

# 1 On the computation of marginals of the cgDNA probability distribution

## 1.1 Marginalise over intra-base-pair variables

1. i) See Figure (1).

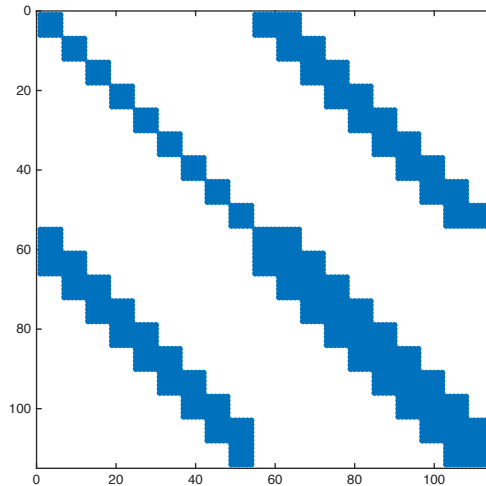


Figure 1: Sparsity of  $\tilde{K}_D$

- ii) The marginal stiffness matrix  $K_1^{(u,u)}$  is dense.
2. i)  $\tilde{\Sigma}_D$  does not have any specific pattern as the covariance matrix  $\Sigma_D$  is dense.
- ii) The marginal stiffness matrix  $K_2^{(u,u)}$  is dense.

## 1.2 A localized cgDNA model: marginalise over the configurations of the flanking sequences

The sparsity pattern of the marginalised stiffness matrix is the same as the cgDNA one, i.e, 18 times 18 blocks with 6 times 6 overlaps. Just be aware that if you want to use the MATLAB function `spy` to visualize the sparsity pattern of the matrix you should use the following combination : `spy(abs(K_marginal)> 1e-10)`. You can now use your Kullback-Leibler divergence to estimate the difference between the marginal and the cgDNA reconstruction of  $S_D$ , or you can visualize the difference using the function `plotMatrix2D` given in exercise 2 of this sheet. It is possible to prove why the localized marginal still banded, but this prove is beyond the purpose of this course as it involves the prove for the Maximum Entropy fit of banded Gaussian. For the motivated students we refer to the Phd thesis of Jaroslaw Glowacki, "Computation and Visualization in Multiscale Modelling of DNA Mechanics", 2016, EPFL Thesis #7062, Chapter P1.1 where you can find the prove of the Maximum Entropy fit and make yourself and idea about the prove of the localized marginal.

Here ([http://lcvwww.epfl.ch/teaching/modelling\\_dna/public\\_files/LocalizedCgDNA.m](http://lcvwww.epfl.ch/teaching/modelling_dna/public_files/LocalizedCgDNA.m)) you can download the code for computing the marginal over the configurations of the flanking sequence.

## 2 Gaussian Integral III

Let define  $f_{X,Y}(x, y)$  the joint distribution of the two random variables  $X$  and  $Y$ . The distribution of  $X$  given that  $Y$  is fixed to a value  $y$ , can be written as

$$f_X(x|Y=y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad (1)$$

where  $f_Y(y)$  is the marginal distribution for  $Y$ . In the case of Gaussian distribution we can explicitly compute the latter ratio and find a close form for the conditional distribution. The main idea is to decompose the quadratic form in the argument of the exponential into the marginal plus the conditional. This computation has already been done in the solution of the Exercise 1, Session 9, where we aimed at computing the marginal distribution of a Gaussian. Here we propose a similar decomposition than the one presented in the solution of the Exercise 1 of Session 9:

$$(\mathbf{x} - \hat{\mathbf{x}}) \cdot K(\mathbf{x} - \hat{\mathbf{x}}) = (\mathbf{x}_2 - \hat{\mathbf{x}}_2) \cdot \Sigma_{22}^{-1}(\mathbf{x}_2 - \hat{\mathbf{x}}_2) + (\mathbf{x}_1 - (\hat{\mathbf{x}}_1 + \eta)) \cdot K_{11}(\mathbf{x}_1 - (\hat{\mathbf{x}}_1 + \eta)), \quad (2)$$

