Projection Back を距離射影で扱う制約付き優決定ブラインド音源分離 *

◎ 松本和樹 (早大), 山田宏樹, 矢田部浩平 (農工大)

1 はじめに

分離行列推定に基づく優決定ブラインド音源分離（BSS）では，時間周波数領域における観測信号に対し分離行列を適用することで分離音を得られる．分離行列は，スパース性など音源の持つ性質をモデル化し，対応する最小化問題を解くことで得られる．このとき，多くの BSS 手法は三段階の手続きをとる（図–1，3-staged）：(i) 白色化，(ii) 分離行列の最適化，(iii) Projection Back（PB）．ただし，白色化は最適化を安定化させる前処理，PB は振幅や位相の不定性を解消し，周波数特性の歪みを除去する後処理である [1]．最適化には補助関数法 [2–4] や，交互方向乗数法（ADMM）などの近接分離法 [5–7] が用いられる．

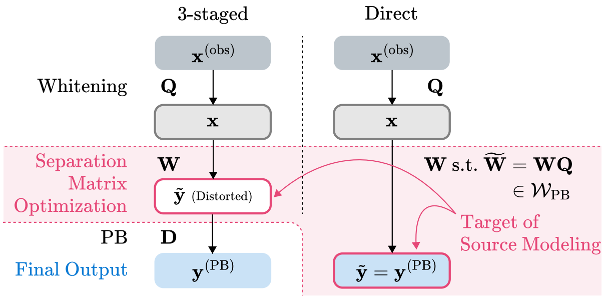
高性能な音源分離手法の開発は，音源モデルの緻密化により実現できる．このとき，前述した三段階の手続きでは，最適化の段階における暫定的な分離音の持つ歪みが問題となる場合がある．例えば，音源モデルに学習済みの DNN を活用する手法 [3, 7] において，分離音の歪みは学習データとのミスマッチを生み，DNN の挙動を不安定にしてしまう．この問題は，最適化の段階で PB 後の分離音，すなわち，周波数特性の歪みを解消した後の分離音を扱うことで対処できる．補助関数法に基づく手法では，最適化の各反復でヒューリスティックに PB を施す処理が用いられている [3, 4]．しかし，後の実験で示すように，近接分離法 [5–7] に対し同様の処理を取り入れると最適化の安定性が損なわれてしまう．

本稿では，近接分離法による分離行列推定において，最適化の段階で PB 後の分離音を扱うための枠組み（図–1，Direct）を提案する．提案法は，先行研究 [8, 9] と同様に PB 済みの分離行列の持つ性質に着目する．特に我々は，PB 制約集合 $\mathcal{W}_{\mathrm{PB}}$ 及びそこへの距離を利用した凸な正則化項を導入することで，最適化後の分離音を PB 後の分離音と一致させる．実験の結果，PB 制約や正則化が最適化の段階で得られる分離音の歪みを解消することが示された．

2 優決定 BSS

2.1 分離行列による音源分離

優決定条件下では，分離行列の適用により音源を分離できる [2]．時間周波数領域における N 個の音源を $\mathbf{s}[f,t] = [s_1[f,t], \cdots, s_N[f,t]]^{\mathsf{T}} \in \mathbb{C}^N$，



図–1 通常の三段階の手続きによる音源分離（3-staged）と，本稿で提案する最終出力の直接的な最適化（Direct）の比較．3-staged では，最適化後の分離音 $\tilde{\mathbf{y}}$ の持つ歪みに PB で対処する．一方，本稿で扱う Direct では，最適化後の分離行列 $\widetilde{\mathbf{W}}$（白色化行列 $\mathbf{Q}$ と分離行列 $\mathbf{W}$ の積）を PB 制約集合 $\mathcal{W}_{\mathrm{PB}}$ の元に制約することで，最適化後の分離音 $\tilde{\mathbf{y}}$ と PB 後の分離音 $\mathbf{y}^{(\mathrm{PB})}$ を一致させ，最適化の段階で歪みの解消された分離音を扱えるようにする．

