作用素論的アプローチによる時系列データ解析・ニューラルネットワークの汎化誤差解析

橋本　悠香

機械学習における重要なトピックである，時系列データ解析・ニューラルネットワークの理論解析の作用素論的アプローチが注目を集めている．作用素論的アプローチにより，時系列データの長期にわたる振る舞いの解析や，ニューラルネットワークの汎化誤差解析が可能となる．本稿では，これらの話題に関して，Krylov 部分空間法とのつながりを含めて解説する．

キーワード：Koopman 作用素，Perron–Frobenius 作用素，Krylov 部分空間法，ニューラルネットワーク

1. 序論

時系列データ解析・ニューラルネットワークの理論解析は，機械学習分野における最も重要なトピックの中に含まれ，活発に研究が進められている．これらに対する，作用素論を用いた方法によるアプローチが近年注目を集めている [1–4]．このような作用素論的アプローチにおいては，非線形関数を，Koopman 作用素または Perron–Frobenius 作用素と呼ばれる線形作用素により表現することで，線形代数的な方法により解析することを可能としている．

一方，数値線形代数分野においては，大規模行列の振る舞いを近似的に調べるために，Krylov 部分空間法の研究が活発に行われてきた [5–8]．これらの方法においては，Krylov 部分空間と呼ばれる部分空間を与えられた行列を用いて構成し，行列を Krylov 部分空間に射影することによって近似する．行列を有限次元空間上の線形作用素と考えれば，Krylov 部分空間法を，一般には無限次元な Hilbert 空間上の線形作用素に対して拡張することが可能である [9, 10]．

本稿では，作用素論的アプローチによる時系列データ解析・ニューラルネットワーク解析を，Krylov 部分空間法とのつながりを含めて解説する．Koopman 作用素や Perron–Frobenius 作用素を Krylov 部分空間法を用いて近似する方法について述べるとともに，近似作用素を用いて，時系列データの長期にわたる振る舞いが解析できることを述べる．また，Koopman 作用素を用いたニューラルネットワークの汎化誤差解析について解説し，上記のような近似作用素の応用可能性について議論する．

はしもと　ゆか
NTT ネットワークサービスシステム研究所
〒 180–8585　東京都武蔵野市緑町 3–9–11
yuka.hashimoto@ntt.com

2. 準備

本節では，基礎的な概念を導入し，その性質について解説する．

2.1 再生核 Hilbert 空間

再生核 Hilbert 空間 (RKHS) は，データの非線形性や高次モーメントを扱うために用いられてきた [11, 12]．本稿では，Koopman 作用素・Perron–Frobenius 作用素が定義される関数空間として，RKHS を用いる．

$\mathcal{X}$ を，空でない集合とする．集合 $\mathcal{X}$ はデータの集合に対応する．RKHS は，正定値カーネルと呼ばれる複素数値関数から構成される．

定義 2.1（正定値カーネル）**.** 関数 $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ が以下の 2 条件を満たすとき，正定値カーネルと呼ばれる．

1. $x, y \in \mathcal{X}$ に対して，$k(x, y) = \overline{k(y, x)}$
2. $n \in \mathbb{N}$, $\alpha_i \in \mathbb{C}$,$x_i \in \mathcal{X}$ に対して，$\sum_{i,j=1}^{n} \overline{\alpha}_i \alpha_j k(x_i, x_j) \geq 0$

ただし，複素数 $\alpha \in \mathbb{C}$ に対して $\overline{\alpha}$ は，α の複素共役を表す．

写像 $\phi : \mathcal{X} \to \mathbb{C}^{\mathcal{X}}$ を，$\phi(x) = k(\cdot, x)$ により定義する．ただし，$\mathbb{C}^{\mathcal{X}}$ は，$\mathcal{X}$ 上の複素数値関数全体の集合である．これを用いて以下の関数空間を定義する．

$$\mathcal{H}_{k,0} = \left\{ \sum_{i=1}^{n} \alpha_i \phi(x_i) \,\middle|\, n \in \mathbb{N},\ \alpha_i \in \mathbb{C},\ x_i \in \mathcal{X} \right\}$$

