

# Estimation du MAP Marginal d'un signal Bernoulli-Gaussien: une approche par relaxation continue

Pierre BARBAULT<sup>1</sup> Matthieu KOWALSKI<sup>2</sup> Charles SOUSSEN<sup>1</sup>

<sup>1</sup>L2S, Université Paris-Saclay, CNRS, CentraleSupélec, Gif-sur-Yvette, France

<sup>2</sup>Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numériques, Gif-sur-Yvette, France

**Résumé** – Cet article traite de l'estimation du support de signaux parcimonieux dans le cadre de problèmes inverses. A l'aide d'un *a priori* Bernoulli-Gaussien, le support est estimé en maximisant la loi *a posteriori* marginalisée par rapport aux amplitudes des sources. On propose un algorithme d'Espérance-Maximisation (EM), où la variable discrète codant le support est relâchée dans un espace continu. Une évaluation empirique sur signaux simulés avec un opérateur issu de la magnétoencephalographie montre la pertinence de cette approche en comparaison avec l'approche de maximisation de la loi jointe du support et des amplitudes, interprétable comme la minimisation d'un critère avec pénalisation  $\ell_0$ .

**Abstract** – We focus on recovering the support of sparse signals in inverse problems. Using a Bernoulli-Gaussian prior to model sparsity, we propose to estimate the support using the so-called marginal maximum *a posteriori* estimator, where the nonzero coefficients are marginalized out. 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 data using a magnetoencephalographic lead field matrix shows that this approach outperforms the joint posterior maximization of the support and coefficients, seen as the minimization of an  $\ell_0$ -penalized criterion.

## 1 Introduction

On considère le problème inverse linéaire  $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$  avec  $\mathbf{H} \in \mathbb{R}^{M \times N}$ ,  $\mathbf{x} \in \mathbb{R}^N$  et  $\mathbf{y} \in \mathbb{R}^M$ . Le vecteur  $\mathbf{x}$  inconnu est supposé parcimonieux, et modélisé par un processus Bernoulli-Gaussien (BG). Pour tout  $n$ , les éléments  $x[n]$  sont supposés indépendants, identiquement distribués (i.i.d.), avec

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

où  $\delta$  est la distribution de Dirac. Le bruit  $\mathbf{n}$  est supposé indépendant de  $\mathbf{x}$ , blanc et Gaussien :  $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_0^2 \mathbf{I})$ .

On écrit  $\mathbf{x}$  comme produit de deux variables aléatoires indépendantes :  $\forall n, x[n] = q[n]r[n]$ , où  $q[n]$  est une variable binaire représentant le support de  $\mathbf{x}$  ( $q[n] = 1$  si  $x[n] \neq 0$ ) et  $r[n]$  représente les amplitudes du signal.  $\mathbf{q}[n]$  suit une distribution de Bernoulli :  $q[n] \sim \mathcal{B}(p)$  et  $r[n]$  une loi gaussienne :  $r[n] \sim \mathcal{N}(0, \sigma_x^2)$ . En notation matricielle, cela donne

$$\mathbf{x} = \mathbf{Q}\mathbf{r} \quad \text{avec} \quad \mathbf{Q} = \text{Diag}(\mathbf{q}). \quad (1)$$

L'estimateur du MAP de  $\mathbf{x} = \{\mathbf{q}, \mathbf{r}\}$  peut être formulé comme le maximum de la vraisemblance jointe *a posteriori* de  $(\mathbf{q}, \mathbf{r})$  ou comme la vraisemblance de la loi marginale *a posteriori* de  $\mathbf{q}$ . Ces approches conduisent à deux estimateurs distincts [4] : le MAP Joint (JMAP) et le MAP Marginal (MMAP). L'expression du JMAP est obtenue à partir de  $p(\mathbf{q}, \mathbf{r} | \mathbf{y})$  :

