音響信号を対象とした信号分離

(DIagonalization using Equivalent Matrices(DIEM) を用いた同時対角化問題の解法)

西山 慶 (古川利博 教授,　高橋智博 助教)

1　はじめに

入手可能な情報が観測信号のみで，元の信号源が統計的に独立であるという仮定の下で，観測信号から元の信号を取り出す技術としてブラインド音源分離 (Blind Source Separation:BSS) が知られている．所望の音声が周囲の雑音に埋もれて聞き取りにくくなる場合，BSS を用いて混合音声の中から特定の音声だけを抽出して取り出すことができれば，様々な場面での応用が期待できる．例えば，ノイズキャンセルを行ったり，BSS を前処理として行うことで音声認識や音声記録の精度や品質を高めることができる．

文献 [1] では，最適化問題の分野の 1 つである同時対角化問題の解法に DIEM を提唱している．DIEM は，計算に反復を必要としないため，演算量が少なくて解が求まるという特徴をもつ．そこで本研究では，同時対角化問題の解法の一種である DIEM を用いて新しい BSS の提案をし，その特性を評価する．

2　BSS 問題の定式化

図 1 のように 1 つの部屋に 2 つの信号源と 2 つのマイクロフォンがある状況について考える．時刻tにおける信号源が統計的に独立な信号源ベクトルを $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_J(t)]^T$, 観測信号ベクトルを $\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_I(t)]^T$,$s_j$ から x_i までのインパルス応答を $a_{ij}(\tau)$，雑音ベクトルを $\boldsymbol{n}(t) = [s_1(t), s_2(t), \ldots, s_I(t)]^T$ とする.$\boldsymbol{n}(t)$ は平均 0 のガウス性白色雑音とする. 今研究で扱う問題は瞬時混合であり，この場合 a_{ij} のインパルス応答長は 1 となり，(τ) は省略する．以上のことを踏まえて，図 1 の信号源の個数が J 個，マイクロフォンの個数が I 個のときの観測信号と信号源と雑音の関係を定式化すると (1) 式のようになる．

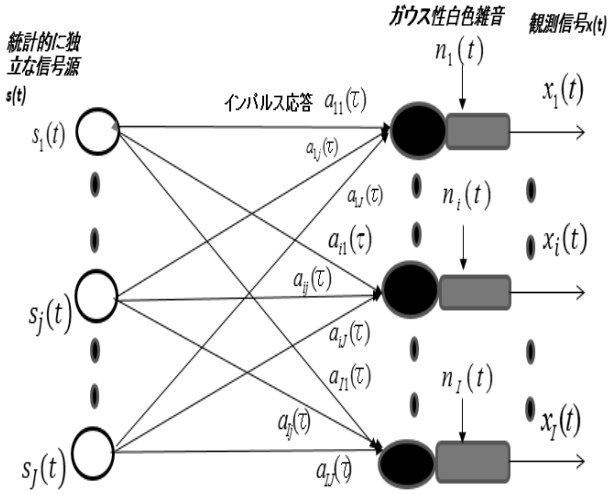

図 1　畳み込み混合モデル

$$\begin{pmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_I(t) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1J} \\ a_{21} & a_{22} & \ldots & a_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ a_{I1} & a_{I2} & \ldots & a_{IJ} \end{pmatrix} \begin{pmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_J(t) \end{pmatrix} + \begin{pmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_I(t) \end{pmatrix}$$

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{1}$$

$\boldsymbol{A}$ は混合行列とよばれ，$\boldsymbol{A} \in \mathbb{R}^{I \times J}$ である．$\boldsymbol{A}$ は正則を仮定し，信号源の個数が既知である仮定で信号源の個数 J とマイクロフォンの個数 I の関係は $I \geqq J \geqq 2$ が成り立つとする．BSS は観

