# BSS applied on EEG signal

SISEA - Problèmes Inverses

February 14, 2018

## 1 Prerequisites

We aime to reconstruct the unknown source vector $\boldsymbol{s}$ given its observation :

$$\boldsymbol{x} = \mathbf{A}\boldsymbol{s} + \epsilon, \tag{1}$$

where $\mathbf{A}$ is the known $M \times N$ mixing matrix ($M \ll N$), $\boldsymbol{x}$ the observation column vector ($M \times 1$), and $\epsilon$ the white Gaussian noise of unknown variances $\sigma_\epsilon^2$.
To impose sparsity in the solution, it is further supposed that $\boldsymbol{s}$ follows a Bernoulli-Gaussian (BG) model, an independent, identically distributed (iid) process defined in two stages. Firstly, the sparse nature is governed by the Bernoulli law:

$$P(\boldsymbol{q}) = \lambda^L (1-\lambda)^{N-L} \tag{2}$$

with $\boldsymbol{q} = [q_1, \ldots, q_N]^{\mathrm{t}}$ a binary sequence and $L = \sum_{n=1}^{N} q_n$, the number of non-zero entries of $\boldsymbol{q}$. Secondly, amplitudes $\boldsymbol{s} = [s_1, \ldots, s_N]^{\mathrm{t}}$ are assumed iid zero-mean Gaussian conditionally to $\boldsymbol{q}$:

$$\boldsymbol{s} \mid \boldsymbol{q} \sim \mathcal{N}(\mathbf{0}, \sigma_s^2 \operatorname{diag}(\boldsymbol{q})), \tag{3}$$

where $\operatorname{diag}(\boldsymbol{q})$ denotes a diagonal matrix whose diagonal is $\boldsymbol{q}$.
The problem becomes the estimation of $\Theta = \{\boldsymbol{s}, \boldsymbol{q}, \lambda, \sigma_\epsilon^2, \sigma_s^2\}$ given $\boldsymbol{x}$, for which the joint posterior distribution writes :

$$P(\Theta \,|\, \boldsymbol{x}) \propto \; g(\boldsymbol{x} - \mathbf{A}\boldsymbol{s}; \sigma_\epsilon^2 \mathbf{I}_M)\, P(\boldsymbol{q}; \lambda)\, g(\boldsymbol{s}; \sigma_s^2 \operatorname{diag}(\boldsymbol{q}))\, P(\sigma_\epsilon^2) P(\sigma_s^2) P(\lambda) \tag{4}$$

where $g(\cdot; \mathbf{R})$ denotes the centered Gaussian density of covariance $\mathbf{R}$. Conjugate prior laws are adopted for $\lambda$, $\sigma_\epsilon^2$ and $\sigma_s^2$:

$$\lambda \sim Be(1,1)?, \sigma_\epsilon^2 \sim IG(1,1)? \quad \sigma_\epsilon^2 \sim IG(1, 1\mathrm{e}{-4})?$$

where $IG$ and $Be$ represents respectively the **inverse Gamma** and **Beta** distributions.

## 2 Gibbs Sampler

According to the Monte Carlo principle, a posterior mean estimator of the unknown random variables can be approximated by:

$$\widehat{\Theta} = \frac{1}{I-J} \sum_{k=J+1}^{I} \Theta^{(k)}, \tag{5}$$

where the sum extends over the last $I-J$ samples. In the MCMC framework, the samples are generated iteratively, so that asymptotically converges in distribution to the joint posterior probability in eq. (4).

| Metropolis-Hastings algorithm |
|---|
| ① current configuration $\Theta^{(k)}$ |
| ② draw $\Theta'$ with a proposition law $Q(\Theta' \mid \Theta^{(k)})$ |
| ③ accept $\Theta^{(k+1)} = \Theta'$ with probability $\min\left\{1, \frac{P(\Theta')Q(\Theta^{(k)} \mid \Theta')}{P(\Theta^{(k)})Q(\Theta' \mid \Theta^{(k)})}\right\}$ |
| ④ repeat from ① |

| Gibbs algorithm |
|---|
| ① current configuration $\Theta^{(k)}$ |
| ② choose (cyclically or randomly) $i$ |
| ③ draw $\Theta_i^{(k+1)} \sim \Theta_i \mid \Theta^{(k)} \setminus \Theta_i^{(k)}, \boldsymbol{z}$ |
| ④ repeat from ① |

