Regularization of Reverberant Multichannel Audio

İlker Bayram and Aziz Koçanaoğulları

***Abstract*—Different classes of natural audio signals exhibit structured sparse appearance in the time-frequency domain. However, recordings obtained in reverberant environments are also affected by the room impulse response. Consequently, structured sparse time-frequency models that do not take into account reverberation effects are not directly feasible for such recordings. When the room impulse responses are unknown, enforcing structured sparsity on the underlying source signal calls for estimating the room impulse response first, which can be challenging. In this paper, we propose a regularization function for reverberant multichannel audio signals that does not require to estimate the room impulse responses. We demonstrate the utility of the proposed regularization function on a denoising formulation and discuss other potential applications.**

I. Introduction

Sparse or structured sparse models have found application in audio processing, due to the appearance of audio signals in time-frequency domains such as the short-time Fourier transform (STFT) or constant-Q transforms (see e.g. [29, 21, 5, 20, 9, 8] for a sample of a vast literature). However, if an audio source is recorded in a reverberant room, the recording is also affected by the room impulse response (RIR), and the resulting time-frequency representation ceases to be sparse. This is illustrated in Fig. 1. The left panel shows the spectrogram of an audio source. This spectrogram contains special structures due to the harmonics, but these structures are scarcely distributed in the image, rendering valid a structured sparse model. The spectrogram on the right belongs to the same signal recorded in a reverberant room. Reverberation smears the harmonics and the appearance is no longer (structured) sparse. If and when the RIR is known, it is possible to employ sparse models for reverberant signals. However, in many practical scenarios, information about the RIR is not available. In addition, RIRs are known to vary a lot with respect to the positions of the source and/or the microphone. Consequently, in a dynamic scene (with a moving source or microphone), one needs to constantly keep track of the RIRs. Therefore, it is of interest to work under the assumption that the RIR is unknown. But then, in order to employ sparse models, one needs to resort to blind methods, which estimate the signal and the RIR jointly. This in turn leads to non-convex problems, extracting good solutions from which can be challenging. Fortunately, when multiple microphones are used, it is possible to avoid such blind formulations by making use of the relation between the different RIRs. In this paper, we thus propose a *convex* regularizer for multichannel *reverberant* signals.

In order to set the stage, consider recordings of an audio source with M microphones in a reverberant room. Let

The authors are with the Department of Electronics and Communications Engineering, Istanbul Technical University, Istanbul, Turkey. E-mail : ibayram@itu.edu.tr, azizkocana@gmail.com. This research is supported by TÜBİTAK (Project No : 113E511).

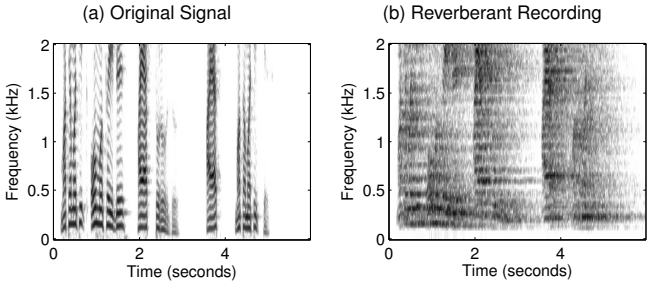

Fig. 1. (a) The spectrogram of a speech signal, (b) the spectrogram of the recording of the signal in (a) played in a reverberant room.

$y_m(n,s)$ and $z(n,s)$ denote the short-time Fourier transform (STFT) coefficients of the m^{th} observation and the source at the time-frequency point (n,s). Neglecting the effect of noise, at the s^{th} frequency band, y_m and z are related to each other approximately through,

$$y_m(n,s) \approx \sum_t z(n-t,s)\, h_{m,s}(t), \tag{1}$$

where convolution with $h_{m,s}$ in the STFT domain approximates the effect of convolution with the the RIR in the time-domain followed by the STFT [27]. The length of the filter $h_{m,s}$ depends on the the duration of the RIR as well as the STFT parameters and can extend up to a few tens of non-zero samples. Therefore, even if z is sparse, y_m may fail to be sparse due to smoothing with $h_{m,s}$. In such a setting, if the observations are noisy, then in order to reduce the effect of noise, we could use a formulation of the form

$$\min_x \sum_{m=1}^{M} \frac{1}{2}\|y_m - H_m\, x\|_2^2 + p(x), \tag{2}$$

where p is a sparsity inducing prior and H_m is a reverberation operator in agreement with (1). In fact, a time-domain version of this formulation was discussed in [21]. Unfortunately, this formulation is hindered by the fact that, in many practical scenarios, h_m, therefore H_m, is unknown and is hard to predict, especially for frequencies above a few hundred Hertz [13, 12, 24]. Thus, even though the sparsity of x is valuable prior information, it cannot be directly utilized in this setting. In this scenario, even though we are interested in *denoising* y_m, utilization of the sparsity of x calls for *dereverberation*. The regularizer proposed in this paper allows to denoise without dereverberating and does not require knowledge of the room impulse responses.

The motivation of the proposed regularizer comes from the properties of relative transfer functions (RTF), which relate RIRs from different microphones (see Fig. 2). Even though RIRs belonging to different microphones can differ significantly, the RTFs are fast decaying sequences [28, 19, 18]. This in turn implies that the observation from a microphone can be represented as a linear combination of a few time-shifted

recording from another microphone. Based on this observation, we propose to construct a matrix using the observations such that many singular values of the matrix are close to zero. This construction can be achieved by applying a linear operator to the set of observations. In order to penalize matrices with a complete set of non-zero singular values, we employ the nuclear norm on the constructed matrix to define a regularization function. Similar to the model in (1), the regularizer is also separable with respect to frequency bands and it does not make use of the relationship between different frequency bands.

Related Work and Contribution

The idea of collecting together similar patches and using the nuclear norm for enforcing a low rank was used for image/video denoising in [16, 11, 26]. However, these papers collect together different patches from the image by a search. Therefore, the 'linear operator' that collects together the patches is signal dependent. Consequently, the overall approach is not convex. In contrast, we use a fixed linear operator and the resulting regularization function is convex.

As noted above, although typical room impulse responses can be long, the relative transfer functions that relate observations of the sources from different microphones can be represented with short filters. This observation, which motivates the proposed regularizer has been utilized in recent work [28, 19, 18] to estimate RTFs. However, these papers do not aim to propose regularizing functions for reverberant signals and differ significantly from the approach in this paper. In fact, we do not estimate the RTFs at all and only make use of the parsimony of the RTFs for defining a regularization function.

Finally, the nuclear norm, which constitutes one of the two main ingredients for our regularizer, has been previously used in audio processing for denoising and component separation [15, 5]. However, the low rank matrices considered in these works are motivated by common features found in music (like a repetitive background) and are not related with reverberation. In addition, the examples of recent work on audio processing cited above mainly work with a single signal and do not consider the interactions between different observations of a source. In contrast, the regularization function proposed in this paper targets multichannel audio and it exploits the unknown correlation between the different recordings. In addition, our target is signal is reverberant because we would like to avoid blind formulations that call for non-convex formulations. As far as we are aware, a regularizer for reverberant multichannel audio with unknown RIRs has not been proposed in the literature.

