1 On the main outputs of the cgDNA+ package

1.1 Stiffness matrices

1. In Figure 1 we show the sparsity pattern of the stiffness matrix K for the sequence $S = CGCGAATTCGCG$. The sparsity pattern of the cgDNA+ stiffness matrix is 42×42 blocks with 18×18 overlaps in the interior of the sequence (ie for interior dimer steps), and 36×36 blocks with 18×18 overlaps for first and last dimer steps (ie for end dimer steps). Note that the first and last blocks are smaller, only 36×36, but still with 18×18 overlaps. Due to the overlapping blocks the inverse of K is dense.

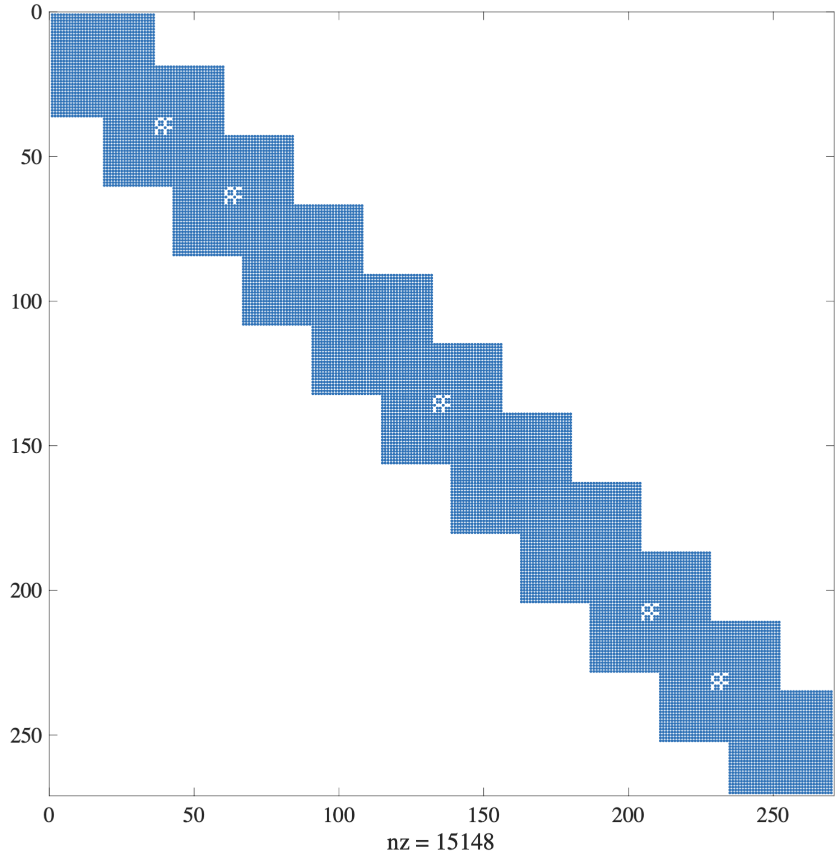

Figure 1: Use `spy(abs(K)>tolerance)` with tolerance $= 10^{-10}$.

2. We performed a single point mutation on the sequence S which leads to the following sequence $\tilde{S} = CGCGCATTCGCG$. We basically changed the fifth base A with a C. In Figure 2 we can see that the latter modification leads to the modification of two 42×42 blocks. In particular the single point mutation we perform changed the middle trimer GAA of sequence S into GCA and left the rest unchanged. Finally in Figure 2 we see the difference between the trimer GAA and the trimer GCA in the same context (i.e with the same flanking sequence).

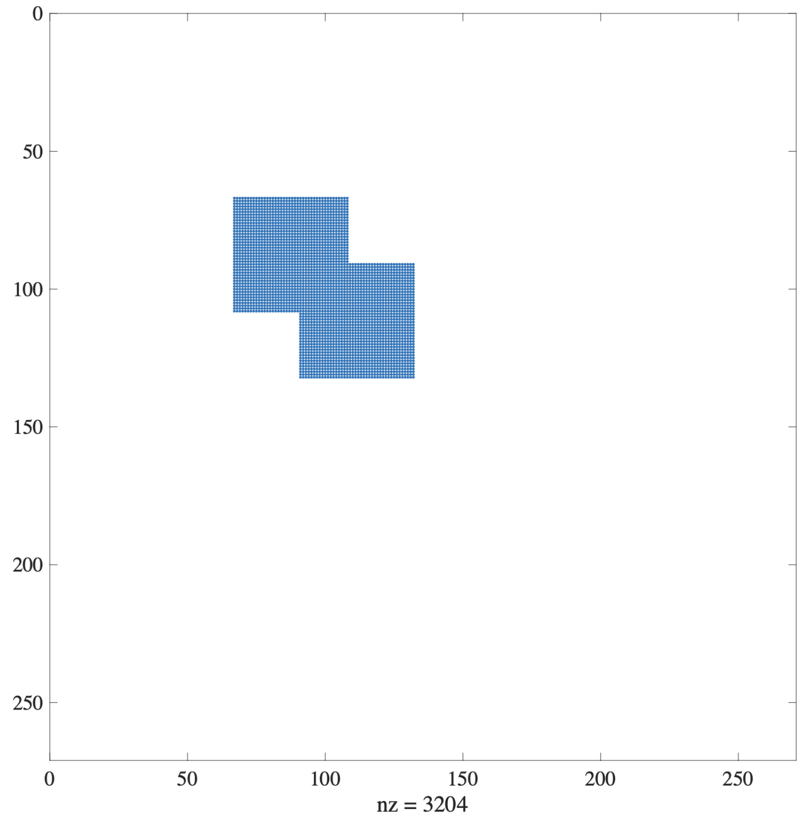

Figure 2: `spy` of $K - \tilde{K}$ with tolerance $= 10^{-10}$ where $\tilde{S} = CGCGCATTCGCG$.

