JOINT ESTIMATION OF SOUND SOURCE LOCATION AND NOISE COVARIANCE IN SPATIALLY COLORED NOISE

Futoshi Asano and Hideki Asoh

Intelligent systems R. I., AIST
Central 2, 1-1-1 Umezono Tsukuba 305-8568, Japan
email: f.asano@aist.go.jp
web://staff.aist.go.jp/f.asano/

ABSTRACT

In this paper, sound source localization in spatially colored noise such as room reverberation is discussed. Two iterative algorithms for jointly estimating signal source parameters and noise covariance are proposed. Experimental results show that the estimation of noise covariance improves the spatial resolution of source localization.

1. INTRODUCTION

In the source localization problem, the additive noise in the environment is often assumed to be spatially white for the sake of convenience in deriving an algorithm. In the case of spatially colored noise, a noise-whitening technique such as the generalized eigenvalue decomposition (GEVD)[1] was proposed as briefly reviewed in Section 3. For noise-whitening, information about the noise such as covariance matrix must be available. In speech and audio applications, however, the noise sometimes consists of room reverberation and thus cannot be observed independently. In this paper, two algorithms for jointly estimating source location and noise covariance are proposed. The first one is the maximum-likelihood (ML) approach, while the second one is the Bayesian approach. The experimental results show that the spatial resolution of source localization was improved by the proposed methods as compared to the method using a spatially white assumption.

2. MODEL OF SIGNAL AND NOISE

The frequency-domain observation vector is defined as $\mathbf{z}_k = [Z_1(\omega,k),\cdots,Z_M(\omega,k)]^T$, where $Z_m(\omega,k)$ is the short-time Fourier transform (STFT) of the mth sensor input at the time frame k. The observation vector $\mathbf{z}_k$ can be modeled as

$$\mathbf{z}_k = \sum_{i=1}^{N} \mathbf{a}(\theta_i)s_{i,k} + \mathbf{v}_k = \mathbf{A}(\theta)\mathbf{s}_k + \mathbf{v}_k \qquad (1)$$

where $\mathbf{s}_k = [s_1,\cdots,s_N]^T$ and $\mathbf{v}_k$ are the source and noise vector, respectively. The noise $\mathbf{v}_k$ is assumed to be Gaussian with the distribution $\mathcal{N}(\mathbf{0},\mathbf{K})$ where $\mathbf{K} = E[\mathbf{v}_k\mathbf{v}_k^H]$. The vector $\mathbf{a}(\theta_i)$ is the array manifold vector for the ith source located in the direction θ_i. $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1),\cdots,\mathbf{a}(\theta_N)]$ and $\theta = [\theta_1,\cdots,\theta_N]^T$.

Assuming that $\mathbf{s}_k$ and $\mathbf{v}_k$ are uncorrelated, the correlation matrix of the observation can be modeled as

$$\mathbf{R} = E[\mathbf{z}_k\mathbf{z}_k^H] = \mathbf{A}(\theta)\Gamma\mathbf{A}^H(\theta) + \mathbf{K} \qquad (2)$$

where $\Gamma = E[\mathbf{s}_k\mathbf{s}_k^H]$. Generally, $\mathbf{s}_k$ and $\mathbf{v}_k$ may have some correlation when $\mathbf{v}_k$ consists of room reverberation. For the observation $\mathbf{z}_k$ obtained by STFT as employed in this paper, however, the coherency between the direct sound and the reverberation consisting of replicas of source signal in the previous frames is usually low. A typical example is the direct sound of a consonant in speech overlapped by the reverberation of a vowel in the previous frames. In this paper, thus, $\mathbf{s}_k$ and $\mathbf{v}_k$ are assumed to be uncorrelated for the sake of ease in deriving an algorithm.

