Theorem 2: If $x(t)$ is a real-valued signal, $M_1 = (a_1, b_1, c_1, d_1)$, $M_2 = (a_2, b_2, c_2, d_2)$, then

$$\Delta u_{M_1}^2 \Delta u_{M_2}^2 \geq \left(a_1 a_2 \Delta t^2 + \frac{b_1 b_2}{4\Delta t^2} \right)^2 + \frac{(a_1 b_2 - a_2 b_1)^2}{4} \tag{12}$$

and the equality is achieved iff $x(t) = (1/\pi\sigma^2)^{1/4} \exp(-((t - t_0)^2/2\sigma^2))$, where σ is an arbitrary real constant.

Proof: With the results of Lemma 2 and 3, we can obtain

$$\begin{aligned}\Delta u_M^2 &= \frac{1}{E} \int_{-\infty}^{+\infty} |u X_M(u)|^2 \, du - u_{M0}^2 \\ &= \frac{1}{E} \left[a^2 \int_{-\infty}^{+\infty} |t x(t)|^2 \, dt + b^2 \int_{-\infty}^{\infty} |\omega X(\omega)|^2 \, d\omega \right. \\ &\qquad \left. + jab(I^* - I) \right] - a^2 t_0^2 - b^2 \omega_0^2 - 2ab t_0 \omega_0 \\ &= a^2 \Delta t^2 + b^2 \Delta\omega^2 + \frac{jab(I^* - I)}{E} - 2ab t_0 \omega_0 \\ &= a^2 \Delta t^2 + b^2 \Delta\omega^2 \end{aligned} \tag{13}$$

because $x(t)$ is real, which means $\omega_0 = 0$ and I is real. Therefore, using (13) and the uncertainty relation in the FT domain, the spread in any LCT domain for a real signal is lower bounded by its spreads in the time and frequency domains, i.e.,

$$\Delta u_M^2 \geq a^2 \Delta t^2 + \frac{b^2}{4\Delta t^2}. \tag{14}$$

For arbitrary two LCT domains we have

$$\begin{aligned}\Delta u_{M_1}^2 \Delta u_{M_2}^2 &\geq \left[a_1^2 \Delta t^2 + \frac{b_1^2}{4\Delta t^2} \right] \cdot \left[a_2^2 \Delta t^2 + \frac{b_2^2}{4\Delta t^2} \right] \\ &= (\Delta t^2)^2 a_1^2 a_2^2 + \frac{b_1^2 b_2^2}{16(\Delta t^2)^2} + \frac{a_1^2 b_2^2 + b_1^2 a_2^2}{4} \\ &= \left(a_1 a_2 \Delta t^2 + \frac{b_1 b_2}{4\Delta t^2} \right)^2 + \frac{(a_1 b_2 - a_2 b_1)^2}{4}\end{aligned}$$

and the equality is achieved iff $x(t)$ is a Gaussian signal.

IV. Conclusion

In this correspondence, we discuss the uncertainty relations in the LCT domain. A lower bound for complex signals in two LCT domains is derived, which can be achieved by a complex chirp signal with Gaussian envelope. Moreover, the tighter lower bound for real signals in two LCT domains given in [15] is also proven to hold for arbitrary LCT parameters based on the properties of moments in the LCT domain. The uncertainty principle in the FrFT domain is a special case of the achieved results.

References

[1] M. Moshinsky and C. Quesne, "Linear canonical transformations and their unitary representations," *J. Math. Phys.*, vol. 12, pp. 1772–1783, Aug. 1971.

[2] L. M. Bernardo, "ABCD matrix formalism of fractional Fourier optics," *Opt. Eng.*, vol. 35, no. 3, pp. 732–740, Mar. 1996.

[3] S. C. Pei and J. J. Ding, "Eigenfunctions of linear canonical transform," *IEEE Trans. Signal Processing*, vol. 50, no. 1, pp. 11–26, 2002.

[4] R. Tao, L. Qi, and Y. Wang, *Theory and Applications of the Fractional Fourier Transform*. Beijing: Tsinghua University Press, 2004.

[5] H. M. Ozaktas, M. A. Kutay, and Z. Zalevsky, *The Fractional Fourier Transform With Applications in Optics and Signal Processing*. New York: Wiley, 2000.

[6] L. B. Almeida, "The fractional Fourier transform and time-frequency representations," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3084–3091, 1994.

[7] B. Barshan, M. A. Kutay, and H. M. Ozaktas, "Optimal filters with linear canonical transformations," *Opt. Commun.*, vol. 135, pp. 32–36, 1997.

[8] K. K. Sharma and S. D. Joshi, "Signal separation using linear canonical and fractional Fourier transforms," *Opt. Commun.*, vol. 265, pp. 454–460, 2006.

[9] D. Mustard, "Uncertainty principle invariant under fractional Fourier transform," *J. Austral. Math. Soc. Ser. B*, vol. 33, pp. 180–191, 1991.

[10] H. M. Ozaktas and O. Aytur, "Fractional Fourier domains," *Signal Processing*, vol. 46, pp. 119–124, 1995.

[11] T. Alieva and M. J. Bastiaans, "On fractional Fourier transform moments," *IEEE Signal Processing Letters*, vol. 7, no. 11, pp. 320–323, 2000.

[12] S. Shinde and V. M. Gadre, "An uncertainty principle for real signals in the fractional Fourier transform domain," *IEEE Trans. Signal Processing*, vol. 49, no. 11, pp. 2545–2548, 2001.

[13] C. Capus and K. Brown, "Short-time fractional Fourier methods for the time-frequency representation of chirp signals," *J. Acoust. Soc. Am*, vol. 113, no. 6, pp. 3253–3263, 2003.

[14] A. Stern, "Uncertainty principles in linear canonical transform domains and some of their implications in optics," *J. Opt. Soc. Am. A*, vol. 25, no. 3, pp. 647–652, 2008.

[15] K. K. Sharma and S. D. Joshi, "Uncertainty principles for real signals in linear canonical transform domains," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2677–2683, 2008.

A Multiplicative Algorithm for Convolutive Non-Negative Matrix Factorization Based on Squared Euclidean Distance

Wenwu Wang, Andrzej Cichocki, and Jonathon A. Chambers

***Abstract*—Using the convolutive nonnegative matrix factorization (NMF) model due to Smaragdis, we develop a novel algorithm for matrix decomposition based on the squared Euclidean distance criterion. The algorithm features new formally derived learning rules and an efficient update for the reconstructed nonnegative matrix. Performance comparisons in terms of computational load and audio onset detection accuracy indicate the advantage of the Euclidean distance criterion over the Kullback–Leibler divergence criterion.**

***Index Terms*—Audio object separation, convolutive nonnegative matrix factorization, multiplicative algorithm, squared Euclidean distance.**

I. Introduction

Non-negative matrix factorization (NMF), an emerging technique for data analysis [1], [2], has found many potentially useful appli-

Manuscript received August 28, 2008; accepted January 27, 2009. First published March 10, 2009; current version published June 17, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Mark J. Coates. This work was presented in part at the IEEE International Symposium on Signal Processing and Information Technology, Cairo, Egypt, 2007.

