$h(s) = \mathbb{E}_\pi[\bar{S}(Y;\bar{\theta}(s))] - s.$

Ass1: for any $s \in \mathcal{S}$, the function $s \mapsto \ell(s,\theta) = +\psi(\theta) - \langle s, \phi(\theta)\rangle + \lambda P(\theta)$ has a unique minimum at $\bar{\theta}(s)$, characterized by the first-order condition.

$$\nabla_\theta \psi(\bar{\theta}(s)) - J_\phi(\bar{\theta}(s))\, s + \lambda \nabla_\theta P(\bar{\theta}(s)) = 0.$$

Proposition: $h(s_*) = 0 \iff \nabla_\theta KL(\pi \| g_{\theta^*}) + \lambda \nabla_\theta P(\theta^*) = 0$ with $\theta^* = \bar{\theta}(s^*)$.

Let s^* be a root of h: $h(s^*) = 0$ and set $\theta^* = \bar{\theta}(s^*)$

$$(1) \quad +\nabla_\theta \psi(\bar{\theta}(s)) - J_\phi^\theta(\bar{\theta}(s))\, s + \nabla_\theta R(\bar{\theta}(s)) = 0 \quad \Big\}\ \text{characterization of the stationary point.}$$

where $J_\phi^\theta(\theta)$ is the Jacobian of the function $\vartheta \mapsto \phi(\vartheta)$ at $\theta \in \Theta$.

From the Fisher identity

$$\nabla_\theta \log g_\theta(y) = -\nabla_\theta \psi(\theta) + J_\phi^\theta(\theta)\, \bar{S}(y;\theta) \qquad \text{where } \bar{S}(y;\theta) = \mathbb{E}[S(Y,Z) \mid Y;\theta]$$

$$KL(\pi \| g_\theta) = \mathbb{E}_\pi\left[\log \frac{\pi(Y)}{g_\theta(Y)}\right] = \mathbb{E}_\pi[\log \pi(Y)] - \mathbb{E}_\pi[\log g_\theta(Y)]$$

which implies

$$\nabla_\theta KL(\pi \| g_\theta) = -\mathbb{E}_\pi[\nabla_\theta \log g_\theta(Y)]$$ (under assumptions allowing to switch diff and integration).

Therefore, taking the expectation w.r.t π (the law of the observation).

$$\nabla_\theta KL(\pi \| g_\theta) = +\nabla_\theta \psi(\theta) - J_\phi^\theta(\theta)\, \mathbb{E}_\pi[\bar{S}(Y;\theta)].$$

if $h(s^*) = 0 \Rightarrow s^* = \mathbb{E}_\pi[\bar{S}(Y;\bar{\theta}(s_*))]$

and $\nabla_\theta KL(\pi \| g_{\theta^*}) + \nabla_\theta R(\theta^*) = \nabla_\theta \psi(\theta^*) - J_\phi^\theta(\theta^*)\, s^* + \nabla_\theta R(\theta^*) = 0$

Conversely if $\nabla_\theta KL(\pi \| g_{\theta^*}) + \lambda \nabla_\theta R(\theta^*) = 0$ Set $s_* = \mathbb{E}_\pi[\bar{S}(Y;\theta^*)]$

$$\nabla_\theta \psi(\theta^*) - J_\phi^\theta(\theta^*)\, s_* + \nabla_\theta R(\theta^*) = 0$$

Since (1) characterizes the extremum: $\theta^* = \bar{\theta}(s^*)$.

Set $w(s) = KL(\pi \| g_{\bar{\theta}(s)}) + R(\bar{\theta}(s))$.

Recall that: $\nabla_\theta KL(\pi \| g_\theta) = +\nabla_\theta \psi(\theta) - \{J_\phi^\theta(\theta)\}^T \mathbb{E}_\pi[\bar{S}(Y,\theta)]$

and thus $\nabla_s w(s) = +\{J_{\bar{\theta}}^s(s)\}^T \{\nabla_\theta \psi(\bar{\theta}(s)) - \{J_\phi^\theta(\bar{\theta}(s))\}^T \mathbb{E}_\pi[\bar{S}(Y;\bar{\theta}(s))] + R(\bar{\theta}(s))\}$

where $J_{\bar{\theta}}^s(s)$ is the Jacobian of the function $s \mapsto \bar{\theta}(s)$ at $s \in \mathcal{S}$.

For any $s \in \mathcal{S}$, $\bar{\theta}(s)$ is the minimum of $\theta \mapsto \psi(\theta) - \langle \phi(\theta), s\rangle + R(\theta)$.

$$\nabla_\theta \psi(\bar{\theta}(s)) - \{J_\phi^\theta(\bar{\theta}(s))\}^T s + \nabla_\theta R(\bar{\theta}(s)) = 0$$

$$\Rightarrow \boxed{\nabla_s w(s) = -\{J_{\bar{\theta}}^s(s)\}^T \{J_\phi^\theta(\bar{\theta}(s))\}^T h(s).}$$

For all $s \in \mathcal{S}$, $\nabla_\theta \psi(\bar\theta(s)) - J_\phi^\theta(\bar\theta(s))s + \lambda \nabla_\theta P(\bar\theta(s)) = 0 . = \nabla_\theta \ell(s,\theta)$.

We set $\Phi(s,\theta) = \nabla_\theta \psi(\theta) - \{J_\phi^\theta(\theta)\}^T s + \nabla_\theta R(\theta)$.

$$D_s \Phi(s,\theta) = -\{J_\phi^\theta(\theta)\}^T$$

$$D_\theta \Phi(s,\theta) = H_\ell^\theta(s,\theta)$$ which is the Hessian of

$$\theta \mapsto \ell(s,\theta) = \psi(\theta) - \langle \phi(\theta), s\rangle + R(\theta).$$

Differentiating w.r.t s: $s \mapsto \nabla_\theta \ell(s, \bar\theta(s))$ we get:

$$-\{J_\phi^\theta(\bar\theta(s))\}^T + H_\ell^\theta(s, \bar\theta(s))\, J_{\bar\theta}^s(s) = 0.$$

showing that: $\{J_\phi^\theta(\bar\theta(s))\}^T = H_\ell^\theta(s,\bar\theta(s))\, J_{\bar\theta}^s(s)$.

$$\Rightarrow \quad J_{\bar\theta}^s(s) = [H_\ell^\theta(s,\bar\theta(s))]^{-1}\{J_\phi^\theta(\bar\theta(s))\}^T$$

$$\nabla_s w(s) = -\{J_{\bar\theta}^s(s)\}^T\{J_\phi^\theta(\bar\theta(s))\}^T h(s) =$$

$$\Rightarrow \quad \boxed{\nabla_s w(s) = -J_\phi^\theta(\bar\theta(s))\,[H_\ell^\theta(s,\bar\theta(s))]^{-1}\{J_\phi^\theta(\bar\theta(s))\}^T h(s)}$$

and $\langle h(s), \nabla_s w(s)\rangle = -\, h(s)^T J_\phi^\theta(\bar\theta(s))\,[H_\theta \ell(s,\bar\theta(s))]^{-1}\{J_\phi^\theta(\bar\theta(s))\}^T h(s)$.