An iterative procedure for the estimation of drift and diffusion coefficients of Langevin processes

D. Kleinhans [a,*], R. Friedrich [a], A. Nawroth [b], J. Peinke [b]

[a] *Institute for Theoretical Physics, University of Münster, D-48149 Münster, Germany*
[b] *Institute for Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany*

Abstract

A general method is proposed which allows one to estimate drift and diffusion coefficients of a stochastic process governed by a Langevin equation. It extends a previously devised approach [R. Friedrich et al., Phys. Lett. A 271 (2000) 217], which requires sufficiently high sampling rates. The analysis is based on an iterative procedure minimizing the Kullback–Leibler distance between measured and estimated two time joint probability distributions of the process.
© 2005 Elsevier B.V. All rights reserved.

PACS: 87.23.Cc; 02.50.Ey; 05.40.Jc

1. Introduction

Complex behavior in systems far from equilibrium can quite often be traced back to rather simple laws due to the existence of processes of selforganization [1]. Since complex systems are composed of a huge number of subsystems, however, fluctuations stemming from the microscopic degrees of freedom play an important role introducing a temporal variation on a fast time scale which quite often can be considered as fluctuations. The consequence is the ex-

istence of evolution equations of a set of macroscopic order parameters $\mathbf{q}(t)$ which are governed by nonlinear Langevin equations [2,3]:

$$\frac{d}{dt}q_i = D_i^1(\mathbf{q}) + \sum_l g_{il}(\mathbf{q})F_l(t), \quad (1)$$

where $\mathbf{q}(t)$ denotes the n-dimensional state vector, $\mathbf{D}^1(\mathbf{q})$ is the drift vector and the matrix $g(\mathbf{q})$ is related to the diffusion matrix according to $(D^2(\mathbf{q}))_{ij} = \sum_k g_{ik}(\mathbf{q})g_{jk}(\mathbf{q})$. $\mathbf{F}(t)$ are fluctuating forces with Gaussian statistics delta-correlated in time: $\langle F_l(t) \rangle = 0$, $\langle F_l(t)F_k(t') \rangle = 2\delta_{lk}\delta(t - t')$. Here and in the following we adopt Itô's interpretation of stochastic integrals [2,3].

* Corresponding author.
  E-mail address: kleinhan@uni-muenster.de (D. Kleinhans).

Analyzing complex systems, which can be described by stochastic equations of the form (1), therefore, amounts to assess the underlying Langevin equations or the corresponding Fokker–Planck equations from an inspection of experimentally determined time series [4]. Recently, an operational method [5,6] has been devised, which allows one to estimate drift and diffusion coefficients of the stochastic processes from experimental data. This method has been successfully applied to various problems in the field of complex systems like the analysis of noisy electrical circuits [6], stochastic dynamics of metal cutting [7], systems with feedback delay [8], meteorological processes like wind-driven Southern Ocean variability [9], traffic flow data [10] and physiological time series [11]. Furthermore it has been applied to problems like turbulent flows [12,13], passive scalar advection [14], financial time series [15], analysis of rough surfaces [16,17], which can be characterized as a stochastic process with respect to a scale variable exhibiting Markovian properties in scale.

The method is based on the evaluation of the time limits the first and second conditional moments,

$$\mathbf{D}^1(\mathbf{q}) = \lim_{\tau \to 0} \frac{1}{\tau} \langle \mathbf{q}(t+\tau) - \mathbf{q}(t) \mid \mathbf{q}(t) = \mathbf{q} \rangle, \qquad (2a)$$

$$D_{ij}^2(\mathbf{q}) = \lim_{\tau \to 0} \frac{1}{2\tau} \langle \left[\mathbf{q}(t+\tau) - \mathbf{q}(t) \right]_i \\ \times \left[\mathbf{q}(t+\tau) - \mathbf{q}(t) \right]_j \mid \mathbf{q}(t) = \mathbf{q} \rangle. \qquad (2b)$$

From these expressions it becomes evident that the sampling rate in the experiments has to be sufficiently high in order to allow for a reliable evaluation of the limit $\tau \to 0$. Therefore, in all applications mentioned above the results have been checked in a selfconsistent manner by a recalculation of conditional pdf's from the estimated Fokker–Planck equation. Possible problems in estimating drift and diffusion coefficients related with low sampling frequencies have been addressed by Sura [18], Ragwitz and Kantz [19], [21] and Friedrich et al. [20].