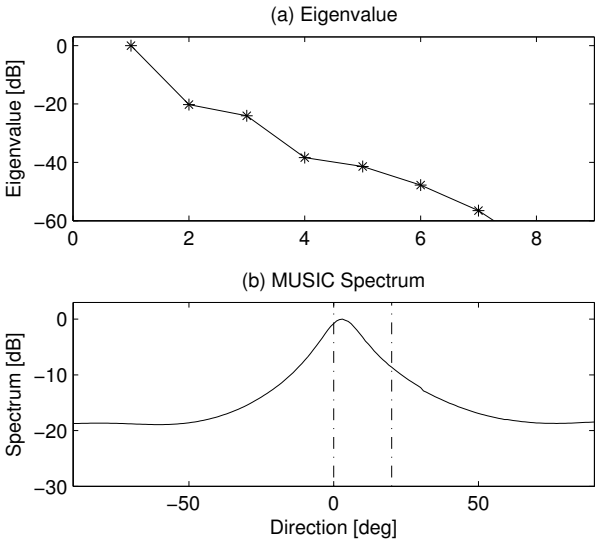


Figure 1: Eigenvalues and the MUSIC spectrum using SEVD.

3. ROLE OF NOISE COVARIANCE IN SOURCE LOCALIZATION

In this section, it is briefly shown how information in the noise covariance affects source localization using the MUSIC estimator as an example. The spatial spectrum of the MUSIC estimator is given by $P(\varphi) = \frac{1}{\|\mathbf{a}^H(\varphi)\mathbf{E}_N\|^2}$, where $\mathbf{a}^H(\varphi)$ denotes the array manifold vector for aribitrary direction φ. The matrix $\mathbf{E}_N = [\mathbf{e}_{N+1},\cdots,\mathbf{e}_M]$ consists of the eigenvector of $\mathbf{R}$ corresponding to the noise subspace. The standard eigenvalue decomposition (SEVD) is usually used under the assumption that the noise is spatially white.

Fig. 1 shows an example of the eigenvalue distribution and the MUSIC spectrum. For obtaining the observation, the impulse responses for two closely located sound sources

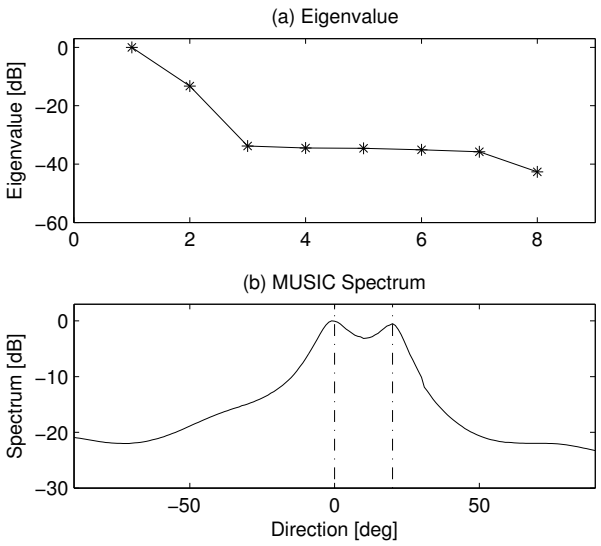


Figure 2: Eigenvalues and MUSIC spectrum using GEVD with the known noise covariance.

$(0^\circ, 20^\circ)$ were measured in a meeting room with the reverberation time of 0.5 s and were convolved with a source signal (speech). The powers of the two sound sources were the same. A microphone array with eight elements mounted on the head of a robot (HRP-2) was used. The spacing between microphones is 50-80 mm (not uniformly spaced.) The length of the frame in STFT is 32 ms (512 points). The length of data for calculating covariance matrix is 2 s (250 frames with 128-point frame shift). The data at 1500 Hz were used. It can be seen that the two peaks that should appear at $(0^\circ, 20^\circ)$ were merged into a single peak.

