COMPLEXITY REDUCTION OF EIGENVALUE DECOMPOSITION-BASED DIFFUSE POWER SPECTRAL DENSITY ESTIMATORS USING THE POWER METHOD

Marvin Tammen[⋆] *Ina Kodrasi*[⋆†] *Simon Doclo*[⋆]

[⋆] University of Oldenburg, Department of Medical Physics and Acoustics
and Cluster of Excellence Hearing4All, Oldenburg, Germany
[†] Idiap Research Institute, Speech and Audio Processing Group, Martigny, Switzerland
{marvin.tammen,ina.kodrasi,simon.doclo}@uni-oldenburg.de

ABSTRACT

In noisy and reverberant environments speech enhancement techniques such as the multi-channel Wiener filter (MWF) can be used to improve speech quality and intelligibility. Assuming that reverberation and ambient noise can be modeled as diffuse sound fields, such techniques require an estimate of the diffuse power spectral density (PSD). Recently a multi-channel diffuse PSD estimator based on the eigenvalue decomposition (EVD) of the prewhitened signal PSD matrix was proposed. The EVD-based PSD estimator is advantageous in comparison to other state-of-the-art PSD estimators, since it does not require knowledge of the relative early transfer functions of the target signal. However, computing the EVD can be computationally expensive, particularly when the number of microphones is large. In this paper we propose to reduce the complexity of the EVD-based PSD estimator by using the iterative power method to compute the eigenvalues. Since the EVD-based PSD estimator only requires the largest eigenvalues, the full EVD is not required and the power method is a well suited computationally efficient technique to estimate these eigenvalues. Experimental results show that using the PSD estimated via the power method in an MWF yields a very similar performance as using the PSD estimated via the full EVD.

Index Terms— dereverberation, PSD estimation, EVD, power method, complexity reduction

1. INTRODUCTION

The microphone signals recorded in many hands-free speech communication applications such as teleconferencing, voice-controlled systems or hearing aids are often corrupted by reverberation and ambient noise. Reverberation and noise cause the recorded signals to sound distant and spectrally distorted and typically result in a degradation of speech quality and intelligibility [1, 2] as well as a performance deterioration of automatic speech recognition systems [3, 4]. In order to mitigate these detrimental effects, dereverberation and noise reduction techniques are required. Both single-microphone as well as multi-microphone techniques exist, where multi-microphone techniques are generally preferred, since they are also able to take the spatial characteristics of the microphone signals into account [5]. A commonly used dereverberation and noise reduction technique is the multi-channel Wiener Filter (MWF), which minimizes the mean-square error between the output signal and a target signal [6–9]. The MWF can be implemented as a minimum variance distortionless response (MVDR) beamformer, which takes spatial information into account, followed by a single-channel spectral postfilter, which requires an estimate of the late reverberation and noise power spectral densities (PSDs) [9–14]. Since late reverberation is commonly modeled as a diffuse sound field [10–15] and since diffuse background noise is commonly encountered in many speech communication applications, implementing the MWF requires an estimate of the diffuse PSD. Several multi-channel diffuse PSD estimators have been proposed [10–15], which typically require an estimate of the relative early transfer functions (RETFs) of the target signal between the reference microphone and all microphones. The RETFs may be difficult to estimate accurately, particularly in highly reverberant and noisy scenarios. Recently we proposed a multi-channel diffuse PSD estimator based on the eigenvalue decomposition (EVD) of the prewhitened signal PSD matrix, which does not require knowledge of the RETFs [16, 17]. Experimental results in [17] show the advantages of using the EVD-based PSD estimator in an MWF, both when the RETFs are perfectly estimated as well as in the presence of RETF estimation errors. However, since the late reverberation and noise are nonstationary and their PSD needs to be estimated in each time-frequency bin, i.e., the EVD needs to be computed for each time-frequency bin, the EVD-based PSD estimator may be computationally unsuitable for real-time applications, particularly when the number of microphones is large.

