

学位申請論文

近似同時対角化問題の解法を用いた
ブラインド信号源分離に関する研究

2019年9月

齋藤 晋哉

目次

第 1 章	序論	1
1.1	本研究の背景と目的	1
1.2	本論文の概要	2
第 2 章	畳み込み混合モデルとブラインド信号源分離	4
2.1	まえがき	4
2.2	周波数領域における畳み込みブラインド信号分離	4
2.2.1	畳み込み混合モデル	4
2.2.2	ブラインド信号分離	5
2.3	同時対角化問題	6
2.4	分離行列の推定	7
2.4.1	勾配法に基づく分離行列の推定	7
2.4.2	同時対角化問題解法に基づく分離行列の推定	7
2.5	パーミュテーション問題の解法	8
2.5.1	信号の到来方向を利用した信号源の割り当て	8
2.5.2	信号源信号の相関を利用した信号源の割り当て	8
2.6	スケーリング問題の解法	9
2.7	分離性能の指標と非パーミュテーションブラインド手法	10
2.8	Rahbar 氏の従来の ALSP アルゴリズムの概要と問題点	11
2.9	まとめ	11
第 3 章	最小 2 乗型同時対角化問題に基づくインダイレクトモデルブラインド信号源分離	12
3.1	まえがき	12
3.2	逆行列の展開式を用いたインダイレクトモデルブラインド信号分離	12
3.3	最小 2 乗型同時対角化問題のブロック処理	14
3.4	シミュレーション	16
3.5	まとめ	18
第 4 章	DMDI アルゴリズムを用いた同時対角化問題の解法に基づくブラインド信号源分離	20
4.1	まえがき	20
4.2	ブラインド信号分離のための DMDI アルゴリズムの導出	21

4.2.1	観測信号のクロススペクトル密度行列の推定と正規化	22
4.2.2	CDLS 評価量の最小化による分離行列の推定	23
4.2.3	正則な分離行列の構築	23
4.2.4	ILS 評価量の最小化による対角行列の推定	24
4.2.5	遅延分離行列	25
4.3	シミュレーションと実環境による実験	25
4.3.1	シミュレーション条件	25
4.3.2	評価基準と非パーミュテーションブラインド手法	28
4.3.3	残響環境下における性能比較	28
4.3.4	雑音環境下における性能比較	29
4.3.5	実環境における性能評価	29
4.4	まとめ	37
第 5 章	DMDI アルゴリズムの高速化手法とブラインド信号抽出への応用	41
5.1	まえがき	41
5.2	DMDI アルゴリズムの特徴と問題点	42
5.3	FDMDI アルゴリズムのブラインド信号抽出への適用	42
5.3.1	観測信号のクロススペクトル密度行列の推定	43
5.3.2	対角行列の推定	43
5.3.3	混合行列の推定	43
5.3.4	正則な分離行列の構築	44
5.3.5	信号源信号の復元	46
5.3.6	演算量の比較	47
5.4	シミュレーション	48
5.4.1	DMDI アルゴリズムと FDMDI アルゴリズムの基本性能を調べる事前シミュレーション	48
5.4.2	音声信号を用いた畳み込み混合におけるシミュレーション条件	50
5.4.3	Determined の場合	53
5.4.4	Overdetermined の場合	53
5.5	実環境における分離性能	54
5.6	まとめ	60
第 6 章	結論	61
6.1	本研究で得られた結果の概要	61
6.2	今後の課題	61
	謝辞	62

研究業績	69
付録	72
付録 A (3.2) の導出	72
付録 B 最小 2 乗型同時対角化問題のブロック処理と ALSP アルゴリズムによる解法	73
付録 C 観測信号のクロススペクトル密度行列の推定	75
付録 D 式 (4.3) の導出	76
付録 E 式 (5.12) の導出	78
付録 F FDMDI アルゴリズムの演算量	79
付録 G DMDI アルゴリズムの演算量	82

目次

1.1	学位論文の構成	2
2.1	畳み込み混合モデルとブラインド信号源分離	5
3.1	小部屋における音源とマイクロフォンの配置図	17
3.2	平均出力 SIR とブロック長 L の関係 ($J = N = 2$ 、 $M = 12000$ 、 $T_s = 8$ 、 $\text{SNR} \approx 20$ dB)	17
3.3	平均出力 SIR と SNR の関係 ($J = N = 2$ のとき平均入力 SIR は約 -0.14 dB、 $J = N = 3$ のとき約 -3.52 dB、 $J = N = 4$ のとき約 -5.87 dB)	19
4.1	周波数領域における混合・分離モデルと DMDI アルゴリズム (エポック m 、周波数ビン ω_k における J 個の観測信号は $\mathbf{x}(m) = \mathbf{B}\mathbf{s}(m)$ と表記している。周波数ビンの添字を $\mathbf{x}(m) \triangleq \mathbf{x}(m, \omega_k)$ 、 $\mathbf{B} \triangleq \mathbf{B}(\omega_k)$ 、 $\mathbf{s}(m) \triangleq \mathbf{s}(m, \omega_k)$ のように省略している。 N 個の信号源信号をベクトルで $\mathbf{s}(m)$ と表している。 \mathbf{B} は周波数ビン ω_k における混合行列である。また、 N 個の分離フィルタの出力信号を $\mathbf{y}(m) = \mathbf{W}\mathbf{x}(m) = \mathbf{W}\mathbf{B}\mathbf{s}(m)$ と表記している。ここで、周波数ビンの添字を $\mathbf{y}(m) \triangleq \mathbf{y}(m, \omega_k)$ 、 $\mathbf{W} \triangleq \mathbf{W}(\omega_k)$ のように省略している。 \mathbf{W} は周波数ビン ω_k における分離行列である。)	21
4.2	DMDI アルゴリズムのステップのダイアグラム	22
4.3	各音源の信号波形	27
4.4	小部屋における音源とマイクロフォンの配置図 ($N = 3$ と $N = 4$)	27
4.5	実環境におけるスピーカとマイクロフォンの配置図 ($N = 4$)	30
4.6	スピーカ A から図 4.5 に描かれた円の中心を通る垂線上で円の中心より下側に位置する無指向性マイクロフォンまでの室内空間インパルス応答とその減衰曲線	30
4.7	実環境において DMDI BSS によって推定した i 番目のスピーカと j 番目のマイクロフォン間の混合分離縦続接続フィルタの振幅特性 $ \gamma_{ij} $ (SNR ≈ 20 dB、8192 ポイント FFT、 $N = J = 4$)	35
4.8	実環境において DMDI BSS によって推定した i 番目のスピーカと j 番目のマイクロフォン間の混合分離縦続接続フィルタのインパルス応答 $\gamma_{ij}(t)$ (SNR ≈ 20 dB、8192 ポイント FFT、 $N = J = 4$)	38

4.9	音源信号、実環境における観測信号、DMDI BSS による分離信号のスペクトログラム (SNR \approx 20 dB、8192 ポイント FFT、 $N = J = 4$)	39
4.10	音源信号、DMDI BSS と ALUJA BSS による分離信号のスペクトログラム (SNR \approx 20 dB、8192 ポイント FFT、 $N = J = 4$)	40
5.1	$\text{cond}(\mathbf{H}) \leq 5, N = 5, J = 5$ における Determined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較	49
5.2	$\text{cond}(\mathbf{H}) \leq 5, N = 5, J = 7$ における Overdetermined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較	50
5.3	$\text{cond}(\mathbf{H})_{\text{free}}, N = 5, J = 5$ における Determined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較	50
5.4	$\text{cond}(\mathbf{H})_{\text{free}}, N = 5, J = 7$ における Overdetermined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較	51
5.5	$\text{cond}(\mathbf{H})_{\text{free}}, N = 5, J = 5$ から 8 における FDMDI アルゴリズムの分離性能と演算量の比較	51
5.6	小部屋における音源とマイクロフォンの配置図 ($N = 3$)	52
5.7	SNR \approx 20 dB、平均入力 SIR -3.16 dB、 $N = J = 3$ 、残響時間 300-ms におけるエポック数と平均出力 SIR、演算時間の比較	54
5.8	SNR \approx 20 dB、 $M = 400$ 、 $N = 3$ におけるブラインド信号抽出の平均出力 SIR、周波数ビン当たりの平均演算時間と反復回数、平均 PESQ の比較	56
5.9	SNR \approx 20 dB、 $M = 800$ 、 $N = 4$ におけるブラインド信号抽出の平均出力 SIR、周波数ビン当たりの平均演算時間と反復回数、平均 PESQ の比較	57
5.10	実環境におけるスピーカとマイクロフォンの配置図 ($N = 5$)	57
5.11	音源信号と FDMDI, ALUJA, PARAFAC の各アルゴリズムを用いた実環境におけるブラインド信号抽出法から得た出力信号のスペクトログラム (16,384 ポイント FFT、 $M = 1200$ 、 $\epsilon_C = 10^{-12}$ 、ALUJA においては図 5.10 における点 A, B, C, D, E のマイクロフォン使用)	59

目 次

3.1 勾配法に基づく従来のブラインド信号分離と提案法の分離性能の比較 ($J = N = 2$, $L = 200$, $\text{SNR} \approx 30 \text{ dB}$)	18
4.1 $N = J = 2$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 ($\text{SNR} \approx 20 \text{ dB}$, 8192 ポイント FFT)	31
4.2 $N = J = 3$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 ($\text{SNR} \approx 20 \text{ dB}$, 8192 ポイント FFT)	32
4.3 $N = J = 4$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 ($\text{SNR} \approx 20 \text{ dB}$, 8192 ポイント FFT)	33
4.4 残響時間 300 ms の室内環境における平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 (8192 ポイント FFT, $N = J = 2$)	34
4.5 実環境における平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 ($\text{SNR} \approx 20 \text{ dB}$, 8192 ポイント FFT, $N = J = 4$)	36
4.6 実環境における DMDI BSS と ALUJA BSS の平均 SDR と PESQ スコアの比較 ($\text{SNR} \approx 20 \text{ dB}$, 8192 ポイント FFT, $N = J = 4$)	37
5.1 ALS 反復当たりの演算量の比較	47
5.2 ALS 反復当たりの DMDI, FDMDI, PARAFAC の演算量の比較 ($p_{r1} = 1851.95$, $p_{r2} = 0.0016$, $p_f = 13.12$ については、残響時間 300 ms において DMDI と FDMDI アルゴリズムを実行した平均結果に基づき算出している。)	48
5.3 図 5.6 における音源とマイクロフォンの座標	52
5.4 $\text{SNR} \approx 20 \text{ dB}$, $M = 400$, $N = 3$, 音源とマイクロフォンの個数が等しい畳み込み混合におけるブラインド信号抽出の平均 PESQ、周波数ビン当たりの平均反復回数、演算時間、出力 SIR	55
5.5 $\text{SNR} \approx 20 \text{ dB}$, $M = 800$, $N = 4$, 音源とマイクロフォンの個数が等しい畳み込み混合におけるブラインド信号抽出の平均 PESQ、周波数ビン当たりの平均反復回数、演算時間、出力 SIR	56
5.6 図 5.10 におけるスピーカとマイクロフォンの座標	58

5.7 実環境におけるブラインド信号抽出の周波数ビン当たりの平均演算時間と反復回数、平均 PESQ の比較 (16,384 ポイント FFT、 $N = 5$ 、 $M = 1200$ 、 $\epsilon_C = 10^{-12}$ 、5 本のマイクロフォンは図 5.10 の点 A, B, C, D, E に配置、7 本のマイクロフォンは 5 本のマイクロフォンの配置に点 F, G に 2 本のマイクロフォンを追加)	58
F.1 FDMDI アルゴリズムの演算量	81
G.1 DMDI アルゴリズムの演算量	83

記号の定義

- 大文字太字と小文字太字のアルファベットは行列と列ベクトルをそれぞれ表す。
- 上付き添字 $(\cdot)^T$ と $(\cdot)^H$ は行列とベクトルの転置、エルミート行列をそれぞれ表す。
- 上付き添字 $(\cdot)^{-1}$ と $(\cdot)^*$ は正則行列の逆行列、行列並びにベクトルの複素共役をそれぞれ表す。
- 上付き添字 $(\cdot)^+$ は行列の疑似逆行列を表す。
- 記号 \mathbb{R}^N と \mathbb{C}^N は実数及び複素数 N 次元ベクトル空間をそれぞれ表す。
- 記号 $\mathbb{R}^{J \times N}$ と $\mathbb{C}^{J \times N}$ は実数及び複素数 $J \times N$ 次元行列空間をそれぞれ表す。
- ベクトルと行列を $\mathbf{a} \in \mathbb{C}^N$ と $\mathbf{A} \in \mathbb{C}^{J \times N}$ で表記すると、 $\|\mathbf{a}\|_2$ と $\|\mathbf{A}\|_F$ はそれぞれユークリッドノルムとフロベニウスノルムを表す。
- 単位行列は \mathbf{I} で表す。
- $E[\cdot]$ は期待値を表す。 j は虚数単位 $j = \sqrt{-1}$ を表す。
- $\text{Re}(a)$ と $\text{Im}(a)$ は複素数 a の実数部と虚数部をそれぞれ表す。
- $|\cdot|$ は絶対値を表す。
- $[x]$ は x に等しい、または、 x より小さい最大の整数値を表す。
- 演算子 $*$ は時間領域畳み込み演算を表す。
- $\mathbf{A} = \text{diagonal}(a_1, a_2, \dots, a_N)$ は対角要素 a_1, a_2, \dots, a_N の対角行列を表す。
- $\sigma_1(\mathbf{A})$ と $\sigma_i(\mathbf{A})$ は行列 \mathbf{A} の最大、 i 番目に大きい固有値を表す。
- $\text{rank}(\mathbf{A})$ は \mathbf{A} のランクを表す。
- $\text{tr}[\mathbf{A}]$ は \mathbf{A} のトレースを表す。
- $\text{Zdiag}(\mathbf{A})$ は行列 \mathbf{A} の対角要素を零に設定する演算を表す。
- $\text{diag}[\mathbf{A}]$ は行列 \mathbf{A} の対角要素を取り出して対角行列を生成する演算を表す。
- 演算子 \otimes と \odot はクロネッカ積とカトリー・ラオ積 (KR 積) をそれぞれ表す。

- $\text{vec}\{\mathbf{A}\}$ は $J \times J$ の行列の列ベクトルを順番に積み重ねて J^2 次元の列ベクトルを生成する演算を表す。
- $\text{vec}^{-1}\{\mathbf{a}\}$ は J^2 次元の列ベクトル \mathbf{a} から $J \times J$ の行列を生成する演算を表す。

第1章 序論

1.1 本研究の背景と目的

ブラインド信号分離 (blind source separation, BSS) とは混合過程が未知であることを前提に観測信号から信号源信号を復元することである。BSS は多数の手法がこれまでに提案されている。大きく分けると3つに大別される。1つ目は音声信号のスパース性を利用した時間周波数マスクングである [1]。信号がスパースであるとは信号のほとんどの時間周波数において0であることが前提である。この手法はクラスタリングを用いて音源の到来方向を推定する必要があるが、残響が存在する環境では、形成されるクラスタが重複してしまい音源毎に分類することができないことが報告されている。2つ目は非負値因子分解 (non-negative matrix factorization, NMF) に基づく手法で、基底となる個数のスペクトルとそれらに対応する強度の時間変化に分解する手法である [2]。この手法は信号源の周波数特性に依存し、同じ周波数特性の信号源同士の分解は困難である。3つ目は独立成分分析 (independent component analysis, ICA) で信号源同士が互いに統計的独立という仮定のみで分離が可能で、残響が存在する状況や信号源の周波数特性に依存しないことが特徴である。ICA の中でも行列の同時対角化 (joint diagonalization, JD) はブラインドシステム同定、ブラインド信号源分離、ブラインド信号抽出等に主に応用されている [3–12]。未知な M 個の対角行列を $\{\mathbf{P}_s(m)\}_{m=1}^M$ で定義すると、 $\mathbf{T}(m) = \mathbf{H}\mathbf{P}_s(m)\mathbf{H}^H$ の関係で与えられる M 個の正方なエルミート行列 $\{\mathbf{T}(m)\}_{m=1}^M$ を対角化する行列 \mathbf{H} を求めることが同時対角化問題である。

混合モデルでは、時刻 t において未知の信号源信号 $\mathbf{s}(t) \in \mathbb{R}^N$ が未知系 \mathbf{H} で混合され、観測信号 $\mathbf{x}(t) = \mathbf{H}\mathbf{s}(t)$ として受信される。ここで、 $\mathbf{s}(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ で、信号源信号 $\{s_k(t)\}_{k=1}^N$ は互いに無相関とする。信号源信号が平均零で疑似的な定常過程であると仮定すると [8, 9, 11, 13, 14]、 $\mathbf{P}_s(m)$ は $\mathbf{P}_s(m) = E[\mathbf{s}(t)\mathbf{s}(t)^T]$, $\forall t \in [(m-1)T_b, mT_b - 1]$ と定義される。ここで、 T_b はエポックサイズと呼ばれ、定常過程が成り立つ区間である。ブラインド信号源分離では、できる限り $\{\mathbf{T}(m)\}_{m=1}^M$ が対角行列になるように混合行列 \mathbf{H} を推定する。一方、ブラインド信号抽出では、同時対角化問題の解法によって分離行列 \mathbf{W} を求め、 $\hat{\mathbf{s}}(t) = \mathbf{W}\mathbf{x}(t)$ によって信号源信号を推定する。

同時対角化問題の解法は、直交同時対角化手法 (orthogonal JD approach, OJD approach) と非直交同時対角化手法 (non-orthogonal JD approach, NOJD approach) に大別される。直交同時対角化手法では、観測信号の白色化を必要とし、ユニタリ行列を計算する手法に基づく [4, 15]。一方、非直交同時対角化手法では、観測信号を白色化することなく反復法が用いられる [5, 12, 16–28]。実際には、雑音、推定誤差、入手個数の制限などで $\{\mathbf{T}(m)\}_{m=1}^M$ の代わりに推定値 $\{\hat{\mathbf{T}}(m)\}_{m=1}^M$

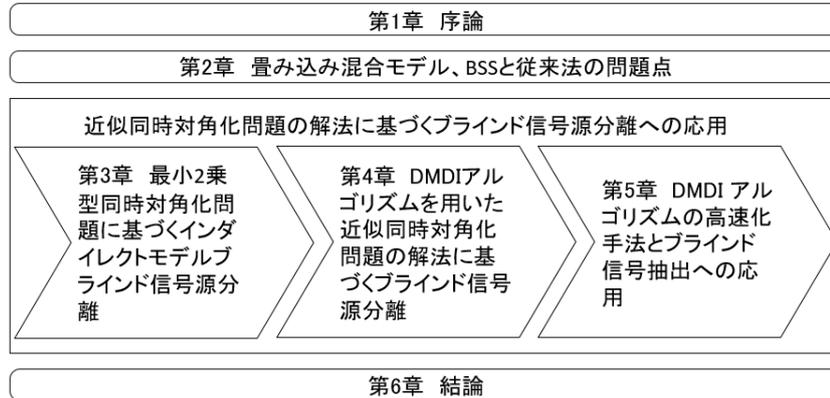


図 1.1: 学位論文の構成

が対角化される。ダイレクト最小 2 乗型同時対角化評価量 (direct least-squares (DLS) criterion) [29] は対角化行列 \mathbf{B} と対角行列 $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ の関数である。一方、インダイレクト最小 2 乗型同時対角化評価量 (indirect LS (ILS) criterion) [30] は分離行列 \mathbf{W} と対角行列 $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ の関数である。

ALS (alternating least-squares) アルゴリズム [5, 30–33] は DLS 評価量、または、ILS 評価量を用いて同時対角化問題を解方法する反復法であり、2 つの手順から成る。手順 1 では、対角行列 $\{\mathbf{P}_s(m)\}_{m=1}^M$ を固定して、DLS 評価量、または、ILS 評価量を最小する対角化行列 \mathbf{B} 、または、分離行列 \mathbf{W} をそれぞれ推定する。手順 2 では、対角化行列 \mathbf{B} 、または、分離行列 \mathbf{W} を固定して、DLS 評価量、または、ILS 評価量を最小する対角行列 $\{\mathbf{P}_s(m)\}_{m=1}^M$ を推定する。これらの 2 つの手順を収束するまで交互に繰り返す。

文献 [13, 34] で提案された ALSP (alternating least-squares with projection) アルゴリズムでは、最小 2 乗法を用いて DLS 評価量を最小する対角化行列 \mathbf{B} を求め、制約条件を写像によって課している。ALSP アルゴリズムの特徴は Khatri-Rao (KR) 積 [17] を利用すると、事実上、対角化行列 \mathbf{B} に関して DLS 評価量を 2 次形式で表現でき、既存の最小化手法が利用できることであるが、収束性が悪いことが指摘されている。本論文では、ALSP アルゴリズムの利点を生かし、欠点を克服する方法を提案し、ブラインド信号源分離とブラインド信号抽出に応用する。

1.2 本論文の概要

学位論文の構成を図 1.1 に記載する。本論文の概要は以下の通りである。第 2 章では、畳み込み混合モデルと同時対角化問題によるブラインド信号源分離、スケーリング、並びにパーミュテーション問題の解法、分離性能の評価指標、Rahbar 氏の従来の ALSP アルゴリズムの概要と問題点について議論する。

KR 積を用いると、DLS 評価量を対角化行列 \mathbf{B} に関して 2 次形式で表現でき、ALSP アルゴリズムにより効率良く評価量を最小にすることができる。他方、ブラインド信号源分離やブライ

ド信号抽出では、分離行列 \mathbf{W} が不可欠であるので、本論文では、ILS 評価量に KR 積を適用して分離行列を直接求める方法について 3 章で述べる。但し、ILS 評価量に直接 ALSP アルゴリズムを適用することは困難であるので、逆行列の補題を用いて観測信号のクロススペクトル密度行列の逆行列を展開式で表し、これに ALSP アルゴリズムを適用する方法を提案し、性能を評価する。逆行列の算出に必要な演算量はブロック処理を導入して低減させる。本論文では、第 3 章で提案するアルゴリズムを最小 2 乗型同時対角化問題に基づくインダイレクトモデルブラインド信号源分離と表記する。

ALSP アルゴリズムは収束性が悪いことが指摘されている。対角化行列 \mathbf{B} に関して 2 次形式で評価量を表現できる特徴を有する反面、実際には、評価量は \mathbf{B} の 4 次関数となっているので、対角化行列 \mathbf{B} の 2 次関数を \mathbf{B} に変換する必要がある。文献 [13, 34] では、累乘法を使用しているが、ALSP 反復の手順 1 に何度も累乘法を適用しなければならず、これが高速な収束と安定な動作の障害となっている。そこで、第 4 章では、対角化行列 \mathbf{B} に関して 2 次形式の評価量に直接制約を課し、ラグランジェの未定乗数法を導入することによって ALSP アルゴリズムの問題を克服する。また、同時対角化問題では対角化行列が正則であることが前提とされているが、最小 2 乗型同時対角化問題の解法の結果、対角化行列が正則となることは保証されていない。そこで、推定された対角化行列が正則でない場合は、これを正則になるように補正する操作を ALS アルゴリズムに加えて、正則な対角化行列を保証する。この操作は、混合過程から分離過程に至るモデルにおいて、分離信号のおける歪みの低減、発振の抑制にも効果があることを示す。本論文では、第 4 章で提案するアルゴリズムを DMDI (Diagonalization implemented by Minimizing the Direct and Indirect least-squares criteria) アルゴリズムと表記する。

4 章の DMDI アルゴリズムは ALSP アルゴリズムを基に作られている為、全体の ALS 反復の中に混合行列を推定する為のループが入れ子構造で存在している。この入れ子構造の演算量はマイクロフォンの個数の 2 乗で増加する為、信号源信号の個数よりマイクロホンの個数が多い混合過程では、演算量が大幅に増加することが明らかになっている。5 章では、入れ子構造を用いることなく非反復法で混合行列を推定できるアルゴリズムを提案する。これによって、演算量の低減によるアルゴリズムの高速化が期待できる。また、信号源信号の個数よりマイクロホンの個数を増やして、実環境下で分離性能の改善の有無を検証する。第 5 章で提案する高速化アルゴリズムを FDMDI (fast diagonalization, implemented by minimizing the direct and indirect LS criteria) アルゴリズムと表記する。信号源信号の個数よりマイクロホンの個数が多い場合と FDMDI アルゴリズムの組み合わせによって、信号源信号の個数とマイクロホンの個数が等しい場合と FDMDI アルゴリズムの組み合わせよりも計算時間が短縮できることをシミュレーションによって明らかにする。

第2章 畳み込み混合モデルとブラインド信号源分離

2.1 まえがき

本章では、記号の定義について述べた後、信号源からマイクロホンまでを畳み込み混合モデルで表現する。次いで、観測信号から信号源信号を復元するブラインド信号分離について議論した後、分離フィルタの推定法、周波数領域におけるブラインド信号分離では不可欠なパーミュテーション問題とその解法、スケールリング問題とその解法について紹介する。最後に、ブラインド信号分離法の性能評価指標、ブラインド信号抽出法によって復元した信号源信号の評価指標について述べる。

2.2 周波数領域における畳み込みブラインド信号分離

2.2.1 畳み込み混合モデル

図 2.1 のように N 個の信号源 $s_1(t), s_2(t), \dots, s_N(t)$ から J 個のマイクロフォン $x_1(t), x_2(t), \dots, x_J(t)$ までの時間領域畳み込み混合において、時刻 t における観測信号は

$$x_i(t) = \sum_{j=1}^N h_{ij}(t) * s_j(t) + n_i(t) \quad (2.1)$$

によって表される。信号源 $s_i(t)$ は平均零、疑似定常過程である。疑似定常過程において近似的に定常過程で扱える短時間区間をエポックと呼び、エポック間では疑似定常過程は非定常となる。信号源 $s_i(t)$ は互いに独立で $J \geq N \geq 2$ とする。 j 番目の信号源から i 番目のマイクロフォンまでの室内空間インパルス応答 $h_{ij}(t)$ は安定で因果性を満たすと仮定する。また、観測信号の測定期間中、室内空間インパルス応答は時不変とする。 $n_i(t)$ は平均零、分散 σ^2 のガウス性白色雑音をそれぞれ表す。ガウス性白色雑音と信号源は互いに独立な過程とする。離散フーリエ変換 (discrete Fourier transform, DFT) のポイント数 K が室内空間インパルス応答長よりも大きい場合、時間領域畳み込み演算は短時間フーリエ変換 (short-time Fourier transform, STFT) を用いると

$$\mathbf{x}(m, \omega_k) \approx \mathbf{H}(\omega_k) \mathbf{s}(m, \omega_k) + \mathbf{n}(m, \omega_k) \quad (2.2)$$

と表現することができる。ここで、 $k = 0, 1, \dots, K-1$ であるので角周波数は $\omega_k = 2\pi k/K$ となる。 $s_j(m, \omega_k)$ と $n_i(m, \omega_k)$ はそれぞれエポック m における $s_j(t)$ と $n_i(t)$ の STFT である。また、 $h_{ij}(\omega_k)$ は $h_{ij}(t)$ の DFT である。エポック m における信号ベクトルを $\mathbf{s}(m, \omega_k) =$

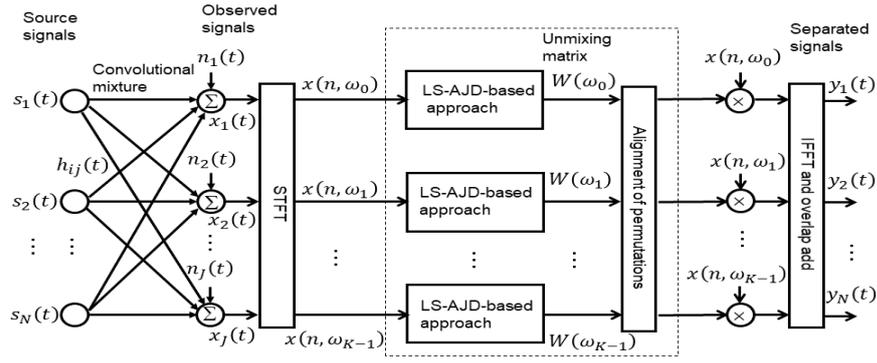


図 2.1: 畳み込み混合モデルとブラインド信号源分離

$[s_1(m, \omega_k), s_2(m, \omega_k), \dots, s_N(m, \omega_k)]^T$ 、 N 個の信号源から J 個のマイクロフォンまでの伝達関数を混合行列 $\mathbf{H}(\omega_k)$ でそれぞれ表記する。混合行列 $\mathbf{H}(\omega_k)$ は全角周波数においてフル列ランクと仮定する。信号源とマイクロフォンの個数が等しい混合、即ち、 $J = N \geq 2$ のとき、混合行列は determined、マイクロフォン個数が信号源の個数より多い混合、即ち、 $J > N \geq 2$ のとき、混合行列を overdetermined という。観測信号ベクトルを $\mathbf{x}(m, \omega_k) = [x_1(m, \omega_k), x_2(m, \omega_k), \dots, x_J(m, \omega_k)]^T$ 、雑音ベクトルを $\mathbf{n}(m, \omega_k) = [n_1(m, \omega_k), n_2(m, \omega_k), \dots, n_J(m, \omega_k)]^T$ と表記し、測定可能な全エポック数を M とすると、観測信号はエポック区間 $1 \leq m \leq M$ で測定可能となる。

2.2.2 ブラインド信号分離

周波数領域における畳み込みブラインド信号分離では

$$\hat{\mathbf{s}}(m, \omega_k) = \mathbf{W}(\omega_k)\mathbf{x}(m, \omega_k) \quad (2.3)$$

のように観測信号 $\mathbf{x}(m, \omega_k)$ に分離行列 $\mathbf{W}(\omega_k)$ を乗算することによって信号源信号 $\mathbf{s}(m, \omega_k)$ を推定する。ここで

$$\mathbf{W}(\omega_k)\mathbf{H}(\omega_k) = \mathbf{\Pi}(\omega_k)\mathbf{P}(\omega_k) \quad (2.4)$$

が成り立ち、パーミュテーション行列 $\mathbf{P}(\omega_k) \in \mathbb{R}^{N \times N}$ は周波数ビン毎に独立して割り当てられた音源の割り当てを揃えるために、スケーリング行列 $\mathbf{\Pi}(\omega_k) \in \mathbb{C}^{N \times N}$ は周波数ビン毎に異なる音源の電力、及び位相を揃えるためにそれぞれ用いられる。ブラインド信号抽出の目的は、混合行列 $\mathbf{H}(\omega_k)$ と信号源信号 $\mathbf{s}(m, \omega_k)$ が未知な状態から分離行列 $\mathbf{W}(\omega_k)$ 、パーミュテーション行列 $\mathbf{P}(\omega_k)$ 、及び、スケーリング行列 $\mathbf{\Pi}(\omega_k)$ を推定することである。

観測信号 $\mathbf{x}(m, \omega_k)$ のクロススペクトル密度行列を

$$\hat{\mathbf{T}}(m, \omega_k) = E[\mathbf{x}(m, \omega_k)\mathbf{x}(m, \omega_k)^H] = \mathbf{H}(\omega_k)\mathbf{P}_s(m, \omega_k)\mathbf{H}(\omega_k)^H + \sigma^2\mathbf{I} \quad (2.5)$$

