Power method for robust diagonal unloading localization beamforming

Published
DOI:10.1109/LSP.2019.2908245

Power Method for Robust Diagonal Unloading Localization Beamforming

Daniele Salvati, Carlo Drioli, *Member, IEEE*, and Gian Luca Foresti, *Senior Member, IEEE*

***Abstract*—We propose a robust version of the diagonal unloading (DU) beamforming for the acoustic source localization problem in high noise conditions. The DU beamformer exploits the subspace orthogonality property by the removal or the attenuation of the signal subspaces, obtained through the subtraction of an opportune diagonal matrix from the covariance matrix. As a result, it provides high resolution directional response with low computational complexity. We show that a robust DU beamformer can be implemented by subtracting the largest eigenvalue of the estimated covariance matrix from the diagonal elements, and that this implementation is valid in general (i.e., for both the single-source and the multiple-source case). We propose the use of the power method for the estimation of the largest eigenvalue in the DU procedure. We show with numerical simulations that the proposed method improves the localization performance in high noise conditions without substantial increment of the computational cost. Applications for this method include a number of scenarios involving multirotor aerial systems due to its robustness to the noise and its low computational complexity.**

***Index Terms*—Robust diagonal unloading beamforming, power method, acoustic source localization, microphone array, noisy environment, multirotor aerial system, drone.**

I. Introduction

ACOUSTIC source localization (ASL) is an important task in microphone array processing and it is of interest in an increasing number of applications such as teleconferencing, surveillance, animal ecology, human-computer interaction, hearing aid, volcanology, medicine, robotics [1]–[13]. Recently, ASL has been recognized to provide interesting application perspectives in a number of scenarios involving multirotor aerial systems [14]–[22]. For example, in aerial surveillance for ground security or search and rescue operations, the localization and recognition of acoustic sources is highly desirable in case of visual occlusion. In these scenarios, the ASL is performed in a low signal-to-noise ratio (SNR) environment. Basically, two main characteristics are strongly required for multichannel signal processing in open air drone applications: 1) robustness to noise; 2) low-complexity for real-time processing.

To address these requirements, we propose a robust diagonal unloading (DU) localization beamforming based on the power method [23] for the estimation of the largest eigenvalue of the available covariance matrix. The DU beamforming, recently proposed in [24], provides high resolution directional response and noise robustness comparable to those of the multiple signal classification (MUSIC) method [25] and the minimum variance distortionless response (MVDR) beamformer [26] while requiring less computational resources. In fact, the MUSIC method and the MVDR beamformer require an eigendecomposition and a matrix inversion operation, respectively, that lead to an increased computation cost in broadband applications. Conversely, the DU beamforming has the same complexity of the conventional steered response power beamforming [27], since the DU beamformer is based on the subtraction of an opportune diagonal matrix from the covariance matrix. The DU beamformer is hence very attractive in drone applications, in which noise robustness and real-time processing are highly desirable. However, the DU beamformer was proposed in [24] with an optimal solution in single-source scenario with spatially white noise and true covariance matrix, and with a suboptimal solution, valid for a broader class of acoustic conditions (single-source, multi-source, anechoic or reverberant environment) when the covariance matrix is estimated (as it is in real applications). This suboptimal solution has been shown to be effective in reverberant and moderate noisy conditions with speech signals [24].

The DU beamformer exploits the orthogonality property between signal and noise subspaces by removing in practice the signal subspace (or subspaces) from the covariance matrix of the input signals of the array, i.e., by subtracting an opportune diagonal matrix from the covariance matrix. We will refer here to a robust implementation of this process as the best solution to the diagonal removal problem to achieve the subspace orthogonality property given the estimated covariance matrix. The novelty of this letter is two-fold. First, we show that the robust DU beamformer can be implemented by subtracting the largest eigenvalue from the diagonal elements of the estimated covariance matrix. Then, we propose the use of the power method in the DU procedure for the estimation of the largest eigenvalue. Beside that, we show that the proposed algorithm improves the localization performance, in terms of direction of arrival (DOA) estimation, in high noise conditions if compared to the DU suboptimal solution of [24].

