階層ベイズ推定を用いた有色雑音環境下での音源定位

Sound localization in colored noise using hierarchical Bayesian estimation

浅野太 [1,2]、麻生英樹 [1]、中臺一博 [2]
Futoshi Asano[1,2], Hideki Asoh[1] and Kazuhiro Nakadai[2]

産業技術総合研究所 [1], ホンダリサーチインスティチュートジャパン [2]
AIST[1], HRI[2]

Abstract

When sound localization is conducted in an enclosure such as a room, reverberation (sum of multiple reflections) behaves as an additive spatially colored noise, resulting in poor spatial resolution. The main reason for this is that the model used in the estimation is assumed to be spatial white. The authors reported that the joint estimation of the sound source parameters and the noise covariance using a Bayesian framework contributes to the improvement of the spatial resolution. In this report, a method of estimating the common factor of the noise covariance from a different observation data set using hierarchical modeling is discussed. This method is considered to be useful in the case such as a dynamic environment when a number of samples included in a single observation is limited.

1　はじめに

部屋などの閉空間内で音源定位を行う場合、部屋の残響（壁などに多重反射した音源信号）が空間的に有色な雑音として加わるため、音源位置推定の分解能が低下する場合がある。これは、推定に用いているモデルが、雑音の空間的白色性を仮定しているためである。有色雑音に対しては、一般化固有値分解 (GEVD) を用いて雑音を白色化する方法が提案されているが [1]、これには有色雑音の共分散行列が既知であることが条件となる。残響の場合、残響だけを単独に観測することができないため、雑音の事前白色化は困難である。著者らは、ベイズ推定の枠組みを用いて、音源位置などのパラメタに加え、雑音の共分散行列を同時推定することにより、空間分解能が改善することを示した [2]。本報告では、共分散行列の推定を階層化 [3] することにより、動的環境のように一度に得られる観測サンプルが少ない場合でも、共分散行列を安定して推定する手法について考える。音源やセンサが閉空間内を移動したとしても、環境（部屋）が同じであれば、部屋に特有の共振周波数は変化せず、このため、部屋内の２点間の伝達関数を共通の極零モデルを用いて表すことができることが報告されている [4]。本報告では、この知見に基づき、雑音の共分散行列を、音源やセンサ位置などが異なる複数のデータセットに共通する項と、個々のデータ（特定の音源-センサ配置）に特有の項に分解できるものと仮定する。このうち、データセットに共通する項を階層モデルを用いて推定する。

2　音源定位の問題

2.1　信号と雑音のモデル

観測値は、マイクロホン入力の短区間フーリエ変換 (STFT) により、$\boldsymbol{z}_{j,k} = [Z_1(\omega, j, k), \cdots, Z_M(\omega, j, k)]^T$ のように構成される。ここで、$Z_m(\omega, j, k)$ は、m 番目のマイクロホン入力の STFT である。j および k はブロックおよびフレームのインデックスを表す。フレームは、STFT を行う単位である。K 個の連続するフレームを $\boldsymbol{Z}_j = [\boldsymbol{z}_1, \cdots, \boldsymbol{z}_K]$ のようにまとめたものを、ここではブロックと呼ぶ。ブロック内では、音源位置 $\boldsymbol{\theta}_j = [\theta_{j,1}, \cdots, \theta_{j,N}]^T$ は定常と見なせるものと仮定する。観測値は、次式のようにモデル化されるものとする。

$$\boldsymbol{z}_{j,k} = \boldsymbol{A}_j(\boldsymbol{\theta}_j)\boldsymbol{s}_{j,k} + \boldsymbol{v}_{j,k} \tag{1}$$

ここで、$\boldsymbol{A}_j(\boldsymbol{\theta}_j) = [\boldsymbol{a}_j(\theta_1), \cdots, \boldsymbol{a}_j(\theta_N)]$ はアレイ・マニフォールド・ベクトル $\boldsymbol{a}_j(\theta_{j,n})$ を列ベクトルに持つマニフォールド行列、$\boldsymbol{s}_{j,k}$ は音源信号、$\boldsymbol{v}_{j,k}$ は雑音を表す。本報告では、$\boldsymbol{v}_{j,k}$ が部屋の残響である場合を考える。$\boldsymbol{v}_{j,k}$ は、以前のフレームにおける信号源 $\boldsymbol{s}_{j,l} (l < k)$ のレプリカであるが、遅延時間がある程度大きい場合は、$\boldsymbol{s}_{j,k}$ と $\boldsymbol{v}_{j,k}$ は無相関と考えてよい。この場合、共分散行列は次式のようにモデル化される。

$$\boldsymbol{R}_j = E[\boldsymbol{z}_{j,k}\boldsymbol{z}_{j,k}^H] = \boldsymbol{A}_j\boldsymbol{\Gamma}_j\boldsymbol{A}_j^H + \boldsymbol{K}_j \tag{2}$$

