



Marginal MAP estimation of a Bernoulli-Gaussian signal: continuous relaxation approach

Pierre Barbault, Matthieu Kowalski, Charles Soussen

► To cite this version:

Pierre Barbault, Matthieu Kowalski, Charles Soussen. Marginal MAP estimation of a Bernoulli-Gaussian signal: continuous relaxation approach. EUSIPCO 2023 - the 31st European Signal Processing Conference, Sep 2023, Helsinki, Finland. pp.1833-1837. hal-04383978

HAL Id: hal-04383978

<https://centralesupelec.hal.science/hal-04383978>

Submitted on 10 Jan 2024

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.

Marginal MAP estimation of a Bernoulli-Gaussian signal: continuous relaxation approach

Pierre Barbault*, Matthieu Kowalski**, Charles Soussen*

*L2S, Université Paris-Saclay, CNRS, CentraleSupélec

**Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numériques
Gif-sur-Yvette, France

pierre.barbault@centralesupelec.fr, matthieu.kowalski@universite-paris-saclay.fr, charles.soussen@centralesupelec.fr

Abstract—We focus on recovering the support of sparse signals for sparse inverse problems. Using a Bernoulli-Gaussian prior to model sparsity, we propose to estimate the support of the sparse signal using the so-called Marginal Maximum a Posteriori estimate after marginalizing out the values of the nonzero coefficients. To this end, we propose an Expectation-Maximization procedure in which the discrete optimization problem in the M-step is relaxed into a continuous problem. Empirical assessment with simulated Bernoulli-Gaussian data using magnetoencephalographic lead field matrix shows that this approach outperforms the usual ℓ_0 Joint Maximum a Posteriori estimation in Type-I and Type-II error for support recovery, as well as in SNR for signal estimation

Index Terms—Sparse coding, inverse problem, Bernoulli-Gaussian model, Marginal-MAP, Joint-MAP.

I. INTRODUCTION

Let us consider a linear inverse problem described by an operator $\mathbf{H} \in \mathbb{R}^{M \times N}$ that generates a set of observations

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \quad (1)$$

with $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{y} \in \mathbb{R}^M$. We focus on the sparse signal setting, where vector \mathbf{x} contains a few nonzero entries. This approach is now part of the state-of-the-art for inverse problems where many convex and non-convex optimization methods are available, see e.g., [1], [2]. In this paper, we make use of the Bernoulli-Gaussian (BG) statistical prior to model sparse signals \mathbf{x} , with known parameters $p \in (0, 1)$ and $\sigma_x^2 > 0$ coding for the rate and variance of the nonzero entries. For all n , the entries $x[n]$ are independent, identically distributed (i.i.d.), with

$$p(x[n]) = \frac{p}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{x[n]^2}{2\sigma_x^2}\right) + (1-p)\delta(x[n])$$

where δ stands for the Dirac distribution centered on zero. The noise \mathbf{n} is assumed white and Gaussian: $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I})$.

Maximum *a posteriori* (MAP) estimation of \mathbf{x} using the BG prior combined with the white Gaussian noise leads to the usual minimization of the $\ell_2 + \ell_0$ cost function when it comes to minimizing the negative log posterior likelihood [3]:

$$\mathbf{x}^{\text{MAP}} = \underset{\mathbf{x}}{\operatorname{argmin}} \frac{1}{2\sigma_0^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{1}{2\sigma_x^2} \|\mathbf{x}\|_2^2 + \rho \|\mathbf{x}\|_0 \quad (2)$$

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

with $\rho = \log\left(\frac{1-p}{p}\right)$, $\|\mathbf{x}\|_2^2 = \sum_{n=1}^N |x[n]|^2$ and $\|\mathbf{x}\|_0 = \#\{n, x[n] \neq 0\}$. The latter optimization problem being highly non-convex, classical solvers yield sub-optimal solutions, including proximal descent algorithms [4] akin to iterative hard thresholding [5] and greedy algorithms [3].

A convenient reformulation of the BG prior is to write \mathbf{x} as the product of two independent random variables:

$$\forall n, x[n] = q[n]r[n] \quad (3)$$

where $q[n]$ is a binary variable equal to 0 when $x[n] = 0$ and 1 otherwise, distributed according to the Bernoulli distribution: $q[n] \sim \mathcal{B}(p)$ and $r[n]$ equals the signal amplitudes: $r[n] \sim \mathcal{N}(0, \sigma_x^2)$. Using matrix notations, one has

$$\mathbf{x} = \mathbf{Q}\mathbf{r} \quad \text{where} \quad \mathbf{Q} = \operatorname{Diag}(\mathbf{q}). \quad (4)$$

In the literature, the MAP estimation of $\mathbf{x} = \{\mathbf{q}, \mathbf{r}\}$ can be formulated by either maximizing the joint posterior likelihood of (\mathbf{q}, \mathbf{r}) or the marginal posterior likelihood of \mathbf{q} . This leads to two distinct estimators: the joint and marginal MAP, respectively (JMAP and MMAP), see [6].

