

Analyse Factorielle parcimonieuse avec priors Bernoulli-Gaussiens sur les facteurs

Décembre 2018

I. Modèle 1 : Parcimonie par priors exponentiels

L'objectif est d'améliorer le modèle suivant :

$$\mathbf{y}_i = \mathbf{L} \lceil \mathbf{q} \rceil \mathbf{c}_i + \mathbf{n}_i \quad (1)$$

Priors	Hyperpriors
$\mathbf{L} \sim \mathcal{N}_c(0, \frac{\mathbf{I}_{MK}}{K})$	
$\mathbf{c} \sim \mathcal{N}_c(0, \lceil \gamma^2 \rceil)$	$\gamma^2 \sim \mathcal{IG}(a_\gamma, b_\gamma)$
$\mathbf{q} \sim \mathcal{Exp}(\mathbf{a}_q)$	
$\mathbf{n} \beta^2 \sim \mathcal{N}_c(0, \beta^2 \lceil \sigma^2 \rceil)$	$\sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$ $\beta^2 \sim \mathcal{IG}(a_\beta, b_\beta)$

Ce modèle pose les problèmes résumés dans le tableau ci-dessous :

Faiblesses rencontrées dans le modèle 1	Solutions proposées par le modèle 2
\mathbf{q} n'est pas normalisé, on ne peut donc pas observer la stabilité de la chaîne	\mathbf{q} vaut 0 ou 1 ; toute la variance est portée par \mathbf{c}
\mathbf{q} et \mathbf{c} risquent d'être fortement corrélés, ce qui peut nuire à la mélangeance des chaînes	Marginalisation partielle de l'échantillonneur
Sensibilité des résultats au choix des hyperparamètres \mathbf{a}_q	Échantillonnage de l'hyperparamètre dans une loi Bêta.

II. Bibliographie

Le chapitre 4 de la thèse de Charly Faure y est dédié, avec le modèle suivant :

$$\mathbf{y} = \mathbf{H} \mathbf{f} + \mathbf{n} \quad (2)$$

où $\mathbf{f} \sim \prod_i \mathcal{BernGauss}(f_i ; l, \gamma^2)$. L'échantillon f_i est donc tiré soit dans une Gaussienne, soit dans un Dirac en 0 (cas particulier d'une Gaussienne de variance nulle). Ce qui peut s'écrire :

$$\mathcal{BernGauss}(l, \gamma^2) = (1 - l) \mathcal{N}_c(0, 0) + l \mathcal{N}_c(0, \gamma^2) \quad (3)$$

? ont une approche similaire, en notant $\mathbf{f} \sim \mathcal{N}_c(0, \lceil \mathbf{q} \rceil \gamma^2)$, où les éléments de \mathbf{q} suivent une loi de Bernoulli.

III. Calculs

3.1. Nomenclature

M	Nombre de micros
K	Nombre de facteurs
I_s	Nombre de moyennes pour le calculs des CSM
i	indice associé à une moyenne
L_q	nombre d'éléments non-nuls du vecteur \mathbf{q}

3.2. Modèle 2 : Parcimonie par priors bernoulliens

On choisit un modèle hiérarchique avec la matrice de mixage \mathbf{L} , les facteurs \mathbf{c} et les binaires \mathbf{q} au même étage :

$$\mathbf{y}_i = \mathbf{L} \lceil \mathbf{q} \rceil \mathbf{c}_i + \mathbf{n}_i \quad (4)$$

Priors	Hyperpriors
$\mathbf{L} \sim \mathcal{N}_c(0, \frac{\mathbf{I}_{MK}}{K})$	
$\mathbf{c} \sim \mathcal{N}_c(0, \lceil \gamma^2 \rceil)$	$\gamma^2 \sim \mathcal{IG}(a_\gamma, b_\gamma)$
$\mathbf{q} \sim \mathcal{Bern}(l)$	$l \sim \mathcal{Beta}(a_l, b_l)$
$\mathbf{n} \beta^2 \sim \mathcal{N}_c(0, \beta^2 \lceil \sigma^2 \rceil)$	$\sigma^2 \sim \mathcal{IG}(a_\sigma, b_\sigma)$ $\beta^2 \sim \mathcal{IG}(a_\beta, b_\beta)$

On note par la suite $\mathbf{\Omega}_n = \beta^2 \lceil \sigma^2 \rceil$. Le paramètre β est introduit par anticipation d'un modèle adapté au cas multirégime d'Airbus.

3.3. Posteriors

3.3.1 Échantillonnage du paramètre de parcimonie l

$$\begin{aligned}
[l | \mathbf{q}] &\propto \mathcal{Beta}(l; a_l, b_l) \prod_k^K \mathcal{Bern}(q_k; l) \\
&\propto \frac{\Gamma(a_l + b_l)}{\Gamma(a_l)\Gamma(b_l)} l^{a_l-1} (1-l)^{b_l-1} \prod_k^K l^{q_k} (1-l)^{1-q_k} \\
&\propto \frac{\Gamma(a_l + b_l)}{\Gamma(a_l)\Gamma(b_l)} l^{a_l-1} (1-l)^{b_l-1} l^{\sum_k q_k} (1-l)^{\sum_k 1-q_k} \\
&\propto \mathcal{Beta}(a_l + L_q, b_l + K - L_q)
\end{aligned}$$

3.3.2 Échantillonnage en bloc de \mathbf{c}

$$\begin{aligned}
[\mathbf{c}_i | \infty_{-\mathbf{c}_i}] &\propto [\mathbf{y}_i | \infty_{-\mathbf{y}_i}] \mathcal{N}_c(\mathbf{c}_i; 0, \mathbf{q}\boldsymbol{\gamma}^2) \\
&\propto \frac{e^{-(\mathbf{y}_i - \mathbf{L}\mathbf{q}\mathbf{c}_i)^H \boldsymbol{\Omega}_n^{-1} (\mathbf{y}_i - \mathbf{L}\mathbf{q}\mathbf{c}_i)}}{|\boldsymbol{\Omega}_n|} e^{\mathbf{c}_i^H \boldsymbol{\gamma}^{-2} \mathbf{c}_i} \\
&\propto e^{-(\mathbf{c}_i - \boldsymbol{\mu}_i)^H \boldsymbol{\Omega}_{\mathbf{c}_i}^{-1} (\mathbf{c}_i - \boldsymbol{\mu}_i)}
\end{aligned}$$

