arXiv:2304.00490v1 [cs.LG] 2 Apr 2023

无限维储备池计算

Lukas Gonon[1], Lyudmila Grigoryeva[2,3],和 Juan-Pablo Ortega[4]

摘要

本文针对一类新的输入/输出系统概念类，证明了储备池计算的逼近界和泛化界。该概念类将所谓的广义 Barron 泛函扩展到了动态语境中，其特征是读出层具有基于无限维状态空间系统构建的某种积分表示。研究表明，这一概念类十分丰富，具有实用特性和通用逼近性质。用于逼近和估计该类中元素的储备池架构，是随机生成的回声状态网络，并采用线性或 ReLU 激活函数。其读出层由随机生成的神经网络构建，且只训练输出层（即极限学习机或随机特征神经网络）。本文的结果给出了一种可完全实现的、基于循环神经网络的学习算法，该算法具有可证明的收敛保证，且不受维数灾难的影响。

关键词：循环神经网络、储备池计算、回声状态网络、ESN、极限学习机、ELM、循环线性网络、机器学习、Barron 泛函、循环 Barron 泛函、普适性、有限记忆泛函、逼近界、卷积滤波器。

目录

1 引言 **2**

2 广义 Barron 函数的动态类比 3
2.1 记号.3 2.2 循环 Barron 泛函的定义与性质.4 2.3 循环 Barron 泛函的示例.7

3 逼近 **11**
3.1 通过有限维系统进行逼近.11 3.2 归一化实现.13 3.3 回声状态网络对归一化截断泛函的逼近.15
3.4 主要逼近结果.17 3.5 普适性.22 3.6 特例：静态情形.24

4 从单条轨迹中学习 25

5 结论 30

致谢 **31**

参考文献 **31**

[1]帝国理工学院。数学系。伦敦。英国。 l.gonon@imperial.ac.uk[2]圣加仑大学。数学与统计学院。圣加仑。瑞士。 lyudmila.grigoryeva@unisg.ch[3]华威大学。统计系。英国。 lyudmila.grigoryeva@warwick.ac.uk[4]南洋理工大学。数学科学部。新加坡。 Juan-Pablo.Ortega@ntu.edu.sg

1 引言

储备池计算（RC） [Jaeg 10, Maas 02, Jaeg 04, Maas 11] ，尤其是回声状态网络（ESN） [Matt 92, Matt 93, Jaeg 04] ，近年来因其在动力系统预测方面的卓越表现而广受欢迎 [Grig 14, Jaeg 04, Path 17, Path 18, Lu 18, Wikn 21, Arco 22] ，并且实现起来十分简便。储备池计算旨在利用随机生成的状态空间系统（称为储备池）来逼近非线性输入/输出系统，其中只需估计线性读出层。理论研究已证实，在多种确定性和随机情形下，这确实是可行的 [Grig 18b, Grig 18a, Gono 20c, Gono 21b, Gono 23] ；在这些情形下，RC系统已被证明具有通用逼近性质。

本文重点推导刚才提及的某类架构变体的误差界，并将随机生成的线性系统作为逼近器，其读出层由随机生成的神经网络构成，且仅训练输出层。因此，从学习的角度看，我们结合了线性回声状态网络与文献中所称的随机特征[Rahi 07] /极限学习机（ELM） [Huan 06]。我们为一个新引入的概念类给出了明确且易于计算的逼近界和估计界；我们将其元素称为递归（广义）Barron 泛函，因为它们可视为 [Barr 92、Barr 93] 中引入并在后来由 [E 20b、E 20a、E 19]扩展的（广义）Barron 函数的动力学类比。与文献中已有的其他概念类相比，例如确定性设置下的衰减记忆类 [Boyd 85] ，或随机情形下的 L^p 泛函，该概念类的主要新颖之处在于，其元素具有显式的无限维状态空间表示，因此在解析上易于处理。正如我们稍后将看到的，许多有趣的输入/输出系统族都属于这一概念类；此外，该类在 L^p-意义下具有普适性。

从逼近理论的角度来看，早已有人研究读出层属于稠密族的线性系统的普适性，无论是在确定性设置中 [Boyd 85, Grig 18b] ，还是在随机设置中 [Gono 20c] 。相应的动力学性质和记忆性质也已得到广泛研究（例如，参见 [Herm 10, Coui 16, Tino 18, Tino 20, Gono 20a, Li 22]及其中的参考文献）。因此，本文的贡献可视为将这些工作扩展至循环 Barron 类。

迄今为止所引用的动态学习工作都只采用有限维状态空间。因此，我们这项工作的一大创新之处，在于所提出的概念类中包含无限维成分。值得一提的是，尽管在静态设置中已有许多关于神经函数逼近的结果（例如，参见 [Chen 95, Stin 99, Krat 20, Bent 22, Cuch 22, Neuf 22]，以及上述关于Barron函数的研究），但无限维状态空间系统尚未得到广泛应用，直到最近才开始受到认真发展。少数例子可参见 [Herm 12, Bouv 17b, Bouv 17a, Kira 19, Li 20, Kova 21, Acci 22, Gali 22, Hu 22, Salv 22] 。利用量子系统实现 RC系统的专用硬件，是这些扩展的一种潜在应用；由于缺乏适当的工具，迄今为止，大多数理论研究都仅在有限维设置中开展 [Chen 19, Chen 20, Tran 20, Tran 21, Chen 22, Mart 23]。

为了更详细地介绍我们的贡献，我们首先回顾一下：回声状态网络（ESN）是由以下两个方程定义的输入/输出系统:

$$\mathbf{x}_t = \sigma(A\mathbf{x}_{t-1} + C\mathbf{z}_t + \boldsymbol{\zeta}), \tag{1.1}$$
$$\mathbf{y}_t = W\mathbf{x}_t, \tag{1.2}$$

对于 $t \in \mathbb{Z}_-$，其中 $d, m, N \in \mathbb{N}$，$\mathbf{z} \in (\mathbb{R}^d)^{\mathbb{Z}_-}$，$\mathbf{x}_t \in \mathbb{R}^N$，矩阵 $W \in \mathbb{M}_{m,N}$ 可训练，σ 表示对给定激活函数 $\sigma\colon \mathbb{R} \to \mathbb{R}$ 的逐分量应用，而 $A \in \mathbb{M}_N$, $C \in \mathbb{M}_{N,d}$ 和 $\boldsymbol{\zeta} \in \mathbb{R}^N$ 是随机生成的。若对每个 $\mathbf{z} \in (\mathbb{R}^d)^{\mathbb{Z}_-}$，方程 (1.1) 都存在唯一解 $\mathbf{x} = (\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in (\mathbb{R}^N)^{\mathbb{Z}_-}$，则 (1.1)-(1.2) 定义了一个映射 $H_W\colon (\mathbb{R}^d)^{\mathbb{Z}_-} \to \mathbb{R}^m$，其定义方式为 $H_W(\mathbf{z}) = \mathbf{y}_0$，其中 $\mathbf{y} \in (\mathbb{R}^m)^{\mathbb{Z}_-}$ 由 (1.2) 给出，而 $\mathbf{x}$ 是与 $\mathbf{z} \in (\mathbb{R}^d)^{\mathbb{Z}_-}$ 相关联的 (1.1) 的唯一解。给定一个待学习的（通常未知的）泛函 $H\colon (\mathbb{R}^d)^{\mathbb{Z}_-} \to \mathbb{R}^m$，随后训练读出层 W，使得H_W 是 H 的良好逼近。

对于回声状态网络 (1.1)–(1.2)，文献 [Gono 23] 给出了逼近界，适用于具有如下性质的映射 H：对于每个 $T \in \mathbb{N}$，其在有限长度输入序列 T 上的限制属于某个 Sobolev 空间，且 (1.1) 中的权重 A, C, ζ 取自具有特定结构的通用分布。本文考虑 (1.1)–(1.2) 的一种新型架构：不再在 (1.2) 中应用线性函数，而是应用随机生成的前馈神经网络，也就是说，(1.2) 被替换为

$$\mathbf{y}_t = \sum_{i=1}^{N} W_i \sigma(\mathbf{a}^{(i)} \cdot \mathbf{x}_{t-1} + \mathbf{c}^{(i)} \cdot \mathbf{z}_t + b_i) \tag{1.3}$$

对于 $t \in \mathbb{Z}_-$，并且系数 $\mathbf{a}^{(i)}$、$\mathbf{c}^{(i)}$ 分别在 $\mathbb{R}^N$、$\mathbb{R}^d$ 和 $\mathbb{R}$ 中取值，且由随机方式生成。大多数情况下，(1.1) 中的激活函数 σ 将只是恒等算子，或最终采用整流线性单元。

为了推导这种架构的学习误差界，我们将分几个步骤进行。首先，我们考察一个新的概念类 $\mathcal{C}$，即由满足特定积分表示的储备池泛函构成的循环（广义）Barron 泛函。事实证明，这个类 $\mathcal{C}$ 具有非常有用的性质：首先，我们证明一大类泛函 $H\colon (\mathbb{R}^d)^{\mathbb{Z}_-} \to \mathbb{R}^m$，其中包括但不限于具有“足够光滑性”的泛函，都属于 $\mathcal{C}$；并且在温和条件下，$\mathcal{C}$ 在 L^2意义下稠密。随后，我们考察储备池计算系统对 $\mathcal{C}$ 中元素的逼近，并推导逼近误差界。这表明，大量泛函都可以由循环神经网络逼近，其隐藏权重从一般分布中随机抽样；还可以为这些网络推导出显式的逼近率。

第二步是获得与这些系统的逼近结果所产生的参数限制相匹配的泛化误差界。这里的关键挑战在于，观测数据并非独立同分布，因此无法直接采用经典统计学习技术 [Bouc 13]，参见 [Gono 20b]。结合这些误差界，我们进而获得这类随机循环神经网络用于学习一般映射类 $\mathcal{C}$ 的学习误差界。作为副产品，我们还为储备池系统得到了新的普适性结果，参见 [Grig 18b, Grig 18a, Gono 20c, Gono 23, Gono 21b]，其中系数通常是随机生成的。值得强调的是，本文所述的构造能够实现一种完全可实施的、基于循环神经网络的学习算法，并且具有可证明的收敛保证，且不会遭遇维数灾难。

2 广义 Barron 函数的动力学类比

本节将介绍循环（广义）Barron 泛函类 $\mathcal{C}$，它构成了本文研究和逼近的概念类。我们证明了 $\mathcal{C}$ 的基本性质，并确定了其中的重要组成部分。具体而言，我们证明 $\mathcal{C}$ 是一个在平方可积泛函集合中稠密的向量空间，并且包含许多重要的映射类 $H\colon (\mathbb{R}^d)^{\mathbb{Z}_-} \to \mathbb{R}$，例如线性映射或“足够光滑”的映射。

2.1 记号

设 $D_d \subset \mathbb{R}^d$、$p \in [1, \infty]$，且 q 是 p 的 H¨older 共轭。回顾记号 $\ell^p = \ell^p(\mathbb{R})$，它表示满足 $\sum_{i\in\mathbb{N}} |x_i|^p < \infty$ 的序列 $(x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ 所构成的序列空间（当 $p < \infty$ 时），以及满足 $\sup_{i\in\mathbb{N}} |x_i| < \infty$ 的序列空间（当 $p = \infty$ 时）。符号 $\ell^p_- = \ell^p_-(\mathbb{R})$ 表示由序列 $(x_n)_{n\in\mathbb{Z}_-} \in \mathbb{R}^{\mathbb{Z}_-}$ 构成的空间，其中这些序列满足与 ℓ^p 中相同的可求和条件，且 $\mathbb{Z}_-$ 为包含零的负整数。我们用 $\ell^\infty(\ell^q)$ 表示满足 $\sup_{t\in\mathbb{Z}_-} \|\mathbf{x}_t\|_q < \infty$ 的序列 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in (\ell^q)^{\mathbb{Z}_-}$ 所构成的集合。对于 $\mathbf{x} \in \ell^p$ 和 $\mathbf{y} \in \ell^q$，记作 $\mathbf{x} \cdot \mathbf{y} = \sum_{i\in\mathbb{N}} x_i y_i$。设 σ_1、$\sigma_2\colon \mathbb{R} \to \mathbb{R}$ 为两个称作激活函数的函数。对于这两个激活函数，我们用相同符号表示它们对序列或向量的逐分量应用。此外，假设 σ_2 是 Lipschitz 连续的，常数为 L_{σ_2}，且 $\sigma_1(0) = 0$。给定 $N \in \mathbb{N}$，记 $\mathcal{X} = \mathbb{R} \times \ell^p \times \mathbb{R}^d \times \mathbb{R}$ 和 $\mathcal{X}^N = \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}^d \times \mathbb{R}$。我们将考虑取值于 $\mathcal{I}_d \subset (D_d)^{\mathbb{Z}_-}$ 的输入。

2.2 循环Barron泛函的定义与性质

本节定义递归广义 Barron 泛函类，并研究这些泛函的一些性质。

定义 2.1. 若存在一个（博雷尔）概率测度 μ 于 $\mathcal{X}$，$\mathbf{B} \in \ell^q$ 以及线性映射 $A\colon \ell^q \to \ell^q$、$C\colon \mathbb{R}^d \to \ell^q$，使得对于每个 $\mathbf{z} \in \mathcal{I}_d$，系统

$$\mathbf{x}_t = \sigma_1(A\mathbf{x}_{t-1} + C\mathbf{z}_t + \mathbf{B}), \quad t \in \mathbb{Z}_-, \tag{2.1}$$

存在唯一解 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in \ell^\infty_-(\ell^q)$，$\mu$ 满足

$$I_{\mu,p} := \int_{\mathcal{X}} |w|(\|\mathbf{a}\|_p + \|\mathbf{c}\| + |b|)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) < \infty$$

并记 $\mathbf{x}_t = \mathbf{x}_t(\mathbf{z})$，

$$H(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in \mathcal{I}_d. \tag{2.2}$$

我们将所有循环广义 Barron 泛函，或简称循环 Barron 泛函，构成的类记为 $\mathcal{C}$。

请注意，读出层 (2.2) 的构造方式是，输出由 $\mathbf{x}_{-1}$ 和 $\mathbf{z}_0$共同构成，而不是像储备池计算中的惯常做法那样仅由 $\mathbf{x}_0$, 构成（见 (1.2)）。之所以作此选择，是因为它使我们稍后能够在第 3.6 节中，以直接明了的方式将静态情形视为循环设置的一个特例。

唯一解性质（USP）。对于 (2.1) 这类状态方程，其解的存在性与唯一性是循环 Barron 泛函定义的一部分，也是一个经过深入研究的问题。文献中将这一性质称为回声状态性质（ESP）[Jaeg 10, Yild 12, Manj 13]，或称为唯一解性质（USP）[G Ma 20, Manj 22a]；相关问题已在确定性设置（例如，见上述参考文献及 [Grig 18a], ）和随机设置 [Manj 22b]中得到广泛研究。下述命题是一个常用充分条件的无限维推广；该条件保证 (2.1) 中的 USP 成立。此外，当激活函数 σ_1 为恒等算子时，(2.1) 是一个线性系统，此时该命题还给出了 USP 的完整刻画。两种情形下均假设输入有界；关于无界情形的可能推广，可参见 [Grig 19]。

Pr命题 2.2。考虑由 (2.1) 给出的状态方程以及以下两种情形 :

(i) 假设输入定义在 $\mathcal{I}_d \subset (D_d)^{\mathbb{Z}_-}$ 中，其中 D_d 是 $\mathbb{R}^d$ 的紧子集。此外，假设 σ_1 是 Lipschitz 连续的，常数为 L_{σ_1}，并且 $L_{\sigma_1}|||A||| < 1$ 满足 $|||A|||$ 为线性映射 $A\colon \ell^q \to \ell^q$ 的算子范数。那么，对于每个 $\mathbf{z} \in \mathcal{I}_d$，(2.1) 都具有 USP，即存在唯一的 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in \ell^\infty_-(\ell^q)$ 满足 (2.1)。

(ii) 假设 σ_1 是恒等算子，并且 $A\colon \ell^q \to \ell^q$、$C\colon \mathbb{R}^d \to \ell^q$ 是有界线性算子。如果算子 A 的谱半径 $\rho(A)$ 满足 $\rho(A) < 1$，那么线性状态方程 (2.1) 关于 $\ell^\infty_-(\mathbb{R}^d)$ 中的输入具有 USP。在这种情况下，(2.1) 的唯一解为

$$\mathbf{x}_t(\mathbf{z}) = \sum_{k=0}^{\infty} A^k(C\mathbf{z}_{t-k} + \mathbf{B}), \quad t \in \mathbb{Z}. \tag{2.3}$$

证明。第 (i) 部分的证明是对 [Grig 18a, 定理 3.1(ii)] 的直接推广，同时注意到假设 $L_{\sigma_1}|||A||| < 1$ 意味着状态方程在状态项上是一个压缩映射。第 (ii) 部分可通过仿照 [Grig 21b, 命题 4.2(i)] 的证明得到；该证明的一个关键要素是使用谱半径的盖尔范德公式，该公式对无限维巴拿赫空间之间的算子成立 [Lax 02, 定理 4, 第 195] 页。 □

循环 Barron 泛函的概念类 $\mathcal{C}$ 是一个向量空间。

命题 2.3。设 $H_1, H_2 \in \mathcal{C}$ 且 $\lambda_1, \lambda_2 \in \mathbb{R}$。那么 $\lambda_1 H_1 + \lambda_2 H_2 \in \mathcal{C}$。

证明。无妨假设 $\lambda_i \neq 0$ 对于 $i = 1, 2$。对于 $i = 1, 2$，令 $\mu^{(i)}$ 为 $\mathcal{X}$上的（博雷尔）概率测度，令 $\mathbf{B}^{(i)} \in \ell^q$，令 $A^{(i)}\colon \ell^q \to \ell^q$, $C^{(i)}\colon \mathbb{R}^d \to \ell^q$ 为线性映射，使得对于每个 $\mathbf{z} \in \mathcal{I}_d$，系统

$$\mathbf{x}_t^{(i)} = \sigma_1(A^{(i)}\mathbf{x}_{t-1}^{(i)} + C^{(i)}\mathbf{z}_t + \mathbf{B}^{(i)}), \quad t \in \mathbb{Z}_-, \tag{2.4}$$

存在唯一解 $(\mathbf{x}_t^{(i)})_{t\in\mathbb{Z}_-} \in \ell_-^\infty(\ell^q)$，$\mu^{(i)}$ 满足 $I_{\mu^{(i)},\,p} < \infty$，并且

$$H_i(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{(i)}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu^{(i)}(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in \mathcal{I}_d. \tag{2.5}$$

定义 $\pi_1\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$，令 $\pi_1((y_n)_{n\in\mathbb{N}}) = (y_{\frac{n+1}{2}}\mathbb{I}^{\mathbb{N}}(\frac{n+1}{2}))_{n\in\mathbb{N}}$，并定义 $\pi_2\colon \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$，令 $\pi_2((y_n)_{n\in\mathbb{N}}) = (y_{\frac{n}{2}}\mathbb{I}^{\mathbb{N}}(\frac{n}{2}))_{n\in\mathbb{N}}$。注意，$\pi_1, \pi_2$是线性的，二者都将 ℓ^p和 ℓ^q映射到自身，$\pi_1 + \pi_2 = \mathbb{I}$，且 $\pi_i \circ \sigma_1 = \sigma_1 \circ \pi_i$，$i = 1, 2$。定义 $\mathbf{B} \in \ell^q$以及线性映射 $A\colon \ell^q \to \ell^q$, $C\colon \mathbb{R}^d \to \ell^q$，令$A\mathbf{x} = \pi_1(A^{(1)}\mathbf{x}) + \pi_2(A^{(2)}\mathbf{x})$，$\mathbf{B} = \pi_1(\mathbf{B}^{(1)}) + \pi_2(\mathbf{B}^{(2)})$，且 $C\mathbf{z} = \pi_1(C^{(1)}\mathbf{z}) + \pi_2(C^{(2)}\mathbf{z})$。

现在假设，对于输入 $\mathbf{z} \in \mathcal{I}_d$ 以及我们刚刚定义的参数A、$\mathbf{B}$ C，$(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in \ell_-^\infty(\ell^q)$ 是 (2.1) 的一个解。对于每个 $t \in \mathbb{Z}_-$，记 $\mathbf{x}_t^{(1)} = (\mathbf{x}_{t,\,2n-1})_{n\in\mathbb{N}}$ 为 $\mathbf{x}_t$ 的奇数分量，记 $\mathbf{x}_t^{(2)} = (\mathbf{x}_{t,\,2n})_{n\in\mathbb{N}}$ 为偶数分量。于是，根据 A、$\mathbf{B}$ C 的构造，$\mathbf{x}_t^{(i)} \in \ell^q$ 是 $i = 1$ 下 (2.4) 的一个解。由 (2.4) 的解的唯一性可知，(2.1) 至多有一个解。另一方面，若现在记 $(\mathbf{x}_t^{(i)})_{t\in\mathbb{Z}_-} \in \ell_-^\infty(\ell^q) i = 1$、2 为输入 $\mathbf{z} \in \mathcal{I}_d$ 下 (2.4) 的唯一解，那么令 $\mathbf{x}_t = \pi_1(\mathbf{x}_t^{(1)}) + \pi_2(\mathbf{x}_t^{(2)})$ 定义了一个序列 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in \ell_-^\infty(\ell^q)$；同样，根据 A、$\mathbf{B}$ C 的构造，该序列是 (2.1) 的一个解。

接下来，定义 $\phi_i\colon \mathcal{X} \to \mathcal{X}$为 $\phi_i(w, \mathbf{a}, \mathbf{c}\ b) = (w(\lambda_1+\lambda_2)\ \pi_i(\mathbf{a}), \mathbf{c}\ b)$，其中 $i = 1, 2$，并令 $\mu = \frac{\lambda_1}{\lambda_1+\lambda_2}(\mu^{(1)}\circ\phi_1^{-1}) + \frac{\lambda_2}{\lambda_1+\lambda_2}(\mu^{(2)}\circ\phi_2^{-1})$，其中 $\mu^{(i)}\circ\phi_i^{-1}$ 表示 $\mu^{(i)}$ 在 ϕ_i下的推前测度。然后，积分变换定理表明

$$\int_{\mathcal{X}} |w|(\|\mathbf{a}\|_p + \|\mathbf{c}\| + |b|)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) = \sum_{i=1}^{2}\lambda_i \int_{\mathcal{X}} |w|(\|\pi_i(\mathbf{a})\|_p + \|\mathbf{c}\| + |b|)\mu^{(i)}(dw, d\mathbf{a}, d\mathbf{c}, db) < \infty$$

并且，根据 (2.5)，对所有 $\mathbf{z} \in \mathcal{I}_d$，

$$\begin{aligned}\int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1} + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) &= \sum_{i=1}^{2}\lambda_i \int_{\mathcal{X}} w\sigma_2(\pi_i(\mathbf{a})\cdot\mathbf{x}_{-1} + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu^{(i)}(dw, d\mathbf{a}, d\mathbf{c}, db)\\ &= \sum_{i=1}^{2}\lambda_i \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{(i)} + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu^{(i)}(dw, d\mathbf{a}, d\mathbf{c}, db)\\ &= \lambda_1 H_1(\mathbf{z}) + \lambda_2 H_2(\mathbf{z}).\end{aligned}$$

因此，$\lambda_1 H_1 + \lambda_2 H_2 \in \mathcal{C}$，如所述。 □

有限维循环 Barron 泛函。对于 $N \in \mathbb{N}$，用 $\mathcal{C}_N$ 表示泛函$H^N\colon \mathcal{I}_d \subset (D_d)^{\mathbb{Z}_-} \to \mathbb{R}$ 的集合，该集合满足以下性质：存在一个定义在$\mathcal{X}^N$上的（博雷尔）概率测度 μ^N，$A \in \mathbb{M}_{N,\,N}$, $\mathbf{B} \in \mathbb{R}^N$, $C \in \mathbb{M}_{N,\,d}$，使得对每个 $\mathbf{z} \in \mathcal{I}_d$，系统

$$\mathbf{x}_t^N = \sigma_1(A\mathbf{x}_{t-1}^N + C\mathbf{z}_t + \mathbf{B}), \quad t \in \mathbb{Z}_-, \tag{2.6}$$

有唯一解 $(\mathbf{x}_t^N)_{t\in\mathbb{Z}_-} \in \ell_-^\infty(\mathbb{R}^N)$，$\mu^N$ 满足

$$\int_{\mathcal{X}^N} |w|(\|\mathbf{a}\| + \|\mathbf{c}\| + |b|)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) < \infty$$

并且，记 $\mathbf{x}_t^N = \mathbf{x}_t^N(\mathbf{z})$，

$$H^N(\mathbf{z}) = \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^N(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z}\in\mathcal{I}_d. \tag{2.7}$$

我们将 $\mathcal{C}_N$ 中的元素称为有限维循环 Barron 泛函。在下一个命题中，我们将说明，有限维循环 Barron 泛函的集合可以自然地视为 $\mathcal{C}$ 的一个子空间。在命题中，我们使用系统态射的概念及其性质；例如， [Grig 21b] 中回顾了这些内容。

命题 2.4. 对于任意的 $N\in\mathbb{N}$，令 $\iota\colon \mathbb{R}^N \to \ell^q$ 为自然嵌入 $(x_1,\dots,x_N)^\top \longmapsto (x_1,\dots,x_N,0,\dots)$。给定一个元素 $H^N\in\mathcal{C}_N$，存在一个类型为 (2.1)-(2.2) 的系统，其实现为 H^N，且映射 ι 是一个系统态射。由此可得

$$\mathcal{C}_N \subset \mathcal{C}, \tag{2.8}$$

也就是说，每个有限维循环 Barron 泛函都存在一种作为循环 Barron 泛函的实现。

证明. 设 $H^N\in\mathcal{C}_N$ 使得 (2.7) 成立，其中 $\mathbf{x}^N\in\ell^\infty_-(\mathbb{R}^N)$ 满足 (2.6)。首先，定义 $\bar{A}\colon \ell^q\to\ell^q$，$\bar{\mathbf{B}}\in\ell^q$，$\bar{C}\colon \mathbb{R}^d\to\ell^q$，令 $(\bar{A}\mathbf{x})_i = \mathbb{1}_{\{i\le N\}}\sum_{j=1}^N A_{ij}x_j$，$\bar{B}_i = \mathbb{1}_{\{i\le N\}}B_i$，$(\bar{C}\mathbf{z})_i = \mathbb{1}_{\{i\le N\}}\sum_{j=1}^d C_{ij}z_j \in\ell^q$，$\mathbf{x}\in\mathbb{R}^d$，$i\in\mathbb{N}$，$\iota$ 为典范嵌入，我们用同一符号表示从 $\mathcal{X}^N$ 到 $\mathcal{X}$ 的嵌入，并以 $\mu=\mu^N\circ\iota^{-1}$ 表示 μ^N 在 ι 下的推前测度。