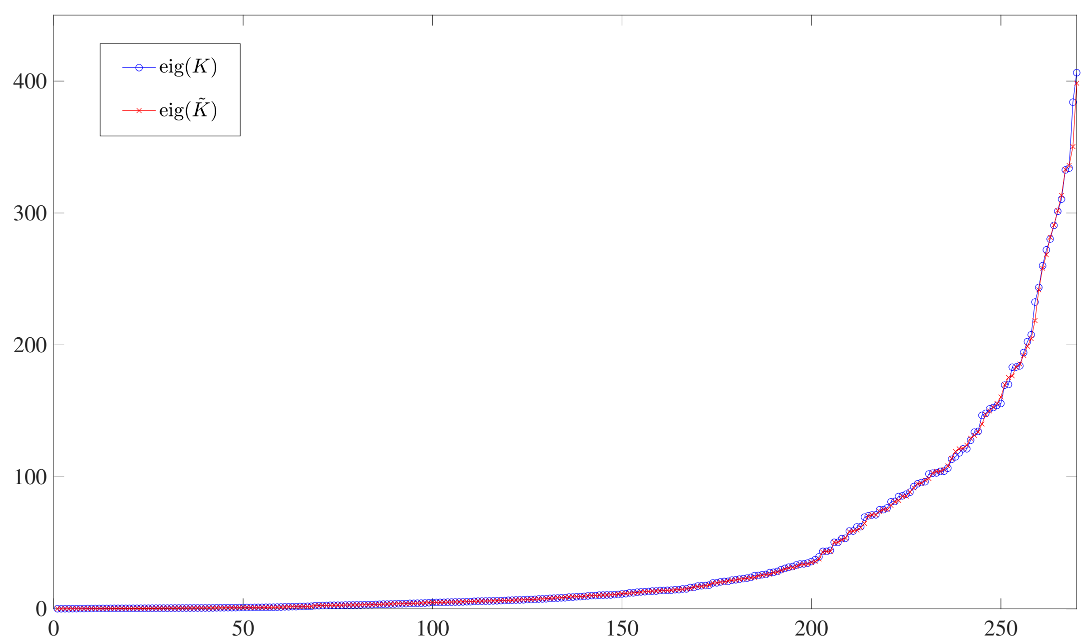

Figure 3: Sorted eigenvalues of K and $\tilde{K}$ as a function of degrees of freedom (ndof=270).

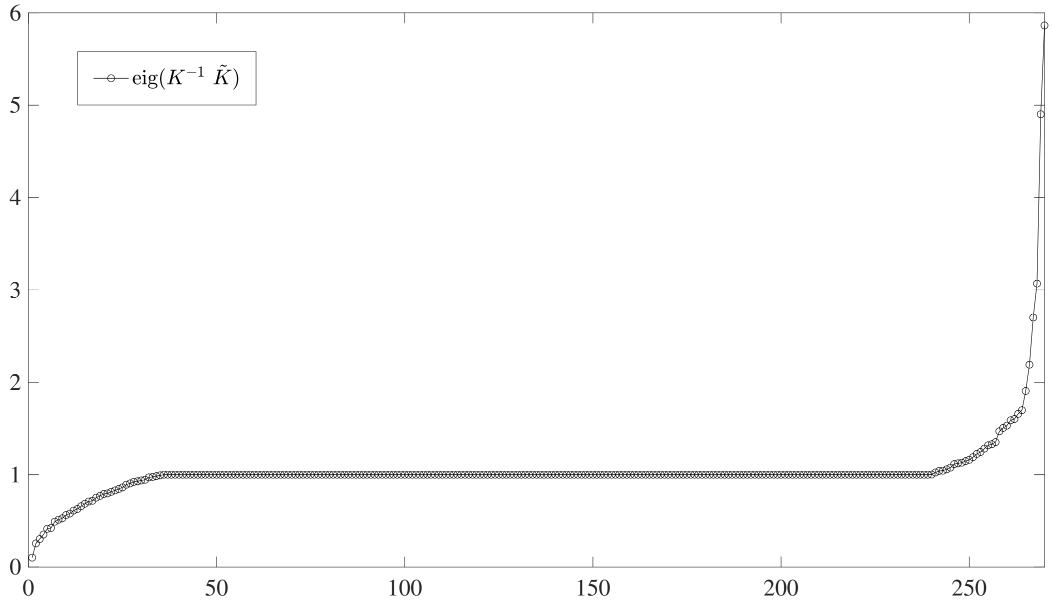

Figure 4: Sorted eigenvalues of $K^{-1}\tilde{K}$ as a function of degrees of freedom (ndof=270).

