Cover Page

Universiteit Leiden

The handle http://hdl.handle.net/1887/56250 holds various files of this Leiden University dissertation

Author: Wel, Casper van der
Title: Lipid mediated colloidal interactions
Date: 2017-10-05

Chapter 3

Force measurement from sparse trajectories in curved geometries

Abstract

The quantitative measurement of forces between microscopic particles is crucial for understanding soft materials. Here, we describe three methods to extract a potential energy landscape from the trajectories of thermally excited particles. We study the measurement accuracy of these methods using simulated data and we describe how to apply them to arbitrarily curved geometries. The first method uses direct position sampling, for which we establish what determines the optimum sampling time. The second uses uses displacement sampling. We devise selection rules that allow us to for instance isolate two-body interactions. Then we study the accuracy of maximum likelihood estimation of a piecewise interpolated interaction force using an approximate analytical form for the transition probability. The accuracy of this second method improves with increasing sampling time up to a limiting value that is set by the gradient of the force. The third method is derived from a master equation. This approach does not involve an approximate analytical model for the transition probability, but instead finds a stationary solution from binned transition probabilities. With this chapter, we provide a reference for how to extract forces from the trajectories of thermally activated particles in arbitrary geometries.

3.1 Introduction

Soft materials such as food, ceramics, cosmetics, pharmaceutics, paints, and even living organisms typically contain particles in the micrometer range. The properties of these complex materials depend sensitively on the interaction forces between these micrometer-sized constituents. Therefore, knowledge of the microscopic interaction forces is crucial for understanding soft material properties. For example, the phase behaviour of colloidal suspensions is governed by the forces between individual particles, which is described by DLVO theory*. This theory has been established experimentally using video microscopy: by following the thermal fluctuations in the positions of individual particles, interaction forces were directly observed.[65,98,99] This approach has also been successful for measuring forces between colloidal particles and a wall,[100] forces mediated by oil-water interfaces,[101] by lipid membranes (see Chapter 5), and even forces acting on single proteins in nerve cells.[102]

Throughout these works, several techniques have been employed to extract interaction forces from particle trajectories. Here, we will summarize these methods and evaluate which is most accurate, using a Brownian dynamics simulation. We will extend these methods for use in curved geometries, to be able to interpret particle movements on the surface of curved lipid membranes. We restrict ourselves to Brownian motion at low Reynolds number, which means that particle inertia does not play a role and particle dynamics are completely described by the overdamped Langevin equation.

The problem is stated as follows: given N particle trajectories of length M and sampling interval τ, what is the interaction energy $U(\vec{x})$? The meaning of the independent coordinate $\vec{x}$ can vary from experiment to experiment: for instance, in the case of particle-wall interaction,[100] $\vec{x}$ is the distance between particle and wall, and for particles that interact via a radial force, $\vec{x}$ is the distance between two particles.[99,103] As the force measurement methods remain conceptually the same for all these cases, we will first discuss the methods in one dimension only, and afterwards extend them to arbitrary number of dimensions and curved geometries.

First we will describe the direct position sampling method, in which independent observations of $\vec{x}$ are counted and $U(\vec{x})$ is induced by assuming the Boltzmann equilibrium distribution.[98] Second, we will discuss how to sample displacements instead of positions and use these to induce local forces via either an analytic model of the transition probabilities[103,104] or via a master equation.[65,99] Finally, we will discuss how to apply these methods in arbitrary geometries.

3.2 Methods

Particle trajectories were generated using a Brownian dynamics simulation implemented in Python 3.4 and Numpy 1.11. Displacements of single particles in each dimension i

*DLVO is an acronym for Derjaguin, Landau, Verweij, and Overbeek, who formulated the theory that combines the Van der Waals and electrostatic interactions between liquid-immersed charged surfaces.

were drawn from a normal distribution with width $\sigma = \sqrt{2D\tau_S}$ and mean $\mu_i = \beta F_i D \tau_S$, where the diffusion coefficient D was fixed at 0.5, the simulation time step τ_S at 0.001, and the inverse thermal energy β at 1. The force F_i was given by $-\nabla_i U$. We chose the energy U as follows:

$$\beta U(r) = \frac{5\epsilon}{4}\left[5\left(\frac{r}{\ell}\right)^4 - 4\left(\frac{r}{\ell}\right)^2\right].$$

Here, r denotes the distance to the origin, ϵ the energy well depth, and ℓ the extent of the potential field. If not stated otherwise, we used $\epsilon = 2$ and $\ell = 5$. See Fig. 3.1c for the shape of $U(r)$.

3.3 Position-based force measurement

The most straightforward way to induce forces from particle trajectories involves direct sampling of particle positions.[98] For measuring interactions between particles, the result of this technique is better known as the radial distribution function, which is typically denoted as $g(r)$. Starting from N independent observations, the number of occurrences N_i inside bin i with centre x_i and size δx are counted to estimate the probability density $\rho(x)$:

$$\frac{N_i}{N} = \int_{x_i - \frac{\delta x}{2}}^{x_i + \frac{\delta x}{2}} \rho(x)dx \approx \rho(x_i)\delta x. \tag{3.1}$$

Assuming that the observations of x are sampled at equilibrium, $\rho(x)$ can be related directly to the interaction energy $U(x)$ via the Boltzmann distribution:

$$U(x_i) = U_0 - k_B T \ln \rho(x_i), \tag{3.2}$$

in which the energy U is determined with respect to a reference value U_0.

To analyse the accuracy of this technique, we sampled $N = 2000$ non-interacting particles in a radial potential field.[†] In Figures 3.1a–b two of these simulations are shown, along with the employed radial potential wells $U(r)$ in solid lines in Fig. 3.1c. From these 2000 particle positions we straightforwardly recovered the energy landscape $U(r)$ that we used in the simulation (see points in Fig. 3.1c). The precision of this measurement is computed directly from the number of observations per bin: it can be shown that the absolute standard error in energy equals $N_i^{-1/2}$, in units of $k_B T$ (see Appendix 1 on page 46).

