INSTITUTE OF PHYSICS PUBLISHING INVERSE PROBLEMS

Inverse Problems **19** (2003) 725–738 PII: S0266-5611(03)53160-4

Determining the Gaussian probability distribution of the best-fit ellipsoid of revolution for a polymer chain from planar projections

Yu Zhou[1], **Denis Wirtz**[2] **and Gregory S Chirikjian**[1]

[1] Department of Mechanical Engineering, The Johns Hopkins University, 3400 N Charles Street, Baltimore, MD 21218, USA
[2] Department of Chemical and Biomolecular Engineering, The Johns Hopkins University, 3400 N Charles Street, Baltimore, MD 21218, USA

E-mail: yuzhou@titan.me.jhu.edu, wirtz@jhu.edu and gregc@jhu.edu

Received 9 September 2002, in final form 7 April 2003
Published 2 May 2003
Online at stacks.iop.org/IP/19/725

Abstract

This paper introduces an algorithm to obtain the probability distribution of the best-fit ellipsoid of an individual random-walk polymer chain from planar projections. This algorithm allows the true three-dimensional behaviour of polymer chains to be inferred from planar experimental measurements such as those obtained by light microscopy. Our algorithm is effective in retrieving the Gaussian probability density functions (PDFs) of evolving best-fit prolate and oblate ellipsoids with axial symmetry. The implementation depends on the properties of the projections of rigid ellipsoids from which the desired PDFs can be retrieved by optimization. Detailed derivation of the relevant formulae is provided, and examples are given to help to explain the algorithm. An error analysis shows that the algorithm has good properties of convergence and robustness. Therefore, the proposed algorithm is applicable to the analysis of data obtained from experiments.

1. Introduction

As one of the most important gross physical characteristics of an individual random-walk polymer, its shape has attracted many researchers' attention. Theoretical predictions [1], computational studies [2–7] and experiments [8, 9] have indicated that the shape of an individual long random-walk polymer is anisotropic. However, previous research has focused mainly on the general properties of the shape of instantaneous polymer conformations. An analysis of the distribution of polymer shapes is lacking. This is important because, in solution, the shape of a polymer evolves over a certain range of sizes and aspect ratios. Some researchers simulated conformations of a random-walk chain with numerical methods and calculated the shape distribution [2, 3], but the purpose of those studies was to prove the asphericity

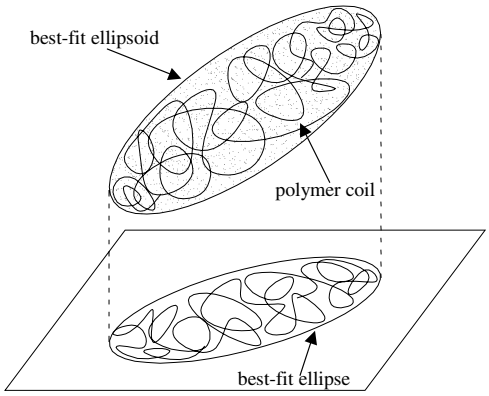

Figure 1. The relation between the best-fit ellipse and the best-fit ellipsoid.

of random walks. So far little effort has been devoted to studying the probability distribution of the shape of an individual random-walk polymer from experimental data. This is partially because the quantities of interest cannot be directly measured experimentally. As a step forward, this paper proposes a computational method that will allow one to reconstruct the distribution of spatial behaviours from planar experimental data.

Recently, researchers have used biopolymers such as actin [10, 11], keratin [12] and DNA [13, 14] to test basic models of polymer physics. These polymers are typically fluorescently labelled, allowed to fluctuate in solution and observed using light microscopy. Microscopy captures two-dimensional (2D) projections of the three-dimensional (3D) shape and motion of the polymer (figure 1) [13]. However, light microscopy techniques can only provide a sequence of planar projections [7, 8]. No currently available method allows one to recover 3D shape and dynamical information from 2D projections. Hence, an inverse problem must be solved to recover the true range of 3D shape variations.

To verify the shape anisotropy of random walks, previous researchers have used the so-called best-fit-ellipse method to approximate the shape of a long random-walk polymer observed from experiments [8, 9]. The general idea is that from the planar projection of an individual polymer, one can find a best-fit ellipse that encompasses the entire projection of the polymer. This idea is easily extended to the 3D situation. That is, to use an ellipsoid to fit the spatial conformation of the polymer. A best-fit ellipse can be considered as a planar projection of a best-fit ellipsoid of revolution (figure 1), which is either prolate or oblate.

However, because a polymer in solution keeps changing its shape, the size, orientation and aspect ratio of the best-fit ellipsoid, which encompasses the polymer, also change over a certain range. The changing best-fit ellipsoid reflects the changing shape of the polymer. Because of optical diffraction limitations, it is very difficult, if not impossible, to depict the precise contour of a tangled polymer chain in experiments. Therefore, the best-fit ellipsoid provides valuable information about the shape evolution of the polymer chain. Because the best-fit ellipsoid keeps changing, it is significant from the perspective of statistical mechanics to determine the probability distribution of its shape.

This paper shows that it is possible to retrieve the probability distribution of the evolving shape of the best-fit ellipsoid with a large enough set of best-fit ellipses constructed from the observed planar projections of an individual polymer. Due to limits on computing power and the amount of data that can be reasonably collected from microscopy experiments, this paper makes the simplifying assumption that the best-fit ellipsoid remains prolate or oblate during its evolution process, and that the evolution of the best-fit ellipsoid follows a Gaussian distribution.

As one can see from the following sections, a robust algorithm is developed based on these assumptions. Given an axisymmetric ellipsoid which is randomly tumbling in space and whose shape is changing over time, and given a finite number of snapshots of planar projections of this ellipsoid, we seek to reconstruct the statistics of the 3D shape fluctuations. An effective algorithm is introduced to retrieve the corresponding probability density function (PDF).

