UNGROUNDED INDEPENDENT NON-NEGATIVE FACTOR ANALYSIS

Bhiksha Raj[1], *Kevin W. Wilson*[2], *Alexander Krueger*[3] *and Reinhold Haeb-Umbach*[3]

1. Carnegie Mellon University, Pittsburgh, PA, USA
2. Mitsubishi Electric Research Labs, Cambridge, MA, USA
3. University of Paderborn, Paderborn, Germany

ABSTRACT

We describe an algorithm that performs regularized non-negative matrix factorization (NMF) to find independent components in non-negative data. Previous techniques proposed for this purpose require the data to be grounded, with support that goes down to 0 along each dimension. In our work, this requirement is eliminated. Based on it, we present a technique to find a low-dimensional decomposition of spectrograms by casting it as a problem of discovering independent non-negative components from it. The algorithm itself is implemented as regularized non-negative matrix factorization (NMF). Unlike other ICA algorithms, this algorithm computes the *mixing* matrix rather than an unmixing matrix. This algorithm provides a better decomposition than standard NMF when the underlying sources are independent. It makes better use of additional observation streams than previous nonnegative ICA algorithms.

Index Terms— matrix decomposition, ICA

1. INTRODUCTION

The problem of finding low-dimensional non-negative decompositions of spectrograms has been an topic of increased interest in recent times. Such decompositions represent spectrograms as *non-cancelling* combinations of bases and are useful when the spectrogram of interest was generated as a sum of independent sources. They find use in a variety of applications such as signal characterization, denoising, signal separation etc. Most methods for computing these decompositions have been based on variants of non-negative matrix factorization (NMF) [1] or its statistical analogues [2].

One of the attractions of NMF-based representations is that they provide a representation by parts, where the observed data are explained as a constructive combination of nominally unrelated parts. However, NMF does not have any innate guarantee that the discovered bases will be independent, as a result of which the discovery of the right bases is often a result of chance, rather than method. Various modifications have therefore been proposed that impose constraints on the manner in which the bases are learned. Among these are techniques represent the data as sparse combinations of bases [3], those that decorrelate the contributions of the bases [4] or maximize the distance between the bases themselves [2]. None of these actually guarantee independence between the bases, however.

The problem of learning *independent* bases to explain non-negative data by construction can be formulated as a problem of independent component analysis (ICA) of non-negative mixtures. ICA aims to extract statistically independent components from observations through linear transformations. Given a collection of (column) vectors $\{\mathbf{v}\}$, which we can represent jointly as a matrix $\mathbf{V}$, ICA attempts to estimate an "unmixing" matrix $\mathbf{M}$ such that the the rows of $\mathbf{H} = \mathbf{MV}$, *i.e.* the components of the vectors $\mathbf{h}$ that form the columns of $\mathbf{H}$ are statistically independent. If the vectors $\mathbf{v}$ were themselves obtained through a linear operation on independent sources, *i.e.* if $\mathbf{V} = \mathbf{WX}$, where $\mathbf{X}$ is a matrix composed of vectors $\mathbf{x}$ with statistically independent components, then $\mathbf{M}$ will be a scaling and permutation of the inverse of the original matrix $\mathbf{W}$, *i.e.* $\mathbf{M} \approx \mathbf{RW}^{-1}$, where $\mathbf{R}$ is a scaling and permutation matrix, and $\mathbf{H} \approx \mathbf{RX}$. Alternately stated, the components of the observed vectors $\mathbf{v}$ are said to be *mixtures* of the original independent random variables represented in $\mathbf{x}$, where $\mathbf{W}$ is the *mixing matrix* that mixes the components of $\mathbf{x}$. ICA as it is normally performed [5, 6] aims to estimate an *unmixing* matrix $\mathbf{M}$ that can recover the original independent components (to within a permutation and scaling) from the mixed data in the observations.

