

白色化の影響を考慮したスパース独立ベクトル分析*

○矢田部浩平 (早大), 北村大地 (香川高専)

1 まえがき

N 個の音源が M 本のマイクロホンによって観測されている状況を, 時間周波数領域において $\mathbf{x}[t, f] \approx A[f]\mathbf{s}[t, f]$ と近似し, 分離フィルタ $W[f]$ を用いて

$$W[f]\mathbf{x}[t, f] \approx W[f]A[f]\mathbf{s}[t, f] = \mathbf{s}[t, f]$$

と分離する問題を考える. この問題に対する一般的な戦略は, 音源信号に対する何らかの先験情報を関数 \mathcal{P} として導入し, 分離フィルタに関する最適化問題

$$\text{Minimize}_{\{W[f]\}_{f=1}^F} \mathcal{P}(W[f]\mathbf{x}[t, f]) - \sum_{f=1}^F \log |\det(W[f])|$$

を解くことであり, 独立成分分析 (ICA) や独立ベクトル分析 (IVA) など様々な手法が提案されている.

我々はこの問題に対し, 近接分離最適化を適用したアルゴリズムを提案している [1, 2]. 提案法の特徴は, ペナルティ関数 \mathcal{P} に関する最小化を, 近接作用素と呼ばれる「より簡単な部分問題」に置き換えることで, 多くの音源モデルを単一のアルゴリズムで統一的に扱える点にある. さらに, アルゴリズムの解釈を緩和することで, 音源モデルを「一般の時間周波数マスクング手法」で間接的に与える拡張手法も提案している [3, 4]. **Algorithm 1** に示すように, 音源を強調するマスク生成関数 $\mathcal{M}_\theta(\cdot)$ さえ定義すれば, それを代入するのみで新たな音源分離手法を実現でき, 定式化や解法に囚われない提案が可能である.

前回の音響学会 [3] では, 一般の時間周波数マスクングに基づく音源分離アルゴリズムの他に, そのアルゴリズムを利用した IVA も合わせて提案するつもりであったが, 結果として未発表のまま現在に至っている. そこで本稿では, 提案アルゴリズムを用いて IVA を拡張した, スパース IVA を改めて提案する.

2 時間周波数マスクングに基づく音源分離

観測データを行列 X で表現し, 分離フィルタをベクトル化して \mathbf{w} とすると, 音源分離問題は $\mathcal{P}(X\mathbf{w}) - \log |\det(\text{mat}(\mathbf{w})[f])|$ の最小化なので, それらの和に対して近接分離アルゴリズムを適用し [1, 2], \mathcal{P} の近接作用素を時間周波数マスク $\mathcal{M}_\theta(\cdot)$ で置き換えることで **Algorithm 1** が得られる [3, 4]. ここで $\mathcal{M}_\theta(\cdot)$ は, 音源モデルを事前分布として Gauss 雑音を除去する MAP 推定に対応しており [3, 4], そのような雑

Algorithm 1 Masking-based PDS-BSS

```

1: Input:  $X, \mathbf{w}^{[1]}, \mathbf{y}^{[1]}, \mu_1, \mu_2, \alpha$ 
2: Output:  $\mathbf{w}^{[K+1]}$ 
3: for  $k = 1, \dots, K$  do
4:    $\tilde{\mathbf{w}} = \text{prox}_{\mu_1 \mathcal{I}}[\mathbf{w}^{[k]} - \mu_1 \mu_2 X^H \mathbf{y}^{[k]}]$ 
5:    $\mathbf{z} = \mathbf{y}^{[k]} + X(2\tilde{\mathbf{w}} - \mathbf{w}^{[k]})$ 
6:    $\tilde{\mathbf{y}} = \mathbf{z} - \mathcal{M}_\theta(\mathbf{z}) \odot \mathbf{z}$ 
7:    $\mathbf{y}^{[k+1]} = \alpha \tilde{\mathbf{y}} + (1 - \alpha) \mathbf{y}^{[k]}$ 
8:    $\mathbf{w}^{[k+1]} = \alpha \tilde{\mathbf{w}} + (1 - \alpha) \mathbf{w}^{[k]}$ 
9: end for

```

音除去効果があれば, どんなマスクでも音源分離アルゴリズムを実現可能である. よって, 新手法の開発をマスク設計問題として捉えなおすことができる.

