

信号のスパース性と rank-1 制約を利用したブラインド残響除去*

☆饒平名文希, 矢田部浩平 (農工大)

1 まえがき

ブラインド残響除去は、残響音声のみを入力として残響を抑圧する技術である。現在では残響成分を予測するフィルタを推定する手法 [1] が主流となっているが、単チャンネル信号からフィルタを精度良く推定することは難しい。そこで本稿では最適化に基づいて、単チャンネル観測信号から音源信号を直接推定する手法を提案する。評価実験により、特に残響時間が長い場合における提案手法の有効性を確認した。

2 問題設定

単一のマイクロホンで音響信号を観測する場合を考える。観測した信号に短時間フーリエ変換を適用し、時間周波数領域にて信号を扱う。\$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を

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

で近似的に考える。ここで、\$\mathbf{x}_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スペクトログラムにおける \$f\$ 番目の周波数ビンでの時系列信号を、* は時間方向の巡回畳み込みをそれぞれ表している。この観測信号 \$\mathbf{y}_f \in \mathbb{C}^N\$ のみから、\$\mathbf{x}_f\$ と \$\mathbf{h}_f\$ 両方を推定する問題を考える。以降、周波数ビンごと独立に処理を行うため、周波数インデックス \$f\$ は省略する。

3 提案手法

2章で述べた問題を最適化問題として定式化して解く。音響信号の時間周波数領域でのスパース性を仮定するが、\$\mathbf{x}\$ の \$l_1\$ ノルム最小化問題として定式化した場合、\$\mathbf{x}\$ と \$\mathbf{h}\$ にスケールの不定性を残してしまう。このスケールの不定性によって生じる最適化の困難さを回避するため、畳み込み演算を行列 \$\mathbf{Z} := \mathbf{x}\mathbf{h}^T\$ を用いて表現する。これにより、仮定を \$\mathbf{x}\$ のスパース性から行列 \$\mathbf{Z}\$ のグループスパース性に置き換え、rank-1 制約の下で \$\mathbf{Z}\$ の混合 \$l_{2,1}\$ ノルムを最小化する非凸最適化問題として定式化する [2]。

3.1 非凸最適化問題としての定式化

rank-1 行列 \$\mathbf{Z}\$ を用いて、式 (1) を以下のように書き換える。

$$\mathbf{y} = \mathbf{x} * \mathbf{h} = \mathcal{S}_{\text{CL}}(\mathbf{Z})\mathbf{1} \quad (2)$$

Algorithm 1 Proposed method

Input: \$\mathbf{y} \in \mathbb{C}^N, M\$

Output: \$\mathbf{x} \in \mathbb{C}^N\$

- 1: While exit condition unsatisfied
- 2: Update \$\mathbf{P}, \mathbf{Q}, \mathbf{Z}\$ by (5)–(7)
- 3: \$\mathbf{\Lambda}^{l+1} = \mathbf{\Lambda}_1^l + \mathbf{Z}^{l+1} - \mathbf{P}^{l+1} \in \mathbb{R}^{N \times M}\$
- 4: \$\mathbf{\Lambda}_2^{l+1} = \mathbf{\Lambda}_2^l + \mathbf{Z}^{l+1} - \mathbf{Q}^{l+1} \in \mathbb{R}^{N \times M}\$
- 5: \$\lambda_0^{l+1} = \lambda_0^l + \mathcal{S}_{\text{CL}}(\mathbf{Z})\mathbf{1} - \mathbf{y} \in \mathbb{R}^N\$
- 6: Perform SVD \$\mathbf{Z} = \sigma\mathbf{u}\mathbf{v}^T\$ to obtain \$\mathbf{x} = \sigma\mathbf{u}, \mathbf{h} = \mathbf{v}\$
- 7: \$\mathbf{x} = \mathbf{x} / \exp(i\text{Arg}(h_1))\$

ここで、\$\mathbf{1} \in \mathbb{R}^M\$ は要素が全て 1 のベクトル、\$\mathcal{S}_{\text{CL}}(\mathbf{Z})\$ は \$\mathbf{Z}\$ の周期的なシフトを行う線形作用素である。

$$\mathcal{S}_{\text{CL}}(\mathbf{Z}) = [\mathbf{Z}_{:,1}, \mathcal{S}_1(\mathbf{Z}_{:,2}), \dots, \mathcal{S}_{M-1}(\mathbf{Z}_{:,M})]$$

ただし、\$\mathcal{S}_\tau(\cdot)\$ は列ベクトルを \$\tau\$ の分だけ下方向へ巡回シフトする演算、\$\mathbf{Z}_{:,m}\$ は \$\mathbf{Z}\$ の \$m\$ 番目の列ベクトルである。