ここで、$\boldsymbol{\Gamma}_j = E[\boldsymbol{s}_{j,k}\boldsymbol{s}_{j,k}^H]$ および $\boldsymbol{K}_j = E[\boldsymbol{v}_{j,k}\boldsymbol{v}_{j,k}^H]$ は、音源および雑音の共分散行列である。雑音が白色の場合は、

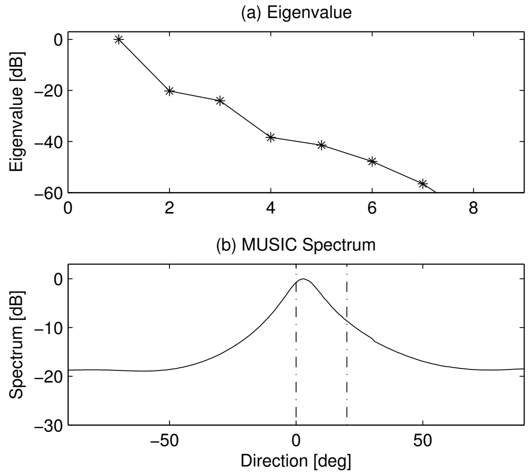

Figure 1: SEVD-MUSIC spatial spectrum. It is assumed that the noise $\boldsymbol{v}_k$ is spatially white with the covariance matrix $\sigma\boldsymbol{I}$.

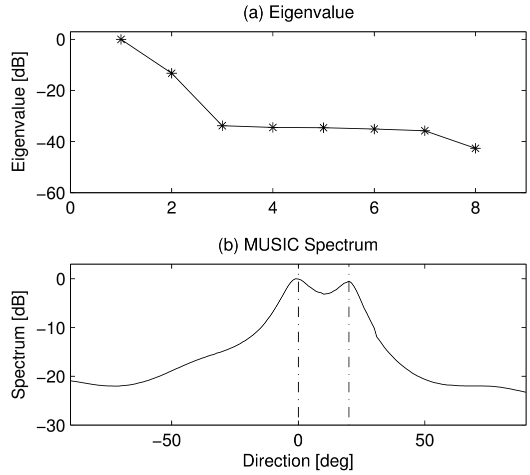

Figure 2: GEVD-MUSIC spatial spectrum. It is assumed that the noise $\boldsymbol{v}_k$ is spatially colored and its covariance matrix $\boldsymbol{K}$ is known.

雑音の共分散行列は、対角行列 $\boldsymbol{K}_j = \sigma_j^2\boldsymbol{I}$ となり、問題は大幅に簡略化される。本報告では、雑音の共分散行列が未知の非対角行列の場合を考える。

2.2 雑音の共分散行列推定の効果

ここでは、雑音の共分散に含まれる情報の効果を、空間スペクトル推定の一手法である MUSIC 法を用いて見ていく。図 1 は、残響時間 0.5 秒程度の会議室で測定したインパルス応答に、白色雑音を畳み込んで生成した観測信号に対して、共分散行列の標準固有値分解 (SEVD) を行い、MUSIC 空間スペクトルを算出したものである。音源方向は図 (b) 中の点線で示すように、$(0°, 20°)$ である。分析パラメタは、表 1 に示してある。インパルス応答測定に用いたマイクロホンアレイは、ロボット (HRP-2) に搭載された、8 素子のものである。同図 (a) から、音源数が $N = 2$ であるにも関わらず、音源数の指標となる大きな固有値は 1 つだけであり、対応した MUSIC 空間スペクトル (同図 (b)) にもピークは 1 つしか現れていない。この原因は、SEVD を用いた MUSIC 法では雑音の空間的白色性を仮定しており、実際の残響の空間的有色性との不整合によるものと考えられる。

Table 1: Parameters for analysis.

Parameter	Value
Sampling frequency	16 kHz
Number of microphones	8
Frame length(STFT length)	512 points
Frame shift	128 points
Block length（observation time）	32000 points
Frequency	1500 Hz

図 2 は、雑音（残響）の共分散行列を既知として、共分散行列のペア $(\boldsymbol{R}, \boldsymbol{K})$ に対して、一般化固有値問題 (GEVD) を解き、これにより MUSIC スペクトルを求めたものである。GEVD に拡張することにより、雑音を空間的に白色化することができる。雑音が残響の場合、実際の応用では $\boldsymbol{K}$ は直接観測できないが、ここでは、ちょっとずるをして、インパルス応答から反射・残響の部分だけを切り出し、これにソース信号を畳み込んで観測値を生成して、$\boldsymbol{K}$ を求めた。この図から、固有値分布では、音源数 $N = 2$ に対応した大きな固有値が現れ、それ以外は、平坦な分布となっている。これは、GEVD による雑音の白色化の効果である。また、空間スペクトルも音源位置に対応した 2 つのピークが現れている。このことから、雑音の共分散行列 $\boldsymbol{K}$ の情報を利用することにより、有色雑音下での空間分解能の向上が期待される。

