重み付きスパース行列復元に基づく単チャネルブラインド残響除去*

◎饒平名文希，山田宏樹，矢田部浩平 (農工大)

1 はじめに

単チャネルブラインド残響除去とは，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制する技術である。我々はこれまで，スパース行列復元と行列リフティングの組み合わせによる手法 [1] を提案した。この手法では行列のスパース性誘導にグループソフト閾値処理を適用しており，大きい要素が過剰に抑制されてしまうことが課題であった。そこで本稿では，この過剰な抑制を緩和するためにスパース性誘導に重みを導入し，性能面での効果を検証した。

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 $\boldsymbol{x}_f \in \mathbb{C}^N$ を以下のように近似する。

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

ここで，$\boldsymbol{s}_f \in \mathbb{C}^N$ 及び $\boldsymbol{h}_f \in \mathbb{C}^M$ 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。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時間方向の長さである。このとき，各周波数ビンで $\boldsymbol{s}_f$，$\boldsymbol{h}_f$ の両方を同時に推定し，それぞれを逆短時間フーリエ変換して時間領域での音源信号とインパルス応答を得る。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 $\boldsymbol{x} = \boldsymbol{s} * \boldsymbol{h} = \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}$ として書き換える。ここで，$\boldsymbol{Z}$ は $\boldsymbol{s}\boldsymbol{h}^{\mathrm{T}}$ 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。音源信号 $\boldsymbol{s}$ 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，$\boldsymbol{Z}$ 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。

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

ここで，$\|\cdot\|_{2,1}$ は $\ell_{2,1}$ 混合ノルム，$\mathrm{rank}(\boldsymbol{Z}) = 1$ は $\boldsymbol{Z}$ をランク 1 行列全体の集合に帰属させる非凸制約である。式 (2) に交互方向乗数法 (ADMM) を適用すると Algorithm 1 が得られる [1]。$\mathcal{R}(\cdot)$ は $\mathcal{S}_\tau(\cdot)\,\mathbf{1}$ の随伴作用素であり，$\mathcal{R}(\boldsymbol{x}) = \left[\boldsymbol{x}, \mathcal{S}_{-1}(\boldsymbol{x}), \ldots, \mathcal{S}_{-(M-1)}(\boldsymbol{x})\right]$ として表される。τ はシフト量を示す。$\boldsymbol{Y}_1 \in \mathbb{C}^{N \times M}$ の更新に用いる作用素 $\mathcal{T}^{\rho}_{\text{block}}$ は，$\ell_{2,1}$ 混合ノルムの近接作用素であるグループソフト閾値処理として

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

で与えられる。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。

Algorithm 1 ADMM algorithm

Input: $\boldsymbol{x} \in \mathbb{C}^N, M \in \mathbb{N}, \rho > 0$
Output: $\hat{\boldsymbol{s}} \in \mathbb{C}^N$

$\boldsymbol{x} = \boldsymbol{x}/\|\boldsymbol{x}\|_2$ ▷ ℓ_2 normalization
$(\boldsymbol{Z}^1, \boldsymbol{\Lambda}_1^1, \boldsymbol{\Lambda}_2^1, \boldsymbol{\lambda}^1) = \textsc{Initialization}$
for $k = 1, 2, \ldots$ **do**
  $\boldsymbol{Y}_1^{k+1} = \mathcal{T}^{\rho}_{\text{block}}(\boldsymbol{Z}^k - \boldsymbol{\Lambda}_1^k)$ ▷ Eq. (3)
  $\boldsymbol{Y}_2^{k+1} = \mathcal{P}_{\text{rank-1}}(\boldsymbol{Z}^k - \boldsymbol{\Lambda}_2^k)$
  $\boldsymbol{C} = \boldsymbol{\Lambda}_1^k + \boldsymbol{\Lambda}_2^k + \boldsymbol{Y}_1^{k+1} + \boldsymbol{Y}_2^{k+1}$
  $\boldsymbol{w} = 2(\boldsymbol{x} + \boldsymbol{\lambda}^k) - \mathcal{S}(\boldsymbol{C})\,\mathbf{1}$
  $\boldsymbol{Z}^{k+1} = \frac{1}{2}(\boldsymbol{C} + \mathcal{R}(\frac{1}{M+2}\boldsymbol{w}))$
  $\boldsymbol{\Lambda}_1^{l+1} = \boldsymbol{\Lambda}_1^l + \boldsymbol{Y}_1^{l+1} - \boldsymbol{Z}^{l+1}$
  $\boldsymbol{\Lambda}_2^{l+1} = \boldsymbol{\Lambda}_2^l + \boldsymbol{Y}_2^{l+1} - \boldsymbol{Z}^{l+1}$
  $\boldsymbol{\lambda}^{l+1} = \boldsymbol{\lambda}^l + \boldsymbol{x} - \mathcal{S}(\boldsymbol{Z})\,\mathbf{1}$
$\hat{\boldsymbol{s}} = \sigma_1(\boldsymbol{Z}^{k+1})\,\boldsymbol{u}_1(\boldsymbol{Z}^{k+1})$ ▷ $\boldsymbol{h} = \boldsymbol{v}_1(\boldsymbol{Z}^{k+1})$
$\hat{\boldsymbol{s}} = \hat{\boldsymbol{s}}/\exp(\mathrm{i}\mathrm{Arg}(h_1))$ ▷ Phase normalization
$\hat{\boldsymbol{s}} = \|\boldsymbol{x}\|_2\,\hat{\boldsymbol{s}}$ ▷ Scale recovery