The JMAP expression can be straightforwardly derived by writing $p(\mathbf{q}, \mathbf{r} | \mathbf{y})$:

$$(\hat{\mathbf{q}}^{\text{JMAP}}, \hat{\mathbf{r}}^{\text{JMAP}}) = \underset{\mathbf{q}, \mathbf{r}}{\operatorname{argmin}} \frac{1}{2\sigma_0^2} \|\mathbf{y} - \mathbf{H}\mathbf{Q}\mathbf{r}\|^2 + \frac{1}{2\sigma_x^2} \|\mathbf{r}\|^2 + \rho \|\mathbf{q}\|_0 \quad (5)$$

One can notice that for fixed \mathbf{q} , the latter criterion is quadratic with respect to \mathbf{r} . Thus, the minimizer of (5) with respect to \mathbf{r} has a closed-form expression $\mathbf{r}(\mathbf{q})$. Plugging this expression into the cost function (5), the JMAP problem can be reformulated (up to technical rearrangements) as:

$$\hat{\mathbf{q}}^{\text{JMAP}} = \underset{\mathbf{q} \in \{0,1\}^N}{\operatorname{argmin}} \mathbf{y}^{(t)} \Gamma_y^{-1}(\mathbf{q}) \mathbf{y} + \rho \|\mathbf{q}\|_0 \quad (6)$$

where

$$\Gamma_y(\mathbf{q}) = \sigma_0^2 \mathbf{I} + \sigma_x^2 \mathbf{H}\mathbf{Q}\mathbf{Q}^{(t)} \mathbf{H}^{(t)}. \quad (7)$$

Once the support \mathbf{q}^{JMAP} has been estimated, the nonzero amplitudes can be deduced by solving a least-squares problem (i.e., by minimizing the Mean Squared Error (MSE) $\mathbb{E}_{\mathbf{x}|\mathbf{y}, \mathbf{q}}[\|\mathbf{x} - \hat{\mathbf{x}}\|^2]$), which reads:

$$\mathbf{x}(\mathbf{q}) = \underset{\mathbf{x}}{\operatorname{argmax}} p(\mathbf{x}|\mathbf{y}, \mathbf{q}) = \sigma_x^2 \mathbf{Q}\mathbf{H}^{(t)} \Gamma_y^{-1}(\mathbf{q}) \mathbf{y}. \quad (8)$$

M MAP estimation consists of minimizing the marginal posterior likelihood $p(\mathbf{q}|\mathbf{y})$ after marginalizing out the signal coefficients \mathbf{r} [7]. It is a natural estimator for applications where the support may bear more interest than the amplitudes, *e.g.*, for source localization of brain activity using Magneto/ElectroEncephalography (M/EEG) [8], [9], [10]. For such problems, it seems more appropriate to first estimate the support using the MMAP estimator, which is the Bayes estimator for the 0–1 loss with discrete random variables, and then retrieve the coefficients \mathbf{r} . It is noticeable that the JMAP and MMAP optimization problems are highly non-convex.

Contributions and outline of the paper. We propose an Expectation-Maximization (EM) approach dedicated to Marginal-MAP estimation. The algorithm is derived in Section II. In Section III, the resulting binary optimization problem is relaxed into a continuous problem over $[0, 1]^N$. We first derive the appropriate algorithm to reach a local minimizer over $[0, 1]^N$. The latter is then used as warm start initialization of the binary EM algorithm. Finally, the numerical experiments in Section IV demonstrate the validity of our approach.

II. MARGINAL-MAP ESTIMATION OF THE SUPPORT

Marginal-MAP estimation relies on the maximization of $p(\mathbf{q}|\mathbf{y}) = \int p(\mathbf{q}, \mathbf{r}|\mathbf{y}) d\mathbf{r}$ over $\{0, 1\}^N$. According to [6], the Marginal-MAP estimate can be found by minimizing

$$-\log p(\mathbf{q}|\mathbf{y}) = \frac{1}{2}\mathbf{y}^{(t)}\mathbf{\Gamma}_y^{-1}(\mathbf{q})\mathbf{y} + \frac{1}{2}\log |\mathbf{\Gamma}_y(\mathbf{q})| + \rho\|\mathbf{q}\|_0 + \kappa \quad (9)$$

where the constant κ does not depend on \mathbf{q} . However, this minimization problem is *NP*-Hard. We propose a sub-optimal approach based on the Expectation-Maximization (EM) algorithm. Following [11], the observation model (1) is rewritten as

$$\mathbf{y} = \mathbf{H}\mathbf{z} + \mathbf{e} \quad \text{and} \quad \mathbf{z} = \mathbf{Q}\mathbf{r} + \mathbf{b} \quad (10)$$

where

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_e) \quad \text{and} \quad \mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma_b^2\mathbf{I}) \quad (11)$$

are independent Gaussian vectors, such that

$$\mathbf{\Gamma}_e + \sigma_b^2\mathbf{H}\mathbf{H}^{(t)} = \sigma_0^2\mathbf{I}. \quad (12)$$

