Estimation of the Noise Covariance Matrix for Rotating Sensor Arrays

Alastair H. Moore, Wei Xue, Patrick A. Naylor, and Mike Brookes,

***Abstract*—Estimation of the noise covariance matrix is addressed for a sensor array that rotates during desired source activity. Applications include beamformer design for head-mounted microphone arrays in assistive hearing devices. We propose a parametric model which leads to an analytical expression for the sensor signal covariance as a function of the array orientation and array manifold. The model allows the estimated noise covariance matrix to be updated in response to array rotation even during desired source activity. Simulation results demonstrate the efficacy of the method compared to a conventional, recursive estimation approach in which the estimate is not updated during desired source activity. The proposed method yields 18 dB lower error in the estimated noise covariance matrix and the resulting beamformer achieves noise reduction which is within 0.2 dB of an oracle beamformer.**

I. Introduction

Spatial filtering is a fundamental tool for multichannel signal enhancement in noisy and reverberant environments and is used in many applications, such as telecommunications, automatic speech recognition, human-robot interaction, assistive listening devices and hearing aids. The widely used minimum variance distortionless response (MVDR) beamformer [1] requires knowlege of two quantities: the steering vector, which defines the distortionless constraint, and the noise covariance matrix (NCM), which describes the interchannel relationship of the undesired signal. The focus of this contribution is the estimation of the NCM encountered by a microphone array in a non-isotropic sound field when the array can rotate freely in three dimensions.

It is common to calculate the NCM based on an assumed model of the noise field. Commonly-used models are spatially white noise [2], [3], spherically isotropic noise [4], [5] or cylindrically isotropic noise [6]–[8]. These models do not account for the true spatial distribution of the acoustic noise field since they are independent of the observed microphone signals.

Adaptive estimation of the NCM normally requires noise-only segments to be identified or an estimate of the speech-absence probability to be determined [9]. In [10], for example, it is assumed that the spatial characteristics of the noise do not change while the desired source is active. This allows an MVDR beamformer designed during noise-only segments to be used during speech activity.

A major source of non-stationarity in the NCM which arises in real-world situations is due to movement of the microphone array. We consider in particular the case of array rotation in response to desired source activity, for example, a robot turning to face a new talker. Since the desired source is active during the array rotation an immediate update of the NCM estimate is not possible.

The convenience of spherical harmonics (SHs) for accomodating rotations has been widely exploited in acoustic analysis for spherical microphone arrays [11]–[13] and in binaural rendering of sound scenes [14], [15]. However, in general it is the plane-wave density (PWD) of the sound field which is considered. In [16] the SH domain covariance matrix is used to estimate the diffuseness of the sound field which is modelled as an isotropic, and so rotation-invariant, background with individual coherent components. Like [16], we consider the sound field's SH covariance matrix. However, in this work, non-isotropic directionally-uncorrelated sound fields are considered.

Beamforming for rotating microphone arrays using a generalized sidelobe canceller structure is proposed in [17], [18], avoiding the need for an explicit estimate of the NCM. In contrast, in this work we propose a SH representation of the noise field from which the NCM can be determined under arbitrary rotations of the microphone array. An adaptive method for estimating the parameters of the proposed model along with further results and analysis can be found in [19].

The remainder of the paper is organized as follows. In Section II the problem is formulated. In Section III the notation and some key properties of SH analysis are briefly reviewed. In Section IV the proposed model of the non-isotropic directionally-uncorrelated field is presented and an analytical expression for the resulting NCM is derived. Simulation experiments which confirm the efficacy of the method under ideal and non-ideal conditions are presented in Section V. Finally, conclusions are drawn in Section VI.

II. Problem formulation

We consider a sound field which is sampled in successive time frames by an array of Q microphones. During speech absence, denoted $\mathcal{H}_0 = 1$, the sound field is considered to be noise-only and is assumed to be in the far-field. The power of the noise incident from each direction, or noise power distribution (NPD), is assumed to vary only slowly with time and so, over the time-scales considered in this paper, is treated as constant. Since, in general, the NPD varies with direction, the goal of this work is to estimate the time-varying NCM from knowledge of the microphone signals, the time-varying array orientation and the free-field array manifold.

Rather than estimate the NCM directly it is proposed to estimate the parameters of a model for the NPD since these are independent of array rotation. The estimated NPD is then

978-1-5386-9218-9/18/$31.00 ©2018 IEEE

used to derive the NCM as a function of the known array orientation. An advantage of this approach is that the estimated NCM can be updated in response to array rotation even during speech presence, i.e. $\mathcal{H}_0 = 0$.