现在考虑系统 (2.6)–(2.7)，以及由 (2.1) 和 (2.2) 给出的系统，其参数为 $\bar{A},\bar{\mathbf{B}},\bar{C}$ 我们刚刚定义的，以及由测度 $\mu=\mu^N\circ\iota^{-1}$ 给出的读出层。我们将证明映射 $\iota\colon\mathbb{R}^N\to\ell^q$ 是这些系统之间的系统态射。为此，需要证明 ι 是系统等变且读出不变的（参见 [Grig 21b] 中的术语）。

首先，我们注意到 ι 是系统等变的，因为

$$\iota\big(\sigma_1(A\mathbf{x}^N + C\mathbf{z}+\mathbf{B})\big) = \sigma_1(\iota(A\mathbf{x}^N + C\mathbf{z}+\mathbf{B})) = \sigma_1(\bar{A}\iota(\mathbf{x}^N) + \bar{C}\mathbf{z}+\bar{\mathbf{B}}),\ \text{for any } \mathbf{x}^N\in\mathbb{R}^N,\ \mathbf{z}\in\mathbb{R}^d,$$

如所需。由此可知（参见 [Grig 21b，命题 2.2]），如果 $(\mathbf{x}_t^N)_{t\in\mathbb{Z}_-}\in\ell^\infty_-(\mathbb{R}^N)$ 是由 (2.6) 和 (2.7) 确定的系统在 $\mathbf{z}\in\mathcal{I}_d$ 下的一个解，那么 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-}\in\ell^\infty_-(\ell^q)$ 也是一个解，其中 $\mathbf{x}_t=\iota(\mathbf{x}_t^N)$ 对应于参数为 $\bar{A}$，$\bar{\mathbf{B}}$、$\bar{C}$、由 (2.1) 和 (2.2) 给出的系统。这个解是唯一的，因为如果 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-}\in\ell^\infty_-(\ell^q)$ 是该系统的另一个解，并且我们将其前 N 个分量记作 $\mathbf{x}_t^N\in\mathbb{R}^N$，那么由 (2.1) 可得，对于分量 $i=1,\dots,N$，有 $\overline{x}_t^N{}_{,i} = \sigma_1((\bar{A}\mathbf{x}_{t-1})_i + (\bar{C}\mathbf{z}_t)_i + \bar{B}_i) = \sigma_1((A\mathbf{x}_{t-1}^N)_i + (C\mathbf{z}_t)_i + B_i)$，这表明 $\mathbf{x}^N$ 是 (2.6) 的一个解，因此由唯一性可知，$\mathbf{x}^N=\mathbf{x}^N$ 必然成立。对于分量 $i>N$，由 (2.1) 可得 $\overline{x}_{t,i}=\sigma_1(0)=0$。综上，这证明了对于每个 $\mathbf{z}\in\mathcal{I}_d$，由参数 $A,\mathbf{B}\ C$ 下的 (2.1) 给出的系统都有唯一解 $\mathbf{x}$，且每一项 $\mathbf{x}_t$ 的前 N 个元素由 $\mathbf{x}_t^N$ 给出。

注意到 变量替换定理和 $\mathbb{R}^N$ 上的范数等价性表明

$$\int_{\mathcal{X}} |w|(\|\mathbf{a}\|_p + \|\mathbf{c}\| + |b|)\mu(dw,d\mathbf{a},d\mathbf{c},db) = \int_{\mathcal{X}^N} |w|(\|\iota(\mathbf{a})\|_p + \|\mathbf{c}\| + |b|)\mu^N(dw,d\mathbf{a},d\mathbf{c},db) < \infty,$$

这证明了 (2.2) 中的 H 定义了 $\mathcal{C}$中的一个元素。最后，我们利用每个解 $\mathbf{x}_t$ 的前 N 个分量由 $\mathbf{x}_t^N$ 给出这一点，并再次应用变量替换定理，证明映射 ι 是读出不变的：

$$\begin{aligned} H(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\mu(dw,d\mathbf{a},d\mathbf{c},db) &= \int_{\mathcal{X}^N} w\sigma_2(\iota(\mathbf{a})\cdot\mathbf{x}_{-1}(\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\mu^N(dw,d\mathbf{a},d\mathbf{c},db) \\ &= \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^N(\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\mu^N(dw,d\mathbf{a},d\mathbf{c},db) = H^N(\mathbf{z}) \end{aligned}$$

对于 $\mathbf{z}\in\mathcal{I}_d$，满足所需条件。特别地，$H^N$ 和 H 实际上是同一个泛函，这证明了包含关系 (2.8)。 □

2.3 递归 Barron 泛函示例

现在我们展示 大量输入/输出系统自然属于循环 Barron 类 s.
有限记忆泛函。下述命题表明，具有某些正则性的有限记忆因果及时不变系统，可以表示为具有线性状态方程的有限维循环 Barron 泛函。

命题 2.5。假设 $\sigma_1(x) = x$, $\sigma_2(x) = \max(x, 0)$, D_d 有界，令 $T \in \mathbb{N}$, 并假设 $h\colon \mathbb{R}^{Td} \to \mathbb{R}$ 属于 Sobolev 空间 $H^s(\mathbb{R}^{Td})$, 其中 $s > \frac{Td}{2} + 2$。那么，泛函 $H(\mathbf{z}) = h(\mathbf{z}_0, \ldots, \mathbf{z}_{-T+1}), \mathbf{z} \in \mathcal{I}_d$, 是一个有限维循环 Barron 泛函。

证明。取 $N = dT$, $\mathbf{B} = \mathbf{0}$,

$$A = \begin{pmatrix} \mathbf{0}_{d,d(T-1)} & \mathbf{0}_{d,d} \\ \mathbb{I}_{d(T-1)} & \mathbf{0}_{d(T-1),d} \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \mathbb{I}_d \\ \mathbf{0}_{d(T-1),d} \end{pmatrix}. \tag{2.9}$$

那么该系统

$$\mathbf{x}_t^N = \sigma_1(A\mathbf{x}_{t-1}^N + C\mathbf{z}_t + \mathbf{B}), \quad t \in \mathbb{Z}_-, \tag{2.10}$$

以 $\mathbf{x}_t^N = (\mathbf{z}_t^\top, \ldots, \mathbf{z}_{t-T+1}^\top)^\top$, $t \in \mathbb{Z}_-$ 为唯一解。接下来，根据 [E 21, 定理 3.1]（该定理基于 [Barr 93, 命题 1 和第 IX.15 节）]；或者，也可以利用 [Gono 23, 定理 1] 的证明中得到的表示推出该结果），存在一个定义在 $\mathbb{R} \times \mathbb{R}^{Td+1}$ 上的（博雷尔）概率测度 $\tilde{\mu}$，使得 $\int_{\mathbb{R}\times\mathbb{R}^{Td+1}} |w|\|\mathbf{v}\|\tilde{\mu}(dw, d\mathbf{v}) < \infty$，并且对勒贝格几乎处处的 $\mathbf{u} \in (D_d)^T$,

$$h(\mathbf{u}) = \int_{\mathbb{R}\times\mathbb{R}^{Td+1}} w\sigma_2(\mathbf{v} \cdot (\mathbf{u}^\top, 1)^\top)\tilde{\mu}(dw, d\mathbf{v}). \tag{2.11}$$

但是，根据 [Evan 10, 第 5 章定理 6(ii), h 关于 $(D_d)^T$ 连续6]，且 (2.11) 右侧关于 $\mathbf{u}$ 也连续，因此表示式 (2.11) 对所有 $\mathbf{u} \in (D_d)^T$ 均成立。

令 $\pi\colon \mathbb{R} \times \mathbb{R}^{Td+1} \to \mathcal{X}^N$ 定义如下：对于 $w \in \mathbb{R}$, 我们令 $\in \mathbb{R}^{Td+1}$ $\pi(w, \mathbf{v}) = (w, (v_{d+1}, \ldots, v_{Td}, 0, \ldots, 0), (v_1, \ldots, v_d)\ v_{Td+1})$。记 $\mu^N = \tilde{\mu} \circ \pi^{-1}$ 为 $\tilde{\mu}$ 在 π 下的推前测度。于是，根据变量替换公式，μ^N 满足

$$\begin{aligned}\int_{\mathcal{X}^N} |w|(\|\mathbf{a}\| + \|\mathbf{c}\| + |b|)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) &\le \sqrt{3}\int_{\mathcal{X}^N} |w|(\|\mathbf{a}\|^2 + \|\mathbf{c}\|^2 + |b|^2)^{1/2}\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &= \sqrt{3}\int_{\mathbb{R}\times\mathbb{R}^{Td+1}} |w|\|\mathbf{v}\|\tilde{\mu}(dw, d\mathbf{v}) < \infty\end{aligned}$$

并且对于所有 $\mathbf{z} \in (D_d)^{\mathbb{Z}_-}$, 有 $g_{\mathbf{z}}(w, \mathbf{a}, \mathbf{c}\ b) = w\sigma_2((\mathbf{c}^\top, \mathbf{a}^\top, b)(\mathbf{z}_0^\top, \mathbf{z}_{-1}^\top, \ldots, \mathbf{z}_{-T}^\top, \mathbf{1})^\top)$,

$$\begin{aligned}H(\mathbf{z}) = h(\mathbf{z}_0, \ldots, \mathbf{z}_{-T+1}) &= \int_{\mathbb{R}\times\mathbb{R}^{Td+1}} w\sigma_2(\mathbf{v} \cdot (\mathbf{z}_0^\top, \ldots, \mathbf{z}_{-T+1}^\top, 1)^\top)\tilde{\mu}(dw, d\mathbf{v}) \\ &= \int_{\mathbb{R}\times\mathbb{R}^{Td+1}} g_{\mathbf{z}}(\pi(w, \mathbf{v}))\tilde{\mu}(dw, d\mathbf{v}) = \int_{\mathcal{X}^N} w\sigma_2((\mathbf{c}^\top, \mathbf{a}^\top, b)(\mathbf{z}_0^\top, \mathbf{z}_{-1}^\top, \ldots, \mathbf{z}_{-T}^\top, 1)^\top)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &= \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a} \cdot \mathbf{x}_{-1}^N(\mathbf{z}) + \mathbf{c} \cdot \mathbf{z}_0 + b)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db),\end{aligned} \tag{2.12}$$

这表明 H 是一个有限维循环 Barron 泛函，即 $H \in \mathcal{C}_N$. □

线性泛函。以下命题表明，某些线性泛函也属于循环 Barron 类 $\mathcal{C}$。

命题 2.6. 假设 $\sigma_1(x) = x$ 且 $\sigma_2(x) = x$ 或 $\sigma_2(x) = \max(x,0)$。令 $\mathbf{w} \in \ell^1_-$, $\mathbf{a}_i \in \mathbb{R}^d$ 适用于 $i \in \mathbb{Z}_-$，并假设 D_d 有界，且 $\sup_{i\in\mathbb{Z}_-} \|\mathbf{a}_i\| < \infty$。假设以下条件之一成立

(i) $\sum_{i\in\mathbb{Z}_-} |w_i|\lambda^i < \infty$ *for some* $\lambda \in (0, 1)$

(ii) 或 $\mathcal{I}_d \subset \ell^q_-(D_d)$。

那么该泛函

$$H(\mathbf{z}) = \sum_{i\in\mathbb{Z}_-} w_i \mathbf{a}_i \cdot \mathbf{z}_i, \quad \mathbf{z} \in \mathcal{I}_d,$$

满足 $H \in \mathcal{C}$。

证明。设 $\lambda \in (0,1)$ 如 (i) 中所述；或者在 (ii) 成立的情况下，设 $\lambda = 1$。

首先考虑 $\sigma_2(x) = x$ 的情形。令 $\mathrm{B} = \mathbf{0}$，并定义 $A\colon \ell^q \to \ell^q$, $C\colon \mathbb{R}^d \to \ell^q$，使得对 $\mathrm{x} \in \ell^q$、$\mathrm{z} \in \mathbb{R}^d$, $i \in \mathbb{N}$，有 $(A\mathrm{x})_i = \mathbb{1}_{\{i>d\}} \lambda x_{i-d}$, $(C\mathrm{z})^i = \mathbb{1}_{\{i\le d\}} z_i$。现在来看 (2.1) 的第 i 个分量。代入定义可得

$$x_{t,i} = \mathbb{1}_{\{i>d\}} \lambda x_{t-1,i-d} + \mathbb{1}_{\{i\le d\}} z_{t,i}, \quad t \in \mathbb{Z}_-. \tag{2.13}$$

通过迭代 (2.13)，我们看到对于 $k \in \mathbb{N}$, $j = 1,\dots,d$，必有 $x_{t,(k-1)d+j} = \lambda^{k-1} z_{t-k+1,j}$。以此定义 $\mathbf{x}$，便可得出 (2.1) 有唯一解。此外，该解满足条件 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in \ell^\infty_-(\ell^q)$：假设 (i) 成立，则在 $q < \infty$ 的情形下，我们验证对任意 $t \in \mathbb{Z}_-$，都有 $\sum_{k\in\mathbb{N}} |x_{t,k}|^q = \sum_{i\in\mathbb{N}} \lambda^{qi} \|\mathbf{z}_{t-i}\|_q^q \le dM^q \lambda^q (1-\lambda^q)^{-1}$，其中 $M > 0$ 的选取满足 $\|\mathbf{u}\|_\infty \le M$，对所有 $\mathbf{u} \in D_d$ 均成立。在 $q = \infty$ 的情形下，我们验证 $\sup_{k\in\mathbb{N}} |x_{t,k}| \le \sup_{s,k\in\mathbb{N}} |z_{s,k}| \le M$。类似地，若 (ii) 成立，则由 $\mathbf{z} \in \mathcal{I}_d \subset \ell^q_-(D_d)$ 可得 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} \in \ell^\infty_-(\ell^q)$。

对于每个 $i \in \mathbb{Z}_- \setminus \{0\}$, $\bar{\mathbf{a}}_i \in \ell^p$，定义如下：令 $\bar{a}_{i,\ j+d(|i|-1)} = \lambda^{i+1} a_{i,\ j}$，其中 $j = 1,\dots,d$；并将 $\bar{\mathbf{a}}_i$ 的其余分量设为零。令 $\mu = \|\mathbf{w}\|_1^{-1} \sum_{i\in\mathbb{Z}_-} |w_i| \delta_{(\mathrm{sign}(w_i)\|\mathbf{w}\|_1,\ \overline{\mathbf{a}^i \mathbb{1}_{\{i<0\}}},\ \mathbf{a}^0 \mathbb{1}_{\{i=0\}},\ 0)}$，则 μ 确实是 $\mathcal{X}$ 上的一个（博雷尔）概率测度，并且 μ 满足

$$\begin{aligned} I_{\mu,p} = \int_{\mathcal{X}} |\bar{w}|(\|\mathbf{a}\|_p + \|\mathbf{c}\| + |b|)\mu(d\bar{w}, d\mathbf{a}, d\mathbf{c}, db) &= \sum_{i\in\mathbb{Z}_-} |w_i| (\|\bar{\mathbf{a}}_i \mathbb{1}_{\{i<0\}}\|_p + \|\mathbf{a}_0 \mathbb{1}_{\{i=0\}}\|) \\ &\le \sqrt{d} \sum_{i\in\mathbb{Z}_-} |w_i| \lambda^{i+1} \|\mathbf{a}_i\| \\ &\le \sqrt{d} \sup_{i\in\mathbb{Z}_-} \{\|\mathbf{a}_i\|\} \sum_{i\in\mathbb{Z}_-} |w_i| \lambda^{i+1} < \infty, \end{aligned}$$

因为 $\|\cdot\|_p \le \|\cdot\|$ 对 $p \ge 2$ 成立，且 $\|\cdot\|_p \le d^{\frac{1}{p}-\frac{1}{2}} \|\cdot\|$ 对 $p \in [1,2]$ 成立。此外，对于任意 $\mathbf{z} \in \mathcal{I}_d$,

$$\begin{aligned} \int_{\mathcal{X}} \bar{w}\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(d\bar{w}, d\mathbf{a}, d\mathbf{c}, db) &= \sum_{i\in\mathbb{Z}_-} w_i (\bar{\mathbf{a}}_i \mathbb{1}_{\{i<0\}} \cdot \mathbf{x}_{-1} + \mathbf{a}_0 \mathbb{1}_{\{i=0\}} \cdot \mathbf{z}_0) \\ &= \mathbf{a}_0 \cdot \mathbf{z}_0 w_0 + \sum_{i\in\mathbb{Z}_-\setminus\{0\}} w_i \lambda^{i+1} \sum_{j=1}^d a_{i,j} x_{-1,j+d(|i|-1)} \\ &= \mathbf{a}_0 \cdot \mathbf{z}_0 w_0 + \sum_{i\in\mathbb{Z}_-\setminus\{0\}} w_i \sum_{j=1}^d a_{i,j} z_{-|i|,j} \\ &= H(\mathbf{z}), \end{aligned} \tag{2.14}$$

这证明了在 $\sigma_2(x) = x$ 的情况下，$H \in \mathcal{C}$。

对于 $\sigma_2(x) = \max(x, 0)$ 的情形，令 $\varphi\colon \mathcal{X} \to \mathcal{X}$ 由 $\varphi(\bar{w}, \mathbf{a}, \mathbf{c}, b) = (-\bar{w}, -\mathbf{a}, -\mathbf{c}, -b)$ 定义，并考虑 $\bar{\mu} = \frac{1}{2}\mu + \frac{1}{2}(\mu \circ \varphi^{-1})$。于是，变量替换定理以及 $I_{\mu,p} < \infty$（如上所证）这一事实表明 $I_{\bar{\mu},p} < \infty$，再由 (2.14) 和 $\sigma_2(x) - \sigma_2(-x) = x$ 可得

$$\begin{aligned} &\int_{\mathcal{X}} \bar{w}\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\bar{\mu}(d\bar{w}, d\mathbf{a}, d\mathbf{c}, db) \\ &= \frac{1}{2}\int_{\mathcal{X}} \bar{w}\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(d\bar{w}, d\mathbf{a}, d\mathbf{c}, db) - \frac{1}{2}\int_{\mathcal{X}} \bar{w}\sigma_2(-\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) - \mathbf{c}\cdot\mathbf{z}_0 - b)\mu(d\bar{w}, d\mathbf{a}, d\mathbf{c}, db) \\ &= \frac{1}{2} H(\mathbf{z}), \end{aligned} \tag{2.15}$$

因此 $\frac{1}{2}H \in \mathcal{C}$，并且根据命题 2.3，也有 $H \in \mathcal{C}$。 □

下一个命题在略有不同的假设下也涵盖线性情形。

命题 2.7。假设 $\sigma_1(x) = x$ 且满足以下二者之一：$\sigma_2(x) = x$ 或 $\sigma_2(x) = \max(x, 0)$。令 $\mathbf{h} \in \ell^p_-(\mathbb{R}^d)$ 且 $\mathcal{I}_d \subset \ell^q_-(D_d)$。则该泛函

$$H(\mathbf{z}) = \sum_{i \in \mathbb{Z}_-} \mathbf{h}_i \cdot \mathbf{z}_i, \quad \mathbf{z} \in \mathcal{I}_d,$$

满足 $H \in \mathcal{C}$。

证明。情形 $\sigma_2(x) = \max(x, 0)$ 可由情形 $\sigma_2(x) = x$ 推出，方法与命题 2.6 的证明完全相同，因此只需考虑情形 $\sigma_2(x) = x$。按命题 2.6 的证明，取 $\lambda = 1$ 和 A，B C，即可得到 (2.1) 的唯一解 $(\mathbf{x}_t)_{t \in \mathbb{Z}_-} \in \ell^\infty_-(\ell^q)$；对于 $k \in \mathbb{N}$, $j = 1, \ldots, d$，我们有 $x_{t,\ (k-1)d+j} = z_{t-k+1,\ j}$。现在定义一个 $\in \ell^p$，令 $a_{(k-1)d+j} = h_{-k,\ j}$，其中 $k \in \mathbb{N}$, $j = 1, \ldots, d$，并考察 $\mu = \delta_{(1, \mathbf{a}, \mathbf{h}_0, 0)}$。于是 $I_{\mu,\ p} < \infty$，且对任意 $\mathbf{z}$ $\in \mathcal{I}_d$,

$$\int_{\mathcal{X}} \bar{w}\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\mu(d\bar{w}, d\mathbf{a}, d\mathbf{c}, db) = \mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z})+\mathbf{h}_0\cdot\mathbf{z}_0 = \mathbf{h}_0\cdot\mathbf{z}_0+\sum_{k\in\mathbb{N}}\sum_{j=1}^{d} h_{-k,j}z_{-k,j} = H(\mathbf{z}), \tag{2.16}$$

这证明了 $H \in \mathcal{C}$。 □

注记 2.8。该命题尤其涵盖了与卷积滤波器相关的泛函。事实上，令 $\mathbf{h}$ $\in \ell^p_-(\mathbb{R}^d)$，并考虑由 $U_{\mathbf{h}}(\mathbf{z})_t = \sum_{j \in \mathbb{Z}_-} \mathbf{h}_j \cdot \mathbf{z}_{t+j}$ 定义的卷积滤波器 $U_{\mathbf{h}}$：$\ell^q_-(D_d) \to \mathbb{R}^{\mathbb{Z}_-}$。那么，其关联泛函 H：$\ell^q_-(D_d) \to \mathbb{R}$ 由 $H(\mathbf{z}) = \sum_{j \in \mathbb{Z}_-} \mathbf{h}_j \cdot \mathbf{z}_j$ 定义，是 $\mathcal{C}$ 的一个元素，正如上文命题 2.7 所示。

作为进一步的例子，我们现在看到，可分希尔伯特空间上线性状态空间系统所关联的任意泛函，都可以实现为 $\mathcal{C}$ 中的一个元素，其中 $p = 2$。

命题 2.9。设 $\mathcal{Y}$ 是一个可分希尔伯特空间，令 $\bar{A}$：$\mathcal{Y} \to \mathcal{Y}$、$\bar{C}$：$\mathbb{R}^d \to \mathcal{Y}$ 为线性映射，且 $\bar{\mathrm{B}} \in \mathcal{Y}$，并假设对于每个 $\mathbf{z}$ $\in \mathcal{I}_d$，该系统

$$\bar{\mathbf{x}}_t = \bar{A}\bar{\mathbf{x}}_{t-1} + \bar{C}\mathbf{z}_t + \bar{\mathbf{B}}, \quad t \in \mathbb{Z}_-, \tag{2.17}$$

存在唯一解 $(\bar{\mathrm{x}}_t)_{t \in \mathbb{Z}_-} \in \ell^\infty_-(\mathcal{Y})$，即满足 $\sup_{t \in \mathbb{Z}_-} \|\bar{\mathrm{x}}_t\|_{\mathcal{Y}} < \infty$。令 $\bar{\mu}$ 为 $\mathbb{R} \times \mathcal{Y} \times \mathbb{R}^d \times \mathbb{R}$ 上的（博雷尔）概率测度，且有 $\int_{\mathbb{R}\times\mathcal{Y}\times\mathbb{R}^d\times\mathbb{R}} |w|(\|\mathrm{a}\|_{\mathcal{Y}} + \|\mathrm{c}\| + |b|)\bar{\mu}(dw, d\mathrm{a}\ d\mathrm{c}\ db) < \infty$，并考虑

$$H(\mathbf{z}) = \int_{\mathbb{R}\times\mathcal{Y}\times\mathbb{R}^d\times\mathbb{R}} w\sigma_2(\langle \mathbf{a}, \bar{\mathbf{x}}_{-1}(\mathbf{z})\rangle_{\mathcal{Y}} + \mathbf{c}\cdot\mathbf{z}_0 + b)\bar{\mu}(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in \mathcal{I}_d. \tag{2.18}$$

则 $H \in \mathcal{C}$ 满足 $p = 2$。

证明。记 T：$\mathcal{Y} \to \ell^2$ 为 $\mathcal{Y}$ 与 ℓ^2之间的等距同构。定义 $\mathbf{x}_t := T\bar{\mathrm{x}}_t$, $A = T\bar{A}T^{-1}$, $C = T\bar{C}$, $\mathrm{B} = T\bar{\mathrm{B}}$, ϕ：$\mathbb{R} \times \mathcal{Y} \times \mathbb{R}^d \times \mathbb{R} \to \mathcal{X}$, $\phi(w, \mathrm{a}, \mathrm{c}\ b) = (w, T\mathrm{a}, \mathrm{c}\ b)$，并令 $\mu = \bar{\mu} \circ \phi^{-1}$ 为 $\bar{\mu}$ 在 ϕ下的推前测度。

那么，μ 是 $\mathbb{R} \times \ell^2 \times \mathbb{R}^d \times \mathbb{R}$ 上的概率测度，且 A：$\ell^2 \to \ell^2$、C：$\mathbb{R}^d \to \ell^2$ 是线性映射。此外，根据 A的选择，我们有 B C，因此 $\mathbf{x}$ $\in \ell^\infty_-(\ell^2)$ 是方程 (2.1) 的一个解。另一方面，若 $(\mathbf{x}_t)_{t \in \mathbb{Z}_-} \in \ell^\infty_-(\ell^2)$ 是 (2.1) 的任意一个解，则 $\bar{\mathrm{x}}_t := T^{-1}\mathbf{x}_t$ 定义了 (2.17) 的一个解。但根据假设，后一个系统有唯一解，因此 (2.1) 的解也是唯一的。

