Stochastic Differential Equations with Additive Colored Noise

K. S. Hausknecht
8.334 Class Project, Massachusetts Institute of Technology
(Dated: May 17, 2025)

We develop a scalable, gradient-based framework for inferring linear stochastic differential equations (SDEs) driven by additive colored noise directly from data. Building on recent work in linear inverse modeling, we model colored noise as a weighted sum of Ornstein–Uhlenbeck noise with distinct correlation times and embed the dynamics into a higher-dimensional Markovian system. We derive moment-matching equations based on the correlation function and enforce them as differentiable soft constraints in a single optimization problem in order to learn SDE parameters from data. All algorithms are implemented in JAX, which uses automatic differentiation to make gradient computations essentially independent of parameter count. In simulation studies with bi-exponential and power law memory kernels, our method robustly recovers drift matrices, diffusion amplitudes, and multiple correlation times with low error.

I. INTRODUCTION

Stochastic differential equations (SDEs) underpin modern quantitative descriptions of noise-driven phenomena in statistical physics, biophysics, and many other fields. There is a large body of literature on inferring SDEs from data. Although some of the most interesting physical phenomena emerge from memory effects and nonlinear dynamics, most of the theoretical and computational machinery for data-driven inference of SDEs invoke strong simplifying assumptions, such as the Gaussian white noise (Markovian) approximation, that limit their applicability. If, for example, a system's dynamics involve non-Markovianity, non-Gaussianity, state-dependent multiplicative noise, non-linearity, or non-stationarity, most methods break down. Here, we focus on SDEs with additive colored noise (also referred to as correlated noise), with an eye toward extensions to multiplicative noise and nonlinear dynamics in future work.

In the white noise limit, the stochastic forcing is delta-correlated in time. This is a powerful simplification because analytical results can be obtained in this limit, and they yield accurate approximations for systems in which the noise decorrelates on timescales much shorter than either the system's intrinsic dynamics or the temporal resolution of the observations. For example, the Kramers-Moyal expansion and its Fokker-Planck truncation assume Markovian dynamics, and computational frameworks for inferring dynamics from data have been developed based on this formalism, e.g. inferring Kramers-Moyal coefficients [1].

Callaham *et al.* [2] discuss how naively applying Markovian frameworks to systems with colored noise leads to inaccurate characterization of the dynamics. To remedy this, they introduce a technique called Langevin Regression, which identifies and coarse-grains a time series beyond its "Einstein–Markov" timescale—the lag time beyond which the autocorrelation decays and the dynamics become effectively white—to infer effective stochastic models of systems with colored noise.

Colored noise processes exhibit nontrivial autocorrelation structures that cannot be captured by white-noise approximations—see, e.g., Łuczka [3] for a review. Systems in statistical physics that display anomalous diffusion or memory effects (e.g., transport in complex fluids [4]), and statistical models of neuronal activity [5, 6] often require colored-noise descriptions. Although learning an effective white noise model by coarse graining can be useful for reduced order modeling, it is not helpful for characterizing the true governing equations. A technique called Colored Linear Inverse Modeling (CLIM) presented in a recent paper [7] made progress toward data-driven recovery of SDE dynamics with colored noise, but only for the restricted case of linear dynamics driven by a single timescale Ornstein-Uhlenbeck process. We elaborate more on the details of their work in Section II.

In this project, we generalize their framework by extending it to a richer set of memory kernels and by providing a flexible gradient-based inference framework. The objectives of this project are twofold: (1) infer general non-Markovian temporal dynamics from data using Markovian embedding, and (2) implement the method with low computational cost. To achieve these goals, we implement all of the algorithms in JAX [8], which provides end-to-end automatic differentiation software, making gradient computations efficient and scalable to high-dimensional spaces.

