Revisiting linear regression of dynamical systems within the context of Zwanzig-Mori theory: tests on a simple system

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

24-10-2021 / 26-10-2021

CITATION

Hsu, David; Mazrooyisebdani, Mohsen; Sears, Lucas Alan; Singh, Anshika; Silver, Mateo N; Boyarski, Makayla; et al. (2021): Revisiting linear regression of dynamical systems within the context of Zwanzig-Mori theory: tests on a simple system. TechRxiv. Preprint. <https://doi.org/10.36227/techrxiv.16864543.v1>

DOI

[10.36227/techrxiv.16864543.v1](https://doi.org/10.36227/techrxiv.16864543.v1)

Revisiting linear regression of dynamical systems within the context of Zwanzig-Mori theory: tests on a simple system

David A. Hsu, Mohsen Mazrooyisebdani, Lucas Alan Sears, Anshika Singh, Mateo N. Silver, Makayla Boyarski, Murielle A. Hsu, Shirui Chen, and Raheel Ahmed

Abstract—Linear regression can be applied to time series data to extract model parameters such as the effective force and friction constant matrices of the system. Even highly nonlinear systems can be analyzed by linear regression, if the total amount of data is broken up into shorter “time windows”, so that the dynamics is considered to be piece-wise linear. Traditionally, linear regression has been performed on the equation of motion itself (which approach we refer to as LRX). There has been surprisingly little published on the accuracy and reliability of LRX as applied to time series data. Here we show that linear regression can also be applied to the time correlation function of the dynamical observables (which approach we refer to as LRC), and that this approach is better justified within the context of statistical physics, namely, Zwanzig-Mori theory. We test LRC against LRX on a simple system of two damped harmonic oscillators driven by Gaussian random noise. We find that LRC allows one to improve the signal to noise ratio in a way that is not possible within LRX. Linear regression using time correlation functions (LRC) thus appears to be not only better justified theoretically, but it is more accurate and more versatile than LRX.

Index Terms—Causality, Langevin equation, linear regression, neural networks, Zwanzig-Mori theory.

I. INTRODUCTION

First introduced by Legendre and Gauss, least squares regression is a common first step in building a mathematical model of natural phenomena [1]. Least squares regression when applied to dynamical systems usually involves regression applied to the dynamical equation of motion itself. When the model parameters appear linearly in the postulated equation of motion, least squares regression becomes linear regression. Granger causality [2] is an example of linear least squares regression, while dynamic causal modeling [3] is an example of potentially nonlinear least squares regression. Complex systems such as the brain are nonlinear at macroscopic timescales. However, even highly nonlinear systems can be

considered to be piece-wise linear at shorter timescales, because if the energy surfaces on which the dynamics moves can be assumed to be differentiable, then the effective potential energy can be Taylor-expanded to second order at every point. One can then divide macroscopic timescales into shorter time windows, within each of which dynamics is linear. Such an “instantaneous normal mode” approach has proved useful for analysis of liquid state dynamics [4-6].

In applications to experimental time series data, such dynamical analysis can help us not only to better understand the dynamics of these systems, but also how to modify the dynamics through externally applied forces. For example, if by dynamical analysis of electroencephalogram data, we understand the effective macroscopic “forces” at play in triggering and maintaining a “seizure trajectory” in the brain, we may be able to design “counter-forces” to disrupt that trajectory using invasive electrical stimulation of the brain. This kind of analysis applied to brain stimulation can result in intelligent prediction and control algorithms, for example, using Kalman filters [7, 8], that would move the field beyond the current empirical trial-and-error standard [9, 10]. More broadly, dynamical analysis can be applied to any other time series data, such as functional magnetic resonance imaging or magnetoencephalogram data. It can help identify spontaneous networks underlying both normal and abnormal brain function, which would lead to better understanding of brain function in health and in disease.

Although least squares regression is widely applied to dynamical data, there is surprisingly little published on how well it performs on test systems. Such testing would need to be performed on computer generated trajectories of mathematical models for which the correct model parameters are known. In this study, we investigate how well least squares regression works for nearly the simplest possible dynamical model, that of two simple harmonic oscillators which move under the

This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE. (*Corresponding author: David Hsu*). David Hsu and Mohsen Mazrooyisebdani are co-first authors.

David Hsu and Murielle Hsu are with the Department of Neurology at the University of Wisconsin, Madison, WI 53705, USA (emails: hsu@neurology.wisc.edu, mvhsu@wisc.edu).

Mohsen Mazrooyisebdani is with the Department of Radiology at the University of Wisconsin, Madison, WI, 53792 USA (email: mazrooyisebd@wisc.edu).

Lucas Alan Sears, Anshika Singh, Mateo N. Silver, and Makayla Boyarski are undergraduate students at the University of Wisconsin, Madison WI, USA (emails: lasears@wisc.edu, singh94@wisc.edu, mnsilver@wisc.edu, boyarski@wisc.edu).

