Generating equilateral random polygons in confinement III

Y. Diao[†], C. Ernst[*], A. Montemayor[*], and U. Ziegler[*]

[*]Department of Mathematics and Computer Science
Western Kentucky University
Bowling Green, KY 42101, USA
[†]Department of Mathematics and Statistics
University of North Carolina Charlotte
Charlotte, NC 28223

Abstract. In this paper we continue our earlier studies [5, 6] on the generation methods of random equilateral polygons confined in a sphere. The first half of the paper is concerned with the generation of confined equilateral random walks. We show that if the selection of a vertex is uniform subject to the position of its previous vertex and the confining condition, then the distributions of the vertices are not uniform, although there exists a distribution such that if the initial vertex is selected following this distribution, then all vertices of the random walk follow this same distribution. Thus in order to generate a confined equilateral random walk, the selection of a vertex cannot be uniform subject to the position of its previous vertex and the confining condition. We provide a simple algorithm capable of generating confined equilateral random walks whose vertex distribution is almost uniform in the confinement sphere. In the second half of the paper we show that any process generating confined equilateral random walks can be turned into a process generating confined equilateral random polygons with the property that the vertex distribution of the polygons approaches the vertex distribution of the walks as the polygons get longer and longer. In our earlier studies, the starting point of the confined polygon is fixed at the center of the sphere. The new approach here allows us to move the the starting point of the confined polygon off the center of the sphere.

AMS classification scheme numbers: 92D20, 92B05, 46N30, 92C40

1. Introduction

There have been numerous studies, both theoretical and numerical, on equilateral random polygons (also known as ideal random polygons), which are often used to model ring polymers under the θ-conditions where polymer segments that are not in a direct contact neither attract nor repel each other. Many theoretical aspects of the equilateral random polygons are now well understood. For example, the mean squared distance between two vertices on an equilateral random polygon of length n that is k vertices apart is $k(n-k)/(n-1)$ and the mean squared radius of gyration of such a random polygon is $(n+1)/12$ [16]. Furthermore, certain measurements with a topological flavor such as the mean ACN (average crossing number) of an equilateral random polygon are also well researched [2, 4]. In the case of numerical studies, several well tested and reliable algorithms have been developed for the purpose of generating equilateral random polygons. These include the polygon folding [9, 10, 11], crankshaft rotation [1, 9], the hedgehog algorithm [9, 12], and the generalized hedgehog algorithm [15]. A new approach using quaternions has been reported recently [3]. In this paper, we continue our earlier research on equilateral random polygons that are confined inside a sphere of fixed radius [5, 6]. The motivation of such an equilateral random polygon model is the well known fact of the highly compact packing of genomic material (long DNA chains) inside living organisms observed in macromolecular self-assembly processes in the complex network of interactions that take place in every organism. Even in the case of a simple organism such as viruses, the DNA packing is of high density. For example, in the prototypic case of the P4 bacteriophage virus, the $3\mu m$-long double-stranded DNA is packed within a viral capsid with a caliper size of about $50nm$, corresponding to a 70-fold linear compaction [8]. Such tight confinement and how the DNA is packed inside the capsid can greatly affect the topological structure of the DNA [14]. Unlike equilateral random polygons without confinement, the confined equilateral random polygons have not been thoroughly studied and there are many unanswered questions. The first issue is how to define the models to reflect the various packing properties the DNA or polymer chains may have. Once a confined random polygon model is defined, the next issue is determining the probability distributions of the random polygons based on the model, and the third issue is the actual generation of the random polygons in accordance with these (theoretical) probability distributions. This line of study, combined with a careful study of the specific knotting structures of the random polygons generated using these probability distributions, can potentially identify the probability distributions and bias preferred by certain topological structures (such as the torus knots), the main problem studied in [14].

In [5] and [6], we have introduced and studied two models of confined equilateral random polygons. In both models, the polygons are "rooted" at the center of the confining sphere. There is no biological or other reason for the polygons to be rooted this way. It is rather a choice for simplicity: as it turns out, equilateral random polygons defined this way are much easier to generate due to the symmetry of the confining sphere (relative to the root) imposed on the equilateral random polygons. The ultimate goal of this paper is to define and study confined equilateral random polygons that are not rooted (at any particular point of the confining sphere). Let us keep in mind that one basic premise on random polygons is that no vertex is more special than the other. Under this premise, all vertices of a confined equilateral random polygon (that is not rooted) should have the same distribution. This condition leads to two new challenging questions: the first is to identify possible distributions of the vertices, and the second is to find an algorithm to generate the polygons following this distribution. These seem to be very hard questions and are the main focus of this paper. Here, we demonstrate that it is possible to generate confined equilateral random walks such that all vertices of the random walk follow the same distribution. This is relevant to our quest and turns out to be very useful. We demonstrate how to use that generation process (of the confined equilateral random walks) to obtain the confined equilateral random polygons with the desired property, namely that their vertices share the same (pre-described) distribution.

The paper is organized as follows: In Section 2, we provide some necessary theoretical background of the probability and conditional probability density distributions needed. We also give an outline of the general idea in using the conditional probability density functions of confined equilateral random walks (conditioned on that its end points are fixed) to generate a confined equilateral random polygon. These results are either well known results or have been established in [5, 6]. But they are essential for this paper to be self-contained and are helpful for the reader to understand our methods and arguments. In Section 3, we present an algorithm for generating confined equilateral random walks such that all vertices of the generated random walk follow the same probability distribution. We also present a case study where we generate the confined random walks in such a way that all vertices of the random walk follow an almost uniform distribution in the confining sphere. In Section 4, we discuss how to adopt the algorithms developed in Section 3 to generate confined equilateral random polygons such that at least most vertices of the generated random polygon follow the desired probability approximately. We also address the problem of generated polygons with starting vertices that are not at the center of the confinement sphere. In Section 5, we discuss certain computational issues in the implementation of our algorithms. Finally, in Section 6, we present some numerical evidence comparing the vertex distribution of random walks with random polygons. We conclude in Section 7 with some open questions and indications of future research.

2. The theoretical background

Let U_1, U_2, ... , U_n be n independent random vectors uniformly distributed on S^2 (so the joint probability density function of the three coordinates of each U_j is simply $\frac{1}{4\pi}$ on the unit sphere). An equilateral random walk of n steps (rooted at the origin), denoted by EW_n, is defined as the polygon whose vertices are given (in a consecutive order along the polygon) by $X_0 = O$, $X_1 = U_1$, $X_2 = U_1+U_2$, ..., $X_n = U_1+U_2+\cdots+U_n$. If the last vertex X_n of EW_n is fixed at X, then we have a conditioned random walk $EW_n|_{X_n=X}$. In particular, EW_n becomes a polygon EP_n if $X_n = O$. In this case, it is called an equilateral random polygon and is denoted by EP_n. It is a well known result that the density function of X_k has the closed form [13]

$$f_k(X_k) = \frac{1}{2\pi^2 r_k}\int_0^\infty x \sin r_k x \left(\frac{\sin x}{x}\right)^k dx, \qquad (1)$$

where $r_k = |X_k|$. Notice that $f_k(X_k)$ is in fact a function of r_k and k.

