Alpha-Stable Matrix Factorization

Umut Şimşekli, *Student Member, IEEE*, Antoine Liutkus, *Member, IEEE*, and Ali Taylan Cemgil, *Member, IEEE*

Abstract—**Matrix factorization (MF) models have been widely used in data analysis. Even though they have been shown to be useful in many applications, classical MF models often fall short when the observed data are impulsive and contain outliers. In this study, we present αMF, a MF model with α-stable observations. Stable distributions are a family of heavy-tailed distributions that is particularly suited for such impulsive data. We develop a Markov Chain Monte Carlo method, namely a Gibbs sampler, for making inference in the model. We evaluate our model on both synthetic and real audio applications. Our experiments on speech enhancement show that αMF yields superior performance to a popular audio processing model in terms of objective measures. Furthermore, αMF provides a theoretically sound justification for recent empirical results obtained in audio processing.**

Index Terms—**Markov chain monte carlo, matrix factorization, stable distributions.**

I. INTRODUCTION

MATRIX FACTORIZATION (MF) models have been a central topic in various research fields such as audio processing, finance, bioinformatics, and computer vision [1], [2]. In MF, the aim is to decompose a matrix as $\mathbf{X} \approx \mathbf{WH}$, where $\mathbf{X}$, $\mathbf{W}$, and $\mathbf{H}$ are of size $F \times N$, $F \times K$, and $K \times N$, respectively. Here, the approximation is in the sense of reducing a suitable cost function, for example $J(\mathbf{W}, \mathbf{H}) = D(\mathbf{X}, \mathbf{WH}) + R(\mathbf{W}, \mathbf{H})$, where $D(\cdot)$ is a *divergence* function that measures the approximation error and $R(\cdot)$ is a regularization term that enforces prior knowledge on the factors. This topic has a long and still active history in linear algebra, since classical problems such as truncated singular value decompositions and related algorithms fall into this category [3], [4], with the principal components analysis being an ubiquitous example.

An alternative approach for developing approximate MF models consists of using a probabilistic framework that has the following hierarchical generative model:

$$p(\mathbf{W}) = \prod_{fk} p(w_{fk}), \quad p(\mathbf{H}) = \prod_{kn} p(h_{kn})$$

$$p(\mathbf{X}|\mathbf{WH}) = \prod_{ft} p(x_{fn}|\mathbf{w}_f, \mathbf{h}_n) \quad (1)$$

where, $\mathbf{w}_f$ denotes the fth row of $\mathbf{W}$ and $\mathbf{h}_n$ denotes the nth column of $\mathbf{H}$. In this context, the cost function $J(\mathbf{W}, \mathbf{H})$ is selected as $-\log p(\mathbf{W}, \mathbf{H}|\mathbf{X})$ and minimizing it corresponds to finding the mode of the posterior. Depending on the choice of the prior distributions $p(\mathbf{W})$, $p(\mathbf{H})$, and the observation model $p(\mathbf{X}|\mathbf{WH})$, one can obtain a plethora of MF models with drastically different statistical properties. Typical choices for the observation models can be listed as the Gaussian distribution [5], [6], [7], Poisson distribution [8], [9], and compound Poisson distribution [10]. Even though MF with the above observation models have shown to be useful in various applications, these models may fall short when the observations are very impulsive and contain outliers, which is a common case in many domains such as audio processing and finance.

A popular approach for modeling impulsive data is to use heavy-tailed observation models, such as the t distribution [11], [12]. Instead of sticking to a particular observation model, in this study, we develop a novel MF model, called as αMF, that makes use of a family of heavy-tailed distributions as the observation model, so called the α-stable distributions. As we will describe in Section II, stable distributions have a rich structure and cover a broad range of noise distributions, where several important distributions appear as special cases. Besides, as opposed to many popular heavy-tailed observation models, α-stable distributions have rigorous statistical interpretations when they are used for modeling audio signals, as we will describe in Section V-B. Stable distributions have been used in signal processing, especially in robust time-series modeling [13], [14], [15], [16], [17], [18]. However, to the best of our knowledge, this is the first study to develop a MF framework with α-stable observations.