M ($M \geq N$) チャネルの観測信号を $\mathbf{x}^{(\mathrm{obs})}[f,t] = [x_1^{(\mathrm{obs})}[f,t], \cdots, x_M^{(\mathrm{obs})}[f,t]]^{\mathsf{T}} \in \mathbb{C}^M$ とする．ただし，$1 \leq t \leq T$ 及び $1 \leq f \leq F$ は時間及び周波数のインデックスであり，$(\cdot)^{\mathsf{T}}$ は転置を表す．時間領域の畳み込み混合は，混合行列 $\mathbf{H}[f] = (h_{ij}[f]) \in \mathbb{C}^{M \times N}$ で

$$\mathbf{x}^{(\mathrm{obs})}[f,t] \approx \mathbf{H}[f]\,\mathbf{s}[f,t] \tag{1}$$

と近似できる．このとき，混合行列 $\mathbf{H}[f]$ の左逆行列に対応する分離行列 $\widetilde{\mathbf{W}}[f] = (\tilde{w}_{ij}[f]) \in \mathbb{C}^{N \times M}$ を用い，分離音 $\tilde{\mathbf{y}}[f,t] = \left[\tilde{y}_1[f,t], \cdots, \tilde{y}_N[f,t]\right]^{\mathsf{T}} \in \mathbb{C}^N$ は

$$\tilde{\mathbf{y}}[f,t] = \widetilde{\mathbf{W}}[f]\mathbf{x}^{(\mathrm{obs})}[f,t] \tag{2}$$

のように得られる．

2.2 白色化と PB を用いる分離行列推定の流れ

図–1 に示すように，多くの BSS 手法は白色化，分離行列の最適化，PB の三段階の手続きをとる [1]．本節では，観測信号 $\mathbf{x}^{(\mathrm{obs})}[f,t] \in \mathbb{C}^M$ に対し白色化行列 $\mathbf{Q}[f] \in \mathbb{C}^{N \times M}$，分離行列 $\mathbf{W}[f] \in \mathbb{C}^{N \times N}$，PB 行列 $\mathbf{D}[f] \in \mathbb{C}^{N \times N}$ を適用し分離音 $\mathbf{y}[f,t]^{(\mathrm{PB})} \in \mathbb{C}^N$ を得る流れを述べる．ただし $M = N$ とし，各段階の分離行列，観測信号，分離音を表–1 に従って区別する．また，$\mathbf{y} = (\mathbf{y}[f,t])_{f=1,t=1}^{F,T}$ のようにインデックスを略すことで全成分をまとめて表記し，線形演算は $\mathbf{Wx} = (\mathbf{W}[f]\mathbf{x}[f,t])_{f=1,t=1}^{F,T}$ と略記する．

2.2.1 白色化行列 Q

白色化は目的関数の形状を均一化することで最適化アルゴリズムを安定化させるための前処理である．

*Constrained determined BSS realizing projection back through metric projection. By Kazuki MATSUMOTO (Waseda University), Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).

表–1　本稿で用いる記号とそれらの関係

名称	表記	役割（他の変数との関係）
白色化行列	$\mathbf{Q}$	目的関数を均一化
分離行列	$\mathbf{W}$	白色化後の目的関数を最小化
PB 行列	$\mathbf{D}$	振幅や位相の不定性を解消
最適化後の分離行列	$\widetilde{\mathbf{W}}$	$(=\mathbf{WQ})$
PB 後の分離行列	$\mathbf{W}^{(\mathrm{PB})}$	$(=\mathbf{D}\widetilde{\mathbf{W}})$
観測信号	$\mathbf{x}^{(\mathrm{obs})}$	
白色化後の観測信号	$\mathbf{x}$	$(=\mathbf{Q}\mathbf{x}^{(\mathrm{obs})})$
最適化後の分離音	$\tilde{\mathbf{y}}$	$(=\mathbf{W}\mathbf{x},\quad \widetilde{\mathbf{W}}\mathbf{x}^{(\mathrm{obs})})$
PB 後の分離音	$\mathbf{y}^{(\mathrm{PB})}$	$(=\mathbf{D}\tilde{\mathbf{y}},\quad \mathbf{W}^{(\mathrm{PB})}\mathbf{x}^{(\mathrm{obs})})$