II. Robust Diagonal Unloading Localization Beamforming

A. Model

Let us refer to a microphone array with M omnidirectional sensors, and to a far-field model for the sound source wave propagation. Suppose that the sound wave from an acoustic source impinges upon the array with a direction $\boldsymbol{\Omega}_s = [\theta_s, \phi_s]$ (θ_s and ϕ_s are the azimuth and elevation angles). In the short-time Fourier transform domain, the data model of the

D. Salvati, C. Drioli, and G.L. Foresti are with the Department of Mathematics, Computer Science and Physics, University of Udine, Udine 33100, Italy, e-mail: daniele.salvati@uniud.it, carlo.drioli@uniud.it, gianluca.foresti@uniud.it.

array signals can be expressed in single-source scenario with spatially white noise as

$$\mathbf{x}(k,f)=\mathbf{a}(f,\mathbf{\Omega}_s)S(k,f)+\mathbf{v}(k,f), \tag{1}$$

where k is the block time index, f is the frequency bin, $S(k,f)$ is the source signal at the reference sensor, $\mathbf{v}(k,f)$ is the noise that is assumed to be spatially white Gaussian with zero mean and variance equals to σ^2 for all sensors, and $\mathbf{a}(f,\mathbf{\Omega}_s)$ is the array steering vector for the source direction.

The output of a beamformer $Y(k,f,\mathbf{\Omega})$ at block time k, for frequency f and look direction $\mathbf{\Omega}=[\theta,\phi]$, is obtained as

$$Y(k,f,\mathbf{\Omega})=\mathbf{w}^H(k,f,\mathbf{\Omega})\mathbf{x}(k,f), \tag{2}$$

where $\mathbf{w}(k,f,\mathbf{\Omega})$ is a column vector containing the beamformer coefficients for time-shifting, weighting, and summing the data, so to steer the array in the direction $\mathbf{\Omega}$, and H denotes the Hermitian transpose. Then, the power spectral density (PSD) of the spatially filtered signal is

$$\begin{aligned}P(k,f,\mathbf{\Omega})&=E\{|Y(k,f,\mathbf{\Omega})|^2\}\\&=\mathbf{w}^H(k,f,\mathbf{\Omega})\mathbf{\Phi}(k,f)\mathbf{w}(k,f,\mathbf{\Omega}),\end{aligned} \tag{3}$$

where $\mathbf{\Phi}(k,f)=E\{\mathbf{x}(k,f)\mathbf{x}^H(k,f)\}$ is the covariance matrix of the array signal, which is symmetric and positive definite, and $E\{\cdot\}$ denotes mathematical expectation. In the conventional steered response power (SRP) beamformer, whose implementation reflects the delay-and-sum scheme, all its weights are equal in magnitude, i.e. $\mathbf{w}_{\text{SRP}}(k,f,\mathbf{\Omega})=\mathbf{a}(f,\mathbf{\Omega})$. The $P(k,f,\mathbf{\Omega})$ is related to the power contribution of a single frequency bin, and a function providing the steered response power information of the whole frequency spectrum can be obtained by merging the contribution by some fusion criterion, such as the normalized frequency fusion proposed in [28], and defined as

$$P(k,\mathbf{\Omega})=\sum_{f=f_{\min}}^{f_{\max}}\frac{P(k,f,\mathbf{\Omega})}{||\mathbf{p}(k,f)||_\infty}, \tag{4}$$

where $||\cdot||_\infty$ denotes the uniform norm of the vector $\mathbf{p}(k,f)=[P(k,f,\mathbf{\Omega}_1),P(k,f,\mathbf{\Omega}_2),\dots,P(k,f,\mathbf{\Omega}_D)]$ that contains all the PSDs for the considered directions D, and $f_{\min}$ and $f_{\max}$ denote the frequency range for the computation of the broadband SRP. The broadband SRP resulting from the fusion is characterized by high energy peaks corresponding to those directions from which acoustic energy is sensed. For the single-source case, the DOA estimation of the source is provided by the maximum energy peak search

$$\hat{\mathbf{\Omega}}_s(k)=\underset{\mathbf{\Omega}}{\text{argmax}}[P(k,\mathbf{\Omega})]. \tag{5}$$

In the multi-source case, a given number (known a priori or estimated) of local maxima energy peaks are searched instead.

