AUGMENTED COMPLEX MATRIX FACTORISATION

David Looney and Danilo P. Mandic

Imperial College London, UK

ABSTRACT

A novel framework for the factorisation of complex-valued data is derived using recent developments in complex statistics. Unlike existing factorisation tools the algorithms can cater for noncircularity of the input - a necessary feature in applications for modelling real-world data. It is furthermore shown how the framework can be constrained to incorporate nonnegativity, helping generate results which allow a more realistic interpretation. Simulations illustrate the usefulness and enhanced accuracy for modelling synthetic data and a mixture of acoustic stimuli.

Index Terms— complex matrix factorisation, widely linear model, nonnegativity

1. INTRODUCTION

Matrix factorisation is a data-driven tool which decomposes an input matrix to reflect some underlying data structure. When combined with constraints such as sparsity or nonnegativity the operation yields a decomposition model with improved physical interpretability compared to, for example, PCA or ICA [1] and has great appeal in computer vision [2] and acoustic scene analysis [3].

Standard acoustic analysis considers only real valued power spectra, recently however, the benefits of including phase information has been emphasized (for example in speech perception and detection [4, 5, 6]), highlighting the need for robust and physically meaningful complex matrix factorisation (CMF) algorithms. Recent work has, for instance, combined the concept of nonnegative constraints with CMF by obtaining factors for input magnitude and phase information separately [7]. However, it is clearly desirable to process phase and amplitude information simultaneously to cater for cross-dynamics in the real and imaginary parts of the input. Additionally, following results in widely linear adaptive signal processing [8, 9, 10], standard CMF algorithms appear limited in scope as they are optimal only for signals with rotation invariant probability distributions (circular data).

Real-world signals are almost invariably noncircular, and their second order statistical properties need to be considered within the framework of so-called augmented complex statistics and based on the widely linear model. To this end, we propose a general model termed augmented complex matrix factorisation (ACMF), capable of dealing with all available second order information in the data, contained in both the covariance and pseudocovariance. Theoretical justification is also given for the flexibility of the model to impose practical constraints. It is shown for instance how the model can be modified to force one of the decomposition matrices to be real-valued, yielding an algorithm termed real-complex matrix factorisation (RCMF). In this way, it facilitates the application of nonnegativity constraints in $\mathbb{C}$ and also caters for the improperness of complex-valued sources.

Simulations illustrate the benefits of applying a widely linear model in a real-world setting. The general augmented framework is used to model a mixture of acoustic sources and, in contrast to standard CMF, allows an accurate estimation of the noncircular mixing arrangement - crucial in analysis.

2. COMPLEX MATRIX FACTORISATION

Existing complex matrix factorisation algorithms represent a straightforward extension of the real-valued algorithm. Given a complex-valued data matrix, $\mathbf{Y} \in \mathbb{C}^{I \times T}$, complex matrix factorisation (CMF) operates by finding two complex matrices $\mathbf{A} \in \mathbb{C}^{I \times J}$ and $\mathbf{X} \in \mathbb{C}^{J \times T}$ where $J \leq \min(I, T)$ such that $\mathbf{Y}$ is factorised as well as possible, that is

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{E} \tag{1}$$

where $\mathbf{E} \in \mathbb{C}^{I \times T}$ denotes an approximation error. Matrix factorisation can be achieved by minimising the following cost function

$$\mathcal{J}_{\text{CMF}} = \frac{1}{2}\text{tr}(\mathbf{Y} - \mathbf{A}\mathbf{X})^H(\mathbf{Y} - \mathbf{A}\mathbf{X}) = \tag{2}$$
$$\frac{1}{2}\text{tr}(\mathbf{Y}^H\mathbf{Y} - (\mathbf{A}\mathbf{X})^H\mathbf{Y} - \mathbf{Y}^H\mathbf{A}\mathbf{X} + (\mathbf{A}\mathbf{X})^H(\mathbf{A}\mathbf{X}))$$

where $\text{tr}(\cdot)$ denotes the trace operator and $(\cdot)^H$ the Hermitian transpose operator. A standard approach is to perform alternating minimisation with respect to matrices $\mathbf{A}$ and $\mathbf{X}$, each time optimising one set of parameters while keeping the other one fixed; this is known as the Alternating Least Squares (ALS) approach [11]. The stationary points for each matrix-wise optimisation procedure can be found by equating the gradient components, $\nabla_{\mathbf{A}}\mathcal{J}_{\text{CMF}}$ and $\nabla_{\mathbf{X}}\mathcal{J}_{\text{CMF}}$, to zero.

