1-Lipschitz 畳み込み層を用いた DNN 雑音除去 *

☆ 内田蓮, 山田宏樹, 矢田部浩平 (農工大)

1 はじめに

Lipschitz 連続性とは，写像として極端な変化を生じさせない性質のことであり，その性質をもった層を DNN に用いると，ロバストな DNN を実現できることが知られている．また，DNN を最適化アルゴリズムと組み合わせる場合に，DNN が 1-Lipschitz 連続であれば収束性や安定性を保証できる．特に畳み込み層に関しては今までにいくつか手法が提案されてきた．しかし，音響信号処理において，1-Lipschitz 連続な層を用いた DNN に関する検討は進められていない．本稿では，1-Lipschitz 連続層を持つ DNN を音響信号の雑音除去に応用したときの影響を調査する．

2 1-Lipschitz 連続層

写像 $f : \mathbb{R}^n \to \mathbb{R}^m$ の ℓ_2 ノルムに関する Lipschitz 定数 L は，すべての $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ に対して

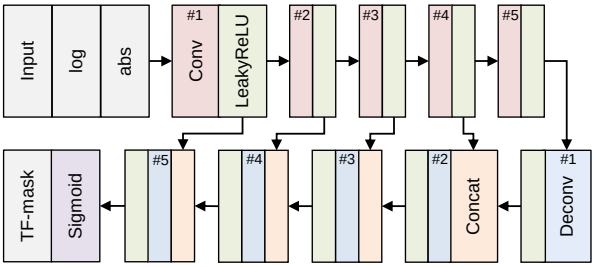
$$\|f(\boldsymbol{x}) - f(\boldsymbol{y})\|_2 \leq L\|\boldsymbol{x} - \boldsymbol{y}\|_2 \qquad (1)$$

を満たす有限の定数である．また，これを満たす f を L-Lipschitz 連続という．DNN が 1-Lipschitz 連続であれば，入力に対する小さな変化が出力に大きな変化をもたらすことはないため，敵対的サンプルなどに対するロバスト性を高めることができる．

2.1 1-Lipschitz 連続畳み込み層

DNN を 1-Lipschitz 連続にするための構成要素として，いくつかの 1-Lipschitz 連続な畳み込み層が提案されている．畳み込みは Lipschitz 連続な写像であるため，Lipschitz 定数の正規化で 1-Lipschitz 連続にすることができる．しかし，畳み込みの Lipschitz 定数を求めるには時間がかかるため，より簡単な 1-Lipschitz 連続な畳み込みの手法がいくつか提案されている．その手法には近似的な手法もあるが，厳密に畳み込みを 1-Lipschitz 連続にする手法として本稿では 2 つの 1-Lipschitz 畳み込み層を調査することにした．

一つ目の手法である Almost Orthogonal Lipschitz layers (AOL) [1] はカーネルのリスケーリングにより，スペクトルノルムを 1 以下に制約し，1-Lipschitz 連続とする．二つ目の手法である Cayley convolution [2] は循環畳み込みが Fourier 領域での積に一致することを利用し，Cayley 変換を用いてカーネルを直交化することで畳み込みを 1-Lipschitz 連続とする．



図–1 UNet の構成．LeakyReLU のスケール係数は 0.01 とし，Concat 層は 1-Lipschitz 連続にするために出力がスケーリングされる．畳み込み層と逆畳み込み層には従来の畳み込み，AOL，Cayley convolution を用いた．

表–1 各畳み込み層と各逆畳み込み層のパラメータ．

#Conv	#channel	kernel size	padding size	stride
1	(1,64)	(7,5)	(3,2,2,1)	(2,2)
2	(64,128)	(7,5)	(3,2,2,1)	(2,2)
3, 4, 5	(128,128)	(5,3)	(2,1,1,0)	(2,2)
#Deconv				
1	(128,128)	(5,3)	(2,1,1,0)	(2,2)
2, 3	(256,128)	(5,3)	(2,1,1,0)	(2,2)
4	(256,64)	(7,5)	(3,2,2,1)	(2,2)
5	(128,1)	(7,5)	(3,2,2,1)	(2,2)

