

# Décontamination des spectres de galaxies en exploitant quatre directions de dispersion \*

Mostafa BELLA<sup>1,2</sup>, Shahram HOSSEINI<sup>1</sup>, Hicham SAYLANI<sup>2</sup>, Thierry CONTINI<sup>1</sup>, Tristan GRÉGOIRE<sup>1</sup>, Andréa GUERRERO<sup>1</sup>, Yannick DEVILLE<sup>1</sup>†

<sup>1</sup>IRAP, Université de Toulouse, UPS, CNRS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France

<sup>2</sup>LETSMP, Faculté des Sciences, Université Ibn Zohr, BP 8106 Cité Dakhla, Agadir, Maroc

**Résumé** – Cet article traite le problème de la décontamination des spectres en spectroscopie sans fente dans le cadre de la mission spatiale *Euclid*. Ce problème peut être formulé comme un problème de séparation de sources qui consiste à estimer un ensemble de signaux inconnus, appelés sources, à partir d'un ensemble de leurs mélanges. Nous présentons d'abord un modèle de mélange reliant les données observées aux spectres de sources qui prend en compte simultanément quatre directions de dispersion du grisme, puis nous proposons deux méthodes pour décontaminer ces spectres à partir des mélangés. L'exploitation de toutes ces directions de dispersion permet d'améliorer l'estimation du spectre d'un objet d'intérêt. Les résultats préliminaires obtenus en utilisant des données bruitées réalistes montrent l'efficacité des méthodes proposées.

**Abstract** – This article treats the problem of spectra decontamination in slitless spectroscopy in the framework of the *Euclid* space mission. This problem can be formulated as a source separation problem consisting in estimating a set of unknown signals, called sources, from a set of their mixtures. We begin by presenting a mixing model linking observed data to source spectra that simultaneously considers four dispersion directions of the grism, and then we propose two methods to decontaminate these spectra from the observed data. Using all these dispersion directions improves the estimation of the spectrum of an object of interest. Preliminary results obtained using realistic noisy data demonstrate the efficiency of the proposed methods.

## 1 Introduction

*Euclid* est un télescope spatial de l'Agence Spatiale Européenne [1], dont le lancement est prévu en 2023. Son objectif principal est de comprendre la nature de l'énergie noire et d'expliquer comment cette énergie est responsable de l'accélération croissante de l'expansion de l'Univers. *Euclid* sera équipé d'un spectrographe sans fente proche infrarouge qui mesurera les spectres de plus de 50 millions de galaxies. Ces spectres seront ensuite analysés pour estimer les décalages spectraux vers le rouge (*redshifts*) des galaxies, ce qui devrait permettre de mieux comprendre comment l'énergie noire contribue à cette accélération. La spectroscopie est réalisée à l'aide de plusieurs grismes qui sont une combinaison de prismes et de réseaux de diffraction. Un grisme disperse différemment, dans une direction de dispersion, les différentes longueurs d'onde de la lumière émise. Comme un objet s'étale généralement sur plusieurs pixels, la sortie du grisme est une image avec plusieurs lignes, qui sera appelée par la suite *spec-*

*trogramme*<sup>1</sup>. Cependant, comme le montre la figure 1, la spectroscopie sans fente utilisée dans *Euclid* conduit à la superposition de spectrogrammes d'objets astronomiques voisins (galaxies et étoiles), ce qui pourrait entraîner une mauvaise estimation des redshifts [2]. Afin de surmon-

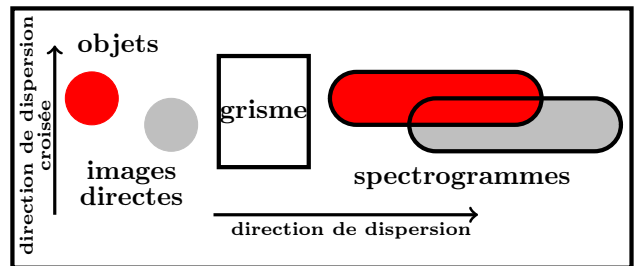


FIGURE 1 – Contamination par des objets voisins.

ter ce problème, plusieurs spectrogrammes seront générés dans différentes directions de dispersion. Par conséquent, le spectre d'un objet d'intérêt est contaminé par différents spectres dans différentes directions. Ensuite, l'exploitation de toutes ces directions permet de mieux estimer le spectre de l'objet d'intérêt.

