

スパース行列復元に基づくブラインド残響除去におけるバイアス軽減の効果*

©饒平名文希, 山田宏樹, 矢田部浩平 (農工大)

1 まえがき

単チャンネルブラインド残響除去とは、単チャン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技術である。我々はこれまで、スパース行列復元と行列リフティングを組み合わせた手法 [1] を提案した。この手法ではグループ軟閾値関数を用いて行列をスパースにしていたが、全ての要素の振幅を縮小する軟閾値関数は残響除去において不利なバイアスを持っていると考えられる。そこで本稿では、バイアスを軽減する目的で硬閾値関数を近似した連続関数を導入し、その性能面における効果を検証した。

2 スパース行列復元に基づく手法

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観測信号 $\mathbf{x}_f \in \mathbb{C}^N$ を

$$\mathbf{x}_f = \mathbf{s}_f * \mathbf{h}_f \quad (1)$$

として近似する。ここで、 $\mathbf{s}_f \in \mathbb{C}^N$ および $\mathbf{h}_f \in \mathbb{C}^M$ 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。また 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周波数方向の長さである。このとき、各周波数ビンで $\mathbf{s}_f, \mathbf{h}_f$ の両方を同時に推定し、それぞれ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。以降、周波数インデックス f は省略する。

2つの変数を同時に扱うことは困難であるため、行列リフティングにより式 (1) を $\mathbf{x} = \mathbf{s} * \mathbf{h} = \mathcal{S}(\mathbf{Z}) \mathbf{1}$ として書き換える。ここで、 \mathbf{Z} は $\mathbf{s}\mathbf{h}^T$ 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。音源信号 \mathbf{s} 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、 \mathbf{Z} は行ごとにスパース性を持つ。以上のことから、残響除去の問題をグループスパースなランク 1 行列復元問題として以下のように定式化する。

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

ここで、 $\|\cdot\|_{2,1}$ は $\ell_{2,1}$ 混合ノルム、 $\text{rank}(\mathbf{Z}) = 1$ は \mathbf{Z} をランク 1 行列全体の集合に帰属させる非凸

Algorithm 1 Proposed Method

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

- 1: **Function** DEREVERB_1CH(\mathbf{x}, M, ρ)
- 2: $\mathbf{X} = \text{STFFT}(\mathbf{x})$ $\triangleright \mathbf{X} \in \mathbb{C}^{F \times N}$
- 3: **for** $f = 1, 2, \dots, F$ **do**
- 4: $\mathcal{S}(f, \cdot) = \text{ADMM.SOLVER}(\mathbf{X}(f, \cdot)^T, M, \rho)^T$
- 5: $\mathbf{s} = \text{iSTFFT}(\mathcal{S})$
- 6: **return** \mathbf{s}
- 7: **Function** ADMM.SOLVER(\mathbf{x}, M, ρ)
- 8: $a = \|\mathbf{x}\|_2$ $\triangleright \mathbf{x} \in \mathbb{C}^N$
- 9: $\mathbf{x} = \mathbf{x}/a$ $\triangleright \ell_2$ normalization
- 10: $\mathbf{Z}^1 = \mathbf{x} [1, 0, 0, \dots, 0]$ $\triangleright \mathbf{Z}^1 \in \mathbb{C}^{N \times M}$
- 11: $(\Lambda_1^1, \Lambda_2^1, \lambda^1) = \text{INITIALIZATION}$ \triangleright Zeros are fine
- 12: **for** $l = 1, 2, \dots$ **do**
- 13: $\mathbf{Y}_1^{l+1} = \mathcal{T}_{\text{block}}^\rho(\mathbf{Z}^l - \Lambda_1^l)$ \triangleright Eq. (3)
- 14: $\mathbf{Y}_2^{l+1} = \mathcal{P}_{\text{rank-1}}(\mathbf{Z}^l - \Lambda_2^l)$
- 15: $\mathbf{C} = \Lambda_1^l + \Lambda_2^l + \mathbf{Y}_1^{l+1} + \mathbf{Y}_2^{l+1}$
- 16: $\mathbf{w} = 2(\mathbf{x} + \lambda^l) - \mathcal{S}(\mathbf{C}) \mathbf{1}$
- 17: $\mathbf{Z}^{l+1} = \frac{1}{2}(\mathbf{C} + \mathcal{R}(\frac{1}{M+2}\mathbf{w}))$
- 18: $\Lambda_1^{l+1} = \Lambda_1^l + \mathbf{Y}_1^{l+1} - \mathbf{Z}^{l+1}$
- 19: $\Lambda_2^{l+1} = \Lambda_2^l + \mathbf{Y}_2^{l+1} - \mathbf{Z}^{l+1}$
- 20: $\lambda^{l+1} = \lambda^l + \mathbf{x} - \mathcal{S}(\mathbf{Z}) \mathbf{1}$
- 21: $\mathbf{s} = \sigma_1(\mathbf{Z}^{l+1}) \mathbf{u}_1(\mathbf{Z}^{l+1})$ $\triangleright \mathbf{h} = \mathbf{v}_1(\mathbf{Z}^{l+1})$
- 22: $\mathbf{s} = \mathbf{s} / \exp(i\text{Arg}(h_1))$ \triangleright Phase normalization
- 23: $\mathbf{s} = a \mathbf{s}$ \triangleright Scale recovery
- 24: **return** \mathbf{s}

