Estimation of Linear Space-Invariant Dynamics

Helmuth Naumer and Farzad Kamalabadi

***Abstract*—We propose a computationally efficient estimator for multi-dimensional linear space-invariant system dynamics with periodic boundary conditions that attains low mean squared error from very few temporal steps. By exploiting the inherent redundancy found in many real-world spatiotemporal systems, the estimator performance improves with the dimensionality of the system. This paper provides a detailed analysis of maximum likelihood estimation of the state transition operator in linear space-invariant systems driven by Gaussian noise. The key result of this work is that, by incorporating the space-invariance prior, the mean squared error of a estimator normalized to the number of parameters is upper bounded by $N^{-1}M^{-1} + O(N^{-1}M^{-2})$, where N is the number of spatial points, and M is the number of observed timesteps after the initial value.**

***Index Terms*—Space-invariant, system identification, spatiotemporal, dynamical systems.**

I. Introduction

Knowledge of the dynamics generating a signal enables numerous state-space based algorithms to provide both better estimates of the signal of interest in fields such as dynamic tomography [1], as well as predictions of the future which have critical implications such as weather patterns [2], [3]. Additionally, estimating physical dynamics is one of the core goals of basic scientific research, where such dynamics may represent fundamental governing equations in numerous fields. The inherent high-dimensionality of spatiotemporal systems normally requires both significant amounts of data and computation to create accurate estimates. Despite the high-dimensionality of the system, the dynamics themselves are often very structured.

Numerous approaches to estimating low-dimensional approximations of spatiotemporal dynamics exist. One successful method is Dynamic Mode Decomposition (DMD), which is inspired by the observation that a Linear Dynamical System (LDS) inherently constructs a Krylov Subspace, a space frequently exploited in efficient numerical algorithms for linear algebra [4]–[6]. While very useful, DMD inherently searches for low-dimensional approximations of the collective system dynamics, rather than the fundamental governing equations. Given two vibrating membranes, a square and a circle, DMD may find sinusoids in the square membrane and Bessel functions for the circular membrane, rather than identifying the wave equation and associated parameter.

Manuscript received September 4, 2020; accepted November 21, 2020. Date of publication November 26, 2020; date of current version December 18, 2020. This work was supported in part by the National Science Foundation under AGS-ATM Grant 1936663 and by NASA under Grant 80NSSC18M0052. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jun Liu. *(Corresponding author: Helmuth Naumer.)*

The authors are with the Department of Electrical and Computer Engineering and Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL 61801 USA (e-mail: hnaumer2@illinois.edu; farzadk@illinois.edu).

Digital Object Identifier 10.1109/LSP.2020.3040941

While understanding global behavior is very important for engineering design and interpreting the impact of boundary conditions, the universe is inherently governed by local interactions which together form these sophisticated behaviors. For this reason, it is sometimes preferable in basic scientific research to identify the fundamental governing equations and associated parameters. Recently proposed alternative formulations use approximations of derivatives to fit nonlinear functions to identify a governing partial differential equation (PDE) [7], [8], use neural networks to identify parameters [9], [10], or build particle simulation models using so-called “interaction networks” [11], [12], oftentimes using hundreds of timesteps. While these methods work incredibly well in practice, they often lack rigorous performance bounds due to the inclusion of necessary, but statistically opaque, steps, such as the sophisticated numerical differentiation procedure found in [13]. Thus, the goal of this work is to begin to understand the source of the remarkable performance of the numerous modern methods for estimating PDE-defined dynamics from data through the careful analysis of the most basic class of systems in the model family.

In this work, we derive an approximate mean squared error (MSE) upper bound for an estimator of linear system dynamics under a space-invariance prior, an assumption that inherently enforces the notion that collective behavior is determined by a set of identical interactions, without the requirement to first convert to continuous time. The paper begins by describing the basic system model and providing an approximate Cramer-Rao lower bound on the variance of unbiased estimators, before deriving the maximum likelihood (ML) estimator for a system that begins at rest. We then derive the bias and MSE of the estimator when applied to a system in steady state, and conclude with simulation results of a diffusion system.

