# 1 Positive Definiteness of the cgDNA stiffness matrix and Invariance of the parameter set

Let $\mathcal{P} = \left\{K^{\alpha\beta}, \sigma^{\alpha\beta}, K^{\alpha}, \sigma^{\alpha} | \alpha\beta \in D \text{ and } \alpha \in M\right\}$ be a cgDNA parameter set where $D$ is the set of all dimers (two base pairs) formed using $M = \{A, T, C, G\}$, and let $S = X_1, X_2, \ldots, X_n$ be a sequence with $X_i \in \{A, T, C, G\}$

**1.** Assume that

$$K^{\alpha\beta} + \frac{1}{2}\begin{bmatrix} K^{\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & K^{\beta} \end{bmatrix} > 0, \tag{1}$$

$$K^{\alpha\beta} + \begin{bmatrix} K^{\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2}K^{\beta} \end{bmatrix} > 0, \tag{2}$$

$$K^{\alpha\beta} + \begin{bmatrix} \frac{1}{2}K^{\alpha} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & K^{\beta} \end{bmatrix} > 0, \tag{3}$$

for all $K^{\alpha\beta}$, $K^{\alpha}$, and $K^{\beta}$ in $\mathcal{P}$. Show that the cgDNA stiffness matrix $K(S, \mathcal{P})$ is positive definite for any sequence $S$.

**2.** Define by $E_{2m-1}$ the trailing-diagonal matrix with $2m-1$ repetitions of $E = diag(-1, 1, 1, -1, 1, 1)$, and define by $\overline{S}$ the complementary sequence of $S$. By assuming

$$\begin{aligned} K^{\overline{\alpha\beta}} &= E_3 K^{\alpha\beta} E_3, \\ \sigma^{\overline{\alpha\beta}} &= E_3 \sigma^{\alpha\beta}, \\ K^{\overline{\alpha}} &= E K^{\alpha} E, \\ \sigma^{\overline{\alpha}} &= E \sigma^{\alpha}, \end{aligned}$$

where $\overline{\beta}\overline{\alpha}$ and $\overline{\alpha}$ are the complementaries sequence to, respectively, $\alpha\beta$ and $\alpha$, show that for any sequence $S$:

$$\begin{aligned} K(\overline{S}, \mathcal{P}) &= E_{2n-1} K(S, \mathcal{P}) E_{2n-1}, \\ \mu(\overline{S}, \mathcal{P}) &= E_{2n-1} \mu(S, \mathcal{P}), \end{aligned}$$

so that $\rho(\overline{\mathbf{w}}; \overline{S}, \mathcal{P}) = \rho(\mathbf{w}; S, \mathcal{P})$, where $\overline{\mathbf{w}} = E_{2n-1}\mathbf{w}$.

**3.** Using part 2), how many of the conditions in part 1) are independent?

## 2 Zero entries in the cgDNA parameter set

i) Describe all of the restrictions on the cgDNA parameters $K^{\alpha\overline{\alpha}}, \sigma^{\alpha\overline{\alpha}}$, and $\hat{x}^{\alpha\overline{\alpha}}$ for each of the palindromic dimer steps $\alpha\overline{\alpha}$, $\alpha \in \{A, T, C, G\}$.

ii) If $S$ is a palindromic sequence, i.e., $\overline{S} = S$, what are the corresponding restrictions on the reconstructions, $K(S), \sigma(S)$, $\mu(S)$.

## 3 Total number of unknowns in a cgDNA parameter set

Let $\mathcal{P} = \left\{K^{\alpha\beta}, \sigma^{\alpha\beta}, K^{\alpha}, \sigma^{\alpha} | \alpha\beta \in D \text{ and } \alpha \in M\right\}$, where $D$ is the set of all dimers formed using the alphabet $M = \{A, T, G, C\}$, be a cgDNA parameter set. Using all the symmetry conditions ( i.e. Crick-Watson symmetries (Ex. 1) and palindrome restrictions (Ex. 2) ) find the total number of independent scalars in the parameter set $\mathcal{P}$.

