Adaptive Weights for NMF with Additional Priors

Julian M. Becker, Martin Rohbeck
and Christian Rohlfing
Institut für Nachrichtentechnik
RWTH Aachen University
D-52056 Aachen, Germany
Email: becker@ient.rwth-aachen.de

***Abstract*—Nonnegative matrix factorization (NMF) has become a very popular method in various signal processing applications. Supporting NMF with additional cost functions, so called priors, is very helpful to adapt the factorization to specific tasks. Additional priors are usually multiplied by fixed weights to adjust the influence of the prior. The question how to adapt these weights to the needs of specific factorization scenarios is yet unsolved. In this paper, we present a method to adjust the weights iteratively throughout the NMF process. We evaluate our method in an audio source separation environment and show, that it is more robust than the recently used method with fixed weights and that it leads to better separation results.**

I. Introduction

Nonnegative matrix factorization (NMF) is a frequently used method in audio source separation, e.g. [1], [2]. It was introduced by Paatero [3], but only became popular after Lee and Seung published efficient algorithms for its computation [4]. NMF is able to factorize audio signals into a specified number of components which correspond to individual sound events. These events can be assigned to the original sources by clustering, resulting in estimated separated sources.

As NMF was not originally developed for source separation, there are various options to extend it, to better adapt it to the task of audio source separation. In the past, several extensions to NMF have been proposed for this purpose. Some extensions use convolutive bases instead of multiplicative ones [5], [6], others extend the matrix factorization model to a tensor factorization model, to separate multichannel recordings [7]. Yet others introduce additonal constraints such as sparsity [8], [9], temporal continuity [9] or spectral continuity [10]. An overview over different versions and extensions of NMF can be found in [11].

In this paper, we focus on NMF with the previously mentioned additional constraints, so-called priors. These additional cost terms are usually added to a reconstruction cost term and multiplied with a weighting factor to adjust the influence on the factorization compared to the reconstruction cost term. The weights are usually fixed before starting the factorization. Choosing a good weight is difficult, because different signals need very different values for the weights. Temporal continuity for example is only helpful for harmonic signals. Thus, the optimal weigth for this prior depends on the harmonicity of the signal. For mixtures of harmonic and non-harmonic signal choosing a good weight becomes quite difficult. However, choosing the correct weight is essential for good separation results. If a prior is used too weak on a signal, the positive effect of the prior is not fully used. On the other hand, if a prior is used to strong, the effect might even be negative, compared to using NMF without additional prior. We propose a method to adapt the value of the weight while performing NMF, so that in the end a value is chosen that is well fitting for the specific signal.

The paper is structured as follows: In Section II, we provide basic information about NMF and its application for audio source separation. In Section III, we describe how additional priors can be used on NMF and analyze, how the weighting of these priors influence the factorization results. This analysis serves as motivation for our algorithm for adaptive weights, which is described in Section IV. Experimental results which show, that our approach has several advantages compared to the method with fixed weights are provided in Section V. Finally, in Section VI we give our conclusions.

II. Fundamentals

A. Nonnegative Matrix Factorization

NMF approximates a nonnegative matrix $\mathbf{X}$ of size $K \times N$ by a product of two nonnegative matrices $\mathbf{B}$ and $\mathbf{G}$

$$\mathbf{X} \approx \tilde{\mathbf{X}} = \mathbf{B}\mathbf{G}, \tag{1}$$

with $\mathbf{B}$ of size $K \times I$ and $\mathbf{G}$ of size $I \times N$. I is a user defined parameter, which is usually chosen smaller than K and N.

The matrices $\mathbf{B}$ and $\mathbf{G}$ are iteratively calculated by minimizing an adequate distance function $c(\mathbf{B}, \mathbf{G})$ between $\mathbf{X}$ and $\tilde{\mathbf{X}}$. Commonly used distance functions are the Euclidean distance, the Kullback-Leibler (KL) divergence and the Itakura-Saito (IS) distance. Lee and Seung [4] introduced efficient multiplicative update rules for the square of the Euclidean distance as well as for the KL divergence, resulting in convergence to a local minimum of the distance function. These update rules can be calculated using the gradient of $c(\mathbf{B}, \mathbf{G})$ with respect to $\mathbf{B}$,

$$\nabla_{\mathbf{B}} c(\mathbf{B}, \mathbf{G}) = \nabla^{+}_{\mathbf{B}} c(\mathbf{B}, \mathbf{G}) - \nabla^{-}_{\mathbf{B}} c(\mathbf{B}, \mathbf{G}), \tag{2}$$