where  $\Sigma_{22}^{-1} = K_{22} - K_{12}^T K_{11}^{-1} K_{12}$  and  $\eta = -K_{11}^{-1} K_{12}(\mathbf{x}_2 - \hat{\mathbf{x}}_2)$ . By using the first result obtained in the solution of Exercise 1 of Session 9 one can define the blocks of  $K$  in term of the blocks of  $\Sigma$ , and one can find that  $K_{11} = (\Sigma_{11} - \Sigma_{21}^T \Sigma_{22}^{-1} \Sigma_{21})^{-1}$ , and that  $\eta = \Sigma_{12} \Sigma_{22}^{-1}(\mathbf{x}_2 - \hat{\mathbf{x}}_2)$ . Finally by defining :

$$f(\mathbf{x}) = f(\mathbf{x}_1, \mathbf{x}_2) = \frac{1}{Z} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}}) \cdot K(\mathbf{x} - \hat{\mathbf{x}}) \right\}, \quad (3)$$

$$f_2(\mathbf{x}_2) = \frac{1}{Z_2} \exp \left\{ -\frac{1}{2}(\mathbf{x}_2 - \hat{\mathbf{x}}_2) \cdot \Sigma_{22}^{-1}(\mathbf{x}_2 - \hat{\mathbf{x}}_2) \right\}, \quad (4)$$

We obtain that

$$f(\mathbf{x}_1|\mathbf{x}_2 = \mathbf{a}) = \frac{f(\mathbf{x}_1, \mathbf{a})}{f_2(\mathbf{a})} = \frac{1}{\bar{Z}} \exp \left\{ -\frac{1}{2}(\mathbf{x}_1 - \bar{\mathbf{x}}) \cdot \bar{\Sigma}^{-1}(\mathbf{x}_1 - \bar{\mathbf{x}}) \right\}, \quad (5)$$

where  $\bar{\mathbf{x}} = \hat{\mathbf{x}}_1 + \Sigma_{12} \Sigma_{22}^{-1}(\mathbf{a} - \hat{\mathbf{x}}_2)$ ,  $\bar{\Sigma} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}$ .

## 3 On the computation of conditionals of the cgDNA probability distribution

1. Let us recall that we are working on the following Gaussian:

$$\rho(\mathbf{w}; S, \mathcal{P}) = \frac{1}{Z} \exp \left\{ -\frac{1}{2} \begin{bmatrix} w_1 - \hat{w}_1 \\ y_i - \hat{y}_i \\ w_2 - \hat{w}_2 \end{bmatrix} \begin{bmatrix} A & B & 0 \\ B^T & C & D^T \\ 0 & D & E \end{bmatrix} \begin{bmatrix} w_1 - \hat{w}_1 \\ y_i - \hat{y}_i \\ w_2 - \hat{w}_2 \end{bmatrix} \right\}. \quad (6)$$

We can now define the following matrices :

$$\begin{aligned} K_{11} &= \begin{bmatrix} A & 0 \\ 0 & E \end{bmatrix}, \\ K_{12} &= \begin{bmatrix} B \\ D \end{bmatrix}, \\ K_{22} &= C, \\ K &= \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^T & K_{22} \end{bmatrix}, \end{aligned}$$

and the following vectors:

$$\begin{aligned}\mathbf{x}_1 &= \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \\ \mathbf{x}_2 &= y_i, \\ \mathbf{x} &= \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix}.\end{aligned}$$

Finally we define  $\hat{\mathbf{x}}$  as done for  $\mathbf{x}$  but for the corresponding "hat" variables. Finally, we have rearranged the stiffness matrix and the mean of (6) in such a way that we can reuse the decomposition obtained in the previous Exercise, Equation (2), in order to get the conditional distribution for the variable  $\hat{\mathbf{x}}_1 = (w_1, w_2)$ . Finally we obtain that

$$\begin{aligned}\bar{w}_1 &= \hat{w}_1 - A^{-1}B(\mathbf{a} - \hat{y}_i) \\ \bar{w}_2 &= \hat{w}_2 - E^{-1}D(\mathbf{a} - \hat{y}_i)\end{aligned}$$

where we fixed  $y_i = \mathbf{a}$ . You can now modify the code and run the computation for the given arguments.