Shirui Chen was an undergraduate student at the University of Wisconsin, Madison WI, USA. He is now with the Department of Applied Mathematics at the University of Washington, Seattle WA, USA (email: sc256@uw.edu).

Raheel Ahmed is with the Department of Neurosurgery at the University of Wisconsin, Madison WI, USA (email: raheel.ahmed@neurosurgery.wisc.edu).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>

influence of friction and Gaussian random forces. We find that the conventional approach to least squares regression uniformly fails in its estimate of the effective friction constant for a dynamical system at thermal equilibrium. We suggest an alternative approach to least squares regression, where regression is applied to the equation of motion for the time correlation functions of the experimental observables (rather than on the experimental observables themselves) that may yield more accurate results. The motivation for looking at time correlation functions comes from revisiting the physical justification of least squares regression.

II. THEORY

Least squares regression can be grounded in physical law through Zwanzig-Mori theory [11-13], which connects microscopic dynamics to macroscopic observables. We refer to macroscopic observables as the explicit degrees of freedom, and all other degrees of freedom as implicit degrees of freedom. Projection operators then “project out” the implicit degrees of freedom such that one is left with an equation of motion for the explicit degrees of freedom that depends on the implicit degrees of freedom through complicated functions. This equation has a form equivalent to the generalized Langevin equation [14], which can be written:

$$\delta\ddot{X}(t) = -K\delta X(t) - G\delta\dot{X}(t) - \sum_{m=1}^M H(m)\delta\dot{X}(t-m\delta t) + R(t) \quad (1)$$

Here $X(t)$ is an N -dimensional time-dependent vector where each element of the vector is an explicit degree of freedom; $\delta X(t) = X(t) - \langle X(t) \rangle$ where $\langle \dots \rangle$ indicates a time average over a particular time window; $\delta\dot{X}(t)$ and $\delta\ddot{X}(t)$ are the first and second time derivatives of $\delta X(t)$; $[K, G, H(m)]$ are the harmonic force constant, friction and memory kernel matrices, respectively; $R(t)$ is a random force term; and the discretized time step is δt . The memory kernels $H(m)$ are time-lag matrices that determine the effect of prior values of $\delta\dot{X}(t-m\delta t)$ from m time steps in the past on the current value of the force, $\delta\ddot{X}(t)$. We are using mass weighted coordinates so that mass is absorbed into the coordinate.

In the engineering literature, it is more common to postulate an equivalent linear equation:

$$\delta X(t + \delta t) = \sum_{m=0}^M A(m)\delta X(t - m\delta t) + R'(t). \quad (2)$$

Here $A(m)$ are $N \times N$ time-lag matrices that determine the effect of prior values $\delta X(t - m\delta t)$ from m time steps in the past on the next future value of the explicit degree of freedom, $\delta X(t + \delta t)$, and $R'(t)$ is an N -dimensional error term. The matrices $A(m)$ can be related to the matrices $[K, G, H(m)]$ by comparing terms in (1) and (2). For example, for $M = 1$, we have:

$$A(0) = [2 - K(\delta t)^2]/[1 + (G\delta t/2)] \quad (3)$$

$$A(1) = -[1 - (G\delta t/2)]/[1 + (G\delta t/2)]. \quad (4)$$

Conventional least squares regression then consists of regarding either $R(t)$ or $R'(t)$ as the errors in the equations of motion in (1) and (2), then choosing either the K, G and $H(m)$'s so as to minimize the sum of the squares of $R(t)$ over some finite time window, or the $A(m)$'s so as to minimize the sum of the squares of $R'(t)$ over the same time window. Let us refer to this (conventional) type of least squares regression as LRX. Solutions for $B = [K, G, H]$ using LRX are derived in Appendix A.

Within Zwanzig-Mori theory, the random force terms $R(t)$ and, equivalently, $R'(t)$, represent contributions from the implicit degrees of freedom which are not necessarily small. Thus there is no physical justification for the assumption underlying LRX [8]. On the other hand, a defining characteristic of the random force, $R(t)$, is that it is not correlated with $X(t)$ when averaged over phase space [11-13]. If we further invoke the ergodic hypothesis [15], which replaces a phase space average with a time average, then one can apply least squares regression to time correlation functions of the explicit degrees of freedom, rather than the explicit degrees of freedom themselves. The error to be minimized is then the square of the time correlation between $X(t)$, which represents signal, and $R(t)$, which represents noise. In effect we are defining noise to be that which is not time-correlated with signal. Let us refer to this type of least squares regression as LRC. Solutions for $B = [K, G, H]$ using LRC are derived in Appendix B.

Notably, both LRX and LRC are constrained by the ergodic hypothesis; that is, if the system is trapped in a local part of phase space by energy barriers, then one cannot obtain information about any part of phase space that is not being sampled in that time window. In comparing the theoretical justifications of LRX and LRC, the main difference is that LRX additionally requires that the random force be small compared to the forces due to $B = [K, G, H]$.

