

APPROCHE BAYÉSIENNE EN TRAITEMENT D'ANTENNE

Ali Mohammad-Djafari*

Laboratoire des signaux et systèmes (CNRS-ESE-UPS)

École supérieure d'électricité

Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France.

email : djafari@lss.supelec.fr

2 juillet 1997

Résumé

Dans ce rapport nous proposons de considérer les problèmes du traitement d'antenne (localisation de source ou la calibration d'antenne) comme des problèmes inverses. L'objectif principal de ce papier est de présenter ces problèmes dans un cadre unique et de proposer une approche bayésienne pour les résoudre. Nous verrons en particulier que les solutions proposées par toutes les méthodes classiques du traitement d'antenne se trouvent comme des solutions particulières obtenues par cette approche. De plus nous verrons comment cette nouvelle approche peut fournir des outils nécessaires pour résoudre les problèmes encore non résolus dans ce domaine. Et par curiosité nous verrons comment ce problème inverse se trouve dans plusieurs autres applications très différentes, comme par exemple, dans l'estimation des paramètres d'une mélange de lois par la méthode des moments ou dans l'estimation des coordonnées des sommets d'un polygone à partir de ses moments. Finalement, quelques résultats de simulation montrant l'intérêt de l'approche choisie.

1 Position du problème

1.1 Un problème typique

Considérons le problème suivant :

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{b}(t) \quad (1)$$

où

- $\mathbf{x}(t) = [x_0(t), \dots, x_{N-1}(t)]$ est un vecteur de fonctions (réels ou complexes) inconnus (par exemple les amplitudes des sources ou les gains complexes des antennes),
- $\boldsymbol{\theta} = [\theta_0, \dots, \theta_{N-1}]$ est un vecteur de réels inconnus (par exemple les directions d'arrivées des ondes),
- $\mathbf{y}(t) = [y_0(t), \dots, y_{M-1}(t)]$ un vecteur contenant les observations, en général des fonctions complexes,
- $\mathbf{b}(t) = [b_0(t), \dots, b_{M-1}(t)]$ un vecteur représentant le bruit des observations, et
- $\mathbf{A}(\boldsymbol{\theta})$ une matrice de dimensions $(M \times N)$ dépendant des paramètres inconnus $\boldsymbol{\theta}$ avec une structure bien définie qui est :

$$\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_0), \dots, \mathbf{a}(\theta_{N-1})] = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{N-1} \\ z_0^2 & z_1^2 & \dots & z_{N-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{M-1} & z_1^{M-1} & \dots & z_{N-1}^{M-1} \end{pmatrix}, \text{ avec } z_n = f(\theta_n), \quad (2)$$

et où la fonction $f(\theta_n)$ est une fonction en général complexe et inversible de θ_n , par exemple en traitement d'antenne $z_n = f(\theta_n) = \exp[-j(2\pi\delta/\lambda) \sin(\theta_n)]$ où λ est la longueur d'onde et δ est la distance entre deux antennes.

Les problèmes à résoudre sont :

- la détermination du nombre de sources M (sélection du modèle);
- l'estimation de $\boldsymbol{\theta}$ ou de \mathbf{z} (direction d'arrivées); et
- l'estimation de \mathbf{x} (amplitudes des sources);

à partir des observations \mathbf{y} .

Ces problèmes ont souvent été traité soit dans le cadre de l'algèbre linéaire soit dans le cadre de l'estimation au sens des moindres carrés soit encore dans le cadre de l'estimation au sens du maximum de vraisemblance, mais à ce jour les travaux utilisant l'approche bayésienne pour la résolution de ce problème restent peu nombreux [1, 2, 3, 4, 5, 6] [7, 8, 9, 10, 11, 12],

Avant d'aller plus loin, il n'est pas inutile de préciser un certain nombre de notations et d'hypothèses dans le cas particulier du traitement d'antenne pour bien comprendre les méthodes classiques.

2 Approches classiques

2.1 Formation de voies et goniomètre

Tout d'abord, en traitement d'antenne classique (réseau linéaire, ondes planes et invariance temporelle), $z_n = \exp[-j(2\pi\delta/\lambda) \sin\theta_n]$ où λ est la longueur d'onde et δ est la distance entre les deux antennes.

Supposons maintenant qu'on a choisit $\delta = \lambda/2$ et que l'espace des angles d'arrivées est discrétisé de telle sorte que $M = N + 1$ et $\sin\theta_n = \frac{n}{N}$, $n = 0, \dots, N - 1$, alors on aura les propriétés suivantes :

$$z_n = \exp[-j\pi n/N], \quad z_0 = 1, \quad z_n^k = \exp[-j\pi kn/N] = z_k^n, \quad k, n = 0, \dots, N - 1.$$

Dans ces conditions la matrice \mathbf{A} devient une matrice de la TFD et on a

$$y_m = [\mathbf{A}\mathbf{x}]_m = \sum_{n=0}^{N-1} z_n^m x_n = \langle \mathbf{a}(\theta_m), \mathbf{x} \rangle, \quad m = 0, \dots, M - 1, \quad (3)$$

et

$$[\mathbf{A}^t \mathbf{y}]_n = \sum_{m=0}^{M-1} z_n^m y_m = \langle \mathbf{y}, \mathbf{a}(\theta_n) \rangle, \quad n = 0, \dots, N. \quad (4)$$

Le vecteur $\mathbf{A}\mathbf{x}$ correspond alors à la TFD du vecteur \mathbf{x} et le vecteur $\frac{1}{N}\mathbf{A}^t \mathbf{y}$ à la TFD inverse de \mathbf{y} , i.e.; $\mathbf{A}^t \mathbf{A} = \mathbf{A}\mathbf{A}^t = N\mathbf{I}$.

La méthode de **goniomètre** correspond alors à estimer \mathbf{x} tout simplement par

$$\hat{\mathbf{x}} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{y} = \mathbf{A}^t (\mathbf{A}\mathbf{A}^t)^{-1} \mathbf{y} = \frac{1}{N} \mathbf{A}^t \mathbf{y} \quad (5)$$

Dans le cas plus général où $z_n = \exp[-j(2\pi\delta/\lambda) \sin\theta_n]$ et $M \geq N$ on peut toujours estimer \mathbf{x} par

$$\hat{\mathbf{x}} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{y}$$

qui correspond en fait à l'estimation au sens des moindres carrés de \mathbf{x} . Sachant de plus que $\mathbf{A}^t \mathbf{A}$ n'est pas trop loin d'une matrice identité ou au moins d'une matrice diagonale $\mathbf{P} = \mathbf{A}^t \mathbf{A} \simeq \text{diag}\{\alpha_1, \dots, \alpha_n\}$, on peut approximer cette solution par

$$\hat{x}_n = \frac{1}{\alpha_n} (\mathbf{A}^t \mathbf{y})_n = \alpha_n^{-1} \langle \mathbf{y}, \mathbf{a}(\theta_n) \rangle. \quad (6)$$

C'est la méthode de **Capon** [13].

En résumé, on peut dire que dans la méthode du goniomètre on approxime $\mathbf{A}^t \mathbf{A}$ par une matrice identité et dans la méthode de Capon on l'approxime par une matrice diagonale.

2.2 Approche algébrique

Le principe de base des méthodes algébriques se trouve dans le fait que la matrice \mathbf{A} , dans le cas général, est une matrice de Vandermonde et on peut montrer les propriétés suivantes :

Considérons le système d'équation $\mathbf{A}_{2M} \mathbf{x}_N = \mathbf{y}_{2M}$:

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ z_0 & z_1 & \dots & z_{N-1} \\ z_0^2 & z_1^2 & \dots & z_{N-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{2M-1} & z_1^{2M-1} & \dots & z_{N-1}^{2M-1} \end{pmatrix} \begin{pmatrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{N-1} \end{pmatrix} = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{M-1} \end{pmatrix} \quad (7)$$

L'idée de base, due à **Prony** [14], se trouve dans le fait que ce système d'équations admet une solution unique (au sens classique) si $M = N$ et si $z_i \neq z_j$, et lorsque $M \geq N$, on peut toujours définir une solution unique au sens d'inverse généralisée. En effet si on considère un polynome $P(z)$ défini par les coefficients $\mathbf{p} = [p_N, p_{N-1}, \dots, p_1]^t$ telle que

$$P(z) = \prod_{i=0}^N (z - z_i) = \sum_{i=0}^N p_i z^{N-i}, \quad p_0 = 1, \quad (8)$$

et si on définit la matrice \mathbf{K}_{2M} de dimensions $(2M \times 2M)$:

$$\mathbf{K}_{2M} = \begin{pmatrix} p_N & \cdots & p_1 & 1 & \mathbf{0} \\ & \ddots & & \ddots & \ddots \\ \mathbf{0} & & p_N & \cdots & p_1 & 1 \end{pmatrix}, \quad (9)$$

il est facile de constater (utilisant la définition du polynome $P(z)$) que

$$\mathbf{K}_{2M} \mathbf{A}_{2M} \mathbf{x}_N = \mathbf{K}_{2M} \mathbf{y}_{2M} = \mathbf{0}.$$

L'équation $\mathbf{K}_{2M} \mathbf{y}_{2M} = \mathbf{0}$ peut être réécrite sous la forme

$$\begin{pmatrix} y_0 & y_1 & \cdots & y_{N-1} \\ y_1 & y_2 & \cdots & y_N \\ \vdots & \vdots & \ddots & \vdots \\ y_{N-1} & y_N & \cdots & y_{2N-2} \end{pmatrix} \begin{pmatrix} p_N \\ p_{N-1} \\ \vdots \\ p_1 \end{pmatrix} = - \begin{pmatrix} y_N \\ y_{N+1} \\ \vdots \\ y_{2N-1} \end{pmatrix}, \quad (10)$$

ou sous forme matricielle $\mathbf{Y}_N \mathbf{p} = -\mathbf{h}_N$, avec $\mathbf{h}_N = [y_N, y_{N+1}, \dots, y_{2N-1}]^t$.

La matrice \mathbf{Y}_N est une matrice de Hankel de dimensions $(N \times N)$ qui peut être diagonalisée par

$$\mathbf{Y}_N = \mathbf{A}_N \text{diag}[\mathbf{x}_N] \mathbf{A}_N^t \quad (11)$$

où $\mathbf{x}_N = [x_0, x_1, \dots, x_{N-1}]^t$ et

$$\mathbf{A}_N = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_0 & z_1 & \cdots & z_{N-1} \\ z_0^2 & z_1^2 & \cdots & z_{N-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{N-1} & z_1^{N-1} & \cdots & z_{N-1}^{N-1} \end{pmatrix} \longrightarrow \det(\mathbf{A}_N) = \prod_{i>j} (z_i - z_j). \quad (12)$$

On constate que si $z_i \neq z_j$ alors $\det(\mathbf{A}_N) \neq 0$ et par conséquent la matrice \mathbf{Y}_N est inversible et on peut calculer \mathbf{p} d'une manière unique.

L'idée de base est donc d'utiliser les observations pour former le vecteur \mathbf{h}_N et la matrice \mathbf{Y}_N et ensuite en l'inversant de calculer le vecteur \mathbf{p} qui contient les coefficients du polynome $P(z)$, et finalement, calculer les zéros de ce polynome qui fournira le vecteur \mathbf{z} .

Deux grandes difficultés en pratique sont les faits que la matrice \mathbf{Y}_N peut être très mal conditionnée et que le calcul des zéros du polynome $P(z)$ peut aussi être très sensible aux erreurs sur ses paramètres. Mais en pratique on ne fait pas cela et on passe par l'intermédiaire de la matrice de corrélation des données comme on le verra dans le paragraphe suivant.

D'autre part, si on définit les deux matrices

$$\mathbf{Y}_0 = \begin{pmatrix} y_{n-1} & y_{n-2} & \cdots & y_0 \\ y_n & y_{n-1} & \cdots & y_1 \\ \vdots & \vdots & \ddots & \vdots \\ y_{2n-2} & y_{2M-3} & \cdots & y_{n-1} \end{pmatrix}, \quad \mathbf{Y}_1 = \begin{pmatrix} y_n & y_{n-1} & \cdots & y_1 \\ y_{n-1} & y_n & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{2n-1} & y_{2M-2} & \cdots & y_n \end{pmatrix} \quad (13)$$

on peut remarquer que

$$\mathbf{Y}_0 = \mathbf{A}_n \text{diag}[\mathbf{x}_n] \mathbf{B}_n^t, \quad \mathbf{Y}_1 = \mathbf{A}_n \text{diag}[\mathbf{x}_n] \text{diag}[\mathbf{z}_n] \mathbf{B}_n^t \quad (14)$$

où

$$\mathbf{A}_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_0 & z_1 & \cdots & z_{n-1} \\ z_0^2 & z_1^2 & \cdots & z_{n-1}^2 \\ \vdots & \vdots & \ddots & \vdots \\ z_0^{n-1} & z_1^{n-1} & \cdots & z_{n-1}^{n-1} \end{pmatrix}, \quad \mathbf{B}_n = \begin{pmatrix} z_0^{n-1} & z_1^{n-1} & \cdots & z_{n-1}^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ z_0^2 & z_1^2 & \cdots & z_{n-1}^2 \\ z_0 & z_1 & \cdots & z_{n-1} \\ 1 & 1 & \cdots & 1 \end{pmatrix}. \quad (15)$$

Ainsi, le vecteur \mathbf{z} recherché peut être calculé en recherchant les valeurs propres de la matrice $\mathbf{Y}_1^{-1} \mathbf{Y}_0$ ou les zéros de la matrice $\mathbf{Y}_0 - z \mathbf{Y}_1$ (Pencil matrix?).

Lorsque le nombre de points M dans le vecteur \mathbf{y} est supérieur à $2n - 1$, on peut toujours former les deux matrices suivantes

$$\mathbf{Y}_0 = \begin{pmatrix} y_{n-1} & y_{n-2} & \cdots & y_0 \\ y_n & y_{n-1} & \cdots & y_1 \\ \vdots & \vdots & \ddots & \vdots \\ y_{2M-2} & y_{2M-3} & \cdots & y_{2M-n-1} \end{pmatrix}, \quad \mathbf{Y}_1 = \begin{pmatrix} y_n & y_{n-1} & \cdots & y_1 \\ y_{n-1} & y_n & \cdots & y_2 \\ \vdots & \vdots & \ddots & \vdots \\ y_{2M-1} & y_{2M-2} & \cdots & y_{2M-n} \end{pmatrix} \quad (16)$$

et en remarquant que seules les nombres de lignes de ces matrices ont été augmenté on peut montrer que les z_j recherchés correspondent cette fois-ci aux valeurs propres de la matrice $(\mathbf{Y}_1^t \mathbf{Y}_1)^{-1} \mathbf{Y}_1^t \mathbf{Y}_0$. Le problème, dans les deux cas, devient celui de la recherche des valeurs propres d'une matrice.

2.3 Estimation au sens des moindres carrés

L'idée des méthodes basées sur l'estimation au sens des moindres carrés (MC) est de remplacer $\mathbf{p} = -\mathbf{Y}^{-1} \mathbf{h}$ par son estimation au sens des MC :

$$\hat{\mathbf{p}} = - \left(\mathbf{Y}^t \mathbf{Y} \right)^{-1} \mathbf{Y}^t \mathbf{h}, \quad (17)$$

ou encore par son estimation au sens des MC pondérés :

$$\hat{\mathbf{p}} = - \left(\mathbf{Y}^t \mathbf{W} \mathbf{Y} \right)^{-1} \mathbf{Y}^t \mathbf{W} \mathbf{h}, \quad (18)$$

où \mathbf{W} est une matrice de pondération.

D'autres auteurs ont défini une solution au sens d'inverse généralisée de $\mathbf{Y} \mathbf{p} = -\mathbf{h}$ en utilisant la décomposition en valeurs singulières de $\mathbf{Y}^t \mathbf{Y}$. Toute la difficulté est alors dans le choix du seuil de troncature de ces valeurs singulières.

D'autres auteurs ont réalisé la nécessité de la régularisation et ont proposé d'estimer \mathbf{p} par

$$\hat{\mathbf{p}} = - \left(\mathbf{Y}^t \mathbf{Y} + \lambda \mathbf{I} \right)^{-1} \mathbf{Y}^t \mathbf{h}, \quad (19)$$

où λ est la valeur singulière la plus petite de la matrice $[\mathbf{Y}, \mathbf{h}]$.

2.4 Estimation au sens du maximum de vraisemblance

L'idée de base ici est de considérer le problème dans le cadre de l'estimation en faisant un certain nombre d'hypothèses sur la nature du bruit [15, 16, 17, 18, 19, 20, 21, 22]. En effet, si on considère le problème dans son cadre général :

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{x} + \mathbf{b} \quad (20)$$

et si on fait l'hypothèse que le bruit est centré blanc (circulaire) gaussien de variance σ_b^2 et indépendant des inconnus $\boldsymbol{\theta}$ et \mathbf{x} , alors on peut écrire

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2) = (2\pi\sigma_b^2)^{M/2} \exp\left[-\frac{1}{2\sigma_b^2}\|\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}\|^2\right] \quad (21)$$

Ensuite on peut envisager l'estimation au sens du MV qui est :

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}, \hat{\sigma}_b^2) = \arg \min_{(\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2)} \left\{ V(\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2) \right\}, \quad (22)$$

avec

$$V(\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2) = -\ln p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2) = cte + M \ln(\sigma_b^2) + \frac{1}{\sigma_b^2} \|\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}\|^2. \quad (23)$$

Le problème devient alors un problème d'optimisation. Certains auteurs ont proposé une minimisation successive de ce critère (sans trop se soucier si la solution ainsi obtenue correspond au maximum global du critère) qui consiste à minimiser le critère successivement par rapport à \mathbf{x} , σ_b^2 et $\boldsymbol{\theta}$.

