

# Reconstruction of stochastic nonlinear dynamical models from trajectory measurements

D.G. Luchinsky<sup>a,b</sup>, V.N. Smelyanskiy<sup>a</sup>, M. Millonas<sup>a</sup>, P.V.E. McClintock<sup>a</sup>

<sup>a</sup>NASA Ames Research Center, Mail Stop 269-2, Moffett Field, CA 94035, USA

<sup>b</sup> Department of Physics, Lancaster University, Lancaster LA1 4YB, UK

## ABSTRACT

We consider the following general problem of applied stochastic nonlinear dynamics (see e.g.<sup>1</sup>). We observe a time series of signals  $\mathbf{y}(t) = \mathbf{y}(t_0 + hn)$  corrupted by noise. The actual state and the nonlinear vector field of the dynamical system is not known. The question is how and with what accuracy can we determine  $\mathbf{x}(t)$  and functional form of  $\mathbf{f}(\mathbf{x})$ . In this talk we discuss a novel approach to the solution of this problem based on the application of the path-integral approach to the full Bayesian inference. We demonstrate a reconstruction of a dynamical state of a system from corrupted by noise measurements. Next we reconstruct the corresponding nonlinear vector field. The emphasis are on the theoretical analysis. The results are compared with the results of earlier research.

**Keywords:** Bayesian inference, nonlinear model reconstruction

## 1. INTRODUCTION

We consider the following typical experimental situation. A corrupted by noise measurement  $\mathbf{y}(t) = \mathbf{y}(t_0 + hn)$  represents time evolution of an unknown dynamical system measured at each  $h$  seconds starting at some time  $t_0$ . In a particular case of a continuous dynamical system this situation can be usually modelled as follows

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}|\mathbf{c}) + \sqrt{\hat{\mathbf{D}}}\xi(t), \quad (1)$$

$$\mathbf{y} = \hat{\mathbf{\Gamma}}\mathbf{x} + \sqrt{\hat{\mathbf{M}}}\eta(t), \quad (2)$$

where  $\mathbf{x} \in R^L$  is an unknown  $L$ -dimensional state vector of a dynamical system, which continuous stochastic dynamics is modelled by equation (1) represent with essentially nonlinear unknown vector field  $\mathbf{f}(\mathbf{x}|\mathbf{c})$ . The system is driven by white Gaussian noise  $\xi$  with diffusion matrix  $\hat{\mathbf{D}}$ . The second equation (2) represents the model of measurements with white Gaussian noise  $\eta(t)$  mixed by the measurement noise  $\hat{\mathbf{M}}$ . The vector of measurements  $\mathbf{y} \in R^K$  has in general dimension  $K < L$  and therefore  $\hat{\mathbf{\Gamma}}$  is  $K \times L$  the measurement matrix.

Such situation is typical in a broad range of applications, including e.g. physics,<sup>2</sup> biology,<sup>3</sup> and ecology<sup>4</sup> to mention a few.

The central problem in parametric model reconstruction (i.e. when the dimension of the unknown dynamical model is given and fixed) is to infer: (i) unknown dynamical state  $x(t)$ , (ii) unknown nonlinear stochastic dynamical model  $\mathbf{f}(\mathbf{x}|\mathbf{c})$ , and (iii) to estimate accuracy of the reconstruction.

Owing to broad range of applications this problem has received much attention recently in the statistical physics and applied stochastic dynamics communities.<sup>5,6</sup> However, there is no currently general efficient algorithm that can reconstruct unknown state and model of nonlinear dynamical system from noisy incomplete measurements (see e.g. a collection of publications on<sup>7</sup> and the corresponding classification of the algorithms).

---

Further author information: (Send correspondence to D.G.L.)  
D.G.L.: E-mail: d.luchinsky@lancaster.ac.uk, Telephone: +1 (650)6042330  
V.N.S.: E-mail: vadim@email.arc.nasa.gov, Telephone: +1 (650)6042261

The main difficulty in the reconstruction of stochastic nonlinear systems stems, of course, from the fact that there are no efficient method to estimate probabilities for a nonlinear system to have a given dynamical state and given model parameters. The bulk of methods developed in earlier work can be, accordingly, roughly divided into two main streams of algorithms. One is based on various linearizations (e.g. Kalman Filter) and another based on extensive numerical calculations (e.g. Markov Chains Monte Carlo (MCMC))

An important numerical technique, based on the estimation of the drift and diffusion coefficients at a number of points in the phase space of the dynamical system, was suggested in.<sup>8–10</sup> This method was further extended in<sup>11</sup> to an analysis incorporating dynamical and measurement noise. These results allow, at least in principle, subsequent application of the least-squares method for the estimation of the model parameters. Such an empirical method, however, requires a considerable amount of data and an intensive computational effort even for a simple stochastic equation. In particular, if measured trajectory does not pass through the same point in the phase space a number of times the method will fail to estimate drift coefficient. A more general theoretical approach is therefore very desirable.