where $\nabla^{+}_{\mathbf{B}} c(\mathbf{B}, \mathbf{G})$ and $\nabla^{-}_{\mathbf{B}} c(\mathbf{B}, \mathbf{G})$ are elementwise nonnegative terms of the gradient. Aquivalently, $\nabla_{\mathbf{G}} c(\mathbf{B}, \mathbf{G})$ is the gradient with respect to $\mathbf{G}$. The update rules are

$$\mathbf{B} \leftarrow \mathbf{B} \otimes \frac{\nabla^{-}_{\mathbf{B}} c(\mathbf{B}, \mathbf{G})}{\nabla^{+}_{\mathbf{B}} c(\mathbf{B}, \mathbf{G})} \tag{3}$$

and

$$\mathbf{G} \leftarrow \mathbf{G} \otimes \frac{\nabla^{-}_{\mathbf{G}} c(\mathbf{B}, \mathbf{G})}{\nabla^{+}_{\mathbf{G}} c(\mathbf{B}, \mathbf{G})}, \tag{4}$$

where $\otimes$ denotes elementwise multiplication and the divisions are also elementwise. For the methods presented in this paper, this generalized formulation of the update rules is sufficient. However, the exact update rules for KL-divergence and squared Euclidean distance can be found in [4] and for the IS-distance in [11].

B. NMF for Audio Source Separation

To use NMF for audio source separation, the time signal $\mathbf{x}$, consisting of M sources $\mathbf{s}_m$, first has to be transformed to time-frequency domain using short time Fourier transform (STFT). This results in a complex valued spectrogram $\underline{\mathbf{X}}$, which has a spectral and a temporal dimension. The NMF can be applied to the magnitude $\mathbf{X} = |\underline{\mathbf{X}}|$ of this spectrogram for audio source separation. Figure 1 shows the result of the factorization of the magnitude spectrogram of a mixture of one harmonic note (horizontal structure) and one percussive tone (vertical structure), using NMF with $I = 2$. The matrices $\mathbf{B}$ (on the left) and $\mathbf{G}$ (on top) are the result of the NMF. The columns of $\mathbf{B}$ (these vectors will from now on be denoted $\mathbf{b}_i$) capture the spectral shape of the acoustical events and can therefore be interpreted as spectral bases. The row vectors $\mathbf{g}_i$ of the matrix $\mathbf{G}$ can be interpreted as temporal activation vectors. The matrices $\tilde{\mathbf{C}}_i = \mathbf{b}_i\mathbf{g}_i$ form the spectrograms of individual acoustical events.
$\tilde{\mathbf{X}} = \sum_i \tilde{\mathbf{C}}_i$ is only an approximation for $\mathbf{X}$. However, it is desirable that the factorized spectrograms exactly sum up to the original spectrogram $\mathbf{X}$. This is done in a filtering step, which is also used to restore phase, using the phase of the mixture:

$$\hat{\underline{\mathbf{C}}}_i = \underline{\mathbf{X}} \otimes \left(\frac{\tilde{\mathbf{C}}_i}{\sum_i \tilde{\mathbf{C}}_i} \right). \quad (5)$$

This Wiener-like filtering is a frequently used postprocessing step of the results of the NMF (e.g. [2],[12]).
In a more complex mixture, I would have to be chosen corresponding to the number of the acoustical events in the mixture. Usually I is higher than the number of sources M, therefore the resulting spectrograms $\hat{\underline{\mathbf{C}}}_i$ have to be clustered to the melodies of the original sources. This is done in a clustering step, resulting in the estimated spectrograms $\hat{\underline{\mathbf{S}}}_m$ for the original sources.
These spectrograms are then transformed into time domain by inverse short-time Fourier transform (ISTFT). This step results in the estimations for the original sources in time domain $\hat{\mathbf{s}}_m$.