In this paper we propose to mitigate this problem by relying on the iterative power method [18] to compute the eigenvalues. Considering that the EVD-based PSD estimator requires only the first or the second largest eigenvalue, a full EVD computation is not necessary and the iterative power method can be used as a computationally efficient procedure to compute the largest eigenvalues, which is similar to the approach proposed in [19] to compute the RETF. Experimental results for several realistic acoustic scenarios show that the iterative power method converges quickly, yielding a similar performance as the full EVD while reducing the computational complexity.

2. SIGNAL MODEL AND NOTATION

We consider a reverberant and noisy acoustic scenario with one speech source and $M \geq 2$ microphones. The signals are considered in their time-frequency representation obtained via the short-time Fourier transform (STFT) with frequency index k and frame index l. In vector notation, the M-dimensional stacked vector of the microphone signals $\mathbf{y}(k,l) = [Y_1(k,l), Y_2(k,l), ..., Y_M(k,l)]^T$ is given by

$$\mathbf{y}(k,l) = \mathbf{x}(k,l) + \mathbf{d}(k,l) + \mathbf{v}(k,l), \tag{1}$$

This work was supported by the Cluster of Excellence Hearing4All, funded by the German Research Foundation (DFG), and the joint Lower Saxony-Israeli Project ATHENA, funded by the State of Lower Saxony.

with $\mathbf{x}(k,l)$ the direct and early reverberation component, $\mathbf{d}(k,l)$ the diffuse sound component, and $\mathbf{v}(k,l)$ the noise component. The vectors $\mathbf{x}(k,l)$, $\mathbf{d}(k,l)$, and $\mathbf{v}(k,l)$ are defined similarly as $\mathbf{y}(k,l)$. The diffuse sound component $\mathbf{d}(k,l)$ models the late reverberation as well as any noise which can be well approximated by a diffuse sound field, such as background noise in large crowded rooms. The noise component $\mathbf{v}(k,l)$ accounts for any noise which cannot be modeled by a diffuse sound field, such as uncorrelated sensor noise. For simplicity, in the remainder of this paper we assume that the non-diffuse noise component $\mathbf{v}(k,l)$ is equal to zero, i.e.,

$$\mathbf{y}(k,l) = \mathbf{x}(k,l) + \mathbf{d}(k,l). \tag{2}$$

However, the diffuse PSD estimators considered in this paper can also be used in acoustic scenarios where non-diffuse noise is present, as long as an estimate of the non-diffuse noise PSD matrix is available [16]. Since all processing is performed separately for each frequency bin, the index k is omitted in the remainder of this paper.

Following the widely-used assumption that the components in (2) are mutually uncorrelated, the $M \times M$-dimensional PSD matrix of the microphone signals can be written as

$$\mathbf{\Phi}_{\mathbf{y}}(l) = \mathcal{E}\{\mathbf{y}(l)\mathbf{y}^H(l)\} = \mathbf{\Phi}_{\mathbf{x}}(l) + \mathbf{\Phi}_{\mathbf{d}}(l) \tag{3}$$

where $\mathcal{E}$ denotes the expected value operator and $\mathbf{\Phi}_{\mathbf{x}}(l)$ and $\mathbf{\Phi}_{\mathbf{d}}(l)$ are the PSD matrices of $\mathbf{x}(l)$ and $\mathbf{d}(l)$, respectively. The direct and early reverberation component is given by

$$\mathbf{x}(l) = S(l)\mathbf{a}(l), \tag{4}$$

where $S(l)$ denotes the target signal as received by a reference microphone and $\mathbf{a}(l)$ is a vector of (possibly) time-varying RETFs of the target signal between the reference microphone and all microphones. Based on (4) and since $\mathbf{\Phi}_{\mathbf{d}}(l)$ is the PSD matrix of the diffuse sound component, (3) can be written as

$$\mathbf{\Phi}_{\mathbf{y}}(l) = \underbrace{\Phi_s(l)\mathbf{a}(l)\mathbf{a}^H(l)}_{\mathbf{\Phi}_{\mathbf{x}}(l)} + \underbrace{\Phi_d(l)\mathbf{\Gamma}}_{\mathbf{\Phi}_{\mathbf{d}}(l)}, \tag{5}$$

with $\Phi_s(l)$ the time-varying target signal PSD, i.e, $\Phi_s(l) = \mathcal{E}\{|S(l)|^2\}$, $\Phi_d(l)$ the time-varying diffuse PSD, and $\mathbf{\Gamma}$ the time-invariant spatial coherence matrix of a diffuse sound field, which can be analytically computed based on the microphone array geometry [20].