制約である。問題 (2) に交互方向乗数法 (ADMM) を適用すると Algorithm 1 が得られる [1]。ここで、 $\Lambda_1, \Lambda_2, \lambda$ はラグランジュ乗数、 $\mathbf{Y}_1, \mathbf{Y}_2$ は補助変数、 $\mathcal{R}(\cdot)$ は $\mathcal{S}_\tau(\cdot) \mathbf{1}$ の随伴作用素であり $\mathcal{R}(\mathbf{x}) = [\mathbf{x}, \mathcal{S}_{-1}(\mathbf{x}), \dots, \mathcal{S}_{-(M-1)}(\mathbf{x})]$ として表される。 τ はシフト量を示す。 $\mathbf{Y}_1 \in \mathbb{C}^{N \times M}$ の更新に用いる作用素 $\mathcal{T}_{\text{block}}^\rho$ は、 $\ell_{2,1}$ 混合ノルムの近接作用素であるグループ軟閾値処理として以下のように与えられる。

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

ここで、 ρ は閾値処理の効果を調整するパラメータ、 $(\cdot)_+ = \max(0, \cdot)$ である。

3 スパース正則化におけるバイアスの軽減

前章で述べたように、提案手法のアルゴリズムではグループ軟閾値関数を用いることで行列のグループスパース性を誘導している。この処理により、残響によって生じた余分な成分を取り除いている。しかし、軟閾値関数は入力された信号の全ての要素の振幅を縮小するため、音響信号に含まれる残響成分だけで

*Effect of bias reduction in blind dereverberation based on sparse matrix recovery. By Fumiki YOHENA, Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).

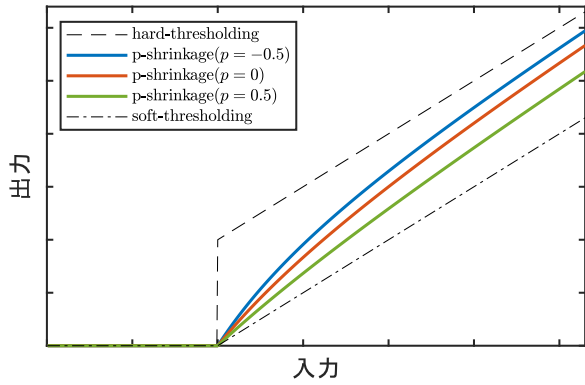


図-1 軟閾値関数 (soft-thresholding), 硬閾値関数 (hard-thresholding), p -shrinkage の概形。 p -shrinkage は, $p = 1$ で軟閾値関数と一致し, p が 1 より小さくなるにつれて硬閾値関数に近づく。

なく直接音成分にまで影響を与えてしまう。そこで今回は, 軟閾値関数の代わりに, バイアスを軽減した閾値関数を用いた処理に置き換える。

3.1 p -shrinkage による硬閾値関数の近似

軟閾値関数と硬閾値関数の概形を図-1 に示す。硬閾値処理は閾値以上の値を保持できる一方で出力信号が不連続となる。軟閾値処理は連続性を保つが, 閾値以上の入力値も 0 の方向に縮小する作用が働く。このバイアスは, 音響信号中の直接音成分も小さくするため, 残響除去に対して不利に働く可能性がある。そこで, 両者の欠点を補うような閾値処理が望まれる。

