On the Linearizing Effect of Temporal Averaging in Nonlinear Dynamical Systems

Sabbir Ahmed   Erfan Nozari

Abstract— **Application of low pass filters (LPFs) to remove noise components is a widely used methodology for processing signals acquired from diverse systems. LPFs are also intrinsic components of many natural and man-made systems, both intended and epiphenomenal, from electromechanical systems to traffic networks and the brain. Across all cases, the effects of LPFs are often studied in a pure filtering sense, such as temporal smoothing and removing high-pass noise components, causing delays and phase distortions, or limiting bandwidths. In this work, we instead show that low-pass filtering and the temporal averaging that underlies it can also have a major and fundamental impact on the linearity of the dynamics. We show using rigorous analysis that across a wide range of stochastic nonlinear systems, temporal averaging dampens nonlinearities and leads to more and more linear dynamics with stronger temporal averaging (lower LPF cutoff frequency), leading asymptotically to a completely linear system as the width of the window over which temporal averaging occurs tends to infinity (LPF cutoff frequency tends to zero). Our results have major implications in a wide range of application areas, including the study of the nervous system whereby LPFs are biologically and algorithmically abundant and a growing body of empirical evidence has found linear models as capable as nonlinear ones in describing neuronal time series.**

I. Introduction

A critical aspect in data-driven analysis of systems and control is the behavior of the signals obtained from the system. While it is expected that signals acquired from complex nonlinear systems behave nonlinearly and are as such best described using nonlinear models, increasing empirical evidence has found linear models to equally well or even better describe data collected from fundamentally nonlinear systems [1]–[8]. The potential implications of such linearity are indeed major, from simplifying model development and control design to enhancing mechanistic understanding of the system's behavior. In this work, we provide fundamental theoretical support for the role that low-pass filtering and temporal averaging, whether due to inherent system characteristics or preprocessing steps after signal acquisition, can have linearizing effect on the dynamical content of behavioral system measurements.

Literature review. Filtering techniques or components are integral to the functioning of many engineered systems [9]–[16]. In data-driven systems and control, for example, filters are used ubiquitously to preprocess input-output data, where pre-filtering using low pass (LPF) or band pass filters (BPF) is the most commonly adopted strategy [17]. Aside from the fact that LPFs are widely used as part of the preprocessing step, there are also numerous inherent LPFs that exists within various natural networks and systems [11]–[13], [15], [18]. In brain networks, for example, neural dynamics are frequently observed or even defined through the signals that are low-pass filtered versions of micro- and meso-scale variables. Among these, the most notable signal is perhaps the blood-oxygen-level-dependent (BOLD) signal captured by functional magnetic resonance imaging (MRI), which can be viewed as an observation of neural activity through the brain's hemodynamic response function (HRF). The latter has a strong low-pass filtering effect, among others, with a cutoff frequency of the order 0.1Hz compared to the up to 200Hz bandwidth of neuronal activity. Also, the field potentials detected by various forms of intra- and extra-cranial electroencephalography (EEG) are for the most part simply aggregates (averages) of neuronal post-synaptic currents [11]. The latter are in turn low-pass filtered observations of neurons' spiking activity through synaptic transmission and neuronal membranes' resistive-capacitive circuits. A similar finding has been reported in [12], in which it has been shown that information transfer in neural systems occurs by filtering through integrator dynamics alongside uncorrelated intrinsic noise, which results in a low-pass filter overall. A further example of an inherent low pass filtering effect in the brain comes from [13] where the thalamus was shown to behave as a low pass filter. This study used a computational approach, based on a simplified, yet biologically reasonable model, and suggested that the thalamus functions as a low-pass filter in order to stabilize sensory representations.

The present study is further motivated by our previous work [1] on the dynamic modeling of brain networks. Therein, we provided empirical evidence using real neuroimaging and neurophysiological time series for an unexpected observation that linear models may indeed provide more accurate prediction of brain recordings compared to various nonlinear models. To support this observation, we showed therein using simulations that averaging, both spatial and temporal, may indeed linearize the dynamics of strongly nonlinear systems. These results were nevertheless only based on simulations of one specific neuronal model, hence only supporting a conjecture on the linearizing effect of spatiotemporal averaging and providing motivation for further rigorous analysis.

In our recent work [19], we provided the first mathematical proof of this conjecture in the case of spatial averaging. However, our proof therein relies heavily on an assumption of statistical independence across space, i.e., independence between different subsystems whose activity is spatially averaged. While feasible in the context of spatial averaging, an assumption of independence cannot be made across time

Sabbir Ahmed is with the Department of Electrical and Computer Engineering, University of California, Riverside, sahme094@ucr.edu. Erfan Nozari is with the Departments of Mechanical Engineering, Electrical and Computer Engineering, and Bioengineering, University of California, Riverside, erfan.nozari@ucr.edu.

as it would require the averaged time series (state trajectories of the system under study) to be white, and hence devoid of *any* dynamics. In this work, we therefore follow the general methodology of [19] but extend it to averaging over statistically correlated sequences, hence providing the first mathematical proof for the linearizing effect of temporal averaging.

