Lecture notes for the tangent-tangent correlation and the persistence length

AEG, ASP, JHM

In this document, we discuss analytical results that can be obtained for idealised chain models, called here idealised chains. For a more extensive presentation see for instance *Statistical Mechanics of Chain Molecules* (P. J. Flory, 1969) and *Flexibility of DNA* (J. A. Schellman in Biopolymers,1974).

1 Discussion for idealised chains

Let $\boldsymbol{r}_1, ..., \boldsymbol{r}_N$, with $\boldsymbol{r}_n \in \mathbb{R}^3$, be a set of chain positions and denote by $\boldsymbol{t}_n$ the unit chord between the points $\boldsymbol{r}_{n+1}$ and $\boldsymbol{r}_n$. The chain positions are then assumed to be governed by a recurrence relation of the form

$$\boldsymbol{r}_{n+1} = \boldsymbol{r}_n + l\boldsymbol{t}_n \text{ with } \boldsymbol{r}_0 = 0 \tag{1}$$

with $l > 0$ fixed and $||\boldsymbol{t}_n|| = 1$ for all $n = 0, ..., N-1$. The very specific form of (1) implies that in such idealised models it is equivalent to prescribe a set of positions $\boldsymbol{r}_1, ..., \boldsymbol{r}_N$ or a set of unit chords $\boldsymbol{t}_0, ..., \boldsymbol{t}_{N-1}$ and a length l to define a chain configuration. We will assume moreover that the unit tangents $\boldsymbol{t}_n$ satisfy the Markov property, i.e.

$$\rho(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n, \boldsymbol{t}_{n-1}, ..., \boldsymbol{t}_0) = \rho(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) \tag{2}$$

where $\rho(\boldsymbol{t}_{n+1}|...)$ denotes the conditional probability density on the sphere $\mathbb{S}^2$. The hypothesis (2) implies the following transition identity

$$\rho(\boldsymbol{t}'_{n+1}|\boldsymbol{t}_{n-1}) = \int_{\mathbb{S}^2} \rho(\boldsymbol{t}'_{n+1}|\boldsymbol{t}'_n)\rho(\boldsymbol{t}'_n|\boldsymbol{t}_{n-1})d\boldsymbol{t}'_n, \tag{3}$$

where $d\boldsymbol{t}'_n$ stands for a local sphere element, or more generally that

$$\rho(\boldsymbol{t}'_m|\boldsymbol{t}_n) = \int_{\mathbb{S}^2} ... \int_{\mathbb{S}^2} \rho(\boldsymbol{t}'_m|\boldsymbol{t}'_{m-1})...\rho(\boldsymbol{t}'_{n+1}|\boldsymbol{t}_n)d\boldsymbol{t}'_m...d\boldsymbol{t}'_n \tag{4}$$

for all $m > n$. The conditional expectation of an observable $F(\boldsymbol{t}_m)$ is defined as

$$\langle F(\boldsymbol{t}_m) \,|\, \boldsymbol{t}_n \rangle = \int_{\mathbb{S}^2} F(\boldsymbol{t}'_m)\rho(\boldsymbol{t}'_m|\boldsymbol{t}_n)d\boldsymbol{t}'_m. \tag{5}$$

Using the definition (5) together with the identity (4) we observe the following relations

$$\begin{aligned}\langle F(\boldsymbol{t}_{n+1}) \,|\, \boldsymbol{t}_0\rangle &= \int_{\mathbb{S}^2} \left\langle F(\boldsymbol{t}'_{n+1}) \,|\, \boldsymbol{t}'_n\right\rangle \rho(\boldsymbol{t}'_n|\boldsymbol{t}_0) d\boldsymbol{t}'_n \\ &= \left\langle \langle F(\boldsymbol{t}_{n+1}) \,|\, \boldsymbol{t}_n\rangle \,|\, \boldsymbol{t}_0\right\rangle .\end{aligned} \tag{6}$$

Another set of simplifying assumptions concern the physical constitution of the chain. We will assume that the chain is

i. *transversally isotropic*: i.e.

$$\rho(Q(\phi;\boldsymbol{t}_n)\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) = \rho(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) \text{ for all } n \text{ and for all angle } \phi \tag{7}$$

where $Q(\phi;\boldsymbol{t}_n)$ denotes a rotation of angle ϕ around an axis parallel to $\boldsymbol{t}_n$.

ii. *homogeneous*: i.e.

$$\rho(\boldsymbol{t}_{n+1}|\boldsymbol{t}_n) = \rho(\boldsymbol{t}_n|\boldsymbol{t}_{n-1}) \text{ for all } n \tag{8}$$

