

Séparation de signaux bruités à spectre de raies à l'aide de statistiques d'ordre deux

Pierre FABRY, Christine SERVIÈRE, Jean-Louis LACOUME

LIS, UPRESA CNRS 5083

ENSIEG, BP 46, 38402 Saint-Martin d'Hères Cedex, France

pierre.fabry@lis.inpg.fr, christine.serviere@lis.inpg.fr, jean-louis.lacoume@lis.inpg.fr

Résumé – Ce papier traite de la séparation de mélanges convolutifs bruités. Les bruits sont temporellement et spatialement corrélés. Leurs supports de corrélation sont supposés finis. Les traitements sont réalisés à chaque canal de fréquence à l'aide de statistiques d'ordre deux. La séparation est obtenue par une étape de blanchiment suivie d'une transformation orthogonale. La technique employée pour estimer la matrice de blanchiment est spécifique aux sources à spectre de raies tandis que la transformation orthogonale est obtenue dans un cadre plus général. Ces deux transformations sont estimées à partir de matrices interspectrales des observations décalées pour supprimer l'influence des bruits. Après séparation les signaux sont débruités par un filtre de Wiener multi-références. Ce filtre utilise des références bruit seul obtenues par projection des observations sur le sous-espace bruit. De bons résultats sont obtenus par simulation dans un canal de mélange où le RSB est de 0 dB.

Abstract – This paper presents a blind source separation technique for noisy convolutive mixtures. The noises may be spatially and temporally correlated ; their correlation lengths are finite. The operations are processed at each frequency bin, using only second-order statistics. The separation is achieved from a whitening of the mixture and an orthogonal transformation. In this paper, the whitening stage is specific to line spectrum signals whereas the orthogonal transformation stage is more general. Both stages are based on interspectral matrices computed from delayed observation vectors to cancel the noise influence. The decorrelation step is slightly adapted to obtain also noise outputs. All the outputs are further used in a Wiener multi-references noise reduction step. Good results are obtained in a frequency bin with SNR equal to 0 dB.

1 Introduction

Nous considérons un système linéaire et stationnaire à p entrées et n ($n \geq p$) sorties. Les sources présentes en entrée sont des sommes de sinusoides. Toutes les fréquences pures étant distinctes, les sources sont orthogonales. La difficulté du problème de séparation provient de la présence de fréquences pures proches non résolues. Les capteurs observent des mélanges convolutifs additivement bruités. Les bruits sont décorrélés des sources mais corrélés entre eux. Tous les signaux sont centrés. Les traitements sont réalisés dans le domaine fréquentiel après Transformée de Fourier Discrète (TFD) sur des blocs de N échantillons. L'analyse à durée limitée provoque un étalement spectral et les composantes proches se mélangent. On dispose alors d'un mélange fréquentiel instantané par canal de fréquence. Notons c le nombre de composantes mélangées au canal discret f et supposons cette quantité connue. La TFD des observations temporelles sur la fenêtre $[t, \dots, t + N - 1]$ constitue le vecteur $\mathbf{r}^t(f)$ modélisée par :

$$\mathbf{r}^t(f) = \underbrace{\mathbf{H}(f) \mathbf{s}^t(f)}_{\mathbf{y}^t(f)} + \mathbf{b}^t(f)$$

- $\mathbf{H}(f)$ est la matrice ($n \times c$) de mélange affectée des densités spectrales des sources,
- $\mathbf{s}^t(f)$ est le vecteur ($c \times 1$) des sources normalisées,
- $\mathbf{b}^t(f)$ est le vecteur ($n \times 1$) des bruits.

Omettons la notation fréquentielle (f). La Décomposition en Valeurs Singulières (DVS) de \mathbf{H} peut s'écrire

$$\mathbf{H} = \mathbf{V}\mathbf{\Lambda}^{1/2}\mathbf{\Pi} \quad \text{avec} \quad \begin{cases} \mathbf{V} : \text{matrice unitaire } n \times c \\ \mathbf{\Pi} : \text{matrice unitaire } c \times c \\ \mathbf{\Lambda}^{1/2} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_c}) \end{cases}$$

Les valeurs singulières sont rangées par ordre décroissant. La séparation de sources consiste à déterminer la pseudo-inverse de \mathbf{H} notée $\mathbf{H}^\#$. L'inversion s'obtient en deux étapes. La première étape, décrite en (1), consiste en un blanchiment de la partie utile des signaux.

$$\mathbf{x}^t = \mathbf{W}\mathbf{r}^t \quad \text{avec} \quad \mathbf{W} = \mathbf{\Lambda}^{-1/2}\mathbf{V}^+ \quad (1)$$

$^+$ représente le transposé-conjugué. La seconde étape permet de déterminer $\mathbf{\Pi}$ à partir de \mathbf{x}^t .

