Audio Engineering Society

Convention Express Paper 100

Presented at the 154th Convention
2023 May 13–15, Espoo, Helsinki, Finland

Detecting simultaneous directions of arrival in an Ambisonic signal with REVEB-ESPRIT

Franz Zotter[1] and Thomas Deppisch[2]

[1]*Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria*
[2]*Division of Applied Acoustics, Chalmers University of Technology, Gothenburg, Sweden*

Correspondence should be addressed to Franz Zotter (zotter@iem.at)

ABSTRACT

We present a purely real-valued variant of the extended vector-based EB-ESPRIT (REVEB-ESPRIT), an algorithm that estimates multiple simultaneous directions of arrival (DOAs) from Ambisonic signals, which are either encoded mono sounds or captured via a spherical microphone array. Our proposal uses fully real-valued spherical harmonics and DOA vectors and presents the required extended set of recurrence relations. Moreover, we propose a real-valued joint Schur decomposition using inverse iterations to efficiently solve the simultaneous diagonalization problem that is inherent in EB-ESPRIT algorithms. We evaluate the proposed algorithm in free-field conditions with a varying number of simultaneously estimated DOAs and varying signal-to-noise ratios. Our analysis shows a slight increase in speed and accuracy due to the proposed real-valued formalism, and in particular a noticeable increase in speed and accuracy when detecting many simultaneous DOAs. A reference implementation of the proposed algorithm is provided online.

1 Introduction

Works by Teutsch and Kellermann [1, 2] introduced EB-ESPRIT, an algorithm that facilitates the simultaneous estimation of multiple directions of arrival (DOAs) from Ambisonic signals, i.e., coefficient signals of the spherical harmonics (SH). In a nutshell, EB-ESPRIT integrates the SH-domain signal covariance matrix into SH recurrence relations and, by exploiting the linear relationship between the covariance matrix and the signal subspace, solves for the desired DOAs using a suitable matrix factorization. Despite its mathematical elegance, the method suffers from numerical instability for source directions in the horizontal plane.

Several methods based on other recurrence relations were introduced to circumvent the numerical instability, e.g. [3, 4], but they either exhibit ambiguities or need an additional matching of parameter pairs. More recently, the vector-based EB-ESPRIT (VEB-ESPRIT) method was developed independently by Jo and Choi [5] and Herzog and Habets [6]. It solves the deficiencies of the earlier algorithms by employing three recurrence relations. As a consequence, the algorithm relies on a simultaneous diagonalization of three matrices to reveal the DOAs as their joint eigenvalues.

In [7], the vector-based EB-ESPRIT was extended to achieve a simultaneous estimation of up to $N^2 + \lfloor 4N/3 \rfloor$ instead of N^2 DOAs, where N is the maximum order of the SH decomposition. While this extended vector-based EB-ESPRIT (EVEB-ESPRIT) is able to deliver the largest number of simultaneously detectable DOAs, it is designed using complex-valued SHs and complex-valued entries for the DOA vectors. In [8, 9, 10], variants of the VEB-ESPRIT and its extension were proposed that aim for higher computational efficiency by employing real-valued computations. The semi-real-valued algorithm from [8] uses real-valued DOA parameters and real-valued covariance matrices but employs complex-valued SHs and steering matrices. In [9] and [10], a real-valued VEB-ESPRIT and EVEB-ESPRIT are defined using conversion matrices to translate complex-valued recurrence relations and complex-valued SHs to a real-valued formulation. Multiarray EB-ESPRIT (MEB-ESPRIT) [11] introduced a fully real-valued, non-extended formalism and the joint generalized Schur decomposition to solve coupled DOA parameter detection with multiple distributed arrays.

So far, the simultaneous diagonalization was typically solved using an ad-hoc method that diagonalizes the three problems individually and then applies the eigenvectors of a single problem to estimate the eigenvalues of the joint problem. This ad-hoc method is computationally efficient and was found to perform similarly to the iterative simultaneous diagonalization algorithms from [12] in free-field scenarios [6, 7]. JAPAM (joint eigenvalue decomposition algorithms using a parametrized matrix) [13] is a class of block coordinate algorithms that is defined using different kinds of 2×2 matrix factorizations. Amongst others, JAPAM-5 was shown to outperform the algorithms from [12] while being more computationally efficient than the algorithms based on the polar decomposition [13]. In the context of the vector-based EB-ESPRIT, it was shown that JAPAM-5 outperforms the ad-hoc method in a test case including a room simulation [8].
The joint Schur decomposition (JSD) was used in [14, 15] to solve the joint eigenvalue problem, however with 2×2 iterations that tend to converge slowly. In [16], the joint generalized Schur decomposition uses inverse iterations that converge quickly.

This paper describes a simplified, fully real-valued extended vector-based EB-ESPRIT algorithm for the single-array setting to detect what is currently known as the highest possible number of simultaneous DOAs in EB-ESPRIT [7]. The maximum number of estimated DOAs is reached via the extension of the SH recurrences as in EVEB-ESPRIT, but is provided here for fully real-valued SHs. In comparison to the MEB-ESPRIT framework, the simplifications concern the single-array case and the specialization of the joint generalized Schur decomposition to the joint Schur decomposition (JSD). We compare the proposed REVEB-ESPRIT algorithm using JSD diagonalization to real- and complex-valued formulations of EVEB-ESPRIT using the ad-hoc diagonalization method, and to the real-valued formulation using diagonalization via JAPAM-5. Our results from anechoic simulations with different numbers of simultaneous DOAs and varying signal-to-noise ratios (SNRs) suggest a high performance that is better or on par with the best of the compared algorithms while being either more computationally efficient than or comparable to the fastest algorithm.