II. System Model

The system follows a discrete-time linear state-space model

$$\mathbf{x}_{i+1} = \mathbf{F}\mathbf{x}_i + \mathbf{u}_i, \tag{1}$$

where $\mathbf{x}_i \in \mathbb{R}^N$ represents the state of the system at a given time i, and $\mathbf{u}_i \in \mathbb{R}^N$ represents i.i.d. centered Gaussian noise driving the system. Each element of the vectors correspond to a physical point in a space, discretized into N points. N is typically quite large, as even constructing a coarse $1000 \times 1000 \times 1000$ grid corresponds to the dimensionality of $N = 10^9$. Each vector is spatially wide sense stationary (WSS), and we assume knowledge of $\mathbf{x}_i$ in our estimator. The Linear Space-Invariant (LSI) model with periodic boundary conditions enforces that the state transition matrix $\mathbf{F} \in \mathbb{R}^{N\times N}$ is a convolution with a fixed convolutional kernel $\mathbf{f} \in \mathbb{R}^N$, represented by

$$\mathbf{F} = \begin{bmatrix}\mathbf{f} & \mathbf{S}\mathbf{f} & \mathbf{S}^2\mathbf{f} & \dots & \mathbf{S}^{N-1}\mathbf{f}\end{bmatrix}^\top, \tag{2}$$

where $\mathbf{S}$ represents a circular shift operator.

III. Cramer-Rao Bound

Before introducing a specific estimator, we derive an asymptotic Cramer-Rao bound [14] for the space-invariance prior. This bound shows the potential reduction in estimator variance enabled by the model assumption.

Lemma 3.1: Given M timesteps beyond the initial state of a linear, space-invariant system partitioned into N spatial points, the Cramer-Rao bound for the steady-state system asymptotically becomes

$$\mathbb{E}\left[\hat{\mathbf{f}}\hat{\mathbf{f}}^\top\right] \geq \frac{\sigma^2}{NM}\mathbf{C}_0^{-1}$$

for large M, where $\hat{\mathbf{f}}$ is an unbiased estimate of the convolutional kernel $\mathbf{f}$, $\hat{\mathbf{f}}^\top$ is the transpose of the estimate, σ^2 is the variance of the noise on a given timestep, and $\mathbf{C}_0 = \mathbb{E}[\mathbf{x}_i\mathbf{x}_i^\top]$ for all i is the spatial covariance matrix.

Proof: We will use the second derivative form of Fisher information,

$$\mathbf{J}(\mathbf{f}) = -\mathbb{E}\left[\frac{\partial^2\ell(\mathbf{f};\mathbf{x})}{\partial\mathbf{f}^\top\partial\mathbf{f}}\middle|\mathbf{f}\right], \tag{3}$$

where ℓ is the log likelihood function of the parameters $\mathbf{f}$ given observations $\mathbf{x}$, and $\mathbb{E}[\cdot]$ denotes the expected value. By taking note of the bipartite graph structure between timesteps, the log likelihood function takes the form

$$\ell(\mathbf{f};\mathbf{x}) = \log p(\mathbf{x}_0|\mathbf{f}) + \sum_{i=1}^{M}\sum_{j=1}^{N}\log p(x_j^{(i)}|\mathbf{x}_{i-1},\mathbf{f}), \tag{4}$$

where $x_j^{(i)}$ represents the state associated with location j at timestep i, and

$$\log p(x_j^{(i)}|\mathbf{x}_{i-1},\mathbf{f}) = -\frac{1}{2\sigma^2}(x_j^{(i)} - \mathbf{f}^\top\mathbf{S}^{1-j}\mathbf{x}_{i-1})^2 + c, \tag{5}$$

where c is a constant that does not depend on $\mathbf{f}$. For sufficiently many timesteps, we approximate the bound by removing $\log p(\mathbf{x}_1|\mathbf{f})$, using $\approx$ to denote the change to asymptotic equivalence. Computing the appropriate derivatives:

$$\frac{\partial^2\ell}{\partial\mathbf{f}^\top\partial\mathbf{f}} \approx \frac{1}{2\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{N} -2\mathbf{S}^{1-j}\mathbf{x}_{i-1}\mathbf{x}_{i-1}^\top\mathbf{S}^{j-1} \tag{6}$$

Finally, taking the expectation results in the Fisher information

$$\mathbf{J}(\mathbf{f}) \approx \frac{1}{2\sigma^2}\sum_{i=1}^{M}\sum_{j=1}^{N} 2\mathbf{S}^{1-j}\mathbb{E}\left[\mathbf{x}_{i-1}\mathbf{x}_{i-1}^\top\right]\mathbf{S}^{j-1} \tag{7}$$

$$= \frac{1}{\sigma^2}NM\mathbf{C}_0. \tag{8}$$

Equation (8) comes from noting the circulant structure of the covariance. We conclude by inverting the Fisher information.

