



Diffuse noise robust multiple source localization based on noise reduction in covariance matrix domain

Nobutaka Ito, Emmanuel Vincent, Nobutaka Ono, Rémi Gribonval, Shigeki Sagayama

► To cite this version:

Nobutaka Ito, Emmanuel Vincent, Nobutaka Ono, Rémi Gribonval, Shigeki Sagayama. Diffuse noise robust multiple source localization based on noise reduction in covariance matrix domain. IEICE EA Technical Committee Meeting, Dec 2010, Tsukuba, Japan. pp.31-36. inria-00596150

HAL Id: inria-00596150

<https://hal.inria.fr/inria-00596150>

Submitted on 30 Sep 2011

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.

共分散行列領域での雑音除去に基づく拡散性雑音に頑健な複数音源定位

伊藤 信貴[†] エマニュエル・ヴァンソン^{††} 小野 順貴[†] レミ・グリボンヴァル^{††}

嵯峨山茂樹[†]

[†] 東京大学大学院 情報理工学系研究科 〒113-0033 東京都文京区本郷 7-3-1

^{††} フランス国立情報学自動制御研究所 Campus de Beaulieu, 35042 Rennes Cedex, France

E-mail: [†]{ito,onono,sagayama}@hil.t.u-tokyo.ac.jp, ^{††}{emmanuel.vincent,remi.gribonval}@inria.fr

あらまし 本稿では、拡散性雑音の存在下でも正確に複数音源の方位角を推定する手法を提案する。複数音源の方位角を推定するための手法として知られる MUSIC (Multiple Signal Classification) は、空間的に白色な雑音に対しては頑健だが、拡散性雑音が存在すると推定精度が低下する。提案手法では、方向性信号の共分散行列が低ランクであるという仮定と、拡散性雑音の共分散行列が行列空間内の或る部分空間に属するという仮定に基づき、観測された共分散行列から方向性信号の共分散行列を推定し、それに対して MUSIC を適用する。拡散性雑音の共分散行列の部分空間モデルは、素子間で無相関な雑音、コヒーレンス行列が既知の雑音、結晶型アレイで観測された等方雑音といった従来の雑音モデルを特別な場合として含む。実環境で録音された雑音を用いた実験により、提案法を用いて従来の MUSIC よりも高精度に複数音源の方位角を推定できることが示された。

キーワード 音源定位、共分散行列、マイクロフォンアレイ、MUSIC、拡散性雑音

Diffuse noise robust multiple source localization based on noise reduction in covariance matrix domain

Nobutaka ITO[†], Emmanuel VINCENT^{††}, Nobutaka ONO[†], Rémi GRIBONVAL^{††}, and Shigeki SAGAYAMA[†]

[†] Graduate School of Information Science and Technology, the University of Tokyo 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

^{††} INRIA, Centre de Rennes - Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes Cedex, France

E-mail: [†]{ito,onono,sagayama}@hil.t.u-tokyo.ac.jp, ^{††}{emmanuel.vincent,remi.gribonval}@inria.fr

Abstract In this paper, we propose a method for estimating the azimuths of multiple sound sources accurately even in the presence of diffuse noise. MUSIC (Multiple Signal Classification) for the estimation of the azimuths of multiple sources is robust against spatially white noise but the estimation performance degrades in the presence of diffuse noise. Based on a low-rank assumption on the covariance matrix of directional signals and the assumption that the covariance matrix of diffuse noise belongs to a subspace in a matrix space, the proposed method estimates the covariance matrix of the directional signals and applies MUSIC to the estimated matrix. The subspace model on the covariance matrix of diffuse noise includes as special cases noise models such as spatially uncorrelated noise, noise with a given coherence matrix, and isotropic noise observed with a crystal array. We showed through experiments with real-world noise recordings that the proposed method estimated the azimuths of multiple sources more accurately than the conventional MUSIC.

Key words Source localization, covariance matrix, microphone arrays, MUSIC, diffuse noise

1. はじめに

音響信号の到来方向を推定する音源定位は、テレビ会議や防

犯におけるビデオカメラの制御を始めとして多くの応用を持つ。実環境では多くの場合、複数の方向性信号に加え、様々な方向から到来する拡散性雑音が存在するため、そのような環境

において各方向性信号の方位角を正確に推定することが重要な課題である。

最も基本的な音源定位法の一つは、遅延和ビームフォーマの出力が最大となる方向を探索する方法であるが、アレイ開口や素子数が限られている場合には複数の音源の正確な定位は困難である。また、相互相関関数を用いた到来時間差推定に基づく音源定位法 [1] もよく用いられるが、複数の音源が存在すると到来時間差の推定値の精度が低下する。これに対し、MUSIC [2]~[4] は、方向性信号のステアリングベクトルと直交する雑音部分空間が同定できれば、それを用いて複数の音源の方位角を同時に推定することが可能である。

MUSIC による音源定位では、雑音部分空間の同定が重要である。雑音が存在しない場合には、観測された共分散行列の零空間として雑音部分空間を得ることができる。また、空間的に白色な雑音は共分散行列の固有ベクトルの方向を変えないので、このような雑音が存在する場合にも、観測された共分散行列の最小固有値に属する固有空間として雑音部分空間を同定することができる。しかしながら、拡散性雑音が存在すると、共分散行列の固有ベクトルの方向が変化するため、観測された共分散行列から雑音部分空間を直接同定することは困難である。