After a brief introduction to α-stable distributions, we describe αMF in detail. Then, we develop a Gibbs sampler for sampling from the posterior distributions of the latent variables. We evaluate our model on both synthetic and real audio data, where αMF outperforms a popular MF model on a speech enhancement application in terms of objective measures.

II. α-STABLE DISTRIBUTIONS

Stable distributions are heavy-tailed distributions and appear as the limiting distributions in the generalized central limit theorem [19]. They are characterized by four parameters: $\mathcal{S}(x; \alpha, \beta, \sigma, \mu)$, where **(1)** $\alpha \in (0, 2]$ is called the characteristic exponent and determines the tail thickness of the distribution. As this parameter gets smaller, the distribution will be heavier-tailed, and therefore the observations will be more impulsive. **(2)** $\beta \in [-1, 1]$ is called the skewness parameter and determines whether the distribution is left- or right-skewed. The distribution is called symmetric α-stable ($\mathcal{S}\alpha\mathcal{S}$) if $\beta = 0$. **(3)** $\sigma \in (0, \infty)$ is called the scale or the dispersion parameter.

It measures the spread of the random variable around its mode. (4) $\mu \in (-\infty, \infty)$ is the location parameter. The probability density function of a stable distribution cannot be written in closed-form except for certain special cases, which are the Gaussian distribution ($\alpha = 2$, $\beta = 0$), the Cauchy distribution ($\alpha = 1$, $\beta = 0$), and the Lévy distribution ($\alpha = 0.5$, $\beta = 1$). However, the characteristic function of the distribution can be written in closed form (see [19]).

α-stable distributions are readily extended to the case of vectors, and in particular to complex random variables $x \in \mathbb{C}$. In this study, we will make use of the complex isotropic α-stable distribution, which is shortly noted as $\mathcal{S}\alpha\mathcal{S}_c(x; \sigma)$, that reduces to $\mathcal{S}(x; \alpha, 0, \sigma, 0)$ in the real case [19], [20].

III. THE MODEL

In this section, we describe the α-Stable Matrix Factorization (αMF) model in detail. αMF models all the entries of an $F \times N$ complex matrix $\mathbf{X}$ as independent and $\mathcal{S}\alpha\mathcal{S}_c$ distributed with dispersion parameter decomposed as follows:

$$x_{fn}|\mathbf{w}_f, \mathbf{h}_n, \alpha \sim \mathcal{S}\alpha\mathcal{S}_c\left(x_{fn}; [\sum_k w_{fk}h_{kn}]^{1/\alpha}\right).$$

An equivalent formulation using augmentation leads to the following composite model [21]:

$$s_{fnk}|w_{fk}, h_{kn}, \alpha \sim \mathcal{S}\alpha\mathcal{S}_c\left(s_{fnk}; [w_{fk}h_{kn}]^{1/\alpha}\right)$$
$$x_{fn} = \sum_k s_{fnk} \qquad (2)$$

where $\mathbf{S} = \{s_{fnk}\}$ are called the latent sources. To be described in more detail in the next section, we will develop a Gibbs sampler for making inference in αMF, where we will need to sample from the conditional distributions of the latent variables. Therefore, we express αMF as conditionally Gaussian by making use of the product property of the stable distributions [15], [16], as follows:

$$\phi_{fnk}|\alpha \sim \mathcal{S}\left(\phi_{fnk}; \frac{\alpha}{2}, 1, 2(\cos\frac{\pi\alpha}{4})^{2/\alpha}, 0\right)$$
$$s_{fnk}|\phi_{fnk}, w_{fk}, h_{kn}, \alpha \sim \mathcal{N}_c\left(s_{fnk}; 0, \phi_{fnk}[w_{fk}h_{kn}]^{2/\alpha}\right)$$
$$x_{fn} = \sum_k s_{fnk}, \qquad (3)$$

where $\mathcal{N}_c$ denotes the complex isotropic Gaussian distribution and $\mathbf{\Phi} \equiv \{\phi_{fnk}\}$ is the *impulse* variable. This formulation will allow us to analytically derive the conditional distributions of $\mathbf{S}$, $\mathbf{W}$, and $\mathbf{H}$. Besides, now we can clearly see the impulsive structure of the model, where the variances of the Gaussian observations are modulated by infinite variance stable random variables, whose impulsiveness is controlled by α.