と定義すると、同時対角化問題では $\hat{\mathbf{T}}(m, \omega_k)$ を対角化することによって $\mathbf{H}(\omega_k)$ または $\{\mathbf{P}_s(m, \omega_k)\}_{m=1}^M$ を求める。信号源信号 $\mathbf{s}(m, \omega_k)$ のクロススペクトル密度行列 $\mathbf{P}_s(m, \omega_k) = E[\mathbf{s}(m, \omega_k)\mathbf{s}(m, \omega_k)^H] \in$

$\mathbb{R}^{N \times N}$ は対角行列である。また、 $\mathbf{T}(m, \omega_k) = \mathbf{H}(\omega_k) \mathbf{P}_s(m, \omega_k) \mathbf{H}(\omega_k)^H$ と定義している。行列 $\mathbf{B}(\omega_k)$ と対角行列 $\mathbf{\Lambda}(m, \omega_k)$ をそれぞれ

$$\mathbf{B}(\omega_k) = \mathbf{H}(\omega_k) \mathbf{P}(\omega_k)^{-1} \mathbf{\Pi}(\omega_k)^{-1} \quad (2.6)$$

$$\mathbf{\Lambda}(m, \omega_k) = \mathbf{\Pi}(\omega_k) \mathbf{P}(\omega_k) \mathbf{P}_s(m, \omega_k) \mathbf{P}(\omega_k)^T \mathbf{\Pi}(\omega_k)^H \quad (2.7)$$

と定義すると、式 (2.5) は

$$\hat{\mathbf{T}}(m, \omega_k) = \mathbf{B}(\omega_k) \mathbf{\Lambda}(m, \omega_k) \mathbf{B}(\omega_k)^H + \sigma^2 \mathbf{I} \quad (2.8)$$

と表現することができる。ここで、行列 $\mathbf{B}(\omega_k)$ の i 番目の行ベクトル $\mathbf{b}_i(\omega_k)$ には制約条件 $\|\mathbf{b}_i(\omega_k)\|_2 = 1$ が課せられている。式 (2.4) と (2.6) より

$$\mathbf{W}(\omega_k) \mathbf{B}(\omega_k) = \mathbf{I} \quad (2.9)$$

の関係が得られる。 $\mathbf{B}(\omega_k)$ を推定した後、分離行列 $\mathbf{W}(\omega_k)$ は式 (2.9) を満足するように求められる。パーミュテーション問題とスケーリング問題を解法した後、逆離散フーリエ変換 (inverse DFT, IDFT) によって時間領域に $\mathbf{P}(\omega_k)^{-1} \mathbf{\Pi}(\omega_k)^{-1} \mathbf{W}(\omega_k)$ を変換すると、分離フィルタのインパルス応答を算出することができる。

2.3 同時対角化問題

M 個のエルミートな対象行列 $\{\mathbf{T}(m, \omega_k)\}_{m=1}^M$ が

$$\mathbf{T}(m, \omega_k) = \mathbf{B}(\omega_k) \mathbf{\Lambda}(m, \omega_k) \mathbf{B}(\omega_k)^H \quad (2.10)$$

の形式で与えられたとき、 $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ が未知な状態で $\{\mathbf{T}(m, \omega_k)\}_{m=1}^M$ から $\mathbf{B}(\omega_k)$ を推定する問題が同時対角化問題 (joint diagonalization, JD) である。通常、加法性雑音、推定誤差などのために、対象行列は推定値 $\{\hat{\mathbf{T}}(m, \omega_k)\}_{m=1}^M$ が入手可能で、しかも、入手個数には限りがある。近似同時対角化問題 (approximate JD, AJD) はダイレクト最小 2 乗評価量 (DLS 評価量) [29, 30]

$$C_{\text{DLS}} \left(\mathbf{B}(\omega_k), \{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M \right) = \sum_{m=1}^M \left\| \hat{\mathbf{T}}(m, \omega_k) - \mathbf{B}(\omega_k) \mathbf{\Lambda}(m, \omega_k) \mathbf{B}(\omega_k)^H \right\|_F^2 \quad (2.11)$$

を最小にすることによって $\mathbf{B}(\omega_k)$ と M 個の対角行列 $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ を推定する問題である。制約条件 $\|\mathbf{b}_i(\omega_k)\|_2 = 1$ が課される場合には、制約条件付きダイレクト型評価量 (CDLS 評価量)

$$\begin{aligned} C_{\text{CDLS}} \left(\mathbf{B}(\omega_k), \{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M \right) &= \sum_{m=1}^M \left\| \hat{\mathbf{T}}(m, \omega_k) - \mathbf{B}(\omega_k) \mathbf{\Lambda}(m, \omega_k) \mathbf{B}(\omega_k)^H \right\|_F^2 \\ \text{s.t. } \|\mathbf{b}_i(\omega_k)\|_2 &= 1 \end{aligned} \quad (2.12)$$

の最小化が用いられる。ブラインド信号分離、並びにブラインド信号抽出の目的は行列 $\left\{ \mathbf{W}(\omega_k) \hat{\mathbf{T}}(m, \omega_k) \mathbf{W}(\omega_k)^H \right\}_{m=1}^M$ ができる限り対角行列になるように分離行列 $\mathbf{W}(\omega_k)$ を推定する

ことであるが、 $\mathbf{W}(\omega_k)$ は DLS 評価量、または CDLS 評価量を最小化しても直接推定することができない。 $\mathbf{W}(\omega_k)$ を直接推定する方法の 1 つは

$$C_{\text{ILS}}(\mathbf{W}(\omega_k)) = \sum_{m=1}^M \left\| \text{Zdiag} \left(\mathbf{W}(\omega_k) \hat{\mathbf{T}}(m, \omega_k) \mathbf{W}(\omega_k)^H \right) \right\|_F^2 \quad (2.13)$$

を最小にすることである。解が $\mathbf{W}(\omega_k) = \mathbf{0}$ となることを避けるために、ILS 評価量に制約条件が課せられる。また、インダイレクト最小 2 乗評価量 (ILS 評価量)

$$C_{\text{ILS}} \left(\mathbf{W}(\omega_k), \{ \mathbf{\Lambda}(m, \omega_k) \}_{m=1}^M \right) = \sum_{m=1}^M \left\| \mathbf{W}(\omega_k) \hat{\mathbf{T}}(m, \omega_k) \mathbf{W}(\omega_k)^H - \mathbf{\Lambda}(m, \omega_k) \right\|_F^2 \quad (2.14)$$

を最小化することによって $\mathbf{W}(\omega_k)$ と対角行列 $\{ \mathbf{\Lambda}(m, \omega_k) \}_{m=1}^M$ を直接推定する方法も提案されている [30]。

2.4 分離行列の推定

2.4.1 勾配法に基づく分離行列の推定

勾配法に基づく FDSOS (frequency-domain second-order-statistics) アルゴリズム [35, 36] は、式 (2.13) を逐次的に最小にする。反復 $\ell + 1$ における分離行列 $\mathbf{W}^{(\ell+1)}(\omega_k)$ は反復 ℓ における分離行列 $\mathbf{W}^{(\ell)}(\omega_k)$ に負の勾配 $-\partial C_{\text{ILS}}(\mathbf{W}(\omega_k)) / \partial \mathbf{W}(\omega_k)^{(\ell)*}$ に比例した変化を加算することによって、FDSOS アルゴリズム

$$\mathbf{W}^{(\ell+1)}(\omega_k) = \mathbf{W}^{(\ell)}(\omega_k) - \mu(\omega_k) \frac{\partial C_{\text{ILS}}(\mathbf{W}(\omega_k))}{\partial \mathbf{W}(\omega_k)^{(\ell)*}} \quad (2.15)$$

が導出される。ここで、 $\mu(\omega_k)$ は

$$\mu(\omega_k) = \frac{\alpha}{\sum_{m=1}^M \left\| \hat{\mathbf{T}}(m, \omega_k) \right\|_F^2} \quad (2.16)$$

で与えられる。文献 [36] では正規化ステップサイズは $\alpha = 0.5$ に設定されている。

2.4.2 同時対角化問題解法に基づく分離行列の推定

最小 2 乗型同時対角化問題問題の解法は、式 (2.11) の評価量を次の 2 つの手順

手順 1) CDLS 評価量を最小にする混合行列 $\mathbf{B}(\omega_k)$ を推定する。このとき、対角行列 $\{ \mathbf{\Lambda}(m, \omega_k) \}_{m=1}^M$ は手順 2 で推定した値に固定する

手順 2) CDLS 評価量を最小にする対角行列 $\{ \mathbf{\Lambda}(m, \omega_k) \}_{m=1}^M$ を推定する。このとき、混合行列 $\mathbf{B}(\omega_k)$ は手順 1 で推定した値に固定する

を交互に繰り返し、反復することによって混合行列と対角行列を逐次的に推定する。反復終了後

$$\mathbf{W}(\omega_k) = (\mathbf{B}(\omega_k)^H \mathbf{B}(\omega_k))^{-1} \mathbf{B}(\omega_k)^H \quad (2.17)$$

によって分離行列を算出する。

2.5 パーミュテーション問題の解法

周波数ビン毎に分離フィルタの出力に信号源信号を割り当てると、周波数ビン毎に割り当てが異なり、全周波数帯域で割り当てを揃える必要がある。これをパーミュテーション問題という。代表的なパーミュテーション問題の解法として、信号の到来方向を用いる方法と同一信号源から同一時刻に発せられた信号源信号の相関を利用した方法が存在する。

2.5.1 信号の到来方向を利用した信号源の割り当て

信号の到来方向を用いた方法では、周波数ビン毎に信号源からの信号の到来方向を推定して、全周波数帯域で分離フィルタの出力への信号源を割り当てを揃える方法である [37]。到来方向推定の精度がパーミュテーション問題の解法精度に影響を与え、特に残響が存在する環境下では到来方向推定に誤りが生じ易いことが指摘されている。

2.5.2 信号源信号の相関を利用した信号源の割り当て

同一信号源から同一時刻に発せられた信号は、近傍周波数ビン間で信号包絡に高い相関がある。2つの信号源を仮定すると、周波数ビン ω_k と ω_ℓ の分離信号 $v_q(m, \omega_k)$ と $v_p(m, \omega_k)$ の相関係数は

$$\rho_{qp}(\omega_k, \omega_\ell) = \frac{\sum_{m=1}^M v_q(m, \omega_k) v_p(m, \omega_\ell)}{\sqrt{\sum_{m=1}^M v_q^2(m, \omega_k)} \sqrt{\sum_{m=1}^M v_p^2(m, \omega_\ell)}} \quad (2.18)$$

で与えられる。ここで、 $v_q(m, \omega_k)$ と $v_p(m, \omega_\ell)$ は $\mathbf{V}(m, \omega_k)$ の q 番目の要素、 $\mathbf{V}(m, \omega_\ell)$ の p 番目の要素 $v_p(m, \omega_\ell)$ である。また、 M はフレームの総数である。

$$r = \frac{\rho_{qq}(\omega_k, \omega_\ell) + \rho_{pp}(\omega_k, \omega_\ell)}{\rho_{pq}(\omega_k, \omega_\ell) + \rho_{qp}(\omega_k, \omega_\ell)} \quad (2.19)$$

を用いると、 r が大きい値のとき ω_k と ω_ℓ の間の信号源割り当ては合致している。一方、 r が小さい値のとき ω_k の信号源の割り当てを変える必要がある。 N 個の信号源を仮定すると、パーミュテーション行列は

$$\hat{\mathbf{P}}(\omega_k) = \underset{\mathbf{P}(\omega_k)}{\operatorname{argmax}} \operatorname{tr} (\mathbf{P}(\omega_k) \mathbf{E}(\omega_k) \mathbf{E}(\omega_\ell)^T) \quad (2.20)$$

によって推定される [13]。ただし

$$\mathbf{E}(\omega_k) = \left(\sum_{m=1}^M \mathbf{V}(m, \omega_k)^2 \right)^{-1/2} \Theta \quad (2.21)$$

$$\mathbf{V}(m, \omega_k) = \text{diagonal}(v_1(m, \omega_k), v_2(m, \omega_k), \dots, v_N(m, \omega_k)) \quad (2.22)$$

$$\Theta = \begin{pmatrix} v_1(1, \omega_k), v_1(2, \omega_k), \dots, v_1(M, \omega_k) \\ \vdots & \vdots & \ddots & \vdots \\ v_N(1, \omega_k), v_N(2, \omega_k), \dots, v_N(M, \omega_k) \end{pmatrix} \quad (2.23)$$

である。次に、相関係数の算出対象について述べる [13, 38, 39]。

- a) 分離フィルタの出力信号 $y_q(m, \omega_k)$ と $y_p(m, \omega_\ell)$ の包絡 $v_q(m, \omega_k) = |y_q(m, \omega_k)|$ と $v_p(m, \omega_\ell) = |y_p(m, \omega_\ell)|$ の相関係数
- b) 分離フィルタの出力信号の電力 $v_q(m, \omega_k) = |y_q(m, \omega_k)|^2$ と $v_p(m, \omega_\ell) = |y_p(m, \omega_\ell)|^2$ の相関係数
- c) 電力の比の相関係数、即ち、分離フィルタの出力信号の全電力と q 番目の出力信号の電力の比 [38]

$$v_q(m, \omega_k) = \frac{|y_q(m, \omega_k)|^2}{\sum_{k=1}^J |y_k(m, \omega_k)|^2} \quad (2.24)$$

の最大値をから時間周波数領域 (m, ω_k) において支配的な信号源信号を求める。信号源が音声である場合、音声信号のスパース性を利用すると、電力比が 1 に近い値であれば q 番目の出力信号が支配的で十分に分離されていることを表し、零に近い値であれば他の出力信号が支配的となる。そこで、隣接周波数ビン間の電力比の相関係数を利用する。

2.6 スケーリング問題の解法

周波数ビン毎に分離信号の振幅は異なる。この影響を無くすために minimal distortion principle [40]

$$\mathbf{W}(\omega_k) = \text{diag}[\mathbf{W}(\omega_k)^{-1}] \mathbf{W}(\omega_k) \quad (2.25)$$

が用いられる。また、文献 [14] では、分離行列の各列ベクトルをその大ききで正規化することによってスケーリング問題を解いている。

2.7 分離性能の指標と非パーミュテーションブラインド手法

分離性能の指標には、入力所望信号対干渉信号比 (input signal-to-interference ratio, input SIR) と出力所望信号対干渉信号比 (output SIR)

$$\text{SIR}_{x_i} = 10 \log_{10} \frac{\sum_t [(h_{ii}(t) * s_i(t))^2]}{\sum_t \left[\left(\sum_{\substack{j=1 \\ j \neq i}}^N h_{ij}(t) * s_j(t) + n_i(t) \right)^2 \right]} \quad (2.26)$$

$$\text{SIR}_{y_i} = 10 \log_{10} \frac{\sum_t [(\gamma_{ii}(t) * s_i(t))^2]}{\sum_t \left[\left(\sum_{\substack{j=1 \\ j \neq i}}^N \gamma_{ij}(t) * s_j(t) + \sum_{j=1}^J w_{ij}(t) * n_j(t) \right)^2 \right]} \quad (2.27)$$

を用いる [14]。ここで、ここで、 $w_{ij}(t)$ と $\gamma_{ij}(t)$ は分離行列 $\mathbf{W}(\omega_k)$ と

$$\mathbf{\Gamma}(\omega_k) = e^{-j\pi k} \mathbf{P}(\omega_k)^{-1} \mathbf{\Pi}(\omega_k)^{-1} \mathbf{W}(\omega_k) \mathbf{H}(\omega_k) \quad (2.28)$$

を IDFT によって時間領域に変換したものである。また、 $\gamma_{ij}(t)$ は i 番目の信号源から合成フィルタの j 番目の出力までのインパルス応答を表す。信号対雑音比 (signal-to-noise ratio, SNR) は、最適な分離フィルタ $\mathbf{W}(\omega_k)_{\text{opt}}$ と最適なパーミュテーション行列 $\mathbf{P}(\omega_k)_{\text{opt}}$ を用いて分離フィルタの出力信号における所望信号成分と雑音成分の電力比によって決定した [14]。即ち、室内空間インパルス応答 $h_{ij}(t)$ が既知のとき、最適な分離フィルタ $\mathbf{W}(\omega_k)_{\text{opt}}$ は $\mathbf{H}(\omega_k)^{-1}$ の各列ベクトルを正規化することによって求められる。正規化は数学的にスケーリング問題の解法と等価である。同様に、 $\mathbf{H}(\omega_k)$ と $\mathbf{W}(\omega_k)_{\text{opt}}$ が既知のとき、最適なパーミュテーション行列 \mathbf{P}_{opt} は

$$\mathbf{P}(\omega_k)_{\text{opt}} = \underset{\mathbf{P}(\omega_k)}{\text{argmax}} \|\text{diag} [\mathbf{P}(\omega_k) \mathbf{C}(\omega_k)_{\text{opt}}]\|_F^2 = \mathbf{I} \quad (2.29)$$

によって求めることができる。ここで、 $\mathbf{C}(\omega_k)_{\text{opt}} = \mathbf{W}(\omega_k)_{\text{opt}} \mathbf{H}(\omega_k)$ である。このように、SN 比は最適な出力 SIR に等しい。ブラインドで抽出された音質は PESQ スコア [41] によって評価する。PESQ は通信路における音質の評価方法の 1 つである。

分離行列の推定精度だけを評価するために非パーミュテーションブラインド手法を使用する。これは、パーミュテーション行列の算出に限り $\mathbf{H}(\omega_k)$ が既知と仮定しており、このとき、パーミュテーション行列は

$$\hat{\mathbf{P}}(\omega_k) = \underset{\mathbf{P}(\omega_k)}{\text{argmax}} \left\| \text{diag} \left[\mathbf{P}(\omega_k) \hat{\mathbf{C}}(\omega_k) \right] \right\|_F^2 \quad (2.30)$$

によって求められる [14]。ここで、 $\hat{\mathbf{W}}(\omega_k)$ は推定値で、 $\hat{\mathbf{C}}(\omega_k) = \hat{\mathbf{W}}(\omega_k) \mathbf{H}(\omega_k)$ である。

2.8 Rahbar 氏の従来の ALSP アルゴリズムの概要と問題点

Rahbar 氏の従来の ALSP アルゴリズム [13] は行列 $\hat{\mathbf{T}}(m, \omega_k)$ をベクトル変換すると行列 $\mathbf{B}(\omega_k)^* \odot \mathbf{B}(\omega_k)$ と $\mathbf{d}(m, \omega_k)$ の掛け算で表せることに着目している。これにより $\mathbf{B}(\omega_k)^* \odot \mathbf{B}(\omega_k)$ に関して 2 乗になるので最小 2 乗法が使えるというのが従来の ALSP アルゴリズムの概要である。集合 Ω は付録 B の式 (B.14) で定義される複素領域の集合 $J^2 \times 1$ に含まれる、 $J^2 \times 1$ の部分集合である。この制約条件を $\mathbf{B}(\omega_k)^* \odot \mathbf{B}(\omega_k)$ の列ベクトルに課すことによって $\mathbf{b}_i(\omega_k)$ への写像を可能にしている。

従来の ALSP アルゴリズムは CDLS 評価量を基にしている為、直接分離行列 $\mathbf{W}(\omega_k)$ を推定出来ない。またスケールリング問題も別途解く必要がある。更に $\hat{\mathbf{T}}(m, \omega_k)$ が 3×3 以上だと精度良く分離行列を推定出来ない。最後にアルゴリズムが収束するまでに多くの反復回数を要することによる計算時間の増加等が問題点として挙げられる。

2.9 まとめ

本章では、信号源からマイクロフォンまでの畳み込み混合モデル、同時対角化問題、分離フィルタの推定方法、パーミュテーション問題とその解法、スケールリング問題とその解法、分離性能の評価指標と非パーミュテーションブラインド信号分離、Rahbar 氏の従来の ALSP アルゴリズムの概要と問題点について述べた。

第3章 最小2乗型同時対角化問題に基づくインダイレクトモデルブラインド信号源分離

3.1 まえがき

従来の ALSP アルゴリズムは CDLS 評価量を基にしている為、直接分離行列 $\mathbf{W}(\omega_k)$ を推定出来ない。またスケールリング問題も別途解く必要がある。更に $\hat{\mathbf{T}}(m, \omega_k)$ が 3×3 以上だと精度良く分離行列を推定出来ない。これら問題点を克服する手法を3章で提案する。

インダイレクトモデルブラインド信号分離は、ILS 評価量を最小にすることによって分離行列を直接推定する方法である。本章では、畳み込み混合過程にインダイレクトモデルブラインド信号分離を導入し、同時対角化問題の解法に基づき、観測信号のクロススペクトル密度行列の逆行列から対角化行列と対角行列を推定することによって分離行列を推定する。逆行列を展開式で表現すると、ALSP (alternating least-squares with projection) アルゴリズム [13] の適用が可能となり、制約条件付き同時対角化問題を ALSP アルゴリズムによって解法すると、スケールリング問題を解くことなく分離行列を推定することができる。更に、この解法にブロック処理 [42, 43] を導入して、少ない演算量で高い分離性能を有するインダイレクトモデルブラインド信号分離法を提案する。提案法は

ステップ 1) ALSP アルゴリズムによって分離行列を直接推定する

ステップ 2) パーミュテーション問題を解法する

によって実行できることが特徴である。シミュレーションでは、従来法と比較して提案法が高い分離性能を達成できることを示す。

3.2 逆行列の展開式を用いたインダイレクトモデルブラインド信号分離

ILS 評価量を最小にする従来のインダイレクトモデルブラインド信号分離は文献 [30, 35, 36] で提案されている。観測信号のクロススペクトル密度行列に左から分離行列、右から分離行列のエルミートを乗算すると、観測誤差を $\mathbf{W}(\omega_k)\hat{\mathbf{T}}(m, \omega_k)\mathbf{W}(\omega_k)^H$ から $\mathbf{\Lambda}(m, \omega_k)$ を減算することによって求めることができる。文献 [35, 36] では、分離行列を求めるために最急降下法によって観測誤差

の総和を最小にしているが、同時に、エポック m における音源のクロススペクトル密度行列を推定する必要がある。文献 [30] では、瞬時混合のための最小 2 乗型同時対角化問題を解法するために観測誤差の総和を最小化している。ALSP アルゴリズム [13] は、 $\mathbf{W}(\omega_k)\hat{\mathbf{T}}(m, \omega_k)\mathbf{W}(\omega_k)^H$ において行列 $\hat{\mathbf{T}}(m, \omega_k)$ が対角行列でないためにインダイレクトモデルに基づく最小 2 乗問題を解法することができない。一方、ダイレクトモデルに基づく最小 2 乗問題では、 $\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H$ において行列 $\mathbf{\Lambda}(m, \omega_k)$ は対角行列であるので、ALSP アルゴリズムは適用可能である。ここでは、ブロック毎に指数関数的に重み付けされた観測信号のクロススペクトル密度行列の逆行列を用いてブロック観測誤差の総和を最小化することによって分離行列を直接推定するインダイレクトモデルブラインド信号分離アルゴリズムを提案する。

従来のインダイレクトモデルブラインド信号分離 [35, 36] は周波数領域循環畳み込み演算によって時間領域畳み込み演算を実現するために、FFT ポイントと分離フィルタの次数に依存する部分空間に勾配ベクトルを写像することによってパーミュテーション問題を解法する必要があった。一方、提案法では、FFT ポイントと分離フィルタの次数に制約は無く、隣接周波数ビン間における分離信号の電力比の相関係数を利用してパーミュテーション問題を解法する。

マイクロフォンの個数が音源の個数より多い場合、雑音を加法されていない純粋な観測信号のクロススペクトル密度行列は

$$\hat{\mathbf{T}}(m, \omega_k) - \sigma^2 \mathbf{I} = \mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H \quad (3.1)$$

で与えられる。ただし、 σ^2 は $\hat{\mathbf{T}}(m, \omega_k)$ の最小固有値、即ち、加法性雑音の分散で、一般にマイクロフォンと音源の個数が等しい混合過程では求めることができない。ここでは、マイクロフォンと音源の個数が等しい混合過程におけるブラインド信号分離の性能を評価する。 $\mathbf{B}(\omega_k)$ と $\mathbf{\Lambda}(m, \omega_k)$ が正則行列であると仮定すると、時間周波数領域 (m, ω_k) における観測信号のクロススペクトル密度行列の逆行列は

$$\begin{aligned} \hat{\mathbf{T}}(m, \omega_k)^{-1} &= (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H + \sigma^2 \mathbf{I})^{-1} \\ &= (\mathbf{B}(\omega_k)^H)^{-1} \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{B}(\omega_k)^{-1} \\ &\quad - \sigma^2 \sum_{\ell=0}^{\infty} (-1)^\ell \sigma^{2\ell} \left[(\mathbf{B}(\omega_k)^H)^{-1} \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{B}(\omega_k)^{-1} \right]^{\ell+2} \end{aligned} \quad (3.2)$$

のように展開できる。式 (3.2) の証明は付録 A に記す。対角化行列 $\mathbf{W}(\omega_k) \in \mathbb{C}^{N \times N}$ と対角行列 $\mathbf{\Lambda}(m, \omega_k)^{-1} \in \mathbb{C}^{N \times N}$ は制約条件 $\|\mathbf{w}_i(\omega_k)\|_2 = 1$ で

$$\begin{aligned} \hat{\mathbf{T}}(m, \omega_k)^{-1} + \sigma^2 \sum_{\ell=0}^{\infty} (-1)^\ell \sigma^{2\ell} \left[(\mathbf{B}(\omega_k)^H)^{-1} \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{B}(\omega_k)^{-1} \right]^{\ell+2} \\ = \mathbf{W}(\omega_k)^H \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{W}(\omega_k) \end{aligned} \quad (3.3)$$

となる。但し、 $\mathbf{w}_i(\omega_k)$ は $\mathbf{W}(\omega_k)$ の i 番目の列ベクトルである。

時間周波数領域 (m, ω_k) における観測誤差を

$$\mathcal{E}(m, \omega_k) = \hat{\mathbf{T}}(m, \omega_k)^{-1} - \mathbf{W}(\omega_k)^H \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{W}(\omega_k) \quad (3.4)$$

と定義すると、対角化行列 $\mathbf{W}(\omega_k)$ と対角行列 $\{\mathbf{\Lambda}(m, \omega_k)^{-1}\}_{m=1}^M$ は、最小 2 乗型同時対角化問題の解法に基づき、周波数ビン毎に制約条件 $\|\mathbf{w}_i(\omega_k)\|_2 = 1$ 付きで 2 乗観測誤差の総和

$$\hat{\mathbf{W}}(\omega_k), \hat{\mathbf{\Lambda}}(m, \omega_k)^{-1} = \underset{\substack{\mathbf{W}(\omega_k), \mathbf{\Lambda}(m, \omega_k)^{-1} \\ \|\mathbf{w}_i(\omega_k)\|_2 = 1}}{\operatorname{argmin}} \sum_{m=1}^M \|\mathcal{E}(m, \omega_k)\|_F^2 \quad (3.5)$$

を最小にすることによって推定することができる。ダイレクトモデルブラインド信号分離では混合行列を推定するのに対し、提案するインダイレクトモデルブラインド信号分離法では分離行列を推定する点が異なる。

音声や音楽信号のような非定常な信号源では、 $\hat{\mathbf{T}}(m, \omega_k)$ の短時間区間での推定に指数関数的に重み付けされた観測信号のクロススペクトル密度行列

$$\hat{\mathbf{T}}(m, \omega_k) = \sum_{\tau=1}^m \beta^{m-\tau} \mathbf{x}(\tau, \omega_k) \mathbf{x}(\tau, \omega_k)^H \quad (3.6)$$

が有効である。ここで、 β は忘却係数で、1 未満の正の定数である。 $\mathbf{T}(m, \omega_k)$ は

$$\hat{\mathbf{T}}(m, \omega_k) = \beta \hat{\mathbf{T}}(m-1, \omega_k) + \mathbf{x}(m, \omega_k) \mathbf{x}(m, \omega_k)^H \quad (3.7)$$

によって反復的に更新される。 $\mathbf{T}(0, \omega_k) = c\mathbf{I}$ で、 c は小さな正の定数である。 $\hat{\mathbf{T}}(n, \omega_k)^{-1}$ は $\hat{\mathbf{T}}(m, \omega_k)^{-1}$ を正規化

$$\tilde{\mathbf{T}}(m, \omega_k)^{-1} = \frac{\hat{\mathbf{T}}(m, \omega_k)^{-1}}{\|\hat{\mathbf{T}}(m, \omega_k)^{-1}\|_F} \quad (3.8)$$

することによって求められる。

3.3 最小 2 乗型同時対角化問題のブロック処理

隣接フレーム間のシフトサイズ T_s を STFT のポイント数 K よりも短くすると、限られた個数の観測信号から多くの同時対角化対象行列を生成することができる。一方、同時対角化対象行列の個数を増やすと、フレーム毎の $\mathbf{P}(m, \omega_k)^{-1}$ の算出、対角行列 $\{\mathbf{\Lambda}(m, \omega_k)^{-1}\}_{m=1}^M$ 、分離行列 $\mathbf{W}(\omega_k)$ の推定に膨大な演算が必要となる。演算量を削減するために、最小 2 乗型同時対角化問題にブロックフレーム毎に指数関数的に重み付けされた観測信号のクロススペクトル密度行列の逆行列

$$\tilde{\mathbf{T}}(\tau L, \omega_k)^{-1} = \frac{\hat{\mathbf{T}}(\tau L, \omega_k)^{-1}}{\|\hat{\mathbf{T}}(\tau L, \omega_k)^{-1}\|_F} \quad (3.9)$$

の計算にブロック処理 [42, 43] を導入する。ブロック長 L は 1、または 1 以上の正の整数である。式 (3.4) と同様に、ブロック処理を用いてブロックフレームにおける観測誤差を

$$\mathcal{E}(\tau L, \omega_k) = \tilde{\mathbf{T}}(\tau L, \omega_k)^{-1} - \mathbf{W}(\omega_k)^H \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \mathbf{W}(\omega_k) \quad (3.10)$$

と定義すると、逆行列 $\{\mathbf{T}(\tau L, \omega_k)^{-1}\}_{\tau=1}^{\lfloor M/L \rfloor}$ を用いた最小 2 乗型同時対角化問題のブロック処理は、制約条件 $\|\mathbf{w}_i(\omega_k)\|_2 = 1$ 付きでブロック毎の 2 乗観測誤差の総和

$$\hat{\mathbf{W}}(\omega_k), \hat{\mathbf{\Lambda}}(\tau L, \omega_k)^{-1} = \underset{\substack{\mathbf{W}(\omega_k), \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \\ \|\mathbf{w}_i(\omega_k)\|_2 = 1}}{\operatorname{argmin}} \sum_{\tau=1}^{\lfloor M/L \rfloor} \|\mathcal{E}(\tau L, \omega_k)\|_F^2 \quad (3.11)$$