Notice that one necessarily has $\sigma_b^2 \leq \frac{\sigma_0^2}{\|\mathbf{H}\mathbf{H}^{(t)}\|}$ (where $\|\cdot\|$ refers to the spectral norm of a matrix) to ensure a non-degenerate normal distribution for \mathbf{e} . This model has been proposed in [11] to derive the Iterative Shrinkage/Thresholding Algorithm for the LASSO [1]/Basis Pursuit Denoising [2] problem, and re-used in [12], [13] to estimate the parameters (ρ, σ_x^2) of a Bernoulli-Gaussian model.

Note that for a given \mathbf{z} , the recovery of \mathbf{q} from $\mathbf{z} = \mathbf{Q}\mathbf{r} + \mathbf{b}$ boils down to a simple denoising problem. For the latter, the MMAP estimator of \mathbf{q} yields a closed-form component-wise expression for each binary variable $q[n]$ [14], [12], [13]:

Proposition II.1. *Let $\mathbf{z} = \mathbf{Q}\mathbf{r} + \mathbf{b}$ as in Eq. (10) with the corresponding prior. The MMAP estimator of $\mathbf{q}|\mathbf{z}$ can be written component-wise. For all n ,*

$$q[n] = 1 \quad \text{iff} \quad P(q[n] = 1|z[n]) \geq P(q[n] = 0|z[n])$$

$$\text{i.e. iff} \quad z[n]^2 \geq 2\sigma_b^2 \frac{\sigma_x^2 + \sigma_b^2}{\sigma_x^2} \left(\rho + \frac{1}{2} \log \left(\frac{\sigma_b^2 + \sigma_x^2}{\sigma_b^2} \right) \right)$$

Let us derive an EM algorithm for MMAP estimation by considering \mathbf{z} as a hidden variable. The EM approach then reads:

$$\begin{aligned} \mathbf{q}^{(t+1)} &= \underset{\mathbf{q} \in \{0,1\}^N}{\operatorname{argmin}} \mathcal{Q}(\mathbf{q}, \mathbf{q}^{(t)}) \\ \mathcal{Q}(\mathbf{q}, \mathbf{q}^{(t)}) &= \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}[-\log p(\mathbf{z}, \mathbf{q}|\mathbf{y})] \end{aligned} \quad (13)$$

The E-step and M-step are derived hereafter.

A. E-step

For conciseness, we will use the following writing:

$$f(\mathbf{q}) + \kappa \stackrel{\kappa}{=} f(\mathbf{q}).$$

to refer to an equality up to any constant κ (which does not depend on \mathbf{q}). Using the linearity of the expectation and the fact that $p(\mathbf{q}|\mathbf{z}, \mathbf{y}) = p(\mathbf{q}|\mathbf{z})$, we have that

$$\mathcal{Q}(\mathbf{q}, \mathbf{q}^{(t)}) \stackrel{\kappa}{=} \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}[-\log p(\mathbf{q}|\mathbf{z})] \quad (14)$$

where $\kappa = \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}[-\log p(\mathbf{z}|\mathbf{y})]$. Then, Bayes' rule yields:

$$\mathcal{Q}(\mathbf{q}, \mathbf{q}^{(t)}) \stackrel{\kappa}{=} \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}[-\log p(\mathbf{z}|\mathbf{q})] + \rho\|\mathbf{q}\|_0 \quad (15)$$

with now $\kappa = \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}[\log p(\mathbf{z}) - \log p(\mathbf{z}|\mathbf{y})]$. Using Eqs. (10) to (12), we directly have

$$\mathbf{z}|\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}_z(\mathbf{q})) \quad \text{with} \quad \mathbf{\Gamma}_z(\mathbf{q}) = \sigma_b^2\mathbf{I} + \sigma_x^2\mathbf{Q}^{(t)}\mathbf{Q}. \quad (16)$$