1.1 Ambisonics

Ambisonics [17] is a scene-based multi-channel audio format consisting of spherical harmonic (SH) coefficient signals $\chi_n^m(t)$. Their expansion over the spherical harmonics (SHs) $Y_n^m(\boldsymbol{\theta})$ delivers a continuous surround-with-height signal depending on the DOA unit-vector $\boldsymbol{\theta} = [x, y, z]^\mathrm{T}$ that can be evaluated at any direction $\boldsymbol{\theta}$,

$$s(\boldsymbol{\theta}, t) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \chi_n^m(t) Y_n^m(\boldsymbol{\theta}). \tag{1}$$

The SHs $Y_n^m(\boldsymbol{\theta})$ of order n and degree m are a sequence of orthogonal spherical polynomials of the degree n and in their typical formulation they contain trigonometric functions of the degree m in the azimuth $\varphi = \arctan \frac{y}{x}$ and the associated Legendre functions P_n^m of the cosine of the zenith angle $\vartheta = \arctan \frac{\sqrt{x^2+y^2}}{z}$. Here, we use a real-valued definition of the SHs [18]

$$Y_n^m(\boldsymbol{\theta}) = N_n^{|m|} P_n^{|m|}(\cos\vartheta) \begin{cases} \cos(m\varphi) & \text{for } m \geq 0 \\ \sin(|m|\varphi) & \text{for } m < 0 \end{cases}, \tag{2}$$

using the N3D normalization factor $N_n^{|m|} = (-1)^m \sqrt{\frac{(2n+1)(2-\delta_m)}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}}$ that involves the binary Kronecker Delta: $\delta_m = 1$ for $m = 0$, and 0 elsewhere.

1.1.1 Encoded directional and diffuse signals

Ambisonic signals can be composed of a mixture of directional and diffuse signals. The vector $\boldsymbol{\chi}(t) = [\chi_0^0(t), \chi_1^{-1}(t), \ldots, \chi_{\mathrm{N}}^{\mathrm{N}}(t)]^{\mathrm{T}}$ of its coefficient signals is obtained as

$$\boldsymbol{\chi}(t) = \boldsymbol{Y}\boldsymbol{s}(t) + \boldsymbol{\nu}(t), \tag{3}$$

where the matrix $\boldsymbol{Y}$ encodes a vector of corresponding J directional signals $\boldsymbol{s}(t)$. The columns of $\boldsymbol{Y} = [\boldsymbol{y}_{\mathrm{SH}}(\boldsymbol{\theta}_1), \ldots, \boldsymbol{y}_{\mathrm{SH}}(\boldsymbol{\theta}_J)]$ contain encoding vectors $\boldsymbol{y}_{\mathrm{SH}}(\boldsymbol{\theta}) = [Y_0^0(\boldsymbol{\theta}), Y_1^{-1}(\boldsymbol{\theta}), \ldots, Y_{\mathrm{N}}^{\mathrm{N}}(\boldsymbol{\theta})]^{\mathrm{T}}$ of vertically stacked $(N+1)^2$ SHs up to the highest order $n \le \mathrm{N}$. Additive noise and diffuse sound field content $\boldsymbol{\nu}(t)$ is directly described as a set of incoherent or uncorrelated Ambisonic signals.

1.1.2 Encoded spherical array signals

A spherical microphone array (SMA) records sounds from multiple directions with its microphones that are arranged on the surface of a sphere. The recorded signal $p(\boldsymbol{\theta}_i, t)$ of a microphone at position $\boldsymbol{\theta}_i$ can be encoded into Ambisonic signals $\chi_n^m(t)$ by inverting the model equation

$$p(\boldsymbol{\theta}_i, t) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} [\chi_n^m(t) \star b_n(t)]\, Y_n^m(\boldsymbol{\theta}_i). \tag{4}$$

SMA theory to estimate $\chi_n^m(t)$ is well covered, e.g. in [19, 20, 21, 17]. The model includes the sampled spherical harmonics $Y_n^m(\boldsymbol{\theta}_i)$, which are invertible as a matrix for a finite $n \le N$, and the sensitivities $b_n(t)$ in the frequency domain $b_n(\omega) = \mathrm{i}^n[(ka)^2 h_n'(ka)]^{-1}$, which are invertible filters when accepting realistic white noise gain limits at low frequencies and the spatial aliasing limits at high frequencies. It is useful to focus on this operational frequency range, here.

1.2 Signal subspace of observed signals

If only the signals $\boldsymbol{\chi}(t)$ are known, be it from the encoded content of an SMA recording or a directional mix of mono signals, DOAs are only contained implicitly. Considering an observed block of B time samples

$$\boldsymbol{X} = [\boldsymbol{\chi}_0, \ldots, \boldsymbol{\chi}_{B-1}], \tag{5}$$

we estimate the sample covariance as $\boldsymbol{R} = 1/B\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$. We denote the eigenvectors of $\boldsymbol{R}$ corresponding to the largest J eigenvalues α_j in $\boldsymbol{\alpha}$ as signal subspace eigenvectors and collect them in the columns of $\boldsymbol{U}$,

$$\boldsymbol{R}_\chi = [\boldsymbol{U}, \boldsymbol{U}_0]\, \mathrm{diag}\{\boldsymbol{\alpha}\}\, [\boldsymbol{U}, \boldsymbol{U}_0]^{\mathrm{T}}. \tag{6}$$

While encoding vectors $\boldsymbol{Y}$ in (3) cannot be estimated directly from the signals $\boldsymbol{\chi}(t)$, the signal subspace $\boldsymbol{U}$ and the steering vectors $\boldsymbol{Y}$ span the same subspace $S_X = \mathrm{span}(\boldsymbol{Y}) = \mathrm{span}(\boldsymbol{U})$. Accordingly, there must be a non-singular matrix $\boldsymbol{T} \in \mathbb{R}^{J \times J}$ relating both $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{T}$. This is a hint towards $\boldsymbol{Y}$, with no DOA vectors yet, that is exploited by all EB-ESPRIT algorithms.

