HAL
open science

Statistical Analysis of the Multichannel Wiener Filter Using a Bivariate Normal Distribution for Sample Covariance Matrices

Chengshi Zheng, Antoine Deleforge, Xiaodong Li, Walter Kellermann

To cite this version:

Chengshi Zheng, Antoine Deleforge, Xiaodong Li, Walter Kellermann. Statistical Analysis of the Multichannel Wiener Filter Using a Bivariate Normal Distribution for Sample Covariance Matrices. IEEE/ACM Transactions on Audio, Speech and Language Processing, 2018, 26 (5), pp.951 - 966. 10.1109/TASLP.2018.2800283 . hal-01909612

HAL Id: hal-01909612

https://inria.hal.science/hal-01909612v1

Submitted on 5 Nov 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Statistical Analysis of the Multichannel Wiener Filter Using a Bivariate Normal Distribution for Sample Covariance Matrices

Chengshi Zheng, Antoine Deleforge, Xiaodong Li, and Walter Kellermann

***Abstract*—This paper studies the statistical performance of the multichannel Wiener filter (MWF) when the weights are computed using estimates of the sample covariance matrices of the noisy and the noise signals. It is well known that the optimal weights of the minimum variance distortionless response (MVDR) beamformer are only determined by the noisy sample covariance matrix or the noise sample covariance matrix, while those of the MWF are determined by both of them. Therefore, the difficulty increases dramatically in statistically analyzing the MWF when compared to analyzing the MVDR, where the main reason is that expressing the general joint probability density function (p.d.f.) of the two sample covariance matrices presented a hitherto unsolved problem, to the best of our knowledge. For a deeper insight into the statistical performance of the MWF, this paper first introduces a bivariate normal distribution to approximately model the joint p.d.f. of the noisy and the noise sample covariance matrices. Each sample covariance matrix is approximately modeled by a random scalar multiplied by its true covariance matrix. This approximation is designed to preserve both the bias and the mean squared error of the matrix with respect to a natural distance on covariance matrices. The correlation of the bivariate normal distribution, referred to as the *sample covariance matrices intrinsic correlation coefficient*, captures all second-order dependencies of the noisy and the noise sample covariance matrices. By using the proposed bivariate normal distribution, the performance of the MWF can be predicted from the derived analytical expressions and many interesting results are revealed. As an example, the theoretical analysis demonstrates that the MWF performance may degrade in terms of noise reduction and signal-to-noise-ratio improvement when using more sensors in some noise scenarios.**

***Index Terms*—Statistical analysis, multichannel Wiener filter, sample covariance matrix, bivariate normal distribution**

I. Introduction

The multichannel Wiener filter (MWF) is one of the most popular microphone-array speech enhancement (MASE) algorithms for several reasons [1]-[3]. Compared with the generalized sidelobe canceller (GSC) [4]-[6], the MWF is insensitive to signal model mismatch and it does not need any knowledge regarding the direction of arrival (DOA) of the desired speech [7], [8]. Compared with the minimum variance distortionless response (MVDR) filter [9]-[12], the MWF can automatically and simultaneously steer a beam to the DOA of the desired signal and suppress the noise at the output of the MVDR. This results from the fact that, given the noise and the noisy covariance matrices, the MWF can theoretically be decomposed into an optimum MVDR beamformer cascaded with a single-channel postfilter when there is only a single desired signal (see [13] and references therein).

Theoretical analysis of MASE algorithms has attracted a fast-growing interest over the last three decades. These theoretical studies can be classified into two categories. The first one is based on deterministic signal models [8], [14]-[19]. In [8], the robustness of the GSC and that of the MWF have been compared by both theory and experiment. In [15], the theoretical performance of the GSC is expressed as a function of the complex coherence and theoretical limits of multichannel noise reduction algorithms are examined in different noise fields. In [19], the MWF has been theoretically studied for second order statistics estimation errors, where the error matrix is assumed to be Hermitian and invertible. The other category of studies is based on stochastic signal models [20]-[25]. In [21], the amounts of noise reduction and speech distortion are theoretically analyzed via higher-order statistics when using a structure-generalized parametric blind spatial subtraction array. In [22] and [23], some two-channel post-filter estimators are statistically studied in isotropic noise fields. The former category treats signals as deterministic processes and provides asymptotical analysis of the performance of MASE algorithms. The latter treats signals as stochastic processes and reveals how estimation parameters influence performance. Stochastic signal models can also predict the performance of speech enhancement algorithms in transient conditions [26]. For speech and audio signals, it is often more reasonable to use stochastic signal models than deterministic signal models. In this area, stochastic signal models have already been widely used in both MASE algorithms and single-channel speech enhancement algorithms [1]. Moreover, numerous distributions of speech coefficients have already been proposed and used in recent years (see [27] and references therein).

Let $\mathbf{X}(k,l)$, $\mathbf{S}(k,l)$, and $\mathbf{N}(k,l)$ denote the noisy, the target, and the noise multichannel signals in the complex-valued short-time discrete Fourier domain such that $\mathbf{X}(k,l) = \mathbf{S}(k,l) + \mathbf{N}(k,l)$, where k and l denote frequency and frame indices, respectively, and each (k,l) bin contains an M-

C. Zheng and X. Li are with the Key Laboratory of Noise and Vibration Research, Institute of Acoustics, Chinese Academy of Science, Beijing, 100190, China, and also with University of Chinese Academy of Sciences, Beijing, 100049, China (email: {cszheng, lxd}@mail.ioa.ac.cn)

W. Kellermann is with the Chair of Multimedia Communications and Signal Processing, Friedrich-Alexander-University Erlangen-Nürnberg, 91058 Erlangen, Germany (e-mail: {walter.kellermann}@fau.de).

A. Deleforge is with Inria Rennes - Bretagne Atlantique, 35000 Rennes, France (email : antoine.deleforge@inria.fr).

This work was supported by NSFC (National Science Fund of China) under Grant No. 61571435.

Manuscript received July XX, 2017; revised XXXX XX, XX.

sensor observation, represented by a vector, e.g., $\mathbf{X}(k,l) = \begin{bmatrix} X_1(k,l) & \cdots & X_M(k,l) \end{bmatrix}^T$. $X_i(k,l)$ can be computed by

$$X_i(k,l) = \sum_{n=0}^{N-1} x_i(n+lR) w(n) e^{-j\frac{2\pi nk}{N}}, \tag{1}$$

where $x_i(n)$ is the time-domain noisy signal at the sensor i. N is the frame length, R is the frame shift and $w(n)$ is a window function. $S_i(k,l)$ and $N_i(k,l)$ can be computed in the same way.

Let us assume that $\mathbf{S}(k,l)$ and $\mathbf{N}(k,l)$ are two statistically independent, zero-mean circularly-symmetric complex Gaussian random vectors, i.e. $\mathbf{S}(k,l) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{ss}(k,l))$ and $\mathbf{N}(k,l) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{nn}(k,l))$, where $\mathbf{R}_{ss}(k,l)$ and $\mathbf{R}_{nn}(k,l)$ are, respectively, the inter-sensor covariance matrix of the desired signal and that of the noise. This assumption is only an approximation when the target is speech. Other models [27] could be considered to improve the model accuracy for speech applications. We further assume that the noisy observations are statistically independent over frequency and time for mathematical convenience. Note that this statistical independence assumption could only be valid if there is no overlap ($R \geq N$) between frames when computing the short-time discrete Fourier transform, and if the spectral leakage is ignored. By definition, the optimal multichannel Wiener filter $\mathbf{w}_{\text{opt}} \in \mathbb{C}^M$ is such that $\mathbf{w}_{\text{opt}}^H \mathbf{X}(k,l)$ represents the minimum mean squared error (MMSE) estimate of $S_1(k,l)$. Note that the first channel is chosen as the reference channel for the desired source and then, by this convention, the optimal Wiener filter can be written as

$$\mathbf{w}_{\text{opt}}(k,l) = [\mathbf{R}_{xx}(k,l)]^{-1} (\mathbf{R}_{xx}(k,l) - \mathbf{R}_{nn}(k,l)) \mathbf{e}_1, \tag{2}$$

where $\mathbf{e}_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T \in \mathbb{C}^M$ and $\mathbf{R}_{xx}(k,l) \in \mathbb{C}^{M\times M}$ and $\mathbf{R}_{nn}(k,l) \in \mathbb{C}^{M\times M}$ denote the covariance matrices of the noisy and the noise signals, respectively [12]. In practice, these matrices are unknown and sample covariance matrix (SCM) estimates $\widehat{\mathbf{R}}_{xx}(k,l)$ and $\widehat{\mathbf{R}}_{nn}(k,l)$ are used instead [8], [18], [19], [28]-[34]. It is well-known that the MWF performance critically depends on the quality of those estimates, e.g., [8]. In the stationary case, this quality is mostly governed by the number of temporal snapshots used to compute the estimates, and by the number of sensors. In this article, we also examine the influence of a third, much less-studied factor: the mutual correlation between $\widehat{\mathbf{R}}_{xx}$ and $\widehat{\mathbf{R}}_{nn}$. To see why this factor is important, let us consider the example where the observed noisy signal consists of noise only, i.e., $\mathbf{X}(k,l) = \mathbf{N}(k,l)$. In that case, the Wiener filter should ideally completely cancel the signal, and we should have $\widehat{\mathbf{w}}_{\text{opt}} = \mathbf{0}$. In practice, however, the noisy and the noise sample covariance matrices (SCMs) may be estimated from different sets of samples, and $\widehat{\mathbf{w}}_{\text{opt}} \neq \mathbf{0}$ in general. This example suggests that a perfect correlation between $\widehat{\mathbf{R}}_{xx}$ and $\widehat{\mathbf{R}}_{nn}$ would yield a perfect filter, while a decreasing correlation will likely decrease performances. Such effects are very difficult to study in theory, because the general joint probability density function (p.d.f.) of two SCMs is likely to be very intricate, and its general expression is still unknown in mathematics, to the best of our knowledge. In this article, we circumvent this problem by proposing a novel approach to model such dependencies. We first approximate the joint p.d.f of the two SCMs using a carefully designed bivariate normal distribution. This new p.d.f. has the advantage of capturing mutual correlations by a single variable ρ, that we will refer to as *sample covariance matrices intrinsic correlation coefficient* (SCMs-ICC). Under this simplified model, the theoretical amount of noise reduction and the signal-to-noise-ratio (SNR) improvement of the MWF can be expressed as a closed-form function of the SCMs-ICC ρ, the number of sensors M and the number of temporal snapshots used for estimation. We show that our model correctly predicts the MWF performance in a number of practical noise reduction scenarios, including situations when noise-only frames are erroneously detected as noisy frames containing the target signal. Throughout this paper, all the theoretical results are verified by comparison with performance obtained on Monte Carlo simulations or real-world recordings.

The remainder of this paper is organized as follows. Section II introduces an approximate bivariate normal distribution for the noisy and the noise SCMs. In Section III, we apply this approximate model to statistically analyze the MWF in noise-only segments, where the amount of noise reduction can be predicted correctly by an analytical expression. Section IV presents the experimental results in reverberant and noisy environments to verify the theoretical results for noise-only periods. In Section V, the proposed approximate model is further applied to analyze the MWF in noisy periods to show the amount of SNR improvement versus different parameters. Some conclusions are presented in Section VI.

II. An Approximate Bivariate Model for Sample Covariance Matrices

A. Signal Model and Problem Formulation

In the present study, we assume that the noise and the observed (noisy) SCMs are calculated as follows

$$\widehat{\mathbf{R}}_{nn}(k,l) = \frac{1}{L_n} \sum_{l_n=0}^{L_n-1} \mathbf{N}(k,l-l_n) \mathbf{N}^H(k,l-l_n), \tag{3}$$

and

$$\widehat{\mathbf{R}}_{xx}(k,l) = \frac{1}{L_x} \sum_{l_x=0}^{L_x-1} \mathbf{X}(k,l-l_x) \mathbf{X}^H(k,l-l_x), \tag{4}$$

where L_x and L_n denote the number of frames available for estimation[1]. Generally, the noise-only signal is not available and thus the noise SCM needs to be estimated from noise-only observations during target pauses. Therefore, it is often necessary to use a robust target activity detection algorithm on the observed noisy signal [28]. The target activity detector $\mathcal{P}$ decides here on whether a frame is considered for the

[1]In practical applications, it is more common to estimate $\widehat{\mathbf{R}}_{xx}(k,l)$ and $\widehat{\mathbf{R}}_{nn}(k,l)$ using a recursive averaging method. Statistical equivalences between non-recursive smoothing and first-order recursive smoothing for the smoothed periodograms are theoretically studied in [35], and an extension of the present results to recursive estimates is left for future work.

computation of $\widehat{\mathbf{R}}_{nn}(k,l)$, which can be described by:

$$\widehat{\mathbf{R}}_{nn}(k,l) = \widehat{\mathbf{R}}_{nn}(k,l-1) + \frac{1-\mathcal{P}(k,l)}{L_n} \times \left(\mathbf{X}(k,l)\mathbf{X}^H(k,l) - \mathbf{X}(k,l_{\min})\mathbf{X}^H(k,l_{\min})\right), \quad (5)$$

where $\mathcal{P}(k,l) = 1$ when the target signal is detected at the frequency index k of the frame index l; and $\mathcal{P}(k,l) = 0$, otherwise. For speech applications, when the target activity detector is often replaced by a soft speech presence probability (SPP) [36], $\mathcal{P}(k,l)$ can vary from 0 to 1, which is left for future research. $l_{\min}$ corresponds to the oldest frame index used to estimate the noise SCM. Fig. 1 plots an example of the estimation of the noise SCM using (5). The upper and the lower parts show how $\widehat{\mathbf{R}}_{nn}(k,l-1)$ and $\widehat{\mathbf{R}}_{nn}(k,l)$ are estimated, respectively. If $\mathcal{P}(k,l) = 1$, $\widehat{\mathbf{R}}_{nn}(k,l) = \widehat{\mathbf{R}}_{nn}(k,l-1)$ holds. If $\mathcal{P}(k,l) = 0$, the current noise-only observation should be included and the oldest frame indexed by $l_{\min}$ should be discarded when estimating the noise SCM. (5) is presented here to show how the noise SCM is estimated when only the noisy observation is available. Ideally, during noise-only periods, we have $\widehat{\mathbf{R}}_{nn}(k,l) = \widehat{\mathbf{R}}_{xx}(k,l)$ if we assume that $L_x = L_n$ and that the target activity detector is perfect[2]. However, we cannot expect to have a perfect target activity detector even during noise-only periods in practice, which leads to $\widehat{\mathbf{R}}_{nn}(k,l) \neq \widehat{\mathbf{R}}_{xx}(k,l)$. Thus, the following question arises: *with an imperfect target activity detector, what is the performance of the MWF in the case $L_x = L_n$?* More precisely, if there are L_o frames out of L_n frames (see Fig. 15 for illustration, the subscript 'o' means overlap of frames) used for estimating both $\widehat{\mathbf{R}}_{xx}(k,l)$ and $\widehat{\mathbf{R}}_{nn}(k,l)$, we want to predict the performance of the MWF for relevant signal models. For nonstationary noise sources, the assumption on $\mathbf{N}(k,l)$ will be violated and the noise covariance matrix $\mathbf{R}_{nn}(k,l)$ may change rapidly and it becomes difficult to quantitatively predict the performance of the MWF. However, this paper explains the impact of the number of frames on the estimation and therefore allows conclusions on how sensitive the MWF is if nonstationarity demands short averaging periods for the noise. Furthermore, this paper shows the impact of the SCMs-ICC on the estimation and thus can explain the importance of continuously updating the noise covariance matrix for the MWF, especially for nonstationary noise sources.

