ADMM を用いた ILRMA における スパース正則化項およびアルゴリズムの比較検討 *

◎渡會博子 (農工大)，松本和樹 (早大)，山田宏樹，矢田部浩平 (農工大)

1　はじめに

我々は前回，交互方向乗数法 (ADMM) を用いた独立低ランク行列分析 (ILRMA) 型ブラインド音源分離 (BSS) 手法を提案し，適切なパラメータの下で，補助関数法に基づく ILRMA と同程度の性能を示すことを確認した [1]．しかし，アルゴリズムを単純にするため，通常の ILRMA と異なる目的関数を用いていた．本稿では，目的関数を通常の ILRMA と合わせるために，スパース正則化項を変更した更新式 (ADMM-ILRMA) を導出した．実験により，ADMM-ILRMA では，低ランク近似に対して適度なスパース性が誘導されることで分離が促進される可能性が示された．

2　ADMM を用いた BSS [1]

ILRMA は分離行列 $\mathbf{W}[f]\in\mathbb{C}^{N\times M}$ の推定に基づく BSS 手法である．ただし，N, M $(N=M)$ は順に音源数とマイク数である．推定された分離行列を用いて，時間周波数領域の観測信号 $\mathbf{x}[t,f]\in\mathbb{C}^{M}$ から分離音 $\mathbf{y}[t,f]=\mathbf{W}[f]\,\mathbf{x}[t,f]\in\mathbb{C}^{N}$ が得られる．文献 [1] では，ILRMA の最適化問題を以下で定めた．

$$
\begin{aligned}
\underset{(\mathbf{w},\widetilde{\mathbf{w}},\mathbf{Y},\boldsymbol{\Upsilon},\mathbf{T},\mathbf{V},\widetilde{\mathbf{T}},\widetilde{\mathbf{V}})}{\text{Minimize}}\quad & \mathcal{I}(\widetilde{\mathbf{w}})+\mathcal{D}(\mathbf{Y},\boldsymbol{\Upsilon})\\
\text{subject to}\quad & \widetilde{\mathbf{w}}=\mathbf{w},\ \mathbf{Y}=\mathbf{X}\mathbf{w},\ \boldsymbol{\Upsilon}=\mathbf{T}\mathbf{V},\\
& \mathbf{T}=\widetilde{\mathbf{T}},\ \mathbf{V}=\widetilde{\mathbf{V}},\ \widetilde{\mathbf{T}}\geq 0,\ \widetilde{\mathbf{V}}\geq 0
\end{aligned}
\tag{1}
$$

ここで，$\mathbf{w}\in\mathbb{C}^{NMF}$ はベクトル化した分離行列，$\mathbf{X}\in\mathbb{C}^{NTF\times NMF}$ は観測信号からなるブロック対角行列，$\mathcal{I}(\mathbf{w})=-\sum_{f=1}^{F}\sum_{n=1}^{N}\log\sigma_n(\mathrm{mat}(\mathbf{w}[f]))$ である．ただし，$\sigma_n(\cdot)$ は n 番目の特異値を得る関数，$\mathrm{mat}(\cdot)$ は行列化作用素である．また，$\mathbf{T}=(\mathbf{T}_n)_{n=1}^{N}, \mathbf{V}=(\mathbf{V}_n)_{n=1}^{N}$ は NMF 変数 $\mathbf{T}_n\in\mathbb{R}_{+}^{F\times R}, \mathbf{V}_n\in\mathbb{R}_{+}^{R\times T}$ の組，R は NMF の基底数，$\mathbf{T}\mathbf{V}=(\mathbf{T}_n\mathbf{V}_n)_{n=1}^{N}\in\mathbb{R}_{+}^{NFT}$ は低ランク近似の組である．また，$\mathcal{D}$ は，$|\mathbf{Y}|^2$ と $\boldsymbol{\Upsilon}$ 間の構造を関連付け，$\boldsymbol{\Upsilon}$ のスパース性を誘導する関数である．我々は前回，変数分離の簡略化とアルゴリズムの単純化のために，通常の ILRMA と異なる $\mathcal{D}$ を用いていた．前回用いた $\mathcal{D}$ では，式 (1) の変数分離によって，各変数の更新を解析的に計算可能であった．

