Parameter Estimation in Sparse Inverse Problems Using Bernoulli-Gaussian Prior

Pierre Barbault, Matthieu Kowalski, Charles Soussen

PARAMETER ESTIMATION IN SPARSE INVERSE PROBLEMS USING BERNOULLI-GAUSSIAN PRIOR

Pierre Barbault, Matthieu Kowalski, Charles Soussen

L2S, University Paris-Saclay, CNRS, CentraleSupélec

ABSTRACT

Sparse coding is now one of the state-of-art approaches for solving inverse problems. In combination with (Fast) Iterative Shrinkage Thresholding Algorithm (ISTA), among other algorithms, one can efficiently get a nice estimator of the sought sparse signal. However, the major drawback of these methods is the tuning of the so-called hyperparameter. In this paper, we first provide an Expectation-Maximization (EM) algorithm to estimate the parameters of a Bernoulli-Gaussian model for denoising a sparse signal corrupted by a white Gaussian noise. Then, building on the Expectation-Maximization interpretation of ISTA, we provide a simple iterative algorithm to blindly estimate all the model parameters in the linear inverse problem context, including the hyperparameter involved in the popular ℓ_0 regularized minimization. Moreover, the algorithm directly yields an estimator of the sparse signal.

Index Terms— Sparse coding, Inverse Problem, ISTA, Bernoulli-Gaussian model

1. INTRODUCTION

Sparse coding is now one of the state-of-the-art approaches for linear inverse problems [1, 2] of the form

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e} \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^{N_\mathbf{y}}$ is the observed/measured signal, $\mathbf{x} \in \mathbb{R}^{N_\mathbf{x}}$ is the original signal, $\mathbf{H} \in \mathbb{R}^{N_\mathbf{y} \times N_\mathbf{x}}$ is a linear operator, and $\mathbf{e} \in \mathbb{R}^{N_\mathbf{y}}$ is some noise supposed to be white and Gaussian. Usually, the signal $\mathbf{x}$ is supposed to admit a sparse representation in a chosen, or learned, dictionary of atoms [3, 4]. Without loss of generality, we suppose here that the signal $\mathbf{x}$ is sparse.

A popular approach to estimate sparse signals relies on the minimization of

$$\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \lambda \mathcal{R}(\mathbf{x}) \tag{2}$$

where $\|\,.\,\|$ refers to the ℓ_2 norm, $\mathcal{R}$ is a regularization term enforcing sparsity, such as ℓ_1 or ℓ_0 (quasi)-norms, and $\lambda > 0$ is a fixed hyperparameter. Solving (2) can be efficiently done using proximal descent methods [5] and more particularly the (Fast) Iterative Shrinkage/Thresholding (FISTA) [6] and Iterative Hard Thresholding (IHT) [7] algorithms for the ℓ_1 and ℓ_0 problems, respectively.

One of the main drawbacks of the optimization approach (2), is the tuning of the hyperparameter λ. In the pure denoising case, one can use the Stein Unbiased Risk Estimator (SURE) for choosing the λ with the ℓ_1 regularization [8]. In [9], the authors proposed the Generalized-SURE (GSURE) to extend the SURE approach to inverse problems solving with ℓ_1 regularization. For the ℓ_0 problem, the SURE approach no longer applies, and the SCORE approach has been proposed [10] in the denoising case.

An alternative is the Bayesian framework, where sparsity is modeled through a suitable prior. A popular choice is the Bernoulli-Gaussian (BG) model [11], which is strongly related to the ℓ_0 regularized problem through the maximum *a posteriori* estimator [12]. In [11], the authors proposed a stochastic-EM procedure to estimate the sparse solution of a deconvolution problem. Indeed, a classical EM procedure appears intractable because of the combinatorial aspect due to the BG prior. Following the EM approach developed in [13] which relies on an additional latent variable, [14] proposed an iterative procedure to estimate the parameters of a BG model when the matrix $\mathbf{H}$ is a union of unitary dictionaries. However, with the chosen prior on the hyperparameters, this approach leads to a non-convergent estimate even in the denoising case as shown in [15]. Moreover, preliminary experiments show that the method is not robust for the general inverse problem. Finally, MCMC approaches for the BG sparse model do not scale well in practice [16] although recent efforts have been carried out to improve their efficiency [17].