本稿では，白色化後の観測信号

$$\mathbf{x}=\mathbf{Q}\mathbf{x}^{(\mathrm{obs})} \tag{3}$$

について，行列 $\sum_{t=1}^{T}\mathbf{x}[f,t](\mathbf{x}[f,t])^{\mathsf{H}}$ が単位行列になるような白色化行列

$$\mathbf{Q}[f]=\left(\sum_{t=1}^{T}\mathbf{x}^{(\mathrm{obs})}[f,t]\,(\mathbf{x}^{(\mathrm{obs})}[f,t])^{\mathsf{H}}\right)^{-1/2} \tag{4}$$

を用いる [1]．

2.2.2　分離行列 W

分離行列 $\mathbf{W}$ は，最適化問題

$$\min_{\mathbf{W},\tilde{\mathbf{y}}}\ \mathcal{I}(\mathbf{W})+\mathcal{P}(\tilde{\mathbf{y}}),\quad \text{s.t.}\quad \tilde{\mathbf{y}}=\mathbf{W}\,\mathbf{x} \tag{5}$$

を解くことで得られる [1]．ただし，$\mathcal{I}(\mathbf{W})=-\sum_{f=1}^{F}\log(|\det(\mathbf{W}[f])|)$ であり，$\mathcal{P}:\mathbb{C}^{NTF}\to\mathbb{R}\cup\{+\infty\}$ は音源モデルに応じた関数である．音源モデルとして局所ガウス分布（LGM）を用いる場合，

$$\mathcal{P}_{\mathrm{LGM}}(\mathbf{y})=\sum_{n=1}^{N}\sum_{f=1}^{F}\sum_{t=1}^{T}\frac{|y_n[f,t]|^2}{2\sigma_n^2[f,t]} \tag{6}$$

と書ける [2]．ただし，$\sigma_n^2[f,t]>0$ は分散パラメータである．前段の白色化の影響を考慮すれば，最適化後の分離行列 $\widetilde{\mathbf{W}}$ 及び分離音 $\tilde{\mathbf{y}}$ は，

$$\widetilde{\mathbf{W}}=\mathbf{WQ},\qquad \tilde{\mathbf{y}}=\mathbf{Wx} \tag{7}$$

と書ける．

2.2.3　Projection Back 行列 D

最適化後の分離行列 $\widetilde{\mathbf{W}}$ や分離音 $\tilde{\mathbf{y}}$ は，異なる周波数間で振幅や位相の整合性が保たれておらず，周波数特性が歪んでいる場合がある．PB はこの問題に対処するための後処理であり，分離音の周波数特性を k $(1\le k\le M)$ チャネル目の観測信号に合わせることで特性の歪みを解消する [1]．本稿では $k=1$ として議論を進める．PB 行列 $\mathbf{D}[f]$ は，式 (5) の時点で推定された混合行列 $\widetilde{\mathbf{H}}[f]=(\tilde{h}_{ij}[f])=(\widetilde{\mathbf{W}}[f])^{-1}$ の 1 行目を対角に並べた以下の行列になる．

$$\mathbf{D}[f]=\begin{pmatrix}\tilde{h}_{1,1}[f] & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & \tilde{h}_{1,N}[f]\end{pmatrix} \tag{8}$$

PB 後の分離行列 $\mathbf{W}^{(\mathrm{PB})}$ 及び分離音 $\mathbf{y}^{(\mathrm{PB})}$ は，

$$\mathbf{W}^{(\mathrm{PB})}=\mathbf{D}\widetilde{\mathbf{W}},\qquad \mathbf{y}^{(\mathrm{PB})}=\mathbf{D}\tilde{\mathbf{y}} \tag{9}$$

と書ける．

3　提案手法

三段階の分離行列推定では最適化の段階で扱う分離音 $\tilde{\mathbf{y}}$ の周波数特性が歪んでいる場合がある（図–1，3-staged）．前述のとおり，この歪みは学習済みの DNN を活用する場合など [3,7]，緻密な音源モデルを考える際に不都合を生じる可能性がある．本稿ではこの問題を解決するために，分離行列を最適化する段階で PB 制約 $\widetilde{\mathbf{W}}\in\mathcal{W}_{\mathrm{PB}}$ を課す（図–1，Direct）．これにより，最適化の段階で扱う分離音 $\tilde{\mathbf{y}}$ を周波数特性の歪みが解消された PB 後の分離音 $\mathbf{y}^{(\mathrm{PB})}$ と一致させることができる．以下では，近接分離法で PB 制約を扱うために，PB 制約集合 $\mathcal{W}_{\mathrm{PB}}$ への距離射影を導出する．その後，距離射影に基づく PB を利用した正則化項や ADMM による最適化アルゴリズムを提案する．

