Fast Bayesian inference of the multivariate Ornstein-Uhlenbeck process

Rajesh Singh,[1] Dipanjan Ghosh,[2] and R. Adhikari[1,3,*]

[1] *The Institute of Mathematical Sciences-HBNI, CIT Campus, Taramani, Chennai 600113, India*
[2] *Dept of Chemical Engineering, Jadavpur University, Kolkata 700032, India*
[3] *DAMTP, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*

The multivariate Ornstein-Uhlenbeck process is used in many branches of science and engineering to describe the regression of a system to its stationary mean. Here we present a fast Bayesian method to estimate the drift and diffusion matrices of the process from discretely observed sample paths. We use exact likelihoods, expressed in terms of four sufficient statistic matrices, to derive explicit maximum a posteriori parameter estimates and their standard errors. We apply the method to the Brownian harmonic oscillator, a bivariate Ornstein-Uhlenbeck process, to jointly estimate its mass, damping, and stiffness and to provide Bayesian estimates of the correlation functions and power spectral densities. We present a Bayesian model comparison procedure, embodying Ockham's razor, to guide a data-driven choice between the Kramers and Smoluchowski limits of the oscillator. These provide novel methods of analysing the inertial motion of colloidal particles in optical traps.

I. INTRODUCTION

The multivariate Ornstein-Uhlenbeck process is widely used in many branches of science and engineering to describe the regression of a system to its stationary mean. Its importance arises from the fact that it is the only process that is simultaneously Markovian, Gaussian and stationary. The process, therefore, is fully characterized by its stationary and conditional distribution, each of which are multivariate normal distributions. This simplicity implies that the probabilities of discretely sampled paths can be calculated exactly and explicitly in terms of the process parameters, the drift and diffusion matrices. In turn, this implies an exact expression for the likelihood of paths, given the parameters, and the possibility of exact Bayesian inference.

Here we present the details of such an exact Bayesian estimation method for the M-dimensional Ornstein-Uhlenbeck process. From the exact likelihood of the discrete path, we identify four matrices of sufficient statistics, in terms of which we derive MAP estimates and the error bars of the process parameters. In addition we examine the problem of model selection, that is, of choosing the Ornstein-Uhlenbeck model with the least number of parameters that best explains the data. Both these aspects are illustrated using the Brownian harmonic oscillator, a bivariate Ornstein-Uhlenbeck process, as an example. Inter alia, this provides a Bayesian method for estimating the mass, friction, spring constant, correlation functions and power spectral densities of colloidal particles in optical traps.

The remainder of the paper is organized as follows. In section II, we briefly review key properties of the Ornstein-Uhlenbeck process and make use of them in section III to obtain MAP estimates for the parameters and the model odds. In section IV we present an exact path sampling algorithm which is used in V to generate sample paths of the Brownian harmonic oscillator and to validate the results of section III. We conclude in section VI with a discussion on further applications.

II. MULTIVARIATE ORNSTEIN-UHLENBECK PROCESS

The multivariate Ornstein-Uhlenbeck is defined by the Ito stochastic differential equation [1]

$$dx_i = -\lambda_{ij} x_j dt + (\sqrt{2D})_{ij} \cdot dW_j, \tag{1}$$

where λ_{ij} is a stable matrix of mean regression rates, D_{ij} is a symmetric positive semi-definite matrix of diffusion coefficients, $W_i(t)$ are Wiener process and $i, j = 1, \ldots, M$. We denote $(x_1, \ldots, x_M)^T$ by the vector $\boldsymbol{x}$ and λ_{ij} by the matrix $\boldsymbol{\lambda}$, with similar bold-face notation for other vectors and matrices, when convenient.

The probability density of a displacement from $\boldsymbol{x}$ at time t to $\boldsymbol{x}'$ at time t', $P_{1|1}(\boldsymbol{x}'t'|\boldsymbol{x}\,t)$, obeys the Fokker-Planck equation $\partial_t P_{1|1} = \mathcal{L} P_{1|1}$, where the Fokker-Planck operator is

$$\mathcal{L}(\boldsymbol{x}) = -\frac{\partial}{\partial x_i}\lambda_{ij} x_j + \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}. \tag{2}$$

The solution is a multivariate normal distribution

$$\boldsymbol{x}'t'|\boldsymbol{x}\,t \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{3}$$

where

$$\boldsymbol{\mu} = \boldsymbol{\Lambda}\boldsymbol{x}, \quad \boldsymbol{\Sigma} = \boldsymbol{c} - \boldsymbol{\Lambda}\boldsymbol{c}\,\boldsymbol{\Lambda}^T, \quad \boldsymbol{\Lambda} = e^{-\boldsymbol{\lambda}\Delta t} \tag{4}$$

and $\Delta t = t' - t$. This solution is exact and holds for arbitrary values of $|\Delta t|$. The stationary distribution $P_1(\boldsymbol{x})$ obeys the steady state Fokker-Plank equation $\mathcal{L}P_1 = 0$ and the solution is, again, a normal distribution,

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{c}). \tag{5}$$

* rjoy@imsc.res.in

Then, $\boldsymbol{c} = \langle \boldsymbol{x}\boldsymbol{x}^T \rangle$ can be identified as the matrix of covariances in the stationary state and $\mathcal{L}P_1 = 0$ implies that the matrices $\boldsymbol{\lambda}$, $\boldsymbol{D}$ and $\boldsymbol{c}$ are not all independent but are related by the stationarity condition

$$\boldsymbol{\lambda}\boldsymbol{c} + (\boldsymbol{\lambda}\boldsymbol{c})^T = 2\boldsymbol{D}. \tag{6}$$

This is a Lyapunov matrix equation for $\boldsymbol{c}$, given $\boldsymbol{\lambda}$ and $\boldsymbol{D}$. Solutions are considerably simplified when the Fokker-Planck operator obeys detailed balance $\mathcal{L}(\boldsymbol{x})P_1(\boldsymbol{x}) = P_1(\boldsymbol{\epsilon}\boldsymbol{x})\mathcal{L}^\dagger(\boldsymbol{\epsilon}\boldsymbol{x})$, where $\mathcal{L}^\dagger$ is the adjoint Fokker-Planck operator, $\boldsymbol{\epsilon}$ is a diagonal matrix of the parities $\epsilon_i = \pm 1$ of x_i under time reversal, and the stationary distribution is time-reversal invariant, $P(\boldsymbol{x}) = P(\boldsymbol{\epsilon}\boldsymbol{x})$. This implies Onsager-Casimir symmetry $\boldsymbol{\epsilon}(\boldsymbol{\lambda}\boldsymbol{c}) = (\boldsymbol{\lambda}\boldsymbol{c})^T\boldsymbol{\epsilon}$ for the regression matrix and $\boldsymbol{\epsilon}\boldsymbol{c} = \boldsymbol{c}\boldsymbol{\epsilon}$ for the covariance matrix.