B. Optimal DU Beamformer in Single-Source Case with Spatially White Noise and True Covariance Matrix

The DU beamformer [24] is a data-dependent spatial filtering model that aims at exploiting the orthogonality property between signal and noise subspaces by subtracting an opportune diagonal matrix from the covariance matrix $\mathbf{\Phi}(k,f)$ of the array output vector. It follows from following the minimization problem

$$\begin{aligned}\text{minimize}\quad & ||\mathbf{w}(k,f,\mathbf{\Omega})-\mathbf{a}(f,\mathbf{\Omega})||^2,\\ \text{subject to}\quad & \mathbf{u}_s^H(k,f)\mathbf{w}(k,f,\mathbf{\Omega})=0,\end{aligned} \tag{6}$$

where $\mathbf{u}_s(k,f)$ is the signal subspace of $\mathbf{\Phi}(k,f)$, and $||\cdot||$ denotes the Euclidean norm. Using the method of Lagrange multipliers, the solution of (6) for the beamforming coefficients $\mathbf{w}$ is:

$$\mathbf{w}_{\text{DU}}(k,f,\mathbf{\Omega})=\Big(\frac{1}{\lambda}\mathbf{I}\Big)\mathbf{\Phi}_{\text{DU}}(k,f)\mathbf{a}(f,\mathbf{\Omega}), \tag{7}$$

where λ is the noise eigenvalue of the transformed matrix $\mathbf{\Phi}_{\text{DU}}(k,f)$ that can be written as

$$\mathbf{\Phi}_{\text{DU}}(k,f)=\mathbf{\Phi}(k,f)-\mu(k,f)\mathbf{I}, \tag{8}$$

where $\mu(k,f)$ is a real-valued, positive scalar, selected in such a way that its eigenvalue corresponding to the signal subspace is null, and $\mathbf{I}$ denotes identity matrix. The value of μ that satisfies such constraints in a single source case with spatially white noise can be shown to be [24]

$$\mu(k,f)=\text{tr}[\mathbf{\Phi}(k,f)]-(M-1)\sigma^2, \tag{9}$$

where tr$[\cdot]$ is the operator that computes the trace of a matrix. In fact, the covariance matrix can be decomposed in its eigenvalues and their associated eigenvectors through a subspace decomposition $\mathbf{\Phi}(k,f)=\mathbf{U}\text{diag}(MP_s(f)+\sigma^2,\sigma^2,\dots,\sigma^2)\mathbf{U}^H$, where $\mathbf{U}$ is the square matrix of eigenvectors, and $P_s(f)=E\{|S(f)|^2\}$ is the power of the signal. By applying the diagonal removal in (8) and (9), we have that the signal eigenvalue of $\mathbf{\Phi}_{\text{DU}}(k,f)$ becomes null, i.e., $MP_s(f)+\sigma^2-\text{tr}[\mathbf{\Phi}(k,f)]+(M-1)\sigma^2=MP_s(f)+\sigma^2-M(P_s(f)+\sigma^2)+(M-1)\sigma^2=0$, and the noise eigenvalue becomes negative, i.e., $\lambda=\sigma^2-\text{tr}[\mathbf{\Phi}(k,f)]+(M-1)\sigma^2=-\sigma^2-M(P_s(f)+\sigma^2)+(M-1)\sigma^2=-MP_s(f)$. Hence, the transformed matrix $\mathbf{\Phi}_{\text{DU}}(k,f)$ contains only the noise eigenvectors and it is negative semidefinite. Substituting (7) in (3), we have $P'_{\text{DU}}(f,\mathbf{\Omega})=\frac{\sigma^2}{\lambda^3}\mathbf{a}^H(f,\mathbf{\Omega})\mathbf{\Phi}_{\text{DU}}(f)\mathbf{a}(f,\mathbf{\Omega})$, where the quantity $\frac{\sigma^2}{\lambda^3}$ is a scalar factor that can be omitted since it has no influence on the DOA estimation. Since $\mathbf{\Phi}_{\text{DU}}(k,f)$ is negative semidefinite, i.e., $P'_{\text{DU}}(f,\mathbf{\Omega})\le 0$, we can write the pseudo-spectrum in the equivalent form

$$P_{\text{DU}}(f,\mathbf{\Omega})=\frac{-1}{\mathbf{a}^H(f,\mathbf{\Omega})\mathbf{\Phi}_{\text{DU}}(k,f)\mathbf{a}(f,\mathbf{\Omega})}. \tag{10}$$