General Notation

As noted above, the regularizer proposed in this paper is separable with respect to the STFT frequency bands. Therefore, in order to simplify notation, we assume that the frequency band is fixed and do not use an index to specify the frequency band. We assume throughout the paper that there are M microphones producing M reverberant observations. The STFT coefficients of the m^{th} observation at the selected frequency band is denoted by $y_m(n)$. We take y_m to be a column vector. We also define Y to be the matrix with Mcolumns as

$$Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_M \end{bmatrix}. \tag{3}$$

The set of $n \times m$ matrices with complex entries is denoted by $\mathbb{C}^{n,m}$. For $\mathbb{C}^{n,m}$, we use an inner product defined as

$$\langle X, Y \rangle = \operatorname{Re}\left(\sum_{k,j} X_{k,j}\, Y^*_{k,j} \right). \tag{4}$$

Notice that $\|X\|_F^2 = \langle X, X \rangle$, where $\|\cdot\|_F$ denotes the Frobenius norm.

Suppose $X \in \mathbb{C}^{N,M}$ and $J = \{j_1, \ldots j_{|J|}\}$ is an ordered collection of integers in the range $[1, N]$, where $|J|$ denotes the cardinality of J. Then X_J denotes the $|J| \times M$ matrix whose k^{th} row is equal to the j_k^{th} row of X.

Some of the frequently used symbols and terms (some of which will be defined in the sequel) are listed in Table I.

TABLE I
FREQUENTLY USED TERMS/SYMBOLS

M : Number of observations (microphones)
$\mathbb{C}^{n,m}$: set of $n \times m$ complex matrices
$\sigma(X)$: greatest singular value of X.
B : set of matrices defined in (18)
S_k : Advance operator by k samples – see (12)
$S\,x = \begin{bmatrix} S_{-K}(x) & S_{-K+1}(x) & \cdots & S_K(x) \end{bmatrix}$
x_i : i^{th} observation vector (length N) obtained from the i^{th} mic.
$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \end{bmatrix}$
$\mathcal{S}X = \begin{bmatrix} Sx_1 & Sx_2 & \cdots & Sx_M \end{bmatrix}$
$\mathcal{J}$: a partition of the set of integers in the range $[1, N]$
X_J : Matrix obtained by selecting the rows of X indexed by J
SVC : Singular value clipping operator – see Defn. 1

Outline

In Section II, we recall the definition of the nuclear norm and discuss some of its properties that will be used. A motivation for the proposed regularization function, followed by a formal definition and some discussion is provided in Section III. Section IV discusses the application of the proposed regularizer in a denoising scenario. Section V is the conclusion.

II. PRELIMINARY ON THE NUCLEAR NORM

The nuclear norm is one of the main ingredients for the proposed regularization function. In this section, we briefly recall some facts about the nuclear norm, that will be used in the sequel.

The rank of a matrix A may be interpreted as the ℓ_0 count of the singular values of A. The rank does not constitute a norm for matrices and it is a non-convex function on $\mathbb{C}^{n,m}$. On the other hand, the ℓ_1 norm of the singular values turns out to be a matrix norm and is referred to as the nuclear norm, trace norm or Schatten-1 norm [14]. The nuclear norm of A is denoted in this paper as $\|A\|_*$. The relation between the

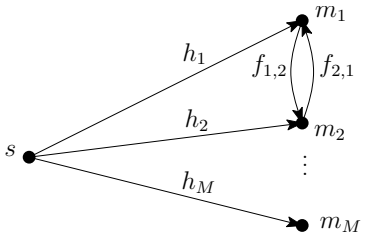

Fig. 2. The scene of interest for the proposed regularization function. Here, s denotes the source and m_k denotes the k^{th} microphone. The impulse response from the source to the k^{th} microphone is denoted by h_k and the relative transfer function from the k^{th} microphone to the j^{th} microphone is denoted by $f_{k,j}$.

nuclear norm and the rank is similar to the relation between the ℓ_1 norm and the ℓ_0 count. We refer to [4] for a discussion of this aspect.

The dual norm of the nuclear norm, defined as

$$\|Z\|_D = \sup_{\|X\|_* \leq 1} \langle X, Z \rangle, \tag{5}$$

turns out to be the spectral norm $\sigma(Z)$ (that is, the greatest singular value of Z) [14]. This gives a dual characterization of the nuclear norm as,

$$\|X\|_* = \sup_{\sigma(Z) \leq 1} \langle X, Z \rangle. \tag{6}$$

The proximity operator associated with a convex function g is denoted by $J_{\alpha g}$ and is defined as [6]

$$J_{\alpha g}(y) = \arg\min_x \frac{1}{2\alpha}\|x - y\|_2^2 + g(x). \tag{7}$$

In the current context, given $Y \in \mathbb{C}^{m,n}$, the proximity operator of $g(\cdot) = \|\cdot\|_*$ is given as

$$J_{\alpha g}(Y) = \arg\min_{X \in \mathbb{C}^{m,n}} \frac{1}{2\alpha}\|Y - X\|_F^2 + \|X\|_*. \tag{8}$$

To describe the solution, let the SVD of Y be $Y = U\,\Sigma\,V^*$, where Σ is the diagonal matrix holding the singular values of Y. Also let $\tilde{\Sigma}$ be the $m \times n$ diagonal matrix obtained by applying a soft threshold to the diagonal entries of Σ. That is,

$$\tilde{\Sigma}_{i,i} = \max\big(\Sigma_{i,i} - \alpha, 0\big). \tag{9}$$

Then, $J_{\alpha g}(Y) = U\,\tilde{\Sigma}\,V^*$. The mapping that takes Y to $U\,\tilde{\Sigma}\,V^*$ is called 'singular value thresholding' [3] and is denoted as $\mathrm{SVT}_\alpha(Y)$.

In the following, we will introduce a new operation called singular value clipping and discuss its relation with SVT.

III. The Proposed Regularization Function

Before giving the definition of the proposed regularization function, we start with some motivating discussion.

A. Motivation for the Proposed Regularizer

Recall that y_m denotes the STFT coefficients of a specific frequency band for the m^{th} observation. Let z be the STFT coefficients of the original source for the same frequency band. Neglecting the effects of noise, y_m is related to z through some filter h_m as $y_m \approx z * h_m$ (see Fig. 2). We remark that even if the microphones are close to each other, h_m and h_j may differ significantly [25]. Nevertheless, the relative transfer function (RTF) mapping y_m to y_j is associated with a filter that decays fast [28, 19, 18]. Therefore we can write $h_j \approx h_m * f_{m,j}$ for some short filter $f_{m,j}$. This is illustrated in Fig. 3. Fig. 3a shows the real part of the impulse responses h_1, h_2, for a specific STFT band, obtained with microphones close to each other. Notice that the two sequences differ noticeably. Using these h_i's, we obtained a non-causal RTF $f_{2,1}$ with six taps, by solving