The Gauss-Markov property of the Ornstein-Uhlenbeck process ensures that the correlation function

$$\boldsymbol{C}(t-t') \equiv \langle \boldsymbol{x}(t)\boldsymbol{x}^T(t') \rangle = e^{-\boldsymbol{\lambda}|\Delta t|}\boldsymbol{c}, \tag{7}$$

decays exponentially and that its Fourier transform, the power spectral density

$$\boldsymbol{C}(\omega) = (-i\omega\mathbf{1} + \boldsymbol{\lambda})^{-1}\,(2\boldsymbol{D})\,(i\omega\mathbf{1} + \boldsymbol{\lambda}^{\boldsymbol{T}})^{-1}, \tag{8}$$

is a multivariate Lorentzian in the angular frequency ω [2].

In what follows, we shall take $\boldsymbol{\Lambda}$, the matrix exponential of the mean regressions rates and $\boldsymbol{c}$, the covariance matrix, to be the independent parameters. Estimates of the parameters in the diffusion matrix can then be obtained from the estimates of $\boldsymbol{\Lambda}$ and $\boldsymbol{c}$ through the stationarity condition. For notational brevity the set of all unknown parameters is collected in $\boldsymbol{\theta} = (\boldsymbol{\Lambda}, \boldsymbol{c})$.

III. BAYESIAN INFERENCE

Parameter Estimation: Consider now the discrete time series $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N)$, consisting of N observations of the sample path $\boldsymbol{x}(t)$ at the discrete times $t = n\Delta t$ with $n = 1, \ldots, N$. Each observation $\boldsymbol{x}_n$ is an M-dimensional vector corresponding to the number of components of the multivariate Ornstein-Uhlenbeck process. From the Markov property of the process, the probability of the path, given the parameters in $\boldsymbol{\theta}$, is

$$P\left(\boldsymbol{X}|\boldsymbol{\theta}\right) = \prod_{n=1}^{N-1} P_{1|1}\left(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n, \boldsymbol{\theta}\right) P_1\left(\boldsymbol{x}_1|\boldsymbol{\theta}\right). \tag{9}$$

The probability $P\left(\boldsymbol{\theta}|\boldsymbol{X}\right)$ of the parameters, given the sample path, is given by Bayes theorem to be

$$P\left(\boldsymbol{\theta}|\boldsymbol{X}\right) = \frac{P\left(\boldsymbol{X}|\boldsymbol{\theta}\right) P\left(\boldsymbol{\theta}\right)}{P\left(\boldsymbol{X}\right)}. \tag{10}$$

The denominator $P\left(\boldsymbol{X}\right)$ is a normalization independent of the parameters and can, so, be ignored in parameter estimation. Using informative uniform priors for $P\left(\boldsymbol{\theta}\right)$, the logarithm of the posterior probability, after using the explicit forms of $P_{1|1}$ and P_1, is

$$\begin{aligned}\ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right) = &-\frac{1}{2}\sum_{n=1}^{N-1} \boldsymbol{\Delta}_n^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}_n - \frac{1}{2}\boldsymbol{x}_1^T \boldsymbol{c}^{-1}\,\boldsymbol{x}_1 \\ &+ \frac{N-1}{2}\ln\frac{1}{(2\pi)^M\,|\boldsymbol{\Sigma}|} + \frac{1}{2}\ln\frac{1}{(2\pi)^M\,|\boldsymbol{c}|}\end{aligned}$$

where

$$\boldsymbol{\Delta}_n \equiv \boldsymbol{x}_{n+1} - \boldsymbol{\Lambda}\,\boldsymbol{x}_n. \tag{11}$$

With some elementary manipulations, detailed in the Appendix, the posterior probability can be written in terms of the following four matrix sufficient statistics

$$\boldsymbol{T}_1 = \sum_{n=1}^{N-1} \boldsymbol{x}_{n+1}\boldsymbol{x}_{n+1}^T, \quad \boldsymbol{T}_2 = \sum_{n=1}^{N-1} \boldsymbol{x}_{n+1}\boldsymbol{x}_n^T, \tag{12a}$$

$$\boldsymbol{T}_3 = \sum_{n=1}^{N-1} \boldsymbol{x}_n\boldsymbol{x}_n^T, \qquad \boldsymbol{T}_4 = \boldsymbol{x}_1\boldsymbol{x}_1^T. \tag{12b}$$

and the maximum a posteriori (MAP) estimates for mean regression rates and the covariances come out to be

$$\boldsymbol{\Lambda}^* = \boldsymbol{T}_2\boldsymbol{T}_3^{-1}, \tag{13a}$$

$$\boldsymbol{\Sigma}^* = \frac{1}{N}\left(\boldsymbol{T}_1 - \boldsymbol{T}_2\boldsymbol{T}_3^{-1}\boldsymbol{T}_2^T\right). \tag{13b}$$

The standard errors are obtained from the Hessian matrix $\boldsymbol{A}$ of the posterior probability evaluated at the maximum [3–5]. Its explicit form is provided in the Appendix. The four $M \times M$ sufficient statistic matrices $\boldsymbol{T}_i$ rather than the considerably larger $M \times N$ times series $\boldsymbol{X}$ contains all the information relevant for inference. Their use reduces both the storage and computational cost of the inference algorithm.

The estimate for the mean regression rate is easily recognized to be the multivariate generalization of our earlier result for the univariate Ornstein-Uhlenbeck process. An explicit estimate for $\boldsymbol{c}$ can be obtained from that of $\boldsymbol{\Sigma}$ when the Fokker-Planck operator obeys detailed balance. Then Onsager-Casimir symmetry $\boldsymbol{\epsilon}(\boldsymbol{\lambda}\boldsymbol{c}) = (\boldsymbol{\lambda}\boldsymbol{c})^T\boldsymbol{\epsilon}$ implies $\boldsymbol{\Lambda}\boldsymbol{c} = \boldsymbol{c}\boldsymbol{\epsilon}\boldsymbol{\Lambda}^T\boldsymbol{\epsilon}$ and, from the definition of $\boldsymbol{\Sigma}$, it follows that the MAP estimate for $\boldsymbol{c}$ is

$$\boldsymbol{c}^* = \boldsymbol{\Sigma}^*\left[\mathbf{1} - (\boldsymbol{\Lambda}^*\boldsymbol{\epsilon})^T(\boldsymbol{\Lambda}^*\boldsymbol{\epsilon})^T\right]^{-1}. \tag{14}$$

The MAP estimate for the matrix of diffusion coefficients then follows from the stationarity condition. This is the multivariate generalization of our earlier result for the diffusion coefficient for the univariate Ornstein-Uhlenbeck process. We refer to this method as "Bayes I", following the terminology of [6].