To assess the accuracy of this method for determining the potential energy, we define the squared deviation of the measured $U(r_i)$ with respect to the model U, averaged over all bins: $\langle \sigma_U^2 \rangle$. We will use this quantity throughout this chapter and call it "the average accuracy". In Fig. 3.1d it can be seen that this $\langle \sigma_U^2 \rangle$ is inversely proportional to the

[†]Two-dimensional radial coordinates are not directly described by Eq. 3.1. The Jacobian determinant of this coordinate system needs to be incorporated, which in this case amounts to $\delta\vec{x} = r\delta r$. See Section 3.5.

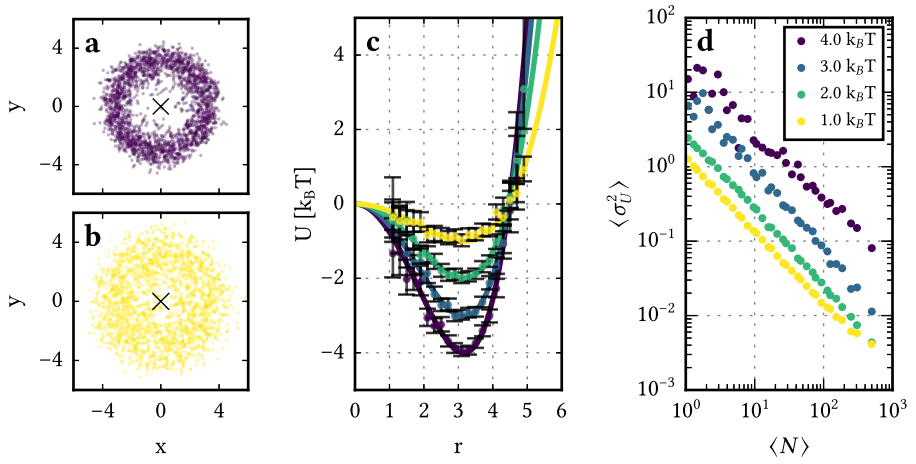

Figure 3.1. Interaction energy measurement of particles in a radial potential field at $\ell = 5$. (a, b) Simulations of $N = 2000$ particles equilibrated in a radial potential field with a well depth ϵ of (a) $4\,k_BT$ and (b) $1\,k_BT$. The potential energy $U(r)$ for four different well depths (see legend in d) are shown in (c) as solid lines. The energy was measured from the simulations and plotted as dots in (c), with error bars that denote the 2σ confidence intervals. (d) The mean squared deviation of the measured energy with respect to the model function, averaged over 200 independent simulations and over all bins. We varied the number of positions in a single simulation N between 10 and 10000, and the bin width δx between 0.05 and 0.2. On the horizontal axis, the average number of particles per bin is plotted, so that the measurements at different δx collapse onto a single curve. The observed precision follows the expected $\langle \sigma_U^2 \rangle \propto 1/\langle N \rangle$ relation.

average number of particles per bin. The absolute value of this accuracy depends on how the particles are distributed over the bins. Because the number of measurements inside a particular bin scales with the Boltzmann factor, high values of U become exponentially less frequent and therefore, the average accuracy deteriorates rapidly for increasing well depths ϵ. This is demonstrated in Fig. 3.1d: for $\epsilon = 1\,k_BT$ we need 140 independent observations per bin to reach an accuracy of $0.1\,k_BT$, while for $\epsilon = 4\,k_BT$, we already require several thousands. Thus, direct position sampling is only viable when the involved energy differences do not exceed a few units of k_BT.

Up to now, we have considered independent observations of particle positions. If we however use positions on the same particle trajectory, these are only independent at considerably long sample times, as positions only change with the square root of time. We here ask the question: what is the average accuracy when coordinates are sampled from a single trajectory with time difference τ between measurements?

Theoretically, the correlation in particle positions is governed by the transition probability $P(x, \tau|x', 0)$, which depends on $U(x)$ that is not known analytically. Therefore,

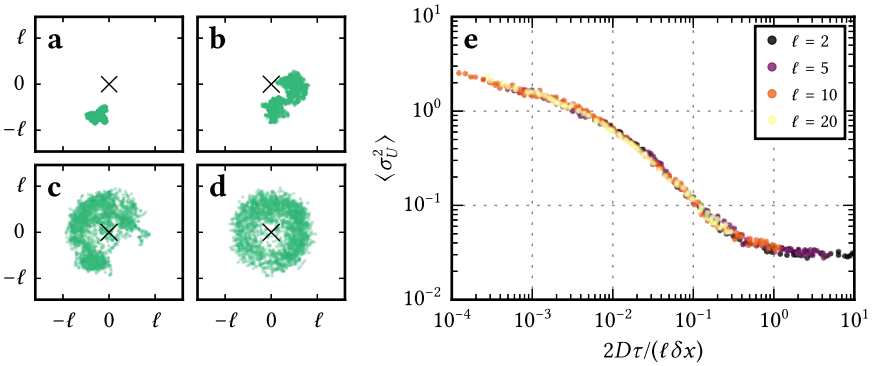

Figure 3.2. Direct sampling method applied to a single trajectory in a radial potential field. (a) to (d) show simulations of $N = 2000$ time steps with different sampling times τ of 0.001, 0.01, 0.1, and 1, respectively. The extent of correlation in position can be observed by eye from these trajectories. (e) We computed the corresponding U and averaged the deviations from the model potential over 200 independent trajectories, at a fixed $\langle N \rangle$ of 100, and for three bin widths δx ranging from 0.05 to 0.2 (plotted in the same colour). If we plot this against the dimensionless parameter $2D\tau/(\ell\delta x)$, we observe a collapse of the data onto a single curve.

we approach this question numerically: for a range of sampling times τ, we simulated 200 independent trajectories of length 2000 in a radial potential field. See Figures 3.2a–d of example trajectories with increasing sampling times. To probe the dependence of the average accuracy on τ, we fixed the average number of particles per bin $\langle N \rangle$ and simulated trajectories for different values of sampling time τ, bin width δx, and potential range ℓ. As can be seen in Figure 3.2e, we observe a data collapse when we plot the average accuracy against $2D\tau/(\ell\delta x)$. Naturally, this rescaling includes $2D\tau$, which is the squared typical distance that a particle displaces in time τ. This is then divided by two length scales: the range of the potential field ℓ, and the bin width δx, yielding a dimensionless number. From these observations we conclude that measurements are uncorrelated if $2D\tau/(\ell\delta x) \gg 1$. Such a dimensional analysis giving the minimum value of τ has been reported before,[100] however they compared $2D\tau$ with a different length scale. With our simulations, we have proven that the optimum sampling time depends on both the bin width and range of the interaction force.