Under Gaussian i.i.d. assumptions in the complex Fourier domain, the SCMs in (3) and (4) follow complex Wishart distributions [37], i.e.,

$$\widehat{\mathbf{R}}_{nn} \sim \mathcal{W}_M^{\mathcal{C}}(\mathbf{R}_{nn}, L_n) \quad \text{and} \quad \widehat{\mathbf{R}}_{xx} \sim \mathcal{W}_M^{\mathcal{C}}(\mathbf{R}_{xx}, L_x), \quad (6)$$

where k and l are omitted as long as misunderstandings can be precluded. By substituting (3) and (4) into (2), the estimated

[2]In practical applications, it is more common to use $L_n \gg L_x$ with the assumption that the noise signal is more stationary than the target signal. Throughout this paper, only $L_x = L_n$ is studied for the following considerations. First, it can still give valuable insight into the MWF under a stochastic signal model, such as the amount of noise reduction and the signal-to-noise-ratio (SNR) improvement of the MWF. Second, this is the simplest relevant case to completely model the correlation of the noisy and the noise covariances, where the correlation can range from zero to one.

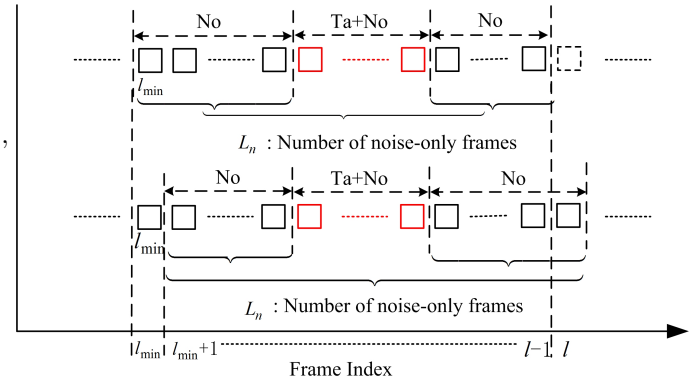

Fig. 1. Example of the estimation of the noise SCM using (5) when only the noisy observation is available. 'No' means noise-only frames and 'Ta+No' indicates frames with detected target activity and noise.

Wiener filter is given by:

$$\widehat{\mathbf{w}}_{\text{opt}} = \left[\widehat{\mathbf{R}}_{xx}\right]^{-1}\left(\widehat{\mathbf{R}}_{xx} - \widehat{\mathbf{R}}_{nn}\right)\mathbf{e}_1. \quad (7)$$

It is very difficult to rigorously derive the p.d.f. of $\widehat{\mathbf{w}}_{\text{opt}}$ for several reasons: Firstly, the joint p.d.f. of the noisy and the noise SCMs is still unknown and is likely to be very intricate, since the SCMs jointly contain $M \times (M+1)$ variables totally. Secondly, (7) contains three matrix operations, such as subtraction, inversion and multiplication, which can make derivations even more intricate. As we know, the most difficult derivation comes from the product of two matrices, especially as one matrix is the inverse of a complex Wishart-distributed matrix.

It should be mentioned that there are many ways to improve the estimation accuracy of the noise covariance matrix. For example, the Hermitian symmetric property has been exploited to reduce the estimation errors in [38]. In most recent works [24], [25], [39]-[42], the spatial coherence matrix of the noise and the acoustic transfer functions of the desired signal are assumed to be known and thus only one parameter, i.e., the time-varying noise power spectral density (PSD), needs to be estimated. This paper only studies the statistical performance of the MWF using the averaged noise SCM. The above methods which estimate noise covariance matrices by exploiting additional prior knowledge or assumptions are out of the scope of this paper.

B. A Bivariate Model for Covariance Matrices

To facilitate the estimation of the p.d.f. of $\widehat{\mathbf{w}}_{\text{opt}}$, we propose to approximate the joint p.d.f. of $\widehat{\mathbf{R}}_{nn}$ and $\widehat{\mathbf{R}}_{xx}$ using the following bivariate normal distribution:

$$\begin{bmatrix} \beta_x \\ \beta_n \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} \bar{\beta}_x \\ \bar{\beta}_n \end{bmatrix}, \begin{bmatrix} \sigma_{\beta_x}^2 & \rho\sigma_{\beta_x}\sigma_{\beta_n} \\ \rho\sigma_{\beta_x}\sigma_{\beta_n} & \sigma_{\beta_n}^2 \end{bmatrix}\right), \quad (8)$$

which we use to define the following random matrices to approximately model $\widehat{\mathbf{R}}_{nn}$ and $\widehat{\mathbf{R}}_{xx}$:

$$\widetilde{\mathbf{R}}_{xx} = \exp(\beta_x)\mathbf{R}_{xx} \quad \text{and} \quad \widetilde{\mathbf{R}}_{nn} = \exp(\beta_n)\mathbf{R}_{nn}. \quad (9)$$

This approximation is motivated by the fact that if a variable β is Gaussian, then its exponential function $\exp(\beta)$ is log-

normally distributed, which is known to approximately correspond to a χ^2 distribution [43]. The Wishart distribution can be seen as a generalization of the χ^2 distribution to matrices, and the diagonal elements of Wishart-distributed matrices are χ^2-distributed. Note that in (8), the factor ρ is introduced as the sample covariance matrices intrinsic correlation coefficient (SCMs-ICC) to capture all dependencies between $\widetilde{\mathbf{R}}_{xx}$ and $\widetilde{\mathbf{R}}_{nn}$, while $\bar{\beta}_x, \bar{\beta}_n, \sigma^2_{\beta_x}$ and $\sigma^2_{\beta_n}$ must be carefully chosen so that $\widetilde{\mathbf{R}}_{xx}$ and $\widetilde{\mathbf{R}}_{nn}$ are "good" approximations of $\widehat{\mathbf{R}}_{xx}$ and $\widehat{\mathbf{R}}_{nn}$, respectively. In the following section and Appendix C, the factor ρ in the log domain will be derived from the correlations in the linear domain, which is similar to the derivation in [44].

For conciseness of the discussion of these approximations, we use $\widehat{\mathbf{R}}$ to represent $\widehat{\mathbf{R}}_{xx}$ or $\widehat{\mathbf{R}}_{nn}$. Correspondingly, $\mathbf{R}$ represents $\mathbf{R}_{xx}$ or $\mathbf{R}_{nn}$ and $\widetilde{\mathbf{R}}$ corresponds to $\widetilde{\mathbf{R}}_{xx}$ or $\widetilde{\mathbf{R}}_{nn}$, and L represents L_x or L_n. Accordingly, we have

$$\widetilde{\mathbf{R}} = \exp(\beta)\,\mathbf{R}, \tag{10}$$

where β represents β_x or β_n. According to (8), β follows the normal distribution. In other words, we only need to obtain the mean, $\bar{\beta}$, and the variance, σ^2_β, to determine the p.d.f. of β. As mentioned above, $\widetilde{\mathbf{R}}$ needs to be a "good" approximation of $\widehat{\mathbf{R}}$ if we want to use $\widetilde{\mathbf{R}}$ instead of $\widehat{\mathbf{R}}$ to statistically analyze the MWF. In [8], it is emphasized that the estimation of the second order statistics of $\mathbf{R}_{nn}$ is important for the MWF, so we measure the goodness of $\widetilde{\mathbf{R}}$ in the second-order sense. That is to say, $\widetilde{\mathbf{R}}$ should have the same bias and the same mean square error (MSE) as $\widehat{\mathbf{R}}$. Using the proposed model (10), $\widetilde{\mathbf{R}}$ has only one variable, while $\widehat{\mathbf{R}}$ according to (5) has $M \times (M+1)/2$ variables. If we can use $\widetilde{\mathbf{R}}$ instead of $\widehat{\mathbf{R}}$ to analyze the statistical properties of the estimated Wiener filter in (7), a complicated problem can be transformed into a much simpler one. While the proposed model should thus suffice to analyze the second-order moment based Wiener filter, it will generally not be able to describe the statistical behaviour of higher-order statistics-based algorithms, e.g., blind source separation algorithms exploiting non-Gaussianity [45], [46].

We note that (10) has a very similar form to the estimation problem posed in [24], [25], [39]-[42]. In [47], the Cramér-Rao bounds of the reverberation PSD estimators used in [39] and [42] are analyzed. In [48], a Bayesian refinement of the ML-based postfilter for the MWF is further derived, which outperforms the ML-based estimator and a single-channel speech enhancement algorithm presented in [49]. However, there are some essential differences between (10) and the model posed in [24], [25], [39]-[42]: Firstly, the scalar $\exp(\beta)$ in (10) is not a time-varying PSD, while it follows a log-normal distribution. Secondly, $\mathbf{R}$ in (10) equals $E\{\widehat{\mathbf{R}}\}$ and thus is not only the spatial coherence matrix that is assumed to be known in previous works. Finally, the model in (10) is introduced to replace $\widehat{\mathbf{R}}$ to significantly simplify the statistical analysis of the estimated Wiener filter $\widehat{\mathbf{w}}_{\text{opt}}$ in (7). The aim of this paper is not to propose a new implementation algorithm of the MWF for practical applications.

There are at least two ways to measure the bias and the MSE of $\widehat{\mathbf{R}}$. One is the flat metric on $\mathscr{P}_M \cong \boldsymbol{Gl}_M(\mathbf{R})/\boldsymbol{O}_M(\mathbf{R})$ and the other is the natural metric on $\mathscr{P}_M$, where $\mathscr{P}_M$ is the space of covariance matrices [37], $\boldsymbol{Gl}_M(\mathbf{R})$ is the general linear group of degree M over $\mathbf{R}$ and $\boldsymbol{O}_M(\mathbf{R})$ is the Lie group of unitary matrices[3]. In this paper, we propose to use the natural metric on $\mathscr{P}_M$ to measure the second-order statistics of $\widehat{\mathbf{R}}$. The main reason is that the MSE depends on the actual underlying covariance matrix $\mathbf{R}$ if the flat metric on $\mathscr{P}_M$ is chosen, while the MSE is independent of $\mathbf{R}$ when using the natural metric on $\mathscr{P}_M$. Using the natural metric on $\mathscr{P}_M$[37], the bias and the MSE of $\widehat{\mathbf{R}}$, respectively, can be given by

$$\mathbf{B}(\mathbf{R}) = E\left\{\exp_{\mathbf{R}}^{-1}\widehat{\mathbf{R}}\right\}, \tag{11}$$

and

$$\varepsilon^2_{\text{cov}} = E\left\{d^2_{\text{cov}}\left(\widehat{\mathbf{R}}, \mathbf{R}\right)\right\}, \tag{12}$$

where $\mathbf{B}(\mathbf{R})$ in (11) defines the bias vector field of $\widehat{\mathbf{R}}$ with respect to $\mathbf{R}$ as defined in [37, (100)] and $\varepsilon^2_{\text{cov}}$ is the MSE of $\widehat{\mathbf{R}}$ that relates to the root MSE as defined in [37, (65)]. "$\exp_{\mathbf{R}}$" is the exponential map and "$\exp_{\mathbf{R}}^{-1}$" is its inverse, which is defined as $\exp_{\mathbf{R}}^{-1}\widehat{\mathbf{R}} = \mathbf{R}^{1/2}\left(\log \mathbf{R}^{-1/2}\widehat{\mathbf{R}}\mathbf{R}^{-1/2}\right)\mathbf{R}^{1/2}$. The inverse exponential map corresponds to the square root of general matrices and to the logarithm of positive-definite Hermitian matrices [37]. $d_{\text{cov}}\left(\widehat{\mathbf{R}}, \mathbf{R}\right) = \left(\sum_k (\log \lambda_k)^2\right)^{1/2}$ in (12) is the root MSE of $\widehat{\mathbf{R}}$ as defined in [37, (65)], where λ_k, with $k = 1, 2, \cdots, M$, are the generalized eigenvalues of the linear pencil $\widehat{\mathbf{R}} - \lambda\mathbf{R}$. (11) and (12) define the bias and the MSE of a (sample covariance) matrix $\widehat{\mathbf{R}}$ with respect to its true covariance matrix $\mathbf{R}$, which are similar to the bias and the MSE of an estimation with respect to its true value.

As mentioned above, $\widetilde{\mathbf{R}}$ should have the same bias and the same MSE as $\widehat{\mathbf{R}}$ using the natural metric on $\mathscr{P}_M$, thus

$$\mathbf{B}(\mathbf{R}) = E\left\{\exp_{\mathbf{R}}^{-1}\widehat{\mathbf{R}}\right\} = E\left\{\exp_{\mathbf{R}}^{-1}\widetilde{\mathbf{R}}\right\}, \tag{13}$$

and

$$\varepsilon^2_{\text{cov}} = E\left\{d^2_{\text{cov}}\left(\widehat{\mathbf{R}}, \mathbf{R}\right)\right\} = E\left\{d^2_{\text{cov}}\left(\widetilde{\mathbf{R}}, \mathbf{R}\right)\right\}. \tag{14}$$

The expected value of β, which can be derived from (13) (see Appendix A), is given by

$$\bar{\beta} = E\{\beta\} = -\log L + \frac{1}{M}\sum_{i=1}^{M}\psi(L-i+1), \tag{15}$$

where $\psi(\bullet)$ is the Digamma function [50], which is the expectation of logarithm of the chi-square distribution χ^2_{L-i+1} with $2(L-i+1)$ degrees of freedom [37]. Eq. (15) implicitly requires $L \geq M$, because the degrees of freedom of χ^2_{L-i+1} for all $i = 1, \cdots, M$ should be positive. If $L < M$, $\widehat{\mathbf{R}}_{xx}$ is a

[3]The flat metric on the space covariance matrices is expressed using the Frobenius norm, while the natural metric is expressed using the 2-norm of the vectors of logarithms of the generalized eigenvalues between two positive-definite matrices [37]. As the covariance matrix $\mathbf{R}$ is Hermitian and positive-definite, $\mathbf{R}$ has the Cholesky decomposition $\mathbf{R} = \mathbf{A}\mathbf{A}^H$, where $\mathbf{A} \in \boldsymbol{Gl}_M(\mathbf{R})$ (the general linear group) is an invertible matrix. A general linear group of degree M is defined as the set of $M \times M$ invertible matrices, together with the operation of ordinary matrix multiplication. $\mathbf{A}$ has the polar decomposition $\mathbf{A} = \mathbf{U}\mathbf{P}$, where $\mathbf{U} \in \boldsymbol{O}_M(\mathbf{R})$ is the Lie group of unitary matrices and $\mathbf{P} \in \mathscr{P}_M$ is a positive-definite symmetric matrix. A Lie group is defined as a manifold with differentiable group operations.

