OT Homework 3

Due December 3rd

This homework studies two notions of projection on the simplex[1], which is the set of probability vectors:

$$\Delta_n = \{x \in \mathbb{R}^d, x \geq 0, \sum_{i=1}^{n} x_i = 1\} \tag{0.1}$$

The natural (Euclidean) projection onto the simplex is defined as:

$$\Pi(y) = \operatorname*{argmin}_{x \in \Delta_d} \frac{1}{2}\|x - y\|^2 \tag{0.2}$$

Q1. In dimension 2, do a plot of the simplex showing what $\Pi(y)$ is for $y = (3, 4)$. If you were given the vector $(3, 4)$ and asked to transform it into a probability vector, would your answer be the value given by $\Pi(y)$? Would you suggest something else?

In the sequel, we'll define a better notion of projection onto the simplex. It is based on three very useful objects in optimization and statistics: the entropy, Bregman divergences, and the Kullback-Leibler divergence.

Definition 0.1 (Negative entropy)**.** *The following function is called the* negative entropy*:*

$$\begin{aligned} H : \mathbb{R}^d_+ &\to \mathbb{R} \\ x &\mapsto \sum_1^d x_i \log x_i \end{aligned} \tag{0.3}$$

(with the convention $0 \log 0 = 0$*)*

Q2. Remember that a twice differentiable function is convex provided its Hessian is positive definite (i.e. has positive eigenvalues). Show that the Hessian of the negative entropy is diagonal. Show that the negative entropy is convex.

Definition 0.2 (Bregman divergence)**.** *For a convex differentiable function* J*, the* Bregman divergence *associated to* J *is:*

$$D : x, y \mapsto J(x) - J(y) - \langle \nabla J(y), x - y \rangle \tag{0.4}$$

Q3. Show that for a fixed y, $D(\cdot, y)$ is convex.

Q4. Compute the Bregman divergence associated to $J = \frac{1}{2}\|\cdot\|^2$.

Q5. Show that Bregman divergence associated to H is the *Kullback-Leibler divergence*[2]:

$$\mathrm{KL}(x, y) = \sum_{i=1}^d x_i \log \frac{x_i}{y_i} - \sum_{i=1}^d x_i + \sum_{i=1}^d y_i \tag{0.5}$$

[1] aucun lien, fils unique

[2] if there exists i such that $x_i \neq 0$ and $y_i = 0$, the KL is $+\infty$; ignore this technicality

Note: the KL divergence is very often considered for vectors x and y on the simplex, in which case the last two sums cancel.

For $y \in \mathbb{R}^d_{++}$ (the case where y has vanishing entries is easy to handle similarly), we take the convex objective function in (0.2) and replace it by the convex (in x) $\mathrm{KL}(x, y)$. We thus want to solve:

$$\underset{x \in \Delta_d}{\operatorname{argmin}} \sum_{i=1}^{d} x_i \log \frac{x_i}{y_i} - \sum_{i=1}^{d} x_i + \sum_{i=1}^{d} y_i \tag{0.6}$$

Q6. Are the KKT sufficient for this optimization problem? Write its Lagrangian.

Q7. Show that the KKT conditions can be written:

$$\begin{cases} x_i = y_i \exp(\lambda - \mu_i) \\ \sum_{i=1}^{d} x_i = 1 \\ \mu_i x_i = 0 \end{cases} \tag{0.7}$$

Q8. Show that for a KKT point (x, λ, μ) we must have $1 = \exp(\lambda) \sum_{i=1}^{d} y_i$.

Q9. Show that a solution of (0.6) is given by:

$$\left(\frac{y_i}{\sum_{i=1}^{d} y_i} \right)_{i \in [d]} \tag{0.8}$$