En effet, à $\boldsymbol{\theta}$ et σ_b^2 fixé la solution s'obtient très facilement par :

$$\nabla_{\mathbf{x}} V = \mathbf{0} \longrightarrow \hat{\mathbf{x}} = \left(\mathbf{A}^t \mathbf{A}\right)^{-1} \mathbf{A}^t \mathbf{y} = \mathbf{P} \mathbf{A}^t \mathbf{y} \quad (24)$$

En reportant cette solution dans le critère $V(\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2)$ on obtient

$$V_1(\boldsymbol{\theta}, \sigma_b^2) = V(\hat{\mathbf{x}}, \boldsymbol{\theta}, \sigma_b^2) = cte + M \ln(\sigma_b^2) + \frac{1}{\sigma_b^2} \|\mathbf{Q}\mathbf{y}\|^2 = cte + M \ln(\sigma_b^2) + \frac{1}{\sigma_b^2} \text{Tr} \left\{ \mathbf{Q} \mathbf{y} \mathbf{y}^t \right\} \quad (25)$$

avec

$$\mathbf{Q} = \mathbf{I} - \mathbf{A} \left(\mathbf{A}^t \mathbf{A}\right)^{-1} \mathbf{A}^t = \mathbf{I} - \mathbf{A} \mathbf{P} \mathbf{A}^t.$$

Dérivant $V_1(\boldsymbol{\theta}, \sigma_b^2)$ par rapport à σ_b^2 donne

$$\frac{\partial V_1}{\partial \sigma_b^2} = 0 \longrightarrow \hat{\sigma}_b^2 = \frac{1}{M} \text{Tr} \left\{ \mathbf{Q} \mathbf{y} \mathbf{y}^t \right\} = \frac{1}{M} \|\mathbf{Q}\mathbf{y}\|^2. \quad (26)$$

Reportant $\hat{\sigma}_b^2$ dans V_1 on obtient

$$V_2(\boldsymbol{\theta}) = V_1(\boldsymbol{\theta}, \hat{\sigma}_b^2) = M \ln \|\mathbf{Q}\mathbf{y}\|^2 + cte \quad (27)$$

Le minimum de $V_2(\boldsymbol{\theta})$ s'obtient pour une valeur de $\|\mathbf{Q}\mathbf{y}\|^2$ proche de zéro. En effet, le vecteur de $\boldsymbol{\theta}$ optimal s'obtient en cherchant la valeur propre nulle (ou la plus petite possible) de la matrice \mathbf{Q} qui dépend de ce vecteur. En pratique l'optimisation par rapport à $\boldsymbol{\theta}$ est

abordée en utilisant les notions de sous-espace signal et sous-espace bruit qui s'expliquent si on effectue une décomposition en valeurs singulière de la matrice $\mathbf{A}^t \mathbf{A}$:

$$\mathbf{A} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^t, \quad \mathbf{A}^t \mathbf{A} = \mathbf{V} \mathbf{\Lambda}^2 \mathbf{V}^t = \sum_{i=1}^n \lambda_i^2 \mathbf{v}_i \mathbf{v}_i^t, \quad \mathbf{A} \mathbf{A}^t = \mathbf{U} \mathbf{\Lambda}^2 \mathbf{U}^t = \sum_{i=1}^M \lambda_i^2 \mathbf{u}_i \mathbf{u}_i^t, \quad (28)$$

$$\mathbf{P} = \left(\mathbf{A}^t \mathbf{A} \right)^{-1} = \mathbf{V} \mathbf{\Sigma} \mathbf{V}^t = \sum_{i=1}^M \frac{1}{\lambda_i^2} \mathbf{v}_i \mathbf{v}_i^t, \quad \mathbf{Q} = \mathbf{I} - \mathbf{U} \mathbf{\Lambda} \mathbf{\Sigma} \mathbf{\Lambda}^t \mathbf{U}^t = \mathbf{I} - \sum_{i=1}^n \mathbf{u}_i \mathbf{u}_i^t = \sum_{i=N+1}^M \mathbf{u}_i \mathbf{u}_i^t \quad (29)$$

Lorsque $M > N$, $\hat{\boldsymbol{\theta}}$ optimal qui minimise $V_1(\boldsymbol{\theta})$ correspond aux valeurs propres non nulles de $\mathbf{A} \mathbf{P} \mathbf{A}^t$ ou les valeurs propres nulles de \mathbf{Q} .

A ce stade on peut comprendre finalement que pratiquement toutes les méthodes classiques du traitement d'antenne peuvent correspondre à une solution approchée de la méthode du MV. Par exemple la méthode de la **formation de voies** correspond à faire l'hypothèse que $\mathbf{P} = (\mathbf{A}^t \mathbf{A})^{-1} \simeq \mathbf{I}$. La solution $\hat{\mathbf{x}}$ devient alors

$$\hat{\mathbf{x}} = \mathbf{A}^t \mathbf{y} \longrightarrow \hat{x}_n = \langle \mathbf{y}, \mathbf{a}(\theta_n) \rangle = \mathbf{a}^t(\theta_n) \mathbf{y} \quad (30)$$

La méthode de **Capon** correspond à faire l'hypothèse que $\mathbf{P} = (\mathbf{A}^t \mathbf{A})^{-1} = \text{diag} \{ \alpha_1, \dots, \alpha_n \}$. La solution $\hat{\mathbf{x}}$ devient alors

$$\hat{x}_n = \alpha_n \langle \mathbf{y}, \mathbf{a}(\theta_n) \rangle = \alpha_n \mathbf{a}^t(\theta_n) \mathbf{y} \quad (31)$$

Les méthodes de **Pisarenko**, **Tuft-Kumaresman** [23, 24] correspondent à faire l'hypothèse

$$\mathbf{P} = \left(\mathbf{A}^t \mathbf{A} \right)^{-1} = \sum_{i=1}^P \frac{1}{\lambda_i^2} \mathbf{v}_i \mathbf{v}_i^t, \quad \mathbf{Q} = \mathbf{I} - \mathbf{A} \mathbf{P} \mathbf{A}^t = \mathbf{I} - \sum_{i=1}^P \mathbf{u}_i \mathbf{u}_i^t = \sum_{i=P+1}^M \mathbf{u}_i \mathbf{u}_i^t. \quad (32)$$

Les matrices \mathbf{P} et \mathbf{Q} sont appelées respectivement les matrices de projection sur le sous-espace signal et sur le sous-espace bruit.

La solution $\hat{\mathbf{x}}$ devient alors

$$\hat{\mathbf{x}} = \mathbf{P} \mathbf{A}^t \mathbf{y} = \sum_{i=1}^P \frac{\langle \mathbf{y}, \mathbf{u}_i \rangle}{\lambda_i} \mathbf{v}_i \quad (33)$$

et bruit peut être estimé par

$$\|\hat{\mathbf{b}}\|^2 = \|\mathbf{Q} \mathbf{y}\|^2 = \left\| \mathbf{y} - \sum_{i=P+1}^n \langle \mathbf{y}, \mathbf{u}_i \rangle \mathbf{u}_i \right\|^2. \quad (34)$$

$P = M = N - 1$ correspond à la méthode de **Pisarenko** et $P = M > N - 1$ correspond à la méthode de **Tuft-Kumaresman**.

D'autres choix pour la décomposition des matrices \mathbf{P} et \mathbf{Q} aussi ont été proposés [25, 26, 27, 28], [29, 30, 31]. L'idée de base étant de choisir $N - P$ vecteurs \mathbf{u}_i dans le sous-espace bruit et former la matrice \mathbf{Q} telle que $\|\mathbf{Q} \mathbf{y}\|^2$ soit le plus faible que possible.

Bien que cette approche permette de prendre en compte la nature du bruit et définit la solution du problème comme celle minimisant un critère, lors de sa mise en œuvre les auteurs se heurtent aux difficultés qu'ils considèrent d'ordre numériques ! Après ils

cherchent à contourner ces difficultés par exemple en remplaçant le calcul théorique de $\hat{\mathbf{x}} = (\mathbf{A}^t \mathbf{A})^{-1} \mathbf{A}^t \mathbf{y}$ par $\hat{\mathbf{x}} = (\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I})^{-1} \mathbf{A}^t \mathbf{y}$! Certains auteurs d'ailleurs justifie cela en parlant de la méthode du maximum de vraisemblance avec un modèle aléatoire pour le signal \mathbf{x} . Est-ce un début pour appréhender l'approche bayésienne ?

Une question plus grave qui se pose concerne l'existence même d'un minimum global du critère $V(\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2)$ ainsi que l'existence des minima locaux. En effet, si on considère ce critère pour σ_b^2 et \mathbf{x} fixés et on étudie le gradient du critère par rapport à \mathbf{z} :

$$\frac{\partial V}{\partial z_k} = \sum_{i=0}^{M-1} -2i z_k^{i-1} x_k \left(y_i - \sum_{k=0}^{N-1} x_k z_k^i \right) \quad (35)$$

on s'aperçoit que le critère peut avoir un très grand nombre de minima locaux, ce qui peut être problématique lors de son optimisation. Minimiser ces critères exige la mise en œuvre des algorithmes d'optimisation multidimensionnels, la plupart du temps très coûteux. Certains auteurs ont proposé des algorithmes stochastiques comme le recuit simulé [32] ou des algorithmes génétiques [20, 33].

Enfin, quelles que soit la méthode, l'implantation réelle et efficace de ces algorithmes restent délicate, ce qui explique pourquoi ces méthodes restent encore peu utilisées.

Un des objectifs de cette communication est de montrer l'apport de l'approche bayésienne pour la résolution de ce problème, non seulement d'un point de vue théorique mais aussi d'un point de vue pratique.

3 Approche bayésienne

Considérons le problème dans son cadre général :

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\theta})\mathbf{x} + \mathbf{b} \quad (36)$$

L'idée de base dans l'approche bayésienne se trouve dans le fait que non seulement on définit une loi de probabilité pour \mathbf{b} pour traduire notre connaissance *a priori* sur le bruit, mais aussi on définit de la même manière des lois de probabilités pour \mathbf{x} et pour \mathbf{b} et même pour le nombre de source N . Ensuite on utilise la règle de Bayes pour déterminer des lois *a posteriori* pour ces inconnues et finalement, on définit des estimateurs et des estimés pour ces inconnues à partir de ces lois.

3.1 Cadre général

Soit H_k l'hypothèse que le nombre de sources N soit k . Sous cette hypothèse on a

$$\mathbf{y} = \mathbf{A}(\boldsymbol{\theta}_k)\mathbf{x}_k + \mathbf{b} \quad (37)$$

Si on note par $Pr(H_k|\mathbf{y})$ la loi *a posteriori* de H_k , par $p(\boldsymbol{\theta}_k|\mathbf{y}, H_k)$ la loi *a posteriori* de $\boldsymbol{\theta}_k$ sous l'hypothèse H_k et par $p(\mathbf{x}_k|\mathbf{y}, \boldsymbol{\theta}_k, H_k)$ la loi *a posteriori* de \mathbf{x}_k sous l'hypothèse H_k alors on peut établir les relations suivantes :

$$\begin{aligned} p(\mathbf{y}, \mathbf{x}_k, \boldsymbol{\theta}_k | H_k) &= p(\mathbf{y} | \mathbf{x}_k, \boldsymbol{\theta}_k, H_k) p(\mathbf{x}_k | H_k) p(\boldsymbol{\theta}_k | H_k), \\ p(\mathbf{y}, \boldsymbol{\theta}_k | H_k) &= \int p(\mathbf{y}, \mathbf{x}_k, \boldsymbol{\theta}_k | H_k) d\mathbf{x}, \\ p(\mathbf{y} | H_k) &= \int p(\mathbf{y}, \boldsymbol{\theta}_k | H_k) d\boldsymbol{\theta}. \end{aligned}$$

De même on a :

$$\begin{aligned} p(\mathbf{x}_k, \boldsymbol{\theta}_k | \mathbf{y}, H_k) &= p(\mathbf{y}, \mathbf{x}_k, \boldsymbol{\theta}_k | H_k) / p(\mathbf{y} | H_k), \\ p(\boldsymbol{\theta}_k | \mathbf{y}, H_k) &= p(\mathbf{y}, \boldsymbol{\theta}_k | H_k) / p(\mathbf{y} | H_k), \\ Pr(H_k | \mathbf{y}) &= p(\mathbf{y} | H_k) Pr(H_k) / \sum_k p(\mathbf{y} | H_k) Pr(H_k). \end{aligned}$$

Notons que toutes ces grandeurs sont définies à partir de l'expression de $p(\mathbf{y} | \mathbf{x}_k, \boldsymbol{\theta}_k, H_k)$, $p(\mathbf{x}_k | H_k)$, $p(\boldsymbol{\theta}_k | H_k)$ et $Pr(H_k)$. La première s'obtient à partir de la loi du bruit et les trois autres sont les lois *a priori*. Une fois les lois *a posteriori* sont déterminées on peut ensuite envisager la procédure d'estimation suivante :

$$\begin{aligned} \hat{N} &= \arg \min_k \{-\ln Pr(H_k | \mathbf{y})\} = \arg \min_k \{-\ln p(\mathbf{y} | H_k) - \ln Pr(H_k)\}, \\ \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}_k} \{-\ln p(\boldsymbol{\theta}_k | \mathbf{y}, H_k)\} = \arg \min_{\boldsymbol{\theta}_k} \{-\ln p(\mathbf{y}, \boldsymbol{\theta}_k | H_k)\}, \\ \hat{\mathbf{x}} &= \arg \min_{\mathbf{x}_k} \{-\ln p(\mathbf{x}_k | \mathbf{y}, \hat{\boldsymbol{\theta}}_k, H_k)\} = \arg \min_{\mathbf{x}_k} \{-\ln p(\mathbf{y}, \mathbf{x}_k, \hat{\boldsymbol{\theta}}_k | H_k)\}. \end{aligned}$$

D'autres choix sont possibles, mais plutôt que détailler ces choix (que l'on peut trouver dans d'autres travaux plus général sur l'approche bayésienne) je préfère illustrer cette approche par l'étude d'un cas particulier qui correspond à choisir des lois *a priori* spécifiques pour le bruit et pour les inconnues.

Dans un premier temps, pour comprendre l'essentiel de cette approche, supposons que le nombre de source soit connue. Nous reviendrons sur cette hypothèse plus loin. Commençons par faire l'hypothèse que notre connaissance sur le bruit se limite à dire qu'il est centré et il n'est pas envisageable de connaître autre chose que sa variance σ_b^2 . Utilisant le principe du maximum d'entropie (ou tout simplement le bon sens) nous conduit à choisir une loi gaussienne pour \mathbf{b} et par conséquent on peut écrire

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, \sigma_b^2) = (2\pi\sigma_b^2)^{-M/2} \exp\left[-\frac{1}{2\sigma_b^2}\|\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}\|^2\right] \quad (38)$$

Supposons aussi que notre connaissance sur le vecteur \mathbf{x} se limite à sa moyenne $\mathbf{x}_0 = x_0\mathbf{1}$ et à sa variance σ_x^2 . Là encore utilisant le principe du maximum d'entropie nous conduit à choisir une loi gaussienne pour \mathbf{x} :

$$p(\mathbf{x}|\sigma_x^2) = (2\pi\sigma_x^2)^{-k/2} \exp\left[-\frac{1}{2\sigma_x^2}\|\mathbf{x} - x_0\mathbf{1}\|^2\right] \quad (39)$$

Faisant les mêmes hypothèses sur $\boldsymbol{\theta}$ nous conduit à :

$$p(\boldsymbol{\theta}|\sigma_\theta^2) = (2\pi\sigma_\theta^2)^{-k/2} \exp\left[-\frac{1}{2\sigma_\theta^2}\|\boldsymbol{\theta} - \theta_0\mathbf{1}\|^2\right] \quad (40)$$

Notons qu'il est fort possible de revenir sur ces hypothèses (nous le verrons plus tard), mais pour l'instant supposons que ces hypothèses sont réalistes. Dans ce cas il nous reste à appliquer la règle de Bayes pour obtenir :

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}, \sigma_b^2, \sigma_x^2, \sigma_\theta^2, x_0, \theta_0) = \frac{(2\pi\sigma_b^2)^{-M/2} (4\pi^2\sigma_x^2\sigma_\theta^2)^{-k/2}}{p(\mathbf{y}|\sigma_b^2, \sigma_x^2, \sigma_\theta^2, x_0, \theta_0)} \exp\left[-\frac{1}{2\sigma_b^2}J_1(\mathbf{x}, \boldsymbol{\theta})\right] \quad (41)$$

avec

$$J_1(\mathbf{x}, \boldsymbol{\theta}) = \|\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}\|^2 + \lambda\|\mathbf{x} - x_0\mathbf{1}\|^2 + \mu\|\boldsymbol{\theta} - \theta_0\mathbf{1}\|^2, \quad \text{où } \lambda = \frac{\sigma_b^2}{\sigma_x^2}, \quad \text{et } \mu = \frac{\sigma_b^2}{\sigma_\theta^2}. \quad (42)$$

On peut envisager ensuite au moins trois possibilités :

- Estimation jointe au sens du maximum *a posteriori* (MAP) de \mathbf{x} et $\boldsymbol{\theta}$:

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \arg \max_{(\mathbf{x}, \boldsymbol{\theta})} \{p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y})\} = \arg \min_{(\mathbf{x}, \boldsymbol{\theta})} \{J_1(\mathbf{x}, \boldsymbol{\theta})\} \quad (43)$$

- Estimation de $\boldsymbol{\theta}$ par

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\mathbf{y})\} = \arg \max_{\boldsymbol{\theta}} \{p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta})\} \quad (44)$$

et ensuite, celle de \mathbf{x} par

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y}, \hat{\boldsymbol{\theta}})\}. \quad (45)$$

Nous appellerons cette méthode MAP marginale (MAPM).

- Estimation de $\boldsymbol{\theta}$ comme dans le cas précédent, mais celle de \mathbf{x} par

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x}} \{p(\mathbf{x}|\mathbf{y})\}, \quad (46)$$

mais, en général le calcul de $p(\mathbf{x}|\mathbf{y})$ n'est pas très facile.

Par la suite nous envisagerons seulement les deux premières possibilités. Par ailleurs, pour simplifier les écritures, nous fixons $x_0 = 0$ et $\theta_0 = 0$.

3.2 Estimation au sens du maximum *a posteriori* (MAP)

L'estimation au sens du maximum *a posteriori* (MAP) de \mathbf{x} et $\boldsymbol{\theta}$ devient :

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \arg \min_{(\mathbf{x}, \boldsymbol{\theta})} \{J_1(\mathbf{x}, \boldsymbol{\theta})\} \quad (47)$$

Notant que lorsque le bruit est négligeable $\sigma_b^2 = 0$ ou lorsque notre connaissance *a priori* sur \mathbf{x} et sur $\boldsymbol{\theta}$ est négligeable (c'est à dire lorsque $\sigma_\theta^2 = \sigma_x^2 = \infty$) on a $\lambda = \mu = 0$ et on retrouve l'estimation au sens du MV.