The acoustic noise field is described by the PWD, $\underline{a}(\ell, \underline{\Omega})$, where $\underline{\Omega}$ is the direction of incident plane waves in world coordinates. Throughout this paper, we use an underbar to denote quantities that are defined in world coordinates and are therefore unaffected by array rotation. The NPD, $\underline{s}(\underline{\Omega})$, gives the direction dependence of the noise-field power and is

$$\underline{s}(\underline{\Omega}) = \mathbb{E}\{|\underline{a}(\ell, \underline{\Omega})|^2\} \tag{1}$$

where ℓ is the time-frame index, $\mathbb{E}\{\cdot\}$ denotes expectation and the expectation is over time.

The vector of noise signals, $\mathbf{v}(\nu, \ell)$, recorded by an array of Q microphones at frequency index ν and time-frame index ℓ is expressed directly in the short time Fourier transform (STFT) domain

$$\mathbf{v}(\nu, \ell) = \mathbf{x}(\nu, \ell) + \boldsymbol{u}(\nu, \ell) \tag{2}$$

where $\mathbf{x}(\nu, \ell)$ and $\boldsymbol{u}(\nu, \ell)$ are the vectors of Q complex-valued microphone signals due to the acoustic noise and sensor noise, respectively. Since each frequency bin is processed independently, the dependence on ν is omitted below.

Assuming $\mathbf{x}(\ell)$ and $\boldsymbol{u}(\ell)$ are uncorrelated, the NCM of the microphone noise signals is

$$\mathbf{R}_v(\ell) = \mathbf{R}_x(\ell) + \mathbf{R}_u \tag{3}$$

where $\mathbf{R}_v(\ell) \triangleq \mathbb{E}\left\{\mathbf{v}(\ell)\mathbf{v}^H(\ell)\right\}$ and $(\cdot)^H$ is the conjugate transpose. The acoustic NCM, $\mathbf{R}_x(\ell)$, and sensor NCM, $\mathbf{R}_u$ are similarly defined. Note that, like the NPD, the sensor NCM is assumed to be quasi-stationary. Therefore, the time-dependence in (3) arises only from array rotation.

In Section IV an expression which relates the acoustic NCM to a parametric model of the NPD is developed for a single realization of the array rotation, Λ. In Section V an online algorithm is used with time varying array rotation to adaptively estimate the model parameters.

In this work directions are expressed both in world coordinates and array coordinates. For an arbitrary array rotation, Λ, the relation between Ω and $\underline{\Omega}$ is given in [13, eq 1.65] and is here denoted by $\underline{\Omega}(\Omega, \Lambda)$.

III. Key properties of spherical harmonics

This section briefly presents the key properties of SH analysis to introduce the required notation. For a comprehensive introduction the reader is referred to [13] or [20].

The complex SH functions of order $n \geq 0$ and degree m with $|m| \leq n$ are defined for $\Omega = (\vartheta, \varphi) \in S^2$, as

$$Y_n^m(\Omega) = \sqrt{\frac{2n+1}{4\pi}\frac{(n-m)!}{(n+m)!}} P_n^m(\cos\vartheta)\, e^{\imath m\varphi} \tag{4}$$

where $P_n^m(\cdot)$ is the associated Legendre function and $\imath = \sqrt{-1}$. The SH functions form an orthonormal basis where increasing n and m results in functions, $Y_n^m(\Omega)$, with higher spatial frequency. The spherical Fourier transform (SFT) of a square-integrable function, $\mathcal{K}(\Omega)$, is given by

$$\mathcal{K}_{n,m} = \int_{\Omega \in S^2} \mathcal{K}(\Omega)\left[Y_n^m(\Omega)\right]^* d\Omega. \tag{5}$$

Assuming $\mathcal{K}(\Omega)$ is spatially bandlimited, its order, $N_{\mathcal{K}}$, is the maximum n for which any $\mathcal{K}_{n,m} > 0$. To avoid a plethora of subscripts, we index the individual SH functions with the single index $p = n^2 + n + m + 1$ such that $Y_p(\Omega) \equiv Y_n^m$ and $\mathcal{K}_p \equiv \mathcal{K}_{n,m}$, where $1 \leq p \leq P_{\mathcal{K}} = (N_{\mathcal{K}} + 1)^2$.

The inverse SFT (ISFT) decomposes $\mathcal{K}(\Omega)$ in terms of the $Y_p(\Omega)$ and may be written in vector form as

$$\mathcal{K}(\Omega) = \boldsymbol{\kappa}^T \mathbf{y}(\Omega) \tag{6}$$

where the SH coefficient vector $\boldsymbol{\kappa} = \begin{bmatrix}\mathcal{K}_1 & \dots & \mathcal{K}_{P_{\mathcal{K}}}\end{bmatrix}^T$, the SH function vector $\mathbf{y}(\Omega) = \begin{bmatrix}Y_1(\Omega) & \dots & Y_{P_{\mathcal{K}}}(\Omega)\end{bmatrix}^T$ and $(\cdot)^T$ denotes the transpose.