Applying Bayes' rule, the posterior distribution of \mathbf{z} reads

$$\mathbf{z}|\mathbf{y}, \mathbf{q} \sim \mathcal{N}(\hat{\mathbf{z}}, \mathbf{\Sigma}) \quad (17)$$

with

$$\begin{aligned} \hat{\mathbf{z}} &= \mathbf{\Gamma}_z(\mathbf{q})\mathbf{H}^{(t)}\mathbf{\Gamma}_y^{-1}(\mathbf{q})\mathbf{y} \\ \mathbf{\Sigma} &= \mathbf{\Gamma}_z(\mathbf{q}) - \mathbf{\Gamma}_z(\mathbf{q})\mathbf{H}^{(t)}\mathbf{\Gamma}_y^{-1}(\mathbf{q})\mathbf{H}\mathbf{\Gamma}_z(\mathbf{q}). \end{aligned} \quad (18)$$

So, we get

$$\begin{aligned} \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}[-\log p(\mathbf{z}|\mathbf{q})] &\stackrel{\kappa}{=} \frac{1}{2}\mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^{(t)}}\left[\mathbf{z}^{(t)}\mathbf{\Gamma}_z^{-1}(\mathbf{q})\mathbf{z}\right] + \frac{1}{2}\log |\mathbf{\Gamma}_z(\mathbf{q})| \\ &\stackrel{\kappa}{=} \frac{1}{2}(\hat{\mathbf{z}}^{(t)})^{(t)}\mathbf{\Gamma}_z^{-1}(\mathbf{q})(\hat{\mathbf{z}}^{(t)}) + \frac{1}{2}\operatorname{Trace}[\mathbf{\Gamma}_z^{-1}(\mathbf{q})\mathbf{\Sigma}^{(t)}] + \frac{1}{2}\log |\mathbf{\Gamma}_z(\mathbf{q})|. \end{aligned}$$

Finally, the E-step reads

$$\begin{aligned} \mathcal{Q}(\mathbf{q}, \mathbf{q}^{(t)}) &\stackrel{\kappa}{=} \frac{1}{2}(\hat{\mathbf{z}}^{(t)})^{(t)}\mathbf{\Gamma}_z^{-1}(\mathbf{q})(\hat{\mathbf{z}}^{(t)}) + \frac{1}{2}\operatorname{Trace}[\mathbf{\Gamma}_z^{-1}(\mathbf{q})\mathbf{\Sigma}^{(t)}] \\ &\quad + \frac{1}{2}\log |\mathbf{\Gamma}_z(\mathbf{q})| + \rho\|\mathbf{q}\|_0 \\ &\stackrel{\kappa}{=} \frac{1}{2} \sum_{n=1}^N \mathcal{Q}_n(q[n]) \end{aligned} \quad (19)$$

with

$$\mathcal{Q}_n(q[n]) = \frac{(\hat{z}^{(t)}[n])^2 + \Sigma^{(t)}[n, n]}{\sigma_b^2 + \sigma_x^2 q[n]^2} + \log(\sigma_b^2 + \sigma_x^2 q[n]^2) + 2\rho q[n]$$

As expected, $\mathcal{Q}(\mathbf{q}, \mathbf{q}^{(t)})$ is a decoupled sum on $q[n]$. Thus, the minimization of \mathcal{Q} boils down to the separate minimization of $\mathcal{Q}_n(q[n])$ for all n . This minimization is derived in the M-step described hereafter.

B. M-Step

Given the previous observation on the decoupling of the support variables and using the notation $\Gamma_n^{(t)} = (\hat{z}^{(t)}[n])^2 + \Sigma^{(t)}[n, n]$, the M-step simplifies to

$$q^{(t+1)}[n] = \operatorname{argmin}_{q \in \{0,1\}} \frac{\Gamma_n^{(t)}}{\sigma_b^2 + \sigma_x^2 q^2} + \log(\sigma_b^2 + \sigma_x^2 q^2) + 2\rho q \quad (20)$$

It appears that $\hat{q}^{(t+1)}[n] = 1$ corresponds to the case where

$$\Gamma_n^{(t)} \geq \sigma_b^2 \frac{\sigma_x^2 + \sigma_b^2}{\sigma_x^2} \left(\log \left(\frac{\sigma_b^2 + \sigma_x^2}{\sigma_b^2} \right) + 2\rho \right). \quad (21)$$

This operation is similar to the thresholding formula in the Marginal-MAP denoising problem with $\Gamma_n^{(t)}$ instead of $z[n]$ in [Prop. II.1](#).

C. Summary of the algorithm

The whole EM procedure is summarized in [Alg. 1](#). One can notice that when $\mathbf{H} = \mathbf{I}$, \mathbf{e} can be set to $\mathbf{0}$ in (10), thus $\sigma_b^2 = \sigma_0^2$, and the covariance matrices Γ_z and Γ_y are equal. Then, [Alg. 1](#) retrieves the MMAP estimator in the denoising case (that is, for $\mathbf{z} = \mathbf{y}$ in [Prop. II.1](#)).