Thus, if we assume some level of indistinguishability between the points in space, the points are each reused to estimate parameters, reducing the variance by a factor of N.

IV. LSI ML Estimator

In this section, we derive the maximum likelihood estimator of system dynamics in an LSI system. The derivation follows from expressing the state transition matrix $\mathbf{F}$ as being comprised of rows of the generating kernel $\mathbf{f}$ shifted by repeated application of a circular shift operator $\mathbf{S}$.

Theorem 4.1: The ML estimator of the state transition operator $\mathbf{F}$ of a discrete-time LSI system driven by centered i.i.d. Gaussian noise and initialized as $x_0 = u_0$ is

$$\hat{F}(\mathbf{\Omega}) = \frac{\sum_{i=1}^{M}\hat{S}_{i-1,i}(\mathbf{\Omega})}{\sum_{i=1}^{M}\hat{S}_{i-1,i-1}(\mathbf{\Omega})},$$

where $\hat{S}_{i,j}(\mathbf{\Omega}) = X_i^*(\mathbf{\Omega})X_j(\mathbf{\Omega})$ represents the empirical cross spectral density between timesteps i and j, and $\hat{F}(\mathbf{\Omega})$ is an estimate of the Fourier transform of the convolutional kernel of the system dynamics at spatial frequency $\mathbf{\Omega}$.

Proof: Starting from (4), the maximum likelihood estimator is

$$\arg\max_{\mathbf{f}\in\mathbb{R}^N}\ell(\mathbf{f}|\mathbf{x}) = \arg\max_{\mathbf{f}\in\mathbb{R}^N}\sum_{i=1}^{M}\sum_{j=1}^{N}\log p(x_j^{(i)}|\mathbf{x}_{i-1},\mathbf{f}), \tag{9}$$

where the first term was removed due to the lack of dependence on $\mathbf{f}$. Under our model of i.i.d. additive Gaussian noise, $p(x_j^{(i)}|\mathbf{x}_{i-1},\mathbf{f})$ becomes Gaussian with mean $(\mathbf{S}^{j-1}\mathbf{f})^\top\mathbf{x}_{i-1}$. By our model the noise is i.i.d., thus the scalars become irrelevant and we reduce our problem to

$$\arg\max_{\mathbf{f}\in\mathbb{R}^N}\sum_{i=1}^{M}\sum_{j=1}^{N} -(x_j^{(i)} - \mathbf{f}^\top\mathbf{S}^{1-j}\mathbf{x}_{i-1})^2. \tag{10}$$

Differentiating with respect to $\mathbf{f}$ and setting equal to 0:

$$\sum_{i=1}^{M}\sum_{j=1}^{N} 2(x_j^{(i)} - \mathbf{f}^\top\mathbf{S}^{1-j}\mathbf{x}_{i-1})(\mathbf{x}_{i-1}^\top\mathbf{S}^{j-1}) = 0 \tag{11}$$

Multiplying out the expression and shifting to opposite sides results in the maximum likelihood estimation of $\mathbf{f}$ to be the solution of the following equation:

$$\left(\sum_{i=1}^{M}\sum_{j=1}^{N}\mathbf{S}^{1-j}\mathbf{x}_{i-1}\mathbf{x}_{i-1}^\top\mathbf{S}^{j-1}\right)\mathbf{f} = \sum_{i=1}^{M}\sum_{j=1}^{N}x_j^i\mathbf{S}^{1-j}\mathbf{x}_{i-1} \tag{12}$$

We can consider $\sum_{i=1}^{M}\mathbf{x}_{i-1}\mathbf{x}_{i-1}^\top$ to be a scaled empirical spatial correlation matrix. Summing over our shift operators represents summing over all of the circular diagonal in our matrix, enforcing our prior of the circulant structure. Thus, if we denote our correctly normalized empirical spatial covariance matrix to be $\hat{\mathbf{C}}_0$, our equation becomes

$$NM\hat{\mathbf{C}}_0\mathbf{f} = \sum_{i=1}^{M}\sum_{j=1}^{N}x_j^i\mathbf{S}^{1-j}\mathbf{x}_{i-1}. \tag{13}$$