If $\mathcal{K}(\Omega)$ is expressed in a rotated frame of reference as $\underline{\mathcal{K}}(\underline{\Omega}(\Omega, \Lambda))$, then the resultant SH coefficients may be obtained in the SH domain as

$$\underline{\boldsymbol{\kappa}} = \mathbf{D}(\Lambda)\boldsymbol{\kappa} \tag{7}$$

where $\mathbf{D}(\Lambda)$ is the Wigner-D rotation matrix [13], which is block-diagonal and sparse.

IV. Proposed parametric model

In Section IV-A the acoustic NCM is related to the covariance matrix of a SH domain representation of the PWD of the sound field. In Section IV-B a parametric model for the NPD is presented and its relationship to the acoustic NCM derived. Throughout this section, for clarity of notation, a single realization of the microphone array orientation, Λ, is assumed such that the acoustic NCM, $\mathbf{R}_x$, is time-invariant.

A. Acoustic NCM from sound-field covariance

Assuming the effective support of the array manifold is short compared to the STFT frame length, the acoustic noise, $x_q(\ell)$, observed by the q^{th} microphone may be expressed as the array response to an infinite sum of plane waves over S^2

$$x_q(\ell) = \int_{\Omega \in S^2} h_q(\Omega)\underline{a}\big(\ell, \underline{\Omega}(\Omega, \Lambda)\big) d\Omega \tag{8}$$

where $\underline{a}(\ell, \underline{\Omega})$ is expressed in world coordinates and represents the PWD of the sound field at the origin in the absence of the microphone array and $h_q(\Omega)$ for $1 \leq q \leq Q$ is the array manifold.

In order to express (8) in terms of SHs, let $\mathbf{a}(\ell)$ be the SH coefficient vector representing the PWD of the sound field in array coordinates. In world coordinates, $\underline{a}(\ell, \underline{\Omega}(\Omega, \Lambda)) = a\big(\ell, \Omega\big)$ may be decomposed using the ISFT from (6)

$$\underline{a}\big(\ell, \underline{\Omega}(\Omega, \Lambda)\big) = a(\ell, \Omega) = \mathbf{a}^T(\ell)\mathbf{y}(\Omega). \tag{9}$$

To express $h_q(\Omega)$ in the SH domain ammenable to simplification, the SFT of the conjugate array manifold, $h_q^*(\Omega)$, of microphone q is defined as [14]

$$\tilde{h}_{q,p} = \int_{\Omega \in S^2} h_q^*(\Omega) Y_p^*(\Omega) d\Omega \tag{10}$$

with the corresponding ISFT as

$$h_q^*(\Omega) = \tilde{\mathbf{h}}_q^T \mathbf{y}(\Omega) \tag{11}$$

where $\tilde{\mathbf{h}}_q = \begin{bmatrix} \tilde{h}_{q,1} & \ldots & \tilde{h}_{q,P_h} \end{bmatrix}^T$. Note that properties of the SFT imply that, in general, $\tilde{h}_{q,p}^* \neq h_{q,p}$.

Substituting (9) and the conjugate of (11) into (8) gives

$$x_q(\ell) = \int_{\Omega \in S^2} \tilde{\mathbf{h}}_q^H \mathbf{y}^*(\Omega)\mathbf{y}^T(\Omega)\mathbf{a}(\ell) d\Omega \tag{12}$$

$$= \tilde{\mathbf{h}}_q^H \mathbf{a}(\ell) \tag{13}$$

where the simplification follows from the orthonormality of SHs [13]. The vector, $\mathbf{x}(\ell)$, of all Q microphone signals is therefore given by

$$\mathbf{x}(\ell) = \widetilde{\mathbf{H}}^H \mathbf{a}(\ell) \tag{14}$$

where $\widetilde{\mathbf{H}} = \begin{bmatrix} \tilde{\mathbf{h}}_1 & \tilde{\mathbf{h}}_2 & \ldots & \tilde{\mathbf{h}}_Q \end{bmatrix}$.

The acoustic NCM can therefore be expressed as

$$\mathbf{R}_x = \widetilde{\mathbf{H}}^H \mathbf{R}_a \widetilde{\mathbf{H}} \tag{15}$$

where

$$\mathbf{R}_a = \mathbb{E}\left\{\mathbf{a}(\ell)\mathbf{a}^H(\ell)\right\} \tag{16}$$

is the SH covariance matrix of the acoustic noise field in array coordinates. Expressed in vectorized form [21], (15) is

$$\mathbf{r}_x = \left(\widetilde{\mathbf{H}}^T \otimes \widetilde{\mathbf{H}}^H\right) \mathbf{r}_a \tag{17}$$

where $\mathbf{r}_x = \overrightarrow{\mathbf{R}_x}$, $\mathbf{r}_a = \overrightarrow{\mathbf{R}_a}$, $\otimes$ denotes the Kronecker product and $\overrightarrow{\cdot}$ denotes the vectorization of a matrix obtained by concatenating its columns. Thus (15) and (17) express the acoustic NCM in terms of the SH covariance matrix of the acoustic noise field.

