Physics Constrained Nonlinear Regression Models for Time Series

Andrew J. Majda[1] **and John Harlim**[2]

[1] Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, NY 10012, USA
[2] Department of Mathematics, North Carolina State University, NC 27695, USA

E-mail: `jonjon@cims.nyu.edu`, `jharlim@ncsu.edu`

Abstract. A central issue in contemporary science is the development of data driven statistical nonlinear dynamical models for time series of partial observations of nature or a complex physical model. It has been established recently that ad-hoc quadratic multi-level regression models can have finite-time blow up of statistical solutions and/or pathological behavior of their invariant measure. Here a new class of physics constrained multi-level quadratic regression models are introduced, analyzed, and applied to build reduced stochastic models from data of nonlinear systems. These models have the advantages of incorporating memory effects in time as well as the nonlinear noise from energy conserving nonlinear interactions. The mathematical guidelines for the performance and behavior of these physics constrained multi-level regression models as well as filtering algorithms for their implementation are developed here. Data driven applications of these new multi-level nonlinear regression models are developed for test models involving a nonlinear oscillator with memory effects and the difficult test case of the truncated Burgers-Hopf (TBH) model. These new physics constrained quadratic multi-level regression models are proposed here as process models for Bayesian estimation through Markov Chain Monte Carlo algorithms of low frequency behavior in complex physical data.

AMS classification scheme numbers: 93E11, 62M20, 93E11, 62L12, 65C20,62J02,62M10

Submitted to: *Nonlinearity*

1. Introduction

A central issue in contemporary science is the development of data driven statistical-dynamical models for the time series of a partial subset of observed variables, $u_I(t) \in \mathbb{R}^{N_1}$, which arise from observations of nature or from an extremely complex physical model [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. This is an important issue in systems ranging from bio-molecular dynamics to climate science to engineering turbulence. Examples of such data driven dynamical models are multi-level linear autoregressive models with external factors [2, 6] as well as ad-hoc quadratic nonlinear regression models [6, 11, 12, 13, 14]

Such purely data driven ad-hoc regression models are developed through various criteria to fit the data but by design, do not respect the underlying physical dynamics of the partially observed system or the causal processes in the dynamics; nevertheless, the goal of purely data driven statistical modeling is to provide simplified low order models with high predictive skill for central features of the underlying physical system and not just fit (or over-fit, see [2]) the given data.

Indeed, Majda and Yuan [15] provide rigorous mathematical theory and examples with straightforward numerical experiments where the ad-hoc quadratic multi-level regression models proposed in [11, 12, 13] necessarily have non-physical finite-time blow up of statistical solutions and also pathological behavior of the related invariant measure even though these models match a long time series of the observed data produced from the physical model with high accuracy. The goal of the present paper is to develop new physics constrained multi-level quadratic regression models which simultaneously reflect the causality and energy conserving principles of the underlying nonlinear physics and, by design, mitigate the non-physical finite-time blow up or pathology in the invariant measure of ad-hoc quadratic regression strategies. The approach developed here builds on earlier work for single level models without memory effects which uses physical analytic properties [16, 17, 18, 19] to constrain data driven methods [9, 20, 21, 22]. The objective here is to develop theory which blends natural physical analytic constraints with the attractive memory effects of multi-level quadratic regression in a seamless fashion and then to illustrate this new approach on a suite of models. A detailed outline of the remainder of this paper is presented next.

The physics constrained quadratic multi-level regression models are introduced in Section 2 including motivation and connections with earlier work. Mathematical guidelines for the properties of these models are developed in Section 3 which ends with a brief summary of algorithms [23] for filtering and parameter estimation in order to implement the new regression strategies. Section 4 contains applications of the physics constrained regression strategies to test models involving a nonlinear oscillator with memory effects as well as the difficult test problem involving the first mode of the truncated Burgers-Hopf (TBH) model [24, 25].

2. Physics Constrained Multi-Level Quadratic Regression Models

First, consider a variable $u \in \mathbb{R}^N$ and the class of linear multi-level regression models with the form

$$\begin{aligned} \frac{du}{dt} &= Lu + F + L_{0,1} r_1, \\ \frac{dr_i}{dt} &= Q_i u + \sum_{j=1}^{i} a_{ij} r_j + r_{i+1}, \quad 1 \le i \le p-1, \\ \frac{dr_p}{dt} &= Q_p u + \sum_{j=1}^{p} a_{pj} r_j + \tilde{\sigma} \dot{W}, \end{aligned} \tag{1}$$

where $r_i \in \mathbb{R}^M$, $L_{0,1} \in \mathbb{R}^{N\times M}$ has rank M, and $\vec{r} = (r_1, \ldots, r_p)^T \in \mathbb{R}^{Mp}$, while $\tilde{\sigma}$ is an $M \times M$ matrix and $\dot{W} \in \mathbb{R}^M$ are independent white noises (physicist's notation for white noise is used but the Itô form is understood throughout this paper for all SDE's). In (1), F denotes large-scale forcing term. With concise notation (1) can be written as

$$\begin{aligned} \frac{du}{dt} &= Lu + F + L_{0,1} r_1, \\ \frac{d\vec{r}}{dt} &= Qu + A\vec{r} + \sigma \dot{W}, \end{aligned} \tag{2}$$

where $\sigma \in \mathbb{R}^{Mp\times M}$ is equals to $\tilde{\sigma}$ in its last M rows and zero everywhere else. Regarding u as the fundamental variables, there are p levels of memory associated with the models in (1). The case with no memory in the model noise, $p = 0$, involves the identification of $r_1 = \dot{W}$ and the stochastic model

$$\frac{du}{dt} = Lu + F + L_{0,1}\dot{W}. \tag{3}$$

In the development of the physics constrained multi-level quadratic regression models developed here $u = (u_I, u_{II})$ where $u_I \in \mathbb{R}^{N_1}$ is the original observed variable and $u_{II} \in \mathbb{R}^{N_2}$ is a hidden variable incorporating physics constrained primary nonlinear interaction with $N_1 + N_2 = N$. The physics constrained multi-level regression models proposed here have the same general structural form as (2) but include quadratic nonlinear interactions in u, $B(u, u)$, which conserve energy:

$$\langle u, B(u, u)\rangle = 0. \tag{4}$$

In (4), the inner product can be general but when convenient below, we utilize the standard Euclidean inner product. Let $\Pi_2(u) = (0, u_{II})^T$ denote the projection on the variable u_{II}, then the **Physics Constrained Multi-Level Regression Models** proposed here have the form

$$\begin{aligned} \frac{du}{dt} &= Lu + B(u, u) + F + \Pi_2 r_1, \\ \frac{d\vec{r}}{dt} &= Qu + A\vec{r} + \sigma \dot{W}, \end{aligned} \tag{5}$$

together with the physical constraint in (4) on the nonlinear terms, and $r_1 \in \mathbb{R}^{N_2}$, the memory dependent noise. The solutions of the linear equation for $\vec{r}$ can be written as

$$\vec{r}(t) = e^{At}\vec{r}(0) + \int_0^t e^{A(t-s)} Qu(s)ds + \int_0^t e^{A(t-s)} \sigma dW(s). \tag{6}$$

The first component, $r_1(t)$, can be substituted into the first equation in (5) to yield a memory dependent equation for u alone with correlated noise. Such a Zwanzig-Mori interpretation of the dynamics is useful below.