2.3 観測値の尤度と最尤法

ここでは、観測値に対する尤度を導入する。また、後述のベイズ推定の理解を助けるため、最尤法についても簡単に述べておく。

式 (1) において、$\boldsymbol{v}_{j,k}$ が多次元複素ガウス分布に従うものとし、また、観測値 $\boldsymbol{z}_{j,k}$ は互いに独立であると仮定すると、尤度関数は次式のようになる。

$$
\begin{aligned}
p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) &\propto |\boldsymbol{K}_j|^{-K} \times \\
&\exp\left(-\sum_{k=1}^{K}\left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j\boldsymbol{s}_{j,k}\right]^H \boldsymbol{K}_j^{-1}\left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j\boldsymbol{s}_{j,k}\right]\right) \\
&= |\boldsymbol{K}_j|^{-K}\exp\left\{-\mathrm{tr}\left(\boldsymbol{C}_j\boldsymbol{K}_j^{-1}\right)\right\} \qquad (3)
\end{aligned}
$$

ここで、

$$
\boldsymbol{C}_j = \sum_{k=1}^{K}\left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j\boldsymbol{s}_{j,k}\right]\left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j\boldsymbol{s}_{j,k}\right]^H \qquad (4)
$$

$\boldsymbol{S}_j = [\boldsymbol{s}_{j,1}, \cdots, \boldsymbol{s}_{j,K}]$ はブロック内の音源信号を表す。

本節では、簡単のため、$\boldsymbol{K}_j$ を既知の固定値と考え、式 (3) を $\boldsymbol{s}^*_{j,k}$ について偏微分して $\mathbf{0}$ とおくと、

$$\boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j \boldsymbol{s}_{j,k}\right] = \mathbf{0} \tag{5}$$

これから、$\boldsymbol{s}_{j,k}$ の最尤推定値は次式で与えられる。

$$\hat{\boldsymbol{s}}_{j,k} = \left[\boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{A}_j\right]^{-1} \boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{z}_{j,k} \tag{6}$$

式 (6) を式 (3) に代入して対数をとり、$\boldsymbol{\theta}$ に無関係な項を取り除いて $\boldsymbol{\theta}$ についての対数尤度を求めると、次式のようになる。

$$\begin{aligned} LL(\boldsymbol{\theta}) &\propto -\mathrm{tr}\left(\hat{\boldsymbol{C}}_j \boldsymbol{K}_j^{-1}\right) \\ &= -\mathrm{tr}\left(\boldsymbol{G}(\boldsymbol{\theta}) \boldsymbol{C}_{z,j} \boldsymbol{G}^H(\boldsymbol{\theta}) \boldsymbol{K}_j^{-1}\right) \end{aligned} \tag{7}$$

ここで、

$$\hat{\boldsymbol{C}}_j = \sum_{k=1}^{K} \left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j(\boldsymbol{\theta}) \hat{\boldsymbol{s}}_{j,k}\right] \left[\boldsymbol{z}_{j,k} - \boldsymbol{A}_j(\boldsymbol{\theta}) \hat{\boldsymbol{s}}_{j,k}\right]^H \tag{8}$$

また、$\boldsymbol{G}(\boldsymbol{\theta})$ は、次式で定義されるように、音源方向を $\boldsymbol{\theta}$ と仮定したときに、残差 $\boldsymbol{z}_{j,k} - \boldsymbol{A}_j(\boldsymbol{\theta}) \boldsymbol{s}_{j,k}$ を与えるフィルタである。

$$\boldsymbol{G}(\boldsymbol{\theta}) = \boldsymbol{I} - \boldsymbol{A}_j(\boldsymbol{\theta}) \left[\boldsymbol{A}_j^H(\boldsymbol{\theta}) \boldsymbol{K}_j^{-1} \boldsymbol{A}_j(\boldsymbol{\theta})\right]^{-1} \boldsymbol{A}_j^H(\boldsymbol{\theta}) \boldsymbol{K}_j^{-1} \tag{9}$$

$\boldsymbol{C}_{z,j}$ は、次式で定義される観測値 $\boldsymbol{z}_{j,k}$ のサンプル共分散行列である。

$$\boldsymbol{C}_{z,j} = \sum_{k=1}^{K} \boldsymbol{z}_{j,k} \boldsymbol{z}_{j,k}^H \tag{10}$$

$\boldsymbol{\theta}_j$ の最尤推定値は、次式のようになる。

$$\hat{\boldsymbol{\theta}}_j = \arg\max_{\boldsymbol{\theta}_j} LL(\boldsymbol{\theta}_j) \tag{11}$$