The aim of the present Letter is to devise an extension of the above method in order to overcome problems related with the time limit $\tau \to 0$. These problems immediately show up for low sampling rates. We also want to point out that for the case of stochastic forces $\mathbf{F}(t)$ with small but finite temporal correlations the process is not Markovian in the limit $\tau \to 0$. In this case, however, one should use the Stratonovich interpretation of stochastic processes [2].

2. Description of the method

The starting point is a first estimate of drift and diffusion coefficients by the expressions (2) evaluated for the smallest reliably possible values of τ. The second step is an embedding of drift and diffusion coefficients into a family of functions $\mathbf{D}^1(\mathbf{q}, \sigma)$, $\mathbf{D}^2(\mathbf{q}, \sigma)$ parameterized by a set of free parameters σ. The expressions obtained in the first step already yield a crude estimate of the parameters σ. The third step consists in optimizing the free parameters σ.

Optimization of the free parameters can be performed in the following way. One determines the conditional probability distribution

$$p(\mathbf{q}, t | \mathbf{q}_0, t_0; \sigma) \qquad (3)$$

for the parameter set σ either by a simulation of the Langevin equations or by a numerical solution of the corresponding Fokker–Planck equation. In each case, one can determine the two point pdf $f(\mathbf{q}, t; \mathbf{q}_0, t_0; \sigma) = p(\mathbf{q}, t | \mathbf{q}_0, t_0; \sigma) f(\mathbf{q}_0, t_0)$. The reader should note that this may be done for various finite values of $t - t_0$. The obtained two time pdf can now be compared with the experimental one. A suitable measure for the distance is the Kullback–Leibler information [4] defined according to

$$K(\sigma, t, t_0) = \int d\mathbf{q} \int d\mathbf{q}_0 \, f_{\exp}(\mathbf{q}, t; \mathbf{q}_0, t_0) \\ \times \ln \frac{f_{\exp}(\mathbf{q}, t; \mathbf{q}_0, t_0)}{f(\mathbf{q}, t; \mathbf{q}_0, t_0, \sigma)}. \qquad (4)$$

The minimum of the Kullback–Leibler information with respect to the parameters σ yields estimates of drift and diffusion of a stochastic process. This process is the best approximation with respect to this measure in the class of stochastic processes characterized by the parameters σ. The problem of identifying a stochastic process is then equivalent to determining a minimum of the Kullback information. In practice the minimum can be determined by gradient or genetic algorithms and solved by standard methods [22]. In the following we shall consider cases, where it is possible to obtain a parametrization of the stochastic processes by only few parameters σ such that

the Kullback–Leibler measure can be investigated by graphical means.

3. Examples

For certain classes of stochastic processes the above procedure can be reduced considerably by the fact that only few free parameters for the parametrization of drift and diffusion terms have to be introduced. As a consequence the minimization procedure of the Kullback–Leibler information is greatly facilitated.

3.1. One-dimensional systems

The case of one-dimensional systems allows for the following treatment due to the fact that the stationary pdf, which is assumed to exist, can be determined analytically:

$$f(q) = \frac{N}{D^2(q)} e^{\int^q dq' \frac{D^1(q')}{D^2(q')}}. \tag{5}$$

As a consequence, we have the relationship

$$D^1(q) = D^2(q) \frac{d}{dq} \ln f(q) + \frac{d}{dq} D^2(q). \tag{6}$$

Since $f(q)$ can be determined from the time series an estimate in terms of a parameterized ansatz for the diffusion term suffices. In fact, one may use the ansatz $D^2(q) = Q + aq^2 + bq^4 + \cdots$, which helps in lowering the number of parameters σ to be estimated by the above procedure of minimization the Kullback–Leibler information. The drift then follows from (6).