3 スパース独立ベクトル分析

IVA は, 各窓をグループとするグループスパース誘導関数を \mathcal{P} として選んだ場合に対応する. 例えば,

$$(\mathcal{M}_\lambda(\mathbf{z}))_m[t, f] = (1 - \lambda / (\sum_{f=1}^F |z_m[t, f]|^2))^{\frac{1}{2}}$$

のような閾値処理が, $\ell_{2,1}$ ノルムを選ぶとマスクとして導かれ, これを提案アルゴリズムに代入すれば, 球対称 Laplace 分布に基づく IVA が実現される. このマスクを拡張すれば, 音源毎に独立なマスクを考える限り IVA の拡張を与える (音源毎に独立でない場合は, もはや独立性基準と解釈はできないが, 独立性が必ずしも適切な仮定とは限らないので, 考える価値はある). 任意のマスクに対応する関数 \mathcal{P} は陽に書けないので, 特別な場合を除いて音源の事前分布も閉形式で表せないが, 提案手法を用いれば, そのような数式上の都合に制約されることなく IVA を拡張できる.

3.1 白色化が音源分離に与える影響

多くの音源分離アルゴリズムで, 安定性や収束速度の意味で, 白色化が有用な前処理として利用されている. 周波数領域 ICA のように, 各周波数を独立に取り扱う場合は, 各周波数を個別に白色化して問題ないが, IVA のように周波数間の共起性を考慮する場合は, 白色化の副作用に注意する必要がある. 図-1 に, 白色化前後の音声のスペクトログラムを示す. 各周波数で独立に白色化を行った結果, 元からエネルギーの小さい周波数帯域が持ち上がり, ノイズが支配的な帯域が発生している. これを周波数方向に足し合わせたエネルギーが下段にあるが, 白色化の前後で特徴

*Sparse independent vector analysis considering side effect of whitening. By Kohei YATABE (Waseda University), Daichi KITAMURA (National Institute of Technology, Kagawa College).

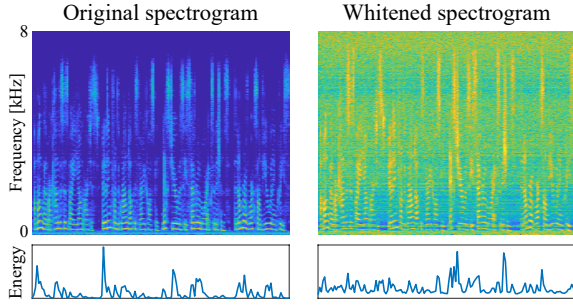


図-1 白色化前後のスペクトログラムと窓毎のエネルギー

が大きく異なっている。球対称 Laplace 分布は、この下段の図がスパースであることを仮定しているの、白色化がその前提を崩し得ることを示唆している¹。

3.2 白色化の影響を考慮した独立ベクトル分析

白色化の影響を考慮した上で、より強くスパースに誘導するマスクを提案することで、IVA を拡張する。まず、白色化によって崩されたスパース構造を復元するために、重み $\Theta_\eta[\cdot]$ を導入する。ここでは、周波数毎のスパース性を正規化 ℓ_1 ノルムで計り、その逆数

$$(\Theta_\eta[\mathbf{x}])_f = \Upsilon_\eta \left[\frac{(\sum_{m=1}^M \sum_{t=1}^T |x_m[t, f]|^2)^{\frac{1}{2}}}{\sum_{m=1}^M \sum_{t=1}^T |x_m[t, f]|} \right]$$

を利用する。 $\Upsilon_\eta[\cdot]$ は ℓ_1 正規化を表し、 $\eta \geq 0$ により

$$\Upsilon_\eta[\xi] = \xi_\eta / (\|\xi_\eta\|_1 / F), \quad \xi_\eta = (\xi - \eta)_+$$

のように閾値処理を行う。この重み $\Theta_\eta[\cdot]$ はスパースでない周波数帯域で小さな値やゼロとなり、ノイズが支配的な帯域を無視する効果がある。

さらに、グループスパース項とは別に、要素毎のスパース誘導項も足すことを想定し、soft thresholding のバイアスを軽減した firm thresholding を用いて

$$(\mathcal{M}_{\text{sparseIVA}}^{\mathbf{x}, \eta, \lambda, \kappa}(\mathbf{z}))_m[t, f] = \zeta_m^{\mathbf{z}, \kappa}[t, f] \times \Xi_\kappa \left[\left(1 - \frac{\lambda_1}{(\sum_{f=1}^F (\Theta_\eta[\mathbf{x}])_f |\zeta_m^{\mathbf{z}, \kappa}[t, f] z_m[t, f]|^2)^{\frac{1}{2}}} \right)_+ \right]$$

