A General Theory of Emergent Linearity in Complex Dynamical Systems: The Role of Spatial Averaging and Vanishing Correlations

Sabbir Ahmed Hafiz Fareed Ahmed Erfan Nozari

***Abstract*—Various natural and engineered systems, from urban traffic flow to the human brain, have been described by large-scale networked dynamical systems. Despite their vast differences, these systems are often similar in being comprised of numerous microscopic subsystems with complex nonlinear dynamics and interactions that give rise to diverse emergent macroscopic behaviors. As such, a long-standing question across various fields has been to understand why and how various forms of macroscopic behavior emerge from underlying microscopic dynamics. Motivated by a growing body of empirical observations, in this work we focus on linearity as one of the most fundamental aspects of system dynamics, and develop a general theoretical framework for the interplay between spatial averaging, decaying microscopic correlations, and emergent macroscopic linearity. Using and extending the theory of mixing sequences, we show that *in a broad class of autonomous nonlinear networked systems, the dynamics of the average of all subsystems' states becomes asymptotically linear as the number of subsystems grows to infinity, provided that (in addition to technical assumptions) pairwise correlations between subsystems decay to 0 as their pairwise distance grows to infinity*. We prove this result when the latter distance is between subsystems' linear indices or spatial locations, and provide extensions to linear time-invariant (LTI) limit dynamics, finite-sample analysis of rates of convergence, and networks of spatially-embedded subsystems with random locations. To our knowledge, this work is the first rigorous analysis of macroscopic linearity in large-scale heterogeneous networked systems, and provides a solid foundation for further theoretical and empirical analyses in various domains of science and engineering.**

I. Introduction

The spatial scale of analysis is a decisive factor in studying large-scale systems, from engineered systems to natural phenomena, with vast implications for modeling, system analysis, and control [2]–[7]. In many large-scale systems, individual microscopic subsystems exhibit complex, nonlinear dynamics that give rise, often in ways that we do not fully understand, to qualitatively distinct emergent dynamics at the macroscale [8]–[13]. Therefore, in principle, macroscopic dynamics can be studied in hyper-dimensional models with microscopic resolution [14], [15]. However, the sheer complexity of this approach often makes it infeasible, leading many researchers to use linear models *as approximations or local linearizations* for studying complex networks [16]–[18].

In a recent empirical study [19], on the other hand, we found linear models to be surprisingly effective–in fact, seemingly optimal–for capturing the macroscopic dynamics of large-scale brain networks across hundreds of subjects and different data modalities. Remarkably, this linearity appears to be in great contrast to the well-established theoretical and empirical nonlinearity of individual neuron dynamics [20], [21]. Nevertheless, this observation has since been replicated by others [22] as well as various simulations we have carried [1], and has remained begging for theoretical understanding and a rigorous mathematical framework that explains precisely why and when it holds. The goal of this work is to provide such a theoretical foundation for general stochastic nonlinear systems.

Literature Review: Our initial study [19] further aligns with other recent observations in large-scale biological [22]–[26] and artificial [27]–[29] neural networks. In the data-driven modeling of large-scale brain dynamics linked to electrical stimulation [23] or motor behavior [24], nonlinear models have repeatedly failed to outperform linear models, challenging the role and extent of nonlinearity in large-scale brain dynamics. More recently, it was found in [22] that nonlinear readouts can indeed improve the accuracy of predicting behavioral outcomes from neural recordings, but linear models are as accurate as nonlinear ones in predictive modeling of neural data itself. Similarly, in structural and functional brain scans, nonlinear models have performed on par with linear ones in tasks such as age or sex prediction [25], and correlations between visual stimuli and macroscopic brain recordings were observed to be linear in [26]. Recent studies also support linear models in resting-state fMRI analysis, showing strong performance in neuropsychiatric case-control comparisons [30], and that resting-state dynamics are explained well by stationary, linear properties of the data [31]. In the realm of artificial neural networks (ANNs) with nonlinear activation functions, the linearity of gradient descent dynamics concerning network parameters has been demonstrated in the limit of infinite network width [27], [28]. Earlier studies [29], [32] also established asymptotic linearity of ANNs with respect to their parameters. Collectively, this growing body of evidence strongly suggests that, counter-intuitively, the collective dynamics of millions to billions of highly-nonlinear microscopic elements can in fact be *less nonlinear* than each individual subsystem.

In [19], we demonstrated through simulations that spatial averaging can be a key factor in explaining emergent macroscopic linearity. However, this was only supported by simulations of two neuron models. In the preliminary version of this work [1], we extended our simulations to more general and complex forms of microscopic nonlinear dynamics, and presented the first mathematical proof supporting the conjecture that spatial averaging induces macroscopic linearity in nonlinear systems. However, this result was shown under the strong assumption of independent and identically-distributed

A preliminary version of this work appeared at the 61st IEEE Conference on Decision and Control [1].

Sabbir Ahmed is with the Computer Science Department at State University of New York (SUNY) Binghamton, sahmed9@binghamton.edu. Hafiz Fareed Ahmed is with the Department of Mechanical Engineering at the University of California, Riverside (UCR), hahme031@ucr.edu. Erfan Nozari is with the UCR's Departments of Mechanical Engineering, Electrical and Computer Engineering, and Bioengineering and the Neuroscience Graduate Program, erfan.nozari@ucr.edu.

(i.i.d.) subsystems, in particular ruling out all forms of network interactions. The present work extends the methodology of [1] to systems with heterogeneous subsystems and different forms of spatially correlated dynamics. Our approach leverages tools from probability theory [33]–[36], particularly the central limit theorem, and incorporates concepts from the theory of strong mixing sequences [37], which are critical for addressing the complexities introduced by spatial correlations.

Statement of Contributions: The main contribution of this work is the introduction of a general theoretical framework that explains the linearizing effects of spatial averaging in populations of nonlinear dynamical systems. This can be broken into four sets of technical contributions, as follows. *First*, we extend the well-established theory of mixing stochastic processes in a number of ways, including (i) generalizing the central limit theorem (CLT) for ρ-mixing sequences of random variables to multivariate processes, (ii) defining and theoretically characterizing the notion of a residual factor for ρ-mixing sequences, and (iii) extending the definition and properties of ρ-mixing sequences, including the CLT, to ρ^*-mixing for spatially-embedded sequences with fixed as well as random Euclidean locations. Building on these, *our second and main contribution* consists of the mathematical formulation and proof of the asymptotic linearity of averaged state dynamics under spatial averaging. We prove this result under two general settings–with and without spatial embeddings–and only mild technical restrictions on the form of microscopic nonlinearity, noise distributions, and network connectivity patterns. Notably, however, the limit dynamics are in general linear time-varying (LTV). *Our third contribution* then consists of extending the above results to scenarios where, under additional stationarity-related assumptions on the microscopic dynamics, the spatially-averaged limit dynamics are linear and time-invariant (LTI). Finally, *our fourth contribution* pertains to the finite-sample analysis of the rate at which averaged dynamics converge to linearity. In essence, our analysis shows that in the extreme case of independent subsystems (where averaging is most effective), convergence occurs at the parametric rate ($O(1/\sqrt{N})$), and the convergence slows down as the correlations between subsystems strengthen (with the opposite extreme of no convergence if all the subsystems are perfectly correlated). Notably, among these technical contributions the only one that existed in our preliminary version [1] is the third. The first and the fourth sets of contributions are completely novel, while our second contribution is now proven under significantly more general conditions–particularly, interconnected and heterogeneous systems–whereas its preliminary version was shown only for homogeneous and disconnected (i.i.d.) subsystems. To our knowledge, our results provide the first theoretical account of emergent linearity in complex dynamical systems, and open the door to a broad range of subsequent empirical and theoretical studies.

II. Notation

We use $\mathbb{R}$ to denote the set of real numbers. For a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, $\mathbf{A}^\dagger \in \mathbb{R}^{n\times m}$ denotes its pseudoinverse. 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. For two random variables ξ and η, $\mathrm{Var}(\xi)$ and $\rho(\xi,\eta)$ denote the variance of ξ and the Pearson correlation coefficient between ξ and η, respectively. For two random vectors $\boldsymbol{\xi} \in \mathbb{R}^n$ and $\boldsymbol{\eta} \in \mathbb{R}^m$, $\mathrm{Cov}(\boldsymbol{\xi},\boldsymbol{\eta}) \in \mathbb{R}^{n\times m}$ is the covariance between them, and $\mathrm{Cov}(\boldsymbol{\xi}) = \mathrm{Cov}(\boldsymbol{\xi},\boldsymbol{\xi})$. For sequences of random variables, we use both $\xrightarrow{d}$ and dlim to denote their convergence in distribution and $\xrightarrow{a.s.}$ to denote almost sure convergence.

III. Preliminaries: Mixing Sequences

In this section we review some fundamental notions and properties of mixing sequences which form the theoretical basis for the ensuing discussion. In essence, for a discrete-time stochastic process $\xi_i, i = 1, 2, \dots$ different notions of 'mixing' characterize the case when the statistical dependence between ξ_i and ξ_j diminishes as $|i-j|$ increases. As such, mixing conditions generalize the notion of a pairwise independent (a.k.a. white) sequence to one in which nearby elements can be dependent but their dependence decays as the distance between them grows. Various versions of mixing sequences have been 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 [38]).*** *Consider a sequence of random variables $\xi_1, \xi_2, \dots$ in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$ and define*

$$\varrho(n) = \sup_i \ \rho(\sigma(\xi_1,\dots,\xi_i), \sigma(\xi_{i+n}, \xi_{i+n+1}, \dots)), \tag{1}$$

where $\sigma(\cdot)$ denotes the smallest σ-algebra of Ω generated by a set of random variables, and, for any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$,

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

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

$$\varrho(n) \to 0 \quad as \quad n \to \infty. \qquad \square$$

Clearly, any i.i.d. sequence is ρ-mixing and the latter is a generalization of the former. Thus, many properties of i.i.d. sequences, such as the laws of large numbers [39], [40] and the central limit theorem (CLT) [35], [41], have been generalized to mixing sequences as long as the decay of dependence is sufficiently fast. Of particular relevance to this work is the CLT for ρ-mixing sequences, as presented below. But we first need a technical definition, as follows.

Definition III.2. ***(Slowly-varying function).*** *A function $h : \mathbb{N} \to \mathbb{R}$ is called* slowly varying *if*

$$\lim_{N\to\infty} \frac{h(kN)}{h(N)} = 1$$

for all $k \in \mathbb{N}$. □

Intuitively, if $h(N)$ is slowly varying it implies that the asymptotic behavior of $h(N)$ becomes insensitive to constant multiplicative factors, resulting in a very slow rate of growth/decay. Some examples of slowly-varying functions are constant functions $h(N) = c$, logarithm functions such as $h(N) = \log N$ or $h(N) = \log\log N$, and bounded functions with nonzero constant limits such as $h(x) = 1 + \frac{\sin(x)}{x}$.

We are now ready to present the following result, which is a generalization of the standard CLT for i.i.d. sequences [33, Thm 3.4.1] to ρ-mixing sequences of random variables.

Proposition III.3. ***(CLT for ρ-mixing sequences [35, Thm B]).*** *Consider a ρ-mixing sequence of random variables $\xi_1, \xi_2, \ldots$ and define its cumulative sum and cumulative variance as*

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

Assume

$$\mathbb{E}[\xi_i] = 0 \text{ and } \mathrm{Var}(\xi_i) < \infty \text{ for all } i, \tag{4a}$$

$$\sup_{M \geq 0, N \geq 1} \frac{1}{\sigma_N^2} \mathbb{E}[(S_{M+N} - S_M)^2] < \infty, \tag{4b}$$

$$h(N) = \frac{\sigma_N^2}{N} \text{ is a slowly-varying function.} \tag{4c}$$

Then

$$\frac{S_N}{\sigma_N} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad N \to \infty. \quad \square \tag{5}$$

It is instructive to compare Proposition III.3 with the standard CLT for i.i.d. sequences [33, Thm 3.4.1]. Both versions assume that all ξ_i have finite mean and variance; they slightly differ in that Proposition III.3 assumes ξ_i's are already mean-subtracted while the standard version mean-subtracts the sum. Condition (4b) asks that the ratio of the variance of shifted partial sums to that of initial partial sums remains bounded. This is a generalization of the 'identically distributed' assumption in the standard version, in which case the left hand side of (4b) is always 1. Finally, ρ-mixingness is a generalization of independence, as noted earlier. Importantly, however, condition (4c) ensures that the sequence $\{\xi_i\}$ is not any ρ-mixing sequence, but one in which the correlations between neighboring elements decays sufficiently fast. This is a subtle point and we will get back to it later in Section IV when we define residual factors. For now, note the normalization by σ_N in (5) 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 can vary, whereas σ_N^2 always grows as N for i.i.d. sequences.

In what follows, we will extend Proposition III.3 in a number of ways, and use the results to prove the Gaussianity of joint state-noise distributions under spatial averaging, using which we will prove the linearity of spatially-averaged dynamics.