The usual algorithms for ICA are agnostic to the polarity of the data. As a result, if $\mathbf{M}$ is a valid unmixing matrix, then $\mathbf{ZM}$ is also a solution, where $\mathbf{Z} = diag(1, 1, \ldots, -1, \cdots,)$ is a diagonal matrix where some diagonal terms are 1 and the rest are -1. When both the original data $\mathbf{X}$ and their mixed observations $\mathbf{V}$ are known *a priori* to be strictly non-negative then the solutions obtained may not be satisfactory, since $\mathbf{H}$ is not guaranteed to be non-negative. Exploiting nonnegativity constraint for both the signal and the mixing matrix can improve signal recovery in the presence of noise.

In [7] Plumbley presents a "non-negative ICA" algorithm that recovers non-negative independent components from mixtures of non-negative sources. Given a mixture of independent *well-grounded* sources $\mathbf{V}$, it can derive an unmixing matrix $\mathbf{M}$ such that the entries of $\mathbf{MV}$ are strictly non negative and its rows are independent. This solution however remains unsatisfactory for many reasons. The columns of the *mixing matrix* in the standard formalism ($\mathbf{V} = \mathbf{WH}$) represent the bases we referred to earlier in this section. The algorithm of [7] does not guarantee non-negativity of the mixing matrix. Secondly, there is a strong requirement for *grounding*, which implies that each basis has a non-zero probability of not being present (having 0 weight) in any vector $\mathbf{v}$. A DC bias is common in most naturally occurring data, which will make them ungrounded.

In [8] we presented an alternate mechanism that addressed at least one of these issues. We recast the problem of deriving independent non-negative bases for non-negative data as a regularized NMF based on the observation that if a mixed non-negative matrix can be expressed as the product of two non-negative matrices such that the rows of the one of them are decorrelated and grounded, then the rows of that matrix are also independent. In other words, if we were to simultaneously estimate a non-negative *mixing* matrix $\mathbf{W}$ and a matrix of non-negative uncorrelated vectors $\mathbf{H}$ such that $\mathbf{V} = \mathbf{WH}$ and $\mathbf{H}$ is grounded, the rows of $\mathbf{H}$ will also be independent.

Although the algorithm was shown to be effective, it still retained one of the deficiencies from [7] – the rows of $\mathbf{H}$ were required to be grounded. Thus, the algorithm was still unable to deal with data that had a DC offset.

In this paper we extend our algorithm to explicitly account for un-grounded data. It can be proved theoretically that independent components $\mathbf{W}$ can be derived for ungrounded data in the same manner as in [8] by shifting $\mathbf{V}$ to lie close to the origin, such that $\mathbf{W}^{+}\mathbf{V}_{\text{inf}}$ remains non-negative. Based on this theorem we develop a NMF-based algorithm for deriving independent components from generic non-negative data that has no restrictions of grounding. Maintaining the nomenclature of [8] we refer to this algorithm as "Ungrounded Independent Non-Negative Factor Analysis" or U-INFA.

Simulations show that U-INFA is able to estimate mixing matrices accurately, and results in estimates of unmixed independent components that are comparable (in terms of SNR) to or better than those obtained by other ICA algorithms, particularly in the presence of noise. When the data are not grounded, U-INFA is able to perform while other previous non-negative ICA algorithms fail. Further, when the mixing matrix $\mathbf{W}$ is not square (and has more rows than the number of independent sources) we achieve superior results to other ICA techniques in cases with significant additive noise.

2. UNCORRELATED GROUNDED NMF AS NON-NEGATIVE ICA

Before presenting the actual algorithm, we present two theorems that form the basis for the development of our algorithm. We briefly present an intuitive explanation for the theorems; the actual proofs (which will not fit in this paper) can be found in [9].