3.1　PB 制約集合 $\mathcal{W}_{\mathrm{PB}}$ への距離射影

本節では，最適化の段階で PB 後の分離音を扱うための制約集合を定義し，そこへの距離射影を導入する．先行研究 [8,9] でも議論されているように，各周波数における分離行列 $\widetilde{\mathbf{W}}[f]$ が可逆であり[1]，条件式

$$\sum_{k=1}^{N}\tilde{w}_{kj}[f]=\begin{cases}1 & (j=1)\\ 0 & (j\neq 1)\end{cases} \tag{10}$$

を満たすとき，$\widetilde{\mathbf{W}}$ は PB の操作に対する不動点となる，すなわち，$\mathbf{W}^{(\mathrm{PB})}=\widetilde{\mathbf{W}}$ が成り立つ[2]．そこで本稿では，最適化後の分離音 $\tilde{\mathbf{y}}$ と PB 後の分離音 $\mathbf{y}^{(\mathrm{PB})}$ を一致させる制約を，最適化後の分離行列 $\widetilde{\mathbf{W}}$ を

$$\mathcal{W}_{\mathrm{PB}}=\left\{\hat{\mathbf{W}}\in\mathbb{C}^{F\times N\times M}\ \middle|\ \sum_{k=1}^{N}\hat{w}_{kj}[f]=\delta_{1j}\,(1\le j\le M,\ 1\le f\le F)\right\} \tag{12}$$

[1] $\widetilde{\mathbf{W}}[f]$ の可逆性は LogDet 項により満たされる．

[2] 条件式 (10) の導出を記す．分離行列を $\widetilde{\mathbf{W}}[f]$ とし，対応する PB 行列を $\mathbf{D}[f]$ とする．$\widetilde{\mathbf{W}}[f]$ が PB の不動点であるとすると，$\widetilde{\mathbf{W}}[f]=\mathbf{D}[f]\,\widetilde{\mathbf{W}}[f]$ を満たす．さらに，$\widetilde{\mathbf{W}}[f]$ の可逆性から $\mathbf{D}[f]=\mathbf{I}$ を得る．混合行列を $\widetilde{\mathbf{H}}[f]=(\tilde{h}_{ij}[f])=(\widetilde{\mathbf{W}}[f])^{-1}$ とおき，式 (8) と照らし合わせれば $\tilde{h}_{1j}[f]=1$ が成り立つ．よって

$$\widetilde{\mathbf{W}}^{-1}[f]=\begin{pmatrix}1 & \cdots & 1\\ * & \cdots & *\\ \vdots & \ddots & \vdots\\ * & \cdots & *\end{pmatrix}=\begin{pmatrix}\mathbf{1}^{\mathsf{T}}\\ *\\ \vdots\\ *\end{pmatrix} \tag{11}$$

を得る．ここで，逆行列の定義 $(\widetilde{\mathbf{W}}[f])^{-1}\,\widetilde{\mathbf{W}}[f]=\mathbf{I}$ の 1 行目に着目すれば $\mathbf{1}^{\mathsf{T}}\widetilde{\mathbf{W}}[f]=\mathbf{e}_1^{\mathsf{T}}$ が得られ，これは式 (10) と一致する．ただし，$\mathbf{1}\in\mathbb{C}^N$ は全要素が 1 のベクトル，$\mathbf{e}_1\in\mathbb{C}^M$ は第 1 成分が 1，その他の成分が 0 のベクトルである．同様の議論は先行研究 [8,9] でなされている．

で定める PB 制約集合 $\mathcal{W}_{\rm PB}$ の元に制約することで実現する．ただし，δ_{ij} はクロネッカーのデルタであり，$i=j$ のとき 1，$i \neq j$ のとき 0 を返す．式 (12) より，$\mathcal{W}_{\rm PB}$ はアフィン集合なので，そこへの距離射影

$$\Pi_{\mathcal{W}_{\rm PB}}(\mathbf{W}) = \mathop{\arg\min}_{\hat{\mathbf{W}} \in \mathcal{W}_{\rm PB}} \sum_{f=1}^{F} \|\hat{\mathbf{W}}[f] - \mathbf{W}[f]\|_{\rm Fro}^2 \quad (13)$$