When multiple particles are present in the measurement box, many-body effects have to be taken into account. Extracting two-body forces from many-particle radial distribution functions requires molecular dynamic simulations, and functions only under assumption of pairwise force additivity.[98,105] In the next section, we will show that displacement-sampling based force measurement does not have this limitation.

To summarize, direct position sampling is a straightforward method that provides interaction energies directly from independent equilibrium positions. However, the

method is only applicable for energy differences of up to a few k_BT and particle trajectories need to be sampled at sampling times $\tau \gg \ell\delta x/(2D)$ in order to ensure independent measurements.

3.4 Displacement-based force measurement

An entirely different approach to induce local forces from particle trajectories makes use of the fundamental Markov property of Brownian motion: as Brownian particles have no memory, their displacements are uncorrelated in time. Therefore, particles can act as local probes of the force field. As long as the sampling interval is above the Brownian timescale of the particles, the measured displacements are uncorrelated. The advantage of this approach is clear: the minimum sampling time is now set by the Brownian timescale, so that many independent measurements can be acquired within a limited time. Also, we need not to assume equilibrium, which allows for the measurement of larger energy differences.

3.4.1 Sampling

In literature, displacement sampling has been employed exclusively in combination with optical tweezers.[99,103] In these type of experiments, trajectories are typically of high quality, meaning that particles are observed continuously and do not disappear. If we however use displacement sampling to extract forces from particle trajectories in a limited field-of-view and with out-of-focus movement, we have to take care not to introduce a bias in the measured forces. Here, we will describe sampling selection rules that effectively remove this "sampling bias". These selection rules allow the use of displacement sampling techniques for analysing sparse trajectories.

Starting from observed particle trajectories, positions are rearranged into displacement pairs (x_0, x_1) with a given lag time. If a particle is positioned initially in the middle of the field of view, it will always be observed in the next frame (see triangle in Fig. 3.3). This allows for complete sampling of the particle displacements. However, for a particle that is positioned close to a field of view edge, we cannot sample all displacements, because we do not observe the particle if it steps outside (see diamond in Fig. 3.3). In this case, the mean of this distribution will be biased away from the field of view edge, so that there appears to exist an inwards force that acts on particles near a boundary.

To overcome this "sampling bias", we include only displacement pairs in which the initial position x_0 is such that we are certain of observing it in the next frame. In this example, we would not include x_0 that is closer than L_τ to a boundary, in which L_τ is the maximum distance a particle can move in the sampling interval. In this way, we remove the sampling bias at the expense of ignoring part of the displacement data. This displacement pair selection is crucial to employ displacement-based sampling methods to sparse trajectories.

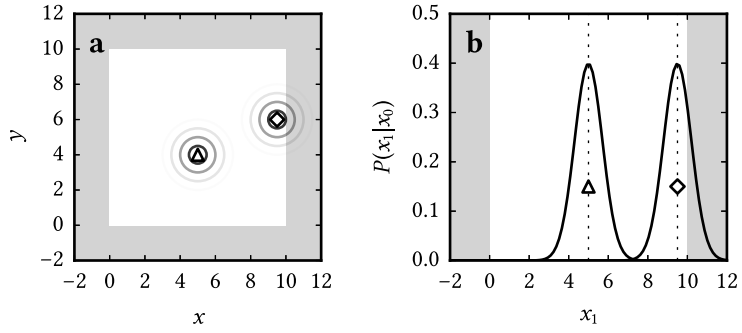

Figure 3.3. Displacement sampling near an edge. (a) A measurement box (white) with two Brownian particles. The circles denote 1, 2, and 3σ contours of the probability to find particles in the next frame. (b) Transition probability densities $P(x_1|x_0)$ of the x coordinate of the two particles, with their initial positions x_0 shown in dashed lines. The particle in the centre (triangle) will remain visible in the next frame, while the particle at the edge (diamond) can step outside of the field of view, in which case its mean displacement will be biased away from the edge. Therefore, transitions starting close to the edge should be omitted to obtain unbiased displacement samples.

As long as we do not involve the final position x_1 in the selection of displacement pairs, we do not bias the measured distribution $P(x_1|x_0)$. In other words: as long as we base the selection on x_0 only, we can apply any selection rule. This is for instance useful in the case in which particles interact via forces that are not pairwise additive, in which case we look at $P(r_1|r_0)$, with r is the distance between two particles. To isolate the two-body interaction, do not want to incorporate displacement pairs in which the two particles between which r is measured have a third particle closer by than ℓ, the length scale of the interaction. The selection rule that excludes many-body interaction is then formulated as follows: ignore displacement pairs in which r_0 has more than one particle closer than $\ell + 2L_\tau$. Thus, displacement sampling can effectively isolate two-body interactions by adopting an appropriate data selection rule, which poses a clear advantage over the direct position sampling method.

