Crystal-MUSIC: Accurate Localization of Multiple Sources in Diffuse Noise Environments Using Crystal-Shaped Microphone Arrays

Nobutaka Ito[1,2], Emmanuel Vincent[1], Nobutaka Ono[2], Rémi Gribonval[1], and Shigeki Sagayama[2]

[1] INRIA, Centre de Rennes - Bretagne Atlantique
Campus de Beaulieu, 35042 Rennes Cedex, France
{nobutaka.ito,emmanuel.vincent,remi.gribonval}@inria.fr
[2] The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
{ito,onono,sagayama}@hil.t.u-tokyo.ac.jp

Abstract. This paper presents crystal-MUSIC, a method for DOA estimation of multiple sources in the presence of diffuse noise. MUSIC is well known as a method for the estimation of the DOAs of multiple sources but *is not very robust to diffuse noise from many directions, because the covariance structure of such noise is not spherical.* Our method makes it possible for MUSIC to accurately estimate the DOAs by removing the contribution of diffuse noise from the spatial covariance matrix. This denoising is performed in two steps: 1) denoising of the off-diagonal entries via a blind noise decorrelation using crystal-shaped arrays, and 2) denoising of the diagonal entries through a low-rank matrix completion technique. *The denoising process does not require the spatial covariance matrix of diffuse noise to be known, but relies only on an isotropy feature of diffuse noise.* Experimental results with real-world noise show that the DOA estimation accuracy is substantially improved compared to the conventional MUSIC.

Key words: Diffuse noise, DOA estimation, microphone arrays, MUSIC, source localization

1 Introduction

DOA estimation of sound sources is an important issue with many applications such as beamforming and speaker tracking. Real-world sound environments typically contain multiple directional sounds as well as diffuse noise, which comes from many directions like in vehicles or cafeterias. In this paper, we present crystal-MUSIC, a method for accurate estimation of the DOAs of multiple sources in the presence of diffuse noise.

One of the most fundamental approaches to DOA estimation is to maximize the output power of the delay-and-sum or other fixed beamformers with respect to the steering direction. However, since a sharp beam cannot be achieved with a practical small-sized array, DOA estimates may be inaccurate in the presence of multiple sources. Methods based on Time Delay Of Arrival (TDOA) estimates [1] widely in use today assume a single target source and again the performance can be degraded when more than one source is present. In comparison, MUSIC [2–4] estimates the DOAs of multiple sources as directions in which the corresponding steering vector becomes most nearly orthogonal to the noise subspace.

It is important in MUSIC to accurately identify the noise subspace. When there is no noise, it is easily obtained as the null space of the observed covariance matrix. It can be obtained also in the presence of spatially white noise, since such noise only adds its power to all eigenvalues uniformly without changing the eigenvectors because of its spherical covariance structure. Therefore, the basis vectors of the noise subspace coincide with the eigenvectors of the observed covariance matrix belonging to the smallest eigenvalue. Directional noise can be dealt with as well, for it can be regarded as one of the directional signals. In contrast, diffuse noise can significantly degrade the identification of the noise subspace, because the noise spans the whole observation space, and unlike spatially white noise, its covariance structure is not spherical.

Aiming to make MUSIC robust to diffuse noise, this paper proposes a method for removing the contribution of diffuse noise from the spatial covariance matrix. The denoising is performed in two steps. In the first step, the contribution of diffuse noise is removed from the off-diagonal entries through the diagonalization of the covariance matrix of diffuse noise. This is performed through a technique of Blind Noise Decorrelation (BND) [5, 6], in which any covariance matrix of isotropic noise is diagonalized by a single unitary matrix based on the use of symmetrical arrays called crystal arrays. We do not assume the coherence matrix of diffuse noise like in Ref. [7], but only assume an isotropy defined later, aiming to adapt more to various environments in the real world. In the second step of the denoising, thus obtained off-diagonal entries are completed to be the full matrix with the diagonal entries filled in via a low-rank matrix completion technique [8–10]. We present a modified version of the method in Ref. [8] with a positive semi-definite constraint on the estimated covariance matrix.