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

さらに，写像 $\langle\cdot,\cdot\rangle_k : \mathcal{H}_{k,0} \times \mathcal{H}_{k,0} \to \mathbb{C}$ を以下のように定義する．

$$\left\langle \sum_{i=1}^{n} \alpha_i \phi(x_i), \sum_{j=1}^{l} \beta_j \phi(y_j) \right\rangle_k = \sum_{i=1}^{n} \sum_{j=1}^{l} \overline{\alpha_i} \beta_j k(x_i, y_j)$$

定義 2.1 より，$\langle\cdot,\cdot\rangle_k$ は well-defined で，内積の性質を満たし，さらに，$v \in \mathcal{H}_{k,0}$ と $x \in \mathcal{X}$ に対して再生性

$$\langle \phi(x), v\rangle_k = v(x)$$

を満たすことが示せる．

空間 $\mathcal{H}_{k,0}$ の完備化は k に関する RKHS と呼ばれ，本稿では $\mathcal{H}_k$ と表す．内積 $\langle\cdot,\cdot\rangle_k$ は $\mathcal{H}_k$ に連続に拡張され，写像 $\mathcal{H}_k \ni v \mapsto (x \mapsto \langle\phi(x), v\rangle_k) \in \mathbb{C}^{\mathcal{X}}$ は単射であることが示せる．よって，$\mathcal{H}_k$ もまた $\mathbb{C}^{\mathcal{X}}$ の部分集合で，かつ再生性を満たすことがわかる．

写像 ϕ は，データを，一般には無限次元の Hilbert 空間 $\mathcal{H}_k$ に移す写像で，特徴写像と呼ばれる．データの空間は有限次元であることが多いが RKHS $\mathcal{H}_k$ は無限次元の空間であるため，データの非線形な振る舞いを，ϕ により，$\mathcal{H}_k$ 上の線形な振る舞いに変換することができる [12].

2.2 Koopman 作用素，Perron–Frobenius 作用素

Koopman 作用素は，非線形な力学系を表現する線形作用素である [1, 2, 13]．以下のような，空間 $\mathcal{X}$ 上の離散力学系を考える．

$$x_{t+1} = f(x_t) \quad (t = 0, 1, \ldots) \tag{1}$$

ただし，$f : \mathcal{X} \to \mathcal{X}$ は一般には非線形な力学系である．Hilbert 空間 $\mathcal{H} \subseteq \mathbb{C}^{\mathcal{X}}$ を考える．式 (1) に関する $\mathcal{H}$ 上の Koopman 作用素 K は，以下のように定義される．

$$Kv = v \circ f, \quad v \in \mathcal{D}$$

ただし，$\mathcal{D} = \{v \in \mathcal{H} \mid v \circ f \in \mathcal{H}\}$ である．Hilbert 空間 $\mathcal{H}$ として，RKHS [2] や測度空間 $\mathcal{X}$ に対する $L^2(\mathcal{X})$ [14] が採用されることが多い．RKHS 上では，Koopman 作用素の随伴作用素にあたる Perron–Frobenius 作用素も考えられている．正定値カーネル $k : \mathcal{X} \times \mathcal{X} \to \mathbb{C}$ に関する RKHS $\mathcal{H}_k$ 上の Perron–Frobenius 作用素 L は，特徴写像 $\phi(x) = k(\cdot, x)$ を用いて以下のように定義される線形作用素である．

$$L\phi(x) = \phi(f(x)), \quad x \in \mathcal{X} \tag{2}$$

2.3 Krylov 部分空間法

Krylov 部分空間法は，線形作用素の振る舞いを近似する方法としてさまざまな研究が行われてきた [5, 8, 9]．Hilbert 空間 $\mathcal{H}$ 上の線形作用素 A とベクトル $v \in \mathcal{H}$ に対して，m 次元 Krylov 部分空間 $\mathcal{K}_m(A, v)$ は以下のように定義される．