Figure 2 shows a schematic of the methodology developed here. In Section II, we review the concept of Markovian embedding and specify the class of colored noise we will focus on. In section III, we obtain the moment-matching equations that will become soft constraints in our optimization framework. In Section IV, we provide the details of the full optimization procedure. In Section V, we present the results of two numerical experiments to test the methodology. Finally, in section VI, we summarize the main findings, limitations, and discuss areas for future research.

II. MARKOVIAN EMBEDDING

One approach to colored noise in SDEs that has been previously adopted in the literature is Markovian em-

bedding, a technique for mapping a given non-Markovian problem into a higher dimensional Markovian one by augmenting the state space with auxiliary variables (e.g., [5, 9–14]). The canonical example of this embedding is Ornstein-Uhlenbeck noise, a model for Gaussian colored noise that can be dynamically generated via

$$\frac{d}{dt}\boldsymbol{\eta}(t) = -\frac{1}{\tau}\boldsymbol{\eta}(t) + \frac{1}{\tau}\boldsymbol{\xi}(t) \tag{1}$$

where $\xi(t)$ is zero-mean Gaussian white noise with the property that $\langle \boldsymbol{\xi}(t+s)\boldsymbol{\xi}(t)^T \rangle = \delta(s)I$. It is straightforward to show that

$$\langle \boldsymbol{\eta}(t+s)\boldsymbol{\eta}(t)^T \rangle = \frac{1}{2\tau} \exp\left(-\frac{s}{\tau}\right) I. \tag{2}$$

When this is coupled into the SDE

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + \sqrt{2Q}\boldsymbol{\eta}(t), \tag{3}$$

the joint system $(\mathbf{x}, \boldsymbol{\eta})$ becomes Markovian in a state space of twice the original dimension.

The CLIM framework was introdueced specifically for the single timescale Ornstein-Uhlenbeck forcing described by Equation 3. In this special case, the moment equations close at the level of the 2nd derivative, and can be solved directly for A and Q. They then sweep through the parameter space to find the best τ to match the correlation function at non-zero lag.

Extending CLIM to richer colored-noise kernels that require higher-dimensional Markovian embeddings is non-trivial. One of the main limitations of Markovian embeddding, discussed more broadly in the literature (e.g., [15, 16]), is that augmenting the state space to achieve Markovianity can become analytically and/or computationally intractable.

In this work, we consider constructing more complex forms of colored noise by summing Ornstein-Uhlenbeck noise processes with distinct time scales by considering SDEs of the following form:

$$\frac{d\mathbf{x}}{dt} = A\mathbf{x} + \sum_{m=1}^{M} \sqrt{2Q_m}\boldsymbol{\eta}^{(m)}(t), \tag{4}$$

for $\mathbf{x} \in \mathbb{R}^N$ and

$$\frac{\mathrm{d}\boldsymbol{\eta}^{(m)}}{\mathrm{d}t} = -\frac{1}{\tau_m}\boldsymbol{\eta}^{(m)} + \frac{1}{\tau_m}\boldsymbol{\xi}^{(m)} \tag{5}$$

$$\left\langle \boldsymbol{\xi}^{(m)}(t+s)\boldsymbol{\xi}^{(n)}(t)^T \right\rangle = \delta_{mn}\delta(s)I. \tag{6}$$

In this setup, A is a Hurwitz matrix, i.e., the real part of all its eigenvalues must be negative while the Q_m are symmetric.

This can be embedded into a higher dimensional problem as follows

$$\frac{d}{dt}\tilde{\mathbf{x}} = \mathcal{A}\tilde{\mathbf{x}} + \mathcal{Q}\tilde{\mathbf{x}} \tag{7}$$

with

$$\tilde{\mathbf{x}} = \left[\mathbf{x} \;\; \boldsymbol{\eta}^{(1)} \;\; \boldsymbol{\eta}^{(2)} \;\; \cdots \;\; \boldsymbol{\eta}^{(M-1)} \;\; \boldsymbol{\eta}^{(M)},\right]^T$$

and the augmented drift and diffusion in block matrix form as follows.