IV. Linearizing Effect of Spatial Averaging on Sequences of Dynamical Systems

A. Problem Formulation

Consider a heterogeneous population of N dynamical subsystems, where each has the general discrete-time nonlinear form

$$\begin{aligned} \mathbf{x}_i(t+1) &= f_i(\mathbf{x}_{\mathcal{N}_i}(t), \mathbf{w}_i(t)), \qquad i = 1, 2, \ldots N, \\ \mathbf{x}(0) &= \begin{bmatrix} \mathbf{x}_1(0)^T & \cdots & \mathbf{x}_N(0)^T \end{bmatrix}^T \sim p_0, \\ \mathbf{w}(t) &= \begin{bmatrix} \mathbf{w}_1(t)^T & \cdots & \mathbf{w}_N(t)^T \end{bmatrix}^T \sim p_w(t). \end{aligned} \tag{6}$$

$\mathbf{x}_i(t) \in \mathbb{R}^n$ is the state of subsystem i with initial joint distribution p_0, and $\mathbf{w}_i(t) \in \mathbb{R}^m$ is the noise process of subsystem i with the joint distribution $p_w(t)$. The vector

$$\mathbf{x}_{\mathcal{N}_i}(t) = \begin{bmatrix} \mathbf{x}_{\max\{1, i-\tau\}}^T & \cdots & \mathbf{x}_i^T & \cdots & \mathbf{x}_{\min\{i+\tau, N\}}^T \end{bmatrix}^T \tag{7}$$

denotes the state of all subsystems on which f_i *can* depend (i.e., a superset of the in-neighbor set of node i). $\tau < \infty$ can be arbitrarily large, but finite, and controls the spatial range of neighborhoods, such that subsystems i and j can (but do not need to) be neighbors if $|i - j| \leq \tau$.

The problem we tackle in this section, motivated by our prior empirical observations [19] and those of others [24], [25], [42], is as follows.

Problem 1. ***(Linearizing Effect of Spatial Averaging on Sequences of Dynamical Systems).*** *Consider a heterogeneous population of nonlinear dynamical systems described by* (6)*, and define the population's average state vector as*

$$\bar{\mathbf{x}}(t) = \frac{1}{\phi(N)} \sum_{i=1}^{N} \mathbf{x}_i(t) - \mathbb{E}[\mathbf{x}_i(t)], \tag{8}$$

where $\phi(N)$ is a normalization factor. Prove, under appropriate assumptions and choice of $\phi(N)$, that the dynamics of $\bar{\mathbf{x}}(t)$ becomes asymptotically linear as $N \to \infty$. □

The normalization factor $\phi(N)$ plays the same role that σ_N plays in (5), but it is simpler as it only captures the *growth rate* of σ_N and not σ_N itself. For comparison, in the standard CLT $\phi(N) = \sqrt{N}$ for all i.i.d. sequences while σ_N varies from one sequence to another. As we move to multivariate processes (see below) this distinction will become bolder. A naive extension of (5) would require computing the exact covariance matrix of S_N, which often lacks a closed-form solution, as well as normalizing S_N by $\mathrm{Cov}(S_N)^{-\frac{1}{2}}$, which would inter-mix different dimensions and lose the notion of an 'average'.

In the rest of this section we will first lay the statistical foundation of our framework in Section IV-B, where we extend and further characterize the properties of ρ-mixing sequences. In Section IV-C we then use this foundation to present our main result of this section, i.e., the solution to Problem 1. Under additional stationarity assumptions, we further strengthen this result to prove linear *time-invariant* (LTI) limit dynamics, and finally characterize the rate of this convergence to linearity in Section IV-E for the special case of i.i.d. sequences.

B. Multivariate ρ-Mixing Sequences

In this subsection we extend Definition III.1 to vector-valued stochastic processes, and then prove some of their properties, including a more practical CLT. These results will play a central role in our proof of the linearizing effect of spatial averaging in Theorem IV.8.

The following is a natural generalization of Definition III.1.

Definition IV.1. ***(Multivariate ρ-mixing sequences).*** *Consider a sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots \in \mathbb{R}^q$ in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, and define*

$$\varrho(n) = \sup_j \ \rho(\sigma(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_j), \sigma(\boldsymbol{\xi}_{j+n}, \dots)), \tag{9}$$

where$\rho(\mathcal{A}, \mathcal{B})$ for two σ-algebras $\mathcal{A}$ and $\mathcal{B}$ is defined in (2)*. The sequence $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots$ is called ρ-mixing if*

$$\varrho(n) \to 0 \quad as \quad n \to \infty. \tag{10}$$

Before we can present the generalized CLT for multivariate ρ-mixing sequences we also need the following definition. This will help, in particular, in determining the appropriate normalization factor in Problem 1.

Definition IV.2. ***(Residual factor).*** *Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots$ be a ρ-mixing sequence and $S_N = \sum_{i=1}^N \boldsymbol{\xi}_i$. The function $h(N)$ is called a* residual factor *for $\{\boldsymbol{\xi}_i\}_{i=1}^\infty$ if*

$$\lim_{N\to\infty} \frac{1}{Nh(N)} \mathrm{Cov}(S_N) < \infty, \tag{11}$$

i.e., the limit exists and is finite. □

To put this definition in perspective, if $\{\boldsymbol{\xi}_i\}_{i=1}^\infty$ is i.i.d. then $\mathrm{Cov}(S_N) = N\mathrm{Cov}(\boldsymbol{\xi}_1)$ and so $h(N) = 1$ is a residual factor. Note, also, that residual factors are not unique. In the i.i.d. case, $h(N) = c$ is also a residual factor for any constant c. Further, note that Definition IV.2 does not require the limit to be nonzero. Thus, $h(N) = N$, $\log(N)$, and e^N are all valid residual factors for an i.i.d. sequence. Finally, and related to the last point, a residual factor always exists for any ρ-mixing sequence. The following Theorem formalizes this existence.

Theorem IV.3. ***(Existence of residual factors).*** *Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots \in \mathbb{R}^q$ be a ρ-mixing sequence and assume that $\mathrm{Var}((\boldsymbol{\xi}_i)_\ell) \le \bar{\sigma}^2$ for all i, ℓ and some $\bar{\sigma} < \infty$. Then, $h(N) = N$ is a residual factor for $\{\boldsymbol{\xi}_i\}_{i=1}^\infty$.*

Proof. Let $S_N = \sum_{i=1}^N \boldsymbol{\xi}_i$ and let $\varrho(n)$ be as in (9). Then, for any $\ell = 1, \dots, q$,

$$\begin{aligned}\mathrm{Var}((S_N)_\ell) &= \sum_{i=1}^N \mathrm{Var}((\boldsymbol{\xi}_i)_\ell) + 2 \sum_{1\le i<j\le N} \mathrm{Cov}((\boldsymbol{\xi}_i)_\ell, (\boldsymbol{\xi}_j)_\ell) \\ &\le N\bar{\sigma}^2 + 2\bar{\sigma}^2 \sum_{1\le i<j\le N} \varrho(|i-j|). \end{aligned} \tag{12}$$

Changing the order of summation, we have:

$$\begin{aligned}\sum_{1\le i<j\le N} \varrho(|i-j|) &= \sum_{k=1}^{N-1} \sum_{i=1}^{N-k} \varrho(k) = \sum_{k=1}^{N-1} (N-k)\varrho(k) \\ &\le N \sum_{k=1}^N \varrho(k). \end{aligned} \tag{13}$$

Assume, without loss of generality, that $\varrho(x)$ is defined and infinitely differentiable for all real-valued $x \in \mathbb{R}_{>0}$. This is without loss of generality since $\varrho(n)$ can always be replaced by an upper bound that satisfies this assumption and still decays to 0 as $n \to \infty$. Then, using the Euler–Maclaurin formula [43],

$$\sum_{k=1}^N \varrho(k) = O\Big(\int_1^N \varrho(x)dx\Big) \tag{14}$$

Combining (12)-(14), we get

$$\lim_{N\to\infty} \frac{1}{N^2} \mathrm{Var}((S_N)_\ell) \le \lim_{N\to\infty} \frac{N\bar{\sigma}^2}{N^2} + \frac{2\bar{\sigma}^2 C}{N} \int_1^N \varrho(x)dx \tag{15}$$

where C is a bounding constant from (14). Limit of the first term is clearly 0. The second limit has two possibilities. If the non-negative quantity

$$\int_1^\infty \varrho(x)dx \tag{16}$$

is finite ($\varrho(n)$ decays fast), then the second limit on the right hand side of (15) is also clearly 0. If, on the other hand, the limit in (16) is infinite ($\varrho(n)$ decays slowly), then by the L'Hopital's rule and (10), still

$$\lim_{N\to\infty} \frac{\int_1^N \varrho(x)dx}{N} = \lim_{N\to\infty} \frac{\varrho(N)}{1} = 0.$$

Therefore, either way, the limits in (15) are all 0. Since this is true for all $\ell = 1, \dots, q$, we also have

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{N^2} |\mathrm{Cov}((S_N)_\ell, (S_N)_m)| \\ &\qquad \le \lim_{N\to\infty} \frac{\sqrt{\mathrm{Var}((S_N)_\ell)}}{N} \frac{\sqrt{\mathrm{Var}((S_N)_m)}}{N} = 0,\end{aligned}$$

for all $\ell, m = 1, \dots, q$. Put together, $\lim_{N\to\infty} \frac{1}{N^2} \mathrm{Cov}(S_N) = 0$, completing the proof. ■

A corollary to Theorem IV.3 is that any $h(N) \ge N$ is also a valid residual factor for any ρ-mixing sequence with uniformly-bounded variance. However, residual factors are practically useful only when they grow at the same rate as $\mathrm{Cov}(S_N)$ itself and the limit in (11) is nonzero. This "smallest" residual factor is what we will use in the sequel, even though the forthcoming results will still be correct, but not necessarily useful, for all valid residual factors.

Going back to Proposition III.3, note also the similarity between (4c) and (11), where $h(N)$ in (4c) is clearly a residual factor for $\{\xi_k\}_{i=1}^\infty$ in Proposition III.3. As we will see later, whether a residual factor of a ρ-mixing sequence is slowly-varying remains to be a critical condition for validity of CLT (and in turn asymptotic linearity) for that sequence.

To better clarify the notion of a residual factor and how it depends on the statistics of $\{\boldsymbol{\xi}_i\}_{i=1}^\infty$, in the following result we provide closed-form expressions for "smallest" residual factors of a few representative ρ-mixing sequences.

Theorem IV.4. ***(Relationship between decay r of correlations and growth of residual factors).*** *Let $\xi_1, \xi_2, \dots \in \mathbb{R}$ be a ρ-mixing sequence of random variables, and let $\rho(n)$ be as in* (9)*. Assume, for simplicity, that $\mathrm{Var}(\xi_i) = \sigma^2$ for all i and*

$\rho(\xi_i, \xi_j) = \varrho(|i-j|)$. *Then the following relationships exist, where $h(N)$ is a residual factor in each case:*

(i) if $\varrho(n) = \frac{1}{n}$ then $h(N) = \log(N)$;
(ii) if $\varrho(n) = \frac{1}{\log(1+n)}$ then $h(N) = \frac{N}{\log(N)}$;
(iii) if $\varrho(n) = \frac{1}{n^p}, p \in (0,1)$, then $h(N) = N^{1-p}$.

Proof. In all cases, let $S_N = \sum_{i=1}^N \xi_i$.

(i) Similar to (12) and (13), we have

$$\mathrm{Var}(S_N) = N\sigma^2 + 2\sigma^2 \sum_{1\le i<j\le N} \frac{1}{|i-j|},$$

$$\sum_{1\le i<j\le N} \frac{1}{|i-j|} = \sum_{k=1}^{N-1} \frac{N-k}{k} = N\sum_{k=1}^{N-1}\frac{1}{k} - (N-1).$$

For large N, the harmonic sum $\sum_{k=1}^{N-1}\frac{1}{k}$ is approximately $\log(N)+\gamma$, where γ is the Euler's constant. Thus, for large N,

$$\sum_{1\le i<j\le N} \frac{1}{|i-j|} \simeq N(\log(N)+\gamma) - (N-1)$$
$$= N\log(N) + \text{lower order terms},$$

and

$$\mathrm{Var}(S_N) \simeq N\sigma^2 + 2\sigma^2 N\log(N) + \text{lower order terms}$$
$$= 2\sigma^2 N\log(N) + \text{lower order terms}.$$

Therefore, for large N,

$$\frac{\mathrm{Var}(S_N)}{N} \simeq 2\sigma^2\log(N) + \text{lower order terms}$$

and so $h(N) = \log(N)$ is a residual factor. This is also a slowly-varying function, since for $k \in \mathbb{N}$,

$$\lim_{N\to\infty}\frac{h(kN)}{h(N)} = \lim_{N\to\infty}\frac{2\sigma^2\log(kN)}{2\sigma^2\log(N)} = 1.$$

In other words, the correlation decay rate of $\frac{1}{|i-j|}$ is "sufficiently fast" so that $\mathrm{Var}(S_N)$ grows sufficiently slowly and $h(N)$ becomes slowly varying. As we will see in Theorem IV.6, this allows for CLT to hold for this sequence.