2. General principles

The main objective of this work is to develop an algorithm to retrieve the spatial probability distribution of the evolving best-fit ellipsoid of a fluctuating polymer chain. A natural characterization of an ellipsoid is the lengths of its three semi-axes, denoted by a, b and c respectively with $a \leqslant b \leqslant c$. Therefore, the tri-variate PDF $f(a, b, c)$ is used to describe the evolution of its shape. Meanwhile, a number of planar projections of the evolving best-fit ellipsoid, the best-fit ellipses, can be obtained from microscopy experiments. Denoting the length of the minor semi-axis of the best-fit ellipse as $l_{\min}$, and the length of the major semi-axis as $l_{\max}$, one can obtain the corresponding bi-variate PDF $f(l_{\min}, l_{\max})$ by calculating the normalized frequencies of occurrence of the ellipses with different $l_{\min}$ and $l_{\max}$ [16]. With $f(l_{\min}, l_{\max}|a, b, c)$ denoting the PDF of $l_{\min}$ and $l_{\max}$ of a rotating rigid ellipsoid with specfic values of a, b and c, there exists the relationship

$$f(l_{\min}, l_{\max}) = \int_a \int_b \int_c f(l_{\min}, l_{\max}|a, b, c) f(a, b, c)\, \mathrm{d}a\, \mathrm{d}b\, \mathrm{d}c. \tag{1}$$

This is the starting point for all the computations in this paper. The objective is to solve the inverse problem of computing $f(a, b, c)$ given $f(l_{\min}, l_{\max})$ and $f(l_{\min}, l_{\max}|a, b, c)$.

For an evolving ellipsoid of revolution where two axes have equal length all the time, $f(a, b, c)$ is reduced to the bi-variate PDF $f(a, b)$, where a denotes the short semi-axis/axes, and b the long semi-axis/axes. Correspondingly, equation (1) is reduced to

$$f(l_{\min}, l_{\max}) = \int_b \int_a f(l_{\min}, l_{\max}|a, b) f(a, b)\, \mathrm{d}a\, \mathrm{d}b. \tag{2}$$

An evolving prolate ellipsoid has one long axis and two equal short axes. As a result, $l_{\min} = a$, and $l_{\max}$ changes between a and b (figure 2). Moreover, the relationship between a and $l_{\min}$ is bijective, which results in the univariate PDF

$$f(a) = f(l_{\min}) = \int_{l_{\max}} f(l_{\min}, l_{\max})\, \mathrm{d}l_{\max}. \tag{3}$$

The conditional PDF $f(l_{\max}|a)$ is computed from Bayes' rule as

$$f(l_{\max}|a) = f(l_{\max}|l_{\min}) = \frac{f(l_{\min}, l_{\max})}{f(l_{\min})} = \frac{f(a, l_{\max})}{f(a)}. \tag{4}$$

One can obtain $f(l_{\max}|a)$ by

$$f(l_{\max}|a) = \int_b f(l_{\max}, b|a)\, \mathrm{d}b = \int_b f(l_{\max}|a, b) f(b|a)\, \mathrm{d}b. \tag{5}$$

Given $f(l_{\min}, l_{\max})$ generated from experiments, one can compute $f(a)$ by equation (3), and then $f(l_{\max}|a)$ by equation (4). As derived in the following sections, $f(l_{\max}|a, b)$ has a closed form formula, and $f(b|a)$ can be retrieved by using an optimization method [15, 17, 18]. Then, the desired PDF $f(a, b)$ is obtained by

$$f(a, b) = f(a) f(b|a). \tag{6}$$

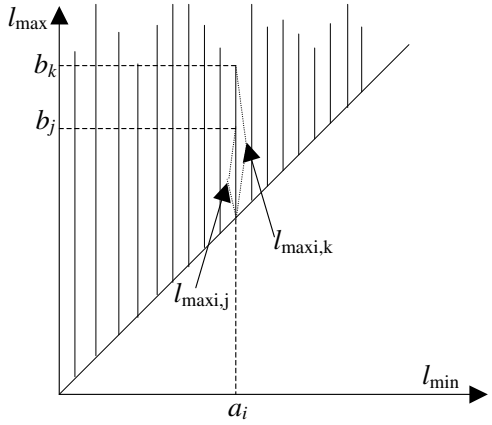

Figure 2. Relationship between the axes of prolate ellipsoids and those of the projected ellipses.

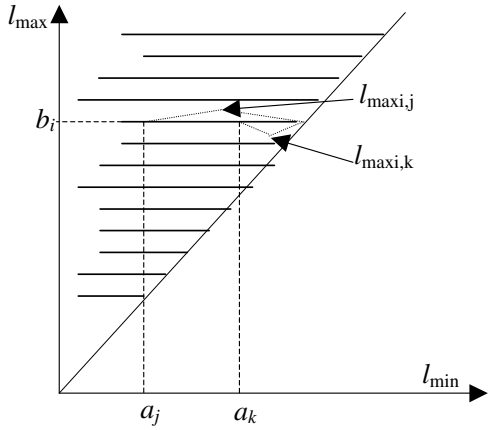

Figure 3. Relationship between the axes of oblate ellipsoids and those of the projected ellipses.

In contrast, an evolving oblate ellipsoid has one short axis and two equal long axes. As a result, $l_{max} = b$, and l_{min} varies between a and b (figure 3). Equations (3)–(6) change correspondingly into

$$f(b) = f(l_{max}) = \int_{l_{min}} f(l_{min}, l_{max})\,\mathrm{d}l_{min}, \tag{7}$$

$$f(l_{min}|b) = f(l_{min}|l_{max}) = \frac{f(l_{min}, l_{max})}{f(l_{max})} = \frac{f(l_{min}, b)}{f(b)}, \tag{8}$$

$$f(l_{min}|b) = \int_a f(l_{min}, a|b)\,\mathrm{d}a = \int_a f(l_{min}|a, b)\,f(a|b)\,\mathrm{d}a, \tag{9}$$

$$f(a, b) = f(b)\,f(a|b). \tag{10}$$

The following section focuses on the computation of $f(b|a)$ of the evolving prolate ellipsoid and $f(a|b)$ of the evolving oblate ellipsoid.