$$\mathcal{A} = \begin{bmatrix} \mathbf{A} & \sqrt{2\mathbf{Q}_1} & \sqrt{2\mathbf{Q}_2} & \cdots & \sqrt{2\mathbf{Q}_{M-1}} & \sqrt{2\mathbf{Q}_M} \\ 0 & -\frac{1}{\tau_1}\mathbf{I} & 0 & \cdots & 0 & 0 \\ 0 & 0 & -\frac{1}{\tau_2}\mathbf{I} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -\frac{1}{\tau_{M-1}}\mathbf{I} & 0 \\ 0 & 0 & 0 & \cdots & 0 & -\frac{1}{\tau_M}\mathbf{I} \end{bmatrix}$$

$$\mathcal{Q} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \frac{1}{\tau_1}\mathbf{I} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \frac{1}{\tau_2}\mathbf{I} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{\tau_{M-1}}\mathbf{I} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \frac{1}{\tau_M}\mathbf{I} \end{bmatrix}.$$

Every additional memory timescale introduces another auxiliary variable. To circumvent the two problems of (1) moment equations that are too complicated to solve in closed form and (2) the increased size of the state space, we formulate this as an optimization problem in JAX. We impose the CLIM-style equations, derived in Section III, as differentiable soft constraints in the loss function and let the optimizer satisfy them. An important feature of automatic differentiation, which has made it indispensable for training large neural networks, is that the cost of gradient computations can be independent of the number of parameters, allowing us to accommodate larger Markovian embeddings without combinatorial penalty.

In Section V, we consider a bi-exponential memory kernel: to capture stochastic forcing with both a fast and slow memory timescale, we superpose two OU noises ($M = 2$) with distinct correlation times—a short-lived and a long-lived component.

A. Prony Series Approximation

Another motivation for pursuing the formulation in Equation 4, beyond its generality, is that a linear combination of decaying exponentials can approximate a power law kernel to any degree of accuracy [17]. This is known in the literature as a Prony Series, and it is of the general form

$$t^{-\alpha} = \sum_i w_i e^{-t/\tau_i}, \tag{8}$$

where the τ_i are logarithmically spaced. This approximation has been used extensively in the literature, often in the context of Markovian embedding as well [10–12, 18, 19]. In practice, because of the finite simulation

resolution, we choose the shortest τ_i to be the size of the integration step, and we choose the largest one based on the full length of the simulation. Figure 1 displays the finite Prony series approximation for a power law scaling of -2 with $M \in \{3, 4, 5, 6\}$ terms in the series. With more terms in the series, the approximation becomes increasingly accurate. In Section IV, we use $M = 6$ terms for our numerical experiments.

III. THEORY

In this section, we derive analogous equations to those in [7] for identifying the SDE parameters from data based on the correlation function. We find that many of the simplifications made in [7] to directly solve the equations are not possible in our more general setting. This motivates, as we discuss further in Section IV, implementing the analogous moment-matching equations as soft constraints in our JAX-based optimizer.

Because the augmented process is Markovian, its joint density satisfies the Fokker-Planck equation. Stationarity $\partial_t P_{\text{st}} = 0$ gives $\langle \mathcal{L}^* g \rangle = 0$ for any smooth g where $\mathcal{L}$ is the adjoint Fokker-Planck operator. Choosing $g = xx^T$, we find

$$\sum_{m=1}^{M} \left(\left\langle \mathbf{x}\boldsymbol{\eta}^{(m)T} \right\rangle \sqrt{2Q_m}^T + \sqrt{2Q_m} \left\langle \boldsymbol{\eta}^{(m)} \mathbf{x}^T \right\rangle \right) + AC_{xx} + C_{xx}A^T = 0 \tag{9}$$

where $C_{xx} = \langle \mathbf{x}\mathbf{x}^T \rangle$ and $C_{x\eta^{(m)}} = \langle \mathbf{x}\boldsymbol{\eta}^{(m)T} \rangle$. We can evaluate

$$\frac{d}{dt} C_{x\eta^{(m)}} = AC_{x\eta^{(m)}} - \frac{1}{\tau_m} C_{x\eta^{(m)}} + \frac{1}{\tau_m} \sqrt{\frac{Q_m}{2}} = 0, \tag{10}$$