(ii) Similar to case (i),

$$\mathrm{Var}(S_N) = N\sigma^2 + 2\sigma^2 \sum_{1\le i<j\le N} \frac{1}{\log(1+|i-j|)}$$

and

$$\sum_{1\le i<j\le N} \frac{1}{\log(1+|i-j|)} = \sum_{k=1}^{N-1}\frac{N-k}{\log(1+k)} = \sum_{k=2}^{N}\frac{N+1-k}{\log(k)}$$
$$= (N+1)\sum_{k=2}^{N}\frac{1}{\log(k)} - \sum_{k=2}^{N}\frac{k}{\log(k)}. \tag{17}$$

For large N, this can be approximated (with the same growth rate) using the Euler-Maclaurin formula [43] as

$$(N+1)\int_2^N \frac{1}{\log(x)}dx - \int_2^N \frac{x}{\log(x)}dx. \tag{18}$$

By definition, $\int_2^N \frac{1}{\log(x)}dx = \mathrm{li}(N) - \mathrm{li}(2)$ where $\mathrm{li}(u) = \int_0^u \frac{dx}{\log(x)}$ is the logarithmic integral function. To compute the second integral, we can use the change of variables $u = \log(x)$ and $v = 2u$ so that

$$\int_2^N \frac{x}{\log(x)}dx = \int_{\log(2)}^{\log(N)} \frac{e^u}{u}e^u du = \int_{2\log(2)}^{2\log(N)} \frac{e^v}{v}dv$$
$$= \mathrm{Ei}(\log(N^2)) - \mathrm{Ei}(\log(4))$$

where $\mathrm{Ei}(u) = \int_{-\infty}^u \frac{e^x}{x}dx$ is the exponential integral function. Since $\mathrm{li}(u) = \mathrm{Ei}(\log(u))$, (18) further simplifies to

$$N\mathrm{li}(N) - \mathrm{li}(N^2) + \mathrm{li}(N) - (N+1)\mathrm{li}(2) + \mathrm{li}(4) \tag{19}$$

For large N,

$$\mathrm{li}(N) = \frac{N}{\log(N)}\Big[1 + O\big(\frac{1}{\log(N)}\big)\Big] \simeq \frac{N}{\log(N)}$$

Therefore, (19) is approximately equal to

$$\frac{N^2}{\log(N)} - \frac{N^2}{\log(N^2)} + \frac{N}{\log(N)} - (N+1)\mathrm{li}(2) + \mathrm{li}(4)$$
$$= \frac{N^2}{2\log(N)} + \text{lower order terms}.$$

Combining these results, for large N we get

$$\frac{\mathrm{Var}(S_N)}{N} \simeq \sigma^2 + 2\sigma^2\frac{N}{2\log(N)} + \text{lower order terms}$$
$$\simeq \frac{\sigma^2 N}{\log(N)} + \text{lower order terms}.$$

and so $h(N) = \frac{N}{\log(N)}$ is a residual factor. This is not a slowly-varying function, however, since for $k \in \mathbb{N}$,

$$\lim_{N\to\infty}\frac{h(kN)}{h(N)} = \lim_{N\to\infty}\frac{\frac{\sigma^2 kN}{\log(kN)}}{\frac{\sigma^2 N}{\log(N)}} = k.$$

Therefore in this case, unlike case (i), correlations decay too slowly. As a result, $\mathrm{Var}(S_N)$ grow too fast and $h(N)$ is not slowly-varying. As we will see in Theorem IV.6, this prevents CLT to hold for this sequence, despite being ρ-mixing.

(iii) This case falls in between cases (i) and (ii) in terms of the decay rate of $\rho(n)$. Proceeding as before,

$$\mathrm{Var}(S_N) = N\sigma^2 + 2\sigma^2 \sum_{1\le i<j\le N} \frac{1}{|i-j|^p}$$

and

$$\sum_{1\le i<j\le N} \frac{1}{|i-j|^p} = \sum_{k=1}^{N-1}\frac{N-k}{k^p} = N\sum_{k=1}^{N-1}k^{-p} - \sum_{k=1}^{N-1}k^{1-p}$$

Using the Euler-Maclaurin formula, we get for large N,

$$\sum_{k=1}^{N-1}k^{-p} \simeq \int_1^{N-1} x^{-p}dx = \frac{(N-1)^{1-p}-1}{1-p}$$
$$\sum_{k=1}^{N-1}k^{1-p} \simeq \int_1^{N-1} x^{1-p}dx = \frac{(N-1)^{2-p}-1}{2-p}$$

Thus, for large N,

$$\sum_{1\le i<j\le N} \frac{1}{|i-j|^p} \simeq \frac{N(N-1)^{1-p}}{1-p} - \frac{(N-1)^{2-p}}{2-p}$$
$$= \frac{(N-1)^{2-p}}{(1-p)(2-p)} + \frac{(N-1)^{1-p}}{1-p}$$

and so

$$\frac{\mathrm{Var}(S_N)}{N} \simeq \sigma^2 + \frac{2\sigma^2}{N}\Big[\frac{(N-1)^{2-p}}{(1-p)(2-p)} + \frac{(N-1)^{1-p}}{1-p}\Big]$$
$$= \frac{2\sigma^2}{(1-p)(2-p)}\frac{(N-1)^{2-p}}{N} + \text{lower order terms.}$$

Therefore, $h(N) = \frac{N^{2-p}}{N} = N^{1-p}$ is a residual factor. This $h(N)$ is not a slowly-varying function, since for $k \in \mathbb{N}$,

$$\lim_{N\to\infty}\frac{h(kN)}{h(N)} = \lim_{N\to\infty}\frac{(kN)^{1-p}}{N^{1-p}} = k^{1-p}.$$

In other words, even the polynomial decay rate of $\frac{1}{|i-j|^p}, p \in (0,1)$ is too slow for $h(N)$ to be slowly varying and for CLT to hold. ■

In summary, it follows from Theorem IV.4 that $\rho(n) \propto \frac{1}{n}$ is the rough "boundary" between decay rates that are sufficiently fast and those that are too slow.

Before we are ready to present the generalized CLT for multivariate ρ-mixing sequences, we need to prove another property of ρ-mixing sequences, namely, that sequences formed through transformations of ρ-mixing sequences are also ρ-mixing.

Lemma IV.5. ***(Sequence formed through transformation of ρ-mixing sequence is ρ-mixing).*** *Consider a sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots \in \mathbb{R}^q$ in a probability space $(\Omega, \mathcal{F}, \mathcal{P})$. Let $\tau, k \geq 0$ be arbitrary fixed integers and assume that for any $i \geq 1$, h_i is a measurable function from $\mathbb{R}^{q\min\{\tau+i, 2\tau+1\}}$ to $\mathbb{R}^k$. Define*

$$\boldsymbol{\zeta}_i = h_i(\boldsymbol{\xi}_{\min\{1,i-\tau\}}, \dots, \boldsymbol{\xi}_i, \dots, \boldsymbol{\xi}_{i+\tau}), \qquad i \geq 1.$$

If the sequence $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots$ is ρ-mixing, then so is $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \dots$.

Proof. Without loss of generality let $n \geq \tau + 1$. Then, by the definition of the σ-algebra generated by a collection of random variables and the measurability of all h_i, we get

$$\sigma(\boldsymbol{\zeta}_1, \dots, \boldsymbol{\zeta}_j) \subseteq \sigma(\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_{j+\tau}),$$

and

$$\sigma(\boldsymbol{\zeta}_{j+n}, \dots) \subseteq \sigma(\boldsymbol{\xi}_{j+n-\tau}, \dots),$$

for all $j \geq 1$. Therefore, by (2),

$$0 \leq \rho_\zeta(n) \leq \rho_\xi(n - 2\tau), \qquad n \geq \tau + 1. \tag{20}$$

where $\rho_\zeta(n)$ and $\rho_\xi(n)$ are defined as in (1) for the respective sequences $\{\boldsymbol{\zeta}_i\}$ and $\{\boldsymbol{\xi}_i\}$. The theorem then follows by letting $n \to \infty$ in (20) and using the mixing property of $\{\boldsymbol{\xi}_i\}$. ■

The next result extends and improves the CLT in Proposition III.3 to the case of multivariate ρ-mixing sequences.

Theorem IV.6. ***(Multivariate CLT for ρ-mixing sequence).*** *Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \dots \in \mathbb{R}^q$ be a ρ-mixing sequence where each $\boldsymbol{\xi}_i$ has zero mean and finite variance. Let $h(N)$ be a residual factor for $\{\boldsymbol{\xi}_i\}_{i=1}^\infty$ and assume that it is slowly-varying. Also, assume that for all $\boldsymbol{\theta} \neq 0$,*

$$\sup_{M,N}\frac{\boldsymbol{\theta}^T\mathrm{Cov}\Big(\sum_{i=M+1}^{M+N}\boldsymbol{\xi}_i\Big)\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathrm{Cov}\Big(\sum_{i=1}^{N}\boldsymbol{\xi}_i\Big)\boldsymbol{\theta}} < \infty. \tag{21}$$

Then, the "mean" variable

$$\bar{\boldsymbol{\xi}} = \frac{1}{\sqrt{Nh(N)}}\sum_{i=1}^N \boldsymbol{\xi}_i, \tag{22}$$

satisfies

$$\bar{\boldsymbol{\xi}} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}^*_{\bar{\xi}}) \quad \textit{as} \quad N \to \infty, \tag{23}$$

where (cf. (11)*)*

$$\boldsymbol{\Sigma}^*_{\bar{\xi}} = \lim_{N\to\infty}\frac{1}{Nh(N)}\mathrm{Cov}\Big(\sum_{i=1}^N \boldsymbol{\xi}_i\Big). \tag{24}$$

Proof. Let $\boldsymbol{\theta} \neq \mathbf{0}$ be an arbitrary constant vector. By Lemma IV.5, the sequence of random variables $\{\boldsymbol{\theta}^T\boldsymbol{\xi}_i\}_{i\geq 1}$ is ρ-mixing. Define the cumulative sum and cumulative variance of this sequence as

$$S_N = \sum_{i=1}^N \boldsymbol{\theta}^T\boldsymbol{\xi}_i, \qquad \sigma_N^2 = \mathrm{Var}(S_N). \tag{25}$$

The sequence $\{\boldsymbol{\theta}^T\boldsymbol{\xi}_i\}_{i\geq 1}$ follows (4a) by assumption. It also follows (4b) due to (21). We need to show that it also satisfies (4c) before we can use the CLT in Proposition III.3.

Note that

$$\sigma_N^2 = \boldsymbol{\theta}^T\mathrm{Cov}\Big(\sum_{i=1}^N \boldsymbol{\xi}_i\Big)\boldsymbol{\theta}$$

so we get

$$\lim_{N\to\infty}\frac{\sigma_N^2}{Nh(N)} = \boldsymbol{\theta}^T\boldsymbol{\Sigma}^*_{\bar{\xi}}\boldsymbol{\theta}. \tag{26}$$

First consider the case where $\boldsymbol{\theta}^T\boldsymbol{\Sigma}^*_{\bar{\xi}}\boldsymbol{\theta} \neq 0$. Using the fact that $h(N)$ is slowly varying, we have, for any $k \in \mathbb{N}$,

$$\lim_{N\to\infty}\frac{\frac{\sigma_{kN}^2}{kNh(kN)}h(kN)}{\frac{\sigma_N^2}{Nh(N)}h(N)} = \frac{\lim_{N\to\infty}\frac{\sigma_{kN}^2}{kNh(kN)}}{\lim_{N\to\infty}\frac{\sigma_N^2}{Nh(N)}} \cdot \lim_{N\to\infty}\frac{h(kN)}{h(N)}$$
$$= \frac{\boldsymbol{\theta}^T\boldsymbol{\Sigma}^*_{\bar{\xi}}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\boldsymbol{\Sigma}^*_{\bar{\xi}}\boldsymbol{\theta}} \cdot 1 = 1.$$

This proves that

$$\tilde{h}(N) = \frac{\sigma_N^2}{N} \tag{27}$$

is slowly varying, and thus σ_N^2 satisfies (4c). Therefore, Proposition III.3 holds and, from (5) and (26), as $N \to \infty$

$$\frac{S_N}{\sqrt{Nh(N)}} = \frac{S_N}{\sigma_N}\cdot\frac{\sigma_N}{\sqrt{Nh(N)}} \xrightarrow{d} \mathcal{N}(0,1)\cdot\sqrt{\boldsymbol{\theta}^T\boldsymbol{\Sigma}^*_{\bar{\xi}}\boldsymbol{\theta}}$$
$$= \mathcal{N}(0, \boldsymbol{\theta}^T\boldsymbol{\Sigma}^*_{\bar{\xi}}\boldsymbol{\theta}) \tag{28}$$

This means that

$$\boldsymbol{\theta}^T\bar{\boldsymbol{\xi}} \xrightarrow{d} \boldsymbol{\theta}^T\bar{\boldsymbol{\xi}}^*, \quad \text{as} \quad N \to \infty, \tag{29}$$

where $\bar{\boldsymbol{\xi}}^*$ is a multivariate random variable distributed as $\mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}^*_{\bar{\boldsymbol{\xi}}}\right)$.

On the other hand, if $\boldsymbol{\theta}^T \boldsymbol{\Sigma}_{\bar{\xi}}^* \boldsymbol{\theta} = 0$ (including when $\boldsymbol{\theta} = \mathbf{0}$), then (29) follows immediately without the need to use any CLT since $\boldsymbol{\theta}^T \bar{\boldsymbol{\xi}}^* = 0$ and

$$\lim_{N\to\infty} \mathrm{Var}(\boldsymbol{\theta}^T \bar{\boldsymbol{\xi}}) = \lim_{N\to\infty} \frac{\sigma_N^2}{Nh(N)} \overset{(26)}{=} \boldsymbol{\theta}^T \boldsymbol{\Sigma}_{\bar{\xi}}^* \boldsymbol{\theta} = 0.$$

Therefore, (29) holds for all $\boldsymbol{\theta} \in \mathbb{R}^q$. According to Cramer-Wold device theorem [33, Thm 3.9.5], it then follows that

$$\bar{\boldsymbol{\xi}} \xrightarrow{d} \bar{\boldsymbol{\xi}}^* \quad \text{as} \quad N \to \infty, \tag{30}$$

completing the proof. ■

Theorem IV.6 provides the foundation for the asymptotic analysis of spatially averaged dynamics formulated in Problem 1, which we undertake next.