The special choice of $p = 0$ from (3) in (5) yields the physics constrained models without memory

$$\frac{du}{dt} = Lu + B(u, u) + F + \Pi_2 \sigma \dot{W}, \tag{7}$$

which are the starting point for developing normal forms for single level stochastic regression models [15, 20] for the variable u_I alone based on stochastic mode elimination

of the u_{II} variables [16, 17, 18, 19]; these normal form reduced models incorporate the physics constrained effects of conservation of energy as well as both additive and multiplicative noises in the dynamics of the u_I variables under the assumption of separation of time scales between the dynamics of u_I and u_{II}. The advantage of the more general data driven models with physical constraints from (4) and (5) is that they retain both the nonlinear additive and multiplicative noise effects contained in (7) and in addition, allow for p-level memory effects between u_I and u_{II}.

2.1. Comparison to Other Methods

The multi-level quadratic regression models introduced in [11] assume $u \equiv u_I$ and do not have the hidden variable u_{II} so $u_{II} = 0$. They have the structural form

$$\begin{aligned} \frac{du_I}{dt} &= Lu_I + B(u_I, u_I) + F + r_1, \\ \frac{d\vec{r}}{dt} &= Qu_I + A\vec{r} + \sigma\dot{W}, \end{aligned} \tag{8}$$

where $B(u_I, u_I)$ is a general quadratic nonlinearity which does not impose the physical constraint of energy conservation from (4) on the nonlinear terms. By substituting the integral formula for $r_1(t)$ from (6) into (8), we see that the regression models in (8) capture linear memory effects in u_I and correlated noise but cannot capture the nonlinear memory dependent noise effects present in (5) through the hidden variable u_{II}. Furthermore, the quadratic multi-level regression models from (8) which do not impose physical constraints on the nonlinear terms can fit long time series of physical data for u_I very well yet suffer from non-physical finite-time blow up for statistical solutions and pathological behavior of the invariant measure [15]. In Section 3, we provide mathematical guidelines which mitigate such pathological behavior for the physics constrained regression models in (4), (5). Note that when compared with (8), the physics constrained nonlinear regression models from (4), (5) impose an additional layer of memory through u_{II} with nonlinear interactions with u_I which satisfy the physical constraints in (4) in a natural fashion, unlike the general procedure in (8).

Finally, consider the extreme limit of the models in (8) with no memory levels but impose the physical constraint in (4); the result is the single level nonlinear regression model

$$\begin{aligned} &\frac{du_I}{dt} = Lu_I + B_I(u_I, u_I) + F + \sigma\dot{W}, \\ &\langle u_I, B_I(u_I, u_I)\rangle = 0. \end{aligned} \tag{9}$$

Such kinds of regression model have been introduced in [26]. These models in (9) can be derived as a special limiting case of the general physics constrained regression procedure introduced in (4), (5) above in a straightforward fashion. Begin with the limiting model in (7) and assume that the quadratic term, B_I, maps the subspace of u_I into u_I alone so that (4) becomes $\langle u_I, B_I(u_I, u_I)\rangle = 0$; with time scale separation in (7) and applying stochastic mode reduction [16, 17, 18, 19], we arrive at the regression form in (9).

3. Mathematical Guidelines for Physics Constrained Quadratic Regression Models

Here we develop mathematical guidelines for the physics constrained multi-level regression models introduced in (4), (5) of Section 2. First, we discuss the natural necessary and sufficient conditions for nondegenerate (non-blow up) solutions of the linear regression model in (1), which arise from ignoring the energy conserving nonlinear terms in the physics constrained model in (4), (5). Then, we demonstrate the necessity of these linear conditions for the nonlinear multi-level regression models in (4), (5) with an explicit nonlinear example. Finally, we use the geometric ergodic theory (Theorem 4.4 of Mattingly-Stuart-Higham [27]) to establish the necessary and sufficient conditions for nondegenerate solutions of the nonlinear regression model in (4), (5). We will see that the mathematical guidelines for the linear regression models in (1) are special cases of those for the nonlinear regression models in (4), (5).

3.1. Natural Mathematical Constraints on Multi-Level Linear Regression Models

Introduce the notation $x = (u, \vec{r})^T$ so that (1) or (2) becomes the linear system

$$\frac{dx}{dt} = \mathcal{L}x + \Sigma\dot{W} + F, \tag{10}$$

where the explicit definition of the matrices $\mathcal{L}, \Sigma$, is evident from (2). This is a linear system with Gaussian statistics and degenerate noise matrix N and we would like to guarantee that the statistical dynamics for (10) is stable in time and approaches a nondegenerate stationary Gaussian measure as $t \to \infty$. Necessary and sufficient conditions for this behavior are the following:

$$\begin{aligned} &\text{(i) } \textbf{Stability}\text{: All eigenvalues, } \lambda_j, \text{ of } \mathcal{L} \text{ satisfy Re } \{\lambda_j\} < 0. \\ &\text{(ii) } \textbf{Controllability:} \text{ The entire space is spanned by the set} \\ &\qquad \text{of matrices given by } \mathcal{L}^k\Sigma, k = 0, 1, \ldots \end{aligned} \tag{11}$$

Condition (ii) is the standard controllability condition for linear systems [28].

Let $p_{\mathcal{L}}(x, t)$ denote the Gaussian statistical solution of the Fokker-Planck equation associated with (10) so that $p_{\mathcal{L}}(x, t) = \mathcal{N}(\bar{x}(t), R(t))$ where $\bar{x}(t)$ is the mean and $R(t)$ is the covariance. The conditions in (11)(i), (ii) guarantee that there is a non-degenerate Gaussian invariant measure $p_{\mathcal{L},eq} = \mathcal{N}(\bar{x}_\infty, R_\infty)$ satisfying

$$\begin{aligned} &\mathcal{L}\bar{x}_\infty + F = 0, \\ &R_\infty > 0 \text{ and } \mathcal{L}R_\infty + R_\infty\mathcal{L}^T + \Sigma\Sigma^T = 0, \end{aligned} \tag{12}$$

so that the solution of (10) is **geometrically ergodic**, that is,

$$\bar{x}(t) \to \bar{x}_\infty \text{ and } R(t) \to R_\infty \text{ as } t \to \infty, \tag{13}$$

with an exponential rate of convergence.

Intuitively, we expect the linear conditions (11)(i) and (11)(ii) to be necessary for nondegenerate solutions of the nonlinear regression models in (5), (4). First, we will use

a nonlinear example to demonstrate that this intuition is indeed correct. In particular, we will show in the next section that if the stability condition in (11)(i) is violated in the mildest fashion and replaced by neutral stability of a single variable u_I, the physics constrained regression models from (4), (5) can exhibit pathological unbounded growth of the mean statistics as time evolves.