In what follows, we drop the memory terms $H(m)$ with the goal of understanding the simpler system first. The memory terms are certain to be important in complex systems such as the brain, but G may be viewed as the zeroth order memory term in (1); it is typically larger than higher order memory terms.

It is useful to monitor the total instantaneous energy and time rate of change of the energy. Within any given time window, the total energy of the explicit degrees of freedom is given by:

$$E(t) = \frac{1}{2}\delta\dot{X}(t)^T\delta\dot{X}(t) + \frac{1}{2}\delta X(t)^TK\delta X(t). \quad (5)$$

Here the superscript T indicates a transpose. The time-averaged time-rate of change of the total energy is then:

$$\langle \dot{E} \rangle = Tr[\bar{K}\langle\delta\dot{X}\delta X^T\rangle] - Tr[\bar{G}\langle\delta\dot{X}\delta\dot{X}^T\rangle] + \langle R^T\delta\dot{X} \rangle. \quad (6)$$

Here Tr indicates a matrix trace and:

$$\bar{K} = \frac{1}{2}(K - K^T) \quad (7)$$

$$\bar{G} = \frac{1}{2}(G + G^T). \quad (8)$$

The first term in (6) involves $\bar{K}$, which is an antisymmetric matrix. Force constant matrices that are derivable from a differentiable potential energy are symmetric, in which case $\bar{K} = 0$. In the general case for a macroscopic system, we leave open the possibility that $\bar{K}$ is not necessarily symmetric. The second term in (6) is the time rate of energy loss due to friction, i.e., the dissipation:

$$\langle \dot{E}_{diss} \rangle = Tr[\bar{G}\langle \delta \dot{X} \delta \dot{X}^T \rangle]. \quad (9)$$

At steady state, the dissipation term must be balanced by the third term in (6), which is the rate of energy influx due to the random force. The lowest order contribution to the third term requires expansion of the current velocity to first order in the time step, δt . The velocity at the current time is then:

$$\delta \dot{X}(t) = \delta \dot{X}(t - \delta t) + \frac{1}{2} \delta t [\delta \ddot{X}(t - \delta t) + \delta \ddot{X}(t)] \quad (10)$$

Inserting (10) into (6), then using (1) and keeping only the lowest order terms in δt , yields the fluctuation-dissipation theorem for the simple Langevin equation [14]:

$$\langle R^T R \rangle = \frac{2}{\delta t} Tr[\bar{G}\langle \delta \dot{X} \delta \dot{X}^T \rangle] \quad (11)$$

III. COMPUTER-GENERATED MODEL

We tested LRX vs LRC on computer generated modeling of a system of two damped harmonic oscillators obeying the equation of motion:

$$\delta \ddot{X}(t) = -K\delta X(t) - G\delta \dot{X}(t) + R(t) \quad (12)$$

where:

$$K = \begin{pmatrix} 986.96 & 0.39478 \\ 1.57913 & 19107.55 \end{pmatrix} \quad (13)$$

$$G = \begin{pmatrix} G_0 & 0 \\ 0 & G_0 \end{pmatrix}. \quad (14)$$

The choice of K in (13) corresponds to two asymmetrically coupled harmonic oscillators, one with resonant frequency of 5 Hz, the other with resonant frequency of 22 Hz. The random force is taken to be Gaussian with zero mean and root mean amplitude of R_0 ; R_0 and G_0 are each varied over values of 0.01, 1 and 100. The actual values for $[K, G, R_0]$ used to generate the data are referred to as the true values. The initial conditions are $X(0) = 1$ and $\dot{X}(0) = 0$. Trajectories for $X(t)$ are calculated using a leap-frog algorithm with time steps of $\delta t = 0.001$ sec for a total of 250 seconds per free-running simulation run. LRX and LRC are then applied to consecutive 10-second time-windows yield estimated values for $[K, G]$, one set of $[K, G]$, for each 10-second time window, which are then compared to the

true values. The time rate of change of the energy (6) is tracked during each run so that we know whether the system has reached steady state. For the values that we tested, a time of 40 seconds is long enough to insure that thermal equilibrium has been reached.

In order to simulate the effect of an external perturbation on the performance of LRX and LRC, at time $t = 250$ sec for each simulation run, we turn on 50 Hz sinusoidal stimulation with amplitude of 1000. This choice of parameters models cortical mapping as performed on patients undergoing epilepsy surgery workup [16, 17]. The stimulation is turned on and off in consecutive alternating time windows, and LRX and LRC are applied to time windows when the stimulation is off. An additional 1500 time windows are analyzed in this way, for a total of 15,250 seconds of data analyzed (250 seconds of free-running data, plus 15,000 seconds of post-stimulation data) for each net simulation run.