This article deals with the unsupervised estimation of the BG parameters. In the denoising case, we first develop an EM algorithm to estimate all the parameters of the Bernoulli model in section 2. Then, in section 3, we revisit the approach [13], using an additional latent variable, to extend the EM estimation in the linear inverse problem context. Furthermore, we propose an iterative algorithm similar to ISTA that estimates the model parameters and provides an estimation of the sparse signal $\mathbf{x}$. These approaches are evaluated on both simulated and real data in section 4.

2. BG PRIOR AND DENOISING

We first study the denoising case, *i.e.* $\mathbf{H} = \mathbf{I}_N$ with $N = N_\mathbf{y} = N_\mathbf{x}$:

$$\mathbf{y} = \mathbf{x} + \mathbf{e}\,. \tag{3}$$

We assume a BG prior on $\mathbf{x}$ denoted by $\mathbf{x} \sim \mathcal{BN}(p, 0, \sigma_x^2 \mathbf{I}_N)$. That is, for each coordinate n of $\mathbf{x}$, we have

$$p(x_n | p, \sigma_\mathbf{x}^2) = (1-p)\delta(x_n) + p\frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{x_n^2}{2\sigma_x^2}} \tag{4}$$

where δ stands for the Dirac delta function. The parameter $p \in (0, 1)$ is directly related to the level of sparsity, *i.e.*, the rate of non-zero components in vector $\mathbf{x}$ – the smaller, the sparser – while parameter σ_x^2 is related to the range of the nonzero components. The noise is assumed to be white and Gaussian:

$$\mathbf{e} \sim \mathcal{N}(0, \sigma_e^2 \mathbf{I}_N)\,. \tag{5}$$

The goal is to estimate the model parameters $\theta = (\sigma_e^2, p, \sigma_x^2)$ by maximizing the likelihood of data $\mathbf{y}$, in an empirical Bayes fashion.

This work was supported by the French National Agency for Research through the BMWs project (ANR-20-CE45-0018).

Using $\mathbf{x}$ as a hidden variable, we can derive an EM algorithm:

$$\theta^{(t+1)} = \underset{\theta}{\operatorname{argmin}}\, \mathbb{E}_{\mathbf{x}|\mathbf{y},\theta^{(t)}}\left[\mathcal{L}(\mathbf{y},\mathbf{x}|\theta)\right]. \tag{6}$$

where $\mathcal{L}(\mathbf{y},\mathbf{x}|\theta) = -\log p(\mathbf{y},\mathbf{x}|\theta)$. The following proposition summarizes the E-step of the algorithm. For the sake of conciseness, we denote by $\mathcal{N}_x\left(\mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$.

Proposition 2.1 (E-step for denoising)**.** *Consider the BG denoising model* (3)–(5) *and denote by* $\theta^{(t)} = \{(\sigma_e^2)^{(t)}, p^{(t)}, (\sigma_x^2)^{(t)}\}$ *the estimated parameters at iteration* t*. Furthermore, define*

$$\phi_n^{(t)} = \frac{p^{(t)}\mathcal{N}_{y_n}(0,(\sigma_x^2)^{(t)}+(\sigma_e^2)^{(t)})}{p^{(t)}\mathcal{N}_{y_n}(0,(\sigma_x^2)^{(t)}+(\sigma_e^2)^{(t)}) + (1-p^{(t)})\mathcal{N}_{y_n}(0,(\sigma_e^2)^{(t)})}$$

with

$$\mu^{(t)} = \frac{(\sigma_x^2)^{(t)}}{(\sigma_x^2)^{(t)}+(\sigma_e^2)^{(t)}},\;\; \nu^{(t)} = \frac{(\sigma_x^2)^{(t)}(\sigma_e^2)^{(t)}}{(\sigma_x^2)^{(t)}+(\sigma_e^2)^{(t)}}. \tag{7}$$