W. Wang is with the Centre for Vision, Speech and Signal Processing, Department of Electronic Engineering, University of Surrey, Guildford, GU2 7XH, U.K. (e-mail: w.wang@surrey.ac.uk).

A. Cichocki is with the Laboratory for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, Saitama, 351-0198, Japan (e-mail: cia@brain.riken.jp).

J. A. Chambers is with the Advanced Signal Processing Group, Department of Electronic and Electrical Engineering, Loughborough University, Loughborough, Leics LE11 3TU, U.K. (e-mail: j.a.chambers@lboro.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2016881

cations in signal and image processing, e.g., [3]–[19]. For example, in audio signal processing, based on spectrogram factorization, NMF has been applied to music transcription [5], [9] and audio source separation [16]–[19]. The standard NMF model given in [1] has been shown to be satisfactory and sufficient in certain tasks provided that the spectral frequencies of the analyzed audio signal do not change dramatically over time, which is however not the case for many realistic audio signals. As a result, the single basis obtained via the standard NMF decomposition may not be adequate to capture the temporal dependency of the frequency patterns within the signal. Moreover, a single basis function is typically required for the representation of each note of a given instrument in music audio, and therefore a clustering step needs to be used for source separation of instruments playing melodies [20], [19] and [18]. However, as identified by [18], it may be difficult to perform a reliable clustering in many situations.

To overcome these issues, the approaches of convolutive NMF (or similar methods called shifted NMF) have been introduced in [6]–[8], [13], [18], and [19]. As a common characteristic of these approaches, the data to be analyzed are modelled as a linear combination of a group of shifted matrices. However, the developed learning algorithms have different characteristics due to the various strategies adopted in their derivation or different applications being addressed. For example, a multiplicative learning algorithm has been developed in [6] and [7] for the adaptation of a criterion based on the Kullback–Leibler (KL) divergence, and no restrictions are enforced on the frequency resolution of the spectrogram. In [18], translated versions of a single basis function are used to represent the typical frequency spectrum of any notes belonging to a single music source, which however requires the spectrogram to be logarithmic in the frequency scale. In [19], the learning algorithms are developed based on explicit constraints of temporal continuity and sparseness of the signals.

The convolutive learning rules developed in [6] and [7] are essentially an extension of the multiplicative rules in [1] for the minimization of the KL error norm. However, no formal mathematical derivations are given therein. Moreover, although there are learning algorithms for the standard NMF [1] based on the Euclidean distance, few such algorithms have been proposed for the convolutive case. The aim of this correspondence is to address these issues. To this end, we formally derive a multiplicative algorithm for the gradient adaptation of the error norm measured by the squared Euclidean distance. In addition, we provide an efficient procedure for the update of the reconstructed data matrix at each iteration, which can considerably reduce the computational load required by existing algorithms in the literature. The original idea of this work has been presented in [10].

We have applied our proposed algorithm to the audio object separation problem (i.e., the detection of repeating patterns, or note events from music audio signals), which is one of the core issues underlying many applications in audio engineering such as music transcription, audio information retrieval, low bit-rate audio coding, and automatic auditory scene analysis. In order to show its general performance, we have compared the proposed algorithm with the methods in [6] and [11], respectively, in this application context. Our experimental results reveal its superior performance to that of the two benchmark methods, especially in terms of its computational efficiency and note detection accuracy.

The remainder of the correspondence is organized as follows. The next section briefly reviews the standard NMF and Smaragdis' original work on convolutive NMF. The proposed convolutive NMF algorithm is described in detail in Section III. Section IV investigates the performance of the proposed algorithm using numerical experiments, and conclusions are drawn in Section V.

II. Preliminaries

A. Standard NMF

Given an $M \times N$ nonnegative matrix $\mathbf{X} \in \mathbb{R}_+^{M \times N}$, the goal of NMF is to find nonnegative matrices $\mathbf{W} \in \mathbb{R}_+^{M \times R}$ and $\mathbf{H} \in \mathbb{R}_+^{R \times N}$, such that $\mathbf{X} \approx \mathbf{WH}$, where R is the rank of the factorization, generally chosen to be smaller than M (or N), or akin to $(M+N)R < MN$, which results in the extraction of some latent features whilst reducing some redundancies in the input data. To find $\mathbf{W}$ and $\mathbf{H}$, several error functions have been proposed [1]–[4], one of which, denoted $\mathcal{L}(\cdot)$, is based on the squared Euclidean distance

$$(\hat{\mathbf{W}}, \hat{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \mathcal{L}(\mathbf{W}, \mathbf{H}) = \arg\min_{\mathbf{W},\mathbf{H}} \frac{1}{2} \left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_F^2 \qquad (1)$$

where $\hat{\mathbf{W}}$ and $\hat{\mathbf{H}}$ are the estimated optimal values of $\mathbf{W}$ and $\mathbf{H}$, $\|\cdot\|_F$ denotes the Frobenius norm, and $\hat{\mathbf{X}}$ is given by $\hat{\mathbf{X}} = \mathbf{WH}$. Alternatively, we can also minimize the error function based on the extended KL divergence

$$(\hat{\mathbf{W}}, \hat{\mathbf{H}}) = \arg\min_{\mathbf{W},\mathbf{H}} \sum_{m=1}^{M} \sum_{n=1}^{N} \mathbf{D}_{m,n} \qquad (2)$$

where $\mathbf{D}_{m,n}$ is the m, nth element of the matrix $\mathbf{D}$ given by $\mathbf{D} = \mathbf{X} \odot \log[\mathbf{X} \oslash \hat{\mathbf{X}}] - \mathbf{X} + \hat{\mathbf{X}}$, where $\odot$ and $\oslash$ denote the Hadamard (element-wise) product and division, respectively. We are particularly interested in the multiplicative algorithm of Lee and Seung [1], [2] due to its simplicity. In matrix form, the algorithm for minimizing criterion (1) can be written as

$$\mathbf{H}^{q+1} = \mathbf{H}^q \odot ((\mathbf{W}^q)^T \mathbf{X}) \oslash ((\mathbf{W}^q)^T \mathbf{W}^q \mathbf{H}^q) \qquad (3)$$

$$\mathbf{W}^{q+1} = \mathbf{W}^q \odot (\mathbf{X} (\mathbf{H}^{q+1})^T) \oslash (\mathbf{W}^q \mathbf{H}^{q+1} (\mathbf{H}^{q+1})^T) \qquad (4)$$

where q is the iteration index, and $(\cdot)^T$ is the matrix transpose operator. These rules are easy to implement. In addition, a step-size parameter which is normally required for gradient algorithms [4], is not necessary in these rules.