C. Robust DU Beamformer with Estimated Covariance Matrix Using the Power Method

In real-world applications, the covariance matrix $\mathbf{\Phi}(k,f)$ is unknown and it has to be estimated. In general, the estimation can be computed through the averaging of the array signal blocks [29]

$$\widehat{\mathbf{\Phi}}(k,f)=\frac{1}{B}\sum_{k_b=0}^{B-1}\mathbf{x}(k-k_b,f)\mathbf{x}^H(k-k_b,f), \tag{11}$$

where B is the number of snapshots for the averaging. There is always a certain mismatch between the estimated and the true covariance matrix, due to the finite sample size (number of snapshots), to the signal model mismatches, and to the nonstationary nature of the source. The solution in (9) is based on an ideal model of single source with spatially white noise, which is easily violated in practice due to the model mismatch or when operated in multi-source scenarios.

A general data model of the array signals can be expressed as

$$\mathbf{x}(k,f) = \sum_{n=1}^{N} (\mathbf{a}(f, \mathbf{\Omega}_{s_n}) S_n(k,f)) + \mathbf{v}(k,f), \qquad (12)$$

where N denotes the number of sources, $\mathbf{a}(f, \mathbf{\Omega}_{s_n})$ is the array steering vector for the n-th source direction $\mathbf{\Omega}_{s_n}$, and $\mathbf{v}(k,f)$ is the noise component. A practical DU solution is given in [24] by assuming $\mu(k,f)' = \mathrm{tr}[\widehat{\mathbf{\Phi}}(k,f)]$, which is a suboptimal solution:

$$P_{\mathrm{DU}}^{\mathrm{subopt}}(f,\mathbf{\Omega}) = \frac{-1}{\mathbf{a}^H(f,\mathbf{\Omega})(\widehat{\mathbf{\Phi}}(k,f) - \mathrm{tr}[\widehat{\mathbf{\Phi}}(k,f)]\mathbf{I})\mathbf{a}(f,\mathbf{\Omega})}. \qquad (13)$$

This solution guarantees that the transformed covariance matrix is negative semidefinite, allowing the exploitation of the orthogonality property, which is however affected by a certain quantity of signal subspace (or signal subspaces) in the transformed covariance matrix.

The proposed robust diagonal unloading beamforming is based on the estimation, computed by the power method, of the largest eigenvalue of the covariance matrix. We can write the estimated eigenvalue matrix of the estimated covariance matrix $\widehat{\mathbf{\Phi}}(k,f)$ at time block k, organizing the eigenvalues of $\widehat{\mathbf{\Phi}}(k,f)$ in descending order ($\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_M$) as $\widehat{\mathbf{\Lambda}} = \mathrm{diag}(\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_M)$. The eigenvalue matrix of the transformed covariance matrix can be written as $\widehat{\mathbf{\Lambda}}_{\mathrm{DU}} = \mathrm{diag}(\widehat{\lambda}_1 - \mu(k,f), \widehat{\lambda}_2 - \mu(k,f), \ldots, \widehat{\lambda}_M - \mu(k,f))$. We can easily see that the robust DU implementation in the single-source case is obtained by assuming that the parameter $\mu(k,f)$ is equal to the largest eigenvalue $\widehat{\lambda}_1$. This allows the best removal of the signal subspace in the transformed covariance matrix $\widehat{\mathbf{\Phi}}_{\mathrm{DU}}(k,f)$. We now assume the case of N sources, i.e., $\widehat{\lambda}_n$ ($n = 1, 2, \ldots, N$) are signal eigenvalues. By considering that the transformed covariance matrix has to be negative semidefinite to exploit the subspace orthogonality property, meaning that each eigenvalue of the matrix has a value less than or equal to zero, we have that the parameter $\mu(k,f)$ has to be greater than or equal to the largest eigenvalue $\widehat{\lambda}_1$. We can thus write a generalized parameter $\mu(k,f) = \widehat{\lambda}_1 + \alpha$, where α is a real positive value. The optimal solution, which aims at reducing as maximum as possible the N signal eigenvectors, can be computed by solving the following maximization problem for the signal eigenvalues of $\widehat{\mathbf{\Phi}}_{\mathrm{DU}}(k,f)$:

