LINEAR RESPONSE BASED PARAMETER ESTIMATION IN THE PRESENCE OF MODEL ERROR

A PREPRINT

He Zhang
Department of Mathematics
The Pennsylvania State University, University Park, PA 16802, USA.
hqz5159@psu.edu

John Harlim
Department of Mathematics, Department of Meteorology and Atmospheric Science, Institute for CyberScience
The Pennsylvania State University, University Park, PA 16802, USA.
jharlim@psu.edu

Xiantao Li
Department of Mathematics
The Pennsylvania State University, University Park, PA 16802, USA.
xxl12@psu.edu

November 1, 2019

ABSTRACT

Recently, we proposed a method to estimate parameters in stochastic dynamics models based on the linear response statistics. The method rests upon a nonlinear least-squares problem formulated in [Harlim et al. [1] and Zhang et al. [2]] that takes into account the response properties that stem from the Fluctuation-Dissipation Theory. In this article, we address an important issue that arises in the presence of model error. In particular, when the equilibrium density function is high dimensional and non-Gaussian, and in some cases, is unknown, the linear response statistics are inaccessible. We show that this issue can be resolved by fitting the imperfect model to appropriate marginal linear response statistics that can be approximated using the available data and parametric or nonparametric models. The effectiveness of the parameter estimation approach is demonstrated in the context of molecular dynamics models (Langevin dynamics), where the modeling error is due to coarse-graining, and a PDE (non-Langevin dynamics) that exhibits spatio-temporal chaos, where the model error arises from severe spectral truncations. In these examples, we show how the imperfect models, the Langevin equation with parameters estimated using the proposed scheme, can predict the nonlinear response statistics of the underlying dynamics under admissible external disturbances.

Keywords Parameter Estimation · Linear Response Theory · Missing Dynamics · Kernel Mean Embedding

1 Introduction

An important and routine task in scientific research is to determine parameters in a given mathematical model. Whether the model are built from direct empirical observations, or constructed from an underlying dynamic that is much more comprehensive, model error is inevitable in general. Even in the context of first-principle approaches, e.g., quantum-mechanical descriptions [3, 4], they are rarely implemented in its full form. Instead, various reductions are introduced so that the models fit into practical computations, e.g., tight-binding methods [5] and
We circumvent this difficulty by introducing the marginal linear response statistics that are practically computable. We have in mind a class of stochastic models, whose parameters can be classified as follows. There is a subspace of parameters directly responsible to the equilibrium statistics. This subspace of the parameters can be determined by many standard statistical methods. Once this subspace is determined, the full model parameters can be estimated by fitting the quotient parameter space to the non-equilibrium properties of the dynamics. In this paper, we consider fitting the dynamical behavior through an impulse/response approach by ‘poking’ the system with a small external force. This approach is known as the linear response in statistical physics as the first step to study non-equilibrium phenomena. A hallmark in linear response theory is the Fluctuation-Dissipation Theory (FDT). Basically, it states that for a dynamical system at equilibrium, the full response to small external perturbations can be characterized by a first-order linear response through only the information of appropriate time correlation functions of the unperturbed dynamical system. Motivated by the accessibility of FDT, we have developed a parameter estimation approach \[1\] as well as an efficient numerical scheme \[2\] under the perfect model assumption. By inferring the estimates from the linear response statistics, we are able to reproduce both the equilibrium statistics and the response properties under small perturbations.

This paper focuses primarily on the response-based method in the presence of model error. In general, the model error may be a result of an empirical assumption, truncation, coarse-graining, multiscale expansion, discounting the memory effect, etc. In practice, the highly desirable parameter estimation methods are those that are not aware of these effects. This is particularly relevant when the only available information is a data set of some observables of interest, which poses several new challenges for the parameter estimation methods. First, the external forcing, when applied to the underlying dynamics and then carried over to the imperfect model, may depend on variables other than the observed ones. We clarify this issue by introducing a concept called admissible external forcing. Another emerging issue is that the linear response involves a variable conjugate to the external force \[7\] (also interpreted as a flux variable \[8\]), which, in general, is only accessible when the statistics of the full model is known. We circumvent this difficulty by introducing the marginal linear response statistics that are practically computable given the available data and parametric/nonparametric modeling.

These ideas are best explained with concrete examples. Our first example is motivated by an important problem in computational chemistry, i.e., coarse-graining (CG) molecular dynamics (MD). The needs for CGMD is driven by the observation that full molecular models are limited by the small intrinsic time scale and it is difficult for direct simulations to reach the time scale of interest \[9\]-\[11\], where important biological processes occur \[12\]. CGMD circumvents this problem by projecting the dynamics to the CG variables, a much smaller set, often defined by the local averages of the atomic coordinates and momentum \[13\]-\[14\]-\[15\]. The interactions among the CG variables are represented via the free energy (a.k.a. the potential of mean force (PMF)), which in turn provides an explicit form of the probability density function. Statistical methods have played a central role in determining the parameters in the PMF \[16\]. Within our discussions, these parameters will be called equilibrium parameters.

Meanwhile, there are parameters in the CG dynamics that do not appear in the PMF; a well-known example of which is the damping coefficient in the Langevin equation, and more generally, the friction kernel in the generalized Langevin equations. In the latter case, parameters have to be determined by following the statistical properties of the trajectories of the CG variables. Various techniques have been considered, including the Bayesian approach \[17\], relative entropy \[18\], as well as the Padé-type of algorithms that match the time correlation properties \[19\]-\[20\]-\[21\]. The application of our response-based approach for CGMD models has been motivated by the steered MD simulations \[22\]-\[23\], where forces are applied to proteins to facilitate the folding and unfolding processes. With a one-dimensional chain model as a simple test problem, we demonstrate how the response-based approach is formulated and implemented to determine the damping coefficient, which is assumed to be a more general band matrix.

Another important class of problems is PDEs with solutions that exhibit chaotic behavior. One well known type of examples is turbulence. These dynamics are often projected to Fourier modes to study the energy transfer among different modes. From a reduced-modeling viewpoint, an outstanding challenge is the model closure that retains a finite number of modes. Part of the challenge stems from the fact that the equilibrium statistics of the full model is high-dimensional and unknown. We will consider the Kuramoto-Sivashinsky (KS) equation as such an example, and demonstrate how to circumvent this difficulty with a semi-parametric method using the kernel mean embedding (KME) method.

In principle, the linear response theory is a prediction tool by itself, and the imperfect models would seem to be of little value in this case. However, as we previously alluded to, the conjugate variables might not be computable. The semi-parametric approach addresses precisely this issue. In addition, we demonstrate with examples that our approach yields models that also predict the full response with good accuracy.
The paper is organized as follows: We start with a short review of the linear response theory and the parameter estimation approach based on linear response in Section 2. To motivate a general parameter estimation scheme in the presence of model error, a linear fast-slow model will be presented in Section 3 for which explicit calculations can be carried out. In Section 4 we elaborate the general difficulties and outline a parameter estimation scheme to address those issues. The numerical scheme will be followed by two examples: a 1-D chain model (Section 5) and the KS equation (Section 6). We will close by a short summary and discussions in Section 7.

2 The parameter estimation method via linear response statistics

The FDT is a mathematical framework for quantifying the linear response of a dynamical system subjected to small external forcing. The linear response statistics, determined based on two-point equilibrium statistics of the unperturbed dynamics, provide estimates for the non-equilibrium properties. In statistical mechanics literature, FDT is known as the linear response approach, which is the foundation for defining transport coefficients, e.g., viscosity, diffusion constant, heat conductivity etc.

We begin by reviewing the linear response theory and the concept of the response statistics, which is a core component of the response-based estimation method. This will be presented in the context of a stochastic dynamics, expressed in the general form of an n-dimensional (time-homogeneous) stochastic differential equations (SDEs), also known as the Itô diffusion. The SDEs, together with its perturbed dynamics, are written as follows,

\[ \dot{X} = b(X;\theta) + \sigma(X;\theta) \dot{W}, \]

\[ \dot{X}^\delta = [b(X^\delta;\theta) + c(X^\delta) \delta f(t)] + \sigma(X^\delta;\theta) \dot{U}, \]

respectively, where \( W_t \) and \( U_t \) are standard Wiener processes. In the unperturbed system, the vector field \( b(X;\theta) \) denotes the drift and \( \sigma(X;\theta) \) is the diffusion tensor; while in the perturbed system, an order-\( \delta \) \( (\delta \ll 1) \) external perturbation is introduced, in the form of \( f(x,t) = c(x)\delta f(t) \). In both equations, \( \theta \in D \) denotes the model parameters with parameter domain \( D \subset \mathbb{R}^N \).

Throughout this section, which concerns with the perfect model case, our goal is estimating the underlying true parameter value \( \theta^1 \) based on the time series of the unperturbed dynamics at the equilibrium. In the case of imperfect model, which is the central problem concerned in this paper, there are no true parameter values since the parameters in the imperfect model are not consistent with the parameters in the true model. Before we introduce the concept of the response statistics, we pose the following assumptions for all \( \theta \) in \( D \):

1. The system governed by Eq. 1 is ergodic with equilibrium density \( p_{eq}(x;\theta) \), that is, \( p_{eq}(x;\theta) \) is the unique steady-state solution of the corresponding Fokker-Planck equation. In particular, we denote \( p_{eq}^1(x) := p_{eq}(x;\theta^1) \).

2. The statistical properties associated with Eq. 2 can be characterized by a perturbed density \( p^\delta(x,t;\theta) \), which follows the corresponding Fokker-Planck equation under initial condition \( p^\delta(x,0;\theta) = p_{eq}(x;\theta) \).

There is an abuse of notation since for certain SDEs, \( p_{eq} \) can depend on a subspace of the parameter. We use the notation \( \mathbb{E}_{p_{eq}(\theta)}[\cdot] \) to emphasize that the corresponding statistics are computed from the sample generated by the model at parameter value \( \theta \) at equilibrium.

For any integrable observable \( A(\cdot) \), one can define the difference

\[ \Delta \mathbb{E}[A](t) := \mathbb{E}_{p^\delta} [A(X^\delta)](t) - \mathbb{E}_{p_{eq}} [A(X)] \]

as the full response statistics. The linear response theory allows us to estimate the order-\( \delta \) term of \( \Delta \) by a convolution integral, that is,

\[ \Delta \mathbb{E}[A](t) = \int_0^t k_A(t-s) \delta f(s) \, ds + O(\delta^2). \]

The FDT formulates the linear response operator, \( k_A(t) \) in 14, as the following two-point statistics:

\[ k_A(t;\theta) := \mathbb{E}_{p_{eq}(\theta)} [A(X(t)) \otimes B(X(0);\theta)], \quad B_i(X;\theta) := - \frac{\partial X_i}{\partial \theta} [c_i(X)p_{eq}(X;\theta)] {p_{eq}(X;\theta)} \]

where \( B_i \) and \( c_i \) denote the \( i \)th components of \( B \) and \( c \), respectively. The variable \( B \) will be called the conjugate variable to \( A \).
The significance of FDT is that the response operator is defined without involving the perturbed density \( p^\theta(x, t; \theta) \). Rather, it can be evaluated at equilibrium of the unperturbed dynamics. To turn this into a practical tool, we first notice that for a given \( t \geq 0 \), the value of \( k_A(t; \theta) \) can be computed using a Monte-Carlo sum based on the time series of the unperturbed system (1) at \( p_{eq}^\theta \). For example, let \( \{X_i = X(t_i)\}_{i=1}^M \) be the time series generated at \( p_{eq}^\theta \) with step length \( \Delta t = t_{i+1} - t_i \), then for \( t = k\Delta t \), the Monte-Carlo approximation can be written as
\[
k_A(t; \theta) \approx \frac{1}{M-k} \sum_{i=1}^{M-k} A(X_{i+k}) \otimes B_i(X_i).
\]

In practice, the computation of (6) can be done more efficiently using the block averaging algorithm [26]. Meanwhile, the same response statistics can be computed for any \( \Delta \) value. In particular, essential statistics named as
\[
\text{essential statistics for } k_A(t; \theta).
\]

In practice, the computation of (6) can be done more efficiently using the block averaging algorithm [26]. Meanwhile, the same response statistics can be computed for any \( \Delta \) value. In particular, essential statistics named as
\[
\text{essential statistics for } k_A(t; \theta).
\]

Motivated by the accessibility of \( k_A(t; \theta) \) as well as the finite-dimensional Padé approximation introduced in [19], we have proposed to infer the true parameter value \( \theta^\dagger \) from a discretization of \( k_A(t; \theta) \), \( \{k_A^i(t) := k_A(t; \theta^i)\}_{i=1}^K \), named as essential statistics [12]. The estimate \( \hat{\theta} \) solves the following nonlinear least-squares problem:
\[
\hat{\theta} := \arg \min_{\theta \in \mathcal{D}} \sum_{i=1}^{K} f_i^2(\theta), \quad f_i(\theta) := k_A^i(t_i) - \hat{k}_A(t_i; \theta), \quad i = 1, 2, \ldots, K.
\]

To ensure the solvability of (7), we assume that the total number of involved essential statistics is always strictly greater than the dimension of \( \theta \), that is, \( K > N \).