Let us consider system I with drift and diffusion functions

$$D^1(q) = q - q^3 \quad \text{and} \quad D^2(q) = 1 + q^2 \tag{7}$$

driven by a multiplicative noise term. We use synthetic data obtained by numerical integration of the corresponding Langevin equation [2],

$$
\begin{aligned}
q(t + \tilde{\tau}) = {} & q(t) + \tilde{\tau} D^1\big[q(t)\big] \\
& + \sqrt{\tilde{\tau}} D^2\big[q(t)\big] \Gamma(t).
\end{aligned}
\tag{8}
$$

A time series containing 10^6 points with time increment 10^{-2} was generated. The intrinsic increment $\tilde{\tau}$ used for numerical integration of the corresponding Langevin equation was 10^{-5}. A time segment of the
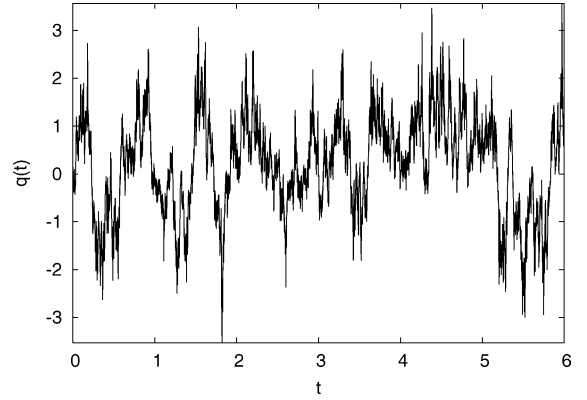

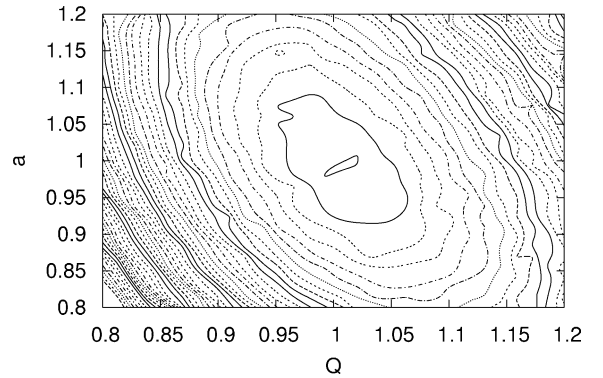Fig. 1. Segment of the one-dimensional synthetic time series I.

Fig. 2. Kullback distance $K(Q, a)$ as function of the parameters Q and a for time series I. The lines are equidistant contour lines starting from 2.6×10^{-4} in the center. The distance between contour lines is 5×10^{-5}. A clear minimum is located at $(Q, a) = (1, 1)$.

data is presented in Fig. 1. Since the stochastic process is stationary and ergodic all statistical quantities can be retrieved from this data.

For the estimation of the pdf's from data state space has to be divided into bins. We used 100 equidistant bins for the stationary pdf. A very accurate way to calculate the integral yielding the Kullback–Leibler distance without running out of memory even for higher-dimensional data is to use an adequate local grid for the first argument (the destination) of the conditional pdf's. The conditional pdf then locally can be retrieved from the data for any $(\mathbf{q}, \mathbf{q}_0)$ with high accuracy. The local grid used in this example covered 20 equidistant bins.

During the iteration procedure the two point pdf's have to be calculated. We again use the numerical sim-

ulation of Langevin processes as a very efficient way to generate these pdf's.

Starting from the estimates (2) the ansatz $D^2(Q, a, q) = Q + aq^2$ is reasonable. The drift immediately follows from (6) and, for each parameter set (Q, a), one obtains a stationary distribution that equals the experimental one. Due to this fact the evaluation of the conditional pdf $p(q, t + \tau | q_0, t; Q, a)$ suffices to calculate the Kullback–Leibler distance. A clear minimum of the distance is found at $(Q, a) = (1, 1)$ corresponding to the original set of parameters. The Kullback distance close to this minimum in the two-dimensional parameter space is exhibited in Fig. 2.

3.2. Application to potential systems

The procedure for one-dimensional systems can be immediately extended to higher dimensions if one restricts the analysis to the so-called class of potential systems for which the drift vector $\mathbf{D}^1(\mathbf{q})$ is obtained from a potential $V(\mathbf{q})$ and $g_{ik} = \sqrt{Q} \delta_{ik}$. The central point of our analysis is the following exact expression for the stationary pdf

$$f(\mathbf{q}) = N e^{-V(\mathbf{q})/Q}. \tag{9}$$

Since the stationary pdf can be estimated from experimental data one may parameterize the class of stochastic processes by the single variable Q. Thus the drift function can be taken to be fixed except for the value Q:

$$\mathbf{D}^1(\mathbf{q}) = Q \nabla \ln f(\mathbf{q}). \tag{10}$$

As an example we consider the two-dimensional system

$$\mathbf{D}^1(\mathbf{q}) = \begin{pmatrix} \epsilon q_1 - q_1[q_1^2 + B q_2^2] \\ \epsilon q_2 - q_2[B q_1^2 + q_2^2] \end{pmatrix}. \tag{11}$$