As the cost function in (2) is a real function of complex variables, the complex gradients with respect to $\mathbf{A}$ and $\mathbf{X}$ can be defined based on the conjugate gradient and $\mathbb{CR}$ calculus[1],

[1]For more details concerning $\mathbb{CR}$ calculus, see [10, 12].

that is

$$\nabla_{\mathbf{A}}\mathcal{J}_{\text{CMF}} = \frac{\partial \mathcal{J}_{\text{CMF}}}{\partial \mathbf{A}^*} \text{ and } \nabla_{\mathbf{X}}\mathcal{J}_{\text{CMF}} = \frac{\partial \mathcal{J}_{\text{CMF}}}{\partial \mathbf{X}^*}. \quad (3)$$

This way, gradient evaluation with respect to $\mathbf{A}$ gives

$$\frac{\partial}{\partial \mathbf{A}^*}\text{tr}((\mathbf{AX})^H\mathbf{Y}) = \mathbf{YX}^H \quad (4)$$

$$\frac{\partial}{\partial \mathbf{A}^*}\text{tr}((\mathbf{AX})^H\mathbf{AX}) = \mathbf{AXX}^H \quad (5)$$

whereas the gradients for the other terms in (2) are zero. Finally, the complex gradient with respect to $\mathbf{A}$ is given by

$$\nabla_{\mathbf{A}}\mathcal{J}_{\text{CMF}} = -\mathbf{YX}^H + \mathbf{AXX}^H \quad (6)$$

In a similar fashion, it can be shown that the gradient with respect to $\mathbf{X}$ is given by

$$\nabla_{\mathbf{X}}\mathcal{J}_{\text{CMF}} = -\mathbf{A}^H\mathbf{Y} + \mathbf{A}^H\mathbf{AX} \quad (7)$$

Setting the gradients in (6) and (7) to zero, the complex ALS updates become

$$\begin{aligned} \mathbf{A} &\leftarrow [\mathbf{YX}^H(\mathbf{XX}^H)^{-1}] \\ \mathbf{X} &\leftarrow [(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H\mathbf{Y}] \end{aligned} \quad (8)$$

3. CIRCULARITY

For a complex random variable $\mathbf{z}$, the covariance matrix is given by $\mathcal{C}_{\mathbf{zz}} = E[\mathbf{zz}^H]$ and is normally used in second-order statistical signal processing. However, it has been shown [8, 9, 10] that complex statistics are not a straightforward extension of real-valued statistics and to cater fully for all the second-order statistical information of $\mathbf{z}$ it is necessary also to consider the so-called pseudo-covariance matrix given by $\mathcal{P}_{\mathbf{zz}} = E[\mathbf{zz}^T]$.

Consider a random variable $z = z_r + jz_i$. Its pseudo-covariance is given by $E[zz^T] = E[z_r^2] - E[z_i^2] + 2jE[z_r z_i]$ and vanishes for uncorrelated real and imaginary parts of equal variances. Signals for which the pseudo-covariance is zero are called second-order circular or proper. In practice however, due to short observation windows, anisotropic noises, unequal powers of data channels and reflections, it is only natural to assume and $E[zz^T] \neq 0$. To cater for noncircularity, work [8, 9] has proposed a widely linear framework in which the signal model is based on the augmented form of the complex variable given by $\widehat{\mathbf{z}} = [\mathbf{z}^T; \mathbf{z}^H]^T$. Then, the augmented covariance matrix $\mathcal{C}_{\widehat{\mathbf{z}}\widehat{\mathbf{z}}}$ is given by

$$\mathbf{C}_{\widehat{\mathbf{z}}\widehat{\mathbf{z}}} = \begin{bmatrix} \mathbf{z} \\ \mathbf{z}^* \end{bmatrix} [\mathbf{z}^H \mathbf{z}^T] = \begin{bmatrix} \mathcal{C}_{\mathbf{zz}} & \mathcal{P}_{\mathbf{zz}} \\ \mathcal{C}^*_{\mathbf{zz}} & \mathcal{P}^*_{\mathbf{zz}} \end{bmatrix} \quad (9)$$

where $(\cdot)^*$ denotes the complex conjugate, and contains information from both the covariance and pseudo-covariance matrices. Augmented statistics have been employed successfully in, for example, the design of adaptive filters [10] where it was shown that using a widely linear model given by

$$y = \mathbf{a}^T\mathbf{z} + \mathbf{b}^T\mathbf{z}^*, \quad (10)$$

it is possible to cater for general complex processes (both second-order circular and noncircular).