3　ADMM-ILRMA

本稿では，目的関数を通常の ILRMA と合わせるために，スパース正則化項を変更する．各更新式を解析的に計算するため，スパース正則化項に対して変数分離を行う．補助変数 $\boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2$ を追加することで

$$
\begin{aligned}
\underset{\substack{(\mathbf{w},\widetilde{\mathbf{w}},\mathbf{Y},\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}_1,\boldsymbol{\Upsilon}_2,\\ \mathbf{T},\mathbf{V},\widetilde{\mathbf{T}},\widetilde{\mathbf{V}})}}{\text{Minimize}}\quad & \mathcal{I}(\mathbf{w})+\Phi(\mathbf{Y},\boldsymbol{\Upsilon}_1)+\gamma\log(|\boldsymbol{\Upsilon}_2|+\epsilon)\\
\text{subject to}\quad & \widetilde{\mathbf{w}}=\mathbf{w},\ \mathbf{Y}=\mathbf{X}\mathbf{w},\\
& \boldsymbol{\Upsilon}=\mathbf{T}\mathbf{V},\ \boldsymbol{\Upsilon}_1=\boldsymbol{\Upsilon},\ \boldsymbol{\Upsilon}_2=\boldsymbol{\Upsilon},\\
& \mathbf{T}=\widetilde{\mathbf{T}},\ \mathbf{V}=\widetilde{\mathbf{V}},\ \widetilde{\mathbf{T}}\geq 0,\ \widetilde{\mathbf{V}}\geq 0
\end{aligned}
\tag{2}
$$

と変形できる．ただし，$\boldsymbol{\Upsilon}$ の要素ごとの log の総和を $\log(\boldsymbol{\Upsilon})$ と略記した．ここで，$\gamma>0$ は log 項の重み，$\epsilon>0$ は log の発散を回避するための微小定数である．また，Φ は 2 変数 $(\mathbf{Y},\boldsymbol{\Upsilon})$ に対する関数

$$
\Phi(\mathbf{Y},\boldsymbol{\Upsilon})=\sum_{t=1}^{T}\sum_{f=1}^{F}\sum_{n=1}^{N}\phi(y_n[t,f],\upsilon_n[f,t]) \tag{3}
$$

であり，ϕ は以下の $|\cdot|^2$ の perspective である [2]．

$$
\phi(y,\upsilon)=\begin{cases}\dfrac{|y|^2}{\upsilon} & (\upsilon>0)\\ 0 & (y=0 \text{ and } \upsilon=0)\\ +\infty & (\text{otherwise})\end{cases} \tag{4}
$$

式 (2) に ADMM を適用すると，Alg. 1, 2 が得られる．ここで，$\max\{\cdot,0\}$ は要素ごとの非負射影，prox は近接作用素 [2]，$\rho>0$，$\alpha\in(0,2)$ である．また，

Algorithm 1 ADMM-ILRMA

Input: $\mathbf{X},\mathbf{w},\widetilde{\mathbf{w}},\mathbf{Y},\boldsymbol{\Upsilon},\boldsymbol{\Upsilon}_1,\boldsymbol{\Upsilon}_2,\mathbf{T},\mathbf{V},\widetilde{\mathbf{T}},\widetilde{\mathbf{V}}$, $\mathbf{u}_{\widetilde{\mathbf{w}}},\mathbf{u}_{\mathbf{Y}},\mathbf{u}_{\boldsymbol{\Upsilon}},\mathbf{u}_{\boldsymbol{\Upsilon}_1},\mathbf{u}_{\boldsymbol{\Upsilon}_2},\mathbf{u}_{\mathbf{T}},\mathbf{u}_{\mathbf{V}},\rho,\alpha,\gamma,\epsilon$
Output: $\mathbf{w}$