For the post-stimulation analysis using LRC, there is an additional option of cumulatively averaging the time correlation functions calculated in each post-stimulation time window. For instance, let $C_i(t)$, indicate the time correlation function corresponding to the i^{th} time window, and let N_W indicate the number of post-stimulation time windows acquired thus far, where N_W ranges from 1 to 15,000. Then define:

$$C_{avg}(t; N_W) = \frac{1}{N_W} \sum_{i=1}^{N_W} C_i(t). \quad (15)$$

An effective K and G are then calculated as in Appendix B, but using $C_{avg}(t; N_W)$ rather than the individual $C_i(t)$'s. Such a cumulative average is not available within LRX.

IV. RESULTS

Figure 1 shows the relative errors in LRX and LRC estimates of K vs time with R_0 varied over values of 0.01, 1 and 100, and with the true G_0 held fixed at 1 Hz. Each point on these graphs represents the relative error of an estimated value for K within one time window, with each time window being 10 seconds in duration. External stimulation is turned on for consecutive alternating time windows starting at a time of 250 seconds. For LRC, a cumulative average of the time correlation function as defined in (15) is calculated, starting with the first post-stimulation time window. This cumulatively averaged time correlation function is then used to calculate the LRC estimates of K and G .

Figure 1 shows that both LRX and LRC produce reasonably accurate estimates of K in the first time window ($t = 10$ sec), at thermal equilibrium ($t = 40$ to 250) and in the post-stimulation time windows ($t = 250$ to 15,250). However, the LRX estimate for K shows progressive deterioration in the post-stimulation time windows. This effect is due to a heating-up effect of the externally applied stimulation. With repeated applications of the externally applied stimulation, the overall thermal energy increases, which would then be expected to reduce the signal to noise ratio. In contrast, the LRC estimate for K shows an abrupt improvement with the post-stimulation time-windows which is maintained with increasing trials of stimulation. As the root

mean square amplitude of the random force, R_0 , is increased, the main effect is that larger R_0 leads to less accurate estimates of G , which is to be expected as larger R_0 results greater thermal noise and a smaller signal to noise ratio.

Fig. 1. Top panel: relative error in K using LRX with $R_0 = 0.01$, 1, and 100. Bottom panel: relative error in K using LRC with $R_0 = 0.01$, 1, and 100. Note Log-Log scales for relative error and time. The true G_0 is held fixed at 1 Hz.

Figure 2 shows the relative errors in LRX and LRC estimates of G vs time with R_0 varied over values of 0.01, 1 and 100, and with G_0 held fixed at 1 Hz. Each point on these graphs represents the relative error of an estimated value for G within one time window, with each time window being 10 seconds in duration. External stimulation is turned on for consecutive alternating time windows starting at a time of 250 seconds. For LRC, a cumulative average of the time correlation function as defined in (15) is calculated, starting with the first post-stimulation time window. This cumulatively averaged time correlation function is then used to calculate the LRC estimates of K and G .

Figure 2 shows that both LRX and LRC produce reasonably accurate estimates of G in the first time window, when the system is not thermally equilibrated, and that both perform poorly after the system is thermally equilibrated ($t = 40$ to 250), with errors of $\sim 100\%$ at thermal equilibrium. When external stimulation is applied, thus driving the system away from thermal equilibrium, both LRX and LRC improve in their estimates of G , more so when the random noise is small or moderate ($R_0 = 0.01$ or 1), less well when it is large ($R_0 = 100$). LRX again shows a heating-up effect where repeated applications of externally applied stimulation result in gradually degraded estimates of G . In contrast, not only is the heating-up effect absent for LRC, but there is a slow gradual improvement

in the estimated G as more post-stimulation time windows are accumulated.

Fig. 2. Top panel: relative error in G using LRX with $R_0 = 0.01$, 1, and 100. Bottom panel: relative error in G using LRC with $R_0 = 0.01$, 1, and 100. Note Log-Log scales for relative error and time. The true G_0 is held fixed at 1 Hz.

Figure 3 shows the relative errors in LRX and LRC estimates of K vs time with the true G_0 varied over values of 0.01, 1 and 100, and with R_0 held fixed at 1. Each point on these graphs represents the relative error of an estimated value for K within one time window, with each time window being 10 seconds in duration. External stimulation is turned on for consecutive alternating time windows starting at a time of 250 seconds. For LRC, a cumulative average of the time correlation function as defined in (15) is calculated, starting with the first post-stimulation time window. This cumulatively averaged time correlation function is then used to calculate the LRC estimates of K and G .

Figure 3 shows that both LRX and LRC produce reasonably accurate estimates of K in the first time window ($t = 10$ sec), at thermal equilibrium ($t = 40$ to 250) and in the post-stimulation time windows ($t = 250$ to 15,250), as long as G_0 is not too large. At large G_0 , the signal is damped too fast, thus reducing the number of oscillatory cycles the signal exhibits before the oscillation becomes too small to be detected. In addition, the LRX estimate for K shows progressive deterioration in the post-stimulation time windows. This effect is due to a heating-up effect of the externally applied stimulation. With repeated applications of the externally applied stimulation, the overall thermal energy increases, which then reduces the signal to noise ratio. In contrast, the LRC estimate for K shows an abrupt

improvement with the post-stimulation time-windows which is maintained with increasing trials of stimulation.