$$\mathcal{K}_m(A, v) = \mathrm{Span}\{v, Av, \ldots, A^{m-1}v\}$$

ただし，$\{v, Av, \ldots, A^{m-1}v\}$ は一次独立であるとする．空間 $\mathcal{K}_m(A, v)$ は，A に関する多項式の空間とみなすこともできる．ベクトル $q_1, \ldots, q_m \in \mathcal{H}$ を，Krylov 部分空間 $\mathcal{K}_m(A, v)$ の正規直交基底とし，$Q_m : \mathbb{C}^m \to \mathcal{H}$ を，$Q_m[c_1, \ldots, c_m] = \sum_{i=1}^{m} c_i q_i$ により定める．このとき，$Q_m Q_m^*$ は，$\mathcal{H}$ から $\mathcal{K}_m(A, v)$ への射影となっていることがわかる．ここで，Q^* は Q の随伴作用素を表す．Krylov 部分空間法では，作用素 A のベクトル $u \in \mathcal{H}$ への作用を，u を $\mathcal{K}_m(A, v)$ へ射影した上で A を作用させ，得られたベクトルをさらにもう一度 $\mathcal{K}_m(A, v)$ へ射影することによって近似する．つまり，A を $Q_m Q_m^* A Q_m Q_m^*$ により近似する．このとき，近似である作用素 $Q_m Q_m^* A Q_m Q_m^*$ の正規直交基底 $\{q_1, \ldots, q_m\}$ に関する表現行列は，$Q_m^* A Q_m \in \mathbb{C}^{m \times m}$ となる．

3. Koopman 作用素と Krylov 部分空間法：時系列データ解析

Koopman 作用素や Perron–Frobenius 作用素の特徴の一つは，力学系 (1) を知らなかったとしても，それに従う時系列データが得られれば，作用素の近似が得られることである．本節では，特に RKHS 上の Perron–Frobenius 作用素に焦点を当て，その近似が Krylov 部分空間法によって実現されることを解説する．さらに，作用素の近似を固有値分解することにより，力学系の長期にわたる挙動を捉えることができることを紹介する．

3.1 Krylov 部分空間法による Perron–Frobenius 作用素の近似

時系列データ $x_0, x_1, \ldots \in \mathcal{X}$ が与えられているときに，そのデータが従う力学系に関する RKHS $\mathcal{H}_k$ 上の Perron–Frobenius 作用素 L を推定することを考える．以下のような Krylov 部分空間 $\mathcal{K}_m(L, \phi(x_0))$ を考える．

$$\mathcal{K}_m(L, \phi(x_0)) = \mathrm{Span}\{\phi(x_0), \ldots, \phi(x_{m-1})\}$$

ただし，$\{\phi(x_0), \ldots, \phi(x_{m-1})\}$ は一次独立とする．Gram–Schmidt の正規直交化により，$\mathcal{K}_m(L, \phi(x_0))$ の正規直交基底 $\{q_1, \ldots, q_m\}$ を得る．正則行列 $R_m \in \mathbb{C}^{m \times m}$ が存在して，$\Phi(x_0, \ldots, x_{m-1})R_m = Q_m$ が成

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

り立つ．ここで，$\Phi(x_0,\ldots,x_{m-1}) : \mathbb{C}^m \to \mathcal{H}$ は，$\Phi(x_0,\ldots,x_{m-1})[c_1,\ldots,c_m] = \sum_{i=1}^m c_i\phi(x_i)$ により定められる線形作用素である．2.3節で議論したように，Perron–Frobenius 作用素 L の近似 L_m を，$Q_mQ_m^*$ を用いて $L_m = Q_mQ_m^*LQ_mQ_m^*$ で定義する．近似 L_m の正規直交基底 $\{q_1,\ldots,q_m\}$ に関する行列表現 $\mathbf{L}_m$ は，L の定義を用いると，以下のように与えられたデータのみを用いて計算できる．

$$\begin{aligned}\mathbf{L}_m &= R_m^*\Phi(x_0,\ldots,x_{m-1})^*L\Phi(x_0,\ldots,x_{m-1})R_m\\ &= R_m^*\Phi(x_0,\ldots,x_{m-1})^*\Phi(x_1,\ldots,x_m)R_m\\ &= R_m^*G_mR_m\end{aligned}$$

ただし，$G \in \mathbb{C}^{m\times m}$ は，(i,j) 成分が $k(x_i,x_{j+1})$ により定義される行列である．

3.2 モード分解への応用

Perron–Frobenius 作用素の近似 L_m の固有値 $\lambda_1,\ldots,\lambda_l$ と対応する固有ベクトルを用いた分解を考える．この分解はモード分解と呼ばれ，力学系の長期にわたる挙動を捉えることを可能とする．固有値 λ_i に対応する Jordan ブロックを J_i，広義固有空間の次元を n_i とする．行列 $\mathbf{L}_m$ の Jordan 分解を $V_m\,\mathrm{diag}\{J_1,\ldots,J_l\}V_m^{-1}$ とする．観測関数 $g \in \mathcal{H}_k$ に対して $g(x_t)$ は以下のように近似的に分解される．

