1 Banded matrices and their inverses

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

$$C = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & \Psi & I \end{bmatrix} \begin{bmatrix} a & e & 0 \\ e^T & b & 0 \\ 0 & 0 & H \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & \Omega \\ 0 & 0 & I \end{bmatrix}, \tag{1}$$

2. In order to compute the inverse of C using the decomposition (1), 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.

$$
\begin{aligned}
K &= \begin{bmatrix} I & 0 & 0 \\ 0 & I & -\Omega \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta & 0 \\ 0 & 0 & H^{-1} \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & -\Psi & I \end{bmatrix} \\
&= \begin{bmatrix} \alpha & \varepsilon & 0 \\ \varepsilon^T & \beta + \Omega H^{-1}\Psi & -\Omega H^{-1} \\ 0 & -H^{-1}\Psi & H^{-1} \end{bmatrix}. \qquad (2)
\end{aligned}
$$

Thus, the blocks outside the stencil are zero.

4. In order to derive the algorithm we first decompose (2) 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.

Note: 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.

Remarks: following points are useful for the computations of next question:

1. if $a \in \mathbb{R}$, then $\operatorname{tr}(a) = a$.

2. The trace is invariant under cycling permutation, i.e, $\operatorname{tr}(ABC) = \operatorname{tr}(BCA) = \operatorname{tr}(CAB)$.

3. Let $A, B \in \mathbb{R}^{n\times n}$, we define the following matrix inner product:

$$
A : B = \sum_{i,j=1}^{n} a_{ij} b_{ij} = \operatorname{tr}(B^T A) = \operatorname{tr}(A^T B).
$$

4. Consequence of the previous remarks:

$$
(x - \widehat{x})^T K (x - \widehat{x}) = \operatorname{tr}\left((x - \widehat{x})^T K (x - \widehat{x})\right) = \operatorname{tr}\left(K (x - \widehat{x})(x - \widehat{x})^T\right) = K : (x - \widehat{x})(x - \widehat{x})^T, \quad (3)
$$

for all shifted quadratic form with $K = K^T > 0$.

2 Entropy and Relative entropy formulas for Gaussians

1) Define

$$p(x) = \frac{1}{(2\pi)^{N/2}|K_1^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \widehat{x}_1) \cdot K_1(x - \widehat{x}_1)\right\}, \ x \in \mathbb{R}^N,$$

and compute the entropy of p:

$$\begin{aligned} h(p) &= -\int_{\mathbb{R}^N} p(x)\ln(p(x))dx \\ &= -\int_{\mathbb{R}^N} \left(\frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right) + \frac{1}{2}(x - \widehat{x}_1) \cdot K_1(x - \widehat{x}_1)\right)p(x)dx \\ &= -\frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right) - \frac{1}{2}\int_{\mathbb{R}^N} (x - \widehat{x}_1) \cdot K_1(x - \widehat{x}_1)p(x)dx \\ &= -\frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right) - \frac{1}{2}\left\langle(x - \widehat{x}_1)^T K_1(x - \widehat{x}_1)\right\rangle_p \qquad (4) \end{aligned}$$

Let us study the second term in (4):

$$\begin{aligned} \left\langle(x - \widehat{x}_1)^T K_1(x - \widehat{x}_1)\right\rangle_p &= \left\langle K_1 : (x - \widehat{x}_1)(x - \widehat{x}_1)^T\right\rangle_p \qquad (5) \\ &= K_1 : \left\langle(x - \widehat{x}_1)(x - \widehat{x}_1)^T\right\rangle_p \qquad (6) \\ &= K_1 : K_1^{-1} = \mathrm{tr}\left(K_1 K_1^{-1}\right) = N. \qquad (7) \end{aligned}$$

We can now conclude that

$$\begin{aligned} h(p) &= -\frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right) - \frac{1}{2}N \\ &= -\frac{1}{2}\ln\left((2\pi e)^N|K^{-1}|\right), \end{aligned} \qquad (8)$$

2) Here we will use the fact that the relative entropy between p and q can be seen as the expectation with respect to p of the quantity $\ln\left(\frac{p}{q}\right)$. Recall that

$$\begin{aligned} p(x) &= \frac{1}{(2\pi)^{N/2}|K_1^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \widehat{x}_1) \cdot K_1(x - \widehat{x}_1)\right\}, \ x \in \mathbb{R}^N, \\ q(x) &= \frac{1}{(2\pi)^{N/2}|K_2^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(x - \widehat{x}_2) \cdot K_2(x - \widehat{x}_2)\right\}, \ x \in \mathbb{R}^N, \end{aligned}$$