4. AUGMENTED CMF

A problem with standard complex matrix factorisation is that it makes no provision for noncircularity. To that end, we propose augmented complex matrix factorisation (ACMF) which operates within a widely linear framework. Given a complex-valued data matrix, $\mathbf{Y} \in \mathbb{C}^{I\times T}$, ACMF calculates three factor matrices $\mathbf{A} \in \mathbb{C}^{I\times J}$, $\mathbf{B} \in \mathbb{C}^{I\times J}$ and $\mathbf{X} \in \mathbb{C}^{J\times T}$ where $J \leq \min(I, T)$ such that $\mathbf{Y}$ is factorised as well as possible by the widely linear model, that is

$$\mathbf{Y} = \mathbf{AX} + \mathbf{BX}^* + \mathbf{E} \quad (11)$$

Then, factorisation is performed by minimizing

$$\begin{aligned} \mathcal{J}_{\text{ACMF}} &= \frac{1}{2}\text{tr}(\mathbf{Y} - \mathbf{AX} - \mathbf{BX}^*)^H(\mathbf{Y} - \mathbf{AX} - \mathbf{BX}^*) = \\ &\frac{1}{2}\text{tr}(\mathbf{Y}^H\mathbf{Y} - \mathbf{X}^H\mathbf{A}^H\mathbf{Y} - \mathbf{X}^T\mathbf{B}^H\mathbf{Y} - \mathbf{Y}^H\mathbf{AX} + \mathbf{X}^H\mathbf{A}^H\mathbf{AX} \\ &+ \mathbf{X}^T\mathbf{B}^H\mathbf{AX} - \mathbf{Y}^H\mathbf{BX}^* + \mathbf{X}^H\mathbf{A}^H\mathbf{BX}^* + \mathbf{X}^T\mathbf{B}^H\mathbf{BX}^*) \end{aligned} \quad (12)$$

Applying similar analysis to before, it is straightforward to show that the gradient with respect to the factor $\mathbf{A}$ is given by

$$\nabla_{\mathbf{A}}\mathcal{J}_{\text{ACMF}} = -\mathbf{YX}^H + \mathbf{AXX}^H + \mathbf{BX}^*\mathbf{X}^H \quad (13)$$

and that the gradient with respect to $\mathbf{B}$ is given by

$$\nabla_{\mathbf{B}}\mathcal{J}_{\text{ACMF}} = -\mathbf{YX}^T + \mathbf{AXX}^T + \mathbf{BX}^*\mathbf{X}^T \quad (14)$$

Evaluation of the gradient with respect to $\mathbf{X}$ gives the non-zero terms

$$\begin{aligned} \frac{\partial}{\partial \mathbf{X}^*}\text{tr}(\mathbf{X}^H\mathbf{A}^H\mathbf{Y}) &= \mathbf{A}^H\mathbf{Y} \\ \frac{\partial}{\partial \mathbf{X}^*}\text{tr}(\mathbf{X}^H\mathbf{A}^H\mathbf{AX}) &= \mathbf{A}^H\mathbf{AX} \\ \frac{\partial}{\partial \mathbf{X}^*}\text{tr}(\mathbf{Y}^H\mathbf{BX}^*) &= \mathbf{B}^T\mathbf{Y}^* \\ \frac{\partial}{\partial \mathbf{X}^*}\text{tr}(\mathbf{X}^H\mathbf{A}^H\mathbf{BX}^*) &= (\mathbf{A}^H\mathbf{B} + (\mathbf{A}^H\mathbf{B})^T)\mathbf{X}^* \end{aligned} \quad (15)$$

Their combination yields

$$\begin{aligned} &\nabla_{\mathbf{X}}\mathcal{J}_{\text{ACMF}} = \\ &-(\mathbf{A}^H\mathbf{Y} + \mathbf{B}^T\mathbf{Y}^*) + \mathbf{A}^H(\mathbf{AX} + \mathbf{BX}^*) + \mathbf{B}^T(\mathbf{B}^*\mathbf{X} + \mathbf{A}^*\mathbf{X}^*) \end{aligned} \quad (16)$$

which can be expressed in a compact form as

$$\nabla_{\mathbf{X}}\mathcal{J}_{\text{ACMF}} = -([\mathbf{A}^H \;\; \mathbf{B}^T]\widehat{\mathbf{Y}}) + [\mathbf{A}^H \;\; \mathbf{B}^T]\mathbf{C}\widehat{\mathbf{X}} \quad (17)$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{bmatrix}, \quad \widehat{\mathbf{Y}} = \begin{bmatrix} \mathbf{Y} \\ \mathbf{Y}^* \end{bmatrix}, \quad \widehat{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{X}^* \end{bmatrix}$$

