2. $\sup_{\theta,\phi} g(\theta,\phi) = \sup_{\theta}\left(\sup_{\phi} g(\theta,\phi)\right)$

$= \sup_{\theta}\left(\sup_{\phi}\left(f(\theta) - h(\theta,\phi)\right)\right)$

$= \sup_{\theta} f(\theta) \qquad (\because\ \forall\theta\ \exists\phi\ \text{s.t. } h(\theta,\phi)=0, \text{ which is its minimum})$

$\theta_0 \in \operatorname{argmax}_{\theta} f \quad \Rightarrow \quad f(\theta_0) = \sup_{\theta} f(\theta)$

$\Rightarrow \quad g(\theta_0,\phi_0) = f(\theta_0) = \sup_{\theta} f(\theta) = \sup_{\theta,\phi} g(\theta,\phi)$

for ϕ_0 s.t. $h(\theta_0,\phi_0) = 0$

$\Rightarrow \quad (\theta_0, \phi_0) \in \operatorname{argmax}_{(\theta,\phi)} g$

$(\theta_0, \phi_0) \in \operatorname{argmax}_{(\theta,\phi)} g \quad \Rightarrow \quad g(\theta_0,\phi_0) \le g(\theta_0,\phi_1) = f(\theta_0)$

for ϕ_1 s.t. $h(\theta_0,\phi_1) = 0$.

That is, $\phi_0 = \phi_1$ and $h(\theta_0,\phi_0) = 0$.

$\Rightarrow \quad f(\theta_0) = g(\theta_0,\phi_0) = \sup_{\theta,\phi} g(\theta,\phi) = \sup_{\theta} f(\theta)$

$\Rightarrow \quad \theta_0 \in \operatorname{argmax}_{\theta} f$

$\square$

3.

```
Curr = i
next = σ(i)

While ( next == i )
  {  Curr = next
     next = σ(next) }

return  Curr
```

$\square$

4.

(a) $P_\sigma^T = [e_{\sigma(1)} \quad e_{\sigma(2)} \quad \cdots \quad \sigma_{\sigma(n)}]$

$P_\sigma^T P_\sigma = e_{\sigma(1)} e_{\sigma(1)}^T + \cdots + e_{\sigma(n)} e_{\sigma(n)}^T$

$= \left(I\,(i,j) = (\sigma(1),\sigma(1))\right)_{\substack{1\le i\le n \\ 1\le j\le n}} + \cdots + \left(I\,(i,j) = (\sigma(n),\sigma(n))\right)_{\substack{1\le i\le n \\ 1\le j\le n}}$

$= I$

$= \left(e^T_{\sigma(i)}\, e_{\sigma(j)} \right)_{\substack{1\le i\le n \\ 1\le j\le n}}$

$= P_\sigma P_\sigma^T$

$\therefore \quad P_\sigma^{-1} = P_\sigma^T$

$$P_{\sigma^{-1}} = \begin{bmatrix} e_{\sigma^{-1}(1)}^T \\ \vdots \\ e_{\sigma^{-1}(n)}^T \end{bmatrix} = \left(I\left(\sigma^{-1}(i) = j \right) \right)_{\substack{1\le i\le n \\ 1\le j\le n}}$$

$$= \left(I\left(\sigma(j) = i \right) \right)_{\substack{1\le i\le n \\ 1\le j\le n}}$$

$$= \left(I\left(\sigma(i) = j \right) \right)^T_{\substack{1\le i\le n \\ 1\le j\le n}}$$

$$= P_\sigma^T$$

(b) $\det P_\sigma = \det P_\sigma^T \underset{(a)}{=} \det P_\sigma^{-1}$

$$(\det P_\sigma)^2 = (\det P_\sigma)(\det P_\sigma^{-1}) = \det P_\sigma P_\sigma^{-1} = \det I = 1$$

$$\therefore \ |\det P_\sigma| = 1$$

$\square$

5. $\dfrac{\partial z}{\partial \lambda} = \left(\dfrac{\partial z_i}{\partial \lambda_j} \right)_{\substack{1\le i\le n \\ 1\le j\le n}}$.