1.1 The tangent-tangent correlation

The tangent-tangent correlation is defined as the conditional expectation

$$\langle \boldsymbol{t}_{n+1} \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0\rangle \in \mathbb{R} \tag{9}$$

in the sense of (5). In order to derive an explicit expression for the tangent-tangent correlation, it is convenient to first consider the conditional expectation $\langle \boldsymbol{t}_{n+1} \,|\, \boldsymbol{t}_n\rangle$. Using the definition in (5) together with the assumption of transverse isotropy in (7) we obtain

$$\langle \boldsymbol{t}_{n+1} \,|\, \boldsymbol{t}_n\rangle = \int_{\mathbb{S}^2} \boldsymbol{t}'_{n+1}\rho(\boldsymbol{t}'_{n+1}|\boldsymbol{t}_n) d\boldsymbol{t}'_{n+1} = \alpha_n \boldsymbol{t}_n \tag{10}$$

where

$$\alpha_n = \langle \boldsymbol{t}_{n+1} \cdot \boldsymbol{t}_n \,|\, \boldsymbol{t}_n\rangle \in \mathbb{R}, \tag{11}$$

which implies, using the identity (6), that

$$\langle \boldsymbol{t}_{n+1} \,|\, \boldsymbol{t}_0\rangle = \alpha_n \langle \boldsymbol{t}_n \,|\, \boldsymbol{t}_0\rangle \text{ for all } n. \tag{12}$$

For homogeneous chains, it can be shown using (8) that the value of α_n in (11) is independent of n i.e. $\alpha_n = \alpha$ for all n. The tangent-tangent correlation satisfy then the recurrence relation

$$\langle \boldsymbol{t}_{n+1} \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0\rangle = \alpha \langle \boldsymbol{t}_n \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0\rangle \tag{13}$$

with $\langle \boldsymbol{t}_0 \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0\rangle = 1$ and we consequently have

$$\langle \boldsymbol{t}_n \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0\rangle = \alpha^n. \tag{14}$$

In such idealised models, the value of $\ln \langle \boldsymbol{t}_n \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0\rangle$ is then linearly decaying with a slope given by $\ln \alpha < 0$ (since $|\alpha| < 1$).

1.2 The Flory persistence vector and the persistence length

The Flory persistence vector is defined as

$$\langle \boldsymbol{r}_n - \boldsymbol{r}_0 | \, \boldsymbol{t}_0 \rangle \in \mathbb{R}^3 \tag{15}$$

and for fixed $\boldsymbol{r}_0$, and the persistence length l_p is defined as the limit

$$l_p = \lim_{n\to\infty} \langle (\boldsymbol{r}_n - \boldsymbol{r}_0) \cdot \boldsymbol{t}_0 \, | \, \boldsymbol{t}_0 \rangle \, . \tag{16}$$

Using the recurrence relation (1) we have the expression

$$\langle \boldsymbol{r}_n - \boldsymbol{r}_0 | \, \boldsymbol{t}_0 \rangle = \sum_{k=0}^{n-1} l \, \langle \boldsymbol{t}_k \, | \, \boldsymbol{t}_0 \rangle \, . \tag{17}$$

For transversally isotropic and homogeneous chains it can be shown, using (12), that

$$\langle \boldsymbol{t}_k \, | \, \boldsymbol{t}_0 \rangle = \alpha^k \boldsymbol{t}_0 \tag{18}$$

which implies that

$$\langle \boldsymbol{r}_n - \boldsymbol{r}_0 | \, \boldsymbol{t}_0 \rangle = l \left(\sum_{k=0}^{n-1} \alpha^k \right) \boldsymbol{t}_0 = l \frac{1-\alpha^n}{1-\alpha} \boldsymbol{t}_0 . \tag{19}$$

The persistence length in (16) is then explicitly given by

$$l_p = \lim_{n\to\infty} l \sum_{k=1}^{n-1} \alpha^k = \frac{l}{1-\alpha} \tag{20}$$

since $|\alpha| < 1$ and $\boldsymbol{t}_0 \cdot \boldsymbol{t}_0 = 1$.

2 Application to the random-ϕ model

The *random-ϕ model* was proposed by J. A. Schellman (1974) and assumes that the conditional distribution of two consecutive unitary chord $\boldsymbol{t}_n$ and $\boldsymbol{t}_{n+1}$ is given by

$$\rho(\boldsymbol{t}_{n+1} | \boldsymbol{t}_n) = \frac{1}{Z} \exp\{B \cos\theta\} \tag{21a}$$

where

$$Z = \int_0^\pi \exp\{B\cos\theta\} \sin\theta d\theta = \frac{1}{B} \left(\exp\{B\} - \exp\{-B\} \right) . \tag{21b}$$