提案手法では、方向性信号の共分散行列が低ランクであるという仮定と、拡散性雑音の共分散行列が行列空間内の或る部分空間に属するという仮定に基づき、観測された共分散行列から拡散性雑音の寄与を排除して、方向性信号の共分散行列を推定する。これにより、推定された行列の零空間として雑音部分空間を同定することができる。

本稿の残りの部分は次のように構成される。2 節では拡散性雑音環境における複数音源定位の問題を定式化する。3 節では従来の MUSIC の原理と課題を確認する。4 節では提案する共分散行列の雑音除去法を示す。5 節で実験により提案法の有効性を示した後、最後に 6 節で結論を述べる。

2. 拡散性雑音環境における複数音源定位の問題設定

本稿を通じて次の記法を用いる。上付き記号 * および H はそれぞれ複素共役とエルミート転置を表す。信号は時間周波数領域における零平均の確率変数で表され、 τ および ω はそれぞれフレーム番号と角周波数を表す。ベクトルの信号 $\boldsymbol{\alpha}(\tau, \omega)$ の共分散行列を

$$\Phi_{\boldsymbol{\alpha}\boldsymbol{\alpha}}(\tau, \omega) \triangleq \mathcal{E}[\boldsymbol{\alpha}(\tau, \omega)\boldsymbol{\alpha}^{\text{H}}(\tau, \omega)] \quad (1)$$

により表す。但し、 $\mathcal{E}[\cdot]$ は期待値演算である。

M 素子のマイクロフォンアレイが拡散性雑音の存在下で、到来方向未知の L ($L < M$) 個の方向性信号を受信するとする。ここで、これらの方向性信号のいくつかは方向性の干渉音であっても良い。 $\boldsymbol{s}(\tau, \omega) \in \mathbb{C}^L$ を基準点 (アレイ配置の重心など) で観測された L 個の方向性信号からなるベクトル、 $\boldsymbol{x}(\tau, \omega) \in \mathbb{C}^M$ を各マイクロフォンにおける観測信号からなるベクトル、 $\boldsymbol{v}(\tau, \omega) \in \mathbb{C}^M$ を各マイクロフォンにおける拡散性雑音からなるベクトルとする。方位角 θ から水平に入射する平

面波のステアリングベクトルを

$$\boldsymbol{d}(\omega; \theta) \triangleq \left[e^{-j\omega\delta_1(\theta)} \quad e^{-j\omega\delta_2(\theta)} \quad \dots \quad e^{-j\omega\delta_M(\theta)} \right]^{\text{T}} \quad (2)$$

と記す。ここで、 $\delta_m(\theta)$ は方位角 θ の平面波が基準点から m 番目のマイクロフォンに到達するまでに掛かる時間であり、マイクロフォン座標を用いて幾何学的に計算できる θ の既知の関数である。以上より、 θ_l を l 番目の音源の到来方位角とし、行列

$$\boldsymbol{D}(\omega) = \left[\boldsymbol{d}(\omega; \theta_1) \quad \boldsymbol{d}(\omega; \theta_2) \quad \dots \quad \boldsymbol{d}(\omega; \theta_L) \right] \quad (3)$$

を定義すると、観測信号のモデルは次のように書ける。

$$\boldsymbol{x}(\tau, \omega) \triangleq \boldsymbol{D}(\omega)\boldsymbol{s}(\tau, \omega) + \boldsymbol{v}(\tau, \omega) \quad (4)$$

$\boldsymbol{s}(\tau, \omega)$ と $\boldsymbol{v}(\tau, \omega)$ が無相関であるとする、共分散行列の間に次の関係が成立する。

$$\Phi_{\boldsymbol{x}\boldsymbol{x}}(\tau, \omega) = \boldsymbol{D}(\omega)\Phi_{\boldsymbol{s}\boldsymbol{s}}(\tau, \omega)\boldsymbol{D}^{\text{H}}(\omega) + \Phi_{\boldsymbol{v}\boldsymbol{v}}(\tau, \omega) \quad (5)$$

本稿では $\Phi_{\boldsymbol{x}\boldsymbol{x}}(\tau, \omega)$ を観測された共分散行列と呼び、 $\boldsymbol{D}(\omega)\Phi_{\boldsymbol{s}\boldsymbol{s}}(\tau, \omega)\boldsymbol{D}^{\text{H}}(\omega)$ を方向性信号の共分散行列、 $\Phi_{\boldsymbol{v}\boldsymbol{v}}(\tau, \omega)$ を拡散性雑音の共分散行列と呼ぶ。

拡散性雑音環境下での複数音源定位の目的は、 $\boldsymbol{x}(\tau, \omega)$ が与えられたときに $\{\theta_l | l = 1, \dots, L\}$ を推定することである。

3. MUSIC

3.1 MUSIC による複数音源定位の原理

MUSIC による音源定位は、アレイ多様体と信号部分空間の交点の検出に基づく。アレイ多様体とは θ を変化させたときにステアリングベクトル $\boldsymbol{d}(\omega; \theta)$ が描く \mathbb{C}^M 内の曲線

$$\mathcal{M}(\omega) \triangleq \{\boldsymbol{d}(\omega; \theta) | \theta \in [0, 2\pi)\} \quad (6)$$