3.4.2 Maximum likelihood estimation

To extract a local force from the sampled particle displacements, a model is required that describes the probability of observing a particle at position x_1 at time $t + \tau$, given a certain initial position x_0 at time t. This so-called *transition probability* $P(x_1, t + \tau|x_0, t)$ can be approximated analytically for sufficiently small times τ, as follows [106, p. 73]:

$$P_\tau(x_1|x_0) = \frac{1}{2\sqrt{\pi D(x)\tau}} \exp\left[- \frac{(x_1 - x_0 - \beta F(x) D(x)\tau)^2}{4D(x)\tau} \right]. \tag{3.3}$$

Here, we introduced $P_\tau(x_1|x_0)$ as a shorthand for $P(x_1, t+\tau|x_0, t)$. $F(x)$ and $D(x)$ are both allowed to vary as function of x, as long as they do not vary significantly within a single displacement.

To obtain the local force and diffusivity, we can straightforwardly compute a two-dimensional histogram of the displacement pairs $\{x_j, x_{j+1}\}$ and evaluate the mean and spread of the distributions for each initial displacement.[103] A more direct approach was reported recently in ref. [104] and directly fits a model for $F(x)$ and $D(x)$ to the displacements. Here, we employ the latter approach using a constant D and a piecewise interpolated force $F_i = F(x_i)$ on points x_i spaced δx from each other. Given this model, the maximum likelihood of the parameters can then be found by maximizing the log-likelihood $\mathcal{L}$, which is the logarithm of the probability of observing all the displacements $\{x_j, x_{j+1}\}$ given model parameters D and F_i:

$$\mathcal{L}(\{x_j, x_{j+1}\}|D, F_i) = \sum_{j=1}^{N} \log P_\tau(x_{j+1}|x_j, D, F_i), \tag{3.4}$$

where P_τ is given by Eq. 3.3. By maximizing this log-likelihood, the maximum likelihood estimate (mle) of D and F_i can be found. In Appendix 2 it is shown that the thus obtained F_{mle} indeed equals the force, but that D_{mle} slightly underestimates the true D. Therefore, we chose to fix D to the known value that we used in the simulation and only optimize for F_i. Using Eq. 3.4, a Bayesian approach was adopted to sample the piecewise interpolated force. After numerical integration, the energy was obtained up to an arbitrary choice of a reference energy. In Fig. 3.4a an example of this analysis is shown, obtained from $N = 10000$ uniformly sampled displacement pairs in a radial potential field.

We observed that the average accuracy of this method is inversely proportional to the number of observed displacements N, while it does not depend on the distance between points on the piecewise interpolated force (δx). See Figure 3.4b. Below $N = 1000$, we found that the numerical minimization of the log-likelihood has a high probability to diverge.

The average accuracy is also inversely proportional to τ, which can be understood by comparing the diffusive displacement $\sqrt{2D\tau}$ with the force-induced displacement $\beta F D \tau$. For low values of τ, the diffusive displacements dominate which makes the force measurement less precise. This measurement precision in the estimated force can be derived more precisely in the form of the Cramér-Rao lower bound (see Appendix 2). The energy is obtained through numerical integration of this force, so that this fundamental precision limit propagates into the energy.

We numerically evaluated the average accuracy in the energy by changing the sampling time τ at fixed number of measurements, and found that the accuracy is indeed inversely proportional to τ. See Fig. 3.4c. The numerical integration introduced a dependence on the length scale of the potential field, as displayed in Fig. 3.4c. Here we note that the analytical model for P_τ (Eq. 3.3) becomes inaccurate when the force changes during a single particle displacement, e.g. due to a large gradient in the force. Approximately, the corresponding condition is $\tau \ll (\beta D|F'(x_i)|)^{-1}$ (see Appendix 2). Because of

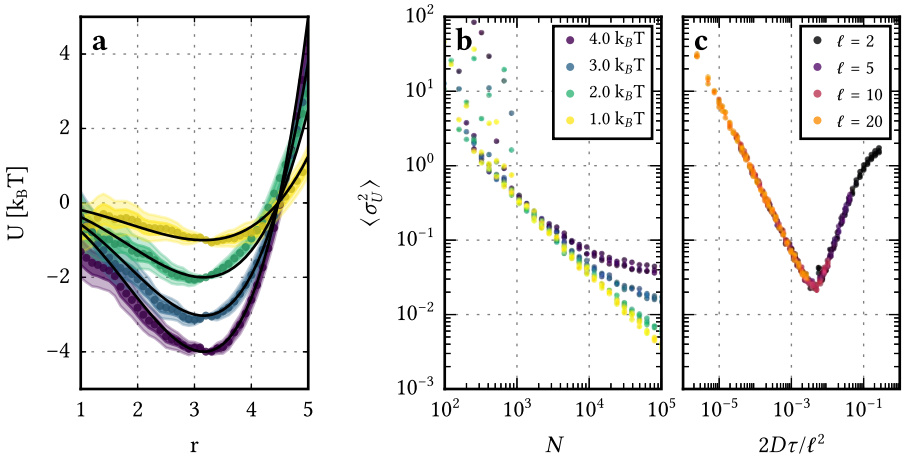

Figure 3.4. Maximum likelihood estimation of the energy applied to a single trajectory in a radial potential field. (a) Real (black line) and measured (points) potential energy curves for different well depths (see legend in b), each with $N = 10000$ measured displacements at $2D\tau = 0.05$. The dark and light shaded regions denote the 1σ and 2σ confidence intervals, respectively. The reference value of U was chosen at the potential minimum, therefore the confidence interval has zero width at that point. (b) The average accuracy of the measured energy $\langle \sigma_U^2 \rangle$, averaged over 200 independent simulations. We varied the number of displacements in a single simulation N between 100 and 100000, and the distance between the points in the piecewise interpolated force (δx) between 0.05 and 0.2, at a sampling time of $2D\tau = 0.05$. We found that the accuracy does not depend on δx and above a minimum sample size of $N = 1000$, the variance is inversely proportional to N, until it levels off due to the inaccuracy of the analytical model for P_τ. For increasing well depths, this latter issue becomes increasingly more pronounced. (c) We varied the sampling time τ for different potential length scales ℓ and for δx between 0.05 and 0.2, at fixed $N = 10000$. The average accuracy $\langle \sigma_U^2 \rangle$ was inversely proportional to $2D\tau/\ell^2$, up to a critical value of τ after which it degrades due to inaccuracy of the analytical model for the transition probability.

this effect, we observed in Fig. 3.4c that the average accuracy increases proportionally to τ for larger values of τ.