for $k=1,\ldots,K$ **do**
　$\mathbf{w} \leftarrow (\mathbf{X}^{\mathsf{H}}\mathbf{X}+\mathbf{I})^{-1}\big((\widetilde{\mathbf{w}}-\mathbf{u}_{\widetilde{\mathbf{w}}})+\mathbf{X}^{\mathsf{H}}(\mathbf{Y}-\mathbf{u}_{\mathbf{Y}})\big)$
　$(\boldsymbol{\xi}_{\widetilde{\mathbf{w}}},\boldsymbol{\xi}_{\mathbf{Y}}) \leftarrow \alpha(\mathbf{w},\mathbf{X}\mathbf{w})+(1-\alpha)(\widetilde{\mathbf{w}},\mathbf{Y})$
　$\widetilde{\mathbf{w}} \leftarrow \mathrm{prox}_{(1/\rho)\mathcal{I}}[\boldsymbol{\xi}_{\widetilde{\mathbf{w}}}+\mathbf{u}_{\widetilde{\mathbf{w}}}]$
　for $n=1,\ldots,N$ **do**
　　Update $\mathbf{T}_n,\mathbf{V}_n,\widetilde{\mathbf{T}}_n,\widetilde{\mathbf{V}}_n,\mathbf{u}_{\mathbf{T}n},\mathbf{u}_{\mathbf{V}n},\boldsymbol{\xi}_{\boldsymbol{\Upsilon}n}$ by Channel-wise Update in Alg. 2
　end for
　$\boldsymbol{\Upsilon} \leftarrow \frac{1}{3}\big((\xi_{\boldsymbol{\Upsilon}}+\mathbf{u}_{\boldsymbol{\Upsilon}})+(\boldsymbol{\Upsilon}_1-\mathbf{u}_{\boldsymbol{\Upsilon}_1})+(\boldsymbol{\Upsilon}_2-\mathbf{u}_{\boldsymbol{\Upsilon}_2})\big)$
　$(\boldsymbol{\xi}_{\boldsymbol{\Upsilon}_1},\boldsymbol{\xi}_{\boldsymbol{\Upsilon}_2}) \leftarrow \alpha(\boldsymbol{\Upsilon},\boldsymbol{\Upsilon})+(1-\alpha)(\boldsymbol{\Upsilon}_1,\boldsymbol{\Upsilon}_2)$
　$(\mathbf{Y},\boldsymbol{\Upsilon}_1) \leftarrow \mathrm{prox}_{(1/\rho)\Phi}[\boldsymbol{\xi}_{\mathbf{Y}}+\mathbf{u}_{\mathbf{Y}},\ \boldsymbol{\xi}_{\boldsymbol{\Upsilon}_1}+\mathbf{u}_{\boldsymbol{\Upsilon}_1}]$
　$\boldsymbol{\Upsilon}_2 \leftarrow \mathrm{prox}_{(\gamma/\rho)\log(|\cdot|+\epsilon)}[\boldsymbol{\xi}_{\boldsymbol{\Upsilon}_2}+\mathbf{u}_{\boldsymbol{\Upsilon}_2}]$
　$(\mathbf{u}_{\widetilde{\mathbf{w}}},\mathbf{u}_{\mathbf{Y}},\mathbf{u}_{\boldsymbol{\Upsilon}},\mathbf{u}_{\boldsymbol{\Upsilon}_1},\mathbf{u}_{\boldsymbol{\Upsilon}_2})$
　　$\leftarrow (\mathbf{u}_{\widetilde{\mathbf{w}}},\mathbf{u}_{\mathbf{Y}},\mathbf{u}_{\boldsymbol{\Upsilon}},\mathbf{u}_{\boldsymbol{\Upsilon}_1},\mathbf{u}_{\boldsymbol{\Upsilon}_2})$
　　　$+(\boldsymbol{\xi}_{\widetilde{\mathbf{w}}}-\widetilde{\mathbf{w}},\boldsymbol{\xi}_{\mathbf{Y}}-\mathbf{Y},\boldsymbol{\xi}_{\boldsymbol{\Upsilon}}-\boldsymbol{\Upsilon},\boldsymbol{\xi}_{\boldsymbol{\Upsilon}_1}-\boldsymbol{\Upsilon}_1,\boldsymbol{\xi}_{\boldsymbol{\Upsilon}_2}-\boldsymbol{\Upsilon}_2)$
end for

* Comparison of sparse regularization terms and algorithms in ADMM-based ILRMA. By Hiroko WATARAI (Tokyo University of Agriculture and Technology), Kazuki MATSUMOTO (Waseda University), Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).