In order to preserve conjugacy, we assume generalized gamma priors on the latent factors:

$$w_{fk}|\alpha \sim \mathcal{GG}(w_{fk}; a_w, b_w, -2/\alpha)$$
$$h_{kn}|\alpha \sim \mathcal{GG}(h_{kn}; a_h, b_h, -2/\alpha), \qquad (4)$$

where the probability density function of the generalized gamma distribution is defined as follows:

$$\mathcal{GG}(x; a, b, c) = \frac{|c|}{\Gamma(a)b^{ca}} x^{ca-1} \exp\left(-(\frac{x}{b})^c\right). \qquad (5)$$
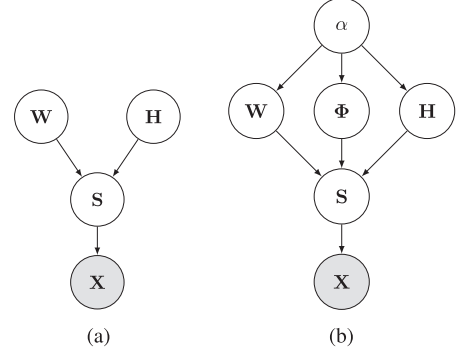

Fig. 1. Graphical models of IS-NMF and αMF. The nodes represent the random variables, the arrows determine the conditional independence structure, and the shaded the nodes represent the observed variables. (a) IS-NMF (b) αMF.

Finally, we assume uniform prior on α: $\alpha \sim \mathcal{U}(\alpha; (0, 2])$. The graphical representation of αMF is given in Fig. 1.

IV. INFERENCE

In practice, depending on the application, we are interested in the posterior distributions of the latent factors $p(\mathbf{W}, \mathbf{H}|\mathbf{X})$ or the latent sources $p(\mathbf{S}|\mathbf{X})$. In this study, we develop a Markov Chain Monte Carlo (MCMC) method for sampling from the posterior distributions of the latent variables $\mathbf{\Theta} \equiv \{\mathbf{W}, \mathbf{H}, \mathbf{S}, \mathbf{\Phi}\}$.

Monte Carlo methods are numerical techniques to approximately compute the expectations of the form:

$$\mathbb{E}[f(\mathbf{\Theta})] = \int f(\mathbf{\Theta})\pi(\mathbf{\Theta})d\mathbf{\Theta} \approx \frac{1}{M}\sum_{m=1}^{M} f(\mathbf{\Theta}^{(i)}) \qquad (6)$$

where $\mathbf{\Theta}^{(i)}$ are the samples drawn from $\pi(\mathbf{\Theta})$, that is the posterior distribution $\pi(\mathbf{\Theta}) = p(\mathbf{\Theta}|X)$ in our case. However, sampling directly from $\pi(\mathbf{\Theta})$ is intractable. MCMC methods generate samples from the target distribution $\pi(\mathbf{\Theta})$ by forming a Markov chain through a transition kernel; $\mathbf{\Theta}^{(i+1)} \sim \mathcal{T}(\mathbf{\Theta}|\mathbf{\Theta}^{(i)})$, whose stationary distribution is $\pi(\mathbf{\Theta})$, so that $\pi(\mathbf{\Theta}) = \int \mathcal{T}(\mathbf{\Theta}|\mathbf{\Theta}')p(\mathbf{\Theta}')d\mathbf{\Theta}'$. In order to design such kernels, various strategies can be applied [22]. In this study, we develop a Gibbs sampler, that implicitly forms a transition kernel by sampling from the full conditional distributions of the latent variables.