\*This is a French adaptation of our English paper "Decontamination of galaxy spectra using four dispersion directions", accepted at Eusipco 2022 on behalf of the Euclid Consortium.

†Ce travail a été partiellement financé par le CNES (Centre National d'Etudes Spatiales, France).

1. Les astronomes l'appellent spectrogramme puisqu'il est généré par un spectrographe optique. Il ne faut pas le confondre avec le spectrogramme résultant de l'analyse temps-fréquence d'un signal.

La décontamination des spectres en spectroscopie sans fente peut être vue comme un problème de séparation de sources qui consiste à estimer un ensemble de signaux inconnus, appelés sources, à partir d'un ensemble de leurs mélanges [3]. En effet, nous avons montré dans [4, 5] que sous certaines hypothèses, le spectrogramme contaminé d'un objet d'intérêt dans chaque direction de dispersion du grisme peut être approximé par un modèle linéaire instantané. Nous avons proposé dans [4, 5] une méthode aveugle basée sur la factorisation en matrices non-négatives (NMF) [6, 7] pour décontaminer les spectres, en ne considérant que les spectrogrammes correspondant aux deux directions de dispersion 0 et 90 degrés. Cependant, le consortium Euclid a récemment adopté une autre stratégie d'observation qui exclut l'utilisation d'un grisme dans la direction 90 degrés. La nouvelle stratégie consiste à générer des spectrogrammes dans quatre directions de dispersion, à savoir les directions 0, 180, 184 et  $-4$  degrés. Par conséquent, les méthodes que nous avons proposées précédemment ne sont plus applicables à cette nouvelle stratégie. Par ailleurs, on sait que les méthodes basées sur la NMF sont sensibles à l'initialisation et ne garantissent pas l'unicité de la solution, ce qui est primordial dans un projet spatial comme Euclid. Dans cet article, nous proposons deux nouvelles méthodes exploitant les données spectrales fournies dans ces 4 directions, ainsi que les images directes fournies par les photomètres proche infrarouge d'*Euclid*. Dans un premier temps, les images directes sont utilisées pour obtenir une estimation des coefficients de mélange, puis le spectre de l'objet d'intérêt est estimé à partir de ces images et des spectrogrammes contaminés observés.

## 2 Méthodes proposées

### 2.1 Modèle de mélange

Nous avons montré dans [4] que pour chaque direction de dispersion du grisme, notée  $d_i$ , le mélange contenant le spectrogramme contaminé d'un objet d'intérêt supposé non bruité peut être approximé par un modèle linéaire instantané comme suit :

$$\mathbf{X}^{(d_i)} = \mathbf{A}^{(i)} \mathbf{E}^{(i)}, \quad (1)$$

où  $\mathbf{X}^{(d_i)}$  est la matrice de taille  $M_i \times K$  contenant les données observées, avec  $M_i$  le nombre de lignes dans la direction orthogonale à la direction de dispersion associée au spectrogramme de l'objet d'intérêt et  $K$  le nombre de bandes spectrales considérées identiques pour toutes les directions de dispersion.  $\mathbf{A}^{(i)} = [\mathbf{a}_s^{(i)} | \mathbf{A}_c^{(i)}]$  est la matrice de mélange de taille  $M_i \times (N_i + 1)$ , définie par :

$$\mathbf{A}^{(i)} = \left( \begin{array}{c|ccc} \overbrace{\mathbf{a}_s^{(i)}} & \overbrace{\mathbf{A}_c^{(i)}} & & \\ \hline a_s^{(i)}(1) & a_{c_1}^{(i)}(1) & \dots & a_{c_{N_i}}^{(i)}(1) \\ \vdots & \vdots & \ddots & \vdots \\ a_s^{(i)}(M_i) & a_{c_1}^{(i)}(M_i) & \dots & a_{c_{N_i}}^{(i)}(M_i) \end{array} \right) \quad (2)$$

où  $N_i$  est le nombre de contaminants considérés dans la direction  $d_i$ ,  $a_s^{(i)}(m)$  et  $a_{c_n}^{(i)}(m)$ ,  $(m, n) \in [1, M_i] \times [1, N_i]$ , sont respectivement les coefficients de mélange de l'objet d'intérêt, constituant le vecteur  $\mathbf{a}_s^{(i)} = [a_s^{(i)}(1), \dots, a_s^{(i)}(M_i)]^T$ , et les contaminants constituant la matrice  $\mathbf{A}_c^{(i)}$ . Quant à  $\mathbf{E}^{(i)}$ , elle est la matrice des sources de taille  $(N_i + 1) \times K$  dont la première ligne correspond au spectre de l'objet d'intérêt, tandis que ses autres lignes contiennent les spectres des contaminants dans la direction  $d_i$ .