Là aussi, une fois le critère de l'estimation choisi, le problème de l'optimisation et de la mise en œuvre reste entier. Par exemple pour calculer la solution $(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}})$ on peut envisager plusieurs procédures d'optimisation dont deux sont présentées par la suite.

Optimisation successive :

La solution $(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}})$ doit satisfaire au moins les conditions suivantes : $\nabla_{\mathbf{x}} J_1(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \mathbf{0}$ et $\nabla_{\boldsymbol{\theta}} J_1(\hat{\mathbf{x}}, \hat{\boldsymbol{\theta}}) = \mathbf{0}$. La première nous fournit :

$$\nabla_{\mathbf{x}} J_1(\hat{\mathbf{x}}) = \mathbf{0} \longrightarrow \hat{\mathbf{x}} = \left(\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I} \right)^{-1} \mathbf{A}^t \mathbf{y} = \mathbf{P} \mathbf{A}^t \mathbf{y} \quad (48)$$

Si on reporte cette solution dans l'expression de $J_1(\mathbf{x}, \boldsymbol{\theta})$ on obtient

$$\begin{aligned} J_2(\boldsymbol{\theta}) = J_1(\hat{\mathbf{x}}, \boldsymbol{\theta}) &= \|\mathbf{y} - \mathbf{A} \mathbf{P} \mathbf{A}^t \mathbf{y}\|^2 + \lambda \|\mathbf{P} \mathbf{A}^t \mathbf{y}\|^2 + \mu \|\boldsymbol{\theta}\|^2 \\ &= \mathbf{y}^t \left[\mathbf{I} + (\lambda - 2) \mathbf{A} \mathbf{P} \mathbf{A}^t + \mathbf{A} \mathbf{P} \mathbf{A}^t \mathbf{A} \mathbf{P} \mathbf{A}^t \right] \mathbf{y} + \mu \|\boldsymbol{\theta}\|^2 \end{aligned} \quad (49)$$

On peut alors envisager de calculer $\hat{\boldsymbol{\theta}}$ et $\hat{\mathbf{x}}$ par

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \{J_1(\boldsymbol{\theta})\}, \quad \hat{\mathbf{x}} = \left(\mathbf{A}^t \mathbf{A} + \lambda \mathbf{I} \right)^{-1} \mathbf{A}^t \mathbf{y}. \quad (50)$$

Malheureusement $J_1(\boldsymbol{\theta})$ n'est, en général, pas une onction quadratique de $\boldsymbol{\theta}$ et sa minimisation nécessite un algorithme d'optimisation global. Pour $\mu = 0$, $\hat{\boldsymbol{\theta}}$ optimal peut être calculer en cherchant les valeurs propres nulles de la matrice $[\mathbf{I} + (\lambda - 2) \mathbf{A} \mathbf{P} \mathbf{A}^t + \mathbf{A} \mathbf{P} \mathbf{A}^t \mathbf{A} \mathbf{P} \mathbf{A}^t]$.

Optimisation itérative :

On peut aussi envisager la procédure itérative suivante :

- Initialiser l'algorithme par $\boldsymbol{\theta}^{(0)}$,
- Garder fixe $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$ et optimiser par rapport à \mathbf{x} :

$$\hat{\mathbf{x}}^{(k)} = \arg \min_{\mathbf{x}} \left\{ J_1(\mathbf{x}, \hat{\boldsymbol{\theta}}^{(k)}) \right\}$$

- Garder fixe $\mathbf{x} = \hat{\mathbf{x}}^{(k)}$ et optimiser par rapport à $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}}^{k+1} = \arg \min_{\boldsymbol{\theta}} \left\{ J_1(\hat{\mathbf{x}}^{(k)}, \boldsymbol{\theta}) \right\}$$

Notons qu'ici aussi, il faut être conscient que $J_1(\hat{\mathbf{x}}^{(k)}, \boldsymbol{\theta})$ n'est pas unimodal. Cependant, il est possible d'effectuer une sorte de relaxation en choisissant une valeur de μ très grande et la réduire progressivement pour essayer d'éviter de rester coincer dans un minimum local.

3.3 Estimation au sens du MAPM

L'idée de base est d'abord estimer $\boldsymbol{\theta}$ en intégrant hors du problème \boldsymbol{x} :

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int p(\boldsymbol{x}, \boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{x} \propto p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \quad (51)$$

Dans le cas gaussien, l'intégrale a une solution analytique et il n'est pas difficile à établir les relations suivantes :

$$\mathbf{y}|\boldsymbol{\theta} \simeq \mathcal{N}(\mathbf{A}\mathbf{x}_0, \mathbf{R}_y), \quad \mathbf{R}_y = \mathbf{A}\mathbf{R}_x\mathbf{A}^t + \mathbf{R}_b = \sigma_x^2(\mathbf{A}\mathbf{A}^t + \lambda\mathbf{I}) \quad (52)$$

ou encore d'une manière plus explicite :

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^M |\mathbf{R}_y|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{y} - \mathbf{A}\mathbf{x}_0)^t \mathbf{R}_y^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_0) \right]. \quad (53)$$

En utilisant la règle de Bayes :

$$p(\boldsymbol{\theta}|\mathbf{y}) = p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})/p(\mathbf{y}), \quad (54)$$

on définit la solution :

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \{p(\boldsymbol{\theta}|\mathbf{y})\} = \arg \max_{\boldsymbol{\theta}} \{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\} = \arg \min_{\boldsymbol{\theta}} \{J_2(\boldsymbol{\theta})\}. \quad (55)$$

avec

$$\begin{aligned} J_2(\boldsymbol{\theta}) &= \left[\ln |\mathbf{R}_y| + (\mathbf{y} - \mathbf{A}\mathbf{x}_0)^t \mathbf{R}_y^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_0) + \frac{1}{2}\mu \|\boldsymbol{\theta}\|^2 \right] \\ &= M \ln \sigma_x^2 + \ln |\mathbf{A}\mathbf{A}^t + \lambda\mathbf{I}| \\ &\quad + \frac{1}{\sigma_x^2}(\mathbf{y} - \mathbf{A}\mathbf{x}_0)^t (\mathbf{A}\mathbf{A}^t + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_0) + \mu \|\boldsymbol{\theta}\|^2 \end{aligned} \quad (56)$$

Une fois $\hat{\boldsymbol{\theta}}$ calculé on peut estimer $\hat{\boldsymbol{x}}$ par

$$\hat{\boldsymbol{x}} = \arg \max_{\boldsymbol{x}} \{p(\boldsymbol{x}, \hat{\boldsymbol{\theta}}|\mathbf{y})\} = \arg \min_{\boldsymbol{x}} \{J_1(\boldsymbol{x}, \hat{\boldsymbol{\theta}})\}. \quad (57)$$

La principale difficulté dans cette approche est d'ordre calculatoire qui nécessite le calcul du déterminant $|\mathbf{R}_y|$ et l'inversion de \mathbf{R}_y .

Notons là encore que $J_2(\boldsymbol{\theta})$ n'est, en général, pas unimodal, mais que là encore, il est possible d'effectuer une sorte de relaxation en choisissant une valeur de μ très grande et la réduire progressivement pour essayer d'éviter de rester coincer dans un minimum local.

On peut aussi se permettre le luxe d'estimer σ_x^2 par

$$\frac{\partial J_2}{\partial \sigma_x^2} = 0 \longrightarrow \hat{\sigma}_x^2 = \frac{1}{M}(\mathbf{y} - \mathbf{A}\mathbf{x}_0)^t (\mathbf{A}\mathbf{A}^t + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}_0). \quad (58)$$

Nous reviendrons sur ce point lors de la discussion sur l'estimation des hyperparamètres.

Notons que si $\mu = 0$ cette méthode correspond à ce que l'on appelle la méthode du maximum de vraisemblance avec un modèle d'amplitude aléatoire. Je préfère l'appellation maximum de vraisemblance marginale (MVM).

4 Apport de l'approche bayésienne

Arrivé à ce stade on peut se poser la question et alors ! Voici quelques arguments pour répondre à cette exclamation :

- Cela nous simplifie la vie. Toutes les méthodes précédentes se trouvent être un cas particulier correspondant à une approximation de la solution proposée par cette approche.
- Cette approche nous ouvre la voie vers de nouvelles méthodes qui peuvent à la fois prendre en compte notre connaissance *a priori* sur le bruit et sur les paramètres inconnus sous une forme assez souple (probabiliste), mais il est aussi possible d'introduire des contraintes dures comme par exemple la connaissance *a priori* du support (par exemple $x_j > 0$ ou encore $a < x_j < b$).
- Cette approche nous permet naturellement de fournir des renseignements sur la qualité de la solution, car tout est basé sur la loi *a posteriori* ce qui, par définition, nous permettra de mesurer notre incertitude après avoir épuisé notre connaissance *a priori* et les données qui étaient à notre disposition.
- Un inconvénient cependant : il y a des hyperparamètres ($\sigma_b^2, \sigma_\theta^2, \sigma_x^2$) ou plutôt (λ, μ) qu'il faut soit connaître *a priori* soit les estimer. Mais pour cela il existe des outils (voir la discussion sur la détermination des hyperparamètres).

On peut aussi se poser des questions sur le choix des différentes lois $p(\mathbf{x})$ et $p(\boldsymbol{\theta})$. Ci-dessous quelques idées sous forme de question-réponse.

Question : Que faire si l'hypothèse bruit blanc n'est pas réaliste?

Réponse : La réponse en théorie est très simple. Si on connaît la matrice de covariance du bruit \mathbf{R}_b il suffit de remplacer l'équation (38) par

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, \mathbf{R}_b) \propto \exp \left[-\frac{1}{2}[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}]^t \mathbf{R}_b^{-1}[\mathbf{y} - \mathbf{A}(\boldsymbol{\theta})\mathbf{x}] \right] \quad (59)$$

et continuer la suite. Les calculs deviennent à peine plus complexes.

Question : Et si l'hypothèse gaussien pour \mathbf{x} n'est pas convenable?

Réponse : Effectivement si on sait que x_j correspond à l'intensité d'une source de rayonnement et par conséquent on sait que $x_j > 0$ on peut envisager de choisir une loi Gamma multivariable pour \mathbf{x} :

$$p(\mathbf{x}|a, b) = \prod_{j=1}^n \frac{b^{-a}}{\Gamma(a)} x_j^{a-1} \exp \left[-\frac{x_j}{b} \right] \quad (60)$$

ce qui donne

$$\ln p(\mathbf{x}|a, b) \propto \sum_{j=1}^n (a-1) \ln(x_j) - \frac{x_j}{b} \quad (61)$$

D'autres choix sont aussi possibles. Ici par contre les calculs deviennent plus complexes car on n'a plus de solution analytique pour l'optimisation par rapport à \mathbf{x} .

Question : Et l'hypothèse gaussien pour $\boldsymbol{\theta}$?

Réponse : Effectivement si θ_j représentent les directions d'arrivées ($\theta_j \in [0, \pi]$), cela n'est peut être pas convenable? A la limite une loi uniforme sur l'intervat $[0, \pi]$ serait plus convenable. Mais on peut faire mieux sachant que c'est souhaitable d'obtenir $\theta_i \neq \theta_j$. Pour cela je propose de choisir la loi $p(\boldsymbol{\theta})$ comme suit :

$$\begin{aligned} p(\theta_1) &= \frac{1}{\pi} && \text{si } \theta_1 \in [0, \pi], && 0 \text{ ailleurs} \\ p(\theta_2|\theta_1) &= \frac{1}{(\pi-\theta_1)} && \text{si } \theta_2 \in [\theta_1, \pi], && 0 \text{ ailleurs} \\ p(\theta_3|\theta_2, \theta_1) &= p(\theta_3|\theta_2) = \frac{1}{(\pi-\theta_2)} && \text{si } \theta_3 \in [\theta_2, \pi], && 0 \text{ ailleurs} \\ &\vdots && && \\ p(\theta_n|\theta_{N-1}, \dots, \theta_1) &= p(\theta_n|\theta_{N-1}) = \frac{1}{(\pi-\theta_{N-1})} && \text{si } \theta_n \in [\theta_{N-1}, \pi], && 0 \text{ ailleurs} \end{aligned}$$

ce qui donne

$$p(\boldsymbol{\theta}) = \begin{cases} p(\theta_1) \prod_{j=2}^n p(\theta_j|\theta_{j-1}) = \frac{1}{\pi} \prod_{j=2}^n \frac{1}{(\pi - \theta_{j-1})} & \text{si } 0 < \theta_1 < \theta_2 < \dots < \theta_n < \pi, \\ 0 & \text{ailleurs} \end{cases}$$

ou encore

$$-\ln p(\boldsymbol{\theta}) = \begin{cases} \sum_{j=1}^{N-1} \ln(\pi - \theta_{j-1}) & \text{si } \theta_0 = 0 < \theta_1 < \theta_2 < \dots < \theta_n < \pi \\ \infty & \text{ailleurs} \end{cases} \quad (62)$$

Un autre choix peut être

$$\begin{aligned} p(\theta_1) &= \frac{\ln \pi}{\theta_1} && \text{si } \theta_1 \in [0, \pi], && 0 \text{ ailleurs} \\ p(\theta_2|\theta_1) &= \frac{\ln(\pi-\theta_1)}{(\theta_2-\theta_1)} && \text{si } \theta_2 \in [\theta_1, \pi], && 0 \text{ ailleurs} \\ p(\theta_3|\theta_2, \theta_1) &= p(\theta_3|\theta_2) = \frac{\ln(\pi-\theta_2)}{(\theta_3-\theta_2)} && \text{si } \theta_3 \in [\theta_2, \pi], && 0 \text{ ailleurs} \\ &\vdots && && \\ p(\theta_n|\theta_{N-1}, \dots, \theta_1) &= p(\theta_n|\theta_{N-1}) = \frac{\ln(\pi-\theta_{N-1})}{(\theta_n-\theta_{N-1})} && \text{si } \theta_n \in [\theta_{N-1}, \pi], && 0 \text{ ailleurs} \end{aligned}$$

ce qui donne

$$p(\boldsymbol{\theta}) = \begin{cases} p(\theta_1) \prod_{j=2}^n p(\theta_{j-1}|\theta_{j-1}) = \frac{\ln \pi}{\theta_1} \prod_{j=2}^n \frac{\ln(\pi - \theta_j)}{(\theta_j - \theta_{j-1})} & \text{si } 0 < \theta_1 < \theta_1 < \dots < \theta_n < \pi, \\ 0 & \text{ailleurs} \end{cases}$$

ou encore

$$-\ln p(\boldsymbol{\theta}) = \begin{cases} -\sum_{j=1}^n \ln(\ln(\pi - \theta_j)) + \ln(\theta_j - \theta_{j-1}) & \text{si } \theta_0 = 0 < \theta_1 < \theta_1 < \dots < \theta_n < \pi, \\ \infty & \text{ailleurs} \end{cases} \quad (63)$$

Un troisième choix peut être

$$\begin{aligned} p(\theta_1) &\propto \theta_1 \exp[-\theta_1] && \text{si } \theta_1 \in [0, \pi], && 0 \text{ ailleurs} \\ p(\theta_2|\theta_1) &\propto (\theta_2 - \theta_1) \exp[-(\theta_2 - \theta_1)] && \text{si } \theta_2 \in [\theta_1, \pi], && 0 \text{ ailleurs} \\ p(\theta_3|\theta_2, \theta_1) &= p(\theta_3|\theta_2) \propto (\theta_3 - \theta_2) \exp[-(\theta_3 - \theta_2)] && \text{si } \theta_3 \in [\theta_2, \pi], && 0 \text{ ailleurs} \\ &\vdots && && \\ p(\theta_n|\theta_{N-1}, \dots, \theta_1) &\propto (\theta_n - \theta_{N-1}) \exp[-(\theta_n - \theta_{N-1})] && \text{si } \theta_n \in [\theta_{N-1}, \pi], && 0 \text{ ailleurs} \end{aligned}$$

ce qui donne

$$-\ln p(\boldsymbol{\theta}) = \begin{cases} cte + \sum_{j=1}^n (\theta_j - \theta_{j-1}) \ln(\theta_j - \theta_{j-1}) & \text{si } \theta_0 = 0 < \theta_1 < \theta_2 < \dots < \theta_n < \pi, \\ \infty & \text{ailleurs} \end{cases} \quad (64)$$

Question : Et les hyperparamètres $\boldsymbol{\alpha} = (\sigma_x^2, \sigma_\theta^2, \sigma_b^2)$?

Réponse : Personne ne nous empêche d'attribuer une loi de probabilité à $\boldsymbol{\alpha}$ et les estimer ensuite. Par exemple on peut envisager

$$(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\alpha}}) = \arg \max_{(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\alpha})} \{p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{y})\}.$$

Principales difficultés restent le choix judicieux de $p(\boldsymbol{\alpha})$ et le fait que la loi *a posteriori* $p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{y})$ peut ne pas être unimodale, les deux étant liées. En effet, il faut choisir $p(\boldsymbol{\alpha})$ de telle sorte que $p(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{\alpha} | \boldsymbol{y})$ soit unimodale ou qu'au moins $p(\hat{\boldsymbol{x}}, \hat{\boldsymbol{\theta}}, \boldsymbol{\alpha} | \boldsymbol{y})$ soit unimodale par rapport à $\boldsymbol{\alpha}$. Là encore d'autres choix pour l'estimation sont possibles (marginalisation ou pas).

Question : Et le choix de N le nombre de sources?

Réponse : Personne ne nous empêche d'attribuer une loi de probabilité à N et à l'estimer au même titre que les autres inconnus. Par exemple un choix judicieux est une distribution binomiale

$$P(N = k) = \binom{M}{k} p^k (1-p)^{M-k}, \quad (65)$$

avec M le nombre des points mesures \boldsymbol{y} et $p \in [0, 1]$. Un choix de $p < 1/2$ favorisera des petites valeurs pour N et un choix $p > 1/2$ favorisera des grandes valeurs pour N , et si on n'a pas de favori on peut choisir $p = 1/2$. En tous cas un tel choix assure que N soit entre 0 et M . D'autres choix sont aussi possibles : lois géométrique, binomiale-négative, Poisson, etc.

5 Présentation de nouveaux algorithmes

Pour illustrer les propos tenus précédemment, nous proposons deux nouvelles méthodes basées sur cette approche dans un cas spécifique où $\boldsymbol{\theta}$ représente des fréquences des sinusoïdes et \mathbf{x} leurs amplitudes.

