Spatial Localization for Nonlinear Dynamical Stochastic Models for Excitable Media*

Nan CHEN[1]   Andrew J. MAJDA[2]   Xin T. TONG[3]

Abstract Nonlinear dynamical stochastic models are ubiquitous in different areas. Their statistical properties are often of great interest, but are also very challenging to compute. Many excitable media models belong to such types of complex systems with large state dimensions and the associated covariance matrices have localized structures. In this article, a mathematical framework to understand the spatial localization for a large class of stochastically coupled nonlinear systems in high dimensions is developed. Rigorous mathematical analysis shows that the local effect from the diffusion results in an exponential decay of the components in the covariance matrix as a function of the distance while the global effect due to the mean field interaction synchronizes different components and contributes to a global covariance. The analysis is based on a comparison with an appropriate linear surrogate model, of which the covariance propagation can be computed explicitly. Two important applications of these theoretical results are discussed. They are the spatial averaging strategy for efficiently sampling the covariance matrix and the localization technique in data assimilation. Test examples of a linear model and a stochastically coupled FitzHugh-Nagumo model for excitable media are adopted to validate the theoretical results. The latter is also used for a systematical study of the spatial averaging strategy in efficiently sampling the covariance matrix in different dynamical regimes.

Keywords Large state dimensions, Diffusion, Mean field interaction, Spatial averaging strategy, Efficiently sampling
2000 MR Subject Classification 65C40, 68Q17, 60G20

1 Introduction

Nonlinear dynamical stochastic models are ubiquitous in different areas (see [35–36, 50, 52]). Key features of these complex systems are multiscale dynamics, high-dimensional phase space, nonlinear energy transfers, highly non-Gaussian probability density functions (PDFs for short), intermittent instability, random internal and external forcing as well as extreme events. Understanding and predicting these complex turbulent dynamical systems have both scientific and practical significance. Many excitable media models belong to such types of

nonlinear dynamical stochastic systems (see [23, 33, 45, 56]). These excitable media models are ubiquitous in nature and can be found in physical, chemical, and biological systems that are far from thermodynamic equilibrium. One important feature of these excitable medium is that they are susceptible to finite perturbations, where a variety of wave patterns can be triggered. Examples of excitable medium include solitary waves, target like patterns, and spiral waves.

For nonlinear dynamical stochastic models, the high dimensionality imposes great challenges to the understanding of the underlying mechanisms, the design of effective prediction schemes and various uncertainty quantification problems (see [4, 47]). In particular, the high dimensionality leads to both a large computational cost and an enormous storage even for one single model simulation. Moreover, it is often of interest to calculate the time evolution of the statistics, especially the covariance matrix that includes the information of the interactions between different components (see [1, 29]). Since analytical solutions for complex nonlinear systems are typically unavailable, Monte Carlo methods are often applied. However, the affordable sample size is very limited due to the computational cost while the size of the covariance matrix, which is the square of the state variables, can be many magnitudes larger in many real applications that involve millions or billions of state variables. Furthermore, non-Gaussian features due to the nonlinear nature of the underlying dynamics such as extreme events can greatly affect the calculation of the covariance matrix and introduce extra difficulties. Therefore, understanding the covariance structures in nonlinear dynamical stochastic models and then developing efficient computing strategies that can exploit these structures become extremely important.

To facilitate the calculation of the covariance matrix, various approximations are applied in practice. One widely used strategy is to adopt Gaussian approximations for the underlying dynamics since there are many well-developed analytical and numerical methods for sampling high-dimensional Gaussian distributions. However, completely ignoring the non-Gaussian information in the underlying dynamics often leads to large model errors. One effective strategy is to decompose the entire phase space into a high dimensional Gaussian part and a low dimensional non-Gaussian part using the conditional Gaussian framework (see [9–12, 32, 38]). Another strategy is to exploit the fact that the majority of the problem uncertainty is distributed within a low dimensional subspace of the Karhuen-Loève expansion, then uncertainty quantification (UQ for short) can be operated just in this subspace for sampling and computational simplicity (see [7, 13–15, 19, 46, 51]).

This paper focuses on a different structure-exploiting strategy known as localization, which is widely used in many areas including the numerical weather prediction (see [22, 25–26]) and efficiently quantifying the uncertainty in excitable media (see [42, 53][6, 44]). In many applications, the system state variables represent the information at different spatial locations, and the dependence of certain information between different grid points decays quickly when their distance becomes large. In other words, two grid points can be regarded as nearly independent if they are far from each other. The localization strategy exploits this feature and adopts a band matrix as the approximation of the covariance (see [2, 48]). Thus, the localization allows the development of much simpler approaches in analyzing the covariance matrix and thus facilitates the study of many important UQ problems.

While the localization strategy has a wide range of real-world applications, it remains unclear about the role of different dynamical components on affecting the covariance localization. In addition, despite many empirical evidences, there is no rigorous justifiction of the covariance decay as a function of the spatial distance. Nevertheless, these are crucial topics in practice for at least two reasons. First, it is known that applying localization by approximating the full covariance matrix with a band matrix may introduce some biases in quantifying the uncertainty of the underlying system (see [42, 53]) and the current bias estimates (see [41, 53]) require the knowledge of the underlying covariance matrix as well as a quantitative description of the decay

in the covariance. Second, in order to apply localization in practice, one often needs to specify a localization radius L, which represents the typical distance for two components to be roughly independent. Unfortunately, in the absence of a rigorous theory to understand the covariance decay behavior, determining L often relies on exhaustive tuning and data fitting, which can be expensive and inaccurate.

This paper aims at developing a theoretical framework to understand the covariance behavior for a large class of stochastically coupled nonlinear systems in high dimensions that include many excitable media models and other nonlinear complex models. Rigorous mathematical analysis shows that the local effect from the diffusion results in an exponential decay of the components in the covariance matrix as a function of the distance while the global effect due to the mean field interaction synchronizes different components and contributes to a global covariance. To achieve this, a covariance comparison principle is established. Then by explicitly computing the covariance propagation in a linear surrogate model, an upper bound on the covariance entries can be obtained.

Two important applications of these theoretical results are illustrated. First, a spatial averaging sampling strategy is developed. It makes use of the statistical symmetry to replace the samples from repeated experiments using ensemble methods by samples at different spatial grid points (see [8, 49]), which greatly reduces the number of repeated experiments while the accuracy remains the same. In other words, combining the spatial averaging strategy with a small number of Monte Carlo simulation of the underlying nonlinear complex systems, the covariance with a localization structure can be sampled in an efficient and accurate fashion, at least for a finite time. Secondly, the same structure can be exploited to improve the accuracy of data assimilation (see [18, 30–31, 37]). The theory developed here provides a practical guideline to choose the localization radius L in a systematical way that facilitates an effective data assimilation for large dimensional systems.

The rest of the article is organized as follows. Section 2 states the main theoretical results for the spatial localization for nonlinear dynamical stochastic models and the two applications. Section 3 includes the detailed analysis of the covariance matrix with both local and nonlocal effects. Numerical simulations of a linear model and a stochastically coupled FitzHugh-Nagumo (FHN for short) model for excitable media in different dynamical regimes are shown in Section 4. The article is concluded in Section 5. Details of proofs are included in Appendix.

2 Problem Setup and Main Results

To describe the behavior of nonlinear dynamical stochastic models for an excitable media, we consider a general stochastic diffusion reaction model with mean field interaction on a one dimensional torus $[0, 1]$:

$$\frac{\mathrm{d}}{\mathrm{d}t}u_t = \nu\Delta u_t + F(u_t) + \int h(u_t(x))\mathrm{d}x + \dot{W}(t, x), \tag{2.1}$$

where $W(t, x)$ is a spatial white noise. We can discretize (2.1) in space, and write down the dynamics at the i-th grid point $u_i(t) = u_t\left(\frac{i}{N}\right)$ as

$$\mathrm{d}u_i(t) = \frac{\nu}{2}N^2(u_{i+1}(t) + u_{i-1}(t) - 2u_i(t))\mathrm{d}t + F(u_i(t))\mathrm{d}t + \frac{1}{N}\sum_{j=1}^{N} h(u_j(t)) + \mathrm{d}w_i(t), \tag{2.2}$$

where $w_i(t)$ are independent standard Wiener processes.

The deterministic forcings in (2.2) can be classified into two groups based on their range of action. Here $\frac{\nu}{2}N^2(u_{i+1}(t) + u_{i-1}(t) - 2u_i(t)) + F(u_i(t))$ describes how does a component u_i

force its neighbors and itself while $\frac{1}{N}\sum_{j=1}^{N}h(u_j(t))$ models the McKean Vlasov type of mean field interaction (see [39]). We will adapt this classification to simplify our notation. Moreover, we will consider the general setting where u_i is not univariate. This will be useful when we discuss the FHN model, as we need both the membrane voltage and the recovery forcing.

Specifically, we let $\mathbf{x}(t) = (\mathbf{x}_1(t), \cdots, \mathbf{x}_N(t))$ be a stochastic process in $\mathbb{R}^d$. For simplicity, we assume all block share the same dimension, that is $\mathbf{x}_i(t) \in \mathbb{R}^q$ and $d = qN$. Suppose each block follows an SDE,

$$d\mathbf{x}_i = \mathbf{f}(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})dt + \frac{1}{N}\sum_{j=1}^{N}\mathbf{h}(t, \mathbf{x}_j)dt + \Sigma d\mathbf{w}_i(t), \quad i = 1, \cdots, N. \qquad (2.3)$$

In (2.3), $\mathbf{w}_i(t)$ are independent standard Wiener processes of dimension q, and Σ is symmetric. The index i should be interpreted in a cyclic fashion, that is $\mathbf{x}_0 = \mathbf{x}_N, \mathbf{x}_1 = \mathbf{x}_{N+1}$. The natural distance between indices is given by

$$\mathbf{d}(i, j) = \min\{|i - j|, |i + N - j|, |j - i + N|\}.$$

For simplicity, we assume $\{\mathbf{x}_i(0)\}_{i=1,\cdots,N}$ are i.i.d. samples from $\mathcal{N}(m_0, \Sigma_0\Sigma_0^\mathrm{T})$.

Many high dimensional stochastic models can be formulated as (2.3). Examples include the Lorenz 96 model (see [34]) and the FHN model (see[33]).

Our main results can be summarized as the follows.

Theorem 2.1 *Suppose the following constants are independent of N,*

$$\begin{aligned}
\lambda_0 &:= \sup_{i,\mathbf{x},s\leq t} \{\lambda_{\max}(\nabla_{\mathbf{x}_i}\mathbf{f}(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) + \nabla_{\mathbf{x}_i}\mathbf{f}(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})^\mathrm{T})\}, \\
\lambda_F &:= \sup_{i,\mathbf{x},s\leq t} \{\|\nabla_{\mathbf{x}_{i-1}}\mathbf{f}(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\|, \|\nabla_{\mathbf{x}_{i+1}}\mathbf{f}(s, \mathbf{x}_{i+1}, \mathbf{x}_i, \mathbf{x}_{i+1})\|\}, \\
\lambda_H &:= \sup_{i,\mathbf{x}_i,s\leq t} \|\nabla_{\mathbf{x}_i}\mathbf{h}(s, \mathbf{x}_i)\|.
\end{aligned} \qquad (2.4)$$

Here $\lambda_{\max}$ denotes the maximum eigenvalue. Then for any test function g and $\beta > 0$, there is a constant $C_{\beta,t}$ independent on N, such that

$$\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t))) \leq C_{\beta,t}\|\nabla g\|_\infty^2\left(\mathrm{e}^{-\beta\mathbf{d}(i,j)} + \frac{1}{N}\right),$$

where $\|\nabla g\|_\infty := \sup_{\mathbf{x}}\|\nabla g(\mathbf{x})\|$. And if $\mathbf{h} \equiv 0$, the bound can be improved to $C_{\beta,t}\|\nabla g\|_\infty^2\mathrm{e}^{-\beta\mathbf{d}(i,j)}$.

The exact value of $C_{\beta,t}$ can be found below in Theorem 3.2. The main conclusion here is that when the dimension N is large, for components that are far from each other, their covariance is close to zero. In other words, the covariance between model components is significant only for nearby components. Following [42], we call such a covariance structure to be local. Below, we illustrate two important applications that show such a property is crucial for uncertainty quantification (UQ for short).

2.1 Sampling with spatial averaging

One fundamental UQ question regarding SDEs such as (2.3) is how to efficiently compute $\mathbb{E}g(\mathbf{x}_i(t))$ for a given test function and ensure that the estimation error is small. The standard Monte Carlo approach involves simulating (2.3) multiple times, and use the sample average $\widehat{g}_{mc} = \frac{1}{K}\sum_{j=1}^{K}g(\mathbf{x}_i^{(j)}(t))$ as an estimator. However, with the increase of the dimension N, the

computational costs to generate each independent sample $\mathbf{x}^{(j)}(t)$ increases dramatically. Therefore, only a small number of samples K can be adopted in practice.

One important observation is that (2.3) is invariant under index shifting. Therefore, by symmetry, $\mathbf{x}_i(t)$ shares a common distribution for all i. As a consequence, each of them can be viewed as an individual, despite not independent, sample. Exploiting this perspective, it has been proposed in [8] the following spatial average as an estimator of $\mathbb{E}g(\mathbf{x}_i(t))$,

$$\widehat{g}_{K,N} = \frac{1}{K} \sum_{j=1}^{K} \frac{1}{N} \sum_{i=1}^{N} g(\mathbf{x}_i^{(j)}(t)).$$

Since the samples are independent, the statistical properties of $\widehat{g}_{M,N}$ is determined by the ones of

$$\widehat{g}_N = \frac{1}{N} \sum_{i=1}^{N} g(\mathbf{x}_i(t)). \tag{2.5}$$

It is straightforward to verify that $\widehat{g}_N$ and $\widehat{g}_{K,N}$ are unbiased:

$$\mathbb{E}\widehat{g}_{K,N} = \mathbb{E}\widehat{g}_N = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}g(\mathbf{x}_i(t)) = \mathbb{E}g(\mathbf{x}_1(t)).$$

The accuracy of these estimators can be measured by their variance. By independence, we have $\text{var}\widehat{g}_{K,N} = \frac{1}{K}\widehat{g}_N$, which is given by

$$\text{var}\widehat{g}_N = \frac{1}{N^2} \sum_{i,j=1}^{N} \text{cov}(g(\mathbf{x}_i), g(\mathbf{x}_j)) = \frac{1}{N} \sum_{j=1}^{N} \text{cov}(g(\mathbf{x}_1), g(\mathbf{x}_j)). \tag{2.6}$$

Then plugging in the estimate from Theorem 2.1, the result in (2.6) can be bounded by

$$\text{var}\widehat{g}_N \leq \frac{1}{N} \sum_{j=1}^{N} C_{\beta,t} \|\nabla g\|_\infty^2 \left(e^{-\beta \mathbf{d}(1,j)} + \frac{1}{N} \right) \leq \frac{1}{N} \frac{C_{\beta,t} \|\nabla g\|_\infty^2}{1 - e^{-\beta}}.$$

Therefore we have

$$\text{var}\widehat{g}_{K,N} \leq \frac{1}{NK} \frac{C_{\beta,t} \|\nabla g\|_\infty^2}{1 - e^{-\beta}}.$$

In contrast, if we consider the standard estimator $\widehat{g}_{mc}$, its variance is $\frac{1}{K}\text{var}(g(\mathbf{x}_i(t)))$, which is of $O(N)$ multiple of the one of $\text{var}\widehat{g}_{M,N}$. In other words, the spatial averaging has increase the effective sample size by a factor of N.

2.2 Localization in data assimilation

In data assimilation (DA for short), sequential partial observations are incorporated to improve predictions of a stochastic dynamical system (see [18, 30–31, 37]). One major application of DA is numerical weather prediction, where forecasts are obtained by combining simulations of the dynamical models with satellite data. The associated dimensions of such data assimilation problems are extremely high, often in orders of millions in order to describe variables in different temporal and spatial scales of the underlying nonlinear complex systems. Moreover, for these refined models, the current computational power can only provide around one hundred simulation samples. DA algorithms, for example the ensemble Kalman filter (EnKF for short), need to use these 100 samples to represent the covariance matrix of model components.

The classic sample covariance,

$$\widehat{C} := \frac{1}{K-1}\sum_{j=1}^{K}(\mathbf{x}^{(j)}(t) - \overline{\mathbf{x}}) \cdot (\mathbf{x}^{(j)}(t) - \overline{\mathbf{x}}(t))^{\mathrm{T}}, \quad \overline{\mathbf{x}}(t) = \frac{1}{K}\sum_{j=1}^{K}\mathbf{x}_t^{(j)}$$

is not accurate in this setting. A classic random matrix results (see [3]) indicates that if $\mathbf{x}_t^{(i)}$ are i.i.d. samples from $\mathcal{N}(0, C)$, the covariance estimation error $\|C - \widehat{C}\| = O(\sqrt{N/K})$. Therefore, the estimator is very inaccurate in the situations with the small samples and high dimensions, i.e., $K \ll N$. The intuition behind this mathematical phenomenon is that, spurious correlation may exist between each pair of components with a small probability because of the sampling effect, but there are $O(N^2)$ pairs of components. Thus, the resulting errors are very likely to be significant.