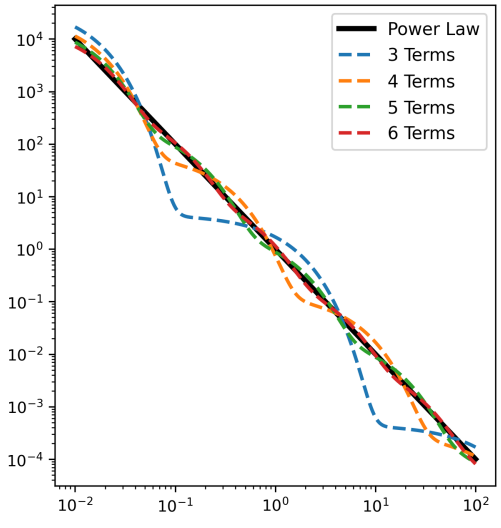

FIG. 1. The effect of adding successive terms to a Prony series approximation of a power law.

which implies

$$C_{x\eta^{(m)}} = (I - \tau_m A)^{-1} \sqrt{\frac{Q_m}{2}} \equiv B_m \sqrt{\frac{Q_m}{2}}. \tag{11}$$

Altogether this yields the Fluctuation-Dissipation Relation

$$AC_{xx} + C_{xx}A^T + \sum_{m=1}^{M} \left(B_m Q_m + Q_m B_m^T \right) = 0. \tag{12}$$

Furthermore, the correlation function is given by

$$\begin{aligned} C(s) &= \langle \mathbf{x}(t+s)\mathbf{x}(t)^T \rangle \\ &= e^{sA} C_{xx} + e^{sA} \sum_{m=1}^{M} \int_0^s e^{-u\left(A + \frac{1}{\tau_m} I\right)} Q_m B_m^\top \, du \end{aligned} \tag{13}$$

From this form, we obtain the following derivative hierarchy for the n-th derivative evaluated at 0 lag,

$$C^{(n)}(0) = AC^{(n-1)}(0) + \sum_m \left(-\frac{1}{\tau_m} \right)^{n-1} Q_m B_m^T. \tag{14}$$

For stability purposes, we only use up to the third derivative.

Unlike the single time scale OU noise case, where A and Q can be solved for analytically for a given τ, this cannot be achieved as simply here because we cannot eliminate Q_m from the second derivative equation. Thus, we implement the equations as soft constraints in an optimization problem to learn the dynamics from data.

IV. OPTIMIZATION

A schematic of the full optimization procedure is shown in Figure 2. We simulate trajectories for Equation 4 by constructing the higher dimensional Markovian system given by Equation 7 and using MATLAB's built-in Milstein method solver as an integrator. Once we have the trajectories, we compute the correlation functions and use a finite differences scheme to obtain the derivatives of the correlation function. Before proceeding to the optimization steps, we were able to check the validity of the equations derived in Section III by plugging in the parameters A, $\{Q_m\}$, $\{\tau_m\}$ used in the simulations and the correlation function and its derivatives calculated from the generated data. We found the equations to hold to a high degree of accuracy (up to the numerical error of the integration schemes). This step of the methodology is shown in Figure 2(a) and (b).