は一意に定まる．ただし，$\|\cdot\|_{\rm Fro}$ はフロベニウスノルムである．さらに，距離射影 $\Pi_{\mathcal{W}_{\rm PB}}$ は

$$(\Pi_{\mathcal{W}_{\rm PB}}(\mathbf{W})[f])_{ij} = w_{ij}[f] + \frac{1}{N}\left(\delta_{1j} - \sum_{k=1}^{N} w_{kj}[f]\right) \quad (14)$$

と解析的に書ける．特筆すべき点として，式 (8) に示した従来の PB 行列の適用は逆行列演算を含む不連続な写像であるのに対し，式 (14) に示す距離射影に基づく PB は非拡大な写像であることがあげられる．収束性が作用素の非拡大性に依拠する近接分離法では，距離射影に基づく PB の利用がアルゴリズムの安定した挙動に寄与すると期待される．

3.2 PB 正則化

提案法では，式 (5) に示した従来の最適化問題に対し，最適化後の分離行列を PB 制約集合 $\mathcal{W}_{\rm PB}$ に近づける効果を持つ正則化項 $\mathcal{Q}(\mathbf{W}^{\rm (PB)})$ を導入する．提案する最適化問題は

$$\begin{aligned} \min_{\mathbf{W},\mathbf{W}^{\rm (PB)},\mathbf{y}^{\rm (PB)}} \quad & \mathcal{I}(\mathbf{W}) + \mathcal{P}(\mathbf{y}^{\rm (PB)}) + \gamma\mathcal{Q}(\mathbf{W}^{\rm (PB)}) \\ \text{s.t.} \quad & \mathbf{y}^{\rm (PB)} = \mathbf{W}\mathbf{x},\ \mathbf{W}^{\rm (PB)} = \mathbf{W}\mathbf{Q} \end{aligned} \quad (15)$$

と書ける．ただし，$\gamma > 0$ は PB 正則化項の重みであり，$\mathcal{Q}(\mathbf{W}^{\rm (PB)})$ は周波数ごとの $\mathcal{W}_{\rm PB}$ との距離 $\mathcal{D}(\mathbf{W}^{\rm (PB)}[f]) = \|\Pi_{\mathcal{W}_{\rm PB}}(\mathbf{W}^{\rm (PB)}[f]) - \mathbf{W}^{\rm (PB)}[f]\|_{\rm Fro}$，及び関数 $\mathcal{R}: \mathbb{R}_+ \to \mathbb{R} \cup \{+\infty\}$ を用い

$$\mathcal{Q}(\mathbf{W}^{\rm (PB)}) = \sum_{f=1}^{F} \mathcal{R}(\mathcal{D}(\mathbf{W}^{\rm (PB)}[f])) \quad (16)$$

と定義する．$\mathcal{R}$ が下半連続で真凸な単調非減少関数のとき，$\mathcal{Q}(\mathbf{W}^{\rm (PB)})$ は $\mathbf{W}^{\rm (PB)} \in \mathcal{W}_{\rm PB}$ で最小値をとるような下半連続真凸関数になる．本稿では正則化の強さと分離性能の関連を調査するために，関数 $\mathcal{R}$ として $\mathcal{R}_{\rm none}$（正則化なし），$\mathcal{R}_{\rm soft}$（二乗距離正則化），$\mathcal{R}_{\rm hard}$（一致制約）の三種類を用いる．これらの関数は

$$\mathcal{R}_{\rm none}(x) = 0 \quad (17)$$

$$\mathcal{R}_{\rm soft}(x) = x^2/2 \quad (18)$$

$$\mathcal{R}_{\rm hard}(x) = \begin{cases} 0 & (x = 0) \\ +\infty & (x \neq 0) \end{cases} \quad (19)$$

