1 Principle of maximum entropy parameter estimation for banded stiffness matrices

According to the maximum entropy principle, the distribution $\rho_{ME}(x)$ can be defined as

$$\rho_{ME} = \text{argmin}_{\rho \in C} S[\rho] \text{ where } S[\rho] = \int_\Omega \rho(x) \ln \rho(x) dx.$$

The Lagrange multiplier method allows to write the distribution $\rho_{ME}(x) \in C$ as the solution of

$$\int_\Omega \{(1 + \ln \rho_{ME}(x)) - \lambda_0 - \lambda_1 \cdot x - [[\lambda_2]] : (x \otimes x)\} \delta\rho(x) dx = 0 \tag{1}$$

for any $\delta\rho \in \mathrm{L}^1(\Omega)$ and for some Lagrange multipliers $\lambda_0 \in \mathbb{R}$, $\lambda_1 \in \mathbb{R}^{24n-18}$ and $\lambda_2 \in \mathbb{R}^{(24n-18)\times(24n-18)}$. Note that we have used that the first variation of the functional $S[\rho]$ can be written as

$$\delta S[\rho]\delta\rho = \int_\Omega (1 + \ln \rho(x))\delta\rho(x) dx$$

for any $\rho, \delta\rho \in \mathrm{L}^1(\Omega)$ and that

$$\delta \left\{ \int_\Omega \phi(x)\rho(x) dx \right\} \delta\rho = \int_\Omega \phi(x)\delta\rho(x) dx$$

for any $\phi \in \mathrm{L}^1(\Omega)$ to deduce 1. A sufficient condition is then that the distribution $\rho_{ME}(x)$ is normal, i.e. that it is of the form

$$\rho_{ME}(x) = \frac{1}{Z} \exp\left\{ -\frac{1}{2}(x - a) \cdot A(x - a) \right\}. \tag{2}$$

with

$$(\lambda_0 - 1) + \lambda_1 \cdot x + [[\lambda_2]] : (x \otimes x) = -\frac{1}{2}(x - a) \cdot A(x - a) - \ln Z \tag{3}$$

for all $x \in \Omega$. Moreover, since we have the constraints $\rho_{ME}(x) \in C$, we can directly deduce $\rho_{ME}(x)$ but we have to define

$$a = \mu, A = K_{ME} \text{ and } Z = \sqrt{\det(2\pi K_{ME})}$$

accordingly the identities regarding the first and second moment of a normal distribution. We note that the equality 3 allows then to compute explicitly the values of the Lagrange multipliers λ_0, λ_1 and λ_2.

2 Gaussian integrals II (Part 1 in Session 1)

First we will find expressions for the inverse and the determinant of a symmetric block matrix. Note that inverse of a symmetric matrix is also symmetric. Let $\mathbf{A} = \mathbf{A}^T \in \mathbb{R}^{n\times n}$, $\mathbf{C} = \mathbf{C}^T \in \mathbb{R}^{m\times m}$, $\mathbf{B} \in \mathbb{R}^{n\times m}$, $\alpha = \alpha^T \in \mathbb{R}^{n\times n}$, $\gamma = \gamma^T \in \mathbb{R}^{m\times m}$ and $\beta \in \mathbb{R}^{n\times m}$. We want to solve the equation

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{4}$$

which is equivalent to the system

$$\begin{cases} \mathbf{A}\alpha & + & \mathbf{B}\beta^{\mathbf{T}} & = & \mathbf{I} \\ \mathbf{A}\beta & + & \mathbf{B}\gamma & = & \mathbf{0} \\ \mathbf{B}^T\alpha & + & \mathbf{C}\beta^{\mathbf{T}} & = & \mathbf{0} \\ \mathbf{B}^T\beta & + & \mathbf{C}\gamma & = & \mathbf{I}, \end{cases} \tag{5}$$

and

$$\begin{cases} \beta^T = -\mathbf{C}^{-1}\mathbf{B}^T\alpha \\ \gamma = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{B}^T\beta \\ \mathbf{A}\alpha - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T\alpha = \mathbf{I} \\ \mathbf{A}\beta + \mathbf{B}\mathbf{C}^{-1} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T\beta = \mathbf{0}, \end{cases} \tag{6}$$

which gives

$$\begin{aligned} \alpha &= (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}, \\ \beta &= -(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{C}^{-1}, \\ \gamma &= \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}^T(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{C}^{-1}. \end{aligned} \tag{7}$$