Summarizing, we have shown that the maximum likelihood estimation is capable of inducing accurate energy profiles from particle displacement data. For optimal precision, sampling time as well as the number of observations should be chosen as high as possible. There however exists an upper limit for τ, which is due to the assumption that F is constant during a displacement. Further refinement of this model may be possible by adapting the chosen model of the transition probability. For instance, the transition probability in a parabolic potential can be derived analytically (the so-called Ornstein-Uhlenbeck process[106]), which provides an analytical description up to the second derivative of the energy. Additionally, a means of combining analyses at dif-

ferent lag times into a single log-likelihood function might further improve the precision of this method.[103]

3.4.3 Master equation method

When the previously discussed conditions for the approximate model of the transition probability does not hold, the maximum likelihood estimation becomes inaccurate. To address this issue we will here discuss another approach that makes use of the master equation of ρ. This approach to measuring interaction forces has been reported by Crocker and Grier [65, 99], although not in combination with the here described sampling selection rules, which make this method more widely applicable. Here, we will shortly summarize the method and evaluate its accuracy.

In general, the evolution of a probability density can be described with a master equation, as follows:

$$\rho(x, t + \tau) = \int P_\tau(x|x')\rho(x', t)dx', \tag{3.5}$$

in which the transition probability $P_\tau(x|x')$ is related to $U(x)$, which we want to measure. To find $U(x)$, we do not need to know this relation analytically, as we can use the stationary solution of the master equation:

$$\rho_s(x) = \int P_\tau(x|x')\rho_s(x')dx'. \tag{3.6}$$

If this solution exists, then ρ_s is also a stationary solution of the Smoluchowski equation, which for constant D is equivalent to the Boltzmann distribution in Eq. 3.2. See Appendix 3 for the corresponding derivation. Thus, we can obtain $U(x)$ by sampling the transition probability matrix $P_\tau(x_1|x_0)$ from the measured displacement pairs $\{x_0, x_1\}$ and finding the eigenvector of P_τ with eigenvalue 1.

Using our simulated data, we indeed recovered an accurate $U(x)$ through this method. See Figure 3.5a. The average accuracy $\langle \sigma_U^2 \rangle$ is again inversely proportional to the number of measurements, as can be seen in Fig. 3.5b. The lower boundary of the number of observations N is determined by how the observed displacements pairs $\{x_0, x_1\}$ are distributed over the bins: every initial position bin must contain at least 1 observation to be able to compute the energy. We observed that the accuracy is inversely proportional to $\langle N \rangle$, up to a certain point after which the spatial binning starts to play a role.

In Figure 3.5c it can be seen that for low sampling time τ, the average accuracy improves for increasing $2D\tau/(\ell\delta x)$. The reason for this is the same as for the maximum likelihood method: for low τ, random displacements dominate over the force-induced displacements and therefore the force measurement is less precise. For very large τ, we observed that the accuracy converges to a constant value. In that case, the master equation method is equivalent to the direct position sampling method, because $P_\tau(x_1|x_0)$ then directly equals $\rho(x_1)$, independent of x_0.

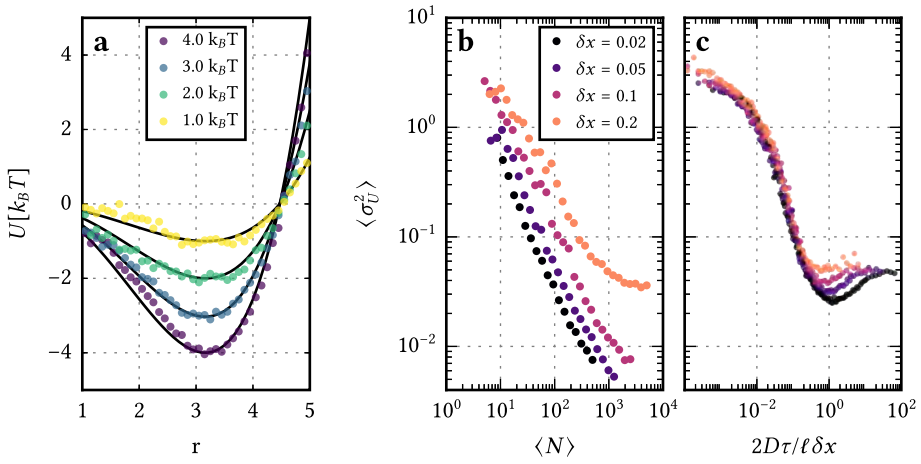

Figure 3.5. The master equation method applied to a single trajectory in a radial potential field. (a) Four examples with $N = 10000$ measured displacements, with the model potential in a black solid line and the measurement potential in coloured dots. (b) The average accuracy $\langle \sigma_U^2 \rangle$ was quantified by averaging the squared deviations of 200 independent simulations with the model function. The sampling time was fixed at $2D\tau = 0.05$, the length scale at $\ell = 5$, and the well depth at $\epsilon = 2k_BT$. The bin width was varied in between 0.02 and 0.2. The average accuracy is inversely proportional to the number of measurements. (c) We measured the average accuracy for four bin widths δx ranging from 0.02 to 0.2 (see legend in b) and four length scales ℓ (plotted in the same colour), at fixed $\langle N \rangle = 100$ and well depth $\epsilon = 2k_BT$. The accuracy improves with increasing $2D\tau/(\ell\delta x)$. Contrary to the maximum likelihood method, it does not deteriorate at larger values of τ. This is because the master equation method does not use an approximation for the transition probability.

3.5 Estimating forces in curved geometries

We have discussed three approaches for force measurement from trajectories in one dimension. Most experiments are, however, performed in multiple dimensions. For example, in sections 3.3 and 3.4, we analysed particle trajectories in a two-dimensional potential field, in which the force was dependent on the distance to the origin r. Therefore, it was useful to describe this system with radial coordinates r and θ (see Figure 3.6a). As a radial force does not depend on the angle θ, the dimensionality of the problem conveniently reduced to one.