Finally, setting the gradients in (13), (14) and (17) to zero, the augmented complex ALS updates are given by

$$\begin{aligned} \mathbf{A} &\leftarrow [(\mathbf{YX}^H - \mathbf{BX}^*\mathbf{X}^H)(\mathbf{XX}^H)^{-1}] \\ \mathbf{B} &\leftarrow [(\mathbf{YX}^T - \mathbf{AXX}^T)(\mathbf{X}^*\mathbf{X}^T)^{-1}] \\ \widehat{\mathbf{X}} &\leftarrow [(\mathbf{C}^H\mathbf{C})^{-1}\mathbf{C}^H\widehat{\mathbf{Y}}] \end{aligned} \quad (18)$$

5. REAL-COMPLEX MATRIX FACTORISATION

Most practical applications of matrix factorisation are based on prior knowledge of the factors, such as their nonnegativity. Since this assumption is not possible to implement in the complex domain directly, we propose a modification of the widely linear framework that supports nonnegativity by constraining the general ACMF model given in (11) such that $\mathbf{A} = \mathbf{B}$, that is

$$\mathbf{Y} = \mathbf{AX}+\mathbf{AX}^*+\mathbf{E} = \mathbf{A}(\mathbf{X}+\mathbf{X}^*)+\mathbf{E} = \mathbf{AX}_{\Re}+\mathbf{E} \quad (19)$$

where $\mathbf{X}_{\Re} = 2\Re\{\mathbf{X}\}$ and real-valued. The factorisation matrices, $\mathbf{A} \in \mathbb{C}^{I\times J}$ and $\mathbf{X}_{\Re} \in \mathbb{R}^{J\times T}$ where $J \leq \min(I,T)$, are now obtained by minimising

$$\mathcal{J}_{\text{RCMF}} = \frac{1}{2}\text{tr}(\mathbf{Y}-\mathbf{AX}-\mathbf{AX}^*)^H(\mathbf{Y}-\mathbf{AX}-\mathbf{AX}^*) \quad (20)$$

with the respective gradients

$$\begin{aligned}\nabla_{\mathbf{X}}\mathcal{J}_{\text{RCMF}} =& -\mathbf{A}^H\mathbf{Y} - \mathbf{A}^T\mathbf{Y}^* + \mathbf{A}^H\mathbf{AX} + \mathbf{A}^T\mathbf{A}^*\mathbf{X}\\ &+ (\mathbf{A}^H\mathbf{A} + (\mathbf{A}^H\mathbf{A})^T)\mathbf{X}^*\\ =& -(\mathbf{A}^H\mathbf{Y}+\mathbf{A}^T\mathbf{Y}^*) + \mathbf{A}^H\mathbf{A}(\mathbf{X}+\mathbf{X}^*)\\ &+ \mathbf{A}^T\mathbf{A}^*(\mathbf{X}+\mathbf{X}^*)\\ =& -\widehat{\mathbf{A}}^H\widehat{\mathbf{Y}} + \widehat{\mathbf{A}}^H\widehat{\mathbf{A}}\mathbf{X}_{\Re} \end{aligned} \quad (21)$$

where $\widehat{\mathbf{A}} = [\mathbf{A}; \mathbf{A}^*]$, and

$$\begin{aligned}\nabla_{\mathbf{A}}\mathcal{J}_{\text{RCMF}} =& \mathbf{YX}^H + \mathbf{AXX}^H + \mathbf{AX}^*\mathbf{X}^H\\ &- \mathbf{YX}^T + \mathbf{AXX}^T + \mathbf{AX}^*\mathbf{X}^T\\ =& -\mathbf{YX}_{\Re}^T + \mathbf{AX}_{\Re}\mathbf{X}_{\Re}^T \end{aligned} \quad (22)$$

Setting the gradients in (21) and (22) equal to zero gives the ALS updates for the real-complex matrix factorisation (RCMF) in the form

$$\begin{aligned}\mathbf{A} &\leftarrow [\mathbf{YX}_{\Re}{}^T(\mathbf{X}_{\Re}\mathbf{X}_{\Re}{}^T)^{-1}]\\ \mathbf{X}_{\Re} &\leftarrow [(\widehat{\mathbf{A}}^H\widehat{\mathbf{A}})^{-1}\widehat{\mathbf{A}}^H\widehat{\mathbf{Y}}]\end{aligned} \quad (23)$$