Throughout, the superscript $^{\mathsf{H}}$ denotes Hermitian transposition. Signals are represented in the time-frequency domain with τ and ω denoting the frame index and the angular frequency. The covariance matrix of a zero-mean random vector $\boldsymbol{\gamma}(\tau, \omega)$ is denoted by $\boldsymbol{\Phi}_{\boldsymbol{\gamma}\boldsymbol{\gamma}}(\tau, \omega) \triangleq \mathcal{E}[\boldsymbol{\gamma}(\tau, \omega)\boldsymbol{\gamma}^{\mathsf{H}}(\tau, \omega)]$, where $\mathcal{E}[\cdot]$ is expectation.

2 Review of MUSIC

2.1 Observation model

We assume that an array of M microphones receives $L(< M)$ directional signals (some of them can be unwanted directional interferences) from unknown directions in the presence of diffuse noise. We assume the number of directional sources, L, to be known in this paper. Let $\boldsymbol{s}(\tau, \omega) \in \mathbb{C}^L$ be the vector comprising the directional signals observed at a reference point (*e.g.* the array centroid), and $\boldsymbol{x}(\tau,\omega) \in \mathbb{C}^M$ and $\boldsymbol{v}(\tau,\omega) \in \mathbb{C}^M$ be the vectors comprising the observed signals and the diffuse noise at the microphones, respectively. Assuming planewave propagation and static sources of the directional signals, we can model the transfer function from $s_l(\tau,\omega)$ to $x_m(\tau,\omega)$ as $D_{ml}(\omega) \triangleq \mathrm{e}^{-\mathrm{j}\omega\delta_{ml}}$, where δ_{ml} denotes the delay in arrival of the directional signal $s_l(\tau,\omega)$ from the reference point to the m-th microphone. Consequently, our observation model is given by

$$\boldsymbol{x}(\tau,\omega) = \boldsymbol{D}(\omega)\boldsymbol{s}(\tau,\omega) + \boldsymbol{v}(\tau,\omega) \tag{1}$$

$$= \sum_{l=1}^{L} \boldsymbol{d}(\omega;\theta_l)s_l(\tau,\omega) + \boldsymbol{v}(\tau,\omega), \tag{2}$$

where θ_l denotes the DOA of the l-th directional sound,

$$\boldsymbol{d}(\omega;\theta) \triangleq \left[\mathrm{e}^{-\mathrm{j}\omega\delta_1(\theta)}\ \mathrm{e}^{-\mathrm{j}\omega\delta_2(\theta)}\ \dots\ \mathrm{e}^{-\mathrm{j}\omega\delta_M(\theta)}\right]^{\mathsf{T}} \tag{3}$$

denotes the steering vector corresponding to DOA θ, and $\delta_m(\theta)$ denotes the time delay of arrival for DOA θ from the reference point to the m-th microphone. We assume $\boldsymbol{s}(\tau,\omega)$ and $\boldsymbol{v}(\tau,\omega)$ to be uncorrelated zero-mean random vectors. As a result, $\boldsymbol{x}(\tau,\omega)$ is a zero-mean random vector with covariance matrix

$$\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega) = \boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega) + \boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega). \tag{4}$$

2.2 DOA estimation

The orthogonal projection of $\boldsymbol{d}(\omega;\theta)$ onto the noise subspace, *i.e.* the orthogonal complement of $\mathrm{span}\{\boldsymbol{d}(\omega;\theta_l)\}_{l=1}^{L}$, becomes zero when θ coincides with θ_l. Therefore, the MUSIC spectrum

$$f_{\mathrm{MUSIC}}(\omega;\theta) \triangleq \|\boldsymbol{V}^{\mathsf{H}}(\omega)\boldsymbol{d}(\omega;\theta)\|_2^{-2} \tag{5}$$

attains peaks at θ_l, where $\boldsymbol{V}$ is a matrix whose columns are orthonormal basis vectors of the noise subspace. Since the MUSIC spectrum (5) is defined for each ω, it is needed to integrate the information from all frequency bins in order to obtain a single estimate of the DOAs. A common approach is to average Eq. (5) over frequencies [3, 4]. For example, the geometric mean [3] gives

$$\overline{f}_{\mathrm{MUSIC}}(\theta) \triangleq \left[\prod_{\omega} f_{\mathrm{MUSIC}}(\omega;\theta)\right]^{\frac{1}{K}}, \tag{6}$$

with K denoting the number of averaged frequency bins. The DOAs are estimated as peaks in $\overline{f}_{\mathrm{MUSIC}}(\theta)$.

