

Lipschitz 連続性を保証した時間周波数マスキング型 DNN *

© 松本和樹, 内田蓮, 矢田部浩平 (農工大)

1 はじめに

Deep Neural Network (DNN) の Lipschitz 連続性は、敵対的サンプルに対する頑健性保証や、最適化に基づく信号処理アルゴリズムの内部で DNN を用いる Plug-and-Play (PnP) の収束性解析など、DNN の挙動に対して理論的な保証を与えるための重要な性質である。しかしながら、音響信号処理の分野でよく用いられる、複素スペクトログラムの振幅のみを操作する DNN は一般に Lipschitz 連続でないため、既存の理論の多くを有効に活用できない。

この課題に対処するために、我々は振幅を操作する Lipschitz 連続な DNN の構造として Lipschitz-continuous Amplitude Modifier (LipsAM) を提案した [1]。ただし、これまでに提案してきた LipsAM は線形層と活性化関数からなる Feed-Forward 型の DNN で、ターゲット成分または雑音成分を直接推定するものに限定されていた。そのため、多くの音響信号処理タスクで有効性が確認されている時間周波数マスキングに基づく DNN には対応していなかった。本稿では、時間周波数マスキング型の DNN を Lipschitz 連続にするための特別な機構を提案し、その有効性を PnP 残響除去のタスクで実験的に確認した。

2 振幅操作型 DNN の Lipschitz 連続性

2.1 PnP と DNN の Lipschitz 連続性

DNN による信号処理は高い性能を達成する一方、出力の根拠が不明瞭であり、解釈性が要求される場面には適さない場合がある。このブラックボックス性を緩和するための枠組みとして PnP がある [2]。PnP では、観測系の数理モデルに基づいた反復的な信号処理アルゴリズムの内部モジュールとして DNN を利用する。具体的には、PnP では信号に関する先験情報を反映する近接作用素 (e.g., スパース性を誘導するソフト閾値作用素) を、雑音除去タスクで学習した DNN で置き換えることで、データ駆動型の高性能な信号処理と解釈可能なアルゴリズムを両立する。

ただし、不安定な挙動をする DNN を PnP で利用するとアルゴリズムが収束しない場合がある。PnP アルゴリズムの収束性保証に関する理論研究の多くは、「数学的に良い性質」として DNN の Lipschitz 連続性を仮定する [2]。ただし、 $\mathcal{H}_1, \mathcal{H}_2$ を Hilbert 空

間とすると、写像 $f: \mathcal{H}_1 \rightarrow \mathcal{H}_2$ が L -Lipschitz 連続であるとは、任意の点 $\mathbf{x}, \mathbf{y} \in \mathcal{H}_1$ に対して不等式 $\|f(\mathbf{x}) - f(\mathbf{y})\|_2 \leq L\|\mathbf{x} - \mathbf{y}\|_2$ が成立することをいう。また、この不等式を満たす L の下限を f の Lipschitz 定数と呼び、

$$\text{Lip}(f) := \sup_{\mathbf{x} \neq \mathbf{y} \in \mathcal{H}_1} \frac{\|f(\mathbf{x}) - f(\mathbf{y})\|_2}{\|\mathbf{x} - \mathbf{y}\|_2} \quad (1)$$

と書く。Lipschitz 定数は写像の安定性に関する定量的な指標であり、Lipschitz 定数が小さいほど、入力に対する摂動が出力に与える影響が小さいことを意味する。実ベクトル空間上の微分可能な写像 $f: \mathbb{R}^N \rightarrow \mathbb{R}^M$ の Lipschitz 定数は f の Jacobi 行列

$$\mathbf{J}_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_N}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_M}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_M}{\partial x_N}(\mathbf{x}) \end{pmatrix} \in \mathbb{R}^{M \times N} \quad (2)$$

の作用素ノルムの上限に対応する。すなわち、

$$\text{Lip}(f) = \sup_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{J}_f(\mathbf{x})\|_{\text{op}} < +\infty \quad (3)$$

が成立する。ただし、 $\|\cdot\|_{\text{op}}$ は作用素ノルムであり、行列の場合は最大特異値と一致する。ReLU などの微分不可能な関数では、各点において微分の代わりに Clarke 劣微分を用い、その上限として Lipschitz 定数を与えることができる。例えば、ReLU は原点を除いて微分可能であり、勾配の最大値が 1 である。したがって、ReLU は 1-Lipschitz 連続である。また、複素ベクトル空間上の写像 $f: \mathbb{C}^N \rightarrow \mathbb{C}^M$ は、適当な実ベクトル空間上の写像 $\tilde{f}: \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2M}$ と同一視して議論することができる。

