Fast Bayesian inference of the multivariate Ornstein-Uhlenbeck process

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

[1] *Dept of Chemical Engineering, Jadavpur University, Kolkata 700032, India*

[2] *The Institute of Mathematical Sciences-HBNI, CIT Campus, Taramani, Chennai 600113, India*

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 coefficients of the process from discretely observed sample paths. We use exact likelihoods to derive simple expressions for the maximum a posteriori estimates in terms of four sufficient statistic matrices, . We show that the dimensionality of the estimation problem is reduced when microscopic reversibility enforces Onsager-Casimir symmetry on the process parameters. We apply the method to estimate the parameters of the Langevin equation describing the inertial motion of a harmonically bound particle in a fluctuating Stokesian fluid.

I. MULTIVARIATE ORNSTEIN-UHLENBECK PROCESS

The multivariate Ornstein-Uhlenbeck is defined by the Ito stochastic differential equation [3]

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

where λ_{ij} is a positive matrix of mean regression rates, D_{ij} is a symmetric positive-definite matrix of diffusion coefficients, $W_i(t)$ are Wiener process and $i, j = 1, \ldots, M$.

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

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

The solution is a multivariate normal distribution

$$\boldsymbol{x}'t'|\boldsymbol{x}\,t \sim \mathcal{N}\left(e^{-\boldsymbol{\lambda}\Delta t}\boldsymbol{x}, \boldsymbol{c} - e^{-\boldsymbol{\lambda}\Delta t}\boldsymbol{c}\, e^{-\boldsymbol{\lambda}^T \Delta t}\right), \tag{3}$$

where $\Delta t = t - t'$ and $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma^2})$ is the multivariate normal distribution with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma^2}$. This solution is exact and holds for arbitrary values of $|\Delta t|$. The matrix $\boldsymbol{c} = \langle \boldsymbol{x}\boldsymbol{x}^T \rangle$ contains the variances and covariances in the stationary state. The correlation function decays exponentially,

$$\boldsymbol{C}(t - t') \equiv \langle \boldsymbol{x}(t)\boldsymbol{x}^T(t') \rangle = e^{-\boldsymbol{\lambda}|\Delta t|}\boldsymbol{c}, \tag{4}$$

a property guaranteed by Doob's theorem for any Gauss-Markov processes [6]. The power spectral density

$$\boldsymbol{P}(\omega) = (-i\omega \boldsymbol{I} + \boldsymbol{\lambda})^{-1} (2\boldsymbol{D}) (i\omega \boldsymbol{I} + \boldsymbol{\lambda^T})^{-1}, \tag{5}$$

is a multivariate Lorentzian in the angular frequency ω.

The matrices $\boldsymbol{\lambda}$, $\boldsymbol{D}$ and $\boldsymbol{c}$ are not all independent but are related by the stationarity condition

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

This is a Lyapunov matrix equation for $\boldsymbol{c}$, given $\boldsymbol{\lambda}$ and $\boldsymbol{D}$. When detailed balance, $\boldsymbol{\lambda c} = (\boldsymbol{\lambda c})^T$, is satisfied, the solution is $\boldsymbol{c} = \boldsymbol{\lambda}^{-1}\boldsymbol{D}$. Alternatively, if the covariance matrix is known, as is typically the case in physical applications, the variance of the noise can be obtained given the mean regression rates.

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,

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{c}). \tag{7}$$

In the following, we shall assume that process is specified by the matrix of mean regression rates $\boldsymbol{\lambda}$ and the stationary covariance matrix $\boldsymbol{c}$.

II. BAYESIAN INFERENCE

Consider now the data matrix $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_N)$, which is a time series consisting of N observations of the sample path of the multidimensional general p-variate Ornstein-Uhlenbeck process at discrete times $t = n\Delta t$ with $n = 1, 2, ..., N$. Each of the observations $\boldsymbol{x_i}$ is a p-dimensional vector, such that $\boldsymbol{x_i} = \left[x_i^{(1)}, x_i^{(2)}, ..., x_i^{(p)}\right]^T$. From the Markov property of the process, the probability of the path, given the parameters $\boldsymbol{\lambda}$ and $\boldsymbol{c}$, is

$$P(\boldsymbol{X}|\boldsymbol{\lambda}, \boldsymbol{c}) = \prod_{n=1}^{N-1} P_{1|1}(\boldsymbol{x}_{n+1}|\boldsymbol{x}_n, \boldsymbol{\lambda}, \boldsymbol{c})\, P_1(\boldsymbol{x}_1|\boldsymbol{\lambda}, \boldsymbol{c}). \tag{8}$$