Nevertheless, if we are aware that covariance is local as described in Theorem 2.1, then the erroneous estimation can be made accurate. Note that the far-off-diagonal entries are close to zero. Thus, a natural way is to truncate these entries. In particular, with a given bandwidth L and covariance matrix C, we define the localization of C as

$$[C^L]_{i,j} = [C]_{i,j}\mathbf{1}_{\mathbf{d}(i,j)\leq L}.$$

It has been shown in [5] that such a localization technique can significantly improve the sampling accuracy. A simplified version in [53] shows that if $\mathbf{x}(t)$ is Gaussian distributed, then there exists a uniform constant c, so that for any $\epsilon > 0$,

$$\mathbb{P}(\|\widehat{C}^L - C^L\| > \epsilon) \leq 8\exp\left(2\log N - cK\min\left\{\frac{\epsilon}{2L\|C\|}, \frac{\epsilon^2}{4L^2\|C\|^2}\right\}\right). \tag{2.7}$$

In other words, if L is a fixed bandwidth, it requires only $L^2 O(\log N)$ many samples to find an accurate estimate of C^L.

On the other hand, one needs to notice that (2.7) indicates $\widehat{C}^L$ is a good estimator of C^L, but not necessarily C. In order for it to also be a good estimator of C, one needs $\|C - C^L\|$ to be small as well. Theorem 2.1 can guarantee this if there is only local interaction, i.e., $\mathbf{h} \equiv 0$. Since a covariance matrix's l_2-norm is bounded by its l_1-norm, we can plug in the estimate in Theorem 2.1 with $g(x) = x$ and find

$$\|C - C^L\| \leq \max_i \sum_{\mathbf{d}(i,j)>L} |[C]_{i,j}| \leq \frac{2C_{\beta,t}\mathrm{e}^{-\beta L}}{1 - \mathrm{e}^{-\beta}}. \tag{2.8}$$

In practice, given an error threshold $\epsilon > 0$, one can combine (2.8) and (2.7) together to find a bandwidth L and sample size $K = O(L^2\log N)$, such that the covariance estimation error $\|\widehat{C}^L - C\|$ is below 2ϵ with high probability.

Since localization involves using very simple numerical operation to significantly improve the estimation of the covariance matrix, it has been widely applied in all DA algorithms, in particular, the ensemble Kalman filter (see [17, 21, 24, 27, 40, 54–55, 57]). Moreover, adaptation of the localization strategy can be applied to Bayesian inverse problems where the associated covariance matrices follow the description of Theorem 2.1. In particular, it has been shown in [42] that blocking the high dimensional components, and then applying Gibbs sampler, the resulting Markov chain Monte Carlo algorithms will converge to the desired distribution with a rate independent of the dimension. One fundamental assumption of all these practical computational methods is that the underlying system is spatially localized, yet there has been no rigorous justification of this assumption. This article closes this gap for systems of type of (2.3).

2.3 Preliminaries

To facilitate our discussion, we will use the following notations: $\|v\|$ denotes the l_2 norm of v, while $\langle u, v \rangle = u^{\mathrm{T}} v$ is the inner product; $\mathbf{I}_q$ denotes the identity matrix of dimension q. Given a matrix A, the norm $\|A\|$ is the l_2 operator norm of A; $\lambda_{\max}(A)$ denotes the maximum eigenvalue of A and $\|A\|_F$ is its Frobenius norm. We use $[A]_{i,j}$ to denote the (i, j)-entry of A.

When we have a vector $\mathbf{x}$ of qN dimension, we often view it as the concatenation of N sub-blocks of dimension q. The decomposition can be written as $\mathbf{x} = (\mathbf{x}_1, \cdots, \mathbf{x}_N)$. Given a $qN \times qN$ matrix A, we write its (i, j)-th $q \times q$ sub-block as $A_{i,j}$ or $\{A\}_{i,j}$. The readers should not confuse these notations with the matrix entries $[A]_{i,j}$. In particular, we can write down the entries of $A_{i,j}$:

$$[A_{i,j}]_{m,n} = [\{A\}_{i,j}]_{m,n} = [A]_{(i-1)q+m,(j-1)q+n}, \quad i, j = 1, \cdots, N, \ m, n = 1, \cdots, q.$$

With a function f, ∇f denotes its Jacobian matrix. $\nabla_{\mathbf{x}_i} f(\mathbf{x})$ denotes the partial derivative with respect to $\mathbf{x}_i$, which is the i-th block-column of ∇f.

For two symmetric matrices $A \preceq B$ indicates that $B - A$ is positive semidefinite.

3 Analysis of Covariance Matrices

We provide the explicit statements of our result in this section. The detailed proofs are allocated in Appendix.

3.1 Linear covariance propagation

Consider the following multivariate SDE:

$$\mathrm{d}\mathbf{x}_i(t) = \mathbf{f}_i(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\mathrm{d}t + \frac{1}{N}\sum_{j=1}^{N} \mathbf{h}_i(t, \mathbf{x}_j)\mathrm{d}t + \Sigma \mathrm{d}\mathbf{w}_i(t), \quad i = 1, \cdots, N. \tag{3.1}$$

Our main result is established through a covariance comparison principle between (3.1) and a linear surrogate.

Theorem 3.1 *Suppose the covariance propagation of* (3.1) *is dominated by* $(F_*(t), H_*(t)) \in (\mathbb{R}^{N \times N}, \mathbb{R}^{N \times N})$ *in the following sense:*
(1) $\nabla_{\mathbf{x}_i} \mathbf{f}_i(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) + \nabla_{\mathbf{x}_i} \mathbf{f}_i(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})^{\mathrm{T}} \preceq 2[F_*(t)]_{i,i}\mathbf{I}$.
(2) $\|\nabla_{\mathbf{x}_{i-1}} \mathbf{f}_i(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\| \leq [F_*(t)]_{i,i-1}, \|\nabla_{\mathbf{x}_{i+1}} \mathbf{f}_i(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\| \leq [F_*(t)]_{i,i+1}$.
(3) $\frac{1}{N}\|\nabla_{\mathbf{x}_j} \mathbf{h}_i(t, \mathbf{x}_j)\| \leq [H_*(t)]_{i,j}$.
Let $G_*(s) = F_*(s) + H_*(s)$, *and*

$$Q(s) = \exp \Big(\int_{t-s}^{t} G_*(r)\mathrm{d}r \Big) \exp \Big(\int_{t-s}^{t} G_*(r)\mathrm{d}r \Big)^{\mathrm{T}}.$$

Then for any test function g with bounded gradient,

$$|\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \leq \sqrt{q}\|\nabla g\|_\infty^2 \Big(\|\Sigma_0^2\|_F [Q(t)]_{i,j} + \|\Sigma^2\|_F \int_0^t [Q(s)]_{i,j}\mathrm{d}s \Big). \tag{3.2}$$

This theorem shows that if the covariance propagation of (3.1) is dominated by $G_* = F_* + H_*$, then the spatial covariance of (3.1) is also dominated by the linear surrogate model associated with G_*:

$$\mathrm{d}\mathbf{x}_i = [G_*]_{i,i}\mathbf{x}_i\mathrm{d}t + [G_*]_{i,i-1}\mathbf{x}_{i-1}\mathrm{d}t + [G_*]_{i,i+1}\mathbf{x}_{i+1}\mathrm{d}t + \Sigma \mathrm{d}\mathbf{w}_i(t), \quad i = 1, \cdots, N. \tag{3.3}$$

To see that (3.3) has the spatial covariance structure as in (3.2), we concatenate (3.3) into one equation for $\mathbf{x}(t)$

$$\mathrm{d}\mathbf{x}(t) = D_{G_*}\mathbf{x}\mathrm{d}t + D_\Sigma \mathrm{d}\mathbf{w}(t), \quad \mathbf{x}(0) \sim \mathcal{N}(0, D_{\Sigma_0}).$$

Here, D_{G_*}, D_Σ and D_{Σ_0} are $qN \times qN$ dimensional matrices. Their $q \times q$ sub-blocks are given by

$$\{D_{G_*}\}_{i,j} = [G_*]_{i,j}\mathbf{I}_q, \quad \{D_\Sigma\}_{i,j} = \mathbb{1}_{i=j}\Sigma, \quad \{D_{\Sigma_0}\}_{i,j} = \mathbb{1}_{i=j}\Sigma_0.$$

Since this equation is linear, we can apply the Duhamel's formula and write the solution as

$$\mathbf{x}(t) = q(t)\mathbf{x}(0) + \int_0^t q(s)D_\Sigma \mathrm{d}\mathbf{w}(s), \quad q(s) := \exp\Big(\int_{t-s}^t D_{G_*}(r)\mathrm{d}r\Big).$$

By Itô's isometry,

$$\mathrm{cov}\,\mathbf{x}(t) = q(t)D_{\Sigma_0}D_{\Sigma_0}^{\mathrm{T}}q(t)^{\mathrm{T}} + \int_0^t q(s)D_\Sigma D_\Sigma^{\mathrm{T}}q(s)^{\mathrm{T}}\mathrm{d}s.$$

In particular, the covariance between $\mathbf{x}_i(t)$ and $\mathbf{x}_j(t)$ is given by $\{\mathrm{cov}\,\mathbf{x}(t)\}_{i,j}$. Since D_{Σ_0} and D_Σ are block diagonal, it is not difficult to establish (3.2) for system (3.3).

3.2 A representation of the covariance matrix

One difficulty in reaching Theorem 3.1 is how to compute the covariance matrix for a multivariate SDE. We will establish a framework for covariance matrix computation, starting from a simple Stein's lemma type of formula:

Lemma 3.1 *Suppose X, Y are independent standard Gaussian random variables of dimension d, f and g are smooth functions with bounded first order derivatives. Let $X^\theta = \cos\theta\,X + \sin\theta\,Y$, then*

$$\mathrm{cov}(f(X), g(X)) = \mathbb{E}f(X)g(X) - f(X)g(Y) = \int_0^{\frac{\pi}{2}} \sin\theta\mathbb{E}\langle\nabla f(X), \nabla g(X^\theta)\rangle\mathrm{d}\theta.$$

In order to apply Lemma 3.1, we need to represent $\mathbf{x}(t)$ as a function of certain Gaussian random variables. These Gaussian random variables are the increments of the Wiener processes driving $\mathbf{x}(t)$. To setup this connection, recall that one way to simulate (3.1) is to perturb its ordinary differential equation (ODE for short) version with random noises. In particular, consider the ODE system:

$$\mathrm{d}\mathbf{x}_i(t) = \mathbf{f}_i(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\mathrm{d}t + \frac{1}{N}\sum_{j=1}^N \mathbf{h}_i(t, \mathbf{x}_j)\mathrm{d}t, \quad i = 1, \cdots, N. \tag{3.4}$$

Let $\Psi_h : \mathbb{R}^{qN} \mapsto \mathbb{R}^{qN}$ be the mapping from $\mathbf{x}(t)$ to $\mathbf{x}(t+h)$. We suppress the dependence of Ψ_h on t, for notational simplicity. Then one way to simulate (3.1) is generate the following iterates

$$\mathbf{X}(0) = m_0 + \Sigma_0\mathbf{W}_0, \quad \mathbf{X}(n+1) = \Psi_h(\mathbf{X}(n)) + \Sigma_h\mathbf{W}_n, \quad \Sigma_h = \sqrt{h}D_\Sigma. \tag{3.5}$$

Then $\mathbf{X}(T)$ is a numerical approximation of $\mathbf{x}(t)$ if $Th = t$ when T is large. A standard verification lemma of this statement is provided in Appendix. We write the i-th block of $\mathbf{X}(n)$ as $\mathbf{X}_i(n) \in \mathbb{R}^q$. $\mathbf{W}_n$ are qN dimensional standard Gaussian noise for $n = 0, 1, \cdots, T$.

The strategy here is to consider $\mathbf{X}(T)$ instead of $\mathbf{x}(t)$, because its dependence on the noise realization $\mathbf{W}_i$ is easier to handle using Lemma 3.1. We can show a similar result for $\mathbf{X}(T)$ as

in Theorem 3.1. Then in the limit $h \to 0$, we have our desired results. The details are shown in Appendix. As a remark, it might be possible and more elegant to represent the covariance of $\mathbf{x}_t$ directly using tools from Malliavin calculus (see [43]), and avoid the procedures of discretization and taking continuous limit. Yet the current proof strategy can be easier to understand for an applied math audience.

3.3 Diffusion and mean field interaction

While (3.2) provides a general upper bound, from its formulation is not easy to see the covariance decay we expect. To make the covariance decay explicit, we try to use instead a homogenous linear surrogate. We update the constant definition in (2.4) to the spatial inhomogeneous setting:

$$
\begin{aligned}
\lambda_0 &:= \sup_{i,\mathbf{x},s \leq t} \{\lambda_{\max}(\nabla_{\mathbf{x}_i} \mathbf{f}_i(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) + \nabla_{\mathbf{x}_i} \mathbf{f}_i(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})^{\mathrm{T}})\}, \\
\lambda_F &:= \sup_{i,\mathbf{x},s \leq t} \{\|\nabla_{\mathbf{x}_{i-1}} \mathbf{f}_i(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\|, \|\nabla_{\mathbf{x}_{i+1}} \mathbf{f}_i(s, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1})\|\}, \qquad (3.6) \\
\lambda_H &:= \sup_{i,j,\mathbf{x},s \leq t} \|\nabla_{\mathbf{x}_j} \mathbf{h}_i(s, \mathbf{x}_j)\|.
\end{aligned}
$$

They are finite under the conditions of Theorem 3.1. We can build a homogeneous linear surrogate by letting

$$
[F_*]_{i,i} = \lambda_0, \quad [F_*]_{i,i\pm 1} = \lambda_F, \quad [H_*]_{i,j} = \frac{1}{N}\lambda_H.
$$

Then the linear surrogate (3.3) corresponds

$$
\mathrm{d}\mathbf{x}_i = (\lambda_0 + 2\lambda_F)\mathbf{x}_i \mathrm{d}t + \lambda_F \Delta \mathbf{x}_i \mathrm{d}t + \frac{\lambda_H}{N} \sum_{j=1}^{N} \mathbf{x}_j + \Sigma \mathrm{d}\mathbf{w}_i(t), \quad i = 1, \cdots, N, \qquad (3.7)
$$

where the discrete Laplacian is given by

$$
\Delta \mathbf{x}_i := \mathbf{x}_{i-1} + \mathbf{x}_{i+1} - 2\mathbf{x}_i,
$$

which is closely related to the heat equation.

The spatial covariance of $\mathbf{x}(t)$ is built up by the diffusion effect of $\Delta \mathbf{x}_i$ and the mixing effect of mean field interaction $\frac{1}{N} \sum_{j=1}^{N} \mathbf{x}_j$. These two effects operate in different ways. The diffusion propagates the diagonal entries in the covariance matrix to the off-diagonal ones, it contributes to the exponential decay term $\mathrm{e}^{-\beta \mathbf{d}(i,j)}$ in Theorem 2.1. The mean field interaction synchronizes the components, and contributes to a global covariance of scale $\frac{1}{N}$.

We can establish the following explicit bounds in terms of $\lambda_G, \lambda_F, \lambda_H$, instead of the matrix $Q(t)$.

Theorem 3.2 *Under the same conditions of Theorem 3.1, for any $\beta > 0$, let*

$$
\lambda_\beta = \lambda_0 + \lambda_F(\mathrm{e}^\beta + \mathrm{e}^{-\beta}), \quad \eta_\beta = \lambda_0 + \lambda_F(\mathrm{e}^\beta + \mathrm{e}^{-\beta}) + \lambda_H
$$

with λ_0, λ_F and λ_H defined by (3.6). Then with any function g,

$$
\begin{aligned}
&|\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \\
&\leq 2\sqrt{q}\|\nabla g\|_\infty^2 (\mathrm{e}^{\lambda_\beta t}\|\Sigma_0^2\|_F + (\mathrm{e}^{\lambda_\beta t} - 1)\|\Sigma^2\|_F/\lambda_\beta)\mathrm{e}^{-\beta \mathbf{d}(i,j)} \\
&\quad + \frac{2\sqrt{q}(1 + \mathrm{e}^{-\beta})\|\nabla g\|_\infty^2}{(1 - \mathrm{e}^{-\beta})N}\left((\mathrm{e}^{\eta_\beta t} - \mathrm{e}^{\lambda_\beta t})\|\Sigma_0^2\|_F + \left(\frac{\mathrm{e}^{\eta_\beta t} - 1}{\eta_\beta} - \frac{\mathrm{e}^{\lambda_\beta t} - 1}{\lambda_\beta}\right)\|\Sigma^2\|_F\right).
\end{aligned}
$$

When there is only mean field interaction, i.e., $\lambda_F = 0$, the result can be simplified by choosing $\beta \to \infty$,

$$
\begin{aligned}
&|\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \\
&\le \frac{2\sqrt{q}\|\nabla g\|_\infty^2}{N}\Big((\mathrm{e}^{(\lambda_0+\lambda_H)t} - \mathrm{e}^{\lambda_0 t})\|\Sigma_0^2\|_F + \Big(\frac{\mathrm{e}^{(\lambda_0+\lambda_H)t}-1}{\lambda_0+\lambda_H} - \frac{\mathrm{e}^{\lambda_0 t}-1}{\lambda_0}\Big)\|\Sigma^2\|_F\Big).
\end{aligned}
\tag{3.8}
$$

When there is only diffusion effect, that is $\mathbf{h}_i \equiv 0$, the result can be simplified as

$$
|\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \le 2\sqrt{q}\|\nabla g\|_\infty^2(\mathrm{e}^{\lambda_\beta t}\|\Sigma_0^2\|_F + (\mathrm{e}^{\lambda_\beta t} - 1)\|\Sigma^2\|_F/\lambda_\beta)\mathrm{e}^{-\beta\mathbf{d}(i,j)}.
\tag{3.9}
$$

It is worth noticing the result holds for all $\beta > 0$. This indicates the covariance structure actually decays faster than exponential. In practice, one can try to minimize the right hand side with respect to β, or directly compute the covariance of (3.7) to find the exact local covariance structure. While sharper theoretical upper bound maybe obtainable by applying formulas as in [28], they are complicated and work only asymptotically. Our current upper bound is simpler, and sufficient for applications discussed in Subsections 2.1 and 2.2.