$$\begin{aligned}g(x_t) &= \langle\phi(x_t),g\rangle = \langle L^t\phi(x_0),g\rangle \approx \langle L_m^t\phi(x_0),g\rangle\\ &= \langle Q_mV_m\,\mathrm{diag}\{J_1,\ldots,J_l\}^tV_m^{-1}Q_m^*\phi(x_0),g\rangle\\ &= \sum_{i=1}^{l}\langle \mathbf{v}_iJ_i^t\mathbf{u}_i,Q_m^*g\rangle \qquad (3)\end{aligned}$$

ただし，$\mathbf{u} = [u_1,\ldots,u_m]^T = V_m^{-1}Q_m^*\phi(x_0)$ に対し，$\mathbf{u}_i = [u_{\sum_{j=1}^{i-1}n_j+1},\ldots,u_{\sum_{j=1}^{i}n_j}]^T$ である．また，$V_m = [v_1,\ldots,v_m]$ に対し，$\mathbf{v}_i = [v_{\sum_{j=1}^{i-1}n_j+1},\ldots,v_{\sum_{j=1}^{i}n_j}]$ である．式 (3) において時刻 t に依存するのは J_i^t のみである．よって，$t\to\infty$ における J_i^t の振る舞いによって，$g(x_t)$ を成分に分解することができる．具体的には，λ_i が単純固有値でかつ大きさ 1 である場合，式 (3) の対応する項は時刻 t が十分大きくなっても残り，振動する成分である．一方，λ_i の大きさが 1 より真に小さい場合，対応する項は時刻 t の増加とともに減衰する成分である．

3.3 ランダム性のある力学系に対する Perron–Frobenius 作用素

これまで，決定的な力学系に対する Koopman 作用素や Perron–Frobenius 作用素を考えてきたが，これをランダム性のある力学系に拡張することに対しても研究がされている [14–16]．ここでは，文献 [15] における定義を紹介する．確率空間 $(\Omega,\mathcal{F},P)$，局所コンパクト Hausdorff 空間 $\mathcal{X}$，$\mathcal{X}$ 上の Borel 集合体 $\mathcal{B}$ に対して，$\mathcal{X}$ 上の確率変数 $X_t : \Omega \to \mathcal{X}$ が以下を満たすとする．

$$X_{t+1} = f(X_t,\xi_t)\quad (t=0,1,\ldots) \qquad (4)$$

ただし，$f : \mathcal{X}\times\mathcal{X}\to\mathcal{X}$ は可測で一般には非線形な写像である．また，$\xi_t : \Omega\to\mathcal{X}$ はノイズを表す，X_0 と独立な確率変数で，$\{\xi_t\}_{t=0}^\infty$ は独立同分布であるとする．確率変数 X_t に対する押し出し測度（分布）を $X_{t*}P$ で表す．測度 $X_{t*}P$ に関する Perron–Frobenius 作用素を RKHS $\mathcal{H}_k$ 上で定義するために，カーネル平均埋め込みを用いる．局所コンパクト Hausdorff 空間 $\mathcal{X}$ 上の有限かつ正則な複素数値 Borel 測度を $\mathcal{M}(\mathcal{X})$ とすると，$\mu\in\mathcal{M}(\mathcal{X})$ のカーネル平均埋め込み $\Phi(\mu)$ は，任意の $v\in\mathcal{H}$ に対して $\langle u,v\rangle = \int_{x\in\mathcal{X}}v(x)\mathrm{d}\mu(x)$ を満たす一意なベクトル $u\in\mathcal{H}_k$ によって定義される．RKHS $\mathcal{H}_k$ を構成する正定値カーネル k が連続かつ有界で，c_0-universal と呼ばれるクラスに属していれば，写像 $\Phi : \mathcal{M}(\mathcal{X})\to\mathcal{H}_k$ は well-defined でかつ単射になる．これを用いて，$\mathcal{H}_k$ 上の Perron–Frobenius 作用素 L は以下のように定義される．

$$L\Phi(\mu) = \Phi(f_{t*}(\mu\otimes P)),\quad \mu\in\mathcal{M}(\mathcal{X}) \qquad (5)$$