測信号 $\boldsymbol{x}(t)$ のみから未知の信号源 $\boldsymbol{s}(t)$ を得る技術であり，以下の (2) 式に示す分離行列 $\boldsymbol{W}$ を推定して $\hat{\boldsymbol{s}}(t)$ を得ることになる．以下の (2) 式により，本研究では $\boldsymbol{W} = \hat{\boldsymbol{A}}^{-1}$ とする．

$$\min_{\boldsymbol{W}} \parallel \boldsymbol{W}\boldsymbol{A} - \boldsymbol{I} \parallel_F^2, \quad \hat{\boldsymbol{s}}(t) = \boldsymbol{W}(\boldsymbol{A}\boldsymbol{s}(t) + \boldsymbol{n}(t)) \tag{2}$$

3 同時対角化問題と DIEM

3.1 同時対角化問題

$\boldsymbol{T}_x(k) = \boldsymbol{A}\boldsymbol{\Lambda}_s(k)\boldsymbol{A}^T$ に対し，正則な行列 $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ と，K 個の必ずしも正則とは限らない対角行列 $\boldsymbol{\Lambda}_s(k) \in \mathbb{R}^{N\times N}$ が以下の K 個の等式を全て満たす場合を考える．

$$\boldsymbol{T}_x(k) = \boldsymbol{A}\boldsymbol{\Lambda}_s(k)\boldsymbol{A}^T \quad (k = 1, \ldots, K)$$

この K 個の等式を全て満たすとき，K 個の $\boldsymbol{T}_x(k) \in \mathbb{R}^{N\times N}$ は同時対角可能な行列と定義される．同時対角化問題とは，$\{\boldsymbol{T}_x(k)\}_{k=1}^K$ しかない状況で K 個の等式を満たす正則行列 $\boldsymbol{A}$ を推定する問題である．

3.2 DIEM

文献 [1] では，同時対角化問題の解法の一種として DIEM が提唱されている．DIEM は計算に反復を必要とせず，また演算量が少ないという特徴をもつ． $\hat{\boldsymbol{T}}_x(k) \in \mathbb{R}^{N\times N}$ とすれば，(3) 式は $\boldsymbol{A}$ に関して 4 乗，N 次元の最小化問題となる．しかし，$\boldsymbol{A}$ に関して 4 乗のままでは解が一意に定まらないので (3) 式から変形を施す．ここで (3) 式と (4) 式は等価な関係になる．

$$C_{DLS}(\boldsymbol{A}, \{\boldsymbol{\Lambda}_s(k)\}_{k=1}^K) = \sum_{k=1}^{K} \parallel \hat{\boldsymbol{T}}_x(k) - \boldsymbol{A}\boldsymbol{\Lambda}_s(k)\boldsymbol{A}^T \parallel_F^2 \tag{3}$$

$$C_{DLS}(\boldsymbol{A}, \{\boldsymbol{\Lambda}_s(k)\}_{k=1}^K) = \sum_{k=1}^{K} \parallel vec(\hat{\boldsymbol{T}}_x(k)) - (\boldsymbol{A} \circ \boldsymbol{A})\boldsymbol{\delta}_k \parallel^2 \tag{4}$$

$\boldsymbol{\delta}_k$ は $\boldsymbol{\Lambda}_s(k)$ の対角成分を縦に N 個並べたベクトルである．
$\hat{\boldsymbol{t}}_k \triangleq vec(\hat{\boldsymbol{T}}_x(k))$,$\dot{\boldsymbol{A}} \triangleq \boldsymbol{A} \circ \boldsymbol{A}$ ($\circ$ は Khatri-Rao 積) を用いて (4) 式を変形すると (5) 式になる．

$$C_{DLS}(\boldsymbol{A}, \{\boldsymbol{\Lambda}_s(k)\}_{k=1}^K) = \sum_{k=1}^{K} \parallel \hat{\boldsymbol{t}}_k - \dot{\boldsymbol{A}}\boldsymbol{\delta}_k \parallel^2 \tag{5}$$