3. Core of the algorithm

In the case of the evolving prolate ellipsoid, we assume that $f(b|a)$ is a Gaussian distribution

$$f(b|a) = \frac{1}{\sqrt{2\pi}\sigma(a)} \exp\left(-\frac{(b-\mu(a))^2}{2\sigma^2(a)}\right), \tag{11}$$

where $\mu(a)$ denotes the mean value of b given a, and $\sigma(a)$ is the standard deviation. To determine $f(b|a)$, one only needs to compute $\mu(a)$ and $\sigma(a)$. To compute $\mu(a)$ and $\sigma(a)$, an effective algorithm is derived below. The core of the algorithm is an optimization method derived from equation (5), with knowledge of $f(l_{\max}|a)$ and $f(l_{\max}|a,b)$.

To solve for $f(b|a)$, one can discretize the domain of b into N b_i values with equal increment Δb. Equation (5) changes into a discrete form

$$\begin{aligned} f(l_{\max}|a) &= \sum_{i=1}^{N} f(l_{\max}|a,b_i) f(b_i|a)\Delta b \\ &= \sum_{i=1}^{N} f(l_{\max}|a,b_i) w(b_i|a) = \boldsymbol{f}(l_{\max}|a,\boldsymbol{b})^{\mathrm{T}} \boldsymbol{w}(\boldsymbol{b}|a), \end{aligned} \tag{12}$$

where $w(b_i|a) = f(b_i|a)\Delta b$ is the probability of occurrence of the intermediate rigid prolate ellipsoid with short semi-axis of length a and long semi-axis of length b_i, $\boldsymbol{b}$ is the vector of N b_i values, $\boldsymbol{f}(l_{\max}|a,\boldsymbol{b})$ is the vector of N $f(l_{\max}|a,b_i)$ values and $\boldsymbol{w}(\boldsymbol{b}|a)$ is the vector of N $w(b_i|a)$ values. The set of ellipses generated from the evolving best-fit ellipsoid can be considered as originated from *a set of intermediate rigid ellipsoids*. To find the desired $\boldsymbol{w}(\boldsymbol{b}|a)$, an optimization method is applied.

From equation (12), one obtains

$$\int_{l_{\max}} (f(l_{\max}|a) - \boldsymbol{f}(l_{\max}|a,\boldsymbol{b})^{\mathrm{T}} \boldsymbol{w}(\boldsymbol{b}|a))^2 \,\mathrm{d}l_{\max} = 0, \tag{13}$$

of which the discretized form is

$$\sum_{j=1}^{M} (f(l_{\max\, j}|a) - \boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})^{\mathrm{T}} \boldsymbol{w}(\boldsymbol{b}|a))^2 \Delta l_{\max} = 0, \tag{14}$$

where $l_{\max}$ is discretized into M $l_{\max\, j}$ values.

Equation (14) can be expanded as

$$\begin{aligned} &\sum_{j=1}^{M} f(l_{\max\, j}|a)^2 - 2\left(\sum_{j=1}^{M} f(l_{\max\, j}|a)\boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})^{\mathrm{T}}\right)\boldsymbol{w}(\boldsymbol{b}|a) \\ &\qquad + \boldsymbol{w}(\boldsymbol{b}|a)^{\mathrm{T}} \sum_{j=1}^{M} \boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})\boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})^{\mathrm{T}} \boldsymbol{w}(\boldsymbol{b}|a) = 0, \end{aligned} \tag{15}$$

which can be written in the simple form

$$C - 2\boldsymbol{d}^{\mathrm{T}}\boldsymbol{w}(\boldsymbol{b}|a) + \boldsymbol{w}(\boldsymbol{b}|a)^{\mathrm{T}}\boldsymbol{H}\boldsymbol{w}(\boldsymbol{b}|a) = 0, \tag{16}$$

where $C = \sum_{j=1}^{M} f(l_{\max\, j}|a)^2$ does not contain $\boldsymbol{w}(\boldsymbol{b}|a)$, $\boldsymbol{d} = \sum_{j=1}^{M} f(l_{\max\, j}|a)\boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})^{\mathrm{T}}$ and $\boldsymbol{H} = \sum_{j=1}^{M} \boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})\boldsymbol{f}(l_{\max\, j}|a,\boldsymbol{b})^{\mathrm{T}}$. From equation (16), a quadratic form optimization problem can be constructed: namely, to find the vector $\boldsymbol{w}(\boldsymbol{b}|a)$ which minimizes

$$M = \tfrac{1}{2}\boldsymbol{w}(\boldsymbol{b}|a)^{\mathrm{T}}\boldsymbol{H}\boldsymbol{w}(\boldsymbol{b}|a) - \boldsymbol{d}^{\mathrm{T}}\boldsymbol{w}(\boldsymbol{b}|a), \tag{17}$$

subject to the constraints

$$0 \leqslant w(b_i|a) \leqslant 1, \tag{18}$$

and

$$\sum_{i=1}^{N} w(b_i|a) = 1. \tag{19}$$

Equations (18) and (19) must be satisfied because the w values are probabilities. By using a quadratic programming optimization method, the desired w values can be computed.

From the computed w values, the mean of b can be obtained by

$$\mu(a) = \sum_{i=1}^{N} b_i w(b_i|a), \tag{20}$$

and the standard deviation of b can be obtained by

$$\sigma(a) = \sqrt{\sum_{i=1}^{N} (b_i - \mu(a))^2 w(b_i|a)}. \tag{21}$$

Then the desired Gaussian PDF is constructed by equation (11).