This dynamical system arises as order parameter equations for instabilities in nonequilibrium systems and has applications from the fields of pattern formation in nonequilibrium systems to pattern recognition [1]. It exhibits the features of multistability and selection. We considered the case $\epsilon = 0.25$ and $B = 2$ (time series II). These parameters yield four stable fixpoints of the dynamics on the axes at $|\mathbf{q}| = 1/2$ and unstable fixpoints at the origin and on the bisectional lines at $|\mathbf{q}| = \sqrt{6}/6$.
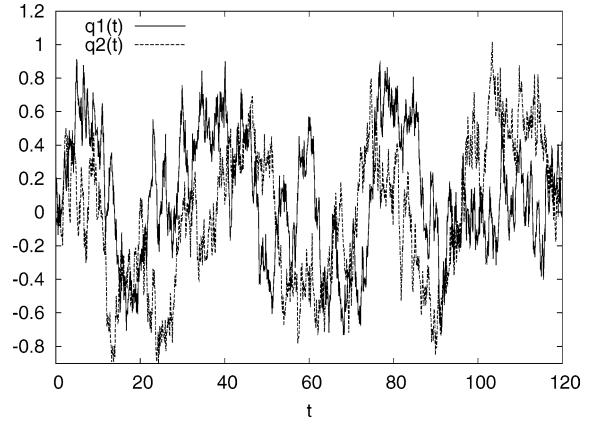

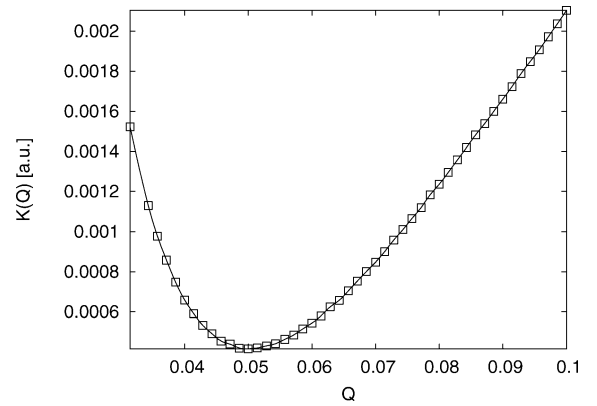Fig. 3. Segment of the two-dimensional synthetic time series II.

Fig. 4. The Kullback distance $K(Q)$ as a function of the noise strength Q (time series II). A minimum is clearly visible at the value $Q = 0.05$.

Data with time increments 10^{-1} for the datapoints has been generated with a time step 10^{-5} for the integration of the Langevin equations. The simulated time series II with $Q = 0.05$ consists of 5×10^6 data points. Fig. 3 exhibits a segment of the generated data.

We analyzed the time series as outlined above. State space in this case is divided in 100×100 equidistant bins. Since the drift $\mathbf{D}^1(\mathbf{q})$ can be evaluated from (10) all parameters are fixed except for the noise strength Q.

After evaluating the Kullback measure for various values of Q this value has to be optimized. The optimal value is determined by the minimum of the Kullback distance. For the present case the minimum can easily be determined by graphical means.
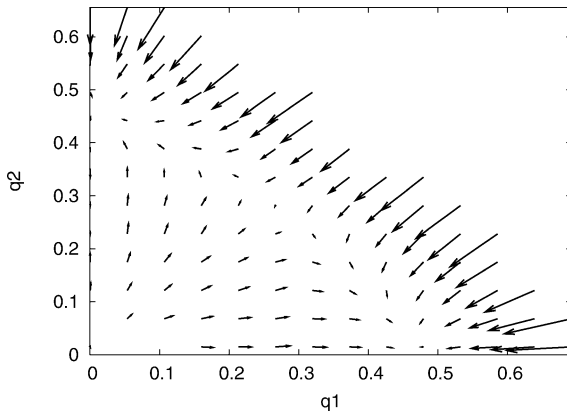
Fig. 5. Time series II: Drift vector field extracted from data using the optimal value of Q. Unstable fixpoints in the center and on the bisectional line as well as the attractive fixpoints are clearly visible.