The most general approach to the solution of this problem is arguably based on Bayes' theorem.<sup>12–14</sup> In particular, it was shown in<sup>13</sup> that parameters of stochastic map with weak measurement noise can be estimated using extended Kalman filter with subsequent application of the MCMC algorithm. The approach discussed in<sup>13</sup> in fact highlights the advantages and disadvantages of the algorithms introduced in earlier work on the nonlinear model reconstruction. The algorithms work well when the system is discrete, the noise is small, the number of measured points is less than few thousands, the model has known functional form, and one has to estimate only a few model parameters. If any of these simplifications is lifted, the algorithm becomes much less reliable and/or very expensive numerically. In particular, the results obtained in<sup>12–14</sup> for the maps can not be immediately adopted to the continuous systems, because of the dynamical noise-induced systematic error of the Bayesian estimators introduce in earlier work.

As we will show, a prefactor accounting for the Jacobian of the transformation has to be included in the likelihood function in the continuous case. Such a prefactor was considered in<sup>15</sup> in the context of Bayesian inference for continuous systems; however, an *ad hoc* likelihood function was used in this paper instead of the correct form derived here. As a result a number of additional limitations on the measured data points and on the initial conditions had to be imposed in this work to obtain good estimation.

A common drawback of these earlier works is their reliance on extensive numerical optimization for finding the minimum of some cost function and intensive numerical multi-dimensional integration for evaluating certain normalization constants. These disadvantages become increasingly more pronounced when systems with ever larger numbers of unknown parameters are investigated.

In our earlier work<sup>16</sup> we have introduced a novel algorithm that does not require extensive global optimization, provides optimal compensation for noise-induced errors, and is robust for a broad range of dynamical models. The key features of the suggested algorithm were the following: (*i*) a likelihood function written in the form of a path-integral over stochastic system trajectories, properly accounting for the Jacobian prefactor term that optimally compensates for dynamical noise; (*ii*) a parametrization of the unknown vector field that reduces the problem of nonlinear dynamical inference to an essentially linear one. These features allowed us to write an efficient Bayesian inference algorithm for nonlinear noise-driven dynamical models. However, it was assumed that the measurement noise is absent, i.e. the dynamical state of the system (1) is known.

In this paper, we extend our technique for inferring the parameters of nonlinear stochastic dynamical systems and encompass the case of arbitrary strong measurement noise. The method is based on the global optimization of the likelihood function using conjugate gradient algorithm. In doing so we avoid the limitation on a small measurement noise introduced by the linearization in the extended Kalman filter algorithm.<sup>13</sup> Because of the correct analytical form of the likelihood function the corresponding gradient and the Hessian matrix can be calculated analytically. This allows us to write a very fast and efficient algorithm for optimization of the likelihood function in the multidimensional space of stochastic dynamical trajectories. The dynamical trajectory with filtered out measurement noise can then be used to infer model parameters.

The paper is organized as follows. In the Sec. 2 we discuss the analytical results of the two step optimization procedure. In the Sec. 3 we illustrate the performance of the algorithm using as an example a stochastic FitzHugh-Nagumo system. Finally, we discuss the obtained results in the Sec. 4.

## 2. THEORY

As we have mentioned above, in a typical experimental situation where the noisy trajectory  $\mathbf{x}(t)$  of a dynamical system is measured at sequential time instants a series  $\mathcal{Y} = \{\mathbf{y}_n \equiv \mathbf{y}(t_n); n = 0, 1, \dots, N\}$  is obtained. The goal of reconstruction of stochastic dynamical systems from noisy measurements is then to find both: (i) the probability density function  $\rho(\mathcal{X}|\mathcal{Y})$  to have a given stochastic trajectory  $\mathcal{X} = \{\mathbf{x}_n \equiv \mathbf{x}(t_n); n = 0, 1, \dots, N\}$  conditioned on observation and (ii) the conditional probability density function  $\rho(\mathcal{M}|\mathcal{Y})$  for the complete set  $\mathcal{M} = \{\mathbf{c}, \hat{\mathbf{D}}, \hat{\mathbf{M}}\}$  of unknown parameters.

The key to the solution of this problem is to find the correct form of the sampling distribution or *likelihood* function  $\ell(\mathcal{Y}|\mathcal{X}, \mathcal{M})$  to have given set of observations  $\mathcal{Y}$  under the condition that the stochastic dynamical trajectory  $\mathcal{X}$  and the model  $\mathcal{M}$  are known. In our earlier work<sup>16</sup> (see also<sup>17</sup> for further discussion and an example of application) we have assumed that there is no measurement noise, i.e. the state of the dynamical system is known. Consequently, we have used the model of observation in the form of product of delta-functions  $p_o(\mathcal{Y}|\mathcal{X}) \simeq \prod_{n=0}^N \delta(\mathbf{y}_n - \mathbf{x}_n)$  and integrated dynamical variables out of the likelihood function. However, in general the presence of measurement noise this is not possible, because dynamical model (1) is essentially nonlinear in  $\mathbf{x}(t)$ .