In the case of the evolving oblate ellipsoid, we assume that $f(a|b)$ is a Gaussian distribution

$$f(a|b) = \frac{1}{\sqrt{2\pi}\sigma(b)} \exp\left(-\frac{(a-\mu(b))^2}{2\sigma^2(b)}\right), \tag{22}$$

where $\mu(b)$ denotes the mean of a given b, and $\sigma(b)$ the standard deviation. To compute $f(a|b)$, one only needs to compute $\mu(b)$ and $\sigma(b)$. Discretizing the domain of a into N a_i values with equal increment Δa, and denoting by $w(a_i|b) = f(a_i|b)\Delta b$, $\boldsymbol{a}$ the vector of N a_i values, $\boldsymbol{f}(l_{\min}|\boldsymbol{a}, b)$ the vector of N $f(l_{\min}|a_i, b)$ values and $\boldsymbol{w}(\boldsymbol{a}|b)$ the vector of N $w(a_i|b)$ values, the quadratic form optimization problem changes into

$$\text{minimizing } M = \tfrac{1}{2}\boldsymbol{w}(\boldsymbol{a}|b)^{\mathrm{T}}\boldsymbol{H}\boldsymbol{w}(\boldsymbol{a}|b) - \boldsymbol{d}^{\mathrm{T}}\boldsymbol{w}(\boldsymbol{a}|b), \tag{23}$$

subject to

$$0 \leqslant w(a_i|b) \leqslant 1, \tag{24}$$

and

$$\sum_{i=1}^{N} w(a_i|b) = 1, \tag{25}$$

where $\boldsymbol{d} = \sum_{j=1}^{M} f(l_{\min\, j}|b)\boldsymbol{f}(l_{\min\, j}|\boldsymbol{a}, b)^{\mathrm{T}}$, and $\boldsymbol{H} = \sum_{j=1}^{M} \boldsymbol{f}(l_{\min\, j}|\boldsymbol{a}, b)\boldsymbol{f}(l_{\min\, j}|\boldsymbol{a}, b)^{\mathrm{T}}$.

To construct the quadratic form (17) and (23), one must know $f(l_{\max}|a, b)$ for prolate ellipsoids and $f(l_{\min}|a, b)$ for oblate ellipsoids in advance. The next section derives the closed form formula of $f(l_{\max}|a, b)$ and $f(l_{\min}|a, b)$.

4. Derivation of $f(l_{\max}|a, b)$ and $f(l_{\min}|a, b)$

In this section, the closed form formula for $f(l_{\max}|a, b)$ for prolate ellipsoids and that of $f(l_{\min}|a, b)$ for oblate ellipsoids are derived. To derive them, the closed form formulae of $l_{\max}$ and $l_{\min}$ are required respectively.

4.1. Closed form formulae for $l_{\max}$ and $l_{\min}$

The equation of a general ellipsoid is

$$(\boldsymbol{r} - \boldsymbol{r_0})^{\mathrm{T}} \boldsymbol{R} \boldsymbol{A} \boldsymbol{R}^{\mathrm{T}} (\boldsymbol{r} - \boldsymbol{r_0}) = 1, \tag{26}$$

where $\boldsymbol{r}$ denotes the position vector of a point on the surface of the ellipsoid, $\boldsymbol{r}_0$ the position vector of the centre, $\boldsymbol{R} = \boldsymbol{R}(\alpha, \beta, \gamma)$ is the rotation matrix describing the orientation of the ellipsoid in space, α, β and γ are the *ZXZ* Euler angles which define the orientation [19], and

$$\boldsymbol{A} = \begin{pmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{b^2} & 0 \\ 0 & 0 & \frac{1}{c^2} \end{pmatrix}. \tag{27}$$

The contour of the planar projection of the ellipsoid is an ellipse which consists of all the planar projections of those points on the surface of the ellipsoid (figure 1). For a prolate ellipsoid, $a < b$ and

$$\boldsymbol{A} = \begin{pmatrix} \frac{1}{a^2} & 0 & 0 \\ 0 & \frac{1}{a^2} & 0 \\ 0 & 0 & \frac{1}{b^2} \end{pmatrix}. \tag{28}$$

One can obtain the lengths of the two semi-axes of the ellipse as

$$l_{\min} = a, \tag{29}$$

and

$$l_{\max} = \sqrt{a^2 + (b^2 - a^2)\sin^2\beta}. \tag{30}$$

Equations (29) and (30) reflect the fact that when a rigid prolate ellipsoid rotates in space, the length of the minor axis of its planar projection does not change. Only the length of the major axis changes.

For an oblate ellipsoid, one simply switches the positions of a and b in equation (28), and obtains the following $l_{\min}$ and $l_{\max}$:

$$l_{\max} = b, \tag{31}$$

and

$$l_{\min} = \sqrt{b^2 + (a^2 - b^2)\sin^2\beta}, \tag{32}$$

where, as the rigid oblate ellipsoid rotates, the length of the major axis of its planar projection does not change, only the length of the minor axis changes.

4.2. Closed form formulae of $f(l_{\max}|a, b)$ and $f(l_{\min}|a, b)$

To compute $f(l_{\max}|a, b)$ for a prolate ellipsoid, one can sample its orientation uniformly, and calculate the frequency of occurrence of each $l_{\max}$ value. Sampling the orientation uniformly means the uniform distribution of the orientation, i.e.

$$\rho(\boldsymbol{R}(\alpha, \beta, \gamma)) = c, \tag{33}$$

where c is a constant. From

$$\begin{aligned} \int_{SO(3)} \rho(\boldsymbol{R}(\alpha, \beta, \gamma))\, \mathrm{d}\boldsymbol{R} &= \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \rho(\boldsymbol{R}(\alpha, \beta, \gamma)) \sin\beta \, \mathrm{d}\alpha \, \mathrm{d}\beta \, \mathrm{d}\gamma \\ &= \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} c \sin\beta \, \mathrm{d}\alpha \, \mathrm{d}\beta \, \mathrm{d}\gamma \\ &= 1, \end{aligned} \tag{34}$$

one can obtain

$$\rho(\boldsymbol{R}(\alpha, \beta, \gamma)) = \frac{1}{8\pi^2}. \tag{35}$$

Moreover,

$$\begin{aligned}\rho(\beta) &= \int_0^{2\pi}\int_0^{2\pi} \rho(\boldsymbol{R}(\alpha, \beta, \gamma)) \sin\beta \,\mathrm{d}\alpha \,\mathrm{d}\gamma \\ &= \int_0^{2\pi}\int_0^{2\pi} \frac{1}{8\pi^2} \sin\beta \,\mathrm{d}\alpha \,\mathrm{d}\gamma \\ &= \frac{\sin\beta}{2}.\end{aligned} \tag{36}$$

