

Méthode de Factorisation en Matrices Non négatives pour des mélanges linéaires quadratiques

Inès MEGANEM^{1,2}, Yannick DEVILLE¹, Shahram HOSSEINI¹, Philippe DÉLIOT², Xavier BRIOTTET², Leonardo T. DUARTE³

¹IRAP, CNRS, Université de Toulouse, UPS-OMP, 14 av. Edouard Belin, 31400 Toulouse, France.

²ONERA, 2 av. Edouard Belin, 31055 Toulouse, France.

³Laboratoire DSPCom, Université de Campinas (UNICAMP), av. Albert Einstein, 400, CEP: 13083-852, Campinas, Brésil

{ imeganem, ydeville, shosseini } @ast.obs-mip.fr
{ philippe.deliot, xavier.briottet }@onera.fr, ltduarte@dmo.fee.unicamp.br

Résumé – Nous nous intéressons au problème de séparation de sources dans le cadre d’un mélange linéaire quadratique. Les mélanges linéaires quadratiques sont peu étudiés dans la littérature, en comparaison des mélanges linéaires. Ce type de mélange est néanmoins intéressant pour certaines applications, notamment en imagerie hyperspectrale, pour des scènes de télédétection en milieu urbain. Nous proposons ici une méthode originale fondée sur la Factorisation en Matrices Non négatives (NMF). La NMF est bien adaptée à certaines applications à données positives. Nous étendons ici une méthode de type NMF, qui est une méthode pour mélanges linéaires, à des mélanges linéaire quadratiques. La méthode de séparation de sources proposée est testée avec des mélanges de signaux artificiels et les résultats sont très encourageants. Nous montrons aussi ce que donne son application dans le cas de démixage d’images hyperspectrales, en la testant sur des mélanges de spectres de réflectances.

Abstract – We here focus on source separation for linear-quadratic mixing models. Linear-quadratic models have been less studied in the literature than linear ones. This class of mixing models is interesting since it is encountered in some real applications, such as hyperspectral unmixing for urban environments. We here propose a source separation method based on Non-negative Matrix Factorization (NMF). NMF methods are well-suited for some applications where data are positive. The originality of our work is that we here developed an extension of NMF, which is initially a method for linear mixtures, for our linear-quadratic model. To validate our method for linear-quadratic models we present tests with mixtures of artificial signals that yield attractive performance. We also show how we can apply it to hyperspectral unmixing, by means of an example with artificial mixtures of reflectance spectra.

1 Introduction

Nous nous intéressons dans cet article à la séparation de sources dans le cadre de mélanges linéaires quadratiques instantanés. Ces mélanges ont été moins étudiés que les mélanges linéaires. Il existe néanmoins quelques approches dans la littérature, essentiellement fondées sur l’Analyse en Composantes Indépendantes (ACI) [3] ou des approches bayésiennes [2]. Le modèle de mélange linéaire quadratique est intéressant car il intervient dans différentes applications réelles telles que la télédétection à laquelle nous nous intéressons ici dans le cadre du démixage hyperspectral [5]. Ce modèle de mélange est alors justifié lorsque les réflexions entre les différents matériaux interviennent, c-à-d quand les scènes étudiées ne sont pas planes. C’est le cas par exemple des milieux urbains.

Nous proposons une méthode de type NMF (Non-negative Matrix Factorization) [1], que nous développons spécifiquement pour notre type de mélange. Le choix de ce type de méthode est justifié par le fait qu’il est bien adapté à des applications réelles où les données sont positives et qu’il reste envisageable dans le cas où les sources ne sont pas indépendantes, à la différence des méthodes de type ACI par exemple. C’est

notamment le cas des images hyperspectrales en télédétection.

Pour valider notre méthode, nous présentons d’abord des tests sur des signaux artificiels. Nous nous intéressons ensuite au démixage hyperspectral et présentons dans ce cadre un exemple des résultats obtenus avec des mélanges de spectres de réflectances.

Nous présenterons dans la section 2 notre méthode NMF pour le modèle de mélange linéaire quadratique. Dans la section 3 nous montrons les résultats obtenus avec des signaux artificiels, puis dans le cas de démixage hyperspectral quand le modèle est linéaire quadratique.