To guarantee that Q_m is symmetric positive semi-definite, we parameterize it by the $N(N+1)/2$ entries of a lower triangular matrix L_m so that in our objective function we compute Q_m as $Q_m = L_m L_m^T$. In our first implementation, we tried enforcing the Hurwitz condition on A as an inequality constraint, which was too computationally expensive. In the current implementation of the

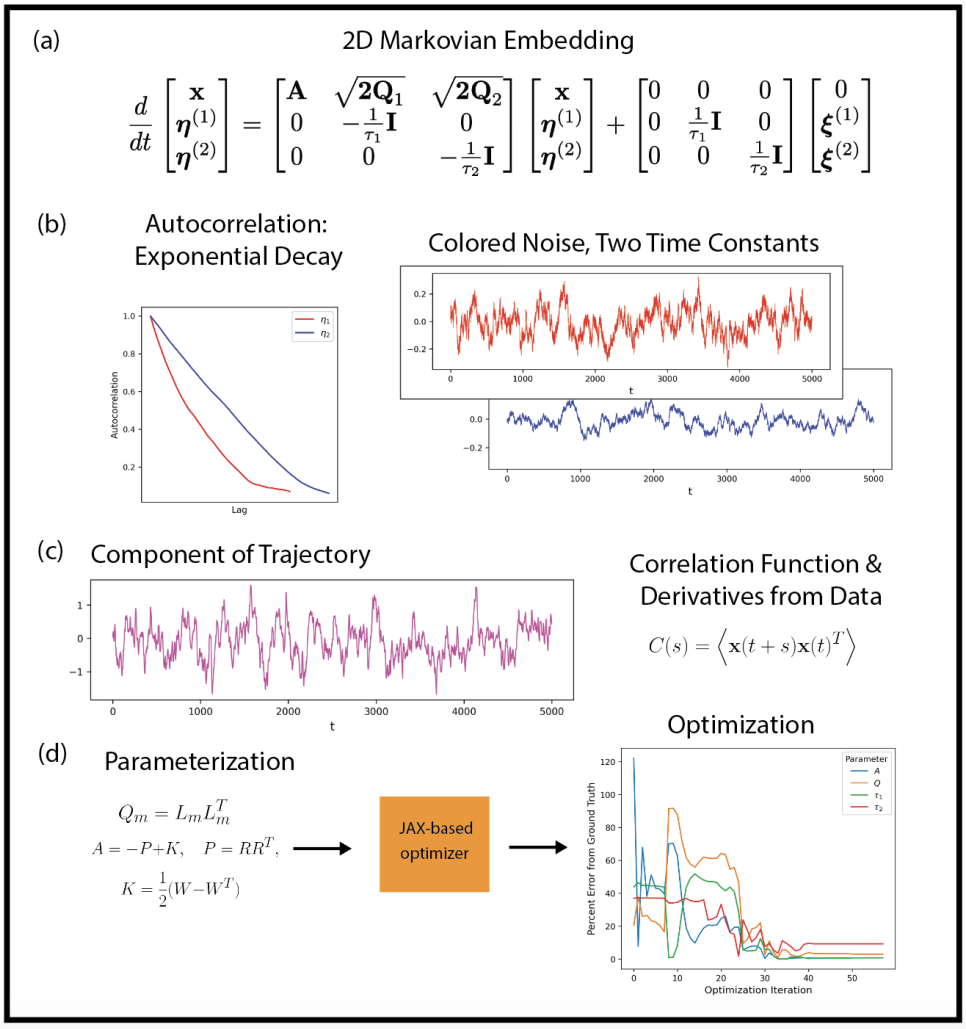

FIG. 2. Schematic of methodology: (a) Embedding the non-Markovian dynamics of interest into a higher dimensional Markovian problem, (b) simulating trajectories given the SDE, (c) using the trajectory to compute the correlation function and its derivatives, which become a part of the objective function, and (d) parameterizing the optimization variables and using the JAX-based optimizer to to recover the SDE parameters

methodology, we parameterize the set of Hurwitz matrices, by decomposing A as

$$A = -P + K, \quad P = RR^T, \quad K = \frac{1}{2}(W - W^T), \quad (15)$$

where R is lower traingular and W is unconstrained so that P is the symmetric positive-definite, and K is skew symmetric. While Q_m is an $N \times N$ matrix parameterized by fewer than N^2 parameters, the parameterization for A requires more than N^2 parameters. In practice, we found that this parameterization greatly improves the optimization. A key benefit of the method is that we have the freedom to increase the size of parameter space with virtually no slow-down in the optimizer.