Because $0 \leqslant \beta \leqslant \pi$, one can obtain two solutions of β from equation (30),

$$\beta_1 = \arcsin\sqrt{\frac{l_{\max}^2 - a^2}{b^2 - a^2}}, \tag{37}$$

and

$$\beta_2 = \pi - \arcsin\sqrt{\frac{l_{\max}^2 - a^2}{b^2 - a^2}}. \tag{38}$$

This results in the following relationship of probability:

$$|f(l_{\max}|a, b)\,\mathrm{d}l_{\max}| = |\rho(\beta_1)\,\mathrm{d}\beta_1| + |\rho(\beta_2)\,\mathrm{d}\beta_2|, \tag{39}$$

from which one obtains

$$\begin{aligned}f(l_{\max}|a, b) &= \rho(\beta_1)\left|\frac{\mathrm{d}\beta_1}{\mathrm{d}l_{\max}}\right| + \rho(\beta_2)\left|\frac{\mathrm{d}\beta_2}{\mathrm{d}l_{\max}}\right| \\ &= \frac{l_{\max}}{\sqrt{(b^2 - a^2)(b^2 - l_{\max}^2)}}.\end{aligned} \tag{40}$$

For an oblate ellipsoid, one obtains two solutions of β from equation (32),

$$\beta_1 = \arcsin\sqrt{\frac{b^2 - l_{\min}^2}{b^2 - a^2}}, \tag{41}$$

and

$$\beta_2 = \pi - \arcsin\sqrt{\frac{b^2 - l_{\min}^2}{b^2 - a^2}}. \tag{42}$$

Following a similar derivation as the prolate case, one obtains

$$f(l_{\min}|a, b) = \frac{l_{\min}}{\sqrt{(b^2 - a^2)(l_{\min}^2 - a^2)}}. \tag{43}$$

5. Error analysis

The main source of error in our reconstruction method arises in the computation of $f(b|a)$ and $f(a|b)$. There are two categories of errors that should not be ignored. One is the error caused by incompleteness of the set of intermediate rigid ellipsoids which is employed to compute the $w(b_i|a)$ values and $w(a_i|b)$ values; the other is the result of error in the input data which is mostly caused by inaccuracies in experimental measurements.

In the following subsections, we present the results of numerical experiments with synthetic measured data and examine the robustness of our technique to these two sources of error.

5.1. Error from incompleteness

The evolving best-fit ellipsoid changes its size continuously, which means that a large number of intermediate rigid ellipsoids need to be used to retrieve the accurate $f(b|a)$ and $f(a|b)$. In practice, one cannot determine the number of those intermediate ellipsoids in advance. Therefore, an approximation with fewer intermediate ellipsoids is expected. In this situation, some error will arise because the set of intermediate ellipsoids may not be large enough to reflect the evolution of the best-fit ellipsoid completely. It is desirable for the retrieved $f(b|a)$ and $f(a|b)$ to converge to the actual ones as the number of intermediate ellipsoids increases.

In the case of prolate ellipsoids, the convergence of the proposed algorithm can be tested as follows.

(1) Values of a and the upper bound of b, $b_{\max}$, are chosen so that b changes between a and $b_{\max}$. The domain of b is discretized into N b_i values with equal increment Δb, which corresponds to N intermediate ellipsoids, where N should be a certain large integer.
(2) The mean $\mu_0(a)$ and standard deviation $\sigma_0(a)$ of the corresponding Gaussian PDF $f(b|a)$ are chosen. Corresponding probability density values of $f(b_i|a)$ are calculated. Then $w(b_i|a)$ values are calculated by $w(b_i|a) = f(b_i|a)\Delta b$.
(3) Having derived the closed form formulae of $f(l_{\max}|a, b_i)$ values, one constructs $f(l_{\max}|a)$ using equation (12).
(4) Assuming that the PDF $f(l_{\max}|a)$ computed in step (3) is obtained from experiments, the domain of b is re-discretized into K b_j values which correspond to K intermediate ellipsoids with $1 \leqslant K \leqslant N$. Then the corresponding quadratic form is constructed and the $w(b_j|a)$ values are computed by a quadratic programming optimization algorithm, and the μ and σ are calculated. The calculated μ and σ are compared with μ_0 and σ_0 to see the difference.
(5) Repeating step (4) with $K = 1$ to N, one obtains the relationship between the number of intermediate ellipsoids and the errors in μ and σ. It reflects the property of convergence of the algorithm.

As an example, we set $a = 2$, $b_{\max} = 12$, $N = 100$, $\mu_0 = 7$ and $\sigma_0 = 1.5$, and construct the simulated experimental PDF as step (3). Then we retrieve the μ and σ from a set with one intermediate ellipsoid to a set with 100 intermediate ellipsoids, and calculate the errors in μ and σ. The plots in figure 4 show the quick convergence in μ and σ, which means that the algorithm can be implemented with a relatively small set of intermediate ellipsoids. As one would expect, the errors are close to zero when the number of intermediate ellipsoids increases to 100.

The convergence in the case of oblate ellipsoids is similar to that in the case of prolate ellipsoids, and similar plots of convergence are obtained as those in figure 4.

5.2. Error from input

The process of making experimental measurements also introduces errors that affect the computation of μ and σ. It is important to know the impact of the input error on the output error. That is, it is important to know how sensitive our reconstruction method is to input error. The test procedure for the output error caused by input error is similar to that for testing the error from incompleteness. Here, the simulated error-corrupted $f(l_{\max}|a, b)$ and $f(l_{\min}|a, b)$ are used in the test.

In the case of prolate ellipsoids, by an input error e, it is meant that for an actual size $l_{\max}$, the measured size $l'_{\max}$ could be a value between $l_{\max} - e$ and $l_{\max} + e$ with a certain

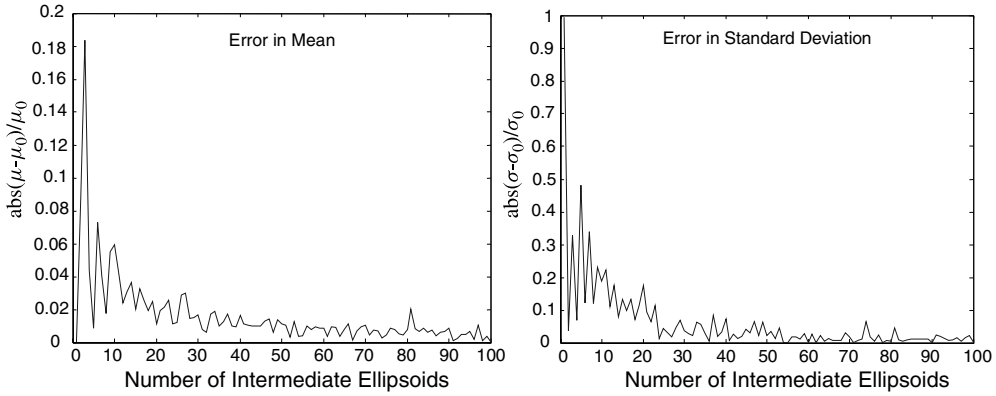

Figure 4. Convergence in mean and standard deviation.

