Generalized Constraints for NMF with Application to Informed Source Separation

Christian Rohlfing and Julian M. Becker

Institut für Nachrichtentechnik

RWTH Aachen University, D-52056 Aachen, Germany

Email: rohlfing@ient.rwth-aachen.de

Abstract—**Nonnegative matrix factorization (NMF) is a widely used method for audio source separation. Additional constraints supporting e.g. temporal continuity or sparseness adapt NMF to the structure of audio signals even further.**

In this paper, we propose generalized NMF constraints which make use of prior information gathered for each component individually. In general, this information could be obtained blindly or by a training step. Here we make use of these novel constraints in an algorithm for informed audio source separation (ISS). ISS uses source separation to code audio objects by assisting a source separation step in the decoder with parameters extracted with knowledge of the sources in the encoder. In [1], a novel algorithm for ISS was proposed which makes use of an NMF step in the decoder. We show in experiments that the generalized constraints enhance the separation quality while keeping the additionally needed bit rate very low.

I. INTRODUCTION

Additional constraints are a powerful extension to nonnegative matrix factorization (NMF) as they adapt NMF to specific applications such as audio source separation. In this field, most prominent constraints enforce either continuity or sparseness on the NMF matrices [2], [3]. In [4], constraints were adapted to each NMF component since the components are often structured differently and should not be constrained equally. Usually, NMF tries to minimize these constraints.

In this paper, we assume that some prior information exists which gives us the opportunity to calculate certain reasonable *target values* for the constraint. Instead of minimizing the constraint, we let the constraint on each component converge to the corresponding target value.

We evaluate this novel constraint family in the field of informed audio source separation (ISS). ISS uses source separation in the setting of audio object coding and usually consists of two stages: The original sources are perfectly known at the *encoder* which calculates a compact set of side-information. This information is transmitted alongside the mixture of the sources to the *decoder* which uses the side-information to estimate the sources by a source separation step. ISS was initially proposed in [5] and adapted to NMF inspired algorithms in e.g. [6] and [7]. Refer to [8] for a comparative study of ISS algorithms.

In [1], a novel ISS algorithm was proposed which makes use of a semi-blind source separation (SBSS) algorithm using NMF in the decoder. In this scenario, the constraint target values are extracted from an NMF model calculated from the sources directly which are perfectly known at the encoder. After transmission, the target values are then used to guide the decoder NMF at run-time.

This paper is structured as follows. In Section II, we discuss source separation with constrained NMF. We review the recently proposed NMF-ISS algorithm based on semi-blind source separation using NMF in Section III. In Section IV, we introduce novel generalized constraints for NMF and show the application to NMF-ISS. We present experimental results in Section V and conclude this paper with Section VI.

II. NMF-BASED SOURCE SEPARATION

A. Overview

In the following, we assume a linear mixture of M sources

$$\underline{\mathbf{X}} = \sum_{m=1}^{M} \underline{\mathbf{S}}_m \qquad (1)$$

where $\underline{\mathbf{X}}$ and $\underline{\mathbf{S}}_m$ denote the complex spectrograms of the mixture and the mth source in time-frequency (TF) domain.

The source separation algorithm described in the following is based on the algorithm proposed in [9], [10]. First, the mixture is separated by NMF into acoustical events also denoted as components. Second, these components are used for TF Wiener-like masking to yield the estimated sources.

NMF is used to factorize the magnitude mixture spectrogram $\mathbf{X} = |\underline{\mathbf{X}}| \in \mathbb{R}_+^{F \times T}$ into I components

$$x_{ft} \approx \sum_{i=1}^{I} b_{fi} \, g_{ti} \qquad (2)$$

with frequency basis $\mathbf{B} \in \mathbb{R}_+^{F \times I}$, temporal gain $\mathbf{G} \in \mathbb{R}_+^{T \times I}$, spectral index $f \in [1, F]$, time bin $t \in [1, T]$, component index $i \in [1, I]$ and x_{ft} one element of matrix $\mathbf{X}$.