ただし，$f_t(x,\omega) = f(x,\xi_t(\omega))$ で，$\{\xi_t\}_{t=0}^\infty$ は独立同分布であることから，L は時刻 t に依存しない．また，式 (4) と，X_0 と ξ_t が独立であることより，$L\Phi(X_{t*}P) = \Phi(X_{t+1*}P)$ が成り立ち，L は測度 $X_{t*}P$ の時間発展を表しているといえる．写像 f が ξ_t に依存しない場合，式 (5) において μ を $x\in\mathcal{X}$ における Dirac 測度とすると，式 (2) が復元されることに注意する．

4. Koopman 作用素とニューラルネットワーク：汎化誤差解析

Koopman 作用素の時系列データ解析以外への応用として，ニューラルネットワークの解析がある．ニューラルネットワークは関数の合成を用いて構成されるが，Koopman 作用素を用いると，合成を線形作用素の積として記述することが可能となる．本節では，この性質を用いて，Koopman 作用素をニューラルネットワークの汎化誤差解析に応用することについて述べる．

4.1 Koopman 作用素を用いたニューラルネットワークの表現

ここでは，L レイヤーから成るニューラルネット

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

ワークを考える．整数 d_0 を，入力データ空間の次元とする．$j = 1, \dots, L$ に対して，$d_j \in \mathbb{N}$ とし，$W_j : \mathbb{R}^{d_{j-1}} \to \mathbb{R}^{d_j}$ を線形作用素，$b_j : \mathbb{R}^{d_j} \to \mathbb{R}^{d_j}$ を，$x \mapsto x + a_j$ により定義されるシフト作用素とする．ここで，$a_j \in \mathbb{R}^{d_j}$ はバイアスと呼ばれる．また，非線形関数 $\sigma_j : \mathbb{R}^{d_j} \to \mathbb{R}^{d_j}$ を考える．関数 σ_j は，活性化関数と呼ばれる．さらに，$g : \mathbb{R}^{d_L} \to \mathbb{C}$ を，非線形関数とする．以下のような，線形変換と非線形変換を繰り返す，最も基本的なニューラルネットワーク f を考える．

$$f = g \circ b_L \circ W_L \circ \sigma_{L-1} \circ b_{L-1} \circ W_{L-1} \circ \cdots \circ \sigma_1 \circ b_1 \circ W_1 \tag{6}$$

関数空間 $H_p(\mathbb{R}^d) = \{f \in L^2(\mathbb{R}^d) \mid \hat{f}/\sqrt{p} \in L^2(\mathbb{R}^d)\}$ を，内積 $\langle f, g \rangle_{H_p(\mathbb{R}^d)} = \int_{\mathbb{R}^d} \hat{f}(\omega)\hat{g}(\omega)/p(\omega) \mathrm{d}\omega$ の構造をもつ Hilbert 空間とする．ここで，$\hat{f}$ は，f の Fourier 変換である．このとき，$k_p(x, y) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}(x-y)\cdot\omega} p(\omega) \mathrm{d}\omega$ は，$H_p(\mathbb{R}^d)$ を構成する正定値カーネルであることがわかる．$j = 0, \dots, L$ に対して，p_j を，$\mathbb{R}^{d_j}$ 上の有限な非負測度の密度関数とする．変換 W_j，b_j，σ_j に関する Koopman 作用素 $K_{W_j} : H_{p_j}(\mathbb{R}^{d_j}) \to H_{p_{j-1}}(\mathbb{R}^{d_{j-1}})$，$K_{b_j}, K_{\sigma_j} : H_{p_j}(\mathbb{R}^{d_j}) \to H_{p_j}(\mathbb{R}^{d_j})$ を用いると，ニューラルネットワーク f は以下のように表すことができる．

$$f = K_{W_1} K_{b_1} K_{\sigma_1} \cdots K_{W_{L-1}} K_{b_{L-1}} K_{\sigma_{L-1}} K_{W_L} K_{b_L} g$$

2.2 節では，Koopman 作用素は同じ空間から同じ空間への作用素としてのみ考えていたが，一般には，定義されている空間と行先の空間が異なってもよい．ここでは，たとえば K_{W_j} は，$h \in H_{p_j}(\mathbb{R}^{d_j})$ を $h \circ W$ に移す線形作用素である．このように，ニューラルネットワークの合成の構造を Koopman 作用素の積を用いて表すことで，ニューラルネットワーク全体の情報を各関数 W_j，b_j，σ_j の情報に分割することが容易となり，次節で議論するように，汎化誤差の解析に役立つ．記法の単純化のため，以降では，$H_{p_j}(\mathbb{R}^{d_j})$ を H_j と書くこととする．以降では，次のような仮定をおく．