Algorithm 2 Channel-wise Update

Input: $\boldsymbol{\Upsilon}, \mathbf{T}, \mathbf{V}, \widetilde{\mathbf{T}}, \widetilde{\mathbf{V}}, \mathbf{u}_{\boldsymbol{\Upsilon}}, \mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}, \alpha$

Output: $\mathbf{T}, \mathbf{V}, \widetilde{\mathbf{T}}, \widetilde{\mathbf{V}}, \mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}, \boldsymbol{\xi}_{\boldsymbol{\Upsilon}}$

$\mathbf{T} \leftarrow ((\boldsymbol{\Upsilon} - \mathbf{u}_{\boldsymbol{\Upsilon}})\mathbf{V}^{\mathsf{T}} + (\widetilde{\mathbf{T}} - \mathbf{u}_{\mathbf{T}}))(\mathbf{V}\mathbf{V}^{\mathsf{T}} + \mathbf{I})^{-1}$

$\boldsymbol{\xi}_{\mathbf{T}} \leftarrow \alpha\mathbf{T} + (1-\alpha)\widetilde{\mathbf{T}}$

$\mathbf{V} \leftarrow ({\boldsymbol{\xi}_{\mathbf{T}}}^{\mathsf{T}}\boldsymbol{\xi}_{\mathbf{T}} + \mathbf{I})^{-1}({\boldsymbol{\xi}_{\mathbf{T}}}^{\mathsf{T}}(\boldsymbol{\Upsilon} - \mathbf{u}_{\boldsymbol{\Upsilon}}) + (\widetilde{\mathbf{V}} - \mathbf{u}_{\mathbf{V}}))$

$(\boldsymbol{\xi}_{\mathbf{V}}, \boldsymbol{\xi}_{\boldsymbol{\Upsilon}}) \leftarrow \alpha(\mathbf{V}, \mathbf{T}\mathbf{V}) + (1-\alpha)(\widetilde{\mathbf{V}}, \boldsymbol{\Upsilon})$

$(\widetilde{\mathbf{T}}, \widetilde{\mathbf{V}}) \leftarrow \max\{(\boldsymbol{\xi}_{\mathbf{T}} + \mathbf{u}_{\mathbf{T}}, \boldsymbol{\xi}_{\mathbf{V}} + \mathbf{u}_{\mathbf{V}}), 0\}$

$(\mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}) \leftarrow (\mathbf{u}_{\mathbf{T}}, \mathbf{u}_{\mathbf{V}}) + (\boldsymbol{\xi}_{\mathbf{T}} - \mathbf{T}, \boldsymbol{\xi}_{\mathbf{V}} - \mathbf{V})$

log 項の近接作用素は，$0 < \lambda \leq \epsilon^2$ のとき

$$\mathrm{prox}_{\lambda \log(|\cdot|+\epsilon)}[v] = \begin{cases} 0 & (|v| \leq \frac{\lambda}{\epsilon}) \\ \mathrm{sgn}(v) r(|v|) & (|v| > \frac{\lambda}{\epsilon}) \end{cases} \tag{5}$$

であり，$0 < \epsilon^2 < \lambda$ のとき

$$\mathrm{prox}_{\lambda \log(|\cdot|+\epsilon)}[v] = \begin{cases} 0 & (|v| < v_*) \\ \{0,\ \mathrm{sgn}(v) r(v_*)\} & (|v| = v_*) \\ \mathrm{sgn}(v) r(|v|) & (|v| > v_*) \end{cases} \tag{6}$$

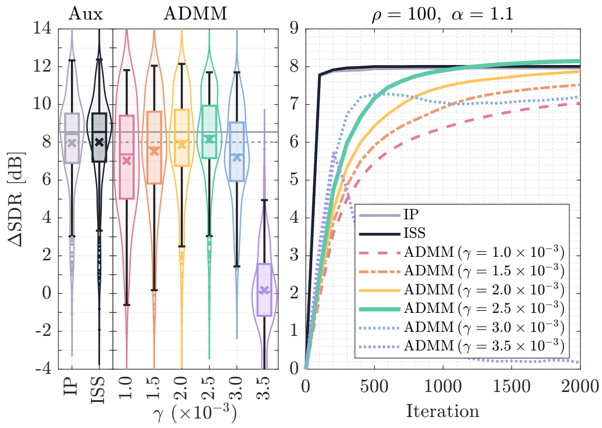
である [3]．ここで，$\mathrm{sgn}(\cdot)$ は符号関数，$r(z) = (1/2)(z-\epsilon) + \sqrt{(1/4)(z-\epsilon)^2 - \lambda}$ である．また，v_* は区間 $[2\sqrt{\lambda} - \epsilon,\ \frac{\lambda}{\epsilon}]$ 上の関数 $g(v) = q_{\lambda,z}(r(v)) - q_{\lambda,z}(0)$ の零点である．ただし，$q_{\lambda,z}(v) = (1/2\lambda)(v-z)^2 + \log(|v| + \epsilon)$ である．

