique group [23]. The prior mean $\boldsymbol{\mu}$ is therefore a mix of the auxiliary variables. Eq. (13) is sufficient statistics form but Eq. (12) is more natural where each clique have an unknown mean and, consequently, each clique has a LMG. The full marginal *a priori* law for the image writes

$$\mathrm{p}(\boldsymbol{x}) = \int \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{b}) \prod_{c=1}^{2N} \mathrm{p}(b_c)\,\mathrm{d}\boldsymbol{b}, \tag{14}$$

remains an LMG and, by choosing a Laplace distribution for b_c, the partition function is expressed wrt. hyper-parameters as

$$K(\gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}) \propto \gamma_{\boldsymbol{x}}^{-\frac{N}{2}} \gamma_{\boldsymbol{b}}^{-2N}. \tag{15}$$

Finally, the existence condition is not that the number of clique equals the number of pixels but that $\boldsymbol{Q}$ is not singular. Since the gradient operator leads to singular $\boldsymbol{Q}$ and improper prior with a null eigen-value for the mean level, we refer to papers [22] and [21] and the change to $K(\gamma_{\boldsymbol{x}}, .) \propto \gamma_{\boldsymbol{x}}^{-(N-1)/2}$.

B. Posterior Law and Estimators

A full joint *a posteriori* law of the unknown can now be expressed, with the horizontal and vertical gradient,

$$\mathrm{p}(\boldsymbol{x}, \boldsymbol{b}, \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}} \mid \boldsymbol{y}; \gamma_{\boldsymbol{n}}) \propto \mathrm{p}(\boldsymbol{y} \mid \boldsymbol{x}; \gamma_{\boldsymbol{n}}) \, \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{b}, \gamma_{\boldsymbol{x}})\,\mathrm{p}(\boldsymbol{b} \mid \gamma_{\boldsymbol{b}})\,\mathrm{p}(\gamma_{\boldsymbol{x}})\,\mathrm{p}(\gamma_{\boldsymbol{b}}). \tag{16}$$

The *a priori* laws is correlated Gaussian for $\boldsymbol{x}$, Laplacian for $\boldsymbol{b}$ and conjugate for $\gamma_{\boldsymbol{x}}$ and $\gamma_{\boldsymbol{b}}$, that is the non-informative Jeffrey's distribution [26]. The *a posteriori* law writes

$$\mathrm{p}(\boldsymbol{x}, \boldsymbol{b}, \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}} \mid \boldsymbol{y}; \gamma_{\boldsymbol{n}}) \propto \gamma_{\boldsymbol{x}}^{-\frac{(N-1)}{2}} \gamma_{\boldsymbol{b}}^{-2N} \exp\Big(-\frac{\gamma_{\boldsymbol{n}}}{2}\|\boldsymbol{y} - \boldsymbol{H}\boldsymbol{x}\|^2\Big) \exp\Big(-\frac{\gamma_{\boldsymbol{x}}}{2}\|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{b}\|^2\Big) \exp\Big(-\frac{\gamma_{\boldsymbol{b}}}{2}\|\boldsymbol{b}\|_1\Big). \tag{17}$$

Several estimators are studied:

1) The Supervised Expectation *a posteriori* (SEAP)

$$\boldsymbol{x}_{\text{SEAP}} = \int \boldsymbol{x}\, \mathrm{p}(\boldsymbol{x} \mid \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}, \boldsymbol{y})\, \mathrm{d}\boldsymbol{x} \tag{18}$$

$$= \iint \boldsymbol{x}\, \mathrm{p}(\boldsymbol{x}, \boldsymbol{b} \mid \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}, \boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{b}. \tag{19}$$

This estimator is the main reference for the proposed unsupervised EAP estimator. It allows finding the best hyper parameter values $\gamma_{\boldsymbol{x}}^{\star}$ and $\gamma_{\boldsymbol{b}}^{\star}$ given the true signal $\boldsymbol{x}^{\star}$ and a measure. Then the image

$$\boldsymbol{x}_{\text{SEAP}}^{\star} = \int \boldsymbol{x}\, \mathrm{p}(\boldsymbol{x} \mid \gamma_{\boldsymbol{x}}^{\star}, \gamma_{\boldsymbol{b}}^{\star}, \boldsymbol{y})\, \mathrm{d}\boldsymbol{x} \tag{20}$$

is considered as a reference, being the best possible reconstructed image given that model.

