Proof of proposition 2 (section 4.1) : best inner product related to a kernelized empirical distribution

This is the fully detailed version of proposition 2, which could not fit in the article for length constraint reasons.

[...]

One might however wonder whether minimizing the Kullback-Leibler distance to a sum of Dirac peaks makes sense. Luckily, the previous proposition can be extended to the case of symmetric translation-invariant unit-mass kernels $\mathcal{K}(\cdot - \cdot)$ defined on the space T of deformations. We replace $\overrightarrow{\mathcal{D}}_{emp}$ by the kernel-smoothed empirical distribution

$$\overrightarrow{\mathcal{D}}^{\mathcal{K}}_{emp}(\mathbf{f}) \;=\; \sum_j w_j\, \mathcal{K}(\mathbf{f}_j - \mathbf{f}).$$

Note : The family $(\mathbf{f}_j)$ is finite, so we work in a finite-dimensioned subspace of the tangent space T, and $\mathcal{K}$ can be understood, in the simple case, as a real function multiplied by the usual Lebesgue measure $d\mathbf{f}$. In the infinite dimension case, $\mathcal{K}$ cannot be isotropic (because it has finite mass).

Proposition 2. *The inner product P which leads to the probability distribution $\overrightarrow{\mathcal{D}}_P$ the closest to the empirical distribution $\overrightarrow{\mathcal{D}}^{\mathcal{K}}_{emp} = \sum_j w_j \mathcal{K}(\mathbf{f}_j - \cdot)$ for the Kullback-Leibler divergence, is obtained by diagonalization of the sum of the correlation matrix F and the second moment of $\mathcal{K}$.*

Proof. The energy E becomes:

$$\begin{aligned}
E(\overrightarrow{\mathcal{D}}_P | \overrightarrow{\mathcal{D}}^{\mathcal{K}}_{emp}) &= -\sum_j w_j \int_{\mathbf{f}\in T} \ln\big[\overrightarrow{\mathcal{D}}_P(\mathbf{f}_j + \mathbf{f})\big]\, \mathcal{K}(\mathbf{f}) \\
&= \sum_j w_j \int_T \left(\|\mathbf{f}_j + \mathbf{f}\|^2_P - \frac{1}{2}\sum_n \ln \alpha_n \right) \mathcal{K}(\mathbf{f}) \;+ c^{st} \\
&= \sum_j w_j \|\mathbf{f}_j\|^2_P \;+\; \mu^2_{\mathcal{K},P} - \frac{1}{2}\sum_n \ln \alpha_n \quad + c^{st}
\end{aligned}$$

which is the same expression as previously (equation **2**) up to the second scalar moment of $\mathcal{K}$, $\mu^2_{\mathcal{K},P} = \int_T \|\mathbf{f}\|^2_P \mathcal{K}(\mathbf{f})$, which depends on P. This moment can be rewritten as:

$$\mu^2_{\mathcal{K},P} \quad = \quad \sum_n \alpha_n \int_T \langle\, \mathbf{f} \,|\, \mathbf{e}_n \,\rangle^2_{P_0} \mathcal{K}(\mathbf{f}) \quad = \quad \sum_n \alpha_n \, \mathbf{e}_n \, M_{\mathcal{K}} \, \mathbf{e}_n$$

where $M_{\mathcal{K}}$ is the second moment matrix of $\mathcal{K}$ for P_0:

$$M_{\mathcal{K}} = \int_T \mathbf{f} \otimes \mathbf{f} \;\; \mathcal{K}(\mathbf{f}).$$

Thus, up to a constant, the energy rewrites:

$$E = \sum_n \left(\alpha_n \, \langle\, \mathbf{e}_n \,| F + M_{\mathcal{K}} \,|\, \mathbf{e}_n \,\rangle_{P_0} - \frac{1}{2} \ln \alpha_n \right)$$

which is similar to equation (**2**) except that F is replaced by $F + M_{\mathcal{K}}$. The proof ends with the same reasoning as previously. Note that when the kernel $\mathcal{K}$ gets closer to a Dirac peak, $M_{\mathcal{K}}$ gets closer to 0, and we obtain proposition **1** again. □