2 Extended REVEB-ESPRIT

The challenge mastered by EB-ESPRIT is to retrieve the multiple DOAs $\boldsymbol{\theta}_j$ from a short-term observation of the sigfnal space $\boldsymbol{U}$. We propose a real-valued, extended, vector-based EB-ESPRIT (REVEB-ESPRIT).

2.1 DOA parameters as diagonal, linear factors

Vector-based EB-ESPRIT involves re-expanding the SHs $\boldsymbol{Y}$ multiplied by entries of the DOA vector $\boldsymbol{\theta}_j = [x_j, y_j, z_j]^{\mathrm{T}}$, denoted as $\boldsymbol{x} = [x_j]$, $\boldsymbol{y} = [y_j]$, $\boldsymbol{z} = [z_j]$, as diagonal factors, $\boldsymbol{Y}\mathrm{diag}\{\boldsymbol{x}\}$, $\boldsymbol{Y}\mathrm{diag}\{\boldsymbol{y}\}$, $\boldsymbol{Y}\mathrm{diag}\{\boldsymbol{z}\}$. Multiplying any SH $Y_n^m(\boldsymbol{\theta})$ by x, y, or z increases the degree of the resulting spherical polynomial to $n+1$. As with other families of orthogonal degree-n polynomials, re-expansion of any SH multiplied with $z = \cos\vartheta$ must only involve SHs with the indices $n \pm 1$, due to the associated Legendre functions involved. Multiplication with x or y affects the azimuth harmonics by trigonometric addition theorems and therefore additionally involves the shift $m \pm 1$, yielding altogether

$$Y_n^m x = \tfrac{(1-\delta_m)\,\mathrm{sgn}_{-m}}{\sigma_m \sigma_{m-1}} [-w_n^m Y_{n-1}^{m-1} + w_{n+1}^{-m+1} Y_{n+1}^{m-1}] \tag{7}$$
$$- \tfrac{(1-\delta_{m+1})\,\mathrm{sgn}_m}{\sigma_m \sigma_{m+1}} [w_n^{-m} Y_{n-1}^{m+1} - w_{n+1}^{m+1} Y_{n+1}^{m+1}],$$
$$Y_n^{-m} y = \tfrac{\mathrm{sgn}_{-m}}{\sigma_m \sigma_{m-1}} [-w_n^m Y_{n-1}^{m-1} + w_{n+1}^{-m+1} Y_{n+1}^{m-1}] \tag{8}$$
$$+ \tfrac{(1-\delta_m)(1-\delta_{m+1})\,\mathrm{sgn}_m}{\sigma_m \sigma_{m+1}} [w_n^{-m} Y_{n-1}^{m+1} - w_{n+1}^{m+1} Y_{n+1}^{m+1}],$$
$$Y_n^m z = v_n^m Y_{n-1}^m + v_{n+1}^m Y_{n+1}^m. \tag{9}$$

The lengthy derivation of these relations is made available online, together with a MATLAB script for numerical validation[1]. The coefficients are equivalent to

[1] https://git.iem.at/thomasdeppisch/real-sh-recurrence-relations

those in [22, 11], and involve a non-zero sign function sgn_m and re-normalization σ_m:

$$w_n^m = \sqrt{\tfrac{(n+m-1)(n+m)}{(2n-1)(2n+1)}}, \qquad v_n^m = \sqrt{\tfrac{(n-m)(n+m)}{(2n-1)(2n+1)}},$$
$$\mathrm{sgn}_m = \begin{cases} 1, & m \geq 0, \\ -1, & \text{else}, \end{cases} \qquad \sigma_m = \sqrt{2-\delta_m}.$$

Similar as with complex-valued SHs in [7], we extend the $3N^2$ relations for the orders $n \leq N-1$ by further $4N$ equations for $n = N$, after eliminating the $Y_{n+1}^{m\pm1}$ terms contained within, as a contribution of this paper. The resulting extended recurrences are

$$\tfrac{h_n^m \mathrm{sgn}_{-m}}{\sigma_m}[(1-\delta_m)Y_n^m x + Y_n^{-m} y] - \tfrac{h_n^{-m+1}}{\sigma_{m-1}} Y_n^{m-1} z = -\tfrac{h_{n-1}^m}{\sigma_{m-1}} Y_{n-1}^{m-1}, \tag{10}$$

$$\tfrac{h_n^{-m} \mathrm{sgn}_m}{\sigma_m}[Y_n^m x - (1-\delta_m)Y_n^{-m} y] - \tfrac{(1-\delta_{m+1})h_n^{m+1}}{\sigma_{m+1}} Y_n^{m+1} z = -\tfrac{(1-\delta_{m+1})h_{n-1}^{-m}}{\sigma_{m+1}} Y_{n-1}^{m+1}, \tag{11}$$

and use the coefficient $h_n^m = \sqrt{\frac{n-m}{2n+1}}$. Similar as in [7], the recurrences (7)–(9), (10)–(11) relate the steering vectors $\boldsymbol{Y}$ to the desired DOA parameters as diagonal matrices $\boldsymbol{Y}\mathrm{diag}\{\boldsymbol{x}\}, \boldsymbol{Y}\mathrm{diag}\{\boldsymbol{y}\}, \boldsymbol{Y}\mathrm{diag}\{\boldsymbol{z}\}$,

$$\begin{bmatrix} \boldsymbol{M} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{M} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{M} \\ \boldsymbol{A} & \boldsymbol{B} & \boldsymbol{C} \end{bmatrix} \begin{bmatrix} \boldsymbol{Y} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{Y} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Y} \end{bmatrix} \begin{bmatrix} \mathrm{diag}\{\boldsymbol{x}\} \\ \mathrm{diag}\{\boldsymbol{y}\} \\ \mathrm{diag}\{\boldsymbol{z}\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{M}_x \\ \boldsymbol{M}_y \\ \boldsymbol{M}_z \\ \boldsymbol{D} \end{bmatrix} \boldsymbol{Y}. \tag{12}$$

