Online DOA estimation using real eigenbeam ESPRIT with propagation vector matching

Adrian Herzog, Emanuël Habets

ONLINE DOA ESTIMATION USING REAL EIGENBEAM ESPRIT WITH PROPAGATION VECTOR MATCHING

Adrian Herzog
International Audio Laboratories Erlangen*
adrian.herzog@audiolabs-erlangen.de

Emanuël A.P. Habets
International Audio Laboratories Erlangen*
emanuel.habets@audiolabs-erlangen.de

ABSTRACT

Eigenbeam ESPRIT (EB-ESPRIT) is a method to estimate multiple directions-of-arrival (DOAs) of sound sources from the spherical harmonics domain (SHD) coefficients of a spherical microphone array recording. Recently, an EB-ESPRIT variant based on three types of recurrence relations of complex spherical harmonics was proposed (DOA-vector EB-ESPRIT). However, due to the signal subspace computation and the joint diagonalization procedure, the computational cost might be too large for many real-time applications. In this work, we propose a computationally more efficient real-valued DOA-vector EB-ESPRIT. The signal subspace is estimated by the deflated projection approximation subspace tracking (PASTd) method. To avoid the joint diagonalization, we propose a subspace propagation-vector matching for the estimation of two DOAs. In the evaluation, we compare the performance of the complex and real DOA-vector EB-ESPRIT with an existing robust B-format DOA estimation method under noisy and reverberant conditions.

1. INTRODUCTION

For parametric time-frequency-domain spatial audio coding, the directions-of-arrival (DOAs) of sound sources have to be estimated from the microphone signals for each time-frame and frequency band. Accurately estimating these parameters in real-time is a challenging task.

For directional audio coding (DirAC) [1], one DOA has to be estimated per time-frame and frequency band, which can be done with low computational cost using the pseudointensity vector (PIV) [2]. The accuracy of the PIV is, however, quite limited compared to subspace-based DOA estimators [3]. For high angular resolution plane-wave expansion (HARPEX) [4], two DOAs are estimated from a B-format signal per time-frame and frequency band using properties of plane-wave propagation vectors. In [5], a robust version of this DOA estimator is proposed which computes a signal subspace prior to the DOA estimation.

Eigenbeam ESPRIT [6] uses recurrence relations of spherical harmonics to estimate multiple DOAs from a spherical harmonics domain signal. In [7–10], robust and unambiguous EB-ESPRIT variants have been developed. In [10], the authors proposed the DOA-vector EB-ESPRIT which can accurately estimate the DOAs. However, the computational complexity of the DOA-vector EB-ESPRIT is high due to the signal subspace estimation and the joint diagonalization procedure. Hence, a computationally more efficient version of the DOA-vector EB-ESPRIT is needed.

In this work, we propose a new DOA-vector EB-ESPRIT based on real spherical harmonics recurrence relations and the computationally efficient deflated projection approximation subspace tracking (PASTd) algorithm [11]. If only one DOA has to be estimated per time-frame and frequency band, the EB-ESPRIT equations can be simplified. For estimating two DOAs, we propose to first estimate the plane-wave propagation vectors using properties of the real spherical harmonics and then apply the simplified EB-ESPRIT equations for one DOA to both estimated propagation vectors, thereby, avoiding the joint diagonalization. This procedure is referred to as subspace propagation-vector matching in the remainder of this work.

In Sec. 2, spherical harmonic domain signals are introduced. In Sec. 3, the real DOA-vector EB-ESPRIT is derived. In Sec. 4, simplifications for an efficient online implementation are discussed. In Sec. 5, the proposed method is evaluated and compared to [10] and [5].

2. SPHERICAL HARMONICS DOMAIN

Let $p(k; r, \Omega)$ denote the sound pressure field on the surface of a spherical microphone array (SMA) with radius r, where k denotes the wavenumber and $\Omega = (\theta, \phi)$ the angular position on the sphere specified by the elevation $\theta \in [0, \pi]$ and azimuth $\phi \in [-\pi, \pi)$ angles. The pressure field can be expanded using the spherical harmonics expansion [12, 13] as

$$p(k; r, \Omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} b_l(kr) P_{lm}(k) Y_{lm}(\Omega) \,, \quad (1)$$

where l and m denote the *order* and *mode*, respectively. The radial dependencies $b_l(kr)$ are denoted as *mode strengths* which depend on the SMA properties only,

*A joint institution of the Friedrich Alexander University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

doi:10.25836/sasp.2019.18

$P_{lm}(k)$ are the spherical harmonic domain (SHD) coefficients and $Y_{lm}(\Omega)$ the complex spherical harmonic functions [12, 13].

The real spherical harmonics $R_{lm}(\Omega)$ are related to $Y_{lm}(\Omega)$ as follows [13]:

$$R_{lm}(\Omega) = \begin{cases} \frac{i}{\sqrt{2}}\left(Y_{lm}(\Omega) - (-1)^m Y_{l(-m)}(\Omega)\right) & m < 0 \\ Y_{l0}(\Omega) & m = 0 \\ \frac{1}{\sqrt{2}}\left(Y_{l(-m)}(\Omega) + (-1)^m Y_{lm}(\Omega)\right) & m > 0 \end{cases} \tag{2}$$

with $i^2 = -1$. Using the vector notation

$$\begin{aligned} \mathbf{y}(\Omega) &:= [Y_{00}, Y_{1-1}, Y_{10}, ..., Y_{LL}]^T(\Omega) \text{ and} \\ \mathbf{r}(\Omega) &:= [R_{00}, R_{1-1}, R_{10}, ..., R_{LL}]^T(\Omega)\,, \end{aligned} \tag{3}$$