を最小にすることで対角化行列 $\mathbf{W}(\omega_k)$ と対角行列 $\{\mathbf{\Lambda}(\tau L, \omega_k)^{-1}\}_{\tau=1}^{\lfloor M/L \rfloor}$ を推定することである。式 (3.11) を解くために ALSP アルゴリズムを使用する。ALSP アルゴリズムは 2 つの手順から成る。手順 1 では、 $\mathbf{W}(\omega_k)^H \odot \mathbf{W}(\omega_k)$ を推定するためにブロック毎に 2 乗観測誤差の総和を最小 2 乗法を用いて最小化にする。このとき、反復当たり $O(\lfloor M/L \rfloor N^3)$ の演算を要する。累乘法 [44] によって $\mathbf{W}(\omega_k)^H \odot \mathbf{W}(\omega_k)$ から $\mathbf{W}(\omega_k)$ を推定した後、手順 2 で $\{\mathbf{\Lambda}(\tau L, \omega_k)^{-1}\}_{\tau=1}^{\lfloor M/L \rfloor}$ を最小 2 乗法によって推定する。このとき、疑似逆行列の演算に反復当たり $O(\lfloor M/L \rfloor N^3)$ の演算を要する。近似同時対角化問題のブロック処理による解法については付録 B で述べる。ブロック処理を用いたインダイレクトモデルブラインド信号分離アルゴリズムを Algorithm 1 に要約する [32]。

Algorithm 1 ブロック処理を用いたインダイレクトモデルブラインド信号分離アルゴリズム

- Initialization:**
- 1: Set $\hat{\mathbf{T}}(0, \omega_k)$ to $c\mathbf{I}$.
- Estimation of the cross-spectral density matrices of the observed signal:**
- 2: Update $\hat{\mathbf{T}}(m, \omega_k)$ at each frame, as given by (3.7).
 - 3: **while** $\tau < \lfloor M/L \rfloor$ **do**
- Normalization of $\hat{\mathbf{T}}(\tau L, \omega_k)$:**
- 4: Compute the inverse of $\hat{\mathbf{T}}(\tau L, \omega_k)$.
 - 5: Normalize $\hat{\mathbf{T}}(\tau L, \omega_k)^{-1}$ at each block frame using (3.9).
- Estimation of the unmixing matrix:**
- 6: Estimate $\mathbf{W}(\omega_k)$ by the ALSP algorithm at each block frame as shown in (14), (18), (20), and (22) in [13].
- 7: **end while**
- Realization of the delayed unmixing filter:**
- 8: Multiply $\mathbf{W}(\omega_k)$ by $e^{-j\pi k}$.
- Solution of the permutation problems:**
- 9: Resolve the permutation ambiguity to find $\mathbf{\Pi}(\omega_k)^{-1}\mathbf{W}(\omega_k)$.
-

3 章の提案するブロック処理を用いたインダイレクトモデルブラインド信号分離アルゴリズムの概要を箇条書きで下記に記す。

1. クロススペクトル密度行列の推定
2. ブロック処理による対象行列の削減
3. 2 で得た行列に対し逆行列の計算
4. インダイレクト型 ALSP アルゴリズムの実行
5. パーミュテーション問題の解法

3.4 シミュレーション

提案法と従来のダイレクトモデルブラインド信号分離 [13] の性能を比較する。シミュレーションプログラムは C 言語を用いて実装した。 N 個のスピーカと J 個のマイクロホンの配置を図 3.1 に示す。マイクロフォンの間隔は $d = 0.06$ m に設定した。標本化周波数 8 kHz、量子化ビット数 16 ビットで TSP (time stretched pulse) 信号を使用してスピーカ・マイクロフォン間の室内空間インパルス応答を測定し、音声データと室内空間インパルス応答を畳み込み混合して観測信号として使用した。室内の残響時間は約 300 ms であった。10 名の男性と 10 名の女性の音声データ [45] を使用し、 $J = N = 2$ の場合には 12 s、 $J = N = 3$ では 30 s、 $J = N = 3$ では 100 s の長さの音声データをそれぞれ使用した。マイクロフォンが 2 つの直線アレーでは 30° と -30° の方向から、マイクロフォンが 3 つの直線アレーでは 30° 、 0° 、 -30° の方向から、マイクロフォンが 4 つの直線アレーでは 40° 、 30° 、 0° 、 -30° の方向から、音声到来するようにスピーカが設置されている。STFT には 2,048 ポイントの FFT (fast Fourier transform) とハニング窓を使用し、フレーム数は窓関数を $T_s = 512$ シフトさせ、 $J = N = 2$ のときは $\mathcal{T} = 187$ 、 $J = N = 3$ のときは $\mathcal{T} = 468$ 、 $J = N = 4$ のときは $\mathcal{T} = 1562$ にそれぞれ設定した。クロススペクトル密度行列 $\mathbf{T}(m, \omega_k)$ の更新式において忘却係数は $\beta = 0.99$ 、初期値は $c = 10^{-2}$ に設定した。ALSP の初期値は文献 [13] に基づき設定した。ALSP アルゴリズムの反復終了基準は $\epsilon_P = 10^{-15}$ と $\epsilon_G = \epsilon_C = 10^{-6}$ にそれぞれ設定した。パーミュテーション行列は分離信号の近傍周波数ビン間における電力比の相関係数を用いて求めた [38]。従来法では、ウォルシュペリオドグラム法 [46] によって観測信号のクロススペクトル密度行列を近似し、スケーリング問題は各周波数ビンにおいて $\mathbf{W}(\omega_k)$ の列ベクトルを正規化することによって解法した。尚、提案法では、制約条件 $\|\mathbf{w}_i(\omega_k)\|_2 = 1$ のために、別途、スケーリング問題を解法する必要は生じない。従って、提案法と従来法は分離行列の推定方法だけが異なる。

平均出力 SIR とブロック長 L の関係を図 3.2 に示す。但し、 $J = N = 2$ 、 $M = 12000$ 、 $T_s = 8$ 、 $\text{SNR} \approx 20$ dB に設定している。平均入力 SIR は約 0.01 dB である。図 3.2 には、非パーミュテーション提案ブラインド手法の平均出力 SIR も示している。ブロック長を $L \leq 200$ に設定すると分離性能は $L = 1$ のときとほとんど変わらず、 $L > 200$ に設定するとブロック長を長くするほど分離性能は劣化した。 $L = 200$ のとき、本手法は非パーミュテーション提案ブラインド手法より分離性能が若干劣るが、演算量は $L = 1$ のときの 1/200 倍である。

次に、音源数 J とマイクロフォン N の個数を $J = N = 2$ から $J = N = 4$ に変化させて分離性能を比較する。比較対象は、ダイレクトモデルに基づく従来のブラインド信号分離 [13] とその非パーミュテーションブラインド手法、提案法とその非パーミュテーション提案ブラインド手法の 4 つである。音源数 J とマイクロフォン N の個数を $J = N = 2$ に設定したとき、図 3.3(a) に分離性能が得られた。ここで、 T_e はエポックサイズを表す。15 dB より高い SNR では、非パーミュテーション従来法と非パーミュテーション提案法は共に出力 SIR が最適値に達しなかった。これは、近似同時対角化推定の誤差が加法性雑音に比べて大きいためである。一方、SNR が低い場合

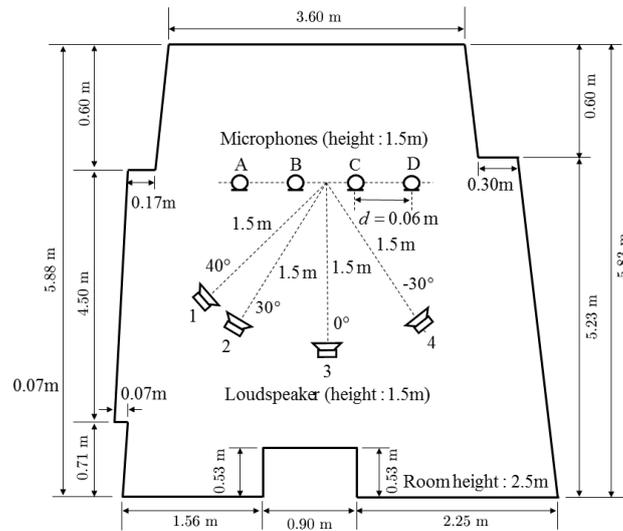


図 3.1: 小部屋における音源とマイクロフォンの配置図

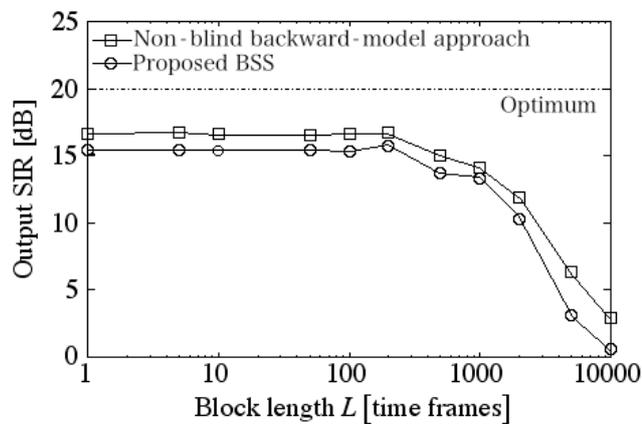


図 3.2: 平均出力 SIR とブロック長 L の関係 ($J = N = 2$, $M = 12000$, $T_s = 8$, $\text{SNR} \approx 20$ dB)

には、加法的雑音が推定誤差に比べ支配的になると考えられる。従来法と比較すると、提案法はロバスト的な観点から最大 1.31 dB 高い出力 SIR を得た。音源数 J とマイクロフォン N の個数を $J = N = 3$ に設定すると、図 3.3(b) の特性が得られた。SNR が 10 dB よりも高いとき、非パーミュテーション従来法と非パーミュテーション提案法は共に出力 SIR が最適値に達しなかった。SNR が高い範囲と低い範囲では、提案法は従来法よりも約 0.08 から 1.2 dB 高い分離性能を得た。音源数 J とマイクロフォン N の個数を $J = N = 4$ に設定すると、図 3.3(c) に示す分離性能

表 3.1: 勾配法に基づく従来のブラインド信号分離と提案法の分離性能の比較 ($J = N = 2$, $L = 200$, $\text{SNR} \approx 30 \text{ dB}$)

FDSOS BSS [35, 36]	Proposed BSS
1.65 dB	16.72 dB

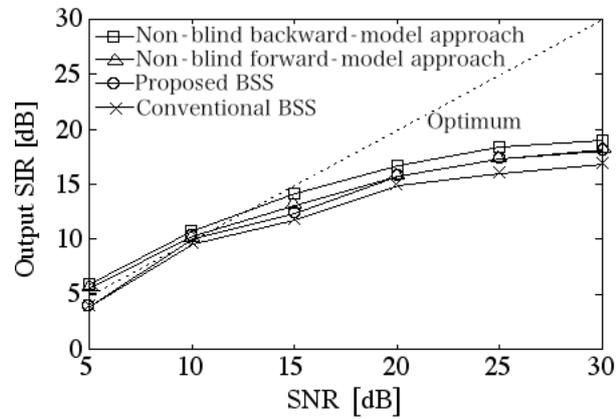
能が得られた。従来法より約 0.55 から 1.62 dB 高い分離性能が提案法によって達成された。

最後に、勾配法に基づく従来のブラインド信号分離 [35] と提案法の分離性能を表 3.1 に比較する。勾配法に基づく従来法については文献 [36] に記載された設定値を用いた。点音源から理想点マイクロフォンまでの室内空間インパルス応答は、寸法 $16.6 \times 11.2 \times 8.0 \text{ ft}$ の部屋を想定して標準化周波数 8 kHz で image method [47, 48] と呼ばれるシミュレータを用いて発生させた。2 つの点音源と 2 つの理想点マイクロフォンは文献 [36] と同じ位置に設置した。小部屋の反射係数も文献 [36] と同じ値 0.7、即ち、残響時間に換算して約 270 ms を採用した。平均入力 SIR は約 -0.67 dB であった。勾配法に基づく従来のブラインド信号分離に比べ、提案法は極めて高い分離性能を示した。

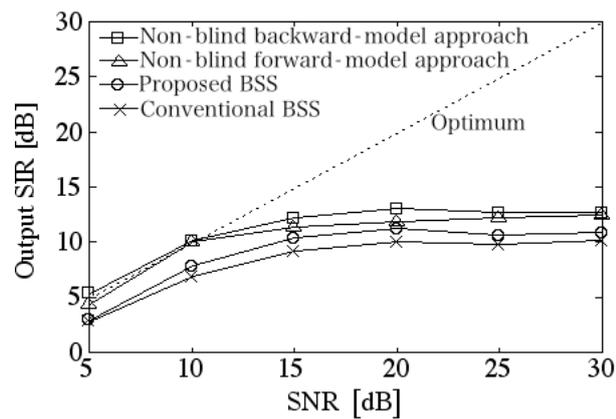
3.5 まとめ

同時対角化問題の解法にブロック処理を導入して、少ない演算量で高い分離性能を得るインダイレクトモデルブラインド信号分離法を提案した。指数関数的に重み付けされた観測信号のクロススペクトル密度行列の逆行列をブロック毎に計算し、その逆行列を展開式で表現することによって ALSF アルゴリズムの適用を可能とした。また、ブロック処理が演算量の削減に有効であること、隣接周波数ビン間における分離信号の電力比から求めた相関係数を利用する方法がパーミュテーション行列の解法に効果的であることを明らかにした。シミュレーションでは、提案法が従来法に比べ、SNR に依存することなく高い分離性能を発揮できることを示した。

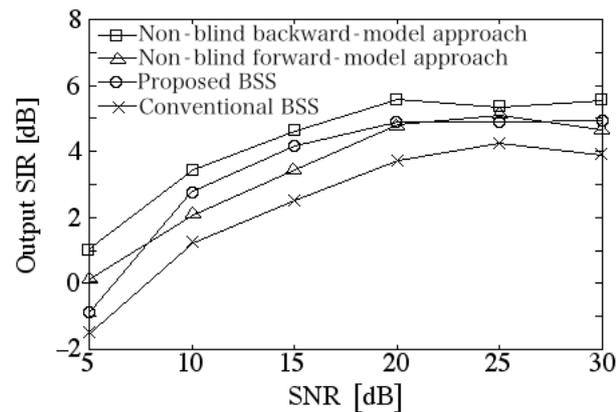
しかし、従来手法よりも高い SIR を得ることに成功したが、3 音源以上で高い精度で分離ができたとはいえず、問題点を全て克服できなかった。



(a) $J = N = 2$, $T_s = 8$, $L = 200$, and $T_e = 500$



(b) $J = N = 3$, $T_s = 15$, $L = 10$, and $T_e = 10$



(c) $J = N = 4$, $T_s = 30$, $L = 13$, and $T_e = 13$

図 3.3: 平均出力 SIR と SNR の関係 ($J = N = 2$ のとき平均入力 SIR は約 -0.14 dB、 $J = N = 3$ のとき約 -3.52 dB、 $J = N = 4$ のとき約 -5.87 dB)

第4章 DMDI アルゴリズムを用いた同時対角化問題の解法に基づくブラインド信号源分離

4.1 まえがき

この章では、3章で克服できなかった3音源以上で高い分離性能を得られる手法を提案する。4章では、非直交化 AJD 問題に DLS 評価量と ILS 評価量の両方を導入する。DLS 評価量は混合行列の関数で、ILS 評価量は対角行列の関数である。尚、本章以降では、原則、周波数ビンを表す添字を省略する。図 4.1 は周波数領域における畳み込み混合と分離過程を混合・分離モデルで表現したものである。観測信号 $\{x_i(m)\}_{i=1}^J$ のクロススペクトル密度行列を近似した後、これを正規化して $\tilde{\mathbf{P}}_x(m)$ を得る (図 4.1 のステップ 1)。混合モデルでは、混合行列 \mathbf{B} を求めるために、ラグランジェの未定乗数法によって CDLS 評価量 $C_{\text{DLS}}(\mathbf{B})$ を最小化する (ステップ 2)。混合行列 \mathbf{B} の推定値は近似同時対角化問題では正則でなければならない [10, 49]。そこで、混合行列 \mathbf{B} の推定値がフルランクでない場合は、これをフルランクな行列で置き換える。分離モデルでは、分離行列と混合行列の積 \mathbf{WB} が単位行列 \mathbf{I} でなければならないので [13]、 $\|\mathbf{I} - \mathbf{WB}\|_F^2$ を最小化するように混合行列から分離行列を求め (ステップ 3)、対角行列 $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ を求めるために、ILS 評価量 $C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M)$ を最小 2 乗法を用いて最小化する (ステップ 4)。提案する DMDI (Diagonalization implemented by Minimizing the Direct and Indirect least-squares criteria) アルゴリズムは、前の反復過程で推定した $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ を固定し \mathbf{B} に関して DLS 評価量 $C_{\text{DLS}}(\mathbf{B})$ を最小化し、前の反復過程で推定した \mathbf{B} を固定して $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ に関して ILS 評価量 $C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M)$ を最小化する手順を交互に繰り返す。ステップ 2~4 は収束するまで繰り返される。ステップ 5 では、混合・分離モデルが因果性となるように、 $e^{-j\pi k}\mathbf{W}$ によって分離行列を遅延分離行列に変換させる。

周波数領域ブラインド信号分離では、更に2つのステップが追加されます。DMDI アルゴリズムでは、各周波数ビンにおいて分離行列の各列ベクトル正規化することによってスケール問題を解法します。隣接周波数ビン間の電力比の相関係数を利用してパーミュテーション問題を解法する。音源とマイクロホンの個数が等しい状態で、DMDI アルゴリズムの分離性能を室内空間インパルス応答を用いて算出する。

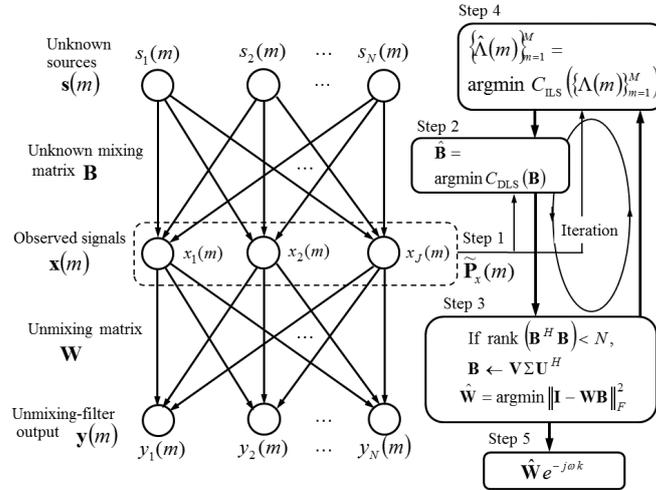


図 4.1: 周波数領域における混合・分離モデルと DMDI アルゴリズム (エポック m 、周波数ビン ω_k における J 個の観測信号は $\mathbf{x}(m) = \mathbf{B}\mathbf{s}(m)$ と表記している。周波数ビンの添字を $\mathbf{x}(m) \triangleq \mathbf{x}(m, \omega_k)$ 、 $\mathbf{B} \triangleq \mathbf{B}(\omega_k)$ 、 $\mathbf{s}(m) \triangleq \mathbf{s}(m, \omega_k)$ のように省略している。 N 個の信号源信号をベクトルで $\mathbf{s}(m)$ と表している。 \mathbf{B} は周波数ビン ω_k における混合行列である。また、 N 個の分離フィルタの出力信号を $\mathbf{y}(m) = \mathbf{W}\mathbf{x}(m) = \mathbf{W}\mathbf{B}\mathbf{s}(m)$ と表記している。ここで、周波数ビンの添字を $\mathbf{y}(m) \triangleq \mathbf{y}(m, \omega_k)$ 、 $\mathbf{W} \triangleq \mathbf{W}(\omega_k)$ のように省略している。 \mathbf{W} は周波数ビン ω_k における分離行列である。)

4.2 ブラインド信号分離のための DMDI アルゴリズムの導出

DMDI アルゴリズム [14, 33] は最小 2 乗型同時対角化問題の推定に基づき、DLS 評価量と ILS 評価量の両方を交互に最小化することが従来の同時対角化アルゴリズムとの違いである。ここで、DLS 評価量は混合行列の関数、ILS 評価量は対角行列の関数である。図 4.2 に示すように、DMDI アルゴリズムは次の 5 つのステップから成る。

- ステップ 1) 観測信号のクロススペクトル密度行列を推定し、これを正規化する
 - ステップ 2) CDLS 評価量を最小化することによって混合行列を推定する
 - ステップ 3) 推定した混合行列がランク落ちしている場合は、これをフルランク行列に置き換え、分離行列を算出する
 - ステップ 4) ILS 評価量を最小化することによって対角行列を推定する
 - ステップ 5) 加法性雑音の影響を低減するために分離行列を遅延させる。
- 周波数領域における畳み込み混合に DMDI アルゴリズムを適用する場合は
- ステップ 6) スケーリング問題を解法する
 - ステップ 7) パーミュテーション問題を解法する
- を付け加える。次の項では、DLS 評価量と ILS 評価量の役割について説明する。

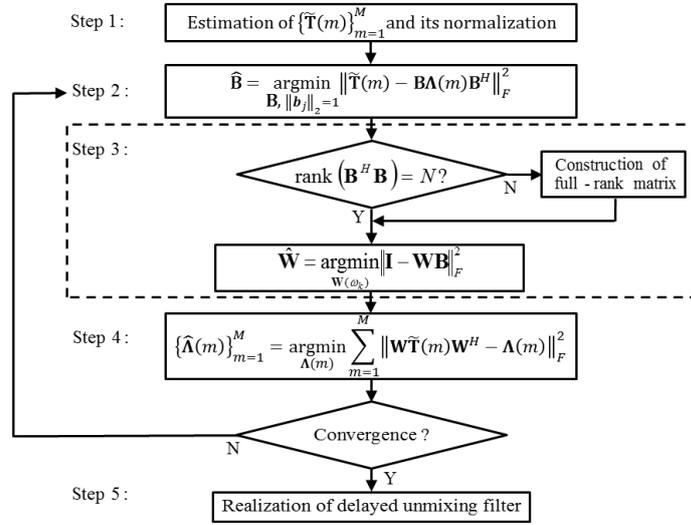


図 4.2: DMDI アルゴリズムのステップのダイアグラム

4.2.1 観測信号のクロススペクトル密度行列の推定と正規化

この節では、周波数ビンを表す添字を再導入する。ブラインド信号分離、ブラインド同定では、スケージング依存性を避けるために、観測信号のクロススペクトル密度行列を正規化する。観測信号のクロススペクトル密度行列の推定値 $\hat{\mathbf{T}}(m, \omega_k)$ をその振幅で周波数ビン毎に正規化することによって、スペクトル包絡は周波数帯域でフラットになる [13]。マイクロフォンの個数が増したとき、 $\hat{\mathbf{T}}(m, \omega_k)$ の非対角要素数も増加するので、結果として、スペクトル包絡は更に周波数帯域でフラットになる。これは、 $\hat{\mathbf{T}}(m, \omega_k)$ の全要素の電力に対して $\hat{\mathbf{T}}(m, \omega_k)$ の非対角要素の電力の比が大きくなるためである。

本提案手法では、周波数ビン毎ではなくエポック毎に観測信号のクロススペクトル密度行列の推定値 $\hat{\mathbf{T}}(m, \omega_k)$ を正規化する。音声信号は、低周波数帯のフォルマントによって特徴付けられる。スペクトル包絡で表現される特徴を失うことなく正規化を実現するために、エポック m において推定された観測信号のクロススペクトル密度行列 $\hat{\mathbf{T}}(m, \omega_k)$ をその最大振幅で

$$\hat{\mathbf{T}}(m, \omega_k) = \frac{\hat{\mathbf{T}}(m, \omega_k)}{\max_{\omega_k} \left\| \hat{\mathbf{T}}(m, \omega_k) \right\|_F} \quad (4.1)$$

のように除算する。観測信号のクロススペクトル密度行列の推定については付録で議論する。正規化によって求められたスペクトル包絡は、観測信号のスペクトル包絡の近似として扱うことができる。

4.2.2 CDLS 評価量の最小化による分離行列の推定

音源からマイクロフォンまでの混合過程において、音源信号が混合フィルタを通過することによって観測信号として受信される。それ故、CDLS 評価量 $C_{\text{CDLS}}(\mathbf{B})$ のラグランジェの未定乗数法による最小化は混合行列の推定に有効である。 M 個のエポックにおいて周波数ビン ω_k 毎の CDLS 評価量 $C_{\text{CDLS}}(\mathbf{B})$ の最小化問題は

$$C_{\text{CDLS}}(\mathbf{B}) = \sum_{m=1}^M \|\tilde{\mathbf{T}}(m) - \mathbf{B}\boldsymbol{\Lambda}(m)\mathbf{B}^H\|_F^2 + 2 \sum_{j=1}^N \gamma_j (\mathbf{b}_j^H \mathbf{b}_j - 1) \quad (4.2)$$

を最小化することによって混合行列 \mathbf{B} を求めるために用いられる。ここで、 γ_j はラグランジェの未定乗数である。この問題の解は

$$\hat{\mathbf{g}}_i = \frac{1}{\|\mathbf{z}_i\|_2^2} \left[\mathbf{F}_i \mathcal{R}e(\mathbf{z}_i) - \frac{1}{J} \left(\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{F}_i) \mathcal{R}e(\mathbf{z}_i) - \|\mathbf{z}_i\|_2^2 \right) \text{vec}\{\mathbf{I}\} \right] \quad (4.3)$$

である。式 (4.3) の導出については付録を参照されたい。 \mathbf{G} の各列ベクトルに関する式 (D.9) を最小化した後、式 (4.3) が繰り返し求められる。このとき、最小化の対象とならない列ベクトルは固定される。算出前と算出後の \mathbf{G} の差を誤差行列として記憶する。この繰り返しは、誤差行列のフロベニウスノルムが反復終了基準 ϵ_G を下回るまで続けられる。反復終了後、 \mathbf{b}_i は累乘法 [44] によって $\|\text{vec}^{-1}\{\hat{\mathbf{g}}_i\} - \mathbf{b}_i \mathbf{b}_i^H\|_F^2$ を最小化することによって ALS 反復に 1 度求められる [13]。

4.2.3 正則な分離行列の構築

AJD 問題では、推定された分離行列は正則でなければならない [10, 49]。そこで、 $\mathbf{B}^H \mathbf{B}$ が正則な行列でないとき、DMDI アルゴリズムでは、対角行列を推定する前に式 (4.2) の最小化を満たすことを前提に \mathbf{B} を正則な行列で置き換える。

\mathbf{B} の特異値分解は

$$\mathbf{B} = \mathbf{V}_r \boldsymbol{\Sigma}_r \mathbf{U}_r^H \quad (4.4)$$

によって与えられる。ここで、 \mathbf{V}_r と \mathbf{U}_r はユニタリ行列

$$\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r] \quad (4.5)$$

$$\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \quad (4.6)$$

$\boldsymbol{\Sigma}_r$ は対角行列

$$\boldsymbol{\Sigma}_r = \text{diagonal}(\sigma_1, \sigma_2, \dots, \sigma_r) \quad (4.7)$$

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 \quad (4.8)$$

をそれぞれ表す。条件 $\|\mathbf{b}_i\|_2 = 1$ は $\boldsymbol{\Sigma}_r$ のトレースを \sqrt{N} に等しくする、即ち、 $\text{tr}[\boldsymbol{\Sigma}_r] = \sqrt{N}$ によって保証される。正規直交基底 $\{\mathbf{v}_i\}_{i=1}^r$ に直交する正規直交基底 $\{\mathbf{v}_i\}_{i=r+1}^N$ と、正規直交基底

$\{\mathbf{u}_i\}_{i=1}^r$ に直交する正規直交基底 $\{\mathbf{u}_i\}_{i=r+1}^N$ を使用して、行列 \mathbf{B} のランクが $r < N$ のとき、行列 \mathbf{B} を正則行列

$$\mathbf{V}\Sigma\mathbf{U}^H = \frac{\sqrt{N}}{\sqrt{N} + \delta N} [\mathbf{V}_r, \bar{\mathbf{V}}_r] \left(\begin{bmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \delta\mathbf{I} \right) [\mathbf{U}_r, \bar{\mathbf{U}}_r]^H \quad (4.9)$$

に置き換える。ここで、 $\delta > 0$ とし、式 (4.9) の右辺の $\sqrt{N}/(\sqrt{N} + \delta N)$ は

$$\text{tr}[\Sigma] = \sqrt{N} \quad (4.10)$$

を満たさなければならない。また

$$\bar{\mathbf{V}}_r = [\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_N] \quad (4.11)$$

$$\bar{\mathbf{U}}_r = [\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_N] \quad (4.12)$$

であり、 $\bar{\mathbf{V}}$ と添字 r は、 $\{\mathbf{v}_i\}_{i=r+1}^N$ が $\{\mathbf{v}_i\}_{i=1}^r$ の直交補空間であることを示している。 $\{\mathbf{v}_i\}_{i=1}^r$ と $\{\mathbf{u}_i\}_{i=1}^r$ はそれぞれ $\{\mathbf{v}_i\}_{i=r+1}^N$ と $\{\mathbf{u}_i\}_{i=r+1}^N$ に直交関係にあること、また、式 (4.2) の条件が満たされていることから、 $\mathbf{V}\Sigma\mathbf{U}^H$ によって \mathbf{B} を置き換えても式 (4.2) による最小化には影響を与えない。

式 (2.9) より、分離行列 \mathbf{W} は

$$\varepsilon(\mathbf{W}) = \mathbf{I} - \mathbf{W}\mathbf{B} \quad (4.13)$$

を最小にすることによって推定される。即ち、混合フィルタと分離フィルタを連続に接続した合成フィルタの応答は

$$\xi(\mathbf{W}) = \|\varepsilon(\mathbf{W})\|_F^2 = \|\mathbf{I} - \mathbf{W}\mathbf{B}\|_F^2 \quad (4.14)$$

を最小にすることによって求めることができる。 $\mathbf{B}^H\mathbf{B}$ のランクが $r < N$ のとき、式 (4.14) の値は最小値 $\xi_{\min} = 0$ より大きくなる。一方、 $\mathbf{B}^H\mathbf{B}$ が正則のとき、式 (4.14) の値は最小値 $\xi_{\min} = 0$ に等しくなる。このように、正則な分離行列は正則な混合行列から得られ、このとき、 $\mathbf{W}\mathbf{B}$ は単位行列に等しくなる。

4.2.4 ILS 評価量の最小化による対角行列の推定

音源から分離フィルタ出力までの過程において、 $\mathbf{W}\mathbf{B} = \mathbf{I}$ と

$$\hat{\mathbf{T}}(m) = \mathbf{B}\Lambda(m)\mathbf{B}^H + \sigma^2\mathbf{I} \quad (4.15)$$

を用いると

$$\Lambda(m) = \mathbf{W}\mathbf{B}\Lambda(m)\mathbf{B}^H\mathbf{W}^H = \mathbf{W}(\mathbf{P}_x(m) - \sigma^2\mathbf{I})\mathbf{W}^H \quad (4.16)$$

が成り立つ。誤差

$$\mathbf{Z}(m) = \mathbf{W}\tilde{\mathbf{T}}(m)\mathbf{W}^H - \mathbf{\Lambda}(m) \quad (4.17)$$

を用いて ILS 評価量

$$C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M) = \sum_{m=1}^M \|\mathbf{W}\tilde{\mathbf{T}}(m)\mathbf{W}^H - \mathbf{\Lambda}(m)\|_F^2 \quad (4.18)$$