仮定 4.1. 非線形関数 g は H_L に含まれ，さらに，K_{σ_j} は有界である．

仮定 4.2. ある定数 $B > 0$ が存在して，任意の $x \in \mathbb{R}^d$ に対して $|k_{p_0}(x, x)| \le B^2$ である．

集合 F を，式 (6) の形で表現され，仮定 4.1 を満たすすべての関数からなる集合とする．

4.2 ニューラルネットワークの汎化誤差解析への応用

汎化誤差（汎化ギャップとも呼ばれる）とは，経験的な誤差（有限個のデータにより得られる誤差）と真の誤差（無限個のデータがあったと仮定した場合の誤差）との差である．汎化誤差を評価するために用いられる一つの有用な指標として，Rademacher 複雑度がある．Rademacher 複雑度は，関数クラスがどの程度複雑な関数を含むかを表す指標である．複雑な関数が含まれるほど，経験的な誤差は真の誤差から乖離する可能性が高くなるため，Rademacher 複雑度の上界を求めることで，汎化誤差の上界を求めることが可能となる [17]．本節では，簡単のため，線形変換 W_j が可逆である場合（つまり，ある $d \in \mathbb{N}$ に対して $d_j = d$ の場合）に着目する．

空間 (Ω, P) を，確率測度空間とする．可測関数 s に対して，その Ω 上での積分 $\int_\Omega s(\omega) \mathrm{d}P(\omega)$ を，$\mathrm{E}[s]$ と表すこととする．確率変数 $s_1, \dots, s_n$ を，Rademacher 変数，つまり，$\{-1, 1\}$ 上の一様分布に従う確率変数とする．関数クラス G と，データ $x_1, \dots, x_n \in \mathbb{R}^d$ に対して，経験 Rademacher 複雑度は，以下で定義される．

$$\hat{R}_n(\mathbf{x}, G) = \frac{1}{n} \mathrm{E}\left[\sup_{g \in G} \sum_{i=1}^{n} s_i g(x_i) \right]$$

ただし，$\mathbf{x} = (x_1, \dots, x_n)$ である．確率変数 s_i と $g(x_i)$ の符号が一致すれば，$\sum_{i=1}^n s_i g(x_i)$ の値は大きくなる．よって，Rademacher 複雑度は，-1 と 1 を一様にランダムに取る確率変数 $s_1, \dots, s_n$ に，関数 $g \in G$ がどのくらいフィットできるかの度合を表している．確率変数 s_i と $g(x_i)$ の符号を一致させるような関数 g が存在するほど，関数クラス G は複雑な関数を含んでいるといえる．

Rademacher 複雑度の上界に関して，次の定理が成り立つ [4](Theorem 2)．正の定数 $C, D > 0$ に対して，線形変換の集合 $\mathcal{W}(C, D) = \{W \in \mathbb{R}^{d \times d} \mid \|W\| \le C, |\det(W)| \ge D\}$ と，関数クラス $F_{\mathrm{inv}}(C, D) = \{f \in F \mid W_j \in \mathcal{W}(C, D)\}$ を定める．

定理 4.3. 関数クラス $F_{\mathrm{inv}}(C, D)$ に対する Rademacher 複雑度 $\hat{R}_n(\mathbf{x}, F_{\mathrm{inv}}(C, D))$ は，以下のように評価できる．

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

$$\hat{R}_n(\mathbf{x}, F_{\mathrm{inv}}(C,D)) \leq \frac{B\|g\|_{H_L}}{\sqrt{n}} \sup_{W_j \in \mathcal{W}(C,D)} \prod_{j=1}^{L} \left\| \frac{p_j}{p_{j-1} \circ W_j^*} \right\|_\infty^{1/2} \frac{\|K_{\sigma_j}\|}{\sqrt{|\det(W_j)|}} \tag{7}$$