2) The supervised Maximum *a posteriori* (MAP)

$$\boldsymbol{x}_{\text{MAP}} = \arg\min_{\boldsymbol{x}} -\ln(\mathrm{p}(\boldsymbol{x} \mid \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}, \boldsymbol{y})) \tag{21}$$

$$= \arg\min_{\boldsymbol{x}} -\ln\left(\int \mathrm{p}(\boldsymbol{x}, \boldsymbol{b} \mid \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}, \boldsymbol{y})\, \mathrm{d}\boldsymbol{b}\right). \tag{22}$$

Champagnat *et. al.* shown in [25] that this estimator can be computed by an EM algorithm equivalent to the efficient HQ Geman & Yang optimization scheme. The MAP also allow to qualify and quantify the differences, if any, with the SEAP, and can also be used to determine best $\gamma_{\boldsymbol{x}}^{\star}$ and $\gamma_{\boldsymbol{b}}^{\star}$.

3) The semi-unsupervised Expectation *a posteriori* (EAP)

$$\boldsymbol{x}_{\text{EAP}} = \int \boldsymbol{x}\, \mathrm{p}(\boldsymbol{x} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{x} \tag{23}$$

$$= \iint \boldsymbol{x}\, \mathrm{p}(\boldsymbol{x}, \boldsymbol{b}, \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{x}\, \mathrm{d}\boldsymbol{b}\, \mathrm{d}\gamma_{\boldsymbol{x}}\, \mathrm{d}\gamma_{\boldsymbol{b}}. \tag{24}$$

The EAP estimator is the objective and will be compared to supervised reconstructions. The algorithm, presented Sec. III-C, also provides the EAP $\hat{\gamma}_{\boldsymbol{x}}$ and $\hat{\gamma}_{\boldsymbol{b}}$ that can be compared to $\gamma_{\boldsymbol{x}}^{\star}$ and $\gamma_{\boldsymbol{b}}^{\star}$.

4) The *unsupervised* Expectation *a posteriori* with quadratic prior, that is with prior model

$$\mathrm{p}(\boldsymbol{x} \mid \gamma_{\boldsymbol{x}}) \propto \exp\left(-\gamma_{\boldsymbol{x}} \|\boldsymbol{D}\boldsymbol{x}\|^2/2\right). \tag{25}$$

This estimator $\boldsymbol{x}_{\text{QUAD}}$ is known to produce near optimal hyper-parameter estimation when a quadratic penalization is used [21].

C. MCMC Algorithm for EAP and SEAP

The computation of the EAP estimator is based on a Gibbs sampler that successively simulate, after an initialization $k = 0$, $\boldsymbol{b}^{(0)} = \boldsymbol{0}$ and $\gamma_{\boldsymbol{x}}^{(0)} = \gamma_{\boldsymbol{b}}^{(0)} = 1$, the conditional *a posteriori* laws as

1) $\boldsymbol{x}^{(k+1)} \sim \mathrm{p}\left(\boldsymbol{x} \mid \boldsymbol{b}^{(k)}, \gamma_{\boldsymbol{x}}^{(k)}, \boldsymbol{y}\right)$,
2) $\boldsymbol{b}^{(k+1)} \sim \mathrm{p}\left(\boldsymbol{b} \mid \boldsymbol{x}^{(k+1)}, \gamma_{\boldsymbol{x}}^{(k)}, \gamma_{\boldsymbol{b}}^{(k)}\right)$,
3) $\gamma_{\boldsymbol{x}}^{(k+1)} \sim \mathrm{p}\left(\gamma_{\boldsymbol{x}} \mid \boldsymbol{x}^{(k+1)}, \boldsymbol{b}^{(k+1)}\right)$,
4) $\gamma_{\boldsymbol{b}}^{(k+1)} \sim \mathrm{p}\left(\gamma_{\boldsymbol{b}} \mid \boldsymbol{b}^{(k+1)}\right)$,
5) $k \leftarrow k + 1$.

For the SEAP estimator, step 3 and 4 are not undertaken as $\gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}$ are fixed, eventually to $\gamma_{\boldsymbol{x}}^{\star}$ and $\gamma_{\boldsymbol{b}}^{\star}$.