The following results provide some necessary theoretical background needed in this paper. The proofs can be found in [5].

Lemma 1 *Let Y be uniformly distributed on the unit sphere centered at a fixed point $Y_0 \neq O$ and let $r = |Y|$. Then the probability density function $p(r)$ of r is given by ($r_0 = |Y_0|$):*

$$p(r) = \begin{cases} r/(2r_0), & r \in [|r_0 - 1|, r_0 + 1] \\ 0, & \text{otherwise} \end{cases}$$

Theorem 1 *Let $X_0(= O)$, X_1, ..., X_n be the vertices of an EW_n and $r_j = |X_j|$ for $1 \leq j \leq n$. Let $g_j(r_j)$ be the probability density function of r_j. Then*

(a) $g_j(r_j) = 4\pi r_j^2 f_j(X_j)$ where f_j is given in (1);

(b) In case that r_{k+1} is fixed, then the conditional probability density function of r_k is given by

$$h_k(r_k|r_{k+1}) = \frac{r_{k+1}}{2r_k}\frac{g_k(r_k)}{g_{k+1}(r_{k+1})}, \qquad (2)$$

where $r_k \in [r_{k+1} - 1, r_{k+1} + 1]$ if $r_{k+1} \geq 1$ and $r_k \in [1 - r_{k+1}, 1 + r_{k+1}]$ if $r_{k+1} < 1$.

(c) In case that $r_{k+1} = \tau$ is relatively small compared to k, $h_k(r_k|r_{k+1})$ can be approximated by $r_k/(2\tau)$. In fact, $h_k(r_k|r_{k+1}) = r_k/(2\tau)(1 + O(\tau_0^2/k))$ for $r_k \in [|\tau - 1|, \tau + 1]$ (and it equals 0 otherwise), where $\tau_0 = \max\{\tau, 1\}$. Furthermore, under these conditions

$$f_k(X_k) = \left(\sqrt{\frac{3}{2\pi k}}\right)^3 (1 + O(\frac{\tau_0^2}{k})).$$

In theory, an equilateral random walk with fixed ends can be generated using the conditional probability distribution functions $h_k(r_k|r_{k+1})$. In particular, an equilateral random polygon of given length n (≥ 3) can be generated using such conditional probability distribution functions. This has been described in [5, 6]. For the convenience of our reader, we repeat it here.

Initial step: The starting and ending point of the polygon are set to be the origin by default. X_{n-1} is chosen uniformly on the unit sphere centered at the origin. Once X_{n-1} is chosen, we are left with an equilateral random walk (of $n-1$ edges) with end points fixed at O and X_{n-1}.

Recursive steps: Starting with $j = 2$, choose r_{n-j} according to its distribution (with the condition that r_{n-j+1} has been chosen in the prior step). Specifically, if $H_{n-j}(r_{n-j}|r_{n-j+1})$ is the cumulative probability distribution of r_{n-j} (under the condition that r_{n-j+1} is fixed), then r_{n-j} is chosen to be the solution of the equation

$$H_{n-j}(r_{n-j}|r_{n-j+1}) = u, \tag{3}$$

where u is a random number uniformly chosen from $[0, 1]$. Once $r_{n-j} = |X_{n-j}|$ is chosen, X_{n-j} is chosen uniformly on the intersection circle of the unit sphere centered at X_{n-j+1} and the sphere centered at O with radius r_{n-j}. The last value for j is $n - 2$.

Final step: X_2 is already chosen. At this point X_1 is simply chosen uniformly from the intersection circle of the unit sphere centered at X_2 and the unit sphere centered at O. This allows the random walk to return to the origin.

In [6] the effects of different confinement enforcement strategies on the $H_{n-j}(r_{n-j}|r_{n-j+1})$ functions were discussed. Specifically, whether r_{n-j} is chosen given that $r_{n-k} \leq R$ for $k \leq j$ or for just $k = j$. The former can be seen to treat the confinement sphere as an absorbing boundary, and the latter treats the confinement sphere as a reflective one. In this latter case we follow an accept/reject approach in our algorithm. That is, we generate r_{n-j} without considering confinement, and if $r_{n-j} > R$ we generate a new u and hence a new r_{n-j} until a confined point is found. The advantage to the reflective boundary approach is the great simplicity in its implementation in the algorithm. Furthermore, we can easily incorporate other types of boundaries than a confinement sphere centered at the origin of the polygon. Of note is a confinement sphere centered away from the origin. This simulates the generation of unrooted polygons, see Section 4.2.

3. The case of confined equilateral random walks

As we pointed out in Section 1, our main interest is to develop a method to generate confined equilateral random polygons such that all vertices of the polygon follow the same distribution. Of course, the first question one needs to answer is whether such equilateral random polygons exist. That is, if R is the radius of the confining sphere and $f(r)$ is a probability distribution for $0 \leq r \leq R$, is it possible to define an equilateral random polygon confined in the sphere of radius R such that the probability density function of each vertex X_j of the random polygon is of the form $f(r)/(4\pi r^2)$ (where $r = |X_j|$). And if so, how do we generate such a polygon? Notice also that there is an additional hidden condition here: the generating algorithm must generate the regular equilateral random polygons when the confining condition does not apply. More specifically, suppose that we have chosen the starting point of the polygon (using the pre-described density function) and that the point is close to the center of the confining sphere and the confining sphere has a large radius compared to the length of the polygon. In this case the polygon has no chance of getting out of the confining sphere, hence the confining condition does not apply. Thus the rest of the vertices must follow the distribution of a regular unconfined equilateral random polygon and the generating algorithm has to follow this distribution in this case. As we pointed out in the introduction section, this is a very difficult problem. In Subsection 3.1, we show that at least in the case of equilateral random walks, this is possible for some specially chosen density function $f(r)$. In Subsection 3.2 we demonstrate numerically that this is possible for a special density function $f(r)$ that is very close to the uniform distribution in the confining sphere. In the next section, we use these results to demonstrate that we can at least generate confined equilateral random polygons in which most vertices follow a distribution that is approximately uniform in the confining sphere, which sheds some light on the main question of our concern in this paper.