線形層と Lipschitz 連続な活性化関数からなる Feed-Forward 型の DNN は全体としても Lipschitz 連続である。また、線形層の重みに対して制約を課すことでアーキテクチャ全体の Lipschitz 定数を所望の値以下に抑えることも可能である。例えば、ReLU などの 1-Lipschitz 連続な活性化関数を用い、かつ線形層の重みの最大特異値を 1 に制約すれば、アーキテクチャ全体の Lipschitz 定数を 1 以下に抑えられ、非拡大な DNN を実現できる [2]。

2.2 LipsAM: 振幅操作型 DNN [1]

音響信号処理の手法には、時間領域信号を短時間 Fourier 変換 (STFT) することで得られる複素スペクトログラムを処理対象とするものが数多く存在する。

*Deep neural networks based on time-frequency masking with guaranteed Lipschitz-continuity. By Kazuki MATSUMOTO, Ren UCHIDA and Kohei YATABE (Tokyo University of Agriculture and Technology).

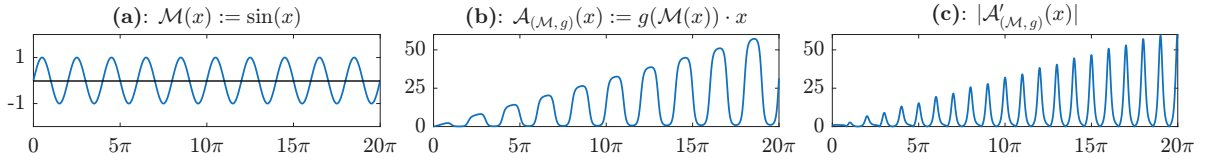


図-1 従来のマスキング型アーキテクチャが Lipschitz 連続にならない一変数関数の例. $g(t) = \text{sigmoid}(4t)$ を用いており, これは 1-Lipschitz 連続である. (a) は式 (8) における写像 $\mathcal{M}$ であり, ここでは $\mathcal{M}(x) = \sin(x)$ とした. このとき $\mathcal{M}$ は 1-Lipschitz 連続である. (b) はマスキング操作 $\mathcal{A}_{(\mathcal{M},g)}$, (c) は (b) の導関数の絶対値である. (c) が非有界であることから (b) は Lipschitz 連続でないことが確認できる.

その典型的なアーキテクチャとして, 振幅のみを操作するものがある. 我々はこれを Amplitude Modifier (AM) と呼び,

$$\mathcal{D}_A : \mathbb{C}^N \rightarrow \mathbb{C}^N : \mathbf{z} \mapsto \mathcal{A}(|\mathbf{z}|) \odot \text{sign}(\mathbf{z}) \quad (4)$$

と書く. ここで, $\mathbf{z} = [z_1, \dots, z_N]^T \in \mathbb{C}^N$ は複素スペクトログラムの全要素をベクトル化したものである. 振幅成分 $|\mathbf{z}| \in \mathbb{R}_+^N$ 及び位相成分 $\text{sign}(\mathbf{z}) \in \mathbb{C}^N$ は, 要素ごとの絶対値関数及び複素符号関数を用いて

$$(|\mathbf{z}|)_n = |z_n|, \quad (\text{sign}(\mathbf{z}))_n = \begin{cases} \frac{z_n}{|z_n|} & (z \neq 0) \\ 0 & (z = 0) \end{cases} \quad (5)$$

と書ける. また, $\mathcal{A} : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ は振幅を操作する作用素であり, $\odot$ は要素ごとの乗算を表す.

一般の AM は, 仮に $\mathcal{A}$ (AM の振幅操作部) を Lipschitz 連続なものに制約したとしても, $\mathcal{D}_A$ (AM 全体) の Lipschitz 連続性を保証できず, PnP の収束性保証などの DNN の Lipschitz 連続性に依拠した理論を適用できない. この問題を解決するために, 我々は $\mathcal{D}_A$ が Lipschitz 連続となるように $\mathcal{A}$ に対して制約を課す LipsAM を提案した [1].