The full conditional distributions of $\mathbf{W}$ and $\mathbf{H}$ are given as follows:

$$p(w_{fk}|\alpha, \mathbf{S}, \mathbf{H}, \mathbf{\Phi}) = \mathcal{GG}(w_{fk}; a_w', b_w', -2/\alpha)$$
$$p(h_{kn}|\alpha, \mathbf{S}, \mathbf{W}, \mathbf{\Phi}) = \mathcal{GG}(h_{kn}; a_h', b_h', -2/\alpha)$$

where

$$a_w' = a_w + \frac{N}{2}, \quad b_w' = \left(b_w^{2/\alpha} + \sum_n \frac{|s_{fnk}|^2}{2h_{kn}^{2/\alpha}\phi_{fnk}}\right)^{\alpha/2},$$

$$a_h' = a_h + \frac{F}{2}, \quad b_h' = \left(b_h^{2/\alpha} + \sum_n \frac{|s_{fnk}|^2}{2w_{fk}^{2/\alpha}\phi_{fnk}}\right)^{\alpha/2}.$$

To sample the latent sources, it is possible to develop a 'block' Gibbs sampler, where we would need to sample $\{s_{fnk}\}_{k=1}^{K}$ jointly at each iteration [21]. However, this approach requires

sampling from a degenerate multivariate Gaussian and could yield computational inefficiencies in certain cases. Therefore, we utilize a *partially collapsed* Gibbs sampler [23] for sampling **S** where we draw samples from the marginal conditional distribution of s_{fnk} instead of the full conditional distribution, given as follows:

$$p(s_{fnk}|\alpha, \mathbf{W}, \mathbf{H}, \boldsymbol{\Phi}, \mathbf{X}) = \mathcal{N}_c(s_{fnk}; \mu'_s, \sigma'_s)$$

where

$$\mu'_s = g_{fnk}x_{fn}, \quad \sigma'_s = (1 - g_{fnk})[w_{fk}h_{kn}]^{2/\alpha}\phi_{fnk},$$

$$g_{fnk} = \frac{[w_{fk}h_{kn}]^{2/\alpha}\phi_{fnk}}{\sum_{k'}[w_{fk'}h_{k'n}]^{2/\alpha}\phi_{fnk'}}.$$

This approach provides computational advantages over block Gibbs sampling, since it needs to generate only univariate random variables.

Unfortunately, the full conditional distributions of α and $\boldsymbol{\Phi}$ cannot be derived analytically, therefore we resort to Metropolis Hastings (MH) algorithm for sampling from their full conditional distributions. For α, we choose a symmetric proposal distribution $q(\alpha'|\alpha) = \mathcal{N}(\alpha'; \alpha, \sigma_\alpha^2)$, that would explore the state space of α by a random walk. The acceptance probability for α then becomes:

$$a(\alpha \to \alpha') = \min\left(1, \frac{p(\mathbf{S}, \mathbf{W}, \mathbf{H}, \boldsymbol{\Phi}, \alpha')}{p(\mathbf{S}, \mathbf{W}, \mathbf{H}, \boldsymbol{\Phi}, \alpha)}\right)$$

Evaluating this probability requires stable densities to be evaluated twice at each epoch. Therefore, we have developed a fast numerical method for evaluating stable densities by making use of their power series representation [24], [16]. The details of this method is given in the online supplementary document [25].

We follow a similar procedure for $\boldsymbol{\Phi}$, where we choose the prior distribution of ϕ_{fnk} as its proposal distribution: $q(\phi_{fnk}) = p(\phi_{fnk})$. Accordingly, the acceptance probability simplifies and we obtain the following expression:

$$a(\phi_{fnk} \to \phi'_{fnk}) = \min\left(1, \frac{\mathcal{N}_c(s_{fnk}; 0, [w_{fk}h_{kn}]^{2/\alpha}\phi'_{fnk})}{\mathcal{N}_c(s_{fnk}; 0, [w_{fk}h_{kn}]^{2/\alpha}\phi_{fnk})}\right).$$

V. EXPERIMENTS

In this section, we will evaluate our method on both synthetic and audio data. Our implementations are mostly in Matlab, apart from α-stable density evaluation and random number generation algorithms, which are implemented in C.