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

$$P(\boldsymbol{\lambda}, \boldsymbol{c}|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|\boldsymbol{\lambda}, \boldsymbol{c})\, P(\boldsymbol{\lambda}, \boldsymbol{c})}{P(\boldsymbol{X})}.$$

* rjoy@imsc.res.in

The denominator $P(\boldsymbol{X})$ is an unimportant normalization, independent of the parameters and can be ignored for the purpose of parameter estimation. Using informative uniform priors for $P(\boldsymbol{\lambda}, \boldsymbol{c})$, the logarithm of the posterior probability, after using the explicit forms of $P_{1|1}$ and P_1, is

$$\ln P(\boldsymbol{\lambda}, \boldsymbol{c}|\boldsymbol{X}) = \frac{N-1}{2} \ln \frac{1}{(2\pi)^p |\boldsymbol{\Sigma}|} + \frac{1}{2} \ln \frac{1}{(2\pi)^p |\boldsymbol{c}|} - \frac{1}{2} \sum_{n=1}^{N-1} \boldsymbol{\Delta}_n^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Delta}_n - \frac{1}{2} \boldsymbol{x}_1^T \boldsymbol{c}\, \boldsymbol{x}_1, \tag{9}$$

where the two quantities $\boldsymbol{\Sigma}$ and $\boldsymbol{\Delta}_n$ are

$$\boldsymbol{\Sigma} \equiv \boldsymbol{c} - e^{-\boldsymbol{\lambda}\Delta t} \boldsymbol{c}\, e^{-\boldsymbol{\lambda}^T \Delta t}, \ \boldsymbol{\Delta}_n \equiv \boldsymbol{x}_{n+1} - e^{-\boldsymbol{\lambda}\Delta t} \boldsymbol{x}_n. \tag{10}$$

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

$$\boldsymbol{T}_1 = \sum_{n=1}^{N-1} \boldsymbol{x}_{n+1} \boldsymbol{x}_{n+1}^T, \quad \boldsymbol{T}_2 = \sum_{n=1}^{N-1} \boldsymbol{x}_{n+1} \boldsymbol{x}_n^T, \tag{11a}$$

$$\boldsymbol{T}_3 = \sum_{n=1}^{N-1} \boldsymbol{x}_n \boldsymbol{x}_n^T, \qquad \boldsymbol{T}_4 = \boldsymbol{x}_1 \boldsymbol{x}_1^T. \tag{11b}$$

and the maximum a posteriori estimates for the matrix of mean regression rates and the transition variation-covariance matrix can be written as

$$\boldsymbol{\lambda}^* = -\frac{1}{\Delta t} \ln(\boldsymbol{T}_2 \boldsymbol{T}_3^{-1}), \tag{12a}$$

$$\boldsymbol{\Sigma}^* = \frac{1}{N} \left(\boldsymbol{T}_1 - \boldsymbol{T}_2 \boldsymbol{T}_3^{-1} \boldsymbol{T}_2^T \right). \tag{12b}$$

The error bars can be derived by expanding the log posterior to quadratic order around the MAP estimate.

III. UNDERDAMPED BROWNIAN HARMONIC OSCILLATOR

In previous work, we have demonstrated the application of Bayesian inference to determine the harmonic trap stiffness as well as the particle friction coefficient from a time series of observed positions of an overdamped Brownian harmonic oscillator. In the overdamped limit, the inertia of the particle is neglected and hence the mass of the particle cannot be considered as a parameter for Bayesian inference. In the underdamped harmonic oscillator however, the characteristic time scale of momentum dissipation is finite and hence mass of the particle can be treated as a physical parameter for the problem. Thus, using the bivariate Ornstein Uhlenbeck process to model the position-velocity data of an underdamped Brownian harmonic oscillator, we can get a Bayesian MAP estimate for the mass of the trapped particle, alongside the trap stiffness and particle friction.

The Langevin equation for the Brownian motion of a damped harmonic oscillator of potential U is given by

$$m\dot{v} + \gamma v + \nabla U = \xi. \tag{13}$$

Here $v(t)$ is the instantaneous velocity of the Brownian particle as a function of time, m its inertial mass, γ its friction coefficient. The harmonic potential is given by $U = \frac{1}{2} k x^2$, where k is Hooke's constant and $x(t)$ is the coordinate of the oscillator as a function of time. $\xi(t)$is a zero-mean Gaussian white noise with variance $\langle \xi(t)\, \xi(t') \rangle = 2 k_B T \gamma \delta(t - t')$ as required by the fluctuation-dissipation theorem.