The conditional *a posteriori* law of $\boldsymbol{x}$ is Gaussian. Since all operators are circulant convolution, this law can be simulated very efficiently with diagonal matrix in Fourier space

$$\mathrm{p}\left(\mathring{\boldsymbol{x}} \mid \mathring{\boldsymbol{b}}, \gamma_{\boldsymbol{x}}, \boldsymbol{y}\right) \propto \exp\left(-\frac{\gamma_{\boldsymbol{n}}}{2}\|\mathring{\boldsymbol{y}} - \boldsymbol{\Lambda}_{\boldsymbol{H}}\mathring{\boldsymbol{x}}\|^2 - \frac{\gamma_{\boldsymbol{x}}}{2}\|\boldsymbol{\Lambda}_h\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{b}}_h\|^2 - \frac{\gamma_{\boldsymbol{x}}}{2}\|\boldsymbol{\Lambda}_v\mathring{\boldsymbol{x}} - \mathring{\boldsymbol{b}}_v\|^2\right). \tag{26}$$

A sample is $\boldsymbol{x}^{(k+1)} = \boldsymbol{F}^{\dagger}\mathring{\boldsymbol{x}}^{(k+1)}$ where

$$\mathring{\boldsymbol{x}}^{(k+1)} = \gamma_{\boldsymbol{n}}\boldsymbol{\Sigma}\boldsymbol{\Lambda}_{\boldsymbol{H}}^{\dagger}\boldsymbol{F}\boldsymbol{y} + \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{F}\boldsymbol{\epsilon} \tag{27}$$

with $\boldsymbol{\Sigma}^{-1} = \gamma_{\boldsymbol{n}}|\boldsymbol{\Lambda}_{\boldsymbol{H}}|^2 + \gamma_{\boldsymbol{x}}(|\boldsymbol{\Lambda}_h|^2 + |\boldsymbol{\Lambda}_v|^2)$, and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$.

The conditional *a posteriori* law of auxiliary variables $\boldsymbol{b}$ is more delicate but is *a posteriori* conditionally independent

$$\mathrm{p}(\boldsymbol{b} \mid \boldsymbol{x}, \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}) \propto \prod_{c=1}^{2N} \exp\left(-\frac{\gamma_{\boldsymbol{x}}}{2}|\bar{x}_c - b_c|^2 - \gamma_{\boldsymbol{b}}|b_c|\right) \tag{28}$$

where $\bar{\boldsymbol{x}} = \boldsymbol{D}\boldsymbol{x}$. The choice of the LogErf allows fast simulation by inversion of the cumulative density function (icdf) as described in [22].

However, practical use of the icdf shows that numerical error and floating point arithmetic overflow may arise, even if small fraction of $\boldsymbol{b}$ is concerned. In that case we propose a fallback on a random-walk Metropolis-Hastings (MH) step. This fallback algorithm rarely occurs, is for independent scalar variables, and is used only for the elements of $\boldsymbol{b}$ where the icdf has failed. In practice, the MH appears not to be difficult to tune, is efficient, and produce satisfactory results.

Lastly, the conditional *a posteriori* laws for hyper-parameters are Gamma distribution

$$\mathrm{p}(\gamma_{\boldsymbol{x}} \mid \boldsymbol{x}, \boldsymbol{b}) \propto \gamma_{\boldsymbol{x}}^{-\frac{(N-1)}{2}} \exp\left(-\gamma_{\boldsymbol{x}}\left(\frac{\|\boldsymbol{D}\boldsymbol{x} - \boldsymbol{b}\|^2}{2}\right)\right) \tag{29}$$

and

$$\mathrm{p}(\gamma_{\boldsymbol{b}} \mid \boldsymbol{b}) \propto \gamma_{\boldsymbol{b}}^{-2N} \exp\left(-\gamma_{\boldsymbol{b}}\left(\sum_{c=1}^{2N}\frac{|b_c|}{2}\right)\right). \tag{30}$$

These scalar distribution are simulated with standard toolbox available in Matlab or Numpy for instance.