B. Convolutive NMF

To take into account the potential dependency between the neighboring columns of the input data matrix $\mathbf{X}$, the standard (instantaneous) NMF can be extended to a convolutive model [6]

$$\hat{\mathbf{X}} = \sum_{p=0}^{P-1} \mathbf{W}(p) \overset{p\rightarrow}{\mathbf{H}} \qquad (5)$$

where $\mathbf{W}(p) \in \mathbb{R}_+^{M \times R}$, $p = 0, \ldots, P-1$, are a set of basis matrices, $\mathbf{H} \in \mathbb{R}_+^{R \times N}$ is a weighting matrix, and $\overset{p\rightarrow}{\mathbf{H}}$ shifts the columns of $\mathbf{H}$ by p spots to the right, with the columns shifted in from outside the matrix set to zero. Effectively, $\hat{\mathbf{X}}$ is now expressed as a sum of shifted matrix products. These column shifts are noncircular. Analogously, $\overset{\leftarrow p}{\mathbf{H}}$ shifts the columns of $\mathbf{H}$ by p spots to the left. These notations will also be used for the shifting operations of other matrices throughout the work. Note that, $\overset{0\rightarrow}{\mathbf{H}} = \overset{\leftarrow 0}{\mathbf{H}} = \mathbf{H}$. With the convolutive model, the temporal continuity possessed by many audio signals can be expressed more effectively in the time-frequency domain, especially for those signals whose frequencies vary with time [6].

To find a decomposition with the form of (5), Smaragdis has further proposed multiplicative learning rules based on the extended KL divergence (2), which can be rewritten as

$$\mathbf{H}^{q+1} = \mathbf{H}^q \odot (((\mathbf{W}^q(p))^T \overset{\leftarrow p}{\tilde{\mathbf{X}}^q}) \oslash ((\mathbf{W}^q(p))^T \mathbf{\Xi})) \qquad (6)$$

$$\mathbf{W}^{q+1}(p) = \mathbf{W}^q(p) \odot ((\tilde{\mathbf{X}}^q (\overset{p\rightarrow}{\mathbf{H}^{q+1}})^T) \oslash (\mathbf{\Xi} (\overset{p\rightarrow}{\mathbf{H}^{q+1}})^T)) \qquad (7)$$

where $\mathbf{\Xi}$ is an $M \times N$ matrix whose elements are all set to unity, $\tilde{\mathbf{X}}^q = \mathbf{X} \oslash \hat{\mathbf{X}}^q$.

C. Motivations and Contributions of Our Work

Two technical issues in [6] are worth, however, further investigation. First, no formal mathematical derivation for (6) and (7) was provided. As a result, the theoretical analysis of its convergence performance may become difficult to achieve. Second, no learning rules were developed for convolutive NMF under the criterion (1). In this correspondence, we develop a novel algorithm by formally deriving the learning rules using the criterion (1). After comparisons with the KL divergence based criterion, we have observed that the Euclidean distance based criterion provides better performance for convolutive NMF in terms of computational load and onset detection accuracy. To the best of our knowledge, it is probably the first study to address the performance difference between algorithms based on the two criteria in this context. Another contribution in our work is using efficient recursions in the adaptation of the reconstructed nonnegative matrix for further reducing the computational load of our algorithm, which can be readily applied to a category of similar algorithms. Note, that similar symbolic notations as those in [6] are used here for matrix shifts and convolutions, simply for making easy comparison between the algorithms.

III. A Novel Convolutive NMF Algorithm

A. Derivation of the Learning Rules

We first derive the update rules for $\mathbf{W}(p)$. According to (5), we have the derivative $\hat{\mathbf{X}}_{i,j}$ with respect to (w.r.t) $\mathbf{W}_{m,n}(p)$,

$$\frac{\partial \hat{\mathbf{X}}_{i,j}}{\partial \mathbf{W}_{m,n}(p)} = \frac{\partial \sum_p \sum_d \mathbf{W}_{i,d}(p) \overset{p\rightarrow}{\mathbf{H}}_{d,j}}{\partial \mathbf{W}_{m,n}(p)}$$
$$= \frac{\partial \sum_d \mathbf{W}_{i,d}(p) \overset{p\rightarrow}{\mathbf{H}}_{d,j}}{\partial \mathbf{W}_{m,n}(p)} = \delta_{i,m} \overset{p\rightarrow}{\mathbf{H}}_{n,j} \tag{8}$$

where $\delta_{i,m}$ is the Kronecker delta, and the right-bottom subscripts represent the element indices of a matrix, e.g., $\hat{\mathbf{X}}_{i,j}$ is the i,jth element of the matrix $\hat{\mathbf{X}}$. Similarly, according to (1) and (8), we have the derivative of $\mathcal{L}$ w.r.t $\mathbf{W}_{m,n}(p)$,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_{m,n}(p)} = \sum_i \sum_j (\hat{\mathbf{X}}_{i,j} - \mathbf{X}_{i,j}) \frac{\partial \hat{\mathbf{X}}_{i,j}}{\partial \mathbf{W}_{m,n}(p)}$$
$$= \sum_j (\hat{\mathbf{X}}_{m,j} - \mathbf{X}_{m,j}) \overset{p\rightarrow}{\mathbf{H}}_{n,j}. \tag{9}$$

Let the element-wise step-size[1] be

$$[\mu_{\mathbf{W}(p)}]_{m,n} = \frac{\mathbf{W}_{m,n}(p)}{\sum_j \overset{p\rightarrow}{\mathbf{H}}_{n,j} \hat{\mathbf{X}}_{m,j}}. \tag{10}$$

Using (9) and (10), we can derive the adaptation equation of $\mathbf{W}_{m,n}(p)$ as

$$\mathbf{W}_{m,n}(p) = \mathbf{W}_{m,n}(p) - [\mu_{\mathbf{W}(p)}]_{m,n} \frac{\partial \mathcal{L}}{\partial \mathbf{W}_{m,n}(p)} \tag{11}$$
$$= \mathbf{W}_{m,n}(p)$$
$$- \frac{\mathbf{W}_{m,n}(p)}{\sum_j \overset{p\rightarrow}{\mathbf{H}}_{n,j} \hat{\mathbf{X}}_{m,j}} \sum_j (\hat{\mathbf{X}}_{m,j} - \mathbf{X}_{m,j}) \overset{p\rightarrow}{\mathbf{H}}_{n,j} \tag{12}$$
$$= \mathbf{W}_{m,n}(p) \frac{(\mathbf{X}(\overset{p\rightarrow}{\mathbf{H}})^T)_{m,n}}{(\hat{\mathbf{X}}(\overset{p\rightarrow}{\mathbf{H}})^T)_{m,n}}. \tag{13}$$