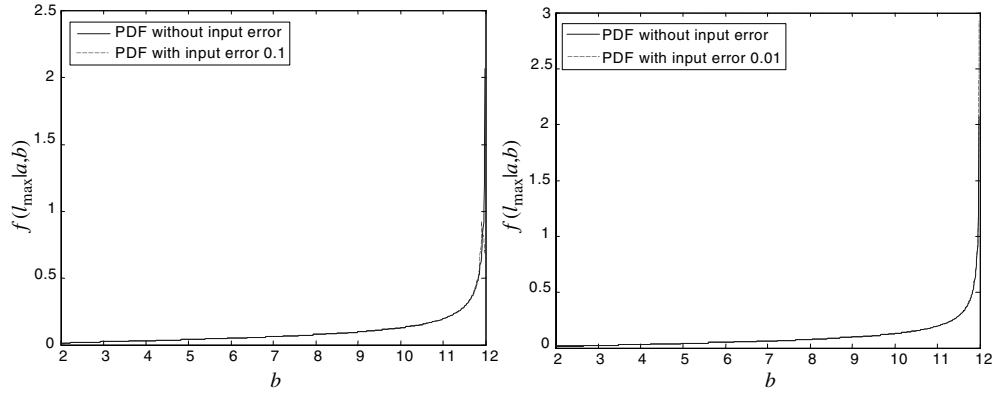

Figure 5. Comparison between $f(l_{\max}|a, b)$ with and without input error e ($a = 2, b = 12$).

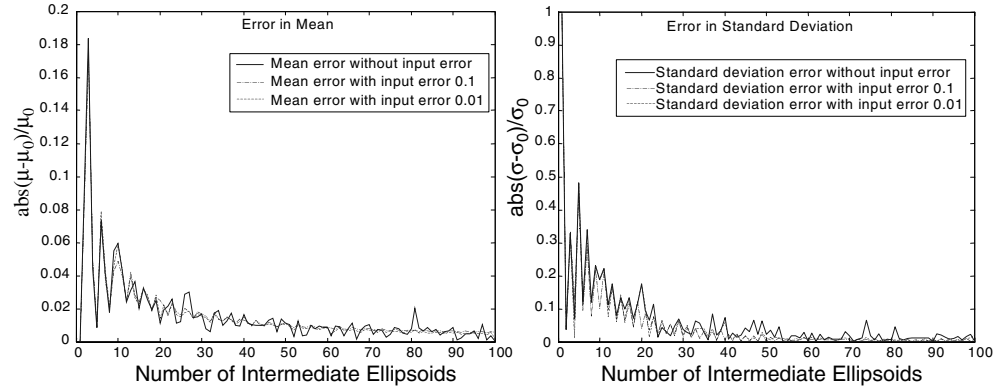

Figure 6. Comparison between convergence of cases with and without input error e.

probability distribution. As a result, the PDF $f(l'_{\max}|a, b)$ is the convolution of $f(l_{\max}|a, b)$ and the PDF of the error $g(x)$, i.e.

$$f(l'_{\max}|a, b) = \int_a^b f(l_{\max}|a, b)g(l'_{\max} - l_{\max})\,\mathrm{d}l_{\max}. \tag{44}$$

A usual type of error is the uniformly distributed error with a fixed window

$$g(x) = \begin{cases} \dfrac{1}{2e}, & -e \leqslant x \leqslant e \\ 0, & \text{otherwise.} \end{cases} \tag{45}$$

From equation (44), one obtains the corresponding error-corrupted PDF (figure 5)

$$\begin{aligned} f(l'_{\max}|a,b) &= \int_a^b f(l_{\max}|a,b)\, g(l'_{\max} - l_{\max})\, \mathrm{d}l_{\max} \\ &= \int_{\max(a,l'_{\max}-e)}^{\min(b,l'_{\max}+e)} \frac{l_{\max}}{\sqrt{(b^2-a^2)(b^2-l_{\max}^2)}} \frac{1}{2e}\, \mathrm{d}l_{\max} \\ &= \frac{1}{2e\sqrt{b^2-a^2}} \Big(-\sqrt{b^2 - \min(b, l_{\max}+e)^2} + \sqrt{b^2 - \max(a, l_{\max}-e)^2} \Big). \end{aligned} \tag{46}$$

With the error-corrupted PDF, following the same testing procedure as stated in section 5.1, the plots of convergence in figure 6 are obtained with 0.1 and 0.01 input errors respectively.

Another type of uniformly distributed error has a changing window with a constant relative error e with respect to its size

$$g(x-y|y) = \begin{cases} \dfrac{1}{2ey}, & -ey \leqslant x-y \leqslant ey \\ 0, & \text{otherwise} \end{cases} = \begin{cases} \dfrac{1}{2ey}, & \frac{x}{1+e} \leqslant y \leqslant \frac{x}{1-e} \\ 0, & \text{otherwise.} \end{cases} \tag{47}$$

From equation (44), one obtains the corresponding error-corrupted PDF (figure 7)

$$\begin{aligned} f(l'_{\max}|a,b) &= \int_a^b f(l_{\max}|a,b)\, g(l'_{\max} - l_{\max}|l_{\max})\, \mathrm{d}l_{\max} \\ &= \int_{\max(a,\frac{l'_{\max}}{1+e})}^{\min(b,\frac{l'_{\max}}{1-e})} \frac{l_{\max}}{\sqrt{(b^2-a^2)(b^2-l_{\max}^2)}} \frac{1}{2el_{\max}}\, \mathrm{d}l_{\max} \\ &= \frac{1}{2e\sqrt{b^2-a^2}} \left(\arcsin \frac{\min\big(b, \frac{l'_{\max}}{1-e}\big)}{b} - \arcsin \frac{\max\big(a, \frac{l'_{\max}}{1+e}\big)}{b} \right). \end{aligned} \tag{48}$$

With the error-corrupted PDF, the plots of convergence in figure 8 are obtained with $0.1 \times l_{\max}$ and $0.01 \times l_{\max}$ input errors respectively.