Theorem 1. *Let $\mathbf{V} = (V_1, ..., V_n)^T$ be a vector of n real-valued, non-negative random variables V_i, which is obtained as the product $\mathbf{V} = \mathbf{ZU}$ of a non-negative real-valued matrix $\mathbf{Z}$ and a vector $\mathbf{U} = (U_1, ..., U_n)^T$ of n non-negative, real valued,* **well-grounded** *and independent random variables U_i. Further, assume that V can be also expressed as the product $\mathbf{V} = \mathbf{WH}$, where $\mathbf{W}$ is an* **invertible**, *non-negative and real-valued matrix and $\mathbf{H} = (H_1, ..., H_n)^T$ is a vector of n non-negative, real-valued and* **uncorrelated** *random variables H_i.*

Then the components of $\mathbf{H}$ can be expressed as a non-negative, generalized permutation of the components of $\mathbf{U}$ and are also independent.

The intuition behind the proof of Theorem 1 is as follows: let $\mathbf{V}$ is obtained by non-negative transformation $\mathbf{W}$ of some matrix $\mathbf{H}$ such that the rows of $\mathbf{H}$ are independent and grounded, then it can be shown that it is not possible to find some other matrix $\mathbf{H}'$ with uncorrelated, but dependent rows that could also be transformed by a non-negative matrix $\mathbf{W}'$ into $\mathbf{V}$. For $\mathbf{H}'$ to exist, $\mathbf{WH} = \mathbf{W}'\mathbf{H}'$, *i.e.* $\mathbf{W}'^{+}\mathbf{WH} = \mathbf{H}'$, where $\mathbf{W}'^{+}$ is the pseudo inverse of $\mathbf{W}'$. However any transformation $\mathbf{W}'^{+}\mathbf{W}$ will necessarily rotate $\mathbf{H}$ such that at least some components become negative, thereby contradicting the hypothesis that $\mathbf{H}' = \mathbf{W}'^{+}\mathbf{WH}$ is non-negative, unless $\mathbf{W}'$ is a permutation (and scaling) of $\mathbf{W}$.

Theorem 2. *Let $\tilde{\mathbf{V}} = (\tilde{V}_1, ..., \tilde{V}_m)^T$ be a vector of m real-valued, non-negative random variables $\tilde{V}_i$, which is obtained as the product $\tilde{\mathbf{V}} = \mathbf{Z}\tilde{\mathbf{U}}$ of a non-negative real-valued $m \times n$ matrix $\mathbf{Z}$ and a vector $\tilde{\mathbf{U}} = (\tilde{U}_1, ..., \tilde{U}_n)^T$ of n non-negative, real valued and independent random variables $\tilde{U}_i$. Define the vector $\tilde{\mathbf{V}}_{inf}$ by $\tilde{\mathbf{V}}_{inf} := (\tilde{V}_{inf,1}, ..., \tilde{V}_{inf,m})^T$. Further assume that $\mathbf{V} := \tilde{\mathbf{V}} - \tilde{\mathbf{V}}_{inf}$ can be expressed as the product $\mathbf{V} = \mathbf{WH}$, where $\mathbf{W}$ is a non-negative and real-valued $m \times n$ matrix with rank$(\mathbf{W}) = n$ and $\mathbf{H} = (H_1, ..., H_n)^T$ is a vector of n non-negative, real-valued and* **uncorrelated** *random variables H_i.*

Then there exists a non-negative, generalized permutation matrix $\mathbf{P}$ such that $\mathbf{H} = \mathbf{PU}$, where $\mathbf{U} := \tilde{\mathbf{U}} - \tilde{\mathbf{U}}_{inf}$ with $\tilde{\mathbf{U}}_{inf} = (\tilde{U}_{inf,1}, ..., \tilde{U}_{inf,n})^T$ Particularly, the components of $\mathbf{H}$ are independent. Further, the product $\mathbf{P}\tilde{\mathbf{U}}$ can be computed by $\mathbf{P}\tilde{\mathbf{U}} = \mathbf{H} + \mathbf{W}^{+}\tilde{\mathbf{V}}_{inf}$, where $\mathbf{W}^{+}$ is the pseudo inverse of $\mathbf{W}$.

Here the subscript "inf" represents the infimum. The proof for Theorem 2. actually follows directly from the fact that if the rows of $\mathbf{H}$ are independent, $\mathbf{V}_{\text{inf}} = \mathbf{WH}_{\text{inf}}$.

