九州大学学術情報リポジトリ
Kyushu University Institutional Repository

# ニューラルネットワークの万能近似定理

西島, 隆人
九州大学大学院数理学府数理学専攻

https://hdl.handle.net/2324/7329795

出版情報：Kyushu University, 2020, 修士, 修士
バージョン：
権利関係：

令和2年度 修士論文

# ニューラルネットワークの万能近似定理

九州大学数理学府数理学専攻
2MA19011T
西島 隆人
指導教員　溝口佳寛 教授

令和3年1月29日

# 目 次

**1 概要** **3**

**2 はじめに** **3**
2.1 記号 . . . . . 3
2.2 万能近似定理の歴史 . . . . . 5

**3 リッジ関数とその万能近似定理への応用** **8**
3.1 リッジ関数の性質 . . . . . 8
3.2 Lin と Pinkus の結果の証明 . . . . . 12

**4 ニューラルネットワークの $C(\mathbb{R}^r)$ における万能近似定理** **15**
4.1 Leshno et al. の結果 . . . . . 15
4.2 Chui と Li の結果 . . . . . 24
4.3 Sun と Cheney の結果 . . . . . 27
4.4 Hornik の結果 . . . . . 32

**5 $C(\mathbb{R}^r)$ における万能近似定理の応用** **38**
5.1 多出力・多層への拡張 . . . . . 38
5.2 $L^p$ 空間における万能近似定理 . . . . . 40
5.3 ニューラルネットワークによる補間 . . . . . 46

**6 RBF ネットワークの万能近似定理** **49**

**7 近似レートの評価** **55**
7.1 Sobolev 空間における近似レート . . . . . 55
7.2 Barron space と Barron-Maurey-Jones 評価 . . . . . 63

**8 まとめと今後の課題** **73**

**9 付録** **74**
9.1 測度の基本性質 . . . . . 74
9.2 単関数近似定理 . . . . . 76
9.3 Lebesgue 積分の基本定理 . . . . . 77
9.4 Hahn-Banach の拡張定理 . . . . . 80
9.5 Riesz-Markov-角谷の表現定理 . . . . . 82
9.6 Stone-Weierstrass の定理 . . . . . 83
9.7 畳み込み積の性質 . . . . . 85
9.8 $\mathbb{R}^n$ 上の有限 Borel 測度の正則性 . . . . . 87
9.9 $C_0(\mathbb{R}^n)$ の $L^p$ 空間における稠密性 . . . . . 90

9.10 Lusin の定理 . . . . . 94
9.11 Clarkson の不等式 . . . . . 95
9.12 Jackson 評価 . . . . . 104
9.12.1 一般 Jackson 核の性質 . . . . . 104
9.12.2 差分, 滑率の性質 . . . . . 108
9.12.3 $L^p$ かつ $2\pi$ 周期で 1 変数の場合 . . . . . 112
9.12.4 $L^p$ かつ $2\pi$ 周期で多変数の場合 . . . . . 114
9.12.5 Sobolev 空間の元の場合 . . . . . 117
9.13 Baire のカテゴリー定理 . . . . . 122
9.14 $C^\infty$ 級関数が多項式と一致するための必要十分条件 . . . . . 124

# 1 概要

今日, ニューラルネットワークは関数近似器として様々な分野で盛んに利用され, 多くの成功を収めている. では, その成功は単なる偶然であろうか. すなわち, ニューラルネットワークの近似能力に理論的な保証はあるだろうか. その答えが「ニューラルネットワークの万能近似定理」である. この定理は, 適当な設定のもとで, ある関数空間においてニューラルネットワークが稠密であることを述べたものである. 例えば, ある条件下で, $\mathbb{R}^r$ のコンパクト集合上の連続関数を sup ノルムの意味で所望の精度で近似するニューラルネットワークが存在することを主張する. したがって, この定理は, 応用で現れるほとんどすべての関数をニューラルネットワークによって (適当な意味で) 近似できる裏付けを与える. よって, これまでの応用での成功は単なる偶然ではなく, こうしたニューラルネットワークの普遍的な関数近似能力の賜物であると考えてよいだろう. feedforward 型 3 層ニューラルネットワークの万能近似定理は 1989 年に Cybenko, Funahashi, Hornik など複数の研究者により独立に証明された. これらの結果は有界な活性化関数に対するものだったが, 1993 年には Leshno et al. により非有界の場合にも成立することが示された. 本論文は, こういった feedforward 型ニューラルネットワークの万能近似定理や, その近似レートの問題 (中間ユニット数と近似誤差の関係) について総合的にまとめた解説論文である.

# 2 はじめに

## 2.1 記号

本節では本論文全体で共通して用いる記号の定義を述べる.

$\mathbb{N} = \{1, 2, 3, \ldots\}$ を自然数全体の集合とし, $\mathbb{Z}$ を整数全体の集合とし, $\mathbb{Q}$ を有理数全体の集合とする. また $\mathbb{R}$ で実数全体の集合を表す. 特に断らなければ, $\mathbb{R}^n$ は通常の距離によって位相空間とみなす.

**定義 2.1.1.** 位相空間 $X, Y$ に対して,

$$C(X,Y) := \{f : X \to Y \mid f \text{ は連続}\}$$

と定義する. $C(X,\mathbb{R})$ は単に $C(X)$ と表す. また, 関数 $f : X \to \mathbb{R}$ に対して,

$$\mathrm{supp}(f) := \mathrm{cl}(\{x \in X \mid f(x) \neq 0\})$$

と定義し, 開集合 $\Omega \subset \mathbb{R}^n$ と $m = 0, 1, 2, \ldots$ に対して

$$C_0(X) := \{f \in C(X) \mid \mathrm{supp}(f) \text{ はコンパクト}\}$$
$$C^m(\Omega) := \{f \in C(\Omega) \mid f \text{ は}\Omega\text{上で } m \text{ 階偏微分可能で全ての偏導関数が連続}\}$$
$$C^\infty(\Omega) := \bigcap_{m=0}^{\infty} C^m(\Omega)$$

と定義する.

**定義 2.1.2 (生成される $\sigma$-加法族).**
$X$ を集合とし, $\mathcal{A}_0 \subset \mathcal{P}(X)$ とする ($\mathcal{P}(X)$ は $X$ の部分集合全体の集合). このとき, $\mathcal{A}_0$ で生成される $X$ 上の $\sigma$-加法族 $\sigma[\mathcal{A}_0]$ を

$$\sigma[\mathcal{A}_0] := \bigcap \{\mathcal{A} \mid \mathcal{A} \supset \mathcal{A}_0, \mathcal{A} \text{ は } X \text{ 上の}\sigma\text{-加法族}\}$$

と定義する.

**定義 2.1.3 (Borel $\sigma$-加法族, Borel 測度).**
$(X, \mathcal{O})$ を位相空間とする. このとき, $X$ 上の Borel $\sigma$-加法族 $\mathcal{B}_X$ を $\mathcal{B}_X = \sigma[\mathcal{O}]$ と定義する. そして, $X$ 上の Borel 測度とは, 可測空間 $(X, \mathcal{B}_X)$ 上の測度のことをいう. $\mathbb{R}^n$ は $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ として可測空間とみなす.

**定義 2.1.4 (可測写像, $L^p$ 空間).**
$(X, \mathcal{M}), (Y, \mathcal{N})$ を可測空間とする. このとき, $X$ から $Y$ への可測写像全体の集合を $\mathbb{M}(X, Y)$ で表す:

$$\mathbb{M}(X,Y) := \{f : X \to Y \mid \forall E \in \mathcal{N}, f^{-1}(E) \in \mathcal{M}\}$$

また, $(X, \mathcal{M})$ 上の測度 $\mu$ と $1 \leq p < \infty$ に対して,

$$L^p(X, \mathcal{M}, \mu) := \{f \in \mathbb{M}(X, \mathbb{R}) \mid \int_X |f|^p d\mu < \infty\}$$

と定義し, $f \in L^p(X, \mathcal{M}, \mu)$ に対して,

$$\|f\|_{L^p} = \left(\int_X |f|^p d\mu\right)^{1/p}$$

と定義する．文脈から推察できる場合には $L^p(X, \mathcal{M}, \mu)$ を単に $L^p(X)$ と表す．さらに,

$$L^\infty(X, \mathcal{M}, \mu) := \{f \in \mathbb{M}(X, \mathbb{R}) \mid \mu\text{-a.e. で有界}\}$$

とおき, $f \in L^\infty(X, \mathcal{M}, \mu)$ に対して,

$$\|f\|_{L^\infty} = \inf\{\lambda \geq 0 \mid \mu(|f| \geq \lambda) = 0\}$$

と定義する．文脈から推察できる場合には $L^\infty(X, \mathcal{M}, \mu)$ を単に $L^\infty(X)$ と表す.

なお, $L^p(X, \mathcal{M}, \mu)$ に属する関数は a.e. で一致するとき等しいとみなす．つまり, $f, g \in L^p(X, \mathcal{M}, \mu)$ に対して, $f = g$ とは $\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0$ となることをいう.

**定義 2.1.5 (アフィン関数).** $r \in \mathbb{N}$ に対して,

$$\mathbb{A}^r := \{f : \mathbb{R}^r \to \mathbb{R} \mid f \text{ はアフィン関数}\}$$

とおく．ただし, $f : \mathbb{R}^r \to \mathbb{R}$ がアフィン関数であるとは, ある $w \in \mathbb{R}^r$ と $b \in \mathbb{R}$ が存在して,

$$f(x) = w^{\mathrm{T}} x + b \quad (x \in \mathbb{R}^r)$$

となることをいう.

## 2.2 万能近似定理の歴史

本節では本論文で主に扱う feedforward 型 3 層ニューラルネットワークの定義を与え, 詳しく歴史を振り返る．証明は章を変えて行う.

feedforward 型 3 層ニューラルネットワークを次により定義する.

**定義 2.2.1.** $r \in \mathbb{N}, W \subset \mathbb{R}^r, \Theta \subset \mathbb{R}$ と関数 $\Psi : \mathbb{R} \to \mathbb{R}$ に対して,

$$\mathcal{N}_r(\Psi, W, \Theta) := \mathrm{span}\{\mathbb{R}^r \ni x \mapsto \Psi(w^{\mathrm{T}} x + \theta) \in \mathbb{R} \mid w \in W, \theta \in \Theta\}$$
$$\Sigma^r(\Psi) := \mathcal{N}_r(\Psi, \mathbb{R}^r, \mathbb{R}) = \left\{ \sum_j^q \beta_j \Psi \circ A_j \mid \beta_j \in \mathbb{R}, A_j \in \mathbb{A}^r, q \in \mathbb{N} \right\}$$

と定義する．これらを $\Psi$ を活性化関数とする feedforward 型 3 層ニューラルネットワークと呼ぶ.

この定義において中間ユニット数 $q$ に制限を設けないことに注意する．この事情からこのようなニューラルネットワークは shallow wide network と呼ばれることがある (なお, 逆に $q$ に制限を設け, 層の数に制限を設けないニューラルネットワーク, deep narrow network についても近年考察が進められている [16]).

明らかに, $\Psi \in \mathbb{M}(\mathbb{R}, \mathbb{R})$, つまり $\Psi$ が $\mathbb{R}$ から $\mathbb{R}$ への Borel 可測関数ならば $\Psi$ を活性化関数とする feedforward 型 3 層ニューラルネットワークは $\mathbb{M}(\mathbb{R}^r, \mathbb{R})$ に含まれる. また, $\Psi \in C(\mathbb{R}, \mathbb{R})$ ならば $\Psi$ を活性化関数とする feedforward 型 3 層ニューラルネットワークは $C(\mathbb{R}^r, \mathbb{R})$ に含まれる.

次に, feedforward 型 3 層ニューラルネットワークの万能近似定理の意味を正確に述べるために, 距離空間における稠密性の定義をする.

**定義 2.2.2.** $(X, \rho)$ を距離空間とし, $S, T \subset X$ とする. このとき, $S$ が $T$ において $\rho$-稠密であるとは, 任意の $t \in T$ と $\varepsilon > 0$ に対して, ある $s \in S$ が存在して, $\rho(s, t) < \varepsilon$ が成り立つことをいう.

なお, しばしばこの概念を濫用して, 必ずしも $X$ に含まれない集合 $S, T$ について $\rho$-稠密という表現をすることがある (その場合は $S, T$ の元について $\rho$ の値が定まることを前提とする).

**定義 2.2.3.** $K \subset \mathbb{R}^r$ を空でないコンパクト集合とする. $f, g \in C(K)$ に対して,

$$\rho_K(f, g) := \sup_{x \in K} |f(x) - g(x)|$$

と定義する. また, $S \subset C(\mathbb{R}^r)$ に対して,

$$S_K := \{f|_K \mid f \in S\}$$

と定義する. ここで $f|_K$ は $f$ の $K$ への制限である. $(S_K, \rho_K)$ は距離空間となる.

**定義 2.2.4.** $C(\mathbb{R}^r)$ の元の列 $(f_n)_{n=1}^{\infty}$ が $f \in C(\mathbb{R}^r)$ に広義一様収束するとは, 任意の空でないコンパクト集合 $K \subset \mathbb{R}^r$ に対して, $\rho_K(f_n, f) \to 0$ $(n \to \infty)$ が成り立つことをいう. また, $S \subset C(\mathbb{R}^r)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるとは, 任意の空でないコンパクト集合 $K \subset \mathbb{R}^r$ に対して, $S_K$ が $C(K)$ において $\rho_K$-稠密であることをいう.

1960 年代以降のニューラルネットワークの応用における数多くの成功から, どのような条件下で $\mathcal{N}_r(\Psi, W, \Theta)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるか？という問いが立てられた. Hornik[12] によれば, この問いを研究者が研究し始めたのは 1980 年代になってからである. 歴史的にはまず $\Sigma^r(\Psi)$ に関して考察が進められた. つまり, 活性化関数 $\Psi$ がどのようなものならば万能近似能力を持つかという問題が考えられた. この問題は 1989 年に複数の研究者により独立に異なる手法で解かれた. Cybenko[8] は $\Psi$ が連続な sigmoidal 関数であるときに万能近似能力を持つことを Hahn-Banach の拡張定理や Riesz-Markov-角谷の表現定理などを用いて示した. ここで, sigmoidal 関数の定義は次の通りである.

**定義 2.2.5 (sigmoidal 関数).**
関数 $\sigma : \mathbb{R} \to \mathbb{R}$ が次の条件をみたすとき, sigmoidal 関数と呼ぶ:

$$\sigma(t) \to \begin{cases} 1 & (t \to +\infty) \\ 0 & (t \to -\infty) \end{cases}$$

一方, Hornik[12] は $\Psi$ が squashing 関数 (単調非減少な sigmoidal 関数) の場合を Stone-Wierstrass の定理を使って証明した．また, 船橋 [7] は $\Psi$ が非定数かつ有界かつ単調非減少な連続関数の場合を入江・三宅 [14] が示した事実を用いて証明した．そして, 1993 年には Leshno et al. [20] によって以下の定理が証明された.

**定義.** $\Omega \subset \mathbb{R}^n$ を開集合とする．また, $\mu$ を $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ 上の Lebesgue 測度とする．このとき,

$$L^\infty_{\mathrm{loc}}(\Omega) := \{f \mid \text{任意のコンパクト集合 } K \subset \Omega \text{に対して } f|_K \in L^\infty(K)\}$$
$$\mathcal{M}(\Omega) := \{f \in L^\infty_{\mathrm{loc}}(\Omega) \mid \mu(\mathrm{cl}\{f \text{ の不連続点 }\}) = 0\}$$

と定義する.

**定理 (Leshno et al. 1993 [20]).** $\mu$ を $\mathbb{R}$ 上の Lebesgue 測度とする．$\Psi \in \mathcal{M}(\mathbb{R})$ に対して, $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるための必要十分条件は, いかなる一変数多項式関数 $P : \mathbb{R} \to \mathbb{R}$ についても $\Psi(x) = P(x)$ ($\mu-$a.e.$x$) とならないことである.

さらに, この定理が発表されたのち, Hornik[13] により連続性に関する仮定を緩和した次の結果が得られた.

**定理 (Hornik 1993 [13]).** $\Psi : \mathbb{R} \to \mathbb{R}$ は各閉区間上で有界かつ Riemann 積分可能であるとすると, $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることと, $\Psi$ が多項式関数と a.e. で一致しないことは同値である.

閉区間上で有界な Riemann 可積分関数は Lebesgue 可測かつその不連続点全体の集合の測度が 0 になることに注意する (猪狩 [44] 定理 3.24 を参照されたい). Leshno et al. の結果では「不連続点全体の集合の閉包」の測度が 0 になることが仮定されていたので, 確かに連続性に関する仮定が緩和されている.

Leshno et al. や Hornik により示されたこれらの定理の適用範囲は広く, 万能近似定理をめぐる一連の議論には一応の決着が付けられたと考えてよいだろう．ただし, 以上で述べた結果のうち, 1989 年の Cybenko, Hornik および 1993 年の Leshno et al., Hornik の結果は適用できる範囲に共通部分はあれど包含関係は無い.

万能近似能力が明らかになってからは近似レートの問題 (中間ユニットの数と近似精度の関係) に焦点が移り, Mhaskar による Sobolev 空間における近似レートの研究 [10] や Barron[1] に始まる Barron space と呼ばれる「ニューラルネットワークでよく近似できる空間」の研究など複数のアプローチでの結果が出ており, 現在でも研究が続いている (Weinan et al.[6] や Siegel, Xu[31] を参照されたい).

本論文ではまず第 3 章で万能近似定理と関わりが深いリッジ関数の性質についていくつか述べる．そして第 4 章のはじめに 1993 年の Leshno et al., Hornik の結果を証明する．次いで, Cybenko の結果を一般化した Chui と Li の結果 [2](下表参照) を解説する．さらに Chui と Li の結果をノルム空間へ一般化した, Sun と Cheney による結果 [33] を紹介し, その後, 1989 年の Hornik の結果の証明を与える.

次に章を変え, 5 章では $C(\mathbb{R}^r)$ における万能近似定理の応用として以上の結果を多層・多出力にした場合や $L^p$ 空間における万能近似定理, 関数の補間について述べる. さらに, 6 章では feedforward 型ニューラルネットワークから一旦離れ RBF(Radial-Basis-Function) ネットワークの万能近似定理を証明する. その後, 7 章では話題を近似レートの問題に移し, feedforward 型 3 層ニューラルネットワークの近似誤差解析について解説する.

表 1: $\mathcal{N}_r(\Psi, W, B)$ が $C(\mathbb{R}^r)$ において万能近似能力を持つ条件

| 西暦 | 証明者 | $\Psi$ | $\Psi$ の連続性 | $W$ | $B$ | 本論文 定理番号 |
|---|---|---|---|---|---|---|
| 1989 | Cybenko | sigmoidal | 連続 | $\mathbb{R}^r$ | $\mathbb{R}$ | 4.2.1 |
| | Funahashi | 有界 単調非減少 | 連続 | $\mathbb{R}^r$ | $\mathbb{R}$ | 4.1.11 |
| | Hornik | sigmoidal 単調非減少 | 仮定なし | $\mathbb{R}^r$ | $\mathbb{R}$ | 4.4.6 |
| 1992 | Chui, Li | sigmoidal | 連続 | $\mathbb{Z}^r$ | $\mathbb{Z}$ | 4.2.1 |
| 1993 | Leshno et al. | 非多項式, $L^\infty_{\mathrm{loc}}$ | $\mu(\mathrm{cl}\{$ 不連続点 $\}) = 0$ | $\mathbb{R}^r$ | $\mathbb{R}$ | 4.1.11 |
| | Hornik | $B$ 上非多項式 局所有界 | $\mu(\{$ 不連続点 $\}) = 0$ | $0 \in W^\circ$ | 開区間 | 4.1.12 |

# 3 リッジ関数とその万能近似定理への応用

本章では万能近似定理と関わりが深いリッジ関数の性質について述べ, 万能近似定理を証明するためには 1 次元の場合のみ考えれば十分であることを示す.

## 3.1 リッジ関数の性質

**定義 3.1.1 (有界線形写像, 共役空間).**
$X, Y$ をノルム空間とする. $f : X \to Y$ が有界線形写像であるとは, $f$ が線形写像であって, かつ $M \geq 0$ が存在して, 任意の $x \in X$ に対して,

$$\|f(x)\|_Y \leq M\|x\|_X$$

となることをいう. 有界線形写像はノルムに関して連続である. また, $X$ の共役空間を

$$X^* := \{f : X \to \mathbb{R} \mid f \text{ は有界線形写像 }\}$$

と定義する. そして, $f \in X^*$ に対して,

$$\|f\| = \inf\{M \geq 0 \mid \forall x \in X, |f(x)| \leq M\|x\|_X\}$$

と定義する. $\|\cdot\|$ は $X^*$ 上のノルムであり, 作用素ノルムと呼ばれる.

**定義 3.1.2 (リッジ関数).**
$X$ をノルム空間とする．関数 $f : X \to \mathbb{R}$ がリッジ関数であるとは, $\varphi \in X^*$ と $g : \mathbb{R} \to \mathbb{R}$ が存在して, $f = g \circ \varphi$ となることをいう．特に $X$ が Hilbert 空間であるときは Riesz の表現定理より, $v \in X$ と $g : \mathbb{R} \to \mathbb{R}$ が存在して, $f(x) = g(\langle x, v \rangle)$ となることをいう．

**定義 3.1.3.** ノルム空間 $X$ と $\mathcal{F} \subset X^*$ に対して,

$$C(\mathbb{R}) \circ \mathcal{F} = \{g \circ f \mid g \in C(\mathbb{R}), f \in \mathcal{F}\}$$

と定義する．$C(\mathbb{R}) \circ \mathcal{F} \subset C(X)$ である．

**補題 3.1.4.** $r, s \in \mathbb{N} \cup \{0\}$ と相異なる $\beta_j \in \mathbb{R} \setminus \{0\}$ $(j = 0, \ldots, r+s)$ に対して, ある $c_j \in \mathbb{R}$ $(j = 0, \ldots, r+s)$ が存在して, 任意の $(x_1, x_2) \in \mathbb{R}^2$ に対して,

$$\sum_{j=0}^{r+s} c_j (x_1 + \beta_j x_2)^{r+s} = x_1^r x_2^s$$

となる．

**証明.** もし, そのような $c_j \in \mathbb{R}$ が存在するなら, 任意の $(x_1, x_2) \in \mathbb{R}^2$ に対して,

$$\begin{aligned}\sum_{j=0}^{r+s} c_j (x_1 + \beta_j x_2)^{r+s} &= \sum_{k=0}^{r+s} \binom{r+s}{k} \sum_{j=0}^{r+s} c_j \beta_j^k x_1^{r+s-k} x_2^k \\ &= \sum_{k=0}^{r+s} \delta_{k,s} x_1^{r+s-k} x_2^k\end{aligned}$$

となるので, $c_j$ は連立一次方程式

$$\binom{r+s}{k} \sum_{j=0}^{r+s} c_j \beta_j^k = \delta_{k,s} \quad (k = 0, \ldots, r+s)$$

をみたす．ところが, この連立一次方程式は Vandermonde 行列の正則性から一意的に解ける ($\beta_j$ は相異なり, かつすべて 0 でないことに注意)．そして, この解 $c_j$ は所望の性質を持つ． □

**定理 3.1.5.** $X$ をノルム空間とするとき, $X$ 上のリッジ関数全体の集合の線形包は $C(X)$ において広義一様収束の意味で稠密である．

**証明.** コンパクト集合 $K \subset X$ に対して, 部分空間 $P \subset C(K)$ を次で定める．

$$P = \mathrm{span}\{K \ni x \mapsto (\varphi(x))^n \in \mathbb{R} \mid n \in \mathbb{N} \cup \{0\},\ \varphi \in X^*\}$$

すると, $P$ は通常の和とスカラー倍, および積に関して代数をなす. そのことを示すには, 任意の $\varphi, \theta \in X^*$ と $r, s \in \mathbb{N} \cup \{0\}$ に対して, $x \mapsto \varphi(x)^r \theta(x)^s$ が $P$ に属することを言えばよいが, 補題 3.1.4 より, $\beta_j, c_j \in \mathbb{R}$ $(j = 0, \ldots, r+s)$ が存在して, 任意の $x \in K$ に対して,

$$\varphi(x)^r \theta(x)^s = \sum_{j=0}^{r+s} c_j (\varphi(x) + \beta_j \theta(x))^{r+s}$$

となるので $x \mapsto \varphi(x)^r \theta(x)^s$ は $P$ に属する. そして, $x, y \in K$ が $x \neq y$ をみたすならば, Hahn-Banach の拡張定理より $\varphi(x) \neq \varphi(y)$ なる $\varphi \in X^*$ が取れる. さらに, $n$ として 0 をとることで $1 \in P$ がわかる. よって, Stone-Weierstrass の定理より, $P$ は $C(K)$ において一様収束の意味で稠密である. これより $X$ 上のリッジ関数全体の集合の線形包は $C(X)$ において広義一様収束の意味で稠密である. □

**補題 3.1.6.** $V = \{\mathbb{R}^r \ni x \mapsto \sum_{i=1}^{l} f_i(a_i^{\mathrm{T}} x) \in \mathbb{R} \mid l \in \mathbb{N}, a_i \in \mathbb{R}^r, f_i \in C(\mathbb{R})\}$ とおくと, $V$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

**証明.** 定理 3.1.5 による (連続関数を定数倍しても連続関数であることに注意). □

次の命題により, 連続関数の空間でのニューラルネットワーク $\Sigma^r(\Psi)$ の稠密性を考える上では 1 次元の場合を考えれば十分であることがわかる.

**命題 3.1.7.** 任意の $\Psi : \mathbb{R} \to \mathbb{R}$ に対して, $\Sigma^1(\Psi)$ が $C(\mathbb{R})$ において広義一様収束の意味で稠密ならば, 任意の $r$ について $\Sigma^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

**証明.** $g \in C(\mathbb{R}^r)$ とし, 任意に空でないコンパクト集合 $K \subset \mathbb{R}^r$ をとる. 任意の $\varepsilon > 0$ に対して, すぐ上の補題により $k \in \mathbb{N}$ と $a_i \in \mathbb{R}^r, f_i \in C(\mathbb{R})$ $(i = 1, \ldots, k)$ が存在して,

$$\sup_{x \in K} |g(x) - \sum_{i=1}^{k} f_i(a_i^{\mathrm{T}} x)| < \varepsilon/2$$

となる. $K$ はコンパクトなので, 各 $i$ について $\{a_i^{\mathrm{T}} x \mid x \in K\} \subset [\alpha_i, \beta_i]$ をみたす $\alpha_i, \beta_i \in \mathbb{R}$ が取れる. そこで, $\Sigma^1(\Psi)$ の稠密性から, $A_{i,j} \in \mathbb{A}^1$ と $c_{i,j} \in \mathbb{R}$ が存在して,

$$\sup_{y \in [\alpha_i, \beta_i]} |f_i(y) - \sum_{j=1}^{m_i} c_{i,j} \Psi(A_{i,j}(y))| < \varepsilon/2k$$

となる. よって, 任意の $x \in K$ に対して,

$$\begin{aligned}
&|g(x) - \sum_{i=1}^{k} \sum_{j=1}^{m_i} c_{i,j} \Psi(A_{i,j}(a_i^{\mathrm{T}} x))| \\
&\leq |g(x) - \sum_{i=1}^{k} f(a_i^{\mathrm{T}} x)| + \sum_{i=1}^{k} |f_i(a_i^{\mathrm{T}} x) - \sum_{j=1}^{m_i} c_{i,j} \Psi(A_{i,j}(a_i^{\mathrm{T}} x))| \\
&< \varepsilon/2 + \varepsilon/2 = \varepsilon
\end{aligned}$$

となる．しかも $x \mapsto A_{i,j}(a_i^{\mathrm{T}}x)$ は $\mathbb{A}^r$ に属する． □

さらにリッジ関数について次の定理が成り立つ．

**定理 3.1.8 (Lin and Pinkus [21]).**
$\mathcal{A} \subset \mathbb{R}^r$ に対して，

$$R(\mathcal{A}) = \{\mathbb{R}^r \ni x \mapsto f(w^{\mathrm{T}}x) \in \mathbb{R} \mid f \in C(\mathbb{R}), w \in A\}$$

とおくと，$\mathrm{span}R(\mathcal{A})$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることと，$\mathcal{A}$ 上で恒等的に 0 を取る非自明な $r$ 変数斉次多項式が存在しないことは同値である．

**証明.** 次節参照． □

この定理により次のことがわかる．

**命題 3.1.9.** $\Psi : \mathbb{R} \to \mathbb{R}$ と $W, B \subset \mathbb{R}$ に対して，

$$\mathcal{N}_1(\Psi, W, B) = \mathrm{span}\{x \mapsto \Psi(wx + b) \mid w \in W, b \in B\}$$

は $C(\mathbb{R})$ において広義一様収束の意味で稠密であるとする．このとき，$\mathcal{A} \subset \mathbb{R}^r$ 上で 0 を取る非自明な $r$ 変数斉次多項式が存在しないならば，

$$\mathcal{N}_r(\Psi, W\mathcal{A}, B) = \mathrm{span}\{\Psi(aw^{\mathrm{T}}x + b) \mid a \in \mathcal{A}, w \in W, b \in B\}$$

は $C(\mathbb{R}^r)$ において広義一様稠密の意味で稠密である．

**証明.** 命題 3.1.7 と同様に示せる．詳しくは次の通りである．$g \in C(\mathbb{R}^r)$ とし，任意に空でないコンパクト集合 $K \subset \mathbb{R}^r$ をとる．任意の $\varepsilon > 0$ に対して，すぐ上の定理により $k \in \mathbb{N}$ と $a_i \in A, f_i \in C(\mathbb{R})$ $(i = 1, \ldots, k)$ が存在して，

$$\sup_{x \in K} |g(x) - \sum_{i=1}^{k} f_i(a_i^{\mathrm{T}}x)| < \varepsilon/2$$

となる．$K$ はコンパクトなので，各 $i$ について $\{a_i^{\mathrm{T}}x \mid x \in K\} \subset [\alpha_i, \beta_i]$ をみたす $\alpha_i, \beta_i \in \mathbb{R}$ が取れる．そこで，$\mathcal{N}_1(\Psi, W, B)$ の稠密性から，$m_i \in \mathbb{N}, w_{i,j} \in W, b_{i,j} \in B$ と $c_{i,j} \in \mathbb{R}$ が存在して，

$$\sup_{y \in [\alpha_i, \beta_i]} |f_i(y) - \sum_{j=1}^{m_i} c_{i,j}\Psi(w_{i,j}y + b_{i,j})| < \varepsilon/2k$$

となる．よって，任意の $x \in K$ に対して，

$$\begin{aligned}
&|g(x) - \sum_{i=1}^{k}\sum_{j=1}^{m_i} c_{i,j}\Psi(w_{i,j}(a_i^{\mathrm{T}}x) + b_{i,j})| \\
&\leq |g(x) - \sum_{i=1}^{k} f(a_i^{\mathrm{T}}x)| + \sum_{i=1}^{k} |f_i(a_i^{\mathrm{T}}x) - \sum_{j=1}^{m_i} c_{i,j}\Psi(w_{i,j}(a_i^{\mathrm{T}}x) + b_{i,j})| \\
&< \varepsilon/2 + \varepsilon/2 = \varepsilon
\end{aligned}$$

となる． □

## 3.2 Lin と Pinkus の結果の証明

本節では前節に述べた Lin と Pinkus の結果 (定理 3.1.8) の証明を述べる.

**定義 3.2.1.** $\alpha = (\alpha_1, \ldots, \alpha_n) \in (\mathbb{Z}_{\geq 0})^n$ に対して,

$$|\alpha| := \sum_{j=1}^{n} \alpha_j, \quad \alpha! := \alpha_1! \cdots \alpha_n!$$

と定義する. $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ に対して,

$$x^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

と定義する. 偏微分作用素 $\partial := (\partial_1, \ldots, \partial_n) = (\partial/\partial x_1, \ldots, \partial/\partial x_n)$ に対する $\partial^\alpha$ も同様に定める.

**定義 3.2.2.** $n, k \in \mathbb{N}$ に対して, $n$ 変数の $k$ 次斉次多項式関数の集合を

$$\mathcal{H}_k(\mathbb{R}^n) := \left\{ \mathbb{R}^n \ni x \mapsto \sum_{|\alpha|=k} c_\alpha x^\alpha \in \mathbb{R} \mid c_\alpha \in \mathbb{R} \right\}$$

と定義する. ただし, 和は $|\alpha| = k$ なる $\alpha \in (\mathbb{Z}_{\geq 0})^n$ 全体にわたるものとする. また,

$$\mathcal{P}_k(\mathbb{R}^n) = \bigcup_{s=0}^{k} \mathcal{H}_s(\mathbb{R}^n), \quad \mathcal{P}(\mathbb{R}^n) = \bigcup_{s=0}^{\infty} \mathcal{H}_s(\mathbb{R}^n)$$

と定義する.

**命題 3.2.3.** $\mathcal{H}_k(\mathbb{R}^n)$ は通常の和とスカラー倍に関して $\mathbb{R}$ 上の有限次元線形空間をなす. また, $\mathcal{H}_k(\mathbb{R}^n)$ の元 $p(x) = \sum c_\alpha x^\alpha, q(x) = \sum c'_\alpha x^\alpha$ に対して,

$$\langle p, q \rangle := p(\partial) q = \sum_{|\alpha|=k} \alpha! c_\alpha c'_\alpha$$

と定義すると, これは $\mathcal{H}_k(\mathbb{R}^n)$ 上の内積であり, この内積に関して $\mathcal{H}_k(\mathbb{R}^n)$ は Hilbert 空間をなす.

**証明.** $\mathcal{H}_k(\mathbb{R}^n)$ が $\mathbb{R}$ 上の線形空間であること, $\langle \cdot, \cdot \rangle$ が内積の公理をみたすことは明らかである. 有限次元であることは有限集合 $\{x^\alpha \mid |\alpha| = k\}$ が $\mathcal{H}_k(\mathbb{R}^n)$ を生成することからわかる. 完備性については, ノルム空間の有限次元部分空間が完備であることによる ([38] の §1.5 (a) を参照されたい). □

**定義 3.2.4.** 自然数 $m, n \geq 1$ に対して, $m \times n$ の実行列全体の集合を $\mathrm{Mat}(m, n; \mathbb{R})$ で表す. $A \in \mathrm{Mat}(m, n; \mathbb{R})$ に対して,

$$L(A) := \{(y^{\mathrm{T}} A)^{\mathrm{T}} \mid y \in \mathbb{R}^m\} \subset \mathbb{R}^n$$

と定義する．$\Omega \subset \mathrm{Mat}(m,n;\mathbb{R})$ に対して，

$$L(\Omega) := \bigcup_{A\in\Omega} L(A)$$

と定義する．また，集合 $R(\Omega)$ を

$$R(\Omega) = \{\mathbb{R}^n \ni x \mapsto g(Ax) \in \mathbb{R} \mid A \in \Omega, g \in C(\mathbb{R}^m)\}$$

と定義する．

**定理 3.2.5 (射影定理).**
$X$ を Hilbert 空間とし，$L \subset X$ を閉部分空間とする．このとき，任意の $x \in X$ は

$$x = y + z \quad (y \in L,\ z \in L^{\perp})$$

の形に一意的に分解可能である．ただし，

$$L^{\perp} := \{x \in X \mid \langle x, y\rangle = 0\ (\forall y \in L)\}$$

である．

**証明.** [38] の定理 3.2 を参照されたい． □

**補題 3.2.6.** $k \geq 0$ とし $\Omega \subset \mathrm{Mat}(n,m;\mathbb{R})$ とする．また，$L(\Omega)$ 上で恒等的に 0 を取る $\mathcal{H}_k(\mathbb{R}^n)$ の元は 0 に限るとする．このとき，

$$\mathcal{H}_k(\mathbb{R}^n) = \mathrm{span}\{\mathbb{R}^n \ni x \mapsto (d^{\mathrm{T}}x)^k \in \mathbb{R} \mid d \in L(\Omega)\} =: L$$

となる．

**証明.** $\mathcal{H}_k(\mathbb{R}^n) \subset L$ を示せばよい．$L$ はノルム空間 $\mathcal{H}_k(\mathbb{R}^n)$ の有限次元部分空間なので閉である ([38] の §1.5 (a) 参照)．したがって，任意の $p \in \mathcal{H}_k(\mathbb{R}^n)$ に対して，射影定理より，$q \in L, r \in L^{\perp}$ が存在して $p = q + r$ となる．ところが，$r(x) = \sum c_\alpha x^\alpha$ と書くことにすると，$r \in L^{\perp}$ より，任意の $d \in L(\Omega)$ に対して，

$$\begin{aligned}
0 &= \langle r(x), (d^{\mathrm{T}}x)^k\rangle \\
&= r(\partial)(d^{\mathrm{T}}x)^k \\
&= \sum_{|\alpha|=k} c_\alpha \partial^\alpha (d^{\mathrm{T}}x)^k \\
&= \sum_{|\alpha|=k} c_\alpha k! d^\alpha \\
&= k! r(d)
\end{aligned}$$

となる．ゆえに，$r(d) = 0\ (\forall d \in L(\Omega))$ であるから，仮定より $r = 0$ である．よって，$p = q \in L$ である． □

**定理 3.2.7.** $\Omega \subset \mathrm{Mat}(n, m; \mathbb{R})$ とする．このとき, $\mathrm{span}R(\Omega)$ が $C(\mathbb{R}^n)$ において広義一様収束の意味で稠密であることと, $L(\Omega)$ 上で恒等的に 0 を取る $\mathcal{P}(\mathbb{R}^n)$ の元が 0 以外にないことは同値である.

**証明.** $L(\Omega)$ 上で 0 を取る $\mathcal{P}(\mathbb{R}^n)$ の元は 0 に限るとする．このとき, 任意の $k \geq 0$ に対して $\mathcal{H}_k(\mathbb{R}^n) \subset \mathrm{span}R(\Omega)$ となることを示そう．これが示されれば, すべての多項式が $\mathrm{span}R(\Omega)$ に含まれることがわかるので, Stone-Weierstrass の定理により $\mathrm{span}R(\Omega)$ は $C(\mathbb{R}^n)$ において広義一様収束の意味で稠密である．さて, すぐ上の補題により

$$\mathcal{H}_k(\mathbb{R}^n) = \mathrm{span}\{\mathbb{R}^n \ni x \mapsto (d^{\mathrm{T}}x)^k \in \mathbb{R} \mid d \in L(\Omega)\}$$

となる．また, $d \in L(\Omega)$ ならば $A \in \Omega$ と $y \in \mathbb{R}^m$ が存在して $d = A^{\mathrm{T}}y$ となるので, $g(z) := (z^T y)^k$ と定義すれば $g(Ax) = (A^{\mathrm{T}}x)y)^k = (x^{\mathrm{T}}Ay)^k = (x^{\mathrm{T}}d)^k$ となる．よって, $x \mapsto (d^{\mathrm{T}}x)^k$ は $R(\Omega)$ に属するので,

$$\mathcal{H}_k(\mathbb{R}^n) = \mathrm{span}\{\mathbb{R}^n \ni x \mapsto (d^{\mathrm{T}}x)^k \in \mathbb{R} \mid d \in L(\Omega)\} \subset \mathrm{span}R(\Omega)$$

である．逆の証明については Lin, Pinkus, 1992 [21] を参照されたい． □

最後に 7 章で使う命題を証明する．証明には「斉次多項式の決定問題」に関する次の事実を使う.

**命題 3.2.8.** $n$ 変数 $d$ 次斉次多項式はある $r(n, d) := \binom{n-1+d}{d}$ 個の相異なる 0 でない点で決定される．つまり, ある $\xi^{(1)}, \ldots, \xi^{(r(n,d))} \in \mathbb{R}^n \setminus \{0\}$ が存在して, 任意の $p \in \mathcal{H}_d(\mathbb{R}^n)$ に対して,

$$(\forall i \in \{1, \ldots, r(n, d)\},\ p(\xi^{(i)}) = 0) \Rightarrow p \equiv 0$$

となる．さらに, このような $\xi^{(1)}, \ldots, \xi^{(r(n,d))} \in \mathbb{R}^n \setminus \{0\}$ として, 任意の $k = 1, \ldots, d$ と $p \in \mathcal{H}_k(\mathbb{R}^n)$ に対して,

$$(\forall i \in \{1, \ldots, r(n, d)\},\ p(\xi^{(i)}) = 0) \Rightarrow p \equiv 0$$

となるようなものが取れる.

**証明.** [34] を参照されたい． □

**命題 3.2.9.** $r = \dim\mathcal{H}_k(\mathbb{R}^n) = \binom{n-1+k}{k}$ とおき, $n$ 変数の $k$ 次以下の多項式全体の集合を $\pi_k(\mathbb{R}^n)$ と表すことにする．このとき, $a_1, \ldots, a_r \in \mathbb{R}^n$ が存在して,

$$\pi_k(\mathbb{R}^n) = \left\{ x \mapsto \sum_{i=1}^{r} g_i(a_i^{\mathrm{T}}x) \mid g_i \in \pi_k(\mathbb{R}) \right\}$$

となる.

**証明.** 命題 3.2.8 より $a_1, \ldots, a_r \in \mathbb{R}^n$ が存在して, 各 $s = 0, \ldots, k$ に対して, $A = \{a_1, \ldots, a_r\}$ 上で恒等的に 0 を取る $\mathcal{H}_s(\mathbb{R}^n)$ の元は 0 に限る. したがって, 補題 3.2.6 からわかるように

$$\mathcal{H}_s(\mathbb{R}^n) = \mathrm{span}\{\mathbb{R}^n \ni x \mapsto (d^{\mathrm{T}}x)^s \in \mathbb{R} \mid d \in A\}$$

となる. このことと多項式を斉次多項式の和として表せることより

$$\begin{aligned}\pi_k(\mathbb{R}^n) &= \mathrm{span}\{x \mapsto (d^{\mathrm{T}}x)^s \mid d \in A, s = 0, \ldots, k\} \\ &= \left\{x \mapsto \sum_{i=1}^{r} g_i(a_i^{\mathrm{T}}x) \mid g_i \in \pi_k(\mathbb{R})\right\}\end{aligned}$$

となる. □

# 4 ニューラルネットワークの $C(\mathbb{R}^r)$ における万能近似定理

本章ではまず 1993 年の Leshno et al., Hornik の結果を証明する. 次いで, Cybenko の結果を一般化した Chui と Li の結果を解説する. さらに Chui と Li の結果をノルム空間へ一般化した, Sun と Cheney による結果を紹介し, その後, 1989 年の Hornik の結果の証明を与える.

## 4.1 Leshno et al. の結果

本節では, Leshno et al. の結果 [20] およびその改良版について述べる. 本節では特に断らなければ a.e. は Lebesgue 測度の意味で用いる.

**定義 4.1.1.** $\Omega \subset \mathbb{R}^n$ を開集合とする. また, $\mu$ を $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ 上の Lebesgue 測度とする. このとき,

$$\begin{aligned}L^\infty_{\mathrm{loc}}(\Omega) &:= \left\{f : \Omega \to \mathbb{R} \;\middle|\; \begin{array}{l}\text{任意のコンパクト集合 } K \subset \Omega \text{に対して} \\ f|_K \in L^\infty(K)\end{array}\right\} \\ \mathcal{M}(\Omega) &:= \{f \in L^\infty_{\mathrm{loc}}(\Omega) \mid \mu(\mathrm{cl}\{f \text{ の不連続点 }\}) = 0\}\end{aligned}$$

と定義する.

Leshno et al. [20] で示された主要な結果を再掲しよう.

**定理 4.1.2 (Leshno et al. 1993 [20]).**
$\mu$ を $\mathbb{R}$ 上の Lebesgue 測度とする. $\Psi \in \mathcal{M}(\mathbb{R})$ に対して, $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるための必要十分条件は, いかなる一変数多項式関数 $P : \mathbb{R} \to \mathbb{R}$ についても $\Psi = P$ ($\mu$−a.e.) とならないことである.

命題 3.1.7 よりこの定理を示すには $r = 1$ の場合を示せばよい．まず必要性を示そう．必要性の証明は非常に単純である．

**補題 4.1.3.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ が a.e. で多項式関数であるならば, $\Sigma^1(\Psi)$ は $C(\mathbb{R})$ において広義一様収束の意味で稠密でない．

**証明.** $\deg(\Psi) = n$ とすると, $\Sigma^1(\Psi)$ は $n$ 次以下の 1 変数多項式と a.e. で一致する関数全体の集合に含まれる．したがって, $n+1$ 次の 1 変数多項式は広義一様収束の意味で近似できない．なぜなら, 実数 $a < b$ に対して, $[a, b]$ 上で定義された $n$ 次以下の 1 変数多項式関数全体の集合を $P_n([a, b])$ で表すと, $P_n([a, b])$ は $C([a, b])$ の部分線形空間であり, 有限次元 ($n+1$ 次元) である．したがって, sup ノルムに関する閉包について $\overline{P_n([a, b])} = P_n([a, b])$ となる．そこで, もしある $n+1$ 次の 1 変数多項式関数 $f$ が $n$ 次以下の 1 変数多項式関数で $[a, b]$ 上で一様近似できると仮定すると, $f \in \overline{P_n([a, b])} = P_n([a, b])$ となるが, これは $f$ が $n+1$ 次の 1 変数多項式であることに反する． □

十分性は次の定理に含まれるため本節では代わりに次の定理を証明する．

**定理 4.1.4.** $W \subset \mathbb{R}$ は孤立点を持たず $0 \in W$ であるとする．また, $\Psi : \mathbb{R} \to \mathbb{R}$ は $\mathcal{M}(\mathbb{R})$ に属するとする．さらに, ある開区間 $I$ に対して $\Psi$ が $I$ 上 a.e. に多項式と一致しないとする．このとき, $\mathcal{A} \subset \mathbb{R}^r$ 上で 0 を取る非自明な $r$ 変数斉次多項式が存在しないならば,

$$\mathcal{N}_r(\Psi, W\mathcal{A}, I) = \mathrm{span}\{\mathbb{R}^r \ni x \mapsto \Psi(wa^T x + b) \in \mathbb{R} \mid w \in W, a \in \mathcal{A}, b \in I\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である．

さらにこの定理は命題 3.1.9 と次の定理より従うので次の定理を証明すればよい．

**定理.** $W \subset \mathbb{R}$ は孤立点を持たず $0 \in W$ であるとする．また, $\Psi : \mathbb{R} \to \mathbb{R}$ は $\mathcal{M}(\mathbb{R})$ に属するとする．このとき, 開区間 $I$ に対して $\Psi$ が $I$ 上 a.e. に多項式と一致しないなら, $\mathcal{N}_1(\Psi, W, I)$ は $C(\mathbb{R})$ において広義一様収束の意味で稠密である．

まず $\Psi$ が $C^\infty$ 級である場合に成立することを証明しよう．

**補題 4.1.5.** $W \subset \mathbb{R}$ は孤立点を持たず $0 \in W$ であるとする．また, $\Psi : \mathbb{R} \to \mathbb{R}$ はある開区間 $I$ 上で $C^\infty$ かつ多項式と一致しないとする．このとき, 各整数 $n \geq 0$ について多項式 $x \mapsto x^n$ は $\mathcal{N}_1(\Psi, W, I)$ で広義一様近似される．したがって, 特に, $\mathcal{N}_1(\Psi, W, I)$ は $C(\mathbb{R})$ において広義一様収束の意味で稠密である．

**証明.** 後半については $C(\mathbb{R})$ の元が多項式関数で広義一様近似できること, および前半より $\mathcal{N}_1(\Psi, W, I)$ が全ての 1 変数の $\mathbb{R}$ 係数多項式を広義一様近似できることからわかる．$K \subset \mathbb{R}$ を空でないコンパクト集合とする．また, $b \in I$ に対して, $f_b : \mathbb{R} \times \mathbb{R} \ni (x, w) \mapsto \Psi(wx + b) \in \mathbb{R}$ と定義する．このとき, 任意の $k \in \mathbb{N}, w_0 \in W, b \in I$ に対

して, 関数 $x \mapsto ((\partial w)^k f_b)(x, w_0)$ が $\mathcal{N}_1(\Psi, W, I)$ の元で $K$ 上一様近似できることを示せば十分である. なぜなら, もしそれが示されれば, $\Psi$ が $I$ 上で多項式関数でないことから, 任意の $k$ について $\Psi^{(k)}(b_k) \neq 0$ なる $b_k \in I$ が取れるので, $w_0 = 0, b = b_k$ とすることで, 関数 $x \mapsto ((\partial w)^k f_{b_k})(x, w_0) = x^k \Psi^{(k)}(w_0 x + b_k) = x^k \Psi^{(k)}(b_k)$ が $\mathcal{N}_1(\Psi, W, I)$ の元で $K$ 上一様近似できることがわかる. さて, まず $k = 1$ の場合を示そう. 任意に $\delta_0, \varepsilon > 0$ を取る. いま,

$$K \times [w_0 - \delta_0, w_0 + \delta_0] \ni (x, w) \mapsto ((\partial w) f_b)(x, w) \in \mathbb{R}$$

は連続である. $K \times [w_0 - \delta_0, w_0 + \delta_0]$ はコンパクトだから, $0 < \delta < \delta_0$ が存在して, 任意の $(x_1, w_1), (x_2, w_2) \in K \times [w_0 - \delta_0, w_0 + \delta_0]$ について, $|x_1 - x_2|, |w_1 - w_2| < \delta$ ならば, $|((\partial w) f_b)(x_1, w_1) - ((\partial w) f_b)(x_2, w_2)| < \varepsilon$ となる. そこで, $-\delta < h < \delta$ かつ $w_0 + h \in W$ となるように $h$ をとると ($W$ は孤立点を持たないのでとれる), 任意の $x \in K$ に対して, 平均値の定理より,

$$\begin{aligned}
&\left| \frac{\Psi((w_0 + h)x + b) - \Psi(w_0 x + b)}{h} - ((\partial w) f_b)(x, w_0) \right| \\
&\left| \frac{f_b(x, w_0 + h) - f_b(x, w_0)}{h} - ((\partial w) f_b)(x, w_0) \right| \\
&= |((\partial w) f_b)(x, w_0 + \theta h) - ((\partial w) f_b)(x, w_0)| \quad (\exists \theta \in [0, 1]) \\
&\leq \varepsilon
\end{aligned}$$

となる. これで $k = 1$ の場合が証明された. 次に $k$ で成立するとするとして $k + 1$ での成立を示そう. $k = 1$ の場合の証明と同様にして, 任意の $\varepsilon > 0$ に対してある $h \in \mathbb{R}$ が存在して, $w_0 + h \in W$ かつ任意の $x \in K$ に対して,

$$\left| \frac{((\partial w)^k f_b)(x, w_0 + h) - ((\partial w)^k f_b)(x, w_0)}{h} - ((\partial w)^{k+1} f_b)(x, w_0) \right| < \varepsilon$$

となることがわかる. 帰納法の仮定から,

$$\begin{aligned}
&x \mapsto ((\partial w)^k f_b)(x, w_0 + h) \\
&x \mapsto ((\partial w)^k f_b)(\cdot, w_0)
\end{aligned}$$

は $\mathcal{N}_1(\Psi, W, I)$ で $K$ 上一様近似できる. ゆえに, 上式と合わせると $((\partial w)^{k+1} f_b)(\cdot, w_0)$ は $\mathcal{N}_1(\Psi, W, I)$ で $K$ 上一様近似される. これで示せた. □

次に一般の $\Psi$ の場合を証明するためにいくつか補題を示す.