3.2. Neutrally Stable Physics Constrained Regression Models with Energy Conservation and Evolving Blow up of Mean Statistics

Consider the stochastic triad model introduced and analyzed in [29],

$$\begin{aligned} \frac{du_1}{dt} &= A_1 u_2 u_3, \\ \frac{du_2}{dt} &= A_2 u_1 u_3 - d_2 u_2 + \sigma_2 \dot{W}_2, \\ \frac{du_3}{dt} &= A_3 u_1 u_2 - d_3 u_3 + \sigma_3 \dot{W}_3. \end{aligned} \tag{14}$$

We see that the model in (14) has the quadratic regression form in (5) with $p = 0$ as presented in (7) provided we identify $u_I = u_1$ and $u_{II} = (u_2, u_3)$ in (14). The physical constraint of conservation of energy from (4) is satisfied provided that

$$A_1 + A_2 + A_3 = 0. \tag{15}$$

Strong linear stability is satisfied for u_2, u_3 but there is only neutral stability of u_1; thus, the stability condition in (11)(i) is violated in the mildest fashion. Furthermore, it is easy to check Hormander's condition [30] that the nonlinear terms in (14) have a hypoelliptic Fokker-Planck generator so that the nonlinear version of the controllability requirement in (11)(ii) is satisfied (we will discuss the nonlinear version of the controllability condition in Section 3.3). In [29] p. 209-210, elementary calculations are utilized to show that the system in (14) has a Gaussian invariant measure, p_{eq}, if and only if this measure has the form,

$$p_{eq}(u) = C \exp\Big(-\frac{1}{2}(\frac{u_1^2}{E_1} + \frac{u_2^2}{E_2} + \frac{u_3^2}{E_3})\Big), \tag{16}$$

where $E_2 = \sigma_2^2/(2d_2)$, $E_2 = \sigma_3^2/(2d_3)$. The coefficient E_1 is necessarily given in terms of the other energy parameters E_2, E_3 by

$$E_1 = -A_1 E_2 E_3 (A_2 E_3 + A_3 E_2)^{-1}. \tag{17}$$

Thus, there is a Gaussian invariant measure for (14) if and only if $E_1 > 0$. Now, it is easy to arrange the values E_2, E_3 with

$$A_1, A_2 < 0, A_3 = -(A_1 + A_2) > 0, \text{ and } A_2 E_3 + A_3 E_2 < 0 \tag{18}$$

or equivalently

$$A_3 < \frac{E_3}{E_2}|A_2|.$$

This automatically guarantees that $E_1 < 0$ in (14) for an energy conserving nonlinearity so there is no Gaussian invariant measure as required in (16). Simple numerical

experiments with (14) in regimes with (18) satisfied and reported in Fig. 5 from [29] show that the mean of a statistical solution of (14) becomes unbounded as time evolves and clearly has pathological dynamical behavior. This example illustrates the necessity of the strict linear stability condition in (11)(i) in mitigating unphysical behavior for (4), (5) as quadratic multi-level regression models. Even the mildest violation of (11)(i) through neutral stability can lead to pathological behavior in the dynamics.

3.3. Rigorous Nondegenerate Conditions for the Physics-Constrained Nonlinear Regression Models

Our goal here is to establish rigorous mathematical conditions to avoid pathological non-physical blow up behavior for (4), (5) as quadratic multi-level regression models. In particular, we are looking for necessary and sufficient conditions for ergodicity of the dynamical systems in (4), (5); a generalization of the linear conditions in (11). This problem, especially for systems of differential equations with degenerate noise and locally Lipcshitz vector fields, has been studied in detail by Mattingly-Stuart-Higham [27]. For our specific need, we will use the ergodic theorem 4.4 of Mattingly-Stuart-Higham [27] to determine the nondegenerate conditions for the physics-constrained nonlinear regression models in (4), (5). For a self-contained presentation, we restate the geometric ergodic theorem 4.4 of [27] in accordance with what we need below (see [27] for the detailed version as well as the proof).

Geometric Ergodicity theorem. (see [27] for the proof) *Consider the following system of SDE,*

$$dx = f(x)dt + \Sigma dW, \quad x(0) = x_0. \tag{19}$$

where $x \in \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}^d$, $\Sigma \in \mathbb{R}^{d\times m}$ *with* $m \leq d$, *and* W *is a standard* m*-dimensional Brownian motion. Without loss of generality, let us define* $\Sigma = [\rho_1, \ldots, \rho_m]$, *where* $\{\rho_j \in \mathbb{R}^d, j = 1, \ldots, m\}$ *are linearly independent. Assume that* x *can be characterized with transition kernel*

$$P_t(x, A) \equiv \mathbb{P}(x(t) \in A | x(0) = x),$$

that satisfies the following condition:

(A) Hypoellipticity (or Hormander's condition [30]): Given a fixed compact set $C \in \mathcal{B}(\mathbb{R}^d)$, $P_t(x, A)$ *has a density* $p_t(x, y)$ *defined as*

$$P_t(x, A) = \int_A p_t(x, y)dy, \quad \forall x \in C,\ A \in \mathcal{B}(\mathbb{R}^d) \cap \mathcal{B}(C),$$

and $p_t(x, y)$ *is jointly continuous at* $(x, y) \in C \times C$.

Also, assume also that (19) *satisfies the following:*

(B) The deterministic dynamical system without noise, $\Sigma = 0$, *in* (19) *has a global Lyapunov function,* $V(x) = \frac{1}{2}\|x\|^2$, *i.e., there exists* $\beta > 0$ *and an inner product* $\langle \cdot, \cdot \rangle$, *such that* $\langle f(x), x\rangle \leq -\beta\|x\|^2$.

Then (19) *has a unique invariant measure,* π. *Furthermore, for* $g \in \mathcal{G}_\ell \equiv \{$*measurable* $g : \mathbb{R}^d \to \mathbb{R}$ *with* $|g(x)| \leq 1 + \|x\|^{2\ell}\}$, $\mathbb{E}_{x_0} g(x(t)) \longrightarrow \pi(g)$ *as* $t \to \infty$ *with an exponential rate of convergence.*

In compact notation, we can rewrite the unforced ($F = 0$) nonlinear multi-level model in (5) as

$$\frac{dx}{dt} = \mathcal{L}x + \mathcal{N}(x) + \Sigma \dot{W}, \tag{20}$$

where, $x = (u, \vec{r}) \in \mathbb{R}^d$, $d = N + Mp$, $\vec{r} = (r_1, \ldots, r_p)$, $r_i \in \mathbb{R}^M$.

Assume as in (i) from (11) that all eigenvalues of $\mathcal{L}$ have negative real parts; then it is well known (see Chapter 7 of [31]) that there is an inner product $\langle \cdot, \cdot \rangle$ and $\beta > 0$ such that

$$\langle x, \mathcal{L}x \rangle = \langle x, \frac{\mathcal{L} + \mathcal{L}^T}{2} x \rangle \leq -\beta \|x\|^2$$

with $\mathcal{L}^T$ the transpose matrix in this inner product. We require that the quadratic terms in (4) conserve energy in this inner product, i.e.,

$$\langle u, B(u, u) \rangle = 0. \tag{21}$$

Under these assumptions, the condition required in (B) above is immediate since by using (21),

$$\frac{dV}{dt} = \langle x, f(x) \rangle = \langle x, \mathcal{L}x \rangle \leq -\beta \|x\|^2.$$

This condition immediately implies non-blowup for the second moments of the SDE without further assumptions.

To verify the smoothness of the density $p_t(x, y)$ or the hypoellipticity condition in (A), we define Lie algebra $Lie(\rho_0, \rho_1, \ldots, \rho_m)$ where $\rho_0 \equiv f$ with Lie bracket $[\rho_i, \rho_j] = (\partial_x \rho_j)\rho_i - \rho_i(\partial_x \rho_j)$. Based on the results by Kunita [32], the smoothness of density p_t is satisfied if the ideal generated by $\{\rho_1, \ldots, \rho_m\}$ in $Lie(\rho_0, \rho_1, \ldots, \rho_m)$ spans d-dimensional space. In other words, one simply needs to check whether the following operators

$$\begin{array}{ll} \rho_j, & j = 1, \ldots, M, \\ [\rho_j, \rho_k], & j = 0, 1, \ldots, M, \\ [\rho_j, [\rho_k, \rho_\ell]], & j = 0, 1, \ldots, M, \\ \vdots & \end{array}$$

span $\mathbb{R}^d$. Notice that $\rho_0 \equiv f$ is only available in the brackets. Note that if (19) is linear, then this smoothness condition is nothing else but the linear controllability condition in (11)(ii). For the general nonlinear multi-level regression model in (20), verification of the smoothness condition involves cumbersome algebraic expressions in the higher order brackets. To summarize, our mathematical guideline for nondegenerate solutions of the physics-constrained nonlinear regression models can be formally stated as follows:

Theorem. *Consider the centered physics-constrained nonlinear regression models in* (4), (5) *with* $F = 0$. *If the Fokker-Planck generator of this system of SDE's is hypoelliptic (that is, condition (A) is satisfied) and all eigenvalues of* $\mathcal{L}$ *in* (20) *have negative real part, then solutions of* (4), (5) *are geometrically ergodic provided energy conservation in* (4) *is imposed in an appropriate inner product as in* (21).