En théorie, les éléments permettant de construire la matrice de blanchiment \mathbf{W} (de taille $c \times n$) sont issus de la Décomposition en Valeurs Propres (DVP) de la matrice interspectrale des mélanges non bruités :

$$\mathbf{R}_{\mathbf{y}\mathbf{y}} = E \{ \mathbf{y}^t \mathbf{y}^{t+} \} = \mathbf{H}\mathbf{H}^+ = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^+ \quad (2)$$

Tout le problème consiste donc à estimer $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ (de taille $n \times n$) à partir des observations bruitées. L'Analyse en Composantes Principales est proscrite puisque les bruits ne sont pas spatialement blancs. Lorsque les bruits sont spatialement blancs l'ACP est tout de même déconseillée en raison de l'instabilité des valeurs et vecteurs propres à une perturbation (erreur d'estimation de $\mathbf{R}_{\mathbf{r}\mathbf{r}}$) [4]. Le support de corrélation des fréquences pures étant infini et

ceux des bruits étant supposés finis, nous montrons que \mathbf{W} peut être estimée à l'aide des TFD des observations temporellement décalées, \mathbf{r}^t et $\mathbf{r}^{t+\tau}$. Ces observations permettent de construire des matrices interspectrales dissymétriques (i.e. non hermitiennes) qui ne subissent pas l'influence des bruits. Nous montrons (section 2) que deux matrices interspectrales dissymétriques particulières suffisent pour déterminer $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ et donc \mathbf{W} . La matrice \mathbf{W} réalise la projection des observations sur une base du sous-espace signal. En réalisant aussi une projection sur les $(n-c)$ directions complémentaires nous obtenons des références bruit seul en vue de l'étape de réduction de bruit.

A la section suivante nous nous intéressons à la détermination de Π . Les techniques classiques permettent d'estimer Π en absence de bruit à l'aide de statistiques d'ordre deux de \mathbf{x}^t [1], et en présence de bruits gaussiens à l'aide de statistiques d'ordre supérieur à deux (SOS) de \mathbf{x}^t [3]. Malheureusement, lorsque les bruits ne sont pas gaussiens ou lorsque les niveaux de bruits sont élevés, ces techniques génèrent des erreurs d'estimation importantes. Nous montrons, dans la section 3, que lorsque les supports de corrélation des sources sont supérieurs à ceux des bruits, il est possible d'estimer Π à l'aide de statistiques d'ordre deux.

Après cette étape chaque sortie contient une source et du bruit et nous disposons de $(n-c)$ références bruit. La dernière étape, décrite en section 4, consiste à débruiter chacune des c premières voies à l'aide des $(n-1)$ autres.

Enfin, nous présentons des résultats de simulation dans un contexte fortement bruité.

2 Décorrélacion et construction de références bruit seul

Définissons la matrice interspectrale dissymétrique des observations par :

$$\mathbf{R}_{\mathbf{r}\mathbf{r}}^\tau = E \{ \mathbf{r}^t (\mathbf{r}^{t+\tau})^+ \} = \mathbf{R}_{\mathbf{y}\mathbf{y}}^\tau + \mathbf{R}_{\mathbf{b}\mathbf{b}}^\tau$$

Dans cette écriture \mathbf{r}^t et $\mathbf{r}^{t+\tau}$ sont les TFD des observations temporelles sur les fenêtres débutant respectivement aux indices t et $t+\tau$. Notons τ_B le plus grand support de corrélation des bruits. Si l'on choisit un retard τ tel que $N + \tau_B < \tau$ alors \mathbf{r}^t et $\mathbf{r}^{t+\tau}$ sont issus d'échantillons de bruit décorrélés et d'échantillons de signal corrélés. Par conséquent $\mathbf{R}_{\mathbf{b}\mathbf{b}}^\tau = \mathbf{0}$ et

$$\mathbf{R}_{\mathbf{r}\mathbf{r}}^\tau = \mathbf{R}_{\mathbf{y}\mathbf{y}}^\tau = \mathbf{H}\mathbf{R}_{\mathbf{s}\mathbf{s}}^\tau\mathbf{H}^+ \quad (3)$$