**Remark 2.1.** The accessibility of \( k_A(t; \theta) \) also allows us to use its time derivatives as the essential statistics to infer \( \theta^\dagger \). For example, under specific external forcing and observable \( A(\cdot) \), the first-order time derivative of \( k_A \) at \( t = 0 \) of a Langevin dynamics model provides direct estimates for the damping coefficients [11]. We will return to this in Section 5. But it is worthwhile to mention that estimating higher order derivatives of \( k_A(t; \theta) \) requires time series with sufficient accuracy, which is not necessarily available, especially when the data are from experimental observations.

**Remark 2.2.** It is important to point out that, in general, the quantity \( B(X; \theta) \) in (5) depends on the unknown parameter \( \theta \), which provides difficulties in computing \( k_A(t; \theta^\dagger) \) given only a time series of \( X \) at \( p_{eq}^\theta \). Qi and Majda addressed this issue by computing the ‘kicked response’ based on direct simulation of the underlying dynamics [27, 28, 29]. However, since the underlying dynamics is unknown in our scenario, we will consider both parametric and nonparametric estimators for a marginal invariant density \( p^\delta \) defined on the resolved (and identifiable unresolved) variables, in place of the high-dimensional unknown equilibrium density \( p_{eq}^\theta \), in Sections 5 and 6, respectively.

One practical challenge in solving (7) is that, except for very special cases, there is no explicit expressions for \( \hat{k}_A(t; \theta) \). Therefore an iterative method, e.g., the Gauss-Newton method, would necessarily require solving the model (1) repeatedly in a sequential manner, which can be rather computationally demanding. This issue has been mitigated by a polynomial based surrogate model [2], where the cost functions \( f_i(\theta) \) in (7) are approximated by linear combinations of orthogonal polynomials, denoted as \( f_i^{M}(\theta) \), based on the precomputed values \( \{f_i(\theta_j)\} \) over a set of grid points \( \Theta \subset D \), and such pre-computation can be proceed in parallel. Here, \( M \) stands for the order of the polynomials involved. In the subsequently approximations, we will replace \( f_i(\theta) \) in (7) by \( f_i^{M}(\theta) \), and formulate a new least-squares problem for the order-\( M \) polynomial based surrogate model.

So far, we have reviewed a parameter estimation approach [11, 22] for Itô diffusions [11] as well as an efficient numerical scheme under the perfect model assumption. By inferring from the linear response statistics, our estimates are able to reproduce both the equilibrium statistics and the response under certain perturbation. A much more difficult problem, however, is when the available model is a truncated dynamics or only partially known. In this case, several interesting issues will emerge, which will be investigated in the remaining part of the paper. In the next section, we start with a simple two-scale system to demonstrate the importance of the response statistics in determining the model parameters.

### 3 A linear fast-slow model for insight

Consider the following 2-D linear fast-slow dynamics as the underlying model
\[
\dot{x} = (a_{11}x + a_{12}y) + \sigma_x W_x
\]
\[
\dot{y} = \frac{1}{\epsilon}(a_{21}x + a_{22}y) + \frac{\sigma_y}{\sqrt{\epsilon}} W_y
\]
where $x$ and $y$ are the slow and the fast variables, respectively. Here, $W_x$ and $W_y$ are two independent Wiener processes and $\epsilon \ll 1$ characterizes the time-scale gap. Such multi-scale Ornstein-Uhlenbeck process has been used as a linear test model in a variety of situations [30]. For the sake of non-degeneracy and ergodicity, we assume that $\sigma_x, \sigma_y > 0$ and

$$a_{11}a_{22} - a_{12}a_{21} > 0, \quad a_{11} + \frac{a_{22}}{\epsilon} < 0, \quad a_{22} < 0,$$

which are enough to ensure the drift coefficients

$$A_\epsilon := \begin{pmatrix} a_{11} & a_{12} \\ a_{21}/\epsilon & a_{22}/\epsilon \end{pmatrix}$$

to be negative definite as well as the existence of the averaged (effective) dynamics [31]

$$\dot{X} = \tilde{a}X + \sigma_3 \dot{W}_t, \quad \tilde{a} := \frac{a_{11}}{a_{22}} - \frac{a_{12}a_{21}}{a_{22}} < 0.$$  \hfill (10)

We will keep the form of (10) by posing

$$\dot{X} = bX + \sigma_3 \dot{W}_t, \quad b < 0, \quad \sigma > 0,$$  \hfill (11)

as our imperfect model of the fast-slow dynamics with $\theta = (b, \sigma)$ as the unknown parameters. Applying the parameter estimation scheme reviewed in Section 2 to $\theta$ by taking observable $A(X) = X$ and a constant external forcing $f = \delta$, we find that (see the proof of Theorem 1 in our previous work [1] for computational details)

$$\hat{k}_A(t; \theta) = \mathbb{E}_{\mu_{eq}}[X(t)X(0)] S^\top = e^{tb} S,$$

where $S = -\sigma^2/2b$ is the equilibrium variance of $X$. In contrast, $S^\top$ is the variance that corresponds to the ‘true parameter’ value $\theta^\top$, when the underlying dynamics is exactly of the form (11) with $\theta = \theta^\top$. Fitting the essential statistics ($k_A^*(t) = \hat{k}_A(t; \theta)$) leads to

$$\mathbb{E}_{\mu_{eq}}[X(t)X(0)] = e^{tb} S = e^{tb} S^\top.$$  \hfill (12)

In sharp contrast to the perfect model situation, in our case, the partial observations are from an underlying 2-D dynamics [3-9] while the essential statistics in (12) are derived based on an 1-D imperfect model (11). Thus, the left-hand side of (12), legitimately, is not available, and will be replaced by the two-point statistics of the slow variable $x$ in (6). Eq. (12) now becomes,

$$\mathbb{E}_{\mu_{eq}}^\top [x(t)\varsigma(0)] = [1, 0] e^{t \lambda_\Sigma} \varsigma[1, 0]^\top \approx e^{tb} S = -e^{tb} \frac{\sigma^2}{2b},$$  \hfill (13)

where the covariance matrix $\Sigma = (\Sigma_{ij})$ is determined by the stationary Lyapunov equation [25]

$$A_\epsilon \Sigma + \Sigma A_\epsilon^\top = -Q, \quad Q = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2/\epsilon \end{pmatrix}.$$  \hfill (14)

One can interpret Eq. (13) as the consistency between the two time auto-covariance functions of the slow variable $x$ and the variable $X$ in the imperfect model. Obviously, seeking an exact fit for (13) is impossible since its left-hand side has a total of 6 degrees of freedom ($A_\epsilon$ and $\Sigma$), which is much greater than that of the right-hand side. However, taking advantage of the multi-scale structure of the underlying dynamics, we can still find estimates for $\theta$ when $\epsilon \ll 1$ such that (13) is fulfilled up to an error of certain order. To see this, we first simplify the matrix exponential in (13) by computing the eigenvalues of $A_\epsilon$, \begin{align*}
\lambda_{1,2} = \frac{1}{2} \left( a_{11} + \frac{a_{22}}{\epsilon} \pm \sqrt{\Delta} \right), \quad \Delta = \left( a_{11} + \frac{a_{22}}{\epsilon} \right)^2 - \frac{4a_{12}a_{21}}{\epsilon},
\end{align*}
where the $\tilde{a}$ is the same as in Eq. (10). Assuming that the two eigenvalues are real, via a direct asymptotic expansions with respect to $\epsilon$, one can show that

$$
\lambda_1 = \tilde{a} - \frac{a_{12}a_{21}}{a_{22}} \frac{\tilde{a}}{\epsilon} + O(\epsilon^2), \quad \lambda_2 = \frac{a_{22}}{\epsilon} + \frac{a_{12}a_{21}}{a_{22}} + \frac{a_{12}a_{21} \tilde{a}}{a_{22}^2} \epsilon + O(\epsilon^2).
$$

Notice that $a_{22} < 0$ due to our assumption, and we are able to truncate the term containing $e^{t\lambda_2}$ as long as $t$ is large enough. Thus, up to a negligible error (13) becomes

$$e^{t\lambda_1} \Sigma_{11}(\lambda_2 - a_{11}) - \Sigma_{21} a_{12} \overline{\lambda_2 - \lambda_1} = -e^{tb} \frac{\sigma^2}{2b},$$

5
which can be achieved by taking
\[
\begin{align*}
\dot{b} &= \lambda_1 = \tilde{a} - \frac{a_{12}a_{21}}{a_{22}^2} \tilde{a} + O(\epsilon^2) \\
\dot{\sigma}^2 &= -2\lambda_1 \frac{\Sigma_{11}(\lambda_2 - a_{11}) - \Sigma_{21}a_{12}}{a_{22}^2} = \sigma_x^2 - \frac{2\sigma_x^2 a_{12}a_{21} - \sigma_y^2 a_{22}^2}{a_{22}^2} \epsilon + O(\epsilon^2),
\end{align*}
\]
(15)
as our estimates. The same estimates have been obtained based on the Riccati equation [32].

Compared with the averaged dynamics (10), our estimates (15), based on matching the linear response statistics, exhaust the capability of the imperfect model (11) in the sense that the two-point statistics
\[ E_{p_{eq}}[x(t)x(0)] \] (\( t \gg \epsilon \)) of the slow variable \( x \) can be reproduced up to any order of \( \epsilon \) as \( \epsilon \to 0^+ \). As a test, we set \( a_{11} = a_{21} = a_{22} = -1, a_{12} = 1, \) and \( \sigma_x^2 = \sigma_y^2 = 2 \) in (8) and (9). Figure 1 shows the absolute error in reproducing the two-point statistics under two different choices of \( \epsilon \), in comparison with various estimation methods of the imperfect model. For small \( \epsilon \), the estimate in (15) (which we called order-\( \infty \)) produces the optimal result. However, when \( \epsilon \) is relatively large, we are no longer allowed to truncate the term \( e^{t\lambda_2} \) in the simplification of (13), which in turn invalidates the estimate in (15) as shown in the second panel of Figure 1. However, the solution of (13) in the least-squares sense is accurate and this justifies the importance of solving the nonlinear least-squares problem in (7) for parameter estimation in the presence of model error.

