

# リザーバー計算による力学系の時系列予測

京都大学大学院理学研究科数学・数理解析専攻 原 誠人

Hara Masato

Department of Mathematics, Graduate School of Science, Kyoto University

## 1 はじめに

リザーバー計算は回帰型ニューラルネットワークを用いた機械学習手法である。回帰型ニューラルネットワークはそれ自身が大自由度力学系であり、時系列データの扱いに適している。本稿ではリザーバー計算による時系列予測に関して、「ダイナミクスの再現」という観点に立った研究の結果を述べる。

なお、本稿の元になった研究集会での発表は福田悠登・國府寛司（いずれも京都大学）との共同研究に基づいて行われた。また、本稿は著者の修士論文 [3] の一部を研究集会での発表に即して構成したものに相当する。

### 記号

- $\mathbb{N}$  で正の整数全体の集合、 $\mathbb{N}_0$  で非負整数全体の集合を表す。
- 正方行列  $A$  に対し、 $A$  のスペクトル半径（＝固有値の絶対値の最大値）を  $\rho(A)$  で表す。

## 2 リザーバー計算の概要

リザーバー計算 (reservoir computing: RC) は、ニューラルネットワークの中でも特に時系列データの扱いに優れた回帰型ニューラルネットワーク (recurrent neural network: RNN) を用いた機械学習法の一つである。一般にニューラルネットワークは役割の異なる 3 種類の層、入力層・中間層 (隠れ層)・出力層から構成されるが、リザーバー計算の文脈では中間層を「リザーバー」と呼ぶ。

以下、主に [2] に基づいてリザーバー計算の一般的な枠組みを解説する。 $K \in \mathbb{N}_0; N, L \in \mathbb{N}$  に対し、入力  $\mathbf{u}(t) \in \mathbb{R}^K$ 、リザーバー  $\mathbf{r}(t) \in \mathbb{R}^N$ 、出力  $\boldsymbol{\eta}(t) \in \mathbb{R}^L$  をもつ RNN を考える。このとき、入力層からリザーバーへの結合を  $W^{\text{in}} \in M_{N,K}(\mathbb{R})$ 、リザーバー内の結合を  $W \in M_{N,N}(\mathbb{R})$ 、各層から出力層への結合を  $W^{\text{out}} \in M_{L,K+N+L}(\mathbb{R})$ 、出力層からリザーバーへの結合 (フィードバック) を  $W^{\text{fb}} \in M_{N,L}(\mathbb{R})$  によって表現する。

ここで、ニューラルネットワークを用いた（教師あり）機械学習は一般に、

Step1：既知の入出力に基づいて結合行列（ $W^{\text{in}}, W, \dots$ ）などのパラメータを適切に決定する

Step2：新たな入力に対し、Step1 で確定したネットワークを用いて出力を得る

という枠組みをもつ。Step1 のことを「学習」という。リザーバー計算の場合、学習によって決定するのは  $W^{\text{out}}$  だけで、 $W^{\text{in}}, W, W^{\text{fb}}$  はそれぞれあるクラスからランダムに取って固定しておく。従って  $W^{\text{in}}, W, W^{\text{fb}}$  が属するクラスの設定が問題になるが、Echo State Network (ESN) [2] では、 $W$  として RNN に “Echo State Property (ESP)” を付与するようなものを取る。ESP の正確な定義は抽象的なので、通常は簡便な代替条件である  $\rho(W) < 1$  を  $W$  の満たすべき条件と考える。

いま、「学習」を経て  $W^{\text{out}}$  が決定された状況を考える。入力は各時刻において与えられるが、リザーバー及び出力層の各ニューロン（層を構成する情報処理の要素）は、次の式に従って時間発展する：

$$\begin{aligned} r(t+1) &= \Phi(W^{\text{in}}u(t+1) + Wr(t) + W^{\text{fb}}\eta(t)) \\ \eta(t+1) &= \Psi(W^{\text{out}}[u(t+1), r(t+1), \eta(t)]). \end{aligned}$$

但し  $\Phi: \mathbb{R}^N \rightarrow \mathbb{R}^N, \Psi: \mathbb{R}^L \rightarrow \mathbb{R}^L$ ,  $[u(t+1), r(t+1), \eta(t)]$  は  $u(t+1), r(t+1), \eta(t)$  を縦に連結した  $(K + N + L)$ -次元ベクトルである。