Next, the right hand side of the equation represents a single spatial cross-covariance vector between two adjacent timesteps. By introducing the notation $\hat{\mathbf{c}}_1 = M^{-1}N^{-1}\sum_{i=1}^{M}\sum_{j=1}^{N}x_j^i\mathbf{S}^{1-j}\mathbf{x}_{i-1}$ to be a correctly normalized estimate, the expression becomes

$$\hat{\mathbf{C}}_0\mathbf{f} = \hat{\mathbf{c}}_1. \tag{14}$$

The expression reduces to exactly what we would expect for such a task. We normalize the cross-correlation by the steady-state spatial distribution. The enforced circulant structure allows the estimation to be computed in the frequency domain through a spectral decomposition, and thus our estimate becomes

$$\hat{F}(\mathbf{\Omega}) = \frac{\sum_{i=1}^{M} \hat{S}_{i-1,i}(\mathbf{\Omega})}{\sum_{i=1}^{M} \hat{S}_{i-1,i-1}(\mathbf{\Omega})}, \tag{15}$$

where $\hat{S}_{i,j}(\mathbf{\Omega})$ is the Fourier transform of our empirical cross-correlation between $\mathbf{x}_i$ and $\mathbf{x}_j$ at spatial frequency $\mathbf{\Omega}$.

V. Properties

In this section, we derive the MSE of the estimator for a system in steady-state. It is important to note that this differs from the derivation of the estimator, and for steady-state, the estimator more closely resembles an ordinary least squares (OLS) estimator, rather than the actual ML estimator.

We begin by developing lemmas on the OLS estimator for complex circular autoregressive processes. The derivations use the Taylor Series approach from the classical work on real-valued AR(1) processes [15], [16]. Ultimately, the complex process reduces the bias slightly, and has little impact on the MSE. Rather than work with covariances, as was standard in the classical work, we work directly with moments due to ease of converting the relevant terms into each other.

We define a stable process that evolves according to $x_{i+1} = fx_i + u_i$, where $x_i \in \mathbb{C}$ takes complex scalar values and u_i are i.i.d. circular complex Gaussian random variables with $\mathbb{E}[|u_i|^2] = \sigma^2$, and $|f| < 1$, which we will refer to as a stable, circular CAR(1) process [17], [18]. We additionally define two random variables of interest relating to the OLS estimator, $X = \frac{1}{M}\sum_{i=0}^{M-1} x_{i+1}x_i^*$ and $Y = \frac{1}{M}\sum_{i=0}^{M-1} x_i x_i^*$, where x_i^* is the complex conjugate of x_i.

Lemma 5.1: Using the above definitions of X and Y, $\mathbb{E}[X]/\mathbb{E}[Y] = f$ and $\mathbb{E}[Y] = \frac{\sigma^2}{1-|f|^2}$.

The proof of lemma 5.1 is omitted as it follows directly from linearity of expectation and a geometric sum. We now define E_{XY} and E_{XX} which represent additive terms required to convert $\mathbb{E}[XY]$ or $\mathbb{E}[XX^*]$ to $\mathbb{E}[Y^2]$.

Lemma 5.2: Let $E_{XY} = \mathbb{E}[XY] - f\mathbb{E}[Y^2]$, then

$$E_{XY} = \frac{f\sigma^4}{M(1-|f|^2)^2} - \frac{f\sigma^4(1-|f|^{2M})}{M^2(1-|f|^2)^3}.$$

Proof: By expanding X, we note $\mathbb{E}[XY] = f\mathbb{E}[Y^2] + \frac{1}{M}\sum_{i=0}^{M-1}\mathbb{E}[u_i x_i^* Y]$. Expanding Y creates the expression

$$E_{XY} = \frac{1}{M^2}\sum_{i=0}^{M-1}\sum_{j=0}^{M-1}\mathbb{E}\left[u_i x_i^* x_j x_j^*\right]. \tag{16}$$

Now, note that u_i must be "matched" with a u_i^* term. Otherwise, independence between timesteps and $\mathbb{E}[u_i] = \mathbb{E}[u_i^2] = 0$ cause the full term to be zero. The only source of a u_i^* is x_j^* for $j > i$, so we reduce our expression to

$$E_{XY} = \frac{1}{M^2}\sum_{i=0}^{M-1}\sum_{j=i+1}^{M-1} f^{*|j-i-1|}\mathbb{E}\left[|u_i|^2 x_i^* x_j\right]. \tag{17}$$