Then, the expectation term within (6) *reads:*

$$\begin{aligned}\mathbb{E}_{\mathbf{x}|\mathbf{y},\theta^{(t)}}\left[\mathcal{L}(\mathbf{y},\mathbf{x}|\theta)\right] &= C + \frac{1}{2\sigma_e^2}\sum_{n=1}^{N} y_n^2(1-\phi_n^{(t)}) \\ &+ N\log\left(\frac{\sigma_e}{1-p}\right) + \frac{1}{2\sigma_e^2}\sum_{n=1}^{N}\left((\mu^{(t)}-1)^2y_n^2+\nu^{(t)}\right)\phi_n^{(t)} \\ &+ \frac{1}{2\sigma_x^2}\sum_{n=1}^{N}\left((\mu^{(t)}y_n)^2+\nu^{(t)}\right)\phi_n^{(t)} + \log\left(\frac{1-p}{p}\sigma_x\right)\sum_{n=1}^{N}\phi_n^{(t)}\end{aligned} \tag{8}$$

where C *is some constant which does not depend on* θ*.*

Sketch of proof. Thanks to the componentwise independence of the prior and likelihood functions, the expectation reduces to a sum of coordinate-wise operations:

$$\mathbb{E}_{\mathbf{x}|\mathbf{y},\theta^{(t)}}\left[\mathcal{L}(\mathbf{y},\mathbf{x}|\theta)\right] = \sum_{n=1}^{N}\mathbb{E}_{x_n|y_n,\theta^{(t)}}\left[\mathcal{L}(y_n,x_n|\theta)\right].$$

The coordinate neg-log-likelihood is further developed into:

$$\mathcal{L}(y_n,x_n|\theta) = \frac{(y_n-x_n)^2}{2\sigma_e^2} + \frac{1}{2}\log[\sigma_e^2] - \log p(x_n|\theta) + C$$

Let us now calculate the law of $x_n|y_n,\theta^{(t)}$. To do so, we apply Bayes' rule with the prior distribution (4). Up to a few rearrangements (to isolate the factors depending on x_n), we get:

$$\begin{aligned}p(x_n|y_n,\theta^{(t)}) &= (1-p^{(t)})\delta(x_n)\frac{\mathcal{N}_{y_n}(0,(\sigma_e^2)^{(t)})}{p(y_n|\theta^{(t)})} \\ &+ p^{(t)}\mathcal{N}_{x_n}\left(\mu^{(t)}y_n,\nu^{(t)}\right)\frac{\mathcal{N}_{y_n}(0,(\sigma_x^2)^{(t)}+(\sigma_e^2)^{(t)})}{p(y_n|\theta^{(t)})}\end{aligned}$$

with $\mu^{(t)}$ and $\nu^{(t)}$ given by (7), which rereads:

$$p(x_n|y_n,\theta^{(t)}) = (1-\phi_n^{(t)})\,\delta(x_n) + \phi_n^{(t)}\,\mathcal{N}_{x_n}\left(\mu^{(t)}y_n,\nu^{(t)}\right).$$

It follows that

$$\mathbb{E}_{x_n|y_n,\theta^{(t)}}\left[x_n\right] = \mu^{(t)}y_n\phi_n^{(t)} \tag{9}$$

$$\mathbb{E}_{x_n|y_n,\theta^{(t)}}\left[x_n^2\right] = \left((\mu^{(t)}y_n)^2+\nu^{(t)}\right)\phi_n^{(t)} \tag{10}$$

Finally, using that

$$\begin{cases}-\log p(x_n|x_n\neq 0,\theta) = \frac{1}{2}\log[\sigma_x^2] + \frac{x_n^2}{2\sigma_x^2} - \log[p] + C\\ -\log p(x_n=0|\theta) = -\log[1-p]\end{cases}$$

one gets

$$\begin{aligned}\mathbb{E}_{x_n|y_n,\theta^{(t)}}\left[-\log p(x_n|\theta)\right] &= \log\left(\frac{1-p}{p}\sigma_x\right)\phi_1^{(t)} + \\ &\frac{(\mu^{(t)}y_n)^2+\nu^{(t)}}{2\sigma_x^2}\phi_1^{(t)} - \log(1-p) + C.\end{aligned} \tag{11}$$

∎

To complete the EM algorithm, it remains to minimize the negative log-likelihood. By setting the derivative to zero w.r.t. θ of (8), we get a closed-form expression of the EM summarized now.