$$\begin{aligned} &\text{maximize} \quad \sum_{n=1}^{N} \widehat{\lambda}_n - N(\widehat{\lambda}_1 + \alpha), \\ &\text{subject to} \quad \alpha \geq 0. \end{aligned} \qquad (14)$$

We have that the cost function can be written as $\sum_{n=2}^{N} \widehat{\lambda}_n - (N-1)\widehat{\lambda}_1 - N\alpha$. Since $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_N$, we have that $(N-1)\widehat{\lambda}_1 > \sum_{n=2}^{N} \widehat{\lambda}_n$. The sum of eigenvalues in the cost function is always negative, and thus the solution is given for $\alpha = 0$. Hence, we can say that the robust DU solution for an available covariance matrix with a general model is obtained by imposing $\mu^{\mathrm{rob}}(k,f) = \widehat{\lambda}_1(k,f)$.

Hence, the proposed robust DU beamforming becomes

$$P_{\mathrm{DU}}^{\mathrm{rob}}(f,\mathbf{\Omega}) = \frac{-1}{\mathbf{a}^H(f,\mathbf{\Omega})(\widehat{\mathbf{\Phi}}(k,f) - \widehat{\lambda}_1(k,f)\mathbf{I})\mathbf{a}(f,\mathbf{\Omega})}. \qquad (15)$$

The robust DU implementation requires that the largest eigenvalue of $\widehat{\mathbf{\Phi}}(k,f)$ has to be estimated. To avoid the use of the eigendecomposition, which has $O(M^3)$ complexity, we use herein the power method, which has $O(M^2)$ complexity. The power method is an iterative procedure for approximating the largest eigenvalue and the corresponding eigenvector of a matrix [23]. The iterative sequence is given by ($i = 0, 1, \ldots$):

$$1) \quad \mathbf{g}_u(i+1) = \widehat{\mathbf{\Phi}}(k,f)\mathbf{g}(i), \quad 2) \quad \mathbf{g}(i+1) = \frac{\mathbf{g}_u(i+1)}{||\mathbf{g}_u(i+1)||}, \qquad (16)$$

where $\mathbf{g}$ is a weight vector. After I iterations the largest eigenvalue is estimated as:

$$\widehat{\lambda}_1(k,f) = \frac{\mathbf{g}_u^H(I+1)\mathbf{g}(I)}{\mathbf{g}^H(I)\mathbf{g}(I)}. \qquad (17)$$

The weight vector $\mathbf{g}$ is initialized with arbitrary nonzero values, and the iteration is computed until a convergence criterion is satisfied. We adopt the threshold criterion $\mathbf{e}(i+1) = ||\mathbf{g}(i+1) - \mathbf{g}(i)|| < \epsilon$. Note that the weight vector needs to be normalized during the iterations to prevent it from becoming too large or too small. The rate of convergence of the power method depends upon the ratio $\widehat{\lambda}_2/\widehat{\lambda}_1$ (i.e., it has linear convergence). If $\widehat{\lambda}_1 = \widehat{\lambda}_2$ the method may not converge.