![Figure 1: The absolute error in reproducing \( E_{p_{eq}}[x(t)x(0)] \) under two different scale gaps \( \epsilon = 0.09 \) (top) and \( \epsilon = 0.15 \) (bottom). In terms of the estimates, order-0: \( \dot{b} = \tilde{a} \) and \( \dot{\sigma}^2 = \sigma_x^2 \); order-1: \( \dot{b} = \tilde{a} - \epsilon a_{12}a_{21} \tilde{a}/a_{22}^2 \) and \( \dot{\sigma}^2 = \sigma_x^2 - \epsilon(2\sigma_x^2 a_{12}a_{21} - \sigma_y^2 a_{22}^2)/a_{22}^2 \); order-\( \infty \): \( \dot{b} \) and \( \dot{\sigma}^2 \) are given by (15) without truncation, and LS: least-squares estimates.](image)

To summarize, we have elucidated the importance of the proposed parameter estimation scheme in the presence of model error on an idealistic problem. In particular, we have showed that the linear response-based parameter estimation scheme produces accurate linear mean response statistics, outperforming the conventional method, the averaging theory [31], even in a regime when such theory is valid.

### 4 Parameter estimation in the presence of model error

Inspired by the result in Section 3, we consider extending the response-based parameter estimation scheme in the presence of model error arises from coarse-graining or missing dynamics. In current section, we provide a formal discussion on the key issues in such a difficult scenario and propose a strategy to address them.
To elaborate the general ideas, let us use the abstract notations
\[
\dot{u} = F(u), \quad X = P(u),
\]
to denote the underlying full dynamics with partial observation of the variable of interest, \(X\). Here, \(P\) denotes a projection operator to the observables that is not necessarily known. We assume that \(F\) is unknown, instead, the available imperfect model is given as follows,
\[
X = R_1(X; Y; \theta), \quad \dot{Y} = R_2(X; Y; \theta).
\]
Here, we have included an additional variable \(Y\) as possible auxiliary variables and \(\theta \in \mathcal{D}\) are the unknown model parameters. Our goal is to estimate the parameter \(\theta\) in (17) using the time series of \(X\) observed from (16) such that the resulting model in (17) can predict the full nonlinear response of the underlying dynamics (16) under admissible external forcing, as explained below. While the choice of model in (17) is central to the accuracy of the prediction, our focus is not on the specification of this model. Instead, we assume that the model in (17) is given and our focus, again, is to generalize the proposed parameter estimation scheme reviewed in Section 2 on this scenario.

In this case, the linear response operator (5) associated with the underlying dynamics (16) is no longer accessible since, in general, it is a two-point statistic that can be computed only if the full data set of \(u\) is available. Thus, to infer the parameter \(\theta\) in (17) from the linear response statistics of the underlying dynamics (16), we need to answer the following two questions:

1. How to find an alternative for the essential statistics \(k^1_X(t_i)\) in (7) which is computable given only the time series of \(X\)?
2. How to formulate a computationally feasible nonlinear least-squares problem analogous to (7) for the imperfect model (17)?

To address these issues, we introduce the concepts of (i) admissible external forcing, and (ii) marginal linear response (MLR). Using the notations in (16), similar to (1) and (2), the perturbed underlying dynamics can be written in the form
\[
\dot{u}^\delta = F(u^\delta) + c(u^\delta)\delta f(t), \quad X^\delta = P(u^\delta).
\]
Taking the time derivative of \(X\) in (16) and \(X^\delta\) in (18), we have
\[
\dot{X} = \nabla P(u)F(u), \quad \dot{X}^\delta = \nabla P(u^\delta)F(u^\delta) + \delta f(t)\nabla P(u^\delta)c(u^\delta),
\]
where \(\nabla P\) denotes the Jacobian matrix of \(P\). Thus, the term \(\delta f(t)\nabla P(u^\delta)c(u^\delta)\) can be interpreted as the external force on the dynamics of \(X\). In the case when the full dynamics (16) is a system of SDEs with an external force in the drift term, a similar observation can be made by using the Itô’s formula. To this end, we define:

**Definition 4.1.** An external forcing \(f(u, t) = c(u)\delta f(t)\) to the full dynamics in (16) is admissible with respect to the dynamics (16) if \(\nabla P(u)c(u)\) can be written as a function of \(X\) denoted as \(c_R(X)\).

Basically, an admissible external forcing \(c(u^\delta)\delta f(t)\) of the underlying dynamics (16) introduces a perturbation \(c_R(X^\delta)\delta f(t)\) to the imperfect model (17),
\[
\dot{X}^\delta = R_1(X^\delta, Y; \theta) + c_R(X^\delta)\delta f(t), \quad \dot{Y}^\delta = R_2(X^\delta, Y; \theta).
\]
Intuitively, this condition rules out other external forces that depend on \(u\), which is not accessible in our scenario. To simplify the discussion, let us define \(\theta_{eq} := \theta_{eq}(\theta)\) as the components of the parameters that directly appear in the equilibrium density \(\rho_{eq}(X, Y; \theta_{eq})\) of the imperfect model (17); they usually can be estimated directly by matching equilibrium statistics. For example, in the linear problem in (11), while the model parameters \(\theta\) consist of the damping coefficient \(b\) and noise amplitude \(\sigma\), the parameter that appears in the equilibrium density, \(\theta_{eq}\), is \(S = -\sigma^2/2b\), which can be directly obtained by matching the equilibrium variance.

Once \(\theta_{eq}\) has been estimated (by \(\hat{\theta}_{eq}\)), given an observable \(A(X)\) the linear response operator (5) of the perturbed imperfect dynamics (19) is given as,
\[
\dot{k}_A(t; \theta) = \mathbb{E}_{\rho_{eq}(\theta)}[A(X(t)) \otimes B\{X(0), Y(0); \hat{\theta}_{eq}\}], \quad B_1(X, Y; \hat{\theta}_{eq}) = -\frac{\partial X}{\rho_{eq}(X, Y; \theta_{eq})}, \quad B_2(X, Y; \hat{\theta}_{eq}) = \frac{\partial X}{\rho_{eq}(X, Y; \theta_{eq})},
\]
where the expectation over \(\rho_{eq}(\theta)\) will be realized through Monte-Carlo averages with respect to the solutions of the unperturbed imperfect model in (17) for parameter \(\theta\) that satisfies the constrained of \(\theta_{eq}(\theta) = \hat{\theta}_{eq}\).

Given an observable \(A(X)\) and an admissible external forcing \(f(u, t)\) in Definition 4.1, in comparison with the linear response operator (5) and linear response statistics (4), we define:
Definition 4.2. Let $p_R(X) := \int \delta(X - P(u)) p_{eq}^\perp(u) \, du$ be the marginal invariant density of $X$, which can be estimated from the given data of $X$ at the equilibrium. We define the MLR operator,

$$\hat{k}_A^\perp(t) := \mathbb{E}_{p_{eq}} \left[ A(X(t)) \otimes B_R^\perp(X(0)) \right], \quad (B_R^\perp)_i := -\frac{\partial_{X_i} \left[ (c_R(X)_i p_R(X) \right]}{p_R(X)}.$$ (21)

We call $\left\{ \hat{k}_A^\perp(t) \right\}_{i=1}^K$ the marginal essential statistics. Furthermore, we define,

$$\hat{\Delta}E[A](t) = \int_0^t \hat{k}_A^\perp(t - s) \delta f(s) \, ds$$ (22)

as the MLR statistics.

The two-point statistics in (21), defined with respect to the equilibrium distribution of the underlying dynamics (16), can be approximated via Monte-Carlo, like (6), based on the available time series of $X$. We require the external forcing to be admissible; otherwise, $c_R$ in (21) would depend on $u$ and the marginal essential statistics are not accessible when the data of $u$ is not available. In practice, as we will show in the next two sections, the marginal density $p_R$ can be learned from the available time series of $X$ by using standard statistical methods. This way, the two-point statistics in (21) is computable as opposed to (6).

To infer the parameter $\theta$ in (17) from (21), we formulate the following new nonlinear least-squares problem

$$\hat{\theta} := \arg \min_{\theta \in \mathcal{D}: \theta_{eq}(\theta) = \hat{\theta}_{eq}} \sum_{i=1}^K f_i^2(\theta), \quad f_i(\theta) := \hat{k}_A^\perp(t_i) - \hat{k}_A^\perp(t_i; \theta), \quad i = 1, 2, \ldots, K,$$ (23)

where $\hat{k}_A^\perp(t_i)$ is the marginal essential statistics (21) and $\hat{k}_A^\perp(t_i; \theta)$ is the linear response operator (20) of the imperfect model (17). This is a dynamic constrained nonlinear least-squares problem through the imperfect model in (17). We should point out the minimization is over a subspace of the parameter set, characterized by the constraint, $\theta_{eq}(\theta) = \hat{\theta}_{eq}$, where $\hat{\theta}_{eq}$ is the estimate for the equilibrium parameters, $\theta_{eq}$, obtained from direct matching to the available equilibrium statistics.

We validate the estimate $\hat{\theta}$ by comparing the nonlinear response of the imperfect model at the estimated parameters with that of the underlying model under admissible external forcing. These statistics are computed via Monte-Carlo averages over solutions of the perturbed underlying dynamics (16) and perturbed imperfect model (19). We should point out the minimization is over a subspace of the parameter set, characterized by the constraint, $\theta_{eq}(\theta) = \hat{\theta}_{eq}$, where $\hat{\theta}_{eq}$ is the estimate for the equilibrium parameters, $\theta_{eq}$, obtained from direct matching to the available equilibrium statistics.

Remark 4.3. In the fast-slow model discussed in the previous section, the underlying dynamics are governed by 8-9, while the imperfect model is given by 11. The constant external forcing we proposed in the previous analysis is admissible since it acts only on the slow variable $x$.

Remark 4.4. While the choice of imperfect model in (17) depends on the specific applications, in Sections 6 and 8 we will consider Langevin dynamics, regardless of whether the underlying dynamics are Langevin or not. In the molecular dynamics example (Section 5), the imperfect Langevin model, as a result of coarse-graining, is a closed equation of $X$, with no dependent on any auxiliary variables $Y$. In this case, the associated equilibrium density has a natural parametric form, $p_{eq}(X; \theta_{eq})$, given by the Gibbs measure. Thus, we simply use the direct estimates together with maximum entropy principle to estimate $\theta_{eq}$.

In the PDE example (Section 6), choosing Langevin dynamic as the imperfect model is motivated largely by the separable structure of its equilibrium distribution, $p_{eq}(X, Y; \theta_{eq}) = \hat{p}(X)p_{G}(Y; \theta_{eq})$, where $\hat{p}(X)$ denotes the (possibly) non-Gaussian density of the variable of interest, while $p_{G}(Y; \theta_{eq})$ denotes the Gaussian auxiliary components. In this case, $\theta_{eq}$ can be estimated directly from samples of $X$ and we will use a nonparametric kernel mean embedding (KME) method to recover $\hat{p}$ from the data of $X$.