In matrix form, the update (13) can be written as (marked with iteration number q)

$$\mathbf{W}^{q+1}(p) = \mathbf{W}^q(p) \odot ((\mathbf{X}(\overset{p\rightarrow}{\mathbf{H}^q})^T) \oslash (\hat{\mathbf{X}}^q (\overset{p\rightarrow}{\mathbf{H}^q})^T)) \tag{14}$$

where $p = 0, \ldots, P-1$. Similarly, using the gradient

$$\frac{\partial \mathcal{L}}{\partial \mathbf{H}_{m,n}} = \sum_j \mathbf{W}_{j,m}(p) (\overset{\leftarrow p}{\hat{\mathbf{X}}}_{j,n} - \overset{\leftarrow p}{\mathbf{X}}_{j,n}) \tag{15}$$

we can obtain the update equation for $\mathbf{H}$ in matrix form as

$$\mathbf{H}^{q+1} = \mathbf{H}^q \odot (((\mathbf{W}^{q+1}(p))^T \overset{\leftarrow p}{\mathbf{X}}) \oslash ((\mathbf{W}^{q+1}(p))^T \overset{\leftarrow p}{\hat{\mathbf{X}}^q})). \tag{16}$$

It is worth noting that (6) and (7) can be obtained using a similar derivation method and the KL divergence.

B. Further Practical Improvements

1) Update of $\mathbf{H}$*:* It can be observed from (14) and (16) that $P+1$ matrices in total, i.e., a set P of $\mathbf{W}(p)$, $p = 0, \ldots, P-1$, requires to be updated first, followed by $\mathbf{H}$, at each iteration. Specifically, $\mathbf{H}$ is updated using $\mathbf{W}(P-1)$, which is the last update in the P loop for updating $\mathbf{W}(p)$. As a result, the update of $\mathbf{H}$ can be dominated by $\mathbf{W}(P-1)$. In order to mitigate this effect, as suggested in [6], we can first update all $\mathbf{W}(p)$, $p = 0, \ldots, P-1$, and then take the average of all the updates for $\mathbf{H}$, that is

$$\mathbf{H}^{q+1} = \frac{1}{P} \sum_{p=0}^{P-1} \mathbf{H}^q(p) \tag{17}$$

where $\mathbf{H}^q(p)$ is given by

$$\mathbf{H}^q(p) = \mathbf{H}^q \odot (\mathbf{W}^{q+1}(p)^T \overset{\leftarrow p}{\mathbf{X}}) \oslash (\mathbf{W}^{q+1}(p)^T \overset{\leftarrow p}{\hat{\mathbf{X}}^q}). \tag{18}$$

In contrast to (16), the update (17) also exploits the information from $\mathbf{W}(0), \ldots, \mathbf{W}(P-2)$ [see (18)], which has the practical advantage of reducing the dominant effect of $\mathbf{W}(P-1)$ for the update of $\mathbf{H}^{q+1}$ in (16). Note in this correspondence, that (17) is an intuitive operation which is obtained empirically, rather than justified theoretically.

2) Update of $\hat{\mathbf{X}}$*:* From (16) and (14), it is clear that the updates of $\mathbf{H}$ and $\mathbf{W}(p)$ both rely on the update of $\hat{\mathbf{X}}$, which, on the other hand, depends on the instantaneous values of $\mathbf{W}(p)$ and $\mathbf{H}$, according to (5). This means that $\hat{\mathbf{X}}$ should be updated correspondingly once for each $\mathbf{W}(p)$ update. Nevertheless, updating the whole (5) is computationally demanding if only an individual $\mathbf{W}(p)$ has a new value. Therefore, instead of directly using (5), we use the following simpler formulation

$$\hat{\mathbf{X}}^q = \hat{\mathbf{X}}^q - \mathbf{W}^q(p) \overset{p\rightarrow}{\mathbf{H}^q} + \mathbf{W}^{q+1}(p) \overset{p\rightarrow}{\mathbf{H}^q} \quad (p = 0, \ldots, P-1) \tag{19}$$

where $\hat{\mathbf{X}}^q$ is updated to accommodate the new values of each $\mathbf{W}(p)$ (inside the P loops), and the initial value of $\hat{\mathbf{X}}^q$ $(q > 1)$ in the right-hand side (RHS) of (19) is obtained at the end of the $(q-1)$th iteration (outside the P loops), when the recursions are completed. For $q = 1$, $\hat{\mathbf{X}}^q$ in the RHS of (19) is still calculated via (5). In practice, we found that the nonnegative property of $\hat{\mathbf{X}}^q$ may not be guaranteed, due to the subtraction operation and small numerical errors. The small negative values can be prevented by using the following projection:

$$\hat{\mathbf{X}}^q_{i,j} = \max(\epsilon, \hat{\mathbf{X}}^q_{i,j}) \tag{20}$$

[1] Note that, the selection of such a learning rate follows from the similar rescaling operation used in [2], and the convergence of our proposed algorithm under such a learning rate can be proved similarly for the convolutive model.

TABLE I
SUMMARY OF THE PROPOSED ALGORITHM: CONVNMF-ED

1) Generate $\mathbf{X}$ from the audio data. Initialize P, $\mathbf{W}(p)$, $\mathbf{H}(p)$, ζ and Q (i.e. the predetermined iteration number). Calculate $\hat{\mathbf{X}}$ using (5), update (1), and set $q = 0$. Run steps 2-6.
2) Set $q = q + 1$, $p = 0$, $\mathbf{H}_{sum} = \mathbf{0}$, where $\mathbf{0} \in \mathbb{R}_{+}^{R \times N}$, i.e., a matrix with each element equal to zero.
3) Set $p = p + 1$. Calculate $\mathbf{W}^q(p)$ using (14). Update $\hat{\mathbf{X}}^q$ using (19) and (20).
4) Calculate $\mathbf{H}^q(p)$ using (18), and $\mathbf{H}_{sum} = \mathbf{H}_{sum} + \mathbf{H}^q(p)$. If $p < P$, return to step 3, otherwise, go to step 5.
5) Calculate $\mathbf{H}^q$ as $\mathbf{H}_{sum}/P$, i.e., in terms of (17). Normalize $\mathbf{H}^q$ and $\mathbf{W}^q(p)$.
6) Update $\hat{\mathbf{X}}^q$ using (5), update (1) and (21). If (21) is satisfied or $q \geq Q$, stop iterations and output $\mathbf{W}^o(p)$ and $\mathbf{H}^o$, otherwise, return to step 2.

where $\hat{\mathbf{X}}_{i,j}^q$ is the i, jth element of the matrix $\hat{\mathbf{X}}$ at the iteration q, $\max(\cdot)$ takes the maximum value of its arguments, and ϵ is a trivial constant, typically, $\epsilon = 10^{-9}$ in our implementation. The algorithm stops iterations when the following criterion is satisfied:

$$\frac{\left\| \hat{\mathbf{X}}^{q+1} - \hat{\mathbf{X}}^{q} \right\|_F}{\left\| \hat{\mathbf{X}}^{q} \right\|_F} < \zeta \tag{21}$$

where ζ is a small constant.