である。信号部分空間は真の音源方向に対応するステアリングベクトルが張る \mathbb{C}^M の部分空間

$$\mathcal{S}(\omega) \triangleq \text{span}\{\boldsymbol{d}(\omega; \theta_l)\}_{l=1}^L \quad (7)$$

である。 $\theta = \theta_1, \theta_2, \dots, \theta_L$ において曲線 $\mathcal{M}(\omega)$ は $\mathcal{S}(\omega)$ と交わるから、このとき $\mathcal{S}(\omega)$ の直交補空間である雑音部分空間 $\mathcal{N}(\omega)$ への $\boldsymbol{d}(\omega; \theta)$ の正射影は零となる。従って、正射影の 2 ノルムの 2 乗の逆数

$$f_{\mathcal{N}}(\omega; \theta) \triangleq \frac{1}{\|\boldsymbol{V}^{\text{H}}(\omega)\boldsymbol{d}(\omega; \theta)\|_2^2} \quad (8)$$

は $\theta = \theta_1, \theta_2, \dots, \theta_L$ においてピーク値をとる。ここで、 $\boldsymbol{V}(\omega)$ は $\mathcal{N}(\omega)$ の正規直交基底を列として持つ $M \times (M - L)$ 行列である。本稿では $f_{\mathcal{N}}(\omega; \theta)$ を狭帯域 MUSIC スペクトルと呼ぶ。

単一の方位角推定結果を得るためには、狭帯域 MUSIC スペクトルを周波数方向に平均して単一の広帯域 MUSIC スペクトルを計算する必要がある。平均の取り方は複数考えられるが、例えば算術、幾何、調和平均を用いた場合はそれぞれ以下のようになる [3], [4]。

$$f_{W,A}(\theta) \triangleq \frac{1}{|\Omega|} \sum_{\omega \in \Omega} f_N(\omega; \theta) \quad (\text{算術平均}) \quad (9)$$

$$f_{W,G}(\theta) \triangleq \left[\prod_{\omega \in \Omega} f_N(\omega; \theta) \right]^{1/|\Omega|} \quad (\text{幾何平均}) \quad (10)$$

$$f_{W,H}(\theta) \triangleq \frac{|\Omega|}{\sum_{\omega \in \Omega} 1/f_N(\omega; \theta)} \quad (\text{調和平均}) \quad (11)$$

ここで Ω は平均に用いられる ω の値全体の集合である。方位角の推定値は広帯域 MUSIC スペクトルのピークを検出することにより得られる。

3.2 雑音部分空間の同定

狭帯域 MUSIC スペクトル (8) の計算においては、雑音部分空間 $\mathcal{N}(\omega)$ の正確な同定が重要である。 $\mathcal{N}(\omega)$ は方向性信号の共分散行列の零空間に一致するから、方向性信号のみが存在する場合にはそれは容易に求まる。一般には雑音が存在するために方向性信号の共分散行列自体は観測できないが、空間的に白色な雑音は共分散行列の固有空間を変えないため、この場合には観測された共分散行列の最小固有値に属する固有空間として $\mathcal{N}(\omega)$ を求めることができる。しかしながら、拡散性雑音はマイクロフォン間で相関を持つため、共分散行列の固有ベクトルの方向を変え、拡散性雑音の存在下では観測信号から直接 $\mathcal{N}(\omega)$ を正確に同定することが困難である。

4. 共分散行列の雑音除去のための提案法

本節では、MUSIC を拡散性雑音に対して頑健にすることを旨として、観測された共分散行列から方向性信号の共分散行列を復元する方法を示す。これができれば、雑音部分空間 $\mathcal{N}(\omega)$ は復元された行列の零空間として同定することができる。

拡散性雑音の共分散行列は一般に必ず観測できるとは限らないが、空間的な無相関性など、なんらかのモデルが仮定できる場合も多い。本研究では、こうしたモデルの多くが行列空間内の部分空間として表現できることに着目し、まず観測された共分散行列をこの雑音共分散行列部分空間の直交補空間に正射影することにより、拡散性雑音の寄与を含まない成分を得る。この成分は、理想的には方向性信号の共分散行列のみの寄与で決まるが、一方、雑音共分散行列部分空間に含まれる信号共分散成分は得られない。そこで次に、方向性信号の共分散行列が低ランク（音源数に一致）であることを仮定し、行列の要素（もしくはそれらの線形結合）の一部分から全体を復元する行列補完と呼ばれる数理的技術を用いて、方向性信号の共分散行列を推定する。

4.1 拡散性雑音の共分散行列の部分空間モデル

本稿では、拡散性雑音の共分散行列が、 $M \times M$ エルミート行列の全体がなす \mathbb{R} 上のベクトル空間 $\mathcal{H} \subset \mathbb{C}^{M \times M}$ の或る部分空間 $\mathcal{V}(\omega)$ に属すると仮定する。 \mathcal{H} には内積

$$\langle \mathbf{A}, \mathbf{B} \rangle \triangleq \sum_{m=1}^M \sum_{n=1}^M a_{mn} b_{mn}^* \quad (12)$$

$$= \text{tr}(\mathbf{A}\mathbf{B}) \quad (13)$$

とノルム

$$\|\mathbf{A}\| \triangleq \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle} \quad (14)$$