The EM algorithm is known as a local ascent method, converging towards a local maximizer of the MMAP criterion. Since the MMAP criterion is highly non-convex (due to the presence of the ℓ_0 cost operator), the algorithm may reach a poor local maximizer. In the next section, we propose to relax the MMAP problem in the continuous setting, that is, for $\mathbf{q} \in [0, 1]^N$ to reach warm start support, which may be further used as an initial solution for [Alg. 1](#).

Algorithm 1: EM algorithm for MMAP estimation of support \mathbf{q}

Result: $\mathbf{q} \in \{0, 1\}^N$

Input: $t = 0, \mathbf{q}^{(t)} \in [0, 1]^N, \rho > 0$

while not converged do

$\mathbf{Q}^{(t)} = \text{Diag}(\mathbf{q})$;

$\Gamma_z^{(t)} = \sigma_b^2 \mathbf{I} + \sigma_x^2 \mathbf{Q}^{(t)} \mathbf{Q}^{(t)^{(t)}};$

$\Gamma_y^{(t)} = \sigma_0^2 \mathbf{I} + \sigma_x^2 \mathbf{H} \mathbf{Q}^{(t)} \mathbf{Q}^{(t)^{(t)}} \mathbf{H}^{(t)};$

$\Sigma^{(t)} = \Gamma_z^{(t)} - \Gamma_z^{(t)} \mathbf{H}^{(t)} \Gamma_y^{-1}(\mathbf{q}) \mathbf{H} \Gamma_z^{(t)};$

$\Gamma_n^{(t)} = \hat{z}^{(t)}[n]^2 + \Sigma^{(t)}[n, n];$

for $n = 1$ **to** N **do**

if $\Gamma_n^{(t)} \geq 2\sigma_b^2 \frac{\sigma_b^2 + \sigma_x^2}{\sigma_x^2} \left(\rho + \log \sqrt{1 + \frac{\sigma_x^2}{\sigma_b^2}} \right)$ **then**

$q^{(t+1)}[n] = 1;$

else

$q^{(t+1)}[n] = 0;$

end

end

$t = t + 1$

end

III. OPTIMIZATION BY CONTINUOUS RELAXATION

Hereafter, the binary optimization problem defined in (13) is relaxed into a continuous optimization problem, in which the objective function is unchanged and the domain $\{0, 1\}^N$ is replaced by $[0, 1]^N$. According to [Section II](#), the cost function $\mathbf{Q}(\mathbf{q}, \mathbf{q}^{(t)})$ reads as the decoupled sum (19). So, the M-step still consists of solving 1D problems akin to (20).

First, let us point out that when $q[n] \in \{0, 1\}$, we have $q[n]^2 = q[n]$. Then, minimizing $\mathcal{Q}_n(q[n])$ over $\{0, 1\}$ is equivalent to minimizing:

$$\frac{\Gamma_n^{(t)}}{\sigma_b^2 + \sigma_x^2 q[n]^2} + \log(\sigma_b^2 + \sigma_x^2 q[n]^2) + 2\rho q[n]^2.$$

Now, consider the continuous relaxation of $q[n]$ over $[0, 1]$:

$$q_n^{(t+1)} = \operatorname{argmin}_{q \in [0,1]} \frac{\Gamma_n^{(t)}}{\sigma_b^2 + \sigma_x^2 q^2} + \log(\sigma_b^2 + \sigma_x^2 q^2) + 2\rho q^2. \quad (22)$$

This is a 1D problem with bound constraints. The related unconstrained minimizer can be found by calculating the roots of the first-order derivative w.r.t $u = q^2$, written

$$\frac{-\sigma_x^2 \Gamma_n^{(t)}}{(\sigma_b^2 + \sigma_x^2 u)^2} + \frac{\sigma_x^2}{\sigma_b^2 + \sigma_x^2 u} + 2\rho = 0. \quad (23)$$

When $\rho > 0$, there is a single positive root:

$$u[n] = \frac{1}{4\rho} \left(\sqrt{1 + 8\rho \frac{\Gamma_n^{(t)}}{\sigma_x^2}} - 1 \right) - \frac{\sigma_b^2}{\sigma_x^2}. \quad (24)$$

Moreover, a careful study of the cost function within (22) (seen as a function of $u = q^2$) reveals that its second-order derivative is non-negative for the positive root u . Taking into account the bound constraints $q \in [0, 1]$, we get:

$$q^{(t+1)}[n] = \begin{cases} 0 & \text{if } u[n] \leq 0 \\ 1 & \text{if } u[n] \geq 1 \\ \sqrt{u[n]} & \text{otherwise.} \end{cases} \quad (25)$$

Algorithm 2: EM algorithm for continuous relaxed \mathbf{q}

Result: $\mathbf{q} \in [0, 1]^N$

Input: $t = 0, \mathbf{q}^{(t)} \in [0, 1]^N$

while not converged do

$\mathbf{Q}^{(t)} = \text{Diag}(\mathbf{q})$;