Par identification, on a :

- $\mathbf{c}_i^H \boldsymbol{\Omega}_{\mathbf{c}_i}^{-1} \mathbf{c}_i = \mathbf{c}_i^H (\mathbf{q}\mathbf{L}^H \boldsymbol{\Omega}_n^{-1} \mathbf{L}\mathbf{q} + \boldsymbol{\gamma}^{-2}) \mathbf{c}_i$
- $\boldsymbol{\Omega}_{\mathbf{c}_i}^{-1} \boldsymbol{\mu}_i = \mathbf{q}^H \mathbf{L}^H \boldsymbol{\Omega}_n^{-1} \mathbf{y}_i$

Finalement :

$$[\mathbf{c}_i | \infty_{-\mathbf{c}_i}] \propto \mathcal{N}_c(\boldsymbol{\mu}_i, \boldsymbol{\Omega}_{\mathbf{c}_i}) \quad (5)$$

où : $\boldsymbol{\Omega}_{\mathbf{c}_i} = (\mathbf{q}\mathbf{L}^H \boldsymbol{\Omega}_n^{-1} \mathbf{L}\mathbf{q} + \boldsymbol{\gamma}^{-2})^{-1}$
 et $\boldsymbol{\mu}_i = \boldsymbol{\Omega}_{\mathbf{c}_i} \mathbf{q}^H \mathbf{L}^H \boldsymbol{\Omega}_n^{-1} \mathbf{y}_i$

Dans le cas où le nombre de moyenne est suffisamment grand, on peut échantillonner directement la matrice interspectrale $\mathbf{S}_{cc} = \mathbb{E}\{\mathbf{c}_i \mathbf{c}_i^H\}$. On a :

$$\mathbf{c}_i = \boldsymbol{\mu}_i + \mathbf{x}_i \quad \text{avec} \quad \mathbf{x}_i \sim \mathcal{N}_c(0, \boldsymbol{\Omega}_{\mathbf{c}_i}),$$

donc :

$$\mathbb{E}\{\mathbf{c}_i \mathbf{c}_i^H\} = \mathbb{E}\{\boldsymbol{\mu}_i \boldsymbol{\mu}_i^H\} + \mathbb{E}\{\mathbf{x}_i \mathbf{x}_i^H\} + \underbrace{2\mathbb{E}\{\mathbf{x}_i \boldsymbol{\mu}_i^H\}}_{\rightarrow 0 \text{ quand } I_s \rightarrow \infty},$$

avec :

- $\mathbb{E}\{\mathbf{x}_i \mathbf{x}_i^H\} = \mathbf{W}_c \sim \mathcal{W}(\boldsymbol{\Omega}_{\mathbf{c}_i}, I_s)$
- $\mathbb{E}\{\boldsymbol{\mu}_i \boldsymbol{\mu}_i^H\} = \boldsymbol{\Omega}_{\mathbf{c}_i} \mathbf{q}^H \mathbf{L}^H \boldsymbol{\Omega}_n^{-1} \mathbf{S}_{yy} \mathbf{L}\mathbf{q} \boldsymbol{\Omega}_{\mathbf{c}_i}^H$

3.3.3 Échantillonnage de \mathbf{L}

La matrice \mathbf{L} est vectorisée de manière à écrire une matrice de covariance pour la postérieure qui soit à 2 dimensions. On note $\boldsymbol{\lambda} = \text{vec}(\mathbf{L})$

$$[\boldsymbol{\lambda} | \infty_{-\boldsymbol{\lambda}}] \propto \prod_i^{I_s} [\mathbf{y}_i | \infty_{-\mathbf{y}_i}] [\boldsymbol{\lambda}]$$

Or :

$$\text{vec}(\mathbf{y}_i) = \mathbf{y}_i = \text{vec}(\mathbf{L}\mathbf{q}\mathbf{c}_i) + \mathbf{n}_i = (\mathbf{c}_i^T \mathbf{q} \otimes \mathbf{I}_M) \boldsymbol{\lambda} + \mathbf{n}_i, \quad (6)$$

donc :

$$\begin{aligned} [\boldsymbol{\lambda}|\infty_{-\boldsymbol{\lambda}}] &\propto e^{-\sum_i (\mathbf{y}_i - (\mathbf{c}_i^T \mathbf{q} \otimes \mathbf{I}_K) \boldsymbol{\lambda})^H \boldsymbol{\Omega}_n^{-1} (\mathbf{y}_i - (\mathbf{c}_i^T \mathbf{q} \otimes \mathbf{I}_M) \boldsymbol{\lambda})} e^{-\boldsymbol{\lambda}^H K \mathbf{I}_{MK} \boldsymbol{\lambda}} \\ &\propto e^{-(\boldsymbol{\lambda} - \boldsymbol{\mu}_\lambda)^H \boldsymbol{\Omega}_\lambda^{-1} (\boldsymbol{\lambda} - \boldsymbol{\mu}_\lambda)} \end{aligned}$$

