26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

Time-Uniform Universal Approximation and Recurrent Neural Networks

Adrian N. Bishop [a]

[a]*CSIRO, Australia*
Email: adrian.bishop@csiro.au

Abstract: This work examines the ability of recurrent neural networks to approximate the trajectories of random dynamical systems with stochastic inputs, on non-compact domains, and over infinite or indefinite time horizons. The main result states such random trajectories can be uniformly approximated in time to any desired accuracy. Unlike much of the existing literature—which typically focuses on compact state spaces and finite time intervals—this formulation addresses broader settings. The model assumptions are natural, minimal, and straightforward to verify.

Keywords: *Recurrent neural networks; Time-uniform approximation; Universal approximation.*

1 INTRODUCTION

We study **universal *time-uniform* trajectory approximation** of random dynamical systems with recurrent neural networks (RNN): i.e. we consider the properties of very general (random) dynamical systems and the RNN characteristics that enable trajectory approximation to any desired accuracy, uniformly in time, over an indefinite or infinite time interval. This work is also presented in Bishop (2022); Bishop and Bonilla (2023).

2 RANDOM DYNAMICAL SYSTEMS AND RECURRENT NEURAL NETWORK ARCHITECTURES

Let $\mathbb{X} \subseteq \mathbb{R}^{d_x}$, $\|\cdot\| : \mathbb{X} \to [0,\infty)$ be any norm on $\mathbb{X}$, and $t \in \mathbb{N} := \{0,1,2,\ldots\}$. Fix throughout the standard probability space $(\Omega, \mathcal{B}(\Omega), \mathsf{P})$. We denote by $\mathfrak{Lip}(\mathbb{X})$ the space of Lipschitz endomorphisms on $\mathbb{X}$. Consider the $\mathfrak{Lip}(\mathbb{X})$-valued random process defined by the indexed family $\{F_t \,|\, t \in \mathbb{N}\}$ of measurable maps $F_t : \Omega \to \mathfrak{Lip}(\mathbb{X})$. This is well defined, see Diaconis and Freedman (1999).

A random dynamical system is defined by the compositional iteration of random F_t acting on an initial $x_0 \in \mathbb{X}$,

$$X_t^{x_0} \;=\; F_t \cdots F_2 \cdot F_1(x_0) \;=\; F_t(X_{t-1}^{x_0}) \tag{1}$$

Let $\mathbb{U} \subseteq \mathbb{R}^{d_u}$. The indexed family $\{U_t \,|\, t \in \mathbb{N}\}$ of random variables $U_t : \Omega \to \mathbb{U}$ is a random process. We consider the maps in (1) of the form,

$$F_t(\omega)(\cdot) \;:=\; f(\cdot, U_t(\omega)) \tag{2}$$

where $f : \mathbb{X} \times \mathbb{U} \to \mathbb{X}$ is a deterministic, time-invariant, measurable map, Lipschitz in $\mathbb{X}$. See Diaconis and Freedman (1999); Arnold (2003) for further background on these systems.

Let $\sigma(\cdot) := \max\{\cdot, 0\} : \mathbb{R} \to \mathbb{H} \subseteq \mathbb{R}$; which acts component wise on a vector $v \in \mathbb{R}^d$ so $\sigma(v) \in \mathbb{H}^d$. A general deep recurrent neural network (RNN), with $L \in \mathbb{N}$ layers, is given in the form,

$$\widehat{X}_t \;=\; h_{L,t} \;=\; \tau_L(h_{L-1,t}) \tag{3}$$

$$h_{l,t} \;=\; \sigma\left(\tau_l(h_{l-1,t}) + \varphi_l(h_{t-1})\right), \qquad l \in \{1,\ldots,L-1\} \tag{4}$$

$$h_{0,t} \;:=\; U_t \tag{5}$$

where $h_{l,t} \in \mathbb{H}^{d_{h_l}}$ with $\mathbb{H}^{d_{h_0}} := \mathbb{U}$ and $\mathbb{H}^{d_{h_L}} := \mathbb{X}$. The feedback vector is $h_t = (h_{1,t},\ldots,h_{L,t}) \in \mathbb{H}^{d_h}$. The maps $\tau_l : \mathbb{H}^{d_{h_{l-1}}} \to \mathbb{R}^{d_{h_l}}$ and $\varphi_l : \mathbb{H}^{d_h} \to \mathbb{R}^{d_{h_l}}$ are finite affine maps. We write a RNN like (3), (4), (5) as,

$$\widehat{X}_t^{h_0} \;=\; \widehat{f}(h_{t-1}, U_t) \;=:\; \widehat{F}_t(h_{t-1}) \tag{6}$$

Specifying the parameters of the affine maps defines the topology of the network $\widehat{f}$. Specifying also the initial condition $h_0 \in \mathbb{H}^{d_h}$ defines the resulting RNN output for a given input sequence.