Dans la première méthode, il s'agit d'estimer N , $\boldsymbol{\theta}$ et \mathbf{x} par le critère MAP:

$$(\hat{N}, \hat{\boldsymbol{\theta}}, \hat{\mathbf{x}}) = \arg \max_{(k, \boldsymbol{\theta}, \mathbf{x})} \{p(\mathbf{x}_k, \boldsymbol{\theta}_k, H_k | \mathbf{y})\} \quad (66)$$

où

$$p(\mathbf{x}_k, \boldsymbol{\theta}_k, H_k | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{x}_k, \boldsymbol{\theta}_k, H_k) p(\mathbf{x}_k | H_k) p(\boldsymbol{\theta}_k | H_k) Pr(H_k), \quad (67)$$

Dans la deuxième méthode, il s'agit d'estimer N , $\boldsymbol{\theta}$ et \mathbf{x} par le critère MAPM:

$$\begin{aligned} \hat{N} &= \arg \min_k \{-\ln Pr(H_k | \mathbf{y})\} = \arg \min_k \{-\ln p(\mathbf{y} | H_k) - \ln Pr(H_k)\}, \\ \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}_k} \{-\ln p(\boldsymbol{\theta}_k | \mathbf{y}, H_k)\} = \arg \min_{\boldsymbol{\theta}_k} \{-\ln p(\mathbf{y}, \boldsymbol{\theta}_k | H_k)\}, \\ \hat{\mathbf{x}} &= \arg \min_{\mathbf{x}_k} \{-\ln p(\mathbf{x}_k | \mathbf{y}, \hat{\boldsymbol{\theta}}_k, H_k)\} = \arg \min_{\mathbf{x}_k} \{-\ln p(\mathbf{y}, \mathbf{x}_k, \hat{\boldsymbol{\theta}}_k | H_k)\}. \end{aligned}$$

Dans les deux cas nous avons fait l'hypothèse que le bruit est blanc, centré et gaussien. Ceci nous ramène à :

$$p(\mathbf{y} | \mathbf{x}_k, \boldsymbol{\theta}_k, H_k) = (2\pi\sigma_b^2)^{-M/2} \exp \left[-\frac{1}{2\sigma_b^2} \|\mathbf{y} - \mathbf{A}_k \mathbf{x}_k\|^2 \right]. \quad (68)$$

Pour la loi *a priori* de \mathbf{x} nous avons choisi trois cas:

1. Une loi gaussienne :

$$p(\mathbf{x}_k | H_k) = (2\pi\sigma_x^2)^{-k/2} \exp \left[-\frac{1}{2\sigma_x^2} \|\mathbf{x}_k\|^2 \right] = (2\pi\sigma_x^2)^{-k/2} \exp \left[-\frac{1}{2\sigma_x^2} \sum_{j=1}^k |x_j|^2 \right]$$

2. Une loi de Cauchy :

$$p(\mathbf{x}_k | H_k) \propto \prod_{j=1}^k \frac{1/\alpha}{1 + |x_j/\alpha|^2} = \alpha^{-k} \exp \left[-\sum_{j=1}^k \ln(1 + |x_j/\alpha|^2) \right]$$

3. Une loi entropique :

$$p(\mathbf{x}_k | H_k) = \left(\frac{\alpha^{\alpha+1}}{\Gamma(\alpha+1)} \right)^k \exp \left[\alpha \sum_{j=1}^k \ln x_j - x_j \right]$$

Pour la loi *a priori* de $\boldsymbol{\theta}$ nous avons aussi choisi trois cas:

1. Une loi gaussienne :

$$p(\boldsymbol{\theta}_k | H_k) = (2\pi\sigma_\theta^2)^{-k/2} \exp \left[-\frac{1}{2\sigma_\theta^2} \|\boldsymbol{\theta}_k\|^2 \right]$$

2. Une loi uniforme $\theta_j \in [\theta_{max} - \theta_{min}]$:

$$p(\boldsymbol{\theta}_k) = cte = \frac{1}{(\theta_{max} - \theta_{min})^k}$$

3. Une loi entropique :

$$p(\boldsymbol{\theta}_k | H_k) = \left(\frac{\beta^{\beta+1}}{\Gamma(\beta+1)} \right)^k \exp \left[\beta \sum_{j=1}^k \ln(\theta_j) - \theta_j \right]$$

Pour la loi *a priori* de H_k nous avons choisi deux cas:

1. Une loi uniforme $k \in [0, M]$:

$$P(H_k) = P(N = k) = cte = \frac{1}{M}$$

2. Une loi multinomiale $k \in [0, M]$ de paramètre p :

$$Pr(H_k) = Pr(N = k) = \binom{M}{k} p^k (1-p)^{M-k}.$$

Remplaçant ces expressions dans l'expression de la loi *a posteriori* on a:

– dans la première méthode :

$$(\hat{N}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{x}}) = \arg \min_{(k, \boldsymbol{\theta}_k, \boldsymbol{x}_k)} \{J_1(\boldsymbol{x}_k, \boldsymbol{\theta}_k, k)\}, \quad (69)$$

– et dans la deuxième méthode :

$$\begin{aligned} \hat{N} &= \arg \min_k \{J_3(k)\} \\ \hat{\boldsymbol{\theta}} &= \arg \min_{\boldsymbol{\theta}} \{J_2(\boldsymbol{\theta}, \hat{N})\} \\ \hat{\boldsymbol{x}} &= \arg \min_{\boldsymbol{x}} \{J_1(\boldsymbol{x}, \hat{\boldsymbol{\theta}}, \hat{N})\}. \end{aligned}$$

avec

$$J_1(\boldsymbol{x}_k, \boldsymbol{\theta}_k, k) = \frac{1}{2\sigma_b^2} \|\boldsymbol{y} - \mathbf{A}_k \boldsymbol{x}_k\|^2 - \ln p(\boldsymbol{x}_k | H_k) - \ln p(\boldsymbol{\theta}_k | H_k) - \ln P(H_k) \quad (70)$$

Par exemple, dans le cas du choix: loi gaussienne pour $p(\boldsymbol{x}_k | H_k)$, loi uniforme pour $p(\boldsymbol{\theta}_k | H_k)$ et loi uniform pour $P(H_k)$ on a

$$\begin{aligned} J_1(\boldsymbol{x}_k, \boldsymbol{\theta}_k, k) &= \frac{1}{2\sigma_b^2} \|\boldsymbol{y} - \mathbf{A}_k \boldsymbol{x}_k\|^2 \\ &+ \frac{1}{2\sigma_x^2} \|\boldsymbol{x}_k\|^2 + \frac{k}{2} \ln(2\pi\sigma_x^2) \\ &+ k \ln(\theta_{max} - \theta_{min}) \end{aligned}$$

et dans le cas du choix: loi gaussienne pour $p(\mathbf{x}_k|H_k)$, loi entropique pour $p(\boldsymbol{\theta}_k|H_k)$ et loi multinomiale pour $P(H_k)$ on a

$$\begin{aligned} J_1(\mathbf{x}_k, \boldsymbol{\theta}_k, k) &= \frac{1}{2\sigma_b^2} \|\mathbf{y} - \mathbf{A}_k \mathbf{x}_k\|^2 \\ &+ \frac{1}{2\sigma_x^2} \|\mathbf{x}_k\|^2 + \frac{k}{2} \ln(2\pi\sigma_x^2) \\ &- \beta \left[\sum_{j=1}^k (\ln \theta_j - \theta_j) \right] - k(\beta \ln(1 + \beta) - \ln(\Gamma(\beta))) \\ &- [\ln(k!) + \ln((M - k)!) - k \ln p - (M - k) \ln(1 - p)]. \end{aligned}$$

Dans la deuxième méthode, la principale difficulté est dans le calcul $p(\boldsymbol{\theta}|\mathbf{y}, H_k)$ et $P(H_k|\mathbf{y})$ qui nécessite:

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}, H_k) &= \int p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}, H_k) d\mathbf{x} \\ P(H_k|\mathbf{y}) &= \int p(\boldsymbol{\theta}|\mathbf{y}, H_k) d\boldsymbol{\theta} \end{aligned}$$

Notons

$$J_1(\mathbf{x}, \boldsymbol{\theta}, k) = -\ln p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}, H_k) - \ln p(\mathbf{x}) - \ln p(\boldsymbol{\theta}) - \ln P(H_k)$$

et faisons l'hypothèse que J_1 est quadratique autour de son minimum. Notons aussi

$$\begin{aligned} J_2(\boldsymbol{\theta}, k) &= J_1(\mathbf{x}, \boldsymbol{\theta}, k) - \ln |\mathbf{H}_1| \\ J_3(k) &= J_2(\boldsymbol{\theta}, k) - \ln |\mathbf{H}_2| \end{aligned}$$

où \mathbf{H}_1 et \mathbf{H}_2 sont, respectivement, des matrices Hessiennes de J_1 par rapport à \mathbf{x} et de J_2 par rapport à $\boldsymbol{\theta}$:

$$[\mathbf{H}_1]_{ij} = \frac{\partial^2 J_1}{\partial x_i \partial x_j}, \quad [\mathbf{H}_2]_{ij} = \frac{\partial^2 J_2}{\partial \theta_i \partial \theta_j}.$$

Les expressions de J_2 et de J_3 , dans le cas général sont très complexes. Cependant, dans le cas d'une loi gaussienne pour \mathbf{x} , on a une expression analytique pour J_2 . En effet, dans ce cas J_1 est quadratique en \mathbf{x} et on a

$$\begin{aligned} \mathbf{H}_1 &= \frac{1}{2\sigma_b^2} \mathbf{A}_k^t \mathbf{A}_k + \frac{1}{2\sigma_x^2} \mathbf{I} = \frac{1}{2\sigma_b^2} \mathbf{P}, \\ \hat{\mathbf{x}} &= \arg \min_{\mathbf{x}} \{J_1(\mathbf{x}_k, \boldsymbol{\theta}_k, k)\} = \mathbf{P}^{-1} \mathbf{A}^t \mathbf{y}, \end{aligned}$$

$$J_2(\boldsymbol{\theta}_k, k) = \|\mathbf{y} - \mathbf{A} \mathbf{P} \mathbf{A}^t \mathbf{y}\|^2 + \lambda \|\mathbf{P} \mathbf{A}^t \mathbf{y}\|^2 + \ln p(\boldsymbol{\theta}_k) + \ln P(H_k) + \frac{k}{2} \ln |\mathbf{P}|$$

L'expression de J_3 reste encore trop complexe. C'est pourquoi on peut envisager d'estimer N et $\boldsymbol{\theta}$ par

$$(\hat{N}, \hat{\boldsymbol{\theta}}) = \arg \min_{(k, \boldsymbol{\theta})} \{J_2(\boldsymbol{\theta}, k)\} \quad (71)$$

avec

$$J_k(\boldsymbol{\theta}) = \ln |\mathbf{P}| + \frac{1}{\sigma_x^2} \mathbf{y}^t \mathbf{P}^{-1} \mathbf{y} - \ln(p(\boldsymbol{\theta})) \quad (72)$$

où $\mathbf{P} = [\mathbf{A} \mathbf{A}^t + \lambda \mathbf{D}]$ et où $\mathbf{D} = -\nabla^2 (\ln p(\mathbf{x}))$ est égale à l'identité dans le cas gaussien et une matrice diagonale dont les éléments diagonaux dépendent de la solution $\hat{\mathbf{x}}$ calculée.

En tous cas, la minimisation par rapport à k se fait par énumération entre 1 et M et la minimisation par rapport à θ se fait sur une grille discrète $\theta_j = j(\pi/M)$, $j = 0, \dots, M$, et finalement, la minimisation par rapport à \mathbf{x} , dans le cas gaussien, se fait d'une manière analytique. L'algorithme proposé est le suivant :

Algorithme 1 :

0. Initialiser $\theta = [\theta_1, \dots, \theta_m]$ avec $\theta_j = j(\pi/M)$,

1. Pour $k = 1, \dots, M$:

1.1 Pour toutes les combinaisons c possibles de k sources parmi M sources :

1.1.1 construire la matrice \mathbf{A}_k ;

1.1.2 calculer $\hat{\mathbf{x}}_k = \arg \min_{\mathbf{x}} \{J_{kc}(\mathbf{x}_k, \theta_k)\}$ et $J_{kc}(\hat{\mathbf{x}}_k, \theta_k)$ (eq:70) ou $J_{kc}(\theta_k)$ (eq:72);

(Dans le cas gaussien : $\hat{\mathbf{x}}_k = (\mathbf{A}_k^t \mathbf{A}_k + \lambda \mathbf{I}_k)^{-1} \mathbf{A}_k^t \mathbf{y}$)

fin 1.1

1.2 Sélectionner le vecteur $\hat{\theta}_k$ qui correspond au $\min_c(J_{kc})$.

fin 1.

2. Sélectionner \hat{k} qui correspond au $\min_k(\min_c(J_k))$

L'étape 2 de cet algorithme est très coûteux. On peut proposer un algorithme sous-optimal qui consiste à fixer les $\theta_j, j = 1, \dots, k-1$ et ne réactualiser que θ_k . Ceci revient à effectuer une recherche unidimensionnelle pour θ_k parmi les $M-k$ valeurs possibles alors que dans l'algorithme optimal il s'agit d'effectuer une recherche k -dimensionnelle dans un espace de dimension M qui comporte $\frac{M!}{k!(M-k)!}$ combinaisons au lieu de $M-k$ dans le cas sous-optimal.

Algorithme 2 :

0. Initialiser $\theta = [\theta_1, \dots, \theta_m]$ avec $\theta_j = j(\pi/M)$ et $\mathbf{A}_k = []$.

1. Pour $k = 0, \dots, M$:

1.1 Pour $c = k+1, \dots, M$:

1.1.1 réactualiser la dernière colonne de \mathbf{A}_k par $\mathbf{a}(\theta_c)$: $\mathbf{A}_k = [\mathbf{A}_k, \mathbf{a}(\theta_c)]$,

1.1.2 calculer $\hat{\mathbf{x}}_k = \arg \min_{\mathbf{x}} \{J_{kc}(\mathbf{x}_k, \theta_k)\}$ et $J_{kc}(\hat{\mathbf{x}}_k, \theta_k)$ ou $J_{kc}(\theta_k)$ (eq:72);

(Dans le cas gaussien : $\hat{\mathbf{x}}_k = (\mathbf{A}_k^t \mathbf{A}_k + \lambda \mathbf{I}_k)^{-1} \mathbf{A}_k^t \mathbf{y}$)

fin 1.1

1.2 Sélectionner le $\hat{\theta}_c$ qui correspond au $\min_c(J_{kc})$.

(Notons que seule le composant θ_c se réactualise.)

1.3 Réactualiser la dernière colonne de \mathbf{A}_k par $\mathbf{a}(\hat{\theta}_c)$,

fin 1.

2. Sélectionner \hat{k} qui correspond au $\min_k(\min_c(J_{kc}))$

En langage parlé cet algorithme se décrit :

1. On choisit un nombre maximum M pour le nombre des sources.
2. On choisit M fréquences parmi celles dont les modules de leurs TFD sont les plus grandes.
3. On fait l'hypothèse qu'il n'y a qu'une seule sources et on sélectionne, parmi celles choisit en 2, la fréquence la plus vraisemblable ($J(f_1)$ minimum) et on estime son amplitude. On obtient ainsi (\hat{f}_1, \hat{x}_1) et J_1 .

4. Ensuite, parmi les fréquences restantes, on sélectionne le couple des fréquences (\hat{f}_1, f_2) la plus vraisemblable, *i.e.*; $J(\hat{f}_1, f_2)$ minimum, \hat{f}_1 étant celle qui avait été sélectionnée dans l'étape précédente. On estime aussi les amplitudes correspondantes en minimisant $J(x_1, x_2)$. On obtient ainsi (\hat{f}_1, \hat{f}_2) , (\hat{x}_1, \hat{x}_2) et J_2 .
5. On continue cette procédure avec les triplets, quadruplets, etc. parmi les fréquences restantes, *i.e.*; $J(\hat{f}_1, \hat{f}_2, f_3)$ minimum, \dots ; $J(\hat{f}_1, \hat{f}_2, \hat{f}_3, \dots, f_m)$ minimum.
6. Et, finalement, on choisit parmi cet ensemble $\{(\hat{f}_1), (\hat{f}_1, \hat{f}_2), (\hat{f}_1, \hat{f}_2, \hat{f}_3), \dots, (\hat{f}_1, \hat{f}_2, \dots, \hat{f}_m)\}$ l'élément qui est la plus vraisemblable et on estime les amplitudes correspondantes.

Le coût de calcul de cet algorithme est aussi important. C'est pourquoi on peut aussi envisager un troisième algorithme:

Algorithme 3 :

0. Initialiser $\boldsymbol{\theta} = [\theta_1, \dots, \theta_m]$ avec $\theta_j = j(\pi/M)$.
1. Pour $k = M : -1 : 1$:
 - 1.1 Construire la matrice \mathbf{A}_k et calculer la matrice $\mathbf{P} = (\mathbf{A}_k \mathbf{A}_k^t + \lambda \mathbf{I}_k)$,
 - 1.2 Calculer $\hat{\mathbf{x}}_k = \mathbf{A}_k^t \mathbf{P}^{-1} \mathbf{y}$ et le critère $J_k(\hat{\mathbf{x}}_k, \boldsymbol{\theta}_k)$ (eq:70) ou $J_k(\boldsymbol{\theta}_k)$ (eq:72).
 - 1.3 Identifier l'élément θ_c qui correspond à $\hat{x}_c = \min_c(\hat{x}_1, \dots, \hat{x}_k)$
(Rechercher la fréquence dont l'amplitude est la plus petite et éliminer la colonne correspondante dans la matrice \mathbf{A} .)
- fin 1.
2. Sélectionner le vecteur $\hat{\boldsymbol{\theta}}_k$ qui correspond au $\min_k(J_k)$.
3. Reconstruire la matrice \mathbf{A}_k et calculer $\hat{\mathbf{x}}_k$ correspondant.

En langage parlé ceci signifie :

1. On choisit un nombre maximum M pour le nombre des sources.
2. On choisit M fréquences parmi celles dont les modules de leurs TFD sont les plus grandes.
3. On fait l'hypothèse qu'il y a $k = M$ fréquences, on estime leurs amplitudes en minimisant le critère par rapports à \mathbf{x}_k et on calcule le critère $J_k(\hat{\mathbf{x}}_k, \boldsymbol{\theta}_k)$ (eq:70) ou $J_k(\boldsymbol{\theta}_k)$ (eq:72) ;
4. On identifie la fréquence dont l'amplitude estimée est la plus petite et on la supprime.
5. On continue cette procédure pour $k = M - 1, M - 2, \dots, 0$, et finalement on choisit comme solution $(\hat{\mathbf{x}}_k, \hat{\mathbf{f}}_k)$ correspondant au $\min_k J_k$

Dans la section suivante quelques résultats de simulations sont fournis pour illustrer les performances de ces deux derniers algorithmes.