C. Asymptotic Linearity Under Spatial Averaging

In this section we use the results in Section IV-B to tackle Problem 1. Throughout this section, we make the following assumptions on the population dynamics (6). For ease of notation, for all i, t let

$$\mathbf{y}_i(t) = \begin{bmatrix} \mathbf{x}_i(t)^T & \mathbf{w}_i(t)^T \end{bmatrix}^T, \tag{31}$$

and

$$\mathbf{z}_i(t) = \begin{bmatrix} \mathbf{x}_i(t+1)^T & \mathbf{x}_i(t)^T & \mathbf{w}_i(t)^T \end{bmatrix}^T. \tag{32}$$

Assumption IV.7. ***(Standing assumptions).***

(A1) *The noise processes are zero mean and finite-variance, i.e., there exists $C < \infty$ such that for all i and t,*

$$\mathbb{E}[\mathbf{w}_i(t)] = \mathbf{0} \quad \textit{and} \quad \left\|\mathbb{E}[\mathbf{w}_i(t)\mathbf{w}_i(t)^T]\right\| \le C.$$

(A2) *For all $\boldsymbol{\theta} \neq 0$ and all t,*

$$\sup_{M,N} \frac{\boldsymbol{\theta}^T \mathrm{Cov}\Big(\sum_{i=M+1}^{M+N} \mathbf{z}_i(t)\Big)\boldsymbol{\theta}}{\boldsymbol{\theta}^T \mathrm{Cov}\Big(\sum_{i=1}^{N} \mathbf{z}_i(t)\Big)\boldsymbol{\theta}} < \infty. \tag{33}$$

(A3) *For each t, all noises $\{\mathbf{w}_i(t)\}_{i=1}^\infty$ are independent of contemporaneous states $\{\mathbf{x}_i(t)\}_{i=1}^\infty$ at the same time t.*

(A4) *The functions $f_i(.)$ are uniformly bounded, i.e., there exists $M < \infty$ such that for all $i, \mathbf{x}_{\mathcal{N}_i}, \mathbf{w}_i$*

$$||f_i(\mathbf{x}_{\mathcal{N}_i}, \mathbf{w}_i)|| \le M.$$

(A5) *For all t, the sequence of random vectors $\{\mathbf{y}_i(t)\}_{i=1}^\infty$ in* (31) *is ρ-mixing with a slowly-varying residual factor $h(N)$.*

The key assumption in Assumption IV.7 is the ρ-mixing assumption in (A5) which, in essence, prevents global synchrony. In other words, it prevents a case where all $\{\mathbf{y}_i(t)\}_{i=1}^\infty$ are strongly correlated with each other, in which case averaging does not do much—the average of $\{\mathbf{y}_i(t)\}_{i=1}^\infty$ becomes very similar to any one of them.

The remaining 4 assumptions are technical and automatically satisfied in most practical applications. Assumptions (A1) and (A2) are parallel to (4a) and (4b) in Proposition III.3 and described thereafter. Next, assumption (A3) asks for lack of instantaneous effect, where each noise variable takes at least one time step before affecting the state. Finally, note that Assumption (A4) is practically equivalent to uniform boundedness of *state-noise trajectories*. If $(\mathbf{x}_{\mathcal{N}_i}, \mathbf{w}_i)$ are guaranteed to remain within a compact region $\mathcal{D}$, as is the case for any real-world state/noise variable, then f_i needs to be bounded only on $\mathcal{D}$ and its value can be replaced by an arbitrary finite value (say, 0) outside of $\mathcal{D}$ without affecting the solutions of (6). This is true, e.g., if f_i is any analytic function on $\mathcal{D}$.

With the stated assumptions above, we are ready to characterize the asymptotic spatial aggregate dynamics of the population of dynamical systems in (6), as follows.

Theorem IV.8. ***(Linearizing effect of spatial averaging on sequences of dynamical systems).*** *Consider the population dynamics* (6)*, and assume that Assumptions* (A1)-(A5) *hold. Define the population average variables*

$$\bar{\mathbf{x}}(t) = \frac{1}{\sqrt{Nh(N)}} \sum_{i=1}^N \mathbf{x}_i(t) - \mathbb{E}[\mathbf{x}_i(t)], \tag{34}$$
$$\bar{\mathbf{w}}(t) = \frac{1}{\sqrt{Nh(N)}} \sum_{i=1}^N \mathbf{w}_i(t),$$

and $\bar{\mathbf{y}}(t) = \begin{bmatrix} \bar{\mathbf{x}}(t)^T & \bar{\mathbf{w}}(t)^T \end{bmatrix}^T$ and assume that $\lim_{N\to\infty} \mathrm{Cov}\big(\bar{\mathbf{x}}(t+1), \bar{\mathbf{y}}(t)\big)$ exists. Then the relationship between $\bar{\mathbf{x}}(t+1)$, $\bar{\mathbf{x}}(t)$ and $\bar{\mathbf{w}}(t)$ becomes asymptotically linear as $N \to \infty$, i.e., for all $\boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\omega} \in \mathbb{R}^m$,

$$\mathbb{E}\left[\bar{\mathbf{x}}(t+1) \middle| \begin{bmatrix} \bar{\mathbf{x}}(t) \\ \bar{\mathbf{w}}(t) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\xi} \\ \boldsymbol{\omega} \end{bmatrix}\right] \to \mathbf{A}_\infty(t)\boldsymbol{\xi} + \mathbf{B}_\infty(t)\boldsymbol{\omega} \quad \textit{as } N \to \infty, \tag{35}$$

where

$$\begin{aligned}
\mathbf{A}_\infty(t) &= \mathrm{Cov}\big(\bar{\mathbf{x}}_\infty(t+1), \bar{\mathbf{x}}_\infty(t)\big)\mathrm{Cov}\big(\bar{\mathbf{x}}_\infty(t)\big)^\dagger \\
\mathbf{B}_\infty(t) &= \mathrm{Cov}\big(\bar{\mathbf{x}}_\infty(t+1), \bar{\mathbf{w}}_\infty(t)\big)\mathrm{Cov}\big(\bar{\mathbf{w}}_\infty(t)\big)^\dagger \\
\bar{\mathbf{x}}_\infty(t) &= \operatorname*{dlim}_{N\to\infty} \bar{\mathbf{x}}(t) \\
\bar{\mathbf{w}}_\infty(t) &= \operatorname*{dlim}_{N\to\infty} \bar{\mathbf{w}}(t).
\end{aligned} \tag{36}$$

Proof. Fix t, and let $\mathbf{z}_i(t)$ be as in (32). From (6), Assumption (A5), and Lemma IV.5, the sequence $\{\mathbf{z}_i(t)\}_{i=1}^\infty$ is also ρ-mixing. Since mean-centering preserves correlation coefficients, $\{\mathbf{z}_i(t) - \mathbb{E}[\mathbf{z}_i(t)]\}_{i=1}^\infty$ is ρ-mixing as well. The latter sequence is mean-zero by construction and finite-variance by Assumptions (A4) and (A1). Furthermore, $h(N)$ is a residual factor for both $\{\mathbf{x}_i(t+1)\}_{i=1}^\infty$ and $\{\begin{bmatrix} \mathbf{x}_i(t)^T & \mathbf{w}_i(t)^T \end{bmatrix}^T\}_{i=1}^\infty$, so both of the limits

$$\lim_{N\to\infty} \mathrm{Cov}\big(\bar{\mathbf{x}}(t+1)\big) \quad \text{and} \quad \lim_{N\to\infty} \mathrm{Cov}\big(\bar{\mathbf{y}}(t)\big)$$

exist. Since the limit of the covariance between $\bar{\mathbf{x}}(t+1)$ and $\bar{\mathbf{y}}(t)$ also exists by assumption, $h(N)$ is a (slowly-varying) residual factor for $\{\mathbf{z}_i(t) - \mathbb{E}[\mathbf{z}_i(t)]\}_{i=1}^\infty$. These, together with Assumption (A2), satisfy all the requirements of Theorem IV.6. Therefore,

$$\bar{\mathbf{z}}(t) = \frac{1}{\sqrt{Nh(N)}} \sum_{i=1}^N \mathbf{z}_i(t) - \mathbb{E}[\mathbf{z}_i(t)] = \begin{bmatrix} \bar{\mathbf{x}}(t+1) \\ \bar{\mathbf{x}}(t) \\ \bar{\mathbf{w}}(t) \end{bmatrix}$$

converges in distribution to $\bar{\mathbf{z}}_\infty(t) = \begin{bmatrix} \bar{\mathbf{x}}_\infty(t+1)^T & \bar{\mathbf{x}}_\infty(t)^T & \bar{\mathbf{w}}_\infty(t)^T \end{bmatrix}^T \sim \mathcal{N}(\mathbf{0}, \mathrm{Cov}(\bar{\mathbf{z}}_\infty(t)))$,

where $\mathrm{Cov}(\bar{\mathbf{z}}_\infty(t)) = \lim_{N\to\infty}\mathrm{Cov}(\bar{\mathbf{z}}(t))$. This, together with the fact that $\bar{\mathbf{z}}(t)$ is uniformly integrable (since it has finite variance), implies convergence of expectations over these distributions. In particular, for any $\boldsymbol{\xi},\boldsymbol{\omega}$,

$$\lim_{N\to\infty}\mathbb{E}\left[\bar{\mathbf{x}}(t+1)\middle|\begin{bmatrix}\bar{\mathbf{x}}(t)\\ \bar{\mathbf{w}}(t)\end{bmatrix}=\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\omega}\end{bmatrix}\right] = \mathbb{E}\left[\bar{\mathbf{x}}_\infty(t+1)\middle|\begin{bmatrix}\bar{\mathbf{x}}_\infty(t)\\ \bar{\mathbf{w}}_\infty(t)\end{bmatrix}=\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\omega}\end{bmatrix}\right]. \tag{37}$$

Since $\mathbf{z}_\infty$ is normally distributed, the right hand side in (37) equals

$$\mathrm{Cov}\left(\bar{\mathbf{x}}_\infty(t+1),\begin{bmatrix}\bar{\mathbf{x}}_\infty(t)\\ \bar{\mathbf{w}}_\infty(t)\end{bmatrix}\right)\mathrm{Cov}\left(\begin{bmatrix}\bar{\mathbf{x}}_\infty(t)\\ \bar{\mathbf{w}}_\infty(t)\end{bmatrix}\right)^\dagger\begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\omega}\end{bmatrix}$$

which simplifies to $\mathbf{A}_\infty(t)\boldsymbol{\xi}+\mathbf{B}_\infty(t)\boldsymbol{\omega}$ in (35) because of the independence of $\bar{\mathbf{x}}_\infty(t)$ and $\bar{\mathbf{w}}_\infty(t)$ resulting from Assumption (A3). ■

As far as linearity of the limit dynamics is concerned, it is important to note that the matrices $\mathbf{A}_\infty(t)$ and $\mathbf{B}_\infty(t)$ in Theorem IV.8 depend on the *distributions* of states $\{\mathbf{x}_i(t)\}_{i=1}^\infty,\{\mathbf{x}_i(t+1)\}_{i=1}^\infty$ and noises $\{\mathbf{w}_i(t)\}_{i=1}^\infty$, but not on their specific values (realizations). In other words, for any choice of model in (6), i.e., for any choice of $\{f_i(\cdot)\}_{i=1}^\infty$, p_0, and p_w, the matrices $\mathbf{A}_\infty(t)$ and $\mathbf{B}_\infty(t)$ can be in theory pre-computed for all $t\geq 0$, and would be the same for all realizations of noises and states generated by this system. This is important for the dynamics to be actually linear, as illustrated by the following simple example.

Example IV.9. ***(LTV vs. nonlinear dynamics).*** *Consider a dynamical system described by*

$$y(t+1)=a(t)y(t),\quad y(0)\sim\mathcal{N}(e,1)$$

and three cases, as follows.

(i) $a(t)=e^{2^t}$. *Here the system is clearly LTV.*

(ii) $a(t)=y(t)$. *Here the system is clearly nonlinear.*

(iii) $a(t)=\mathbb{E}[y(t)]$. *Although the system here may look nonlinear, the coefficient* $\mathbb{E}[y(t)]$ *can be pre-computed for all* $t\geq 0$ *based only on the distribution of* $y(0)$. *In fact, it is straightforward to see that here* $a(t)=e^{2^t}$, *making this system equivalent to that in (i).* □

D. Time-Invariant Limit Dynamics

Interestingly, if the dynamics in (6) satisfy additional assumptions, the average dynamics tend not only to a linear system but further to an LTI one. For ease of notation, let

$$\mathbf{x}(t+1)=F(\mathbf{x}(t),\mathbf{w}(t)) \tag{38}$$

denote the combined dynamics of all sub-systems in (6), where $\mathbf{x}(t)\in\mathbb{R}^{Nn}$ and $\mathbf{w}(t)\in\mathbb{R}^{Nm}$ are the concatenations of all states and noises, respectively. We also need to make additional assumptions, as follows.

Assumption IV.10. ***(Additional assumptions for time-invariance of limit dynamics).***

(A6) *The noise distribution* $p_w(t)\equiv p_w$ *is time-invariant.*

(A7) *The function* $F(\cdot)$ *is globally Lipschitz in* $\mathbf{x}$, *i.e., for any* $\mathbf{w}\in\mathbb{R}^{Nm}$ *there exists* $L(\mathbf{w})\geq 0$ *such that*

$$\|F(\mathbf{x}_1,\mathbf{w})-F(\mathbf{x}_2,\mathbf{w})\|\leq L(\mathbf{w})\|\mathbf{x}_1-\mathbf{x}_2\|, \tag{39}$$

for all $\mathbf{x}_1,\mathbf{x}_2\in\mathbb{R}^{Nn}$.

(A8) *the initial distribution* p_0 *is such that* $p_0(A)=0$ *for any set* A *where* $p^*(A)=0$;

(A9) *for any set* A *with* $p^*(A)>0$, *the noise distribution* p_w *is such that for all* $\mathbf{x}\in\mathbb{R}^{Nn}$, $\mathbb{P}\{F(\mathbf{x},\mathbf{w})\in A\}>0$.

