

Lipschitz 連続な振幅操作型 DNN を用いた Plug-and-Play アルゴリズムの収束性保証*

© 松本和樹, 内田蓮, 吉野夏樹, 矢田部浩平 (農工大)

1 はじめに

数理最適化に基づく信号処理アルゴリズムの内部に Deep Neural Network (DNN) で構成した雑音除去器を組み込む Plug-and-Play (PnP) は、数理モデルに基づく解釈性の高い処理とデータ駆動型の処理を両立する。しかし、一般の DNN はアルゴリズムの収束を保証するための数学的な性質 (e.g., Lipschitz 連続性) を満たさないことが多く、そのような DNN を組み込んだ PnP アルゴリズムは不安定な挙動に陥る場合がある。この課題の解決に向けて、我々はこれまで、Lipschitz 連続な音響用 DNN である LipsAM (Lipschitz-continuous Amplitude Modifier) を提案したが [1], Lipschitz 定数の制御が不十分であり、PnP アルゴリズムの収束を理論的に保証するには至っていなかった。本稿では、Lipschitz 定数が 1 未満の DNN を構成し、これを残差成分の推定に利用する **CoReM-LipsAM** (Contractive-Residual Map via LipsAM) を提案する。さらに、CoReM-LipsAM を用いた音響 PnP アルゴリズムの収束を定理として示す。

2 PnP と Lipschitz 連続な音響用 DNN

PnP は近接分離アルゴリズムに雑音除去 DNN $\mathcal{D}$ を組み込む枠組みである。例えば、残響を付加する行列 $\mathbf{H} \in \mathbb{R}^{T \times T}$ とガウス雑音 $\mathbf{n} \in \mathbb{R}^T$ により劣化した観測信号 $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \in \mathbb{R}^T$ から原信号 $\mathbf{x} \in \mathbb{R}^T$ を復元する残響除去には、交互方向乗数法 (ADMM) に基づく以下の PnP アルゴリズム

$$\begin{cases} \mathbf{x}^{[k+1]} = \mathbf{G}^H \mathcal{D}(\mathbf{G}(\boldsymbol{\xi}^{[k]} - \boldsymbol{\eta}^{[k]})) \\ \boldsymbol{\xi}^{[k+1]} = \text{prox}_{\alpha F_{\mathbf{y}}}(\mathbf{x}^{[k+1]} + \boldsymbol{\eta}^{[k]}) \\ \boldsymbol{\eta}^{[k+1]} = \boldsymbol{\eta}^{[k]} + \mathbf{x}^{[k+1]} - \boldsymbol{\xi}^{[k+1]} \end{cases} \quad (\text{A})$$

を適用できる。ただし、 $\mathbf{G} \in \mathbb{C}^{N \times T}$ は短時間フーリエ変換行列、 $\text{prox}_{\alpha F_{\mathbf{y}}} : \mathbb{R}^T \rightarrow \mathbb{R}^T$ はデータ忠実性項 (e.g., $F_{\mathbf{y}}(\mathbf{x}) = (1/2) \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2$) の近接作用素、 $\alpha > 0$ はデータ忠実性項の重み、 $(\cdot)^H$ は複素共役転置、 $k \in \mathbb{N}$ は反復の添え字、 $\boldsymbol{\xi}, \boldsymbol{\eta} \in \mathbb{R}^T$ である。1 行目の DNN $\mathcal{D} : \mathbb{C}^N \rightarrow \mathbb{C}^N$ は複素スペクトログラムを処理する雑音除去器であり、正則化関数の近接作用素 (e.g., ソフト閾値作用素) の代わりとして作用する。