Finally, our proposed framework relies on tools from probability theory [20]–[23] including, in particular, the central limit theorem, as well as concepts from the theory of strong mixing sequences [24].

Statement of contributions. In this paper, we develop a theoretical framework to explain the empirically observed linearizing effects of temporal averaging in nonlinear dynamical systems. Our contributions to this end are three-fold. First, we introduce a new notion of strong mixing for sequences of random variables, to which we refer to as exponential ρ-mixing. This notion provides a stronger (more restrictive) version of the standard ρ-mixing from the literature whereby an exponential decay rate is required for the correlation coefficient between elements of the sequence with increasing distance. Using extensive simulations, we nevertheless demonstrate that state trajectories from a wide range of nonlinear dynamical systems indeed seem to be all exponentially ρ-mixing. In addition, we prove that unlike standard ρ-mixing sequences for which the variance of their cumulative sum may grow at any rate in the range $O(1)$ to $O(N^2)$, the variance of the cumulative sum of exponentially ρ-mixing sequences grows at a rate of $O(N)$, i.e., at the same rate as independent and identically distributed (i.i.d) sequences. Building on this result, our second contribution then consists of proving a central limit theorem (CLT) for exponentially ρ-mixing sequences. As opposed to the existing CLT for ρ-mixing sequences in the literature which is only applicable to stationary sequences, our proposed CLT applies to any exponentially ρ-mixing sequence, whether stationary or not, as long as the variables in the sequence have a uniformly bounded variance. This proposed CLT serves as the foundation of our final and main contribution, where we prove that under very mild assumptions, the dynamical system that results from the temporal averaging of any stochastic nonlinear system converges to a linear time-invariant (LTI) system in the asymptotic limit of infinite averaging.

II. Notation

We use $\mathbb{R}$ and $\mathbb{Z}$ to denote the set of reals and integers, respectively. $\mathbf{0}$ denotes zero vectors and matrices of appropriate size. When a vector $\mathbf{y}$ or matrix $\mathbf{A}$ are block-partitioned, $\mathbf{y}_i$ and $\mathbf{A}_{ij}$ refer to the ith block of $\mathbf{y}$ and (i,j)th block of $\mathbf{A}$, respectively. Throughout this work, all probabilities are defined on measurable spaces consisting of a Euclidean space (or a subset thereof) and the associated Borel σ-algebra. Hence, when clear from the context, the space over which each probability is defined is omitted. $\mathbb{E}[\cdot]$ and $\mathbb{P}\{\cdot\}$ denote expectation and probability, respectively. $\mathrm{Var}(\xi)$ denotes variance of random variable ξ. For two random vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$, $\mathrm{Cov}(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{n\times m}$ denotes the covariance between them. For simplicity, $\mathrm{Cov}(\mathbf{x}) = \mathrm{Cov}(\mathbf{x}, \mathbf{x})$. For sequences of random variables, $\xrightarrow{d}$ and $\xrightarrow{p}$ denote their convergence in distribution and probability, respectively. Finally, for any $t \in \mathbb{R}$, $\lfloor t \rfloor = \max\{k \in \mathbb{Z} | k \le t\}$.

III. Preliminaries

Here we review basic notions and properties of of mixing sequences that will be used in our ensuing discussion.

Mixing Sequences

For a stochastic process $\xi(t)$, the notion of mixing implies that the statistical dependence between $\xi(t_1)$ and $\xi(t_2)$ diminishes as $|t_1 - t_2|$ increases. In other words, mixing conditions generalize the notion of a pairwise independent (white) sequence to one in which nearby elements can be dependent but their dependence decays as the distance between them grows. Various alternative versions of mixing sequences are then proposed, corresponding to different measures of dependence which has to decay with distance. One of the most practical and empirically verifiable versions is that of ρ-mixing, as defined next.

Definition III.1. ***(ρ-mixing sequence [25]).*** *Consider a sequence of random variables $\xi_1, \xi_2, \dots$ in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and define*

$$\rho(n) = \sup_j \ \rho(\sigma(\xi_1, \dots, \xi_j), \sigma(\xi_{j+n}, \dots, \xi_\infty)), \quad (1)$$

where $\sigma(\xi_1, \dots, \xi_j)$ is the smallest σ-algebra of Ω generated by the variables $\xi_1, \dots, \xi_j$, similarly for $\sigma(\xi_{j+n}, \dots, \xi_\infty)$, and for any two σ-fields $\mathcal{A}$ and $\mathcal{B}$,

$$\rho(\mathcal{A}, \mathcal{B}) = \sup \frac{|\mathrm{Cov}(y, z)|}{\mathrm{Var}(y)^{1/2}\mathrm{Var}(z)^{1/2}} \quad (2)$$

where the supremum is taken over all pairs of squared-integrable random variables y and z such that y is $\mathcal{A}$-measurable and z is $\mathcal{B}$-measurable. The sequence $\xi_1, \xi_2, \dots$ is then ρ-mixing if

$$\rho(n) \to 0 \quad \textit{as} \quad n \to \infty. \qquad \square$$

Clearly, i.i.d sequences are special cases of mixing sequences. Thus, many properties of i.i.d sequences, such as the laws of large numbers [26], [27] and CLT [22], [28] continue to apply to mixing sequences as long as the decay of dependence is sufficiently rapid. Of particular relevance here is the following CLT for ρ-mixing sequences.