The functional form of the distribution in (21a) can be motivated by the identity $\frac{1}{2} (\boldsymbol{t}_{n+1} - \boldsymbol{t}_n) \cdot (\boldsymbol{t}_{n+1} - \boldsymbol{t}_n) = \frac{1}{2} - (\boldsymbol{t}_{n+1} \cdot \boldsymbol{t}_n)$ which expresses that the value of $\cos\theta =$

$\boldsymbol{t}_{n+1} \cdot \boldsymbol{t}_n$ is related to an approximation of the square of the local chain curvature. The angle θ is then the bending angle between $\boldsymbol{t}_n$ and $\boldsymbol{t}_{n+1}$ and the constant B is referred as the bending stiffness parameter, since it encodes how difficult it is to have the angle θ deviating from the value 0. This model is called random-ϕ since the probability in (21a) is independent of the longitudinal angle ϕ between $\boldsymbol{t}_n$ and $\boldsymbol{t}_{n+1}$, when $\boldsymbol{t}_{n+1}$ is expressed in spherical coordinates with respect to $\boldsymbol{t}_n$. The constant Z is the normalization constant, and it has to be stressed that the appropriate Jacobian factor $\sin\theta$, related to the spherical coordinates, appears in this definition and that it makes explicit computations *easier* thanks to the use of the change of variables $u = \cos\theta$.

The value of the parameter $\alpha = \alpha_n$ for all n, defined in (11), can then explicitly be computed to be

$$\alpha = \frac{1}{Z}\int_0^\pi \cos\theta \exp\{B\cos\theta\}\sin\theta d\theta = \frac{\cosh B}{\sinh B} - \frac{1}{B}. \tag{22}$$

The function $L(B) = \frac{\cosh B}{\sinh B} - \frac{1}{B}$ is called the Langevin function.

2.1 The random flight limit

The random flight limit is characterised by a very *small* value of bending stiffness parameter B, i.e. $B << 1$. In this case, the random-ϕ model (21) is governed approximately by a probability which is uniform in the angle θ. We have then

$$\alpha = 0 + O(B), \tag{23}$$

which implies that the tangent-tangent correlation

$$\langle \boldsymbol{t}_n \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0 \rangle = 0 + O(nB) \tag{24}$$

for all n, and that the persistence length

$$l_p = l + O(lB). \tag{25}$$

These results are interpreted as a *fast* complete randomisation of the unitary chords.

2.2 The semi-flexible limit

The semi-flexible limit is characterised by a very *large* value of bending stiffness parameter B, i.e. $B >> 1$. In this case, the random-ϕ model (21) is governed by a probability which is very peaked in $\theta = 0$. We have then

$$\alpha = 1 - \frac{1}{B} + O(\exp\{-2B\}), \tag{26}$$

which implies that the tangent-tangent correlation

$$\langle \boldsymbol{t}_n \cdot \boldsymbol{t}_0 \,|\, \boldsymbol{t}_0 \rangle = \exp\left\{-\frac{nl}{B}\right\} + O\left(n\exp\{-2B\}\right) \tag{27}$$

for all n, and that the persistence length

$$l_p = Bl + O(l \exp\{-2B\}). \tag{28}$$

These results are interpreted as a *slow* complete randomisation of the unitary chords. It is interesting to note that expression (27) shows that the larger the bending stiffness parameter B is, the slower is the randomisation.

3 Remark for the cgDNA model

More realistic descriptions of chains involve more complicated relations than (1) to describe the position $\boldsymbol{r}_n$ of each element in the chain, and more complicated functional forms than (21) to model the chain probability density distribution. For instance, in the cgDNA model, six degrees of freedom are necessary to describe the relative base pair positions and $12(N-6)$ to write the probability density distribution. Moreover the local intrinsic twist and curvature (encoded by the shift vector of the quadratic form) as well as the sequence dependence of the coefficients (encoded in the sequence dependent parameter set) implies the failure of transverse isotropy (7) and homogeneity (8) assumptions. Finally, we also emphasize that the 18×18 overlapping block structure of the stiffness matrices predicted by the cgDNA model implies that the Markovian assumption (2) is no longer valid. The cgDNA setting is consequently much more complicated than the one in (21) and makes analytical computations challenging.

It should be stressed however that, even if none of the basic hypotheses of the ideal chain model are consistent with the cgDNA model, we still observe using numerical Monte Carlo computation that the limiting value of the Flory persistence vector is finite, as in (19), and that the tangent-tangent correlation is decaying approximately like (14) but with corrections due to the intrinsic shape. A large discussion of this phenomena is presented for the rigid base pair model in section 2.4 & 7.2 in the Ph. D thesis [1] «On the statistical physics of chains and rods, with application to multi-scale sequence dependent DNA modelling» (A. E. Grandchamp, EPFL, 2016).

[1] The full document is available on http://lcvmwww.epfl.ch/publications/#1-13