Comme mentionné dans la section 1, pour chaque objet d'intérêt, nous disposons de quatre observations pour les quatre directions de dispersion (0, 180, 184, et  $-4$  degrés). En supposant que l'objet d'intérêt a le même spectre dans toutes les directions de dispersion, nous pouvons combiner ces quatre observations afin d'améliorer l'estimation du spectre de cet objet. On redéfinit alors la matrice d'observation  $\mathbf{X}$  d'un objet d'intérêt, dont on cherche à décontaminer le spectre, en concaténant ses spectrogrammes contaminés dans les quatre directions comme suit :

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}^{(d_1)} \\ \mathbf{X}^{(d_2)} \\ \mathbf{X}^{(d_3)} \\ \mathbf{X}^{(d_4)} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^{(0)} \\ \mathbf{X}^{(180)} \\ \mathbf{X}^{(184)} \\ \mathbf{X}^{(-4)} \end{bmatrix} = \mathbf{A} \mathbf{E}, \quad (3)$$

où  $\mathbf{A} = [\mathbf{a}_s | \mathbf{A}_c]$  est la matrice de mélange totale de taille  $M \times (N + 1)$ , avec  $M = \sum_{i=1}^4 M_i$ ,  $N = \sum_{i=1}^4 N_i$ ,  $\mathbf{a}_s$  est le vecteur de mélange de l'objet d'intérêt, défini par :

$$\mathbf{a}_s = [\mathbf{a}_s^{(1)T}, \mathbf{a}_s^{(2)T}, \mathbf{a}_s^{(3)T}, \mathbf{a}_s^{(4)T}]^T \quad (4)$$

et  $\mathbf{A}_c$  est la matrice de mélange des contaminants. En supposant que les contaminants dans chaque direction de dispersion sont distincts, cette matrice peut s'écrire :

$$\mathbf{A}_c = \begin{bmatrix} \mathbf{A}_c^{(1)} & 0 & 0 & 0 \\ 0 & \mathbf{A}_c^{(2)} & 0 & 0 \\ 0 & 0 & \mathbf{A}_c^{(3)} & 0 \\ 0 & 0 & 0 & \mathbf{A}_c^{(4)} \end{bmatrix}. \quad (5)$$

$\mathbf{E}$  est la matrice de taille  $(N + 1) \times K$  contenant le spectre de l'objet d'intérêt dans sa première ligne et les spectres des contaminants dans chacune des quatre directions dans les lignes suivantes.

Dans les deux sous-sections suivantes, nous présenterons deux méthodes qui peuvent être utilisées pour estimer le spectre de l'objet d'intérêt en exploitant la matrice d'observation  $\mathbf{X}$  et les informations disponibles sur les objets.

### 2.2 Première méthode

La première étape de cette méthode consiste à estimer la matrice de mélange  $\mathbf{A}$  en utilisant l'image directe de l'objet d'intérêt ainsi que celles de ses contaminants. En effet, *Euclid* est également équipé d'un photomètre qui fournit des images directes de tous les objets astronomiques dans le champ de vue [2]. Comme on peut le voir sur la figure 1, le flux de l'image directe est dispersé par un grisme pour

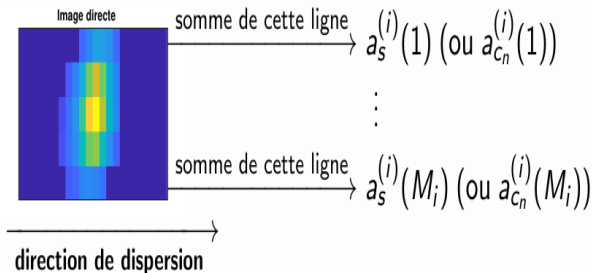


FIGURE 2 – Estimation des coefficients de mélange à l'aide d'images directes.