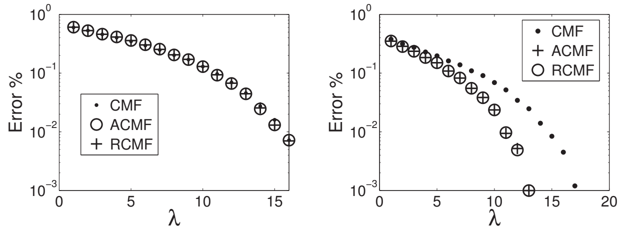
As $\mathbf{X}_{\Re}$ is real-valued and $\mathbf{A}$ can be written as $|\mathbf{A}|e^{j\phi}$, where ϕ is a $I \times J$ matrix that contains the phase information of $\mathbf{A}$, nonnegative constraints can now be placed on $\mathbf{X}_{\Re}$, together with normalisation and sparsity constraints on both $\mathbf{X}_{\Re}$ and on $|\mathbf{A}|$.

6. EXPERIMENTS

Simulations were first conducted on synthetic data to illustrate the benefits of using ACMF and RCMF for improper complex sources. In the second set of simulations, the convenience of using ACMF for modelling acoustic mixtures is shown.

6.1. Synthetic Data

The performances of CMF, ACMF and RCMF were compared for factorising complex constellations, a typical communications scenario. An example of a circular constellation arrangement contains the elements $[1+j, 1-j, -1+j, -1-j]$ and a noncircular one contains the elements $[1+j, -1-j]$. For each of these two arrangements a 100×100 matrix was created containing constellation values selected in a random fashion. With the constellation matrix as an input, $\mathbf{Y}$, the CMF, ACMF and RCMF algorithms were applied to obtain factorisations for different values of λ. The parameter λ controls the dimensions of the factorisation matrices for the algorithms as $J = 2\lambda$ for ACMF, $J = 3\lambda$ for CMF and $J = 4\lambda$ for RCMF. This ensures an equal number of factorisation elements and a fair performance comparison.



(a) Factorisation performance for circular arrangement (b) Factorisation performance for noncircular arrangement

Fig. 1. Factorisation performances for complex constellations using CMF, ACMF and RCMF

The error rates, that is the percentage of constellation points estimated incorrectly after synthesis of the input, are shown respectively in Fig. 1(a) and Fig. 1(b) for the circular and noncircular constellation arrangements. Conforming with the analysis, in the case of the circular constellation the performances of the algorithms were equivalent, while in the case of the noncircular constellation the performances of the ACMF and RCMF algorithms outperformed CMF.

6.2. Modelling Acoustic Mixtures

Two speakers were placed approximately 30cm apart, facing a line of 16 microphones[2], with each speaker playing a single sound stimulus only. Stimulus A was a sinusoid of frequency 441Hz and stimulus B was a sinusoid of frequency 1764Hz. Thus the pair of speakers can be regarded as a single complex source, with A and B as its real and imaginary parts respectively. To obtain a suitable complex representation for the recordings[3], each neighbouring pair of microphones were treated as the real and imaginary parts of a single complex observation, thus giving 8 complex observations which formed the rows of the observation matrix $\mathbf{Y}$.

Both the ACMF and CMF algorithms were used to factorise $\mathbf{Y}$ with J equal to the number of sources ($J = 1$) to obtain physically meaningful results. To mitigate for the problem of uniqueness associated with matrix factorisation[4], the algorithms were combined with an additional sparsity con-

[2]The distance between each microphone was 3cm and the distance between the microphone line and the speakers was approximately 40cm.

[3]Recordings were made at a sampling frequency of 44100Hz.

[4]Several local minima exist for the cost functions of both CMF and ACMF.

straint. The 'correct' source estimates were those therefore which maximised

$$\mathcal{J}_{\text{sparsity}} = -\sum_i \log(1 + \mathrm{P}_{\Re(\mathbf{X})}(i)) - \sum_i \log(1 + \mathrm{P}_{\Im(\mathbf{X})}(i))$$

where $\mathrm{P}_{\Re(\mathbf{X})}$ and $\mathrm{P}_{\Im(\mathbf{X})}$ denote respectively the power spectra for the real and imaginary parts of the estimated source matrix $\mathbf{X}$.