Proposition III.2. ***(CLT for stationary ρ-mixing sequences [22, Thm B]).*** *Consider a stationary ρ-mixing sequence $\xi_1, \xi_2, \dots$ each having zero mean and finite variance. Define the cumulative sum and cumulative variance*

$$S_N = \sum_{i=1}^N \xi_i, \quad \sigma_N^2 = \mathrm{Var}(S_N) \quad (3)$$

and assume that

$$\mathbb{E}[\xi_i] = 0, \qquad \mathrm{Var}(\xi_i) < \infty, \qquad \forall i, \quad (4)$$

$$\sup_{M\ge 0, N\ge 1} \frac{1}{\sigma_N^2}\ \mathbb{E}[(S_{M+N} - S_M)^2] < \infty. \quad (5)$$

Then

$$\frac{S_N}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad N \to \infty. \quad \square \tag{6}$$

In Section V, we extend this result to strongly ρ-mixing sequences relaxing the need for stationarity.

IV. Problem Statement

Consider the general discrete-time stochastic nonlinear system

$$\begin{aligned} \mathbf{x}(t+1) &= f(\mathbf{x}(t), \mathbf{w}(t)), \qquad t \geq 0 \\ \mathbf{w}(t) &\sim p_w, \\ \mathbf{x}(0) &\sim p_0, \end{aligned} \tag{7}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$ represents the state of the dynamical system at time t, $\mathbf{w}(t) \in \mathbb{R}^m$ represents a stationary noise process (not necessarily white or Gaussian) with the distribution p_w, and p_0 represents the distribution of the initial state $\mathbf{x}(0)$. Throughout this work, we make the following standing assumptions about (7).

Assumption 1. ***(Standing assumptions).***

(A1) *The noise has finite variance, i.e., for all i and t*

$$\mathrm{Var}(w_i(t)) \leq \bar{\sigma}_{w_i}^2 < \infty.$$

(A2) *The solution of* (7) *has a finite variance, i.e., for all i and t*

$$\mathrm{Var}(x_i(t)) \leq \bar{\sigma}_{x_i}^2 < \infty.$$

(A3) *The sequence of random vectors*

$$\mathbf{z}(t) = \begin{bmatrix} \mathbf{x}(t+1)^T & \mathbf{x}(t)^T & \mathbf{w}(t)^T \end{bmatrix}^T, \quad t \geq 0 \tag{8}$$

is a regular exponentially ρ-mixing sequence (cf. Definitions V.1 and V.3). $\square$

Assumptions (A1) and (A2) are satisfied by most, if not all, models of real-world systems due to the inherent boundedness of their noise and state trajectories. The main assumption of our work is therefore (A3), which we discuss in details in Section V and Appendix A.

We are now ready to formulate our problem of interest.

Problem 1. ***(Effects of temporal averaging on aggregate dynamics of a nonlinear dynamical system).*** *Consider a dynamical system in* (7) *and assume that Assumption 1 holds. Define the mean-centered and low-pass filtered state trajectory*

$$\bar{\mathbf{x}}(t) = \frac{1}{\sqrt{T}} \sum_{\tau=t}^{t+T-1} (\mathbf{x}(\tau) - \mathbb{E}[\mathbf{x}(\tau)]). \tag{9}$$

Show that the dynamics of $\bar{\mathbf{x}}(t)$ converges to a linear system as $T \to \infty$. $\square$

Notice that the signal $\bar{\mathbf{x}}(t)$ is the result of applying a linear filter with impulse response (assuming for simplicity that T is odd)

$$g(t) = \frac{1}{\sqrt{T}} \mathrm{rect}\Big(\frac{t-(T-1)/2}{T}\Big) \tag{10}$$

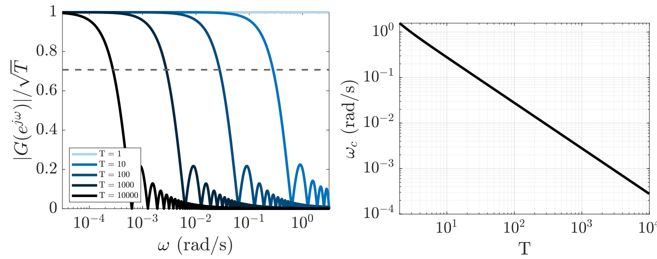

Fig. 1: The frequency response (left) and cutoff frequency (right) of the rectangular low-pass filter in (10). The horizontal line in the left panel shows the 3dB threshold.

to the mean-centered signal $\mathbf{x}(t) - \mathbb{E}[\mathbf{x}(t)]$, where $\mathrm{rect}(n/T) = 1$ if $|n| \leq T/2$ and $\mathrm{rect}(n/T) = 0$ otherwise. This filter has the low-pass frequency response [29]

$$G(e^{j\omega}) = \frac{1}{\sqrt{T}} \frac{\sin(T\omega/2)}{\sin(\omega/2)} e^{-j\omega(T-1)/2}$$

the amplitude and cutoff frequency of which are shown in Figure 1. Also note that for simplicity of notation, we have taken a forward average in (9) which makes the filter in (10) non-causal. However, everything that follows would still hold if $\mathbf{x}(t)$ is defined over $(-\infty, \infty)$ and $\bar{\mathbf{x}}(t)$ is defined using a backwards (causal) average over $\tau = t - T + 1, \ldots, t$. Finally, we highlight that our choice of a rectangular LPF low-pass filter is only for analytical tractability while, based on our extensive simulations, we expect all LPFs to impose the linearization effect noted in Problem 1 (cf., e.g., the effects of a Gaussian LPF in our prior work [1]).