Comme précédemment, par identification, on a finalement :

$$[\boldsymbol{\lambda}|\infty_\lambda] \propto \mathcal{N}_c(\boldsymbol{\mu}_\lambda, \boldsymbol{\Omega}_\lambda) \quad (7)$$

où :

$$\begin{aligned} \boldsymbol{\Omega}_\lambda &= \left(\sum_i (\mathbf{c}_i^T \mathbf{q} \otimes \mathbf{I}_M)^H \boldsymbol{\Omega}_n^{-1} (\mathbf{c}_i^T \mathbf{q} \otimes \mathbf{I}_M) + K \mathbf{I}_{MK} \right)^{-1} \\ &= ((\mathbf{q} \mathbf{S}_{cc}^* \mathbf{q}) \otimes (\boldsymbol{\Omega}_n^{-1}) + K \mathbf{I}_{MK})^{-1} \\ \text{et } \boldsymbol{\mu}_\lambda &= \boldsymbol{\Omega}_\lambda \sum_i (\mathbf{c}_i^T \mathbf{q} \otimes \mathbf{I}_M)^H \boldsymbol{\Omega}_n^{-1} \mathbf{y}_i \\ &= \boldsymbol{\Omega}_\lambda \text{vec}(\boldsymbol{\Omega}_n^{-1} \mathbf{S}_{yy} \boldsymbol{\Omega}_n^{-1} \mathbf{L} \mathbf{q} \boldsymbol{\Omega}_{c_i} \mathbf{q}) \end{aligned}$$

Notes : - Ce dernier résultats s'obtient en remplaçant \mathbf{c}_i par $\boldsymbol{\mu}_i$.

- Les propriétés du produit de Kronecker sont les suivantes :

$$\begin{aligned} \text{vec}(ABC) &= (C^T \otimes A) \text{vec}(B) \\ (A \otimes B)^T &= A^T \otimes B^T \\ (A \otimes B)^* &= A^* \otimes B^* \\ (A \otimes B)(C \otimes D) &= (AC) \otimes (BD) \end{aligned}$$

si les dimensions des matrices A , B , C et D le permettent.

3.3.4 Échantillonnage de γ^2

Cas où $\gamma^2 = \gamma^2 \mathbf{I}_K$

$$\begin{aligned} [\gamma^2 | \mathbf{S}_{cc}] &\propto \prod_i [\mathbf{c}_i | \gamma^2] [\gamma^2] \\ &\propto \frac{e^{-\gamma^{-2} \sum_i \mathbf{c}_i^H \mathbf{c}_i} e^{-\gamma^{-2} b_\gamma}}{\gamma^{2I_s K} \gamma^{2(a_\gamma - 1)}} \end{aligned}$$

$[\gamma^2 | \mathbf{S}_{cc}] \propto \mathcal{IG}(a_\gamma + K I_s, b_\gamma + \text{Trace}(\mathbf{S}_{cc}))$

(8)

Cas où $\gamma^2 = [\gamma^2] = \text{diag}(\gamma_1^2, \dots, \gamma_k^2, \dots, \gamma_K^2)$

$$[\gamma_k^2 | \mathbf{S}_{cc}] \propto \prod_i [\mathbf{c}_{ik} | \gamma_k^2] [\gamma_k^2]$$

$[\gamma_k^2 | \mathbf{S}_{cc}] \propto \mathcal{IG}(a_\gamma + I_s, b_\gamma + \mathbf{S}_{cc_{kk}})$

(9)

3.3.5 Échantillonnage de σ^2

On considère un bruit hétéroskédastique : $\sigma^2 = \text{diag}(\sigma_1^2, \dots, \sigma_m^2, \dots, \sigma_M^2)$.

$$\begin{aligned} [\sigma^2 | \infty_{-\sigma^2}] &\propto \prod_i [\mathbf{y}_i | \infty_{-\mathbf{y}_i}] [\sigma] \\ &\propto \frac{e^{-\sum_i (\mathbf{y}_i - \mathbf{L} \mathbf{q} \mathbf{c}_i)^H \beta^{-2} \sigma^{-2} (\mathbf{y}_i - \mathbf{L} \mathbf{q} \mathbf{c}_i)} e^{-\sigma^{-2} b_\sigma}}{\beta^2 |\sigma|^{I_s}} \frac{e^{-\sigma^{-2} b_\sigma}}{\sigma^{2(a_\sigma-1)}} \\ &\propto \frac{e^{-\text{Trace} \left(\beta^{-2} \sigma^{-2} \sum_i \left\{ \mathbf{y}_i \mathbf{y}_i^H + \mathbf{L} \mathbf{q} \mathbf{c}_i \mathbf{c}_i^H \mathbf{q} \mathbf{L}^H - \mathbf{y}_i \mathbf{c}_i^H \mathbf{q} \mathbf{L}^H - \mathbf{L} \mathbf{q} \mathbf{c}_i \mathbf{y}_i^H \right\} \right)}}{|\sigma^2|^{I_s}} \frac{e^{-\sigma^{-2} b_\sigma}}{\sigma^{2(a_\sigma-1)}} \end{aligned}$$

On réinjecte $\mathbf{c}_i = \boldsymbol{\mu}_i + \mathbf{x}_i$ et on suppose que $\mathbb{E}(\mathbf{x}_i \mathbf{y}_i^H) \approx 0$. En notant :

$$\mathbf{P} = \mathbf{L} \mathbf{q} \boldsymbol{\Omega}_{c_i} \mathbf{q} \mathbf{L}^H \boldsymbol{\Omega}_n^{-1} = \mathbf{P}^H,$$

on a : $\mathbf{L} \mathbf{q} \boldsymbol{\mu}_i = \mathbf{P} \mathbf{y}_i$. Finalement :

$$\begin{aligned} \mathbf{T} &= \mathbf{S}_{yy} + \mathbf{L} \mathbf{q} \mathbf{W}_c \mathbf{q} \mathbf{L}^H + \mathbf{P} \mathbf{S}_{yy} \mathbf{P} - \mathbf{S}_{yy} \mathbf{P} - \mathbf{P} \mathbf{S}_{yy} \\ &= (\mathbf{I}_M - \mathbf{P}) \mathbf{S}_{yy} (\mathbf{I}_M - \mathbf{P}) + \mathbf{L} \mathbf{q} \mathbf{W}_c \mathbf{q} \mathbf{L}^H \end{aligned}$$

$$\boxed{[\sigma_m^2 | \infty_{-\sigma_m^2}] \propto \mathcal{IG}(a_\sigma + M I_s, b_\sigma + \beta^{-2} \mathbf{T}_{mm})} \quad (10)$$

