

混合過程推定にスパース性を利用したブラインド音源分離

Blind Source Separation Using Sparsity Prior on Mixing Matrix Estimation

立命館大学情報理工学部 ○ 北原 大地, 小田 亮太, 平林 晃

D. Kitahara, R. Oda, and A. Hirabayashi

College of Information Science and Engineering, Ritsumeikan University

Abstract In this paper, we propose a blind source separation (BSS) algorithm using *a priori* information on the mixing process. In the conventional BSS algorithms, including the state-of-the-art method called *independent low-rank matrix analysis (ILRMA)*, the separation performance is sometimes degraded by inappropriate estimation of the mixing matrix. To solve this problem, we estimate the mixing matrix by considering the sparsity and the decreasing envelope of impulse responses. Numerical experiments using publicly available impulse responses show that the proposed method outperforms ILRMA with respect to evaluation measures such as SDR, SIR, and SAR.

1 はじめに

音源分離とは、複数の音源信号が混合された観測信号から、混合前の各音源信号を推定する技術である。特に、音源の混合過程に関する事前情報（音源やマイクの位置など）がほとんど未知である場合には、ブラインド音源分離と呼ばれる。音源信号の混合過程は、時間領域では畳み込み演算となり取り扱いが難しいため、周波数領域で乗算としてモデル化することが多い。過決定条件（マイク数 \geq 音源数）下におけるブラインド音源分離では、周波数ビンごとに、混合系（混合行列）の逆写像である分離系（分離行列）を求める分離手法が数多く提案されている [1]–[3]。これらの手法では、音源信号の生成モデルを統計的独立性と優ガウス性（ガウス関数と比較してピークが鋭い性質）に基づいて設計した上で、最尤推定を行うことで分離行列を決定している。

例えば、独立成分分析 (independent component analysis: ICA) [1] では、音源信号の生成モデルとして複素ラプラス分布を仮定して最尤推定を行う。しかしながら、この生成モデルには周波数ビンごとに分離信号の順序を入れ替えても尤度に変化しないという問題があり、分離信号の順序を適切に並べて揃える複雑な後処理が必要となる [4]。上記後処理を回避するため、独立ベクトル分析 (independent vector analysis: IVA) [2] では、球対称な複素ラプラス分布を音源信号の生成モデルとする。これにより、同一音源の異なる周波数間の高次相関が考慮され、分離信号の順序が自動的に揃うようになる。また、独立低ランク行列分析 (independent low-rank matrix analysis: ILRMA) [3] は、各音源信号のパワースペクトログラムが非負値行列の積で低ランク近似できる（非負値

行列分解と呼ばれる）ことに基づいて、時変複素ガウス分布を音源信号の生成モデルとしており、ICA や IVA よりも高い分離性能を示すことが知られている。

上記のように、音源信号の生成モデルを改良することで様々な分離手法が開発されてきた。一方で、マイクロホンアレイを用いた場合において、音源信号の到来方向を利用した分離手法が提案されている [5]–[7]。これらの手法では、到来方向から推定された混合行列の逆行列を考慮することで、高い分離性能を持つ分離行列を求めている。しかしながら、通常は音源やマイクの位置関係は未知である場合が多く、上記手法を適用できない。

本研究では、混合過程に関する一般的な性質である「音源からマイクへのインパルス応答のスパース性」を利用する。実際、室内のインパルス応答を推定する研究において、スパース性やブロックスパース性を考慮したコスト関数の値を小さくすることで、推定精度を改善できることが報告されている [8]–[10]。この考えを利用して、従来の最尤推定に基づく指標に、インパルス応答のスパース性を評価する指標を加えたコスト関数を定義し、これを最小化することで音源分離を行う手法を提案する。提案法では、分離行列とインパルス応答を同時に推定し、スパースなインパルス応答から計算される混合行列の逆行列を考慮することで、より優れた分離行列を求めることができる。分離行列とインパルス応答の推定値は交互に更新され、それぞれ反復射影法 [7] とハード閾値処理を用いることで、提案する最適化問題を比較的高速に解くことができる。公開されている音源信号とインパルス応答を用いた数値実験により、提案法が音楽信号と音声信号のいずれに対しても、従来法より高品質な分離結果を与えることを示す。

2 過決定条件下におけるブラインド音源分離

2.1 ブラインド音源分離問題

音源数を $N (\geq 2)$ とし、マイク数を $M (\geq 1)$ とする。 n 番目 ($n = 1, 2, \dots, N$) の音源から時刻 t に生成される音を $s_n[t]$ とし、 m 番目 ($m = 1, 2, \dots, M$) のマイクで時刻 t に収録 (観測) される音を $x_m[t]$ とする。 n 番目の音源から m 番目のマイクへのインパルス応答を $h_{m,n}[\tau]$ ($\tau = 0, 1, \dots, T-1$) とすると、混合音 $x_m[t]$ は

$$x_m[t] = \sum_{n=1}^N (h_{m,n} * s_n)[t] = \sum_{n=1}^N \sum_{\tau=0}^{T-1} h_{m,n}[\tau] s_n[t-\tau] \quad (1)$$

のように表される。ここで、 $*$ は畳み込み演算である。離散時間信号 $s_n[t]$ と $x_m[t]$ の短時間フーリエ変換を

$$\begin{cases} \hat{s}_{i,j,n} := \sum_{\tau=0}^{L-1} \psi[\tau] s_n[(j-1)\eta + \tau] \exp(-i\frac{2\pi(i-1)\tau}{L}) \\ \hat{x}_{i,j,m} := \sum_{\tau=0}^{L-1} \psi[\tau] x_m[(j-1)\eta + \tau] \exp(-i\frac{2\pi(i-1)\tau}{L}) \end{cases}$$

とする。ここで、 L はフレーム長、 $\psi[\tau]$ は窓関数、 η はフレームシフトであり、 i は虚数単位である。また、 $i = 1, 2, \dots, L$ は周波数インデックス、 $j = 1, 2, \dots, J$ はフレームインデックスを表す。更に、インパルス応答 $h_{m,n}[\tau]$ の離散時間フーリエ変換を

$$\hat{h}_{m,n}(\omega) := \sum_{\tau=0}^{T-1} h_{m,n}[\tau] \exp(-i\omega\tau)$$

とし、これを $\omega_i := \frac{2\pi(i-1)}{L}$ でサンプリングしたものを

$$\hat{a}_{i,m,n} := \hat{h}_{m,n}(\omega_i) = \sum_{\tau=0}^{T-1} h_{m,n}[\tau] \exp(-i\frac{2\pi(i-1)\tau}{L}) \quad (2)$$