**補題 4.1.6.** $\Psi \in \mathcal{M}(\mathbb{R})$ とし, $\mu$ を $\mathbb{R}$ 上の Lebesgue 測度とする. このとき, $a \leq b$ と $\delta > 0$ に対して有限個の開区間 $I_1, \ldots, I_n$ が存在して, $U = \bigcup_{k=1}^n I_k$ について, $\mu(U) < \delta$ かつ $\Psi$ は $[a, b] \setminus U$ 上で一様連続である.

**証明.** $A := [a,b] \cap \mathrm{cl}\{\Psi$の不連続点$\}$ とおく．$\Psi \in \mathcal{M}(\mathbb{R})$ なので $\mu(A) = 0$ である．すると, Lebesgue 測度の外正則性より, 開集合 $U \supset A$ で $\mu(U) < \delta$ なるものが取れる．$\mathbb{R}$ の開集合は開区間の和集合で表せるので, $U = \bigcup_{j \in J} I_j$ と表すことにすると, $A$ のコンパクト性より $j_1, \ldots, j_m \in J$ が存在して $A \subset \bigcup_{k=1}^m I_{j_k}$ となる．$I = \bigcup_{k=1}^m I_{j_k}$ とおくと $[a,b] \setminus I$ は $\Psi$ の不連続点を含まないコンパクト集合である．コンパクト集合上の連続関数は一様連続であるので $I_{j_1}, \ldots, I_{j_m}$ が求めるものである． □

**補題 4.1.7.** $\Psi \in \mathcal{M}(\mathbb{R})$ とする．このとき, 任意の $\rho \in C_0^\infty(\mathbb{R})$ に対して, $\Psi$ と $\rho$ の畳み込み積

$$(\Psi * \rho)(x) = \int_{\mathbb{R}} \Psi(x-y)\rho(y)dy$$

は $\mathcal{N}_1(\Psi, \{1\}, \mathbb{R})$ により広義一様近似される．

**証明.** コンパクト集合 $K \subset \mathbb{R}$ をとる．$\alpha > 0$ を適当に取り $K$ と $\mathrm{supp}(\rho)$ が $[-\alpha, \alpha]$ に含まれるようにする．$[-\alpha, \alpha]$ 上で一様近似できることを示せば十分である．そこで, $m \in \mathbb{N}$ に対して $y_j = -\alpha + 2j\alpha/m, \Delta y_i = 2\alpha/m$ $(i = 1, \ldots, m)$ とおき,

$$x \mapsto \sum_{j=1}^m \Psi(x - y_j)\rho(y_j)\Delta y_j$$

が $\Psi * \rho$ に $[-\alpha, \alpha]$ 上一様収束することを示そう．$\rho \neq 0$ としてよい．$\varepsilon > 0$ を任意に取る．$0 < \delta < \alpha$ を次をみたすように取る．

$$10\delta \|\Psi\|_{L^\infty([-\alpha,\alpha])} \|\rho\|_{L^\infty} \leq \varepsilon.$$

いま, $\Psi \in \mathcal{M}(\mathbb{R})$ なので, $r(\delta)$ 個の開区間が存在して, それらの和集合を $U$ とすると, その Lebesgue 測度 $\mu(U)$ は $\mu(U) < \delta$ をみたし, かつ $\Psi$ は $[-2\alpha, 2\alpha] \setminus U$ 上で一様連続となる．そして, $m \in \mathbb{N}$ を $m\delta > \alpha r(\delta)$ かつ,

$$\begin{aligned} &|s-t| < 2\alpha/m \Rightarrow |\rho(s) - \rho(t)| \leq \frac{\varepsilon}{2\alpha \|\Psi\|_{L^\infty([-2\alpha,2\alpha])}} \\ &s, t \in [-2\alpha, 2\alpha] \setminus U, |s-t| < 2\alpha/m \Rightarrow |\Psi(s) - \Psi(t)| \leq \frac{\varepsilon}{\|\rho\|_{L^1}} \end{aligned}$$

となるように取る ($\rho$ は一様連続であることに注意)．ここで $x \in [-\alpha, \alpha]$ を固定する．$\Delta_j = [y_{j-1}, y_j]$ とおく (ただし, $y_0 = \alpha$ とする) と, $\mathrm{supp}(\rho) \subset [-\alpha, \alpha]$ であるから,

$$\begin{aligned} &\left| \int_{\mathbb{R}} \Psi(x-y)\rho(y)dy - \sum_{j=1}^m \int_{\Delta_j} \Psi(x-y_j)\rho(y)dy \right| \\ &\leq \sum_{j=1}^m \int_{\Delta_j} |\Psi(x-y) - \Psi(x-y_j)||\rho(y)|dy \end{aligned}$$

となる．$y \in \Delta_j$ について $|y - y_j| < 2\alpha/m$ であり，$x - y \in [-2\alpha, 2\alpha]$ であるので，もし $(x - \Delta_j) \cap U = \emptyset$ であるならば，

$$\int_{\Delta_j} |\Psi(x-y) - \Psi(x-y_j)||\rho(y)|dy \leq \frac{\varepsilon}{\|\rho\|_{L^1}} \int_{\Delta_j} |\rho(y)|dy$$

となる．ゆえに，そのような $j$ すべてについて和を取ると，その和は $\varepsilon$ で抑えられる．また，$(x - \Delta_j) \cap U \neq \emptyset$ となる $\Delta_j$ を $\tilde{\Delta}_j$ でかくことにすると，$U$ の Lebesgue 測度が $\delta$ 未満であること及び $U$ が $r(\delta)$ 個の開区間の和集合であることから，$\tilde{\Delta}_j$ の Lebesgue 測度の総和は $\delta + (4\alpha/m)r(\delta)$ で抑えられる．したがって，$m$ の選び方から $\tilde{\Delta}_j$ の Lebesgue 測度の総和は $5\delta$ で抑えられる．よって，$\delta$ の取り方から，

$$\sum \int_{\tilde{\Delta}_i} |\Psi(x-y) - \Psi(x-y_j)||\rho(y)|dy \leq 2\|\Psi\|_{L^\infty([-2\alpha,2\alpha])}\|\rho\|_{L^\infty} 5\delta \leq \varepsilon$$

となる．最後に，$\rho$ に対する $m$ の取り方から，

$$\begin{aligned}
&\left|\sum_{j=1}^m \int_{\Delta_j} \Psi(x-y_j)\rho(y)dy - \sum_{j=1}^m \Psi(x-y_j)\rho(y_j)\Delta y_j\right| \\
&= \left|\sum_{j=1}^m \int_{\Delta_j} \Psi(x-y_j)(\rho(y) - \rho(y_j))dy\right| \\
&\leq \sum_{j=1}^m \int_{\Delta_j} |\Psi(x-y_j)|\ |\rho(y) - \rho(y_j)|dy \\
&\leq \sum_{j=1}^m \int_{\Delta_j} \|\Psi\|_{L^\infty([-2\alpha,2\alpha])} \frac{\varepsilon}{2\alpha\|\Psi\|_{L^\infty([-2\alpha,2\alpha])}} dy \\
&= \varepsilon
\end{aligned}$$

となる．以上から，任意の $x \in [-\alpha, \alpha]$ に対して，三角不等式により，

$$\left|\int_{\mathbb{R}} \Psi(x-y)\rho(y)dy - \sum_{j=1}^m \Psi(x-y_j)\rho(y_j)\Delta y_j\right| \leq 3\varepsilon.$$

□

**注意 4.1.8.** 上の定理で「$\mathcal{N}_1(\Psi, \{1\}, \mathbb{R})$ により広義一様近似される」と述べたが，$y_j \notin \mathrm{supp}(\rho)$ のとき $\Psi(x - y_j)\rho(y_j) = 0$ であるので，$\mathcal{N}_1(\Psi, \{1\}, \mathrm{supp}(\rho))$ により広義一様近似されることがわかる．

**補題 4.1.9.** $N \in \mathbb{N}$ とする．$p_j : \mathbb{R} \to \mathbb{R}$ を $\deg(p_j) \leq N$ なる多項式関数の列とする．このとき，$f : \mathbb{R} \to \mathbb{R}$ に対して，$p_j$ が $f$ に各点収束するならば，$f$ はある $N$ 次以下の多項式と一致する．また，$p_j$ が $f$ に概収束する (つまり，a.e.$x \in \mathbb{R}$ に対して $p_j(x) \to f(x)$ となる) ならば，$f$ はある $N$ 次以下の多項式と a.e. で一致する．

**証明.** まず, 各点収束の場合を考える. $p_j(x) = \sum_{k=0}^{N} a_{j,k}x^k$ と表すことにする. 相異なる $x_0, x_1, \ldots, x_N \in \mathbb{R}$ を任意に取り,

$$\mathbf{p}^{(j)} = \begin{pmatrix} p_j(x_0) \\ p_j(x_1) \\ \vdots \\ p_j(x_N) \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_0^1 & \cdots & x_0^N \\ 1 & x_1^1 & \cdots & x_1^N \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_N^1 & \cdots & x_N^N \end{pmatrix},$$

$$\mathbf{a}^{(j)} = \begin{pmatrix} a_{j,0} \\ a_{j,1} \\ \vdots \\ a_{j,N} \end{pmatrix}, \quad \mathbf{f} = \begin{pmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_N) \end{pmatrix}$$

とおく. すると, Vandermonde 行列 $\mathbf{X}$ は正則であり,

$$\mathbf{p}^{(j)} - \mathbf{f} = \mathbf{X}\mathbf{a}^{(j)} - \mathbf{f}$$

である. したがって, $p_j(x_i) \to f(x_i)$ であることと行列写像が連続であることから,

$$\mathbf{a}^{(j)} = \mathbf{X}^{-1}\mathbf{f} + \mathbf{X}^{-1}(\mathbf{p}^{(j)} - \mathbf{f}) \to \mathbf{X}^{-1}\mathbf{f} \quad (j \to \infty)$$

となる. よって, $(a_1, \ldots, a_N)^{\mathrm{T}} = \mathbf{X}^{-1}\mathbf{f}$ とおき, $p(x) = \sum_{k=0}^{N} a_k x^k$ とおくと, 各点で $p_j \to p$ となる. よって, $f = p$ となる. 概収束の場合も同様に示せる (Lebesgue 測度で考えているので収束点を $N+1$ 個取れることに注意). □

**補題 4.1.10.** $\Psi \in \mathcal{M}(\mathbb{R})$ とする. このとき, $\rho_j \in C_0^\infty(\mathbb{R})$ $(j = 1, 2, \ldots)$ が存在して, $\mathrm{supp}(\rho_j) \subset [-1/j, 1/j]$ かつ $\Psi * \rho_j$ は $\Psi$ に $\mathbb{R}$ 上概収束する.

**証明.** $\rho \in C_0^\infty(\mathbb{R})$ で $\int_{\mathbb{R}} \rho(x)dx = 1$ かつ $\mathrm{supp}(\rho) \subset [-1, 1]$ をみたすものをとる. 例えば,

$$k(t) := \begin{cases} e^{-\frac{1}{t}} & (t > 0) \\ 0 & (t \leq 0) \end{cases}$$

とおき, $H(x) := k(1 - |x|)$, $\rho(x) := H(x)/\int_{\mathbb{R}} H(y)dy$ とおけばこの条件を満たす ($k \in C^\infty(\mathbb{R})$ の丁寧な証明が松本 [40] の p.176〜178 にあるので参照されたい). そして,

$$\rho_j(x) := j\rho(jx) \quad (x \in \mathbb{R}, j \in \mathbb{N})$$

とおく. すると, $\mathrm{supp}(\rho_j) \subset [-1/j, 1/j]$ ゆえ $\rho_j \in C_0^\infty(\mathbb{R})$ であり, $\Psi$ の連続点 $x \in \mathbb{R}$ に対して,

$$\begin{aligned} (\Psi * \rho_j)(x) &= \int_{\mathbb{R}} \Psi(x - y)\rho_j(y)dy \\ &= \int_{\mathbb{R}} \Psi(x - z/j)\rho(z)dz \qquad (z = jy) \\ &\to \int_{\mathbb{R}} \Psi(x)\rho(z)dz = \Psi(x) \qquad (\text{優収束定理}) \end{aligned}$$

である．$\Psi$ に関する仮定より $\Psi$ の不連続点全体の集合の測度は 0 なのでこれで証明できた． □

以上で準備が整ったので定理を証明していこう．

**定理 4.1.11.** $W \subset \mathbb{R}$ は孤立点を持たず $0 \in W$ であるとする．また, $\Psi : \mathbb{R} \to \mathbb{R}$ は $\mathcal{M}(\mathbb{R})$ に属するとする．このとき, 開区間 $I$ に対して $\Psi$ が $I$ 上 a.e. に多項式と一致しないなら, $\mathcal{N}_1(\Psi, W, I)$ は $C(\mathbb{R})$ において広義一様収束の意味で稠密である．

**証明.** 開区間 $J = (c, d)$ と $\varepsilon$ を $J^\varepsilon := (c - \varepsilon, d + \varepsilon) \subset I$ となるように取る．$\Psi$ についての仮定より任意の $\rho \in C_0^\infty(\mathbb{R})$ に対して $\Psi * \rho$ は $C^\infty$ 級関数である (付録の命題 9.7.2 を参照)．また, 補題 4.1.7 より, $\mathrm{supp}(\rho) \subset (-\varepsilon, \varepsilon)$ なる任意の $\rho \in C_0^\infty(\mathbb{R})$ と任意の $w \in W, b \in I$ に対して, $x$ の関数

$$(\Psi * \rho)(wx + b) = \int_{\mathbb{R}} \Psi(wx + b - y)\rho(y)dy$$

は $y_i \in \mathrm{supp}(\rho) \subset (-\varepsilon, \varepsilon)$ を適当に取れば $\sum_i \Psi(wx + b - y_i)\rho(y_i)\Delta y_i$ で広義一様近似される．ゆえに,

$$\overline{\mathcal{N}_1(\Psi * \rho, W, I)} \subset \overline{\mathcal{N}_1(\Psi, W, J^\varepsilon)} \subset \overline{\mathcal{N}_1(\Psi, W, I)}$$

がわかる (ただし, ここでの閉包は広義一様収束の意味である)．いま $\Psi * \rho$ は $C^\infty$ 級だから, 補題 4.1.5 の証明より, 任意の自然数 $k$ に対して $x \mapsto x^k(\Psi * \rho)^{(k)}(b)$ は $\overline{\mathcal{N}_1((\Psi * \rho, W, I)}$ に含まれることがわかる．Stone-Weierstrass の定理より多項式関数全体は $C(\mathbb{R})$ において広義一様収束の意味で稠密であるので, もし $\mathcal{N}_1(\Psi, W, I)$ が $C(\mathbb{R})$ において広義一様収束の意味で稠密でないならば, ある $m$ が存在して $\mathrm{supp}(\rho) \subset (-\varepsilon, \varepsilon)$ なる任意の $\rho \in C_0^\infty(\mathbb{R})$ に対して $(\Psi * \rho)^{(m)}(b) = 0$ $(\forall b \in I)$ とならなければならない．したがって, このとき, $\mathrm{supp}(\rho) \subset (-\varepsilon, \varepsilon)$ なる任意の $\rho \in C_0^\infty(\mathbb{R})$ に対して $\Psi * \rho$ は $I$ 上で $m$ 次以下の多項式となる．ところが, 補題 4.1.10 より $\rho_j \in C_0^\infty(\mathbb{R})$ が存在して, $\mathrm{supp}(\rho_j) \subset (-\varepsilon, \varepsilon)$ かつ a.e.$x \in I$ に対して $\Psi * \rho_j(x) \to \Psi(x)$ となる．よって, 補題 4.1.9 より $\Psi$ は $I$ 上である多項式と a.e. で一致する．しかし, これは我々の仮定に反する． □

以上で Leshno らの主結果が証明できた．なお, 第 2 章で述べたように Leshno らの結果における連続性の仮定は Hornik の指摘 [13] により次の形に緩和された．

**定理 (Hornik 1993 [13]).** $W \subset \mathbb{R}^r$ は 0 を内点として持つとする．また, $\Psi : \mathbb{R} \to \mathbb{R}$ は各閉区間上で有界かつ Riemann 積分可能であるとする．さらに, ある開区間 $I$ に対して $\Psi$ が $I$ 上で多項式と一致しないとする．このとき, $\mathcal{A} \subset \mathbb{R}^r$ 上で 0 を取る非自明な $r$ 変数斉次多項式が存在しないならば, $\mathcal{N}_r(\Psi, W, I)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である．

この定理は次の定理に含まれる．なぜなら, 0 を内点として持つ $W \subset \mathbb{R}^r$ に対して, $\mathcal{A} \subset \mathbb{R}^r$ として原点中心の開球をとり十分小さい $\varepsilon > 0$ に対して $W' = (-\varepsilon, \varepsilon)$ とおけば $W'\mathcal{A} \subset W$ となるからである (開球上で多項式は一意に決定されることに注意).

**定理 4.1.12.** $W \subset \mathbb{R}$ は孤立点を持たず $0 \in W$ であるとする．また, $\Psi : \mathbb{R} \to \mathbb{R}$ は各閉区間上で有界かつ Riemann 積分可能であるとする．さらに, ある開区間 $I$ に対して $\Psi$ が $I$ 上で多項式と一致しないとする．このとき, $\mathcal{A} \subset \mathbb{R}^r$ 上で 0 を取る非自明な $r$ 変数斉次多項式が存在しないならば,

$$\mathcal{N}_r(\Psi, W\mathcal{A}, I) = \mathrm{span}\{\mathbb{R}^r \ni x \mapsto \Psi(wa^T x + b) \in \mathbb{R} \mid w \in W, a \in \mathcal{A}, b \in I\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

さらにこの定理を示すためには, Leshno らの結果の証明と同様にして次が成り立つことを示せばよい.

**命題 4.1.13.** 各閉区間上で有界かつ Riemann 積分可能な $\Psi : \mathbb{R} \to \mathbb{R}$ に対して, 任意の $\rho \in C_0^\infty(\mathbb{R})$ に対して, $\Psi$ と $\rho$ の畳み込み積

$$(\Psi * \rho)(x) = \int_{\mathbb{R}} \Psi(x - y)\rho(y)dy$$

は $\mathcal{N}_1(\Psi, \{1\}, \mathrm{supp}(\rho))$ により広義一様近似される

**証明.** 空でないコンパクト集合 $K \subset \mathbb{R}$ を任意に取る．$\alpha > 0$ を適当に取り $K$ と $\mathrm{supp}(\rho)$ が $[-\alpha, \alpha]$ に含まれるようにする．$m \in \mathbb{N}$ に対して

$$y_j = -\alpha + 2j\alpha/m, \quad \Delta y_j = 2\alpha/m \qquad (j = 1, \ldots, m)$$

とおき,

$$x \mapsto \sum_{j=1}^{m} \Psi(x - y_j)\rho(y_j)\Delta y_j$$

が $\Psi * \rho$ に $[-\alpha, \alpha]$ 上一様収束することを示そう．任意の $x \in [-\alpha, \alpha]$ に対して,

$$\begin{aligned}
&\left| \int_{\mathbb{R}} \Psi(x - y)\rho(y)dy - \sum_{j=1}^{m} \Psi(x - y_j)\rho(y_j)\Delta y_j \right| \\
&\leq \left| \int_{\mathbb{R}} \Psi(x - y)\rho(y)dy - \sum_{j=1}^{m} \int_{\Delta_j} \Psi(x - y_j)\rho(y)dy \right| \\
&\quad + \left| \sum_{j=1}^{m} \int_{\Delta_j} \Psi(x - y_j)\rho(y)dy - \sum_{j=1}^{m} \Psi(x - y_j)\rho(y_j)\Delta y_j \right|
\end{aligned}$$

である. 右辺の二つの項がそれぞれ $[-\alpha,\alpha]$ 上一様に 0 に収束することを示せばよい. まず, 第一項について考える. $\mathrm{supp}(\rho)\subset[-\alpha,\alpha]$ であるから, 任意の $x\in[-\alpha,\alpha]$ に対して,

$$\begin{aligned}
&\left|\int_{\mathbb{R}}\Psi(x-y)\rho(y)dy-\sum_{j=1}^{m}\int_{\Delta_j}\Psi(x-y_j)\rho(y)dy\right|\\
&\le\sum_{j=1}^{m}\int_{\Delta_j}|\Psi(x-y)-\Psi(x-y_j)||\rho(y)|dy\\
&\le\|\rho\|_{L^\infty}\sum_{j=1}^{m}\left(\sup_{y\in\Delta_j}\Psi(x-y)-\inf_{y\in\Delta_j}\Psi(x-y)\right)\frac{2\alpha}{m}
\end{aligned}$$

となる. $[-2\alpha,2\alpha]$ を $2m$ 等分する分割 $\Delta'_j$ $(j=1,\dots,2m)$ について,

$$\begin{aligned}
&\sum_{j=1}^{m}\left(\sup_{y\in\Delta_j}\Psi(x-y)-\inf_{y\in\Delta_j}\Psi(x-y)\right)\\
&\le\sum_{j=1}^{2m}\left(\sup_{z\in\Delta'_j}\Psi(z)-\inf_{z\in\Delta'_j}\Psi(z)\right)
\end{aligned}$$

となるので, $\Psi$ が $[-2\alpha,2\alpha]$ 上 Riemann 積分可能であることより, $[-\alpha,\alpha]$ 上一様に

$$\begin{aligned}
&\left|\int_{\mathbb{R}}\Psi(x-y)\rho(y)dy-\sum_{j=1}^{m}\int_{\Delta_j}\Psi(x-y_j)\rho(y)dy\right|\\
&\le\|\rho\|_{L^\infty}\sum_{j=1}^{2m}\left(\sup_{z\in\Delta'_j}\Psi(z)-\inf_{z\in\Delta'_j}\Psi(z)\right)\frac{2\alpha}{m}\\
&\to\|\rho\|_{L^\infty}\left(\int_{-2\alpha}^{2\alpha}\Psi(z)dz-\int_{-2\alpha}^{2\alpha}\Psi(z)dz\right)=0\quad(m\to\infty)
\end{aligned}$$

となる. 次に第二項について考えよう. $\varepsilon>0$ を任意に取る. 任意の $x\in[-\alpha,\alpha]$ に対して,

$$\begin{aligned}
&\left|\sum_{j=1}^{m}\int_{\Delta_j}\Psi(x-y_j)\rho(y)dy-\sum_{j=1}^{m}\Psi(x-y_j)\rho(y_j)\Delta y_j\right|\\
&=\left|\sum_{j=1}^{m}\int_{\Delta_j}\Psi(x-y_j)(\rho(y)-\rho(y_j))dy\right|\\
&\le\sum_{j=1}^{m}\int_{\Delta_j}|\Psi(x-y_j)|\ |\rho(y)-\rho(y_j)|dy\\
&\le\|\Psi\|_{L^\infty([-2\alpha,2\alpha])}\sum_{j=1}^{m}\int_{\Delta_j}|\rho(y)-\rho(y_j)|dy
\end{aligned}$$

である. そこで, $\rho$ が $\mathbb{R}$ 上一様連続であることに注意して, $N \in \mathbb{N}$ を次をみたすように取る.

$$|s-t| < 2\alpha/N \Rightarrow |\rho(s)-\rho(t)| \leq \frac{\varepsilon}{2\alpha\|\Psi\|_{L^\infty([-2\alpha,2\alpha])}}.$$

すると, 任意の $m \geq N$ と $x \in [-\alpha, \alpha]$ に対して,

$$\begin{aligned}&\left|\sum_{j=1}^{m}\int_{\Delta_j}\Psi(x-y_j)\rho(y)dy - \sum_{j=1}^{m}\Psi(x-y_j)\rho(y_j)\Delta y_j\right| \\ &\leq \|\Psi\|_{L^\infty([-2\alpha,2\alpha])}\sum_{j=1}^{m}\int_{\Delta_j}|\rho(y)-\rho(y_j)|dy \leq \varepsilon\end{aligned}$$

となる. 以上で証明できた. □

**系 4.1.14.** $\Psi : \mathbb{R} \to \mathbb{R}$ は各閉区間上で有界かつ Riemann 積分可能であるとする. このとき, $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることと, $\Psi$ がいかなる多項式関数とも a.e. で一致することはないことは同値である.

## 4.2 Chui と Li の結果

本節では Cybenko の結果 [8] を含む Chui と Li による次の結果を証明する.

**定理 4.2.1 (Chui and Li 1992 [2]).**
$\sigma : \mathbb{R} \to \mathbb{R}$ を連続な sigmoidal 関数とすると,

$$\mathcal{N}_r(\sigma, \mathbb{Z}^r, \mathbb{Z}) = \mathrm{span}\{\mathbb{R}^r \ni x \mapsto \sigma(m^{\mathrm{T}}x + k) \in \mathbb{R} \mid m \in \mathbb{Z}^r, k \in \mathbb{Z}\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

この定理をふたつの補題に分けて証明する. 補題の主張を述べるために言葉と記号の定義をしておく.

$X \subset \mathbb{R}^r$ に対して, 位相空間 $(X, \mathcal{O})$ を $\mathbb{R}^r$ の部分空間とする. すなわち $\mathcal{O} = \{X \cap U \mid U \subset \mathbb{R}^r, \mathrm{open}\}$ とする. また, $(X, \mathcal{O})$ 上の正則符号付き Borel 測度全体の集合を $M(X)$ で表すことにする (正則符号付き測度の定義については付録を参照されたい).

**定義 4.2.2 (discriminatory 関数).**
関数 $\sigma : \mathbb{R} \to \mathbb{R}$ が次の条件をみたすとき, discriminatory 関数と呼ぶ: 任意の空でないコンパクト集合 $X \subset \mathbb{R}^r$ に対して,

$$\forall \mu \in M(X), \left(\forall m \in \mathbb{Z}^r, k \in \mathbb{Z}, \int_X \sigma\left(m^{\mathrm{T}}x + k\right)d\mu(x) = 0\right) \Rightarrow \mu = 0.$$

この定義のもとで以下のふたつの命題が成り立つ.

**補題 4.2.3.**
$\sigma : \mathbb{R} \to \mathbb{R}$ を連続な discriminatory 関数とすると, $\mathcal{N}_r(\sigma, \mathbb{Z}^r, \mathbb{Z})$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

**補題 4.2.4.**
有界で可測な sigmoidal 関数 $\sigma$ は discriminatory 関数である. 特に連続な sigmoidal 関数は discriminatory 関数である.

上記ふたつの補題から定理 4.2.1 が導かれることは明らかであろう. 以下でこれらの補題の証明をしていく.

**補題 4.2.3 の証明.**
$K \subset \mathbb{R}^r$ を空でないコンパクト集合とする. $\mathcal{N}_r(\sigma, \mathbb{Z}^r, \mathbb{Z})$ の元 (関数) の $K$ 上への制限すべてを集めた集合を $S$ とおく. $S$ は $C(K)$ の部分線形空間である. 補題 4.2.3 の主張は $\overline{S} = C(K)$ が成り立つということである. そこで, $\overline{S} \neq C(K)$ を仮定する (背理法). すると, Hahn-Banach の拡張定理により, $C(K)$ 上の有界線形汎関数 $L \neq 0$ で, $L = 0$ (on $\overline{S}$) なるものが取れる (付録の系 9.4.5 参照). $K$ はコンパクト Hausdorff 空間であるので, Riesz-Markov-角谷の表現定理から, $\mu \in M(K)$ で,

$$\forall f \in C(K), L(f) = \int_K f d\mu$$

をみたすものが取れる. 任意の $m \in \mathbb{Z}^n, k \in \mathbb{Z}$ に対して, 写像

$$K \ni x \mapsto \sigma(m^{\mathrm{T}} x + k) \in \mathbb{R}$$

は $S$ に属するので,

$$\int_K \sigma(m^{\mathrm{T}} x + k) d\mu(x) = 0$$

となる. ゆえに, $\sigma$ が discriminatory 関数であることから $\mu = 0$ となり, したがって $L = 0$ となるが, これは $L \neq 0$ に反する. よって, $\overline{S} = C(K)$ でなければならない. □

**補題 4.2.4 の証明.**
後半については連続性から可測性がわかり, sigmoidal 関数であることと連続性から有界性がわかるので成立する. 前半を証明する. $X \subset \mathbb{R}^r$ を空でないコンパクト集合とする. $\mu \in M(X)$ を任意にとり固定する. そして,

$$\forall m \in \mathbb{Z}^r, k \in \mathbb{Z}, \quad \int_X \sigma\left(m^{\mathrm{T}} x + k\right) d\mu(x) = 0$$

が成り立つと仮定する. このとき, $\mu = 0$ を示すのが目標である. さて, いま, 任意の $m \in \mathbb{Z}^r, q \in \mathbb{Q}, k \in \mathbb{Z}$ に対して,

$$\Pi_{m,q} = \left\{x \in X \mid m^{\mathrm{T}} x + q = 0\right\}, \quad H_{m,q} = \left\{x \in X \mid m^{\mathrm{T}} x + q > 0\right\}$$

とおき,

$$\gamma(x) := \begin{cases} 1 & x \in H_{m,q} \\ \sigma(k) & x \in \Pi_{m,q} \\ 0 & \text{otherwise} \end{cases}$$

と定めると, $\sigma$ は sigmoidal 関数なので, $+\infty$ に発散する自然数列 $l_n$ を任意に取ると, 任意の $x \in \mathbb{R}^r$ に対して,

$$\lim_{n\to\infty} \sigma(l_n(m^{\mathrm{T}}x + q) + k) \to \gamma(x)$$

となる. そこで, $q = q_1/q_2,\ q_1 \in \mathbb{Z}, q_2 \in \mathbb{N}$ と表すことにすると, $\sigma$ の有界性と $\mu$ の有限性および仮定より,

$$\begin{aligned} 0 &= \lim_{n\to+\infty} \int_X \sigma(nq_2(m^{\mathrm{T}}x + q) + k) d\mu(x) \\ &= \int_X \lim_{n\to+\infty} \sigma(nq_2(m^{\mathrm{T}}x + q) + k) d\mu(x) \\ &= \int_X \gamma(x) d\mu(x) \\ &= \sigma(k)\mu\left(\Pi_{m,q}\right) + \mu\left(H_{m,q}\right) \end{aligned}$$

となる (2 つ目の等号は優収束定理による). よって, $\sigma$ が sigmoidal 関数であることと $k \in \mathbb{Z}$ が任意であることに注意すると, $k \to -\infty$ とすることで, $\mu\left(H_{m,q}\right) = 0$ が得られ, さらに $k \to +\infty$ とすることで $\mu\left(\Pi_{m,q}\right) = 0$ が得られる. したがって, $\mathbb{R}$ 上の有界な可測関数のなす線形空間上の線形汎関数 $F$ を

$$F(h) = \int_X h(m^{\mathrm{T}}x) d\mu(x)$$

により定めると, 定義関数

$$h(x) = \chi_{[q,\infty)}(x) := \begin{cases} 1 & (x \in [q, \infty)) \\ 0 & (x \notin [q, \infty)) \end{cases}$$

に対して,

$$F(h) = \int_X h(m^{\mathrm{T}}x) d\mu(x) = \mu\left(\Pi_{m,-q}\right) + \mu\left(H_{m,-q}\right) = 0$$

となる. したがって, $q \in \mathbb{Q}$ は任意であったから, $F$ の線形性より有理数を端点とする任意の $\mathbb{R}$ の区間 $J$ について $F(\chi_J) = 0$ となる. したがって, 有限個の $\mathbb{R}$ の区間の定義関数の線形結合を $F$ で写すと 0 になる. 以下, 一旦 $m \in \mathbb{Z}^r$ を固定して考える. このとき $A = \{m^{\mathrm{T}}x \mid x \in X\}$ は $\mathbb{R}$ のコンパクト集合である. 何故なら $A$ はコンパクト集合 $X$ の連続写像による像だからである. そこで有理数を端点とする

有界閉区間 $J$ を $A \subset J$ をみたすように取れる．いま $J$ 上の任意の連続関数 $g$ は有理数を端点とする $\mathbb{R}$ の区間の定義関数たちの線形結合 $\sum_{k=1}^{N} a_k \chi_{J_k}$ により一様近似できる．すなわち任意の正数 $\varepsilon$ に対して, 適当な $a_k \in \mathbb{R}$ と有理数を端点とする区間 $J_k$ を 取れば,

$$\forall x \in J, \quad \left| g(x) - \sum_{k=1}^{N} a_k \chi_{J_k}(x) \right| < \varepsilon$$

とできる．実際, コンパクト集合上の連続関数は一様連続なので, 任意の $\varepsilon > 0$ に対して, $\delta > 0$ を適当に取れば,

$$\forall x, y \in J, |x - y| < \delta \Rightarrow |g(x) - g(y)| < \varepsilon$$

が成り立つ．そこで, 一つの区間の幅が $\delta$ に収まるような分割 $J = \sum_{k=1}^{N} J_k$ であって, 各区間 $J_k$ の端点が有理数になるものを取る．そして各 $k$ に対して $x_k \in J_k$ を任意に取れば,

$$\forall x \in J, \quad \left| g(x) - \sum_{k=1}^{N} g(x_k) \chi_{J_k}(x) \right| < \varepsilon$$

となる．したがって,

$$\begin{aligned}
|F(g)| &= \left| \int_X g(m^{\mathrm{T}} x) d\mu(x) \right| \\
&= \left| \int_X g(m^{\mathrm{T}} x) d\mu(x) - \int_X \sum_{k=1}^{N} a_k \chi_{J_k}(m^{\mathrm{T}} x) d\mu(x) \right| \\
&\leq \int_X \left| g(m^{\mathrm{T}} x) - \sum_{k=1}^{N} a_k \chi_{J_k}(m^{\mathrm{T}} x) \right| d\,|\mu|\,(x) \\
&\leq \int_X \varepsilon d|\mu|(x) = \varepsilon\,|\mu|\,(X)
\end{aligned}$$

となる．ただし $|\mu| = \mu^+ + \mu^-$ である．よって, $|\mu|\,(X)$ が $\varepsilon$ に依存しない定数 (有限値) であることと $\varepsilon$ の任意性から, $F(g) = 0$ がわかる．特に $\cos(x), \sin(x)$ についてもこれが成立するので,

$$0 = \int_X \cos(m^{\mathrm{T}} x) + i \sin(m^{\mathrm{T}} x) d\mu(x) = \int_X \exp(i m^{\mathrm{T}} x) d\mu(x)$$

がわかる．よって, $m \in \mathbb{Z}$ は任意だったので Stone-Weierstrass の定理 (付録の系 9.6.6) から $\mu = 0$ となる．これが示したいことであった． □

## 4.3 Sun と Cheney の結果

本節では Chui と Li の結果を自然にノルム空間 $X$ 上の連続関数空間 $C(X)$ に拡張した Sun と Cheney による次の結果を証明する．また, mean-periodic 関数などニューラルネットワークの万能近似定理に関連する事柄をいくつか紹介する．

**定理 4.3.1 (Sun and Cheney 1992 [33]).**
$X \neq \{0\}$ をノルム空間とし, $\mathcal{F} \subset X^*$ は

$$\{f/\|f\| \mid f \in \mathcal{F}, f \neq 0\}$$

が $X^*$ の単位球面 $\{f \in X^* \mid \|f\| = 1\}$ において作用素ノルム $\|\cdot\|$ に関して稠密であるものとする. このとき, 連続な sigmoidal 関数 $\sigma : \mathbb{R} \to \mathbb{R}$ に対して,

$$\mathcal{G} = \{\mathbb{R} \ni t \mapsto \sigma(kt + j) \in \mathbb{R} \mid k, j \in \mathbb{Z}\}$$

とおくと,

$$\mathcal{G} \circ \mathcal{F} = \{X \ni x \mapsto \sigma(kf(x) + j) \in \mathbb{R} \mid k, j \in \mathbb{Z}, f \in \mathcal{F}\}$$

の線形包は $C(X)$ において広義一様収束の意味で稠密である.

この定理を示すためにひとつ定理を示す. なお本節では記述を簡潔にするために次の言葉を用いる.

**定義 4.3.2.** $X$ をノルム空間とするとき, 部分集合 $S \subset C(X)$ が基礎的であるとは, $S$ の線形包が広義一様収束の意味で $C(X)$ において稠密であることをいう.

**定理 4.3.3.** $X \neq \{0\}$ をノルム空間とし, $\mathcal{G} \subset C(\mathbb{R})$ は基礎的であるとする. このとき, $\mathcal{F} \subset X^*$ に対して,

$$\{f/\|f\| \mid f \in \mathcal{F}, f \neq 0\}$$

が $X^*$ の単位球面 $\{f \in X^* \mid \|f\| = 1\}$ において作用素ノルム $\|\cdot\|$ に関して稠密であるならば, $\mathcal{G} \circ \mathcal{F} = \{g \circ f \mid g \in \mathcal{G}, f \in \mathcal{F}\}$ は $C(X)$ において基礎的である.

これが示されれば, Chui と Li の結果より $\mathcal{G}$ は $C(\mathbb{R})$ において基礎的であるので本節のメインの定理の成立は明らかである.

**証明.** $h \in C(X)$ とし, 任意に空でないコンパクト集合 $K \subset X$ と $\varepsilon > 0$ をとる. すると, 定理 3.1.5 より, $h_i \in C(\mathbb{R})$ と $\psi_i \in X^*$ が存在して,

$$\sup_{x \in K} \left| h(x) - \sum_{i=0}^{n} h_i(\psi_i(x)) \right| < \varepsilon/3$$

となる. $h_i$ たちを定数倍することで $\|\psi_i\| = 1$ と仮定してよい (0 がある場合は $\mathbb{R} \ni t \mapsto h_i(0) \in \mathbb{R}$ という定数関数を導入すればよい). ここで, $M = \sup_{x \in K} \|x\|$ とおき, $\delta > 0$ を次をみたすように取る ($h_i$ は連続なので取れる).

$$|h_i(s) - h_i(t)| < \varepsilon/3n \quad (|s| \leq M, |t| \leq M, |s - t| < M\delta, 1 \leq i \leq n).$$

そして, $f_i \in \mathcal{F}$ を $\|f_i/\|f_i\| - \psi_i\| < \delta$ をみたすように選ぶ. また, 各 $i$ について $\lambda_i = 1/\|f_i\|$ とおき, $a_{i,j} \in \mathbb{R}$ と $g_{i,j} \in \mathcal{G}$ を

$$\sup_{|t| \le M/\lambda_i} \left| h_i(\lambda_i t) - \sum_{j=0}^{N_i} a_{i,j} g_{i,j}(t) \right| < \varepsilon/3n$$

となるようにとる. すると, 任意の $x \in K$ に対して,

$$|f_i(x)| \le \|f_i\|\|x\| \le M\|f_i\| = M/\lambda_i$$

であり, また, $|\psi_i(x)| \le \|\psi_i\|\|x\| \le M$ かつ

$$|\lambda_i f_i(x) - \psi_i(x)| \le \|f_i/\|f_i\| - \psi_i\|\|x\| < M\delta$$

より $|h_i(\lambda_i f_i(x)) - h_i(\psi_i(x))| < \varepsilon/3n$ となる. よって, 三角不等式より,

$$\begin{aligned}
&\left| h(x) - \sum_{i=0}^{n} \sum_{j=0}^{N_i} a_{i,j} g_{i,j}(f_i(x)) \right| \\
&\le \left| h(x) - \sum_{i=0}^{n} h_i(\psi_i(x)) \right| + \left| \sum_{i=0}^{n} h_i(\psi_i(x)) - \sum_{i=0}^{n} h_i(\lambda_i f_i(x) \right| \\
&\quad + \left| \sum_{i=0}^{n} h_i(\lambda_i f_i(x)) - \sum_{i=0}^{n} \sum_{j=0}^{N_i} a_{i,j} g_{i,j}(f_i(x)) \right| \\
&\le \varepsilon/3 + \varepsilon/3 + \varepsilon/3 = \varepsilon.
\end{aligned}$$

□

**系 4.3.4.** 上の定理は $\mathcal{F} \subset X^*$ が単位球面において作用素ノルムに関して稠密であるという仮定のもとで成り立つ.

**系 4.3.5.** $X \neq \{0\}$ をノルム空間とし, $\mathcal{F} \subset X^*$ を上の定理と同様のものとする. また, $g \in C(\mathbb{R})$ とし, $A \subset \mathbb{R}$ とする. このとき,

$$\mathcal{G} = \{g_a : \mathbb{R} \ni t \mapsto g(t+a) \in \mathbb{R} \mid a \in A\}$$

が $C(\mathbb{R})$ において基礎的であるならば, $\mathcal{G} \circ \mathcal{F} = \{g_a \circ f \mid a \in A, f \in \mathcal{F}\}$ は $C(X)$ において基礎的である.

この系に関連して mean-periodic 関数について紹介する.

**定義 4.3.6.** 関数 $f \in C(\mathbb{R}^r)$ が mean-periodic であるとは,

$$\mathcal{N}_r(f, \{1\}, \mathbb{R}) = \mathrm{span}\{x \mapsto f(x+a) \mid a \in \mathbb{R}^r\}$$

が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密でないことをいう.

**定理 4.3.7 (Schwartz 1947 [30] p.907).**
$\Psi \in C(\mathbb{R})$ とする．このとき, ある $1 \leq p < \infty$ について $\Psi \in L^p(\mathbb{R})$ であるか, または $\Psi$ は有界で非定数かつ $x \to \infty$ か $x \to -\infty$ としたときに極限値を持つならば, $\Psi$ は mean-periodic でない.

これより上の定理の仮定をみたす $\Psi$ については, 任意の実数 $\lambda \neq 0$ について, $\mathbb{N}(\Psi, \{\lambda\}, \mathbb{R})$ は $C(\mathbb{R})$ において広義一様収束の意味で稠密である．さらに, Lin と Pinkus の結果 (定理 3.1.8) を用いると次の命題の成立もわかる.

**命題 4.3.8.** $\Psi$ は上の定理の仮定をみたすものとする．このとき, $A \subset \mathbb{R}^r$ 上で 0 を取る非自明な $r$ 変数斉次多項式が存在しないならば,

$$\mathcal{N}_r(\Psi, A, \mathbb{R}) = \mathrm{span}\{\mathbb{R}^r \ni x \mapsto \Psi(\langle x, v\rangle + b) \mid v \in A, b \in \mathbb{R}\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

この命題を二段階に分けて証明しよう.

**補題 4.3.9.** $X$ をノルム空間とし, $E \subset C(\mathbb{R})$ は $C(\mathbb{R})$ において基礎的であるとする．また, $\mathcal{F} \subset X^*$ は $C(\mathbb{R}) \circ \mathcal{F}$ が $C(X)$ において基礎的なものとする．このとき, $\{g \circ f \mid g \in E, f \in \mathcal{F}\}$ は $C(X)$ において基礎的である.

**証明.** $K \subset X$ を空でないコンパクト集合とし, 任意に $\varphi \in C(X)$ と $\varepsilon > 0$ を取る．仮定より, $\beta_i \in \mathbb{R}, g_i \in C(\mathbb{R}), f_i \in \mathcal{F}$ $(i = 1, \ldots, q)$ が存在して,

$$\sup_{x \in K} |\sum_{i=1}^{q} \beta_i g_i(f_i(x)) - \varphi(x)| < \varepsilon/2$$

となる．各 $i$ について $f_i$ は連続であるから $f_i(K) \subset \mathbb{R}$ はコンパクトである．そこで, 仮定より各 $i$ について $\alpha_{i,j} \in \mathbb{R}, g_{i,j} \in E$ $(j = 1, \ldots, m_i)$ が存在して

$$\sup_{y \in f_i(K)} |\sum_{j=1}^{m_i} \alpha_{i,j} g_{i,j}(y) - \beta_i g_i(y)| < \varepsilon/(2q)$$

となる．すると, 任意の $x \in K$ に対して,

$$\begin{aligned}
&\left|\sum_{i=1}^{q} \beta_i g_i(f_i(x)) - \sum_{i=1}^{q}\sum_{j=1}^{m_i} \alpha_{i,j} g_{i,j}(f_i(x))\right| \\
&\leq \sum_{i=1}^{q} \left|\beta_i g_i(f_i(x)) - \sum_{j=1}^{m_i} \alpha_{i,j} g_{i,j}(f_i(x))\right| \\
&\leq \sum_{i=1}^{q} \sup_{y \in f_i(K)} \left|\beta_i g_i(y) - \sum_{j=1}^{m_i} \alpha_{i,j} g_{i,j}(y)\right| \\
&< \varepsilon/2
\end{aligned}$$

であるから, $\sum_{i=1}^{q}\sum_{j=1}^{m_i} \alpha_{i,j} g_{i,j} \circ f_i$ が求めるものである. □

**命題 4.3.10.** $\Psi \in C(\mathbb{R})$ とし, $B \subset \mathbb{R}$ とする. また, 集合

$$S := \{\mathbb{R} \ni t \mapsto \Psi(t+b) \in \mathbb{R} \mid b \in B\}$$

は $C(\mathbb{R})$ において基礎的であるとする. このとき, $A \subset \mathbb{R}^r$ 上で $0$ を取る非自明な $r$ 変数斉次多項式が存在しないならば,

$$\mathcal{N}_r(\Psi, A, B) = \mathrm{span}\{\mathbb{R}^r \ni x \mapsto \Psi(\langle x, v\rangle + b) \mid v \in A, b \in B\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

**証明.** 仮定と Lin と Pinkus の結果より $C(\mathbb{R}) \circ \{x \mapsto \langle x, v\rangle \mid v \in A\}$ が基礎的であるのですぐ上の補題から明らかである. □

さて, 本節の冒頭で「Chui と Li の結果を自然にノルム空間 $X$ 上の連続関数空間 $C(X)$ に拡張した」と述べた. きちんと拡張になっていることを以下で確認しておく. つまり, $X = \mathbb{R}^s$ とおくとき, $\mathcal{F} = \{x \mapsto \langle x, v\rangle \mid v \in \mathbb{Z}^s\}$ とおくと,

$$\{f/\|f\| \mid f \in \mathcal{F}, f \neq 0\}$$

が $X^*$ の単位球面 $\{f \in X^* \mid \|f\| = 1\}$ において作用素ノルム $\|\cdot\|$ に関して稠密であることを確かめておく.

**補題 4.3.11.** $X, Y$ を位相空間とし, $f : X \to Y$ を全射連続写像とする. このとき, $A \subset X$ が稠密であるならば, $f(A) \subset Y$ は稠密である.

**証明.** $y \in Y$ をとり, $y$ の近傍 $V$ を任意に取る. $f$ は全射なので $x \in X$ で $y = f(x)$ をみたすものが取れる. すると, $f^{-1}(V)$ は $x$ の近傍となる. そこで, 仮定より $a \in A$ で $a \in f^{-1}(V)$ なるものが取れる. したがって $f(a) \in V$ であるから, $f(A) \cap V \neq \emptyset$ である. これは $f(A)$ が $Y$ で稠密であることを意味する. □

**命題 4.3.12.** $X = \mathbb{R}^s$ とおくとき, $\mathcal{F} = \{x \mapsto \langle x, v\rangle \mid v \in \mathbb{Z}^s\}$ とおくと,

$$\{f/\|f\| \mid f \in \mathcal{F}, f \neq 0\}$$

は $X^*$ の単位球面 $\{f \in X^* \mid \|f\| = 1\}$ において作用素ノルム $\|\cdot\|$ に関して稠密である.