Proposition 2.2 (M-step for denoising)**.** *Consider the BG denoising model* (3)–(5)*. Then, the EM solution* $\theta^{(t+1)}$ *at iteration* $t+1$ *is given by*

$$\begin{aligned}p^{(t+1)} &= \frac{1}{N}\sum_{n=1}^{N}\phi_n^{(t)},\; (\sigma_x^2)^{(t+1)} = \nu^{(t)} + \frac{(\mu^{(t)})^2}{Np^{(t+1)}}\sum_{n=1}^{N}y_n^2\phi_n^{(t)} \\ (\sigma_e^2)^{(t+1)} &= \frac{1}{N}\sum_{n=1}^{N}y_n^2 - \frac{2\mu^{(t)}}{N}\sum_{n=1}^{N}y_n^2\phi_n^{(t)} + p^{(t+1)}(\sigma_x^2)^{(t+1)}\end{aligned}$$

Once all the θ parameters are estimated using the EM procedure, one can choose any estimator for $\mathbf{x}$. In the following, we will be explicitly interested in the two estimates studied in [18] for denoising, that is, the so-called *marginal MAP* (MMAP) estimator and the posterior mean estimate. The MMAP estimator consists of estimating first the support of the sparse signal by marginalizing out the nonzero coefficients of x. It reads [18]:

$$x_n^{MMAP} = \begin{cases}\frac{\sigma_x^2}{\sigma_x^2+\sigma_e^2}y_n & \text{if } |y_n| > \lambda^{MMAP}\\ 0 & \text{otherwise}\end{cases} \tag{12}$$

with

$$\lambda^{MMAP} = \sqrt{2\sigma_e^2\frac{\sigma_e^2+\sigma_x^2}{\sigma_x^2}\log\left(\frac{1-p}{p}\sqrt{1+\frac{\sigma_x^2}{\sigma_e^2}}\right)}. \tag{13}$$

The posterior mean estimator is given by Eq.(11). One can notice that this is generally a non-sparse signal.

3. EXTENSION TO INVERSE PROBLEMS

As stressed in [11], the EM approach with BG prior is intractable for general inverse problems. Following Figueiredo and Nowak's approach [13], we introduce a latent variable $\mathbf{z}\in\mathbb{R}^{N_\mathbf{x}}$ such that

$$\mathbf{y} = \mathbf{H}\mathbf{z} + \mathbf{n},\; \mathbf{z} = \mathbf{x} + \mathbf{b} \tag{14}$$

where $\mathbf{b}\sim\mathcal{N}(0,\sigma_b^2\mathbf{I}_{N_\mathbf{x}})$ and $\mathbf{n}\sim\mathcal{N}(0,\sigma_e^2\mathbf{I}_{N_\mathbf{y}} - \sigma_b^2\mathbf{H}\mathbf{H}^T)$ are independent random vectors, with $\sigma_b^2 \leq \frac{\sigma_e^2}{\|\mathbf{H}\mathbf{H}^T\|}$ to ensure a non-degenerate distribution. In [13], $\mathbf{z}$ was used as the hidden variable in an EM algorithm, which turns on to the popular ISTA, to estimate the sparse signal $\mathbf{x}$: the E-step reduces to the gradient descent on the ℓ_2 data-fidelity term while the M-step reduces to the proximal operator associated with the regularizer. Here, we use $\mathbf{z}$ as a latent variable to be estimated (as in Sparse Bayesian Learning [19]) and consider $\mathbf{x}$ as the hidden variable that is marginalized out. Let $\alpha \leq 1$ be a fixed parameter such that $\frac{\sigma_b^2}{\sigma_e^2} = \frac{\alpha}{\|\mathbf{H}\mathbf{H}^T\|}$, and let $\theta = (\sigma_b^2, \sigma_x^2, p)$. Using $\mathbf{x}$ as the hidden variable, the EM approach reads

$$\min_{\mathbf{z},\theta,\sigma_e^2} \mathbb{E}_{\mathbf{x}|\mathbf{y},\mathbf{z}^{(t)},\theta^{(t)}} \left\{-\log p(\mathbf{y},\mathbf{x},\mathbf{z}|\theta,\sigma_e^2)\right\}$$

$$\text{s.t. } \sigma_e^2 = \frac{\|\mathbf{H}\mathbf{H}^T\|}{\alpha}(\sigma_b^2)^{(t)} \tag{15}$$