定義 1. 写像 $\mathcal{D}_A(\cdot) = \mathcal{A}(|\cdot|) \odot \text{sign}(\cdot)$ が LipsAM であるとは, その振幅操作部分 $\mathcal{A} : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ がある定数 $L_1, L_2 \geq 0$ のもとで以下を満たすことをいう.

(i) $\mathcal{A}$ は L_1 -Lipschitz 連続

(ii) 任意の $n = 1, \dots, N$ 及び $\mathbf{x} \in \mathbb{R}_+^N$ に対して

$$(\mathcal{A}(\mathbf{x}))_n \leq L_2 x_n \quad (6)$$

次に示すように, LipsAM は Lipschitz 連続であり, その Lipschitz 定数の見積もりが可能である [1].

定理 2. 写像 $\mathcal{D}_A : \mathbb{R}^N \rightarrow \mathbb{R}^N$ が LipsAM ならば $\max(L_1, L_2)$ -Lipschitz 連続である.

この定理に基づき, 我々はこれまでに Lipschitz 連続な振幅操作型 DNN を二種類提案した.

$$\mathcal{D}_S^{(\text{Lips})}(\mathbf{z}) = \text{ReLU}(\min(\mathcal{S}(|\mathbf{z}|), |\mathbf{z}|)) \odot \text{sign}(\mathbf{z})$$

$$\mathcal{D}_R^{(\text{Lips})}(\mathbf{z}) = \text{ReLU}(|\mathbf{z}| - \text{ReLU}(\mathcal{R}(|\mathbf{z}|))) \odot \text{sign}(\mathbf{z})$$

前者を LipsAM-SE (Signal Estimator) と呼び, これは写像 $\mathcal{S} : \mathbb{R}_+ \rightarrow \mathbb{R}$ を用いてターゲット成分の振幅を推定する. 後者を LipsAM-RE (Residual Estimator)

と呼び, これは写像 $\mathcal{R} : \mathbb{R}_+ \rightarrow \mathbb{R}$ を用いて雑音成分の振幅を推定し, それを差し引く. これらのアーキテクチャは min および $\mathcal{R}$ 直後の ReLU の導入により LipsAM の条件 (ii) を満たし, $\mathcal{S}$ 及び $\mathcal{R}$ が Lipschitz 連続ならば全体としても Lipschitz 連続になる.

2.3 マスキング型 DNN は非 Lipschitz 連続

音響信号処理分野の典型的な DNN として, 時間周波数マスキングに基づくものがある [3]. 本稿では, これを AM-SM (Signal Masker) と呼び,

$$\text{AM-SM} : \mathcal{D}_{(\mathcal{M},g)}(\mathbf{z}) = G_g(\mathcal{M}(|\mathbf{z}|)) \odot \mathbf{z} \quad (7)$$

と書く. ただし, 写像 $G_g : \mathbb{R}^N \rightarrow [0, B_g]^N : \mathbf{x} \mapsto (g(x_n))_{n=1}^N$ は要素ごとに有界な非負値関数 $g : \mathbb{R} \rightarrow [0, B_g]$ を適用する写像, $B_g \in \mathbb{R}_+$ は関数 g の上限, $\mathcal{M} : \mathbb{R}_+^N \rightarrow \mathbb{R}^N$ はマスク推定 DNN の学習対象部分である. 関数 g としては sigmoid 関数がよく用いられ, このときマスク値の上限 $B_g = 1$ となる. このようなマスキング型の DNN は「成分を残す部分では 1 に近い値を, 抑圧する部分では 0 に近い値を取るマスクを複素スペクトログラムに乗算する」という直観的な構造を有し, かつ振幅の値を直接推定する方法と比べて学習が容易であるという利点がある.

しかしながら, AM-SM は一般には Lipschitz 連続でない. 実際, AM-SM の振幅操作部分は

$$\mathcal{A}_{(\mathcal{M},g)}(\mathbf{x}) = G_g(\mathcal{M}(\mathbf{x})) \odot \mathbf{x} \quad (8)$$

と書けるが, 仮に学習対象部分 $\mathcal{M}$ 及び関数 g の Lipschitz 連続性を仮定しても, $\mathcal{A}_{(\mathcal{M},g)}$ が Lipschitz 連続とならない場合があり, LipsAM の条件 (i) の成立を保証できない¹. 図-1 に $\mathcal{M}(x) = \sin(x)$, $g(t) = \text{sigmoid}(4t)$ の例を示す. これらの $\mathcal{M}$ 及び g はいずれも 1-Lipschitz 連続である. しかし, 図-1 (b) に示すマスキング型の振幅操作は Lipschitz 連続でない. 実際, 式 (3) に示すように, Lipschitz 連続な関数では微分値が有界となるべきだが, 図-1 (c) に示す関数 $|\mathcal{A}'_{(\mathcal{M},g)}(x)|$ は非有界である.