**証明.** シュワルツの不等式から $\mathbb{R}^s \ni x \mapsto \langle x, u\rangle \in \mathbb{R}$ の作用素ノルムは $\|u\|$ と一致するので,

$$\{f/\|f\| \mid f \in \mathcal{F}, f \neq 0\}$$

が $\{\mathbb{R}^s \ni x \mapsto \langle x, u\rangle \in \mathbb{R} \mid u \in \mathbb{R}^s, \|u\| = 1\}$ において作用素ノルムの意味で稠密であることを示せばよい. そして, このことを示すには,

$$S = \{m/\|m\| \mid m \in \mathbb{Z}, z \neq 0\}$$

が $U = \{u \in \mathbb{R}^s \mid \|u\| = 1\}$ においてユークリッド距離に関して稠密であることを示せば十分である．いま, 写像 $\varphi$ を

$$\varphi : \mathbb{R}^s \setminus \{0\} \to U,\ \varphi(x) = x/\|x\|$$

と定義すると, これは連続かつ全射である．そして, $\mathbb{Q}^s \setminus \{0\}$ は $\mathbb{R}^s \setminus \{0\}$ において稠密であるから, すぐ上の補題より

$$\varphi(\mathbb{Q}^s \setminus \{0\}) = \{q/\|q\| \mid q \in \mathbb{Q}^s, q \neq 0\}$$

は $U$ において稠密である．しかも, $\varphi(\mathbb{Q}^s \setminus \{0\}) = S$ である．実際, $q \in \mathbb{Q}^s \setminus \{0\}$ を取ると, $n \in \mathbb{Z}^s$ と $m \in \mathbb{Z}$ を使って $q = n/m$ と表せるので, $q/\|q\| = n/\|n\| \in S$ となる．逆向きの包含関係は明らかである．よって, $S$ は $U$ において稠密である．以上で系は示された． □

なお, この命題から次のことも直ちにわかる．

**命題 4.3.13.** $\mathcal{G} \subset C(\mathbb{R})$ が基礎的であるとすると,

$$\{\mathbb{R}^s \ni x \mapsto g(\langle x, v\rangle) \in \mathbb{R} \mid g \in \mathcal{G}, v \in \mathbb{Z}^s\}$$

は $C(\mathbb{R}^s)$ において基礎的である．

## 4.4 Hornik の結果

本節では Hornik の結果, つまり, 活性化関数 $\Psi$ が squashing 関数であるときに $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることの証明を与える．ただし, squashing 関数の定義は次の通りである．

**定義 4.4.1 (squashing 関数).**
関数 $\Psi : \mathbb{R} \to [0, 1]$ が squashing 関数であるとは, $\Psi$ が単調非減少であり,

$$\Psi(t) \to \begin{cases} 1 & (t \to +\infty) \\ 0 & (t \to -\infty) \end{cases}$$

が成り立つことをいう．つまり, 単調非減少な sigmoidal 関数を squashing 関数という．

**注意 4.4.2.** squashing 関数は Borel 可測である．なぜなら, 単調非減少関数は Borel 可測だからである．実際, $f : \mathbb{R} \to \mathbb{R}$ を単調非減少関数とするとき, 任意の $a \in \mathbb{R}$ に対して, $x = \sup\{y \in \mathbb{R} \mid f(y) \leq a\}$ とおくと,

$$f^{-1}((-\infty, a]) = (-\infty, x) \text{ or } (-\infty, x] \in \mathcal{B}_{\mathbb{R}}$$

となるから, $f$ は Borel 可測である．

squashing 関数には連続性が要求されていないことに注意されたい．Hornik は連続性の仮定をなくすことができるかということに関心があったようである [12]. Leshno et al. の結果の節で紹介した Hornik の結果にもこの姿勢は表れている．

さて, まず次の ΣΠ-ネットワークについて考察していく．

**定理 4.4.3.** $G : \mathbb{R} \to \mathbb{R}$ を定数でない連続関数とする．このとき,

$$\Sigma\Pi^r(G) := \left\{ f : \mathbb{R}^r \to \mathbb{R} \;\middle|\; \begin{array}{l} f = \displaystyle\sum_j^q \beta_j \prod_{k=1}^{l_j} G \circ A_{j,k} \\ \beta_j \in \mathbb{R}, A_{j,k} \in \mathbb{A}^r, l_j \in \mathbb{N}, q \in \mathbb{N} \end{array} \right\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である．

**証明.** 任意に空でないコンパクト集合 $K \subset \mathbb{R}^r$ を取る．すると $\Sigma\Pi^r(G)_K$ は $C(K)$ の部分代数である．実際, 線形部分空間であることは明らかである．また, $f, g \in \Sigma\Pi^r(G)_K$ ならば $fg \in \Sigma\Pi^r(G)_K$ であることも有限個の組 $(i, j)$ を一列化することで示される．いま, $G$ は定数でないので $G(a) \neq 0$ なる $a \in \mathbb{R}$ が取れる．このとき, $A(x) := 0^{\mathrm{T}}x + a = a$ と定義すると, $G \circ A/G(a) \in \Sigma\Pi^r(G)_K$ となり, $G(A(x))/G(a) = 1 \quad (\forall x \in K)$ となる．次に, $x, y \in K,\ x \neq y$ をとる．$G$ は定数でないので, $G(a) \neq G(b)$ なる $a, b \in \mathbb{R}$ が取れる．いま, $A(z) = w^{\mathrm{T}}z + \beta$ で $A(x) = a, A(y) = b$ なるものが取れる．実際, $x \neq y$ ゆえ, ある $i$ で $x_i - y_i \neq 0$ となることに注意して, $w_i = (a-b)/(x_i - y_i), w_j = 0\ (j \neq i)$ として, $\beta = (bx_i - ay_i)/(x_i - y_i)$ とおけば良い．このとき, $G(A(x)) = a \neq b = G(A(y))$ である．よって, Stone-Weierstrass の定理によって, $\Sigma\Pi^r(G)_K$ は $C(K)$ において一様収束の意味で稠密である．以上から, $\Sigma\Pi^r(G)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である． □

**補題 4.4.4.** $\Psi : \mathbb{R} \to \mathbb{R}$ を squashing 関数とする．このとき, 任意の連続な squashing 関数 $F$ と $\varepsilon > 0$ に対して, ある $H \in \Sigma^1(\Psi)$ が存在して,

$$\sup_{x \in \mathbb{R}} |F(x) - H(x)| < \varepsilon.$$

**証明.** 任意に $\varepsilon >$ を取る．$\varepsilon < 1$ としてよい．$\beta_j \in \mathbb{R}$ と $A_j \in \mathbb{A}^1$ で,

$$\sup_{x \in \mathbb{R}} |F(x) - \sum_{j=0}^{Q-1} \beta_j \Psi(A_j(x))| < \varepsilon$$

なるものを見つけたい．そこで $Q \in \mathbb{N}$ を $1/Q < \varepsilon/2$ を満たすように取る．そして, $\beta_j = 1/Q$ とおく．いま $\Psi$ は squashing 関数なので $M > 0$ で

$$\Psi(-M) < \frac{1}{Q}, \quad \Psi(M) > 1 - \frac{1}{Q}$$

なるものが取れる．また, $F$ は連続な squashing 関数なので

$$
\begin{aligned}
r_j &= \sup\{x \in \mathbb{R} \mid F(x) = j/Q\} \quad (j = 1, \ldots, Q-1) \\
r_Q &= \sup\{x \in \mathbb{R} \mid F(x) = 1 - 1/(2Q)\}
\end{aligned}
$$

が存在する．そして, $r_0 \in (-\infty, r_1)$ を任意に取る．さて, $r < s$ に対して $A_{r,s} \in \mathbb{A}^1$ を $A_{r,s}(r) = -M, A_{r,s}(s) = M$ なるものとする (これは一意に定まる)．このとき, $A_j = A_{r_j, r_{j+1}}$ とすると,

$$
H = \sum_{j=0}^{Q-1} \beta_j \Psi \circ A_j \ \in \Sigma^1(\Psi)
$$

が求めるものであることがわかる．実際, $x \in (-\infty, r_0]$ の場合は, $0 \leq F(x) \leq 1/Q$ であり, かつ $\Psi$ が squashing 関数であることより,

$$
0 \leq H(x) \leq \sum_{j=0}^{Q-1} \frac{1}{Q} \Psi(-M) \leq \frac{1}{Q}
$$

となる．よって, $|F(x) - H(x)| \leq 1/Q < \varepsilon$ となる．次に, $x \in (r_0, r_1]$ の場合を考える．このとき, $0 \leq F(x) \leq 1/Q$ であり,

$$
\begin{aligned}
0 \leq H(x) &\leq \frac{1}{Q} + \sum_{j=1}^{Q-1} \frac{1}{Q} \Psi(-M) \\
&\leq \frac{1}{Q} + \frac{Q-1}{Q} \frac{1}{Q} \\
&\leq \frac{1}{Q} + \frac{1}{Q}
\end{aligned}
$$

となるので, $|F(x) - H(x)| \leq 1/Q + 1/Q \leq \varepsilon$ となる．次に $x \in (r_j, r_{j+1}]$ $(j = 1, \ldots, Q-2)$ の場合を考える．このとき, $r_j$ の定義より

$$
\frac{j}{Q} \leq F(x) \leq \frac{j+1}{Q}
$$

である．また,

$$
H(x) \geq \sum_{k=0}^{j-1} \frac{1}{Q} \Psi(M) \geq \frac{j}{Q} - \frac{j}{Q^2} \geq \frac{j-1}{Q}
$$

かつ

$$
\begin{aligned}
H(x) &\leq \sum_{k=0}^{j} \frac{1}{Q} + \sum_{k=j+1}^{Q-1} \frac{1}{Q} \Psi(-M) \\
&\leq \frac{j+1}{Q} + \frac{1}{Q} = \frac{(j+2}{Q}
\end{aligned}
$$

である．したがって，$|F(x)-H(x)| \le 2/Q \le \varepsilon$ となる．次に $x \in (r_{Q-1}, r_Q]$ の場合を考える．このとき，

$$1-\frac{1}{Q}=\frac{Q-1}{Q} \le F(x) \le 1-\frac{1}{2Q}$$

である．また，

$$H(x) \ge \sum_{j=0}^{Q-2}\frac{1}{Q}\Psi(M) \ge \frac{Q-1}{Q}-\frac{Q-1}{Q^2} \ge \frac{Q-2}{Q}=1-\frac{2}{Q}$$

かつ，

$$H(x) \le \sum_{j=0}^{Q-1}\frac{1}{Q}=1$$

である．したがって，$|F(x)-H(x)| \le \max\{1/Q, 2/Q-1/(2Q)\} \le \varepsilon$ となる．最後に，$x \in (r_Q, \infty)$ の場合を考える．このとき，$1-1/(2Q) \le F(x) \le 1$ であり，

$$1 \ge H(x) \ge \sum_{j=0}^{Q-1}\frac{1}{Q}\Psi(M) \ge 1-\frac{1}{Q}$$

である．したがって，$|F(x)-H(x)| \le 1/Q \le \varepsilon$ となる． □

**定理 4.4.5.** $\Psi : \mathbb{R} \to \mathbb{R}$ を squashing 関数とする．このとき，$\Sigma\Pi^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である．

**証明.** 連続な squashing 関数 $F$ に対して，$\Sigma\Pi^r(\Psi)$ が $\Sigma\Pi^r(F)$ において広義一様収束の意味で稠密であることを示せば十分である．なぜなら，それが示されれば定理 4.4.3 により $\Sigma\Pi^r(F)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるので，$\Sigma\Pi^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密となる．

さて，このことを示すためには，$\prod_{k=1}^l F \circ A_k$ の形の関数を $\Sigma\Pi^r(\Psi)$ の元で一様近似できることを示せばよい．そこで，$\varepsilon > 0$ を任意に取る．積 $\prod_{k=1}^l : [0,1]^l \ni (a_1, \ldots, a_l) \mapsto \prod_{k=1}^l a_l \in \mathbb{R}$ は連続で $[0,1]^l$ はコンパクトなので，ある $\delta$ が存在して任意の $a_k, b_k \in [0,1]$ に対して，$|a_k - b_k| < \delta \quad (k=1,\ldots,l)$ ならば $|\prod_{k=1}^l a_k - \prod_{k=1}^l b_k| < \varepsilon$ となる．一方，補題 4.4.4 により，ある $\beta_j \in \mathbb{R}$ と $A_j^1 \in \mathbb{A}^1$ が存在して，$H=\sum_{j=1}^m \beta_j \Psi \circ A_j^1$ について

$$\sup_{x\in\mathbb{R}}|F(x)-H(x)| < \delta$$

が成り立つ．したがって，

$$\sup_{x\in\mathbb{R}^r}\left|\prod_{k=1}^l F(A_k(x))-\prod_{k=1}^l H(A_k(x))\right| \le \varepsilon$$

となる．いま，$A_j^1 \circ A_k \in \mathbb{A}^r$ であり，$\Sigma\Pi(\Psi)$ は積と和に関して閉じているので，$\prod_{k=1}^l H \circ A_k \in \Sigma\Pi(\Psi)$ である．これで証明は完了した． □

次に本題の定理を示していく.

**定理 4.4.6 (Hornik 1989 [12]).**
$\Psi : \mathbb{R} \to \mathbb{R}$ を squashing 関数とする. このとき, $\Sigma^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

この定理を証明するために 2 つ補題を示す.

**補題 4.4.7.** $\Psi : \mathbb{R} \to \mathbb{R}$ を squashing 関数とする. このとき, 任意の $\varepsilon, M > 0$ に対して, $F \in \Sigma^1(\Psi)$ が存在して,

$$\sup_{x \in [-M,M]} |F(x) - \cos(x)| < \varepsilon$$

が成立する. 便宜的にこのような $F$ を任意に一つ取り $\cos_{M,\varepsilon}$ で表すことにする.

**証明.** $n$ を $M \leq (2n - 1/2)\pi$ となるように取る. このとき, $[-M, M] \subset [-(2n + 1/2)\pi, (2n - 1/2)\pi]$ である. いま, 連続な squashing 関数 $g, h : \mathbb{R} \to \mathbb{R}$ を

$$g(x) = \begin{cases} 0 & (x \leq -\pi/2) \\ \cos(x) & (-\pi/2 \leq x \leq 0) \\ 1 & (0 \leq x) \end{cases}$$

および,

$$h(x) = \begin{cases} 0 & (x \leq -\pi) \\ 1 + \cos(x) & (-\pi \leq x \leq -\pi/2) \\ 1 & (-\pi/2 \leq x) \end{cases}$$

で定義する. そして $g_\alpha(x) = g(x - \alpha\pi) \quad (\alpha \in \mathbb{R})$ と定義する. $h_\alpha$ も同様に定義する. すると, $g_\alpha, h_\alpha$ は連続な squashing 関数なので補題 4.4.4 より, 任意の $\varepsilon > 0$ に対して $G_\alpha, H_\alpha \in \Sigma^1(\Psi)$ が存在して,

$$\sup_{x \in \mathbb{R}} |g_\alpha(x) - G_\alpha(x)| < \varepsilon$$
$$\sup_{x \in \mathbb{R}} |h_\alpha(x) - H_\alpha(x)| < \varepsilon$$

となる. 一方, $\tilde{h}_\alpha := h_\alpha - 1$ とおくと, 任意の $x \in [-(2n + 1/2)\pi, (2n - 1/2)\pi]$ に対して,

$$\cos(x) = \sum_{j=0}^{4n-1} \Big( (-1)^j g_{-2n+j}(x) + (-1)^{j+1} \tilde{h}_{-2n+1+j}(x) \Big)$$

である. また, $1 \geq \Psi(K) > 1 - \varepsilon$ なる $K$ を取り, $\tilde{H}_\alpha := H_\alpha - \Psi(K)$ とおくと, $\Sigma^1(\Psi)$ は和に関して閉じているので $\tilde{H}_\alpha \in \Sigma^1(\Psi)$ であり, $\sup_{x \in \mathbb{R}} |\tilde{H}_\alpha(x) - \tilde{h}_\alpha| < 2\varepsilon$ となる. よって,

$$F = \sum_{j=0}^{4n-1} \Big( (-1)^j G_{-2n+j}(x) + (-1)^{j+1} \tilde{H}_{-2n+1+j}(x) \Big) \quad \in \Sigma^1(\Psi)$$

が求めるものである．実際,

$$\begin{aligned}\sup_{x\in[-M,M]}|F(x)-\cos(x)| &\le \sum_{j=0}^{4n-1}\sup_{x\in[-M,M]}|G_{-2n+j}(x)-g_{-2n+j}(x)| \\ &\quad+\sum_{j=0}^{4n-1}\sup_{x\in[-M,M]}|\tilde{H}_{-2n+1+j}(x)-\tilde{h}_{-2n+1+j}(x)| \\ &\le 4n\varepsilon+8n\varepsilon=12n\varepsilon.\end{aligned}$$

□

**補題 4.4.8.** $g=\sum_{j=1}^{Q}\beta_j\cos\circ A_j$, $A_j\in\mathbb{A}^r$ とする．また, $\Psi:\mathbb{R}\to\mathbb{R}$ を squashing 関数とする．このとき, 任意の $\varepsilon>0$ と任意のコンパクト集合 $C\subset\mathbb{R}^r$ に対して, $f\in\Sigma^r(\Psi)$ が存在して, $\sup_{x\in C}|f(x)-g(x)|<\varepsilon$ となる．

**証明.** $C$ がコンパクトで各 $A_j$ は連続なので $A_j(C)\subset\mathbb{R}$ はコンパクトである．したがって, $M>0$ を適当に取れば $A_j(K)\subset[-M,M]$ $(j=1,\ldots,Q)$ となる．そこで $K=Q\sum_{j=1}^{Q}|\beta_j|$ とおくと,

$$\sup_{x\in C}\left|\sum_{j=1}^{Q}\beta_j\cos_{M,\varepsilon/K}(A_j(x))-g(x)\right|<\varepsilon$$

となる．そして, $\cos_{M,\varepsilon/K}\in\Sigma^1(\Psi)$ なので, $\cos_{M,\varepsilon/K}\circ A_j\in\Sigma^r(\Psi)$ となる．よって, $\sum_{j=1}^{Q}\beta_j\cos_{M,\varepsilon/K}\circ A_j\in\Sigma^r(\Psi)$ であるので, これで補題の成立が確かめられた． □

**定理 4.4.6 の証明..** $\cos:\mathbb{R}\to\mathbb{R}$ は定数でない連続関数なので, 定理 4.4.3 により,

$$\left\{\sum_{j=1}^{Q}\beta_j\prod_{k=1}^{l_j}\cos\circ A_{j,k}\mid Q,l_j\in\mathbb{N},\beta_j\in\mathbb{R},A_{j,k}\in\mathbb{A}^r\right\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である．ところが, 任意の $a,b\in\mathbb{R}$ に対して,

$$\cos(a)\cos(b)=\frac{\cos(a+b)-\cos(a-b)}{2}$$

であり, $\mathbb{A}^r$ は加法とスカラー倍に関して閉じているので, ある $\alpha_j\in\mathbb{R},A_j\in\mathbb{A}^r$ が存在して,

$$\sum_{j=1}^{Q}\beta_j\prod_{k=1}^{l_j}\cos\circ A_{j,k}=\sum_{j}\alpha_j\cos\circ A_j$$

となる．よって, 補題 4.4.8 より, $\Sigma^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である． □

# 5　$C(\mathbb{R}^r)$ における万能近似定理の応用

## 5.1　多出力・多層への拡張

前章で述べた結果はすべて, スカラーを出力するネットワークに関するものであったが, これらはベクトル値を出力するネットワークについても成り立つ. 本節ではこのことを正確に述べよう.

**定義 5.1.1.** 関数 $G : \mathbb{R} \to \mathbb{R}$ に対して,

$$\Sigma^{r,s}(G) := \{f : \mathbb{R}^r \to \mathbb{R}^s \mid f = \sum_j^q \beta_j G \circ A_j,\ \beta_j \in \mathbb{R}^s, A_j \in \mathbb{A}^r, q \in \mathbb{N}\}$$

$$\Sigma\Pi^{r,s}(G) := \left\{ f : \mathbb{R}^r \to \mathbb{R}^s \;\middle|\; \begin{array}{l} f = \displaystyle\sum_j^q \beta_j \prod_{k=1}^{l_j} G \circ A_{j,k}\ , \beta_j \in \mathbb{R}^s \\ A_{j,k} \in \mathbb{A}^r, l_j \in \mathbb{N}, q \in \mathbb{N} \end{array} \right\}$$

と定義する.

**定理 5.1.2.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r, \mathbb{R})$ において広義一様収束の意味で稠密になるものとする. このとき, $\Sigma^{r,s}(\Psi)$ は $C(\mathbb{R}^r, \mathbb{R}^s)$ において広義一様収束の意味で稠密である. すなわち, 任意に空でないコンパクト集合 $K \subset \mathbb{R}^r$ を取るとき, 任意の $g \in C(\mathbb{R}^r, \mathbb{R}^s)$ と $\varepsilon > 0$ に対して, $f \in \Sigma^{r,s}(\Psi)$ が存在して,

$$\sum_{j=1}^s \sup_{x \in K} |f_j(x) - g_j(x)| < \varepsilon$$

となる. ここで, $f = (f_1, \ldots, f_s)^{\mathrm{T}}, g = (g_1, \ldots, g_s)^{\mathrm{T}}$ である.

**証明.** 各 $g_i \in C(\mathbb{R}^r, \mathbb{R})$ を $\Sigma^r(\Psi)$ の元

$$f_i = \sum_{j=1}^{Q_i} \beta_{i,j} \Psi \circ A_{i,j}$$

で近似する. すると,

$$f = \sum_{i=1}^s \sum_{j=1}^{Q_i} \beta_{i,j} (\Psi \circ A_{i,j}) \mathbf{e}_i$$

が $g$ の近似を与える. ここに $\mathbf{e}_i$ は第 $i$ 成分が 1 で他の成分がすべて 0 である $\mathbb{R}^s$ の元である. □

さて, これまで述べてきた結果はすべて中間層が 1 つののみからなるネットワークに関するものであった. そこで次に, 多層かつ多出力のネットワークの万能近似定理について述べよう. ニューラルネットワークの層数を $l \geq 2$ で表すことにする. ただし, ここでは入力層は層数に含めないことにする. つまり, これまで扱っていたのは $l = 2$ の場合ということである.

**定義 5.1.3 ($\Sigma_l^{r,s}$ ネットワーク).**
関数 $\Psi : \mathbb{R} \to \mathbb{R}$ に対して, 次のように集合 $I_1^{(q)}, \ldots, I_{l-1}^{(q)}, I_l$ $(q \in \mathbb{N})$ を帰納的に定義する.

$$\begin{aligned} I_1^{(q)} &= \{(\Psi \circ A_1, \ldots, \Psi \circ A_q)^{\mathrm{T}} \mid A_1 \ldots, A_q \in \mathbb{A}^r\}, \\ I_{k+1}^{(q)} &= \{(\Psi \circ A_1, \ldots, \Psi \circ A_q)^{\mathrm{T}} \circ f \mid q' \in \mathbb{N}, f \in I_k^{(q')}, A_j \in \mathbb{A}^{(q')}\}, \\ I_l &= \{(A_1, \ldots, A_s)^{\mathrm{T}} \circ f \mid q' \in \mathbb{N}, f \in I_{l-1}^{(q')}, A \in \mathbb{A}^{(q')}\}. \end{aligned}$$

このとき, $I_l$ を $\Psi$ を活性化関数とする $\Sigma_l^{r,s}$ ネットワークと呼び, $\Sigma_l^{r,s}(\Psi)$ で表す.

より具体的に書けば $\Sigma_l^{r,s}(\Psi)$ の元は次のような $\mathbb{R}^r$ から $\mathbb{R}^s$ への関数である: $x \in \mathbb{R}^r$ に対して, $a_0 = x$ とおく. そして, $k = 1, \ldots, l$ に対して, $q_0 = r, q_l = s, q_j \in \mathbb{N}$ とし, $A_{k,i} \in \mathbb{A}^{q_{k-1}}$ $(i = 1, \ldots, q_k)$ とする. また, $G_1, \ldots, G_{l-1} = \Psi$ とし, $G_l = \mathrm{id}$ とする. このとき, 次の式

$$a_{k,i} = G_k(A_{k,i}(a_{k-1})) \quad (i = 1, \ldots, q_k, k = 1, \ldots, l)$$

にしたがって計算を繰り返した結果 $a_l = (a_{l,1}, \ldots, a_{l,s})^{\mathrm{T}} \in \mathbb{R}^s$ を出力する.

この定義から $\Sigma^{r,s}(\Psi) \subset \Sigma_2^{r,s}(\Psi)$ がわかる. 実際, $f = \sum_{j=1}^{Q} \beta_j \Psi \circ A_j$ $(\beta_j \in \mathbb{R}^s, A_j \in A^r)$ とするとき, 各 $k = 1, \ldots, Q$ に対して $B_k \in \mathbb{A}^Q$ を $B_k(x) = \sum_{j=1}^{Q} \beta_j^{(k)} x_j$ と定義すれば, $f = (B_1, \ldots, B_Q)^{\mathrm{T}} \circ (\Psi \circ A_1, \ldots, \Psi \circ A_Q)^{\mathrm{T}}$ となる. このことから, $\Sigma^{r,s}(\Psi)$ が万能近似能力を持つとき, $\Sigma_2^{r,s}(\Psi)$ も万能近似能力を持つことがわかる. より一般に次が成り立つ.

**定理 5.1.4.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r, \mathbb{R})$ において広義一様収束の意味で稠密になるものとする. このとき, $\Sigma_l^{r,s}(\Psi)$ は $C(\mathbb{R}^r, \mathbb{R}^s)$ において広義一様収束の意味で稠密である.

この定理を示すためにひとつ補題を証明しよう.

**補題 5.1.5.** $\mathcal{F}$ は「$\mathbb{R}$ から $\mathbb{R}$ への関数全体の集合」の部分集合で, $C(\mathbb{R})$ において広義一様収束の意味で稠密であるとし, $\mathcal{G}$ は「$\mathbb{R}^r$ から $\mathbb{R}$ への関数全体の集合」の部分集合で, $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるとする. このとき,

$$F \circ G = \{f \circ g \mid f \in \mathcal{F}, g \in \mathcal{G}\}$$

は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である.

**証明.** 空でないコンパクト集合 $K \subset \mathbb{R}^r$ を任意に取る. $h \in C(\mathbb{R}^r)$ とし, $\varepsilon > 0$ を任意に取ると, 仮定より $g \in \mathcal{G}$ が存在して, $\sup_{x \in K} |g(x) - h(x)| < \varepsilon$ となる. このとき, $h(K) \subset \mathbb{R}$ はコンパクト, 特に有界であるので, $g(K)$ も有界である. そこで, $g(K)$ の閉包 $S := \mathrm{cl}(g(K))$ はコンパクトである. すると, 仮定より連続関数

$\mathbb{R} \ni s \mapsto s \in \mathbb{R}$ に対して, $f \in C(\mathbb{R})$ が存在して $\sup_{s\in S}|f(s)-s| < \varepsilon$ となる. ゆえに, 任意の $x \in K$ に対して,

$$\begin{aligned}|f(g(x)) - h(x)| &\le |f(g(x)) - g(x)| + |g(x) - h(x)| \\ &\le \sup_{s\in S}|f(s) - s| + \sup_{z\in K}|g(z) - h(z)| \\ &< \varepsilon + \varepsilon = 2\varepsilon\end{aligned}$$

である. よって, $\sup_{x\in K}|f(g(x)) - h(x)| \le 2\varepsilon$ となる. □

**定理 5.1.4 の証明..** $s = 1$ の場合を証明すれば十分である. 次のように集合 $J_1, \ldots, J_l$ を帰納的に定義する.

$$J_1 = \{\sum_j^Q \beta_j \Psi \circ A_j \mid Q \in \mathbb{N}, A_j \in \mathbb{A}^r, \beta_j \in \mathbb{R}\},$$

$$J_{k+1} = \{\sum_j^Q \beta_j \Psi \circ A_j \circ f \mid Q \in \mathbb{N}, f \in J_k, A_j \in \mathbb{A}^1, \beta_j \in \mathbb{R}\},$$

このとき, $x \mapsto \sum_j \beta_j x_j$ は線形写像なので $J_l \subset \Sigma_{l+1}^{r,s}(\Psi)$ である. そこで, $J_l$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることを示せばよい. このことを $l$ に関する帰納法で示そう. まず $\Sigma^r(\Psi) = J_1$ であるので $J_1$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である. 次に, $J_k$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるとすると, $J_{k+1} = \Sigma^1(\Psi) \circ J_k$ であり, $\Sigma^1(\Psi)$ は $C(\mathbb{R})$ において広義一様収束の意味で稠密であるので, 前補題より $J_{k+1}$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密である. □

## 5.2 $L^p$ 空間における万能近似定理

本節では, 前章の結果を使って $\mathbb{R}^r$ 上の Borel 可測関数全体の空間 $\mathbb{M}^r := \mathbb{M}(\mathbb{R}^r, \mathbb{R})$ や $L^p$ 空間における $\Sigma^r(\Psi)$ の万能近似定理を述べる. まず, $C(\mathbb{R}^r)$ が $\mathbb{M}^r$ において稠密であることについて述べる.

**定義 5.2.1.** $\mu$ を $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度とする. このとき, $f, g \in \mathbb{M}^r$ が $\mu$-同値であるとは, $\mu(\{f \ne g\}) = 0$ となることをいう. 以後, この同値関係による $\mathbb{M}^r$ の商集合をそのまま $\mathbb{M}^r$ で表す場合がある.

**定義 5.2.2.** $\mu$ を $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度とする. このとき, $\rho_\mu : \mathbb{M}^r \times \mathbb{M}^r \to [0, \infty)$ を,

$$\rho_\mu(f, g) := \inf\{\varepsilon > 0 \mid \mu(|f - g| > \varepsilon) < \varepsilon\}$$

により定義する (有限測度なので $\rho_\mu(f, g) \le \mu(\mathbb{R}^r) < \infty$ となることに注意).

**補題 5.2.3.** $\mu$ を $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度とする．また, $(f_n)_{n=1}^{\infty}$ を $\mathbb{M}^r$ の元の列とし, $f \in \mathbb{M}^r$ とする．このとき, 次の条件は同値である．

(1): $\rho_\mu(f_n, f) \to 0$.

(2): $\forall \varepsilon > 0,\ \mu(\{|f_n - f| > \varepsilon\}) \to 0$.

(3): $\int_{\mathbb{R}^r} \min(|f_n - f|, 1) d\mu \to 0$.

**証明.**
(1) $\Rightarrow$ (2): 任意に $\varepsilon > 0$ を取る．いま, 仮定よりある $N \in \mathbb{N}$ が存在して任意の $n \geq N$ に対して $\rho_\mu(f_n, f) < \varepsilon$ となる．すると, 任意の $n \geq N$ に対して, $\rho_\mu$ の定義より $\mu(|f_n - f| > \varepsilon_n) < \varepsilon_n$ なる $0 < \varepsilon_n < \varepsilon$ が取れる．ゆえに, 任意の $n \geq N$ に対して, $\mu(|f_n - f| > \varepsilon) \leq \mu(|f_n - f| > \varepsilon_n) < \varepsilon_n < \varepsilon$ となる．したがって, 任意に $\delta > 0$ を取ると, ある $M \in \mathbb{N}$ が存在して任意の $n \geq M$ に対して $\mu(|f_n - f| > \delta) < \delta$ となる．そこで, $\varepsilon \leq \delta$ の場合は, 任意の $n \geq N$ に対して, $\mu(|f_n - f| > \varepsilon) < \varepsilon \leq \delta$ となる．$\varepsilon > \delta$ の場合は, 任意の $n \geq M$ に対して $\mu(|f_n - f| > \varepsilon) \leq \mu(|f_n - f| > \delta) < \delta$ となる．これは $\mu(\{|f_n - f| > \varepsilon\}) \to 0$ を意味する．
(2) $\Rightarrow$ (1): 任意に $\varepsilon > 0$ を取る．いま, 仮定よりある $N \in \mathbb{N}$ が存在して任意の $n \geq N$ に対して $\mu(|f_n - f| > \varepsilon) < \varepsilon$ となる．ゆえに, 任意の $n \geq N$ に対して $\rho_\mu(f_n, f) \leq \varepsilon$ となる．
(2) $\Rightarrow$ (3): $\varepsilon > 0$ を任意に取る．このとき, 仮定よりある $N \in \mathbb{N}$ が存在して任意の $n \geq N$ に対して $\mu(|f_n - f| > \varepsilon) < \varepsilon$ となる．そこで, 任意の $n \geq N$ に対して,

$$\begin{aligned}\int_{\mathbb{R}^r} \min(|f_n - f|, 1) d\mu &\leq \int_{|f_n - f| > \varepsilon} 1 d\mu + \int_{|f_n - f| \leq \varepsilon} |f_n - f| d\mu \\ &\leq \varepsilon + \varepsilon \mu(\mathbb{R}^r)\end{aligned}$$

となる．$\mu$ は有限測度なのでこれで (3) の成立が確かめられた．
(3) $\Rightarrow$ (2): 任意に $1 > \varepsilon > 0$ を取る．このとき, 任意の $\delta > 0$ に対して, ある $N \in \mathbb{N}$ が存在して任意の $n \geq N$ に対して, $\int_{\mathbb{R}^r} \min(|f_n - f|, 1) d\mu < \varepsilon\delta$ となる．そこで, 任意の $n \geq N$ に対して,

$$\begin{aligned}\mu(|f_n - f| > \varepsilon) &\leq \frac{1}{\varepsilon} \int_{|f_n - f| > \varepsilon} \min(|f_n - f|, 1) d\mu \\ &\leq \frac{1}{\varepsilon} \int_{\mathbb{R}^r} \min(|f_n - f|, 1) d\mu \\ &\leq \delta\end{aligned}$$

となる．よって, $\mu(\{|f_n - f| > \varepsilon\}) \to 0$ である．$\varepsilon \geq 1$ の場合も $0 < \varepsilon' < 1$ に対して,

$$\mu(|f_n - f| > \varepsilon) \leq \mu(|f_n - f| > \varepsilon') \to 0.$$

となるので成立する． □

**補題 5.2.4.** $\mu$ を $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度とするとき, $C(\mathbb{R}^r)$ は $\mathbb{M}^r$ において $\rho_\mu$-稠密である.

**証明.** $f \in \mathbb{M}^r$ とし, $\varepsilon > 0$ を任意に取る. 優収束定理により, 十分大きな $M \geq 0$ に対して,

$$\int_{\mathbb{R}^r} \min(|f \cdot \chi_{\{|f| \geq M\}} - f|, 1) d\mu < \varepsilon$$

が成立する. ここで, $f \cdot \chi_{\{|f| \geq M\}} \in L^1(\mathbb{R}^r, \mathcal{B}^r, \mu)$ であるから, $C_0(\mathbb{R}^r)$ が $L^1(\mathbb{R}^r, \mathcal{B}^r, \mu)$ において $L^1$ ノルムの意味で稠密であること (付録の補題 9.9.4) より, ある $g \in C(\mathbb{R}^r)$ で $\int_{\mathbb{R}^r} |f \cdot \chi_{\{|f| \geq M\}} - g| d\mu < \varepsilon$ となる. よって,

$$\begin{aligned}
\int_{\mathbb{R}^r} \min(|f - g|, 1) d\mu &\leq \int_{R^r} \min(|f - f \cdot \chi_{\{|f| \geq M\}}|, 1) d\mu \\
&\quad + \int_{R^r} \min(|f \cdot \chi_{\{|f| \geq M\}} - g|, 1) d\mu \\
&\leq \varepsilon + \int_{R^r} |f \cdot \chi_{\{|f| \geq M\}} - g| d\mu \\
&\leq \varepsilon + \varepsilon = 2\varepsilon
\end{aligned}$$

となる. これより任意の $\varepsilon > 0$ に対して $g \in C(\mathbb{R}^r)$ が存在して $\rho_\mu(f, g) < \varepsilon$ となることがわかる (補題 5.2.3 の (3) $\Rightarrow$ (2) の証明参照). □

次に $\Sigma^r(\Psi)$ が $\mathbb{M}^r$ において稠密であることの証明に使う補題を示す.

**補題 5.2.5.** $(f_n)_{n=1}^\infty$ を $\mathbb{M}^r$ の元の列とし, $f \in \mathbb{M}^r$ とする. このとき, $(f_n)_{n=1}^\infty$ が $f$ に広義一様収束するならば, 任意の $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度 $\mu$ に対して, $\rho_\mu(f_n, f) \to 0$ となる.

**証明.** 任意に $\varepsilon > 0$ を取る. $\mu$ は有限測度なので付録の系 9.8.7 より正則である. したがって, あるコンパクト集合 $K \in \mathbb{R}^r$ が存在して $\mu(\mathbb{R}^r \setminus K) < \varepsilon$ となる. そこで仮定より, ある $N \in \mathbb{N}$ が存在して, 任意の $n \geq N$ に対して,

$$\sup_{x \in K} |f_n(x) - f(x)| < \varepsilon$$

となる. ゆえに, 任意の $n \geq N$ に対して,

$$\begin{aligned}
\int_{\mathbb{R}^r} \min(|f_n - f|, 1) d\mu &= \int_K \min(|f_n - f|, 1) d\mu + \int_{\mathbb{R}^r \setminus K} \min(|f_n - f|, 1) d\mu \\
&\leq \int_K |f_n - f| d\mu + \int_{\mathbb{R}^r \setminus K} 1 d\mu \\
&\leq \varepsilon \mu(\mathbb{R}^r) + \varepsilon
\end{aligned}$$

となる. よって, $\int_{\mathbb{R}^r} \min(|f_n - f|, 1) d\mu \to 0$ なので, 補題 5.2.3 より $\rho_\mu(f_n, f) \to 0$ である. □

**定理 5.2.6.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密になるものとする．このとき, 任意の $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度 $\mu$ に対して, $\Sigma^r(\Psi)$ は $\mathbb{M}^r$ において $\rho_\mu$-稠密である.

**証明.** $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることと補題 5.2.5 より $\Sigma^r(\Psi)$ は $C(\mathbb{R}^r)$ において $\rho_\mu$-稠密である．このことと 補題 5.2.4 を合わせれば $\Sigma^r(\Psi)$ は $\mathbb{M}^r$ において $\rho_\mu$-稠密であることがわかる. □

次の定理は $\mathbb{M}^r$ を sup ノルムの意味で $\Sigma^r(\Psi)$ により十分よく近似できることを示すものである.

**定理 5.2.7.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密になるものとする．また $\mu$ を $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度とする．このとき, 任意の $g \in \mathbb{M}^r$ と $\varepsilon > 0$ に対して, $f \in \Sigma^r(\Psi)$ とコンパクト集合 $K \subset \mathbb{R}^r$ が存在して, $\mu(K) > \mu(\mathbb{R}^r) - \varepsilon$ かつ $\forall x \in K, |f(x) - g(x)| < \varepsilon$ となる.

証明には次の事実を使う (宮島 [36] の第 7 章 2 節などを参照されたい).

**定理 5.2.8 (Tietze の拡張定理).**
$X$ は Hausdorff 空間で, 任意の交わらない閉集合 $F_1, F_2 \subset X$ に対して $F_1 \subset U_1, F_2 \subset U_2$ かつ $U_1 \cap U_2 = \emptyset$ となる開集合 $U_1, U_2 \subset X$ が存在するとする．このとき, 閉集合 $A \subset X$ 上の任意の実数値連続関数 $f$ に対して $X$ 上の実数値連続関数 $f'$ が存在して $f = f'|_A$ が成立する．つまり, $f$ を $X$ 上の実数値連続関数に拡張できる.

**定理 5.2.7 の証明.** Lusin の定理 (付録の命題 9.10.2) より, コンパクト集合 $K \subset \mathbb{R}^r$ が存在して, $\mu(K) > \mu(\mathbb{R}^r) - \varepsilon$ かつ $K$ 上の関数 $g|K$ は連続となる．そこで Tietze の拡張定理より, $g' \in C(\mathbb{R}^r)$ が存在して $g|K = g'|K$ となる．いま $\Sigma^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるので, $f \in \Sigma^r(\Psi)$ が存在して,

$$\sup_{x \in K} |f(x) - g(x)| = \sup_{x \in K} |f(x) - g'(x)| < \varepsilon.$$

□

さて, $L^p$ 空間における万能近似定理に話を移そう.

**定義 5.2.9.** $f, g \in L^p(\mathbb{R}, \mathcal{B}^r, \mu)$ に対して,

$$d_\mu^p(f, g) = \left( \int_{\mathbb{R}^r} |f - g|^p d\mu \right)^{1/p}$$

と定義する．$d_\mu^p$ は $L^p(\mathbb{R}^r, \mathcal{B}^r, \mu)$ 上の距離である.

**定理 5.2.10.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密になるものとする．また $\mu$ を $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度とする．このとき, $\mu(K) = \mu(\mathbb{R}^r)$ なるコンパクト集合 $K \subset \mathbb{R}^r$ が存在すれば, 任意の $p \in [1, \infty)$ について, $\Sigma^r(\Psi)$ は $L^p(\mathbb{R}^r, \mathcal{B}^r, \mu)$ において $d_\mu^p$-稠密である.

**証明.** $g \in L^p(\mathbb{R}^r, \mathcal{B}^r, \mu)$ と $\varepsilon > 0$ を任意に取る．付録の補題 9.9.4 より $C_0(\mathbb{R}^r)$ は $L^p(\mathbb{R}^r, \mathcal{B}^r, \mu)$ において $d_\mu^p$-稠密であるので，$h \in C(\mathbb{R}^r)$ で $d_\mu^p(g,h) < \varepsilon$ なるものが取れる．いま，$\Sigma^r(\Psi)$ は $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であるので，$f \in \Sigma^r(\Psi)$ が存在して $\sup_{x\in K} |h(x) - f(x)| < \varepsilon$ となる．よって，$\mu(K) = \mu(\mathbb{R}^r)$ であることに注意すると，

$$d_\mu^p(g,f) \leq d_\mu^p(g,h) + d_\mu^p(h,f) \leq \varepsilon + (\int_K \varepsilon^p d\mu)^{1/p} \leq \varepsilon + \varepsilon(\mu(\mathbb{R}^r))^{1/p}$$

となる． □

**系 5.2.11.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密になるものとする．また，$K \subset \mathbb{R}^r$ をコンパクト集合とし $\mu$ を $(K, \mathcal{B}_K)$ 上の有限測度とする．このとき，任意の $p \in [1,\infty)$ について，$\Sigma^r(\Psi)$(を $K$ に制限したもの) は $L^p(K, \mathcal{B}_K, \mu)$ において $d_\mu^p$-稠密である．ただし，$\mathcal{B}_K$ は $K$ 上の Borel $\sigma$-加法族であり，$\mathcal{B}_K = \{K \cap A \mid A \in \mathcal{B}^r\}$ を満たす．

**証明.** $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度 $\tilde{\mu}$ を $\tilde{\mu}(A) = \mu(A \cap K)$ で定め，$f \in L^p(K, \mathcal{B}_K, \mu)$ に対し $\tilde{f} \in L^o(\mathbb{R}^r, \mathcal{B}^r, \tilde{\mu})$ を

$$\tilde{f}(x) = \begin{cases} f(x) & (x \in K) \\ 0 & (\text{otherwise}) \end{cases}$$

で定める．このとき，すぐ上の定理より，任意の $\varepsilon > 0$ に対して．$g \in \Sigma^r(\Psi)$ が存在して，

$$\varepsilon > d_{\tilde{\mu}}^p(\tilde{f}, g) = \left(\int_{\mathbb{R}^r} |\tilde{f} - g|^p d_{\tilde{\mu}}\right)^{1/p} = \left(\int_K |f - g|^p d\mu\right)^{1/p} = d_\mu^p(f, g|_K)$$

となる． □

以下は Leshno et al. の結果の系である．

**系 5.2.12.** $\mu$ を $(\mathbb{R}^r, \mathcal{B}_{\mathbb{R}^r})$ 上の有限測度で，$\mu(K) = \mu(\mathbb{R}^r)$ なるコンパクト集合 $K \subset \mathbb{R}^r$ が存在するとする．このとき，$\Psi \in \mathcal{M}(\mathbb{R})$ に対して，$\Psi$ がどんな多項式関数とも a.e. で一致することがないことと $\Sigma^r(\Psi)$ が $L^p(\mathbb{R}^r, \mathcal{B}_{\mathbb{R}^r}, \mu)$ $(p \geq 1)$ で稠密であることは同値である．

**証明.** $\Psi$ がある $m$ 次多項式関数と *a.e.* で一致するとすると，$\Sigma^r(\Psi)$ は $r$ 変数の $m$ 次以下の多項式関数と a.e. で一致する関数全体の集合になるので $L^p(\mathbb{R}^r, \mathcal{B}_{\mathbb{R}^r}, \mu)$ において $m+1$ 次の多項式関数を近似することはできない ($\mu(K) = \mu(\mathbb{R}^r)$ ゆえ多項式関数は $L^p(\mu)$ に属することに注意)．逆に $\Psi$ がどんな多項式関数とも a.e. で一致

することがないとし, $f \in L^p(\mathbb{R}^r, \mathcal{B}_{\mathbb{R}^r}, \mu)$ をとる. いま, 付録の命題 9.9.4 より $C(K)$ は $L^p(K, \mu)$ で稠密であるので, $g \in C(K)$ が存在して,

$$\|f - g\|_{L^p}^p = \int_{\mathbb{R}^r} |f - g|^p d\mu = \int_K |f - g| d\mu < (\varepsilon/2)^p$$

となる. 定理 4.1.2 より $\Sigma^r(\Psi)$ は $C(K)$ において稠密であるので, $h \in \Sigma^r(\Psi))$ が存在して,

$$\sup_{x \in K} |g(x) - h(x)| < \varepsilon/\mu(K)^{1/p}$$

となる. よって,

$$\|f - h\|_{L^p} \leq \|f - g\|_{L^p} + \|g - h\|_{L^p} \leq \varepsilon.$$

□

以上の結果は多層・多出力の場合も成立する. 証明は $C(\mathbb{R}^r, \mathbb{R}^s)$ の場合と同様にできる.

**定義 5.2.13.** 次のように定める.

$$M^{r,s} = M(\mathbb{R}^r, \mathbb{R}^s) := \{f : \mathbb{R}^r \to \mathbb{R}^s \mid f \text{ は } (\mathcal{B}^r, \mathcal{B}^s) \text{ 可測 }\}$$
$$L^p(r, s, \mu) := \{f = (f_1, \ldots, f_s) \in M^{r,s} \mid \int_{\mathbb{R}^r} |f_j|^p d\mu < \infty \quad (j = 1, \ldots, s)\}$$

**定理 5.2.14.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r, \mathbb{R})$ において広義一様収束の意味で稠密になるものとする. このとき, $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度 $\mu$ に対して, $\Sigma^{r,s}(\Psi)$ は $\mathbb{M}^{r,s} := \mathbb{M}(\mathbb{R}^r, \mathbb{R}^s)$ において $\rho_\mu^s$-稠密である. ただし, $f, g \in \mathbb{M}^{r,s}$ に対して,

$$\rho_\mu^s(f, g) = \sum_{j=1}^s \inf\{\varepsilon > 0 \mid \mu(|f_j - g_j| > \varepsilon) < \varepsilon\}$$

である. さらに, $\mu(K) = \mu(\mathbb{R}^r)$ なるコンパクト集合 $K \subset \mathbb{R}^r$ が存在すれば, 任意の $p \geq 1$ に対して, $\Sigma^{r,s}(\Psi)$ は $L^p(r, s, \mu)$ において $d_\mu^{p,s}$-稠密である. ただし, $f, g \in L^p(r, s, \mu)$ に対して,

$$d_\mu^{p,s}(f, g) = \sum_{j=1}^s \left( \int_{\mathbb{R}^r} |f_j - g_j|^p d\mu \right)^{1/p}$$

である.

**定理 5.2.15.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r, \mathbb{R})$ において広義一様収束の意味で稠密になるものとする. このとき, 任意の $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度 $\mu$ に対して, $\Sigma_l^{r,s}(\Psi)$ は $\mathbb{M}(\mathbb{R}^r, \mathbb{R}^s)$ において $\rho_\mu^s$-稠密である. さらに, $\mu(K) = \mu(\mathbb{R}^r)$ なるコンパクト集合 $K \subset \mathbb{R}^r$ が存在すれば, 任意の $p \geq 1$ に対して, $\Sigma_l^{r,s}(\Psi)$ は $L^p(r, s, \mu)$ において $d_\mu^{p,s}$-稠密である.

## 5.3 ニューラルネットワークによる補間

本節では feedforward 型ニューラルネットワークによる補間について述べる. つまり, 与えられた有限個の点を (近似的または真に) 通るような feedforward 型ニューラルネットワークの存在について述べる.

**定理 5.3.1.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r, \mathbb{R})$ において広義一様収束の意味で稠密になるものとする. また $(\mathbb{R}^r, \mathcal{B}^r)$ 上の有限測度 $\mu$ は有限個の点の上でのみ値を取るものとする. つまり, 相異なる $x_1, \ldots, x_n \in \mathbb{R}^r$ が存在して,

$$\sum_{j=1}^{n} \mu(\{x_j\}) = \mu(\mathbb{R}^r), \quad \mu(\{x_j\}) > 0 \quad (j = 1, \ldots, n)$$

となるとする. このとき, 任意の $g \in \mathbb{M}^r$ と $\varepsilon > 0$ に対して, $f \in \Sigma^r(\Psi)$ が存在して, $\mu(|f - g| > \varepsilon) = 0$ となる.

**証明.** $\tilde{\varepsilon} := \min_{1 \le j \le n} \mu(\{x_j\})$ とおく. $\varepsilon < \tilde{\varepsilon}$ としてよい. 定理 5.2.6 より, $\Sigma^r(\Psi)$ は $\mathbb{M}^r$ において $\rho_\mu$-稠密であるので, $f \in \Sigma^r(\Psi)$ が存在して,

$$\rho_\mu(f, g) = \inf\{\varepsilon' > 0 \mid \mu(|f - g| > \varepsilon') < \varepsilon'\} < \varepsilon$$

となる. このとき, ある $\varepsilon' > 0$ で, $\mu(|f - g| > \varepsilon') < \varepsilon'$ かつ $\varepsilon' < \varepsilon$ なるものが取れる. すると,

$$\mu(|f - g| > \varepsilon) \le \mu(|f - g| > \varepsilon') < \varepsilon' < \varepsilon < \tilde{\varepsilon}$$

となるので, $\mu(|f - g| > \varepsilon) = 0$ でなければならない. □

**系 5.3.2.** 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Sigma^r(\Psi)$ が $C(\mathbb{R}^r, \mathbb{R})$ において広義一様収束の意味で稠密になるものとする. このとき, 任意の関数 $g : \{0,1\}^r \to \{0,1\}$ と $\varepsilon > 0$ に対して, $f \in \Sigma^r(\Psi)$ が存在して,

$$\sup_{x \in \{0,1\}^r} |f(x) - g(x)| < \varepsilon$$

となる.

**証明.** $\mu$ を $\{0,1\}^r$ の各点で $1/2^r$ を値に取る $(\mathbb{R}^r, \mathcal{B}^r)$ 上の測度とする. $g$(正確には $g$ を $\mathbb{R}^r$ に拡張したもの) は $\mathbb{M}^r$ に属するので, すぐ上の定理より $f \in \Sigma^r(\Psi)$ が存在して, $\mu(|f - g| \ge \varepsilon) = 0$ となる. このとき, もし $x \in \{0,1\}^r$ が存在して, $|f(x) - g(x)| \ge \varepsilon$ となるなら, $\mu((|f - g| \ge \varepsilon) \ge \mu(\{x\}) = 1/2^r > 0$ となり不合理である. よって, $\sup_{x \in \{0,1\}^r} |f(x) - g(x)| < \varepsilon$ である. □

さらに, $\Psi$ がある条件を満たせば, 任意の関数 $g : \mathbb{R}^r \to \mathbb{R}$ と任意に与えられた有限個の点に対して, その点の上で $g$ と値が一致するような $f \in \Sigma^r(\Psi)$ が存在する. このことを示すために次の補題を示す.

**補題 5.3.3.** $n \in \mathbb{N}$ とし, $K$ を $|K| \geq n+1$ なる体とする. このとき, 任意の $K$ 上の線形空間 $V$ は, $k \leq n$ 個の真線形部分空間の和集合で表せない. つまり, どんな線形部分空間 $V_1, \ldots, V_k \subsetneq V$ についても $V = \bigcup_{j=1}^{k} V_j$ とはならない. 特に, $\mathbb{R}$ 上の線形空間 $V$ は真線形部分空間の有限和集合で表せない.