(A10) *The sequence* $\{\mathbf{y}_i(t)\}_{i=1}^\infty$ *is* ρ*-mixing* uniformly across time, *i.e.,* $\sup_t\varrho(n,t)\to 0$ *as* $n\to\infty$, *where* $\varrho(n,t)$ *is defined as in* (10) *for the sequence* $\{\mathbf{y}_i(t)\}_{i=1}^\infty$. □

Assumption (A6) is clearly necessary if we expect (6) to admit a stationary solution. Assumption (A7), while being restrictive on the space of all functions, becomes mild when considering only bounded functions as required by assumption (A4). Assumptions (A8) and (A9) are technical and ensure the distributions p_0 and p_w are well-behaved. Assumption (A10) is also technical and prevents contrived cases where, e.g., the system gradually converges to a state of global synchrony.

The next result uses Assumptions (A6)-(A9) to show the existence and attractivity of stationary solutions, which will then be used to prove the LTI version of Theorem IV.8.

Theorem IV.11. ***(Existence and attractivity of stationary solutions).*** *Consider the population dynamics in* (6) *and assume that Assumptions* (A1)-(A7) *hold. Then the population has a stationary solution*

$$\mathbf{x}^*(t)\sim p^*,\qquad t\geq 0. \tag{40}$$

where p^* *is independent of time. If, further, Assumptions* (A8)-(A9) *hold, then* $\mathbf{x}(t)$ *converge to* $\mathbf{x}^*(t)$ *in distribution as* $t\to\infty$.

Proof. The proof is the same as proof of [1, Thm IV.4]. ■

Combining the linearity of Theorem IV.8 and the stationarity of Theorem IV.11 ensures the convergence of the average population dynamics to an LTI system, as shown next.

Theorem IV.12. ***(LTI average population dynamics).*** *Consider the population dynamics* (6) *and assume that assumptions* (A1)-(A10) *hold. Then,*

$$\mathbb{E}\left[\bar{\mathbf{x}}(t+1)\middle|\begin{bmatrix}\bar{\mathbf{x}}(t)\\ \bar{\mathbf{w}}(t)\end{bmatrix}=\begin{bmatrix}\boldsymbol{\upsilon}_0\\ \boldsymbol{\zeta}_0\end{bmatrix}\right]\to\mathbf{A}^*_\infty\boldsymbol{\upsilon}_0+\mathbf{B}^*_\infty\boldsymbol{\zeta}_0$$

as $N,t\to\infty$, *where*

$$\mathbf{A}^*_\infty=\lim_{t\to\infty}\mathbf{A}_\infty(t),\quad \mathbf{B}^*_\infty=\lim_{t\to\infty}\mathbf{B}_\infty(t), \tag{41}$$

and $\mathbf{A}_\infty(t),\mathbf{B}_\infty(t)$ *are as in* (36).

Proof. By Theorem IV.11 each subsystem $\mathbf{x}_i(t)$ converges in distribution to the corresponding stationary solution $\mathbf{x}_i^*\sim p_i^*$ as $t\to\infty$, which are ρ-mixing (in the limit) by (A10). Define

$$\bar{\mathbf{x}}^*=\frac{1}{\sqrt{Nh(N)}}\sum_{i=1}^N\mathbf{x}_i^*-\mathbb{E}[\mathbf{x}_i^*],$$

and let $\bar{\mathbf{x}}^*_\infty \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{\bar{x}^*})$ where $\mathbf{\Sigma}_{\bar{x}^*} = \mathrm{Cov}(\bar{\mathbf{x}}^*)$. Then by Theorem IV.6, $\bar{\mathbf{x}}^*$ converges in distribution to $\bar{\mathbf{x}}^*_\infty$ as $N \to \infty$ and, therefore,

$$\operatorname*{dlim}_{N,t\to\infty} \bar{\mathbf{x}}(t) = \operatorname*{dlim}_{t\to\infty} \bar{\mathbf{x}}_\infty(t) = \bar{\mathbf{x}}^*_\infty. \tag{42}$$

It then follows from (42) that all the time-dependent covariances in (36) converge to their respective limits and, hence, so do the matrices $\mathbf{A}_\infty(t)$ and $\mathbf{B}_\infty(t)$. This completes the proof. ■

Similar to the guaranteed linearity of Theorem IV.8, the result of Theorem IV.12 is asymptotic and LTI dynamics are only *approached* as N and t grow to infinity. In many real-world systems, however, dynamics settle to stationary solutions after only a few time steps. Further, and more importantly for our discussion, nonlinearities often vanish rather quickly with averaging. We next focus on formalizing the latter dimension, namely, the rate of convergence to linearity under spatial averaging.

E. Rate of Convergence to Linearity

In this section we seek to estimate the rate at which spatially-averaged dynamics converge to linearity. We have empirically shown that this rate is often rather fast [1], but rigorous characterizations are missing. The main result of this section is the following, where we prove that convergence of conditional expectations (similar to (35)) occurs at the $O(1/\sqrt{N})$ rate. We present this result for the simplest case of i.i.d. sequences of static random vectors and discuss generalizations afterwards.

Theorem IV.13. ***(Rate of convergence to linearity).*** *Consider a sequence of N i.i.d random vectors $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_N \in \mathbb{R}^{n+q}$ where, for all i,*

$$\mathbf{z}_i = \begin{bmatrix}\mathbf{x}_i \\ \mathbf{y}_i\end{bmatrix}, \quad \mathbf{x}_i \in \mathbb{R}^n, \quad \mathbf{y}_i \in \mathbb{R}^q \tag{43}$$
$$\mathbb{E}[\mathbf{z}_i] = \boldsymbol{\mu}_z, \quad \mathrm{Cov}(\mathbf{z}_i) = \mathbf{\Sigma}_z, \quad \|\mathbf{z}_i\|_\infty \le M < \infty.$$

Let

$$\bar{\mathbf{z}} = \frac{1}{N}\sum_{i=1}^N \mathbf{z}_i \qquad \textit{and} \qquad \mathbf{z}^* = \begin{bmatrix}\mathbf{x}^* \\ \mathbf{y}^*\end{bmatrix} \sim \mathcal{N}\Big(\boldsymbol{\mu}_z, \frac{1}{N}\mathbf{\Sigma}_z\Big)$$

Then, for any $\boldsymbol{\upsilon} \in \mathbb{R}^q$ and any $\delta > 0$,

$$\Big\|\mathbb{E}[\bar{\mathbf{x}}|\bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})] - \mathbb{E}[\mathbf{x}^*|\mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})]\Big\| = O\Big(\frac{1}{\sqrt{N}}\Big), \tag{44}$$

where $\mathcal{B}_\delta(\boldsymbol{\upsilon}) = \{\mathbf{y} \in \mathbb{R}^q : ||\mathbf{y} - \boldsymbol{\upsilon}|| \le \delta\}$ is a δ-ball around $\boldsymbol{\upsilon}$.

Proof. According to Berry-Essen theorem [44], for any convex sets $U \subseteq \mathbb{R}^{n+q}$,

$$\big|\mathbb{P}(\bar{\mathbf{z}} \in U) - \mathbb{P}(\mathbf{z}^* \in U)\big| \le \frac{C}{\sqrt{N}} \tag{45}$$

where C is a constant independent of N. So consider, in particular, the sets of the form

$$U = \big\{\mathbf{x} \in \mathbb{R}^n | x_i \le \xi\big\} \times \mathcal{B}_\delta(\boldsymbol{\upsilon})$$

for any $i \in \{1, \ldots, n\}$ and any $\xi \in \mathbb{R}$. Plugging this into (45) we get

$$\big|\mathbb{P}\big\{\bar{x}_i \le \xi, \bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\} - \mathbb{P}\big\{x^*_i \le \xi, \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}\big| \le \frac{C}{\sqrt{N}}. \tag{46}$$

Similarly, by plugging $U = \mathbb{R}^n \times \mathcal{B}_\delta(\boldsymbol{\upsilon})$ in (45) we get

$$\big|\mathbb{P}\big\{\bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\} - \mathbb{P}\big\{\mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}\big| \le \frac{C}{\sqrt{N}}. \tag{47}$$

Further, let

$$\epsilon = \min\Big\{\mathbb{P}\big\{\bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}\ ,\ \mathbb{P}\big\{\mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}\Big\} > 0. \tag{48}$$

Combining (46), (47) and (48) we get

$$\begin{aligned}
&\Big|F_{\bar{x}_i|\bar{\mathbf{y}}}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) - F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)\Big| \\
&\triangleq \Big|\mathbb{P}\big\{\bar{x}_i \le \xi \mid \bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\} - \mathbb{P}\big\{x^*_i \le \xi \mid \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}\Big| \\
&= \left|\frac{\mathbb{P}\big\{\bar{x}_i \le \xi\ ,\ \bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}}{\mathbb{P}\big\{\bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}} - \frac{\mathbb{P}\big\{x^*_i \le \xi\ ,\ \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}}{\mathbb{P}\big\{\mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\}}\right| \\
&\le \frac{2C}{\sqrt{N}\epsilon^2}
\end{aligned} \tag{49}$$

where in the latter inequality we used $\big|\frac{a}{b} - \frac{c}{d}\big| \le \frac{|d||a-c|+|c||b-d|}{|b||d|}$. Now, using the standard relationship between $\mathbb{E}[X] = -\int_0^\infty F_X(x)dx + \int_0^\infty (1 - F_X(x))dx$ between the cumulative distribution function (CDF) and expected value of any random variable, we can translate the bound on the difference of conditional CDFs in (49) into a bound on the difference of conditional expectations, as follows.

$$\begin{aligned}
&\Big|\mathbb{E}\big[\bar{x}_i \mid \bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big] - \mathbb{E}\big[x^*_i \mid \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big]\Big| \\
&= \Big|\int_{-\infty}^0 \Big(F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) - F_{\bar{x}_i|\bar{\mathbf{y}}}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)\Big)d\xi \\
&\qquad + \int_0^\infty \Big(F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) - F_{\bar{x}_i|\bar{\mathbf{y}}}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)\Big)d\xi\Big| \\
&\le \int_{-\infty}^0 \Big|F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) - F_{\bar{x}_i|\bar{\mathbf{y}}}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)\Big|d\xi \\
&\qquad + \int_0^\infty \Big|F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) - F_{\bar{x}_i|\bar{\mathbf{y}}}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)\Big|d\xi
\end{aligned} \tag{50}$$

Note that, because the bound in (49) is independent of ξ, substituting it directly into (50) will give a useless bound of ∞. However, because $x_i \in [-M, M]$ by assumption, $F_{\bar{x}_i|\bar{\mathbf{y}}}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) = 0$ for all $\xi \notin [-M, M]$ and we can use this to achieve a finite bound, as follows. For reasons that we will see later, let $M' \ge M$ be a (yet-to-be-determined) more conservative bound on x_i. Then,

$$\begin{aligned}
&\Big|\mathbb{E}\big[\bar{x}_i \mid \bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big] - \mathbb{E}\big[x^*_i \mid \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big]\Big| \\
&\le \int_{-\infty}^{-M'} F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)d\xi + \int_{-M'}^0 \frac{2C}{\sqrt{N}\epsilon^2}d\xi \\
&\qquad + \int_0^{M'} \frac{2C}{\sqrt{N}\epsilon^2}d\xi + \int_{M'}^\infty \Big(1 - F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)\Big)d\xi \\
&= \frac{4M'C}{\sqrt{N}\epsilon^2} + 2\int_{-\infty}^{-M'} F_{x^*_i|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big)d\xi.
\end{aligned} \tag{51}$$

To evaluate the integral in (51), note that

$$\begin{aligned}
&F_{x_i^*|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) = \mathbb{P}\big\{x_i^* \le \xi \mid \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big\} \\
&\le \sup_{\mathbf{y}_0 \in \mathcal{B}_\delta(\boldsymbol{\upsilon})} \mathbb{P}\big\{x_i^* \le \xi \mid \mathbf{y}^* = \mathbf{y}_0\big\} \\
&= \sup_{\mathbf{y}_0 \in \mathcal{B}_\delta(\boldsymbol{\upsilon})} \Phi\left(\frac{\xi - \mu_{x_i} - \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}(\mathbf{y}_0 - \boldsymbol{\mu}_{\mathbf{y}})}{\big[\frac{1}{N}\big(\Sigma_{x_i} - \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{y}x_i}\big)\big]^{-\frac{1}{2}}}\right) \\
&= \Phi\left(\frac{\xi - \mu_{x_i} - \inf_{\mathbf{y}_0 \in \mathcal{B}_\delta(\boldsymbol{\upsilon})} \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}(\mathbf{y}_0 - \boldsymbol{\mu}_{\mathbf{y}})}{\big[\frac{1}{N}\big(\Sigma_{x_i} - \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{y}x_i}\big)\big]^{\frac{1}{2}}}\right) \\
&= \Phi\left(\frac{\xi - \mu_{x_i} - \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}(\boldsymbol{\upsilon} - \boldsymbol{\mu}_{\mathbf{y}}) + \delta\|\boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\|}{\big[\frac{1}{N}\big(\Sigma_{x_i} - \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{y}x_i}\big)\big]^{\frac{1}{2}}}\right)
\end{aligned}$$

where $\Phi(\cdot)$ is the standard normal CDF. For ease of notation let $\xi_0 = \mu_{x_i} + \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}(\boldsymbol{\upsilon} - \boldsymbol{\mu}_{\mathbf{y}}) - \delta\|\boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\|$ and $\sigma_0^2 = \Sigma_{x_i} - \boldsymbol{\Sigma}_{x_i\mathbf{y}}\boldsymbol{\Sigma}_{\mathbf{y}}^{-1}\boldsymbol{\Sigma}_{\mathbf{y}x_i}$. Then

$$F_{x_i^*|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) \le \Phi\Big(\frac{\xi - \xi_0}{\sigma_0}\sqrt{N}\Big) = \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_0}\int_0^{\xi} e^{-\frac{(\zeta-\xi_0)^2}{2\sigma_0^2}N} d\zeta$$