produire son spectrogramme. Ainsi, pour chaque direction de dispersion, les coefficients de mélange ( $a_s^{(i)}(m)$  ou  $a_{c_n}^{(i)}(m)$ ) relatifs à un objet (cible ou contaminant) peuvent être estimés en calculant la somme des valeurs de pixels sur les différentes lignes de son image directe, ce qui nous permettra d'estimer la matrice de mélange totale  $\mathbf{A}$ . Pour estimer ces coefficients, nous sur-échantillonons d'abord l'image directe de chaque objet dans la direction de la dispersion croisée, ensuite nous décalons cette image dans la direction de dispersion croisée pour la recentrer avec le spectrogramme de l'objet considéré, et enfin nous la sous-échantillonons par le même taux d'échantillonnage. Ceci nous permet de corriger le décalage entre l'image directe et le spectrogramme de cet objet. Après ce prétraitement, nous choisissons les  $M_i \times R$  valeurs des pixels de l'image centrée sur la position de l'objet d'intérêt, où  $R$  est le nombre de colonnes dans l'image directe. Ensuite, la somme des valeurs de pixels de la  $m$ -ième ligne fournit la valeur  $a_s^{(i)}(m)$  ou  $a_{c_n}^{(i)}(m)$  de cet objet. Ce processus est illustré en figure 2. La matrice de mélange  $\mathbf{A}$  est ensuite construite à l'aide des coefficients estimés. Ensuite, une première solution pour estimer la matrice des sources  $\mathbf{E}$  consiste à minimiser le critère suivant, où  $\|\cdot\|_2$  représente la norme de Frobenius :

$$J_1 = \|\mathbf{X} - \mathbf{A}\mathbf{E}\|_2^2, \quad (6)$$

ce qui conduit à l'estimation des moindres carrés (LSQ) :

$$\mathbf{E} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}. \quad (7)$$

Une deuxième solution pour estimer  $\mathbf{E}$  consiste à utiliser l'algorithme des moindres carrés non négatifs (NNLSQ) présentée dans [8]. En effet, les spectres à estimer étant par définition non-négatifs, cette méthode prend en compte cette non-négativité par la minimisation du critère :

$$J_2 = \|\mathbf{X} - \mathbf{A}\mathbf{E}\|_2^2 \text{ tel que } \mathbf{E} \geq 0 \quad (8)$$

où  $\mathbf{E} \geq 0$  signifie que toutes les valeurs de la matrice  $\mathbf{E}$  sont non-négatives. Enfin, le spectre de l'objet d'intérêt, noté  $\hat{\mathbf{E}}_s$ , est la première ligne de la matrice estimée  $\hat{\mathbf{E}}$ .

### 2.3 Seconde méthode

Contrairement à la méthode présentée dans la sous-section 2.2 qui nécessite l'image directe de l'objet d'intérêt

et celles de tous ses contaminants, la méthode proposée dans cette sous-section ne nécessite que l'image de l'objet d'intérêt afin d'estimer son spectre. En effet, cette image permet d'estimer les coefficients de mélange  $a_s^{(i)}(m)$  de l'objet d'intérêt dans toutes les directions de dispersion, ce qui nous permettra d'estimer le vecteur de mélange  $\mathbf{a}_s$  de cet objet par la même procédure que celle décrite dans la sous-section 2.2. Ce vecteur sera ensuite exploité par une méthode de formation de voie ou "beamformer" pour estimer le spectre de l'objet d'intérêt. Le but du beamformer est d'extraire un signal utile (spectre de l'objet d'intérêt) tout en atténuant les interférences (spectres des contaminants) et le bruit [9, 10]. Dans cet article, nous avons utilisé le beamformer baptisé MPDR [10]. Ce beamformer vise à estimer un filtre optimal noté  $\mathbf{w}_{\text{MPDR}}$ , dont la sortie minimise la puissance totale sous la contrainte  $\mathbf{w}^H \mathbf{a}_s = 1$ . Le filtre souhaité est la solution du problème de minimisation suivant :

$$\mathbf{w}_{\text{MPDR}} = \underset{\mathbf{w}}{\text{argmin}} \{ \mathbf{w}^H \mathbf{R}_X \mathbf{w} \} \text{ tel que } \mathbf{w}^H \mathbf{a}_s = 1, \quad (9)$$