In this paper, speech enhancement is achieved using the MWF, which is implemented as an MVDR beamformer followed by a single-channel spectral postfilter. Applying the MVDR beamformer with filter coefficients $\mathbf{w}(l) = [W_1(l), ..., W_M(l)]^T$, the MVDR output signal is given by

$$\hat{X}(l) = \mathbf{w}^H(l)\mathbf{y}(l) = \left(\frac{\mathbf{\Gamma}^{-1}\mathbf{a}(l)}{\mathbf{a}^H(l)\mathbf{\Gamma}^{-1}\mathbf{a}(l)}\right)^H \mathbf{y}(l). \tag{6}$$

To obtain the final target signal estimate, a spectral postfilter $G(l)$ is applied to the MVDR output signal as

$$\hat{S}(l) = G(l)\hat{X}(l) = \frac{\hat{\Phi}_s(l)}{\hat{\Phi}_s(l) + \hat{\Phi}_d(l)/(\mathbf{a}^H(l)\mathbf{\Gamma}^{-1}\mathbf{a}(l))}\hat{X}(l), \tag{7}$$

with $\hat{\Phi}_s(l)$ and $\hat{\Phi}_d(l)$ estimates of the target signal and diffuse PSDs. As can be seen from (7), an estimate of the diffuse PSD $\hat{\Phi}_d(l)$ is required to obtain the final target signal estimate $\hat{S}(l)$.

3. EVD-BASED DIFFUSE PSD ESTIMATOR

Using the structure of the PSD matrix $\mathbf{\Phi}_{\mathbf{y}}(l)$ in (5), an EVD-based estimator for the diffuse sound PSD $\hat{\Phi}_d(l)$ has been proposed in [17]. First, the Cholesky decomposition of the spatial coherence matrix $\mathbf{\Gamma}$ is computed, i.e.,

$$\mathbf{\Gamma} = \mathbf{L}\mathbf{L}^H, \tag{8}$$

with $\mathbf{L}$ an $M \times M$-dimensional lower triangular matrix. Using (8), the signal PSD matrix $\mathbf{\Phi}_{\mathbf{y}}(l)$ is prewhitened as

$$\begin{aligned}
\mathbf{\Phi}^w_{\mathbf{y}}(l) =& \mathbf{L}^{-1}\mathbf{\Phi}_{\mathbf{y}}(l)\mathbf{L}^{-H} && (9)\\
=& \Phi_s(l)\underbrace{\mathbf{L}^{-1}\mathbf{a}(l)}_{\mathbf{b}(l)}\underbrace{\mathbf{a}^H(l)\mathbf{L}^{-H}}_{\mathbf{b}^H(l)} + \Phi_d(l)\mathbf{L}^{-1}\mathbf{\Gamma}\mathbf{L}^{-H} && (10)\\
=& \Phi_s(l)\mathbf{b}(l)\mathbf{b}^H(l) + \Phi_d(l)\mathbf{I}_M, && (11)
\end{aligned}$$

where $\mathbf{I}_M$ is the $M \times M$-dimensional identity matrix. Due to the structure in (11), the eigenvalues of $\mathbf{\Phi}^w_{\mathbf{y}}(l)$ are given by

$$\begin{aligned}
&\lambda_1\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\} = \sigma(l) + \Phi_d(l)\\
&\lambda_i\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\} = \Phi_d(l) \quad \forall i \in \{2, ..., M\},
\end{aligned} \tag{12}$$

where $\lambda_i\{\cdot\}$ denotes the i-th eigenvalue (arranged in descending order) and $\sigma(l)$ denotes the only non-zero eigenvalue of the rank-1 term $\Phi_s(l)\mathbf{b}(l)\mathbf{b}^H(l)$. Based on (12), in [17] we proposed to estimate the diffuse PSD by computing the EVD of the prewhitened PSD matrix $\mathbf{\Phi}^w_{\mathbf{y}}(l)$ and using either the second eigenvalue, i.e.,