In the absence of detailed balance such explicit expressions can no longer be found and linear systems have to be solved to obtain the MAP estimate of $\boldsymbol{c}$ from that

of $\boldsymbol{\Sigma}$ and, then, to relate it to the matrix of diffusion coefficients. We shall pursue this elsewhere.

An alternative Bayesian procedure for directly estimating $\boldsymbol{c}$ is arrived at by interpreting the time series $\boldsymbol{X}$ as independent repeated samples from the stationary distribution $\mathcal{N}(\boldsymbol{0}, \boldsymbol{c})$. Using non-informative priors, the expression of the logarithms of the posterior probability in this approach is given as

$$\ln P(\boldsymbol{c}\,|\boldsymbol{X}) = \frac{N}{2} \ln \frac{1}{(2\pi)^M\,|\boldsymbol{c}|} - \frac{1}{2}\sum_{n=1}^{N} \boldsymbol{x}_n^T \boldsymbol{c}^{-1}\,\boldsymbol{x}_n, \quad (15)$$

from which the MAP estimate

$$\boldsymbol{c}^* = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{x}_n \boldsymbol{x}_n^T = \frac{1}{N}\boldsymbol{T}_3, \quad (16)$$

follows straightforwardly. The Bayesian inference method described above, using the stationary distribution of the Ornstein-Uhlenbeck process, is referred as "Bayes II". Consistency between the above two methods of estimating the covariance matrix provides a stringent test of the appropriateness of the Ornstein-Uhlenbeck process as the data generating model. The preceding equations (13), (14) and (16) are the central results of this paper.

Model comparison: Thus far we assumed that the data generating model was given and that only the parameters of the model needed to be estimated. In certain circumstances, though, the model itself may be uncertain and it becomes necessary to estimate the probability of different models $\mathcal{M}_\alpha$ [3, 7–10]. The probability of a model, given the data, is

$$P(\mathcal{M}_\alpha|\boldsymbol{X}) \propto P(\boldsymbol{X}|\mathcal{M}_\alpha)P(\mathcal{M}_\alpha), \quad (17)$$

where the first term on the right is the "evidence" of the model and the second term is the prior probability of the model. We shall assume all models to be, a priori, equally likely. The evidence is the normalizing constant in Eq.(10), given as an integral over the space of parameters $\boldsymbol{\theta}$ contained in the drift and diffusion matrices:

$$P(\boldsymbol{X}|\mathcal{M}_\alpha) = \int P(\boldsymbol{X}|\boldsymbol{\theta}, \mathcal{M}_\alpha)P(\boldsymbol{\theta}|\mathcal{M}_\alpha)d\boldsymbol{\theta}. \quad (18)$$

For unimodal posterior distributions, the height at the MAP value $\boldsymbol{\theta}^*$ times the width $\Delta\boldsymbol{\theta}$ of the distribution is, often, a very good approximation for the evidence,

$$P(\boldsymbol{X}|\mathcal{M}_\alpha) \simeq P(\boldsymbol{X}|\boldsymbol{\theta}^*, \mathcal{M}_\alpha)P(\boldsymbol{\theta}^*|\mathcal{M}_\alpha)\Delta\boldsymbol{\theta}. \quad (19)$$

The first term is the best fit likelihood while the second term, the product of the prior for the MAP estimate and the standard error of this estimate is called the Ockham factor. Thus models which achieve a compromise between the degree of fit to the data and the number of parameters required for the fit are ones which are favoured by the Bayesian model selection procedure. This avoids the over-fitting that would occur if the degree of fit was made the sole criterion for model selection and encodes the commonsense "principle of parsimony", attributed to William of Ockham, that between two models that fit the data equally well, the simpler one is to be preferred. Both the evidence and the Ockham factor can be obtained straightforwardly for the Ornstein-Uhlenbeck models and we shall make use of this below for model selection within a family of Ornstein-Uhlenbeck models.

IV. PATH SAMPLING

The solution of the Fokker-Planck equation, Eq.(3), provides a method for sampling paths of the multivariate Ornstein-Uhlenbeck process *exactly*. Given an initial state $\boldsymbol{x}$ at time t and final state $\boldsymbol{x}'$ at time t' , the quantity $\boldsymbol{x}' - \boldsymbol{\Lambda}\boldsymbol{x}$ is normally distributed with mean zero and variance $\boldsymbol{\Sigma}$, a property that was first recognized by Uhlenbeck and Ornstein. Therefore, a sequence of states at times $t = n\Delta t$, where n is a positive integer, forming a discrete sampling of a path can be obtained from the following iteration

$$\boldsymbol{x}_{n+1} = \boldsymbol{\Lambda}\boldsymbol{x}_n + \sqrt{\boldsymbol{\Sigma}}\,\boldsymbol{\xi}_n, \quad (20)$$

where $\sqrt{\boldsymbol{\Sigma}}$ is a matrix square-root of $\boldsymbol{\Sigma}$ and $\boldsymbol{\xi}_n \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{1})$ is an M-dimensional uncorrelated normal variate with zero mean and unit variance. The exponential of the regression matrix and the square-root of the variance matrix are the two key quantities in the iteration. They can be obtained analytically in low dimensional problems but for high dimensional problems they will, in general, have to be obtained numerically. The sampling interval must satisfy $\lambda_{\max}\Delta t \ll 1$ such that the shortest time scale in the dynamics, corresponding to the inverse of the largest eigenvalue $\lambda_{\max}$ of the regression matrix, is resolved in the samples. In the following section we use the method above to sample paths of the Brownian harmonic oscillator, which is equivalent to a bivariate Ornstein-Uhlenbeck process, to test the accuracy of our Bayesian estimates.