3 1-Lipschitz 連続層を用いた雑音除去

1-Lipschitz 連続な DNN は画像処理の分野で検討が進んでいるが，音響信号処理の分野では十分に検討されていない．また，最適化アルゴリズムと組み合わせて使う場合，1-Lipschitz 連続な DNN は雑音除去タスクを解くことになるため，1-Lipschitz 連続層を持つ DNN が雑音除去に与える影響を調査する．

UNet で時間周波数マスクを推定し，雑音除去を行う．図–1 に今回用いる UNet の構造を示す．畳み込み層と逆畳み込み層のパラメータはそれぞれ 表–1 に示す通りである．実験では，従来の畳み込み層，AOL，Cayley convolution を用いた 3 種類の UNet (Standard-UNet，AOL-UNet，Cayley-UNet) で雑音除去を行い，推論結果から 1-Lipschitz 畳み込み層の表現力を調査する．

エポック数は 300，バッチサイズは 32 とした．初期学習率は 0.001 とし，50 エポックごとに 0.5 倍した．データセットは 48 kHz から 16 kHz にダウンサンプリングした VoiceBank-DEMAND を用いて，非重複かつ無作為に選び，約 1.5 秒に切り出した．損失関数は時間領域での平均絶対値誤差 (MAE) を用いた．STFT の解析窓と合成窓には窓長 512 のルートハン窓，オーバーラップは 128，時間フレーム数は 64

* DNN-based denoising with 1-Lipschitz convolutional layers. By Ren UCHIDA, Koki YAMADA and Kohei YATABE (Tokyo University of Agriculture and Technology).

表–2　UNet ごとの畳み込み層のスペクトルノルム．反復回数 1000 回のべき乗法でスペクトルノルムを求めた．

	#Conv				
	1	2	3	4	5
Standard	1.7214	15.4779	11.1872	9.3498	11.2728
AOL	0.5714	0.6185	0.6525	0.6523	0.6129
Cayley	0.2678	0.9999	0.9999	1.0000	0.9998

表–3　UNet ごとの逆畳み込み層のスペクトルノルム．

	#Deconv				
	1	2	3	4	5
Standard	12.6228	16.9374	16.3381	22.9048	3.4715
AOL	0.9136	0.6519	0.8143	0.8863	0.8570
Cayley	0.2316	0.5863	0.5073	0.1757	1.0000

とした．雑音除去の評価は SI-SDR [dB] を用いた．

また，SI-SDR を損失に用いた FGSM [3] を用いて敵対的サンプルを作成し，ロバスト性を調査した．敵対的サンプル $\widetilde{\boldsymbol{x}}$ は損失を最大化するような摂動を加えられた入力のことであり，摂動の大きさを ϵ，入出力信号 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ として，

$$\widetilde{\boldsymbol{x}} = \boldsymbol{x} + \epsilon \cdot \mathrm{sign}(\nabla_x \mathrm{SI\text{-}SDR}(\boldsymbol{x}, \boldsymbol{y})) \qquad (2)$$

と計算される．ϵ による SI-SDR の変化を評価するために，ϵ を 0.000 から 0.010 までを 0.001 間隔，0.010 から 1.000 までを 0.010 間隔で取り，FGSM を用いてテストテータに摂動を加えた．