以上を踏まえて、リザーバー計算の模式図を図 1 に示す。図中の丸印はニューロンを表し ( $K = 2, N = 10, L = 3$ )、実線矢印は固定する結合行列、破線矢印は学習により決定する結合行列を表す。なお、 $\Phi, \Psi$  の作用は図中に明示していない。

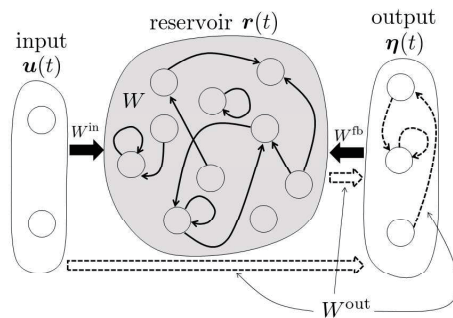


図 1: リザーバー計算の模式図

### 3 リザーバー計算による時系列予測

本章では、リザーバー計算を用いた時系列予測タスク、特に実 1 次元離散力学系  $x(t+1) = f(x(t))$  の定めるカオス的時系列  $\{x(t)\}$  の予測を扱うための枠組みを解説する。主に [1, 2] に基づくが、独自に工夫を加えている。

### 3.1 ネットワークの設計

機械学習におけるネットワーク設計では細部に関して幅広い選択の余地があるが、以下では本節の設計を標準とする。

入力なし、リザーバー  $\mathbf{r}(t) \in \mathbb{R}^N$ 、出力  $\eta(t) \in \mathbb{R}^1$  の RNN を考える。結合行列として  $W \in \{X \in M_{N,N}(\mathbb{R}) \mid \rho(X) = R, X_{ij} \geq 0 (i, j = 1, \dots, N)\}$ ,  $\mathbf{w}^{\text{fb}} \in \{\mathbf{x} \in \mathbb{R}^N \mid |x_i| = a (i = 1, \dots, N)\}$  はランダムに選んで固定し、 $\mathbf{w}^{\text{out}} \in M_{1,N}(\mathbb{R})$  を学習で決定する。ただし、 $N \in \mathbb{N}$  及び  $R, a \geq 0$  は実験ごとに選択・固定する。ここで、特に ESP の代替条件  $\rho(W) < 1$  は課していない。このことは、第 4 章の数値実験にとって重要な点である。また、リザーバー計算では  $W$  を疎行列とすることも一般的であるが、本稿ではそれも仮定しない。

学習終了後はリザーバー及び出力層の各ニューロンは、次の式に従って時間発展する：

$$\begin{aligned} \mathbf{r}(t+1) &= \overline{\tanh(W\mathbf{r}(t) + \eta(t)\mathbf{w}^{\text{fb}})}, \\ \eta(t+1) &= \tanh(\mathbf{w}^{\text{out}}\mathbf{r}(t+1)). \end{aligned}$$

但し  $\overline{\tanh}$  は要素ごとに  $\tanh$  を返す写像、すなわち

$$\overline{\tanh}(\mathbf{x}) := (\tanh(x_1) \dots \tanh(x_N))^{\text{T}} \quad (\mathbf{x} \in \mathbb{R}^N)$$

である。

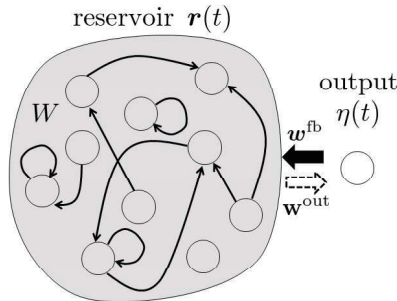


図 2: 本稿におけるネットワーク設計

### 3.2 アルゴリズムの概要

本稿で考える時系列予測のアルゴリズムの概略は、次のようにまとめられる：

#### リザーバー計算による時系列予測

Step 0 : 準備

$N \in \mathbb{N}$  及び  $R, a \geq 0$  を選び、

$$\begin{aligned} W &\in \{X \in M_{N,N}(\mathbb{R}) \mid \rho(X) = R, X_{ij} \geq 0 (i, j = 1, \dots, N)\}, \\ \mathbf{w}^{\text{fb}} &\in \{\mathbf{x} \in \mathbb{R}^N \mid |x_i| = a (i = 1, \dots, N)\} \end{aligned}$$

をランダムに選んで固定する。

また,  $T_0, T_1, T_2 \in \mathbb{N}$  を選ぶ ( $T_0 < T_1 < T_2$ )。

更に,  $x(0), \mathbf{r}(0)$  を選ぶ. 選び方は問題に応じて適宜工夫する.