III. Additional Priors for NMF

The NMF update rules (3) and (4) usually minimize a reconstruction error cost function. However, since audio signals have some special properties, these properties can be used to improve NMF factorization. This can for example be done by adding new cost functions, so called priors, to the reconstruction terms, to penalize factorizations, which do not have these typical properties of audio signals. In the following, we will describe additional priors which only depend on the matrix $\mathbf{G}$. However, priors can equivalently be defined depending on the matrix $\mathbf{B}$.
With the additional prior, the cost term to be minimized by NMF transforms to

$$c(\mathbf{B}, \mathbf{G}) = c_r(\mathbf{B}, \mathbf{G}) + \alpha_t c_t(\mathbf{G}), \quad (6)$$

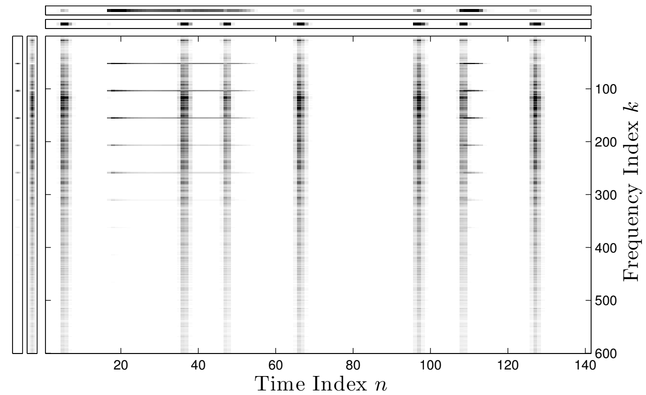


Fig. 1. Factorization of an audio mixture using NMF. Basis vectors $\mathbf{b}_i$ (on the left) and activation vectors $\mathbf{g}_i$ (on top) form the approximated spectrogram $\tilde{\mathbf{X}} = \sum_i \mathbf{b}_i\mathbf{g}_i$ (in the center).

where $c_r(\mathbf{B}, \mathbf{G})$ is the reconstruction error term (e.g. KL-divergence), $c_t(\mathbf{G})$ is the additional cost term and α_t is a weight to adjust the influence of the additional cost term. For $\alpha_t = 0$, this model equals the standard NMF.
The update rule for $\mathbf{G}$ transforms to

$$\mathbf{G} \leftarrow \mathbf{G} \otimes \frac{\nabla^-_{\mathbf{G}} c_r(\mathbf{B}, \mathbf{G}) + \alpha_t \nabla^-_{\mathbf{G}} c_t(\mathbf{G})}{\nabla^+_{\mathbf{G}} c_r(\mathbf{B}, \mathbf{G}) + \alpha_t \nabla^+_{\mathbf{G}} c_t(\mathbf{G})}, \quad (7)$$

while the update term for $\mathbf{B}$ stays the same as in Equation (3). It is important to mention that, while for single reconstruction error terms, such as KL-divergence, Euclidean distance or IS-distance, convergence to a local minimum has been mathematically proofen for the original update rules (3) and (4), this is not guaranteed anymore for the given update rules with additional prior. However, as long as the prior is weighted relatively small, convergence is usually assumed.

A. Analysis of the Temporal Continuity Prior

In this section, we evaluate one prior examplarily, with focus on the influence of the weighting factor α_t. We chose the temporal continuity prior proposed by Virtanen [9], since this was one of the first proposed priors and is still one of the most widely used ones. NMF with temporal continuity was proposed by Virtanen [9] to prevent incorrect factorizations. In the factorization in Fig. 1 it can be observed, that two small temporal gaps appear in the activation vector of the harmonic note, at the time, where it is overlapped by the percussive tone. This happens, because the complete energy of the signal is only assigned to the percussive source at this frame, or because of phase cancellations. The temporal continuity term proposed by Virtanen to solve this problem is a squared temporal difference (STD) cost term, which can be calculated as

$$c_t(\mathbf{G}) = \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{n=2}^{N} (g_{i,n} - g_{i,n-1})^2, \quad (8)$$

with $\sigma_i = \sqrt{(1/N)\sum_{n=1}^{N} g_{i,n}^2}$ being the standard deviation of each row of $\mathbf{G}$. $g_{i,n}$ denotes one element of the matrix $\mathbf{G}$ at indizes i and n. The negative and positive gradient terms of this cost function are

$$[\nabla^-_{\mathbf{G}} c_t(\mathbf{G})]_{i,n} = \frac{2N(g_{i,n-1} + g_{i,n+1})}{\sum_{l=1}^{N} g_{i,l}^2} + \frac{2N g_{i,n} \sum_{l=2}^{N} (g_{i,l} - g_{i,l-1})^2}{\left(\sum_{l=1}^{N} g_{i,l}^2\right)^2} \tag{9}$$

and

$$[\nabla^+_{\mathbf{G}} c_t(\mathbf{G})]_{i,n} = \frac{4N g_{i,n}}{\sum_{l=1}^{N} g_{i,l}^2}. \tag{10}$$