Notice that, from (51), always $\xi \le -M'$, where $M' \ge M$ is yet to be determined. Choose $M' = \max\{M, 1 - \xi_0\}$. Then, for all $\xi \le -M'$ we have

$$F_{x_i^*|\mathbf{y}^*}\big(\xi|\mathcal{B}_\delta(\boldsymbol{\upsilon})\big) \le \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_0}\int_0^{\xi} e^{-\frac{|\zeta-\xi_0|}{2\sigma_0^2}N} d\zeta = \frac{\sqrt{2}\sigma_0}{\sqrt{\pi N}} e^{\frac{\xi-\xi_0}{2\sigma_0^2}N}$$

Substituting this into (51) and integrating from $-\infty$ to $-M'$ gives

$$\begin{aligned}
&\Big|\mathbb{E}\big[\bar{x}_i \mid \bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big] - \mathbb{E}\big[x_i^* \mid \mathbf{y}^* \in \mathcal{B}_\delta(\boldsymbol{\upsilon})\big]\Big| \\
&\le \frac{4M'C}{\sqrt{N}\epsilon^2} + \frac{2}{\sqrt{\pi}}\Big(\frac{2\sigma_0}{N}\Big)^{\frac{3}{2}} e^{-\frac{M'+\xi_0}{2\sigma_0^2}N} = O\Big(\frac{1}{\sqrt{N}}\Big).
\end{aligned}$$

Because this is true for all $i = 1, \ldots, n$, and $\|\cdot\| \le \sqrt{n}\|\cdot\|_\infty$, we get (44), completing the proof. ■

A number of remarks about Theorem IV.13 are in order. First, this is the only result in this paper where we assume independence among the subsystems, in order to simplify the analysis. If the subsystems form a ρ-mixing sequence, similar results can be obtained using extensions of the Berry-Esseen theorem (see, e.g., [45]) but the analysis becomes more involved and the rates of convergence become variable depending on the rate at which correlations decay (e.g., $n^{-\frac{1}{6}}\log n$ if correlations decay as $n^{-\frac{3}{2}}$). Second, the L_∞ nature of the Berry-Esseen theorem (i.e., that the bound in (45) does not depend on the set U) has forced us to move from the singleton-conditioning in (35) to neighborhood-conditioning in (44). To the best of our knowledge this is not avoidable for continuous-valued conditioning variables (see [46] for conditioning on discrete variables) but the radius of $\mathcal{B}_\delta(\boldsymbol{\upsilon})$ can be made sufficiently small so that conditioning on $\bar{\mathbf{y}} \in \mathcal{B}_\delta(\boldsymbol{\upsilon})$ is a good approximation of conditioning on $\bar{\mathbf{y}} = \boldsymbol{\upsilon}$. Finally, we presented Theorem IV.13 for static random variables, but its extension to dynamical systems is straightforward, with $\mathbf{x}_i$ and $\mathbf{y}_i$ in (43) being replaced by $\mathbf{x}_i(t+1)$ and $\mathbf{y}_i(t) = \big[\mathbf{x}_i(t)^T \quad \mathbf{w}_i(t)^T\big]^T$, respectively. Under the same assumptions, namely, i.i.d. subsystems with uniformly bounded states and noises, the same conclusion as in (44) would then hold for conditional expectation of average states at any time $t+1$ given neighborhood (not singleton) conditions on average states and noises at the previous time t.

V. Linearizing Effect of Spatial Averaging on Spatially-Embedded Dynamical Systems

A. Problem Formulation

While the results of Section IV are general in many respects (form of nonlinearity, noise characteristics, etc.) they are still limited in requiring a decay of correlations with the difference in two subsystems' linear indices (i.e., $\varrho(|i-j|) \to 0$ as $|i-j| \to \infty$). In brain networks that initially motivated the present work [19], e.g., it is well-known that correlations between neurons decay as their distance grows [47]–[49], but this distance is physical (Euclidean) rather than between linear indices. Similarly, common forms of spatial averaging that occur in the brain, such as those underlying functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), occur over neurons that are close to each other in a physical sense rather than index-wise. Therefore, in this section we generalize our results of Section IV to systems where both the decay of pairwise correlations and the domains of spatial averaging occur in Euclidean space.

Consider the same form of nonlinear population dynamics (6) but now assume that each subsystem i lies at a location $\mathbf{r}_i \in \mathbb{R}^d$. These locations could be deterministic and fixed, or themselves be stochastic, and we will return to this later. For the sake of generality we keep d arbitrary, though we are often interested in $d = 3$. Instead of (7) let

$$\mathbf{x}_{\mathcal{N}_i}(t) = \big[\mathbf{x}_j(t) \ : \ \|\mathbf{r}_i - \mathbf{r}_j\| \le \tau\big], \tag{52}$$

where $\tau > 0$ can still be arbitrarily large but finite. We can now formulate the problem that we will tackle in this section, which closely parallels Problem 1.

Problem 2. ***(Linearizing Effect of Spatial Averaging on Spatially-Embedded Dynamical Systems).*** *Consider a heterogeneous population of nonlinear dynamical systems described by* (6) *and* (52)*. Define the population's average state vector as*

$$\bar{\mathbf{x}}(t) = \frac{1}{\phi(N_{\mathcal{R}})}\sum_{i:\|\mathbf{r}_i\|\le\mathcal{R}} \mathbf{x}_i(t) - \mathbb{E}[\mathbf{x}_i(t)], \tag{53}$$

where, for any $\mathcal{R} > 0$,

$$N_{\mathcal{R}} = \big|\{i \mid \|\mathbf{r}_i\| \le \mathcal{R}\}\big|, \tag{54}$$

and $\phi(\cdot)$ is a normalization factor as in (8)*. Prove, under appropriate assumptions and choice of $\phi(\cdot)$, that the dynamics of $\bar{\mathbf{x}}(t)$ becomes asymptotically linear as $\mathcal{R} \to \infty$.* □

The rest of this section is concerned with addressing Problem 2. We will follow the same general steps as in Section IV, focusing on what needs to be done differently in order to address the challenges that arise from the new formulation.

B. ρ^*-Mixing Sequences

Consider a spatially-embedded sequence of dynamical systems as in Problem 2, which gradually grows in size as $\mathcal{R}, N_{\mathcal{R}} \to \infty$. Regardless of how the subsystems are enumerated, if all we know is that their correlations decay with their physical distance, there is no guarantee that the same happens if $|i-j| \to \infty$. Therefore, to address Problem 2 we need a different notion of mixing, as introduced next.

Definition V.1. ***(Multivariate ρ^*-mixing sequence).*** *Consider a sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots \in \mathbb{R}^q$ with corresponding Euclidean positions $\mathbf{r}_1, \mathbf{r}_2, \cdots \in \mathbb{R}^d$. For any $r > 0$, define*

$$\varrho^*(r) = \sup_{A,B} \ \rho\Big(\sigma\big(\{\boldsymbol{\xi}_i \mid \mathbf{r}_i \in A\}\big), \sigma\big(\{\boldsymbol{\xi}_j \mid \mathbf{r}_j \in B\}\big)\Big), \quad (55)$$

where the supremum is taken over all nonempty sets $A, B \subset \mathbb{R}^d$ such that

$$\inf_{\mathbf{r}_i \in A,\ \mathbf{r}_j \in B} \|\mathbf{r}_i - \mathbf{r}_j\| \geq r,$$

and $\rho(\cdot,\cdot)$ is defined in (2)*. The sequence is ρ^*-mixing if*

$$\varrho^*(r) \to 0 \quad \textit{as} \quad r \to \infty. \quad (56)$$

ρ^*-mixing sequences share many of the same properties with ρ-mixing ones. For example, the notion of a residual factor (Definition IV.2) extends naturally to ρ^*-mixing sequences. The formal definition, for clarity, is as follows.

Definition V.2. ***(Residual factor for spatially-embedded sequences).*** *Let $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ be a ρ^*-mixing sequence with corresponding Euclidean positions $\mathbf{r}_1, \mathbf{r}_2, \ldots$. For any $\mathcal{R} > 0$ let*

$$S_{\mathcal{R}} = \sum_{i:\|\mathbf{r}_i\| \leq \mathcal{R}} \boldsymbol{\xi}_i,$$

and $N_{\mathcal{R}}$ be as in (54)*. The function $h(N_{\mathcal{R}})$ is called a* residual factor *for $\{\boldsymbol{\xi}_i\}_{i=1}^{\infty}$ if*

$$\lim_{\mathcal{R}\to\infty} \frac{1}{N_{\mathcal{R}} h(N_{\mathcal{R}})} \mathrm{Cov}(S_{\mathcal{R}}) < \infty,$$

i.e., the limit exists and is finite. □

In what follows we also need the following result, which parallels Lemma IV.5 and can be proved using the same approach.

Lemma V.3. ***(Sequence formed through transformation of ρ^*-mixing sequence is ρ^*-mixing).*** *Consider a sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots \in \mathbb{R}^q$ with corresponding Euclidean positions $\mathbf{r}_1, \mathbf{r}_2, \cdots \in \mathbb{R}^d$. For each i, let $\boldsymbol{\xi}_{\mathcal{N}_i} \in \mathbb{R}^{\nu_i}$ be as in* (52)*, h_i be a measurable function from $\mathbb{R}^{\nu_i}$ to $\mathbb{R}^k$, and*

$$\boldsymbol{\zeta}_i = h(\boldsymbol{\xi}_{\mathcal{N}_i}) \in \mathbb{R}^k.$$

If the sequence $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots$ is ρ^*-mixing, then so is the sequence $\boldsymbol{\zeta}_1, \boldsymbol{\zeta}_2, \ldots$. □

Next we move to the CLT, which lies at the core of our methodology. Here the main challenge lies in proving a result similar to Proposition III.3 for ρ^*-mixing sequences, as tackled next. Once that is shown, generalization to multivariate sequences will be straightforward.

Theorem V.4. ***(CLT for scalar ρ^*-mixing sequences).*** *Consider a scalar ρ^*-mixing sequence of random variables $\xi_1, \xi_2, \cdots \in \mathbb{R}$ with corresponding Euclidean positions $\mathbf{r}_1, \mathbf{r}_2, \cdots \in \mathbb{R}^d$. For any $\mathcal{R} > 0$, let*

$$S_{\mathcal{R}} = \sum_{i:\|\mathbf{r}_i\| \leq \mathcal{R}} \xi_i, \qquad \sigma_{\mathcal{R}}^2 = \mathrm{Var}(S_{\mathcal{R}}) = N_{\mathcal{R}} h(N_{\mathcal{R}}), \quad (57)$$

where $N_{\mathcal{R}}$ is defined in (54) *and $h(N_{\mathcal{R}})$ is a slowly-varying function. Assume, further, that*

$$\mathbb{E}[\xi_i] = 0 \textit{ and } |\xi_i| \leq M \textit{ for all } i \textit{ and some } M < \infty \quad (58a)$$

$$\sup_{\mathcal{R}>0} \frac{\mathbb{E}\big[|S_{\gamma\mathcal{R}}|^p\big]}{\sigma_{\mathcal{R}}^p} < \infty \textit{ for all } \gamma \in [0,1] \textit{ and some } p > 2 \quad (58b)$$

$$N_{\mathcal{R}} \propto \mathcal{R}^d, \textit{ i.e., } \lim_{\mathcal{R}\to\infty} \frac{N_{\mathcal{R}}}{\mathcal{R}^d} = C_v \textit{ for some } C_v > 0. \quad (58c)$$

Then

$$\frac{S_{\mathcal{R}}}{\sigma_{\mathcal{R}}} \xrightarrow{d} \mathcal{N}(0,1) \quad \textit{as} \quad \mathcal{R} \to \infty. \quad (59)$$

Proof. Consider the partial-sum stochastic process

$$W_{\mathcal{R}}(\gamma) = \frac{S_{\gamma^{\frac{1}{d}}\mathcal{R}}}{\sigma_{\mathcal{R}}}, \quad \text{for all } \gamma \in [0,1].$$

In what follows we will prove the weak convergence of $W_{\mathcal{R}}(\gamma)$ to the standard Brownian process $W(\gamma)$, from which the statement of the theorem follows as a special case. By (58a) we have

$$\mathbb{E}[W_{\mathcal{R}}(\gamma)] = 0, \quad (60)$$

and by (57) and (58c) we get

$$\begin{aligned}
\lim_{\mathcal{R}\to\infty} \mathbb{E}\big[W_{\mathcal{R}}^2(\gamma)\big] &= \lim_{\mathcal{R}\to\infty} \frac{1}{\sigma_{\mathcal{R}}^2} \mathbb{E}\Big[S_{\gamma^{\frac{1}{d}}\mathcal{R}}^2\Big] \\
&= \lim_{\mathcal{R}\to\infty} \frac{N_{\gamma^{\frac{1}{d}}\mathcal{R}} h\Big(N_{\gamma^{\frac{1}{d}}\mathcal{R}}\Big)}{N_{\mathcal{R}} h(N_{\mathcal{R}})} \\
&= \lim_{\mathcal{R}\to\infty} \frac{N_{\gamma^{\frac{1}{d}}\mathcal{R}}}{\gamma \mathcal{R}^d} \cdot \gamma \cdot \frac{\mathcal{R}^d}{N_{\mathcal{R}}} \cdot \frac{h\Big(N_{\gamma^{\frac{1}{d}}\mathcal{R}}\Big)}{h(N_{\mathcal{R}})} = \gamma, \quad (61)
\end{aligned}$$

where in the last equality we used the fact that $\lim_{\mathcal{R}\to\infty} h\big(N_{\gamma^{1/d}\mathcal{R}}\big)/h(N_{\mathcal{R}}) = 1$ by the uniform convergence of slowly-varying functions [50, Thm. 1.2.1].