Starting with an initialization for the NMF matrices, $\mathbf{B}_0$ and $\mathbf{G}_0$, NMF uses multiplicative update rules to minimize the β-Divergence d_β between the left and the right hand side of Equation (2). This reconstruction cost function can be extended by constraints enforcing different structures of $\mathbf{B}$ or $\mathbf{G}$ such as sparseness or continuity.

Figure 1 shows the magnitude spectrogram $\mathbf{X}$, the spectral basis $\mathbf{B}$ and the temporal gain matrix $\mathbf{G}$ for an exemplary trumpet-tambourine mixture.
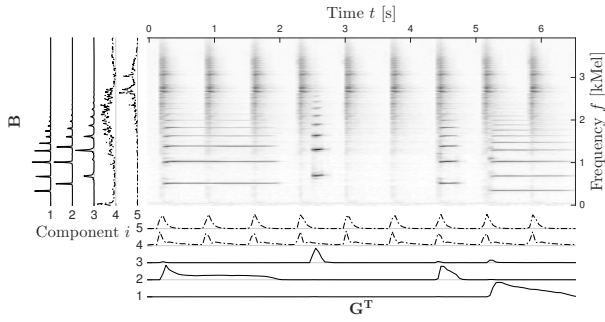
Fig. 1. Mixture spectrogram $\mathbf{X}$, frequency basis $\mathbf{B}$ and temporal gain $\mathbf{G}$ for exemplary trumpet-tambourine mixture. The trumpet is modeled by the first three and the tambourine by the other two components.

For synthesis, Wiener-like TF masking with the complex mixture spectrogram $\underline{\mathbf{X}}$ is performed to restore phase information and finer structures for each component spectrogram. Finally, a clustering step is needed to group the component spectrograms to M estimated sources $\underline{\tilde{\mathbf{S}}}_m$.

B. Constrained NMF

To adapt NMF further to the application of audio source separation, additional constraints are very helpful. In [2], two very popular constraints on the temporal gain matrix $\mathbf{G}$ were proposed: Temporal continuity and sparseness favor either smoothness or sparseness of the column vectors $\mathbf{g}_i$

$$c_{\text{tc}}(\mathbf{G}) = \sum_{i=1}^{I} \frac{1}{\sigma_i^2} \sum_{t=2}^{T} [g_{ti} - g_{t-1,i}]^2 = \sum_i c_{\text{tc}}(\mathbf{g}_i) \quad (3)$$

$$c_{\text{s}}(\mathbf{G}) = \sum_{i=1}^{I} \sum_{t=1}^{T} \frac{g_{ti}}{\sigma_i} = \sum_i c_{\text{s}}(\mathbf{g}_i) \quad (4)$$

with $\sigma_i = \sqrt{\frac{1}{T} \sum_t g_{ti}^2}$ estimating the variance of the ith column vector $\mathbf{g}_i$ of $\mathbf{G}$. These constraints describe either typical behaviour of harmonic or percussive components respectively.

In the following derivations, we assume arbitrary constraints $c(\mathbf{G})$ on the temporal gain $\mathbf{G}$ with the property that the constraint can be calculated componentwise

$$c(\mathbf{G}) = \sum_i c(\mathbf{g}_i) \ . \quad (5)$$

Please note that the following derivations and our proposed method (cf. Section IV) are also applicable to constraints on $\mathbf{B}$ as long as they can be evaluated on each component separately as shown in Equation (5) for $\mathbf{G}$. However, we only consider constraints on $\mathbf{G}$ for conciseness in this paper.

Usually, NMF constraints are weighted with a factor $\alpha \geq 0$ and added to the β-Divergence d_β taking the NMF reconstruction error into account thus yielding a combined cost function

$$c_{\text{c}}(\mathbf{B}, \mathbf{G}) = d_\beta(\mathbf{X}, \mathbf{B}, \mathbf{G}) + \alpha \, c(\mathbf{G}) \ . \quad (6)$$

