

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

1. For proving the positiveness of $K = K(S, \mathcal{P})$ we will prove that $z^T K z > 0$, for any $z \in \mathbb{R}^{24N-18}$. Lets decompose the matrix K into 42×42 blocks for interior dimer steps and 36×36 blocks for end dimes steps:

$$z^T K z = \sum_{i=2}^{N-2} z_i^T K_i z_i + z_1^T K_1 z_1 + z_{N-1}^T K_{N-1} z_{N-1} \quad (1)$$

where

$$\begin{aligned} K_i &= K^{X_i X_{i+1}}, \forall i = 2, \dots, N-2, \\ K_1 &= K^{5' X_1 X_2}, \\ K_{N-1} &= K^{X_{N-1} X_N 3'}, \end{aligned}$$

and $z_i \in \mathbb{R}^{42}$ for interior dimer steps and $z_i \in \mathbb{R}^{36}$ for end dimes steps. Here, z_i are the column vector that forms z when combined on the corresponding positions with 18×1 overlap. This is very similar to computing oligomer based $\sigma(S) \in \mathbb{R}^{24N-18}$ for any sequence that is required to compute the groundstate for any sequence S . Using the assumptions we can conclude that $z^T K z > 0$, for any $z \in \mathbb{R}^{24N-18}$.

2. We denote by $\bar{S} = \bar{X}_N \bar{X}_{N-1} \dots \bar{X}_2 \bar{X}_1$ the complementary of the sequence S , and $\bar{K} := K(\bar{S}, \mathcal{P})$. The position of the base j of \bar{S} in the original sequence is $\epsilon(j) = N - j + 1$, for example, the first base in \bar{S} is the last one in S . Now, with the decomposition used in part 1) we have, $\forall z \in \mathbb{R}^{24N-18}$:

$$z^T \bar{K} z = \sum_{j=2}^{N-2} z_j^T \bar{K}_j z_j + z_1^T \bar{K}_1 z_1 + z_{N-1}^T \bar{K}_{N-1} z_{N-1}, \quad (2)$$

where the definition of $\bar{K}_j, j = 2, \dots, N-2, \bar{K}_1, \bar{K}_{N-1}$ are defined in the same way as in the previous point. The only thing one has to be careful is to associate the right base pair to the matrices \bar{K}_j using the linear change of indices ϵ . Using the symmetry assumption and the latter remark, we obtain :

$$\bar{K}_j = K^{\bar{X}_{\epsilon(j)} \bar{X}_{\epsilon(j+1)}} \quad (3)$$

$$= E^{int} K^{X_{\epsilon(j+1)} X_{\epsilon(j)}} E^{int} \quad (4)$$

$$= E^{int} K_{\epsilon(j+1)} E^{int}. \quad (5)$$

A similar computation leads to

$$\bar{K}_1 = E^{3'} K_{N-1} E^{3'} \quad (6)$$

$$\bar{K}_{N-1} = E^{5'} K_1 E^{5'}. \quad (7)$$

Finally we obtain that

$$\begin{aligned}
z^T \overline{K} z &= \sum_{j=2}^{N-2} z_j^T E^{int} K_{\epsilon(j)} E^{int} z_j + z_1^T E^{3'} K_{N-1} E^{3'} z_1 + z_{N-1}^T E^{5'} K_1 E^{5'} z_{N-1} \\
&= \sum_{j=2}^{N-2} (E^{int} z_j)^T K_{\epsilon(j)} (E^{int} z_j) + (E^{3'} z_1)^T K_{N-1} (E^{3'} z_1) + (E^{5'} z_{N-1})^T K_1 (E^{5'} z_{N-1}) \\
&= \sum_{i=1}^{N-2} (E^{int} z_{N-i+1})^T K_i (E^{int} z_{N-i+1}) + (E^{3'} z_1)^T K_{N-1} (E^{3'} z_1) + (E^{5'} z_{N-1})^T K_1 (E^{5'} z_{N-1}) \\
&= (E_N z)^T K (E_N z) = z^T E_N K E_N z
\end{aligned}$$