2 Méthode proposée

2.1 Le modèle de mélange et son adaptation à la NMF

Nous nous intéressons au modèle de mélange linéaire quadratique instantané et nous nous limiterons à ce stade à un mélange de deux sources s_1 et s_2 . Ce mélange s’écrit donc

comme suit, pour un échantillon n de l'observation \mathbf{x}_i

$$x_i(n) = a_1(i)s_1(n) + a_2(i)s_2(n) + a_{1,2}(i)s_1(n)s_2(n). \quad (1)$$

Cela donne sous forme matricielle

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (2)$$

où :

- $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_P]^T$, avec $\mathbf{x}_i = [x_i(1) \cdots x_i(K)]^T$, P étant le nombre d'observations ($P \geq 2$) et K le nombre d'échantillons
- $\mathbf{S} = [\mathbf{s}_1 \quad \mathbf{s}_2 \quad \mathbf{s}_1 \odot \mathbf{s}_2]^T$, avec \odot la multiplication élément par élément, et où $\mathbf{s}_j = [s_j(1) \cdots s_j(K)]^T$
- $\mathbf{A} = \begin{bmatrix} a_1(1) & a_2(1) & a_{1,2}(1) \\ \vdots & \vdots & \vdots \\ a_1(P) & a_2(P) & a_{1,2}(P) \end{bmatrix}$ est la matrice de mélange.

Nous supposons que nos données (sources et coefficients de mélange) sont non négatives. L'expression (2) va donc nous permettre de développer une méthode de type NMF, en considérant que la troisième ligne de la matrice \mathbf{S} correspond à une pseudo-source égale au produit de nos deux sources réelles. Nous appellerons dans la suite cette pseudo-source, la "troisième source", mais sans oublier que notre mélange est en fait réellement composé d'uniquement deux sources, \mathbf{s}_1 et \mathbf{s}_2 .

2.2 Méthode proposée : LQ-NMF

Principe de la NMF : étant donné une matrice non négative \mathbf{V} , la NMF consiste à trouver des matrices non négatives \mathbf{W} et \mathbf{H} telles que $\mathbf{V} \approx \mathbf{W}\mathbf{H}$.

La manière dont nous avons écrit notre modèle de mélange dans l'équation (2) nous permet d'étendre le principe de cette méthode pour mélanges linéaires à notre cas linéaire quadratique.

Différents algorithmes de type NMF ont été proposés dans la littérature pour des mélanges linéaires (voir [1]), basés sur différents critères à minimiser avec ou sans contraintes et donnant lieu à différentes règles de mise à jour (algorithmes itératifs). Nous avons opté pour le critère classique de norme euclidienne (ou de Frobenius), avec un simple algorithme de descente en gradient (voir [1]). Mais notre modèle nécessite un nouveau calcul du gradient qui aboutit à de nouvelles expressions de mise à jour pour l'estimation de notre matrice \mathbf{S} .

Pour faciliter le calcul nous réécrivons nos matrices comme suit

$$\bullet \mathbf{A} = [\mathbf{A}_a \quad \mathbf{A}_b], \text{ où } \mathbf{A}_a = \begin{bmatrix} a_1(1) & a_2(1) \\ \vdots & \vdots \\ a_1(P) & a_2(P) \end{bmatrix}$$

$$\text{et } \mathbf{A}_b = \begin{bmatrix} a_{1,2}(1) \\ \vdots \\ a_{1,2}(P) \end{bmatrix}$$

$$\bullet \mathbf{S} = \begin{bmatrix} \mathbf{S}_a \\ \mathbf{S}_b \end{bmatrix}, \text{ où } \mathbf{S}_a = [\mathbf{s}_1 \quad \mathbf{s}_2]^T \text{ et } \mathbf{S}_b = [\mathbf{s}_1 \odot \mathbf{s}_2]^T.$$

Notre critère à minimiser s'écrit alors comme suit

$$J = \frac{1}{2} \|\mathbf{X} - \mathbf{A}\mathbf{S}\|_F^2 = \frac{1}{2} \|\mathbf{X} - \mathbf{A}_a\mathbf{S}_a - \mathbf{A}_b\mathbf{S}_b\|_F^2. \quad (3)$$

Le calcul de gradient par rapport à la matrice de mélange n'est pas altéré par notre modèle, par rapport au modèle linéaire.