**証明.** $n$ に関する帰納法による. $n = 1$ のときは明らかに成立する (線形空間と真線形部分空間は一致しない). $n$ で成立すると仮定する. このとき, もしある $|K| \geq n+2$ なる体 $K$ と $K$ 上の線形空間 $V$ および $k \leq n+1$ について, 線形部分空間 $V_1, \ldots, V_k \subsetneq V$ が存在して $V = \bigcup_{j=1}^{k} V_j$ となったとする. ここですべての $V_j$ は $\{0\}$ でないとしてよい. いま, $x \in V_k \setminus \{0\}$ をとり, $y \in V \setminus V_k$ をとると, 任意の $\alpha \in K \setminus \{0\}$ に対して $x + \alpha y \in V \setminus V_k = \bigcup_{j=1}^{k-1} V_j$ である. したがって, $|K| \geq n+2$ に注意して相異なる $\alpha_1, \ldots, \alpha_k \in K \setminus \{0\}$ を取ると, 鳩の巣原理により, ある $l \neq k$ と $i, j$ について $x + \alpha_i y, x + \alpha_j y \in V_l$ となる. すると, $V_l$ は線形部分空間であるので,

$$(\alpha_i - \alpha_j)y = (x + \alpha_i y) - (x + \alpha_j y) \in V_l$$

となり, $\alpha_i - \alpha_j \neq 0$ ゆえ, この逆元をかけることで $y \in V_l$ がわかる. よって, $x = (x + \alpha_i y) - \alpha_i y \in V_l$ である. 以上から, $V = \bigcup_{j=1}^{k-1} V_j$ となる. これは帰納法の仮定に反する. ゆえに $n+1$ のときも成立しなければならない. □

**定理 5.3.4.** $\Psi : \mathbb{R} \to \mathbb{R}$ を squashing 関数とする. もし $\Psi(z_0) = 0, \Psi(z_1) = 1$ なる $z_0, z_1 \in \mathbb{R}^r$ が存在するならば, 相異なる任意の点 $x_1, \ldots, x_n \in \mathbb{R}^r$ と任意の関数 $g : \mathbb{R}^r \to \mathbb{R}$ に対して, $f \in \Sigma^r(\Psi)$ が存在して $f(x_j) = g(x_j)$ $(j = 1, \ldots, n)$ となる.

**証明.** 二段階に分けて証明する. Step1 で $r = 1$ の場合を証明し, Step2 で $r \geq 2$ の場合を証明する.

**Step1**: 添字を付け替えることで $x_1 < \cdots < x_n$ としてよい. いま, $\Psi$ は単調非減少なので $M > 0$ を適当に取れば任意の $x \in \mathbb{R}$ に対して, $x \leq -M$ ならば $\Psi(x) = 0$ となり, $x \geq M$ ならば $\Psi(x) = 1$ となる. そこで, $A_1(x) = M$ $(x \in \mathbb{R})$ とし, $\beta_1 = g(x_1)$ とおく. また, $A_j \in \mathbb{A}^1$ $(j = 2, \ldots, n)$ を $A_j(x_{j-1}) = -M$ かつ $A_j(x_j) = M$ となるものとし, $\beta_j = g(x_j) - g(x_{j-1})$ とする. すると,

$$f = \sum_{j=1}^{n} \beta_j \Psi \circ A_j \ \in \Sigma^1(\Psi)$$

が求めるものである.

**Step2**: $r \geq 2$ とする. このとき, ある $p \in \mathbb{R}^r$ が存在して, $i \neq j$ ならば $p^{\mathrm{T}}(x_i - x_j) \neq 0$ となる. 実際, $\bigcup_{i \neq j} \{q \in \mathbb{R}^r \mid q^{\mathrm{T}}(x_i - x_j) = 0\}$ は $\mathbb{R}^r$ の超平面の有限和であるので補題 5.3.3 より $\mathbb{R}^r$ を覆い尽くすことはできない. ここで, 添字を付け替えることで $p^{\mathrm{T}}x_1 < \cdots < p^{\mathrm{T}}x_n$ としてよい. すると, Step1 と同等にして, $\beta_j \in \mathbb{R}$, $A_j \in \mathbb{A}^1$ が存在して,

$$\sum_{j=1}^{n} \beta_j \Psi(A_j(p^{\mathrm{T}}x_k)) = g(x_k) \quad (k = 1, \ldots, n)$$

となる. $A'_j(x) = A_j(p^{\mathrm{T}}x)$ $(x \in \mathbb{R}^r)$ とおくと $A'_j \in \mathbb{A}^r$ であるので,

$$f = \sum_{j=1}^{n} \beta_j \Psi \circ A'_j$$

が求めるものである. □

次の定理は Pinkus[28] による.

**定理 5.3.5.** $\Psi \in C(\mathbb{R})$ はある開区間上で多項式でないとする. このとき, 任意の $\alpha_1, \ldots, \alpha_k \in \mathbb{R}$ と相異なる任意の点 $x_1, \ldots, x_n \in \mathbb{R}^r$ に対して, $w_1, \ldots, w_k \in \mathbb{R}^r$ と $c_1, \ldots, c_k$,$b_1, \ldots, b_k \in \mathbb{R}$ が存在して,

$$\sum_{j=1}^{k} c_j \Psi(w_j^{\mathrm{T}} x_i + b_j) = \alpha_i \quad (i = 1, \ldots, k)$$

となる.

**証明.** $K := \{x_1, \ldots, x_k\} \subset \mathbb{R}^r$ とおく. いま, すべてが 0 とはならない $d_1, \ldots, d_k \in \mathbb{R}$ が存在して

$$\sum_{i=1}^{k} d_i \Psi(w^{\mathrm{T}} x_i + b) = 0 \quad (\forall w \in \mathbb{R}^r, b \in \mathbb{R})$$

となると仮定しよう. このとき,

$$G : C(K) \ni f \mapsto \sum_{i=1}^{k} d_i f(x_i) \in \mathbb{R}$$

とおくと, ある $d_i$ は 0 でないので $G$ は 0 でない有界線形汎関数であり,

$$G(f) = 0 \quad (\forall f \in \mathrm{span}\{K \ni x \mapsto \Psi(w^{\mathrm{T}} x + b) \mid w \in \mathbb{R}^r, b \in \mathbb{R}\})$$

となる. よって, $\mathrm{span}\{K \ni x \mapsto \Psi(w^{\mathrm{T}} x + b) \mid w \in \mathbb{R}^r, b \in \mathbb{R}\}$ は $C(K)$ において稠密でない (稠密なら $G = 0$ でければならない). これは万能近似定理 (定理 4.1.2) に反する. よって, $w, b$ の連続関数 $\Psi(w^{\mathrm{T}} x_1 + b), \ldots, \Psi(w^{\mathrm{T}} x_k + b)$ は一次独立である. ゆえに, $w_1, \ldots, w_k \in \mathbb{R}^r$ 及び $b_1, \ldots, b_k$ が存在して

$$\det((\Psi(w_j x_i + b_j))_{i,j=1,\ldots,k}) \neq 0$$

となる. ゆえに, $c_1, \ldots, c_k \in \mathbb{R}$ に関する連立一次方程式

$$\sum_{j=1}^{k} c_j \Psi(w_j^{\mathrm{T}} x_i + b_j) = \alpha_i \quad (i = 1, \ldots, k)$$

は解を持つ. □

# 6　RBF ネットワークの万能近似定理

本章では以下で定義する RBF(Radial-Basis-Function) ネットワークの万能近似能力について述べる．すなわち, 関数 $K$ がある条件をみたすときに $K$ を基底関数とする RBF ネットワークが $L^p(\mathbb{R}^r)$ で稠密であることや, $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることを示す.

**定義 6.1.** 関数 $K : \mathbb{R}^r \to \mathbb{R}$ に対して,

$$q(x) = \sum_{i=1}^{M} w_i K\left(\frac{x - z_i}{\sigma}\right) \quad (M \in \mathbb{N}, w_i \in \mathbb{R}, \sigma > 0, z_i \in \mathbb{R}^r)$$

という形の関数 $q : \mathbb{R}^r \to \mathbb{R}$ 全体の集合を $S_K$ と表し, $S_K$ を $K$ を基底関数とする RBF ネットワークという.

まず次の補題を用意しておく.

**補題 6.2.** $f \in L^p(\mathbb{R}^r),\ p \in [1, \infty)$ とし, $\phi \in L^1(\mathbb{R}^r)$ は $\int_{\mathbb{R}^r} \phi(x)dx = 1$ をみたすとする．このとき, $\varepsilon > 0$ に対して,

$$\phi_\varepsilon(x) := \frac{1}{\varepsilon^r}\phi\left(\frac{x}{\varepsilon}\right) \quad (x \in \mathbb{R}^r)$$

と定義すると, $\|\phi_\varepsilon * f - f\|_{L^p} \to 0 \ \ (\varepsilon \to 0)$ となる.

**証明.** $\phi_\varepsilon \in L^1(\mathbb{R}^r) \ \ (\varepsilon > 0)$ であるので, 付録の命題 9.7.1 より $\phi_\varepsilon * f \in L^p(\mathbb{R}^r)$ である．また, a.e.$y \in \mathbb{R}^r$ に対して, 変数変換により,

$$(\phi_\varepsilon * f)(y) = \int_{\mathbb{R}^r} \phi_\varepsilon(x) f(y - x)dx = \int_{\mathbb{R}^r} f(y - \varepsilon x)\phi(x)dx$$

である．したがって, $\int_{\mathbb{R}^r} \phi(x)dx = 1$ より,

$$(\phi_\varepsilon * f)(y) - f(y) = \int_{\mathbb{R}^r} (f(y - \varepsilon x) - f(y))\phi(x)dx$$

である．そこで, $q \in [1, \infty]$ を $1/p + 1/q = 1$ をみたすものとすると, Hölder の不等式 (付録の命題 9.11.2) より,

$$\begin{aligned}
|(\phi_\varepsilon * f)(y) - f(y)| &\le \int_{\mathbb{R}^r} |f(y - \varepsilon x) - f(y)|\,|\phi(x)|dx \\
&= \int_{\mathbb{R}^r} |f(y - \varepsilon x) - f(y)|\,|\phi(x)|^{1-1/q}|\phi(x)|^{1/q}dx \\
&\le \||\phi|^{1/q}\|_{L^q} \left(\int_{\mathbb{R}^r} |f(y - \varepsilon x) - f(y)|^p|\phi(x)|dx\right)^{1/p}
\end{aligned}$$

となる ($p=1$ のときは $\||\phi|^{1/q}\|_{L^q}=1$ と解釈する). ゆえに, Fubini の定理より,

$$\begin{aligned}\|\phi_\varepsilon * f - f\|_{L^p}^p &\le \||\phi|^{1/q}\|_{L^q}^p \int_{\mathbb{R}^r}\int_{\mathbb{R}^r} |f(y-\varepsilon x)-f(y)|^p|\phi(x)|dxdy \\ &= \||\phi|^{1/q}\|_{L^q}^p \int_{\mathbb{R}^r} |\phi(x)|\|f(\cdot-\varepsilon x)-f(\cdot)\|_{L^p}^p dx\end{aligned}$$

となる. 任意の $x\in\mathbb{R}^r, \varepsilon>0$ に対して, 三角不等式より

$$|\phi(x)|\|f(\cdot-\varepsilon x)-f(\cdot)\|_{L^p} \le 2|\phi(x)|\|f\|_{L^p}$$

であり, $L^p$ ノルムの平行移動に関する連続性 (付録の命題 9.9.5) より,

$$\|f(\cdot-\varepsilon x)-f(\cdot)\|_{L^p}\to 0 \quad (\varepsilon\to 0)$$

であるから, 優収束定理より

$$\|\phi_\varepsilon * f - f\|_{L^p}\to 0 \quad (\varepsilon\to 0)$$

である. □

**定理 6.3 (J. Parkn and I. W. Sandberg, 1991 [26]).**
$\mu$ を $\mathbb{R}^r$ 上の Lebesgue 測度とする. 関数 $K:\mathbb{R}^r\to\mathbb{R}$ が以下を満たすとする.

(1) $K$ は有界.

(2) $K\in L^1(\mathbb{R}^r,\mu)$ かつ $\int_{\mathbb{R}^r} Kd\mu\neq 0$.

(3) $\mu(\{K$ の不連続点$\})=0$

このとき, 任意の $p\ge 1$ に対して, $S_K$ は $L^p(\mathbb{R}^r,\mu)$ において $L^p$ ノルムについて稠密である.

**証明.** 任意に $f\in L^p(\mathbb{R}^r,\mu)$ と $\varepsilon>0$ をとる. $C_0(\mathbb{R}^r)$ は $L^p(\mathbb{R}^r, u)$ において $L^p$ ノルムについて稠密であるから, $f_c\in C_0(\mathbb{R}^r)$ が存在して, $\|f-f_c\|_{L^p}<\varepsilon/3$ となる. $0\in S_K$ であるから $f_c\neq 0$ と仮定してよい. ここで関数 $\phi:\mathbb{R}^r\to\mathbb{R}$ を

$$\phi(x)=\frac{1}{\int_{\mathbb{R}^r} Kd\mu}K(x)$$

と定義する. そして, $\sigma>0$ に対して, $\phi_\sigma(x)=\phi(x/\sigma)/\sigma^r$ と定義する. このとき, $\|\phi_\sigma * f_c - f_c\|_{L^p}\to 0$ $(\sigma\to 0)$ となる. そこで, $\sigma>0$ を $\|\phi_\sigma * f_c - f_c\|_{L^p}<\varepsilon/3$ となるようにとっておく. 次に $\mathrm{supp} f_c\subset[-T,T]^r$ なる $T>0$ をとる. 各 $\alpha\in\mathbb{R}^r$ に対して $\phi_\sigma(\alpha-\cdot)f_c(\cdot)$ は有界かつほとんど至る所連続であるから $[-T,T]^r$ 上 Riemann

積分可能である．したがって, 各 $[-T,T]$ を $n$ 等分することで得られる $[-T,T]^r$ の分割を $\Delta_n$ とおき, $n^r$ 個ある $\Delta_n$ の各小領域から任意に $\alpha_i$ を選ぶと,

$$v_n(\alpha) := \sum_{i=1}^{n^r} \phi_\sigma(\alpha-\alpha_i) f_c(\alpha_i) \left(\frac{2T}{n}\right)^r$$

は $\int_{[-T,T]^r} \phi_\sigma(\alpha-x) f_c(x)dx$ に収束する．一方, $\mathrm{supp} f_c \subset [-T,T]^r$ であるので,

$$\int_{[-T,T]^r} \phi_\sigma(\alpha-x) f_c(x)dx = \int_{\mathbb{R}^r} \phi_\sigma(\alpha-x) f_c(x)dx = (\phi_\sigma * f_c)(\alpha)$$

である．よって, 各点 $\alpha \in \mathbb{R}^r$ について $v_n(\alpha) \to (\phi_\sigma * f_c)(\alpha)$ となる．$\phi_\sigma * f_c \in L^p(\mathbb{R}^r, \mu)$ に注意すると, ある $T_1 > 0$ について,

$$\int_{\mathbb{R}^r \setminus [-T_1,T_1]^r} |\phi_\sigma * f_c|^p d\mu < (\varepsilon/9)^p$$

となる．また, $\phi_\sigma \in L^1(\mathbb{R}^r, \mu)$ であり, かつ $\phi_\sigma$ は有界であることより $|\phi_\sigma|^p \in L^1$ となるから, ある $T_2 > 0$ について,

$$\int_{\mathbb{R}^r \setminus [-T_2,T_2]^r} |\phi_\sigma|^p d\mu < \left(\frac{\varepsilon}{9\|f_c\|_{L^\infty}(2T)^r}\right)^p$$

となる．$[0,\infty) \ni y \to y^p \in \mathbb{R}$ が凸関数であることより, 任意の $\alpha \in \mathbb{R}^r$ に対して,

$$\left(\frac{1}{n^r}\sum_{i=1}^{n^r} |\phi_\sigma(\alpha-\alpha_i)|\right)^p \leq \frac{1}{n^r}\sum_{i=1}^{n^r} |\phi_\sigma(\alpha-\alpha_i)|^p$$

となることに注意すると,

$$|v_n(\alpha)|^p \leq (\|f_c\|_{L^\infty}(2T)^r)^p \frac{1}{n^r}\sum_{i=1}^{n^r} |\phi_\sigma(\alpha-\alpha_i)|^p$$

である．ここで, $T_0 = \max\{T_1, T_2+T\}$ とおくと, $\alpha_i$ の各成分 $\alpha_{i,j}$ は $|\alpha_{i,j}| \leq T$ をみたすので,

$$\int_{[-T_2,T_2]^r} |\phi_\sigma(\alpha)|^p d\alpha \leq \int_{[-T_0,T_0]^r} |\phi_\sigma(\alpha-\alpha_i)|^p d\alpha$$

となる．したがって,

$$\int_{\mathbb{R}^r \setminus [-T_0,T_0]^r} |\phi_\sigma(\alpha-\alpha_i)|^p d\alpha \leq \int_{\mathbb{R}^r \setminus [-T_2,T_2]^r} |\phi_\sigma(\alpha)|^p d\alpha$$

であるので,

$$
\begin{aligned}
&\int_{\mathbb{R}^r\setminus[-T_0,T_0]^r} |v_n(\alpha)|^p d\alpha \\
&\leq (\|f_c\|_{L^\infty}(2T)^r)^p \frac{1}{n^r}\sum_{i=1}^{n^r}\int_{\mathbb{R}^r\setminus[-T_0,T_0]^r} |\phi_\sigma(\alpha-\alpha_i)|^p d\alpha \\
&\leq (\|f_c\|_{L^\infty}(2T)^r)^p \frac{1}{n^r}\sum_{i=1}^{n^r}\int_{\mathbb{R}^r\setminus[-T_2,T_2]^r} |\phi_\sigma(\alpha)|^p d\alpha \\
&< (\varepsilon/9)^p
\end{aligned}
$$

となる. また, $T_0 \geq T_1$ より,

$$
\int_{\mathbb{R}^r\setminus[-T_0,T_0]^r} |\phi_\sigma * f_c|^p d\mu < (\varepsilon/9)^p
$$

である. いま, $\phi_\sigma * f_c \in L^p$ と $v_n$ の有界性より優収束定理が使えて,

$$
\int_{[-T_0,T_0]^r} |\phi_\sigma * f_c - v_n|^p d\mu \to 0 \quad (n\to\infty)
$$

となる. ゆえに, 適当に $N \in \mathbb{N}$ を取ると,

$$
\int_{[-T_0,T_0]^r} |\phi_\sigma * f_c - v_N|^p d\mu < (\varepsilon/9)^p
$$

となる. 以上から,

$$
\begin{aligned}
\|v_N - \phi_\sigma * f_c\|_{L^p} &\leq \|v_N\cdot\chi_{\mathbb{R}^r\setminus[-T_0,T_0]^r}\|_{L^p} + \|(v_N-\phi_\sigma * f_c)\cdot\chi_{[-T_0,T_0]^r}\|_{L^p} \\
&\quad + \|(\phi_\sigma * f_c)\cdot\chi_{\mathbb{R}^r\setminus[-T_0,T_0]^r}\|_{L^p} \\
&< \varepsilon/9+\varepsilon/9+\varepsilon/9 = \varepsilon/3
\end{aligned}
$$

となる. よって,

$$
\|f-v_N\|_{L^p} \leq \|f-f_c\|_{L^p} + \|f_c-\phi_\sigma * f_c\|_{L^p} + \|\phi_\sigma * f_c - v_N\|_{L^p} < \varepsilon
$$

である. そして,

$$
v_N = \sum_{i=1}^{N^r}\phi_\sigma(\cdot-\alpha_i)f_c(\alpha_i)\left(\frac{2T}{N}\right)^r = \sum_{i=1}^{N^r} w_i K\left(\frac{\cdot-\alpha_i}{\sigma}\right) \in S_K
$$

である. ただし,

$$
w_i = \frac{1}{\sigma^r} f_c(\alpha_i)\left(\frac{2T}{N}\right)^r \frac{1}{\int_{\mathbb{R}^r} K d\mu}
$$

である. これで示せた. □

**定理 6.4.** $\mu$ を $\mathbb{R}^r$ 上の Lebesgue 測度とする．関数 $K:\mathbb{R}^r\to\mathbb{R}$ が以下を満たすとする．

(1) $K$ は有界.

(2) $K\in L^1(\mathbb{R}^r,\mu)$ かつ $\int_{\mathbb{R}^r}Kd\mu\neq 0$.

(3) $K$ は連続関数である.

このとき, $S_K$ は $C(\mathbb{R}^r,\mathbb{R})$ において広義一様収束の意味で稠密である.

**証明.** まず, 先と同様に, 関数 $\phi:\mathbb{R}^r\to\mathbb{R}$ を

$$\phi(x)=\frac{1}{\int_{\mathbb{R}^r}Kd\mu}K(x)$$

と定義し, $\sigma>0$ に対して, $\phi_\sigma(x)=\phi(x/\sigma)/\sigma^r$ と定義する．任意に $M>0$ をとる．$[-M,M]^r$ 上の関数 $f$ と $\varepsilon>0$ に対して, $g\in S_K$ が存在して, $\sup_{x\in[-M,M]^r}|f(x)-g(x)|<\varepsilon$ となることを示せば十分である．$T>M$ をとると, Tietze の拡張定理より

$$\tilde{f}(x)=f(x)\ (x\in[-M,M]^r),\quad \tilde{f}(x)=0\ (x\in\mathbb{R}^r\setminus[-T,T])$$

なる連続関数 $\tilde{f}:\mathbb{R}^r\to\mathbb{R}$ がとれる．明らかに $\tilde{f}$ は一様連続であるので, ある $\delta>0$ が存在して,

$$\forall x,y\in\mathbb{R}^r,|x-y|<\delta\Rightarrow|\tilde{f}(x)-\tilde{f}(y)|<\frac{\varepsilon}{4\|\phi\|_{L^1}}$$

となる．一方, $\phi\in L^1(\mathbb{R}^r,\mu)$ なので, ある $T_0>0$ について,

$$\int_{\mathbb{R}^r\setminus[-T_0,T_0]^r}|\phi|d\mu<\frac{\varepsilon}{8\|\tilde{f}\|_{L^\infty}}$$

となる．そこで, 適当に $\sigma>0$ をとり, $\forall x\in[-T_0,T_0]^r,\ |\sigma x|<\delta$ となるようにすると, 任意の $\alpha\in[-M,M]^r$ について,

$$\begin{aligned}
|(\phi_\sigma*\tilde{f})(\alpha)-\tilde{f}(\alpha)|&=\left|\int_{\mathbb{R}^r}\frac{1}{\sigma^r}\phi\left(\frac{x}{\sigma}\right)\tilde{f}(\alpha-x)dx-\int_{\mathbb{R}^r}\tilde{f}(\alpha)\phi(x)dx\right|\\
&=\left|\int_{\mathbb{R}^r}(\tilde{f}(\alpha-\sigma x)-\tilde{f}(\alpha))\phi(x)dx\right|\\
&\leq\int_{\mathbb{R}^r}|\tilde{f}(\alpha-\sigma x)-\tilde{f}(\alpha)|\cdot|\phi(x)|dx\\
&\leq\int_{[-T_0,T_0]^r}|\tilde{f}(\alpha-\sigma x)-\tilde{f}(\alpha)|\cdot|\phi(x)|dx\\
&\quad+\int_{\mathbb{R}^r\setminus[-T_0,T_0]^r}2\|\tilde{f}\|_{L^\infty}|\phi(x)|dx\\
&<\varepsilon/2
\end{aligned}$$

となる．ここで, 前定理の証明と同様に, $[-T,T]^r$ を等分割し, 各 $\alpha \in \mathbb{R}^r$ に対して,

$$\tilde{v}_n(\alpha) = \sum_{i=1}^{n^r} \phi_\sigma(\alpha - \alpha_i)\tilde{f}(\alpha_i)\left(\frac{2T}{n}\right)^r$$

と定義する．関数 $(s,x) \mapsto \phi_\sigma(s-x)\tilde{f}(x)$ は $[-M,M]^r \times [-T,T]^r$ 上で一様連続なので, ある $\delta_0 > 0$ が存在して, 任意の $s \in [-M,M]^r, x, y \in [-T,T]^r$ について, $|x-y| < \delta_0$ ならば,

$$|\phi_\sigma(s-x)\tilde{f}(x) - \phi_\sigma(s-y)\tilde{f}(y)| < \frac{\varepsilon}{2(2T)^r}$$

となる．$N$ を十分大きく取ると, $\Delta_N$ の各小領域の任意の 2 点間の距離は $\delta_0$ 未満になる．ゆえに, $[-T,T]^r = \bigcup_{i=1}^{N^r} J_i$ と表すことにすると, 任意の $\alpha \in [-M,M]^r$ に対して,

$$\begin{aligned}
&|v_N(\alpha) - \int_{[-T,T]^r} \phi_\sigma(\alpha - x)\tilde{f}(x)dx| \\
&\leq \sum_{i=1}^{N^r} \int_{J_i} |\phi_\sigma(\alpha - \alpha_i)\tilde{f}(\alpha_i) - \phi_\sigma(\alpha - x)\tilde{f}(x)|dx \\
&\leq \sum_{i=1}^{N^r} \int_{J_i} \frac{\varepsilon}{2(2T)^r} dx \leq \varepsilon/2
\end{aligned}$$

となる．任意の $\alpha \in [-M,M]^r$ に対して,

$$\int_{[-T,T]^r} \phi_\sigma(\alpha - x)\tilde{f}(x)dx \quad = (\phi_\sigma * \tilde{f})(\alpha)$$

であることに注意すると,

$$\begin{aligned}
|f(\alpha) - v_N(\alpha)| &= |\tilde{f}(\alpha) - v_N(\alpha)| \\
&\leq |\tilde{f}(\alpha) - (\phi_\sigma * \tilde{f})(\alpha)| + |(\phi_\sigma * \tilde{f})(\alpha) - v_N(\alpha)| \\
&< \varepsilon
\end{aligned}$$

となる．よって, $\sup_{x\in[-M,M]^r} |f(x) - v_N(x)| \leq \varepsilon$ である．これで示せた． □

なお, 以上の結果に関連して, Pinkus により次の結果が得られている．

**定理 6.5 (Pinkus 1996 [27] Theorem12).** $\sigma \in C(\mathbb{R}_+)$ に対して,

$$\mathrm{span}\{x \mapsto \sigma(\rho\|x-a\|) \mid \rho > 0, a \in \mathbb{R}^r\}$$

が $C(\mathbb{R}^r)$ において広義一様収束の意味で稠密であることと, $\sigma$ が偶多項式と一致しないことは同値である．

# 7　近似レートの評価

本章では, 与えられた関数空間の部分集合を 3 層ニューラルネットワークにより近似することを考えるとき, 中間層のユニット数に関してどれくらいの効率で近似誤差が減少していくかという問題を考える. つまり, $X$ を

$$\mathcal{N}_r^n(\Psi) := \left\{ x \mapsto \sum_{i=1}^{n} c_i \Psi(w_i^{\mathrm{T}} x + b_i) \,\middle|\, w_i \in \mathbb{R}^r, c_i, b_i \in \mathbb{R} \right\}$$

を含む関数空間とするとき, 部分集合 $\mathcal{F} \subset X$ に対して, 近似誤差

$$\sup_{f \in \mathcal{F}} \inf_{g \in \mathcal{N}_r^n(\Psi)} \|f - g\|_X$$

は $n$ を大きくしていくときどのようなレートで減少するか？という問題を考える.

## 7.1　Sobolev 空間における近似レート

本節では Pinkus,1999 [28] で紹介されている, Sobolev 空間における近似レートの結果を見ていく. まず近似誤差を一般的な形で定義しておく.

**定義 7.1.1 (近似誤差).**
$X$ をノルム空間とする. $f \in X$ と空でない $Y \subset X$ に対して,

$$\|f - Y\|_X := \inf_{g \in Y} \|f - g\|_X$$

と定義する. 付録の補題 9.8.5 より $X \ni f \mapsto \|f - Y\|_X \in \mathbb{R}$ は連続である. さらに, 空でない $Z \subset X$ に対して,

$$E(Z, Y, X) := \sup_{z \in Z} \|z - Y\|_X = \sup_{z \in Z} \inf_{y \in Y} \|z - y\|_X$$

と定義する.

次に Sobolev 空間の定義を述べよう. Sobolev 空間についての詳しい説明は宮島 [37] などを参照されたい.

**定義 7.1.2 (Sobolev 空間).**
開集合 $\Omega \subset \mathbb{R}^r$ と $p \in [1, \infty]$ および整数 $m \geq 1$ に対して,

$$W^{m,p}(\Omega) := \left\{ f \in L^p(\Omega) \,\middle|\, \begin{array}{l} |\alpha| \leq m \text{ なる任意の} \alpha \in \mathbb{Z}_{\geq 0}^r \text{に対して} \\ f \text{ の弱導関数} \partial^\alpha f \text{ が存在して} \partial^\alpha f \in L^p(\Omega) \end{array} \right\}$$

と定義する．ここで, $h:\Omega\to\mathbb{R}$ が $\partial^\alpha$ に関する $f$ の弱導関数であるとは, 任意の $\varphi\in C_0^\infty(\Omega)$ に対して,

$$\int_\Omega f(x)(\partial^\alpha\varphi)(x)dx=(-1)^{|\alpha|}\int_\Omega h(x)\varphi(x)dx$$

が成り立つことをいう．さらに, $f\in W^{m,p}(\Omega)$ に対して,

$$\|f\|_{W^{m,p}(\Omega)}:=\sum_{|\alpha|\le m}\|\partial^\alpha f\|_{L^p(\Omega)}$$

と定義する．また,

$$\mathcal{B}^{m,p}(\Omega):=\{f\in W^{m,p}(\Omega)\mid \|f\|_{W^{m,p}(\Omega)}\le 1\}$$

と定義する．

Pinkus[28] では近似誤差 $E(\mathcal{B}^{m,p},\mathcal{N}_r^n(\Psi),L^p)$ の $n$ に関する減少率について 1999 年までの結果がいくつか紹介されている．

さて, まず多項式による Sobolev 空間の近似について述べるために多項式の空間の定義を思い出しておく．

**定義 7.1.3.** $r,k\in\mathbb{N}$ に対して, $r$ 変数の $k$ 次斉次多項式関数の集合を

$$\mathcal{H}_k(\mathbb{R}^r):=\left\{\mathbb{R}^r\ni x\mapsto\sum_{|\alpha|=k}c_\alpha x^\alpha\in\mathbb{R}\mid c_\alpha\in\mathbb{R}\right\}$$

と定義する．ただし, 和は $|\alpha|=k$ なる $\alpha\in(\mathbb{Z}_{\ge 0})^r$ 全体にわたるものとする．また,

$$\mathcal{P}_k(\mathbb{R}^r):=\bigcup_{s=0}^{k}\mathcal{H}_s(\mathbb{R}^r),\quad \mathcal{P}(\mathbb{R}^r):=\bigcup_{s=0}^{\infty}\mathcal{H}_s(\mathbb{R}^r)$$

と定義する．さらに $r$ 変数の $k$ 次以下の $\mathbb{R}$ 係数多項式全体の集合を

$$\pi_k(\mathbb{R}^r):=\left\{\mathbb{R}^r\ni x\mapsto\sum_{|\alpha|\le k}c_\alpha x^\alpha\in\mathbb{R}\mid c_\alpha\in\mathbb{R}\right\}$$

と定義する．

多項式による Sobolev 空間の近似について次が成り立つことが知られている．

**定理 7.1.4 (Jackson 評価).**
$I=(-1,1)$ とおく．$p\in[1,\infty],m,r\ge 1$ とするとき, ある $C>0$ が存在して,

$$E(\mathcal{B}^{m,p}(I^r),\pi_k(\mathbb{R}^r),L^p(I^r))\le Ck^{-m}\qquad(k=1,2,3,\ldots)$$

となる．

**証明.** 付録の定理 9.12.27 を参照されたい. □

次に活性化関数が $C^\infty$ 級で多項式と一致しない場合のニューラルネットワークによる近似レートについて述べる. そのためにまず補題 4.1.5 を精密化しよう.

**補題 7.1.5.** $n \in \mathbb{N}$ とすると, 次の式が成り立つ.

$$\sum_{i=0}^{n}(-1)^i\binom{n}{i}(n-2i)^j = \begin{cases} 0 & (j=0,\dots,n-1) \\ 2^n n! & (j=n) \end{cases}$$

**証明.**

$$\begin{aligned}
\sum_{i=0}^{n}(-1)^i\binom{n}{i}(n-2i)^j &= \sum_{i=0}^{n}(-1)^i\binom{n}{i}\left(\frac{d}{dx}\right)^j \exp((n-2i)x)\Bigg|_{x=0} \\
&= \left(\frac{d}{dx}\right)^j \left(\sum_{i=0}^{n}(-1)^i\binom{n}{i}\exp((n-2i)x)\right)\Bigg|_{x=0} \\
&= \left(\frac{d}{dx}\right)^j \left(\sum_{i=0}^{n}\binom{n}{i}\exp(x)^{n-i}(-\exp(-x))^i\right)\Bigg|_{x=0} \\
&= \left(\frac{d}{dx}\right)^j (\exp(x)-\exp(-x))^n\Bigg|_{x=0} \\
&= 2^n\left(\frac{d}{dx}\right)^j (\sinh(x))^n\Bigg|_{x=0}
\end{aligned}$$

である. ここに

$$\sinh(x) = \frac{\exp(x)-\exp(-x)}{2}$$

である. よって

$$\left(\frac{d}{dx}\right)^j (\sinh(x))^n\Bigg|_{x=0} = \begin{cases} 0 & (j=0,\dots,n-1) \\ n! & (j=n) \end{cases}$$

を示せばよい. これを $n$ についての帰納法で示そう. $n=1$ での成立は明らか. $n$ で成立すると仮定すると, $j=0,\dots,n-1$ に対しては, 積の微分に関する Leibniz 則より

$$\begin{aligned}
&\left(\frac{d}{dx}\right)^j (\sinh(x))^{n+1}\Bigg|_{x=0} \\
&= \sum_{k=0}^{j}\binom{j}{k}\left(\left(\frac{d}{dx}\right)^k (\sinh(x))^n\right)\left(\left(\frac{d}{dx}\right)^{j-k}\sinh(x)\right)\Bigg|_{x=0} = 0.
\end{aligned}$$

また,

$$\left(\frac{d}{dx}\right)^n (\sinh(x))^{n+1}\bigg|_{x=0} = \left(\left(\frac{d}{dx}\right)^n (\sinh(x))^n\right)\sinh(x)\bigg|_{x=0} = 0$$

であり,

$$\begin{aligned}\left(\frac{d}{dx}\right)^{n+1} (\sinh(x))^{n+1}\bigg|_{x=0} &= \binom{n+1}{n}\left(\left(\frac{d}{dx}\right)^n (\sinh(x))^n\right)\left(\frac{d}{dx}\sinh(x)\right)\bigg|_{x=0} \\ &\quad + \left(\left(\frac{d}{dx}\right)^{n+1}\sinh(x)\right)\sinh(x)\bigg|_{x=0} \\ &= (n+1)(n!) + 0 = (n+1)!\end{aligned}$$

である. これで示せた. □

**補題 7.1.6.** $W \subset \mathbb{R}$ は $0$ を内点として持つとする. また, $\Psi : \mathbb{R} \to \mathbb{R}$ はある開区間 $I$ 上で $C^\infty$ かつ多項式と一致しないとする. このとき, 各 $n$ について多項式 $x \mapsto x^n$ は

$$\mathcal{N}_1^{n+1}(\Psi, W, I) = \left\{ x \mapsto \sum_{i=1}^{n+1} c_i \Psi(w_i x + b_i) \mid c_i \in \mathbb{R}, w_i \in W, b_i \in I \right\}$$

で広義一様近似される. さらに, 1 変数の $n$ 次以下の $\mathbb{R}$ 係数多項式は $\mathcal{N}_1^{2n+1}(\Psi, W, I)$ で広義一様近似される.

**証明.** $K \subset \mathbb{R}$ を空でないコンパクト集合とする. $\Psi$ は $I$ 上 $C^\infty$ 級で多項式と一致しないので付録の定理 9.14.1 よりある $b \in I$ が存在して任意の整数 $n \geq 0$ について $\Psi^{(n)}(b) \neq 0$ となる. $f : \mathbb{R} \times \mathbb{R} \ni (x, w) \mapsto \Psi(wx + b) \in \mathbb{R}$ と定義する. 前半を示すには, 任意の $n \in \mathbb{N}$ に対して, 関数 $x \mapsto ((\partial w)^n f)(x, 0)$ が $\mathcal{N}_1^{n+1}(\Psi, W, I)$ の元で $K$ 上一様近似できることを示せば十分である. 実際, それが示されれば

$$x \mapsto ((\partial w)^n f)(x, 0) = x^n \Psi^{(n)}(0x + b) = x^n \Psi^{(n)}(b)$$

なので $x \mapsto x^n$ は $\mathcal{N}_1^{n+1}(\Psi, W, I)$ の元で $K$ 上一様近似される. いま, $h > 0$ に対して,

$$F_x(h) := \sum_{i=0}^{n} (-1)^i \binom{n}{i} f(x, (n-2i)h) = \sum_{i=0}^{n} (-1)^i \binom{n}{i} \Psi((n-2i)hx + b)$$

とおくと, Taylor の定理より

$$F_x(h) = F_x(0) + hF_x'(0) + \cdots + \frac{h^n}{n!} F_x^{(n)}(0) + \frac{h^{n+1}}{(n+1)!} F_x^{(n+1)}(\xi_x) \quad (\exists \xi_x \in (0, h))$$

となる．前命題より

$$F_x^{(j)}(0) = ((\partial w)^j f)(x,0)\sum_{i=0}^{n}(-1)^i\binom{n}{i}(n-2i)^j = \begin{cases} 0 & (j=0,\dots,n-1) \\ n!2^n((\partial w)^n f)(x,0) & (j=n) \end{cases}$$

となる．従って, $M_1 = \sup\{((\partial w)^{n+1}f)(x,w) \mid x\in K, w\in[-n,n]\}$ とおき,

$$M_2 = M_1\left|\sum_{i=0}^{n}(-1)^i\binom{n}{i}(n-2i)^{n+1}\right|$$

とおけば, 任意の $\varepsilon>0$ に対して $\delta>0$ を $\delta<\min\{\varepsilon/M_2,1\}$ となるよう取ることで, 任意の $0<h<\delta$ と $x\in K$ に対して

$$\left|\frac{F_x(h)}{(2h)^n} - ((\partial w)^n f)(x,0)\right| \le \frac{h}{(n+1)!}|F^{(n+1)}(\xi_x)| \le hM_2 < \varepsilon$$

となる．よって, $0<h<\delta$ を満たし且つ $[-nh,nh]\subset W$ を満たす $h$ を取れば

$$\mathbb{R}\ni x \to \frac{F_x(h)}{(2h)^n} = \frac{1}{(2h)^n}\sum_{i=0}^{n}(-1)^i\binom{n}{i}\Psi((n-2i)hx+b)\in\mathbb{R}$$

は $\mathcal{N}_1^{n+1}(\Psi,W,I)$ に属し, $((\partial w)^n f)(\cdot,0)$ を $K$ 上一様近似する．これで前半が示せた．次に $k\le n$ に対して $n$ 次以下の多項式 $p(x)=a_nx^n+\cdots+a_0$ を任意に取る．このとき, 上で示したことにより任意の $\varepsilon>0$ に対して $h>0$ が存在して $[-nh,nh]\subset W$ 且つ

$$\sup_{x\in K}\left|x^j - \frac{\sum_{i=0}^{j}(-1)^i\binom{j}{i}\Psi((j-2i)hx+b)}{\Psi^{(k)}(b)(2h)^j}\right| < \varepsilon \quad (j=0,\dots,n)$$

となる．よって,

$$G(x) = \sum_{j=0}^{n} a_j\frac{\sum_{i=0}^{j}(-1)^i\binom{j}{i}\Psi((j-2i)hx+b)}{\Psi^{(k)}(b)(2h)^j}$$

とおくと,

$$\sup_{x\in K}|p(x)-G(x)| < \sum_{i=0}^{n}|a_i|\varepsilon$$

となる．そして,

$$S = \{(j-2i)h \mid \ j=0,\dots,n,\ i=0,\dots,j\}$$

の要素数は $2n+1$ であるので,

$$G(x) = \sum_{j=1}^{2n+1} c_j\Psi(w_jx+b) \quad (c_j\in\mathbb{R}, w_j\in W)$$

と表される．これで後半も示せた． □

**定理 7.1.7.** $I=(-1,1)$ とおく．$\Psi:\mathbb{R}\to\mathbb{R}$ はある開区間 $J$ 上で $C^\infty$ 級かつ多項式と一致しないとする．このとき, 各 $p\in[1,\infty], m,r\geq 1$ に対してある $C>0$ が存在して, 任意の整数 $n\geq 1$ に対して

$$E(\mathcal{B}^{m,p}(I^r),\mathcal{N}_r^n(\Psi),L^p(I^r))\leq Cn^{-m/r}$$

となる．

**証明.** $C$ は $r$ に依存してよいので $3r\leq n$ としてよい．$\binom{r-1+k}{k}(2k+1)\leq n$ なる最大の整数 $k\geq 1$ を取る．注意 3.2.9 より, $l:=\dim\mathcal{H}_k(\mathbb{R}^r)=\binom{r-1+k}{k}$ とおくと, $a_1,\ldots,a_l\in\mathbb{R}^r$ が存在して,

$$\pi_k(\mathbb{R}^r)=\left\{x\mapsto\sum_{i=1}^{l}g_i(a_i^{\mathrm{T}}x)\mid g_i\in\pi_k(\mathbb{R})\right\}$$

となる．また, 補題 7.1.6 より, $\pi_k(\mathbb{R})$ の元は $\mathcal{N}_1^{2k+1}(\Psi)$ により広義一様近似される．したがって,

$$\pi_k(\mathbb{R}^r)\subset\overline{\mathcal{N}_r^{l(2k+1)}(\Psi)}$$

となる (ただし, 広義一様収束の位相に関する閉包をとっている)．ゆえに, 定理 7.1.4 より $k$ に依存しない定数 $C>0$ が存在して,

$$\begin{aligned}E(\mathcal{B}^{m,p}(I^r),\overline{\mathcal{N}_r^n(\Psi)},L^p(I^r))&E(\mathcal{B}^{m,p}(I^r),\overline{\mathcal{N}_r^{l(2k+1)}(\Psi)},L^p(I^r))\\&\leq E(\mathcal{B}^{m,p}(I^r),\pi_k(\mathbb{R}^r),L^p(I^r))\leq Ck^{-m}\end{aligned}$$

となる．ここで,

$$E(\mathcal{B}^{m,p}(I^r),\mathcal{N}_r^n(\Psi),L^p(I^r))\leq E(\mathcal{B}^{m,p}(I^r),\overline{\mathcal{N}_r^n(\Psi)},L^p(I^r))$$

を示す．任意に $f\in\overline{\mathcal{N}_r^n(\Psi)}, g\in\mathcal{B}^{m,p}(I^r)$ をとる．$I^r$ 上で $f$ に一様収束する列 $f_j\in\mathcal{N}_r^n(\Psi)$ がとれる．すると,

$$\begin{aligned}\|g-f\|_{L^p(I^r)}&=\lim_{j\to\infty}\|g-f_j\|_{L^p(I^r)}\\&\geq\inf_{h\in\mathcal{N}_r^n(\Psi)}\|g-h\|_{L^p(I^r)}=\|g-\mathcal{N}_r^n(\Psi)\|_{L^p(I^r)}\end{aligned}$$

となる．$f\in\overline{\mathcal{N}_r^n(\Psi)}$ は任意なので,

$$\|g-\overline{\mathcal{N}_r^n(\Psi)}\|_{L^p(I^r)}\geq\|g-\mathcal{N}_r^n(\Psi)\|_{L^p(I^r)}$$

である．よって,

$$E(\mathcal{B}^{m,p}(I^r),\mathcal{N}_r^n(\Psi),L^p(I^r))\leq E(\mathcal{B}^{m,p}(I^r),\overline{\mathcal{N}_r^n(\Psi)},L^p(I^r))$$

である．以上より，

$$E(\mathcal{B}^{m,p}(I^r), \mathcal{N}_r^n(\Psi), L^p(I^r)) \leq Ck^{-m}$$

となる．いま $k$ の取り方から

$$n \leq \binom{(r-1)+(k+1)}{k+1}(2k+3)$$

となるが, $k$ に依存しない定数 $C_1 > 0$ が存在して

$$\begin{aligned}\binom{(r-1)+(k+1)}{k+1} &= \frac{((k+1)+(r-1))((k+1)+(r-2))\cdots((k+1)+1)}{(r-1)(r-2)\cdots 1} \\ &= \left(\frac{k+1}{r-1}+1\right)\left(\frac{k+1}{r-2}+1\right)\cdots\left(\frac{k+1}{1}+1\right) \\ &\leq C_1(k+1)^{r-1}\end{aligned}$$

となるので, $k$ に依存しない定数 $C_2 > 0$ が存在して,

$$n \leq \binom{r+k}{k+1}(2k+3) \leq C_2 k^r$$

となる．よって, $CC_2$ を改めて $C$ とおけば

$$E(\mathcal{B}^{m,p}(I^r), \mathcal{N}_r^n(\Psi), L^p(I^r)) \leq Cn^{-m/r}$$

である． □

次にリッジ関数による近似について以下が成り立つ．

**定理 7.1.8.** $I = (-1, 1)$ とおく．また, $p \in [1, \infty], m \geq 1, r \geq 2$ とする．このとき, ある $C > 0$ が存在して, 任意の整数 $n \geq 1$ に対して

$$E(\mathcal{B}^{m,p}(I^r), \mathcal{R}_r^n, L^p(I^r)) \leq Cn^{-m/(r-1)}$$

が成り立つ．ここで

$$\mathcal{R}_r^n := \left\{ x \mapsto \sum_{i=1}^n g_i(a_i^{\mathrm{T}} x) \mid a_i \in \mathbb{R}^r, g_i \in C(\mathbb{R}) \right\}$$

である．

**証明.**
$n \geq r$ としてよい．$k$ を $\binom{r-1+k}{k} \leq n$ なる最大の整数とする．注意 3.2.9 より, $l := \dim \mathcal{H}_k(\mathbb{R}^r) = \binom{r-1+k}{k}$ とおくと, $a_1, \ldots, a_l \in \mathbb{R}^r$ が存在して,

$$\pi_k(\mathbb{R}^r) = \left\{ x \mapsto \sum_{i=1}^l g_i(a_i^{\mathrm{T}} x) \mid g_i \in \pi_k(\mathbb{R}) \right\}$$

となる．したがって, $\pi_k(\mathbb{R}^r) \subset \mathcal{R}_r^l$ となる．ゆえに, 定理 7.1.4 より $k$ に依存しない定数 $C > 0$ が存在して,

$$\begin{aligned} E(\mathcal{B}^{m,p}(I^r), \mathcal{R}_r^n, L^p(I^r)) &\leq E(\mathcal{B}^{m,p}(I^r), \mathcal{R}_r^l, L^p(I^r)) \\ \leq E(\mathcal{B}^{m,p}(I^r), \pi_k(\mathbb{R}^r), L^p(I^r)) &\leq Ck^{-m} \end{aligned}$$

となる．いま $k$ に依存しない $C' > 0$ があって $n \leq C'k^{r-1}$ であるから, $CC'$ を改めて $C$ とおけば

$$E(\mathcal{B}^{m,p}(I^r), \mathcal{R}_r^n, L^p(I^r)) \leq Cn^{-m/(r-1)}$$

となる． □

上の定理と次の結果を合わせることでニューラルネットワークによる近似に関する結果が得られる．

**定理 7.1.9 (Maiorov and Pinkus 1999 [22]).**
狭義単調増加な sigmoidal 関数 $\Psi \in C^\infty(\mathbb{R})$ であって, 次の性質をみたすものが存在する: 任意の整数 $n, r \geq 1$ と $g \in \mathcal{R}_r^n, \varepsilon > 0$ に対して,

$$\sup_{x \in B(0,1)} \left| g(x) - \sum_{i=1}^{r+n+1} c_i \Psi(w_i^{\mathrm{T}} x + b_i) \right| < \varepsilon$$

なる $c_i, b_i \in \mathbb{R}, w_i \in \mathbb{R}^r$ が存在する．

**系 7.1.10.** $I = (-1, 1)$ とおく．狭義単調増加な sigmoidal 関数 $\Psi \in C^\infty(\mathbb{R})$ であって, 次の性質をみたすものが存在する: $p \in [1, \infty], m \geq 1, r \geq 2$ に対して,

$$E(\mathcal{B}^{m,p}(I^r), \mathcal{N}_r^{r+n+1}(\Psi), L^p(I^r)) \leq Cr^{-m/(n-1)}.$$

以上の結果は近似誤差の上からの評価であった．一方で次の結果は近似誤差を下から評価するものである．

**定理 7.1.11 (Maiorov and Meir 1999 [23], Theorem5).**
$I = (-1, 1)$ とし $p \in [1, \infty], m \geq 1, r \geq 2$ とする．活性化関数 $\Psi : \mathbb{R} \to \mathbb{R}$ が

$$\Psi(t) = \frac{1}{1 + \exp(-t)}$$

であるとき, ある定数 $C > 0$ が存在して, 任意の $n \geq 1$ に対して

$$E(\mathcal{B}^{m,p}(I^r), \mathcal{N}_r^n(\Psi), L^p(I^r)) \geq C(n \log(n))^{-m/r}$$

が成り立つ．

## 7.2 Barron space と Barron-Maurey-Jones 評価

前節の結果は, Sobolev 空間をニューラルネットで近似する際のレートの評価であった. そして, Sobolev 空間の近似誤差の上からの評価については, 求められる精度 $\varepsilon$ に対して中間ユニット数 $n$ を $\varepsilon^{-r/m}$ 以上に取る必要があり, 入力の次元 $r$ に関して指数関数的に増大することがわかる. これは次元の呪いとしばしば呼ばれる現象である. では, 逆に, この次元の呪いから解放される, つまり目標の精度を達成するために必要な中間ユニット数が $r$ に関してより緩やかに増大する, あるいは $r$ と無関係であるような関数空間はどのようなものであろうか？このような関数空間は Barron space と呼ばれ, 現在も研究が続いている (Weinan et al.[6], Siegel,Xu[31]). 本節では, 「G-variation(後述) に関して有界な空間」がこの問題のひとつの答えを与えることを示す. これは Kainen et al.[15] で紹介されている事実である.