For colored noise, GEVD that satisfies

$$\mathbf{R}\mathbf{e}_i = \lambda_i \mathbf{K}\mathbf{e}_i \tag{3}$$

can be used[1] instead of SEVD. The difference between SEVD and GEVD is that the noise-whitening process is included in eigenvalue decomposition. Fig. 2 shows the case when GEVD is employed. For using GEVD, the noise covariance $\mathbf{K}$ must be known. For this example, the impulse responses were divided into the direct sound and reflection, and then the responses corresponding to the reflection were convolved with the source signal to obtain the noise observation $\mathbf{v}_k$ separately. From Fig. 2(a), two dominant eigenvalues corresponding to the number of sources $N = 2$ can be seen while the other eigenvalues are almost flat. This is the effect of noise whitening by GEVD. In Fig. 2(b), two peaks appear at $(0^\circ, 20^\circ)$. From these, it can be seen that the spatial resolution of sound localization is improved by the information of noise covariance. In a real application, however, the noise $\mathbf{v}_k$ cannot be observed separately in a case such as noise consisting of room reverberation.

4. ESTIMATION OF NOISE COVARIANCE

In this section, the conditional distribution and expectation of the noise covariance $\mathbf{K}$ is derived [2].

It is assumed that the covariance matrix $\mathbf{K}$ has an inverse-Wishart distribution, a conjugate prior distribution when $\mathbf{v}_k$ is Gaussian, i.e.,

$$p(\mathbf{K}) \propto \det(\mathbf{K})^{-(\nu_0+M+1)} \exp\left\{-\mathrm{tr}(\mathbf{C}_0\mathbf{K}^{-1})\right\} \tag{4}$$

where $\mathbf{C}_0$ is the prior covariance. ν_0 is the virtual sample size for obtaining $\mathbf{C}_0$. The conditional distribution of $\mathbf{K}$ is also the following inverse-Wishart distribution:

$$\begin{aligned} p(\mathbf{K}|\mathbf{Z},\mathbf{S},\theta) &\propto p(\mathbf{K})p(\mathbf{Z}|\theta,\mathbf{S},\mathbf{K}) \\ &\propto \det(\mathbf{K})^{-(\nu_0+K+M+1)} \exp\left\{-\mathrm{tr}([\mathbf{C}_0+\mathbf{C}_1]\mathbf{K}^{-1})\right\} \end{aligned} \tag{5}$$

where $\mathbf{Z} = [\mathbf{z}_1, \cdots, \mathbf{z}_K]$ and $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_K]$. The matrix $\mathbf{C}_1$ is defined as

$$\mathbf{C}_1 = \sum_{k=1}^{K} [\mathbf{z}_k - \mathbf{A}(\theta)\mathbf{s}_k][\mathbf{z}_k - \mathbf{A}(\theta)\mathbf{s}_k]^H \tag{6}$$

The likelihood $p(\mathbf{Z}|\theta,\mathbf{S},\mathbf{K})$ is given by

$$\begin{aligned} p(\mathbf{Z}|\theta,\mathbf{S},\mathbf{K}) &\propto \det(\mathbf{K})^{-K} \cdot \\ &\exp\left(-\sum_{k=1}^{K} [\mathbf{z}_k - \mathbf{A}(\theta)\mathbf{s}_k]^H \mathbf{K}^{-1} [\mathbf{z}_k - \mathbf{A}(\theta)\mathbf{s}_k]\right) \end{aligned} \tag{7}$$

From (5), the conditional expectation of $\mathbf{K}$ is

$$\begin{aligned} E[\mathbf{K}|\mathbf{Z},\mathbf{S},\theta] &= \frac{1}{\nu_0+K-M-1}(\mathbf{C}_0+\mathbf{C}_1) \\ &= \alpha_0 \left(\frac{1}{\nu_0-M-1}\mathbf{C}_0\right) + \alpha_1 \left(\frac{1}{K}\mathbf{C}_1\right) \end{aligned} \tag{8}$$

where $(1/K)\mathbf{C}_1$ is the sample estimate of $\mathbf{K}$. $\alpha_0 = (\nu_0 - M - 1)/(\nu_0 + K - M - 1)$ and $\alpha_1 = K/(\nu_0 + K - M - 1)$ can be interpreted as weights.

5. METHOD I: ML-BASED ALGORITHM

In this section, an algorithm for jointly estimating θ and $\mathbf{K}$, based on the ML method, is proposed.