Cela donne : $\frac{\partial J}{\partial \mathbf{A}} = -(\mathbf{X} - \mathbf{A}\mathbf{S})\mathbf{S}^T$, ce qui aboutit à la mise à jour suivante (où α est un gain d'adaptation positif) :

$$\mathbf{A}^{(m+1)} = \mathbf{A}^{(m)} - \alpha \frac{\partial J}{\partial \mathbf{A}}. \quad (4)$$

Le fait que la troisième source soit le produit de \mathbf{s}_1 et \mathbf{s}_2 change par contre le calcul du gradient par rapport aux sources, en comparaison du modèle linéaire. Nous écrivons donc d'abord notre critère sous forme scalaire comme suit

$$\begin{aligned} J &= \frac{1}{2} \sum_{i,n} [\mathbf{X} - \mathbf{A}_a\mathbf{S}_a - \mathbf{A}_b\mathbf{S}_b]_{in}^2 \\ &= \frac{1}{2} \sum_{i,n} [x_{in} - (\mathbf{A}_a\mathbf{S}_a)_{in} - (\mathbf{A}_b\mathbf{S}_b)_{in}]^2 \\ &= \frac{1}{2} \sum_{i,n} [x_{in} - \sum_{j=1}^2 a_j(i)S_{jn} - a_{1,2}(i)S_{1n}S_{2n}]^2. \end{aligned}$$

Le gradient de J par rapport à S_{1n} (où $S_{jn} = [\mathbf{S}]_{jn}$) se calcule alors comme suit :

$$\begin{aligned} \frac{\partial J}{\partial S_{1n}} &= \sum_i (x_{in} - \sum_{j=1}^2 a_j(i)S_{jn} - a_{1,2}(i)S_{1n}S_{2n}) \\ &\quad \times (-a_1(i) - a_{1,2}(i)S_{2n}) \\ &= -\sum_i a_1(i)(x_{in} - \sum_{j=1}^2 a_j(i)S_{jn} - a_{1,2}(i)S_{1n}S_{2n}) \quad (5) \\ &\quad - S_{2n} \sum_i a_{1,2}(i)(x_{in} - \sum_{j=1}^2 a_j(i)S_{jn} - a_{1,2}(i)S_{1n}S_{2n}). \end{aligned}$$

L'expression du gradient par rapport à S_{2n} peut se déduire facilement de (5) en inversant, pour $a_j(i)$ et S_{jn} , les indices 1 et 2.

En remplaçant l'écriture scalaire par des produits matriciels, nous obtenons

$$\begin{aligned} \frac{\partial J}{\partial S_{1n}} &= -[\mathbf{A}_a^T(\mathbf{X} - \mathbf{A}\mathbf{S})]_{1n} - [\mathbf{S}_a]_{2n} \times [\mathbf{A}_b^T(\mathbf{X} - \mathbf{A}\mathbf{S})]_n \\ \frac{\partial J}{\partial S_{2n}} &= -[\mathbf{A}_a^T(\mathbf{X} - \mathbf{A}\mathbf{S})]_{2n} - [\mathbf{S}_a]_{1n} \times [\mathbf{A}_b^T(\mathbf{X} - \mathbf{A}\mathbf{S})]_n. \end{aligned}$$

Cela nous donne, pour une itération $(m+1)$, les règles de mise à jour suivantes pour les sources

$$S_{1n}^{(m+1)} = S_{1n}^{(m)} - \alpha \frac{\partial J}{\partial S_{1n}} \quad (6)$$

$$S_{2n}^{(m+1)} = S_{2n}^{(m)} - \alpha \frac{\partial J}{\partial S_{2n}} \quad (7)$$

$$S_{3n}^{(m+1)} = S_{1n}^{(m+1)} S_{2n}^{(m+1)}. \quad (8)$$

Comme nous pouvons le voir dans (8), pour la troisième ligne de la matrice \mathbf{S} , nous copions le produit élément par élément des deux premières lignes. Nous forçons ainsi la 3^{ème} source à coller à la réalité du mélange en lui imposant d'être égale au produit des deux premières.