B. Spherical harmonic model of noise power distribution

If the acoustic noise sound field is directionally-uncorrelated, the covariance between two directions, $\underline{\Omega}$ and $\underline{\Omega}'$, of the sound-field PWD is

$$\mathbb{E}\left\{\underline{a}(\ell,\underline{\Omega})\underline{a}^*(\ell,\underline{\Omega}')\right\} = \underline{s}(\underline{\Omega})\delta(\underline{\Omega} - \underline{\Omega}') \tag{18}$$

where $\underline{s}(\underline{\Omega})$ is the noise power distribution (NPD) defined in (1) and $\delta(\Omega - \Omega')$ is the delta function on the sphere at $\Omega' = (\vartheta', \varphi')$. The parameters of the proposed model are the SFT coefficients, $\underline{\mathbf{s}}$, of $\underline{s}(\underline{\Omega})$ which satisfy the ISFT relation

$$\underline{s}(\underline{\Omega}) = \underline{\mathbf{s}}^T \mathbf{y}(\underline{\Omega}) = \mathbf{y}^T(\underline{\Omega})\underline{\mathbf{s}} \tag{19}$$

where $\underline{\mathbf{s}}$ is a P_s element vector of SH coefficients. The valid range of $\underline{\mathbf{s}}$ is constrained such that the NPD, $\underline{s}(\underline{\Omega})$, is real-valued and non-negative for all $\underline{\Omega}$.

The model parameters, $\underline{\mathbf{s}}$, are defined in world coordinates which means that they are independent of the array rotation, Λ. Using (7) and (19), $s(\Omega)$ can be written in array coordinates as

$$s(\Omega) = \underline{s}(\underline{\Omega}(\Omega, \Lambda)) \tag{20}$$

$$= \mathbf{y}^T(\Omega)\mathbf{D}(\Lambda^{-1})\underline{\mathbf{s}} \tag{21}$$

where Λ^{-1} denotes the inverse rotation of Λ.

It is shown in [19] that element (p', p'') of $\mathbf{R}_a$ in (16) is given by

$$\mathbb{E}\left\{a_{p'}(\ell)a_{p''}^*(\ell)\right\} = \mathbf{g}_{p',p''}^T \mathbf{D}^T(\Lambda)\underline{\mathbf{s}}^* \tag{22}$$

where $\mathbf{g}_{p',p''} = \begin{bmatrix} G_{1,p',p''} & \ldots & G_{P_s,p',p''} \end{bmatrix}^T$ and

$$G_{p,p',p''} = \int_{\Omega \in S^2} Y_p(\Omega)Y_{p'}(\Omega)Y_{p''}^*(\Omega) d\Omega \tag{23}$$

is the Gaunt coefficient, for which a closed form solution is given in [22, pp. 39–40]. Therefore the vectorized SH covariance matrix, $\mathbf{r}_a$ from (17), of a directionally-uncorrelated sound field can be written as

$$\mathbf{r}_a = \mathbf{G}\mathbf{D}^T(\Lambda)\underline{\mathbf{s}}^* \tag{24}$$

where $\mathbf{G}$ is a $P_h^2 \times P_s$ matrix in which row $p' + (p'' - 1)P_h$ is equal to $\mathbf{g}_{p',p''}^T$ and P_h and P_s are the number of SH coefficients used to describe the microphone array manifold, $h_q(\Omega)$, and the noise power distribution (NPD), $\underline{s}(\underline{\Omega})$, respectively.

Substituting (24) into (17) the acoustic NCM is

$$\mathbf{r}_x = \left(\widetilde{\mathbf{H}}^T \otimes \widetilde{\mathbf{H}}^H\right) \mathbf{G}\mathbf{D}^T(\Lambda)\underline{\mathbf{s}}^* \tag{25}$$

$$= \mathbf{B}\mathbf{D}^T(\Lambda)\underline{\mathbf{s}}^* \tag{26}$$

where $\mathbf{B} = \left(\widetilde{\mathbf{H}}^T \otimes \widetilde{\mathbf{H}}^H\right)\mathbf{G}$. Note that $\mathbf{B}$ is a $Q^2 \times P_s$ matrix and is independent of both array rotation and fluctuations in the sound field. This time-invariance means it can be calculated once for a given array manifold. As a result, for a given array manifold, the cost of calculating $\mathbf{R}_x$ from (26) is independent of P_h, the number of SH coefficients used to describe the array manifold.