**Exercise 2.1** *Show that the Gibbs algorithm*

1. *is a special case of the* ***Metropolis-Hastings*** *algorithm for which $\Theta'$ is always accepted;*
2. *satisfies the equilibrium condition and thus $\Theta^{(k)}$ converges to $P(\Theta \mid \boldsymbol{z})$ in distribution.*

**Exercise 2.2** *write the following conditional probabilities from the joint probability in Eq* (4)*:*

1. $q_n \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}$,
2. $s_n \mid \Theta \setminus \{s_n\}, \boldsymbol{x}$,
3. $\sigma_\epsilon^2 \mid \Theta \setminus \sigma_\epsilon^2, \boldsymbol{x}$,
4. $\sigma_s^2 \mid \Theta \setminus \sigma_s^2, \boldsymbol{x}$,
5. $\lambda \mid \Theta \setminus \lambda, \boldsymbol{x}$,

*N.B. the Beta distribution writes $Beta(x;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}$, and the inverse Gamma $IG(x;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}x^{-(\alpha+1)}\exp(-\beta/x)$.*

Conditional laws:

We first detail the joint probability in Eq (4) by :

$$P(\Theta \mid \boldsymbol{x}) \propto \left(2\pi\sigma_\epsilon^2\right)^{-M/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{x} - \mathbf{A}\boldsymbol{s}\|^2\right\} \cdot \lambda^{\sum q_n}(1-\lambda)^{N-\sum q_n}$$
$$\cdot \left((2\pi)^N \det(\sigma_s^2 \operatorname{diag}(\boldsymbol{q}))\right)^{-1/2} \exp\left\{-\frac{1}{2}\boldsymbol{s}^{\mathrm{t}}(\sigma_s^2 \operatorname{diag}(\boldsymbol{q}))^{-1}\boldsymbol{s}\right\} \cdot P(\sigma_s^2) \cdot P(\sigma_\epsilon^2).$$

Notice that $\det(\sigma_s^2 \operatorname{diag}(\boldsymbol{q})) = (\sigma_s^2)^N \prod q_n$ due to its diagonal structure, while a zero-variance normal distribution can be considered a Dirac distribution. Notice also that for each binary variable $q_n$, we should have the normalization constraint

$$P(q_n = 0 \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}) + P(q_n = 1 \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}) = 1. \tag{6}$$

Its conditional probability can thus be derived by calculating respectively the two terms up to a normalization factor :

- for $P(q_n = 0 \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x})$, according to the B-G model, $s_n = 0$, thus

  $$P(q_n = 0, s_n \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{x} - \mathbf{A}_{-n}\boldsymbol{s}_{-n} - \mathbf{A}_n s_n\|^2\right\} \cdot (1-\lambda) \cdot \delta_0(s_n);$$

  for which the $\mathbf{A}_{-n}$ and $\boldsymbol{s}_{-n}$ represents the matrix $\mathbf{A}$ without the $n$-th column and the vector $\boldsymbol{s}$ without its $n$-th element respectively; we can easily integrate out $s_n$ and have the marginal form:

  $$P(q_n = 0 \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{x} - \mathbf{A}_{-n}\boldsymbol{s}_{-n}\|^2\right\} \cdot (1-\lambda); \tag{7}$$

- likewise for $P(q_n = 1, s_n \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x})$:

$$P(q_n = 1, s_n \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{x} - \mathbf{A}_{-n}\boldsymbol{s}_{-n} - \mathbf{A}_n s_n\|^2\right\} \cdot \lambda \cdot \frac{1}{\sqrt{2\pi\sigma_s^2}} \exp(-\frac{s_n^2}{2\sigma_s^2})$$

  the quadratic form inside the exponential terms suggests that $s_n$ follows a normal distribution ($\mathcal{N}(\mu_n, \sigma_n^2)$) and can be integrated out analytically to yield :

$$P(q_n = 1 \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{x} - \mathbf{A}_{-n}\boldsymbol{s}_{-n}\|^2\right\} \cdot \frac{\lambda\sigma_n}{\sigma_s} \cdot \exp\left\{\frac{\mu_n^2}{2\sigma_n^2}\right\}, \tag{8}$$

  where $\sigma_n^2 = \frac{\sigma_\epsilon^2\sigma_s^2}{\sigma_\epsilon^2 + \sigma_s^2\|\mathbf{A}_n\|^2}$ and $\mu_n = \frac{\sigma_n^2}{\sigma_\epsilon^2}\mathbf{A}_n^{\mathrm{t}}(\boldsymbol{x} - \mathbf{A}_{-n}\boldsymbol{s}_{-n})$.