The objective function that we optimize is exactly based on the equations in Section III. One term in the objective function is based on the FDR–namely we want to minimize the LHS Equation 12 with C_{xx} computed directly from the data and A, Q_m, and B_m, which depends on τ_m are optimized. Three other terms in the objective function come from minimizing the difference between the RHS and LHS of Equation 14 for the first, second, and third derivatives ($n = 1, 2, 3$). Here, $C^{(n)}(0)$ and $C^{(n-1)}(0)$ are computed directly from the data while the other terms in the equation are optimization variables. Another term that can be included in the objective function is a penalty on the mismatch between the modeled and empirical $C(s)$ for a few lag times versus.

Evaluating this full objective function, i.e., computing both sides of each equation, only requires linear algebra operations (multiplication, inversion, etc) that are implemented quite efficiently and in a way that is fully differentiable in JAX. So, we are able to compute accurate gradients of the objective function with respect to the SDE parameters. We use the interior point optimization algorithm as

our gradient-based optimizer [20]. A Python notebook containing the optimization codebase is provided here: `https://colab.research.google.com/drive/1QfLbTyMtrMjnBAcitRQ68EmUXdiWRjhf?usp=sharing`.

V. RESULTS

Here, we report the results from two experiments performed with our methodology (the linked python notebook contains others). In the first experiment, we consider a system with bi-exponential temporal correlations. That is, a system governed by Equation 4 with $M = 2$. We simulate 100 trajectories where we randomly sample values for τ_1, τ_2, A, and Q within some parameter range and use our optimizer to simultaneously recover all the parameters. Figure 3 shows the percent error from the ground truth of the parameters during the optimization. In this example, by 50 iterations, the optimizer has converged.

In Table 1, we report statistics on the 100 examples for which we ran the optimizer. These results demonstrate that the optimization framework developed here, which uses Markovian embedding and moment matching as soft constraints, is effective at recovering the SDE dynamics. We computed the statistics separately for the two time scales with $\tau_1 < \tau_2$ because from the equations in Section III, we expect the model to have higher error for the larger τ. This can be seen in the derivative hierarchy of Equation 14 where there is $\frac{1}{\tau}$ or $\frac{1}{\tau^2}$ dependence. We also sometimes encountered issues when the τ values were approximately equal.

In the second experiment, we simulate systems with power law temporal correlations. We vary the power law exponent, and use the optimizer with $M = 6$ to estimate the full dynamics of the SDE (Q, A, and the power law exponent). The goal of this experiment was to see how accurate the optimizer was at recovering the exponent.

Parameter	Median Percent Error
A	1.57%
Q	5.29%
τ_1	4.60%
τ_2	12.53%

TABLE I. Median percent errors of inferred SDE parameters on 100 different examples ($\tau_1 < \tau_2$)

We run the optimizer for 30 iterations on each example, and Figure 4 shows the estimated exponents versus the ground truth for a range of exponents. Besides a few cases, where the optimizer was likely stuck in a local minima and/or needed to be run for more iterations, we see that the framework is effective at learning the form of the stochastic forcing from the data.