3.1. A confined equilateral random walk whose vertices have a simple distribution function in the confining sphere

Let S be the confining sphere with radius $R \geq 1$ and consider an equilateral random walk W_k of length k confined in S. Let X_0, X_1, ..., X_k be the (consecutive) vertices of the random walk. Here, the random walk is defined as a Markov chain: each X_{j+1} depends only on X_j in the following way: once X_j is chosen, X_{j+1} is chosen uniformly over the portion of the unit sphere centered at X_j that is contained within S. Notice that this algorithm generates the regular equilateral random walks when the confining condition vanishes. We have the following theorem.

Theorem 2 *Let $f(r)$ be a probability density function defined by*

$$f(r) = \begin{cases} ar^2, & 0 < r \leq R-1; \\ \frac{ar}{4}(R^2 - (r-1)^2), & R-1 < r \leq R; \end{cases}$$

where $a = 48/(16R^3 - 12R^2 + 1) = 48/((2R-1)^2(4R+1))$. If the initial vertex X_0 of W_k is chosen with the distribution $f(|X_0|)/(4\pi|X_0|^2)$, then each vertex X_j of W_k follows the same distribution $f(|X_j|)/(4\pi|X_j|^2)$.

Proof. Given the way W_k is defined, it suffices to prove the following: Let X and Y be two random points in S that are a unit distance apart (think of $X = X_j$ and $Y = X_{j+1}$) such that X follows the probability distribution $f(|X|)/(4\pi|X|^2)$ and that Y is uniformly distributed over the portion of the unit sphere centered at X that is contained in S, then Y follows the probability distribution $f(|Y|)/(4\pi|Y|^2)$.

Let $g(u)$ be the probability density function of $u = |Y|$, $f(r)$ be the probability density function of $r = |X|$ defined in (1) and let $g(u|r)$ be the conditional probability density function of $|Y| = u$ under the condition that $|X| = r$. We have:

$$g(u) = \int_0^R g(u|r)f(r)dr.$$

Let θ be the angle between $\overrightarrow{XO}$ and $\overrightarrow{XY}$, where O is the center of S. It is a well known fact that $\cos\theta \sim U[-1,1]$ if $r \leq R-1$ (hence the confining condition does not apply to Y) and that $\cos\theta \sim U[(1+r^2-R^2)/(2r), 1]$ if $R-1 < r \le R$. Notice that $|Y|^2 = 1 + r^2 - 2r\cos\theta$.

For a fixed r such that $r \le R-1$ it follows that

$$\begin{aligned} P(|Y| \le u|r) &= P(1+r^2-2r\cos\theta \le u^2) \\ &= P(\cos\theta \ge \frac{1+r^2-u^2}{2r}) \\ &= \frac{1}{2}\left(1 - \frac{1+r^2-u^2}{2r}\right). \end{aligned}$$

Differentiating with respect to u yields

$$g(u|r) = u/(2r), \tag{4}$$

where $|r-1| \le u \le r+1$.

On the other hand, for a fixed r such that $R-1 < r \le R$ it follows that

$$\begin{aligned} P(|Y| \le u|r) &= P(1+r^2-2r\cos\theta \le u^2) \\ &= P(\cos\theta \ge \frac{1+r^2-u^2}{2r}) \\ &= \frac{2r}{R^2-(r-1)^2}\left(1 - \frac{1+r^2-u^2}{2r}\right). \end{aligned}$$

It follows that if $R-1 < r \le R$, then

$$g(u|r) = \frac{2u}{R^2-(r-1)^2} \tag{5}$$

where $|r-1| \le u \le R$.

Let us now consider 3 cases: $|u-R| \ge 2$, $2 > |R-u| \ge 1$, and $|R-u| < 1$. The first two cases cover all situations where $0 < u \le R-1$ and we need to show that $g(u) = au^2$ and the third case is for $R-1 < u \le R$ and we need to show that $g(u) = \frac{au}{4}(R^2-(u-1)^2)$.

Case 1. $|u-R| \ge 2$ (of course this can only occur if $R \ge 2$). It follows that $|u-1| \le r \le u+1 \le R-1$ hence $g(u|r) = \frac{u}{2r}$ for $|u-1| \le r \le u+1$ and $g(u|r) = 0$ otherwise. Thus we have

$$g(u) = \int_{|u-1|}^{u+1} \frac{u}{2r} f(r)dr = \int_{|u-1|}^{u+1} \frac{u}{2r} ar^2 dr = au^2.$$

Case 2. $2 > |R-u| \ge 1$. For $|u-1| \le r \le R-1$, we have $g(u|r) = u/2r$, but for $R-1 < r \le u+1$ $(\le R)$ we have $g(u|r) = \frac{2u}{R^2-(r-1)^2}$. It follows that

$$\begin{aligned} g(u) &= \int_{|u-1|}^{R-1} \frac{u}{2r} ar^2 dr + \int_{R-1}^{u+1} \frac{2u}{R^2-(r-1)^2} \frac{ar(R^2-(r-1)^2)}{4} dr \\ &= \int_{|u-1|}^{u+1} \frac{aur}{2} dr = au^2. \end{aligned}$$

For $|u-1|>R-1$, we only obtain one integral and it follows that

$$g(u) = \int_{|u-1|}^{u+1} \frac{2u}{R^2-(r-1)^2} \frac{ar(R^2-(r-1)^2)}{4} dr$$
$$= \int_{|u-1|}^{u+1} \frac{aur}{2} dr = au^2.$$

Case 3. $|R-u|<1$. For $|u-1| \le r \le R-1$, again we have $g(u|r) = u/2r$, but for $R-1 < r \le R$ (since now $u+1 > R$ cannot be the upper bound for r anymore) we have $g(u|r) = \frac{2u}{R^2-(r-1)^2}$. It follows that

$$g(u) = \int_{|u-1|}^{R-1} \frac{u}{2r} ar^2 dr + \int_{R-1}^{R} \frac{2u}{R^2-(r-1)^2} \frac{ar(R^2-(r-1)^2)}{4} dr$$
$$= \int_{|u-1|}^{R} \frac{aur}{2} dr = au(R^2-(u-1)^2)/4.$$

For $|u-1|>R-1$, we only obtain one integral and it follows that

$$g(u) = \int_{|u-1|}^{R} \frac{2u}{R^2-(r-1)^2} \frac{ar(R^2-(r-1)^2)}{4} dr$$
$$= \int_{|u-1|}^{R} \frac{aur}{2} dr = au(R^2-(u-1)^2)/4.$$

This finishes the proof. □

Note that the vertex distribution of the random walk generated by the probability density function of Theorem 2 is non-uniform. The density declines for r-values in the range $R-1<r<R$, see Figure 1.