where $(\cdot)^T$ denotes the transpose and L the maximum SHD order considered, we can write

$$\mathbf{r}(\Omega) = \mathbf{U}_L \mathbf{y}(\Omega) = \mathbf{U}_L^* \mathbf{y}^*(\Omega) \tag{4}$$

with the unitary $(L+1)^2 \times (L+1)^2$ matrix $\mathbf{U}_L$ and $(\cdot)^*$ denoting the complex conjugate. Let us define:

$$\begin{aligned} \mathbf{p}(k) &:= [P_{00}, P_{1-1}, P_{10}, P_{11}, ..., P_{LL}]^T(k) \text{ and} \\ \mathbf{B}(kr) &:= \text{diag}\left\{[b_0, b_1, b_1, b_1, ..., b_L]\right\}(kr)\,, \end{aligned} \tag{5}$$

where in $\mathbf{B}(kr)$, each mode strength $b_l(kr)$ appears $2l+1$ times on the diagonal. The complex and real *spherical harmonics transforms* can then be defined as follows:

$$\begin{aligned} \mathbf{p}(k) &= \mathbf{B}^{-1}(kr) \int_{\mathcal{S}^2} p(k; r, \Omega) \mathbf{y}^*(\Omega)\, \mathrm{d}\Omega \\ \mathbf{p}^{\mathfrak{R}}(k) &= \mathbf{B}^{-1}(kr) \int_{\mathcal{S}^2} p(k; r, \Omega) \mathbf{r}(\Omega)\, \mathrm{d}\Omega\,. \end{aligned} \tag{6}$$

In practice, the mode-strength compensation $\mathbf{B}^{-1}(kr)$ has to be regularized and the integral has to be approximated by a weighted sum over the microphone directions $\Omega_1, ..., \Omega_P$ of the SMA. Moreover, the sound pressure at these directions has to be replaced with the respective microphone signals $X_1, ..., X_P$. This yields the discrete spherical harmonics transform:

$$\mathbf{x}(k) = \mathbf{B}_{\text{reg}}^{-1}(kr) \sum_{p=1}^{P} \mathfrak{q}_p X_p(k) \mathbf{y}^*(\Omega_p) \tag{7}$$

and analogously for the real spherical harmonics transform, where $\mathbf{B}_{\text{reg}}^{-1}$ denotes the regularized inverse of $\mathbf{B}$ and $\mathfrak{q}_1, ..., \mathfrak{q}_P$ the sampling weights, which depend on the microphone distribution of the SMA [12, 13]. For uniform spatial sampling, one yields $\mathfrak{q}_p = \frac{4\pi}{P}$ [14].

In the following sections, we assume that all signals have been transformed to the short-time Fourier transform (STFT) domain, where time and frequency indices are omitted for brevity.

3. DOA-VECTOR EB-ESPRIT

3.1 Complex DOA-Vector EB-ESPRIT

Let $\mathbf{x} = [X_{00}, X_{1-1}, ..., X_{LL}]$ denote the mode strength compensated complex SHD coefficients of a SMA recording including J plane-wave sources and additive noise, which are assumed to be mutually uncorrelated. The power spectral density (PSD) matrix of $\mathbf{x}$ is defined as $\mathbf{\Phi}_{\mathbf{x}} := E\{\mathbf{x}\mathbf{x}^H\}$, where $(\cdot)^H$ denotes the conjugate transpose and $E\{\cdot\}$ the statistical expectation. It can be shown that the plane-wave propagation vectors are proportional to $\mathbf{y}^*(\Omega_j)$ for $j = 1, ..., J$, where Ω_j denotes the DOA of source j [12]. The signal subspace span$\{\mathbf{y}^*(\Omega_1), ..., \mathbf{y}^*(\Omega_J)\}$ can be estimated from the eigenvectors $[\mathbf{u}_1, ..., \mathbf{u}_J] =: \mathbf{U}_s$ corresponding to the J largest eigenvalues using the following relation [6]:

$$\mathbf{U}_s = [\mathbf{y}^*(\Omega_1), ..., \mathbf{y}^*(\Omega_J)]\, \mathbf{T}\,, \tag{8}$$

where $\mathbf{T}$ is an invertible matrix of size $J \times J$.

EB-ESPRIT uses (8) and recurrence relations of spherical harmonics to estimate the plane-wave DOAs $\Omega_1, ..., \Omega_J$ from the signal subspace eigenmatrix $\mathbf{U}_s$. The DOA-vector EB-ESPRIT [10] estimates the DOA-vectors $\mathbf{n}(\Omega_1), ..., \mathbf{n}(\Omega_J)$, defined as

$$\mathbf{n}(\Omega) = \begin{bmatrix} n_x(\Omega) \\ n_y(\Omega) \\ n_z(\Omega) \end{bmatrix} = \begin{bmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{bmatrix} \tag{9}$$

using the following three recurrence relations:

$$n_a(\Omega)\mathbf{D}_0 \mathbf{y}^*(\Omega) = \mathbf{D}_a \mathbf{y}^*(\Omega) \text{ for } a = x, y, z \tag{10}$$

with

$$\mathbf{D}_x := \tfrac{1}{2}(\mathbf{D}_- + \mathbf{D}_+) \quad , \quad \mathbf{D}_y := \tfrac{1}{2i}(\mathbf{D}_- - \mathbf{D}_+)\,, \tag{11}$$