However, it is important to note that in this curvilinear coordinate system, the infinitesimal area element dA depends on the coordinates: $dA = rdrd\theta$. This has the consequence that a particle has greater probability to move away from the origin, leading to an apparent force that is merely caused by the choice of coordinate system. In this section, we will describe how to accurately extract forces present in such coordin-

ate system for the three discussed force measurement methods. Also, we will show how to extend this to case that not only the coordinates are curvilinear, but also the geometry itself is curved, for example when a particle is confined to a spherical surface (see Figure 3.6b).

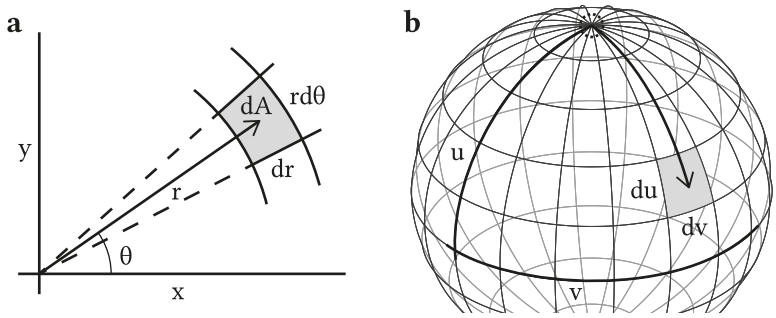

Figure 3.6. Two examples of integration in a non-Euclidean coordinate system. (a) Two-dimensional Euclidean space with a radial coordinate system $\vec{y} = \{r, \theta\}$. The infinitesimal area element dA (shaded region) is given in radial coordinates by $rd\theta dr$. (b) An intrinsically curved coordinate system $\vec{y} = \{u, v\}$ on the surface of a sphere with radius R. The origin of the coordinate system is denoted with a dotted circle. The infinitesimal area element (shaded region) at coordinates $\{u, v\}$ is given by $R^2 \sin(u) du dv$.

Position sampling When determining the probability density from the number of observations, we need to take into account the proper area element. This can be seen through substitution of variables when integrating the probability density:

$$\frac{N_i}{N} = \iint_{\delta A_i} \rho(\vec{x}) d\vec{x} = \iint_{\delta A_i} \rho(\vec{y}) J(\vec{y}) d\vec{y}, \tag{3.7}$$

where N_i is the number of observations in the two-dimensional bin δA_i and $J(\vec{y}) = |d\vec{x}/d\vec{y}|$ the determinant of the Jacobian matrix. $J(\vec{y}) = r$ in two-dimensional radial coordinates (see Figure 3.6a). For sufficiently small bins we can approximate the probability density and obtain:

$$\frac{N_i}{N} \approx \rho(\vec{y}) J(\vec{y}) \delta A, \tag{3.8}$$

where δA denotes the multiplication of the bin sizes in all dimensions. We can use this expression to extract $\rho(\vec{y})$ from a series of observations. From this perspective, the Jacobian determinant acts as a weighing function, as follows:

$$\rho(\vec{y}_i) \approx \frac{1}{N \delta A} \sum_{j=1}^{N_i} \frac{1}{J(\vec{y}_j)}, \tag{3.9}$$

where the summation sign sums over the N_i observations inside bin i. For Euclidean coordinate systems, $J = 1$, and this equation indeed reduces to Eq. 3.1.

Maximum likelihood estimation The extension of the maximum likelihood method into multiple dimensions requires an analytical form of the multidimensional transition probability. In an Euclidean geometry, this is achieved simply by multiplying the transition probabilities of the separate coordinates (Eq. 3.3). In curvilinear coordinate systems, however, we have to use a more complex expression instead, because the diffusion constant becomes a matrix. As a reference, we here provide this expression which was obtained from ref. [106, pp. 81–95].

$$\begin{aligned} P_\tau(\vec{y}|\vec{y}') = & \frac{1}{2^M\sqrt{(\pi\tau)^M \det[\tilde{D}(\vec{y})]}} \times \\ & \exp\left[-\frac{1}{4\tau}[\tilde{D}(\vec{y})]^{-1}_{ij}(\Delta y_i - \tilde{F}_i(\vec{y})\tau)(\Delta y_j - \tilde{F}_j(\vec{y})\tau)\right], \end{aligned} \tag{3.10}$$

in which repeated indices are summed and M denotes the number of dimensions. $\tilde{F}_i(\vec{y})$ and $\tilde{D}_{ij}(\vec{y})$ are obtained from the diffusion constant $D(\vec{x})$ and force vectors $F_i(\vec{x})$ in a Euclidean metric, as follows:

$$\tilde{F}_i(\vec{y}) = \sum_j \frac{\partial y_i}{\partial x_j}\beta F_j(\vec{x})D(\vec{x}) + \sum_j \frac{\partial^2 y_i}{\partial x_j^2}D(\vec{x}), \tag{3.11}$$

$$\tilde{D}_{ij}(\vec{y}) = \sum_k \frac{\partial y_i}{\partial x_k}\frac{\partial y_j}{\partial x_k}D(\vec{x}). \tag{3.12}$$

In the example of a two-dimensional radial force, it follows that $\tilde{D}_{rr} = D$, $\tilde{D}_{\theta\theta} = D/r^2$, $\tilde{F}_r = D(\beta F_r + 1/r)$, and $\tilde{D}_{\theta r} = \tilde{D}_{r\theta} = \tilde{F}_\theta = 0$. In the corresponding analysis of trajectories in a two-dimensional radial potential (Fig. 3.4) we observed that this approach provided the same results as when performing the analysis in Euclidean coordinates.