まず, Barron-Maurey-Jones 評価と呼ばれる次の結果がある.

**定義 7.2.1.** $X$ を $\mathbb{R}$ 上の線形空間とするとき, $n \in \mathbb{N}$ と部分集合 $G \subset X$ に対して,

$$\mathrm{conv}_n(G) := \left\{ \sum_{i=1}^{n} a_i g_i \mid a_i \geq 0,\ \sum_{i=1}^{n} a_i = 1,\ g_i \in G \right\}$$

と定める. また,

$$\mathrm{conv}(G) := \bigcup_{n=1}^{\infty} \mathrm{conv}_n(G)$$

と定める. $\mathrm{conv}(G)$ は $G$ を含む最小の凸集合である.

**定理 7.2.2 (Barron-Maurey-Jones [17]).**
$X$ を内積空間とする. $G \subset X$ を空でない有界集合とし, $s_G := \sup_{g \in G} \|g\|_X$ とおくと, 任意の $f \in \mathrm{cl}(\mathrm{conv}(G))$ と $n \in \mathbb{N}$ に対して,

$$\|f - \mathrm{conv}_n(G)\|_X \leq \sqrt{\frac{s_G^2 - \|f\|_X^2}{n}}$$

が成り立つ.

**証明.** $X \ni f \mapsto \|f - G\|_X \in \mathbb{R}$ の連続性より, $f \in \mathrm{conv}(G)$ の場合に証明すれば十分である. そこで,

$$f = \sum_{i=1}^{m} a_i h_i \quad (a_i \geq 0, \sum_{i=1}^{m} a_i = 1, h_i \in G)$$

と表す. すると, 三角不等式より

$$\|f\| \leq \sum_{i=1}^{m} a_i \|h_i\| \leq \sum_{i=1}^{m} s_G = s_G$$

である. したがって,

$$c := s_G^2 - \|f\|^2 \geq 0$$

である. ここで, 以下の性質を持つ列 $g_i \in G$ を帰納的に構成しよう.

$$k \in \mathbb{N}, f_k := \sum_{i=1}^{k} \frac{g_i}{k} \Rightarrow e_k^2 := \|f - f_n\|_X^2 \leq \frac{c}{k}.$$

このようなものが構成できれば定理が成立することは明らかである. さて, いま,

$$\begin{aligned}
\sum_{j=1}^{m} a_j \|f - h_j\|^2 &= \sum_{j=1}^{m} a_j \|f\|^2 - 2\langle f, \sum_{j=1}^{m} a_j h_j \rangle + \sum_{j=1}^{m} a_j \|h_j\|^2 \\
&= \|f\|^2 - 2\|f\|^2 + \sum_{j=1}^{m} a_j \|h_j\|^2 \\
&\leq -\|f\|^2 + \sum_{j=1}^{m} a_j s_G^2 \\
&= s_G^2 - \|f\|^2 = c
\end{aligned}$$

であるので, ある $j$ について $\|f - h_j\|^2 \leq c$ である. そこで, $g_1 = h_j$ とおく. 次に $g_1, \ldots, g_k \in G$ が上記条件を満たすとする. $g_{n+1} \in G$ を選ぶために $e_{n+1}^2$ を計算すると,

$$\begin{aligned}
e_{n+1}^2 = \|f - f_{n+1}\|^2 &= \|\frac{n}{n+1}(f - f_n) - \frac{1}{n+1}(f - g_{n+1})\|^2 \\
&= \frac{n^2}{(n+1)^2} e_n^2 + \frac{2n}{(n+1)^2} \langle f - f_n, f - g_{n+1} \rangle + \frac{1}{(n+1)^2} \|f - g_{n+1}\|^2
\end{aligned}$$

となる. 一方,

$$\begin{aligned}
&\sum_{j=1}^{m} a_j \left( \frac{2n}{(n+1)^2} \langle f - f_n, f - h_j \rangle + \frac{1}{(n+1)^2} \|f - h_j\|^2 \right) \\
&= \frac{1}{(n+1)^2} \left( 2n \langle f - f_n, \sum_{j=1}^{m} a_j (f - h_j) \rangle + \sum_{j=1}^{m} a_j \|f - h_j\|^2 \right) \\
&= \frac{1}{(n+1)^2} \left( 2n \langle f - f_n, f - f \rangle + \sum_{j=1}^{m} a_j \|f - h_j\|^2 \right) \\
&\leq \frac{c}{(n+1)^2}
\end{aligned}$$

であるから, ある $j$ について

$$\frac{2n}{(n+1)^2} \langle f - f_n, f - h_j \rangle + \frac{1}{(n+1)^2} \|f - h_j\|^2 \leq \frac{c}{(n+1)^2}$$

である．そこで $g_{n+1} = h_j$ とおけば,

$$\begin{aligned} e_{n+1}^2 &= \frac{n^2}{(n+1)^2} e_n^2 + \frac{2n}{(n+1)^2} \langle f - f_n, f - g_{n+1} \rangle + \frac{1}{(n+1)^2} \|f - g_{n+1}\|^2 \\ &\leq \frac{n^2}{(n+1)^2} \frac{c}{n} + \frac{c}{(n+1)^2} = \frac{c}{n+1} \end{aligned}$$

となる．これで示せた． □

次に，この結果の $L^p$ 空間における類似である Darken-Donahue-Gurvits-Sontag による評価を述べる．証明には次の不等式を使う．

**定理 7.2.3 (Clarkson の不等式).**
$(X, \mathcal{M}, \mu)$ を測度空間とする．$p \in (1, \infty), q = p/(p-1), a = \min\{p, q\}$ とすると，任意の $f, g \in L^p(X, \mathcal{M}, \mu)$ に対して,

$$\|f + g\|_{L^p(X)}^a + \|f - g\|_{L^p(X)}^a \leq 2(\|f\|_{L^p(X)}^a + \|g\|_{L^p(X)}^a).$$

**証明.** 付録参照. □

**定義 7.2.4.** $X$ を $\mathbb{R}$ 上の線形空間とするとき, $n \in \mathbb{N}$ と部分集合 $G \subset X$ に対して,

$$\mathrm{span}_n(G) := \left\{ \sum_{i=1}^n a_i g_i \mid a_i \in \mathbb{R},\ g_i \in G \right\}$$

と定める．

**定理 7.2.5 (Darken-Donahue-Gurvits-Sontag 1993 [4]).**
$(X, \mathcal{M}, \mu)$ を測度空間とする．$G \subset L^p(X, \mathcal{M}, \mu)$, $p \in (1, \infty)$ とし, $f \in \mathrm{cl}(\mathrm{conv} G)$ とする．このとき, $r > 0$ が存在して, $G \subset B(f; r) = \{h \in L^p(X, \mathcal{M}, \mu) \mid \|f - h\|_{L^p(X)}\}$ となるならば, 任意の整数 $n \geq 1$ について,

$$\|f - \mathrm{span}_n G\|_{L^p(X)} \leq \frac{2^{1/a} r}{n^{1/b}}$$

となる．ここで, $q := p/(p-1), a := \min(p, q), b := \max(p, q)$ である．

**証明.** $L^p(X, \mathcal{M}, \mu) \ni h \mapsto \|h - \mathrm{span} G\|_{L^p(X)} \in \mathbb{R}$ は連続であるから, $f \in \mathrm{conv} G$ の場合を証明すれば十分である．$f = \sum_{j=1}^m w_j h_j$ と表す．このとき, $G$ の元の列 $(g_i)$ であって, $f_n = \sum_{i=1}^n g_i / n$ とおくと,

$$e_n := \|f - f_n\|_{L^p(X)} \leq \frac{2^{1/a} r}{n^{1/b}}$$

となるものが存在することを帰納法により示す．まず, 任意に $g \in G$ をとり $g_1 = g$ とする．すると, $G \subset B(f; r)$ および $a > 0$ より $\|f - g_1\|_{L^p(X)} \subset r \leq 2^{1/a} r$ であ

る．$g_1, \ldots, g_n \in G$ が得られているとする．$f - f_n = 0$ ならば $g_i = g_n \quad (i \geq n+1)$ とすればよい．$f - f_n \neq 0$ ならば Hahn-Banach の拡張定理より線形汎関数 $\Pi_n : L^p(X, \mathcal{M}, \mu) \to \mathbb{R}$ で $\|\Pi_n\| = 1$ かつ $\Pi_n(f - f_n) = \|f - f_n\|_X$ なるものが取れる (付録の系 9.4.6). $\sum_{j=1}^m w_j(f - h_j) = 0$ であるから,

$$0 = \Pi_n \left( \sum_{j=1}^m w_j(f - h_j) \right) = \sum_{j=1}^m w_j \Pi_n(f - h_j)$$

である．したがって, ある $j$ について $\Pi_n(f - h_j) \leq 0$ である．このような $j$ をひとつとり $g_{n+1} = h_j$ とおく．

$$f_{n+1} = \frac{1}{n+1} \sum_{i=1}^{n+1} g_i = \frac{n}{n+1} f_n + \frac{1}{n+1} g_{n+1}$$

であるから, Clarkson の不等式より,

$$\begin{aligned}
(e_{n+1})^a = \|f - f_{n+1}\|_{L^p(X)}^a &= \left\| \frac{n}{n+1}(f - f_n) + \frac{1}{n+1}(f - g_{n+1}) \right\|_{L^p(X)}^a \\
&= 2 \left( \left\| \frac{n}{n+1}(f - f_n) \right\|_{L^p(X)}^a + \left\| \frac{1}{n+1}(f - g_{n+1} \right\|_{L^p(X)}^a \right) \\
&\quad - \left\| \frac{n}{n+1}(f - f_n) - \frac{1}{n+1}(f - g_{n+1}) \right\|_{L^p(X)}^a
\end{aligned}$$

となる．一方, いま, $\|\Pi_n\| = 1$ かつ $\Pi_n(f - g_{n+1}) \leq 0$ であるから,

$$\begin{aligned}
&\left\| \frac{n}{n+1}(f - f_n) - \frac{1}{n+1}(f - g_{n+1}) \right\|_{L^p(X)} \\
&\geq \left| \Pi_n \left( \frac{n}{n+1}(f - f_n) - \frac{1}{n+1}(f - g_{n+1}) \right) \right| \\
&= \left| \frac{n}{n+1} \|f - f_n\|_{L^p(X)} - \frac{1}{n+1} \Pi_n(f - g_{n+1}) \right| \\
&\geq \frac{n}{n+1} \|f - f_n\|_{L^p(X)}
\end{aligned}$$

である．よって, このことと $g_{n+1} \in G$ および帰納法の仮定より,

$$
\begin{aligned}
(e_{n+1})^a &\leq 2\left(\left\|\frac{n}{n+1}(f-f_n)\right\|_{L^p(X)}^a + \left\|\frac{1}{n+1}(f-g_{n+1}\right\|_{L^p(X)}^a\right) \\
&\quad - \left(\frac{n}{n+1}\|f-f_n\|_{L^p(X)}\right)^a \\
&= \left(\frac{n}{n+1}\right)^a \|f-f_n\|_{L^p(X)}^a + \frac{2}{(n+1)^a}\|f-g_{n+1}\|_{L^p(X)}^a \\
&= \left(\frac{n}{n+1}\right)^a (e_n)^a + \frac{2}{(n+1)^a}\|f-g_{n+1}\|_{L^p(X)}^a \\
&\leq \left(\frac{n}{n+1}\right)^a \left(\frac{2^{1/a}r}{n^{1/b}}\right)^a + \frac{2r^a}{(n+1)^a} = \frac{2r^a}{(n+1)^a}\left(1+\frac{n^a}{n^{a/b}}\right)
\end{aligned}
$$

となる．$1/a+1/b=1$ より $a-a/b=1$ であるから,

$$
(e_{n+1})^a \leq \frac{2r^a}{(n+1)^a}(1+n^{a-a/b}) = \frac{2r^a}{(n+1)^{a-1}} = \frac{2r^a}{(n+1)^{a/b}}
$$

である．これで示せた． □

以上のふたつの結果はいずれも $f \in \mathrm{cl}(\mathrm{conv}G)$ に対する評価であり, 一般の $f \in X$ については適用されない．そこで $X$ の元すべてに対して適用できるような評価を与えよう．そのために次の概念を導入する．

**定義 7.2.6 ($G$-variation).**
$X$ をノルム空間とする．空でない $G \subset X$ と $f \in X$ に対して,

$$
\|f\|_G := \inf\{c > 0 \mid f/c \in \mathrm{cl}(\mathrm{conv}(G \cup -G))\}
$$

と定義する．ここで, $-G = \{-g \mid g \in G\}$ である．$\|f\|_G$ を $f$ の $G$-variation という．

**注意 7.2.7.** $G \subset X$ を空でない有界集合とし, $s_G := \inf_{g\in G}\|g\|_X$ とおくと, $\|f\|_X \leq s_G\|f\|_G$ となる．実際, $b > 0$ で $f/b \in \mathrm{cl}(\mathrm{conv}(G \cup -G))$ なるものを任意に取ると, $f/b = \lim_{n\to\infty} h_n$ なる $h_n \in \mathrm{conv}(G \cup -G)$ が取れる．すると, $\|h_n\|_X \leq s_G$ であるので $n \to \infty$ として $\|f\|_X \leq bs_G$ となる．よって, $b$ の任意性から $\|f\|_X \leq s_G\|f\|_G$ である．

**補題 7.2.8.** $X$ を $\mathbb{R}$ 上のノルム空間とし, $C \subset X$ は $0$ を含む凸集合とし, $m$ を $C$ の Minkowski 汎関数, つまり $x \in X$ に対して $m(x) := \inf\{c > 0 \mid x/c \in C\}$ とすると, 任意の $x, y \in X,\ \lambda \geq 0$ に対して $m(\lambda x) = \lambda m(x)$, $m(x+y) \leq m(x)+m(y)$ であり,

$$
\{x \in X \mid m(x) < 1\} \subset C
$$

となる．さらに $-C = C$ であるならば, $\lambda \in \mathbb{R}$ に対して $m(\lambda x) = |\lambda| m(x)$ となる．特に, $G$-variation は以上で述べた性質をすべて持ち,

$$\{f \in X \mid \|f\|_G < 1\} \subset \mathrm{cl}(\mathrm{conv}(G \cup -G))$$

となる．

**証明.** まず $\lambda > 0$ に対して,

$$\begin{aligned}\lambda m(x) &= \inf\{\lambda c \mid c > 0, x/c \in C\} \\ &= \inf\{\lambda c \mid (\lambda c) > 0, \lambda x/(\lambda c) \in C\} \\ &= \inf\{b \mid b > 0, \lambda x/b \in X\} = m(\lambda x)\end{aligned}$$

である．また, $\lambda = 0$ のときは $m(0) = 0$ より $m(\lambda x) = \lambda m(x)$ である．そして, $m(x) = +\infty$ または $m(y) = +\infty$ の場合は明らかに $m(x+y) \leq m(x) + m(y)$ であり, そうでない場合は $x/c, x/b \in C$ なる任意の $c, b \in (0, \infty)$ に対し, $C$ が凸集合であることから,

$$\frac{1}{c+b}(x+y) = \frac{c}{c+b}\frac{1}{c}x + \frac{b}{c+b}\frac{1}{b}y \in C$$

である．よって $m(x+y) \leq c+b$ であり, $c, b$ の任意性より $m(x+y) \leq m(x)+m(y)$ となる．次に $m(x) < 1$ ならば $m$ の定義より $c \in (0,1)$ で $x/c \in C$ なるものが取れる．すると, $0 \in C$ と $C$ の凸性より,

$$x = c\frac{x}{c} + (1-c)0 \in C$$

である．最後に, $-C = C$ ならば $\lambda \in \mathbb{R}$ と $b > 0$ に対して,

$$\frac{\lambda x}{b} \in C \Longleftrightarrow \frac{|\lambda| x}{b} \in C$$

であるので $m(\lambda x) = |\lambda| m(x)$ となる． □

**注意 7.2.9.** Minkowski 汎関数という言葉を用いれば $G$-variation は $\mathrm{cl}(\mathrm{conv}(G \cup -G))$ の Minkowski 汎関数である．そして上の補題より $G$-variation が有限な $X$ の元全体の集合は $X$ の部分空間をなし, $G$-variation はその上の半ノルムとなる．また $G$ が $X$ のノルムに関して有界ならその上のノルムとなる．

**補題 7.2.10.** $X$ をノルム空間とし, $\emptyset \neq G \subset X$ とする．$c \geq 0$ に対して $G(c) := \{wg \in \mid |w| \leq c, g \in G\}$ とおくと, $\|f\|_G < \infty$ なる $f \in X$ に対して,

$$f \in \mathrm{cl}(\mathrm{conv}(G(\|f\|_G)))$$

となる．

**証明.** すぐ上の補題より, 任意の $\varepsilon > 0$ に対して,

$$\frac{f}{\|f\|_V + \varepsilon} \in \mathrm{cl}(\mathrm{conv}(G \cup -G))$$

である. ゆえに, $\varepsilon \to 0$ とすると $f/\|f\|_G \in \mathrm{cl}(\mathrm{conv}(G \cup -G))$ である. よって,

$$f \in \|f\|_G \mathrm{cl}(\mathrm{conv}(G \cup -G)) = \mathrm{cl}\left(\|f\|_G \mathrm{conv}(G \cup -G)\right) \subset \mathrm{cl}(\mathrm{conv} G(\|f\|_G)).$$

□

**定理 7.2.11.** $X$ をノルム空間とする. また, $G \subset X$ を空でない有界集合とし, $s_G := \sup_{g \in G} \|g\|_X$ とおく. このとき, $f \in X$ と整数 $n \geq 1$ に対して,

(1) $X$ が内積空間ならば,

$$\|f - \mathrm{span}_n G\|_X \leq \sqrt{\frac{s_G^2 \|f\|_G^2 - \|f\|_X^2}{n}}$$

となる. 特に, $B_R(\|\cdot\|_G) := \{g \in X \mid\mid \|g\|_G \leq R\}$ に対して,

$$E(B_R(\|\cdot\|_G), \mathrm{span}_n G, X) \leq s_G \frac{R}{n^{1/2}}.$$

(2) 測度空間 $(Y, \mathcal{N}, \nu)$ と $p \in (1, \infty)$ に対して $X = L^p(Y, \mathcal{N}, \nu)$ であるとき,

$$\|f - \mathrm{span}_n G\|_X \leq \frac{2^{1+1/a} s_G \|f\|_G}{n^{1/b}}$$

となる. ここで, $a := \min(p, p/(p-1)), b := \max(p, p/(p-1))$ である. 特に,

$$E(B_R(\|\cdot\|_G), \mathrm{span}_n G, L^p(Y, \nu)) \leq 2^{1+1/a} s_G \frac{R}{n^{1/b}}.$$

**証明.** $c \geq 0$ に対して $G(c) := \{wg \in \mid |w| \leq c,\ g \in G\}$ とおく. このとき, $\mathrm{conv}_n G(c) \subset \mathrm{span}_n G(c) = \mathrm{span}_n G$ である. したがって, $f \in X$ に対して,

$$\|f - \mathrm{span}_n G\|_X \leq \|f - \mathrm{conv}_n G(\|f\|_G)\|_X$$

である. 一方, すぐ上の補題より, $\|f\|_G < \infty$ のとき $f \in \mathrm{cl}(\mathrm{conv}(G(\|f\|_G)))$ である. また, $\|f\|_X \leq s_G \|f\|_G$ より $G(\|f\|_G) \subset B(f, 2s_G\|f\|_G)$ である. よって, 定理 7.2.2 と定理 7.2.5 より主張は成り立つ. □

以上の結果は直接的にニューラルネットによる近似レートを述べたものではないが, 活性化関数 $\Psi : \mathbb{R} \to \mathbb{R}$ および $W \subset \mathbb{R}^r, B \subset \mathbb{R}$ に対して $G = G_r(\Psi, W, B) := \{x \mapsto \Psi(w^{\mathrm{T}} x + b) \mid w \in W, b \in B\}$ とおくとき, $\Psi, W, B$ や問題とする関数空

間に適当な条件を加えれば定理の仮定 ($G$ が有界) をみたすようにできる．また, $b_j \in B, c_j \in \mathbb{R}, w_j \in W$ に対して,

$$f(x) = \sum_{j=1}^{m} c_j \Psi(w_j^{\mathrm{T}} x + b_j) \quad (x \in \mathbb{R}^r)$$

とおくと,

$$\begin{aligned} f(x) &= \sum_{c_j > 0} c_j \Psi(w_j^{\mathrm{T}} x + b_j) + \sum_{c_j \le 0} c_j \Psi(w_j^{\mathrm{T}} x + b_j) \\ &= \sum_{c_j > 0} |c_j| \Psi(w_j^{\mathrm{T}} x + b_j) + \sum_{c_j \le 0} |c_j| (-\Psi(w_j^{\mathrm{T}} x + b_j)) \end{aligned}$$

であるので,

$$\|f\|_G = \inf\{c > 0 \mid f/c \in \mathrm{cl}(\mathrm{conv}(G \cup -G))\} \le \sum_{j=1}^{m} |c_j|$$

であることを注意しておく．

次に近似誤差の下限に関する定理を紹介しよう．

**定理 7.2.12 (Makovoz 1996 [24] theorem4).**
sigmoidal 関数 $\Psi : \mathbb{R} \to \mathbb{R}$ は $\Psi = \chi_{[0,\infty)}$ であるか, Lipschitz 連続且つある $\eta, C > 0$ が存在して

$$|\Psi(t) - \chi_{[0,\infty)}(t)| \le C|t|^{-\eta} \quad (\forall t \in \mathbb{R} \setminus \{0\})$$

となるとする．このとき, $G := G_r(\Psi, \mathbb{R}^r, \mathbb{R})$ とおけば, 開凸集合 $D \subset \mathbb{R}^r$ と $\xi > 0$ に対して定数 $C(D, \xi) > 0$ が存在して, 任意の整数 $n \ge 1$ に対して

$$E(B_1(\|\cdot\|_G), \mathrm{span}_n G, L^2(D)) \ge C(D, \xi) n^{-1/2-1/r-\xi}.$$

さて, 以上の結果は冪レートの結果であった．最後に指数レートの結果を述べよう．

**定理 7.2.13 (Kůrková-Sanguineti 2008 [18]).**
$X$ を Hilbert 空間とする．また, $G \subset X$ を空でない有界集合とし, $s_G = \sup_{g \in G} \|g\|_X$ とおく．このとき, 任意の $f \in \mathrm{conv} G$ に対して, $\tau_f \in [0, 1)$ が存在して, 任意の整数 $n \ge 1$ について

$$\|f - \mathrm{conv}_n G\|_X \le \sqrt{\tau_f^{n-1}(s_G^2 - \|f\|_X^2)}$$

が成り立つ．

**証明.** $\sum_{j=1}^{m} a_j = 1,\ a_j > 0,\ g_j \in G$ により $f = \sum_{j=1}^{m} a_j g_j$ と表す．$m = 1$ なら $\tau_f = 0$ として成立するので $m > 1$ としてよい．また, ある $j$ について $f = g_j$

となる場合も除いてよい．$G' := \{g_1, \ldots, g_m\}$ とおく．各 $n = 1, \ldots, m$ に対して $f_n \in \mathrm{conv}_n G', \rho_n \in (0,1)$ が存在して

$$\|f - f_n\|_X^2 \leq (1 - \rho_n^2)^{n-1}(s_G^2 - \|f\|_X^2)$$

が成り立つことを $n$ に関する帰納法で示す．まず, $g_{j_1} \in G'$ を

$$\|f - g_{j_1}\|_X = \min_{g \in G'} \|f - g\|_X$$

なるものとし $f = g_{j_1}$ とおけば,

$$\begin{aligned}
\|f - f_1\|_X^2 = \sum_{j=1}^m a_j \|f - f_1\|_X^2 &\leq \sum_{j=1}^m a_j \|f - g_j\|_X^2 \\
&= \|f\|_X^2 - 2\langle f, \sum_{j=1}^m a_j g_j \rangle + \sum_{j=1}^m a_j \|g_j\|_X^2 \\
&= \|f\|_X^2 - 2\|f\|_X^2 + \sum_{j=1}^m a_j \|g_j\|_X^2 \\
&\leq s_G^2 - \|f\|_X^2
\end{aligned}$$

となる．よって, $\rho_1 \in (0,1)$ を任意に取れば $n = 1$ の場合は成り立つ．次に $n-1$ で成り立つとする．$f_{n-1} = f$ ならば $f_n = f_{n-1}, \rho_n = \rho_{n-1}$ と取ればよいので $f_{n-1} \neq f$ とする．このとき,

$$\sum_{j=1}^m a_j \langle f - f_{n-1}, f - g_j \rangle = \langle f - f_{n-1}, f - \sum_{j=1}^m a_j g_j \rangle = \langle f - f_{n-1}, f - f \rangle = 0$$

であるので, 次のいずれかが成り立つ ($m > 1$ に注意).

(1) $\exists g \in G', \langle f - f_{n-1}, f - g \rangle < 0.$

(2) $\forall g \in G', \langle f - f_{n-1}, f - g \rangle < 0.$

(2) はあり得ないことを示そう．$f_{n-1} \in \mathrm{conv}_{n-1} G'$ であるので,

$$f_{n-1} = \sum_{k=1}^{n-1} b_k g_{j_k} \quad (\sum_{k=1}^{n-1} b_k = 1,\ b_k > 0, g_{j_1} \in G')$$

と表される．従って, (2) が成り立つならば,

$$\begin{aligned}
\|f - f_{n-1}\|_X^2 &= \langle f - f_{n-1}, f - \sum_{k=1}^{n-1} b_k g_{j_k} \rangle \\
&= \sum_{k=1}^{n-1} b_k \langle f - f_{n-1}, f - g_{j_k} \rangle = 0
\end{aligned}$$

であるので $f = f_{n-1}$ となる．これは $f \neq f_{n-1}$ に反する．よって, (2) はあり得ない．ゆえに (1) が成り立つ．そこで $g_{j_n} \in G'$ を $\langle f - f_{n-1}, f - g_{j_n} \rangle < 0$ を満たすものとする．そして, $f \neq f_{n-1}, g_{j_n}$ に注意して

$$\rho_n := \min \left\{ \rho_{n-1}, -\frac{\langle f - f_{n-1}, f - g_{j_n} \rangle}{\|f - f_{n-1}\|_X \|f - g_{j_n}\|_X} \right\} \in (0, 1)$$

とおく．また, $\alpha_n \in [0, 1]$ に対して

$$f_n = \alpha_n f_{n-1} + (1 - \alpha_n) g_{j_n} \in \mathrm{conv}_n G'$$

とおく．以下, $\alpha_n$ をうまく取ることで $f_n, \rho_n$ が条件を満たすようにできることを示す．$q_n = -\langle f - f_{n-1}, f - g_{j_n} \rangle > 0$, $r_n = \|f - g_{j_n}\|_X > 0$ とおき, 各 $k = 1, \ldots, n$ について $e_k := \|f - f_k\|_X$ とおくと,

$$\begin{aligned}
e_n^2 &= \|\alpha_n (f - f_{n-1}) + (1 - \alpha_n)(f - g_{j_n})\|_X^2 \\
&= \alpha_n^2 e_{n-1}^2 + 2\alpha_n (1 - \alpha_n) \langle f - f_{n-1}, f - g_{j_n} \rangle + (1 - \alpha_n)^2 \|f - g_{j_n}\|_X^2 \\
&= \alpha_n^2 e_{n-1}^2 - 2\alpha_n (1 - \alpha_n) q_n + (1 - \alpha_n)^2 r_n^2 \\
&= \alpha_n^2 (e_{n-1}^2 + 2q_n + r_n^2) - 2\alpha_n (q_n + r_n^2) + r_n^2
\end{aligned}$$

となる．従って,

$$\frac{de_n^2}{d\alpha_n} = 2\alpha_n (e_{n-1}^2 + 2q_n + r_n^2) - 2(q_n + r_n^2)$$

である．よって, $e_{n-1}^2 + 2q_n + r_n^2 > 0$ より $e_n^2$ は

$$\alpha_n = \frac{q_n + r_n^2}{e_{n-1}^2 + 2q_n + r_n^2} \in [0, 1]$$

で最小値をとる．この $\alpha_n$ が求めるものである．実際, そのとき

$$\begin{aligned}
e_n^2 &= \frac{(q_n + r_n^2)^2}{e_{n-1}^2 + 2q_n + r_n^2} - 2\frac{(q_n + r_n^2)(q_n + r_n^2)}{e_{n-1}^2 + 2q_n + r_n^2} + r_n^2 \\
&= \frac{-q_n^2 - 2q_n r_n^2 - r_n^4 + e_{n-1}^2 r_n^2 + 2q_n r_n^2 + r_n^4}{e_{n-1}^2 + 2q_n + r_n^2} \\
&= \frac{-q_n^2 + e_{n-1}^2 r_n^2}{e_{n-1}^2 + 2q_n + r_n^2}
\end{aligned}$$

であり, $\rho_n = \min \{\rho_{n-1}, q_n / (e_{n-1} r_n)\}$ より $\rho_n e_{n-1} r_n \leq q_n$ であるので,

$$\begin{aligned}
e_n^2 &\leq \frac{-\rho_n^2 e_{n-1}^2 r_n^2 + e_{n-1}^2 r_n^2}{e_{n-1}^2 + 2q_n + r_n^2} < \frac{(1 - \rho_n^2) e_{n-1}^2 r_n^2}{r_n^2} = (1 - \rho_n^2) e_{n-1}^2 \\
&\leq (1 - \rho_n^2)(1 - \rho_{n-1}^2)^{n-2} (s_G^2 - \|f\|_X^2) \leq (1 - \rho_n^2)^{n-1} (s_G^2 - \|f\|_X^2)
\end{aligned}$$

となる (最後の不等号は $\rho_n \leq \rho_{n-1}$ による). よって $n$ で成立する.
さて, $\rho_f := \min\{\rho_n \mid n = 1, \ldots, m\}$ とおく. すると, $n = 1, \ldots, m$ に対して

$$\|f - \mathrm{conv}_n G\|_X^2 \leq \|f - f_n\|_X^2 \leq (1-\rho_n)^{n-1}(s_G^2 - \|f\|_X^2) \leq (1-\rho_f)^{n-1}(s_G^2 - \|f\|_X^2)$$

となり, $n > m$ に対しては $\|f - \mathrm{conv}_n G\|_X = 0$ となるので $\tau_f = 1 - \rho_f^2$ が求める定数である. □

# 8 まとめと今後の課題

### まとめ

活性化関数が非多項式かつ局所有界で Riemann 積分可能の場合に feedforward 型ニューラルネットワークが万能近似能力を持つなど, ニューラルネットワークの万能近似定理について, 複数の結果を記号を統一して解説した. また, その近似レートの問題において, 近似する空間として Sobolev 空間を取った場合を扱った. さらに, その場合には次元の呪いという, 所望の精度を達成するために必要な中間ユニット数が入力の次元に関して指数関数的に増大する現象が発生することを説明した. そして, 次元の呪いから解放される空間として Barron space の概念を導入し, Barron space のひとつとして「G-variation に関して有界な空間」を紹介した.

### 今後の課題

本論文で触れることができなかったことに, Deep narrow network の万能近似定理 [16] や深層ニューラルネットワークの近似レート (Yarotsky 2016 [35] など) がある. また, Barron space に関しては, Barron 自身の結果やごく最近の Weinan et al.[6], Siegel, Xu[31] の結果をまとめることができなかった. 今後は, こうした深層ニューラルネットワークや Barron space の近年の進展をまとめたい. さらに, 近年注目を集めている, 日本の研究者が創始したニューラルネットワークの積分表現理論 (Murata 1996 [25] や Sonoda, Murata 2015 [32] など) の解説を試みたい.

# 謝辞

当初, 論理や計算に興味があるということで研究室に入ったのに拘らず, 結局それと関係のないテーマで修士論文を執筆しました. そんな興味が二転三転するような私を辛抱強くご指導いただき, 自由に研究させてくださった指導教官の溝口佳寛教授に深く感謝いたします. また, 本論文のいくつかの命題の証明の完成にご協力頂いた, 片岡佑太さん, 九州大学大学院数理学研究院の高田了准教授, 慶應義塾大学経済学部の服部哲弥教授に厚く御礼申し上げます. 最後になりますが, 同期の友人たちをはじめ, 論文作成にご協力いただいた全ての方々に感謝の意を表します.

# 9 付録

## 9.1 測度の基本性質

本節では, 測度の基本性質を証明する. まず, 可測空間, 測度の定義を述べよう.

**定義 9.1.1 (可測空間).**
集合 $X$ と $X$ の部分集合系 $\mathcal{M}$ の組 $(X, \mathcal{M})$ が可測空間であるとは,

(1) $\emptyset \in \mathcal{M}$.

(2) $\forall E \in \mathcal{M}, X \setminus E \in \mathcal{M}$.

(3) 任意の $\mathcal{M}$ の元の列 $(E_n)_{n\in\mathbb{N}}$ について $\bigcup_{n\in\mathbb{N}} E_n \in \mathcal{M}$.

が成り立つことをいう.

**注意 9.1.2.** $(X, \mathcal{M})$ を可測空間とすると, $\mathcal{M}$ は可算個までの共通部分をとる操作に関して閉じている. 実際, $(E_n)_{n\in\mathbb{N}}$ を $\mathcal{M}$ の元の列とすると,

$$\bigcup_{n=1}^{\infty} E_n = X \setminus \left( \bigcup_{n=1}^{\infty} (X \setminus E_n) \right) \in \mathcal{M}.$$

**定義 9.1.3 (測度, 測度空間).**
$(X, \mathcal{M})$ を可測空間とするとき, 写像 $\mu : \mathcal{M} \to [0, \infty]$ が $(X, \mathcal{M})$ 上の測度であるとは,

(1) $\mu(\emptyset) = 0$.

(2) $\mathcal{M}$ の元の列 $(E_n)_{n\in\mathbb{N}}$ が $E_n \cap E_m = \emptyset$ $(n \neq m)$ を満たすならば,

$$\mu\left( \bigcup_{n\in\mathbb{N}} E_n \right) = \sum_{n=1}^{\infty} \mu(E_n).$$

が成り立つことをいう. このとき, 三つ組 $(X, \mathcal{M}, \mu)$ を測度空間という.

**補題 9.1.4.** $(X, \mathcal{M}, \mu)$ を測度空間とするとき, 以下が成り立つ.

(1) $A, B \in \mathcal{M}$ が $B \subset A$ を満たすとき, $\mu(B) \leq \mu(A)$.

(2) $A, B \in \mathcal{M}$ が $B \subset A$ を満たすとき $\mu(B) < \infty$ ならば $\mu(A \setminus B) = \mu(A) - \mu(B)$.

(3) $\mathcal{M}$ の元の列 $(A_n)_{n=1}^{\infty}$ に対して,

$$\mu\left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \mu(A_n).$$

(4) $\mathcal{M}$ の元の列 $(A_n)_{n=1}^{\infty}$ が $A_n \subset A_{n+1}$ $(\forall n)$ を満たすならば,

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right).$$

(5) $\mathcal{M}$ の元の列 $(A_n)_{n=1}^{\infty}$ が $A_n \supset A_{n+1}$ $(\forall n)$ を満たすとき $\mu(A_1) < \infty$ ならば

$$\lim_{n\to\infty} \mu(A_n) = \mu\left(\bigcap_{n=1}^{\infty} A_n\right).$$

(1) を測度の単調性, (3) を測度の劣加法性, (4), (5) を測度の連続性という.

**証明.**
(1): $\mu(A) = \mu((A \setminus B) \cup B) = \mu(A \setminus B) + \mu(B) \geq \mu(B)$.
(2): (1) の証明において $\mu(B)$ を両辺から引けばよい.
(3): $B_1 = A_1$ とし, $B_{n+1} = A_{n+1} \setminus \bigcup_{j=1}^{n} A_j$ とする. このとき,

$$\bigcup_{n=1}^{\infty} B_n = \bigcup_{n=1}^{\infty} A_n$$

であり, $n \neq m$ ならば $B_n \cap B_m = \emptyset$ である. したがって,

$$\begin{aligned}
\mu\left(\bigcup_{n=1}^{\infty} A_n\right) &= \mu\left(\bigcup_{n=1}^{\infty} B_n\right) \\
&= \sum_{n=1}^{\infty} \mu(B_n) \\
&\leq \sum_{n=1}^{\infty} \mu(A_n) \quad (\because (1)).
\end{aligned}$$

(4):$A_0 = \emptyset$ とし, $B_n := A_n - A_{n-1}$ とおく. すると, $B_{\in}\mathcal{M}$ で, $n \neq m$ に対して $B_n \cap B_m = \emptyset$ となる. さらに, $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} B_n$ である. したがって, 測度の完全加法性より,

$$\begin{aligned}
\mu(\bigcup_{n=1}^{\infty} A_n) &= \mu(\bigcup_{n=1}^{\infty} B_n) \\
&= \sum_{n=1}^{\infty} \mu(B_n) \\
&= \sum_{n=1}^{\infty} (\mu(A_n) - \mu(A_{n-1})) \quad (\because (2)) \\
&= \lim_{n\to\infty} (\mu(A_n) - \mu(A_0)) \\
&= \lim_{n\to\infty} \mu(A_n).
\end{aligned}$$

(5): $(\mu(A_n))_{n=1}^{\infty}$ は非負な単調減少列なので極限が存在すること, また, $\mu(A_1) < \infty$ より $\mu(\bigcap_{j=1}^{\infty} A_j) \leq \mu(A_n) < \infty$ $(\forall n)$ であることに注意する. いま, $C_n := X \setminus A_n$ とおくと, $C_n \in \mathcal{M}$ であり, $C_n \subset C_{n+1}$ $(\forall n)$ であるから,

$$\begin{aligned}
\mu(X) - \mu\left(\bigcap_{n=1}^{\infty} A_n\right) &= \mu\left(X \setminus \bigcap_{n=1}^{\infty} A_n\right) \quad (\because (2)) \\
&= \mu\left(\bigcup_{n=1}^{\infty} X \setminus A_n\right) \\
&= \lim_{n\to\infty} \mu(X \setminus A_n) \quad (\because (4)) \\
&= \lim_{n\to\infty} (\mu(X) - \mu(A_n)) \quad (\because (2)) \\
&= \mu(X) - \lim_{n\to\infty} \mu(A_n)
\end{aligned}$$

となる. よって, 求める等式を得る. □

## 9.2 単関数近似定理

本節では非負値可測関数が単調増加な非負単関数列で近似できることを示す. ここで可測空間 $(X, \mathcal{M})$ 上の単関数とは,

$$f = \sum_{j=1}^{m} c_j \chi_{E_j} \quad (m \in \mathbb{N}, c_j \in \mathbb{R}, E_j \in \mathcal{M})$$

と表される関数のことをいう. ただし $\chi_E$ で $E$ の定義関数を表す.

**定理 9.2.1 (単関数近似定理).**
$f \geq 0$ を可測空間 $(X, \mathcal{M})$ 上の可測関数とする. このとき, $n = 1, 2, \ldots$ に対して,

$$\begin{aligned}
&A_k^n := f^{-1}([(k-1)/2^n, k/2^n)) \quad (k = 1, \ldots, 2^n n), \\
&A_0^n := f^{-1}([n, \infty)), \\
&\varphi_n := \sum_{k=1}^{2^n n} \frac{k-1}{2^n} \chi_{A_k^n} + n\chi_{A_0^n}
\end{aligned}$$

と定義すると, 任意の $x \in X$ に対して,

$$\begin{aligned}
&0 \leq \varphi_1(x) \leq \varphi_2(x) \leq \cdots \leq \varphi_n(x) \leq \cdots \leq f(x), \\
&f(x) = \lim_{n\to\infty} \varphi_n(x)
\end{aligned}$$

が成り立つ.

**証明.** 任意に $x \in X$ を取る．また, 任意に $n$ を取る．このとき, $\varphi_n(x) \leq \varphi_{n+1}(x)$ を示そう．まず $f(x) \geq n+1$ の場合は, $\varphi_n(x) = n \leq n+1 = \varphi_{n+1}(x)$ で成立する．次に $f(x) < n$ のときは, $f(x) \in [(k-1)/2^n, k/2^n)$ なる $k \in \{1, \ldots, 2^n n\}$ が取れる．すると,

$$\frac{2k-2)}{2^{n+1}} \leq f(x) \leq \frac{2k-1}{2^{n+1}}$$

または,

$$\frac{2k-1)}{2^{n+1}} \leq f(x) \leq \frac{2k}{2^{n+1}}$$

である．前者なら $\varphi_n(x) = (k-1)/2^n = 2(k-1)/2^{n+1} = \varphi_{n+1}(x)$ で成立する．後者なら $\varphi_n(x) = (k-1)/2^n = 2(k-1)/2^{n+1} \leq (2k-1)/2^{n+1}$ で成立する．最後に $n \leq f(x) < n+1$ の場合は, $(k-1)/2^{n+1} \leq f(x) < k/2^{n+1}$ なる $k \in \{1 + 2^{n+1}n, \ldots, 2^{n+1}(n+1)\}$ が取れる．したがって,

$$\varphi_n(x) = n \leq (k-1)/2^{n+1} = \varphi_{n+1}(x)$$

となって成立する．次に $\varphi_n(x) \to f(x)$ を示す．任意の $\varepsilon > 0$ に対して, $f(x) > N$ かつ $1/2^N < \varepsilon$ なる $N$ が取れる．このとき, $n \geq n$ に対して,

$$\frac{k-1}{2^n} \leq f(x) < \frac{k}{2^n}$$

なる $k \in \{1, \ldots, 2^n n\}$ がとれ, $\varphi_n(x) = (k-1)/2^n$ となるので,

$$0 \leq f(x) - \varphi_n(x) \leq \frac{1}{2^n} < \frac{1}{2^N} < \varepsilon$$

となる．よって $\varphi_n(x) \to f(x)$ である．同時に $\varphi_n(x) \leq f(x)$ も示された． □

## 9.3 Lebesgue 積分の基本定理

本節では Lebesgue 積分論の基本定理を説明する．詳しい解説や証明は猪狩 [44] などを参照されたい．まず積分の定義を述べる．

**定義 9.3.1.** $(X, \mathcal{M}, \mu)$ を測度空間とする． 非負単関数 $f : X \to [0, \infty)$, つまり $a_1, \ldots, a_k \geq 0$ と $E_1, \ldots, E_k \in \mathcal{M}$ により $f = \sum_{j=1}^k a_j \chi_{E_j}$ と表される $f$ に対しては

$$\int_X f d\mu = \int_X f(x) d\mu(x) := \sum_{j=1}^k a_j \mu(E_j)$$

と定義する．非負の可測関数 $f : X \to [0, \infty)$ に対しては

$$\int_X f(x) d\mu(x) := \sup \left\{ \int_X \varphi d\mu \,\middle|\, 0 \leq \varphi \leq f, \varphi \text{ は単関数} \right\}$$

と定義する．可測関数 $f : X \to \mathbb{R}$ に対しては $\int_X |f(x)|d\mu(x) < \infty$ である場合に

$$\int_X f(x)d\mu(x) := \int_X f^+(x)d\mu(x) - \int_X f^-(x)d\mu(x)$$

と定義する．ここで $f^+ = \max(0, f),\ f^- = \max(0, -f)$ である．

次に収束定理について述べる．

**定理 9.3.2 (単調収束定理).**
$(\Omega, \mathcal{F}, \mu)$ を測度空間とする．$(f_n)$ は $(\Omega, \mathcal{F}, \mu)$ 上の非負値可測関数列で関数 $f : \Omega \to \mathbb{R}$ に対して

$$f_1(x) \leq f_2(x) \leq \cdots \leq f_n(x) \to f(x) \quad (\forall x \in \Omega)$$

を満たすとすると,

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu$$

が成り立つ．

**定理 9.3.3 (優収束定理).**
$(\Omega, \mathcal{F}, \mu)$ を測度空間とする．$(f_n)$ が $(\Omega, \mathcal{F}, \mu)$ 上の可測関数列で, 関数 $f$ に各点収束し, かつある $(\Omega, \mathcal{F}, \mu)$ 上の可積分関数 $g$ が存在して任意の $n, x$ に対して $\|f_n(x)\| \leq g(x)$ が成り立つならば, $f$ は可積分で,

$$\lim_{n\to\infty} \int_\Omega f_n d\mu = \int_\Omega f d\mu$$

が成り立つ．

**定理 9.3.4 (積分記号下の微分).**
$(\Omega, \mathcal{F}, \mu)$ を測度空間とし $a < b$ とする．$f : \Omega \times (a, b) \to \mathbb{R}$ は次の二条件を満たすとする．

(1) 各 $t \in (a, b)$ について $f(\cdot, t) : \Omega \ni x \mapsto f(t, x) \in R$ が $L^1(\Omega, \mathcal{F}, \mu)$ に属し, 各点 $(x, t)$ において $(\partial_t f)(x, t)$ が存在する．

(2) $0 \leq g \in L^1(\Omega, \mathcal{F}, \mu)$ が存在して $|(\partial_t f)(x, t)| \leq g(x) \quad (\forall x \in \Omega, t \in (a, b))$.

このとき, $F(t) := \int_\Omega f(x, t)d\mu(x)$ は $(a, b)$ 上で微分可能であって,

$$F'(t) = \int_\Omega \frac{\partial f}{\partial t}(x, t)d\mu(x).$$

**証明.** $t_0 \in (a, b)$ を固定し $t_0$ に収束する数列 $t_n \in (a, b),\ t_n \neq t_0$ を取り, $x \in \Omega$ に対して

$$h_n(x) := \frac{f(x, t_n) - f(x, t_0)}{t_n - t_0} \quad (n = 1, 2, \ldots, )$$

とおく．このとき各 $x$ について $(\partial_t f)(x,t_0) = \lim_{n\to\infty} h_n(x)$ であり，平均値の定理より

$$|h_n(x)| \leq \sup_{t\in(a,b)} \left|\frac{\partial f}{\partial t}(x,t)\right| \leq g(x)$$

である．よって，優収束定理より

$$\lim_{n\to\infty} \frac{F(t_n)-F(t_0)}{t_n-t_0} = \lim_{n\to\infty} \int_\Omega h_n(x)d\mu(x) = \int_\Omega \frac{\partial f}{\partial t}(x,t_0)d\mu(x).$$

ゆえに数列 $(t_n)$ の任意性から定理は示された． □

最後に Fubini の定理について説明しよう．

**定義 9.3.5 ($\sigma$-有限測度空間).**
測度空間 $(X,\mathcal{M},\mu)$ が $\sigma$-有限であるとは，$E_n \in \mathcal{M} \quad (n = 1,2,\ldots)$ が存在して $\mu(E_n) < \infty$ 且つ $X = \bigcup_{n\in\mathbb{N}} E_n$ となることをいう．

**定義 9.3.6 (直積可測空間).**
$(X,\mathcal{M}),(Y,\mathcal{N})$ を可測空間とする．$X\times Y$ 上の $\sigma$-加法族 $\mathcal{M}\otimes\mathcal{N}$ を次のように定める．

$$\mathcal{M}\otimes\mathcal{N} := \sigma\left[\{E\times F \mid E\in\mathcal{M}, F\in\mathcal{N}\}\right]$$

このとき，可測空間 $(X\times Y,\mathcal{M}\otimes\mathcal{N})$ を $(X,\mathcal{M}),(Y,\mathcal{N})$ の直積可測空間という．

**定義 9.3.7 (直積測度空間).**
$(X,\mathcal{M},\mu),(Y,\mathcal{N},\nu)$ を $\sigma$-有限な測度空間とする．このとき，直積可測空間 $(X\times Y,\mathcal{M}\otimes\mathcal{N})$ 上の測度 $\mu\otimes\nu$ で

$$(\mu\otimes\nu)(E\times F) = \mu(E)\nu(F) \quad (E\in\mathcal{M}, F\in\mathcal{N})$$

を満たすものが一意的に存在する．$\mu\otimes\nu$ を $\mu,\nu$ の直積測度といい，測度空間 $(X\times Y,\mathcal{M}\otimes\mathcal{N},\mu\otimes\nu)$ を $(X,\mathcal{M},\mu),(Y,\mathcal{N},\nu)$ の測度空間という．

**定理 9.3.8 (Tonelli の定理).**
$(X,\mathcal{M},\mu),(Y,\mathcal{N},\nu)$ を $\sigma$-有限な測度空間とする．このとき，$(X\times Y,\mathcal{M}\otimes\mathcal{N}$ 上の可測関数 $f\geq 0$ に対して次が成り立つ．

(1) $F_1(x) := \int_Y f(x,y)d\nu(y)$ は $X$ 上の可測関数であり，$F_2(y) := \int_X f(x,y)d\mu(x)$ は $Y$ 上の可測関数である．

(2) さらに，

$$\int_{X\times Y} f(x,y)d(\mu\otimes\nu)(x,y) = \int_X F_1(x)d\mu(x) = \int_Y F_2(y)d\nu(y).$$

**定理 9.3.9 (Fubini の定理).**
$(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ を $\sigma$-有限な測度空間とする. $f \in L^1(X \times Y, \mathcal{M} \otimes \mathcal{N}, \mu \otimes \nu)$ に対して次が成り立つ.

(1) $\mu$-a.e.$x$ で $f(x, \cdot) \in L^1(Y, \mathcal{N}, \nu)$ 且つ $\nu$-a.e.$y$ で $f(\cdot, y) \in L^1(X, \mathcal{M}, \mu)$.

(2) $\mu$-a.e.$x$ で定義される $F_1(x) := \int_Y f(x, y) d\nu(y)$ について $F \in L^1(X, \mathcal{M}, \mu)$. また, $\nu$-a.e.$y$ に対して定義される $F_2(y) := \int_X f(x, y) d\mu(x)$ について $F_2 \in L^1(Y, \mathcal{N}, \nu)$ となる.

(3) さらに,

$$\int_{X \times Y} f(x, y) d(\mu \otimes \nu)(x, y) = \int_X F_1(x) d\mu(x) = \int_Y F_2(y) d\nu(y).$$

## 9.4 Hahn-Banach の拡張定理

本節では Hahn-Banach の定理の主張の説明をし, 本論文で用いたいくつかの事実を Hahn-Banach の定理から導出する.