3 重み付きスパース正則化

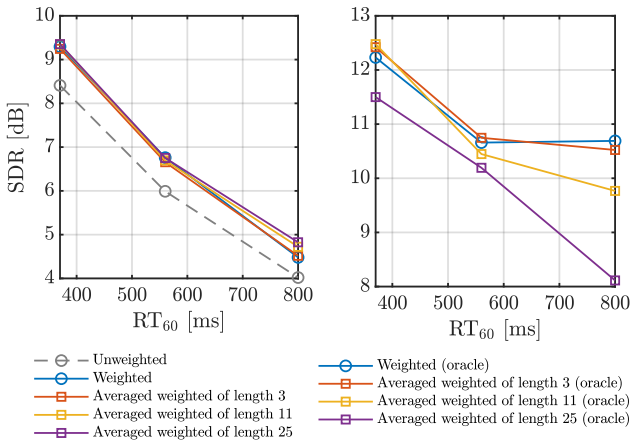
提案手法では式 (3) を用いて行列のスパース性を誘導しているが，全ての要素が閾値の分だけ均一に縮小されるため，大きな要素にも過剰にペナルティが課されてしまう。このような ℓ_1 ノルム正則化におけるバイアスを軽減するために，各要素に対して閾値の大きさを個別に調整する重み付き ℓ_1 ノルム正則化が提案されている [2]。これを提案手法に用いて，直接音成分には大きなペナルティを課さずに残響成分を抑制するような重みを設計することで，良い推定結果が得られることが期待される。

式 (2) を重み付き $\ell_{2,1}$ 混合ノルム最小化問題として再度定式化すると以下のようになる。

$$\min_{\boldsymbol{Z} \in \mathbb{C}^{N \times M}} \|\boldsymbol{w} \odot \boldsymbol{Z}\|_{2,1} \ \text{s.t.}\ \boldsymbol{x} = \mathcal{S}(\boldsymbol{Z})\,\mathbf{1},\ \mathrm{rank}(\boldsymbol{Z}) = 1 \tag{4}$$

ここで，$\boldsymbol{w} \in \mathbb{R}^N$ は非負の重みベクトル，$\odot$ は $\boldsymbol{w}$ と $\boldsymbol{Z}$ の各列の要素ごとの乗算である。この問題におけ

*Single-channel blind dereverberation based on weighted sparse matrix recovery. By Fumiki YOHENA, Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).



図–1　(左) 重み無し・重み付き・平均した重み付きを適用した際の性能比較。(右) 真の音源信号を使った重み付き・平均した重み付きを適用した際の性能比較。

る作用素 $\mathcal{T}^{\rho}_{\mathrm{block}}$ は以下のように与えられる。

$$\mathcal{T}^{\rho,w}_{\mathrm{block}}(\boldsymbol{X})(n,:) = \left(1 - \frac{w_n}{\rho \|\boldsymbol{X}(n,:)\|_2}\right)_+ \boldsymbol{X}(n,:) \quad (5)$$