V. Exponential ρ-Mixing

In this section we present the main notion of mixing for sequences of random variables that will be the basis for our main result in Section VI. In particular, we introduce the new notion of exponentially ρ-mixing sequences and characterize some of its key properties, including the growth rate of the cumulative variance of, as well as a CLT for, exponentially ρ-mixing sequences. These results will play a central role in our proof of the linearizing effect of temporal averaging in Theorem VI.1.

Definition V.1. ***(Exponentially ρ-mixing sequence).*** *Consider a sequence of random variables $\xi_1, \xi_2, \ldots$ and let $\rho(n)$ be defined as in* (1). *This sequence is exponentially ρ-mixing if there exists constants $C \geq 0$ and $r \in [0,1)$ such that*

$$\rho(n) \leq C r^n, \quad \forall n \geq 1. \tag{11}$$

Accordingly, a sequence of random vectors $\mathbf{z}_1, \mathbf{z}_2, \cdots \in \mathbb{R}^k$ is exponentially ρ-mixing if the sequences of random variables $\boldsymbol{\theta}^T \mathbf{z}_1, \boldsymbol{\theta}^T \mathbf{z}_2, \ldots$ are exponentially ρ-mixing for all $\boldsymbol{\theta} \in \mathbb{R}^k$. $\square$

Exponential ρ-mixing is a considerably stronger assumption than the ρ-mixing itself, as is the case, e.g., with exponential vs. asymptotic stability. However, we are yet to find a *bounded-variance* stochastic nonlinear dynamical system of the form (7) with practical relevance whose solutions are not exponentially ρ-mixing. To express this observation more formally, we present in Appendix A the results of our simulations of three nonlinear systems with fundamentally different behaviors (asymptotic stability, limit cycle, chaos) and their respective estimates $\hat{\rho}(n)$. In all cases, we observe an exponential decay in $\hat{\rho}(n)$, suggesting that Assumption 1 is indeed mild. In the remainder of this section, we will characterize this new mixing definition before using it to solve Problem 1 in Section VI.

As one may have noticed, a major difference between Proposition III.2 and the standard CLT for i.i.d. sequences is the normalization by σ_N in (6) vs. the standard normalization by $\sqrt{N}$ for i.i.d. sequences. This difference stems from the fact that the growth rate of the cumulative variance (σ_N^2) of ρ-mixing sequences cannot be determined a priori, except for a global upper bound of $O(N^2)$ applicable to all sequences of random variables with or without mixing conditions (based on a simple application of the Cauchy-Schwarz inequality omitted here). In contrast, we next show that the cumulative variance of exponentially ρ-mixing sequences satisfies the same growth rate of $O(N)$ possessed by i.i.d. sequences.

Theorem V.2. ***(Growth rate of cumulative variance of exponentially ρ-mixing sequences).*** *Consider a zero-mean exponentially ρ-mixing sequence $\xi_1, \xi_2, \dots$. Assume that*

$$\mathrm{Var}(\xi_i) \le \bar{\sigma}^2 < \infty, \qquad \forall i \tag{12}$$

for some constant $\bar{\sigma}$. Then the cumulative variance σ_N^2 as defined in (3) *is $O(N)$.* □

Theorem V.2 shows that, unlike standard ρ-mixing sequences, the cumulative variance of an exponentially ρ-mixing sequence cannot grow any faster than the cumulative variance of an i.i.d. sequence. In other words, exponential ρ-mixing ensures that the correlations between nearby elements decay fast enough so that the sum of their pairwise covariances cannot dominate the sum of their variances. Nevertheless, it is still possible for an exponentially ρ-mixing sequence to have a σ_N^2 that grows slower than N. In other words, the correlation between nearby elements can still be sufficiently negative for them to cancel out individual variances. However, this is an extremely rare event that we have not observed in any dynamical systems with practical relevance. In the following, we characterize the scope of this possibility before assuming that it does not happen in (7).

Consider again an exponentially ρ-mixing sequence $\xi_1, \xi_2, \dots$ satisfying (12). For any $i \ge 1$, the infinite series $\sum_{j=1}^{\infty} \mathrm{Cov}(\xi_i, \xi_j)$ is absolutely convergent by the comparison test, the facts that $|\mathrm{Cov}(\xi_i, \xi_j)| \le \bar{\sigma}^2 C r^{|i-j|}$, and the fact that

$$\sum_{j=1}^{\infty} \bar{\sigma}^2 C r^{|i-j|} = \bar{\sigma}^2 C \frac{r - r^i + 1}{1 - r}. \tag{13}$$

For the simplicity of notation, let

$$C_{\boldsymbol{\xi}}^i \triangleq \sum_{j=1}^{\infty} \mathrm{Cov}(\xi_i, \xi_j)$$

denote the convergence value of this series. We can then present the following definition.