In [13], a modification of this prior was presented. It was

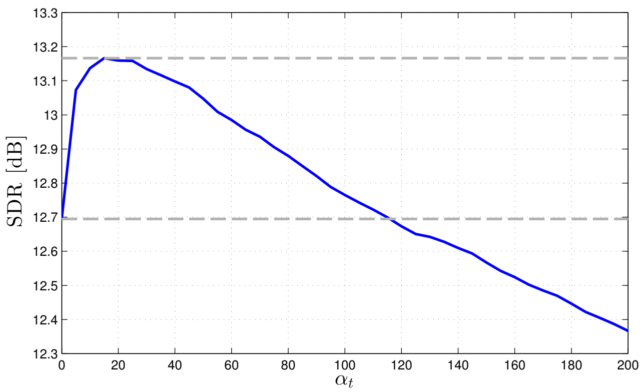


Fig. 2. Results for different values of α_t for the STD prior. The lower dashed line is a reference to the separation quality of the standard NMF. The upper dashed line shows the highest reached separation quality.

reasoned, that taking the logarithm of the cost function c_t leads to an adaption of the cost function to different components. The results showed an improvement of separation quality for this modified version of the prior. Thus, we also used this prior for evaluation. Since this prior uses the logarithm of the STD cost function as new cost function, we will denote it ln(STD) in the following. To analyse the influence of the fixed weights α_t, we first performed source separation with NMF with the STD prior for different values of α_t. The testset consisted of 60 audio signals, including harmonic and percussive signals, speech, vocals and noise, each being sampled with 44.1 kHz. These signals were mixed in every possible two-source combination, resulting in 1770 mixtures. The testset is identical to the one used in [2]. As quality measure of the separated sources, we used the signal-to-distortion ratio (SDR) [14]. The results of the mean SDR over all mixtures for the different values of α_t are shown in Fig. 2. The following observations can be made:

- For several values of α_t, the NMF with STD prior leads to better results, than the standard NMF ($\alpha_t = 0$, SDR $= 12.69$ dB).
- The best SDR with fixed α_t is achieved for $\alpha_t = 15$ (SDR $= 13.17$ dB).
- If the fixed value of α_t is chosen too high, ($\alpha_t > 120$), the SDR drops below the value of the standard NMF.

To evaluate the potential, of an algorithm to adapt α_t to the mixture, we calculated the highest SDR for each mixture for any α_t. The corresponding weight α_t is denoted $\alpha_{t,\text{opt}}$. Taking the mean of these maximum SDR values results in a

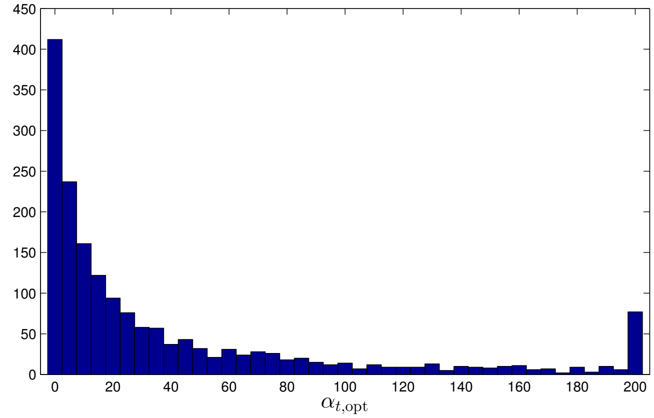


Fig. 3. Histogram of the distribution of the best values for α_t.

separation quality of 14.09 dB, almost 1 dB more, than the best quality with fixed α_t. This result shows the potential of an adaption of the weight to the mixture. Fig. 3 shows a histogram of the values of $\alpha_{t,\text{opt}}$, that reached the highest SDR over all 1770 mixtures. It can be observed, that the optimal α_t is very different for different mixtures. While a lot of mixtures reach the best results for $\alpha_t = 0$ (the standard NMF), several others reach their maximum for higher values. There is even a notable number of mixtures, for which the best SDR is reached for $\alpha_t > 120$, the values, for which the mean SDR drops below the mean SDR of the standard NMF for fixed α_t (see Figure 2). The bin $\alpha_{t,\text{opt}} = 200$ actually contains all cases, where the optimal weight would be 200 or greater and is therefore higher than neighbouring ones.