(5) 式を最小にする $\boldsymbol{\delta}_k$ は,$\boldsymbol{\delta}_k = \dot{\boldsymbol{A}}^{\dagger}\hat{\boldsymbol{t}}_k$($\dot{\boldsymbol{A}}^{\dagger}$ は，$\dot{\boldsymbol{A}}$ の疑似逆行列) なので (6) 式に書き換えられる．

$$C_{DLS}(\dot{\boldsymbol{A}}) = \sum_{k=1}^{K} \parallel \hat{\boldsymbol{t}}_k - \dot{\boldsymbol{A}}\dot{\boldsymbol{A}}^{\dagger}\hat{\boldsymbol{t}}_k \parallel^2 \tag{6}$$

$P^{\perp}(\boldsymbol{A}) \triangleq \boldsymbol{I} - \dot{\boldsymbol{A}}\dot{\boldsymbol{A}}^{\dagger}$ とおくと、$\boldsymbol{A}$ を解く最小化問題は次の (7) 式に書き換えることができる。

$$C_{DLS}(\boldsymbol{A}) = trace\left(P^{\perp}(\boldsymbol{A})\sum_{k=1}^{K}\hat{\boldsymbol{t}}_k\hat{\boldsymbol{t}}_k^T\right) \tag{7}$$

(6) 式により $\boldsymbol{\Lambda}_s(k)$ を K 個求める必要がなくなる．また，$\dot{\boldsymbol{A}}$ に関して 2 乗，N^2 次元の最小化問題になる．これは，(3) 式が N 次元の最小化問題であることから最小値は一致しない．そこで，(7) 式の $\hat{\boldsymbol{T}} \triangleq \sum_{k=1}^{K}\hat{\boldsymbol{t}}_k\hat{\boldsymbol{t}}_k^T \in \mathbb{R}^{N^2\times N^2}$ の N^2 個の固有値 ($\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_N \geqq \ldots \geqq \lambda_{N^2}$) の中の影響力の大きい N 個の固有値 $\lambda_1, \ldots, \lambda_N$ に対応する固有ベクトル $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N$ を求める．その各固有ベクトルの $jN+1$ 行成分から $(j+1)N$ 行成分 ($j = 0, \ldots, N-1$) を，1 行 $j+1$ 列成分から N 行 $j+1$ 列成分まで上から N 個代入すれば，N 次元の最小化問題として解くことができる．

4　同時対角化問題の音源分離への適用 (提案)

音声は様々な音素から構成されていて信号の統計的性質が時間とともに変化する非定常信号であるが，短時間区間 (epoch) では定常信号とみなすことができる．これを信号の quasistationary 性という．信号源の平均値が $\mathbf{0}$，信号源が統計的に独立であること，quasistationary 性の 3 つの仮定の下，音声信号を信号源として用いた環境下で，(2) 式で示した BSS の混合行列 $\boldsymbol{W}$ の推定に同時対角化問題の解法の一種の DIEM を用いることを提案する．

時刻 t_0,t_1,t_2,...,t_K のそれぞれを epoch の区切りとする．時刻 t_0 から t_1 までの信号は定常であり，同様に各 epoch の t_{k-1} から t_k までの区間がすべて定常であるとする．このとき t_{k-1} から t_k までの観測信号の共分散行列 $\boldsymbol{T}_x(k)$ を (1) 式の BSS の観測信号 $\boldsymbol{x}(t)$ を用いて (8) 式に定義する．