3 Denoising of the Spatial Covariance Matrix

To calculate (5), it is important to accurately estimate $\boldsymbol{V}(\omega)$, namely basis vectors of the noise subspace. However, diffuse noise can significantly degrade the estimation, because it spans the whole observation space, and unlike spatially white noise, its covariance structure is not spherical. Our idea therefore consists in restoring the covariance matrix of the directional signals, $\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)$, from the observed covariance matrix $\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)$ contaminated by diffuse noise, so that we can obtain $\boldsymbol{V}(\omega)$ as eigenvectors of the restored matrix belonging to the eigenvalue 0. The matrix denoising is performed in two steps. Firstly, the contribution of diffuse noise to the off-diagonal entries is removed using BND [5, 6] as explained in Section 3.1. Secondly, the diagonal entries are denoised via a low-rank matrix completion technique [8–10] as explained in Section 3.2.

3.1 Diffuse noise removal from the off-diagonal entries

Coming from many directions, diffuse noise can be regarded as more isotropic than directional. Therefore, we make the following assumptions:

1) Diffuse noise has the same power spectrogram at all microphones:

$$[\boldsymbol{\Phi}_{\boldsymbol{vv}}]_{11}(\tau,\omega) = [\boldsymbol{\Phi}_{\boldsymbol{vv}}]_{22}(\tau,\omega) = \cdots = [\boldsymbol{\Phi}_{\boldsymbol{vv}}]_{MM}(\tau,\omega). \tag{7}$$

2) The inter-channel cross-spectrogram of diffuse noise is identical for all microphone pairs with an equal distance:

$$r_{mn} = r_{pq} \;\Rightarrow\; [\boldsymbol{\Phi}_{\boldsymbol{vv}}]_{mn}(\tau,\omega) = [\boldsymbol{\Phi}_{\boldsymbol{vv}}]_{pq}(\tau,\omega), \tag{8}$$

where r_{mn} is the distance between the m-th and n-th microphones. It was shown that there exist such array geometries that any $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$ satisfying these assumptions is diagonalized by a single unitary matrix [5, 6]. So far, we have found five classes of geometries enabling such diagonalization, namely, regular polygonal, (twisted) rectangular, (twisted) regular polygonal prism, rectangular solid, and regular polyhedral arrays. They are called crystal arrays from their shapes.

Making use of a crystal array, we can remove the contribution of the diffuse noise to the off-diagonal entries as follows:

$$\boldsymbol{P}^{\mathsf{H}}\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)\boldsymbol{P} = \boldsymbol{P}^{\mathsf{H}}\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)\boldsymbol{P} + \boldsymbol{P}^{\mathsf{H}}\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)\boldsymbol{P} \tag{9}$$

where $\boldsymbol{P}$ is a unitary diagonalization matrix of $\boldsymbol{\Phi}_{\boldsymbol{vv}}(\tau,\omega)$.

3.2 Denoising of the diagonal entries

Now that the off-diagonal entries of $\boldsymbol{P}^{\mathsf{H}}\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)\boldsymbol{P}$ has been obtained, the problem has reduced to that of completing its missing diagonal elements. Once this is done, the desired matrix $\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)$ will be computed by the transformation $\boldsymbol{P}(\cdot)\boldsymbol{P}^{\mathsf{H}}$. Since $\boldsymbol{P}^{\mathsf{H}}\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)\boldsymbol{P}$ is of rank at most L, the technique of low-rank matrix completion [8–10] can be applied. We present here a variant of an EM-based method by Srebro *et al.* [8] with a positive semi-definite constraint on the matrix to be completed. This is because MUSIC identifies the noise subspace based on the property that the eigenvectors of $\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)$ belonging to the positive and zero eigenvalues form bases of the signal and noise subspaces, respectively. Therefore, if the estimated matrix has some negative eigenvalues, there is no way of assigning the corresponding eigenvectors to one of these subspaces in a reasonable way.

We consider that we obtain via BND an imcomplete observation $\boldsymbol{Y}$ of $\boldsymbol{\Theta} \triangleq \boldsymbol{P}^{\mathsf{H}}\boldsymbol{D}(\omega)\boldsymbol{\Phi}_{\boldsymbol{ss}}(\tau,\omega)\boldsymbol{D}^{\mathsf{H}}(\omega)\boldsymbol{P}$, where the obtained off-diagonal elements $\{x_{mn}\}$ are regarded as observables, and the missing diagonal elements $\{z_{mm}\}$ as latent variables. (Therefore, the diagonal entries of the basis-transformed observed covariance matrix $\boldsymbol{P}^{\mathsf{H}}\boldsymbol{\Phi}_{\boldsymbol{xx}}(\tau,\omega)\boldsymbol{P}$ is just abandoned here.) The observation $\boldsymbol{Y}$ is considered to contain some errors because BND is generally not perfect. Therefore, $\boldsymbol{Y}$ is modeled as follows:

$$\underbrace{\begin{bmatrix} z_{11} & y_{12} & \cdots & y_{1M} \\ y_{21} & z_{22} & \cdots & y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ y_{M1} & y_{M2} & \cdots & z_{MM} \end{bmatrix}}_{\boldsymbol{Y}} = \underbrace{\begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1M} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{M1} & \theta_{M2} & \cdots & \theta_{MM} \end{bmatrix}}_{\boldsymbol{\Theta}} + \underbrace{\begin{bmatrix} \epsilon_{11} & \epsilon_{12} & \cdots & \epsilon_{1M} \\ \epsilon_{21} & \epsilon_{22} & \cdots & \epsilon_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \epsilon_{M1} & \epsilon_{M2} & \cdots & \epsilon_{MM} \end{bmatrix}}_{\boldsymbol{E}}, \tag{10}$$

where $\boldsymbol{E}$ is the error term and its entries ϵ_{mn} are assumed to be i.i.d. complex-valued Gaussian random variables. The criterion is the maximization of the log-likelihood of the observed data subject to the constraint that $\boldsymbol{\Theta}$ is positive semi-definite and of rank at most L:

$$\hat{\boldsymbol{\Theta}} = \arg\max_{\boldsymbol{\Theta}\in\Omega} \ln P(\{y_{mn}\}_{m\neq n}|\boldsymbol{\Theta}), \tag{11}$$

where Ω denotes the set of the $M \times M$ positive semi-definite matrices of rank at most L.

The E-step amounts to the calculating the new estimate $\hat{\boldsymbol{Y}}^{(i+1)}$ of $\boldsymbol{Y}$ by completing the diagonal entries of $\boldsymbol{Y}$ by those of the current estimate $\hat{\boldsymbol{\Theta}}^{(i)}$ of $\boldsymbol{\Theta}$:

$$\hat{\boldsymbol{Y}}^{(i+1)} = \begin{bmatrix} \hat{\theta}_{11}^{(i)} & y_{12} & \cdots & y_{1M} \\ y_{21} & \hat{\theta}_{22}^{(i)} & \cdots & y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ y_{M1} & y_{M2} & \cdots & \hat{\theta}_{MM}^{(i)} \end{bmatrix}. \tag{12}$$

The M-step amounts to calculating the new estimate $\hat{\boldsymbol{\Theta}}^{(i+1)}$ of $\boldsymbol{\Theta}$ as the best approximation of $\hat{\boldsymbol{Y}}^{(i+1)}$ in the Frobenius sense subject to $\boldsymbol{\Theta} \in \Omega$:

$$\hat{\boldsymbol{\Theta}}^{(i+1)} = \arg\min_{\boldsymbol{\Theta}\in\Omega} \|\hat{\boldsymbol{Y}}^{(i+1)} - \boldsymbol{\Theta}\|_{\mathrm{F}}. \tag{13}$$

The solution is written explicitly using the eigenvalue decomposition of $\hat{\boldsymbol{Y}}^{(i+1)}$:

$$\hat{\boldsymbol{Y}}^{(i+1)} = \boldsymbol{U}^{(i+1)}\boldsymbol{\Lambda}^{(i+1)}\boldsymbol{U}^{\mathsf{H}(i+1)}, \tag{14}$$

where $\boldsymbol{U}^{(i+1)}$ is unitary and the eigenvalues in the diagonal of $\boldsymbol{\Lambda}^{(i+1)}$ are ordered from largest to smallest (possibly negative). Then, the solution to (13) is

$$\hat{\boldsymbol{\Theta}}^{(i+1)} = \boldsymbol{U}^{(i+1)}\boldsymbol{\Lambda}_{\mathrm{T}}^{(i+1)}\boldsymbol{U}^{\mathsf{H}(i+1)}. \tag{15}$$

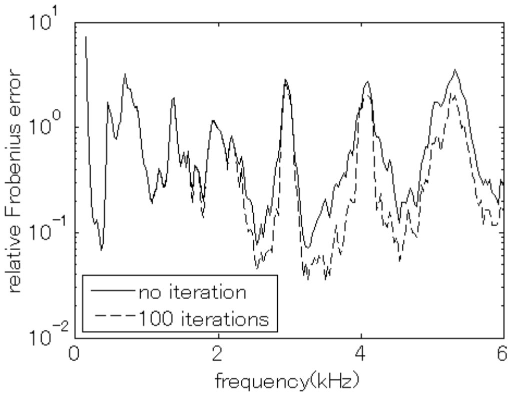


Fig. 1. The relative Frobenius error as a function of the frequency before and after the the covariance matrix denoising.