を最小にする対角行列は

$$\hat{\mathbf{\Lambda}}(m) = \text{diag}[\mathbf{W}\tilde{\mathbf{T}}(m)\mathbf{W}^H] \quad (4.19)$$

と推定される。このとき、 \mathbf{W} は定数として扱っている。式 (D.5) によって \mathbf{B} と $\hat{\mathbf{\Lambda}}(m)$ から $\mathbf{Gd}(m)$ と求め、式 (D.1) の変化が ALS 反復終了基準 ϵ_C を下回るまで、ALS 反復は繰り返される。

4.2.5 遅延分離行列

分離フィルタは信号源分離だけでなく逆フィルタの役割も有する。混合フィルタが安定で、因果性を満たし、最小位相であるとき、分離フィルタのインパルス応答を遅延させることによって加法性雑音の影響を低減させることができる。このため、 \mathbf{W} に $e^{-j\pi k}$ を乗算する。スケーリング行列 $\mathbf{\Pi}^{-1}\mathbf{W}$ とパーミュテーション行列 $\mathbf{P}^{-1}\mathbf{\Pi}^{-1}\mathbf{W}$ を求めた後、 $e^{-j\pi k}\mathbf{\Pi}^{-1}\mathbf{D}^{-1}\mathbf{W}$ を逆離散フーリエ変換によって時間領域に変換して分離フィルタのインパルス応答を算出する。DMDI アルゴリズムを Algorithm 2 に要約する。

4.3 シミュレーションと実環境による実験

4.3.1 シミュレーション条件

雑音と残響が存在する環境下で分離性能を比較するために、FDSOS [35,36]、AC-DC [5]、ALSP [13]、PARAFAC with enhanced line search (ELS) [24]、ALUJA [27] を条件

- 1) 音源からマイクロフォンまでの混合過程が畳み込み混合で、ガウス性白色雑音を加法していること
- 2) 混合行列、または、分離行列が周波数ビン毎に評価量を反復的に最小化することによって推定されること
- 3) 入手できる観測信号行列の個数が有限であること

に基づき選択した。尚、文献 [35,36] で提案された FDSOS は FDSOS I と表記した。また、分離行列の推定に最急勾配法 [35,36]、パーミュテーション問題の解法に文献 [38] を用いた方法を FDSOS II と表記する。シミュレーションプログラムは C 言語を用いて実装した。

Algorithm 2 : DMDI アルゴリズム

-
- Estimation of the cross-spectral density matrices of the observed signal:**
- 1: Compute $\hat{\mathbf{T}}(m)$ at each epoch using the Welch periodogram method for $m = 1, 2, \dots, M$.
- Normalization of $\hat{\mathbf{T}}(m)$:**
- 2: Normalize $\hat{\mathbf{T}}(m)$ by (4.1) for $m = 1, 2, \dots, M$.
- Initialization:**
- 3: Set the initial values of $\mathbf{B}(\omega_k)$ and $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ to the LS-AJD estimate at frequency bin ω_{k-1} with $\mathbf{B}(\omega_0) = \mathbf{I}$ and $\{\mathbf{\Lambda}(m, \omega_0) = \mathbf{I}\}_{m=1}^M$.
- 4: **while** The absolute value of the difference between the current and previous values of $C_{\text{DLS}}(\mathbf{G})$ in the ALS iterations is larger than ϵ_C . **do**
- First phase (Estimation of the mixing matrix):**
- 5: **while** The absolute value of the difference between the current and previous values of \mathbf{G} is larger than ϵ_G . **do**
 - 6: Minimize $C_{\text{DLS}}(\mathbf{g}_i)$ to find $\hat{\mathbf{g}}_i$ by the method of Lagrange multipliers for $i = 1, 2, \dots, N$ according to (4.3).
 - 7: **end while**
 - 8: Minimize $\|\text{vec}^{-1}\{\hat{\mathbf{g}}_i\} - \mathbf{b}_i \mathbf{b}_i^H\|_F^2$ to find \mathbf{b}_i by the power method for $i = 1, 2, \dots, N$.
- Construction of the full-rank unmixing matrix:**
- 9: **if** $\text{rank}(\mathbf{B}^H \mathbf{B}) = r < N$ **then**
 - 10: Construct the orthonormal sets $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_N$ and $\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_N$ by the Gram-Schmidt process.
 - 11: Replace \mathbf{B} with $\mathbf{V} \Sigma \mathbf{U}^H$ as given by (4.9).
 - 12: **end if**
 - 13: Compute \mathbf{W} from \mathbf{B} according to (5.26).
- Second phase (Estimation of the diagonal matrices):**
- 14: Minimize $C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M)$ to find $\hat{\mathbf{\Lambda}}(m)$ through the method of least squares according to (4.19).
 - 15: Compute $\mathbf{Gd}(m)$ from \mathbf{B} and $\hat{\mathbf{\Lambda}}(m)$ by (D.5).
 - 16: **end while**
- Realization of the delayed unmixing filter:**
- 17: Multiply \mathbf{W} by $e^{-j\pi k}$.
Solution of the scaling and permutation problems:
 - 18: Solve the scaling problem to find $\mathbf{D}^{-1} \mathbf{W}$ after obtaining the LS-AJD estimate in each frequency bin independently.
 - 19: Resolve the permutation ambiguity to find $\mathbf{\Pi}^{-1} \mathbf{D}^{-1} \mathbf{W}$.
 - 20: Convert $e^{-j\pi k} \mathbf{\Pi}^{-1} \mathbf{D}^{-1} \mathbf{W}$ into the time domain by the inverse FFT.
-

音声コーパス [45] から入手した男女の音声データを繋ぎ合わせて音源として使用した。図 4.3 に各音源の信号波形を一部示す。音声データ長は 1000 s で、各マイクロオンで集音したデータの振幅が等しくなるように正規化した。点音源から理想点マイクロフォンまでの室内空間インパルス応答は、小部屋を想定して標準化周波数 8 kHz で image method [47, 48] と呼ばれるシミュレータを用いて発生させた。残響時間は 100, 300, 500, 700, 900 ms に設定した。

文献 [36] と同じ大きさ $5.06 \times 3.41 \times 2.44$ m の部屋を用いて $J = N = 2$ の各 BSS の性能を比較した。2 つのマイクロフォンは座標 (0.30, 2.74, 1.68) と (4.75, 2.74, 1.68)、2 つの音源は座標 (2.13, 0.91, 1.52) と (3.35, 0.91, 1.52) にそれぞれ配置した [36]。

文献 [50] と同様に、 $J = N = 3$ と $J = N = 4$ については寸法 $4.45 \times 3.55 \times 2.50$ m の部屋を用いて各 BSS の性能を比較した。マイクロホンの指向性の影響を無くすために、 $J = N = 3$ では図 4.4 に示すように一辺の長さが 0.2 m の正三角形の頂点 (2.34, 1.78, 1.2)、(2.17, 1.88, 1.2)、(2.17, 1.68, 1.2) に合計 3 つのマイクロフォンを設置した。3 つの音源は同図の点 A(3.35, 1.36, 1.2)、

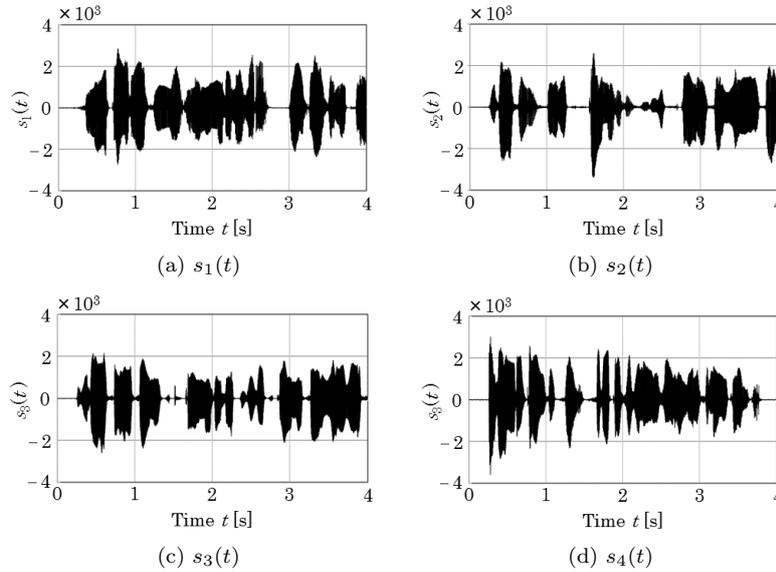
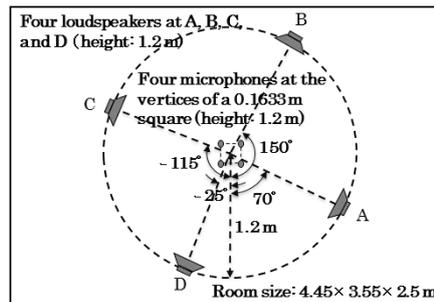


図 4.3: 各音源の信号波形


 図 4.4: 小部屋における音源とマイクロフォンの配置図 ($N = 3$ と $N = 4$)

点 B(2.83, 2.81, 1.2)、点 C(1.14, 2.28, 1.2) に設置した。正三角形の重心は半径 1.2m の円の中心に重ねた。 $J = N = 4$ では、図 4.4 の一辺の長さが 0.1633m の正方形の頂点 (2.34, 1.78, 1.2)、(2.23, 1.89, 1.2)、(2.11, 1.78, 1.2)、(2.23, 1.66, 1.2) に合計 4 つのマイクロフォンを配置した。4 つの音源は同図の点 A(3.35, 1.36, 1.2)、点 B(2.83, 2.81, 1.2)、点 C(1.14, 2.28, 1.2)、点 D(1.72, 0.69, 1.2) に設置した。

STFT にはハニング窓を使用し、重複率 80% のフレームで 8,192 ポイントの FFT を用いた [13]。DMDI では、ALS 反復終了基準を $\epsilon_G = \epsilon_C = 10^{-6}$ に設定した。また、正則な分離行列の構築のために $\delta = \sigma_r$ を用いた。FDSOS I を除き、観測信号 $\mathbf{x}(m, \omega_k)$ のクロススペクトル密度行列は、エポック当たり重複率 80% の 2 つのフレームを用いたウォルシュペリオドグラム法 [46] によって近似した。FDSOS I では、エポック当たりの非重複フレーム数は 9 に設定した [35]。DMDI BSS では、推定した観測信号 $\mathbf{x}(m, \omega_k)$ のクロススペクトル密度行列はエポック毎に、また、FDSOS I と ALUJA を除く他の BSS では各ポイント毎に振幅で正規化した。PARAFAC では、 $\mathbf{B}(\omega_0)$ と $\{\mathbf{\Lambda}(m, \omega_0)\}_{m=1}^M$ は白色雑音で初期化した [24]。その他の ALS アルゴリズムでは、単位行列で

$\mathbf{B}(\omega_0)$ と $\{\mathbf{\Lambda}(m, \omega_0)\}_{m=1}^M$ を初期化した。ALUJA を除く全 ALS アルゴリズムでは、 $\mathbf{B}(\omega_1)$ と $\{\mathbf{\Lambda}(m, \omega_1)\}_{m=1}^M$ は、周波数ビン ω_0 における推定値に初期化した。この初期化手順は、周波数ビン $\omega_{K/2}$ における推定の初期化まで続けた [13]。ALUJA では、 $\mathbf{B}(\omega_k)$ と $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ は全周波数ビンにおいて単位行列に初期化した。最急降下法に基づく FDSOS I と II では、周波数ビン毎に最も高い性能が得られる値にステップゲインを設定した。全 BSS において、 \mathbf{W} に $e^{-j\pi k}$ を乗算した後、FDSOS I と ALUJA を除く他の BSS では、各周波数ビンにおいて \mathbf{W} の列ベクトルを正規化することによってスケールリング問題を解法した¹。また、FDSOS I を除く BSS では、パーミュテーション行列は近傍周波数ビン間の電力比の相関係数を用いて求めた [38]。FDSOS I を除く従来法と DMDI の違いは、観測信号のクロススペクトル密度行列の正規化方法と分離行列の推定法である。実環境での性能評価は節 4.3.5 で述べる。

4.3.2 評価基準と非パーミュテーションブラインド手法

DMDI アルゴリズムと従来の代表的なアルゴリズムの出力 SIR を分離性能の評価基準として採用する。非パーミュテーションブラインド手法は FDSOS I では実現する不可能である。これは、分離フィルタ長における制約が分離フィルタの音源割り当てを満足するためである [35, 36]。即ち、FDSOS I において分離フィルタ長を Q で表現すると、 K/Q を 8 に設定した [25, 35, 36]。プログラムは C 言語で作成し、インテル製の Core i7-2600 3.4 GHz で倍精度浮動小数点で演算した。パーミュテーションとスケールリング問題解法に費やす時間は演算時間評価には含まれていない。

4.3.3 残響環境下における性能比較

表 4.1、4.2、4.3 にそれぞれ $N = J = 2$ 、 $N = J = 3$ 、 $N = J = 4$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数を示す。非パーミュテーションブラインド手法の性能も同表に列記する。DMDI は ALUJA に基づく非パーミュテーションブラインド手法を除き、残響時間 100 ms の室内環境において $N = J = 3$ のとき最も高い分離性能を得た。分離性能は非パーミュテーションブラインド手法の出力 SIR に近い結果となった。これは、近傍周波数ビン間の電力比の相関係数を用いて求めたパーミュテーション問題が正しく解法できていることを意味する。残響時間 300 ms の室内環境においては、 $N = J = 2$ のとき DMDI は PARAFAC BSS より 2.97 dB 高い出力 SIR を得た。残響時間 300 ms の室内環境においては、 $N = J = 3$ と $N = J = 4$ のとき DMDI は ALUJA BSS よりそれぞれ 1.13 dB と 0.36 dB 高い出力 SIR を得た。残響時間を 100 ms から 900 ms に長くすると、全ての BSS の性能は悪化した。残響時間が分離性能に多大な影響を与えることが確認できた。

残響時間 300 ms の室内環境において、DMDI アルゴリズムは $N = J = 2$ のとき周波数ビン当たり平均 3.81 の反復回数、 $N = J = 3$ のとき周波数ビン当たり平均 5.25 の反復回数、 $N = J = 4$ のとき周波数ビン当たり平均 7.77 の反復回数で収束した。一方、 $N = J = 2$ のとき、ALSP アル

¹ALUJA においては、分離行列の対角要素が 1 に設定されるので、スケールリング問題は生じない [27]。

ゴリズムは収束に時間が掛かった。DMDI アルゴリズムの演算速度は、全アルゴリズムの中で最も速いか、2 番目に速かった。残響時間 300 ms の室内環境において $N = J = 2$ のとき、PARAFAC アルゴリズムは 2 番目に高い出力 SIR であったが、演算時間は DMDI アルゴリズムの 34.37 倍であった。残響時間 300 ms の室内環境において $N = J = 3$ と $N = J = 4$ のときは、ALUJA アルゴリズムは 2 番目に高い出力 SIR であったが、演算時間は $N = J = 3$ のときは DMDI アルゴリズムの 2.20 倍、 $N = J = 4$ のときは DMDI アルゴリズムの 1.77 倍であった。

4.3.4 雑音環境下における性能比較

残響時間 300 ms の室内環境において、SNR を変えて各 BSS の性能を表 4.4 で比較した。全 BSS 共に SNR が良くなると、反復回数は減っていった。DMDI アルゴリズムが、最も少ない、或いは 2 番目に少ない演算量、また、最も少ない反復回数で最も高い分離性能を示し、非ブラインド非パーミュテーション手法との性能比較においても大差が無かった。これは、近傍周波数ビン間の電力比の相関係数がパーミュテーション問題の解法に有効であることを意味している。SNR が 20 dB よりも高い雑音環境においては、非パーミュテーションブラインド手法は最適な出力 SIR に達しなかった。これは、SNR が低い雑音環境下ではガウス性白色雑音が支配的になるが、SNR が高い雑音環境下では推定誤差が支配的になるためである。DMDI BSS は出力 SIR が ALUJA BSS と比較して最大 3.77 dB 高く、雑音に強いという結果となった。

4.3.5 実環境における性能評価

4 個の指向性スピーカと 4 本のマイクロフォンを図 4.5 の空室に配置して実環境における DMDI BSS の性能を評価した。部屋の形状、大きさは異なるが、図 4.4 と同じになるようスピーカとマイクロフォンの配置は決めた。スピーカとマイクロフォンの特性に影響されることなく分離性能を評価するために、TSP 信号を用いて 4 個の指向性スピーカから 4 本のマイクロフォンまでの室内空間インパルス応答を標準化周波数 8 kHz、量子化ビット数 16 ビットで 16 回測定した後、これらを平均し、音声データと畳み込み演算して観測信号を生成した。図 4.5 のスピーカ A からマイクロフォンまでの室内空間インパルス応答とその残響特性を図 4.6 に示す。同図より残響時間は約 350 ms であった。観測信号にはガウス性白色雑音を加法した。シミュレーション条件は、前節のシミュレーション条件と同じである。この節では、Algorithm 2 に示した DMDI アルゴリズムを用いて実環境における性能を評価した。

式 (2.28) によって定義される $\mathbf{\Gamma}$ の i 行 j 列の要素の大きさ $|\gamma_{ij}|$ 、即ち、振幅特性を図 4.7 に示す。分離性能が 4 個の指向性スピーカと 4 本のマイクロフォン間の混合フィルタと分離フィルタの縦続接続の振幅特性によって示されている。図 4.7 では、分離フィルタの 1, 2, 4, 3 番目の出力が 1, 2, 3, 4 番目の行の振幅特性にそれぞれ割り当てられている。

同様に、図 4.8 に i 番目のスピーカから j 番目のマイクロフォンまでの混合分離縦続接続フィルタのインパルス応答を示す。これらのインパルス応答は、 $\mathbf{\Gamma}$ の IDFT によって得られる。全て

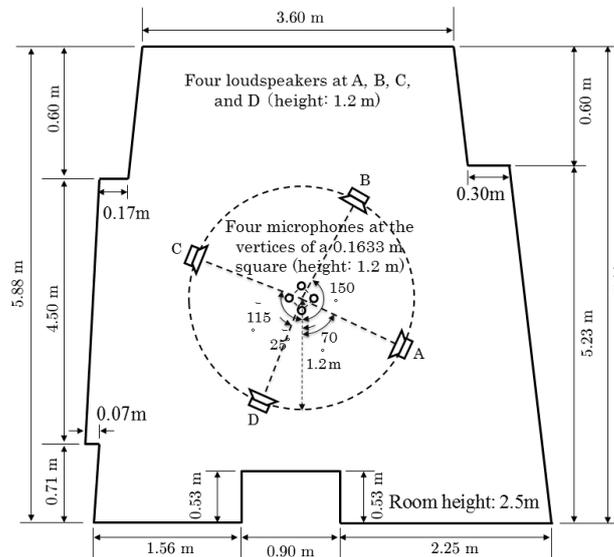


図 4.5: 実環境におけるスピーカとマイクロフォンの配置図 ($N = 4$)

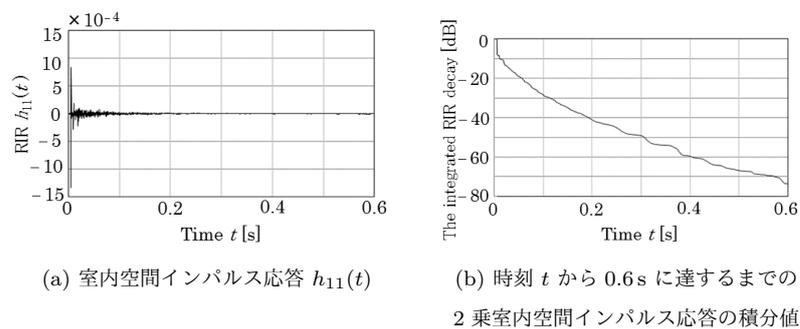


図 4.6: スピーカ A から図 4.5 に描かれた円の中心を通る垂線上で円の中心より下側に位置する無指向性マイクロフォンまでの室内空間インパルス応答とその減衰曲線

表 4.1: $N = J = 2$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 (SNR \approx 20 dB, 8192 ポイント FFT)

Method	Reverberation time [ms]				
	100	300	500	700	900
	Overall input SIR [dB]				
	0.25	-0.57	-1.02	-1.34	-1.58
DMDI					
Output SIR [dB]	19.80	19.75	18.96	17.66	16.30
CPU time [s]	0.033	0.035	0.037	0.039	0.040
Iterations	3.54	3.81	3.99	4.22	4.33
Nonblind approach					
Output SIR [dB]	19.95	19.90	19.25	17.95	16.71
ALUJA [27]					
Output SIR [dB]	18.07	16.18	14.27	12.77	11.02
CPU time [s]	0.049	0.061	0.067	0.069	0.072
Iterations	24.85	31.06	34.02	34.75	36.65
Nonblind approach					
Output SIR [dB]	18.38	16.99	14.55	13.49	11.83
PARAFAC [24]					
Output SIR [dB]	18.41	16.78	16.49	15.07	14.01
CPU time [s]	1.110	1.203	1.226	1.229	1.281
Iterations	56.45	61.19	62.36	62.53	65.17
Nonblind approach					
Output SIR [dB]	18.50	17.36	16.88	15.59	14.55
ALSP [13]					
Output SIR [dB]	14.09	12.42	11.97	11.05	10.72
CPU time [s]	1.396	1.501	1.510	1.515	1.539
Iterations	162.76	177.89	178.91	179.46	182.26
Nonblind approach					
Output SIR [dB]	14.89	12.93	12.42	11.20	11.03
AC-DC [5]					
Output SIR [dB]	6.40	6.05	5.81	5.26	4.68
CPU time [s]	0.586	0.621	0.631	0.651	0.652
Iterations	157.21	166.70	169.44	174.82	175.18
Nonblind approach					
Output SIR [dB]	7.23	6.67	6.26	5.78	5.13
FDSOS I [35, 36]					
Output SIR [dB]	4.35	3.03	2.91	0.92	0.63
CPU time [s]	0.033	0.034	0.037	0.040	0.043
Iterations	179.91	184.91	199.90	217.50	236.20
Nonblind approach					
Output SIR [dB]	—	—	—	—	—
FDSOS II [36, 38]					
Output SIR [dB]	3.63	2.24	1.94	1.84	1.77
CPU time [s]	0.538	0.567	0.571	0.574	0.576
Iterations	361.21	380.68	383.39	386.00	386.78
Nonblind approach					
Output SIR [dB]	4.22	3.43	2.98	2.59	2.45

表 4.2: $N = J = 3$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 (SNR ≈ 20 dB、8192 ポイント FFT)

Method	Reverberation time [ms]				
	100	300	500	700	900
	Overall input SIR [dB]				
	-2.59	-3.39	-3.63	-3.66	-3.62
DMDI					
Output SIR [dB]	17.43	17.23	15.27	13.62	11.93
CPU time [s]	0.124	0.179	0.182	0.190	0.206
Iterations	4.77	5.25	6.04	6.33	7.03
Nonblind approach					
Output SIR [dB]	17.77	17.39	15.71	14.02	12.55
ALUJA [27]					
Output SIR [dB]	17.75	16.10	13.89	10.10	8.48
CPU time [s]	0.390	0.393	0.399	0.405	0.422
Iterations	39.00	39.32	39.90	40.48	42.23
Nonblind approach					
Output SIR [dB]	18.05	17.32	13.97	11.65	10.38
PARAFAC [24]					
Output SIR [dB]	12.12	10.12	8.07	7.23	6.73
CPU time [s]	6.619	9.539	9.799	10.454	10.560
Iterations	169.66	244.50	251.16	268.47	270.70
Nonblind approach					
Output SIR [dB]	12.53	10.46	9.00	7.90	7.02
ALSP [13]					
Output SIR [dB]	10.13	4.65	3.25	2.67	1.56
CPU time [s]	3.544	13.905	14.979	16.336	17.246
Iterations	110.41	147.05	157.12	158.18	162.81
Nonblind approach					
Output SIR [dB]	10.17	4.84	3.67	2.75	1.72

表 4.3: $N = J = 4$ の平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 (SNR ≈ 20 dB、8192 ポイント FFT)

Method	Reverberation time [ms]				
	100	300	500	700	900
	Overall input SIR [dB]				
	-4.23	-5.19	-5.39	-5.54	-5.90
DMDI					
Output SIR [dB]	14.89	14.71	13.88	11.86	8.86
CPU time [s]	0.952	1.284	1.481	1.714	1.788
Iterations	5.76	7.77	8.96	10.37	10.82
Nonblind approach					
Output SIR [dB]	16.59	16.31	14.82	12.73	10.73
ALUJA [27]					
Output SIR [dB]	14.55	14.35	10.54	7.09	6.12
CPU time [s]	2.246	2.277	2.291	2.305	2.346
Iterations	43.63	44.24	44.52	44.79	46.08
Nonblind approach					
Output SIR [dB]	17.44	15.82	12.98	11.27	9.91
PARAFAC [24]					
Output SIR [dB]	8.10	2.39	1.38	0.58	-1.03
CPU time [s]	62.491	77.357	81.470	83.366	91.967
Iterations	1063.01	1315.90	1385.86	1418.11	1451.67
Nonblind approach					
Output SIR [dB]	9.23	3.88	2.78	2.28	1.38

表 4.4: 残響時間 300 ms の室内環境における平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 (8192 ポイント FFT、 $N = J = 2$)

Method	SNR [dB]						
	0	5	10	15	20	25	30
	Overall input SIR [dB]						
	-1.83	-1.17	-0.90	-0.66	-0.57	-0.55	-0.54
DMDI							
Output SIR [dB]	-0.02	4.92	9.73	14.21	19.75	20.64	23.11
CPU time [s]	0.126	0.086	0.058	0.043	0.035	0.031	0.030
Iterations	14.01	9.57	6.43	4.73	3.81	3.30	3.16
Nonblind approach							
Output SIR [dB]	-0.01	4.98	9.79	14.75	19.90	21.30	23.57
ALUJA [27]							
Output SIR [dB]	-0.66	3.41	9.18	13.51	16.18	19.02	19.34
CPU time [s]	0.176	0.159	0.147	0.107	0.061	0.039	0.031
Iterations	89.62	81.20	74.85	54.47	31.06	19.77	15.55
Nonblind approach							
Output SIR [dB]	-0.01	4.90	9.70	13.65	16.99	19.37	20.80
PARAFAC [24]							
Output SIR [dB]	-0.05	4.45	8.65	12.35	16.78	19.65	21.56
CPU time [s]	1.645	1.555	1.446	1.367	1.203	0.951	0.776
Iterations	83.66	79.08	73.55	69.51	61.19	48.42	39.59
Nonblind approach							
Output SIR [dB]	-0.02	4.73	9.05	13.70	17.36	19.99	22.06
ALSP [13]							
Output SIR [dB]	-1.11	2.07	7.50	10.63	12.42	13.23	14.61
CPU time [s]	1.624	1.598	1.562	1.540	1.501	1.445	1.387
Iterations	192.40	189.30	185.08	182.46	177.89	171.28	164.38
Nonblind approach							
Output SIR [dB]	-0.31	3.92	7.91	11.26	12.93	13.72	15.21
AC-DC [5]							
Output SIR [dB]	-1.12	2.03	5.42	5.70	6.05	6.34	6.67
CPU time [s]	0.884	0.822	0.776	0.684	0.621	0.511	0.324
Iterations	237.42	220.73	208.38	183.72	166.70	137.19	86.94
Nonblind approach							
Output SIR [dB]	-0.33	3.15	5.44	5.93	6.67	6.79	7.09
FDSOS I [35, 36]							
Output SIR [dB]	-1.26	1.18	2.17	2.53	3.03	3.11	3.42
CPU time [s]	0.045	0.042	0.040	0.036	0.034	0.033	0.033
Iterations	244.83	226.94	219.55	198.42	184.91	182.05	181.24
Nonblind approach							
Output SIR [dB]	—	—	—	—	—	—	—
FDSOS II [36, 38]							
Output SIR [dB]	-1.60	-0.02	1.56	1.95	2.24	2.45	2.49
CPU time [s]	0.647	0.600	0.580	0.578	0.567	0.564	0.561
Iterations	434.11	402.39	389.29	388.062	380.68	378.45	376.76
Nonblind approach							
Output SIR [dB]	-1.13	0.07	2.89	3.30	3.43	3.49	3.50

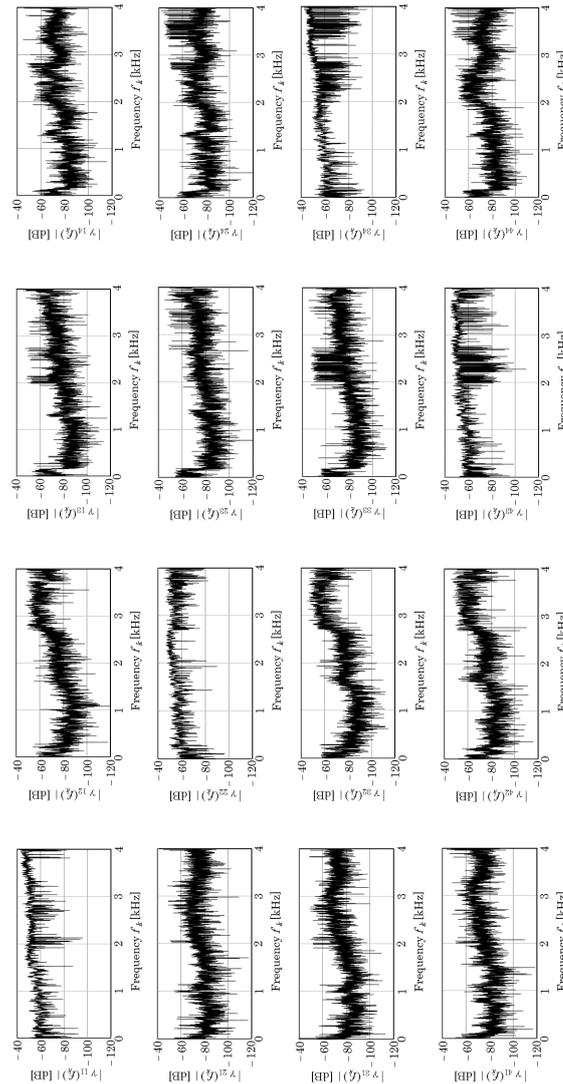


図 4.7: 実環境において DMDI BSS によって推定した i 番目のスピーカと j 番目のマイクロフォン間の混合分離接続フィルタの振幅特性 $|\gamma_{ij}|$ (SNR ≈ 20 dB, 8192 ポイント FFT, $N = J = 4$)

表 4.5: 実環境における平均出力 SIR、周波数ビン当たりの平均演算時間、収束までに要する周波数ビン当たりの平均反復回数の比較 (SNR \approx 20 dB、8192 ポイント FFT、 $N = J = 4$)

Method	Overall input SIR: -4.27 dB
DMDI	
Overall output SIR [dB]	14.65
CPU time [s]	1.227
Iterations	7.16
Nonblind approach	
Output SIR [dB]	15.72
ALUJA [27]	
Overall output SIR [dB]	11.06
CPU time [s]	1.057
Iterations	26.47
Nonblind approach	
Output SIR [dB]	14.97
PARAFAC [24]	
Overall output SIR [dB]	1.39
CPU time [s]	77.383
Iterations	1313.25
Nonblind approach	
Output SIR [dB]	3.28

