

Une analyse de l'orthogonal least squares par la propriété d'isométrie restreinte en présence de signaux décroissants

Jonathan BOSSE

DEMR, ONERA

Université Paris-Saclay F-91123 Palaiseau - France

jonathan.bosse@onera.fr

Résumé – L'orthogonal least squares (OLS) est un algorithme glouton très répandu pour l'approximation ou la reconstruction de signaux parcimonieux. Nous abordons ici le problème des conditions suffisantes de reconstruction exacte du support d'un signal via OLS en absence de bruit. Sur la base de l'inégalité de Wielandt généralisée, de nouvelles bornes sur la règle de sélection de l'OLS sont fournies qui permettent d'établir une condition suffisante sur la constante d'isométrie (RIC en anglais) du dictionnaire par une approche géométrique simple. Étant donné que ces nouvelles bornes prennent naturellement en compte la distribution d'amplitude du signal, elles permettent de dériver une nouvelle condition qui s'avère particulièrement pertinente lorsque l'amplitude du signal décroît.

Abstract – The Orthogonal Least Squares (OLS) is a widespread greedy algorithm for sparse signal approximation or reconstruction. We here address the problem of sufficient condition for exact signal support recovery via OLS. Based on the generalized Wielandt inequality, new bounds on the OLS selection rule are provided that enable a simple geometric approach of the Restricted Isometry Property based analysis of OLS. Since these new bounds naturally take the signal amplitude distribution into account, they allow a derivation of a new relaxed exact support recovery condition in the noise-free case.

1 Introduction

La reconstruction de signaux qui admettent une représentation parcimonieuse a toujours suscité beaucoup d'intérêt, notamment depuis l'émergence du compressed sensing (CS) [1, 2]. Il s'agit, partant d'un signal mesuré $\mathbf{y} \in \mathbb{C}^m$, de retrouver le signal K -parcimonieux (qui contient au plus K composantes non nulles) $\mathbf{x} \in \mathbb{C}^N$ (en général $m < N$) tel que

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \quad (1)$$

où $\mathbf{A} \in \mathbb{C}^{m \times N}$ est le dictionnaire et ses colonnes sont nommées atomes. Une méthode très répandue pour résoudre ce problème est l'orthogonal least squares (OLS) [3].

Les conditions de reconstruction de l'algorithme OLS sont étudiées dans [4] mais ne sont pas toujours commode à vérifier en pratique. Les études de performances de ce type de méthodes sont ainsi généralement conduites à l'aide de deux outils privilégiés : la cohérence [4] ainsi que la constante d'isométrie restreinte (RIC en anglais)[2]. Cette dernière permet une analyse plus fine des performances, la cohérence produisant généralement des bornes assez strictes et se destinant plutôt à des degrés de parcimonie peu élevés [5].

Une matrice satisfait la propriété d'isométrie restreinte (RIP en anglais) d'ordre K si, pour chaque vecteur K -parcimonieux \mathbf{x} , nous avons

$$(1 - \delta_K) \|\mathbf{x}\|_2^2 \leq \|\mathbf{A}\mathbf{x}\|_2^2 \leq (1 + \delta_K) \|\mathbf{x}\|_2^2.$$

La plus petite constante $\delta_K \in [0, 1[$ est la constante d'isométrie restreinte (RIC en anglais).

La première étude d'un algorithme glouton à l'aide de la RIP est due à [6]. Depuis les résultats ont été sans cesse améliorés et les conditions suffisantes sont devenues de plus en plus pertinentes. Récemment, l'étude [7] donne une condition suffisante optimale sur la RIC d'ordre $K + 1$ pour reconstruire tout signal K -parcimonieux. Cette condition est la suivante :

$$\delta_{K+1} < C_K = \begin{cases} \frac{1}{\sqrt{K+\frac{1}{4}}} & K = 2 \\ \frac{1}{\sqrt{K+\frac{1}{16}}} & K = 3 \\ \frac{1}{\sqrt{K}} & K = 1, K \geq 4 \end{cases}. \quad (2)$$

La condition est optimale en ce sens que l'on peut trouver des dictionnaires et des signaux qui ne permettent pas la reconstruction si elle n'est pas vérifiée. Cette condition demeure assez pessimiste puisqu'elle considère un fonctionnement au pire cas et est souvent éloignée du comportement moyen observé en pratique de l'OLS.