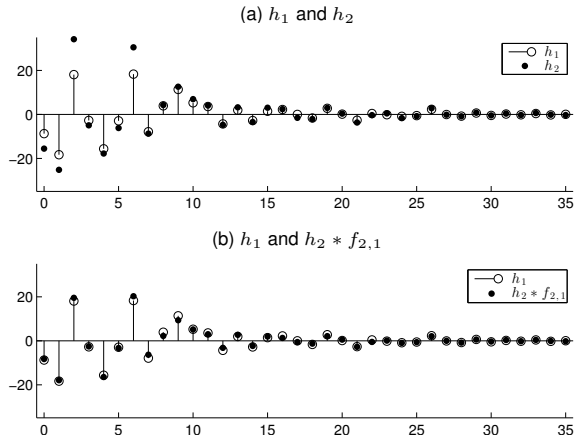

Fig. 3. (a) The impulse responses h_1 and h_2 in the STFT domain for a specific channel. (b) h_1 and its approximation $h_2 * f_{2,1}$ for a filter $f_{2,1}$ with six non-zero coefficients.

$$f_{2,1} = \arg\min_f \|h_1 - h_2 * f\|_2^2. \tag{10}$$

Fig. 3b shows h_1 and $h_2 * f_{2,1}$. In contrast to the sequences in Fig. 3a, these sequences are much more closer. We remark that, the formulation (10) is, although plausible, somewhat arbitrary and one could possibly find better RTFs for the same number of taps by solving, for instance, a weighted ℓ_∞ minimization formulation. Here, our purpose is not to find the best possible RTF, but rather to show that short RTFs exist. We are not going to use the RTFs in the proposed regularization function.

The approximate equality $h_j \approx h_m * f_{m,j}$ implies that $y_j \approx y_m * f_{m,j}$. Let us assume for the sake of discussion that $f_{m,j}$ is causal and has K non-zero coefficients. Consider the $N \times (K+1)$ matrix obtained by augmenting y_j and the shifts of y_m as,

$$\begin{bmatrix} y_j(n) & y_m(n) & y_m(n-1) & \dots & y_m(n-K+1) \end{bmatrix}. \tag{11}$$

Since $y_j(n)$ can be approximately written as a linear combination of $y_m(n), \dots, y_m(n-K+1)$, (where $n = 0, \dots, N-1$), at least one of the singular values of this matrix will be close to zero. Suppose now that we collect together the shifts of all the observations and form a larger matrix. By the foregoing discussion, this matrix, which has yet more columns, will have a number of singular values close to zero. This observation is the main motivation behind the definition given in the sequel.

B. Definition of the Proposed Regularizer

We now provide a precise description of the proposed regularization function. For this, suppose x_m denotes the STFT coefficients of a specific channel for the m^{th} observation, where

$$X = \begin{bmatrix} a & e \\ b & f \\ c & g \\ d & h \end{bmatrix} \Longrightarrow \mathcal{S}X = \begin{bmatrix} 0 & a & b & 0 & e & f \\ a & b & c & e & f & g \\ b & c & d & f & g & h \\ c & d & 0 & g & h & 0 \end{bmatrix}$$

Fig. 4. Demonstration of the operator $\mathcal{S}$ for $M = 2$, $K = 1$, $N = 4$.

$m = 1, 2, \ldots, M$. We take x_m to be a column vector of length-N, indexed as $x_m^T = \begin{bmatrix} x_m(0) & x_m(1) & \ldots & x_m(N-1) \end{bmatrix}$.

In this setting, for $k \geq 0$, we define S_k to be the linear operator that advances a length-N vector x by k samples and inserts k zeros at the end so that $S_k(x)$ is also length-N. More precisely, if $u = S_k(x)$, then

$$u(n) = \begin{cases} x(n+k), & \text{if } 0 \leq n < N-k, \\ 0, & \text{if } N-k \leq n \leq N-1. \end{cases} \tag{12}$$

Note that $S_0 = I$. For $k < 0$, we define $S_k = S_{|k|}^T$. Thus, for $k < 0$, S_k delays the signal by $|k|$ samples and inserts $|k|$ zeros at the onset.

For a fixed K, we define S to be the operator that maps a length-N signal x to

$$S\,x = \begin{bmatrix} S_{-K}(x) & S_{-K+1}(x) & \ldots & S_K(x) \end{bmatrix}. \tag{13}$$

Notice that according to our description, $S\,x$ is a matrix of size $N \times (2K+1)$. Now, let

$$X = \begin{bmatrix} x_1 & x_2 & \ldots & x_M \end{bmatrix} \tag{14}$$

and define $\mathcal{S}$ to be the operator that applies to X, the action of which can be described as,

$$\mathcal{S}\,X = \begin{bmatrix} S\,x_1 & S\,x_2 & \ldots & S\,x_M \end{bmatrix}. \tag{15}$$

Notice that $\mathcal{S}\,X$ is an $N \times M \cdot (2K+1)$ matrix. Examples demonstrating the action of the operators $\mathcal{S}$ and $\mathcal{S}^T$ are provided in Fig. 4 and Fig. 5.

By the discussion in Sec. III-A, we expect $\mathcal{S}X$to have rank that is much smaller than the number of its columns. In order to penalize a high rank, we could therefore employ the nuclear norm on $\mathcal{S}X$ and obtain a convex regularizer. If the source and the microphones are stationary, then such a regularizer might be feasible. However, in order to cope with a dynamical scene, we can also consider partitioning the signals along the time-axis. For this, let $\mathcal{N}$ denote the set of integers in the range $[1, N]$ and suppose $\mathcal{J}$ is a partition of $\mathcal{N}$, where each element of $\mathcal{J}$contains integers in some interval. That is $\mathcal{J} = \{J_1, \ldots, J_m\}$, where each $J_i \subset \mathcal{N}$ consists of the integers in an interval such that $J_i \cap J_k = \emptyset$ when $i \neq k$ and $\cup_i J_i = \mathcal{N}$. In this setting, the proposed regularization function is defined as,

$$q(X) = \sum_{J \in \mathcal{J}} \left\| (\mathcal{S}X)_J \right\|_*. \tag{16}$$

A dual description of this function will also be used in the sequel. By the discussion in Sec. II, we obtain a dual description as,

$$q(X) = \sum_{J \in \mathcal{J}} \max_{Z \in B_\sigma} \langle (\mathcal{S}X)_J, Z \rangle. \tag{17}$$

$$Z = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \end{bmatrix}$$

$$\mathcal{S}^T Z = \begin{bmatrix} a_{12} + a_{21} & a_{15} + a_{24} \\ a_{31} + a_{22} + a_{13} & a_{34} + a_{25} + a_{16} \\ a_{41} + a_{32} + a_{23} & a_{44} + a_{35} + a_{26} \\ a_{42} + a_{33} & a_{45} + a_{36} \end{bmatrix}$$

Fig. 5. Demonstration of the operator $\mathcal{S}^T$ for $M = 2$, $K = 1$, $N = 4$. Notice that each column of $\mathcal{S}^T Z$ is obtained by partitioning the columns of Z and summing along the antidiagonals.