Thus, we obtained that $\overline{K} = E_N K E_N$. We can prove in a similar manner that $\overline{\sigma} = E_N \sigma$. Finally, by using the fact that $E_N = E_N^T = E_N^{-1}$ we obtain that

$$\mu(\overline{S}, \mathcal{P}) = \overline{K}^{-1} \overline{\sigma} = E_N K^{-1} E_N E_N \sigma = E_N K^{-1} \sigma = E_N \mu(S, \mathcal{P}). \quad (8)$$

3. The number of independent dimers is 10 for the interior blocks (using the symmetry condition given in the part 2 of this exercise we can consider only 10 independent conditions). For the end blocks we gave two sets of 16 conditions and again using the symmetry conditions given in the part 2 we can consider only one set of conditions as a independent set for the end blocks. Thus, the total number of independent conditions are 26 (10 for the interior blocks plus 16 for the end blocks).

2 Zero entries in the cgDNA+ parameter set

- i) We consider first $\sigma^{\alpha\bar{\alpha}}$ with $\alpha \in \{A, T, G, C\}$. We use the following notation:

$$\sigma^{\alpha\bar{\alpha}} = (y_1^{pW} \in \mathbb{R}^6, y_1 \in \mathbb{R}^6, y_1^{pC} \in \mathbb{R}^6, x \in \mathbb{R}^6, y_2^{pW} \in \mathbb{R}^6, y_2, y_2^{pC} \in \mathbb{R}^6) \in \mathbb{R}^{42}$$

The symmetry condition leads to

$$\sigma^{\alpha\bar{\alpha}} = E^{int} \sigma^{\alpha\bar{\alpha}}, \quad (9)$$

Finally by computing the right hand side of (9) we get the following restrictions:

$$\begin{aligned}
x(1) &= 0 \\
x(4) &= 0 \\
y_1 &= E y_2 \\
y_1^{pW} &= y_2^{pC} \\
y_2^{pW} &= y_1^{pC}
\end{aligned} \quad (10)$$

Here $x(1), x(4)$ are first and fourth elements of x . Thus for $\sigma^{\alpha\bar{\alpha}}$ we obtain total 20 restrictions. We consider now the stiffness matrix $K^{\alpha\bar{\alpha}}$ and the symmetry condition $K^{\alpha\bar{\alpha}} = E^{int} K^{\alpha\bar{\alpha}} E^{int}$. The first thing to understand is how the blocks of $K^{\alpha\bar{\alpha}}$ permute with the double multiplication with E^{int} . Schematically, we can divide $K^{\alpha\bar{\alpha}}$ into fourty-nine 6×6 block matrices and, for sake of simplicity, we can number each block by taking into account the symmetry of the matrix $K^{\alpha\bar{\alpha}}$. Note that $K^{\alpha\bar{\alpha}}$ is symmetric it means all the diagonal blocks are also symmetric. So,

will denote by "-" the entries that do change sign:

$$\begin{aligned}
EAE &= \begin{bmatrix} -1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix} A \begin{bmatrix} -1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & -1 & & \\ & & & & 1 & \\ & & & & & 1 \end{bmatrix} \\
&= \begin{bmatrix} + & - & - & + & - & - \\ - & + & + & - & + & + \\ - & + & + & - & + & + \\ + & - & - & + & - & - \\ - & + & + & - & + & + \\ - & + & + & - & + & + \end{bmatrix} \tag{11}
\end{aligned}$$

Finally, we remark that the inter block of $K^{\alpha\bar{\alpha}}$, the block K_{44} , do not change position when multiplied twice by E^{int} . Thus, we have the following condition

$$K_{44} = EK_{44}E,$$

thanks to the +/- scheme of (11), we obtain the following restrictions on the entries of $K^{\alpha\bar{\alpha}}$:

$$\begin{aligned}
K_{([19,22],[20,21,23,24])}^{\alpha\bar{\alpha}} &= 0 \\
K_{([20,21,23,24],[19,22])}^{\alpha\bar{\alpha}} &= 0
\end{aligned}$$