In this paper we extend our earlier results and introduce explicitly the model of measurements with arbitrary large measurement noise. The complexity of the reconstruction problem is substantially increased in this case, since both the state and the model are now unknown. The state of the dynamical system has now to be found in a multidimensional space of stochastic dynamical trajectories. In traditional approach to the solution of this problem (see e.g.<sup>4, 13, 14</sup> and<sup>7</sup> for further references) two basic techniques were exploited: linearization and extensive numerical calculations. Accordingly, the stronger is the measurement noise and the larger is the dynamical model these techniques become less and less reliable. Furthermore, results of the earlier work are mainly related to the maps and can not be immediately applied to the continuous systems, in which dynamical noise introduces additional systematic error in parameters estimation. Here we provide an analytical solution of this problem, which is stable both for strong noise and high-dimensional model space.

The main steps of the method include: (i) discretization of the model and writing an explicit analytical form of the likelihood function; (ii) global optimization of the likelihood function in the space of stochastic dynamical trajectories conditioned on given model parameters; (iii) analytical calculation of the probability density function for the model parameters conditioned on a given stochastic dynamical trajectory. Two last steps can be repeated iteratively until convergence is achieved.

Below we consider this procedure in details.

### 2.1. The likelihood function

A convenient framework for derivation of the correct form of the likelihood function is a pre-point Euler discretization scheme of the model (1), (2) in the form

$$\left. \begin{aligned} \mathbf{x}_{n+1} &= \mathbf{x}_n + h \mathbf{f}(\mathbf{x}_n^* | \mathbf{c}) + \mathbf{z}_n, \\ \mathbf{y}_n &= \hat{\mathbf{F}} \mathbf{x}_n + \nu_n, \end{aligned} \right\} \quad (3)$$

where we have adopted a uniform sampling scheme  $t_n = t_0 + nh$  with constant and small time step  $h \equiv (t_N - t_0)/N$ . In the notations of (3)  $\mathbf{x}_n^* \equiv \frac{1}{2} (\mathbf{x}_{n+1} + \mathbf{x}_n)$  while  $\mathbf{z}_n$  and  $\nu_n$  are independent, zero-mean Gaussian random variables with  $\langle \mathbf{z}_n \mathbf{z}_{n'}^\top \rangle = h \hat{\mathbf{D}} \delta_{nn'}$  and  $\langle \nu_n \nu_{n'}^\top \rangle = \hat{\mathbf{M}} \delta_{nn'}$ .

Then the sampling distribution can be factorized as a product of two main factors

$$\ell(\mathcal{Y}|\mathcal{X}, \mathcal{M}) = p_o(\mathcal{Y}|\mathcal{X}) \mathcal{F}_{\mathcal{M}}[\mathbf{x}(t)], \quad (4)$$

where  $p_o(\mathcal{Y}|\mathcal{X})$  represents the conditional probability density of observations and  $\mathcal{F}_{\mathcal{M}}[\mathbf{x}(t)]$  is the probability density functional to have given dynamical trajectory conditioned on fixed values of the model parameters.

The conditional probability density of observations can be readily written in the form

$$p_o(\mathcal{Y}|\mathcal{X}) = \prod_{n=0}^N \frac{1}{\sqrt{(2\pi h)^L |\hat{\mathbf{M}}|}} \exp\left(-\frac{1}{2} [\mathbf{y}_n - \hat{\mathbf{\Gamma}}\mathbf{x}_n]^T \hat{\mathbf{M}}^{-1} [\mathbf{y}_n - \hat{\mathbf{\Gamma}}\mathbf{x}_n]\right), \quad (5)$$

The explicit form of the probability density functional  $\mathcal{F}_{\mathcal{M}}[\mathbf{x}(t)]$  has been given in.<sup>18–21</sup> Here for convenience of the reader we derive this results for a particular case (3). The starting point of the derivation is an observation that probability of a particular realization of the random noise process is simply given by

$$\mathcal{P}[\{\mathbf{z}_n\}] = \prod_{n=0}^{N-1} \frac{d\mathbf{z}_n}{\sqrt{(2\pi h)^L |\hat{\mathbf{D}}|}} \exp\left(-\frac{1}{2h} \mathbf{z}_n^T \hat{\mathbf{D}}^{-1} \mathbf{z}_n\right). \quad (6)$$