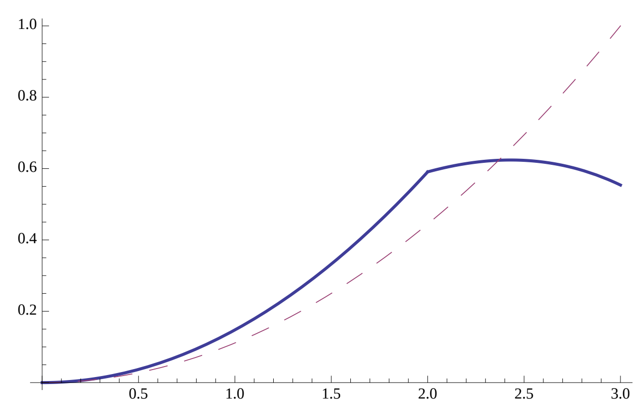

Figure 1. For a confinement radius $R=3$ the probability density function of Theorem 2 is shown together with the uniform distribution function $3r^2/R^3$ (dashed line).

Now that Theorem 2 has confirmed the existence of at least one way to generate confined equilateral random walks whose vertices have the same distribution, we may ask the next question: is it possible for us to select a distribution ahead of time (presumably some distribution with properties that we desire)? For example, is it possible to generate confined equilateral random walks whose vertices all follow the uniform distribution in S? We address this special case in the next subsection.

3.2. A confined equilateral random walk whose vertices are almost uniformly distributed in the confining sphere

Let X be a random point uniformly distributed in S and assume that S is centered at the origin as before. Let $f(r)$ be the corresponding probability density function of $r = |X|$, that is, $f(r) = 3r^2/R^3$ for $0 \le r \le R$ and 0 otherwise. To generate a confined equilateral random walk in S such that all vertices of the polygon follow the uniform distribution in S, the starting point X_0 has to be chosen uniformly in S. Since X_1 is selected conditioned on the choice of X_0, in order for it to have a uniform distribution in S, its conditional probability distribution $h(X_1|X_0)$ has to satisfy certain conditions. Keep in mind that X_1 can be determined in two steps (as we explained earlier towards the end of Section 2), first choosing $u = |X_1|$ (conditioned on the already chosen $r = |X_0|$), then choosing X_1 uniformly from the intersection circle of the spheres centered at O and X_0 with radii r and 1 respectively. Let $g(u|r)$ be the corresponding conditional probability distribution of $|X_1| = u$ under the fixed $r = |X_0|$. Since g vanishes if $u < |r-1|$ or $u > \min\{R, r+1\}$, we must have

$$1 = \int_{|r-1|}^{\min\{R,r+1\}} g(u|r)du.$$

On the other hand, the overall probability distribution of u is given by $\int_0^R g(u|r)f(r)dr$. In order for X_1 to have the same uniform distribution as X_0 in S, we have to have

$$f(u) = \int_0^R g(u|r)f(r)dr.$$

If we can determine the function $g(u|r)$ satisfying the above two conditions, then we are done: the generation of X_2 can be repeated (by repeated use of the function $g(u|r)$) since X_1 follows the uniform distribution, and so on. Unfortunately we have not been able to find an analytic solution to above equations. However we have been able to come up with a function $g(u|r)$ that - at least numerically - comes close. The idea is that we use the same probability density function as in section 3.1 as long as we are at least one unit step away from the boundary of S, but modify the function once we are within one unit of the boundary of S. More specifically, let X be the current position with $|X| = r$ and we take a step to Y with $|Y| = u$. Recall that $u^2 = 1 + r^2 - 2r\cos\theta$ where θ is the angle between $\overrightarrow{XO}$ and $\overrightarrow{XY}$. As in 3.1 we choose $\cos\theta$ with uniform probability density whenever possible. That means if $r \le R-1$ we choose $\cos\theta$ with uniform probability of 1/2 as before. If $r > R-1$ then the unit sphere $S(X,1)$ centered at X with $|X| = r$ intersects not only the confinement sphere $S(O,R)$ but also the sphere $S(O,R-1)$ of radius $R-1$ centered at the origin. On the part of the sphere $S(X,1)$ that is contained within $S(O,R-1)$ we choose any point with uniform probability as before, but for its part between $S(O,R)$ and $S(O,R-1)$ the point is chosen with a different probability density function. So we need to determine which values of $-\cos\theta$ will result in a point on the circular sector between $S(O,R)$ and $S(O,R-1)$ in the next step for $r > R-1$. Solving $u^2 = 1 + r^2 - 2r\cos\theta$ with $u = R-1$ for $-\cos\theta$ shows that for the next point to be outside of $S(O,R-1)$, $-\cos\theta$ needs to be at least $a = \frac{R^2-r^2-2R}{2r}$ if $S(X,1)$ and $S(O,R-1)$ intersect. Otherwise $S(O,R-1)$ is contained in $S(X,1)$ (this may happen for R values less than 1.5) and the next point will be outside $S(O,R-1)$. Similarly, solving the above equation with $u = R$ for $-\cos\theta$ shows that the maximal value of $-\cos\theta$ is $b = \frac{R^2-r^2-1}{2r}$.

A complete specification of a and b is as follows:

$$a(r,R) = \begin{cases} \frac{R^2-r^2-2R}{2r}, & r > R-1 \textit{ and } R-1 > Min(r,|r-1|); \\ -1, & r > R-1 \textit{ and } R-1 \le Min(r,|r-1|); \\ 1, & r \le R-1, \end{cases}$$

$$b(r,R) = \begin{cases} \frac{R^2-r^2-1}{2r}, & r > R-1; \\ 1, & r \le R-1, \end{cases}$$

We now define the pdf for $-\cos\theta$ to have uniform density $1/2$ for values $-1 \le -\cos\theta \le a$ and to be linearly increasing for values of $a < -\cos\theta \le b$. To obtain a pdf function we fix the slope of the linear part as shown below:

$$pdf(-\cos\theta) = \begin{cases} \frac{1}{2}, & -1 \le -\cos\theta \le a; \\ \frac{1}{2}(1 + c(-\cos\theta - a)), & a < -\cos\theta \le b, \end{cases}$$

where

$$c = \frac{4r((r+1)^2 - R^2)}{(2R-1)^2}.$$

Of course, for values $r \le R-1$ our probability density function remains the constant $1/2$. Figure 2 shows the result using this modified conditional probability density function with a confinement radius $R = 2$. We see that the vertex distribution of the vertices in the random walk matches the uniform vertex density function very well. We believe that this is enough evidence that one could indeed generate confined equilateral random walks with uniformly distributed vertices, although the formulation of the required conditional probability distribution function may be convoluted.