$$\arg\min_{\mathbf{q}, \mathbf{r}} \frac{1}{2\sigma_0^2} \|\mathbf{y} - \mathbf{H}\mathbf{Q}\mathbf{r}\|^2 + \frac{1}{2\sigma_x^2} \|\mathbf{r}\|^2 + \rho \|\mathbf{q}\|_0, \quad (2)$$

avec  $\rho = \log\left(\frac{1-p}{p}\right)$ . En utilisant le changement de variable (1), l'estimation au sens du JMAP se ramène à un pro-

blème de minimisation  $\ell_2 + \ell_0$  [15] :

$$\mathbf{x}^{\text{JMAP}} = \arg\min_{\mathbf{x}} \frac{1}{2\sigma_0^2} \|\mathbf{y} - \mathbf{H}\mathbf{x}\|^2 + \frac{1}{2\sigma_x^2} \|\mathbf{x}\|^2 + \rho \|\mathbf{x}\|_0. \quad (3)$$

À  $\mathbf{q}$  fixé, le minimiseur de (2) selon  $\mathbf{r}$  admet une forme explicite  $\mathbf{r}(\mathbf{q})$ . En réintroduisant l'expression dans la fonction coût (2), le critère JMAP se réécrit :

$$\hat{\mathbf{q}}^{\text{JMAP}} = \arg\min_{\mathbf{q} \in \{0,1\}^N} \mathbf{y}^T \Gamma_y(\mathbf{q})^{-1} \mathbf{y} + \rho \|\mathbf{q}\|_0 \quad (4)$$

avec

$$\Gamma_y(\mathbf{q}) = \sigma_0^2 \mathbf{I} + \sigma_x^2 \mathbf{H}\mathbf{Q}\mathbf{Q}^T \mathbf{H}^T. \quad (5)$$

Une fois le support  $\mathbf{q}^{\text{JMAP}}$  estimé, les amplitudes non nulles sont déduites au sens des moindres carrés :

$$\mathbf{x}(\mathbf{q}) = \arg\max_{\mathbf{x}} p(\mathbf{x} | \mathbf{y}, \mathbf{q}) = \sigma_x^2 \mathbf{Q}\mathbf{H}^T \Gamma_y^{-1}(\mathbf{q}) \mathbf{y} \quad (6)$$

avec  $\mathbf{q} = \mathbf{q}^{\text{JMAP}}$ .

L'estimateur du MMAP consiste à minimiser  $p(\mathbf{q} | \mathbf{y})$ , après marginalisation des amplitudes  $\mathbf{r}$  du signal [4]. C'est un estimateur naturel dans les applications où le support contient l'information principale, comme pour la localisation de sources cérébrales en Magneto/ElectroEncephalographie (M/EEG) [10]. Il semble plus approprié d'estimer d'abord le support par le MMAP, qui correspond à l'estimateur de Bayes associé à l'erreur discrète 0 – 1, puis de reconstruire les valeurs de  $\mathbf{r}$ .

**Contributions et plan de l'article.** On propose en Section 2 une approche par Espérance-Maximisation (EM) dédiée à l'estimation du MMAP. En Section 3, les contraintes d'optimisation binaire sont relâchées dans un domaine continu,  $[0, 1]^N$ . La solution du problème relâché sert ensuite d'initialisation à chaud de l'algorithme EM pour variables binaires. Enfin, les expériences présentées en Section 4 viennent appuyer la validité de l'approche.

Ce travail est supporté par l'Agence Nationale de la Recherche à travers le projet BMWs (ANR-20-CE45-0018)