最后，变量替换公式以及 T 是等距映射这一事实表明，μ 满足$I_{\mu,\ 2} = \int_{\mathbb{R}\times\mathcal{Y}\times\mathbb{R}^d\times\mathbb{R}} |w|(\|T\mathrm{a}\|_2 + \|\mathrm{c}\| + |b|)\bar{\mu}(dw, d\mathrm{a}\ d\mathrm{c}\ db) < \infty$ 且

$$\begin{aligned} H(\mathbf{z}) &= \int_{\mathbb{R}\times\mathcal{Y}\times\mathbb{R}^d\times\mathbb{R}} w\sigma_2(T\mathbf{a}\cdot T\bar{\mathbf{x}}_{-1}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\bar{\mu}(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &= \int_{\mathbb{R}\times\ell^2\times\mathbb{R}^d\times\mathbb{R}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db). \end{aligned} \tag{2.19}$$

这表明，泛函 H 可以表示为系统 (2.1) 的读出层 (2.2)，因此 $H \in \mathcal{C}$。这也证明，T 是由 A、C、B 和 $\bar{\mu}$ 确定的系统与 $\mathcal{C}$ 中由 A、C、B 和 μ确定的系统之间的系统同构。

□

下一个命题将 [Gono 20c 中的定理 III.13] 的一个普适性结果推广到了循环 Barron 泛函的情形。

命题 2.10. 假设 $\sigma_1(x) = x$，并且满足以下条件之一：$\sigma_2(x) = \max(x, 0)$，或 σ_2 有界、连续且非常数。设 $\bar{p} \in [1, \infty)$，并设 γ 是 $\mathcal{I}_d \subset \ell^\infty_-(D_d)$ 上的一个概率测度。那么，$\mathcal{C} \cap L^{\bar{p}}(\mathcal{I}_d, \gamma)$ 在 $L^{\bar{p}}(\mathcal{I}_d, \gamma)$ 中稠密。

证明. 设 $H \in L^{\bar{p}}(\mathcal{I}_d, \gamma)$，且 $\varepsilon > 0$。先考虑 $\sigma_2(x) = \max(x, 0)$ 的情形。定义激活函数 $\bar{\sigma}(x) = \sigma_2(x+1) - \sigma_2(x)$，则对 $x \in \mathbb{R}$，有 $|\bar{\sigma}(x)| \leq 1$。因此，根据 [Gono 20c 中的定理 III.13]，存在一个储备池泛函 H^{RC}，它与一个带有神经网络读出层的线性系统相关联，且满足 $H^{RC} \in L^{\bar{p}}(\mathcal{I}_d, \gamma)$ 和 $\|H - H^{RC}\|_{L^{\bar{p}}(\mathcal{I}_d, \gamma)} < \varepsilon$。更具体地说，存在 $N \in \mathbb{N}$、$A^N \in \mathbb{R}^{N \times N}$、$C^N \in \mathbb{R}^{N \times d}$，使得该系统

$$\mathbf{x}_t^N = A^N \mathbf{x}_{t-1}^N + C^N \mathbf{z}_t, \quad t \in \mathbb{Z}_- \tag{2.20}$$

具有回声状态性质，且对于某个神经网络 $h\colon \mathbb{R}^N \to \mathbb{R}$，有 $H^{RC}(\mathbf{z}) = h(\mathbf{x}_0^N(\mathbf{z}))$，其定义为

$$h(\mathbf{x}) = \sum_{j=1}^{k} \beta_j \bar{\sigma}(\boldsymbol{\alpha}_j \cdot \mathbf{x} - \theta_j)$$

对于某些 $k \in \mathbb{N}$、β_j、$\theta_j \in \mathbb{R}$ 和 $\boldsymbol{\alpha}_j \in \mathbb{R}^N$。我们声称 $H^{RC} \in \mathcal{C}_N$。为证明这一点，注意到取 $A = A^N$、$\mathbf{B} = \mathbf{0}$ 和 $C = C^N$ 时，由于 (2.20) 的回声状态性质，过程 $\mathbf{x}^N$ 是系统 (2.6) 的唯一解。根据 [Gono 20c 的证明、定理 III.13]，以及假设 $\mathcal{I}_d \subset \ell^\infty_-(D_d)$，我们得到唯一解满足 $\mathbf{x}^N \in \ell^\infty_-(\mathbb{R}^N)$。现在令$\mu^N = \frac{1}{2k}\sum_{j=1}^{k} \delta_{(2k\beta_j, A^\top \boldsymbol{\alpha}_j, C^\top \boldsymbol{\alpha}_j, -\theta_j+1)} + \delta_{(-2k\beta_j, A^\top \boldsymbol{\alpha}_j, C^\top \boldsymbol{\alpha}_j, -\theta_j)}$。那么 μ^N 是 $\mathcal{X}^N$ 上的概率测度，并且对于 $\mathbf{z} \in \mathcal{I}_d$

$$\begin{aligned}
&\int_{\mathcal{X}^N} w\sigma_2(\mathbf{a} \cdot \mathbf{x}_{-1}^N(\mathbf{z}) + \mathbf{c} \cdot \mathbf{z}_0 + b)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&= \frac{1}{2k}\sum_{j=1}^{k} 2k\beta_j \sigma_2(\boldsymbol{\alpha}_j \cdot (A\mathbf{x}_{-1}^N(\mathbf{z}) + C\mathbf{z}_0) - \theta_j + 1) - 2k\beta_j \sigma_2(\boldsymbol{\alpha}_j \cdot (A\mathbf{x}_{-1}^N(\mathbf{z}) + C\mathbf{z}_0) - \theta_j) \\
&= \sum_{j=1}^{k} \beta_j \bar{\sigma}(\boldsymbol{\alpha}_j \cdot \mathbf{x}_0^N(\mathbf{z}) - \theta_j) = h(\mathbf{x}_0^N(\mathbf{z})) = H^{RC}(\mathbf{z}).
\end{aligned}$$

这表明 $H^{RC} \in \mathcal{C}_N$ 和 $\|H - H^{RC}\|_{L^{\bar{p}}(\mathcal{I}_d, \gamma)} < \varepsilon$。结合命题 2.4 可得所需结论。

当 σ 有界、连续且非常数时，我们可以直接使用 σ（而不是 $\bar{\sigma}$），并采用类似的方法继续推导。

□

广义卷积泛函。最后，我们证明类 $\mathcal{C}$ 包含可推广卷积滤波器的重要变换族的泛函。

命题 2.11。假设 $\sigma_1(x) = x$, $p \in (1, \infty)$，并且 $\mathcal{I}_d \subset \ell^q_-(D_d)$。设 μ 是定义在 $\mathcal{X}$ 上的概率测度，满足 $I_{\mu,p} < \infty$。则泛函 $H\colon \mathcal{I}_d \to \mathbb{R}$,

$$H(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a} \cdot \mathbf{z}_{-1:-\infty} + \mathbf{c} \cdot \mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in \mathcal{I}_d \tag{2.21}$$

在 $\mathcal{C}$中。

证明。我们只需构造 $\mathbf{B} \in \ell^q$ 以及线性映射 $A\colon \ell^q \to \ell^q$, $C\colon \mathbb{R}^d \to \ell^q$，使得对于每个 $\mathbf{z} \in \mathcal{I}_d$，系统

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + C\mathbf{z}_t + \mathbf{B}, \quad t \in \mathbb{Z}_-, \tag{2.22}$$

存在唯一解 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-} = (\mathbf{z}_{t:-\infty})_{t\in\mathbb{Z}_-} \in \ell^\infty_-(\ell^q)$。于是 $\mathbf{x}_{-1} = \mathbf{z}_{-1:-\infty}$，因此 H 具有表示式 (2.2)。取 $\mathbf{B} = \mathbf{0}$，令 $C\mathbf{y}$ 为一个序列，其前 d个分量与 $\mathbf{y}$ 相同，其余分量为 0；令 A 为一个算子，将输入的第 i 个分量移至第 $i+d$ 个分量，并在前 d 个分量处插入 0。于是 $A\colon \ell^q \to \ell^q$、$C\colon \mathbb{R}^d \to \ell^q$ 均为线性算子；事实上，对所有 $t \in \mathbb{Z}_-$，$\mathbf{x}_t = \mathbf{z}_{t:-\infty}$是方程 (2.22) 的唯一解，且 $\mathbf{x} \in \ell^\infty_-(\ell^q)$。 □

注意，当 μ 是狄拉克测度且 σ_2 是恒等算子时，标准卷积滤波器是 (2.21) 中泛函的一种特殊情形。

3 逼近

本节将为循环 Barron 泛函的概念类 $\mathcal{C}$ 建立逼近结果。作为逼近系统，我们采用对有限维回声状态网络 (1.1)–(1.2) 的一种改造：将 (1.2) 中的线性读出层 W 替换为极限学习机 (ELM) / 随机特征神经网络，即一种具有 (1.3) 形式的神经网络，其参数均为随机生成，只有由权重 $(W_i)_{i=1,\ \ldots\ldots,\ N}$ 给出的输出层是可训练的。为推导此类界，我们分几个逼近步骤进行：(a) 用 (1.1) 形式的有限系统逼近无限系统 (2.1)；(b) 对 (2.2) 中的积分进行蒙特卡洛逼近；(c) 使用重要性抽样程序，以确保神经网络权重可从一般分布生成，而非从（未知的）测度 μ 生成。

3.1 通过有限维系统进行逼近

我们首先证明，对于任意 $N \in \mathbb{N}$，$\mathcal{C}$ 中的元素都可以通过有限维截断来逼近，也就是说，可以通过 $\mathcal{C}_N$ 中的有限维循环 Barron 泛函来逼近。

命题 3.1. 假设 $q \in [1,\infty)$ 和 $\mathcal{I}_d \subset \ell^\infty_-(\mathbb{R}^d)$。设 $H \in \mathcal{C}$ 具有表示式 (2.2)，其中概率测度 μ 定义在 $\mathcal{X}$ 上，$\mathbf{B} \in \ell^q$，且 $A\colon \ell^q \to \ell^q$、$C\colon \mathbb{R}^d \to \ell^q$ 是有界线性映射。令 $N \in \mathbb{N}$，对于 $j \in \mathbb{N}$, $k = 1,\ldots,d$，令 $\boldsymbol{\epsilon}^j = (\delta_{ij})_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$ 和 $\mathbf{e}^k = (\delta_{ik})_{i=1,\ldots,d} \in \mathbb{R}^d$。令 $\bar{\mathbf{B}} \in \mathbb{R}^N$, $\bar{A} \in \mathbb{R}^{N\times N}$, $\bar{C} \in \mathbb{R}^{N\times d}$ 如下给出：$\bar{B}_i = B_i$, $\bar{A}_{ij} = (A\boldsymbol{\epsilon}^j)_i$, $\bar{C}_{ik} = (C\mathbf{e}^k)_i$，其中 $i,j = 1,\ldots,N$, $k = 1,\ldots,d$。假设 σ_1 是 L_{σ_1}-Lipschitz 且 $L_{\sigma_1}|||A||| < 1$。则以下陈述成立：

(i) 对每个 $\mathbf{z}, \in \mathcal{I}_d$ 该系统

$$\mathbf{x}^N_t = \sigma_1(\bar{A}\mathbf{x}^N_{t-1} + \bar{C}\mathbf{z}_t + \bar{\mathbf{B}}), \quad t \in \mathbb{Z}_-, \tag{3.1}$$

存在唯一解 $(\mathbf{x}^N_t)_{t\in\mathbb{Z}_-} \in \ell^\infty_-(\mathbb{R}^N)$。

(ii) 记 $\mathbf{x}^N_t(\mathbf{z}) = \mathbf{x}^N_t$，并定义泛函 $H^N\colon \mathcal{I}_d \to \mathbb{R}$ 如下：

$$H^N(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\iota(\mathbf{x}^N_{-1}(\mathbf{z})) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in \mathcal{I}_d. \tag{3.2}$$

那么存在一个与 N 无关的常数 $C_{\mathrm{fin}} > 0$，使得对所有 $\mathbf{z} \in \mathcal{I}_d$,

$$|H(\mathbf{z}) - H^N(\mathbf{z})| \le C_{\mathrm{fin}} \sum_{k=0}^{\infty} (L_{\sigma_1}|||A|||)^k \left(\sum_{i=N+1}^{\infty} |x_{-1-k,i}|^q \right)^{1/q}. \tag{3.3}$$

特别地，$\lim_{N\to\infty} H^N(\mathbf{z}) = H(\mathbf{z})$。该常数为 $C_{\mathrm{fin}} = L_{\sigma_2}\int_{\mathcal{X}} |w|\|\mathbf{a}\|_p \mu(dw, d\mathbf{a}\ d\mathbf{c}\ db)$。

(iii) $H^N \in \mathcal{C}_N$。

证明。部分 (i) 是命题 2.2 以及不等式 $|||\bar{A}||| \leq |||A|||$ 的推论，而该不等式又意味着 $L_{\sigma_1}|||\bar{A}||| < 1$。关于 (ii)，首先注意到，根据我们对 $\bar{C}$ 的选择，可得对于所有 $\mathbf{u} \in \mathbb{R}^d$, $i = 1, \dots, N$，$(C\mathbf{u})_i = \sum_{k=1}^d u_k (C\mathbf{e}^k)_i = \sum_{k=1}^d u_k \bar{C}_{ik} = (\bar{C}\mathbf{u})_i$。同样，对于 $\mathbf{y} \in \mathbb{R}^N$, $i = 1, \dots, N$，有 $(\bar{A}\mathbf{y})_i = \sum_{j=1}^N y_j \bar{A}_{ij} = \sum_{j=1}^N y_j (A\boldsymbol{\epsilon}^j)_i = (A\iota(\mathbf{y}))_i$。因此，对于 $i = 1, \dots, N$:

$$\begin{aligned} |x_{t,i} - x_{t,i}^N| &\leq L_{\sigma_1} |(A\mathbf{x}_{t-1})_i + (C\mathbf{z}_t)_i + B_i - (\bar{A}\mathbf{x}_{t-1}^N)_i - (\bar{C}\mathbf{z}_t)_i - \bar{B}_i| \\ &= L_{\sigma_1} |(A(\mathbf{x}_{t-1} - \iota(\mathbf{x}_{t-1}^N)))_i|. \end{aligned} \tag{3.4}$$

记 $\pi\colon\ \ell^q \to \mathbb{R}^N$ 投影到前 N 个坐标上。则

$$\begin{aligned} \|\mathbf{x}_t(\mathbf{z}) - \iota(\mathbf{x}_t^N(\mathbf{z}))\|_q &\leq \|\mathbf{x}_t(\mathbf{z}) - \iota(\pi(\mathbf{x}_t(\mathbf{z})))\|_q + \|\iota(\pi(\mathbf{x}_t(\mathbf{z}))) - \iota(\mathbf{x}_t^N(\mathbf{z}))\|_q \\ &= \left(\sum_{i=N+1}^{\infty} |x_{t,i}|^q \right)^{1/q} + \left(\sum_{i=1}^{N} |x_{t,i} - x_{t,i}^N|^q \right)^{1/q} \\ &\leq \left(\sum_{i=N+1}^{\infty} |x_{t,i}|^q \right)^{1/q} + L_{\sigma_1} \left(\sum_{i=1}^{N} |(A(\mathbf{x}_{t-1} - \iota(\mathbf{x}_{t-1}^N)))_i|^q \right)^{1/q} \\ &\leq \left(\sum_{i=N+1}^{\infty} |x_{t,i}|^q \right)^{1/q} + L_{\sigma_1} \|A(\mathbf{x}_{t-1} - \iota(\mathbf{x}_{t-1}^N))\|_q \\ &\leq \left(\sum_{i=N+1}^{\infty} |x_{t,i}|^q \right)^{1/q} + L_{\sigma_1} |||A||| \|\mathbf{x}_{t-1} - \iota(\mathbf{x}_{t-1}^N)\|_q. \end{aligned} \tag{3.5}$$

因此，迭代（3.5）可得

$$\|\mathbf{x}_t(\mathbf{z}) - \iota(\mathbf{x}_t^N(\mathbf{z}))\|_q \leq \sum_{k=0}^{\infty} (L_{\sigma_1} |||A|||)^k \left(\sum_{i=N+1}^{\infty} |x_{t-k,i}|^q \right)^{1/q}. \tag{3.6}$$

现在可以利用这一点估计 H 与其截断之间的差异:

$$\begin{aligned} |H(\mathbf{z}) - H^N(\mathbf{z})| &\leq \int_{\mathcal{X}} |w| |\sigma_2(\mathbf{a} \cdot \mathbf{x}_{-1}(\mathbf{z}) + \mathbf{c} \cdot \mathbf{z}_0 + b) - \sigma_2(\mathbf{a} \cdot \iota(\mathbf{x}_{-1}^N(\mathbf{z})) + \mathbf{c} \cdot \mathbf{z}_0 + b)| \mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &\leq L_{\sigma_2} \int_{\mathcal{X}} |w| |\mathbf{a} \cdot \mathbf{x}_{-1}(\mathbf{z}) - \mathbf{a} \cdot \iota(\mathbf{x}_{-1}^N(\mathbf{z}))| \mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &\leq L_{\sigma_2} \int_{\mathcal{X}} |w| \|\mathbf{a}\|_p \mu(dw, d\mathbf{a}, d\mathbf{c}, db) \|\mathbf{x}_{-1}(\mathbf{z}) - \iota(\mathbf{x}_{-1}^N(\mathbf{z}))\|_q \\ &\leq C_{\text{fin}} \sum_{k=0}^{\infty} (L_{\sigma_1} |||A|||)^k \left(\sum_{i=N+1}^{\infty} |x_{-1-k,i}|^q \right)^{1/q} \end{aligned} \tag{3.7}$$

其中 $C_{\text{fin}} = L_{\sigma_2} \int_{\mathcal{X}} |w| \|\mathbf{a}\|_p \mu(dw, d\mathbf{a}\ d\mathbf{c}\ db)$。

尚需证明 $H^N \in \mathcal{C}_N$。回顾 $\pi\colon\ \ell^p \to \mathbb{R}^N$表示向前 N个分量的投影，而令 $\phi\colon\ \mathcal{X} \to \mathcal{X}^N$为乘积映射 $\phi = \mathrm{Id}_{\mathbb{R}} \times \pi \times \mathrm{Id}_{\mathbb{R}^d} \times \mathrm{Id}_{\mathbb{R}}$，即$\phi(w, \mathbf{a}, \mathbf{c}\ b) = (w, (a_i)_{i=1,\dots,N}, \mathbf{c}\ b)$。记 μ在ϕ下的推前测度为 $\mu^N = \mu \circ \phi^{-1}$。则 μ^N 是 $\mathcal{X}^N$ 上的一个（博雷尔）概率测度，并且在 (i) 中我们证明了，对于每个 $\mathbf{z} \in \mathcal{I}_d$，系统 (3.1) 都有唯一解。此外，积分变换定理蕴含μ^N 满足

$$\int_{\mathcal{X}^N} |w| (\|\mathbf{a}\| + \|\mathbf{c}\| + |b|) \mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) = \int_{\mathcal{X}} |w| (\|\pi(\mathbf{a})\| + \|\mathbf{c}\| + |b|) \mu(dw, d\mathbf{a}, d\mathbf{c}, db) < \infty$$

并记 $\mathbf{x}_t^N = \mathbf{x}_t^N(\mathbf{z})$，

$$\begin{aligned} H^N(\mathbf{z}) &= \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\iota(\mathbf{x}_{-1}^N(\mathbf{z})) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &= \int_{\mathcal{X}} w\sigma_2(\pi(\mathbf{a})\cdot\mathbf{x}_{-1}^N(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &= \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^N(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z}\in\mathcal{I}_d. \end{aligned} \tag{3.8}$$

这证明了 $H^N \in \mathcal{C}_N$。 □

3.2 归一化实现

接下来的结果表明，概念类 $\mathcal{C}$ 中的一大类元素可以转化为我们称之为归一化实现的形式，其中方程 (2.1) 的解映射是一个乘法缩放算子。

命题 3.2. 假设 $\sigma_1(x) = x$, $p\in(1,\infty)$，且 D_d 有界。设 $H\in\mathcal{C}$ 有一种类型为 (2.1)-(2.2) 的实现，其中 C 有界，且有界的 A: $\ell^q\to\ell^q$ 满足 $\rho(A) < 1$，这保证存在一个正整数 $k_0\in\mathbb{N}$，使得对所有 $k\geq k_0$ 都有 $|||A^k||| < 1$。固定 $\lambda\in\left(|||A^{k_0}|||^{1/k_0}, 1\right)$。那么，存在一个定义在 $\mathcal{X}$ 上的（博雷尔）概率测度 $\bar{\mu}$，满足 $\int_{\mathcal{X}}|w|(\|\mathbf{a}\|_p+\|\mathbf{c}\|+|b|)\bar{\mu}(dw, d\mathbf{a}\ d\mathbf{c}\ db) < \infty$，使得 H 具有另一种表示

$$H(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot(\bar{\Lambda}\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\bar{\mu}(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z}\in\mathcal{I}_d, \tag{3.9}$$

其中 $\bar{\Lambda}$: $(D_d)^{\mathbb{Z}_-}\to\ell^q$，$(\bar{\Lambda}\mathbf{z})_{d(k-1)+j} = \lambda^{k-1}z_{-k,\ j}$ 对于 $k\in\mathbb{N}$、$j = 1$、……、d。

注记 3.3。前一陈述中关于 D_d 有界的假设可以放宽为要求 $\sum_{k\in\mathbb{N}}\|\mathbf{z}_{-k}\|\lambda^k < \infty$。

证明。首先，我们回顾一下，根据命题 2.2，陈述中的假设保证系统 (2.1) 对每个 $\mathbf{z}\in\mathcal{I}_d$ 都有唯一解 $(\mathbf{x}_t(\mathbf{z}))_{t\in\mathbb{Z}_-}\in\ell^\infty_-(\ell^q)$，并且 $\mathbf{x}_t(\mathbf{z})$ 由表达式 (2.3) 明确给出。根据盖尔范德公式，对于谱半径 $\rho(A)$ 的表达式，存在一个正整数 $k_0\in\mathbb{N}$，使得对所有$k\geq k_0$ 均有 $|||A^k||| < 1$。

接下来，我们定义 $\tilde{A} = \lambda^{-1}A$，并注意到选择 $\lambda\in\left(|||A^{k_0}|||^{1/k_0}, 1\right)$ 保证了 $\left|\left|\left|\tilde{A}^{k_0}\right|\right|\right| < 1$。此外，对于每个 $k\in\mathbb{N}_0$，映射 $\tilde{A}^kC$: $\mathbb{R}^d\to\ell^q$ 是两个巴拿赫空间之间的有界线性映射，因此其伴随算子是从 $(\ell^q)^*\to(\mathbb{R}^d)^*$ 到 $\ell^p\to\mathbb{R}^d$的有界线性映射。我们可以将其识别为 $(\tilde{A}^kC)^*$。更具体地说，对于 $y = [y\circ\tilde{A}^kC\,] = y\,\tilde{A}^kC$ $y = \mathbf{y}\cdot\ \mathbf{y}\in\ell^p$ $\in\ell^q(\ell^q)^*$，$\mathbf{u}\in\mathbb{R}^d$，

伴随算子定义为 $[(\tilde{A}^kC)^*(\ )](\mathbf{u})(\ )(\mathbf{u})(\mathbf{u})$，因此，根据识别 $(\mathbf{b})\,\mathbf{a}\,\mathbf{b}$（对某个 $\mathbf{a}$ 以及所有 $\mathbf{b}$），我们得到伴随性质可表述为 $(\tilde{A}^kC)^*(\mathbf{a}\mathbf{y})\cdot\mathbf{u} = \mathbf{a}\mathbf{y}\cdot(\tilde{A}^kC)\mathbf{u}$。

因此，我们可以考虑由 $[\mathcal{L}(\mathbf{a})]_k = (\tilde{A}^kC)^*\mathbf{a}$ 定义的映射 $\mathcal{L}$: $\ell^p\to((\mathbb{R}^d)^{\mathbb{N}_0})$，其中 $k\in\mathbb{N}_0$，$\mathbf{a}\in\ell^p$。我们声称，像 $\mathcal{L}(\ell^p)$ 可以视为 ℓ^p 的一个子集。事实上，

$$\begin{aligned} \sum_{k\in\mathbb{N}_0}\|[\mathcal{L}(\mathbf{a})]_k\|^p &= \sum_{k\in\mathbb{N}_0}\|(\tilde{A}^kC)^*\mathbf{a}\|^p \leq \sum_{k\in\mathbb{N}_0}\left|\left|\left|(\tilde{A}^kC)^*\right|\right|\right|^p\|\mathbf{a}\|_p^p = \sum_{k\in\mathbb{N}_0}\left|\left|\left|\tilde{A}^kC\right|\right|\right|^p\|\mathbf{a}\|_p^p \leq (|||C|||\|\mathbf{a}\|_p)^p\sum_{k\in\mathbb{N}_0}\left|\left|\left|\tilde{A}^k\right|\right|\right|^p \\ &\leq (|||C|||\|\mathbf{a}\|_p)^p\sum_{j\in\mathbb{N}_0}\sum_{i=0}^{k_0-1}\left|\left|\left|\tilde{A}^{k_0}\right|\right|\right|^{jp}\left|\left|\left|\tilde{A}^i\right|\right|\right|^p = (|||C|||\|\mathbf{a}\|_p)^p\sum_{i=0}^{k_0-1}\frac{\left|\left|\left|\tilde{A}^i\right|\right|\right|^p}{1-\left|\left|\left|\tilde{A}^{k_0}\right|\right|\right|^p} < \infty, \end{aligned} \tag{3.10}$$

由于 $\left|\!\left|\!\left|\tilde{A}^{k_0}\right|\!\right|\!\right| = \lambda^{-k_0}\left|\!\left|\!\left|A^{k_0}\right|\!\right|\!\right| < 1$。因此，由 $[\bar{\mathcal{L}}(\mathbf{a})]_{d(k-1)+j} = [(\tilde{A}^{k-1}C)^*\mathbf{a}]_j$ 定义的映射 $\bar{\mathcal{L}}\colon \ell^p \to \ell^p$ 对于 $k \in \mathbb{N}$, $j = 1, \dots, d$ 是良定义的，并且由 (3.10) 可得

$$\|\bar{\mathcal{L}}(\mathbf{a})\|_p = \left(\sum_{k\in\mathbb{N}}\sum_{j=1}^{d}[(\tilde{A}^{k-1}C)^*\mathbf{a}]_j^p\right)^{1/p} = \left(\sum_{k\in\mathbb{N}_0}\|[\mathcal{L}(\mathbf{a})]_k\|_p^p\right)^{1/p} \le \left(d\sum_{k\in\mathbb{N}_0}\|[\mathcal{L}(\mathbf{a})]_k\|^p\right)^{1/p} \le c_0\|\mathbf{a}\|_p, \tag{3.11}$$