On the other hand, the ad-hoc quadratic regression models proposed in [11] can satisfy linear stability at the origin and still exhibit pathological finite-time blow up of statistical solutions since condition (B) above is not satisfied with general nonlinear quadratic terms (that is, $B(x, x)$ that does not satisfy (4)). This fact was shown rigorously and numerically in [15].

3.4. Mathematical Guidelines and Filtering Algorithms for the Physics Constrained Regression Models

In practical implementation of the multi-level regression models in (4) and (5), we have to estimate the primary variables, $x = (u_I, u_{II}, \vec{r})^T$, from partial observations of only the variable u_I as well as estimate the regression parameter coefficients in the model. As in many practical settings, we assume these parameters are constants even if the hidden true parameters (slowly) change in time since we don't know them. A standard strategy to achieve all of this parameter estimation is to use an extended Kalman filter algorithm as utilized for empirical-dynamical quadratic regression in [23]; the algorithm utilizes the parameter estimation scheme developed in [33] and the residual noise method developed in [34]. This is the algorithm which we utilize in the numerical experiments presented in Section 4 and the reader is referred to [23, 33, 34] for algorithmic details. Let $\Pi_1(x)$ be defined by $\Pi_1(x) = u_I$; then, the natural additional mathematical requirement to impose on the multi-level regression model in (4) and (5) is that the partial observation $\Pi_1(x)$ is observable [35, 36] for the linear operator in (10) which arises from dropping the nonlinear terms in (5). With the notation in (10) for this linear operator, the observability condition for the primary variables x requires:

$$\begin{aligned}&\text{The entire space is spanned by the set of matrices}\\&\text{given by } (\mathcal{L}^k)^T \Pi_1^T, k = 0, 1, \ldots\end{aligned} \tag{22}$$

If the dynamics in (5) is strongly nonlinear so that B in (5) is large, there are important caveats and perhaps limited skill in using an extended Kalman filter which is based on successive linearization for both the dynamics for the primary variables and the parameters [37]. A more flexible algorithm for filtering and parameter estimation in the present context is based on finite ensemble filters [38] but this approach is deferred to future applications.

4. Applying the Physics Constrained Regression Algorithms for Data Driven Models

Here we apply the physics constrained regression strategies from (4), (5) with the EKF algorithm in [23, 33, 34] (as we mentioned in Section 3.3) to test models involving a nonlinear oscillator with memory effects as well as the difficult test problem involving observations of the first mode of the TBH model [19, 25] without any direct knowledge of the dynamics which created this time series.

4.1. Nonlinear Oscillator with Memory as a Test Model

Here, we test the ability of the physics constrained quadratic multi-level regression strategies in (4), (5) with the numerical filtering algorithm mentioned in Section 3.3 to recover the statistics of a time series in a nonlinear test problem. The starting point is the nonlinear oscillator

$$\begin{aligned} \frac{du_1}{dt} &= -\,\mathrm{i}a u_1^* u_2^*, \\ \frac{du_2}{dt} &= \mathrm{i}a(u_1^*)^2, \end{aligned} \tag{23}$$

for complex valued components u_1, u_2 so that (23) is a four-dimensional dynamical system. It is easy to see directly that the nonlinear terms in (23) conserve energy for any complex valued number a so that (4) is satisfied; furthermore, (23) is completely integrable and its dynamics is equivalent to the dynamics of a constant phase shift in time and the integrable behavior of a particle in a quartic potential [24] so (23) is essentially a nonlinear oscillator; moreover, (23) arises with a special choice of the coefficient, a, from Galerkin truncation of the TBH model to the first two Fourier modes [24]. The test model considered here consists of adding stable linear couplings plus a level of colored noise to the model in (23) so that the test model is essentially a nonlinear oscillator with memory with the form,

$$\begin{aligned} \frac{du_1}{dt} &= -\,\mathrm{i}a u_1^* u_2^* + a_{11}u_1 + a_{12}u_2, \\ \frac{du_2}{dt} &= \mathrm{i}a(u_1^*)^2 + a_{21}u_1 + a_{22}u_2 + r, \\ \frac{dr}{dt} &= \alpha_1 u_1 + \alpha_2 u_2 + \beta_1 r + \sigma \dot{W}, \end{aligned} \tag{24}$$

With a slight change of notation, the test model in (24) has the structure of the normal form for the physics constrained quadratic multi-level regression model in (4), (5) with $p = 1$ with coefficients in this perfect model satisfying the linear stability, controllability, and observability conditions in (11) and (22). Here we take a long time series generated by the perfect model in (24) with truth parameters, $\theta = \{a, a_{11}, a_{12}, a_{21}, a_{22}, \alpha_1, \alpha_2, \beta_1, \sigma\}$, specified in Table 1 and ask whether the regression models in (4), (5) together with the filtering algorithm in 3.3 can produce a new physics constrained regression model with the same structural form as in (24), perhaps with

different coefficients, $\hat{\theta}$, so that the statistical behavior of $u_I = u_1$ in the nonlinear regression model reproduces that in the original perfect model. Mathematically, given any integrable function $g(u_1)$, we want to verify whether,

$$\int g(u_1)\pi(dx|\hat{\theta}) \approx \int g(u_1)\pi(dx|\theta), \qquad x \equiv (u_1, u_2, r) \tag{25}$$

In (25), $\pi(\cdot|\theta)$ and $\pi(\cdot|\hat{\theta})$ denote the equilibrium measures of (24), conditional to the true, θ, and estimated, $\hat{\theta}$, parameters, respectively. The condition in (25) is only a minimal requirement for the regression and in addition one would also like to match unequal time statistics [19, 29] such as the autocorrelation function, $R(\tau) = \langle u_1(t+\tau)u_1(t)\rangle$. Note that determining this regression model from filtering is an underdetermined nonlinear procedure when given only the time series of u_1 alone and one cannot expect exact recovery of the coefficients of the dynamics of the original model in (24). Such nonuniqueness occurs even for linear regression models and explicit examples are given in [15]. This is a stringent test model for the capabilities of the physics constrained regression procedure.

Here and below, we pick the observational noise variance to be 10% of the observed variance of u_1 from the perfect model and always utilize the fixed time step $T_{obs} = 0.01$ between observations of u_1. These same values are also utilized for the TBH model results reported later in this section.