Since $p(\mathbf{y},\mathbf{x},\mathbf{z}|\theta) = p(\mathbf{y}|\mathbf{z},\theta)p(\mathbf{z},\mathbf{x}|\theta)$, the latter reads:

$$\min_{\mathbf{z}} \left\{-\log p(\mathbf{y}|\mathbf{z},(\sigma_e^2)^{(t)}) + \min_{\theta}(\mathbb{E}_{\mathbf{x}|\mathbf{z}^{(t)},\theta^{(t)}}\mathcal{L}(\mathbf{z},\mathbf{x}|\theta))\right\}$$

where $(\sigma_e^2)^{(t)} = \frac{\|\mathbf{H}\mathbf{H}^T\|}{\alpha}(\sigma_b^2)^{(t)}$ and $\min_{\theta}(\mathbb{E}_{\mathbf{x}|\mathbf{z}^{(t)},\theta^{(t)}}\mathcal{L}(\mathbf{z},\mathbf{x}|\theta))$ is the denoising problem in $\mathbf{z}^{(t)}$ treated in the previous section 2. Then, denoting by $\theta^{(t+1)}$ the current estimate of θ, the minimization step with respect to $\mathbf{z}$ becomes

$$\min_{\mathbf{z}} \mathbb{E}_{\mathbf{x}|\mathbf{z}^{(t)},\theta^{(t)}} \left\{-\log p(\mathbf{y},\mathbf{x},\mathbf{z}|\theta^{(t+1)})\right\} = \tag{16}$$

$$\min_{\mathbf{z}} -\log p(\mathbf{y}|\mathbf{z},\sigma_e^{2(t+1)}) + \mathbb{E}_{\mathbf{x}|\mathbf{z}^{(t)},\theta^{(t)}} \left\{\frac{1}{2\sigma_b^{2(t+1)}}\|\mathbf{z}-\mathbf{x}\|^2\right\} + C$$

where C is some constant which does not depend on $\mathbf{z}$. Let $\hat{\mathbf{x}}^{(t)} = \mathbb{E}_{\mathbf{x}|\mathbf{z}^{(t)},\theta^{(t)}}\{\mathbf{x}\}$ be the posterior mean estimate of the BG denoising problem with respect to $\mathbf{z}^{(t)}$ given by (11). Using the linearity of the expectation, the minimization with respect to $\mathbf{z}$ reads

$$\min_{\mathbf{z}} -\log p(\mathbf{y}|\mathbf{z},\sigma_e^{2(t+1)}) + \frac{1}{2\sigma_b^{2(t+1)}}\left\{\|\mathbf{z}\|^2 - 2\mathbf{z}^T\hat{\mathbf{x}}^{(t)}\right\} \tag{17}$$

$$= \min_{\mathbf{z}} -\log p(\mathbf{y}|\mathbf{z},\sigma_e^{2(t+1)}) + \frac{1}{2\sigma_b^{2(t+1)}}\left\{\|\mathbf{z}-\hat{\mathbf{x}}^{(t)}\|^2\right\}$$

$$= \min_{\mathbf{z}} -\log\left\{p(\mathbf{y}|\mathbf{z},\sigma_e^{2(t+1)})p(\mathbf{z}|\hat{\mathbf{x}}^{(t)},\theta^{(t+1)})\right\} \tag{18}$$

$$= \min_{\mathbf{z}} -\log p(\mathbf{z}|\mathbf{y},\hat{\mathbf{x}}^{(t)},\sigma_e^{2(t+1)},\theta^{(t+1)}) \tag{19}$$

$p(\mathbf{z}|\mathbf{y},\hat{\mathbf{x}}^{(t)},\sigma_e^2,\theta^{(t+1)})$ being Gaussian, the minimizer, denoted by $\mathbf{z}^{(t+1)}$, reads [20]

$$\mathbf{z}^{(t+1)} = \hat{\mathbf{x}}^{(t)} + \frac{\alpha}{\|\mathbf{H}\mathbf{H}^T\|}\mathbf{H}^T\left(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}^{(t)}\right) \tag{20}$$

Algorithm 1 summarizes this EM procedure.