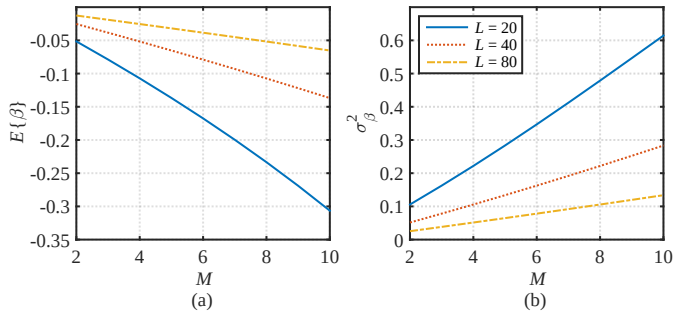

Fig. 2. The mean and the variance of β versus M for different values of L: (a) $\bar{\beta} = E\{\beta\}$; (b) σ_β^2.

singular matrix that is not invertible, because summing up L rank-1 matrices of type $\mathbf{X}\mathbf{X}^H$ leads to rank L (at best). This will not be further discussed.

From (14), the variance of β can be given by

$$\sigma_\beta^2 \approx \operatorname{var}\{\beta\} = M/L + 5\left(M/L\right)^2 \Big/ 6 - \left(E\{\beta\}\right)^2, \quad (16)$$

as is derived in Appendix B.

Before studying the behavior of β, it is explained why the χ^2 distribution cannot be used directly to model $\exp(\beta)$ in (10). In [44], a bivariate χ^2 distribution is introduced in modeling two correlated periodogram bins and the relation between the correlation in the log domain and that in the linear domain is derived in a rigorous way. If we used the bivariate χ^2 distribution for modeling $\exp(\beta_x)$ and $\exp(\beta_n)$ as in [44], the derivation would become much simpler than the bivariate normal distribution used in this paper. Indeed, the degrees of freedom of the χ^2 distribution can capture the impact of the number of frames L, and the correlation ρ has already been derived formally in [44]. However, it is still unknown how to find the relation between the degrees of freedom and the number of sensors M. Furthermore, it is well-known that the χ^2 distribution has only one parameter, i.e., the number of degrees of freedom, while the normal distribution has two parameters, where one is the mean and the other is the variance. As mentioned above, the first and second order statistics of $\widetilde{\mathbf{R}}$ and $\widehat{\mathbf{R}}$ should be equal, and thus both (13) and (14) need to be satisfied. If we used the χ^2 distribution for modeling $\exp(\beta)$, there were two equations and only one unknown parameter, i.e., the number of degrees of freedom, which leads to an overdetermined system that has no solution. If the normal distribution is introduced for β, there are two equations and two unknown parameters and thus we can determine the mean and the variance uniquely with (15) and (16), respectively.

From (15) and (16), we note that both the mean and the variance of β are independent of $\mathbf{R}$, and they both only depend on the number of sensors M and the number of frames L. This is expected when using the natural metric rather than the flat metric (see above). To give an intuitive idea of the behaviour of β, Fig. 2 plots its mean and its variance versus the number of sensors M for different values of L. One can see that both the mean and the variance of β tend to 0 when increasing L. However, for the same value of L, the absolute values of the mean and the variance increase when increasing M. It is quite intuitive that the absolute values of the mean and the variance should reduce when increasing L for a fixed value of M, since $\widehat{\mathbf{R}} \to \mathbf{R}$ results when $L \to \infty$. As shown by (10), when the mean and the variance of β approach zero, $\widetilde{\mathbf{R}} \to \mathbf{R}$ also results. For a fixed value of L, the Cramér-Rao bound on the natural distance between $\mathbf{R}$ and $\widehat{\mathbf{R}}$ increases as the number of sensors M increases [37, (82)] and so does the bias [37, (102)]. β is introduced to capture both the impact of L and M on estimation accuracy of $\mathbf{R}$ using $\widetilde{\mathbf{R}}$.

Turning to estimate the p.d.f. of $\widehat{\mathbf{w}}_{\text{opt}}$ in (7) using the approximations of (8) and (9), we note that given M, L_x, and L_n we can compute $\bar{\beta}_x$ and $\bar{\beta}_n$ using (15). Meanwhile, $\sigma_{\beta_x}^2$ and $\sigma_{\beta_n}^2$ can be calculated from (16). Now only the SCMs-ICC ρ is missing, which will be discussed in the next section.

III. Statistical Analysis of the MWF during Noise-Only Periods

This section studies the amount of noise reduction using (7) with the proposed model presented in (8) and (9). We will analyze the dependency of noise reduction on the number of frames, that on the number of sensors, and that on the SCMs-ICC in both theory and simulation. Monte Carlo simulations in reverberation-free environments will be considered to validate the analytical results.

For $L_x = L_n$, (8) reduces to

$$\begin{bmatrix} \beta_x \\ \beta_n \end{bmatrix} \sim \mathcal{N}_2\left(\begin{bmatrix} \bar{\beta} \\ \bar{\beta} \end{bmatrix}, \begin{bmatrix} \sigma_\beta^2 & \rho\sigma_\beta^2 \\ \rho\sigma_\beta^2 & \sigma_\beta^2 \end{bmatrix}\right), \quad (17)$$

where $\bar{\beta} = E\{\beta_x\} = E\{\beta_n\}$ is the mean and $\sigma_\beta^2 = \operatorname{var}\{\beta_x\} = \operatorname{var}\{\beta_n\}$ is the variance of β_x or β_n. β_x and β_n have the same means and variances because they are only determined by the number of frames and the number of sensors as shown in (15) and (16). Note that even in noise-only periods, $\widehat{\mathbf{R}}_{nn}(k,l) = \widehat{\mathbf{R}}_{xx}(k,l)$ cannot be guaranteed. This is because $\widehat{\mathbf{R}}_{nn}(k,l)$ and $\widehat{\mathbf{R}}_{xx}(k,l)$ may be estimated using different frames of $\mathbf{X}$ (for instance, when $\mathcal{P}(k,l) = 0$ is erroneously detected as $\mathcal{P}(k,l) = 1$), which is due to an imperfect target activity detector in noise-only periods. Thus, we need β_x and β_n to capture the differences between $\widehat{\mathbf{R}}_{nn}(k,l)$ and $\widehat{\mathbf{R}}_{xx}(k,l)$ for statistical analysis in noise-only periods. The parameter ρ is determined by the number of frames, L_o, in estimating both the noise and the noisy SCMs, which is $\rho \approx L_o/L_x = L_o/L_n$ with $L_o \in [0 \quad L_x]$ (support of this assumption is provided by Appendix C). ρ captures all second-order dependencies between $\widehat{\mathbf{R}}_{xx}$ and $\widehat{\mathbf{R}}_{nn}$, which are represented by $\widetilde{\mathbf{R}}_{xx}$ and $\widetilde{\mathbf{R}}_{nn}$, respectively. The parameter ρ in (17) is the SCMs-ICC as introduced in (8).

Since $\widehat{\mathbf{R}}_{xx}$ and $\widehat{\mathbf{R}}_{nn}$ are modeled by $\widetilde{\mathbf{R}}_{xx}$ and $\widetilde{\mathbf{R}}_{nn}$ regarding the first and second-order statistics, this paper proposes to substitute $\widetilde{\mathbf{R}}_{xx}$ and $\widetilde{\mathbf{R}}_{nn}$ into (7) to allow for a performance analysis of the MWF in a statistical way. For noise-only segments (i.e., $\mathbf{R}_{xx} = \mathbf{R}_{nn}$), this yields

$$\begin{aligned} \widetilde{\mathbf{w}}_{\text{opt}} &= \left[\widetilde{\mathbf{R}}_{xx}\right]^{-1}\left(\widetilde{\mathbf{R}}_{xx} - \widetilde{\mathbf{R}}_{nn}\right)\mathbf{e}_1 \\ &= \left(1 - \frac{\exp\left(\beta_n\right)}{\exp\left(\beta_x\right)}\right)\mathbf{e}_1 = G_0\mathbf{e}_1, \end{aligned} \quad (18)$$

where

$$G_0 = \left(1 - \frac{\exp(\beta_n)}{\exp(\beta_x)}\right). \tag{19}$$

We note that the estimated Wiener filter reduces to a gain function G_0 multiplying a vector $\mathbf{e}_1$ when a bivariate model for sample covariance matrices is proposed, and thus it becomes simpler to study the behavior of the MWF using $\widetilde{\mathbf{w}}_{\text{opt}}$ than using $\widehat{\mathbf{w}}_{\text{opt}}$.

In (19), G_0 could be not only negative but also smaller than -1, which may result in amplifying the noise instead of suppressing it. This is because no constraint is introduced when computing $\widehat{\mathbf{R}}_{xx} - \widehat{\mathbf{R}}_{nn}$ in (7). To prevent this unexpected result, some improved versions of the MWF have already been proposed, such as speech distortion weighted-MWF (SDW-MWF) [28], rank-one SDW-MWF [8], [19], spatial prediction SDW-MWF [30], parametric MWF [31], and eigenvalue decomposition-based MWF [32]. Among these improved versions, a positive semi-definite constraint is introduced when computing $\widehat{\mathbf{R}}_{ss} = \widehat{\mathbf{R}}_{xx} - \widehat{\mathbf{R}}_{nn}$, where $\widehat{\mathbf{R}}_{ss}$ is the estimated covariance matrix of the desired signal. Although the proposed model for SCMs in Section II can also be applied to study the performance of such constrained versions of the MWF, e.g., by setting $G_0 = 0$ when $G_0 < 0$, we concentrate on the essential problems linked to the basic MWF solution (7).

The power transfer function (PTF) from the first sensor to the output of the MWF during noise-only periods [8] can be derived directly with the help of (18), which is given by

$$\frac{\widetilde{\mathbf{w}}_{\text{opt}}^H \mathbf{R}_{nn} \widetilde{\mathbf{w}}_{\text{opt}}}{\mathbf{e}_1^H \mathbf{R}_{nn} \mathbf{e}_1} = \frac{G_0^2 \mathbf{e}_1^H \mathbf{R}_{nn} \mathbf{e}_1}{\mathbf{e}_1^H \mathbf{R}_{nn} \mathbf{e}_1} = G_0^2. \tag{20}$$

From (20), the theoretical amount of noise reduction [23] is given by

$$\text{NR}_{\text{MWF}}\,[\text{dB}] = -10 \log_{10}\left(\int_{-\infty}^{1} G_0^2 p_1(G_0)\, dG_0\right), \tag{21}$$

where $p_1(G_0)$ is the p.d.f. of G_0. As shown in (19), G_0 is only a function of β_x and β_n. Hence, $p_1(G_0)$ can be derived from the joint p.d.f. of β_x and β_n in (17), in a similar way as the derivation in [23] and using [55, (8-8)]. Finally, $p_1(G_0)$ is given by

$$p_1(G_0) = \frac{1}{2\pi\sigma_\beta^2\sqrt{1-\rho^2}} \int_{g=0}^{\infty} \frac{1}{(1-G_0)\, g} \exp\left(f_1(G_0, g)\right) dg, \tag{22}$$

if $G_0 \leq 1$, where $f_1(G_0, g)$ is

$$\begin{aligned} f_1(G_0, g) = &\left(\log g - \bar{\beta}\right)^2 + \left(\log(1-G_0) + \log g - \bar{\beta}\right)^2 \\ &-2\rho\left(\log g - \bar{\beta}\right)\left(\log(1-G_0) + \log g - \bar{\beta}\right), \end{aligned} \tag{23}$$

and $p_1(G_0) = 0$ if $G_0 > 1$.

A. Data Generation in the Noise-Only Case

Before studying the impact of the number of frames L_n to estimate the noise and the noisy SCMs, the number of sensors M and the SCMs-ICC ρ on the performance of the MWF, this part describes how to generate simulation data and how to obtain the simulation results to verify the theoretical results.

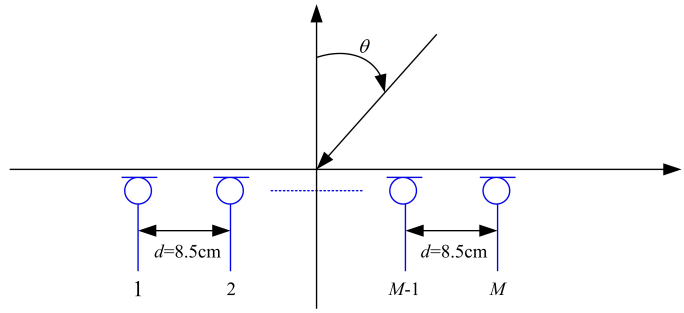

Fig. 3. Example of an M sensors broadside array with the distance of two adjacent sensors $d = 8.5$ cm, θ is the incident angle of a point source.

From (21), one can infer that the theoretical amount of noise reduction is independent of the frequency and it only depends on the number of frames L_n, the number of sensors M, and the SCMs-ICC ρ. Furthermore, the derivation of (21) does not account for the geometry of the sensors and the direction of arrival of the noise. To analytically validate the theoretical performance of the MWF, a linear array is chosen for simplicity and three frequencies are chosen to verify that the amount of noise reduction is independent of the frequency in the noise-only case.