VI. CONCLUSIONS & FUTURE WORK

In this work, we have introduced a general, gradient-based approach for inferring linear SDEs with a rich variety of additive colored-noise forms. We utilize Markovian embedding to map non-Markovian dynamics (SDEs with a stochastic forcing represented by a sum of Ornstein-Uhlenbeck noise with different timescales) into a higher-dimensional Markovian problem. We derive CLIM-style moment-derivative relations for these equations in Section III. Then, we enforce these equations as soft constraints in a JAX-based objective function. We generate simulated data and use the optimizer to to recover drift matrices, diffusion amplitudes, and multiple correlation times simultaneously. Our two numerical studies of bi-exponential memory kernels ($M = 2$) and power-law kernels (Prony series with $M = 6$) demonstrate that even with up to six Ornstein-Ullenbeck components, the method robustly recovers all parameters with low median

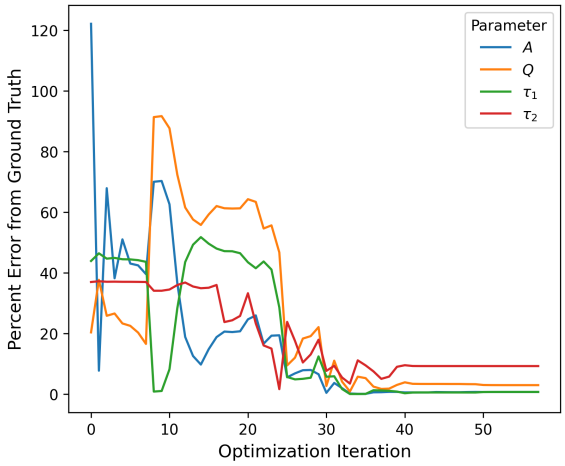

FIG. 3. Percent error of SDE parameters throughout optimization iterations where $\tau_1 < \tau_2$.

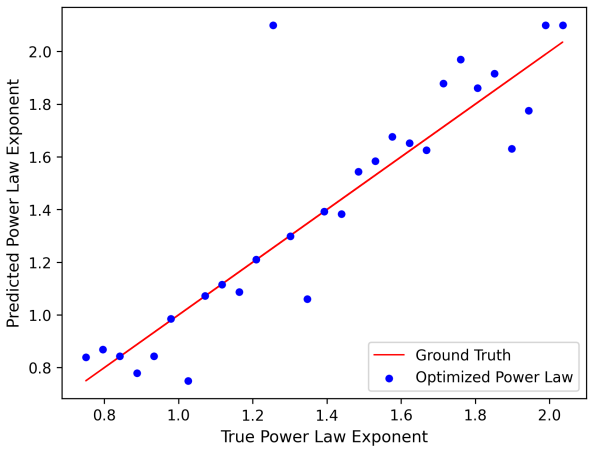

FIG. 4. Estimating the power law exponent of the temporal correlations from data

errors and converges in fewer than 50 iterations.

One limitation of our approach is that—even though the cost of each gradient computation is nearly independent of the number of OU components—embedding into higher dimensions can introduce additional degeneracies or local minima in the loss landscape. Addressing these may require adding moment-derivative constraints, which can increase the computational cost of the optimizer. Indeed, in my experiments, I saw examples of some degeneracies that can arise when time scales are chosen too close to each other for example. While the problem setting of linear SDEs with additive colored noise is restrictive, recent extensions of CLIM to nonlinear SDEs derive analogous moment equations that can be directly ported into our framework [21]. Since we have already formulated this as an optimization problem, we may be able to extend their work to learning arbitrary polynomials drifts, which they were not able to pursue in the paper. In future work, we hope to integrate these nonlinear constraints into our JAX-based loss to create a unified methodology for data-driven inference of both linear and nonlinear stochastic dynamics driven by colored noise.

VII. ACKNOWLEDGMENTS

I would like to thank Professor Kardar and Daniel Swartz for a wonderful semester. I have also benefited from conversations with Saeed Mahdisoltani and Justin Lien while working on this project.

[1] L. Gorjão and F. Meirinhos, kramersmoyal: Kramers–moyal coefficients for stochastic processes, Journal of Open Source Software **4**, 1693 (2019).