## 2 Estimation du support via le MMAP

L'estimateur du MMAP est basé sur la maximisation de  $p(\mathbf{q}|\mathbf{y}) = \int p(\mathbf{q}, \mathbf{r}|\mathbf{y}) d\mathbf{r}$  sur  $\{0, 1\}^N$ . Comme décrit dans [4], il conduit à minimiser

$$-\log p(\mathbf{q}|\mathbf{y}) = \frac{1}{2} \mathbf{y}^T \Gamma_y^{-1}(\mathbf{q}) \mathbf{y} + \frac{1}{2} \log |\Gamma_y(\mathbf{q})| + \rho \|\mathbf{q}\|_0. \quad (7)$$

qui s'avère être un problème NP-difficile. On propose une résolution sous-optimale basée sur un algorithme EM. En suivant [6], le modèle d'observation est décomposé en deux modèles liés à l'opérateur direct et au bruitage additif :

$$\mathbf{y} = \mathbf{H}\mathbf{z} + \mathbf{e} \quad \text{et} \quad \mathbf{z} = \mathbf{Q}\mathbf{r} + \mathbf{b}. \quad (8)$$

On impose les distributions

$$\mathbf{e} \sim \mathcal{N}(0, \Gamma_e) \quad \text{et} \quad \mathbf{b} \sim \mathcal{N}(0, \sigma_b^2 \mathbf{I}) \quad (9)$$

définissant des vecteurs indépendants, gaussiens, tels que

$$\Gamma_e + \sigma_b^2 \mathbf{H}\mathbf{H}^T = \sigma_0^2 \mathbf{I} \quad \text{avec} \quad \sigma_b^2 \leq \frac{\sigma_0^2}{\|\mathbf{H}\mathbf{H}^T\|}. \quad (10)$$

où  $\|\cdot\|$  est la norme spectrale d'une matrice. Cette décomposition permet de dériver l'algorithme *Iterative Shrinkage/Thresholding* pour les problèmes de Lasso [16]/Basis Pursuit Denoising [5]. Elle est aussi utilisée dans [9, 1] afin d'estimer les paramètres  $(\rho, \sigma_x^2)$  d'un modèle Bernoulli-Gaussien.

Pour maximiser  $p(\mathbf{q}|\mathbf{y})$ , considérons  $\mathbf{z}$  comme variable cachée admissible. L'algorithme EM s'écrit :

$$\mathbf{q}^{t+1} = \underset{\mathbf{q} \in \{0,1\}^N}{\operatorname{argmin}} \mathcal{Q}(\mathbf{q}, \mathbf{q}^t) \quad (11)$$

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

L'étape E et l'étape M sont développées dans la suite.

### 2.1 Etape-E

Dans un but de concision, la notation suivante sera adoptée :

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

pour faire référence à toute égalité à une constante  $\kappa$  près (qui ne dépend pas de  $\mathbf{q}$ ). Par la linéarité de l'espérance et le fait que  $p(\mathbf{q}|\mathbf{z}, \mathbf{y}) = p(\mathbf{q}|\mathbf{z})$ , on a

$$\mathcal{Q}(\mathbf{q}, \mathbf{q}^t) \stackrel{\kappa}{=} \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^t}[-\log p(\mathbf{q}|\mathbf{z})] \quad (12)$$

où  $\kappa = \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^t}[-\log p(\mathbf{z}|\mathbf{y})]$ . La règle de Bayes mène à :

$$\mathcal{Q}(\mathbf{q}, \mathbf{q}^t) \stackrel{\kappa}{=} \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^t}[-\log p(\mathbf{z}|\mathbf{q})] + \rho \|\mathbf{q}\|_0 \quad (13)$$

avec  $\rho$  défini en (2), avec maintenant  $\kappa = \mathbb{E}_{\mathbf{z}|\mathbf{y}, \mathbf{q}^t}[\log p(\mathbf{z}) - \log p(\mathbf{z}|\mathbf{y})]$ . En utilisant Eqs. (8) à (10), on a directement

$$\mathbf{z}|\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \Gamma_z(\mathbf{q})) \quad \text{avec} \quad \Gamma_z(\mathbf{q}) = \sigma_b^2 \mathbf{I} + \sigma_x^2 \mathbf{Q}^T \mathbf{Q}. \quad (14)$$