の応答は約 0.52 s から始まる。これは、分離フィルタのインパルス応答に遅延 0.512 s を付加しているためである。表 4.5 に実環境における DMDI BSS の特性を示す。実環境の特性は、表 4.3 の残響時間 350 ms の仮想環境とほとんど同じである。

図 4.9 (a)、(b)、(c) にそれぞれ音源信号、観測信号、分離信号のスペクトログラムを示す。雑音と残響の抑圧、信号分離によって音声データにおいて弱い周波数成分が十分に抑制されて、音声データが観測信号から復元できていることが分かる。2.5 kHz より高い周波数ビンにおいては分離性能が十分ではないが、全体の分離性能に大影響は与えていない。これらの結果から、DMDI BSS は仮想環境と実環境の両方において十分な分離性能を発揮できることが示された。

図 4.10(a)、(b)、(c) にそれぞれ音源信号、DMDI BSS による分離信号、ALUJA BSS による分離信号のスペクトログラムを示す。但し、分離信号と音源信号は各行に揃えている。図 4.10(a) の音声信号波形と同図 (b) の分離信号波形を比較すると、無音区間の雑音と残響は DMDI BSS によって十分に低減されていることが分かる。分離信号の音質を平均 SDR [51] と PESQ スコア [41] によって評価した。SDR は観測信号と分離信号の差の電力と観測信号電力の比によって定義されている [51]。PESQ は通信路における音質の評価方法の 1 つである。表 4.6 に平均 SDR と PESQ スコアを示す。ここで、音源信号と観測信号間の遅延時間は約 0.006 s、音源信号と分離信号間の遅延時間は約 0.517 s である。これらの結果から、分離信号の音質は DMDI BSS が ALUJA BSS よりも良いことが分かる。DMDI BSS は信号分離だけでなく分離信号の音質からも有効である。一方、図 4.10(c) と表 4.6 に示したように、ALUJA BSS においては、分離信号の音質は十分に改善されていない。ALUJA アルゴリズムは行列 $\left\{ \mathbf{W}\hat{\mathbf{T}}(m)\mathbf{W}^H \right\}_{m=1}^M$ ができる限り対角行列になる

表 4.6: 実環境における DMDI BSS と ALUJA BSS の平均 SDR と PESQ スコアの比較 (SNR \approx 20 dB, 8192 ポイント FFT, $N = J = 4$)

Method	Improvement in SDR [dB]	PESQ score
DMDI	12.56	2.09
ALUJA [27]	0.77	1.70

ように分離行列 \mathbf{W} を求めるアルゴリズムであるが、残響や歪みを十分に除去できないことが示された。

4.4 まとめ

残響と雑音環境下において観測信号から信号源信号を復元する方法を提案した。提案手法は干渉抑制だけでなく、分離信号から残響の影響と歪みを低減させる効果がある。CDLS 評価量をラグランジェの未定乗数法を用いて最小化することにより混合行列を推定し、推定した混合行列がフルランクでない場合は、これをフルランク行列に置き換える。分離行列は、分離行列と混合行列の乗算が単位行列になるように設定し、ILS 評価量が最小になるように対角行列を推定する。これらの演算は、ALS アルゴリズムによって実行される。シミュレーションでは、2つの評価量を組み合わせることによって、残響と雑音が存在する環境下において提案手法が優れた収束特性と分離特性を発揮できることを示した。また、スペクトログラムによって実環境においても提案手法の有効性が実証された。

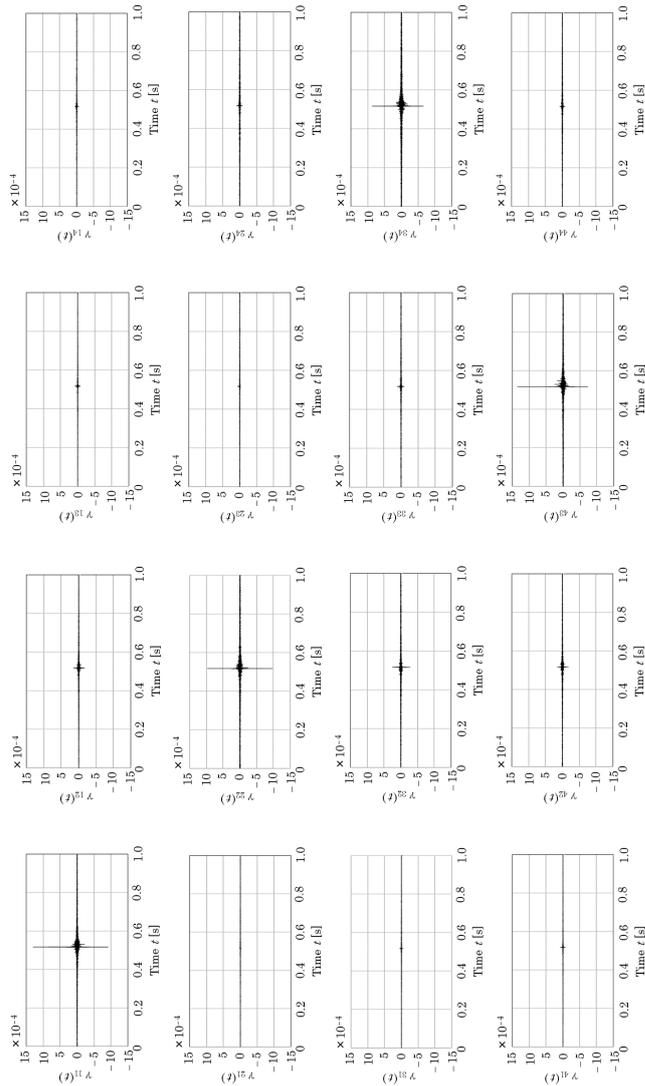
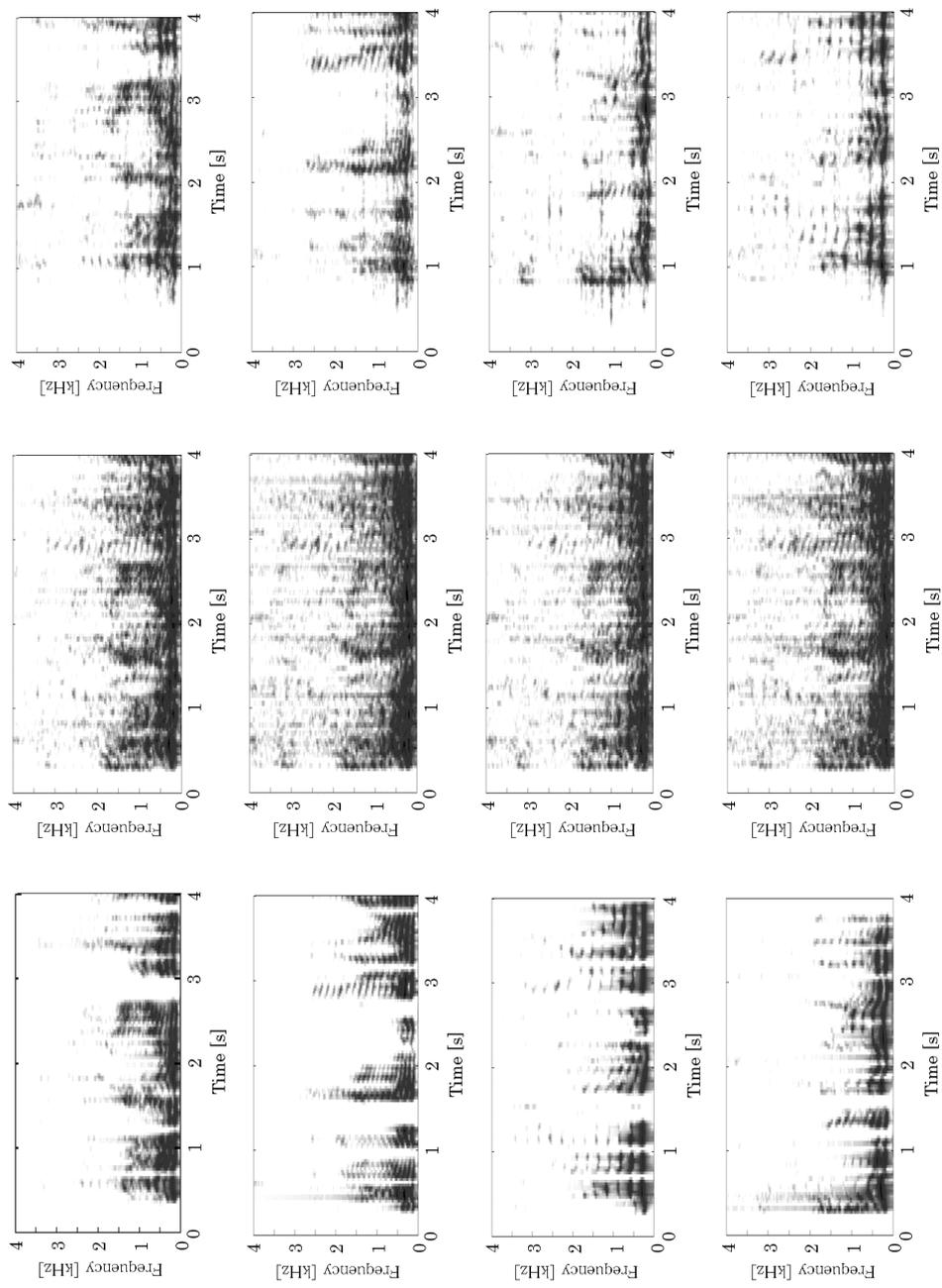


図 4.8: 実環境において DMDI BSS によって推定した i 番目のスピーカと j 番目のマイクロフォン間の混合分離連続接続フィルタのインパルス応答 $\gamma_{ij}(t)$ (SNR ≈ 20 dB, 8192 ポイント FFT, $N = J = 4$)

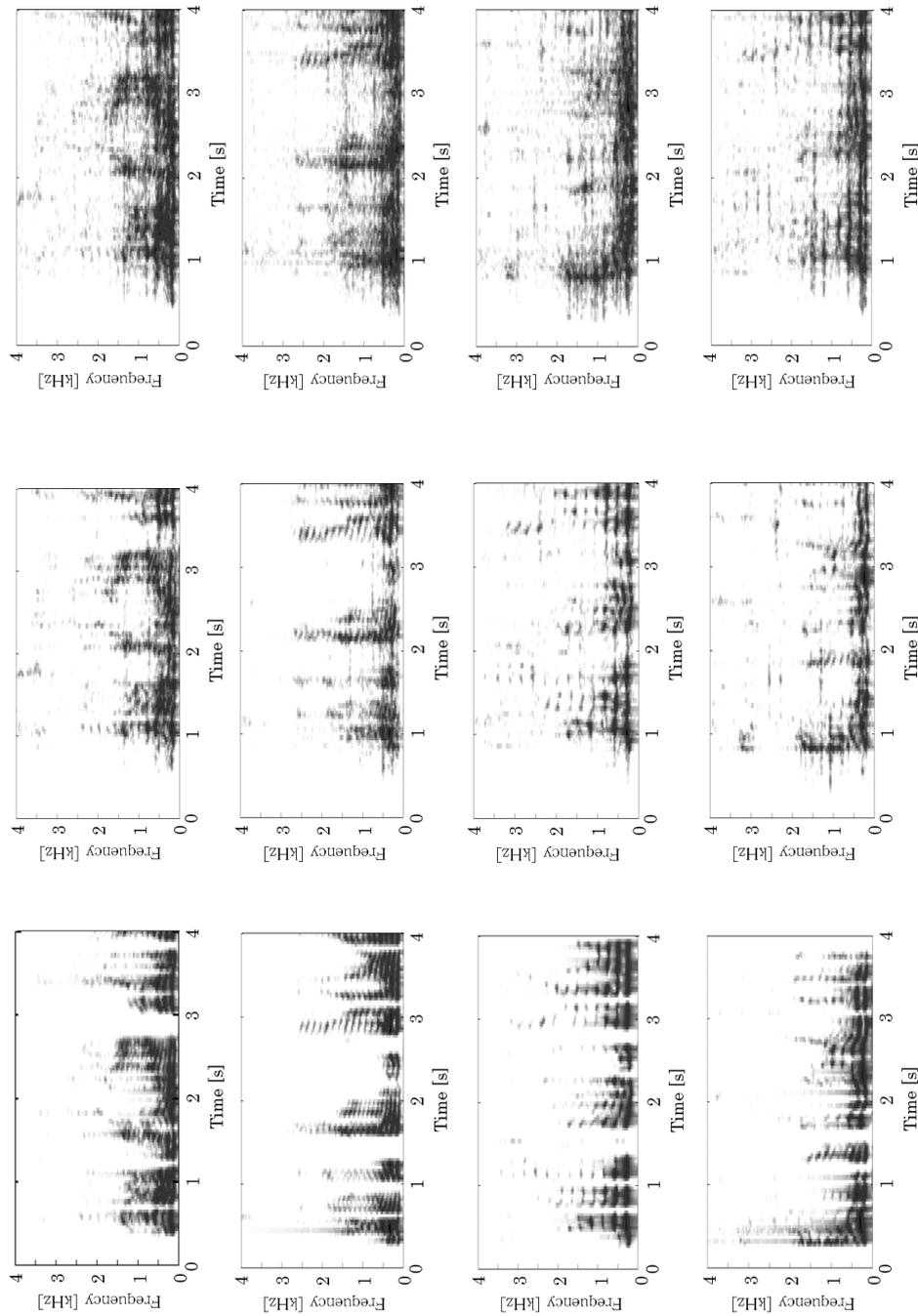


(a) The speech signals.

(b) The observed signals.

(c) The unmixing-filter output signals.

図 4.9: 音源信号、実環境における観測信号、DMDI BSS による分離信号のスペクトログラム (SNR ≈ 20 dB, 8192 ポイント FFT, $N = J = 4$)



(a) The speech signals. (b) The separated signals by the DMDI BSS. (c) The separated signals by the ALUJA BSS.

図 4.10: 音源信号、DMDI BSS と ALUJA BSS による分離信号のスペクトログラム (SNR ≈ 20 dB, 8192 ポイント FFT, $N = J = 4$)

第5章 DMDI アルゴリズムの高速化手法と ブラインド信号抽出への応用

5.1 まえがき

近年、対象行列の同時対角化 (joint diagonalization, JD) はブラインド同定、補聴器、音声強調、ブラインド信号分離等に応用されている。実環境の音声・音響への応用においては、雑音、並びにスピーカや増幅器等の非線形性の影響により純粋な対象行列は入手は不可能であり、しかも、入手可能な対象行列の個数は制限される。このような対象行列の同時対角化を近似同時対角化問題 (approximate JD, AJD) という。近似同時対角化問題の一例にブラインド信号抽出 (blind source extraction, BSE) がある。これは、雑音、残響が存在下で畳み込み混合過程が既知でない環境において、観測雑音のみから音源信号を復元するために使用される [7-9, 11]。

近似同時対角化問題は ALS (alternating least-squares) アルゴリズム [5, 9, 12, 13, 29, 30, 35, 36, 52] によって解法される。ALS アルゴリズムは最小 2 乗 (least-squares, LS) 評価量を最小化する一手法である [5, 13, 29, 30]。一般的に、ALS アルゴリズムは 2 つの手順からなる。手順 1 では、対角化行列を固定し、評価量を最小化することによって対角行列を推定する。手順 2 では、対角行列を固定し、評価量を最小化することによって対角化行列を推定する。これら 2 つの手順を交互に繰り返すことによって AJD 問題を解法する。ALS アルゴリズムの欠点は収束速度が遅いことである。少ない演算量で ALS アルゴリズムの収束速度を改善することが熱望されている。

近似同時対角化問題の解の精度は、同時対角化対象行列の個数に依存する。対象行列の個数が十分に多ければ、近似同時対角化問題の推定精度は改善されるが、演算量は増大する。対象行列の個数が少なければ演算的な負担は低減するが、近似同時対角化問題の推定精度は劣化する。このように、近似同時対角化問題の推定精度と演算量の間にはトレードオフの関係が存在する。音源とマイクロフォンの個数が等しい Determined の混合過程と音源の個数よりマイクロフォンの個数が多い Overdetermined の混合過程がブラインド信号抽出において存在する。Overdetermined の混合過程をブラインド信号抽出に導入すると分離性能が向上することが文献 [53, 54] に示されている。これは、入手可能な対象行列の個数が限られた応用において分離性能の改善に効果的であると考えられる。

4 章の DMDI アルゴリズムは ALSP アルゴリズムを基に作られている為、全体の ALS 反復の中に混合行列を推定する為のループが入れ子構造で存在している。この入れ子構造の演算量はマイクロフォンの個数の 2 乗で増加する為、Overdetermined の混合過程では、演算量が大幅に増加することが表 5.2 より明らかになっている。従って DMDI アルゴリズムで Overdetermined の混合過

程では、分離性能の改善と引き換えに膨大な演算量が必要となる。

第 5 章では、混合行列のために CDLS 評価量、対角行列のために ILS 評価量を交互に最小化する FDMDI (fast diagonalization, implemented by minimizing the direct and indirect LS criteria) アルゴリズムを提案する。DMDI アルゴリズムは混合行列推定の際、入れ子構造を用いて推定しているのに対し、FDMDI アルゴリズムは入れ子構造を用いることなく非反復法で混合行列を推定できることが特徴である。この手法により、信号抽出性能の劣化を伴わずにアルゴリズムの高速化ができる。更に、Overdetermined と FDMDI アルゴリズムの組み合わせによって、Determined よりも計算時間が短縮できることがシミュレーション結果によって明らかになった。

FDMDI アルゴリズムを用いて、Overdetermined の AJD 問題を解くことによって、DMDI アルゴリズムよりも少ない演算量で信号抽出性能の精度を高めることを目的とする。実環境において、音源数よりもマイクロフォンの個数を増やすことによって FDMDI アルゴリズムにより高い分離性能を実現できることを明らかにする。

5.2 DMDI アルゴリズムの特徴と問題点

DMDI アルゴリズムは \mathbf{G} の各列ベクトルに関する式 (4.3) が繰り返し求められる。このとき、最小化の対象とならない列ベクトルは固定される。更新前後の \mathbf{G} の差を誤差行列として記憶する。この繰り返しは、全体の ALS 反復ループの中で実行され、誤差行列のフロベニウスノルムが反復終了基準 ϵ_G を下回るまで続けられる。

Overdetermined の場合、マイクロフォンの個数の 2 乗で \mathbf{G} の各列ベクトルの個数は増大する。この増大によって、 \mathbf{G} の要素数は大幅に増える。この影響によって、誤差行列のフロベニウスノルムが反復終了基準 ϵ_G を下回るまでにより多く反復を要することによる計算時間増加や、あるいは反復終了基準 ϵ_G を下回らずに設定した反復上限値で打ち切られ、DMDI アルゴリズムが収束しないまま \mathbf{G} の解が算出されることによる分離性能低下の可能性がある。

5.3 FDMDI アルゴリズムのブラインド信号抽出への適用

\mathbf{G} の各列ベクトルに関する式 (4.3) を繰り返し求める入れ子構造を解消し、マイクロフォンの個数が多数の場合でも計算時間が短縮できる FDMDI アルゴリズム [14,52] を提案する。FDMDI アルゴリズムは、次の 2 の手順から成る。

手順 1) ILS 評価量を最小にして対角行列 $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ を推定する。このとき、混合行列 $\mathbf{B}(\omega_k)$ は前回の反復において手順 2 の反復で推定した値に固定する。

手順 2) CDLS 評価量を最小にして混合行列 $\mathbf{B}(\omega_k)$ を推定する。このとき、対角行列 $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ は手順 1 で推定した値に固定する。

CDLS 評価量は分離行列の関数ではないので、CDLS 評価量を最小化することによって分離行列を直接推定することはできない。FDMDI アルゴリズムでは、手順 2 で求めた混合行列から分離行列を求め、ILS 評価量を算出している。即ち、推定された混合行列が正則で、混合フィルタに分離フィルタを連続に接続した合成フィルタのインパルス応答が単位インパルス関数に近似できると仮定して、ILS 評価量を最小にすることによって対角行列 $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ を推定する。

5.3.1 観測信号のクロススペクトル密度行列の推定

ウォルシュペリオドグラム [46] を用いて観測信号を M 個のエポックに分割することによって、観測信号のクロススペクトル密度行列 $\{\hat{\mathbf{T}}(m, \omega_k)\}_{m=1}^M$ を推定することができる。推定されたクロススペクトル密度行列は

$$\tilde{\mathbf{T}}(m, \omega_k) = \frac{\hat{\mathbf{T}}(m, \omega_k)}{\max_{\omega_k} \|\hat{\mathbf{T}}(m, \omega_k)\|_F} \quad (5.1)$$

によって正規化する。

5.3.2 対角行列の推定

手順 1 では、 M 個の対角行列 $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ は ILS 問題

$$C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M) = \sum_{m=1}^M \|\mathbf{W}\tilde{\mathbf{T}}(m)\mathbf{W}^H - \mathbf{\Lambda}(m)\|_F^2 \quad (5.2)$$

を解くことによって推定する。ここで、周波数ビン ω_k を表す記号は全ての記号において共通であるので省略している。評価量 $C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M)$ を $\mathbf{\Lambda}(m)$ に関して最小化すると

$$\hat{\mathbf{\Lambda}}(m) = \text{diag}[\mathbf{W}\tilde{\mathbf{T}}(m)\mathbf{W}^H] \quad (5.3)$$

を得る。

5.3.3 混合行列の推定

手順 2 では、周波数ビン毎に M 個のエポックにおける CDLS 評価量

$$C_{\text{CDLS}}(\mathbf{B}) = \sum_{m=1}^M \|\tilde{\mathbf{T}}(m) - \mathbf{B}\mathbf{\Lambda}(m)\mathbf{B}^H\|_F^2 + 2 \sum_{j=1}^N \gamma_j (\mathbf{b}_j^H \mathbf{b}_j - 1) \quad (5.4)$$

を \mathbf{B} に関して最小化することによって混合行列 \mathbf{B} を求める [52]。ここで、 γ_j はスケーリング制約条件に関するラグランジェの未定乗数である。式 (5.4) に KR 積を適用すると、CDLS 評価量は

$$C_{\text{CDLS}}(\mathbf{G}) = \sum_{m=1}^M \|\tilde{\mathbf{t}}(m) - \mathbf{G}\mathbf{d}(m)\|_2^2 + 2 \sum_{j=1}^N \gamma_j (\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{g}_j) - 1) \quad (5.5)$$

と表現できる。ここで

$$\tilde{\mathbf{t}}(m) = \text{vec} \{ \tilde{\mathbf{T}}(m) \} \quad (5.6)$$

$$\mathbf{G} = [\mathbf{b}_1^* \otimes \mathbf{b}_1, \dots, \mathbf{b}_N^* \otimes \mathbf{b}_N] \quad (5.7)$$

$$\mathbf{d}(m) = [\lambda_1(m), \lambda_2(m), \dots, \lambda_N(m)]^T \quad (5.8)$$

$$\mathbf{G}\mathbf{d}(m) = \text{vec} \{ \mathbf{B}\boldsymbol{\Lambda}(m)\mathbf{B}^H \} = [\mathbf{B}^* \odot \mathbf{B}] \cdot \mathbf{d}(m) \quad (5.9)$$

である。行列 \mathbf{F} と \mathbf{R} を

$$\mathbf{F} = \sum_{m=1}^M \mathbf{d}(m)\mathbf{d}(m)^T \quad (5.10)$$

$$\mathbf{R} = \sum_{m=1}^M \tilde{\mathbf{t}}(m)\mathbf{d}(m)^T \quad (5.11)$$

と定義すると、ラグランジェの未定乗数法を用いて $C_{\text{CDLS}}(\mathbf{G})$ を最小化する解

$$\hat{\mathbf{G}} = \mathcal{R}e(\hat{\mathbf{G}}) + j\mathcal{I}m(\hat{\mathbf{G}}) \quad (5.12)$$

を得る。ここで

$$\mathcal{R}e(\hat{\mathbf{G}}) = \mathcal{R}e(\mathbf{R})\mathbf{F}^{-1} - \frac{1}{J}\text{vec}\{\mathbf{I}\} \left[\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{R})\tilde{\mathbf{f}}_1 - 1, \dots, \text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{R})\tilde{\mathbf{f}}_N - 1 \right] \quad (5.13)$$

$$\mathcal{I}m(\hat{\mathbf{G}}) = \mathcal{I}m(\mathbf{R})\mathbf{F}^{-1} \quad (5.14)$$

である。また、 $\tilde{\mathbf{f}}_k$ は行列 \mathbf{F}^{-1} の列ベクトルを表す。

式 (5.5) を最小化して $\hat{\mathbf{g}}_i$ を求めた後、CDLS 推定問題は $\hat{\mathbf{g}}_i$ を部分空間 $\Omega \subset \mathbb{C}^{J^2 \times 1}$ 上に写像することによって解法される。ここで、 Ω は

$$\Omega = \{ \text{vec}\{\boldsymbol{\Phi}\} | \boldsymbol{\Phi} = \mathbf{v}\mathbf{v}^H, \mathbf{v} \in \mathbb{C}^{J \times 1}, \|\mathbf{v}\|_2^2 = 1 \} \quad (5.15)$$

によって定義される [13, 34]。CDLS 推定問題は、累乘法 [44] を適用することによって、ALS 反復毎に 1 度 $\|\text{vec}^{-1}\{\hat{\mathbf{g}}_i\} - \mathbf{b}_i\mathbf{b}_i^H\|_F^2$ を最小化する \mathbf{b}_i を求めることによって解法することができる [13, 34]。

5.3.4 正則な分離行列の構築

同時対角化問題では、推定された混合行列は正則でなければならない [10, 49]。推定された混合行列が非正則であるである場合、これを正則な行列に置き換えた後、分離行列を算出する。行列 \mathbf{B} の特異値分解を

$$\mathbf{B} = \mathbf{V}_r \boldsymbol{\Sigma}_r \mathbf{U}_r^H \quad (5.16)$$

によって表現する。ここで、 \mathbf{V}_r と \mathbf{U}_r はユニタリ行列

$$\mathbf{V}_r = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r] \quad (5.17)$$

$$\mathbf{U}_r = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_r] \quad (5.18)$$

Σ_r は対角行列

$$\Sigma_r = \text{diagonal}(\sigma_1, \sigma_2, \dots, \sigma_r) \quad (5.19)$$

$$\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0 \quad (5.20)$$

をそれぞれ表す。行列 \mathbf{B} のランクが $r < N$ の場合、正規直交基底 $\{\mathbf{v}_i\}_{i=1}^r$ に直交する正規直交基底 $\{\mathbf{v}_i\}_{i=r+1}^N$ と、正規直交基底 $\{\mathbf{u}_i\}_{i=1}^r$ に直交する正規直交基底 $\{\mathbf{u}_i\}_{i=r+1}^N$ を使用して、行列 \mathbf{B} を正則行列

$$\mathbf{v}\Sigma\mathbf{u}^H = \frac{\sqrt{N}}{\sqrt{N} + \delta N} \cdot [\mathbf{V}_r, \bar{\mathbf{V}}_r] \left(\begin{bmatrix} \Sigma_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \delta \mathbf{I} \right) [\mathbf{U}_r, \bar{\mathbf{U}}_r]^H \quad (5.21)$$

に置き換える。ここで、 $\delta > 0$ とし、式 (5.21) の右辺の $\sqrt{N}/(\sqrt{N} + \delta N)$ は

$$\text{tr}[\Sigma] = \sqrt{N} \quad (5.22)$$

を満たさなければならない。また

$$\bar{\mathbf{V}}_r = [\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_N] \quad (5.23)$$

$$\bar{\mathbf{U}}_r = [\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_N] \quad (5.24)$$

であり、 $\bar{\mathbf{V}}$ と添字 r は、 $\{\mathbf{v}_i\}_{i=r+1}^N$ が $\{\mathbf{v}_i\}_{i=1}^r$ の直交補空間であることを示している。 $\{\mathbf{v}_i\}_{i=1}^r$ と $\{\mathbf{u}_i\}_{i=1}^r$ はそれぞれ $\{\mathbf{v}_i\}_{i=r+1}^N$ と $\{\mathbf{u}_i\}_{i=r+1}^N$ に直交関係にあること、また、式 (5.22) の条件が満たされていることから、 $\mathbf{v}\Sigma\mathbf{u}^H$ によって \mathbf{B} を置き換えても式 (5.5) による最小化には影響を与えない。

式 (2.9) より、分離行列 \mathbf{W} は

$$\xi(\mathbf{W}) = \|\mathbf{I} - \mathbf{W}\mathbf{B}\|_F^2 \quad (5.25)$$

を最小にすることによって推定される。式 (5.25) の最小化は、混合フィルタと分離フィルタを縦続に接続した合成フィルタの応答を求めることと等価であり [14]、分離行列は

$$\mathbf{W} = (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \quad (5.26)$$

によって推定される。式 (5.9) によって \mathbf{B} と $\Lambda(m)$ から $\mathbf{Gd}(m)$ を算出した後、評価基準 ϵ_C を下回るまで手順 1 と手順 2 は繰り返される。FDMDI アルゴリズムを Algorithm 3 に要約する。

Algorithm 3 : FDMDI アルゴリズム

Estimation of the cross-spectral density matrices of the observed signal:

- 1: Compute
- $\hat{\mathbf{T}}(m)$
- at each epoch using the Welch periodogram method for
- $m = 1, 2, \dots, M$
- .

Normalization of $\hat{\mathbf{T}}(m)$:

- 2: Normalize
- $\hat{\mathbf{T}}(m)$
- for
- $m = 1, 2, \dots, M$
- according to (5.1).
-
- 3: Convert
- $\hat{\mathbf{T}}(m)$
- into
- $\tilde{\mathbf{t}}(m)$
- according to (5.6).

Initialization:

- 4: Set the initial values of
- $\mathbf{W}(\omega_k)$
- to the indirect LS estimate at frequency bin
- ω_{k-1}
- with
- $\mathbf{W}(\omega_0) = \mathbf{I}$
- .
-
- 5:
- while**
- The absolute value of the difference between the current and previous values of
- $C_{\text{CDLS}}(\mathbf{G})$
- in the ALS iterations is larger than
- ϵ_C
- .
- do**

First phase (Estimation of the diagonal matrices):

- 6: Minimize
- $C_{\text{ILS}}(\{\Lambda(m)\}_{m=1}^M)$
- to find
- $\hat{\Lambda}(m)$
- by the method of least squares according to (5.3).

Second phase (Estimation of the mixing matrix):

- 7: Compute
- \mathbf{F}
- and
- \mathbf{R}
- according to (5.10) and (5.11), respectively.
-
- 8: Compute the inverse of
- \mathbf{F}
- .
-
- 9: Minimize
- $C_{\text{CDLS}}(\mathbf{G})$
- to find
- $\hat{\mathbf{G}}$
- by the method of Lagrange multipliers according to (5.12).
-
- 10: Minimize
- $\|\text{vec}^{-1}\{\hat{\mathbf{g}}_i\} - \mathbf{b}_i \mathbf{b}_i^H\|_F^2$
- to find
- \mathbf{b}_i
- by the power method for
- $i = 1, 2, \dots, N$
- .