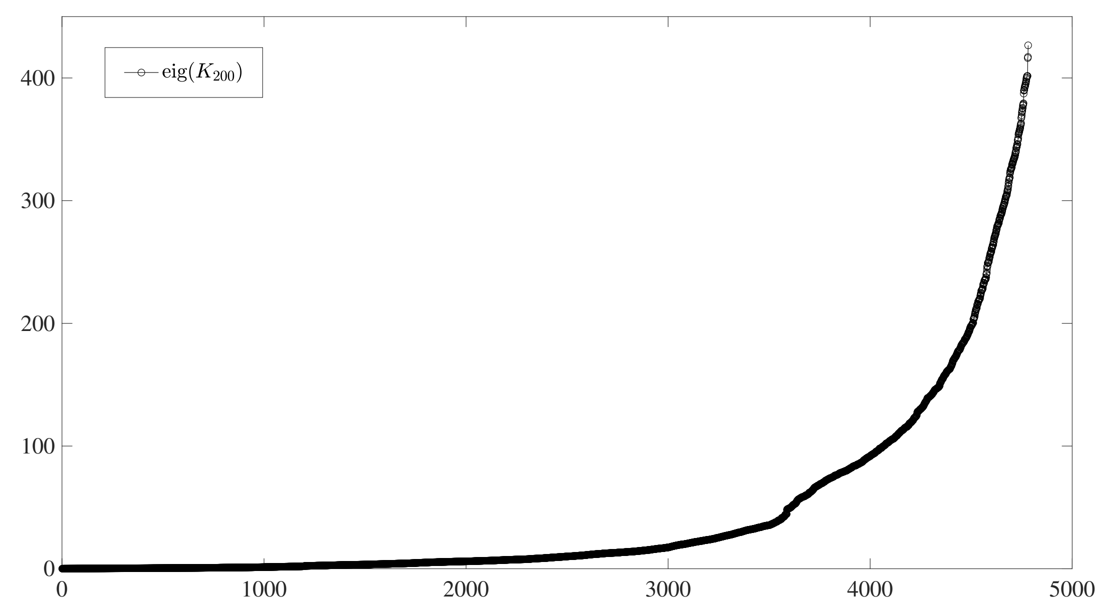

Figure 5: Sorted eigenvalues of K_{200} for a random 200 basepair long sequence, as a function of degrees of freedom (ndof=4782).

3. In Figure 3 we compare the eigenvalues of K and $\tilde{K}$. We can observe that the single point mutation did not influence the eigenvalues. As said in the statement of this exercise, comparing two symmetric matrix is not a simple task. Another way to compare two symmetric positive definite matrices $K_1, K_2 \in \mathbb{R}^{n\times n}$ in a non-dimensional way that captures both differences in eigenvalues and in eigenvectors is to consider the generalised eigenvalue problem

$$K_1\mathbf{x} = \lambda K_2\mathbf{x},$$

This problem always has n positive eigenvalues, and the two matrices K_1 and K_2 are identical if and only if all the eigenvalues are 1. We will discuss this idea extensively later in the course. Here we merely show in Figure 4 the eigenvalues for the generalised eigenvalue problem between the original and point mutation cgDNA+ stiffness matrices.

4. In order to generate a random sequence we first set that each of the four bases have same probability to get picked. Then the sequence is constructed by picking randomly a number p from 0 to 1 and by assigning it a base, for example if $p < 0.25$ next base is A, if $p \geq 0.25$ and $p < 0.5$ next base is C, and so on. Let define K_{200} to be the stiffness matrix for the computed random sequence. In Figure 5 we plot its sorted eigenvalues. An interesting observation is that the range of the values of the eigenvalues do not increase by increasing the number of basepair. This property is a property of the cgDNA+ stiffness matrix.

1.2 Visualization of the ground state

In Figure 6 & 7 we plot the helical parameters and phosphate coordinates. The first observation is that the sequence $S = CGCGAATTCGCG$ is a palindrome and that can be observed easily by inspecting the helical parameters of the ground-state shown in Figure 6. One can observe that the first column (Tilt-Shift-Buckle-Shear) are odd functions with respect to the middle junction while the other two columns are even with respect to the middle junction. The latter property is a property of the palindromes and is due to the change of reading strand properties of the cgDNA+ coordinates. This property will be explored in details in the coming weeks with more exercises.

1. The visualization of the 3D shape can be a useful tools for the study and analysis of the ground-state. In Figure 8 we show the bases and phosphates as rigid bodies whose color scheme (for bases) depend upon the sequence. By observing closely we can see some features of the palindromic sequence already observed in Figure 6. For example we can observe the change in sign of the Buckle for the 4th base pair.

2. Another way of visualizing the ground-state is just to plot all the entries of the vector at once in one Figure. This way could be useful when one have to compare two or more groundstates in a quick way. Also you can construct, visualize and compare the groundstate of more than one sequence directly on the web server cgDNAweb+ (`http://cgdnaweb.epfl.ch`).

3. By running the `main.m` function of the cgDNA+ package it automatically generates and saves a '.pdb' file that contains the Cartesian coordinates of all the atoms forming all the bases and phosphates of the predicted fragment. You can generate a '.pdb' file for any sequence you like and then visualize it using molviewer in Matlab.