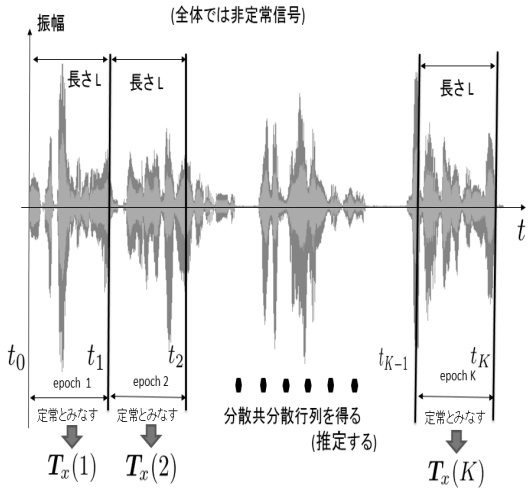

図 2　epoch 毎の共分散行列の模式図

$$\begin{aligned}\boldsymbol{T}_x(k) &= E[\boldsymbol{x}(t)\boldsymbol{x}(t)^T] \\ &= \boldsymbol{A}E[\boldsymbol{s}(t)\boldsymbol{s}^T(t)]\boldsymbol{A}^T + E[\boldsymbol{n}(t)\boldsymbol{n}(t)^T] \\ &= \boldsymbol{A}\boldsymbol{\Lambda}_s(k)\boldsymbol{A}^T + \boldsymbol{\Lambda}_n \qquad (L(k-1) < t < Lk)\end{aligned} \tag{8}$$

信号源ベクトル $\boldsymbol{s}(t) = [s_1(t), s_2(t), \ldots, s_J(t)]^T$ は，それぞれ互いに独立を仮定しているために (8) 式で示した対角共分散行列 $\boldsymbol{\Lambda}_s(k)$ が成立している．E[·] は期待値を表している．実際には共分散行列は未知なので，長さ L の epoch の 1 つ毎に時間平均をとって (9) 式の推定値 $\hat{\boldsymbol{T}}_x(k)$ を得る．これを (3) 式に代入して DIEM を用いて混合行列 $\hat{\boldsymbol{A}}$ を求める．したがって，BSS の目標である (2) 式の $\boldsymbol{W} = \hat{\boldsymbol{A}}^{-1}$ および $\hat{\boldsymbol{s}}(t)$ を求めることができる．

$$\hat{\boldsymbol{T}}_x(k) = \frac{1}{L}\sum_{t=L(k-1)}^{Lk-1} \boldsymbol{x}(t)\boldsymbol{x}^T(t) \tag{9}$$

5　計算機シミュレーション

5.1　K(個数)，L(長さ) を変化させたときの DIEM の分離精度の比較

信号源とマイクロフォンの個数は 2 個ずつとして，混合行列は，$\boldsymbol{A} = \begin{pmatrix} -0.395 & 0.276 \\ 0.187 & 0.695 \end{pmatrix}$ を用いた．$\boldsymbol{A}$ の成分は平均 0，分散 1 のガウス乱数により与えた．信号源には日本音響学会研究用連続音声データベース (ASJ-JIPDEC) を用いた [2]．その中の同一人物の短文読み上げ音声を繋げて長い音声データ (男性と女性の声を 2 つずつ) 用意し，(男性 1, 男性 2),(女性 1, 女性 2),(男性 1, 女性 1) の 3 通りをシミュレーションに用いた．今回はサンプリングレート 8kHz，最大 130 分のデータを用いた．$\hat{\boldsymbol{W}}\boldsymbol{A} = \boldsymbol{G}$ としたとき，分離精度を示す評価量として (10) 式に示した performance index(PI(G)) を用いて 3 通りの平均を算出した．ただし g_{ij} は $\boldsymbol{G}$ の i 行 j 列目の要素を表す．$PI(G)$ は $-\infty$[dB] に近いほど $\boldsymbol{G}$ が対角行列に近くなるため，分離精度が高いという指標になる．

以上の条件で，BSS の解法として DIEM を用いることが有効であるかどうかを確かめた．

$$PI(G) = 10\log_{10}\left(\frac{1}{J(J-1)}\left[\sum_i^J\left(\sum_j^J \frac{|g_{ij}|^2}{\max_l|g_{il}|^2} - 1\right) + \sum_j^J\left(\sum_i^J \frac{|g_{ij}|^2}{\max_l|g_{lj}|^2} - 1\right)\right]\right) \tag{10}$$