The binary mask $\boldsymbol{M} = \begin{bmatrix} \boldsymbol{I}_{\{N^2\times N^2\}} & \boldsymbol{0}_{\{N^2\times 2N+1\}} \end{bmatrix}$ truncates the SHs in $\boldsymbol{Y}$ to the order $N-1$ while $\boldsymbol{M}_{\{x,y,z\}}$ apply weights and shifts according to the x, y, or z recurrences of (7)–(9). The matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ contain the $2N+2N$ rows for $n = N$ as defined by the recurrences of (10)–(11). Due to the index shift for m in the z relation, the first extending recurrence only exists for $-n+1 \leq m \leq n$, while the second one only exists for $-n \leq m \leq n-1$, as shown by the blue patches in the two $n = 3$ rows of Fig. 1. The figure displays all the recurrence coefficients utilized on the left-hand side for $Y_n^m x$, $Y_n^m y$, $Y_n^m z$ and on the right-hand side for Y_n^m in grayscale, according to the matrix representation (12).

2.2 Identifying $\boldsymbol{T}$ from joint diagonalization

Following the observation model in Sec. 1.2 we insert $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{T}$ into the system of (12), cf. [7].

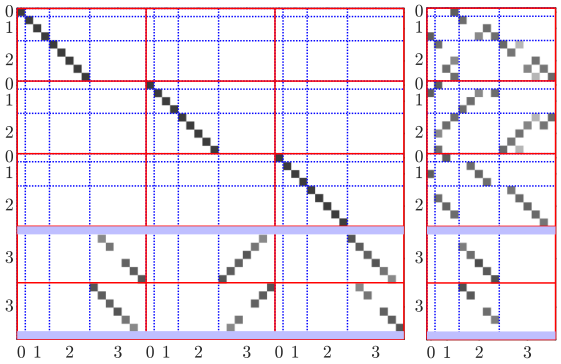


Fig. 1: Image showing the entries of the recurrence coefficient matrices of the left and right side of (12) for a maximum Ambisonic order $N = 3$.

Left-inversion numerically leaves a $3J \times J$ right-hand matrix whose $J \times J$ partitions $\boldsymbol{\Psi}_x$, $\boldsymbol{\Psi}_y$, $\boldsymbol{\Psi}_z$ relate to a joint eigenvalue problem $\boldsymbol{\Psi}_x = \boldsymbol{T}\mathrm{diag}\{\boldsymbol{x}\}\boldsymbol{T}^{-1}$, $\boldsymbol{\Psi}_y = \boldsymbol{T}\mathrm{diag}\{\boldsymbol{y}\}\boldsymbol{T}^{-1}$, $\boldsymbol{\Psi}_z = \boldsymbol{T}\mathrm{diag}\{\boldsymbol{z}\}\boldsymbol{T}^{-1}$ with the DOA parameters as eigenvalues and $\boldsymbol{T}$ containing the joint eigenvectors:

$$\begin{bmatrix} \boldsymbol{T}\mathrm{diag}\{\boldsymbol{x}\}\boldsymbol{T}^{-1} \\ \boldsymbol{T}\mathrm{diag}\{\boldsymbol{y}\}\boldsymbol{T}^{-1} \\ \boldsymbol{T}\mathrm{diag}\{\boldsymbol{z}\}\boldsymbol{T}^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{MU} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{MU} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{MU} \\ \boldsymbol{AU} & \boldsymbol{BU} & \boldsymbol{CU} \end{bmatrix}^{\dagger} \begin{bmatrix} \boldsymbol{M}_x\boldsymbol{U} \\ \boldsymbol{M}_y\boldsymbol{U} \\ \boldsymbol{M}_z\boldsymbol{U} \\ \boldsymbol{DU} \end{bmatrix} =: \begin{bmatrix} \boldsymbol{\Psi}_x \\ \boldsymbol{\Psi}_y \\ \boldsymbol{\Psi}_z \end{bmatrix}. \tag{13}$$

The superscript $(\cdot)^{\dagger}$ denotes the pseudoinverse. The performance of EB-ESPRIT algorithms is highly influenced by the accuracy and robustness of the algorithm that is employed to solve this simultaneous diagonalization problem. The evaluation below compares the JAPAM-5 algorithm [13] to the ad-hoc method, cf. Sec. 2.3, and to our proposed, simplified joint Schur decomposition that is described in Sec. 2.4.