and

$$\begin{aligned} [\mathbf{D}_z(\cdot)]_{lm} &= \sqrt{(l-m)(l+m)/(2l-1)(2l+1)}\, [(\cdot)]_{(l-1)m} \\ &\quad + \sqrt{(l+1-m)(l+1+m)/(2l+1)(2l+3)}\, [(\cdot)]_{(l+1)m} \\ [\mathbf{D}_\pm(\cdot)]_{lm} &= \pm\sqrt{(l-1\mp m)(l\mp m)/(2l-1)(2l+1)}\, [(\cdot)]_{(l-1)(m\pm 1)} \\ &\quad \mp\sqrt{(l+1\pm m)(l+2\pm m)/(2l+1)(2l+3)}\, [(\cdot)]_{(l+1)(m\pm 1)} \\ \mathbf{D}_0(\cdot) &= [[(\cdot)]_{00}, ..., [(\cdot)]_{(L-1)(L-1)}]^T \end{aligned} \tag{12}$$

for $l = 0, ..., L-1$ and $m = -l, ..., l$, where $[(\cdot)]_{lm} := 0$ if $|m| > l$. Using relation (8), we get from (10)

$$(\mathbf{D}_0 \mathbf{U}_s)\mathbf{\Psi}_a = \mathbf{D}_a \mathbf{U}_s \tag{13}$$

with $\mathbf{\Psi}_a = \mathbf{T}^{-1}\text{diag}\{n_a(\Omega_1), ..., n_a(\Omega_J)\}\mathbf{T}$ for $a = x, y, z$. The matrices $\mathbf{\Psi}_x$, $\mathbf{\Psi}_y$ and $\mathbf{\Psi}_z$ can be estimated in the least-squares sense by:

$$\hat{\mathbf{\Psi}}_a = (\mathbf{D}_0 \mathbf{U}_s)^+ \mathbf{D}_a \mathbf{U}_s\,, \tag{14}$$

where $(\cdot)^+$ denotes the pseudo-inverse. The DOA-vectors $\mathbf{n}(\Omega_1), ..., \mathbf{n}(\Omega_J)$ can then be estimated by jointly diagonalizing $\hat{\mathbf{\Psi}}_x$, $\hat{\mathbf{\Psi}}_y$ and $\hat{\mathbf{\Psi}}_z$ as follows [10]:

$$[\hat{\mathbf{n}}(\Omega_1), ..., \hat{\mathbf{n}}(\Omega_J)] = \text{Re}\left\{[\boldsymbol{\lambda}_x, \boldsymbol{\lambda}_y, \boldsymbol{\lambda}_z]^T\right\}\,, \tag{15}$$

where $\boldsymbol{\lambda}_x$, $\boldsymbol{\lambda}_y$ and $\boldsymbol{\lambda}_z$ denote the eigenvalue vectors of $\hat{\mathbf{\Psi}}_x$, $\hat{\mathbf{\Psi}}_y$ and $\hat{\mathbf{\Psi}}_z$, respectively, derived using a joint diagonalization method.

doi:10.25836/sasp.2019.18

3.2 Real DOA-Vector EB-ESPRIT

Let $\mathbf{x}^{\mathfrak{R}}$ denote the real SHD signal corresponding to $\mathbf{x}$. Using $\mathbf{x}^{\mathfrak{R}} = \mathbf{U}_L^* \mathbf{x}$ and (4), one can show that the plane-wave propagation vectors are proportional to $\mathbf{r}(\Omega_j)$ for $j = 1, ..., J$. Note that, although the propagation vectors are real valued, the signal vectors are still complex valued due to the complex STFT. Analogously to [5], we split $\mathbf{x}^{\mathfrak{R}}$ into real and imaginary parts and construct a real valued PSD matrix thereof:

$$\mathbf{X} := [\mathrm{Re}\{\mathbf{x}^{\mathfrak{R}}\}, \mathrm{Im}\{\mathbf{x}^{\mathfrak{R}}\}] \qquad \mathbf{\Phi}_{\mathbf{X}}^{\mathfrak{R}} := E\{\mathbf{X}\mathbf{X}^H\} \,. \quad (16)$$

One can show that $\mathbf{\Phi}_{\mathbf{X}}^{\mathfrak{R}}$ coincides with the PSD matrix of $\mathbf{x}^{\mathfrak{R}}$, if the noise PSD matrix is real valued. Theoretically, this is the case for diffuse noise and microphone self-noise.

The real signal subspace is constructed from the real eigenvectors $[\mathbf{o}_1, ..., \mathbf{o}_J] =: \mathbf{O}_s$ corresponding to the J largest eigenvalues of $\mathbf{\Phi}_{\mathbf{X}}^{\mathfrak{R}}$, which is related to the plane-wave propagation vectors as follows:

$$\mathbf{O}_s = [\mathbf{r}(\Omega_1), ..., \mathbf{r}(\Omega_J)]\, \mathbf{T}^{\mathfrak{R}} \,, \quad (17)$$

where $\mathbf{T}^{\mathfrak{R}}$ is a real-valued invertible $J \times J$ matrix. Using relation (4), the recurrence relations (10) can be formulated for real spherical harmonics:

$$\begin{aligned} n_a(\Omega)\mathbf{D}_0\mathbf{y}^*(\Omega) &= \mathbf{D}_a\mathbf{y}^*(\Omega) \\ n_a(\Omega)\mathbf{D}_0\mathbf{U}_L^T\mathbf{r}(\Omega) &= \mathbf{D}_a\mathbf{U}_L^T\mathbf{r}(\Omega) \\ n_a(\Omega)\mathbf{U}_{L-1}^T\mathbf{D}_0\mathbf{r}(\Omega) &= \mathbf{D}_a\mathbf{U}_L^T\mathbf{r}(\Omega) \\ n_a(\Omega)\mathbf{D}_0\mathbf{r}(\Omega) &= \mathbf{U}_{L-1}^*\mathbf{D}_a\mathbf{U}_L^T\mathbf{r}(\Omega) =: \mathbf{D}_a^{\mathfrak{R}}\mathbf{r}(\Omega) \,, \end{aligned} \quad (18)$$

where we defined $\mathbf{D}_a^{\mathfrak{R}} := \mathbf{U}_{L-1}^*\mathbf{D}_a\mathbf{U}_L^T$ which must be real valued for $a = x, y, z$. Using (17), the EB-ESPRIT equations become:

$$(\mathbf{D}_0\mathbf{O}_s)\mathbf{\Psi}_a^{\mathfrak{R}} = \mathbf{D}_a^{\mathfrak{R}}\mathbf{O}_s \,. \quad (19)$$

The DOA-vectors can then be estimated analogously to (14)-(15). In contrast to the complex DOA-vector EB-ESPRIT, the matrices involved are real valued reducing the computational complexity.

4. ONLINE IMPLEMENTATION

In principle, one can use the real or complex DOA-vector EB-ESPRIT to estimate the source DOAs per time-frequency bin in an online manner. However, due to the eigendecompostion of the PSD matrix and the joint diagonalization procedure, the computational cost might be too large for many real-time applications. Therefore, we propose various modifications in the following sections.

4.1 Recursive Subspace Tracking

Let n and k denote the time-frame and frequency indices, respectively. The PSD matrix of $\mathbf{x}(n,k)$ or $\mathbf{X}(n,k)$ can be estimated recursively as follows:

$$\begin{aligned} &\hat{\mathbf{\Phi}}_{\mathbf{x}}(n,k) = \beta\hat{\mathbf{\Phi}}_{\mathbf{x}}(n-1,k) + (1-\beta)\mathbf{x}(n,k)\mathbf{x}^H(n,k) \text{ or} \\ &\hat{\mathbf{\Phi}}_{\mathbf{X}}^{\mathfrak{R}}(n,k) = \beta\hat{\mathbf{\Phi}}_{\mathbf{X}}^{\mathfrak{R}}(n-1,k) + (1-\beta)\mathbf{X}(n,k)\mathbf{X}^H(n,k) \,, \end{aligned} \quad (20)$$

Algorithm 1: PASTd for real DOA-vector EB-ESPRIT

```
X = X(n,k);
for j = 1, ..., J do
    z^T = o_j^T(n-1,k) X;
    λ_j(n,k) = β λ_j(n-1,k) + ||z||^2;
    E = X - o_j(n-1,k) z^T;
    o_j(n,k) = o_j(n-1,k) + E z^* / λ_j(n,k);
    X = X - o_j(n,k) z^T;
```

where $\beta \in [0, 1)$ is a forgetting factor. The signal subspace can be constructed by performing a singular value decomposition (SVD) to $\hat{\mathbf{\Phi}}_{\mathbf{x}}(n,k)$ or $\hat{\mathbf{\Phi}}_{\mathbf{X}}^{\mathfrak{R}}(n,k)$ and then selecting the eigenvectors corresponding to the J largest eigenvalues. However, this involves a SVD of a $(L+1)^2 \times (L+1)^2$ matrix per time-frequency bin.

A computationally more efficient method to recursively estimate the signal subspace is the deflated projection approximation subspace tracking (PASTd) algorithm [11]. The PASTd algorithm for real SHD signals is summarized in Alg. 1. Note, that $\mathbf{X}(n,k)$ is a $(L+1)^2 \times 2$ matrix and thus the PASTd from [11] has been adjusted accordingly.

Additionally, we othonormalize the estimated eigenvectors using a QR decomposition which can be implemented with low computational cost using e.g. the modified Gram-Schmidt algorithm [15].

4.2 Simplifications for $J = 1$

For $J = 1$, the signal subspace is one dimensional and the matrices $\hat{\mathbf{\Psi}}_x$, $\hat{\mathbf{\Psi}}_y$ and $\hat{\mathbf{\Psi}}_z$ become scalars $\hat{\Psi}_x$, $\hat{\Psi}_y$ and $\hat{\Psi}_z$. Therefore, no joint diagonalization is necessary. The DOA-vector can be estimated directly via:

$$\begin{aligned} \hat{\mathbf{n}}(\Omega_1) &= \mathrm{Re}\{[\hat{\Psi}_x, \hat{\Psi}_y, \hat{\Psi}_z]^T\} \\ &= \|\mathbf{D}_0\mathbf{u}_1\|^{-2}\mathrm{Re}\left\{[\mathbf{D}_x\mathbf{u}_1, \mathbf{D}_y\mathbf{u}_1, \mathbf{D}_z\mathbf{u}_1]^T\right\} \,, \end{aligned} \quad (21)$$

where $\mathbf{u}_1$ is the dominant eigenvector of $\mathbf{\Phi}_{\mathbf{x}}$. The factor $\|\mathbf{D}_0\mathbf{u}_1\|^{-2}$ can be replaced by a normalization to ensure $\|\hat{\mathbf{n}}(\Omega_1)\| = 1$. These simplifications can be made for the real DOA-vector EB-ESPRIT analogously.