4　実験結果

4.1　Lipschitz 定数の確認

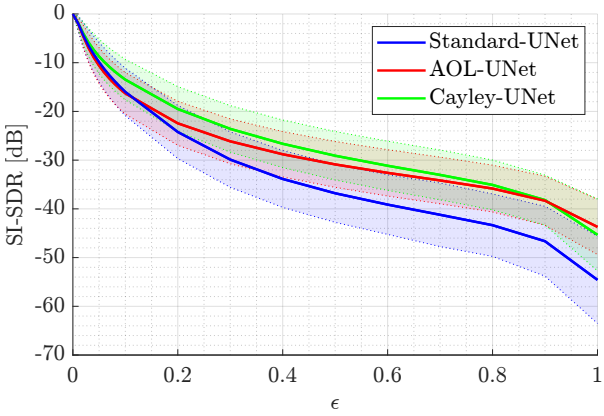
表–2 と 表–3 は学習した UNet の畳み込み層と逆畳み込み層のスペクトルノルムを示す．AOL と Cayley convolution のスペクトルノルムが 1 以下であるため，1-Lipschitz 連続であることが分かる．また，AOL の Lipschitz 定数はすべてほぼ同程度であるが，Cayley convolution は一部の層でスペクトルノルムは 1 に近い値となっていた．

4.2　雑音除去評価

表–4 は学習済みの UNet を用いてテストデータを推論した結果の SI-SDR 改善量の平均値，中央値，標準偏差を示す．表–4 より，畳み込みの 1-Lipschitz 制約によって雑音除去精度の低下が読み取れる．このことから，畳み込みの 1-Lipschitz 制約によって表現力が低下していることが分かる．AOL-UNet よりも Cayley-UNet の方が SI-SDR 改善量が大きいため，Cayley convolution の方が表現力が高いと言える．これらの表現力の低下は，Lipschitz 定数が小さい場合，異なる入力の差に対してそれぞれの出力の差が小さくなり，出力が似ていくことが原因であると考えられる．実際に 表–2 と 表–3 より，Lipschitz 定数に制限のない従来の畳み込みが Lipschitz 定数と SI-SDR 改善量が最も高い．AOL と Cayley convolution では Cayley convolution の方が値が 1 に近い層が多いため，より SI-SDR 改善量が高いと考えられる．

表–4　テストデータを推論した結果の SI-SDR 改善量．

	SI-SDR		
	平均値	中央値	標準偏差
Standard	8.5238	8.0736	4.5237
AOL	5.7945	5.3133	3.9026
Cayley	6.8225	6.4036	4.2901



図–2　敵対的攻撃による推論結果の SI-SDR 変化量．実線が平均値で，破線が標準偏差である．

図–2 はテストデータに対して敵対的攻撃を施した際の推論結果の SI-SDR 変化量を示す．図–2 より，全ての UNet において ϵ が 0 から 0.2 の間で SI-SDR が大きく減少し，その後緩やかに SI-SDR が減少していることが分かる．全体を通して Standard-UNet より，AOL-UNet と Cayley-UNet の方がロバストであった．また，AOL-UNet と Cayley-UNet を比較して，AOL よりも Cayley convolution の方がロバストであると言える．これは Cayley convolution が 1-Lipschitz 連続かつ，より表現力の高いためであると考えられる．

5　むすび

本稿では 1-Lipschitz 畳み込み層が雑音除去に与える影響を調査した．実験結果から，1-Lipschitz 制約を持つ畳み込みが従来の畳み込みよりも表現力が低下することを確認した．また，敵対的攻撃に対しては従来の畳み込みよりもロバストであることも確認した．1-Lipschitz 連続な畳み込み手法を比較して，AOL よりも Cayley convolution の方が高い表現力と敵対的攻撃に対するロバスト性を有することが分かった．

参考文献

[1] B. Prach and C. H. Lampert, "Almost-orthogonal layers for efficient general-purpose Lipschitz networks," *Eur. Conf. Comput. Vis. (ECCV)*, pp. 350–365 (2022)

[2] A. Trockman and J.Z. Kolter, "Orthogonalizing convolutional layers with the Cayley transform," *Int. Conf. Learn. Represent. (ICLR)* (2021).

[3] I. J. Goodfellow, J. Shlens and C. Szegedy, "Explaining and earnessing adversarial examples," *Int. Conf. Learn. Represent. (ICLR)* (2015).