V. BROWNIAN HARMONIC OSCILLATOR

Inference problem: We now apply the results above to the physically important case of a massive Brownian particle confined in a harmonic potential described by the Langevin equation

$$m\dot{v} + \gamma v + \nabla U(x) = \xi. \quad (21)$$

Here the pair (x, v) describes the state of the particle in its phase space of position and velocity, while m and γ are the particle mass and friction coefficient respectively. The potential $U = \frac{1}{2}kx^2$ is harmonic with a stiffness k. $\xi(t)$is a zero-mean Gaussian white noise with variance $\langle \xi(t)\,\xi(t')\rangle = 2k_BT\gamma\delta(t - t')$ that satisfies the fluctuation-dissipation relation [11]. The inference

problem is to jointly estimate the triplet of parameters (m, γ, k) from discrete observations of the position and velocity and to estimate the correlation functions and the spectral densities from these observations.

Bivariate Ornstein-Uhlenbeck process: The Langevin equation can be recast as a bivariate Ornstein-Uhlenbeck process in phase space

$$d\begin{pmatrix} x \\ v \end{pmatrix} = -\boldsymbol{\lambda}\begin{pmatrix} x \\ v \end{pmatrix} dt + \sqrt{2\boldsymbol{D}}\begin{pmatrix} dW_x \\ dW_v \end{pmatrix}, \tag{22}$$

where the mean regression matrix is

$$\boldsymbol{\lambda} = \begin{pmatrix} 0 & -1 \\ \omega_0^2 & 1/\tau \end{pmatrix}, \tag{23}$$

and the diffusion matrix is

$$\boldsymbol{D} = \begin{pmatrix} 0 & 0 \\ 0 & D/\tau^2 \end{pmatrix}. \tag{24}$$

Here $\omega_0 = \sqrt{k/m}$ is the natural frequency of the undamped harmonic oscillator, $\tau = m/\gamma$ is the characteristic time scale associated with the thermalization of the momentum due to viscous dissipation, and $\omega = \sqrt{k/m - \gamma^2/(2m)^2} = \sqrt{\omega_0^2 - 1/(2\tau)^2}$ is the frequency of the damped oscillator. The diffusion coefficient of the particle is defined, as usual, by the Einstein relation $D = k_B T \gamma^{-1}$ [12, 13]. The structure of the diffusion matrix ensures that the positional Wiener process dW_x does not enter the dynamics.

At thermal equilibrium, the joint distribution of position and velocity factorize into the Gibbs distribution for the position and the Maxwell-Boltzmann distribution for the velocity to give a diagonal covariance matrix

$$\boldsymbol{c} = \begin{pmatrix} k_B T/k & 0 \\ 0 & k_B T/m \end{pmatrix}. \tag{25}$$

It is easily verified that the stationarity condition, Eq. (6), is satisfied by the above matrices. The condition of micro-reversibility translates, here, into Onsager-Casimir symmetry, $\boldsymbol{\lambda}^{\mathrm{ir}}\boldsymbol{c} = \boldsymbol{D}$, where $\lambda^{\mathrm{ir}}_{ij} = \frac{1}{2}(\lambda_{ij} + \epsilon_i\epsilon_j\lambda_{ij})$ is the irreversible part of the drift coefficient with $\epsilon_i = \pm 1$ for variables that are, respectively, even or odd under time reversal. Then, the only non-zero entry of $\boldsymbol{\lambda}^{\mathrm{ir}}$ is $\lambda^{\mathrm{ir}}_{22} = \tau^{-1}$ and it trivial to verify the Onsager-Casimir symmetry.

Path sampling: To sample paths of the Brownian harmonic oscillator exactly, it is necessary to obtain the exponential of the regression matrix $\boldsymbol{\lambda}$ and the square-root of the variance matrix $\boldsymbol{\Sigma}$. From the Cayley-Hamilton theorem the former is easily found to be

$$\boldsymbol{\Lambda} = e^{-\boldsymbol{\lambda}\Delta t} = \Lambda_1 \mathbf{1} + \Lambda_2 \boldsymbol{\lambda} \tag{26}$$

where

$$\Lambda_1 = \exp(\tfrac{-\Delta t}{2\tau})\left[\cos(\omega\Delta t) + \frac{1}{2\omega\tau}\sin(\omega\Delta t)\right], \tag{27a}$$

$$\Lambda_2 = \exp(\tfrac{-\Delta t}{2\tau})\left[-\frac{1}{\omega}\sin(\omega\Delta t)\right]. \tag{27b}$$

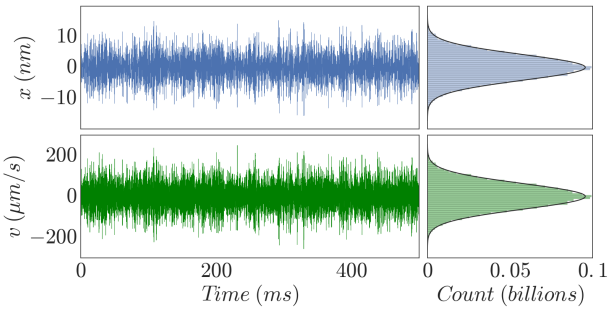

Figure 1. Sample paths (left) and histograms (right) of the position and velocity of the Brownian harmonic oscillator obtained from exact path sampling, Eq. (20). Parameters used are $m = 1\mathrm{ng}$, $k = 225\mathrm{mg}/s^2$, $\gamma = \gamma_c/10$, $\gamma_c = \sqrt{4mk}$ and $T = 275K$. The path is sampled at 2^{16} Hz.

The latter is obtained from a Cholesky factorization of $\boldsymbol{\Sigma}$ into a lower triangular matrix and its transpose,

$$\sqrt{\boldsymbol{\Sigma}} = \begin{pmatrix} s_1 & 0 \\ s_2 & s_3 \end{pmatrix}, \tag{28}$$

where the elements are

$$s_1 = \left[\frac{k_B T}{k} - \frac{k_B T}{k}\exp(\tfrac{-\Delta t}{\tau})\Big(\frac{k}{m\omega^2}\sin^2(\omega\Delta t) + \big[\cos(\omega\Delta t) + \tfrac{1}{2\omega\tau}\sin(\omega\Delta t)\big]^2\Big)\right]^{1/2}, \tag{29a}$$

$$s_2 = \frac{1}{s_1}\frac{k_B T}{\gamma\omega^2\tau^2}\exp(\tfrac{-\Delta t}{\tau})\sin^2(\omega\Delta t), \tag{29b}$$

$$s_3 = \left[\frac{k_B T}{m} - \frac{k_B T}{m}\exp(\tfrac{-\Delta t}{\tau})\Big(\frac{k}{m\omega^2}\sin^2(\omega\Delta t) + \big[\cos(\omega\Delta t) - \tfrac{1}{2\omega\tau}\sin(\omega\Delta t)\big]^2\Big) - s_2^2\right]^{1/2}. \tag{29c}$$

We use the above to explicit results in Eq. (20) to obtain exactly sampled trajectories of the Brownian harmonic oscillator.