Master equation The transition probabilities in a curved geometry can be sampled in a fashion similar to the direct position sampling method, as is shown conveniently from the master equation (Eq. 3.5). If we transform from an Euclidean coordinate system ($\vec{x}$) to another ($\vec{y}$), we obtain the following master equation:

$$\rho(\vec{y}, t+\tau)J(\vec{y}) = \int P_\tau(\vec{y}|\vec{y}')\rho(\vec{y}', t)J(\vec{y}')d\vec{y}' \tag{3.13}$$

$$\rho(\vec{y}, t+\tau) = \int \left[P_\tau(\vec{y}|\vec{y}')\frac{J(\vec{y}')}{J(\vec{y})}\right]\rho(\vec{y}', t)d\vec{y}'. \tag{3.14}$$

Therefore, $\rho_s(\vec{y})$ is not an eigenvector of the probability matrix $P_\tau(\vec{y}|\vec{y}')$, but of the rescaled probability matrix shown between brackets in Eq. 3.14. In our example of 2D radial coordinates, we thus have to weigh a displacement from r_0 to r_1 with r_0/r_1 when estimating the transition probability matrix. This approach is reminiscent of the projection method that is described in ref. [99], however by expressing this in the Jacobian determinant, we extended the method into arbitrarily curved geometries.

Curved geometry Some processes not only require a curvilinear coordinate system, but also take place on a geometry that has an intrinsic curvature. For example in Chapter 5, we extract a two-body interaction force from particle trajectories that are confined to a spherical surface. See Figure 3.6b for the corresponding coordinate system, in which we fixed one of the particles in the origin. In this case the intrinsic curvature of the geometry should be accounted for, which is possible through the embedding of the curved surface into three-dimensional flat space. In the example of a spherical surface, this embedding is formulated as follows:

$$\begin{cases} x_1 = & R\cos v \sin u \\ x_2 = & R\sin v \sin u \\ x_3 = & R\cos u \end{cases} \tag{3.15}$$

Here, $\{x_1, x_2, x_3\}$ denote the three-dimensional Euclidean coordinates, R the curvature of the sphere, and $\{u, v\}$ the internal coordinates of the surface. An analytical model for the transition probability on the curved geometry can be acquired directly from these equations by evaluating the partial derivatives and substituting them in equations 3.10–3.12.

For the direct position sampling (Eq. 3.9) and the master equation (Eq. 3.14) methods, the infinitesimal area element J on the surface needs to be evaluated. This can be done through the metric tensor $g_{\mu\nu}$ as follows [107, pp. 88-90]:

$$g_{\mu\nu} = \sum_i \frac{\partial x_i}{\partial y_\mu}\frac{\partial x_i}{\partial y_\nu}, \tag{3.16}$$

$$J = \sqrt{\det g_{\mu\nu}}. \tag{3.17}$$

In the example of a spherical surface parametrized by $\vec{y} = \{u, v\}$, it follows that $g_{\mu\nu} = \{\{R^2, 0\}, \{0, R^2 \sin^2 u\}\}$, so that $J = R^2 \sin u$.

3.6 Summary and Conclusion

We have discussed three different techniques of measuring forces from the trajectories of Brownian particles in arbitrary geometries.

The first approach samples the equilibrium positions and computes the energy profile directly through the Boltzmann factor. We found that this approach requires that

particle positions are uncorrelated in time, with the condition $2D\tau \gg \ell\,\delta x$, with τ the sampling time, D the diffusion coefficient, ℓ the length scale of the potential, and δx the bin width. Provided that particle positions are uncorrelated, the measurement precision can be estimated directly from N_i, the number of particles per bin: $\sigma_{U,i} = k_B T/\sqrt{N_i}$. If enough independent samples from an equilibrium distribution are available, this straightforward approach is the method of choice.

A fundamentally different method uses displacement pairs instead of positions. This does not require the particle positions to be equilibrated, and as displacements are uncorrelated in time, higher sampling times are possible. This allows for force measurement from sparse data sets as well as selection of particular parts of the trajectory if necessary. Here presented data selection rules are able to remove artefacts such as three-body interactions without introducing a bias.

Displacements can be fitted directly to an analytical model of the transition probability. As this maximum likelihood approach is based on a log-likelihood maximization, we can use Bayesian methods to estimate the confidence intervals for the force. We found that the accuracy of the method improves with increasing sampling time τ, as long as forces are constant during each observed displacement. The corresponding condition for τ was found to be $\tau \ll (\beta D|F'(x_i)|)^{-1}$, with β the reciprocal thermal energy, and F' the gradient of the force.

If this condition cannot be met, a master equation approach can be used. This method is also based on the measurement of particle displacements, however it does not assume a model for the transition probability. Although this is less constrained than the direct fit to an analytical model of the transition probability, it does not provide a straightforward way to estimate the measurement precision. Therefore, when the sampling can be performed sufficiently fast, the transition probability method is preferred.

Appendix 1: Uncertainty in position sampling

The variance in the number of occurrences N_i inside a single bin is given by the Poisson distribution. For sufficiently large N_i, this is approximated by $\sigma_N^2 = N_i$. As ρ is proportional to N_i, the relative standard deviation in ρ is given by:

$$\frac{\sigma_\rho}{\rho} = \frac{\sigma_N}{N_i} = \frac{1}{\sqrt{N_i}}. \tag{3.18}$$

Using a Taylor series and Eq. 3.2, the standard deviation in U is then given as follows:

$$\sigma_U \approx \sigma_\rho \left| \frac{d}{d\rho} \left[k_B T \ln\left(\frac{\rho}{\rho_0}\right) \right] \right| = k_B T \frac{\sigma_\rho}{\rho} = \frac{k_B T}{\sqrt{N_i}}. \tag{3.19}$$

The average accuracy $\langle \sigma_u^2 \rangle$ is then directly given by Eq. 3.19 and depends on how the particles are distributed over the bins:

$$\langle \sigma_u^2 \rangle_{bins} = k_B T \langle N_i^{-1} \rangle_{bins}. \tag{3.20}$$