PnP の課題として、(A) 中の DNN $\mathcal{D}$ が不安定な挙動を示すとアルゴリズムが収束しないことが挙げられる。この課題の解決に向けて、PnP アルゴリズムの収束性保証に必要な数学的性質を備えた DNN アーキテクチャに関する理論が進展している [2]。これらの理論において、Lipschitz 連続な DNN を構成し、その Lipschitz 定数を制御することが、PnP アルゴリズムの収束を保証するうえで重要な役割を果たす。ここで、写像 $\mathcal{D}$ が L -Lipschitz 連続であるとは、任意の $\mathbf{x}, \mathbf{y}$ に対して $\|\mathcal{D}(\mathbf{x}) - \mathcal{D}(\mathbf{y})\| \leq L \|\mathbf{x} - \mathbf{y}\|$ が成り立つことをいう。また、そのような L の下限を $\mathcal{D}$ の Lipschitz 定数と呼び、 $\text{Lip}(\mathcal{D})$ と書く。

2.1 LipsAM: Lipschitz 連続な振幅操作型 DNN

我々はこれまで、音響 PnP アルゴリズムの収束性保証を念頭に、Lipschitz 連続な音響用 DNN である LipsAM を提案した [1]。具体的には、複素スペクトログラム $\mathbf{z} \in \mathbb{C}^N$ の振幅のみを操作する DNN $\mathcal{D}_{\mathcal{A}}(\mathbf{z}) = \mathcal{A}(|\mathbf{z}|) \odot \text{sign}(\mathbf{z})$ が Lipschitz 連続となるために、振幅操作部分 $\mathcal{A} : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ が満たすべき条件を導出し、以下の 2 つのアーキテクチャ

- (1) **LipsAM-E** (Estimator, $\mathcal{D}_{\mathcal{E}}^{(\text{Lips})}$) : 所望信号の振幅を直接的に推定するアーキテクチャ
- (2) **LipsAM-M** (Masker, $\mathcal{D}_{(\mathcal{M}, g, a, b)}^{(\text{Lips})}$) : 時間周波数マスキングに基づくアーキテクチャ

を提案した。これらはそれぞれ以下の式で書ける。

$$\mathcal{D}_{\mathcal{E}}^{(\text{Lips})}(\mathbf{z}) = (\min(\mathcal{L}(|\mathbf{z}|), |\mathbf{z}|))_+ \odot \text{sign}(\mathbf{z}) \quad (1)$$

$$\mathcal{D}_{(\mathcal{M}, g, a, b)}^{(\text{Lips})}(\mathbf{z}) = g(\max(\mathcal{L}(|\mathbf{z}|), a|\mathbf{z}| + b1)) \odot \mathbf{z} \quad (2)$$

ただし、 $\mathcal{L} : \mathbb{R}_+^N \rightarrow \mathbb{R}^N$ は L -Lipschitz 連続な DNN、 $a > 0$, $b \in \mathbb{R}$ である。また、 $|\cdot|$ 及び $\text{sign}(\cdot)$ は要素ごとの絶対値及び複素符号関数、 $\odot$ は要素ごとの乗算、 $(\cdot)_+$ は ReLU、 g は sigmoid を表す。

3 提案法 : CoReM-LipsAM [3]

式 (1), (2) に示した LipsAM は Lipschitz 定数が 1 以上になる場合があり、Banach の不動点定理に基づく PnP アルゴリズムの収束性保証 [2] を適用できなかった。本稿ではこの問題を解決するために、LipsAM をスケールリングすることで Lipschitz 定数が 1 未満の写

* A Plug-and-Play algorithm with guaranteed convergence using Lipschitz-continuous amplitude modifiers.
By Kazuki MATSUMOTO, Ren UCHIDA, Natsuki YOSHINO and Kohei YATABE (Tokyo University of Agriculture and Technology).