Il est couramment admis que les signaux d'amplitude égales constituent un pire cas [8]. En pratique, il est pourtant assez courant (par exemple en radar) d'observer des signaux qui ont peu de chances d'être d'amplitude égales. Exploiter la décroissance potentielle des signaux est un moyen d'obtenir des conditions de reconstruction moins sévères, plus à même de rendre compte du comportement observé de l'OLS. Il semble que des garanties de reconstruction de l'OLS relaxées exploitant la décroissance des signaux n'aient pas été proposées par la littérature.

Dans cet article, partant d'une interprétation géométrique de la règle de sélection de l'OLS, nous fournissons de nouvelles

bornes sur cette quantité permettant une analyse par le biais de la RIP de l'OLS. Ces bornes sont basées sur l'inégalité de Wielandt généralisée [9, 10] et permettent une analyse géométrique assez simple de l'OLS. La vertu de la condition suffisante énoncée est de prendre naturellement en compte la répartition d'amplitude du signal, contrairement à la littérature [7]. La condition sur δ_{K+1} dépend alors aussi de δ_K et de la distribution de l'amplitude du signal. Nous montrons que, dès que les coefficients du signal obéissent à une décroissance suffisante, la condition suffisante proposée est moins sévère que l'état de l'art.

L'article est organisé de la manière suivante. Dans le paragraphe 2, nous présentons la base de l'algorithme OLS ainsi que de nouvelles bornes sur la règle de sélection. Puis, dans la section 3, nous en déduisons une nouvelle condition suffisante de reconstruction pour l'OLS et illustrons son apport sur un exemple de la littérature.

Dans l'article, $\langle \cdot, \cdot \rangle$ fait référence au produit scalaire classique entre vecteurs complexes. \mathbf{A}_Λ et \mathbf{x}_Λ désignent la matrice restreinte aux colonnes appartenant à l'ensemble Λ et le vecteur restreint à Λ . \mathbf{P}_Λ représente la projection orthogonale sur \mathbf{A}_Λ et $\mathbf{P}_\Lambda^\perp = \mathbf{I} - \mathbf{P}_\Lambda$. $S \subset \Omega = \{1, 2, \dots, N\}$ représente le vrai support du signal à récupérer \mathbf{x} . $|S|$ est ici le cardinal de l'ensemble S . Les coefficients non nuls de \mathbf{x} sont ordonnés ici de manière décroissante : $|x_{S(1)}| \geq |x_{S(2)}| \geq \dots \geq |x_{S(K)}|$ où $S(k)$ désigne le k -ième élément de S . Nous supposons aussi, sans perte de généralité (puisque le signal \mathbf{x} peut toujours être normalisé en conséquence), que les colonnes de \mathbf{A} notées \mathbf{a}_i sont de norme unité.

2 Algorithme OLS et nouvelles bornes sur la règle de sélection

L'algorithme OLS est résumé dans l'Algorithme 1. Cet algorithme itératif vise à trouver une approximation parcimonieuse dans un certain dictionnaire \mathbf{A} d'un signal discret donné \mathbf{y} . Pour cela, il sélectionne les atomes selon une règle de sélection puis se débarrasse de leur contribution grâce à une projection orthogonale. Si la règle d'arrêt n'est pas satisfaite, alors l'algorithme continue. Ce principe de base est commun aux autres approches gloutonnes.

La règle d'arrêt dépend généralement du contexte (par exemple, sans bruit, bruit gaussien, bruit borné) et peut par exemple être formulée en termes de nombre d'itérations maximales.