4. Confined equilateral random polygons: an asymptotic distribution approximation approach

We apply the results obtained in the last section to try to obtain models of confined equilateral random polygons with a "desired" vertex distribution. A random polygon is a special case of a random walk with fixed end points. In general, the condition that the end points are fixed greatly affects the distributions of the vertices of the random walk. Consequently, the vertex distribution of a random polygon is generally very different from the vertex distribution of a random walk (from the corresponding random walk

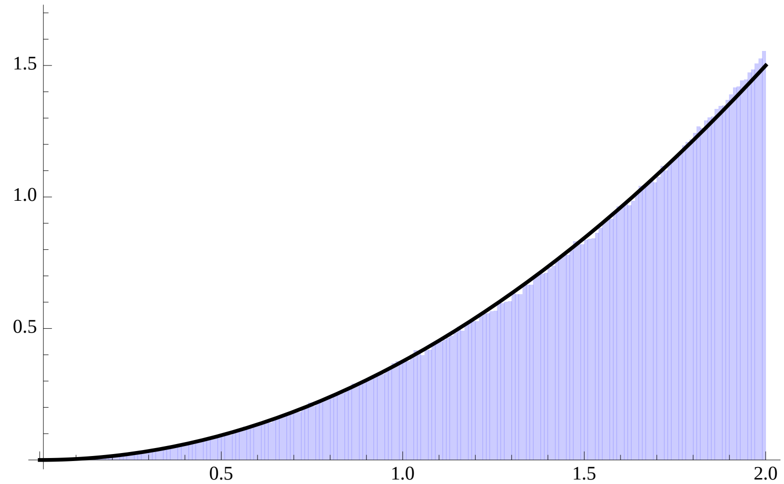

Figure 2. For a confinement radius $R = 2$ the vertex distribution is shown together with the dashed uniform distribution function $3r^2/R^3$. The random walk sample consists of 300 walks of a length of $3,000$ vertices each.

model). However, in the case of confinement, the vertices are close to the origin. In the case that the radius of the confining sphere is much less than the length of the polygon, we say that the polygon is "strongly confined". In the case that the confining sphere behaves like a reflective boundary rather than an absorbing boundary (which is the case in the way we generated the equilateral random walks in the last section), there is strong evidence that the vertices of a "strongly confined" polygon behave much like the vertices of a random walk (from which the polygon model is derived), see Theorem 1(c). Furthermore, even if the starting vertex of the polygon is rooted at a fixed point (not necessarily the center of the confining sphere), the polygon quickly "forgets" where it started, meaning that the distributions of the vertices that are reasonably far away from the starting vertex (in terms of the distance measured along the polygon) behave similarly to the vertices from a random walk of corresponding model. In the following we describe two approaches used to convert a generation process for random walks to a process for random polygons.

4.1. Approach 1: the polygon is rooted at the center of the confining sphere

A model of a confined random walk (like the ones described in the previous section) is given by conditional probability density functions $p(t|s)$ which describe the distance $s = r_{k+1} = |X_{k+1}|$ of the vertex X_{k+1} given that we are currently at distance $t = r_k = |X_k|$ away from the origin. The functions $p(t|s)$ can now be used to construct conditional probability density functions $h_k(t|s)$ that can be used to construct equilateral random polygons as follows:

$$h_k(t|s) = \frac{p(t|s)}{C(k,s)} \frac{s^2}{t^2} \frac{g_k(t)}{g_{k+1}(s)}, \tag{6}$$

where

$$C(k,s)=\int_{|s-1|}^{\min\{R,s+1\}} p(u|s)\frac{s^2}{u^2}\frac{g_k(u)}{g_{k+1}(s)}du$$

and the functions g_k are given by Theorem 1(b). The purpose of the constant C is to normalize $h_k(r|s)$ to ensure it is a probability density function. We have the following Lemma.

Lemma 2 $$\lim_{k\to\infty} h_k(t|s)=p(t|s)$$

Proof. From Theorem 1(a) and (c) we have

$$\begin{aligned}\frac{g_k(t)}{g_{k+1}(s)} &= \frac{4\pi t^2 f_k(t)}{4\pi s^2 f_{k+1}(s)}\\ &= \frac{t^2}{s^2}\frac{\left(\sqrt{\frac{3}{2\pi k}}\right)^3(1+O(\frac{\tau_0^2}{k}))}{\left(\sqrt{\frac{3}{2\pi(k+1)}}\right)^3(1+O(\frac{\tau_0^2}{k}))}\\ &= \frac{t^2}{s^2}\left(1+O(\frac{\tau_0^2}{k})\right).\end{aligned}$$

Thus

$$\lim_{k\to\infty}\frac{s^2}{t^2}\frac{g_k(t)}{g_{k+1}(s)}=1.$$

The result now follows since $\lim_{k\to\infty} C(k,s)=1$ by the fact that

$$\int_{|s-1|}^{\min\{R,s+1\}} p(t|s)dt=1.$$

We have the following theorem:

Theorem 3 *For large n, the joint vertex distribution of the random walks EW_n generated using the conditional probability density functions $p(t|s)$ is the same as the joint vertex distribution of the random polygons EP_n generated using the conditional probability density functions $h(t|s)$.*

This suggests that we can generate random polygons with different joint vertex distributions by experimenting with the much simpler problem of generating randoms walks with different joint vertex distributions. Of course one of the key steps is to find an efficient method of computing the much more complicated conditional probability density functions $h(t|s)$ which is addressed in the next section.

Lemma 2 can be used to explain the connection with the simple model of a random walk in Section 3.1 and the equilateral polygon model used in [5]. Recall that the random walk model in Section 3.1 has the following conditional probability density functions, see equation (4) and (5):

$$p(t|s)=\begin{cases} t/(2s), & s\le R-1;\\ t/(2sc), & R-1<s\le R,\end{cases} \tag{7}$$

where $c = \frac{R^2-(s-1)^2}{4s}$ is a scaling factor to preserve the probability density function property. Putting this conditional probability density function (7) into equation (6) results in the following equations for conditional probability density functions for a random polygon: For $s \le R-1$ we get the equation (2) of Theorem 1

$$h_k(t|s) = \frac{s}{2t}\frac{g_k(t)}{g_{k+1}(s)}, \tag{8}$$

For $R-1 < s \le R$ we get the following

$$h_k(t|s) = \frac{1}{cC}\frac{s}{2t}\frac{g_k(t)}{g_{k+1}(s)},$$

which simplifies slightly to

$$h_k(t|s) = \frac{1}{C'}\frac{s}{2t}\frac{g_k(t)}{g_{k+1}(s)}, \tag{9}$$

where

$$C' = \int_{s-1}^{\min(R,s+1)} \frac{s}{2u}\frac{g_k(u)}{g_{k+1}(s)}du.$$

Equations (8) and (9) are exactly the method that was implemented in [5].