By comparison, one can see the same trend of quick convergence in mean and standard deviation of both cases with and without input error. The retrieved mean and standard deviation with an input error as high as 10% of the actual size are still very close to those without input error. This means that the proposed algorithm can generate a fairly accurate estimate of the actual probability distribution of the evolving best-fit ellipsoid under a potentially large input error.

In the case of oblate ellipsoids, with the input error (45), one can obtain the corresponding error-corrupted PDF (figure 9)

$$\begin{aligned} f(l'_{\min}|a,b) &= \int_a^b f(l_{\min}|a,b)\, g(l'_{\min} - l_{\min})\, \mathrm{d}l_{\min} \\ &= \int_{\max(a,l'_{\min}-e)}^{\min(b,l'_{\min}+e)} \frac{l_{\min}}{\sqrt{(b^2-a^2)(l_{\min}^2-a^2)}} \frac{1}{2e}\, \mathrm{d}l_{\min} \\ &= \frac{1}{2e\sqrt{b^2-a^2}} \Big(\sqrt{\min(b, l_{\min}+e)^2 - a^2} - \sqrt{\max(a, l_{\min}-e)^2 - a^2} \Big). \end{aligned} \tag{49}$$

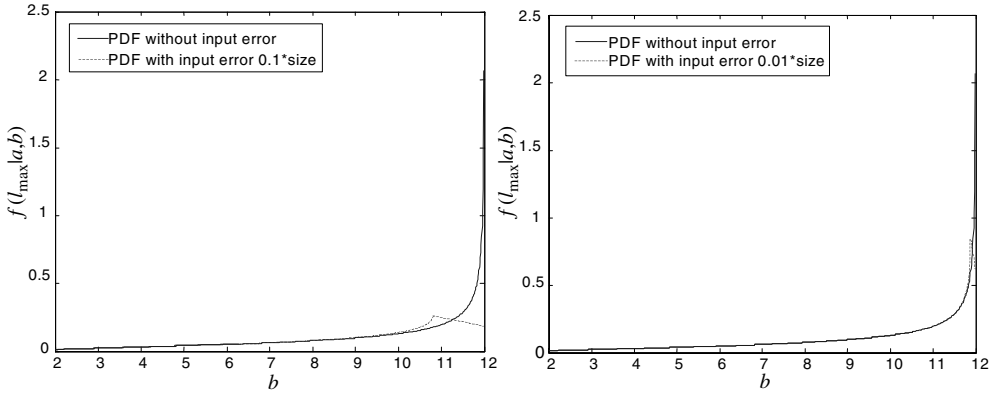

Figure 7. Comparison between $f(l_{\max}|a, b)$ with and without input error $e \times l_{\max}$ ($a = 2, b = 12$).

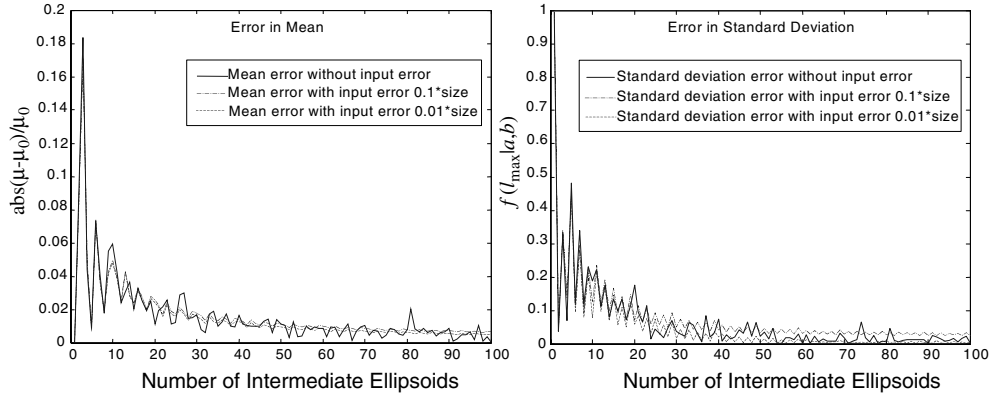

Figure 8. Comparison between convergence of cases with and without input error $e \times l_{\max}$.

With input error (47), one can obtain (figure 10)

$$\begin{aligned} f(l'_{\min}|a,b) &= \int_a^b f(l_{\min}|a,b)\, g(l'_{\min} - l_{\min}|l_{\min})\, \mathrm{d}l_{\min} \\ &= \int_{\max(a,\frac{l'_{\min}}{1+e})}^{\min(b,\frac{l'_{\min}}{1-e})} \frac{l_{\min}}{\sqrt{(b^2-a^2)(l_{\min}^2-a^2)}} \frac{1}{2el_{\min}}\, \mathrm{d}l_{\min} \\ &= \frac{1}{2e\sqrt{b^2-a^2}} \log \frac{\min\left(b, \frac{l'_{\min}}{1-e}\right) + \sqrt{\min\left(b, \frac{l'_{\min}}{1-e}\right)^2 - a^2}}{\max\left(a, \frac{l'_{\min}}{1+e}\right) + \sqrt{\max\left(a, \frac{l'_{\min}}{1+e}\right)^2 - a^2}}. \end{aligned} \tag{50}$$

With the error-corrupted PDFs, similar plots of convergence are obtained to figures 6 and 8.