4 実験

ADMM-ILRMA を AuxILRMA (IP，ISS) と比較する実験を行った．データセットは，SiSEC 2011 の `dev1 liverec` を用い，2 チャネルの 2 話者混合音 (計 24 組) を作成した．STFT の窓関数は 2048 サンプルの Hann 窓，シフト長は 1024 サンプルとした．反復回数は 2000 回とし，評価指標は ΔSDR を用いた．NMF の基底数 R は 2，NMF 変数 $\mathbf{T}, \mathbf{V}$ の初期値は 0 から 1 の一様分布乱数とし，乱数シードを変えて各手法をそれぞれ 10 回試行した．ADMM-ILRMA のパラメータのうち，スパース正則化項の重み γ を 0.001，0.0015，0.002，0.0025，0.003，0.0035 とし，ρ を 100，α を 1.1，ϵ を 1.0×10^{-3} とした．

最終的な ΔSDR の値及び 100 反復ごとの ΔSDR の平均値の推移を図–1 に示す．ADMM-ILRMA では，$\gamma = 0.0025$ のとき AuxILRMA と近い性能が得られ，1200 回程度の反復で平均値が上回った．適切なパラメータ設定下において平均値が向上する点は前回と同様であったが，変数分離の影響により，前回より多くの反復回数を要した [1]．また，図–1 左図より，ADMM-ILRMA では従来法と比較して高い分離性能を示すサンプル数が増加した一方で，その他のサンプルではわずかに性能が低下する傾向が見られた．

続いて，実験結果と γ の役割から，γ の値が分離音 $\mathbf{X}\mathbf{w}$ に及ぼす影響について考察する．図–1 右図から，



図–1　2000 反復後の ΔSDR (左) と反復に伴う ΔSDR の平均値の推移 (右)．左図では，各手法の最終的な ΔSDR の平均値を×で示し，AuxILRMA-ISS の中央値と平均値を左図を横断する実線と点線で示している．2000 反復後の ΔSDR の平均値は，AuxILRMA-ISS で 8.01 dB，ADMM-ILRMA ($\gamma = 0.0025$) で 8.15 dB であった．

300 反復以前では γ の値が大きいほど収束が速い一方で，300 反復以降では γ が大きすぎると ΔSDR が低下することが分かる．また，γ はスパース正則化項の重みであり，値が大きいほど変数 $\boldsymbol{\Upsilon}_2$ のスパース性が誘導される．これにより，間接的に低ランク近似 $\boldsymbol{\Upsilon}$ のスパース性も誘導される．$\boldsymbol{\Upsilon}$ のスパース性は，Φ を通じて変数 $\mathbf{Y}$ に反映される．さらに，変数 $\mathbf{Y}$ と分離音 $\mathbf{X}\mathbf{w}$ は等式制約によって関連付けられている．以上から，低ランク近似 $\boldsymbol{\Upsilon}$ に適度なスパース性が誘導された場合，それが間接的に分離音に反映されることで分離が促進される可能性が示唆された．一方で，γ の値が大きく，低ランク近似に対して過剰にスパース性が課された場合では，最適化が不安定になっていた．実際，$\gamma = 0.0035$ における最終反復の変数 $\boldsymbol{\Upsilon}_2$ を確認すると，半数以上の要素が 0 になっていた．このことから，変数 $\boldsymbol{\Upsilon}_2$ に対する過剰な閾値処理によって最適化が不安定となった可能性が示唆された．

5 むすび

本稿では，目的関数を通常の ILRMA と合わせるために，スパース正則化項を変更して ADMM-ILRMA を実現し，比較実験を行った．今後の展望として，提案手法の柔軟性を活かした正則化項や制約の追加や，アルゴリズムの高速化及び安定化が挙げられる．

参考文献

[1] 渡會 博子, 松本 和樹, 山田 宏樹, 矢田部 浩平, “ADMM アルゴリズムを用いた ILRMA 型ブラインド音源分離,” 音講論集, pp. 91–94 (2024.3).

[2] P. L. Combettes and C. L. Müller, “Perspective maximum likelihood-type estimation via proximal decomposition,” *Electron. J. Stat.*, **14**(1), 207–238 (2020).

[3] A. Prater-Bennette, L. Shen and E. E. Tripp, “The proximity operator of the log-sum penalty,” *J. Sci. Comput.*, **93**(3), 34 pages (2022).