行列リフティングを用いた単チャネルブラインド残響除去における ランク 1 制約の有効性 *

◎饒平名文希, 山田宏樹, 矢田部浩平 (農工大)

1 まえがき

本稿では，単チャネルの音響信号に含まれる残響を補助情報なしで抑圧するブラインド残響除去を扱う。前回の音響学会では，行列リフティングを用いた定式化に基づく手法を提案した [1]。本稿では，最適化問題におけるランク 1 制約を緩めた際の性能の変化を検証する。評価実験により，残響除去性能と実行時間におけるランク 1 制約の妥当性を確認した。

2 行列リフティングに基づく手法

時間周波数領域における f 番目の周波数ビンを取り出した，時系列信号としての観測信号 $\boldsymbol{x}_f \in \mathbb{C}^N$ を

$$\boldsymbol{x}_f = \boldsymbol{s}_f * \boldsymbol{h}_f \tag{1}$$

として近似する。ここで，$\boldsymbol{s}_f \in \mathbb{C}^N$ および $\boldsymbol{h}_f \in \mathbb{C}^M$ は音源信号と室内インパルス応答の時間周波数表現における f 番目の周波数ビンでの時系列信号を，$*$ は時間方向の畳み込みをそれぞれ表している。ブラインド残響除去では，各周波数で $\boldsymbol{s}_f$，$\boldsymbol{h}_f$ の両方を同時に推定し，逆短時間フーリエ変換によりまとめて時間領域に戻すことで目的信号を得ることを考える。以降，周波数インデックス f は省略する。

二つの変数を同時に扱うことで問題が双線形となることによる難しさを回避するために，提案手法では行列リフティングを用いて式 (1) を $\boldsymbol{x} = \boldsymbol{s} * \boldsymbol{h} = \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}$ として書き換える。ここで，$\boldsymbol{Z}$ は $\boldsymbol{s}$ と $\boldsymbol{h}$ のリフティングによって $\boldsymbol{Z} = \boldsymbol{s}\boldsymbol{h}^{\mathrm{T}}$ として得られたランク 1 行列，$\mathcal{S}$ は行列の各列ベクトルに対して周期的にシフトを行う線形作用素，$\mathbf{1} \in \{1\}^M$ は要素が全て 1 のベクトルである。このとき，残響除去問題を

$$\min_{\boldsymbol{Z} \in \mathbb{C}^{N \times M}} \|\boldsymbol{Z}\|_{2,1} \quad \text{s.t.} \quad \boldsymbol{x} = \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}, \quad \mathrm{rank}(\boldsymbol{Z}) = 1 \tag{2}$$

として定式化する。音源信号は時間周波数領域においてスパースであるという仮定から，$\boldsymbol{Z}$ は行ごとにスパース性を持つため，$\ell_{2,1}$ 混合ノルム正則化を課している。この定式化では，観測信号モデル $\boldsymbol{x} = \boldsymbol{s} * \boldsymbol{h}$ の双線形性による難しさを非凸制約 $\mathrm{rank}(\boldsymbol{Z}) = 1$ による難しさに置き換えている．式 (2) に ADMM を適用した提案手法のアルゴリズムを Algorithm 1 に示す [1]。

Algorithm 1 Proposed Method

Input: $\boldsymbol{x} \in \mathbb{R}^{N_{\mathrm{TD}}}$, $M \in \mathbb{N}$, $\rho > 0$
Output: $\boldsymbol{s} \in \mathbb{R}^{N_{\mathrm{TD}}}$