C. Nonnegative Decomposition of Magnitude Spectra

In our problem, the nonnegative matrix $\mathbf{X}$ is generated as the magnitude spectra of the input audio data. For a signal with L samples, using a T-point short-term Fourier transform (STFT), it can be segmented to K frames, where $K = \lfloor (L - T)/\delta \rfloor$, δ is the time shift between the adjacent windows and $\lfloor \cdot \rfloor$ is an operator taking the maximum integer no greater than its argument. Concatenating the absolute value of the spectrum of each frame, we can generate $\mathbf{X}$ with a dimension of $(T/2 + 1) \times K$ (see details in [9]). The proposed algorithm can be summarized in Table I, where $\mathbf{W}(p)$ and $\mathbf{H}(p)$ are typically initialized with random nonnegative elements. For convenience, we name it as ConvNMF-ED (i.e., convolutive NMF based on squared Euclidean distance).

Upon convergence of the algorithm, $\mathbf{X}$ is decomposed into the convolution of P nonnegative matrices, denoted as $\mathbf{W}^o(p) \in \mathbb{R}_{+}^{(T/2+1) \times R}$ and $\mathbf{H}^o \in \mathbb{R}_{+}^{R \times K}$, i.e., the corresponding local optimum values of $\mathbf{W}(p)$ and $\mathbf{H}$, respectively. And $\mathbf{H}^o$ contains the bases of the temporal patterns while $\mathbf{W}^o(p)$ contains the frequency patterns of the original signal. All the set of P $\mathbf{W}^o(p)$ matrices together contain both frequency and temporal information of the time-frequency patterns (i.e., audio objects) of the original audio signal.

D. Comparisons to Existing Methods

If $P = 1$, ConvNMF-ED reduces to a standard NMF algorithm similar to that represented by (3) and (4) with the difference in the update of $\hat{\mathbf{X}}^q$. If $P > 1$, its computational load is approximately P times that of the standard NMF. ConvNMF-ED is different from Smaragdis' algorithm [6] in steps 3–4 (see Table I), where $\mathbf{W}^q(p)$, $\hat{\mathbf{X}}^q$ and $\mathbf{H}^q(p)$ in [6] are instead updated by (7), (5) and (6), respectively. As a result, compared with [6], ConvNMF-ED requires $(2RP - 4R + P + 2)MN$ fewer element operations (multiplications and additions) in each iteration. We also consider a faster version of Smaragdis' algorithm [6], denoted as ConvNMF-KL, which is similar to [6] except that $\hat{\mathbf{X}}^q$ is updated using (19) and (20), instead of (5). Apart from its higher computational efficiency, ConvNMF-KL has very similar performance to [6]. Therefore, it is also used as an alternative to [6] in our simulations. Although the algorithm in [7] uses the same learning rules as those in [6], it is a supervised speech separation approach using additional constraints from the application domain. Making direct comparisons to [7] is not straightforward; therefore, it is not explored further in this work.

SNMF2D [11], [12] considers a two-dimensional deconvolution scheme, together with sparseness constraints. We will examine the most related learning rules based on the least squares (LS) criterion in [12], i.e., SNMF2D-LS. In contrast, SNMF2D-LS uses the shifted versions of $\mathbf{W}^q$ and $\mathbf{H}^q$ at all time lags $p = 0, \ldots, P - 1$, for updating $\mathbf{W}^q(p)$ and $\mathbf{H}^q(p)$ with an individual time lag at each iteration. For example, according to [11], the update of $\mathbf{H}$ in (16) may be written as

$$\begin{aligned}\mathbf{H}^{q+1} = \mathbf{H}^{q} \odot & [(\mathbf{W}^{q+1}(0)^{T} \overset{\leftarrow 0}{\mathbf{X}} + \cdots + \mathbf{W}^{q+1}(P-1)^{T} \overset{\leftarrow P-1}{\mathbf{X}})] \\ & \oslash [(\mathbf{W}^{q+1}(0)^{T} \overset{\leftarrow 0}{\hat{\mathbf{X}}^{q}}) + \cdots + (\mathbf{W}^{q+1}(P-1)^{T} \overset{\leftarrow P-1}{\hat{\mathbf{X}}^{q}})].\end{aligned} \tag{22}$$

The resulting representation of $\mathbf{W}^o(p)$ and $\mathbf{H}^o$ using SNMF2D-LS has actually broken the structure of audio objects, i.e., the time-frequency signature in the spectrogram has been shifted more than it actually requires. As a result, it becomes difficult to detect the event or onset directly from $\mathbf{W}^o(p)$ and $\mathbf{H}^o$ (as shown in our experiments in Section IV). Furthermore, SNMF2D-LS is computationally much more expensive, as compared with both ConvNMF-ED and ConvNMF-KL. The time required for computing (22) would be approximately $P - 1$ times than that for computing (16). These observations will be further confirmed in numerical experiments.

IV. NUMERICAL EXPERIMENTS

In this section, we study numerically the performance of ConvNMF-ED in the context of audio object detection,[2] and perform comparisons with ConvNMF-KL, SNMF2D-LS, and Smaragdis' algorithm [6].

A. Music Audio

Two music audio signals with each containing repeating musical notes G4 and A3 played by a guitar are mixed together. The mixed signal is approximately 6.8 s sampled at $f_s = 22050$ Hz. Note, that for illustrative purpose, the signals used in this section are relatively simple, however, realistic audio signals have also been tested in this work, see, e.g., Section IV-D. Some parameters used in the experiments are set as: $T = 4096$, $P = 105$, $R = 2$, $\zeta = 0.0001$. The described algorithms ConvNMF-ED and ConvNMF-KL, together with SNMF2D-LS,[3] were all applied to decompose $\mathbf{X}$, where $\mathbf{W}(p)$ and $\mathbf{H}$ are the absolute values of random matrices with elements drawn from a

[2] Numerical examples for applying the proposed algorithm to artificial audio signals containing audio patterns whose frequency changes linearly with time have been shown in [10], where we have shown the failure of the standard NMF for such a scenario, and the advantage of our proposed algorithm.

[3] The MATLAB code of the SNMF2D algorithm was downloaded from Morup's webpage [14]. In our evaluations, the 2-D deconvolution is reduced to one dimension by setting $\phi = 0$. The convolutional frame length τ was held identical to P used in ConvNMF-KL and ConvNMF-ED in all experiments.