The other conditions on the diagonal blocks are given by the two following relations:

$$\begin{aligned}
K_{11} &= K_{77}, \\
K_{22} &= EK_{66}E, \\
K_{33} &= K_{55}
\end{aligned} \tag{12}$$

which means that by knowing K_{77} one can retrieve K_{11} and similarly for K_{22} and K_{33} . Further the conditions on anti-diagonal blocks are given by the two following relations:

$$\begin{aligned}
K_{17} &= K_{17}^T, \\
K_{26} &= EK_{26}^T E, \\
K_{35} &= K_{35}^T,
\end{aligned} \tag{13}$$

This relation implies that, in absolute values, the upper triangular part of K_{17} is equal to the upper triangular part of K_{17}^T , thus the total number of independent entries of K_{17} are the same as for a symmetric matrix. We stress that K_{17} is not symmetric, the entries in its upper triangular part are equal to the entries in the lower triangular part except for the entries corresponding to the +/- scheme (11). Similarly, for K_{26} and K_{35} blocks. Finally, there are

conditions on rest of the off-diagonal blocks. For example,

$$\begin{aligned}
K_{12} &= IK_{67}^T E = (EK_{67}I)^T, \\
K_{13} &= K_{57}^T, \\
K_{14} &= IK_{47}^T E = (EK_{47}I)^T, \\
K_{15} &= K_{37}^T, \\
K_{16} &= IK_{27}^T E = (EK_{27}I)^T, \\
K_{23} &= EK_{56}^T I, \\
K_{24} &= EK_{46}^T E = (EK_{46}E)^T, \\
K_{25} &= EK_{36}^T I, \\
K_{26} &= EK_{26}^T E = (EK_{26}E)^T, \\
K_{34} &= IK_{45}^T E = (EK_{45}I)^T, \\
K_{35} &= IK_{35}^T I,
\end{aligned} \tag{14}$$

- ii) For any palindromic sequence S of arbitrary length N we can say that N is even and that the middle dimer (step between the $\frac{N}{2}$ and $\frac{N}{2} + 1$ base-pair) of S must be a palindromic dimer. We recall also that the reconstructions of $K(S)$ and $\sigma(S)$ is local but $\mu(S)$ has a non local reconstruction because it is computed using the inverse of K that is dense. Before finding all the restrictions on K, σ , and μ , we recall that

$$K(\bar{S}) = E_N K(S) E_N, \tag{15}$$

$$\sigma(\bar{S}) = E_N \sigma(S), \tag{16}$$

$$\mu(\bar{S}) = E_N \mu(S). \tag{17}$$

We remark that the mid dimer $X_{\frac{N}{2}} X_{\frac{N}{2}+1}$ of S correspond to the mid dimer of \bar{S} and that $\bar{X}_{\frac{N}{2}} \bar{X}_{\frac{N}{2}+1} = X_{\frac{N}{2}} X_{\frac{N}{2}+1}$ because it is a palindromic dimer. Denote now $K_{mid} \in \mathbb{R}^{42 \times 42}$, $\sigma_{mid} \in \mathbb{R}^{42}$, and $\mu_{mid} \in \mathbb{R}^{42}$ the components of respectively K, σ , and μ , associated to the mid dimer step. Using the relations (15,16,17) we obtain that :