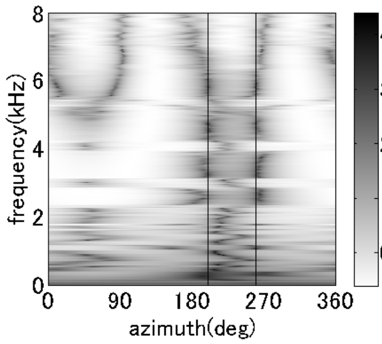
Here, $\boldsymbol{\Lambda}_{\mathrm{T}}^{(i+1)}$ is the truncated version of $\boldsymbol{\Lambda}^{(i+1)}$ whose diagonal entry (eigenvalue of $\hat{\boldsymbol{Y}}^{(i+1)}$) is kept if and only if it is positive and among the L largest and replaced by zero otherwise. The parameters are initialized by $\hat{\boldsymbol{Y}}^{(0)} = \hat{\boldsymbol{\Theta}}^{(0)} = \boldsymbol{P}^{\mathsf{H}}\boldsymbol{\Phi}_{\boldsymbol{xx}}\boldsymbol{P}$. Using the resulting estimate $\hat{\boldsymbol{\Theta}}$, the estimate of $\boldsymbol{D\Phi_{ss}D}^{\mathsf{H}}$ is given by $\boldsymbol{P}\hat{\boldsymbol{\Theta}}\boldsymbol{P}^{\mathsf{H}}$, from which the vector $\boldsymbol{V}$ in Eq. (5) is calculated.

4 Experimental Results

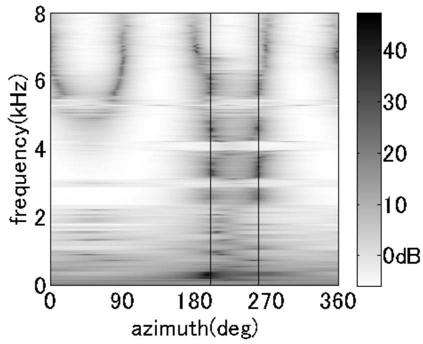
We present experimental results to show the effectiveness of crystal-MUSIC. We recorded noise in a station building in Tokyo with a square array of diameter 5 cm [11]. Two target speech signals were added to this noise recording under the assumption of plane wave propagation. The speech data were taken from the ATR Japanese speech database [12]. The duration of thus generated observed signals was 7 s, and the sampling frequency was 16 kHz. We used STFT for subband decomposition, where the frame length and the frame shift were 512 and 64, respectively, and the Hamming window was used. $\boldsymbol{\Phi}_{\boldsymbol{xx}}$ for both methods was calculated by averaging $\boldsymbol{x}(\tau,\omega)\boldsymbol{x}^{\mathsf{H}}(\tau,\omega)$ temporally over all frames.

To see how well the covariance matrix denoising works, we plot in Fig. 1 a relative Frobenius error defined by $\| \cdot - \boldsymbol{D\Phi_{ss}D}^{\mathsf{H}}\|_{\mathrm{F}}/\|\boldsymbol{D\Phi_{ss}D}^{\mathsf{H}}\|_{\mathrm{F}}$ as a function of the frequency. The solid and dashed lines are the results for the covariance matrices before and after the denoising (*i.e.* $\boldsymbol{\Phi}_{\boldsymbol{xx}}$ and $\boldsymbol{P}\hat{\boldsymbol{\Theta}}\boldsymbol{P}^{\mathsf{H}}$). The SNR at the first microphone was adjusted to -5 dB and the number of iterations of the EM algorithm was 100. The true DOAs of the target signals were 200° and 260°. We see that the error was effectively reduced through the denoising.

Figure 2 is an example of (a) the conventional MUSIC spectra and (b) the crystal-MUSIC spectra, for each frequency. The SNR at the first microphone was adjusted to -5 dB and the number of iterations of the EM was 100. The lines show the true DOAs, namely 200° and 260°. We see that the crystal-MUSIC gave more accurate peak positions and much less spurious peaks.