式 (2) を満たす rank-1 行列 \$\mathbf{Z}\$ のなかで、全ての要素が 0 である行ベクトルを最も多く持つ \$\mathbf{Z}\$ を推定する次の非凸最適化問題を考える。

$$\min_{\mathbf{Z}} \|\mathbf{Z}\|_{2,1} \text{ s.t. } \mathbf{y} = \mathcal{S}_{\text{CL}}(\mathbf{Z})\mathbf{1}, \mathbf{Z} \in \mathcal{R}_1 \quad (3)$$

ここで、\$\|\cdot\|_{2,1}\$ は各行ベクトルの \$l_2\$ ノルムを並べたベクトルの \$l_1\$ ノルムであり、\$\mathbf{Z} \in \mathcal{R}_1\$ は \$\mathbf{Z}\$ を rank-1 行列に帰属させる非凸制約である。式 (3) を交互方向乗数法 (ADMM) を用いて解く。

3.2 ADMM による解法

ADMM を適用するため、補助変数 \$\mathbf{P}, \mathbf{Q}\$ を導入し、式 (3) を以下のように変形する。

$$\begin{aligned} \min_{\mathbf{P}, \mathbf{Q}, \mathbf{Z}} \|\mathbf{P}\|_{2,1} + \iota_{\text{rank}(\cdot) \leq 1}(\mathbf{Q}) \\ \text{s.t. } \mathbf{y} = \mathcal{S}_{\text{CL}}(\mathbf{Z})\mathbf{1}, \mathbf{P} = \mathbf{Z}, \mathbf{Q} = \mathbf{Z} \end{aligned} \quad (4)$$

ただし、\$\iota_{\text{rank}(\cdot) \leq 1}(\mathbf{Q})\$ は rank-1 制約をコスト関数に組み込むための指示関数である。

$$\iota_{\text{rank}(\cdot) \leq 1}(\mathbf{Q}) := \begin{cases} 0 & \text{if } \text{rank}(\mathbf{Q}) \leq 1 \\ +\infty & \text{otherwise} \end{cases}$$

この問題に ADMM を適用した提案手法のアルゴリズムを Algorithm 1 に示す。ただし、2 行目は \$\mathbf{P}, \mathbf{Q}, \mathbf{Z}\$ それぞれについての拡張ラグランジュ関数の最小化で与えられ、以下のように更新する。

*Blind dereverberation with signal sparsity and rank-1 constraint. By Fumiki YOHENA and Kohei YATABE (Tokyo University of Agriculture and Technology).

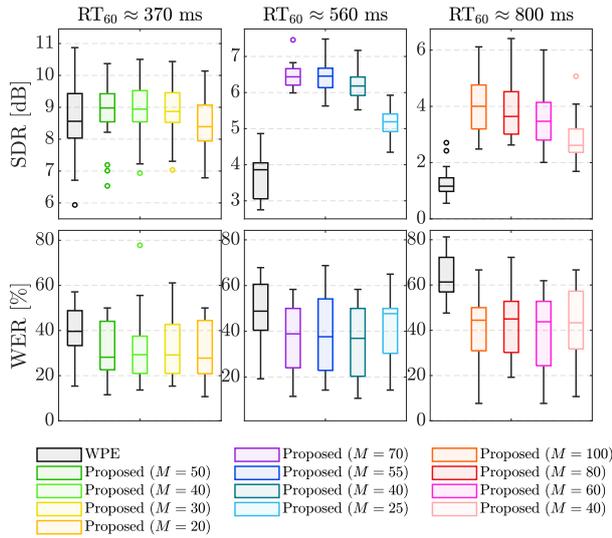


図-1 SDR (上段) と WER (下段) による性能比較

$$\mathbf{P}_{n,:}^{l+1} = \max \left(0, 1 - \frac{1}{\rho \|\hat{\mathbf{P}}_{n,:}^l\|_2} \right) \hat{\mathbf{P}}_{n,:}^l \quad (5)$$

$$\mathbf{Q}^{l+1} = \sigma_1 \mathbf{U}_{:,1} \mathbf{V}_{:,1}^T \quad (6)$$

$$\mathbf{Z}^{l+1} = \frac{1}{2} \mathcal{S}_{\text{-CL}} \left(\left(\mathcal{S}_{\text{CL}}(\mathbf{C}_1^l + \mathbf{C}_2^l) + \mathbf{C}_3^l \right) \mathbf{W} \right) \quad (7)$$