其中 $c_0 = d^{1/p}|\!|\!|C|\!|\!|\left(\sum_{i=0}^{k_0-1}\frac{|\!|\!|\tilde{A}^i|\!|\!|^p}{1-|\!|\!|\tilde{A}^{k_0}|\!|\!|^p}\right)^{1/p}$，并且我们利用了 $\|\cdot\|_p \le \|\cdot\|$（对于 $p \ge 2$）以及 $\|\cdot\|_p \le d^{\frac{1}{p}-\frac{1}{2}}\|\cdot\|$（对于 $p \in [1,2]$）。现在可以使用 (2.3) 将 $\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z})$ 改写为

$$\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z}) = \mathbf{a}\cdot\bar{\mathbf{B}}_0 + \sum_{i\in\mathbb{N}}a_i\sum_{k=0}^{\infty}(A^kC\mathbf{z}_{-1-k})_i = \mathbf{a}\cdot\bar{\mathbf{B}}_0 + \sum_{k=0}^{\infty}\mathbf{a}\cdot(\tilde{A}^kC\mathbf{z}_{-1-k}\lambda^k), \tag{3.12}$$

其中 $\bar{\mathbf{B}}_0 = \sum_{k=0}^{\infty}A^k\mathbf{B} \in \ell^q$，并且由 Fubini 定理可知级数可以交换，具体可通过估计下式看出

$$\begin{aligned}\sum_{k=0}^{\infty}\sum_{i\in\mathbb{N}}|a_i||(A^kC\mathbf{z}_{-1-k})_i| &\le \sum_{k=0}^{\infty}\|\mathbf{a}\|_p\|A^kC\mathbf{z}_{-1-k}\|_q \le \sum_{k=0}^{\infty}\|\mathbf{a}\|_p\left|\!\left|\!\left|A^k\right|\!\right|\!\right||C|||\mathbf{z}_{-1-k}\|_q \\ &\le \|\mathbf{a}\|_p|C|\!|\!|\sum_{j\in\mathbb{N}_0}\sum_{i=0}^{k_0-1}\left|\!\left|\!\left|A^{k_0}\right|\!\right|\!\right|^j\left|\!\left|\!\left|A^i\right|\!\right|\!\right|\|\mathbf{z}_{-1-k}\|_q < \infty,\end{aligned}$$

因为 $\left|\!\left|\!\left|A^{k_0}\right|\!\right|\!\right| < 1$，且 D_d 有界。(3.12) 中的第二项可以改写为

$$\begin{aligned}\sum_{k=0}^{\infty}\mathbf{a}\cdot(\tilde{A}^kC\mathbf{z}_{-1-k}\lambda^k) &= \sum_{k=0}^{\infty}((\tilde{A}^kC)^*\mathbf{a})\cdot(\mathbf{z}_{-1-k}\lambda^k) = \sum_{k=0}^{\infty}\sum_{j=1}^{d}[(\tilde{A}^kC)^*\mathbf{a}]_j[z_{-1-k}]_j\lambda^k \\ &= \sum_{k=1}^{\infty}\sum_{j=1}^{d}[(\tilde{A}^{k-1}C)^*\mathbf{a}]_j[z_{-k}]_j\lambda^{k-1} = \sum_{k=1}^{\infty}\sum_{j=1}^{d}[\bar{\mathcal{L}}(\mathbf{a})]_{d(k-1)+j}(\bar{\Lambda}\mathbf{z})_{d(k-1)+j} \\ &= \bar{\mathcal{L}}(\mathbf{a})\cdot(\bar{\Lambda}\mathbf{z}).\end{aligned} \tag{3.13}$$

将 (3.13) 与 (3.12) 结合并代入 (2.2)，即可得到

$$H(\mathbf{z}) = \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\bar{\mathbf{B}}_0 + \bar{\mathcal{L}}(\mathbf{a})\cdot(\bar{\Lambda}\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in \mathcal{I}_d. \tag{3.14}$$

设 $\bar{\phi}\colon \mathcal{X} \to \mathcal{X}$ 由 $\bar{\phi}(w, \mathbf{a}, \mathbf{c}\ b) = (w, \bar{\mathcal{L}}(\mathbf{a}), \mathbf{c}, \mathbf{a}\cdot\bar{\mathbf{B}}_0 + b)$ 给出，并将 $\bar{\mu} = \mu\circ\bar{\phi}^{-1}\ \mu$ 在 ϕ 下的推前测度记为 $\bar{\mu} = \mu\circ\bar{\phi}^{-1}$。那么，由变量替换公式和 (3.11) 可知

$$\begin{aligned}I_{\bar{\mu},p} &= \int_{\mathcal{X}}|w|(\|\bar{\mathcal{L}}(\mathbf{a})\|_p + \|\mathbf{c}\| + |\mathbf{a}\cdot\bar{\mathbf{B}}_0 + b|)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &\le \int_{\mathcal{X}}|w|(c_0\|\mathbf{a}\|_p + \|\mathbf{c}\| + \|\mathbf{a}\|_p\|\bar{\mathbf{B}}_0\|_q + |b|)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &\le \max(c_0 + \|\bar{\mathbf{B}}_0\|_q, 1)I_{\mu,p} < \infty\end{aligned}$$

类似地，由 (3.14) 可得，对于任意 $\mathbf{z} \in \mathcal{I}_d$,

$$\begin{aligned}H(\mathbf{z}) &= \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\bar{\mathbf{B}}_0 + \bar{\mathcal{L}}(\mathbf{a})\cdot(\bar{\Lambda}\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\ &= \int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot(\bar{\Lambda}\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\bar{\mu}(dw, d\mathbf{a}, d\mathbf{c}, db).\end{aligned} \tag{3.15}$$

□

3.3 Echo 归一化截断泛函的状态网络逼近 s

现在考虑储备池计算的情形：我们的目标是使用有限维 ELM-ESN 对 $\mathcal{C}$ 类中的未知元素进行近似。也就是说，使用随机生成的回声状态网络式状态方程，并以神经网络作为读出层，其中仅训练输出层。本节首先考虑目标泛函属于有限维循环 Barron 泛函类 $\mathcal{C}_N$ 的情形。

更具体地，考虑回声状态网络

$$\mathbf{x}_t^{\mathrm{ESN}} = \sigma_1(A^{\mathrm{ESN}}\mathbf{x}_{t-1}^{\mathrm{ESN}} + C^{\mathrm{ESN}}\mathbf{z}_t + \mathbf{B}^{\mathrm{ESN}}), \quad t \in \mathbb{Z}_-, \tag{3.16}$$

它具有随机生成矩阵 $A^{\mathrm{ESN}} \in \mathbb{M}_N$, $C^{\mathrm{ESN}} \in \mathbb{M}_{N,\ d}$ 和 $\mathbf{B}^{\mathrm{ESN}} \in \mathbb{R}^N$，用于捕捉动力学，然后接入一个随机神经网络读出层。因此，$\mathcal{C}_N$ 中的元素可近似为

$$\widehat{H}(\mathbf{z}) = \sum_{i=1}^{N} w^{(i)}\sigma_2(\mathbf{a}^{(i)} \cdot \mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{z}) + \mathbf{c}^{(i)} \cdot \mathbf{z}_0 + b_i) \tag{3.17}$$

其系数 $\mathbf{a}^{(i)}$, $\mathbf{c}^{(i)}$, b_i 分别取值于 $\mathbb{R}^N$, $\mathbb{R}^d$ 和 $\mathbb{R}$，其中 $\mathbf{W} = \left(w^{(1)}, .., w^{(N)}\right)^\top \in \mathbb{R}^N$ 是可训练的。

令 $T = \lceil \frac{N}{d} \rceil$ 并定义 Kalman 可控性矩阵

$$K = \pi_{N\times N}(C^{\mathrm{ESN}}|A^{\mathrm{ESN}}C^{\mathrm{ESN}}|\cdots|(A^{\mathrm{ESN}})^{T-2}C^{\mathrm{ESN}}|(A^{\mathrm{ESN}})^{T-1}C^{\mathrm{ESN}}), \tag{3.18}$$

其中，在 $T > \frac{N}{d}$ 的情况下，映射 $\pi_{N\times N}$ 会移除最后 $dT - N$ 列。

命题 3.4. 假设 $H^N\colon (D_d)^{\mathbb{Z}_-} \to \mathbb{R}$ 具有如下类型的归一化实现

$$H^N(\mathbf{z}) = \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a} \cdot (\Lambda_N\mathbf{z}) + \mathbf{c} \cdot \mathbf{z}_0 + b)\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in (D_d)^{\mathbb{Z}_-}, \tag{3.19}$$

对于某个定义在 $\mathcal{X}^N$ 上的 μ^N，满足 $I_{\mu^N} = \int_{\mathcal{X}^N} |w|(\|\mathbf{a}\|+\|\mathbf{c}\|+|b|)\mu^N(dw, d\mathbf{a}\ d\mathbf{c}\ db) < \infty$，且其中 $\Lambda_N\colon (D_d)^{\mathbb{Z}_-} \to \mathbb{R}^N$ 满足 $(\Lambda_N\mathbf{z})_{d(k-1)+j} = \lambda^{k-1}(z_{-k})_j$，对所有满足 $d(k-1)+j \le N$ 的 $k \in \mathbb{N}, j = 1, \ldots, d$ 均成立。假设 $\sigma_1(x) = x$, D_d 有界，并考虑任意一个满足 $\rho\left(A^{\mathrm{ESN}}\right) < 1$ 的 (3.16) 型系统，且 (3.18) 中的矩阵 K 可逆。那么存在定义在 $\mathcal{X}^N$ 上的博雷尔测度 $\tilde{\mu}^N$，使得 $I_{\tilde{\mu}^N} < \infty$，并且映射

$$\bar{H}^N(\mathbf{z}) = \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a} \cdot \mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{z}) + \mathbf{c} \cdot \mathbf{z}_0 + b)\tilde{\mu}^N(dw, d\mathbf{a}, d\mathbf{c}, db), \quad \mathbf{z} \in (D_d)^{\mathbb{Z}_-} \tag{3.20}$$

对所有 $\mathbf{z} \in (D_d)^{\mathbb{Z}_-}$ 均满足以下界

$$\begin{aligned}|H^N(\mathbf{z}) - \bar{H}^N(\mathbf{z})| \le L_{\sigma_2}&\left(|||C^{\mathrm{ESN}}|||M + \|\mathbf{B}^{\mathrm{ESN}}\|\right)\left|||\left(A^{\mathrm{ESN}}\right)^{k_0}|||^{\lfloor \frac{T}{k_0} \rfloor}|||A^{\mathrm{ESN}}|||^{k_0-1}\right.\\ &\times\left(\sum_{i=0}^{k_0-1}\frac{|||\left(A^{\mathrm{ESN}}\right)^i|||}{1-|||\left(A^{\mathrm{ESN}}\right)^{k_0}|||}\right)|||K^{-\top}\Lambda|||\int_{\mathcal{X}^N}|w|\|\mathbf{a}\|\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db),\end{aligned} \tag{3.21}$$

其中，$k_0 \in \mathbb{N}$ 是满足 $|||\left(A^{\mathrm{ESN}}\right)^{k_0}||| < 1$ 的最小整数，$M > 0$ 的选取使得对所有 $\mathbf{u} \in D_d$，$\|\mathbf{u}\| \le M$；$\Lambda \in \mathbb{R}^{N\times N}$ 是对角矩阵，其元素为 $\Lambda_{d(k-1)+j,\ d(k-1)+j} = \lambda^{k-1}$ 对所有$k \in \mathbb{N}$、$j = 1$、.....、d均成立，且满足 $d(k-1) + j \le N$。

注 3.5. 前述结论的一个特例是，选择 A^{ESN} 使得 $|||A^{\mathrm{ESN}}||| < 1$。在这种情况下，不再使用界 (3.21)，而是得到

$$|H^N(\mathbf{z}) - \bar{H}^N(\mathbf{z})| \le L_{\sigma_2}\left(|||C^{\mathrm{ESN}}|||M + \|\mathbf{B}^{\mathrm{ESN}}\|\right)\frac{|||A^{\mathrm{ESN}}|||^T}{1-|||A^{\mathrm{ESN}}|||}|||K^{-\top}\Lambda|||\int_{\mathcal{X}^N}|w|\|\mathbf{a}\|\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db). \tag{3.22}$$

注 3.6. 可以证明（参见 [Grig 21a，命题 4.4]），如果 A^{ESN} 和 C^{ESN} 是使用取值于相应空间中的正则随机变量随机抽取的，那么关于 K 可逆性的假设几乎处处成立。

证明。首先注意到，根据我们的假设和 (2.3)，方程 (3.16) 在$\ell^\infty(\mathbb{R}^N)$ 中有唯一解，其表达式为

$$\mathbf{x}_t^{\mathrm{ESN}} = \sum_{k=0}^{\infty} (A^{\mathrm{ESN}})^k (C^{\mathrm{ESN}} \mathbf{z}_{t-k} + \mathbf{B}^{\mathrm{ESN}}), \quad \text{for any } t \in \mathbb{Z}_-. \tag{3.23}$$

令 $\mathbf{x}_t^N = \sum_{k=0}^{T-1} (A^{\mathrm{ESN}})^k (C^{\mathrm{ESN}} \mathbf{z}_{t-k} + \mathbf{B}^{\mathrm{ESN}})$。则

$$\begin{aligned}
\|\mathbf{x}_t^{\mathrm{ESN}} - \mathbf{x}_t^N\| &\leq \sum_{k=T}^{\infty} \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^k \right|\!\right|\!\right| \left(|||C^{\mathrm{ESN}}||| \|\mathbf{z}_{t-k}\| + \|\mathbf{B}^{\mathrm{ESN}}\|\right) \\
&\leq \left(|||C^{\mathrm{ESN}}|||M + \|\mathbf{B}^{\mathrm{ESN}}\|\right) \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^T \right|\!\right|\!\right| \sum_{k=0}^{\infty} \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^k \right|\!\right|\!\right| \\
&\leq \left(|||C^{\mathrm{ESN}}|||M + \|\mathbf{B}^{\mathrm{ESN}}\|\right) \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^{k_0} \right|\!\right|\!\right|^{\lfloor \frac{T}{k_0} \rfloor} |||A^{\mathrm{ESN}}|||^{k_0-1} \left(\sum_{i=0}^{k_0-1} \frac{\left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^i \right|\!\right|\!\right|}{1 - \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^{k_0} \right|\!\right|\!\right|} \right).
\end{aligned} \tag{3.24}$$

为了得到前一个表达式中的最后一个不等式，我们将 $T = \lfloor \frac{T}{k_0} \rfloor k_0 + r$ 分解为 $0 \leq r < k_0$，从而可以写成

$$\left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^T \right|\!\right|\!\right| = \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^{\lfloor \frac{T}{k_0} \rfloor k_0 + r} \right|\!\right|\!\right| \leq \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^{k_0} \right|\!\right|\!\right|^{\lfloor \frac{T}{k_0} \rfloor} \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^r \right|\!\right|\!\right| \leq \left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^{k_0} \right|\!\right|\!\right|^{\lfloor \frac{T}{k_0} \rfloor} |||A^{\mathrm{ESN}}|||^{k_0-1}.$$

在最后一个不等式中，我们用到了这样一个事实：$k_0 \in \mathbb{N}$ 是满足 $\left|\!\left|\!\left| \left(A^{\mathrm{ESN}}\right)^{k_0} \right|\!\right|\!\right| < 1$ 的最小整数，因此当 $r < k_0$ 时，必有 $|||\left(A^{\mathrm{ESN}}\right)^r||| \geq 1$。

记 $\tilde{\mathbf{z}} \in \mathbb{R}^N$ 为其分量为 $\tilde{\mathbf{z}}^{d(k-1)+j} = (z_{-k})_j$ 的向量，并记 $\Lambda \in \mathbb{R}^{N \times N}$ 为满足 $\Lambda_{d(k-1)+j, d(k-1)+j} = \lambda^{k-1}$ 的对角矩阵，对所有 $k \in \mathbb{N}$, $j = 1, \ldots, d$，其中 $d(k-1) + j \leq N$。于是 $\mathbf{x}_{-1}^N = \sum_{k=1}^T (A^{\mathrm{ESN}})^{k-1} (C^{\mathrm{ESN}} \mathbf{z}_{-k} + \mathbf{B}^{\mathrm{ESN}}) = K\tilde{\mathbf{z}} + \tilde{\mathbf{B}}$，其中 $\tilde{\mathbf{B}} = \sum_{k=1}^T (A^{\mathrm{ESN}})^{k-1} \mathbf{B}^{\mathrm{ESN}}$。令 $\tilde{\phi}\colon \mathcal{X}^N \to \mathcal{X}^N$ 定义为 $\tilde{\phi}(w, \mathbf{a}, \mathbf{c}, b) = (w, K^{-\top}\Lambda \mathbf{a}, \mathbf{c} - (K^{-\top}\Lambda \mathbf{a}) \cdot \tilde{\mathbf{B}} + b)$，并记 $\tilde{\mu}^N = \mu^N \circ \tilde{\phi}^{-1}$ 为 μ^N 在 $\tilde{\phi}$ 下的推前测度。于是，由变量替换公式可得

$$\begin{aligned}
I_{\tilde{\mu}^N} &= \int_{\mathcal{X}^N} |w| (\|K^{-\top}\Lambda \mathbf{a}\| + \|\mathbf{c}\| + |-(K^{-\top}\Lambda \mathbf{a}) \cdot \tilde{\mathbf{B}} + b|) \mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&\leq \max(\left|\!\left|\!\left| K^{-\top}\Lambda \right|\!\right|\!\right|, 1)(1 + \|\tilde{\mathbf{B}}\|) I_{\mu^N} < \infty
\end{aligned}$$

and

$$\begin{aligned}
H^N(\mathbf{z}) &= \int_{\mathcal{X}^N} w \sigma_2((\Lambda \mathbf{a}) \cdot \tilde{\mathbf{z}} + \mathbf{c} \cdot \mathbf{z}_0 + b) \mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&= \int_{\mathcal{X}^N} w \sigma_2((K^{-\top}\Lambda \mathbf{a}) \cdot K\tilde{\mathbf{z}} + \mathbf{c} \cdot \mathbf{z}_0 + b) \mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&= \int_{\mathcal{X}^N} w \sigma_2((K^{-\top}\Lambda \mathbf{a}) \cdot \mathbf{x}_{-1}^N - (K^{-\top}\Lambda \mathbf{a}) \cdot \tilde{\mathbf{B}} + \mathbf{c} \cdot \mathbf{z}_0 + b) \mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&= \int_{\mathcal{X}^N} w \sigma_2(\mathbf{a} \cdot \mathbf{x}_{-1}^N + \mathbf{c} \cdot \mathbf{z}_0 + b) \tilde{\mu}^N(dw, d\mathbf{a}, d\mathbf{c}, db).
\end{aligned} \tag{3.25}$$

结合式 (3.25) 和式 (3.24)，可得

$$
\begin{aligned}
|H^N(\mathbf{z}) - \bar{H}^N(\mathbf{z})| &\leq \int_{\mathcal{X}^N} w|\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^N + \mathbf{c}\cdot\mathbf{z}_0 + b) - \sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)|\tilde{\mu}^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&\leq \int_{\mathcal{X}^N} wL_{\sigma_2}|\mathbf{a}\cdot(\mathbf{x}_{-1}^N - \mathbf{x}_{-1}^{\mathrm{ESN}})|\tilde{\mu}^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&\leq \|\mathbf{x}_{-1}^N - \mathbf{x}_{-1}^{\mathrm{ESN}}\|L_{\sigma_2}\int_{\mathcal{X}^N} w\|K^{-\top}\Lambda\mathbf{a}\|\mu^N(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&\leq L_{\sigma_2}(|||C^{\mathrm{ESN}}|||M + \|\mathbf{B}^{\mathrm{ESN}}\|)\left|\left|\left|\left(A^{\mathrm{ESN}}\right)^{k_0}\right|\right|\right|^{\lfloor \frac{T}{k_0}\rfloor}|||A^{\mathrm{ESN}}|||^{k_0-1} \\
&\quad\cdot\left(\sum_{i=0}^{k_0-1}\frac{\left|\left|\left|\left(A^{\mathrm{ESN}}\right)^i\right|\right|\right|}{1-\left|\left|\left|\left(A^{\mathrm{ESN}}\right)^{k_0}\right|\right|\right|}\right)|||K^{-\top}\Lambda|||I_{\mu^N}.
\end{aligned}
\tag{3.26}
$$

□

3.4 主要逼近结果

本节给出我们的主要逼近结果。在该结果中，我们将有限维 ELM-ESN 对逼近有限维循环 Barron 泛函类 $\mathcal{C}_N$ 中元素的可能性，扩展到循环 Barron 泛函的整个类 $\mathcal{C}$。

在这种情况下，我们假设用于构造 ELM 层（3.17）的随机变量，即读出权重的初始化 $w^{(i)}$ 以及隐藏权重 $\mathbf{a}^{(i)}$、$\mathbf{c}^{(i)}$, b_i（分别取值于 $\mathbb{R}$、$\mathbb{R}^N$、$\mathbb{R}^d$和 $\mathbb{R}$），满足 $(w^{(1)}, \mathbf{a}^{(1)}, \mathbf{c}^{(1)}, b_1)\ldots,(w^{(N)}, \mathbf{a}^{(N)}, \mathbf{c}^{(N)}, b_N)$ 独立同分布（IID）；记其共同分布为 ν。

令 $\Lambda \in \mathbb{R}^{N\times N}$ 为命题 3.4 中引入的对角矩阵，令 π: $\ell^p \to \mathbb{R}^N$ 为投影映射 $\pi(\mathbf{a}) = (a_i)_{i=1,\ldots,N}$；对于 $\mathcal{X}$上的给定博雷尔测度 μ，$\mathbf{B} \in \ell^q$，以及满足 $\rho(A) < 1$的有界线性映射 C: $\mathbb{R}^d \to \ell^q$, A: $\ell^q \to \ell^q$，令 $\mu_{A,B,C} = \mu \circ \psi_{A,B,C}^{-1}$，其中 $\psi_{A,B,C}$: $\mathcal{X} \to \mathcal{X}^N$为映射

$$
\psi_{A,B,C}(w, \mathbf{a}, \mathbf{c}, b) = \left(w, K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a})), \mathbf{c}, -(K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a})))\cdot\left[\sum_{k=1}^{T}(A^{\mathrm{ESN}})^{k-1}\mathbf{B}^{\mathrm{ESN}}\right] + \mathbf{a}\cdot\left[\sum_{k=0}^{\infty}A^k\mathbf{B}\right] + b\right)
$$

其中 $[\bar{\mathcal{L}}(\mathbf{a})]_{d(k-1)+j} = [(\tilde{A}^{k-1}C)^*\mathbf{a}]_j$ 对于 $k \in \mathbb{N}$, $j = 1,\ldots,d$。此外，令

$$
I_{\mu,p}^{(2)} = \left(\int_{\mathcal{X}} w^2\left(\|\mathbf{a}\|_p^2 + \|\mathbf{c}\|^2 + |b|^2 + 1\right)\mu(dw, d\mathbf{a}, d\mathbf{c}, db)\right)^{1/2}.
$$

使用这一记号，我们现在给出一个结果：对于 $H \in \mathcal{C}$ 中满足以下条件的元素，即其表示所对应的状态方程 (2.4) 是线性的且具有唯一解性质的元素，可以用随机生成的 ESN-ELM 对任意精度地逼近；在这些对中，仅训练 ELM 的输出层。我们给出一个逼近界，并明确说明其对 ESN 状态空间维数 N 和输入维数 d 的依赖关系。

定理 3.7。此处假设 $\sigma_1(x) = x$, $p \in (1,\infty)$，且输入集 D_d 有界。设 $H \in \mathcal{C}$ 可由类型 (2.2) 的实现表示，其中 $\mathbf{B} \in \ell^q$，有界线性映射C: $\mathbb{R}^d \to \ell^q$, A: $\ell^q \to \ell^q$满足 $|||A||| < 1$ 和 $I_{\mu}^{(2),\,p} < \infty$。使用本陈述前引入的记号，假设 $\mu_{A,\,B,\,C} \ll \nu$，且 $d\mu_{A,\quad B\,C}$，有界。令 $\lambda \in (|||A|||, 1)$。现在考虑一个随机生成的 ESN-ELM 对，使得 $|||A^{\mathrm{ESN}}||| < 1$，且其中 (3.18) 中的可控性矩阵$K$ 可逆。

那么，存在一个可测函数 f: $(\mathcal{X}^N)^N \to \mathbb{R}^N$，使得具有相同参数和新读出层 $\mathbf{W} = f((w^{(i)}, \mathbf{a}^{(i)}, \mathbf{c}^{(i)}, b_i)_{i=1,\ldots,N})$ 的 ESN-ELM 对 $\widehat{H}$ 满足：对于任意 $\mathbf{z} \in \mathcal{I}_d$，其

逼近误差界

$$\mathbb{E}[|H(\mathbf{z})-\widehat{H}(\mathbf{z})|^2]^{1/2}\leq C_{H,\mathrm{ESN}}\left[\lambda^{\frac{N}{d}}+|||A^{\mathrm{ESN}}|||^T+\frac{1}{N^{\frac{1}{2}}}\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty^{\frac{1}{2}}\right] \tag{3.27}$$

with

$$C_{H,\mathrm{ESN}}=\tilde{c}_1\max\left(\frac{|||C|||}{(1-|||\lambda^{-1}A|||^p)^{1/p}},\frac{\|\mathbf{B}\|_q}{1-|||A|||},1\right)\times\max\left(I_{\mu,p},I_{\mu,p}^{(2)}\right)\cdot\left(1+\frac{|||C^{\mathrm{ESN}}|||+\|\mathbf{B}^{\mathrm{ESN}}\|}{1-|||A^{\mathrm{ESN}}|||}|||K^{-\top}\Lambda|||\right) \tag{3.28}$$