Now, for the same reason, the remaining u_i terms inside of the x_i^* and x_j must match, resulting in the following:

$$E_{XY} = \frac{1}{M^2}\sum_{i=0}^{M-1}\sum_{j=i+1}^{M-1} f^{*|j-i-1|} f^{|j-i|}\mathbb{E}\left[|u_i|^2|x_i|^2\right] \tag{18}$$

$$= \frac{f}{|f|^2 M^2}\sum_{i=0}^{M-1}\sum_{j=i+1}^{M-1}|f|^{2|j-i|}\mathbb{E}\left[|u_i|^2|x_i|^2\right] \tag{19}$$

By construction u_i is independent of x_i, and so, by noting $\mathbb{E}[|u_i|^2] = \sigma^2$ and $\mathbb{E}[|x_i|^2] = \frac{\sigma^2}{1-|f|^2}$,

$$E_{XY} = \frac{f\sigma^4}{|f|^2 M^2(1-|f|^2)}\sum_{i=0}^{M-1}\sum_{j=i+1}^{M-1}|f|^{2|j-i|}. \tag{20}$$

By viewing the double summation as summing upper diagonals of a Toeplitz matrix populated with power of $|f|^2$, we note that the above can be simplified to our original claim. ∎

Lemma 5.3: Let $E_{XX} = \mathbb{E}[|X|^2] - |f|^2\mathbb{E}[Y^2]$, then

$$E_{XX} = f^* E_{XY}^* + fE_{XY} + \frac{\sigma^4}{M(1-|f|^2)}.$$

Proof: By expanding X and noting that Y is a real, non-negative random variable,

$$\mathbb{E}\left[XX^*\right] = \mathbb{E}\left[\left(fY + \sum_{i=0}^{M-1}\frac{u_i x_i^*}{M}\right)\left(f^*Y + \sum_{i=0}^{M-1}\frac{u_i^* x_i}{M}\right)\right]. \tag{21}$$

Then, by multiplying, recognizing the form of E_{XY}, and noting that complex conjugation commutes with expectation:

$$E_{XX} = f^* E_{XY} + fE_{XY}^* + \frac{1}{M^2}\mathbb{E}\left[\left(\sum_{i=0}^{M-1} u_i x_i^*\right)\left(\sum_{j=0}^{M-1} u_j^* x_j\right)\right]. \tag{22}$$

Finally, note that all terms in the final expectation other than when $i = j$ are zero, and we reduce to

$$E_{XX} = f^* E_{XY} + fE_{XY}^* + \frac{1}{M}\mathbb{E}\left[|u_i|^2|x_i|^2\right], \tag{23}$$

from which we conclude the desired result. ∎

Now, noting that the OLS estimator of an AR(1) correlation parameter takes the form $\hat{f} = X/Y$, we analyze the bias and MSE of our estimator.

Lemma 5.4: The bias of an OLS estimator using M timesteps of a stable circular CAR(1) process is

$$\mathbb{E}\left[\hat{f} - f\right] = -\frac{f}{M} + O(M^{-2}).$$

Proof: Note that the OLS estimator takes the form $\hat{f} = X/Y$. By using a Taylor series expansion of X/Y around the point $X = \mathbb{E}[X]$ and $Y = \mathbb{E}[Y]$, we find that

$$\frac{X}{Y} \approx \frac{\mathbb{E}[X]}{\mathbb{E}[Y]} - \frac{1}{\mathbb{E}[Y]^2}(Y - \mathbb{E}[Y])(X - \mathbb{E}[X]) + \frac{\mathbb{E}[X]}{\mathbb{E}[Y]^3}(Y - \mathbb{E}[Y])^2,$$

where the linear terms have been removed in anticipation of taking the expected value. By expanding the quadratic terms and applying the lemma 5.1, we find

$$\mathbb{E}\left[\frac{X}{Y}\right] \approx f\left(1 + \frac{\mathbb{E}\left[Y^2\right]}{\mathbb{E}[Y]^2}\right) - \frac{\mathbb{E}[XY]}{\mathbb{E}[Y]^2}. \tag{24}$$

Applying lemmas 5.1 and 5.2 results in

$$\mathbb{E}\left[\frac{X}{Y}\right] \approx f\left(1 - \frac{1}{M} + \frac{1}{M^2}\frac{(1-|f|^{2M})}{(1-|f|^2)}\right). \quad (25)$$

The method of statistical differentials using the 2nd order expansion in such averaging applications is known to be accurate to $O(M^{-2})$, and so we combine the final term with the inaccuracy of the expansion [15], [19]. ■

It is worth noting that, by multiplying by $M/(M-1)$, we can reduce the bias to $O(M^{-2})$. We now compute the variance of the OLS estimator through a similar method.