$$\hat{\Phi}_{d,\mathrm{EIG2}}(l) = \lambda_2\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\}, \tag{13}$$

or the mean of the last $M-1$ eigenvalues, i.e.,

$$\hat{\Phi}_{d,\mathrm{EIG1}}(l) = \frac{1}{M-1}\left(\mathrm{tr}\left\{\mathbf{\Phi}^{\mathrm{w}}_{\mathbf{y}}(l)\right\} - \lambda_1\left\{\mathbf{\Phi}^{\mathrm{w}}_{\mathbf{y}}(l)\right\}\right), \tag{14}$$

with $\mathrm{tr}\{\cdot\}$ denoting the trace and (14) derived using the fact that the trace of a matrix is equal to the sum of its eigenvalues. Obviously, when the true spatial coherence matrix $\mathbf{\Gamma}$ and the true PSD matrix $\mathbf{\Phi}_{\mathbf{y}}(l)$ are known, the EVD-based PSD estimates in (13) and (14) are equal. However, since in practice the model in (5) does not perfectly hold, the EVD-based PSD estimates in (13) and (14) are different. Note that in contrast to other state-of-the-art diffuse PSD estimators [10–15], the EVD-based estimator does not suffer from performance degradation caused by RETF estimation errors. However, since the EVD needs to be performed for every time-frequency bin, the computational cost may become unsustainable for real-time applications, particularly when the number of microphones is large. To mitigate this problem, in the following section we propose to estimate the two largest eigenvalues $\lambda_1\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\}$ and $\lambda_2\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\}$ using the computationally efficient iterative power method.

4. ITERATIVE POWER METHOD

The power method is a well-known iterative procedure for numerically solving eigenproblems [18]. It is applicable to estimating the largest eigenvalue of an $M \times M$-dimensional matrix $\mathbf{A}$ under the condition that

$$|\lambda_1\{\mathbf{A}\}| > |\lambda_2\{\mathbf{A}\}| \geq ... \geq |\lambda_M\{\mathbf{A}\}|. \tag{15}$$

As already mentioned in Section 3, since the model in (5) does not perfectly hold, the eigenvalues of $\mathbf{\Phi}^w_{\mathbf{y}}(l)$ are typically different, i.e.,

$$|\lambda_1\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\}| > |\lambda_2\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\}| > ... > |\lambda_M\{\mathbf{\Phi}^w_{\mathbf{y}}(l)\}|. \tag{16}$$

Hence, since the eigenvalues of the prewhitened PSD matrix $\mathbf{\Phi}_{\mathbf{y}}^{w}(l)$ satisfy (15), the power method can be applied to obtain an estimate of $\lambda_1\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}$. In addition, since $\lambda_2\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}$ also typically differs from the remaining eigenvalues, the power method can be slightly modified to obtain an estimate of $\lambda_2\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}$ as well (cf. Section 4.1). In the remainder of this section, the power method is presented and some insights on the computational complexity reduction compared to the full EVD are provided.

4.1. Algorithm

The power method for computing the two largest eigenvalues of the matrix $\mathbf{\Phi}_{\mathbf{y}}^{w}(l)$ is shown in Algorithm 1. It should be noted that in principle the power method can be used to compute all M eigenvalues of the matrix $\mathbf{\Phi}_{\mathbf{y}}^{w}(l)$, however, in this paper only the largest two eigenvalues are of interest. The justification of this iterative procedure, including a proof of convergence, can be found in [18]. The convergence speed depends on the eigenvalue ratios $\left(\frac{\lambda_m\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}}{\lambda_1\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}}\right)^n$, $2 \leq m \leq M$, where n is the iteration index. Since $|\lambda_1\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}| > |\lambda_m\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}| \quad \forall m \in \{2, ..., M\}$, the eigenvalue ratios approach zero as the number of iterations n increases.

The modification that allows to compute not only the largest but also the second largest eigenvalue is based on matrix rank reduction. Once the largest eigenvalue has been computed, the column space of the input matrix $\mathbf{\Phi}_{\mathbf{y}}^{w}(l)$ is reduced using the corresponding estimated eigenvector and the same iterations as for computing the largest eigenvalue estimate are repeated. Since the power method is an iterative procedure, a termination criterion needs to be imposed, which is chosen to be a fixed number of iterations.