3. THE UNGROUNDED INFA ALGORITHM

The import of the two theorems in Section 2 is that in order to derive non-negative independent factors from a matrix $\mathbf{V}$, it is sufficient to obtain the non-negative factorization $\mathbf{V} - \mathbf{V}_{\text{inf}} = \mathbf{WH}$, subject to the constraint that the rows of $\mathbf{H}$ are uncorrelated and $\mathbf{H}^{+}$ is non-negative. This means that after a preprocessing step to "artificially" ground the observations, we can apply our previous INFA algorithm [8] to these preprocessed observations.

Starting from ungrounded observations $\tilde{\mathbf{V}}$, the first step is thus to generate artificially grounded observations $\mathbf{V} = \tilde{\mathbf{V}} - \tilde{\mathbf{V}}_{\text{inf}}$ using the element-wise minimums of the ungrounded observations. We then apply the INFA algorithm, which we review below, to these artificially grounded observations.

We seek $\mathbf{W}$ and $\mathbf{H}$ such that

$$\begin{aligned} \mathbf{V} &= \mathbf{WH} \\ W_{ab} &\geq 0 \ \forall\, a, b \\ H_{bc} &\geq 0 \ \forall\, b, c \\ V_{ac} &\geq 0 \ \forall\, a, c \\ P(H_{ic}H_{jc}) &= P(H_{ic})P(H_{jc}) \ \forall\, i, j, c \end{aligned} \quad (1)$$

where W_{ab} and H_{bc} are components of $\mathbf{W}$ and $\mathbf{H}$ respectively. The fifth condition in Equation 1 expresses independence of the rows of $\mathbf{H}$. We solve the above as a problem of regularized non-negative matrix factorization. The general form of the NMF update with regularization on $\mathbf{H}$ from [10] is:

$$\begin{aligned} W_{ab} &\leftarrow W_{ab} \frac{[\mathbf{VH}^{\mathsf{T}}]_{ab}}{[\mathbf{WHH}^{\mathsf{T}}]_{ab}} \\ H_{bc} &\leftarrow H_{bc} \frac{\left[[\mathbf{W}^{\mathsf{T}}\mathbf{V}]_{bc} - \alpha\varphi(H_{bc})\right]_{\varepsilon}}{[\mathbf{W}^{\mathsf{T}}\mathbf{WH}]_{bc} + \varepsilon} \end{aligned} \quad (2)$$

where ε is a small positive constant and $[\,]_{\varepsilon}$ indicates that any values within the brackets less than ε should be replaced with ε to prevent violations of the nonnegativity constraint. $\varphi(H)$ is the gradient of $J(\mathbf{H})$ with respect to H.

For our problem,

$$\varphi(H_{bc}) = \frac{\partial J(\mathbf{H})}{\partial H_{bc}} \quad (3)$$

$$= \sum_i \sum_j C_{ij} \frac{\partial C_{ij}}{\partial H_{bc}} \quad (4)$$

It is the straightforward to show that $\partial C_{ij}/\partial H_{bc}$ has the form:

$$\frac{\partial C_{ij}}{\partial H_{bc}} = \frac{B_{ij}(\partial A_{ij}/\partial H_{bc}) - A_{ij}(\partial B_{ij}/\partial H_{bc})}{B_{ij}^2} \quad (5)$$

where we define intermediate variables $\mathbf{A}$ and $\mathbf{B}$ as follows for notational convenience:

$$\mathbf{A} = \mathbf{H}\mathbf{H}^{\mathsf{T}} \quad (6)$$
$$\mathbf{B} = \mathbf{n}\mathbf{n}^{\mathsf{T}} \quad (7)$$
$$\mathbf{n}_b = ||\mathbf{H}_b|| \quad (8)$$
$$\partial A_{ij}/\partial H_{bc} = \mathbf{1}_b \mathbf{H}_c^{\mathsf{T}} + \mathbf{H}_c \mathbf{1}_b^{\mathsf{T}} \quad (9)$$
$$\partial B_{ij}/\partial H_{bc} = H_{bc}(\mathbf{U}\mathbf{1}_b\mathbf{1}_b^{\mathsf{T}} + \mathbf{1}_b\mathbf{1}_b^{\mathsf{T}}\mathbf{U}^{\mathsf{T}}) \quad (10)$$
$$\mathbf{U} = \mathbf{n}(\mathbf{n}^{-1})^{\mathsf{T}} \quad (11)$$

