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*

○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

$$W[f]\mathbf{x}[t,f] \approx W[f]A[f]\mathbf{s}[t,f] = \mathbf{s}[t,f] \quad (1)$$

と分離する問題を考える．すなわち，データのみからより良い $W[f]$（$A[f]$ の左逆行列）を求める方法について考える．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ペナルティ関数 $\mathcal{P}$ として導入し，分離フィルタに関する最適化問題

$$\underset{\{W[f]\}_{f=1}^{F}}{\text{Minimize}} \ \mathcal{P}(W[f]\mathbf{x}[t,f]) - \sum_{f=1}^{F} \log|\det(W[f])| \quad (2)$$

として定式化して解くことであり，独立成分分析や独立ベクトル分析など，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

$$\mathrm{prox}_{\mu\mathcal{P}}[\mathbf{y}] = \arg\min_{\mathbf{z}} \left[\mathcal{P}(\mathbf{z}) + \frac{1}{2\mu}\|\mathbf{y}-\mathbf{z}\|_2^2 \right] \quad (3)$$

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．すなわち，音源モデルを変えた際のアルゴリズム導出の手間を減らすことができ，新たな音源モデルの探求がより容易になったと考えている．一方で，音源を強調するような適当な手続きを考えた場合に，その手続きをペナルティ関数の形で書き下すことが必ずしも単純とは限らず，近接作用素を考えること自体が簡単ではないことがあり得る．

そこで本稿では，提案アルゴリズムの解釈を緩めることで，従来の音源モデルを拡張する新たなコンセプトを提案する [3]．一般の時間周波数マスキング手法を代入するのみで，新たな音源分離手法を実現でき，より良い音源分離モデルの発見が期待される．

2 近接分離最適化に基づく音源分離

観測データを行列 X として表現し，分離フィルタをベクトル化して $\mathbf{w}$ とすると，式 (2) は $-\log|\det(\cdot)|$ と $\mathcal{P}(X\mathbf{w})$ の最小化と捉えることができる（文献 [1, 2] 参照）．$-\log|\det(\cdot)|$ は各分離手法において共通なので，$\mathcal{P}\circ X$ に対処するのみで問題 (2) を解くことができれば，$\mathcal{P}$ を変える度に新たなアルゴリズムを考える手間が減り，より幅広い $\mathcal{P}$ を気軽に試すことができるはずである．しかし，X と $\mathcal{P}$ の合成関数に対処するのは一般に容易ではなく，特別な取り扱いが必要になることが多いので，その手間を省くために，近接分離最適化を用いて X と $\mathcal{P}$ を分離したアルゴリズムを提案した [2]．その結果，$\mathcal{P}$ の最小化のみを考えればよくなり，$\mathcal{P}$ の近接作用素 (3) が解析的に解ければ効率的なアルゴリズムが直ちに実現できるようになった．また，解析的でなくとも，式 (3) を反復法で解くことで問題 (2) に対処できるので，複雑な $\mathcal{P}$ を手軽に試すことができるようになったと言える．

しかし，簡単化されているとはいえ，式 (3) 自体も最適化問題なので，それを解く労力は $\mathcal{P}$ に依存して決まるし，そもそも $\mathcal{P}$ を何かしら定義しなければいけないという煩わしさもある．もし仮に，ペナルティ関数 $\mathcal{P}$ やそれに伴う最適化問題を考えずとも，音源の事前情報に基づく適当な手続きを許容するアルゴリズムができれば，非常に多様な音源分離アルゴリズムを気軽に実現することができるはずである．

3 時間周波数マスキングに基づく音源分離

最適化を考える煩わしさを排除するために，提案アルゴリズムの一部を時間周波数マスキングとして再解釈し，ヒューリスティックな拡張を行う．

3.1 スパース誘導項の近接作用素と閾値処理

多くの有用なペナルティ関数 $\mathcal{P}$ について，近接作用素 (3) が解析的に与えられることが知られている．例えば，Laplace 分布に基づく独立成分分析を考えれば，$\mathcal{P}$ は ℓ_1 ノルムとなり，その近接作用素は

$$\mathrm{prox}_{\mu\|\cdot\|_1}[\mathbf{y}] = (1-\mu/|\mathbf{y}|)_+ \odot \mathbf{y} \quad (4)$$