Construction of the full-rank unmixing matrix:

- 11:
- if**
- $\text{rank}(\mathbf{B}^H \mathbf{B}) = r < N$
- then**
-
- 12: Construct the orthonormal sets
- $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_N$
- and
- $\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_N$
- by the Gram-Schmidt process.
-
- 13: Replace
- \mathbf{B}
- with
- $\mathbf{V} \Sigma \mathbf{U}^H$
- as given by (5.21).
-
- 14:
- end if**
-
- 15: Compute
- \mathbf{W}
- from
- \mathbf{B}
- according to (5.26).
-
- 16:
- end while**
-

5.3.5 信号源信号の復元

FDMDI アルゴリズムを用いたブラインド信号抽出を以下の 5 段階で実現する。

- a) 分離行列 \mathbf{W} を、Algorithm 3 によって周波数ビン毎に独立に推定する。
- b) スケーリング問題は $\mathbf{\Pi}^{-1} \mathbf{W}$ によって解かれる。
- c) パーミュテーション行列 $\mathbf{P}^{-1} \mathbf{\Pi}^{-1} \mathbf{W}$ を算出して全周波数ビンで信号源への割り当てを揃える。
- d) $\mathbf{P}^{-1} \mathbf{\Pi}^{-1} \mathbf{W}$ に $e^{-j\pi k}$ を乗算する。
- e) 逆離散フーリエ変換によって時間領域に $e^{-j\pi k} \mathbf{P}^{-1} \mathbf{\Pi}^{-1} \mathbf{W}$ を変換して分離フィルタのインパルス応答を算出する。

分離フィルタは信号源分離だけでなく逆フィルタの役割も有する。混合フィルタが安定で、因果性を満たし、最小位相であるとき、分離フィルタのインパルス応答を遅延させることによって加法性雑音の影響を低減させることができる [55]。このため、上記 d) のよって分離フィルタのインパルス応答に遅延要素を付加している。信号源信号は

$$\hat{s}_i(t) = \sum_{j=1}^J w_{ij}(t) * x_j(t) \quad (5.27)$$

表 5.1: ALS 反復当たりの演算量の比較

Method	Multiplications	
	First phase	Second phase
DMDI	$(J^3 + J^2)M$	$(M + 5)J^3 + (M + 2N + 3)J^2$ $+ \{J^2MN + (4J^2 + 2)NM + 2J^2\} p_f^a$ $+ (2J^2 + 4J)Np_{r1}^b$ $+ 2(2J^2 + 4J)(N - r)p_{r2}^c$ $(Np_f + Np_{r1} + 2(N - r)p_{r2} + 1 \text{ divisions})$
FDMDI	$(J^3 + J^2)M$	$(N^2 + 2J^2N)M$ $+ 5J^3 + 3J^2 + N^3/2 - N^2/2 + N$ $+ 3(JN)^2 + (2J^2 + 4J)Np_{r1}$ $+ 2(2J^2 + 4J)(N - r)p_{r2}$ $(Np_{r1} + 2(N - r)p_{r2} + 1 \text{ divisions})$
PARAFAC [24]	$\{NJ(1 + 7N)(J + 2M) + 3J^2MN + 11N^3$ $+ 2N^2(2J + M) + (8N + 10)J^2M\}/3$	$2\{NJ(1 + 7N)(J + 2M) + 3J^2MN + 11N^3$ $+ 2N^2(2J + M) + (8N + 10)J^2M\}/3$

^a p_f denotes the number of iterations for estimating \mathbf{G} in DMDI.

^b p_{r1} denotes the number of iterations for the power method to find \mathbf{b}_i .

^c p_{r2} denotes the number of iterations for the power method to construct the full-rank unmixing matrix.

によって復元される。ここで、 $w_{ij}(t)$ は遅延分離行列 $e^{-j\pi k} \mathbf{P}^{-1} \mathbf{\Pi}^{-1} \mathbf{W}$ を逆離散フーリエ変換したものである。

5.3.6 演算量の比較

ALSP アルゴリズム [13] においては、混合行列を一度に推定するのではなく、混合行列の列ベクトルに分けて、列ベクトル毎に推定している。このため、ALS 反復内に混合行列の列ベクトルを推定する反復が存在する。一方、FDMDI アルゴリズムでは、行列 \mathbf{F} の逆行列を用いて混合行列を直接推定している。このとき、完全ピボット選択付きガウスの消去法 [44] を用いている。表 5.1 で FDMDI アルゴリズムと PARAFAC with ELS [24] の演算量を比較する。表中の p_{r1} と p_{r2} は、 \mathbf{b}_i を推定するために用いた累乗法の反復回数と正則な分離行列の構築するために用いた累乗法の反復回数をそれぞれ表す。FDMDI アルゴリズムと DMDI アルゴリズムの演算量算出については付録を参照されたい。また、PARAFAC の ALS 反復当たりの乗算回数の詳細については文献 [24] を参照されたい。表 5.1 より FDMDI は DMDI に比べ演算量を削減できること、FDMDI と PARAFAC の演算量に大差がないことが分かる。

表 5.2 に DMDI、FDMDI、PARAFAC の演算量を比較する。 $p_{r1} = 1851.95$ 、 $p_{r2} = 0.0016$ 、 $p_f = 13.12$ については、残響時間 300 ms において DMDI と FDMDI アルゴリズムを実行した平均反復回数を採用している。累乗法の実平均反復回数を考慮すると、FDMDI の演算量は PARAFAC よりも少ないことが分かる。

表 5.2: ALS 反復当たりの DMDI、FDMDI、PARAFAC の演算量の比較 ($p_{r1} = 1851.95$, $p_{r2} = 0.0016$, $p_f = 13.12$ については、残響時間 300 ms において DMDI と FDMDI アルゴリズムを実行した平均結果に基づき算出している。)

N	J	M	DMDI	FDMDI	PARAFAC	
3	3	200	544355.08	186892.36	161199.0	
		400	921624.59	206692.36	321399.0	
		600	1298894.09	226492.36	481599.0	
		800	1676163.60	246292.36	641799.0	
		1000	2053433.10	266092.36	801999.0	
	5	400	2450145.42	513896.18	703327.0	
		800	4510095.51	637496.18	1404527.0	
	7	400	4749315.30	981233.53	1223383.0	
		800	8795782.00	1259233.53	2442983.0	
	4	4	400	2110123.91	446338.16	732416.0
			800	3863865.72	535938.16	1462016.0
		5	400	3246481.92	666873.66	1020724.0
800			5973082.04	813273.66	2037524.0	
6		400	4635406.12	936492.33	1352464.0	
		800	8557633.06	1158892.33	2699664.0	
5	5	1200	10829318.71	1160812.13	4176375.0	
	7		21088469.41	2260719.83	6916895.0	
	9		34857329.69	3817447.44	10282855.0	

5.4 シミュレーション

FDMDI、DMDI アルゴリズムと従来の代表的なアルゴリズム ALSP [13], PARAFAC with ELS [24], ALUJA [27], and GNJD [56] の分離性能と演算量をシミュレーションによって比較する。シミュレーションプログラムは C 言語を用いて実装した。

5.4.1 DMDI アルゴリズムと FDMDI アルゴリズムの基本性能を調べる事前シミュレーション

対象行列の集合 $\{\hat{\mathbf{T}}(m)\}_{m=1}^M$ は

$$\hat{\mathbf{T}}(m) = \mathbf{H}\mathbf{P}_s(m)\mathbf{H}^H + \beta(m)\mathbf{Q}(m) \quad (5.28)$$

によって生成される。 \mathbf{H} 、 $\mathbf{Q}(m)$ の要素は複素数で、独立、そして平均 0、標準偏差 1 のガウス分布である。 $\mathbf{P}_s(m)$ の要素は実数で、独立、そして平均 0、標準偏差 1 のガウス分布である。 $\beta(m)$ は SNR を決めることによって与えられる。SNR は

$$\text{SNR} = 10 \log_{10} \frac{\|\mathbf{H}\mathbf{P}_s(m)\mathbf{H}^H\|_F^2}{\|\beta(m)\mathbf{Q}(m)\|_F^2} \quad (5.29)$$

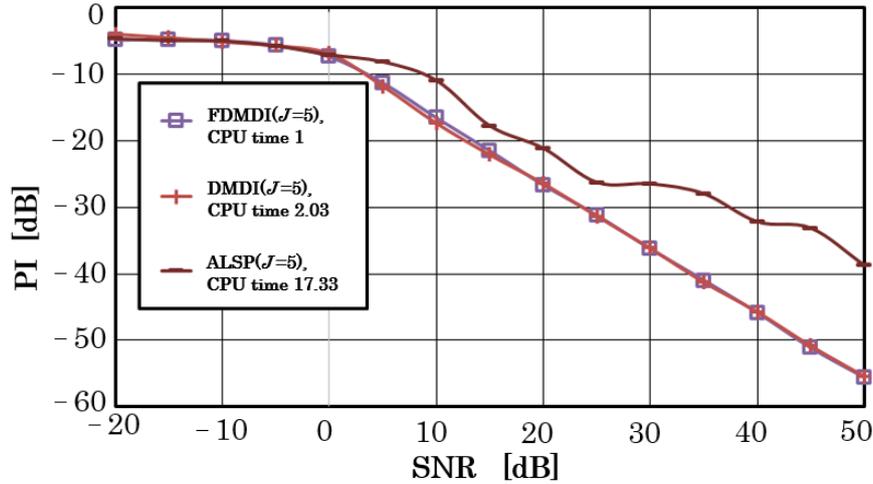


図 5.1: $\text{cond}(\mathbf{H}) \leq 5, N = 5, J = 5$ における Determined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較

で与えられる。試行回数は 300 回で、閾値 ϵ_G と ϵ_C は 10^{-6} とした。分離性能は performance index (PI) を用いた。PI は

$$\text{PI}(\mathbf{\Gamma}) = \frac{1}{2N(N-1)} \sum_{i=1}^N \left(\sum_{j=1}^N \frac{|\gamma_{ij}|^2}{\max_m |\gamma_{im}|^2} - 1 \right) + \frac{1}{2N(N-1)} \sum_{j=1}^N \left(\sum_{i=1}^N \frac{|\gamma_{ij}|^2}{\max_m |\gamma_{mj}|^2} - 1 \right) \quad (5.30)$$

で定義される。 $\mathbf{\Gamma} = \mathbf{WH}$ で与えられ、 γ_{ij} は行列 $\mathbf{\Gamma}$ の i 行 j 列の要素を表す。図 5.1、5.2、5.3、5.4 に DMDI、FDMDI、ALSP アルゴリズムの比較結果を示す。シミュレーションでは、DMDI と FDMDI アルゴリズムはほぼ同等の結果を示している。計算時間は入れ子構造がない FDMDI の方が計算時間が短い結果となった。

また、Determined の結果の図 5.1、5.3 と、Overdetermined の結果の図 5.2、5.4 を比較すると、Determined よりも Overdetermined の方が FDMDI と DMDI との計算時間の差が大きくなっている。

更に図 5.5 に J の値を 5 から 8 まで変化させた時にの FDMDI の分離性能を示す。この結果より $J = 5$ の Determined よりも $J = 6, J = 7$ の方が計算時間が短いことが判明した。従って FDMDI アルゴリズムは Overdetermined にすることによって、計算時間を短縮できる特性をもつことが分かった。

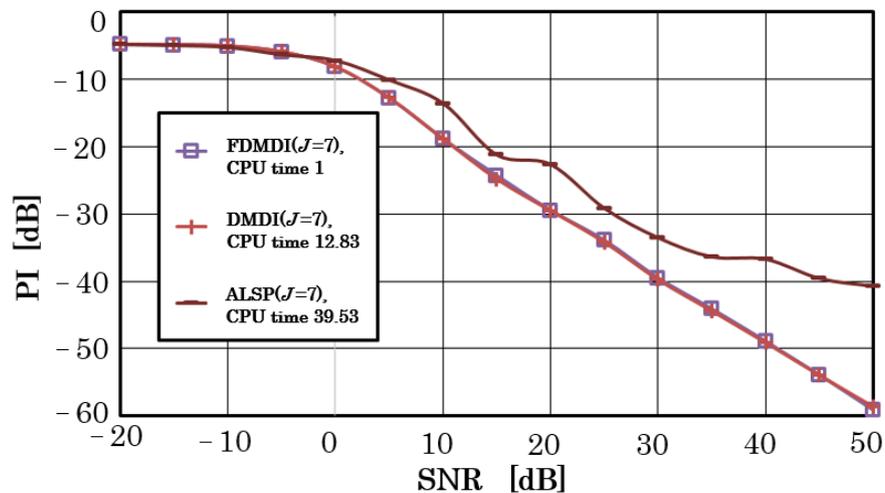


図 5.2: $\text{cond}(\mathbf{H}) \leq 5, N = 5, J = 7$ における Overdetermined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較

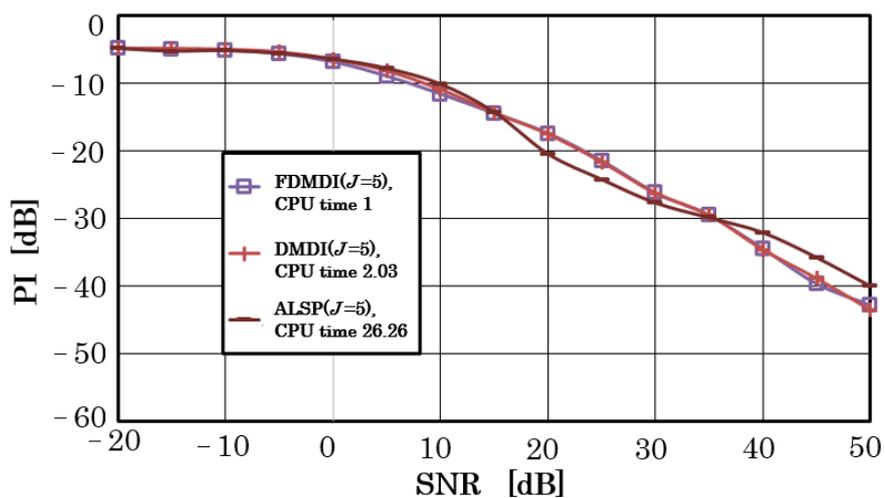


図 5.3: $\text{cond}(\mathbf{H}) \text{ free}, N = 5, J = 5$ における Determined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較

5.4.2 音声信号を用いた畳み込み混合におけるシミュレーション条件

音声コーパス [45] から入手した男女の音声データを繋ぎ合わせて音源として使用した。点音源から理想点マイクロフォンまでの室内空間インパルス応答は、寸法 $4.45 \times 3.55 \times 2.50 \text{ m}$ の部屋を想定して標本化周波数 8 kHz で image method [47, 48] を用いて発生させた。残響時間は $100, 300, 500, 700, 900 \text{ ms}$ に設定した。 $N = 3$ のときは、3 つの音源を図 5.6 の点 α, β, γ の位置に設置した。図中の三角形は、その重心が半径 1.20 m の球の中心 $(2.23, 1.78, 1.20 \text{ m})$ に一致するように設置した。表 5.3 に各ラベルの音源とマイクロフォンの座標を示す。マイクロホンの指

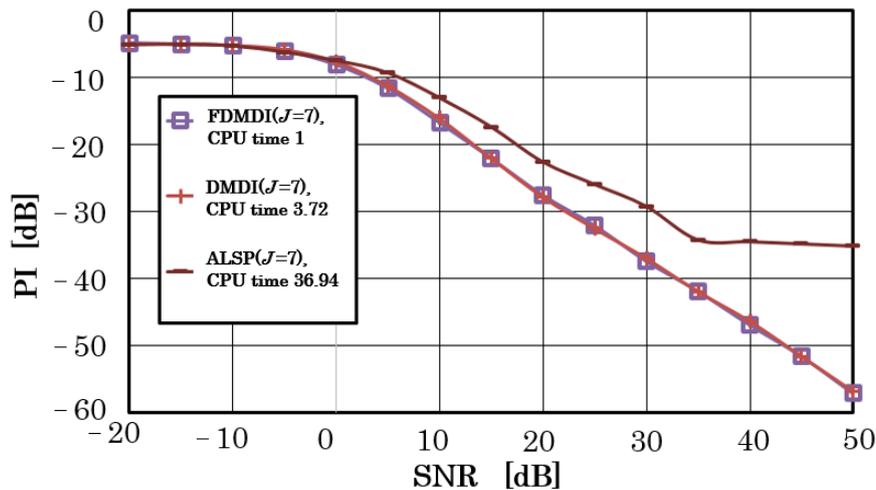


図 5.4: $\text{cond}(\mathbf{H})_{\text{free}}, N = 5, J = 7$ における Overdetermined の DMDI、FDMDI、ALSP アルゴリズムの分離性能と演算量の比較

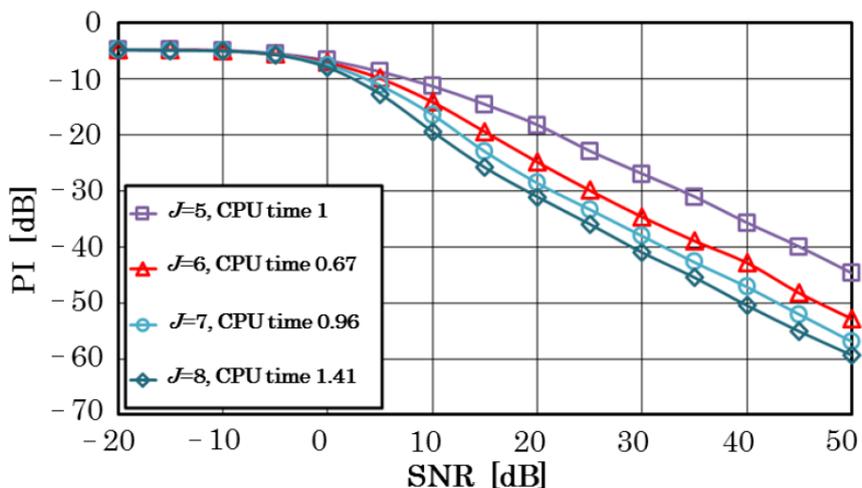


図 5.5: $\text{cond}(\mathbf{H})_{\text{free}}, N = 5, J = 5$ から 8 における FDMDI アルゴリズムの分離性能と演算量の比較

向性の影響を無くすために、7つのマイクロフォンは図 5.6 の点 A, B, C, D, E, F, G の位置に配置した。J=3 のときは、図 5.6 の点 A, C, D に設置したマイクロフォンで信号を観測した。J=5 のときは、図 5.6 の点 A, B, C, D, E に設置したマイクロフォンで信号を観測した。N=4 のときは、音源は文献 [14] に示された配置と同じとした。J=4 のときは、マイクロフォンは文献 [14] と同じ位置に配置した。J=5 のときは、J=4 の配置に1つのマイクロフォンを座標 (2.23, 1.78, 1.32 m) に追加した。J=6 のときは、J=5 の配置に1つのマイクロフォンを座標 (2.23, 1.78, 1.08 m) に追加した。空間インパルス応答と音声データを畳み込みした後、平均零、SN

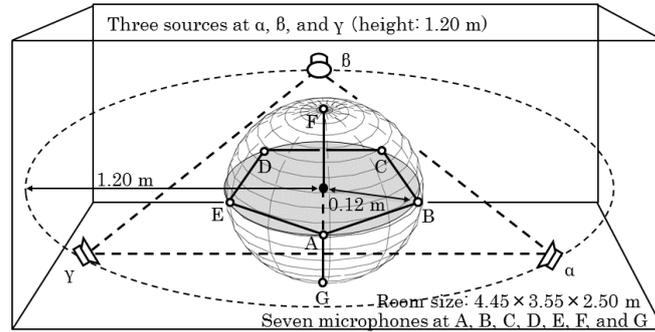

 図 5.6: 小部屋における音源とマイクロフォンの配置図 ($N = 3$)

表 5.3: 図 5.6 における音源とマイクロフォンの座標

Room size: 4.45 × 3.55 × 2.50 m			
		Label in Fig. 5.6	Coordinate [m]
Sources	$s_1(t)$	α	(3.26, 1.18, 1.20)
	$s_2(t)$	β	(2.23, 2.98, 1.20)
	$s_3(t)$	γ	(1.19, 1.18, 1.20)
Microphones	$x_1(t)$	A	(2.23, 1.66, 1.20)
	$x_2(t)$	B	(2.34, 1.74, 1.20)
	$x_3(t)$	C	(2.30, 1.87, 1.20)
	$x_4(t)$	D	(2.15, 1.87, 1.20)
	$x_5(t)$	E	(2.11, 1.74, 1.20)
	$x_6(t)$	F	(2.23, 1.78, 1.32)
	$x_7(t)$	G	(2.23, 1.78, 1.08)

比 20 dB のガウス性白色雑音を加法して観測信号として使用した。音声データ長は 1500s で、各マイクロフォンで集音したデータの振幅が等しくなるように正規化した。STFT にはハニング窓を使用し、重複率 80% のフレームで 8,192 ポイントの FFT を用いた [14]。観測信号 $\mathbf{x}(m, \omega_k)$ のクロススペクトル密度行列は、エポック当たり重複率 80% の 2 つのフレームを用いたウォルシュペリオドグラム法 [46] によって近似した。エポック区間は 10,000 データサンプルに固定した [13]。FDMDI では、推定した観測信号 $\mathbf{x}(m, \omega_k)$ のクロススペクトル密度行列はエポック毎に、また、ALUJA と GNJD を除く他の BSE では各ポイント毎に振幅で正規化した。

PARAFAC では、 $\mathbf{B}(\omega_0)$ と $\{\mathbf{\Lambda}(m, \omega_0)\}_{m=1}^M$ は白色雑音で初期化した [24]。その他の ALS アルゴリズムでは、単位行列で $\mathbf{B}(\omega_0)$ と $\{\mathbf{\Lambda}(m, \omega_0)\}_{m=1}^M$ を初期化した。ALUJA と GNJD を除く全 ALS アルゴリズムでは、 $\mathbf{B}(\omega_1)$ と $\{\mathbf{\Lambda}(m, \omega_1)\}_{m=1}^M$ は、周波数ビン ω_0 における推定値に初期化した。この初期化手順は、周波数ビン $\omega_{K/2}$ における推定の初期化まで続けた [13]。ALUJA と GNJD では、 $\mathbf{B}(\omega_k)$ と $\{\mathbf{\Lambda}(m, \omega_k)\}_{m=1}^M$ は全周波数ビンにおいて単位行列に初期化した。GNJD では、行列 R の数は 4 に設定した [56]。FDMDI では、ALS 反復終了基準を $\epsilon_C = 10^{-6}$ に設定した。また、正則な分離行列の構築のために $\delta = \sigma_r$ を用いた。PARAFAC の最大反復回数は 2000

に選んだ。プログラムは C 言語で作成し、インテル製の Core i7-2600 3.4 GHz で倍精度浮動小数点で演算した。全ブラインド信号抽出法で、分離行列推定後に $e^{-j\pi k}$ を乗算し、加法雑音の影響を低減させた。ALUJA と GNJD を除き、スケールリング問題は各周波数ビンにおいて \mathbf{W} の列ベクトルを正規化することによって解法した。また、パーミュテーション行列は近傍周波数ビン間の電力比の相関係数を用いて求めた [38]。

5.4.3 Determined の場合

ALS アルゴリズムの演算量を評価するため、ALS アルゴリズムの両手順の演算時間を計測した。測定条件は残響時間を 300 ms、 $N = J = 3$ に設定した。図 5.7 に FDMDI アルゴリズムにおいて手順 1 の演算時間を示す。完全ピボット選択付きガウスの消去法を用いて逆行列 \mathbf{F}^{-1} を直接計算することによって混合行列を推定している。即ち、Algorithm 3 のステップ 8 に J^2 次元のベクトルのみが含まれているので、図 5.7 に示すように FDMDI アルゴリズムの手順 1 の演算量は総エポック数 M に依存しない。一方、FDMDI アルゴリズムの Step 7 における \mathbf{F} と \mathbf{R} の計算は M に依存して増大する。PARAFAC アルゴリズムにおける ELS の計算は手順 2 に含まれている。FDMDI アルゴリズムでは混合行列の推定には反復法を使用していないので、PAFARAC アルゴリズムの演算量が最も多い結果となった。一方、FDMDI アルゴリズムの演算量は最も少ない結果となった。表 5.4 と 5.5 に示したように、混合行列の推定に反復法を利用した ALSP アルゴリズムよりも演算量が少なく、全演算量は FDMDI アルゴリズムが最も少なかった。これは、混合行列の推定に反復法を使用せず、逆行列を用いて推定値を得ていることに依る。

図 5.7 でエポック数と平均出力 SIR、演算時間の比較する。FDMDI アルゴリズムを用いたブラインド信号抽出が少ない演算量で最も高い分離性能を示した。表 5.4 と表 5.5 では音源とマイクロフォンの個数が等しい $N = 3$ と $N = 4$ の畳み込み混合における分離性能を従来法と比較する。PESQ による音質評価では、ALUJA と GNJD と比較して、それぞれ 0.94 と 0.82、提案方法による音質が上回った。

5.4.4 Overdetermined の場合

図 5.8 に音源とマイクロフォンの個数が等しい畳み込み混合、また、マイクロフォン個数が音源数より多い畳み込み混合における FDMDI 並びに PARAFAC を用いたブラインド信号抽出の平均 PESQ、周波数ビン当たりの平均反復回数、演算時間、出力 SIR を示す。残響時間が 300 ms より短い環境では、マイクロフォンの個数を 5 本より多くしても FDMDI アルゴリズムでは、ALS 反復回数はほとんど増大せず分離性能も向上が見られないが、演算時間は増大する。図 5.7 と 5.8 から明らかなように、7 個のマイクロフォンを使用して $M = 400$ の場合、FDMDI の分離性能は、5 個のマイクロフォンを使用し、 $M = 800$ の場合よりも低くなった。エポック長 $M = 400$ は音声データ長 500 s、 $M = 800$ は音声データ長 1000 s に相当する。更に、残響時間 300 ms、 $M = 800$ 、

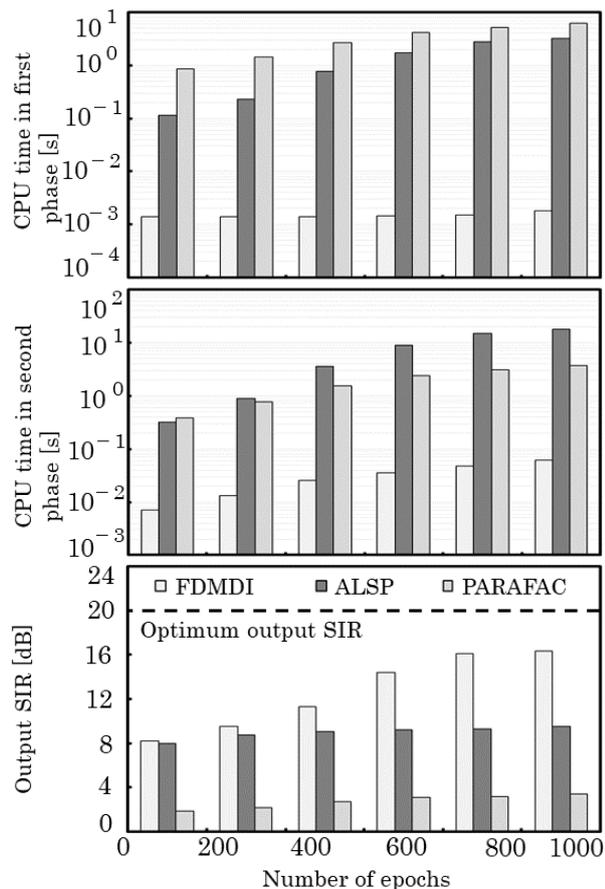


図 5.7: SNR ≈ 20 dB、平均入力 SIR -3.16 dB、 $N = J = 3$ 、残響時間 300-ms におけるエポック数と平均出力 SIR、演算時間の比較

5 個のマイクロフォンを使用した場合、FDMDI アルゴリズムの演算時間は、7 個のマイクロフォンを使用した場合の約 60% であった。

図 5.9 に音源とマイクロフォンの個数が等しい畳み込み混合、また、マイクロフォン個数が音源数より多い畳み込み混合におけるブラインド信号抽出の性能を示す。ALUJA と GNJD については J から N への次元削減法を使用しない場合、マイクロフォン個数が音源数より多い畳み込み混合の分離問題に適用できない [57]。FDMDI アルゴリズムを用いたブラインド信号抽出法が最も高い分離性能を得た。また、同抽出法では、残響時間が長くエポック数が十分に長い場合、マイクロフォンの個数を音源数より多くすると、分離性能が向上することが確認できた。また、図 5.9 から明らかのように、マイクロフォンの個数を音源数より多くすると、FDMDI アルゴリズムを用いたブラインド信号抽出により復元した音声の PESQ 値が最も高くなった。

5.5 実環境における分離性能

室内空間インパルス応答は線形モデルを仮定して、小部屋を想定して点音源から理想点マイクロフォンまでの室内空間インパルス応答を image method [47, 48] と呼ばれるシミュレータを用いて

表 5.4: SNR \approx 20 dB、 $M = 400$ 、 $N = 3$ 、音源とマイクロフォンの個数が等しい畳み込み混合におけるブラインド信号抽出の平均 PESQ、周波数ビン当たりの平均反復回数、演算時間、出力 SIR

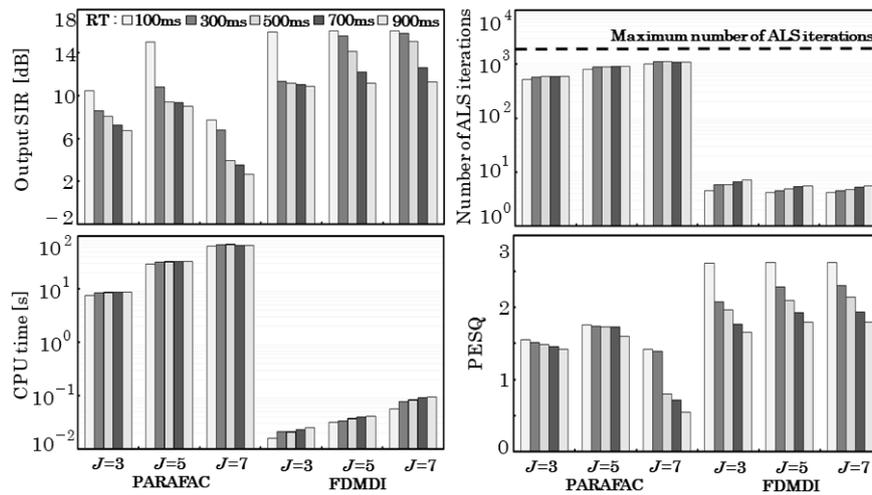
Method	Reverberation time (RT) [ms]				
	100	300	500	700	900
	Overall input SIR [dB]				
	-2.59	-3.39	-3.63	-3.66	-3.62
FDMDI					
Output SIR [dB]	15.94	11.33	11.16	11.03	10.88
CPU time [s]	0.016	0.021	0.021	0.023	0.025
ALS iterations	4.52	5.82	5.82	6.67	7.18
PESQ score	2.61	2.08	1.96	1.77	1.65
GNJD [56]					
Output SIR [dB]	11.18	11.08	9.16	8.81	7.88
CPU time [s]	8.658	12.167	15.935	20.260	22.900
Iterations	171.0	227.0	305.0	343.0	395.0
PESQ score	1.76	1.26	0.92	0.89	0.84
ALUJA [27]					
Output SIR [dB]	12.41	11.10	6.38	5.54	5.38
CPU time [s]	0.206	0.224	0.227	0.228	0.237
Iterations	37.62	40.91	41.50	41.57	43.36
PESQ score	1.67	1.14	1.04	0.87	0.87
PARAFAC [24]					
Output SIR [dB]	10.44	8.60	8.08	7.28	6.74
CPU time [s]	7.631	8.522	8.629	8.661	8.693
ALS iterations	514.81	574.94	582.16	584.32	586.48
PESQ score	1.55	1.51	1.49	1.46	1.42
ALSP [13]					
Output SIR [dB]	7.96	3.68	2.33	1.35	0.83
CPU time [s]	13.720	27.558	27.909	28.232	28.324
ALS iterations	96.24	193.37	195.83	198.10	198.75
PESQ score	2.12	1.53	1.46	1.36	1.31