We assume that a uniform broadside linear array (ULA) consists of M sensors and the distance of two adjacent sensors is 8.5 cm. The sampling frequency of each sensor is $f_s = 16$ kHz. The ULA is plotted in Fig. 3 and θ indicates the incident angle of a point source. Three frequency bins that correspond to 1000 Hz, 2000 Hz and 4000 Hz, respectively, are chosen to calculated the empirical amount of noise reduction. The three frequencies are chosen so that the distance of two adjacent sensors is equal to a quarter wavelength at 1000 Hz, a half wavelength at 2000 Hz, and a full wavelength at 4000 Hz. Note that similar results can be obtained by using a nonuniform linear array or an arbitrary array geometry. To support this hypothesis, Section IV will present experimental results using real-world data recorded by a circular array in a meeting room.

We further assume that there are J mutually independent interfering point sources in a plane, which also contains the sensor array, radiating from θ_j with $j = 1, \cdots, J$, where each is assumed to be a white Gaussian random process with zero mean and equal variance of $\sigma^2 = 1000$. The sensor noise is assumed to be a white Gaussian process with zero mean and unit variance. The number of interfering point sources in the far field J can be varied from 0 to any integer value and the direction of arrival $\theta_j \in [0^\circ\ 360^\circ]$ can also be arbitrary. In this section, we fix both J and θ_j, where $J = M$ and $\theta_j = j \times 10^\circ$ for $J < 36$. The simulation results are nearly the same by varying the values of J and θ_j. In all simulations, we only consider reverberation-free environments to validate the analytical results. Results for real-world recordings using a circular array as discussed in Section IV will further verify the analytical results.

Given the array geometry and the DOAs of the interferences, the simulated data for each sensor can be generated accordingly. With the simulated data, we can compute the estimated Wiener filter using (7). We emphasize that the

Wiener filter is estimated by (7) throughout this paper and (18) is only used to statistically analyze the theoretical performance of (7). In all simulations, the noisy SCM $\widehat{\mathbf{R}}_{xx}$ is estimated using (1) and (4), while the noise SCM $\widehat{\mathbf{R}}_{nn}$ is also estimated using a delayed version of the simulated data, where the number of relatively delayed samples is $L_n(1-\rho)R$. This ensures that a prescribed SCMs-ICC is realized in the simulations. It corresponds to the situations where there are $L_n(1-\rho)$ noise-only frames that are erroneously detected as $\mathcal{P}(k,l) = 1$ for every L_n noise-only frames. Note that $\rho = 1$ implies a perfect target activity detector in noise-only periods. For example, the noisy SCM $\widehat{\mathbf{R}}_{xx}(k,l)$ is computed using $\left[\begin{array}{ccc}\mathbf{X}(k,l-L_n+1) & \cdots & \mathbf{X}(k,l)\end{array}\right]$, while the noise SCM $\widehat{\mathbf{R}}_{nn}(k,l)$ is computed using $\left[\begin{array}{ccc}\mathbf{X}(k,l-2L_n+L_o+1) & \cdots & \mathbf{X}(k,l-L_n+L_o)\end{array}\right]$ with $L_o = \rho L_n = \rho L_x$. By comparing the input of the first sensor to the output of the MWF, the empirical amount of noise reduction can be obtained. Finally, the theoretical results according to (21) are compared to the empirical results, where a Monte Carlo simulation with 1000 trials is used to obtain each empirical result.

B. Noise Reduction versus the Number of Frames

Fig. 4 plots the theoretical amount of noise reduction $\mathrm{NR}_{\mathrm{MWF}}$ [dB] versus the number of frames L_n to estimate the SCM for the SCMs-ICC $\rho = 0$ and $\rho = 0.9$, where six sensors (i.e., $M = 6$) are considered with $J = 6$. Note that the simulation results fit very well the theoretical results, where the theoretical amount of noise reduction can be calculated by evaluating (21) numerically. As can be seen from Fig. 4(a), the amount of noise reduction is only about 10 dB even for a large value of L_n (for instance, $L_n = 100$) with $\rho = 0$. Comparing Fig. 4(b) with Fig. 4(a), the amount of noise reduction for $\rho = 0.9$ is about 10 dB higher than that for $\rho = 0$. That is to say, the amount of noise reduction of the MWF increases dramatically when increasing the SCMs-ICC ρ. Empirical noise reduction is nearly the same for the three frequencies, which demonstrates that the amount of noise reduction of the MWF is independent of the frequency in the noise-only case. When L_n is less than about 20 for the six-sensor case, the amount of noise reduction (in decibels) is negative, which means that the noise is not suppressed but, instead, is amplified. When the number of frames is not much larger than the number of sensors, the noise may be amplified by using the MWF. This poor noise reduction performance is due to the fact that no constraint is introduced in estimating the covariance matrix of the desired signal [32], where this phenomenon can also be well predicted by the proposed model in this paper. Such constraints could also be well predicted by the proposed model, which is left for future work.

C. Noise Reduction versus the Number of Sensors

Fig. 5 shows the theoretical noise reduction versus the number of sensors M for the SCMs-ICC $\rho = 0$ and $\rho = 0.9$, for a fixed number of frames, $L_n = 50$. The difference between the theoretical and the simulation results is less than 1 dB. It is interesting to see that the amount of noise reduction

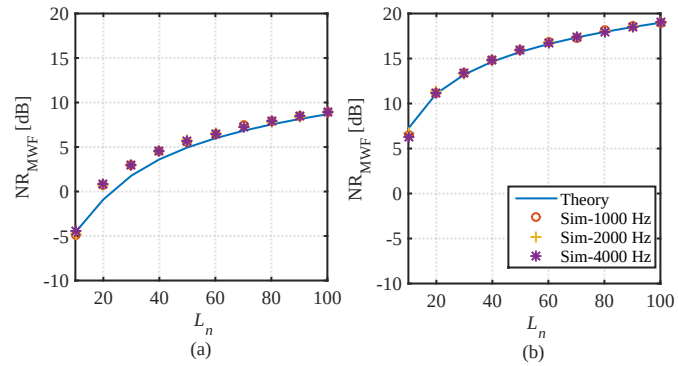

Fig. 4. Theoretical and empirical noise reduction $\mathrm{NR}_{\mathrm{MWF}}$ [dB] versus the number of frames L_n to estimate the noise and the noisy SCMs for the $M = 6$ with the SCMs-ICC (a) $\rho = 0$; (b) $\rho = 0.9$. Empirical and theoretical results at frequencies 1000 Hz, 2000 Hz and 4000 Hz are given.

reduces gradually as the number of sensors M increases during noise-only periods.

It is well-known that some MASE algorithms improve the amount of noise reduction as the number of sensors increases, such as the delay-and-sum beamformer (DSB) and the superdirective beamformer. However, in this part, we show that, for a fixed number of frames, the amount of noise reduction in the MWF decreases when the number of sensors M increases, where the main reason is that the amount of noise reduction is highly correlated to the natural distance from the noisy SCM $\widehat{\mathbf{R}}_{xx}$ to the noise SCM $\widehat{\mathbf{R}}_{nn}$ during noise-only periods. From (51), we see that both the noisy and the noise SCMs increase their MSE as the number of sensors M increases for a given number of frames L_n. This effect can also be interpreted as resulting from the strongly increasing variance of β in (16) with increasing the number of sensors. This results in increasing their natural distance statistically. One can also see this phenomenon in [18, Fig. 5], where a greater forgetting factor is required to achieve the same performance as the Wiener filter for noise reduction when using more sensors. Therefore, when using more sensors, we need more frames to estimate the noise and the noisy SCMs to ensure that the performance of the MWF does not degrade.

During noise-only periods, the results above suggest that the amount of noise reduction does not always increase with increasing the number of sensors when the number of available frames is limited. This conclusion can be especially supported for wide-sense stationary processes, where increasing the number of frames may be another option to reduce the natural distance of the covariance matrix estimate. For non-stationary processes, however, it seems often useful to increase the number of sensors to make up for the lack of a long time interval for estimating the covariance matrices, because the target activity detector may work better by using more sensors. If the target activity detector becomes more accurate, the noise covariance matrix can be updated more quickly, which results in improving the noise reduction performance. In the following part, we will show the importance of updating the noise covariance matrix continuously in both theory and simulation.

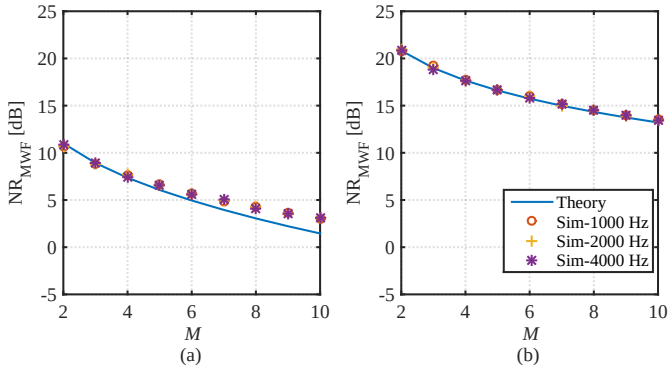

Fig. 5. Theoretical and empirical noise reduction $\mathrm{NR}_{\mathrm{MWF}}$ [dB] versus the number of sensors M for a fixed number of frames ($L_n = 50$) with the SCMs-ICC (a) $\rho = 0$; (b) $\rho = 0.9$. Empirical and theoretical results at frequencies 1000 Hz, 2000 Hz and 4000 Hz are given.

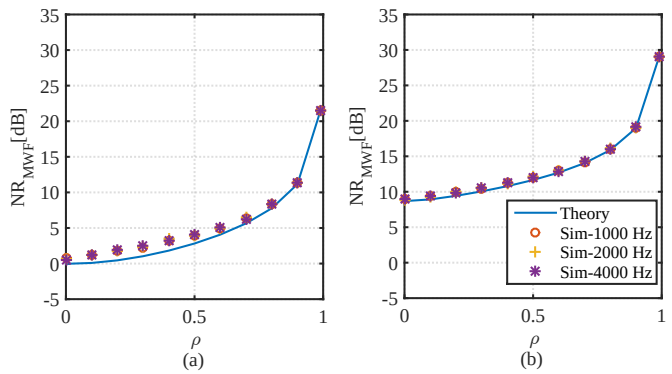

Fig. 6. Theoretical and empirical noise reduction $\mathrm{NR}_{\mathrm{MWF}}$ [dB] versus the SCMs-ICC ρ for $M = 6$ with the number of independent frames (a) $L_n = 20$; (b) $L_n = 100$. Empirical and theoretical results at frequencies 1000 Hz, 2000 Hz and 4000 Hz are given.

D. Noise Reduction versus the SCMs-ICC

Fig. 6 shows both the theoretical and the empirical noise reduction of the MWF versus the SCMs-ICC ρ with the number of frames $L_n = 20$ and $L_n = 100$ for the six-sensor case ($M = 6$). Note that the amount of noise reduction becomes infinite if $\rho = 1$, so we only plot $\rho = 0.99$ instead of $\rho = 1$ in order to give a quantitative value in this figure. Given the number of frames L_n, the amount of noise reduction increases as the SCMs-ICC ρ increases. As an example, the amount of noise reduction increases from about 0 dB to 11 dB when ρ increases from 0 to 0.9 for $L_n = 20$. Comparing Fig. 6(b) with Fig. 6(a), the amount of noise reduction is about 10 dB higher than that for $L_n = 20$.

In practice, if both the number of frames L_n and the number of sensors M are fixed, there are two ways of increasing the amount of noise reduction. First, it is better to update the noise SCM continuously, and thus ensure that the SCMs-ICC ρ is as close as possible to one during noise-only periods. Second, we should consider a good trade-off between frequency resolution and time resolution, since the number of frames can be potentially increased if sacrificing the frequency resolution using the same time interval of input data. For instance, one can consider to apply Bartlett's method for this purpose [18].

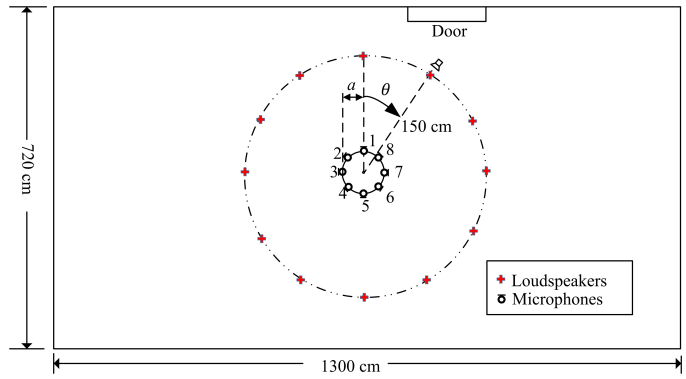

Fig. 7. Experiment setup. A circular array with eight microphones is used to record data in a meeting room. Microphones positions and loudspeakers positions are represented by ○ with digits and +, respectively.

IV. Validation using Real-World Recordings for Noise-Only Periods

The theoretical analysis results, presented in Section III, are only verified by Monte Carlo simulation results. In this section, experimental results using real-world recordings are given to further validate the theoretical analysis results for completeness. Fig. 7 describes the array configuration and the loudspeaker positions. The sound sources are positioned at twelve angles with 30° interval around the circular array and thus we obtain twelve audio files that are sampled at 16 kHz with eight-channel signals. To be consistent with the hypothesis in Section I, the sound sources used for each recording are white Gaussian processes with zero mean and unit variance. All the data are recorded in a meeting room with a reverberation time of 0.7 second, where the length, the width and the height of this meeting room are 13.0 m, 7.2 m and 2.3 m, respectively. The circular array is placed on a table in the center of the meeting room with the height 1.5 m and the loudspeaker is placed at the same height of the array. This section concentrates on results which are not so intuitive to validate the theoretical analysis results using real-world data. For example, in noise-only periods, the amount of noise reduction is independent of frequency and it reduces when increasing the number of sensors. Since both ambient noise and microphone self-noise are inevitable in practice, we do not need to add sensor noise as we did in the Monte Carlo simulations [8].

A. Noise Reduction versus Frequency

In the first experiment, all eight microphones are used. The microphone signals are obtained by adding five audio files together, where these files are recorded separately by placing five uncorrelated sound sources at 30°, 60°, 90°, 120° and 150°. Using the eight-microphone signals, we can perform the MWF using (1), (3), (4) and (7) to compute the experimental results of the amount of noise reduction. Fig. 8 plots the amount of noise reduction versus frequency for different values L_n with the SCMs-ICC $\rho = 0$ and $\rho = 0.9$. One can see that there are only slight differences between the experimental results and the theoretical results and also

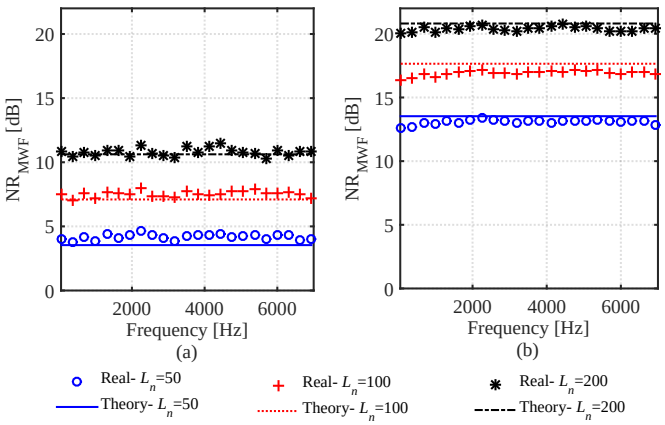

Fig. 8. Theoretical and experimental noise reduction $\mathrm{NR}_{\mathrm{MWF}}$ [dB] versus frequency for $M = 8$ and different values of L_n, i.e., 50, 100, 200, with the SCMs-ICC (a) $\rho = 0$; (b) $\rho = 0.9$.