と定義する。このとき、短時間フーリエ変換のフレーム長 L がインパルス応答の長さ T より十分大きいならば、各フレームで式 (1) の「時間領域における畳み込み演算」を「周波数領域における乗算」に変換することができる。したがって、 $i = 1, 2, \dots, \lfloor \frac{L+2}{2} \rfloor =: I$ と $j = 1, 2, \dots, J$ に対して¹,

$$\hat{x}_{i,j,m} = \sum_{n=1}^N \hat{a}_{i,m,n} \hat{s}_{i,j,n} + \hat{\epsilon}_{i,j,m} \quad (3)$$

が成立する。ここで、 $\hat{\epsilon}_{i,j,m}$ は「短時間フーリエ変換で得られる周波数成分」と「真の周波数成分」の差に起因

¹ $i = 2, 3, \dots, \lfloor \frac{L+1}{2} \rfloor$ に対して、 $\hat{s}_{i,j,n}$, $\hat{x}_{i,j,m}$, $\hat{a}_{i,m,n}$ はそれぞれ $\hat{s}_{L-i+2,j,n}$, $\hat{x}_{L-i+2,j,m}$, $\hat{a}_{L-i+2,m,n}$ と複素共役な関係にあるので、 $i = 1, 2, \dots, \lfloor \frac{L+2}{2} \rfloor$ のときのみを考えれば十分である。

するモデル誤差である。ベクトル $\mathbf{s}_{i,j}$, $\mathbf{x}_{i,j}$, $\boldsymbol{\epsilon}_{i,j}$, $\mathbf{a}_{i,j}$ と行列 A_i を

$$\begin{cases} \mathbf{s}_{i,j} := (\hat{s}_{i,j,1}, \hat{s}_{i,j,2}, \dots, \hat{s}_{i,j,N})^T \in \mathbb{C}^N \\ \mathbf{x}_{i,j} := (\hat{x}_{i,j,1}, \hat{x}_{i,j,2}, \dots, \hat{x}_{i,j,M})^T \in \mathbb{C}^M \\ \boldsymbol{\epsilon}_{i,j} := (\hat{\epsilon}_{i,j,1}, \hat{\epsilon}_{i,j,2}, \dots, \hat{\epsilon}_{i,j,M})^T \in \mathbb{C}^M \\ \mathbf{a}_{i,n} := (\hat{a}_{i,1,n}, \hat{a}_{i,2,n}, \dots, \hat{a}_{i,M,n})^T \in \mathbb{C}^M \\ A_i := (\mathbf{a}_{i,1}, \mathbf{a}_{i,2}, \dots, \mathbf{a}_{i,N}) \in \mathbb{C}^{M \times N} \end{cases}$$

のように定義すれば、式 (3) から

$$\mathbf{x}_{i,j} = A_i \mathbf{s}_{i,j} + \boldsymbol{\epsilon}_{i,j} \quad (i = 1, 2, \dots, I; j = 1, 2, \dots, J) \quad (4)$$

が成立する。ブラインド音源分離とは、式 (4) における混合行列 A_i が未知の状態、 $\mathbf{x}_{i,j}$ から $\mathbf{s}_{i,j}$ を推定し、最終的に各音源信号 $s_n[t]$ ($n = 1, 2, \dots, N$) を推定する問題²である。

過決定条件 ($M \geq N$)³下においては、 A_i の一般逆行列

$$W_i := (\mathbf{w}_{i,1}, \mathbf{w}_{i,2}, \dots, \mathbf{w}_{i,N})^H := (A_i^H A_i)^{-1} A_i^H \in \mathbb{C}^{N \times M}$$

を求めることで、 $\mathbf{s}_{i,j}$ を

$$\begin{aligned} \mathbf{y}_{i,j} &:= (\hat{y}_{i,j,1}, \hat{y}_{i,j,2}, \dots, \hat{y}_{i,j,N})^T \in \mathbb{C}^N \\ &:= (\mathbf{w}_{i,1}^H \mathbf{x}_{i,j}, \mathbf{w}_{i,2}^H \mathbf{x}_{i,j}, \dots, \mathbf{w}_{i,N}^H \mathbf{x}_{i,j})^T = W_i \mathbf{x}_{i,j} \end{aligned}$$

のように推定する手法が主流である [1]-[3]。ここで、 W_i は分離行列と呼ばれ、特に $\boldsymbol{\epsilon}_{i,j} = \mathbf{0}$ である場合には、真の分離行列が求めれば $\mathbf{y}_{i,j} = \mathbf{s}_{i,j}$ が成立する。また、 $\boldsymbol{\epsilon}_{i,j} \neq \mathbf{0}$ の場合でも、二乗誤差 $\|\mathbf{x}_{i,j} - A_i \mathbf{y}_{i,j}\|_2^2$ を最小にする $\mathbf{y}_{i,j}$ が $\mathbf{s}_{i,j}$ の推定値として得られる。

2.2 音源信号の独立性と優ガウス性を利用する従来法

音源信号 $s_n[t]$ の複素スペクトログラム \bar{S}_n は

$$\begin{cases} \bar{\mathbf{s}}_{j,n} := (\hat{s}_{1,j,n}, \hat{s}_{2,j,n}, \dots, \hat{s}_{I,j,n})^T \in \mathbb{C}^I \\ \bar{S}_n := (\bar{\mathbf{s}}_{1,n}, \bar{\mathbf{s}}_{2,n}, \dots, \bar{\mathbf{s}}_{J,n}) \in \mathbb{C}^{I \times J} \end{cases}$$

と定義される。代表的な従来法である、独立成分分析 (ICA) [1], 独立ベクトル分析 (IVA) [2], 独立低ランク行列分析 (ILRMA) [3] はいずれも、「複素スペクトログラム \bar{S}_n が優ガウ斯的な分布⁴から独立に生成される」と仮定して分離行列 W_i を求める手法である。以下では、

²実際には、 $(h_{m,n} * s_n)[t]$ ($m = 1, 2, \dots, M; n = 1, 2, \dots, N$) を推定する場合がほとんどである。

³より厳密には、 $\text{rank}(A_i) = N$ ($\forall i = 1, 2, \dots, I$) の場合であり、更に A_i の条件数は大きくないことが望まれる。