Changing to dynamical variables using (3), we obtain the PDF for the system (1) to have an arbitrary dynamical trajectory  $\{\mathbf{x}_n\}$  in the form

$$\begin{aligned} \mathcal{F}_{\mathcal{M}}[\{\mathbf{x}_n\}] &= p_s(\mathbf{x}_0) J(\{\mathbf{x}_n\}) \\ &\times \prod_{n=0}^{N-1} \frac{1}{\sqrt{(2\pi h)^L |\hat{\mathbf{D}}|}} \exp\left(-\frac{1}{2h} [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]^T \hat{\mathbf{D}}^{-1} [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]\right), \end{aligned} \quad (7)$$

where  $p_s(\mathbf{x})$  signifies the stationary distribution of  $\mathbf{x}(t)$ ,  $\dot{\mathbf{x}}_n \equiv (\mathbf{x}_{n+1} - \mathbf{x}_n)/h$  is the “velocity”, and the Jacobian of the transformation is given by

$$J(\{\mathbf{x}_n\}) = \left| \left\{ \frac{\partial z_{ln}}{\partial x_{l'n'}} \right\} \right| \simeq \prod_{n=1}^N \prod_{l=1}^L \left[ 1 - \frac{h}{2} \frac{\partial f_l(\mathbf{x}_n^*|\mathbf{c})}{\partial x_{ln}} \right] \simeq \exp\left[-\frac{h}{2} \sum_{n=1}^N \text{tr } \hat{\Phi}(\mathbf{x}_n^*|\mathbf{c})\right], \quad (8)$$

approximated to leading order in  $h$ , with  $\Phi_{ll'}(\mathbf{x}|\mathbf{c}) \equiv \partial f_l(\mathbf{x}|\mathbf{c})/\partial x_{l'}$ .

With these substitutions, (4) is easily evaluated, yielding

$$\begin{aligned} -\frac{2}{N} \ln \ell(\mathcal{Y}|\mathcal{M}) &= -\frac{2}{N} p_s(\mathbf{y}_0) + 2L \ln(2\pi h) + \ln |\hat{\mathbf{D}}| + \ln |\hat{\mathbf{M}}| \\ &+ \frac{1}{N} \sum_{n=0}^N [\mathbf{y}_n - \hat{\mathbf{\Gamma}}\mathbf{x}_n]^T \hat{\mathbf{M}}^{-1} [\mathbf{y}_n - \hat{\mathbf{\Gamma}}\mathbf{x}_n] + \frac{h}{N} \sum_{n=0}^{N-1} \left\{ \text{tr } \hat{\Phi}(\mathbf{x}_n|\mathbf{c}) + [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]^T \hat{\mathbf{D}}^{-1} [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})] \right\}. \end{aligned} \quad (9)$$

It is important to note that this likelihood function is asymptotically exact in the limit  $h \rightarrow 0$  and  $N \rightarrow \infty$  while  $T = Nh$  remains constant.

It is the term  $\text{tr } \hat{\Phi}(\mathbf{y}_n|\mathbf{c})$  in the above that provides optimal compensation for the detrimental effects of dynamical noise, and distinguishes our likelihood function from those introduced in earlier works. Formally, this term emerges from the path integral as the Jacobian of the transformation from noise variables to dynamical variables.<sup>19, 22</sup>

## 2.2. Global optimization in the space of dynamical trajectories

To find the global minimum of the minus likelihood function (13) derived above in the space of dynamical trajectories we employ a conjugate gradient method.<sup>23</sup> The computational efficiency is achieved by analytical calculations of the gradient and the Hessian matrix of  $S = -\ln \ell(\mathcal{Y}|\mathcal{M})$ .

The gradient function is given by the following expression

$$\begin{aligned} \frac{\partial S}{\partial \mathbf{x}_n} &= -(y_n - \mathbf{\Gamma}\mathbf{x}_n)^T \hat{\mathbf{M}}^{-1} \mathbf{\Gamma} \\ &+ \frac{h}{2} \frac{\partial}{\partial \mathbf{x}_n} \left[ \text{tr } \hat{\Phi}(\mathbf{x}_n|\mathbf{c}) \right] + [\dot{\mathbf{x}}_{n-1} - \mathbf{f}(\mathbf{x}_{n-1}^*|\mathbf{c})]^T \hat{\mathbf{D}}^{-1} - h[\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^*|\mathbf{c})]^T \hat{\mathbf{D}}^{-1} \left[ \hat{\mathbf{I}}/h + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n} \right]. \end{aligned} \quad (10)$$