$$K_{mid}(\bar{S}) = E^{int} K_{mid}(S) E^{int},$$

$$\sigma_{mid}(\bar{S}) = E^{int} \sigma_{mid}(S),$$

$$\mu_{mid}(\bar{S}) = \mu_{mid}(S).$$

and we can reuse the restrictions founded in the previous parts. Thus, the components of K, σ , and μ related to the inter block are zero for a palindrome. Next step is to study the difference between a local reconstruction and a non local reconstruction of a palindrome S of length N when S has additional palindromic dimers different from the middle one. In fact one can check that if $X_i X_{i+1}$ is a palindromic dimer with $i \neq \frac{N}{2}$, we have that $\bar{X}_{N-i} \bar{X}_{N-i+1} = X_i X_{i+1}$, because is a palindrome. Thus, we will obtain the same restrictions we obtain for the middle components only for K and σ because for μ we will not have the same components for the related base-pairs because μ has a non local reconstruction. Use the cgDNA+ matlab package to verify the findings. You can find the reconstruction of σ vector, in the function `constructSeqParams`.

3 Total number of unknowns in a cgDNA+ parameter set

Let $\mathcal{P} = \{K^{\alpha\beta} \in \mathbb{R}^{42 \times 42}, \sigma^{\alpha\beta} \in \mathbb{R}^{42}, K^{5'\alpha\beta} \in \mathbb{R}^{36 \times 36}, \sigma^{5'\alpha\beta} \in \mathbb{R}^{36} | \alpha\beta \in D\}$. Using the Crick and Watson symmetries we decrease the numbers of dimers to 10 independent (four palindromic plus six non

palindromic dimers) dimers for the interior blocks. For the end blocks we have one sets (we are saying that lets take all the dimers corresponding to 5' end as a independent and parameters for dimes corresponding to 3' end will be computed using the symmetry) of 16 dimers. So in total we have 26 dimer blocks. In this way we decrease the total number of element in the parameter set to $20682(10 \times 42 \times 43/2 + 16 \times 36 \times 37/2 + 10 \times 42 + 16 \times 36)$. This number comes out by computing:

- 1) the entries of the 10 symmetric matrices $K^{\alpha\beta} \in \mathbb{R}^{42 \times 42}$,
- 2) the entries of the 16 symmetric matrices $K^{5'\alpha\beta} \in \mathbb{R}^{36 \times 36}$,
- 3) the entries of the 10 vectors $\sigma^{\alpha\beta} \in \mathbb{R}^{42}$,
- 4) the entries of the 16 vectors $\sigma^{5'\alpha\beta} \in \mathbb{R}^{36}$.

Next we will count how many independent entries the cgDNA+ paramset has. For this we use the restrictions for the palindromic dimers founded in exercise 2 of this session. One can easily found that for the fours $\sigma^{\alpha\bar{\alpha}}$ we have 20 restrictions. For the stiffness matrices $K^{\alpha\bar{\alpha}}$ is slightly more difficult. Using the same notation as it is used in the solution of the exercise 2 of this session, we have that the following numbers of entries are not independent for $K^{\alpha\bar{\alpha}}$:

- 1) 8 entries of the block K_{44} ,
- 2) 21 entries of the blocks $K_{17}, K_{26}, K_{35}, K_{11}, K_{22}, K_{33}$,
- 3) 36 entries of the blocks $K_{12}, K_{13}, K_{14}, K_{15}, K_{16}, K_{23}, K_{24}, K_{25}, K_{26}, K_{34}, K_{35}$

Finally we obtained that for each $\alpha\bar{\alpha}$ we have 550 not independent entries ($K^{\alpha\bar{\alpha}}$ plus $\sigma^{\alpha\bar{\alpha}}$) and we have total 4 palindromic dimers ($\alpha\bar{\alpha}$), which lead to a total of $18482(20682 - 4 \times 530 - 4 \times 20)$ independent entries in the cgDNA+ parameter set $\mathcal{P} = \{K^{\alpha\beta}, \sigma^{\alpha\beta}, K^{5'\alpha\beta}, \sigma^{5'\alpha\beta} | \alpha\beta \in \tilde{D}\}$, where \tilde{D} is a set having 10 independent dimers for the interior blocks plus 16 independent dimers corresponding to 5' end blocks.