With Theorem 3.2, it is straightforward to show the spatial averaging estimator is consistent:

Corollary 3.1 *Under the same settings of Theorem* 3.2, *the following bound holds for the block estimator* (2.5):

$$
\begin{aligned}
\mathrm{var}\widehat{g}_N \le{}& \frac{2\|\nabla g\|_\infty^2}{(1-\mathrm{e}^{-\beta})N}(2\mathrm{e}^{2\lambda_\beta t}\|\Sigma_0^2\|_F + (\mathrm{e}^{2\lambda_\beta t} - 1)\|\Sigma^2\|_F/\lambda_\beta) \\
&+ \frac{2\|\nabla g\|_\infty^2}{(1-\mathrm{e}^{-\beta})N}\Big((\mathrm{e}^{\eta_\beta t} - \mathrm{e}^{\lambda_\beta t})\|\Sigma_0^2\|_F + \Big(\frac{\mathrm{e}^{\eta_\beta t}-1}{\eta_\beta} - \frac{\mathrm{e}^{\lambda_\beta t}-1}{\lambda_\beta}\Big)\|\Sigma^2\|_F\Big).
\end{aligned}
\tag{3.10}
$$

Proof Since in (2.6), $\sum_{j=1}^{N} \mathrm{e}^{-\beta\mathbf{d}(1,j)} \le 2\sum_{j=0}^{\infty} \mathrm{e}^{-\beta j} = \frac{2}{1-\mathrm{e}^{-\beta}}$.

3.4 Long time stability

In many applications, the long time $t \to \infty$ scenario is of interest. For example, if we want to find the equilibrium measure of $\mathbf{x}(t)$, we often simulate $\mathbf{x}(t)$ for a long time, and use its distribution as an approximate. Likewise, if we are interested in data assimilation with models as (2.3), the algorithms are usually iterated for an extended period. In order to apply the spatial averaging strategy in Section 2.1, and the localization strategy in Section 2.2 for these operations, we need the local covariance structure to be stable in time.

In the view of Theorem 3.2, we can guarantee the local covariance structure is stable if $\lambda_0 + 2\lambda_F + \lambda_H < 0$. In particular we have the following corollary by letting $t \to \infty$ in the estimate.

Corollary 3.2 *If $\lambda_0 + \lambda_H + 2\lambda_F < 0$, then there is a $\beta_0 > 0$, so that if $0 < \beta < \beta_0$, $\lambda_\beta \le \eta_\beta < 0$ we have*

$$
\begin{aligned}
&\limsup_{t\to\infty} |\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \\
&\le -\sqrt{q}\|\nabla g\|_\infty^2\|\Sigma^2\|_F\Big(\frac{2}{\lambda_\beta}\mathrm{e}^{-\beta\mathbf{d}(i,j)} - \frac{2(1+\mathrm{e}^{-\beta})}{(1-\mathrm{e}^{-\beta})N}(\eta_\beta^{-1} - \lambda_\beta^{-1})\Big).
\end{aligned}
$$

4 Test Examples

In this section, we demonstrate our theoretical findings with several numerical test examples.

4.1 A linear model

We start with a linear model, where the analytic solutions associated with this linear model can be written down explicitly (See Appendix B). Therefore, this model is able to provide insights and validation of the theories developed in the previous sections.

The linear model is as follows:

$$\frac{\mathrm{d}u_i}{\mathrm{d}t} = -au_i + d_u(u_{i+1} - 2u_i + u_{i-1}) + w(\overline{u} - u_i) + \sigma_u \dot{W}_{u_i} \quad \text{for } i = 1, \cdots, N \tag{4.1}$$

with periodic boundary conditions. Here $\overline{u}$ is the averaged value of all the state variables,

$$\overline{u} = \frac{1}{N} \sum_{i=1}^{N} u_i.$$

In the linear model (4.1), the coefficients of damping a, diffusion d_u, mean field interaction w and the stochastic noise σ_u are all constants. Thus, the flow field is statistically homogeneous. Below, we fix

$$a = 1 \quad \text{and} \quad \sigma_u = 0.5 \tag{4.2}$$

while the diffusion and mean field interaction coefficients d_u and w vary in different cases.

In the linear model (4.1), it is easy to write down the constants in (3.6),

$$\lambda_0 = -a - 2d_u - w, \quad \lambda_F = d_u, \quad \lambda_H = w. \tag{4.3}$$

In addition, only one-layer model is utilized here and we consider the covariance of different u_i themselves. Therefore

$$q = 1 \quad \text{and} \quad \nabla g = 1. \tag{4.4}$$

Furthermore, $\|\Sigma^2\|_F = \sigma_u^2$. Below, we study the role of the mean field interaction and the diffusion. All the results shown in this subsection are computed using the exact solution given in Appendix B, which allows us to understand Theorem 3.2 without being interfered by numerical errors.

(1) Linear model with only mean field interaction First, we study the case with only mean field interaction. In other words, we set the diffusion coefficient $d_u = 0$. The constants in (4.3) reduce to

$$\lambda_0 = -a - w, \quad \lambda_F = 0, \quad \lambda_H = w. \tag{4.5}$$

According to (3.8) in Theorem 3.2, the covariance bound is proportional to $\frac{1}{N}$ and the coefficient is provided by letting $\beta \to \infty$. Therefore, assuming the initial covariance is zero, we have the following result

$$|\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \leq 2\Big(\frac{\mathrm{e}^{-at} - 1}{-a} - \frac{\mathrm{e}^{-(a+w)t} - 1}{-a - w}\Big)\sigma_u^2 \frac{1}{N}. \tag{4.6}$$

In Figure 1, we show the covariance between u_1 and u_i for $i = 1, \cdots, N$ at $t = 5$. Here the initial conditions are zero everywhere. Different columns show the covariance with different N. The spatiotemporal simulations are also included to provide intuitions. Without the diffusion and the initial covariance, the covariance between u_1 and all other u_i with $i \neq 1$ has the same non-zero value due to the global effect from the mean field interaction.
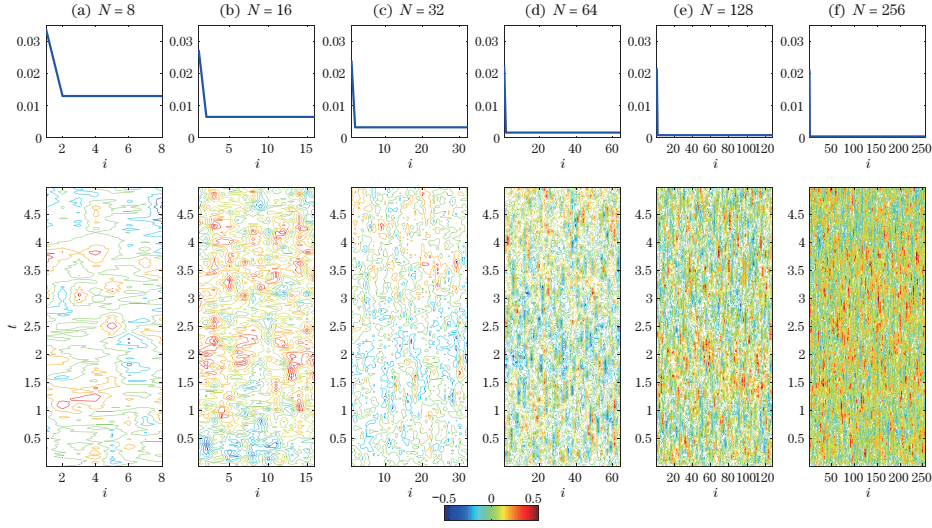
Figure 1 Linear model with only mean field interaction. The covariance between u_1 and u_i for $i = 1, \cdots, N$ at $t = 5$. The spatiotemporal simulations are also included.
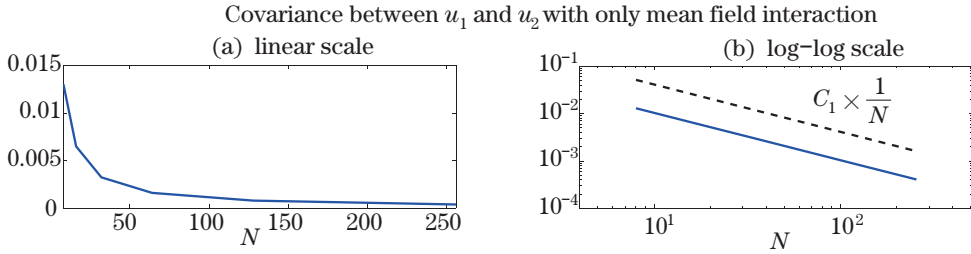

Figure 2 Linear model with only mean field interaction. The dependence of the covariance between u_1 and u_2 as a function of N. The black dashed line in panel (b) (log-log scale) shows the bound in (4.6), where the constant C_1 is given by the coefficient on the right hand side of (4.6) in front of $\frac{1}{N}$.

In Figure 2, the dependence of the covariance between u_1 and u_2 as a function of N is shown. It is clear that the covariance decays as a function of $\frac{1}{N}$, which validates the results in Theorem 3.2 (and that in (4.6)). The black dashed line in panel (b) (log-log scale) shows the bound in (4.6), where the constant C_1 is given by the coefficient on the right hand side of (4.6) in front of $\frac{1}{N}$.

(2) Linear model with only diffusion Next, we study the case with only diffusion. In other words, we set the mean field interaction coefficient $w = 0$. The constants in (4.3) reduce to

$$\lambda_0 = -a - 2d_u, \quad \lambda_F = d_u, \quad \lambda_H = 0. \tag{4.7}$$

According to (3.9) in Theorem 3.2, the covariance bound is proportional to $\mathrm{e}^{-\beta \mathbf{d}(i,j)}$. Again, assuming the initial covariance is zero, we have the following result

$$|\mathrm{cov}(g(\mathbf{x}_i(t)), g(\mathbf{x}_j(t)))| \le \frac{2(\mathrm{e}^{\lambda_\beta t} - 1)\sigma_u^2}{\lambda_\beta} \mathrm{e}^{-\beta \mathbf{d}(i,j)}, \tag{4.8}$$

where

$$\lambda_\beta = \lambda_0 + \lambda_F(e^\beta + e^{-\beta}).$$

In Figure 3, we show the covariance between u_1 and u_i for $i = 1, \cdots, N$ at $t = 5$ with $N = 64$ fixed. Here the initial conditions are zero everywhere. Different columns show the covariance with different strengths of the diffusion coefficient d_u. The spatiotemporal simulations are also included to provide intuitions. Without the mean field interaction and the initial covariance, the covariance between u_1 and u_i for u_i being far from u_1 decays to zero, reflecting the local contribution due to only the diffusion (without mean field interaction). On the other hand, the increase of d_u allows the increase of the covariance between u_1 and u_i for i far from 1. This is clearly illustrated in the spatiotemporal patterns and is consistent with our intuition as well.

Figure 4 shows the covariance between u_1 and u_i for $i = 1, \cdots, \frac{N}{2}$ with $d_u = 20$ and $N = 64$ at $t = 5$. In panel (b), the logarithm scale is shown, which indicates a linear dependence of the log covariance on the distance $\mathbf{d}(1, i)$. This validates the exponential decay of the covariance as a function of the distance as shown in (4.8). The black dashed line shows the bound in (4.8), where the constant C_2 is given by the coefficient on the right hand side of (4.8) in front of $e^{-\beta\mathbf{d}(i,j)}$ by taking $\beta = \frac{1}{5}$. We have tested other β, with which the bounds all above the actual covariance. Therefore, we validate the theoretical results with only diffusion, described by (3.9).
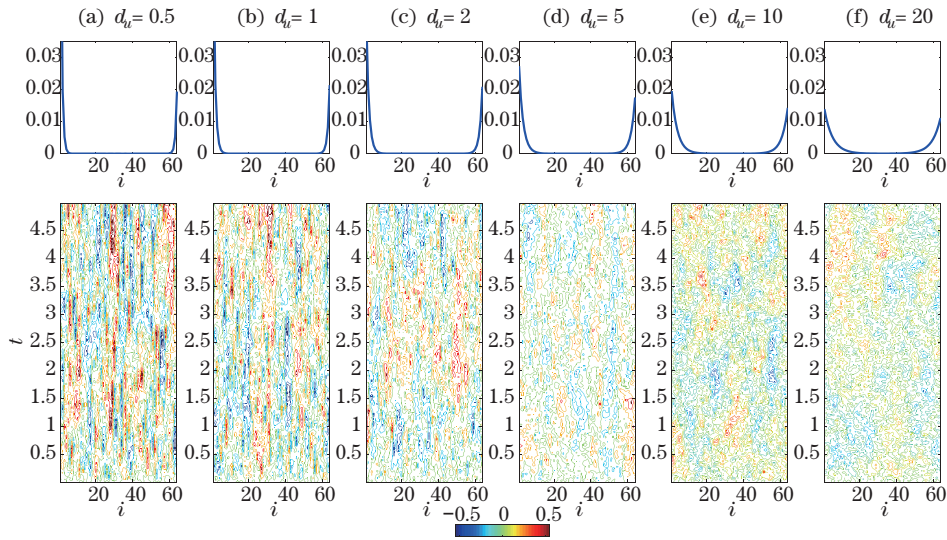

Figure 3　Linear model with only diffusion. The covariance between u_1 and u_i for $i = 1, \cdots, N$ at $t = 5$ with $N = 64$ fixed.

(3) Linear model with both the mean field interaction and diffusion Finally, we study the case with both mean field interaction and diffusion. Figure 5 is similar to Figure 3 but with a non-zero mean field interaction $w = 5$. Therefore, unlike the results in Figure 5, the covariance between u_1 and u_i where i is far from 1 is non-zero due to the mean field interaction that has a global impact. In Figure 6, it is also clear that when i increases the covariance between u_i and u_1 first experience an exponential decay and then the covariance remains as a constant. Thus, the exponential decay is the main contribution to the covariance behavior for grid points that are close to $i = 1$ while the mean field interaction plays the dominant role at the location that is far away. These simulations are consistent with the theoretical results shown in Theorem 3.2.
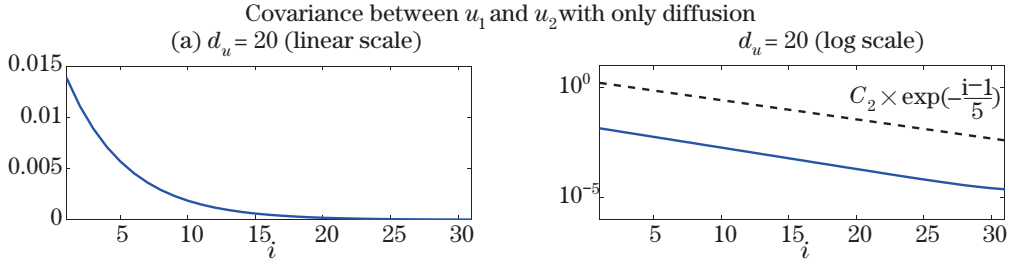
Covariance between u_1 and u_2 with only diffusion

Figure 4 Linear model with only diffusion. The covariance between u_1 and u_i for $i = 1, \cdots, \frac{N}{2}$ with $d_u = 20$ and $N = 64$ at $t = 5$. The black dashed line in panel (b) (logarithm scale) shows the bound in (4.8), where the constant C_2 is given by the coefficient on the right hand side of (4.8) in front of $\mathrm{e}^{-\beta \mathbf{d}(i,j)}$ by taking $\beta = \frac{1}{5}$.

4.2 A stochastically coupled FHN model

In this subsection, we explore the covariance structure in a stochastically coupled FitzHugh-Nagumo (FHN for short) model, which is a nonlinear model that describes activation and deactivation dynamics of spiking neurons (see [8, 33]). The model reads,

$$\epsilon \mathrm{d}u_i = \left(u_i - \frac{1}{3}u_i^3 - v_i + d_u(u_{i+1} - 2u_i + u_{i-1}) + w(\overline{u} - u_i)\right)\mathrm{d}t + \sqrt{\epsilon}\delta_1 \mathrm{d}W_{u_i},$$
$$\mathrm{d}v_i = (u_i + a)\mathrm{d}t + \delta_2 \mathrm{d}W_{v_i}, \quad i = 1, \cdots, N, \tag{4.9}$$

where u_i and v_i are activator and inhibitor variables. Periodic boundary conditions are imposed on u_i variables. In (4.9), the timescale ratio $\epsilon = 0.01 \ll 1$ leads to a slowfast structure of the model. The parameter $a = 1.05 > 1$ such that the system has a global attractor in the absence of noise and diffusion (see [20]). The random noise is able to drive the system above the threshold level of global stability and triggers limit cycles intermittently. Note that with $N = 1$, the model reduces to the classical FHN model with a single neuron and it contains the model families with both coherence resonance and self-induced stochastic resonance (see [16]). With different choices of the noise strength δ_1, δ_2 and the diffusion coefficient d_u, the system in (4.9) exhibits rich dynamical behaviors. Finally, in (4.9), the variable $\overline{u} = \frac{1}{N}\sum_{i=1}^{N} u_i$ is the mean field interaction.

The goal of this subsection is to study the spatial structure of the covariance due to both the local effect (i.e., diffusion) and the global effect (i.e., mean field interaction).