A. Experiments on Synthetic Data

We first conduct experiments on synthetic data, where the aim is to validate our inference procedure. In these experiments, given a fixed α, we generate the latent variables $\mathbf{W}, \mathbf{H}, \boldsymbol{\Phi}, \mathbf{S}$, and the observed complex matrix $\mathbf{X}$ by using the generative model given in (3). Then, given the observed matrix $\mathbf{X}$, we run our inference algorithm after initializing all the latent variables randomly. In our experiments, we set $F = 50$, $N = 80$, $K = 2$, and $\sigma_\alpha^2 = 0.001$ and we repeat this experiment for several values of α.

Due to space limitations, we only report the results of the estimation of α, since it is the most prominent variable, determining
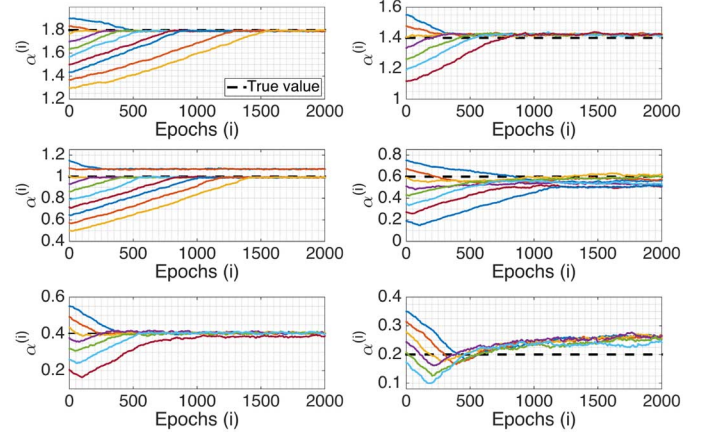

Fig. 2. Results of the synthetic data experiments.

the structure of the distribution. In Fig. 2, we visualize the samples $\alpha^{(i)}$ that are generated by our algorithm for different true α values. The results show that, even though the initial samples, $\alpha^{(0)}$, might be far from the true value of the variable, our inference algorithm can successfully locate the mode near the true α and starts sampling around that mode, even when the observations are coming from an extremely heavy-tailed distribution ($\alpha = 0.2$).

B. Experiments on Audio

In our next set of experiments, we evaluate αMF on real audio data. We compare αMF with Itakura-Saito NMF [5]; a MF model that is often used in audio processing, having the following underlying probabilistic model:

$$w_{fk} \sim \mathcal{IG}(w_{fk}; a_w, b_w), \quad h_{kn} \sim \mathcal{IG}(h_{kn}; a_h, b_h)$$
$$s_{fnk}|w_{fk}, h_{kn} \sim \mathcal{N}_c(s_{fnk}; 0, w_{fk}h_{kn})$$
$$x = \sum_k s_{fnk} \qquad (7)$$

where $\mathcal{IG}$ denotes the inverse gamma distribution. Here, $\mathbf{X}$ is taken as a time-frequency representation of the audio signal, with the indices f and n denoting the frequencies and the time-frames, respectively. IS-NMF appears as a special case of αMF: if we set $\alpha = 2$, the generalized gamma distribution becomes the inverse gamma distribution, $\boldsymbol{\Phi}$ becomes deterministic, and therefore αMF reduces to IS-NMF (see Fig. 1).

IS-NMF is considered as an important model for audio modeling since there is a rigorous statistical interpretation of the model from the waveform level to the power spectra level: if we assume that all the time-frames are independent and wide-sense stationary (WSS), we can show that all the entries of the short-time Fourier transform (STFT) of the signal are indeed independent and distributed with a complex centered isotropic Gaussian distribution [26] whose variances correspond to the power spectral density (PSD) of the signal. In this sense, IS-NMF models the PSD of a WSS signal by using a low rank approximation.