In: $\mathbf{\Phi}_{\mathbf{y}}^{w}(l) \in \mathbb{C}^{M \times M}$, number of iterations N
Out: 2 eigenvalue estimates $\hat{\lambda}_1\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}$ and $\hat{\lambda}_2\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}$
for $m = 1$ **to** 2 **do**
  initialize $\mathbf{u}_m^{(0)} \in \mathbb{C}^M$;
  for $n = 1$ **to** N **do**
    $\mathbf{t} = \mathbf{\Phi}_{\mathbf{y}}^{w}(l)\mathbf{u}_m^{(n-1)}$;
    $\mathbf{u}_m^{(n)} = \mathbf{t}/||\mathbf{t}||_2$;
    `/* Rayleigh quotient */`
    $\lambda_m^{(n)} = \mathbf{u}_m^{(n),H}\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\mathbf{u}_m^{(n)}$;
  end
  $\hat{\lambda}_m\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\} = \lambda_m^{(N)}$;
  `/* matrix rank reduction */`
  $\mathbf{\Phi}_{\mathbf{y}}^{w}(l) = \mathbf{\Phi}_{\mathbf{y}}^{w}(l) - \hat{\lambda}_m\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}\mathbf{u}_m^{(N)}\mathbf{u}_m^{(N),H}$;
end

Algorithm 1: Power method for computing the first two largest eigenvalues

4.2. Computational Complexity

In this section we provide some insights on the complexity reduction that is achieved when using the power method to compute the required eigenvalues instead of the full EVD. The computational complexity is given in terms of the number of real floating-point operations (flops), with each basic arithmetic operation counted as 1 flop.

For the power method, one iteration of the inner for-loop requires $8M^2 - 2M - 3$ additions, $8M^2 + 2M$ multiplications, $2M$ divisions, and 1 square-root operation, which results in total in $16M^2 + 2M - 2$ flops. Hence, if only the largest eigenvalue is computed using N iterations, $N(16M^2 + 2M - 2)$ flops are required. If the second largest eigenvalue is computed, additional operations are required for the matrix rank reduction and N iterations of the inner for-loop should be repeated. Reducing the rank of $\mathbf{\Phi}_{\mathbf{y}}^{w}(l)$ requires $2M^2$ additions and $3M^2$ multiplications, yielding in total $5M^2$ flops. Hence, using the power method to estimate $\lambda_2\{\mathbf{\Phi}_{\mathbf{y}}^{w}(l)\}$ requires $N(16M^2+2M-2)+5M^2$ flops. Overall, the complexity of the power method is $O(M^2)$.

Although many algorithms exist for computing the full EVD, we consider the QR decomposition-based algorithm [18], which is one of the most widely used algorithms to compute eigenvalues. The complexity of the QR decomposition-based algorithm for Hermitian matrices is $O(M^3)$ flops [21], also when the matrix is first transformed into real tridiagonal form using Householder reflections [18].

Hence, using the power method to compute the eigenvalues of interest instead of the full EVD reduces the complexity from $O(M^3)$ to $O(M^2)$, which can be advantageous for a real-time implementation of the diffuse PSD estimator, particularly when the number of microphones M is large.

5. EXPERIMENTAL VALIDATION

In this section, the diffuse PSD estimation accuracy using the power method is evaluated for different numbers of iterations. Furthermore, the performance of an MWF using the diffuse PSD estimate based on either the full EVD or the power method is compared.

5.1. Evaluation Setup and Algorithmic Settings

Three different acoustic scenarios are investigated [22–24]. Each scenario consists of a single speech source and a microphone array with $M = \{4, 6\}$ microphones. The details for each scenario are summarized in Table 1. The reverberant microphone signals are obtained by convolving a 38 s long clean speech signal with the measured room impulse responses (RIRs) at a sampling frequency of 16 kHz. Diffuse babble noise generated as in [25] is added at different input signal-to-noise ratios (SNRs) $\in \{10, 20, 30, 40\}$ dB.