La règle de Bayes donne la loi *a posteriori* de  $\mathbf{z}$

$$\mathbf{z}|\mathbf{y}, \mathbf{q} \sim \mathcal{N}(\hat{\mathbf{z}}, \Sigma) \quad (15)$$

dont les paramètres sont

$$\begin{aligned} \hat{\mathbf{z}} &= \Gamma_z(\mathbf{q}) \mathbf{H}^T \Gamma_y^{-1}(\mathbf{q}) \mathbf{y} \\ \Sigma &= \Gamma_z(\mathbf{q}) - \Gamma_z(\mathbf{q}) \mathbf{H}^T \Gamma_y^{-1}(\mathbf{q}) \mathbf{H} \Gamma_z(\mathbf{q}). \end{aligned} \quad (16)$$

C'est pourquoi l'espérance s'écrit

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

Enfin, l'étape-E s'écrit comme la somme séparable

$$\mathcal{Q}(\mathbf{q}, \mathbf{q}^t) \stackrel{\kappa}{=} \frac{1}{2} \sum_{n=1}^N \mathcal{Q}_n(q[n]) \quad (17)$$

avec

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

Ainsi, la minimisation de  $\mathcal{Q}$  se réduit à  $N$  problèmes de minimisation 1D, *i.e.*, la minimisation des critères  $\mathcal{Q}_n$ .

### 2.2 Etape-M

Introduisons la notation  $\gamma_n^t = (\hat{z}^t[n])^2 + \Sigma^t[n, n]$ . La minimisation de  $\mathcal{Q}_n$  s'écrit :

$$q^{t+1}[n] = \underset{q \in \{0,1\}}{\operatorname{argmin}} \frac{\gamma_n^t}{\sigma_b^2 + \sigma_x^2 q^2} + \log(\sigma_b^2 + \sigma_x^2 q^2) + 2\rho q \quad (18)$$

Après calculs, la solution est  $q^{t+1}[n] = 1$  si et seulement si

$$\gamma_n^t \geq \sigma_b^2 \frac{\sigma_x^2 + \sigma_b^2}{\sigma_x^2} \left( \log \left( \frac{\sigma_b^2 + \sigma_x^2}{\sigma_b^2} \right) + 2\rho \right). \quad (19)$$

Cette opération n'est pas sans rappeler la formule de seuillage dans le cadre du problème de débruitage en MAP-Marginal (avec  $\gamma_n^t$  en lieu et place de  $z[n]$  ci-dessous).

**Proposition 2.1** ([13, 9, 1]). *Soit  $\mathbf{z} = \mathbf{Q}\mathbf{r} + \mathbf{b}$ , avec les lois a priori définies plus haut pour  $(\mathbf{q}, \mathbf{r})$ . L'estimateur du MMAP de  $\mathbf{q}|\mathbf{z}$  est donné explicitement, pour tout  $n$ , par*

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

$$\text{c.à.d. ssi} \quad z[n]^2 \geq 2\sigma_b^2 \frac{\sigma_x^2 + \sigma_b^2}{\sigma_x^2} \left( \rho + \frac{1}{2} \log \left( \frac{\sigma_b^2 + \sigma_x^2}{\sigma_b^2} \right) \right).$$

### 2.3 Résumé de l'algorithme

La méthode complète est résumé dans Alg. 1. Lorsque  $\mathbf{H} = \mathbf{I}$ ,  $\mathbf{e}$  peut être mis à  $\mathbf{0}$  dans (8), et donc  $\sigma_b^2 = \sigma_0^2$ . Dans ce cas, les matrices de covariance  $\Gamma_z$  et  $\Gamma_y$  sont égales, et l'algorithme se réduit au cas de l'estimateur du MMAP en débruitage (avec  $\mathbf{z} = \mathbf{y}$  dans Prop. 2.1). Dans la section suivante, une relaxation continue du problème MMAP est proposée, où les variables binaires  $q[n]$  sont remplacées par des variables dans  $[0, 1]$ . L'idée est d'obtenir une solution « à chaud » qui permettra de mieux initialiser Alg. 1.