2.1 Map Approximation versus Trajectory Approximation

Map approximation seeks for any $\epsilon > 0$ a network $\widehat{f}(h,\cdot)$ and a network state $h \in \mathbb{H}^{d_h}$ such that,

$$\mathsf{E}\big[\,\|\widehat{f}(h,\cdot) - f(x,\cdot)\|^p\,\big]^{1/p} \;\leq\; \epsilon \tag{7}$$

for all x and all inputs in $\mathbb{X} \times \mathbb{U}$. The network initialisation is explicitly a part of the approximation. Map approximation theory follows from classical universal approximation results for feedforward networks, e.g. with the RNN frozen in time.

The problem of *trajectory approximation* seeks for any $\epsilon > 0$ a network $\widehat{f}$ and initialisation $h_0 \in \mathbb{H}^{d_h}$ such that,

$$\mathsf{E}\big[\, \big\| \widehat{F}_t \cdots \widehat{F}_2 \cdot \widehat{F}_1(h_0) \;-\; F_t \cdots F_2 \cdot F_1(x_0) \big\|^p \,\big]^{1/p} \;\leq\; \epsilon \tag{8}$$

for all times in a given interval $t \in \{0, \ldots, T\}$, $T \in \mathbb{N}$ and for all x_0 and all inputs in $\mathbb{X} \times \mathbb{U}^T$. Trajectory approximation over known fixed-length intervals is studied in Jin et al. (1995); Schäfer and Zimmermann (2007).

Finally, a trajectory approximation error bound that holds $\underline{\underline{\text{for all } t \in \mathbb{N}}}$ is a *time-uniform* error bound.

2.2 Results

Assume there is a finite $C > 0$, and $\lambda \in (0, 1)$, all independent of $s, t \in \mathbb{N}$, such that,

$$\mathsf{E}\big[\, \|F_t \cdots F_{s+1}(X_s^x) - F_t \cdots F_{s+1}(X_s^{x_0})\|^p \,\big] \;\leq\; C\,\lambda^{(t-s)}\,\mathsf{E}\big[\, \|X_s^x - X_s^{x_0}\|^p \,\big] \tag{9}$$

exists for all $x, x_0 \in \mathbb{X}$ and all $t > s$. A sequence $\{F_t \,|\, t \in \mathbb{N}\}$ that satisfies (9) is called *exponentially p-contractive*.

Theorem 1. *Consider the map $f : \mathbb{X} \times \mathbb{U} \to \mathbb{X}$ of (2) and a RNN $\widehat{f}$ of the form (3), (4), (5), see also (6). Let $X_0 : \Omega \to \mathbb{X}$ be random, independent, and have finite $p \geq 1$ moments. Assume $\{F_t(\omega) \,|\, t \in \mathbb{N}\}$ is stationary, and exponentially p-contractive, and $\mathsf{E}\,[\|F(x) - x\|^p] < \infty$. Then for any $\epsilon > 0$ there is a finite $\widehat{f}$ and an initial $h_0 \in \mathbb{H}^{d_h}$ such that,*

$$\mathsf{E}\big[\, \big\| \widehat{F}_t \cdots \widehat{F}_2 \cdot \widehat{F}_1(h_0) \;-\; F_t \cdots F_2 \cdot F_1(x_0) \big\|^p \,\big]^{1/p} \;\leq\; \epsilon \tag{10}$$

for all $t \in \mathbb{N}$ and the sequence $\{\widehat{F}_t(\cdot) := \widehat{f}(\cdot, U_t) \,|\, t \in \mathbb{N}\}$ is exponentially p-contractive.