The time-domain signals are transformed into the time-frequency domain using an STFT with 64 ms frame size and 75 % overlap. The MVDR is calculated as in (6), where the RETF vector $\mathbf{a}(l)$ is computed from the first 8 ms of the RIRs using the first microphone as the reference microphone and the diffuse coherence matrix $\mathbf{\Gamma}$ is constructed assuming spherically isotropic noise. The postfilter is calculated as in (7), where the minimum gain is set to -20 dB and the speech PSD estimate $\hat{\Phi}_s(l)$ is obtained via the decision-directed approach [26]. An estimate of the signal PSD matrix is obtained using recursive smoothing as

$$\hat{\mathbf{\Phi}}_{\mathbf{y}}(l) = \alpha\hat{\mathbf{\Phi}}_{\mathbf{y}}(l-1) + (1-\alpha)\mathbf{y}(l)\mathbf{y}^H(l), \tag{17}$$

with smoothing constant $\alpha = 0.67$ corresponding to approximately 40 ms. We consider four different methods to estimate the diffuse PSD $\hat{\Phi}_d(l)$:

- EIG_1 and EIG_2 denote the PSD estimates corresponding to (14) and (13) obtained via the *full* EVD. Note that the full EVD is obtained utilizing the MATLAB function `eig`, which utilizes a QR-decomposition-based algorithm.
- PI_1 and PI_2 denote the PSD estimates where the two largest eigenvalues are computed with the power method.

The accuracy of the PSD estimates is evaluated using the total PSD estimation error (averaged over all frames and frequencies) with respect to the true PSD [27]. The true PSD is determined intrusively from the late reverberant and diffuse noise component. Hereby, late

	array geometry	mic. distance	θ	T_{60}
AS_1 [22]	linear	$d = 8\,\text{cm}$	45°	$0.61\,\text{s}$
AS_2 [23]	circular	$r = 10\,\text{cm}$	45°	$0.73\,\text{s}$
AS_3 [24]	linear	$d = 6\,\text{cm}$	-15°	$1.25\,\text{s}$

Table 1: Configuration of considered acoustic scenarios; d: inter-microphone distance, r: circle radius, θ: speaker direction of arrival

reverberant component refers to the speech component originating from the portion of the RIRs excluding the first 8 ms. The performance of the MWF using the considered PSD estimators is evaluated in terms of the frequency-weighted segmental SNR (fwsSNR) [28] as well as the perceptual evaluation of speech quality (*PESQ*) [29] measure.

5.2. Initialization and Convergence Speed

In this section, the influence of the initialization $\mathbf{u}_m^{(0)}$ and the number of iterations N on the PSD estimation accuracy of the power method is investigated. Here, we only consider AS_1 with $M = 4$ and an input SNR of 20 dB, but similar results are obtained for all other scenarios.

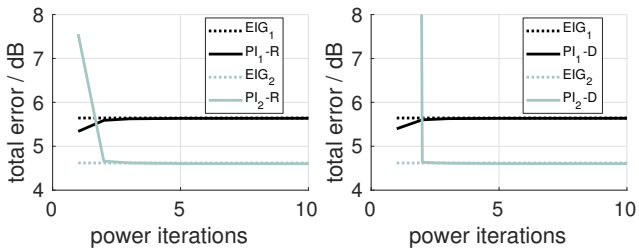
We either initialize $\mathbf{u}_m^{(0)}$ for each time-frequency bin deterministically using $\mathbf{u}_m^{(0)} = [1\,0\,0\,...\,0]^T$ or randomly using an M-dimensional real-valued vector with Gaussian distributed elements and variance 1. In the latter case, the simulations are repeated 5 times.

Fig. 1 depicts the total PSD estimation error based on the full EVD (EIG_1 and EIG_2) as well as the power method, either with random initialization (PI_1-R and PI_2-R) or with deterministic initialization (PI_1-D and PI_2-D). For the random initialization in Fig. 1, both the mean value (solid line) as well as the standard deviation (shaded area) are shown. First, it can be observed that for all considered initializations the convergence speed is fast, i.e., after only 2-3 iterations the power method-based PSD estimates converge to the PSD estimates obtained using the full EVD. The specific number of iterations required for convergence depends on the considered acoustic scenario; however, it has never exceeded $N = 3$ in the acoustic scenarios we have considered. Second, it can be observed that the standard deviation for the random initialization is small and not even visible at this scale.