---

**Algorithme 1** : Estimateur MMAP du support  $\mathbf{q}$ .

---

**Entrées** :  $t = 0, \mathbf{q}^{(t)} \in [0, 1]^N, \rho > 0$ ;  
**Sortie** :  $\mathbf{q} \in \{0, 1\}^N$ ;  
**while not converged do**  
     $\mathbf{Q}^{(t)} = \text{Diag}(\mathbf{q})$ ;  
     $\Gamma_z^{(t)} = \sigma_b^2 \mathbf{I} + \sigma_x^2 \mathbf{Q}^{(t)} \mathbf{Q}^{(t)T}$ ;  
     $\Gamma_y^{(t)} = \sigma_0^2 \mathbf{I} + \sigma_x^2 \mathbf{H} \mathbf{Q}^{(t)} \mathbf{Q}^{(t)T} \mathbf{H}^T$ ;  
     $\Sigma^{(t)} = \Gamma_z^{(t)} - \Gamma_z^{(t)} \mathbf{H}^T \Gamma_y^{(t)-1} \mathbf{H} \Gamma_z^{(t)}$ ;  
     $\hat{\mathbf{z}}^{(t)} = \Gamma_z^{(t)} \mathbf{H}^T \Gamma_y^{(t)-1} \mathbf{y}$ ;  
    **for**  $n = 1, \dots, N$  **do**  
         $\gamma_n^{(t)} = \hat{\mathbf{z}}^t[n]^2 + \Sigma^t[n, n]$ ;  
        **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**

---

### 3 Optimisation par relaxation continue

On relâche le problème d'optimisation binaire (11) en un problème d'optimisation continue. Le domaine de définition  $\{0, 1\}^N$  de la fonction coût est remplacé par  $[0, 1]^N$ . Comme la fonction coût  $\mathcal{Q}(\mathbf{q}, \mathbf{q}^t)$  est la somme découplée (17), l'étape M reste un problème 1D semblable à (18).

On remarque que pour  $q[n] \in \{0, 1\}$ , alors  $q[n]^2 = q[n]$ . Ainsi, minimiser  $\mathcal{Q}_n$  sur  $\{0, 1\}$  revient à minimiser :

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

Considérons la relaxation continue de  $q[n]$  sur  $[0, 1]$  :

$$q_n^{t+1} = \underset{q \in [0, 1]}{\text{argmin}} \frac{\gamma_n^{(t)}}{\sigma_b^2 + \sigma_x^2 q^2} + \log(\sigma_b^2 + \sigma_x^2 q^2) + 2\rho q^2. \quad (20)$$

Il s'agit d'un problème 1D avec des contraintes de borne. Le minimiseur non contraint peut être obtenu en cherchant les racines de la dérivée première par rapport à  $u = q^2$ , qui s'écrit

$$\frac{-\sigma_x^2 \gamma_n^{(t)}}{(\sigma_b^2 + \sigma_x^2 u^2)^2} + \frac{\sigma_x^2}{\sigma_b^2 + \sigma_x^2 u} + 2\rho = 0. \quad (21)$$

Lorsque  $\rho > 0$ , il existe une seule racine positive :

$$u[n] = \frac{1}{4\rho} \left( \sqrt{1 + 8\rho \frac{\gamma_n^{(t)}}{\sigma_x^2}} - 1 \right) - \frac{\sigma_b^2}{\sigma_x^2}. \quad (22)$$

Une étude minutieuse de la fonction coût dans (20) (considérée comme une fonction de  $u = q^2$ ) révèle que la dérivée seconde est positive pour le point  $u[n]$ . En prenant en compte que  $q \in [0, 1]$ , la solution devient :

$$q^{t+1}[n] = \begin{cases} 0 & \text{si } u[n] \leq 0, \\ 1 & \text{si } u[n] \geq 1, \\ \sqrt{u[n]} & \text{sinon.} \end{cases} \quad (23)$$