⁴ X を実数値確率変数とする。 X の平均を $\mu := E[X]$, 分散を $\sigma^2 := E[(X - \mu)^2]$ とすると、尖度は $v := E[(X - \mu)^4] / \sigma^4 - 3$ と定義される。 X は $v > 0$ のとき優ガウ斯的であるといい、 $v < 0$ のとき劣ガウ斯的であるという (ガウス分布は $v = 0$ である)。

式 (4) において $\epsilon_{i,j} = \mathbf{0}$ が成立しているものとして議論を進める。

各成分が $\hat{s}_{i,j,n}, \hat{x}_{i,j,m}$ であるテンソルを $\mathcal{S} \in \mathbb{C}^{I \times J \times N}$, $\mathcal{X} \in \mathbb{C}^{I \times J \times M}$ とし, それぞれの確率密度関数を p_s, p_x とする. $\epsilon_{i,j} = \mathbf{0}$ のとき, A_i ($i = 1, 2, \dots, I$) は \mathcal{S} から \mathcal{X} への 1 対 1 の線形写像となり, p_x は p_s に比例する. 更に \bar{S}_n の独立性から,

$$p_x(\mathcal{X}) = \frac{p_s(\mathcal{S})}{\prod_{i=1}^I (\det(A_i^H A_i))^J} \\ = \prod_{i=1}^I (\det(W_i W_i^H))^J \prod_{n=1}^N p_n(\bar{S}_n)$$

が成立する. ここで, p_n は \bar{S}_n の確率密度関数である. W_i を真の分離行列とすると, $\hat{s}_{i,j,n} = \hat{y}_{i,j,n} = \mathbf{w}_{i,n}^H \mathbf{x}_{i,j}$ であるため, p_x は全ての分離行列 W_i ($i = 1, 2, \dots, I$) の組からなるテンソル $\mathcal{W} \in \mathbb{C}^{N \times M \times I}$ をパラメータとする関数 $p_x(\mathcal{X} | \mathcal{W})$ になる. したがって, 最尤推定の考えに基づき, 尤度 $p_x(\mathcal{X} | \mathcal{W})$ の最大化問題, つまり負対数尤度 $-\log(p_x(\mathcal{X} | \mathcal{W}))$ の最小化問題

$$\underset{\mathcal{W}}{\text{minimize}} - \sum_{n=1}^N \log(p_n(\bar{S}_n)) - J \sum_{i=1}^I \log(\det(W_i W_i^H)) \quad (5)$$

を解くことで分離行列 W_i を決定することができる.

p_n には優ガウスの確率密度関数, すなわち, ガウス関数に比べてピークが鋭く大きな裾を持つ確率密度関数を仮定する. 例えば, ICA では複素ラプラス分布⁵, IVA では球対称複素ラプラス分布⁶, ILRMA では (分散が非負値行列分解で定義されている) 時変複素ガウス分布⁷ を仮定した状態で, 式 (5) の最適化問題を解いている. いずれの場合においても, 必要であれば補助関数法 [11] を用いることで, 分離行列 W_i ごとの最適化問題

$$\underset{W_i}{\text{minimize}} \sum_{n=1}^N \mathbf{w}_{i,n}^H U_{i,n}^{(l+1)} \mathbf{w}_{i,n} - \log(\det(W_i W_i^H)) \quad (6)$$

に帰着させて \mathcal{W} の推定値を更新する. 式 (6) は $(l+1)$ 回目の更新に用いる最適化問題であり, $U_{i,n}^{(l+1)} \in \mathbb{C}^{M \times M}$

⁵ 複素ラプラス分布

$$p_n(\bar{S}_n) := \prod_{i=1}^I \prod_{j=1}^J \frac{1}{2\pi\sigma_n^2} \exp\left(-\frac{|\hat{s}_{i,j,n}|}{\sigma_n}\right).$$

⁶ 球対称複素ラプラス分布

$$p_n(\bar{S}_n) := \prod_{j=1}^J \frac{2^{I-1}(I-1)!}{(2\pi\sigma_n^2)^I (2I-1)!} \exp\left(-\frac{\|\bar{s}_{j,n}\|_2}{\sigma_n}\right).$$

⁷ 時変複素ガウス分布

$$p_n(\bar{S}_n) := \prod_{i=1}^I \prod_{j=1}^J \frac{1}{\pi r_{i,j,n}} \exp\left(-\frac{|\hat{s}_{i,j,n}|^2}{r_{i,j,n}}\right).$$

は

$$U_{i,n}^{(l+1)} := \frac{1}{J} \sum_{j=1}^J \frac{X_{i,j}}{r_{i,j,n}^{(l+1)}} := \frac{1}{J} \sum_{j=1}^J \frac{\mathbf{x}_{i,j} \mathbf{x}_{i,j}^H}{r_{i,j,n}^{(l+1)}}$$

と定義される. ここで, $r_{i,j,n}^{(l+1)}$ の値はアルゴリズムごとに

$$r_{i,j,n}^{(l+1)} := \begin{cases} 2\sigma_n |\hat{y}_{i,j,n}^{(l)}| & \text{(ICA のとき)} \\ 2\sigma_n \|\bar{\mathbf{y}}_{j,n}^{(l)}\|_2 & \text{(IVA のとき)} \\ \sum_{k=1}^{K_n} b_{i,k,n}^{(l+1)} c_{k,j,n}^{(l+1)} & \text{(ILRMA のとき)} \end{cases}$$

と与えられる⁸. 分離行列 W_i は, 式 (6) のコスト関数の停留点 (局所最適解) の近似値を計算することにより,

$$\begin{cases} \mathbf{v}_{i,n}^{(l+1)} = (U_{i,n}^{(l+1)})^{-1} \mathbf{a}_{i,n}^{(l)} \\ \mathbf{w}_{i,n}^{(l+1)} = \frac{\mathbf{v}_{i,n}^{(l+1)}}{\sqrt{\mathbf{v}_{i,n}^{(l+1)H} U_{i,n}^{(l+1)} \mathbf{v}_{i,n}^{(l+1)}}} \end{cases} \quad (7)$$

のように更新される. ここで, $\mathbf{a}_{i,n}^{(l)}$ は $W_i^{(l)}$ の一般逆行列 $A_i^{(l)} = W_i^{(l)H} (W_i^{(l)} W_i^{(l)H})^{-1}$ の第 n 列ベクトルである.