where $\mathbf{1}_b$ is an indicator vector that is zero everywhere except for having the b^{th} element equal to one. $\mathbf{n}$ is a vector whose elements are the norms of the rows of $\mathbf{H}$, and $\mathbf{U}$ is an outer product of $\mathbf{n}$ with its element-wise inverse. For further details and justification of this core INFA algorithm, see [8].

After iterating the core INFA algorithm to convergence, we must then undo the preprocessing that was done to artificially ground the sources. We do this as described above by defining the ungrounded signal reconstruction $\tilde{\mathbf{H}} = \mathbf{H} + \mathbf{W}^{+}\tilde{\mathbf{V}}_{\text{inf}}$.

The ungrounded INFA algorithm thus consists of a preprocessing stage which grounds the observations, the application of the INFA algorithm, and then a postprocessing stage to recover the ungrounded sources.

4. RELATION TO OTHER METHODS

In contrast to UINFA or INFA, NMF by itself is not guaranteed to recover independent signals; any such recovery is purely incidental. The decorrelating regularization term is critical for independence.

Decorrelation in itself does not work as an independence criterion for ICA, and most ICA algorithms actually attempt to manipulate higher-order moments of the data either directly or indirectly to achieve independence in the unmixed outcome. Under some conditions, however, decorrelation can directly result in independence. Fancourt and Parra [4] have previously employed decorrelation as an independence criterion for nonstationary (although not non-negative) signals where they seek a solution that decorrelates the reconstructed sources at multiple points in time. Oja and Plumbley [11] also use decorrelation (without enforcing higher order independence) as part of their nonnegative ICA algorithm. They prove that this decorrelation criterion is sufficient for use as an independence criterion for nonnegative ICA as long as the source PDFs are "well-grounded". The key contrast between our work and these prior approaches is that we aim to estimate the mixing matrix, whereas prior methods have invariably attempted to estimate the unmixing matrix. We will show in our results that our approach yields better results in the noisy, overdetermined case.

Additional distinctions exist with respect to prior algorithms for ICA of non-negative data. In contrast to Plumbley's approach which only ensures that $\mathbf{H}$ is non-negative, and requires it to be "well-grounded", our approach ensures that both $\mathbf{W}$ and $\mathbf{H}$ are nonnegative and does not require $\mathbf{H}$ to be grounded. For some applications, this may be an important distinction. On the other hand, Plumbley's approach leads to a problem with no local minima, whereas our algorithm, like all NMF formulations, is guaranteed only to find a local minimum. The two constraints of decorrelation and nonnegativity will only be achieved when a perfect decomposition $\mathbf{V} = \mathbf{W}\mathbf{H}$ is found. For locally optimal solutions the decorrelation may not be complete, from which it follows that $\mathbf{H}$ will not be truly independent.

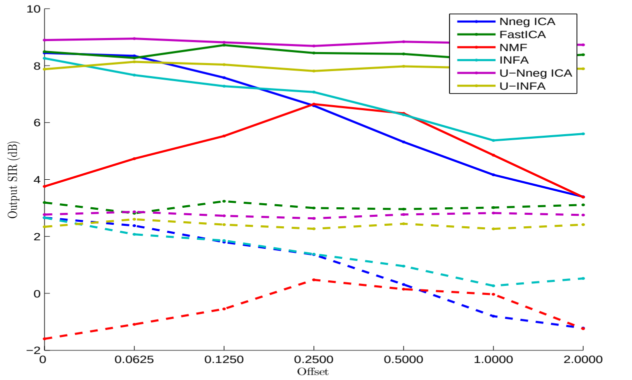

Fig. 1. Output SIR vs. offset for a square mixing matrix in the presence of additive noise with an input SNR of 19 dB. ($n_o = n_s = 3$). Solid lines show average performance. Dashed lines show the worst SNR out of the three recovered sources.

