

ℓ_1/ℓ_2 分数ペナルティによるスパース正則化を用いた Audio Declipping*

☆ 角田清香 (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

音響機器に振幅の大きな信号が入力されると、振幅制限を超えた部分で歪みが生じる。このような歪みを除去する処理を audio declipping と呼ぶ。Audio declipping は、時間周波数領域における真の信号のスパース性を仮定することで最適化問題として定式化できる。代表的な手法としては ℓ_1 ノルム最小化に基づくものがあるが、 ℓ_1 ノルムは信号のスケールに比例するペナルティであり、復元信号の振幅が小さいほど目的関数値が下がってしまう。この性質は、真の信号の振幅が大きい部分を推定するという declipping の問題設定と矛盾したものであり、結果として復元性能が損なわれることが知られている [1]。本稿では、スケール不変なスパースペナルティである ℓ_1/ℓ_2 分数ペナルティを用いることで上記の問題への対処を試みた。結果として、 ℓ_1 ノルムより高い復元性能が得られた一方で、 ℓ_1/ℓ_2 分数ペナルティに特有の課題が確認された。

2 スパース最適化に基づく declipping

本稿では hard-clipping を扱う [1]。時間領域における真の信号を $\mathbf{x}^{(\text{GT})} = [x^{(\text{GT})}[1], \dots, x^{(\text{GT})}[T]] \in \mathbb{R}^T$ とすると、clipping 後の観測信号 $\mathbf{x}^{(\text{obs})} = [x^{(\text{obs})}[1], \dots, x^{(\text{obs})}[T]] \in \mathbb{R}^T$ の各要素は閾値 τ により

$$x^{(\text{obs})}[t] = \begin{cases} \tau & (x^{(\text{GT})}[t] \geq \tau) \\ x^{(\text{GT})}[t] & (-\tau < x^{(\text{GT})}[t] < \tau) \\ -\tau & (x^{(\text{GT})}[t] \leq -\tau) \end{cases} \quad (1)$$

のように表せる。ただし、 t ($1 \leq t \leq T$) は時刻を表すインデックスである。式 (1) の通り、観測信号 $\mathbf{x}^{(\text{obs})}$ のインデックスは 3 つの集合 $H = \{t | x[t] \geq \tau\}$, $R = \{t | -\tau < x[t] < \tau\}$, $L = \{t | x[t] \leq -\tau\}$ に分割できる。これらを用いて観測信号 $\mathbf{x}^{(\text{obs})}$ に矛盾しない時間領域信号の集合

$$\Gamma = \left\{ \mathbf{x} \in \mathbb{R}^T \left| \begin{array}{ll} x[t] \geq \tau & (t \in H) \\ x[t] = x^{(\text{obs})}[t] & (t \in R) \\ x[t] \leq -\tau & (t \in L) \end{array} \right. \right\} \quad (2)$$

を定義することで、audio declipping は

$$\underset{\mathbf{x} \in \mathbb{R}^T}{\text{minimize}} \mathcal{F}(\mathbf{G}\mathbf{x}) \quad \text{s.t.} \quad \mathbf{x} \in \Gamma \quad (3)$$

Algorithm 1 ADMM アルゴリズム

Input: $\mathbf{z}^{[1]}, \mathbf{u}^{[1]}, \rho$
Output: $\mathbf{x}^{[I+1]}$

- 1: **for** $i = 1, 2, \dots, I$ **do**
- 2: $\mathbf{x}^{[i+1]} = \text{proj}_{\Gamma}(\mathbf{G}^*(\mathbf{z}^{[i]} - \mathbf{u}^{[i]}))$
- 3: $\mathbf{z}^{[i+1]} = \text{prox}_{(1/\rho)\mathcal{F}}(\mathbf{G}\mathbf{x}^{[i+1]} + \mathbf{u}^{[i]})$
- 4: $\mathbf{u}^{[i+1]} = \mathbf{u}^{[i]} + \mathbf{G}\mathbf{x}^{[i+1]} - \mathbf{z}^{[i+1]}$
- 5: **end for**

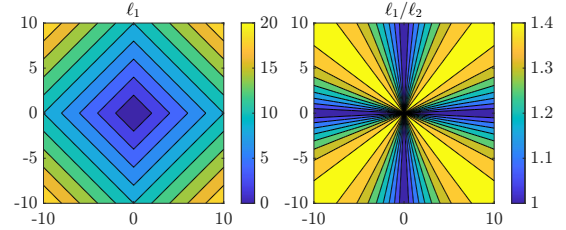


図-1 二次元ベクトル空間における ℓ_1 ノルム (左) と ℓ_1/ℓ_2 分数ペナルティ (右) の比較. $(x, y) \in \mathbb{R}^2$ に対して順に $|x| + |y|$ と $(|x| + |y|)/\sqrt{x^2 + y^2}$ で計算される。