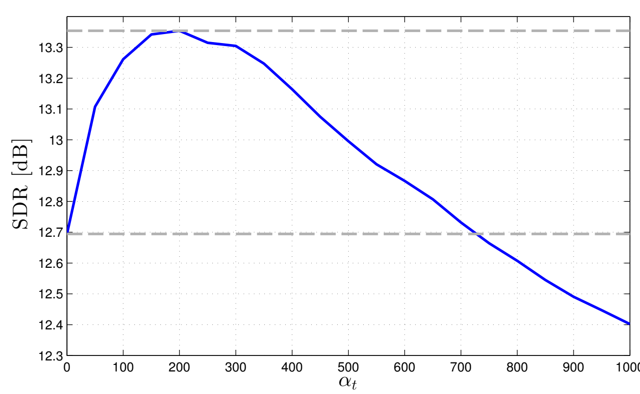
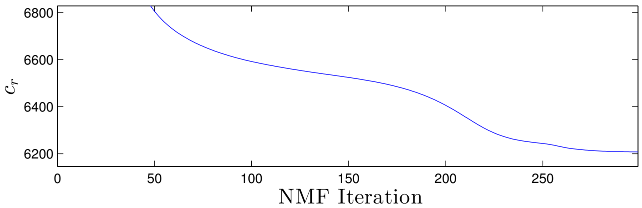
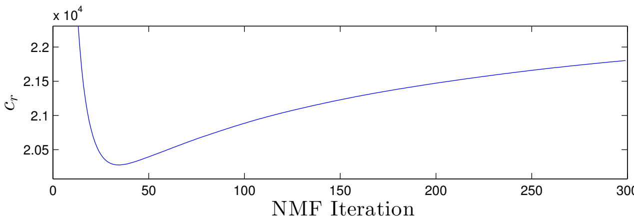


Fig. 4. Results for different values of α_t for the ln(STD) prior. The lower dashed line is a reference to the separation quality of the standard NMF. The upper dashed line shows the highest reached separation quality.

We also performed source separation with NMF with the ln(STD) prior for different values of α_t on the same testset. The results are shown in Figure 4. The characteristics of the SDR look similar than for the STD prior, however, the range of α_t differs from the one for the first prior. This shows, that the choice of α_t does not only depend on the signal, but also on the prior. The highest reached SDR value was 13.36 dB for this prior. Choosing $\alpha_{t,\text{opt}}$ for each mixture, the mean SDR increased to 14.25 dB.



(a) $\alpha_t = 70$



(b) $\alpha_t = 840$

Fig. 5. Comparison of the convergence behaviour of the reconstruction cost term c_r for different values of α_t. If α_t is chosen too high, c_r does not converge anymore.

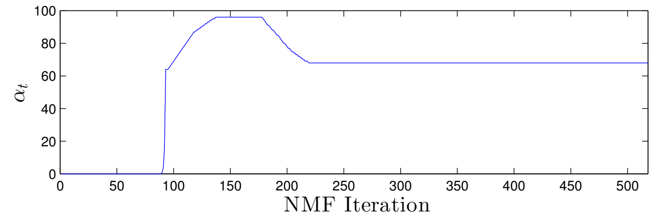


Fig. 6. Development of the estimation for the weight α_t for the proposed method.

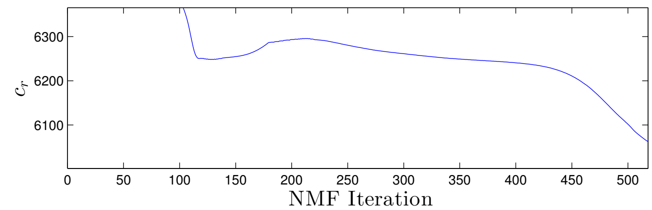


Fig. 7. Reconstruction error term c_r over NMF iterations for the proposed method with adaptive weights.

In the following, we will analyze the impact of an additional prior c_t on one specific example, to show, how different values of α_t change the results. For this example, we used the ln(STD) prior. We performed source separation with NMF with this prior on a mixture of electric guitar and drums. For the standard NMF, the SDR of the estimated sources was $17.5\,\text{dB}$: The highest separation quality with prior was reached for $\alpha_t = 70$ with $18.5\,\text{dB}$. For higher values of α_t, the separation quality decreased. For $\alpha_t = 840$, the SDR had again fallen to $17.5\,\text{dB}$ (the separation quality of the standard NMF). For higher α_t, the results were even lower than for standard NMF. Figure 5 shows the reconstruction error term c_r over the different iterations of NMF. This cost term can not be guaranteed to converge anymore, when an additional prior is used. However, since the standard NMF (which only uses the cost function c_r) has proven to be a suitable method for source separation, it can be concluded, that convergence of this cost term is desirable. Figure 5 indicates, that this assumption is correct. For the best weight ($\alpha_t = 70$), c_r still decreases in each iteration step. However, for the case, where the weight is already chosen too high ($\alpha_t = 840$), c_r increases for higher iterations.