続いて, 式変形を通じて一般のマスキング型の振幅操作 $\mathcal{A}_{(\mathcal{M},g)}$ が Lipschitz 連続でない要因を確認

¹LipsAM の条件 (ii) は常に満たされる. 実際, $G_g(t) \leq B_g$ より $(\mathcal{A}_{(\mathcal{M},g)}(\mathbf{x}))_n \leq B_g x_n$ が成立する.

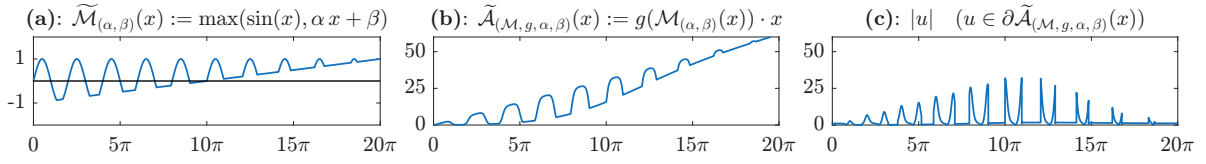


図-2 提案手法による Lipschitz 連続性保証の例. (a) は図-1 における $\mathcal{M}(x) = \sin(x)$ を $\mathcal{M}(x)$ と $\alpha x + \beta$ との最大値に変更したものである. ただし, $\alpha = 1/(10\pi)$, $\beta = -1$ とした. これにより, (c) は有界な関数に, (b) は Lipschitz 連続に制約された. $\max$ 関数は微分不可能なので, (c) には Clarke 劣勾配の絶対値を示した.

する. 簡単のために $\mathcal{M}$ と g が定義域全体で微分可能であることを仮定する. このとき, 振幅操作部分 $\mathcal{A}_{(\mathcal{M}, g)}$ の Jacobi 行列は

$$\mathbf{J}_{\mathcal{A}_{(\mathcal{M}, g)}}(\mathbf{x}) = \text{diag}(G_g(\mathcal{M}(\mathbf{x}))) + \underbrace{\text{diag}(\mathbf{x}) \mathbf{J}_{G_g}(\mathcal{M}(\mathbf{x})) \mathbf{J}_{\mathcal{M}}(\mathbf{x})}_{:= \Phi_{(\mathcal{M}, g)}(\mathbf{x})} \quad (9)$$

と書ける. ただし, $\Phi_{(\mathcal{M}, g)}(\mathbf{x})$ は対角行列であり,

$$\Phi_{(\mathcal{M}, g)}(\mathbf{x}) = \begin{pmatrix} x_1 \cdot g'((\mathcal{M}(\mathbf{x}))_1) & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & x_N \cdot g'((\mathcal{M}(\mathbf{x}))_N) \end{pmatrix}$$

と書ける. $\mathcal{A}_{(\mathcal{M}, g)}$ が Lipschitz 連続となるためには, 式 (9) に示した $\mathbf{J}_{\mathcal{A}_{(\mathcal{M}, g)}}(\mathbf{x})$ の作用素ノルムが有界でなければならない. 第一項は, 関数 g の有界性を仮定すれば $\|\text{diag}(G_g(\mathcal{M}(\mathbf{x})))\|_{\text{op}} \leq B_g$ が成り立つ. 第二項に現れる行列 $\mathbf{J}_{G_g}(\mathcal{M}(\mathbf{x}))$ 及び $\mathbf{J}_{\mathcal{M}}(\mathbf{x})$ に関しても, g 及び $\mathcal{M}$ の Lipschitz 連続性を仮定すれば $\|\mathbf{J}_{G_g}(\mathcal{M}(\mathbf{x}))\|_{\text{op}} \leq \text{Lip}(g)$ 及び $\|\mathbf{J}_{\mathcal{M}}(\mathbf{x})\|_{\text{op}} \leq \text{Lip}(\mathcal{M})$ が成立する. しかしながら, 入力されるベクトル $\mathbf{x}$ 自体は一般に非有界であり, それゆえ第二項全体 $\text{diag}(\mathbf{x}) \mathbf{J}_{G_g}(\mathcal{M}(\mathbf{x})) \mathbf{J}_{\mathcal{M}}(\mathbf{x})$ の有界性は保証されない. 特に, 行列 $\Phi_{(\mathcal{M}, g)}(\mathbf{x})$ の作用素ノルムは

$$\|\Phi_{(\mathcal{M}, g)}(\mathbf{x})\|_{\text{op}} = \max_{n=1, \dots, N} |x_n \cdot g'((\mathcal{M}(\mathbf{x}))_n)| \quad (10)$$