The computational cost for the broadband robust DU can be expressed in terms of the approximated floating-point operation (FLOP, either a real multiplication or a real summation). Let L denote the frame size for the fast Fourier transform (FFT), we obtain $BM(4L\log_2 L - 6L + 8)$ FLOPs for the FFTs of M channels for B snapshots. Let F denote the number of frequency bins, we obtain $M^2F(6+2B)$ FLOPs for the estimation of covariance matrices. The steered response power (3) requires $FD(7M^2+7M-2)$ FLOPs with D being the number of considered search directions. The suboptimal DU operation adds $F(M-1)$ summations for the trace operation and the diagonal removal requires FM subtractions. The broadband fusion adds $3FD - F - D$ FLOPs. The robust DU requires the estimation of the largest eigenvalue with the power method that requires $IF(8M^2 + 6M)$ FLOPs. The MUSIC [25], [28] instead requires an eigendecomposition and the product of the noise subspace with the corresponding conjugate transpose that have $F(21M^3 - 8M^2 + 2M)$ FLOPs. The MUSIC has thus a cubic complexity $O(M^3)$ that becomes significant at increasing of the array size. The overall computational cost for each method is summarized in Table I.

TABLE I
THE COMPUTATIONAL COST EXPRESSES IN TERMS OF THE APPROXIMATED NUMBER OF FLOPS.

Suboptimal DU	$BM(4L\log_2 L - 6L + 8) + M^2F(7D + 2B + 6) + MF(7D + 2) + F(D - 2) - D$
Robust DU	$BM(4L\log_2 L - 6L + 8) + M^2F(7D + 2B + 8I + 6) + MF(7D + 1 + 6I) + F(D - 1) - D$
MUSIC	$BM(4L\log_2 L - 6L + 8) + 21M^3F + M^2F(7D + 2B - 2) + MF(7D + 2) + F(D - 1) - D$

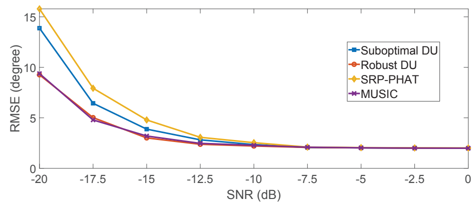

Fig. 1. Localization performance of a single source at variation of SNR level. The number of snapshots was 25.

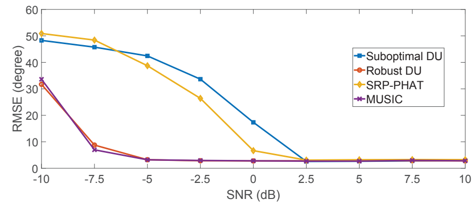

Fig. 2. Localization performance of two sources at variation of SNR level. The number of snapshots was 25.