## 4 On the parametrization of junction displacement using quaternions (a final exercise pertaining to cgDNAmc)

Nowadays it is rather fast to samples from multivariate distribution, thus the main part on a Monte Carlo code is in the evaluation of the chose deterministic function. For example in the cgDNAmc we have made two different choice of functions of the cgDNA coordinates:

**1.** $\phi(\mathbf{x}) = (R_1^T R_n)_{(3,3)}$,

**2.** $\phi(\mathbf{x}) = R_1^T (r_n - r_1)$,

where $(A)_{(3,3)}$ means the (3,3) entry of a matrix $A \in \mathbb{R}^{3\times 3}$, $(R_1, r_1)$ is a fixed base-pair frame chosen to be the first (but not necessarily the first one of the DNA fragment), and $(R_n, r_n)$ is the $n$th base-pair frames after the fixed one. By choosing the above function, in cgDNAmc we have to perform many matrix multiplications in $SO(3)$ in order to be able to evaluate the functions for each sampled configuration. In the cgDNAmc code for efficiency there are implemented using the quaternions multiplication. We have already seen how to parametrize a rotation matrix using three numbers or the Cayley vectors. In this exercise we will study the parametrization of a rotation matrix by four numbers called Euler-Rodrigues parameters or quaternions.

Any vector $q = (q_0, q_1, q_2, q_3) \in \mathbb{S}^3 = \{x \in \mathbb{R}^4 | x \cdot x = 1\}$ can be interpreted as a right-handed rotation in $\mathbb{R}^3$ through an angle $\theta$ and around a unit axis $\mathbf{n} \in \mathbb{R}^3$, where $\theta$ and $\mathbf{n}$ solve :

$$\cos\frac{\theta}{2} = q_0, \quad \text{and} \quad \mathbf{n}\sin\frac{\theta}{2} = \begin{bmatrix} q_1 \\ q_2 \\ q_3 \end{bmatrix} = \mathbf{q}. \tag{4}$$

**1.** Let $Q \in$ SO(3) a rotation matrix about a unit axis $\mathbf{n}$ through an angle $0 \leq \theta < \pi$. Let $u = Cay(Q) \in \mathbb{R}^3$ be the Cayley parametrisation of $Q$. Find the quaternion parametrisation of $Q$ in term of the Cayley vector $u$. [Hint: We recall that $\|u\| = \tan\frac{\theta}{2}$].

**2.** Using the previous part, show that the Euler-Rodrigues formula (2) of exercise 1.2 session 3 implies the following quaternion parametrisation:

$$Q(q) = \begin{bmatrix} q_1^2 - q_2^2 - q_3^2 + q_0^2 & 2(q_1q_2 - q_3q_0) & 2(q_1q_3 + q_2q_0) \\ 2(q_1q_2 + q_3q_0) & -q_1^2 + q_2^2 - q_3^2 + q_0^2 & 2(-q_1q_0 + q_2q_3) \\ 2(q_1q_3 - q_2q_0) & 2(q_1q_0 + q_2q_3) & -q_1^2 - q_2^2 + q_3^2 + q_0^2 \end{bmatrix} \tag{5}$$

**3.** From a computational point of view, the efficiency of quaternion, instead of rotation matrix, is that if $Q_i = Q(q_i) \in \mathrm{SO}(3)$ , for $i = 1, 2, 3$, we have

$$Q_3 = Q_1 Q_2 \Longleftrightarrow q_3 = q_1 \circ q_2, \tag{6}$$

where the symbol $\circ$ mean the multiplication operator for the quaternion that for two quaternions $q = (q_0, \mathbf{q})$ and $p = (p_0, \mathbf{p})$ reads

$$q \circ p = \left(q_0 p_0 - \mathbf{q} \cdot \mathbf{p}, q_0 \mathbf{p} + p_0 \mathbf{q} + \mathbf{q} \times \mathbf{p}\right). \tag{7}$$

We could derive (7) and prove the equivalence (6) but here we will just check them numerically. For that purpose use the cgDNA package to reconstruct the ground-state of a short fragment of DNA (10-12 base-pair). Then, check the following

$$R_3 = R_2 Q_2 \Longleftrightarrow q^{R_3} = q^{R_2} \circ q^{Q_2}, \tag{8}$$

where $R_i$ is the orientation of the $i$th base-pair, and $Q_2$ is the rotational part of the second junction displacement, and $q^{M_i}$ is the quaternion related to the rotation matrix $M_i$.
**Remark:** In the cgDNA model the Cayley vectors are scaled in a way that their norm equal $10 \tan \frac{\theta}{2}$, where $\theta$ is the angle of the rotation. Thus, you have to rescale the cgDNA Cayley vectors such that their norm are $\tan \frac{\theta}{2}$ if you want to use the relation between Cayley vectors and quaternions.