In the following, we briefly review the derivation of the multiplicative update rules for calculating the NMF matrices $\mathbf{B}$ and $\mathbf{G}$. The gradients of the combined NMF cost function

given in Equation (6) with respect to $\mathbf{B}$ and $\mathbf{G}$ can be expressed as $\nabla_{\mathbf{B}} c_{\text{c}}(\mathbf{B}, \mathbf{G}) = \nabla_{\mathbf{B}}^+ c_{\text{c}}(\mathbf{B}, \mathbf{G}) - \nabla_{\mathbf{B}}^- c_{\text{c}}(\mathbf{B}, \mathbf{G})$ with the gradient terms $\nabla_{\mathbf{B}}^+ c_{\text{c}}(\mathbf{B}, \mathbf{G})$ and $\nabla_{\mathbf{B}}^- c_{\text{c}}(\mathbf{B}, \mathbf{G})$ element-wise positive. $\nabla_{\mathbf{G}} c_{\text{c}}(\mathbf{B}, \mathbf{G})$ and the gradients of d_β and $c(\mathbf{G})$ can be expressed equivalently. With these expressions, the multiplicative update rules for constrained NMF yield

$$b_{fi} \leftarrow b_{fi} \frac{\left[\nabla_{\mathbf{B}}^- d_\beta(\mathbf{X}, \mathbf{B}, \mathbf{G})\right]_{fi}}{\left[\nabla_{\mathbf{B}}^+ d_\beta(\mathbf{X}, \mathbf{B}, \mathbf{G})\right]_{fi}} \quad (7)$$

$$g_{ti} \leftarrow g_{ti} \frac{\left[\nabla_{\mathbf{G}}^- d_\beta(\mathbf{X}, \mathbf{B}, \mathbf{G})\right]_{ti} + \alpha \left[\nabla_{\mathbf{G}}^- c(\mathbf{G})\right]_{ti}}{\left[\nabla_{\mathbf{G}}^+ d_\beta(\mathbf{X}, \mathbf{B}, \mathbf{G})\right]_{ti} + \alpha \left[\nabla_{\mathbf{G}}^+ c(\mathbf{G})\right]_{ti}} \quad (8)$$

For constraints on $\mathbf{B}$, the corresponding positive and negative gradient terms are weighted and added to the denominator and enumerator of the update rule in Equation (7). The gradient terms for d_β are given e.g. in [10] and for temporal continuity and sparseness constrains in [2]. Note that for $\alpha = 0$ unconstrained NMF is performed.

C. Component-adaptive constraints for NMF

Not all constraints are suitable for all components in general. For example, temporal continuity favors components modeling harmonic notes which are more often continuous in time than percussive ones.

To cope with differently structured components at the same time, component-adaptive constraints were introduced in [4]. The main principle is to amplify the effect of the constraint c for components with a low value of $c(\mathbf{g}_i)$ which means that the constraint is already favored by these components. Therefore, the gradient of the component-adaptive constraint $\tilde{c}$ is set to

$$[\nabla_{\mathbf{G}} \tilde{c}(\mathbf{G})]_{s,j} = [\nabla_{\mathbf{G}} c(\mathbf{G})]_{s,j} / c(\mathbf{g}_j) \ . \quad (9)$$

The corresponding constraint term $\tilde{c}$ can be derived as

$$\tilde{c}(\mathbf{G}) = \sum_i \ln\left(c(\mathbf{g}_i)\right) \ . \quad (10)$$

III. NMF-BASED INFORMED SOURCE SEPARATION

NMF-based informed source separation (NMF-ISS) was proposed recently in [1]. The general idea is to use a semi-blind source separation (SBSS) algorithm in the decoder to separate the sources out of their mixture. NMF, which is used for SBSS, is assisted by side-information extracted in the encoder with knowledge of the sources.