2.3 Best disjoint eigendecomposition (ad-hoc)

Previous publications found ad-hoc methods to produce sufficiently accurate results [6][7]. In particular, the ad-hoc method diagonalizes the three individual problems $\boldsymbol{\Psi}_x$, $\boldsymbol{\Psi}_y$, $\boldsymbol{\Psi}_z$ disjointedly, and from the resulting three eigenvector matrices $\boldsymbol{T}_x$, $\boldsymbol{T}_y$, $\boldsymbol{T}_z$, the one $\hat{\boldsymbol{T}}$ is chosen that is able to best diagonalize all three matrices jointly, based on the least off-diagonal squares,

$$\hat{\boldsymbol{T}} = \arg\min_{\boldsymbol{T}} \sum_{\mu\in\{x,y,z\}} \left\| \mathrm{zdiag}(\boldsymbol{T}^{-1}\boldsymbol{\Psi}_\mu \boldsymbol{T}) \right\|_{\mathrm{F}}^2, \tag{14}$$

where zdiag zeroes the diagonal and $\|\cdot\|_{\mathrm{F}}^2$ is the squared Frobenius norm. The resulting matrices contain the DOA parameters on their diagonals, $\boldsymbol{x} = \mathrm{diag}\{\hat{\boldsymbol{T}}^{-1}\boldsymbol{\Psi}_x\hat{\boldsymbol{T}}\}$, $\boldsymbol{y} = \mathrm{diag}\{\hat{\boldsymbol{T}}^{-1}\boldsymbol{\Psi}_y\hat{\boldsymbol{T}}\}$, $\boldsymbol{z} = \mathrm{diag}\{\hat{\boldsymbol{T}}^{-1}\boldsymbol{\Psi}_z\hat{\boldsymbol{T}}\}$.

2.4 Joint Schur decomposition

The real-valued formulation permits to employ an approximate joint Schur decomposition (JSD) that only uses orthogonal, real-valued Schur eigenvectors to jointly reduce all three matrices $\boldsymbol{\Psi}_x$, $\boldsymbol{\Psi}_y$, $\boldsymbol{\Psi}_z$ to upper triangular form, revealing the direction parameters as eigenvalues on the diagonals. Oseledets et al. [16] proposed a fast algorithm for a joint generalized Schur decomposition (JGSD) involving left and right subspace eigenvectors and successive deflation.

As listed in Alg. 1, we propose a non-generalized JSD as a simplification of the JGSD. It involves a single, both left and right subspace eigenvector $\boldsymbol{v}$ per deflation. The objective of JSD($\{\boldsymbol{\Psi}_x, \boldsymbol{\Psi}_y, \boldsymbol{\Psi}_z\}$) is to find the eigenvalues associated with the joint eigenvectors of the REVEB-ESPRIT matrix set. The JSD successively deflates the matrices that are processed as $\{\boldsymbol{A}_k\}$ by their first dimension $\boldsymbol{e}_1 = [1,0,0,\ldots]^{\mathrm{T}}$ after Householder-mapping the newly discovered joint eigenvector $\boldsymbol{v}$ there. The eigenvalues to $\boldsymbol{v}$ constitute the i^{th} DOA vector $\boldsymbol{\theta}_i = [\lambda_1, \lambda_2, \lambda_3]^{\mathrm{T}}$. A full succession of deflations starts with the size $i = J$, and when $i = 1$ is reached, matrices are scalars and directly identify the last DOA $\boldsymbol{\theta}_1 = [\boldsymbol{A}_1, \boldsymbol{A}_1, \boldsymbol{A}_3]^{\mathrm{T}}$. Steps to form explicit upper triangular matrices are skipped as only eigenvalues are needed. Until convergence or reaching an iteration limit, the eigenvector estimate $\boldsymbol{v}$ for one deflation is initially guessed, then repeatedly refined by: (i), a Rayleigh-quotient estimating the eigenvalues λ_k, assuming $\boldsymbol{v}$ is a joint eigenvector, (ii), superimposed shifted matrices $\boldsymbol{B}_k = \boldsymbol{A}_k - \lambda_k \boldsymbol{I}$ squared that ought to vanish when multiplied by a joint eigenvector, assuming its eigenvalues are λ_k, and, (iii), a corresponding inverse iteration to improve $\boldsymbol{v}$.

Algorithm 1 Joint Schur Decomposition for REVEB-ESPRIT

procedure JSD($\{\boldsymbol{A}_k\}$)
  $i = \mathrm{size}(\boldsymbol{A}_1)$ ▷ initial size counter $i = J$
  while $i > 1$ **do** ▷ deflate $\{\boldsymbol{A}_k\}$ by j.eigenvec.
    $\boldsymbol{v} = \mathrm{randn}(i, 1)$ ▷ init. joint eigenvec. estim.
    $\boldsymbol{v} = \boldsymbol{v}/\|\boldsymbol{v}\|$ ▷ L2-normalize it
    repeat ▷ iteratively refine $\boldsymbol{v}$ for all k
      $\lambda_k = \boldsymbol{v}^{\mathrm{T}}\boldsymbol{A}_k\boldsymbol{v}, \forall k$ ▷ estimate eigenvals.
      $\boldsymbol{B}_k = \boldsymbol{A}_k - \lambda_k \boldsymbol{I}, \forall k$ ▷ shifted matrices
      $\boldsymbol{G} = \sum_k \boldsymbol{B}_k^{\mathrm{T}}\boldsymbol{B}_k$ ▷ sum of squares
      $\boldsymbol{v}_{\mathrm{old}} = \boldsymbol{v}$, ▷ store old estimate
      $\boldsymbol{v} = \boldsymbol{G}^{-1}\boldsymbol{v}_{\mathrm{old}}$ ▷ re-estimate j.eigenvec.
      $\boldsymbol{v} = \boldsymbol{v}/\|\boldsymbol{v}\|$ ▷ L2-normalize it
    until $1 - |\boldsymbol{v}^{\mathrm{T}}\boldsymbol{v}_{\mathrm{old}}| < \mathrm{tol}$ ▷ $\boldsymbol{v}$ converged?
    store $\{\lambda_k\}$ ▷ i^{th} eigenvalue set
    $\boldsymbol{q} = \mathrm{sign}(v_1)\boldsymbol{e}_1 + \boldsymbol{v}$ ▷ Householder v. of $\boldsymbol{v}$
    $\boldsymbol{Q} = \boldsymbol{I} - 2\frac{\boldsymbol{q}\boldsymbol{q}^{\mathrm{T}}}{\|\boldsymbol{q}\|^2}$ ▷ Householder reflection
    $\boldsymbol{A}_k = \boldsymbol{Q}_{[2:\mathrm{end},:]}\boldsymbol{A}_k\boldsymbol{Q}_{[:,2:\mathrm{end}]}, \forall k$ ▷ defl. by $\boldsymbol{v}$
    i- - ▷ decrease size counter
  end while
  store $\{\lambda_k\} \equiv \{\boldsymbol{A}_k\}$ ▷ eigenval. set for $i = 1$
end procedure