4.2.1 Basic properties

For the convenience of stating the theoretical results, we slightly change of notation of (4.9) in this paragraph,

$$\mathrm{d}u_i = \left(\frac{1}{\epsilon}u_i - \frac{1}{3\epsilon}u_i^3 - \frac{1}{\sqrt{\epsilon}}v_i^\epsilon + \frac{d_u}{\epsilon}(u_{i+1} + u_{i-1} - 2u_i) + \frac{\widetilde{w}}{\epsilon}(\overline{u} - u_i)\right)\mathrm{d}t + \frac{1}{\sqrt{\epsilon}}\delta_1 \mathrm{d}W_{u_i},$$
$$\mathrm{d}v_i^\epsilon = \left(\frac{1}{\sqrt{\epsilon}}u_i + \frac{a}{\sqrt{\epsilon}}\right)\mathrm{d}t + \frac{\delta_2}{\sqrt{\epsilon}}\mathrm{d}W_{v_i}. \tag{4.10}$$

To apply the framework we developed, it is natural to let $\mathbf{x}_i = (u_i, v_i^\epsilon)^{\mathrm{T}}$ such that $q = 2$, and

$$\nabla_{\mathbf{x}_i}\mathbf{f}(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) = \begin{bmatrix} \epsilon^{-1} - \epsilon^{-1}u_i^2 - 2\epsilon^{-1}d_u - \epsilon^{-1}\widetilde{w} & -\epsilon^{-\frac{1}{2}} \\ \epsilon^{-\frac{1}{2}} & 0 \end{bmatrix} \tag{4.11}$$

(a) $d_u = 0.5, w = 5$ (b) $d_u = 1, w = 5$ (c) $d_u = 2, w = 5$ (d) $d_u = 5, w = 5$ (e) $d_u = 10, w = 5$ (f) $d_u = 20, w = 5$
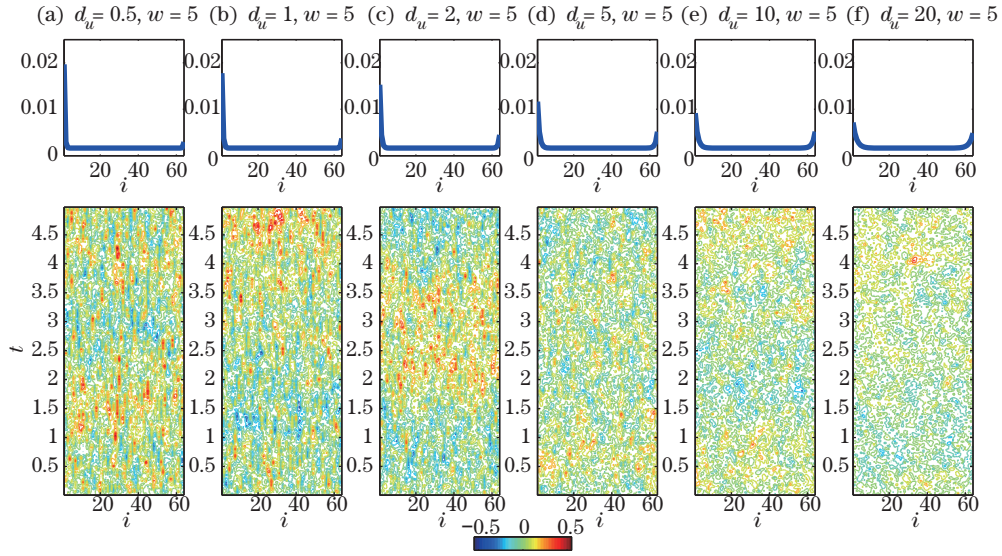

Figure 5 Linear model with both mean field interaction and diffusion. The covariance between u_1 and u_i for $i = 1, \cdots, N$ at $t = 5$ with $N = 64$ fixed and the mean field interaction $w = 5$.

Covariance between u_1 and u_2 with both mean field interaction and diffusion
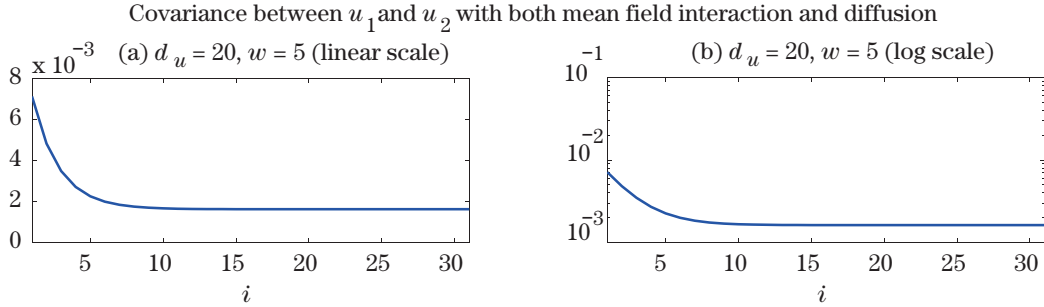

Figure 6 Linear model with both mean field interaction and diffusion. The covariance between u_1 and u_i for $i = 1, \cdots, \frac{N}{2}$ with $d_u = 20, w = 5$ and $N = 64$ at $t = 5$. The local interaction is the main contributor to the covariance behavior at the beginning while the mean field interaction plays the dominant role at the location that is far away.

and

$$\nabla_{\mathbf{x}_{i\pm1}}\mathbf{f}(t, \mathbf{x}_{i-1}, \mathbf{x}_i, \mathbf{x}_{i+1}) = \begin{bmatrix} \epsilon^{-1}d_u & 0 \\ 0 & 0 \end{bmatrix}, \quad \nabla_{\mathbf{x}_i}h = \begin{bmatrix} \widetilde{w} & 0 \\ 0 & 0 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \epsilon^{-\frac{1}{2}}\delta_1 & 0 \\ 0 & \epsilon^{-\frac{1}{2}}\delta_2 \end{bmatrix}.$$

Since the $-\epsilon^{-1}u_i^2$ is always negative, we can take

$$\lambda_0 = \epsilon^{-1}\max\{1 - 2d_u - \widetilde{w}, 0\}, \quad \lambda_F = \epsilon^{-1}d_u, \quad \lambda_H = \epsilon^{-1}\widetilde{w}.$$

Therefore, using the spatial averaging sampling strategy for high dimensional systems, we have the following corollary.

Corollary 4.1 *The block average estimator is consistent for the FHN system* (4.10). *In particular, estimate* (3.10) *holds with* $\|\Sigma^2\|_F = \frac{\sqrt{\delta_1^4 + \delta_2^4}}{\epsilon}$, *and*

$$\eta_\beta = \epsilon^{-1} \max\{1 - 2d_u, \widetilde{w}\} + 2\epsilon^{-1} d_u(e^\beta + e^{-\beta}), \quad \lambda_\beta = \eta_\beta - \epsilon^{-1}\widetilde{w}.$$

In particular, from the bound (3.10), we learn that the estimator consistency deteriorates with larger t, d_u and $\widetilde{w}$, which correspond to longer simulation time, stronger local interaction, and stronger mean field interaction regimes.

4.2.2   The FHN model with different diffusion and mean field interaction coefficients

(1) FHN model with only diffusion First, we focus on the situation in the FHN model where the mean field interaction is zero ($w = 0$). Therefore, the covariance comes purely from the local effect, namely the diffusion. We set $\delta_1 = \delta_2 = 0.4$ but choose different values of the diffusion parameter d_u which provides different dynamical behavior:

$$\begin{aligned}
&\text{(a) Strongly mixed regime:} && d_u = 0.02; \\
&\text{(b) Weakly coherent regime:} && d_u = 0.5; \\
&\text{(c) Strongly coherent regime:} && d_u = 10.
\end{aligned} \qquad (4.12)$$

The simulation of the model in these three dynamical regimes is shown in Figure 7. Figure 8 shows the covariance between u_1 and u_i for $i = 1, \cdots, N$ at $t = 5$. Here we use a Monte Carlo simulation with $M = 8192$ samples at each grid points (namely, running the Monte Carlo for M times). With the increase of d_u from Regime (a) to Regime (c), the spatial covariance of u increases. In particular, in Regime (c) with a strong diffusion, the covariance between u_1 and all other u_i for $i = 1, \cdots, N$ remains significant, and is above 0.5. Note that there is almost no mixing of the underlying stochastic system with such a large diffusion coefficient in Regime (c) (See Figure 7), which may cause some issues of the controllability of the system. In fact, with a large diffusion coefficient, the role of the stochastic forcing becomes relatively weaker and the control of the system via the external stochastic forcing becomes more difficult. This also leads to some potential issues in applying the spatial averaging strategy for computing the covariance, especially for long time. Note that although Corollary 4.1 is always valid, the constant η_β becomes very large with a large diffusion coefficient (due to the second term) and thus the consistency becomes nearly degenerated very quickly. See the remarks below Corollary 4.1 and the statement of the bound in (3.10). The numerical results using the spatial averaging strategy will be illustrated in the next subsection. On the other hand, the spatial covariance of v changes in the same fashion as u but is weaker. This is because the diffusion appears in the u equations and it only has an indirect impact on v. In all three regimes, the spatial covariance decays in an exponential rate (see the subpanels). All these are consistent with the result in (3.9) of Theorem 3.2.

(2) FHN model with only mean field interaction Now we study the situation in the FHN model where the diffusion is zero ($d_u = 0$). Therefore, the covariance comes purely from the global effect (i.e., the mean field interaction). We set again $\delta_1 = \delta_2 = 0.4$ but choose different values of the mean field interaction parameter w which provides different dynamical behavior:

$$\begin{aligned}
&\text{(a) Weak mean field interaction regime:} && w = 0.1; \\
&\text{(b) Moderate mean field interaction regime:} && w = 0.3; \\
&\text{(c) Strong mean field interaction regime:} && w = 0.5.
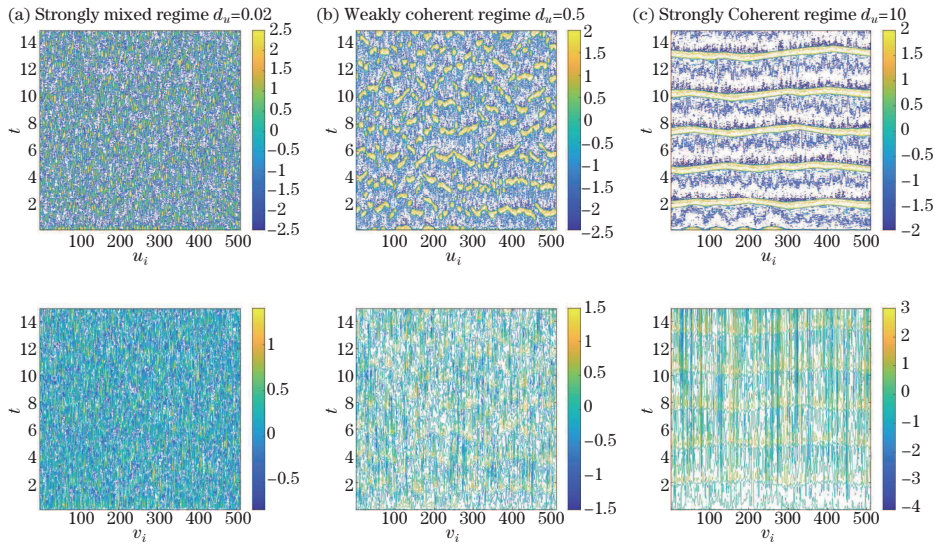\end{aligned} \qquad (4.13)$$

Figure 7 Simulations of the FHN model (4.9) with only diffusion but no mean field inter-action ($w = 0$) in (a) Strongly mixed regime, (b) Weakly coherent regime, and (c) Strongly coherent regime. The first row is the simulation of u_i and the second is v_i. Here $N = 512$.
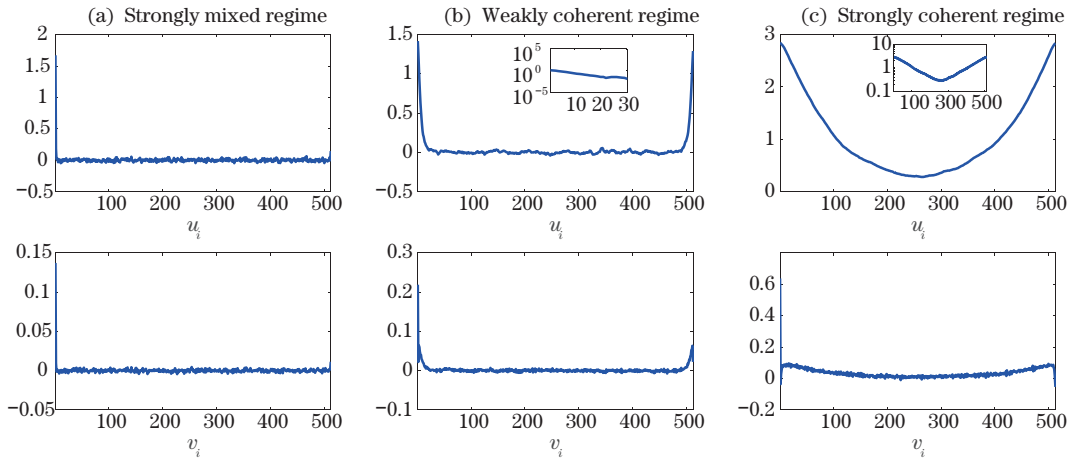

Figure 8 Covariance between u_1 and u_i for $i = 1, \cdots, N$ in the FHN model (4.9) with only diffusion but no mean field interaction ($w = 0$) in the three regimes described in Figure 7 at time $t = 5$. The first row is the simulation of u_i and the second is v_i. The subpanels show the covariance in a logarithm scale.

The simulations of the three regimes are shown between Figure 9. Again, these results are computed using a Monte Carlo simulation with $M = 8192$ samples at each grid points. In the moderate mean field interaction regime (Regime (b)), some global effects can already been seen. In the strong mean field interaction regime (Regime (c)), it is clear that the variables at different grid points are highly synchronized with each other, which is consistent with the theoretic analysis described in the previous section.

Note the difference in Figures 7 and 9. In Figure 7 the diffusion leads to the development

of local interaction and extends it to the global scale. In Figure 9 the mean field interaction starts directly with a change in the global scale and such a change becomes significant with the increase of the mean field interaction strength w. Again, the constant η_β in Corollary 4.1 with a large mean field interaction coefficient can become very large (due to the first term) and thus the consistency becomes nearly degenerated.

Figure 10 shows the covariance $\mathrm{Cov}(u_1, u_2)$ and $\mathrm{Cov}(v_1, v_2)$ as a function of N in two different regimes with either moderate or strong mean field interactions. With the increase of N, the covariance decreases and the decrease rate is proportion to $\frac{1}{N}$ (see the log-log plot), which is consistent with the theoretical conclusion in (3.8) of Theorem 3.2.
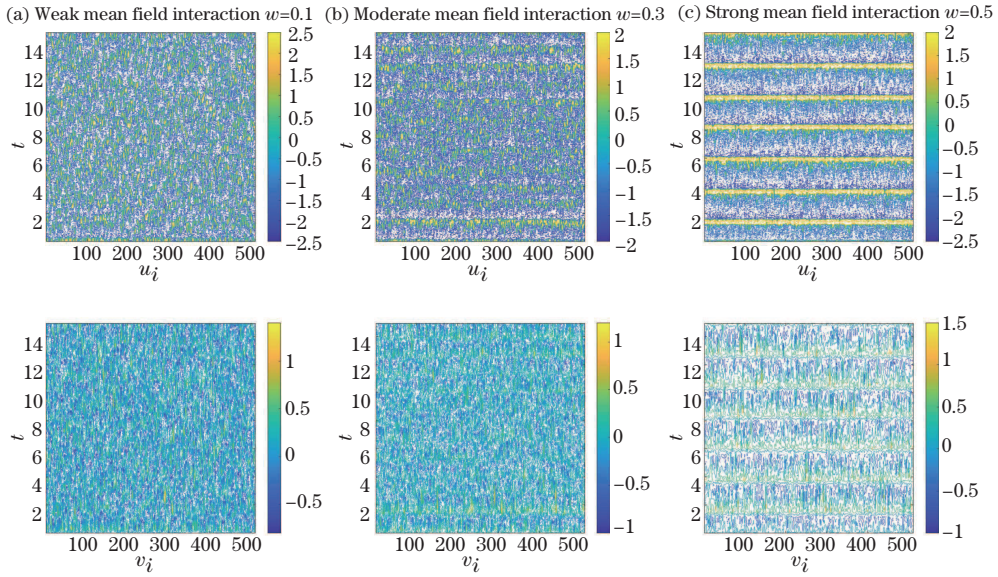

Figure 9   Simulations of the FHN model (4.9) with only mean field interaction but no diffusion $d_u = 0$ in (a) Weak mean field interaction regime, (b) Moderate mean field interaction regime, and (c) Strong mean field interaction regime. The first row is the simulation of u_i and the second is v_i. Here $N = 512$.

4.2.3   Recovery of the covariance using the spatial averaging strategy

In this subsection, we compare the covariance computed from the spatial averaging strategy with the truth in order to study the consistency of the spatial averaging strategy. Here the truth is formed by using direct Monte Carlo simulations.