像を構成し、これを残差成分の推定に用いる **CoReM-LipsAM** を提案する。提案法は任意の Lipschitz 連続な写像 $\mathcal{D}_A^{(\text{Lips})} : \mathbb{C}^N \rightarrow \mathbb{C}^N$, その Lipschitz 定数の上界 $B \geq \text{Lip}(\mathcal{D}_A^{(\text{Lips})})$, 及び定数 $\epsilon \in [0, 1)$ を用いて,

$$\mathcal{D}_{(A, \epsilon)}^{(\text{CoReM-Lips})}(\mathbf{z}) = \mathbf{z} - (\epsilon/B)\mathcal{D}_A^{(\text{Lips})}(\mathbf{z}) \quad (3)$$

と定める。また、上式右辺の $\mathcal{D}_A^{(\text{Lips})}$ に式 (1) 及び式 (2) を代入することで、**CoReM-LipsAM-E**, 及び **CoReM-LipsAM-M** を提案する。さらに、我々は $\mathcal{D}$ として CoReM-LipsAM を用いたアルゴリズム (A) の収束を以下の定理として示した [3]。

定理. $\alpha > 0$, $\epsilon \in [0, 1)$, $\mathcal{D} = \mathcal{D}_{(A, \epsilon)}^{(\text{CoReM-Lips})}$ とし, $F_{\mathbf{y}}$ を μ -強凸関数とする。また, $\mathbf{G}^H \mathbf{G} = \mathbf{I}$ 及び $\epsilon/(1 + \epsilon - 2\epsilon^2) < \alpha\mu$ を仮定する¹。このとき, (A) により生成される点列は唯一の不動点に収束する。

3.1 Lipschitz 定数の上界 B の計算

式 (3) 中の上界 $B \geq \text{Lip}(\mathcal{D}_A^{(\text{Lips})})$ としては、我々がこれまでに求めた LipsAM-E/-M の Lipschitz 定数の上限を用いる [3]。これらは、学習対象部分の Lipschitz 定数の上限 $L \geq \text{Lip}(\mathcal{L})$ の関数として、それぞれ

$$B_{\mathcal{E}}(L) = \sup_{\text{Lip}(\mathcal{L}) \leq L} \left(\text{Lip}(\mathcal{D}_{\mathcal{E}}^{(\text{Lips})}) \right) = \sqrt{L^2 + 1} \quad (4)$$

$$B_{(\mathcal{M}, g)}(L, a, b) = \sup_{\text{Lip}(\mathcal{L}) \leq L} \left(\text{Lip}(\mathcal{D}_{(\mathcal{M}, g, a, b)}^{(\text{Lips})}) \right) \quad (5)$$

と書ける。式 (4) は閉形式で与えられるため、自動微分による DNN の学習に直接組み込める。また、式 (5) は解析的に評価することが困難であるものの、7 変数の最大化問題に帰着できるため、粒子群最適化などの大域的探索手法により数値的に計算できる。なお、 $B_{(\mathcal{M}, g)}$ は自動微分による DNN の学習に組み込むことができないため、学習時は $B_{(\mathcal{M}, g)}$ を近似する微分可能な関数 (e.g., スプライン多項式) を用いる。

3.2 学習対象部分 $\mathcal{L}$ の構造

Lipschitz 定数の上界 $L \geq \text{Lip}(\mathcal{L})$ を明示的に扱うために、学習対象部分 $\mathcal{L}$ は 1-Lipschitz 連続な写像 $\tilde{\mathcal{L}} : \mathbb{R}_+^N \rightarrow \mathbb{R}^N$ を用いて $\mathcal{L}(\mathbf{x}) = \tilde{\mathcal{L}}(\text{sigmoid}(s) \cdot \mathbf{x})$ のように構成する。ただし、 $s \in \mathbb{R}$ は学習可能なパラメータであり、 $\tilde{\mathcal{L}}$ は直交畳み込み層と 1-Lipschitz 連続な活性化関数を直列に接続することで実装する。これにより、学習対象部分 $\mathcal{L}$ の Lipschitz 定数の上界を $L = \text{sigmoid}(s)$ により計算でき、また学習を通じてその値を最適化できるようになる。