3 Numerical evaluation study

We evaluate REVEB-ESPRIT using the ad-hoc diagonalization from Sec. 2.3, JAPAM-5 [13], and the proposed JSD from Sec. 2.4, and compare them to the complex-valued EVEB-ESPRIT algorithm using the ad-hoc method in a plane-wave, free-field scenario following the signal model in (3). We use unit-variance white-noise signals $\boldsymbol{s}(t)$ and adjust the variance of the additive white Gaussian noise $\boldsymbol{\nu}(t)$ to modify the signal-to-noise ratio (SNR). The J test directions $\{\hat{\boldsymbol{\theta}}_j\}$ are drawn randomly from a 48-point t-design [23]. Despite ad-hoc methods only decompose the disjoint problems $\boldsymbol{\Psi}_x$,$\boldsymbol{\Psi}_y$,$\boldsymbol{\Psi}_z$ separately, they work well under realistic conditions [6]. Under rigorous conditions, only the joint methods can resolve, for example, 3 distinct directions with $x_1 = x_2$, $y_2 = y_3$, $z_1 = z_3$, because algebraic multiplicity leaves an undetermined degree of freedom to mix the corresponding eigenvectors in the x,y,z sub-problems. In the respective 2 dimensions of the example with algebraic multiplicity, ad-hoc methods will yield a $\boldsymbol{T}_x$ that does not diagonalize $\boldsymbol{\Psi}_y$, $\boldsymbol{\Psi}_z$, a $\boldsymbol{T}_y$ that does not diagonalize $\boldsymbol{\Psi}_x$, $\boldsymbol{\Psi}_z$, and a $\boldsymbol{T}_z$ that does not diagonalize $\boldsymbol{\Psi}_x$, $\boldsymbol{\Psi}_y$. For the comparison to become realistic, normally-distributed noise of 2° variance is added to every set of test directions to avoid such perfectly-matching pairs in x,y,z. The maximum Ambisonic order is kept constant at $N = 3$ in all tests. For every combination of SNR and number of sources J, the DOA estimation is repeatedly instanti-

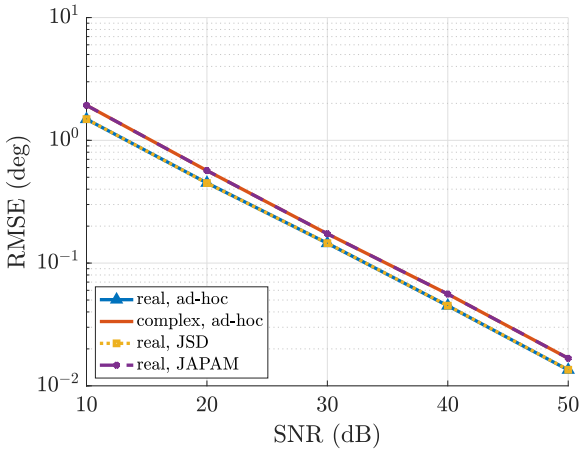
ated $K = 400$ times with source directions, source signals, and noise generated randomly. The algorithmic performance is evaluated using the root-mean-square error RMSE $= \sqrt{\sum_{k=1}^{K}\sum_{j=1}^{J}\left|\varepsilon_{\theta,j}^{(k)}\right| / (KJ)}$ of the great-circle distance $\varepsilon_{\theta,l}^{(k)} = \arccos(\boldsymbol{\theta}_j^{(k)\mathrm{T}}\hat{\boldsymbol{\theta}}_j^{(k)})$ between the estimated direction $\boldsymbol{\theta}_j^{(k)}$ and the true direction $\hat{\boldsymbol{\theta}}_j^{(k)}$. For every instance k, the estimated source directions are iteratively assigned to the true source directions with the smallest great-circle distance before evaluating RMSE.

The JSD and JAPAM-5 algorithms both used a tolerance tol $= 10^{-7}$ and were limited to 100 iterations.