III. Simulations

The localization performance is illustrated through a set of simulated experiments in noisy conditions adding mutually independent white Gaussian noise to each channel with a wideband SNR defined as $10\log_{10}\frac{E\{\sum_{n=1}^{N}|s_n(t)|^2\}}{\sigma_v^2}$ ($s_n(t)$ is the time-domain n-th source signal at time t in the first channel, and σ_v^2 is the wideband noise power). A circular uniform array of 8 microphones and radius 20 cm was used. The spatial resolution was 5 degrees. The sampling frequency was 48 kHz, the block size L was 2048 samples, and a Hann windowing was used. The tolerance ϵ for the power method was set to 10^{-3}. We considered 25 random source positions with 10 trial repetitions for each position. Recording of a drone sound was used as source signal. The drone sound has a concentrated energy up to 6000 Hz, which consists of a broadband aerodynamic noise induced by the propellers and nonstationary narrowband components originated by the electrical engines. We have compared the DOA localization performance of the robust DU, the suboptimal DU [24], the MUSIC [25], [28], and the conventional SRP beamforming [27] with the phase transform (PHAT) normalization [30]. The localization beamforming was limited to the [150-6000] Hz frequency range for all methods. Performance is reported in terms of the root mean square error (RMSE) for all the estimates (RMSE $=\sqrt{\frac{\sum_k\sum_{n=1}^{N}((\theta_{s_n}-\widehat{\theta}_n(k))^2+(\phi_{s_n}-\widehat{\phi}_n(k))^2)}{K}}$, K is total number of estimates). In the first set of simulations, the single-source localization performance at variation of SNR level was evaluated. The number of snapshots was 25 for all methods. The results are reported in Figure 1. As we can observe, all methods have the same performance at an SNR of 0 dB, and the proposed robust DU outperforms the suboptimal DU and SRP-PHAT when the noise level increases. The robust DU performance is similar to that of the MUSIC. Next, the localization performance in a multi-source case was evaluated, by using two drone signal sources. The number of snapshots was 25. The results at variation of SNR level are shown in Figure 2. The proposed robust DU significantly outperforms the suboptimal DU and the SRP-PHAT at increasing of the noise level, achieving an RMSE comparable to that of the MUSIC. The trace-based diagonal removal for the suboptimal DU (13) becomes ineffective for the suppression of the signal subspaces from the covariance matrices at low SNR levels. Finally, simulations using two sources at variation of number of snapshots were performed. The SNR was -5 dB. The results are depicted in Figure 3. The robust DU improves the performance if compared to the suboptimal DU at increasing of number of snapshots. We can see that the robust and the suboptimal DU provide the same performance in the case of single snapshot. This is due to the fact that the estimated covariance matrix has only one non-null eigenvalue. In this case the trace of the covariance matrix, which is needed for the suboptimal DU implementation, is equal to the largest eigenvalue. The suboptimal DU degrades at increasing of number of snapshots (i.e., when the estimated covariance matrix becomes more accurate), as it does not provide a sufficient attenuation of the signal subspaces in low SNR conditions since the trace contains both signals and noise eigenvalues. Considering $M=8$, $L=2048$, $B=25$, $F=251$, and $D=1296$, we have that the suboptimal DU requires $1.807\cdot10^8$ FLOPs. We measured an average number of iterations $I=20$ for the power method in the simulations, and hence, the robust DU requires $1.836\cdot10^8$ FLOPs without adding significant computational cost.

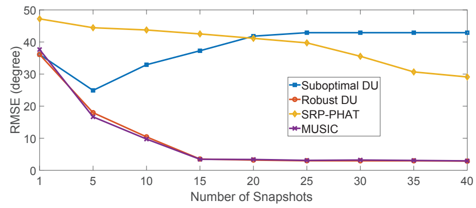

Fig. 3. Localization performance of two sources (SNR=-5 dB) at variation of number of snapshots.

IV. Conclusions

We have proposed a robust DU beamforming that improves acoustic DOA estimation in high noise conditions. We discussed how to set the diagonal removal procedure of the covariance matrix to obtain the best suppression of signal subspaces in a general model, exploiting as much as possible the subspace orthogonality property that provides high resolution directional response and noise robustness. We demonstrated that the robust DU beamformer can be implemented by subtracting the largest eigenvalue from the diagonal elements of the estimated covariance matrix. We have proposed the use of the power method for the estimation of the largest eigenvalue without adding significant computational cost. The proposed method can be attractive for microphone array applications in a number of scenarios involving multirotor aerial systems due to the noise robustness similar to that of the MUSIC without however the computational cost of the eigendecomposition for each narrowband component.

REFERENCES

[1] D. Salvati and S. Canazza, "Incident signal power comparison for localization of concurrent multiple acoustic sources," *The Scientific World Journal*, vol. 2014, pp. 1–13, 2014.

[2] D. J. Mennill, M. Battiston, D. R. Wilson, J. R. Foote, and S. M. Doucet1, "Field test of an affordable, portable, wireless microphone array for spatial monitoring of animal ecology and behaviour," *Methods in Ecology and Evolution*, vol. 3, no. 4, pp. 704–712, 2012.

[3] D. Salvati, S. Canazza, and A. Rodà, "A sound localization based interface for real-time control of audio processing," in *Proceedings of the International Conference on Digital Audio Effects*, 2011, pp. 177–184.

[4] S. Doclo, M. Moonen, T. V. den Bogaert, and J. Wouters, "Reduced bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 38–51, 2009.

[5] D. Salvati, C. Drioli, and G. L. Foresti, "Exploiting a geometrically sampled grid in the steered response power algorithm for localization improvement," *Journal of the Acoustical Society of America*, vol. 141, no. 1, pp. 586–601, 2017.