Algorithm 1: EM algorithm for Inverse Problem with BG prior

Result: Estimation of the model parameters.
$t = 0, \mathbf{z}^{(t)} = \mathbf{0}, \mathbf{x}^{(t)} = \mathbf{0}$;
while $t < N_t$ **do**
  $\mathbf{z}^{(t+1)} = \mathbf{x}^{(t)} + \frac{\alpha}{\|\mathbf{H}\mathbf{H}^T\|}\mathbf{H}^T\left(\mathbf{y}-\mathbf{H}\mathbf{x}^{(t)}\right)$;
  Estimate $\theta^{(t+1)}$ using prop.2.2 with $\mathbf{y} = \mathbf{z}^{(t)}$;
  $\mathbf{x}^{(t+1)} = \mathbb{E}_{\mathbf{x}|\mathbf{z}^{(t+1)},\theta^{(t+1)}}\{\mathbf{x}\}$;
  $t = t + 1$;
end

The main purpose of Algorithm 1 is to provide an estimation of the parameters of the model, that is, the parameters p and σ_x^2 of the BG prior as well as the variance $\sigma_e^2 = \sigma_b^2\frac{\|\mathbf{H}\mathbf{H}^T\|}{\alpha}$ of the noise. It turns out that by construction, the algorithm provides an estimate for $\mathbf{x}$, which is the posterior mean of the denoising problem for $\mathbf{z}$. Furthermore, one can also compute the threshold associated with the MMAP estimator in the denoising step related to $\mathbf{z}^{(t)}$, that is λ^{MMAP} as given in (13). The following numerical experiments show that these values can be used to tune the hyperparameter λ in the classical ℓ_0 minimization.

4. NUMERICAL EXPERIMENTS

Evaluation is carried out for denoising and deconvolution of signals and images. To this aim, we simulate BG sparse signals as follows. For an arbitrary sparsity level $p \in (0,1)$, we set $\sigma_x^2 = 1$, and σ_e^2 is chosen according to the desired input Signal-to-Noise Ratio (iSNR $= 10\log_{10}\left(\frac{\|\mathbf{H}\mathbf{x}\|^2}{\|\mathbf{e}\|^2}\right)$). We generate 100 realizations of a vector $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{e}$ of size 40 000. We also tested our approach on the fishing boat test image of size 512×512. We used a Daubechies-4 orthogonal wavelet transform to get sparse coefficients. We initialize the parameters to be estimated by EM using the method of moments, and the sparse coefficients by zeros. We have fixed $\alpha = 1$. The comparisons are mainly based on the output SNR (oSNR $= 10\log_{10}\left(\frac{\|\mathbf{x}\|^2}{\|\mathbf{x}-\hat{\mathbf{x}}\|^2}\right)$, where $\hat{\mathbf{x}}$ is the current estimate of $\mathbf{x}$). The experiments are run with Python on a laptop with an Intel Core i7 CPU at 1.8 GHz with 16 GB of RAM.

4.1. Denoising

Fig. 1 illustrates the ability of the EM procedure, given by Prop. 2.2, to recover the parameters of simulated signals in the case of a denoising problem. We display the relative error on a given parameters, *i.e.* $\frac{\hat{\theta}_i - \theta_i}{\theta_i}$, denoting by $\hat{\theta}_i$ the estimated value.

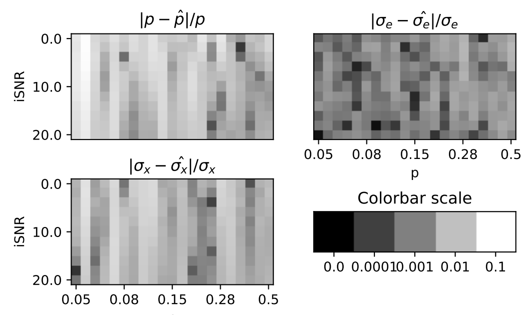

Fig. 1. Recovery of the parameters on noisy simulated signals.

Fig. 2 deals with the denoising problem of simulated signals with $p = 0.01$. The proposed approach is compared with denoising by hard thresholding, *i.e.* the set of ℓ_0 solutions obtained for a wide range of hyperparameter λ, see (2). The parameter estimates provided by our procedure turn out to be very accurate. In turn, our unsupervised approach yields a posterior mean estimate that reaches the best oSNR.