となる．ただし，$(\cdot)_+ = \max\{\cdot, 0\}$，$\odot$ は要素毎の積を表し，絶対値と商も要素毎にスカラーの意味で作用するとする．これは，soft-thresholding として知られる閾値処理の一種である．球対称 Laplace 分布に基づく独立ベクトル分析で現れる $\ell_{2,1}$ 混合ノルムや，スパース最適化で現れる ℓ_0 擬ノルムについても

$$\mathrm{prox}_{\mu\|\cdot\|_{2,1}}[\mathbf{y}] = (1-\mu/\Sigma_{\mathcal{G}}(|\mathbf{y}|))_+ \odot \mathbf{y} \quad (5)$$

$$\mathrm{prox}_{\mu\|\cdot\|_0}[\mathbf{y}] = \chi_{\geq\sqrt{2\mu}}(|\mathbf{y}|) \odot \mathbf{y} \quad (6)$$

*Independent vector analysis based on general time-frequency masking. By Kohei YATABE (Waseda University) and Daichi KITAMURA (National Institute of Technology, Kagawa College).

のように，閾値処理として解析的に近接作用素を与えることができる．ただし，$\Sigma_{\mathcal{G}}$ はグループごとに ℓ_2 ノルムを計算した上でベクトルを拡張する作用素，$\chi_{\geq\alpha}$ は α 以上の要素を 1 で置き換え，残りを 0 で置き換える作用素を表す．このように，ℓ_p 準ノルムや $\log(|\cdot|+\varepsilon)$ なども含め，実用されている多くの $\mathcal{P}$ について，その近接作用素は閾値処理の形式となる．

3.2 閾値処理の時間周波数マスキングとしての解釈

上述の近接作用素を眺めると，全て要素毎の積になっていることがわかる．すなわち，$\mathcal{M}_\theta(\cdot)$ を各要素 0 から 1 の間の実数で置き換える操作として，

$$\mathcal{M}_\theta(\mathbf{y}) \odot \mathbf{y} \tag{7}$$

のような形式で式 (4)–(6) を書くことができる．これは，$\mathbf{y}$ がスペクトログラムであれば，時間周波数マスキングと呼ばれ，パラメータ θ と入力に依存して生成されたソフトマスクをかける操作となっている（式 (6) の場合はバイナリマスク）．実際，音源分離問題 (2) ではチャンネル数分のスペクトログラムを扱うので，上に挙げた独立成分分析や独立ベクトル分析に対して，文献 [2] の近接分離アルゴリズムは，反復毎に時間周波数マスクをかけるアルゴリズムになる．

3.3 一般の時間周波数マスキングに基づく音源分離

音源分離に現れる多くのペナルティ関数の近接作用素が時間周波数マスキングになるということは，逆に，一般の時間周波数マスキングを近接作用素の代わりに用いても，何らかの音源分離が可能であると考えられる．そこで，アルゴリズムに現れる近接作用素を，時間周波数マスキングに置き換えるヒューリスティックを提案する．すなわち，$\mathrm{prox}_{\mathcal{P}/\mu 2}[\mathbf{z}]$ を $\mathcal{M}_\theta(\mathbf{z}) \odot \mathbf{z}$ に置き換えた **Algorithm 1** を提案する．ここで，$\mathcal{M}_\theta(\mathbf{z})$ は，$\mathbf{z}$ や，（例えばマスクの効果を調節する）何らかのパラメータ θ に依存して，各要素 0 から 1 の間の実数で構成された時間周波数マスクを返す関数である（その他の記号や，複数のマスク生成関数 $\mathcal{M}$ を用いる拡張については [1, 2] 参照）．