4.3 Subspace Propagation-Vector Matching

In [4, 5], general properties of plane-wave propagation vectors are employed to estimate two DOAs per time-frequency bin from a B-format signal. In this section, we develop a similar method to estimate the mixing matrix $\mathbf{C} := (\mathbf{T}^{\mathfrak{R}})^{-1}$ for the real DOA-vector EB-ESPRIT with two sources ($J = 2$). From (17) we get:

$$[\mathbf{r}(\Omega_1), \mathbf{r}(\Omega_2)] = \mathbf{O}_s\mathbf{C} \,. \quad (22)$$

We use the following properties of real spherical harmonics:

$$R_{00}(\Omega) = \frac{1}{\sqrt{4\pi}} \qquad \|\mathbf{r}_1(\Omega)\|^2 = \frac{1}{\pi} \,, \quad (23)$$

where $\mathbf{r}_1(\Omega)$ is the vector of real spherical harmonics up to order 1. Note, that the conditions (23) are sufficient to ensure that $\mathbf{r}_1(\Omega)$ is a real spherical harmonic vector. To ensure that $\mathbf{r}(\Omega)$ describes a real spherical harmonic vector,

we would need more conditions. Let us denote with $\mathbf{O}_s^{(1)}$ the coefficients of $\mathbf{O}_s$ up to first order and with $\mathbf{Q}^{(1)}\mathbf{R}^{(1)}$ the QR decomposition thereof. Using (22), we find

$$[\mathbf{r}_1(\Omega_1), \mathbf{r}_1(\Omega_2)] = \mathbf{Q}^{(1)}\tilde{\mathbf{C}} = \mathbf{Q}^{(1)} \begin{bmatrix} \tilde{c}_{11} & \tilde{c}_{12} \\ \tilde{c}_{21} & \tilde{c}_{22} \end{bmatrix} , \quad (24)$$

where $\tilde{\mathbf{C}} = \mathbf{R}^{(1)}\mathbf{C}$. Inserting (24) into the conditions (23) and using the orthonormal property of $\mathbf{Q}^{(1)}$ yields:

$$\mathbf{q}^T\tilde{\mathbf{c}}_j = \frac{1}{\sqrt{4\pi}} \qquad \text{and} \qquad \|\tilde{\mathbf{c}}_j\|^2 = \frac{1}{\pi} \quad (25)$$

for $j = 1, 2$, where we defined $\mathbf{q}^T := [Q^{(1)}_{00,1}, Q^{(1)}_{00,2}]$ and $\tilde{\mathbf{c}}_j := [\tilde{c}_{1j}, \tilde{c}_{2j}]^T$. Using the second condition, one can write $\tilde{\mathbf{c}}_j$ in the following form:

$$\tilde{\mathbf{c}}_j = \tfrac{1}{\sqrt{\pi}} \left[\cos(\varphi_j), \sin(\varphi_j)\right]^T . \quad (26)$$

Writing $\mathbf{q}$ in terms of magnitude q and phase φ_q:

$$\mathbf{q} = q \left[\cos(\varphi_q), \sin(\varphi_q)\right]^T \quad (27)$$

and inserting (26) into the first condition of (25) one can derive

$$\frac{1}{\sqrt{4\pi}} = \mathbf{q}^T\tilde{\mathbf{c}}_j = \frac{q}{\sqrt{\pi}}\cos(\varphi_q - \varphi_j)$$
$$\Rightarrow \varphi_j = \varphi_q \mp \operatorname{acos}\left(\frac{1}{2q}\right) . \quad (28)$$

We, therefore, get two solutions which can be assigned to φ_1 and φ_2. The results are real valued if $q \geq \frac{1}{2}$. Otherwise, we estimate one DOA only from the dominant eigenvector $\mathbf{o}_1$. The mixing matrix $\mathbf{C}$ can be obtained from φ_1 and φ_2 as follows:

$$\mathbf{C} = (\mathbf{R}^{(1)})^{-1}\tilde{\mathbf{C}} = \frac{1}{\sqrt{\pi}}(\mathbf{R}^{(1)})^{-1} \begin{bmatrix} \cos(\varphi_1) & \cos(\varphi_2) \\ \sin(\varphi_1) & \sin(\varphi_2) \end{bmatrix} . \quad (29)$$

The full higher-order plane-wave propagation vectors $\mathbf{r}(\Omega_1)$ and $\mathbf{r}(\Omega_2)$ can then be estimated by applying $\mathbf{C}$ to the full signal subspace $\mathbf{O}_s$, i.e.,

$$[\hat{\mathbf{r}}(\Omega_1), \hat{\mathbf{r}}(\Omega_2)] = \mathbf{O}_s\mathbf{C} . \quad (30)$$

Finally, the simplified DOA-vector EB-ESPRIT equations discussed in Sec. 4.2 can be used to estimate the DOA-vectors from $\hat{\mathbf{r}}(\Omega_1)$ and $\hat{\mathbf{r}}(\Omega_2)$ separately. Hence, the joint diagonalization is avoided by estimating the source propagation vectors before the EB-ESPRIT equations are used.