Given the partition $\mathcal{J}$, let us define a set of matrices

$$B = \left\{ Z \in \mathbb{C}^{N, M \cdot (2K+1)} : \sigma(Z_J) \leq 1 \text{ for all } J \in \mathcal{J} \right\}. \tag{18}$$

We remark that B is a convex set. Using B, we can rewrite (17) as,

$$q(X) = \max_{Z \in B} \langle \mathcal{S}X, Z \rangle = \max_{Z \in B} \langle X, \mathcal{S}^T Z \rangle. \tag{19}$$

C. Validating the Assumption on the Singular Values of $\mathcal{S}X$

Let us now numerically check if the assumptions about the singular values of $\mathcal{S}X$ are valid. For this, we recorded a source with eight microphones in a reverberant room. These recordings comprise our clean observations. We computed the STFT coefficients of the observations, chose a specific band and formed the matrix X. Then for $K = 1$ and $K = 6$, we computed the singular values of $\mathcal{S}X$. Note that the number of columns in $\mathcal{S}X$ is $8 \cdot (2K+1) = 24$ for $K = 1$ and 104 for $K = 6$. The number of rows is greater than the number of columns so that the columns determine the upper bound for the rank. For both choices of K, we computed the singular values of $\mathcal{S}X$. These singular values, in decreasing order, are shown in Fig. 6 (thick lines). We observe that for a higher K value, the ratio of the significant singular values to the number of columns is lower. This is in agreement with the claim in Sec. III-A. However, in order for the proposed regularization function to be useful, it should also be able to help separate noisy observations. To test this, we added white Gaussian noise to the time-domain signals and repeated the computations. If we denote the new observation matrix in the STFT domain as Y, the SNR of Y is 10 dB. The singular values of $\mathcal{S}Y$ for $K = 1$ and $K = 6$ are shown in Fig. 6 (thin lines). We observe that the singular values of $\mathcal{S}Y$ decrease significantly slower than the singular values of $\mathcal{S}X$, justifying the proposed definition.

D. Properties of the Operator $\mathcal{S}$

We end this section with two properties of $\mathcal{S}$, which will be useful in the following discussions.

We first consider the operator $\mathcal{S}^T\mathcal{S}$. It turns out that this operator is diagonal.

Lemma 1. For an X with N rows, where $N > 2K$, $\mathcal{S}^T\mathcal{S}\,X = \Lambda X$, where Λ is a diagonal matrix whose diagonal

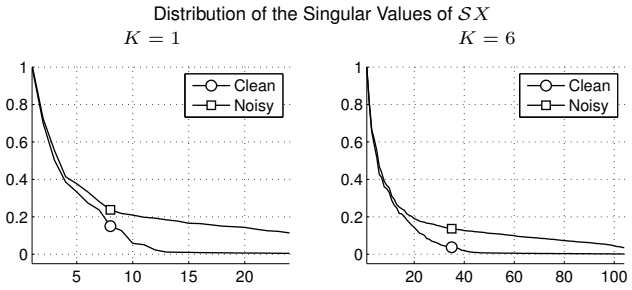

Fig. 6. Distribution of the Singular Values of $\mathcal{S}X$ for (a) $K = 1$, (b) $K = 6$.

entries are,

$$\Lambda_{k,k} = \begin{cases} K + k, & \text{if } 1 \le k \le K, \\ 2K + 1, & \text{if } K + 1 \le k \le N - K, \\ N + K + 1 - k, & \text{if } N - K + 1 \le k \le N. \end{cases} \quad (20)$$

Proof. Suppose x and y are length-N signals with $N > 2K$. Observe that

$$\langle S_k\, x, S_k y\rangle = \begin{cases} \sum\limits_{n=0}^{N-|k|-1} \operatorname{Re} x(n)\, y^*(n), & \text{if } k < 0, \\ \sum\limits_{n=k}^{N-1} \operatorname{Re} x(n)\, y^*(n), & \text{if } 0 \le k. \end{cases} \quad (21)$$

Therefore $S_k^T\, S_k\, x = D^{(k)}\, x$ where

$$D^{(k)} = \begin{cases} \operatorname{diag}\big(\underbrace{1, \ldots, 1}_{N-|k| \text{ ones}}, \underbrace{0, \ldots, 0}_{k \text{ zeros}}\big), & \text{if } k < 0, \\ \operatorname{diag}\big(\underbrace{0, \ldots, 0}_{k \text{ zeros}}, \underbrace{1, \ldots, 1}_{N-k \text{ ones}}\big), & \text{if } 0 \le k. \end{cases} \quad (22)$$

Observe that

$$S^T\, S\, x = \sum_{k=-K}^{K} S_k^T\, S_k x = \left(\sum_{k=-K}^{K} D^{(k)} \right) x = \Lambda\, x, \quad (23)$$

where Λ is the diagonal operator defined in (20). Finally, let $X, Y, \in C^{N,M}$ and let x_m and y_m denote the m^{th} column of X and Y respectively. Then, we have

$$\langle \mathcal{S}^T \mathcal{S} X, Y\rangle = \sum_{m=1}^{M} \langle \Lambda x_m, y_m\rangle = \langle \Lambda\, X, Y\rangle. \quad (24)$$

Thus follows the claim. □

As a corollary of this lemma, we can write down the spectral norms of $\mathcal{S}^T\, \mathcal{S}$ and $\mathcal{S}\, \mathcal{S}^T$ easily.

Corollary 1. If $N > 2K$, then $\sigma(\mathcal{S}^T\, \mathcal{S})$ and $\sigma(\mathcal{S}\, \mathcal{S}^T)$ are equal to $2K + 1$.

IV. Denoising with the Proposed Regularizer

We now discuss the application of the proposed regularization function for denoising reverberant observations, without having to estimate RIRs. For this, we consider the formulation

$$\min_{X \in \mathbb{C}^{N,M}} \frac{1}{2} \|Y - X\|_F^2 + \lambda\, q(X), \quad (25)$$

where the columns of Y contain observations from different microphones. The problem (25) is strictly convex and has a unique minimizer $\hat{X}$. In order to derive an algorithm that obtains $\hat{X}$, we study the dual problem.

A. Denoising Algorithm via the Dual Problem

Using the dual description of the proposed prior given in (19), the problem in (25) may be expressed as a saddle point problem as

$$\min_{X \in \mathbb{C}^{n,m}} \max_{Z \in B} \frac{1}{2} \|Y - X\|_F^2 + \langle X, \lambda\, \mathcal{S}^T Z\rangle \quad (26)$$

Changing the order of minimization and maximization, we obtain the dual problem as,

$$\min_{Z \in B} g(Z), \quad (27)$$

where

$$g(Z) = -\left[\min_X \frac{1}{2} \|Y - X\|_F^2 + \langle X, \lambda\, \mathcal{S}^T Z\rangle \right] \quad (28)$$

The solution of the minimization problem in (28), namely $\hat{X}$ satisfies,

$$\hat{X} = Y - \lambda\, \mathcal{S}^T\, Z. \quad (29)$$

Plugging this for X in the expression for $g(Z)$, we find

$$\begin{aligned} g(Z) &= -\frac{1}{2} \|\lambda \mathcal{S}^T Z\|_F^2 - \langle Y, \lambda\, \mathcal{S}^T Z\rangle \\ &\qquad + \langle \lambda\, \mathcal{S}^T Z, \lambda\, \mathcal{S}^T Z\rangle \quad (30) \\ &= \frac{1}{2} \|Y - \lambda\, \mathcal{S}^T Z\|_F^2 - \frac{1}{2} \|Y\|_F^2. \quad (31) \end{aligned}$$

Discarding the constant with respect Z, the dual problem may be written as,

$$\min_{Z \in B} \frac{1}{2} \|Y - \lambda \mathcal{S}^T Z\|_F^2. \quad (32)$$

This problem asks to project Y onto $\lambda \mathcal{S}^T\, B$. Even though this projection is unique, since $\mathcal{S}^T$ has a non-trivial null-space, the problem in (32) may have multiple solutions. Nevertheless, a solution always exists because the convex cost function in (32) is bounded from below. The foregoing discussion may be summarized as follows.