3 混合過程の先験情報を考慮した分離手法

3.1 混合過程に関する先験情報の導入

第 2 節で紹介した従来法では, 「複素スペクトログラムの各成分の振幅値 $|\hat{s}_{i,j,n}|$ は高い頻度で 0 になる」という先験情報に基づき, \bar{S}_n の確率密度関数 p_n に優ガウスの分布を仮定した. 結果として, これらの手法は, 音源信号の先験情報に基づいて設計された最適化問題を解くことにより, 分離行列 W_i を決定していることになる. 一方で, マイクロホンアレイを用いて収録を行う場合, 音源信号の到来方向が既知であれば, 混合行列 A_i の各列ベクトル (ステアリングベクトルと呼ばれる) を推定できる. 推定された混合行列の一般逆行列を利用して, 高い分離性能を持つ分離行列 W_i を求める手法が提案されている [5]–[7]. しかしながら, ブラインド音源分離

⁸ ILRMA では, 音源信号 $s_n[t]$ のパワースペクトログラム $|\bar{S}_n|^{-2}$ が非負値行列 $B_n \in \mathbb{R}_+^{I \times K_n}$ (基底行列) と $C_n \in \mathbb{R}_+^{K_n \times J}$ (係数行列) の積で低ランク近似できると仮定する. ここで, K_n は基底ベクトルの数を表し, 低ランク近似のため $K_n \ll \min(I, J)$ を満たしている. $b_{i,k,n}$ と $c_{k,j,n}$ は B_n と C_n の各成分であり, 時変複素ガウス分布の分散は $r_{i,j,n} := \sum_{k=1}^{K_n} b_{i,k,n} c_{k,j,n}$ と定義される. $b_{i,k,n}$ と $c_{k,j,n}$ は, 補助関数法を用いて

$$\begin{cases} b_{i,k,n}^{(l+1)} = b_{i,k,n}^{(l)} \sqrt{\frac{\sum_{j=1}^J |\hat{y}_{i,j,n}^{(l)}|^2 c_{k,j,n}^{(l)} (\sum_{k'=1}^{K_n} b_{i,k',n}^{(l)} c_{k',j,n}^{(l)})^{-2}}{\sum_{j=1}^J c_{k,j,n}^{(l)} (\sum_{k'=1}^{K_n} b_{i,k',n}^{(l)} c_{k',j,n}^{(l)})^{-1}}} \\ c_{k,j,n}^{(l+1)} = c_{k,j,n}^{(l)} \sqrt{\frac{\sum_{i=1}^I |\hat{y}_{i,j,n}^{(l)}|^2 b_{i,k,n}^{(l+1)} (\sum_{k'=1}^{K_n} b_{i,k',n}^{(l+1)} c_{k',j,n}^{(l)})^{-2}}{\sum_{i=1}^I b_{i,k,n}^{(l+1)} (\sum_{k'=1}^{K_n} b_{i,k',n}^{(l+1)} c_{k',j,n}^{(l)})^{-1}}} \end{cases}$$

のように更新される.

では、通常、音源信号の到来方向は未知である。また、到来方向が既知であったとしても、マイクロホンアレイを用いていない場合には、上記の手法は適用できない。

混合過程において、音源やマイクの位置によらず一般的に成り立つ性質として、インパルス応答のスパース性が挙げられる。実際に、スパース性やブロックスパース性を考慮することで、室内のインパルス応答の推定精度を向上させる研究も報告されている [8]–[10]。本研究では、「インパルス応答 $h_{m,n}[\tau]$ はスパースである」という混合過程の先験情報に基づいて混合行列 A_i を推定する。そして、この行列の一般逆行列を考慮しながら分離行列 W_i の更新を行う分離手法を提案する。

3.2 インパルス応答のスパース性を利用する提案法

インパルス応答 $h_{m,n}[\tau]$ をベクトルで

$$\mathbf{h}_{m,n} := (h_{m,n}[0], h_{m,n}[1], \dots, h_{m,n}[T-1])^T \in \mathbb{R}^T$$

と表し、全てのインパルス応答 $\mathbf{h}_{m,n}$ ($m = 1, 2, \dots, M$; $n = 1, 2, \dots, N$) の組からなるテンソルを $\mathcal{H} \in \mathbb{R}^{T \times M \times N}$ と定義する。分離行列 W_i から計算される混合行列を $A_i = W_i^H (W_i W_i^H)^{-1}$ とする。 $i = 2, 3, \dots, \lfloor \frac{L+1}{2} \rfloor$ に対して、 A_i の各成分 $\hat{a}_{i,m,n}$ の複素共役をとることで $\hat{a}_{L-i+2,m,n}$ を定義し、 $\hat{a}_{i,m,n}$ ($i = 1, 2, \dots, L$) を並べて

$$\bar{\mathbf{a}}_{m,n} := (\hat{a}_{1,m,n}, \hat{a}_{2,m,n}, \dots, \hat{a}_{L,m,n})^T \in \mathbb{C}^L$$

を定義する。

本研究では、負対数尤度 $-\log(p_x(\mathcal{X} | \mathcal{W}))$ に加えてインパルス応答のスパース性も考慮した最適化問題

$$\begin{aligned} & \underset{\mathcal{W}, \mathcal{H}, (B), (C)}{\text{minimize}} && -\log(p_x(\mathcal{X} | \mathcal{W})) + \lambda J[\mathcal{F}(\mathcal{W}, \mathcal{H}) + \nu \mathcal{G}(\mathcal{H})] \\ & \text{s.t.} && \forall n \sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L \quad \text{and} \quad \forall n \sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1 \end{aligned} \quad (8)$$

を解く手法を提案する。ここで、 \mathcal{F} は \mathcal{W} と \mathcal{H} の間の整合性を評価する関数であり、 \mathcal{G} はインパルス応答 $\mathbf{h}_{m,n}$ のスパース性を評価する関数である。制約条件 $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$ と $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$ は音源信号と混合過程の間に存在するスケールの任意性を取り除くために追加している。非負値行列 $B := (B_1, B_2, \dots, B_N)$ と $C := (C_1^T, C_2^T, \dots, C_N^T)^T$ は、 \bar{S}_n の確率密度関数 p_n に ILRMA と同様の時変複素ガウス分布を仮定した際に必要となる変数である。また、 $\lambda > 0$ と $\nu > 0$ は各項の重みを調整するパラメータである。提案法では、 \mathcal{W} と \mathcal{H} の更新を交互に行うことで、式 (8) の問題の近似解を求める。 $\mathbf{s}_{i,j}$ の推定値は従来法と同様に、 $\mathbf{y}_{i,j} = W_i \mathbf{x}_{i,j}$ により与えられるものとする。