4.2. Approach 2: polygon is obtained through induced closure of a random walk

The last approach used equation (6) to obtain a given polygon distribution. This result only works for a confinement sphere centered at origin or root of the polygon. In this second approach we no longer require this condition, but we use the assumption that the polygon is relatively long. Given any root X_0 in the confinement sphere we generate the polygon as if we are generating a confined random walk as we discussed in Section 3. We do this for all the vertices except, say, the last $k = \lfloor 4R \rfloor$ vertices. At that point, what remains would be a confined equilateral random walk with fixed end points starting at X_{n-k} and ending in $X_0 = X_n$. Since there are relative few vertices left, we can use the accept/rejection method here. That is, we generate (unconfined) random walks with these fixed end points X_{n-k} and X_0 until we find one that is confined in S. More specifically we use the algorithm in Section 2 with the unmodified h_k functions as in equation (2). Here we are using a translated coordinate system such that X_0 is the origin since otherwise we cannot apply the functions given in equation (2). Of course, in doing so, we lose the control of the distribution of the last k vertices. However, since this only happens to a few vertices (even these few are not biased in any obvious way), we may argue that polygons so generated may possess similar geometric and topological characteristics of their counterparts (where all vertices share exactly the same distribution). One big advantage of this approach is the easy implementation of it. This allows fast and easy generation of long, unrooted polygons.

4.3. Approach 3: polygon is obtained through accept/ reject method

There is another rather obvious approach to generated unrooted polygons assuming the the conditional probability density functions given by equation (6) are available and computable. Note that these probability density functions assume that the polygon is rooted at the origin and this cannot be changed since the functions $g_j(r_j)$ of Theorem 1 need this assumption. So instead of shifting the root of the polygon we shift the center of the confinement sphere along the positive z axis using a randomly generated shift that is equivalent to choosing the root according to a desired probability distribution for the vertex density. Now we generate our polygon step-by-step, implementing an accept/reject method on each step. In more detail, assuming that X_{n-j+1} is fixed, we construct X_{n-j} ignoring the confinement sphere. If X_{n-j} is not in confinement then we reject X_{n-j} and start the process over. We repeat this until a point X_{n-j} is constructed that is in confinement. This method also allows the generation of long, unrooted polygons however it is not as fast as the method described in the previous subsection.

5. On the computation of the needed probability density functions

The direct application of the algorithms outlined in the last section, unfortunately, is not so easy. Although the density functions $g_j(r_j) = 4\pi r_j^2 f_j(X_j)$ have an explicit formula as given in (1), the formula involves an improper integral. For any value of n, we can in principle compute the integral exactly using a software package like Mathematica. Thus we have an exact expression for the functions $h_k(r_k|r_{k+1})$ in Theorem 1. However the length of these expressions increases very quickly and their size and the computation time involved makes it difficult to increase the size of k arbitrarily. In this section we replace the improper integral with a summation formula and thus make our algorithm much more suited to a direct computation.

Recall that:

$$h_k(t|s) = \frac{p(t|s)}{C(k,s)} \frac{s^2}{t^2} \frac{g_k(t)}{g_{k+1}(s)}, \tag{10}$$

where

$$C(k,s) = \int_{|s-1|}^{\min\{R,s+1,k\}} p(u|s) \frac{s^2}{u^2} \frac{g_k(u)}{g_{k+1}(s)} du,$$

and

$$g_k(t) = \frac{2t}{\pi} \int_0^\infty x \sin(tx) \left(\frac{\sin x}{x}\right)^k dx.$$

By [7], for $k \geq 2$ there is an explicit expression for $g_k(t)$:

$$g_k(t) = \frac{t}{2^{k-1}(k-2)!} \sum_{j=0}^{k} (-1)^j \binom{k}{j} (t+k-2j)^{k-2} \chi_{(0,\infty)}(2j-k-t), \tag{11}$$

where $\chi_{(0,\infty)}$ is the characteristic function over $(0,\infty)$:

$$\chi_{(0,\infty)}(t) = \begin{cases} 0, & t \leq 0 \\ 1, & 0 < t. \end{cases}$$

Thus if we rewrite $h_k(t|s)$ in the following way

$$h_k(t|s) = \frac{\frac{1}{t^2}p(t|s)g_k(t)}{\int_{|s-1|}^{\min\{R,s+1,k\}} \frac{1}{u^2}p(u|s)g_k(u)du}, \tag{12}$$

we might obtain exact summation formulae for several classes of $p(t|s)$-functions. Recall that the key step of the algorithm is given in equation (3). Thus for a given random value u we need to compute the inverse cumulative probability density function $H_k(t|s)$. The previous equation only give a formula for $h_k(t|s)$ and needs to be integrated with respect to t to obtain a formula for $H_k(t|s)$. If the function $p(t|s)$ is simple enough then this can be done easily. For example, assuming that $s < R-1$ in case of the random walk model in Section 3.1 the equation (12) simplifies to

$$h_k(t|s) = \frac{I_k(t)}{\int_{|s-1|}^{\min\{R,s+1,k\}} I_k(u)du},$$

where

$$I_k(t) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} (t+k-2j)^{k-2} \chi_{(0,\infty)}(2j-k-t).$$

Since $I_k(t)$ is polynomial it is no problems to compute the integral in the denominator and to integrate $h_k(t|s)$ with respect to t. The computation of the inverse $(H_k(t|s))^{-1}(u)$ then needs to be handled numerically.

6. Numerical results

The convergence of vertex distributions given by Theorem 3 is slow due to the fact that even in confinement the root of the random polygon influences the distribution of its vertices. In Figure 3 we illustrate this for the random walk model described in Section 3. The distribution shown is from $10,000$ polygons for each length all of which are rooted at the origin. Clearly we can see that the vertex distribution of the 150-step polygons is closer to the distribution of the random walks than the vertex distribution of the 100-step polygons. However it is also obvious that they are still quite different.

In Figure 4 we illustrate this for the random walk model described in Section 3.2. Again we can see that the vertex distribution of the 150-step polygons is closer to the distribution of the random walks than the vertex distribution of the 100-step polygons. To discount the effect of the root of the polygon we show in Figure 5 the distribution of the middle 20 vertices only. It can be seen that the vertex distribution approximates the uniform distribution much better. However it is also obvious that even for the middle 20 vertices they are still quite different.