Definition V.3. ***(Regular exponentially ρ-mixing sequence).*** *An exponentially ρ-mixing sequence satisfying* (12) *is called regular if the limit*

$$\bar{C}_{\boldsymbol{\xi}} = \lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} C_{\boldsymbol{\xi}}^i \tag{14}$$

exists and is nonzero. Likewise, an exponentially ρ-mixing sequence of random vectors $\mathbf{z}_1, \mathbf{z}_2, \dots \in \mathbb{R}^k$ satisfying $\|\mathrm{Cov}(\mathbf{z}(t))\| \le \bar{\sigma}_{\mathbf{z}}^2 < \infty$ for all t is called regular if $\boldsymbol{\theta}^T \mathbf{z}(1), \boldsymbol{\theta}^T \mathbf{z}(2), \dots$ is regular for all $\boldsymbol{\theta} \in \mathbb{R}^k$. □

While examples of bounded-variance exponentially ρ-mixing sequences can be built which are not regular, they are often highly contrived. For instance,

- a sequence of independent random variables $\xi_1, \xi_2, \dots$ with $\mathrm{Var}(\xi_i) = 2 + \cos(\log(i))$ is an example whereby the limit in (14) fails to exist[1];
- a sequence of the form $\xi_1, -\xi_1, \xi_2, -\xi_2, \dots$ with independent $\xi_i, \xi_j, i \ne j$ is an example whereby the limit in (14) is zero.

Otherwise, notice that all the terms $C_{\boldsymbol{\xi}}^i$ are uniformly bounded between $\pm \bar{\sigma}^2 C \frac{1+r}{1-r}$ from (13) and, thus, regularity of $\{\xi_i\}_{i \ge 1}$ only requires this uniformly bounded sequence to have an infinite-horizon average and for the terms of that average not to precisely cancel each other. This has been the case for the state trajectories of all dynamical systems of interest that we have examined, and can also be proved in limited cases such as the state trajectories of dynamical systems that converge to a stationary distribution and have (a sufficiently strong noise input such that) $r < \frac{1}{3}$.[2]

The following Corollary to Theorem V.2 makes its assertion sharper when assuming regularity.

Corollary V.4. ***(Growth rate of cumulative variance of regular exponentially ρ-mixing sequences).*** *Consider a regular zero-mean exponentially ρ-mixing sequence $\xi_1, \xi_2, \dots$ satisfying* (12)*. Then the cumulative variance σ_N^2 as defined in* (3) *satisfies*

$$\lim_{N \to \infty} \frac{\sigma_N^2}{N} = \bar{C}_{\boldsymbol{\xi}}. \quad \square \tag{15}$$

While it may not have been obvious from (14), Corollary V.4 implies in particular that $\bar{C}_{\boldsymbol{\xi}} > 0$. Together with (13) this ensures that $\bar{C}_{\boldsymbol{\xi}}$ satisfies the bounds

$$0 < \bar{C}_{\boldsymbol{\xi}} \le \bar{\sigma}^2 C \frac{1 + r}{1 - r}.$$

Before proving the last and main result of this section, i.e., a CLT for regular exponentially ρ-mixing sequences, we need a lemma as follows.

Lemma V.5. ***(Invariance of regularity under finite shifts).*** *Consider a regular exponentially ρ-mixing sequence*

[1] Note that similar oscillatory examples with $\mathrm{Var}(\xi_i) = 2 + \cos(i)$ or $\mathrm{Var}(\xi_i) = 2 + (-1)^i$ are nevertheless regular.

[2] In brief, this is due to the facts that in such a system $C_{\boldsymbol{\xi}}^i$ converges to a limit $C_{\boldsymbol{\xi}}^{\infty}$, which would then also be the value of the limit in (14), and $C_{\boldsymbol{\xi}}^{\infty} > 0$ because $\frac{2r}{1-r} < 1$ and hence the variance term in $C_{\boldsymbol{\xi}}^{\infty}$ dominates all the covariance terms.

$\xi_1, \xi_2, \ldots$ *satisfying* (12)*. Then, for any* $k < \infty$, $\zeta_i = \xi_{i+k}, i = 1, 2, \ldots$, *is also exponentially* ρ*-mixing and regular with* $\bar{C}_{\boldsymbol{\zeta}} = \bar{C}_{\boldsymbol{\xi}}$. □

We are now ready to prove our CLT for regular exponentially ρ-mixign sequences. Unlike Proposition III.2, this CLT holds for stationary and non-stationary sequences alike and relies on the precise growth rate obtained in Corollary V.4 to scale the cumulative sum of the sequence and obtain the variance of the limit normal distribution. We also no longer need to assume (5) as it follows from the regularity of the sequence.