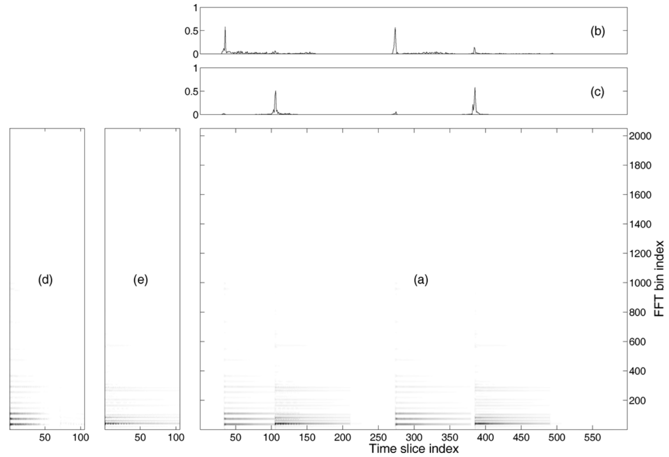

Fig. 1. Decomposition result of $\mathbf{X}$ using the proposed ConvNMF-ED algorithm. (a) is the plot of magnitude spectrum matrix $\mathbf{X}$ of the music audio signal. (b) and (c) are the plots of the rows of the factorized $\mathbf{H}^o$. (d) and (e) are the plots of the columns of the factorized $\mathbf{W}^o(p)$ as a collection for all $p = 0, \ldots, 104$, where (d) represents time-frequency signature of note G4, while (e) denotes note A3.

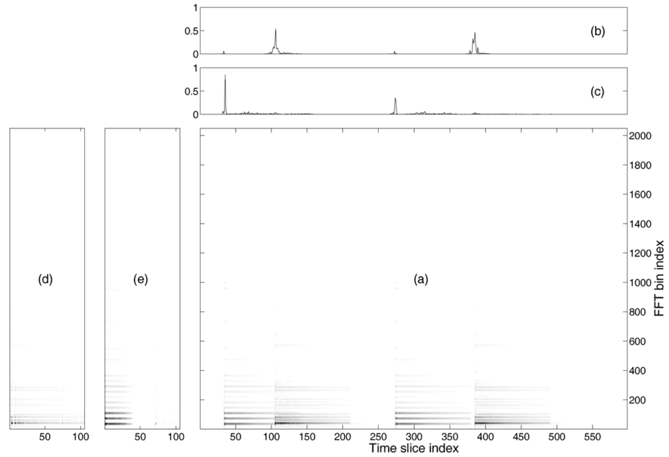

Fig. 2. Decomposition result of $\mathbf{X}$ using the ConvNMF-KL algorithm. (a) is the plot of magnitude spectrum matrix $\mathbf{X}$ of the music audio signal. (b) and (c) are the plots of the rows of the factorized $\mathbf{H}^o$. (d) and (e) are the plots of the columns of the factorized $\mathbf{W}^o(p)$ as a collection for all $p = 0, \ldots, 104$, where (d) represents the time-frequency signature of note A3, while (e) denotes note G4. Note that, the decomposed $\mathbf{W}^o$ and $\mathbf{H}^o$ by ConvNMF-KL differ from those by the ConvNMF-ED algorithm with a permutation ambiguity, which is inherent due to the signal model considered, although the ambiguity does not always occur in many of our random tests.

standardized Gaussian probability density function. All tests were run on a computer whose CPU speed is 1.8 GHz.

Figs. 1–3 show the decomposition results of these algorithms, respectively. In all these algorithms, plot (a) is the spectrogram matrix $\mathbf{X}$. Plots (b) and (c) visualize the first and second row of $\mathbf{H}^o$, and plots (d) and (e) visualize the first and second column of $\mathbf{W}^o(p)$ for the collection of p. We can observe from these figures that the performance of ConvNMF-ED is approximately identical to that of ConvNMF-KL, except for a difference in the permutation. The notes G4 and A3 were separated correctly by both algorithms, where $\mathbf{W}^o(p)$ represents the magnitude spectrum of the notes, while $\mathbf{H}^o$ denotes the onset locations of the notes. However, the SNMF2D-LS algorithm does not correctly identify the note events. For example, it can be seen from Fig. 3(b) and (c) that, after the convergence of the algorithm, the onset locations in $\mathbf{H}^o$ have been actually *over-shifted*, i.e., shifted more than is required. As a consequence, $\mathbf{H}^o$ does not reveal the correct onset locations. Moreover, the magnitude spectrum of the notes described by $\mathbf{W}^o(p)$ is less accurate than obtained by ConvNMF-KL and ConvNMF-ED.

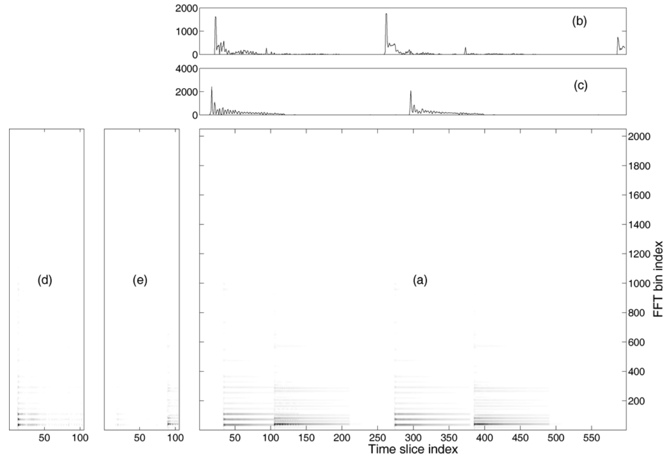

Fig. 3. Decomposition result of $\mathbf{X}$ using the algorithm SNMF2D-LS. (a) is the plot of magnitude spectrum matrix $\mathbf{X}$ of the music audio signal. (b) and (c) are the plots of the rows of the factorized $\mathbf{H}^o$. (d) and (e) are the plots of the columns of the factorized $\mathbf{W}^o(p)$ as a collection for all $p = 0, \ldots, 104$, where (d) represents the time-frequency signature of note G4, while (e) denotes note A3.

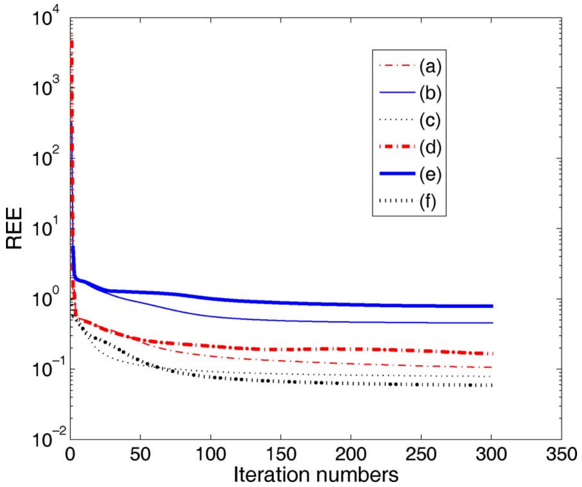

Fig. 4. Convergence comparison between the algorithms ConvNMF-ED (plots (a) and (d)), ConvNMF-KL (plots (b) and (e)), and SNMF2D-LS (plots (c) and (f)). Two random tests were performed for each algorithm with $T = 256$ (plots (a) (b) and (c)), and 1024 (plots (d) (e) (f)) respectively.