Recall in Subsection 2.1, the spatial averaging strategy makes use of the statistical symmetry of the system. Comparing with the covariance computed by using the direct Monte Carlo simulations with a large sample size, the spatial averaging strategy requires only a small number of repeated experiments to reach the same accuracy (providing that the consistency is valid). In fact, due to the spatial homogeneity, the covariance between u_1 and u_2 is the same as that between u_i and u_{i+1} for all $i = 1, \cdots, N$. Therefore, regarding $u_1, u_2, \cdots, u_N$ as a sample set, the covariance between u_1 and u_2 can be computed by taking the covariance between $u_1, u_2, \cdots, u_N$ and its shifted sample set $u_2, u_3, \cdots, u_N, u_1$. Note that if M repeat experiments
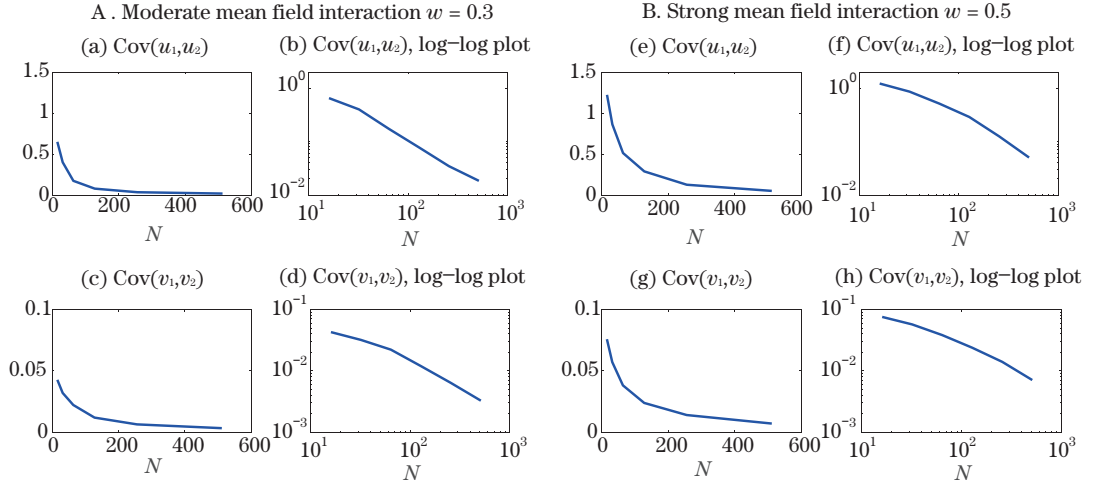
A . Moderate mean field interaction $w = 0.3$

(a) Cov(u_1,u_2)

(b) Cov(u_1,u_2), log–log plot

B. Strong mean field interaction $w = 0.5$

(e) Cov(u_1,u_2)

(f) Cov(u_1,u_2), log–log plot

(c) Cov(v_1,v_2)

(d) Cov(v_1,v_2), log–log plot

(g) Cov(v_1,v_2)

(h) Cov(v_1,v_2), log–log plot

Figure 10 Covariance between u_1 and u_2 (panel (a)) and between u_1 and u_2 (panel (c)) as a function of N in the regime with moderate mean field interaction $w = 0.3$ and $d_u = 0$. Panels (b) and (d) show the same results but are plotted in the log-log scale in order to illustrate the decay rate $\frac{1}{N}$ (in log-log scale, this corresponds to a linear decay rate). Panels (e)–(h) are similar but are for the regime with strong mean field interaction $w = 0.5$ and $d_u = 0$.
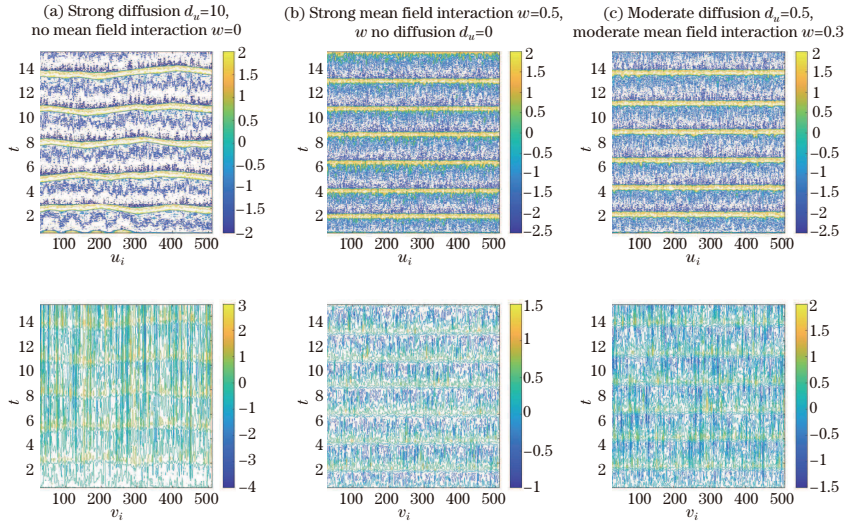
(a) Strong diffusion d_u=10, no mean field interaction w=0

(b) Strong mean field interaction w=0.5, w no diffusion d_u=0

(c) Moderate diffusion d_u=0.5, moderate mean field interaction w=0.3

Figure 11 Spatiotemporal patterns of different dynamical regimes: (a) Strong diffusion $d_u = 10$ and no mean field interaction $w = 0$; (b) Strong mean field interaction $w = 0.5$ and no diffusion $d_u = 0$; and (c) Moderate diffusion $d_u = 0.5$ and moderate mean field interaction $w = 0.3$.

are combined (namely, M Monte Carlo simulations) with the spatial averaging strategy, then the effective sample size is $M_{eff} = MN$. One potential issue of the spatial averaging strategy is that the samples $u_1, u_2, \cdots, u_N$ are not independent with each other. Therefore, when these samples are strongly correlated, then the spatial averaging strategy in computing the covariance
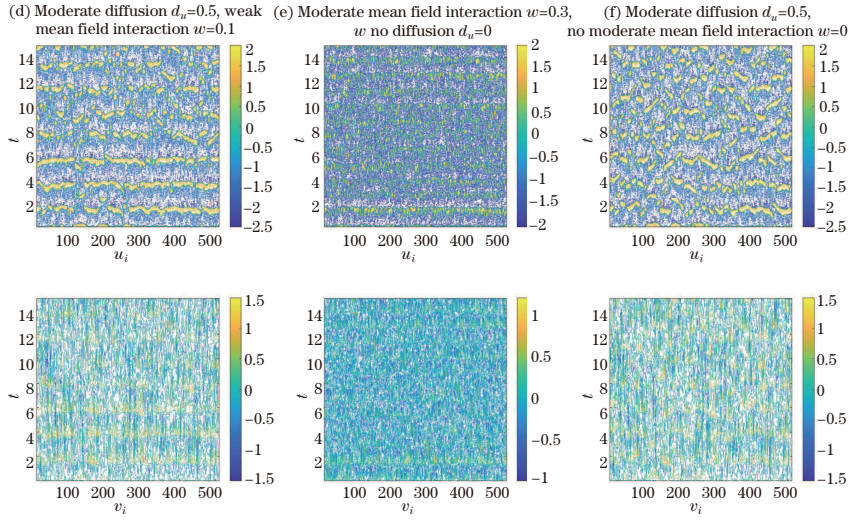
Figure 12 Spatiotemporal patterns of different dynamical regimes (continued): (d) Moderate diffusion $d_u = 0.5$ and weak mean field interaction $w = 0.1$; (e) Moderate mean field interaction $w = 0.3$ and no diffusion $d_u = 0$; and (f) Moderate diffusion $d_u = 0.5$ and no mean field interaction $w = 0$.

may lead to biases. In this subsection, we compare the covariance computed from the spatial averaging strategy with the truth that is generated from the Monte Carlo simulations in different dynamical regimes. The goal is to explore the consistency of this spatial averaging strategy in computing the covariance in different dynamical regimes.

Below, we always take $N = 512$ spatial grid points. For a fair comparison, we generate the truth using the direct Monte Carlo simulation with $M_{MC} = N = 512$ samples, which equals to the effective sample number using the spatial averaging strategy with simply one simulation. In all the tests shown below, we compare the covariance computed using the above two strategies at time $t = 0.5, t = 1, t = 2$ and $t = 5$. The regimes to be tested contain both local and non-local effects with various strengths. See Figures 11–12 for the spatiotemporal simulations of different dynamical regimes.

In Figure 11, strong spatial correlations are seen in all three dynamical regimes and the underlying stochastic processes is not mixing. Yet, there are some fundamental difference in these dynamical regimes. In Regime (a), there is a strong diffusion but there is no mean field interaction. Thus, the correlation comes from only the local effect but the diffusion is so strong that each grid point affects the other points that are far from itself. In Regime (b), only the mean field interaction plays an role since the diffusion is zero but the mean field interaction coefficient is quite large. Thus, a strong synchronization is seen in the spatiotemporal patterns. Regime (c) involves both a moderate diffusion and a moderate mean field interaction. The combined local and non-local effects also contribute to the strong correlation in the spatiotemporal patterns. Figure 13 shows the calculated covariance based on the spatial averaging strategy (red) and the truth generated from the direct Monte Carlo simulation (green). For a short time up to $t = 1$, the consistency is found in Regimes (a) and (b). But the covariance computed from the spatial averaging strategy becomes significantly different from the truth after $t = 2$, which indicates the

loss of the long-term consistency using the spatial averaging strategy. These results are related to the loss of practical controllability as was discussed in the previous subsection. In fact, the constant η_β in Corollary 4.1 becomes very large in both regimes due to the strong diffusion and/or the strong mean field interaction and thus the consistency of the spatial averaging strategy becomes nearly degenerated very quickly, which has been pointed out in the remarks below Corollary 4.1. For Regime (b), due to the strong global effect (via mean field interaction), the consistency is again lost very quickly at $t = 1$. Note in all these three regimes the covariance is not localized after a certain time due to either strong diffusion or strong mean field interaction (or both). Thus, the spatial averaging strategy does not work well for long time.
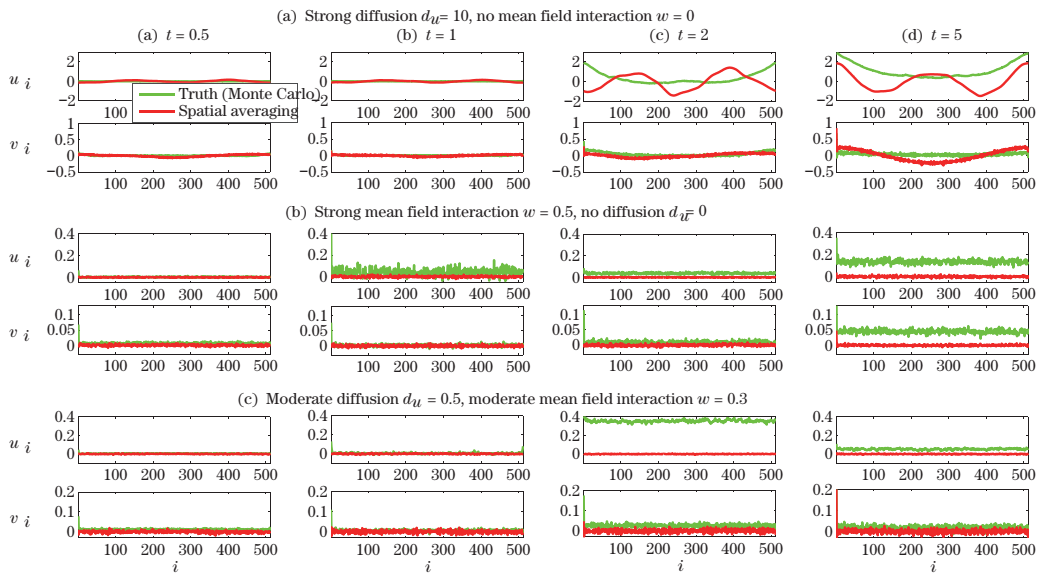

Figure 13   Comparison of the Monte Carlo and spatial average strategies in computing the covariance of the FHN model in the regimes with (a) Strong diffusion $d_u = 10$ and no mean field interaction $w = 0$; (b) Strong mean field interaction $w = 0.5$ and no diffusion $d_u = 0$; (c) Moderate diffusion $d_u = 0.5$ and moderate mean field interaction $w = 0.3$.

On the other hand, no strong spatial synchronization is observed in the three dynamical regimes shown in Figure 12. Some weak spatial synchronization is observed in Regimes (d) and (e), where the former involves a moderate diffusion and a weak mean field interaction while the latter is given by a moderate mean field interaction but no diffusion. Regarding the spatial covariance, Figure 14 shows that the spatial averaging strategy is able to provide accurate solutions up to $t = 2$. At $t = 5$, a slight inconsistency in the solution computed from the spatial averaging strategy is found, which is due to the degeneration of the consistency for long time as indicated in Corollary 4.1. Thus, the spatial averaging strategy is skillful at short and moderate times but the long-term consistency cannot be guaranteed in these regimes.

Finally, in Regime (f), which involves only a moderate diffusion but no mean field interaction, the consistency of the solution using the spatial averaging strategy lasts longer and is at least up to $t = 5$ (see Figure 15).

To summarize, both a strong global effect (mean field interaction) and a strong local effect (diffusion) will lead to the loss of the long-term consistency in the solutions using the spatial
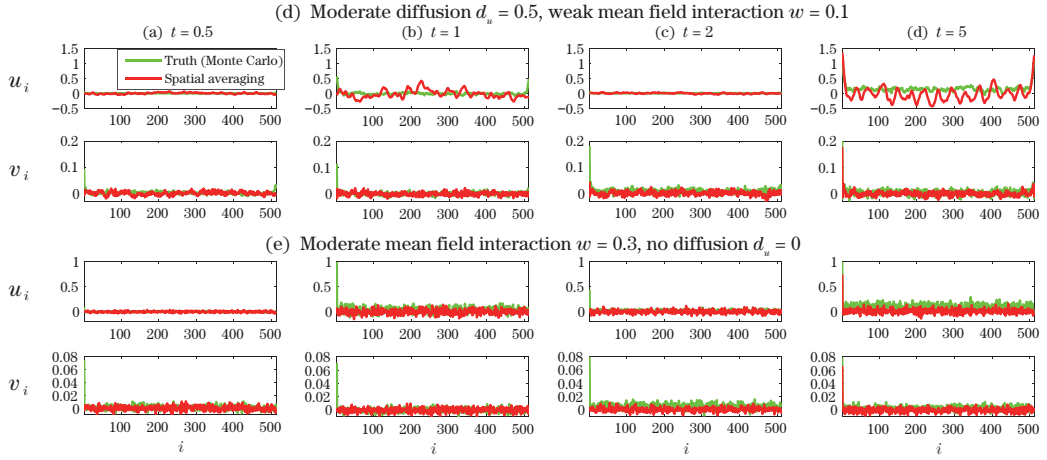
Figure 14   Comparison of the Monte Carlo and spatial average strategies in computing the covariance of the FHN model in the regimes with (d) Moderate diffusion $d_u = 0.5$ and weak mean field interaction $w = 0.1$; and (e) Moderate mean field interaction $w = 0.3$ and no diffusion $d_u = 0$.
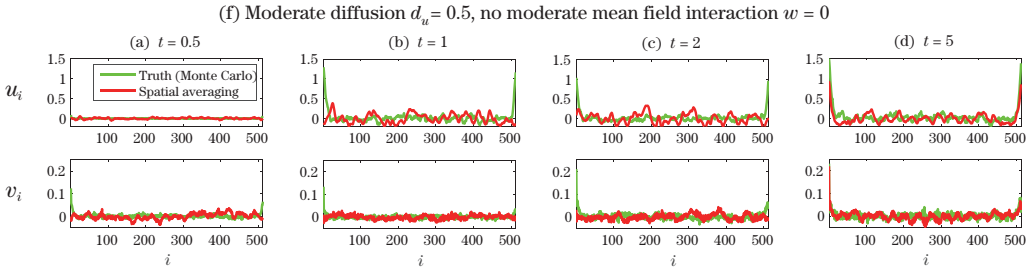

Figure 15   Comparison of the Monte Carlo and spatial average strategies in computing the covariance of the FHN model in the regimes with (f) Moderate diffusion $d_u = 0.5$ and no mean field interaction $w = 0$.

averaging strategy, which has been clearly indicated in Corollary 4.1 and the remarks below it. The spatial averaging strategy is skillful and can be applied for a relatively long time when the FHN model has a weak global or a moderate local effect.

5 Conclusions

In this article, a theoretical framework to understand the spatial localization for a large class of stochastically coupled nonlinear systems in high dimensions is developed and is applied to nonlinear complex models for excitable media. Rigorous mathematical theories (Theorem 3.2) show covariance decay behavior due to both local and nonlocal effects, which result from the diffusion and the mean field interaction, respectively. The analysis is based on a comparison with an appropriate linear surrogate model, of which the covariance propagation can be computed explicitly. Two important applications of these theoretical results are illustrated. They are the spatial averaging strategy for efficiently sampling the state variables and the localization technique in data assimilation. Test examples of a linear model and a stochastically coupled FitzHugh-Nagumo model for excitable media are adopted to validate the theoretical results.

The later is also used for a systematical study of the spatial averaging strategy in efficiently sampling the covariance matrix in different dynamical regimes. The spatial averaging strategy is always skillful for short time. The long-term consistency of the spatial averaging strategy requires that the mean field interaction and the diffusion cannot be too strong and the covariance matrix has localized structures.

A Proof

We provide the proofs of the theoretical results in this appendix. In particular for Theorem 3.1, its arguments are broken into subsections.