However, the assumption of the time-frames being WSS can be restrictive for various types of audio signals that have impulsive nature, such as speech. The interest of αMF in this context is that it generalizes IS-NMF by relaxing the WSS assumption
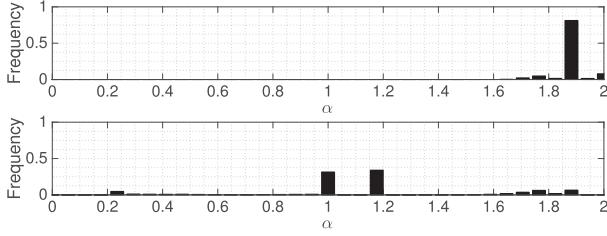
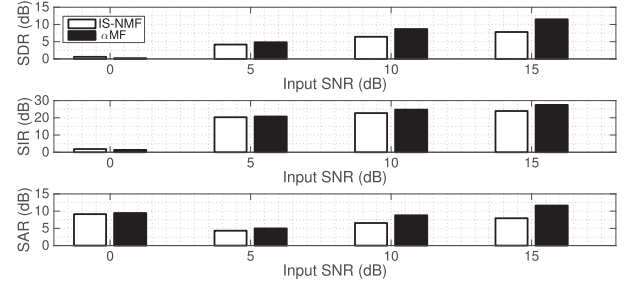Fig. 3. Histograms of α for noise (top) and speech (bottom).

Fig. 4. Evaluation results of IS-NMF and αMF on speech enhancement. Note that, IS-NMF coincides with αMF when $\alpha^{\mathrm{sp}} = \alpha^{\mathrm{no}} = 2$.

and assumes that all the time-frames are independent and stationary harmonizable α-stable processes. With such an assumption, we can show that the STFT coefficients are still independent but distributed with a $\mathcal{S}\alpha\mathcal{S}_c$ distribution, generalizing the WSS case $\alpha = 2$ [20].

We conduct our experiments on NOIZEUS noisy speech corpus [27]. This dataset contains 30 sentences that are uttered by 3 female and 3 male speakers. These sentences are corrupted by using 8 different real noise signals (train, babble, car, exhibition hall, restaurant, street, airport, train-station) at 4 different signal-to-noise ratio (SNR) levels. We analyze the signals by using the STFT with a Hamming window of length 512 samples and 75% overlap.

Firstly, we run αMF on each audio signal (30 clean speech and 8 noise signals). For each signal, we generate 2000 samples where we discard the first 100 of them as the burn-in period. We repeat this procedure three times with different initializations and combine all the samples in two groups: clean speech and noise. We use $K = 5$ for each noise signal and $K = 10$ for each speech signal, and we set $\sigma_\alpha^2 = 0.01$.

Fig. 3 shows the histograms of α for speech and noise. We can observe that, for the noise signals, the posterior distribution of α is concentrated near $\alpha = 1.89$, i.e. almost Gaussian, whereas we obtain two modes at $\alpha = 1.2$ and $\alpha = 1$ for the clean speech. This is expected because it has long been observed that informative signals such as speech tend to exhibit heavier tails than noises occurring in practice, justifying the use of α-stable models in audio [15]. More interestingly, this outcome provides a sound foundation to the recent empirical results obtained in [20], where the authors demonstrated that $\alpha = 1.2$ is the best performing exponent of the generalized Wiener filter, that implicitly assumes that the audio signals are stable distributed.

Secondly, we compare αMF with IS-NMF on a speech enhancement application, where the aim is to recover the clean speech signal, given a noisy speech signal. In this experiment, we follow a semi-supervised approach and use a slightly different model for the noisy mixtures, given as follows: $x_{fn}^{\mathrm{mix}} = x_{fn}^{\mathrm{sp}} + x_{fn}^{\mathrm{no}}$, where

$$x_{fn}^{\mathrm{sp}} \sim \mathcal{S}\alpha\mathcal{S}_c\left(x_{fn}^{\mathrm{sp}}; [\sum_k w_{fk}^{\mathrm{sp}} h_{kn}^{\mathrm{sp}}]^{1/\alpha^{\mathrm{sp}}}\right),$$

$$x_{fn}^{\mathrm{no}} \sim \mathcal{S}\alpha\mathcal{S}_c\left(x_{fn}^{\mathrm{no}}; [\sum_k w_{fk}^{\mathrm{no}} h_{kn}^{\mathrm{no}}]^{1/\alpha^{\mathrm{no}}}\right).$$

Here, 'sp' denotes the speech and 'no' denotes the noise. For IS-NMF we set $\alpha^{\mathrm{sp}} = \alpha^{\mathrm{no}} = 2$. For αMF, we set $\alpha^{\mathrm{sp}} = 1.2$ and $\alpha^{\mathrm{no}} = 1.89$, as suggested by the results above.