Both algorithms were capable of estimating the underlying source (acoustic stimuli). The estimate using ACMF ($\mathbf{X}_{\text{ACMF}}$) is shown in Fig. 2 where the real and imaginary parts reflect stimuli A and B respectively. Given that the estimate for the source was correct using both the CMF and ACMF algorithms, the synthesised observations (that is $\mathbf{Y}_{\text{CMF}} = \mathbf{A}_{\text{CMF}}\mathbf{X}_{\text{CMF}}$ in the case of CMF and $\mathbf{Y}_{\text{ACMF}} = \mathbf{A}_{\text{ACMF}}\mathbf{X}_{\text{ACMF}} + \mathbf{B}_{\text{ACMF}}\mathbf{X}^*_{\text{ACMF}}$ in the case of ACMF) were compared with the original observation $\mathbf{Y}$ to illustrate their performance in modelling the source mixture. It was found that CMF could not synthesise $\mathbf{Y}$ accurately, indicating an inadequate mixing model. This is evident, for example, in Fig. 3 which shows the observation of the fifth microphone in the top panel, or the real part of the third complex observation (real part of row 3 of matrix $\mathbf{Y}$), and the corresponding synthesis using CMF and ACMF in the middle and lower panels. Note the increase in accuracy using the ACMF model. This result is consistent with synthesised estimates for all other rows in $\mathbf{Y}$.

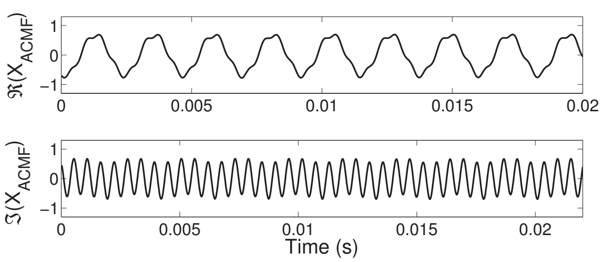


Fig. 2. Real and imaginary parts of source estimate (sound stimuli at 441Hz and 1764Hz) using ACMF ($\mathbf{X}_{\text{ACMF}}$)

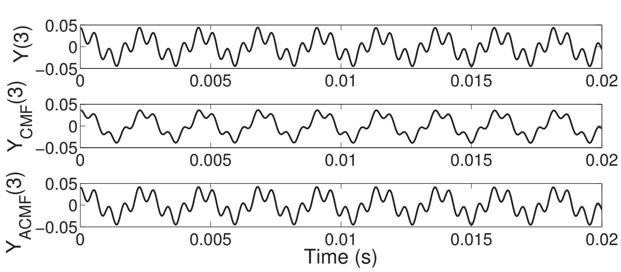


Fig. 3. Real part of the third row of the observation matrix (top) and its synthesised estimate using CMF (middle) and ACMF (bottom)

7. CONCLUSIONS

Two novel complex-valued factorisation algorithms that operate within the widely linear framework have been presented which, in the role of blind source separation, provide a more robust model when the mixing arrangement is noncircular. For insight, one of the algorithms constrains one of the factors to be real-valued allowing nonnegative constraints to enhance the uniqueness of the operation and to help generate more physically meaningful results.

8. REFERENCES

[1] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*, Wiley, 2009.

[2] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[3] P. Smaragdis and J. C. Brown, "Non-negative matrix factorization for polyphonic music transcription," in *2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2003, pp. 177–180.

[4] R. Schluter and H. Ney, "Using phase spectrum information for improved speech recognition performance," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2001, vol. 1, pp. 133–136.

[5] G. Shi, M. M. Shanechi, and P. Aarabi, "On the importance of phase in human speech recognition," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 5, pp. 1867–1874, 2006.

[6] P. Aarabi et. al, *Phase-Based Speech Processing*, World Scientific, 2005.

[7] H. Kameoka, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF: A new sparse representation for acoustic signals," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 3437–3440.

[8] P. J. Schreier and L. L. Scharf, "Second-order analysis of improper complex random vectors and processes," *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 714–725, 2003.

[9] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Transactions on Signal Processing*, vol. 43, no. 8, pp. 2030–2033, 1995.

[10] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*, Wiley, 2009.

[11] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*, Wiley, 2002.

[12] K. Kreutz-Delgado, "The complex gradient operator and the $\mathbb{CR}$-calculus," in *Deptartment of Electrical and Computer Engineering, UC San Diego, Course Lecture Supplement No. ECE275A*, 2006, pp. 1–74.