À la k -ième itération, étant donné un support estimé S_{k-1} , l'OLS sélectionne un atome \mathbf{a}_{i_k} qui minimise la norme du résidu $\mathbf{P}_{S_{k-1} \cup i}^\perp \mathbf{y}$,

$$i_k = \arg \min_{i \notin S_{k-1}} \left\| \mathbf{P}_{S_{k-1} \cup i}^\perp \mathbf{y} \right\|_2,$$

qui peut se réécrire

$$i_k = \arg \max_{i \notin S_{k-1}} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\left\| \mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i \right\|_2}. \quad (3)$$

Comme il a été observé précédemment, la règle de sélection de l'OLS (3) peut être interprétée comme le cosinus d'un angle

[11, 4]. En effet, $\|\mathbf{r}_{k-1}\|_2$ ne dépend pas de i , on peut toujours diviser (3) par cette quantité. Puisque $\mathbf{P}_{S_k}^\perp$ est un projecteur orthogonal, il est idempotent et hermitien, de sorte que $|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle| = \left| \left\langle \mathbf{a}_i, \mathbf{P}_{S_{k-1}}^\perp \mathbf{r}_{k-1} \right\rangle \right| = \left| \left\langle \mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i, \mathbf{r}_{k-1} \right\rangle \right|$. On a donc

$$i_k = \arg \max_{i \notin S_{k-1}} \cos \Theta \left(\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i, \mathbf{r}_{k-1} \right), \quad (4)$$

où pour toute paire de vecteurs (\mathbf{x}, \mathbf{y}) nous définissons comme dans [10]

$$\cos \Theta(\mathbf{x}, \mathbf{y}) = \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|}{\|\mathbf{x}\|_2 \|\mathbf{y}\|_2}.$$

Par conséquent, la règle de sélection de l'OLS revient à comparer le cosinus de l'angle entre chaque atome projeté $\mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i$ et le résidu \mathbf{r}_{k-1} et à sélectionner le plus petit angle. Cette interprétation géométrique de la règle de sélection est à la base du présent article, où l'inégalité de Wielandt généralisée entre des paires de vecteurs complexes [9, 10] a permis de dériver de nouvelles bornes pertinentes.

Data: \mathbf{y}, \mathbf{A}

Result: $\hat{S} = S_k, \hat{\mathbf{x}}_{\hat{S}} = \mathbf{x}_k$

$\mathbf{r}_0 \leftarrow \mathbf{y}, \mathbf{x}_0 \leftarrow \mathbf{0}, S_0 \leftarrow \emptyset, k \leftarrow 0;$

while *Stopping rule is not met* **do**

$k \leftarrow k + 1$

$i_k \leftarrow \arg \max_{i \notin S_{k-1}} \frac{|\langle \mathbf{a}_i, \mathbf{r}_{k-1} \rangle|}{\left\| \mathbf{P}_{S_{k-1}}^\perp \mathbf{a}_i \right\|_2}$

$S_k \leftarrow i_k \cup S_{k-1}$

$\mathbf{P}_{S_k} \leftarrow \mathbf{A}_{S_k} (\mathbf{A}_{S_k}^H \mathbf{A}_{S_k})^{-1} \mathbf{A}_{S_k}^H$

$\mathbf{r}_k \leftarrow \mathbf{y} - \mathbf{P}_{S_k} \mathbf{y} = \mathbf{P}_{S_k}^\perp \mathbf{y}$

$\mathbf{x}_k \leftarrow (\mathbf{A}_{S_k}^H \mathbf{A}_{S_k})^{-1} \mathbf{A}_{S_k}^H \mathbf{y}$

end

Algorithm 1: Algorithme OLS

Lemme 1 *Pour tout ensemble $S_k \subset S$, nous avons les bornes supérieures et inférieures suivantes de la règle de sélection en absence de bruit :*

$$\max_{i \notin S} \cos \Theta \left(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k \right) \leq \delta_{K+1} \quad (5)$$

$$\max_{i \in S \setminus S_k} \cos \Theta \left(\mathbf{P}_{S_k}^\perp \mathbf{a}_i, \mathbf{r}_k \right) \geq \frac{\cos \phi_k - \delta_K}{1 - \delta_K \cos \phi_k}, \quad (6)$$

où

$$\cos \phi_k = \max_{i \in S \setminus S_k} \frac{|x_i|}{\|\mathbf{x}_{S \setminus S_k}\|_2}. \quad (7)$$

La borne supérieure a déjà été proposée pour l'OMP [12], mais elle est plus pertinente pour l'OLS. La borne inférieure est un nouveau résultat.