Proposition 1. Let $\hat{X}$ denote the solution of (25). Also let $\hat{Z}$ denote a solution of (32). Then $\hat{X} = Y - \lambda \mathcal{S}^T\, \hat{Z}$.

The dual problem may be solved via forward-backward splitting (FBS) [7, 6] iterations which coincide with the projected gradient algorithm (PGA) [10] for this problem.

For a constrained minimization problem involving a smooth cost function, such as,

$$\min_{x \in C} f(x), \quad (33)$$

where C is a convex set, the FBS iterations produce a sequence as

$$x^{k+1} = P_C\big(x^k - \alpha\, \nabla f(x^k)\big), \quad (34)$$

where P_C denotes the projection operator onto C.

For our problem (32), the FBS iterations construct a sequence as,

$$Z^{k+1} = P_B\Big(Z^k - \alpha\, \lambda\, \mathcal{S}\big(\lambda\, \mathcal{S}^T\, Z^k - Y\big)\Big), \quad (35)$$

where P_B is the projection operator onto B. In this algorithm, the convergence of the algorithm is ensured if the 'step-size' α is small enough.

Proposition 2. The algorithm in (35) converges if $\alpha < 2/\big(\lambda^2\,(2K+1)\big)$.

Proof. FBS iterations in (34) are known to converge if $\alpha < 2/L$, when ∇f is L-Lipschitz continuous, i.e., when

$$\|\nabla f(z) - \nabla f(u)\|_2 \leq L\|z-u\|_2, \tag{36}$$

for any z, u [7, 1].

In the current context, $f(Z) = \frac{1}{2}\|Y - \lambda\,\mathcal{S}^T\,Z\|_F^2$. Therefore,

$$\|\nabla f(Z) - \nabla f(U)\|_2 = \|\lambda^2\,\mathcal{S}\,\mathcal{S}^T\,(Z-U)\|_2. \tag{37}$$

Since $\sigma(\mathcal{S}^T\,\mathcal{S}) = 2K+1$ by Corollary 1, the claim follows. □

In order to implement the algorithm in (35), we need to realize P_B, the projection operator onto B. To describe this operator we introduce the 'singular value clipping' operation.

Definition 1. Let $Y \in \mathbb{C}^{n,m}$. Suppose the SVD of Y is $Y = U\Sigma V^*$. Let $\bar{\Sigma}$ to be the diagonal matrix whose k^{th} diagonal entry $\bar{\Sigma}_k$ is obtained by clipping the k^{th} diagonal entry of Σ as,

$$\bar{\Sigma}_k = \min\big(\Sigma_k, 1\big). \tag{38}$$

Now let $Z = U\,\bar{\Sigma}\,V^*$. The mapping that takes Y to Z is called 'singular value clipping' and is denoted as $\mathrm{SVC}(Y)$.

Singular value clipping enjoys an optimality property which will be relevant for projecting onto the set B.

Proposition 3. Let $B_1 \subset \mathbb{C}^{n,m}$denote the set of matrices whose singular values lie in the range $[0,1]$. Then, the projection of $Y \in \mathbb{C}^{n,m}$onto $B_1$is $\mathrm{SVC}(Y)$.

Proof. We first consider a diagonal Σ with non-negative (real) entries and show that the projection of Σ onto B_1 has to be diagonal. For this, assume the contrary and let $Z \in B_1$ be the sought projection with a non-zero off-diagonal entry. Since $Z \in B_1$, its SVD is of the form $Z = U_Z\,\Sigma_Z\,V_Z^*$ where each diagonal entry of $\Sigma_Z$is real and lies in the range $[0,1]$. Since U_Z is orthogonal, it follows that each row of the matrix $U_Z\,\Sigma_Z$ has ℓ_2 norm bounded by unity. Since $V_Z^*$is also orthogonal, it follows by the Cauchy-Schwarz inequality that each entry of Z (regarded as the inner products of the rows of $U_Z\,\Sigma_Z$ and columns of V_Z^*) has absolute value bounded by unity. Thus, if we were to set the off-diagonal entries of Z to zero to obtain a new matrix $\tilde{Z}$, then $\tilde{Z}$ would also be in B. But since Y is diagonal, we would have $\|Y - \tilde{Z}\|_F < \|Y - Z\|_F$, a contradiction. Thus, Z must be diagonal. But among the set of diagonal matrices in B_1, $\bar{\Sigma}$, defined as in (38), minimizes $\|\Sigma - \bar{\Sigma}\|_F$ and therefore it must be the projection of Σonto B_1.

Now let the SVD of Y be $Y = U\Sigma V^*$. Also, let Z be the projection of Y onto B_1. Then,

$$\|Y - Z\|_F = \|\Sigma - U^*\,Z\,V\|_F. \tag{39}$$

But by the previous discussion, $U^*\,Z\,V$ must be equal to $\bar{\Sigma}$. Thus, $Z = U\,\bar{\Sigma}\,V^* = \mathrm{SVC}(Y)$. □

Singular value clipping and singular value thresholding are closely related.

Proposition 4. $\mathrm{SVC}(Y) + \mathrm{SVT}_1(Y) = Y$.

Proof. For a non-negative number t, observe that

$$\min(t,1) + \max(t-1,0) = t. \tag{40}$$

Therefore, if the SVD of Y is $Y = U\,\Sigma\,V$, then

$$\mathrm{SVC}(Y) + \mathrm{SVT}_1(Y) = U\,\big(\bar{\Sigma} + \tilde{\Sigma}\big)\,V, \tag{41}$$

where $\bar{\Sigma}$ and $\tilde{\Sigma}$ are diagonal matrices whose k^{th} diagonals are given as, $\min(\Sigma_k, 1)$ and $\max(\Sigma_k - 1, 0)$ respectively. Using (40), we find $\bar{\Sigma} + \tilde{\Sigma} = \Sigma$, from which the claim follows. □

As a corollary of Prop. 3, we obtain an expression for the projector onto B.

Proposition 5. Let a partition $\mathcal{J}$ be given. Then $Z = P_B(Y)$ if and only if

$$Z_J = \mathrm{SVC}(Y_J) \text{ for all } J \in \mathcal{J}. \tag{42}$$

Following this discussion, pseudocode for an algorithm minimizing (25) is given in Algorithm 1.