TABLE II

Comparison of Computing Time Required for the Algorithms Running 100 Iterations. For Each T, the Time Consumed (in Seconds) Was Averaged Over 50 Random Tests

T	256	512	1024	204	4096
ConvNMF-ED	250	584	1102	2043	4952
ConvNMF-KL	262	607	1231	2424	5800
SNMF2D-LS	694	1094	2079	3841	9212
Algorithm [6]	673	1109	2115	3912	10545

B. Convergence Performance

In this section, we compare numerically the convergence performance of the three algorithms. We perform two random tests for each

TABLE III
COMPARISON OF TP/FP FOR NOTE ONSET DETECTION BETWEEN THE THREE ALGORITHMS. FOR EACH T, THE RESULT WAS AVERAGED OVER 50 RANDOM REALIZATIONS. THE VALUES OF TP ARE ALSO SHOWN IN BRACKETS.

T	256	512	1024	2048
ConvNMF-ED	13.49 (93.1%)	3.46 (77.6%)	3.56 (78.1%)	3.39 (77.2%)
ConvNMF-KL	3.95 (79.1%)	3.34 (70.1%)	2.45 (71.0%)	0.92 (47.8%)
SNMF2D-LS	0.06 (5.7%)	0.04 (4.0%)	0.04 (3.9%)	0.04 (3.4%)

algorithm with T set to 256 and 1024, respectively, and run each algorithm for 300 iterations. The evolution of the relative estimation error (REE) versus iteration number is drawn in Fig. 4, where REE is defined as

$$\mathrm{REE} = \frac{\left\| \mathbf{X} - \hat{\mathbf{X}} \right\|_F}{\|\mathbf{X}\|_F}. \tag{23}$$

This performance index is relatively less sensitive to the signal dynamics as compared with the absolute estimation error due to the adopted normalization. All the three algorithms were initialized randomly with the same starting points. The three algorithms have very similar convergence behavior, with SNMF2D-LS having lowest REE, followed by ConvNMF-ED and ConvNMF-KL.

C. Computational Load

To compare the computational complexity of the proposed algorithm with that of ConvNMF-KL, SNMF2D-LS and the algorithm in [6], we ran each algorithm 50 times for each T, where T was set to be 256, 512, 1024, 2048, and 4096, respectively. We measured the computing time required for 100 iterations. All the algorithms were initialized randomly. Other parameters remained the same as those in Section IV-A. The average results over 50 tests for each algorithm are listed in Table II. More specifically, if we further average the results over different T, the reductions of computing load of ConvNMF-ED as compared with ConvNMF-KL, SNMF2D-LS and Smaragdis' algorithm in [6] are respectively 13%, 47%, and 51%. Clearly, the proposed algorithm ConvNMF-ED is the most efficient algorithm. When T becomes larger or the signal to be analyzed becomes longer, the advantage of ConvNMF-ED becomes more significant. This numerical result supports our theoretical analysis results in Section III-D.

D. Audio Object Separation Accuracy

In order to evaluate the convolutive NMF algorithms more objectively, it can be useful to measure the accuracy of the detected note events from $\mathbf{H}^o$ or $\mathbf{W}^o(p)$ that is obtained using those algorithms. Here, we use the same peak-picking algorithm described [9] to detect the note onsets from $\mathbf{H}^o$. The threshold used for peak-picking was set to 0.4 and held constant for all the tests. We evaluate the performance using the ratio between the percentage of true positives (i.e., the number of correct detections relative to that of total existing onsets, denoted as TP for brevity) and the percentage of the false positives (i.e., the number of erroneous onsets relative to that of the total detected onsets, denoted as FP for brevity) [21], i.e., TP/FP. A higher value of TP/FP indicates a relatively better performance. A detected note is considered to be a true positive if it falls into one analysis window within the original onset. Otherwise, it is considered as a false positive. In practice, there may exist a few missing notes that are not detected at all, which nevertheless does not affect the accuracy of evaluation using TP/FP. A subset of a commercial music audio database was used for the evaluation, for which 25 testing signals, each containing various numbers of notes, were used [15]. For each signal, T was set to be 256, 512, 1024, 2048, and 4096, respectively, and 50 random realizations were run for each T. Table III shows the average results of TP/FP, with TP also listed in the brackets. It can be seen from these figures that SNMF2D-LS almost totally fails for onset detections of note events in the signals. This is not surprising regarding the discussion we had earlier in Sections III-D and IV-A. In contrast to both SNMF2D-LS and ConvNMF-KL, the proposed algorithm ConvNMF-ED provides the better decompositions that are particularly suitable for note onset detection.

V. CONCLUSION

A new multiplicative learning algorithm for convolutive NMF has been presented. The algorithm features novel learning rules derived from the squared Euclidean distance, together with an efficient method for computing the estimate of the low-rank approximation. The proposed algorithm has advantages over both Smaragdis' algorithm and the algorithm by Schmidt and Morup, in the context of audio object separation and note onset detection, in terms of the performance measurement of computational complexity and detection accuracy. The proposed algorithm can be a useful tool for a wide range of applications including the analysis of complex auditory scenes.

ACKNOWLEDGMENT

The authors would like to thank the associate editor and the anonymous reviewers for their constructive comments.

REFERENCES

[1] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[2] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing*. Cambridge, MA: MIT Press, 2001, vol. 13.

[3] P. O. Hoyer, "Non-negative matrix factorization with sparseness constraints," *J. Mach. Learn. Res.*, no. 5, pp. 1457–1469, 2004.

[4] A. Cichocki, R. Zdunek, and S. Amari, "Csiszar's divergences for non-negative matrix factorization: family of new algorithms," *Springer Lecture Notes in Comput. Sci.*, vol. 3889, pp. 32–39, 2006.

[5] P. Smaragdis and J. C. Brown, "Nonnegative matrix factorization for polyphonic music transcription," in *Proc. IEEE Int. Workshop Applications Signal Process. Audio Acoustics*, New Paltz, NY, Oct. 2003, pp. 177–180.

[6] P. Smaragdis, "Non-negative matrix factor deconvolution, extraction of multiple sound sources from monophonic inputs," in *Proc. 5th Int. Conf. Independent Component Analysis Blind Signal Separation*, Granada, Spain, Sep. 22–24, 2004, vol. LNCS 3195, Lecture Notes on Computer Science, pp. 494–499.

[7] P. Smaragdis, "Convolutive speech bases and their application to supervised speech separation," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 1, pp. 1–12, 2007.

[8] P. D. O'Grady and B. A. Pearlmutter, "Convolutive non-negative matrix factorisation with a sparseness constraint," in *Proc. IEEE Int. Workshop Machine Learning Signal Process.*, Maynooth, Ireland, Sep. 6–8, 2006, pp. 427–432.

[9] W. Wang, Y. Luo, S. Sanei, and J. A. Chambers, "Non-negative matrix factorization for note onset detection of audio signals," in *Proc. IEEE Int. Workshop Machine Learning Signal Process.*, Maynooth, Ireland, Sep. 6–8, 2006, pp. 447–452.