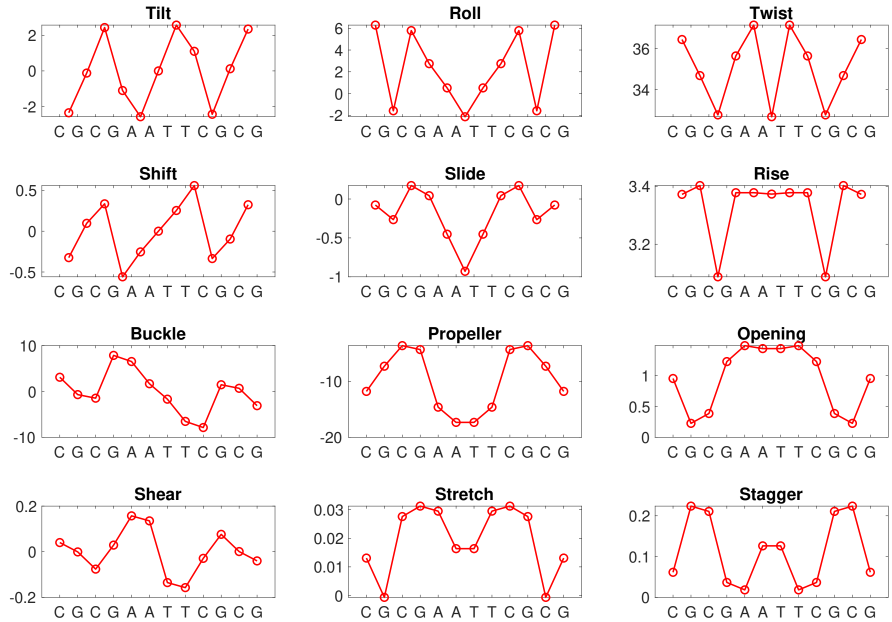

Figure 6: Helical parameters of the groundstate for the sequence S in Curves+ sense (translations in Angstroms and rotations in degrees). The rows 1 and 2 are the components of the inters while rows 3 and 4 are components of the intras. Rows 1 and 3 are the rotation parts while rows 2 and 4 are the translation parts. See Figure 1 of Serie 5 for the cartoon of the helical parameters.

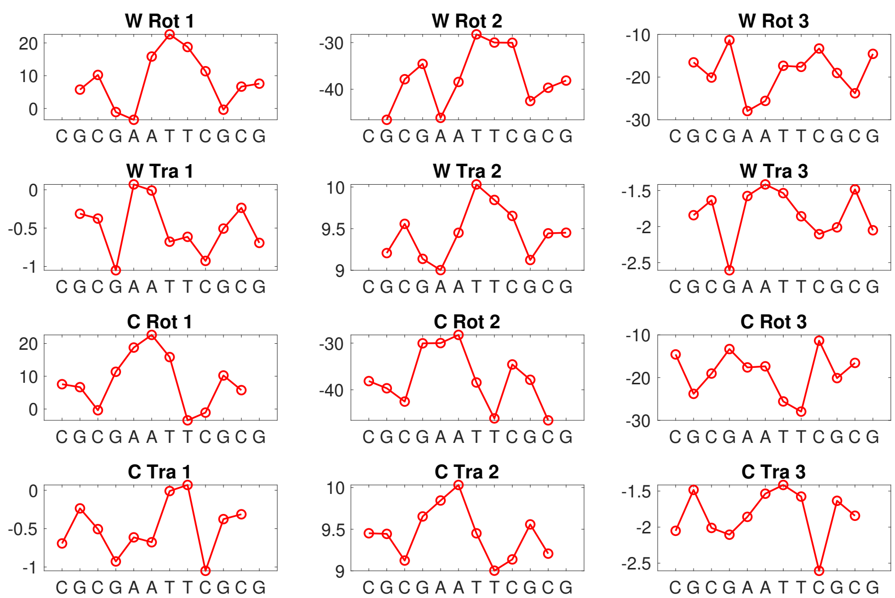

Figure 7: Phosphate coordinates of the ground-state for the sequence S (translations in Angstroms and rotations in degrees). The rows 1 and 2 correspond to Watson (W) phosphate while rows 3 and 4 correspond to Crick (C) phosphate. Rows 1 and 3 are the rotation parts while rows 2 and 4 are the translation parts. See Figure 2 Serie 5 for the cartoon of phosphates.

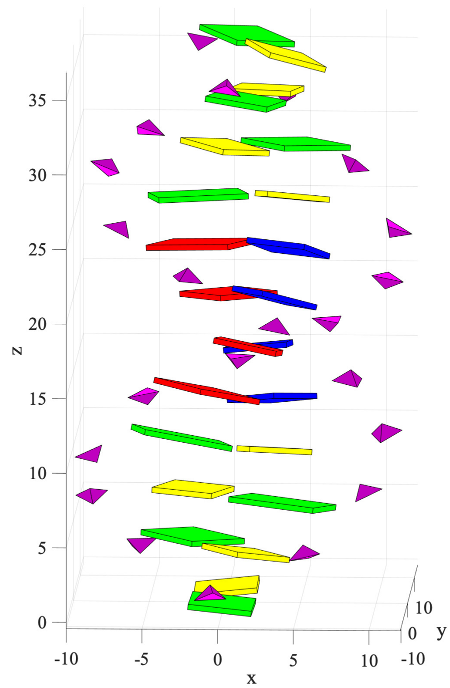

Figure 8: 3D view of the sequence S with the color scheme : yellow - C , green - G , blue - T , red - A and purple for phosphates.