In Fig.(1) we show a typical sample path of the positions and velocities and their corresponding histograms. The histograms of the positions and the velocities clearly show that the distributions are normal.

Parameter estimation: We now take the time series $\boldsymbol{X} = (x_1 v_1, \ldots, x_N v_N)$ of discrete observations of positions and velocities obtained from the exact path sampling and construct from it the sufficient statistics

$$\boldsymbol{T}_1 = \sum_{n=1}^{N-1}\begin{pmatrix} x_{n+1}^2 & x_{n+1}v_{n+1} \\ v_{n+1}x_{n+1} & v_{n+1}^2 \end{pmatrix}, \tag{30a}$$

$$\boldsymbol{T}_2 = \sum_{n=1}^{N-1}\begin{pmatrix} x_{n+1}x_n & x_{n+1}v_n \\ v_{n+1}x_n & v_{n+1}v_n \end{pmatrix}, \tag{30b}$$

$$\boldsymbol{T}_3 = \sum_{n=1}^{N-1}\begin{pmatrix} x_n^2 & x_n v_n \\ v_n x_n & v_n^2 \end{pmatrix}. \tag{30c}$$

$m\ (ng)$			$k\ (mg/s^2)$			$\gamma\ (\mu g/s^2)$	
Simulation	Bayes I	Bayes II	Simulation	Bayes I	Bayes II	Simulation	Bayes I
1	0.994 ± 0.006	0.993 ± 0.008	225	224.81 ± 1.72	224.49 ± 1.75	3	2.966 ± 0.021
1.5	1.521 ± 0.014	1.517 ± 0.012	250	251.97 ± 1.93	251.82 ± 1.97	1.936	1.974 ± 0.016
2	2.107 ± 0.015	2.092 ± 0.016	300	316.41 ± 2.37	314.88 ± 2.46	1.639	1.788 ± 0.015

Table I. Bayesian MAP estimates and standard errors of the parameters of the Brownian harmonic oscillator. There is excellent agreement between the two Bayesian methods and with the parameter values used to generate the paths.

From these, we compute the MAP estimate for the regression matrix

$$\boldsymbol{\lambda}^* = -\frac{1}{\Delta t} \ln \boldsymbol{\Lambda}^* = -\frac{1}{\Delta t} \ln(\boldsymbol{T}_2 \boldsymbol{T}_3^{-1}),$$

which yields the MAP estimates for the natural frequency and the relaxation time scale

$$\left(\frac{k}{m}\right)^* = \lambda_{21}^*, \quad \left(\frac{\gamma}{m}\right)^* = \lambda_{22}^*. \tag{31}$$

The MAP estimate of the covariance matrix

$$\boldsymbol{c}^* = \frac{1}{N}\left(\boldsymbol{T}_1 - \boldsymbol{T}_2 \boldsymbol{T}_3^{-1} \boldsymbol{T}_2^T\right)\left[\mathbf{1} - (\boldsymbol{T}_2 \boldsymbol{T}_3^{-1} \boldsymbol{\epsilon})^T (\boldsymbol{T}_2 \boldsymbol{T}_3^{-1} \boldsymbol{\epsilon})^T\right]^{-1}$$

yields the MAP estimate of the spring constant and the mass, in units of $k_B T$

$$\frac{k_B T}{k^*} = c_{11}^*, \qquad \frac{k_B T}{m^*} = c_{22}^*. \tag{32}$$

The friction constant is estimated by eliminating the mass between two of the previous ratios

$$\gamma^* = \frac{k_B T \lambda_{22}^*}{c_{22}^*}. \tag{33}$$

The three preceding equations provide the Bayes I map estimates of the oscillator parameters.

Next we use Eq.(25) for the covariance matrix in the bivariate analogue of Eq.(15) for the logarithm of the posterior probability. The MAP estimates and the error bars for mass m and spring constant k are then

$$\frac{k^*}{k_B T} = \frac{N}{\sum_{n=1}^N x_n^2}, \qquad \sigma_k = \frac{\sqrt{2}}{\sqrt{N}} k^*, \tag{34a}$$

$$\frac{m^*}{k_B T} = \frac{N}{\sum_{n=1}^N v_n^2}, \qquad \sigma_m = \frac{\sqrt{2}}{\sqrt{N}} m^*. \tag{34b}$$

These correspond to Bayes II estimates of the oscillator parameters.

In Table I, we provide the MAP estimates and corresponding error bars for three different time series data obtained from the simulation of the underdamped Brownian harmonic oscillator. This clearly shows that the MAP estimates of both Bayesian methods and values of the simulation parameters are in excellent agreement. The error bars have been calculated from the Hessian matrix. The Bayesian standard error in estimating the parameters is less than 1% for each case.