---

**Algorithme 2** : Estimation of continuous relaxed  $\mathbf{q}$ 

---

**Entrées** :  $t = 0, \mathbf{q}^{(t)} \in [0, 1]^N$  ;  
**Sortie** :  $\mathbf{q} \in [0, 1]^N$  ;  
**while not converged do**  
     $\mathbf{Q}^{(t)} = \text{Diag}(\mathbf{q})$  ;  
     $\Gamma_z^{(t)} = \sigma_b^2 \mathbf{I} + \sigma_x^2 \mathbf{Q}^{(t)} \mathbf{Q}^{(t)T}$  ;  
     $\Gamma_y^{(t)} = \sigma_0^2 \mathbf{I} + \sigma_x^2 \mathbf{H} \mathbf{Q}^{(t)} \mathbf{Q}^{(t)T} \mathbf{H}^T$  ;  
     $\Sigma^{(t)} = \Gamma_z^{(t)} - \Gamma_z^{(t)} \mathbf{H}^T \Gamma_y^{(t)-1} \mathbf{H} \Gamma_z^{(t)}$  ;  
     $\hat{\mathbf{z}}^{(t)} = \Gamma_z^{(t)} \mathbf{H}^T \Gamma_y^{(t)-1} \mathbf{y}$  ;  
    **for**  $n = 1, \dots, N$  **do**  
         $\gamma_n^{(t)} = \hat{\mathbf{z}}^t[n]^2 + \Sigma^t[n, n]$  ;  
         $u[n] = \frac{1}{4\rho} \left( \sqrt{1 + 8\rho \frac{\gamma_n^{(t)}}{\sigma_x^2}} - 1 \right) - \frac{\sigma_b^2}{\sigma_x^2}$  ;  
         $q[n] = \max\{\min\{\sqrt{u[n]}, 1\}, 0\}$  ;  
    **end**  
     $t = t + 1$  ;  
**end**

---

L'algorithme obtenu Alg. 2 est proche de l'EM Sparse Bayesian Learning (SBL) de [17], exception faite de l'estimation de  $\mathbf{q}$ . Comme  $\mathbf{x}|\mathbf{q} \sim \mathcal{N}(\mathbf{0}, \sigma_x^2 \mathbf{Q})$ , lorsque  $\mathbf{q}$  est continu, on retrouve la loi *a priori* utilisée dans SBL. La relaxation proposée peut être vue comme une approche SBL où les variances sont bornées par  $[0, \sigma_x^2]^N$ . Enfin, quand  $\rho = 0$ , on a  $p = \frac{1}{2}$  dans le modèle BG. Les mises à jour (20) de  $q[n]$  sont données par

$$q[n]^2 = \min \left\{ 1, \left( \frac{\gamma_n^{(t)} - \sigma_b^2}{\sigma_x^2} \right)^+ \right\} \quad (24)$$

qui conduit exactement à l'EM SBL [17] si  $\sigma_x^2 = 1$  et  $\sigma_b^2 = 0$ .

Concrètement, Alg. 2 est lancé pour différentes valeurs de l'hyperparamètre  $\rho$ . Les résultats sont utilisés comme initialisation à chaud de l'EM binaire pour le MMAP (Alg. 1) avec les vraies valeurs des paramètres du modèle. Cette initialisation permet d'éviter que Alg. 1 converge trop rapidement vers un mauvais minimum local.