が付随する。以下に示すように、この部分空間モデルは、空間的に無相関な雑音 [5] やコヒーレンス行列が既知の雑音 [6], [7] といった雑音モデルを特別な場合として含む。

4.1.1 空間的に無相関な雑音

マイクロフォン同士が、波長と比較して十分に離れていれば、拡散性雑音はマイクロフォン間で無相関であると近似できる。このとき拡散性雑音の共分散行列は対角だから、このモデルは次の部分空間に対応する。

$$\mathcal{V} = \{ \mathbf{A} \in \mathcal{H} | \mathbf{A}: \text{対角行列} \} \quad (15)$$

このとき、 \mathcal{V} およびその直交補空間 \mathcal{V}^\perp への射影演算子 $\mathcal{P}[\mathbf{A}]$ および $\mathcal{P}^\perp[\mathbf{A}]$ は以下で与えられる。

$$\mathcal{P}[\mathbf{A}] = \text{diag}(\mathbf{A}) \quad (16)$$

$$\mathcal{P}^\perp[\mathbf{A}] = \text{off-diag}(\mathbf{A}) \quad (17)$$

ここで、 $\text{diag}[\cdot]$ は非対角成分を 0 で置き換える演算、 $\text{off-diag}[\cdot]$ は対角成分を 0 で置き換える演算である。

4.1.2 コヒーレンス行列が既知である雑音

球対称な雑音場など、理想的な拡散性雑音場ではそのコヒーレンス行列（正規化された共分散行列） $\mathbf{\Gamma}(\omega)$ が分かっていることがある。このような場合に対応する部分空間は次で与えられる。

$$\mathcal{V}(\omega) = \{ k\mathbf{\Gamma}(\omega) | k \in \mathbb{R} \} \quad (18)$$

射影演算子は

$$\mathcal{P}(\omega)[\mathbf{A}] = \frac{\text{tr}(\mathbf{A}\mathbf{\Gamma}(\omega))}{\text{tr}(\mathbf{\Gamma}^2(\omega))} \mathbf{\Gamma}(\omega) \quad (19)$$

$$\mathcal{P}^\perp(\omega)[\mathbf{A}] = \mathbf{A} - \frac{\text{tr}(\mathbf{A}\mathbf{\Gamma}(\omega))}{\text{tr}(\mathbf{\Gamma}^2(\omega))} \mathbf{\Gamma}(\omega) \quad (20)$$

により与えられる。

4.1.3 結晶型アレイで観測された等方的雑音

我々は論文 [8] において、結晶型アレイ（図 1 参照）と呼ばれる一定のクラスの対称なアレイ配置を用いるならば、等方的雑音の任意の共分散行列 $\mathbf{\Phi}_{\mathbf{v}\mathbf{v}}(\tau, \omega)$ は定ユニタリ行列 \mathbf{P} により対角化されることを示した。ここで、等方的雑音とは次の 2 つの条件を満たす雑音のことと定義する。

- パワースペクトログラムは全てのマイクロフォンにおいて同一である：

$$[\mathbf{\Phi}_{\mathbf{v}\mathbf{v}}]_{11}(\tau, \omega) = [\mathbf{\Phi}_{\mathbf{v}\mathbf{v}}]_{22}(\tau, \omega) = \dots = [\mathbf{\Phi}_{\mathbf{v}\mathbf{v}}]_{MM}(\tau, \omega) \quad (21)$$

- チャネル間のクロススペクトログラムは、等距離のマイクロフォン対に対しては等しくなる：

$$r_{mn} = r_{pq} \Rightarrow [\mathbf{\Phi}_{\mathbf{v}\mathbf{v}}]_{mn}(\tau, \omega) = [\mathbf{\Phi}_{\mathbf{v}\mathbf{v}}]_{pq}(\tau, \omega) \quad (22)$$

ここで r_{mn} は m 番目および n 番目のマイクロフォン間の距離である。

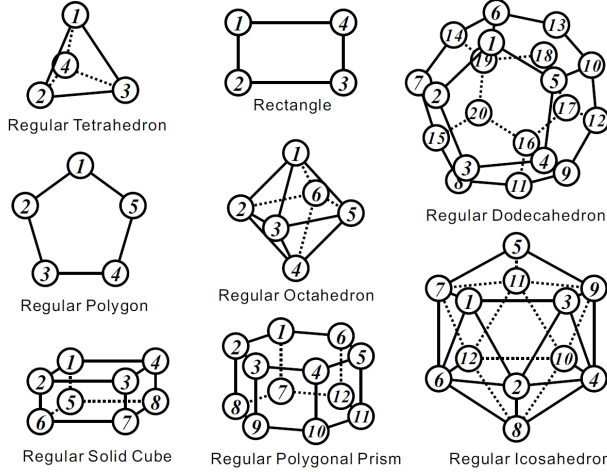


図 1 結晶型アレイの例

上記のモデルは \mathcal{H} の部分空間として下記のように書きなおせる。 $\Phi_{\mathbf{v}\mathbf{v}}(\tau, \omega)$ が \mathbf{P} により対角化されるというのは、 $\Phi_{\mathbf{v}\mathbf{v}}(\tau, \omega)$ が部分空間

$$\mathcal{V} = \{\mathbf{A} \in \mathcal{H} | \mathbf{P}^H \mathbf{A} \mathbf{P}: \text{対角}\} \quad (23)$$