The perfect model parameters, θ, and estimated parameters, $\hat{\theta}$, from multi-level regression (MLR) for the first test case are reported in Table 1. Figure 1 compares the perfect and estimated probability distribution function (pdf) and autocorrelation for u_1 produced by the perfect model and the MLR algorithm. Here the MLR algorithm is highly skillful with true correlation times given by 0.53 and 0.54 for the real and imaginary parts of the perfect model compared with 0.52 for both components from the MLR model; the true variance for u_1 (0.0029) is also estimated very well by the MLR algorithm (0.0032) as is evident in Figure 1. Note from Table 1 that both the perfect model and the MLR model have nonlinear coefficients of non-trivial and comparable magnitude so that the physics constrained MLR algorithm is fully nonlinear here. For completeness, in Figure 2 we report the convergence history of the estimated parameters in the MLR algorithm and note the excellent rapid convergence of all parameters to roughly constant values; this suggests that a relatively short time series of data is required here. The parameters associated with the deterministic operator in (24), $\{a, a_{ij}, \alpha_j, \beta_1\}$ are estimated with extended Kalman filter with stationary parameter models [33]. To estimate the noise parameter σ, we define $q = \sigma^2 T_{obs}/2$ (see Figure 1 and Table 1); $q > 0$ is estimated with a secondary extended Kalman filter algorithm with stationary prior model for q and observation residual square observation model; this algorithm was introduced in [34].

In the second case, we increase the value, a, of the nonlinear coupling coefficient in the perfect model with random choices of similar magnitude as in the first test case for the remaining parameters; thus, the perfect model is more nonlinear in this second test with non-Gaussian pdf (with skewness 0.69 for the real and -0.35 for the imaginary

components and excess kurtosis of roughly 1.09). The true and estimated parameters from MLR, which also infer stronger nonlinear dynamics, are given in Table 2; the true and the estimated statistics for u_1 are presented in Figure 3. In this highly nonlinear test case, the non-Gaussian shape of the pdf's as well as the variance of the pdf's in the true model (0.0034) is estimated very well by the MLR algorithm (0.0039) despite both being strongly nonlinear and non-Gaussian. By sight, the autocorrelations are also well captured in Figure 3 although there are slight discrepancies with perfect model autocorrelation time, 0.55, and the MLR estimate, 0.60. One possible source of discrepancies in this stronger nonlinear regime of the perfect model is the use of linear tangent models in the extended Kalman filter algorithm [37] and this needs careful additional study beyond the scope of the present paper.

4.2. The Nonlinear MLR Algorithm Applied to Time Series from TBH

With the confidence gained in applying the MLR algorithm to nonlinear dynamics in Section 4.1, here we apply the MLR algorithm to a difficult test case involving observation of the first Fourier component of the TBH model. The TBH model [19, 24, 25] is described by the following quadratically nonlinear equation for complex Fourier coefficients, u_k, $1 \leq |k| \leq 50$ with $u_{-k} = u_k^*$,

$$\frac{du_k}{dt} = -\frac{\mathrm{i}k}{2} \sum_{\substack{k+p+q=0 \\ 1 \leq |k| \leq 50}} u_p^* u_q^*. \tag{26}$$

It is a striking example of intrinsic stochastic chaotic dynamics arising in a large deterministic system with 100 variables with a number of remarkable statistical properties documented elsewhere [19, 24, 25]. The first Fourier mode has the longest autocorrelation time and thus the largest statistical memory of the dynamics in (26). The Galerkin truncation of (26) to the first two Fourier modes yields the nonlinear oscillator model in (23) with a specific choice of coefficient, a.

Here we have a difficult application of the physics constrained MLR algorithms from (4), (5): We attempt to recover the statistics of u_1 through the physics constrained MLR algorithm with the structural form in (24) without any further knowledge of the detailed dynamics of TBH. In particular, we parameterize the nonlinear oscillator with memory in (24) with the following observations,

$$v_m = u_1(t_m) + \sigma_m^o, \quad \sigma_m \in \mathcal{N}(0, r^o),$$

where $u_1(t_m)$ are solutions of the TBH model (26). The discrete-time observation v_m are taken at every $t_m = mT_{obs}$, where $T_{obs} = 0.01$ and corrupted by a complex valued Gaussian noise with variance r^o. As in the previous section, we set r^o to be 10% of the empirically estimated variance of u_1. By complex valued noise, we mean $\sigma_m = (\sigma_m^1 + \sigma_m^2)/\sqrt{2}$, where σ_m^j are the standard real-valued white noise with variance r^o. Our goal is to recover the marginal statistics (or/and distribution) of the true u_1 with long time integration of the model in (24). Mathematically, given any integrable

function $g(u_1)$, we would like to verify whether

$$\int g(u_1)\pi(dx|\hat{\theta}) \approx \langle g \rangle. \tag{27}$$

In (27), $\pi(\cdot|\hat{\theta})$ denotes the equilibrium measure of (24), conditional to the estimated parameters $\hat{\theta}$, whereas $\langle g \rangle$ denotes the observed empirically estimated equilibrium statistics from solutions of the TBH model in (26). Also, we would like to check the skill of the regression model in recovering the autocorrelation in time as mentioned below (25). Compared to (25), there is an additional degree of difficulty due to modeling errors with unknown truth equilibrium measure.

The results from the MLR algorithm with (24) for the estimated parameters are given in Table 3 and the statistics of u_1 from TBH and the MLR algorithm are compared in Figure 4. The values of the coefficients in Table 3 satisfy the linear stability condition in (11). The pdf of u_1 is fit quite well with a slight underestimation of the variance; the autocorrelation function is fit perfectly for all lags smaller than one and larger than three but the peculiar positive bulge in the TBH autocorrelation function for u_1 for $1 \leq t \leq 3$ is not captured by the MLR algorithm. From Table 3, the estimated coefficient, a, of the nonlinear oscillator is nearly zero indicating that linear regression models with three-level memory ($p = 1$) are excellent models for the stochastic dynamics of the first mode of TBH. In fact, much higher level stable linear regression models for the first mode u_1 of TBH can match the autocorrelation and variance of TBH exactly [39]. However, a four level nonlinear or linear regression model ($p = 2$) with the same physics constrained structural form as in (24) yields essentially the same performance of the MLR model as in Figure 4 while a two-level model with $p = 0$ exhibits large statistical discrepancies in u_1. These results of statistical linearity of the dynamics of the first mode of TBH predicted from the physics constrained nonlinear MLR algorithm with the normal form in (24) are foreshadowed by earlier work [25]. In [25], a sequence of modified TBH models defined by a parameter ϵ with increasing scale separation as ϵ decreases for $\epsilon \ll 1$ and agreeing with the TBH model for $\epsilon = 1$ were studied through deterministic mode elimination; the predicted stochastic models for u_1 in that context are linear Langevin equations, i.e., linear stochastic models without memory, which fit the autocorrelation and variance for u_1, in the range, $0 < \epsilon \lesssim 0.5$. Thus, the nearly linear three-level stochastic model produced by the MLR algorithm is consistent with this limiting behavior.

5. Concluding Discussion

New physics constrained multi-level quadratic regression models with the advantages of incorporating the nonlinear noise from energy conserving nonlinear dynamics as well as memory effects in time are introduced here in Section 2 and compared with existing methods in the literature. Mathematical guidelines for nondegenerate solutions of these physics constrained multi-level quadratic stochastic models as well as MLR filtering algorithms are presented in Section 3. An interesting direction for further

mathematical research is to extend the rigorous non-blowup theorem in Section 3.3 to more general non-zero forcing and verify the hypoellipticity conditions. Applications of the physics constrained MLR algorithms to a test model involving a nonlinear oscillator with memory and the difficult test case of the TBH model were developed in Section 4 with encouraging results reported there. Improvements in the MLR filtering algorithms through finite ensemble Kalman filters are an interesting future project as mentioned earlier.