仅依赖于 d、 σ_2、 p、$\mathrm{diam}(D_d)$ 和 λ 的 $\tilde{c}_1$ （见 (3.43)）。

证明。令 $\bar{\Lambda}\colon (D_d)^{\mathbb{Z}_-}\to\ell^q$, $(\bar{\Lambda}\mathbf{z})_{d(k-1)+j}=\lambda^{k-1}(z_{-k})_j$，其中 $k\in\mathbb{N}$、 $j=1$、.....、 d。那么命题 3.2 表明，存在定义在 $\mathcal{X}$ 上的（博雷尔）概率测度 $\bar{\mu}$，满足 $I_{\bar{\mu},p}<\infty$，且归一化表示

$$H(\mathbf{z})=\int_{\mathcal{X}}w\sigma_2(\mathbf{a}\cdot(\bar{\Lambda}\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\bar{\mu}(dw,d\mathbf{a},d\mathbf{c},db),\quad \mathbf{z}\in\mathcal{I}_d, \tag{3.29}$$

成立。注意，从命题 2.6 的证明可得 $(\bar{\Lambda}\mathbf{z})_{d(k-1)+j}=\lambda^{k-1}(z_{-k})_j=\bar{x}_{-1,d(k-1)+j}$，其中 $\bar{\mathbf{x}}$ 是以下方程的唯一解

$$\bar{\mathbf{x}}_t=\sigma_1(\bar{A}\bar{\mathbf{x}}_{t-1}+\bar{C}\mathbf{z}_t),\quad t\in\mathbb{Z}_-, \tag{3.30}$$

其中 $\bar{A}\colon \ell^q\to\ell^q$, $\bar{C}\colon \mathbb{R}^d\to\ell^q$ 由 $(\bar{A}\mathbf{x})_i=\mathbb{1}_{\{i>d\}}\lambda x_{i-d}$, $(\bar{C}\mathbf{z})^i=\mathbb{1}_{\{i\leq d\}}z_i$ 给出，其中 $\mathbf{x}\in\ell^q$, $\mathbf{z}\in\mathbb{R}^d$, $i\in\mathbb{N}$。令 $\hat{}A\in\mathbb{R}^{N\times N}$, $\hat{}C\in\mathbb{R}^{N\times d}$ 定义为 $\hat{}A_{ij}=(\bar{A}\boldsymbol{\epsilon}^j)_i=\mathbb{1}_{\{i>d\}}\lambda\delta_{i-d,\ j}$, $\hat{}C_{ik}=(\bar{C}\mathbf{e}^k)_i=\mathbb{1}_{\{i\leq d\}}\delta_{i,\ k}$，其中 i, $j=1,\dots,N$, $k=1,\dots,d$。因此，命题 3.1 表明，对于每个 $\mathbf{z}\in\mathcal{I}_d$，系统

$$\mathbf{x}_t^N=\sigma_1(\hat{A}\mathbf{x}_{t-1}^N+\hat{C}\mathbf{z}_t),\quad t\in\mathbb{Z}_-, \tag{3.31}$$

存在唯一解 $(\mathbf{x}_t^N)_{t\in\mathbb{Z}_-}\in\ell^\infty_-(\mathbb{R}^N)$，且泛函 $H^N\colon \mathcal{I}_d\to\mathbb{R}$,

$$H^N(\mathbf{z})=\int_{\mathcal{X}}w\sigma_2(\mathbf{a}\cdot\iota(\mathbf{x}_{-1}^N(\mathbf{z}))+\mathbf{c}\cdot\mathbf{z}_0+b)\bar{\mu}(dw,d\mathbf{a},d\mathbf{c},db),\quad \mathbf{z}\in\mathcal{I}_d, \tag{3.32}$$

满足 $H^N\in\mathcal{C}_N$，并且

$$|H(\mathbf{z})-H^N(\mathbf{z})|\leq C_{\mathrm{fin}}\sum_{l=0}^\infty|||\bar{A}|||^l\left(\sum_{i=N+1}^\infty|\bar{x}_{-1-l,i}|^q\right)^{1/q} \tag{3.33}$$

对所有 $\mathbf{z}\in\mathcal{I}_d$，根据 (3.11)

$$C_{\mathrm{fin}}=L_{\sigma_2}\int_{\mathcal{X}}|w|\|\mathbf{a}\|_p\bar{\mu}(dw,d\mathbf{a},d\mathbf{c},db)=L_{\sigma_2}\int_{\mathcal{X}}|w|\|\bar{\mathcal{L}}(\mathbf{a})\|_p\mu(dw,d\mathbf{a},d\mathbf{c},db)\leq c_0L_{\sigma_2}I_{\mu,p}.$$

选择 $M>0$，使得对于所有 $\mathbf{u}\in D_d$， $\|\mathbf{u}\|\leq M$。此外，注意到 $|||\bar{A}|||\leq\lambda$ 和 $\bar{x}_{-1-l,\ (k-1)d+j}=\lambda^{k-1}[z_{-l-k}]_j$。因此，利用 $\|\cdot\|_q^q\leq\max(d^{1-\frac{q}{2}},1)\|\cdot\|^q$ 以及对于所有 $\mathbf{u}\in D_d$ 均有 $\|\mathbf{u}\|\leq M$，由 (3.33) 得到

$$\begin{aligned}|H(\mathbf{z})-H^N(\mathbf{z})|&\leq c_0L_{\sigma_2}I_{\mu,p}\sum_{l=0}^\infty(|||\bar{A}|||)^l\left(\sum_{k=\lfloor\frac{N}{d}\rfloor+1}^\infty[\lambda^{k-1}]^q\sum_{j=1}^d|[z_{-l-k}]_j|^q\right)^{1/q}\\&\leq c_0L_{\sigma_2}I_{\mu,p}\sum_{l=0}^\infty(|||\bar{A}|||)^l\left(\sum_{k=\lfloor\frac{N}{d}\rfloor+1}^\infty[\lambda^{k-1}]^q\max(d^{1-\frac{q}{2}},1)M^q\right)^{1/q}\\&\leq c_0L_{\sigma_2}I_{\mu,p}M\max(d^{\frac{1}{q}-\frac{1}{2}},1)\frac{\lambda^{\lfloor\frac{N}{d}\rfloor}}{(1-\lambda)(1-\lambda^q)^{1/q}}.\end{aligned} \tag{3.34}$$

由 (3.31) 可得，对于 $i=1,\dots,N, t\in\mathbb{Z}_-$，有 $(x_t^N)_i = \mathbb{I}_{\{i>d\}}\lambda(x_{t-1}^N)^{i-d} + \mathbb{I}^{\{i\le d\}}(z_t)_{i\circ}$ 因此，对于所有满足 $d(k-1)+j\le N$ 的 $k\in\mathbb{N}, j=1,\dots,d$，都有 $(x_t^N)_{d(k-1)+j} = \lambda^{k-1}(z_{t-k+1})_j$。特别地，令 Λ_N，使得对于所有满足 $d(k-1)+j\le N$ 的 $k\in\mathbb{N}, j=1,\dots,d$，$(D_d)^{\mathbb{Z}_-}\to\mathbb{R}^N$ 满足 $(\Lambda_N \mathbf{z})_{d(k-1)+j} = \lambda^{k-1}(z_{-k})_j$，则可得 $(x_{-1}^N)_{d(k-1)+j} = \lambda^{k-1}(z_{-k})_j = (\Lambda_N \mathbf{z})_{d(k-1)+j}$。令 $\phi\colon \mathcal{X}\to\mathcal{X}^N$ 为投影映射 $\phi(w,\mathbf{a},\mathbf{c},b) = (w,\pi(\mathbf{a}),\mathbf{c},b)$，其中 $\pi(\mathbf{a}) = (a_i)_{i=1,\dots,N}$，并将 $\bar\mu$ 在 ϕ 下的推前测度记为 $\mu^N = \bar\mu\circ\phi^{-1}$。于是 $I_{\mu^N}<\infty$，由 (3.32) 可得

$$H^N(\mathbf{z}) = \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a}\cdot(\Lambda_N\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\mu^N(dw,d\mathbf{a},d\mathbf{c},db), \quad \mathbf{z}\in\mathcal{I}_d. \tag{3.35}$$

因此，命题 3.4 意味着存在一个定义在 $\mathcal{X}^N$ 上的博雷尔测度 $\tilde\mu^N$，使得 $I_{\tilde\mu^N}<\infty$，并且映射

$$\bar H^N(\mathbf{z}) = \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{z}) + \mathbf{c}\cdot\mathbf{z}_0 + b)\tilde\mu^N(dw,d\mathbf{a},d\mathbf{c},db), \quad \mathbf{z}\in(D_d)^{\mathbb{Z}_-} \tag{3.36}$$

对所有 $\mathbf{z}\in\mathcal{I}_d$，均满足界（3.22），即

$$|H^N(\mathbf{z}) - \bar H^N(\mathbf{z})| \le L_{\sigma_2}\Big(|||C^{\mathrm{ESN}}|||M + \|\mathbf{B}^{\mathrm{ESN}}\|\Big)\frac{|||A^{\mathrm{ESN}}|||^T}{1-|||A^{\mathrm{ESN}}|||}|||K^{-\top}\Lambda||| \int_{\mathcal{X}^N}|w|\|\mathbf{a}\|\mu^N(dw,d\mathbf{a},d\mathbf{c},db), \tag{3.37}$$

其中，$\Lambda\in\mathbb{R}^{N\times N}$ 是一个对角矩阵，其对角元素为 $\Lambda_{d(k-1)+j,\ d(k-1)+j} = \lambda^{k-1}$ 对所有 $k\in\mathbb{N}$、$j=1$、……、d，且满足 $d(k-1)+j\le N$。此外，$\|\cdot\|\le \max(1, d^{\frac12-\frac1p})\|\cdot\|_p$ 与（3.11）推出

$$\begin{aligned}\int_{\mathcal{X}^N}|w|\|\mathbf{a}\|\mu^N(dw,d\mathbf{a},d\mathbf{c},db) &= \int_{\mathcal{X}}|w|\|\pi(\bar{\mathcal{L}}(\mathbf{a}))\|\mu(dw,d\mathbf{a},d\mathbf{c},db)\\ &\le \max(1,d^{\frac12-\frac1p})\int_{\mathcal{X}}|w|\|\bar{\mathcal{L}}(\mathbf{a})\|_p\mu(dw,d\mathbf{a},d\mathbf{c},db)\\ &\le \max(1,d^{\frac12-\frac1p})c_0 I_{\mu,p}.\end{aligned} \tag{3.38}$$

注意，（3.36）中的测度 $\tilde\mu^N$ 由 $\tilde\mu^N = \mu^N\circ\tilde\phi^{-1}$ 给出，其中 $\tilde\phi$ 如命题 3.4 的证明所示，而 $\bar\mu = \mu\circ\bar\phi^{-1}$ 满足 $\bar\phi$，如命题 3.2 的证明所示。然后我们验证 $\tilde\mu^N = \mu_{A,\ B,\ C}$，并且变量替换定理表明

$$\begin{aligned}\int_{\mathcal{X}^N} f(w,\mathbf{a},\mathbf{c},b)\tilde\mu^N(dw,d\mathbf{a},d\mathbf{c},db) &= \int_{\mathcal{X}^N} f(w,K^{-\top}\Lambda\mathbf{a},\mathbf{c},-(K^{-\top}\Lambda\mathbf{a})\cdot\tilde{\mathbf{B}}+b)\mu^N(dw,d\mathbf{a},d\mathbf{c},db)\\ &= \int_{\mathcal{X}} f(w,K^{-\top}\Lambda\pi(\mathbf{a}),\mathbf{c},-(K^{-\top}\Lambda\pi(\mathbf{a}))\cdot\tilde{\mathbf{B}}+b)\bar\mu(dw,d\mathbf{a},d\mathbf{c},db)\\ &= \int_{\mathcal{X}} f(w,K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a})),\mathbf{c},-(K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a})))\cdot\tilde{\mathbf{B}}+\mathbf{a}\cdot\bar{\mathbf{B}}_0+b)\mu(dw,d\mathbf{a},d\mathbf{c},db)\end{aligned} \tag{3.39}$$

对于任意函数 f，只要最后一行中的被积函数关于 μ 可积，且其中 $\tilde{\mathbf{B}} = \sum_{k=1}^T (A^{\mathrm{ESN}})^{k-1}\mathbf{B}^{\mathrm{ESN}}$、$\bar{\mathbf{B}}_0 = \sum_{k=0}^\infty A^k\mathbf{B}\in\ell^q$ 和 $\bar{\mathcal{L}}\colon \ell^p\to\ell^p$ 是线性的并满足 (3.11)。

设定 $W_i = \dfrac{w^{(i)}}{N}\dfrac{d\tilde\mu^N}{d\nu}(w^{(i)},\mathbf{a}^{(i)},\mathbf{c}^{(i)},b_i)$，并注意到

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^N W_i\sigma_2(\mathbf{a}^{(i)}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}} + \mathbf{c}^{(i)}\cdot\mathbf{z}_0 + b_i)\right] &= \int_{\mathcal{X}^N} w\frac{d\tilde\mu^N}{d\nu}(w,\mathbf{a},\mathbf{c},b)\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}} + \mathbf{c}\cdot\mathbf{z}_0 + b)\nu(dw,d\mathbf{a},d\mathbf{c},db)\\ &= \int_{\mathcal{X}^N} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}} + \mathbf{c}\cdot\mathbf{z}_0 + b)\tilde\mu^N(dw,d\mathbf{a},d\mathbf{c},db) = \bar H^N(\mathbf{z}).\end{aligned}$$

因此，

$$\begin{aligned}\mathbb{E}[|\bar{H}^N(\mathbf{z})-\widehat{H}(\mathbf{z})|^2] &= \mathbb{E}[|\bar{H}^N(\mathbf{z})-\sum_{i=1}^N W_i\sigma_2(\mathbf{a}^{(i)}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}^{(i)}\cdot\mathbf{z}_0+b_i)|^2]\\ &= \mathrm{Var}\left(\sum_{i=1}^N W_i\sigma_2(\mathbf{a}^{(i)}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}^{(i)}\cdot\mathbf{z}_0+b_i)\right)\\ &= \sum_{i=1}^N \mathrm{Var}(W_i\sigma_2(\mathbf{a}^{(i)}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}^{(i)}\cdot\mathbf{z}_0+b_i))\\ &\le N\mathbb{E}[W_1^2\sigma_2(\mathbf{a}^{(1)}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}^{(1)}\cdot\mathbf{z}_0+b_1)^2].\end{aligned} \tag{3.40}$$

如上所述，$\tilde{\mu}^N=\mu_{A,\ B,\ C}$ 并且根据假设，$d\mu_{A,\quad B\ C}$，</style}有界，因此

$$\frac{}{d\nu}$$

$\mathbb{E}[W_1^2\sigma_2(\mathbf{a}^{(1)}\cdot\mathbf{x}_{-1}^{\text{回声状态网络}}+\mathbf{c}^{(\ )}\cdot\mathbf{z}_0+b_1)\ ]$

$$\begin{aligned}&= \int_{\mathcal{X}^N}\frac{w^2}{N^2}\frac{d\tilde{\mu}^N}{d\nu}(w,\mathbf{a},\mathbf{c},b)\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}\cdot\mathbf{z}_0+b)^2\tilde{\mu}^N(dw,d\mathbf{a},d\mathbf{c},db)\\ &\le 2\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty\int_{\mathcal{X}^N}\frac{w^2}{N^2}[L_{\sigma_2}^2|\mathbf{a}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}\cdot\mathbf{z}_0+b|^2+|\sigma_2(0)|^2]\tilde{\mu}^N(dw,d\mathbf{a},d\mathbf{c},db)\\ &= 2\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty\int_{\mathcal{X}}\frac{w^2}{N^2}[L_{\sigma_2}^2|K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a}))\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}+\mathbf{c}\cdot\mathbf{z}_0-(K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a})))\cdot\tilde{\mathbf{B}}+\mathbf{a}\cdot\bar{\mathbf{B}}_0+b|^2+|\sigma_2(0)|^2]\mu(dw,d\mathbf{a},d\mathbf{c},db)\\ &\le 2\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty\int_{\mathcal{X}}\frac{w^2}{N^2}[L_{\sigma_2}^2|\|\pi(\bar{\mathcal{L}}(\mathbf{a}))\|\|\Lambda K^{-1}(\mathbf{x}_{-1}^{\mathrm{ESN}}-\tilde{\mathbf{B}})\|+\|\mathbf{c}\|\|\mathbf{z}_0\|+\|\mathbf{a}\|_p\|\bar{\mathbf{B}}_0\|_q+|b|^2+|\sigma_2(0)|^2]\mu(dw,d\mathbf{a},d\mathbf{c},db)\\ &\le \frac{c_1}{N^2}\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty\max(c_0^2,\|\bar{\mathbf{B}}_0\|_q^2,1)(1+\left|\left|\left|\Lambda K^{-1}\right|\right|\right|^2\|\mathbf{x}_{-1}^{\mathrm{ESN}}-\tilde{\mathbf{B}}\|^2)\int_{\mathcal{X}}w^2[\|\mathbf{a}\|_p^2+\|\mathbf{c}\|^2+|b|^2+1]\mu(dw,d\mathbf{a},d\mathbf{c},db)\end{aligned}$$

wh此处我们使用了 (3.11)，$\|\cdot\|\le\max(1,d^{\frac{1}{2}-\frac{1}{p}})\|\cdot\|_p$、$c_0=d^{1/(2p)}|||C|||(1-\left|\left|\left|\tilde{A}\right|\right|\right|^p)^{-1/p}$ 以及 se t

$$c_1=8\max(1,d^{\frac{1}{2}-\frac{1}{p}})^2\max(L_{\sigma_2},|\sigma_2(0)|)^2\max(M^2,1). \tag{3.41}$$

将其与 (3.40)、(3.34)、(3.37) 和 (3.38) 结合，便得到

$$\begin{aligned}&\mathbb{E}[|H(\mathbf{z})-\widehat{H}(\mathbf{z})|^2]^{1/2}\\ &\le \mathbb{E}[|H(\mathbf{z})-H^N(\mathbf{z})|^2]^{1/2}+\mathbb{E}[|H^N(\mathbf{z})-\bar{H}^N(\mathbf{z})|^2]^{1/2}+\mathbb{E}[|\bar{H}^N(\mathbf{z})-\widehat{H}(\mathbf{z})|^2]^{1/2}\\ &\le c_0L_{\sigma_2}I_{\mu,p}M\max(1,d^{\frac{1}{q}-\frac{1}{2}})\frac{\lambda^{\lfloor\frac{N}{d}\rfloor}}{(1-\lambda)(1-\lambda^q)^{1/q}}+L_{\sigma_2}\left(\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right|M+\|\mathbf{B}^{\mathrm{ESN}}\|\right)\frac{\left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^T}{1-|||A^{\mathrm{ESN}}|||}\left|\left|\left|K^{-\top}\Lambda\right|\right|\right|\max\left(1,d^{\frac{1}{2}-\frac{1}{p}}\right)c_0I_{\mu,p}\\ &\quad+\frac{c_1^{1/2}}{N^{1/2}}\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty^{1/2}\max(c_0,\|\bar{\mathbf{B}}_0\|_q,1)(1+\left|\left|\left|\Lambda K^{-1}\right|\right|\right|\|\mathbf{x}_{-1}^{\mathrm{ESN}}-\tilde{\mathbf{B}}\|)I_{\mu,p}^{(2)}\\ &\le \tilde{c}_1\max\left(\frac{|||C|||}{(1-|||\lambda^{-1}A|||^p)^{1/p}},\frac{\|\mathbf{B}\|_q}{1-|||A|||},1\right)[I_{\mu,p}\lambda^{\frac{N}{d}}+\left(\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right|+\|\mathbf{B}^{\mathrm{ESN}}\|\right)\frac{\left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^T}{1-|||A^{\mathrm{ESN}}|||}\left|\left|\left|K^{-\top}\Lambda\right|\right|\right|I_{\mu,p}\\ &\quad+\frac{1}{N^{\frac{1}{2}}}\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty^{\frac{1}{2}}(1+\left|\left|\left|\Lambda K^{-1}\right|\right|\right|\left(\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right|+\|\mathbf{B}^{\mathrm{ESN}}\|\right)\frac{1}{1-|||A^{\mathrm{ESN}}|||})I_{\mu,p}^{(2)}]\\ &\le \tilde{c}_1\max\left(\frac{|||C|||}{(1-|||\lambda^{-1}A|||^p)^{1/p}},\frac{\|\mathbf{B}\|_q}{1-|||A|||},1\right)\max(I_{\mu,p},I_{\mu,p}^{(2)})(1+\frac{\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right|+\|\mathbf{B}^{\mathrm{ESN}}\|}{1-|||A^{\mathrm{ESN}}|||}\left|\left|\left|K^{-\top}\Lambda\right|\right|\right|)\\ &\quad\times\left[\lambda^{\frac{N}{d}}+\left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^T+\frac{1}{N^{\frac{1}{2}}}\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty^{\frac{1}{2}}\right]\end{aligned} \tag{3.42}$$

with

$$\tilde{c}_1=8^{\frac{1}{2}}\max(L_{\sigma_2},|\sigma_2(0)|,1)\max(1,d^{\frac{1}{2}-\frac{1}{p}})d^{\frac{1}{2p}}\max(1,d^{\frac{1}{q}-\frac{1}{2}})(1-\lambda)^{-1}(1-\lambda^q)^{-\frac{1}{q}}\lambda^{-1}\max(M,1)^2. \tag{3.43}$$

□

注记 3.8. 如果用作逼近器的 ESN-ELM 满足更一般的条件 $\rho(A^{\mathrm{ESN}})<1$，则可以得到与 (3.27) 中类似的界。在这种情况下，只需在 (3.37) 中用更一般的版本 (3.21) 替换不等式 (3.22)，即可证明该定理。

推论 3.9. 与定理 3.7 中一样，假设循环激活函数为 $\sigma_1(x) = x$, $p \in (1, \infty)$，且输入集 D_d 有界。设 $H \in \mathcal{C}$ 可按 (2.2) 的形式实现，其测度为 μ，并且有界线性映射 $C\colon \mathbb{R}^d \to \ell^q$, $A\colon \ell^q \to \ell^q$ 满足 $|||A||| < 1$ 和 $I_\mu^{(2),\ p} < \infty$。令 $\lambda \in (|||A|||, 1)$。现在考虑一个随机构造的 ESN-ELM 对，使得 $|||A^{\mathrm{ESN}}||| < 1$，并且其 (3.18) 中的可控性矩阵 K 可逆。则存在一个关于读出隐藏层权重的分布 ν，以及一个读出层 W（一个 $\mathbb{R}^N$ 值随机向量），使得 ESN-ELM$\frown H$ 满足：对任意 $\mathbf{z} \in \mathcal{I}_d$，其逼近误差界为

$$\mathbb{E}[|H(\mathbf{z}) - \widehat{H}(\mathbf{z})|^2]^{1/2} \leq C_{H,\mathrm{ESN}} \left[\lambda^{\frac{N}{d}} + |||A^{\mathrm{ESN}}|||^{\frac{N}{d}} + \frac{1}{N^{\frac{1}{2}}}\right] \tag{3.44}$$

其中 $C_{H,\ \mathrm{ESN}}$ 如式 (3.28) 所示。

证明。注意到 $|||A^{\mathrm{ESN}}|||^T \leq |||A^{\mathrm{ESN}}|||^{\frac{N}{d}}$ 由 $T \geq \frac{N}{d}$ 得出，并取 $\nu = \mu_{A,\ B,\ C}$，使得 $d\mu_{A,\ B\ C}/d\nu = 1$，于是由定理 3.7 可得式 (3.44)。 □

注记 3.10。维数灾难。误差界 (3.44) 由常数 $C_{H,\ \mathrm{ESN}}$ 和三个显式依赖于 N 的项组成。常数 $C_{H,\ \mathrm{ESN}}$ 可能通过范数 $|||A^{\mathrm{ESN}}|||$、$\|\mathbf{B}^{\mathrm{ESN}}\|$、$|||C^{\mathrm{ESN}}|||$ 依赖于 N，但对于这些范数，合理的假设是它们不依赖于 N（实际应用中，这些矩阵的范数会经过归一化）。同样，假设范数 $|||K^{-\top}\Lambda|||$ 在 N 中有界似乎是合理的，因为范数 $|||A^{\mathrm{ESN}}|||$ 很可能有下界。这一论证表明，在实际相关的情形下，该常数不依赖于 N；并且从显式表达式 (3.43) 可以看出，它对维数 d 的依赖仅为多项式级。为使逼近误差至多为 ε，因此可取 $N = \left\lceil \max\left(d\log\left(\frac{\varepsilon}{3C_{H,\ \mathrm{ESN}}}\right) \log(\lambda)\ d\log\left(\frac{\varepsilon}{3C_{H,\ \mathrm{ESN}}}\right) \log(|||A^{\mathrm{ESN}}|||)^{-1} 9C_{H,\ \mathrm{ESN}}^2 \varepsilon^{-2}\right)\right\rceil$，其增长仅关于 d 和 ε^{-1} 呈多项式级。因此，在这些情况下，误差界 (3.44) 中不存在维数灾难。

注记 3.11. 实现。从实践角度看，定理 3.7 和推论 3.9 中得到的界适用于两种不同的学习过程：第一种情况下，所选的极限学习机权重分布 ν 满足定理 3.7 中的绝对连续性假设，只需训练极限学习机输出层。第二种情况下，若不满足该条件，则还需要训练所有 ELM 权重。不过请注意，这些并非循环权重，因此这里不会出现梯度消失问题。相比之下，在通过时间反向传播训练所有权重的标准循环神经网络中，有时会出现这一问题。