The Hessian matrix has a tridiagonal block structure. In particular case of two dimensional dynamical model it takes the form

$$\hat{\mathbf{H}} = \begin{pmatrix} \hat{\mathbf{B}}_{00} & \hat{\mathbf{B}}_{01} & \hat{\mathbf{0}} & \dots & \hat{\mathbf{0}} & \hat{\mathbf{0}} \\ \hat{\mathbf{B}}_{10} & \hat{\mathbf{B}}_{11} & \hat{\mathbf{B}}_{12} & \dots & \hat{\mathbf{0}} & \hat{\mathbf{0}} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \hat{\mathbf{0}} & \hat{\mathbf{0}} & \hat{\mathbf{0}} & \dots & \hat{\mathbf{B}}_{n-1,n} & \hat{\mathbf{B}}_{nn} \end{pmatrix}, \quad (11)$$

where matrices  $\hat{\mathbf{B}}_{nm}$  are given by the following expression

$$\begin{aligned} \hat{\mathbf{B}}_{nm} = \frac{\partial S}{\partial \mathbf{x}_n \partial \mathbf{x}_m} &= \mathbf{\Gamma}^T \hat{\mathbf{M}}^{-1} \mathbf{\Gamma} \delta_{nm} + \frac{h}{2} \frac{\partial}{\partial \mathbf{x}_n \partial \mathbf{x}_m} [\text{tr } \hat{\Phi}(\mathbf{x}_n | \mathbf{c})] \delta_{nm} + \frac{1}{h} \hat{\mathbf{D}}^{-1} \delta_{nm} \\ &+ h \left[ \hat{\mathbf{I}}/h + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n} \right]^T \hat{\mathbf{D}}^{-1} \left[ \hat{\mathbf{I}}/h + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n} \right] \delta_{nm} - h [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^* | \mathbf{c})]^T \hat{\mathbf{D}}^{-1} \left[ \frac{\partial^2 \mathbf{f}}{\partial \mathbf{x}_n^2} \right] \delta_{nm} \\ &- h \left[ \hat{\mathbf{I}}/h + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_{n-1}} \right]^T \hat{\mathbf{D}}^{-1} \delta_{m,n-1} - h \hat{\mathbf{D}}^{-1} \left[ \hat{\mathbf{I}}/h + \frac{\partial \mathbf{f}}{\partial \mathbf{x}_n} \right] \delta_{m,n+1} \end{aligned} \quad (12)$$

### 2.3. The algorithm

Given a set of noisy observations  $\mathcal{Y}$  we first minimize  $S$  with respect to  $\mathcal{X}$  keeping fixed values for the model parameters. According to the standard conjugate gradient procedure<sup>23</sup> we follow steps:

1. choose initial values for the state vector  $\mathcal{X}_0$  and choose initial directions  $\mathbf{d}_0 = -\nabla S(\mathcal{X}_0)$
2. update values of the coordinates using  $\mathcal{X}_1 = \mathcal{X}_0 + \alpha \mathbf{d}_0$  where

$$\alpha = - (\mathbf{d}_0^T \nabla S(\mathcal{X}_0)) / (\mathbf{d}_0^T \hat{\mathbf{H}}(\mathcal{X}_0) \mathbf{d}_0)$$

3. update direction, using

$$\mathbf{d}_1 = -\nabla S(\mathcal{X}_1) + \frac{|\nabla S(\mathcal{X}_1)|^2}{|\nabla S(\mathcal{X}_0)|^2} \mathbf{d}_0$$

Once the conjugate gradient algorithm has converged to some global minimum  $\mathcal{X}$  in the space of dynamical trajectories. We use this trajectory to infer model parameters in two steps as was described earlier in.<sup>16</sup> In the first step we assume that the diffusion matrix  $\hat{\mathbf{D}}$  is known and infer model parameters  $\mathbf{c}$ . In the second step we infer  $\hat{\mathbf{D}}$  for obtained values of  $\mathbf{c}$ . This procedure is repeated until convergence is reached.

We start by writing an abbreviated likelihood function to have given dynamical trajectory  $\mathcal{X}$  conditioned on fixed model parameters  $\mathcal{M}$  (cf.<sup>16</sup>)

$$-\frac{2}{N} \ln \tilde{\ell}(\mathcal{Y} | \mathcal{M}) = L \ln(2\pi h) + \ln |\hat{\mathbf{D}}| + \frac{h}{N} \sum_{n=0}^{N-1} \left\{ \text{tr } \hat{\Phi}(\mathbf{x}_n | \mathbf{c}) + [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^* | \mathbf{c})]^T \hat{\mathbf{D}}^{-1} [\dot{\mathbf{x}}_n - \mathbf{f}(\mathbf{x}_n^* | \mathbf{c})] \right\}. \quad (13)$$

Next we choose a prior model PDF that is Gaussian in  $\mathbf{c}$  and uniform in  $\hat{\mathbf{D}}$ :

$$p_{\text{pr}}(\mathcal{M}) \propto \sqrt{\frac{|\hat{\Sigma}_{\text{pr}}|}{(2\pi)^M}} \exp \left[ -\frac{1}{2} (\mathbf{c} - \mathbf{c}_{\text{pr}})^T \hat{\Sigma}_{\text{pr}} (\mathbf{c} - \mathbf{c}_{\text{pr}}) \right], \quad (14)$$

and apply Bayes theorem to obtain the posterior model PDF in the form  $p_{\text{post}}(\mathcal{M} | \mathcal{X}) = \text{const} \times \exp[-\tilde{S}(\mathcal{M} | \mathcal{X})]$ , where