As $\mathbf{C}$ is estimated using zero- and first-order coefficients of $\mathbf{O}_s$ only, $\mathbf{O}_s\mathbf{C}$ is not necessarily close to a set of plane-wave propagation vectors. Therefore, we perform a consistency check between the DOA-vectors $\hat{\mathbf{n}}(\Omega_j)$ estimated with the EB-ESPRIT equations and $\hat{\mathbf{n}}_{\text{fo}}(\Omega_j)$ derived from the first order coefficients as follows:

$$\hat{\mathbf{n}}_{\text{fo}}(\Omega_j) := \sqrt{\tfrac{4\pi}{3}}[\hat{R}_{11}(\Omega_j), \hat{R}_{1-1}(\Omega_j), \hat{R}_{10}(\Omega_j)]^T \quad (31)$$

for $J = 1, 2$. If their angular distance is $\geq \Delta\varphi$, we use $\hat{\mathbf{n}}_{\text{fo}}(\Omega)$ instead of $\hat{\mathbf{n}}(\Omega_j)$ for the DOA-vector estimate. The proposed method can be summarized as follows:

1. Update real signal subspace matrix (Sec. 4.1)
2. Estimate plane-wave propagation vectors using subspace propagation-vector matching (Sec. 4.3)
3. Estimate DOA-vectors with simplified $J = 1$ real DOA-vector EB-ESPRIT (Sec. 4.2)
4. Consistency check of DOA-vectors with first order coefficients of estimated propagation vectors (Sec. 4.3, last paragraph)

5. EVALUATION

5.1 Setup

For the evaluation, third order ($L = 3$) spherical harmonic domain signals with one or two plane-wave sources, reverberation and diffuse stationary noise were simulated. The plane-wave source signals consisted of male and female English speech signals of 3.8 seconds length and sampled at 16 kHz, taken from [16]. The source DOAs were randomly and uniquely selected from a set of 48 uniformly distributed directions. For the dual source scenario, the 48 directions were divided into two sectors, from which the DOAs are selected.

For the non-reverberant scenarios, the plane-wave sources were transformed to the STFT domain. Each time-frequency bin was then multiplied with the plane-wave propagation vector $\mathbf{y}^*(\Omega)$ at the corresponding DOA Ω. For the reverberant scenarios, microphone signals of a rigid spherical microphone array with 32 microphones and 7 cm radius placed at $[4.103\text{ m}, 3.471\text{ m}, 2.912\text{ m}]$ in a $8 \times 7 \times 6$ m^3 shoebox room were simulated using [17]. The sources were placed at a distance of 2 m from the virtual microphone array. Reverberation times (T_{60}) of 0.3 and 0.6 seconds were used. The microphone signals were then transformed to the STFT domain. Finally, the SHD coefficients are derived using the discrete spherical harmonics transform (7) with uniform sampling weights and $\mathbf{B}_{\text{reg}}^{-1} = (\mathbf{B}^H\mathbf{B} + \lambda\mathbf{I})^{-1}\mathbf{B}^H$ with $\lambda = 10^{-6}$.

For all cases, diffuse stationary white noise with signal plus reverberation-to-noise ratio SNR $= 6$ dB was added. The desired source variance, which is needed to determine the variance of the noise, was computed as the mean energy of the noiseless signal, excluding time-frames with energies less than 1% of the maximum frame energy.

For the STFTs, a frame-length of 128 samples (8 ms), 50% overlap, a square-root-Hann window and a discrete Fourier transform size of 256 was chosen.

For the recursive signal-subspace estimation, $\beta = 0.9$ ($\hat{=}$ 38 ms time-constant) was chosen. An angular distance $\Delta\varphi = 0.4\pi$ was chosen for the consistency check of the subspace propagation-vector matching. The DOAs were estimated within the frequency range $[100, 2340]$ Hz, were the lower bound has been chosen to reduce the effect of the regularized mode-strength compensation and the upper bound to avoid spatial aliasing.

For the performance evaluation we computed angular

doi:10.25836/sasp.2019.18

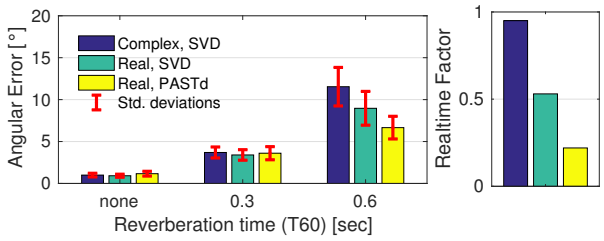

Figure 1. Mean angular estimation errors and mean realtime factors for single source scenario ($J = 1$)

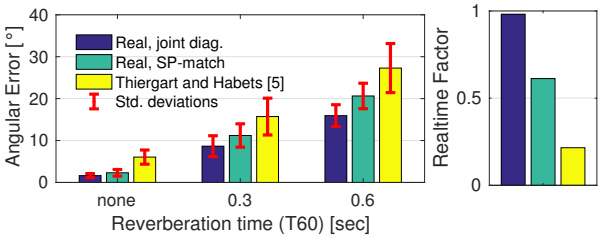

Figure 2. Mean angular estimation errors and mean realtime factors for dual sources scenario ($J = 2$)

estimation errors

$$\Delta\Omega_j(n,k) = \min_{j'} \left\{ \text{acos}\left(\mathbf{n}(\Omega_j)^T \mathbf{n}(\hat{\Omega}_{j'}(n,k)) \right) \right\} , \tag{32}$$

where Ω_j is the true DOA of source j and $\hat{\Omega}_{j'}(n,k)$ the j'-th estimated DOA at time-frequency bin (n,k). The mean angular estimation errors within the regions where the sources are active are defined as

$$\overline{\Delta\Omega_j} = \sum_{n,k} w_j(n,k)\Delta\Omega_j(n,k) , \tag{33}$$

where $w_j(n,k)$ is one if the narrowband frame energy of source j, including reverberation, at bin (n,k) is greater than -30 dB w.r.t. the maximum energy and zero otherwise. To evaluate the computational complexity we computed realtime factors $= \frac{\text{Computation time}}{\text{Signal length}}$. The DOA estimators were implemented with MATLAB [18] in double precision and executed on a computer with a 3.40 GHz CPU.