Fig. 3. Top panel: relative error in K using LRX with true $G_0 = 0.01, 1$, and 100 . Bottom panel: relative error in K using LRC with true $G_0 = 0.01, 1$, and 100 . Note Log-Log scales for relative error and time. R_0 is held fixed at 1 .

Figure 4 shows the relative errors in LRX and LRC estimates of G vs time with the true G_0 varied over values of $0.01, 1$ and 100 , and with R_0 held fixed at 1 . Each point on these graphs represents the relative error of an estimated value for K or G within one time window, with each time window being 10 seconds in duration. External stimulation is turned on for consecutive alternating time windows starting at a time of 250 seconds. For LRC, a cumulative average of the time correlation function as defined in (15) is calculated, starting with the first post-stimulation time window. This cumulatively averaged time correlation function is then used to calculate the LRC estimates of K and G .

Figure 4 shows that both LRX and LRC produce reasonably accurate estimates of G in the first time window, when the system is not thermally equilibrated, and that both perform poorly after the system is thermally equilibrated ($t = 40$ to 250), with errors of $\sim 100\%$ at thermal equilibrium. When external stimulation is applied, thus driving the system away from thermal equilibrium, both LRX and LRC improve in their estimates of G . LRX again shows a heating-up effect where repeated applications of externally applied stimulation result in gradually degraded estimates of G . In contrast, the heating-up effect is absent for LRC; for moderate values of the friction ($G_0=1$), there is a slow gradual improvement in the estimated G as more post-stimulation time windows are accumulated. Note

that relative errors for G_0 are largest both for very small and very large values. For small values of the true G_0 , an accurate estimate is more challenging because the relative error is taken with respect to the true value of G_0 , so that a small absolute error translates to a large relative error. For large values of G_0 , an accurate estimate is challenging because the signal damps out faster, leaving a shorter time duration of the damping dynamics to analyze.

Fig. 4. Top panel: relative error in G using LRX with true $G_0 = 0.01, 1$, and 100 . Bottom panel: relative error in G using LRC with true $G_0 = 0.01, 1$, and 100 . Note Log-Log scales for relative error and time. R_0 is held fixed at 1 .

It is instructive to look at the relative errors using (1) vs (2). If one uses (2) to fit time series data, it is possible to be fooled into thinking one has an accurate solution if one calculates only the relative errors in the estimates for $A(0)$ and $A(1)$. To show this, consider the case of two uncoupled harmonic oscillators with frequencies of 5 and 222 Hz, with $G_0 = 1$ Hz, and $R_0 = 0.005$. The relative errors for $A(0)$ and $A(1)$ are very small whether extracted using either LRX or LRC (see Table 1). However, if these values are converted to the corresponding K 's and G 's, one would see that the friction constant G extracted using LRX is off by 13.98% . The reason that estimates for the relative errors in $A(0)$ and $A(1)$ can be so misleading is because $A(0)$ and $A(1)$ are to lowest order in the timestep, δt , not dependent on either K or G (see (3) and (4)).

TABLE I

Relative errors in estimates of model parameters for the case where the harmonic oscillator frequencies are 5 and 222 Hz, the G_0 is 1 Hz, and $R_0 = 0.005$.

Relative error in:	LRX	LRC
--------------------	-----	-----

A(0)	6.251×10^{-5}	4.851×10^{-5}
A(1)	1.049×10^{-4}	1.906×10^{-7}
K	2.499×10^{-4}	2.499×10^{-4}
G	0.1398	2.547×10^{-4}

In additional efforts to improve the estimates of K and G when the system is at steady state, we tried using both longer but fewer time windows, and shorter and more time windows, such that the total time remained the same. We found that neither longer nor shorter time windows improved the estimates. We also tried accumulating individual K 's and G 's, each one extracted from individual time windows, then averaging these K 's and G 's across time windows. This procedure also failed to improve the estimates. That these two procedures failed to improve the estimated K 's and G 's is not surprising, given that neither procedure improves the signal to noise ratio. For instance, a longer time window will include both more signal and more noise.

We additionally tested other values of the K and G matrices, including a wider frequency gap and both diagonal and nondiagonal versions. Results were concordant with those presented above.