### 4 Expériences

Le signal Bernoulli-Gaussien ( $\mathbf{q}, \mathbf{r}$ ) est simulé avec les paramètres  $p = 0.05$  et  $\sigma_x^2 = 1$ , et le bruit blanc gaussien est de variance  $\sigma_0^2 = 0.01$  ( ce qui correspond à un SNR d'entrée théorique d'environ 7 dB ). L'opérateur  $\mathbf{H}$  utilisé est une sous-matrice issue de l'opérateur direct (*leadfield*) en magnéto-encéphalographie (M/EEG), de taille  $272 \times 600$ . Cette matrice contient des colonnes hautement corrélées. Elle est représentative de la difficulté des problèmes inverses en M/EEG. L'estimation de  $\mathbf{q}$  au sens du MMAP est comparée aux estimations correspondantes  $\mathbf{x}(\mathbf{q})$  telles que dans (6), ainsi qu'au classique MAP Joint de  $\mathbf{q}$  et  $\mathbf{x}$  obtenu en minimisant (3). Pour ces méthodes, on utilise un algorithme de descente proximale similaire à l'algorithme d'Iterative Hard Thresholding (IHT) [2] avec redémarrage à chaud. Les comparaisons sont faites selon les métriques suivantes : le taux de Faux Positifs (erreur de Type-I), le taux de Faux Négatifs (erreur de Type-II), le SNR des  $\mathbf{x}(\mathbf{q})$  estimé selon (6) et calculé en prenant  $SNR = 10 \log_{10} \left( \frac{\|\mathbf{x}\|_2^2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2} \right)$ .

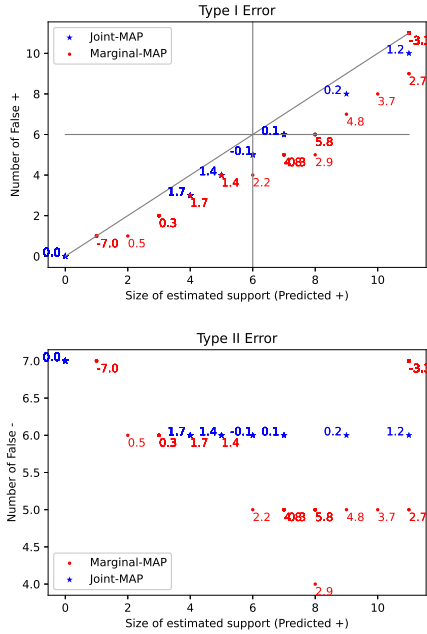


FIGURE 1 : Estimateurs du MAP marginal vs du MAP joint, comparés pour  $p = 0.01$  et  $\sigma_0^2 = 0.01$ . La mesure de SNR des signaux reconstruits  $\mathbf{x}$  est donnée pour certains points (de même que le score du meilleur SNR atteint).

Les erreurs de Type-I et de Type-II sont représentées en fonction de la taille du support prédit  $q$  en Fig. 1. On observe que la solution du MMAP est de meilleure qualité que celle du JMAP, que ce soit en terme d'erreur de type-I ou II et de SNR (surtout, pour les métriques Type-II et RSB). De plus, l'algorithme se comporte bien pour des opérateurs  $\mathbf{H}$  très corrélés. D'autres expérimentations (non montrées ici) montrent que pour des matrices  $\mathbf{H}$  gaussiennes, modérément corrélées, les estimateurs du JMAP et du MMAP donnent des résultats similaires. La version Marginale reste plus gourmande en temps de calcul à cause de la présence de l'inverse de la matrice  $\Sigma^t$  dans Algs. 1 et 2.

## 5 Discussion et conclusion

Cet article propose une approche par MMAP pour sélectionner le meilleur support possible, car il s'agit de l'estimateur bayésien pour l'erreur 0 – 1. Ce problème de sélection de support peut être vu comme un problème de sélection de variables très étudié en statistique [12], aussi connu comme *best subset selection problem* [8]. En classification, ce problème est en général étudié dans un contexte de « prédiction » de la classe de nouvelles données ; ainsi, l'estimation de  $\mathbf{x}$  est primordiale. Dans [8], une comparaison exhaustive est faite entre le Lasso et les solutions optimales  $\ell_0$  en utilisant un solveur MILP venant de [3]. Dans ce contexte, les résultats des approches  $\ell_0$  et Lasso sont similaires. Cependant, la matrice utilisée dans ces simulations n'est pas aussi mal posée que l'opérateur de M/EEG. Notre étude permet d'entamer une réflexion sur les relations existantes entre les minimiseurs du MAP Joint et du MMAP. Nos futurs travaux s'attaqueront à l'impact de la relaxation continue. De plus, une comparaison en profondeur de ces méthodes avec les approches MCMC appliquées aux a