Notons f_i la fréquence de la i ème source vue dans le canal d'analyse. Etant donné que les sources sont normalisées et orthogonales,

$$\mathbf{R}_{\mathbf{s}\mathbf{s}}^\tau = \begin{pmatrix} e^{-j2\pi f_1\tau} & & 0 \\ & \ddots & \\ 0 & & e^{-j2\pi f_c\tau} \end{pmatrix} = \theta$$

En introduisant la DVS de \mathbf{H} dans (3) on obtient

$$\mathbf{R}_{\mathbf{r}\mathbf{r}}^\tau = \mathbf{V}\Lambda^{1/2}\Pi\theta\Pi^+\Lambda^{1/2}+\mathbf{V}^+$$

Une seconde matrice interspectrale dissymétrique, au retard -2τ , permet de reconstruire la matrice interspectrale des mélanges non bruités selon

$$\mathbf{R}_{\mathbf{y}\mathbf{y}} = \mathbf{R}_{\mathbf{r}\mathbf{r}}^\tau (\mathbf{R}_{\mathbf{r}\mathbf{r}}^{\tau+})^{-1} \mathbf{R}_{\mathbf{r}\mathbf{r}}^{-2\tau}$$

L'inversion de $\mathbf{R}_{\mathbf{r}\mathbf{r}}^{\tau+}$ est réalisée par le calcul de la pseudo-inverse en supposant le rang c connu. Lorsque le nombre de composantes c est inconnu, il peut être estimé de façon robuste [2].

La matrice $\mathbf{R}_{\mathbf{y}\mathbf{y}}$ étant de taille $(n \times n)$ sa DVP se met sous la forme :

$$\mathbf{R}_{\mathbf{y}\mathbf{y}} = [\mathbf{V}, \mathbf{V}_\perp] \Delta \begin{bmatrix} \mathbf{V}^+ \\ \mathbf{V}_\perp^+ \end{bmatrix} \quad \text{avec} \quad \Delta = \begin{bmatrix} \Lambda & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

\mathbf{V} de taille $n \times c$ est une base orthonormée du sous-espace signal. Elle représente les directions principales d'énergie des signaux utiles. \mathbf{V}_\perp de taille $n \times (n-c)$ est une base du sous-espace orthogonal complémentaire. Elle représente $(n-c)$ directions de bruit. En vue de l'étape de réduction de bruit, il est opportun de projeter les observations sur ces vecteurs de façon à constituer $(n-c)$ références bruit seul. Le filtrage matriciel (1) est donc remplacé par :

$$\begin{bmatrix} \mathbf{x}^t \\ \mathbf{x}_\perp^t \end{bmatrix} = \mathbf{W}'\mathbf{r}^t \quad \text{avec} \quad \mathbf{W}' = \begin{bmatrix} \mathbf{W} \\ \mathbf{G}^{-1/2}\mathbf{V}_\perp^+ \end{bmatrix}$$

La matrice diagonale $\mathbf{G}^{-1/2}$ fixe le niveau d'énergie des références bruit seul. Arbitrairement nous choisissons :

$$\mathbf{G}^{-1/2} = \alpha\mathbf{I}_{n-c} \quad \text{avec} \quad \alpha = \left(\frac{1}{c} \sum_{i=1}^c \lambda_i \right)^{-1/2}$$

où \mathbf{I}_{n-c} est la matrice identité de rang $(n-c)$. On obtient

$$\begin{aligned} \mathbf{x}^t &= \Pi\mathbf{s}^t + \mathbf{W}\mathbf{b}^t \\ \mathbf{x}_\perp^t &= \mathbf{G}^{-1/2}\underbrace{\mathbf{V}_\perp^+\mathbf{V}\Lambda^{1/2}}_{\mathbf{0}}\Pi\mathbf{s}^t + \mathbf{G}^{-1/2}\mathbf{V}_\perp^+\mathbf{b}^t = \alpha\mathbf{V}_\perp^+\mathbf{b}^t \end{aligned}$$