The next step is to find a close form for the determinant of

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix}.$$

For that purpose we will use the following properties (given without proof) of the determinant of block diagonal matrices: Let Π be the following block matrix

$$\Pi = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \tag{8}$$

with $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D} \in \mathbb{R}^{n\times n}$, we have then that $\det \Pi = \det(\mathbf{A}\mathbf{D} - \mathbf{B}\mathbf{C})$. Now if we have that $\mathbf{C} = \mathbf{B} = \mathbf{0}$ we will obtain that $\det \Pi = \det \mathbf{A} \det \mathbf{D}$. It can be shown that the latter result also holds for square matrices $\mathbf{A} \in \mathbb{R}^{r\times r}$ and $\mathbf{D} \in \mathbb{R}^{n\times n}$ with $r \neq n$. One can further proof that if one of the two blocks $\mathbf{B}$ or $\mathbf{C}$ is non zeros, we will obtain that

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} = \det \mathbf{A} \det \mathbf{D}. \tag{9}$$

Moreover, for all symmetric positive definite matrices the following decomposition holds:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T & \mathbf{0} \\ \mathbf{C}^{-1}\mathbf{B}^T & \mathbf{I} \end{pmatrix}. \tag{10}$$

Then, using the previous remarks, we obtain that the determinant of our matrix is

$$\det \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} = \det \mathbf{C} \, \det[\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T] = \det \mathbf{C} \, \det[\alpha^{-1}]. \tag{11}$$

We will now identify: $\mathbf{A} = K_{11}$, $\mathbf{B} = K_{12}$, $\mathbf{C} = K_{22}$. Hence, we have found the relation between the blocks of the stiffness and the blocks of the covariance matrix.

Now we want to write the argument of the exponential as a sum of two quadratic forms, so that one of them would not depend on $\mathbf{x}_2$:

$$\begin{aligned} U(\mathbf{x}) &= (\mathbf{x} - \hat{\mathbf{x}}) \cdot K(\mathbf{x} - \hat{\mathbf{x}}) = [(\mathbf{x}_1 - \hat{\mathbf{x}}_1)^T, (\mathbf{x}_2 - \hat{\mathbf{x}}_2)^T] \begin{bmatrix} K_{11} & K_{12} \\ K_{12}^T & K_{22} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 - \hat{\mathbf{x}}_1 \\ \mathbf{x}_2 - \hat{\mathbf{x}}_2 \end{bmatrix} \\ &= (\mathbf{x}_1 - \hat{\mathbf{x}}_1)^T K_{11}(\mathbf{x}_1 - \hat{\mathbf{x}}_1) + 2(\mathbf{x}_1 - \hat{\mathbf{x}}_1)^T K_{12}(\mathbf{x}_2 - \hat{\mathbf{x}}_2) \\ &+ (\mathbf{x}_2 - \hat{\mathbf{x}}_2)^T K_{22}(\mathbf{x}_2 - \hat{\mathbf{x}}_2). \end{aligned}$$

We define $\eta = -K_{22}^{-1} K_{12}^T (\mathbf{x}_1 - \hat{\mathbf{x}}_1)$. Then

$$\begin{aligned} U(\mathbf{x}) &= (\mathbf{x}_1 - \hat{\mathbf{x}}_1)^T K_{11} (\mathbf{x}_1 - \hat{\mathbf{x}}_1) - \eta^T K_{22} \eta \\ &+ \big((\mathbf{x}_2 - \hat{\mathbf{x}}_2)^T K_{22} (\mathbf{x}_2 - \hat{\mathbf{x}}_2) - 2\eta^T K_{22} (\mathbf{x}_2 - \hat{\mathbf{x}}_2) + \eta^T K_{22} \eta \big) \\ &= (\mathbf{x}_1 - \hat{\mathbf{x}}_1)^T (K_{11} - K_{12} K_{22}^{-1} K_{12}^T)(\mathbf{x}_1 - \hat{\mathbf{x}}_1) \\ &+ (\mathbf{x}_2 - (\hat{\mathbf{x}}_2 + \eta))^T K_{22} (\mathbf{x}_2 - (\hat{\mathbf{x}}_2 + \eta)). \end{aligned}$$