We end our discussion by proposing the use of the physics constrained multi-level regression models as process models for use with full Bayesian estimation of parameters through Markov Chain Monte Carlo algorithms [6, 14]. These new nonlinear regression models can be applied to low frequency time series for a variety of physical systems. The authors plan to report on research addressing all of these issues with various collaborators in the near future.

Acknowledgments

We thank the anonymous reviewer for pointing out the geometric ergodic theory developed in [27]. The research of A.J.M. is partially supported by the National Science Foundation grant DMS-0456713 and the Office of Naval Research grants DRI N00014-10-1-0554 and N00014-11-1-0306. The research of J.H. is partially supported by the Office of Naval Research Grant N00014-11-1-0310, the NC State startup fund, and the NC State Faculty Research and Professional Development fund. Both authors are also partially supported by the Office of Naval Research Grant MURI N00014-12-1-0912.

References

[1] I. Horenko. Finite element approach to clustering of multidimensional time series. *SIAM J. Sci. Comp.*, 32(1):68–83, 2010.

[2] I. Horenko. On the identification of nonstationary factor models and their application to atmospheric data analysis. *Journal of the Atmospheric Sciences*, 67(5):1559–1574, 2010.

[3] D. Crommelin and E. Vanden-Eijnden. Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, 65:2661–2675, 2008.

[4] M.B. Priestly. *Non-linear and non-stationary time series analysis.* Academic Press, London, UK, 1988.

[5] H. Tong. *Nonlinear time series: A dynamical systems approach.* Oxford University Press, Oxford, UK, 1990.

[6] N. Cressie and C.K. Wikle. *Statistics for spatio-temporal data.* Wiley, New Jersey, 2011.

[7] G.A. Pavliotis and A.M. Stuart. Parameter estimation for multiscale diffusions. *J. Stat. Phys.*, 127(4):741–781, 2007.

[8] R. Azencott, A. Beri, and I. Timofeyev. Adaptive sub-sampling for parametric estimation of gaussian diffusions. *Journal of Statistical Physics*, 139:1066–1089, 2010. 10.1007/s10955-010-9975-y.

[9] A.J. Majda, C.L. Franzke, A. Fischer, and D.T. Crommelin. Distinct metastable atmospheric regimes despite nearly gaussian statistics: A paradigm model. *Proceedings of the National Academy of Sciences*, 103(22):8309–8314, 2006.

[10] C. Franzke, I. Horenko, A.J. Majda, and R. Klein. Systematic metastable atmospheric regime identification in an agcm. *Journal of the Atmospheric Sciences*, 66(7):1997–2012, 2009.

[11] S. Kravtsov, D. Kondrashov, and M. Ghil. Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability. *Journal of Climate*, 18(21):4404–4424, 2005.

[12] D. Kondrashov, S. Kravtsov, A. W. Robertson, and M. Ghil. A hierarchy of data-based enso models. *Journal of Climate*, 18(21):4425–4444, 2005.

[13] D. Kondrashov, S. Kravtsov, and M. Ghil. Empirical mode reduction in a model of extratropical low-frequency variability. *Journal of the Atmospheric Sciences*, 63(7):1859–1877, 2006.

[14] C. Wikle and M. Hooten. A general science-based framework for dynamical spatio-temporal models. *TEST*, 19:417–451, 2010. 10.1007/s11749-010-0209-z.

[15] A.J. Majda and Y. Yuan. Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems. *Discrete and Continuous Dynamical Systems B*, 17(4):1333–1363, 2012.

[16] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. Models for stochastic climate prediction. *Proc. Nat. Acad. Sci.*, 96:15687–15691, 1999.

[17] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. A mathematical framework for stochastic climate models. *Comm. Pure Appl. Math.*, 54:891–974, 2001.

[18] A.J. Majda, C. Franzke, and B. Khouider. An Applied Mathematics Perspective on Stochastic Modelling for Climate. *Philos Transact A Math Phys Eng Sci.*, 366(1875):2429–2455, 2008.

[19] 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.

[20] A.J. Majda, C. Franzke, and D. Crommelin. Normal forms for reduced stochastic climate models. *Proceedings of the National Academy of Sciences*, 106(10):3649–3653, 2009.

[21] C. Franzke, A.J. Majda, and E. Vanden-Eijnden. Low-order stochastic mode reduction for a realistic barotropic model climate. *Journal of the Atmospheric Sciences*, 62(6):1722–1745, 2005.

[22] Y. Yuan and A.J. Majda. Invariant measures and asymptotic gaussian bounds for normal forms of stochastic climate model. *Chinese Annals of Mathematics - Series B*, 32:343–368, 2011. 10.1007/s11401-011-0647-2.

[23] K. Strounine, S. Kravtsov, D. Kondrashov, and M. Ghil. Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance. *Physica D: Nonlinear Phenomena*, 239(34):145 – 166, 2010.

[24] A.J. Majda and I. Timofeyev. Statistical mechanics for truncations of the Burgers-Hopf equation: A model for intrinsic behavior with scaling. *Milan Journal of Mathematics*, 70:39–96, 2002.

[25] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. Stochastic models for selected slow variables in large deterministic systems. *Nonlinearity*, 19:769–794, 2006.

[26] F. Kwasniok. Reduced atmospheric models using dynamically motivated basis functions. *Journal of the Atmospheric Sciences*, 64(10):3452–3474, 2007.

[27] J.C. Mattingly, A.M. Stuart, and D.J. Higham. Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise. *Stochastic Processes and their Applications*, 101(2):185 – 232, 2002.

[28] J. Macki and A. Strauss. *Introduction to Optimal Control Theory*. Undergraduate Texts in Mathematics. Springer, 1981.

[29] A.J. Majda, I. Timofeyev, and E. Vanden-Eijnden. A priori tests of a stochastic mode reduction strategy. *Physica D*, 170:206–252, 2002.

[30] L. Hormander. *The analysis of linear partial differential operators III: Pseudo-Differential operators*. Number v. 3 in Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen. Springer, 1985.

[31] M. Hirsch and S. Smale. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press (New York), 1974.

[32] H. Kunita. Supports of diffusion processes and controllability problems. In Itô, editor, *Proceedings*

of the International Symposium on Stochastic Differential Equations, Kyoto University, pages 163–185. John Wiley & Sons, 1978.

[33] B. Friedland. Treatment of bias in recursive filtering. *IEEE Trans. Automat. Contr.*, AC-14:359–367, 1969.

[34] P.R. Belanger. Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, 10(3):267 – 275, 1974.

[35] B.D. Anderson and J.B. Moore. *Optimal filtering*. Prentice-Hall Englewood Cliffs, NJ, 1979.

[36] C. Chui and G. Chen. *Kalman filtering: with Real-Time Applications*. Springer New York, 1999.

[37] M. Branicki, B. Gershgorin, and A.J. Majda. Filtering skill for turbulent signals for a suite of nonlinear and linear extended kalman filters. *Journal of Computational Physics*, 231(4):1462 – 1498, 2012.

[38] S.-J. Baek, B.R. Hunt, E. Kalnay, E. Ott, and I. Szunyogh. Local ensemble Kalman filtering in the presence of model bias. *Tellus A*, 58(3):293–306, 2006.

[39] E.L. Kang, J. Harlim, and A.J. Majda. Regression models with memory for the linear response of turbulent dynamical systems. *Comm. Math. Sci. (in press)*, 2012.