と定式化できる [1, 2]。ただし、 \mathcal{F} はスパース性を誘導するペナルティ関数であり、 \mathbf{G} は離散 Gabor 変換 (DGT) である。項 $\mathcal{F}(\mathbf{G}\mathbf{x})$ は時間周波数領域において信号のスパース性を誘導する正則化項である。代表的な手法として、関数 \mathcal{F} に ℓ_1 ノルムを用いるものが挙げられる [1]。式 (3) の最適化問題に ADMM を適用することで Alg. 1 が得られる。ただし、 I は反復回数、 $1/\rho > 0$ はステップサイズ、 \mathbf{G}^* は \mathbf{G} の複素共役転置、 proj_{Γ} は集合 Γ への射影、 prox は近接作用素である。関数 \mathcal{F} として ℓ_1 ノルムを用いる場合、 $\text{prox}_{(1/\rho)\mathcal{F}}$ は閾値 $(1/\rho)$ のソフト閾値作用素となる。

3 提案手法： ℓ_1/ℓ_2 分数ペナルティによる declipping

図-1 (左) に示す通り、従来手法で用いられる ℓ_1 ノルムは信号のスケールに比例し、原点で最小値をとる。しかし、audio declipping において復元すべき箇所は、clipping が生じた箇所、すなわち閾値より振幅が大きい箇所である。従って、先ほど述べた ℓ_1 ノルムの性質は declipping の問題設定と部分的に矛盾しており、これが復元性能の低下をもたらすと知られている [1]。そこで本稿では、図-1 (右) に示す ℓ_1/ℓ_2 分数ペナルティの利用を考える。 ℓ_1/ℓ_2 分数ペナルティは軸上で最小値をとるコスト関数であり、コスト関数値が信号の定数倍に依存しない。そのため、 ℓ_1/ℓ_2 分数ペナルティを declipping に利用することで復元信

* Audio declipping using sparse regularization with ℓ_1/ℓ_2 fractional penalty. By Kiyoka TSUNODA (Tokyo University of Agriculture and Technology), Kazuki MATSUMOTO (Waseda University), Koki YAMADA, and Kohei YATABE (Tokyo University of Agriculture and Technology)

号の振幅を過度に小さくすることなく高調波成分を抑圧することが期待できる。本稿では、関数 \mathcal{F} を時間フレームごとの ℓ_1/ℓ_2 分数ペナルティの和

$$\mathcal{F}(\mathbf{y}) = \sum_{n=1}^N \frac{\|\mathbf{y}[n]\|_1}{\|\mathbf{y}[n]\|_2} \quad (4)$$

と定義する。ただし、 $\mathbf{y}[n] = [y[1, n], \dots, y[M, n]]^T \in \mathbb{C}^M$ は時間周波数領域のスペクトログラムにおける n ($1 \leq n \leq N$) 番目の時間フレームを表すベクトルであり、 M は周波数点数である。提案手法では、**Alg. 1** の3行目において、時間フレームごとに ℓ_1/ℓ_2 分数ペナルティの近接作用素 [3]

$$\text{prox}_{\lambda(\ell_1/\ell_2)}(\mathbf{z}) = \arg \min_{\mathbf{v} \in \mathbb{C}^M} \left(\frac{\|\mathbf{v}\|_1}{\|\mathbf{v}\|_2} + \frac{1}{2\lambda} \|\mathbf{v} - \mathbf{z}\|_2^2 \right) \quad (5)$$

を計算する。ただし、 $\lambda > 0$ は近接作用素のパラメータであり、**Alg. 1** の3行目における $1/\rho$ に対応する。

4 実験

提案手法である ℓ_1/ℓ_2 分数ペナルティを従来の ℓ_1 ノルム最小化に基づく手法と比較した。真の信号には7秒間の音楽信号を10種類用いた。サンプリング周波数は44.1 kHzであった。DGTの窓長は8192サンプルとし、シフト幅は4096サンプルとした。窓関数にはHann窓を用いた。評価指標には ΔSDR を用いた。SDRは復元信号 \mathbf{x} 、真の信号 $\mathbf{x}^{(\text{GT})}$ を用いて

$$\text{SDR} = 10 \log_{10} \frac{\sum_{t \in (L \cup H)} (x^{(\text{GT})}[t])^2}{\sum_{t \in (L \cup H)} (x^{(\text{GT})}[t] - x[t])^2} \quad (6)$$