式 (5) において、 l は反復回数、 ρ は正則化項の重みを調整するパラメータ、 $\hat{\mathbf{P}}^l := \mathbf{Z}^l + \Lambda_1^l$ である。式 (6) は $\hat{\mathbf{Q}}^l := \mathbf{Z}^l + \Lambda_2^l$ の特異値分解 $\hat{\mathbf{Q}} = \mathbf{U} \text{diag}(\boldsymbol{\sigma}) \mathbf{V}^T$ により計算され、 $(\cdot)^T$ は転置を示す。式 (7) において、 $\mathbf{C}_1^l := \mathbf{P}^{l+1} - \lambda_1^l$ 、 $\mathbf{C}_2^l := \mathbf{Q}^{l+1} - \lambda_2^l$ 、 $\mathbf{C}_3^l := [\mathbf{c}_3^l, \dots, \mathbf{c}_3^l] \in \mathbb{R}^{N \times M}$ 、 $\mathbf{c}_3^l := \mathbf{y} - \lambda_0^l$ 、 $\mathcal{S}_{\text{-CL}}(\cdot)$ は上方向への巡回シフトである。 \mathbf{W} は $M \times M$ の単位行列 \mathbf{I}_M を用いて $\mathbf{W} = \mathbf{I}_M - \mathbf{1}\mathbf{1}^T / (M+2)$ で与えられる。

Algorithm 1 の 6 行目では、最適化の結果として得られた \mathbf{Z} の特異値分解から \mathbf{x} を求めている。ただし、ここで得られる信号 \mathbf{x} は各周波数で位相がずれているため、7 行目の処理を施して全ての周波数で位相を揃える必要がある。Algorithm 1 の処理を各周波数で行い、逆短時間フーリエ変換によってまとめて時間領域に戻すことで、残響が抑圧された信号を得る。

4 評価実験

提案手法の有効性を確認するため、単チャンネルでも効果的に動作する手法である WPE [1] と残響除去性能を比較した。評価指標には SDR と WER を用いた。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 個ずつ) を音源音声とし、BUT Reverb Database の実測の室内インパルス応答 3 種類 ($\text{RT}_{60} \approx 370, 560, 800$ ms) を畳み込んだ 60 個の観測信号を生成した。サンプリング周波数は 16 kHz

表-1 各遅延フレーム数における SDR 比較

		$D = 0$	$D = 1$	$D = 2$	$D = 3$
WPE	370 ms	—	7.22	8.56	8.80
	560 ms	—	3.97	3.86	3.45
	800 ms	—	0.35	1.16	1.59
Prop.	370 ms	8.98	8.92	9.24	8.79
	560 ms	6.43	5.14	5.10	4.89
	800 ms	4.00	3.46	2.80	2.00

である。推定する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 (それぞれ N と M) は事前に設定する必要があるが、 N の値は観測信号の長さとし、 M の値は各残響時間において 4 パターンで設定した。ADMM の反復回数は 1000 回とし、 ρ の値は観測信号の l_1 ノルムに応じて、各周波数で 5~80 の間を滑らかに変化するようなガウス分布で設定した。また、従来手法と提案手法における遅延フレーム数を変化させた場合の性能の変化についても検証した。

4.1 実験結果

各残響時間における SDR と WER の比較結果を図-1 に示す。提案手法と従来手法の遅延フレーム数はそれぞれ $D = 0, 2$ である。370 ms において、提案手法の SDR および WER は従来手法のものと同程度であった一方で、560 ms と 800 ms ではいずれのパラメータにおいても SDR は上回り WER は下回った。このことは、残響時間が長い音声信号に対して、提案手法の処理は有効であることを示す。

続いて、各遅延フレーム数における SDR の中央値による評価結果を表-1 に示す。このとき、残響時間が短いものから順に $M = 50, 70, 100$ である。提案手法では遅延を付加しない場合で高い SDR が得られた。この結果から、提案手法は性能の安定性のために遅延の導入を必要としないことを確認した。

5 むすび

本稿では、信号のスパース性と rank-1 近似に基づく単チャンネルブラインド残響除去手法を提案し、評価実験によってその有効性を確かめた。今後は雑音を含んだ観測信号に対する残響除去性能の評価を行う。

参考文献

- [1] R. Ikeshita, N. Kamo, and T. Nakatani, "Blind signal dereverberation based on mixture of weighted prediction error models," *IEEE Signal Process. Lett.*, vol. 28, pp. 399–403, 2021.
- [2] C. Cheng and W. Dai, "Short-and-sparse deconvolution via rank-one constrained optimization (ROCO)," *IEEE ICASSP*, pp. 5882–5886, 2022.