によりマスクを定義する。ただし $\lambda = [\lambda_1, \lambda_2]$ は閾値、

$$(\Xi_\kappa[\mathbf{z}])_m[t, f] = (\kappa z_m[t, f] / \max_{m, t, f} \{z_m[t, f]\})_-$$

は $\kappa \geq 1$ によるバイアス低減処理、 $(\cdot)_- = \min\{1, \cdot\}$ 、

$$\zeta_m^{\mathbf{z}, \kappa}[t, f] = \Xi_\kappa[(1 - \lambda_2 / |z_m[t, f]|)_+]$$

は要素毎の firm thresholding を表す。このマスクを **Algorithm 1** に代入して得られるアルゴリズムを、

¹白色化の有無はアルゴリズムのみでなく、問題自体の性質、特に $\log \det$ 項の性質も変えるので、白色化をしなければ副作用を解決可能だと直ちに結論することはできない。白色化の影響は、 $\log \det$ 項と関数 \mathcal{P} と観測データ X の関係で決まるので、本質的にどのような問題を解いているのかは議論する必要がある。

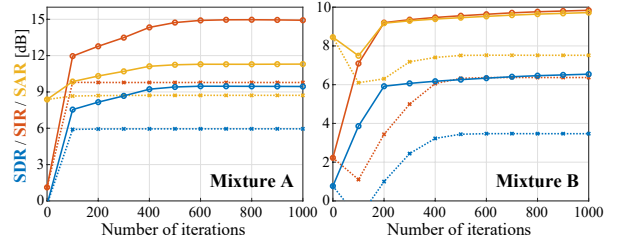


図-2 従来法 (点線) と提案法 (実線) の反復毎の評価

表-1 図-2 右端のスコアおよび反復毎の計算時間

	Mixture A			Mixture B			Time [ms]
	SDR	SIR	SAR	SDR	SIR	SAR	
IVA	6.0	9.8	8.7	3.4	6.3	7.5	55.2
Prop.	9.5	14.9	11.3	6.5	9.8	9.7	67.1
差	3.5	5.1	2.6	3.1	3.5	2.2	11.9
比	1.6 x	1.5 x	1.3 x	1.9 x	1.6 x	1.3 x	1.2 x

スパース IVA として提案する。これは、Laplace IVA に重み $\Theta_\eta[\cdot]$ を導入し、 $(\ell_{2,1}$ ノルムの代わりに) 非凸に拡張した「 $\ell_{2,1}$ ノルムと ℓ_1 ノルムの和」をスパース誘導項 \mathcal{P} として用いた場合に対応する。

4 数値実験

提案したスパース IVA の可能性を示すために、文献 [1, 2] と同様の実験を行った。音源には SiSEC の一部である UND タスク内の dev1 に含まれる liverec の中で、残響時間 130 ms の環境でマイクロホン間隔 5 cm で収録された女声の音源を用い、 -50° と 45° から到来した音源の混合を **Mixture A**、 -10° と 15° からの混合を **Mixture B** とした。128 ms の Hann 窓をずらし幅 64 ms で用い、各パラメータは $\mu_1 = 1$ 、 $\mu_2 = 1$ 、 $\alpha = 1.75$ 、 $\lambda_1 = 2$ 、 $\lambda_2 = 0.01$ 、 $\kappa = 1.1$ 、 $\eta = 0.5$ とした。また、全ての $W[f]$ は単位行列で初期化し、 \mathbf{y} の初期値はゼロベクトルとした。

反復毎の SDR · SIR · SAR を図-2 に示し、その右端での値と 1 反復あたりの計算時間を表-1 に示す。提案したスパース IVA は、元にした従来の Laplace IVA と比べて、SDR が 3.3 dB 程度改善している。一方、(MATLAB 2017a と Core i5-7200U で) 反復あたりの計算時間は 1.2 倍しかかかっておらず、元々 IVA の演算量が低いことを思い返せば、提案法は演算量を低く抑えたまま性能の改善ができていと言える。

参考文献

- [1] 矢田部浩平, 北村大地, “近接分離最適化によるブラインド音源分離,” 日本音響学会講演論文集, pp. 431–434 (2018).
- [2] K. Yatabe and D. Kitamura, “Determined blind source separation via proximal splitting algorithm,” *IEEE Int. Conf. Acoust., Speech, Signal Process.*, pp. 776–780 (2018).
- [3] 矢田部浩平, 北村大地, “一般の時間周波数マスキングに基づく独立ベクトル分析,” 日本音響学会講演論文集, pp. 219–220 (2018).
- [4] K. Yatabe and D. Kitamura, “Time-frequency-masking-based determined BSS with application to Sparse IVA,” — (submitted).