1 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

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix}\begin{pmatrix} \alpha & \beta \\ \beta^T & \gamma \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}, \tag{1}$$

which is equivalent to the system

$$\begin{cases} \mathbf{A}\alpha & + & \mathbf{B}\beta^{\mathbf{T}} & = & \mathbf{I} \\ \mathbf{A}\beta & + & \mathbf{B}\gamma & = & \mathbf{0} \\ \mathbf{B}^T\alpha & + & \mathbf{C}\beta^{\mathbf{T}} & = & \mathbf{0} \\ \mathbf{B}^T\beta & + & \mathbf{C}\gamma & = & \mathbf{I}, \end{cases} \tag{2}$$

and

$$\begin{cases} \beta^T = -\mathbf{C}^{-1}\mathbf{B}^T\alpha \\ \gamma = \mathbf{C}^{-1} - \mathbf{C}^{-1}\mathbf{B}^T\beta \\ \mathbf{A}\alpha - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T\alpha = \mathbf{I} \\ \mathbf{A}\beta + \mathbf{B}\mathbf{C}^{-1} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T\beta = \mathbf{0}, \end{cases} \tag{3}$$

which gives

$$\begin{aligned} \alpha &= (\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}, \\ \beta &= -(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{C}^{-1}, \\ \gamma &= \mathbf{C}^{-1} + \mathbf{C}^{-1}\mathbf{B}^T(\mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T)^{-1}\mathbf{B}\mathbf{C}^{-1}. \end{aligned} \tag{4}$$

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

$$\Pi = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \tag{5}$$

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

$$\det\begin{pmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} = \det\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} = \det \mathbf{A} \det \mathbf{D}. \tag{6}$$

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

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{B} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}\begin{pmatrix} \mathbf{A} - \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T & \mathbf{0} \\ \mathbf{C}^{-1}\mathbf{B}^T & \mathbf{I} \end{pmatrix}. \tag{7}$$

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

$$\det\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{C} \end{pmatrix} = \det\mathbf{C}\,\det[\mathbf{A}-\mathbf{B}\mathbf{C}^{-1}\mathbf{B}^T] = \det\mathbf{C}\,\det[\alpha^{-1}]. \tag{8}$$

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

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

From (4) we remark, that $K_{11}-K_{12}\,K_{22}^{-1}\,K_{12}^T = \Sigma_{11}^{-1}$ and from (8), that $\frac{1}{\det[K^{-1}]} = \det K = \det K_{22}\det[\Sigma_{11}^{-1}]$. We finally get

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

2 Entropy and Relative entropy formulas for Gaussians

Remarks useful for both computations:

1. if $a \in \mathbb{R}$, then $\mathrm{tr}(a) = a$.
2. the trace is invariant under cycling permutation, i.e, $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$. This property is
3. Let $A, B \in \mathbb{R}^{n\times n}$, we define the following matrix inner product: $A : B = \sum_{i,j=1}^n a_{ij}b_{ij} = \mathrm{tr}(B^T A) = \mathrm{tr}(A^T B)$.
4. Consequence of the previous remarks:

$$(x-\widehat{x})^T K(x-\widehat{x}) = \mathrm{tr}\left((x-\widehat{x})^T K(x-\widehat{x})\right) = \mathrm{tr}\left(K(x-\widehat{x})(x-\widehat{x})^T\right) = K : (x-\widehat{x})(x-\widehat{x})^T, \tag{9}$$

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

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 h:

$$\begin{aligned}
h(q) &= -\int_{\mathbb{R}^N} p(x)\ln(p(x))dx \\
&= \int_{\mathbb{R}^N}\left(\frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right)+\frac{1}{2}(x-\widehat{x}_1)\cdot K_1(x-\widehat{x}_1)\right)p(x)dx \\
&= \frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right)+\frac{1}{2}\int_{\mathbb{R}^N}(x-\widehat{x}_1)\cdot K_1(x-\widehat{x}_1)p(x)dx \\
&= \frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right)+\frac{1}{2}\left\langle(x-\widehat{x}_1)^T K_1(x-\widehat{x}_1)\right\rangle_p \qquad (10)
\end{aligned}$$