3.3.6 Échantillonnage de β^2

De manière similaire à l'échantillonnage de σ_m :

$$\begin{aligned} [\beta^2 | \infty_{-\beta^2}] &\propto \prod_i [\mathbf{y}_i | \infty_{-\mathbf{y}_i}] [\beta^2] \\ &\propto \frac{e^{-\beta^{-2} \text{Trace}(\sigma^{-2} \mathbf{T})} e^{-\beta^{-2} b_\beta}}{|\beta^2 \mathbf{I}_M|^{I_s}} \frac{e^{-\beta^{-2} b_\beta}}{\beta^{2(a_\beta-1)}} \end{aligned}$$

$$\boxed{[\beta^2 | \infty_{-\beta^2}] \propto \mathcal{IG}(a_\beta + M I_s, b_\beta + \text{Trace}(\sigma^{-2} \mathbf{T}))} \quad (11)$$

3.3.7 Échantillonnage des binaires du vecteur \mathbf{q}

Sans marginalisation

$$\begin{aligned} [q_k | \infty_{-q_k}] &\propto \prod_i [\mathbf{y}_i | q_k, \infty_{-\mathbf{y}_i}] [q_k] \\ &\propto \frac{e^{-\sum_i (\mathbf{y}_i - \mathbf{L} \mathbf{q} \mathbf{c}_i)^H \boldsymbol{\Omega}_n^{-1} (\mathbf{y}_i - \mathbf{L} \mathbf{q} \mathbf{c}_i)}}{|\boldsymbol{\Omega}_n|^{I_s}} l^{q_k} (1-l)^{1-q_k} \\ &\propto e^{-\text{Trace}(\boldsymbol{\Omega}_n^{-1} \mathbf{T}(q_k)) - q_k \ln(\frac{1}{l} - 1)} \\ &\propto e^{-g(q_k)} \end{aligned}$$

Le changement d'état d'un binaire se traduit par l'ajout de la quantité $\delta_k = (-1)^{q_k}$. On note $q_{k_{mod}} = q_k + \delta_k$. La probabilité que le binaire change d'état est :

$$\begin{aligned} [q_{k_{mod}} | \infty_{-q_k}] &\propto \prod_i [\mathbf{y}_i | q_{k_{mod}}, \infty_{-\mathbf{y}_i, \mathbf{c}}] [q_{k_{mod}}] \\ &\propto e^{-\text{Trace}(\mathbf{\Omega}_n^{-1} \mathbf{T}(q_{k_{mod}})) - q_{k_{mod}} \ln(\frac{1}{l} - 1)} \\ &\propto e^{-g(q_{k_{mod}})} \end{aligned}$$

Pour que la somme de ces deux probabilité soit égale à 1, il faut les normaliser. La probabilité qu'un binaire change d'état est alors :

$$\begin{aligned} [q_{k_{mod}} | \infty_{-q_k}] &= \frac{e^{-g(q_{k_{mod}})}}{e^{-g(q_{k_{mod}})} + e^{-g(q_k)}} \\ &= \frac{1}{1 + e^{-(g(q_k) - g(q_{k_{mod}}))}} \end{aligned}$$

Un échantillon $t \sim \mathcal{U}(0, 1)$ est comparé à $[q_{k_{mod}} | \infty_{-q_k}]$. S'il y est inférieur, le binaire change d'état, sinon il conserve son état actuel.

En pratique, comme \mathbf{T} peut prendre de grandes valeurs (not. si I_s est grand), le changement est accepté si

$$-\ln\left(\frac{1}{t} - 1\right) < g(q_k) - g(q_{k_{mod}})$$

Avec marginalisation sur les facteurs \mathbf{c}

$$[q_k | \infty_{-q_k, -\mathbf{c}}] \propto \prod_i [\mathbf{y}_i | q_k, \infty_{-\mathbf{y}_i, -\mathbf{c}}] [q_k]$$

Or,

$$[\mathbf{y}_i | q_k, \infty_{-\mathbf{y}_i, -\mathbf{c}}] = \mathcal{N}_c(0, \underbrace{\mathbf{L} \mathbf{q} \mathbf{\gamma}^2 \mathbf{q} \mathbf{L}^H}_{\mathbf{B}} + \mathbf{\Omega}_n) \quad (12)$$

Ce résultat est démontré en annexe ??.

La posterior marginalisée sur le binaire q_k est donc :

$$\begin{aligned} [q_k | \infty_{-q_k, -\mathbf{c}}] &\propto \frac{e^{-\sum_i \mathbf{y}_i^H \mathbf{B}(q_k)^{-1} \mathbf{y}_i}}{|\mathbf{B}(q_k)|^{I_s}} l^{q_k} (1 - l)^{1 - q_k} \\ &\propto e^{-\text{Trace}(\mathbf{B}(q_k)^{-1} \mathbf{S}_{yy}) - I_s \ln |\mathbf{B}(q_k)| - q_k \ln(\frac{1}{l} - 1)} \\ &\propto e^{-g(q_k)} \end{aligned}$$

De la même manière, la probabilité que le binaire change d'état est :

$$\begin{aligned} [q_{k_{mod}} | \infty_{-q_{k_{mod}}, -\mathbf{c}}] &\propto \frac{e^{-\sum_i \mathbf{y}_i^H \mathbf{B}(q_{k_{mod}})^{-1} \mathbf{y}_i}}{|\mathbf{B}(q_{k_{mod}})|^{I_s}} l^{q_{k_{mod}}} (1 - l)^{1 - q_{k_{mod}}} \\ &\propto e^{-g(q_{k_{mod}})} \end{aligned}$$