ただし，$\|\cdot\|_\infty$ は，sup ノルムである．

証明には，次の補題と，再生核 Hilbert 空間 H_0 における再生成を用いる．

補題 4.4. $j = 1, \ldots, L$ に対して，線形変換 $W_j : \mathbb{R}^d \to \mathbb{R}^d$ は可逆であるとする．このとき，以下が成立する．

$$\|K_{W_j}\| \leq \left(\left\| \frac{p_j}{p_{j-1} \circ W_j^*} \right\|_\infty \frac{1}{|\det(W_j)|} \right)^{1/2}$$
$$\|K_{b_j}\| = 1$$

Rademacher 複雑度の上界を求める際に Koopman 作用素のノルムが現れ，このノルムに対して補題 4.4 を用いることで，定理 4.3 が導かれる．

Koopman 作用素が定義される空間を定める確率密度関数 p_j が $p_j(\omega) = 1/(1+\|\omega\|^2)^s$ であるとき，つまり，H_j が Sobolev 空間であるとき，式 (7) 中の $\|p_j/(p_{j-1} \circ W_j^*)\|_\infty$ に関して，以下の補題が成立する．

補題 4.5. $j = 0, \ldots, L$ に対して，$p_j = p$，$p(\omega) = 1/(1+\|\omega\|^2)^s$，ただし，$s > d/2$ とする．このとき，$\|p/(p \circ W_j^*)\|_\infty \leq \max\{1, \|W_j\|^{2s}\}$ が成り立つ．

よって，H_j が Sobolev 空間であるとき，式 (7) より，Rademacher 複雑度の上限は $|\lambda_{j,1}|^s/|\lambda_{j,1} \cdots \lambda_{j,d}|$ という値によって支配されることがわかる．ただし，$\lambda_{j,1} \geq \cdots \geq \lambda_{j,d}$ は，W_j の固有値である．これまでに，ニューラルネットワークの汎化誤差の上界を与える研究はさかんに行われてきた [18–20] が，ほとんどの結果は，W_j に小さな固有値または特異値があるほど，小さくなるような上界であった．一方で，本結果は，W_j に大きな固有値があっても汎化誤差が小さくなり得ることを示唆している．特に，W_j がユニタリ行列である場合，$|\lambda_{j,1}|^s/|\lambda_{j,1} \cdots \lambda_{j,d}| = 1$ となり，式 (7) の上界は最小となる．W_j に大きな固有値や特異値があっても汎化誤差が小さくなるという事実は，実験的には知られていた [21] が，定理 4.3 は，その事実に対する理論的な説明を与えたといえる．

上記のように，理論的に新たな結果を与えている一方，定理 4.3 で示されている上界を実際に計算するのは困難な場合もある．まず，W_j が大規模な場合には，その $\det(W_j)$ を計算するのは計算コストが高い．この場合，$\det(W_j)$ を近似的にコスト少なく計算することが求められる．また，活性化関数に関する Koopman 作用素 K_{σ_j} のノルムの値を実際に計算したり，その上界を与えるのは一般には難しい．そこで，3.1 節で述べたように，Krylov 部分空間法を用いて K_{σ_j} を近似するなどの工夫が必要となる．

本節では，簡単のため，線形変換 W_j は可逆であると仮定したが，W_j が単射の場合にも，同様の結果が導かれる [4](Theorem 5)．また，W_j が単射でない場合には，Koopman 作用素 K_{W_j} が非有界になるという困難性があるが，W_j の代わりに $\tilde{W}_j(x_1, x_2) = (W_j x_1, P_j x_1, x_2)$ により定義される単射な線形作用素 $\tilde{W}_j$ や，$\tilde{K}_{W_j} h = \psi_j \cdot h \circ W_j$ により定義される重み付き Koopman 作用素 $\tilde{K}_{W_j}$ を用いて類似の結果を得ることができる [4](Section 4.3)．ただし，P_j は $\ker(W_j)$ への射影，$\psi_j : \mathbb{R}^{d_j} \to \mathbb{C}$ は，$\ker(W_j)$ において遠方で減衰する関数である．

5. 結論

本稿では，作用素論的アプローチによる時系列データ解析・ニューラルネットワーク解析を，Krylov 部分空間法とのつながりを含めて解説した．Krylov 部分空間法を用いることで，与えられた時系列データから，そのデータが従う力学系に関する Perron–Frobenius 作用素が近似できることを述べ，それを用いて力学系の長期にわたる振る舞いを解析する方法について述べた．また，Koopman 作用素を用いてニューラルネットワークの汎化誤差の上界を与えることが可能となることを述べ，具体的な上界の計算における，Koopman 作用素の近似の有用性を議論した．今後，機械学習分野と数値線形代数分野の融合による，各分野のさらなる発展が期待される．

参考文献

[1] I. Mezić, “Spectral properties of dynamical systems, model reduction and decompositions,” *Nonlinear Dynamics*, **41**, pp. 309–325, 2005.

