Утверждение (Pedregosa, 2016)

Пусть L — дифференцируемая функция, такая что все стационарные точки L являются локальными минимумами. Пусть также гессиан H^{-1} функции потерь L является обратимым в каждой стационарной точке. Тогда

$$\nabla_{\mathrm{h}} Q\left(T\left(\theta_0, \mathbf{h}\right), \mathbf{h}\right)=\nabla_{\mathbf{h}} Q\left(\boldsymbol{\theta}^\eta, \mathbf{h}\right)-\nabla_{\mathbf{h}} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}^\eta, \mathbf{h})^{\top} \mathbf{H}^{-1} \nabla_{\boldsymbol{\theta}} Q\left(\boldsymbol{\theta}^\eta, \mathbf{h}\right)$$

Доказательство

Для стационарной точки выполняется, что $\nabla_\theta L(T(\theta_0, h))=0$

Следовательно $\nabla_h\left(\nabla_\theta L\left(T\left(\theta_0, h\right)\right)\right)=\nabla_{\theta, h} L\left(\theta^\eta, h\right)+\nabla_\theta^2 L\left(\theta^\eta, h\right) \nabla_h \theta=0$

Из предыдущего выражаем $\nabla_h \theta=-\left(\nabla_\theta^2 L\left(\theta^\eta, h\right)\right)^{-1} \nabla_{\theta, h} L\left(\theta^\eta, h\right)$

Тогда

$$\begin{aligned}
\nabla_h Q\left(T\left(\theta_0, h\right)\right) & =\nabla_h Q\left(\theta^\eta, h\right)+\nabla_\theta Q(\theta^\eta, h)^{\top} \nabla_h \theta \\
& =\nabla_h Q\left(\theta^\eta, h\right)-\nabla_\theta Q(\theta^\eta, h)^{\top}\left(\nabla_\theta^2 L\left(\theta^\eta, h\right)\right)^{-1} \nabla_{\theta, h} L\left(\theta^\eta, h\right) \\
& =\nabla_h Q\left(\theta^\eta, h\right)-\nabla_{\theta, h} L(\theta^\eta, h)^{\top}\left(\nabla_\theta^2 L\left(\theta^\eta, h\right)\right)^{-1} \nabla_\theta Q\left(\theta^\eta, h\right)
\end{aligned}$$