Ces bornes sont obtenues de façon similaire à celle de l'OMP [12] mais cette fois ci à l'aide de l'inégalité de Wielandt généralisée [9, 10]. Cette dernière propose une borne du cosinus de l'angle entre deux vecteurs lorsqu'une même matrice inversible est appliquée à ces vecteurs. La borne obtenue dépend du cosinus

de l'angle entre ces deux vecteurs et des valeurs propres minimales et maximales de la matrice inversible. Le lien avec la constante RIC est effectuée par ce biais. Pour plus de détails techniques, on pourra consulter [13].

3 Une nouvelle condition suffisante de reconstruction

L'une des conséquences des bornes proposées est la garantie suivante de reconstruction de l'OLS. En effet, pour que OLS reconstruise le signal à chaque itération il suffit que (5) soit inférieur à (6). La distribution des amplitudes du signal se fait naturellement par le terme $\cos \phi_k$ in (6).

Theorème 1 *Si un signal K -parcimonieux \mathbf{x} de support S suit le modèle (1), alors OLS avec règle d'arrêt $\|\mathbf{r}_k\|_2 > 0$, reconstruit exactement \mathbf{x} en K itérations pourvu que*

$$\delta_{K+1} < \frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}, \quad (8)$$

où

$$\begin{aligned} \cos \Phi &= \min_{\Lambda} \left\{ \frac{\max_{i \in S \setminus \Lambda} |x_i|}{\|\mathbf{x}_{S \setminus \Lambda}\|_2}, \Lambda \subset S \right\}, \\ &= \min_{1 \leq k \leq K} \frac{|x_{S(k)}|}{\sqrt{\sum_{i=k}^K |x_{S(i)}|^2}}. \end{aligned} \quad (9)$$

La comparaison de la condition (8) qui implique à la fois δ_K et $\cos \Phi$ avec l'état de l'art (2) n'est pas directe, car ce dernier ne dépend que de δ_{K+1} et K . Il est possible d'obtenir néanmoins les bornes suivantes sur la valeur limite δ_{K+1}^{lim} qui vérifie (8) qui ne dépendent pas de δ_K (puisque $0 \leq \delta_K \leq \delta_{K+1}$, la borne supérieure correspond au cas $\delta_K = 0$, la borne inférieure au cas $\delta_{K+1} = \delta_K$) :

Lemme 2 *La condition (8) s'écrit*

$$\delta_{K+1} < \frac{1 - \sin \Phi}{\cos \Phi}, \quad (10)$$

si $\delta_K = \delta_{K+1}$ ("pire" cas) et

$$\delta_{K+1} < \cos \Phi, \quad (11)$$

si $\delta_K = 0$ (cas le plus favorable).

Par définition, nous avons

$$\frac{1}{\sqrt{K}} \leq \cos \Phi \leq 1, \quad (12)$$

où la borne inférieure correspond au cas où les amplitudes du signal sont égales et la borne supérieure au cas de signaux fortement décroissants : $\sqrt{\sum_{i=k}^K |x_{S(i)}|^2} \approx |x_{S(k)}|$, si bien que $\cos \Phi \approx 1$.

Notons que dans (8) il faut $\cos \Phi \geq \delta_K$ car δ_{K+1} est positif. En examinant (10), puisque $\delta_K \leq \delta_{K+1}$, cette condition est vérifiée.