Fig. 4 shows the Kullback distance $K(Q)$ as a function of the noise strength Q for the time series II. The minimum is clearly visible at $Q = 0.05$ and agrees with the one used for simulation. With this parameter the drift vector field can be recalculated from the stationary distribution based on relation (10). The resulting drift vector field of dataset II is exhibited in Fig. 5.

4. Conclusion

Summarizing, we have outlined an operational method for the estimation of drift and diffusion terms from experimental time series of stochastic Langevin processes. In contrast to previous approaches the present algorithm does not rely on estimating conditional moments in the small time increment limit. Although this limit yields a first approximation an iterative refinement of the estimated stochastic process is performed by minimization of the Kullback–Leibler distance between estimated and measured two time probability distributions. The proposed procedure solves the problem of estimating drift and diffusion terms of Langevin processes from time series. It involves the numerical solution of Langevin equations with parameter dependent drift and diffusion terms, an evaluation of the Kullback–Leibler integral

(which may be determined by means of a Monte Carlo method) and an optimization procedure, for which standard approaches can be used. All involved steps are based on routine calculations. Furthermore, restriction to certain classes of stochastic processes like potential systems can drastically lower the numerical efforts of the procedure. Therefore, the proposed algorithm can be applied also to systems with higher-dimensional state spaces.

References

[1] H. Haken, Synergetics: Introduction and Advanced Topics, Springer-Verlag, New York, 2004.
[2] H. Risken, The Fokker–Planck Equation, Springer-Verlag, New York, 1983.
[3] C.W. Gardiner, Handbook of Stochastic Methods, Springer-Verlag, Berlin, 1983.
[4] H. Haken, Information and Self-Organization: A Macroscopic Approach to Complex Systems, Springer-Verlag, Berlin, 2004.
[5] S. Siegert, R. Friedrich, J. Peinke, Phys. Lett. A 234 (1998) 275.
[6] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindemann, J. Raethjen, G. Deuschl, G. Pfister, Phys. Lett. A 271 (2000) 217.
[7] J. Gradisek, I. Grabec, S. Siegert, R. Friedrich, Mech. Syst. Signal Process. 16 (5) (2002) 831.
[8] T.D. Frank, P.J. Beek, R. Friedrich, Phys. Lett. A 328 (2004) 219;
T.D. Frank, R. Friedrich, P.J. Beek, Stoch. Dyn. 9 (2004) 44.
[9] P. Sura, S.T. Gille, J. Mar. Res. 61 (2003) 313.
[10] S. Kriso, R. Friedrich, J. Peinke, P. Wagner, Phys. Lett. A 299 (2002) 287.
[11] T. Kuusela, Phys. Rev. E 69 (2004) 031916.
[12] R. Friedrich, J. Peinke, Phys. Rev. Lett. 78 (1997) 863.
[13] Ch. Renner, J. Peinke, R. Friedrich, J. Fluid Mech. 433 (2001) 383.
[14] M. Tutkun, L. Mydlarski, New J. Phys. 6 (2004) 49.
[15] R. Friedrich, J. Peinke, Ch. Renner, Phys. Rev. Lett. 84 (2000) 5224.
[16] G.R. Jafari, S.M. Fazeli, F. Ghasemi, S.M.V. Allaei, M.R.R. Tabar, A.I. Zad, G. Kavei, Phys. Rev. Lett. 91 (2003) 226101.
[17] M. Wächter, F. Riess, H. Kantz, J. Peinke, Europhys. Lett. 64 (2003) 579.
[18] P. Sura, J. Barsugli, Phys. Lett. A 305 (2002) 304.
[19] M. Ragwitz, H. Kantz, Phys. Rev. Lett. 87 (2001) 254501.
[20] R. Friedrich, Ch. Renner, M. Siefert, J. Peinke, Phys. Rev. Lett. 89 (2002) 149401.
[21] M. Ragwitz, H. Kantz, Phys. Rev. Lett. 89 (2002) 149402.
[22] E. Weinstein, M. Feder, A.V. Oppenheim, IEEE Trans. Acoustics Speech Signal Process. 38 (1990) 1652.