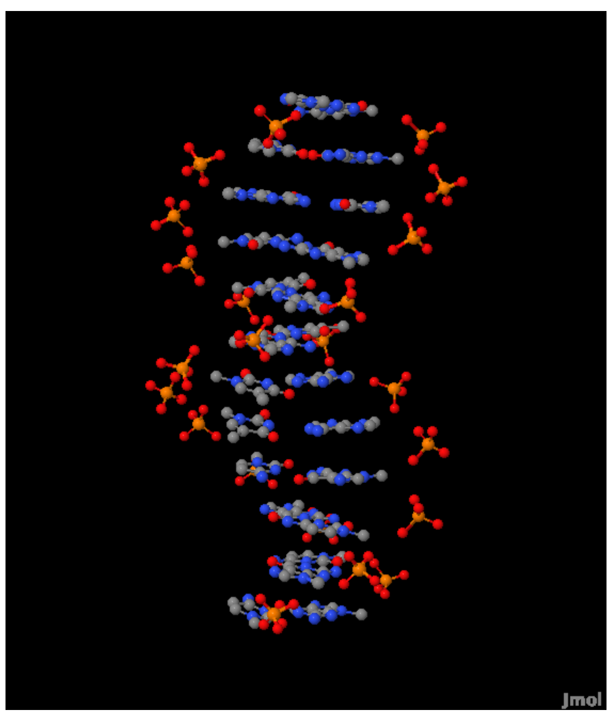

Figure 9: Using molviewer('.pdb') in matlab to visualise ground state predicted by cgDNA+.

1.3 Understanding the function `frames`

1. Let $z = (y_1, y_1^{pC}, x_1, y_2^{pW}, y_2, y_2^{pC}, x_2, \ldots, x_{N-1}, y_N^{pW}, y_N) \in \mathbb{R}^{24N-18}$ be the groundstates for the sequence S_N. We recall that: $y_i = (\eta_i, w_i) \in \mathbb{R}^6$ are the intras, $y_i^{pC} = (\eta_i^{pC}, w_i^{pC}) \in \mathbb{R}^6$ are the coordinates of Crick phosphates, $y_i^{pW} = (\eta_i^{pW}, w_i^{pW}) \in \mathbb{R}^6$ are the coordinates of Watson phosphates and $x_i = (u_i, v_i) \in \mathbb{R}^6$ are the inters. Notice that for the first basepair step there is just Crick phosphate and for last basepair step there is just Watson phosphate (see Figure 8 for 3d cartoon), So i will run : from 1 to $N-1$ for Crick phosphates and inters, 2 to N for Watson phosphates, and 1 to N for intras.

   Here η_i are Cayley vectors encoding the relative intra-basepair rotations while w_i are intra-basepair translations, $\{\eta_i^{pC}, \eta_i^{pW}\}$ are relative base to phosphates rotations while $\{w_i^{pC}, w_i^{pW}\}$ are base to phosphates translations for Crick and Watson phosphates respectively, and u_i are Cayley vectors encoding the relative inter-basepair rotations while v_i are inter-basepair translations.

   The reconstruction rules for computing the absolute coordinates of base and basepair frames $\{(R_i^-, r_i^-), (R_i, r_i), (R_i^+, r_i^+)\}_{i=1}^N$, and for Crick phosphates $\{(Rp_i^-, rp_i^-)\}_{i=1}^{N-1}$ and for Watson phosphates $\{(Rp_i^+, rp_i^+)\}_{i=2}^N$ are the following:

Inter's part:

$$\begin{bmatrix} R_{i+1} & r_{i+1} \\ \mathbf{0}^T & 1 \end{bmatrix} = \begin{bmatrix} R_i & r_i \\ \mathbf{0}^T & 1 \end{bmatrix} \begin{bmatrix} P(u_i) & P(u_i)^{\frac{1}{2}} v_i \\ \mathbf{0}^T & 1 \end{bmatrix} = \prod_{j=1}^{i} \begin{bmatrix} P(u_j) & P(u_j)^{\frac{1}{2}} v_j \\ \mathbf{0}^T & 1 \end{bmatrix}, \tag{1}$$

where we have chose as first basepair frames $(R_1, r_1) = (\mathbf{I}_3, \mathbf{0})$, and $P(u_i) = (I + \frac{1}{10}[u_i \times])(I - \frac{1}{10}[u_i \times])^{-1}$ is the Cayley transform of $[u_i \times]$.

Intra's part:

$$\begin{aligned} R_i^- &= R_i Q(\eta_i)^{-\frac{1}{2}}, \; r_i^- = r_i - \frac{1}{2} R_i w_i && (2) \\ R_i^+ &= R_i^- Q(\eta_i), \; r_i^+ = r_i^- + R_i w_i, && (3) \end{aligned}$$

and here $Q(\eta_i) = (I + \frac{1}{10}[\eta_i \times])(I - \frac{1}{10}[\eta_i \times])^{-1}$ is the Cayley transform of $[\eta_i \times]$.