表 1　K,L を変化させたときの PI(G)　　(PI(G) の単位は dB)

K(epoch の個数) / L(epoch の長さ)	100	500	1000	5000	10000
0.0125 秒 (100 サンプル)	−33.60	−32.19	−34.25	−40.63	−40.09
0.125 秒 (1000 サンプル)	−45.89	−41.42	−45.79	−47.87	−54.45
1.25 秒 (10000 サンプル)	−37.60	−48.72	**−61.40**	−52.74	− − −
2.5 秒 (20000 サンプル)	−30.04	−31.65	−59.49	− − −	− − −

表 1 の結果から，DIEM を音声の混合モデルに用いることができ，有効であることがわかった．また K(epoch の個数)=1000,L(epoch の長さ)=10000 のときに PI(G) の値が最も低くなるため分離精度が高くなると考えられる．

5.2　SNR と分離精度に関するシミュレーション

表 1 により推定した最適なパラメータを用いて，表 2 で示した条件下で観測信号にノイズが混入したときのシミュレーションを行った．SIR は, 所望信号と干渉信号の比を表す評価量であり. 干渉が少ないと高くなる．8KHz のサンプリングレートのデータを用いたので epoch の 1 区間 1.25 秒となる．図 3 により，DIEM で推定した $\boldsymbol{W} = \hat{\boldsymbol{W}} = \hat{\boldsymbol{A}}^{-1}$ と完全な分離行列 $\boldsymbol{W} = \boldsymbol{W}_{opt} = \boldsymbol{A}^{-1}$ で SIR を比較したとき，ほぼ一致していることがわかる．

表 2　最適なパラメータを用いたときの評価量

K(epoch の個数)	1000
L(epoch の長さ)	1.25 秒 (10000 サンプル)
SNR	$\dfrac{E[(g_{ii}s_i(t))^2]}{E\left[\left(\sum_{j=1}^{N} w_{ij}n_j(t)\right)^2\right]}$
SIR	$\dfrac{E[(g_{ii}s_i(t))^2]}{E\left[\left(\sum_{j=1,j\neq i}^{N} g_{ij}s_j(t) + \sum_{j=1}^{N} w_{ij}n_j(t)\right)^2\right]}$

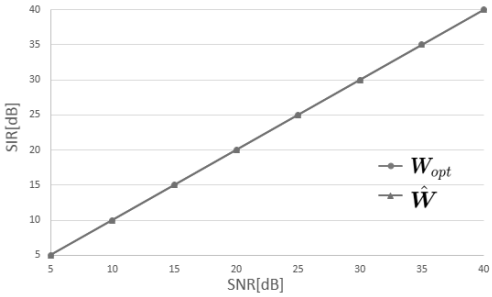

図 3　シミュレーション結果

6　まとめ

同時対角化問題の解法の一種である DIEM を用いて，ブラインド音源分離を行うことができた．また，観測信号にノイズが混入した状況下では，影響をほとんど受けないことが確認できた．これは，N^2 個の中から影響力の大きい N 個の固有ベクトルを取り出して最小化問題を考えた際，N 個の固有ベクトルは信号源の影響力が強く，$N^2 - N$ 個の固有ベクトルがもつノイズの影響は捨てて考えたため，ノイズにロバストな結果になったと考えられる．

今回，瞬時混合の研究を進めたが，畳み込み混合モデルにおいても DIEM の特性を解明したい．

参考文献

[1] Gills Chabriel and Jean Barrere, ”A Direct Algorithm for Non-orthogonal Approximate Joint Diadonalization,” *IEEE Trans. Signal Process.*, vol.60, no.1, pp.39-47, Jan. 2012.

[2] (社) 日本音響学会　連続音声データベース調査委員会,(財) 日本情報処理開発協会　知的音声処理調査委員会, ”研究用連続音声データベース (ASJ-JIPDEC),” (財) 日本情報処理開発協会.