V. DISCUSSION

For the highly simplified dynamical system consisting of two coupled, damped harmonic oscillators which start from a nonequilibrium initial condition and is then allowed to evolve under the influence of a Gaussian random force, both the standard approach to linear regression of dynamical systems (LRX) and the time correlation function approach proposed here (LRC) perform well in extracting an effective force constant matrix (K), but both perform poorly in extracting an effective friction constant (G), *if the system is allowed to come to thermal equilibrium*. The reason that friction constant estimates are poor at thermal equilibrium is because a defining feature of steady state dynamics is that energy flow into a system (largely due to random noise forces) must on average be balanced by the energy flow out of the system (largely due to friction). This equality is the basis of the fluctuation-dissipation theorem [14]. In this case, steady state is reached when the magnitudes of the random and frictional forces are equal, which makes it difficult to separate random from frictional forces. With this insight, it becomes obvious that estimates of the friction constant can be improved by perturbing the system away from thermal equilibrium, for example, by applying an external perturbation, or by averaging the equilibrium time correlation function across multiple time windows, as in (15). Averaging the time correlation function increases the signal to noise ratio in the time correlation function and improves LRC estimates of the friction constant. An analogous averaging process is not available within the LRX framework, as averaging the explicit degrees of freedom themselves merely results in constants which contain no dynamical information.

A caveat for repeat applications of an external stimulation is that over time, the system can “heat up,” and increase the

amount of background thermal noise. This increased thermal noise degrades estimates of both the force constant and friction matrices within the LRX framework but not within the LRC framework. A caveat for averaging the time correlation function across multiple time windows within the LRC framework is that one must insure that the system is in the same “state” within each time window analyzed. For example, if the system is highly nonlinear (as the brain is), then it may enter a different state at certain times, each with its own best force constant and friction matrices. The pairing of a standardized protocol for application of an external stimulus followed by post-stimulus sampling may help insure that the system under study is in the same state for each time window sampled.

The effective friction is an important model parameter for all real-world applications. In applications to neural systems, for instance to the detection and monitoring of neural activity via electroencephalograms, magnetoencephalograms or functional magnetic resonance imaging, we hypothesize that (1) the force constant matrix (K) represent direct interactions between principal neuronal groups, the eigenstates of which represent neural patterns of activation by which information or neural codes may be stored; (2) the friction matrix (G) represent inhibitory interactions that act as a brake on the activity of principal neurons; and (3) the random forces R represent energy influx that activates the neural patterns, and provides a balance to the dissipative forces of the friction matrix. One may hypothesize, for instance, that frictional forces are abnormally low just before onset of an epileptic seizure [18]. Thus it is important to have a more reliable estimate of the effective friction constant.

For dynamical systems that are intrinsically nonlinear, it may not always be possible to choose a short enough time window such that the dynamics becomes piece-wise linear. In this case, one may still be guided by the principle that the random force term is best defined as that which is not time-correlated with the experimental observables, and then the model parameters are again chosen variationally so as to minimize the mean square of the time correlation of the random force with the experimental observables; that is, one can variationally minimize the term in (B5) even when the equation of motion is nonlinear. The minimization process may become much more difficult for nonlinear model systems.

VI. CONCLUSION

Our simulations suggest that linear regression performed on time correlation functions of experimental variables (LRC) is a useful alternative to more traditional methods applying linear regression to the experimental variables themselves (LRX). In particular, we find that estimates of an effective friction constant from dynamical data are unreliable unless the experimental system is either perturbed away from equilibrium using either LRX or LRC, and/or if the estimate is derived from averaging the time correlation function of the dynamic variables across multiple time windows using LRC. If one employs an empirical dynamical model such as (2), as is common in the engineering literature, it is possible to be fooled into thinking one's estimates of model parameters are accurate,

when in fact the corresponding estimated friction constant has a relative error on the order of 100%.

We regard our results as exploratory. Future studies may address the best ways of perturbing the model system away from equilibrium so as to obtain more reliable estimates of the force constant and friction constant matrices. Studies of more complex systems [19-21] and incorporation of memory effects (the H -terms in (1)) are also needed. In preliminary studies, we found that introducing H -terms can corrupt estimates for the force constant matrix (K) and the friction constant matrix (G). For instance, trajectories generated by simple damped harmonic oscillators (without memory, namely, where all the H -terms are zero) can result in LRX and LRC estimates where the estimated H -terms were dominant rather than being zero, with corresponding degradation of the estimated K 's and G 's. Similar results were found in by Darve et al. [22]. L2 regularization [23] failed to remove this flaw. We suspect that further physically-based constraints need to be imposed on the H -terms, for instance, that higher order terms in the H series are damped in time by an exponential envelope.

Finally, in the context of machine learning and artificial intelligence approaches to the analysis of dynamical systems, the LRC approach can be regarded as complementary, in that time averaging of time correlation functions systematically improves signal-to-noise ratio, and the time-averaged time correlation functions themselves can be regarded as inputs to machine learning and artificial intelligence algorithms.