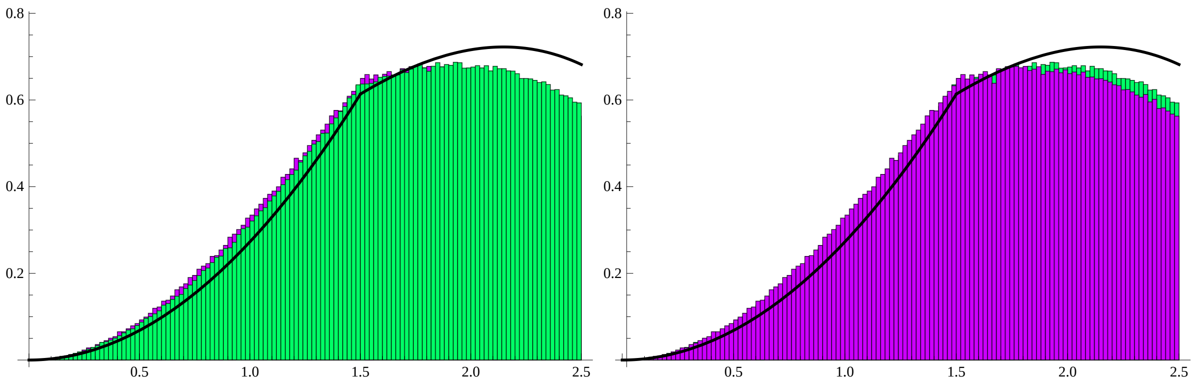

Figure 3. For a confinement radius $R = 2.5$ the vertex distribution function for a random walk is shown together with vertex distributions of $10,000$ polygons of length 100 and length 150. The distribution of the 150-step polygons is in the front on the left while the distribution of the 100-step polygons is in the front on the right. Note the distribution data excludes the first and the last three vertices of the rooted polygons.

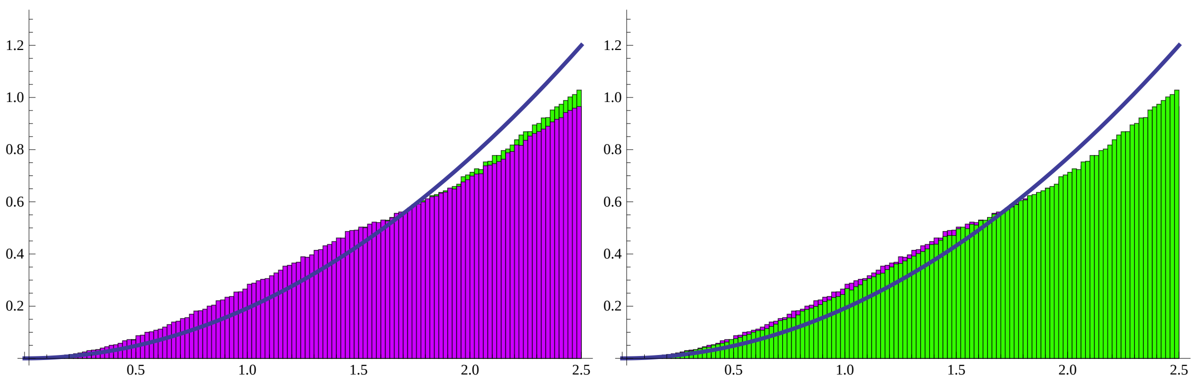

Figure 4. For a confinement radius $R = 2.5$ the uniform vertex distribution function is shown together with vertex distributions of $10,000$ polygons of length 100 and length 150. The distribution of the 150-step polygons is in the front on the left while the distribution of the 100-step polygons is in the front on the right. Note the distribution data excludes the first and the last three vertices of the rooted polygons.

The convergence to the limit vertex distribution can be improved if we vary the starting vertex according to the limit vertex distribution. Figure 6 shows the same size sample of the 100 and 150 steps polygons when the method to generated unrooted polygons via a random walk described in section 4.2 is implemented. The vertex distribution of the polygons is much closer to the target distribution when compared with the rooted polygons.

Finally, Figure 7 shows the three different vertex distributions of the two methods described in Subsections 4.2 and 4.3 together with the polygons of Figure 3. Note we only use the 150 step polygons and that the vertex distribution generated by the method of Subsections 4.3 is between the other two.

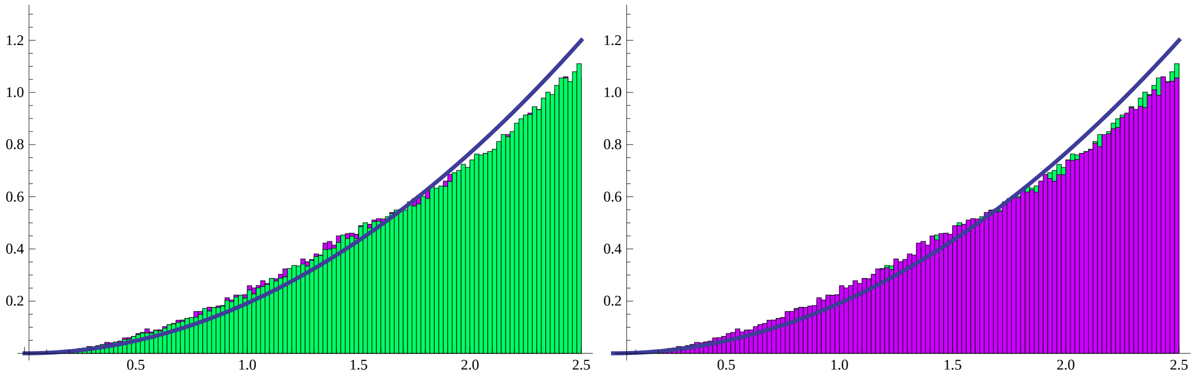

Figure 5. For a confinement radius $R = 2.5$ the uniform vertex distribution function is shown together with vertex distributions of the middle 20 vertices of the polygons used for Figure 4. The distribution of the 150-step polygons is in the front on the left while the distribution of the 100-step polygons is in the front on the right.

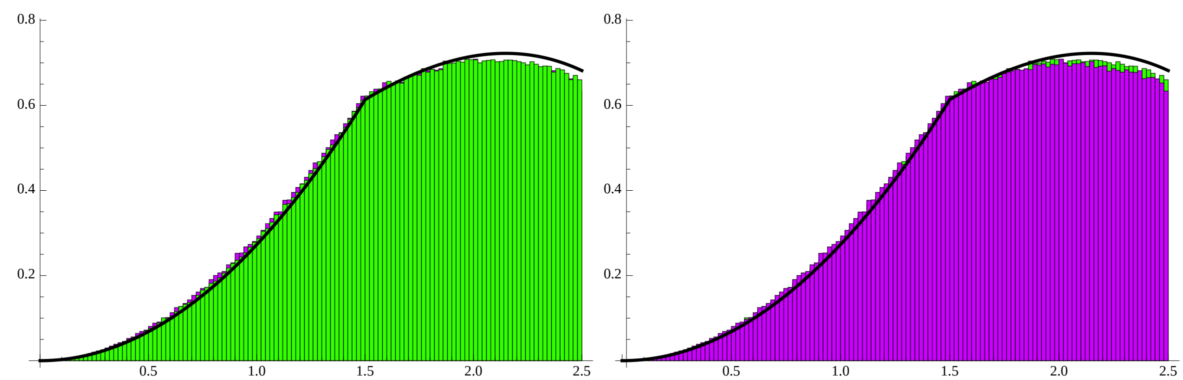

Figure 6. For a confinement radius $R = 2.5$ the vertex distribution function for a random walk is shown together with vertex distributions of $10,000$ polygons of length 100 and length 150 generated using the method described in section 4.2. The distribution of the 150-step polygons is in the front on the left while the distribution of the 100-step polygons is in the front on the right. Note the distribution data still excludes the first and the last three vertices of the rooted polygons so that we have the same sample size for the vertices as in Figure 3.