Remark 4.5. For the molecular dynamics example (Section 5), we consider $p_R(X) = p_{eq}(X; \theta_{eq})$ since $p_{eq}$ is the natural equilibrium density obtained through coarse-graining. We should point out that choosing $p_R$ to depend only on $X$ in (21) is only sensible (in the linear response theory) if $X$ is independent of the orthogonal component, $X^\perp := u - P(u)$, at equilibrium state, in that $p_{eq}(X, X^\perp) = p_R(X)p_{R}^\perp(X^\perp)$. In this case, we have

$$\frac{\partial_{X_i} \left[ c_i(X) p_{eq}(X, X^\perp) \right]}{p_{eq}(X, X^\perp)} = \frac{\partial_{X_i} \left[ c_i(X) p_{R}(X)p_{R}^\perp(X^\perp) \right]}{p_R(X)p_{R}^\perp(X^\perp)} = \frac{\partial_{X_i} \left[ c_i(X) p_{R}(X) \right]}{p_R(X)},$$

which suggests that the MLR operator is identical to the linear response operator of the underlying dynamics, $\hat{k}_A^\perp(t) = \hat{k}_A^\perp(t)$. In general, however, the distributions of $X$ and $X^\perp$ are not independent and fitting to the marginal
statistics defined through $p_R(X)$ may not be the optimal strategy. In the PDE example (Section 5), the variable of interest is the first Fourier mode, $X = u_1$, obtained from a discrete Fourier transform of the PDE solution. In addition to $u_1$, we are allowed to have access to the time series of the second Fourier mode $u_2$, as well as the dynamical equation of the only variable $u_1$. Given these information, we can extract the identifiable irrelevant component, $g$, by regressing $u_1, u_2$ to the dynamical equation for $u_1$. In this case, we will employ the nonparametric KME formulation to estimate $p_R(u_1, u_2, g)$. We will show that the resulting MLR statistics, computed via the formula in [22], which carried some information of $X^2$ through $u_2$ and $g$, is an accurate estimator of the full nonlinear response statistics of the underlying dynamics under admissible external disturbances up to finite time.

5 A one-dimensional molecular model

We consider a molecular dynamics model consisting of a chain of atoms with mass $m$. Let the equilibrium spacing between the particles be $a_0$, and the displacement of the $i^{th}$ particles from its equilibrium position $R_i = i a_0$ be $r_i$. We take the Lennard-Jones (LJ) potential [33], expressed as

$$U_0 = \sum_{i=1}^{N} \sum_{j=i-2}^{i-1} \psi(r_i - r_j + (i-j)a_0), \quad \psi(r) = 4\left(|r|^{-12} - |r|^{-6}\right).$$

(24)

5.1 The underlying dynamics

Our underlying dynamics, as a finite-dimensional approximation of the LJ lattice model, contains a total of $N$ particles equipped with periodic boundary conditions. Let $r = (r_1, r_2, \ldots, r_N)^\top$ and $v = (v_1, v_2, \ldots, v_N)^\top$ be the displacement and the velocities of the particles, respectively. The potential energy of the finite system, as a function of the relative displacement $d := (r_2 - r_1, r_3 - r_2, \ldots, r_N - r_{N-1})^\top \in \mathbb{R}^{N-1}$, can be written as,

$$U(d) = \sum_{i=1}^{N+2} \sum_{j=i-2}^{i-1} \psi(r_j - r_j + (i-j)a_0).$$

(25)

Driven by the potential energy $U(d)$, the frictional, and random forcing, $(r, v)$ obeys the Langevin equation of motion,

$$\begin{aligned}
\dot{r} &= v, \\
\dot{v} &= -V_r U(d) - γv + \sqrt{2k_B T}γ\dot{W}_t,
\end{aligned}$$

(26)

where the mass has been set to unity ($m = 1$), and $W_t$ is an N-dimensional Wiener process; $γ$ and $k_B T$ denote the friction constant and the temperature, respectively. The Langevin dynamics is a means to model canonical ensemble [34]. Also known as the Gibbs measure of [25], the probability distribution of the relative displacement $d$ and the velocity $v$, is of the form

$$p_{eq}(d, v) \propto \exp \left[-\frac{1}{k_B T} \left(U(d) + \frac{1}{2}v^2\right)\right],$$

(27)

under the constraint induced by the periodic boundary condition.

In the numerical simulation, we set $a_0 = 1.1196, k_B T = 0.25, m = 1.0, γ = 0.5$ and $N = 100$ in [25]. A Verlet-type of integration algorithm [35] is applied with sub-sampled 1/10 at the equilibrium state with sample size $5 \times 10^5$ and step length $\Delta t = 0.05$. The temperature $k_B T$ is set to be high enough, so that at the equilibrium, the marginal density of $d$ is non-Gaussian. Throughout the section, the underlying dynamics [26] will only be used for generating CG observations and computing the full response for verification of the parameter estimates. We do not assume the explicit formula of the Gibbs measure $p_{eq}$ [27] is available.

5.2 The coarse-grained model

We first define CG variables by dividing the system equally to consecutive blocks, each of which contains $J$ atoms. The displacement and velocities of the CG particles can be defined as

$$q_i := \frac{1}{J} \sum_{j=(i-1)J+1}^{iJ} r_j, \quad p_i := \frac{1}{J} \sum_{j=(i-1)J+1}^{iJ} v_j, \quad i = 1, 2, \ldots, n_b, \quad n_b := \frac{N}{J},$$

(28)

respectively. Here, $n_b$ denotes the number of CG particles. Formally, the CG Langevin equation of the underlying dynamics [26] can be written as

$$\begin{aligned}
\dot{q}_i &= \dot{p}_i, \\
\dot{p}_i &= -\nabla_q V(d_{CG}; \mathbf{a}) - \Gamma \dot{p}_i + \sqrt{2b}^\top \sigma \dot{W}_t,
\end{aligned}$$

(29)

respectively.
where \( U_t \), to distinguish from \( W_t \) in [29], is an \( n_b \)-dimensional Weiner process, the PMF \( V(\mathbf{d}_{CG}; \mathbf{a}) \), as a function of the CG relatively displacement \( \mathbf{d}_{CG} := (q_2 - q_1, q_3 - q_2, \ldots, q_{n_b} - q_{n_b-1}) \), is of the form

\[
V(\mathbf{d}_{CG}; \mathbf{a}) = \sum_{i=1}^{4} \sum_{j=1}^{4} \alpha_j (q_{i+1} - q_i)^j, \quad q_{k+n_b} = q_k, \quad \alpha_i > 0
\]  

(30)

with unknown parameters \( \mathbf{a} = (a_1, a_2, a_3, a_4) \). In other words, the CG particles are assumed to follow a nearest-neighbor interaction driven by a quartic potential.

On the other hand, rather than assuming a scalar damping coefficient like the underlying dynamics (26), we propose a symmetric positive definite (SPD) damping matrix \( \Gamma \) in the reduced Langevin equation (29), while \( \sigma \) in the diffusion term of (29) satisfies \( \sigma \sigma^T = \Gamma \) due to the fluctuation-dissipation theorem [21]. In practice, such \( \sigma \) can be determined by a Cholesky decomposition of \( \Gamma \). Motivated by the symmetry of the system, the damping matrix \( \Gamma \) is assumed to have a symmetric circulant structure:

\[
\Gamma = \begin{pmatrix}
\gamma_0 & \gamma_1 & \gamma_2 & \cdots & \gamma_2 \\
\gamma_1 & \gamma_0 & \gamma_2 & \cdots & \gamma_2 \\
\gamma_2 & \gamma_1 & \gamma_0 & \cdots & \gamma_2 \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\gamma_2 & \gamma_1 & \gamma_2 & \cdots & \gamma_0
\end{pmatrix} \in \mathbb{R}^{n_b \times n_b},
\]  

(31)

which can be fully determined by the first \( \lfloor \frac{n_b}{2} \rfloor + 1 \) elements \( (\gamma_0, \cdots, \gamma_\lfloor \frac{n_b}{2} \rfloor) \) in the first row. To guarantee that \( \Gamma \) is positive definite, we pose a diagonal dominant constraint:

\[
\gamma_0 > \begin{cases}
2 \sum_{i=1}^{n_b} |\gamma_i|, & n_b \text{ even}, \\
2 \sum_{i=1}^{n_b-1} |\gamma_i|, & n_b \text{ odd}.
\end{cases}
\]

(32)

For simplicity, we will set \( \gamma_3 = \gamma_4 = \cdots = \gamma_{\lfloor \frac{n_b}{2} \rfloor} = 0 \) in (31), and the parameter set of the imperfect model (29) can be summarized as

\[
\theta := (\beta^{-1}, a_1, a_2, a_3, a_4, \gamma_0, \gamma_1, \gamma_2), \quad a_i, \beta > 0, \quad \gamma_0 > 2(|\gamma_1| + |\gamma_2|),
\]  

(32)

where \( a_i \) are the coefficients in the potential (30), \( \gamma_i \) are the component of the damping matrix (31) and \( \beta^{-1} \) is the temperature.

There are many other approaches in the model reduction of Langevin dynamics, e.g. the partitioning technique [36], the Petrov-Galerkin projection [20], and the maximum likelihood estimator [17]. Here, we consider the problem in the presence of model error, and use the one-dimensional system to demonstrate the parameter estimation procedure.

5.3 Parameter estimation and numerical results

So far, we have introduced our underlying dynamics (26) and the corresponding CG model (29) with unknown parameters \( \theta \) (32). In the numerical experiment, the observation of the CG model (29) will be acquired from Eq. (28) based on the time series of \( (\mathbf{d}, \mathbf{v}) \) generated by the underlying dynamics. For instance, the CG relatively displacement \( \mathbf{d}_{CG} \) can be determined by

\[
q_{i+1} - q_i = \frac{1}{I} \sum_{j=(i-1)J+1}^{ij} (r_{i+j}-r_j) = \frac{1}{I} \sum_{j=(i-1)J+1}^{ij} \sum_{k=j}^{j+J-1} (r_{k+1}-r_k), \quad i = 1, 2, \ldots, n_b - 1.
\]  

(33)

Here, Eq. (33), together with the second equation of (29) correspond to the function \( P(\cdot) \) in our general setup (10) with the variable of interest \( X = (\mathbf{d}_{CG}, \mathbf{p}) \). Thus, we are able to obtain the time series of \( (\mathbf{d}_{CG}, \mathbf{p}) \), containing model error, for the imperfect model (29), which will be the only information used in estimating \( \theta \).

Parameters \( \theta_{eq} := (a, \beta) \) appeared in the equilibrium statistics. Therefore they constitute the parameter set \( \theta_{eq} \) that we discussed in Section 4.1 To determine these parameters, consider the Gibbs measure of the imperfect model (29) (following the notation in (20)),

\[
\rho_{eq}(\mathbf{d}_{CG}, \mathbf{p}; \theta_{eq}) \propto \exp \left[ -\beta \left( V(\mathbf{d}_{CG}; \mathbf{a}) + \frac{1}{2} \mathbf{p}^2 \right) \right],
\]

(34)
which suggests that the temperature $\beta^{-1}$ can be directly estimated from the sample variance of the CG velocities. In practice, since $\{p_i\}$ are independent identical distributed at the equilibrium, we take

$$\hat{\beta}^{-1} = E_{p_{eq}}[p_i^2],$$  

(35)

as the estimate for $\beta^{-1}$. Recall that we use the notation $E_{p_{eq}}$ in (35) since the expectation is defined based on the equilibrium distribution (27) of the underlying dynamics (26).

For the parameters in the potential (30), we will apply the maximum entropy method[37, 38]. The CG relative displacements, due to the translational symmetry, are assumed to be independent and identically distributed at the equilibrium. In particular, the marginal distribution of $q_{i+1} - q_i$ with respect to $\rho_{eq}$ (34) is proposed to be of the form,

$$\rho(x; a, \hat{\beta}) \propto \exp\left(-\beta \sum_{j=1}^4 a_i x^j\right),$$

for $i = 1, 2, \ldots, n_b - 1$. By maximizing the Shannon’s entropy under the moment constraints, we formulate the following maximum entropy estimates for $a$,

$$\hat{a} = \arg\min \int -\rho(x; a, \hat{\beta}) \log \rho(x; a, \hat{\beta}) \, dx,$$

(36)

s.t. $\int \rho(x; a, \hat{\beta}) \, dx = \frac{1}{n_b - 1} \sum_{i=1}^{n_b-1} E_{p_{eq}}[(q_{i+1} - q_i)^j], \quad j = 1, 2, 3, 4,$

(37)

where $\hat{\beta}$ is determined by (35) and the moments on the right hand side of (37) will be approximated using Monte-Carlo based on the available observations of $d_{CG}$. Thus, such constrained minimization problem (36) and (37) is equivalent to the reverse Monte Carlo method[39]. Figure 2 shows the comparison between $\rho(x; a, \hat{\beta})$ and the histogram of $q_{i+1} - q_i$, and excellent agreement has been found.