Theorem V.6. ***(CLT for regular exponentially*** ρ***-mixing sequences).*** *Consider a regular zero-mean exponentially* ρ*-mixing sequence* $\xi_1, \xi_2, \ldots, \ldots$ *satisfying* (12)*, and let* S_N *and* $\bar{C}_{\boldsymbol{\xi}}$ *be as defined in* (3) *and* (14)*, respectively. Then*

$$\frac{S_N}{\sqrt{N}} \xrightarrow{d} \mathcal{N}(0, \bar{C}_{\boldsymbol{\xi}}) \quad \textit{as} \quad N \to \infty. \qquad \square$$

Theorem V.6 will play a key role in the proof of the linearizing effect of temporal averaging, as showed next.

VI. Main Result

In this section we present the main result of our paper with regards to Problem 1, as follows.

Theorem VI.1. ***(Linearizing effect of temporal averaging).*** *Consider the nonlinear dynamics* (7) *and assume that Assumption 1 holds. Let* $\bar{\mathbf{x}}(t)$ *be defined as in* (9) *and*

$$\bar{\mathbf{w}}(t) = \frac{1}{\sqrt{T}} \sum_{\tau=t}^{t+T-1} \mathbf{w}(\tau) - \mathbb{E}[\mathbf{w}(\tau)].$$

Then the relationship between $\bar{\mathbf{x}}(t+1)$, $\bar{\mathbf{x}}(t)$ *and* $\bar{\mathbf{w}}(t)$ *becomes asymptotically linear as* $T \to \infty$*, i.e., for any* $\boldsymbol{\xi} \in \mathbb{R}^n$ *and* $\boldsymbol{\omega} \in \mathbb{R}^m$*,*

$$\mathbb{E}\left[\bar{\mathbf{x}}(t+1) \,\middle|\, \begin{bmatrix} \bar{\mathbf{x}}(t) \\ \bar{\mathbf{w}}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\omega} \end{bmatrix} \right] \to \mathbf{A}\boldsymbol{\xi} + \mathbf{B}\boldsymbol{\omega} \quad \textit{as} \quad T \to \infty,$$

where $\mathbf{A}$ *and* $\mathbf{B}$ *are constant time-invariant matrices.* □

Note that the noise input to the resulting average system is the average of the noise inputs $\bar{\mathbf{w}}(t)$. Additionally, among the many notable aspects of Theorem VI.1 is the time-invariant nature of the average dynamics despite the lack of any stationarity requirements on the nonlinear system (7). This is in contrast to our earlier work [19] where the result of spatially-averaged dynamics becomes linear *time-varying*, unless additional stationarity assumptions are imposed. We defer a comprehensive unification of the two linearizing effect (spatial and temporal) to future work.

VII. Conclusions and Future Work

In this paper we presented a theoretical framework for understanding the linearity of temporally averaged dynamics of stochastic nonlinear dynamical systems. This study was inspired by the empirically-observed linearity of dynamics in nonlinear systems subject to low-pass filtering and the ubiquity of LPFs across natural and man-made systems, our recent dual demonstration of the linearizing effects of spatial averaging in large-scale populations of nonlinear systems. We showed using extensive numerical simulations that a wide range of nonlinear systems appear to exhibit an exponential decay in the correlation coefficient of their temporally-separated state variables, and as such extended the standard definition of ρ-mixing to exponentially ρ-mixing sequences. We then provided multiple characterizations of such sequences, including the i.i.d.-like growth rate of their cumulative variance and a corresponding central limit theorem. Using the latter, we then proved that bounded nonlinear systems with exponentially ρ-mixing solutions converge to a linear time-invariant (LTI) system when passed through a LPF with smaller and smaller cutoff frequency. Future work will include the extension of our framework to networked systems with correlated subsystems that explains linearity in both spatially and temporally averaged dynamics.

Appendix A. Numerical Evaluation of Exponential ρ-Mixing in Dynamical Systems

In this Appendix we provide numerical support for the mildness of Assumption (A3). Consider three nonlinear dynamical systems,

$$S_1 : x(t+1) = \frac{10x(t)}{1 + x^2(t)} + w(t), \tag{16}$$

$$S_2 : \mathbf{x}(t+1) = \begin{bmatrix} 0 & 1 \\ 0.4 & 0.4 + \sin(0.2x_2(t)) \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ w(t) \end{bmatrix},$$

$$S_3 : \mathbf{x}(t+1) = \begin{bmatrix} 0 & 1 \\ -0.2 & 2.75 - x_2^2(t) \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ 10^{-\frac{3}{2}} w(t) \end{bmatrix}.$$

System S_1 is bistable while S_2 has a limit cycle and S_3 (the Duffling map) is chaotic, each representing a structurally distinct and purely nonlinear behavior.