The SEAP and EAP are approximated by the empirical mean of K samples, after discarding samples of the burn-in period,

$$\boldsymbol{x}_{\text{EAP}} \approx \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{x}^{(k)}, \qquad \boldsymbol{b}_{\text{EAP}} \approx \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{b}^{(k)}. \tag{31}$$

The pixel variance, diagonal of the posterior covariance, is also approximated by

$$\boldsymbol{\sigma}_{\text{EAP}}^2 = \mathrm{diag}(\boldsymbol{\Sigma}_{\boldsymbol{x},\text{EAP}}) \approx \frac{1}{K-1}\sum_{k=1}^{K}\left(\boldsymbol{x}^{(k)} - \boldsymbol{x}_{\text{EAP}}\right)^2. \tag{32}$$

IV. Results

The proposed method is tested on the "cameraman" image, a $N = 256 \times 256$ pixels image with strong discontinuities. The observation operator has a squared impulsionnal response (IR) of size 5×5 pixels. The convolution is done by filtering in frequencies space as $\boldsymbol{y} = \boldsymbol{F}^\dagger \boldsymbol{\Lambda_H} \boldsymbol{F} \boldsymbol{x}^\star + \boldsymbol{n}$, where $\boldsymbol{x}^\star$ stands for the true image. The tested noise inverse variance $\gamma_{\boldsymbol{n}}$ are 100, 1 and 0.1. The horizontal and vertical gradients are computed via the frequency space $\boldsymbol{x}_h = \boldsymbol{F}^\dagger \boldsymbol{\Lambda}_h \boldsymbol{F} \boldsymbol{x}$ and $\boldsymbol{x}_v = \boldsymbol{F}^\dagger \boldsymbol{\Lambda}_v \boldsymbol{F} \boldsymbol{x}$.

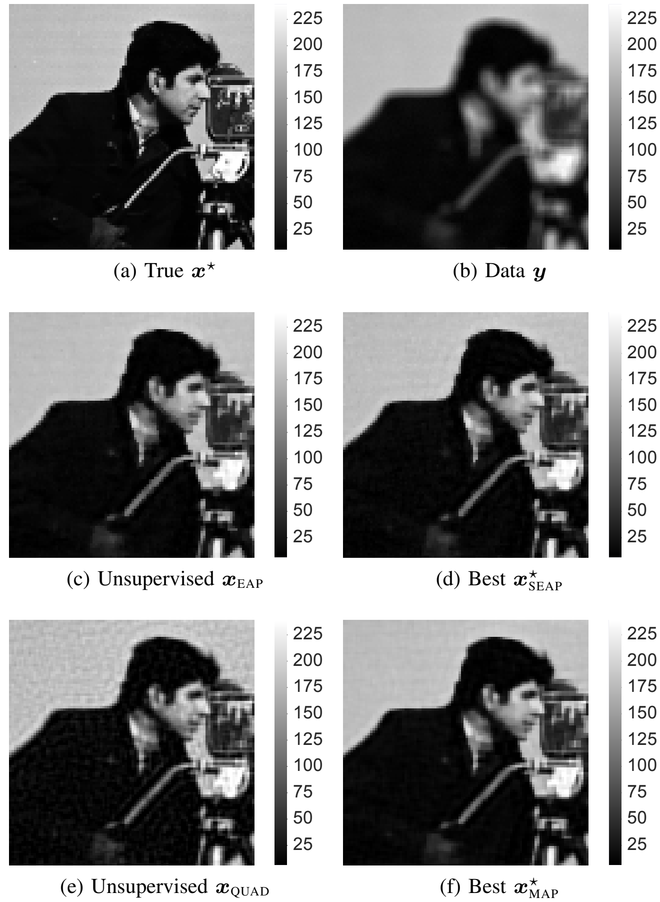
The MCMC algorithm produces 1000 samples $[\boldsymbol{x}, \boldsymbol{b}, \gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}]$, with a burn-in period of 500 samples. These numbers are chosen to have visually a sufficiently good exploration after the burn-in period. The algorithm, implemented with Python and Scipy has been run on a 1.9 GHz processor. The total time to produce 1000 samples is around 80 seconds, with 0.08 seconds per iteration. The majority of time (60%) is spent inside the icdf simulation and special functions evaluation.

Fig. 1 shows the true image $\boldsymbol{x}^\star$ and the data $\boldsymbol{y}$ with $\gamma_{\boldsymbol{n}} = 1$. The best Supervised Expectation *a posteriori* $\boldsymbol{x}_{\text{SEAP}}$ is illustrated in Fig. 1d, where the hyper-parameters $\gamma_{\boldsymbol{x}}^\star, \gamma_{\boldsymbol{b}}^\star$ are determined by minimization of the ℓ_1 error $\|\boldsymbol{x}_{\text{SEAP}}(\gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}) - \boldsymbol{x}^\star\|_1$. The minimization is done by exhaustive search to reach the near optimal SEAP solution. The Fig. 1c shows our proposition where hyper-parameters are automatically estimated from the data without hand-tuning and without knowledge of the true image, except the noise level.