5.1 Maximum-likelihood estimation of θ

First, the ML estimator for θ on the assumption that $\mathbf{K}$ is given is briefly reviewed [3, 4]. The ML estimate of the signal $\mathbf{s}_k$ is given by

$$\hat{\mathbf{s}}_k = \left[\mathbf{A}^H(\theta)\mathbf{K}^{-1}\mathbf{A}(\theta)\right]^{-1} \mathbf{A}^H(\theta)\mathbf{K}^{-1}\mathbf{z}_k \tag{9}$$

By substituting (9) into (7), the log likelihood in which the unnecessary terms are omitted is obtained as

$$LL(\theta) = -\sum_{k=1}^{K} [\mathbf{G}(\theta)\mathbf{z}_k]^H \mathbf{K}^{-1} [\mathbf{G}(\theta)\mathbf{z}_k] \tag{10}$$

$$= -\mathrm{tr}\left[\mathbf{G}(\theta)\mathbf{C}_z\mathbf{G}^H(\theta)\mathbf{K}^{-1}\right] \tag{11}$$

where

$$\mathbf{G}(\theta) = \mathbf{I} - \mathbf{A}(\theta)\left[\mathbf{A}^H(\theta)\mathbf{K}^{-1}\mathbf{A}(\theta)\right]^{-1}\mathbf{A}^H(\theta)\mathbf{K}^{-1} \tag{12}$$

and

$$\mathbf{C}_z = \sum_{k=1}^{K} \mathbf{z}_k\mathbf{z}_k^H \tag{13}$$

The ML estimate of the source location is given by

$$\hat{\theta} = \arg\max_{\theta} LL(\theta) \tag{14}$$

5.2 Iterative algorithm

In this section, an iterative algorithm for the joint estimation of θ and $\mathbf{K}$ is proposed.

1. Set $\mathbf{K}^{(1)} = \sigma^2\mathbf{I}$ as the initial value.
2. Calculate the log likelihood using (11) and (12) as

$$LL(\theta) = -\mathrm{tr}\left[\mathbf{G}(\theta)\mathbf{C}_z\mathbf{G}^H(\theta)(\mathbf{K}^{(p)})^{-1}\right] \quad (15)$$

$$\mathbf{G}(\theta) = \mathbf{I} - \mathbf{A}(\theta)\left[\mathbf{A}^H(\theta)(\mathbf{K}^{(p)})^{-1}\mathbf{A}(\theta)\right]^{-1} \cdot\mathbf{A}^H(\theta)(\mathbf{K}^{(p)})^{-1} \quad (16)$$

3. Obtain the ML estimate of the location $\theta^{(p+1)}$ using (14) as

$$\theta^{(p+1)} = \arg\max_{\theta} LL(\theta) \quad (17)$$

4. Obtain the sample estimate of the noise covariance using (6) and (16) as

$$\frac{1}{K}\mathbf{C}_1^{(p+1)} = \frac{1}{K}\mathbf{G}(\theta^{(p+1)})\mathbf{C}_z\mathbf{G}^H(\theta^{(p+1)}) \quad (18)$$

5. Update the conditional expactation of the noise covariance using (8) as

$$\mathbf{K}^{(p+1)} = \alpha_0\left(\frac{1}{\nu_0 - M - 1}\mathbf{C}_0\right) + \alpha_1\left(\frac{1}{K}\mathbf{C}_1^{(p+1)}\right) \quad (19)$$

6. Go back to Step 2 with $p \leftarrow p+1$ where p is the iteration index.

This method is somewhat similar to the EM algorithm in which the noise $\mathbf{v}_k$ is treated as the latent variable. However, the conditional expectation of the log likelihood with respect to $\mathbf{v}_k$ cannot be easily derived. Thus, the ML estimate of θ is obtained on the assumption that $\mathbf{K}$ is given. Next, the conditional expectation of $\mathbf{K}$ is calculated, and this process is iterated.