3　パラメタの同時推定

複数のパラメタを同時推定する手法としては、変分ベイズ法 [5] やギブスサンプリング [3] などが考えられる。本報告では、観測値が音源方向 $\boldsymbol{\theta}$ の非線形関数となっていることから、ギブスサンプリングとメトロポリス・アルゴリズムを組み合わせて用いる [3, 6, 2]。3.1 節〜3.3 節では、サンプルを得るための条件付き分布について述べる。3.4 節では、これらの条件付き分布を用いて、パラメタのサンプルを反復して求める手続きについて述べる。

3.1　$\boldsymbol{s}_{j,k}$ の条件付き分布

ベイズの定理により次式が成り立つ。

$$p(\boldsymbol{s}_{j,k}|\boldsymbol{Z}_j, \boldsymbol{\theta}_j, \tilde{\boldsymbol{S}}_j, \boldsymbol{K}_j) \propto p(\boldsymbol{s}_{j,k}) p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) \tag{12}$$

ここで、

$$\tilde{\boldsymbol{S}}_j = [\boldsymbol{s}_{j,1}, \cdots, \boldsymbol{s}_{j,k-1}, \boldsymbol{s}_{j,k+1}, \cdots, \boldsymbol{s}_{j,K}] \tag{13}$$

$p(\boldsymbol{s}_{j,k})$ がガウス分布 $\mathcal{N}(\mathbf{0}, \boldsymbol{\Phi}_\mathbf{0})$ であると仮定すると、$\boldsymbol{s}_{j,k}$ の条件付き分布は次式のようになる。

$$p(\boldsymbol{s}_k|\boldsymbol{Z}, \boldsymbol{\theta}, \tilde{\boldsymbol{S}}, \boldsymbol{K}) = \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Phi}) \tag{14}$$

$$\boldsymbol{\Phi}_j = \left(\boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{A}_j + \boldsymbol{\Phi}_0^{-1}\right)^{-1} \tag{15}$$

$$\boldsymbol{\mu}_{j,k} = \boldsymbol{\Phi}_j \boldsymbol{A}_j^H \boldsymbol{K}_j^{-1} \boldsymbol{z}_{j,k} \tag{16}$$

式 (6) と式 (14) を比較すると、条件付き平均値 $\boldsymbol{\mu}_k$ は、最尤推定値 $\hat{\boldsymbol{s}}_k$ に事前分布に関する項 $\boldsymbol{\Phi}_0^{-1}$ を加えたものとなっていることがわかる。

3.2　$\boldsymbol{K}_j$ の条件付き分布

共分散行列 $\boldsymbol{K}_j$ は、次式の inverse-Wishart 分布に従うものと仮定する。

$$\begin{aligned} p(\boldsymbol{K}_j) &= \text{inv-Wishart}\left(\boldsymbol{K}_j; \nu_0, (\nu_0 \boldsymbol{K}_0)^{-1}\right) \\ &\propto |\boldsymbol{K}_j|^{-(\nu_0+M)} \exp\left\{-\mathrm{tr}(\nu_0 \boldsymbol{K}_0 \boldsymbol{K}_j^{-1})\right\} \end{aligned} \tag{17}$$

これから、$\boldsymbol{K}_j$ の条件付き分布も、次式に示すような inverse-Wishart 分布となる。

$$\begin{aligned} & p(\boldsymbol{K}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j, \boldsymbol{\theta}_j) \\ &\propto p(\boldsymbol{K}_j) p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j, \boldsymbol{S}_j, \boldsymbol{K}_j) \\ &\propto |\boldsymbol{K}_j|^{-(\nu_0+M)} \exp\left\{-\mathrm{tr}(\nu_0 \boldsymbol{K}_0 \boldsymbol{K}_j^{-1})\right\} \\ &\quad \times |\boldsymbol{K}_j|^{-K} \exp\left\{-\mathrm{tr}(\boldsymbol{C}_j \boldsymbol{K}_j^{-1})\right\} \\ &= |\boldsymbol{K}_j|^{-(\nu_0+M+K)} \exp\left\{-\mathrm{tr}([\nu_0 \boldsymbol{K}_0 + \boldsymbol{C}_j] \boldsymbol{K}_j^{-1})\right\} \\ &\propto \text{inv-Wishart}\left(\nu_0 + K, [\nu_0 \boldsymbol{K}_0 + \boldsymbol{C}_j]^{-1}\right) \end{aligned} \tag{18}$$