Equation 13 can be rewitten as two-coupled first order differential equations in position-velocity phase space,

$$\frac{d}{dt} \begin{pmatrix} x(t) \\ v(t) \end{pmatrix} = -\boldsymbol{\lambda} \begin{pmatrix} x(t) \\ v(t) \end{pmatrix} + \frac{\sqrt{2D}}{\tau} \begin{pmatrix} 0 \\ \zeta(t) \end{pmatrix}, \tag{14}$$

where $\zeta(t)$is now a zero-mean Gaussian white noise with unit variance. We have introduced Einstein's relation $D = k_B T \gamma^{-1}$ and the 2×2 matrix

$$\boldsymbol{\lambda} = \begin{pmatrix} 0 & -1 \\ \frac{k}{m} & \frac{\gamma}{m} \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ \omega_0^2 & \frac{1}{\tau} \end{pmatrix}. \tag{15}$$

Here $\omega_0 = \sqrt{k/m}$ is the cyclic frequency of the underdamped oscillator, and $\tau = m/\gamma$ is the characteristic time scale of momentum dissipation of the particle in the presence of viscous forces. The cyclic frequency of the damped oscillator is $\omega = \sqrt{k/m - \gamma^2/(4m^2)}$, which is real for less than critical damping, $\gamma^2 < 4mk$.

The stationary covariance matrix for the position-velocity phase space is given by

$$\boldsymbol{c} = \begin{pmatrix} \frac{k_B T}{k} & 0 \\ 0 & \frac{k_B T}{m} \end{pmatrix}, \tag{16}$$

where the diagonal elements are the familiar variances of the Gibbs distribution and Maxwell-Boltzmann distributions, which respectively, correspond to the equilibrium distributions of the position and velocity of the particle. Representing the position-velocity data of the underdamped oscillator as the vector time series $\boldsymbol{x_n} = [x(t = n\Delta t) \quad v(t = n\Delta t)]^T$, we can get the MAP estimates of $\boldsymbol{\lambda}$ and $\boldsymbol{c}$, and subsequently get the scalar MAP estimates for the parameters of interest, the mass of the particle, the trap stiffness and the particle friction coefficient, using elementary algebraic manipulations.

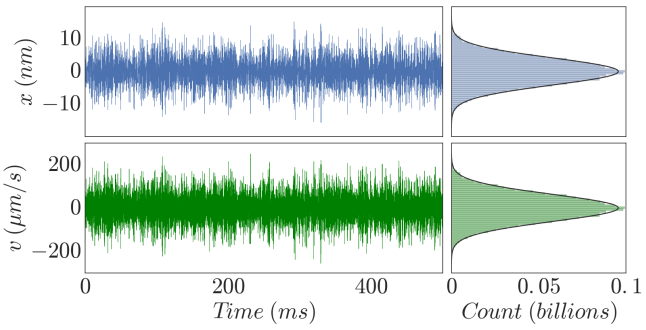

Figure 1. Positions (blue) and velocities (green) of harmonically trapped massive particle. The parameters used for simulation: $m = 1\text{ng}$, $k = 225\text{mg}/s^2$, $T = 275K$, sampling frequency = 65,536 Hz, $(x(0), v(0)) = (0, 0)$, $\gamma = \sqrt{4mk}/10$. On the right, histograms show the distribution of the position and velocity data with Gaussians overlaid.

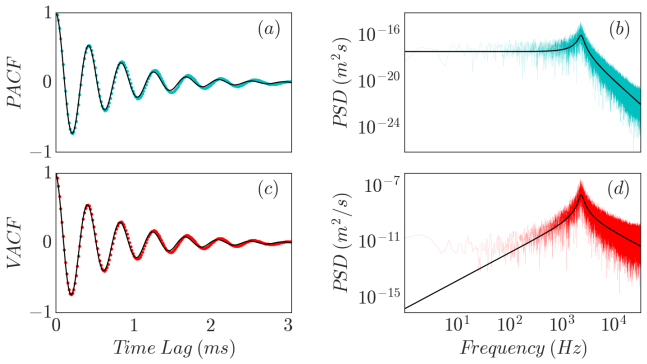

Figure 2. (a) Normalized autocorrelation function for position, (b) Power Spectrum for position, (c) Normalized autocorrelation function for velocity, (d) Power Spctrum for velocity. The autocorrelation and power spectra computed from the position and velocity data are plotted in cyan and red respectively, while the black curves represent the analytical expressions for autocorrelations and power spectra, given in eqns. 30, 31a and 31b, computed using the Bayesian MAP estimates.