**定義 9.4.1 (劣線形汎関数).**
$V$ を $\mathbb{R}$ 上の線形空間とする. 写像 $p : V \to \mathbb{R}$ が劣線形であるとは, 以下の 2 つの条件が成り立つことを言う:

$$(1) \quad \forall \lambda > 0, \forall x \in V, \quad p(\lambda x) = \lambda p(x)$$
$$(2) \quad \forall x, y \in V, \quad p(x + y) \leq p(x) + p(y)$$

例えば, ノルム空間においてノルムは劣線形汎関数である. Hahn-Banach の定理は線形部分空間上の線形汎関数は劣線形汎関数に支配されているのであればその支配を崩さずに全空間に拡張できるという主張である. 正確には以下の通りである.

**定理 9.4.2 (Hahn-Banach の拡張定理).**
$V$ を $\mathbb{R}$ 上の線形空間, $W \subset V$ を線形部分空間, $p : V \to \mathbb{R}$ を劣線形汎関数とするとき, 線形汎関数 $f : W \to \mathbb{R}$ が $\forall x \in W, f(x) \leq p(x)$ をみたすならば, 線形汎関数 $F : V \to \mathbb{R}$ であって, $W$ 上で $F = f$ となり, $V$ 上で $F \leq p$ となるものが存在する.

**証明.** 宮島 [36] を参照されたい. □

この定理から次の形の Hahn-Banach の拡張定理が導ける.

**定理 9.4.3 (Hahn-Banach の拡張定理 2).**
$V$ をノルム空間とし, $W \subset V$ を部分線形空間とする. このとき, $W$ 上の有界線形汎関数 $f : W \to \mathbb{R}$ に対して, $V$ 上の有界線形汎関数 $F$ で

$$F(x) = f(x) \ (x \in W), \ \|F\| = \|f\|$$

を満たすものが存在する.

**証明.** $x \in V$ に対して $p(x) = \|f\|\|x\|$ とおく．明らかに $p$ は劣線形汎関数であり，$f(x) \leq p(x)$ $(x \in W)$ である．従って，Hahn-Banach の拡張定理より $F(x) = f(x)$ $(x \in W)$ かつ $F(x) \leq p(x)$ $(x \in V)$ なる線形汎関数 $F : V \to \mathbb{R}$ が取れる．このとき

$$-F(x) = F(-x) \leq p(-x) = p(x) = \|f\|\|x\| \quad (x \in V)$$

であるので $F$ は有界で $\|F\| \leq \|f\|$ である．また，

$$|f(x)| = |F(x)| \leq \|F\|\|x\| \quad (x \in W)$$

であるので $\|f\| \leq \|F\|$ であるので $\|f\| = \|F\|$ である． □

本論文で用いる Hahn-Banach の拡張定理の系を述べよう．

**系 9.4.4.** $V$ を $\mathbb{R}$ 上のノルム空間，$W \subset V$ を線形部分空間とする．このとき，ある $x_0 \in V$ が $d := d(x_0, W) = \inf_{x \in W} \|x - x_0\| > 0$ を満たすならば，

$$F(x) = 0 \ (x \in W),\ F(x_0) = 1,\ \|F\| = \frac{1}{d}$$

を満たす有界線形汎関数 $F : V \to \mathbb{R}$ が存在する．

**証明.** $X := W + \mathbb{R}x_0$ とおき，$f : X \in w + ax_0 \mapsto a \in \mathbb{R}$ とおく．$f$ は well-defined である．実際，$w_1 + ax_0 = w_2 + bx_0$ とすると，$(b - a)x_0 = w_1 - w_2 \in W$ であるので $d(x_0, W) > 0$ より $a = b$ でなければならない．明らかに $f$ は線形汎関数であり $f(x) = 0$ $(x \in W)$ かつ $f(x_0) = 1$ を満たす．そして，$x \in W, a\mathbb{R} \setminus \{0\}$ に対して，

$$\|x + ax_0\| = |a| \left\| (-\frac{1}{a}x) - x_0 \right\| \leq |a|d$$

であるので $|f(x + ax_0)| = |a| = \|x + ax_0\|/d$ である．よって，$f$ は有界であり $\|f\| \leq 1/d$ である．また，$d = \lim_{n \to \infty} \|x_0 - x_n\|$ なる $W$ の元の列 $(x_n)$ を取れば，

$$1 = f(-x_n + x_0) \leq \|f\|\|x_0 - x_n\| \to \|f\|d$$

であるので $1/d \leq \|f\|$ である．よって $\|f\| = 1/d$ であるので，Hahn-Banach の拡張定理 2 より，有界線形汎関数 $F : V \to \mathbb{R}$ であって

$$F(x) = f(x) \ (x \in X),\ \|F\| = \|f\| = \frac{1}{d}$$

なるものが存在する．この $F$ が求めるものである． □

**系 9.4.5.**
$V$ を $\mathbb{R}$ 上のノルム空間，$W \subset V$ を線形部分空間とする．もし $W$ の $V$ における閉包 $\overline{W}$ が $V$ と一致しなければ，$V$ 上の有界線形汎関数 $f \neq 0$ で，$f = 0$ (on $\overline{W}$) なるものが存在する．

**証明.** $x_0 \in V \setminus \overline{W}$ とすると $d(x_0, W) > 0$ である. 従って, すぐ上の系から $f(x_0) = 1, f(x) = 0$ $(x \in W)$ なる線形汎関数 $f : V \to \mathbb{R}$ が取れる. $x \in \overline{W}$ に対して $x$ に収束する列 $x_n \in W$ を取れば,

$$|f(x)| = |f(x) - f(x_n)| = |f(x - x_n)| \leq \|f\|\|x - x_n\| \to 0$$

であるから $f(x) = 0$ $(x \in \overline{W})$ である. よってこの $f$ が求めるものである. □

**系 9.4.6.**
$x \neq 0$ ならば $f(x) = \|x\|_X$ かつ $\|f\| = 1$ なる有界線形汎関数 $f : X \to \mathbb{R}$ が存在する.

**証明.** $W = \{0\}$ とおけば $x_0 = x$ として系 9.4.4 の仮定が満たされる. ゆえに, $d = d(x, W) = \|x\|$ とおくと, 有界線形汎関数 $F : V \to \mathbb{R}$ で

$$F(x_0) = 1,\ \|F\| = \frac{1}{d}$$

を満たすものが取れる. よって, $f = \|x\|F$ が求めるものである. □

## 9.5 Riesz-Markov-角谷の表現定理

本節ではまず符号付き測度及びそれによる積分の定義を非負値の測度の知識を前提にして述べる. その後, 符号付き Borel 測度の定義を述べ, Riesz-Markov-角谷の表現定理の主張を述べる. 証明については宮島 [36] の第 7 章を参照されたい.

**定義 9.5.1 (符号付き測度).**
$(\Omega, \mathcal{F})$ を可測空間とする. 写像 $\varphi : \mathcal{F} \to \mathbb{R}$ が次の条件をみたすときを $\varphi$ を $(\Omega, \mathcal{F})$ 上の符号付き測度と呼ぶ: $\mathcal{F}$ の元の列 $(A_n)$ が互いに素であるならば,

$$\varphi(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \varphi(A_n)$$

となる.

以下の事実に基づき符号付き測度による積分を定義する.

**命題 9.5.2.**
$\varphi$ を可測空間 $(\Omega, \mathcal{F})$ 上の符号付き測度とするとき, $A \in \mathcal{F}$ に対して,

$$\varphi^+(A) = \sup \varphi(B) \mid B \in \mathcal{F}, B \subset A, \varphi^-(A) = -\inf \varphi(B) \mid B \in \mathcal{F}, B \subset A$$

とおくと, $\varphi^+, \varphi^-$ は $(\Omega, \mathcal{F})$ 上の有限測度となり, $\varphi = \varphi^+ - \varphi^-$ が成立する.

**定義 9.5.3.**
$\varphi$ を可測空間 $(\Omega, \mathcal{F})$ 上の符号付き測度とするとき, $f \in L^1(\Omega, \mathcal{F}, \varphi^+) \cap L^1(\Omega, \mathcal{F}, \varphi^-)$ に対して, $\varphi$ に関する $f$ の積分を以下で定める.

$$\int_\Omega f d\varphi := \int_\Omega f d\varphi^+ - \int_\Omega f d\varphi^-$$

次に正則符号付き Borel 測度の定義を述べる.

**定義 9.5.4 (正則符号付き Borel 測度).**
位相空間 $(\Omega, \mathcal{F})$ 上の符号付き測度 $\varphi : \mathcal{F} \to \mathbb{R}$ は $\varphi^+, \varphi^-$ がともに $(\Omega, \mathcal{F})$ 上の正則 Borel 測度になるとき, $(\Omega, \mathcal{F})$ 上の正則符号付き測度と呼ばれる.

なお, 文脈から推察可能な場合は $S$ に入っている位相を明示せず単に $S$ 上の正則符号付き測度と呼ぶこともある. 以上の定義のもと以下が成り立つ.

**定理 9.5.5 (Riesz-Markov-角谷の表現定理).**
$S$ をコンパクト Hausdorff 空間とし, $C(S)$ を $S$ 上の実数値連続関数全体の集合とするとき, 任意の有界線形汎関数 $\varphi : C(S) \to \mathbb{R}$ に対して, $S$ 上の正則符号付き Borel 測度 $\mu$ であって,

$$\forall f \in C(S), \varphi(f) = \int_S f d\mu$$

をみたすものが一意的に存在する.

## 9.6 Stone-Weierstrass の定理

本節では Stone-Weierstrass の定理という, 連続関数空間における一様収束の意味での稠密性に関する定理の主張を述べる. 証明は宮島 [36] または新井 [41] の付録を参照されたい. なお, 本節では $S$ をコンパクト Hausdorff 空間とし, $C(S, \mathbb{K})$ で $S$ 上の $\mathbb{K}$-値連続関数全体の集合を表す ($K = \mathbb{R}, \mathbb{C}$ とする). $C(S, \mathbb{K})$ は $\|f\| = \sup\{f(x) \mid x \in S\}$ により $\mathbb{K}$ 上のノルム空間になる.

**定義 9.6.1 ($C(S, \mathbb{K})$ の部分代数).**
$A \subset C(S, \mathbb{K})$ が部分代数であるとは, $A$ が $C(S, \mathbb{K})$ の $\mathbb{K}$ 上の線形部分空間であり, かつ任意の $f, g \in A$ について積 $fg$ が $A$ の元になることをいう.

$C(S, \mathbb{K})$ の部分集合 $A$ が与えられたとき, $A$ を含む (包含関係の意味で) 最小の部分代数が存在する ($A$ を含む部分代数全体の共通部分を取ればよい). それを $A$ で生成された部分代数と呼ぶ. Stone-Weierstrass の定理は生成された部分代数の稠密性について述べたものである.

**定理 9.6.2 (Stone-Weierstrass の定理 (実数形)).**
$A \subset C(S, \mathbb{R})$ が次の 2 つの条件をみたしているとする.

(1) 恒等的に 1 を取る関数 $\mathbf{1}$ は $A$ に属する.

(2) 任意の相異なる $x, y \in S$ に対して, $f(x) \neq f(y)$ なる $f \in A$ が存在する.

このとき $A$ が生成する部分代数の閉包は $C(S, \mathbb{R})$ と一致する. すなわち, 任意の $f \in C(S, \mathbb{R})$ と $\varepsilon > 0$ に対して, $A$ が生成する部分代数の元 $g$ で, $|f - g| < \varepsilon$ をみたすものが取れる.

**系 9.6.3.** $K \subset \mathbb{R}$ をコンパクト集合とするとき, 1 変数の $\mathbb{R}$ 係数多項式関数を $K$ 上に制限したもの全体の集合を $A$ とおくと, 明らかに $A$ は上の 2 条件をみたすので Stone-Weierstrass の定理から $A$ が生成する部分代数の閉包は $C(K, \mathbb{R})$ と一致する. ところが $A$ は部分代数なので $A$ が生成する部分代数は $A$ になる. よって, $K$ 上の実数値連続関数は多項式により一様近似できる.

**定理 9.6.4 (Stone-Weierstrass の定理 (複素数形)).**
$A \subset C(S, \mathbb{C})$ が次の 2 つの条件をみたしているとする.

(1) 恒等的に 1 を取る関数 $\mathbf{1}$ は $A$ に属する.

(2) 任意の相異なる $x, y \in S$ に対して, $f(x) \neq f(y)$ なる $f \in A$ が存在する.

(3)$f \in A \Rightarrow \overline{f} \in A$

このとき $A$ が生成する部分代数の閉包は $C(S, \mathbb{C})$ と一致する. すなわち, 任意の $f \in C(S, \mathbb{C})$ と $\varepsilon > 0$ に対して, $A$ が生成する部分代数の元 $g$ で, $|f - g| < \varepsilon$ をみたすものが取れる.

**系 9.6.5.** $K \subset \mathbb{R}^n$ をコンパクト集合とし, $m \in \mathbb{Z}^n$ に対して $e_m : K \ni x \mapsto \exp(i\langle m, x\rangle) \in \mathbb{C}$ とおく. このとき $A := \mathrm{span}\{e_m\}_{m \in \mathbb{Z}^n}$ とおくと, 明らかに $A$ は上の 3 条件をみたすので Stone-Weierstrass の定理から $A$ が生成する部分代数の閉包は $C(S)$ と一致する. ところが $A$ は部分代数なので $A$ が生成する部分代数は $A$ になる. よって, $K$ 上の複素数値連続関数は $A$ の元により一様近似できる.

Riesz-Markov-角谷の定理と Stone-Weierstrass の定理から次のことがわかる.

**系 9.6.6.** $K \subset \mathbb{R}^n$ をコンパクト集合とし, $\mu$ を部分空間 $K$ 上の正則符号付き Borel 測度とすると,

$$\left(\forall m \in \mathbb{Z}^n, \ \int_K \exp(i\,\langle x, m\rangle) d\mu(x) = 0\right) \Rightarrow \mu = 0$$

が成り立つ.

**証明.** 任意に $f \in C(K)$ と $\varepsilon$ をとる．すぐ上の系より $c_k \in \mathbb{C}, m_k \in \mathbb{Z}^n$ $(k = 1, \ldots, N)$ が存在して,

$$\forall x \in K, \quad \left| f(x) - \sum_{k=1}^{N} c_k e_{m_k}(x) \right| < \varepsilon$$

となる．よって,

$$\begin{aligned} |\int_K f(x) d\mu(x)| &= |\int_K f(x) - \sum_{k=1}^{N} c_k e_{m_k}(x) d\mu(x)| \\ &\leq \int_K \left| f(x) - \sum_{k=1}^{N} c_k e_{m_k}(x) \right| d|\mu|(x) \leq |\mu|(K)\varepsilon \end{aligned}$$

となる．$\varepsilon > 0$ は任意なので $\int_K f(x) d\mu(x) = 0$ である．ところが, $K$ 上の正則符号付き Borel 測度 $\nu = 0$ もこれを満たす．よって, Riesz-Markov-角谷の定理の一意性の部分により $\mu = \nu = 0$ である． □

## 9.7 畳み込み積の性質

本節では, $C_0^\infty(\mathbb{R})$ と $L^1_{\mathrm{loc}}(\mathbb{R})$ の畳み込みが $C^\infty(\mathbb{R})$ になること, $L^1(\mathbb{R}^N)$ と $L^p(\mathbb{R}^N)$ の畳み込みが $L^p(\mathbb{R}^N)$ になることを示す.

**命題 9.7.1.** $1 \leq p < \infty$ とする．このとき, $\varphi \in L^1(\mathbb{R}^n)$ とし $f \in L^p(\mathbb{R}^n)$ とすると $\varphi * f \in L^p(\mathbb{R}^n)$ であり,

$$\|\varphi * f\|_{L^p} \leq \|\varphi\|_{L^1} \|f\|_{L^p}.$$

**証明.** $p = 1$ の場合をまず考える．このとき,

$$\begin{aligned} \int_{\mathbb{R}^n} |(\varphi * f)(x)| dx &= \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \varphi(x-y) f(y) dy \right| dx \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |\varphi(x-y)||f(y)| dy dx \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |\varphi(x-y)||f(y)| dx dy \quad \text{(Fubini の定理)} \\ &= \int_{\mathbb{R}^n} \|\varphi\|_{L^1} |f(y) dy \\ &= \|\varphi\|_{L^1} \|f\|_{L^1} \end{aligned}$$

であるので成立する．次に $p > 1$ とし $1/p + 1/q = 1$ とする．すると, 任意の $x \in \mathbb{R}^n$

に対して, Hölder の不等式 (命題 9.11.2) より

$$
\begin{aligned}
|(\varphi * f)(x)| &= \left| \int_{\mathbb{R}^n} \varphi(x-y) f(y) dy \right| \\
&\leq \int_{\mathbb{R}^n} |\varphi(x-y)|^{1-1/p} |\varphi(x-y)|^{1/p} |f(y)| dy \\
&\leq \left( \int_{\mathbb{R}^n} |\varphi(x-y)| d\mu(y) \right)^{1/q} \left( \int_{\mathbb{R}^n} |\varphi(x-y)| |f(y)|^p dy \right)^{1/p} \\
&= \|\varphi\|_{L^1}^{1/q} \left( \int_{\mathbb{R}^n} |\varphi(x-y)| |f(x-y)|^p dy \right)^{1/p}
\end{aligned}
$$

であるから,

$$
\begin{aligned}
&\int_X \left| \int_{\mathbb{R}^n} \varphi(x-y) f(y) dx \right|^p dx \\
&\leq \int_{\mathbb{R}^n} \|\varphi\|_{L^1}^{p/q} \int_{\mathbb{R}^n} |\varphi(x-y)| |f(x-y)|^p dy dx \\
&= \|\varphi\|_{L^1}^{p/q} \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} |\varphi(x-y)| |f(y)| dx dy \quad (\because \text{ Fubini の定理}) \\
&= \|\varphi\|_{L^1}^{p/q} \int_{\mathbb{R}^n} |f(y)|^p \|\varphi\|_{L^1} dy \\
&= \|\varphi\|_{L^1}^{p} \|f\|_{L^p}^{p} < \infty \quad (\because \ 1 + p/q = p)
\end{aligned}
$$

となる. よって, $\varphi * f \in L^p(\mathbb{R}^n)$ であり,

$$
\|\varphi * f\|_{L^p} \leq \|\varphi\|_{L^1} \|f\|_{L^p}.
$$

□

**命題 9.7.2.** $\varphi \in C_0^\infty(\mathbb{R})$ とし $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ とすると $\varphi * f \in C^\infty(\mathbb{R})$ であり, 任意の整数 $k \geq 0$ に対して $(\varphi * f)^{(k)} = \varphi^{(k)} * f$ となる.

**証明.** $K = \mathrm{supp}\varphi$ とおく. 任意の整数 $k \geq 0$ に対して $\mathrm{supp}(\varphi^{(k)}) \subset K$ であることに注意する. 命題を $k$ についての帰納法で示そう. $k = 0$ では自明. $k$ で成立すると仮定する. 任意に $x_0 \in \mathbb{R}$ を取り $a < b$ を $x_0 \in [a, b]$ となるように取る. $K' = [a, b] - K = \{x - y \mid x \in [a, b], y \in K\}$ とおくと $K'$ はコンパクトであり, $x \in (a, b)$ に対して

$$
(\varphi^{(k)} * f)(x) = \int_{\mathbb{R}} \varphi^{(k)}(x-y) f(y) dy = \int_{K'} \varphi^{(k)}(x-y) f(y) dy
$$

となる. いま, $g(x, y) = \varphi^{(k)}(x-y) f(y) \ \ (x \in (a, b), y \in K')$ とおけば,

$$
\left| \frac{\partial g}{\partial x}(x, y) \right| = M |f(y)| \quad (x \in (a, b), y \in K')
$$

となる．ここで, $M = \sup_{t\in\mathbb{R}} |\varphi^{(k+1)}(t)| < \infty$ である．$f$ は $K'$ 上可積分であるので積分記号下の微分 (付録の定理 9.3.4) より $(a, b)$ 上で

$$F(x) := (\varphi * f)^{(k)}(x) = (\varphi^{(k)} * f)(x) = \int_{K'} \varphi^{(k)}(x-y)f(y)dy = \int_{K'} g(x,y)dy$$

は微分可能で,

$$F'(x) = \int_{K'} \frac{\partial g}{\partial x}(x,y)dy = \int_{K'} \varphi^{(k+1)}(x-y)f(y)dy = (\varphi^{(k+1)} * f)(x)$$

となる．よって, $(\varphi * f)^{(k+1)} = \varphi^{(k+1)} * f$ である．これで示せた． □

## 9.8 $\mathbb{R}^n$ 上の有限 Borel 測度の正則性

本節では $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ 上の有限測度が次の意味で正則であることを示す.

**定義 9.8.1 (正則 Borel 測度).**
位相空間 $(X, \mathcal{O})$ 上の Borel 測度 $\mu$ は以下の条件をみたすとき正則であると言われる:

$$\forall A \in \mathcal{B}_X, \quad \mu(A) = \inf\{\mu(U) \mid U \supset A, \text{open}\} = \sup\{\mu(K) \mid K \subset A, \text{cpt.}\}.$$

ただし, 位相空間上の Borel $\sigma$-加法族, および Borel 測度の定義は以下の通りである.

**定義 9.8.2 (生成される $\sigma$-加法族).**
$X$ を集合とし, $\mathcal{A}_0 \subset \mathcal{P}(X)$ とする ($\mathcal{P}(X)$ は $X$ の部分集合全体の集合). このとき, $\mathcal{A}_0$ で生成される $\sigma$-加法族 $\sigma[\mathcal{A}_0]$ を

$$\sigma[\mathcal{A}_0] := \bigcap \{\mathcal{A} \mid \mathcal{A} \supset \mathcal{A}_0, \mathcal{A} \text{ は } X \text{ 上の} \sigma\text{-加法族}\}$$

と定義する. 明らかに $\sigma[\mathcal{A}_0]$ は $X$ 上の $\sigma$-加法族である.

**定義 9.8.3 (Borel $\sigma$-加法族, Borel 測度).**
$(X, \mathcal{O})$ を位相空間とする．このとき, $X$ 上の Borel $\sigma$-加法族 $\mathcal{B}_X$ を $\mathcal{B}_X = \sigma[\mathcal{O}]$ と定義する．そして, $X$ 上の Borel 測度とは, 可測空間 $(X, \mathcal{B}_X)$ 上の測度のことをいう.

**注意 9.8.4.** Borel $\sigma$-加法族は閉集合全体の集合 $\mathcal{D}$ から生成される．実際, $\mathcal{O} \subset \mathcal{A}$ なる $\sigma$-加法族 $\mathcal{A}$ は $E \in \mathcal{A} \Rightarrow X \setminus E \in \mathcal{A}$ をみたすので $\mathcal{D} \subset \mathcal{A}$ をみたし, $\mathcal{D} \subset \mathcal{A}$ なる $\sigma$-加法族 $\mathcal{A}$ は $E \in \mathcal{A} \Rightarrow X \setminus E \in \mathcal{A}$ をみたすので $\mathcal{O} \subset \mathcal{A}$ をみたす.

**補題 9.8.5.** $(X, d)$ を距離空間とし, $A \subset X$ は空でないとする．このとき, $x \in X$ に対して

$$\mathrm{dist}(x, A) := \inf\{d(x, y) \mid y \in A\}$$

と定義すると, $X \ni x \mapsto \mathrm{dist}(x, A) \in \mathbb{R}$ は連続である． また, $x \in X$ に対して $\mathrm{dist}(x, A) = 0$ と $x \in \mathrm{cl}(A)$ は同値である．特に, $A$ が閉集合であるとき, $\mathrm{dist}(x, A) = 0$ と $x \in A$ は同値である.

**証明.** 任意の $x_1, x_2 \in X, y \in A$ に対して

$$\mathrm{dist}(x_1, A) \leq d(x_1, y) \leq d(x_1, x_2) + d(x_2, y)$$

であるので, $\mathrm{dist}(x_1, A) - d(x_1, x_2) \leq d(x_2, y)$ となる. そこで, 左辺が $y$ に依存していないことに注意して $y$ に関して下限を取ると

$$\mathrm{dist}(x_1, A) - d(x_1, x_2) \leq \mathrm{dist}(x_2, A)$$

が得られる. よって, $\mathrm{dist}(x_1, A) - \mathrm{dist}(x_2, A) \leq d(x_1, x_2)$ となる. $x_1$ と $x_2$ の役割を交換すると, $\mathrm{dist}(x_2, A) - \mathrm{dist}(x_1, A) \leq d(x_2, x_1)$ が得られる. 以上から,

$$|\mathrm{dist}(x_1, A) - \mathrm{dist}(x_2, A)| \leq d(x_1, x_2)$$

となり, これより連続性が従う. 次に $x \in X$ に対して $\mathrm{dist}(x, A) = 0$ とする. すると, 任意の $n$ に対して, $d(x, y_n) < 1/n$ なる $y_n \in A$ が取れる. このとき, $y_n \to x$ であるので $x \in \mathrm{cl}(A)$ となる. $x \in \mathrm{cl}(A)$ のとき $d(x, A) = 0$ となることについては上の議論を逆向きにすればよい. □

**命題 9.8.6.** 距離空間 $X$ 上の有限 Borel 測度 $\mu$ は次の意味で正則である: 任意の $A \in \mathcal{B}_X$ に対して,

$$\begin{aligned}\mu(A) &= \sup\{\mu(C) \mid C \subset A, C \text{ は閉集合}\} \\ &= \inf\{\mu(U) \mid U \supset A, U \text{ は開集合}\}\end{aligned}$$

が成立する.

**証明.** $A \in \mathcal{B}_X$ かつ

$$\begin{aligned}\mu(A) &= \sup\{\mu(C) \mid C \subset A, C \text{ は閉集合}\} \\ &= \inf\{\mu(U) \mid U \supset A, U \text{ は開集合}\}\end{aligned}$$

なる集合 $A$ 全体の集合を $\mathcal{R}$ とおく. すると, $\mathcal{R}$ は $X$ 上の $\sigma$-加法族である. まず, $\emptyset \in \mathcal{R}$ は明らか. 次に, $A \in \mathcal{R}$ とすると $X \setminus A \in \mathcal{R}$ であることを示す. 任意に $\varepsilon > 0$ をとる. 閉集合 $C \subset A$ と開集合 $U \supset A$ を

$$\mu(A) < \mu(C) + \varepsilon, \mu(A) > \mu(U) - \varepsilon$$

をみたすように取る. すると, $X \setminus C \supset X \setminus A$ は開集合であるので,

$$\begin{aligned}\mu(X \setminus A) &= \mu(X) - \mu(A) \\ &> \mu(X) - \mu(C) - \varepsilon \\ &= \mu(X \setminus C) - \varepsilon \\ &\geq \inf\{\mu(E) \mid E \supset X \setminus A, E \text{ は開集合}\} - \varepsilon\end{aligned}$$

となり, $X \setminus U \subset X \setminus A$ は閉集合なので,

$$\begin{aligned}\mu(X \setminus A) &= \mu(X) - \mu(A) \\ &< \mu(X) - \mu(U) + \varepsilon \\ &= \mu(X \setminus U) + \varepsilon \\ &\leq \sup\{\mu(E) \mid E \subset X \setminus A, E \text{ は閉集合 }\} + \varepsilon\end{aligned}$$

となる．よって, $\varepsilon > 0$ の任意性によって $X \setminus A \in \mathcal{R}$ となる．次に $\mathcal{R}$ の元の列 $(A_n)_{n=1}^{\infty}$ に対して $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$ となることを示す．そのために $\varepsilon > 0$ を任意にとり, 各 $n$ に対して, 開集合 $U_n$ と閉集合 $C_n$ を $C_n \subset A_n \subset U_n$ かつ

$$\mu(U_n) - \mu(A_n) < \frac{\varepsilon}{2^n}, \quad \mu(A_n) - \mu(C_n) < \frac{\varepsilon}{2^n}$$

となるようにとる．すると, $U := \bigcup_{n=1}^{\infty} U_n \supset \bigcup_{n=1}^{\infty} A_n =: A$ は開集合であり,

$$\begin{aligned}\mu(U) - \mu(A) &= \mu\left((\bigcup U_n) \setminus (\bigcup A_n)\right) \\ &\leq \mu\left(\bigcup (U_n \setminus A_n)\right) \\ &\leq \sum \mu(U_n \setminus A_n) \\ &= \sum (\mu(U_n) - \mu(A_n)) \\ &\leq \sum \frac{\varepsilon}{2^n} = \varepsilon\end{aligned}$$

となる．また, $C = \bigcup C_n$ とおくと, 測度の連続性から

$$\mu(\bigcup C_n) = \lim_{N \to \infty} \mu\left(\bigcup_{n=1}^{N} C_n\right)$$

である．したがって, ある $N$ について $D_N = \bigcup_{n=1}^{N} C_n$ とおくと, $\mu(C) - \mu(D_N) < \varepsilon$ となる．このとき, $D_N \subset A$ は閉集合であり,

$$\begin{aligned}\mu(A) - \mu(D_N) &< \mu(A) - \mu(C) + \varepsilon \\ &\leq \mu(\bigcup (A_n \setminus C_n)) + \varepsilon \\ &\leq \sum (\mu(A_n) - \mu(C_n)) + \varepsilon \\ &\leq \sum \frac{\varepsilon}{2^n} + \varepsilon = 2\varepsilon\end{aligned}$$

となる．以上から $\bigcup A_n \in \mathcal{R}$ となる．
さて, $\mathcal{R}$ が $\sigma$-加法族であるので, すべての $X$ の閉集合が $\mathcal{R}$ に含まれることを示せば $\mathcal{B}_X \subset \mathcal{R}$ となることがわかる．そこで, $A \subset X$ を空でない閉集合とする．このと

き, $U_n = \{x \in X \mid \mathrm{dist}(x, A) < 1/n\}$ とおけば, $x \mapsto \mathrm{dist}(x, A)$ の連続性から $U_n$ は開集合であって, $A$ は閉集合であるから,

$$\bigcap_{n=1}^{\infty} U_n = \{x \in X \mid d(x, A) = 0\} = A$$

となる. しかも, $U_1 \supset U_2 \supset \cdots \supset U_n \supset \cdots$ であるから, 測度の連続性より $\mu(A) = \lim_{n\to\infty} \mu(U_n) = \inf_n \mu(U_n)$ となる. さらに $U_n \supset A$ であるから,

$$\mu(A) \leq \inf\{\mu(U) \mid U \supset A, U \text{ は開集合}\} \leq \inf_n \mu(U_n) = \mu(A)$$

となる. また, $A$ は閉集合なので $\mu(A) = \sup\{\mu(C) \mid C \subset A, C \text{ は閉集合}\}$ となる. よって, $A \in \mathcal{R}$ である. □

**系 9.8.7.** $\mathbb{R}^n$ 上の有限 Borel 測度 $\mu$ は正則である. つまり任意の $A \in \mathcal{B}_{\mathbb{R}}$ に対して,

$$\begin{aligned}\mu(A) &= \sup\{\mu(K) \mid K \subset A, C \text{ はコンパクト}\} \\ &= \inf\{\mu(U) \mid U \supset A, U \text{ は開集合}\}\end{aligned}$$

が成立する.

**証明.** $\mathbb{R}^n$ において有界閉集合とコンパクト集合は一致することに注意する. $\mathbb{R}^n$ は距離空間であるのですぐ上の命題より

$$\mu(A) = \sup\{\mu(C) \mid C \subset A, C \text{ は閉集合}\}$$

である. そこで, 任意の $\varepsilon > 0$ に対して, 閉集合 $C \subset A$ で $\mu(A) - \varepsilon < \mu(C)$ なるものが取れる. このとき, $0 \in \mathbb{R}^n$ 中心の半径 $n$ の閉球を $B(0; n)$ と表すと, $C_n = C \cap B(0; n)$ は有界閉集合であるのでコンパクトである. そして, $C_1 \subset C_2 \subset \cdots$ であり, $C = \bigcup C_n$ であるので, 測度の連続性より $\mu(C) = \lim_{n\to\infty\mu(C_n)}$ となる. したがって, $\mu(C) - \mu(C_N) < \varepsilon$ なる $N$ が取れる. ゆえに,

$$\mu(A) < \mu(C) + \varepsilon < \mu(C_N) + 2\varepsilon \leq \sup\{\mu(K) \mid K \subset A, C \text{ はコンパクト}\} + 2\varepsilon$$

となる. よって, $\varepsilon > 0$ の任意性より

$$\mu(A) \leq \sup\{\mu(K) \mid K \subset A, C \text{ はコンパクト}\}$$

となる. これで示せた. □

## 9.9 $C_0(\mathbb{R}^n)$ の $L^p$ 空間における稠密性

本節では, $\mathbb{R}^n$ 上の正則 Borel 測度 $\mu$ であって任意のコンパクト集合 $K \subset \mathbb{R}^n$ に対して $\mu(K) < \infty$ となるものについて, $C_0(\mathbb{R}^n)$ が $L^p(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mu)$ において稠密であることを証明する.

**補題 9.9.1.** $U \subset \mathbb{R}^n$ を開集合とし, $\emptyset \neq K \subset U$ をコンパクト集合とする．このとき, 開集合 $V \subset \mathbb{R}^n$ で $K \subset V \subset \mathrm{cl}(V) \subset U$ かつ $\mathrm{cl}(V)$ がコンパクトになるものが存在する.

**証明.** 各 $x \in U$ に対して, $B(x; 2\varepsilon_x) \subset U$ なる $\varepsilon_x > 0$ をとる．このとき, $K \subset U = \bigcup_{x \in U} B(x; \varepsilon_x)$ となる．$K$ はコンパクトなのである $x_1, \ldots, x_k$ について $K \subset \bigcup_{j=1}^k B(x_j; \varepsilon_{x_j})$ となる．そこで, $V = \bigcup_{j=1}^k B(x_j; \varepsilon_{x_j})$ とおけば, $V$ は開集合で, $K \subset V$ かつ,

$$\mathrm{cl}(V) \subset \bigcup_{j=1}^k \mathrm{cl}(B(x_j; \varepsilon_{x_j})) \subset \bigcup_{j=1}^k B(x_j; 2\varepsilon_{x_j}) \subset U$$

である. $\mathrm{cl}(V)$ のコンパクト性も上の関係からわかる. □

**定義 9.9.2.** 次のように定義する.

$$C_0(\mathbb{R}^n) := \{f \in C(\mathbb{R}^n) \mid \mathrm{supp}(f) \text{ はコンパクト }\}.$$

**補題 9.9.3.** $U \subset \mathbb{R}^n$ を開集合とし, $\emptyset \neq K \subset U$ をコンパクト集合とする．このとき, ある $\varphi \in C_0(\mathbb{R}^n)$ が存在して次の (1), (2), (3) をみたす.

(1) $\forall x \in K,\ \varphi(x) = 1$

(2) $0 \leq \varphi \leq 1$

(3) $\mathrm{supp}\varphi \subset U$

**証明.** $K \subset V \subset \mathrm{cl}(V) \subset U$ かつ $\mathrm{cl}(V)$ がコンパクトになるような開集合 $V \subset \mathbb{R}^n$ をとる．このとき, $K \cap (\mathbb{R}^n \setminus V)$ であるので, 前節の補題 9.8.5 より, 任意の $x \in \mathbb{R}^n$ に対して, $\mathrm{dist}(x, K) = 0$ かつ $\mathrm{dist}(x, \mathbb{R}^n \setminus V) = 0$ となることはない．したがって, $\mathrm{dist}(x, K) + (x, \mathbb{R}^n \setminus V) > 0$ である．このことに注意して,

$$\varphi(x) = \frac{\mathrm{dist}(x, \mathbb{R}^n \setminus V)}{\mathrm{dist}(x, K) + \mathrm{dist}(x, \mathbb{R}^n \setminus V)}$$

と定義する．すると, 明らかに $0\varphi \leq 1$ であり, dist の連続性から $\varphi \in C(\mathbb{R}^n)$ である．また, $x \in K$ に対して, $\mathrm{dist}(x, K) = 0$ だから $\varphi(x) = 1$ である．そして, $\{x \in \mathbb{R}^n \mid \varphi(x) = 0\} = \mathbb{R}^n \setminus V$ であるので, $\{x \in \mathbb{R}^n \mid \varphi(x) \neq 0\} = V$ である．よって, $\mathrm{supp}\varphi = \mathrm{cl}(V) \subset U$ である．そして, この関係から $\mathrm{supp}\varphi$ のコンパクト性がわかるので $\varphi \in C_0(\mathbb{R}^n)$ である. □

**命題 9.9.4.** $1 \leq p < \infty$ とし, $\mu$ を $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ 上の正則測度で任意のコンパクト集合 $K \subset \mathbb{R}^n$ に対して $\mu(K) < \infty$ となるものとする．このとき, $L^p(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mu)$ において $C_0(\mathbb{R}^n)$ は $L^p$ 距離で稠密である．すなわち, 任意の $f \in L^p(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, \mu)$ と $\varepsilon > 0$ に対して, $g \in C_0(\mathbb{R}^n)$ が存在して,

$$d_\mu^p(f, g) = \|f - g\|_{L^p} = \left( \int_{\mathbb{R}^r} |f - g|^p d\mu \right)^{1/p} < \varepsilon$$

となる.

**証明.** まず, $f=\chi_E$ $(E\in\mathcal{B}_{\mathbb{R}^r})$ の場合を考える. $f^p=f\in L^1$ であるので, $\mu(E)<\infty$ である. すると, $\mu$ の正則性より $\mu(U)<\mu(E)+\varepsilon$ なる開集合 $U\supset E$ と $\mu(E)-\varepsilon<\mu(K)$ なるコンパクト集合 $\emptyset\neq K\subset E$ が取れる. いま, $K\subset U$ に対して, すぐ上の補題を適用して $\varphi\in C_0(\mathbb{R}^n)$ を取る. このとき, $\mu(K)<\infty$ より $\mu(U\setminus K)=\mu(U)-\mu(K)=\mu(U)-\mu(E)+\mu(E)-\mu(K)<2\varepsilon$ であることに注意すると,

$$\begin{aligned}\int_{\mathbb{R}^n}|\chi_E-\varphi|^p d\mu &= \int_U |\chi_E-\varphi|^p d\mu \quad (\because \mathrm{supp}\varphi\subset U, E\subset U)\\ &= \int_{U\setminus K}|\chi_E-\varphi|^p d\mu \quad (\because K\subset E, \varphi\equiv 1 \text{ (on } K))\\ &\leq \int_{U\setminus K} 2^p d\mu = 2^p\mu(U\setminus K)<2^{p+1}\varepsilon\end{aligned}$$

となる. よって, この場合は成立する. 次に $f\geq 0$ の場合を考える. このとき, 単関数近似定理より非負単関数列 $\varphi_n$ が存在して, 各点 $x$ で $0\leq\varphi_1(x)\leq\varphi_2(x)\leq\cdots\leq\varphi_n(x)\to f(x)$ となる. すると, $|f-\varphi_n|^p\leq 2^p f^p\in L^1$ であるので, 優収束定理より

$$d_\mu^p(f,\varphi_n)^p=\int(f-\varphi_n)^p d\mu\to 0$$

となる. そこで, ある非負単関数 $\varphi=\sum_{j=1}^n c_j\chi_{E_j}$ について $d_\mu^p(f,\varphi)<\varepsilon$ となる. ここで, すぐ上で示したことより, 各 $\chi_{E_j}$ に対して適当に $g_i\in C_0(\mathbb{R}^r)$ を選ぶと, $d_\mu^p(\chi_{E_j},g_i)<\varepsilon/n\max_j\{|c_j|\}$ とできる. ゆえに, $g=\sum_{j=1}^n c_j g_j$ とおくと, $g\in C_0(\mathbb{R}^r)$ であり,

$$\|\varphi-g\|_{L^p}\leq\sum_{j=1}^n|c_j|\|\chi_{E_j}-g_j\|_{L^p}<\sum_{j=1}^n|c_j|\varepsilon/n\max_{1\leq j\leq n}\{|c_j|\}\leq\varepsilon$$

となる. よって, $\|f-g\|_{L^p}\leq\|f-\varphi\|_{L^p}+\|\varphi-g\|_{L^p}<2\varepsilon$ となる. 一般の $f\in L^p$ については $f=f^+-f^-, f^\pm\geq 0$ と分解することで $f\geq 0$ の場合に帰着される. □

この定理を使うと次のことが示せる.

**命題 9.9.5.** $1\leq p<\infty$ とし, $\mu$ を $(\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n})$ 上の Lebesgue 測度とする. このとき, $f\in L^p(\mathbb{R}^n,\mathcal{B}_{\mathbb{R}^n},\mu)$ に対して,

$$\|f(\cdot-h)-f(\cdot)\|_{L^p(\mu)}\to 0\quad(h\to 0).$$

**証明.** $\varepsilon>0$ を任意に取る. 定理 9.9.4 より $\varphi\in C_0(\mathbb{R}^n)$ で $\|f-\varphi\|_{L^p}<\varepsilon$ なるものが取れる. $\varphi$ は $\mathbb{R}^n$ 上一様連続なのである $\delta>0$ があって

$$\forall x,y\in\mathbb{R}^n, |x-y|<\delta\Rightarrow|\varphi(x)-\varphi(y)|<\frac{\varepsilon}{(2\mu(\mathrm{supp}(\varphi))^{1/p}}$$

となる．ゆえに, 任意の $0 < h < \delta$ に対して, Lebesgue 測度の平行移動不変性 ($\mu(E+h) = \mu(E)$) より,

$$\begin{aligned}\int_{\mathbb{R}^n} |\varphi(x-h) - \varphi(x)|^p d\mu(x) &= \int_{(\mathrm{supp}(\varphi)+h)\cup \mathrm{supp}(\varphi)} |\varphi(x-h) - \varphi(x)|^p d\mu(x) \\ &\leq \frac{\varepsilon^p}{2\mu(\mathrm{supp}(\varphi))}(\mu((\mathrm{supp}(\varphi)+h)) + \mu(\mathrm{supp}(\varphi))) \\ &= \frac{\varepsilon^p}{2\mu(\mathrm{supp}(\varphi))} 2\mu(\mathrm{supp}(\varphi)) = \varepsilon^p\end{aligned}$$

となる．ゆえに, 任意の $0 < h < \delta$ に対して, Lebesgue 測度の平行移動不変性 (すぐ下の注意) より

$$\begin{aligned}&\|f(\cdot - h) - f(\cdot)\|_{L^p(\mu)} \\ &\leq \|f(\cdot - h) - \varphi(\cdot - h)\|_{L^p(\mu)} + \|\varphi(\cdot - h) - \varphi(\cdot)\|_{L^p(\mu)} + \|\varphi - f\|_{L^p(\mu)} \\ &\leq 2\|\varphi - f\|_{L^p(\mu)} + \varepsilon \leq 2\varepsilon\end{aligned}$$

となる． □

**注意 9.9.6.** 上の証明で可測関数 $f \geq 0$ に対して

$$\int_{\mathbb{R}^n} f(x-h) d\mu(x) = \int f(x) d\mu(x)$$

となることを使った．これを示すには $f$ に収束する単調増加単関数列 $\varphi_j$ を取る (単関数近似定理)．$\varphi_j = \sum_{k=1}^{m_j} c_{j,k}\chi_{E_{j,k}}$ と表すと

$$\varphi_j(x-h) = \sum_{k=1}^{m_j} c_{j,k}\chi_{E_{j,k}+h}(x)$$

となるので, Lebesgue 測度の平行移動不変性より

$$\int_{\mathbb{R}^n} \varphi_j(x-h) d\mu(x) = \sum_{k=1}^{m_j} c_{j,k}\mu(E_{j,k}+h) = \sum_{k=1}^{m_j} c_{j,k}\mu(E_{j,k}) = \int_{\mathbb{R}^n} \varphi_j(x) d\mu(x)$$

となる．ゆえに単調収束定理より

$$\begin{aligned}\int_{\mathbb{R}^n} f(x) d\mu(x) &= \lim_{j\to\infty} \int_{\mathbb{R}^n} \varphi_j(x) d\mu(x) \\ &= \lim_{j\to\infty} \int_{\mathbb{R}^n} \varphi_j(x-h) d\mu(x) \\ &= \int_{\mathbb{R}^n} f(x-h) d\mu(x).\end{aligned}$$

## 9.10 Lusin の定理

本節では Lusin の定理を証明する．そのためにまず Egorov の定理を証明しよう．

**定理 9.10.1 (Egorov の定理).**
$(X, \mathcal{M}, \mu)$ を有限測度空間とする．可測関数列 $(f_n : X \to \mathbb{R})_{n\in\mathbb{N}}$ が可測関数 $f : X \to \mathbb{R}$ に概収束する (つまり，$f_n \to f$ ($\mu$-a.e.)) ならば，任意の $\varepsilon > 0$ に対して $E \in \mathcal{M}$ が存在して $\mu(E) < \varepsilon$ かつ $X \setminus E$ 上で一様に $f_n \to f$ となる．

**証明.** はじめに $f_n$ が $f$ に各点収束する場合を考える．$k \in \mathbb{N}$ に対して

$$E_n(k) := \{x \in X \mid |f_n(x) - f(x)| \geq \frac{1}{k}\} \quad (n \in \mathbb{N})$$

とおき，$F_n(k) := \bigcup_{j=n}^{\infty} E_j(k)$ とおくと，$F_n(k) \supset F_{n+1}(k)$ $(\forall n)$ であり，$f_n \to f$ より $\bigcap_{n=1}^{\infty} F_n(k) = \emptyset$ である．従って，$\mu(X) < \infty$ に注意すると，測度の連続性より $\mu(F_n(k)) \to \mu(\emptyset) = 0$ $(n \to \infty)$ である．ゆえに，任意の $\varepsilon > 0$ に対して，$k \in \mathbb{N}$ ごとに適当な $n_k \in \mathbb{N}$ が存在して $\mu(F_{n_k}(k)) < \varepsilon/2^{k+1}$ となる．そこで $E = \bigcup_{k=1}^{\infty} F_{n_k}(k)$ とおくと，

$$\mu(E) = \sum_{k=1}^{\infty} \mu(F_{n_k}(k)) \leq \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+1}} < \varepsilon$$

となる．また，任意の $\delta > 0$ に対して $1/k_0 < \delta$ なる $k_0 \in \mathbb{N}$ を取ると，任意の $n \geq n_{k_0}$ に対して

$$\forall x \in X \setminus E = \bigcap_{k=1}^{\infty} \bigcap_{j=n_k}^{\infty} \left\{x \in X \mid |f_j(x) - f(x)| < \frac{1}{k}\right\},\ |f_n(x) - f(x)| < \frac{1}{k_0} < \delta$$

となる．よって，$X \setminus E$ 上で一様に $f_n \to f$ となる．次に概収束する場合であるが，$\mu(X \setminus F) = 0$ なる $F \in \mathcal{M}$ 上で各点収束するとすれば，上で示したことから任意の $\varepsilon > 0$ に対して $A \in \mathcal{M}$ が存在して $\mu(F \cap A) < \varepsilon$ かつ $F \setminus (F \cap A)$ 上で一様収束する．よって，$E = (X \setminus F) \cup (F \cap A)$ とおけば $\mu(E) \leq \mu(X \setminus F) + \mu(F \cap A) < \varepsilon$ かつ $X \setminus E \subset F \setminus (F \cap A)$ 上で一様収束する． □

**定理 9.10.2 (Lusin の定理).**
$\mu$ を位相空間 $X$ 上の有限正則 Borel 測度とする．このとき，任意の可測関数 $f : X \to \mathbb{R}$ と $\varepsilon > 0$ に対して，コンパクト集合 $K \subset X$ が存在して，$\mu(K) > \mu(X) - \varepsilon$ かつ $f$ は $K$ 上連続となる．

**証明.** $f \geq 0$ として証明すれば十分である．実際，それが示されれば，$f = f^+ - f^-$ と分解するとき，コンパクト集合 $K_+, K_- \subset X$ が存在して $\mu(X \setminus K_\pm) < \varepsilon/2$ かつ $f^\pm$ は $K_\pm$ 上連続となるから，$K = K_+ \cap K_-$ が所望のものになる．
さて，$f \geq 0$ とする．まず $f$ が単関数である場合を考える．この場合，$a_1, \ldots, a_k \geq 0$ と互いに素な可測集合 $E_1, \ldots, E_k$ により $f = \sum_{j=1}^{k} a_j \chi_{E_j}$ と表される．すると，$\mu$ が

有限正則測度であることよりコンパクト集合 $K_j \subset X$ が存在して $\mu(E_j \setminus K_j) < \varepsilon/k$ となる．$K_1, \ldots, K_k$ は互いに素で $f$ は $K_j$ 上定数であるから $K = \bigcup_{j=1}^k K_j$ が求めるものである．次に一般の $f \geq 0$ について考える．単関数近似定理より $f$ に収束する非負値単関数列 $(\varphi_n)_{n\in\mathbb{N}}$ が取れる．上で示したことから各 $n \in \mathbb{N}$ に対してコンパクト集合 $K_n \subset X$ が存在して $\mu(X \setminus K_n) < \varepsilon/2^{n+1}$ かつ $K_n$ 上 $\varphi_n$ は連続となる．そこで $K = \bigcap_{n=1}^\infty K_n$ とおくと

$$\mu(X \setminus K) \leq \sum_{n=1}^\infty \mu(X \setminus K_n) < \varepsilon/2$$

となる．Egorov の定理よりコンパクト集合 $K_0 \subset K$ で $\mu(K \setminus K_0) < \varepsilon/2$ かつ $K_0$ 上で一様に $\varphi_n \to f$ となるものが取れる．従って, 任意の $n$ について $\varphi_n$ は $K_0$ 上連続であるので, $f$ は $K_0$ 上連続である．そして,

$$\mu(X \setminus K_0) \leq \mu(X \setminus K) + \mu(K \setminus K_0) < \varepsilon$$

となるので $K_0$ が求めるものである． □

**系 9.10.3.** $\mu$ を $\mathbb{R}^r$ 上の有限 Borel 測度とする．このとき, 任意の $f \in \mathbb{M}^r$ と $\varepsilon > 0$ に対して, コンパクト集合 $K \subset \mathbb{R}^r$ が存在して, $\mu(K) > \mu(\mathbb{R}^r) - \varepsilon$ かつ $K$ 上の関数 $f|K$ は連続となる．

**証明.** $\mathbb{R}^r$ 上の有限 Borel 測度は正則であるから上の定理より従う． □

## 9.11 Clarkson の不等式

本節では Clarkson の不等式について述べる．参考文献は [3], [42], [9] である．歴史的には Clarkson の不等式は $L^p$ 空間が一様凸であることを示すために 1936 年の Clarkson[3] により初めて証明された．まず, $0 < p < \infty$ のときの $L^p$ 空間における基本的な不等式について述べよう．本節では特に断らなければ $(X, \mathcal{M}, \mu)$ で測度空間を表す．

**補題 9.11.1.** (1) $1 \leq p < \infty$ とすると,

$$(a+b)^p \leq 2^{p-1}(a^p + b^p) \quad (\forall a, b \geq 0).$$

(2) $0 < p < 1$ とすると,

$$(a+b)^p \leq a^p + b^p \quad (\forall a, b \geq 0).$$

(3) $1 \leq p < \infty, q = p/(p-1)$ とすると,

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q} \quad (\forall a, b \geq 0).$$

(4) $0 < p < 1, q = p/(p-1)$ とすると,

$$\frac{a^p}{p} + \frac{b^q}{q} \leq ab \quad (\forall a \geq 0, b > 0).$$

**証明.** (1): $t \in [0,1]$ に対して $f(t) = (1+t)^p/(1+t^p)$ とおくと,

$$\begin{aligned} f'(t) &= p\frac{(1+t)^{p-1}}{(1+t^p)} - pt^{p-1}\frac{(1+t)^p}{(1+t^p)^2} \\ &= \frac{p(1+t)^{p-1}}{(1+t^p)^2}((1+t^p) - t^{p-1}(1+t))) \\ &= \frac{p(1+t)^{p-1}}{(1+t^p)^2}(1-t^{p-1}) \geq 0 \end{aligned}$$

である. ゆえに $f(1) = 2^{p-1}$ が $f$ の最大値である. よって, 任意の $t \in [0,1]$ に対して

$$(1+t)^p \leq 2^{p-1}(1+t^p)$$

である. これより所望の不等式を得る ($a/b$ または $b/a$ を考えればよい).
(2): 上の $f$ について $0 < p < 1$ より $f'(t) \leq 0$ となるから $f(0) = 1$ が最大値になる. これより従う.
(3) $t \in [0,\infty)$ に対して $f(t) = t^p/p + b^q/q - bt$ とおくと,

$$f'(t) = t^{p-1} - b$$

であるので $t = b^{1/(p-1)}$ で $f$ は最小値をとる.

$$f(b^{1/(p-1)}) = \frac{b^{p/(p-1)}}{p} + \frac{b^q}{q} - b^{p/(p-1)} = \frac{b^q}{p} + \frac{b^q}{q} - b^q = 0$$

であるので,

$$0 = f(b^{1/(p-1)}) \leq f(a) = \frac{a^p}{p} + \frac{b^q}{q} - ab$$

である. よって成立する.
(4): $t \in [0,\infty)$ に対して $f(t) = t^{1/p} - t/p - 1/q$ とおくと,

$$f'(t) = \frac{1}{p}t^{1/p-1} - \frac{1}{p} = \frac{1}{p}(t^{1/p-1} - 1)$$

である. ゆえに, $1/p - 1 > 0$ より $0 \leq t < 1$ のとき $f'(t) \leq 0$ であり, $1 < t < \infty$ のとき $f'(t) \geq 0$ である. そして $f'(1) = 0$ であるので,

$$f(1) = 1 - 1/p - 1/q = 0$$

が $f$ の最小値である ($1/p + 1/q = 1$ に注意). よって,

$$0 = f(1) \leq f(a^p/b^q) = \frac{a}{b^{q/p}} - \frac{a^p}{pb^q} - \frac{1}{q}$$

であるから, 両辺に $b^q$ をかけると,

$$0 \leq ab^{q-q/p} - \frac{a^p}{p} - \frac{b^q}{q}$$

となる. $q/p = q-1$ であるのでこれで所望の不等式が得られた. □

**命題 9.11.2 (Hölder の不等式).**
(1): $1 \leq p < \infty$ とし, $q = p/(p-1)$ とする. このとき, 可測関数 $f, g : X \to \mathbb{R}$ に対して,

$$\|fg\|_1 \leq \|f\|_p \|g\|_q.$$

(2): $0 < p < 1$ とし, $q = p/(p-1)$ とする. このとき, $f \in L^p(X, \mathcal{M}, \mu), g \in L^q(X, \mathcal{M}, \mu)$ に対して,

$$\|f\|_p \|g\|_q \leq \|fg\|_1$$

が成り立つ. ただし, $r < 0$ に対して $|0|^r = +\infty$ と解釈し, $h \in L^r(X, \mathcal{M}, \mu)$ であるとは, $h$ が可測関数であり,

$$\|h\|_r^r := \int_X |h|^r d\mu < \infty$$

であることと定義する (このとき特に $\mu$-a.e. で $|h| > 0$ であり, $\|h\|_r > 0$ である).