Figure 2: Comparison of the estimated density, $\rho(x; a, \hat{\beta})$ (red-dash), with the histogram of $q_{i+1} - q_i$ (blue-solid).

Up to this point, we have estimated $\theta_{eq} = (a, \hat{\beta})$ in the Gibbs measure (34) of the imperfect model (34). We are going to apply our parameter estimation approach to estimate $\{\gamma_i\}$ in the damping matrix $\Gamma$. These parameters are responsible for the non-equilibrium property. To satisfy the assumptions posed in Section 4, the external forcing of the underlying dynamics is set to be a constant external forcing, exerted equally on the first block of atoms, and the observable $A(X)$ are the velocities of the first two CG particles. As a result, following (19), the corresponding external forcing $c_R \delta f$ for the imperfect model satisfies

$$c_R = (1, 0, \ldots, 0)^\top \in \mathbb{R}^{n_b}, \quad \delta f = \delta.$$

(38)

Therefore the forcing is admissible. With observables $A = (p_1, p_2)^\top$, following (20) - (21), we write

$$\hat{k}_A(t) = \hat{\beta} E_{p_{eq}}[(p_1(t), p_2(t))^\top \otimes p_1(0)],$$

$$\hat{k}_A(t; \theta) = \hat{\beta} E_{p_{eq}(\theta)}[(p_1(t), p_2(t))^\top \otimes p_1(0)],$$

(39)
where \( \hat{\beta} \) is the estimates of \( \beta \) in \( \theta_{eq} \) (35). Recall that, as we have mentioned in Remark 4.5 we set \( p_R = \rho_{eq}(dCG,p;\theta_{eq}) \) in computing the marginal essential statistics (21). Interested readers are referred to [2] for the computational details. Dropping \( \hat{\beta} \) in (39), we formulate the following constrained nonlinear least-squares problem:

\[
\min_{\theta} \sum_{i=1}^{K} \sum_{j=1}^{2} \left( g_j^*(t_i) - \tilde{g}_j(t_i;\theta) \right)^2, \quad \text{s.t.} \, \beta = \hat{\beta}, \, a = \tilde{a}, \quad \gamma_0 > 2 (|\gamma_1| + |\gamma_2|),
\]

where \( g_j^*(t) = E_{\rho_{eq}^{*}}[p_j(t)p_1(0)], \quad \tilde{g}_j(t;\theta) = E_{\rho_{eq}^{*}[\theta]}[p_j(t)p_1(0)], \quad j = 1, 2. \)

Recall that both \( \{g_j^*(t_i)\} \) and \( \{\tilde{g}_j(t_i;\theta)\} \) will be approximated by Monte-Carlo of the form (6). The only difference is that \( \{g_j^*(t_i)\} \) is based on the available CG observations, while \( \{\tilde{g}_j(t_i;\theta)\} \) is based on the time series generated by simulating the imperfect model (29) at the corresponding parameter value. In the numerical simulation of the imperfect model (29), we generalized the Verlet-like integration algorithm to the case where the damping coefficient is a SPD matrix in an efficient manner. For more details, please check the Appendix B.

As an example, we set \( J = 10. \) So \( n_b = 10 \) since \( N = 100. \) Setting \( t_i = 0.2i \) with \( i = 1, 2, \ldots, 20, \) we solve (40) using the numerical scheme reviewed in Section 2. Specifically, a polynomial chaos expansion with order-6 Legendre polynomials is used to approximate the cost function (40) to avoid excessive computational cost for evaluating \( \hat{g}_j \) on new parameter \( \theta \) (see [2] for detail). Together with the previous estimates \( \hat{\beta}^{-1} \) (35) and \( \tilde{a} \) (36), Table 1 shows the value of the estimates. Figure 3 and Figure 4 provide the results in fitting the essential statistics and recovering the full response of the underlying dynamics, respectively. One can observe from the figures that the estimates provide an excellent fit on the auto-correlation function. For the cross-correlation estimation, the poor performance at the longer time might be caused by the Monte-Carlo error involved in (40) and possibly the polynomial chaos approximation on the cost function. Interestingly, the imperfect model produces reasonably accurate estimates of the full nonlinear response statistics of the underlying dynamics.

### Table 1: Estimates of the parameter in (29)

| Parameter (\( \theta \)) | \( \hat{\beta}^{-1} \) | \( a_1 \) | \( a_2 \) | \( a_3 \) | \( a_4 \) | \( \gamma_0 \) | \( \gamma_1 \) | \( \gamma_2 \) |
|--------------------------|----------------|--------|--------|--------|--------|----------|----------|----------|
| Estimates (\( \theta \)) | 0.0247 | 0.0193 | 0.148 | -0.0830 | 0.0331 | 1.23 | -0.371 | -0.0420 |

Figure 3: Comparison of the time correlation functions involved in the nonlinear least-squares problem (40). The blue-solids curves are computed from the available CG observations, while the red-dash curves are the approximation of the imperfect model based on the estimates. (sample size: \( 1 \times 10^6 \))

### 6  A severely truncated spatio-temporal system

In this section, we consider a parameter estimation problem where the underlying equilibrium distribution is high-dimensional and unknown. With this constraint, there is a practical issue in the estimation of the linear response
Figure 4: Reconstruction of full response statistics: underlying dynamics (blue-solid) vs. the imperfect model (red-dash). We take $\delta = 0.1$ in (38) as the external forcing. The full response statistics (3) are computed via Monte-Carlo. Check Appendix A for the details.

statistics $k^*_A$ for solving the least-squares problem in (7), since the FDT approximation to these statistics require an explicit form of the underlying density. To overcome this difficulty, we will consider the kernel mean embedding method [40, 41, 42] (KME) as a nonparametric estimator of the equilibrium density. We shall see that the resulting linear response statistics obtained from the kernel mean embedding estimate is more accurate than the Quasi-Gaussian FDT (QG-FDT) linear response [43, 44]. To test the performance of the parameter estimation method in recovering the full response statistics, we consider estimating parameters of an arbitrary closure model of a severely truncated Kuramoto-Shivashinsky (KS) equation, which full solutions exhibit a spatio-temporal chaotic patterns observed in many applications, such as the trapped ion modes in plasma [45] and phase dynamics in reaction-diffusion systems [46].

6.1 The underlying dynamics

The underlying dynamics of the KS equation, defined on a one-dimensional periodic domain $[0, L]$, can be approximated on uniformly distributed spatial nodes [47, 48, 49] with discrete Fourier modes that satisfy,

$$u_0 = 0, \quad \dot{u}_k = (q_k^2 - q_k^4) u_k - \frac{i q_k}{2} \sum_{1 \leq |\ell|, |k-\ell| \leq \frac{N}{2}-1} u_\ell u_{k-\ell}, \quad |k| \leq \frac{N}{2} - 1.$$  (41)

Here, $q_k = \frac{2\pi k}{L}$ and $u_k(t)$ denote the Fourier coefficients that satisfy $u_{-k} = u_k^*$. We take the numerical solution to the system [41] with $N = 48$ as the solutions to the underlying dynamics; this choice of $N$ is sufficient to recover the inertial manifold of the system [49]. This numerical approximate solution is generated using an exponential time difference fourth order Runge-Kutta (ETDRK4) method [48] with the discrete Fourier transform of $(1 + \sin(x)) \cos(x)$ on the grid points $x_i = iL/N |i = 0, \ldots, N-1$ as the initial condition. Here, the ETDRK4 method is applied with $L = 2\pi / \sqrt{0.085}$. The value of $L$ is large enough to yield chaotic solutions [47]. We drop the first $5 \times 10^4$ time units to allow the system to relax, and take $u_1$ and $u_2$ in the next $5 \times 10^5$ time units with step length $\Delta t = 0.1$ as the training data set. We denote this training data set as $\{u_{1,t}, u_{2,t}\}_{t=1,\ldots,T}$, where $T = 5 \times 10^6$.

6.2 FDT response based on kernel mean embedding estimates

Our interest is to predict the full response of the first Fourier mode, $u_1$, under small external forcing, without knowing the full underlying dynamics in [41]. Instead, we are only given the dynamical equation associated to the first Fourier mode,

$$\dot{u}_1 = (q_1^2 - q_1^4) u_1 - i q_1 u_1 u_2 + g, \quad g := -\frac{i q_k}{2} \sum_{1 \leq |\ell|, |k-\ell| \leq \frac{N}{2}-1} u_\ell u_{k-\ell} + i q_1 u_1 u_2$$  (42)
where \( g \) is the \textit{identifiable unresolved component} in the following sense. Given the partial dynamics \([42]\) and the training data set \( \{u_{1,t}, u_{2,t}\}_{t=1,...,T} \), one can extract \( \{g_t\}_{t=1,...,T} \) by fitting these data set into the first equation in \([42]\), where the time derivative is replaced by finite difference approximations \([49]\).

As we previously alluded, the key issue here is that we have no access to the explicit form of the equilibrium distribution of the full dynamics, which is very high-dimensional. Since we only have data set \( \{u_{1,t}, u_{2,t}, g_t\}_{t=1,...,T} \), we propose to consider the linear response operator in \([5]\), computed by averaging over \( B \) that is defined with respect to a marginal density, \( p_R \), of these three variables \( \{u_{1,t}, u_{2,t}, g_t\} \). This is exactly where MLR operator \([21]\) comes in. Specifically, we define

\[
x := \left( \text{Re}(u_1), \text{Im}(u_1) \right)^\top, \quad y := \left( \text{Re}(u_2), \text{Im}(u_2) \right)^\top, \quad \text{and} \quad z := \left( \text{Re}(g), \text{Im}(g) \right)^\top.
\]

Under a constant external forcing on \( u_1 \), which is admissible according to the criteria in Definition IV.1, Eq. \([21]\) suggests that, using the real representation \([43]\), the MLR operator of \([41]\) is given by

\[
\hat{\mathcal{L}}_A(t) = \mathbb{E}_{\rho_{eq}} \left[ A(x(t)) \otimes B(x(0), y(0), z(0)) \right], \quad B(x, y, z) := -\nabla_x \log \{ p_R(x, y, z) \},
\]

where the unknown \( p_R \) corresponds to the marginal distribution of \( (x, y, z) \) at the equilibrium of the underlying dynamics \([21]\). While the expectation over \( \rho_{eq} \) can be realized using Monte-Carlo average over the available data, one needs an actual expression of \( p_R \) which appears in \( B \). This way, we have replaced the problem of estimating a very high-dimensional density of \( \{u_1, ..., u_N\} \) to a moderately low-dimensional (in this case, six-dimensional real-valued) function. In the remaining of this section, we will discuss how to approximate \( p_R \) with a KME estimator, \( \hat{p}_R \), and check the validity of the MLR in predicting the full response statistics of the underlying dynamics under small perturbations.