5.2 Single source scenario

For the single source scenario, 20 SHD signals with one plane-wave source were generated. The source signals consisted of 10 female and 10 male English speech signals.

In Fig. 1, the mean angular estimation errors and realtime factors for the complex DOA-vector EB-ESPRIT with SVD-based subspace estimation and the real DOA-vector EB-ESPRIT with SVD-based subspace estimation or PASTd are shown for different reverberation times. The angular errors have been averaged over the 20 experiments and the corresponding standard deviations are represented with red errorbars.

One can see that the real-valued formulation and the PASTd reduce the computational cost by $\sim 75\%$ without a significant loss of DOA estimation accuracy.

For $T_{60} = 0.6$ seconds, the proposed real DOA-vector EB-ESPRIT yields less estimation errors than the complex DOA-vector EB-ESPRIT. Recall that $\mathbf{\Phi}_{\mathbf{X}}^{\mathfrak{R}} = \text{Re}\{\mathbf{\Phi}_{\mathbf{x}^{\mathfrak{R}}}\}$. The PSD matrices of the direct-path plane-wave sources and the diffuse noise are real-valued in the real SHD. Only the early and late reflections may contribute to $\text{Im}\{\mathbf{\Phi}_{\mathbf{x}^{\mathfrak{R}}}\}$. Therefore, it is plausible that it is more robust to use $\mathbf{\Phi}_{\mathbf{X}}^{\mathfrak{R}}$ instead of $\mathbf{\Phi}_{\mathbf{x}^{\mathfrak{R}}}$ for the subspace estimation under reverberant conditions. Using the PASTd instead of the SVD further improves the estimation accuracy for $T_{60} = 0.6$ seconds, which can be explained by the fact that PASTd can be more robust than the SVD for low signal to noise/reverberation ratios [11].

5.3 Dual sources scenario

For the dual sources scenario, the real DOA vector EB-ESPRIT with PASTd and joint diagonalization (joint diag.) or subspace propagation-vector matching (SP-match) were compared with a robust B-format DOA estimator [5]. Ten different SHD signals with one female and one male plane-wave source were generated.

The mean angular errors and realtime factors are shown in Fig. 1. The results are averaged over the 10 configurations and the two DOAs. The respective standard deviations are represented with red errorbars. The EB-ESPRIT-based methods yield lower estimation errors compared to the robust B-format method, which is expected as the latter does not incorporate higher-order SHD coefficients. The real DOA-vector EB-ESPRIT with SP-matching is less accurate than the joint diagonalization based method. However, the computational cost is reduced by $\sim 40\%$.

So far only mean angular estimation errors have been analysed. In what follows, we analyse a dual source scenario with DOAs $\Omega_1 = [90°, 60°]$, $\Omega_2 = [130°, -80°]$ and $T_{60} = 0.3$ seconds in more detail.

In Fig. 3, the spectrogram of the source signals (a), the source activity per time-frequency bin (b) and angular estimation errors for both source DOAs (c-h) are shown. One can see that the EB-ESPRIT-based methods yield less estimation errors than the robust B-format method, except at time-frequency bins where the number of active sources changes from one to two are two to one. For the robust B-format method, these regions are less critical, however, the overall angular estimation errors are larger.

In Fig. 4, distributions of the estimated azimuth and elevation angles are shown for the three methods. One can see that, for the EB-ESPRIT-based methods, the estimated azimuth and elevation are mostly concentrated around the true DOAs ($\Omega_1 = [90°, 60°]$, $\Omega_2 = [130°, -80°]$), while for the robust B-format method, the estimates are more scattered across the angular space.

6. CONCLUSION

We proposed the real DOA-vector EB-ESPRIT which reduces the computational complexity of the DOA-vector EB-ESPRIT [10] by working with real-valued quantities and by efficiently estimating the signal subspace using the PASTd algorithm [11]. To futher reduce the computational complexity, we replaced the joint diagonalization with a