Step 1 ( $t = 0, \dots, T_1$ ): 学習

1-1: 次の規則でリザバーの内部状態を更新する ( $T_1 + 1$  回):

$$\begin{aligned}x(t+1) &= f(x(t)), \\ \mathbf{r}(t+1) &= \overline{\tanh}(W\mathbf{r}(t) + y(t)\mathbf{w}^{\text{fb}}), \\ y(t) &= \tanh(x(t)).\end{aligned}$$

1-2: 1-1 で得た内部状態列  $\{\mathbf{r}(t)\}_{t=T_0+1}^{T_1}$  から定まる  $\{\mathbf{w}^{\text{out}}\mathbf{r}(t)\}_{t=T_0+1}^{T_1}$  と  $\{x(t)\}_{t=T_0+1}^{T_1}$  の誤差を最小化するように  $\mathbf{w}^{\text{out}}$  を決定する.

Step 2 ( $t = 0, \dots, T_2$ ): 予測

2-1 ( $t = 0, \dots, T_1$ ): リザバーを

$$\begin{aligned}x(t+1) &= f(x(t)), \\ \mathbf{r}(t+1) &= \overline{\tanh}(W\mathbf{r}(t) + y(t)\mathbf{w}^{\text{fb}}), \\ y(t) &= \tanh(x(t))\end{aligned}$$

の規則で更新し ( $T_1 + 1$  回), 予測時系列  $\{\mathbf{w}^{\text{out}}\mathbf{r}(t)\}_{t=T_0+1}^{T_1}$  を得る.

2-2 ( $t = T_1 + 1, \dots, T_2$ ): リザバーを

$$\begin{aligned}\mathbf{r}(t+1) &= \overline{\tanh}(W\mathbf{r}(t) + \eta(t)\mathbf{w}^{\text{fb}}), \\ \eta(t) &= \tanh(\mathbf{w}^{\text{out}}\mathbf{r}(t))\end{aligned}$$

の規則で更新し ( $T_2 - T_1$  回), 予測時系列  $\{\mathbf{w}^{\text{out}}\mathbf{r}(t)\}_{t=T_1+1}^{T_2}$  を得る.

以下, 技術的な注意を述べる.

- Step 0 において, 選んだ  $R, a \geq 0$  に対し  $W, \mathbf{w}^{\text{fb}}$  を「ランダムに選んで固定する」ことは, 実際には乱数シードを一つ固定することで実現する. 同一の乱数シードに対しては常に同一の  $W, \mathbf{w}^{\text{fb}}$  が再現される.
- 1-1, 2-1 において,  $f$  が生成する時系列  $\{x(t)\}$  は適宜平行移動して用いてもよい. たとえば第 4 章の数値実験では,  $f(x(t)) = 3.7x(t)(1 - x(t))$  に対し時系列  $\{x(t) - 0.5\}$  を用いて学習・予測を行い, 結果を表示する段階で元に戻した.
- 1-2 において,  $\{\mathbf{r}(t)\}_{t=1}^{T_0}$  は  $\mathbf{w}^{\text{out}}$  の決定に使わない. これは, 初期値  $x(0), \mathbf{r}(0)$  に依存する過渡的な挙動を無視するためである. また, 「誤差を最小化する」ということをどのように定式化するかはアルゴリズムを決定付ける重要な問題であるが, ここでは自乗誤差の最小化を考

える。これは方程式  $(R^T R)(\mathbf{w}^{\text{out}})^T = R\mathbf{x}$  の求解に帰着することが知られている (線型回帰)。但し  $R := (\mathbf{r}(T_0+1)^T \dots \mathbf{r}(T_1)^T)^T$ ,  $\mathbf{x} := (x(T_0+1) \dots x(T_1))^T$ 。本稿の数値実験における実際の求解には Python (NumPy) の線型代数ライブラリを活用した。

- 2-1 の応用上の主目的は、2-2 で必要なリザーバー及び出力の初期値  $\mathbf{r}(T_1+1), \eta(T_1+1)$  を合理的に与えることにある。また、2-1 では  $\{\mathbf{r}(t)\}$  の生成に毎回「正解」のデータ  $\{x(t)\}$  を用いるので、学習の成否を確認することにも役立つ。