$\Gamma_z^{(t)} = \sigma_b^2 \mathbf{I} + \sigma_x^2 \mathbf{Q}^{(t)} \mathbf{Q}^{(t)^{(t)}};$

$\Gamma_y^{(t)} = \sigma_0^2 \mathbf{I} + \sigma_x^2 \mathbf{H} \mathbf{Q}^{(t)} \mathbf{Q}^{(t)^{(t)}} \mathbf{H}^{(t)};$

$\Sigma^{(t)} = \Gamma_z^{(t)} - \Gamma_z^{(t)} \mathbf{H}^{(t)} \Gamma_y^{-1}(\mathbf{q}) \mathbf{H} \Gamma_z^{(t)};$

for $n = 1$ **to** N **do**

$u[n] = \frac{1}{4\rho} \left(\sqrt{1 + 8\rho \frac{\Gamma_n^{(t)}}{\sigma_x^2}} - 1 \right) - \frac{\sigma_b^2}{\sigma_x^2};$

$q[n] = \max\{\min\{\sqrt{u[n]}, 1\}, 0\};$

end

$t = t + 1;$

end

The resulting algorithm is given in Alg. 2.

One can remark that Alg. 2 is very similar to the EM Sparse Bayesian Learning given in [15], except for the estimation of \mathbf{q} in the last step. Indeed, one has $\mathbf{x}|\mathbf{q} \sim \mathcal{N}(0, \sigma_x^2 \mathbf{Q})$. Then, when \mathbf{q} is continuous valued, we recover the prior used in SBL. The proposed continuous relaxation can be seen as an SBL approach where the variances are constrained in $[0, \sigma_x^2]^N$, with a particular prior leading to the thresholding step. Moreover, when $\rho = 0$, that is $p = \frac{1}{2}$ in the BG model, the updates (22) of $q[n]$ are given by

$$q[n]^2 = \min \left\{ 1, \left(\frac{\Gamma_n^{(t)} - \sigma_b^2}{\sigma_x^2} \right)^+ \right\}. \quad (26)$$

with $(x)^+ = \max(x, 0)$. Then, when $\sigma_x^2 = 1$ and $\sigma_b^2 = 0$, Alg. 2 exactly reduces to the EM SBL proposed by [15].

In Alg. 2, ρ can be seen as a hyperparameter. We propose to run Alg. 2 with various values of this hyperparameter and use the result as an initialization for the binary EM Marginal-MAP (Alg. 1) with the actual value of the model. The resulting procedure is summarized in Alg. 3.

Algorithm 3: Practical algorithm for Marginal-MAP estimation

Result: $\mathbf{q} \in \{0, 1\}^N$

Input: $k = 0$, $\tilde{\mathbf{q}}^{(k)} \in [0, 1]^N$, $\Lambda = (\lambda^0 > \dots > \lambda^K)$

for $k = 0$ to K **do**

 Estimate $\tilde{\mathbf{q}}^{(k+1)} \in [0, 1]^N$ by Alg. 2 initialized by $\tilde{\mathbf{q}}^{(k)}$ for $\rho = \lambda^k$;

 Estimate $\mathbf{q}^{(k+1)} \in \{0, 1\}^N$ by Alg. 1 initialized by $\tilde{\mathbf{q}}^{(k+1)}$ for $\rho = \log\left(\frac{1-p}{p}\right)$;

end

IV. NUMERICAL EXPERIMENTS

In this section, we assess the performance of the proposed algorithm using statistical results on simulated data.

The simulated signals \mathbf{q} and \mathbf{r} are generated with Bernoulli and Gaussian distributions of parameters $p = 0.05, 0.01$ and $\sigma_x^2 = 1$. Then \mathbf{x} is obtained according to $\mathbf{x} = \mathbf{Q}\mathbf{r}$. The observations are degraded by a Gaussian noise of variance $\sigma_0^2 = 0.01$. The operator \mathbf{H} used here is a sub-matrix drawn from a M/EEG leadfield of size 272×600 . This kind of operator contains strongly correlated columns and thus is representative of common M/EEG inverse problems. We compare the Marginal-MAP estimation of \mathbf{q} given by (3), and the corresponding signal estimate $\mathbf{x}(\mathbf{q})$ as in (8), to the usual Joint-MAP estimation of \mathbf{q} and \mathbf{x} obtained by minimizing (2). In practice, we used a proximal descent algorithm similar to the popular Iterative Hard Thresholding (IHT) algorithm [5] with warm restart.