に属することと同値である。 \mathcal{V} はより具体的に

$$\mathcal{V} = \text{span}\{\mathbf{p}_m \mathbf{p}_m^H\}_{m=1}^M \quad (24)$$

とも書ける。ここで、 \mathbf{p}_m は \mathbf{P} の第 m 列である。

射影演算子 \mathcal{P} および \mathcal{P}^\perp は下記のように導くことができる。 $\{\mathbf{p}_m \mathbf{p}_m^H\}_{m=1}^M$ が \mathcal{V} の正規直交基底をなすことに注意すると、

$$\mathcal{P}[\mathbf{A}] = \sum_{m=1}^M \langle \mathbf{A}, \mathbf{p}_m \mathbf{p}_m^H \rangle \mathbf{p}_m \mathbf{p}_m^H \quad (25)$$

$$= \sum_{m=1}^M \text{tr}(\mathbf{A} \mathbf{p}_m \mathbf{p}_m^H) \mathbf{p}_m \mathbf{p}_m^H \quad (26)$$

$$= \sum_{m=1}^M (\mathbf{p}_m^H \mathbf{A} \mathbf{p}_m) \mathbf{p}_m \mathbf{p}_m^H \quad (27)$$

$$= \mathbf{P} \text{diag}(\mathbf{P}^H \mathbf{A} \mathbf{P}) \mathbf{P}^H \quad (28)$$

$$\mathcal{P}^\perp[\mathbf{A}] = \mathbf{A} - \mathcal{P}[\mathbf{A}] \quad (29)$$

$$= \mathbf{P} \text{off-diag}(\mathbf{P}^H \mathbf{A} \mathbf{P}) \mathbf{P}^H \quad (30)$$

4.2 方向性信号の共分散行列の復元アルゴリズム

$\Phi_{\mathbf{x}\mathbf{x}}$ をノイズフリーな空間 \mathcal{V}^\perp に射影するために、 \mathcal{V} の正規直交基底を $\mathbf{Q}_1, \dots, \mathbf{Q}_p$ と書き、 \mathcal{V}^\perp の正規直交基底を $\mathbf{Q}_{p+1}, \dots, \mathbf{Q}_{M^2}$ と書く。但し、 $p \triangleq \dim \mathcal{V}$ である。 $\Phi_{\mathbf{x}\mathbf{x}}$ の $\{\mathbf{Q}_i\}_{i=1}^{M^2}$ による展開を

$$\Phi_{\mathbf{x}\mathbf{x}} = \sum_{i=1}^{M^2} \tilde{\phi}_i \mathbf{Q}_i \quad (31)$$

と書くと、 $\Phi_{\mathbf{x}\mathbf{x}}$ の \mathcal{V}^\perp 成分は

$$\mathcal{P}^\perp[\Phi_{\mathbf{x}\mathbf{x}}] = \sum_{i=p+1}^{M^2} \tilde{\phi}_i \mathbf{Q}_i \quad (32)$$

である。但し、

$$\tilde{\phi}_i \triangleq \langle \Phi_{\mathbf{x}\mathbf{x}}, \mathbf{Q}_i \rangle \quad (33)$$

である。

射影 $\mathcal{P}^\perp[\Phi_{\mathbf{x}\mathbf{x}}]$ は理想的にはノイズフリーであり、方向性信号の共分散行列の \mathcal{V}^\perp 成分に一致するが、一方、 $\Phi_{\mathbf{x}\mathbf{x}}$ の \mathcal{V} 成分は拡散性雑音の共分散行列の寄与も含む。そこで、 \mathcal{V} 成分を未知数 $\{\tilde{z}_i\}_{i=1}^p$ を用いて表した行列

$$\mathbf{Y} \triangleq \sum_{i=1}^p \tilde{z}_i \mathbf{Q}_i + \sum_{i=p+1}^{M^2} \tilde{\phi}_i \mathbf{Q}_i \quad (34)$$

を考え、これを

$$\mathbf{Y} = \Theta + \mathbf{E} \quad (35)$$

のように、真の方向性信号の共分散行列 Θ の誤差 \mathbf{E} を含む観測であるとみなし、低ランク制約 $\Theta \in \Psi$ の下、 Θ を最尤推定する。ここで、 Ψ はランク L 以下の $M \times M$ 半正定値エルミート行列全体の集合である。

いま、

$$\Theta = \sum_{i=1}^{M^2} \tilde{\theta}_i \mathbf{Q}_i \quad (36)$$

$$\mathbf{E} = \sum_{i=1}^{M^2} \tilde{e}_i \mathbf{Q}_i \quad (37)$$

により係数 $\{\tilde{\theta}_i\}$ 、 $\{\tilde{e}_i\}$ を定義すると、式 (35) は

$$\tilde{z}_i = \tilde{\theta}_i + \tilde{e}_i \quad (i = 1, \dots, p) \quad (38)$$

$$\tilde{\phi}_i = \tilde{\theta}_i + \tilde{e}_i \quad (i = p+1, \dots, M^2) \quad (39)$$