と定める．

3.3 ADMM による最適化

最適化問題 (15) は ADMM で最適化できる．式の簡略化のために，全周波数の分離行列 $\mathbf{W}$ をベクトル化したものを $\mathbf{w} \in \mathbb{C}^{NMF}$ とする．また，$\mathbf{X} \in \mathbb{C}^{NTF \times NMF}$ を観測信号からなるブロック対角行列とする [5]．このとき，式 (15) は最適化問題

$$\begin{aligned} \min_{\mathbf{w},\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3} \quad & \mathcal{I}(\mathbf{v}_1) + \mathcal{P}(\mathbf{v}_2) + \mathcal{Q}(\mathbf{v}_3) \\ \text{s.t.} \quad & \mathbf{v}_1 = \mathbf{w}, \quad \mathbf{v}_2 = \mathbf{X}\mathbf{w}, \quad \mathbf{v}_3 = \mathbf{Q}\mathbf{w} \end{aligned} \quad (20)$$

に書き換えられ，これに対し ADMM を適用することで **Alg. 1** を得る [6]．ただし，線形演算 $\mathbf{X}\mathbf{w}$ 及び $\mathbf{Q}\mathbf{w}$ は $\mathbf{W}\mathbf{x}$ 及び $\mathbf{W}\mathbf{Q}$ に対応する演算を略記したものである．また，$\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ は双対変数，$\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3$ は一時変数である．prox は近接作用素であり，関数 $\mathcal{F}: \mathbb{C}^K \to \mathbb{R} \cup \{+\infty\}$ とパラメータ $\mu > 0$ に対し

$$\mathrm{prox}_{\mu\mathcal{F}}(\mathbf{y}) = \mathop{\arg\min}_{\mathbf{s} \in \mathbb{C}^{NFT}} \left(\mathcal{F}(\mathbf{s}) + \frac{1}{2\mu}\|\mathbf{s} - \mathbf{y}\|_2^2\right) \quad (21)$$

で定義される．LogDet 項の近接作用素 $\mathrm{prox}_{\mu\mathcal{I}}$ は，入力された分離行列の n 番目の特異値 $\sigma_n[f]$ を $(\sigma_n[f] + \sqrt{(\sigma_n[f])^2 + 4\mu})/2$ に写す [5]．音源モデル項の近接作用素 $\mathrm{prox}_{\mu\mathcal{P}}$ は，式 (6) の場合，

$$\left(\mathrm{prox}_{\mathcal{P}_{\rm LGM}}(\mathbf{y})\right)_n [f,t] = \frac{y_n[f,t]}{1 + |\sigma_n[f,t]|^2} \quad (22)$$

と書ける．PB 正則化項の近接作用素 $\mathrm{prox}_{\mu\mathcal{Q}}$ は周波数ごとに計算でき，

$$\begin{aligned} \mathrm{prox}_{\mu\mathcal{Q}}(\mathbf{w}[f]) &= \Pi_{\mathcal{W}_{\rm PB}}(\mathbf{w}[f]) \\ &+ \left(\frac{\mathrm{prox}_{\mu\mathcal{R}}(\mathcal{D}(\mathbf{w}[f]))}{\mathcal{D}(\mathbf{w}[f])}\right)(\mathbf{w}[f] - \Pi_{\mathcal{W}_{\rm PB}}(\mathbf{w}[f])) \end{aligned} \quad (23)$$

で得られる．$\mathcal{D}(\mathbf{w}[f]) = 0$ のときに生じるゼロ除算には $\mathrm{prox}_{\mu\mathcal{Q}}(\mathbf{w}[f]) = \mathbf{w}[f]$ とすることで対処できる．式 (17)–(19) に示した関数の近接作用素は

$$\mathrm{prox}_{\mathcal{R}_{\rm none}}(x) = x \quad (24)$$

$$\mathrm{prox}_{\mu\mathcal{R}_{\rm soft}}(x) = x/(1+\gamma) \quad (25)$$

$$\mathrm{prox}_{\mu\mathcal{R}_{\rm hard}}(x) = 0 \quad (26)$$

と書ける [10]．

4 実験

PB 制約や正則化の妥当性を検証するための実験を行った．音源モデルとしては，式 (6) に示した LGM の分散パラメータ $\sigma_n^2[f,t]$ を真の音源の振幅の二乗 $|s_n[f,t]|^2$ で与えたものを用いた．PB 正則化には，式 (17)–(19) に示した None, Soft, Hard の 3 つを用いた．Soft に関しては，正則化項の重み γ として 1, 100, 10000 を用いた．比較手法としては，従来の ADMM-BSS（Conv.）[6]，及び **Alg. 2** に記載の，各反復で