力学系理論の観点からは、次の点は重要である。

- 2-2 はメインタスクであり、この時点でシステムは自励系

$$\mathbf{r}(t+1) = \overline{\tanh}(W\mathbf{r}(t) + \tanh(\mathbf{w}^{\text{out}}\mathbf{r}(t))\mathbf{w}^{\text{fb}}),$$

になる。その相空間を本稿では「リザーバー空間」と呼ぶ。

注: 以上の枠組みでは、たとえ学習済み RNN が正確に  $f$  を模倣していたとしても、初期値に対する鋭敏な依存性のために、時系列そのものを長時間にわたって正確に予測することはできない。

## 4 ロジスティック写像に対する数値計算結果

本章では具体的なタスクとして、ロジスティック写像  $x(t+1) = \alpha x(t)(1-x(t)) =: f(x(t))$  (但し  $\alpha = 3.7$ ) の定める時系列  $\{x(t)\}$  の予測を扱う。以下の数値計算は第 3 章に述べた枠組みを基本とし、反復回数は  $T_0 = 200, T_1 = 1200, T_2 = 2000$  とした。また、 $x(0) = 0.3$  とし、 $\mathbf{r}(0)$  は  $[-1, 1]^N$  からランダムに選んだ。

### 4.1 適切なパラメータの選択

$N = 30$  の場合に  $\rho(W)$  及び  $a := |w_i^{\text{fb}}|$  を変化させて、各場合について 200 組の  $(W, \mathbf{w}^{\text{fb}})$  で予測を行い、以下に述べる方法で「成功」数を評価した。成功数をヒートに対応させることで表現したヒートマップを図 3 に示す。「成功」の定義には以下に述べる問題があるものの、特に高い成功率は  $\rho(W) > 1$  のときに達成されること、よいパラメータを選べば乱数シードによらず安定した結果を得られることが結果から読み取れる。

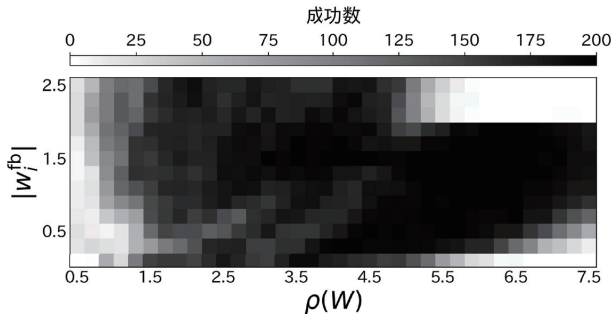


図 3: パラメータごとの成功率 ( $N = 30$ )

注：この実験では、「成功」を

1.  $T_1 + 1 \leq t \leq T_2 - 1$  における誤差  $|\mathbf{w}^{\text{out}}\mathbf{r}(t+1) - f(\mathbf{w}^{\text{out}}\mathbf{r}(t))|$  の平均が  $10^{-1.6}$  以下
2. 十分時間後に  $f$  の平衡点  $1 - 1/\alpha$  に相当する位置に「穴」があいておらず、かつ、 $1 - 1/\alpha$  に収束もしていない (区間  $[0.65, 0.75]$  を 10 等分したどの区間にもデータ点  $\mathbf{w}^{\text{out}}\mathbf{r}(t)$  ( $T_1 + 200 \leq t \leq T_2$ ) が少なくとも一つ含まれている)

とともに満足すること、と定義した。これは「時系列を生成するシステムがリザーバー計算によって再現されたか」を問うためのもので、とくに基準 2 は実験で図 4 のような「失敗」例が典型的に見られることを踏まえたものである。

注意すべきこととして、誤差や「穴」の許容範囲は著者の主観に従い仮に設けた基準に過ぎない。たとえば  $5.5 \leq \rho(W) \leq 7.5$  付近において、 $|w_i^{\text{fb}}| = 1.9$  では高い「成功」率を示しているにもかかわらず  $|w_i^{\text{fb}}| = 2.1$  では全く「成功」していないように見えるのは、前者は 200 例とも基準 2 をわずかに下回るのに対し後者はわずかに上回るためであり、それ以上の特異な差はないことが判明している。