Table 1. Weakly nonlinear regime: The truth and estimated parameters of the test model in (24). The estimated parameters are obtained from filtering noisy time series of the observed component, u_1, at discrete time step $T_{obs} = 0.01$.

parameter	truth (θ)	MLR estimate ($\hat{\theta}$)
a	$0.1 - 0.3\mathrm{i}$	$0.1319 - 0.0556\mathrm{i}$
a_{11}	$-3 + 0.75\mathrm{i}$	$-2.0570 + 1.3106\mathrm{i}$
a_{12}	$0.25 - 0.25\mathrm{i}$	$-0.0459 - 0.1040\mathrm{i}$
a_{21}	$0.2 + 0.2\mathrm{i}$	$-0.4809 + 0.9616\mathrm{i}$
a_{22}	$-2.4 - 1.6\mathrm{i}$	$-1.4187 - 2.0819\mathrm{i}$
α_1	$-2.1 + 2.6\mathrm{i}$	$-1.9971 + 2.9933\mathrm{i}$
α_2	$-2.7 - 0.9\mathrm{i}$	$-2.2236 + 1.0463\mathrm{i}$
β_1	$-1.4 + 1.9\mathrm{i}$	$-2.2975 + 2.1854\mathrm{i}$
$q = \sigma^2 T_{obs}/2$	0.1	0.7304

Table 2. Strongly nonlinear regime: The truth and estimated parameters of the test model in (24). The estimated parameters are obtained from filtering noisy time series of the observed component, u_1, at discrete time step $T_{obs} = 0.01$.

parameter	truth (θ)	estimate ($\hat{\theta}$)
a	$1 + \mathrm{i}$	$-0.1066 - 0.1324\mathrm{i}$
a_{11}	$-3 + 0.75\mathrm{i}$	$-2.2683 + 0.6771\mathrm{i}$
a_{12}	$0.25 - 0.25\mathrm{i}$	$-0.0306 - 0.0246\mathrm{i}$
a_{21}	$0.2 + 0.2\mathrm{i}$	$-0.4810 + 0.9871\mathrm{i}$
a_{22}	$-2.4 - 1.6\mathrm{i}$	$-1.5916 - 1.8842\mathrm{i}$
α_1	$-2.1 + 2.6\mathrm{i}$	$-1.9989 + 2.9953\mathrm{i}$
α_2	$-2.7 - 0.9\mathrm{i}$	$-2.2871 + 1.1881\mathrm{i}$
β_1	$-1.4 + 1.9\mathrm{i}$	$-1.9599 + 2.2733\mathrm{i}$
$q = \sigma^2 T_{obs}/2$	0.1	17.4107

Table 3. The estimated parameters for the three-level MLR model in (24) obtained from filtering noisy time series of the first component, u_1, of the TBH model at discrete time step $T_{obs} = 0.01$.

parameter	MLR estimate ($\hat{\theta}$)
a	$0.0012 + 0.0013\mathrm{i}$
a_{11}	$-3.0507 + 0.7685\mathrm{i}$
a_{12}	$-0.2321 - 0.2636\mathrm{i}$
a_{21}	$0.2118 + 0.2285\mathrm{i}$
a_{22}	$-2.3645 - 1.5840\mathrm{i}$
α_1	$-2.1186 + 2.6744\mathrm{i}$
α_2	$-2.7059 - 0.8956\mathrm{i}$
β_1	$-1.3976 + 1.8711\mathrm{i}$
$q = \sigma^2 T_{obs}/2$	4.39

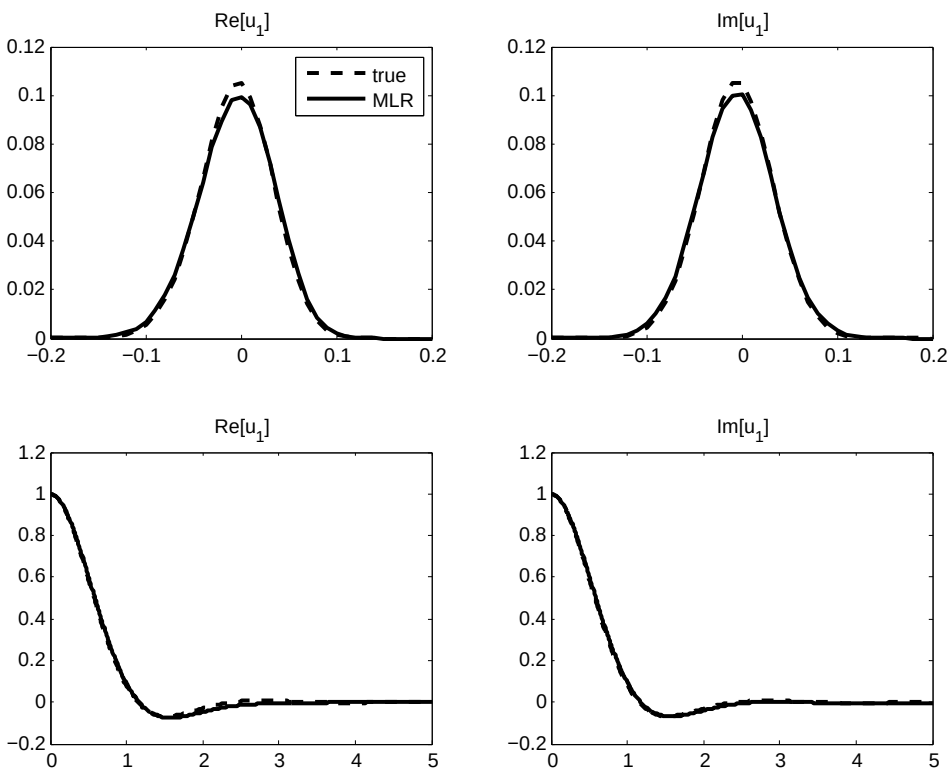

Figure 1. The test model in a weakly nonlinear regime: Marginal distributions of real and imaginary parts (top panels) and autocorrelation functions (bottom panels) of the first component, u_1, of the truth and MLR estimate (24).