Lemma 5.5: The variance of the OLS estimator using M samples of a stable circular CAR(1) process which evolves according to the difference equation $x_{i+1} = fx_i + u_i$, where u_i is a centered circular complex Gaussian is $\frac{1-|f|^2}{M} + O(M^{-2})$.

Proof: By using the first order Taylor series approximation of $\frac{X}{Y}$, we note that the variance takes the form

$$\mathrm{var}\left(\frac{X}{Y}\right) \approx \frac{\mathbb{E}\left[|f|^2Y^2 - f^*XY - fX^*Y + XX^*\right]}{\mathbb{E}\left[Y\right]^2}. \quad (26)$$

By applying the definitions of E_{XX} and E_{XY}:

$$\mathrm{var}\left(\frac{X}{Y}\right) \approx \frac{E_{XX} - f^*E_{XY} - fE^*_{XY}}{\mathbb{E}\left[Y\right]^2} \quad (27)$$

Then, by applying lemmas 5.1 and 5.3, and noting that the first-order expansion was accurate to $O(M^{-1})$ we conclude that

$$\mathrm{var}\left(\frac{X}{Y}\right) = \frac{1-|f|^2}{M} + O(M^{-2}), \quad (28)$$

which is identical to the real-valued case [15]. ■

Next, we conclude the MSE of the OLS estimator.

Lemma 5.6: The MSE of the OLS estimator using M samples of a stable circular CAR(1) process is upper bounded by $M^{-1} + O(M^{-2})$

Proof: Note that the mean squared error can be decomposed in the form

$$\mathrm{mse} = |\mathrm{bias}|^2 + \mathrm{variance}. \quad (29)$$

From this decomposition, lemma 5.5 and lemma 5.4, we immediately conclude the result noting that $|f|^2 < 1$.

Finally, we apply the OLS CAR(1) results to the LSI model.

Theorem 5.1: Given a stable discrete-time LSI system with centered i.i.d. Gaussian noise, the maximum likelihood estimator of the state transition matrix has MSE upper bounded by $N^{-1}M^{-1} + O(N^{-1}M^{-2})$, normalized to the number of parameters.

Proof: We define the normalized mean squared error as

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N}(\hat{f}_j - f_j)^2\right], \quad (30)$$

where the f_j are the spatial values of the convolutional kernel. By the linearity of expectation and Parseval's theorem

$$\mathbb{E}\left[\frac{1}{N}\sum_{j=1}^{N}(\hat{f}_j - f_j)^2\right] = \frac{1}{N^2}\sum_{j=1}^{N}\mathbb{E}\left[(\hat{F}_j - F_j)^2\right] \quad (31)$$

$$\leq N^{-1}M^{-1} + O\left(N^{-1}M^{-2}\right), \quad (32)$$

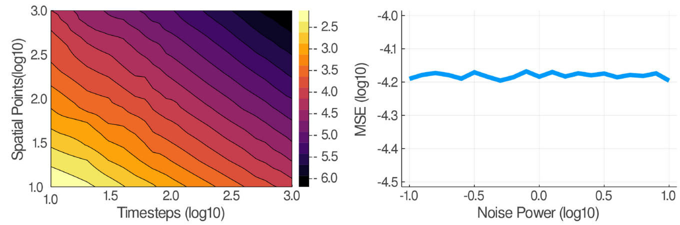

Fig. 1. Left Panel: MSE as a function of timesteps and spatial points. Right Panel: MSE as a function of noise power.

where the final line comes from noting each F_j element in the frequency domain evolves according to a circular CAR(1) process, and then applying lemma 5.5.

VI. Numerical Simulations

In this section, we provide numerical simulations of the maximum likelihood estimator to validate the claims.

A 1D finite-difference approximation of a system governed by diffusion was simulated (i.e. convolving the state vector with $[\alpha, 1-2\alpha, \alpha]$ at each timestep for some $\alpha \in [0, 0.5]$) with i.i.d. Gaussian noise added at each timestep. Because the estimator for higher spatial dimension (i.e 2D or 3D) systems follows directly by flattening the discretized space and treating it as a large 1D space, we expect the numerical results to be representative of higher dimensional problems. We computed an empirical MSE performance surface of the estimator as a function of the number of timesteps and the number of spatial points. Both parameters were varied from 10 to 1000, stepping by 10 from 10 to 100, then by 100 from 100 to 1000. The results can be seen in the left panel of Fig. 1. In the figure, both axis and the values are placed on a log scale to emphasize the linear relationship with mean squared error. The color represents the mean squared error on a log scale.