であり, ある次元 n において x_n が発散するようなケースで $|x_n \cdot g'((\mathcal{M}(\mathbf{x}))_n)|$ の有界性は保証されない. これが AM-SM が Lipschitz 連続にならない根本的な要因である.

3 提案手法: Lipschitz 連続性を保証したマスキング型 DNN

本稿では, Lipschitz 連続性を保証するための特別な機構を取り入れたマスキング型 DNN として LipsAM-SM を提案する. 提案手法のアイデアは, 入力の要素 x_n が発散するとき, これを打ち消すように $g'((\mathcal{M}(\mathbf{x}))_n)$ を 0 に収束させることでこれらの積 $x_n \cdot g'((\mathcal{M}(\mathbf{x}))_n)$ を有界に抑える, というものである. まずはこのアイデアを Lipschitz 連続なマスキング型 DNN の十分条件として記述する (証明は割愛する).

定理 3. $\mathcal{M} : \mathbb{R}_+^N \rightarrow \mathbb{R}^N$ を Lipschitz 連続な写像, $g : \mathbb{R} \rightarrow [0, B_g]$ を Lipschitz 連続かつ有界な写像とする. $t \in \mathbb{R}$ における g の Clarke 劣微分の絶対値の上限を $h_g(t) := \sup_{u \in \partial g(t)} |u|$ と定める. $\mathcal{M}$ と g が次の条件を満たすならば $\mathcal{D}_{(\mathcal{M}, g)}(\cdot) = G_g(\mathcal{M}(|\cdot|)) \odot (\cdot)$ は Lipschitz 連続である.

- (i) 任意の $n = 1, \dots, N$ 及び $\lim_{k \rightarrow +\infty} x_n^{(k)} = +\infty$ となる点列 $(\mathbf{x}^{(k)})_{k \in \mathbb{N}} \subset \mathbb{R}_+^N$ について以下が成立.

$$\limsup_{k \rightarrow +\infty} \left| \frac{x_n^{(k)}}{(\mathcal{M}(\mathbf{x}^{(k)}))_n} \right| < +\infty$$

- (ii) $|t \cdot h_g(t)| < +\infty$ が成立.

定理 3 に対する直感的な解釈を述べる. 簡単のため, g は全域で微分可能であると仮定する. 条件 (i) はある次元 n において入力 x_n が無限大に発散するとき, $(\mathcal{M}(\mathbf{x}))_n$ もまた線形関数以上の速さで無限大に発散することを要求する. 続いて, 条件 (ii) は入力の発散 $t \rightarrow +\infty$ に伴って $g'(t)$ が $1/t$ 以上の速さで 0 に収束することを要求する. これらの組み合わせにより, 入力の発散 $x_n \rightarrow +\infty$ はマスク値の微分の収束 $g'((\mathcal{M}(\mathbf{x}))_n) \rightarrow 0$ に相殺され, 結果として $x_n \cdot g'((\mathcal{M}(\mathbf{x}))_n)$ は有界になる. これにより, AM-SM の Lipschitz 連続を保証できない要因であった $\|\Phi_{(\mathcal{M}, g)}(\mathbf{x})\|_{\text{op}}$ の発散の可能性を排除する.