(a) subband MUSIC spectra



(b) subband crystal-MUSIC spectra

Fig. 2. An example of the subband MUSIC spectra for (a) conventional MUSIC and (b) crystal-MUSIC.

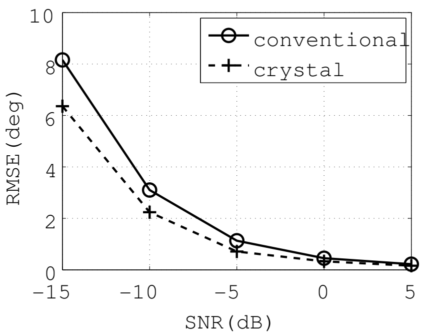


Fig. 3. The RMSE of DOA estimation as a function of SNR.

Finally, we compare the accuracy of DOA estimation by MUSIC and crystal-MUSIC statistically. Fig. 3 shows the Root Mean Square Error (RMSE) of the DOA estimation by the methods as a function of the SNR at the first microphone. The estimates were obtained from the geometric mean of narrowband MUSIC spectra as in Eq. (6). The range of averaging was 80th to 150th frequency bins (approximately 2.5 to 4.7 kHz). The range was determined so as to avoid using low frequencies with a very low SNR and high frequencies with spatial aliasing. The RMSE was calculated from an experiment with various source DOAs, where all the 15 DOA combination from the set $\{0°, 60°, 120°, \cdots, 300°\}$ were tested. The figure shows a substantial improvement in RMSE by crystal-MUSIC.

5 Conclusion

We described crystal-MUSIC, an accurate method for estimating DOAs of multiple sounds in a diffuse noise field. It is based on removal of the contribution of diffuse noise from the observation covariance matrix via BND using crystal arrays and a low-rank matrix completion technique. We presented a new matrix completion method with a positive semi-definite constraint, which is more suitable to MUSIC. The experiment using real-world noise showed the effectiveness of the covariance matrix denoising and the substantial improvement in the DOA estimation accuracy by crystal-MUSIC.

Acknowledgments. This work is supported by INRIA under the Associate Team Program VERSAMUS and by Grant-in-Aid for Young Scientists (B)21760309 from MEXT, Japan.

References

1. C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, no. 4, pp. 320–327, Aug. 1976.
2. R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, pp. 276–280, Mar. 1986.
3. M. Wax, T.-J. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eignestructure methods," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 817–827, Aug. 1984.
4. T. Pham and B. M. Sadler, "Adaptive wideband aeroacoustic array processing," *IEEE Trans. Acoust., Speech, Signal Process.*, pp. 817–827, Aug. 1984.
5. H. Shimizu, N. Ono, K. Matsumoto, and S. Sagayama, "Isotropic noise suppression in the power spectrum domain by symmetric microphone arrays," in *Proc. WASPAA*, New Paltz, NY, Oct. 2007, pp. 54–57.
6. N. Ono, N. Ito, and S. Sagayama, "Five classes of crystal arrays for blind decorrelation of diffuse noise," in *Proc. SAM*, Darmstadt, Germany, July 2008, pp. 151–154.
7. J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds. Berlin: Springer-Verlag, 2001, ch. 2, pp. 19–38.
8. N. Srebro and T. Jaakkola, "Weighted low-rank approximations," in *20th International Conference on Machine Learning*. AAAI Press, 2003, pp. 720–727.
9. E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *The Journal of the Society for the Foundations of Computational Mathematics*, no. 9, pp. 717–772, Apr. 2009.
10. S. Ji and J. Ye, "An accelerated gradient method for trace norm minimization," in *Proceedings of the 26th International Conference on Machine Learning*, Montreal, Canada, 2009, pp. 457–464.
11. N. Ito, N. Ono, E. Vincent, and S. Sagayama, "Designing the Wiener post-filter for diffuse noise suppression using imaginary parts of inter-channel cross-spectra," in *Proc. ICASSP 2010*, Dallas, USA, Mar. 2010.
12. A. Kurematsu, K. Takeda, Y. Sagisaka, S. Katagiri, H. Kuwabara, and K. Shikano, "ATR Japanese speech database as a tool of speech recognition and synthesis," vol. 9, no. 4, pp. 357–363, Aug. 1990.