Additionally, to demonstrate the lack of dependence on noise, we included a plot of the mean squared error for a specific set of parameters in the right panel of Fig. 1. We simulated 100 timesteps with 100 spatial points for a fixed diffusion process and varied the amount of noise.

VII. Conclusion

The estimator presented in this paper is simple to implement and has performance that improves with the dimensionality of the system. A performance bound approximately independent of any unknown parameters exists, and thus it may be relatively easy to integrate the estimator into more sophisticated systems.

Acknowledgment

The author would like to thank P. Karimi for feedback on the clarity of this manuscript.

References

[1] M. D. Butala, R. A. Frazin, Y. Chen, and F. Kamalabadi, "Tomographic imaging of dynamic objects with the ensemble kalman filter," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1573–1587, Jul. 2009.

[2] C. Penland, "Random forcing and forecasting using principal oscillation pattern analysis," *Monthly Weather Rev.*, vol. 117, no. 10, pp. 2165–2185, Oct. 1989.

[3] C. Penland and T. Magorian, “Prediction of Niño 3 sea surface temperatures using linear inverse modeling,” *J. Climate*, vol. 6, no. 6, pp. 1067–1076, Jun. 1993.
[4] P. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *J. Fluid Mech.*, vol. 656, pp. 5–28, 2010.
[5] J. H. Tu, C. W. Rowley, D. M. Luchtenburg, S. L. Brunton, and J. N. Kutz, “On dynamic mode decomposition: Theory and applications,” *J. Comput. Dyn.*, vol. 1, no. 2, pp. 391–421, 2014.
[6] L. N. Trefethen and D. Bau, *Numerical Linear Algebra*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.
[7] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proc. Nat. Acad. Sci.*, vol. 113, no. 15, pp. 3932–3937, 2016.
[8] S. H. Rudy, S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Data-driven discovery of partial differential equations,” *Sci. Adv.*, vol. 3, no. 4, Apr. 2017, Art. no. e1602614.
[9] M. Raissi and G. E. Karniadakis, “Hidden physics models: Machine learning of nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 357, pp. 125–141, Mar. 2018.
[10] M. Raissi, P. Perdikaris, and G. Karniadakis, “Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations,” *J. Comput. Phys.*, vol. 378, pp. 686–707, Feb. 2019.
[11] P. Battaglia, R. Pascanu, M. Lai, D. J. Rezende, and K. Kavukcuoglu, “Interaction networks for learning about objects, relations and physics,” in *Proc. 30th Int. Conf. Neural Inf. Process. Syst.*, 2016, pp. 4509–4517.
[12] A. Sanchez-Gonzalez, J. Godwin, T. Pfaff, R. Ying, J. Leskovec, and P. W. Battaglia, “Learning to simulate complex physics with graph networks,” in *Proc. 37th Int. Conf. Mach. Learn.*, 2020, pp. 8459–8468.
[13] R. Chartrand, “Numerical differentiation of noisy, nonsmooth data,” *ISRN Appl. Math.*, vol. 2011, May 2011, Art. no. 164564.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hooboken, NJ, USA: Wiley-Interscience, 2006.
[15] M. Mudelsee, “Note on the bias in the estimation of the serial correlation coefficient of AR(1) processes,” *Stat. Papers*, vol. 42, no. 4, pp. 517–527, Oct. 2001.
[16] F. H. C. Marriott and J. A. Pope, “Bias in the estimation of autocorrelations,” *Biometrika*, vol. 41, nos. 3/4, pp. 390–402, 1954.
[17] B. Picinbono and P. Bondon, “Second-order statistics of complex signals,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 411–420, Feb. 1997.
[18] A. M. Sykulski, S. C. Olhede, and J. M. Lilly, “A widely linear complex autoregressive process of order one,” *IEEE Trans. Signal Process.*, vol. 64, no. 23, pp. 6200–6210, Dec. 2016.
[19] O. D. Anderson, “C427. The precision of approximations to moments of sample momentss,” *J. Stat. Comput. Simul.*, vol. 52, no. 2, pp. 163–165, Apr. 1995.