Function DEREVERB_1CH($\boldsymbol{x}, M, \rho$)
  $\boldsymbol{X} = \mathrm{STFT}(\boldsymbol{x})$ ▷ $\boldsymbol{X} \in \mathbb{C}^{F \times N}$
  for $f = 1, 2, \ldots, F$ **do**
    $\boldsymbol{S}(f, :) = \text{ADMM\_SOLVER}(\boldsymbol{X}(f, :)^{\mathrm{T}}, M, \rho)^{\mathrm{T}}$
  $\boldsymbol{s} = \mathrm{iSTFT}(\boldsymbol{S})$
  return $\boldsymbol{s}$
Function ADMM_SOLVER($\boldsymbol{x}, M, \rho$)
  $a = \|\boldsymbol{x}\|_2$ ▷ $\boldsymbol{x} \in \mathbb{C}^N$
  $\boldsymbol{x} = \boldsymbol{x}/a$ ▷ ℓ_2 normalization
  $\boldsymbol{Z}^1 = \boldsymbol{x}\,[1, 0, 0, \ldots, 0]$ ▷ $\boldsymbol{Z}^1 \in \mathbb{C}^{N \times M}$
  $(\boldsymbol{\Lambda}_1^1, \boldsymbol{\Lambda}_2^1, \boldsymbol{\lambda}^1) = \text{INITIALIZATION}$ ▷ Zeros are fine
  for $l = 1, 2, \ldots$ **do**
    $\boldsymbol{Y}_1^{l+1} = \mathcal{T}^{\rho}_{\text{block}}(\boldsymbol{Z}^l - \boldsymbol{\Lambda}_1^l)$ ▷ Eq. (3)
    $\boldsymbol{Y}_2^{l+1} = \mathcal{P}_{\text{rank-1}}(\boldsymbol{Z}^l - \boldsymbol{\Lambda}_2^l)$ ▷ Eq. (4)
    $\boldsymbol{C} = \boldsymbol{\Lambda}_1^l + \boldsymbol{\Lambda}_2^l + \boldsymbol{Y}_1^{l+1} + \boldsymbol{Y}_2^{l+1}$
    $\boldsymbol{w} = 2(\boldsymbol{x} + \boldsymbol{\lambda}^l) - \mathcal{S}(\boldsymbol{C})\,\mathbf{1}$
    $\boldsymbol{Z}^{l+1} = \frac{1}{2}(\boldsymbol{C} + \mathcal{R}(\frac{1}{M+2}\boldsymbol{w}))$
    $\boldsymbol{\Lambda}_1^{l+1} = \boldsymbol{\Lambda}_1^l + \boldsymbol{Y}_1^{l+1} - \boldsymbol{Z}^{l+1}$
    $\boldsymbol{\Lambda}_2^{l+1} = \boldsymbol{\Lambda}_2^l + \boldsymbol{Y}_2^{l+1} - \boldsymbol{Z}^{l+1}$
    $\boldsymbol{\lambda}^{l+1} = \boldsymbol{\lambda}^l + \boldsymbol{x} - \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}$
  $\boldsymbol{s} = \sigma_1(\boldsymbol{Z}^{l+1})\,\boldsymbol{u}_1(\boldsymbol{Z}^{l+1})$ ▷ $\boldsymbol{h} = \boldsymbol{v}_1(\boldsymbol{Z}^{l+1})$
  $\boldsymbol{s} = \boldsymbol{s}/\exp(\mathrm{i}\mathrm{Arg}(h_1))$ ▷ Phase normalization
  $\boldsymbol{s} = a\,\boldsymbol{s}$ ▷ Scale recovery
  return $\boldsymbol{s}$

ここで，N, M はそれぞれ音源信号と室内インパルス応答の時間周波数領域での長さ，$\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \boldsymbol{\lambda}$ はラグランジュ乗数，$\boldsymbol{Y}_1, \boldsymbol{Y}_2$ は ADMM を適用するために式 (2) に導入した補助変数，$\mathcal{R}(\cdot)$ は $\mathcal{S}(\cdot)\,\mathbf{1}$ の随伴作用素であり $\mathcal{R}(\boldsymbol{x}) = \left[\boldsymbol{x}, \mathcal{S}_{-1}(\boldsymbol{x}), \ldots, \mathcal{S}_{-(M-1)}(\boldsymbol{x})\right]$ で表される。$\boldsymbol{Y}_1, \boldsymbol{Y}_2 \in \mathbb{C}^{N \times M}$ の更新式はそれぞれグループ軟閾値処理，ランク 1 行列への射影として以下のように与えられる。

$$\mathcal{T}^{\rho}_{\text{block}}(\boldsymbol{X})(n, :) = \left(1 - \frac{1}{\rho\,\|\boldsymbol{X}(n, :)\|_2}\right)_+ \boldsymbol{X}(n, :) \tag{3}$$

$$\mathcal{P}_{\text{rank-1}}(\boldsymbol{X}) = \sigma_1(\boldsymbol{X})\,\boldsymbol{u}_1(\boldsymbol{X})\,\boldsymbol{v}_1(\boldsymbol{X})^{\mathrm{H}} \tag{4}$$

ここで，ρ は閾値処理の効果を調整するパラメータ，$(\cdot)_+ = \max(0, \cdot)$，$(\cdot)^{\mathrm{H}}$ は共役転置であり，$\sigma_1(\cdot)$，$\boldsymbol{u}_1(\cdot)$，$\boldsymbol{v}_1(\cdot)$ はそれぞれ入力された行列の最大特異値とそれに対応する左特異ベクトル，右特異ベクトルである。ρ の値は周波数に応じて調整するのが困難であるため，9 行目で観測信号に正規化を施し，全ての周波数において単一の値を適用する。

* Effectiveness of rank-1 constraint in single-channel blind dereverberation using matrix lifting. By Fumiki YOHENA, Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).

3　ランク 1 制約の緩和

前章で述べたように，提案手法では行列をランク 1 集合へ射影することによって $\boldsymbol{x} = \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}$ で表される畳み込みモデルを保っている。しかし，ランク 1 制約は厳しい制約であるため，制約を緩和した場合の性能と比較することでランク 1 制約の妥当性を確かめる必要がある。制約を緩和する方法として，ランク k 行列への射影や，核ノルム最小化によって行列の低ランク性を誘導する方法が考えられる。

ランク k 行列への射影では，式 (4) を

$$\mathcal{P}_{\text{rank-}k}(\boldsymbol{X}) = \boldsymbol{u}_{1:k}(\boldsymbol{X})\,\boldsymbol{\Sigma}_{1:k}(\boldsymbol{X})\,\boldsymbol{v}_{1:k}(\boldsymbol{X})^{\mathrm{H}} \quad (5)$$