$$\frac{\partial z_i}{\partial \lambda_j} = \begin{cases} 1 & i\in\Omega,\ j\in\Omega \\ 0 & i\in\Omega,\ j\in\Omega^c \\ \text{Something} & i\in\Omega^c,\ j\in\Omega \\ e^{s_\theta(\lambda_\Omega)_j} & i\in\Omega^c,\ j\in\Omega^c \end{cases}$$

Consider $\sigma \in S_n$: increasing s.t. $\sigma^{-1}(\Omega) = \{1, \cdots, |\Omega|\}$

$\Leftrightarrow \sigma^{-1}(\Omega^c) = \{ |\Omega|+1, \cdots, n \}$

$$\underbrace{\begin{bmatrix} e_{\sigma(1)}^T \\ \vdots \\ e_{\sigma(n)}^T \end{bmatrix}}_{P_\sigma} \frac{\partial z}{\partial \lambda} \underbrace{[e_{\sigma(1)} \cdots e_{\sigma(n)}]}_{P_\sigma^T} = \begin{bmatrix} I & 0 \\ * & \operatorname{diag} e^{s_\theta(\lambda_\Omega)} \end{bmatrix}$$

$$\frac{dz}{d\lambda} = P_\sigma^{-1} \begin{bmatrix} I & 0 \\ * & \operatorname{diag} e^{s_\theta(\lambda_\Omega)} \end{bmatrix} \underbrace{(P_\sigma^T)^{-1}}_{P_\sigma}$$

$$\left|\frac{dz}{d\lambda}\right| = |P_\sigma^{-1}| \begin{vmatrix} I & 0 \\ * & \operatorname{diag} e^{s_\theta(\lambda_\Omega)} \end{vmatrix} |P_\sigma| = 1 \cdot |I| \left|\operatorname{diag} e^{s_\theta(\lambda_\Omega)}\right| \cdot 1 = \left|\operatorname{diag} e^{s_\theta(\lambda_\Omega)}\right|$$

$$\log \left|\frac{dz}{d\lambda}\right| = \log \left|\operatorname{diag} e^{s_\theta(\lambda_\Omega)}\right| = 1_{(n-|\Omega|)}^T s_\theta(\lambda_\Omega)$$

$\square$

8.

$$\nabla_\phi \, E_{z \sim q_\phi}\left[\log \frac{h(z)}{q_\phi(z)}\right]$$

$$= \nabla_\phi \int_{\mathbb{R}^K} \left(\log \frac{h(z)}{q_\phi(z)}\right) \cdot q_\phi(z)\, dz$$

$$= \int_{\mathbb{R}^K} \nabla_\phi \left(\left(\log \frac{h(z)}{q_\phi(z)}\right) q_\phi(z)\right) dz$$

$$= \int_{\mathbb{R}^K} \frac{\cancel{q_\phi(z)}}{\cancel{h(z)}} \left(-\frac{\cancel{h(z)}\, \dot{q}_\phi(z)}{\cancel{(q_\phi(z))^2}}\right) \cancel{q_\phi(z)} + \left(\log \frac{h(z)}{q_\phi(z)}\right) \dot{q}_\phi(z) \quad dz$$

$$= \int_{\mathbb{R}^K} \dot{q}_\phi(z) \log \frac{h(z)}{q_\phi(z)}\, dz \quad - \underbrace{\int_{\mathbb{R}^K} \dot{q}_\phi(z)\, dz}_{= \nabla_\phi \int_{\mathbb{R}^K} q_\phi(z)\, dz \;=\; \nabla_\phi 1 = 0}$$

$$= \int_{\mathbb{R}^K} \frac{\dot{q}_\phi(z)}{q_\phi(z)} \left(\log \frac{h(z)}{q_\phi(z)}\right) \cdot q_\phi(z)\, dz$$

$$= E_{z \sim q_\phi}\left[\left(\nabla_\phi \log q_\phi(z)\right) \log\left(\frac{h(z)}{q_\phi(z)}\right)\right]$$

$\square$