6. METHOD II: ALGORITHM USING GIBBS SAMPLER

In this section, a Bayesian approach for jointly estimating θ, $\mathbf{S}$ and $\mathbf{K}$ using the Gibbs sampler (e.g.,[2]) is proposed.

6.1 Conditional distribution of $\mathbf{s}_k$

The conditional distribution of $\mathbf{s}_k$ is given by

$$p(\mathbf{s}_k|\mathbf{Z},\theta,\tilde{\mathbf{S}}_k,\mathbf{K}) \propto p(\mathbf{s}_k)p(\mathbf{Z}|\theta,\mathbf{S},\mathbf{K}) \quad (20)$$

where

$$\tilde{\mathbf{S}}_k = [\mathbf{s}_1,\cdots,\mathbf{s}_{k-1},\mathbf{s}_{k+1},\cdots,\mathbf{s}_K] \quad (21)$$

Assuming that the prior $p(\mathbf{s}_k)$ is the Gaussian distribution $\mathcal{N}(\mathbf{0},\Phi_0)$, the conditional distribution (20) is also the following Gaussian distribution:

$$\begin{aligned} p(\mathbf{s}_k|\mathbf{Z},\theta,\tilde{\mathbf{S}}_k,\mathbf{K}) &\propto \exp\left[-\mathbf{s}_k^H\left(\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}+\Phi_0^{-1}\right)\mathbf{s}_k \right. \\ &\quad \left. +\mathbf{s}_k^H\mathbf{A}^H\mathbf{K}^{-1}\mathbf{z}_k + \mathbf{z}_k^H\mathbf{K}^{-1}\mathbf{A}\mathbf{s}_k\right] \\ &= \mathcal{N}(\mu_k,\Phi) \end{aligned} \quad (22)$$

where

$$\Phi = \left(\mathbf{A}^H\mathbf{K}^{-1}\mathbf{A}+\Phi_0^{-1}\right)^{-1} \quad (23)$$

$$\mu_k = \Phi\mathbf{A}^H\mathbf{K}^{-1}\mathbf{z}_k \quad (24)$$

6.2 Conditional distribution of θ

Since $\mathbf{A}(\theta)$ is a nonlinear function of θ, it is difficult to obtain samples of θ directly from its conditional distribution. In this case, the Metropolis algorithm can be used[2, 5]. In the Metropolis algorithm, a sample θ^* is obtained from a proposal distribution $J(\theta^*|\theta^{(p)})$ where $\theta^{(p)}$ is the previous sample. In this paper, the following uniform distribution is employed:

$$J(\theta^*|\theta^{(p)}) = \mathcal{U}(\theta^{(p)}-\delta,\theta^{(p)}+\delta) \quad (25)$$

where δ is an appropriate constant vector. The new sample θ^* is accepted when the acceptance ratio r defined by (26) exceeds a threshold r_{thr}.

$$r = \frac{p(\mathbf{Z}|\theta^*,\mathbf{S}^{(p+1)},\mathbf{K}^{(p+1)})}{p(\mathbf{Z}|\theta^{(p)},\mathbf{S}^{(p+1)},\mathbf{K}^{(p+1)})}\frac{p(\theta^*)}{p(\theta^{(p)})} \quad (26)$$

6.3 Iterative algorithm

1. Set $\mathbf{K}^{(1)}$ and $\theta^{(1)}$ as the initial value.
2. Sample $\mathbf{s}_k^{(p+1)} \sim p(\mathbf{s}_k|\mathbf{Z},\theta^{(p)},\tilde{\mathbf{S}}_k^{(p)},\mathbf{K}^{(p)}) = \mathcal{N}(\mu_k,\Phi)$ where

$$\Phi = \left(\mathbf{A}^H(\theta^{(p)})(\mathbf{K}^{(p)})^{-1}\mathbf{A}(\theta^{(p)})+\Phi_0^{-1}\right)^{-1} \quad (27)$$

$$\mu_k = \Phi\mathbf{A}^H(\theta^{(p)})(\mathbf{K}^{(p)})^{-1}\mathbf{z}_k \quad (28)$$