3.3 分離行列の更新

インパルス応答の推定値 $\mathbf{h}_{m,n}$ から式 (2) により計算される混合行列を \tilde{A}_i ($i = 1, 2, \dots, L$)⁹ と定義し、対応する分離行列を $\tilde{W}_i := (\tilde{A}_i^H \tilde{A}_i)^{-1} \tilde{A}_i^H$ ($i = 1, 2, \dots, L$) とする。まず、 $\mathcal{H} = \mathcal{H}^{(l)}$ と固定し \mathcal{W} を更新する。 \mathcal{W} の更新時には、 \mathcal{W} と \mathcal{H} の間の整合性を評価する関数 \mathcal{F} を

$$\mathcal{F}(\mathcal{W}, \mathcal{H}) := \sum_{i=1}^L \|W_i - \tilde{W}_i\|_F^2 = \sum_{i=1}^L \sum_{n=1}^N \|\mathbf{w}_{i,n} - \tilde{\mathbf{w}}_{i,n}\|_2^2 \quad (9)$$

と定義する。このとき、制約条件 $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$ を無視すれば、 \bar{S}_n の確率密度関数 p_n に複素ラプラス分布、球対称複素ラプラス分布、時変複素ガウス分布のいずれを採用した場合においても、式 (6) から分離行列 W_i ごとに最適化問題

$$\begin{aligned} & \underset{W_i}{\text{minimize}} && \sum_{n=1}^N \left(\mathbf{w}_{i,n}^H U_{i,n}^{(l+1)} \mathbf{w}_{i,n} + \lambda \|\mathbf{w}_{i,n} - \tilde{\mathbf{w}}_{i,n}^{(l)}\|_2^2 \right) \\ & && - \log(\det(W_i W_i^H)) \end{aligned} \quad (10)$$

を解けばよいことが分かる。式 (10) のコスト関数を $Q_i^{(l+1)}$ とすると、 $Q_i^{(l+1)}$ の値を局所的に最小にする W_i において、勾配 $\nabla Q_i^{(l+1)} := \frac{\partial Q_i^{(l+1)}}{\partial \Re[W_i]} + i \frac{\partial Q_i^{(l+1)}}{\partial \Im[W_i]}$ は零行列となるため、

$$\begin{aligned} \nabla Q_i^{(l+1)}(W_i) = 2 & \begin{pmatrix} \mathbf{w}_{i,1}^H U_{i,1}^{(l+1)} \\ \mathbf{w}_{i,2}^H U_{i,2}^{(l+1)} \\ \vdots \\ \mathbf{w}_{i,N}^H U_{i,N}^{(l+1)} \end{pmatrix} + 2\lambda \begin{pmatrix} \mathbf{w}_{i,1}^H - \tilde{\mathbf{w}}_{i,1}^{(l)H} \\ \mathbf{w}_{i,2}^H - \tilde{\mathbf{w}}_{i,2}^{(l)H} \\ \vdots \\ \mathbf{w}_{i,N}^H - \tilde{\mathbf{w}}_{i,N}^{(l)H} \end{pmatrix} \\ & - 2(W_i W_i^H)^{-1} W_i = O_{N \times M} \end{aligned}$$

が成立する。両辺のエルミート転置をとり、2で割れば、

$$\begin{aligned} & \left(\tilde{U}_{i,2}^{(l+1)} \mathbf{w}_{i,1} \quad \tilde{U}_{i,2}^{(l+1)} \mathbf{w}_{i,2} \quad \dots \quad \tilde{U}_{i,N}^{(l+1)} \mathbf{w}_{i,N} \right) \\ & = W_i^H (W_i W_i^H)^{-1} + \lambda \tilde{W}_i^{(l)H} \end{aligned}$$

が成立する。ここで、 $\tilde{U}_{i,n}^{(l+1)} := U_{i,n}^{(l+1)} + \lambda E_M$ であり、 E_M は M 次の単位行列である。左から W_i を掛ければ、

$$\begin{aligned} & W_i \left(\tilde{U}_{i,1}^{(l+1)} \mathbf{w}_{i,1} \quad \tilde{U}_{i,2}^{(l+1)} \mathbf{w}_{i,2} \quad \dots \quad \tilde{U}_{i,N}^{(l+1)} \mathbf{w}_{i,N} \right) \\ & = E_N + \lambda W_i \tilde{W}_i^{(l)H} \end{aligned}$$

が成立するため、 W_i は $p, q = 1, 2, \dots, N$ に対して、

$$\mathbf{w}_{i,p}^H \tilde{U}_{i,q}^{(l+1)} \mathbf{w}_{i,q} = \delta_{p,q} + \lambda \mathbf{w}_{i,p}^H \tilde{\mathbf{w}}_{i,q}^{(l)} \quad (11)$$

を満たす。ここで、 $\delta_{p,q}$ はクロネッカーのデルタである。

⁹ $i = 1, 2, \dots, L$ でなく、 $i = 1, 2, \dots, L$ である点に注意されたい。

式 (11) を満たす停留点を高速かつ正確に求めることは困難であるため、 $n = 1, 2, \dots, N$ に対して

$$\begin{cases} \mathbf{w}_{i,p}^{(l)H} \tilde{U}_{i,n}^{(l+1)} \mathbf{w}_{i,n}^{(l+1)} = \lambda \mathbf{w}_{i,p}^{(l)H} \tilde{\mathbf{w}}_{i,n}^{(l)} & (\forall p \neq n) \\ \mathbf{w}_{i,n}^{(l+1)H} \tilde{U}_{i,n}^{(l+1)} \mathbf{w}_{i,n}^{(l+1)} = 1 + \lambda \mathbf{w}_{i,n}^{(l+1)H} \tilde{\mathbf{w}}_{i,n}^{(l)} \end{cases} \quad (12)$$