For comparison, the unsupervised quadratic solution, $\boldsymbol{x}_{\text{QUAD}}$ Fig. 1e, has some noise residual, well visible inside flat region, and Gibbs effect is present near the edges. These defects are no more visible on the proposed solution $\boldsymbol{x}_{\text{EAP}}$, see Fig. 1c, as for the other convex solutions. To assess the good results of the semi-unsupervised EAP estimator, Fig. 1f shows the best supervised convex $\boldsymbol{x}_{\text{MAP}}^\star$ estimator, known to provide good results [24]. In that case, the hyper-parameters are also found via the minimization of the ℓ_1 error $\|\boldsymbol{x}_{\text{MAP}}(\gamma_{\boldsymbol{x}}, \gamma_{\boldsymbol{b}}) - \boldsymbol{x}^\star\|_1$. Almost no differences is visible by eyes between supervised $\boldsymbol{x}_{\text{SEAP}}^\star$, $\boldsymbol{x}_{\text{MAP}}^\star$ and our unsupervised proposition $\boldsymbol{x}_{\text{EAP}}$.

Thanks to the MCMC algorithm, the EAP estimator also provides the uncertainties for every quantities, notably the image $\boldsymbol{x}$ illustrated Fig. 2a, where uncertainty is concentrated near region of strong gradient. Fig 2b shows the Expectation *a posteriori* of $\boldsymbol{b}_v$. The values are naturally around zero and variables are clearly able to detect edge as awaited.

Chains of hyper-parameters are illustrated in Fig. 3. The $\gamma_{\boldsymbol{b}}$ chains converge in short time and present a small dispersion in regards to the $\gamma_{\boldsymbol{x}}$ chain. An explanation is the small variation of the auxiliary variables during the iteration. On the contrary, the $\gamma_{\boldsymbol{x}}$ chain presents a longer convergence time with more intra-correlation. The chain still converges after approximately 200 iterations. A possible explanation is the greater sensibility of prior adequacy wrt. the auxiliary variables changes but further investigations are needed. Finally, tests with more and less noise is also presented Fig. 4.

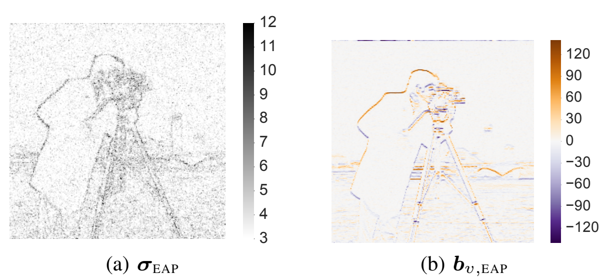


(a) True $\boldsymbol{x}^\star$ (b) Data $\boldsymbol{y}$

(c) Unsupervised $\boldsymbol{x}_{\text{EAP}}$ (d) Best $\boldsymbol{x}_{\text{SEAP}}^\star$

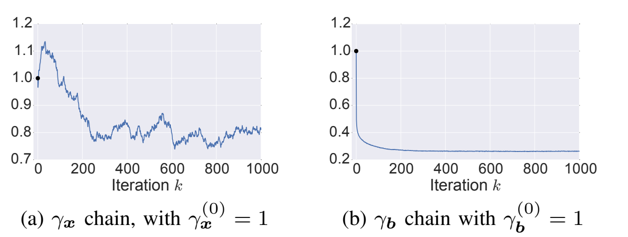
(e) Unsupervised $\boldsymbol{x}_{\text{QUAD}}$ (f) Best $\boldsymbol{x}_{\text{MAP}}^\star$

Fig. 1. Results with noise precision $\gamma_{\boldsymbol{n}} = 1$. The true image $\boldsymbol{x}^\star$ and data $\boldsymbol{y}$ are represented Fig. 1a and 1b. The Fig. 1d is the best SEAP solution and Fig. 1c the proposed unsupervised restoration. For comparison Fig. 1e shows the quadratic unsupervised solution and 1f the best supervised MAP.