Pour garantir la non négativité des résultats dans les équations (4) (6) et (7), nous optons pour la solution (utilisée dans la littérature [1]) qui consiste à comparer, après chaque itération de l'algorithme, les mises à jour avec un ϵ très faible et positif et à garder le maximum des deux :

$$\mathbf{A}^{(m+1)} = \max\{\mathbf{A}^{(m+1)}, \epsilon\} \quad (9)$$

$$\mathbf{S}_a^{(m+1)} = \max\{\mathbf{S}_a^{(m+1)}, \epsilon\}. \quad (10)$$

Cela consiste plus précisément à comparer chaque élément de la matrice à ϵ , et s'il est négatif à le remplacer par ϵ . Nous éliminons ainsi tous les éléments négatifs et garantissons la non négativité sur laquelle reposent les méthodes NMF.

Il est connu que la NMF ne fournit pas de solution unique [4], qu'elle ne garantit pas la convergence vers un minimum global et que les performances peuvent dépendre de l'initialisation de l'algorithme [1]. Pour réduire ce phénomène, des contraintes dépendant des propriétés des signaux impliqués peuvent être ajoutées au critère qu'on minimise, telles que des contraintes de parcimonie par exemple.

A ce stade de notre étude nous nous plaçons dans un cas général, sans contraintes, qui peut donc être facilement adapté à n'importe quelle application et dont les performances peuvent être améliorées par l'ajout de contraintes adéquates.

3 Résultats de tests

3.1 Tests sur signaux artificiels

Nous testons d'abord l'algorithme sur des mélanges de signaux aléatoires mutuellement indépendants, constitués de 100 échantillons i.i.d. distribués uniformément sur $[0, 1]$.

Nous proposons ici une étude des performances en fonction de l'éloignement de l'initialisation de la NMF par rapport aux vraies données.

Nous initialisons donc l'algorithme avec \mathbf{A}^0 et \mathbf{S}^0 obtenues à partir des vraies données auxquelles on ajoute un bruit dont l'importance est variable. Notre initialisation est ainsi plus ou moins différente des données suivant l'importance de ce bruit ajouté. Ceci nous permet de comparer l'erreur d'estimation en sortie de la NMF, $erreur_f$, avec l'erreur en entrée, $erreur_i$, qui sont définies respectivement comme suit :

$$erreur_f = \frac{\|\mathbf{Y} - \widehat{\mathbf{Y}}\|_F}{\|\mathbf{Y}\|_F} \quad (11)$$

$$erreur_i = \frac{\|\mathbf{Y} - \mathbf{Y}^0\|_F}{\|\mathbf{Y}\|_F} \quad (12)$$

où \mathbf{Y} peut être \mathbf{S}_a , \mathbf{A}_a ou \mathbf{AS} . La notation \mathbf{Y}^0 représente l'initialisation et $\widehat{\mathbf{Y}}$ l'estimation finale.

Avant de calculer l'erreur sur les sources et les coefficients de mélange, les matrices estimées sont normalisées pour être mises à la même échelle que les vraies données (chaque ligne de $\widehat{\mathbf{S}}_a$ est divisée par sa puissance et multipliée par la puissance de la vraie source correspondante ; on fait de même pour les colonnes de $\widehat{\mathbf{A}}_a$). Ceci évite de considérer l'estimation à un

