Mathematical Modelling of DNA
Prof. John Maddocks
Session 8: Solutions — A. Patelli

1 Explicit computation of apparent persistence length for a tractable probability density function (the HWLC)

1. Here you have an example of implementation of the Euler–Rodrigues formula.

```
function Q = Euler_Rodrigues( u )

% This Ensure that the vector u is a column vector
u = reshape(u , [ 3 1]) ;

% Compute norm of u
uu = u'*u ;

% Get the matrix [ u x ]
u_cross = [  0     -u(3)    u(2)    ; ...
             u(3)   0      -u(1)    ; ...
            -u(2)   u(1)    0     ] ;

Q = ( 1 - uu ) / ( 1 + uu ) * eye(3)    + ...
             2 / ( 1 + uu ) * u_cross   + ...
             2 / ( 1 + uu ) * (u*u')    ;


end
```

2. You will obtain that the entries $(2,3)$ and $(1,3)$ tend to zero for a enough large N.

3. i) $\langle Q(u)\rangle_{(3,3)} = 0.993337$, $\ell_p = 150.08$

N	$\langle Q(u)\rangle_{(3,3)}$ by MC	Error
100	0.993966	0.0006287
1000	0.993059	0.0002779
10000	0.993446	0.0001086
100000	0.993422	0.0000851

ii) $\langle Q(u)\rangle_{(3,3)} = 0.993333$, $\ell_p = 150.0$

N	$\langle Q(u)\rangle_{(3,3)}$ by MC	Error
100	0.993826	0.0004931
1000	0.993344	0.0000107
10000	0.993320	0.0000141
100000	0.993359	0.0000254

iii) $\langle Q(u)\rangle_{(3,3)} = 0.993338$, $\ell_p = 150.11$

N	$\langle Q(u)\rangle_{(3,3)}$ by MC	Error
100	0.994102	0.000763
1000	0.993473	0.000134
10000	0.993361	0.000022
100000	0.993363	0.000024

4 -7 The value of K_1 and K_2 play the more important role in the approximation of the analytical formula then the value of K_3.

Remarks : The analytic expression (1) in the statement of this exercise has been obtained by making the assumption of independent distribution of the the junction displacements. Moreover by assuming that each junction is identically distributed (uniform chain assumption) we used the analytic expression to derive that $\ln\langle \mathbf{t}^{[1]} \cdot \mathbf{t}^{[n]} \rangle \approx -n\alpha + \mathcal{O}(\frac{n}{K_{min}^2})$ (see pdf file for notations). Hence the plot of the natural logarithm of the tangent-tangent correlation as function of the base-pair n, is linear with a negative slope. We stress here that the latter assumptions (i.i.d. of the junction displacement) are not satisfied by the cgDNA model. In session 6 we computed the tangent-tangent correlation in the context of the cgDNA model by means of the direct Monte Carlo sampling of the cgDNA Gaussian for the 6 distinct poly dimers. What we observed is that the obtained plots in Figure 9 of the $\ln\langle \mathbf{t}^{[1]} \cdot \mathbf{t}^{[n]} \rangle$ vs number of base-pairs has a negative slope and is very close to be straight for all the 6 poly dimers. This is due to the fact that the DNA, in general, is highly twisted and, more important, to the fact that the poly dimers have a very straight ground-states. These two features lead to an almost uniform behaviour of the poly dimers as one can see in the Figures (6-7) of the solution sheet of session 6. In session 7 we consider other sequences, fragments of the lambda phage genome, and we observed that the plots $\ln\langle \mathbf{t}^{[1]} \cdot \mathbf{t}^{[n]} \rangle$ vs number of base-pairs (Figure 6 solution of session 7) behaves dramatically different compared to the poly dimers one and that are far from being straight. The reason of this behaviour is related to the ground-states of the lambda phage fragments that are more bend compared to the poly dimers (they do not show an uniform behaviours, see for instance Figures 4 and 5 in solution session 7). We understood here that the ground-state of a molecule of DNA has an impact on the shape of the decaying of the logarithm of the tangent-tangent correlation. Thus, we introduced the shape factorization (see equation (1) exercise 2 session 7) that reduce the impact of the ground-state on the tangent-tangent correlation and leads to more straight plots even for bended ground-states (see for instance Figure 7 solution session 7). The shape factorization has been introduced in the article "Sequence-Dependent Persistence Lengths of DNA" , J.S. Mitchell et al. To understand why the shape factorization works is beyond the purpose of this lecture, you just need to know how the factorization works.