Recall that KME, as a nonparametric approach to learn a distribution function from the sample, can be formulated with an arbitrary kernel that uniquely determines the reproducing kernel Hilbert space (RKHS). We consider kernels based on orthogonal basis of a weighted \( L^2 \)-space. Specifically, let the target density function \( f \in \mathcal{H}_\rho \subset L^2(\mathbb{R}^d, q^{-1}) \), where \( \mathcal{H}_\rho \) denotes the RKHS induced by a set of orthonormal basis \( \{\psi_m = \psi_m \rho\} \) of \( L^2(\mathbb{R}^d, q^{-1}) \) and \( q \) denotes a positive weight function on \( \mathbb{R}^d \) (see \([41, 42]\) for details), where \( \tilde{m} = (m_1, m_2, ..., m_d) \) is a multi-index notation with nonnegative components. In particular, we take \( q \) to be the standard \( d \)-dimensional Gaussian distribution, and define the kernel

\[
k_\rho(x, y) := \sum_{m \geq 0} \lambda_{\tilde{m}} \psi_m(x) \psi_m(y), \quad \lambda_{\tilde{m}} := \rho^{||\tilde{m}||_1}, \quad ||\tilde{m}||_1 := \sum_{k=1}^d m_k,
\]

for \( \rho \in (0, 1) \). We have shown \([50]\) that equipped with the inner product

\[
\langle f, g \rangle_{\mathcal{H}_\rho} := \sum_{m \geq 0} \frac{\tilde{f}_m \tilde{g}_m}{\rho^{||\tilde{m}||_1}}, \quad f = \sum_{m \geq 0} \tilde{f}_m \psi_m, \quad g = \sum_{m \geq 0} \tilde{g}_m \psi_m,
\]

\( \mathcal{H}_\rho \) is an RKHS with a bounded kernel \( k_\rho \), satisfying

\[
k_\rho(x, y) = (2\pi)^{-d} (1 - \rho^2)^{-\frac{d}{2}} \exp \left\{ -\frac{1}{2(1 - \rho^2)} \left( ||x||^2 + ||y||^2 - 2\rho \sum_{i=1}^d x_i y_i \right) \right\}.
\]

Here, \( k_\rho \) \([45]\) can be interpreted as a generalization of Mehler kernel \([51]\).

Then the order-\( M \) KME estimate of \( f \), denoted by \( f_M \), is given by

\[
f_M := \sum_{||\tilde{m}||_1 \leq M} \tilde{f}_m \psi_m, \quad \tilde{f}_m := \int_{\mathbb{R}^d} f \psi_m q^{-1} \, dx = \int_{\mathbb{R}^d} f \psi_m \, dx.
\]

Since \( \{\psi_m\} \) forms an orthonormal basis for \( L^2(q) \), we take

\[
\psi_m(x) = \prod_{k=1}^d \psi_{m_k}(x_k),
\]

where \( \psi_{m_k} \) is the order-\( m_k \) normalized Hermite polynomial. We should point out with this choice of basis representation, we basically arrive at a polynomial chaos approximation \([52]\) of a density function \( f \). In practice, the integral in \([47]\) can be approximated by Monte-Carlo,

\[
\tilde{f}_m \approx \tilde{f}_{m,N} = \frac{1}{N} \sum_{n=1}^N \psi_m(X_n),
\]

(48)
where \( \{X_n\}_{n=1}^{N} \) are sampled from the target density function \( f \). One can define the order-\( M \) empirical KME estimate of \( f \) as
\[
\hat{f}_{M,N} := \sum_{|\hat{\mathbf{n}}| \leq M} \hat{f}_{\hat{\mathbf{n}}} \Psi_{\hat{\mathbf{n}}}.
\]
(49)

In our application, the sample \( \{X_n\} \), as a time series of \((x,y,z)\) (43), can be interpreted as a stationary process with the target distribution function \( f = p_{R} \).

Take the observable \( A(x,y,z) = x \), and let \( \hat{p}_{R} \) be the order-\( M \) KME estimator of \( p_{R} \). Then, Eq. (44) is approximated by an order-\( M \) MLR estimator,
\[
\hat{k}_{A}^{\dagger}(t) \approx -E_{p_{eq}} \left[ x(t) \otimes \nabla_{x} \log(\hat{p}_{R}(x(0), y(0), z(0))) \right].
\]
(50)

Recall that the expectation is still over \( p_{eq}^{\dagger} \), since we are using the observed time series from the underlying dynamics to compute the statistics (50).

**Remark 6.1.** Theoretically, the two-point statistics on the right-hand side of (50) may not be well-defined, since the basis function \( \Psi_{\hat{\mathbf{n}}} \) used in the order-\( M \) KME estimate (47) can be negative. Our error estimate (50) suggests that for large enough \( M \), there exists a compact domain \( D_{\epsilon} \subset \mathbb{R}^{d} \) for which the \( \hat{k}_{A}^{\dagger} \) can be approximated up to order-\( \epsilon \) and the order-\( M \) MLR estimator (the right-hand-term in (50)) is well-defined. In practice, assuming that \( \epsilon > 0 \) is small enough, the sample data \( \{X_n\} \subset D_{\epsilon} \). Thus, for large enough \( M \), the MLR estimator is well-defined and can be approximated by a Monte-Carlo integral over the entire data set.

Using the approximate response operator in (50), we now compare the resulting approximate linear response statistics with the corresponding full response statistics. The numerical result (Figure 5) suggests that the full response is well captured by the linear response that uses \( \hat{p}_{R}(x,y,z) \) up to finite time. This result empirically validates the MLR response in this model. In fact, the MLR statistic is more accurate compared to the QG-FDT response [43, 44], which is obtained from response operator (6) with \( B^{\dagger} \) replaced by the one defined with a Gaussian approximation of the invariant density.

![Figure 5: The response of \( u_{1} \) component of the underlying dynamics (41) to a constant external forcing \( \delta f = \frac{1}{N}(0.5 + 0.5i) \). The blue solid curves show the full nonlinear response of \( u_{1} \) approximated by Monte-Carlo based on \( 5 \times 10^{5} \) times of simulations; the red dash shows the performance of QG-FDT based on the first six modes \((u_{1}, u_{2}, \ldots, u_{6})\); yellow dot-dash curves correspond to the MLR statistics produced by (50) based on \((u_{1}, u_{2}, f)\) as the resolved variable.](image)

**6.3 The truncated model**

The focus in this section is to devise a simple dynamical model that can improve the estimation of the full response, yet retaining the accuracy of the equilibrium distribution of variable \( u_{1} \) obtained through the KME approach discussed in the previous section. We should point out that while many reduced-order models have been proposed for the KS model (see[47, 53, 49, 54] just to name a few), their constructions typically require the knowledge of more than two Fourier modes. In our case, we only work with the data set of the first two Fourier modes and the resulting identifiable irrelevant variables, \( g \), as defined in (42). While the dynamical equation for the first component in (42)
is given, the fact that the linear term is unstable, since $q_1^2 - q_1^4 > 0$, makes it difficult to devise a closure model that is stable and ergodic with accurate invariant density of the variable $x := (\text{Re}(u_1), \text{Im}(u_1))^T$.

Given such constraints, we propose the following “semi-parametric” extended Langevin equation as our imperfect model

$$\begin{align*}
\dot{x} &= v \\
\dot{v} &= \Lambda \nabla_x \log (\hat{\rho}(x)) - \Gamma v + \sigma W_t,
\end{align*}$$

(51)

where $x \in \mathbb{R}^2$ is the real representation of $u_1$ as in (43) and $W_t$ denotes the standard 2-D Wiener process. By “semi-parametric”, we refer to the combination of linear parametric equation in the right-hand-side with a “non-parametric” term that involves $\hat{\rho}(x)$ that is estimated by the kernel mean embedding approximation of the marginal distribution of $x$ at equilibrium. We should point out that since the density $\hat{\rho}$ is approximated using a specific choice of kernel with Hermite polynomials, as explained in the previous section, the resulting model in (51) is effectively parametric. The following proposition guarantees that Eq. (51) is able to recover the marginal distribution of $x$ at the equilibrium state.

**Proposition 6.2.** Assume the extended Langevin equation (51) is defined with SPD $\Lambda$ and positive definite $\Gamma$. Let $\sigma$ satisfies the Lyapunov equation

$$\Gamma \Lambda + \Lambda \Gamma^\top = \sigma \sigma^\top.$$

(52)

and $\hat{\rho}(x)$ be a continuously differentiable density function controlled by a Gaussian density. That is, there exist positive constants $C$ and $\beta$ such that

$$\hat{\rho}(x) \leq C \exp \left( -\frac{\|x\|^2}{2\beta^2} \right), \quad \forall x \in \mathbb{R}^2.$$

(53)

Then the system in (51) is ergodic with a unique equilibrium density

$$\rho_{eq}(x,v) \propto \hat{\rho}(x) \exp \left( -\frac{1}{2} v^\top \Lambda^{-1} v \right).$$

(54)

See Appendix C for the proof. Here, the decay rate condition (53) is consistent with the assumption $p_R \in \mathcal{H}_p$ (see (50)). Figure 6 shows the performance of the KME estimator. One can see that, compared with the Gaussian approximation, the KME estimator has a better performance near the peaks of the target density functions. Meanwhile, Figure 6 also justifies the decay rate assumption (53) in Proposition 6.2. For the corresponding parameter estimation problem, with the Lyapunov equation (52), the parameter space of (51) can be reduced to

$$\theta := (\Lambda, \Gamma),$$

(55)

where $\Lambda$, as a covariance matrix of the auxiliary variable $v$, is SPD and $\Gamma$ is positive definite. Connecting to Section 4, components of $\Lambda$ are the equilibrium parameters, $\theta_{eq}$.

![Figure 6](image-url)

Figure 6: The KME approximation of the marginal distribution of $x$ in (43). The blue solid curves are the PDFs obtained from the observed time series of $u_1^N$ using kernel density estimates for smoothing. The red dash curves are the order-5 KME estimates. The yellow curves are the density function of the normal distribution of the same mean and variance.
6.4 Parameter estimation and numerical results

Next, we will estimate the parameters \( \theta = (\Lambda, \Gamma) \) using the least-squares formula (7) and check whether the imperfect model in (51) with the estimated parameters can produce a better estimate of the full response statistics compared to the linear responses as shown in Figure 5. First, notice that the imperfect model (51) yields an equilibrium distribution (54), which suggests the auxiliary variable \( v \) is Gaussian with covariance matrix \( \Lambda \) at the equilibrium. Thus, we can use the sampled covariance of \( \dot{x} \), approximated through a finite-difference, to directly estimate \( \Lambda \). We will denote the estimate as \( \hat{\Lambda} \).

For the damping coefficient \( \Gamma \), we take the diagonal elements of the approximated linear response operator (50) as the essential statistics (they are shown as the yellow stars in Figure 7), and we are able to set up the following nonlinear least square problems to estimate \( \Gamma \):

\[
\min_\Gamma \sum_{i=1}^{K} \sum_{j=1}^{2} \left( g_j^\dagger(t_i) - \hat{g}_j(t_i; \Gamma, \hat{\Lambda}) \right)^2, \tag{56}
\]

where

\[
g_j^\dagger(t) = E_{p_{eq}} \left[ x_j(t) \partial x_j \log(\hat{p}) \right], \quad \hat{g}_j(t; \Gamma, \hat{\Lambda}) = E_{p_{eq}(\Gamma, \hat{\Lambda})} \left[ x_j(t) \partial x_j \log(\hat{p}) \right], \quad j = 1, 2.
\]

In particular we take \( t_i = i \) for \( i = 1, 2, \ldots, 20 \). We should point out that the \( E_{p_{eq}} \) will be approximated by Monte-Carlo over the true time series of \( \{u_1, t, u_2, t, g_t\}_{t=1, \ldots, T} \). We use Euler-Maruyama method to generate time series from the imperfect dynamics at \( \Gamma \) and \( \hat{\Lambda} \). Similar to the setup of the numerical scheme of the underlying dynamics (41), we drop the first \( 5 \times 10^4 \) time units to relax the system, and sub-sample \( x \) in the next \( 5 \times 10^5 \) time units with step length \( \Delta t = 0.1 \). Table 2 records the estimated parameter values.