ce qui conduit aux coefficients de beamforming suivants<sup>2</sup> [10] :

$$\mathbf{w}_{\text{MPDR}} = \frac{\mathbf{R}_X^{-1} \mathbf{a}_s}{\mathbf{a}_s^H \mathbf{R}_X^{-1} \mathbf{a}_s}, \quad (10)$$

où  $\mathbf{R}_X = \mathbf{X}\mathbf{X}^H / K$  est la matrice de covariance des observations. Enfin, le spectre de l'objet d'intérêt est estimé comme suit :

$$\hat{\mathbf{E}}_s = \mathbf{w}_{\text{MPDR}}^H \mathbf{X}. \quad (11)$$

## 3 Résultats

Pour évaluer les performances de nos méthodes, nous avons choisi un scénario où le spectrogramme d'un objet d'intérêt (une galaxie avec un redshift de 1,12) est contaminé par les spectres de quatre objets dans les deux directions 0 degré et 180 degrés, par les spectres de deux autres objets dans la direction 184 degrés, et par le spectre d'un autre objet dans la direction -4 degrés. Les spectrogrammes observés et les spectres 1D correspondants<sup>3</sup> concernant cet objet sont représentés sur la figure 3. Pour mesurer les performances de décontamination, nous avons utilisé comme critère le NRMSE (*Normalized Root-Mean-Square Error*) avant et après décontamination, défini par :

$$\text{NRMSE}_{\text{in}} = \frac{1}{4} \sum_{d_i \in \{0, 180, 184, -4\}} \frac{\|\mathbf{E}_s - \mathbf{E}_x^{(d_i)}\|_2}{\|\mathbf{E}_s\|_2}, \quad (12)$$

$$\text{NRMSE}_{\text{out}} = \frac{\|\mathbf{E}_s - \hat{\mathbf{E}}_s\|_2}{\|\mathbf{E}_s\|_2}, \quad (13)$$

où  $\mathbf{E}_x^{(d_i)}$  est le spectre 1D des données observées dans la direction  $d_i$ ,  $\mathbf{E}_s$  est le vrai spectre non bruité de l'objet

2. Notons qu'il est nécessaire de centrer la matrice d'observation  $\mathbf{X}$  avant d'effectuer le beamforming.

3. Le spectre 1D observé est défini ici comme la moyenne de toutes les lignes du spectrogramme.

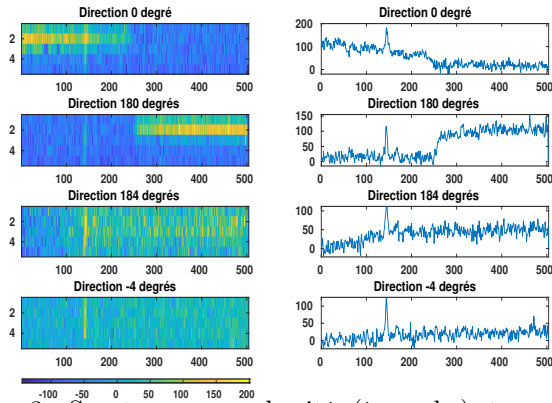


FIGURE 3 – Spectrogrammes bruités (à gauche) et spectres 1D correspondants (à droite).

TABLE 1 – NRMSE avant et après la décontamination.

NRMSE <sub>in</sub>	NRMSE <sub>out</sub>			
	LSQ	NNLSQ	Second method	NMF-ALS
10.67	0.90	<b>0.73</b>	0.92	1.19

d'intérêt et  $\hat{\mathbf{E}}_s$  son estimation. Dans (12) et (13), tous les spectres sont centrés, et  $\hat{\mathbf{E}}_s$  et  $\mathbf{E}_x^{(d_i)}$  sont normalisés pour avoir la même variance que  $\mathbf{E}_s$ . Les méthodes proposées ont été comparées à la méthode NMF Alternating Least Square (NMF-ALS) présentée dans [11]. Le vrai spectre non bruité de l'objet d'intérêt et ses estimations à l'aide de nos méthodes et de la méthode NMF-ALS après normalisation par leur maximum sont présentés dans la figure 4. Cette figure montre clairement l'efficacité des méthodes proposées pour décontaminer le spectre de l'objet d'intérêt. En effet, nos méthodes éliminent avec succès les contaminations des autres objets pour ce scénario. La méthode NMF-ALS a des performances plutôt médiocres comme on peut le voir sur la figure 4. Les résultats numériques sont présentés dans le tableau 1. Notez que bien que le NRMSE<sub>out</sub> des méthodes proposées semble élevé dans ce scénario, le NRMSE<sub>in</sub> des données observées est égal à 10,67, ce qui est beaucoup plus important comparé à ceux obtenus avec nos méthodes. Comme on peut le constater, notre première méthode utilisant l'algorithme NNLSQ donne les meilleurs résultats pour ce scénario.