IV. Weight Adaption

The example in Section III-A indicates, that the convergence behaviour of the reconstruction cost term can be used, to adapt the weights α_t to the mixture. If the weights are chosen too high, the cost function will not converge anymore, hence, α_t should be reduced. We use this result for our agorithm for the adaption of the weights α_t. The weights are changed throughout the NMF iterations. To adapt the weights, the behaviour of the reconstruction cost function c_r is used, to decide if the current value is too high or too low. The proposed weight adaption algorithm works in five phases:

1) In a first phase, standard NMF is started with $\alpha = 0$. This phase is necessary, because the reconstruction cost function c_r decreases very quickly in the first iterations and therefore, the impact of the weight α_t will not be reflected in the convergence behaviour. This is done, by defining a starting iteration i_{start}, until which the standard NMF is performed. After this iteration, the next phase starts.
2) In the second phase, the weight α_t is increased multiplicatively to quicky reach a rough first estimate of a good value for α_t. First, α_t is set to 1. After that, in each iteration step, the weight is multiplied by a factor λ, until the value of the recontruction cost term c_r rises for the first time. With this event, the second phase is finished and the third phase starts.
3) In the next phase, the weight is increased further, but very slowly, to find an upper bound for α_t. This is done, by incrementing the weight by 1 every time, the cost function starts to fall again. This phase stops, when the cost function does not fall any more for a number c_{wait} of iterations.
4) After finding this upper bound for α_t, the weight is then decreased again in the fourth phase, to find a reasonable final value for α_t. This is done similarily to phase 3. The weight is decremented by 1 every time, that the cost function starts to rise again. When the cost function does not rise any more for a number of c_{wait} iterations, this phase stops and the current weight α_t is used as final estimate.
5) In the last phase, i_{final} more iterations are performed with the now fixed final weight α_t, to let the constrained NMF finalize the factorization.

Figure 6 shows, how α_t is estimated for the given example. After $i_{\text{start}} = 90$ NMF iterations (phase 1), the estimation starts and α_t increases quickly in the second phase. After that, α_t is slowly increased to find an upper bound. This is reached after around 140 iterations. In the next phase, the weight is decreased, until the final estimate $\alpha_t = 68$ is reached after 220 iterations. After that, $i_{\text{final}} = 300$ final NMF iterations are performed with the estimated α_t. The SDR of this separation ($18.4\,\text{dB}$) was close to the best separation quality reached with the fixed weight. Figure 7 shows, how the reconstruction error term c_r developed during the estimation process. It can be seen, that the term increases only in the short period, where the upper bound for α_t is found (between iterations 150 and 200), but decreases again, as the final estimate is chosen.

V. Experimental results

We performed source separation as described in Sec. II-A. To evaluate the separation quality of the NMF without being affected by errors of a clustering algorithm, we used a non-blind clustering with knowledge of the original signals, as described in [9]. As measure for separation quality, we used the signal-to-distortion ratio (SDR), signal-to-inference ratio (SIR) and signal-to-artifacts ratio (SAR), as proposed in [14]. All given values are averaged over the complete used testset.

A. Testset & Setup

We used two different testsets for evaluation of the algorithm. The first testset was the same, that we already used in Section III-A, consisiting of 60 signals, that were mixed to 1770 two-source mixtures. We used this testset for general evaluation as well as for training purposes to optimize parameters of the algorithms. The second testset consists of 26 samples of harmonic and percussive instruments as well as singing voice, that were extracted from the QUASI database [15]. All signals were again mixed in every possible two-source combination, resulting in 300 mixtures. This testset was used to evaluate the algorithm without optimizing the algorithms parameters, using the parameters obtained from the first testset.
For the STFT, we used a window size of $s_w = 2^{12}$ and a hop size of $s_h = 2^{11}$ samples. I was set to 20 for every mixture, since this had shown to be a suitable number of components for the testsets. As reconstruction error term, we used KL-divergence, as this produced the best separation results. We initialized the NMF by performing an SVD on the complex spectrogram $\underline{\mathbf{X}}$ as proposed in [16]. For the experiments with fixed weights, we performed 300 NMF iterations. For the proposed method, the total number of iterations varied, depending on how quickly the final estimate for the weight α_t was found. We set α_{final} to 300 to perform the same number of NMF iterations with the final estimate than with the fixed weights. λ was set to 4 for all experiments.