[2] J. L. Callaham, J.-C. Loiseau, G. Rigas, and S. L. Brunton, Nonlinear stochastic modelling with langevin regression, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **477**, 10.1098/rspa.2021.0092 (2021).

[3] J. Łuczka, Non-markovian stochastic processes: Colored noise, Chaos: An Interdisciplinary Journal of Nonlinear Science **15**, 026107 (2005).

[4] S. N. Gomes, G. A. Pavliotis, and U. Vaes, Mean-field limits for interacting diffusions with colored noise: phase transitions and spectral numerical methods (2020), arXiv:1904.05973 [math.NA].

[5] S. Vellmer and B. Lindner, Theory of spike-train power spectra for multidimensional integrate-and-fire neurons, Phys. Rev. Res. **1**, 023024 (2019).

[6] Q.-X. Li, Sheng-Hong; Zhu, Stochastic impact in fitzhugh–nagumo neural system with time delays driven by colored noises, Chinese Journal of Physics (Taipei) **56**, 346 (2018).

[7] J. Lien, Y.-N. Kuo, H. Ando, and S. Kido, Colored linear inverse model: A data-driven method for studying dynamical systems with temporally correlated stochasticity, Physical Review Research **7**, 10.1103/physrevresearch.7.023042 (2025).

[8] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang, JAX: composable transformations of Python+NumPy programs (2018).

[9] B. Farah, George N; Lindner, Exponentially distributed noise—its correlation function and its effect on nonlinear dynamics, Journal of Physics A : Mathematical and Theoretical **54**, 035003 (2020).

[10] N. Zilberstein, A. Sabharwal, and S. Segarra, Solving linear inverse problems using higher-order annealed langevin diffusion (2023), arXiv:2305.05014 [stat.ML].

[11] I. H.-P. Siegle, P.; Goychuk, Markovian embedding of fractional superdiffusion, Europhysics Letters **93**, 20002 (2011).

[12] T. Goychuk, Igor; Pöschel, Finite-range viscoelastic subdiffusion in disordered systems with inclusion of inertial effects, New Journal of Physics **22**, 113018 (2020).

[13] A. M. Barajas-Solano, David A.; Tartakovsky, Probabilistic density function method for nonlinear dynamical systems driven by colored noise, Physical Review E **93**, 052121 (2016).

[14] T. Albash, S. Young, and N. T. Jacobson, Temporal coarse graining for classical stochastic noise in quantum systems (2025), arXiv:2502.12296 [quant-ph].

[15] A. M. Barajas-Solano, David A.; Tartakovsky, Probabilistic density function method for nonlinear dynamical systems driven by colored noise, Physical Review E **93**, 052121 (2016).

[16] Y. M. R. S. J. Li, Hua; Xu, Transition path properties for one-dimensional non-markovian models, Journal of Physics A : Mathematical and Theoretical **57**, 355201 (2024).

[17] D. Bochud, Thierry; Challet, Optimal approximations of power laws with exponentials: application to volatility models with long memory, Quantitative Finance **7**, 585 (2007).

[18] A. D. Baczewski and S. D. Bond, Numerical integration of the extended variable generalized langevin equation with a positive prony representable memory kernel, The Journal of Chemical Physics **139**, 044107 (2013), https://pubs.aip.org/aip/jcp/article-pdf/doi/10.1063/1.4815917/14136018/044107_1_online.pdf.

[19] S. A. McKinley, L. Yao, and M. G. Forest, Transient anomalous diffusion of tracer particles in soft matter, Journal of Rheology **53**, 1487 (2009), https://pubs.aip.org/sor/jor/article-pdf/53/6/1487/16014115/1487_1_online.pdf.

[20] L. T. Wächter, Andreas; Biegler, On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming, Mathematical Programming **106**, 25 (2005).

[21] J. Lien and H. Ando, Beyond gaussian assumptions: A nonlinear generalization of linear inverse modeling (2025), arXiv:2503.24234 [math.NA].