6 Exemples de résultats de simulation

Pour illustrer les propos de cet article considérons le cas de l'estimation des sinusoides. Nous avons effectué deux ensembles de simulations :

1. Dans le premier, nous avons simulé un signal composé de 2 sinusoides avec les fréquences $\mathbf{f} = [1.0, 1.1]$ Hz et les amplitudes $\mathbf{x} = [0.5, 1.0]$:

$$y(t) = \sum_{j=1}^2 x_j * \sin(2\pi f_j t) + b(t) \quad (73)$$

2. Dans le deuxième, nous avons simulé un signal composé de 5 sinusoides avec des fréquences $\mathbf{f} = [0.6, 0.8, 1.0, 1.1, 1.4]$ Hz. et des amplitudes $\mathbf{x} = [0.3, 0.8, 0.5, 1.0, 0.2]$:

$$y(t) = \sum_{j=1}^5 x_j * \sin(2\pi f_j t) + b(t) \quad (74)$$

Le pas d'échantillonnage est choisit $\Delta T = 0.25$ S et nous avons considéré deux situations : la première avec un interval d'observation $T = [0, 10]$ S (M=40 échantillons) et la deuxième avec $T = [0, 5]$ S (M=20 échantillons). Dans les deux cas un bruit blanc, centré, gaussien et de variance $\sigma_b^2 = .01$ est ajouté aux observations.

Fig. 1 montre ces signaux et les modules de leurs TFD.

Fig. 2 montre ces mêmes signaux prolongés par des zéros et les modules de leurs TFD.

Fig. 3 montre la fonction de log-vraisemblance

$$L(f_1, f_2, x_1, x_2) = -\ln p(\mathbf{y}|f_1, f_2, x_1, x_2, H_2) = \frac{1}{2\sigma_b^2} \sum_{i=1}^M \left[y_i - \sum_{j=1}^n x_j * \sin(2\pi f_j t) \right]^2 \quad (75)$$

lorsque les amplitudes sont supposées connues et correctes $x_1 = 0.5$ et $x_2 = 1$.

Fig. 4 montre cette même fonction de vraisemblance lorsque les amplitudes sont supposées connues mais incorrectes $x_1 = 1$ et $x_2 = 1$.

Fig. 5 montre le critère MAP

$$\begin{aligned} J(f_1, f_2, x_1, x_2) &= -\ln p(f_1, f_2, x_1, x_2 | \mathbf{y}, H_2) \\ &= V(f_1, f_2, x_1, x_2) + \frac{1}{2\sigma_x^2} \sum_{j=1}^k (x_j - x_0)^2 + \frac{1}{2\sigma_f^2} \sum_{j=1}^k (f_j - f_0)^2 \end{aligned} \quad (76)$$

avec $k = 2$. et lorsque les amplitudes sont supposées connues et correctes $x_1 = 0.5$ et $x_2 = 1$.

Fig. 6 montre ce même critère lorsque les amplitudes sont supposées connues mais incorrectes $x_1 = 1$ et $x_2 = 1$.

Fig. 7 montre le critère MVM

$$J(f_1, f_2) = -\ln p(\mathbf{y}|f_1, f_2) = \ln |\mathbf{P}| + \frac{1}{2\sigma_b^2} \mathbf{y}^t \mathbf{P}^{-1} \mathbf{y}, \quad \text{avec} \quad \mathbf{P} = [\mathbf{A}\mathbf{A}^t + \lambda \mathbf{I}] \quad (77)$$

en fonction de (f_1, f_2) .

Fig. 8 montre le critère MAPM

$$\begin{aligned} J(f_1, f_2) &= -\ln p(f_1, f_2 | \mathbf{y}) = -\ln p(\mathbf{y} | f_1, f_2) - \ln p(f_1, f_2) \\ &= \ln |\mathbf{P}| + \frac{1}{2\sigma_b^2} \mathbf{y}^t \mathbf{P}^{-1} \mathbf{y} + \frac{1}{2\sigma_f^2} \sum_{j=1}^k (f_j - f_0)^2 \end{aligned}$$

en fonction de (f_1, f_2) .

Fig. 9 et 10 montrent les résultats obtenus en minimisant le critère MAP à l'aide des algorithmes 2 et 3 sur les deux signaux. Nous avons constaté que l'algorithme 3 a fournis toujours de meilleurs résultats que l'algorithme 2.

Fig. 11 montre la variation du critères MAP: $J_k(\mathbf{x}_k, \mathbf{f}_k)$ en fonction de k . Nous avons constaté que le nombre k est bien estimé par l'algorithme 3 et en général sur-estimé par l'algorithme 2, mais par contre les amplitudes correspondant aux fréquences indésirables sont très faibles.

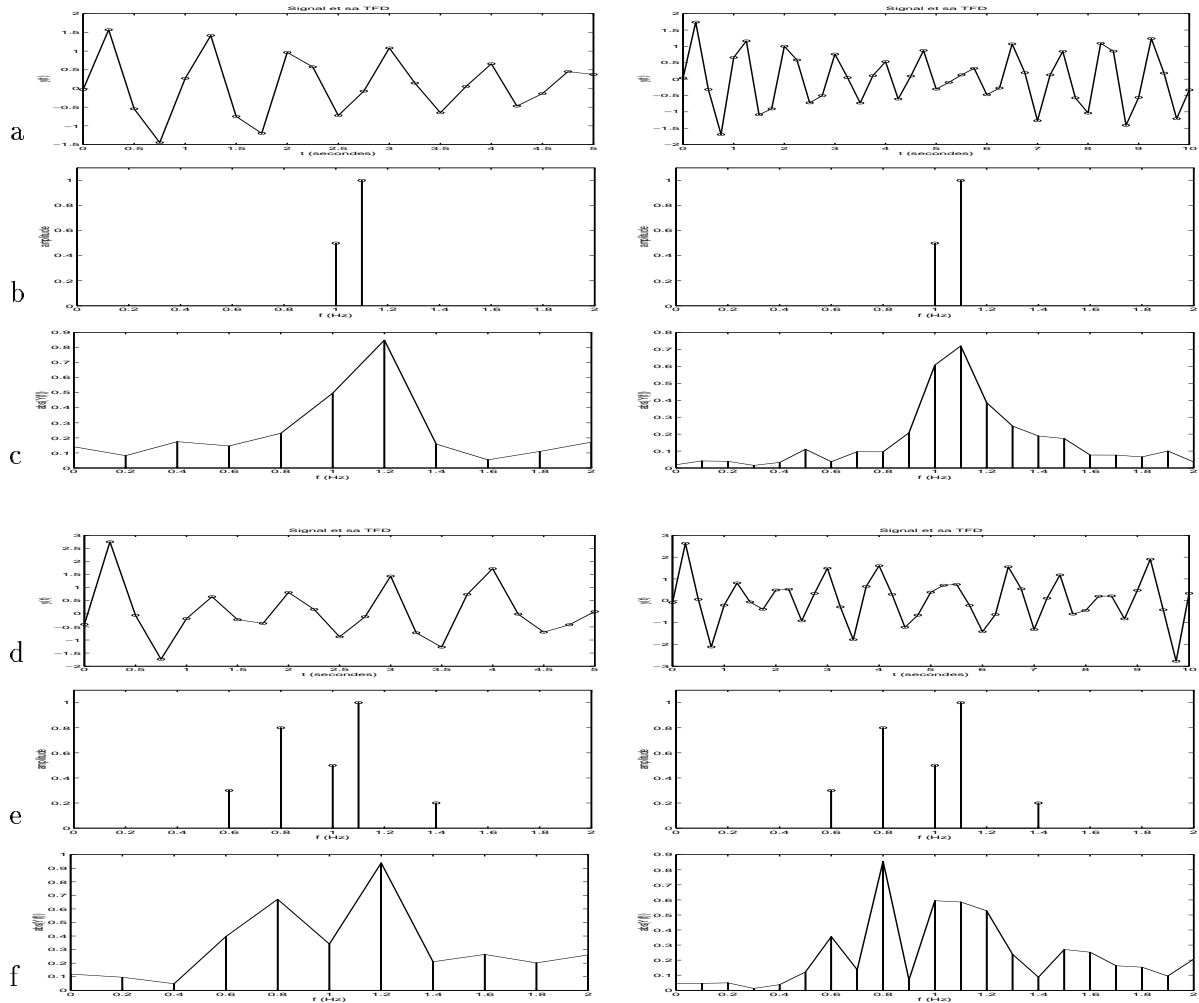


FIG. 1 - Signaux et leurs TFD :

a) signal $y(t) = \sum_{j=1}^2 x_j * \sin(2\pi f_j t) + b(t)$ avec $\mathbf{f} = [1.0, 1.1]$ Hz, $\mathbf{x} = [0.5, 1.0]$, $\Delta T = 0.25$ S et $\sigma_b^2 = .1$, observé sur l'intervall $T = [0, 5]$ S (20 échantillons) et sur l'intervall $T = [0, 10]$ S (40 échantillons).

b) Spectres théoriques.

c) modules de leurs TFD: $|Y(f)|$ calculé sur l'intervall $[0, \frac{1}{2\Delta T} = 2]$ Hz.

d) signal $y(t) = \sum_{j=1}^5 x_j * \sin(2\pi f_j t) + b(t)$ avec $\mathbf{f} = [0.6, 0.8, 1.0, 1.1, 1.4]$ Hz, $\mathbf{x} = [0.3, 0.8, 0.5, 1.0, 0.2]$, $\Delta T = 0.25$ S et $\sigma_b^2 = .1$, observé sur l'intervall $T = [0, 5]$ S (20 échantillons) et sur l'intervall $T = [0, 10]$ S (40 échantillons).

e) Spectres théoriques.

f) modules de leurs TFD: $|Y(f)|$ calculé sur l'intervall $[0, \frac{1}{2\Delta T} = 2]$ Hz.

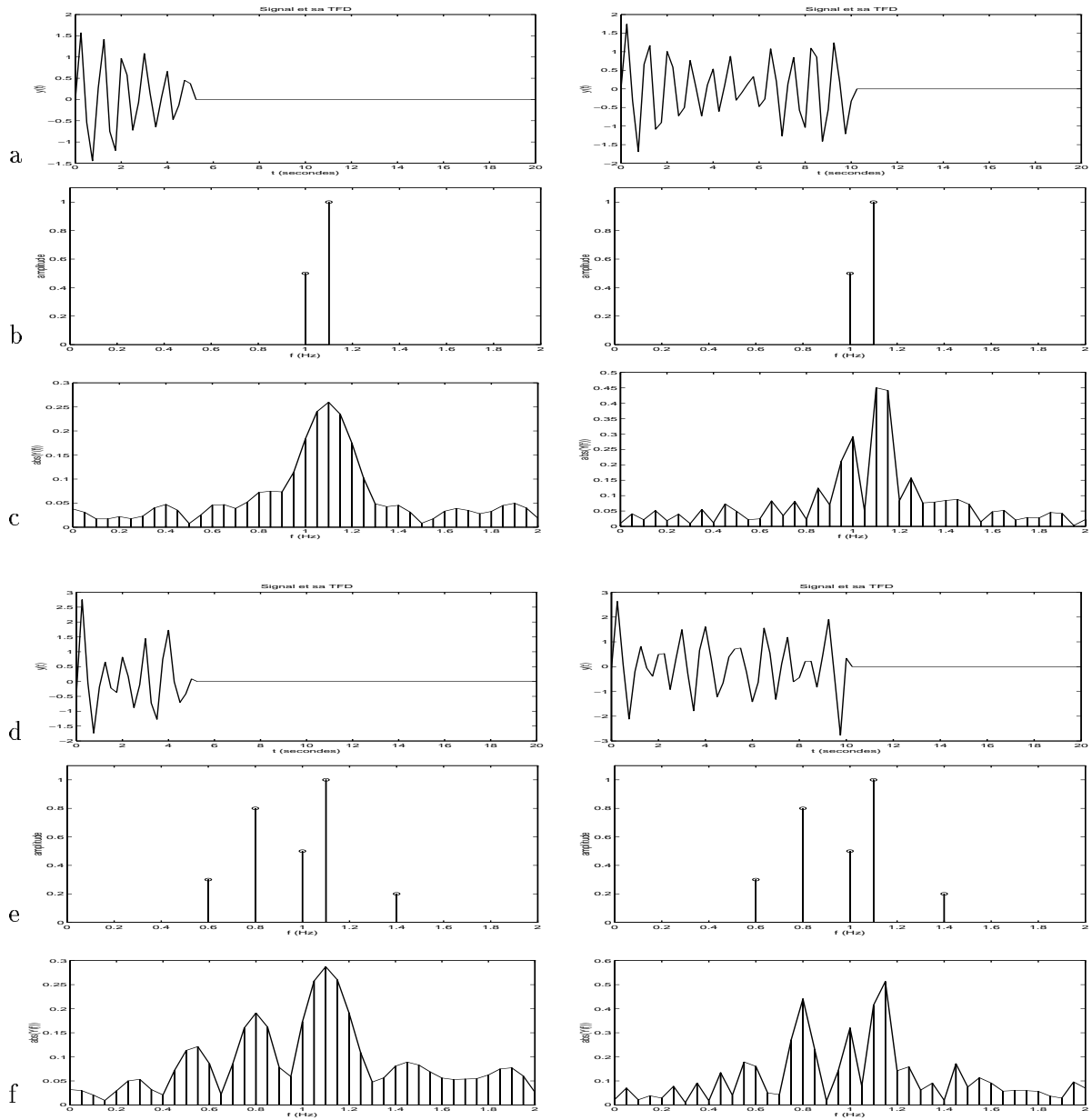


FIG. 2 - Signaux étendus par zéros et leurs TFD :

Les signaux de la figures 1 étendus par zéros et leurs TFD. Le nombre des zéros ajouté est tel que la résolution de la TFD soit de .05 Hz ($n=50$).

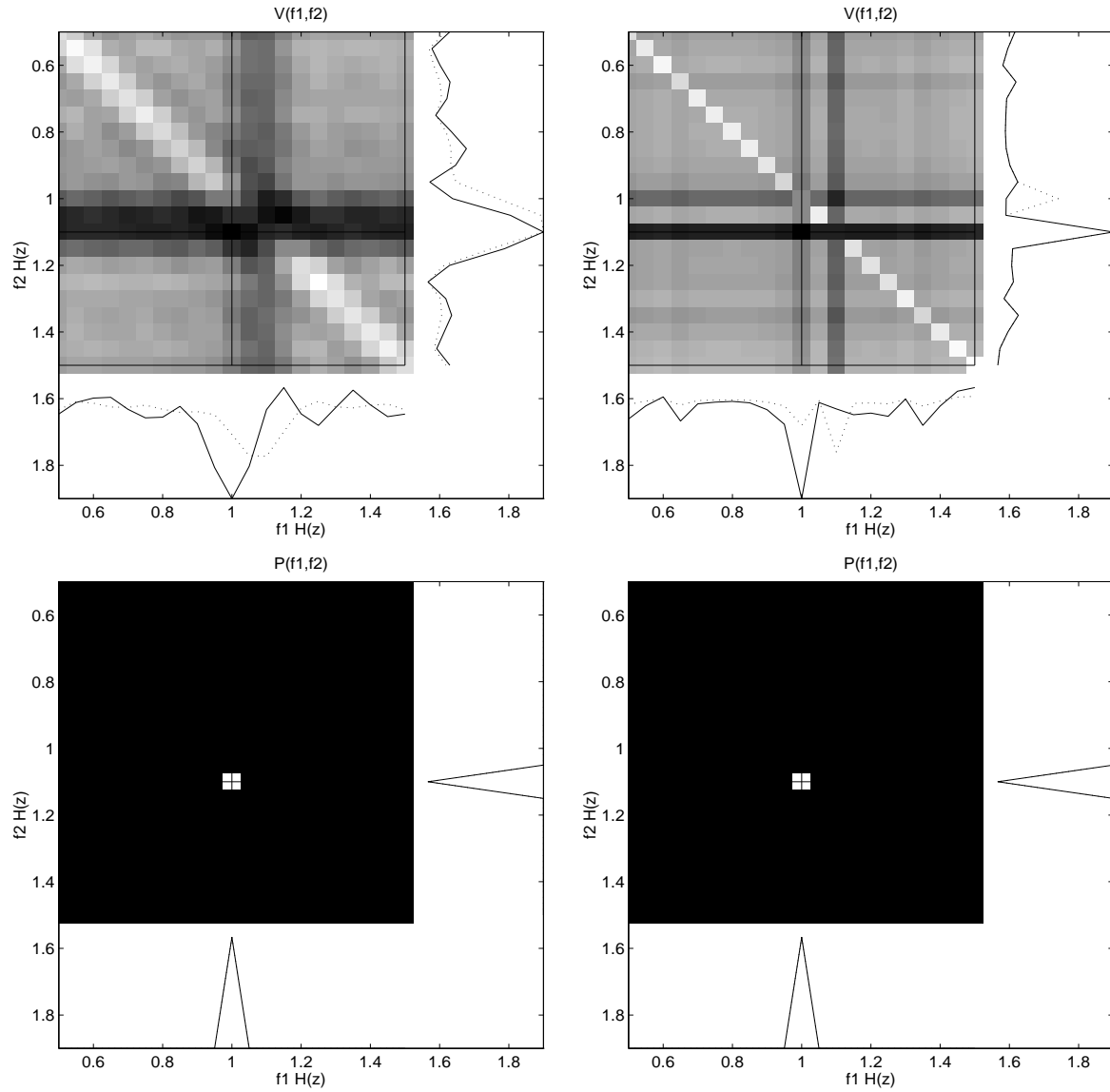


FIG. 3 - Visualisation de la fonction de vraisemblance :

La vraisemblance $V(f_1, f_2)$ et $L(f_1, f_2) = -\ln V(f_1, f_2)$, lorsque les amplitudes sont supposées connues et correctes ($x_1 = .5, x_2 = 1$).