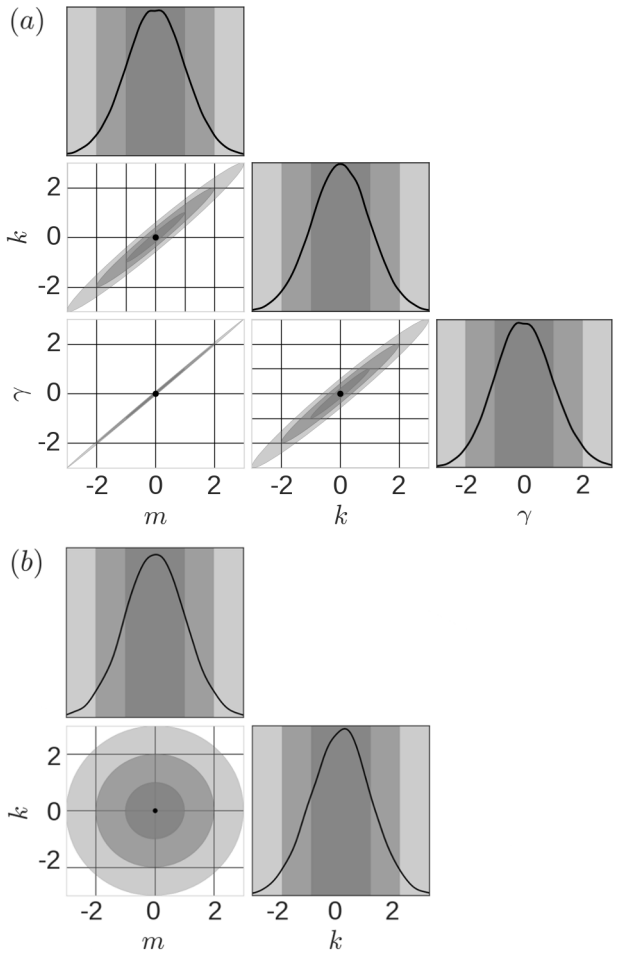

Figure 2. Corner plots of the joint posterior distribution of the mass m, friction γ and stiffness k. Panel (a) shows Bayes I, Eq.(13) while panel (b) shows Bayes II, Eq.(34). The variables have been scaled as $(x_i - \mu_i)/\sigma_{ii}$, where μ_i and σ_{ii}^2 are the MAP estimate and variance of x_i respectively. The MAP estimate is marked by a black dot, and the regions of 70%, 90% and 99% posterior probability have been shaded.

In Fig.(2) we show the posterior distribution of the parameters as corner plots [14], with panel (a) corresponding to Bayes I and panel (b) to Bayes II. The distributions have been shifted to the MAP estimates, which, therefore are always at the origin marked by the black dot, and scaled by the variances. The Bayesian credible intervals corresponding to 70%, 90% and 99% of the posterior probability are shown in shades of gray.

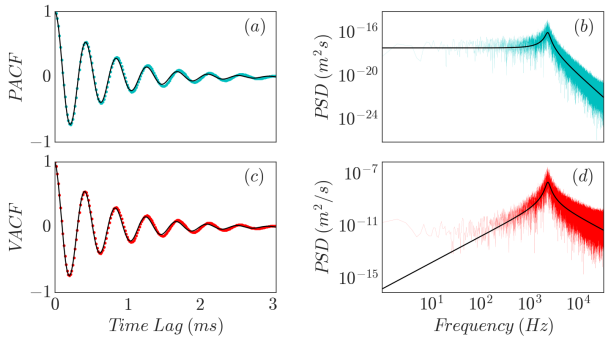

Figure 3. Bayesian estimates of the autocorrelation function and spectral density of position are shown in (a) and (b) respectively as solid lines. There is excellent agreement with simulations shown in cyan. Corresponding estimates for the velocity are shown in (c) and (d) with simulations shown in red.

Correlation functions and spectral densities: The MAP estimates of the parameters provide a novel way of estimating the correlation functions and spectral densities. Their expressions can be obtained from Eq.(7) and Eq.(8) using the explicit forms of the covariance and the regression matrices. The autocorrelations are

$$\langle x\,(t)\,x\,(0)\rangle = \frac{k_B T}{k}\exp\left(\frac{-t}{2\tau}\right)\frac{(2\omega\tau\cos\omega t+\sin\omega t)}{2\omega\tau},$$
$$\langle v\,(t)\,v\,(0)\rangle = \frac{k_B T}{m}\exp\left(\frac{-t}{2\tau}\right)\frac{(2\omega\tau\cos\omega t-\sin\omega t)}{2\omega\tau}.$$

and the spectral densities are

$$C_{xx}(\omega) = \frac{2\gamma k_B T}{m^2(\omega_0^2-\omega^2)^2+\gamma^2\omega^2},$$
$$C_{vv}(\omega) = \frac{2\gamma k_B T\omega^2}{m^2(\omega_0^2-\omega^2)^2+\gamma^2\omega^2}.$$

These expressions are evaluated at the Bayesian MAP estimates for the parameters. In Fig.(3) the result is compared with the autocorrelation function and the discrete Fourier transform of the time series. There is excellent agreement between the two autocorrelations. We emphasize that no numerical function fitting is required to obtain this estimate. The spectral density is, arguably, even more impressive as it interpolates through the noisy discrete Fourier transform in a sensible manner. Our method of estimating the spectral density shares a methodological similarity with Burg's classic maximum entropy method [15]. The principle differences are (i) that we choose the Ornstein-Uhlenbeck process as the data generating model, rather than the discrete autoregressive process assumed by Burg and (ii) that we work directly on the space of trajectories rather than in the space of correlations. From a Bayesian perspective, our estimation procedure encodes the prior information that the data is generated by an Ornstein-Uhlenbeck process and, therefore, will outperform all methods [16, 17] that do not incorporate this prior information, such as the one in [18].

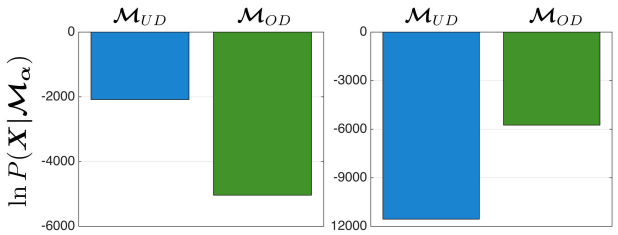

Figure 4. Bayesian model selection procedure for time series data from underdamped ($\mathcal{M}_{UD}$) and overdamped ($\mathcal{M}_{OD}$) Brownian harmonic oscillator. The logarithm of Bayesian evidence has been plotted in the natural log units (nits). The plot for underdamped data ($\gamma = \gamma_c/10$) is shown on the left, while the right panel is for overdamped data ($\gamma = 10\gamma_c$).

Model comparison: To observe the inertial motion of a colloidal particle in an optical trap, it is necessary to observe the motion at time intervals much smaller than the momentum relaxation time, that is, to ensure $\Delta t \ll \tau$. In the opposite limit, of $\Delta t \gg \tau$, the inertia of the colloid is no longer relevant and only purely diffusive motion can be observed. These correspond, respectively, to the Kramers and Smoluchowski limits of the Brownian harmonic oscillator. In experiment, it is often not a priori clear if the observational time scale places the system in the Kramers, Smoluchowski or crossover regime. Bayesian model comparison provides a principled way to answer this question, as we show below.