と書き直せる。最尤法を定式化するために、誤差 \tilde{e}_i は i.i.d. でおのおの複素ガウス分布 $\mathcal{N}(0, \sigma^2)$ に従うと仮定する。 $\tilde{z}_1, \dots, \tilde{z}_p$ は未知であるから隠れ変数とみなし、 $\{\tilde{\phi}_{p+1}, \dots, \tilde{\phi}_{M^2}\}$ を観測 (不完全) データ、 $\{\tilde{z}_1, \dots, \tilde{z}_p, \tilde{\phi}_{p+1}, \dots, \tilde{\phi}_{M^2}\}$ を完全データとみなし、制約条件

$$\Theta = \sum_{i=1}^{M^2} \tilde{\theta}_i \mathbf{Q}_i \in \Psi \quad (40)$$

の下、観測データの尤度を最大化する $\{\tilde{\theta}_i\}_{i=1}^{M^2}$ を EM アルゴリズムにより探索する。

完全データの対数尤度は

$$\begin{aligned} & \ln p(\{\tilde{z}_i\}_{i=1}^p, \{\tilde{\phi}_i\}_{i=p+1}^{M^2}; \{\tilde{\theta}_i\}_{i=1}^{M^2}, \sigma^2) \\ & \triangleq -M^2 \ln \sigma^2 - \sum_{i=1}^p \frac{|\tilde{z}_i - \tilde{\theta}_i|^2}{\sigma^2} - \sum_{i=p+1}^{M^2} \frac{|\tilde{\phi}_i - \tilde{\theta}_i|^2}{\sigma^2} \end{aligned} \quad (41)$$

である。これを隠れ変数に関して平均化して、 Q 関数は、

$$\begin{aligned} & Q(\{\tilde{\theta}_i\}_{i=1}^{M^2}, \sigma^2; \{\hat{\theta}_i^{(t)}\}_{i=1}^{M^2}, \hat{\sigma}^{2(t)}) \\ & = -M^2 \ln \sigma^2 - p \frac{\hat{\sigma}^{2(t)}}{\sigma^2} - \sum_{i=1}^p \frac{|\hat{\theta}_i^{(t)} - \tilde{\theta}_i|^2}{\sigma^2} \\ & \quad - \sum_{i=p+1}^{M^2} \frac{|\tilde{\phi}_i - \tilde{\theta}_i|^2}{\sigma^2} \end{aligned} \quad (42)$$

となる。これを最大化する $\{\hat{\theta}_i\}_{i=1}^{M^2}$ (つまり、 $\{\hat{\theta}_i^{(t+1)}\}_{i=1}^{M^2}$) は、次式を最小化する。

$$\sum_{i=1}^p |\hat{\theta}_i^{(t)} - \tilde{\theta}_i|^2 + \sum_{i=p+1}^{M^2} |\tilde{\phi}_i - \tilde{\theta}_i|^2 \quad (43)$$

$$= \|\hat{\mathbf{Y}}^{(t+1)} - \Theta\|_F^2 \quad (44)$$

ここで、

$$\hat{\mathbf{Y}}^{(t+1)} \triangleq \sum_{i=1}^p \hat{\theta}_i^{(t)} \mathbf{Q}_i + \sum_{i=p+1}^{M^2} \tilde{\phi}_i \mathbf{Q}_i \quad (45)$$

$$= \mathcal{P}[\hat{\Theta}^{(t)}] + \mathcal{P}^\perp[\Phi_{\mathbf{x}\mathbf{x}}] \quad (46)$$

$$\hat{\Theta}^{(t)} \triangleq \sum_{i=1}^{M^2} \hat{\theta}_i^{(t)} \mathbf{Q}_i \quad (47)$$

である。よって、結局

$$\hat{\Theta}^{(t+1)} = \arg \max_{\Theta \in \Psi} \|\hat{\mathbf{Y}}^{(t+1)} - \Theta\|_F^2 \quad (48)$$

とすれば良い。

紙面の都合上証明は省略するが、式 (48) の解は以下のように $\hat{\mathbf{Y}}^{(t+1)}$ のある種の打ち切り固有値分解により与えられる。 $\hat{\mathbf{Y}}^{(t+1)}$ の固有値分解を

$$\hat{\mathbf{Y}}^{(t+1)} = \mathbf{U}^{(t+1)} \Lambda^{(t+1)} \mathbf{U}^{\mathbf{H}(t+1)} \quad (49)$$

とおく。ここで $\mathbf{U}^{(t+1)}$ はユニタリ行列である。式 (48) の解は

$$\hat{\Theta}^{(t+1)} = \mathbf{U}^{(t+1)} \Lambda_T^{(t+1)} \mathbf{U}^{\mathbf{H}(t+1)} \quad (50)$$

である。ここで、 $\Lambda_T^{(t+1)}$ は $\Lambda^{(t+1)}$ の対角要素 ($\hat{\mathbf{Y}}^{(t+1)}$ の固有値) を大きい方から L 番目まで残して、残りは 0 で置き換える打ち切りを行った後、更に負の固有値があればそれを 0 で置き換えることにより得られる行列である。