3. Sample $\mathbf{K}^{(p+1)} \sim p(\mathbf{K}|\mathbf{Z},\mathbf{S}^{(p+1)},\theta^{(p)})$ using (5) and (6).
4. Sample $\theta^* \sim J(\theta^*|\theta^{(p)})$ using (25) and determine the new sample as

$$\theta^{(p+1)} = \begin{cases} \theta^* & r > r_{thr} \\ \theta^{(p)} & \text{otherwise} \end{cases} \quad (29)$$

5. Go back to step 2 with $p \leftarrow p+1$

7. EXPERIMENT

The same example as used in Section 3 was used in this experiment. The conditions of the simulation and the analysis are the same as those in Section 3. As the prior noise covariance, $\mathbf{C}_0 = \sigma_0^2\mathbf{I}$ was employed.

First, Method I was evaluated. Fig. 3 shows the estimated source directions. It can be seen that the estimated directions approach the true directions indicated by the dash-dot lines as the number of iteration increases.

Fig. 4 shows the log likelihood $LL(\theta)$ for $p=1$ and $p=8$. When $p=1$, the spatially white covariance ($\mathbf{K}^{(p)} = \sigma^2\mathbf{I}$) is used. For the spatially white case, the distribution is sharp with maximum at $(8^\circ, 9^\circ)$. For the estimated covariance ($p=8$), the distribution is broad with a maximum at $(1^\circ, 18^\circ)$ that is closer to the true direction $(0^\circ, 20^\circ)$. The probable reason for the broad distribution is that the estimated covariance $\mathbf{K}^{(p)}$ is closer to the true covariance, resulting in an increased likelihood.

In Fig. 5 , Method I was applied to the wide frequency range of $[1000, 2500]$ Hz with 49 frequency bins. Fig. 5(a) shows the estimation error $\varepsilon^{(p)} = (1/N)\sum_{i=1}^{N}|\theta_i^{(p)} - \theta_i^{true}|$. "White" and "Estimated" correspond to the case of $p=1$ and $p=8$, respectively. Fig. 5(b) shows an improvement in

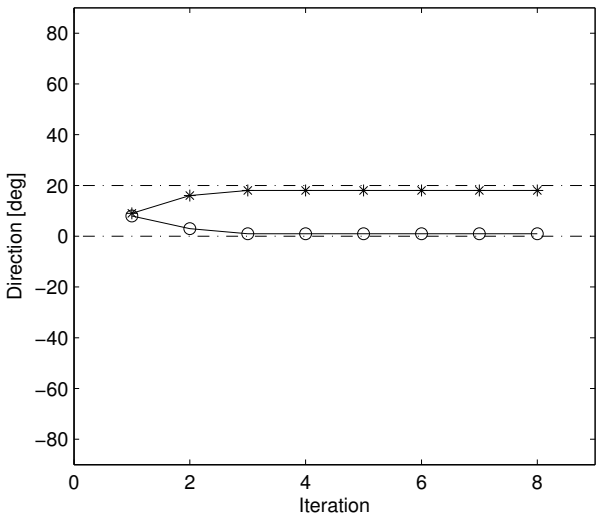
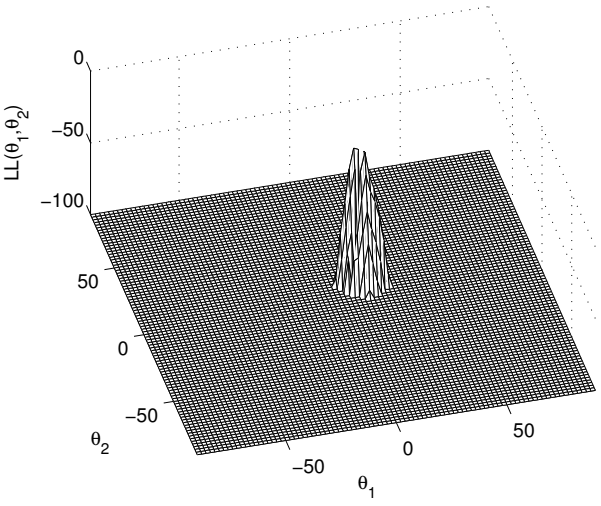


Figure 3: Source directions estimated by Method I.