を満たすような $W_i^{(l+1)}$ を近似値として求める。式 (12) を満たす $W_i^{(l+1)}$ は、 $M > N$ のとき無数に存在するが¹⁰、例えば、

$$\begin{cases} \mathbf{v}_{i,n}^{(l+1)} = (\tilde{U}_{i,n}^{(l+1)})^{-1} \mathbf{a}_{i,n}^{(l)} \\ \tilde{\mathbf{v}}_{i,n}^{(l+1)} = \lambda (\tilde{U}_{i,n}^{(l+1)})^{-1} \tilde{\mathbf{w}}_{i,n}^{(l)} \\ d_{i,n}^{(l+1)} = \mathbf{v}_{i,n}^{(l+1)H} \tilde{U}_{i,n}^{(l+1)} \mathbf{v}_{i,n}^{(l+1)} \\ \tilde{d}_{i,n}^{(l+1)} = \mathbf{v}_{i,n}^{(l+1)H} \tilde{U}_{i,n}^{(l+1)} \tilde{\mathbf{v}}_{i,n}^{(l+1)} \\ \mathbf{w}_{i,n}^{(l+1)} = \begin{cases} \frac{\mathbf{v}_{i,n}^{(l+1)}}{\sqrt{d_{i,n}^{(l+1)}}} + \tilde{\mathbf{v}}_{i,n}^{(l+1)} & \text{if } \tilde{d}_{i,n}^{(l+1)} = 0, \\ \frac{\tilde{d}_{i,n}^{(l+1)}}{2d_{i,n}^{(l+1)}} \left[-1 + \sqrt{1 + \frac{4d_{i,n}^{(l+1)}}{|\tilde{d}_{i,n}^{(l+1)}|^2}} \right] \mathbf{v}_{i,n}^{(l+1)} + \tilde{\mathbf{v}}_{i,n}^{(l+1)} & \text{otherwise,} \end{cases} \end{cases} \quad (13)$$

と与えられ、式 (13) は式 (7) の一般化となっている [7]。しかし、式 (8) の問題を解く際に条件 $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$ を無視していたため、これを満たすように $\mathcal{W}^{(l+1)}$ を修正する。まず、分離行列 $W_i^{(l+1)}$ から混合行列 $A_i^{(l+1)} = W_i^{(l+1)H} (W_i^{(l+1)} W_i^{(l+1)H})^{-1}$ ($i = 1, 2, \dots, I$) を求める。次に、 $i = 2, 3, \dots, \lfloor \frac{L+1}{2} \rfloor$ に対して、 $A_i^{(l+1)}$ の各成分の複素共役をとることで $A_{L-i+2}^{(l+1)}$ を計算する。 $\gamma_n^{(l+1)} := \sqrt{\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}^{(l+1)}\|_2^2} / L$ ($n = 1, 2, \dots, N$) と定義し、 $\mathbf{a}_{i,n}^{(l+1)}$ ($i = 1, 2, \dots, I$) を $\gamma_n^{(l+1)}$ で割る、すなわち $\mathbf{w}_{i,n}^{(l+1)}$ を $\gamma_n^{(l+1)}$ 倍することで、 $\sum_{m=1}^M \|\bar{\mathbf{a}}_{m,n}\|_2^2 = L$ を満たす分離行列の組 $\mathcal{W}^{(l+1)}$ が求まる。

3.4 インパルス応答の更新

\mathcal{W} の更新後は、 $\mathcal{W} = \mathcal{W}^{(l+1)}$ と固定し \mathcal{H} を更新する。 \mathcal{H} の更新時には、関数 \mathcal{F} を

$$\mathcal{F}(\mathcal{W}, \mathcal{H}) := \sum_{i=1}^L \|A_i - \tilde{A}_i\|_F^2 = \sum_{m=1}^M \sum_{n=1}^N \|\bar{\mathbf{a}}_{m,n} - \Phi \mathbf{h}_{m,n}\|_2^2 \quad (14)$$

と定義¹¹する。ここで、 Φ は式 (2) の離散時間フーリエ変換からのサンプリングを表し、 $\phi_{i,\tau} := \exp(i \frac{2\pi(i-1)\tau}{L})$

¹⁰ $M = N$ のときでも、 $\tilde{d}_{i,n}^{(l+1)} = 0$ の場合には無数に存在する。

¹¹ 同じサイズの行列 X と Y 、及びそれぞれの一般逆行列 X^\dagger と Y^\dagger に対して、通常 $\|X - Y\|_F^2 \neq \|X^\dagger - Y^\dagger\|_F^2$ であるため、式 (9) と式 (14) で、実際には別々の関数を定義していることになる。しかしながら、 $X = Y$ のときには $\|X - Y\|_F^2 = \|X^\dagger - Y^\dagger\|_F^2 = 0$ が成立することから、本研究では式 (9) と式 (14) を \mathcal{W} と \mathcal{H} の間の整合性を評価する関数として用いる。

($i = 1, 2, \dots, L; \tau = 0, 1, \dots, T-1$) を用いて、

$$\begin{cases} \boldsymbol{\phi}_\tau := (\phi_{1,\tau}, \phi_{2,\tau}, \dots, \phi_{L,\tau})^T \in \mathbb{C}^L \\ \Phi := (\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_{T-1}) \in \mathbb{C}^{L \times T} \end{cases}$$

と定義される。インパルス応答のスパース性を評価する関数 \mathcal{G} には、重み付き l_0 ノルム

$$\begin{aligned} \mathcal{G}(\mathcal{H}) &:= \sum_{m=1}^M \sum_{n=1}^N \|\mathbf{h}_{m,n}\|_0^\kappa \\ &:= \sum_{m=1}^M \sum_{n=1}^N \sum_{\tau=0}^{T-1} \kappa[\tau] \Gamma(h_{m,n}[\tau]) \end{aligned} \quad (15)$$

を採用する。ここで、重み $\kappa[\tau] > 0$ ($\tau = 0, 1, \dots, T-1$) の値は、「インパルス応答の振幅値 $|h_{m,n}[\tau]|$ は徐々に減少する傾向がある」ことを考慮して、 τ に関して単調増加するように設計される。2値関数 $\Gamma: \mathbb{R} \rightarrow \{0, 1\}$ は

$$\Gamma(h) := \begin{cases} 0 & \text{if } h = 0, \\ 1 & \text{if } h \neq 0, \end{cases}$$

と定義される。このとき、制約条件 $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$ を無視すれば、インパルス応答 $\mathbf{h}_{m,n}$ ごとに最適化問題

$$\text{minimize}_{\mathbf{h}_{m,n}} \|\bar{\mathbf{a}}_{m,n}^{(l+1)} - \Phi \mathbf{h}_{m,n}\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \quad (16)$$

を解けばよいことが分かる。

式 (16) の問題の解 $\mathbf{h}_{m,n}^{(l+1)}$ に関して、 Φ の列ベクトルの直交性 $\boldsymbol{\phi}_p^H \boldsymbol{\phi}_q = L \delta_{p,q}$ ($p, q = 0, 1, \dots, T-1$) から、

$$\begin{aligned} \mathbf{h}_{m,n}^{(l+1)} &= \text{argmin}_{\mathbf{h}_{m,n}} \|\bar{\mathbf{a}}_{m,n}^{(l+1)} - \Phi \mathbf{h}_{m,n}\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \\ &= \text{argmin}_{\mathbf{h}_{m,n}} \left\| \frac{\Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)}}{\sqrt{L}} - \sqrt{L} \mathbf{h}_{m,n} \right\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \\ &= \text{argmin}_{\mathbf{h}_{m,n}} L \left\| \frac{\Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)}}{L} - \mathbf{h}_{m,n} \right\|_2^2 + \nu \|\mathbf{h}_{m,n}\|_0^\kappa \\ &= \text{argmin}_{\mathbf{h}_{m,n}} \left\| \frac{\Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)}}{L} - \mathbf{h}_{m,n} \right\|_2^2 + \frac{\nu}{L} \|\mathbf{h}_{m,n}\|_0^\kappa \end{aligned}$$