Consider the Kalman filter $(\overline{X}_t, \mathbf{C}_t)$, with $\overline{X}_t = \mathsf{E}\,[X_t \,|\, U_1, \cdots, U_t]$ and $\mathbf{C}_t = \mathsf{Cov}\,[X_t \,|\, U_1, \cdots, U_t]$ for the simple underlying scalar system, $X_t = \alpha\, X_{t-1} + V_t$ and $U_t = X_t + \beta\, W_t$ where $\mathbb{X} = \mathbb{R}$, $\alpha \in \{0.98, 1.001\}$, $\beta \in \{1, 2\}$ and V_t, W_t are independent standard Gaussian white noises, independent of X_0 which is zero mean Gaussian with variance 25. The Kalman filter $(\overline{X}_t, \mathbf{C}_t)$, see Bishop and Bonilla (2023), is optimal, is exponentially contractive, and with $|\alpha| < 1$ the driving observation sequence is asymptotically stationary.

We compute the RMSE between $\overline{X}$ and the mean estimate from a basic particle and a separate RNN approximation (see Bishop (2022); Bishop and Bonilla (2023) for details); over $N_{\text{test}} = 1000$ independent examples. The test horizon is $T_{\text{test}} = 2000$ in all cases. In Fig 1 ($\alpha = 0.98$, $\beta \in \{1, 2\}$) and Fig 2 ($\alpha = 1.001$, $\beta = 2$) we plot the errors.

In each case in Fig 1 we train with $T_{\text{train}} = 20$, and $N_{\text{train}} = 5000$. With $\alpha = 0.98$ a version of Theorem 1 holds and the RNN-based method outperforms the particle filter (which does not handle accurate observations $\beta = 1$ well here). In Fig 2 we consider $T_{\text{train}} = 20$, $N_{\text{train}} = 5000$ (as before), and also $T_{\text{train}} = 200$, $T_{\text{train}} = 1000$, $T_{\text{train}} = 2000$.

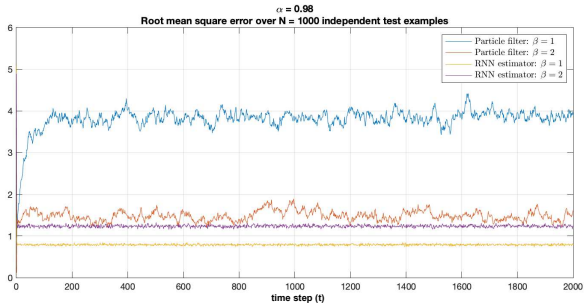

Figure 1. Root mean square errors.

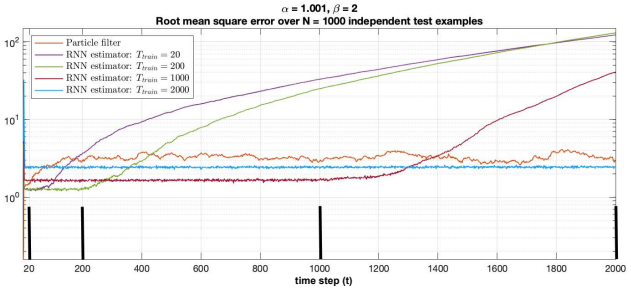

Figure 2. Root mean square errors (log vertical scale).

When the conditions of Theorem 1 are met, we may train on finite-length examples (often very short, maybe one or a handful of steps, depending on the ergodicity), but apply the filter indefinitely at test time with no unbounded accumulated growth of error. This is seen in Fig. 1. If either assumption is unmet, we may learn networks that work well on the length of the training data, but eventually the error may start to accumulate. This is seen in Fig. 2

References

Arnold Ludwig (2003), *Random Dynamical Systems*, corrected 2nd print edn, Springer.

Bishop Adrian N. (2022), Universal Time-Uniform Trajectory Approximation for Random Dynamical Systems with Recurrent Neural Networks, *arXiv e-print, arXiv: 2211.08018* .

Bishop Adrian N. and Bonilla Edwin V. (2023), Recurrent neural networks and universal approximation of Bayesian filters, *in* 'Proceedings of The 26th International Conference on Artificial Intelligence and Statistics', Vol. PMLR 206, pp. 6956–6967.

Diaconis Persi and Freedman David (1999), Iterated Random Functions, *SIAM Review* **41**(1), 45–76.

Jin Liang, Nikiforuk Peter N. and Gupta Madan M. (1995), Approximation of discrete-time state-space trajectories using dynamic recurrent neural networks, *IEEE Transactions on Automatic Control* **40**(7), 1266–1270.

Schäfer Anton Maximilian and Zimmermann Hans-Georg (2007), Recurrent Neural Networks are Universal Approximators, *International Journal of Neural Systems* **17**(04), 253–263.