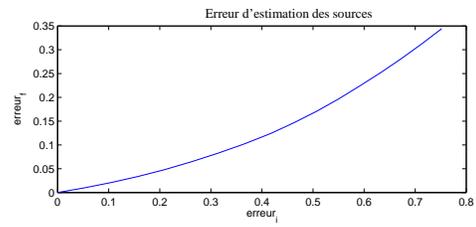


FIG. 1 – Erreur d'estimation des sources : $erreur_f$ en fonction de $erreur_i$

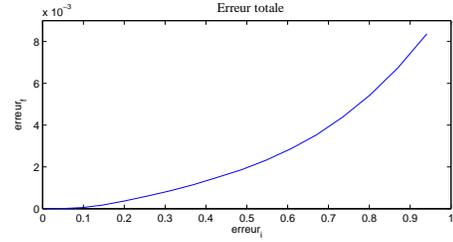


FIG. 2 – Erreur d'estimation totale : $erreur_f$ en fonction de $erreur_i$

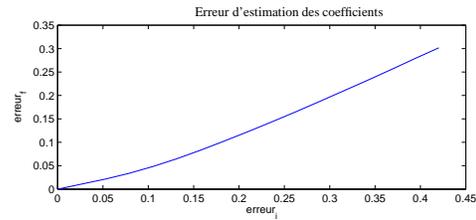


FIG. 3 – Erreur d'estimation des coefficients de mélange : $erreur_f$ en fonction de $erreur_i$

facteur d'échelle près comme une erreur, étant donné que c'est une indétermination connue des méthodes NMF.

Les résultats présentés ici ont été obtenus avec la matrice de mélange suivante

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 0.5 & 0.4 \\ 0.6 & 1 & 0.5 \end{bmatrix}.$$

Les résultats fournis correspondent à une moyenne sur dix couples de signaux aléatoires mélangés avec la même matrice, et avec dix initialisations différentes pour chaque cas (donc une moyenne sur 100 runs).

On étudie donc l'évolution de $erreur_f$ en fonction de $erreur_i$. Cette étude est intéressante car elle montre la robustesse de l'algorithme quand l'initialisation est très éloignée des vraies données, sachant que les performances de la NMF dépendent souvent de l'initialisation. De plus, dans des applications réelles, on dispose souvent d'informations a priori sur les données ou d'une bibliothèque de signaux similaires et ceci peut être utilisé lors de l'initialisation.

On peut voir sur la figure 1 l'erreur d'estimation des sources. L'estimation est bonne en comparaison de l'erreur en entrée, même quand \mathbf{S}^0 est très différente de \mathbf{S} (valeurs élevées de $erreur_i$). L'erreur totale (c-à-d calculée pour \mathbf{AS}) à l'issue de l'estimation reste en dessous de 10^{-2} (figure 2). Cela montre que l'algorithme converge bien. L'estimation des coefficients est quant à elle un peu moins satisfaisante (figure 3).

Comparons à présent l'erreur d'estimation sur les 3 figures en un point donné des courbes, pour de grandes valeurs de $erreur_i$, par exemple le dernier point. L'erreur totale est très faible, l'algorithme converge donc bien. Néanmoins, pour ce point, l'erreur d'estimation des sources et coefficients est non négligeable (environ 0.3). Il est alors évident que la minimisation du critère ne conduit pas forcément à une séparation de sources satisfaisante. Ceci peut être expliqué par la non-unicité des solutions de la NMF et la possibilité de convergence vers des minima locaux. Ce problème peut être réduit en ajoutant des contraintes au critère, comme mentionné précédemment.

3.2 Application au démixage hyperspectral

Nous nous intéressons maintenant au démixage hyperspectral en télédétection. Le principe est, pour un pixel représentant une surface donnée, d'extraire les spectres de réflectances des matériaux qui composent cette surface. Le cas le plus généralement traité est le mélange linéaire. Néanmoins, quand la scène imagée n'est pas plane, comme par exemple dans les milieux urbains, le mélange est linéaire quadratique. Ceci est dû aux réflexions entre les différentes surfaces. Comme les données dans ce cas sont non négatives (réflectances et coefficients de mélange), notre méthode NMF est applicable.

Les résultats présentés ici correspondent à des mélanges artificiels obtenus en mélangeant un couple de spectres réels avec la matrice A_1 présentée ci-dessus.

Pour la NMF, nous initialisons S^0 avec des spectres d'une bibliothèque, correspondant à des classes de matériaux différentes de celles des vraies sources. On suppose donc ne pas connaître les matériaux dans nos mélanges.

Pour initialiser la matrice de mélange, on utilise la méthode des moindres carrés avec contrainte de positivité. Connaissant S^0 , on applique cette méthode séparément à chaque observation pour obtenir l'initialisation A^0 de la matrice de mélange.