本稿で提案するヒューリスティックな拡張は，見た目としては些細な変更だが，アルゴリズムが扱える手続きの意味では大きな変更であると言える．なぜなら，Algorithm 1 は，マスク生成関数 $\mathcal{M}$ さえ陽に与えれば実行可能であり，その際に必ずしも対応するペナルティ関数 $\mathcal{P}$ が陽に書き下せる必要はない．例えば，単チャンネル音源強調として提案されている手法をそのまま Algorithm 1 の 6 行目に代入することで，何らかの音源分離アルゴリズムになるが，対応するペナルティ関数が明らかでない場合も多い．本稿の拡張では，そのように最適化問題として表現しづらい一般の時間周波数マスキング手法を用いて，音源分離を実行できるようになっている．

Algorithm 1 Masking-based PDS-BSS

1: **Input:** X, $\mathbf{w}^{[1]}$, $\mathbf{y}^{[1]}$, μ_1, μ_2, α
2: **Output:** $\mathbf{w}^{[K+1]}$
3: **for** $k = 1, \ldots, K$ **do**
4: $\quad \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1 \mathcal{I}}[\, \mathbf{w}^{[k]} - \mu_1 \mu_2 X^H \mathbf{y}^{[k]} \,]$
5: $\quad \mathbf{z} = \mathbf{y}^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
6: $\quad \widetilde{\mathbf{y}} = \mathbf{z} - \mathcal{M}_\theta(\mathbf{z}) \odot \mathbf{z}$
7: $\quad \mathbf{y}^{[k+1]} = \alpha\widetilde{\mathbf{y}} + (1-\alpha)\mathbf{y}^{[k]}$
8: $\quad \mathbf{w}^{[k+1]} = \alpha\widetilde{\mathbf{w}} + (1-\alpha)\mathbf{w}^{[k]}$
9: **end for**

3.4 近接作用素の MAP 推定としての解釈

本稿で提案するヒューリスティックは，近接作用素を MAP 推定として解釈することで正当化することができる．式 (3) の近接作用素は，二乗誤差によってデータ忠実性を計りつつ，$\mathcal{P}$ を最小化している．これは，$C\exp(-\mathcal{P}(\cdot))$ を事前分布として（C は正規化定数），Gauss ノイズを除去する MAP 推定

$$\mathbf{y}_{\mathrm{MAP}} = \arg\max_{\mathbf{z}} \Big[\exp\big(\tfrac{-1}{2\mu}\|\mathbf{y}-\mathbf{z}\|_2^2\big) \exp\big(-\mathcal{P}(\mathbf{z})\big) \Big] \tag{8}$$

と見ることができる．すなわち，提案アルゴリズムは，音源の分布を $C\exp(-\mathcal{P}(\cdot))$ で与えた場合の音源分離問題を，同じ事前分布を用いたノイズ除去問題に置き換えて逐次解いていると見ることができる．従って，分離対象の音源を強調するような時間周波数マスク生成関数 $\mathcal{M}$ を提案法に代入すれば，そのマスクが仮定している音源分布を用いた音源分離アルゴリズムとなる．このとき，事前分布を陽に書き下せる必要はなく，$\mathcal{M}$ が音源を強調するマスクにさえなっていれば，対応する何らかの音源分離が実現可能である．

4 むすび

音源分離のための新たなコンセプトとして，一般の時間周波数マスキングを用いた主双対分離アルゴリズムを提案した [3]．提案手法では，音源を強調するマスク生成関数 $\mathcal{M}$ を定義すれば，それを代入するのみで音源分離を行えるので，簡単である．ただし，得られたアルゴリズムが適切に動作するかは，$\mathcal{M}$ の性質に依存するので，実験的に確認する必要はある．

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] K. Yatabe and D. Kitamura, “Determined BSS based on general time-frequency masking and its application to sparse IVA” (submitting).