An online algorithm for estimating $\mathbf{G}$ based on an exponentially-weighted least squares (EWLS) cost function is proposed in [19]. The exponential weighting factor, $0 < \lambda \leq 1$, controls the adaptation rate with values closer to 1 causing the parameter estimates to change more slowly.

V. Evaluation

In this section, simulation experiments are reported which demonstrate the efficacy of the proposed method in comparsison to conventional signal dependent and independent methods, highlighting in particular the case when array rotation is in response to desired source activity.

A. Experiment setup

The array manifold in (8), $h_q(\nu, \Omega)$, for an array of microphones on the surface of a rigid sphere with radius 9 cm is calculated analytically using a SH expansion [23], [24]. The expansion order is set to 16, which ensures the worst case reconstruction error across all frequencies considered is less than −80 dB. The microphones are equally spaced on a circle, 20° above the horizontal plane. Experiments 1 to 3 use $Q = 4$ microphones while Experiment 4 uses both $Q = 4$ and also $Q = 16$ to investigate the effect of varying Q.

A sound field with known spatial distribution is simulated directly in the time-frequency (TF) domain using independent zero-mean circularly-symmetric Guassian noise signals incident from 578 directions. These directions form a spherical sampling quadrature grid supporting SH decomposition up to order 16. Using (8) discretized according to the quadrature grid, the power of each plane wave is given by $s(\Omega_i)$, as in (21).

Sensor noise is simulated by adding independent zero-mean circularly-symmetric Gaussian noise to each microphone signal. The sensor noise power at each microphone is drawn from a Gaussian distribution with mean fixed at −20 dB with respect to the acoustic noise power, averaged over all microphones, and variance equal to 10 % of the mean. The simulations therefore represent a typical use case where the received signals are dominated by acoustic noise and the sensor noise is similar, but not identical, across microphones.

The rotation sequence and speech absence state, $\mathcal{H}_0(\ell)$, are chosen to reflect a situation in which array rotation is in response to desired source activity. Rotation of the array is implemented according to piecewise-constant trajectories so that frames in which a change in the array orientation occurs are clearly identifiable. The first four orientations have deterministic yaw angles, $\{0°, 30°, 60°, 90°\}$ while pitch and roll are stochastic, as in sequence 2. In the final orientation, roll, pitch and yaw are all 0°. Each orientation is held for 250 frames. Whenever $\mathcal{H}_0(\ell) = 1$ the proposed method and the conventional method use noise-only microphone signals to update the estimated NCM. Whenever $\mathcal{H}_0(\ell) = 0$, the conventional method does not update but, in contrast, the proposed method uses the previously estimated NPD to estimate the NCM from the current array orientation. For $\ell \leq 950$, $\mathcal{H}_0(\ell) = 1$ while for $\ell > 950$, $\mathcal{H}_0(\ell) = 0$.

The proposed method requires a measurement of the array orientation and $\mathcal{H}_0(\ell)$ at each frame. Errors in the yaw, pitch and roll components of the measured array orientation are simulated as independent identically-distributed zero-mean, additive Gaussian noise with standard deviation, $\sigma_{\text{IMU}} = 1°$ while oracle values of $\mathcal{H}_0(\ell)$ are used.

The power distribution of the synthesized sound fields have an axially-symmetric cardioid shape

$$\underline{s}(\underline{\Omega}) = (1/2)^{N_s}(1 + \cos(\underline{\Omega} - \underline{\Omega}_0))^{N_s} \tag{27}$$

where $\underline{\Omega}_0$ is the direction of the maximal response and N_s is the order, where higher-order cardioids concentrate the energy over a narrower region. In each experiment, evaluations are conducted for 20 different sound fields, each with $\underline{\Omega}_0$ aligned to one of the faces of an icosahedron. An oracle voice activity detector (VAD) is used to determine the speech absence state, $\mathcal{H}_0(\ell)$.

The implementation of the proposed method uses a single parameter to describe the sensor noise, which intentionally introduces a mismatch compared with the simulated conditions. Except where otherwise stated, $N_h = 15$ (from which $P_h = (N_h + 1)^2 = 256$), the frequency is 2200 Hz and $(1 - \lambda) = 1 \times 10^{-5}$.

B. Metrics and baseline approaches

The error in the estimated NCM is assessed as the Frobenius norm of the scale-invariant error

$$\mathcal{E}(\ell) = \min_{\varrho} \|\mathbf{R}_v(\ell) - \varrho\hat{\mathbf{R}}_v(\ell)\|_F \tag{28}$$

where $\mathbf{R}_v(\ell)$ is defined in (3) and ϱ is used as in [25], to make the metric independent of an arbitrary scaling factor. This independence of scale allows direct comparison with the fixed-scale model covariance matrices used as baselines, described below.