$$a) L(f_1, f_2) = \frac{1}{2\sigma_b^2} \sum_{i=1}^M \left[y_i - \sum_{j=1}^2 x_j * \sin(2\pi f_j t) \right]^2.$$

Nous avons aussi tracé les courbes

$L_1(f_1) = L(f_1, f_2 = cte)$ et $L_2(f_2) = L(f_1 = cte, f_2)$ (traits -) ainsi que les marginales

$L_1(f_1) = \sum_{f_2} L(f_1, f_2)$ et $L_2(f_2) = \sum_{f_1} L(f_1, f_2)$ (points .).

$$b) V(f_1, f_2) = p(\mathbf{y} | f_1, f_2) = \frac{1}{Z} \exp[-L(f_1, f_2)]$$

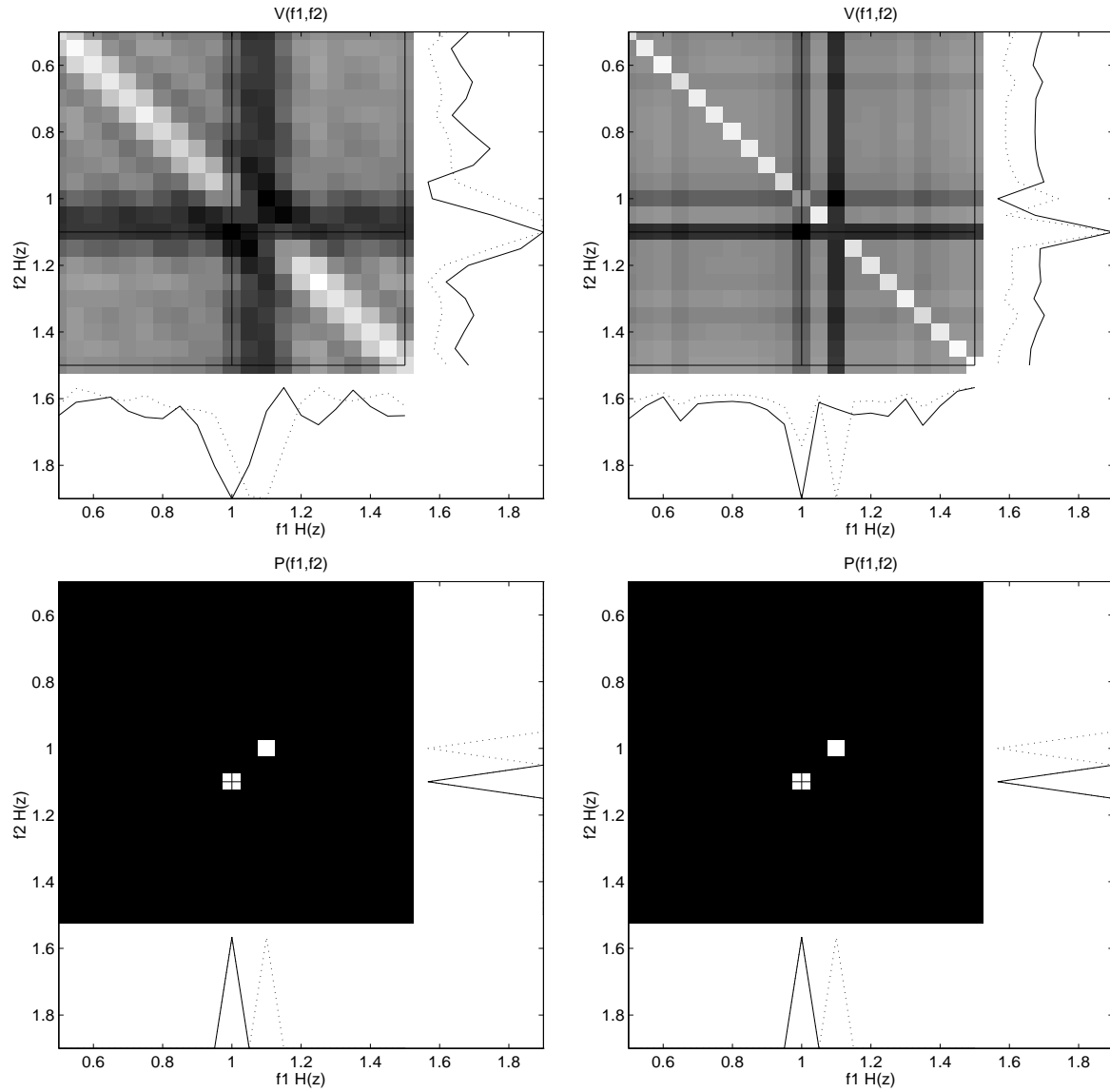


FIG. 4 - Visualisation de la fonction de vraisemblance :

La vraisemblance $V(f_1, f_2)$ et $L(f_1, f_2) = -\ln V(f_1, f_2)$, lorsque les amplitudes sont supposées connues mais incorrectes ($x_1 = 1, x_2 = 1$).

$$a) L(f_1, f_2) = \frac{1}{2\sigma_b^2} \sum_{i=1}^M \left[y_i - \sum_{j=1}^2 x_j * \sin(2\pi f_j t) \right]^2.$$

Nous avons aussi tracé les courbes

$L_1(f_1) = L(f_1, f_2 = cte)$ et $L_2(f_2) = L(f_1 = cte, f_2)$ (traits -) ainsi que les marginales $L_1(f_1) = \sum_{f_2} L(f_1, f_2)$ et $L_2(f_2) = \sum_{f_1} L(f_1, f_2)$ (points .).

$$b) V(f_1, f_2) = p(\mathbf{y} | f_1, f_2) = \frac{1}{Z} \exp[-L(f_1, f_2)]$$

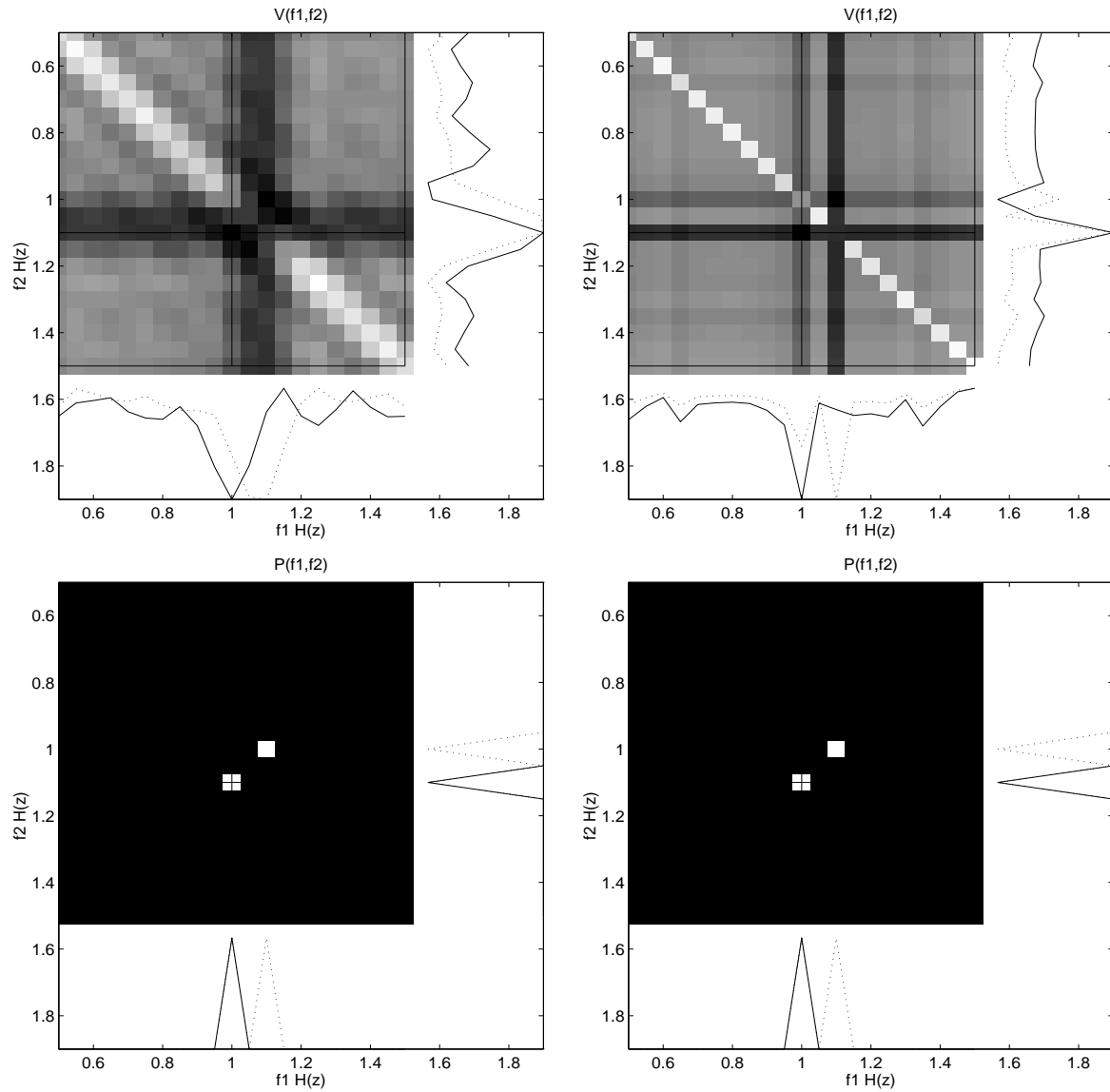


FIG. 5 - Visualisation du critère MAP :

Le critère MAP $J(f_1, f_2)$ lorsque les amplitudes sont supposées connues et correctes ($x_1 = .5, x_2 = 1$).

$$a) J(f_1, f_2) = \frac{1}{2\sigma_b^2} \sum_{i=1}^M \left[y_i - \sum_{j=1}^2 x_j * \sin(2\pi f_j t) \right]^2 + \frac{1}{2\sigma_f^2} \sum_{j=1}^2 (f_j - f_0)^2.$$

Nous avons aussi tracé les courbes

$J_1(f_1) = J(f_1, f_2 = cte)$ et $J_2(f_2) = J(f_1 = cte, f_2)$ (traits -) ainsi que les marginales $J_1(f_1) = \sum_{f_2} J(f_1, f_2)$ et $J_2(f_2) = \sum_{f_1} J(f_1, f_2)$ (points .).

$$b) p(f_1, f_2 | \mathbf{y}) = \frac{1}{Z} \exp[-J(f_1, f_2)]$$

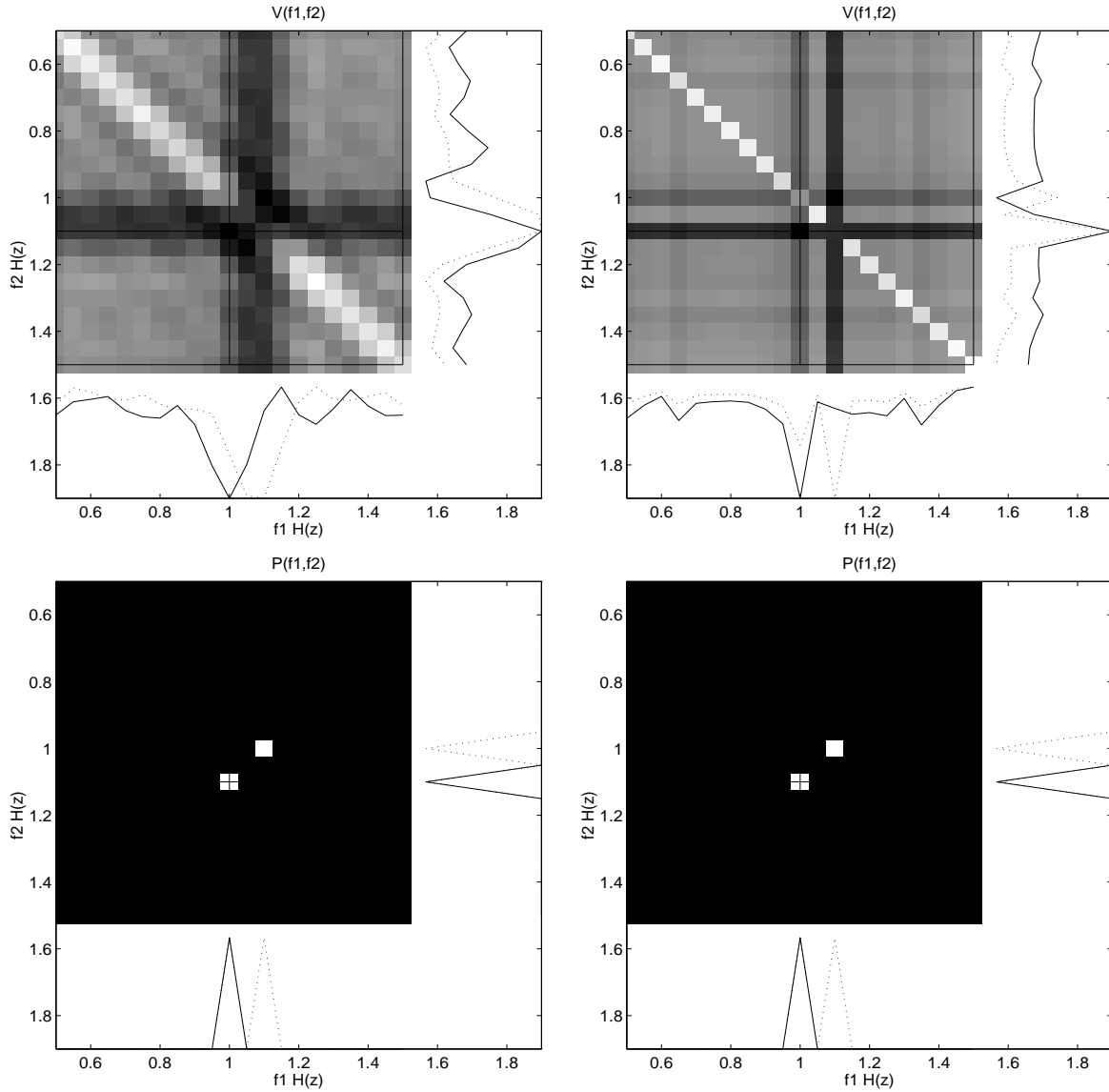


FIG. 6 - Visualisation du critère MAP :

Le critère MAP $J(f_1, f_2)$ lorsque les amplitudes sont supposées connues mais incorrectes ($x_1 = 1, x_2 = 1$).

$$a) J(f_1, f_2) = \frac{1}{2\sigma_b^2} \sum_{i=1}^M \left[y_i - \sum_{j=1}^2 x_j * \sin(2\pi f_j t) \right]^2 + \frac{1}{2\sigma_f^2} \sum_{j=1}^2 (f_j - f_0)^2.$$

Nous avons aussi tracé les courbes

$J_1(f_1) = J(f_1, f_2 = cte)$ et $J_2(f_2) = J(f_1 = cte, f_2)$ (traits -) ainsi que les marginales $J_1(f_1) = \sum_{f_2} J(f_1, f_2)$ et $J_2(f_2) = \sum_{f_1} J(f_1, f_2)$ (points .).

$$b) p(f_1, f_2 | \mathbf{y}) = \frac{1}{Z} \exp[-J(f_1, f_2)]$$

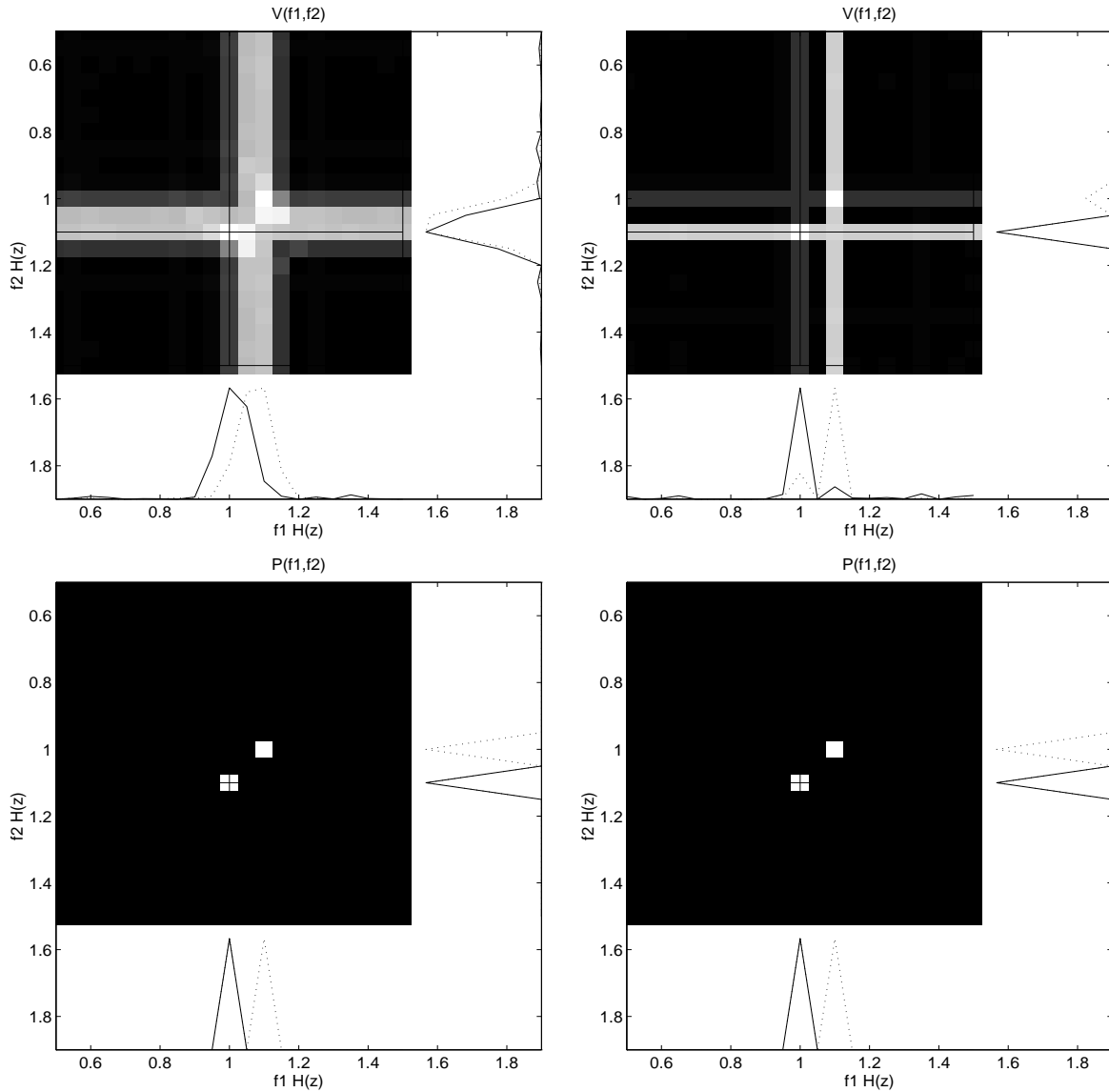


FIG. 7 - Visualisation du critère MVM:

a) $J(f_1, f_2) = -\ln p(\mathbf{y}|f_1, f_2) = \ln |\mathbf{P}| + \frac{1}{2\sigma_b^2} \mathbf{y}^t \mathbf{P}^{-1} \mathbf{y}$, avec $\mathbf{P} = [\mathbf{A}\mathbf{A}^t + \lambda \mathbf{I}]$.

