1.)

$$\Pi_{\mathcal{X}}^{\Phi}(y) = \arg\min_{x} \left(\max_{\substack{\lambda_1, \lambda_2 \\ \in\mathbb{R} \ \leq 0}} \left(\Phi(x) - \Phi(y) - \langle \nabla\Phi(y), x-y \rangle + \lambda_1 \left(\sum_i x_i - 1 \right) + \lambda_2^T x \right) \right)$$

$$= \arg\min_{x} \ f(x, \lambda_1^*, \lambda_2^*)$$

$$f(x, \lambda_1^*, \lambda_2^*) = \sum_i \left[x_i \log(x_i) - y_i \log(y_i) - (\log(y_i)+1)(x_i - y_i) \right] + \lambda_1^*(\cdots) + \lambda_2^{*T} x$$

$$= \sum_i \left[x_i \left(\log(x_i) - [\log(y_i) + 1] \right) - \cancel{y_i \log(y_i)} + \cancel{y_i \log(y_i)} + y_i \right] + (\cdots)$$

$$= \sum_i \left[x_i \left[\log(x_i) - \log(y_i) \right] + y_i - x_i + \lambda_1^* x_i + \lambda_2^*[i]\, x_i \right] - \lambda_1^*$$

By KKT

- Stationarity: $\nabla_x f(x, \lambda_1^*, \lambda_2^*)[i] = \log(x_i) + \cancel{1} - \log(y_i) - \cancel{1} + \lambda_1^* + \lambda_2^*[i] \overset{set}{=} \emptyset$

$$\Rightarrow \log(\tilde{x}_i) = \log(y_i) - \lambda_1^* - \lambda_2^*[i]$$

$$\tilde{x}_i = y_i \cdot e^{-\lambda_1^* - \lambda_2^*[i]}$$

Primal Feasibility:

- $\tilde{x}_i \geq 0$, $\sum_i \tilde{x}_i = 1$

Dual Feasib:

- $\lambda_1^* \in \mathbb{R}$, $\lambda_2^* \geq 0$

Comp. Slack:

- $\lambda_2^*[i] \cdot \tilde{x}_i = \emptyset$

$$\Rightarrow \begin{cases} \lambda_2^*[i] > 0 \text{ iff } \tilde{x}_i = 0 \\ \lambda_2^*[i] = 0 \text{ iff } \tilde{x}_i > 0 \end{cases}$$

- by comp. slackness, we have that:

$\tilde{x}_i > 0 \Rightarrow \lambda_2^*[i] = 0$:

thus $\tilde{x}_i = y_i\, e^{-\lambda_1^*} > 0$ must hold $\forall$ y

this can only be achieved if domain of $\Phi(\cdot)$ is fixed to either $\mathbb{D} = \mathbb{R}_+$ or $\mathbb{D} = \mathbb{R}_-$. assume $\mathbb{D} = \mathbb{R}_+$. Thus, $e^{-\lambda_1^*} > 0$.

- Additionally, by primal feasib we have:

$$\sum_i \tilde{x}_i = \sum_{I_1 = \{i \mid \tilde{x}_i > 0\}} y_i\, e^{-\lambda_1^*} + \sum_{I_2 = \{j \mid \tilde{x}_j = 0\}} \tilde{x}_j^{\nearrow 0} = 1$$

$$\boxed{\therefore\ e^{-\lambda_1^*} = 1 / \sum_{i \in I_1} y_i = 1/\|y\|_1}$$

- Moreover, this choice of λ_1^* is unique. Choice of λ_2^* does not affect the projection's form, as when it has freedom $\tilde{x}_i = \emptyset$.

$$\Rightarrow \boxed{\Pi^{\phi}_{X \cap D}(y) = \frac{y}{\|y\|_1}}$$

(Note: we need restriction of domain of $\phi(\cdot)$ for this result)

2.) By Taylor's Thm, it follows that:

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{(y-x)^T \nabla^2 f(z)(y-x)}{2}$$

for some $z \in \text{conv}(x, y)$

- which implies:

$$\boxed{D(y \| x) = \frac{\|x-y\|_{\nabla^2 \phi(z)}}{2}} \quad \text{for some } z \in \text{conv}(x,y)$$

- the problem as written is incorrect

i.e. $D(y \| x) \neq \|x-y\|_{\nabla^2 \phi(z)}$ for some $z \in \text{conv}(x,y)$

e.g: $f(y) - f(x) + \nabla f(x)(y-x) \neq (y-x)^2 \nabla^2 f(z)$

for any z when $f(z) = z^2$, $z \in \mathbb{R}$, $x \neq y$

- The proof of Taylor's theorem for multi-variate functions is similar to the univariate case, but just more tedious.

- The proof uses Mean-Value Theorem inductively to get the general result.

- E.g: $f(y) = f(x) + \nabla f(z)^T (y-x)$ for some $z \in \text{conv}(x,y)$

Proof: let $g(\lambda) = f(x + \lambda(y-x))$

by MVThm: $g(1) = g(0) + \nabla g(\lambda^*)$ for some $\lambda^* \in [0,1]$

$\Rightarrow g(1) = f(y)$ $\qquad \nabla g(\lambda^*) = \nabla f(x + \lambda^*(y-x))(y-x)$

$g(0) = f(x)$

$$\Rightarrow f(y) = f(x) + \nabla f(z)(y-x)$$

$\square$