Appendix 2: Uncertainty in maximum likelihood estimation

The log-likelihood given by Equations 3.3 and 3.4 provides a maximum likelihood estimation of the parameters D and F. As Eq. 3.3 is a normal distribution with mean $\mu = \beta F D \tau$ and variance $\sigma^2 = 2D\tau$, the expected values of the maximum likelihood estimates of μ and σ^2 are given as follows:[108]

$$E[\mu_{mle}] = \mu, \tag{3.21}$$

$$E[\sigma_{mle}^2] = \frac{N-1}{N}\sigma^2, \tag{3.22}$$

where N denotes the number of measurements. The expected value of μ_{mle} is unbiased, in contrary to the expected value of σ_{mle}^2. Therefore, we prefer to determine the diffusion constant D separately and use this as a fixed constant in the maximum likelihood estimation. In that case, the fundamental limit in the measurement precision of μ_{mle} is given by the Cramér-Rao lower bound:[108]

$$\mathrm{var}[\mu_{mle}] = \frac{\sigma^2}{N}, \tag{3.23}$$

from which the expected variance in the measured force is readily obtained:

$$\mathrm{var}[\beta F_{mle}] = 2(ND\tau)^{-1}. \tag{3.24}$$

Using Eq. 3.24, we can estimate the precision in each point of the piecewise approximation for $F(x)$. The observed accuracies and precisions are plotted against the expected precision $2(N_i D\tau)^{-1}$ in Figure 3.7. We observed that this expression systematically underestimated the precision by a factor of 1.7.

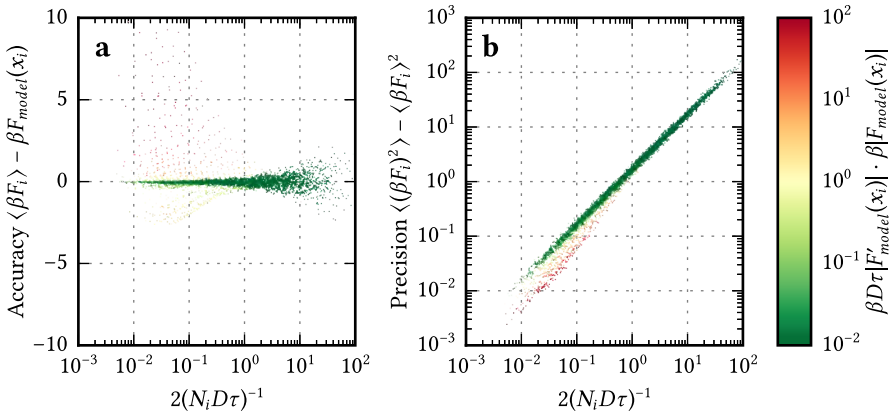

Figure 3.7. Accuracy and precision the maximum likelihood estimate of the piecewise interpolated force, for each $F_i = F_{mle}(x_i)$ separately These were obtained from 200 independent simulations with ℓ ranging from 2 to 20, δx from 0.05 to 0.2, and τ from 0.001 to 1. N_i is the number of observations within δx of x_i. On average, the observed precision was 1.7 times the expected precision. The observations are coloured based on the force gradient in x_i, multiplied with the root mean squared expected displacement in that point. For large force gradients, the accuracy clearly degrades.

In this figure, it is also shown that this method becomes inaccurate if the local force gradients become too large. From a Taylor expansion of F around x_i, we estimate that the inaccuracy due to a gradient in given by:

$$|\langle F_i \rangle - F_{model}| \approx |\Delta x F'(x_i)| = \sqrt{2D\tau + (\beta F(x_i) D\tau)^2} |F'(x_i)|. \tag{3.25}$$

To find a useful expression for the upper limit of τ, we take the approximation $(\beta F)^2 \gg 2/(D\tau)$. When looking for an upper limit for τ, this approximation is often valid in at least one point on the interaction curve. We then obtain the following expression for the accuracy in F_i:

$$\left| \frac{\langle F_i \rangle - F_{model}}{F_{model}} \right| \approx |\beta F(x_i) D\tau F'(x_i)|, \tag{3.26}$$

which provides an upper limit for the sampling time: $\tau \ll (\beta D |F'(x_i)|)^{-1}$.

Appendix 3: Stationary solution of the Smoluchowski equation

We start with the Smoluchowski equation in a curved geometry:

$$\partial_t \rho = \nabla_\mu (D_{ij} \partial_\nu - \beta D_{ij} F_\nu) \rho. \tag{3.27}$$

Here, ∂_i denotes the partial derivative with respect to i, ∇_μ the covariant derivative, D_{ij} the diffusion matrix, and β the inverse of the thermal energy. As the diffusion matrix is a contravariant tensor,[106] we can replace D_{ij} with $D g^{\mu\nu}$ if diffusion is isotropic in an Euclidean coordinate system. Here $g^{\mu\nu}$ is the inverse of the metric tensor. Therefore, using the Einstein notation:

$$\partial_t \rho = D \nabla_\mu (g^{\mu\nu} \partial_\nu - \beta g^{\mu\nu} F_\nu) \rho. \tag{3.28}$$

The force is related to the energy through differentiation, $F_\nu = -\partial_\nu U$. We now find a stationary solution, using the identity $\nabla_\mu g^{\mu\nu} = 0$ [107, p. 99]:

$$\begin{aligned}
0 &= \nabla_\mu (g^{\mu\nu} \partial_\nu - \beta g^{\mu\nu} F_\nu) \rho \\
&= (\partial_\nu \nabla_\mu g^{\mu\nu} + g^{\mu\nu} \nabla_\mu \partial_\nu - \beta F_\nu \nabla_\mu g^{\mu\nu} - \beta g^{\mu\nu} \nabla_\mu F_\nu) \rho \\
&= (g^{\mu\nu} \nabla_\mu \partial_\nu - \beta g^{\mu\nu} \nabla_\mu F_\nu) \rho \\
&= g^{\mu\nu} \nabla_\mu (\partial_\nu \rho + \beta \rho \partial_\nu U) \\
\rho &= \rho_0 \exp(-\beta U).
\end{aligned}$$