3 Transformation orthogonale

Les vecteurs \mathbf{x}^t et $\mathbf{x}^{t+\tau}$ obtenus par blanchiment à l'ordre deux des vecteurs \mathbf{r}^t et $\mathbf{r}^{t+\tau}$ permettent de construire une matrice interspectrale dissymétrique qui ne subit pas l'influence des bruits :

$$\mathbf{R}_{\mathbf{x}\mathbf{x}}^\tau = \Pi\theta\Pi^+$$

On constate que Π est la matrice des vecteurs propres de $\mathbf{R}_{\mathbf{x}\mathbf{x}}^\tau$. La séparation est alors achevée par la projection :

$$\mathbf{a}^t = \Pi^+\mathbf{x}^t = \mathbf{s}^t + \mathbf{v}^t \quad \text{avec} \quad \begin{cases} \mathbf{v}^t = \mathbf{H}\#\mathbf{b}^t \\ \mathbf{H}\# = \Pi^+\Lambda^{-1/2}\mathbf{V}^+ \end{cases} \quad (4)$$

Notons que la DVP de $\mathbf{R}_{\mathbf{x}\mathbf{x}}^\tau$ donne Π à une matrice complexe de permutation près $\hat{\Pi} = \Pi\mathbf{P}$. Cette remarque servira pour l'interprétation des résultats de simulation.

4 Réduction de bruit

Les composantes du vecteur \mathbf{v}^t sont corrélées entre elles et corrélées avec les composantes de \mathbf{x}_\perp^t . De fait, le RSB relatif à chaque composante du vecteur \mathbf{a}^t peut être maximisé par un filtre de Wiener dont les références bruit sont les $(c-1)$ composantes restantes de \mathbf{a}^t et les $(n-c)$ composantes de \mathbf{x}_\perp^t .

Pour simplifier, regroupons tous ces signaux dans un même vecteur \mathbf{z}^t :

$$\mathbf{z}^t = \left[(\mathbf{a}^t)^T, (\mathbf{x}_\perp^t)^T \right]^T$$

T désigne la transposition. Les sources estimées après réduction de bruit s'écrivent :

$$\hat{s}_i^t = z_i^t - \hat{v}_i^t, \quad i = 1 \dots c$$

où \hat{v}_i^t est l'estimée de la i ème composante de bruit. L'estimation est réalisée par combinaison linéaire des $(n-1)$ composantes restantes de \mathbf{z}^t à l'aide de coefficients complexes F_k :

$$\hat{v}_i^t = \sum_{k \neq i} F_k z_k^t = (\mathbf{z}_{-i}^t)^T \mathbf{F}$$

avec

$$\begin{aligned} \mathbf{z}_{-i}^t &= [z_1^t, \dots, z_{i-1}^t, z_{i+1}^t, \dots, z_n^t]^T \\ \mathbf{F} &= [F_1, \dots, F_{i-1}, F_{i+1}, \dots, F_n]^T \end{aligned}$$

La solution qui minimise l'Erreur Quadratique Moyenne $E \left[|v_i^t - \hat{v}_i^t|^2 \right]$ est donnée par :

$$\mathbf{F} = \left(\mathbf{R}_{\mathbf{z}_{-i}\mathbf{z}_{-i}}^* \right)^{-1} \mathbf{R}_{\mathbf{z}_i\mathbf{z}_{-i}}$$

avec

$$\mathbf{R}_{\mathbf{z}_{-i}\mathbf{z}_{-i}} = E \left[\mathbf{z}_{-i}^t \cdot (\mathbf{z}_{-i}^t)^+ \right] \text{ et } \mathbf{R}_{\mathbf{z}_i\mathbf{z}_{-i}} = E \left[z_i^t \cdot (\mathbf{z}_{-i}^t)^* \right]$$

et le gain en terme de RSB est donné par :

$$G_i = \frac{RSB_{sortie}}{RSB_{entrée}} = \frac{1}{1 - \frac{\mathbf{R}_{\mathbf{z}_i\mathbf{z}_{-i}}^+ \cdot \left(\mathbf{R}_{\mathbf{z}_{-i}\mathbf{z}_{-i}}^* \right)^{-1} \cdot \mathbf{R}_{\mathbf{z}_i\mathbf{z}_{-i}}}{R_{v_i v_i}}}$$