Algorithm 1 ADMM-BSS with Metric-PB

Input: $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{w}$, $(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$, $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, ρ, α
Output: $\mathbf{w}$

for $l = 1, ..., L$ **do**
  $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \boldsymbol{\xi}_3) \leftarrow \alpha(\mathbf{w}, \mathbf{X}\mathbf{w}, \mathbf{Q}\mathbf{w}) + (1-\alpha)(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$
  $\mathbf{v}_1 \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{I}}(\boldsymbol{\xi}_1 + \mathbf{u}_1)$
  $\mathbf{v}_2 \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{P}}(\boldsymbol{\xi}_2 + \mathbf{u}_2)$
  $\mathbf{v}_3 \leftarrow \mathrm{prox}_{(\gamma/\rho)\mathcal{Q}}(\boldsymbol{\xi}_3 + \mathbf{u}_3)$
  $\mathbf{u}_i \leftarrow \mathbf{u}_i + \boldsymbol{\xi}_i - \mathbf{v}_i, \quad (i = 1, 2, 3)$
  $\mathbf{w} \leftarrow (\mathbf{I} + \mathbf{X}^{\mathsf{H}}\mathbf{X} + \mathbf{Q}^{\mathsf{H}}\mathbf{Q})^{-1}$
    $((\mathbf{v}_1 - \mathbf{u}_1) + \mathbf{X}^{\mathsf{H}}(\mathbf{v}_2 - \mathbf{u}_2) + \mathbf{Q}^{\mathsf{H}}(\mathbf{v}_3 - \mathbf{u}_3))$
end for

Algorithm 2 ADMM-BSS with Heuristic-PB

Input: $\mathbf{X}$, $\mathbf{Q}$, $\mathbf{w}$, $(\mathbf{u}_1, \mathbf{u}_2)$, $(\mathbf{v}_1, \mathbf{v}_2)$, ρ, α
Output: $\mathbf{w}$

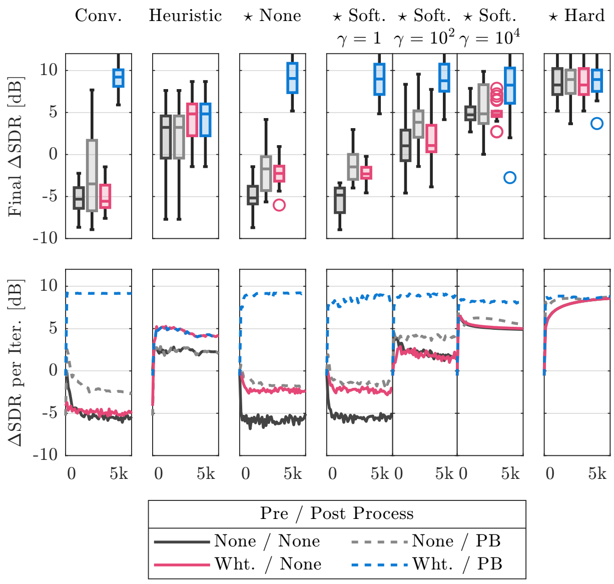
for $l = 1, ..., L$ **do**
  $(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \leftarrow \alpha(\mathbf{w}, \mathbf{X}\mathbf{w}) + (1-\alpha)(\mathbf{v}_1, \mathbf{v}_2)$
  $\mathbf{v}_1 \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{I}}(\boldsymbol{\xi}_1 + \mathbf{u}_1)$
  $\mathbf{v}_2 \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{P}}(\boldsymbol{\xi}_2 + \mathbf{u}_2)$
  $\mathbf{u}_i \leftarrow \mathbf{u}_i + \boldsymbol{\xi}_i - \mathbf{v}_i, \quad (i = 1, 2)$
  $\mathbf{w} \leftarrow (\mathbf{I} + \mathbf{X}^{\mathsf{H}}\mathbf{X})^{-1} ((\mathbf{v}_1 - \mathbf{u}_1) + \mathbf{X}^{\mathsf{H}}(\mathbf{v}_2 - \mathbf{u}_2))$
  Apply $\mathbf{D}$ in Eq.(8) to $(\mathbf{w}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{v}_1, \mathbf{v}_2)$
end for