An alternative Bayesian procedure for directly estimating the mass of the particle and the trap stiffness can be formulated by interpreting the time series of positions and velocities as independent repeated samples from the respective stationary distributions. At equilibrium, the likelihood can be factored into the Gibbs distribution for position and the Maxwell-Boltzmann distribution for the velocity as

$$\begin{aligned} &P(x, v|m, k) \\ &= P(x|k) \times P(v|m) \\ &= \sqrt{\frac{k}{2\pi k_B T}} \exp\left(-\frac{kx^2}{2k_B T}\right) \times \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mv^2}{2k_B T}\right). \end{aligned}$$

Using an informative prior that constrains m and k to positive values, the logarithms of the posterior probability can be written as

$$\ln P(k|x) = \frac{N}{2} \ln \frac{k}{2\pi k_B T} - \frac{1}{2} \frac{k}{k_B T} \sum x_n^2, \tag{17a}$$

$$\ln P(m|v) = \frac{N}{2} \ln \frac{m}{2\pi k_B T} - \frac{1}{2} \frac{m}{k_B T} \sum v_n^2. \tag{17b}$$

The MAP estimates, from the equilibrium distribution of the position and velocity, can be straightforwardly obtained as

$$\frac{k^*}{k_B T} = \frac{N}{\sum_{n=1}^{N} x_n^2}, \quad \frac{m^*}{k_B T} = \frac{N}{\sum_{n=1}^{N} v_n^2}. \tag{18}$$

Following the notation used in our earlier paper, we refer to this alternative method as "Bayes- II", to distinguish it from the framework described in Section II, which we refer to as "Bayes-I".

IV. UNIVARIATE CASE: OVERDAMPED BROWNIAN HARMONIC OSCILLATOR

In the *optical trap limit*, the limit of vanishing particle mass is assumed and the Langevin equation for the position of the Brownian particle reduces to

$$\dot{x} = -\frac{k}{\gamma} x + \sqrt{\frac{2k_B T}{\gamma}} \xi(t). \tag{19}$$

Defining $\lambda = \dfrac{k}{\gamma}$ as the *drift coefficient* and $D = \dfrac{k_B T}{\gamma}$ as the *diffusion coefficient*, we readily identify 19 as the univaraite analogue of 1. The application of Bayesian inference to estimate the parameters λ and D has been discussed at length in earlier work. Here, we demonstrate that the methods described in [1] are effectively the univariate extension of the inference depicted in Section II. Recycling the results from earlier sections, we see for the scalar position variable x, the *matrix* sufficient statistics reduce to corresponding scalar analogues. Finally, the MAP estimates for the drift and diffusion coefficients reduce to

$$\lambda^* = \frac{1}{\Delta t} \ln \frac{\sum x_n^2}{\sum x_{n+1} x_n},$$

$$\frac{k_B T}{k^*} \left(1 - \exp\left(-2\lambda^* \Delta t\right)\right) = \sum x_{n+1}^2 - \frac{\left(\sum x_{n+1} x_n\right)^2}{\sum x_n^2}.$$

which, after elementary manipulations can be shown to be exactly equivalent to the expressions in [1]. Thus,

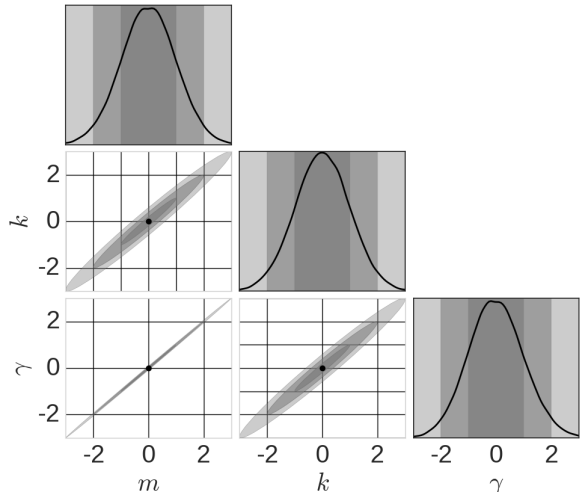

Figure 3. Pairs plot for mass m, trap stiffness k and particle drag coefficient γ obtained from the MAP estimates in 12a and 12b. The variables have been scaled using $x' = {}^{(x-\mu_x)}/_{\sigma^2_{xx}}$, where μ_x and σ^2_{xx} are the MAP estimate and variance of x respectively. The MAP estimates are marked by the black dot. The 1σ, 2σ and 3σ regions have been shaded.