Nevertheless, we believe that our algorithm is useful because it can deal with non-square mixing matrices. We are not aware of an extension to Plumbley's approach to non-square matrices. We will show in the following section that in noisy conditions with non-square mixing matrices, our algorithm can outperform other approaches.

5. RESULTS

We test our algorithm on a simple synthetic problem against three related algorithms. The three methods are unregularized NMF ("NMF"), i.e. Equation 2 with $\alpha = 0$, FastICA [5] ("FastICA"), a popular ICA implementation (that does not include a nonnegativity constraint), and Oja and Plumbley's nonnegative ICA algorithm from [11] ("Nneg ICA"). For INFA and Plumbley's algorithm, we implement both the basic algorithm, which assume grounded data ("INFA" and "Nneg ICA," respectively), as well as ungrounded versions that incorporate the pre- and post-processing described above to handle ungrounded data ("U-INFA" and "U-Nneg ICA," respectively). The same pre- and post-processing that we describe for use in U-INFA can also be used with Plumbley's algorithm.

For the tests we generate synthetic observations $\mathbf{Y}$ by $\mathbf{Y} = [\mathbf{M}\mathbf{X} + Z]_{\varepsilon}$, where the elements of $\mathbf{M}$ are independently chosen from a uniform distribution on $[0, 1]$ and the elements of $\mathbf{X}$ are independently chosen from a uniform distribution on $[q, 1+q]$, where q is the "offset." For example, when $q = 0$, the distribution is on $[0, 1]$, i.e. the well-grounded case. Larger values of q lead to more and more ungrounded data. $\mathbf{Z}$ is IID Gaussian noise. We take as a baseline a problem with 500 samples of a 3-dimensional source and additive noise with a 19 dB input SNR, i.e. $\mathbf{X}$ is 3×500, and we vary the observation dimensionality and offset. We define "input SNR" to be the power ratio between $\mathbf{M}\mathbf{X}$ and $\mathbf{Z}$, the ratio of the mixed source power to the additive noise power. For INFA, U-INFA and for standard NMF, we initialize the entries of $\mathbf{W}$ and $\mathbf{H}$ with uniform random values from [0, 1].

Figure 1 shows results for a square (3×3) mixing matrix, and Figure 2 shows results for a 6×3 mixing matrix, resulting in twice as many observations as sources. Both figures show output SIR as offset is varied. "Output SIR" refers to the signal-to-interferer ratio (SIR), the ratio between the recovered source power and the residual power from other source channels remaining in the reconstruction.

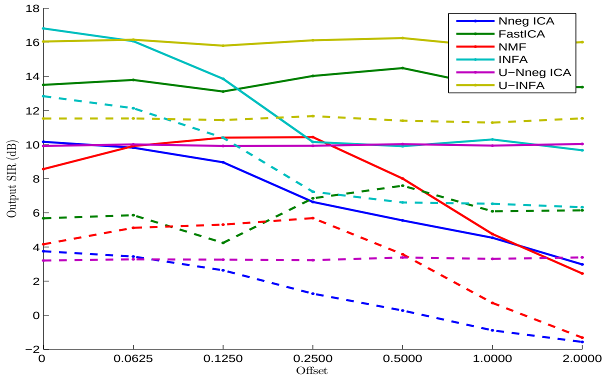

Fig. 2. Output SIR vs. offset for a mixing matrix with twice as many observations as sources in the presence of additive noise with an input SNR of 19 dB. ($n_o = 6, n_s = 3$). Solid lines show average performance. Dashed lines show the worst SNR out of the three recovered sources.

In each figure, solid lines represent mean SIR and dashed lines represent the minimum (worst) SIR of the three recovered sources. This minimum SIR is important because we often want reasonable reconstructions of all sources rather than very good reconstructions of some sources and very poor reconstructions of others. Each point in the figures is an average value over 100 realizations of the problem.