A.1 Consistency of the numerical formulation

Lemma A.1 *Under the conditions of Theorem 3.2, let $\mathbf{X}(0) = \mathbf{x}(0)$, and $\mathbf{W}_n = \mathbf{w}_{hn} - \mathbf{w}_{h(n-1)}$, and consider the corresponding $X(n)$ given by (3.5). Then if we write $\Delta(t) = \mathbf{X}(T) - \mathbf{x}(t)$, there is a constant D_t such that for any sufficiently small h,*

$$\mathbb{E}\|\Delta(s)\|^2 \leq \exp(2(\lambda_0 + 2\lambda_H)t)h\|D_\Sigma\|_F^2, \quad s \leq t.$$

Proof Let $\mathbf{z}(nh) = \mathbf{X}(n)$, and $\mathbf{z}((n+s)h) = \Psi_{sh}(\mathbf{X}(n))$ for $0 \leq s < 1$. Clearly $\mathbf{z}(0) = \mathbf{x}(0)$. Note that

$$\mathrm{d}\Delta(t) = (F(\mathbf{x}(t)) - F(\mathbf{z}(t)))\mathrm{d}t + D_\Sigma \mathrm{d}\mathbf{w}(t), \quad t \in (nh, (n+1)h).$$

So

$$\mathrm{d}\|\Delta(t)\|^2 = 2\langle F(\mathbf{x}(t)) - F(\mathbf{z}(t)), \mathbf{x}(t) - \mathbf{z}(t)\rangle \mathrm{d}t + \|D_\Sigma\|_F^2 \mathrm{d}t + 2\langle \mathbf{x}(t) - \mathbf{z}(t), D_\Sigma \mathrm{d}\mathbf{w}(t)\rangle.$$

Note that under our conditions

$$\langle F(\mathbf{x}(t)) - F(\mathbf{z}(t)), \mathbf{x}(t) - \mathbf{z}(t)\rangle = \int_0^1 \mathrm{d}s \langle \nabla F(\mathbf{z}(t) + s\Delta(t))\Delta(t), \Delta(t)\rangle.$$

But

$$\langle \nabla F(\mathbf{z}(t) + s\Delta(t))\Delta(t), \Delta(t)\rangle$$
$$= \sum_i \Delta_i(t)^{\mathrm{T}}\{\nabla F(\mathbf{z}(t) + s\Delta(t))\}_{i,i}\Delta_i(t) + \sum_{i,m=\pm 1} \Delta_{i+m}(t)^{\mathrm{T}}\{\nabla F(\mathbf{z}(t) + s\Delta(t))\}_{i,i+m}\Delta_i(t)$$
$$\leq \sum_i \lambda_0\|\Delta_i(t)\|^2 + \sum_{i,m=\pm 1} \lambda_H\|\Delta_{i+m}(t)\|\|\Delta_i(t)\| \leq (\lambda_0 + 2\lambda_H)\|\Delta(t)\|^2.$$

Therefore

$$\langle F(\mathbf{x}(t)) - F(\mathbf{z}(t)), \mathbf{x}(t) - \mathbf{z}(t)\rangle \leq (\lambda_0 + 2\lambda_H)\|\Delta(t)\|^2.$$

So by Gronwall's inequality

$$\mathbb{E}\|\Delta(nh + h-)\|^2 \leq \exp(2(\lambda_0 + 2\lambda_H)h)\mathbb{E}\|\Delta(nh)\|^2 + \frac{\exp(2(\lambda_0 + 2\lambda_H)h) - 1}{2\lambda_0 + 4\lambda_H}\|D_\Sigma\|_F^2.$$

Next notice that

$$\Delta((n+1)h) = \Delta(nh) + \int_0^h (F(\mathbf{x}(nh + s)) - F(\mathbf{z}(nh + s)))\,\mathrm{d}s = \Delta((n+1)h-) - \int_0^h D_\Sigma \mathrm{d}\mathbf{w}_{nh+s}.$$

By Itô's isometry

$$\mathrm{d}\Big\langle \Delta(nh+s), \int_0^s D_\Sigma \mathrm{d}\mathbf{w}_{nh+s} \Big\rangle = \langle F(\mathbf{x}(nh+s)) - F(\mathbf{z}(nh+s)), D_\Sigma \mathrm{d}\mathbf{w}_{nh+s} \rangle \mathrm{d}s + \|D_\Sigma\|_F^2 \mathrm{d}s.$$

Therefore

$$\mathbb{E}\Big\langle \Delta(nh+s), \int_0^s D_\Sigma \mathrm{d}\mathbf{w}_{nh+s} \Big\rangle = \|D_\Sigma\|_F^2 s.$$

Therefore, when h is small enough, by Taylor expansion,

$$\begin{aligned}
&\mathbb{E}\|\Delta((n+1)h)\|^2 \\
&= \mathbb{E}\|\Delta((n+1)h-)\|^2 - \|D_\Sigma\|_F^2 h \\
&\leq \exp(2(\lambda_0+2\lambda_H)h)\mathbb{E}\|\Delta(nh)\|^2 + \Big(\frac{\exp(2(\lambda_0+2\lambda_H)h)-1}{2\lambda_0+4\lambda_H} - h\Big)\|D_\Sigma\|_F^2 \\
&\leq \exp(2(\lambda_0+2\lambda_H)h)\mathbb{E}\|\Delta(nh)\|^2 + 4(\lambda_0+2\lambda_H)h^2\|D_\Sigma\|_F^2.
\end{aligned}$$

Then by the discrete Gronwall's inequality, if $\lambda_0 + 2\lambda_H > 0$, at $Th = t$, since $\exp(a) - 1 > a$ for all $a > 0$,

$$\mathbb{E}\|\Delta(t)\|^2 \leq 2(\lambda_0+2\lambda_H)\frac{\exp(2(\lambda_0+2\lambda_H)t)-1}{\exp(2(\lambda_0+2\lambda_H)h)-1}h^2\|D_\Sigma\|_F^2 \leq \exp(2(\lambda_0+2\lambda_H)t)h\|D_\Sigma\|_F^2.$$

A.2 Decomposition of covariance

Clearly, we can view $\mathbf{X}(T)$ as a function of the realization of $\mathbf{W}_i$,

$$\mathbf{X}(T) = \Phi(\mathbf{W}_0, \mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_T) = \Psi_h(\cdots(\Psi_h(\Psi_h(\mathbf{X}_0) + \Sigma_h\mathbf{W}_1) + \Sigma_h\mathbf{W}_2)\cdots) + \Sigma_h\mathbf{W}_T.$$

For the simplicity of notation, we write $\mathbf{W}_0, \mathbf{W}_1, \mathbf{W}_2, \cdots, \mathbf{W}_T$ as $\mathbf{W}_{0:T}$ in below. And the i-th block of $\Phi \in \mathbb{R}^{qN}$ as $\Phi_i \in \mathbb{R}^q$. Then we can write $\mathbf{X}_i(T) = \Phi_i(\mathbf{W}_{0:T})$.

We can study the dependence of $\mathbf{X}(T)$ on the $\mathbf{W}_i$ through the gradients, written as $\nabla_i\Phi := \nabla_{\mathbf{W}_i}\Phi$. By the chain rule, it has this back propagation formulation,

$$\begin{aligned}
\nabla_T\Phi &= \Sigma_h, \quad \nabla_{T-1}\Phi = \nabla\Psi_h(\mathbf{X}_{T-1})\Sigma_h, \\
\nabla_{T-2}\Phi &= \nabla\Psi_h(\mathbf{X}_{T-1})\nabla\Psi_h(\mathbf{X}_{T-2})\Sigma_h, \\
\nabla_k\Phi &= \nabla\Psi_h(\mathbf{X}_{T-1})\cdots\nabla\Psi_h(\mathbf{X}_k)\Sigma_h, \\
\nabla_0\Phi &= \nabla\Psi_h(\mathbf{X}_{T-1})\cdots\nabla\Psi_h(\mathbf{X}_0)\Sigma_0.
\end{aligned} \tag{A.1}$$

To investigate the covariance, we consider another independent sample

$$\mathbf{X}'(T) = \Phi(\mathbf{W}'_{0:T}) = \Phi(\mathbf{W}'_0, \mathbf{W}'_1, \mathbf{W}'_2, \cdots, \mathbf{W}'_T).$$

Then clearly the covariance can be written as

$$\mathrm{cov}(g(\mathbf{X}_i(T)), g(\mathbf{X}_j(T))) = \mathbb{E}g(\mathbf{X}_i(T))g(\mathbf{X}_j(T)) - \mathbb{E}g(\mathbf{X}_i(T))g(\mathbf{X}'_j(T)).$$

We can decompose it by considering

$$\begin{aligned}
\Delta_k &:= g(\Phi_i(\mathbf{W}_{0:T}))g(\Phi_j(\mathbf{W}_{0:k}, \mathbf{W}'_{k+1:T})) - g(\Phi_i(\mathbf{W}_{0:T}))g(\Phi_j(\mathbf{W}_{0:k-1}, \mathbf{W}'_{k:T})) \\
&= g(\Phi_i(\mathbf{W}_{0:T}))g(\Phi_j(\mathbf{W}_{0:k-1}, \mathbf{W}_k, \mathbf{W}'_{k+1:T})) \\
&\quad - g(\Phi_i(\mathbf{W}_{0:T}))g(\Phi_j(\mathbf{W}_{0:k-1}, \mathbf{W}'_k, \mathbf{W}'_{k+1:T})).
\end{aligned} \tag{A.2}$$

Because $\mathbf{X}(T) = \Phi(\mathbf{W}_{0:T}), \mathbf{X}'(T) = \Phi(\mathbf{W}'_{0:T})$, so

$$g(\mathbf{X}_i(T))g(\mathbf{X}_j(T)) - g(\mathbf{X}_i(T))g(\mathbf{X}'_j(T)) = \sum_{k=0}^{T} \Delta_k.$$

We will bound $\mathbb{E}\Delta_k$ using a Stein's lemma type of formula, Lemma 3.1. Its proof is given below.

A.3 Stein's lemma

Proof of Lemma 3.1 The first part comes simply by

$$\mathrm{cov}(f(X), g(X)) = \mathbb{E}f(X)g(X) - \mathbb{E}f(X)\mathbb{E}g(Y) = \mathbb{E}f(X)g(X) - \mathbb{E}f(X)g(Y).$$

The note that

$$f(X)g(Y) - f(X)g(X) = f(X)g(X^\theta)\Big|_{\theta=0}^{\theta=\frac{\pi}{2}}$$

$$= \int_0^{\frac{\pi}{2}} f(X)\frac{\mathrm{d}}{\mathrm{d}\theta}g(X^\theta)\mathrm{d}\theta$$

$$= \int_0^{\frac{\pi}{2}} f(X)\langle\nabla g(X^\theta), -\sin\theta\, X + \cos\theta\, Y\rangle\mathrm{d}\theta.$$

Note that when $X \sim \mathcal{N}(0, \mathbf{I}_d)$, and smooth vector field $h : \mathbb{R}^d \to \mathbb{R}^d$, let $c = (2\pi)^{-\frac{d}{2}}$, by the divergence theorem of calculus, $\int \nabla \cdot (h(\mathbf{x})e^{-\frac{1}{2}|\mathbf{x}|^2})\mathrm{d}\mathbf{x} = 0$, therefore

$$\mathbb{E}\langle h(X), X\rangle = c\int\langle h(\mathbf{x}), \mathbf{x}\rangle e^{-\frac{1}{2}|\mathbf{x}|^2}\mathrm{d}\mathbf{x} = -c\int\langle h(\mathbf{x}), \nabla e^{-\frac{1}{2}|\mathbf{x}|^2}\rangle\mathrm{d}\mathbf{x}$$

$$= c\int\nabla \cdot h(\mathbf{x})e^{-\frac{1}{2}|\mathbf{x}|^2}\mathrm{d}\mathbf{x} = \mathbb{E}\nabla \cdot h(X). \tag{A.3}$$

Here, $\nabla\cdot$ denotes the divergence operator. Therefore, if we condition on the value of Y and use law of total expectation

$$\mathbb{E}\langle f(X)\nabla g(X^\theta), X\rangle = \mathbb{E}\mathbb{E}_Y\langle f(X)\nabla g(\cos\theta\, X + \sin\theta\, Y), X\rangle$$

$$= \mathbb{E}\mathbb{E}_Y\nabla \cdot (f(X)\nabla g(\cos\theta\, X + \sin\theta\, Y))$$

$$= \mathbb{E}\mathbb{E}_Y\langle\nabla f(X), \nabla g(X^\theta)\rangle + \cos\theta\,\mathbb{E}\mathbb{E}_Y f(X)\nabla \cdot \nabla g(X^\theta)$$

$$= \mathbb{E}\langle\nabla f(X), \nabla g(X^\theta)\rangle + \cos\theta\,\mathbb{E}f(X)\nabla \cdot \nabla g(X^\theta).$$

Likewise, we have the following

$$\mathbb{E}\langle f(X)\nabla g(X^\theta), Y\rangle = \mathbb{E}\mathbb{E}_X\langle f(X)\nabla g(\cos\theta\, X + \sin\theta\, Y), Y\rangle$$

$$= \sin\theta\,\mathbb{E}\mathbb{E}_X f(X)\nabla \cdot \nabla g(\cos\theta\, X + \sin\theta\, Y)$$

$$= \sin\theta\,\mathbb{E}f(X)\nabla \cdot \nabla g(\cos\theta\, X + \sin\theta\, Y).$$

Therefore, we have

$$\mathbb{E}f(X)\langle\nabla g(X^\theta), -\sin\theta\, X + \cos\theta\, Y\rangle = -\sin\theta\mathbb{E}\langle\nabla f(X), \nabla g(X^\theta)\rangle,$$

and this finishes our proof.

To apply Lemma 3.1, we view $\mathbf{W}_k$ as X, and $\mathbf{W}'_k$ as Y in (A.2). Let

$$\mathbf{W}^{\theta,k} := [\mathbf{W}_{1:k-1}, \cos\theta\mathbf{W}_k + \sin\theta\mathbf{W}'_k, \mathbf{W}'_{k+1:T}].$$

Then

$$\Delta_k = g(\Phi_i(\mathbf{W}_{0:T}))g(\Phi_j(\mathbf{W}^{0,k})) - g(\Phi_i(\mathbf{W}_{0:T}))g(\Phi_j(\mathbf{W}^{\frac{\pi}{2},k})).$$

Note that by the chain rule,

$$\nabla_k g(\Phi_j(\mathbf{W}_{0:T})) = \nabla g(\Phi_j(\mathbf{W}_{0:T}))\nabla_k\Phi_j(\mathbf{W}_{0:T}).$$

So by Lemma 3.1, we have

$$
\begin{aligned}
\mathbb{E}\Delta_k &= \int_0^{\frac{\pi}{2}} \sin\theta\,\mathbb{E}\langle\nabla g(\Phi_i(\mathbf{W}_{0:T}))\nabla_k\Phi_i(\mathbf{W}_{0:T}), \nabla g(\Phi_j(\mathbf{W}^{\theta,k}))\nabla_k\Phi_j(\mathbf{W}^{\theta,k})\rangle\mathrm{d}\theta \\
&= \int_0^{\frac{\pi}{2}} \sin\theta\,\mathbb{E}\mathrm{tr}\left(\nabla g(\Phi_j(\mathbf{W}^{\theta,k}))^{\mathrm{T}}\nabla g(\Phi_i(\mathbf{W}_{0:T}))\nabla_k\Phi_i(\mathbf{W}_{0:T})\nabla_k\Phi_j(\mathbf{W}^{\theta,k})^{\mathrm{T}}\right)\mathrm{d}\theta \\
&\leq \|\nabla g\|_\infty^2 \int_0^{\frac{\pi}{2}} \sin\theta\,\mathbb{E}\left|\mathrm{tr}\left(\nabla_k\Phi_i(\mathbf{W}_{0:T})\nabla_k\Phi_j(\mathbf{W}^{\theta,k})^{\mathrm{T}}\right)\right|\mathrm{d}\theta.
\end{aligned}
\tag{A.4}
$$

Note that $\nabla_k\Phi_i(\mathbf{W}_{0:T})\nabla_k\Phi_j(\mathbf{W}^{\theta,k})^{\mathrm{T}} \in \mathbb{R}^{q\times q}$ is the (i,j)-th sub-block of

$$R^k := \nabla_k\Phi(\mathbf{W}_{0:T})\nabla_k\Phi(\mathbf{W}^{\theta,k})^{\mathrm{T}} \in \mathbb{R}^{qN\times qN}.\tag{A.5}$$

So essentially we want bound (A.5) in each of its block. To do that, we need to consider the derivative of the flow Ψ_t.

A.4 Derivative flow

Lemma A.2 *Fix any realization of $\mathbf{W}_{0:T}$ and $\mathbf{W}'_{0:T}$. Let $Q_h(s) = \exp(hG_* s)\exp(hG_* s)^{\mathrm{T}}$. The (i,j)-th sub-block of R^k is bounded by*

$$\|R^k_{i,j}\|_F \leq \|h\Sigma^2\|_F[Q(T-k)]_{i,j}, \quad k \geq 1, \quad \|R^0_{i,j}\|_F \leq \|\Sigma_0^2\|_F[Q(T)]_{i,j}.$$

Proof We will discuss only the $k \geq 1$ case, the analysis $k = 0$ case is identical. For any $s \leq T-k$, we can extend the definition of $\mathbf{X}(t)$ to noninteger t, by letting

$$\mathbf{X}(k+s) = \Psi_{sh}(\mathbf{X}(k+\lfloor s\rfloor)) \quad \text{if } s \neq \lfloor s\rfloor.$$

Here $\lfloor s\rfloor$ denotes the largest integer that is less than s. The derivative flow of Ψ_t can also be extended. We define the following $\mathbb{R}^{qN\times qN}$ matrix:

$$P(s) = \nabla\Psi_{(s-\lfloor s\rfloor)h}(\mathbf{X}(k+\lfloor s\rfloor))\cdots\nabla\Psi_h(\mathbf{X}(k))\Sigma_h.$$

Then following the back propagation formula (A.1), $\nabla_k\Phi = P(T-k)$. $P(s)$ also follows the following differential equation with initial condition $P(0) = \Sigma_h$,

$$\dot{P}(s) = hG(k+s, \mathbf{X}(k+s))P(s).$$

$G(s,\mathbf{x}) \in \mathbb{R}^{qN\times qN}$ is the Jacobian created by the ODE flow (3.4). So its (i,j)-th sub-block is given by

$$G_{i,j}(s,\mathbf{x}) = \nabla_{\mathbf{x}_j}\mathbf{f}_i(s,\mathbf{x}_{i-1},\mathbf{x}_i,\mathbf{x}_{i+1}) + \frac{1}{N}\nabla_{\mathbf{x}_j}h_i(s,\mathbf{x}_j).$$

Clearly, it is a block-tridiagonal matrix.