Sur la figure 4, on peut voir les spectres à l'origine de nos mélanges (ligne épaisse pointillée), les estimations (ligne continue) ainsi que les spectres utilisés pour l'initialisation (ligne

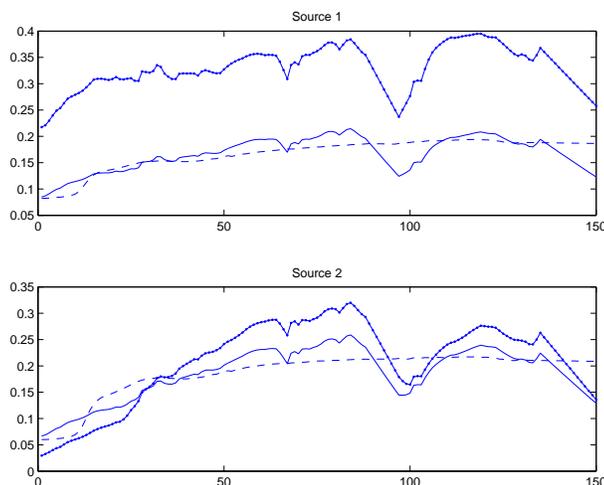


FIG. 4 – Vrais spectres (ligne épaisse pointillée), leurs estimations (ligne continue) et les initialisations (ligne discontinue)

discontinue). Pour cet exemple nous obtenons, en sortie de notre méthode, une erreur d'estimation de 0.16 pour les sources et de 0.46 pour les coefficients de mélange, alors que l'erreur initiale est de 0.4 pour les sources et supérieure à 1 pour les coefficients. Les erreurs sont calculées avec les mêmes formules que précédemment (utilisant les eq. (11) et (12)) et avec la même normalisation préalable.

Notons qu'ici estimer les spectres avec un facteur d'échelle ne pose pas de problème car dans la plupart des applications en télédétection, telles que la classification, seule l'allure du spectre est importante car spécifique à un matériau donné.

Les résultats obtenus ici sont donc encourageants, surtout que séparer des spectres de télédétection n'est pas simple, comparé au cas de signaux artificiels précédent, car ces spectres de réflectances sont souvent très corrélés (le coefficient de corrélation entre nos deux spectres ici est de 0.79) et même les méthodes NMF sont sensibles à ce problème.

4 Conclusion et perspectives

Cet article a présenté une méthode de séparation de sources pour des mélanges linéaires quadratiques, et ceci par extension d'une méthode NMF. Nous avons d'abord illustré les performances de cette méthode sur des mélanges de signaux artificiels. Nous avons aussi montré par un exemple de test la possibilité d'appliquer notre méthode au démixage hyperspectral lorsque l'on est confronté à des mélanges linéaires quadratiques.

La prochaine étape sera d'étendre la méthode au cas général de N sources avec les tests correspondants. Concernant l'application au démixage hyperspectral, il serait aussi intéressant d'ajouter des contraintes au critère de NMF et de tester notre approche sur des cas plus compliqués et de vraies images.

Références

- [1] A. Cichocki, R. Zdunek, A. H. Phan, S.-I. Amari. *Non-negative Matrix and Tensor Factorizations : Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation*, John Wiley and Sons, 2009.
- [2] L. T. Duarte, C. Jutten, S. Moussaoui, *Bayesian source separation of linear-quadratic and linear mixtures through a MCMC method*, Actes de IEEE MLSP, 2009.
- [3] S. Hosseini, Y. Deville, *Blind maximum likelihood separation of a linear-quadratic mixture*, Actes de ICA, 2004. Erratum : <http://arxiv.org/abs/1001.0863>.
- [4] S. Moussaoui, D. Brie, J. Idier, *Non-negative source separation : range of admissible solutions and conditions for the uniqueness of the solution*, Actes de IEEE ICASSP, 2005.
- [5] J. M. P. Nascimento, J. M. Bioucas-Dias. *Nonlinear Mixture Model for Hyperspectral Unmixing*, SPIE Conference on Image and Signal Processing for Remote Sensing XV 2009.