We can see in Fig. 3 that for image denoising, the proposed unsupervised posterior-mean estimator remains very competitive. Here, the best oSNR is reached by the hard-thresholding algorithm with specific values of λ.

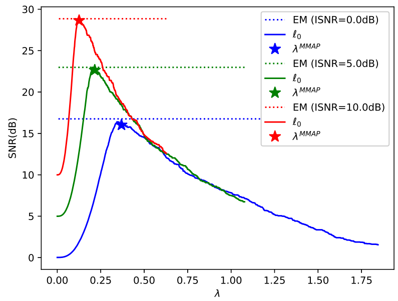

Fig. 2. Comparison of the EM algorithm for denoising simulated BG sparse signal (star shape point) and hard thresholding results (plain curves, showing results for various λ) for three iSNR.

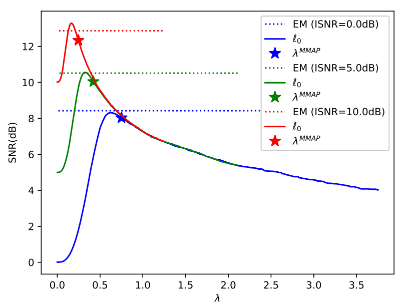

Fig. 3. Comparison of the EM algorithm for image denoising and hard thresholding for denoising an image.

4.2. Deconvolution

In the following experiments, the operator $\mathbf{H}$ corresponds to a convolution operator (which behaves like a low-pass filter) combined to the orthogonal wavelet transform. The normalized convolution kernel is $K(px, py) = \int_{px-0.5}^{px+0.5} \int_{py-0.5}^{py+0.5} \mathcal{N}_x(0, \eta) \mathcal{N}_y(0, \eta)\, dxdy$, where η controls the width of the kernel. In this experiment, BG signals are drawn with $p = 0.01$. We evaluated the ability of the algorithm to recover the parameters for various η. Fig. 4 shows the relative error for each parameters. We can stress that the parameters are recovered precisely whenever $\eta \leq 0.3$.

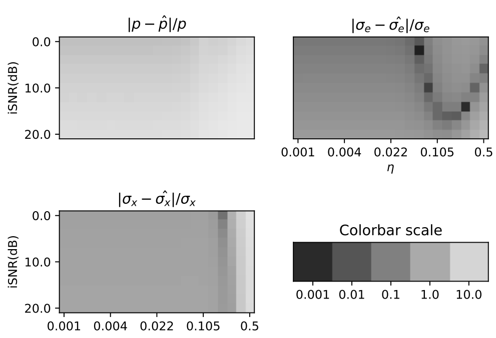

Fig. 4. Recovery of the parameters on blurry simulated signals.

Fig. 5 compares the proposed approach with ℓ_0 minimization, performed using the IHT algorithm with warm restarts [21]. When the kernel becomes wide, the oSNR given by the MMAP value of λ for the ℓ_0 minimization outperforms the posterior-mean estimate. The unsupervised value of λ provided by our approach is closed to the best possible oSNR achieved by the ℓ_0 minimization.

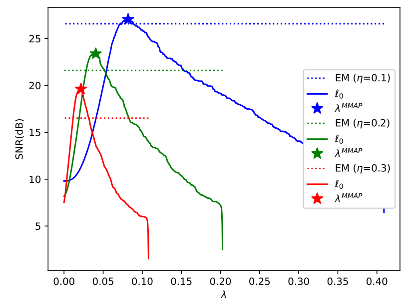

Fig. 5. Comparison of Alg. 1 for deconvolution of simulated BG sparse signal and ℓ_0 minimization,with a fixed iSNR $= 10$ dB.

For image deconvolution, the posterior-mean estimate provides excellent empirical results for small kernels. Its performances collapse for wide convolution kernels. However, the MMAP value of λ appears to be more robust. Using a redundant wavelet transform could help improving the empirical results by providing sparser coefficients. Moreover, the proposed algorithm does not benefit from warm-restarts contrary to the ℓ_0 minimization.