Because of (A.5), we are also interested in $\nabla_k \Phi(\mathbf{W}^{\theta,k})$. We consider a numerical sequence generated by (3.5), but the driving noise is $\mathbf{W}^{\theta,k}$. Fix a $\theta \in \left[0, \frac{\pi}{2}\right]$. Let $\mathbf{Y}(n) = \mathbf{X}(n)$ for $n = 0, \cdots, k$,

$$\mathbf{Y}(k+1) = \Psi_h(\mathbf{X}(k)) + \Sigma_h \cos\theta \mathbf{W}_k + \Sigma_h \sin\theta \mathbf{W}'_k,$$

and $\mathbf{Y}(n+1) = \Psi_h(\mathbf{Y}(n)) + \Sigma_h \mathbf{W}'_n$ for $n \geq k+1$. We also define $\mathbf{Y}(k+s) = \Psi_{sh}(\mathbf{Y}(k+\lfloor s \rfloor))$ for noninteger time $s \neq \lfloor s \rfloor$. Define also the following matrix

$$P'(s) = \nabla\Psi_{(s-\lfloor s \rfloor)h}(\mathbf{Y}_{k+\lfloor s \rfloor}) \cdots \nabla\Psi_h(\mathbf{Y}_k)\Sigma_h.$$

It follows the ODE:

$$\dot{P}'(s) = hG(k+s, \mathbf{Y}(k+s))\dot{P}'(s).$$

By the back propagation formula (A.1), $P'(T-k) = \nabla_k \Phi(\mathbf{W}^{\theta,k})$.

Let $R(s) = P(s)(P'(s))^{\mathrm{T}}$. The quantity we are interested is $R^k = R(T-k)$. $R(s)$ follows the following ODE:

$$\dot{R}(s) = hG(\mathbf{X}_{k+s})R(s) + hR(s)G(\mathbf{Y}_{k+s})^{\mathrm{T}}, \quad R(0) = \Sigma_h\Sigma_h^{\mathrm{T}} = \Sigma_h^2.$$

Here and below, we write $G(k+s, \mathbf{X}_{k+s})$ as $G(\mathbf{X}_{k+s})$ for simplicity. Consider the (i,j)-th sub block of $R_{i,j}(s)$. In below, we use $\{AB\}_{i,j}$ to denote the (i,j)-th sub block of matrix AB. We also define the inner product of matrices as $\langle A, B \rangle = \mathrm{tr}(A^{\mathrm{T}}B)$. Then $\|R_{i,j}(s)\|_F^2 = \langle R_{i,j}(s), R_{i,j}(s) \rangle$. We take time derivative, and recall that G is block tridiagonal, so

$$\frac{\mathrm{d}}{\mathrm{d}s}\langle R_{i,j}(s), R_{i,j}(s) \rangle = 2h\langle \{G(\mathbf{X}_{k+s})R(s)\}_{i,j}, R_{i,j}(s) \rangle + 2h\langle \{R(s)G(\mathbf{Y}_{k+s})^{\mathrm{T}}\}_{i,j}, R_{i,j}(s) \rangle$$

$$= 2h\langle G_{i,i}(\mathbf{X}_{k+s})R_{i,j}(s), R_{i,j}(s) \rangle + 2h\langle R_{i,j}(s)G_{j,j}(\mathbf{Y}_{k+s})^{\mathrm{T}}, R_{i,j}(s) \rangle$$

$$+ 2h\sum_{m=\pm 1}\langle G_{i,i+m}(\mathbf{X}_{k+s})R_{i+m,j}(s), R_{i,j}(s) \rangle$$

$$+ 2h\sum_{m=\pm 1}\langle R_{i,j+m}(s)G_{j,j+m}(\mathbf{Y}_{k+s})^{\mathrm{T}}, R_{i,j}(s) \rangle.$$

Because of the covariance propagation assumption in Theorem 3.1, and an elementary A.3 for matrix inner product, we can bound each term as below

$$\langle G_{i,i}(\mathbf{X}_{k+s})R_{i,j}(s), R_{i,j}(s) \rangle \leq [G_*(s+k)]_{i,i}\|R_{i,j}(s)\|_F^2,$$

$$\langle R_{i,j}(s)G_{j,j}(\mathbf{Y}_{k+s})^{\mathrm{T}}, R_{i,j}(s) \rangle \leq [G_*(s+k)]_{j,j}\|R_{i,j}(s)\|_F^2,$$

$$\langle G_{i,i+m}(\mathbf{X}_{k+s})R_{i+m,j}(s), R_{i,j}(s) \rangle \leq [G_*(s+k)]_{i,i+m}\|R_{i,j}(s)\|_F\|R_{i+m,j}(s)\|_F,$$

$$\langle R_{i,j+m}(s)G_{j,j+m}(\mathbf{Y}_{k+s})^{\mathrm{T}}, R_{i,j}(s) \rangle \leq [G_*(s+k)]_{j,j+m}\|R_{i,j}(s)\|_F\|R_{i,j+m}(s)\|_F.$$

In conclusion, we have the following with the dependence of G_* on $s+k$ suppressed,

$$\frac{\mathrm{d}}{\mathrm{d}s}\|R_{i,j}(s)\|_F^2 \leq 2([G_*]_{i,i} + [G_*]_{j,j})h\|R_{i,j}(s)\|_F^2 + 2\sum_{m=\pm 1}[G_*]_{i,i+m}h\|R_{i,j}(s)\|_F\|R_{i+m,j}(s)\|_F$$

$$+ 2\sum_{m=\pm 1}[G_*]_{j+m,j}h\|R_{i,j}(s)\|_F\|R_{i,j+m}(s)\|_F.$$

Since $\frac{\mathrm{d}}{\mathrm{d}s}\|R_{i,j}(s)\|_F^2 = 2\|R_{i,j}(s)\|_F \frac{\mathrm{d}}{\mathrm{d}s}\|R_{i,j}(s)\|_F$, we have

$$\frac{\mathrm{d}}{\mathrm{d}s}\|R_{i,j}(s)\|_F$$
$$\leq h([G_*]_{i,i} + [G_*]_{j,j})\|R_{i,j}(s)\|_F + h \sum_{m=\pm 1} [G_*]_{i,i+m}\|R_{i+m,j}(s)\|_F + [G_*]_{j+m,j}\|R_{i,j+m}(s)\|_F.$$

Next we consider $Q_h(s)$. Note that $\|h\Sigma^2\|_F[Q_h(0)]_{i,j} = \|R_{i,j}(0)\|_F$. Moreover, Q_h follows the ODE

$$\frac{\mathrm{d}}{\mathrm{d}s}Q_h(s) = hG_*Q_h(s) + hQ_h(s)G_*^{\mathrm{T}}.$$

Use the fact that G_* is tridiagonal,

$$\frac{d}{\mathrm{d}s}[Q_h(s)]_{i,j} = h[G_*Q_h(s)]_{i,j} + h[Q_h(s)G_*^{\mathrm{T}}]_{i,j}$$
$$= h([G_*]_{i,i} + [G_*]_{j,j})[Q_h(s)]_{i,j}$$
$$+ h \sum_{m=\pm 1} [G_*]_{i,i+m}[Q_h(s)]_{i+m,j} + [G_*]_{j,j+m}[Q_h(s)]_{i,j+m}.$$

If we view $[Q_h(s)]_{i,j}$ as a linear ODE, the non diagonal term of this ODE system , $[G_*]_{i,i+m}$ and $[G_*]_{j,j+m}$, are all nonnegative. So by Lemma A.4, we have that

$$\|R_{i,j}(s)\|_F \leq [Q(s)]_{i,j}\|h\Sigma^2\|_F.$$

This leads to our claims.

Lemma A.3 *Suppose* $A + A^{\mathrm{T}} \preceq 2\lambda_A\mathrm{I}$, *then* $\langle AX, X \rangle \leq \lambda_A\langle X, X \rangle$, *and* $\langle XA, X \rangle \leq \lambda_A\langle X, X \rangle$.

Proof Let v_i be the right eigenvectors of X, so $Xv_i = \lambda_i v_i$, then

$$\langle AX, X \rangle = \mathrm{tr}(AXX^{\mathrm{T}}) = \mathrm{tr}(X^{\mathrm{T}}AX) = \sum_i v_i^* X^{\mathrm{T}}AXv_i$$
$$= \sum_i |\lambda_i|^2 v_i^* Av_i \leq \lambda_A \sum_i |\lambda_i|^2 = \lambda_A\langle X, X \rangle.$$

For the second claim, let v_i be the right eigenvectors of X^{T}, so $X^{\mathrm{T}}v_i = \lambda_i v_i$,

$$\langle XA, X \rangle = \mathrm{tr}(XAX^{\mathrm{T}}) = \sum_i v_i^* X^{\mathrm{T}}AXv_i = \sum_i |\lambda_i|^2 v_i^* Av_i \leq \lambda_A \sum_i |\lambda_i|^2 = \lambda_A\langle X, X \rangle.$$

Lemma A.4 *If* $A \in \mathbb{R}^{d \times d}$ *has all non-diagonal entries being nonnegative, then* $\exp(sA)$ *has all entries being nonnegative for any* s. *In particular, if* $X(s) \in \mathbb{R}^d$ *and* $Y(s) \in \mathbb{R}^d$ *satisfies*

$$\frac{\mathrm{d}}{\mathrm{d}s}X(s) \leq AX(s), \quad \frac{\mathrm{d}}{\mathrm{d}s}Y(s) = AY(s),$$

where $\leq$ *is interpreted entry-wise. Then* $X(0) \leq Y(0)$ *leads to* $X(s) \leq Y(s)$ *entry-wise.*

Proof The statement is elementary if all of A diagonal entries are also nonnegative, since

$$\exp(sA) = \mathrm{I} + sA + \frac{1}{2!}s^2a^2 + \frac{1}{3!}s^3A^3 + \cdots.$$

On the other hand, we can always find a λ so that $\lambda \mathbf{I} + sA$ has all diagonal entries being nonnegative. Because $\mathbf{I}$ commute with sA, so

$$\exp(sA) = \exp(-\lambda \mathbf{I}) \exp(\lambda \mathbf{I} + sA) = \mathrm{e}^{-\lambda} \exp(\lambda \mathbf{I} + sA),$$

and all entries of $\exp(sA)$ are nonnegative.

For the ODE system, we first consider the case when all entries of A are positive. Define $\Delta(s) = Y(s) - X(s)$. $\Delta(0) \geq \mathbf{0}$ entry-wise. Also when $\Delta(s) \geq 0$,

$$\frac{\mathrm{d}}{\mathrm{d}s} \Delta(s) \geq A\Delta(s) \geq \mathbf{0},$$

meaning $\Delta(s)$ is increasing entry-wise. Therefore $\Delta(s) \geq \mathbf{0}$ entry-wise for $s \geq 0$.

If the diagonal terms of A are negative, we can consider $X_\lambda(t) = \mathrm{e}^{\lambda t} X(t)$ and $Y_\lambda(t) = \mathrm{e}^{\lambda t} Y(t)$, then

$$\frac{\mathrm{d}}{\mathrm{d}s} X_\lambda(s) \leq (A + \lambda \mathbf{I}) X(s), \quad \frac{\mathrm{d}}{\mathrm{d}s} Y(s) = (A + \lambda \mathbf{I}) Y(s).$$

By our previous analysis, $X_\lambda(t) \leq Y_\lambda(t)$ entry-wise, which leads to $X(t) \leq Y(t)$ entry-wise.

A.5 Summarizing argument

Proof of Theorem 3.1 Recall in (A.4), we have

$$|\mathbb{E}\Delta_k| \leq \|\nabla f\|_\infty^2 \int_0^{\frac{\pi}{2}} \sin\theta |\mathbb{E}\mathrm{tr}(R_{i,j}^k)| \mathrm{d}\theta.$$

Note that $R_{i,j}^k$ is $q \times q$ dimensional, so by Cauchy-Schwartz, $|\mathrm{tr}(R_{i,j}^k)| \leq \sqrt{q} \|R_{i,j}^k\|_F$. Lemma A.2 shows that

$$\|R_{i,j}^k\|_F \leq h\|\Sigma^2\|_F [Q_h(T-k)]_{i,j}, \quad k \geq 1, \quad \|R_{i,j}^0\|_F \leq \|\Sigma_0^2\|_F [Q_h(T)]_{i,j}.$$

Therefore for $k \geq 1$,

$$|\mathbb{E}\Delta_k| \leq h\sqrt{q}\|\nabla g\|_\infty^2 \|\Sigma^2\|_F [Q_h(T-k)]_{i,j} \int_0^{\frac{\pi}{2}} \sin\theta \mathrm{d}\theta = \sqrt{q} h\|\nabla g\|_\infty^2 \|\Sigma^2\|_F [Q_h(T-k)]_{i,j},$$

and likewise $|\mathbb{E}\Delta_0| \leq \sqrt{q}\|\nabla g\|_\infty^2 \|\Sigma_0^2\|_F [Q_h(T)]_{i,j}$. Consequentially, we have

$$|\mathrm{cov}(g(\mathbf{X}_i(T)), g(\mathbf{X}_j(T)))|$$
$$\leq \sqrt{q}\|\nabla g\|_\infty^2 \left(\|\Sigma_0^2\|_F Q_h(T) + h\|\Sigma^2\|_F \sum_{k=1}^T \mathbb{E}[Q_h(T-k)]_{i,j} \right). \tag{A.6}$$

When we let $h \to 0$, $\mathbf{X}(T) \to \mathbf{x}(t)$ by Lemma A.1. Then

$$|\mathbb{E}g(\mathbf{X}_i(T))g(\mathbf{X}_j(T)) - \mathbb{E}g(\mathbf{x}_i(t))g(\mathbf{x}_j(t))|$$
$$\leq \mathbb{E}|g(\mathbf{x}_i(t))(g(\mathbf{X}_j(T)) - g(\mathbf{x}_j(t)))| + \mathbb{E}|g(\mathbf{x}_j(t))(g(\mathbf{X}_i(T)) - g(\mathbf{x}_i(t)))|$$
$$+ \mathbb{E}|(g(\mathbf{X}_i(T)) - g(\mathbf{x}_i(T)))(g(\mathbf{X}_j(T)) - g(\mathbf{x}_j(t)))|$$
$$\leq \sqrt{\mathbb{E}|g(\mathbf{x}_i(T))|^2 \mathbb{E}|g(\mathbf{X}_j(T)) - g(\mathbf{x}_j(t))|^2} + \sqrt{\mathbb{E}|g(\mathbf{x}_j(t))|^2 \mathbb{E}|g(\mathbf{X}_i(T)) - g(\mathbf{x}_i(t))|^2}$$
$$+ \sqrt{\mathbb{E}|g(\mathbf{X}_i(T)) - g(\mathbf{x}_i(t))|^2 \mathbb{E}|g(\mathbf{X}_j(T)) - g(\mathbf{x}_j(t))|^2}.$$

By the Lipschitzness of SDE coefficient, it is standard to show that $\mathbb{E}|g(\mathbf{x}_i(t))|^2 \leq \|\nabla g\|_\infty^2 \mathbb{E}\|\mathbf{x}_i(t)\|^2$ is bounded. By Lemma A.1 we know

$$\mathbb{E}|g(\mathbf{X}_j(T)) - g(\mathbf{x}_j(t))|^2 \leq \mathbb{E}\|\nabla g\|_\infty^2 \|\mathbf{X}_j(T) - \mathbf{x}_j(t)\|^2 = O(h).$$

Also note that $Q_h(n) = Q(hn)$, by letting $h \to 0$, we have our claim.

A.6 Continuous time random walk

Lemma A.5 *The $Q(s)$ in Theorem 3.2 has the following upper bound*

$$[Q(s)]_{i,j} \leq 2\exp(\lambda_\beta s)\Big(\exp(-\beta\mathbf{d}(i,j)) + \frac{(1 + \mathrm{e}^{-\beta})(\mathrm{e}^{\lambda_H s} - 1)}{(1 - \mathrm{e}^{-\beta})N} \Big),$$

and

$$\int_0^t [Q(s)]_{i,j}\mathrm{d}s \leq \frac{2\exp(\lambda_\beta t) - 2}{\lambda_\beta} \Big(\mathrm{e}^{-\beta\mathbf{d}(i,j)} - \frac{2}{(1 - \mathrm{e}^{-\beta})N} \Big) + \frac{2(\exp(\eta_\beta t) - 1)(1 + \mathrm{e}^{-\beta})}{\eta_\beta(1 - \mathrm{e}^{-\beta})N}.$$

Proof Let $F_H = F_* - \lambda_G \mathbf{I}_N$, then its entries are

$$[F_H]_{i,i} = -2\lambda_F, \quad [F_H]_{i,i\pm 1} = \lambda_F.$$

Then $\exp(F_H s) = \mathrm{e}^{s\lambda_G}\exp(F_* s)$. Moreover, because $H_* F_H = F_H H_* = \mathbf{0}$, so $\exp(G_H s) = \exp(F_H s)\exp(H_* s)$. We first study the behavior of $\exp(F_H s)$.