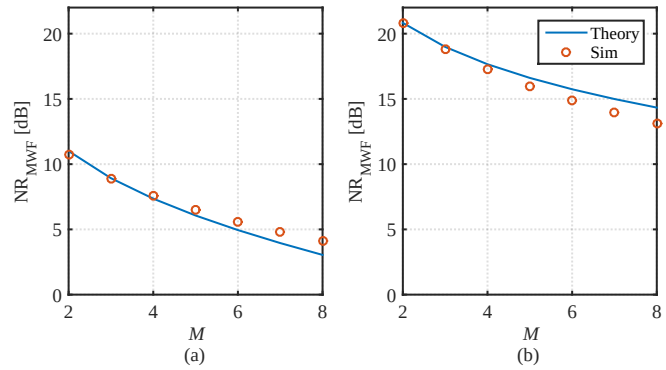

Fig. 9. Theoretical and experimental noise reduction $\mathrm{NR}_{\mathrm{MWF}}$ [dB] versus the number of sensors M for a fixed number of frames ($L_n = 50$) with the SCMs-ICC (a) $\rho = 0$; (b) $\rho = 0.9$.

the amount of noise reduction increases when increasing the number of frames and the SCMs-ICC. Second, the amount of noise reduction in noise-only periods is nearly a constant value over frequency, which means that it is independent of frequency. Finally, this experiment also reveals that the amount of noise reduction does not depend on the array geometry and the noise scenarios in noise-only periods. These phenomena can be well explained by our theoretical analysis results, where the amount of noise reduction of the MWF only depends on the number of smoothing frames, the number of sensors and the SCMs-ICC in noise-only periods.

B. Noise Reduction versus the Number of Sensors

In the second experiment, two to eight microphones are used to show the impact of the number of sensors on the amount of noise reduction. The microphone signals are obtained by adding five audio files together, where these files are recorded separately by locating five uncorrelated sound sources at 30°, 90°, 120°, 180° and 240°. Fig. 9 plots the amount of noise reduction versus the number of sensors M for the number of smoothing frames $L_n = 50$ with SCMs-ICC $\rho = 0$ and $\rho = 0.9$. The experimental results are computed by averaging the amount of noise reduction over frequency, which is due to that the noise reduction is independent of frequency having the same value in both theory and experiment (see Fig. 8). This figure shows that the theoretical results fit well with the experimental results. Both Fig. 9 and Fig. 5 indicate that the amount of noise reduction decreases when increasing the number of sensors in noise-only periods.

V. Statistical Analysis of the MWF during Noisy Periods

The following three assumptions are made when studying theoretical limits of the MWF during noisy periods:

(i) There is only one desired signal.
(ii) The noise covariance matrix is only updated during noise-only period (a perfect target activity detector is assumed[4]).
(iii) The number of frames L_n is large enough so that the cross-covariance of the desired signal vector and the noise vector can be neglected.

The first assumption (i) is commonly used in analyzing the MWF (see [8] and [19] for details), and facilitates the analysis of the PTF of the desired signal. The second assumption (ii) is also reasonable, because it is well-known that the MWF needs a robust target activity detection scheme to avoid serious desired signal distortion [8], [19], [28]. The leakage of the target signal into the noise SCM is left for future research. The third assumption (iii) means that the theoretical analysis herein is only asymptotic. That is to say, the theoretical analysis fits better for larger values of L_n.

Based on the assumption (i), the true desired signal covariance matrix is given by

$$\mathbf{R}_{ss} = P_{ss}\mathbf{H}\mathbf{H}^H, \tag{24}$$

where P_{ss} is the desired signal power spectral density. $\mathbf{H} = [H_1 \;\; H_2 \;\; \cdots H_M]^T$ and H_i is the acoustic transfer function (ATF) between the desired signal and the ith sensor.

We assume that the noise SCM $\widehat{\mathbf{R}}_{nn}$ is estimated during noise-only periods according to assumption (ii). $\widehat{\mathbf{R}}_{xx}$ is the noisy SCM, given by

$$\widehat{\mathbf{R}}_{xx} = \mathbf{R}_{ss} + \widehat{\mathbf{R}}_{n_x n_x}, \tag{25}$$

where $\widehat{\mathbf{R}}_{n_x n_x}$ is the noise SCM computed from the noisy input data directly. In general, $\widehat{\mathbf{R}}_{n_x n_x} \neq \widehat{\mathbf{R}}_{nn}$ holds during noisy periods because $\widehat{\mathbf{R}}_{nn}$ stops updating when the desired signal is present. Note that $\widehat{\mathbf{R}}_{xx}$ in (25) neglects the cross-covariance matrix of the desired signal vector and the noise vector after considering the assumption (iii).

In practical implementations, both P_{ss} and $\mathbf{H}$ are unknown a priori, so (24) and (25) are only introduced to analyze the performance of the MWF in (7). Accordingly, the estimated

[4]Here, "perfect target activity detector" means that the noisy frames are all correctly detected ($\mathcal{P}(k,l) = 1$). Noise-only periods are still allowed to be erroneously detected ($\mathcal{P}(k,l) = 1$). Under this assumption, the leakage of the target signal into the estimated noise covariance matrix will not occur.

Wiener filter can be written as

$$\widehat{\mathbf{w}}_{\text{opt}} = \left(P_{ss}\mathbf{H}\mathbf{H}^H + \widehat{\mathbf{R}}_{n_x n_x}\right)^{-1} \times \left(P_{ss}\mathbf{H}\mathbf{H}^H + \widehat{\mathbf{R}}_{n_x n_x} - \widehat{\mathbf{R}}_{nn}\right)\mathbf{e}_1. \quad (26)$$

As pointed out before, it is difficult to study the behavior of the MWF by using $\widehat{\mathbf{w}}_{\text{opt}}$ directly. In Section III, the bivariate model for sample covariance matrices was introduced to approximately model $\widehat{\mathbf{R}}_{nn}$ and $\widehat{\mathbf{R}}_{n_x n_x}$. This yields

$$\widetilde{\mathbf{w}}_{\text{opt}} = \left(\phi_{ss}\mathbf{H}\mathbf{H}^H + \mathbf{R}_{nn}\right)^{-1}\left(\phi_{ss}\mathbf{H}\mathbf{H}^H + (1-r)\,\mathbf{R}_{nn}\right)\mathbf{e}_1, \quad (27)$$

where $\phi_{ss} = P_{ss}/\exp(\beta_{n_x})$ and $r = \exp(\beta_n)/\exp(\beta_{n_x})$. (27) can be derived due to that $\widehat{\mathbf{R}}_{nn}$ and $\widehat{\mathbf{R}}_{n_x n_x}$ are, respectively, approximated by $\widetilde{\mathbf{R}}_{nn} = \exp(\beta_n)\mathbf{R}_{nn}$ and $\widetilde{\mathbf{R}}_{n_x n_x} = \exp(\beta_{n_x})\mathbf{R}_{nn}$ as proposed in (9). This section uses $\widetilde{\mathbf{w}}_{\text{opt}}$ instead of $\widehat{\mathbf{w}}_{\text{opt}}$ to study the performance of the MWF during noisy periods in a statistical way. The joint p.d.f. of β_n and β_{n_x} is defined in (17). By using the Sherman-Morrison-Woodbury formula [54], (27) can be rewritten as

$$\widetilde{\mathbf{w}}_{\text{opt}} = \widetilde{\mathbf{w}}_{\text{opt},s} + \widetilde{\mathbf{w}}_{\text{opt},n}, \quad (28)$$

where $\widetilde{\mathbf{w}}_{\text{opt},s}$ and $\widetilde{\mathbf{w}}_{\text{opt},n}$ are, respectively, given by

$$\widetilde{\mathbf{w}}_{\text{opt},s} = \left[\frac{\phi_{ss}}{\phi_{ss} + P_{nn}^{\text{MVDR}}}\right]\frac{\mathbf{R}_{nn}^{-1}\mathbf{H}}{\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}}H_1^*, \quad (29)$$

and

$$\widetilde{\mathbf{w}}_{\text{opt},n} = (1-r)\,\mathbf{e}_1 - (1-r)\,\widetilde{\mathbf{w}}_{\text{opt},s}, \quad (30)$$

where $P_{nn}^{\text{MVDR}} = \left(\mathbf{H}^H\mathbf{R}_{nn}^{-1}\mathbf{H}\right)^{-1}$ is the noise PSD of the MVDR output under ideal conditions, i.e., both $\mathbf{H}$ and $\mathbf{R}_{nn}$ are accurately known. We refer to the MVDR with the true covariance matrix $\mathbf{R}_{nn}$ as TCM-MVDR.

After assuming that the noise PSD at each sensor is the same, we can define the input SNR in the first sensor as

$$\xi_{\text{in}} = \frac{P_{ss}|H_1|^2}{P_{nn}}, \quad (31)$$

where P_{nn} is the noise PSD. The SNR improvement of the TCM-MVDR is then given by

$$\xi_{\text{MVDR}} = \frac{P_{nn}}{P_{nn}^{\text{MVDR}}}. \quad (32)$$

With the help of (31) and (32), the output SNR of the TCM-MVDR is given by

$$\xi_{\text{out}}^{\text{MVDR}} = \xi_{\text{in}}\xi_{\text{MVDR}}, \quad (33)$$

and we further define $\xi = \xi_{\text{in}}\xi_{\text{MVDR}}/|H_1|^2$ to describe the the SNR relative to the desired source power.

The following three parts apply (28)-(30) to study the theoretical performances of the MWF in terms of the SNR improvement.

A. PTFs of the Desired Signal and the Noise

To study the SNR improvement of the MWF, PTFs of the desired signal and the noise need to be studied beforehand according to [8]. By using $\widetilde{\mathbf{w}}_{\text{opt}}$ in (28)-(30), the PTF of the desired signal from the first sensor to the output of the MWF during noisy periods is given by

$$G_s^2 = \frac{\widetilde{\mathbf{w}}_{\text{opt}}^H\mathbf{R}_{ss}\widetilde{\mathbf{w}}_{\text{opt}}}{\mathbf{e}_1^H\mathbf{R}_{ss}\mathbf{e}_1} = \left(r\left(\frac{\xi}{\xi + \exp(\beta_{n_x})}\right) + (1-r)\right)^2, \quad (34)$$

and the PTF of the noise is given by

$$G_n^2 = \frac{\widetilde{\mathbf{w}}_{\text{opt}}^H\mathbf{R}_{nn}\widetilde{\mathbf{w}}_{\text{opt}}}{\mathbf{e}_1^H\mathbf{R}_{nn}\mathbf{e}_1} = r^2\left(\frac{\xi}{\xi + \exp(\beta_{n_x})}\right)^2\frac{|H_1|^2}{\xi_{\text{MVDR}}} + (1-r)^2 + 2r(1-r)\left(\frac{\xi}{\xi + \exp(\beta_{n_x})}\right)\frac{|H_1|^2}{\xi_{\text{MVDR}}}. \quad (35)$$

If $\beta_{n_x} = \beta_n = 0$ and thus $r = 1$, (34) and (35) reduce to

$$G_s^2 = \left(\frac{\xi}{\xi + 1}\right)^2, \quad (36)$$

and

$$G_n^2 = \left(\frac{\xi}{\xi + 1}\right)^2\frac{|H_1|^2}{\xi_{\text{MVDR}}}, \quad (37)$$

respectively, where (36) and (37) are identical to [8, (43) and (44)]. It should be emphasized that $\beta_{n_x} = \beta_n \to 0$ holds when L approaches infinity for a given M, which can be deduced from (15) and (16), because both $E\{\beta\}$ and $\text{var}\{\beta\}$ equal zero in this case (see Fig. 2 for details). In [8], (36) and (37) are derived from a deterministic signal model. Using (34)-(35) instead of (36)-(37), we can study the impact of the noise SCM on the desired signal distortion under the above stochastic signal model.

We emphasize the following two extreme cases before studying the SNR improvement of the MWF:

(1) When the input SNR ξ_{in} is 0, (34) and (35) are identical and they all reduce to the noise-only case in (18).
(2) For $\xi \gg 1$, G_s^2 is close to 1. In other words, the desired signal will not be much suppressed when ξ is much larger than 1, i.e., when the TCM-MVDR performs well.

B. SNR Improvement

According to (36) and (37), we can derive the SNR improvement for the deterministic signal model as

$$\xi_{\text{imp}} = \frac{\xi_{\text{MVDR}}}{|H_1|^2} = \frac{P_{nn}}{P_{nn}^{\text{MVDR}}|H_1|^2}. \quad (38)$$

The SNR improvement for the stochastic signal model is given by

$$\widehat{\xi}_{\text{imp}} = E\left\{G_s^2\right\}/E\left\{G_n^2\right\}. \quad (39)$$

By substituting $r = \exp(\beta_n)/\exp(\beta_{n_x})$ into (34), we can rewrite (34) as

$$G_s^2 = \left(1 - \frac{\exp(\beta_n)}{\xi + \exp(\beta_{n_x})}\right)^2. \quad (40)$$

If we further define $\beta_1 = \exp(\beta_n)/(\xi + \exp(\beta_{n_x}))$ and $\beta_2 = \exp(\beta_{n_x})$, the p.d.f. of β_1 can be derived from (17)

directly, and is given by

$$p_2(\beta_1) = \frac{1}{2\pi\sigma_\beta^2\sqrt{1-\rho^2}} \times \int_0^\infty \frac{1}{\beta_1\beta_2}\exp\left(f_2(\beta_1,\beta_2)\right)d\beta_2, \tag{41}$$

where $|\rho| < 1$ and

$$\begin{aligned} f_2(\beta_1,\beta_2) = -\tfrac{1}{2(1-\rho^2)\sigma_\beta^2}\Big(&\left(\log\left((\xi+\beta_2)\beta_1\right)-\bar{\beta}\right)^2 \\ &-2\rho\left(\log\left((\xi+\beta_2)\beta_1\right)-\bar{\beta}\right)\left(\log(\beta_2)-\bar{\beta}\right) \\ &+\left(\log(\beta_2)-\bar{\beta}\right)^2\Big). \end{aligned} \tag{42}$$

With (41) and (42), $E\left\{G_s^2\right\}$ is given by

$$E\left\{G_s^2\right\} = \int_0^\infty (1-\beta_1)^2 p_2(\beta_1)\,d\beta_1. \tag{43}$$

For $|\rho| = 1$, $E\left\{G_s^2\right\} = \left(\xi/\left(\xi+\exp(\bar{\beta})\right)\right)^2$ holds due to $p_2(\beta_1) = \delta(\beta_1 - \exp(\bar{\beta})/\left(\xi+\exp(\bar{\beta})\right))$.