A. Encoder

Fig. 2a shows the block diagram of the NMF-ISS encoder as proposed in [1]. With knowledge of the sources, the encoder first calculates an interference-free NMF model of each source. Instead of transmitting this model directly, we use it to steer the NMF-based SBSS algorithm in the decoder: Given the NMF source model, the encoder calculates an initialization for the decoder NMF as well as a residual NMF model after running the decoder. In this paper, an additional step is introduced: Using the source NMF model, constraint values for each component are calculated which guide the decoder NMF at run-time. This procedure is explained in detail in Section IV.
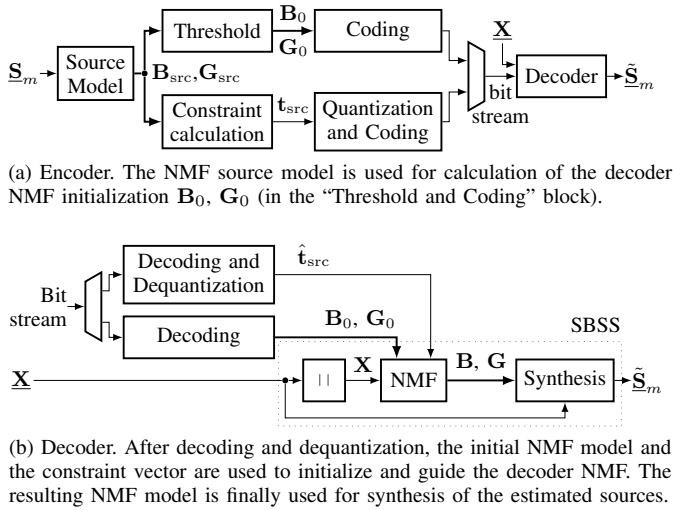
(a) Encoder. The NMF source model is used for calculation of the decoder NMF initialization $\mathbf{B}_0$, $\mathbf{G}_0$ (in the "Threshold and Coding" block).

(b) Decoder. After decoding and dequantization, the initial NMF model and the constraint vector are used to initialize and guide the decoder NMF. The resulting NMF model is finally used for synthesis of the estimated sources.

Fig. 2. Block diagram of NMF-ISS encoder and decoder in time-frequency domain without residual transmission. The mixture spectrogram $\underline{\mathbf{X}}$ is assumed to be perfectly known at both encoder and decoder.

Given the source spectrograms $\mathbf{S}_m$ and a number of NMF components I, the NMF source model is calculated as follows:

1) Estimate number of components per source with NMF with automatic relevance determination (ARD) [11]
2) Evaluate NMF[1] on each source $\mathbf{S}_m$ separately.
3) Concatenate the resulting NMF matrices in the component dimension to yield the NMF source model $\mathbf{B}_{\mathrm{src}}$, $\mathbf{G}_{\mathrm{src}}$.

The initial NMF model for the decoder NMF is calculated by simple thresholding ("Threshold" block in Figure 2a): In the following, the thresholding operation to yield $\mathbf{B}_0$ is shown. The calculation of $\mathbf{G}_0$ is done equivalently. First, the source basis matrix is expressed in dB with $b'_{\mathrm{src},fi} = 10 \log_{10}(b^2_{\mathrm{src},fi}/\max_{f,i} b^2_{\mathrm{src},fi})$. The initial basis matrix $\mathbf{B}_0$ is then obtained by thresholding with threshold $\tau_{\mathbf{B}_0}$ as $b_{0,fi} = 1$ if $b'_{\mathrm{src},fi} > \tau_{\mathbf{B}_0}$ and $b_{0,fi} = 0$ otherwise. The initial gain matrix $\mathbf{G}_0$ is calculated with threshold $\tau_{\mathbf{G}_0}$. The binary initial NMF model matrices are then run-length encoded with subsequent arithmetic encoding of the run-lengths [1], [12].

After running the decoder, the NMF source model is again used for determining a residual NMF model to enhance the separation quality further. In this paper, this step is left out for conciseness.

B. Decoder

The NMF-ISS decoder, depicted in Figure 2b, uses a semi-blind source separation (SBSS) algorithm based on the algorithm explained in Section II. After decoding and dequantization, the initial NMF model and the target vector is passed to the SBSS algorithm and used there for initialization of the NMF (refer to Section II).

---

[1] Nonnegative tensor factorization (NTF) could be used as well to describe the source spectrograms jointly. Components which describe multiple sources could deteriorate the decoder NMF when using NTF.