Algorithm 1 An Algorithm for Solving (25)

Initialize $\alpha \leq 2/\big(\lambda^2(2K+1)\big)$
repeat
  $Z \leftarrow Z - \alpha\,\lambda\,\mathcal{S}\big(\lambda\,\mathcal{S}^T Z - Y\big)$
  $Z_J \leftarrow \mathrm{SVC}(Z_J)$ for all $J \in \mathcal{J}$
until some convergence criterion is met
$\hat{X} \leftarrow Y - \lambda\,\mathcal{S}^T\,Z$

B. Selection of the Regularization Weight

According to Prop. 1, the denoised estimate satisfies $\hat{X} = Y - \lambda\mathcal{S}^T\,\hat{Z}$ for some $\hat{Z} \in B$. If we assume that $\hat{X}$is indeed a good approximation of the underlying clean matrix, then $\lambda\mathcal{S}^T\,\hat{Z} = Y - \hat{X}$ should be a good estimate of the noise in the observation matrix Y. This observation allows to derive a sensible value of λ.

First, let us study the spectral norm of $\mathcal{S}^T Z$. $Z \in B$ means that, for a given partition $\mathcal{J} = \{J_1, J_2, \ldots, J_m\}$,

$$Z = \begin{bmatrix} Z_{J_1} \\ \vdots \\ Z_{J_m} \end{bmatrix}, \tag{43}$$

where $\sigma(Z_{J_k}) \leq 1$ for each k. It follows that,

$$\sigma(Z^T\,Z) = \sigma\left(\sum_{k=1}^{m} Z_{J_k}^T\,Z_{J_k}\right) \leq \sum_{k=1}^{m} \sigma\left(Z_{J_k}^T\,Z_{J_k}\right) \leq m. \tag{44}$$

Therefore, $\sigma(Z) \leq \sqrt{m}$. Now, if $|J_1| = \cdots = |J_m|$, i.e., if each Z_{J_k} has the same number of rows, then, $m = N/G_s$, where $G_s$denotes a 'group-size' defined to be $G_s = |J_k|$ for any k. Finally, in view of Corollary 1, we conclude that

$$\sigma(\mathcal{S}^T\,Z) \leq \sqrt{\frac{N}{G_s}(2K+1)}. \tag{45}$$

Consider now the noise matrix $U \in C^{N,M}$. For the chosen frequency band, the k^{th} column of this matrix holds the STFT coefficients for the k^{th} observation. Suppose now that in the time-domain, the noise is white, with variance σ^2. Also, let g be the window used for computing the STFT. Then, the expected energy of each column of U is equal to $\sigma^2 \|g\|_2^2 N$. If we further assume that the columns of U are statistically independent, then $\|g\|_2 \sigma \sqrt{N}$can be taken as an estimate of $\sigma(U)$.

Setting equal $\lambda \sigma(\mathcal{S}^T Z)$ and $\sigma(U)$, we find,

$$\lambda = \sigma \|g\|_2 \sqrt{\frac{G_s}{2K+1}}. \tag{46}$$

Although this λ value may not be optimal, it provides a quick estimate, that can be fine-tuned using a multiplicative factor around unity.

C. Numerical Experiments

In order to evaluate the proposed denoising scheme, we conducted a batch of experiments. Matlab code for the single-band denoising experiment below is available at "`http://web.itu.edu.tr/ibayram/RegReverb`".

Eight recordings obtained in a reverberant room are available for the experiments. However, to understand the behavior of the proposed prior, we considered two scenarios involving four and eight observations. The microphones are arranged in a linear fashion, with a 2.5 cm spacing in between. The dimensions of the room are approximately $10 \times 10 \times 3$ m and the room is highly reverberant. The RIRs are unknown and we do not try to estimate them. The recordings are sampled at 44.1 KHz. The STFT window size is 40 msec (1764 samples) with a hop size of 20 msec (882 samples). The spectrogram of the clean signal from the first microphone is shown in Fig. 8a.

Single Band Denoising: We first experiment on a single STFT band with center frequency $\omega_c = 650$ Hz. The real part of the 1-D observation from the first microphone is shown in Fig. 7a.

We obtain the noisy observations by adding white Gaussian noise to the time-domain signal and computing the noisy STFT coefficients. All of the SNRs reported below belong to the selected band. The noise energy is evenly distributed among the different observations. In order to define the regularization function, the set $[1, N]$ is partitioned into intervals of duration 2 seconds. This results in a 'group-size' of 100. For varying K values, we ran Algorithm 1 with 100 iterations. We selected the weight λ manually (with a sweep search) so as to achieve the highest SNR in each case. We also normalized each reconstruction with a scalar that maximizes the SNR. Although such normalization can only be realized by an oracle, it preserves the 'shape' of the signal and leads to a better assessment of performance in our opinion. For different K values, the SNR gains of the denoised signals are tabulated in Table- II for when four or eight observations are used.

In order to compare the performance of the proposed regularization function, we also considered a simple sparsity inducing regularization function based on mixed norms [20]. In the current setting, recall that the columns of X contain observations from different microphones and the rows are associated with time. We partition $X \in \mathbb{C}^{N,M}$ along the rows to obtain submatrices of size $L \times M$. Then we add the ℓ_2 norms of the submatrices. This defines a norm on X, dependent on L, which we denote as $\|X\|_L$. By replacing $q(X)$ with $\|X\|_L$ in (25), we obtain alternative reconstuctions. These reconstructions are also normalized with a scalar chosen by an oracle so as to maximize the SNR. The output SNRs thus obtained for this alternative formulation are tabulated in Table III.

TABLE II
SNR GAINS FOR DENOISING A SINGLE BAND (PROPOSED REGULARIZER)

		K					
Mic.	SNR_{in}	0	1	2	4	6	8
4	0	6.71	**7.3**	7.15	6.98	6.86	6.64
4	5	5.6	**5.83**	5.8	5.64	5.46	5.32
4	10	4.61	**4.73**	4.71	4.66	4.57	4.4
4	15	3.64	**3.88**	3.87	3.85	3.78	3.7
4	20	2.95	3.12	3.14	**3.14**	3.12	3.04
8	0	7.95	**8.95**	8.82	8.35	8.07	7.74
8	5	7.17	**7.3**	7.27	7.02	6.69	6.37
8	10	5.77	**6.02**	5.96	5.85	5.64	5.4
8	15	4.75	4.94	**4.96**	4.8	4.63	4.43
8	20	3.9	**4.14**	4.07	3.95	3.79	3.6

TABLE III
SNR GAINS FOR DENOISING A SINGLE BAND (MIXED NORM)

		L					
Mic.	SNR_{in}	1	3	5	7	9	11
4	0	5.81	5.91	5.98	**6.01**	6.01	5.97
4	5	3.84	**3.86**	3.85	3.72	3.69	3.67
4	10	**2.37**	2.32	2.28	2.13	2.09	2.04
4	15	**1.36**	1.29	1.25	1.14	1.09	1.03
4	20	**0.732**	0.673	0.648	0.563	0.521	0.484
8	0	**6.38**	6.37	6.37	6.3	6.29	6.14
8	5	**4.19**	4.15	4.04	3.92	3.87	3.82
8	10	**2.53**	2.44	2.32	2.26	2.17	2.12
8	15	**1.42**	1.32	1.24	1.19	1.12	1.07
8	20	**0.728**	0.657	0.62	0.584	0.526	0.505

Inspecting the tables, we see that the proposed regularizer shows a clear improvement over mixed norm regularization in this experiment. Even though mixed norms can be powerful for regularizing audio signals [29, 2], they are not designed for reverberant signals. Therefore, we expect the proposed reconstruction to be closer to the true signal in the regions with speech activity. To support this claim, we show in Fig. 7 some signals from this experiment, corresponding to the eight microphone, 5 dB SNR case. In Fig. 7d, we show the absolute error of the reconstruction obtained with the proposed regularizer and the mixed norm regularizer respectively. In general, we observed that the proposed regularizer is more effective in regions with audio activity, whereas the mixed norm is more effective in supressing noise in regions of silence.