By combining eqs. (6)(7)(8), we conclude that $q_n \mid \Theta \setminus \{q_n, s_n\}, \boldsymbol{x} \sim \mathrm{Bi}(\frac{\nu_n}{\nu_n + 1 - \lambda})$ for which $\nu_n = \frac{\lambda\sigma_n}{\sigma_s}\exp\left\{\frac{\mu_n^2}{2\sigma_n^2}\right\}$.
The conditional probability of $s_n \mid \Theta \setminus \{s_n\}, \boldsymbol{x}$ is straight forward : if $q_n = 0$, then $s_n = 0$, else $s_n \sim \mathcal{N}(\mu_n, \sigma_n^2)$ as derived previously. The conditional probabilities of the rest three parameters are obvious thanks to the conjugate prior laws.

**Exercise 2.3** *From the Gibbs sampling scheme, derive the pseudocode in Tab. 1*

Table 1: Pseudocode for the Gibbs sampler

`% Initialization`
$\boldsymbol{q} \leftarrow \mathbf{0}$; $\boldsymbol{s} \leftarrow \mathbf{0}$
Sample $\sigma_\epsilon^2, \sigma_s^2, \lambda$ using prior laws
**repeat**
  `% ----------- Step 1: Sample` $(q_i, s_i)$ `-----------`
  $\boldsymbol{e} \leftarrow \boldsymbol{x} - \mathbf{A}\boldsymbol{s}$
  **for** $i = 1$ to $N$ **do**
    $\boldsymbol{e}_i \leftarrow \boldsymbol{e} + \mathbf{A}_i s_i$ `%` $\mathbf{A}_i$ `is the` $i$`-th column of` $\mathbf{A}$
    $\sigma_i^2 \leftarrow \sigma_\epsilon^2\sigma_s^2/(\sigma_\epsilon^2 + \sigma_s^2\|\mathbf{A}_i\|^2)$
    $\mu_i \leftarrow (\sigma_i^2/\sigma_\epsilon^2)\mathbf{A}_i^{\mathrm{t}}\boldsymbol{e}_i$
    $\nu_i \leftarrow \lambda(\sigma_i/\sigma_s)\exp(\mu_i^2/(2\sigma_i^2))$
    $\lambda_i \leftarrow \nu_i/(\nu_i + 1 - \lambda)$
    Sample $q_i \sim Bi(\lambda_i)$
    Sample $s_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$ if $q_i = 1$, $s_i = 0$ otherwise
    $\boldsymbol{e} \leftarrow \boldsymbol{e}_i - \mathbf{A}_i s_i$ `% Update` $\boldsymbol{e}$
  **end for**
  `% ----------- Step 2: Sample` $\sigma_\epsilon^2$ `-----------`
  Sample $\sigma_\epsilon^2 \sim \mathrm{IG}(M/2{+}?, \|\boldsymbol{x} - \mathbf{A}\boldsymbol{s}\|^2/2{+}?)$
  `% ----------- Step 3: Sample` $\lambda$ `------------`
  Sample $\lambda \sim \mathrm{Be}(1 + L, 1 + N - L)$, with $L = \sum_n q_n$
  `% ----------- Step 4: Sample` $\sigma_s^2$ `-----------`
  Sample $\sigma_s^2 \sim \mathrm{IG}(L/2{+}?, \|\boldsymbol{s}\|^2/2{+}?)$
**until** Convergence

## 3 Matlab Experiment :

1. launch the *TP_inverse_problem.m* file in matlab and visualize the source distribution;
2. understand the direct problem process (code given) of linear source mixture and noise addition;
3. use the Gibbs sampler to perform the "deconvolution" and compare the results.