が成立する。分離行列 $W_i^{(l+1)}$ から計算されるインパルス応答 $\tilde{\mathbf{h}}_{m,n}^{(l+1)} := \Phi^H \bar{\mathbf{a}}_{m,n}^{(l+1)} / L$ の第 τ 成分を $\tilde{h}_{m,n}^{(l+1)}[\tau - 1]$ とすれば、 $\mathbf{h}_{m,n}^{(l+1)}$ はハード閾値処理を用いて

$$h_{m,n}^{(l+1)}[\tau] = \begin{cases} \tilde{h}_{m,n}^{(l+1)}[\tau] & \text{if } |\tilde{h}_{m,n}^{(l+1)}[\tau]| \geq \sqrt{\frac{\nu \kappa[\tau]}{L}}, \\ 0 & \text{if } |\tilde{h}_{m,n}^{(l+1)}[\tau]| < \sqrt{\frac{\nu \kappa[\tau]}{L}}, \end{cases} \quad (17)$$

と与えられる。しかし、式 (17) で得られる $\mathcal{H}^{(l+1)}$ は条件 $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$ を無視していたため、これを満たす

ように $\mathbf{H}^{(l+1)}$ を修正する. $\xi_n^{(l+1)} := \sqrt{\sum_{m=1}^M \|\mathbf{h}_{m,n}^{(l+1)}\|_2^2}$ ($n = 1, 2, \dots, N$) と定義し, $\mathbf{h}_{m,n}^{(l+1)}$ ($m = 1, 2, \dots, M$) を $\xi_n^{(l+1)}$ で割ることで, $\sum_{m=1}^M \|\mathbf{h}_{m,n}\|_2^2 = 1$ を満たすインパルス応答の組 $\mathbf{H}^{(l+1)}$ が求まる.

\mathbf{W} と \mathbf{H} の更新を既定の回数繰り返すことで最終的な分離行列 W_i を決定し, $\mathbf{s}_{i,j}$ の推定値 $\mathbf{y}_{i,j} = W_i \mathbf{x}_{i,j}$ を得る. 混合行列 $A_i = W_i^H (W_i W_i^H)^{-1}$ の列ベクトル $\mathbf{a}_{i,n}$ を用いて

$$(\hat{y}_{i,j,1,n}, \hat{y}_{i,j,2,n}, \dots, \hat{y}_{i,j,M,n})^T := \hat{y}_{i,j,n} \mathbf{a}_{i,n} \in \mathbb{C}^M$$

を計算し, $\hat{y}_{i,j,m,n}$ ($i = 1, 2, \dots, I; j = 1, 2, \dots, J$) を逆短時間フーリエ変換 [12] することで, 離散時間信号 $(h_{m,n} * s_n)[t]$ の推定値が分離結果として得られる.

4 数値実験

4.1 実験概要

提案法の有効性を示すために, 音源数 $N = 2$, マイク数 $M = 2$ の場合で数値実験を行う. サンプル周波数 16,000 [Hz] の音源信号として, 無償配布されている音楽信号 (ギター [13] 及びピアノ [14]) を用いた実験と, 音源分離コンペティションで利用される音声信号 (二人の男性の声 [15]) を用いた実験を行った. 図 1 のような音源とマイクの位置関係を想定し, Aachen Impulse Response Database [16] 内にあるインパルス応答を参考に, $\mathbf{h}_{1,1} = \mathbf{h}_{2,2}$ を図 2 のように, $\mathbf{h}_{1,2}$ を図 3 のように, $\mathbf{h}_{2,1}$ を図 4 のように定義した¹². 短時間フーリエ変換を行う際は, フレーム長 $L = 8192$, フレームシフト量 $\eta = 2048$ とし, ハミング窓 $\psi[\tau] := 0.54 - 0.46 \cos(2\pi\tau/L)$ を用いた.

比較する従来法を ILRMA [3] とし, 提案法における \bar{S}_n の確率密度関数 p_n にも ILRMA と同様の時変複素ガウス分布を採用した. 基底行列 B_n の列ベクトル数は, 音楽信号に対する実験では $K_1 = K_2 = 30$ とし, 音声信号に対する実験では $K_1 = K_2 = 5$ とした. 提案法のパラメータは $T = 4096$, $\lambda = 0.075$, $\nu = L = 8192$ とし, 式 (15) で $\kappa[\tau] := -\log_{10}(1 - \exp(-432/(\tau+1)))$ とした. 提案法と ILRMA で共通の初期値から 100 回の更新を行い, 分離行列の初期値を $W_i^{(0)} = E_2$ ($i = 1, 2, \dots, I$), インパルス応答の初期値を $\mathbf{h}_{m,n}^{(0)} = \mathbf{0}$ ($m, n = 1, 2$) とし, 基底行列と係数行列の初期値 $B_n^{(0)}$ と $C_n^{(0)}$ の各成分は $[0, 1]$ の一様分布からランダムに生成した. 音源分離の評価指標には, 文献 [17] で定義されている source-to-distortion ratio (SDR), source-to-interferences ratio (SIR), sources-to-artifacts ratio (SAR) の $x_m[t]$ からの改善量を用い, 10 回の平均値で比較を行った.

¹²各インパルス応答は 2 [m] 分の遅延を削除して定義されている.