From (7) we remark, that $K_{11} - K_{12} K_{22}^{-1} K_{12}^T = \Sigma_{11}^{-1}$ and from (11), that $\frac{1}{\det[K^{-1}]} = \det K = \det K_{22} \det[\Sigma_{11}^{-1}]$. We finally get

$$\begin{aligned} & \frac{\left(\frac{\beta}{\pi}\right)^{\frac{n}{2}}}{\sqrt{\det[K^{-1}]}} \int_{\mathbb{R}^m} e^{-\beta(\mathbf{x}-\hat{\mathbf{x}})\cdot K(\mathbf{x}-\hat{\mathbf{x}})} \, d\mathbf{x}_2 \\ = & \frac{\left(\frac{\beta}{\pi}\right)^{\frac{k}{2}}}{\sqrt{\det \Sigma_{11}}} e^{-\beta(\mathbf{x}_1-\hat{\mathbf{x}}_1)^T \Sigma_{11}^{-1} (\mathbf{x}_1-\hat{\mathbf{x}}_1)} \frac{\left(\frac{\beta}{\pi}\right)^{\frac{m}{2}}}{\sqrt{\det[K_{22}^{-1}]}} \int_{\mathbb{R}^m} e^{-\beta(\mathbf{x}_2-(\hat{\mathbf{x}}_2+\eta))^T K_{22} (\mathbf{x}_2-(\hat{\mathbf{x}}_2+\eta))} \, d\mathbf{x}_2 \\ = & \frac{\left(\frac{\beta}{\pi}\right)^{\frac{k}{2}}}{\sqrt{\det \Sigma_{11}}} e^{-\beta(\mathbf{x}_1-\hat{\mathbf{x}}_1)\cdot \Sigma_{11}^{-1} (\mathbf{x}_1-\hat{\mathbf{x}}_1)}. \end{aligned}$$

3 On the computation of marginals of the cgDNA+ probability distribution

3.1 Marginalise over phosphates variables

You will observe that the marginal stiffness matrix is dense and the sparsity pattern of cgDNA+ is lost while computing the marginals.

3.2 Marginalise over phosphates and intra-base-pair variables

Also for this case the marginal stiffness matrix is dense and do not show any sparsity pattern.

4 A localized cgDNA+ model: marginals over the configurations of the flanking sequences

Here (`http://lcvmwww.epfl.ch/teaching/modelling_dna/protected_files/codes_exercises/LocalizedCgDNAp.m`) you can download the matlab script for computing the marginal over the configurations of the flanking sequence.

In Figure (1) we show the sparsity pattern of marginalised stiffness matrix (K_M). Here sparsity pattern of the marginalised stiffness matrix is the same as the cgDNA+ one, i.e, 42×42 blocks with 18×18 overlaps in the interior of the sequence and 36×36 blocks with 18×18 overlaps for ends. Just be aware that if you want to use the matlab function `spy` to visualize the sparsity pattern of the stiffness matrix then you should use the following combination : `spy(abs(K_M)> 1e-10)`. In this particular example $S_1 = S_2 = poly(AA)_{50}$ is considered but you can choose any random sequences you like. In Figure (2) we show the differences in the groundstate variables between the marginalised groundstate (μ_M) and cgDNA+ groundstate (μ) and one can notice the non-local changes in groundstate due to marginalisation. You can use your Kullback-Leibler divergence script