Given the number of components I, the performance of the SBSS algorithm strongly depends on the NMF parameter β which alters the reconstruction cost function d_β and α which weights the constraints as well as the choice of initial NMF matrices $\mathbf{B}_0$, $\mathbf{G}_0$. The NMF-ISS encoder consists of a full decoder, runs it with different choices for β, α and $\tau_{\mathbf{B}_0}, \tau_{\mathbf{G}_0}$ and chooses the parameter combination which results in the best SBSS performance.

IV. GENERALIZED NMF CONSTRAINTS

Here, we introduce a modification to constrained NMF: Instead of minimizing the constraint, thus driving the constraint towards zero, we propose to steer it towards a certain target value instead. We give mathematical formulations in Sections IV-A and IV-B and show in Section IV-C how to obtain the target values in an ISS setting.

A. Constraint Formulation

We propose to steer the constraint $c(\mathbf{g}_i)$ evaluated on each component to a certain target value t_i stored in a vector $\mathbf{t} \in \mathbb{R}^I$

$$\bar{c}_p(\mathbf{G}) = \sum_{i=1}^{I} |c(\mathbf{g}_i) - t_i|^p = \sum_i \bar{c}_p(\mathbf{g}_i) \ . \qquad (11)$$

The difference between the constraint $c(\mathbf{g}_i)$ and the target value t_i is normalized depending on $p > 0$. For $p = 1$ and $\mathbf{t} = \mathbf{0}$, Equation (11) becomes equal to Equation (5) if $c(\mathbf{g}_i) \geq 0$ for all i. Therefore, we call the novel constraint "generalized". Note that Equation (11) generalizes all constraints fulfilling the componentwise requirement (5). Therefore it is also possible to generalize the component-adaptive constraint $\tilde{c}$ described in Section II-C by replacing $c(\mathbf{g}_i)$ with $\tilde{c}(\mathbf{g}_i)$ in Equation (11).

B. Gradient Terms

In the following, we derive the gradient terms of $\bar{c}_p$ needed for the multiplicative NMF update rules given in Equation (8).

For $p > 0$, the gradient with respect to $\mathbf{G}$ becomes

$$[\nabla \bar{c}_p(\mathbf{G})]_{s,j} = p \cdot |c(\mathbf{g}_j) - t_j|^{p-1}$$
$$\cdot \operatorname{sgn}(c(\mathbf{g}_j) - t_j) \cdot [\nabla c(\mathbf{G})]_{sj} \qquad (12)$$

with $\operatorname{sgn}(x)$ denoting the sign function. Due to the chain rule, the gradient is depending on the gradient of the evaluated constraint, $\nabla_{\mathbf{G}} c(\mathbf{G})$.

In the following, we give the positive and negative gradient terms of the proposed constraint $\bar{c}_p(\mathbf{G})$ for $p = 1$ and $p = 2$.

For $p = 1$, the resulting gradient terms can be interpreted directly

$$\left[\nabla^+ \bar{c}_1(\mathbf{G})\right]_{sj} = \begin{cases} [\nabla^+ c(\mathbf{G})]_{sj} & \text{if } c(\mathbf{g}_j) > t_j \\ [\nabla^- c(\mathbf{G})]_{sj} & \text{otherwise.} \end{cases}$$

$$\left[\nabla^- \bar{c}_1(\mathbf{G})\right]_{sj} = \begin{cases} [\nabla^- c(\mathbf{G})]_{sj} & \text{if } c(\mathbf{g}_j) > t_j \\ [\nabla^+ c(\mathbf{G})]_{sj} & \text{otherwise.} \end{cases} \qquad (13)$$

If the value of $c(\mathbf{g}_i)$ exceeds the target t_i, the constraint is minimized (by using the original positive and negative gradient
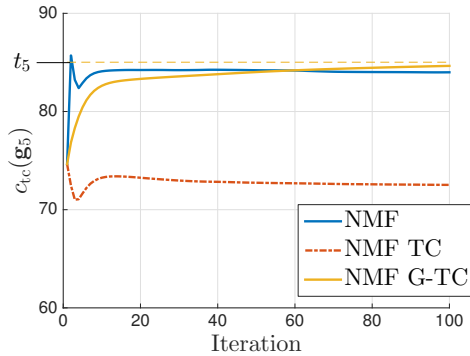
Fig. 3. Temporal continuity cost function $c_{tc}(\mathbf{g}_5)$ of the 5th NMF component calculated at NMF run-time with different constraint configurations for the exemplary trumpet-tambourine mixture.

terms). In the other case, the gradient terms are swapped to maximize the constraint until the target value t_i is exceeded.

