The purpose of this additional sheet is to provide more practice and exam preparation material. N.B. The tutors are not required to work through this material in the tutorial.

Exercise 1. ***Score matching for the exponential family***

In the lecture, we have derived the objective function $J(\boldsymbol{\theta})$ for score matching,

$$J(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\left[\partial_j \psi_j(\mathbf{x}_i;\boldsymbol{\theta}) + \frac{1}{2}\psi_j(\mathbf{x}_i;\boldsymbol{\theta})^2\right], \tag{1}$$

where ψ_j is the partial derivative of the log model-pdf $\log p(\mathbf{x};\boldsymbol{\theta})$ with respect to the j-th coordinate (slope) and $\partial_j\psi_j$ its second partial derivative (curvature). The observed data are denoted by $\mathbf{x}_1,\ldots,\mathbf{x}_n$ and $\mathbf{x}\in\mathbb{R}^m$.

The goal of this exercise is to show that for statistical models of the form

$$\log p(\mathbf{x};\boldsymbol{\theta}) = \sum_{k=1}^{K}\theta_k F_k(\mathbf{x}) - \log Z(\boldsymbol{\theta}), \qquad \mathbf{x}\in\mathbb{R}^m, \tag{2}$$

the score matching objective function becomes a quadratic form, which can be optimised efficiently (see e.g. Barber Appendix A.5.3).

The set of models above are called the (continuous) exponential family, or also log-linear models because the models are linear in the parameters θ_k. Since the exponential family generally includes probability mass functions as well, the qualifier "continuous" may be used to highlight that we are here considering continuous random variables only. The functions $F_k(\mathbf{x})$ are assumed to be known; they are the sufficient statistics (see e.g. Barber Section 8.5).

(a) Denote by $\mathbf{K}(\mathbf{x})$ the matrix with elements $K_{kj}(\mathbf{x})$,

$$K_{kj}(\mathbf{x}) = \frac{\partial F_k(\mathbf{x})}{\partial x_j}, \qquad k=1\ldots K, \quad j=1\ldots m, \tag{3}$$

and by $\mathbf{H}(\mathbf{x})$ the matrix with elements $H_{kj}(\mathbf{x})$,

$$H_{kj}(\mathbf{x}) = \frac{\partial^2 F_k(\mathbf{x})}{\partial x_j^2}, \qquad k=1\ldots K, \quad j=1\ldots m. \tag{4}$$

Furthermore, let $\mathbf{h}_j(\mathbf{x}) = (H_{1j}(\mathbf{x}),\ldots,H_{Kj}(\mathbf{x}))^\top$ be the j–th column vector of $\mathbf{H}(\mathbf{x})$.

Show that for the continuous exponential family, the score matching objective in Equation (1) becomes

$$J(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top\mathbf{r} + \frac{1}{2}\boldsymbol{\theta}^\top\mathbf{M}\boldsymbol{\theta}, \tag{5}$$

where

$$\mathbf{r} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\mathbf{h}_j(\mathbf{x}_i), \qquad \mathbf{M} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{K}(\mathbf{x}_i)\mathbf{K}(\mathbf{x}_i)^\top. \tag{6}$$

(b) The pdf of a zero mean Gaussian parametrised by the variance σ^2 is

$$p(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right), \qquad x \in \mathbb{R}. \tag{7}$$

The (multivariate) Gaussian is a member of the exponential family. By comparison with Equation (2), we can re-parametrise the statistical model $\{p(x; \sigma^2)\}_{\sigma^2}$ and work with

$$p(x; \theta) = \frac{1}{Z(\theta)} \exp\left(\theta x^2\right), \qquad \theta < 0, \qquad x \in \mathbb{R}, \tag{8}$$

instead. The two parametrisations are related by $\theta = -1/(2\sigma^2)$. Using the previous result on the (continuous) exponential family, determine the score matching estimate $\hat{\theta}$, and show that the corresponding $\hat{\sigma}^2$ is the same as the maximum likelihood estimate. This result is noteworthy because unlike in maximum likelihood estimation, score matching does not need the partition function $Z(\theta)$ for the estimation.