![Figure 7](image-url)

**Figure 7:** Result of the nonlinear least squares fitting (56). The values of the essential statistics \( \{g_j^\dagger(t_i)\} \), taken from the blue curves, are highlighted by yellow stars. The red dash curves are the corresponding two-time statistics of the imperfect model (51) the estimated parameters reported in Table 2.

**Table 2:** Estimates of the parameter in (55)

| \( \Lambda_{11} \) | \( \Lambda_{12} \) | \( \Lambda_{22} \) | \( \Gamma_{11} \) | \( \Gamma_{12} \) | \( \Gamma_{21} \) | \( \Gamma_{22} \) |
|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| \( 1.3675 \times 10^{-3} \) | \( -1.01 \times 10^{-3} \) | \( 1.3544 \times 10^{-3} \) | \( 0.5327 \) | \( -0.5038 \) | \( -0.4953 \) | \( 0.6176 \) |

In Figure 7, one can see that the estimated linear response statistics (red dashes) from the imperfect model reproduces the qualitative feature of the true linear response statistics. Figure 8 shows the recovery of the full response using the imperfect model in (51) based on the estimated parameters in Table 2. One can see that the real and imaginary components of \( u_1 \) share similar behavior in both the two-point statistics and the full response statistics. We note that the imperfect model is able to capture such symmetry. Compared with the approximated linear response statistics showed in Figure 5 the full response statistics of the imperfect model produce more accurate estimation for large value of \( t \).
7 Summary and discussion

Motivated by the FDT, we have demonstrated a response-based estimation method to determine parameters that are responsible for the non-equilibrium statistics in stochastic dynamics. Unlike the prior works [1, 2], which were focused on the perfect model case, the emphasis of this paper is placed on more practical scenarios where model error is present. We started with a conventional two-scale model and showed that the response-based approach can outperform a classical averaging method that is based on the equilibrium statistics, even in the case when the method can be justified.

More importantly, we demonstrated that in general, the parameter estimation problems are complicated by the facts that only partial observations are available and the response statistics might not be accessible. To address these difficult issues, we have outlined a general strategy that can be efficiently implemented in practice. In particular, we introduced the concepts of admissible external forcing, for which the forces on the variable of interest (X) can be identified and computed, as well as the MLR statistics, as computable statistics in place of the linear response statistics of the underlying full dynamics that is inaccessible in general.

To illustrate these general strategies in specific context, we considered the Langevin dynamics model, a widely used stochastic model in practice. We studied two examples, including a one-dimensional chain model, coarse-grained from a full molecular dynamics model; and an extended Langevin dynamics, obtained from a severe truncation of the KS equation in the Fourier space. In both cases, we demonstrate how our general strategy is implemented and validated. More specifically, after the model parameters are determined from the least-squares approach, we compared the nonlinear response of the imperfect model with that of the underlying dynamics under admissible forcing. Quantitative agreement has been found.

For the second example, we also proposed the nonparametric KME formulation to estimate the marginal density, and more importantly, the marginal essential statistics, to enable the estimation using the response statistics. We report the convergence analysis of this estimator elsewhere [50].

Acknowledgment

This research was supported by NSF under grant DMS-1619661 and DMS-1819011. JH also acknowledges the support from NSF grant DMS-1854299.

A Approximation of the full response statistics via Monte-Carlo

To test the validity of the parameter estimates, we compare the full response statistics of the underlying dynamics and the imperfect model. For example, in Section 5 to generate Figure 4, one needs to compute the full response of the underlying dynamics [25] and the imperfect dynamics [29] under a constant external forcing on the first block.
of particles and the first CG particle, respectively. In this appendix, we review the Monte-Carlo (MC) approach used in approximating the full response statistics $\Delta E[A](t)$ under a general Itô diffusion setting.

Recall that in Section 2 we have used (1) and (2) to represent the unperturbed and perturbed dynamics respectively. In the verifications, the unknown parameter $\theta$ in the two dynamics will be replaced by an estimate $\hat{\theta}$. To compute $\Delta E[A](t)$, since $E_{\rho^\theta}[A(X)]$ in (3) is an equilibrium statistics of the unperturbed dynamics, it can be directly approximated by an ensemble average

$$E_{\rho^\theta}[A(X)] \approx \frac{1}{N} \sum_{i=1}^{N} A(x_i),$$

where $\{x_i = X(t_i)\}$ denotes a time series of the unperturbed dynamics (1) generated at the equilibrium state. Thus, we are able to reduce our problem to approximating the non-equilibrium statistics $E_{\rho^\theta}[A(X^\delta(t))]$.

Introducing the transition kernel $\rho^\delta(x, t|x_0, 0)$ of the perturbed dynamics (2), $E_{\rho^\theta}[A(X^\delta(t))]$ can be calculated by a double integral

$$E_{\rho^\theta}[A(X^\delta(t))] = \int \int A(x) \rho^\delta(x, t|x_0, 0) p_{eq}(x_0) \, dx \, dx_0,$$

where we have used the fact that the perturbed dynamics (2) is also initiated at the equilibrium state of the unperturbed dynamics (1) with equilibrium distribution $p_{eq}$. Further notice that the transition kernel satisfies the Fokker-Planck equation

$$\frac{\partial \rho^\delta}{\partial t} = (\mathcal{L}^\delta)^* \rho^\delta, \quad \rho^\delta(x, t|x_0, 0) = \delta(x-x_0),$$

where $\mathcal{L}^\delta$ denotes the generator of the perturbed dynamics (2). One can rewrite the transition kernel using the semi-group notation: $\rho^\delta(x, t|x_0, 0) = e^{t(\mathcal{L}^\delta)^*} \delta(x-x_0).$ Then, the double integral (58) becomes

$$E_{\rho^\theta}[A(X^\delta(t))] = \int \int \rho^\delta(x, t|x_0, 0) e^{t(\mathcal{L}^\delta)^*} \delta(x-x_0) p_{eq}(x_0) \, dx \, dx_0$$

$$= \int e^{t(\mathcal{L}^\delta)^*} A(x) \int \delta(x-x_0) p_{eq}(x_0) \, dx_0 \, dx$$

$$= \int e^{t(\mathcal{L}^\delta)^*} A(x) p_{eq}(x) \, dx.$$

Here, $A(x, t) := e^{t(\mathcal{L}^\delta)^*} A(x)$ solves the Kolmogorov’s backward equation (dual form of the Fokker-Planck equation)

$$\frac{\partial u^\delta}{\partial t} = \mathcal{L}^\delta u^\delta, \quad u^\delta(x, 0) = A(x),$$

which suggests that $u^\delta(x, t)$ corresponds to the conditional expectation $E[A(X^\delta(t)) | X^\delta(0) = x]$. Together with Eq. (59), we derive the following two-layer Monte-Carlo approximation,

$$E_{\rho^\theta}[A(X^\delta(t))] = \frac{1}{N} \sum_{i=1}^{N} u^\delta(x_i, t) \approx \frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} X^\delta,x_i(t; U_j),$$

where $X^\delta,x_i(t; U_j)$ denotes the realization of (2) with initial condition $X^\delta(0) = x_i$, given a sample path $U_j$ of the Wiener process $U_s$ ($0 \leq s \leq t$) in (2).

In the numerical implementation, after generating the time series $\{x_i\}$ at $p_{eq}$, one can duplicate $x_i$ up to $M$ times and solve (2) in parallel based on the samples from the time series as initial conditions. Although, such a procedure is very similar to the bootstrapping techniques in statistics, they have totally different motivations. In particular, we take $N = 10^5$ and $M = 10^4$ in estimating the full response of the underlying dynamics and imperfect model shown in Figure 4.

B Verlet-like integration algorithm

In Section 3 the underlying dynamics (25) is a Langevin equation, written as

$$\begin{cases}
\dot{r} = v, \\
\dot{v} = \frac{F(t)}{m} - \gamma v + \sqrt{\frac{2k_BT}{m}} \, \xi(t),
\end{cases}$$

(61)
where the scalar \( \gamma \) denotes the damping coefficients. In this Appendix, we review the Verlet-like integration algorithm \( 85 \) for \( 61 \), and discuss the implementation of the algorithm when the damping coefficient is a symmetric positive definite (SPD) matrix.

The integration scheme \( 85 \) for \( 61 \) can be summarized as

\[
\begin{align*}
    r_{n+1} &= r_n + c_1 \Delta t \dot{v}_n + c_2 \Delta t^2 a_n + \delta r_R, \\
v_{n+1} &= v_n + (c_1 - c_2) \Delta t a_n + c_2 \Delta t a_{n+1} + \delta v_R, \\
c_0 &= e^{-\gamma t}, \quad c_{i+1} = \frac{1-c_0}{\Delta t} e^{-\gamma t}, \quad a_{n+i} = \frac{F(r_{n+i})}{m}, \quad i = 0, 1,
\end{align*}
\]

(62)

where the covariance matrix \( \Sigma \) is given by

\[
\Sigma = \begin{pmatrix}
    \Sigma_{rr} & \Sigma_{rv} \\
    \Sigma_{vr} & \Sigma_{vv}
\end{pmatrix},
\]

\[
\begin{align*}
    \Sigma_{rr} &= \Delta t \frac{k_B T}{m} \gamma^{-1} \left[ 2 - \gamma^{-1} (3 - 4e^{-\gamma \Delta t} + e^{-2\gamma \Delta t}) \right], \\
    \Sigma_{rv} &= \frac{k_B T}{m} (1 - e^{-2\gamma \Delta t}), \\
    \Sigma_{vr} &= \frac{k_B T}{m} \gamma^{-1} (1 - e^{-\gamma \Delta t})^2.
\end{align*}
\]

(63)

In the case where the damping coefficients form a SPD matrix, e.g. the imperfect model \( 29 \) in Section 5, such scheme \( 62 \) still works with the help of the matrix exponential. However, such matrix exponential introduces a number of matrix-vector multiplications (62) and slows down the simulation.

By a change of basis, one can avoid most of the matrix-vector multiplications. We first rewrite \( 61 \) as

\[
\begin{align*}
    \dot{r} &= v, \\
    \dot{v} &= \frac{F}{m} - \Gamma v + \sqrt{2k_B T/m} \sigma W_t,
\end{align*}
\]

(64)

where \( \Gamma = \sigma \sigma^T \) is a SPD matrix. Notice that \( \Gamma \) is SPD and real, consider its eigen-decomposition \( \Gamma = Q \Phi Q^T \), where \( \Phi \) is diagonal and \( Q \) is real and orthonormal. By taking the eigenvectors (columns of \( Q \)) as our basis, we introduce the following variables with respect to the new basis

\[
\begin{align*}
    \dot{r}^* := Q^T r, \\
    \dot{v}^* := Q^T v, \\
    \dot{\tilde{F}}(\tilde{r}) := Q^T F(Qr), \\
    \dot{\tilde{W}}_t := Q^T W_t,
\end{align*}
\]

and the system \( 64 \), after left-multiplying \( Q^T \) to both sides, can be rewritten as

\[
\begin{align*}
    \dot{\tilde{r}} &= \tilde{v}, \\
    \dot{\tilde{v}} &= \frac{\dot{\tilde{F}}(\tilde{r})}{m} - \Phi \tilde{v} + \sqrt{2k_B T/m} \Phi^{1/2} \tilde{W}_t.
\end{align*}
\]

(65)

Due to the the diagonal structure of \( \Phi \), solving \( 65 \) by \( 62 \) involves matrix-vector multiplications only in evaluating \( \tilde{F} \). As an added benefit, all the integrating factors, e.g. \( c_i \) in \( 62 \), can be determined without matrix exponential.

In terms of sampling the random forcing, since \( Q \) is real and orthonormal, \( \tilde{W}_t = Q^T W_t \) is still a standard Wiener process, which can be sampled directly.