The NCM estimation accuracy is also evaluated in terms of the noise reduction, γ, obtained using an MVDR beamformer [26] according to

$$\gamma = \frac{1}{LQ} \sum_{\ell \in \mathcal{L}} \frac{\mathbf{v}^H(\ell)\mathbf{v}(\ell)}{|Z(\ell)|^2} \tag{29}$$

where $Z(\ell)$ is the beamformer output. The MVDR steering vector is $\mathbf{h}(\Omega) = \begin{bmatrix} h_1(\Omega) & \dots & h_Q(\Omega) \end{bmatrix}^T$ and the look direction is fixed in array coordinates to $\Omega = (90°, 0°)$, that is, towards the front of the array. The oracle beamformer, designed using the ground truth NCM, defines the best case performance. The excess noise level, $\Delta\gamma$, of a beamformer is the amount by which the noise power at the output of a beamformer exceeds that of the oracle beamfomer.

For comparison, results are also reported for two typical noise covariance models and a conventional estimation approach. The spatially white model [2], [3], denoted 'White', assumes uncorrelated noise at each microphone. Further assuming that the variance is the same at each microphone and exploiting the scale-invariant error metrics, the spatially white NCM is the identity matrix.

The spherically isotropic model [4], [5], denonted 'Sph iso', assumes the power incident from all directions is the same and so (26) reduces to

$$\hat{\mathbf{r}}_{\text{iso}}(\ell) \propto \overrightarrow{\widetilde{\mathbf{H}}^H\widetilde{\mathbf{H}}}. \tag{30}$$

The recursive smoothing approach [10], denoted 'RS', updates the estimated NCM according to

$$\hat{\mathbf{r}}_{\text{rs}}(\ell) = (1 - \alpha)\,\hat{\mathbf{r}}_{\text{rs}}(\ell - 1) + \alpha\,\overrightarrow{\mathbf{v}(\ell)\mathbf{v}^H(\ell)} \tag{31}$$

only when $\mathcal{H}_0(\ell) = 1$. The smoothing factor, α, controls the trade off between tracking changes in $\mathbf{r}_v(\ell)$ and the variance of the estimate. Results are shown for $\alpha \in \{1, 5, 10, 50, 100\} \times 10^{-3}$.

C. Results

Fig. 1 shows the convergence of $\mathcal{E}(\ell)$ for the conventional RS method, denoted 'RS: α', and the proposed method, denoted 'EWLS: $(1 - \lambda)$', over a range of time constants. As a baseline, the error in the estimated NCM is −11.1 dB for 'White' and −11.4 dB for 'Sph iso'. For RS, the minimum error is obtained with $\alpha = 5 \times 10^{-3}$. In this case, on each orientation change, there is a large increase in $\mathcal{E}(\ell)$ before quickly converging again. The problem with the conventional

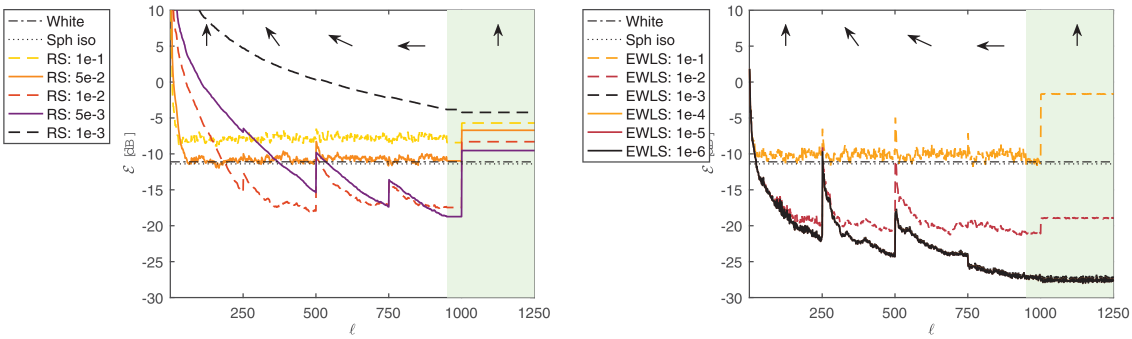

Fig. 1. Noise covariance error over time for (left) recursive smoothing (RS) and (right) proposed method (EWLS) with different time constants. Error using signal independent models, 'White' and 'Sph iso' also shown for comparison. Shaded region indicates $\mathcal{H}_0(\ell) = 0$. Overlayed arrows indicate yaw component of array rotation, which is stepped every 250 frames in sequence $\{0°, 30°, 60°, 90°, 0°\}$.

method is clearly demonstrated at $\ell = 1001$, where the orientation changes but, because $\mathcal{H}_0(\ell) = 0$, the estimate cannot be updated. The error in the estimated noise covariance during desired source activity is −9.5 dB which is larger than for the straightforward model-based estimates.