Further, by (58b) the collection of random variables,

$$\big\{W_{\mathcal{R}}^2(\gamma) \mid \mathcal{R} \in \mathbb{R}\big\} \quad (62)$$

is uniformly integrable. To prove the weak convergence of $W_{\mathcal{R}}(\gamma)$ to $W(\gamma)$, we need to show that $W_{\mathcal{R}}(\gamma)$ satisfies two more conditions. The first is tightness, namely, that for any $\epsilon, \eta > 0$ there exists $\delta > 0$ such that for all $|s-t| < \delta$ and all sufficiently large $\mathcal{R}$,

$$\mathbb{P}\Big(|W_{\mathcal{R}}(s) - W_{\mathcal{R}}(t)| \geq \epsilon\Big) \leq \eta. \quad (63)$$

To show this, note that by Markov's inequality,

$$\mathbb{P}\Big(|W_{\mathcal{R}}(s) - W_{\mathcal{R}}(t)| \geq \epsilon\Big) \leq \frac{\mathbb{E}\big[|W_{\mathcal{R}}(s) - W_{\mathcal{R}}(t)|\big]}{\epsilon}. \quad (64)$$

Further, from (58c) for sufficiently large $\mathcal{R}$ we have $N_{\mathcal{R}} \simeq C_v \mathcal{R}^d$ and hence

$$\begin{aligned}\big|W_{\mathcal{R}}(s)-W_{\mathcal{R}}(t)\big| &= \frac{\Big|S_{s^{\frac{1}{d}}\mathcal{R}} - S_{t^{\frac{1}{d}}\mathcal{R}}\Big|}{\sigma_{\mathcal{R}}} \\ &\leq \frac{\Big|N_{s^{\frac{1}{d}}\mathcal{R}} - N_{t^{\frac{1}{d}}\mathcal{R}}\Big|}{\sigma_{\mathcal{R}}} M \lesssim \frac{C_v \mathcal{R}^d \delta}{\sigma_{\mathcal{R}}} M \xrightarrow{\delta\to 0} 0.\end{aligned}$$

Therefore, $\big|W_{\mathcal{R}}(s) - W_{\mathcal{R}}(t)\big| \to 0$ as $\delta \to 0$ which, together with (64) ensures (63).

The last property of $W_{\mathcal{R}}(\gamma)$ that we need to show is that of asymptotically independent increments, namely, that for all

$$0 \leq s_1 \leq t_1 < s_2 \leq t_2 < \cdots < s_k \leq t_k \leq 1$$

and for all linear Borel sets $H_1, \ldots, H_k$ the difference

$$\mathbb{P}\Big(W_{\mathcal{R}}(t_\ell) - W_{\mathcal{R}}(s_\ell) \in H_\ell, \ell = 1, \ldots, k\Big) - \prod_{\ell=1}^{k} \mathbb{P}\Big(W_{\mathcal{R}}(t_\ell) - W_{\mathcal{R}}(s_\ell) \in H_\ell\Big) \quad (65)$$

converges to 0 as $\mathcal{R} \to \infty$. For any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$, let

$$\alpha(\mathcal{A}, \mathcal{B}) = \sup_{A\in\mathcal{A},\ B\in\mathcal{B}} \Big|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)\Big|,$$

and let E_ℓ be the event $\{W_{\mathcal{R}}(t_\ell) - W_{\mathcal{R}}(s_\ell) \in H_\ell\}$ for any $\ell = 1, \ldots, k$. Then, by definition,

$$\begin{aligned}&\Big|\mathbb{P}\big(E_\ell \cap E_{\ell-1}\big) - \mathbb{P}\big(E_\ell\big)\mathbb{P}\big(E_{\ell-1}\big)\Big| \quad (66)\\ &\leq \alpha\Big(\sigma\Big(\big\{\xi_i \mid s_\ell^{\frac{1}{d}}\mathcal{R} < ||\mathbf{r}_i|| \leq t_\ell^{\frac{1}{d}}\mathcal{R}\big\}\Big), \\ &\qquad \sigma\Big(\big\{\xi_j \mid s_{\ell-1}^{\frac{1}{d}}\mathcal{R} < ||\mathbf{r}_j|| \leq t_{\ell-1}^{\frac{1}{d}}\mathcal{R}\big\}\Big)\Big).\end{aligned}$$

Furthermore, from [51, Prop 3.11], for any two σ-algebras $\mathcal{A}$ and $\mathcal{B}$,

$$\alpha(\mathcal{A}, \mathcal{B}) \leq \frac{1}{4}\rho(\mathcal{A}, \mathcal{B}). \quad (67)$$

Therefore, combining (66), (67), and (56),

$$\Big|\mathbb{P}\Big(E_\ell \cap E_{\ell-1}\Big) - \mathbb{P}\Big(E_\ell\Big)\mathbb{P}\Big(E_{\ell-1}\Big)\Big| \leq \frac{1}{4}\varrho^*(r_\ell)$$

where $r_\ell = \mathcal{R}\big(s_\ell^{\frac{1}{d}} - t_{\ell-1}^{\frac{1}{d}}\big)$. Therefore, as $\mathcal{R} \to \infty$ so does r_ℓ, and hence

$$\Big|\mathbb{P}\Big(E_\ell \cap E_{\ell-1}\Big) - \mathbb{P}\Big(E_\ell\Big)\mathbb{P}\Big(E_{\ell-1}\Big)\Big| \to 0 \quad \text{as} \quad \mathcal{R} \to \infty.$$

Then, by induction,

$$\Big|\mathbb{P}\Big(\bigcap_{\ell=1}^{k} E_\ell\Big) - \prod_{\ell=1}^{k} \mathbb{P}\Big(E_\ell\Big)\Big| \to 0 \quad \text{as} \quad \mathcal{R} \to \infty,$$

which is the same as (65).

Put together, it follows from (60), (61), (62), (63), and (65) that $W_{\mathcal{R}}(\gamma)$ satisfies all the assumptions of [52, Thm 19.2], and hence

$$W_{\mathcal{R}}(\gamma) \xrightarrow{d} W(\gamma) \quad \text{as} \quad \mathcal{R} \to \infty.$$

Then, setting $\lambda = 1$ we get

$$W_{\mathcal{R}}(1) = \frac{S_{\mathcal{R}}}{\sigma_{\mathcal{R}}} \xrightarrow{d} W(1) = \mathcal{N}(0, 1) \quad \text{as} \quad \mathcal{R} \to \infty,$$

which completes the proof. ■

A minor difference between Theorem V.4 and Proposition III.3 is the difference between assumptions (58b) and (4b). The latter are both technical assumptions needed to prove uniform integrability of families of partial sum processes such as that in (62), which in turn prevents pathological cases where some probability mass gradually drifts to infinity as N or $\mathcal{R} \to \infty$. Both assumptions are sufficient but not necessary for uniform integrability, neither is uniformly stronger or weaker than the other, and either can be replaced by the direct (but perhaps more opaque) assumption of uniform integrability.

Similar to Section IV, we next generalize Theorem V.4 to multivariate sequences. The proof technique is the same as that of Theorem IV.6 and hence omitted.

Theorem V.5. ***(Multivariate CLT for ρ^*-mixing sequence).*** *Consider a ρ^*-mixing sequence of random vectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \cdots \in \mathbb{R}^q$ with corresponding Euclidean positions $\mathbf{r}_1, \mathbf{r}_2, \cdots \in \mathbb{R}^d$ satisfying* (58c)*. For any $\mathcal{R} > 0$ let $S_{\mathcal{R}} = \sum_{i:\|\mathbf{r}_i\|\leq\mathcal{R}} \boldsymbol{\xi}_i$, $N_{\mathcal{R}}$ be as in* (54)*, $h(N_{\mathcal{R}})$ be a slowly-varying residual factor for $\{\boldsymbol{\xi}_i\}_{i=1}^{\infty}$, and $\boldsymbol{\Sigma}_{\bar{\xi}}^* = \lim_{R\to\infty} \mathrm{Cov}(S_{\mathcal{R}})/N_{\mathcal{R}}h(N_{\mathcal{R}})$. Assume, further, that*

$$\mathbb{E}[\boldsymbol{\xi}_i] = 0 \textit{ and } \|\boldsymbol{\xi}_i\| \leq M \textit{ for all } i \textit{ and some } M < \infty,$$

$$\sup_{\mathcal{R}>0} \frac{\mathbb{E}\big[|\boldsymbol{\theta}^T S_{\gamma\mathcal{R}}|^p\big]}{\big[\boldsymbol{\theta}^T \mathrm{Cov}(S_{\mathcal{R}})\boldsymbol{\theta}\big]^{\frac{p}{2}}} < \infty \textit{ for all } \boldsymbol{\theta} \in \mathbb{R}^q, \gamma \in [0, 1],$$

and some $p > 2$. Then,

$$\bar{\boldsymbol{\xi}} = \frac{1}{\sqrt{N_{\mathcal{R}}h(N_{\mathcal{R}})}} \sum_{i:\|\mathbf{r}_i\|\leq\mathcal{R}} \boldsymbol{\xi}_i \xrightarrow{d} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\bar{\xi}}^*)$$

as $\mathcal{R} \to \infty$. □

C. Asymptotic Linearity Under Spatial Averaging

Theorem V.5 provides the basis that we need for solving Problem 2. However, note that we have so far assumed, for simplicity, that the Euclidean locations $\{\mathbf{r}_i\}_{i=1}^{\infty}$ are deterministic and fixed–e.g., at the vertices of a lattice–whereas in reality spatially-embedded subsystems are often distributed randomly over space. Therefore, before getting to the main result of this section, we prove the following lemma showing that our only requirement on the subsystems' locations, i.e., assumption (58c), is still satisfied with probability 1 in a proper stochastic setting.

Lemma V.6. ***(Growth rate of number of subsystems distributed according to Poisson point process).*** *Consider a sequence of dynamical subsystems the locations $\{\mathbf{r}_i\}_{i=1}^{\infty}$ of which are distributed according to a homogeneous Poisson point process with rate λ on $\mathbb{R}^d$. For any $\mathcal{R} > 0$, let (the random variable) $N_{\mathcal{R}}$ be as in* (54)*. Then*

$$\mathbb{P}\Big(\lim_{\mathcal{R}\to\infty} \frac{N_{\mathcal{R}}}{\mathcal{R}^d} = \lambda C_v\Big) = 1, \quad (68)$$

where $C_v = \frac{\pi^{\frac{d}{2}}}{\Gamma(1+\frac{d}{2})}$ *is the volume of the unit ball in* $\mathbb{R}^d$.

Proof. For simplicity of notation, let $V_{\mathcal{R}} = C_v \mathcal{R}^d$ be the volume of a ball with radius $\mathcal{R}$, and for any $r > 0$ let

$$\mathcal{R}(r) = \Big(\frac{r}{C_v}\Big)^{\frac{1}{d}}.$$

Clearly $V_{\mathcal{R}(r)} = r$. For all $0 \leq r_0 < r_1 < \cdots < r_k$, define the annuli

$$A_\ell = \big\{\mathbf{r} \in \mathbb{R}^d \;\big|\; R(r_{\ell-1}) < \|\mathbf{r}\| \leq R(r_\ell)\big\},$$

and let $|A_\ell|$ be the volume of A_ℓ. Because A_ℓ's are disjoint, by definition [53, Def. 3.1],

$$N_{\mathcal{R}(r_\ell)} - N_{\mathcal{R}(r_{\ell-1})}, \qquad \ell = 1, \ldots, k$$

are mutually independent Poisson-distributed random variables with rates

$$\lambda|A_\ell| = \lambda\big(V_{R(r_\ell)} - V_{R(r_{\ell-1})}\big) = \lambda\big(r_\ell - r_{\ell-1}\big).$$

Therefore, $N_{\mathcal{R}(r)}, r > 0$ is a standard Poisson point process with rate λ on the positive real line. Let $\delta_1, \delta_2, \ldots$ be the i.i.d. inter-arrival times of this standard Poisson process, and $t_\ell = \delta_1 + \cdots + \delta_\ell$. Each δ_ℓ is exponentially distributed with rate λ, and hence by the strong law of large numbers,

$$\frac{t_\ell}{\ell} = \frac{\delta_1 + \cdots + \delta_\ell}{\ell} \overset{a.s.}{\to} \frac{1}{\lambda} \quad \text{as} \quad \ell \to \infty. \tag{69}$$

For any $r \in [t_\ell, t_{\ell+1})$, by definition $N_{\mathcal{R}(r)} = \ell$ and thus

$$\frac{\ell}{t_{\ell+1}} \leq \frac{N_{\mathcal{R}(r)}}{r} \leq \frac{\ell}{t_\ell}. \tag{70}$$

By (69), as $r \to \infty$ (and thus $\ell \to \infty$) both sides of (70) converge to λ a.s., and thus

$$\frac{N_{\mathcal{R}(r)}}{r} \overset{a.s.}{\to} \lambda \quad \text{as} \quad r \to \infty. \tag{71}$$

The statement of the theorem then follows by replacing $r = C_v \mathcal{R}(r)^d$ in (71) and changing the limiting variable from r to $\mathcal{R}$. ■

Lemma V.6 allows us to restate the preceding CLTs for ρ^*-mixing sequences with fixed Euclidean locations for the case where the locations are Poisson distributed. These restatements are straightforward and hence omitted. Instead, we proceed to use the combination of Lemma V.6 and the preceding CLTs to solve Problem 2. The following is the parallel to Assumption IV.7 for spatially-embedded systems.