Phosphate's part:

$$\begin{aligned} Rp_i^- &= R_i^- P Q(\eta_i^{pC}), \; rp_i^- = r_i^- + R_i^- P w_i^{pC} && (4) \\ Rp_i^+ &= R_i^+ Q(\eta_i^{pW}), \; rp_i^+ = r_i^+ + R_i^+ w_i^{pW}, && (5) \end{aligned}$$

in above equations again $Q(\eta_i^{pC}) = (I + \frac{1}{10}[\eta_i^{pC} \times])(I - \frac{1}{10}[\eta_i^{pC} \times])^{-1}$ is the Cayley transform of $[\eta_i^{pC} \times]$ and $Q(\eta_i^{pW}) = (I + \frac{1}{10}[\eta_i^{pW} \times])(I - \frac{1}{10}[\eta_i^{pW} \times])^{-1}$ is the Cayley transform of $[\eta_i^{pW} \times]$ and P is defined in equation 9.

   Remark: In this exercise we used two different representations of a same SE(3) element, i.e , for us an element $g \in \mathrm{SE}(3)$ can be represented as a couple $g = (R, r)$ where $R \in \mathrm{SO}(3)$ and $r \in \mathbb{R}^3$ or as a matrix $g = \begin{bmatrix} R & r \\ \mathbf{0}^T & 1 \end{bmatrix}$, where again $R \in \mathrm{SO}(3)$ and $r \in \mathbb{R}^3$. The rotation part R and the translation part r are the same in both representation. One can interpret the first representation as a frame which orientation is given by R and position is given by r while the second representation can be interpreted as a rigid body motion of the lab frame, which, in matrix representation, is just the identity matrix $\mathbf{I}_4 \in \mathbb{R}^{4\times4}$ (i.e. rotation part: $\mathbf{I}_3$, translation part $\mathbf{0} \in \mathbb{R}^3$).

2. In function `frames.m` the already added parts are at lines: 14-15.

```
1  nbp = (numel(shapes)+18)/24;
2
3  % absolute coordinates of the first basepair
4  G = eye(3);
5  q = [0, 0, 0]';
6
7  % relative coordinates of the oligomer
8  [eta, w, etapW, wpW, u, v, etapC, wpC] = vector2shapes(shapes);
9
10 bp_level = InitializeStruct(nbp) ;
11
12 for i=1:nbp
13     %%%% store the absolute coordinates of the basepair
14     bp_level(i).R = G ;
15     bp_level(i).r = q ;
16
17     % base pair:
18     r  = cay(eta(i,:));
19     Gw = G * w(i,:)';
20
21     % complimentary strand
22     bp_level(i).Rc = G * (sqrtm(r))';
23     bp_level(i).rc = q - 0.5 * Gw;
24
25     % main strand
26     bp_level(i).Rw = bp_level(i).Rc * r;
27     bp_level(i).rw = bp_level(i).rc + Gw;
28
29     if i<nbp
30         ru = cay(u(i,:));
31         sqrtru = sqrtm(ru);
32         H = G * sqrtru;
33         % next base pair:
34         G = G * ru;
35         q = q + H * v(i,:)';
36
37     end
38 end
39
40 for i = 1:nbp - 1
41     % Watson phosphate
42     bp_level(i+1).Rpw  = bp_level(i+1).Rw*cay(etapW(i,:)) ;
43     bp_level(i+1).rpw  = bp_level(i+1).rw + bp_level(i+1).Rw*wpW(i,:)' ;
44
45     % Crick phosphate
46     Rc = bp_level(i).Rc*diag([1,-1,-1]) ; % unflipping Crick base
47     bp_level(i).Rpc = Rc*cay(etapC(i,:)) ;
48     bp_level(i).rpc = bp_level(i).rc + Rc*wpC(i,:)' ;
49
50 end
```

2 A MATLAB cgDNA viewer

Here `https://lcvmwww.epfl.ch/teaching/modelling_dna/protected_files/codes_exercises/cgDNA_viewer.zip` you can download the complete code of the cgDNA viewer and also the code for `viewer_main_run.m`, `cgDNAviewer_test.m`. Using the parameter set given in cgDNA+ Matlab package run script `viewer_main_run.m`.

```
% code to be added at the line 93 of getRigidBodyConfig.m
RB.conf = (R*verteces_RB')' + repmat(r',[8,1]);
```

3 Scaled Cayley transform

1. Let $\mathbf{u} \in \mathbb{R}^3$ and $\alpha \in \mathbb{R}$ with $\alpha \neq 0$. We have that

$$\begin{aligned} CayTra\left(\frac{1}{\alpha}[\mathbf{u}\times]\right) &= I + \frac{2}{1+\|\frac{1}{\alpha}\mathbf{u}\|^2}\left([\frac{1}{\alpha}\mathbf{u}\times] + [\frac{1}{\alpha}\mathbf{u}\times]^2\right) \\ &= I + \frac{2\alpha^2}{\alpha^2+\|\mathbf{u}\|^2}\frac{1}{\alpha}\left([\mathbf{u}\times] + \frac{1}{\alpha}[\mathbf{u}\times]^2\right) \\ &= I + \frac{2\alpha}{\alpha^2+\|\mathbf{u}\|^2}\left([\mathbf{u}\times] + \frac{1}{\alpha}[\mathbf{u}\times]^2\right). \end{aligned} \tag{6}$$