5 Résultats de simulation

Trois fréquences pures proches filtrées par des filtres AR1 sont reçues sur cinq capteurs. Les observations sont perturbées par cinq bruits obtenus par mélange instantané d'un bruit blanc gaussien et d'un bruit blanc uniforme. Les TFD sont calculées sur des blocs de 64 échantillons fournissant une résolution de $\frac{1}{64} \approx 0.016$ en fréquence réduite. Les fréquences réduites des sources étant $f_1 = 0.10$, $f_2 = 0.11$ et $f_3 = 0.12$, une analyse spectrale à cette résolution ne permet pas de les dissocier. Les traitements sont réalisés au canal $f = 0.11$ pour un RSB de 0dB. Les matrices interspectrales sont estimées sur 900 blocs. Le retard τ est de 67 échantillons. Le nombre de sources est connu.

La qualité d'estimation de la matrice de décorrélation est mesurée par la matrice \mathbf{E} des erreurs relatives sur les composantes de $\widehat{\mathbf{W}}$:

$$\mathbf{E} = (E_{ij}) \quad \text{avec} \quad E_{ij} = \left| \frac{\widehat{W}_{ij} - W_{ij}}{W_{ij}} \right|$$

On obtient

$$\mathbf{E} = \begin{pmatrix} 0.027 & 0.017 & 0.080 & 0.015 & 0.412 \\ 0.021 & 0.064 & 0.049 & 0.022 & 0.035 \\ 0.056 & 0.024 & 0.063 & 0.097 & 0.039 \end{pmatrix}$$

Ces erreurs relatives sont faibles pour un tel contexte de bruit.

La qualité d'estimation de Π est mesurée par la matrice $\mathbf{Q} = \Pi^+ \widehat{\Pi}$. Lorsque l'estimation est parfaite \mathbf{Q} est une matrice de permutation (voir section 3). Ici on obtient

$$(|Q_{ij}|) = \begin{pmatrix} 0.015 & 1.002 & 0.028 \\ 0.028 & 0.013 & 1.067 \\ 0.991 & 0.014 & 0.020 \end{pmatrix}$$

La 1ère source est sur la 3ème sortie (0.991), la 2ème source est sur la 1ère sortie (1.002) et la 3ème source est sur la 2ème sortie (1.067).

Les figures 1, 2 et 3 donnent les enveloppes des sources initiales à retrouver. La figure 4 donne l'enveloppe de l'observation bruitée de l'un des capteurs. Après séparation (étapes 2 et 3) les sorties \mathbf{a}^t contiennent chacune une source estimée et du bruit. L'enveloppe de l'une de ces sorties est donnée à la figure 5. Les figures 6, 7 et 8 donnent les enveloppes des sources estimées débruitées.