Assumption V.7. ***(Updated assumptions for spatially-embedded systems).*** *Consider the population of dynamical systems described by* (6) *and* (52) *and let* $\mathbf{y}_i(t)$, $\mathbf{z}_i(t)$, *and* $N_{\mathcal{R}}$ *be as in* (31), (32), *and* (54), *respectively. We make the following assumptions.*

(A1*) $\mathbb{E}[\mathbf{w}_i(t)] = \mathbf{0}$ *and* $\|\mathbf{w}_i(t)\| \leq C$ *a.s. for all* i, t *and some* $C < \infty$.

(A2*) *For all* $\boldsymbol{\theta} \neq 0$, $t \geq 0$, *and* $\gamma \in [0, 1]$ *and some* $p > 2$,

$$\sup_{\mathcal{R}>0} \frac{\mathbb{E}\big[\big|\boldsymbol{\theta}^T \sum_{i:\|\mathbf{r}_i\| \leq \gamma\mathcal{R}} \mathbf{z}_i(t)\big|^p\big]}{\big[\boldsymbol{\theta}^T \mathrm{Cov}\big(\sum_{i:\|\mathbf{r}_i\| \leq \mathcal{R}} \mathbf{z}_i(t)\big)\boldsymbol{\theta}\big]^{\frac{p}{2}}} < \infty.$$

(A3*) *Same as Assumption* (A3).

(A4*) *Same as Assumption* (A4).

(A5*) *For all* t, *the sequence of random vectors* $\{\mathbf{y}_i(t)\}_{i=1}^\infty$ *is* ρ^**-mixing with a slowly-varying residual factor* $h(N_{\mathcal{R}})$. □

We are now ready to state the main result of this section, as follows. The proof is similar to the proof of Theorem IV.8 and hence omitted.

Theorem V.8. ***(Linearizing effect of spatial averaging on spatially-embedded populations of dynamical systems).*** *Consider the population of dynamical systems described by* (6) *and* (52), *and assume that Assumptions* (A1*)-(A5*) *as well as* (58c) *hold. Define the population average variables* $\bar{\mathbf{x}}(t)$ *and* $\bar{\mathbf{w}}(t)$ *as in* (53), *let* $\bar{\mathbf{y}}(t) = \big[\bar{\mathbf{x}}(t)^T \quad \bar{\mathbf{w}}(t)^T\big]^T$, *and assume that* $\lim_{\mathcal{R}\to\infty} \mathrm{Cov}\big(\bar{\mathbf{x}}(t+1), \bar{\mathbf{y}}(t)\big)$ *exists. Then, for all* $\boldsymbol{\xi} \in \mathbb{R}^n, \boldsymbol{\omega} \in \mathbb{R}^m$,

$$\mathbb{E}\left[\bar{\mathbf{x}}(t+1) \middle| \begin{bmatrix}\bar{\mathbf{x}}(t)\\ \bar{\mathbf{w}}(t)\end{bmatrix} = \begin{bmatrix}\boldsymbol{\xi}\\ \boldsymbol{\omega}\end{bmatrix}\right] \xrightarrow{\mathcal{R}\to\infty} \mathbf{A}_\infty(t)\boldsymbol{\xi} + \mathbf{B}_\infty(t)\boldsymbol{\omega} \tag{72}$$

where $\mathbf{A}_\infty(t)$ *and* $\mathbf{B}_\infty(t)$ *are defined in* (36). *Furthermore, if the subsystem locations* $\{\mathbf{r}_i\}_{i=1}^\infty$ *are distributed according to a homogeneous Poisson point process on* $\mathbb{R}^d$, (72) *holds with probability 1.* □

As we noted in the Introduction,the present work was motivated by our empirical observations of macroscopic linearity in brain networks. Theorem V.8 now rigorously explains these observations, and highlights the key roles of spatial averaging that is inherent in macroscopic recordings such as electroencephalography (EEG) [54] and functional magnetic resonance imaging (fMRI) [55], as well as the weak and decaying correlations that have long been observed between biological neurons [47]–[49]. Whether and to what extent large-scale complex networks in other domains meet the assumptions of this theorem, and are hence bound to follow its consequent macroscopic linearity, remains an interesting topic of inquiry for future research.

VI. Conclusions

In this article we developed a theoretical framework to understand the linearity of spatially averaged dynamics in heterogeneous populations of nonlinear networked dynamical systems. This research was motivated by several observations of linear behavior at the macroscopic level in biological and artificial neural systems, as well as in-silico observations that nonlinear systems nearly universally exhibit (more) linear dynamics when subjected to spatial averaging. To our knowledge, this work is the first to formalize these observations into a unified theoretical framework and support them with rigorous mathematical analysis. By building on and extending the celebrated central limit theorem and the theory of mixing sequences, we proved that averaging has a strong and robust linearizing effect that holds for almost any form of microscopic nonlinearity, noise distribution, and network connectivity patterns, *as long as* pairwise correlations decay with some notion of distance between microscopic subsystems. We proved this result in two general settings, one where pairwise correlations

decay with the distance $|i-j|$ between subsystems' linear indices, and the other when pairwise correlations decay with the Euclidean distance $\|\mathbf{r}_i - \mathbf{r}_j\|$ between the physical locations of spatially-embedded subsystems. These results were further extended to time-invariant limit dynamics, finite-sample averaging with rates of convergence, and networks of spatially-embedded subsystems with random locations. Overall, our results provide significant insights into the macroscopic behavior of large-scale systems, and lay a robust theoretical foundation for future research across various domains.

REFERENCES

[1] 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)*. IEEE, 2022, pp. 641–648.

[2] M. Breakspear, "Dynamic models of large-scale brain activity," *Nature neuroscience*, vol. 20, no. 3, pp. 340–352, 2017.

[3] W. Burghout, H. N. Koutsopoulos, and I. Andreasson, "Hybrid mesoscopic–microscopic traffic simulation," *Transportation Research Record*, vol. 1934, no. 1, pp. 218–225, 2005.

[4] M. Marchiori and L. Possamai, "Micro-macro analysis of complex networks," *PloS one*, vol. 10, no. 1, p. e0116670, 2015.

[5] I. Lampropoulos, G. M. Vanalme, and W. L. Kling, "A methodology for modeling the behavior of electricity prosumers within the smart grid," in *2010 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)*. IEEE, 2010, pp. 1–8.

[6] M. O. Jackson, B. W. Rogers, and Y. Zenou, "The economic consequences of social-network structure," *Journal of Economic Literature*, vol. 55, no. 1, pp. 49–95, 2017.

[7] A. Banos, N. Corson, B. Gaudou, V. Laperrière, and S. R. Coyrehourcq, "Coupling micro and macro dynamics models on networks: Application to disease spread," in *International Workshop on Multi-Agent Systems and Agent-Based Simulation*. Springer, 2015, pp. 19–33.

[8] C. Zang, P. Cui, and C. Faloutsos, "Beyond sigmoids: The nettide model for social network growth, and its applications," in *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016, pp. 2015–2024.

[9] F. Milano and R. Zárate-Miñano, "A systematic method to model power systems as stochastic differential algebraic equations," *IEEE Transactions on Power Systems*, vol. 28, no. 4, pp. 4537–4544, 2013.

[10] M. M. Churchland, J. P. Cunningham, M. T. Kaufman, J. D. Foster, P. Nuyujukian, S. I. Ryu, and K. V. Shenoy, "Neural population dynamics during reaching," *Nature*, vol. 487, no. 7405, pp. 51–56, 2012.

[11] A. A. Ivanova, M. Schrimpf, S. Anzellotti, N. Zaslavsky, E. Fedorenko, and L. Isik, "Is it that simple? linear mapping models in cognitive neuroscience," *bioRxiv*, 2021.

[12] U. A. Ernst and C. W. Eurich, "Cortical population dynamics and psychophysics," *Handbook of brain theory and neural networks*, pp. 294–300, 2002.

[13] M. J. Keeling and J. V. Ross, "On methods for studying stochastic disease dynamics," *Journal of the Royal Society Interface*, vol. 5, no. 19, pp. 171–181, 2008.

[14] H. Markram, "The blue brain project," *Nature reviews neuroscience*, vol. 7, no. 2, pp. 153–160, 2006.

[15] S. Yao, Q. Wang, K. E. Hirokawa, B. Ouellette, R. Ahmed, J. Bomben, K. Brouner, L. Casal, S. Caldejon, A. Cho *et al.*, "A whole-brain monosynaptic input connectome to neuron classes in mouse visual cortex," *Nature Neuroscience*, vol. 26, no. 2, pp. 350–364, 2023.

[16] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," *nature*, vol. 473, no. 7346, pp. 167–173, 2011.

[17] S. Gu, F. Pasqualetti, M. Cieslak, Q. K. Telesford, A. B. Yu, A. E. Kahn, J. D. Medaglia, J. M. Vettel, M. B. Miller, S. T. Grafton *et al.*, "Controllability of structural brain networks," *Nature communications*, vol. 6, no. 1, p. 8414, 2015.

[18] T. Hahn, H. Jamalabadi, E. Nozari, N. R. Winter, J. Ernsting, M. Gruber, M. J. Mauritz, P. Grumbach, L. Fisch, R. Leenings *et al.*, "Towards a network control theory of electroconvulsive therapy response," *PNAS nexus*, vol. 2, no. 2, p. pgad032, 2023.

[19] E. Nozari, M. A. Bertolero, J. Stiso, L. Caciagli, E. J. Cornblath, X. He, A. S. Mahadevan, G. J. Pappas, and D. S. Bassett, "Macroscopic resting-state brain dynamics are best described by linear models," *Nature Biomedical Engineering*, 2023.

[20] V. Booth and J. Rinzel, "A minimal, compartmental model for a dendritic origin of bistability of motoneuron firing patterns," *Journal of computational neuroscience*, vol. 2, no. 4, pp. 299–312, 1995.

[21] H. R. Wilson and J. D. Cowan, "Excitatory and inhibitory interactions in localized populations of model neurons," *Biophysical journal*, vol. 12, no. 1, pp. 1–24, 1972.

[22] O. G. Sani, B. Pesaran, and M. M. Shanechi, "Dissociative and prioritized modeling of behaviorally relevant neural dynamics using recurrent neural networks," *Nature neuroscience*, vol. 27, no. 10, pp. 2033–2045, 2024.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] A. G. Bryant, K. Aquino, L. Parkes, A. Fornito, and B. D. Fulcher, "Extracting interpretable signatures of whole-brain dynamics through systematic comparison," *bioRxiv*, pp. 2024–01, 2024.

[31] Y. Hosaka, T. Hieda, K. Hayashi, K. Jimura, and T. Matsui, "Linear models replicate the energy landscape and dynamics of resting-state brain activity," *bioRxiv*, pp. 2024–05, 2024.

[32] A. Jacot, F. Gabriel, and C. Hongler, "Neural tangent kernel: Convergence and generalization in neural networks," *Advances in neural information processing systems*, vol. 31, 2018.

[33] R. Durrett, *Probability: Theory and Examples*, 5th ed. Cambridge University Press, 2019.

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

[35] 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

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

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

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

[39] 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.

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

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

[42] Y. Yang, O. G. Sani, E. F. Chang, and M. M. Shanechi, "Dynamic network modeling and dimensionality reduction for human ecog activity," *Journal of neural engineering*, vol. 16, no. 5, p. 056014, 2019.

[43] R. L. Graham, *Concrete mathematics: a foundation for computer science*. Pearson Education India, 1994.

[44] M. Raič, "A multivariate berry–esseen theorem with explicit constants," *Bernoulli*, vol. 25, no. 4A, pp. 2824–2853, 2019.

[45] X. J. Wang and S. H. Hu, "The berry–esseen bound for ρ-mixing random variables and its applications in nonparametric regression model,"

Theory of Probability & Its Applications, vol. 63, no. 3, pp. 479–499, 2019.

[46] P. S. Dey and G. Terlov, "Stein's method for conditional central limit theorem," *The Annals of Probability*, vol. 51, no. 2, pp. 723–773, 2023.

[47] M. A. Smith and A. Kohn, "Spatial and temporal scales of neuronal correlation in primary visual cortex," *Journal of Neuroscience*, vol. 28, no. 48, pp. 12 591–12 603, 2008.

[48] M. A. Smith and M. A. Sommer, "Spatial and temporal scales of neuronal correlation in visual area v4," *Journal of Neuroscience*, vol. 33, no. 12, pp. 5422–5432, 2013.

[49] R. Rosenbaum, M. A. Smith, A. Kohn, J. E. Rubin, and B. Doiron, "The spatial structure of correlated neuronal variability," *Nature neuroscience*, vol. 20, no. 1, pp. 107–114, 2017.

[50] N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*. Cambridge university press, 1989, vol. 27.

[51] R. Bradley, *Introduction to Strong Mixing Conditions*, ser. Introduction to Strong Mixing Conditions. Kendrick Press, 2007, no. v. 1. [Online]. Available: https://books.google.com/books?id=b4ovxwEACAAJ

[52] P. Billingsley, K. M. R. Collection, W. Shewhart, W. series in probability, mathematical statistics, and S. Wilks, *Convergence of Probability Measures*, ser. Wiley Series in Probability and Mathematical Statistics. Wiley, 1968. [Online]. Available: https://books.google.com/books?id=O9oQAQAAIAAJ

[53] G. Last and M. Penrose, *Lectures on the Poisson process*. Cambridge University Press, 2018, vol. 7.

[54] 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.

[55] R. B. Buxton, "The physics of functional magnetic resonance imaging (fmri)," *Reports on Progress in Physics*, vol. 76, no. 9, p. 096601, 2013.