For $p = 2$, the following gradient terms result:

$$\left[\nabla^+ \bar{c}_2(\mathbf{G})\right]_{sj} = 2 \left\{ c(\mathbf{g}_j) \left[\nabla^+ c(\mathbf{G})\right]_{sj} + t_j \left[\nabla^- c(\mathbf{G})\right]_{sj} \right\}$$
$$\left[\nabla^- \bar{c}_2(\mathbf{G})\right]_{sj} = 2 \left\{ c(\mathbf{g}_j) \left[\nabla^- c(\mathbf{G})\right]_{sj} + t_j \left[\nabla^+ c(\mathbf{G})\right]_{sj} \right\} . \tag{14}$$

To constrain NMF with the generalized constraint results in replacing the gradient terms of $c(\mathbf{G})$ in the multiplicative update rules (8) with the terms of $\bar{c}_p(\mathbf{G})$ given in Equation (13) and (14).

C. Application to NMF-based Informed Source Separation

In the context of NMF-based informed source separation, the calculation of the target values $\mathbf{t}$ is straightforward by using the source NMF model. In the NMF-ISS encoder, the constraint value for the ith column of the source gain matrix $\mathbf{G}_{src}$ is calculated ("Constraint calculation" in Figure 2a) as

$$t_{src,i} = c(\mathbf{g}_{src,i}) . \tag{15}$$

To indicate the influence of the sources, we denote the target vector as $\mathbf{t}_{src}$. To transmit the target values $\mathbf{t}_{src}$ to the decoder, each value is scalar quantized and the resulting symbols are bit encoded by adaptive arithmetic encoding [12]. Note that the additional bit rate is rather small because only I scalar values have to be transmitted.

D. Example

Figure 3 illustrates the impact of the generalized NMF constraints used in the NMF-ISS decoder NMF for the exemplary mixture shown in Figure 1. We evaluated the temporal continuity (TC) constraint in Equation (3) on the 5th component modeling parts of the percussive tambourine sound. The TC function $c_{tc}(\mathbf{g}_5)$ is plotted over 100 NMF iterations. Although TC is usually not suitable to model percussive components, we use it here to show that the proposed generalized TC constraint prevents the unwanted behaviour of the standard TC constraint: The target value t_5 is insignificantly higher than $c_{tc}(\mathbf{g}_5)$ calculated while running unconstrained NMF

("NMF"). NMF with TC constraint ("NMF TC") damps c_{tc} unnecessarily whereas the proposed generalized TC constraint (NMF "G-TC" with $p = 2$) preserves the original c_{tc} value as it converges to t_5.

The separation quality (measured with signal-to-distortion ratio, SDR, refer to Section V-A) is slightly increased by $0.2\,\mathrm{dB}$ when enabling TC compared to unconstrained NMF for the exemplary mixture. NMF with the proposed generalized constraints increases SDR significantly by around $2\,\mathrm{dB}^2$.

V. Experiments

A. Setup

For evaluation of the proposed method, NMF-ISS is performed on five monaural mixtures sampled at $44\,100\,\mathrm{Hz}$ taken from the QUASI database[3]. The mixtures consist of 3 to 6 sources (e.g. vocals, guitar, drums, effects) and are about $20\,\mathrm{s}$ long. As quality measure, the signal-to-distortion ratio (SDR) is calculated using the "BSS Eval" toolbox [13]. The mean SDR is calculated over all sources per mixture in reference to the performance of an oracle estimator [14] which yields an upper bound for separation with Wiener-like filtering. The resulting measure is denoted as δSDR and given for bit rate R which is normalized per source.