4 実験

残響除去タスクにおいて復元性能と収束性を評価した。 $\mathbf{x}$ は LibriTTS-R の **test-clean** から、 $\mathbf{H}$ は

¹ $\mathbf{G}$ に関する仮定は Parseval タイプ窓により保証できる。

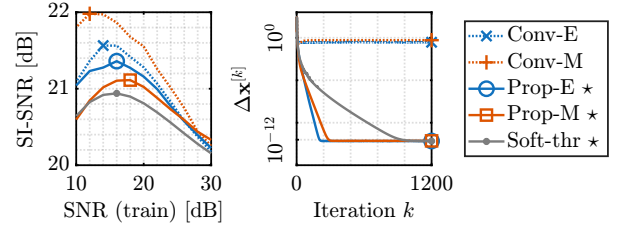


図-1 学習時の SNR に対する復元性能 (左), 及び最適化変数の変化量の推移 (右). Conv は従来法, Prop は提案法, -E は振幅の直接推定, -M はマスキング, * は収束性の理論保証を示す. 右に示す曲線は左において最も高い性能を達成した設定 (マーカー) における結果である。

BUT Reverb Database から、それぞれランダムに 10 サンプルを抽出した。 $\mathbf{n}$ のレベルは $\mathbf{H}\mathbf{x}$ に対して -40 dB とした。定理の条件が成立するように、 $F_{\mathbf{y}}(\mathbf{x}) = (1/2) \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 + (\gamma/2) \|\mathbf{x}\|_2^2$, $\gamma = 0.0005$, $\epsilon = 0.99$, $\alpha = 1.01\epsilon/(\gamma(1 + \epsilon - 2\epsilon^2))$ とした。評価には 1200 反復後の SI-SNR と変数 $\mathbf{x}$ の変化量 $\Delta\mathbf{x}^{[k]} = \|\mathbf{x}^{[k+1]} - \mathbf{x}^{[k]}\|_2$ を用いた。学習には LibriTTS-R の **train-clean-100** に含まれる音声信号を用いた。ガウス雑音除去タスクにおける SNR を 10 dB から 30 dB まで 2 dB 刻みで設定し、各条件について個別に DNN を学習した。学習対象部分 $\mathcal{L}$ は 5 層の直交畳み込み層と Leaky ReLU により構成した。提案法である CoReM-LipsAM (Prop-E/-M) の比較対象として、通常の畳み込み層を用い、写像全体の Lipschitz 連続性を保証しない従来法 (Conv-E/-M), 及びソフト閾値作用素の閾値を学習したもの (Soft-thr) を用いた。

結果を図-1 に示す。図の左に示すように、提案法 (実線) は Lipschitz 連続性に関する制約のない従来法 (点線) と比較して性能が劣るものの、ソフト閾値作用素 (灰色線) を上回る復元性能を得た。また、図の右に示すように、提案法では $\Delta\mathbf{x}$ が 10^{-12} 程度まで減衰しており、定理と整合する結果が得られた。

5 むすび

本稿では、PnP アルゴリズムの収束を保証可能な音響用 DNN として CoReM-LipsAM を提案し、その有効性を残響除去タスクで検証した。今後の課題は、収束性保証と復元性能のトレードオフの解消である。

参考文献

- [1] K. Matsumoto, R. Uchida and K. Yatabe, “LipsAM: Lipschitz-continuous amplitude modifier for audio signal processing and its application to plug-and-play dereverberation,” *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, pp. 816–820 (2026).
- [2] E. Ryu, J. Liu, S. Wang, X. Chen, Z. Wang and W. Yin, “Plug-and-play methods provably converge with properly trained denoisers,” *Int. Conf. Mach. Learn. (ICML)*, vol. 97, pp. 5546–5557 (2019).
- [3] K. Matsumoto, R. Uchida, N. Yoshino and K. Yatabe, “LipsAM: Lipschitz-continuous neural networks for convergent plug-and-play audio signal recovery,” *Preprint*.