Nous avons aussi tracé les courbes

$J_1(f_1) = J(f_1, f_2 = cte)$ et $J_2(f_2) = J(f_1 = cte, f_2)$ (traits -) ainsi que les marginales
 $J_1(f_1) = \sum_{f_2} J(f_1, f_2)$ et $J_2(f_2) = \sum_{f_1} J(f_1, f_2)$ (points .).

b) $p(\mathbf{y}|f_1, f_2) = \frac{1}{Z} \exp[-J(f_1, f_2)]$

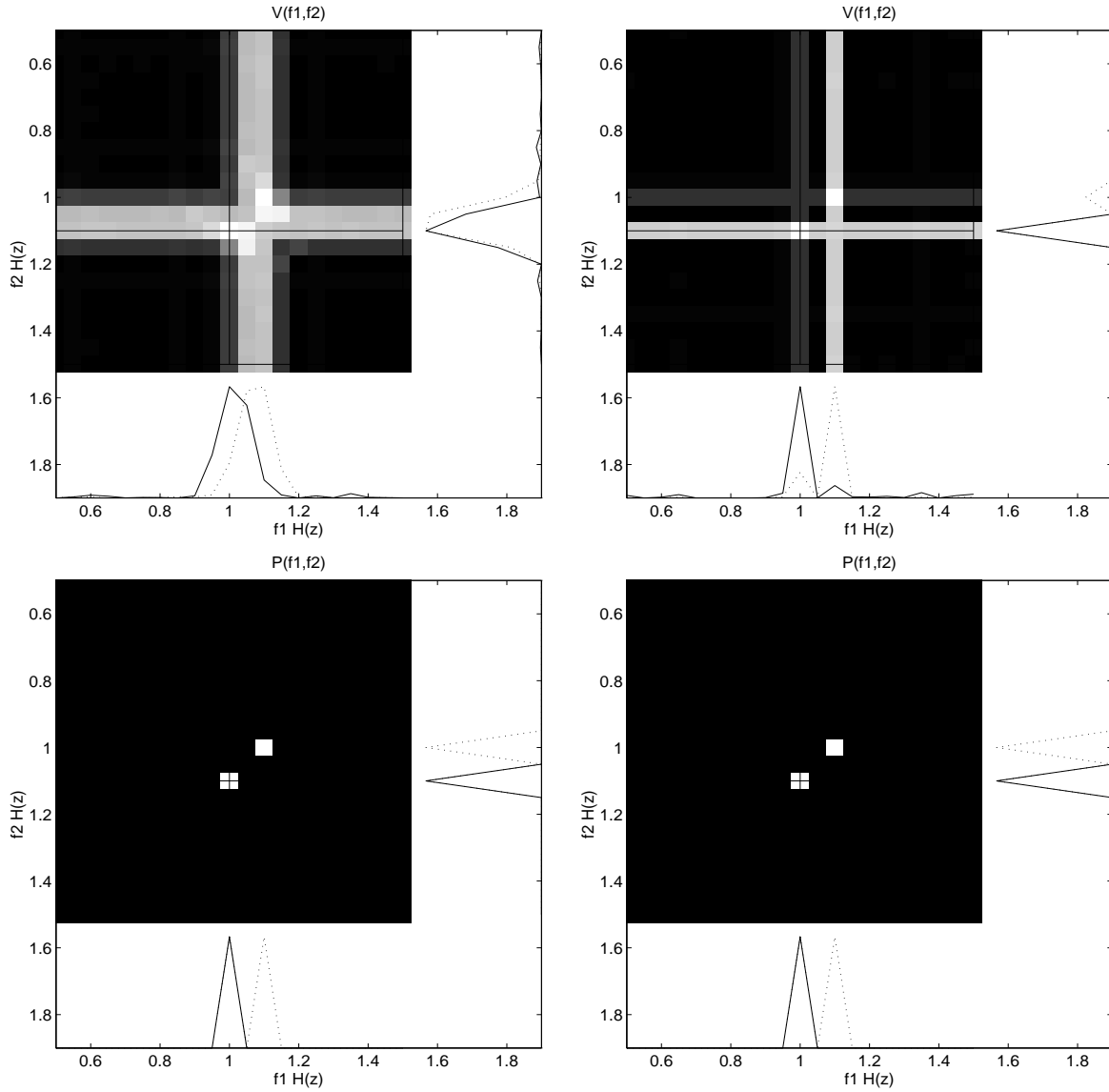


FIG. 8 - Visualisation du critère MAPM:

$$a) J(f_1, f_2) = -\ln p(f_1, f_2 | \mathbf{y}) = \ln |\mathbf{P}| + \frac{1}{2\sigma_b^2} \mathbf{y}^t \mathbf{P}^{-1} \mathbf{y} + \frac{1}{2\sigma_f^2} \sum_{j=1}^k (f_j - f_0)^2.$$

Nous avons aussi tracé les courbes

$J_1(f_1) = J(f_1, f_2 = cte)$ et $J_2(f_2) = J(f_1 = cte, f_2)$ (traits -) ainsi que les marginales

$J_1(f_1) = \sum_{f_2} J(f_1, f_2)$ et $J_2(f_2) = \sum_{f_1} J(f_1, f_2)$ (points .).

$$b) p(f_1, f_2 | \mathbf{y}) = \frac{1}{Z} \exp[-J(f_1, f_2)]$$

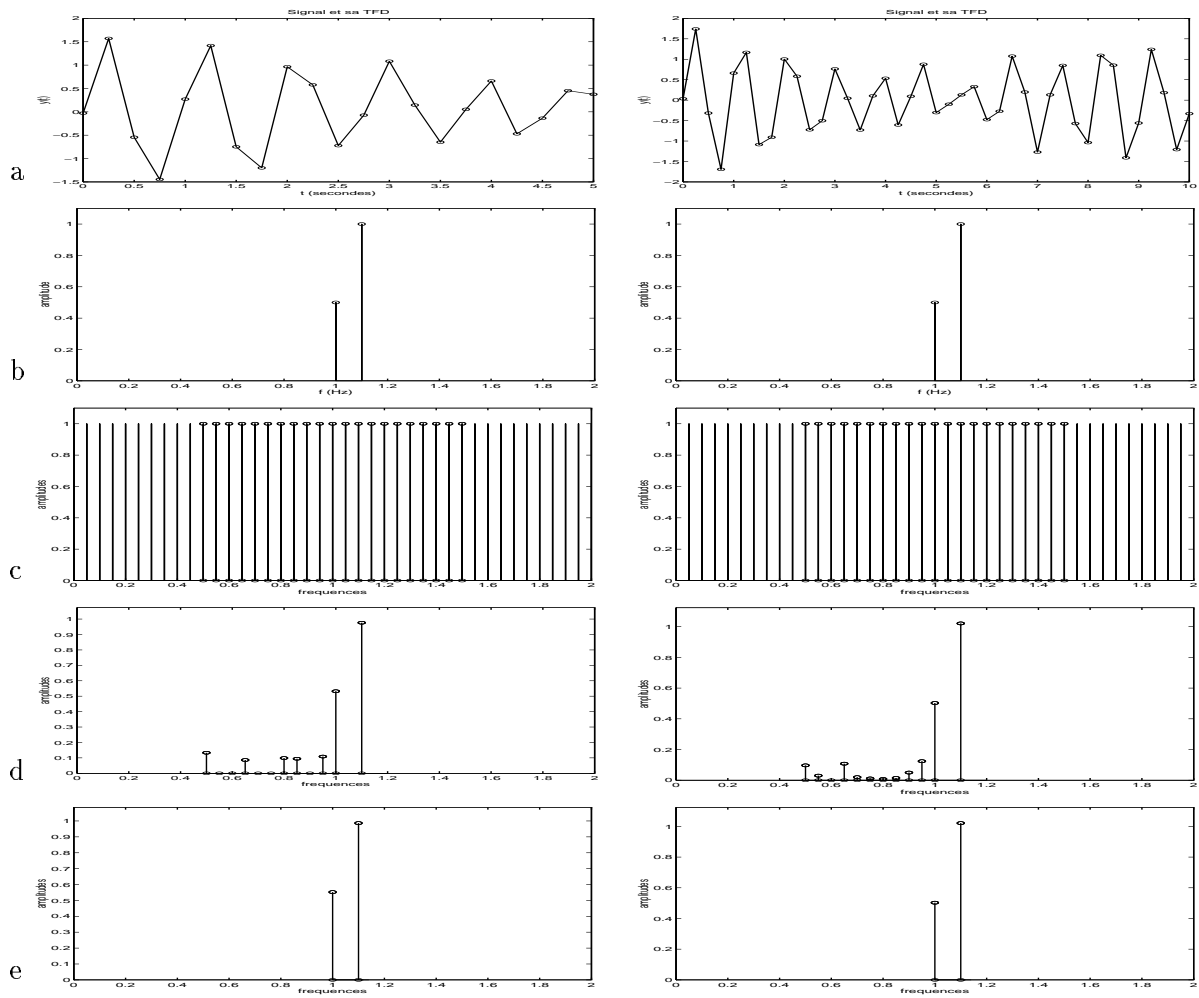


FIG. 9 - Estimation de la solution au sens du MAP:

- a) Signaux composés de 2 fréquences (à droite $N = 20$, à gauche $N = 40$),
- b) Spectres théoriques,
- c) Initialisation,
- d) Résultats obtenus par Algorithme 2.
- e) Résultats obtenus par Algorithme 3.

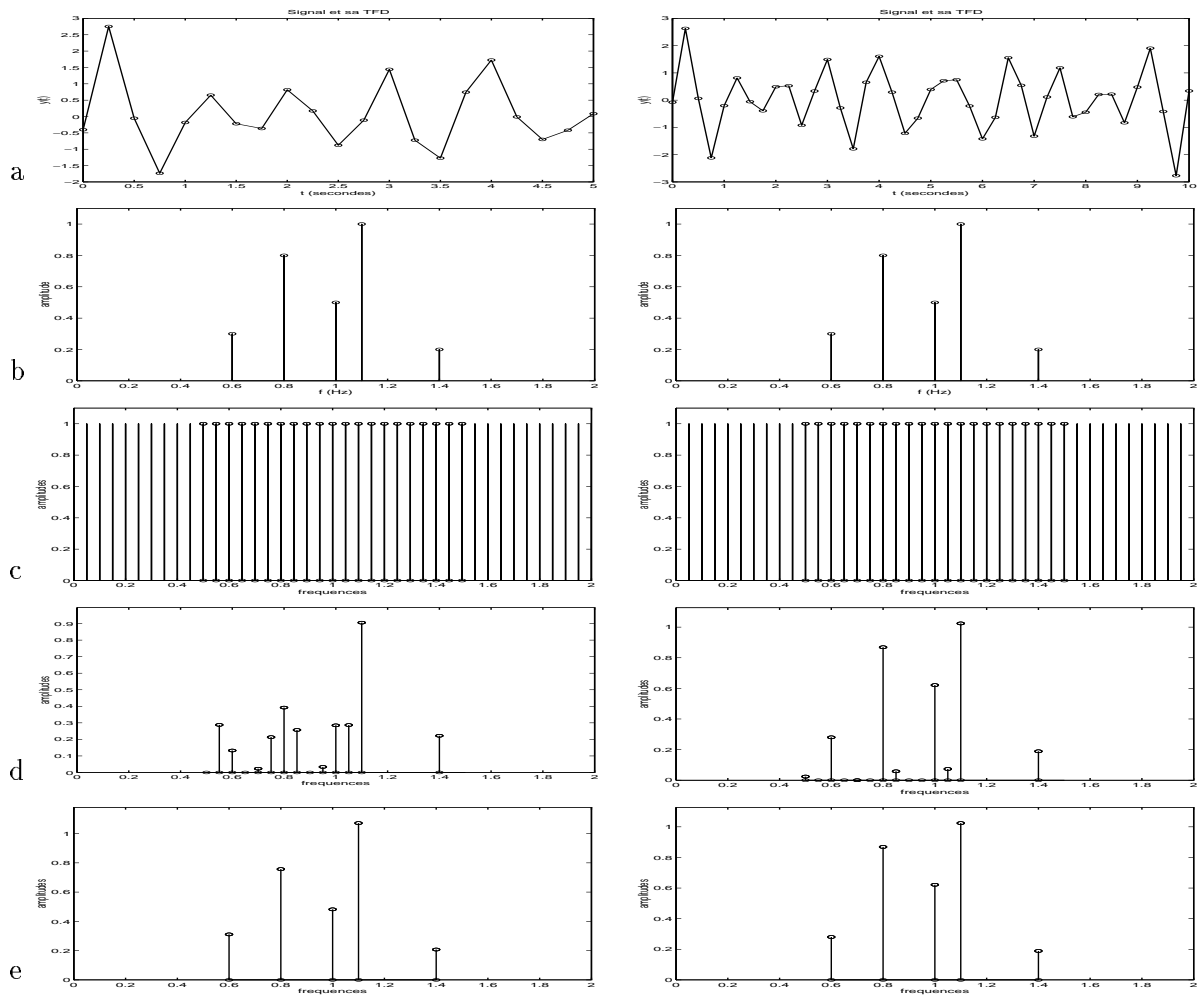


FIG. 10 - Estimation de la solution au sens du MAP:

- a) Signaux composés de 5 fréquences (à droite $N = 20$, à gauche $N = 40$),
- b) Spectres théoriques
- c) Initialisations,
- d) Résultats obtenus par Algorithme 2.
- e) Résultats obtenus par Algorithme 3.

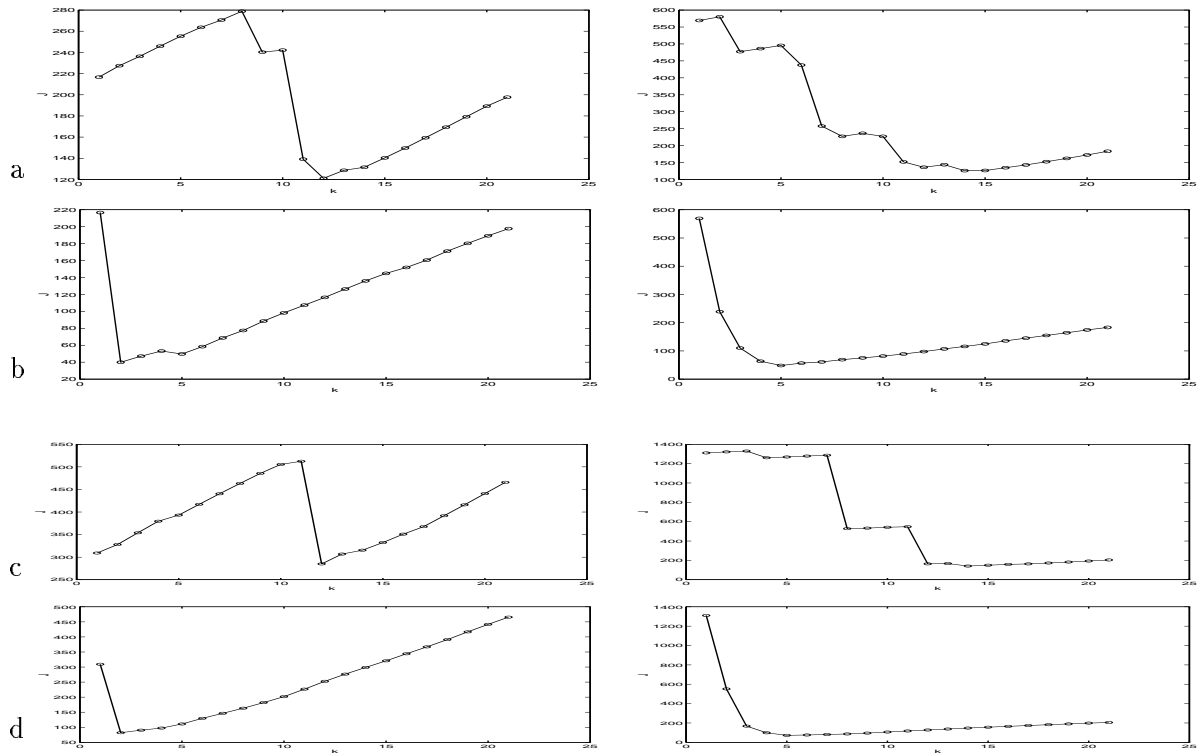


FIG. 11 - Critères en fonction du nombre de sources k
 À droite : Signaux composés de 2 fréquences, À gauche : Signaux composés de 5 fréquences.

$a, b) nT = 21, \quad c, d) nT = 41$
 $a, c) \text{ Résultat obtenu par l'algorithme 2.}$
 $b, d) \text{ Résultat obtenu par l'algorithme 3.}$

7 Annexe

L'objet de cette annexe est de montrer que plusieurs problèmes dans divers domaines peuvent avoir à résoudre le même problème inverse décrit en introduction.