We focus on the quality of the support estimation \mathbf{q} . The signal being sparse, we compare both approaches using:

- the False Positive rate (Type-I error);
- the False Negative rate (Type-II error);

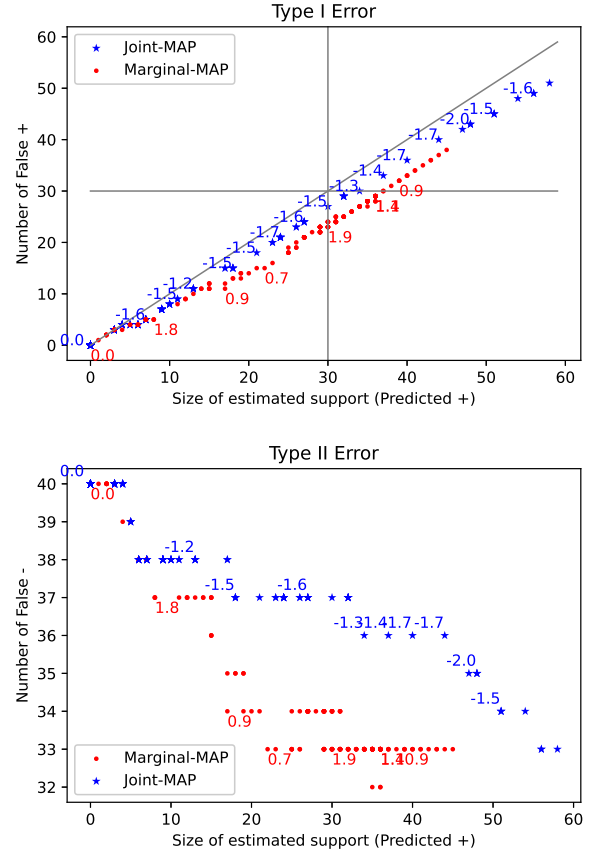


Fig. 1. Marginal-MAP VS Joint-MAP estimation $p = 0.05$, $\sigma_0^2 = 0.01$. The SNR of the associated estimated signal \mathbf{x} is given for some point (including the best reached SNR)

- the SNR of the estimated $\mathbf{x}(\mathbf{q})$ using (8).

The Type-I and Type-II errors are displayed versus the predicted positives (which is the size of the estimated support \mathbf{q}), the latter being directly related to the choice of the hyperparameter ρ in Alg. 3 (the larger ρ , the sparser \mathbf{q}).

The results are presented in Figs. 1 and 2. It can be seen that the Marginal-MAP outperforms the Joint-MAP in both type-I and type-II errors and in SNR (especially Type-II error and SNR). Moreover, the algorithm performs well for highly correlated \mathbf{H} . Other experiments (not displayed here) show that for moderately correlated random Gaussian matrices \mathbf{H} , the Joint-MAP and Marginal-MAP approaches give similar performance, the latter being more computational demanding because of the computation of $\Sigma^{(t)}$ in Algs. 1 and 2.

V. DISCUSSION AND CONCLUSION

We have proposed a Marginal-MAP approach for selecting the best possible support, as it is a Bayesian estimator in the sense that the Marginal-MAP minimizes the 0 – 1 Loss. The selection of the best possible support can be seen as a variable selection problem well studied in statistics [16], also known as the best subset selection problem [17]. The Lasso [1] was initially proposed for variable selection. The problem of

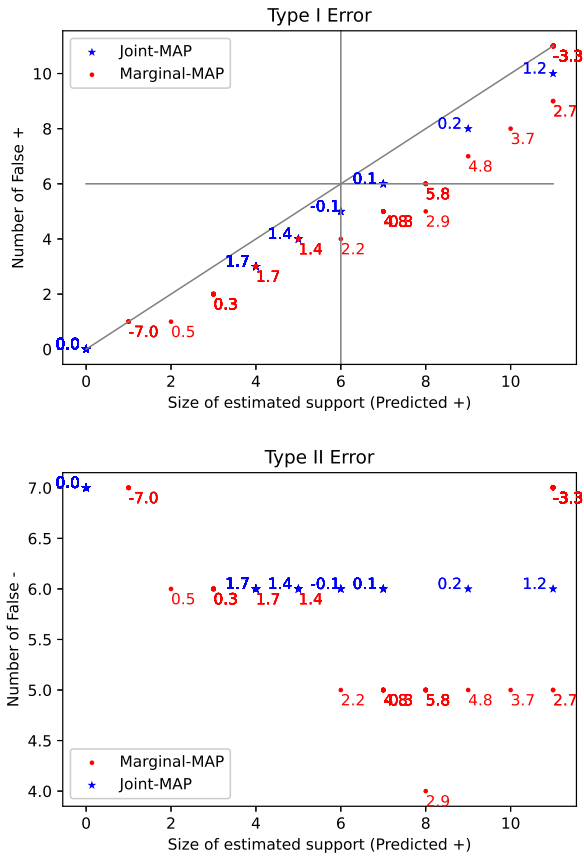


Fig. 2. Marginal-MAP VS Joint-MAP estimation $p = 0.01$, $\sigma_0^2 = 0.01$. The SNR of the associated estimated signal \mathbf{x} is given for some point (including the best reached SNR)