3.3　$\boldsymbol{\theta}_j$ の条件付き分布

$\boldsymbol{A}_j(\boldsymbol{\theta}_j)$ が $\boldsymbol{\theta}_j$ の非線形関数であることから、$\boldsymbol{\theta}_j$ のサンプルをその条件付き分布 $p(\boldsymbol{\theta}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j, \boldsymbol{K}_j)$ から直接得るのは一般に困難である。このような場合、メトロポリス・アルゴリズム [3, 6] により、$\boldsymbol{\theta}_j$ のサンプルを得る手法が一般的である。メトロポリス・アルゴリズムでは、前回 (p) の反復によりサンプル $\boldsymbol{\theta}_j^{(p)}$ が得られているものとし、提案分布 $J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)})$ から新たなサンプル $\boldsymbol{\theta}_j^*$ を得る。本報告では、次式の一様分布を提案分布として用いる。

$$J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)}) = \mathcal{U}(\boldsymbol{\theta}_j^{(p)} - \boldsymbol{\delta}, \boldsymbol{\theta}_j^{(p)} + \boldsymbol{\delta}) \tag{19}$$

ここで、$\boldsymbol{\delta}$ は、適当な定数ベクトルである。新たなサンプルは、次式の採択率が適当な閾値 r_{thr} を超えた場合に採用する。

$$r = \frac{p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j^*, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{K}_j^{(p+1)})}{p(\boldsymbol{Z}_j|\boldsymbol{\theta}_j^{(p)}, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{K}_j^{(p+1)})} \frac{p(\boldsymbol{\theta}_j^*)}{p(\boldsymbol{\theta}_j^{(p)})} \tag{20}$$

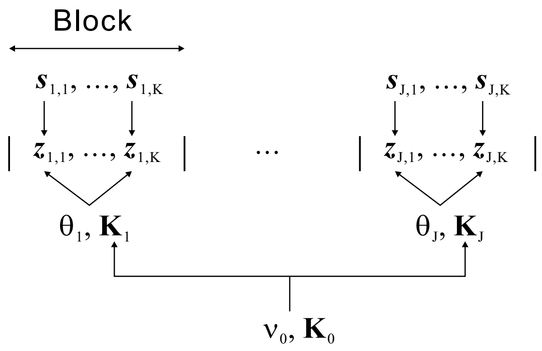

Figure 3: Hierarchical model of the covariance matrix.

3.4 ギブスサンプリング

ここでは、前節までで述べた各パラメタの条件付き分布から反復的にサンプルを得るギブスサンプリングについて述べる。ギブスサンプリングの手続きを以下にまとめる。

1. 初期値 $\boldsymbol{K}_j^{(1)}$ および $\boldsymbol{\theta}_j^{(1)}$ を設定する。

2. $\boldsymbol{s}_{j,k}$ のサンプルを得る。

$$\boldsymbol{s}_{j,k}^{(p+1)} \sim p(\boldsymbol{s}_{j,k}|\boldsymbol{Z}_j, \boldsymbol{\theta}_j^{(p)}, \tilde{\boldsymbol{S}}_{j_k}^{(p)}, \boldsymbol{K}_j^{(p)}) \quad \forall k$$

3. $\boldsymbol{K}_j$ のサンプルを得る。

$$\boldsymbol{K}_j^{(p+1)} \sim p(\boldsymbol{K}_j|\boldsymbol{Z}_j, \boldsymbol{S}_j^{(p+1)}, \boldsymbol{\theta}_j^{(p)})$$

4. $\boldsymbol{\theta}_j$ のサンプルを得る。

$$\begin{aligned} \boldsymbol{\theta}_j^* &\sim J(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(p)}) \\ \boldsymbol{\theta}_j^{(p+1)} &= \begin{cases} \boldsymbol{\theta}_j^* & r > r_{thr} \\ \boldsymbol{\theta}_j^{(p)} & \text{otherwise} \end{cases} \end{aligned}$$

5. $p \leftarrow p+1$ として、ステップ 2 に戻る。

4 階層ベイス推定

4.1 共分散行列の階層モデル

冒頭で述べたように、共分散行列は、J 個の観測値 $\{\boldsymbol{Z}_1, \cdots, \boldsymbol{Z}_J\}$ に共通する項と個々の観測値に特有の項に分解されるものと仮定する。このうち、共通する項を階層モデルを用いて推定する。共分散行列は、次式のサンプリングモデルに従うと仮定する。

$$\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J \sim \text{ i.i.d. inv-Wishart}(\nu_0, (\nu_0 \boldsymbol{K}_0)^{-1}) \quad (21)$$

図 3 は、このモデルを図示したものである。式 (21) において、$\boldsymbol{K}_0$ が観測値に共通する項に相当する。ν_0 は、$\boldsymbol{K}_0$ に対する仮想的なサンプル数であり、式 (18) からわかるように、各データに特有の項に対する共通項の重みの役割を果たす。

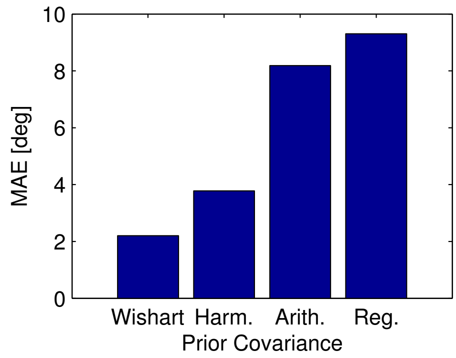

Figure 4: MAE for different estimation method for $\boldsymbol{K}_0$.