7.1 Estimation des paramètres d'une mélange de lois par la méthode des moments

Supposons que nous avons observé n échantillons $\{x_1, \dots, x_n\}$ d'une variable aléatoire X dont sa distribution est supposée être une mélange de K distributions p_k :

$$p(x; a; \boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{k=1}^K \alpha_k p_k(x; a, \theta_k), \quad x > 0, \quad 0 < \alpha_k < 1, \quad \sum_{k=1}^K \alpha_k = 1.$$

et où

$$p_i(x; a, \theta_i) = \exp[-at_1(x) - \theta_i t_2(x) + h(a, \theta_i)]$$

est une loi appartenant à la famille des lois exponentielle à deux paramètres a et θ et où a est le paramètre de forme et θ est le paramètre d'échelle, ce qui signifie que $t_1(x)$ et $t_2(x)$ sont telles que

$$\mu_k = E\{X^k\} = \theta \mu_{k-1}.$$

Le problème est l'estimation des paramètres de la mélange $(a; \boldsymbol{\theta}, \boldsymbol{\alpha})$ par la méthode des moments qui consiste à résoudre le système :

$$\mu_k(a; \boldsymbol{\theta}, \boldsymbol{\alpha}) = \int x^k p(x; a; \boldsymbol{\theta}, \boldsymbol{\alpha}) dx = m_k = \frac{1}{n} \sum_{j=1}^n x_j^k, \quad k = 0, 1, 2, \dots, 2K - 1$$

Remplaçant $p(x; a; \boldsymbol{\theta}, \boldsymbol{\alpha})$ dans ces équations on obtient :

$$\tau_k(m_k, a) = \sum_{i=1}^K \alpha_i \theta_i^k, \quad k = 0, 1, 2, \dots, 2K - 1$$

qui peut s'écrire sous la forme matricielle :

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \theta_1 & \theta_2 & \dots & \theta_K \\ \theta_1^2 & \theta_2^2 & \dots & \theta_K^2 \\ \vdots & \vdots & & \vdots \\ \theta_1^{2K-1} & \theta_2^{2K-1} & \dots & \theta_K^{2K-1} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{pmatrix} = \begin{pmatrix} \tau_0 \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_K \end{pmatrix}$$

Les $\tau_k(m_k, a)$ sont fonctions du paramètres a et des moments empiriques m_k qui sont calculés en utilisant les observations. Par exemple dans le cas des lois exponentielles $a = 0$ et on a $\tau_k = k! m_k$. Dans le cas des lois Gamma on a $\tau_k = a(a+1) \dots (a+k-1) m_k$ et dans le cas des lois de Weibull on a $\Gamma(1+k/a) m_k$.

7.2 Estimation des coordonnées des sommets d'un polygone à partir de ses moments

Considérons la fonction

$$f(x, y) = \begin{cases} 1 & \text{si } (x, y) \in P, \\ 0 & \text{ailleurs,} \end{cases}$$

où P représente la surface d'un polygone de n sommets dont les coordonnées sont : (x_j, y_j) , $j = 1, \dots, n$ et définissons ces moments harmoniques :

$$c_k = c_{k0} = \iint_P f(x, y) z^k dx dy$$

On peut établir une relation entre les c_k et les coordonnées des sommets du polygone P en utilisant le théorème suivant :

Théorème 1 Soit $\{z_j = (x_j + iy_j), j = 1, \dots, n\}$ les coordonnées des sommets d'un polygone P . Alors il existe des coefficients a_j , $j = 1, \dots, n$, tels que pour une fonction $h(z)$ analytique sur la fermeture de P on ait :

$$\iint_P h''(z) dx dy = \sum_{j=1}^n a_j h(z_j).$$

Si le polygone n'est pas dégénéré et si P est une région simplement convexe alors ces coefficients sont non nuls.

Utilisant ce théorème en remplaçant $h(z) = z^k$ on a :

$$\iint_P (z^k)'' dx dy = \sum_{j=1}^n a_j z_j^k$$

La partie gauche de cette équation s'écrit :

$$\iint_P h''(z) dx dy = k(k-1) \iint_P z^{k-2} dx dy = k(k-1)c_{k-2}$$

où c_{k-2} sont les moments harmoniques de $f(x, y)$. Notant que $\tau_k = k(k-1)c_{k-2}$, on a

$$\tau_k = \sum_{j=1}^n a_j z_j^k.$$

Écrivant cette équation pour $k = 0, 1, \dots, M$ on a :

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_n \\ z_1^2 & z_2^2 & \cdots & z_n^2 \\ \vdots & \vdots & & \vdots \\ z_1^M & z_2^M & \cdots & z_n^M \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \tau_0 \\ \tau_1 \\ \tau_2 \\ \vdots \\ \tau_m \end{pmatrix}$$

7.3 Localisation de source ou calibration d'antenne

Considérons le cas d'un réseaux de M antennes positionnés aux points (ξ_m, η_m) , $m = 1, \dots, M$ et supposons qu'il y a N sources placées dans les directions $[\theta_1, \dots, \theta_n]$. Le signal mesuré par le capteur m à un instant donné t est

$$y_m(t) = \sum_{n=1}^N x_n(t) e^{-j\omega_m(\theta_n)} + b_m(t)$$

où $x_n(t)$ est l'amplitude du signal émis par la source n , $b_m(t)$ est le bruit de mesure et où

$$\omega_m(\theta_n) = \frac{2\pi}{\lambda} (\xi_m \sin \theta_n + \eta_m \cos \theta_n)$$

λ est la longueur d'onde. Utilisant les notations

$$\mathbf{y}(t) = [y_1(t), \dots, y_m(t)], \quad \mathbf{b}(t) = [b_1(t), \dots, b_b(t)], \quad \boldsymbol{\theta} = [\theta_1(t), \dots, \theta_n(t)],$$

$$\mathbf{x}(t) = [x_1(t), \dots, x_n(t)], \quad \mathbf{a}(\theta_n) = [e^{-j\omega_1(\theta_n)}, \dots, e^{-j\omega_m(\theta_n)}], \quad \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1) | \dots | \mathbf{a}(\theta_n)]$$

on obtient

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta}) \mathbf{x}(t) + \mathbf{b}(t).$$

Si de plus les antennes sont placées sur l'axe des ξ : ($\eta_m = 0, \xi_m = (m-1)\delta\xi$) on a

$$\omega_m(\theta_n) = \frac{2\pi \sin \theta_n}{\lambda} (m-1)\delta x \longrightarrow \mathbf{a}(\theta_n) = [1, e^{-j\omega(\theta_n)}, e^{-j2\omega(\theta_n)}, \dots, e^{-jM\omega(\theta_n)}]$$

et si on note par $z_n = e^{-j\omega(\theta_n)}$ on a $\mathbf{a}(z_n) = [1, z_n^1, z_n^2, \dots, z_n^{(M-1)}]$ et finalement l'équation

$$\begin{pmatrix} y_0(t) \\ y_1(t) \\ y_2(t) \\ \vdots \\ y_m(t) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_n \\ z_1^2 & z_2^2 & \dots & z_n^2 \\ \vdots & \vdots & & \vdots \\ z_1^{M-1} & z_2^{M-1} & \dots & z_n^{M-1} \end{pmatrix} \begin{pmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \\ \vdots \\ x_n(t) \end{pmatrix} + \begin{pmatrix} b_0(t) \\ b_1(t) \\ b_2(t) \\ \vdots \\ b_m(t) \end{pmatrix}$$

Localisation :

Étant donnée $\mathbf{y}(t)$ estimer les directions $\boldsymbol{\theta}$ et les amplitudes $\mathbf{x}(t)$

Calibration d'antenne :

Supposons que les amplitudes \mathbf{x} soient connues mais que chaque antenne introduit un gain et une déphasage inconnue. Notant par $\mathbf{G} = \text{diag}\{g_1, \dots, g_m\}$ les M nombres complexes désignant ces gains, le problème devient celui de l'estimation de $\boldsymbol{\theta}$ et de $\mathbf{g} = [g_1, \dots, g_m]$.

$$\mathbf{y}(t) = \mathbf{G} \mathbf{A} \mathbf{x}(t) + \mathbf{b}(t)$$

On peut aussi envisager d'estimer simultanément les trois vecteurs \mathbf{x} , $\boldsymbol{\theta}$ et \mathbf{g} .

7.4 Déconvolution impulsionnelle

Dans un problème de déconvolution impulsionnelle on a

$$y(t) = \sum_{n=1}^n x_n h(t - \theta_n) + b(t),$$

où $h(t)$ est la réponse impulsionnelle. Passant dans le domaine de Fourier on a

$$Y(f) = \sum_{n=1}^n x_n H(f) \exp[-j2\pi f \theta_n] + B(f),$$

où $H(f)$ est la réponse fréquentielle du système de mesure. Si $h(t)$ est connue on peut envisager d'estimer les positions $\boldsymbol{\theta}$ et les amplitudes \boldsymbol{x} des impulsions par :

$$Y(f_m) = H(f_m) \sum_{n=1}^n x_n \exp[-j2\pi f_m \theta_n] + B(f_m), \quad m = 0, \dots, M-1$$

En particulier si on choisit $f_m = m f_0$, $m = 0, \dots, M-1$ et si on note par $z_n = \exp[-j2\pi f_0 \theta_n]$ on aura à résoudre $\boldsymbol{y} = \boldsymbol{H} \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{x} + \boldsymbol{b}$ avec

$$\boldsymbol{y} = [Y(f_0), \dots, Y(f_{M-1})]^t, \quad \boldsymbol{b} = [B(f_0), \dots, B(f_{M-1})]^t, \quad \boldsymbol{x} = [x_1, \dots, x_N]^t,$$

et

$$\boldsymbol{H} = \text{diag} \{H(f_0), \dots, H(f_{M-1})\}, \quad \boldsymbol{A} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_n \\ z_1^2 & z_2^2 & \dots & z_n^2 \\ \vdots & \vdots & & \vdots \\ z_1^{M-1} & z_2^{M-1} & \dots & z_n^{M-1} \end{pmatrix}$$

et on retrouve l'analogie avec le problème de la localisation de source.

On peut même envisager d'estimer aussi la réponse impulsionnelle lorsqu'elle n'est pas connue sous réserve de faire l'hypothèse que

$$h(t) = \sum_{m=0}^{M-1} h_m \exp[-j2\pi f_m t] \longrightarrow H(f_m) = \sum_{m=0}^{M-1} h_m \delta(f - f_m)$$

on aura à résoudre $\boldsymbol{y} = \boldsymbol{H} \boldsymbol{A}(\boldsymbol{\theta}) \boldsymbol{x}(t) + \boldsymbol{b}$ où $\boldsymbol{H} = \text{diag} \{h_0, \dots, h_{M-1}\}$ et on retrouve l'analogie avec la calibration d'antenne.

7.5 Séparation d'ondes en sismique

Considérons une antenne linéaire composée de M capteurs et supposons qu'il y a N ondes planes. La trace sismique reçue sur le capteur k s'écrit :

$$d_k(t) = \sum_{n=1}^n w_n(t - \theta_{kn}) + b_k(t),$$

où $w_n(t)$ est la forme d'onde émise par la $n^{\text{ième}}$ source et θ_{kn} est le retard de propagation de cette onde entre le capteur de référence et le $k^{\text{ième}}$ capteur. Passant dans le domaine de Fourier on a

$$D_k(f) = \sum_{n=1}^n W_n(f) \exp[-j2\pi f\theta_{kn}] + B_k(f),$$

où $D_k(f)$ est la TF de $d_k(t)$, $W_n(f)$ est la TF de $w_n(t)$, $B_k(f)$ est la TF de $b_k(t)$. Utilisant les notations

$$\mathbf{y}(f) = [D_1(f), \dots, D_m(f)]^t, \quad \mathbf{b}(f) = [B_1(f), \dots, B_m(f)]^t,$$

$$\mathbf{w}(f) = [w_1(f), \dots, w_n(f)], \quad \mathbf{a}(\theta_n) = [e^{-j2\pi f\theta_{1n}}, \dots, e^{-j2\pi f\theta_{Mn}}], \quad \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1) | \dots | \mathbf{a}(\theta_n)]$$

on obtient

$$\mathbf{y}(f) = \mathbf{A}(\boldsymbol{\theta}, f)\mathbf{w}(f) + \mathbf{b}(f).$$

Ensuite on cherche à résoudre cette équation pour chaque fréquence $f = \{0, f_0, 2f_0, \dots, Kf_0\}$ en faisant l'hypothèse que les ondes sismiques $w_n(t)$ peuvent être caractérisée par

$$w_n(t) = \sum_{l=0}^K w_{nk} \exp[-j2\pi f_k t].$$

7.6 Estimation de sinusoides

Considérons un signal composé de N sinusoides :

$$y(t) = \sum_{i=1}^n x_n \exp[-j\omega_n t] + b(t), \quad t = 0, \Delta t, 2\Delta t, \dots, M\Delta t$$

Si on note $y(k) = y(k\Delta t)$, et $z_n = \exp[-j\omega_n \Delta t]$ on a

$$\begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ z_1 & z_2 & \dots & z_n \\ z_1^2 & z_2^2 & \dots & z_n^2 \\ \vdots & \vdots & & \vdots \\ z_1^{M-1} & z_2^{M-1} & \dots & z_n^{M-1} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Références

- [1] K. M. Wong, J. P. Reilly, Q. Wu, and S. Qiao, “Estimation of directions of arrival of signals in unknown correlated noise, part i: The MAP approach and its implementation,” *IEEE Transactions on Signal Processing*, vol. SP-40, pp. 2007–2017, Aug. 1992. JFG.
- [2] K. M. Wong, J. P. Reilly, Q. Wu, and S. Qiao, “Estimation of directions of arrival of signals in unknown correlated noise, part ii: Asymptotic behavior and performances of the MAP approach,” *IEEE Transactions on Signal Processing*, vol. SP-40, pp. 2018–2028, Aug. 1992. JFG.
- [3] L. Bretthorst, “Bayesian analysis. i. parameter estimation using quadrature nmr models,” *J. of Magnetic Resonance*, vol. 88, pp. 533–551, 1990.
- [4] L. Bretthorst, “Bayesian analysis. ii. signal detection and model selection,” *J. of Magnetic Resonance*, vol. 88, pp. 552–570, 1990.
- [5] L. Bretthorst, “Bayesian analysis. iii. model selection and parameter estimation,” *J. of Magnetic Resonance*, vol. 88, pp. 571–595, 1990.
- [6] L. Bretthorst, “Bayesian analysis. iv. noise and computing time considerations,” *J. of Magnetic Resonance*, vol. 93, pp. 369–394, 1993.
- [7] A. Nallanathan and W. J. Fitzgerald, “Bayesian model selection applied to spatial signal processing,” *Proceedings of the IEE*, vol. 141, pp. 76–80, February 1994. AMD.
- [8] J.-F. Giovannelli, J. Idier, B. Querleux, A. Herment, and G. Demoment, “Maximum likelihood and maximum a posteriori estimation of Gaussian spectra. Application to attenuation measurement and color Doppler velocimetry,” in *Proceedings of International Ultrasonics Symposium*, vol. 3, (Cannes, France), pp. 1721–1724, novembre 1994. JFG.
- [9] G. Le Besnerais, “Méthodes de calibration et d’autocalibration en radio-astronomie par interférométrie,” *Traitement du Signal*, vol. 10, no. 5, pp. –, 1994. numéro spécial sur la calibration d’antenne.
- [10] J.-F. Giovannelli, A. Herment, and G. Demoment, “A Bayesian method for long AR spectral estimation: a comparative study,” *IEEE Transactions on Ultrasonics Ferroelectrics and Frequency Control*, vol. 43, pp. 220–233, March 1996. JFG.
- [11] F. Dublanchet, P. Duvaut, and J. Idier, *Complex sinusoid analysis by Bayesian deconvolution of the discrete Fourier transform*, pp. 323–328. Santa Fe, U.S.A.: Kluwer Academic Publishers, K. Hanson ed., 1995.
- [12] F. Dublanchet, *Contribution de la méthodologie bayésienne à l’analyse spectrale de raies pures et à la goniométrie haute résolution*. PhD thesis, Université de Paris-Sud, Orsay, octobre 1996.
- [13] J. Capon, “High-resolution frequency wavenumber spectrum analysis,” *Proc. of IEEE*, vol. 57, pp. 1408–1418, 1969.

- [14] H. Ouibrahim, "Prony, Pisarenko, and the matrix pencil: A unified presentation," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-37, no. 1, p. 133, 1989.
- [15] H. Clergeot, "Estimation du spectre d'un signal aléatoire gaussien par le critère du maximum de vraisemblance ou du maximum *a posteriori*," Master's thesis, Université de Paris-Sud, 1982. JFG.
- [16] S. M. Kay, "Recursive maximum likelihood estimation of autoregressive processes," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-21, pp. 56–65, 1983. JFG.
- [17] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. 33, pp. 387–392, avril 1985.
- [18] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-34, pp. 1081–1089, October 1986. JFG.
- [19] S. M. Kay, *Modern Spectral Estimation*. Englewood Cliffs: Prentice-Hall, 1988.
- [20] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-36, pp. 1553–1560, octobre 1988.
- [21] S. Haykin, J. Reilly, V. Kezys, and E. Vertatschitsch, "Some aspects of array signal processing," *IEE Proc.*, vol. 139, pp. 1–26, février 1992.
- [22] P. Stoica, B. Ottersten, M. Viberg, and R. L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *Signal Processing*, vol. 44, pp. 96–105, January 1996.
- [23] V. Pisarenko, "The retrieval of harmonics from a covariance function," *J. of the Royal Astronomical Society*, vol. 33, pp. 347–360, 1973.
- [24] D. Tufts and R. Kumaresan, "Estimating the angle of arrival of multiple plane waves," *IEEE Trans. on Aerospace and Electronic System*, vol. 19, pp. 134–139, 1983.
- [25] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. AP-34, pp. 281–290, Mar. 1986. JFG.
- [26] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-31, pp. 1235–1248, octobre 1983.
- [27] S. Marcos and M. Benidir, "Source bearing estimation and sensor positioning with the propagator method," in *Proc. of SPIE*, pp. 312–323, 1990.
- [28] A. Marsal and S. Marcos, "Rapprochement de la méthode de propagateur et music pour la localisation de sources," in *GRETSI*, pp. 681–684, 1991.

- [29] R. Roy and T. Kailath, "ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Transactions on Acoustics Speech and Signal Processing*, vol. ASSP-37, pp. 984–995, July 1989.
- [30] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal - part 1: Fundamentals," *Proceedings of the IEEE*, vol. 80, pp. 519–538, April 1992.
- [31] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal - part 2: Algorithms and applications," *Proceedings of the IEEE*, vol. 80, pp. 539–568, April 1992.
- [32] K. Sharman, "Maximum likelihood estimation by simulated annealing," in *Proc. of IEEE ICASSP*, pp. 2241–2244, 1988.
- [33] K. Sharman and G. McClurkin, "Genetic algorithms for maximum likelihood parameter estimation," in *Proc. of IEEE ICASSP*, pp. 2716–2719, 1989.