6. Conclusion

This paper introduces an algorithm to study the probability distribution of the best-fit ellipsoid of an individual random-walk polymer chain from observed planar projections. The algorithm

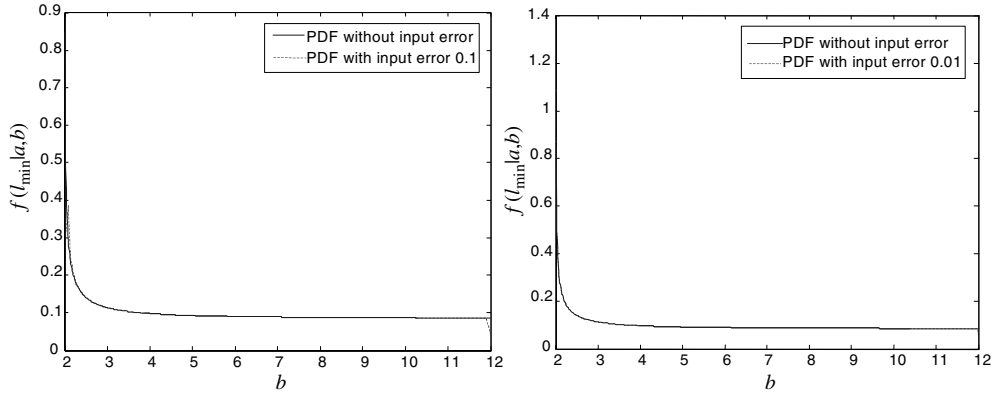

Figure 9. Comparison between $f(l_{\min}|a, b)$ with and without input error e ($a = 2, b = 12$).

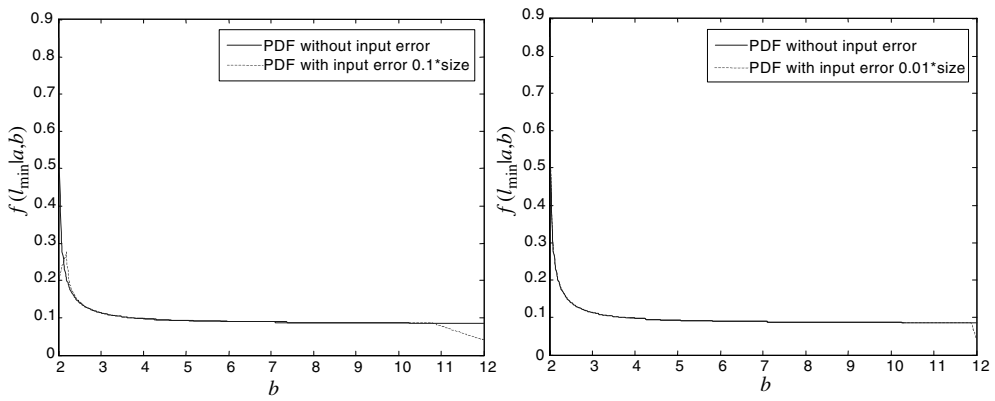

Figure 10. Comparison between $f(l_{\min}|a, b)$ with and without input error $e \times l_{\min}$ ($a = 2, b = 12$).

is effective in retrieving the Gaussian PDFs of evolving best-fit prolate ellipsoids and oblate ellipsoids. From the error analysis, one can conclude that the algorithm has good convergence properties with respect to the number of intermediate ellipsoids employed in the approximation procedure, and good robustness properties with respect to input error. All these mean that the proposed algorithm is applicable to real experimental data.

The algorithm developed in this paper is a promising tool for computing the probability distribution of the shape of a random-walk polymer chain from experimental data. An objective of future work is to retrieve the probability distribution of evolving best-fit ellipsoids with full 3D anisotropy.

References

[1] Kuhn W 1934 Über die gestalt fadenförmiger moleküle in lösungen *Kolloid-Z* **68** 2–11
[2] Solc K and Stockmayer W H 1971 Shape of a random-flight chain *J. Chem. Phys.* **54** 2756–7
[3] Solc K 1971 Shape of a random-flight chain *J. Chem. Phys.* **55** 335–44
[4] Bishop M and Michels J P J 1985 The shape of ring polymers *J. Chem. Phys.* **82** 1059–61
[5] Rundnick J and Gaspari G 1987 The shapes of random walks *Science* **237** 384–9
[6] Rundnick J and Gaspari G 1986 The asphericity of random walks *J. Phys. A: Math. Gen.* **9** L191–3

[7] Aronovitz J A and Nelson D R 1986 Universal features of polymer shapes *J. Physique* **47** 1445–56
[8] Wirtz D 1995 Direct measurement of the transport properties of a single DNA molecule *Phys. Rev. Lett.* **75** 2436–9
[9] Harber C, Ruiz S A and Wirtz D 2000 Shape anisotropy of a single random-walk polymer *Proc. Natl Acad. Sci. USA* **97** 10792–5
[10] Kas J, Strey H and Sackmann E 1994 Direct imaging of reptation for semiflexible actin filaments *Nature* **368** 226–9
[11] Palmer A, Xu J, Kuo S C and Wirtz D 1999 Diffusing wave spectroscopy microrheology of actin filament networks *Biophys. J.* **76** 1063–71
[12] Ma L, Yamada S, Wirtz D and Coulombe P A 2001 A hot-spot mutation alters the mechanical properties of keratin filament networks *Nat. Cell Biol.* **3** 503–6
[13] Leduc P, Harber C, Bao G and Wirtz D 1999 Dynamics of individual flexible polymers in a shear flow *Nature* **399** 564–6
[14] Haber C and Wirtz D 2000 Shear-induced λ-phage DNA assembly *Biophys. J.* **79** 1530–6
[15] Press W H, Flannery B P, Teukolsky S A and Vettering W T 1988 *Numerical Recipes in C: the Art of Scientific Computing* (Cambridge: Cambridge University Press)
[16] Silverman B W 1986 *Density Estimation for Statistics and Data Analysis* (New York: Chapman and Hall)
[17] Boot J C G 1964 *Quadratic Programming* (New York: Elsevier)
[18] Ignizio J P and Cavalier T M 1994 *Linear Programming* (Englewood Cliffs, NJ: Prentice-Hall)
[19] Chirikjian G S and Kyatkin A B 2001 *Engineering Applications of Noncommutative Harmonic Analysis* (Boca Raton, FL: CRC Press)