1 Relative entropy for Gaussians II

1. Let $M \in \mathbb{R}^{N\times N}$ and X_1 and X_2 the normal random variables with respectively probability density functions p and q. Define the change of variable $Y_i = MX_i$, $i = 1, 2$.

i) A linear transformation of a normally distributed random variable is another normally distributed random variable. Using the property of the expected value one can show that

$$\begin{aligned} E[MX_i] &= ME[X_i] = M\widehat{x}_i \\ E\left[(MX_i - E[MX_i])(MX_i - E[MX_i])^T\right] &= ME\left[(X_i - \widehat{x}_i)(X_i - \widehat{x}_i)^T\right]M^T = MK_i^{-1}M^T. \end{aligned}$$

for $i = 1, 2$.

ii) A direct computation gives:

$$\begin{aligned} D(\tilde{p}, \tilde{q}) &= \frac{1}{2}\left[\text{tr}(M^{-T}K_2M^{-1}MK_1^{-1}M^T) - \ln\frac{\text{det}M^{-T}K_2M^{-1}}{\text{det}M^{-T}K_2M^{-1}}\right] \\ &+ \frac{1}{2}M(\widehat{x}_1 - \widehat{x}_2)\cdot M^{-T}K_2M^{-1}M(\widehat{x}_1 - \widehat{x}_2) \\ &= \frac{1}{2}\left[\text{tr}(M^{-T}K_2K_1^{-1}M^T) - \ln\frac{\text{det}K_2}{\text{det}K_2}\right] \\ &+ \frac{1}{2}(\widehat{x}_1 - \widehat{x}_2)\cdot M^{-T}M^{-T}K_2M^{-1}M(\widehat{x}_1 - \widehat{x}_2) \\ &= \frac{1}{2}\left[\text{tr}(K_2K_1^{-1}M^TM^{-T}) - \ln\frac{\text{det}K_2}{\text{det}K_2}\right] + \frac{1}{2}(\widehat{x}_1 - \widehat{x}_2)\cdot K_2(\widehat{x}_1 - \widehat{x}_2) \\ &= \frac{1}{2}\left[\text{tr}(K_2K_1^{-1}) - \ln\frac{\text{det}K_2}{\text{det}K_2}\right] + \frac{1}{2}(\widehat{x}_1 - \widehat{x}_2)\cdot K_2(\widehat{x}_1 - \widehat{x}_2) \\ &= D(p, q). \end{aligned}$$

2. The generalised eigenvalue problem can be rewritten as

$$K_2v_i = \mu_iK_1v_i \Rightarrow K_1^{-1}K_2v_i = \mu_iv_i \Rightarrow \text{tr}(K_2K_1^{-1}) = \sum_{i=1}^N \mu_i\ ,\ \ln\det K_2K_1^{-1} = \sum_{i=1}^N \ln\mu_i \tag{1}$$

We can also rewrite the definition of $D^\dagger$ as

$$D^\dagger(K_1, K_2) = \frac{1}{2}\left[\text{tr}\left(K_2K_1^{-1}\right) - \ln\det K_2K_1^{-1} - N\right]. \tag{2}$$

By combining the consequences of the generalized eigenvalues problem and the latter formula we obtain:

$$D^\dagger(K_1, K_2) = \frac{1}{2}\sum_{i=1}^N(\mu_i - \ln\mu_i - 1)\,. \tag{3}$$

3. i) By using the fact that $\ln\frac{\det K_2}{\det K_1} = \ln\det K_2 - \ln\det K_1$, we get that the symmetrized version of the stiffness part of the relative entropy between two Gaussians is :

$$D^\dagger_{sym} = \frac{1}{4}\text{tr}\left(K_2K_1^{-1} + K_1K_2^{-1}\right) - \frac{N}{2}. \tag{4}$$

ii) By using the same argument used in part 2 of this exercise we find that

$$D^{\dagger}_{sym} = \frac{1}{4}\sum_{i=1}^{N}\left(\sqrt{\mu_i} - \frac{1}{\sqrt{\mu_i}}\right)^2. \tag{5}$$

2 Maximum entropy fit for the stiffness matrix

Recall that the matrix C is the following partitioned matrix:

$$C = \begin{bmatrix} a & e & \mathrm{x} \\ e^T & b & d \\ \mathrm{x}^T & d^T & c \end{bmatrix}, \ a = a^T, \ b = b^T, \ c = c^T,$$

where we have assumed $C > 0$. We want to show that if $\mathrm{x} = eb^{-1}d$, then $K := C^{-1}$ has zeros blocks in the $(1,3)$ and $(3,1)$ entries. For that we will prove the following statements:

1. By direct computation we have

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & \Psi & I \end{bmatrix}\begin{bmatrix} a & e & 0 \\ e^T & b & 0 \\ 0 & 0 & H \end{bmatrix}\begin{bmatrix} I & 0 & 0 \\ 0 & I & \Omega \\ 0 & 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & \Psi & I \end{bmatrix}\begin{bmatrix} a & e & e\Omega \\ e^T & b & b\Omega \\ 0 & 0 & H \end{bmatrix}$$

$$= \begin{bmatrix} a & e & e\Omega \\ e^T & b & b\Omega \\ \Psi e^T & \Psi b & \Psi b\Omega + H \end{bmatrix}$$

Now by using the definition of x, Ψ, Ω, and H we can conclude that

$$C = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & \Psi & I \end{bmatrix}\begin{bmatrix} a & e & 0 \\ e^T & b & 0 \\ 0 & 0 & H \end{bmatrix}\begin{bmatrix} I & 0 & 0 \\ 0 & I & \Omega \\ 0 & 0 & I \end{bmatrix}, \tag{6}$$

2. In order to compute the inverse of C using the decomposition (6), one have to check that the two following matrices

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\Psi & I \end{bmatrix}, \begin{bmatrix} I & 0 & 0 \\ 0 & I & -\Omega \\ 0 & 0 & I \end{bmatrix}$$

are actually the inverses of

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & \Psi & I \end{bmatrix} \text{ and } \begin{bmatrix} I & 0 & 0 \\ 0 & I & \Omega \\ 0 & 0 & I \end{bmatrix},$$

Moreover, the inverse of a block diagonal matrix (without overlaps) is the inverse of each block. Thus, we obtain that $C^{-1} := K$ is

$$K = \begin{bmatrix} I & 0 & 0 \\ 0 & I & -\Omega \\ 0 & 0 & I \end{bmatrix}\begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta & 0 \\ 0 & 0 & H^{-1} \end{bmatrix}\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\Psi & I \end{bmatrix},$$

where $\begin{bmatrix} \alpha & \varepsilon \\ \varepsilon^T & \beta \end{bmatrix} = \begin{bmatrix} a & e \\ e^T & b \end{bmatrix}^{-1}$

3.

$$
\begin{aligned}
K &= \begin{bmatrix} I & 0 & 0 \\ 0 & I & -\Omega \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta & 0 \\ 0 & 0 & H^{-1} \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\Psi & I \end{bmatrix} \\
&= \begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta + \Omega H^{-1}\Psi & -\Omega H^{-1} \\ 0 & H^{-1}\Psi & H^{-1} \end{bmatrix}.
\end{aligned} \tag{7}
$$

Thus, the blocks outside the stencil are zero.

4. In order to derive the algorithm we first decompose (7) in the following manner:

$$
\begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta + \Omega H^{-1}\Psi & -\Omega H^{-1} \\ 0 & -H^{-1}\Psi & H^{-1} \end{bmatrix} = \begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & -b^{-1} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & b^{-1} + \Omega H^{-1}\Psi & -\Omega H^{-1} \\ 0 & -H^{-1}\Psi & H^{-1} \end{bmatrix},
$$

where the first term of the right hand side is the inverse of $\begin{bmatrix} a & e \\ e^T & b \end{bmatrix}$, the second term is minus the inverse of b and the last term is the inverse of $\begin{bmatrix} b & d \\ d^T & c \end{bmatrix}$. The algorithm then is easy:

Step 1: Create an empty (only zero entries) matrix K of the same size as C,

Step 2: Invert each block of C inside the stencil and add them to K in the same position as they appear in C,

Step 3: Invert each overlap in the stencil and subtract them to the overlaps of K, again the position of the overlap must stay the same.

Remark: We recall that the inverse of a partitioned matrix $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is

$$
\begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}
$$

5. No, it is not possible to capture the five block $a, b, c, d,$ and e from the non zeros blocks of K.

3 Kullback-Leibler divergence between : $\rho_{obs}(S)$, $\rho_{band}(S)$, $\rho_{cgDNA}(S, \mathcal{P})$

1. Here (`http://lcvmwww.epfl.ch/teaching/modelling_dna/public_files/KL_Div.m`) you can find a possible way of coding the Kullback-Leibler divergence. We stress here that the formula of the Kullback-Leibler divergence for Gaussian implies the computation of the log of a determinant. In general in MATLAB one has to avoid the computation of the determinant of large positive definite matrix as the value could easily hit the MATLAB infinity. As the determinant of a matrix is equal to the product of its eigenvalues, the log of the determinant can de computed as the sum of the logs of the eigenvalues. The latter strategy is implemented in `KL_Div.m`.

2. In the following table we reported the result of the computations:

	KLD	stiffness part	mean part
$D(\rho_{band}(S), \rho_{obs}(S))$	11.6101	11.6101	0
$D(\rho_{cgDNA}(S, \mathcal{P}), \rho_{band}(S))$	5.5201	3.4391	2.0811