Denoising the Whole Spectrogram: We also experimented with the method applied to the whole spectrogram. Recall that

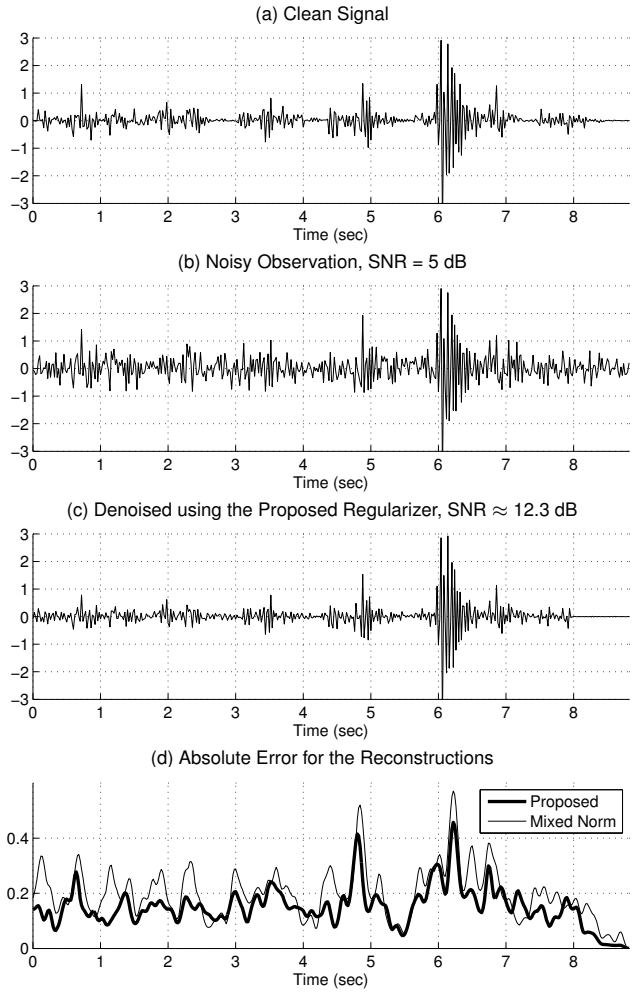

Fig. 7. Signals for the single band denoising experiment. (a) The real part of an STFT band of the clean reverberant signal, (b) the noisy observation, (c) denoised signal using the proposed iterative SVC algorithm, (d) absolute error for the proposed reconstruction and mixed norm regularized reconstruction with the optimal parameters. The absolute errors in (d) are smoothed for a better view.

for such a task, each STFT band is procesed independently. We took the number of observations as eight. The noisy observation is obtained by adding real-valued white Gaussian noise to the time-domain signal. The input SNR is 5 dB. The spectrograms of the clean signal and the noisy signal for the first microphone are shown in Fig. 8a,b respectively. For denoising with the proposed prior, we chose $K = 9$ and set λ to be one thirds of the value proposed in (46). This choice approximately outputs the highest SNR. We use the same partition $\mathcal{J}$ as in the single band case. In this setting, we obtain eight reconstructions. The overall SNR (considering all eight reconstructions) is 12.24 dB. The spectrogram of the reconstruction is shown in Fig. 8c.

For comparison, we also tried denoising with mixed norms, as in the single band case. For the mixed norm, we took $L = 5$ and set the regularization weight λ manually so as to maximize SNR. The best reconstruction in this case yielded an SNR of 10.48 dB. The spectrogram of this signal is visually similar to the one obtained with the proposed regularizer and is not shown. Instead, in order to show the difference in reconstructions, we show in Fig. 8d the energy of the

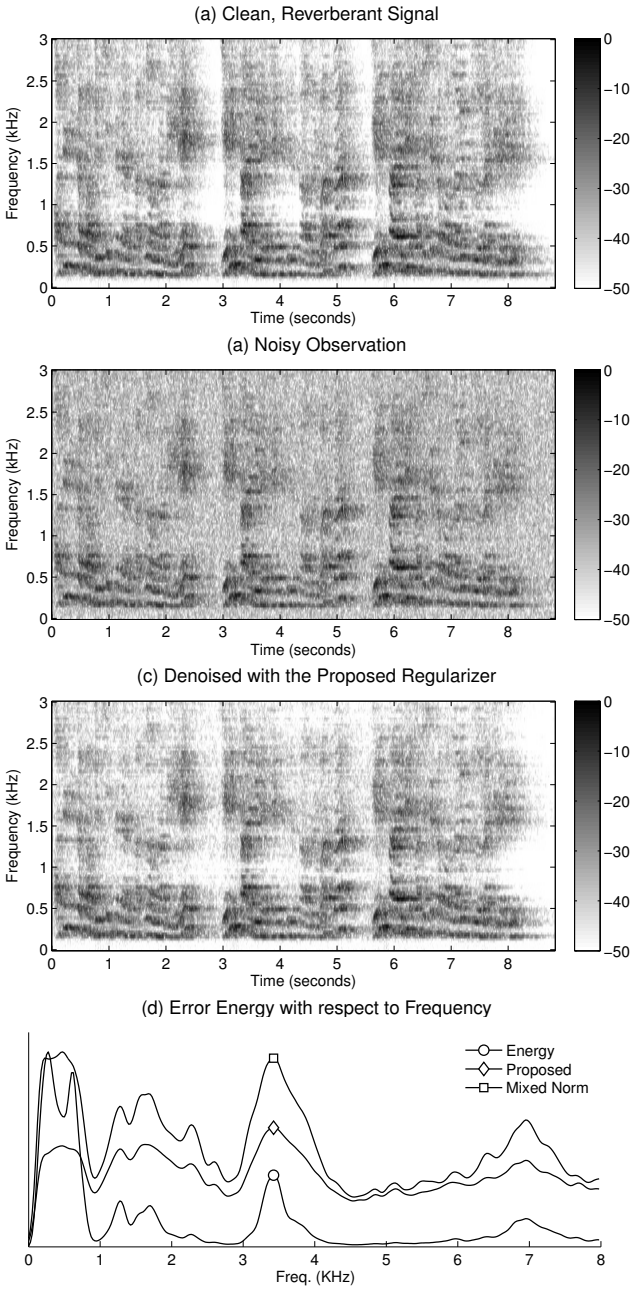

Fig. 8. Reverberant signal denoising using eight input signals with unknown room impulse response. (a) The original clean, reverberant signal for the first microphone. (a) The noisy observation for the first microphone, SNR $\approx$ 5 dB, (c) the reconstruction obtained with the proposed denoising algorithm, SNR $\approx$ 12.24 dB, (d) energy of the error with respect to frequencies for the reconstructions with the proposed regularizer and mixed norms.

reconstruction error in each frequency band, for both methods. We also show the energy with respect to frequency bands in this figure. The figure suggests that, in bands with significant activity, the proposed regularizer performs significantly better than mixed norms.