We consider two Ornstein-Uhlenbeck data generating models

$$\begin{aligned}\mathcal{M}_{UD} :\longrightarrow \quad & m\dot{v}+\gamma v+\nabla U=\xi \\ \mathcal{M}_{OD} :\longrightarrow \quad & \gamma\dot{x}+\nabla U=\xi\end{aligned} \tag{35}$$

corresponding, respectively, to motion on the Kramers (**u**nder**d**amped) and Smoluchoswki (**o**ver**d**amped) time scales. The latter is the formal $m \to 0$ limit of Eq.(21) and can be obtained systematically by adiabatically eliminating the velocity as a fast variable [19]. The overdamped oscillator is an univariate Ornstein-Uhlenbeck model with two parameters, γ and k, and Bayesian parameter estimation for it was presented in [6]. The two-parameter overdamped model is simpler than the three-parameter underdamped model. The question we try to answer is which of these models provides the best explanation of the data for the least number of parameters.

To do so, we compare the posterior probability of the model, given the data, by approximating the model evidence in terms of the maximum likelihood and the Ockham factors, assuming equal prior probabilities for the models as in section III. The evidence for both models is straightforwardly computed from the general expression in the Appendix. For the overdamped model the

sufficient statistics are the scalars

$$T_1 = \sum_{n=1}^{N-1} x_{n+1}^2,\ T_2 = \sum_{n=1}^{N-1} x_{n+1}x_n,\ T_3 = \sum_{n=1}^{N-1} x_n^2. \quad (36)$$

as was first pointed out in [6].

In Fig.(4), we plot the logarithm of the Bayesian evidence $\ln P(\boldsymbol{X}|\mathcal{M}_\alpha)$ in the natural log units (nits) for overdamped and underdamped Brownian motion. The panel on the left contains the plot of the logarithm of the evidence for an underdamped data while the right panel is for an overdamped data. The comparison of the $\ln P(\boldsymbol{X}|\mathcal{M}_\alpha)$ for overdamped and underdamped model clearly shows that the evidence is higher in each case for the true model of the data.

VI. SUMMARY

In summary, we have presented two Bayesian methods for inferring the parameters of a multivariate Ornstein-Uhlenbeck process, given discrete observations of the sample paths. An exact path sampling procedure has been presented and utilized to validate the Bayesian methods for the Brownian harmonic oscillator. The problem of Bayesian model comparison has been addressed and applied to select between Kramers and Smoluchowski limits of the Brownian harmonic oscillator. Future work will address the problem of parameter inference and model selection when either or both of the Markov and Gaussian properties of the process are relaxed.

APPENDIX

MAP estimates

Consider the quadratic form $\boldsymbol{\Delta_n^T \Sigma^{-1} \Delta_n} = \left(\boldsymbol{x_{n+1}} - \boldsymbol{\Lambda x_n}\right)^{\boldsymbol{T}} \boldsymbol{\Sigma^{-1}} \left(\boldsymbol{x_{n+1}} - \boldsymbol{\Lambda x_n}\right)$ that occurs in the expression of the logarithm of the posterior probability in Eq.(11). Expanding terms and completing the summation we have the following identity

$$-\frac{1}{2}\sum_{n=1}^{N-1} \boldsymbol{\Delta_n^T \Sigma^{-1} \Delta_n} = -\frac{1}{2}\boldsymbol{\Sigma}^{-1} : \Big[\Big(\boldsymbol{\Lambda} - \boldsymbol{T_2 T_3^{-1}}\Big) (\boldsymbol{T_3}) \Big(\boldsymbol{\Lambda} - \boldsymbol{T_2 T_3^{-1}}\Big)^{\boldsymbol{T}} + \Big(\boldsymbol{T_1} - \boldsymbol{T_2 T_3^{-1} T_2^T}\Big) \Big]. \quad (37)$$

Here, we have used the definitions of sufficient statistics in Eq. (12). With this identification, the expression of the logarithm of the posterior probability in terms of the sufficient statistics is

$$\ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right) = -\frac{1}{2}\boldsymbol{\Sigma}^{-1} : \Big[\Big(\boldsymbol{\Lambda} - \boldsymbol{T_2 T_3^{-1}}\Big) (\boldsymbol{T_3}) \Big(\boldsymbol{\Lambda} - \boldsymbol{T_2 T_3^{-1}}\Big)^{\boldsymbol{T}} + \Big(\boldsymbol{T_1} - \boldsymbol{T_2 T_3^{-1} T_2^T}\Big) \Big] + \frac{N}{2} \ln \frac{|\boldsymbol{\Sigma^{-1}}|}{(2\pi)^M}. \quad (38)$$

Inspecting the above expression, we recognize that the posterior is normal in $\boldsymbol{\Lambda}$ which directly yields its MAP estimate $\boldsymbol{\Lambda^* = T_2 T_3^{-1}}$. Taking the derivative of above equation with respect to $\boldsymbol{\Sigma}^{-1}$ and using the matrix identity $\partial \ln\left(\det \boldsymbol{A}\right)/\partial \boldsymbol{A} = \left(\boldsymbol{A}^T\right)^{-1}$, we obtain

$$\frac{\partial \ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right)}{\partial \boldsymbol{\Sigma^{-1}}} = -\frac{1}{2}\Big[\Big(\boldsymbol{\Lambda} - \boldsymbol{T_2 T_3^{-1}}\Big) (\boldsymbol{T_3}) \Big(\boldsymbol{\Lambda} - \boldsymbol{T_2 T_3^{-1}}\Big)^{\boldsymbol{T}} + \Big(\boldsymbol{T_1} - \boldsymbol{T_2 T_3^{-1} T_2^T}\Big) \Big] + \frac{N}{2}\boldsymbol{\Sigma}. \quad (39)$$

At the maximum, we obtain the MAP estimate as $\boldsymbol{\Sigma^*} = \frac{1}{N}\Big(\boldsymbol{T}_1 - \boldsymbol{T}_2\boldsymbol{T}_3^{-1}\boldsymbol{T}_2^T\Big)$. The estimates of the parameters of the underdamped oscillator follow from $\boldsymbol{\Lambda^*}$ and $\boldsymbol{\Sigma^*}$ after algebraic manipulations.