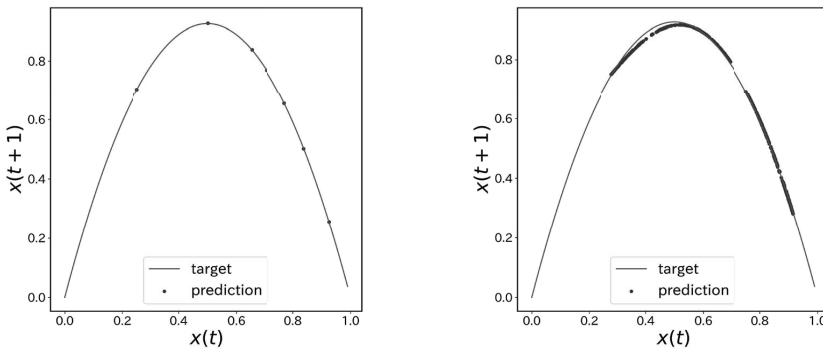


図 4: ( $\mathbf{w}^{\text{out}}\mathbf{r}(t), \mathbf{w}^{\text{out}}\mathbf{r}(t+1)$ )-グラフ (失敗例)

## 4.2 リザーバー空間の構造

$N = 30, \rho(W) = 5.0, |w_i^{\text{fb}}| = 1.5$  ( $i = 1, \dots, N$ ) としたある場合の ( $\mathbf{w}^{\text{out}}\mathbf{r}(t), \mathbf{w}^{\text{out}}\mathbf{r}(t+1)$ )-グラフを図 5 に示す。これはロジスティック写像の  $(x(t), x(t+1))$ -グラフとよく一致していることから、リザーバー計算によって学習対象のダイナミクスがよく再現されたと考えられる。

この例についてリザーバー空間で主成分分析 (PCA) を行ったところ、第 1 主成分の寄与率が約 94.8%、第 2 主成分の寄与率が約 5.2% であった。すなわち、点列  $\{\mathbf{r}(t)\}$  は 30 次元空間内でほぼ 2 次元的に分布していることが判明した。図 6a, 6b はそれぞれ、3, 2 次元までの主成分を抽出して図示したものである。

図 6a, 6b から、 $\{\mathbf{r}(t)\}$  はほぼ 1 次元多様体上に分布していることが観察される。特に  $\{\mathbf{r}(t)\}$  はその第 1 主成分と 1 対 1 に対応しているので、第 1 主成分への射影  $\pi_1: \mathbb{R}^N \rightarrow \mathbb{R}$  がこの 1 次元多

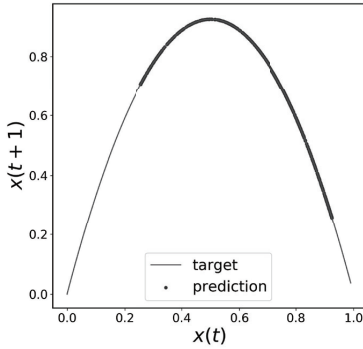
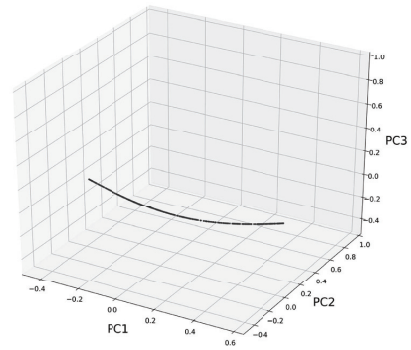
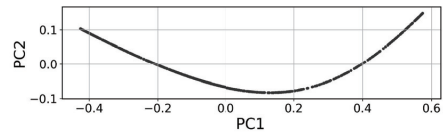


図5:  $(\mathbf{w}^{\text{out}}\mathbf{r}(t), \mathbf{w}^{\text{out}}\mathbf{r}(t+1))$ -グラフ (成功例)  
 $N = 30, \rho(W) = 5.0, |w_i^{\text{in}}| = 1.5 (i \leq N)$



(a) 第3主成分まで



(b) 第2主成分まで

図6: 主成分で可視化したリザーバー空間 ( $N = 30$ )

様体の座標を与えると考えてよい. そこで  $\{\mathbf{r}(t)\}$  の第1主成分を横軸に,  $\{\mathbf{w}^{\text{out}}\mathbf{r}(t)\}$  を縦軸にプロットしたところ, 図7を得た. これによれば  $\{\mathbf{r}(t)\}$  の第1主成分, 従って  $\{\mathbf{r}(t)\}$  は  $\{\mathbf{w}^{\text{out}}\mathbf{r}(t)\}$  と1対1に対応している. 更に, 既に図5で見た通り,  $\{\mathbf{w}^{\text{out}}\mathbf{r}(t)\}$  の挙動はロジスティック写像を精度よく近似している. 従って, 図7をある可逆な連続関数  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  のグラフと見なせば,  $\varphi \circ \pi_1$  はリザーバー空間の力学系とロジスティック写像の間の位相共役を与えていると見なせる.