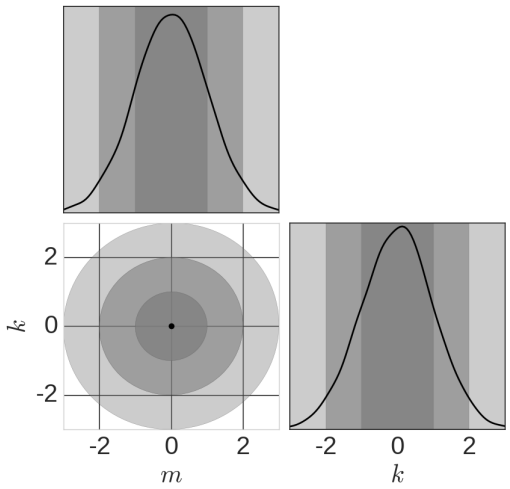

Figure 4. Pairs plot for mass m and trap stiffness k obtained from the MAP estimates in 17a and 17b. The variables have been scaled using $x' = {}^{(x-\mu_x)}/_{\sigma^2_{xx}}$, where μ_x and σ^2_{xx} are the MAP estimate and variance of x respectively. The MAP estimates are marked by the black dot. The 1σ, 2σ and 3σ regions have been shaded.

we demonstrate that our method for finding the MAP estimates is valid for arbitrary dimensionality of the data generating Ornstein Uhlenbeck process. Physically, the displacements of the Brownian particle in the overdamped harmonic oscillator constitute an univariate Ornstein Uhlenbeck process, whereas a physical example of the bivariate analogue is the position-velocity phase space for the underdamped Brownian harmonic oscillator.

m (ng)		
Simulation	Bayes-I	Bayes-II
1	0.994	0.993
1.5	1.521	1.517
2	2.107	2.092

Table I. MAP Estimates of mass

k (mg/s^2)			γ $(\mu g/s^2)$	
Simulation	Bayes-I	Bayes-II	Simulation	Bayes-I
225	224.81	224.49	3	2.966
250	251.97	251.82	1.936	1.974
300	316.41	314.88	1.639	1.788

Table II. MAP estiamtes of trap stiffness

V. OCCAM'S RAZOR

We now focus on the problem of model selection for optical trap data. In previous work, we had demonstrated parameter inference from overdamped harmonic Brownian oscillator data. In this section, the selection of the most probable data generating model between overdamped and underdamped harmonic Brownian oscillators is considered.

According to Mackay[4], the evidence or the probability $P(D|H_i)$, of obtaining data D given the hypothesis H_i is given by

$$P(D|H_i) \simeq P(D|\boldsymbol{w}_{MP}, H_i)\, P(\boldsymbol{w}_{MP}|H_i)\, \Delta\boldsymbol{w}, \qquad (20)$$

where $P(D|\boldsymbol{w}_{MP}, H_i)$ is the 'best fit likelihood' and $P(\boldsymbol{w}_{MP}|H_i)\,\Delta\boldsymbol{w}$ is known as the Occam factor. The latter term penalizes H_i for having the parameter $\boldsymbol{w}$. In the above expression, $\boldsymbol{w}_{MP}$ corrsponds to the MAP estimates of the parameters. For multiple parameters, 20 reduces to

$$P(D|H_i) \simeq P(D|\boldsymbol{w}_{MP}, H_i)\, P(\boldsymbol{w}_{MP}|H_i)\, (2\pi)^{k/2} \left(\det \boldsymbol{A}\right)^{-1/2}, \qquad (21)$$

where $\boldsymbol{A} = -\nabla\nabla \ln P(\boldsymbol{w}|D, H_i)$ is the Hessian matrix. We can recast the problem in terms of optical trap, that is we have a given position-velocity dataset, we would like to determine if the underlying Brownian motion is overdamped (H_{OD}) or underdamped (H_{UD}). Thus, based on the relative magnitude of the evidence $P(D|H_i)$, we can determine whether the data generating process is

σ^2	m	k	γ
m	1.341×10^{-27}	3.015×10^{-19}	8.069×10^{-24}
k	3.015×10^{-19}	7.143×10^{-11}	1.815×10^{-15}
γ	8.069×10^{-24}	1.815×10^{-15}	4.928×10^{-20}