(a) $p = 1$



(b) $p = 8$

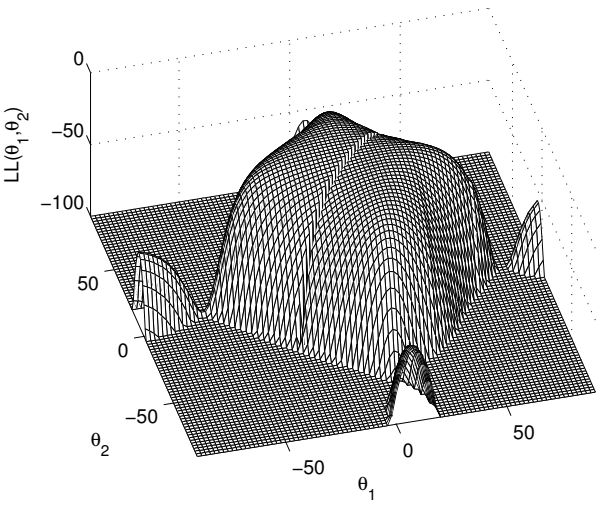


Figure 4: Log likelihood $LL(\theta)$.

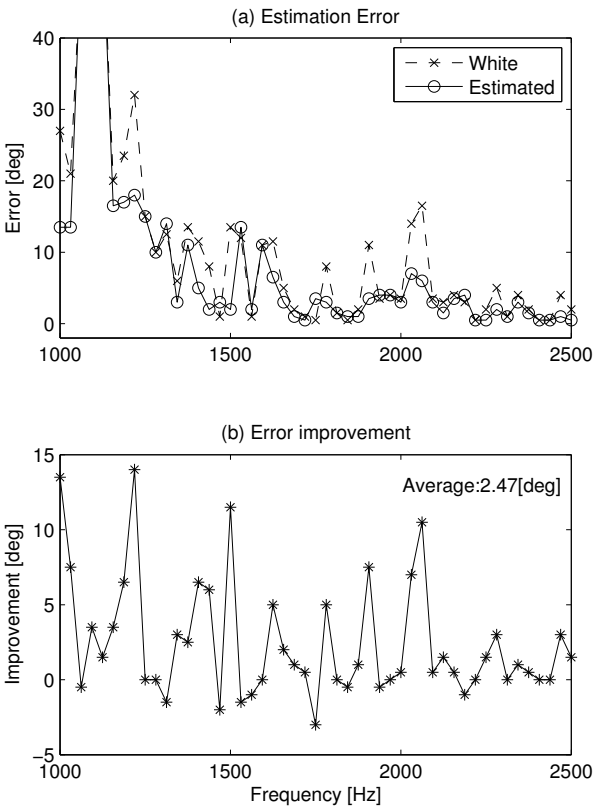


Figure 5: Estimation error for different frequencies.

estimation error, $\varepsilon^{(8)} - \varepsilon^{(1)}$. It can be seen that the estimation precision was improved by the proposed method ($p = 8$) at several frequency bins in the middle frequencies. In the higher frequency range, the estimation was precise even for the initial guess ($p = 1$) with the spatially white assumption. This is mainly due to the physical reason that the spatial resolution is higher for shorter wavelengths. For the lower frequency range, on the other hand, the initial guess was poor at some frequency bins, resulting in a small improvement by the iteration.

In Fig. 6, the estimated covariance was applied to the GEVD-MUSIC method. In the proposed method, the source direction and the noise covariance are jointly estimated. Therefore, there is no need to use GEVD-MUSIC for estimating the source directions. By comparing Fig. 6 with Fig. 2, however, the precision of estimating noise covariance can be known. The eigenvalue distribution and the MUSIC spectrum shown in Fig. 2, which were obtained by using the true noise covariance, were recovered to some extent by using the estimated noise covariance.