$$\tilde{S}(\mathcal{M} | \mathcal{X}) \equiv \tilde{S}_{\mathcal{X}}(\mathbf{c}, \hat{\mathbf{D}}) = \frac{1}{2} \rho_{\mathcal{X}}(\hat{\mathbf{D}}) - \mathbf{c}^T \mathbf{w}_{\mathcal{X}}(\hat{\mathbf{D}}) + \frac{1}{2} \mathbf{c}^T \hat{\Sigma}_{\mathcal{X}}(\hat{\mathbf{D}}) \mathbf{c} \quad (15)$$

is the cost function whose global minimum yields the most probable model  $\mathcal{M}^*$ . Here, use was made of the definitions

$$\rho_{\mathcal{X}}(\hat{\mathbf{D}}) = h \sum_{n=0}^{N-1} \dot{\mathbf{x}}_n^T \hat{\mathbf{D}}^{-1} \dot{\mathbf{x}}_n + N \ln |\hat{\mathbf{D}}|, \quad (16)$$

$$\mathbf{w}_{\mathcal{X}}(\hat{\mathbf{D}}) = \hat{\Sigma}_{\text{pr}} \mathbf{c}_{\text{pr}} + h \sum_{n=0}^{N-1} \left[ \hat{\mathbf{U}}_n^T \hat{\mathbf{D}}^{-1} \dot{\mathbf{x}}_n - \frac{\mathbf{v}(\mathbf{x}_n)}{2} \right], \quad (17)$$

$$\hat{\Xi}_{\mathcal{X}}(\hat{\mathbf{D}}) = \hat{\Sigma}_{\text{pr}} + h \sum_{n=0}^{N-1} \hat{\mathbf{U}}_n^T \hat{\mathbf{D}}^{-1} \hat{\mathbf{U}}_n, \quad (18)$$

where  $\hat{\mathbf{U}}_n \equiv \hat{\mathbf{U}}(\mathbf{x}_n)$ , and the components of the vector  $\mathbf{v}(\mathbf{x})$  are

$$v_m(\mathbf{x}) = \sum_{l=1}^L \frac{\partial U_{lm}(\mathbf{x})}{\partial x_l}, \quad m = 1, \dots, M. \quad (19)$$

For a given block of data  $\mathcal{X}$  of length  $N$ , the best estimates for the model parameters are given by the posterior means of  $\mathbf{c}$  and  $\hat{\mathbf{D}}$ , which coincide with the global minimum of  $S_{\mathcal{X}}(\mathbf{c}, \hat{\mathbf{D}})$ . We handle this optimization problem in the following way. Assume for the moment that  $\mathbf{c}$  is known in (15); for the first iteration, take  $\mathbf{c} = \mathbf{c}_{\text{pr}}$ . Then, minimizing  $S_{\mathcal{X}}(\mathbf{c}, \hat{\mathbf{D}})$  with respect to  $\hat{\mathbf{D}}$ , we find that the posterior distribution over  $\hat{\mathbf{D}}$  has a mean

$$\langle \hat{\mathbf{D}} \rangle = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \dot{\mathbf{x}}_n - \hat{\mathbf{U}}(\mathbf{x}_n) \mathbf{c} \right] \left[ \dot{\mathbf{x}}_n - \hat{\mathbf{U}}(\mathbf{x}_n) \mathbf{c} \right]^T. \quad (20)$$

Assume next that  $\hat{\mathbf{D}}$  is known, and note from (15) that in this case, the posterior distribution over  $\mathbf{c}$  is Gaussian. Its covariance is given by  $\hat{\Xi}_{\mathcal{X}}(\hat{\mathbf{D}})$ , and its mean

$$\langle \mathbf{c} \rangle = \hat{\Xi}_{\mathcal{X}}^{-1}(\hat{\mathbf{D}}) \mathbf{w}_{\mathcal{X}}(\hat{\mathbf{D}}) \quad (21)$$

minimizes  $S_{\mathcal{X}}(\mathbf{c}, \hat{\mathbf{D}})$  with respect to  $\mathbf{c}$ . Thus, for the second iteration,  $\mathbf{c}_{\text{pr}}$  and  $\hat{\Sigma}_{\text{pr}}$  are replaced with  $\langle \mathbf{c} \rangle$  and  $\hat{\Xi}_{\mathcal{X}}(\hat{\mathbf{D}})$ , respectively. This two-step (analytical) optimization procedure is continued iteratively until convergence, which is typically much faster than brute-force numerical optimization that has been attempted in earlier works.