Standard errors and evidence

The second partial derivatives of the logarithm of the posterior distribution with respect to $\boldsymbol{\Sigma}^{-1}$ and $\boldsymbol{\Lambda}$, at the maximum, are

$$\frac{\partial^2 \ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right)}{\partial (\boldsymbol{\Sigma^{-1}})^2} = -\frac{N}{2}\left(\boldsymbol{\Sigma}^*\right)^2, \quad (40a)$$

$$\frac{\partial^2 \ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right)}{\partial \boldsymbol{\Lambda}^2} = -\boldsymbol{\Sigma}^{*-1}(\boldsymbol{T_3} + N\boldsymbol{c}*) - \frac{N}{2}\boldsymbol{\Sigma}^{*-2}((\boldsymbol{\Lambda}^*\boldsymbol{c}^* + \boldsymbol{c}^*\boldsymbol{\Lambda}^{*T}))^2. \quad (40b)$$

The mixed partial derivative at the maximum is

$$\frac{\partial^2 \ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right)}{\partial \boldsymbol{\Lambda} \partial \boldsymbol{\Sigma^{-1}}} \eqsim -\frac{N}{2}(\boldsymbol{\Lambda}^* \boldsymbol{c}^* + \boldsymbol{c}^* \boldsymbol{\Lambda}^{*T}). \tag{41}$$

The Hessian matrix, $\boldsymbol{A} = -\nabla\nabla \ln P\left(\boldsymbol{\theta}|\boldsymbol{X}\right)$, at the maximum is then

$$\boldsymbol{A} = \begin{pmatrix} \frac{N}{2}(\boldsymbol{\Sigma}^*)^2 & \frac{N}{2}(\boldsymbol{\Lambda}^* \boldsymbol{c}^* + \boldsymbol{c}^* \boldsymbol{\Lambda}^{*T}) \\ \frac{N}{2}(\boldsymbol{\Lambda}^* \boldsymbol{c}^* + \boldsymbol{c}^* \boldsymbol{\Lambda}^{*T}) & \boldsymbol{\Sigma}^{*-1}(\boldsymbol{T_3} - N\boldsymbol{c}^*) - \frac{N}{2}\boldsymbol{\Sigma}^{*-2}(\boldsymbol{\Lambda}^* \boldsymbol{c}^* + \boldsymbol{c}^* \boldsymbol{\Lambda}^{*T})^2 \end{pmatrix}. \tag{42}$$

The above expression of the Hessian matrix have been used to obtain the standard errors of the MAP estimates and the evidence of a given model.

The expression of the Bayesian evidence of a model $\mathcal{M}_\alpha$ with several parameters is given in terms of the likelihood and the Hessian matrix as [20]

$$P(\boldsymbol{X}|\mathcal{M}_\alpha) \simeq P(\boldsymbol{X}|\boldsymbol{\theta}^*, \mathcal{M}_\alpha)\, P(\boldsymbol{\theta}^*|\mathcal{M}_\alpha)[\det(\boldsymbol{A}/2\pi)]^{-1/2}. \tag{43}$$

This expression has been used to obtain the logarithm of the evidence for a M-dimensional multivariate Ornstein-Uhlenbeck model $\mathcal{M}_\alpha$. Using explicit forms of the likelihood and the Hessian matrix, at the maximum, the expression of the logarithm of the evidence for a model $\mathcal{M}_\alpha$ is given as

$$\ln P\left(\boldsymbol{X}|\mathcal{M}_\alpha\right) \simeq -\frac{N}{2}\left(\boldsymbol{\Sigma}^{*^{-1}} : \boldsymbol{\Sigma}^* + \ln((2\pi)^M |\boldsymbol{\Sigma}^*|)\right) - \frac{1}{2}\ln\left(\det(\boldsymbol{A}_{22})\det(\boldsymbol{A}_{11} - \boldsymbol{A}_{12}\boldsymbol{A}_{22}^{-1}\boldsymbol{A}_{21})\right) + \tfrac{M}{2}\ln 2\pi. \tag{44}$$

Here $\boldsymbol{A}_{ij}$ are the elements of the Hessian matrix.

[1] C. W. Gardiner. *Handbook of stochastic methods*. Springer Berlin, 1985.
[2] N. G. van Kampen. *Stochastic processes in physics and chemistry*. Elsevier, 1992.
[3] H. Jeffreys. *The theory of probability*. Oxford University Press, Oxford, 1939.
[4] E. T. Jaynes. *Probability theory: The logic of science*. Cambridge University Press, 2003.
[5] D. Sivia and J. Skilling. *Data analysis: a Bayesian tutorial*. Oxford University Press, Oxford, 2006.
[6] S. Bera, S. Paul, R. Singh, D. Ghosh, A. Kundu, A. Banerjee, and R. Adhikari. Fast Bayesian inference of optical trap stiffness and particle diffusion. *Sci. Rep.*, 7:41638, 2017.
[7] R Kashyap. A bayesian comparison of different classes of dynamic models using empirical data. *IEEE Trans. Automat. Cont.*, 22(5):715–727, 1977.
[8] D. J. C. MacKay. Bayesian interpolation. *Neur. Computn.*, 4:415–447, 1992.
[9] A. Zellner. *An introduction to Bayesian inference in econometrics*. Johm Wiley and Sons, New York, 1996.
[10] P. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences*. Cambridge University Press, 2005.
[11] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15:1–89, 1943.
[12] A. Einstein. The theory of the Brownian movement. *Ann. Phys. (Berlin)*, 322:549, 1905.
[13] R. Kubo. The fluctuation-dissipation theorem. *Rep. Prog. Phys.*, 29(1):255, 1966.
[14] D. Foreman-Mackey. corner.py: Scatterplot matrices in Python. *The Journal of Open Source Software*, 1(2):1–2, 2016.
[15] J. P. Burg. Maximum entropy spectral analysis. In *37th Meeting, Soc. of Explor. Geophys., Oklahoma City*, 1967.
[16] K. Berg-Sørensen and H. Flyvbjerg. Power spectrum analysis for optical tweezers. *Rev. Sci. Inst.*, 75(3):594–612, 2004.
[17] M. Tassieri, R. Evans, R. L. Warren, N. J. Bailey, and J. M. Cooper. Microrheology with optical tweezers: data analysis. *New J. Phys.*, 14(11):115032, 2012.
[18] S. Bera, A. Kumar, S. Sil, T. K. Saha, T. Saha, and A. Banerjee. Simultaneous measurement of mass and rotation of trapped absorbing particles in air. *Opt. Lett.*, 41(18):4356–4359, 2016.
[19] C. W. Gardiner. Adiabatic elimination in stochastic systems. i. formulation of methods and application to few-variable systems. *Phys. Rev. A*, 29(5):2814–2822, 1984.
[20] D. J. C. MacKay. *Information theory, inference and learning algorithms*. Cambridge University Press, 2003.