1. Given a strongly convex function $\Phi(\boldsymbol{x})$, the *Bregman-Divergence* associated to $\Phi(\cdot)$ is defined via

$$D_\Phi(\boldsymbol{x}\|\boldsymbol{y}) = \Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y}) - \langle \nabla\Phi(\boldsymbol{y}), \boldsymbol{x}-\boldsymbol{y}\rangle.$$

Given a convex function $\Phi(\cdot)$, and hence the associated Bregman divergence, we can define a projection operation onto a convex set $\mathcal{X}$, with respect to this Bregman divergence:

$$\Pi_{\mathcal{X}}^{\Phi}(\boldsymbol{y}) = \arg\min : D_\Phi(\boldsymbol{x}\|\boldsymbol{y}), \quad \text{s.t.} \quad \boldsymbol{x}\in\mathcal{X}.$$

For $\Phi(\boldsymbol{x})$ given by

$$\Phi(\boldsymbol{x}) = \sum x_i \log x_i.$$

show that the projection onto the simplex

$$\Delta_n = \{\boldsymbol{x}\in\mathbb{R}^n : \sum x_i = 1,\, x_i \geq 0\}.$$

is given by $L1$ renormalization:

$$\boldsymbol{y} \mapsto \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|_1}.$$

Xueying Lu
xl4888

Pf. For $\phi(x) = \sum x_i \log x_i$. $\quad y \geq 0$.

$$D_\phi(x\|y) = \sum x_i \log \frac{x_i}{y_i}.$$

$$\min_x \sum x_i \log \frac{x_i}{y_i}$$

$$\text{s.t. } \sum x_i = 1.$$

$$x_i \geq 0.$$

The Lagrangian is

$$L(x,\lambda,\mu) = \sum x_i \log\frac{x_i}{y_i} - \lambda^T x + \mu\left(\sum x_i - 1\right), \quad \lambda \geq 0$$

By the KKT condition.

$$\begin{cases} \log \frac{x_i}{y_i} + 1 - \lambda_i + \mu = 0. \quad \forall i. \\ x \geq 0. \ \lambda \geq 0. \\ \lambda_i x_i = 0. \quad \forall i. \\ \sum x_i - 1 = 0 \end{cases}$$

$\Rightarrow$ If $x_i \neq 0$. then $\lambda_i = 0$. $x_i = e^{-1-\mu} y_i$.

$\Rightarrow e^{-1-\mu}\left(\sum y_i\right) = 1.$

$\Rightarrow x_i = \dfrac{y_i}{\|y_i\|_1}$ $\square$

2. For $\Phi$ strongly convex and twice differentiable, and for the Bregman divergence defined as above, show that:

$$D_\Phi(\boldsymbol{x}\|\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_{\nabla^2\Phi(\boldsymbol{z})},$$

for some $\boldsymbol{z} \in [\boldsymbol{x}, \boldsymbol{y}]$, i.e., for some $\boldsymbol{z}$ in the convex combination of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Recall that for a positive definite matrix $M$, the Euclidean norm with respect to $M$ is given by

$$\|\boldsymbol{x}\|_M^2 = \boldsymbol{x}^\top M \boldsymbol{x}.$$

Pf. $D\phi(x\|y) = \Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x-y \rangle.$

$\exists\, z \in [x, y]$ s.t.

$\Phi(x) = \Phi(y) + \langle \nabla\Phi(y), x-y \rangle + \frac{1}{2}\langle (x-y)^\top \nabla^2\Phi(z), x-y \rangle.$

$\Rightarrow D\phi(x\|y) = \frac{1}{2}\langle (x-y)^\top \nabla^2\Phi(z), x-y \rangle = \frac{1}{2}\|x-y\|^2_{\nabla^2\Phi(z)}$

$\square$