3.1 Signaux d'amplitude constante

La borne inférieure dans (12) est atteinte pour des signaux d'amplitude constante. Puisque la quantité $\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}$ diminue lorsque $\cos \Phi$ diminue, il s'avère que ce cas représente un pire cas dans notre analyse : il conduit à une contrainte plus forte imposée à δ_{K+1} en (8). Dans ce cas, en utilisant le lemme 2 et (12), on voit que pour les signaux d'amplitude constante, les conditions pour δ_{K+1} dans (8) s'écrivent $\delta_{K+1} < \sqrt{K} - \sqrt{K-1}$ dans le cas le plus défavorable et $\delta_{K+1} < \frac{1}{\sqrt{K}}$ dans le cas le plus favorable. Cela montre, comme prévu, que la borne proposée (10) ne peut pas améliorer la condition (2) dans ce contexte. Nous parlerons de signaux décroissants lorsque $\cos \Phi > \frac{1}{\sqrt{K}}$, c'est à dire que les amplitudes du signal ne sont pas toutes égales.

3.2 Signaux d'amplitude décroissante

La quantité $\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi}$ augmente lorsque $\cos \Phi$ augmente, de sorte que la condition imposée à δ_{K+1} dans (8) devient de plus en plus lâche. Notons que lorsque $\cos \Phi$ tend vers un, quelle que soit la valeur de δ_K , la contrainte (8) disparaît pour δ_{K+1} car elle se transforme en $\delta_{K+1} < 1$.

Le cas des valeurs $\cos \Phi$ croissantes correspond à des signaux présentant une décroissance d'amplitude. À un certain point, la condition (10) devient supérieure à C_K , et la condition (8) devient nécessairement moins stricte que (2).

Plus précisément, lorsque

$$\frac{\cos \Phi - \delta_K}{1 - \delta_K \cos \Phi} \geq C_K,$$

c'est à dire quand $\cos \Phi \geq \frac{\delta_K + C_K}{1 + \delta_K C_K}$, la condition (8) est moins stricte que (2).

4 Un exemple de la littérature

Examinons un exemple réel, issu de [7], construit pour illustrer le caractère optimal de la condition (2). Soit $K = 2$, considérons le dictionnaire $\mathbf{A} \in \mathbb{R}^{3 \times 3}$ tel que

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} 1 & \frac{\delta}{2} & \frac{\delta}{2} \\ \frac{\delta}{2} & 1 & -\frac{\delta}{2} \\ \frac{\delta}{2} & -\frac{\delta}{2} & 1 \end{bmatrix}. \quad (13)$$

ainsi que le signal $\mathbf{x} = [0 \ 1 \ x]^T$ avec $x \in \mathbb{R}$, $x \geq 1$. Dans ce cas, $\delta_{K+1} = \delta_3 = \delta$. Puisque $\delta_K = \delta_2$ est égal à la cohérence, nous avons $\delta_2 = \frac{\delta}{2}$. Le résultat (2) implique que la reconstruction par OLS est possible dès lors que

$$\delta_3 < \frac{1}{\sqrt{\frac{1}{4} + 2}}, \quad (14)$$

quelle que soit la valeur du signal. Le cas limite (14) est obtenu pour $x = 1$. La valeur limite $\bar{\delta}_3$ pour laquelle (8) n'est plus respectée dans ce contexte, est telle que $\bar{\delta}_3 = \frac{\cos \Phi - \frac{\delta_3}{2}}{1 - \frac{\delta_3}{2} \cos \Phi}$,

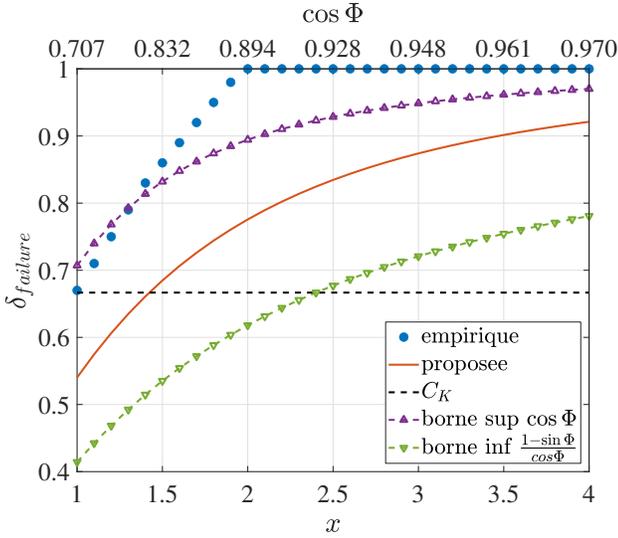


FIGURE 1 – Valeur $\delta_{failure}$ de δ au delà de laquelle OLS échoue à reconstruire le support du signal $\mathbf{x} = [0 \ 1 \ x]^T$ avec le dictionnaire défini en (13).