Let us study the second term in (10):

$$\begin{aligned}
\left\langle(x-\widehat{x}_1)^T K_1(x-\widehat{x}_1)\right\rangle_p &= \left\langle K_1 : (x-\widehat{x}_1)(x-\widehat{x}_1)^T\right\rangle_p \qquad (11)\\
&= K_1 : \left\langle (x-\widehat{x}_1)(x-\widehat{x}_1)^T\right\rangle_p \qquad (12)\\
&= K_1 : K_1^{-1} = \mathrm{tr}\left(K_1K_1^{-1}\right) = N. \qquad (13)
\end{aligned}$$

We can now conclude that

$$\begin{aligned}
h(q) &= \frac{1}{2}\ln\left((2\pi)^N|K^{-1}|\right)+\frac{1}{2}N \\
&= \frac{1}{2}\ln\left((2\pi e)^N|K^{-1}|\right),
\end{aligned} \qquad (14)$$

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

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

$$
\begin{aligned}
D(p,q) = \left\langle \ln \frac{p}{q} \right\rangle_p & \\
&= \frac{1}{2}\left\langle -\ln|K_1^{-1}| - (x-\widehat{x}_1)\cdot K_1(x-\widehat{x}_1) + \ln|K_2^{-1}| + (x-\widehat{x}_2)\cdot K_2(x-\widehat{x}_2)\right\rangle \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} + \langle -(x-\widehat{x}_1)\cdot K_1(x-\widehat{x}_1)\rangle + \langle (x-\widehat{x}_2)\cdot K_2(x-\widehat{x}_2)\rangle\right] \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} - K_1 : \left\langle (x-\widehat{x}_1)(x-\widehat{x}_1)^T\right\rangle + K_2 : \left\langle xx^T - x\widehat{x}_2^T - \widehat{x}_2 x^T + \widehat{x}_2\widehat{x}_2^T\right\rangle\right] \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} - \mathrm{tr}(K_1K_1^{-1}) + K_2 : \left(K_1^{-1} + \widehat{x}_1\widehat{x}_1^T - \widehat{x}_1\widehat{x}_2^T - \widehat{x}_2\widehat{x}_1^T + \widehat{x}_2\widehat{x}_2^T)\right)\right] \\
&= \frac{1}{2}\left[\ln\frac{|K_2^{-1}|}{|K_1^{-1}|} - N + K_2 : K_1^{-1} + K_2 : (\widehat{x}_1-\widehat{x}_2)(\widehat{x}_1-\widehat{x}_2)^T\right] \\
&= \frac{1}{2}\left[-\ln\frac{|K_2|}{|K_1|} - N + K_2 : K_1^{-1} + (\widehat{x}_1-\widehat{x}_2)\cdot K_2(\widehat{x}_1-\widehat{x}_2))\right] \\
&= D^{\dagger}(p,q) + \frac{1}{2}(\widehat{x}_1-\widehat{x}_2)\cdot K_2(\widehat{x}_1-\widehat{x}_2) \qquad (15)
\end{aligned}
$$

An alternative form for $D^{\dagger}(p,q)$ is

$$
D^{\dagger}(p,q) = \frac{1}{2}\left[K_2 : K_1^{-1} - \ln\frac{|K_2|}{|K_1|} - \mathrm{I} : \mathrm{I}\right]. \qquad (16)
$$

3 Jensen's inequality

Let $\Omega \subset \mathbb{R}^n$ be a bounded open set, $u \in L^1(\Omega)$, with range $R(u(\Omega))$ and $\phi : R(u(\Omega)) \to \mathbb{R}$ be convex. Then $\forall y^* \in R(u(\Omega)) \exists a, b, \in \mathbb{R}$ such that $\phi(y) \geq ay + b$ with $\phi(y^*) = ay^* + b$.

Ansatz: Let $y^* = \dfrac{\int_\Omega u dx}{|\Omega|}$.

Thus we have that

$$
\int_\Omega \phi(u)dx \geq \int_\Omega [au+b]dx = a\int_\Omega udx + |\Omega|b = |\Omega|\left(a\frac{\int_\Omega udx}{|\Omega|} + b\right).
$$

Finally we have that

$$
\frac{1}{|\Omega|}\int_\Omega \phi(u)dx \geq a\frac{\int_\Omega udx}{|\Omega|} + b = ay^* + b = \phi(y^*).
$$