ヒューリスティックにPBを適用する手法（Heuristic）を用いた．前後処理の白色化とPBは，それぞれ適用する場合としない場合を比較した．観測信号にはSiSEC2011の dev1 データセットから作成した計24組の2チャネル2話者混合音を用いた．反復回数は5000回とし，ADMMのパラメータ ρ は1000，α は1に設定した．分離性能は ΔSDR で評価した．

実験結果は図–2に示す．Conv. は白色化とPBをともに利用する場合（Wht. / PB），すなわち三段階の手続きでのみ分離を実現した．後処理のPBを省いた場合の分離性能（Wht. / None）を比較すると，Heuristic は Conv. の分離性能を改善するものの三段階の手続きには劣った．この結果はヒューリスティックなPBが最適化を不安定にすることを示唆した．一方，⋆印を付けた提案法では，PB制約や正則化を強く課すほど性能が向上し，三段階の手続きと同程度の分離性能を達成した．したがって，距離射影に基づくPBを利用することで，近接分離法の安定性を保ったままPB後の分離音を直接的に最適化できることが分かった．さらに，PB制約は白色化を利用しない場合（None / None，None / PB）の性能を改善することも確認された．これは，PB制約による探索空間の制限が収束性の向上に寄与したものと考えられる．

5　むすび

本稿では，距離射影に基づくPBを利用し，最適化の段階でPB後の分離音を扱うための枠組みを提案した．提案フレームワークの特徴を生かした音源モデルの検討や加速の導入による反復回数の削減が今後の課題である．



図–2　各アルゴリズムの分離性能．上段には最終反復における分離性能を，下段には分離性能の推移（横軸は反復回数）を示している．提案アルゴリズム（Alg.1）には⋆印をつけている．色は前後処理の有無を示す．

謝辞　我々は2023年12月に独自に式(10)を導出しましたが，それを発表する前の2024年3月に東京都立大学の小野研究室から同様の結果が発表されました[8]．今回本研究を発表するにあたり，内容の重複度合いや互いの進捗について小野順貴先生に問い合わせ，発表して問題ないことを確認して頂きました．お忙しい中お時間を頂きありがとうございました．

参考文献

[1] N. Murata, S. Ikeda and A. Ziehe, “An approach to blind source separation based on temporal structure of speech signals,” *Neurocomputing*, **41**(1-4), 1–24 (2001).
[2] H. Sawada, N. Ono, H. Kameoka, D. Kitamura and H. Saruwatari, “A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF,” *APSIPA Trans. Signal Inf. Process.*, **8**(1), (2019).
[3] N. Makishima, S. Mogami, N. Takamune, D. Kitamura, H. Sumino, S. Takamichi, H. Saruwatari and N. Ono, “Independent deeply learned matrix analysis for determined audio source separation,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **27**(10), 1601–1615 (2019).
[4] D. Kitamura and K. Yatabe, “Consistent independent low-rank matrix analysis for determined blind source separation,” *EURASIP J. Adv. Signal Process.*, **2020**(1), 46 (2020).
[5] K. Yatabe and D. Kitamura, “Determined BSS based on time-frequency masking and its application to harmonic vector analysis,” *IEEE/ACM Trans. Audio Speech Lang. Process.*, **29**, 1609–1625 (2021).
[6] 渡會博子, 山田宏樹, 矢田部浩平, “ADMM アルゴリズムを用いた優決定ブラインド音源分離,” 音講論集, pp. 145–146 (2023.9).
[7] K. Matsumoto and K. Yatabe, “Determined BSS by combination of IVA and DNN via proximal average,” *Proc. IEEE ICASSP*, pp. 871–875 (2024).
[8] 栗城結衣, 中嶋大志, 小野順貴, “プロジェクションバックされた分離行列の直接更新,” *Proc. SPEASIP*, pp. 31–46 (2023).
[9] Y. Kuriki, T. Nakashima and N. Ono, “Direct update of back-projected demixing matrices in blind source separation,” *Proc. EUSIPCO* (2024).
[10] N. Parikh and S. Boyd, “Proximal algorithms,” *Found. Trends Optim.*, **1**(3), 127–239 (2014).