**証明.**
(1): $\|f\|_p = 0$ ならば $f = 0$ (a.e.) となるので $\|f\|_p > 0$ としてよい. 同様に $\|g\|_q > 0$ としてよい. さらに, このとき $f \in L^p(X, \mathcal{M}, \mu), g \in L^q(X, \mathcal{M}, \mu)$ としてよい. まず $1 < p < \infty$ の場合を考える. $x \in X$ に対して $a = |f(x)|/\|f\|_p$,$b = |g(x)|/\|g\|_q$ とおいて, すぐ上の補題の (3) を適用すると,

$$\frac{|f(x)||g(x)|}{\|f\|_p \|g\|_q} \leq \frac{|f(x)|^p}{p\|f\|_p^p} + \frac{|g(x)|^q}{q\|g\|_q}$$

となる. よって, 両辺を積分して $1/p + 1/q = 1$ を使えば所望の不等式が得られる. 次に $p = 1$ の場合を考える. このとき, $\mu$-a.e.$x$ に対して $|f(x)g(x)| \leq |f(x)|\|g\|_\infty$ が成り立つので両辺を積分すれば所望の不等式を得る.
(2): $\|f\|_p > 0$ としてよい. このとき, $a = |f(x)|/\|f\|_p, b = |g(x)|/\|g\|_q$ とおき, すぐ上の補題の (4) を適用すると, $\mu$-a.e.$x \in X$ に対して

$$\frac{|f(x)|^p}{p\|f\|_p^p} + \frac{|g(x)|^q}{q\|g\|_q^q} \leq \frac{|f(x)||g(x)|}{\|f\|_p \|g\|_q}$$

となる. したがって, 両辺を積分すれば所望の不等式を得る. □

**命題 9.11.3 (Minkowski の不等式).**
$0 < p < 1$ とする．このとき, $f, g \in L^p(X, \mathcal{M}, \mu)$ に対して,

$$\|f+g\|_p^p \leq \||f|+|g|\|_p^p \leq \|f\|_p^p + \|g\|_p^p,$$
$$\|f\|_p + \|g\|_p \leq \||f|+|g|\|_p.$$

**証明.** 補題 9.11.1(2) より,

$$|f(x)+g(x)|^p \leq (|f(x)|+|g(x)|)^p \leq |f(x)|^p + |g(x)|^p$$

であるので積分することで第一の不等式を得る．次に第二の不等式については a.e.$x$ で $f(x), g(x) \neq 0$ と仮定してよいので, a.e.$x$ に対して

$$(|f(x)|+|g(x)|)^p = (|f(x)|+|g(x)|)^{p-1}|f(x)| + (|f(x)|+|g(x)|)^{p-1}|g(x)|$$

であるから, Hölder の不等式より

$$\begin{aligned}\||f|+|g|\|_p^p &= \|(|f|+|g|)^{p-1}f\|_1 + \|(|f|+|g|)^{p-1}g\|_1 \\ &\geq \|(|f|+|g|)^{p-1}\|_q(\|f\|_p + \|g\|_p) \\ &= \||f|+|g|\|_p^{p/q}(\|f\|_p + \|g\|_p)\end{aligned}$$

となる．なお, $(|f|+|g|)^{p-1} \in L^q(X, \mathcal{M}, \mu)$ については第一の不等式からわかる．よって,

$$\||f|+|g|\|_p = \||f|+|g|\|_p^{p-p/q} \geq \|f\|_p + \|g\|_p.$$

□

**補題 9.11.4.** $1 < p \leq 2$ とし, $q = p/(p-1)$ とすると,

$$(1+x)^q + (1-x)^q \leq 2(1+x^p)^{q-1} \quad (\forall x \in [0,1]).$$

**証明.** 任意に $0 < x < 1$ を取り,

$$f_x(s) := (1+s^{1-q}x)(1+sx)^{q-1} + (1-s^{1-q}x)(1-sx)^{q-1} \quad (s \in [0,1])$$

とおく．$1 < p \leq 2$ より $0 < x^{p-1} < 1$ であり, $(q-1)(p-1) = 1$ より

$$f_x(x^{p-1}) = (1+x^0)(1+x^p)^{q-1} + (1-x^0)(1-x^p)^{q-1} = 2(1+x^p)^{q-1}$$

である．また, $f_x(1) = (1+x)^q + (1-x)^q$ である．よって, $s \in (0,1)$ に対して $f_x'(s) \leq 0$ が成り立つことを示せば,

$$(1+x)^q + (1-x)^q = f_x(1) \leq f_x(x^{p-1}) = 2(1+x^p)^{q-1}$$

がわかる．さて, $s \in (0,1)$ に対して,

$$\begin{aligned}
f_x'(s) &= (1-q)s^{-q}x(1+sx)^{q-1} + (q-1)x(1+s^{1-q}x)(1+sx)^{q-2} \\
&\quad - (1-q)s^{-q}x(1-sx)^{q-1} - (q-1)x(1-s^{1-q}x)(1-sx)^{q-2} \\
&= (q-1)x \begin{bmatrix} (1+sx)^{q-2}(-s^{-q}(1+sx) + (1+s^{1-q}x)) \\ + (1-sx)^{q-2}(s^{-q}(1-sx) - (1-s^{1-q}x)) \end{bmatrix} \\
&= (q-1)x(1-s^{-q})[(1+sx)^{q-2} - (1-sx)^{q-2}]
\end{aligned}$$

である．$1 < p \leq 2$ より $2 \leq q < \infty$ であるので, $s \in (0,1)$ に対して, $1 - s^{-q} < 0$ かつ $(1+sx)^{q-2} - (1-sx)^{q-2} > 0$ である．よって, $f_x'(s) \leq 0$ である． □

**補題 9.11.5.** 以下の不等式が成り立つ.

(1) $1 < p \leq 2$ とし, $q = p/(p-1)$ とすると,

$$|z+w|^q + |z-w|^q \leq 2(|z|^p + |w|^p)^{q-1} \quad (\forall z, w \in \mathbb{C}).$$

(2) $2 \leq p < \infty$ とし, $q = p/(p-1)$ とすると,

$$|z+w|^q + |z-w|^q \geq 2(|z|^p + |w|^p)^{q-1} \quad (\forall z, w \in \mathbb{C}).$$

**証明.** (1): $z, w = 0$ のときは成り立つので $z, w \neq 0$ とする．$0 < |z| \leq |w|$ としてよい．このとき, $z/w = r\exp(i\theta), 0 < r \leq 1, 0 \leq \theta \leq 2\pi$ と表せる．ゆえに, $q/(q-1) = p$ に注意すると, 所望の不等式を得るためには任意の $\theta \in [0, 2\pi]$ に対して

$$|1 + r\exp(i\theta)|^q + |-1 + r\exp(i\theta)|^q \leq 2(r^p + 1)^{q-1}$$

が成り立つことを示せば十分である．$\theta = 0$ の場合はすぐ上の補題から成り立つ．そのほかの $\theta$ について見るために,

$$\begin{aligned}
f(\theta) :&= |1 + r\exp(i\theta)|^q + |-1 + r\exp(i\theta)|^q \\
&= ((1 + r\cos(\theta))^2 + r^2\sin^2(\theta))^{q/2} + ((-1 + r\cos(\theta))^2 + r^2\sin^2(\theta))^{q/2} \\
&= (1 + 2r\cos(\theta) + r^2)^{q/2} + (1 - 2r\cos(\theta) + r^2)^{q/2}
\end{aligned}$$

とおく．$f(0)$ が $f$ の最大値であることを示せば証明は完了する．対称性から $\theta \in [0, \pi/2]$ の場合を示せば十分である．$\theta \in [0, \pi/2]$ に対して

$$\begin{aligned}
f'(\theta) &= \frac{q}{2}(-2r\sin(\theta))(1 + 2r\cos(\theta) + r^2)^{\frac{q}{2}-1} \\
&\quad + \frac{q}{2}(2r\sin(\theta))(1 - 2r\cos(\theta) + r^2)^{\frac{q}{2}-1} \\
&= qr\sin(\theta)\left((1 - 2r\cos(\theta) + r^2)^{\frac{q}{2}-1} - (1 + 2r\cos(\theta) + r^2)^{\frac{q}{2}-1}\right)
\end{aligned}$$

である．さらに, $1 < p \leq 2$ より $q \geq 2$ であるから $\cos(\theta) \geq 0$ より

$$(1 - 2r\cos(\theta) + r^2)^{\frac{q}{2}-1} - (1 + 2r\cos(\theta) + r^2)^{\frac{q}{2}-1} \leq 0$$

である．よって, $f'(\theta) \leq 0$ である．これより $f$ は $\theta = 0$ で最大値をとる．
(2): $1 < q \leq 2$ かつ $p = q/(q-1)$ である．したがって, (1) より,

$$|z + w|^p + |z - w|^p \leq 2(|z|^q + |w|^q)^{p-1} \quad (\forall z, w \in \mathbb{C}).$$

ゆえに, $q(p-1) = p$ に注意すると, 任意の $u, v \in \mathbb{C}$ に対して $z = (u+v)/2, w = (u-v)/2$ とおくことで,

$$|u|^p + |v|^p \leq 2\frac{1}{2^p}(|u+v|^q + |u-v|^q)^{p-1} = 2^{1-p}(|u+v|^q + |u-v|^q)^{p-1}$$

となる．$p - 1 = 1/(q-1)$ であるから,

$$2^{1/(q-1)}(|u|^p + |v|^p) \leq (|u+v|^q + |u-v|^q)^{1/(q-1)}$$

となる．よって, 両辺を $q-1$ 乗することで所望の不等式を得る． □

**定理 9.11.6 (Clarkson の不等式 1).**
$(X, \mathcal{M}, \mu)$ を測度空間とする．$p \in (1, \infty), q = p/(p-1)$ とする．このとき, 任意の $f, g \in L^p(X, \mathcal{M}, \mu)$ に対して,

(1) $1 < p \leq 2$ ならば,

$$\|f + g\|_p^q + \|f - g\|_p^q \leq 2\left(\|f\|_p^p + \|g\|_p^p\right)^{q-1}.$$

(2) $2 \leq p < \infty$ ならば,

$$\|f + g\|_p^q + \|f - g\|_p^q \geq 2\left(\|f\|_p^p + \|g\|_p^p\right)^{q-1}.$$

**注意 9.11.7.** 上の定理で $p = 2$ の場合は内積空間の中線定理に他ならない．これらの不等式は $L^p$ 空間が Hilbert 空間に近い性質を持つことを含意する [39].

**証明.** (1): $p < 2$ としてよい．$q(p-1) = p$ に注意する．$0 < p - 1 < 1$ であるから Minkowski の不等式より,

$$\begin{aligned}
\||f+g|^q + |f-g|^q\|_{p-1} &\geq \||f+g|^q\|_{p-1} + \||f-g|^q\|_{p-1} \\
&= \|f+g\|_p^{p/(p-1)} + \|f-g\|_p^{p/(p-1)} \\
&= \|f+g\|_p^q + \|f+g\|_p^q
\end{aligned}$$

である．一方, $(q-1)(p-1)=1$ および補題 9.11.5(1) より,

$$\begin{aligned}\||f+g|^q+|f-g|^q\|_{p-1} &\leq \|2(|f|^p+|g|^p)^{q-1}\|_{p-1}\\ &= 2\left(\int_X (|f|^p+|g|^p)^{(q-1)(p-1)}d\mu\right)^{1/(p-1)}\\ &= 2\left(\|f\|_p^p+\|g\|_p^p\right)^{q-1}\end{aligned}$$

である．よって, 所望の不等式を得る．
(2): $q(p-1)=p$ に注意する．$p-1\geq 1$ であるから 三角不等式より,

$$\begin{aligned}\||f+g|^q+|f-g|^q\|_{p-1} &\leq \||f+g|^q\|_{p-1}+\||f-g|^q\|_{p-1}\\ &= \|f+g\|_p^{p/(p-1)}+\|f-g\|_p^{p/(p-1)}\\ &= \|f+g\|_p^q+\|f+g\|_p^q\end{aligned}$$

である．一方, $(q-1)(p-1)=1$ および補題 9.11.5(2) より,

$$\begin{aligned}\||f+g|^q+|f-g|^q\|_{p-1} &\geq \|2(|f|^p+|g|^p)^{q-1}\|_{p-1}\\ &= 2\left(\int_X (|f|^p+|g|^p)^{(q-1)(p-1)}d\mu\right)^{1/(p-1)}\\ &= 2\left(\|f\|_p^p+\|g\|_p^p\right)^{q-1}\end{aligned}$$

である．よって, 所望の不等式を得る． □

次に Clarkson により示されたもうひとつの不等式を示していく．

**補題 9.11.8.** (1) $1<p\leq 2$ とすると,

$$x^p+y^p\geq (x^2+y^2)^{p/2}\quad(\forall x,y\geq 0).$$

(2) $2\leq p<\infty$ とすると,

$$x^p+y^p\leq (x^2+y^2)^{p/2}\quad(\forall x,y\geq 0).$$

**証明.** (1): 関数 $f:[0,1]\ni t\mapsto (t^p+1)-(1+t^2)^{p/2}\in\mathbb{R}$ が非負であることを示せばよいが, $1<p\leq 2$ より $s\mapsto s^{p/2-1}$ は単調減少であるので,

$$f'(t)=pt^{p-1}-\frac{p}{2}(2t)(1+t^2)^{p/2-1}\geq pt^{p-1}-ptt^{p-1}=0$$

となる．よって, $f(0)=0$ と合わせて $f\geq 0$ である．
(2): (1) の証明中の $f$ について, $2\leq p<\infty$ より $s\mapsto s^{p/2-1}$ が単調増加であることから,

$$f'(t)=pt^{p-1}-\frac{p}{2}(2t)(1+t^2)^{p/2-1}\leq pt^{p-1}-ptt^{p-1}=0$$

となるので $f\leq 0$ である．これよりわかる． □

**補題 9.11.9.** 以下の不等式が成り立つ.

(1) $1 < p \leq 2$ とすると,

$$2^{p-1}(|z|^p + |w|^p) \leq |z+w|^p + |z-w|^p \leq 2(|z|^p + |w|^p) \quad (\forall z, w \in \mathbb{C}).$$

(2) $2 \leq p < \infty$ とすると,

$$2^{p-1}(|z|^p + |w|^p) \geq |z+w|^p + |z-w|^p \geq 2(|z|^p + |w|^p) \quad (\forall z, w \in \mathbb{C}).$$

**証明.** (1): 第一の不等式については, $1 < p \leq 2$ より $s \mapsto s^{p/2}$ が凹関数であることおよび $\mathbb{C}$ における中線定理と補題 9.11.8(1) から

$$\begin{aligned}
|z+w|^p + |z-w|^p &\leq (|z+w|^2 + |z-w|^2)^{p/2} \\
&= (2(|z|^2 + |w|^2))^{p/2} \\
&= 2^{p/2} 2^{p/2} \left( \frac{|z|^2}{2} + \frac{|w|^2}{2} \right) \\
&\leq 2^p \frac{1}{2}(|z|^p + |w|^p)
\end{aligned}$$

となるので成り立つ. 第二の不等式は第一の不等式において $z = u+v, w = u-v$ とすることで得られる.
(2): 第一の不等式については, $\mathbb{C}$ における中線定理と補題 9.11.8(2) および $r \geq 1, a, b \geq 0$ に対する不等式 $(a+b)^r \leq 2^{r-1}(a^r + b^r)$ より,

$$\begin{aligned}
|z+w|^p + |z-w|^p &\leq (|z+w|^2 + |z-w|^2)^{p/2} \\
&= (2(|z|^2 + |w|^2))^{p/2} \\
&\leq 2^{p/2} 2^{p/2-1}(|z|^p + |w|^p) \\
&= 2^{p-1}(|z|^p + |w|^p)
\end{aligned}$$

となるので成り立つ. 第二の不等式は第一の不等式において $z = u+v, w = u-v$ とすることで得られる. □

**定理 9.11.10 (Clarkson の不等式 2).**
$(X, \mathcal{M}, \mu)$ を測度空間とする. $p \in (1, \infty), q = p/(p-1)$ とする. このとき, 任意の $f, g \in L^p(X, \mathcal{M}, \mu)$ に対して,

(1) $1 < p \leq 2$ ならば,

$$\|f+g\|_p^p + \|f-g\|_p^p \leq 2(\|f\|_p^p + \|g\|_p^p).$$

(2) $2 \leq p < \infty$ ならば,

$$\|f+g\|_p^p + \|f-g\|_p^p \geq 2\left(\|f\|_p^p + \|g\|_p^p\right).$$

**証明.** 前補題を $f(x), g(x)$ に対して適用して全体を積分すればよい. □

次の不等式の証明は [19] による.

**定理 9.11.11 (Clarkson の不等式).**
$(X, \mathcal{M}, \mu)$ を測度空間とする. $p \in (1, \infty), q = p/(p-1), a = \min\{p, q\}$ とすると, 任意の $f, g \in L^p(X, \mathcal{M}, \mu)$ に対して,

$$\|f+g\|_p^a + \|f-g\|_p^a \leq 2(\|f\|_p^a + \|g\|_p^a).$$

**証明.** まず $1 < p \leq 2$ の場合を考える. このとき, $p \leq 2 \leq q < \infty$ であるので $a = \min\{p, q\} = p$ である. したがって, 定理 9.11.10(1) より,

$$\begin{aligned}\|f+g\|_p^a + \|f-g\|_p^a &= \|f+g\|_p^p + \|f-g\|_p^p \\ &\leq 2(\|f\|_p^p + \|g\|_p^p) \\ &= 2(\|f\|_p^a + \|g\|_p^a)\end{aligned}$$

である. 次に $2 \leq p < \infty$ の場合を考える. このとき, $1 < q \leq 2 \leq p$ であるので $a = \min\{p, q\} = q$ である. 定理 9.11.6(2) において, $f = u + v, g = u - v$ とおくことで,

$$2^{q-1}\left(\|u\|_p^q + \|v\|_p^q\right) \geq \left(\|u+v\|_p^p + \|u-v\|_p^p\right)^{q-1}$$

を得る. $(p-1)(q-1) = 1$ であるから, 両辺を $p-1$ 乗して

$$\|u+v\|_p^p + \|u-v\|_p^p \leq 2\left(\|u\|_p^q + \|v\|_p^q\right)^{p-1}$$

を得る. 一方, 補題 9.11.1(1) より任意の $r \in [1, \infty), x, y \geq 0$ に対して, $(x+y)^r \leq 2^{r-1}(x^r + y^r)$ であるから, $q(p-1) = p$ に注意すると,

$$\begin{aligned}\|u+v\|_p^p + \|u-v\|_p^p &= \|u+v\|_p^{q(p-1)} + \|u-v\|_p^{q(p-1)} \\ &\geq \frac{1}{2^{p-2}}\left(\|u+v\|_p^q + \|u-v\|_p^q\right)^{p-1}\end{aligned}$$

である. よって,

$$\left(\|u+v\|_p^q + \|u-v\|_p^q\right)^{p-1} \leq 2^{p-1}\left(\|u+v\|_p^q + \|u-v\|_p^q\right)^{p-1}$$

であるので, $a = q$ より

$$\|u+v\|_p^a + \|u-v\|_p^a \leq 2\left(\|u+v\|_p^a + \|u-v\|_p^a\right).$$

□

## 9.12 Jackson 評価

本節では Jackson 評価の証明を与える．本節を書くにあたって参照した文献は, Mhaskar[10], Mhaskar, Micchelli[11], Schultz[29], 猪狩 [43] である．まず次の不等式を示す.

**定理 9.12.1 (積分形の Minkowski の不等式).**
$1 \leq p < \infty$ とし $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ を $\sigma$-有限な測度空間とする．このとき, 直積可測空間 $X \times Y$ 上の可測関数 $f \geq 0$ に対して,

$$\left\{ \int_X \left( \int_Y f(x,y) d\nu(y) \right)^p d\mu(x) \right\}^{1/p} \leq \int_Y \left( \int_X f(x,y)^p d\mu(x) \right)^{1/p} d\nu(y).$$

**証明.** $p = 1$ の場合は Fubini-Tonelli の定理より明らか. $1 < p < \infty$ とする.

$$F(x) := \int_Y f(x,y) d\nu(y)$$

とおくと, $F$ は $X$ 上の可測関数である (Fubini-Tonelli の定理を参照されたい). そして, Fubini-Tonelli の定理と Hölder の不等式より

$$\begin{aligned}
\int_X F(x)^p d\mu(x) &= \int_X F(x)^{p-1} \int_Y f(x,y) d\nu(y) d\mu(x) \\
&= \int_X \int_Y F(x)^{p-1} f(x,y) d\nu(y) d\mu(x) \\
&= \int_Y \int_X F(x)^{p-1} f(x,y) d\mu(x) d\nu(y) \\
&\leq \int_Y \left( \int_X F(x)^{q(p-1)} d\mu(x) \right)^{1/q} \left( \int_X f(x,y)^p d\mu(x) \right)^{1/p} d\nu(y) \\
&= \left( \int_X F(x)^p d\mu(x) \right)^{1/q} \int_Y \left( \int_X f(x,y)^p d\mu(x) \right)^{1/p} d\nu(y)
\end{aligned}$$

となるので $\left( \int_X F(x)^p d\mu(x) \right)^{1/q}$ で両辺を割れば所望の不等式を得る. □

**系 9.12.2.** $1 \leq p < \infty$ とし $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ を $\sigma$-有限な測度空間とする．このとき, $Y$ 上の可測関数 $g \geq 0$ と直積可測空間 $X \times Y$ 上の可測関数 $f \geq 0$ に対して,

$$\left\{ \int_X \left( \int_Y g(y) f(x,y) d\nu(y) \right)^p d\mu(x) \right\}^{1/p} \leq \int_Y g(y) \left( \int_X f(x,y)^p d\mu(x) \right)^{1/p} d\nu(y).$$

### 9.12.1 一般 Jackson 核の性質

**定義 9.12.3 (三角多項式).** $N = 0, 1, 2, \ldots$ に対して,

$$\sum_{|\alpha| \leq N} c_\alpha \exp(i\alpha^T x) \quad (c_\alpha \in \mathbb{C},\ x \in \mathbb{R}^d)$$

と表される関数を高々$N$次の$d$変数三角多項式という．ここで$\alpha \in \mathbb{Z}^d$であり，

$$|\alpha| = |\alpha_1| + \cdots + |\alpha_d|$$

である．また，$|\alpha| = N$を満たすある$\alpha \in \mathbb{Z}^d$に対して$c_\alpha \neq 0$となるならば$N$次の$d$変数三角多項式という．

**定義 9.12.4 (Dirichlet 核).**
$N = 0, 1, \ldots$に対して次の関数を$N$次の Dirichlet 核という．

$$D_N(x) := \frac{\sin\left(N + \frac{1}{2}\right)}{\sin\frac{x}{2}} \qquad (x \in \mathbb{R})$$

**命題 9.12.5.** 次の式が成り立つ．

$$D_N(x) = 1 + 2\sum_{n=1}^{N} \cos(nx) = \sum_{n=-N}^{N} \exp(inx) \qquad (x \in \mathbb{R}).$$

特に$N$次の Dirichlet 核は$N$次の 1 変数三角多項式である．

**証明.** 複素数$z \neq 1$に対して，

$$1 + 2\sum_{n=1}^{N} z^n = \frac{1 + z - 2z^{n+1}}{1 - z}$$

であるので$z = \exp(ix)$とおき右辺の分母分子に$\exp(-ix/2)$をかけると

$$\frac{\cos(x/2) - \left[\cos\left(N + \frac{1}{2}\right)x + i\sin\left(N + \frac{1}{2}\right)x\right]}{-i\sin(x/2)}$$

となる．よって，実部を比較することで所望の等式を得る． □

**定義 9.12.6 (Fejér 核).**
$N = 0, 1, 2, \ldots$に対して次の関数を$N$次の Fejér 核という．

$$F_N(x) := \frac{1}{N+1}\left[\frac{\sin\frac{1}{2}(N+1)x}{\sin\frac{1}{2}x}\right]^2$$

**命題 9.12.7.** 次の式が成り立つ．

$$F_N(x) = 1 + 2\sum_{n=1}^{N}\left(1 - \frac{n}{N+1}\right)\cos nx = \sum_{n=-N}^{N}\left(1 - \frac{|n|}{N+1}\right)\exp(inx)$$

特に$N$次の Fejér 核は$N$次の 1 変数三角多項式である．

**証明.** $D_N(x) = \sum_{n=-N}^{N} \exp(inx)$ を使うと,

$$\sum_{n=-N}^{N} \left(1 - \frac{|n|}{N+1}\right) \exp(inx) = \frac{1}{N+1}(D_0(x) + D_1(x) + \cdot + D_N(x))$$

がわかる. したがって, $D_N$ の定義より

$$\begin{aligned}
&\sum_{n=-N}^{N} \left(1 - \frac{|n|}{N+1}\right) \exp(inx) \\
&= \frac{1}{N+1} \frac{1}{2\left(\sin(x/2)\right)^2} \left(\sum_{n=0}^{N} 2\sin(n+\frac{1}{2})x \cdot \sin(x/2)\right) \\
&= \frac{1}{N+1} \frac{1}{2\left(\sin(x/2)\right)^2} \left(\sum_{n=0}^{N} (\cos nx - \cos(n+1)x)\right) \\
&= \frac{1}{N+1} \frac{1}{2\left(\sin(x/2)\right)^2} \left(1 - \cos(N+1)x\right) \\
&= \frac{1}{N+1} \frac{1}{2\left(\sin(x/2)\right)^2} 2\sin\left(\frac{N+1}{2}x\right) = F_N(x)
\end{aligned}$$

となる. □

**定義 9.12.8 (一般 Jackson 核).**
$N = 1, 2, 3, \ldots$ と $r = 2, 3, 4, \cdots$ に対して次の関数を一般 Jackson 核という.

$$J_{N,r}(x) := \frac{1}{c_{N,r}} \left[\frac{\sin \frac{N}{2}x}{\sin \frac{1}{2}x}\right]^{2r}.$$

ここに定数 $c_{N,r}$ は

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} J_{N,r}(x)dx = 1$$

となるように選ぶ.

**注意 9.12.9.** 一般 Jackson 核は $N-1$ 次の Fejér 核を $r$ 乗したものなので高々$r(N-1)$ 次の 1 変数三角多項式である.

一般 Jackson 核の定数 $c_{N,r}$ について次が成り立つ.

**命題 9.12.10.** $r = 2, 3, 4, \cdots$ に対して,

$$\left(\frac{2}{\pi}\right)^{2r} N^{2r-1} \leq c_{N,r} \leq \pi^{2r-1} \left(\frac{1}{2^{2r}} + \frac{1}{2r-1}\right) N^{2r-1} \qquad (\forall N = 0, 1, 2, \ldots).$$

**証明.**

$$f(x) = \left[\frac{\sin\frac{N}{2}x}{\sin\frac{1}{2}x}\right]^{2r}$$

とおくと定義から

$$c_{N,r} = \frac{1}{\pi}\int_0^{\pi} f(x)dx$$

である. $2x/\pi \le \sin x \le x$ $(0 \le x \le \pi/2)$ より,

$$\begin{aligned}
c_{N,r} &= \frac{1}{\pi}\left(\int_0^{1/N} f(x)dx + \int_{1/N}^{\pi} f(x)dx\right) \\
&\le \frac{1}{\pi}\left(\int_0^{1/N}\left(\frac{Nx/2}{x/\pi}\right)^{2r} dx + \int_{1/N}^{\pi}\left(\frac{1}{x/\pi}\right)^{2r} dx\right) \\
&\le \frac{1}{\pi}\left(\left(\frac{\pi N}{2}\right)^{2r}\int_0^{1/N} 1dx + \pi^{2r}\int_{1/N}^{\infty}\frac{1}{x^{2r}}dx\right) \\
&= \pi^{2r-1}\left(\frac{N^{2r-1}}{2^{2r}} + \frac{N^{2r-1}}{2r-1}\right)
\end{aligned}$$

である. また,

$$\begin{aligned}
c_{N,r} &\ge \frac{1}{\pi}\int_0^{\pi/N} f(x)dx \ge \frac{1}{\pi}\int_0^{\pi/N}\left(\frac{\frac{2}{\pi}\frac{N}{2}x}{x/2}\right)^{2r} dx \\
&= \frac{1}{\pi}\left(\frac{\frac{2}{\pi}\frac{N}{2}x}{x/2}\right)^{2r}\frac{\pi}{N} = \left(\frac{2}{\pi}\right)^{2r} N^{2r-1}.
\end{aligned}$$

□

この命題より一般 Jackson 核 $J_{N,r}$ について次の評価が成り立つことがわかる.

**命題 9.12.11.** $N = 1, 2, 3, \ldots$ と $r = 2, 3, 4, \ldots$ に対して次が成り立つ.

$$\begin{aligned}
J_{N,r}(x) &\le \left(\frac{\pi}{2}\right)^{4r} N \quad (|x| \le \pi), \\
J_{N,r}(x) &\le \left(\frac{\pi^2}{2x}\right)^{2r} N^{-2r+1} \quad (|x| \le \pi).
\end{aligned}$$

**証明.** 第一の不等式は $c_{N,r}$ の下からの評価と $2x/\pi \le \sin x \le x$ $(0 \le x \le \pi/2)$ より

$$J_{N,r}(x) = \frac{1}{c_{N,r}}\left(\frac{\sin\frac{N}{2}x}{\sin\frac{x}{2}}\right)^{2r} \le \frac{1}{N^{2r-1}}\left(\frac{\pi}{2}\right)^{2r}\left(\frac{\frac{N}{2}x}{\frac{1}{\pi}x}\right)^{2r} = \left(\frac{\pi}{2}\right)^{4r} N$$

となり成立する. 第二の不等式は同様に

$$J_{N,r}(x) \le \frac{1}{N^{2r-1}}\left(\frac{\pi}{2}\right)^{2r}\left(\frac{1}{\frac{1}{\pi}x}\right)^{2r} = \left(\frac{\pi^2}{2x}\right)^{2r} N^{-2r+1}.$$

□

**補題 9.12.12.** $N = 2, 3, \ldots$ と $r = 2, 3, 4, \ldots$ に対して, $q = 0, 1, \ldots, 2r-2$ ごとに定数 $C_q > 0$ が存在して

$$\frac{1}{\pi}\int_0^{\pi} x^q J_{N,r}(x)dx \leq C_q \frac{1}{N^q}$$

となる. さらに定数 $C_q$ として $C_q \leq 2^{-4r+q+2}\pi^{4r-1}$ を満たすものが取れる.

**証明.** 命題 9.12.11 より

$$\begin{aligned}
\frac{1}{\pi}\int_0^{\pi} x^q J_{N,r}(x)dx &= \frac{1}{\pi}\left(\int_0^{2/N} x^q J_{N,r}(x)dx + \int_{2/N}^{\pi} x^q J_{N,r}(x)dx\right) \\
&\leq \frac{1}{\pi}\left(\int_0^{2/N} x^q \left(\frac{\pi}{2}\right)^{4r} N dx + \int_{2/N}^{\infty} x^q \left(\frac{\pi^2}{2x}\right)^{2r} N^{-2r+1} dx\right) \\
&= \frac{\pi^{4r-1}}{2^{4r}}\left(N\int_0^{2/N} x^q dx + 2^{2r}N^{-2r+1}\int_{2/N}^{\infty} x^{q-2r} dx\right) \\
&= \frac{\pi^{4r-1}}{2^{4r}}\left(\frac{N}{q+1}\left(\frac{2}{N}\right)^{q+1} + \frac{2^{2r}N^{-2r+1}}{2r-q-1}\left(\frac{2}{N}\right)^{1+q-2r}\right) \\
&= \frac{\pi^{4r-1}}{2^{4r}}\left(\frac{2^{q+1}}{q+1} + \frac{2^{q+1}}{2r-q-1}\right)\frac{1}{N^q} \leq \pi^{4r-1}2^{-4r+q+2}\frac{1}{N^q}.
\end{aligned}$$

□

### 9.12.2 差分, 滑率の性質

**定義 9.12.13 (差分, 滑率).**
関数 $f : \mathbb{R} \to \mathbb{R}$ と $r = 1, 2, 3, \ldots$ および $h \in \mathbb{R}$ に対して

$$\Delta_h^r(f)(x) := \sum_{k=0}^{r} \binom{r}{k}(-1)^{r-k} f(x+kh) \quad (x \in \mathbb{R})$$

を $f$ の $r$ 階の差分という. また, $f$ が $2\pi$ 周期の周期関数で $I = (-\pi, \pi) \subset \mathbb{R}$ と $p \in [1, \infty]$ に対して $f \in L^p(I)$ のとき

$$\omega_r(f)_{p,I}(t) := \sup_{0<h\leq t} \|\Delta_h^r(f)\|_{L^p(I)} \quad (t > 0)$$

を $f$ の $r$ 階の滑率 (modulus of smoothness) という.

**命題 9.12.14.** 関数 $f : \mathbb{R} \to \mathbb{R}$ と $h \in \mathbb{R}$ に対して次が成り立つ.

$$\Delta_h^1(\Delta_h^{r-1}(f)) = \Delta_h^r(f) = \Delta_h^{r-1}(\Delta_h^1(f)) \quad (r = 1, 2, 3, \ldots)$$

となる. ただし, $\Delta_h^0(f) = f$ とする.

**証明.** $r$ に関する帰納法. $r=1$ での成立は明らか. $r$ で成立すると仮定すると,

$$\binom{r}{k}+\binom{r}{k-1}=\binom{r+1}{k}$$

であるから,

$$\begin{aligned}
\Delta_h^r(\Delta_h^1(f))(x) &= \sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}\Delta_h^1(f)(x+kh)\\
&= \sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}[f(x+(k+1)h)-f(x+kh)]\\
&= \sum_{k=1}^{r}(-1)^{r+1-k}\left(\binom{r}{k}+\binom{r}{k-1}\right)f(x+kh)\\
&\quad +(-1)^{r+1}f(x)+f(x+(r+1)h)\\
&= \sum_{k=0}^{r+1}\binom{r+1}{k}(-1)^{r+1-k}f(x+kh)=\Delta_h^{r+1}(f)(x).
\end{aligned}$$

よって, 第二の等式は成立する. 同様に

$$\begin{aligned}
\Delta_h^1(\Delta_h^r(f))(x) &= \Delta_h^r(f)(x+h)-\Delta_h^r(f)(x)\\
&= \sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}f(x+(k+1)h)-\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}f(x+kh)\\
&= \Delta_h^{r+1}(f)(x)
\end{aligned}$$

であるから第一の等式も成立する. □

**命題 9.12.15.** 関数 $f:\mathbb{R}\to\mathbb{R}$ と整数 $n\geq 1$, $h\in\mathbb{R}$, $r=1,2,3,\ldots$ に対して次の等式が成り立つ.

$$\Delta_{nh}^r(f)(x)=\sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\Delta_h^r(f)(x+k_1h+\cdots+k_rh)\qquad(x\in\mathbb{R}).$$

**証明.** $r$ に関する帰納法による. まず,

$$\sum_{k_1=0}^{n-1}\Delta_h^1(f)(x+k_1h)=\sum_{k=0}^{n-1}[f(x+(k+1)h-f(x+kh)]=f(x+nh)-f(x)=\Delta_{nh}^1(f)(x)$$

であるから $r=1$ で成立する. 次に $r$ で成立すると仮定すると, $\Delta_{nh}^{r+1}=\Delta_{nh}^1\Delta_{nh}^r$ で

あるから,

$$
\begin{aligned}
&\Delta_{nh}^{r+1}(f)(x)\\
&= \Delta_{nh}^{1}\left(\sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\Delta_h^r(f)(\cdot+k_1h+\cdots+k_rh)\right)(x)\\
&= \sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\left[\Delta_h^r(f)(x+nh+k_1h+\cdots+k_rh)-\Delta_h^r(f)(x+k_1h+\cdots+k_rh)\right]\\
&= \sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\sum_{k_{r+1}=0}^{n-1}\begin{bmatrix}\Delta_h^r(f)(x+k_1h+\cdots+k_rh+(k_{r+1}+1)h)\\ -\Delta_h^r(f)(x+k_1h+\cdots+k_rh+k_{r+1}h)\end{bmatrix}\\
&= \sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\sum_{k_{r+1}=0}^{n-1}\Delta_h^1\left(\Delta_h^r(f)(\cdot+k_1h+\cdots+k_rh+k_{r+1}h)\right)(x)\\
&= \sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\sum_{k_{r+1}=0}^{n-1}\Delta_h^{r+1}(f)(x+k_1h+\cdots+k_rh+k_{r+1}h).
\end{aligned}
$$

□

**命題 9.12.16.** $I=(-\pi,\pi)\subset\mathbb{R}$ とおく．周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ であって，ある $p\in[1,\infty]$ について $f\in L^p(I)$ となるものに対して次の不等式が成り立つ．

$$\omega_r(f)_{p,I}(nt)\le n^r\omega_r(f)_{p,I}(t),\quad \omega_r(f)_{p,I}(\lambda t)\le(\lambda+1)^r\omega_r(f)_{p,I}(t)\quad(\lambda>0).$$

**証明.** すぐ上の命題より

$$
\begin{aligned}
w_r(f)_{p,I}(nt) &= \sup_{0<h\le nt}\|\Delta_h^r(f)\|_{L^p(I)} = \sup_{0<h\le t}\|\Delta_{nh}^r(f)\|_{L^p(I)}\\
&\le \sup_{0<h\le t}\sum_{k_1=0}^{n-1}\cdots\sum_{k_r=0}^{n-1}\|\Delta_h^r(f)(\cdot+k_1h+\cdots+k_rh)\|_{L^p(I)}
\end{aligned}
$$

となる．ゆえに $p=\infty$ の場合は $f$ の周期性より成立する．次に $p<\infty$ とする．$f$ の周期性 (およびすぐ下の注意) より

$$
\begin{aligned}
&\|\Delta_h^r(f)(\cdot+k_1h+\cdots+k_rh)\|_{L^p(I)}^p\\
&= \int_{-\pi}^{\pi}\left|\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}f(x+kh+k_1h+\cdots+k_rh)\right|^p dx\\
&= \int_{-\pi+k_1h+\cdots+k_rh}^{\pi+k_1h+\cdots+k_rh}\left|\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}f(x+kh)\right|^p dx\\
&= \int_{-\pi}^{\pi}\left|\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}f(x+kh)\right|^p dx = \|\Delta_h^r(f)\|_{L^p(I)}^p.
\end{aligned}
$$

である．よって，

$$
\begin{aligned}
w_r(f)_{p,I}(nt) &\le \sup_{0<h\le t} \sum_{k_1=0}^{n-1} \cdots \sum_{k_r=0}^{n-1} \|\Delta_h^r(f)(\cdot + k_1 h + \cdots + k_r h)\|_{L^p(I)} \\
&= n^r \sup_{0<h\le t} \|\Delta_h^r(f)\|_{L^p(I)} = n^r \omega_r(f)_{p,I}(t).
\end{aligned}
$$

第二式については $n < \lambda \le n+1$ なる整数 $n$ をとると，前半より

$$
w_r(f)_{p,I}(\lambda t) \le w_r(f)_{p,I}((n+1)t) \le (n+1)^r \omega_r(f)_{p,I}(t) \le (\lambda+1)^r \omega_r(f)_{p,I}(t).
$$

□

**注意 9.12.17.** 周期 $2a$ の関数 $f : \mathbb{R} \to \mathbb{R}$ と $b \in \mathbb{R}$ に対して

$$
\int_{-a+b}^{a+b} f(x)dx = \int_{-a}^{a} f(x)dx
$$

である．実際，

$$
\begin{aligned}
\int_{-a+b}^{a+b} f(x)dx &= \int_{-a+b}^{a} f(x)dx + \int_{a}^{a+b} f(x)dx \\
&= \int_{-a+b}^{a} f(x)dx + \int_{-a}^{-a+b} f(x+2a)dx \\
&= \int_{-a+b}^{a} f(x)dx + \int_{-a}^{-a+b} f(x)dx = \int_{-a}^{a} f(x)dx.
\end{aligned}
$$

**命題 9.12.18.** $I = (-\pi, \pi)$ とおく．周期 $2\pi$ の周期関数 $f : \mathbb{R} \to \mathbb{R}$ であって，ある $p \in [1, \infty]$ について $f \in L^p(I)$ であるものに対して，

$$
\|\Delta_t^r(f)\|_{L^p(I)} \le w_r(f)_{p,I}(|t|) \quad (|t| < \pi).
$$

**証明.** $t \ge 0$ のときは定義より明らかである．$t \in (-\pi, 0)$ とする．$f$ の周期性より

$$
\begin{aligned}
\|\Delta_t^r(f)\|_{L^p(I)} &= \left\| \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f(\cdot + kt) \right\|_{L^p(I)} \\
&= \left\| \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f(\cdot + (k-r)t) \right\|_{L^p(I)} \\
&= \left\| \sum_{k=0}^{r} \binom{r}{r-k} (-1)^{k} f(x + k(-t)) \right\|_{L^p(I)} \\
&= \left\| (-1)^{-r} \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f(\cdot + k(-t)) \right\|_{L^p(I)} \\
&= \|\Delta_{-t}^r(f)\|_{L^p(I)} \le w_r(f)_{p,I}(|t|).
\end{aligned}
$$

□

### 9.12.3 $L^p$ かつ $2\pi$ 周期で 1 変数の場合

**補題 9.12.19.** 関数 $g:\mathbb{R}\to\mathbb{R}$ はある正の整数 $k$ について $2\pi/k$ 周期の関数であるとする. このとき, $k$ で割り切れない任意の整数 $l$ について

$$\int_{-\pi}^{\pi} g(t)\cos(lt)dt = \int_{-\pi}^{\pi} g(t)\sin(lt)dt = 0.$$

**証明.** $g$ は $2\pi$ 周期関数でもあるので,

$$\begin{aligned}\int_{-\pi}^{\pi} g(t)\exp(ilt)dt &= \int_{-\pi+2\pi/k}^{\pi+2\pi/k} g(s+2\pi/k)\exp(il(s+2\pi/k))ds \\ &= \exp(2\pi il/k)\int_{-\pi+2\pi/k}^{\pi+2\pi/k} g(s)\exp(ils)ds \\ &= \exp(2\pi il/k)\int_{-\pi}^{\pi} g(s)\exp(ils)ds\end{aligned}$$

である. $\exp(2\pi il/k)\neq 1$ であるので所望の等式を得る. □

**補題 9.12.20.** $r\geq 2$ とする. 整数 $n\geq 1$ に対して $m(n,r)=\lfloor n/r\rfloor+1$ とおき $K_{n,r}:=J_{m(n,r),r}$ とおく (ただし, $\lfloor x\rfloor$ は $x$ を超えない最大の整数を表し, $J_{N,r}$ は一般 Jackson 核である). すると, 定数 $C_r>0$ が存在して

$$\int_0^{\pi} t^q K_{n,r}(t)dt \leq C_r\frac{1}{n^q} \quad (q=0,1,\ldots,2r-2)$$

となる. また, 周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ で $f\in L^1(I)$ なるものに対して

$$S_{n,r}(f)(x) := \int_{-\pi}^{\pi}\left(\Delta_t^r(f)(x)+f(x)\right)K_{n,r}(t)dt$$

と定めると, $S_{n,r}(f)$ は高々 $n$ 次の三角多項式である.

**証明.** (前半) : 補題 9.12.12 より定数 $C_r>0$ が存在して $q=0,1,\ldots,2r-2$ に対して

$$\int_0^{\pi} t^q K_{n,r}(t)dt = \int_0^{\pi} t^q J_{m(n,r),r}(t)dt \leq C_r\frac{1}{m(n,r)^q} \leq C_r\frac{r^q}{n^q}$$

となる. したがって, 改めて $C_r r^{2r-2}$ を $C_r$ とおくことにすれば

$$\int_0^{\pi} t^q K_{n,r}(t)dt \leq C_r\frac{1}{n^q} \quad (q=0,1,\ldots,2r-2)$$

となる.
(後半) : 一般 Jackson 核の定義より $K_{n,r}=J_{m(n,r),r}$ は高々 $r(m(n,r)-1)=r\lfloor n/r\rfloor\leq$

$n$ 次の三角多項式であり, 偶関数であるので $K_{n,r}(t)$ は $1, \cos(t), \cos(2t), \ldots, \cos(nt)$ の線形結合で表される. ゆえに $S_{n,r}(f)(x)$ は

$$\int_{-\pi}^{\pi} f(x+kt)\cos(lt)dt \quad (k=0,\ldots,r,\ l=1,\ldots,n)$$

の線形結合で表される. $k=0$ のとき上の式は定数である. $k=1,\ldots,r$ に対しては $t \mapsto f(x+kt)$ は $2\pi/k$ 周期なのですぐ上の補題より $k$ で割り切れない任意の $l=1,\ldots,n$ について

$$\int_{-\pi}^{\pi} f(x+kt)\cos(lt)dt = 0$$

となる. また, $k$ で割り切れる任意の $l=1,\ldots,n$ については $m=l/k$ とおくと

$$\begin{aligned}
&\int_{-\pi}^{\pi} f(x+kt)\cos(lt)dt \\
&= \int_{x-k\pi}^{x+k\pi} f(u)\cos\left(\frac{l(u-x)}{k}\right)\frac{1}{k}du \quad (u=x+kt) \\
&= \int_{x-k\pi}^{x+k\pi} \frac{1}{k}f(u)(\cos(mu)\cos(mx)+\sin(mu)\sin(mx))du \\
&= \left(\int_{x-k\pi}^{x+k\pi} \frac{1}{k}f(u)\cos(mu)du\right)\cos(mx) + \left(\int_{x-k\pi}^{x+k\pi} \frac{1}{k}f(u)\sin(mu)du\right)\sin(mx) \\
&= \left(\int_{-k\pi}^{k\pi} \frac{1}{k}f(u)\cos(mu)du\right)\cos(mx) + \left(\int_{-k\pi}^{k\pi} \frac{1}{k}f(u)\sin(mu)du\right)\sin(mx)
\end{aligned}$$

となる. よって, $m \leq n$ に注意すれば $S_{n,r}(f)$ が高々 $n$ 次の三角多項式であることがわかる. □

**定理 9.12.21.** $I=(-\pi,\pi)$ とおき, $r \geq 2$, $p \in [1,\infty]$ とする. このとき, 次をみたす定数 $C$ が存在する: 周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ であって, あるについて $f \in L^p(I)$ であるものに対して

$$\|S_{n,r}(f)-f\|_{L^p(I)} \leq C\omega_r(f)_{p,I}\left(\frac{1}{n}\right) \quad (n=1,2,3,\ldots)$$

となる.