7. Ending remarks

In this section we raise several questions stemming from our work. Given a vertex distribution function $y = f(r)$ with $\int_0^R f(r)dr = 1$ is it possible to derive a conditional probability density function $p(t|s)$ that describes the next position t given that we are in position s, such that the vertex distribution of a random walk generated by $p(t|s)$ is $f(r)$? What are the effects of these vertex distributions on the geometric/topological properties of the random polygons? For example while it seems obvious that if we generate "stiff" random polygons (i.e. polygons with a relatively small curvature at each vertex) then the vertex distribution of these polygons should have most vertices close to the boundary of the confining sphere. However the converse is not so clear, do polygons

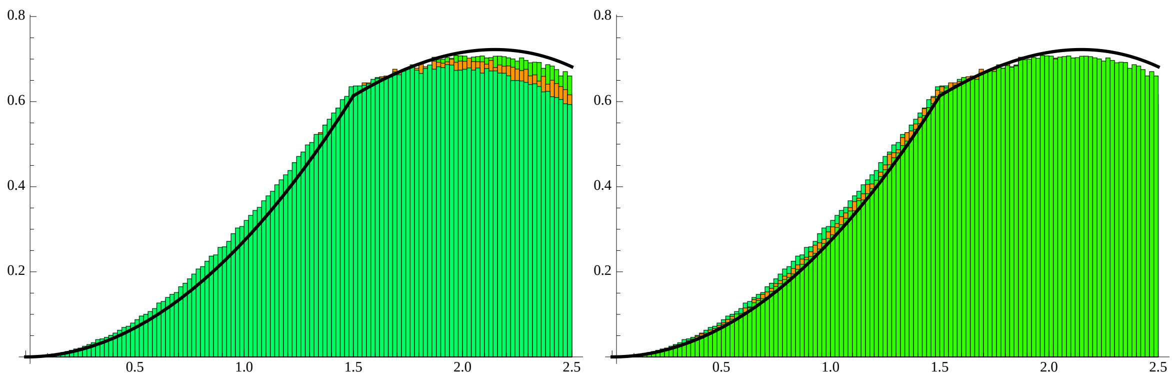

Figure 7. For a confinement radius $R = 2.5$ the vertex distribution function for a random walk is shown together with three vertex distributions of $10,000$ polygons of length 150. Two vertex distributions are generated using the two methods described in Subsections 4.2 and 4.3 (orange). The third vertex distribution represents the rooted polygons of Figure 3.

with a vertex distribution that is denser at the boundary have smaller curvature? One can ask similar question for torsion. Furthermore our approach based on probability density functions might make it possible to write down integral expressions for curvature and torsion.

One can speculate that random polygons with vertex density functions that are denser close to the center of the sphere have a higher complexity of knotting than random polygons with vertex density functions that are denser close to the boundary of the confining sphere. Does the vertex density function have any effect on how many prime knots or composite knots appear? A similar question has already been explored in [14] and the authors plan to investigate some of these questions in their future work.

Acknowledgments

This work is supported in part by NSF Grants #DMS-0920880 and #DMS-1016460 (Y Diao), and by NSF grant #DMS-1016420 (C. Ernst, A. Montemayor, and U. Ziegler).

References

[1] Alvardo S, Calvo J A, Millett K C 2011, *The generation of equilateral polygons*, J. Stat. Phys. **143** 102–38.

[2] Arsuaga J, Borgo B, Diao Y and Scharein R 2009, *The Growth of the Average Crossing Number of Equilateral polygons in Confinement*, J. Phys. A: Math. Theor **42** 465202.

[3] Cantarella J, Deguchi T, Shonkwiler C 2012, *Probability Theory of Random Polygons from the Quaternionic Viewpoint*, arXiv 1206.31612v2, to appear in Com. Pure and Applied Math, 41 pages.

[4] Diao Y, Dobay A, Kusner R, Millett K and Stasiak A 2003, *The Average Crossing Number of Equilateral Random Polygons*, J. Phys. A: Math. Theor **36**(46) 11561–74.

[5] Diao Y, Ernst C, Montemayor A and Ziegler U 2011, *Generating equilateral random polygons in confinement*, *J. Phys. A: Math. Theor*, **44** 405202 doi:10.1088/1751-8113/44/40/405202.

[6] Diao Y, Ernst C, Montemayor A and Ziegler U, *Generating equilateral random polygons in confinement II*, *J. Phys. A: Math. Theor*, **45** 275203 doi:10.1088/1751-8113/45/27/275203.

[7] Hughes B D 1995, *Random walks and random environments. Vol. 1. Random walks.* Oxford Science Publications. The Clarendon Press, Oxford University Press, New York.

[8] Jardine P J and Anderson D L 2006, *DNA packaging in double-stranded DNA phages* The bacteriophages, Ed. Richard Calendar, Oxford University Press, 49–65.

[9] Klenin K V, Vologodskii A V, Anshelevich V V, Dykhne A M and Frank-Kamenetskii M D 1988, *Effect of excluded volume on topological properties of circular DNA, J Biomolec Str and Dyn* **5** 1173–85.

[10] Millett K C 2000, *Knots in Hellas'98 (Delphi), Series on Knots and Everything* **24**, World Scientific, 306–34.

[11] Millett K C 1994, *Knotting of regular polygons in 3-??space*, J. Knot Theory Ram **3** 236–78.

[12] Plunkett P et al 2007, *Total curvature and total torsion of knotted polymers*, *Macromolecules* **40** 3860–7.

[13] Rayleigh L 1919, *Phil. Mag. S. 6.* **37**(220) 321–47.

[14] Reith D, Cifra P, Stasiak A and Virnau P 2012, *Effective stiffening of DNA due to nematic ordering causes DNA molecules packed in phage capsids to preferentially form torus knots*, Nuc. Acids R **40**(11) 5129–37.

[15] Varela R, Hinson K, Arsuaga J and Diao Y 2009, *A Fast Ergodic Algorithm for Generating Ensembles of Equilateral Random Polygons*, J. Phys. A: Math. Theor. **42** 095204.

[16] Zirbel L and Millett K 2012, *Characteristics of shape and knotting in ideal rings*, J. Phys. A: Math. Theor., to appear. arXiv:1204.1366 (April 2012).