Après normalisation, le changement d'état du binaire est accepté si

$$-\ln\left(\frac{1}{t} - 1\right) < g(q_k) - g(q_{k_{mod}})$$

IV. Résultats et analyse

4.1. Paramètres de la simulation

- $M = 20$
- $I_s = 10^4$
- SNR = 0 et -10 dB
- Nombre de fonction de base : 5
- Nombre de facteurs recherché : $M - 1$
- $\text{Trace}(\mathcal{S}_{pp})/M = 2.4$

4.2. Sans marginalisation

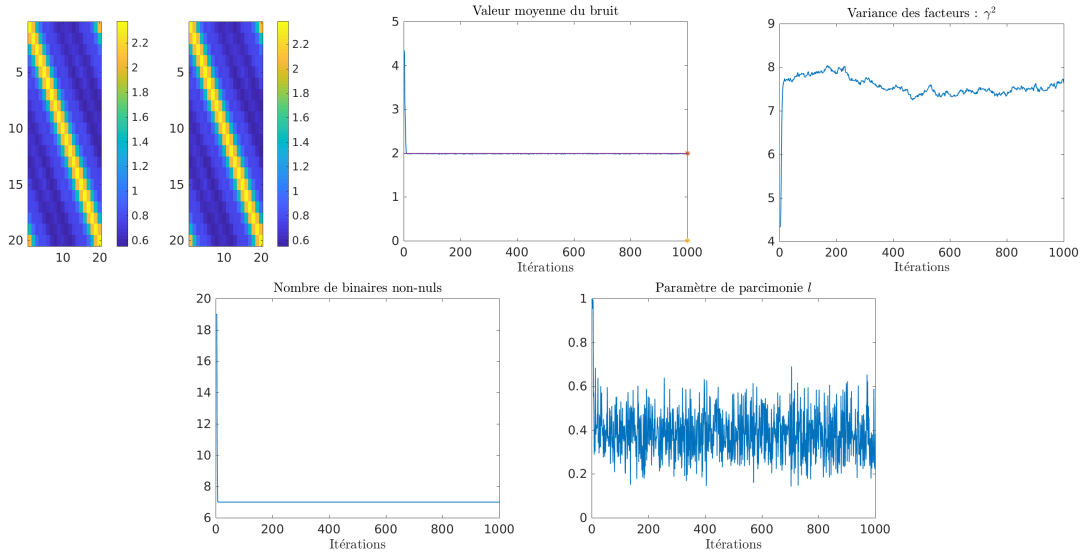