priori BG [7] et les EMVS proposés par [14] (basés sur des mélanges de gaussiennes, et donc s'approchant de  $\mathbf{z}$  dans (8)). Du point de vue pratique, la méthode sera particularisée au cas de la structure spatio-temporelle des signaux issus de M/EEG [11].

## Références

- [1] P. BARBAULT, M. KOWALSKI et C. SOUSSEN : Parameter estimation in sparse inverse problems using Bernoulli-Gaussian prior. *In ICASSP*, pages 5413–5417, 2022.
- [2] T. BLUMENSATH et M. E. DAVIES : Iterative hard thresholding for compressed sensing. *Applied and Computational Harmonic Analysis*, 27(3):265–274, 2009.
- [3] S. BOURGUIGNON, J. NININ, H. CARFANTAN et M. MONGEAU : Exact sparse approximation problems via mixed-integer programming : Formulations and computational performance. *IEEE Trans. Signal Process.*, 64(6):1405–1419, 2015.
- [4] F. CHAMPAGNAT, Y. GOUSSARD et J. IDIER : Unsupervised deconvolution of sparse spike trains using stochastic approximation. *IEEE Trans. Signal Process.*, 44(12):2988–2998, Dec. 1996.
- [5] S. S CHEN, D. L DONOHO et M. A SAUNDERS : Atomic decomposition by basis pursuit. *SIAM review*, 43(1):129–159, 2001.
- [6] M. FIGUEIREDO et R. NOWAK : An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Process.*, 12(8): 906–916, 2003.
- [7] D. GE, J. IDIER et E. LE CARPENTIER : Enhanced sampling schemes for mcmc based blind Bernoulli-Gaussian deconvolution. *Signal Processing*, 91(4):759–772, 2011.
- [8] T. HASTIE, R. TIBSHIRANI et R. TIBSHIRANI : Best subset, forward stepwise or lasso ? analysis and recommendations based on extensive comparisons. *Statistical Science*, 35(4):579–592, 2020.
- [9] M. KOWALSKI et T. RODET : An unsupervised algorithm for hybrid/morphological signal decomposition. *In ICASSP*, pages 4112–4115, 2011.
- [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*, 51(2):642–653, 2010.
- [11] M. LIM, J. M. ALES, B. R. COTTEREAU, T. HASTIE et A. M. NORCIA : Sparse EEG/MEG source estimation via a group lasso. *PLoS ONE*, 12(6):e0176835, 2017.
- [12] R. B. O'HARA et M. J. SILLANPÄÄ : A review of Bayesian variable selection methods : What, how and which. *Bayesian Analysis*, 4(1):85–118, 2009.
- [13] M. PROTTER, I. YAVNEH et M. ELAD : Closed-form MMSE estimation for signal denoising under sparse representation modeling over a unitary dictionary. *IEEE Trans. Signal Process.*, 58(7):3471–3484, 2010.
- [14] V. ROVČKOVÁ et E. I. GEORGE : EMVS : The EM approach to Bayesian variable selection. *J. Am. Stat. Assoc.*, 109(506):828–846, 2014.
- [15] C. SOUSSEN, J. IDIER, D. BRIE et J. DUAN : From Bernoulli-Gaussian deconvolution to sparse signal restoration. *IEEE Trans. Signal Process.*, 59(10):4572–4584, 2011.
- [16] R. TIBSHIRANI : Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society : Series B (Methodological)*, 58(1):267–288, 1996.
- [17] D. P. WIPF et B. D. RAO : Sparse Bayesian learning for basis selection. *IEEE Trans. Signal Process.*, 52(8):2153–2164, 2004.