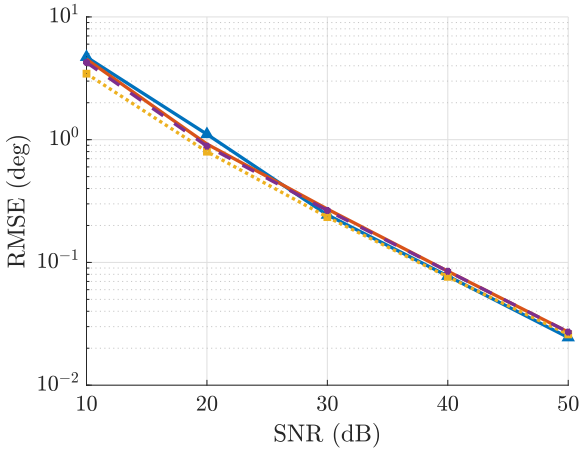
3.1 Results

Fig. 2a shows the obtained RMSEs for the different algorithms and $J = \{2, 6, 13\}$ simultaneous source directions. The proposed real-valued algorithm using the JSD consistently yields the lowest RMSE in all performed scenarios. In the case of $J = 2$ simultaneous DOAs, the real-valued ad-hoc algorithm performs equally well, while the complex-valued ad-hoc algorithm and JAPAM-5 show slightly increased RMSEs. For $J = 6$ sources and low SNRs, the proposed algorithm again achieves the lowest RMSEs. At an SNR of 10 dB it achieves an RMSE that, with 3.4°, is about 0.8° smaller than the next best competitor's RMSE. At high SNRs exceeding 30 dB, all algorithms perform similarly well in this scenario. In the test case with $J = 13$ simultaneously estimated directions, the proposed algorithm outperforms the competitors at all SNRs. At an SNR of 10 dB, it achieves an RMSE of around 27°, which is 3.9° lower than the second best algorithm's RMSE, and at 50 dB an RMSE of 2.9°, which is about 0.8° lower than the next closest RMSE. Interestingly, the complex-valued ad-hoc algorithm achieves lower RMSEs than its real-valued counterpart in this case, which can only be explained by the advantages of its algebraic coupling regarding the x and y related subproblems to $x \pm \mathrm{i}y$. As the JAPAM-5 algorithm did not converge for $J = 13$ sources, there are no corresponding results displayed in this test case.

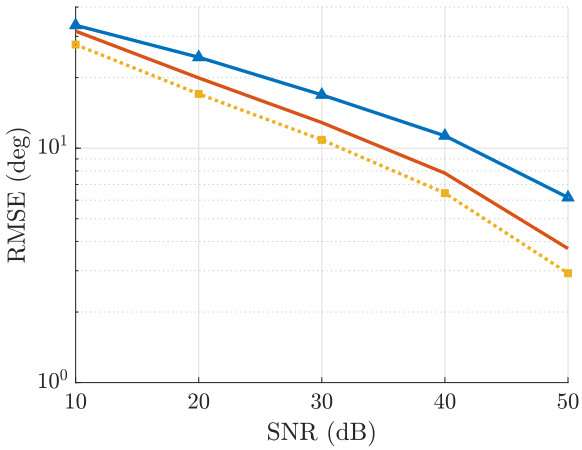
Tab. 1 shows mean execution times in ms of the algorithms for $J = \{2, 6, 13\}$ sources, averaged over the $K = 400$ randomly-instantiated repetitions and over the 5 SNR steps between 10 and 50 dB for which there is only moderate variation. In the test cases with $J = 2$ and $J = 6$ simultaneous source directions, the proposed



(a) $J = 2$



(b) $J = 6$



(c) $J = 13$

Fig. 2: RMSEs produced by the complex and real-valued EVEB-ESPRIT algorithms using different eigenvalue-revealing decompositions for a varying number of sources J. Note the different y-axis in (c).

	$J=2$	$J=6$	$J=13$
real, ad-hoc	0.985	0.760	1.26
complex, ad-hoc	1.12	1.05	2.26
real, JSD	0.579	0.631	1.70
real, JAPAM	1.72	7.41	N/A

Table 1: Mean execution times in ms of the different algorithms for $J = \{2, 6, 13\}$ simultaneous DOAs, averaged over $K = 400$ random initializations and 5 different SNRs between 10 and 50 dB.

algorithm yields the fastest execution times. The execution times of the ad-hoc methods are comparable to the proposed method, and the execution time of the JAPAM-5 algorithm is about 3 times longer than the one of the proposed algorithm with $J = 2$ and about 12 times longer with $J = 13$ simultaneous DOAs. In the test case with $J = 13$ source directions, the real-valued ad-hoc algorithm executes fastest with an average of 1.26 ms, while the real-valued JSD algorithm achieves an average execution time of 1.70 ms and the complex-valued ad-hoc algorithm executes slowest with an average of 2.26 ms. As the JAPAM-5 algorithm did not converge in this test case, no results are displayed. The simulations were carried out on a MacBook Pro with a 1.4 GHz Quad-Core Intel Core i5 processor and 8 GB RAM running MATLAB 2022b.

4 Conclusion

We introduced a fully real-valued formulation of the extended vector-based EB-ESPRIT (REVEB-ESPRIT) for multiple direction-of-arrival estimations from Ambisonic signals. This improves the compliance with Ambisonic format definitions that use real-valued spherical harmonics, requiring normalization to N3D, here, and it enables more efficient ways to retrieve the DOAs in the form of real eigenvalues of the inherent joint eigenvector problem. As solution method we proposed the simplification of an existing joint generalized Schur decomposition algorithm to a joint Schur decomposition (JSD) algorithm that is highly accurate and robust. Its real-valued implementation is more efficient than the complex-valued ad-hoc method and more efficient than the real-valued algorithm using the JAPAM-5 algorithm. For low numbers of simultaneous source directions, it is also more efficient than the real-valued ad-hoc method. The proposed REVEB-ESPRIT algorithm with joint-Schur diagonalization consistently yields the highest estimation accuracy in all tested scenarios. We provide an open MATLAB implementation of our simulation study[2], its recurrences[3], and the JSD/JGSD algorithms[4] including documentation.

5 Acknowledgment

The authors thank Jung-Woo Choi (KAIST, Daejeon, Korea) and Byeongho Jo (ETRI, Daejeon, Korea) for the fruitful collaboration on previous manuscripts, and Byeongho Jo for discussion and sharing his JAPAM-5 implementation.

References

[1] Teutsch, H. and Kellermann, W., "Eigen-beam processing for direction-of-arrival estimation using spherical apertures," *in Proc. Joint Workshop on Hands-Free Communication and Microphone Arrays, Piscataway, NJ*, (0), pp. 2–3, 2005.