FIGURE 1 – $SNR = 0dB$

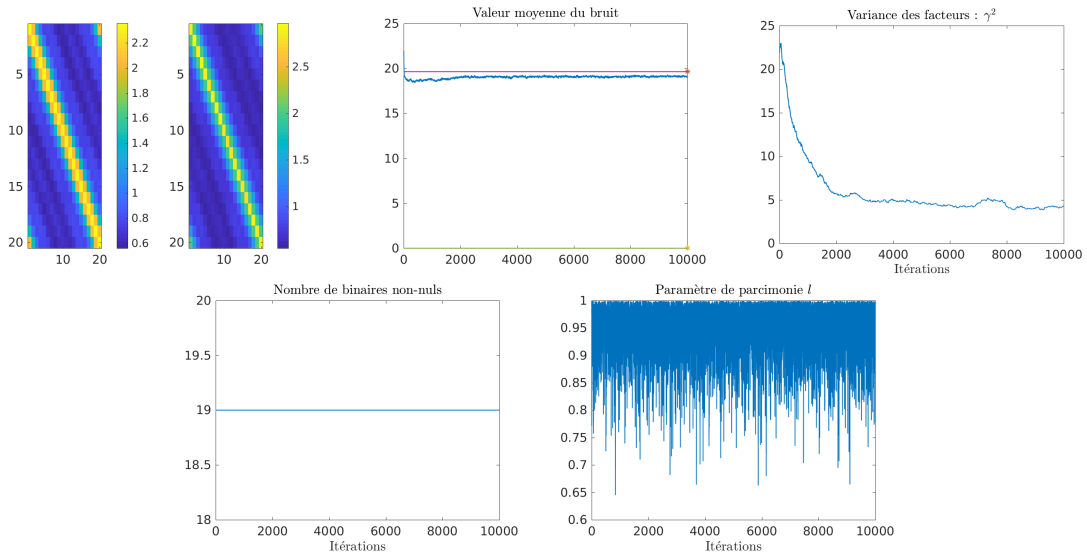


FIGURE 2 – $SNR = -10dB$

4.3. Avec marginalisation

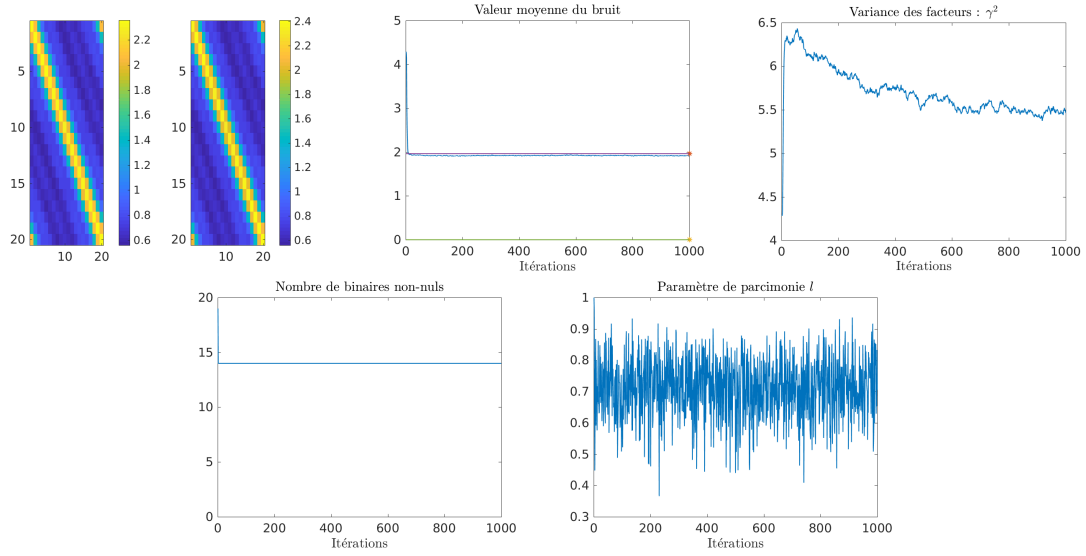


FIGURE 3 – $SNR = 0dB$

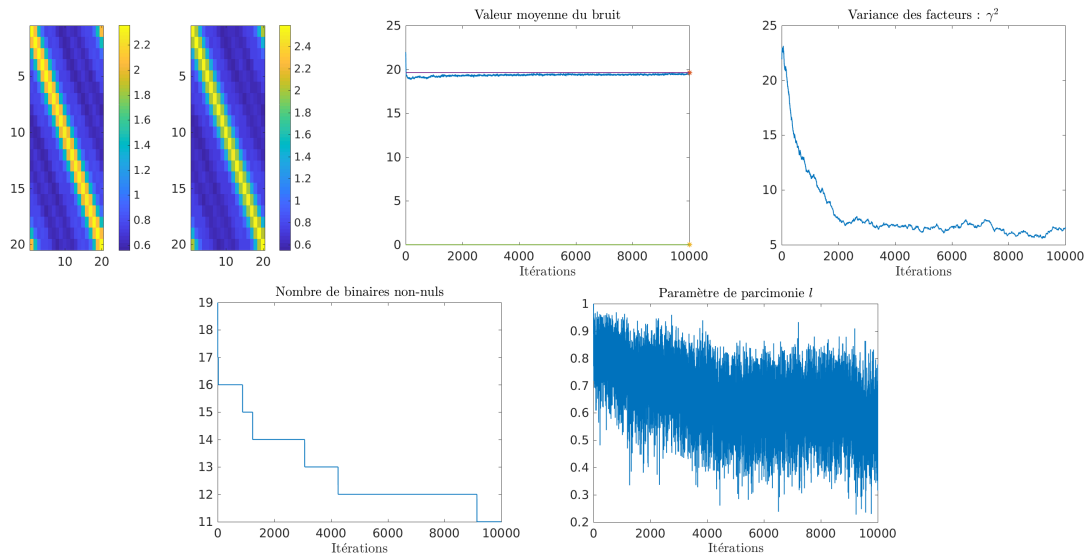


FIGURE 4 – $SNR = -10dB$

4.4. Comparaison avec l'ancienne version

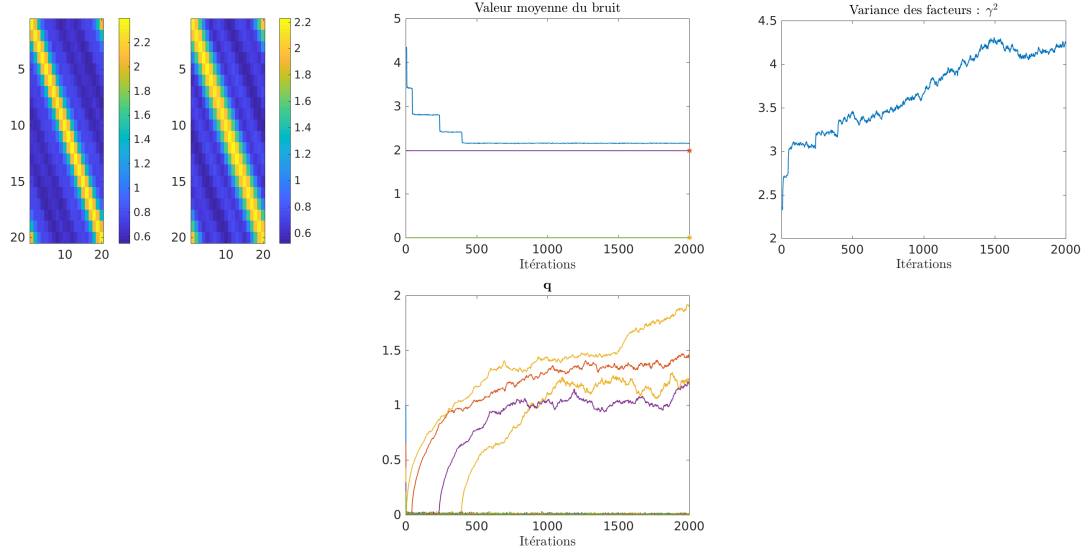


FIGURE 5 – $SNR = 0dB$

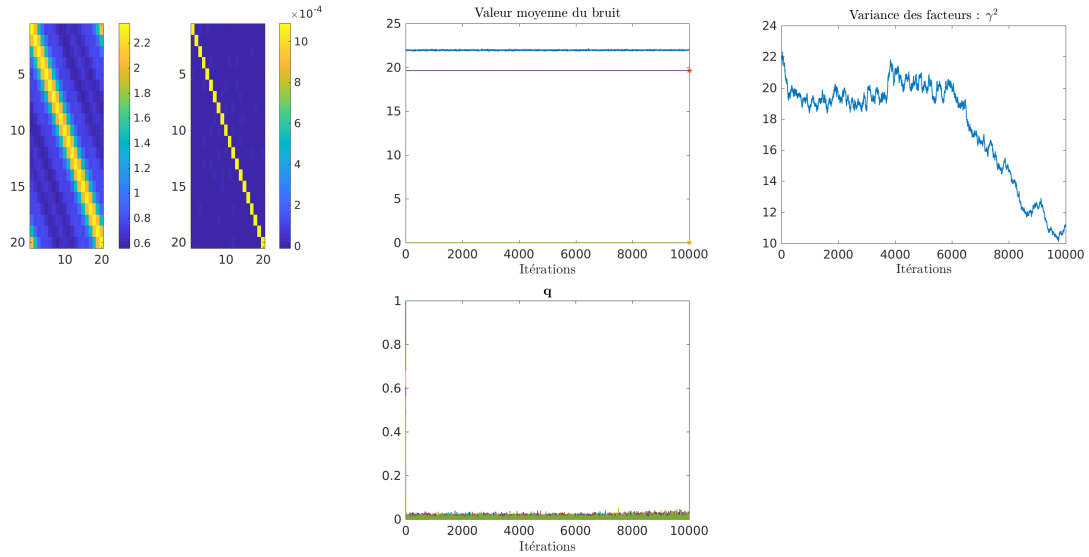


FIGURE 6 – $SNR = -10dB$

Remarques :