Table III. Variance Covriance Matrix

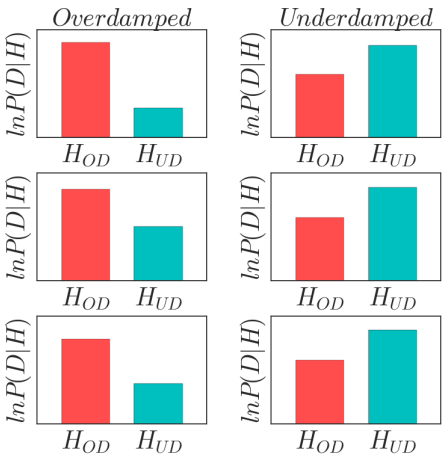

Figure 5. Relative magnitudes of the logarithm of the *evidence* for data being generated by an overdamped model (H_{OD}) and an underdamped model (H_{UD}). The panels on the left show the results of our model selection algorithm for overdamped data, while those on the right correspond to underdamped data. The Occam factor penalizes the reduntant parameter of mass (m) when the data generating model is overdamped. It is seen that the *evidence* for the overdamped model is higher when the data is generated from an overdamped system, and the same is true for the underdamped case.

overdamped Brownian motion or underdamped Brownian motion.

We focus on the stationary Bayesian estimates (labelled Bayes II in our earlier paper). Using some elementary manipulations, it can be shown that

$$\ln P\left(D|H_{UD}\right) = N\ln\frac{k^*}{2\pi k_BT} + N\ln\frac{m^*}{2\pi k_BT} - 2N + \ln\frac{4\pi m^*k^*}{N}, \tag{22a}$$

$$\ln P\left(D|H_{OD}\right) = N\ln\frac{k^*}{2\pi k_BT} - N + \ln\frac{2\pi k^*}{\sqrt{N}}. \tag{22b}$$

where m^* and k^* can be obtained from 18. Thus, if our data generating model is underdamped, we will have $\ln P\left(D|H_{UD}\right) > \ln P\left(D|H_{OD}\right)$. Similarly, for overdamped data, we will have $\ln P\left(D|H_{UD}\right) < \ln P\left(D|H_{OD}\right)$. Consider the case of overdamped data, where, the particle is considered inertia-less and $m^* \to 0$ and $\ln m^* \to -\infty$. Thus the $\ln m^*$ in 22a is severely penalized. Hence Occam's razor indicates that the single parameter k is better suited than the parameter vector (k, m) for overdamped motion.

VI. RESULTS AND DISCUSSIONS

We now present the results of our Bayesian MAP estimates applied to underdamped Brownian harmonic oscillator data. In Figure 1 we show a typical sample path of positions and velocities of one component of motion in the plane of an optical trap, together with its histogram. From the histogram it can be seen that the distributions of positions and velocities are stationary and are independently well-approximated by Gaussians. This fact can be utilized to estimate the mass m from the stationary velocity variance $\langle v^2\rangle$, and spring constant k from the stationary position variance $\langle x^2\rangle$ in the conventional equipartition method. In Figure 2, we show the position and velocity autocorrelation functions and the corresponding power spectral densities. An alternate procedure for estimating mass, trap stiffness and particle friction in optical traps has been explored by [2], where the analytical expressions of the autocorrelation function and power spectral density have been fitted to the data, and best-fit estimates are obtained. However, the error bars for these fitting based methods were as high as 15%. We demonstrate that our methods are more accurate then conventional fitting based methods, as well as order of magnitude faster than Bayesian MCMC based methods. To demonstrate the convergence of the results obtained from fitting analyting expressions to the autocorrelation or power spectral density and our Bayesian MAP estimates, we plot the analytical curves of autocorrelation function and power spectrum, using our Bayesian MAP estimates as parameter values in Figure 2. It can be seen that the results of our Bayesian method agree with the traditional fitting based methods.

The results of Bayesian inference are shown in the pairs-plots of (m, k, γ) in Figure 3. Similarly, Figure 4 represents posterior distribution in the (m, k) plane with the stationary assumption in 17a and 17b. Projections of the posterior distribution are shown with filled contours representing the error-bars. There is a single maximum at (m^*, k^*, γ^*), and the results of both the Bayesian methods are in excellent agreement with our simulation parameters, as demonstrated in Tables I and II. We characterise the error bars for our Bayesian method by the variance-covariance matrix for the posterior distribution in Table III.

In Figure 5, we have compared the logarithm of the evidence for overdamped and underdamped data sets. In each case, it is clearly seen that the relative magnitude of the evidence indicates which is the true data generating model.