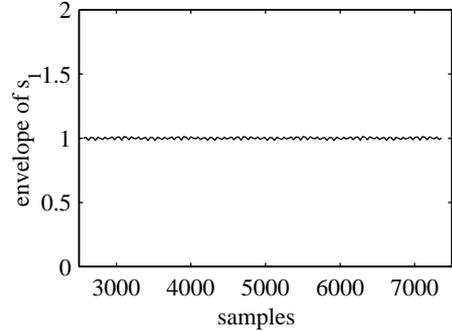


FIG. 1: Enveloppe de la 1ère source

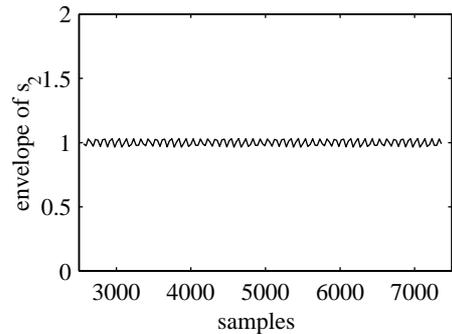


FIG. 2: Enveloppe de la 2ème source

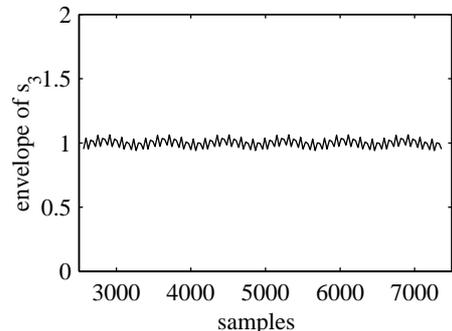


FIG. 3: Enveloppe de la 3ème source

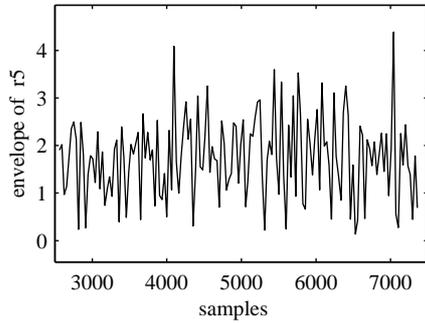


FIG. 4: Enveloppe d'une observation bruitée

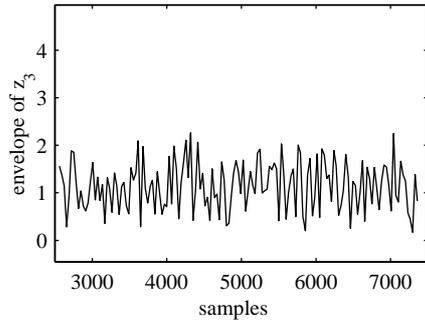


FIG. 5: Enveloppe d'une source estimée bruitée

6 Conclusions

Ce papier traite de la séparation de sources à spectre de raies au travers de mélanges convolutifs bruités. Nous proposons une technique fréquentielle basée sur des statistiques d'ordre deux des observations temporellement décalées. Nous obtenons ainsi des matrices interspectrales « débruitées », à condition que les supports de corrélation des bruits soient inférieurs à ceux des sources. Cette méthode autorise que les bruits soient spatialement et temporellement corrélés. En dernier lieu, une étape de réduction de bruit est mise en oeuvre. La méthode donne des résultats satisfaisants dans un canal de fréquence où le RSB est de 0 dB. L'étape de réduction de bruit est très efficace en raison du nombre et de la qualité des références disponibles.

Références

- [1] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, and E. Moulines. A blind source separation technique using second-order statistics. *IEEE Trans. on Sig. Proc.*, Vol. 45:433–444, 1997.
- [2] P. Fabry, C. Servière, and J.-L. Lacoume. Improving signal subspace estimation and source number detection in the context of spatially correlated noises. In *Proc. EUSIPCO'98, Rhodes*, pages 1809–1812, 1998.
- [3] M. Gaéta and J.-L. Lacoume. Source separation without a priori knowledge : The maximum likelihood solution. In *Proc. EUSIPCO'90, Barcelona*, pages 621–624, 1990.
- [4] G.H. Golub and F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 1983.

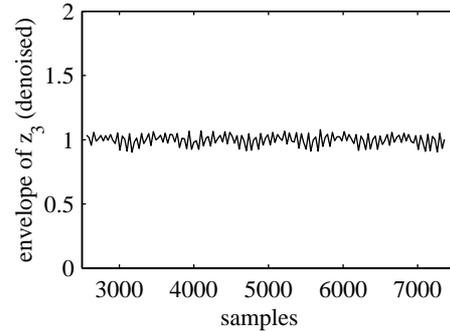


FIG. 6: Enveloppe de la 1ère source estimée

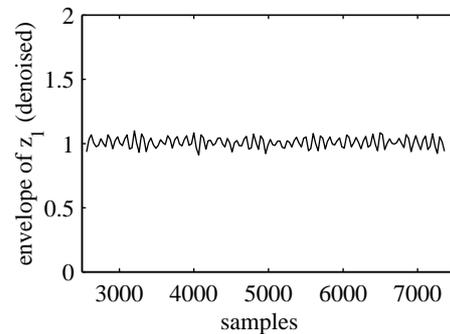


FIG. 7: Enveloppe de la 2ème source estimée

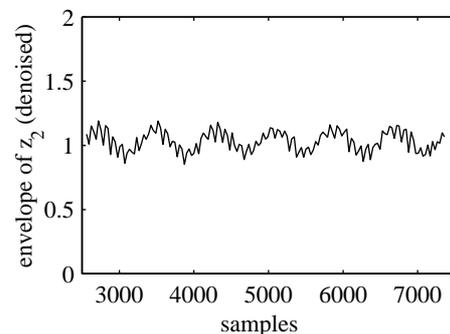


FIG. 8: Enveloppe de la 3ème source estimée