### 3. NUMERICAL EXAMPLE

Consider as an example a nonlinear system with stable limit cycle in the form

$$\dot{x}_1 = x_2 - x_1^2 x_2 + \xi_1(t) \quad \dot{x}_2 = -x_1 + 0.1(1 - x_1^2)x_2 + \xi_2(t). \quad (22)$$

The state of the dynamical system is unknown. We assume for simplicity that both coordinates were measured with measurement noise of amplitude 0.4 in both coordinates, i.e.

$$y_i(t) = x_i(t) + 0.4\nu_i(t), \quad i = 1, 2.$$

Here the measurement matrix has the form  $\hat{\mathbf{F}} = \hat{\mathbf{I}}$  and measurement noise matrix has the form  $\hat{\mathbf{M}} = 0.16\hat{\mathbf{I}}$ .

We further assume that the vector field of (22) is unknown and model it as follows (cf.<sup>16</sup>)

$$\mathbf{f}(\mathbf{x}|\mathbf{c}) = \hat{\mathbf{U}}(\mathbf{x}) \mathbf{c}, \quad (23)$$

where matrix  $\hat{\mathbf{U}}$  has the form

$$\hat{\mathbf{U}} = \left[ \begin{pmatrix} \phi_1 & 0 & \dots & 0 \\ 0 & \phi_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_1 \end{pmatrix} \begin{pmatrix} \phi_2 & 0 & \dots & 0 \\ 0 & \phi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_2 \end{pmatrix} \cdots \begin{pmatrix} \phi_G & 0 & \dots & 0 \\ 0 & \phi_G & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \phi_G \end{pmatrix} \right], \quad (24)$$

**Table 1.** Convergence of some coefficients of the system (25). We have used one block of data with 40000 points.

coefficients	true values	inferred values	updated values
$c_3$	1	1.46	1.08
$c_4$	0	-42.64	0.02
$c_{10}$	-1	-10.93	-1.07
$c_{11}$	0.1	-35.68	0.258
$c_{15}$	0	1.82	0.005
$c_{16}$	-0.1	-27.96	-0.17
$D_{11}$	0.04	325	0.045
$D_{12}$	0	6	0.006
$D_{22}$	0.04	318	0.03

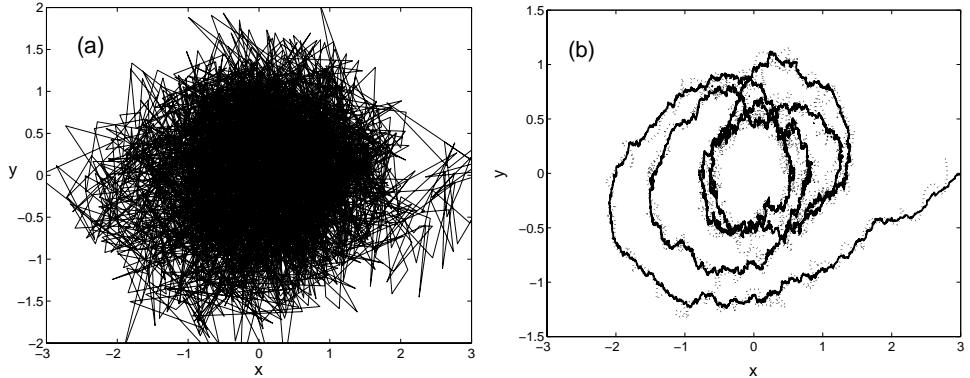
and the set of 8 known base function is taken to be

$$\phi = \{1; x_1; x_2; x_1^2; x_2^2; x_1 x_2; x_1^3; x_1^2 x_2\}.$$

In explicit form the model of the limit cycle system (22) takes the following form

$$\dot{x}_1 = \sum_{i=1}^8 c_{2i-1} \phi_i + \xi_1(t) \quad \dot{x}_2 = \sum_{i=1}^8 c_{2i} \phi_i + \xi_2(t). \quad (25)$$

We now apply algorithm described in the previous section to infer both the unknown state and vector field of this system. An example of corrupted by noise measurements of the (22) is shown in the Fig. 1(a). Our technique allows to recover stochastic dynamics of the system (22) as shown in the Fig. 1(b) and to estimate model parameters. The results of the estimations are shown in the Table 1.



**Figure 1.** (a) Example of corrupted by noise measurements (22) with intensity of dynamical noise  $\langle \xi_x^2(t) \rangle = 0.1$  and  $\langle \xi_y^2(t) \rangle = 0.2$  and the amplitude of measurements noise 0.4 in both coordinates. (b) Recovered stochastic dynamics of the system (22) (dotted line) is shown in comparison with the actual dynamical trajectory (solid line).

#### 4. DISCUSSION

We introduce a novel technique of reconstruction of the unknown state and model of nonlinear stochastic dynamical systems from noisy incomplete measurements. The method is fast and efficient since it does not rely on

extensive numerical calculations. Another important advantage of this technique is that we avoid linearizations and assume an arbitrary strong measurement noise.