APPENDIX

Derivation of MAP Estimates

We define the quantity $\mathbf{\Lambda} = \exp\left(-\boldsymbol{\lambda}\Delta t\right)$. Consider the quantity $\boldsymbol{\Delta_n^T\Sigma^{-1}\Delta_n} \equiv \left(\boldsymbol{x_{n+1}} - \mathbf{\Lambda}\boldsymbol{x_n}\right)^{\boldsymbol{T}}\boldsymbol{\Sigma^{-1}}\left(\boldsymbol{x_{n+1}} - \mathbf{\Lambda}\boldsymbol{x_n}\right)$. Expanding the expression, the quantity $\sum_{n=1}^{N-1}\boldsymbol{\Delta_n^T\Sigma^{-1}\Delta_n}$ can be written as

$$\sum_{n=1}^{N-1} \boldsymbol{\Delta_n^T \Sigma^{-1} \Delta_n}$$
$$\equiv \boldsymbol{\Sigma^{-1}} : \Big[\Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big) \boldsymbol{(T_3)} \boldsymbol{\left(\Lambda - T_2 T_3^{-1} \right)^T} + \Big(\boldsymbol{T_1 - T_2 T_3^{-1} T_2^T} \Big) \Big].$$

Substituting this form in the expression for the log-posterior

$$\begin{aligned} &\ln P\left(\boldsymbol{\lambda}, \boldsymbol{\Sigma} | \boldsymbol{X}\right) \\ &= -\frac{1}{2} \boldsymbol{\Sigma^{-1}} : \Big[\Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big) \boldsymbol{(T_3)} \Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big)^{\boldsymbol{T}} \\ &+ \Big(\boldsymbol{T_1 - T_2 T_3^{-1} T_2^T} \Big) \Big] + \frac{N}{2} \ln \frac{|\boldsymbol{\Sigma^{-1}}|}{4\pi^2}. \end{aligned} \tag{23}$$

Writing only the terms containing λ in the posterior probability density,

$$P\left(\boldsymbol{\lambda}, \boldsymbol{\Sigma} | \boldsymbol{X}\right) \propto \exp\Bigg(-\frac{1}{2} \boldsymbol{\Sigma^{-1}} : \Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big) \boldsymbol{(T_3)} \Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big)^{\boldsymbol{T}} \Bigg).$$

We recognize that the posterior is a Gaussian with respect to $\exp\left(-\boldsymbol{\lambda} \Delta t\right)$. The mean of the Gaussian yields the maximum a posteriori (MAP) estimate of $\boldsymbol{\lambda}$

Taking the derivative of 23 with respect to Σ^{-1} and using the matrix identity $\dfrac{\partial \ln\left(\det \boldsymbol{A}\right)}{\partial \boldsymbol{A}} = \left(\boldsymbol{A}^T\right)^{-1}$, we obtain

$$\begin{aligned} &\frac{\partial \ln P\left(\boldsymbol{\lambda}, \boldsymbol{\Sigma} | \boldsymbol{X}\right)}{\partial \boldsymbol{\Sigma^{-1}}} \\ &= -\frac{1}{2} \Big[\Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big) \boldsymbol{(T_3)} \Big(\boldsymbol{\Lambda - T_2 T_3^{-1}} \Big)^{\boldsymbol{T}} \\ &+ \Big(\boldsymbol{T_1 - T_2 T_3^{-1} T_2^T} \Big) \Big] + \frac{N}{2} \left(\boldsymbol{\Sigma}\right). \end{aligned} \tag{24}$$

Setting this derivative to zero, and noting that $\boldsymbol{\Lambda^* = \exp\left(-\lambda^* \Delta t\right) = T_2 T_3^{-1}}$ at the maximum a posteriori condition, we obtain the MAP estimate for $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma^*} = \frac{1}{N} \left(\boldsymbol{T}_1 - \boldsymbol{T}_2 \boldsymbol{T}_3^{-1} \boldsymbol{T}_2^T \right). \tag{25}$$

Equations 12a and 12b give the MAP estimates of the drift matrix and the transition variance covariance matrix respectively.

Once we have the MAP estimates for $\boldsymbol{\Lambda^*} = \exp\left(-\boldsymbol{\lambda^*} \Delta t\right)$ and $\boldsymbol{\Sigma^*}$, we can get m^*, k^*, γ^* using the following algebraic manipulations

$$\begin{pmatrix} \dfrac{k_B T}{k^*} \\ \dfrac{k_B T}{m^*} \end{pmatrix} = \begin{pmatrix} 1 - \Lambda_{11}^{*2} & -\Lambda_{12}^{*2} \\ -\Lambda_{21}^{*2} & 1 - \Lambda_{22}^{*2} \end{pmatrix}^{-1} \begin{pmatrix} \Sigma_{11}^* \\ \Sigma_{22}^* \end{pmatrix} \tag{26a}$$

$$\gamma^* = \frac{m^{*2} \Sigma_{12}^*}{k_B T \Lambda_{12}^{*2}} \tag{26b}$$

Simulation of position-velocity data in underdamped Brownian harmonic oscillator

The position and velocity data for underdamped Brownian harmonic oscillator were simulated using the exact statistics described in [5]. A brief overview of the method is outlined here. The recursive relation of position and velocity is given as

$$\begin{pmatrix} x_{j+1} \\ v_{j+1} \end{pmatrix} = \exp\left(-\lambda \Delta t\right) \begin{pmatrix} x_j \\ v_j \end{pmatrix} + \begin{pmatrix} \Delta x_j \\ \Delta v_j \end{pmatrix}. \tag{27}$$