以下にアルゴリズムをまとめる。

- $\hat{\mathbf{Y}}^{(0)} = \hat{\Theta}^{(0)} = \Phi_{\mathbf{x}\mathbf{x}}$
- $t = 0, 1, \dots$ に対し、以下をループ停止条件が満たされるまで反復する。
 - $\hat{\mathbf{Y}}^{(t+1)} \triangleq \mathcal{P}[\hat{\Theta}^{(t)}] + \mathcal{P}^\perp[\Phi_{\mathbf{x}\mathbf{x}}]$
 - $\mathbf{Y}^{(t+1)}$ の固有値分解 $\hat{\mathbf{Y}}^{(t+1)} = \mathbf{U}^{(t+1)} \Lambda^{(t+1)} \mathbf{U}^{\mathbf{H}(t+1)}$ を計算する。
 - $\hat{\Theta}^{(t+1)} = \mathbf{U}^{(t+1)} \Lambda_T^{(t+1)} \mathbf{U}^{\mathbf{H}(t+1)}$

なお、観測された共分散行列のみから信号共分散行列を一意に復元できるためには、信号共分散行列のランクが一定以下である必要があると考えられる。そのようなランクの上限については今後の検討課題としたい。

5. 実験による性能評価

本節では、拡散性雑音環境における複数音源定位問題に対する提案法の有効性を検証するため、実環境雑音を用いた音源定位実験の結果を示す。直径 5 cm の正方形アレイを用いて新宿駅で収録した雑音 [9] に、鏡像法 [10] によりシミュレートした目的信号成分を加算することにより観測信号を生成した。部屋のサ

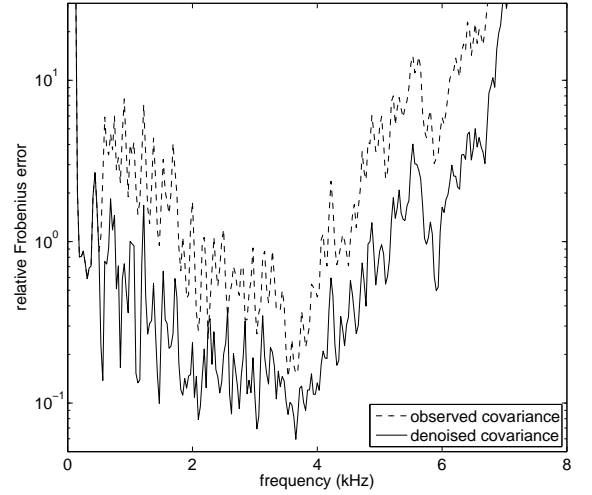


図 2 フロベニウス・ノルムで測った真の方向性音源の共分散行列との相対誤差。破線：観測された共分散行列の相対誤差、実線：推定された共分散行列の相対誤差。

イズは $9\text{m} \times 7\text{m} \times 3\text{m}$ とし、アレイは部屋の中央に水平に配置するものとし、音源はアレイと同一水平面内、1 m の距離にあるものとし、音源数は 2 とした。1 番目のマイクロフォンにおける観測信号の SN 比は、いずれの音源についても -15dB とした。観測信号の信号長は 10 秒、サンプリング周波数は 16kHz であった。観測信号の分析には短時間フーリエ変換を用い、フレーム長を 512 点、フレームシフトを 64 点とし、ハミング窓を用いた。 $\Phi_{\mathbf{x}\mathbf{x}}$ の計算には全フレームを用い、 $\mathbf{x}(\tau, \omega) \mathbf{x}^{\mathbf{H}}(\tau, \omega)$ を時間平均することにより求めた。このように計算された $\Phi_{\mathbf{x}\mathbf{x}}$ に MUSIC を適用した場合 (以下、従来法) と、雑音除去された共分散行列に MUSIC を適用した場合 (以下、提案法) で音源定位の性能を比較する。拡散性雑音の共分散行列のモデル部分空間としては、コヒーレンス行列が既知の雑音 (18) を用い、コヒーレンス行列は球面等方場に対するものを用いた [7]。反復の停止条件は

$$\frac{\|\hat{\Theta}^{(t)} - \hat{\Theta}^{(t-1)}\|_F}{\|\hat{\Theta}^{(t)}\|_F} < 10^{-4} \quad (51)$$

とした。なお、目的音声は ATR 日本語データベース [11] から取った。

方向性信号の共分散行列推定の精度を確認するために、図 2 にフロベニウス・ノルムで測った真の方向性信号の共分散行列 Θ に対する相対誤差

$$\frac{\|\hat{\Theta} - \Theta\|_F}{\|\Theta\|_F} \quad (52)$$

を周波数の関数としてプロットする。破線が観測された共分散行列の相対誤差、実線が推定された方向性信号の共分散行列の相対誤差である。音源の方位角は 60° 、 330° であり、残響時間は 130ms であった。図より、共分散行列の雑音除去により、真の方向性信号の共分散行列との誤差が著しく減少していることが分かる。