In particular case of a dynamical system with global attractor such as Lorenz system or the one considered in (22) the method allows for analytical solution of the problem. Specifically, the efficiency of the global optimization in the multidimensional space of stochastic trajectories (with 40000 points in each trajectory) was achieved by analytical calculations of the gradient and the Hessian of the likelihood function. We illustrate the convergence of our algorithm by reconstructing the unknown dynamical state and 19 unknown model parameters of two-dimensional dynamical system with stable limit cycle.

For a more general case the method can be generalized readily by supplementing the discussed algorithm with simulated annealing technique.

In its present form, however, the method has essential limitation since we have assumed that all the dynamical variables are measured. The extension of the algorithm to encompass the later case will be discussed elsewhere.

## REFERENCES

1. H. D. I. Abarbanel, “Challenges in modelling nonlinear systems: a worked example,” in *Nonlinear dynamics and statistics*, A. I. Mees, ed., ch. I, pp. 1–30, Birkhäuser, Boston, 2001.
2. M. Willemsen, M. P. van Exter, and J. P. Woerdman, “Anatomy of a polarization switch of a vertical-cavity semiconductor laser,” *Phys. Rev. Lett.* **84**(19), pp. 4337–4340, 2000.
3. K. Visscher, M. J. Schnitzer, and S. M. Block, “Single kinezin molecules studied with a molecular force clamp,” *Nature* **400**, pp. 184–189, 1999.
4. C. Calder, M. Lavine, P. Müller, and J. S. Clark, “Incorporating multiple sources of stochasticity into dynamic population models,” *Ecology* **84**(6), pp. 1395–1402, 2003.
5. P. Congdon, *Bayesian statistical modelling*, Wiley series in probability and statistics, Wiley, Chichester, 2001.
6. G. D’Agostini, *Bayesian reasoning in high-energy physics : principles and applications*, CERN, Geneva, 1999.
7. <http://research.microsoft.com/~minka/papers/dynamic.html>
8. S. Siegert, R. Friedrich, and J. Peinke, “Analysis of data sets of stochastic systems,” *Phys. Lett. A* **253**, pp. 275–280, 1998.
9. J. Gradišek, S. Siegert, R. Friedrich, and I. Grabec, “Analysis of time series from stochastic processes,” *Phys. Rev. E* **62**(3), pp. 3146–3155, 2000.
10. R. Friedrich, S. Siegert, J. Peinke, S. Luck, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, and G. Pfister, “Extracting model equations from experimental data,” *Phys. Lett. A* **271**, pp. 217–222, 2000.
11. M. Siefert, A. Kittel, R. Friedrich, and J. Peinke, “On a quantitative method to analyze dynamical and measurement noise,” *Europhys. Lett.* **61**(4), pp. 466–472, 2003.
12. J. P. M. Heald and J. Stark, “Estimation of noise levels for models of chaotic dynamical systems,” *Phys. Rev. Lett.* **84**(11), pp. 2366–2369, 2000.
13. R. Meyer and N. Christensen, “Fast bayesian reconstruction of chaotic dynamical systems via extended kalman filtering,” *Phys. Rev. E* **65**, p. 016206, 2001.
14. C. L. Bremer and D. T. Kaplan, “Markov chain monte carlo estimation of nonlinear dynamics from time series,” *Physica D* **160**, pp. 116–126, 2001.
15. J.-M. Fullana and M. Rossi, “Identification methods for nonlinear stochastic systems,” *Physical Review E* **65**, p. 031107, 2002.
16. V. N. Smelyanskiy and D. G. Luchinsky, “Model reconstruction of nonlinear dynamical systems driven by noise,” in *Noise in complex systems and stochastic dynamics II*, Z. Gingl, J. M. Sancho, L. Schimansky-Geier, and J. Kertesz, eds., *Proc. SPIE* **5471**, pp. 344–354, SPIE, (Bellingham), 2004.
17. V. N. Smelyanskiy, D. G. Luchinsky, A. Stefanovska, and P. V. E. McClintock, “Reconstruction of nonlinear cardiorespiratory interaction from univariate time-series,” *to appear in Phys. Rev. Lett.* .
18. R. Graham Z. *Phys. B* **26**, p. 281, 1977.

19. E. Gozzi, “Functional-integral approach to parisi-wu stochastic quantization: Scalar theory,” *Physical Review D* **28**, pp. 1922–1930, 1983.
20. A. J. McKane, “Noise-induced escape rate over a potential barrier: Results for a general noise,” *Phys. Rev. A* **40**(7), pp. 4050–4053, 1989.
21. M. I. Dykman, “Large fluctuations and fluctuational transitions in systems driven by colored gaussian noise—a high frequency noise,” *Phys. Rev. A* **42**, pp. 2020–2029, 1990.
22. R. Graham, *Tracts in Modern Physics*, vol. 66, ch. Quantum Statistics in Optics and Solid-State Physics. Springer-Verlag, New York, 1973.
23. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, Cambridge University Press, Cambridge, England, 2 ed., 1992.