4.2 $\boldsymbol{K}_0$ の条件付き分布

$p(\boldsymbol{K}_0) = \text{Wishart}(\eta, \boldsymbol{\Psi})$ であると仮定すると (inverse-Wishart ではないことに注意)、$\boldsymbol{K}_0$ の条件付き分布は次式のようになる。

$$\begin{aligned} & p(\boldsymbol{K}_0|\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J, \nu_0) \\ \propto\ & p(\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J|\boldsymbol{K}_0, \nu_0) p(\boldsymbol{K}_0, \nu_0) \\ =\ & p(\boldsymbol{K}_0) \prod_{i=1}^{J} p(\boldsymbol{K}_j|\boldsymbol{K}_0, \nu_0) \\ \propto\ & |\boldsymbol{\Psi}|^{-\eta} |\boldsymbol{K}_0|^{\eta - M} \exp\left\{-\mathrm{tr}(\boldsymbol{K}_0 \boldsymbol{\Psi}^{-1})\right\} \times \\ & \prod_{j=1}^{J} |\boldsymbol{K}_0^{-1}|^{-\nu_0} |\boldsymbol{K}_j|^{-(\nu_0+M)} \exp\left\{-\mathrm{tr}(\nu_0 \boldsymbol{K}_0 \boldsymbol{K}_j^{-1})\right\} \\ \propto\ & |\boldsymbol{K}_0|^{\eta + J\nu_0 - M} \exp\left\{-\mathrm{tr}\left(\boldsymbol{K}_0 \boldsymbol{\Lambda}^{-1}\right)\right\} \\ =\ & \text{Wishart}\left(\boldsymbol{K}_0; \eta + J\nu_0, \boldsymbol{\Lambda}\right) \end{aligned} \quad (22)$$

ここで、

$$\boldsymbol{\Lambda} := \left(\boldsymbol{\Psi}^{-1} + \nu_0 \sum_{j=1}^{J} \boldsymbol{K}_j^{-1}\right)^{-1} \quad (23)$$

$\boldsymbol{\Psi}$ の影響が小さい場合、式 (23) は、$\{\boldsymbol{K}_j\}$ の調和平均を求めるような操作となり、この値（の定数倍）が式 (22) に示す Wishart 分布から得られるサンプルの平均値となる。

4.3 ν_0 の条件付き分布

ν_0 の事前分布に次式を仮定すると、

$$p(\nu_0) \propto \exp(-\alpha \nu_0) \quad (24)$$

条件付き分布は次式のようになる。

$$\begin{aligned} & p(\nu_0|\boldsymbol{K}_0, \boldsymbol{K}_1, \cdots, \boldsymbol{K}_J) \\ \propto\ & p(\nu_0) p(\boldsymbol{K}_1, \cdots, \boldsymbol{K}_J|\nu_0, \boldsymbol{K}_0) \\ \propto\ & \exp\left(-\alpha\nu_0\right) \prod_{j=1}^{J} \frac{|\nu_0 \boldsymbol{K}_0|^{\nu_0}}{\Gamma_M(\nu_0)} |\boldsymbol{K}_j|^{-(\nu_0+M)} \times \\ & \exp\left\{-\mathrm{tr}(\nu_0 \boldsymbol{K}_0 \boldsymbol{K}_j^{-1})\right\} \end{aligned} \quad (25)$$

ここで、$\Gamma_M(\nu_0) = \pi^{M(M-1)/2} \prod_{m=1}^{M} \Gamma(\nu_0 - m + 1)$ である。$\Gamma(\cdot)$ はガンマ関数を表す。

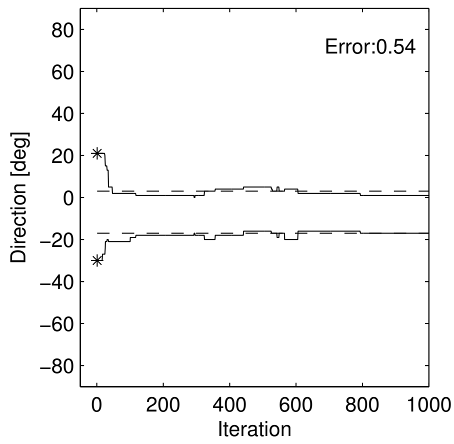

Figure 5: Variation of sample $\boldsymbol{\theta}^{(p)}$ in the Gibbs sampling.