2 On the parametrization of junction displacement using quaternions

1. Let define $u = Cay(Q)$ the Cayley vector of $Q \in$SO(3) with $\|u\| = \tan\frac{\theta}{2}$, $0 \le \theta < \pi$. We need to find $q \equiv q(u) \in \mathbb{R}^4$ such that $\|q\| = 1$ and q satisfy the relations (4) in the statement of this exercise. Thus, we have that

$$\begin{aligned}
\|(q_0, \mathbf{q})\| &= \sqrt{q_0^2 + \|\mathbf{q}\|^2} = \sqrt{\cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}} = \cos\frac{\theta}{2}\sqrt{1 + \tan^2\frac{\theta}{2}} \\
&= \cos\frac{\theta}{2}\sqrt{1 + \|u\|^2} \\
&= \cos\frac{\theta}{2}\|(1, u)\|.
\end{aligned}$$

Finally we found that $q = \cos\frac{\theta}{2}(1, u) \in \mathbb{R}^4$ is a quaternion that satisfy $\|q\| = 1$ and the relations (4).

2. We recall that the Euler-Rodrigues formula reads :

$$Q(u) = \frac{1 - \|u\|^2}{1 + \|u\|^2}\mathrm{I} + \frac{2}{1 + \|u\|^2}[u\times] + \frac{2}{1 + \|u\|^2}u \otimes u, \tag{1}$$

see exercise 1.2 session 3 for details. We will use now the previous part and we will define $q = (q_0, \mathbf{q}) = (q_0, q_1, q_2, q_3) = (\cos\frac{\theta}{2}, \cos\frac{\theta}{2}u)$ where $\dfrac{\mathbf{q}}{\cos\frac{\theta}{2}} = u$. Then the following identities holds

$$\begin{aligned}
\frac{1-\|u\|^2}{1+\|u\|^2} &= \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2} = q_0^2 - q_1^2 - q_2^2 - q_3^2,\\
\frac{2}{1+\|u\|^2}[u\times] &= 2q_0[\mathbf{q}\times] = \begin{bmatrix} 0 & -2q_0q_3 & 2q_0q_2 \\ 2q_0q_3 & 0 & -2q_0q_1 \\ -2q_0q_2 & 2q_0q_1 & 0 \end{bmatrix},\\
\frac{2}{1+\|u\|^2}u\otimes u &= 2\mathbf{q}\otimes\mathbf{q} = \begin{bmatrix} q_1^2 & q_1q_2 & q_1q_3 \\ q_1q_2 & q_2^2 & q_2q_3 \\ q_1q_3 & q_2q_3 & q_3^2 \end{bmatrix}.
\end{aligned}$$

The result is obtained just by using the previous identities.

3. We recall that in the cgDNA model we have a specific scaling for the rotations, thus one has to be careful when using, for example, the Euler-Rodrigues formula or the Cayley transform. In fact in the cgDNA model we use a scaling in such a way that if u is a Cayley vector, $\|u\| = 10\tan\frac{\theta}{2}$. It will be explained in class later on why we are using such a scaling. For the purpose of this exercise, you have to set $U = u_2/10$ where u_2 is the Cayley vector related to the second inter which is related to the rotation Q_2. In order to get the quaternions for the matrices $R_{2,3}$ you have to use the inverse Cayley transform of part 3), exercise 1 session 3 with $M = R_{2,3}$ to get the Cayley vector $\rho_{2,3}$. Once you have the Cayley vectors $\rho_{2,3}$ and U you can use the part 1) of this exercise to compute the quaternions $q^{R_{2,3}}$, and q^{Q_2}. Compute the quaternion multiplication to verify that $q^{R_2}\circ q^{Q_2} = q^{R_3}$ and then use the part 2) to verify that $R_3 = R(q^{R_3})$.