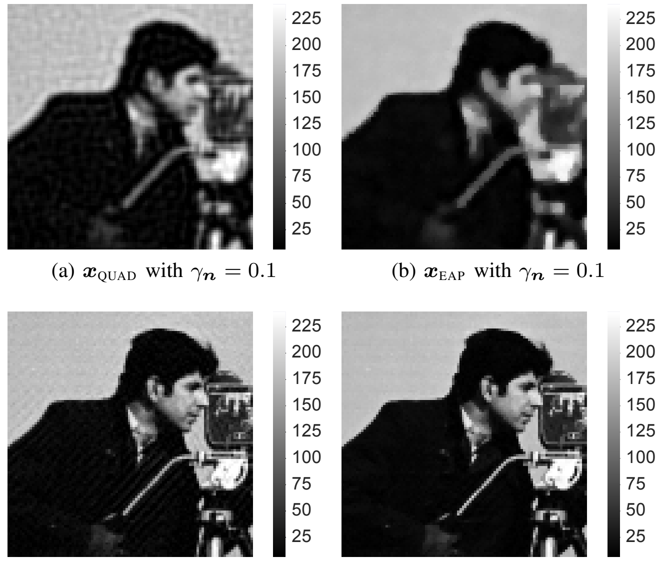
(a) $\boldsymbol{\sigma}_{\text{EAP}}$ (b) $\boldsymbol{b}_{v,\text{EAP}}$

Fig. 2. Results with noise precision $\gamma_{\boldsymbol{n}} = 1$. Fig. 2a shows the estimated standard deviation of each pixel. Fig. 2b shows the estimated vertical auxiliary variables. Both image are slightly saturated for display purpose.



(a) $\gamma_{\boldsymbol{x}}$ chain, with $\gamma_{\boldsymbol{x}}^{(0)} = 1$ (b) $\gamma_{\boldsymbol{b}}$ chain with $\gamma_{\boldsymbol{b}}^{(0)} = 1$

Fig. 3. Chains for noise precision $\gamma_{\boldsymbol{n}} = 1$. Fig. 3a shows the chain for the $\gamma_{\boldsymbol{x}}$ parameter and Fig. 3b shows the $\gamma_{\boldsymbol{b}}$ chain. Black dot is the initialization.



(a) $\boldsymbol{x}_{\text{QUAD}}$ with $\gamma_{\boldsymbol{n}} = 0.1$ (b) $\boldsymbol{x}_{\text{EAP}}$ with $\gamma_{\boldsymbol{n}} = 0.1$

(c) $\boldsymbol{x}_{\text{QUAD}}$ with $\gamma_{\boldsymbol{n}} = 100$ (d) $\boldsymbol{x}_{\text{EAP}}$ with $\gamma_{\boldsymbol{n}} = 100$

Fig. 4. Results with different noise precision $\gamma_{\boldsymbol{n}} = 0.1$ and 100. Quality and absence of Gibbs effect on the camera arm is notable.

V. Conclusion

This paper presents an incremental contribution to the unsupervised convex image restoration within a Bayesian approach based on [22]. However, the proposed approach allows to use several regularization operator for better image restoration avoiding crosslike effect. In addition, we propose a fallback algorithm based on a random walk Metropolis-Hastings step to avoid the numerical instability of the original algorithm. The noise level is supposed known and it's estimation jointly with the other is the main perspective.

References

[1] N. Chakrova, B. Rieger, and S. Stallinga, "Deconvolution methods for structured illumination microscopy," EN, *Journal of the Optical Society of America A*, vol. 33, no. 7, B12, 2016, ISSN: 1084-7529.

[2] J. P. Oliveira, M. a. T. Figueiredo, and J. M. Bioucas-Dias, "Parametric blur estimation for blind restoration of natural images: Linear motion and out-of-focus," *IEEE Transactions on Image Processing*, vol. 23, no. 1, pp. 466–477, 2014, ISSN: 10577149.

[3] E. Vera, M. Vega, R. Molina, and A. K. Katsaggelos, "Iterative image restoration using nonstationary priors," *Appl. Opt.*, vol. 52, no. 10, pp. D102–D110, 2013, ISSN: 1559128X.

[4] S. Bongard, F. Soulez, É. Thiébaut, and É. Pecontal, "3D deconvolution of hyper-spectral astronomical data," *Monthly Notices of the Royal Astronomical Society*, vol. 418, no. 1, pp. 258–270, 2011, ISSN: 00358711. arXiv: arXiv:1107.4049v1.