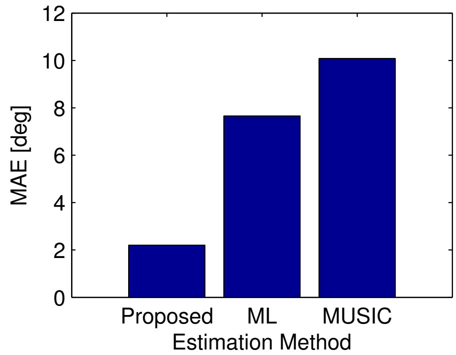

Figure 6: MAE for different parameter estimation method. H-Bayes:hierarchical Bayesian estimator、ML:maximum likelihood estimator、MUSIC:MUSIC estimator。

4.4 反復アルゴリズム

ここでは、$\boldsymbol{K}_0$ および ν_0 のサンプルを得る手続きをまとめる。

1. 初期値 $\boldsymbol{K}_0^{(1)}$ および $\nu_0^{(1)}$ を設定する。
2. 3.4 節で述べたアルゴリズムにより、$\{\boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)}\}$ を得る。
3. $\boldsymbol{K}_0$ のサンプルを得る。
$$\boldsymbol{K}_0^{(p+1)} \sim p(\boldsymbol{K}_0 | \boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)}, \nu_0^{(p)})$$
4. ν_0 のサンプルを得る。
$$\nu_0^{(p+1)} \sim p(\nu_0 | \boldsymbol{K}_0^{(p+1)}, \boldsymbol{K}_1^{(p+1)}, \cdots, \boldsymbol{K}_J^{(p+1)})$$
5. $p \leftarrow p+1$ として、ステップ 2 に戻る。

5 評価実験

5.1 実験 I - 静止音源

実験 1 では、静止音源の環境のシミュレーションとして、20 通りの音源配置 $(J = 20)$ に対するインパルス応答をガウス雑音に畳み込み、観測値を生成した。インパルス応答は、2.2 節で用いたもので同様である。音源間隔を 20° とした音源のペアを用い、音源ペアの方向を乱数を用いて 20 通りに変化させた。分析パラメタは、ブロック長を 3200 点 (0.2 秒) とした以外は、表 1 と同じである。反復アルゴリズムの反復回数は 1000 回とした。パラメタ $\boldsymbol{\theta}$ の初期値としては、最尤法に白色雑音を加えたものを用いた。

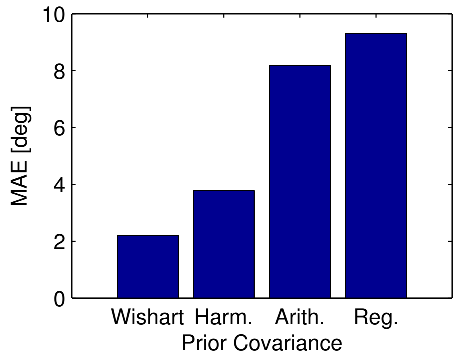

Figure 7: MAE for different estimation method for $\boldsymbol{K}_0$.

図 5 は、反復におけるパラメタ θ の変化の例を表している。この図から、この例では、比較的少ない反復で、推定値が真値に収束しているのがわかる。

図 6 は、階層ベイズ推定を用いた音源定位を、最尤法、MUSIC 法と比較したものである。最尤法および MUSIC 法では、雑音を空間的に白色と仮定してる。図の縦軸は平均絶対誤差 MAE $= 1/N_{avg} \sum |\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|$ であり、20 種の配置 $(J = 20)$ と 30 回のトライアル $(N_{trial} = 30)$ について平均した $(N_{avg} = J \times N_{trial})$。この図から、階層ベイズ推定を用いて雑音の共分散行列を推定した場合の方が、他の 2 つの推定法よりも誤差が小さいのがわかる。

図 7 は、$\boldsymbol{K}_0$ のサンプルを得る方法を変えた場合の影響である。同図の“Wishart”は、4.2 節で述べた Wishart 分布からサンプルを得た場合、“Harm.”は式 (23) における $\boldsymbol{\Lambda}$ を用いた場合、“Arith.”は算術平均 $\sum_j \boldsymbol{K}_j$ を用いた場合、“Reg.”は $\sigma\boldsymbol{I}$ を用いた場合である。“Reg.”の場合は、共分散行列の階層推定は行われず、$\boldsymbol{K}_0 = \sigma\boldsymbol{I}$ の効果は、$\boldsymbol{K}_j$ の正則化となる。この図から、Wishart 分布からサンプルを得た場合、およびその平均値である $\boldsymbol{\Lambda}$ を用いた場合は、MAE が小さいことがわかる。このことから、$\boldsymbol{K}_0$ を $\{\boldsymbol{K}_j\}$ から階層推定することの本質は、式 (23) における調和平均のような操作であることがわかる。