For $\xi \gg 1$, (35) can be approximated by

$$G_n^2 \approx (1-G_0)(1+G_0)/\xi_{\text{imp}} + G_0^2, \tag{44}$$

where G_0 is defined in (19). $E\{G_n^2\}$ is given by

$$E\left\{G_n^2\right\} = \int_{-\infty}^1 \left(\frac{(1-G_0)(1+G_0)}{\xi_{\text{imp}}} + G_0^2\right) p_1(G_0)\,dG_0, \tag{45}$$

where $p_1(G_0)$ is given by (22). For $\xi \ll 1$, $E\left\{G_n^2\right\} = E\left\{G_0^2\right\}$ holds, which is given by (21).

For the special case, $|\rho| = 1$, $E\left\{G_n^2\right\} = 1/\xi_{\text{imp}}$ holds due to $p_1(G_0) = \delta(G_0)$.

C. Simulation Validation

In this part, we show the validity of (39) by using Monte Carlo simulations. In all simulations, we use the same array configuration in Fig. 3. The sensor noise is assumed to be a white Gaussian random process with zero mean and unit variance. The desired signal is assumed to radiate from 0°. Five interferences are assumed to be localized at 15°, 25°, 40°, 60° and 80°, respectively. Both the desired signal and the five interferences are assumed to be independent white Gaussian random processes with zero means and equal variances of $\sigma^2 = 1000$. The sampling frequency of each sensor is $f_s = 16$ kHz and the frequency bins that correspond to 1000 Hz, 2000 Hz and 4000 Hz are chosen to calculate the empirical results. From (39), one can see that the SNR improvement of the MWF not only depends on the number of frames L_n, the number of sensors M, and the SCMs-ICC ρ, but of course also depends on the SNR improvement of the TCM-MVDR. The noise SCM and the desired signal SCM are computed separately and subsequently summed to obtain the noisy SCM for the MWF. Thus, we can ignore the impact of the cross-covariance matrix of the noise vector and the desired signal vector even when the number of frames L_n is small. Note that if the noisy SCM is computed from the noisy vector directly, L_n should be large enough to make sure that the assumption (iii) holds. According to the configuration, the input SNR is $\xi_{\text{in}} = 0.2$ (-6.99 dB). Without loss of generality, the ATF between the desired signal and the first sensor $|H_1|$ is set to 1. The SNR improvement of the TCM-MVDR is shown in Fig. 10. Note that the SNR improvement of the TCM-MVDR varies with the frequency, as it is known for the behaviour of a superdirective beamformer and a delay-and-sum beamformer [1]. Under ideal conditions, i.e., the noise covariance matrix and the ATFs are accurately known, the TCM-MVDR improves its performance when increasing the number of sensors, but the behaviour of the MWF under stochastic models is quite different from the TCM-MVDR. For a frequency of 2000 Hz, the TCM-MVDR has the best performance for $M > 3$, because the distance of two adjacent sensors in this simulation is equal to half the wavelength.

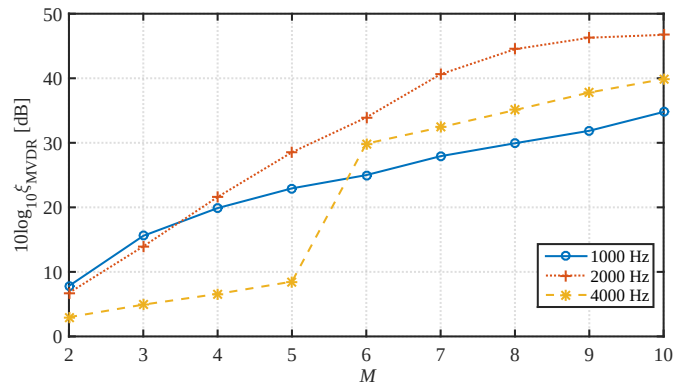

Fig. 10. The SNR improvement of the TCM-MVDR ξ_{MVDR} at 1000 Hz, 2000 Hz and 4000 Hz versus the number of sensors M for five mutually independent interfering point sources in the deterministic case, where $\xi_{\text{MVDR}} = \xi_{\text{imp}}$ and $\xi = 0.2\xi_{\text{imp}}$ hold due to $\xi_{\text{in}} = 0.2$ in this example.

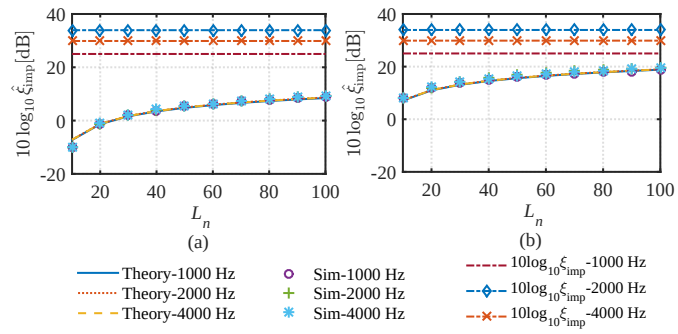

Fig. 11. The SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ (in decibels) at 1000 Hz, 2000 Hz and 4000 Hz versus the number of frames L_n for $M = 6$ with the SCMs-ICC (a) $\rho = 0$ and (b) $\rho = 0.9$.

1) SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ versus the number of frames L_n: We plot $\widehat{\xi}_{\text{imp}}$ versus L_n for the six-sensor case (i.e., $M = 6$) with the SCMs-ICC $\rho = 0$ and $\rho = 0.9$ in Fig. 11. As can be seen from this figure, $\widehat{\xi}_{\text{imp}}$ increases with increasing L_n. Comparing Fig. 11(b) with Fig. 11(a), it is obvious that $\widehat{\xi}_{\text{imp}}$ for $\rho = 0.9$ is much larger than that for $\rho = 0$. Comparing the theoretical results with the simulation results, we see that they match very well. For comparison, the SNR improvement of the MWF under the deterministic signal model ξ_{imp} is also plotted in Fig. 11. As mentioned above, when L_n becomes infinity for a given M, $\widehat{\xi}_{\text{imp}}$ will be

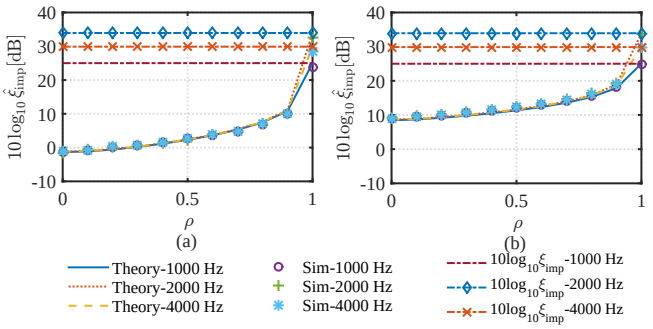

Fig. 12. The SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ (in decibels) at 1000 Hz, 2000 Hz and 4000 Hz versus the SCMs-ICC ρ for $M = 6$ with the number of frames (a) $L_n = 20$ and (b) $L_n = 100$.

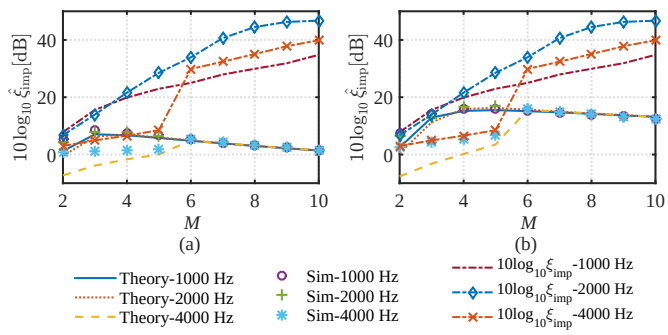

Fig. 13. The SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ (in decibels) at 1000 Hz, 2000 Hz and 4000 Hz versus the number of sensor M for five interferences with the number of frames $L_n = 50$ and the SCMs-ICC (a) $\rho = 0$ and (b) $\rho = 0.9$.

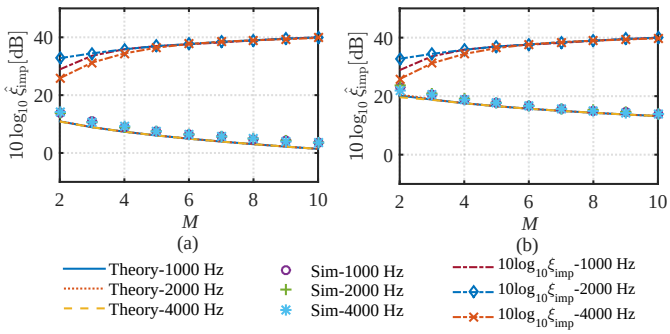

Fig. 14. The SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ (in decibels) versus the number of sensor M for only one interference with the number of frames $L_n = 50$ and the SCMs-ICC (a) $\rho = 0$ and (b) $\rho = 0.9$.

identical to ξ_{imp}. In other words, ξ_{imp} is the upper limit of $\widehat{\xi}_{\text{imp}}$. For completeness, ξ_{imp} is also plotted in Figs. 12-14.

2) SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ versus the SCMs-ICC ρ: We further study the impact of ρ on $\widehat{\xi}_{\text{imp}}$ in Fig. 12, where both the simulation results and the theoretical results show that $\widehat{\xi}_{\text{imp}}$ is significantly improved when ρ increases from 0 to 1. In practice, if we update the noise covariance matrix continuously, $\widehat{\xi}_{\text{imp}}$ can still be a large value that is close to ξ_{imp} at the target signal onsets. $\widehat{\xi}_{\text{imp}}$ reduces when the duration of the target signal is extremely long because of $\rho \to 0$. Comparing Fig. 12(b) with Fig. 12(a), we find that $\widehat{\xi}_{\text{imp}}$ for $L_n = 100$ is much larger than that for $L_n = 20$. Fig. 12 further confirms the validity of the theoretical analysis. For $\rho = 1$, the deterministic values are reached regardless of L_n.

3) SNR improvement of the MWF $\widehat{\xi}_{\text{imp}}$ versus the number of sensors M: Fig. 13 plots $\widehat{\xi}_{\text{imp}}$ versus M with a fixed number of frames $L_n = 50$ for $\rho = 0$ and $\rho = 0.9$. It is interesting to see that $\widehat{\xi}_{\text{imp}}$ increases when M increases from 2 to about 4, and then it reduces gradually when M is larger than 4. Fig. 13 also reveals that using more sensors does not always improve performance of the MWF under a stochastic signal model for a fixed number of frames. The relationship between the SNR improvement of the MWF and the number of sensors is complicated (see Eq. (39)), and it depends on the number of frames L_n, the SCMs-ICC ρ and the noise scenarios. As we know, the SNR improvement of the MWF increases with the number of sensors when the number of frames becomes infinite (i.e., $L_n \to \infty$), since $\widehat{\xi}_{\text{imp}} \to \xi_{\text{MVDR}}$ is true in this extreme case. Note that the single-channel postfilter contained in the MWF does not improve the frequency-domain subband output SNR, which is also implied already in [2, (3.10)]. At 4000 Hz, the simulation results do not fit very well with the theoretical results when the number of sensors is fewer than 5. This is due to that ξ is not much larger than 1 and thus (44) is not accurate enough. When ξ is much larger than 1, the theoretical results can well match the simulation results.

Now only one interference with zero mean and the variance $\sigma^2 = 1000$ is considered, which is localized at 60°. In this example, we have $\xi_{\text{in}} = 1$ (0 dB). Fig. 14 plots the SNR improvement of the TCM-MVDR ξ_{MVDR} versus the number of sensors M, note that $\xi_{\text{imp}} = \xi_{\text{MVDR}}$ holds. Comparing Fig. 14 with Fig. 10, the SNR improvement of the TCM-MVDR ξ_{MVDR} for only one interference is much higher than that for the case of five interferences, especially when the number of sensors M is small. Fig. 14 also plots $\widehat{\xi}_{\text{imp}}$ versus M using L_n=50 frames and the SCMs-ICC $\rho = 0$ and $\rho = 0.9$ in the scenario with one interferer. Both the theoretical results and the simulation results confirm that $M = 2$ achieves the largest SNR improvement with the MWF filter. Under the stochastic signal model, if the SNR improvement of the TCM-MVDR ξ_{MVDR} using more sensors is not much higher than that using fewer sensors, it is better to use fewer sensors when the number of available frames is limited in practical applications.

D. Discussion

The experimental results using real-world recordings in noisy periods will not be presented in this paper since not only the ATFs between the desired signal and the microphones but also the true noise covariance matrix $\mathbf{R}_{nn}$ are unknown beforehand. Hence, the SNR improvement of TCM-MVDR is unknown and the theoretical results cannot be analytically calculated. Instead, we provide come comments based on the results of the Monte Carlo simulation, shown in Figs. 10-14.

As opposed to the deterministic signal model, it is not always true that the SNR improvement of the MWF filter increases with the number of sensors under the stochastic signal model. If the number of available frames to estimate the

SCM is small and the SCMs-ICC ρ is close to 0, it is better to use fewer sensors. Unfortunately, we cannot predict the SNR improvement of the TCM-MVDR ξ_{MVDR} beforehand, so it is difficult to determine the number of sensors that can achieve the highest SNR improvement with the MWF. For practical applications, for maximum SNR, we should use as many frames as the stationarity properties of the involved signals allow. Moreover, we should update the noise covariance matrix estimation as often as possible to reduce the differences between the noise covariance matrix estimated during noise-only periods and that estimated during noisy periods.

This paper only analyzes the statistical performance of the MWF under the Gaussian assumption, which means that the periodogram of the noise is χ^2 distributed with only two degrees of freedom. In practice, the noise may contain some deterministic components, which makes the periodogram follow a noncentral χ^2 distribution. In this case, the proposed statistical analysis method can also be applied, where the only difference is that the noncentral χ^2 distribution should be approximated by a χ^2 distribution with more degrees of freedom [44], [56]. That is to say, it is approximately equivalent to using a larger number of frames to estimate the SCM under a combined stochastic-deterministic signal model.