[5] F. Orieux, E. Sepulveda, V. Loriette, B. Dubertret, and J. C. Olivo-Marin, "Bayesian estimation for optimized structured illumination microscopy," *IEEE Transactions on Image Processing*, vol. 21, no. 2, pp. 601–614, 2012.

[6] J.-F. Giovannelli and A. Coulais, "Positive deconvolution for superimposed extended source and point sources," *Astronomy & Astrophysics*, vol. 412, pp. 401–412, 2005.

[7] J. Idier, Ed., *Bayesian Approach to Inverse Problems*. ISTE Ltd and John Wiley & Sons Inc., 2008.

[8] N. Komodakis and J.-C. Pesquet, "Playing with Duality: An Overview of Recent Primal-Dual Approaches for Solving Large-Scale Optimization Problems," *arXiv.org*, vol. cs.NA, 2014. arXiv: 1406.5429.

[9] A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems.," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2419–34, 2009, ISSN: 1941-0042.

[10] S. Boyd and L. Vandenberghe, *Convex Optimization*, 3. Cambridge University Press, 2010, vol. 25, ch. 1,10,11, pp. 487–487, ISBN: 9780521833783. arXiv: 1111.6189v1.

[11] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia, "Hierarchical Multiple Markov Chain Model for Unsupervised Texture Segmentation," *IEEE Transactions on Image Processing*, vol. 18, no. 8, pp. 1830–1843, 2009.

[12] Y. Zheng, A. Fraysse, and T. Rodet, "Efficient Variational Bayesian Approximation Method Based on Subspace optimization," *IEEE Trans. on Image Processing*, vol. 24, no. 2, pp. 681–693, 2015.

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

[14] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Variational Bayesian Super Resolution," *IEEE Transactions on Image Processing*, vol. 20, no. 4, pp. 984–999, 2011.

[15] 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 Journal of Selected Topics in Signal Processing*, vol. 10, no. 2, pp. 224–241, 2016, ISSN: 1932-4553.

[16] X. Descombes, R. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov Random Field prior parameters using Markov Chain Monte Carlo Maximum Likelihood," *IEEE Transactions on Image Processing*, vol. 8, no. 7, pp. 954–963, 1999.

[17] C. Gilavert, S. Moussaoui, and J. Idier, "Efficient Gaussian Sampling for Solving Large-Scale Inverse Problems Using MCMC," *IEEE Transactions on Signal Processing*, vol. 63, no. 1, pp. 70–80, 2015.

[18] F. Orieux, O. Féron, and J.-F. Giovannelli, "Sampling high-dimensional Gaussian distributions for general linear inverse problems," *IEEE Signal Processing Letters*, vol. 19, no. 5, pp. 251–254, 2012, ISSN: 10709908.

[19] L. Chaari, J.-C. Pesquet, J.-Y. Tourneret, P. Ciuciu, and A. Benazza-Benyahia, "A Hierarchical Bayesian Model for Frame Representation," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5560–5571, 2010, ISSN: 1053-587X.

[20] M. Pereyra, *Maximum-a-posteriori estimation with Bayesian confidence regions*, 2016. arXiv: 1602.08590.

[21] F. Orieux, J.-F. Giovannelli, and T. Rodet, "Bayesian estimation of regularization and psf parameters for wiener-hunt deconvolution," *Journal of the Optical Society of America A*, vol. 27, no. 7, pp. 1593–1607, 2010.

[22] J.-F. Giovannelli, "Unsupervised Bayesian Convex Deconvolution Based on a Field With an Explicit Partition Function," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 16–26, 2008, ISSN: 1057-7149.

[23] R. G. Chinchilla, "Efficiently Sampling from High-Dimensional Mixture of Gaussians Distributions and Applications to Inverse Problems," Master Thesis, Université Paris-Saclay, 2016.

[24] J. Idier, "Convex Half-Quadratic Criteria and Interacting Auxiliary Variables for Image Restoration," *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1001–1009, 2001.

[25] F. Champagnat and J. Idier, "A connection between half-quadratic criteria and EM algorithms," *IEEE Signal Processing Letters*, vol. 11, no. 9, pp. 709–712, 2004, ISSN: 10709908.

[26] M. West, "Outlier Models and Prior Distributions in Bayesian Linear Regression," *Journal of the Royal Statistical Society Series B Methodological*, vol. 46, no. 3, pp. 431–439, 1984, ISSN: 00359246.