For each system, we simulated its dynamics with i.i.d. $w(t) \sim \mathcal{N}(0, 1)$ for $0 \le t \le 999$ and computed $\{\boldsymbol{\theta}^T \mathbf{z}(t)\}_{0 \le t \le 999}$ using a random vector $\boldsymbol{\theta}$ with i.i.d. elements $\theta_i \sim \mathcal{N}(\mu_\theta, \sigma_\theta^2)$ where $\mu_\theta \sim U[0, 10]$ and $\sigma_\theta \sim U[1, 10]$ are themselves drawn from uniform distributions. Then, for computational feasibility, we limit the number of variables generating each σ-algebra in 1 to 3, i.e., $\rho(n) \simeq \rho(\sigma(\xi_1, \xi_2, \xi_3), \sigma(\xi_{3+n}, \xi_{3+n+1}, \xi_{3+n+2}))$. Each of the random variables y and z in (2) can then be any functions of (ξ_1, ξ_2, ξ_3) or $(\xi_{3+n}, \xi_{3+n+1}, \xi_{3+n+2})$. We approximate this functional space basis expansions of the form

$$f(\xi_1, \xi_2, \xi_3) = \sum_{i,j,k} c_{i,j,k} T_i(\xi_1) T_j(\xi_2) T_k(\xi_3),$$

where $c_{i,j,k} \sim \mathcal{N}(0, 1)$ and $T_n(\cdot)$ is the n'th Chebyshev basis function given by

$$T_n(x) = \begin{cases} \cos(n \arccos(x)), & |x| \le 1 \\ \frac{1}{2}((x - \sqrt{x^2 - 1})^n + (x + \sqrt{x^2 - 1})^n), & |x| \ge 1 \end{cases}$$

The absolute value of the correlation coefficient between $f(\xi_1, \xi_2, \xi_3)$ and $f(\xi_{3+n}, \xi_{3+n+1}, \xi_{3+n+2})$ computed analogously provides one approximation for $\rho(n)$. For each $\boldsymbol{\theta}$ and $0 \le n \le 997$, we then repeat this process 10^6 times, and repeat this entire process for 10 random values for $\boldsymbol{\theta}$.

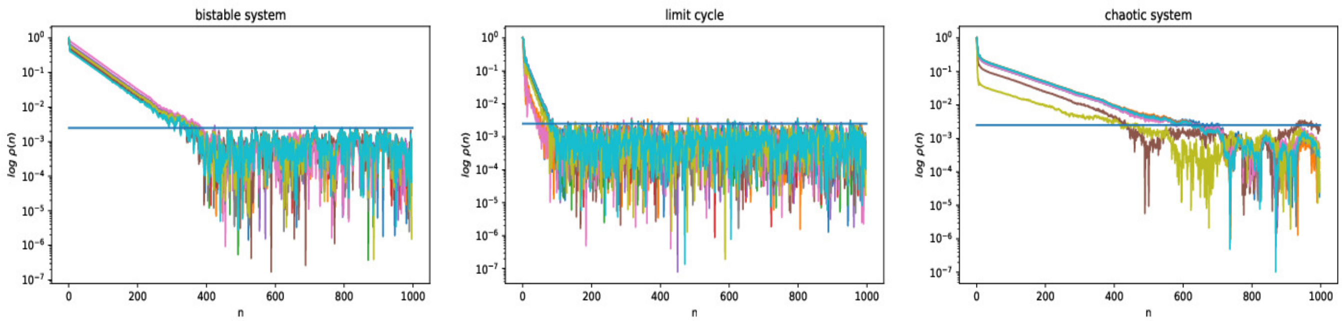

Fig. 2: **Exponential ρ-mixing property of sampled dynamical systems.** From left to right, panels show the exponential decay rate of 10 estimates of $\rho(n)$ for the systems S_1 to S_3 in (16).

Figure 2 shows the resulting approximations of $\rho(n)$ for each system. The horizontal line in each panel shows the theoretical smallest value of the correlation coefficient between any two random variables that can be accurately detected using 10^6 samples. This threshold was computed using standard power analysis [30] with standard bounds of 0.05 and 0.2 on type I and type II errors, respectively.

The semi-logarithmic panels of Figure 2 clearly show an exponential decay in all estimates of $\rho(n)$ for all the three systems, until they reach the theoretical threshold imposed by the finite, though as large as feasible using our computational resources, number of samples 10^6.

REFERENCES

[1] E. Nozari, M. A. Bertolero, J. Stiso, L. Caciagli, E. J. Cornblath, X. He, A. S. Mahadevan, G. J. Pappas, and D. S. Bassett, "Is the brain macroscopically linear? a system identification of resting state dynamics," *Nature Biomedical Engineering*, 2022, under Review.

[2] Y. Yang, S. Qiao, O. G. Sani, J. I. Sedillo, B. Ferrentino, B. Pesaran, and M. M. Shanechi, "Modelling and prediction of the dynamic responses of large-scale brain networks during direct electrical stimulation," *Nature biomedical engineering*, vol. 5, no. 4, pp. 324–345, 2021.

[3] O. G. Sani, H. Abbaspourazad, Y. T. Wong, B. Pesaran, and M. M. Shanechi, "Modeling behaviorally relevant neural dynamics enabled by preferential subspace identification," *Nature Neuroscience*, vol. 24, no. 1, pp. 140–149, 2021.

[4] T. Luo, Z.-Q. J. Xu, Z. Ma, and Y. Zhang, "Phase diagram for two-layer relu neural networks at infinite-width limit," *Journal of Machine Learning Research*, vol. 22, no. 71, pp. 1–47, 2021.

[5] Z. Liu, C. Rios, N. Zhang, L. Yang, W. Chen, and B. He, "Linear and nonlinear relationships between visual stimuli, eeg and bold fmri signals," *Neuroimage*, vol. 50, no. 3, pp. 1054–1066, 2010.