It is shown that Δx_j and Δv_j are two correlated random numbers from zero-mean Gaussian distributions following the given relation for $\xi \sim \mathcal{N}\left(0,1\right)$ and $\zeta \sim \mathcal{N}\left(0,1\right)$

$$\Delta x_j = \sigma_{xx} \xi, \tag{28a}$$

$$\Delta v_j = {}^{\sigma_{xv}^2}\!/_{\sigma_{xx}} \xi_j + \sqrt{\sigma_{vv}^2 - {}^{\sigma_{xv}^4}\!/_{\sigma_{xx}^2}}\, \zeta_j. \tag{28b}$$

where

$$\begin{aligned} \sigma_{xx}^2 &= \frac{D}{4\omega^2 \omega_0^2 \tau^3} (4\omega^2\tau^2 + \exp\left(-\frac{\Delta t}{\tau}\right) \\ &\times \left[\cos\left(2\omega\Delta t\right) - 2\omega\tau \sin\left(2\omega\Delta t\right) - 4\omega_0^2\tau^2\right]), \end{aligned} \tag{29a}$$

$$\begin{aligned} \sigma_{vv}^2 &= \frac{D}{4\omega^2 \tau^3} (4\omega^2\tau^2 + \exp\left(-\frac{\Delta t}{\tau}\right) \\ &\times \left[\cos\left(2\omega\Delta t\right) + 2\omega\tau \sin\left(2\omega\Delta t\right) - 4\omega_0^2\tau^2\right]), \end{aligned} \tag{29b}$$

$$\sigma_{xv}^2 = \frac{D}{\omega^2\tau^2} \exp\left(-\frac{\Delta t}{\tau}\right) \sin^2\left(\omega\Delta t\right). \tag{29c}$$

The autocorrelation matrix is given as

$$\begin{pmatrix} \langle x\left(t\right) x\left(0\right)\rangle & \langle x\left(t\right) v\left(0\right)\rangle \\ \langle v\left(t\right) x\left(0\right)\rangle & \langle v\left(t\right) v\left(0\right)\rangle \end{pmatrix} = \frac{D}{\tau} \exp\left(\frac{-t}{2\tau}\right) \begin{pmatrix} \dfrac{\cos\omega t + \dfrac{\sin\omega t}{2\omega\tau}}{\omega_0^2} & \dfrac{\sin\omega t}{\omega} \\ -\dfrac{\sin\omega t}{\omega} & \cos\omega t - \dfrac{\sin\omega t}{2\omega\tau} \end{pmatrix}. \tag{30}$$

The power spectral density (PSD) for position and velocity are given as

$$P^{(x)}\left(f_k\right) = \frac{D/(2\pi^2)}{(2\pi\tau)^2\left(f_k^2 - f_0^2\right)^2 + f_k^2}, \tag{31a}$$

$$P^{(v)}\left(f_k\right) = \frac{2D}{(2\pi\tau)^2\left(f_k - {}^{f_0^2}/_{f_k}\right)^2 + 1}. \tag{31b}$$

[1] Sudipta Bera, Shuvojit Paul, Rajesh Singh, Dipanjan Ghosh, Avijit Kundu, Ayan Banerjee, and R Adhikari. Fast bayesian inference of optical trap stiffness and particle diffusion. *Scientific Reports*, 7, 2017.

[2] Sudipta K. Bera, Avinash Kumar, Souvik Sil, Tushar Kanti Saha, Tanumoy Saha, and Ayan Banerjee. Simultaneous measurement of mass and rotation of trapped absorbing particles in air. *Opt. Lett.*, 41(18):4356–4359, Sep 2016.

[3] Crispin W Gardiner et al. *Handbook of stochastic methods*, volume 3. Springer Berlin, 1985.

[4] David JC MacKay. Bayesian interpolation. *Neural computation*, 4(3):415–447, 1992.

[5] Simon F Nørrelykke and Henrik Flyvbjerg. Harmonic oscillator in heat bath: Exact simulation of time-lapse-recorded data and exact analytical benchmark statistics. *Physical Review E*, 83(4):041103, 2011.

[6] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.