これまで, バイアスを軽減する目的で硬閾値関数を近似した閾値処理がいくつか提案されてきた [2, 3]。今回は, 軟閾値関数と硬閾値関数を一般化した p -shrinkage を用いる。 p -shrinkage の概形を図-1 に示す。パラメータ p を調節することにより, この閾値関数の概形は硬・軟閾値関数の概形を補間し, 両者の特性を併せ持つ。図-1 から分かるように, p -shrinkage は連続でありながらバイアスが小さく, 軟閾値関数に比べて直接音成分に対して小さい影響しか与えないと期待される。 p -shrinkage をグループ閾値関数に適用した際の作用素 $\mathcal{T}_{\text{block}}^{\rho,p}$ は以下のように与えられる。

$$\mathcal{T}_{\text{block}}^{\rho,p}(\mathbf{X})(n,:) = \left(1 - \left(\frac{1}{\rho \|\mathbf{X}(n,:) \|_2}\right)^{2-p}\right)_+ \mathbf{X}(n,:) \quad (4)$$

この処理を Algorithm 1 の 13 行目で Eq. (3) の代わりに用いて, バイアスを軽減した手法を実現した。

4 評価実験

スパース行列復元に基づく手法において, 軟閾値関数と p -shrinkage それぞれを用いた場合の残響除去性能を比較する実験を行った。実験データとして, VCTK

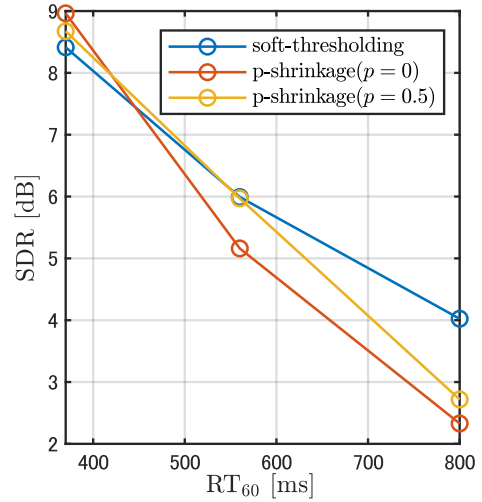


図-2 軟閾値関数と p -shrinkage それぞれを用いた場合の比較。 SDR の中央値を示しており, 値が大きいほど良い。

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) に対して, BUT Reverb Database の室内インパルス応答 ($RT_{60} \approx 370, 560, 800$ ms) をそれぞれ畳み込んだ観測信号 60 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 1000 回, $\rho = 400$, M の値は残響時間が短いものから順にそれぞれ $M = 50, 70, 100$ とした。また p -shrinkage の調整パラメータ p は 0, 0.5 とした。

SDR による比較結果を図-2 に示す。 $RT_{60} \approx 370$ ms のときは p -shrinkage を用いた場合で SDR が向上し, バイアス軽減の効果が見られた。一方で, $RT_{60} \approx 560, 800$ ms のときは軟閾値処理を行った場合よりも SDR が下回った。これは, バイアスを軽減したことで, 残響時間が長いサンプルにおける強い残響成分を取り除く効果が低下したためであると考えられる。

5 むすび

本稿では, スパース行列復元に基づく単チャンネルブラインド残響除去手法において, スパース正則化のバイアスを軽減した際の性能面での効果を検証した。評価実験の結果, 残響時間が短い音響信号でのバイアス軽減の効果を確認できた。今後は残響時間が長い音響信号に対するパラメータの調整や, 他の閾値関数の適用を検討する。

参考文献

- [1] F. Yohena and K. Yatabe, "Single-channel blind dereverberation based on rank-1 matrix lifting in time-frequency domain," in IEEE Int. Conf. Acoust., Speech Signal Proc. (ICASSP), 2024.
- [2] M. Kowalski, "Thresholding rules and iterative shrinkage/thresholding algorithm: A convergence study," in IEEE Int. Conf. Image Proc. (ICIP), pp. 4151–4155, 2014.
- [3] S. Voronin and R. Chartrand, "A new generalized thresholding algorithm for inverse problems with sparsity constraints," in IEEE Int. Conf. Acoust., Speech Signal Proc. (ICASSP), pp. 1636–1640, 2013.