[6] M.-A. Schulz, T. Yeo, J. Vogelstein, J. Mourao-Miranada, J. Kather, K. Kording, B. A. Richards, and D. Bzdok, "Deep learning for brains?: Different linear and nonlinear scaling in uk biobank brain images vs. machine-learning datasets," *BioRxiv*, p. 757054, 2019.

[7] J. Lee, L. Xiao, S. Schoenholz, Y. Bahri, R. Novak, J. Sohl-Dickstein, and J. Pennington, "Wide neural networks of any depth evolve as linear models under gradient descent," *Advances in neural information processing systems*, vol. 32, 2019.

[8] C. Liu, L. Zhu, and M. Belkin, "On the linearity of large non-linear models: when and why the tangent kernel is constant," *Advances in Neural Information Processing Systems*, vol. 33, pp. 15 954–15 964, 2020.

[9] A. K. Tilton, S. Ghiotto, and P. G. Mehta, "A comparative study of nonlinear filtering techniques," in *Proceedings of the 16th International Conference on Information Fusion*. IEEE, 2013, pp. 1827–1834.

[10] C.-L. Hsu, M.-H. Ho, Y.-K. Wu, and T.-H. Chen, "Design of low-frequency low-pass filters for biomedical applications," in *APCCAS 2006-2006 IEEE Asia Pacific Conference on Circuits and Systems*. IEEE, 2006, pp. 690–695.

[11] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—eeg, ecog, lfp and spikes," *Nature reviews neuroscience*, vol. 13, no. 6, pp. 407–420, 2012.

[12] B. Lindner, "Mechanisms of information filtering in neural systems," *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 5–15, 2016.

[13] W. M. Connelly, M. Laing, A. C. Errington, and V. Crunelli, "The thalamus as a low pass filter: filtering at the cellular level does not equate with filtering at the network level," *Frontiers in neural circuits*, vol. 9, p. 89, 2016.

[14] A. Sang and S.-q. Li, "A predictability analysis of network traffic," *Computer networks*, vol. 39, no. 4, pp. 329–345, 2002.

[15] J. Narula, A. M. Smith, B. Gottgens, and O. A. Igoshin, "Modeling reveals bistability and low-pass filtering in the network module determining blood stem cell fate," *PLoS computational biology*, vol. 6, no. 5, p. e1000771, 2010.

[16] Q. Li, Z. Ma, and W. Du, "A low pass filter traffic shaping mechanism for the arrival of traffic," in *2007 International Conference on Wireless Communications, Networking and Mobile Computing*. IEEE, 2007, pp. 1032–1035.

[17] L. Ljung, *System Identification (2nd Ed.): Theory for the User*. USA: Prentice Hall PTR, 1999.

[18] R. B. Cairns and B. D. Cairns, "Social cognition and social networks: A developmental perspective," *The development and treatment of childhood aggression*, pp. 249–278, 1991.

[19] S. Ahmed and E. Nozari, "On the linearizing effect of spatial averaging in large-scale populations of homogeneous nonlinear systems," in *2022 IEEE 61st Conference on Decision and Control (CDC)*, 2022, pp. 641–648.

[20] R. Durrett, *Probability: theory and examples*, 4.1 ed. Cambridge university press, April 2013, vol. 49.

[21] A. Gut and A. Gut, *Probability: a graduate course*. Springer, 2005, vol. 200, no. 5.

[22] N. Herrndorf, "A functional central limit theorem for ρ-mixing sequences," *Journal of Multivariate Analysis*, vol. 15, no. 1, pp. 141–146, 1984. [Online]. Available: https://www.sciencedirect.com/science/article\\/pii/0047259X84900733

[23] J. Jensen, L. Lake, P. Corbett, and D. Goggin, *Statistics for Petroleum Engineers and Geoscientists*. Prentice Hall, 1997.

[24] R. C. Bradley, "Basic properties of strong mixing conditions," in *Dependence in probability and statistics*. Springer, 1986, pp. 165–192.

[25] ——, "Basic properties of strong mixing conditions. a survey and some open questions," *Probability surveys*, vol. 2, pp. 107–144, 2005.

[26] I. Fazekas and O. Klesov, "A general approach to the strong law of large numbers," *Theory of Probability & Its Applications*, vol. 45, no. 3, pp. 436–449, 2001.

[27] D. W. Andrews, "Laws of large numbers for dependent non-identically distributed random variables," *Econometric theory*, vol. 4, no. 3, pp. 458–467, 1988.

[28] M. Ekström, "A general central limit theorem for strong mixing sequences," *Statistics & Probability Letters*, vol. 94, pp. 236–238, 2014.

[29] A. Oppenheim, J. Buck, M. Daniel, A. Willsky, S. Nawab, and A. Singer, *Signals & Systems*, ser. Prentice-Hall signal processing series. Prentice Hall, 1997.

[30] S. B. Hulley, S. R. Cummings, W. S. Browner, D. Grady, N. Hearst, and T. B. Newman, "Designing clinical research: an epidemiologic approach," in *Designing clinical research: an epidemiologic approach*, 2001, pp. 336–336.