図 8 は ν_0 を単一の値に固定して、MAE を算出したものである。この図から、おおむね $\nu_0 = 10^2$ 付近で MAE は最小値をとる。

図 9(b) は、階層ベイズ推定により求めた $\boldsymbol{K}_0$ の推定精度を評価するため、推定された $\boldsymbol{K}_0$ を用いて一般化固有値問題 $(\boldsymbol{C}_{z,j}, \boldsymbol{K}_0)$ を解き、MUSIC スペクトルを求めたものである。比較のために示した、標準固有値分解（雑音の白色性を仮定）の場合 (同図 (a)) と比較すると、空間分解能が向上しているのがわかる。

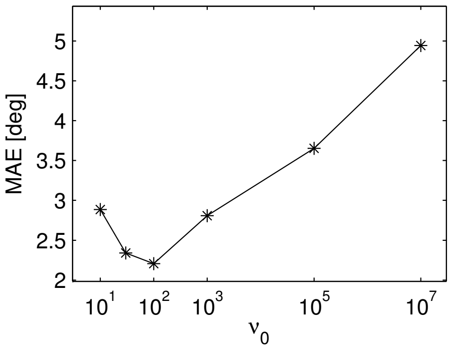

Figure 8: MAE for different ν_0 values.

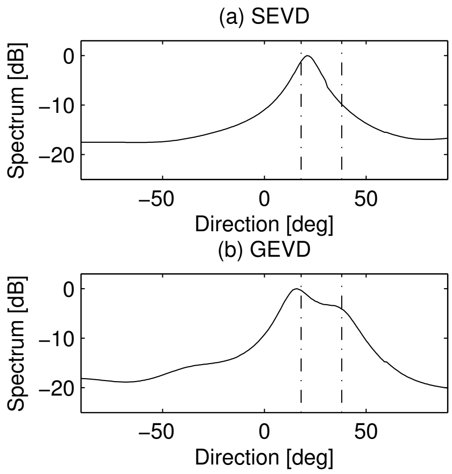

Figure 9: Evaluation of the estimate of $\boldsymbol{K}_0$ using MUSIC spatial spectral estimator.

5.2 実験 II - 移動音源

実験 II では、静止したロボットの周りを 2 名の話者が歩きながら発話した音声を収録し、観測値として用いた。用いたマイクロホンアレイは、ロボット (Honda Hearbo) に搭載した 8 素子のものである。2 名の話者は、約 30° の間隔を保ちながら、ロボットの周囲の半径 1.5m の円周上を等速運動した。分析条件は実験 I と同様である。観測データセットは、連続した 20 ブロックである。ギブスサンプリングの初期値には MUSIC 法の推定値を用いた。実験 I では、周波数を 1500Hz の単一周波数としたが、実験 II では、信号源が音声信号であり、周波数領域でのスパース性のため、周波数により結果が異なる。そこで、800Hz から 3000Hz までの 71 離散周波数について推定を行った。したがって、音源方向の推定値は、20 ブロック × 71 周波数 =1420 の観測データごとに評価を行った。

図 10 は推定誤差をまとめたものである。上述の音声のスパース性のため、1420 の観測データセットについて MAE を計算すると、実効音源数が 0 や 1 の周波数とブロックの組み合わせの場合に、大きな誤差となり、手法の比較が難しい。そこで、ここでは、誤差が ±8° 以内に入るデータの全データに対する割合を評価値として用いた。実験 II では、利用できるアレイマニフォールドベクトルが 5° ごと

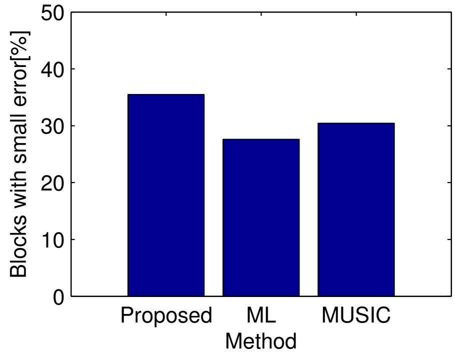

Figure 10: Percentage of blocks with small error .

である。このため、±8° の誤差は、マニフォールドベクトルに換算して概ね 0 ±1 ユニットの誤差に相当する。同図から、階層ベイス推定を行った場合は、ML 法、MUSIC 法に比べ、数%程度推定精度が改善しているのがわかる。

6 結論

本報告では、雑音の共分散行列を、小数のデータからでも安定的に推定する手法として、階層ベイズ推定を用いた手法を検討した。この結果、推定した雑音の共分散行列を用いて音源定位を行うことにより、空間分解能が向上することが示された。今後の課題としては、音源数 N も含めた同時推定の手法の開発が望まれる。

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] F. Asano and H.Asoh, “Joint estimation of sound source location and noise covariance in spatially colored noise,” in *Proc. Eusipco 2011*, 2011.

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

[4] Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, “Common-acoustical-pole and zero modeling of head-related transfer function,” *IEEE Trans. on Speech and Audio Processing*, vol. 7, no. 2, pp. 188–196, 1999.

[5] C. Bishop, *Pattern recognition and machine learning*, Springer, 2006.

[6] 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.