\section{Extended Langevin equation}

In Section 6 we use an extended Langevin equation \( 51 \) as the imperfect model. In this appendix, we discuss the proof of proposition 6.2. The first step is transforming the extended Langevin equation \( 61 \) into a Langevin equation we are familiar with. Consider the following change of variables

\[
q := \Lambda^{-\frac{1}{2}} x, \quad p := \Lambda^{-\frac{1}{2}} v,
\]

then one can show that \( (q, p) \) satisfies

\[
\begin{align*}
    \dot{q} &= p, \\
    \dot{p} &= \nabla_q \log \left( \tilde{\rho}(\Lambda^\frac{1}{2} q) \right) - \Gamma p + \tilde{\sigma} \tilde{W}_t, \\
    \tilde{\Gamma} &= \Lambda^{-\frac{3}{2}} \Gamma \Lambda^\frac{1}{2}, \quad \tilde{\sigma} = \Lambda^{-\frac{1}{2}} \sigma.
\end{align*}
\]

(66)

Here, we have used the fact that \( \Lambda \) is SPD to define \( \Lambda^{-\frac{1}{2}} \). Notice that \( \tilde{\Gamma} \) is positive definite. As a result of the fluctuation-dissipation theorem, a necessary condition for \( 66 \) to form a Langevin equation is

\[
\tilde{\Gamma} + \tilde{\Gamma}^T = \tilde{\sigma} \tilde{\sigma}^T,
\]

which is equivalent to the Lyapunov equation \( 52 \).

Finally, we apply the Theorem 3.2 in \( 55 \) to show the ergodicity of \( 66 \). In \( 66 \), by the decay assumption \( 53 \), the potential function \( F(q) = -\log(\tilde{\rho}(\Lambda^\frac{1}{2} q)) \) grows like \( \|q\|^2 \), which fulfills the assumption in the theorem. Thus, \( 66 \) is ergodic with a unique equilibrium distribution

\[
\tilde{\rho}_{eq}(q, p) = \tilde{\rho}(\Lambda^\frac{1}{2} q) \exp \left( -\frac{1}{2} \|p\|^2 \right),
\]

which leads to the density function \( 54 \) of the extended Langevin dynamics \( 51 \).
References

[1] John Harlim, Xiantao Li, and He Zhang. A parameter estimation method using linear response statistics. Journal of Statistical Physics, 168:146–170, 2017.

[2] He Zhang, Xiantao Li, and John Harlim. A parameter estimation method using linear response statistics: Numerical scheme. Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(3):033101, 2019.

[3] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. Physical Review, 140(4A):A1133–A1138, 1965.

[4] Erich Runge and Eberhard KU Gross. Density-functional theory for time-dependent systems. Physical Review Letters, 52(12):997, 1984.

[5] R Haydock, Volker Heine, and MJ Kelly. Electronic structure based on the local atomic environment for tight-binding bands. Journal of Physics C: Solid State Physics, 5(20):2845, 1972.

[6] DR Hamann, M Schlüter, and C Chiang. Norm-conserving pseudopotentials. Physical Review Letters, 43(20):1494, 1979.

[7] M. Toda, R. Kubo, and N. Hashitsume. Statistical Physics II. Nonquilibrium Statistical Mechanics. Springer, 1983.

[8] Denis J Evans and Gary Morriss. Statistical mechanics of nonequilibrium liquids. Cambridge University Press, 2008.

[9] Marc Baaden and Siewert J Marrink. Coarse-grain modelling of protein-protein interactions. Current Opinion in Structural Biology, 23:878–886, October 2013.

[10] Andrew R Leach. Molecular modelling: principles and applications. Pearson education, 2001.

[11] Gregory A Voth. Coarse-graining of condensed phase and biomolecular systems. CRC press, 2008.

[12] T. Schlick. Molecular Modeling and Simulation: An Interdisciplinary Guide. Springer-Verlag, 2002.

[13] Zhiyong Zhang, Lanyuan Lu, Will G Noid, Vinod Krishna, Jim Pfaendtner, and Gregory A Voth. A systematic methodology for defining coarse-grained sites in large biomolecules. Biophysical journal, 95(11):5073–5083, 2008.

[14] WG Noid. Perspective: Coarse-grained models for biomolecular systems. The Journal of Chemical Physics, 139(9):09B201_1, 2013.

[15] WG Noid, Jhih-Wei Chu, Gary S Ayton, Vinod Krishna, Sergei Izvekov, Gregory A Voth, Avisek Das, and Hans C Andersen. The multiscale coarse-graining method i. a rigorous bridge between atomistic and coarse-grained models. The Journal of Chemical Physics, 128(24):244114, 2008.

[16] Christophe Chipot and Andrew Pohorille. Free energy calculations. Springer, 2007.

[17] Alain Dequidt and Jose G Solano Canchaya. Bayesian parametrization of coarse-grain dissipative dynamics models. The Journal of chemical physics, 143(8):084122, 2015.

[18] Markos A Katsoulakis and Petr Plecháč. Information-theoretic tools for parametrized coarse-graining of nonequilibrium extended systems. The Journal of Chemical Physics, 139(7):074115, 2013.

[19] Huan Lei, Nathan A Baker, and Xiantao Li. Data-driven parameterization of the generalized langevin equation. Proceedings of the National Academy of Sciences, 113(50):14183–14188, 2016.

[20] Lina Ma, Xiantao Li, and Chun Liu. Coarse-graining langevin dynamics using reduced-order techniques. Journal of Computational Physics, 380:170–190, 2019.

[21] Lina Ma, Xiantao Li, and Chun Liu. The derivation and approximation of coarse-grained dynamics from langevin dynamics. The Journal of Chemical Physics, 145(20):204117, 2016.

[22] Robert B Best, Bin Li, Annette Steward, Valerie Daggett, and Jane Clarke. Can non-mechanical proteins withstand force? stretching barnase by atomic force microscopy and molecular dynamics simulation. Biophysical journal, 81(4):2344–2356, 2001.

[23] Hui Lu and Klaus Schulten. Steered molecular dynamics simulations of force-induced protein domain unfolding. Proteins: Structure, Function, and Bioinformatics, 35(4):453–463, 1999.

[24] CE Leith. Climate response and fluctuation dissipation. Journal of the Atmospheric Sciences, 32(10):2022–2026, 1975.

[25] Grigorios A Pavliotis. Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations, volume 60. Springer, 2014.
[26] Henrik Flyvbjerg and Henrik Gordon Petersen. Error estimates on averages of correlated data. *The Journal of Chemical Physics*, 91(1):461–466, 1989.

[27] Andrew J. Majda and Di Qi. Improving prediction skill of imperfect turbulent models through statistical response and information theory. *Journal of Nonlinear Science*, 26(1):233–285, 2016.

[28] Di Qi and Andrew J. Majda. Predicting fat-tailed intermittent probability distributions in passive scalar turbulence with imperfect models through empirical information theory. *Communications in Mathematical Sciences*, 14(6):1687–1722, 2016.

[29] Andrew J. Majda and Di Qi. Strategies for reduced-order models for predicting the statistical responses and uncertainty quantification in complex turbulent dynamical systems. *SIAM Review*, 60(3):491–549, 2018.

[30] Georg A. Gottwald and John Harlim. The role of additive and multiplicative noise in filtering complex dynamical systems. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 469(2155):20130096, 2013.

[31] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization*. Springer Science & Business Media, 2008.

[32] Tyrus Berry and John Harlim. Linear theory for filtering nonlinear multiscale systems with model error. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 470(2167):20140168, 2014.

[33] Yuji Ishimori. Solitons in a one-dimensional lennard-jones lattice. *Progress of Theoretical Physics*, 68(2):402–410, 1982.

[34] Ben Leimkuhler and Charles Matthews. *Molecular Dynamics*. Springer, 2016.

[35] Christoph Dellago, Peter G. Bolhuis, and David Chandler. Efficient transition path sampling: Application to lennard-jones cluster rearrangements. *The Journal of chemical physics*, 108(22):9236–9245, 1998.

[36] Christopher R. Sweet, Paula Petrone, Vijay S. Pande, and Jesús A. Izaguirre. Normal mode partitioning of langevin dynamics for biomolecules. *The Journal of chemical physics*, 128(14):04B606, 2008.

[37] Rafail V. Abramov. The multidimensional maximum entropy moment problem: A review of numerical methods. *Communications in Mathematical Sciences*, 8(2):377–392, 2010.

[38] Ximing Wu. Calculation of maximum entropy densities with application to income distribution. *Journal of econometrics*, 115(2):347–354, 2003.

[39] Alexander P. Lyubartsev and Aatto Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Physical Review E*, 52(4):391, 1995.

[40] Krikamol Muandet, Kenji Fukumizu, Bharath Sriperumbudur, Bernhard Schölkopf, et al. Kernel mean embedding of distributions: A review and beyond. *Foundations and Trends® in Machine Learning*, 10(1-2):1–141, 2017.

[41] Shixiao W. Jiang and John Harlim. Parameter estimation with data-driven nonparametric likelihood functions. *Entropy*, 21(6):559, 2019.

[42] Shixiao W. Jiang and John Harlim. Modeling of missing dynamical systems: Deriving parametric models using a nonparametric framework. *arXiv preprint arXiv:1905.08082*, 2019.

[43] A.J. Majda, R.V. Abramov, and M.J. Grote. *Information theory and stochastics for multiscale nonlinear systems*. CRM Monograph Series v.25, American Mathematical Society, Providence, Rhode Island, USA, 2005.

[44] Andrey Gritsun and Grant Branstator. Climate response using a three-dimensional operator based on the fluctuation–dissipation theorem. *Journal of the Atmospheric Sciences*, 64(7):2558–2575, 2007.

[45] Robert Eugene LaQuey, SM Mahajan, PH Rutherford, and WM Tang. Nonlinear saturation of the trapped-ion mode. *Physical Review Letters*, 34(7):91, 1975.

[46] Yoshiki Kuramoto and Toshio Tsuzuki. Persistent Propagation of Concentration Waves in Dissipative Media Far from Thermal Equilibrium. *Progress of Theoretical Physics*, 55(2):356–369, 02 1976.

[47] Panagiotis Stinis. Stochastic optimal prediction for the kuramoto–sivashinsky equation. *Multiscale Modeling & Simulation*, 2(4):580–612, 2004.

[48] Aly-Khan Kassam and Lloyd N Trefethen. Fourth-order time-stepping for stiff pdes. *SIAM Journal on Scientific Computing*, 26(4):1214–1233, 2005.

[49] Fei Lu, Kevin K Lin, and Alexandre J. Chorin. Data-based stochastic model reduction for the kuramoto–sivashinsky equation. *Physica D: Nonlinear Phenomena*, 340:46–57, 2017.

[50] He Zhang, John Harlim, and Xiantao Li. Kernel mean embedding linear response. *In preparation*, 2019.
[51] F Gustav Mehler. Ueber die entwicklung einer function von beliebig vielen variablen nach laplaceschen functionen höherer ordnung. *Journal für die reine und angewandte Mathematik*, 66:161–176, 1866.

[52] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, 2010.

[53] Markus Schmuck, Marc Pradas, Grigorios A Pavliotis, and Serafim Kalliadasis. A new mode reduction strategy for the generalized kuramoto–sivashinsky equation. *IMA Journal of Applied Mathematics*, 80(2):273–301, 2013.

[54] John Harlim, Shixiao W Jiang, Senwei Liang, and Haizhao Yang. Machine learning for prediction with missing dynamics. *arXiv preprint arXiv:1910.05861*, 2019.

[55] Jonathan C Mattingly, Andrew M Stuart, and Desmond J Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic processes and their applications*, 101(2):185–232, 2002.