## 4 CONCLUSION

Dans cet article, nous avons proposé deux nouvelles méthodes de décontamination qui sont basées sur la concaténation des données issues de quatre directions de dispersion du grisme et qui sont adaptées à la nouvelle stratégie d'observation d'*Euclid*. En effet, l'exploitation de toutes ces observations permet d'améliorer l'estimation du spectre de l'objet d'intérêt. D'après les résultats des tests effectués, notre première méthode utilisant l'algorithme NNLSQ donne les meilleurs résultats en termes de NRMSE mais elle nécessite la connaissance des images directes de tous les objets. Néanmoins, une étude plus complète des performances serait souhaitable pour confirmer ce résultat.

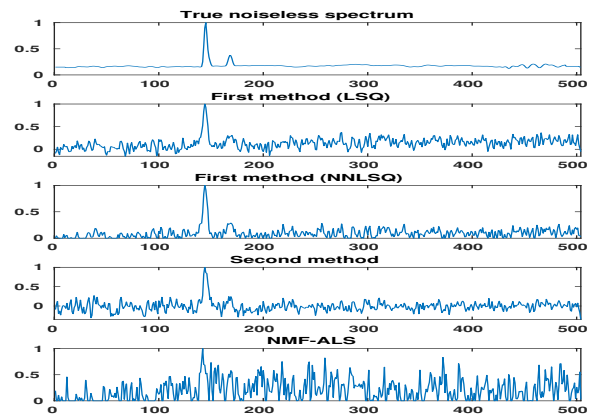


FIGURE 4 – Le vrai spectre non bruité de l'objet d'intérêt et ses estimations en utilisant les trois méthodes.

## Remerciements

Les auteurs remercient leurs collègues au sein des unités organisationnelles SIR et SIM du projet *Euclid*.

## Références

- [1] “Euclid consortium,” <https://www.euclid-ec.org>, 2022.
- [2] “Euclid definition study report,” <https://arxiv.org/pdf/1110.3193.pdf>, 2011.
- [3] Y. Deville, “Blind source separation and blind mixture identification methods,” 2016, pp. 1-33, Wiley Encyclopedia of Electrical and Electronics Engineering.
- [4] A. Selloum, S. Hosseini, Y. Deville, and T. Contini, “Mixing model in slitless spectroscopy and resulting blind methods for separating galaxy spectra,” in 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), 2016, pp. 1-6.
- [5] S. Hosseini, A. Selloum, T. Contini, and Y. Deville, “Separation of galaxy spectra measured with slitless spectroscopy,” *Digit Signal Processing*, vol. 106, pp. 102837, 2020.
- [6] X. Fu, K. Huang, N.D. Sidiropoulos, and W. Ma, “Nonnegative matrix factorization for signal and data analytics : Identifiability, algorithms, and applications,” *IEEE Signal Processing Magazine*, vol. 36, no. 2, pp. 59-80, 2019.
- [7] D.D. Lee and H.S. Seung, “Learning the parts of objects by non-negative matrix factorization,” *Nature*, vol. 401, pp. 788-791, 1999.
- [8] R.J. Lawson, C.L. Hanson, “Solving least squares problems,” 1974.
- [9] J. Capon, “High-resolution frequency-wavenumber spectrum analysis,” *Proceedings of the IEEE*, vol. 57, pp. 1408-1418, August 1969.
- [10] H.L. Van Trees, “Adaptive beamformers,” in *Optimum Array Processing*. John Wiley and Sons, 2002, pp. 710-916.
- [11] P. Paatero and U. Tapper, “Positive matrix factorization : A non-negative factor model with optimal utilization of error estimates of data values,” *Environmetrics*, vol. 5, pp. 111-126, 1994.