本节最后，我们考虑另一种实现情形：在这种情形下，我们固定用于 ELM 采样的测度，并说明仍然可以使用在 1-瓦瑟斯坦度量意义下与该测度任意接近的测度进行采样，代价是误差项可能增大，而这种增大可以通过增加回声状态网络的维数来补偿。令 $\mathcal{W}_1$ 为 1-瓦瑟斯坦度量。假设我们使用测度 ν_0 生成隐藏层权重。下一个推论将说明，当我们从给定测度 ν_0“近似”采样时，误差会如何增大。

推论 3.12。考虑与推论 3.9 相同的情形，并假设 ν_0 具有有限一阶矩，且 $\int_{\mathcal{X}} [|w| + \|\mathbf{a}\|_p + \|\mathbf{c}\| + |b|]\mu(dw, d\mathbf{a}\ d\mathbf{c}\ db) < \infty$。那么，对于任意 $\delta \in (0, 1)$，存在一个概率测度 ν，满足 $\mathcal{W}_1(\nu_0, \nu) \leq \delta$，以及一个读出层 W（取值于 $\mathbb{R}^N$ 的随机向量），使得对于任意 $\mathbf{z} \in \mathcal{I}_d$，具有隐藏层权重分布 ν 的 ESN-ELM 满足如下逼近误差界

$$\mathbb{E}[|H(\mathbf{z}) - \widehat{H}(\mathbf{z})|^2]^{1/2} \leq \tilde{C}_{H,\mathrm{ESN}} \left[\lambda^{\frac{N}{d}} + |||A^{\mathrm{ESN}}|||^{\frac{N}{d}} + \frac{1}{N^{\frac{1}{2}}} \min(\delta^{-\frac{1}{2}}, \mathcal{J}_{\nu_0})\right] \tag{3.45}$$

其中 $\mathcal{J}_{\nu_0} = \left\| d\mu_{A,\ B\ C}/d\nu_0 \right\|_\infty^{\frac{1}{2}}$，$\tilde{C}_{H,\ \mathrm{ESN}} = C_{H,\ \mathrm{ESN}} \max(1, \mathcal{W}_1(\nu_0, \mu_{A,\ B,\ C}))^{\frac{1}{2}}$ 以及 $C_{H,\ \mathrm{ESN}}$，如 (3.28) 所示。

证明。首先注意到 $\left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^T \le \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^{\frac{N}{d}}$，因为 $T \ge \frac{N}{d}$。在 $\mu_{A,\ B,\ C} \ll \nu_0$ 和 $d\mu_{A,\ B\ C}/d\nu_0$ 有界的情况下，也就是说，$\mathcal{J}_{\nu_0} < \infty$，我们可以选择 $\nu = \nu_0$，并由定理 3.7 得到如下界

$$\mathbb{E}[|H(\mathbf{z}) - \widehat{H}(\mathbf{z})|^2]^{1/2} \le \tilde{C}_{H,\mathrm{ESN}} \left[\lambda^{\frac{N}{d}} + \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^{\frac{N}{d}} + \frac{1}{N^{\frac{1}{2}}}\mathcal{J}_{\nu_0}\right]. \tag{3.46}$$

现在我们推导一个替代界（无论 $\mathcal{J}_{\nu_0}$ 是否有限）。结合这两个界，便可得到界 (3.45)。

由 (3.39) 和估计 $\|\pi(\bar{\mathcal{L}}(\mathbf{a}))\| \le \max(1, d^{\frac{1}{2}-\frac{1}{p}})c_0\|\mathbf{a}\|_p$，我们得到

$$\begin{aligned}
&\int_{\mathcal{X}^N} (|w| + \|\mathbf{a}\| + \|\mathbf{c}\| + |b|)\, \mu_{A,B,C}(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&\quad = \int_{\mathcal{X}} \left(|w| + \|K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a}))\| + \|\mathbf{c}\| + |-(K^{-\top}\Lambda\pi(\bar{\mathcal{L}}(\mathbf{a})))\cdot\tilde{\mathbf{B}} + \mathbf{a}\cdot\bar{\mathbf{B}}_0 + b|\right)\mu(dw, d\mathbf{a}, d\mathbf{c}, db) \\
&\quad \le \max(1, \left|\left|\left|K^{-\top}\Lambda\right|\right|\right| \max(1, d^{\frac{1}{2}-\frac{1}{p}})c_0) \\
&\quad \times (1 + \|\tilde{\mathbf{B}}\| + \|\bar{\mathbf{B}}_0\|_q) \int_{\mathcal{X}} (|w| + \|\mathbf{a}\|_p + \|\mathbf{c}\| + |b|)\, \mu(dw, d\mathbf{a}, d\mathbf{c}, db) < \infty
\end{aligned} \tag{3.47}$$

因此 $\mu_{A,\ B,\ C}$ 具有有限一阶矩。因此，$\tilde{\delta} = \delta[\max(1, \mathcal{W}_1(\nu_0, \mu_{A,\ B,\ C}))]^{-1} \in (0, 1)$，且 $\tilde{\nu} = \delta\mu_{A,\ B,\ C} + (1-\delta)\nu_0$ 也具有有限一阶矩。因此，我们可以使用 Kantorovich-Rubinshtein 对偶（参见 [Vill 09，定理 5.10；特殊情况 5.16]）进行计算

$$\begin{aligned}
\mathcal{W}_1(\nu_0, \nu) &= \sup_{f:\ f \text{ is 1-Lipschitz}} \left\{\int_{\mathcal{X}^N} f(x)(\nu_0 - \nu)(dx)\right\} \\
&= \tilde{\delta} \sup_{f:\ f \text{ is 1-Lipschitz}} \left\{\int_{\mathcal{X}^N} f(x)(\nu_0 - \mu_{A,B,C})(dx)\right\} = \tilde{\delta}\mathcal{W}_1(\nu_0, \mu_{A,B,C}) \le \delta.
\end{aligned}$$

此外，$\mu_{A,\ B,\ C} \ll \nu$ 和

$$\tilde{\delta}\frac{d\mu_{A,B,C}}{d\nu} + (1-\tilde{\delta})\frac{d\nu_0}{d\nu} = 1,$$

因此 $\frac{d\nu_0}{d\nu} \ge 0$ 蕴含 $\left\|d\mu_{A,\ B\ C}/d\nu\right\|_\infty \le \frac{1}{\tilde{\delta}}$。因此，定理 3.7 得出 (3.45)。 □

3.5 普适性

作为一项补充结果，我们得到以下普适性结果：它表明，任何平方可积泛函（不一定属于循环 Barron 类）都可以通过适当选择读出层 W，并将隐藏权重从与给定测度 ν_0 任意接近的分布 ν 中生成，从而由 ESN-ELM (3.16)-(3.17) 良好逼近。

推论 3.13. 假设输入集 D_d 有界，并考虑任意泛函 $H\colon \mathcal{I}_d \to \mathbb{R}$，我们将使用按以下规格构建的 ESN-ELM 对其进行逼近：激活函数为 $\sigma_1(x) = x$，且满足以下二者之一：$\sigma_2(x) = \max(x, 0)$，或 σ_2 是有界、Lipschitz 连续、非常值的函数。此外，假设存在 $\bar{c} > 0$ 和 $\underline{l}, l \in (0, 1)$，$\underline{l} < l$，使得对于任意选择 $N \in \mathbb{N}$，ESN 参数均满足：式 (3.18) 中的可控性矩阵 K 可逆，$\underline{l} < \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right| < \bar{l}$，$\|\mathbf{B}^{\mathrm{ESN}}\| \le \bar{c}$，$\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right| \le \bar{c}$，且 $\left|\left|\left|K^{-1}\Lambda\right|\right|\right| \le \bar{c}$，其中 $\Lambda \in \mathbb{R}^{N\times N}$ 是对角矩阵，其对角元素为 $\Lambda_{d(k-1)+j, d(k-1)+j} = \underline{l}^{k-1}$，对所有 $k \in \mathbb{N}, j = 1, \ldots, d$ 均有 $d(k-1)+j \le N$。设 ν_0 为给定的、具有有限一阶矩的隐藏权重分布。那么，对于 $\mathcal{I}_d \subset (D_d)^{\mathbb{Z}_-}$ 上的任意概率测度 γ，只要满足 $H \in L^2(\mathcal{I}_d, \gamma)$，就存在一个概率测度 ν，满足 $\mathcal{W}_1(\nu_0, \nu) < \varepsilon$，以及一个读出层 W（一个 $\mathbb{R}^N$ 值随机向量），使得具有读出层 W 且隐藏层权重分布为 ν 的 ESN-ELM $\widehat{H}$ 满足

$$\left(\int_{\mathcal{I}_d} \mathbb{E}[|H(\mathbf{z}) - \widehat{H}(\mathbf{z})|^2]\gamma(d\mathbf{z})\right)^{1/2} < \varepsilon.$$

证明。首先，根据命题 2.10，存在 $H^{RC}\in\mathcal{C}\cap L^2(\mathcal{I}_d,\gamma)$，使得 $\|H-H^{RC}\|_{L^2(\mathcal{I}_d,\gamma)}<\varepsilon/2$。由命题 2.10 的证明以及 [Gono 20c] 中的构造可知，H^{RC} 具有如下形式

$$H^{RC}(\mathbf{z})=\int_{\mathcal{X}^{\bar N}} w\sigma_2(\mathbf{a}\cdot(\pi_{\bar N}\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\mu^{\bar N}(dw,d\mathbf{a},d\mathbf{c},db),\quad \mathbf{z}\in(D_d)^{\mathbb{Z}_-},\tag{3.48}$$

对于某个 $\bar N\in\mathbb{N}$、一个原子概率测度 $\mu^{\bar N}$，并且有 $\pi_{\bar N}\colon (D_d)^{\mathbb{Z}_-}\to\mathbb{R}^{\bar N}$ 满足

$$(\pi_{\bar N}(\mathbf{z}))_{d(k-1)+j}=(z_{-k})_j,\quad \text{for all } k\in\mathbb{N},\, j=1,\dots,d \text{ with } d(k-1)+j\le\bar N.$$

现在令 $\bar\lambda=\underline{l}\in(0,1)$、$p\in(1,\infty)$、$\mathrm{B}=0$，并定义 $A\colon \ell^q\to\ell^q$、$C\colon \mathbb{R}^d\to\ell^q$，使得 $(A\mathbf{x})_i=\mathbb{1}_{\{i>d\}}\bar\lambda x_{i-d}$、$(C\mathbf{z})_i=\mathbb{1}_{\{i\le d\}}z_i$，其中 $\mathbf{x}\in\ell^q$、$\mathbf{z}\in\mathbb{R}^d$，$i\in\mathbb{N}$。然后，按照命题 2.6 的证明，我们得到 (2.1) 存在唯一解 $(\mathbf{x}_t)_{t\in\mathbb{Z}_-}\in\ell^\infty(\ell^q)$，$\bar\lambda^{k-1}z_{t-k+1,j}$，$\iota\colon \mathbb{R}^{\bar N}\to\ell^p$，$(x_1\ \dots\ x_{\bar N})^\top\mapsto(x_1\ \dots\ x_{\bar N}\ 0\ \dots)$，并且对于 $k\in\mathbb{N}$，$j=1,\dots,d$，有 $\dots$。令 $\dots$ 为自然嵌入 $(\dots)$ $(\dots,\dots)$，并令 $\pi\colon \ell^q\to\mathbb{R}^{\bar N}$ 为自然投影。然后

$$\mathbf{a}\cdot(\pi_{\bar N}\mathbf{z})=\mathbf{a}\cdot(\Lambda^{-1}\pi\mathbf{x}_{-1})=(\iota\Lambda^{-1}\mathbf{a})\cdot\mathbf{x}_{-1}.$$

令 $\phi\colon \mathcal{X}^{\bar N}\to\mathcal{X}$ 由 $\phi(w,\mathbf{a},\mathbf{c},b)=(w,\iota\Lambda^{-1}\mathbf{a},\mathbf{c},b)$ 定义，并将 $\mu=\mu^{\bar N}\circ\phi^{-1}$ 记为 $\mu^{\bar N}$ 在 ϕ 下的推前测度。于是

$$\begin{aligned}H^{RC}(\mathbf{z})&=\int_{\mathcal{X}^{\bar N}} w\sigma_2((\iota\Lambda^{-1}\mathbf{a})\cdot\mathbf{x}_{-1}+\mathbf{c}\cdot\mathbf{z}_0+b)\mu^{\bar N}(dw,d\mathbf{a},d\mathbf{c},db)\\&=\int_{\mathcal{X}} w\sigma_2(\mathbf{a}\cdot\mathbf{x}_{-1}(\mathbf{z})+\mathbf{c}\cdot\mathbf{z}_0+b)\mu(dw,d\mathbf{a},d\mathbf{c},db).\end{aligned}\tag{3.49}$$

因此，$H^{RC}\in\mathcal{C}$ 具有形式为 (2.2) 的表示，其中 $|||A|||=\bar\lambda<1$。此外，由于 $\mu^{\bar N}$ 是原子的，可直接得出 $I^{(2)}_{\mu,p}<\infty$ 且 $\int_{\mathcal{X}}[|w|+\|\mathbf{a}\|_p+\|\mathbf{c}\|+|b|]\mu(dw,d\mathbf{a},d\mathbf{c},db)<\infty$。

接下来，任取 $\lambda\in(|||A|||,1)$，并注意到在关于 ESN 的假设下，(3.28) 中的常数 $C_{H,\mathrm{ESN}}$ 可估计为

$$C_{H,\mathrm{ESN}}\le\tilde c_1\max\Big(\frac{1}{(1-(\lambda^{-1}\bar\lambda)^p)^{\frac1p}},1\Big)\max(I_{\mu,p},I^{(2)}_{\mu,p})\left(1+2\frac{\bar c^2}{1-\bar l}\right)=:\bar C_{H,0},\tag{3.50}$$

这不依赖于 N。类似地，界

$$\mathcal{W}_1(\nu_0,\mu_{A,B,C})\le\int_{\mathcal{X}^N}\|x\|\nu_0(dx)+\int_{\mathcal{X}^N}\|x\|\mu_{A,B,C}(dx),$$

回声状态网络矩阵的假设，以及 (3.47)，可用于得到 $C_{H,1}$ 关于 $\mathcal{W}_1(\nu_0,\mu_{A,B,C})$ 的一个上界，该上界仅依赖于 $\mu,d,p,\lambda,\bar c,\underline{l},\bar l$ 和 ν_0，而不依赖于 N。令 $\bar c_H=\bar C_{H,0}\max(1,\bar C_{H,1})^{\frac12}$。现在我们对 $H=H^{RC}$ 应用推论 3.12。我们选取 $\delta\in(0,1)$，使得 $\delta<\varepsilon$，并且

$$N=\left\lceil\max\left(d\log\left(\frac{\varepsilon}{6\bar c_H}\right)\frac{1}{\log(\lambda)},d\log\left(\frac{\varepsilon}{6\bar c_H}\right)\frac{1}{\log(\bar l)},(6\bar c_H\varepsilon^{-1})^2\delta^{-1}\right)\right\rceil,$$

于是，根据推论 3.12，我们得到存在一个概率测度 ν，满足 $\mathcal{W}_1(\nu_0,\nu)<\varepsilon$，以及一个读出层 W（一个取值于 $\mathbb{R}^N$ 的随机向量），使得具有读出层 W、隐藏层权重分布为 ν 的 ESN-ELM $\widehat H$ 对任意 $\mathbf{z}\in\mathcal{I}_d$ 均满足

$$\mathbb{E}[|H^{RC}(\mathbf{z})-\widehat H(\mathbf{z})|^2]^{1/2}\le\bar c_H\left[\lambda^{\frac Nd}+|||A^{\mathrm{ESN}}|||^{\frac Nd}+\frac{1}{\delta^{\frac12}N^{\frac12}}\right]<\frac{\varepsilon}{2}.\tag{3.51}$$

因此，

$$\left(\int_{\mathcal{I}_d}\mathbb{E}[|H(\mathbf{z})-\widehat H(\mathbf{z})|^2]\gamma(d\mathbf{z})\right)^{1/2}\le\|H-H^{RC}\|_{L^2(\mathcal{I}_d,\gamma)}+\left(\int_{\mathcal{I}_d}\mathbb{E}[|H^{RC}(\mathbf{z})-\widehat H(\mathbf{z})|^2]\gamma(d\mathbf{z})\right)^{1/2}<\varepsilon.$$

□

3.6 特殊情形：静态情形

作为一个特殊情因此，我们考虑无循环的情况，即当 H: $D_d \to \mathbb{R}$ 的形式为 m

$$H(\mathbf{u}) = \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} w\sigma_2(\mathbf{c}\cdot\mathbf{u}+b)\mu_0(dw,d\mathbf{c},db), \quad \mathbf{u}\in D_d\subset\mathbb{R}^d, \tag{3.52}$$

对于满足 $\int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} |w|(\|\mathbf{c}\|+|b|)\mu_0(dw,d\mathbf{c}\ db) < \infty$ 的概率测度 μ_0 on $\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}$。 H 显然是循环广义 Barron 泛函的一个特殊情形，因为它可以表示为式 (2.2)，其中在 $\mathcal{X}$ 上的测度 μ 由 $\mu(dw,d\mathbf{a}\ d\mathbf{c}\ db) = \delta_0(d\mathbf{a})\mu_0(dw,d\mathbf{c}\ db)$ 给出，并且根据 μ_0 的可积性假设，满足 $I_{\mu,\ p}<\infty$。

在这种情况下，我们证明，循环 Barron 类的这类静态元素 H 可以由极限学习机近似，即由前馈神经网络近似

$$\widehat{H}(\mathbf{u}) = \sum_{i=1}^{N} W_i\sigma_2(\mathbf{c}^{(i)}\cdot\mathbf{u}+b_i) \tag{3.53}$$

其系数 $\mathbf{c}^{(i)}$, b_i 分别为取值于 $\mathbb{R}^d$ 和 $\mathbb{R}$ 的随机生成系数，而 $\mathbf{W}\in\mathbb{R}^N$ 可训练。$\widehat{H}$ 可视为泛函 $\mathbf{z}\mapsto\widehat{H}(\mathbf{z}_0)$，其形式为 (3.17)，其中 $\mathbf{a}^{(i)}=0$。设 $w^{(1)},\ldots,w^{(n)}$ 为 $\mathbb{R}$-值随机变量，假设随机变量 $(w^{(1)},\mathbf{c}^{(1)},b_1)\ldots,(w^{(N)},\mathbf{c}^{(N)},b_N)$ 独立同分布，并记 ν_0 为它们的共同分布。

推论 3.14. 假设 H 如 (3.52) 所示，且相应的 μ 满足 $I^{(2)}_{\mu,\ p}<\infty$。假设输入集 D_d 有界。再假设 $\mu_0\ll\nu_0$，且 $\dfrac{d\mu_0}{d\nu_0}$ 有界。那么存在可测函数 f: $(\mathbb{R}\times\mathbb{R}^d\times\mathbb{R})^N\to\mathbb{R}^N$，使得具有读出层 $\mathbf{W}=f((w^{(i)},\mathbf{c}^{(i)},b_i)_{i=1,\ldots,N})\in\mathbb{R}^N$ 的 ELM $\widehat{H}$ 对任意 $\mathbf{u}\in D_d$ 都满足以下逼近误差界

$$\mathbb{E}[|H(\mathbf{u})-\widehat{H}(\mathbf{u})|^2]^{1/2} \le \frac{C_H}{N^{\frac{1}{2}}} \tag{3.54}$$

with

$$C_H = (2\max(2L^2_{\sigma_2},|\sigma_2(0)|^2)\max(1,\sup_{\mathbf{v}\in D_d}\|\mathbf{v}\|^2))^{\frac{1}{2}} I^{(2)}_{\mu,p}\left\|\frac{d\mu_0}{d\nu_0}\right\|^{\frac{1}{2}}_{\infty}. \tag{3.55}$$

证明. 如上所述， H 是一个循环广义 Barron 泛函，且

$$\mu(dw,d\mathbf{a},d\mathbf{c},db) = \delta_0(d\mathbf{a})\mu_0(dw,d\mathbf{c},db)$$

¯，并且我们可以将映射 A、 C 以及序列 B 选为 0。作此选择后，定理 3.7 的证明中出现的映射 $\mathcal{L}$ 和序列 $\mathbf{B}_0$ 均等于 0。因此，由 (3.39) 可得，定理 3.7 的证明中的测度 $\tilde{\mu}^N$ 满足

$$\int_{\mathcal{X}^N} f(w,\mathbf{a},\mathbf{c},b)\tilde{\mu}^N(dw,d\mathbf{a},d\mathbf{c},db) = \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} f(w,0,\mathbf{c},b)\mu_0(dw,d\mathbf{c},db)$$

对于任意 f，只要最后一行中的被积函数关于 μ_0可积即可。这表明，(3.36) 中的 $\bar{H}^N$与映射

$$\bar{H}^N(\mathbf{z}) = \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} w\sigma_2(\mathbf{c}\cdot\mathbf{z}_0+b)\mu_0(dw,d\mathbf{c},db) = H(\mathbf{z}_0), \quad \mathbf{z}\in(D_d)^{\mathbb{Z}_-} \tag{3.56}$$

并且 $\bar{H}^N=H^N$。因此，也将泛函 $\mathbf{z}\mapsto H(\mathbf{z}_0)$ 记为 H，在误差估计 (3.42) 的第一步中，只有最后一项非零。

此外，$\psi_{A,B,C}(w,\mathbf{a},\mathbf{c}\ b) = (w,0,\mathbf{c}\ b)$，因此 $\mu_{A,B,C}(dw,d\mathbf{a}\ d\mathbf{c}\ db) = \delta_0(d\mathbf{a})\mu_0(dw,d\mathbf{c}\ db)$。因此，假设 $\mu_0\ll\nu_0$ 意味着 $\mu_{A,B,C}\ll\delta_0(d\mathbf{a})\nu_0$。

Choosing $W_i = \frac{w^{(i)}}{N}\frac{d\mu_0}{d\nu_0}(w^{(i)}, \mathbf{c}^{(i)}, b_i)$, from (3.40) we thus obtain

$$\mathbb{E}[|H(\mathbf{u}) - \widehat{H}(\mathbf{u})|^2] \leq N\mathbb{E}[W_1^2\sigma_2(\mathbf{c}^{(1)} \cdot \mathbf{u} + b_1)^2]. \tag{3.57}$$

式 (3.57) 中的期望可估计为

$$\begin{aligned}\mathbb{E}[W_1^2\sigma_2(\mathbf{c}^{(1)} \cdot \mathbf{u} + b_1)^2] &= \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} \frac{w^2}{N^2}\frac{d\mu_0}{d\nu_0}(w, \mathbf{c}, b)\sigma_2(\mathbf{c}\cdot\mathbf{u}+b)^2\mu_0(dw, d\mathbf{c}, db)\\ &\leq 2\left\|\frac{d\mu_0}{d\nu_0}\right\|_\infty \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} \frac{w^2}{N^2}[L_{\sigma_2}^2|\mathbf{c}\cdot\mathbf{u}+b|^2 + |\sigma_2(0)|^2]\mu_0(dw, d\mathbf{c}, db)\\ &\leq \frac{c_1}{N^2}\left\|\frac{d\mu_0}{d\nu_0}\right\|_\infty \int_{\mathbb{R}\times\mathbb{R}^d\times\mathbb{R}} w^2[\|\mathbf{c}\|^2 + |b|^2 + 1]\mu_0(dw, d\mathbf{c}, db)\end{aligned}$$

满足 $c_1 = 2\max(2L_{\sigma_2}^2, |\sigma_2(0)|^2)\max(1, \sup_{\mathbf{v}\in D_d}\|\mathbf{v}\|^2)$。 □

4 从单条轨迹中学习

现在考虑这样一种情形：我们旨在根据一条输入/输出对轨迹来学习泛函 H。我们观测到 $n \in \mathbb{N}$ 个数据点 $(\mathbf{z}_t, \mathbf{y}_t)$，其中 $t \in \{0, -1, \ldots, -(n-1)\}$，并旨在据此学习输入/输出关系 H。与静态情形不同，这些数据点并非独立同分布；相反，它们构成一个离散时间随机过程的单次实现。例如，Ziem 22] 曾研究过类似问题。

更形式化地说，我们考虑一个平稳随机过程 $(\mathbf{Z}_t, \mathbf{Y}_t)_{t\in\mathbb{Z}_-}$，并假设可以获得来自该过程单次实现的 n 个连续观测值 $(\mathbf{Z}_t, \mathbf{Y}_t)$，$t \in \{0, -1, \ldots, -(n-1)\}$。假设 $\mathbf{Z}$ 取值于 $\mathcal{I}_d$。令 $H \in \mathcal{C}$ 为未知泛函，并假设数据之间的输入/输出关系表示为 $H(\mathbf{Z}) = \mathbb{E}[\mathbf{Y}_0|\mathbf{Z}]$。例如，若 $\mathbf{Z}$ 是任意平稳过程，且 $\mathbf{Y}_t = H(\ldots, \mathbf{Z}_{t-2}, \mathbf{Z}_{t-1}, \mathbf{Z}_t) + \boldsymbol{\varepsilon}_t$，其中 $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{Z}_-}$ 是与 $\mathbf{Z}$ 独立的平稳过程并满足 $\mathbb{E}[\boldsymbol{\varepsilon}_0] = 0$，则该条件成立。

目标是根据底层输入/输出过程的观测值 $(\mathbf{Z}_t, \mathbf{Y}_t)_{t=0,\,-1,\ldots,\,-n+1}$ 学习泛函 H。回顾一下，$H(\mathbf{Z}) = \mathbb{E}[\mathbf{Y}_0|\mathbf{Z}]$ 在所有可测映射 $G\colon \mathcal{I}_d \to \mathbb{R}^m$ 中，使 $\mathcal{R}(G) = \mathbb{E}[\|G(\mathbf{Z}) - \mathbf{Y}_0\|^2]$ 达到最小。为了从数据中学习 H，因此我们旨在找到一个最小化者，使

$$\mathcal{R}_n(G) = \frac{1}{n}\sum_{i=0}^{n-1}\|G(\mathbf{Z}_{-i}^{-n+1}) - \mathbf{Y}_{-i}\|^2, \tag{4.1}$$

这里我们记 $\mathbf{Z}_{-i}^{-n+1} = (\ldots, \mathbf{0}, \mathbf{0}, \mathbf{Z}_{-n+1}, \ldots, \mathbf{Z}_{-i-1}, \mathbf{Z}_{-i})$。