- Les résultats issus du modèle 2 changent beaucoup en fonction de l'initialisation, mais comme la convergence est rapide (surtout quand le SNR n'est pas trop bas), il serait possible de lancer simultanément plusieurs chaîne avec différentes initialisations, plus courtes.
- Il serait intéressant de comparer les résultats dans le cas où le nombre de fonction de bases est grand (hypothèse de parcimonie non respectée)

5.1. Pour le couple \mathbf{c}, \mathbf{L}

? propose un moyen de gérer l'ambiguïté d'échelle entre les paramètres \mathbf{c} et \mathbf{L} . D'après cet article, lever cette ambiguïté permettrait une exploration de la loi cible plus efficace, avec une meilleure mélangeance, en réduisant le temps de chauffe et la corrélation de ces paramètres. De manière générale, cette ambiguïté se retrouve pour tout problème de la forme :

$$z = h \star x + b$$

où \star est un opérateur bilinéaire quelconque, un produit matriciel dans notre cas. En effet, on a alors, pour tout couple (x, h) et $s \neq 0$:

$$(sx) \star \left(\frac{h}{s}\right) = x \star h$$

? propose donc d'échantillonner le paramètre de changement d'échelle sans modifier la loi cible. Dans notre modèle, il faut un facteur d'échelle s_k par facteur :

$$\mathbf{y} = \mathbf{L}\mathbf{q}\mathbf{c} + \mathbf{n} = (\tilde{\mathbf{L}}\mathbf{S}^{-1})\mathbf{q}(\mathbf{S}\tilde{\mathbf{c}}) + \mathbf{n} \quad \text{avec} \quad \mathbf{S} = \text{diag}(s_1, \dots, s_k, \dots, s_K).$$

On peut montrer par un changement de variable

$$\mathbf{L}_k = \frac{\tilde{\mathbf{L}}_k}{s_k} \quad \text{et} \quad \mathbf{c}_k = \tilde{\mathbf{c}}_k s_k$$

que l'échantillonnage de s_k se fait selon la loi

$$[s_k | \infty_{-s_k}] \propto |s_k|^{2(I_s - M - 1)} [\mathbf{c}_k, \mathbf{L}_k | s_k] \quad (13)$$

Sous l'hypothèse d'indépendance de \mathbf{c}_k et \mathbf{L}_k ,

$$[s_k | \infty_{-s_k}] \propto |s_k|^{2(I_s - M - 1)} \prod_i [\mathbf{c}_{ki}] \prod_m [\mathbf{L}_{mk}]. \quad (14)$$

avec $|s_k|^2 = ss^*$.

En posant $r_k = |s_k|^2$, on en déduit que

$$\begin{aligned} [r_k | \infty] &\propto r_k^{I_s - M - \frac{1}{2} - 1} e^{-r_k \sum_i \tilde{\mathbf{c}}_{ki}^H \gamma_k^{-2} \tilde{\mathbf{c}}_{ki}} e^{-r_k^{-1} \sum_m \tilde{\mathbf{L}}_{mk}^H K \tilde{\mathbf{L}}_{mk}} \\ &\propto r_k^{I_s - M - \frac{1}{2} - 1} e^{-r_k \gamma_k^{-2} \tilde{\mathbf{S}}_{cckk}} e^{-r_k^{-1} K \sum_m \tilde{\mathbf{L}}_{mk}^H \tilde{\mathbf{L}}_{mk}} \end{aligned}$$

r suit donc une loi Gaussienne Inverse Généralisée (GIG) de la forme :

$$f_{GIG}(r) = \left(\frac{a_r}{b_r}\right)^{\frac{p_r}{2}} \frac{r^{p_r-1}}{2 Y_{p_r}(\sqrt{a_r b_r})} e^{-\frac{1}{2}(a_r r + b_r r^{-1})} \quad , r > 0$$

$Y_{p_r}(\cdot)$ est la fonction de Bessel de seconde espèce.

Par identification, les paramètres sont :

$$\begin{aligned} a_{r_k} &= 2\gamma_k^{-2} \tilde{\mathbf{S}}_{cckk} \\ b_{r_k} &= 2K \sum_m \tilde{\mathbf{L}}_{mk}^H \tilde{\mathbf{L}}_{mk} \\ p_r &= I_s - M - \frac{1}{2} \end{aligned}$$

Après tirage de r_k , on calcule $s_k = \sqrt{r_k} e^{j\phi}$ avec $\phi \sim \mathcal{U}([0; 2\pi[)$.

Dans le cas où $M \approx K \ll I_s$, les paramètres sont de l'ordre de grandeur suivant :

$$\begin{aligned} a_{r_k} &\approx 2I_s \\ b_{r_k} &\approx 2M \\ p_r &\approx I_s - M \end{aligned}$$

le mode de la GIG est alors :

$$\begin{aligned} \text{mode}_{GIG} &= \frac{(p_r - 1) + \sqrt{(p_r - 1)^2 + a_r b_r}}{a_r} \\ &\approx \frac{I_s + \sqrt{I_s^2}}{2I_s} \approx 1 \end{aligned}$$

et la variance est très faible. Le paramètre d'échelle oscille donc légèrement autour de 1, ce qui ne modifie pas l'échelle du problème.