APPENDIX A

Here we derive a simple solution to the traditional linear regression problem (LRX). We start with (2):

$$\delta X(t + \delta t) = \sum_{m=0}^M A(m) \delta X(t - m \delta t) + R'(t). \quad (A1)$$

Here $[K, G, H(m)]$ are the $N \times N$ harmonic force constant matrix, zero-time friction and memory kernel, respectively; and $R(t)$ is the random force of the system. Together, $[K, G, H(m)]$ are considered to be the time-invariant properties of the system within a particular time window and are the properties that we wish to extract. For convenience, we define:

$$H_M(t) = [H(1), H(2), H(3), \dots, H(M)] \quad (A2)$$

$$B_M = [K, G, H_M] \quad (A3)$$

$$\psi_M(t) = \begin{bmatrix} X(t) \\ \dot{X}(t) \\ \dot{X}_M(t) \end{bmatrix} \quad (A4)$$

$$\dot{X}_M(t) = \begin{bmatrix} \dot{X}(t-1) \\ \vdots \\ \dot{X}(t-M) \end{bmatrix}. \quad (A5)$$

(A1) can then be rewritten:

$$R(t) = \ddot{X}(t) + B_M \psi_M \quad (A6)$$

The LRX error is the square of the random force $R(t)$, averaged over the time window:

$$E_{LRX} = \frac{1}{2} Tr\{\langle R(t)R(t)^T \rangle\} \quad (A7)$$

Here the superscript T signifies a matrix transpose, $\langle \dots \rangle$ signifies a time average over this time window, and Tr signifies a matrix trace. The optimal B_M that minimizes the total error E_{LRX} is then given by:

$$B_M = [K, G, H_M] = -p/q \quad (A8)$$

$$p = \langle \ddot{X}(t)\psi(t)^T \rangle \quad (A9)$$

$$q = \langle \psi(t)\psi(t)^T \rangle \quad (A10)$$

Equations (A8) through (A10) can be rewritten:

$$\langle [\ddot{X} + B_M \psi]\psi^T \rangle = 0 \quad (A11)$$

Equation (A11) is a statement of the orthogonality principle in linear estimation.

APPENDIX B

Here we derive a simple solution to the linear regression of the time correlation equation of motion. We start with (2):

$$\delta X(t + \delta t) = \sum_{m=0}^M A(m) \delta X(t - m \delta t) + R'(t). \quad (B1)$$

Here $[K, G, H(m)]$ are the $N \times N$ harmonic force constant matrix, zero-time friction and memory kernel, respectively; and $R(t)$ is the random force of the system. Together, $[K, G, H(m)]$ are time-invariant properties of the system within a particular time window and are the properties that we wish to extract.

Next define a time correlation function:

$$C(t) = \frac{2}{N_T} \sum_{i=1}^{N_T/2} \delta X(t+i) \delta X(i). \quad (B2)$$

Taking the second time derivative of (B2) and inserting (B1) yields:

$$\begin{aligned} \ddot{C}(t) = & -KC(t) - G\dot{C}(t) \\ & - \sum_{m=1}^M H(m) \dot{C}(t - m \delta t) + C_{RX}(t) \end{aligned} \quad (B3)$$

where $C_{RX}(t)$ is given by:

$$C_{RX}(t) = \frac{2}{N_T} \sum_{i=1}^{N_T/2} R(t+i) \delta X(i) \quad (B4)$$

Within Zwanzig-Mori theory, the correlation in phase space of the random force with the explicit degrees of freedom is zero [16-18]. If we additionally invoke the *ergodic hypothesis* and replace the phase space correlation with a time correlation, then

$C_{RX}(t)$ in (B4) should be zero. The error function to be minimized is then:

$$E_{LRC} = \frac{1}{2} Tr\{\langle C_{RX}(t)C_{RX}(t)^T \rangle\} \quad (B5)$$

Here the superscript T signifies a matrix transpose, $\langle \dots \rangle$ signifies a time average over this time window, and Tr signifies a matrix trace. For convenience, we define:

$$H_M(t) = [H(1), H(2), H(3), \dots, H(M)] \quad (B6)$$

$$B_M = [K, G, H_M] \quad (B7)$$

$$\Psi(t) = \begin{bmatrix} C(t) \\ \dot{C}(t) \\ \dot{C}_M(t) \end{bmatrix} \quad (B8)$$

$$\dot{C}_M(t) = \begin{bmatrix} \dot{C}(t-1) \\ \vdots \\ \dot{C}(t-M) \end{bmatrix} \quad (B9)$$

Equation (B3) can then be rewritten more simply as

$$C_{RX}(t) = \ddot{C}(t) + B_M \Psi_M \quad (B10)$$

The solution for B_M is then:

$$B_M = [K, G, H_M] = -P/Q \quad (B11)$$

where

$$P = \langle \ddot{C}(t)\Psi(t)^T \rangle \quad (B12)$$

$$Q = \langle \Psi(t)\Psi(t)^T \rangle. \quad (B13)$$

Equations (B11) through (B13) can be rewritten:

$$\langle [\ddot{C} + B_M \Psi] \Psi^T \rangle = 0 \quad (B14)$$

Equation (B14) is a statement of the orthogonality principle in linear estimation as applied to LRC.