为了从数据中学习 H，我们使用 (3.17) 中引入的 H 类逼近函数，并写作 $H_W = \widehat{H}$，以强调 W 是可训练参数。注意，现在 H_W 取值于 $\mathbb{R}^m$，其构造方式很简单：使用一个读出层 $W \in \mathbb{M}_{m,\,N}$ 将 m 个读出向量汇集为一个矩阵；或者等价地，使 (3.17) 中的 $w^{(i)}$ 取值于 $\mathbb{R}^m$。第 3 节的结果，尤其是第 3.5 节的普适性结论，表明 H 可以通过这类系统得到良好的逼近。因此，我们预期，在这类系统上对 $\mathcal{R}_n(\cdot)$ 进行最小化所得的解应能很好地逼近 H。因此，我们着手求解以下最小化问题

$$\widehat{W} = \underset{W\in\mathcal{W}_R}{\arg\min}\frac{1}{n}\sum_{i=0}^{n-1}\|\widehat{H}_W(\mathbf{Z}_{-i}^{-n+1}) - \mathbf{Y}_{-i}\|^2 \tag{4.2}$$

对于 $\mathcal{W}_R$，其定义为满足以下条件的所有随机矩阵 $W\colon \Omega \to \mathbb{R}^{m\times N}$ 对每一行 $W_{i,\,\cdot}$ 均有界 $\|W_{i,\,\cdot}\| \leq R$ 其中正则化常数为 $R > 0$；并且这些矩阵相对于数据 $(\mathbf{Z}_t, \mathbf{Y}_t)_{t=0,\,-1,\,\ldots,\,-n+1}$ 以及随机生成的参数 $(w_i, \mathbf{a}^{(i)}, \mathbf{c}^{(i)}, b_i)_{i=1,\,\ldots,\,N}$, A^{ESN}, $\mathbf{B}^{\mathrm{ESN}}$, C^{ESN}可测。后一项要求考虑到随机生成的参数

在初始化后即固定，因此可训练权重可能依赖于这些参数。 R 的选择将在定理 4.2 的陈述后文中说明。

问题 (4.2) 有显式解（例如，参见 [Gono 21a，第 4.3] 节）。一旦$\widehat{}$计算出来，所学习的泛函即为 $H_{\widehat{W}}$。现在可以通过评估算法的学习误差（或泛化误差）来衡量其学习性能

$$\mathbb{E}[\|H(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{W}}(\bar{\mathbf{Z}})\|^2],$$

$\bar{}$ $\bar{}$其中 Z 是 Z 的一个独立同分布副本，且 Z 与迄今引入的所有其他随机变量相互独立。下文的定理 4.2 将分析学习误差；受 [Gono 20b]，中相关研究的启发，我们假设输入过程和输出过程均属于以下定义中引入的类型。

定义 4.1。若存在 $q \in \mathbb{N}$, $G\colon (\mathbb{R}^q)^{\mathbb{Z}_-} \to \mathbb{R}^k$ 可测，且存在一个由 $\mathbb{R}^q$ 值随机变量构成的独立同分布族 $(\boldsymbol{\xi}_t)_{t\in\mathbb{Z}_-}$，使得一个 $\mathbb{R}^k$ 值随机过程 U 满足以下条件，则称其具有因果伯努利移位结构：

$$\mathbf{U}_t = G(\ldots, \boldsymbol{\xi}_{t-1}, \boldsymbol{\xi}_t), \quad t \in \mathbb{Z}_-.$$

若存在 $C_{\mathrm{dep}} > 0$, $\lambda_{\mathrm{dep}} \in (0,1)$，使得˜弱依赖系数 $\theta(\tau) := \mathbb{E}[\|\mathbf{U}_0 - \mathbf{U}_0^\tau\|]$ 满足 $\theta(\tau) \le C_{\mathrm{dep}}\lambda_{\mathrm{dep}}^\tau$，则称过程 U 具有几何衰减；该条件对所有 $\tau \in \mathbb{N}$均成立，其中

˜ ˜ ˜ ˜$\mathbf{U}_0^\tau = G(\ldots, \boldsymbol{\xi}_{-\tau-1}, \boldsymbol{\xi}_{-\tau}, \boldsymbol{\xi}_{-\tau+1}, \ldots, \boldsymbol{\xi}_0, \quad \boldsymbol{\xi}$是 $\boldsymbol{\xi}$的独立副本。

定理 4.2. 假设 $\sigma_1(x) = x$, $m = 1$, $p \in (1,\infty)$，且输入集 D_d 有界。考虑一个泛函 $H \in \mathcal{C}$，它具有 (2.2) 所示的表示形式，其中 $\mathbf{B} \in \ell^q$，有界线性映射$C\colon \mathbb{R}^d \to \ell^q$, $A\colon \ell^q \to \ell^q$满足 $|||A||| < 1$和 $I_\mu^{(2)},_p < \infty$。令 $\lambda \in (|||A|||, 1)$。现考虑一个 ESN-ELM 系统作为近似器，该系统满足 $\left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right| < 1$，且 (3.18) 中的矩阵 K 可逆。假设 $w^{(1)}$ 有界，$\mu_{A,B,C} \ll \nu$，且 $^{d\mu_A}$，$_{B\ C}$ 被常数 $\kappa > 0$所界。令 $R \ge \frac{\kappa}{\sqrt{N}}\|w^{(1)}\|_\infty$, $r \in (\left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|, 1)$。假设 Y被常数 $\widetilde{M}$所界，并且 $\overline{(Y, \mathrm{Z})}^{d\nu}$ 具有带几何衰减的因果伯努利移位结构，且 $\log(n) < n\log(\lambda_{max}^{-1})$，其中 $\lambda_{max} = \max(r, \lambda_{\mathrm{dep}})$。$\widehat{}$那么，训练后的 ESN-ELM $H_{\widehat{\mathbf{W}}}$ 满足学习误差界

$$\mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{\mathbf{W}}}(\bar{\mathbf{Z}})|^2]^{1/2} \le C_{\mathrm{approx}}\left(\lambda^{\frac{N}{d}} + \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^T + \frac{1}{N^{\frac{1}{2}}}\right) + C_{\mathrm{est}}\left(RN^{\frac{1}{2}}\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)^{\frac{1}{2}} \tag{4.3}$$

with

$$C_{\mathrm{approx}} = \tilde{c}_1 \max(\kappa^{\frac{1}{2}}, 1)\max\left(\frac{|||C|||}{(1 - |||\lambda^{-1}A|||^p)^{1/p}}, \frac{\|\mathbf{B}\|_q}{1 - |||A|||}, 1\right)$$
$$\times \max\left(I_{\mu,p}, I_{\mu,p}^{(2)}\right) \cdot \left(1 + \frac{\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right| + \|\mathbf{B}^{\mathrm{ESN}}\|}{1 - |||A^{\mathrm{ESN}}|||}\left|\left|\left|K^{-\top}\Lambda\right|\right|\right|\right), \tag{4.4}$$

$$C_{\mathrm{est}} = \tilde{c}_2\left[\frac{(\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right| + \|\mathbf{B}^{\mathrm{ESN}}\| + 1)^3}{(1 - |||A^{\mathrm{ESN}}|||)^3\sqrt{\log(\lambda_{max}^{-1})}}\max(R,1)(\mathbb{E}[\|\mathbf{a}^{(1)}\|^2]^{\frac{1}{2}} + \mathbb{E}[\|\mathbf{c}^{(1)}\|^2]^{\frac{1}{2}} + \mathbb{E}[|b_1|^2]^{\frac{1}{2}} + 1)\right.$$
$$\left.\times (r^2 - \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^2)^{-1/2}\max([\kappa\|w^{(1)}\|_\infty]^{-1}, 1)\left(\frac{r}{1-r} + \frac{1}{\lambda_{max}} + \frac{(\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right|^2 + 1)^{\frac{1}{2}}\lambda_{max}^{-1}}{\log(\lambda_{max}^{-1})}\right)\right]^{\frac{1}{2}}, \tag{4.5}$$

$\widetilde{}$对于 $\tilde{c}_1$，其仅取决于 d、σ_2、p、$\mathrm{diam}(D_d)$ 和 λ （见（3.43））；对于 $\tilde{c}_2$，其仅取决于 σ_2、M、$\mathrm{diam}(D_d)$ 和 C_{dep} （见（4.19））。

证明。令 $(\mathbf{Z}\ Y)$ 为 $(\mathbf{Z}\ Y)$ 的一个独立同分布副本，且独立于此前引入的所有其他随机变量。由独立性及假设 $H(\mathbf{Z}) = \mathbb{E}[Y_0|\mathbf{Z}]$ 可知，$\mathbb{E}[H_{\widehat{\mathbf{W}}}(\mathbf{Z})H(\mathbf{Z})] = \mathbb{E}[H_{\widehat{\mathbf{W}}}(\mathbf{Z})Y_0]$，且 $\mathbb{E}[H_{\mathbf{W}}(\mathbf{Z})H(\mathbf{Z})] = \mathbb{E}[H_{\mathbf{W}}(\mathbf{Z})Y_0]$，对任意 $\mathbf{W} \in \mathcal{W}_R$ 均成立。因此，对于任意 $\mathbf{W} \in \mathcal{W}_R$，我们得到

$$\begin{aligned}\mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{\mathbf{W}}}(\bar{\mathbf{Z}})|^2] &= \mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\mathbf{W}}(\bar{\mathbf{Z}})|^2] + \mathbb{E}[|\bar{Y}_0 - \widehat{H}_{\widehat{\mathbf{W}}}(\bar{\mathbf{Z}})|^2] - \mathbb{E}[|\bar{Y}_0 - \widehat{H}_{\mathbf{W}}(\bar{\mathbf{Z}})|^2] \\ &\leq \mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\mathbf{W}}(\bar{\mathbf{Z}})|^2] + \mathbb{E}[\mathcal{R}(\widehat{H}_{\widehat{\mathbf{W}}}) - \mathcal{R}_n(\widehat{H}_{\widehat{\mathbf{W}}}) + \mathcal{R}_n(\widehat{H}_{\mathbf{W}}) - \mathcal{R}(\widehat{H}_{\mathbf{W}})],\end{aligned} \tag{4.6}$$

其中，最后一步用到了 (4.2) 蕴含 $\mathcal{R}_n(H_{\widehat{\mathbf{W}}}) \leq \mathcal{R}_n(H_{\mathbf{W}})$。

右侧 (4.6) 中的第一项可利用定理 3.7 进行估计。事实上，定理 3.7 证明存在一个可测函数 $f\colon (\mathcal{X}^N)^N \to \mathbb{R}^N$，使得具有读出层 $\mathbf{W} = f((w^{(i)}, \mathbf{a}^{(i)}, \mathbf{c}^{(i)}, b_i)_{i=1,\dots,N})$ 的 ESN-ELM $\widehat{H}$ 满足

$$\mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\mathbf{W}}(\bar{\mathbf{Z}})|^2]^{1/2} \leq C_{H,\mathrm{ESN}} \left[\lambda^{\frac{N}{d}} + \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^T + \frac{1}{N^{\frac{1}{2}}} \left\| \frac{d\mu_{A,B,C}}{d\nu} \right\|_\infty^{\frac{1}{2}}\right] \tag{4.7}$$

其中 $C_{H,\mathrm{ESN}}$ 由 (3.28) 给出。这里用到了独立性以及 $\mathbf{Z}$ 取值于 $\mathcal{I}_d$。此外，在定理 3.7 f 的证明中，$\mathbf{W}$ 被明确选为 $W_i = \frac{w^{(i)}}{N} d\mu_{A,B,C}(w^{(i)}, \mathbf{a}^{(i)}, \mathbf{c}^{(i)}, b_i)$。因此，$\mathbb{P}$-几乎必然地，随机向量 $\mathbf{W}$ 满足 $\|\mathbf{W}\| \leq \sqrt{N} \max_{i=1}^N \|W_i\|_\infty \leq \frac{\kappa}{\sqrt{N}} \|w^{(1)}\|_\infty \leq R$，从而 $\mathbf{W} \in \mathcal{W}_R$。

接下来需要分析 (4.6) 右端的第二项。为简化记号，现在我们将 $(w_i, \mathbf{a}^{(i)}, \mathbf{c}^{(i)}, b_i)_{i=1,\dots,N}$, A^{ESN}, $\mathbf{B}^{\mathrm{ESN}}$, C^{ESN} 视为固定值；严格来说，这可以通过在给定这些随机变量的条件下进行后续计算来证明，例如参见[Gono 21a, 定理 4.3]的证明。于是

$$\mathbb{E}[\mathcal{R}(\widehat{H}_{\widehat{\mathbf{W}}}) - \mathcal{R}_n(\widehat{H}_{\widehat{\mathbf{W}}}) + \mathcal{R}_n(\widehat{H}_{\mathbf{W}}) - \mathcal{R}(\widehat{H}_{\mathbf{W}})] \leq 2\mathbb{E}\left[\sup_{\mathbf{w}\in\mathbb{R}^N : \|\mathbf{w}\| \leq R} |\mathcal{R}_n(\widehat{H}_{\mathbf{w}}) - \mathcal{R}(\widehat{H}_{\mathbf{w}})|\right]. \tag{4.8}$$

令 $\bar{\lambda} = (r^2 - \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^2)^{1/2}$ 表示 $\mathcal{C}^{RC} = \{\widehat{H}_{\mathbf{w}} : \mathbf{w} \in \mathbb{R}^N, \|\mathbf{w}\| \leq R\}$，并注意到每个 $\widehat{H}_{\mathbf{w}}$ 都可以写成 $\widehat{H}_{\mathbf{w}} = h_{\mathbf{w}} \circ H^F$，其中 $H^F(\mathbf{z}) = (\mathbf{x}_0^{\mathrm{ESN}}(\mathbf{z}), \mathbf{z}0, \bar{\lambda}\mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{z}))$，$h_{\mathbf{w}}((\mathbf{x}_0, \mathbf{z}_0, \mathbf{x}_{-1})) = \sum_{i=1}^N w_i \sigma_2(\bar{\lambda}^{-1}\mathbf{a}^{(i)} \cdot \mathbf{x}_{-1} + \mathbf{c}^{(i)} \cdot \mathbf{z}_0 + b_i)$。此外，$H^F\colon (D_d)^{\mathbb{Z}_-} \to \mathbb{R}^{N+d+N}$ 是与映射 $F\colon \mathbb{R}^{N+d+N} \times \mathbb{R}^d \to \mathbb{R}^{N+d+N}$, $F((\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2), \mathbf{z}) = (\sigma_1(A^{\mathrm{ESN}}\mathbf{x}_0 + C^{\mathrm{ESN}}\mathbf{z} + \mathbf{B}^{\mathrm{ESN}}), \mathbf{z}, \bar{\lambda}\mathbf{x}_0)$ 相关联的储备池泛函；该映射正是与由 A^{ESN}, $\mathbf{B}^{\mathrm{ESN}}$, C^{ESN} 确定的回声状态网络相关联的储库映射，并加入了当前输入和（经过缩放的）前一状态。

那么，对于每个 $\mathbf{w} \in \mathbb{R}^N$ 满足 $\|\mathbf{w}\| \leq R$，我们有

$$\begin{aligned}&|h_{\mathbf{w}}((\mathbf{x}_0, \mathbf{z}_0, \mathbf{x}_{-1})) - h_{\mathbf{w}}((\bar{\mathbf{x}}_0, \bar{\mathbf{z}}_0, \bar{\mathbf{x}}_{-1}))| \leq \sum_{i=1}^N |w_i| L_{\sigma_2} |\bar{\lambda}^{-1}\mathbf{a}^{(i)} \cdot (\mathbf{x}_{-1} - \bar{\mathbf{x}}_{-1}) + \mathbf{c}^{(i)} \cdot (\mathbf{z}_0 - \bar{\mathbf{z}}_0)| \\ &\leq L_{\sigma_2} R \max((\sum_{i=1}^N \|\bar{\lambda}^{-1}\mathbf{a}^{(i)}\|^2)^{1/2}, (\sum_{i=1}^N \|\mathbf{c}^{(i)}\|^2)^{1/2})[\|\mathbf{x}_{-1} - \bar{\mathbf{x}}_{-1}\| + \|\mathbf{z}_0 - \bar{\mathbf{z}}_0\|] \leq \overline{L_h} \|(\mathbf{x}_0, \mathbf{z}_0, \mathbf{x}_{-1}) - (\bar{\mathbf{x}}_0, \bar{\mathbf{z}}_0, \bar{\mathbf{x}}_{-1})\|,\end{aligned}$$

其中 $\overline{L_h} = 2L_{\sigma_2} R \max((\sum_{i=1}^N \|\bar{\lambda}^{-1}\mathbf{a}^{(i)}\|^2)^{1/2}, (\sum_{i=1}^N \|\mathbf{c}^{(i)}\|^2)^{1/2})$，且 $|h_{\mathbf{w}}(\mathbf{0})| \leq L_{h,\,0}$ 满足 $L_{h,\,0} = R(\sum_{i=1}^N \sigma_2(b_i)^2)^{1/2}$。此外，对于任意 $\mathbf{z}$, $\in D_d$ $F \cdot$ r $r = \left|\left|\left|A^{\mathrm{ESN}}\right|\right|\right|^2 + \bar{\lambda}^{2\ 1/2} <$ 储库映射 $(\mathbf{,}\ \mathbf{z})$ 是一个 -压缩映射，满足 1 $_{0,\ 1,\ 2} \in \mathbb{R}^N \times D_d \times \mathbb{R}^N$ $^- \in \mathbb{R}^d$ $(\mathbf{)}$；对于任意 $(\mathbf{x}\ \mathbf{x}\ \mathbf{x})$ 和 $\mathbf{z}\ \mathbf{z}$，我们有

$$\|F((\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2), \mathbf{z}) - F((\mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2), \bar{\mathbf{z}})\| = \|(C^{\mathrm{ESN}}(\mathbf{z} - \bar{\mathbf{z}}), \mathbf{z} - \bar{\mathbf{z}}, 0)\| = (\left|\left|\left|C^{\mathrm{ESN}}\right|\right|\right|^2 + 1)^{1/2} \|\mathbf{z} - \bar{\mathbf{z}}\|.$$

此外，对于 $\mathbf{z} \in (D_d)^{\mathbb{Z}_-}$，由 (3.23) 可得

$$\|H^F(\mathbf{z})\| = (\|\mathbf{x}_0^{\mathrm{ESN}}(\mathbf{z})\|^2 + \|\mathbf{z}_0\|^2 + \|\bar{\lambda}\mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{z})\|^2)^{1/2} \leq M_{\mathcal{F}},$$

其中 $M_{\mathcal{F}} = (2[(1 - \|\|A^{\mathrm{ESN}}\|\|)^{-1}(\|\|C^{\mathrm{ESN}}\|\|M + \|\mathrm{B}^{\mathrm{ESN}}\|)]^2 + M^2)^{1/2}$，其中 M 的选取满足：对所有 $\mathbf{z} \in D_d$，$\|\mathbf{z}\| \leq M$。最后，令 $C_L = RM_{\mathcal{F}} + M$，并注意到 $L(x, y) = \min(C_L, |x - y|)^2$ 对所有 $H \in \mathcal{C}^{RC}$、$\mathbf{z} \in (D_d)^{\mathbb{Z}_-}$ 以及满足 $|y| \leq \widetilde{M}$ 的 $y \in \mathbb{R}$，都有 $L(H(\mathbf{z}), y) = |H(\mathbf{z}) - y|^2$；且 $z \mapsto [\min(C_L, |z|)]^2$ 是常数为 $2C_L$ 的 Lipschitz 连续函数。这可通过如下写法看出：

$$\begin{aligned}|\min(C_L, |z_1|)^2 - \min(C_L, |z_2|)^2| &= |\min(C_L, |z_1|) - \min(C_L, |z_2|)||\min(C_L, |z_1|) + \min(C_L, |z_2|)| \\ &\leq 2C_L\big||z_2| - |z_1|\big| \leq 2C_L|z_2 - z_1|.\end{aligned}$$

考虑 [Gono 20b] 中定义的 Rademacher 型复杂度：令 $\varepsilon_0, \varepsilon_1, \ldots$ 为独立同分布的 Rademacher 随机变量，且与 $\mathbf{Z}$ 的独立同分布副本 $\tilde{\mathbf{Z}}^{(l)}$, $l \in \mathbb{N}$ 相互独立。对于每个 $k \in \mathbb{N}$，我们定义

$$\mathfrak{R}_k(\mathcal{C}^{RC}) = \frac{1}{k}\mathbb{E}\left[\sup_{H \in \mathcal{C}^{RC}}\left|\sum_{l=0}^{k-1}\varepsilon_l H(\tilde{\mathbf{Z}}^{(l)})\right|\right].$$

用 $\mathbf{X}^l$ 表示其分量为 $X_j^l = \sigma_2(\mathbf{a}^{(j)} \cdot \mathbf{x}_{-1}^{\mathrm{ESN}}(\tilde{\mathbf{Z}}^{(l)}) + \mathbf{c}^{(j)} \cdot \tilde{\mathbf{Z}}_0^{(l)} + b_j)$, $j = 1, \ldots, N$ 的向量。于是，我们可以重写 $\widehat{H}_{\mathbf{w}}(\tilde{\mathbf{Z}}^{(l)}) = \mathbf{w}^\top \mathbf{X}^l$，并利用柯西-施瓦茨不等式、詹森不等式、独立性以及 $\mathbb{E}[\varepsilon_l \varepsilon_j] = \delta_{lj}$ 这一事实进行估计

$$\begin{aligned}\mathfrak{R}_k(\mathcal{C}^{RC}) &= \frac{1}{k}\mathbb{E}\left[\sup_{\mathbf{w} \in \mathbb{R}^N : \|\mathbf{w}\| \leq R}\left|\sum_{l=0}^{k-1}\varepsilon_l \widehat{H}_{\mathbf{w}}(\tilde{\mathbf{Z}}^{(l)})\right|\right] = \frac{1}{k}\mathbb{E}\left[\sup_{\mathbf{w} \in \mathbb{R}^N : \|\mathbf{w}\| \leq R}\left|\mathbf{w}^\top \sum_{l=0}^{k-1}\varepsilon_l \mathbf{X}^l\right|\right] \\ &\leq \frac{R}{k}\mathbb{E}\left[\left\|\sum_{l=0}^{k-1}\varepsilon_l \mathbf{X}^l\right\|\right] \leq \frac{R}{k}\mathbb{E}\left[\left\|\sum_{l=0}^{k-1}\varepsilon_l \mathbf{X}^l\right\|^2\right]^{1/2} = \frac{R}{k}\left(\mathbb{E}\left[\sum_{l=0}^{k-1}\sum_{j=0}^{k-1}\varepsilon_l \varepsilon_j (\mathbf{X}^l)^\top \mathbf{X}^j\right]\right)^{1/2} \\ &= \frac{R}{k}\left(\sum_{l=0}^{k-1}\mathbb{E}\left[(\mathbf{X}^l)^\top \mathbf{X}^l\right]\right)^{1/2} = \frac{R}{\sqrt{k}}\left(\mathbb{E}\left[(\mathbf{X}^1)^\top \mathbf{X}^1\right]\right)^{1/2}.\end{aligned}$$

综上，[Gono 20b, 推论 8(ii)] 的假设均得到满足。将 $\mathcal{R}_n^\infty(G) = \frac{1}{n}\sum_{i=0}^{n-1}|G(\mathbf{Z}_{-i}^{-\infty}) - Y_{-i}|^2$ 记为理想化经验风险，并应用这一结果可得

$$\mathbb{E}\left[\sup_{\mathbf{w} \in \mathbb{R}^N : \|\mathbf{w}\| \leq R}|\mathcal{R}_n^\infty(\widehat{H}_{\mathbf{w}}) - \mathcal{R}(\widehat{H}_{\mathbf{w}})|\right] \leq \frac{C_1}{n} + \frac{C_2 \log(n)}{n} + \frac{C_{3,abs}\sqrt{\log(n)}}{\sqrt{n}} \tag{4.9}$$

对于满足 $\log(n) < n\log(\lambda_{max}^{-1})$ 的所有 $n \in \mathbb{N}$，其中 $\lambda_{max} = \max(r, \lambda_{\mathrm{dep}})$，以及

$$C_1 = 2C_L\frac{2M_{\mathcal{F}}\overline{L_h} + C_{\mathrm{dep}}}{\lambda_{max}}, \qquad C_{3,abs} = \frac{8C_L}{\sqrt{\log(\lambda_{max}^{-1})}}\left[R\left(\mathbb{E}\left[(\mathbf{X}^1)^\top \mathbf{X}^1\right]\right)^{1/2} + \mathbb{E}\left[|Y_0|_2^2\right]^{1/2}\right], \tag{4.10}$$

$$C_2 = \frac{4C_L(M_{\mathcal{F}}\overline{L_h} + L_{h,0} + (\|\|C^{\mathrm{ESN}}\|\|^2 + 1)^{1/2}\overline{L_h}C_{\mathrm{dep}}\lambda_{max}^{-1}) + 2\mathbb{E}[|Y_0|^2]}{\log(\lambda_{max}^{-1})}. \tag{4.11}$$

此外，[Gono 20b，命题 5] 得到 $\mathbb{P}$-几乎处处成立。

$$\sup_{\mathbf{w} \in \mathbb{R}^N : \|\mathbf{w}\| \leq R}\left|\mathcal{R}_n(\widehat{H}_{\mathbf{w}}) - \mathcal{R}_n^\infty(\widehat{H}_{\mathbf{w}})\right| \leq \frac{4C_L r\overline{L_h}M_{\mathcal{F}}}{(1 - r)n}.$$