図 3 において、従来法と提案法による広帯域 MUSIC スペ

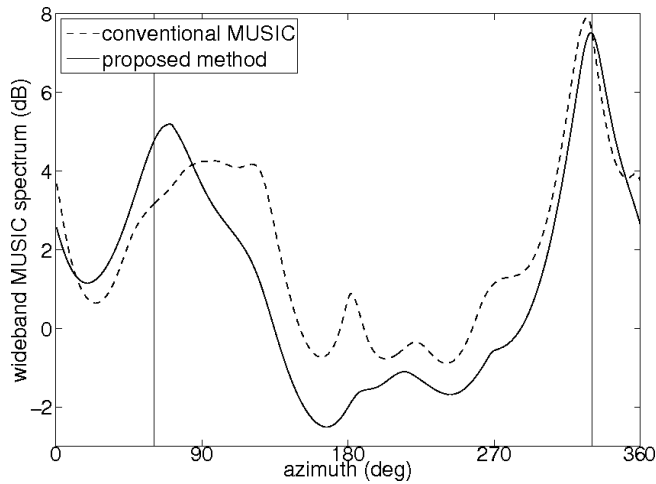


図3 広帯域 MUSIC スペクトルの例。破線：従来の MUSIC、実線：提案法、直線：真の音源の方位角。

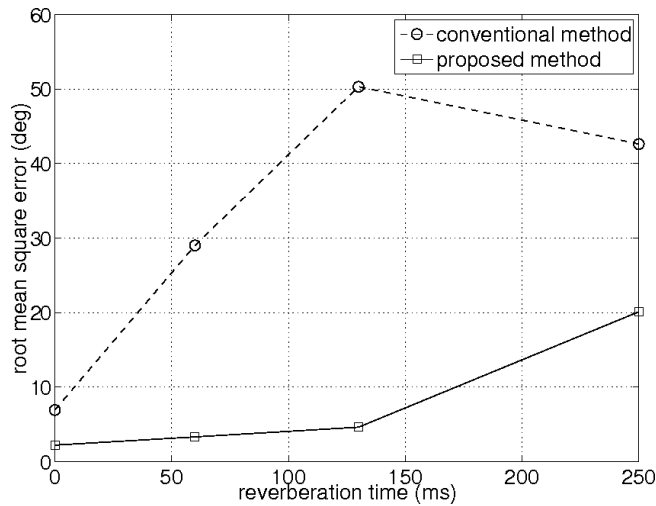


図4 平均二乗誤差の平方根。破線：従来の MUSIC、実線：提案法。

クトルを比較する。諸条件は図2と同一である。SN比の著しく低い3番目の帯域(63Hz)までは捨て、4番目以降の帯域における狭帯域 MUSIC スペクトルを幾何平均した。図中の直線は真の音源の方位角を示す。従来法(破線)は60°の音源に対応するピークが鈍く、ピーク位置が大きくずれているのに加え、偽のピークも存在する。それに対し、提案法(実線)による MUSIC スペクトルにおいては、比較的正確な位置に鋭いピークがあり、偽のピークも抑制されていることが分かる。

最後に、従来法と提案法による方位推定精度の統計的な比較を行った。図4は平均二乗誤差の平方根(root mean square error)を残響時間に対してプロットしたものである。集合 $\{0^\circ, 60^\circ, \dots, 300^\circ\}$ から抽出した任意の2つの方位角に対して実験を行った結果を平均して計算した。図より、提案法は従来法よりも遥かに高精度で音源定位が可能であることが分かる。

6. 結 論

本稿では、観測された共分散行列から方向性信号の共分散行列を復元し、それに対して MUSIC を適用することにより、拡散性雑音の存在下でも正確に複数音源の方位角を推定する手

法を示した。拡散性雑音の共分散行列が行列空間内の或る部分空間に属するというモデルに基づき、観測された共分散行列のモデル部分空間に垂直な成分のみから、方向性信号の共分散行列の低ランク性を用いて、EM アルゴリズムにより方向性信号の共分散行列を復元する。拡散性雑音の共分散行列の部分空間モデルは素子間で無相関な雑音、コヒーレンス行列が既知の雑音、等方的雑音といった従来の雑音モデルを特別な場合として包含する。実環境で録音された雑音と鏡像法によりシミュレートした目的信号成分を用いた実験により、提案法を用いて従来の MUSIC よりも高精度に複数音源の方位角を推定できることが示された。

文 献

- [1] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. ASSP*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.
- [2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
- [3] M. Wax, T. Shan, and T. Kailath, "Spatio-temporal spectral analysis by eigensubstructure methods," *IEEE Trans. ASSP*, vol. ASSP-32, no. 4, pp. 817–827, Aug. 1984.
- [4] T. Pham and B. M. Sadler, "Adaptive wideband aeroacoustic array processing," *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 295–298, Corfu, Jun. 1996.
- [5] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," *Proc. ICASSP '88*, vol. 5, pp. 2578–2581, New York, Apr. 1988.
- [6] R. K. Cook, R. V. Waterhouse, R. D. Berendt, S. Edelman, and M. C. Thompson, Jr., "Measurement of correlation coefficients in reverberant sound fields," *JASA*, vol. 27, no. 6, pp. 1072–1077, Nov. 1955.
- [7] I. A. McCowan and H. Bourslard, "Microphone array post-filter based on noise field coherence," *IEEE Trans. SAP*, vol. 11, no. 6, pp. 709–716, Nov. 2003.
- [8] H. Shimizu, N. Ono, K. Matsumoto, and S. Sagayama, "Isotropic noise suppression in the power spectrum domain by symmetric microphone arrays," *Proc. WASPAA*, pp. 54–57, New Paltz, NY, Oct. 2007.
- [9] 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," *Proc. ICASSP 2010*, pp. 2818–2821, Dallas, Mar. 2010.
- [10] J. A. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *JASA*, vol. 65, no. 4, pp. 943–950, Apr. 1979.
- [11] 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," *Speech Communication*, vol. 9, no. 4, pp. 357–363, Aug. 1990.