Regarding the STFT, we chose a window size of $93\,\mathrm{ms}$ with $50\,\%$ overlap. Additionally, we filtered the spectral dimension of all spectrograms with a Mel-filterbank to speed up the following computation steps and decrease the parameter bit rate. We used $F = 400$ Mel-filters.

The encoder is tested with different numbers of components I per source $I/M \in \{2, 3, 4, 5, 10, 15, 20, 30\}$ with M denoting the number of sources. All $(R, \delta\text{SDR})$ points were optimized per mixture and I/M independently and then smoothed using the locally weighted scatter plot smoothing (LOESS) method to obtain rate/quality curves.

The encoder estimates optimal SBSS parameters at run-time by testing the SBSS algorithm with combinations of the following parameters: Regarding NMF, the β-Divergence parameter and constraint weights are chosen as $\beta \in \{0, 1, 2\}$ and $\alpha \in \{0\} \cup \{10^{-3}, 10^{-2.5} \ldots 10^3\}$. The thresholds for obtaining the initial NMF model $\mathbf{B}_0$, $\mathbf{G}_0$ are either set to $\tau_{\mathbf{B}_0}, \tau_{\mathbf{G}_0} \in \{-15, -30, -60\}\,\mathrm{dB}$ or determined component-wise as $\sum_f b_{src,fi}/F$ or $\sum_t g_{src,ti}/T$. The target values $t_{src,i}$ are coded with $8\,\mathrm{bit}$.

B. Comparison to standard constraints

First, we compare the generalized constraints to standard constraints and evaluate the impact of the normalization factor p used in Equation (11). Therefore, we conducted NMF-ISS without residual transmission. The decoder NMF was either evaluated with temporal continuity or sparseness constraints ("NMF TC" or "NMF S", cf. Equations (3) and (4)) and compared to unconstrained NMF ("NMF"). Additionally, NMF was performed with the proposed generalized temporal

---

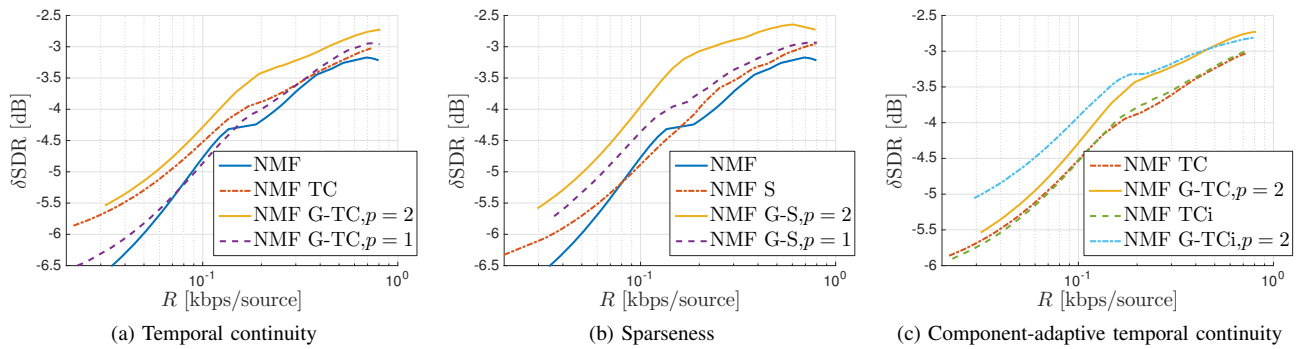[2]Note that we chose the corresponding constraint weights for NMF TC and NMF G-TC maximizing the SDR.

[3]http://www.tsi.telecom-paristech.fr/aao/en/2012/03/12/quasi/

Fig. 4. Separation quality in δSDR. Comparison to standard constraints and to component-adaptive temporal continuity constraint as well as influence of parameter p for proposed generalized constraints.

continuity ("NMF G-TC") and sparseness constraints ("NMF G-S") with normalization factor $p \in \{1, 2\}$ (cf. Equation (11)).

Figure 4a shows the δSDR results for TC: The generalized TC yields worse results than TC for $p = 1$; the performance is similar to unconstrained NMF. For mid-range and high bit rates, G-TC outperforms TC with $p = 2$.