Let Z_t be a continuous time random walk on $\mathbb{Z}$, with rate $\lambda_{z,z\pm 1} = \lambda_H$. Denote its projection on the modular space $\mathbb{Z}/N\mathbb{Z}$ as Y_t, which becomes a continuous time random walk on $\mathbb{Z}/N\mathbb{Z}$. Clearly F_H is the transition rate matrix of Y_t. Therefore,

$$[\exp(F_H s)]_{i,j} = \mathbb{P}(Y_s = j \mid Y_0 = i).$$

Then by the Markov property and the translation invariance of random walk,

$$[\exp(F_H s)]_{i,j} = \sum_{n=-\infty}^{\infty} \mathbb{P}(Z_s = j - i + nN | Z_0 = 0)$$

$$\leq \mathbb{P}(|Z_s| \geq \mathbf{d}(i,j)|Z_0 = 0) = 2\mathbb{P}(Z_s > \mathbf{d}(i,j)|Z_0 = 0).$$

Next we investigate the Laplace transform $\mathrm{e}^{\beta z}$. Applying the generator of Z_t,

$$\mathcal{A}\mathrm{e}^{\beta z} = \lambda_F \mathrm{e}^{\beta(z+1)} + \lambda_F \mathrm{e}^{\beta(z-1)} - 2\lambda_F \mathrm{e}^{\beta z} = \lambda_F(\mathrm{e}^\beta + \mathrm{e}^{-\beta} - 2)\mathrm{e}^{\beta z}.$$

By Dynkin's formula,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}\mathrm{e}^{\beta Z_t} = \lambda_F(\mathrm{e}^\beta + \mathrm{e}^{-\beta} - 2)\mathbb{E}\mathrm{e}^{\beta Z_t},$$

therefore, $\mathbb{E}(\mathrm{e}^{\beta Z_t} \mid Z_0 = 0) = \exp(\lambda_F t(\mathrm{e}^\beta + \mathrm{e}^{-\beta} - 2))$. By Markov inequality,

$$\mathbb{P}(Z_s > \mathbf{d}(i,j) \mid Z_0 = 0) \leq 2\exp(\lambda_F s(\mathrm{e}^\beta + \mathrm{e}^{-\beta} - 2) - \beta\mathbf{d}(i,j)).$$

This leads to our first claim on $[\exp(F_H s)]_{i,j} \leq 2\exp(s\lambda_G + \lambda_F s(\mathrm{e}^\beta + \mathrm{e}^{-\beta} - 2) - \beta\mathbf{d}(i,j))$.

Next we try to compute $\exp(H_* s)$. Because $(H_*/\lambda_H)^n = H_*/\lambda_H = \frac{1}{N}\mathbf{1}\cdot\mathbf{1}^\mathrm{T}$ for any $n \geq 0$, so

$$\exp(H_* s) = \sum_{n=0}^\infty \frac{s^n \lambda_H^n}{n!}\Big(\frac{H_*}{\lambda_H}\Big)^n = \frac{\mathrm{e}^{\lambda_H s} - 1}{\lambda_H}H_* + \mathbf{I}_N.$$

Then

$$[Q(s)]_{i,j} \le \sum_k [\exp(F_H s)]_{i,k}[\exp(H_* s)]_{k,j}$$

$$= [\exp(F_H s)]_{i,j} + \frac{1}{N}(\mathrm{e}^{\lambda_H s} - 1)\sum_{k=1}^{N}[\exp(F_H s)]_{i,k}$$

$$\le 2\exp(\lambda_G s + \lambda_F s(\mathrm{e}^{\beta} + \mathrm{e}^{-\beta} - 2))\Big(\exp(-\beta \mathbf{d}(i,j)) + \frac{\mathrm{e}^{\lambda_H s} - 1}{N}\sum_{k=1}^{N}\exp(-\beta \mathbf{d}(i,k))\Big)$$

$$\le 2\exp(\lambda_\beta s)\Big(\exp(-\beta \mathbf{d}(i,j)) + \frac{(1 + \mathrm{e}^{-\beta})(\mathrm{e}^{\lambda_H s} - 1)}{(1 - \mathrm{e}^{-\beta})N}\Big).$$

Its integral can be further bounded by

$$\int_0^t [Q(s)]_{i,j}\mathrm{d}s \le 2\int_0^t \exp(\lambda_\beta s)\Big(\exp(-\beta \mathbf{d}(i,j)) + \frac{(1 + \mathrm{e}^{-\beta})(\mathrm{e}^{\lambda_H s} - 1)}{(1 - \mathrm{e}^{-\beta})N}\Big)\mathrm{d}s$$

$$= \frac{2\exp(\lambda_\beta t) - 2}{\lambda_\beta}\Big(\mathrm{e}^{-\beta \mathbf{d}(i,j)} - \frac{1 + \mathrm{e}^{-\beta}}{(1 - \mathrm{e}^{-\beta})N}\Big) + \frac{2(\exp(\eta_\beta t) - 1)(1 + \mathrm{e}^{-\beta})}{\eta_\beta(1 - \mathrm{e}^{-\beta})N}.$$

B Analytic Solution of the Linear Model

For the linear model (4.1), collecting all the state variables $\mathbf{u} = (u_1, u_2, \cdots, u_N)$, the abstract form the the system is given by

$$\frac{\mathrm{d}\mathbf{u}}{\mathrm{d}t} = \mathbf{A}\mathbf{u} + \mathbf{\Sigma}\dot{\mathbf{W}}_u. \tag{B.1}$$

According to (4.1), the coefficient matrices $\mathbf{A}$ and $\mathbf{\Sigma}$ are given by

$$\mathbf{A} = \begin{pmatrix} -a & & & & & \\ & -a & & & & \\ & & -a & & & \\ & & & \ddots & & \\ & & & & -a & \\ & & & & & -a \end{pmatrix} + d_u \begin{pmatrix} -2 & 1 & & & & & 1 \\ 1 & -2 & 1 & & & & \\ & 1 & -2 & 1 & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & 1 & -2 & 1 \\ 1 & & & & 1 & -2 \end{pmatrix}$$

$$+ \frac{w}{N}\begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 & 1 \\ 1 & 1 & \cdots & \cdots & 1 & 1 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 1 & 1 & \cdots & \cdots & 1 & 1 \\ 1 & 1 & \cdots & \cdots & 1 & 1 \end{pmatrix} - w\begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{pmatrix},$$

$$\mathbf{\Sigma} = \sigma_u \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}.$$

It is easy to verify that $\mathbf{A}$ is a negative definite matrix. Therefore, eigenvalue decomposition allows the following

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathrm{T}},$$

where $\mathbf{\Lambda}$ is a diagonal matrix with all diagonal entries $\lambda_1, \cdots, \lambda_N$ being negative.

The solution of $\mathbf{u}$ is given by

$$\mathbf{u}(t) = \mathbf{u}(0)\mathrm{e}^{\mathbf{A}t} + \mathrm{e}^{\mathbf{A}t} \int_0^t \mathrm{e}^{-\mathbf{A}t}\mathbf{\Sigma}\mathrm{d}\mathbf{W}_u. \tag{B.2}$$

The time evolution of the covariance matrix is given by

$$\begin{aligned}
\langle \mathbf{u}_t'^2 \rangle &= \langle \mathbf{u}_0'^2 \rangle \mathrm{e}^{2\mathbf{A}t} + \sigma_u^2 \int_0^t \mathrm{e}^{2\mathbf{A}(t-s)}\mathrm{d}s \\
&= \langle \mathbf{u}_0'^2 \rangle \mathbf{Q}\mathrm{e}^{2\mathbf{\Lambda}t}\mathbf{Q}^{\mathrm{T}} + \sigma_u^2 \int_0^t \mathbf{Q}\mathrm{e}^{2\mathbf{\Lambda}(t-s)}\mathbf{Q}^{\mathrm{T}}\mathrm{d}s \\
&= \langle \mathbf{u}_0'^2 \rangle \mathbf{Q}\mathrm{e}^{2\mathbf{\Lambda}t}\mathbf{Q}^{\mathrm{T}} + \sigma_u^2 \mathbf{Q}\left\{ \frac{\mathrm{e}^{2\lambda_i t} - 1}{2\lambda_i} \right\}\mathbf{Q}^{\mathrm{T}},
\end{aligned} \tag{B.3}$$

where

$$\left\{ \frac{\mathrm{e}^{2\lambda_i t} - 1}{2\lambda_i} \right\} = \begin{pmatrix} \dfrac{\mathrm{e}^{2\lambda_1 t} - 1}{2\lambda_1} & & \\ & \ddots & \\ & & \dfrac{\mathrm{e}^{2\lambda_N t} - 1}{2\lambda_N} \end{pmatrix}.$$

References

[1] Anderson, B. D. and Moore, J. B., Optimal filtering, *Englewood Cliffs*, **21**, 1979, 22–95.

[2] Anderson, J. L., Exploring the need for localization in ensemble data assimilation using a hierarchical ensemble filter, *Physica D: Nonlinear Phenomena*, **230**(1–2), 2007, 99–111.

[3] Bai, Z. D. and Yin, Y. Q., Convergence to the semicircle law, *Ann. Probab.*, **16**(2), 1988, 863–875.

[4] Bellman, R. E., Adaptive Control Processes: A Guided Tour, **2045**, Princeton University Press, Princeton, NJ, 2015.

[5] Bickel, P. and Levina, E., Covariance regularization by thresholding, *The Annals of Statistics*, **36**(1), 2008, 199–227.

[6] Biktasheva, I. V., Holden, A. V. and Biktashev, V. N., Localization of response functions of spiral waves in the fitzhugh–nagumo system, *International Journal of Bifurcation and Chaos*, **16**(05), 2006, 1547–1555.

[7] Bui-Thanh, T., Ghattas, O., Martin, J. and Stadler, G., A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with application to global seismic inversion, *SIAM J. Sci. Comput.*, **36**(4), 2013, A2494–A2523.

[8] Chen, N. and Majda, A. J., Beating the curse of dimension with accurate statistics for the Fokker–Planck equation in complex turbulent systems, *Proc. Natl. Acad. Sci. USA*, **114**(49), 2017, 12864–12869.

[9] Chen, N. and Majda, A. J., Efficient statistically accurate algorithms for the Fokker–Planck equation in large dimensions, *J. Comput. Phys.*, **354**, 2018, 242–268.

[10] Chen, N., Majda, A. J. and Tong, X. T., Rigorous analysis for efficient statistically accurate algorithms for solving fokker-planck equations in large dimensions, arXiv:1709.05585.

[11] Chen, N., Majda, A. J. and Tong, X. T., Noisy lagrangian tracers for filtering random rotating compressible flows, *J. Non. Sci.*, **25**(3), 2014, 451–488.

[12] Chen, N., Majda, A. J. and Tong, X. T., Information barriers for noisy lagrangian tracers in filtering random incompressible flows, *Nonlinearity*, **27**, 2014, 2133–2163.

[13] Cotter, S. L., Roberts, G. O., Stuart, A. M. and White. D., MCMC methods for functions: modifying old algorithms to make them faster, *Stat. Sci.*, **28**(3), 2013, 424–446.

[14] Cui, T., Law, K. J. H. and Marzouk, Y. M., Dimension-independent likelihood-informed MCMC, *J. Comput. Phys.*, **304**, 2016, 109–137.

[15] Cui, T., Martin, J., Marzouk, Y. M. and et al., Likelihood-informed dimension reduction for nonlinear inverse problems, *Inverse Probl.*, **29**, 2014, 114015.

[16] DeVille, R. L., Vanden-Eijnden, E. and Muratov, C. B., Two distinct mechanisms of coherence in randomly perturbed dynamical systems, *Physical Review E*, **72**(3), 2005, 031105.

[17] Evensen, G., The ensemble Kalman filter: Theoretical formulation and practical implementation, *Ocean Dynamics*, **53**(4), 2003, 343–367.

[18] Evensen, G., Data Assimilation: The Ensemble Kalman Filter, Springer-Verlag, Berlin, 2009.

[19] Flath, H. P., Wilcox, L. C., Akçcelik, V. and et al., Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations, *SIAM J. Sci. Comput.*, **33**(1), 2011, 407–432.

[20] Guckenheimer, J. and Holmes, P., Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, Applied Mathematical Sciences, **42**, Springer-Verlag, New York, 1983.

[21] Hamill, T. M., Whitaker, C. and Snyder, C., Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, **129**, 2001, 2776–2790.

[22] Hamill, T. M., Whitaker, J. S. and Snyder, C., Distance-dependent filtering of background covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, **129**, 2001, 2776–2790.

[23] Holdenm, A. V., Markus, M. and Othmer, H. G., Nonlinear Wave Processes in Excitable Media, **244**, Springer Science & Buisiness Media, New York, 2013.

[24] Houtekamer, P. L. and Mitchell, H. L., Data assimilation using an ensemble kalman filter technique, *Mon. Wea. Rev.*, **126**(3), 1998, 796–811.

[25] Houtekamer, P. L. and Mitchell, H. L., A sequential ensemble Kalman filter for atmospheric data assimilation, *Mon. Weather Rev.*, **129**, 2001, 123–136.

[26] Houtekamer, P. L., Mitchell, H. L., Pellerin, G. and et al., Atmospheric data assimilation with an ensemble Kalman filter: Results with real observations, *Mon. Weather Rev.*, **133**, 2005, 604–620.

[27] Hunt, B. R., Kostelich, E. J. and Szunyogh, I., Efficient data assimilation for spatiotemporal chaos: a local ensemble transform Kalman filter, *Physica D*, **230**(1), 2007, 112–126.

[28] Iserles, A., How large is the exponential of a banded matrix? *New Zealand Journal of Mathematics*, **29**, 2000, 177–192.

[29] Kalman, R. E., A new approach to linear filtering and prediction problems, *Journal of basic Engineering*, **82**(1), 1960, 35–45.

[30] Kalnay, E., Atmospheric Modeling, Data Assimilation and Predictability, Cambridge university press. Cambridge, New York, 2003.

[31] Law, K., Stuart, A. and Zygalakis, K., Data Assimilation: A Mathematical Introduction, **62**, Springer-Verlag, cham, 2015.

[32] Lee, Y. and Majda, A. J., Multiscale methods for data assimilation in turbulent systems, *Multiscale Modeling & Simulation*, **13**(2), 2015, 691–713.

[33] Lindner, B., Garcıa-Ojalvo, J., Neiman, A. and Schimansky-Geier, L., Effects of noise in excitable systems, *Physics Reports*, **392**(6), 2004, 321–424.

[34] Lorenz, E. N., Predictability: A problem partly solved, *Proceedings of Seminar on Predictability*, **1**, ECMWF, Reading, Berkshire, UK, 1996, 1–18.

[35] Majda, A. J., Introduction to Turbulent Dynamical Systems in Complex Systems, Frontiers in Applied Dynamical Systems: Reviews and Tutorials, **5**, Springer-Verlag, Cham, 2016.

[36] Majda, A. J. and Harlim, J., Filtering Complex Turbulent Systems, Cambridge University Press, Cambridge, New York, 2012.

[37] Majda, A. J. and Wang, X. M., Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows, Cambridge University Press, Cambridge, New York, 2006.

[38] Majda, A. J., Qi, D. and Sapsis, T. P., Blended particle filters for large-dimensional chaotic dynamical systems, *Proceedings of the National Academy of Sciences*, **111**(21), 2014, 7511–7516.

[39] McKean, H. P., A class of markov processes associated with nonlinear parabolic equations, *Proceedings of the National Academy of Sciences*, **56**(6), 1966, 1907–1911.

[40] Miyoshi, T. and Yamane, S., Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution, *Mon. Wea. Rev.*, **135**(11), 2007, 3841–3861.

[41] Morzfeld, M., Tong, T. and Marzouk, Y. M., Localization for MCMC: Sampling high-dimensional posterior distributions with banded structure, arXiv:1710.07747.

[42] Nerger, L., On serial observation processing in localized ensemble kalman filters, *Mon. Wea. Rev.*, **143**(5), 2015, 1554–1567.

[43] Nualart, D., The Malliavin Calculus and Related Topics, Springer-Verlag, New York, 1995.

[44] Ohta, T., Mimura, M. and Kobayashi, R., Higher-dimensional localized patterns in excitable media, *Physica D: Nonlinear Phenomena*, **34**(1–2), 1989, 115–144.

[45] Perc, M., Spatial coherence resonance in excitable media, *Physical Review E*, **72**(1), 2005, 016207.

[46] Petra, N., Martin, J., Stadler, G. and Ghattas, O., A computational framework for infinite-dimensional Bayesian inverse problems, Part II: Stochastic Newton MCMC with application to ice sheet ow inverse problems, *SIAM J. Sci. Comput.*, **36**(4), 2014, 1525–1555.

[47] Powell, W. B., Approximate Dynamic Programming: Solving the Curses of Dimensionality, **703**, John Wiley & Sons, Hoboken, NJ, 2007.

[48] Rekleitis, I. M., A particle filter tutorial for mobile robot localization, Centre for Intelligent Machines, 2004, DOI: 10.1109/IROS.1996.570840.

[49] Ricketson, L. F., A multilevel monte carlo method for a class of mckean-vlasov processes, 2015, arXiv: 1508.02299.

[50] Sheard, S. A. and Mostashari, A., Principles of complex systems for systems engineering, *Systems Engineering*, **12**(4), 2009, 295–311.

[51] Spantini, A., Solonen, A., Cui, T. and et al., Optimal low-rank approximations of Bayesian linear inverse problems, *SIAM J. Sci. Comput.*, **37**(6), 2015, A2451–A2487.

[52] Stelling, J., Kremling, A., Ginkel, M. and et al., Foundations of Systems Biology, MIT Press, Cambridge, MA, 2001.

[53] Tong, X. T., Performance analysis of local ensemble Kalman filter, *J. Nonlinear Science*, **28**(4), 2018, 1397–1442.

[54] Tong, X. T., Majda, A. J. and Kelly, D., Nonlinear stability of the ensemble Kalman filter with adaptive covariance inflation, *Commun. Math. Sci.*, **14**(5), 2016, 1283–1313.

[55] Tong, X. T., Majda, A. J. and Kelly, D., Nonlinear stability and ergodicity of ensemble based Kalman filters, *Nonlinearity*, **29**, 2016, 657–691.

[56] Tyson, J. J. and Keener, J. P., Singular perturbation theory of traveling waves in excitable media (a review), *Physica D: Nonlinear Phenomena*, **32**(3), 1988, 327–361.

[57] Whitaker, J. S. and Hamill, T. M., Ensemble data assimilation without perturbed observations, *Mon. Wea. Rev.*, **130**(7), 2002, 1913–1924.