REFERENCES

1. S. M. Stigler, "Gauss and the invention of least squares," *Ann. Stat.*, vol. 9, no. 3, pp. 465-474, 1981 doi: 10.1214/aos/1176345451
2. C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424-438, 1969, doi: 10.2307/1912791.
3. S. J. Kiebel, M. I. Garrido, R. Moran, C. C. Chen, K. J. Friston, "Dynamic causal modeling for EEG and MEG," *Hum. Brain Mapp.*, vol. 30, no. 6, pp. 1866-1876, Jun 2009, doi: 10.1002/hbm.20775.
4. F. H. Stillinger, T. A. Weber, "Inherent structure in water," *J. Phys. Chem.*, vol. 87, no. 15, pp. 2833-2840, 1983, doi: 10.1021/j100238a027.
5. R. M. Stratt, "The instantaneous normal modes of liquids," *Accounts Chem. Res.*, vol. 28, no. 5, pp. 201-207, May 1995, doi: 10.1021/ar00053a001.
6. T. Keyes, "Instantaneous normal mode approach to liquid state dynamics," *J. Phys. Chem. A*, vol. 101, no. 16, pp. 2921-2930, Apr 1997, doi: 10.1021/jp963706h
7. R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. Am. Soc. Mech. Eng.*, vol. 82, Series D, pp. 35-55, 1960, doi: 10.1115/1.3662552.
8. D. Hsu, M. Hsu, "Zwanzig-Mori projection operators and EEG dynamics: deriving a simple equation of motion," *PMC Biophysics*, vol. 2, article 6, July 2009, doi: 10.1186/1757-5036-2-6.
9. V. Salanova, "Deep brain stimulation for epilepsy," *Epilepsy Behav.* vol. 88S, pp. 21-24, Nov. 2018, doi: 10.1016/j.yebeh.2018.06.041.
10. E. B. Geller, "Responsive neurostimulation: Review of clinical trials and insights into focal epilepsy," *Epilepsy Behav.*, vol. 88S, pp. 11-20, Nov. 2018, doi: 10.1016/j.yebeh.2018.06.042.
11. R. Zwanzig, "Ensemble method in the theory of irreversibility," *J. Chem. Phys.*, vol. 33, no. 5, pp. 1338-1341, 1960, doi: 10.1063/1.1731409.
12. H. Mori, "Transport, collective motion, and Brownian motion," *Prog. Theor. Phys.*, vol. 33, no. 3, pp. 423-455, 1965, doi: 10.1143/PTP.33.423.
13. R. Zwanzig. "Chapter 8: Projection operators," in *Nonequilibrium Statistical Mechanics*, Oxford University Press, 2001.
14. R. Kubo, "The fluctuation-dissipation theorem," *Rep. Prog. Phys.*, vol. 29, pp. 255-284, 1966, doi: 10.1088/0034-4885/29/1/306.
15. C. C. Moore, "Ergodic theorem, ergodic theory, and statistical mechanics," *P. Natl. Acad. Sci. USA*, vol. 112, no. 7, pp. 1907-1911, Feb 2015, doi: 10.1073/pnas.1421798112.
16. S. Van Haerents, B. S. Chang, A. Rotenberg, A. Pascual-Leone, M. M. Shafí, "Noninvasive brain stimulation in epilepsy," *J. Clin. Neurophysiol.*, vol. 37, no. 2, pp. 118-130, Mar 2020, doi: 10.1097/WNP.0000000000000573.
17. P. Jayakar, "Cortical electrical stimulation mapping: special considerations in children," *J. Clin. Neurophysiol.*, vol. 35, no. 2, pp. 106-109, Mar 2018, doi: 10.1097/WNP.0000000000000451.
18. M. Hsu, D. Hsu, "A piece-wise harmonic Langevin model of EEG dynamics: theory and application to EEG seizure detection," *Neurocomputing*, vol. 65, pp. 469-474, Jun 2005, doi: 10.1016/j.neucom.2004.11.003.

> TNNLS-2021-P-19190

19. E.N. Lorenz, “Deterministic nonperiodic flow,” J. Atmos. Sci., vol. 20, no. 2, pp. 130-141, Mar 1963, doi: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.
20. M. Henon, “A two-dimensional mapping with a strange attractor,” Commun. Math. Phys., vol. 50, no. 1, pp 69-77, Feb 1976, doi: 10.1007/BF01608556.
21. Y. Pomeau, P. Manneville, “Intermittent transition to turbulence in dissipative dynamical systems,” Commun. Math. Phys., vol. 74, no. 2, pp. 189-197, Jun 1980, doi: 10.1007/BF01197757.
22. E. Darve, J. Solomon, A. Kia, “Computing generalized Langevin equations and generalized Fokker-Planck equations,” P. Natl. Acad. Sci. USA, vol. 106, no. 27, pp. 10884-10889, Jul 2009, doi: 10.1073/pnas.0902633106.
23. W. H. Press, S. A. Teukolsky, W. T. Vetterling, B. P. Flannery, “Section 19.5: Linear Regularization Method,” in *Numerical Recipes: The Art of Scientific Computing*, 3rd ed., New York, USA: Cambridge University Press, 2007.