[2] Y. Kawahara, “Dynamic mode decomposition with reproducing kernels for Koopman spectral analysis,” In *Proceedings of the 30th Conference on Neural Information Processing Systems (NIPS)*, 2016.

[3] T. Konishi and Y. Kawahara, “Stable invariant models via Koopman spectra,” *Neural Networks*, **165**, pp. 393–405, 2023.

 Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.

[4] Y. Hashimoto, S. Sonoda, I. Ishikawa, A. Nitanda and T. Suzuki, "Koopman-based generalization bound: New aspect for full-rank weights," In *Proceedings of the 12th International Conference on Learning Representations (ICLR)*, 2024.
[5] A. N. Krylov, "On the numerical solution of the equation by which, in technical matters, frequencies of small oscillations of material systems are determined," *Izvestija Akad. Nauk SSSR*, **7**, pp. 491–539, 1931. (in Russian).
[6] M. R Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," *Journal of Research of the National Bureau of Standards*, **49**, pp. 409–436, 1952.
[7] Y. Saad and M. H. Schultz, "GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems," *SIAM Journal on Scientific and Statistical Computing*, **7**, pp. 856–869, 1983.
[8] E. Gallopoulos and Y. Saad, "Efficient solution of parabolic equations by Krylov approximation methods," *SIAM Journal on Scientific Statistics*, **13**, pp. 1236–1264, 1992.
[9] S. Güttel, "Rational Krylov methods for operator functions," PhD Thesis, Techniche Universität Bergakademie Freiberg, 2010.
[10] Y. Hashimoto and T. Nodera, "Shift-invert rational Krylov method for an operator ϕ-function of an unbounded linear operator," *Japan Journal of Industrial and Applied Mathematics*, **36**, pp. 421–433, 2019.
[11] S. Saitoh and Y. Sawano, *Theory of Reproducing Kernels and Applications*. Springer Singapore, 2016.
[12] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*, MIT Press, 2001.
[13] B. O. Koopman, "Hamiltonian systems and transformation in Hilbert space," *Proceedings of the National Academy of Sciences*, **17**, pp. 315–318, 1931.
[14] N. Črnjarić-Žic, S. Maćešić and I. Mezić, "Koopman operator spectrum for random dynamical systems," *Journal of Nonlinear Science*, **30**, pp. 2007–2056, 2020.
[15] Y. Hashimoto, I. Ishikawa, M. Ikeda, Y. Matsuo and Y. Kawahara, "Krylov subspace method for nonlinear dynamical systems with random noise," *Journal of Machine Learning Research*, **21**, pp. 172, 2020.
[16] V. Kostic, K. Lounici, P. Novelli and M. Pontil, "Sharp spectral rates for Koopman operator learning," In *Proceedings of the 37th Conference on Neural Information Processing Systems (NeurIPS)*, pp. 32328–32339, 2023.
[17] M. Mohri, A. Rostamizadeh and A. Talwalkar, *Foundations of Machine Learning*, MIT press, 2018.
[18] B. Neyshabur, R. Tomioka and N. Srebro, "Norm-based capacity control in neural networks," In *Proceedings of the 2015 Conference on Learning Theory (COLT)*, pp. 1376–1401, 2015.
[19] P. L. Bartlett, D. J. Foster and M. J. Telgarsky, "Spectrally-normalized margin bounds for neural networks," In *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS)*, 2017.
[20] N. Golowich, A. Rakhlin and O. Shamir, "Size-independent sample complexity of neural networks," In *Proceedings of the 2018 Conference On Learning Theory (COLT)*, 2018.
[21] M. Goldblum, J. Geiping, A. Schwarzschild, M. Moeller and T. Goldstein, "Truth or backpropaganda? An empirical investigation of deep learning theory," In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2020.

Copyright © by ORSJ. Unauthorized reproduction of this article is prohibited.