avec $\cos \Phi = \frac{x}{\sqrt{1+x^2}}$. Après quelques calculs, on trouve

$$\bar{\delta}_3 = \frac{3 - \sqrt{1 + 8 \sin \Phi}}{2 \cos \Phi}.$$

Dans la figure 1, nous comparons la valeur de δ pour laquelle OLS échoue (notée $\delta_{failure}$). Nous voyons que lorsque le signal est d’amplitude constante (i.e. $x = 1$) la condition (14) est moins restrictive que la limite proposée $\bar{\delta}_3$. Lorsque x grandit, nous pouvons observer que, empiriquement, OLS continue toujours à fournir un support exact pour les valeurs de δ supérieures à la limite (14). Lorsque x augmente, alors $\cos \Phi$ augmente, de même que la valeur limite proposée $\bar{\delta}_3$, contrairement à (14). Par conséquent, la borne proposée devient plus informative lorsque $\cos \Phi$ n’est pas trop faible. Ceci illustre le bénéfice de la condition (8) qui prend naturellement en compte l’amplitude du signal : lorsque le signal l’amplitude diminue, la condition sur δ_{K+1} peut être assouplie.

5 Conclusion

Nous avons proposé une nouvelle condition suffisante sur la constante RIP du dictionnaire permettant la reconstruction pour l’algorithme OLS d’un signal K -parcimonieux. Cette condition permet de prendre en compte la distribution d’amplitude des signaux et s’avère pertinente lorsque les amplitudes des signaux décroissent.

Références

[1] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candes and T. Tao, “Decoding by linear programming,” *IEEE transactions on information theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[3] S. Chen, S. A. Billings, and W. Luo, “Orthogonal least squares methods and their application to non-linear system identification,” *International Journal of control*, vol. 50, no. 5, pp. 1873–1896, 1989.

[4] C. Soussen, R. Gribonval, J. Idier, and C. Herzet, “Joint k-step analysis of orthogonal matching pursuit and orthogonal least squares,” *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 3158–3174, 2013.

[5] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[6] M. A. Davenport and M. B. Wakin, “Analysis of orthogonal matching pursuit using the restricted isometry property,” *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4395–4401, 2010.

[7] J. Kim, J. Wang, and B. Shim, “Optimal restricted isometry condition of normalized sampling matrices for exact sparse recovery with orthogonal least squares,” *IEEE Transactions on Signal Processing*, vol. 69, pp. 1521–1536, 2021.

[8] C. Herzet, A. Drémeau, and C. Soussen, “Relaxed recovery conditions for omp/ols by exploiting both coherence and decay,” *IEEE Transactions on Information Theory*, vol. 62, no. 1, pp. 459–470, 2015.

[9] F. L. Bauer and A. S. Householder, “Some inequalities involving the euclidean condition of a matrix,” *Numerische Mathematik*, vol. 2, no. 1, pp. 308–311, 1960.

[10] M. Lin and G. Sinnamon, “The generalized wielandt inequality in inner product spaces,” *Eurasian Mathematical Journal*, vol. 3, no. 1, pp. 72–85, 2012.

[11] T. Blumensath and M. E. Davies, “On the difference between orthogonal matching pursuit and orthogonal least squares,” 2007.

[12] L.-H. Chang and J.-Y. Wu, “An improved rip-based performance guarantee for sparse signal recovery via orthogonal matching pursuit,” *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5702–5715, 2014.

[13] J. Bosse, “A restricted isometry property based analysis of orthogonal least squares under signal decay,” *Soumis à IEEE Transactions on Signal Processing*, 2022.