Regarding sparseness, the generalized constraint gives better results than the standard constraint for $p = 1$ and $p = 2$ as depicted in Figure 4b. In the case of $p = 2$, a gain of around $1\,$dB for mid-range bit rates can be observed. In this configuration, the generalized TC constraint for $p = 2$ is outperformed by generalized S constraint with $p = 2$ for all bit rates.

C. Generalized component-adaptive constraints

In this section, we compare the proposed generalized constraints to component-adaptive constraints (cf. Section II-C) and also apply the generalization to these constraints.

Figure 4c shows the corresponding results for the temporal continuity constraint. The component-adaptive constraint ("TCi") gives similar results compared to the standard TC constraint ("TC"). The proposed generalized constraint on TC ("G-TC") outperforms TCi and yields an δSDR increase of about $0.5\,$dB for mid- and high-range bit rates. The generalization of TCi ("G-TCi") gives even better results than G-TC for low and mid-range bit-rates (increase of about $0.5\,$dB).

We also conducted experiments for the sparseness (S) constraint which confirm the results obtained for TC: The generalized S constraint gives better results than the component-adaptive S constraint. The quality difference is similar to the difference between "NMF S" and "NMF G-S" for $p = 2$ in Figure 4b. Generalization of the component-wise sparseness constraint does not increase separation quality.

VI. CONCLUSION

In this paper, we proposed to steer constraints for NMF to certain target values componentwise. We derived the corresponding terms necessary to modify the NMF update rules and evaluated these generalized constraints in an informed source separation setup. We showed in experiments, that the generalized constraints outperform a set of existing NMF constraints.

Future work could include evaluating the novel constraints in a more semi-blind or blind setup by estimating the target values at run-time.

REFERENCES

[1] C. Rohlfing, J. Becker, and M. Wien, "NMF-based informed source separation," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 474–478.

[2] T. Virtanen, "Monaural sound source separation by nonnegative matrix factorization with temporal continuity and sparseness criteria," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 3, pp. 1066–1074, 2007.

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

[4] J. Becker and C. Rohlfing, "Component-adaptive priors for NMF," in *Latent Variable Analysis and Signal Separation*. Springer, Heidelberg/Berlin, Aug. 2015.

[5] M. Parvaix, L. Girin, and J.-M. Brossier, "A watermarking-based method for informed source separation of audio signals with a single sensor," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1464–1475, 2010.

[6] A. Liutkus, R. Badeau, and G. Richard, "Informed source separation using latent components," in *Latent Variable Analysis and Signal Separation*. Springer, 2010, pp. 498–505.

[7] A. Ozerov, A. Liutkus, R. Badeau, and G. Richard, "Coding-based informed source separation: Nonnegative tensor factorization approach," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 8, pp. 1699–1712, 2013.

[8] A. Liutkus, S. Gorlow, N. Sturmel, S. Zhang, L. Girin, R. Badeau, L. Daudet, S. Marchand, and G. Richard, "Informed audio source separation: A comparative study," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 2397–2401.

[9] M. Spiertz and V. Gnann, "Beta divergence for clustering in monaural blind source separation," in *128th AES Convention*, London, UK, May 2010.

[10] M. Spiertz, *Underdetermined Blind Source Separation for Audio Signals*, ser. Aachen Series on Multimedia and Communications Engineering. Aachen: Shaker Verlag, July 2012, vol. 10.

[11] V. Y. Tan and C. Févotte, "Automatic relevance determination in nonnegative matrix factorization with beta-divergence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 7, pp. 1592–1605, 2013.

[12] J.-R. Ohm, *Multimedia Signal Coding and Transmission*, ser. Signals and Communication Technology. Springer-Verlag Berlin Heidelberg, 2015.

[13] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, July 2006.

[14] E. Vincent, R. Gribonval, and M. D. Plumbley, "Oracle estimators for the benchmarking of source separation algorithms," *Signal Processing*, vol. 87, no. 8, pp. 1933–1950, Dec 2007.