In general, we do not expect unregularized NMF ("NMF" in the results tables) to perform particularly well because it incorporates no independence constraint. It can be expected to achieve low reconstruction error, i.e. $WH \approx Y$, but the rows of H will not necessarily be a permutation of the rows of Y. Our results in general show that source reconstruction by NMF is poor.

Figure 1 shows that for the square mixing matrix and zero offset, all methods perform reasonably well except for plain NMF. However, as the offset is increased, the performance of INFA and Nneg ICA degrades. U-INFA and U-Nneg ICA maintain their performance in the presence of large offsets.

For the non-square (extra observations) case in Figure 2, U-INFA performs best in all but the zero-offset case. We are unaware of an extension of Oja and Plumbley's nonnegative ICA algorithm to handle non-square matrices, so instead we use only the first 3 observations as a square mixing problem. For this reason, nonnegative ICA performance is comparable in the two figures. Other than the fact that the number of sources was specified, FastICA was used with its default parameters. Note in particular that U-INFA has much better minimum output SIR performance than any other algorithm. As in the square mixing case, the performance of INFA and Nneg ICA degrades with increasing offset, while the performance of U-INFA and U-Nneg ICA does not.

We believe this scenario in which there are many noisy measurements of a relatively small number of independent sources is an important one, for example when a spectrogram with hundreds of frequency bins can be described using a relatively small number of basis functions. In our previous work, we have encountered this scenario while applying NMF-based techniques to speech denoising [12]. In future work, we hope to combine INFA with the regularization techniques in [12] and apply it to denoising and source separation of speech and other nonstationary signals.

6. CONCLUSION

We have presented an NMF-based algorithm for independent component analysis of ungrounded non-negative data. In contrast to previous methods, we estimate a mixing matrix, rather than an unmixing matrix. In contrast to previous work on nonnegative ICA, we do not require that signals be grounded. Experiments show that in the presence of additive noise, we are able to achieve unmixing comparable to other methods of ICA for square mixing matrices, and significantly better when the mixing matrix is not square. Additionally, unlike other approaches, the U-INFA's performance is not impaired by ungrounded sources.

7. REFERENCES

[1] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Neural Information Processing Systems*, 2000, pp. 556–562. [Online]. Available: citeseer.ist.psu.edu/lee01algorithms.html

[2] P. Smaragdis, M. Shashanka, B. Raj, and G. Mysore, "Probabilistic Factorization of Non-Negative Data with Entropic Co-occurrence Constraints," *Independent Component Analysis and Signal Separation*, pp. 330–337.

[3] P. O. Hoyer, "Non-negative sparse coding," *IEEE Workshop on Neural Networks for Signal Processing*, 2002.

[4] C. Fancourt and L. Parra, "A comparison of decorrelation criteria for the blind source separation of non-stationary signals," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2002, pp. 165–168.

[5] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999. [Online]. Available: citeseer.ist.psu.edu/hyv99fast.html

[6] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995. [Online]. Available: citeseer.ist.psu.edu/bell95informationmaximization.html

[7] M. Plumbley, "Conditions for non-negative independent component analysis," *IEEE Signal Processing Letters*, vol. 9, no. 6, pp. 177–180, 2002.

[8] K. W. Wilson and B. Raj, "Spectrogram dimensionality reduction with independence constraints," in *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, 2010.

[9] K. Wilson, A. Kruüger, B. Raj, and R. Haeb-Umbach, "Non-negative Matrix Factorization as Independent Component Analysis," *To be submitted*, 2010.

[10] A. Cichocki, R. Zdunek, and S. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," in *IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, vol. 5, 2006, pp. 621–625.

[11] E. Oja and M. Plumbley, "Blind separation of positive sources by globally convergent gradient search," *Neural Computation*, vol. 16, no. 9, pp. 1811–1825, 2004.

[12] K. Wilson, B. Raj, P. Smaragdis, and A. Divakaran, "Speech denoising using nonnegative matrix factorization with priors," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008.