In summary, using the power method to compute the two largest eigenvalues yields a similar diffuse PSD estimation accuracy as using the full EVD, with convergence reached after only a few iterations and with the initialization method having no significant influence.

5.3. Performance for Different Acoustical Systems

In this section we compare the performance of the MWF with different PSD estimates obtained using either the full EVD or the power method for all considered acoustic systems, microphone configurations, and input SNRs. Based on the results from the previous section, for the power method-based PSD estimates we used a random initialization and a fixed number of iterations ($N = 2$). Fig. 2 shows the performance in terms of ΔfwsSNR and ΔPESQ for different input SNRs for $M = \{4, 6\}$ microphones, averaged over the acoustic scenarios detailed in Table 1. As expected, in general the configuration with $M = 6$ leads to better results than the one with $M = 4$. In terms of both performance measures, it can be observed that there



(a) random initialization (b) deterministic initialization

Fig. 1: Total PSD estimation error vs. number of power iterations for the exemplary acoustic scenario 1; 20 dB input SNR; $M = 4$

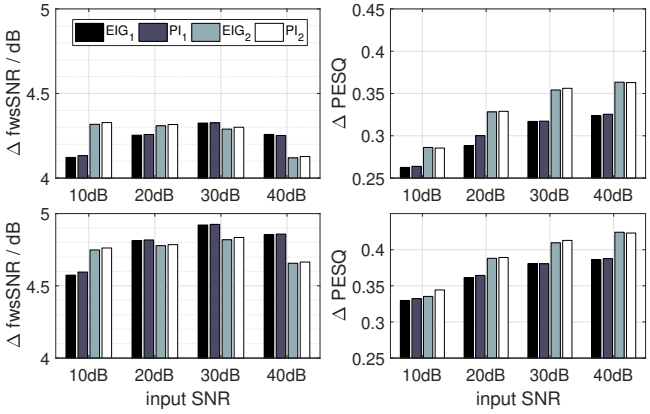


Fig. 2: Average ΔfwsSNR (left) and ΔPESQ (right) of the MWF with different PSD estimates (top: $M = 4$, bottom: $M = 6$)

is no large difference between the performance of the MWF using either the full EVD or the power method. Complying with the experimental findings in [17], the performance of the MWF using either the first or second eigenvalue is similar.

6. CONCLUSION

In this paper we have proposed to use the power method to reduce the computational complexity of the EVD-based diffuse PSD estimator. It is shown that the initialization does not significantly influence the convergence speed of the power method-based PSD estimate and that even for a small number of iterations ($N = 2$), diffuse PSD estimators are obtained that lead to the same performance as the computationally more complex full EVD.

7. REFERENCES

[1] R. Beutelmann and T. Brand, "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," *Journal of the Acoustical Society of America*, vol. 120, no. 1, pp. 331–342, July 2006.

[2] A. Warzybok, I. Kodrasi, J. O. Jungmann, E. A. P. Habets, T. Gerkmann, A. Mertins, S. Doclo, B. Kollmeier, and S. Goetze, "Subjective speech quality and speech intelligibility evaluation of single-channel dereverberation algorithms," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sept. 2014, pp. 333–337.

[3] T. Yoshioka, A. Sehr, M. Delcroix, K. Kinoshita, R. Maas, T. Nakatani, and W. Kellermann, "Making machines understand us in reverberant rooms: Robustness against reverberation for automatic speech recognition," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 114–126, Nov. 2012.

[4] F. Xiong, B. T. Meyer, N. Moritz, R. Rehr, J. Anemüller, T. Gerkmann, S. Doclo, and S. Goetze, "Front-end technologies for robust ASR in reverberant environments–spectral enhancement-based dereverberation and auditory modulation filterbank features," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, Aug. 2015.

[5] P. A. Naylor and N. D. Gaubitch, Eds., *Speech dereverberation*, Springer, London, UK, 2010.

[6] S. Doclo, W. Kellermann, S. Makino, and S. E. Nordholm, "Multichannel signal enhancement algorithms for assisted listening devices: Exploiting spatial diversity using multiple microphones," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 18–30, Mar. 2015.