[2] Teutsch, H. and Kellermann, W., "Detection and localization of multiple wideband acoustic sources based on wavefield decomposition using spherical apertures," in *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, 3, pp. 5276–5279, 2008, ISBN 1424414849, ISSN 15206149, doi: 10.1109/ICASSP.2008.4518850.

[3] Jo, B. and Choi, J.-W., "Direction of arrival estimation using nonsingular spherical ESPRIT," *The Journal of the Acoustical Society of America*, 143(3), pp. EL181–EL187, 2018, ISSN 0001-4966, doi:10.1121/1.5026122.

[4] Huang, Q., Zhang, L., and Fang, Y., "Two-Step Spherical Harmonics ESPRIT-Type Algorithms and Performance Analysis," *IEEE/ACM Transactions on Audio Speech and Language Processing*, 26(9), pp. 1684–1697, 2018, ISSN 23299290, doi: 10.1109/TASLP.2018.2836436.

[5] Jo, B. and Choi, J.-W., "Parametric direction-of-arrival estimation with three recurrence relations of spherical harmonics," *The Journal of the Acoustical Society of America*, 145(1), pp. 480–488, 2019, ISSN 0001-4966, doi:10.1121/1.5087698.

[2] https://git.iem.at/zotter/reveb-esprit/

[3] https://git.iem.at/thomasdeppisch/real-sh-recurrence-relations

[4] https://git.iem.at/zotter/jsd

[6] Herzog, A. and Habets, E. A. P., "Eigenbeam-ESPRIT for DOA-Vector estimation," *IEEE Signal Processing Letters*, 26(4), pp. 572–576, 2019, ISSN 10709908, doi:10.1109/LSP.2019.2898775.

[7] Jo, B., Zotter, F., and Choi, J.-W., "Extended Vector-Based EB-ESPRIT Method," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28, pp. 1692–1705, 2020, doi:10.1109/TASLP.2020.2996090.

[8] Jo, B. and Choi, J.-W., "Robust localization of early reflections in a room using semi real-valued EB-ESPRIT with three recurrence relations and Laplacian constraint," in *Proc. of the 23rd International Congress of Acoustics*, September, pp. 4949–4956, Aachen, 2019.

[9] Herzog, A. and Habets, E. A. P., "Online DOA estimation using real eigenbeam ESPRIT with propagation vector matching," in *EAA Spatial Audio Signal Processing Symposium*, pp. 19–24, Paris, 2019, doi:10.25836/sasp.2019.18.

[10] Jo, B., Zotter, F., and Choi, J. W., "Extended vector-based EB-ESPRIT using additional constraints for increasing the maximum number of detectable sources," in *Proc. of the 49th International Congress and Exposition on Noise Control Engineering (Inter-noise 2020)*, 2020, ISBN 9788994021362.

[11] Choi, J.-W., Zotter, F., Jo, B., and Yoo, J.-H., "Multiarray Eigenbeam-ESPRIT for 3D Sound Source Localization With Multiple Spherical Microphone Arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, pp. 2310–2325, 2022, doi:10.1109/TASLP.2022.3183929.

[12] Balda, E. R., Cheema, S. A., Weiss, A., Yeredor, A., and Haardt, M., "Perturbation analysis of Joint Eigenvalue Decomposition Algorithms," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, pp. 3101–3105, 2017, ISBN 9781509041176, ISSN 15206149, doi:10.1109/ICASSP.2017.7952727.

[13] André, R., Luciani, X., and Moreau, E., "A new class of block coordinate algorithms for the joint eigenvalue decomposition of complex matrices," *Signal Processing*, 145, pp. 78–90, 2018, ISSN 01651684, doi:10.1016/j.sigpro.2017.11.016.

[14] Haardt, M. and Nossek, J. A., "Simultaneous Schur decomposition of several nonsymmetric matrices to achieve automatic pairing in multidimensional harmonic retrieval problems," 46(1), pp. 161–169, 1998.

[15] Abed-Meraim, K. and Hua, Y., "A least-squares approach to joint Schur decomposition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, volume 4, pp. 2541–2544, 1998, doi:10.1109/ICASSP.1998.681669.

[16] Oseledets, I. V., Savostyanov, D. V., and Tyrtyshnikov, E. E., "Fast simultaneous orthogonal reduction to triangular matrices," *SIAM Journal on Matrix Analysis and Applications*, 31(2), pp. 316–330, 2009, ISSN 08954798, doi:10.1137/060650738.

[17] Zotter, F. and Frank, M., *Ambisonics, A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer, 2019, doi:10.1007/978-3-030-17207-7.

[18] Varshalovich, D. A., Moskalev, A. N., and Khersonskii, V. K., *Quantum theory of angular momentum*, World Scientific Pub. Co.,Teaneck, NJ, United States, 1987.

[19] Rafaely, B., *Fundamentals of Spherical Array Processing*, Springer, Berlin, 2015, doi:https://doi.org/10.1007/978-3-662-45664-4.

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

[21] Teutsch, H., *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, Springer, 2007, doi:https://doi.org/10.1007/978-3-540-40896-3.

[22] Ivanic, J. and Ruedenberg, K., "Rotation matrices for real spherical harmonics. Direct determination by recursion," *The Journal of Physical Chemistry*, 100(15), pp. 6342–6347, 1996.

[23] Hardin, R. H. and Sloane, N. J., "McLaren's improved snub cube and other new spherical designs in three dimensions," *Discrete and Computational Geometry*, 15(4), pp. 429–441, 1996, ISSN 01795376, doi:10.1007/BF02711518.