2. If the relation between $Q \in SO(3)$ and $\mathbf{u}$ is known, for any $\alpha \in \mathbb{R}$, $\alpha \neq 0$ we can define

$$CayTra^{-1}(Q) = \frac{\alpha}{(1+\mathrm{tr}(Q))}\left(Q - Q^T\right) = [\alpha\mathbf{u}\times], \tag{7}$$

which implies that

$$\|CayTra^{-1}(Q)\| = \alpha\tan\left(\frac{\phi}{2}\right) = \|\alpha\mathbf{u}\|. \tag{8}$$

3. The routine `cay` in the function `frames.m` should be changed as follow:

```
function [Q] = cay(u)

    I = eye(3) ;
    alpha = 10 ;
    X = [   0    -u(3)  u(2) ;
           u(3)    0   -u(1) ;
          -u(2)  u(1)    0 ] ;
    Q = I + 2*alpha/(alpha^2 + norm(u)^2)*(X + 1/alpha*X^2) ;

end
```

4 Proof of the change of reading strand transformation Part–2

Remark: Let $z = (y_1, y_1^{pC}, x_1, y_1^{pW}, y_2, y_2^{pC}, x_2, \ldots, x_{N-1}, y_N^{pW}, y_N) \in \mathbb{R}^{24n-18}$ be the cgDNA+ model internal coordinates for the sequence S_N, where $y_i = (\eta_i, w_i) \in \mathbb{R}^6$ are the intras, $y_i^{pC} = (\eta_i^{pC}, w_i^{pC}) \in \mathbb{R}^6$ are the coordinates of Crick phosphates, $y_i^{pW} = (\eta_i^{pW}, w_i^{pW}) \in \mathbb{R}^6$ are the coordinates of Watson phosphates and $x_i = (u_i, v_i) \in \mathbb{R}^6$ are the inters. Here η_i are Cayley vectors encoding the relative intra-basepair rotations while w_i are intra-basepair translations, $\{\eta_i^{pC}, \eta_i^{pW}\}$ are relative base to phosphates rotations while $\{w_i^{pC}, w_i^{pW}\}$ are base to phosphates translations for Crick and Watson phosphates, and u_i are Cayley vectors encoding the relative inter-basepair rotations while v_i are inter-basepair translations. Notice that for the first basepair step there is just Crick phosphate and for last basepair step there is just Watson phosphate.

The DNA fragment (with sequence $S_N = X_1, X_2, \dots, X_N$) we are considering is described by the internal coordinates

$$z = \Big((\eta_1, w_1), (\eta_1^{pC}, w_1^{pC}), (u_1, v_1), (\eta_1^{pW}, w_1^{pW}), (\eta_2, w_2), (\eta_2^{pC}, w_2^{pC}), (u_2, v_2), \dots, (u_{N-1}, v_{N-1}),$$
$$(\eta_{N-1}^{pW}, w_{N-1}^{pW}), (\eta_N, w_N)\Big) = \Big(y_1, y_1^{pC}, x_1, y_1^{pW}, y_2, y_2^{pC}, x_2, \dots, x_{N-1}, y_{N-1}^{pW}, y_N\Big)$$

where $y_i = (\eta_i, w_i) \in \mathbb{R}^6$ are the intra variables, $x_l = (u_l, v_l) \in \mathbb{R}^6$ are the inter variables and $y_i^{pC} = (\eta_i^{pC}, w_i^{pC}) \in \mathbb{R}^6$ are the coordinates of Crick phosphates and $y_i^{pW} = (\eta_i^{pW}, w_i^{pW}) \in \mathbb{R}^6$ are the coordinates of Watson phosphates.

We first write down the base–pair frames and junction frames associated to $\{(R_i, r_i)^C, (R_i, r_i)^W\}_{i=1}^N$:

Define the i–th base–pair frame

For all $i = 1, 2, \dots, N$:

- $R_i = R_i^C([R_i^C]^T R_i^W)^{\frac{1}{2}}$
- $r_i = \dfrac{1}{2}(r_i^C + r_i^W)$

Define the l–th junction frame

For all $l = 1, 2, \dots, N-1$:

- $J_l = R_l(R_l^T R_{l+1})^{\frac{1}{2}}$
- $j_l = \dfrac{1}{2}(r_l + r_{l+1})$

For the change of reading strand transformation the matrix $P \in O(3)$ we have to chose is

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}. \tag{9}$$

The next step is to identify the frames of the starting configuration with the frames of the transformed one. It is easy to see that the relation between the two configuration is:

$$\begin{aligned} \overline{R}_{N+1-i}^C &= R_i^W P \quad , \quad \overline{r}_{N+1-i}^C = r_i^W \\ \overline{R}_{N+1-i}^W &= R_i^C P \quad , \quad \overline{r}_{N+1-i}^W = r_i^C \\ \overline{R}_{N+1-i} &= R_i P \quad , \quad \overline{r}_{N+1-i} = r_i \\ \overline{J}_{N-l} &= J_l P \quad , \quad \overline{j}_{N-l} = j_l. \end{aligned} \tag{10}$$

Define now the two linear transformation of the indices: $\sigma(i) = N + 1 - i$ and $\gamma(l) = N - l$. By using Exercise 2.1 of Serie 4, we can define the internal coordinate of the transformed configuration:

Inter variables

$$\forall l = 1, \dots, N-1 \begin{cases} \overline{u}_{\gamma(l)} = -P^T u_l \\ \overline{v}_{\gamma(l)} = -P^T v_l \end{cases} \Rightarrow \overline{x}_{\gamma(l)} = diag(-P^T, -P^T) x_l = E x_l.$$

Here for each l we used the Exercise 5.1 with $\{(R_l, r_l), (J_l, j_l), (R_{l+1}, r_{l+1})\}$ and the transformation P.