For each model, we first train the dictionary matrix $\mathbf{W}^{\mathrm{sp}}$ on the first 20 clean speech signals (2 female and 2 male speakers) by using the following approach. We concatenate the STFTs of the speech signals to obtain $\mathbf{X}$. Then, we run the Gibbs sampler for 3000 epochs where we set $\mathbf{W}^{\mathrm{sp}}$ to the Monte Carlo average (see (6)) by using the last 200 samples. The number of columns of $\mathbf{W}^{\mathrm{sp}}$ is chosen as $K^{\mathrm{sp}} = 100$.

At testing, for each input SNR, we apply both models on 80 different noisy mixtures, where we fix $\mathbf{W}^{\mathrm{sp}}$ and sample the rest of the latent variables, including $\mathbf{W}^{\mathrm{no}}$. Note that, the noisy speech signals are obtained by combining 8 different noise signals with 10 clean speech signals that are not used during training. For each mixture, we set $K^{\mathrm{no}} = 5$ and generate 2500 samples where we use the last 50 samples to estimate the posterior expectations of $\mathbf{X}^{\mathrm{sp}}$ and $\mathbf{X}^{\mathrm{no}}$.

For evaluating the quality of the estimates we use the signal-to-distortion ratio (SDR), signal-to-interference ratio (SIR), and signal-to-artifact ratio (SAR) that are computed with $\mathrm{BSS}_{\mathrm{EVAL}}$ version 3.0 [28]. Fig. 4 shows the results. We can observe that, both models perform poorly when the input SNR is low. However, as we increase the input SNR, the structure of the speech becomes more prominent, and we see that αMF becomes more advantageous in terms of all the objective measures. We obtain 4 dB SDR improvement when the input SNR is 15 dB. Besides, αMF results in less interference and less artifacts as measured by SIR and SAR. These differences are statistically significant with 5% significance level.

VI. CONCLUSION

In this study, we presented αMF, a matrix factorization model with α-stable observations. Due to the heavy-tailed nature of the stable distributions, αMF is particularly suited for impulsive or corrupted data that appear in several domains such as audio processing. We exploited the conditionally Gaussian nature of the stable distribution to develop a Gibbs sampler for sampling from the posterior distributions of the latent variables. We evaluated our model on both synthetic data and real audio signals, where αMF outperformed semi-supervised Itakura Saito-NMF in terms of objective measures on a speech enhancement application.

As a final remark, we note that there have been several extensions on IS-NMF that aim to incorporate the temporal and spatial structure of speech signals into the model [29], [30]. As a possible future direction, we believe that the performance of αMF can be further improved by extending the model in similar aspects.

REFERENCES

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

[2] 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*. Hoboken, NJ, USA: Wiley, Sep. 2009.

[3] G. Golub and C. V. Loan, *Matrix computations*. : JHU Press, 2012, vol. 3.

[4] N. Halko, P. Martinsson, and J. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM Rev.*, vol. 53, no. 2, pp. 217–288, 2011.

[5] C. Févotte, N. Bertin, and J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence. With application to music analysis," *Neural Computat.*, vol. 21, no. 3, pp. 793–830, Mar. 2009.

[6] M. N. Schmidt, O. Winther, and L. K. Hansen, "Bayesian non-negative matrix factorization," in *Independent Component Analysis and Signal Separation*, 2009, pp. 540–547.

[7] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *J. Royal Statist. Soc. B (Statist. Methodol.)*, vol. 61, no. 3, pp. 611–622, 1999.

[8] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems (NIPS)*, Apr. 2001, vol. 13, pp. 556–562.