以上要するに, 本節に述べた解析によってリザーバー空間にはロジスティック写像と位相共役な1次元力学系が埋め込まれていることが数値的に確認されたと考えられる.

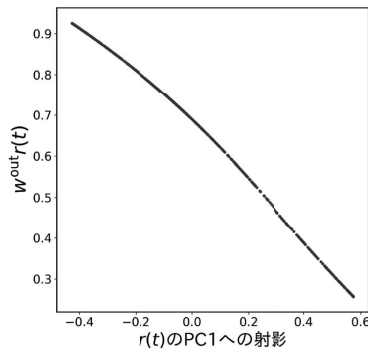


図7: 数値的に得られた共役写像

### 4.3 3次元のリザーバー

$N = 3$  の場合についても 4.1 節と同様の実験を行った。結果を図 8 に示す。これによって、本稿の問題に限れば、 $N = 3$  のリザーバーでも 50% 以上の確率で学習が成功する場合があることが判明した。そこで、 $N = 3, \rho(W) = 4.5, |w_i^{\text{fb}}| = 1.3 (i = 1, \dots, N)$  として成功した場合のリザーバー空間の様子を図 10 に示す。なお、3次元なので PCA は行っていない。

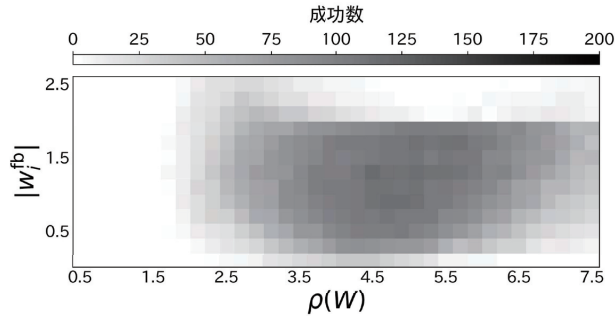


図 8: パラメータごとの成功率 ( $N = 3$ )

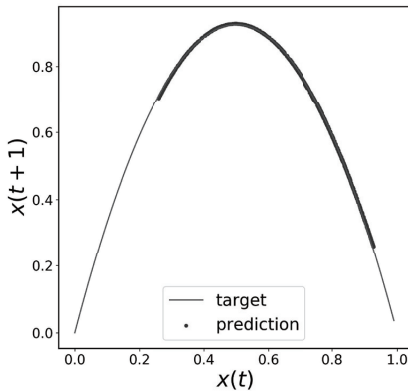


図 9:  $(\mathbf{w}^{\text{out}} \mathbf{r}(t), \mathbf{w}^{\text{out}} \mathbf{r}(t+1))$ -グラフ (成功例)  
 $N = 3, \rho(W) = 4.5, |w_i^{\text{fb}}| = 1.3 (i \leq N)$

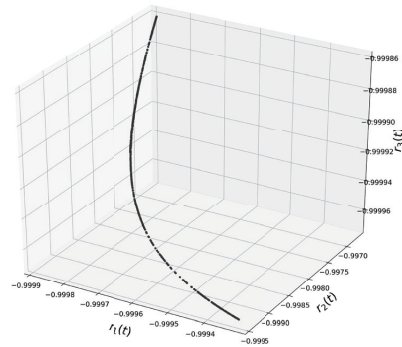


図 10: リザーバー空間 ( $N = 3$ )

## 5 おわりに

本稿では「ダイナミクスの再現」という観点から、第 2, 3 章を通じてリザーバー計算による時系列予測の枠組みを概説し、第 4 章では数値計算結果を示した。

今後の課題として、手法の再検討や双曲型を含む他の力学系での数値計算、本稿で述べた結果を数学解析の立場から整理することなどが挙げられる。



## 参考文献

- [1] Jaeger, H., & Haas, H. (2004). Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667), 78-80.
- [2] Jaeger, H. (2001). The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology GMD Technical Report, 148(34), 13.
- [3] 原誠人 (2021) 「リザーバー計算によるカオスの時系列予測における力学系の学習について」京都大学修士論文 (未公刊).