5.2. Pour le couple β^2, σ

L'ambiguïté se pose également pour le couple β^2, σ dans le cas où on exploite P jeux de données (pour différents régime moteur par ex.). On note β^2 le vecteur contenant les P valeurs réelles de β_p^2 et σ^2 celui contenant les M valeurs réelles σ_m^2 . Le facteur d'échelle s suit alors la loi suivante :

$$\begin{aligned} [s|\infty] &\propto |s|^{P-M-1} \prod_p [s\beta_p^2] \prod_m [s^{-1}\sigma_m^2] \\ &\propto |s|^{P-M-1} e^{-\sum_p \beta_p^{-2} b_\beta s^{-1}} e^{-\sum_m \sigma_m^{-2} b_\sigma s} \\ &\propto \mathcal{GIG}(2 \sum_m \sigma_m^{-2} b_\sigma, 2 \sum_p \beta_p^{-2} b_\beta, P - M) \end{aligned}$$

s doit être réel et son signe est tiré dans une loi uniforme $\mathcal{U}([0; 1])$.

Annexe A. Détail du calcul de la postérieure marginalisée

La marginalisation revient à projeter la loi conjointe $[\mathbf{y}_i, \mathbf{c}_i]$ sur \mathbf{c}_i :

$$\begin{aligned} [\mathbf{y}_i | \infty_{-c}] &= \int [\mathbf{y}_i, \mathbf{c}_i] d\mathbf{c}_i \\ &= \int [\mathbf{y}_i | \mathbf{c}_i][\mathbf{c}_i] d\mathbf{c}_i \end{aligned}$$

Or,

$$\begin{aligned} [\mathbf{y}_i | \mathbf{c}_i][\mathbf{c}_i] &= e^{(\mathbf{y}_i - \mathbf{L}q\mathbf{c}_i)^H \sigma_n^{-2} (\mathbf{y}_i - \mathbf{L}q\mathbf{c}_i)} e^{c^h \gamma^{-2} \mathbf{c}_i} \\ &= e^{-\mathbf{y}_i^H \sigma_n^{-2} \mathbf{y}_i} e^{\underbrace{(\mathbf{L}q\mathbf{c}_i)^H \sigma_n^{-2} \mathbf{y}_i}_{\mathbf{c}_i \Omega_{\mathbf{c}_i}^{-1} \boldsymbol{\mu}_i}} e^{\underbrace{\mathbf{y}_i^H \sigma_n^{-2} \mathbf{L}q\mathbf{c}_i}_{\boldsymbol{\mu}_i^H \Omega_{\mathbf{c}_i}^{-1} \mathbf{c}_i}} e^{\underbrace{-(\mathbf{L}q\mathbf{c}_i)^H \sigma_n^{-2} \mathbf{L}q\mathbf{c}_i}_{-\mathbf{c}_i^H \Omega_{\mathbf{c}_i}^{-1} \mathbf{c}_i}} \\ &= C_1 e^{-(\mathbf{c}_i - \boldsymbol{\mu}_i)^H \Omega_{\mathbf{c}_i} (\mathbf{c}_i - \boldsymbol{\mu}_i)} \end{aligned}$$

avec $\boldsymbol{\mu}_i$ et $\Omega_{\mathbf{c}_i}$ sont donnés dans l'encart (??). Donc,

$$[\mathbf{y}_i | \infty_{-c}] = \int [\mathbf{y}_i | \mathbf{c}_i][\mathbf{c}_i] d\mathbf{c}_i = C_1 \underbrace{\int \mathcal{N}_c(\boldsymbol{\mu}_i, \Omega_{\mathbf{c}_i}) d\mathbf{c}_i}_{=1} = C_1$$

Par identification,

$$\begin{aligned} C_1 e^{-\boldsymbol{\mu}_i^H \Omega_{\mathbf{c}_i} \boldsymbol{\mu}_i} &= e^{-\mathbf{y}_i^H \sigma_n^{-2} \mathbf{y}_i} \\ \Leftrightarrow C_1 &= e^{-\mathbf{y}_i^H \sigma_n^{-2} \mathbf{y}_i} e^{\underbrace{\mathbf{y}_i^H \sigma_n^{-2} \mathbf{L}q\Omega_{\mathbf{c}_i} q^H \mathbf{L}^H \sigma_n^{-2} \mathbf{y}_i}_A} \\ C_1 &= e^{-\mathbf{y}_i^H B^{-1} \mathbf{y}_i} \end{aligned}$$

avec

$$A = \frac{\sigma_n^{-2}}{1 + \underbrace{\mathbf{L}^{H-1} \mathbf{q}^{H-1} \gamma^{-2} \mathbf{q}^{-1} \mathbf{L}^{-1}}_{A_2^{-1}} \sigma_n^2}$$

et

$$\begin{aligned} B^{-1} &= -A + \sigma_n^{-2} = \frac{-\sigma_n^{-2} + \sigma_n^{-2} (1 + A_2^{-1} \sigma_n^2)}{1 + A_2^{-1} \sigma_n^2} \\ &= \frac{1}{A_2 + \sigma_n^2} \end{aligned}$$

Finalement,

$$[\mathbf{y}_i | \infty_{-c}] = C_1 = \mathcal{N}_c(\mathbf{0}, B) \quad \text{où} \quad B = \mathbf{L}q\gamma^2 q^H \mathbf{L}^H + \sigma_n^2.$$

Références

Di Ge, Jérôme Idier, and Eric Le Carpentier. Enhanced sampling schemes for mcmc based blind bernoulli–gaussian deconvolution. *Signal Processing*, 91(4) :759–772, 2011.

Thomas Veit, Jérôme Idier, and Saïd Moussaoui. Rééchantillonnage de l'échelle dans les algorithmes mcmc pour les problèmes inverses bilinéaires. 2008.