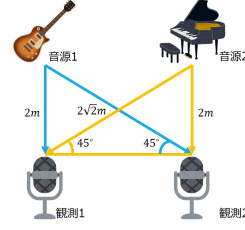


図 1: 音源とマイクの位置

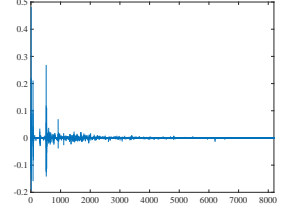


図 2: $\mathbf{h}_{1,1} = \mathbf{h}_{2,2}$

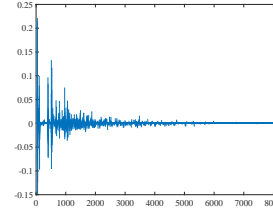


図 3: $\mathbf{h}_{1,2}$

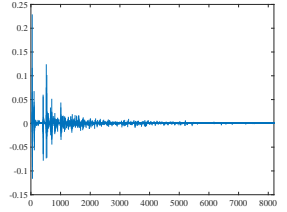


図 4: $\mathbf{h}_{2,1}$

4.2 実験結果

まず音楽信号に対する実験結果を述べる. 図 5 に提案法と従来法 (ILRMA) の音源分離結果を示す. 図 5 から, 提案法による分離結果は, 総合的な分離性能を表す SDR 値で平均約 0.64 [dB], 非目的音の除去性能を表す SIR 値で約 1.18 [dB], 音源分離時に生じた歪みの少なさを表す SAR 値で約 0.39 [dB], 従来法の結果を上回った. また, どちらの手法も 40 回程度の更新回数で SDR 値がほぼ収束していた. 分離結果を実際に聞いてみると, 従来法よりも提案法を用いた場合の方が, 音源間の干渉が少ない, 明瞭な音を得ることができた.

次に音声信号に対する実験結果を述べる. 図 6 に提案法と従来法の音源分離結果を示す. 提案法の分離結果は, SDR 値で約 0.92 [dB], SIR 値で約 0.95 [dB], SAR 値で約 0.84 [dB], 従来法の結果を上回った. また, どちらの手法も 30 回程度の更新回数で SDR 値が収束していた. これらの結果から, 音声信号に対しても提案法によって高品質な分離信号が得られることが確認された.

5 おわりに

本研究では, 過決定条件下のブラインド音源分離問題に対して, インパルス応答のスパース性を考慮しながら分離行列を求める手法を提案した. 提案する最適化問題では, 分離行列とインパルス応答の推定を同時に行い, スパースなインパルス応答から得られる混合行列の一般逆行列と分離行列の推定値を近づけることで, 高品質な分離信号を得ることができる. 公開されている音源信号とインパルス応答を用いた数値実験によって, 提案法が従来法に比べて明瞭な分離信号を生成することを示した.

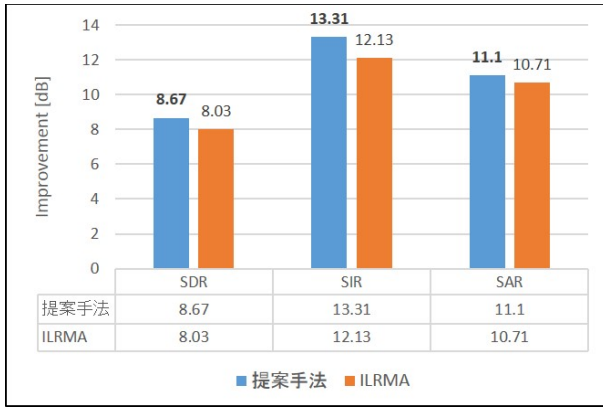


図 5: 音楽信号に対する提案法と ILRMA の分離結果

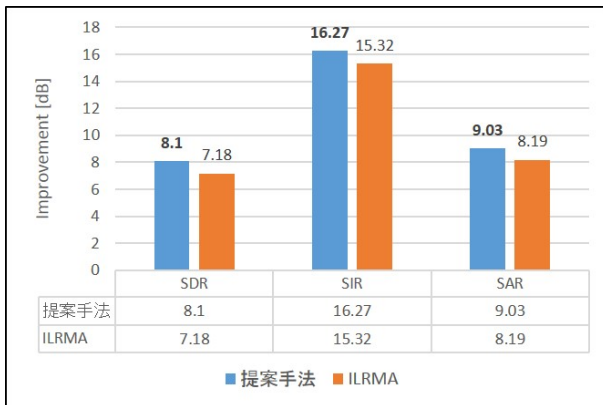


図 6: 音声信号に対する提案法と ILRMA の分離結果

参考文献

- [1] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol. 22, pp. 21–34, 1998.
- [2] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, “Blind source separation exploiting higher-order frequency dependencies,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, 2007.
- [3] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, “Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, 2016.
- [4] 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. Audio, Speech, Lang. Process.*, vol. 12, no. 5, pp. 530–538, 2004.
- [5] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, “Blind source separation based on a fast-convergence algorithm combining ICA and beamforming,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 2, pp. 666–678, 2006.
- [6] L. C. Parra and C. V. Alvino, “Geometric source separation: Merging convolutive source separation with geometric beamforming,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 10, no. 6, pp. 352–362, 2002.
- [7] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, “Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis,” in *Proc. IEEE ICASSP 2018*, Calgary, April, 2018 (to appear).
- [8] B. L. Sturm and G. Defrance, “Detection and estimation of arrivals in room impulse responses by greedy sparse approximation,” in *Proc. EUSIPCO 2010*, pp. 1934–1938, 2010.
- [9] M. Crocco and A. D. Bue, “Room impulse response estimation by iterative weighted L_1 -norm,” in *Proc. EUSIPCO 2015*, pp. 1895–1899, 2015.
- [10] G. Vairetti, E. D. Sena, and M. Catrysse, “A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, pp. 1547–1561, vol. 25, no. 7, 2017.
- [11] N. Ono and S. Miyabe, “Auxiliary-function-based independent component analysis for super-Gaussian sources,” in *Proc. LVA/ICA*, pp. 165–172, 2010.
- [12] D. W. Griffin and J. S. Lim, “Signal estimation from modified short-time Fourier transform,” *IEEE Trans. Acoust. Speech Signal Process.*, vol. 32, no. 2, pp. 236–243, 1984.
- [13] Musicpalette, BGM.012 「せつないギターアルペジオ」, available online at: <https://www.youtube.com/watch?v=GyYUKuJ3JL0&t=3s>
- [14] Musicpalette, BGM.285 「ゴージャス」, available online at: <https://www.youtube.com/watch?v=FVv3v6CH2I8>
- [15] S. Araki, A. Ozerov, V. Gowreesunker, H. Sawada, F. Theis, G. Nolte, D. Lutter, and N. Q. K. Duong, “The 2010 signal separation evaluation campaign (SiSEC2010): Audio source separation,” in *Proc. LVA/ICA*, pp. 114–122, 2010.
- [16] *Aachen Impulse Response Database*, available online at: <https://www.iks.rwth-aachen.de/en/research/tools-downloads/databases/aachen-impulse-response-database/>
- [17] 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, 2006.