[10] W. Wang, "Squared Euclidean distance based convolutive non-negative matrix factorization with multiplicative learning rules for audio pattern separation," in *Proc. IEEE Int. Symp. Signal Process. Info. Tech.*, Cairo, Egypt, Dec. 15–18, 2007.

[11] M. N. Schmidt and M. Morup, "Nonnegative matrix factor 2-D deconvolution for blind single channel source separation," in *Proc. 6th Int. Conf. Independent Component Analysis Blind Signal Separation*, Charleston, SC, 2006, pp. 700–707.

[12] M. Morup and M. N. Schmidt, "Sparse non-negative matrix Factor 2-D deconvolution," Technical Univ., Denmark, 2006.

[13] M. Morup, K. H. Madsen, and L. K. Hansen, "Shifted non-negative matrix factorization," in *Proc. IEEE Int. Workshop Machine Learning Signal Process.*, Maynooth, Ireland, Sep. 6–8, 2007, pp. 427–432.

[14] M. Morup, 2007 [Online]. Available: http://www.mortenmorup.dk/index_files/Page368.htm

[15] W. Wang, 2008 [Online]. Available: http://personal.ee.surrey.ac.uk/Personal/W.Wang/demondata.html

[16] B. Wang and M. D. Plumbley, "Musical audio stream separation by non-negative matrix factorization," in *Proc. DMRN Summer Conf.*, Glasgow, Jul. 23–24, 2005.

[17] R. M. Parry and I. Essa, "Incorporating phase information for source separation via spectrogram factorization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Honolulu, HI, Apr. 15–20, 2007, vol. 2, pp. 661–664.

[18] D. FitzGerald, M. Cranitch, and E. Coyle, "Sound source separation using shifted non-negative tensor factorization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, 2006, vol. V, pp. 653–656.

[19] T. Virtanen, "Monaural sound source separation by non-negative matrix factorization with temporal continuity and sparseness criterion," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1066–1074, Mar. 2007.

[20] M. Casey and A. Westner, "Separation of mixed audio sources by independent subspace analysis," in *Proc. Int. Comput. Music Conf.*, Berlin, Germany, Aug. 2000, pp. 154–161.

[21] J. P. Bello, L. Daudet, S. A. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in musical signals," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 1035–1047, Sep. 2005.

OFDM Joint Data Detection and Phase Noise Cancellation for Constant Modulus Modulations

Yu Gong and Xia Hong

***Abstract*—This correspondence proposes a new algorithm for the OFDM joint data detection and phase noise (PHN) cancellation for constant modulus modulations. We highlight that it is important to address the overfitting problem since this is a major detrimental factor impairing the joint detection process. In order to attack the overfitting problem we propose an iterative approach based on minimum mean square prediction error (MMSPE) subject to the constraint that the estimated data symbols have constant power. The proposed constrained MMSPE algorithm (C-MMSPE) significantly improves the performance of existing approaches with little extra complexity being imposed. Simulation results are also given to verify the proposed algorithm.**

***Index Terms*—Constant modulus modulation, OFDM, phase noise cancellation.**

I. Introduction

The phase noise (PHN) in an orthogonal-frequency-division multiplexing (OFDM) system arises from the imperfections at the receiver's oscillator, damaging the orthogonality among subcarriers [1], [2]. A typical PHN consists of two parts: the common PHN and the random PHN [3]. Most existing PHN cancellation algorithms (e.g., [3] and [4]) mainly consider the common PHN which is an averaging effect over the OFDM transmission and can be mitigated with the help of pilot symbols. The random PHN, on the other hand, is much more difficult to handle as it varies from one symbol to another even for slowly fading channels. Recently, a family of algorithms for joint data detection and PHN cancellation based on the probabilistic approach of variational inference have been proposed [5].

In general, a joint estimation process may suffer from the "overfitting" problem so that the estimate is too close to the received samples and fits into the noise. The overfitting problem is particularly serious in the joint OFDM data detection and PHN cancellation. This is because for an OFDM symbol with N subcarriers, there are N data symbols and N PHN to be determined from $2N$ observations including N from the receiving data model and another N from the PHN model. On another front, the PHN must be mitigated at every symbol since it varies from one symbol to another. This describes a very special case of parameter estimation problem: unlike the classic parameter estimation whereby the estimates can be improved by increasing the number of data samples, here the number of unknown parameters (i.e., the symbols and PHN) always equals the number of the "observation" samples, making it particularly vulnerable to overfitting which thus must be carefully handled as otherwise the whole joint process may be invalidated.

We note that, although it was not explicitly identified, the algorithms described in [5] are in fact equivalent to the Bayesian regularization utilizing the Gaussian distributions as the priors, a common method to combat overfitting [6]. This, however, may not be sufficiently effective for the OFDM PHN cancellation. In our recent correspondence [7] and further in [8], we proposed a new joint data detection and PHN cancellation algorithm based on minimum mean square prediction error (MMSPE), where the hard decision is applied to the symbol estimates at the end of each iteration. The hard decision process can effectively filter the noise out of the symbol estimates and remove the associated uncertainties due to the overfitting which would otherwise be carried forward over the iterations. However, the hard decision imposes as a nonlinear constraint on the iterative procedure which may sometimes be too strong such that some symbol estimates are forced into the wrong direction over the iterations, resulting in performance loss. As will be shown in the simulation later in this correspondence, the MMSPE algorithm has close performance to, if not worse than, those proposed in [5] when the SNR is high (in which case there is little noise to be removed and the negative effect of the hard decision becomes more dominant). This motivates us to explore new methods to combat overfitting problem.

In this correspondence, we focus on the OFDM system with constant modulus modulations such as the PSK. Embedding the deterministic *a priori* information from the modulation that the data symbol must have constant power into the MMSPE cost function as a constraint, we propose a constrained MMSPE (C-MMSPE) algorithm to jointly detect the data symbol and cancel the PHN. The C-MMSPE algorithm can better handle the overfitting and has significantly superior performance to both the MMSPE algorithm and the algorithms described in [5]. The idea of using constant modulus has been well understood in many communications applications such as the Godard blind equalization [9], but this is the first time to be applied to the OFDM PHN cancellation. Although in general any prior information about the system, especially deterministic knowledge, should greatly help to improve the modelling performance, it usually leads to some complicated constrained optimization problems with large computational complexity. Luckily in the case of the OFDM PHN cancellation, we will show that the derived

Manuscript received August 26, 2008; accepted February 06, 2009. First version published March 21, 2009; current version published June 17, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Biao Chen.

The authors are with the School of Systems Engineering, The University of Reading, Reading, RG6 6AY, U.K. (e-mail: y.gong@reading.ac.uk; x.hong@reading.ac.uk).

Color versions of one or more of the figures in this correspondence are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2018362

1053-587X/$25.00 © 2009 IEEE