variable selection is studied in the context of "prediction"; hence the estimation of \mathbf{x} is essential. In [18], an extensive comparison is made between the Lasso and the optimal ℓ_0 solution using a MILP solver as proposed in [19], [20] (as well as few other methods). In the proposed context, the ℓ_0 approach and the Lasso perform very similarly. However, the matrix used in their simulation is not highly correlated as a M/EEG lead field matrix can be. This study opens perspectives on various topics, such as the relations between the Joint-MAP minimizers and the Marginal-MAP ones. Future works will also cover the impact of continuous relaxations. An extensive comparison may be performed between the Marginal-MAP approach, the Joint-MAP, and other approaches for variable selection. In particular, a comparison with MCMC methods for BG prior [21] and the EMVS method proposed in [22]. The latter relies on a mixture of Gaussians, which is very similar to the pdf of the hidden variable \mathbf{z} in Eq. (10). From an application perspective, the method could be extended to structured models to apply to actual M/EEG data [23].

REFERENCES

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM review*, vol. 43, no. 1, pp. 129–159, 2001.

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

[4] H. Li and Z. Lin, "Accelerated proximal gradient methods for nonconvex programming," *Advances in Neural Information Processing Systems*, vol. 28, 2015.

[5] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, 2009.

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

[7] F. Champagnat, Y. Goussard, S. Gautier, and J. Idier, "Deconvolution of spike trains," in *Bayesian Approach to Inverse Problems*, J. Idier, Ed. Apr. 2008, pp. 117–140, ISTE Ltd and John Wiley & Sons Inc.

[8] H. Stefan, C. Hummel, G. Scheler *et al.*, "Magnetic brain source imaging of focal epileptic activity: a synopsis of 455 cases," *Brain*, vol. 126, no. 11, pp. 2396–2405, 2003.

[9] R. Knowlton, R. Elgavish, J. Howell *et al.*, "Magnetic source imaging versus intracranial electroencephalogram in epilepsy surgery: a prospective study," *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, vol. 59, no. 5, pp. 835–842, 2006.

[10] L. Koessler, C. Benar, L. Maillard *et al.*, "Source localization of ictal epileptic activity investigated by high resolution EEG and validated by SEEG," *Neuroimage*, vol. 51, no. 2, pp. 642–653, 2010.

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

[12] M. Kowalski and T. Rodet, "An unsupervised algorithm for hybrid/morphological signal decomposition," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 4112–4115.

[13] P. Barbault, M. Kowalski, and C. Soussen, "Parameter estimation in sparse inverse problems using Bernoulli-Gaussian prior," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 5413–5417.

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

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

[16] R. B. O'Hara and M. J. Sillanpää, "A review of Bayesian variable selection methods: What, how and which," *Bayesian Analysis*, vol. 4, no. 1, pp. 85–118, 2009.

[17] R. R. Hocking and R. N. Leslie, "Selection of the best subset in regression analysis," *Technometrics*, vol. 9, no. 4, pp. 531–540, 1967.

[18] T. Hastie, R. Tibshirani, and R. Tibshirani, "Best subset, forward stepwise or lasso? analysis and recommendations based on extensive comparisons," *Statistical Science*, vol. 35, no. 4, pp. 579–592, 2020.

[19] D. Bertsimas, King A., and R. Mazumder, "Best subset selection via a modern optimization lens," *The Annals of Statistics*, vol. 44, no. 2, pp. 813–852, 2016.

[20] S. Bourguignon, J. Ninin, H. Carfantan, and M. Mongeau, "Exact sparse approximation problems via mixed-integer programming: Formulations and computational performance," *IEEE Transactions on Signal Processing*, vol. 64, no. 6, pp. 1405–1419, 2015.

[21] D. Ge, J. Idier, and E. Le Carpentier, "Enhanced sampling schemes for mcmc based blind Bernoulli-Gaussian deconvolution," *Signal Processing*, vol. 91, no. 4, pp. 759–772, 2011.

[22] V. Roycková and E. I. George, "EMVS: The EM approach to Bayesian variable selection," *Journal of the American Statistical Association*, vol. 109, no. 506, pp. 828–846, 2014.

[23] M. Lim, J. M. Ales, B. R. Cottareau, T. Hastie, and A. M. Norcia, "Sparse EEG/MEG source estimation via a group lasso," *PIOS ONE*, vol. 12, no. 6, pp. e0176835, 2017.