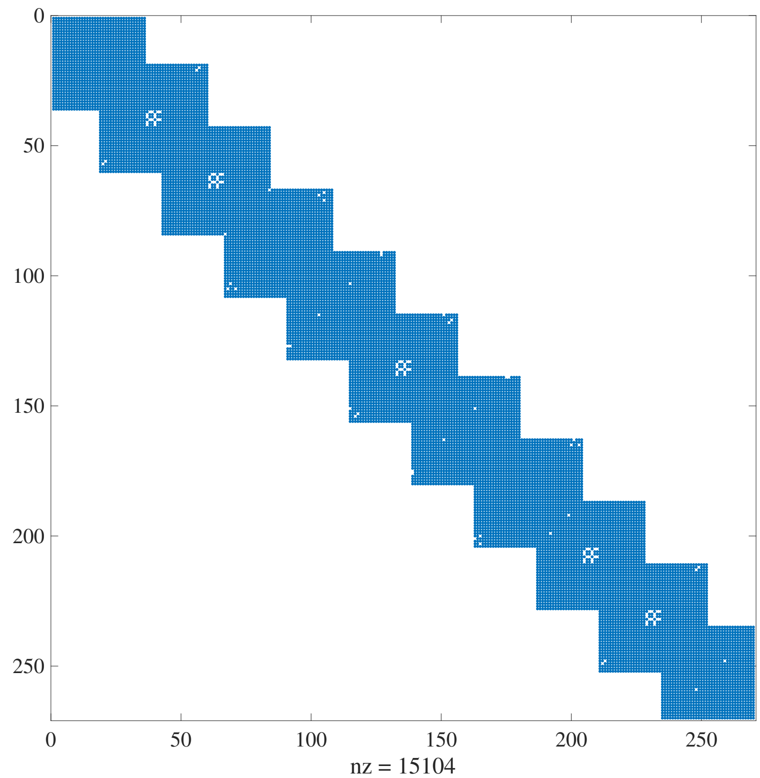

Figure 1: Sparsity pattern of marginalised stiffness matrix (K_M).

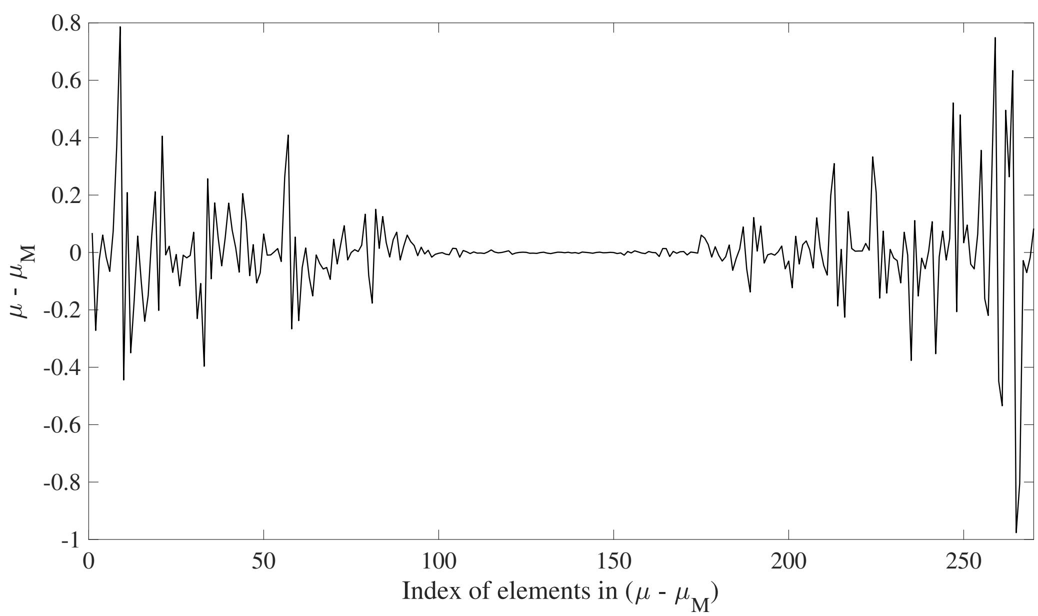

Figure 2: Differences in the groundstate coordinates between the marginalised groundstate (μ_M) and cgDNA+ reconstruction (μ) for S_D.

to estimate the difference between the marginal and the cgDNA+ reconstruction of S_D. In the following table we reported the result of the Kullback-Leibler divergence computations (per degree of freedom) between the marginal and the cgDNA+ reconstruction of S_D:

KLD	stiffness part	mean part
0.0235	0.0131	0.0104

Remark: It is possible to prove why the localised 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 familiar with the idea about the prove of the localised marginal.