この処理を Algorithm 1 の 4 行目で式 (3) の代わりに用いて，重み付きスパース正則化を実現する。

3.1　重みの設計方法

直接音成分に対する重みを小さく，残響成分に対する重みを大きくするには，目的信号 $\boldsymbol{s}$ の逆数を用いることが理想的であるが，実際には目的信号は未知であるため，代わりに反復途中で得られた推定信号を用いる。実験では以下の 2 通りで重みを設計した。

i ADMM の初期の反復では $\boldsymbol{w} = \mathbf{1}$ とし，500 反復終了後，その時点で得られた解 $\hat{\boldsymbol{s}}$ に基づいて

$$\boldsymbol{w}^{k+1} = \frac{1}{\mathrm{rescale}(\hat{\boldsymbol{s}}^k,\ l,\ u)} \quad (6)$$

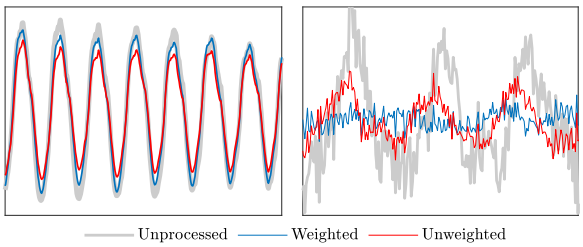
として更新する。以降，100 反復ごとにこの更新を繰り返す。rescale は $\hat{\boldsymbol{s}}^k$ の全ての要素を区間 $[l,\ u]$ にスケーリングする関数である。

ii i の重みを周波数方向に平均化する。ここでは各周波数ビンごとに決定した重みと，その上下に隣接する周波数ビンの重みで平均を取る。

音響信号を各周波数ビン独立に処理する提案手法では，信号の周波数間の連続性が考慮されない。これに対して，重みを周波数方向に平均化することで周波数間の不連続性を緩和し，周波数方向に一貫性のある処理を行うことが期待できる。

4　評価実験

スパース行列復元に基づく手法において，重み無し・重み付き (i)・平均した重み付き (ii) でそれぞれ処理を行った際の性能を比較する実験を行った。実験



図–2　VCTK/310_023.wav に 560 ms の残響を重畳した信号の波形，及び重み付き・重み無しで処理した後の波形の一部。左に直接音成分，右に残響成分を示す。

データとして，VCTK データセットの音声データ (男女 10 データずつ) と，BUT Reverb Database の室内インパルス応答 ($\mathrm{RT}_{60} \approx 370,\ 560,\ 800$ ms) との畳み込みで得られる観測信号 60 データを用いた。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 RT_{60} が短いものから順にそれぞれ $M = 50,\ 70,\ 100$ とした。また式 (6) で $l = 0.5,\ u = 10$ とした。

SDR による比較結果を**図–1** (左) に示す。重み付きの場合は，個別に閾値を調整した効果により全ての RT_{60} 条件下において重み無しの場合よりも高い SDR を実現した。平均した重み付きの場合は，$\mathrm{RT}_{60} = 370,\ 800$ の条件下で重み付きの場合よりもやや向上した。また**図–2** から分かるように，重み無しで処理した信号では直接音成分の振幅が減衰しているのに対し，重み付きの場合は直接音成分を比較的保持しながら残響成分をより軽減できている。

また，周波数方向に平均した重みを適用した際の効果について，簡単のため真の音源信号を使って理想的な重みを設計した際の検証を行った。ここでは真の音源信号を使って式 (6) に基づいて重みを初期化し，全ての反復で用いた。また $l = 0.01,\ u = 5$ とした。比較結果を図–1 (右) に示す。理想的な重みを平均化した場合は，平均を取る長さと RT_{60} が短い条件下で高い SDR を実現した。

5　むすび

本稿では，スパース行列復元に基づく単チャネルブラインド残響除去において重みを導入することを提案し，性能面での効果を確認した。また，重みを平均化することで周波数方向に一貫した処理が可能となることを示した。今後は重みを介さずに複数周波数同時に最適化する手法を検討する。

参考文献

[1]　F. Yohena and K. Yatabe, “Single-channel blind dereverberation based on rank-1 matrix lifting in time-frequency domain,” *IEEE Int. Conf. Acoust., Speech Signal Proc. (ICASSP)*, 891–895 (2024).

[2]　E. Candès, M. Wakin and S. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *J Fourier Anal*, 877–905 (2008).