[6] C. R. Rowell, D. Fee, C. A. S. K. Arnoult, R. S. Matoza, P. P. Firstov, K. Kim, and E. Makhmudov, "Three-dimensional volcano-acoustic source localization at Karymsky Volcano, Kamchatka, Russia," *Journal of Volcanology and Geothermal Research*, vol. 283, pp. 101–115, 2014.

[7] D. Salvati, C. Drioli, and G. L. Foresti, "Sound source and microphone localization from acoustic impulse responses," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1459–1463, 2016.

[8] Y. Li, K. C. Ho, and M. Popescu, "A microphone array system for automatic fall detection," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 5, pp. 1291–1301, 2012.

[9] D. Salvati, C. Drioli, and G. L. Foresti, "Sensitivity-based region selection in the steered response power algorithm," *Signal Processing*, vol. 153, pp. 1–100, 2018.

[10] S. Argentieri, P. Danes, and P. Soueres, "A survey on sound source localization in robotics: from binaural to array processing methods," *Computer Speech and Language*, vol. 34, no. 1, pp. 87–112, 2015.

[11] D. Salvati and S. Canazza, "Improvement of acoustic localization using a short time spectral attenuation with a novel suppression rule," in *Proceedings of the International Conference on Digital Audio Effects*, 2009, pp. 150–156.

[12] J. M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, 2003, pp. 1228–1233.

[13] R. Levorato and E. Pagello, "DOA acoustic source localization in mobile robot sensor networks," in *Proceeding of the IEEE International Conference on Autonomous Robot Systems and Competitions*, 2015, pp. 71–76.

[14] M. Basiri, F. Schill, P. Lima, and D. Floreano, "On-board relative bearing estimation for teams of drones using sound," *IEEE Robotics and Automation Letters*, vol. 1, no. 2, pp. 820–827, 2016.

[15] L. Wang and A. Cavallaro, "Acoustic sensing from a multi-rotor drone," *IEEE Sensors Journal*, vol. 18, no. 1, pp. 4570–4582, 2018.

[16] D. Salvati, C. Drioli, G. Ferrin, and G. L. Foresti, "Beamforming-based acoustic source localization and enhancement for multirotor UAVs," in *Proceedings of the 26th European Signal Processing Conference (EUSIPCO)*, 2018.

[17] L. Wang and A. Cavallaro, "Time-frequency processing for sound source localization from a micro aerial vehicle," in *Proceeding of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 496–500.

[18] K. Hoshiba, K. Washizaki, M. Wakabayashi, T. Ishiki, M. Kumon, Y. Bando, D. Gabriel, K. Nakadai, and H. G. Okuno, "Design of UAV-embedded microphone array system for sound source localization in outdoor environments," *Sensors*, vol. 17, no. 11, 2017.

[19] T. Ishiki and M. Kumon, "Design model of microphone arrays for multirotor helicopters," in *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2015, pp. 6143–6148.

[20] K. Furukawa, K. Okutani, K. Nagira, T. Otsuka, K. Itoyama, K. Nakadai, and H. G. Okuno, "Noise correlation matrix estimation for improving sound source localization by multirotor UAV," in *Proceeding of the International Conference on Intelligent Robots and Systems*, 2013, pp. 3943–3948.

[21] K. Okutani, T. Yoshida, K. Nakamura, and K. Nakadai, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *Proceeding of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 3288–3293.

[22] M. Basiri, F. Schill, P. U. Lima, and D. Floreano, "Robust acoustic source localization of emergency signals from micro air vehicles," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2012, pp. 4737–4742.

[23] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins University Press, 1996.

[24] D. Salvati, C. Drioli, and G. L. Foresti, "A low-complexity robust beamforming using diagonal unloading for acoustic source localization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 609–622, 2018.

[25] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[26] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.

[27] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 1053–5888, 1996.

[28] D. Salvati, C. Drioli, and G. L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 581–585, 2014.

[29] L. Zhang, W. Liu, and L. Yu, "Performance analysis for finite sample MVDR beamformer with forward backward processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2427–2431, 2011.

[30] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.