によって置き換える。ここで，$\boldsymbol{\Sigma}_{1:k}(\cdot)$，$\boldsymbol{u}_{1:k}(\cdot)$，$\boldsymbol{v}_{1:k}(\cdot)$ はそれぞれ入力された行列の上位 k 番目までの特異値が入った対角行列と，それに対応する左特異ベクトル，右特異ベクトルである。核ノルム最小化では，得られた全ての特異値に対して軟閾値処理を行う。これを提案手法においてランク 1 行列への射影と置き換えることで，式 (2) の非凸最適化問題が凸緩和される。

4　評価実験

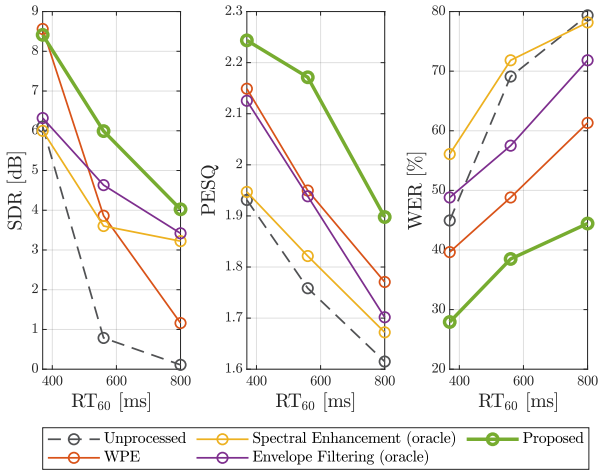
ランク 1 制約を課した提案手法と，上記の緩和手法の性能を比較する実験を行った。VCTK データセットから選んだ 5 s 以上の音声データ (男女 10 データずつ) を音源音声とし，BUT Reverb Database の室内インパルス応答 ($\mathrm{RT}_{60} \approx 370$ ms) を畳み込んだ観測信号 20 データを生成した。サンプリング周波数は 16 kHz である。ADMM の反復回数は 500 回，$\rho = 400$, $M = 50$ とした。ランク k 行列への射影においては，$k = 2, 3$ として検証を行った。核ノルム最小化における閾値は 0.5 とした。

各手法における実行時間と SDR の中央値を**表–1** に示す。核ノルム最小化に基づく手法では，各反復において全ての特異値を計算する必要があるため，計算時間が際立って長くなっている。また，ここで検証された全ての緩和手法は，従来のランク 1 制約を課した手法と比較して SDR が下回っている。この結果は，$\boldsymbol{x} = \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}$ で表される畳み込みモデルを正確に保つべきであることを示している。

また，ランク 1 制約を課した提案手法と他の単チャネルブラインド残響除去の従来手法との性能も比較した。比較手法には WPE [2]，スペクトルエンハンスメント [3]，包絡フィルタリングに基づく手法 [4] を用いた。提案手法と従来手法との SDR，PESQ，WER による比較結果を**図–1** に示す。このとき，残響時間は $\mathrm{RT}_{60} \approx 370, 560, 800$ ms であり，M の値はそれぞれ $M = 50, 70, 100$ である。ADMM の反復回数は 1000 回とした。スペクトルエンハンスメントと包絡フィルタリングでは，複素信号を合成する際の位相情報として，残響のない音源信号の位相を用いた。提案手法は音源信号の位相を用いた手法を含めて他の従来手法を上回り，有効性が確認された。

表–1　各手法における実行時間と SDR の中央値。MATLAB 2022b により実装した。

Type of method	実行時間 [s]	SDR [dB]
提案手法	**119.1**	**8.5947**
ランク 2 射影 ver.	134.6	1.2205
ランク 3 射影 ver.	155.9	-1.5974
核ノルム ver.	2525.6	-9.0376



図–1　各残響時間における従来手法との比較。SDR，PESQ，WER はそれぞれ中央値を示している。SDR，PESQ は値が大きいほど良く，WER は値が小さいほど良い。

5　むすび

本稿では，行列リフティングに基づく単チャネルブラインド残響除去手法におけるランク 1 制約を緩めた際の性能の変化を検証し，従来のランク 1 制約の有効性を確認した。今後は提案手法の高速化や，より正確な残響信号のモデル化を検討する。

参考文献

[1] 饒平名文希，矢田部浩平，“信号のスパース性とランク 1 制約を利用したブラインド残響除去”，日本音響学会講演論文集，pp. 199–200, Mar. 2023.

[2] R. Ikeshita, N. Kamo, and T. Nakatani,“Blind signal dereverberation based on mixture of weighted prediction error models,” in IEEE Signal Process. Lett., vol. 28, pp. 399–403, 2021.

[3] K. Lebart, J.-M. Boucher, and P. N. Denbigh, “A new method based on spectral subtraction for speech dereverberation,” in Acta Acust. United Acust., vol. 87, no. 3, pp. 359–366, 2001.

[4] H. Kameoka, T. Nakatani and T. Yoshioka, “Robust speech dereverberation based on non-negativity and sparse nature of speech spectrograms,” in IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP), pp. 45–48, 2009.