発生させて、ブラインド信号抽出法の性能を比較してきた。一方、実際の室内では、観測信号にはスピーカ、並びにアンプから発生する非線形な成分が含まれる [58–60]。更に、実際のスピーカは点音源ではなく、マイクロフォンは無指向な点マイクロフォンではない。このような非線形性、実際のスピーカ、マイクロホンの悪影響が分離性能に及ぼす影響を、この節で検証することを目的とする。

5 個の指向性スピーカと 9 本のマイクロフォンを図 5.10 の空室に配置した。部屋の形状、大きさ、スピーカとマイクロフォンの個数は異なるが、図 5.6 と同じになるようスピーカとマイクロフォンの配置は決めた。表 5.6 に各ラベルの音源とマイクロフォンの座標を示す。実験に使用した部屋の残響時間は約 350 ms である [14]。観測信号を標本化周波数 16 kHz で 1500 s 間録音した。音声データ長 1500 s はエポック長 $M = 1200$ に相当する。これら以外の測定条件は、前節のシミュレーション条件と同じである。表 5.7 に FDMDI, GNJD, ALUJA, PARAFAC の各アルゴリズム

表 5.5: SNR ≈ 20 dB、 $M = 800$ 、 $N = 4$ 、音源とマイクロフォンの個数が等しい畳み込み混合におけるブラインド信号抽出の平均 PESQ、周波数ビン当たりの平均反復回数、演算時間、出力 SIR

Method	Reverberation time (RT) [ms]				
	100	300	500	700	900
	Overall input SIR [dB]				
	-4.23	-5.19	-5.39	-5.54	-5.90
FDMDI					
Output SIR [dB]	14.96	14.71	13.81	11.77	8.88
CPU time [s]	0.057	0.077	0.091	0.103	0.114
ALS iterations	6.00	8.15	9.73	11.13	12.25
PESQ score	2.64	2.23	2.12	1.93	1.85
GNJD					
Output SIR [dB]	12.87	11.96	11.61	9.59	5.19
CPU time [s]	79.754	85.267	90.779	135.467	190.471
Iterations	162.0	173.0	184.0	275.0	387.0
PESQ score	1.64	1.16	1.03	0.94	0.57
ALUJA					
Output SIR [dB]	14.55	14.35	10.54	7.09	6.12
CPU time [s]	2.246	2.277	2.291	2.305	2.346
Iterations	43.63	44.24	44.52	44.79	46.08
PESQ score	1.78	1.73	1.34	1.25	0.98
PARAFAC					
Output SIR [dB]	8.10	2.39	1.38	0.58	-1.03
CPU time [s]	62.491	77.357	81.470	83.366	91.967
ALS iterations	1063.01	1315.90	1385.86	1418.11	1451.67
PESQ score	0.62	0.57	0.50	0.50	0.50


 図 5.8: SNR ≈ 20 dB、 $M = 400$ 、 $N = 3$ におけるブラインド信号抽出の平均出力 SIR、周波数ビン当たりの平均演算時間と反復回数、平均 PESQ の比較

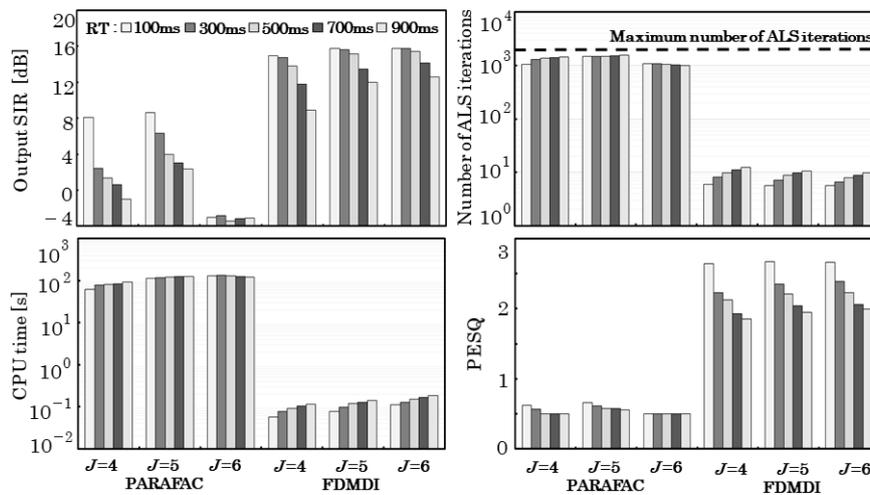


図 5.9: SNR \approx 20 dB, $M = 800$, $N = 4$ におけるブラインド信号抽出の平均出力 SIR、周波数ビン当たりの平均演算時間と反復回数、平均 PESQ の比較

を用いたブラインド信号抽出法の性能を示す。全ブラインド信号抽出法の性能は音源数を 4 個の場合よりも劣化する。また、実環境における性能は、理想的な線形モデルで表した室内環境よりも悪くなった。スピーカとマイクロフォンの個数が等しい畳み込み混合では、FDMDI アルゴリズムの演算時間が最も短くなり、マイクロホンの個数を 9 本に増やした場合には、最も高い PESQ 値を得た。2 番目に高い PESQ 値は、マイクロホンの個数が 7 本の畳み込み混合に FDMDI アルゴリズムを適用した場合であった。マイクロホンの個数がスピーカの個数より多くなると、FDMDI アルゴリズムでは PESQ 値が改善される傾向にある。図 5.11 には、音源信号と FDMDI, ALUJA, PARAFAC の各アルゴリズムを用いたブラインド信号抽出法から得た出力信号のスペクトログラムを示す。FDMDI によって分離された信号は、他の方法に比べ音声のフォルマントが復元できていることが確認できる。ここで、全ブラインド信号抽出法においてスピーカとマイクロフォン間の平均遅延時間は約 0.004s、スピーカと分離信号間の平均遅延時間は約 0.504s であった。

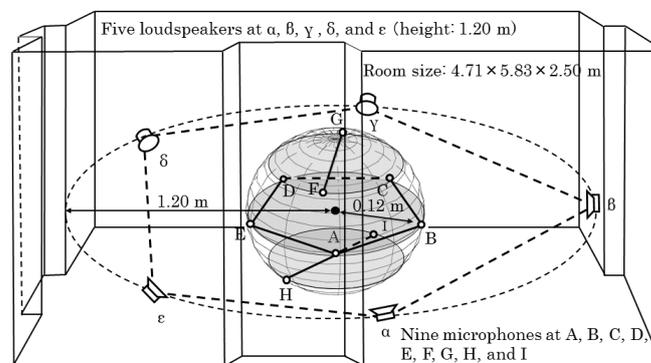


図 5.10: 実環境におけるスピーカとマイクロフォンの配置図 ($N = 5$)

表 5.6: 図 5.10 におけるスピーカとマイクロフォンの座標

Room size: 4.71 × 5.83 × 2.50 m			
		Label in Fig. 5.10	Coordinate [m]
Loudspeakers	$s_1(t)$	α	(2.62, 0.64, 1.20)
	$s_2(t)$	β	(3.42, 1.80, 1.20)
	$s_3(t)$	γ	(2.58, 2.92, 1.20)
	$s_4(t)$	δ	(1.24, 2.46, 1.20)
	$s_5(t)$	ϵ	(1.27, 1.05, 1.20)
Microphones	$x_1(t)$	A	(2.23, 1.66, 1.20)
	$x_2(t)$	B	(2.34, 1.74, 1.20)
	$x_3(t)$	C	(2.30, 1.87, 1.20)
	$x_4(t)$	D	(2.15, 1.87, 1.20)
	$x_5(t)$	E	(2.11, 1.74, 1.20)
	$x_6(t)$	F	(2.23, 1.69, 1.28)
	$x_7(t)$	G	(2.23, 1.86, 1.28)
	$x_8(t)$	H	(2.14, 1.79, 1.12)
	$x_9(t)$	I	(2.31, 1.76, 1.12)

表 5.7: 実環境におけるブラインド信号抽出の周波数ビン当たりの平均演算時間と反復回数、平均 PESQ の比較 (16,384 ポイント FFT、 $N = 5$ 、 $M = 1200$ 、 $\epsilon_C = 10^{-12}$ 、5 本のマイクロフォンは図 5.10 の点 A, B, C, D, E に配置、7 本のマイクロフォンは 5 本のマイクロフォンの配置に点 F, G に 2 本のマイクロフォンを追加)

Method	J	CPU time [s]	Iterations	PESQ score
FDMDI	5	0.382	29.19	1.76
GNJD		1082.555	36.48	1.28
ALUJA		7.841	107.65	1.37
PARAFAC		240.918	1442.147	0.50
FDMDI	7	0.840	20.48	1.90
PARAFAC		447.492	1617.388	0.50
FDMDI	9	1.036	17.01	1.92
PARAFAC		552.538	1343.348	0.50

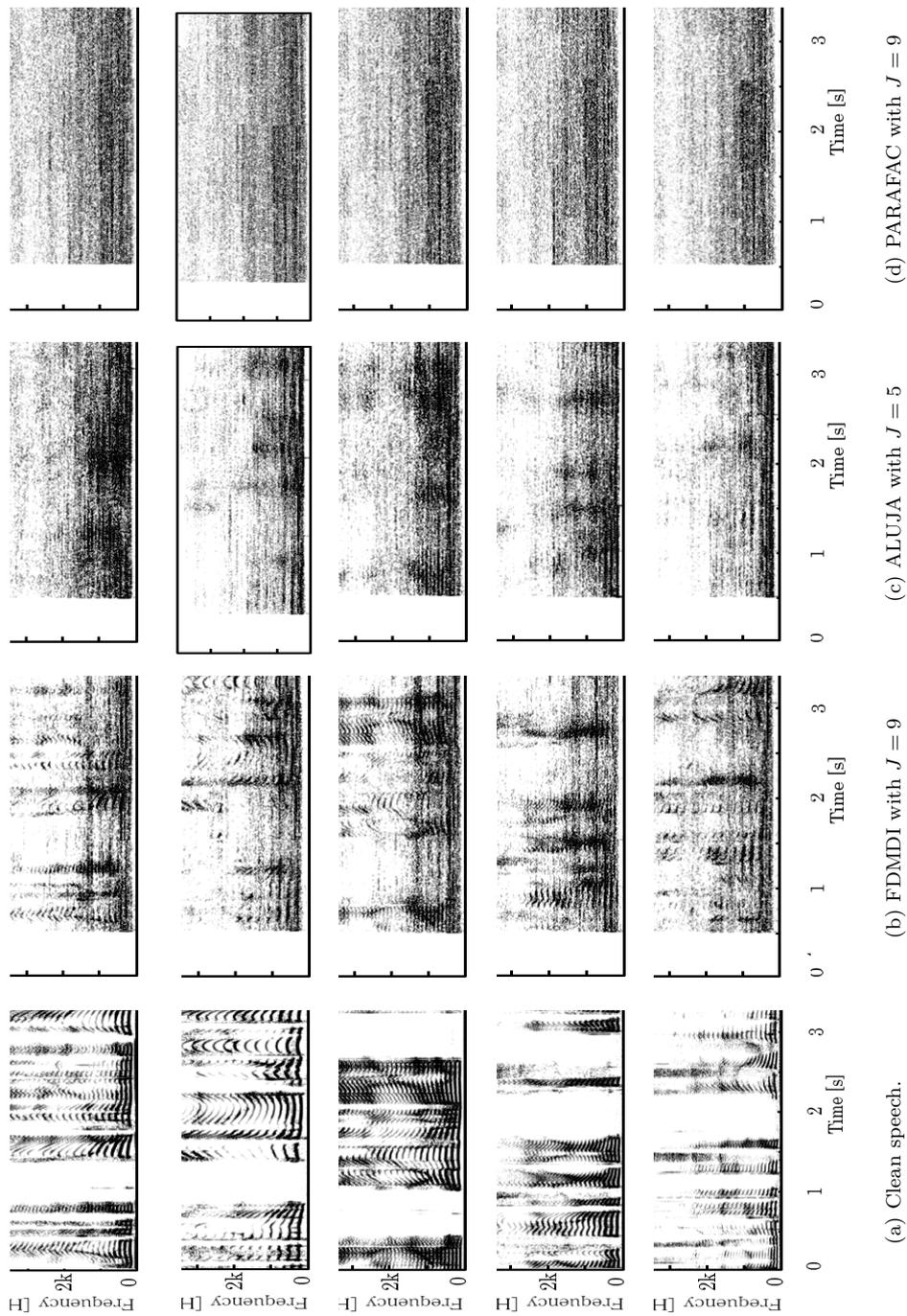


図 5.11: 音源信号と FDMDI, ALUJA, PARAFAC の各アルゴリズムを用いた実環境におけるブラインド信号抽出法から得た出力信号のスペクトログラム (16,384 ポイント FFT、 $M = 1200$ 、 $\epsilon_C = 10^{-12}$ 、ALUJA においては図 5.10 における点 A, B, C, D, E のマイクフォン使用)

5.6 まとめ

DMDI アルゴリズムは混合行列推定の際、入れ子構造を用いて推定しているのに対し、FDMDI アルゴリズムは入れ子構造を用いることなく非反復法による混合行列推定により、信号抽出性能の劣化を伴わずにアルゴリズムの高速化を可能にした。更に、Overdetermined と FDMDI アルゴリズムの組み合わせによって、Determined よりも計算時間が短縮できることがシミュレーション結果によって明らかにした。

演算量の算出と比較、残響が存在する小部屋を想定した計算機シミュレーション結果より、FDMDI アルゴリズムは、従来のアルゴリズムと比較して収束速度が速いという特徴があると同時に、分離性能と反復回数についても従来手法よりも優れていることが確認できた。また、実環境においても、マイクロホンの個数がスピーカの個数より多くすると、入手できる観測雑音のデータ長が短い場合でも FDMDI アルゴリズムで高い分離性能と PESQ 値を達成できることを示した。

PESQ では、従来手法と提案手法との差は大きくは無かったが、実際に試聴してみると提案手法の方が圧倒的に音質が良いことが確認できた。これは人間には不完全な音響的特徴から目標値の推定を行なうという特徴があり、PESQ ではあまり良い数値でなくても、人間の聴覚によって補うことができる為、正確に言葉を認識することができると考えられる。

第6章 結論

6.1 本研究で得られた結果の概要

本論文では、最小 2 乗型同時対角化問題の解法とブラインド信号源分離、並びにブラインド信号抽出への応用について述べた。第 2 章では、周波数畳み込み混合モデルについて示し、最小 2 乗型同時対角化問題の解法がブラインド信号源分離の実現方法の一つであることを示した。また、周波数ビン毎に最小 2 乗型同時対角化問題を解法すると、パーミュテーション問題とスケーリング問題を誘発し、これらの解法も不可欠であることを示した。更に、分離性能の評価指標についても述べた。第 3 章では、分離フィルタを直接推定するために、逆行列の補題を用いて観測信号のクロススペクトル密度行列を展開式で表し、ALSP アルゴリズムに基づく近似同時対角化問題をブラインド信号源分離に適用した。また、ブロック処理を導入して、演算量を削減した。第 4 章では、ALSP アルゴリズムの収束性を改善するために、最小化問題解法にラグランジェの未定乗数法を導入した。同時に、同時対角化問題の前提、即ち、推定した対角化行列が正則でない場合は、この条件を保証する処理を加えることによってアルゴリズムの安定性を向上させた。この結果、分離信号の歪み、発振の抑制の効果的であることをシミュレーションで明らかにした。第 5 章では、マイクロフォンの個数が信号源の個数よりも多い混合過程を対象にブラインド信号抽出の性能を評価した。また、ALSP アルゴリズムの演算量削減のためにラグランジェの未定乗数法を行列演算によって効果的に演算した。提案法である FDMDI アルゴリズムによって、実環境で最大 9 個のマイクロフォンを用いて 5 音源の混合から忠実に音源の復元が可能であることを実証した。

6.2 今後の課題

本研究で提案した ALSP アルゴリズムに基づく同時対角化問題の解法では、5 つの音源のブラインド信号源分離において有効な結果を得た。今後、更に多くの音源数においても十分な分離性能が得られるかを検証する必要がある。overdetermined の場合、マイクロフォンの個数を音源の個数に比べ、どの程度増やせば、能率的に演算でき、かつ、分離性能を向上させることができるのが今後の検討課題の一つである。また、DIEM と IDIEM アルゴリズム [49] は反復法を用いることなく同時対角化問題を解く手法で、演算量が極めて少ないことが特徴である。これらのアルゴリズムと本研究で提案したアルゴリズムの演算量、実環境での分離性能の比較も今後の課題である。

謝辞

本学の博士後期課程入学から今日に至るまで研究を進めるにあたり、御指導、御助言を頂いた東京理科大学工学部情報工学科 古川 利博 教授に心より深く感謝の意を表します。先生には、本学入学以前からの研究の継続を受け入れて頂き、また、その後の研究テーマについて貴重な御提案を頂きました。先生の御提案がなければ本研究をまとめ上げることは叶わなかったと強く感じております。

本論文の審査におきまして、貴重な御意見、御指摘を頂きました東京理科大学工学部情報工学科 宮部 博史 教授、八嶋 弘幸 教授、谷口 行信 教授ならびに池辺 淑子 准教授に深く感謝致します。

元・千葉工業大学工学部電気電子情報工学科 久保田 一 教授には、研究のみならず進路に関しても貴重な御指導、御助言を頂きました。ここに心より感謝の意を表します。

東京工科大学コンピュータサイエンス学部コンピュータサイエンス学科 大石 邦夫 准教授には、学部時代から博士前期課程を経て今日に至るまで、研究の進め方をはじめ物事の考え方や論文の執筆の仕方など、非常に多くのことを御指導頂きました。これまで御指導、御助言を賜り、心より感謝の意を表します。

本研究を進めるにあたり、様々な御支援、御協力を頂いた古川研究室、大石研究室、久保田研究室の皆様に深く感謝致します。

最後に、これまで長い間支えてくれた家族、激励してくれた友人達に感謝致します。

参考文献

- [1] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.
- [2] H. Sawada, H. Kameoka, S. Araki, and N. Ueda, “Multichannel extensions of non-negative matrix factorization with complex-valued data,” *IEEE Trans. Audio Speech Lang. Process.*, vol. 21, no. 5, pp. 971–982, Jul. 2013.
- [3] J.-F. Cardoso and A. Souloumiac, “Blind beamforming for non-Gaussian signals,” *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.
- [4] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [5] A. Yeredor, “Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation,” *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, July 2002. [Online]. Available: <http://www.eng.tau.ac.il/arie/>
- [6] —, “On using exact joint diagonalization for noniterative approximate joint diagonalization,” *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 645–648, Sept. 2005.
- [7] L. Shoker, S. Sanei, and J. Chambers, “Artifact removal from electroencephalograms using a hybrid BSS-SVM algorithm,” *IEEE Signal Process. Lett.*, vol. 12, no. 10, pp. 721–724, Oct. 2005.
- [8] W. D. Clercq, A. Vergult, B. Vanrumste, W. V. Paesschen, and S. V. Huffel, “Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram,” *IEEE Trans. on Biomed. Eng.*, vol. 53, no. 12, pp. 2583–2587, Nov. 2006.
- [9] M. Congedo, C. Gouy-Pailler, and C. Jutten, “On the blind source separation of human electroencephalogram by approximate joint diagonalization of second order statistics,” *Clinical Neurophysiology*, vol. 119, no. 12, pp. 2677–2686, Dec. 2008.
- [10] B. Afsari, “Sensitivity analysis for the problem of matrix joint diagonalization,” *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1148–1171, Sept. 2008.

-
- [11] T. Tsalaile, R. Sameni, S. Sanei, C. Jutten, and J. Chambers, "Sequential blind source extraction for quasi-periodic signals with time-varying period," *IEEE Trans. on Biomed. Eng.*, vol. 56, no. 3, pp. 646–655, Mar. 2009.
- [12] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavsky, and A. Yeredor, "Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications," *Signal Processing Magazine, IEEE*, vol. 31, no. 3, pp. 34–43, May 2014.
- [13] K. Rahbar and J. P. Reilly, "A frequency domain method for blind source separation of convolutive audio mixtures," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 5, pp. 832–844, Sept. 2005.
- [14] S. Saito, K. Oishi, and T. Furukawa, "Convolutional blind source separation using an iterative least-squares algorithm for non-orthogonal approximate joint diagonalization," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 23, no. 12, pp. 2434–2448, Dec. 2015.
- [15] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Matrix Anal. Appl.*, vol. 17, no. 1, pp. 161–164, Jan. 1996.
- [16] D. T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, July 2001.
- [17] A. J. van der Veen, "Joint diagonalization via subspace fitting techniques," in *Proc. 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP2001*, vol. 5, Salt Lake City, USA, pp. 2773–2776, May 2001.
- [18] A. Z. P. Laskov, G. Nolte, and K. R. Müller, "A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation," *J. Mach. Learn. Res.*, vol. 5, pp. 777–800, July 2004.
- [19] R. Vollgraf and K. Obermayer, "Quadratic optimization for simultaneous matrix diagonalization," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3270–3278, Sept. 2006.
- [20] X. L. Li and X. D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, May 2007.
- [21] A. Souloumiac, "Nonorthogonal joint diagonalization by combining givens and hyperbolic rotations," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2222–2231, June 2009.
- [22] P. Tichavsky and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 878–891, Mar. 2009. [Online]. Available: <http://si.utia.cas.cz/CODE/uwedge.html>

-
- [23] D. Nion and L. D. Lathauwer, “An enhanced line search scheme for complex-valued tensor decompositions. application in DS-CDMA,” *Signal Processing*, vol. 88, no. 3, pp. 749–755, May 2008.
- [24] M. Rajih, P. Comon, and R. A. Harshman, “Enhanced line search: A novel method to accelerate PARAFAC,” *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 2, pp. 1128–1147, Sept. 2008.
- [25] D. Nion, K. N. Mokios, N. D. Sidiropoulos, and A. Potamianos, “Batch and adaptive PARAFAC-based blind separation of convolutive speech mixtures,” *IEEE Trans. Audio Speech, Lang. Process.*, vol. 18, no. 6, pp. 1193–1207, Aug. 2010. [Online]. Available: <http://dimitri.nion.free.fr/bss/BSS.html>
- [26] M. Kleinstuber and S. Hao, “Uniqueness analysis of non-unitary matrix joint diagonalization,” *IEEE Trans. Signal Process.*, vol. 61, no. 7, pp. 1786–1796, Apr. 2013.
- [27] V. Maurandi and E. Moreau, “A decoupled Jacobi-like algorithm for non-unitary joint diagonalization of complex-valued matrices,” *Signal Processing Letters, IEEE*, vol. 21, no. 12, pp. 1453–1456, Dec. 2014.
- [28] A. Mesloub, K. Abed-Meraim, and A. Belouchrani, “A new algorithm for complex non-orthogonal joint diagonalization based on shear and givens rotations,” *IEEE Trans. Signal Process.*, vol. 62, no. 8, pp. 1913–1925, Apr. 2014.
- [29] M. Wax and J. Sheinvald, “A least-squares approach to joint diagonalization,” *IEEE Signal Processing Lett.*, vol. 4, no. 2, pp. 52–53, Feb. 1997.
- [30] S. Degerine and E. Kame, “A comparative study of approximate joint diagonalization algorithms for blind source separation in presence of additive noise,” *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 3022–3031, June 2007.
- [31] N. D. Sidiropoulos, G. B. Giannakis, and R. Bro, “Blind PARAFAC receivers for DS-CDMA systems,” *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 810–823, Mar. 2000.
- [32] S. Saito, K. Oishi, and T. Furukawa, “An approach to convolutive backward-model blind source separation based on joint diagonalization,” in *Proc. European Signal Process. Conference EUSIPCO2012*, Bucharest, Romania, pp. 579–583, Aug. 2012.
- [33] —, “An alternating least-squares algorithm for approximate joint diagonalization and its application to blind source separation,” in *Proc. 2014 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP-2014)*, vol. 1, Florence, Italy, pp. 2123–2127, May 2014.

-
- [34] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR systems driven by quasistationary sources using second order statistics," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 406–417, Feb. 2004.
- [35] L. Parra and C. Spence, "Convolutional blind separation of non-stationary sources," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 3, pp. 320–327, May 2000.
- [36] M. Z. Ikram and D. R. Morgan, "Permutation inconsistency in blind speech separation: investigation and solutions," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 1–13, Jan. 2005.
- [37] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Trans. on Speech and Audio Process.*, vol. 12, no. 5, pp. 530–538, Sept. 2004.
- [38] H. Sawada, S. Araki, and S. Makino, "Measuring dependence of bin-wise separated signals for permutation alignment in frequency-domain bss," in *IEEE Int. Symp. Circuits Syst.*, pp. 3247–3250, May 2007.
- [39] R. Mazur and A. Mertins, "An approach for solving the permutation problem of convolutional blind source separation based on statistical signal models," *IEEE Trans. on Audio, Speech, and Lang. Process.*, vol. 17, no. 1, pp. 117–126, Jan 2009.
- [40] K. Matsuoka and S. Nakajima, "Minimal distortion principle for blind source separation," in *Proc. Int. Conf. Independent Compom. Anal. Blind Source Separation*, pp. 722–727, Sept, 2001.
- [41] Y. Hu and P. C. Loizou, "Evaluation of objective quality measures for speech enhancement," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 16, no. 1, pp. 229–238, Jan. 2008.
- [42] G. A. Clark, S. R. Parker, and S. K. Mitra, "A unified approach to time- and frequency-domain realization of FIR adaptive digital filters," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 31, no. 5, pp. 1073–1083, Oct. 1983.
- [43] B. Farhang-Boroujeny and K. S. Chan, "Analysis of the frequency-domain block LMS algorithm," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2332–2342, Aug. 2000.
- [44] G. H. Golub and C. F. V. Loan, "Matrix computations," *The Johns Hopkins University Press*, 1991.
- [45] "ASJ continuous speech corpus for research (ASJ-JIPDEC)," *National Institute of Information*. [Online]. Available: research.nii.ac.jp/src/en/

-
- [46] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70–73, June 1967.
- [47] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Society of America*, pp. 943–950, Apr. 1979.
- [48] "Room impulse response generator for MATLAB." [Online]. Available: <http://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator.html>
- [49] G. Chabriel and J. Barrere, "A direct algorithm for nonorthogonal approximate joint diagonalization," *IEEE Trans. Signal Process.*, vol. 60, no. 1, pp. 39–47, Jan. 2012.
- [50] H. Sawada, S. Araki, and S. Makino, "MLSP 2007 data analysis competition: Frequency-domain blind source separation for convolutive mixtures of speech/audio signals," in *Machine Learning for Signal Processing, 2007 IEEE Workshop on*, Thessaloniki, Greece, pp. 45–50, Aug. 2007.
- [51] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 14, no. 4, pp. 1462–1469, July 2006.
- [52] T. Asamizu, S. Saito, K. Oishi, and T. Furukawa, "Overdetermined blind source separation using approximate joint diagonalization," in *60th IEEE Int. Midwest Symposium on Circuits and Syst.(MWSCAS-2017)*, Boston, USA, pp. 168–171, Aug. 2017.
- [53] A. Westner and V. M. Bove, "Blind separation of real world audio signals using overdetermined mixtures," *Proc. 1st Int. Conf. Ind. Compon. Anal. Blind Signal Separat.*, pp. 11–15, Jan. 1999.
- [54] C. Osterwise and S. L. Grant, "On over-determined frequency domain BSS," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 22, no. 5, pp. 956–966, May 2014.
- [55] B. Widrow and S. D. Stearns, "Adaptive signal processing," *Englewood Cliffs, NJ, Prentice-Hall, Inc.*, 1985.
- [56] X. F. Gong, X. L. Wang, and Q. H. Lin, "Generalized non-orthogonal joint diagonalization with LU decomposition and successive rotations," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1322–1334, Mar. 2015.
- [57] T. Trainini and E. Moreau, "A coordinate descent algorithm for complex joint diagonalization under Hermitian and transpose congruences," *IEEE Trans. Signal Process.*, vol. 57, no. 19, pp. 4974–4983, Oct. 2014.

- [58] A. Stenger and W. Kellermann, “Adaptation of a memoryless preprocessor for nonlinear acoustic echo cancelling,” *Signal Processing*, vol. 80, no. 9, pp. 1747–1760, Sept. 2000.
- [59] A. Guerin, G. Faucon, and R. L. Bouquin-Jeannes, “Nonlinear acoustic echo cancellation based on volterra filters,” *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 672–683, Nov. 2003.
- [60] S. Malik and G. Enzner, “State-space frequency-domain adaptive filtering for nonlinear acoustic echo cancellation,” *IEEE Trans. Audio Speech, Lang. Process.*, vol. 20, no. 7, pp. 2065–2079, Sept. 2012.

研究業績

学術論文

1. S. Saito, K. Oishi, and T. Furukawa, "Convolutional blind source separation using an iterative least-squares algorithm for non-orthogonal approximate joint diagonalization," *IEEE Trans. Audio Speech, Lang. Process.*, vol. 23, no. 12, pp. 2434–2448, Dec. 2015.
2. S. Saito, K. Oishi, and T. Furukawa, "Determined and overdetermined convolutional blind source extractions by approximate joint diagonalization" *Acoust. Sci. and Tech.*, 2019年9月1日掲載決定.

国際会議（査読有）

1. S. Saito, K. Oishi, and T. Furukawa, "An approach to convolutional backward-model blind source separation based on joint diagonalization," in *Proc. European Signal Process. Conference EUSIPCO2012*, Bucharest, Romania, pp. 579–583, Aug. 2012.
2. S. Saito, K. Oishi, and T. Furukawa, "An alternating least-squares algorithm for approximate joint diagonalization and its application to blind source separation," in *Proc. 2014 IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP-2014)*, vol. 1, Florence, Italy, pp. 2123–2127, May 2014.
3. T. Asamizu, S. Saito, K. Oishi, and T. Furukawa, "Overdetermined blind source separation using approximate joint diagonalization," in *60th IEEE Int. Midwest Symposium on Circuits and Syst. (MWSCAS-2017)*, Boston, USA, pp. 168–171, Aug. 2017.