Intra variables

$$\forall i = 1,\dots,N \begin{cases} \overline{\eta}_{\sigma(i)} = -P^T \eta_i \\ \overline{w}_{\sigma(i)} = -P^T w_i \end{cases} \Rightarrow \overline{y}_{\sigma(i)} = diag(-P^T, -P^T) y_i = E y_i.$$

Here for each i we used the Exercise 5.1 with $\{(R_i^C, r_i^C), (R_i, r_i), (R_i^W, r_i^W)\}$ and the transformation P.

Phosphate variables

First recall that phosphate coordinates are defined with respect to corresponding bases (before flipping) for a given strand (see Figure 2 Serie 5), i.e. for i^{th} phosphate of Watson strand, the relative internal coordinates are defined with respect to i^{th} base on Watson strand. Now on switching the strands the i^{th} phosphate on Watson strand will become $N+1-i$ phosphate on Crick strand, so, following linear transformation of indices still holds,

$$\sigma(i) = N + 1 - i$$

Moreover the relative coordinates of phosphates are defined before the base flipping (see Figure 2 Serie 5). It means that the phosphates can be simply mapped from one strand to another strand just by choosing matrix P to be a identity matrix $I_3 \in \mathbb{R}^{3\times 3}$

$$\forall i = 2,\dots,N \begin{cases} \overline{\eta}^{pC}_{\sigma(i)} = I_3 \eta_i^{pW} \\ \overline{w}^{pC}_{\sigma(i)} = I_3 w_i^{pW} \end{cases} \Rightarrow \overline{y}^{pC}_{\sigma(i)} = diag(I_3, I_3) y_i^{pW} = I y_i^{pW}.$$

similarly,

$$\forall i = 2,\dots,N \begin{cases} \overline{\eta}^{pW}_{\sigma(i)} = I_3 \eta_i^{pC} \\ \overline{w}^{pW}_{\sigma(i)} = I_3 w_i^{pC} \end{cases} \Rightarrow \overline{y}^{pW}_{\sigma(i)} = diag(I_3, I_3) y_i^{pC} = I y_i^{pC}.$$

where $I = diag(1,1,1,1,1,1)$. Finally the change of reading strand transformation implies a change also on the sequence of the transformed DNA, i.e, $\overline{S}_N = \overline{X}_N, \overline{X}_{N-1}, \dots, \overline{X}_1$. Now, the internal coordinates of the DNA fragment with sequence $\overline{S}_N$ is described by

$$\overline{z} = (\overline{y}_1, \overline{y}_1^{pC}, \overline{x}_1, \overline{y}_1^{pW}, \overline{y}_2, \overline{y}_2^{pC}, \overline{x}_2, \dots, \overline{x}_{N-1}, \overline{y}_N^{pW}, \overline{y}_N) \tag{11}$$

Using the indices transformation σ and γ, and the relation on the inter, intra and phosphate variables, we can rewrite $\overline{z}$ as

$$\overline{z} = (Ey_N, y_N^{pW}, Ex_{N-1}, \dots, Ex_1, y_1^{pC}, Ey_1) = \begin{bmatrix} & & & & & & E \\ & & & & & I & \\ & & & & E & & \\ & & & \iddots & & & \\ & & E & & & & \\ & I & & & & & \\ E & & & & & & \end{bmatrix} \begin{bmatrix} y_1 \\ y_1^{pC} \\ x_1 \\ \vdots \\ x_{N-1} \\ y_N^{pW} \\ y_N \end{bmatrix} \tag{12}$$

5 On the symmetry of the coordinates system

1. The complementary sequence of S is $\bar{S} = GCTTTTTTTTTCAC$. For checking that $\mathbf{z}(S) = E_N \mathbf{z}(\bar{S})$ one can compute $\|\mathbf{z}(S) - E_N \mathbf{z}(\bar{S})\|$. While for the stiffness one can use the command

`spy` in order to visualize the difference $dK = K(S) - E_N K(\bar{S}) E_N$. Be aware that the command `spy` shows all non zero entries of the matrix or vector passed as argument. In order to filter out small entries one can use the following trick: `spy(abs(dK)>1e-10 )`.

2. The sequence S is a palindrome which means that $S = \bar{S}$. This implies the following:

$$\begin{aligned} \mathbf{z}(S) &= E_N \mathbf{z}(\bar{S}) = E_N \mathbf{z}(S), \\ K(S) &= E_N K(\bar{S}) E_N = E_N K(S) E_N. \end{aligned}$$

From the above equation for the stiffness matrix K, and by using the first part of this question, we can conclude that $K(S) - E_N K(S) E_N$ is the zeros matrix. Finally in order to understand why the Shift in junction 6 is zero it is important to understand what happen to $\mathbf{z}(S)$ when multiplied by E_N. More precisely what happen to the intras and the inters of $\mathbf{z}(S)$. See also the solution of the exercise 4 of this sheet. There will be an exercise later on in the semester where we will better explore the action of the matrix E_N on the ground-state and on the stiffness of palindromic sequences.