V. Conclusion

We proposed a regularization function for multichannel reverberant audio signals. We discussed the application of the proposed regularization function in a denoising scenario in this paper. However, the proposed regularizer can be used in

other applications where the target is multichannel reverberant audio. In the near future, we plan to apply the regularizer in different problems, such as room impulse response shortening [23, 22, 30], source separation or source suppression. Typically, in these applications, when the room impulse responses are not precisely known, blind methods are called for. The proposed regularizer would allow to avoid estimating the unknown room impulse responses.

The proposed regularizer could also be of interest in more general inverse problems, where exact information about the imaging system is not practically available. In that case, in an imaging scheme, if different observations $x_1, \ldots, x_n$ of the source z can be related to each other via simple linear operators, it would be possible to define an operator like $\mathcal{S}$ as in this paper so that $\mathcal{S}\begin{bmatrix} x_1 & \ldots x_n \end{bmatrix}$ is low rank. This in turn brings the nuclear norm into the picture and allows to employ $\left\|\mathcal{S}\begin{bmatrix} x_1 & \ldots x_n \end{bmatrix}\right\|_*$ as a regularizer in the imaging problem.

Another extension might be to employ approximations of rank that are more effective than the nuclear norm, such as the weighted nuclear norm [11, 17] or the weakly-convex function proposed in [26]. These extensions are shown to reduce the bias when utilized for low-rank matrix approximation and they might further improve the reconstructions. We plan to investigate such modifications in future work.

ACKNOWLEDGEMENTS

İ.B. thanks Ivan W. Selesnick, New York University, USA, Pavel Rajmic, Brno University of Technology, Czech Republic, and Savaşkan Bulek, Qualcomm Atheros, Inc., MI, USA, for discussions and comments.

REFERENCES

[1] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.

[2] İ. Bayram. Mixed-norms with overlapping groups as signal priors. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2011.

[3] J.-F. Cai, E. J. Candès, and Z. Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[4] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772, December 2009.

[5] Z. Chen and D. Ellis. Speech enhancement by sparse, low-rank, and dictionary spectrogram decomposition. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2013.

[6] P. L. Combettes and J.-C. Pesquet. Proximal splitting methods in signal processing. In H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, New York, 2011.

[7] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM J. Multiscale Model. Simul.*, 4(4):1168–1200, November 2005.

[8] L. Daudet. Sparse and structured decompositions of signals with the molecular matching pursuit. *IEEE Trans. Audio, Speech and Language Proc.*, 14(5):1808–1816, September 2006.

[9] C. Févotte, B. Torrésani, L. Daudet, and S. J. Godsill. Sparse linear regression with structured priors and application to denoising of musical audio. *IEEE Trans. Audio, Speech and Language Proc.*, 16(1):174–185, January 2008.

[10] A. A. Goldstein. Convex programming in Hilbert space. *Bull. Amer. Math. Soc.*, 70(5):709–710, 1964.

[11] S. Gu, L. Zhang, W. Zuo, and X. Feng. Weighted nuclear norm minimization with application to image denoising. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[12] Y. Haneda, Y. Kaneda, and N. Kitawaki. Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function. *IEEE Trans. Speech and Audio Proc.*, 7(6):709 – 717, November 1999.

[13] Y. Haneda, S. Makino, and Y. Kaneda. Common acoustical pole and zero modeling of room transfer functions. *IEEE Trans. Speech and Audio Proc.*, 2(2):320 – 328, April 1994.

[14] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, 1990.

[15] P.-S. Huang, S. D. Chen, P. Smaragdis, and M. Hasegawa-Johnson. Singing-voice separation from monaural recordings using robust principal component analysis. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, 2012.

[16] H. Ji, S. Huang, Z. Shen, and Y. Xu. Robust video restoration by joint sparse and low rank matrix approximation. *SIAM Journal on Imaging Sciences*, 4(4):1122–1142, 2010.

[17] M. Fazel K. Mohan. Iterative reweighted algorithms for matrix rank minimization. *Journal of Machine Learning Research*, 13:34413473, November 2012.

[18] Z. Koldovský, J. Málek, and S. Gannot. Spatial source subtraction based on incomplete measurements of relative transfer function. *arXiv*, abs/1411.2744v4, 2015.

[19] Z. Koldovský, J. Málek, P. Tichavský, and F. Nesta. Improving relative transfer function estimates using second-order cone programming. *Lecture notes in computer science*, 9237:227–234, 2015.

[20] M. Kowalski and B. Torrésani. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients. *Signal, Image and Video Processing*, 3(3):251–264, 2009.

[21] M. Kowalski, E. Vincent, and R. Gribonval. Beyond the narrowband approximation: Wideband convex methods for under-determined reverberant audio source separation. *IEEE Trans. Audio, Speech and Language Proc.*, 18(7):1818–1829, September 2010.

[22] P. D. Teal L. Krishnan and T. Betlehem. A sparsity based approach for acoustic room impulse response shortening. In *Proc. IEEE Workshop on Statistical Signal Processing (SSP)*, 2014.

[23] A. Mertins, M. Tiemin, and M. Kallinger. Room impulse response shortening/reshaping with infinity- and p-norm optimization. *IEEE Trans. Audio, Speech and Language Proc.*, 18(2):249–259, February 2010.

[24] R. Mignot, G. Chardon, and L. Daudet. Low frequency interpolation of room impulse responses using compressed sensing. *IEEE Trans. Audio, Speech and Language Proc.*, 22(1):205 – 216, January 2014.

[25] J. N. Mourjopoulos. Digital equalization of room acoustics. *Journal of the Audio Engineering Society*, 42(1):884–900, 1994.

[26] A. Parekh and I. W. Selesnick. Enhanced low-rank matrix approximation. *arXiv*, abs/1511.01966, 2015.

[27] J. P. Reilly, M. Wilbur, M. Seibert, and N. Ahmadvand. The complex subband decomposition and its application to the decimation of large adaptive filtering problems. *IEEE Trans. Signal Processing*, 50(11):2730–2743, Nov 2002.

[28] R.Talmon, I. Cohen, and S. Gannot. Relative transfer function identification using convolutive transfer function approximation. *IEEE Trans. Audio, Speech and Language Proc.*, 17(4):546–555, May 2009.

[29] K. Siedenburg and M. Dörfler. Structured sparsity for audio signals. In *Proc. Int. Conf. on Digital Audio Effects (DAFx)*, 2011.

[30] M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor. Application of channel shortening to acoustic channel equalization in the presence of noise and estimation error. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2011.