Next, Method II was evaluated. The initial value of $\theta^{(p)}$ was set to $(-20^\circ, 60^\circ)$. Fig. 7 shows the variation of sample $\theta^{(p)}$ during the iteration. It can be seen that $\theta^{(p)}$ quickly approachs the true direction. The mean value of samples, which is the Monte Carlo approximation of the conditional expectation of θ, is also shown in Fig. 7.

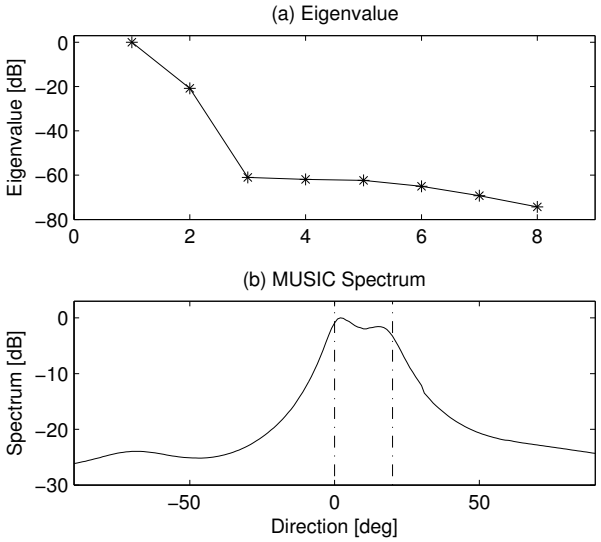


Figure 6: Eigenvalues and MUSIC spectrum using GEVD with the estimated noise covariance.

8. CONCLUSION

In this paper, two algorithms for the joint estimation of signal source parameters and noise covariance were proposed. From the results of the experiment, the spatial resolution for the colored noise environment was improved compared with the method that assumes spatially white-noise.

REFERENCES

[1] R. Roy and T. Kailath, "Esprit - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, no. 7, pp. 984–995, July 1989.

[2] P. D. Hoff, *A first course in Bayesian statistical methods*, Springer, 2009.

[3] M. Miller and D. Fuhrmann, "Maximum-likelihood narrow-band direction finding and the EM algorithm," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 38, no. 9, pp. 1560–1577, 1990.

[4] D. H. Johnson and D. E. Dudgeon, *Array signal processing*, Prentice Hall, Englewood Cliffs NJ, 1993.

[5] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisysinusoids via reversible jump mcmc," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2667–2676, 1999.

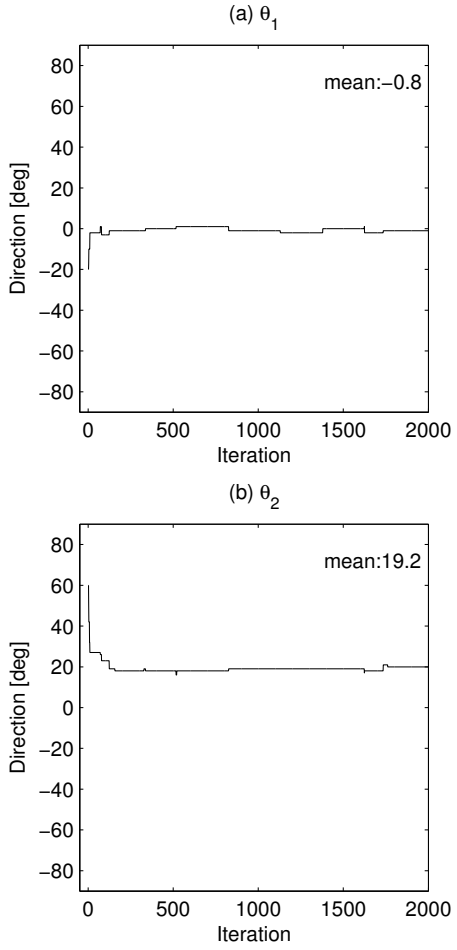


Figure 7: Samples of θ_1 and θ_2 obtained by using Method II.