$$
\begin{aligned}
D(p,q) = \left\langle \ln \frac{p}{q} \right\rangle_p & \\
&= \frac{1}{2}\left\langle -\ln|K_1^{-1}| - (x-\widehat{x}_1)\cdot K_1(x-\widehat{x}_1) + \ln|K_2^{-1}| + (x-\widehat{x}_2)\cdot K_2(x-\widehat{x}_2)\right\rangle \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} + \langle -(x-\widehat{x}_1)\cdot K_1(x-\widehat{x}_1)\rangle + \langle (x-\widehat{x}_2)\cdot K_2(x-\widehat{x}_2)\rangle\right] \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} - K_1 : \left\langle (x-\widehat{x}_1)(x-\widehat{x}_1)^T\right\rangle + K_2 : \left\langle xx^T - x\widehat{x}_2^T - \widehat{x}_2 x^T + \widehat{x}_2\widehat{x}_2^T\right\rangle\right] \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} - \mathrm{tr}(K_1K_1^{-1}) + K_2 : \left(K_1^{-1} + \widehat{x}_1\widehat{x}_1^T - \widehat{x}_1\widehat{x}_2^T - \widehat{x}_2\widehat{x}_1^T + \widehat{x}_2\widehat{x}_2^T\right)\right] \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} - N + K_2 : K_1^{-1} + K_2 : (\widehat{x}_1-\widehat{x}_2)(\widehat{x}_1-\widehat{x}_2)^T\right] \\
&= \frac{1}{2}\left[-\ln\frac{|K_2|}{|K_1|} - N + K_2 : K_1^{-1} + (\widehat{x}_1-\widehat{x}_2)\cdot K_2(\widehat{x}_1-\widehat{x}_2))\right] \\
&= D^{\dagger}(K_1,K_2) + \frac{1}{2}(\widehat{x}_1-\widehat{x}_2)\cdot K_2(\widehat{x}_1-\widehat{x}_2) \qquad (9)
\end{aligned}
$$

An alternative form for $D^{\dagger}(K_1,K_2)$ is

$$
D^{\dagger}(K_1,K_2) = \frac{1}{2}\left[K_2 : K_1^{-1} - \ln\frac{|K_2|}{|K_1|} - \mathrm{I}:\mathrm{I}\right]. \qquad (10)
$$

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

1. You can find here (`http://lcvmwww.epfl.ch/teaching/modelling_dna/protected_files/codes_exercises/KL_Div.m`) 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`. To compute the Kullback-Leibler divergence per degree of freedom divide the computed value by number of degree of freedoms (total number of internal coordinates).

2. In the following table we reported the result of the Kullback-Leibler divergence computations per degree of freedom:

	KLD	stiffness part	mean part
$D(\rho_{band}(S),\rho_{obs}(S))$	0.006845	0.006845	0
$D(\rho_{cgDNAp}(S,\mathcal{P}),\rho_{band}(S))$	0.026043	0.024309	0.001734

4 Jensen's inequality - Part 1

For $u:\Omega\subset\mathbb{R}^n\to\mathbb{R}$, $\mathbb{R}\supset R(u)=\{y : y=u(x),\ x\subset\Omega\subset\mathbb{R}^n\}$ and for simplicity assume u is $C(\Omega)$. Note that $R(u)$ is range of u with an interval (c,d) where $c=-\infty,\ d=+\infty$ is possible. Then

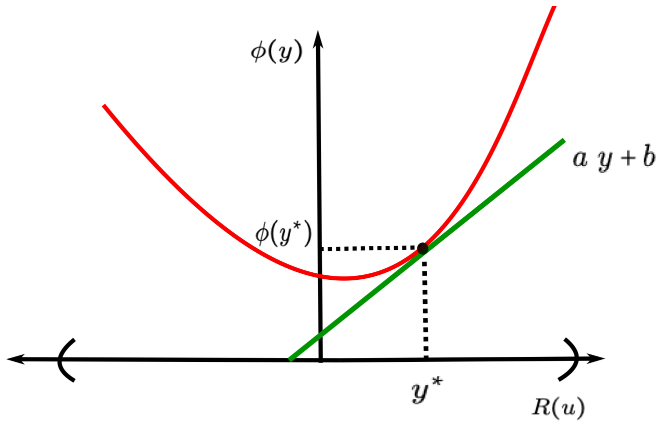

Figure 1: Plot of ϕ vs. y^* and line $a\ y + b$.

convexity of $\phi : R(u) \to \mathbb{R}$ means $\forall\ y\ \&\ y^* \subset R(u)\ \exists\ a, b, \in \mathbb{R}$ such that (see Fig. 1)

$$\phi(y) \geq ay + b \quad \& \quad \phi(y^*) = ay^* + b. \tag{11}$$

Now $u(x) \in R(u) \forall x \in \Omega$ so integration of eq. 11 w.r.t $\mu(x)$ & the choice (note $y^* \in R(u)$ because $R(u)$ is a single connected interval and so convex, because Ω is connected and u is contineous)

$$y^* := \frac{\int_\Omega u(x)\ d\mu(x)}{|\Omega|}, \quad |\Omega| := \int_\Omega d\mu(x)$$

implies

$$\int_\Omega \phi(u(x))\ d\mu(x) \geq a \int_\Omega u(x)\ d\mu(x) + b|\Omega| = |\Omega|\ (ay^* + b) = |\Omega|\ \phi\left(\frac{\int_\Omega u(x)\ d\mu(x)}{|\Omega|}\right),$$

as desired.