と定義され、 ΔSDR は観測信号と復元信号のSDRの差で計算した。Clippingの閾値 τ は観測信号 $\mathbf{x}^{(\text{obs})}$ のSDR (Input SDR) が1, 3, 5, 7, 10, 15, 20 dBになるように調整した。反復回数は15000回、パラメータ ρ は1/5, 1/10, 1/15, 1/20とした。

実験結果を図-2に示す。図-2 (a) に着目すると、従来手法である ℓ_1 ノルムを用いた手法は、ステップサイズによらず一定の ΔSDR に収束した。これに対し、提案手法である ℓ_1/ℓ_2 分数ペナルティを用いた手法は反復の過程で従来手法よりも高い復元性能を達成した。実際、図-2 (c) に示すように、15000反復の段階では全てのInput SDRで ℓ_1 ノルムに基づく手法を5 dB程度上回った。このことから、振幅不定なペナルティがaudio declippingの性能向上に寄与する可能性が示唆された。

一方で、図-2 (a) の右端から分かるように、提案手法では反復を過度に繰り返すと復元性能が低下することが確認された。また、復元信号の波形を確認したところ、従来手法では大半の時刻で真の信号よりも振幅が小さくなっている一方、提案手法では一部の時

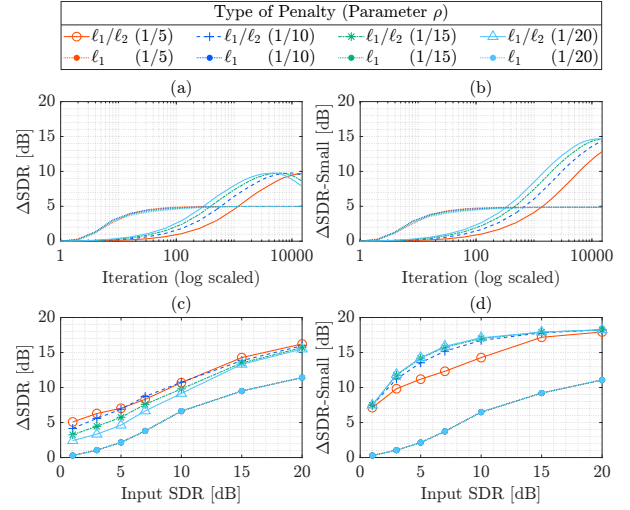


図-2 (a) 反復に伴う ΔSDR の平均値の推移、(b) 反復に伴う $\Delta\text{SDR-Small}$ の平均値の推移、(c) Input SDR ごとの15000反復後の ΔSDR の平均値、(d) Input SDR ごとの15000反復後の $\Delta\text{SDR-Small}$ の平均値。反復回数は対数スケールで表示している。

刻で振幅が過度に大きくなっていた。このことを定量的に確認するために、復元信号 \mathbf{x} が真の信号 $\mathbf{x}^{(\text{GT})}$ を上回る部分を除いたときの復元性能を

$$\text{SDR-Small} = 10 \log_{10} \frac{\sum_{t \in S} (x^{(\text{GT})}[t])^2}{\sum_{t \in S} (x^{(\text{GT})}[t] - x[t])^2} \quad (7)$$

と定め、その傾向を比較した。ただし、 $S = \{t \in (L \cup H) \mid |x[t]| \leq |x^{(\text{GT})}[t]|\}$ は復元信号の振幅が真の信号よりも小さい時刻を表すインデックスの集合である。図-2 (b) から、振幅が増長される部分を除けば提案手法による復元性能も収束することが示された。また、図-2 (d) に示すように、収束後の復元性能は従来手法よりも高い値に収束することが分かった。このことから、提案手法には振幅が過度に増長してしまうという問題点があるものの、これを改善することで復元性能を向上できる可能性が示唆された。

5 むすび

ℓ_1/ℓ_2 分数ペナルティを用いた audio declipping を提案した。提案手法は最適化の過程で ℓ_1 ノルムに基づく手法よりも高い性能を達成したが、反復を過度に繰り返すことで復元性能が低下するという問題点が確認された。今後はこの問題の改善に取り組む。

参考文献

- [1] P. Závřiška, P. Rajmıc, A. Ozerov and L. Rencker, “A survey and an extensive evaluation of popular audio declipping methods,” *IEEE J. Sel. Top. Signal Process.*, **15**(1), 5–24 (2020).
- [2] 田中僚郎, 矢田部浩平, 及川靖広, “制約集合を逐次改良する再重み付きスパース audio declipping,” 音講論集, pp. 207–208 (2024.3).
- [3] M. Tao, “Minimization of L_1 over L_2 for space signal recovery with convergence guarantee,” *SIAM J. Sci. Comput.*, **44**(2), A770–A797 (2022).