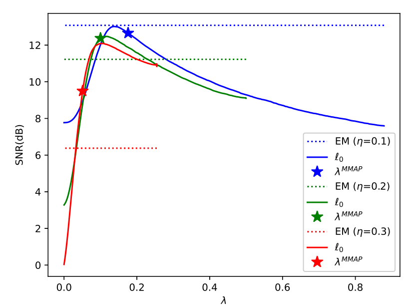

Fig. 6. Comparison of Alg. 1 for image deconvolution and ℓ_0 minimization for various η, with a fixed iSNR $= 10$ dB.

5. CONCLUSION

We proposed fully unsupervised algorithms based on an EM procedure to estimate the parameters of a BG prior. For denoising, we have shown empirically that the estimation is excellent and robust. The posterior-mean estimate reaches then the best possible empirical oSNR as expected. For inverse problems, the EM approach relies on an additional latent variable which reduces to an iterative algorithm such as ISTA, where the previous denoising EM procedure replaces the proximal step. In addition to a posterior-mean estimate, which depends on the latent variable, we also provide a value for the hyperparameter λ used in the ℓ_0 minimization. Future works will study how our approach could exploit warm-start strategies to improve the estimation of the parameters in difficult inverse problems. We also plan to do a proper comparison with Bayesian approaches such as [22, 23]. Results on inverse problems with images should also benefit from a translation-invariant wavelet transform, which produces sparser representations.

6. REFERENCES

[1] A. Ribes and F. Schmitt, "Linear inverse problems in imaging," *IEEE Sig. Proc. Mag.*, vol. 25, no. 4, pp. 84–99, 2008.

[2] J.-L. Starck, "Sparsity and inverse problems in astrophysics," *Journal of Physics: Conference Series*, vol. 699, 2016.

[3] Stéphane Mallat, *A Wavelet Tour of Signal Processing*, Academic press, 2008.

[4] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, 2009.

[5] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, 2005.

[6] Amir Beck and Marc Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[7] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comp. Harmonic Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[8] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2778–2786, 2007.

[9] R. Giryes, M. Elad, and Y. C. Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Applied and Computational Harmonic Analysis*, vol. 30, no. 3, pp. 407–422, 2011.

[10] C.-A. Deledalle, G. Peyré, and J. Fadili, "Stein consistent risk estimator (score) for hard thresholding," *arXiv preprint arXiv:1301.5874*, 2013.

[11] F. Champagnat, Y. Goussard, and J. Idier, "Unsupervised deconvolution of sparse spike trains using stochastic approximation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2988–2998, Dec. 1996.

[12] C. Soussen, J. Idier, D. Brie, and J. Duan, "From Bernoulli-Gaussian deconvolution to sparse signal restoration," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4572–4584, 2011.

[13] M. Figueiredo and R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 906–916, 2003.

[14] M. Kowalski and T. Rodet, "An unsupervised algorithm for hybrid/morphological signal decomposition," in *Proc. IEEE ICASSP*, 2011, pp. 4112–4115.

[15] E. Gassiat, F. Monfront, and Y. Goussard, "On simultaneous signal estimation and parameter identification using a generalized likelihood approach," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 157–162, 1992.

[16] D. Ge, J. Idier, and E. Le Carpentier, "A new MCMC algorithm for blind Bernoulli-Gaussian deconvolution," in *Proc. 16th Eur. Sig. Proc. Conf.*, 2008, pp. 1–5.

[17] M. C. Amrouche, H. Carfantan, and J. Idier, "A partially collapsed Gibbs sampler for unsupervised nonnegative sparse signal restoration," in *Proc. IEEE ICASSP*, Toronto, Canada, Jun. 2021.

[18] M. Protter, I. Yavneh, and M. Elad, "Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3471–3484, 2010.

[19] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, 2004.

[20] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc., 1993.

[21] E. T. Hale, W. Yin, and Y. Zhang, "Fixed-point continuation for ℓ_1-minimization: Methodology and convergence," *SIAM J. Optimization*, vol. 19, no. 3, pp. 1107–1130, 2008.

[22] Lotfi Chaari, Jean-Yves Toumeret, and Caroline Chaux, "Sparse signal recovery using a bernoulli generalized gaussian prior," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 1711–1715.

[23] Juan G Serra, Javier Mateos, Rafael Molina, and Aggelos K Katsaggelos, "Variational em method for blur estimation using the spike-and-slab image prior," *Digital Signal Processing*, vol. 88, pp. 116–129, 2019.