B. Results for the First Testset

We first performed source separation with NMF with Virtanens STD prior as described in Section III-A, but with our algorithm with adaptive weights. The results for different values of i_{start} and c_{wait} are shown in Figure 8. Note, that for all shown parameter combinations, the resulting SDR is higher, than the highest reached value of 13.17 dB ($\alpha_t = 15$) for the method with fixed α_t (see Sec. III-A). From this experiment, we conclude that $i_{\text{start}} = 30$ is a reasonable value for our algorithm. For c_{wait}, the performance of the algorithm is good in general, as long as c_{wait} is not chosen too small. We decided to use $c_{\text{wait}} = 80$ in our following experiments, since higher values only increase the separation quality slightly. A small value of c_{wait} is desirable, because

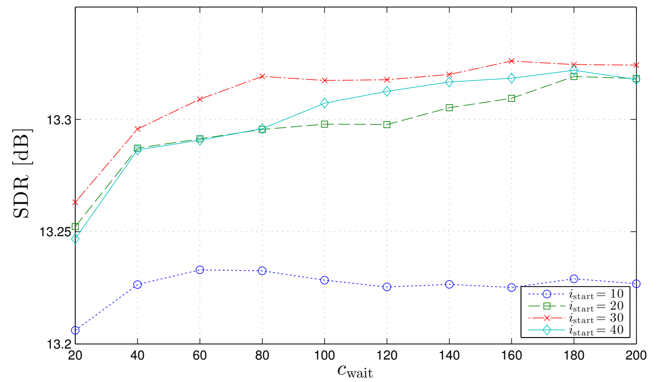


Fig. 8. Results for different values of c_{wait} and i_{start} for the first prior on the first testset.

it leads to a quicker estimation of α_t.
We also performed source separation with NMF with the ln(STD) prior. The results for this prior with different fixed values for α_t are shown in Figure 4. The highest reached separation quality in SDR was 13.36 dB for $\alpha_t = 200$. Note, that while the best α_t for the STD prior was 15, the best value for this second prior is in a completely different range. This means, that with fixed α_t, the weights have to be learned separately for different priors. Since our algorithm should be able to find a good estimate for alpha independently of the used prior, we performed the proposed method with the parameters $i_{\text{start}} = 30$ and $c_{\text{wait}} = 80$, that we got from the experiments with the first prior. No new optimization for the second prior was made. The resulting SDR was 13.47 dB, increasing the best result with fixed weights (13.36 dB) by around 0.1 dB. If we had similarily chosen the best value for α_t, that we had received from the first prior ($\alpha_t = 15$), the SDR for the fixed weight would only have been 12.84 dB. Besides the higher maximum separation quality, this result shows another advantage of our algorithm: The parameters i_{start} and c_{wait} can be trained independently of the prior and still lead to very good results for different priors, while training α_t for the method with fixed weights only leads to improved separation results for the specific prior, for which it was trained.

TABLE I. Results for the first testset

	NMF	STD [9]	STD_{ad}	ln(STD) [13]	$\ln(\text{STD})_{\text{ad}}$
SDR [dB]	12.69	13.17	**13.32**	13.36	**13.47**
SIR [dB]	18.76	19.10	**19.48**	19.34	**19.58**
SAR [dB]	15.64	**16.65**	16.50	**16.66**	16.59

Table I gives an overview over the separation results in SDR, SIR and SAR for the two different priors for the methods with fixed weights and our method with $i_{\text{start}} = 30$ and $c_{\text{wait}} = 80$. STD and ln(STD) are the best results for the two priors with fixed weights ($\alpha_t = 15$ and $\alpha_t = 200$ respectively). STD_{ad} and $\ln(\text{STD})_{\text{ad}}$ are the corresponding results for the same priors with our algorithm with adaptive weights with $i_{\text{start}} = 30$ and $c_{\text{wait}} = 80$. SDR and SIR are increased with our algorithm for both priors, SAR is slightly lower for both priors.

C. Results for the Second Testset

To proof, that our method can also be used more independently of the testset, than the one with fixed weights, we performed source separation on the second testset, using the trained parameters of the first testset. Again, we used the same two priors as in the previous section, to show, that the results are valid for different priors.
For our method, we used the parameters $i_{\text{start}} = 30$ and $c_{\text{wait}} = 80$ for both of the priors. For the method with fixed weights, we used $\alpha_t = 15$ for the first prior and $\alpha_t = 200$ for the second prior, as optimized on the first testset.