[7] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Darmstadt, Germany, Sept. 2001, pp. 31–34.

[8] E. A. P. Habets and J. Benesty, "A two-stage beamforming approach for noise reduction and dereverberation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 945–958, May 2013.

[9] B. Cauchi, I. Kodrasi, R. Rehr, S. Gerlach, A. Jukić, T. Gerkmann, S. Doclo, and S. Goetze, "Combination of MVDR beamforming and single-channel spectral processing for enhancing noisy and reverberant speech," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, 2015.

[10] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using reference signals and a maximum likelihood estimator," in *Proc. European Signal Processing Conference*, Marrakech, Morocco, Sept. 2013.

[11] O. Schwartz, S. Braun, S. Gannot, and E. A. P. Habets, "Maximum likelihood estimation of the late reverberant power spectral density in noisy environments," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct. 2015, pp. 1–5.

[12] O. Schwartz, S. Gannot, and E. A. P. Habets, "Joint maximum likelihood estimation of late reverberant and speech power spectral density in noisy environments," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Shanghai, China, Mar. 2016, pp. 151–155.

[13] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood PSD estimation for speech enhancement in reverberation and noise," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1595–1608, Sept. 2016.

[14] O. Schwartz, S. Gannot, and E. A. P. Habets, "Joint estimation of late reverberant and speech power spectral densities in noisy environments using Frobenius norm," in *Proc. European Signal Processing Conference*, Budapest, Hungary, Sept. 2016, pp. 1123–1127.

[15] O. Thiergart and E. A. P. Habets, "Extracting reverberant sound using a linearly constrained minimum variance spatial filter," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 630–634, May 2014.

[16] I. Kodrasi and S. Doclo, "EVD-based multi-channel dereverberation of a moving speaker using different RETF estimation methods," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, San Francisco, USA, Mar. 2017, pp. 116–120.

[17] I. Kodrasi and S. Doclo, "Late reverberant power spectral density estimation based on an eigenvalue decomposition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, New Orleans, USA, Mar. 2017, pp. 611–615.

[18] G. Golub and C. Van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, USA, 1996.

[19] R. Varzandeh, M. Taseska, and E. A. P. Habets, "An iterative multichannel subspace-based covariance subtraction method for relative transfer function estimation," in *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, San Francisco, USA, Mar. 2017, pp. 11–15.

[20] B. F. Cron and C. H. Sherman, "Spatial-correlation functions for various noise models," *The Journal of the Acoustical Society of America*, vol. 34, no. 11, pp. 1732–1736, Nov. 1962.

[21] J. M. Ortega and H. F. Kaiser, "The LL^T and QR methods for symmetric diagonal matrices," *The Computer Journal*, vol. 6, no. 1, pp. 99–101, Jan. 1963.

[22] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. International Workshop on Acoustic Echo and Noise Control*, Antibes, France, Sept. 2014, pp. 313–317.

[23] K. Kinoshita, M. Delcroix, T. Yoshioka, T. Nakatani, A. Sehr, W. Kellermann, and R. Maas, "The REVERB challenge: A common evaluation framework for dereverberation and recognition of reverberant speech," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2013, pp. 1–4.

[24] J. Eaton, N. D. Gaubitch, A. H. Moore, and P. A. Naylor, "The ACE challenge - Corpus description and performance evaluation," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, USA, Oct. 2015.

[25] E. A. P. Habets, I. Cohen, and S. Gannot, "Generating nonstationary multisensor signals under a spatial coherence constraint," *Journal of the Acoustical Society of America*, vol. 124, no. 5, pp. 2911–2917, Nov. 2008.

[26] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 32, no. 6, pp. 1109–1121, Dec. 1984.

[27] R. C. Hendriks, J. Jensen, and R. Heusdens, "Noise tracking using DFT domain subspace decompositions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 541–553, March 2008.

[28] S. Quackenbush, T. Barnwell, and M. Clements, *Objective measures of speech quality*, Prentice-Hall, New Jersey, USA, 1988.

[29] ITU-T, *Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs P.862*, International Telecommunications Union (ITU-T) Recommendation, Feb. 2001.