For the proposed method with $(1-\lambda) \leq 1e-3$, convergence is insensitive to the choice of λ with no visible difference over three orders of magnitude. Crucially, at $\ell = 1001$ when $\mathcal{H}_0(\ell) = 0$ and the orientation changes, there is no increase in $\mathcal{E}(\ell)$. The proposed method therefore achieves 18 dB lower error than the RS approach.

The fundamental difference between the two approaches is that the proposed method adapts to the properties of the noise field in world coordinates and so the choice of λ depends only on how quickly the NPD changes whereas the RS method must adapt to changes in the observation of the noise field through the microphone signals in array coordinates and so α must be chosen to allow for more rapid adaptation.

The effectiveness of MVDR beamforming depends on the number of microphones. As the number of microphones increases, so does the sensitivity to inaccuracies in the NCM. The effect of model mismatch between the estimation order, $N_{\hat{s}}$, and the true NPD order, N_s, is investigated for arrays with $Q = 4$ and $Q = 16$ microphones. Fig. 2(a) shows the excess noise, $\Delta\gamma$, for the same $Q = 4$ microphone array as in Figure 1. As the noise field becomes more directional there is a small increase in γ obtained using the oracle beamformer, 9.2 dB to 9.7 dB. The proposed method achieves the best performance of $\Delta\gamma \leq 0.05$ dB over all N_s and $N_{\hat{s}}$. This implies that choosing $N_{\hat{s}} \neq N_s$ does not have a detrimental effect on the beamforming performance.

In contrast, the RS method with best-case smoothing parameter, $\alpha = 5 \times 10^{-3}$, achieves $\Delta\gamma = 0.16$ dB for $N_s = 1$, rising to $\Delta\gamma = 0.66$ dB for $N_s = 4$. That is, as the sound field becomes more directional, the 0.5 dB improvement in the oracle beamformer's noise reduction is not matched when using the RS estimate of the NCM. Using either a spatially white or spherically isotropic model of the noise field achieves

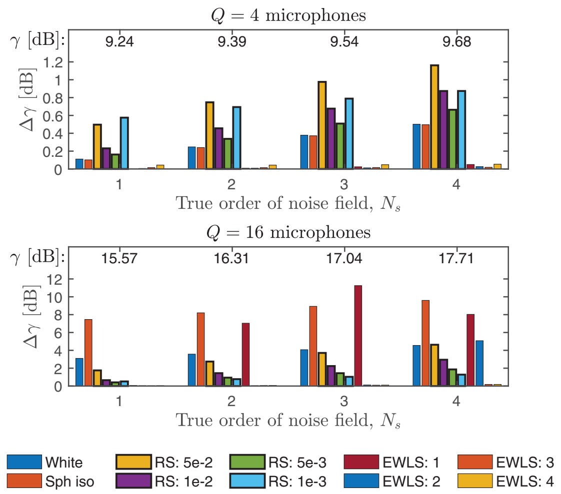

Fig. 2. Excess noise power, $\Delta\gamma$, of MVDR beamformer with (top) $Q = 4$ and (bottom) $Q = 16$ with different noise covariance estimation methods, averaged over 20 N_s-order cardioid-shaped noise fields. Results are grouped by the true NPD order, N_s. The noise reduction, γ, obtained from an oracle MVDR beamformer is shown above each group. Labels for White, Sph iso and RS methods are as for Fig. 1. Labels for the proposed method with different estimation orders are denoted 'EWLS: $N_{\hat{s}}$'.

0.05 dB to 0.16 dB more noise reduction than the best RS estimate.

Fig. 2(b) shows the beamforming performance when the number of microphones is increased to 16. The oracle beamformer can better match the beam pattern to the power distribution of the noise field than with 4 microphones and so more noise reduction is acheived (15.57 dB to 17.71 dB), with the benefit increasing with N_s. With the proposed method, provided $N_{\hat{s}} \geq N_s$, performance very close to that of the oracle beamformer is acheived, with $\Delta\gamma \leq 0.18$ dB. With RS the best-case excess noise rises from 0.43 dB for $N_s = 1$ to 1.28 dB for $N_s = 4$. Since with $Q = 16$ there are more degrees of freedom, the effect of undermodelling is more severe than for $Q = 4$. This is true for the spatially white model – which does not account for the cross-terms in the NCM or the intersensor variations in noise power – the spherically isotropic model – which does not account for the directional variation in the NPD and ignores the sensor noise – and for the proposed method with $N_{\hat{s}} < N_s$. The spatially white assumption is the most robust since, by not modelling the interchannel correlations, it also does not attempt to exploit them in the noise reduction. This leads to 3.10 dB to 4.55 dB more residual noise than the oracle beamformer. In contrast, the effect of errors due to the spherically isotropic assumption and the undermodelled proposed method degrade the beamformer by 7.46 dB to 9.60 dB and 7.04 dB to 11.25 dB, respectively.