结合（4.9），我们得到

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\mathbf{w}\in\mathbb{R}^N\,:\,\|\mathbf{w}\|\leq R}|\mathcal{R}_n(\widehat{H}_{\mathbf{w}})-\mathcal{R}(\widehat{H}_{\mathbf{w}})|\right]\\
&\leq \mathbb{E}\left[\sup_{\mathbf{w}\in\mathbb{R}^N\,:\,\|\mathbf{w}\|\leq R}|\mathcal{R}_n(\widehat{H}_{\mathbf{w}})-\mathcal{R}_n^{\infty}(\widehat{H}_{\mathbf{w}})|\right]+\mathbb{E}\left[\sup_{\mathbf{w}\in\mathbb{R}^N\,:\,\|\mathbf{w}\|\leq R}|\mathcal{R}_n^{\infty}(\widehat{H}_{\mathbf{w}})-\mathcal{R}(\widehat{H}_{\mathbf{w}})|\right]\\
&\leq \frac{4C_L r\overline{L_h}M_{\mathcal{F}}}{(1-r)n}+\frac{C_1}{n}+\frac{C_2\log(n)}{n}+\frac{C_{3,abs}\sqrt{\log(n)}}{\sqrt{n}}\\
&\leq \max(R,1)\max(\overline{L_h},L_{h,0},1,R\left(\mathbb{E}\left[(\mathbf{X}^1)^\top\mathbf{X}^1\right]\right)^{1/2})C_{est,1}\frac{\sqrt{\log(n)}}{\sqrt{n}},
\end{aligned}\tag{4.12}$$

其中

$$\begin{aligned}
C_{est,1}=&\frac{32\max(1,C_{\mathrm{dep}})}{\sqrt{\log(\lambda_{max}^{-1})}}\max(2\frac{(|||C^{\mathrm{ESN}}|||M+\|\mathbf{B}^{\mathrm{ESN}}\|)^2}{(1-|||A^{\mathrm{ESN}}|||)^2}+M^2,1)\max(\widetilde{M},1)\\
&\times\left(\frac{r}{(1-r)}+\frac{1}{\lambda_{max}}+\frac{(2+(|||C^{\mathrm{ESN}}|||^2+1)^{1/2}\lambda_{max}^{-1})+\mathbb{E}[|Y_0|^2]}{\log(\lambda_{max}^{-1})}+\mathbb{E}\left[|Y_0|^2\right]^{1/2}\right).
\end{aligned}\tag{4.13}$$

回想一下，$(w^{(i)},\mathbf{a}^{(i)},\mathbf{c}^{(i)},b_i)_{i=1,\dots,N}$ 曾被视为固定；现在我们也对其分布 ν 取期望，并以与推导（4.12）相同的方式得到

$$\begin{aligned}
\mathbb{E}\left[\sup_{\mathbf{w}\in\mathbb{R}^N\,:\,\|\mathbf{w}\|\leq R}|\mathcal{R}_n(\widehat{H}_{\mathbf{w}})-\mathcal{R}(\widehat{H}_{\mathbf{w}})|\right]\leq\max(R,1)\max\Bigg(&2RL_{\sigma_2}\mathbb{E}\left[\max\left((\sum_{i=1}^N\|\bar{\lambda}^{-1}\mathbf{a}^{(i)}\|^2)^{\frac{1}{2}},(\sum_{i=1}^N\|\mathbf{c}^{(i)}\|^2)^{\frac{1}{2}}\right)\right],\\
&R\mathbb{E}\left[(\sum_{i=1}^N\sigma_2(b_i)^2)^{\frac{1}{2}}\right],1,R\left(\mathbb{E}\left[(\mathbf{X}^1)^\top\mathbf{X}^1\right]\right)^{\frac{1}{2}}\Bigg)C_{est,1}\frac{\sqrt{\log(n)}}{\sqrt{n}}.
\end{aligned}\tag{4.14}$$

（4.14）中的期望可以利用詹森不等式和（3.23）进行估计：

$$\begin{aligned}
\mathbb{E}\left[(\mathbf{X}^1)^\top\mathbf{X}^1\right]^{\frac{1}{2}}=\left(\sum_{j=1}^N\mathbb{E}[|X_j^1|^2]\right)^{\frac{1}{2}}&\leq\max(2L_{\sigma_2}^2,2|\sigma_2(0)|^2)^{\frac{1}{2}}N^{\frac{1}{2}}\mathbb{E}[|\mathbf{a}^{(1)}\cdot\mathbf{x}_{-1}^{\mathrm{ESN}}(\mathbf{Z})+\mathbf{c}^{(1)}\cdot\mathbf{Z}_0+b_1|^2+1]^{\frac{1}{2}}\\
&\leq N^{\frac{1}{2}}C_{est,2}
\end{aligned}$$

with

$$\begin{aligned}
C_{est,2}=&2^{\frac{3}{2}}\max(M,1)\max(L_{\sigma_2},|\sigma_2(0)|,1)\\
&\times\frac{|||C^{\mathrm{ESN}}|||+\|\mathbf{B}^{\mathrm{ESN}}\|+1}{1-|||A^{\mathrm{ESN}}|||}(\mathbb{E}[\|\bar{\lambda}^{-1}\mathbf{a}^{(1)}\|^2]^{\frac{1}{2}}+\mathbb{E}[\|\mathbf{c}^{(1)}\|^2]^{\frac{1}{2}}+\mathbb{E}[|b_1|^2]^{\frac{1}{2}}+1)
\end{aligned}\tag{4.15}$$

因此

$$\begin{aligned}
\max\Bigg(&2RL_{\sigma_2}\mathbb{E}\left[\max\left((\sum_{i=1}^N\|\bar{\lambda}^{-1}\mathbf{a}^{(i)}\|^2)^{\frac{1}{2}},(\sum_{i=1}^N\|\mathbf{c}^{(i)}\|^2)^{\frac{1}{2}}\right)\right],R\mathbb{E}\left[(\sum_{i=1}^N\sigma_2(b_i)^2)^{\frac{1}{2}}\right],1,R\left(\mathbb{E}\left[(\mathbf{X}^1)^\top\mathbf{X}^1\right]\right)^{\frac{1}{2}}\Bigg)\\
&\leq\max\left(RN^{\frac{1}{2}}C_{est,2},1\right).
\end{aligned}\tag{4.16}$$

将此代入 (4.14) 得到

$$\mathbb{E}\left[\sup_{\mathbf{w}\in\mathbb{R}^N\,:\,\|\mathbf{w}\|\leq R}|\mathcal{R}_n(\widehat{H}_{\mathbf{w}})-\mathcal{R}(\widehat{H}_{\mathbf{w}})|\right]\leq\max(R,1)\max\left(RN^{\frac{1}{2}}C_{est,2},1\right)C_{est,1}\frac{\sqrt{\log(n)}}{\sqrt{n}}.\tag{4.17}$$

将此与 (4.6)、(4.7) 和 (4.8) 结合，我们便得到

$$\begin{aligned}\mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{\mathbf{W}}}(\bar{\mathbf{Z}})|^2]^{1/2} &\le \mathbb{E}[|H(\bar{\mathbf{Z}}) - \widehat{H}_{\mathbf{W}}(\bar{\mathbf{Z}})|^2]^{1/2} + \mathbb{E}[\mathcal{R}(\widehat{H}_{\widehat{\mathbf{W}}}) - \mathcal{R}_n(\widehat{H}_{\widehat{\mathbf{W}}}) + \mathcal{R}_n(\widehat{H}_{\mathbf{W}}) - \mathcal{R}(\widehat{H}_{\mathbf{W}})]^{1/2} \\ &\le C_{H,\mathrm{ESN}}[\lambda^{\frac{N}{d}} + |||A^{\mathrm{ESN}}|||^T + \frac{1}{N^{\frac{1}{2}}}\left\|\frac{d\mu_{A,B,C}}{d\nu}\right\|_\infty^{\frac{1}{2}}] \\ &\quad + \sqrt{2}\left(\max(R,1)\max\left(RN^{\frac{1}{2}}C_{est,2},1\right)C_{est,1}\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)^{\frac{1}{2}} \\ &\le C_{\mathrm{approx}}[\lambda^{\frac{N}{d}} + |||A^{\mathrm{ESN}}|||^T + \frac{1}{N^{\frac{1}{2}}}] + \tilde{C}_{\mathrm{est}}\left(RN^{\frac{1}{2}}\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)^{\frac{1}{2}},\end{aligned} \tag{4.18}$$

其中，我们使用 $\max(RN^{\frac{1}{2}},1) \le RN^{\frac{1}{2}}\max([\kappa\|w^{(1)}\|_\infty]^{-1},1)$，并令 $C_{\mathrm{approx}} = C_{H,\mathrm{ESN}}\max(\kappa^{\frac{1}{2}},1)$，$\tilde{C}_{\mathrm{est}} = (2C_{est,1}\max(R,1)C_{est,2}\max([\kappa\|w^{(1)}\|_\infty]^{-1},1))^{\frac{1}{2}}$。令

$$\begin{aligned}C_{\mathrm{est}} = \tilde{c}_2\Bigg[&\frac{(|||C^{\mathrm{ESN}}||| + \|\mathbf{B}^{\mathrm{ESN}}\| + 1)^3}{(1-|||A^{\mathrm{ESN}}|||)^3\sqrt{\log(\lambda_{max}^{-1})}}\max(R,1)(\mathbb{E}[\|\mathbf{a}^{(1)}\|^2]^{\frac{1}{2}} + \mathbb{E}[\|\mathbf{c}^{(1)}\|^2]^{\frac{1}{2}} + \mathbb{E}[|b_1|^2]^{\frac{1}{2}} + 1) \\ &\times (r^2 - |||A^{\mathrm{ESN}}|||^2)^{-1/2}\max([\kappa\|w^{(1)}\|_\infty]^{-1},1)\left(\frac{r}{1-r} + \frac{1}{\lambda_{max}} + \frac{(|||C^{\mathrm{ESN}}|||^2+1)^{\frac{1}{2}}\lambda_{max}^{-1}}{\log(\lambda_{max}^{-1})}\right)\Bigg]^{\frac{1}{2}}\end{aligned}$$

with

$$\tilde{c}_2 = 2^6(\max(1,C_{\mathrm{dep}})\max(M,1)^5\max(\widetilde{M},1)^3\max(L_{\sigma_2},|\sigma_2(0)|,1))^{\frac{1}{2}} \tag{4.19}$$

并代入 (3.28)，即可得到 (4.3)，其中 (4.4) 和 (4.5) 如前所述。 □

注记 4.3. 定理 4.2 中的界可直接推广到多变量情形 $m > 1$，如下所示：若 $H = (H_1,\dots,H_m)$ 且每个 H_i 均满足定理 4.2 中提出的假设，则

$$\begin{aligned}\mathbb{E}[\|H(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{W}}(\bar{\mathbf{Z}})\|^2]^{1/2} &= \mathbb{E}\left[\sum_{i=1}^m |H_i(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{W}_i}(\bar{\mathbf{Z}})|^2\right]^{1/2} \le \sum_{i=1}^m \mathbb{E}\left[|H_i(\bar{\mathbf{Z}}) - \widehat{H}_{\widehat{W}_i}(\bar{\mathbf{Z}})|^2\right]^{1/2} \\ &\le \sum_{i=1}^m C_{\mathrm{approx},i}\left(\lambda_i^{\frac{N}{d}} + |||A^{\mathrm{ESN}}|||^T + \frac{1}{N^{\frac{1}{2}}}\right) + \sum_{i=1}^m C_{\mathrm{est},i}\left(RN^{\frac{1}{2}}\frac{\sqrt{\log(n)}}{\sqrt{n}}\right)^{\frac{1}{2}}\end{aligned} \tag{4.20}$$

其中 $C_{\mathrm{approx},\,i}$、$C_{\mathrm{est},\,i}$ 是定理 4.2 中与 H_i 对应的常数。

5 结论

本文给出了一类新的输入/输出系统概念类的储备池计算逼近界和泛化界。这类概念类将所谓的广义 Barron 泛函扩展到了动力学背景中。我们将这一新类中的元素称为循环广义 Barron 泛函。其元素的特征是存在基于无限维状态空间系统构造、具有特定积分表示的读出层。我们还证明，这些元素构成一个向量空间，且该类包含具有有限维状态空间表示的此类系统作为子集。事实证明，这一类非常丰富：它包含所有足够光滑的有限记忆泛函、线性滤波器和卷积滤波器，以及任何基于可分希尔伯特状态空间并具有积分 Barron 类型读出的输入/输出系统。此外，该类在 L^2-意义下稠密。

用于逼近/估计新类别中元素的储备池架构，是随机生成的回声状态网络，采用线性（用于逼近）或 ReLU（用于估计）激活函数，并使用随机生成的神经网络构建读出层，其中仅

训练输出层（极限学习机或随机特征神经网络）。论文中的结果得出了一种可完全实现的、基于循环神经网络的学习算法，该算法具有可证明的收敛保证，且不受维数灾难的影响。

致谢。作者感谢瑞士国家科学基金会提供的部分经费支持（资助编号 200021 175801/1）。L. Grigoryeva 和 J.-P. Ortega 感谢慕尼黑大学在他们学术访问期间给予的款待；L. Gonon 则感谢南洋理工大学在其学术访问期间给予的款待。他们在这些访问期间开展了本研究的部分工作。我们感谢 Christa Cuchiero 和 Josef Teichmann 与我们进行富有启发性的交流，这些交流促成了本文的部分研究成果。

参考文献

[Acci 22] B. Acciaio、A. Kratsios 和 G. Pammer。“度量超变换器是通用适应映射”。arXiv 预印本 arXiv:2201.13094，2022年。[Arco 22] T. Arcomano、I. Szunyogh、A. Wikner、J. Pathak、B. R. Hunt 和 E. Ott。“一种结合机器学习与基于物理的数值模型的大气建模混合方法”。《地球系统建模进展杂志》，第14卷第3期，页码 e2021MS002712，2022年。[Barr 92] A. R. Barron。“神经网络逼近”。载于：耶鲁大学第七届自适应与学习系统研讨会论文集，第69–72页，1992年。[Barr 93] A. Barron。“S形函数叠加的通用逼近界”。《IEEE 信息论汇刊》，第39卷第3期，第930–945页，1993年5月。[Bent 22] F. E. Benth、N. Detering 和 L. Galimberti。“Fr´echet 空间中的神经网络”。《数学与人工智能年刊》，第1–29页，2022年。[Bouc 13] S. Boucheron、G. Lugosi 和 P. Massart。《集中不等式：独立性的非渐近理论》。牛津大学出版社，2013年。[Bouv 17a] J. Bouvrie 和 B. Hamzi。“用于非线性系统逼近的核方法”。《SIAM 控制与优化杂志》，第55卷第4期，第2460–2492页，2017年。[Bouv 17b] J. Bouvrie 和 B. Hamzi。“用于逼近非线性系统若干关键量的核方法”。《计算动力学杂志》，第4卷第（1-2）期，第1–19页，2017年。[Boyd 85] S. Boyd 和 L. Chua。“衰减记忆与利用 Volterra 级数逼近非线性算子的问题”。《IEEE 电路与系统汇刊》，第32卷第11期，第1150–1161页，1985年。[Chen 19] J. Chen 和 H. I. Nurdin。“利用耗散量子系统学习非线性输入输出映射”。《量子信息处理》，第18卷第7期，第1–36页，2019年。[Chen 20] J. Chen、H. I. Nurdin 和 N. Yamamoto。“含噪量子计算机上的时间信息处理”。《Physical Review Applied》，第14卷第2期，页码 24065，2020年。[Chen 22] J. Chen。《用于新型量子和经典设备时间信息处理的非线性收敛动力学》。博士论文，新南威尔士大学悉尼分校，2022年。[Chen 95] T. Chen 和 H. Chen。“径向基函数神经网络对多变量函数、非线性泛函和算子的逼近能力”。《IEEE 神经网络汇刊》，第6卷第4期，第904–910页，1995年。[Coui 16] R. Couillet、G. Wainrib、H. Sevi 和 H. T. Ali。“线性回声状态神经网络的渐近性能”。《机器学习研究杂志》，第17卷第178期，第1–35页，2016年。

[Cuch 22] C. Cuchiero、F. Primavera 和 S. Svaluto-Ferro。“c`adl`ag 路径连续函数与 L´evy 型签名模型的通用逼近定理”。arXiv 预印本 arXiv:2208.02293，2022年。[E 19] W. E、C. Ma 和 L. Wu。“双层神经网络总体风险的先验估计”。《数学科学通讯》，第17卷第5期，第1407–1425页，2019年。[E 20a] W. E、C. Ma、S. Wojtowytsch 和 L. Wu。“走向对基于神经网络的机器学习的数学理解：我们已知与未知的内容”。预印本，arXiv 2009.10713，2020年。[E 20b] W. E 和 S. Wojtowytsch。“与多层 ReLU 网络相关的巴拿赫空间：函数表示、逼近理论与梯度下降动力学”。预印本，arXiv 2007.15623，2020年。[E 21] W. E 和 S. Wojtowytsch。“Barron 函数的表示公式与逐点性质”。arXiv:2006.05982，2021年。[Evan 10] L. C. Evans。《偏微分方程》。数学研究生教材第19卷，美国数学会，罗德岛州普罗维登斯，第2版，2010年。[G Ma 20] G Manjunath。“稳定性与记忆丧失相辅相成：动力学与计算中的三个结果”。即将发表于《伦敦皇家学会会刊 A 辑：数学、物理与工程科学》，第1–25页，2020年。[Gali 22] L. Galimberti、G. Livieri 和 A. Kratsios。“通用因果深度学习模型的设计：随机分析中无限维动力系统的情形”。arXiv 预印本 arXiv:2210.13300，2022年。[Gono 20a] L. Gonon、L. Grigoryeva 和 J.-P. Ortega。“非线性循环网络的记忆与预测能力”。《Physica D》，第414卷，编号132721，第1–13页，2020年。[Gono 20b] L. Gonon、L. Grigoryeva 和 J.-P. Ortega。“储备池计算的风险界”。《机器学习研究杂志》，第21卷第240期，第1–61页，2020年。[Gono 20c] L. Gonon 和 J.-P. Ortega。“具有随机输入的储备池计算的普适性”。《IEEE 神经网络与学习系统汇刊》，第31卷第1期，第100–112页，2020年。[Gono 21a] L. Gonon。“随机特征神经网络在没有维数灾难的情况下学习 Black-Scholes 型偏微分方程”。预印本，2021年。[Gono 21b] L. Gonon 和 J.-P. Ortega。“渐消记忆回声状态网络具有普适性”。《神经网络》，第138卷，第10–13页，2021年。[Gono 23] L. Gonon、L. Grigoryeva 和 J.-P. Ortega。“随机神经网络与储备池系统的逼近误差估计”。《应用概率年刊》，第33卷第1期，第28–69页，2023年。[Grig 14] L. Grigoryeva、J. Henriques、L. Larger 和 J.-P. Ortega。“ 使用时延储层计算机进行随机时间序列预测：性能与普适性 ”。《神经网络》，第 55 卷，第59–71页，2014年。[Grig 18a] L. Grigoryeva 和 J.-P. Ortega。“回声状态网络具有普适性”。《神经网络》，第108卷，第495–508页，2018年。[Grig 18b] L. Grigoryeva 和 J.-P. Ortega。“具有随机输入和线性读出、基于非齐次状态仿射系统的通用离散时间储层计算机”。《机器学习研究杂志》，第19卷第24期，第1–40页，2018年。

[Grig 19] L. Grigoryeva 和 J.-P. Ortega。“可微储备池计算”。Journal of Machine Learning Research，第20卷，第179期，第1–62页，2019年。[Grig 21a] L. Grigoryeva、A. G. Hart 和 J.-P. Ortega。“利用储备池系统学习奇异吸引子”。arXiv 预印本 arXiv:2108.05024，2021年。[Grig 21b] L. Grigoryeva 和 J.-P. Ortega。“通过规范化降低循环网络的维度”。Journal of Geometric Mechanics，第13卷，第4期，第647–677页，2021年。[Herm 10] M. Hermans 和 B. Schrauwen。“连续时间线性循环神经网络中的记忆”。Neural networks : the official journal of the International Neural Network Society，第23卷，第3期，第341–55页，2010年4月。[Herm 12] M. Hermans 和 B. Schrauwen。“循环核机器：利用无限回声状态网络进行计算”。Neural Computation，第24卷，第104–133页，2012年。[Hu 22] P. Hu、Q. Meng、B. Chen、S. Gong、Y. Wang、W. Chen、R. Zhu、Z.-M. Ma 和 T.-Y. Liu。“利用正则性结构建模随机偏微分方程驱动的动力学的神经算子”。arXiv 预印本 arXiv:2204.06255，2022年。[Huan 06] G.-B. Huang、Q.-Y. Zhu 和 C.-K. Siew。“ 极限学习机：理论与应用 ”。Neurocomputing，第 70 卷，第 1-3 期，第 489–501 页，2006年12月。[Jaeg 04] H. Jaeger 和 H. Haas。“驾驭非线性：预测混沌系统并在无线通信中节约能源”。《科学》，第304卷，第5667期，第78–80页，2004年。[Jaeg 10] H. Jaeger。“分析和训练循环神经网络的‘回声状态’方法（附勘误说明）”。技术报告，德国国家信息技术研究中心，2010年。[Kira 19] F. J. Kir´aly 和 H. Oberhauser。“序列有序数据的核”。Journal of Machine Learning Research，第20卷，2019年。[Kova 21] N. Kovachki、Z. Li、B. Liu、K. Azizzadenesheli、K. Bhattacharya、A. Stuart 和 A. Anand- kumar。“神经算子：学习函数空间之间的映射”。arXiv 预印本 arXiv:2108.08481，2021年。[Krat 20] A. Kratsios 和 I. Bilokopytov。“非欧几里得通用逼近”。第34届神经信息处理系统大会（NeurIPS 2020），加拿大温哥华，2020年。[Lax 02] P. Lax。《泛函分析》。Wiley-Interscience，2002年。[Li 20] Z. Li、N. Kovachki、K. Azizzadenesheli、B. Liu、K. Bhattacharya、A. Stuart 和 A. Anand- kumar。“用于参数偏微分方程的傅里叶神经算子”。arXiv 预印本 arXiv:2010.08895，2020年。[Li 22] Z. Li、J. Han、E. Weinan 和 Q. Li。“线性连续时间循环神经网络的逼近与优化理论”。Journal of Machine Learning Research，第23卷，第41–42页，2022年。[Lu 18] Z. Lu、B. R. Hunt 和 E. Ott。“通过机器学习重构吸引子”。Chaos，第28卷，第6期，2018年。[Maas 02] W. Maass、T. Natschl¨ager 和 H. Markram。“无需稳定状态的实时计算：一种基于扰动的神经计算新框架”。Neural Computation，第14卷，第2531–2560页，2002年。[Maas 11] W. Maass。“液态状态机：动机、理论与应用”。载于：S. S. Barry Cooper 和 A. Sorbi 编，《语境中的可计算性：现实世界中的计算与逻辑》，第8章，第275–296页，2011年。

[Manj 13] G. Manjunath 和 H. Jaeger。“与输入相关联的回声状态性质：探索循环神经网络的一项基本特征”。Neural Computation，第25卷，第3期，第671–696页，2013年。[Manj 22a] G. Manjunath。“将信息嵌入动力系统”。Nonlinearity，第35卷，第3期，第1131页，2022年。[Manj 22b] G. Manjunath 和 J.-P. Ortega。“储备池计算中的传输”。arXiv 预印本 arXiv:2209.07946，2022年。[Mart 23] R. Mart´ınez-Pe˜na 和 J.-P. Ortega。“有限维量子储备池计算”。Physical Review E - 统计物理、等离子体、流体及相关跨学科主题，第107卷，第3期，第035306页，2023年。[Matt 92] M. B. Matthews。《利用神经网络模型对非线性离散时间渐忘系统进行一致逼近》。博士论文，苏黎世联邦理工学院，1992年。¨Matt [M. B. Matthews。“利用神经网络模型逼近非线性渐忘算子”。Circuits, Systems, and Signal Processing，第12卷，第2期，第279–307页，1993年6月。[Neuf 22] A. Neufeld 和 P. Schmocker。“利用迭代积分和神经网络进行混沌对冲”。arXiv 预印本 arXiv:2209.10166，2022年。[Path 17] J. Pathak、Z. Lu、B. R. Hunt、M. Girvan 和 E. Ott。“利用机器学习复现混沌吸引子并根据数据计算李雅普诺夫指数”。Chaos，第27卷，第12期，2017年。[Path 18] J. Pathak、B. Hunt、M. Girvan、Z. Lu 和 E. Ott。“从数据中无模型预测大型时空混沌系统：一种储备池计算方法”。Physical Review Letters，第120卷，第2期，第24102页，2018年。[Rahi 07] A. Rahimi 和 B. Recht。“用于大规模核机器的随机特征”。Advances in neural information，2007年。[Salv 22] C. Salvi、M. Lemercier 和 A. Gerasimovics。“神经随机偏微分方程：连续时空动力学的分辨率不变学习”。载于：Advances in Neural Information Processing Systems，2022年。[Stin 99] M. B. Stinchcombe。“神经网络对紧化空间上的连续泛函和连续函数的逼近”。Neural Networks，第12卷，第3期，第467–477页，1999年。[Tino 18] P. Tino。“随机线性对称回声状态网络的渐近费希尔记忆”。Neurocomputing，第298卷，第4–8页，2018年。[Tino 20] P. Tino。“作为时间特征空间的动力系统”。Journal of Machine Learning Research，第21卷，第1–42页，2020年。[Tran 20] Q. H. Tran 和 K. Nakajima。“高阶量子储备池计算”。arXiv 预印本 arXiv:2006.08999，2020年。[Tran 21] Q. H. Tran 和 K. Nakajima。“时间量子层析学习”。Physical Review Letters，第127卷，第26期，第260401页，2021年。[Vill 09] C. Villani。《最优传输：旧与新》。Springer，2009年。[Wikn 21] A. Wikner、J. Pathak、B. R. Hunt、I. Szunyogh、M. Girvan 和 E. Ott。“利用数据同化训练结合机器学习与基于知识组件的混合预报系统”。Chaos: An Interdisciplinary Journal of Nonlinear Science，第31卷，第5期，第53114页，2021年。

[Yild 12] I. B. Yildiz、H. Jaeger 和 S. J. Kiebel。《重新审视回声状态性质》。《神经网络》，第 35 卷，第 1–9 页，2012年11月。[Ziem 22] I. M. Ziemann、H. Sandberg 和 N. Matni。《单轨迹非参数非线性动力学学习》。载于：P.-L. Loh 和 M. Raginsky 编，《第三十五届学习理论会议论文集》，第 3333–3364 页，PMLR，2022年。