口頭発表（査読無し）

1. 西山慶, 齋藤晋哉, 大石邦夫, 古川利博, "可変 epoch 長を用いた同時対角化問題のための適応ブラインド音源分離法," *信学技報*, vol. 117, no. 515, EA2017-102, pp. 1-6, 2018年3月.
2. 齋藤晋哉, 大石邦夫, 古川利博, "An efficient approximate joint diagonalization algorithm and its application to blind source separation," *信学技報*, vol. 116, no. 475, EA2016-141, pp. 327-332, 2017年3月.

3. 森俊樹, 齋藤晋哉, 大石邦夫, 古川利博, "畳み込み混合下におけるブラインド音源分離のための高速な同時対角化アルゴリズム," 信学技報, vol. 116, no. 475, EA2016-115, pp. 189-194, 2017年3月.
4. 森俊樹, 齋藤晋哉, 大石邦夫, 古川利博, "畳み込み混合下における同時対角化問題を適用した縦続型ブラインド音源分離," 信学技報, vol. 115, no. 521, EA2015-76, pp. 57-62, 2016年3月.

特許

1. 齋藤 晋哉, 大石 邦夫, 久保田 一, "ブラインド信号分離方法およびその装置," 特願 2010-222751(P2010-222751), 2010.09.10, 特開 2011-81373(P2011-81373A), 2011.04.21.
2. 齋藤 晋哉, 大石 邦夫, 古川 利博, "ブラインド信号分離方法およびその装置," 特願 2014-101793(P2014-101793), 2014.4.24, 特開 2015-210512(P2015-210512A), 2015.11.24.
3. 齋藤 晋哉, 大石 邦夫, 古川 利博, "ブラインド信号分離方法およびその装置," 特願 2017-38785(P2017-38785), 2017.2.13, 特開 2017-191309(P2017-191309A), 2017.10.19.

付録

付録 A (3.2) の導出

逆行列の補題を適用すると、 $\mathbf{T}(\omega_k, m)$ の逆行列を

$$\begin{aligned}\mathbf{T}(m, \omega_k)^{-1} &= (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H + \sigma^2\mathbf{I})^{-1} \\ &= (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1} - \sigma^2 (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1} \\ &\quad \cdot \left[\mathbf{I} + \sigma^2 (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1} \right]^{-1} (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1}\end{aligned}\quad (\text{A.1})$$

と表現することができる。 $\left[\mathbf{I} + \sigma^2 (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1} \right]^{-1}$ は

$$\left[\mathbf{I} + \sigma^2 (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1} \right]^{-1} = \sum_{\ell=0}^{\infty} \left[-\sigma^2 (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H)^{-1} \right]^{\ell} \quad (\text{A.2})$$

に展開でき、 σ^2 が $\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H$ の最小固有値よりも小さいと仮定すると、式 (A.2) は収束する。式 (A.1) の右辺第 2 項に式 (A.2) を代入すると

$$\begin{aligned}\mathbf{T}(m, \omega_k)^{-1} &= (\mathbf{B}(\omega_k)\mathbf{\Lambda}(m, \omega_k)\mathbf{B}(\omega_k)^H + \sigma^2\mathbf{I})^{-1} \\ &= (\mathbf{B}(\omega_k)^H)^{-1} \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{B}(\omega_k)^{-1} \\ &\quad - \sigma^2 \sum_{\ell=0}^{\infty} (-1)^{\ell} \sigma^{2\ell} \left[(\mathbf{B}(\omega_k)^H)^{-1} \mathbf{\Lambda}(m, \omega_k)^{-1} \mathbf{B}(\omega_k)^{-1} \right]^{\ell+2}\end{aligned}\quad (\text{A.3})$$

が導出される。

付録B 最小2乗型同時対角化問題のブロック処理とALSPアルゴリズムによる解法

ブロック処理に基づく最小2乗型同時対角化問題は、周波数ビン毎に制約条件 $\|\mathbf{w}_i(\omega_k)\|_2 = 1$ 付き評価量

$$\begin{aligned} C_{\text{DLS}} & \left(\mathbf{W}(\omega_k), \{ \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \}_{\tau=1}^{\lfloor M/L \rfloor} \right) \\ & = \sum_{\tau=1}^{\lfloor M/L \rfloor} \left\| \tilde{\mathbf{T}}(\tau L, \omega_k)^{-1} - \mathbf{W}(\omega_k)^H \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \mathbf{W}(\omega_k) \right\|_F^2 \end{aligned} \quad (\text{B.1})$$

を最小化することによって対角化行列 $\mathbf{W}(\omega_k) \in \mathbb{C}^{N \times N}$ と対角行列 $\{ \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \}_{\tau=1}^{\lfloor M/L \rfloor}$ を求める。このとき $\mathbf{w}_i(\omega_k)$ は i 番目の行ベクトルを表す。

$$\tilde{\mathbf{t}}(\tau L, \omega_k) = \text{vec} \{ \tilde{\mathbf{T}}(\tau L, \omega_k)^{-1} \} \quad (\text{B.2})$$

$$\mathbf{G}(\omega_k) = [\mathbf{w}_1(\omega_k)^T \otimes \mathbf{w}_1(\omega_k)^H, \dots, \mathbf{w}_N(\omega_k)^T \otimes \mathbf{w}_N(\omega_k)^H] \quad (\text{B.3})$$

$$\mathbf{d}(\tau L, \omega_k) = [\lambda_1(\tau L, \omega_k)^{-1}, \lambda_2(\tau L, \omega_k)^{-1}, \dots, \lambda_N(\tau L, \omega_k)^{-1}]^T \quad (\text{B.4})$$

$$\mathbf{G}(\omega_k) \mathbf{d}(\tau L, \omega_k) = \text{vec} \left\{ \mathbf{W}(\omega_k)^H \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \mathbf{W}(\omega_k) \right\} \quad (\text{B.5})$$

$$= [\mathbf{W}(\omega_k)^H \odot \mathbf{W}(\omega_k)] \cdot \mathbf{d}(\tau L, \omega_k) \quad (\text{B.6})$$

を定義すると、式 (B.1) は

$$\sum_{\tau=1}^{\lfloor M/L \rfloor} \left\| \tilde{\mathbf{t}}(\tau L, \omega_k) - \mathbf{G}(\omega_k) \mathbf{d}(\tau L, \omega_k) \right\|_2^2 \quad (\text{B.7})$$

と表現することができる。

$$\mathbf{z}_i(\omega_k) = [\lambda_i(L, \omega_k)^{-1}, \lambda_i(2L, \omega_k)^{-1}, \dots, \lambda_i(L \lfloor M/L \rfloor, \omega_k)^{-1}]^T \quad (\text{B.8})$$

$$\mathbf{C}(\omega_k) = [\tilde{\mathbf{t}}(L, \omega_k), \tilde{\mathbf{t}}(2L, \omega_k), \dots, \tilde{\mathbf{t}}(L \lfloor M/L \rfloor, \omega_k)] \quad (\text{B.9})$$

と定義すると、式 (3.11) は

$$\left\| \mathbf{C}(\omega_k) - \sum_{i=1}^N \mathbf{g}_i(\omega_k) \mathbf{z}_i(\omega_k)^H \right\|_F^2 \quad (\text{B.10})$$

と表現することができる。ここで、式 (B.8) を使用している。ALSP アルゴリズムは手順 1 と手順 2 を交互に繰り返す。

$$\mathbf{F}_j(\omega_k) = \mathbf{C}(\omega_k) - \sum_{\substack{i=1 \\ i \neq j}}^N \mathbf{g}_i(\omega_k) \mathbf{z}_i(\omega_k)^H \quad (\text{B.11})$$

と定義すると、手順 1 では、 $\mathbf{g}_i(\omega_k)$ ($i \neq j$) と $\mathbf{z}_i(\omega_k)$ ($i \neq j$) を固定して $\mathbf{g}_j(\omega_k)$ に関して

$$\|\mathbf{F}_j(\omega_k) - \mathbf{g}_j(\omega_k) \mathbf{z}_j(\omega_k)^H\|_F^2 \quad (\text{B.12})$$

を最小にする。最小 2 乗法を用いて式 (B.12) を解を求めると

$$\hat{\mathbf{g}}_j(\omega_k) = \frac{\mathbf{F}_j(\omega_k) \mathbf{z}_j(\omega_k)}{\mathbf{z}_j(\omega_k)^H \mathbf{z}_j(\omega_k)} \quad (\text{B.13})$$

を得る。制約条件付き最小 2 乗型同時対角化問題では $\hat{\mathbf{g}}_j$ を部分空間 $\Omega \subset \mathbb{C}^{J^2 \times 1}$ 上に写像することによって解法される。ここで、 Ω は

$$\Omega = \{\text{vec}\{\Phi\} | \Phi = \mathbf{v}\mathbf{v}^H, \mathbf{v} \in \mathbb{C}^{J \times 1}, \|\mathbf{v}\|_2 = 1\} \quad (\text{B.14})$$

によって定義される [13, 34]。ALSP アルゴリズムでは、累乘法を適用することによって、 $\|\text{vec}^{-1}\{\hat{\mathbf{g}}_j\} - \mathbf{w}_j(\omega_k)^H \mathbf{w}_j(\omega_k)\|_F^2$ を最小化する $\mathbf{w}_j(\omega_k)$ を求めることができる [13, 34]。 $\mathbf{w}_j(\omega_k)$ の変化が反復終了基準 ϵ_P を下回るまで、累乘法の反復は続けられる。 $\mathbf{G}(\omega_k)$ の全ての列ベクトルの更新が終了するまで上記の最小化は繰り返され、更新前後の $\mathbf{W}(\omega_k)$ の差を誤差行列として記憶する。誤差行列のフロベニウスが反復終了基準 ϵ_G を下回るまで手順 1 は繰り返される。

手順 2 では、 $\mathbf{W}(\omega_k)$ を固定して $\mathbf{\Lambda}(\tau L, \omega_k)^{-1}$ に関して

$$\|\tilde{\mathbf{T}}(\tau L, \omega_k)^{-1} - \mathbf{W}(\omega_k)^H \mathbf{\Lambda}(\tau L, \omega_k)^{-1} \mathbf{W}(\omega_k)\|_F^2 \quad (\text{B.15})$$

を最小化する。式 (B.2) と式 (B.6) を用いると、式 (B.15) は

$$\|\tilde{\mathbf{t}}(\tau L, \omega_k) - \mathbf{G}(\omega_k) \mathbf{d}(\tau L, \omega_k)\|_2^2 \quad (\text{B.16})$$

と表現できる。最小 2 乗法を用いると、式 (B.16) の解

$$\mathbf{d}(\tau L, \omega_k) = \mathbf{G}(\omega_k)^+ \tilde{\mathbf{t}}(\tau L, \omega_k) \quad (\text{B.17})$$

が求められる。ここで、 $\tau = 1, 2, \dots, \lfloor M/L \rfloor$ である。 $C_{\text{DLS}}(\mathbf{W}(\omega_k), \{\mathbf{\Lambda}(\tau L, \omega_k)^{-1}\}_{\tau=1}^{\lfloor M/L \rfloor})$ の変化が ϵ_C を下回るまで ALSP アルゴリズムは繰り返される。

付録C 観測信号のクロススペクトル密度行列の推定

この付録では、周波数ビンを表す添字を再導入する。観測信号のクロススペクトル密度行列を推定する方法としてウォルシュペリオドグラム法 [46] を使用することができる。時間領域の観測信号 $x_i(t)$ は STFT

$$x_i^{(\ell)}(m, \omega_k) = \sum_{t=0}^{K-1} w(t)x_i(t + \ell T_s + mT_b) e^{-j\omega_k t} \quad (\text{C.1})$$

によって時間周波数領域に変換される。ここで、 $w(t)$ は窓関数、 T_s は隣接した2つの窓間のシフトサイズ、 T_b はエポックサイズ、 $N_s + 1$ は条件 $K + N_s T_s \leq T_b$ が不可欠な各エポック内の重複フレーム総数、 $\ell = 0, 1, \dots, N_s$ をそれぞれ表す。疑似定常過程においてエポック m における観測信号のクロススペクトル密度行列は

$$\hat{\mathbf{P}}_x(m, \omega_k) = \frac{1}{N_s + 1} \sum_{\ell=0}^{N_s} \mathbf{x}^{(\ell)}(m, \omega_k) \mathbf{x}^{(\ell)}(m, \omega_k)^H \quad (\text{C.2})$$

によって推定される。但し

$$\mathbf{x}^{(\ell)}(m, \omega_k) = [x_1^{(\ell)}(m, \omega_k), x_2^{(\ell)}(m, \omega_k), \dots, x_J^{(\ell)}(m, \omega_k)]^T \quad (\text{C.3})$$

である。

付録D 式(4.3)の導出

KR 積を用いると、式 (4.2) を

$$C_{\text{DLS}}(\mathbf{G}) = \sum_{m=1}^M \|\tilde{\mathbf{t}}(m) - \mathbf{G}\mathbf{d}(m)\|_2^2 + 2 \sum_{j=1}^N \gamma_j \left(\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{g}_j) - 1 \right) \quad (\text{D.1})$$

と表現することができる。但し

$$\tilde{\mathbf{t}}(m) = \text{vec}\{\tilde{\mathbf{T}}(m)\} \quad (\text{D.2})$$

$$\mathbf{G} = [\mathbf{b}_1^* \otimes \mathbf{b}_1, \dots, \mathbf{b}_N^* \otimes \mathbf{b}_N] \quad (\text{D.3})$$

$$\mathbf{d}(m) = [\lambda_1(m), \lambda_2(m), \dots, \lambda_N(m)]^T \quad (\text{D.4})$$

$$\mathbf{G}\mathbf{d}(m) = \text{vec}\{\mathbf{B}\boldsymbol{\Lambda}(m)\mathbf{B}^H\} = [\mathbf{B}^* \odot \mathbf{B}] \cdot \mathbf{d}(m) \quad (\text{D.5})$$

である。また、 $\lambda_j(m)$ は行列 $\boldsymbol{\Lambda}(m)$ の対角要素、 \mathbf{g}_i は行列 \mathbf{G} の i 番目の列ベクトルをそれぞれ表す。行列 \mathbf{F}_i を

$$\mathbf{F}_i = \mathbf{C} - \sum_{\substack{j=1 \\ j \neq i}}^N \mathbf{g}_j \mathbf{z}_j^H \quad (\text{D.6})$$

と定義する。ここで

$$\mathbf{z}_j = [\lambda_j(1), \lambda_j(2), \dots, \lambda_j(M)]^T \quad (\text{D.7})$$

$$\mathbf{C} = [\tilde{\mathbf{t}}(1), \tilde{\mathbf{t}}(2), \dots, \tilde{\mathbf{t}}(M)] \quad (\text{D.8})$$

である。 \mathbf{g}_i と \mathbf{z}_i を固定し、 \mathbf{g}_i に関して式 (D.1) を最小化すると、制約条件付き DLS 評価量は

$$C_{\text{DLS}}(\mathbf{g}_i) = \|\mathbf{F}_i - \mathbf{g}_i \mathbf{z}_i^H\|_F^2 + 2 \sum_{j=1}^N \gamma_j \left(\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{g}_j) - 1 \right) \quad (\text{D.9})$$

となる。

$$\|\mathbf{F}_i - \mathbf{g}_i \mathbf{z}_i^H\|_F^2 = \text{tr}[\mathbf{F}_i^H \mathbf{F}_i - \mathbf{z}_i \mathbf{g}_i^H \mathbf{F}_i - \mathbf{F}_i^H \mathbf{g}_i \mathbf{z}_i^H + \mathbf{z}_i \mathbf{g}_i^H \mathbf{g}_i \mathbf{z}_i^H] \quad (\text{D.10})$$

であるので、 \mathbf{g}_i の実数部で $C_{\text{DLS}}(\mathbf{g}_i)$ を微分すると

$$\nabla_R(C_{\text{DLS}}(\mathbf{g}_i)) = 2 \left[-\mathcal{R}e(\mathbf{F}_i) \mathcal{R}e(\mathbf{z}_i) + \mathcal{R}e(\mathbf{g}_i) \|\mathbf{z}_i\|_2^2 + \gamma_i \text{vec}\{\mathbf{I}\} \right] \quad (\text{D.11})$$

を得る。

$$\nabla_R (\text{tr} [\mathbf{z}_i \mathbf{g}_i^H \mathbf{F}_i]) = \mathbf{F}_i \mathcal{R}e(\mathbf{z}_i) \quad (\text{D.12})$$

$$\nabla_R (\text{tr} [\mathbf{F}_i^H \mathbf{g}_i \mathbf{z}_i^H]) = \mathbf{F}_i^* \mathcal{R}e(\mathbf{z}_i) \quad (\text{D.13})$$

$$\nabla_R (\text{tr} [\mathbf{z}_i \mathbf{g}_i^H \mathbf{g}_i \mathbf{z}_i^H]) = 2\mathcal{R}e(\mathbf{g}_i) \|\mathbf{z}_i\|_2^2 \quad (\text{D.14})$$

$$\nabla_R \left(\sum_{i=1}^N \gamma_i \left(\text{vec} \{\mathbf{I}\}^T \mathcal{R}e(\mathbf{g}_i) - 1 \right) \right) = \gamma_i \text{vec} \{\mathbf{I}\} \quad (\text{D.15})$$

である。 $\nabla_R (C_{\text{DLS}}(\mathbf{g}_i)) = \mathbf{0}$ にすると

$$\mathcal{R}e(\hat{\mathbf{g}}_i) = \frac{1}{\|\mathbf{z}_i\|_2^2} [\mathcal{R}e(\mathbf{F}_i) \mathcal{R}e(\mathbf{z}_i) \gamma_i \text{vec} \{\mathbf{I}\}] \quad (\text{D.16})$$

となり、制約条件 $\text{vec} \{\mathbf{I}\}^T \mathcal{R}e(\hat{\mathbf{g}}_i) = 1$ に $\mathcal{R}e(\hat{\mathbf{g}}_i)$ を代入すると

$$\gamma_i = \frac{1}{J} \left(\text{vec} \{\mathbf{I}\}^T \mathcal{R}e \{\mathbf{F}_i\} \mathcal{R}e \{\mathbf{z}_i\} - \|\mathbf{z}_i\|_2^2 \right) \quad (\text{D.17})$$

を得る。但し、 $\text{vec} \{\mathbf{I}\}^T \text{vec} \{\mathbf{I}\} = J$ である。 \mathbf{g}_i の虚数部で $C_{\text{DLS}}(\mathbf{g}_i)$ を微分し、 $\nabla_I (C_{\text{DLS}}(\mathbf{g}_i)) = \mathbf{0}$ にすると

$$\mathcal{I}m(\hat{\mathbf{g}}_i) = \frac{1}{\|\mathbf{z}_i\|_2^2} [\mathcal{I}m(\mathbf{F}_i) \mathcal{R}e(\mathbf{z}_i)] \quad (\text{D.18})$$

となる。 $\hat{\mathbf{g}}_i = \mathcal{R}e(\hat{\mathbf{g}}_i) + i\mathcal{I}m(\hat{\mathbf{g}}_i)$ であるので、 $\hat{\mathbf{g}}_i$ を

$$\hat{\mathbf{g}}_i = \frac{1}{\|\mathbf{z}_i\|_2^2} \left[\mathbf{F}_i \mathcal{R}e(\mathbf{z}_i) - \frac{1}{J} \left(\text{vec} \{\mathbf{I}\}^T \mathcal{R}e(\mathbf{F}_i) \mathcal{R}e(\mathbf{z}_i) - \|\mathbf{z}_i\|_2^2 \right) \text{vec} \{\mathbf{I}\} \right] \quad (\text{D.19})$$

と表すことができる。

付録E 式(5.12)の導出

$C_{\text{CDLS}}(\mathbf{G})$ は

$$C_{\text{CDLS}}(\mathbf{G}) = \sum_{m=1}^M \text{tr} \left\{ (\hat{\mathbf{t}}(m) - \mathbf{G}\mathbf{d}(m)) (\hat{\mathbf{t}}(m) - \mathbf{G}\mathbf{d}(m))^H \right\} + 2 \sum_{i=1}^N \eta_i \left(\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{g}_i) - 1 \right) \quad (\text{E.1})$$

と表現することができる。 \mathbf{G} の実数部に関して $C_{\text{CDLS}}(\mathbf{G})$ を微分すると

$$\frac{\partial C_{\text{CDLS}}(\mathbf{G})}{\partial \mathcal{R}e(\mathbf{G})} = 2(\mathcal{R}e(\mathbf{G})\mathbf{F} - \mathcal{R}e(\mathbf{R})) + 2[\eta_1 \text{vec}\{\mathbf{I}\}, \dots, \eta_N \text{vec}\{\mathbf{I}\}] \quad (\text{E.2})$$

となり

$$\mathcal{R}e(\mathbf{G}) = \mathcal{R}e(\mathbf{R})\mathbf{F}^{-1} - [\eta_1 \text{vec}\{\mathbf{I}\}, \dots, \eta_N \text{vec}\{\mathbf{I}\}]\mathbf{F}^{-1} \quad (\text{E.3})$$

を得る。制約 $\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{g}_k) = 1$ に $\mathcal{R}e(\mathbf{G})$ を代入すると

$$\boldsymbol{\eta}^T \tilde{\mathbf{f}}_k = \frac{1}{J} \left(\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{R}) \tilde{\mathbf{f}}_k - 1 \right) \quad (\text{E.4})$$

となる。ここで、 $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_N]^T$ である。また、 $\mathcal{R}e(\mathbf{G}^{(\ell)})$ は

$$\mathcal{R}e(\hat{\mathbf{G}}) = \mathcal{R}e(\mathbf{R})\mathbf{F}^{-1} - \frac{1}{J} \text{vec}\{\mathbf{I}\} \left[\text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{R}) \tilde{\mathbf{f}}_1 - 1, \dots, \text{vec}\{\mathbf{I}\}^T \mathcal{R}e(\mathbf{R}) \tilde{\mathbf{f}}_N - 1 \right] \quad (\text{E.5})$$

と表現することができる。 \mathbf{G} の虚数部に関して $C_{\text{CDLS}}(\mathbf{G})$ を微分すると

$$\frac{\partial C_{\text{CDLS}}(\mathbf{G})}{\partial \mathcal{I}m(\mathbf{G})} = 2(\mathcal{I}m(\mathbf{G})\mathbf{F} - \mathcal{I}m(\mathbf{R})) \quad (\text{E.6})$$

となり

$$\mathcal{I}m(\hat{\mathbf{G}}) = \mathcal{I}m(\mathbf{R})\mathbf{F}^{-1} \quad (\text{E.7})$$

を得る。

付録F FDMDI アルゴリズムの演算量

FDMDI アルゴリズムの各ステップにおける乗算・除算回数を表 F.1 に示す。表 F.1 には、正則な分離行列を構築するための演算量も含まれている。

¹ $N_s + 1$ is the total number of overlapping frames within each epoch [14].
² p_b and p_r denote the number of iterations for the power method.

表 F.1: FDMDI アルゴリズムの演算量

computation	multiplications
<p>Estimation of the cross-spectral density matrices of the observed signal:</p> <p>1) Compute $\hat{\mathbf{P}}_x(m)$ at each epoch using the Welch periodogram method for $m = 1, 2, \dots, M$.</p> <p>Normalization of $\hat{\mathbf{P}}_x(m)$:</p> <p>2) Normalize $\hat{\mathbf{P}}_x(m)$ for $m = 1, 2, \dots, M$ according to (5.1).</p> <p>3) Convert $\hat{\mathbf{P}}_x(m)$ into $\tilde{\mathbf{P}}_x(m)$ according to (5.6).</p> <p>Initialization:</p> <p>4) Set the initial values of $\mathbf{W}(\omega_k)$ to the indirect LS estimate at frequency bin ω_{k-1} with $\mathbf{W}(\omega_0) = \mathbf{I}$.</p> <p>First phase (Estimation of $\{\mathbf{A}(m)\}_{m=1}^M$):</p> <p>5) Minimize $C_{\text{LS}}(\{\mathbf{A}(m)\}_{m=1}^M)$ to find $\mathbf{A}(m)$ by the method of least squares.</p> <p>Second phase (Estimation of \mathbf{B}):</p> <p>6) Compute \mathbf{D} and \mathbf{R}_d according to (5.10) and (5.11), respectively.</p> <p>7) Compute the inverse of \mathbf{D}.</p> <p>8) Minimize $C_{\text{DLs}}(\mathbf{G})$ to find $\hat{\mathbf{G}}$ by the method of Lagrange multipliers according to (5.12).</p> <p>9) Minimize $\ \text{vec}^{-1}\{\hat{\mathbf{g}}_i\} - \mathbf{b}_i \mathbf{b}_i^H\ _F^2$ to find \mathbf{b}_i by the power method for $i = 1, 2, \dots, N$.</p> <p>Construction of the full-rank unmixing matrix:</p> <p>10) If $\text{rank}(\mathbf{B}^H \mathbf{B}) = r < N$, construct the orthonormal sets $\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \dots, \mathbf{v}_N$ and $\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \dots, \mathbf{u}_N$ by the Gram-Schmidt process. Then replace \mathbf{B} with $\mathbf{V}\mathbf{\Sigma}\mathbf{U}^H$.</p> <p>11) Compute \mathbf{W} from \mathbf{B} according to (5.26).</p> <p>12) Go to step 5 until $C_{\text{DLs}}(\mathbf{G})$ changes by less than ϵ_C between iterations.</p> <p>The per-ALS-iteration total:</p>	$2(J^2 + J)MN_s^{-1} (MN_s \text{ divisions})$ $2(J^2 + J)M \quad (M \text{ division})$ $(J^3 + J^2)M$ $MN^2 + 2MJ^2N$ $N^2(N-1)/2$ $3(JN)^2 + N \quad (1 \text{ division})$ $(2J^2 + 4J)Np_b^2 \quad (Np_b \text{ divisions})$ $2J^3 + J^2(J+1)$ $+ 2(2J^2 + 4J)(N-r)p_r^2$ $(2(N-r)p_r \text{ divisions})$ $2J^3$ $2J^2$ $(J^3 + J^2 + N^2 + 2J^2N)M$ $+ 5J^3 + 3J^2 + N^3/2 - N^2/2 + N + 3(JN)^2$ $+ (2J^2 + 4J)Np_b + 2(2J^2 + 4J)(N-r)p_r$ $(Np_b + 2(N-r)p_r + 1 \text{ divisions})$ $2J^2$
<p>Realization of the delayed unmixing filter:</p> <p>13) Multiply \mathbf{W} by $e^{-j\pi k}$</p>	$2J^2$

付録G DMDI アルゴリズムの演算量

DMDI アルゴリズムの各ステップにおける乗算・除算回数を表 G.1 に示す。DMDI アルゴリズムの手順 1 と手順 2 は FDMDI アルゴリズムと反対である。即ち、DMDI アルゴリズムでは、手順 1 で混合行列を推定し、手順 2 で対角行列を推定する。表 G.1 には、正則な分離行列を構築するための演算量も含まれている。

³ p_a denotes the number of iterations for estimating \mathbf{G} .

表 G.1: DMDI アルゴリズムの演算量

computation	multiplications
<p>Estimation of the cross-spectral density matrices of the observed signal:</p> <p>1) Compute $\hat{\mathbf{P}}_x(m, \omega_k)$ at each epoch using the Welch periodogram method for $m = 1, 2, \dots, M$.</p> <p>Normalization of $\hat{\mathbf{P}}_x(m)$:</p> <p>2) Normalize $\hat{\mathbf{P}}_x(m, \omega_k)$ by (5.1) for $m = 1, 2, \dots, M$.</p> <p>Initialization:</p> <p>3) Set the initial values of $\mathbf{B}(\omega_k)$ and $\{\mathbf{\Lambda}(m)\}_{m=1}^M$ to the LS-AJD estimate at frequency bin ω_{k-1} with $\mathbf{B}(\omega_0) = \mathbf{I}$ and $\{\mathbf{\Lambda}(m, \omega_0) = \mathbf{I}\}_{m=1}^M$.</p> <p>First phase (Estimation of \mathbf{B}):</p> <p>4) Compute \mathbf{F}_i according to (44) in [14] and then minimize $C_{\text{DLS}}(\mathbf{g}_i)$ to find $\hat{\mathbf{g}}_i$ by the method of Lagrange multipliers for $i = 1, 2, \dots, N$ according to (16) in [14].</p> <p>5) Go to step 4 until \mathbf{G} changes by less than ϵ_G between iterations.</p> <p>6) Minimize $\ \text{vec}^{-1}\{\hat{\mathbf{g}}_i\} - \mathbf{b}_i \mathbf{b}_i^H\ _F^2$ to find \mathbf{b}_i by the power method for $i = 1, 2, \dots, N$.</p> <p>Construction of the full-rank unmixing matrix:</p> <p>7) If rank $(\mathbf{B}^H \mathbf{B}) = r < N$, construct the orthonormal sets $\{\mathbf{v}_i\}_{i=r+1}^N$ and $\{\mathbf{u}_i\}_{i=r+1}^N$ by the Gram-Schmidt process, and replace \mathbf{B} with $\mathbf{V} \Sigma \mathbf{U}^H$ as given by (5.21).</p> <p>8) Compute \mathbf{W} from \mathbf{B} by (5.26).</p> <p>Second phase (Estimation of $\{\mathbf{\Lambda}(m)\}_{m=1}^M$):</p> <p>9) Minimize $C_{\text{ILS}}(\{\mathbf{\Lambda}(m)\}_{m=1}^M)$ to find $\hat{\mathbf{\Lambda}}(m)$ through the method of least squares according to (5.3).</p> <p>10) Compute $\mathbf{Gd}(m)$ from \mathbf{B} and $\hat{\mathbf{\Lambda}}(m)$ by (5.9).</p> <p>11) Go to step 4 until $C_{\text{DLS}}(\mathbf{G})$ changes by less than ϵ_G between iterations.</p> <p>The per-ALS-iteration total:</p>	$2(J^2 + J)MN_s \quad (MN_s \text{ divisions})$ $2(J^2 + J)M \quad (M \text{ divisions})$ $J^2 MN p_a^3$ $+ (4J^2 + 2)NM p_a \quad (N p_a \text{ divisions})$ $2J^2 p_a$ $(2J^2 + 4J)N p_b \quad (N p_b \text{ divisions})$ $2J^3 + J^2(J + 1)$ $+ 2(2J^2 + 4J)(N - r)p_r$ $(2(N - r)p_r \text{ divisions})$ $2J^3$ $(J^3 + J^2)M$ $2J^2 N$ $2J^2$ $(J^3 + J^2)M + (M + 5)J^3 + (M + 2N + 3)J^2$ $+ \{J^2 MN + (4J^2 + 2)NM + 2J^2\} p_a$ $+ (2J^2 + 4J)N p_b + 2(2J^2 + 4J)(N - r)p_r$ $(N p_a + N p_b + 2(N - r)p_r + 1 \text{ divisions})$ $2J^2$
<p>Realization of the delayed unmixing filter:</p> <p>12) Multiply \mathbf{W} by $e^{-j\pi k}$.</p>	$2J^2$