In practice, the number of available frames to estimate the noisy SCM and that to estimate the noise SCM may be significantly different. Generally, the latter could be much larger than the former to reduce the estimation error of the noise SCM for stationary noise and also to avoid oversmoothing the noisy SCM for nonstationary desired signals. In this case, the SCMs-ICC ρ could be close to 0, the two random variables β_x and β_n have different mean values and different variance values, and both can be computed by (15) and (16). This paper has shown that a SCMs-ICC ρ of 0 is always the worst case for both noise reduction and SNR improvement of the MWF. So we cannot expect that using more frames to estimate the noise SCM than to estimate the noisy SCM can achieve better performances, especially when the noise SCM can be updated continuously and the noise is also extremely nonstationary. According to the theoretical results, the number of frames L_n should be larger than about thrice the number of sensors M, i.e. $L_n > 3M$, to ensure that the noise can be reduced, even for the worst case.

VI. Conclusions

This paper proposes a novel statistical analysis method to study theoretical limits of the MWF under a stochastic signal model. Compared with traditional theoretical analysis on the MWF, this paper quantitatively shows how the parameters can be chosen to influence the performance of the MWF. Future work will concentrate on relaxing some of the assumptions made in the analysis of noisy periods.

Acknowledgement

The authors would like to thank the four anonymous reviewers and the associate editor for their valuable comments that helped to improve the quality of this work.

Appendix A
Derivation of the mean of β

By substituting $\widetilde{\mathbf{R}} = \exp(\beta)\mathbf{R}$ into (13), we get

$$\mathbf{B}(\mathbf{R}) = E\left\{\exp_{\mathbf{R}}^{-1}\widetilde{\mathbf{R}}\right\} = E\{\beta\}\mathbf{R}, \tag{46}$$

because $\exp_{\mathbf{R}}^{-1}\widetilde{\mathbf{R}} = \mathbf{R}^{1/2}\left(\log \mathbf{R}^{-1/2}\widetilde{\mathbf{R}}\mathbf{R}^{-1/2}\right)\mathbf{R}^{1/2} = \beta\mathbf{R}$.

From (13), we also have

$$\begin{aligned}\mathbf{B}(\mathbf{R}) &= E\left\{\exp_{\mathbf{R}}^{-1}\widehat{\mathbf{R}}\right\}\\ &= \frac{1}{M}(-M\log L + \sum_{i=1}^{M}\psi(L-i+1))\mathbf{R},\end{aligned} \tag{47}$$

which has already been derived in [37] and $\psi(\bullet)$ is the Digamma function [50]. With (46) and (47), $E\{\beta\}$ can be easily derived, which is presented in (15).

Appendix B
Derivation of the variance of β

The variance of β is the expected value of the squared derivation from the mean of β, $E\{\beta\}$, which is given by:

$$\begin{aligned}\mathrm{var}\{\beta\} &= E\left\{(\beta - E\{\beta\})^2\right\}\\ &= E\left\{\beta^2\right\} - (E\{\beta\})^2,\end{aligned} \tag{48}$$

where $E\{\beta\}$ is the mean of β in (15). To derive the variance of β, we only need to derive $E\left\{\beta^2\right\}$. With the help of (10), the MSE of $\widetilde{\mathbf{R}}$ in (14) can be further written as

$$\begin{aligned}\varepsilon_{\mathrm{cov}}^2 &= E\left\{d_{\mathrm{cov}}^2\left(\widetilde{\mathbf{R}},\mathbf{R}\right)\right\}\\ &= E\left\{d_{\mathrm{cov}}^2\left(\exp(\beta)\mathbf{R},\mathbf{R}\right)\right\} = ME\left\{\beta^2\right\},\end{aligned} \tag{49}$$

where (49) can be derived due to $\lambda_1 = \cdots = \lambda_M = \exp(\beta)$.

Consider the limit of large L and M with $M/L \le 1$. The p.d.f. of the eigenvalue of $\widehat{\mathbf{R}}_0 = \mathbf{R}^{-1/2}\widehat{\mathbf{R}}\mathbf{R}^{-1/2}$ [51, (1.2)], λ, is given by

$$\begin{aligned}p(\lambda) = &\frac{1}{2\pi\lambda(M/L)}\times\\ &\sqrt{\left(\left(1+\sqrt{M/L}\right)^2 - \lambda\right)\left(\lambda - \left(1-\sqrt{M/L}\right)^2\right)},\end{aligned} \tag{50}$$

where $\lambda \in \left(\left(1-\sqrt{M/L}\right)^2, \left(1+\sqrt{M/L}\right)^2\right)$; otherwise, $p(\lambda) = 0$.

By using (50), the mean square error $\varepsilon_{\mathrm{cov}}^2$ is given by [37, (109)] for large L and M, which is

$$M^{-1}\varepsilon_{\mathrm{cov}}^2 = M/L + 5(M/L)^2\Big/6 + \mathcal{O}\left((M/L)^3\right). \tag{51}$$

By substituting (49) into (48), (48) is written as

$$\mathrm{var}\{\beta\} = E\left\{\beta^2\right\} - (E\{\beta\})^2 = M^{-1}\varepsilon_{\mathrm{cov}}^2 - (E\{\beta\})^2. \tag{52}$$

By substituting (51) into (52), $\mathrm{var}\{\beta\}$ is given by:

$$\begin{aligned}\mathrm{var}\{\beta\} &= E\left\{\beta^2\right\} - (E\{\beta\})^2,\\ &= M^{-1}\varepsilon_{\mathrm{cov}}^2 - (E\{\beta\})^2,\\ &\approx M/L + 5(M/L)^2\Big/6 - (E\{\beta\})^2.\end{aligned} \tag{53}$$

APPENDIX C
A SHORT PROOF OF $|\rho| \approx L_o/L_x = L_o/L_n$

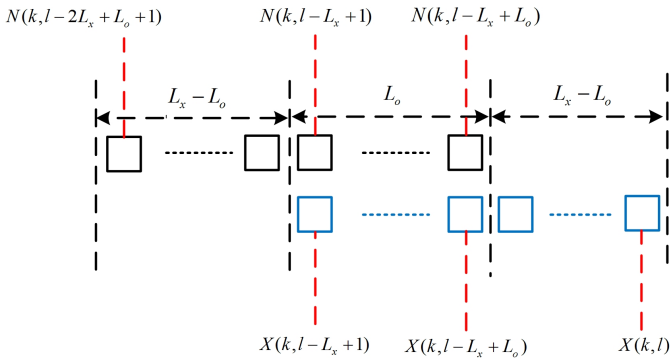

Fig. 15. Correlation of two Gaussian processes with the number of overlapping frames L_o.

We first assume that $N(k,l)$ and $X(k,l)$ are the STFTs of two Gaussian processes. For noise-only segments, we have $X(k,l) = N(k,l)$ and both the real part and the imaginary part of $X(k,l)$ follow Gaussian distributions. When the frame shift is equal to the frame length ($R = K$), each frame is statistically independent of the others, given by

$$E\left\{X\left(k,l_1\right)X^*\left(k,l_2\right)\right\} = \begin{cases} \sigma_x^2, & \text{if } l_1 = l_2 \\ 0, & \text{otherwise} \end{cases} \tag{54}$$

where $\sigma_x^2(k) = E\left\{\left|X\left(k,l\right)\right|^2\right\}$ holds for wide-sense stationary processes.

As shown in Fig. 15, if $X\left(k,l\right)$ and $N\left(k,l\right)$ have L_o frames with the same values, the correlation of the two Gaussian processes is $\rho_x = L_o/L_x$ [52]. If $L_o = 0$, $\rho_x = 0$ indicates that the two Gaussian processes are totally uncorrelated/indepdendent. If $L_o = L_x$, $\rho_x = 1$ means that the two Gaussian processes are perfectly correlated. As defined in (9), the entry in the ith row and jth column of the matrix $\widehat{\mathbf{R}}_{xx}$ and that of the matrix $\widehat{\mathbf{R}}_{nn}$ are respectively given by

$$\left(\widehat{\mathbf{R}}_{xx}\right)_{i,j} = \exp\left(\beta_x\right)\left(\mathbf{R}_{xx}\right)_{i,j}, \tag{55}$$

and

$$\left(\widehat{\mathbf{R}}_{nn}\right)_{i,j} = \exp\left(\beta_n\right)\left(\mathbf{R}_{nn}\right)_{i,j}. \tag{56}$$

For noise-only segments, (55) and (56) reduce to,

$$\left(\widehat{\mathbf{R}}_{xx}\right)_{i,j} = \exp\left(\beta_x\right)\left(\mathbf{R}_{xx}\right)_{i,j} = \exp\left(\beta_x\right)\left(\mathbf{R}_{nn}\right)_{i,j}, \tag{57}$$

and

$$\left(\widehat{\mathbf{R}}_{nn}\right)_{i,j} = \exp\left(\beta_n\right)\left(\mathbf{R}_{nn}\right)_{i,j} = \exp\left(\beta_n\right)\left(\mathbf{R}_{xx}\right)_{i,j}, \tag{58}$$

which is due to $\mathbf{R}_{xx} = \mathbf{R}_{nn}$ for this special case.

According to the bivariate normal distribution approximation in (8), the SCMs-ICC of β_x and β_n is given by

$$\rho = \frac{\operatorname{cov}\left(\beta_x,\beta_n\right)}{\sqrt{\operatorname{var}\left\{\beta_x\right\}\operatorname{var}\left\{\beta_n\right\}}}, \tag{59}$$

where $\operatorname{cov}\left(\bullet,\bullet\right)$ means covariance.

To derive the expression of ρ, the relation between ρ and ρ_x needs to be established first. For $\left(\widehat{\mathbf{R}}_{xx}\right)_{i,j}$ and $\left(\widehat{\mathbf{R}}_{nn}\right)_{i,j}$, the only difference is the factors $\exp(\beta_x)$ and $\exp(\beta_n)$ as shown in (57) and (58). Due to that both $\mathbf{R}_{xx}$ and $\mathbf{R}_{nn}$ are Hermitian matrices, they can be written as $\left(\mathbf{R}_{xx}\right)_{i,j} = \left(\mathbf{R}_{nn}\right)_{i,j} = \sum_{m=1}^{M}\sigma_k^2 u_{im}u_{jm}^*$ for noise-only periods, which is due to that $\mathbf{R}_{xx} = \mathbf{R}_{nn} = \mathbf{U}\mathbf{D}\mathbf{U}^H$, where $\mathbf{U}$ is a unitary matrix with u_{ij} the entry in the ith row and jth column and $\mathbf{D}$ is a diagonal matrix with $\operatorname{diag}\left(\mathbf{D}\right) = \left[\begin{array}{cccc}\sigma_1^2 & \sigma_2^2 & \cdots & \sigma_M^2\end{array}\right]$. For $i = j$, we can further have

$$\left(\widehat{\mathbf{R}}_{xx}\right)_{i,i} = \left(\exp\left(\frac{\beta_x}{2}\right)W_i\right)\left(\exp\left(\frac{\beta_x}{2}\right)W_i\right)^*, \tag{60}$$

and

$$\left(\widehat{\mathbf{R}}_{nn}\right)_{i,i} = \left(\exp\left(\frac{\beta_n}{2}\right)W_i\right)\left(\exp\left(\frac{\beta_n}{2}\right)W_i\right)^*, \tag{61}$$

where $W_iW_i^* = \left|W_i\right|^2 = \sum_{m=1}^{M}\sigma_m^2\left|u_{im}\right|^2$ and thus ρ_x in the log domain is given by

$$\rho_x = \frac{\operatorname{cov}\left(\exp\left(\beta_x/2\right),\exp\left(\beta_n/2\right)\right)}{\sqrt{\operatorname{var}\left\{\exp\left(\beta_x/2\right)\right\}\operatorname{var}\left\{\exp\left(\beta_n/2\right)\right\}}}, \tag{62}$$

where (62) can be derived due to that $\operatorname{cov}(ax,by) = ab\operatorname{cov}(x,y)$ and $\operatorname{var}(ax) = a^2\operatorname{var}(x)$ always hold when a and b are constant values, i.e., $a = W_i$ and $b = W_i^*$. The joint p.d.f. of β_x and β_n is presented in (17). From (59) and (62), one can get that ρ defines in the linear domain, while ρ_x defines in the log domain because of lognormal distributions of both $\exp(\beta_x)$ and $\exp(\beta_n)$, which are similar to [44, (12) and (13)]. The bivariate lognormal distribution has already been well studied, where the relation between ρ and ρ_x can also be derived using some equations in [53]. For completeness, we derived this relation step by step in the following. (62) can be further written as

$$\begin{aligned}\rho_x &= \frac{\operatorname{cov}\left(\exp\left(\frac{\beta_x-\bar{\beta}}{2}\right),\exp\left(\frac{\beta_n-\bar{\beta}}{2}\right)\right)}{\sqrt{\operatorname{var}\left\{\exp\left(\frac{\beta_x-\bar{\beta}}{2}\right)\right\}\operatorname{var}\left\{\exp\left(\frac{\beta_n-\bar{\beta}}{2}\right)\right\}}}, \\ &= \frac{\operatorname{cov}\left(\exp\left(\beta_x^0/2\right),\exp\left(\beta_n^0/2\right)\right)}{\sqrt{\operatorname{var}\left\{\exp\left(\beta_x^0/2\right)\right\}\operatorname{var}\left\{\exp\left(\beta_n^0/2\right)\right\}}},\end{aligned} \tag{63}$$

where the joint p.d.f. of β_x^0 and β_n^0 is given by

$$\left[\begin{array}{c}\beta_x^0 \\ \beta_n^0\end{array}\right] \sim \mathcal{N}_2\left(\left[\begin{array}{c}0 \\ 0\end{array}\right],\left[\begin{array}{cc}\sigma_\beta^2 & \rho\sigma_\beta^2 \\ \rho\sigma_\beta^2 & \sigma_\beta^2\end{array}\right]\right), \tag{64}$$

where β_x^0 and β_n^0 follow a Gaussian distribution with mean zero and variance σ_β^2. Hence, we have

$$\begin{aligned}\operatorname{var}\left\{\exp\left(\beta_x^0/2\right)\right\} &= \operatorname{var}\left\{\exp\left(\beta_n^0/2\right)\right\} \\ &= \exp\left(\sigma_\beta^2/4-1\right)\exp\left(\sigma_\beta^2/4\right).\end{aligned} \tag{65}$$

Using the formula for the exponential of a bivariate normal distribution [53], we have

$$E\left\{\exp\left(\frac{\beta_x^0}{2}\right)\exp\left(\frac{\beta_n^0}{2}\right)\right\} = \exp\left(\frac{\sigma_\beta^2}{4}\left(1+\rho\right)\right), \tag{66}$$

and

$$E\left\{\exp\left(\frac{\beta_x^0}{2}\right)\right\}E\left\{\exp\left(\frac{\beta_n^0}{2}\right)\right\}=\exp\left(\frac{\sigma_\beta^2}{4}\right). \quad (67)$$

By substituting (65), (66) and (67) into (62), one obtains

$$\rho_x=\frac{\exp\left(\rho\sigma_\beta^2\big/4\right)-1}{\exp\left(\sigma_\beta^2\big/4\right)-1}. \quad (68)$$

Since $0<\sigma_\beta^2\leq 1$ and $0<\rho<1$, (68) leads to

$$\rho_x\approx\rho=L_o/L_x=L_o/L_n, \quad (69)$$

where the approximation is deduced from (68) using $\exp(x)\approx 1+x$ for $|x|\ll 1$.