TABLE II. RESULTS FOR THE SECOND TESTSET

	NMF	STD [9]	STD_{ad}	ln(STD) [13]	$\ln(\text{STD})_{\text{ad}}$
SDR [dB]	13.97	13.98	**14.22**	14.46	**14.58**
SIR [dB]	20.34	20.04	**20.42**	20.72	**20.80**
SAR [dB]	16.36	16.81	**16.88**	17.06	**17.22**

Table II shows the separation quality for this experiment in SDR, SIR and SAR. Two different conclusions can be made from these results:

- For the first prior, the method with fixed weights ($13.98\,\text{dB}$) does not significantly increase the SDR separation performance compared to standard NMF ($13.97\,\text{dB}$). Also, the SIR is even lower than for the standard NMF. Only SAR is increased. Thus, we can conclude, that the value for α_t that was optimized on the first testset is not a suitable choice for the second testset.
- For both priors, our method increased the separation results in SDR, SIR and SAR compared to the methods with fixed α_t. We can conclude, that our method does not only make the usage of the additional priors more robust, but also leads to higher separation results.

VI. CONCLUSION

In this paper, we first analyzed, how fixed weights of additional priors for NMF can lead to problems, since different signals need different weights. With this motivation, we presented a method to adapt the weights to the signal on which NMF is performed, using the convergence behaviour of the cost function. We showed by experiment, that our method has several advantages compared to the recently used method with fixed weights: It is more robust regarding choice of priors, parameters do not have to be trained separately for different priors. It is also more robust for different testsets, where the pretrained weights for the method with fixed weights sometimes do not improve results. Finally, our method also leads to higher maximum separation results on average over the whole testsets.
The promising results of our algorithm should motivate future work in this field. Possible modifications of the algorithm could include new ways of deciding on the transitions from one phase to another. While currently the number of iterations are counted, it would also be possible to define a convergence criterion for the cost function as stopping criterion for some of the phases.

REFERENCES

[1] A. Cichocki, R. Zdunek, and S.-I. Amari, "New algorithms for non-negative matrix factorization in applications to blind source separation," in *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, vol. 5. IEEE, 2006, pp. V–V.

[2] M. Spiertz and V. Gnann, "Beta divergence for clustering in monaural blind source separation," in *Audio Engineering Society Convention 128*. Audio Engineering Society, 2010.

[3] P. Paatero and U. Tapper, "Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values," *Environmetrics*, vol. 5, no. 2, pp. 111–126, 1994.

[4] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in neural information processing systems*, 2001, pp. 556–562.

[5] M. N. Schmidt and M. Mørup, "Nonnegative matrix factor 2-D deconvolution for blind single channel source separation," in *Independent Component Analysis and Blind Signal Separation*. Springer, 2006, pp. 700–707.

[6] P. Smaragdis, "Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs," in *Independent Component Analysis and Blind Signal Separation*. Springer, 2004, pp. 494–499.

[7] D. FitzGerald, M. Cranitch, and E. Coyle, "Non-negative tensor factorisation for sound source separation," 2005.

[8] C. Joder, F. Weninger, D. Virette, and B. Schuller, "A comparative study on sparsity penalties for nmf-based speech separation: Beyond lp-norms," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*. IEEE, 2013, pp. 858–862.

[9] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," vol. 15, no. 3. IEEE, 2007, pp. 1066–1074.

[10] J. M. Becker, C. Sohn, and C. Rohlfing, "NMF with spectral and temporal continuity criteria for monaural sound source separation," in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 22nd European*. IEEE, 2014, pp. 316–320.

[11] A. Cichocki, R. Zdunek, A. H. Phan, and S.-I. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[12] R. Jaiswal, D. FitzGerald, D. Barry, E. Coyle, and S. Rickard, "Clustering NMF basis functions using shifted NMF for monaural sound source separation," in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011, pp. 245–248.

[13] J. M. Becker and C. Rohlfing, "Component-adaptive priors for NMF," in *Latent Variable Analysis and Signal Separation*, 2015.

[14] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," vol. 14, no. 4. IEEE, 2006, pp. 1462–1469.

[15] "QUASI database - a musical audio signal database for source separation," http://www.tsi.telecom-paristech.fr/aao/en/2012/03/12/quasi/.

[16] J. M. Becker, M. Menzel, and C. Rohlfing, "Complex SVD initialization for NMF source separation on audio spectrograms," in *DAGA 2015*, Nürnberg, Germany, 2015.