VI. Conclusion

A model for non-isotropic directionally-uncorrelated noise has been proposed based on a SH decomposition of the sound field. An analytical expression for the noise covariance matrix is obtained directly from the proposed model using knowledge of the array manifold and the array orientation. An algorithm for estimating the parameters of the proposed model has been proposed and validated on simulated noise fields with realistic levels of microphone sensor noise. The approach is particularly suited to situations in which changes in array orientation in response to and during desired source activity are expected. In this context, the proposed method achieves 18 dB lower error in the estimated noise covariance matrix than the conventional recursive averaging approach, and noise reduction which is within 0.05 dB of an oracle beamformer using the ground truth noise covariance matrix.

References

[1] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.

[2] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New York, USA, Apr. 1988, pp. 2578–2581.

[3] ——, "Noise reduction based on microphone array with LMS adaptive post-filtering," *IEE Electronics Lett.*, vol. 26, no. 24, pp. 2036–2581, Nov. 1990.

[4] I. McCowan and H. Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 709–716, Nov. 2003.

[5] S. Leukimmiatis and P. Maragos, "Optimum post-filter estimation for noise reduction in multichannel speech processing," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sep. 2006, pp. 1–5.

[6] A. Kuklasiński, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood based multi-channel isotropic reverberation reduction for hearing aids," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Sep. 2014, pp. 61–65.

[7] A. Kuklasiński, S. Doclo, T. Gerkmann, S. H. Jensen, and J. Jensen, "Multi-channel PSD estimators for speech dereverberation - A theoretical and experimental comparison," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 91–95.

[8] A. Schwarz and W. Kellermann, "Coherent-to-diffuse power ratio estimation for dereverberation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 6, pp. 1006–1018, Jun. 2015.

[9] M. Souden, J. Chen, J. Benesty, and S. Affes, "An integrated solution for online multichannel noise tracking and reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 7, pp. 2159–2169, 2011.

[10] U. Kjems and J. Jensen, "Maximum likelihood based noise covariance matrix estimation for multi-microphone speech enhancement," in *Proc. European Signal Processing Conf. (EUSIPCO)*, 2012, pp. 295–299.

[11] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Microphone array measurement system for analysis of directional and spatial variations of sound fields," *J. Acoust. Soc. Am.*, vol. 112, no. 5, pp. 1980–1991, 2002.

[12] V. Tourbabin, H. Barfuss, B. Rafaely, and W. Kellermann, "Enhanced robot audition by dynamic acoustic sensing in moving humanoids," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Apr. 2015, pp. 5625–5629.

[13] B. Rafaely, *Fundamentals of Spherical Array Processing*, ser. Springer Topics in Signal Processing. Berlin Heidelberg: Springer-Verlag, 2015.

[14] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. Am.*, vol. 127, p. 823, 2010.

[15] J. Sheaffer, S. Villeval, and B. Rafaely, "Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction," in *Proc. of the EAA Joint Symposium on Auralization and Ambisonics*, Apr. 2014.

[16] N. Epain and C. T. Jin, "Spherical Harmonic Signal Covariance and Sound Field Diffuseness," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 10, pp. 1796–1807, Oct. 2016.

[17] M. Zohourian, Alan Archer-Boyd, and R. Martin, "Multi-channel speaker localization and separation using a model-based GSC and an inertial measurement unit," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 5615–5619.

[18] M. Zohourian and R. Martin, "Binaural speaker localization and separation based on a joint ITD/ILD model and head movement tracking," in *Proc. IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 430–434.

[19] A. H. Moore, W. Xue, P. A. Naylor, and M. Brookes, "Noise covariance matrix estimation for rotating microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, 2018, (early access).

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

[21] M. Brookes, "The Matrix Reference Manual," Imperial College London, Website, 1998.

[22] M. E. Rose, *Elementary Theory of Angular Momentum*. New York: John Wiley & Sons, Inc., Dec. 1957.

[23] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. San Diego, California, USA; London, UK: Academic Press, 1999.

[24] C. T. Jin, N. Epain, and A. Parthy, "Design, optimization and evaluation of a dual-radius spherical microphone array," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 193–204, 2014.

[25] P. A. Naylor and N. D. Gaubitch, Eds., *Speech Dereverberation*. Springer-Verlag, 2010.

[26] M. S. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.