**証明.** すぐ上の補題の前半より, $n$ に依存しない定数 $C$ が存在して

$$\int_{-\pi}^{\pi} K_{n,r}(t)(n|t|+1)^r dt = 2\int_0^{\pi} K_{n,r}(t)(n|t|+1)^r \leq C$$

となる. まず $p=\infty$ の場合を考える.

$$\int_{-\pi}^{\pi} K_{n,r}(t)dt = 1$$

であることに注意すれば, 任意の $x \in I$ に対して命題 9.12.16 より

$$
\begin{aligned}
|S_{n,r}(f)(x) - f(x)| &= \left| \int_{-\pi}^{\pi} \left[ \Delta_t^r(f)(x) + f(x) \right] K_{n,r}(t)dt - \int_{-\pi}^{\pi} f(x) K_{n,r}(t)dt \right| \\
&= \left| \int_{-\pi}^{\pi} \Delta_t^r(f)(x) K_{n,r}(t)dt \right| \leq \int_{-\pi}^{\pi} |\Delta_t^r(f)(x) K_{n,r}(t)|dt \\
&\leq \int_{-\pi}^{\pi} \|\Delta_t^r(f)(x)\|_{L^\infty(I)} K_{n,r}(t)dt \leq \int_{-\pi}^{\pi} K_{n,r}(t)\omega_r(f)_{\infty,I}(|t|)dt \\
&\leq \int_{-\pi}^{\pi} K_{n,r}(t)\omega_r(f)_{\infty,I}\left(n|t|\frac{1}{n}\right)dt \\
&\leq \omega_r(f)_{\infty,I}\left(\frac{1}{n}\right)\int_{-\pi}^{\pi} K_{n,r}(t)(n|t|+1)^r dt \leq C\omega_r(f)_{\infty,I}\left(\frac{1}{n}\right).
\end{aligned}
$$

次に $1 \leq p < \infty$ の場合を考える. 積分形の Minkowski の不等式および命題 9.12.16 より

$$
\begin{aligned}
&\|S_{n,r}(f) - f\|_{L^p(I)} \\
&= \left( \int_{-\pi}^{\pi} \left| \int_{-\pi}^{\pi} \Delta_t^r(f)(x) K_{n,r}(t)dt \right|^p dx \right)^{1/p} \\
&\leq \int_{-\pi}^{\pi} K_{n,r}(t)\|\Delta_t^r(f)\|_{L^p(I)}dt \leq \int_{-\pi}^{\pi} K_{n,r}(t)\omega_r(f)_{p,I}(|t|)dt \\
&\leq \omega_r(f)_{p,I}\left(\frac{1}{n}\right)\int_{-\pi}^{\pi} K_{n,r}(t)(n|t|+1)^r dt \leq C\omega_r(f)_{p,I}\left(\frac{1}{n}\right).
\end{aligned}
$$

□

### 9.12.4 $L^p$ かつ $2\pi$ 周期で多変数の場合

この節では多変数の場合を扱う.

**補題 9.12.22.** $I = (-\pi, \pi)$ とおき, $1 \leq p \leq \infty$ とする. また, $T$ を $L^p(I)$ から $L^p(I)$ への有界線形写像とする. このとき, $f \in L^p(I^d)$ に対して $f$ の $i$ 番目の変数以外を固定して得られる 1 変数関数に $T$ を作用させたものを $T_i f$ とかくことにする. つまり, $x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_d \in I$ に対して関数 $f_i$ を

$$
f_i : I \ni x \to f(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d) \in \mathbb{R}
$$

により定め $(T_i f)(x_1, \ldots, x_d) := (Tf_i)(x_i)$ と定義する. すると $T_i f \in L^p(I^d)$ である.

**証明.**

$$
\|T_i f\|_{L^p(I^d)} = \|\|T_i f\|_{L^p(I)}\|_{L^p(I^{d-1})} \leq \|\|T\|\|f_i\|_{L^p(I)}\|_{L^p(I^{d-1})} = \|T\|\|f\|_{L^p(I^d)}.
$$

□

**補題 9.12.23.** $I=(-\pi,\pi)$ とおき, $1\le p\le\infty$ とする. また, $T$ を $L^p(I)$ から $L^p(I)$ への有界線形写像とする. このとき, $f\in L^p(I^d)$ に対して, $\varepsilon_i:=\|f-T_if\|_{L^p(I^d)}$ とおき, $N=1,\dots,d$ に対して $T^N=T_NT_{N-1}\cdots T_1$ とおくと,

$$\|f-T^Nf\|_{L^p(I^d)}\le\sum_{i=1}^N\varepsilon_i\|T\|^{N-i}.$$

**証明.** $N$ に関する帰納法. $N=1$ での成立は明らか. $N$ で成立するとすると,

$$\begin{aligned}\|f-T^{N+1}f\|_{L^p(I^d)}&\le\|f-T_{N+1}f\|_{L^p(I^d)}+\|T_{N+1}f-T_{N+1}T^Nf\|_{L^p(I^d)}\\&=\|f-T_{N+1}f\|_{L^p(I^d)}+\|\|T_{N+1}(f-T^Nf)\|_{L^p(I)}\|_{L^p(I^{d-1})}\\&\le\varepsilon_{N+1}+\|\|T\|\|f-T^Nf\|_{L^p(I)}\|_{L^p(I^{d-1})}\\&=\varepsilon_{N+1}+\|T\|\|f-T^Nf\|_{L^p(I^d)}\\&\le\varepsilon_{N+1}+\|T\|\sum_{i=1}^N\varepsilon_i\|T\|^{N-i}=\sum_{i=1}^{N+1}\varepsilon_i\|T\|^{N+1-i}\end{aligned}$$

となるので $N+1$ でも成立する. □

**命題 9.12.24.** $I=(-\pi,\pi)$ とおき, $r,d\ge1$ とする. $n=1,2,3,\dots$ に対して

$$K^d_{n,r}:=J_{\lfloor\frac{n}{rd}\rfloor+1,r}$$

とおく. このとき, 周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ であって, ある $p\in[1,\infty]$ について $f\in L^p(I)$ であるものに対して,

$$S^d_{n,r}(f)(x):=\int_{-\pi}^{\pi}\left(\Delta^r_t(f)(x)+f(x)\right)K^d_{n,r}(t)dt$$

と定めると, $S^d_{n,r}(f)$ は周期 $2\pi$ の周期関数であって $L^p(I)$ に属し,

$$\|S^d_{n,r}(f)\|_{L^p(I)}\le(2^r+1)\|f\|_{L^p(I)}\quad(n=1,2,3,\dots).$$

**証明.** 補題 9.12.20 と同様にして $S^d_{n,r}(f)$ は三角多項式であることがわかるので周期 $2\pi$ の周期関数である. 不等式を示そう. まず $p=\infty$ の場合を考える. このとき, 任意の $x\in[-\pi,\pi]$ に対して

$$\begin{aligned}|S^d_{n,r}(f)(x)|&\le\int_{-\pi}^{\pi}|\Delta^r_t(f)(x)+f(x)|K^d_{n,r}(t)dt\\&\le\left(\sum_{k=0}^r\binom{r}{k}+1\right)\|f\|_{L^\infty(I)}\int_{-\pi}^{\pi}K^d_{n,r}(t)dt=(2^r+1)\|f\|_{L^\infty(I)}\end{aligned}$$

となるのでよい．次に $p < \infty$ の場合を考える．このとき, 積分形の Minkowski の不等式より

$$\begin{aligned}\|S^d_{n,r}(f)\|_{L^p(I)} &\le \int_{-\pi}^{\pi} K^d_{n,r}(t)\left(\int_{-\pi}^{\pi}|\Delta^r_t(f)(x)+f(x)|^p dx\right)^{1/p} dt \\ &\le \int_{-\pi}^{\pi} K^d_{n,r}(t)\left(\sum_{k=0}^{r}\binom{r}{k}+1\right)\|f\|_{L^p(I)}dt = (2^r+1)\,\|f\|_{L^p(I)}\end{aligned}$$

となる． □

**命題 9.12.25.** $I=(-\pi,\pi)$ とおき, $1\le p\le\infty,\ r\ge 2$ とする．$f:\mathbb{R}^d\to\mathbb{R}$ は各変数ごとに周期 $2\pi$ の周期関数で $f\in L^p(I)$ であるとする．このとき, $x=(x_1,\dots,x_d)$ に対して

$$S^d_{n,r,j}(f)(x) := \int_{-\pi}^{\pi}\left[\sum_{k=0}^{r}\binom{r}{k}(-1)^{r-k}f(x_1,\dots,x_j+kt,\dots,x_d)+f(x)\right]K^d_{n,r}(t)dt$$

とおき, $k=1,\dots,d$ に対して $S^k = S^d_{n,r,k}S^d_{n,r,k-1}\cdots S^d_{n,r,1}$ とおくと, $S^d f$ は高々 $n$ 次の $d$ 変数三角多項式であり,

$$\|f-S^d f\|_{L^p(I^d)} \le \sum_{j=1}^{d}\|f-S^d_{n,r,j}f\|_{L^p(I^d)}(2^r+1)^{d-j}.$$

**証明.** $k=1,\dots,d$ に対し $S^k f$ は各変数ごとに $2\pi$ 周期関数であり, $x_{k+1},\dots,x_d$ を固定するとき $S^k f$ は $x_1,\dots,x_k$ の関数として高々 $k\lfloor n/d\rfloor$ 次の $k$ 変数三角多項式であることを帰納法で示す．まず $S^1 f$ が各変数ごとに $2\pi$ 周期関数であることは $f$ の周期性と $S_{n,r,1}$ の定義より明らか．そして $x_2,\dots,x_d$ を固定するとき $f$ の周期性より補題 9.12.20 と同様に $S_{n,r,1}(f)(x_1,\dots,x_d)$ は $x_1$ の関数として高々 $\lfloor n/d\rfloor$ 次の 1 変数三角多項式である．次に $k$ で成り立つと仮定する．$S^{k+1}f$ の周期性は $S^k f$ の周期性より従う．$x_1,\dots,x_i,x_{k+2},\dots,x_d$ を固定する．$S^k f$ の周期性より補題 9.12.20 と同様に $S^{k+1}f$ は $x_{k+1}$ の関数として

$$\cos(mx_{k+1}),\ \sin(mx_{k+1}) \quad (m=0,\dots,\lfloor n/d\rfloor)$$

の線形結合で表されることがわかる．そして, その係数のうち $x_1,\dots,x_k$ に依存する部分は補題 9.12.20 の証明から

$$\begin{aligned}&\int_{-m\pi}^{m\pi}\frac{1}{k}(S^k f)(x_1,\dots,x_k,u,x_{k+2},\dots,x_d)\cos(ju)du \quad (j=0,\dots,\lfloor n/d\rfloor),\\ &\int_{-m\pi}^{m\pi}\frac{1}{k}(S^k f)(x_1,\dots,x_k,u,x_{k+2},\dots,x_d)\cos(ju)du \quad (j=0,\dots,\lfloor n/d\rfloor)\end{aligned}$$

である. 一方, 帰納法の仮定より $S^k f$ は $x_1, \ldots, x_k$ の関数として高々$k\lfloor n/d \rfloor$ 次の $k$ 変数三角多項式であるので

$$(S^k f)(x_1, \ldots, x_k, u, x_{k+2}, \ldots, x_d) = \sum_{|\alpha| \le k\lfloor n/d \rfloor} c^{(\alpha)}_{u, x_{k+2}, \ldots, x_d} \exp(i(\alpha_1 x_1 + \cdots + \alpha_i x_k))$$

と表される. よって,

$$\cos(mx) = \frac{\exp(imx) + \exp(-imx)}{2}, \quad \sin(mx) = \frac{\exp(imx) - \exp(-imx)}{2}$$

であることに注意すれば $S^{k+1} f$ は $x_1, \ldots, x_{k+1}$ の関数として

$$\exp(i(\alpha_1 x_1 + \alpha_{k+1} x_{k+1})) \quad (|\alpha| \le (k+1)\lfloor n/d \rfloor)$$

の線形結合で表される. これで $k+1$ の成立が示された. このことより前半の主張は成立する. 後半の主張については前命題と補題 9.12.23 より従う. □

### 9.12.5 Sobolev 空間の元の場合

本小節では, $r \ge 1$, $1 \le p \le \infty$ に対して定数 $C_{r,p} > 0$ が存在して, 周期 $2\pi$ の周期関数 $f : \mathbb{R} \to \mathbb{R}$ で任意の有界開区間 $J \subset \mathbb{R}$ について $f \in W^{r,p}(J)$ であるものに対して

$$\omega_r(f)_{p,I}\left(\frac{1}{n}\right) \le C_{r,p} \frac{1}{n^r} \|f^{(r)}\|_{L^p(I)} \quad (n = 1, 2, 3, \ldots) \tag{1}$$

となることを示していく (ただし $I = (-\pi, \pi)$ である). これが示されれば次のことがわかる.

**定理 9.12.26.** $I = (-\pi, \pi)$ とおき, $1 \le d < \infty$, $1 \le p \le \infty$, $r \ge 2$ とする. このとき, 次をみたす定数 $C$ が存在する. $f : \mathbb{R}^d \to \mathbb{R}$ が各変数ごとに周期 $2\pi$ の周期関数かつ任意の有界開区間 $J \subset \mathbb{R}$ について $f \in W^{r,p}(J)$ であるならば,

$$\begin{aligned} E_{n,p}(f) &:= \inf\{\|T - f\|_{L^p(I^d)} \mid T \text{ は高々} n \text{ 次の } d \text{ 変数三角多項式}\} \\ &\le C \frac{1}{n^r} \sum_{j=1}^{d} \left\|\partial_j^r f\right\|_{L^p(I^d)} \quad (n = 1, 2, 3, \ldots). \end{aligned}$$

**証明.** 命題 9.12.25 より $S^d f$ は高々$n$ 次の $d$ 変数三角多項式であり,

$$\|f - S^d f\|_{L^p(I^d)} \le \sum_{j=1}^{d} \|f - S^d_{n,r,j} f\|_{L^p(I^d)} (2^r + 1)^{d-j}.$$

となる. $f, S^d_{n,r,j} f$ を $x_j$ の関数とみなすとき定理 9.12.21 と同様にして定数 $C_j$ が存在して

$$\|f - S^d_{n,r,j}\|_{L^p(I)} \le C \omega_r(f)_{p,I}\left(\frac{1}{n}\right) \quad (n = 1, 2, 3, \ldots)$$

となる．したがって, (1) より

$$
\begin{aligned}
\|f - S^d f\|_{L^p(I^d)} &\le \sum_{j=1}^{d} \|f - S^d_{n,r,j} f\|_{L^p(I^d)} (2^r+1)^{d-j} \\
&\le (2^r+1)^d \sum_{j=1}^{d} \| \|f - S^d_{n,r,j} f\|_{L^p(I)} \|_{L^p(I^{d-1})} \\
&\le (2^r+1)^d C \sum_{j=1}^{d} \left\| \omega_r(f)_{p,I}\left(\frac{1}{n}\right) \right\|_{L^p(I^{d-1})} \\
&\le C' \frac{1}{n^r} \sum_{j=1}^{d} \left\| \left\| \partial_j^r f \right\|_{L^p(I)} \right\|_{L^p(I^{d-1})} = C' \frac{1}{n^r} \sum_{j=1}^{d} \left\| \partial_j^r f \right\|_{L^p(I^d)}
\end{aligned}
$$

となる．これで示せた． □

さらにこのことから次のこともわかる．

**定理 9.12.27.** $1 \le d < \infty$, $1 \le p \le \infty$, $r \ge 2$ とする．このとき, $n$ に依存しない定数 $C$ が存在して, 任意の $f \in W^{r,p}((-1,1)^d)$ に対して

$$
\begin{aligned}
E_{n,p}(f) &:= \inf\{\|P - f\|_{L^p((-1,1)^d)} \mid T \text{ は高々} n \text{ 次の } d \text{ 変数多項式}\} \\
&\le C \frac{1}{n^r} \|f\|_{W^{r,p}((-1,1)^d)} \qquad (n = 1, 2, 3, \ldots).
\end{aligned}
$$

**証明.** 次をみたす有界線形作用素が存在する (拡張作用素. 宮島 [37] を参照).

$$
T : W^{r,p}((-1,1)^d) \to W^{r,p}(\mathbb{R}^d), \quad (Tf)(x) = f(x) \quad (\forall x \in (-1,1)^d).
$$

$0 \le \varphi \in C_0^\infty(\mathbb{R}^d)$ を $(-1,1)^d$ 上で 1 をとり $(-3/2, 3/2)^d$ の外で 0 をとるものとする．$x \in \mathbb{R}^d$ に対して $g(x) := \varphi(x)(Tf)(x)$ とおく．そして $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ に対して

$$
h(x) := g(2\cos(x_1), \ldots, 2\cos(x_d))
$$

とおく．すると, $(-3/2, 3/2)^d$ の外で $g = 0$ ゆえ, 任意の空でない有界開区間 $J \subset \mathbb{R}$ について $h$ は各変数ごとに $2\pi$ 周期の周期関数かつ $h \in W^{r,p}(J)$ である．したがって, 特に, 命題 9.12.25 より $S^d h$ は高々 $n$ 次の $d$ 変数三角多項式である．また, $h$ は各変数ごとに偶関数であるので命題 9.12.25 の証明より $S^d h$ は次のように表される:

$$
(S^d h)(x) = \sum_{|\alpha| \le n} c_\alpha \prod_{j=1}^{d} \cos(\alpha_j x_j) \qquad (c_\alpha \in \mathbb{R}).
$$

ここで整数 $k \ge 0$ に対して高々 $k$ 次の 1 変数多項式 $T_k$ が存在して

$$
T_k(2\cos x) = \cos(kx) \qquad (\forall x \in \mathbb{R})
$$

となる (Chebyshev 多項式). そこで高々$n$ 次の $d$ 変数多項式 $P_f$ を

$$P_f(t) = \sum_{|\alpha|\le n} c_\alpha \prod T_{\alpha_j}(t_j)$$

で定め, $\Phi(x) = (2\cos(x_1), \ldots, 2\cos(x_d))$ とおくと, $P(\Phi(x)) = (S^d h)(x)$ となる. よって, 変数変換と定理 9.12.26 の証明より

$$\begin{aligned}
\|f - P_f\|_{L^p((-1,1)^d)} &\le \|g - P_f\|_{L^p((-2,2)^d)} \le C\|g\circ\Phi - P_f\circ\Phi\|_{L^p((-\pi,\pi)^d)} \\
&= \|h - S^d h\|_{L^p((-\pi,\pi)^d)} \le \frac{C'}{n^r}\sum_{j=1}^d \|\partial_j^r h\|_{L^p((-\pi,\pi)^d)} \\
&\le \frac{C''}{n^r}\|g\|_{W^{r,p}((-2,2)^d)} \qquad (\because\ \forall x \notin (-3/2,3/2)^d,\ g(x)=0) \\
&\le \frac{C''}{n^r}\|Tf\|_{W^{r,p}((-2,2)^d)} \le \frac{C''}{n^r}\|T\|\|f\|_{W^{r,p}((-1,1)^d)}
\end{aligned}$$

となる. □

さて, 小節のはじめに述べたことを以下で示していこう.

**命題 9.12.28.** $\mathbb{R}$ 上の関数 $M_r$ $(r = 1, 2, 3, \ldots)$ を次のように帰納的に定める.

$$M_1 := \chi_{[0,1]}, \quad M_r(x) := (M_{r-1} * M_1)(x) = \int_{\mathbb{R}} M_{r-1}(x-y)M_1(y)dy$$

すると, $\mathrm{supp} M_r \subset [0, r]$, $0 \le M_r \le 1$ となる.

**証明.** $r = 1$ での成立は明らか. $r-1$ で成立すると仮定する. すると,

$$0 \le M_r(x) = \int_{\mathbb{R}} M_{r-1}(x-y)M_1(y)dy \le \int_{[0,1]} 1dy = 1 \qquad (\forall x)$$

であり, $x \in (-\infty, 0) \cup (r, \infty)$ とすると $y \in [0,1]$ に対して $x - y \notin [0, r-1]$ であるので $M_r(x) = 0$ である. □

**命題 9.12.29.** $I = (-\pi, \pi) \subset \mathbb{R}$ とおき, $r = 1, 2, 3, \ldots$ に対して,

$$M_r(x,h) := \frac{1}{h}M_r\left(\frac{x}{h}\right) \qquad (x \in \mathbb{R},\ h > 0)$$

と定める. このとき, $f : \mathbb{R} \to \mathbb{R}$ を周期 $2\pi$ の周期関数で, 任意の有界開区間 $J \subset \mathbb{R}$ について $f \in W^{r,p}(J)$ であるものとすると, 任意の $x \in I$ について

$$h^{-r}\Delta_h^r(f)(x) = h^{-r}\sum_{k=0}^r \binom{r}{k}(-1)^{r-k}f(x+kh) = \int_{\mathbb{R}} f^{(r)}(x+t)M_r(t,h)dt$$

が成り立つ.

**証明.** $r$ に関する帰納法による. まず $r=1$ のときは, 周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ で任意の有界開区間 $J\subset\mathbb{R}$ について $f\in W^{r,p}(J)$ であるものに対して

$$\begin{aligned}\int_{\mathbb{R}} f^{(1)}(x+t)M_1(t,h)dt &= \int_{\mathbb{R}} f^{(1)}(x+t)h^{-1}\chi_{[}0,1](h^{-1}t)dt \\ &= \int_{[0,h]} f^{(1)}(x+t)dt = f(x+h)-f(x)\end{aligned}$$

だから成立する. $r-1$ で成立するとする. 周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ で任意の有界開区間 $J\subset\mathbb{R}$ について $f\in W^{r,p}(J)$ であるものを任意に取る. このとき $\Delta_h^1(f)$ は周期 $2\pi$ の周期関数で, 任意の有界開区間 $J\subset\mathbb{R}$ について $f\in W^{r,p}(J)$ であるので帰納法の仮定より

$$\begin{aligned}h^{-r}\Delta_h^r(f,x) &= h^{-1}h^{-(r-1)}\Delta_h^{r-1}(\Delta_h^1(f))(x) \\ &= h^{-1}\int_{\mathbb{R}}(\Delta_h^1(f))^{(r-1)}(x+t)M_{r-1}(t,h)dt \\ &= h^{-1}\int_{\mathbb{R}}\left(f^{(r-1)}(x+h+t)-f^{(r-1)}(x+t)\right)M_{r-1}(t,h)dt \\ &= h^{-1}\int_{\mathbb{R}}\left(\int_0^h f^{(r)}(x+v+t)dv\right)M_{r-1}(t,h)dt \\ &= h^{-1}\int_{\mathbb{R}}\left(\int_0^h f^{(r)}(x+v+uh)dv\right)M_{r-1}(u)du \\ &= \int_{\mathbb{R}}\left(\int_0^1 f^{(r)}(x+(v+u)h)dv\right)M_{r-1}(u)du \\ &= \int_0^1\left(\int_{\mathbb{R}} f^{(r)}(x+(v+u)h)M_{r-1}(u)du\right)dv \\ &= \int_0^1\left(\int_{\mathbb{R}} f^{(r)}(x+wh)M_{r-1}(w-v)dw\right)dv \\ &= \int_{\mathbb{R}}\left(f^{(r)}(x+wh)\int_0^1 M_{r-1}(w-v)dv\right)dw \\ &= \int_{\mathbb{R}} f^{(r)}(x+wh)M_r(w)dw \\ &= \int_{\mathbb{R}} f^{(r)}(x+t)M_r\left(\frac{t}{h}\right)\frac{1}{h}dt \\ &= \int_{\mathbb{R}} f^{(r)}(x+t)M_r(t,h)dt.\end{aligned}$$

□

**注意 9.12.30.** 上の命題の証明で Sobolev 空間の元に対する微分積分学の基本定理を用いた. この証明は例えば宮島 [37] の命題 3.4 を参照されたい.

**定理 9.12.31.** $I=(-\pi,\pi)\subset\mathbb{R}$ とおく. $r\geq 1$, $1\leq p\leq\infty$ に対して定数 $C_{r,p}>0$ が存在して, 周期 $2\pi$ の周期関数 $f:\mathbb{R}\to\mathbb{R}$ で任意の有界開区間 $J\subset\mathbb{R}$ について $f\in W^{r,p}(J)$ であるものに対して

$$\omega_r(f)_{p,I}(t)=\sup_{0<h\leq t}\|\Delta_h^r(f)\|_{L^p(I)}\leq C_{r,p}t^r\|f^{(r)}\|_{L^p(I)}\quad(\forall t\in(0,2\pi/r)).$$

**証明.** 命題 9.12.29 より $x\in I$ に対して

$$\Delta_h^r(f)(x)=h^r\int_{\mathbb{R}}f^{(r)}(x+t)M_r(t,h)dt$$

であるので, $r,p$ のみに依存する定数 $C_{r,p}>0$ が存在して

$$\left\|\int_{\mathbb{R}}f^{(r)}(\cdot+t)M_r(t,h)dt\right\|_{L^p(I)}\leq C_{r,p}\|f^{(r)}\|_{L^p(I)}\quad(\forall h\in(0,2\pi/r))$$

となることを示せばよい. まず $p=\infty$ の場合は

$$\begin{aligned}\int_{\mathbb{R}}f^{(r)}(x+t)M_r(t,h)dt&=\int_{\mathbb{R}}f^{(r)}(x+t)M_r\left(\frac{t}{h}\right)\frac{1}{h}dt\\&=\int_{\mathbb{R}}f^{(r)}(x+wh)M_r(w)dw.\end{aligned}$$

であることと $\mathrm{supp}M_r\subset[0,r]$ より

$$\left\|\int_{\mathbb{R}}f^{(r)}(\cdot+t)M_r(t,h)dt\right\|_{L^\infty(I)}\leq r\|f^{(r)}\|_{L^\infty(I)}$$

となり成立する. 次に $p<\infty$ の場合を考える. Minkowski の不等式 (系 9.12.2) より

$$\begin{aligned}&\left(\int_I\left|\int_{\mathbb{R}}f^{(r)}(x+t)M_r(t,h)dt\right|^p dx\right)^{1/p}\\&\leq\left(\int_I\left(\int_{\mathbb{R}}|f^{(r)}(t)|M_r(t-x,h)dt\right)^p dx\right)^{1/p}\\&\leq\int_{\mathbb{R}}|f^{(r)}(t)|\left(\int_I M_r(t-x,h)^p dx\right)^{1/p}dt\end{aligned}$$

である. ここで,

$$\begin{aligned}\int_I M_r(t-x,h)^p dx&=\int_{-\pi}^{\pi}\frac{1}{h}M_r\left(\frac{t-x}{h}\right)^p dx\\&=\int_{\frac{t-\pi}{h}}^{\frac{t+\pi}{h}}M_r(y)^p dy\end{aligned}$$

であり $\mathrm{supp} M_r \subset [0,r]$ であるので $t \in (-\infty, -\pi] \cup [\pi + rh, \infty)$ のとき

$$\int_I M_r(t-x,h)^p dx = 0$$

である．よって, $h \in (0, 2\pi/r)$ に注意すれば,

$$\begin{aligned}
&\int_{\mathbb{R}} |f^{(r)}(t)| \left( \int_I M_r(t-x,h)^p dx \right)^{1/p} dt \\
&= \int_{-\pi}^{3\pi} |f^{(r)}(t)| \left( \int_{\frac{t-\pi}{h}}^{\frac{t+\pi}{h}} M_r(y)^p dy \right)^{1/p} dt \\
&\leq \int_{-\pi}^{3\pi} |f^{(r)}(t)| \left( \int_0^r 1 dy \right)^{1/p} dt = 2r^{1/p} \int_{-\pi}^{\pi} |f^{(r)}(t)| dt
\end{aligned}$$

となる．これで $p=1$ の場合の成立はわかる．$1 < p < \infty$ の場合は $1/p + 1/q = 1$ とすると Hölder の不等式より

$$\int_{-\pi}^{\pi} |f^{(r)}(t)| dt \leq \left( \int_{-\pi}^{\pi} 1^q dt \right)^q \left( \int_{-\pi}^{\pi} |f^{(r)}(t)|^p dt \right)^p$$

だから成立する． □

**系 9.12.32.** $I = (-\pi, \pi) \subset \mathbb{R}$ とおく．$r \geq 1$, $1 \leq p \leq \infty$ に対して定数 $C_{r,p} > 0$ が存在して, 周期 $2\pi$ の周期関数 $f : \mathbb{R} \to \mathbb{R}$ で任意の有界開区間 $J \subset \mathbb{R}$ について $f \in W^{r,p}(J)$ であるものに対して

$$\omega_r(f)_{p,I} \left( \frac{1}{n} \right) \leq \frac{C_{r,p}}{n^r} \|f^{(r)}\|_{L^p(I)} \quad (n = 1, 2, 3, \ldots).$$

**証明.** $N_{r,p} \in \mathbb{N}$ を十分大きく取れば $n \geq N_{r,p}$ に対して $1/n \in (0, 2\pi/r)$ となる．よって, すぐ上の定理より

$$\omega_r(f)_{p,I} \left( \frac{1}{n} \right) \leq \frac{C_{r,p}}{n^r} \|f^{(r)}\|_{L^p(I)} \quad (\forall n \geq N_{r,p}).$$

となる．そこで $C_{r,p}$ を必要に応じて大きくすれば $n = 1, \ldots, N_{r,p}$ についても

$$\omega_r(f)_{p,I} \left( \frac{1}{n} \right) \leq \frac{C_{r,p}}{n^r} \|f^{(r)}\|_{L^p(I)}$$

となる．これで示せた． □

## 9.13 Baire のカテゴリー定理

本節では完備距離空間における Baire のカテゴリー定理について述べる．

**定理 9.13.1 (Baire のカテゴリー定理).**
$(X,d)$ を完備距離空間とし, $(V_n)_{n\in\mathbb{N}}$ を $X$ において稠密な開集合からなる列とする. このとき $\bigcap_{n\in\mathbb{N}} V_n$ は $X$ において稠密である.

**証明.** 任意の $x_0 \in X$ と $r_0 \in (0,\infty)$ に対し,

$$B(x_0,r_0) \cap \bigcap_{n\in\mathbb{N}} V_n \neq \emptyset \qquad (*)$$

が成り立つことを示せばよい. $x_0 \in X = \overline{V_1}$ より $B(x_0,r_0)\cap V_1 \neq \emptyset$ である. よって,

$$\overline{B}(x_1,r_1) \subset B(x_0,r_0)\cap V_1, \quad 0 < r_1 \leq \frac{r_0}{2}$$

なる閉球 $\overline{B}(x_1,r_1) = \{x \in X : d(x,x_1) \leq r_1\}$ が取れる. $x_1 \in X = \overline{V_2}$ より $B(x_1,r_1)\cap V_2 \neq \emptyset$ であるから,

$$\overline{B}(x_2,r_2) \subset B(x_1,r_1)\cap V_2, \quad 0 < r_2 \leq \frac{r_0}{2^2}$$

なる閉球 $\overline{B}(x_2,r_2)$ が取れる. 同様のことを繰り返せば閉球の列 $(\overline{B}(x_n,r_n))_{n\in\mathbb{N}}$ で,

$$\overline{B}(x_n,r_n) \subset B(x_{n-1},r_{n-1})\cap V_n, \quad 0 < r_n \leq \frac{r_0}{2^n} \quad (\forall n \in \mathbb{N}) \qquad (**)$$

を満たすものが取れる. このとき,

$$\bigcap_{n\in\mathbb{N}} \overline{B}(x_n,r_n) \subset \bigcap_{n\in\mathbb{N}} B(x_{n-1},r_{n-1}) \cap \bigcap_{n\in\mathbb{N}} V_n \subset B(x_0,r_0)\cap \bigcap_{n\in\mathbb{N}} V_n$$

である. よって $(*)$ を示すためには $\bigcap_{n\in\mathbb{N}} \overline{B}(x_n,r_n) \neq \emptyset$ を示せばよい. $m > n$ なる任意の $n,m \in \mathbb{N}$ に対して $(**)$ より,

$$d(x_m,x_n) \leq d(x_m,x_{m-1}) + \ldots + d(x_{n+1},x_n) \leq r_{m-1} + \ldots + r_n \leq 2\frac{r_0}{2^n}$$

である. これより $(x_n)_{n\in\mathbb{N}}$ は Cauchy 列であるので, $X$ の完備性よりある $x \in X$ に収束する. $x \in \bigcap_{n\in\mathbb{N}} \overline{B}(x_n,r_n)$ が成り立つことを示そう. 任意の $n \in \mathbb{N}$ と $\epsilon \in (0,\infty)$ に対して, $m \geq n$ かつ $d(x_m,x) < \epsilon$ を満たす $m \in \mathbb{N}$ を取れば, $x_m \in B(x,\epsilon) \cap \overline{B}(x_n,r_n) \neq \emptyset$ である. よって $\epsilon$ の任意性より $x \in \overline{B}(x_n,r_n)$ であり, $n \in \mathbb{N}$ の任意性より $x \in \bigcap_{n\in\mathbb{N}} \overline{B}(x_n,r_n)$ である. □

**系 9.13.2.** $X$ を空でない完備距離空間, $(F_n)_{n\in\mathbb{N}}$ を $X$ の閉集合からなる列とし, $X = \bigcup_{n\in\mathbb{N}} F_n$ とする. このときある $n \in \mathbb{N}$ に対して $F_n$ は内点を持つ.

**証明.** 任意の $n \in \mathbb{N}$ に対して $F_n$ が内点を持たないと仮定する. このとき, $V_n := X \setminus F_n$ $(\forall n \in \mathbb{N})$ とおけば, $(V_n)_{n\in\mathbb{N}}$ は $X$ において稠密な開集合からなる列である. 実際, ある $x \in X$ と $\varepsilon > 0$ に対して $B(x,\varepsilon)\cap V_n = \emptyset$ とすると $B(x,\varepsilon) \subset F_n$ となり $F_n$ が内点を持たないことに矛盾する. よって, Baire のカテゴリ定理より $\bigcap_{n\in\mathbb{N}} V_n$ は $X$ において稠密である. ところが $\bigcap_{n\in\mathbb{N}} V_n = X \setminus \bigcup_{n\in\mathbb{N}} F_n = \emptyset$ であるからこれは $X$ が空でないことに矛盾する. □

## 9.14 $C^\infty$ 級関数が多項式と一致するための必要十分条件

本節では開区間上の $C^\infty$ 級関数が $\mathbb{R}$ 係数多項式と一致するための必要十分条件について述べる. 以下では, ある開区間上で値が一致する二つの $\mathbb{R}$ 係数多項式は多項式として等しいという事実を断りなく用いる. さて, 簡単な必要十分条件としては, ある整数 $N \geq 0$ が存在して $N$ 階微分が恒等的に 0 となるというものがあるが, 実は次の定理が成立する. なお, 以下の証明は [5] を参考にした.

**定理 9.14.1.** $f$ を開区間 $(c,d)$ 上の $C^\infty$ 級関数とする. このとき, 多項式 $p$ が存在して $(c,d)$ 上 $f=p$ となるための必要十分条件は, 任意の $x \in (c,d)$ に対してある整数 $N_x \geq 0$ が存在して $f^{(N_x)}(x)=0$ となることである.

**証明.** 必要性は明らかである. 十分性を示す. 十分小さいすべての $\varepsilon > 0$ についてある多項式 $p$ が存在して $[c+\varepsilon, d-\varepsilon]$ 上 $f=p$ となることを示せばよいので, はじめから $[c,d]$ 上で $C^\infty$ 級として示せばよい (端点での導値は片側微分の意味で考える). 開集合 $G \subset [c,d]$ を次のように定義し $F := [c,d] \setminus G$ とおく.

$$G := \{x \in [c,d] \mid x \text{ の } [c,d] \text{ における開近傍 } U \text{ と多項式 } p \text{ が存在して } U \text{ 上 } f=p\}$$

$F=\emptyset$ となることを示せばよい. 実際, そのとき $G=[c,d]$ であり, ある多項式 $p_1, p_2$ と $\delta > 0$ があって $[c, c+2\delta)$ 上 $f=p_1$ かつ $(d-2\delta, d]$ 上 $f=p_2$ となる. そして, $a=c+\delta, b=d-\delta$ とおき, $[a,b] \subset (c,d)$ の各元 $x$ に対して $x$ を含む開区間 $I_x \subset (c,d)$ と多項式 $p_x$ で $I_x$ 上 $f=p_x$ なるものを取る. このとき $[a,b] \subset \bigcup_{x\in[a,b]} I_x$ であるので $[a,b]$ のコンパクト性より有限個の $x_1, \ldots, x_n \in [a,b]$ が存在して $[a,b] \subset I_{x_1} \cup \cdots \cup I_{x_n}$ となる. そこで $I_0 = [c, c+2\delta), I_1 = I_{x_1}, \ldots, I_n = I_{x_n}, I_{n+1} = (d-2\delta, d]$ とおくと $[c,d] = \bigcup_{k=0}^{n+1} I_k$ となる. よって $[c,d]$ の連結性より多項式として

$$p_1 = p_{x_1} = \cdots = p_{x_n} = p_2$$

がわかり $[c,d]$ 上 $f=p_1$ となる. よって, $F=\emptyset$ を示せばよい. 以下では $F \neq \emptyset$ と仮定して矛盾を導く.

**(Step1)**:$F$ が孤立点を持たないことを示す. $x_0 \in F$ が孤立点であるとして矛盾を導く.

(1):$x_0 = d$ の場合. ある $b \in (c,d)$ が存在して $(a, x_0) \subset G$ となる. このとき, ある $\delta > 0$ と多項式 $P$ が存在して $(a, a+\delta) \subset (a, x_0)$ 上 $f=P$ となる. そこで

$$y_0 := \sup\{y \in (a, x_0] \mid \forall x \in (a,y),\ f=P\}$$

とおくと $y_0 = x_0$ となる. 実際, $y_0 < x_0$ とすると $y_0 \in (a, x_0) \subset G$ であるので, ある $\varepsilon > 0$ と多項式 $Q$ が存在して $(y_0-\varepsilon, y_0+\varepsilon) \subset (a, x_0]$ 上で $f=Q$ となる. 一方, $y_0$ の定義から $y-\varepsilon < z \leq y_0$ で $(a,z)$ 上で $f=P$ となるものが取れる. 従って, $(y-\varepsilon, z)$ 上で $P=f=Q$ となるので多項式として $P=Q$ である. よって,

$(a, y_0+\varepsilon)$ 上で $f = Q$ となるがこれは $y_0$ の定義に反する．以上から $(a, x_0]$ 上で $f = P$ となる．よって, $(a.x_0]$ は $x_0$ の $[c, d]$ における開近傍であるので $x_0 \in G$ となる．これは不合理.
(2):$x_0 = c$ の場合．(1) と同様に矛盾が導かれる.
(3):$x_0 \in (c, d)$ の場合．ある $a, b \in (c, d)$ が存在して $[a, x_0), (x_0, b] \subset G$ となる．このとき, (1) の証明と同様にある多項式 $P$ が存在して $(a, x_0]$ 上で $f = P$ となる．また, ある多項式 $P'$ が存在して $[x_0, b)$ 上で $f = P'$ となる．このとき $P = P'$ でなければならない．このことを示すためにまず, $n := \deg(P) = \deg(P') =: m$ を示す．いま $(a, x_0]$ 上で $f$ は $n$ 次多項式なので $f^{(k)}(x_0) = 0$ $(k \geq n+1)$ である．一方, $[x_0, b)$ 上で $f$ は $m$ 次多項式なので $f^{(m)}(x_0) \neq 0$ である．そこでもし $n < m$ ならば $0 = f^{(m)}(x_0) \neq 0$ となる．これは不合理であるので $n \geq m$ である．同様に $n \leq m$ なので $n = m$ である．次に $x_0$ 中心の $f$ の Taylor 展開を考えると

$$P(x) = f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \quad (x \in (a, x_0])$$
$$P'(x) = f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \quad (x \in [x_0, b))$$

となる．これより

$$P(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \quad (x \in \mathbb{R})$$
$$P'(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \quad (x \in \mathbb{R})$$

となるから $P = P'$ である．以上から $(a, b)$ 上で $f = P$ である．これは $x_0 \in G$ を意味する．ところが $x_0 \in F$ なのでこれは不合理．よって $F$ は孤立点を持たない.
**(Step2)**:整数 $N \geq 0$ と長さ正の閉区間 $I \subset (c, d)$ が存在して $F \cap I \neq \emptyset$ かつ $F \cap I$ 上 $f^{(k)} = 0$ $(\forall k \geq N)$ となることを示す．整数 $n \geq 0$ に対して $E_n := \{x \in F \mid f^{(n)}(x) = 0\}$ とおく．$E_n$ は $F$ の閉集合であり, 定理の仮定より $F = \bigcup_{n=0}^{\infty} E_n$ である．いま $F \neq \emptyset$ は閉区間 $[c, d]$ の閉集合なので完備距離空間である．よって, Baire のカテゴリー定理 (系 9.13.2) よりある整数 $N \geq 0$ が存在して $E_N$ は内点を持つ．ゆえに長さ正の閉区間 $I \subset (c, d)$ が存在して $\emptyset \neq F \cap I \subset E_N$ となる．Step1 より $F$ は孤立点を持たないので $F \cap I$ も孤立点を持たない．ゆえに $x \in F \cap I$ に対して $x_n \to x$ なる $(F \cap I) \setminus \{x\}$ の元からなる列 $(x_n)_{n \in \mathbb{N}}$ が取れるので,

$$f^{(N+1)}(x) = \lim_{n \to \infty} \frac{f^{(N)}(x_n) - f^{(N)}(x)}{x_n - x} = \lim_{n \to \infty} \frac{0-0}{x_n - x} = 0$$

となる．よって帰納的に $F \cap I$ 上 $f^{(k)} = 0$ $(\forall k \geq N)$ となる.
**(Step3)**:Step2 の $N$ と $I$ について $I$ 上で $f^{(N)} = 0$ となることを示す．これが示さ

れれば $I \subset G$ であるので $\emptyset \neq F \cap I = \emptyset$ なる矛盾に到達する．さて, Step2 より $F \cap I$ 上で $f^{(N)} = 0$ であるので $G \cap I \neq \emptyset$ として $G \cap I$ 上で $f^{(N)} = 0$ となることを示せばよい．任意に $x_0 \in G \cap I$ をとる．$x_0$ を含む開区間 $(a, b) \subset G \cap I$ と多項式 $p$ が存在して $(a, b)$ 上 $f = p$ となる．ここで

$$a_0 = \inf\{y \in I \mid \forall x \in (y, b),\ f(x) = p(x)\}$$

とおく．すると, $a_0 \in F \cap I$ である．実際, $a_0 \in G \cap I$ であるならば, $I \subset (c, d)$ より $a_0$ は端点ではないので, 多項式 $p'$ と $\delta > 0$ が存在して $(a_0 - \delta, a_0 + \delta) \subset G \cap I$ 上で $f = p'$ となる．すぐわかるように $(a_0, a_0 + \delta)$ 上で $p' = f = p$ であるので多項式として $p = p'$ となり, 従って, $(a_0 - \delta, b)$ 上で $f = p$ となる．これは $a_0$ の定義に反する．よって $a_0 \in F \cap I$ である．ゆえに, Step2 より $f^{(k)}(a_0) = 0$ $(\forall k \geq N)$ となる．一方, $(a_0, b)$ 上で $f = p$ であるので $\deg(p) = m$ とおくと,

$$f(z) = \sum_{k=0}^{m} \frac{f^{(k)}(a_0)}{k!}(z - a_0)^k \quad (z \in (a_0, b))$$

となる．これと $f^{(k)}(a_0) = 0$ $(\forall k \geq N)$ を合わせると

$$f(z) = \sum_{k=0}^{N-1} \frac{f^{(k)}(a_0)}{k!}(z - a_0)^k \quad (z \in (a_0, b))$$

となる．よって, $f^{(N)}(x_0) = 0$ である． □

# 参考文献


[1] A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):930–945, 1993.

[2] Charles K. Chui and Xin Li. Approximation by ridge functions and neural networks with one hidden layer. *Journal of Approximation Theory*, 70:131–141, 1992.

[3] James A Clarkson. Uniformly convex spaces. *Transactions of the American Mathematical Society*, vol. 40(no. 3):396–414, 1936.

[4] C. Darken, M. Donahue, L. Gurvits, and E. Sontag. Rate of approximation results motivated by robust neural network learning. In *Proceedings of the Sixth Annual ACM Conference on Computational Learning Theory*, pages 303–309. ACM Press, New York, NY, 1993.

[5] William F. Donoghue. *DISTRIBUTION AND FOURIER TRANSFORMS*. Academic press New York and London, 1969.

[6] Weinan E, Chao Ma, and Lei Wu. Barron spaces and the compositional function spaces for neural network models, 2019.

[7] Ken-Ichi Funahashi. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 2(3):183 – 192, 1989.

[8] G.Cybenko. Approximation by superpositions of a sigmoidal function. *Control SignalsSystems*, 2, 1989.

[9] Edwin Hewitt and Karl Stromberg. *Real and Abstract Analysis*. Springer-Verlag Berlin Heidelberg, 1965.

[10] H.N.Mhaskar. Neural networks for optimal approximation of smooth and analytic functions. *Neural Computation*, 8:164–177, 1996.

[11] H.N.Mhaskar and Charles A. Micchelli. Degree of approximation by neural and translation networks with a single hidden layer. *advances in applied mathematics*, 16:151–183, 1995.

[12] Kurt Hornik. Multilayer feedforward networks are universal approximators. *Neural Networks*, Vol.2:359–366, 1989.

[13] Kurt Hornik. Some new results on neural network approximation. *Neural Networks*, Vol.6:1069–1072, 1993.

[14] Irie and Miyake. Capabilities of three-layered perceptrons. In *IEEE 1988 International Conference on Neural Networks*, pages 641–648 vol.1, 1988.

[15] Paul C. Kainen, Věra Kůrková, and Marcello Sanguineti. Approximating multivariable functions by feedforward neural nets. In M. Maggini M. Bianchini, editor, *Handbook on Neural Information Processing, vol. 49 of Intelligent Systems Reference Library*, chapter 5, page 143–181. Springer Berlin Heidelberg, 2013.

[16] Patrick Kidger and Terry Lyons. Universal approximation with deep narrow networks, 2019.

[17] V. Kůrková. High-dimensional approximation and optimization by neural networks. *Advances in Learning Theory: Methods, Models and Applications*, pages 69–88, 01 2003.

[18] V. Kůrková and M. Sanguineti. Geometric upper bounds on rates of variable-basis approximation. *IEEE Transactions on Information Theory*, 54(12):5681–5688, 2008.

[19] Věra Kůrková and Marcello Sanguineti. : Error estimates for approximate optimization by the extended ritz method. *SIAM J. on Optimization*, 15:261–287, 2005.

[20] Moshe Leshno, Vladimir YA. Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, vol.6:861–867, 1993.

[21] Vladimir Ya. Lin and Allan Pinkus. Fundamentality of ridge functions. *Journal of Approximation Theory*, 75:295–311, 1993.

[22] Maiorov and Pinkus. Lower bounds for approximation by MLP neural networks. *Neurocomputing*, 25:81–91, 1999.

[23] V. E. Maiorov and R. Meir. On the near optimality of the stochastic approximation of smooth functions by neural networks. *Advances in Computational Mathematics*, 13:79–103, 2000.

[24] Y. Makovoz. Random approximants and neural networks. *Journal of Approximation Theory*, 85(1):98 – 109, 1996.

[25] Noboru Murata. An integral representation of functions using three-layered networks and their approximation bounds. *Neural Networks*, 9(6):947–956, 1996.

[26] J. Park and I. W. Sandberg. Universal approximation using radial-basis-function networks. *Neural Computation*, 3:246–257, 1991.

[27] Allan Pinkus. TDI-Subspaces of $C(\mathbb{R}^d)$ and some Density Problems from Neural Networks. *Journal of Approximation Theory*, 85:269–287, 1996.

[28] Allan Pinkus. Approximation theory of the mlp model in neural networks. *Acta Numerica*, pages 143–195, 1999.

[29] Martin H. Schultz. $L^\infty$-multivariate approximation theory. *SIAM Journal on Numerical Analysis*, 6(2):161–183, 1969.

[30] Laurent Schwartz. Theorie generale des fonctions moyenne-periodiques. *Annals of Mathematics*, 48(4):857–929, 1947.

[31] Jonathan W. Siegel and Jinchao Xu. Approximation rates for neural networks with general activation functions. *Neural Networks*, 128:313 – 321, 2020.

[32] Sho Sonoda and Noboru Murata. Neural network with unbounded activation functions is universal approximator. 2015.

[33] Xingping Sun and E. W. Cheney. The fundamentality of sets of ridge functions. *Aequationes Mathematica*, 44:226–235, 1992.

[34] Itaï Ben Yaaco. A multivariate version of the vandermonde determinant identity. `https://hal.archives-ouvertes.fr/file/index/docid/983931/filename/MVVD.pdf`, 2014.

[35] Dmitry Yarotsky. Error bounds for approximations with deep relu networks, 2016.

[36] 宮島静雄. **関数解析**. 横浜図書, 2005.

[37] 宮島静雄. **ソボレフ空間の基礎と応用**. 共立出版, 2006.

[38] 藤田 宏, 黒田 成俊, and 伊藤清三. **岩波講座 基礎数学 関数解析**. 岩波書店, 1978.

[39] 高橋泰嗣. Norm inequalities and geometry of banach spaces. **京都大学数理解析研究所講究録**, 第 1399 巻:112–122, 2004.

[40] 松本幸夫. **多様体の基礎**. 東京大学出版会, 1988.

[41] 新井仁之. **新・フーリエ解析と関数解析学**. 培風館, 2010.

[42] 水口洋康 and 星啓介. Clarkson 不等式の最良定数について. **京都大学数理解析研究所講究録**, 第 1615 巻:135–142, 2008.

[43] 猪狩惺. **フーリエ級数**. 岩波書店, 1975.

[44] 猪狩惺. **実解析入門**. 岩波書店, 1996.