続いて, 定理 3 の条件を満たす具体的なアーキテクチャとして LipsAM-SM を提案する. 提案アーキテクチャは, パラメータ $\alpha \in \mathbb{R}_{++}$ と $\beta \in \mathbb{R}$ を用いて

$$\text{LipsAM-SM} : \mathcal{D}_{(\mathcal{M}, g, \alpha, \beta)}^{(\text{Lips})}(\mathbf{z}) = g(\max(\mathcal{M}(|\mathbf{z}|), \alpha|\mathbf{z}| + \beta\mathbf{1})) \odot \mathbf{z} \quad (11)$$

と定める. ただし $\mathbf{1}$ は全要素が 1 のベクトルである. 写像 $\widetilde{\mathcal{M}}_{(\alpha, \beta)}(\cdot) := \max(\mathcal{M}(|\cdot|), \alpha|\cdot| + \beta\mathbf{1}) \geq \alpha|\cdot| + \beta\mathbf{1}$ は定理 3 の条件 (i) を満たす. また, g として sigmoid 関数を用いれば, 定理 3 の条件 (ii) を満たす. よって, これらを用いたマスキング型のアーキテクチャ $\mathcal{D}_{(\mathcal{M}, g, \alpha, \beta)}^{(\text{Lips})} = \mathcal{D}_{(\widetilde{\mathcal{M}}_{(\alpha, \beta)}, g)}$ は全体として Lipschitz 連続である. 図-2 に示す例から, 図-1 に示した Lipschitz 連続でないアーキテクチャが, 提案手法によって Lipschitz 連続に制約されることが確認できる. なお, 提案アーキテクチャは α を固定しながら $\beta \rightarrow -\infty$ の極限をとることで従来の AM-SM

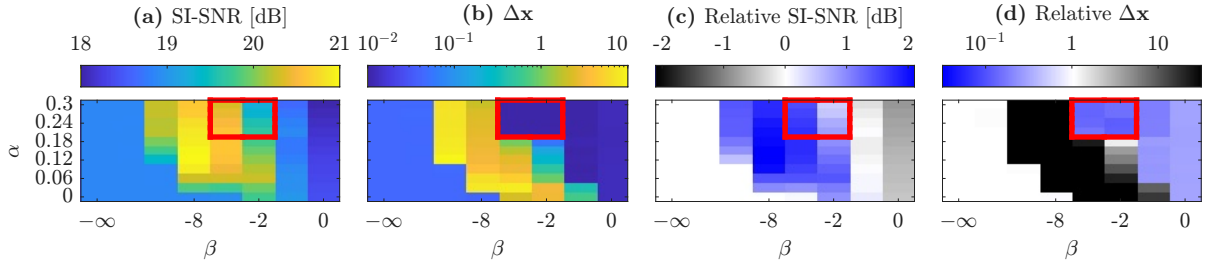


図-3 各パラメータ (α , β) における LipsAM-SM を用いた PnP 残響除去の結果. (a) (b) は SI-SNR と最適化変数の変化量 Δx であり, $\beta = -\infty$ の列は従来の AM-SM に対応する. (c) (d) は従来の AM-SM を基準とした相対量 (SI-SNR は差, Δx は比) であり, 青色は改善, 黒色は劣化を表す. 赤枠は両指標が共に改善した領域を示す.

に有界集合上一様収束する.

4 残響除去における実験的評価

提案アーキテクチャの有効性を確認するために, 雑音除去タスクで学習したマスキング型 DNN を最適化に基づく残響除去アルゴリズムに組み込み, その残響除去性能と収束性を確認した. 本稿では, 観測信号 $\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \in \mathbb{R}^T$ から源信号 $\mathbf{s} \in \mathbb{R}^T$ を推定する残響除去タスクを考える. ただし, $\mathbf{H} \in \mathbb{R}^{T \times T}$ は畳み込み行列, $\mathbf{n} \in \mathbb{R}^T$ は Gauss 性雑音であり, 残響除去問題は

$$\underset{\mathbf{x} \in \mathbb{R}^T}{\text{Minimize}} \left(\lambda \Omega(\mathbf{G}\mathbf{x}) + \frac{1}{2} \|\mathbf{H}\mathbf{x} - \mathbf{y}\|_2^2 \right) \quad (12)$$

として定式化する. $\mathbf{x} \in \mathbb{R}^T$ は復元信号, $\mathbf{G} \in \mathbb{C}^{N \times T}$ は STFT 行列, $\lambda > 0$ は正則化パラメータである. Ω は信号に関する先験情報を反映する正則化関数である. PnP の枠組みではアルゴリズム中の近接作用素 $\text{prox}_{\lambda\Omega}(\mathbf{v}) := \arg \min_{\mathbf{z} \in \mathbb{C}^N} (\Omega(\mathbf{z}) + 1/(2\lambda) \|\mathbf{z} - \mathbf{v}\|_2^2)$ を事前学習済みの雑音除去 DNN $\mathcal{D} : \mathbb{C}^N \rightarrow \mathbb{C}^N$ で置き換える. 交互方向乗数法 (ADMM) に基づくアルゴリズムを用いる場合, 更新式は