[9] A. T. Cemgil, "Bayesian inference for nonnegative matrix factorisation models," *Computat. Intell. Neurosci.*, vol. 2009, pp. 4:1–4:17, 2009.

[10] U. Simsekli, A. Cemgil, and Y. K. Yilmaz, "Learning the beta-divergence in tweedie compound poisson matrix factorization models," in *Proc. 30th Int. Conf. Machine Learning (ICML-13)*, 2013, pp. 1409–1417.

[11] P. J. Wolfe, S. J. Godsill, and W.-J. Ng, "Bayesian variable selection and regularization for time–frequency surface estimation," *J. Royal Statist. Soc. B (Statist. Methodol.)*, vol. 66, no. 3, pp. 575–589, 2004.

[12] C. Fevotte and S. J. Godsill, "A Bayesian approach for blind separation of sparse sources," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 6, pp. 2174–2188, 2006.

[13] J. P. Nolan, "Modeling financial data with stable distributions," in *Handbook of Heavy Tailed Distributions in Finance, Handbooks in Finance*. Amsterdam, The Netherlands: North Holland, 2003, vol. 1, pp. 105–130.

[14] C. Menn and S. T. Rachev, "A garch option pricing model with α-stable innovations," *Eur. J. Oper. Res.*, vol. 163, no. 1, pp. 201–209, 2005.

[15] S. Godsill and E. Kuruoglu, "Bayesian inference for time series with heavy-tailed symmetric α-stable noise processes," in *Heavy Tails 99, Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, 1999, pp. 3–5.

[16] E. E. Kuruoglu, "Signal processing in α-stable noise environments: A least LP-norm approach," Ph.D. thesis, Univ. Cambridge, Cambridge, U.K., 1999.

[17] M. J. Lombardi and S. J. Godsill, "On-line bayesian estimation of signals in symmetric α-stable noise," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 775–779, 2006.

[18] N. Bassiou, C. Kotropoulos, and E. Koliopoulou, "Symmetric α-stable sparse linear regression for musical audio denoising," in *8th Int. Symp. Image and Signal Processing and Analysis (ISPA)*, 2013, pp. 382–387, IEEE.

[19] G. Samoradnitsky and M. Taqqu, *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Boca Raton, FL, USA: CRC Press, 1994, vol. 1.

[20] A. Liutkus and R. Badeau, "Generalized wiener filtering with fractional power spectrograms," in *40th Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[21] C. Févotte and A. T. Cemgil, "Nonnegative matrix factorisations as probabilistic inference in composite models," in *Proc. 17th Eur. Signal Processing Conf. (EUSIPCO)*, 2009, pp. 1913–1917.

[22] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Berlin, Germany: Springer, 2008.

[23] C. Févotte, O. Cappe, and A. Cemgil, "Efficient markov chain monte carlo inference in composite models with space alternating data augmentation," in *IEEE Statistical Signal Processing Workshop (SSP)*, 2011, pp. 221–224.

[24] H. Bergström, "On some expansions of stable distribution functions," *Ark. Matematik*, vol. 2, no. 4, pp. 375–378, 1952.

[25] Alpha-stable matrix factorization supplementary material [Online]. Available: www.loria.fr/aliutkus/amf

[26] A. Liutkus, R. Badeau, and G. Richard, "Gaussian processes for underdetermined source separation," *IEEE Trans. Signal Process.*, vol. 59, no. 7, pp. 3155–3167, 2011.

[27] Y. Hu and P. C. Loizou, "Subjective comparison and evaluation of speech enhancement algorithms," *Speech Commun.*, vol. 49, no. 7, pp. 588–601, 2007.

[28] E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. P. Rosca, "First stereo audio source separation evaluation campaign: Data, algorithms and results," in *Proc. Independent Component Analysis and Signal Separation*, 2007, pp. 552–559, Springer.

[29] C. Févotte, J. L. Roux, and J. R. Hershey, "Non-negative dynamical system with application to speech and audio," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 3158–3162.

[30] U. Şimşekli, J. Le Roux, and J. Hershey, "Non-negative source-filter dynamical system for speech enhancement," in *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6206–6210.