References

[1] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*. Berlin: Springer-Verlag, 2001.

[2] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise reduction in speech processing*. Berlin: Springer-Verlag, 2009.

[3] I. Cohen, J. Benesty, and S. Gannot, *Speech processing in modern communication: challenges and perspectives*. Berlin: Springer-Verlag, 2010.

[4] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 1, pp. 27-34, Jan. 1982.

[5] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 1999, pp. 2965-2968.

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Theoretical analysis of the general transfer function GSC," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2001, pp. 103-106.

[7] S. Doclo, "Multi-microphone noise reduction and dereverberation techniques for speech applications," Ph.D. dissertation, Faculty of Engineering, K.U. Leuven, Leuven, Belgium, 2003.

[8] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 487-503, Jul. 2005.

[9] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 158-170, Jan. 2010.

[10] C. Pan, J. Chen, and J. Benesty, "A multistage minimum variance distortionless response beamformer for noise reduction", *J. Acoust. Soc. Amer.*, vol. 137, no. 3, pp. 1377-1388, Mar. 2015.

[11] H. Huang, L. Zhao, J. Chen, and J. Benesty, "A minimum variance distortionless response filter based on the bifrequency spectrum for single-channel noise reduction," *Digit. Signal Process.*, vol. 33, pp. 169-179, Oct. 2014.

[12] E. A. P. Habets, J. Benesty, S. Gannot, and I. Cohen, "The MVDR beamformer for speech enhancement," in *Speech Processing in Modern Communication*, I. Cohen, J. Benesty, and S. Gannot, Eds. Berlin: Springer-Verlag, 2009, ch. 9, pp. 225-254.

[13] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone arrays: signal processing techniques and applications*, M. Brandstein and D. Ward, Eds. New York: Springer, 2001, ch. 3, pp. 39-60.

[14] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 3, pp. 240-259, May 1998.

[15] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multichannel noise reduction - Algorithms and theoretical limits," in *Proc. 9th Eur. Signal Process. Conf.*, Sep. 1999, pp. 105-108.

[16] W. Herbordt and W. Kellermann, "Analysis of blocking matrices for generalized sidelobe cancellers for non-stationary broadband signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2002.

[17] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218-1234, Jul. 2006.

[18] Y. Huang, J. Benesty, and J. Chen, "Analysis and comparison of multichannel noise reduction methods in a common framework," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 5, pp. 957-968, Jul. 2008.

[19] B. Cornelis, M. Moonen, and J. Wouters, "Performance analysis of multichannel Wiener filter-based noise reduction in hearing aids under second order statistics estimation errors," *IEEE Trans. Speech Audio Process.*, vol. 19, no. 5, pp. 1368-1381, Jul. 2011.

[20] H. Saruwatari and R. Miyazaki, "Blind source separation: advances in theory, algorithms and applications," in *Statistical analysis and evalution of blind speech extraction algorithms*, G. R. Naik and W. Wang, Eds. Berlin: Springer-Verlag, 2014, ch. 10, pp. 291-322.

[21] R. Miyazaki, H. Saruwatari, and K. Shikano, "Theoretical analysis of musical noise and speech distortion in structure-generalized parametric blind spatial subtraction array," in *INTERSPEECH*, Aug. 2011, pp. 341-344.

[22] C. Zheng, Y. Zhou, X. Hu, and X. Li, "Two-channel post-filtering based on adaptive smoothing and noise properties," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, May 2011, pp. 1745-1748.

[23] C. Zheng, H. Liu, R. Peng, and X. Li, "A statistical analysis of two-channel post-filter estimators in isotropic noise fields," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 2, pp. 336-342, Feb. 2013.

[24] A. Kuklasinski, 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 Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 91-95.

[25] A. Kuklasinski, S. Doclo, S. H. Jensen, and J. Jensen, "Maximum likelihood PSD estimation for speech enhancement in reverberation and noise," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1599-1612, Sep. 2016.

[26] C. Breithaupt and R. Martin, "Analysis of the decision-directed SNR estimator for speech enhancement with respect to low-SNR and transient conditions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 2, pp. 277-289, Feb. 2011.

[27] T. Gerkmann and R. Martin, "Empirical distributions of DFT-domain speech coefficients based on estimated speech variances," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Aug. 2010.

[28] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230-2244, Sep. 2002.

[29] M. Souden, J. Benesty, and S. Affes, "On optimal frequency-domain multichannel linear filtering for noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 2, pp. 260-276, Feb. 2010.

[30] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367-2387, Dec. 2004.

[31] S. Braun, K. Kowalczyk, and E. A. P. Habets, "Residual noise control using a parametric multichannel Wiener filter," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2015, pp. 360-364.

[32] R. Serizel, M. Moonen, B. V. Dijk, and J. Wouters, "Low-rank approximation based multichannel Wiener filter algorithms for noise reduction with applications in cochlear implants," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 4, pp. 785-799, Apr. 2014.

[33] M. Rahmani, M. H. Bastani, and S. Shahraini, "Two layers beamforming robust against direction-of-arrival mismatch," *IET Signal Process.*, vol. 8, no. 1, pp. 49-58, Feb. 2014.

[34] H. Huang, C. Hofmann, W. Kellermann, J. Chen, and J. Benesty, "A multiframe parametric Wiener filter for acoustic echo suppression," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2016.

[35] R. Martin, "Bias compensation methods for minimum statistics noise power spectral density estimation," *Signal Process.*, vol. 86, no. 6, pp. 1215-1229, Jun. 2006.

[36] T. Gerkmann and R. C. Hendriks, "Unbiased MMSE-based noise power estimation with low complexity and low tracking delay," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1383-1393, May 2012.

[37] S. T. Smith, "Covariance, subspace, and intrinsic Cramer-Rao bounds," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1610-1630, May 2005.

[38] R. C. Hendriks and T. Gerkmann, "Noise correlation matrix estimation for multi-microphone speech enhancement," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 1, pp. 223-233, Jan. 2012.

[39] S. Braun and E. A. P. Habets, "Dereverberation in noisy environments using references signals and a maximum likelihood estimator," in *Proc. 21st Eur. Signal Process. Conf.*, Sep. 2013, pp. 1-5.

[40] K. Reindl, Y. Zheng, A. Schwarz, S. Meier, R. Maas, A. Sehr, and W. Kellermann, "A stereophonic acoustic signal extraction scheme for noisy and reverberant environments," *Comput. Speech Lang.*, vol. 27, no. 3, pp. 726-745, May 2013.

[41] W. Kellermann, K. Reindl, and Y. Zheng, "Method for evaluating a useful signal and audio device," US Patent US20160029130 A1, Jan. 2016.
[42] S. Braun and E. A. P. Habets, "A multichannel diffuse power estimator for dereverberation in the presence of multiple sources," *EURASIP J. Audio, Speech, Music Process.*, vol. 2015, no. 34, pp. 1-14, Dec. 2015.
[43] W. Jouini, D. Le Guennec, C. Moy, and J. Palicot, "Log-normal approximation of Chi-square distributions for signal processing," in *Proc. XXXth URSI Gen. Assem. Sci. Symp.*, Aug. 2011, pp. 1-4.
[44] T. Gerkmann and R. Martin, "On the statistics of spectral amplitudes after variance reduction by temporal cepstrum smoothing and cepstral nulling," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 4165-4174, Nov. 2009.
[45] H. Buchner, R. Aichner, and W. Kellermann, "Bind source separation algorithms for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," in *Proc. Int. Workshop Acoust. Echo Noise Control*, Sep. 2003, pp. 275-278.
[46] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds. Norwell, MA, USA: Kluwer, 2004, ch. 10, pp. 255-293.
[47] O. Schwartz, S. Gannot, and E. A. P. Habets, "Cramér-Rao bound analysis of reverberation level estimators for dereverberation and noise reduction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 8, pp. 1680-1693, Aug. 2017.
[48] P. Thüne and G. Enzner, "Maximum-likelihood approach with Bayesian refinement for multichannel-Wiener postfiltering," *IEEE Trans. Signal Process.*, vol. 65, no. 13, pp. 3399-3413, Jul. 2017.
[49] P. Thüne and G. Enzner, "Maximum-likelihood approach to adaptive multichannel-Wiener postfiltering for wind-noise reduction," in *Proc. ITG Speech Commun.*, Oct. 2016, pp. 302-306.
[50] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals series and products*, in: A.Jeffrey, D.Zwillinger (Eds.), 6th ed., New York: Academic, 2000.
[51] I. M. Johnstone, "On the distribution of the largest eigenvalue in principal components analysis," *Ann. Statist.*, vol. 29, no. 2, pp. 295-327, Apr. 2001.
[52] G. C. Carter, "Coherence and time delay estimation," *Proc. of the IEEE*, vol. 75, no. 2, pp. 236-255, Feb. 1987.
[53] E. L. Crow and K. Shimizu, *Lognormal distributions: theory and applications*. New York: CRC Press, 1987.
[54] J. Sherman and W. J. Morrison, "Adjustment of an inverse matrix corresponding to a change in one element of a given matrix," *Ann. Math. Statist.*, vol. 21, no. 1, pp. 124-127, Mar. 1950.
[55] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed., New York: McGraw-Hill, 1991.
[56] C. Zheng, "On second-order statistics of log-periodogram and cepstral coefficients for processes with mixed spectra," *Signal Process.*, vol. 92, no. 10, pp. 2560-2565, Oct. 2012.

Chengshi Zheng received the B.S. degree in Electronic Engineering and Information Science from University of Science and Technology of China (USTC), Hefei, China, in 2004, and the Ph.D degree in Acoustics from Institute of Acoustics, Chinese Academy of Sciences, Beijing, in 2009.

Since then, he has been working with the Institute of Acoustics, Chinese Academy of Sciences, where he is currently an Associate Professor at the Key Laboratory of Noise and Vibration Research. From 2014 to 2015, he was a Visiting Scientist at the Chair of Multimedia Communications and Signal Processing in the University Erlangen-Nuremberg. He is a Senior Member of the Institute of Electrical and Electronics Engineers (IEEE). His research interests include speech enhancement, array signal processing, and spectral estimation.

Antoine Deleforge received the B.Sc. (2008), M.Sc. (2010) and Ph.D. (2013) degrees in computer sciences and applied mathematics from the engineering school Ensimag and the Université Joseph Fourier (Grenoble, France). He worked as a post-doctoral fellow (2014-2015) at the Friedrich-Alexander University (Erlangen, Germany) and is a tenured research scientist with Inria (Rennes, France) since 2016. He serves as a member of the IEEE SPS Technical Committee for Audio and Acoustic Signal Processing since 2018. His research interests lie in audio signal processing, machine learning and their interactions.

Xiaodong Li received the B.S. degree from Nanjing University, Nanjing, China, in 1988, the M.Eng.degree from the Harbin Engineering University, Harbin, China, in 1991, and the Ph.D degree in physical acoustics from the Institute of Acoustics, Chinese Academy of Sciences (IACAS), Beijing, China, in 1995.

After a short period as a Research Fellow at the Hong Kong Polytechnic University working on active control of sound, he was appointed Assistant Professor at IACAS in 1997. He was made Associate Professor at IACAS in 1998 and Professor in 2002. He is currently a vice director at the Key Laboratory of Noise and Vibration Research in IACAS. His research interests include acoustic signal processing, active control of sound and vibration, and engineering acoustics.

Walter Kellermann is a professor for communications at the University of Erlangen-Nuremberg, Germany, since 1999. He received the Dipl.-Ing. (univ.) degree in Electrical Engineering from the University of Erlangen-Nuremberg, in 1983, and the Dr.-Ing. degree from the Technical University Darmstadt, Germany, in 1988. From 1989 to 1990, he was a postdoctoral Member of Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ. In 1990, he joined Philips Kommunikations Industrie, Nuremberg, Germany, to work on hands-free communication in cars. From 1993 to 1999, he was a Professor at the Fachhochschule Regensburg, where he also became Director of the Institute of Applied Research in 1997. In 1999, he cofounded DSP Solutions, a consulting firm in digital signal processing, and he joined the University Erlangen-Nuremberg as a Professor and Head of the Audio Research Laboratory. He authored or coauthored 21 book chapters, 300+ refereed papers in journals and conference proceedings, as well as 70+ patents, and is a co-recipient of ten best paper awards. His current research interests include speech signal processing, array signal processing, adaptive filtering, and its applications to acoustic human-machine interfaces. Dr. Kellermann served as an Associate Editor and Guest Editor to various journals, including the IEEE Transactions on Speech and Audio Processing from 2000 to 2004, the IEEE Signal Processing Magazine in 2015, and presently serves as Associate Editor to the EURASIP Journal on Applied Signal Processing. He was the General Chair of seven mostly IEEE-sponsored workshops and conferences. He served as a Distinguished Lecturer of the IEEE Signal Processing Society (SPS) from 2007 to 2008. He was the Chair of the IEEE SPS Technical Committee for Audio and Acoustic Signal Processing from 2008 to 2010, a Member of the IEEE James L. Flanagan Award Committee from 2011 to 2014, a Member of the SPS Board of Governors (2013-2015), and is currently Vice President Technical Directions of the IEEE Signal Processing Society (2016-2018). He was awarded the Julius von Haast Fellowship by the Royal Society of New Zealand in 2012 and the Group Technical Achievement Award of the European Association for Signal Processing (EURASIP) in 2015. In 2016, he was a Visiting Fellow at Australian National University, Canberra, Australia. He is an IEEE Fellow.