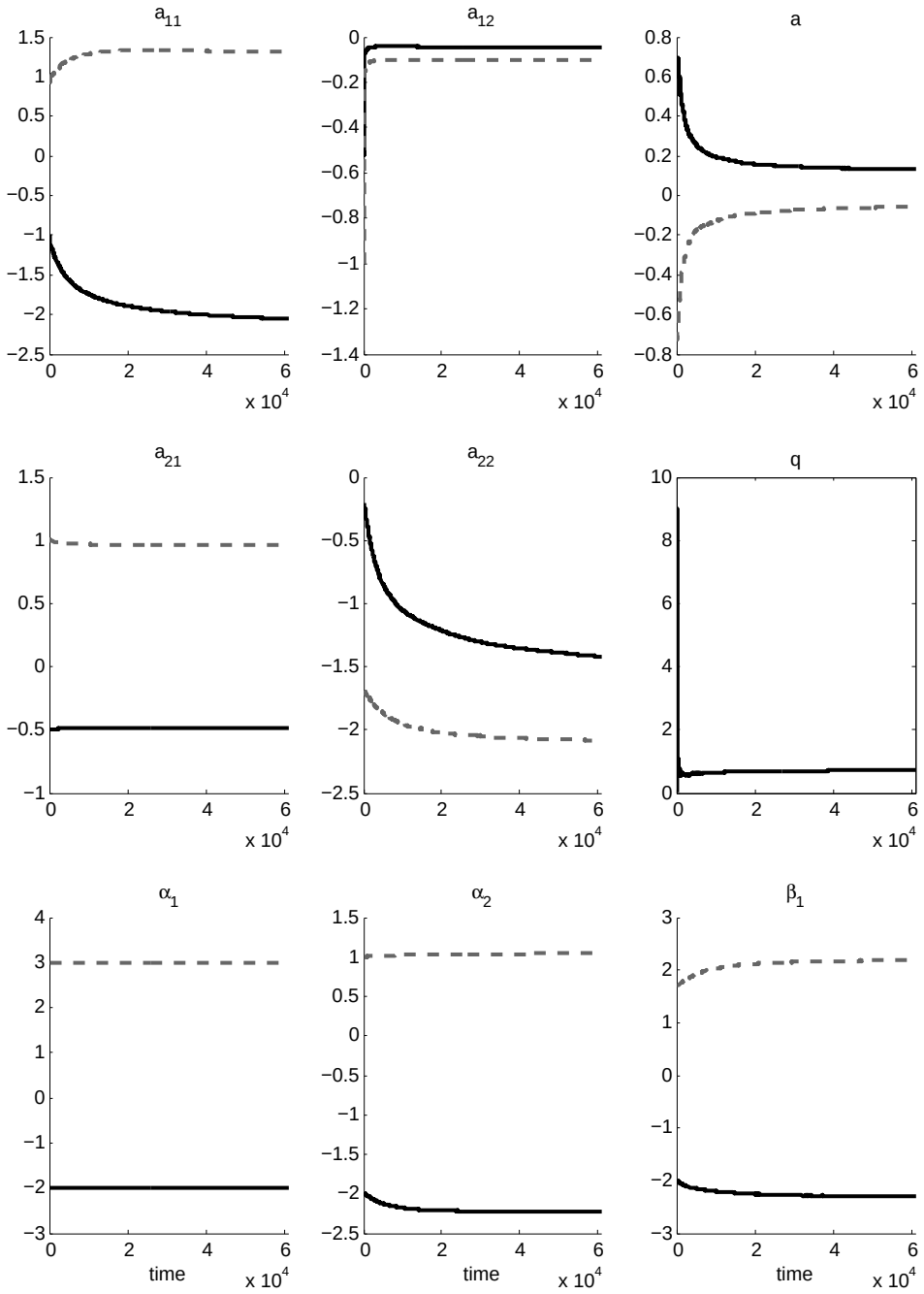

Figure 2. The test model in a weakly nonlinear regime: Posterior parameter estimates as functions of time. In each panel the solid line denotes the real component and the grey dashes denote the imaginary component.

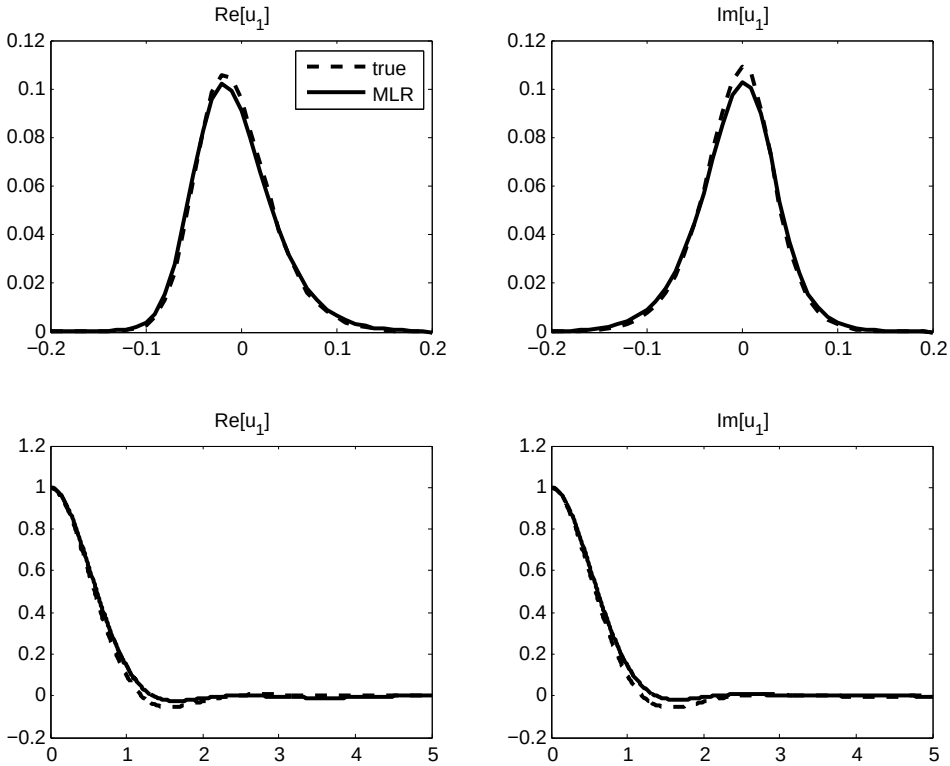

Figure 3. The test model in a strongly nonlinear regime: Marginal distributions of real and imaginary parts (top panels) and autocorrelation functions (bottom panels) of the first component, u_1, of the truth and MLR estimate (24).

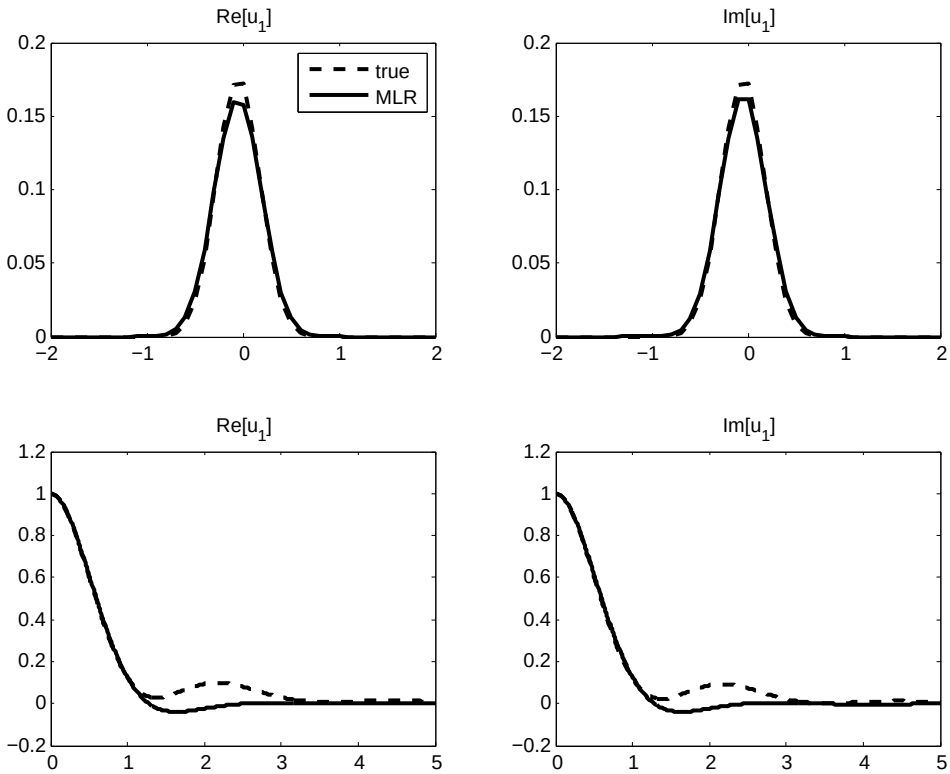

Figure 4. The TBH time series: Marginal distributions of real and imaginary parts (top panels) and autocorrelation functions (bottom panels) of the first component, u_1, of the truth and MLR estimate (24).