$$\begin{cases} \mathbf{x} \leftarrow (\mathbf{H}^T \mathbf{H} + \mathbf{G}^H \mathbf{G})^{-1} (\mathbf{H}^T (\mathbf{u} - \boldsymbol{\xi}_1) + \mathbf{G}^H (\mathbf{v} - \boldsymbol{\xi}_2)) \\ \mathbf{u} \leftarrow \text{prox}_{\frac{1}{2\lambda} \|\cdot - \mathbf{y}\|_2^2} (\mathbf{H}\mathbf{x} + \boldsymbol{\xi}_1) \\ \mathbf{v} \leftarrow \mathcal{D}(\mathbf{G}\mathbf{x} + \boldsymbol{\xi}_2) \\ (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) \leftarrow (\boldsymbol{\xi}_1, \boldsymbol{\xi}_2) + (\mathbf{H}\mathbf{x} - \mathbf{u}, \mathbf{G}\mathbf{x} - \mathbf{v}) \end{cases}$$

となる. ただし, $\mathbf{u}, \boldsymbol{\xi}_1 \in \mathbb{R}^T$, $\mathbf{v}, \boldsymbol{\xi}_2 \in \mathbb{C}^N$ である.

雑音除去 DNN の学習時には, 式 (7) に示す従来の AM-SM $\mathcal{D}_{(\mathcal{M}, g)}$ を学習した. DNN は音声信号に対する Gauss 性雑音除去タスクで学習した. 学習対象部分 $\mathcal{M}$ のアーキテクチャはカーネルサイズ 5 の 1 次元畳み込み層を 3 層積み重ねたモデルとした. 中間層のチャンネル数は 512 とした. 活性化関数としては slope が 0.1 の LeakyReLU (1-Lipschitz 連続) を用いて $\mathcal{M}$ の Lipschitz 連続性を保証した. DNN は SNR が 20 dB から 40 dB の範囲の Gauss 性雑音除去タスクで学習した. 損失関数には時間領域における SNR を用いた. 学習データセットには LibriTTS-R

の train-clean-100 を用いた.

残響除去タスクでの評価では, 学習済みの AM-SM に $\alpha|\cdot| + \beta\mathbf{1}$ との max を取る層を追加することで式 (11) に示す LipsAM-SM を構成し, これを PnP した. LipsAM-SM のパラメータは $0 \leq \alpha \leq 0.3$, $\beta \in \{-\infty, -16, -8, -4, -2, -1, 0\}$ の範囲で変化させた. 各パラメータの組について, 正則化項の重み λ は 10^{-3} から 10^2 の範囲で SI-SNR が最良となるものを用いた. 評価用の源信号 $\mathbf{s}$ には LibriTTS-R の test-clean からランダムに選出したものを用い, 畳み込み行列 $\mathbf{H}$ には BUT Reverb Database からランダムに選出したものを用いた. Gauss 性雑音 $\mathbf{n}$ は SNR が 30 dB となるように重畳した. 反復回数は 2000 回とし, 残響除去性能として SI-SNR を, 収束の指標として最適化変数の変化量 $\Delta \mathbf{x}^{(k)} = \|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|_2$ を使用した. ただし k は反復回数である.

実験の結果を 図-3 に示す. 提案アーキテクチャ LipsAM-SM は従来の AM-SM (i.e., $\beta = -\infty$) と比較して高い SI-SNR を達成し, かつ 2000 反復後における最適化変数の変化量も小さな値に抑えられる場合があることが確認された. 特に赤枠に示す範囲のパラメータ α, β ではこれらの利点を両立した.

5 むすび

本稿では, Lipschitz 連続なマスキング型 DNN として LipsAM-SM を提案した. 実験の結果から提案手法が PnP 残響除去タスクにおける復元性能と収束性を改善できることが示された. 適切なパラメータ α, β を自動的に設定すること, アーキテクチャの Lipschitz 定数を評価することが今後の課題である.

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,” *submitted to IEEE ICASSP* (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,” *ICML*, Vol. 97, pp. 5546–5557 (2019).
- [3] A. Narayanan and D. Wang, “Ideal ratio mask estimation using deep neural networks for robust speech recognition,” *IEEE ICASSP*, pp. 7092–7096 (2013).