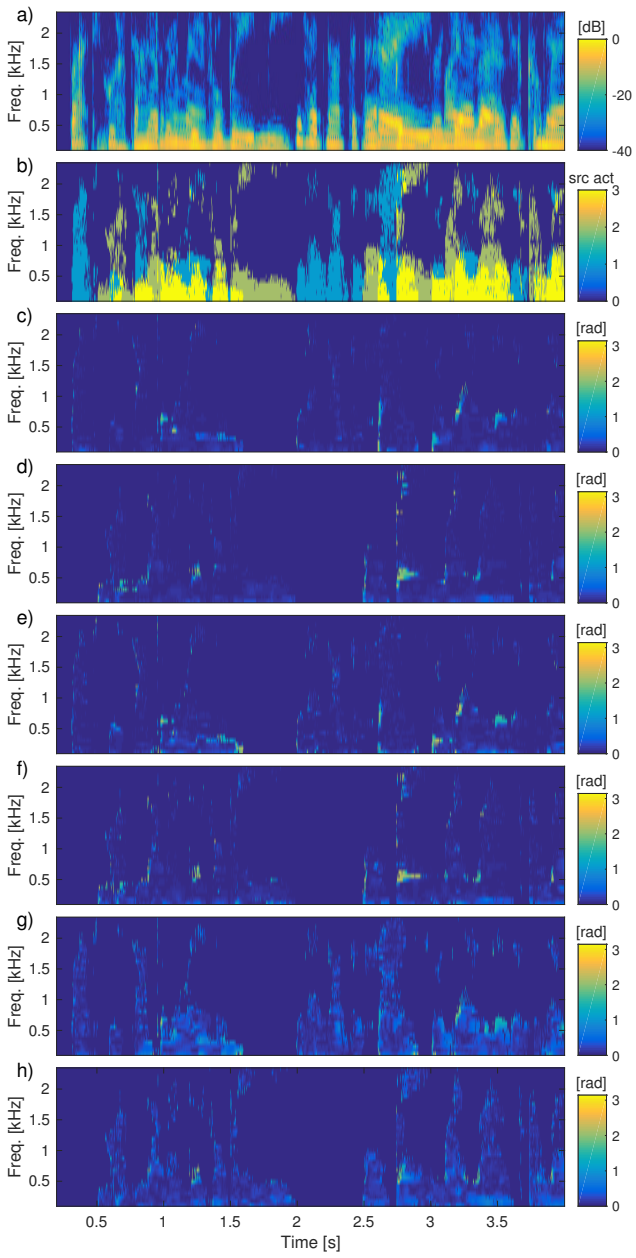

Figure 3. a) Spectrogram of source signals (sum), b) active time-frequency bins with 0: no source active, 1: source 1 active, 2: source 2 active, 3: both sources active, c) - h): Angular estimation errors at active bins for Ω_1 and Ω_2, c) and d) Real, joint diag., e) and f) Real, SP-match, g) and h) Thiergart and Habets [5]

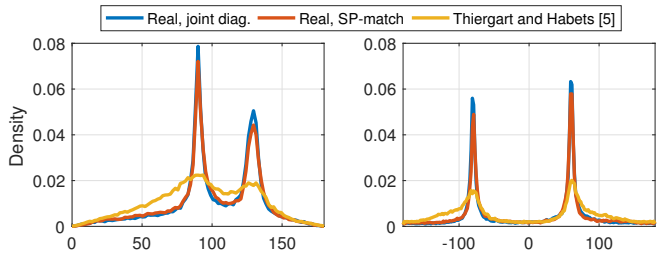

Figure 4. Distributions of elevation and azimuth estimates

subspace propagation-vector matching method for estimating two DOAs. In the evaluation, we showed that the computational cost of the proposed method is significantly reduced compared to the complex DOA-vector EB-ESPRIT and that the method can estimate the source DOAs more accuratly compared to [5].

7. REFERENCES

[1] V. Pulkki, "Spatial sound reproduction with directional audio coding," *Journal Audio Eng. Soc.*, vol. 55, pp. 503–516, June 2007.

[2] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, "3D source localization in the spherical harmonic domain using a pseudointensity vector," in *Proc. European Signal Processing Conf. (EUSIPCO)*, (Aalborg, Denmark), pp. 442–446, Aug. 2010.

[3] A. Herzog and E. A. P. Habets, "On the relation between DOA-vector eigenenbeam ESPRIT and subspace-pseudointensity-vector," in *Proc. European Signal Processing Conf. (EUSIPCO)*, (A Coruña, Spain), Sept. 2019.

[4] S. Berge and N. Barrett, "High angular resolution planewave expansion," in *2nd Intl. Symp. on Ambisonics and Spherical Acoustics*, (Paris, France), May 2010.

[5] O. Thiergart and E. A. P. Habets, "Robust direction-of-arrival estimation of two simultaneous plane waves from a B-format signal," in *Proc. IEEE Convention of Electrical & Electronics Engineers in Israel (IEEEI)*, Nov. 2012.

[6] H. Teutsch and W. Kellermann, "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5276–5279, Mar. 2008.

[7] B. Jo and J. W. Choi, "Direction of arrival estimation using nonsingular spherical ESPRIT," *J. Acoust. Soc. Am.*, vol. 143, pp. EL181–EL187, Mar. 2018.

[8] Q. Huang, L. Zhang, and Y. Fang, "Two-step spherical harmonics ESPRIT-type algorithms and performance analysis," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 26, Sept. 2018.

[9] B. Jo and J. W. Choi, "Nonsingular EB-ESPRIT for the localization of early reflections in a room," *J. Acoust. Soc. Am.*, vol. 144, Sept. 2018.

[10] A. Herzog and E. A. P. Habets, "Eigenbeam-ESPRIT for DOA-vector estimation," *IEEE Signal Process. Lett.*, 2019.

[11] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, pp. 95–107, Jan. 1995.

[12] B. Rafaely, *Fundamentals of Spherical Array Processing*, vol. 8. Springer, 2015.

[13] D. P. Jarrett, E. A. P. Habets, and P. A. Naylor, *Theory and Applications of Spherical Microphone Array Processing*. Springer, 2017.

[14] J. Meyer and G. W. Elko, "Spherical microphone arrays for 3d sound recordings," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems* (Y. Huang and J. Benesty, eds.), pp. 67–89, Norwell, MA, USA: Kluwer Academic Publishers, 2004.

[15] A. Björck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA: SIAM, 1996.

[16] V. Panayotov, G. Chen, D. Povey, and S. Khudanpur, "Librispeech: An ASR corpus based on public domain audio books," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, (South Brisbane, QLD), pp. 5206–5210, Apr. 2015.

[17] D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, and P. A. Naylor, "Rigid sphere room impulse response simulation: algorithm and applications," *J. Acoust. Soc. Am.*, vol. 132, pp. 1462–1472, Sept. 2012.

[18] *MATLAB 2016b*. Natrick, Massachusetts, US: The MathWorks, Inc.

doi:10.25836/sasp.2019.18