Теорема 1. *Пусть $\mathcal{L}$ — дифференцируемая функция, такая что все стационарные точки $\mathcal{L}$ являются локальными минимумами. Пусть также гессиан $\mathbf{H}^{-1}$ функции потерь $\mathcal{L}$ является обратимым в каждой стационарной точке, тогда:*

$$\nabla_{\mathbf{h}}\mathcal{Q}\left(\mathsf{T}\left(\mathbf{\Theta}_0, \mathbf{h}\right), \mathbf{h}\right) = \nabla_{\mathbf{h}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) - \nabla_{\mathbf{h}}\nabla_{\mathbf{\Theta}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)^{\mathsf{T}}\mathbf{H}^{-1}\nabla_{\mathbf{\Theta}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right). \tag{1}$$

Доказательство.

$$\begin{aligned} &\nabla_{\mathbf{\Theta}}\mathcal{L}\left(\mathsf{T}\left(\mathbf{\Theta}_0, \mathbf{h}\right)\right) = 0 \Rightarrow \\ \Rightarrow &\nabla_{\mathbf{h}}\left(\nabla_{\mathbf{\Theta}}\mathcal{L}\left(\mathsf{T}\left(\mathbf{\Theta}_0, \mathbf{h}\right)\right)\right) = \nabla_{\mathbf{\Theta},\mathbf{h}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) + \nabla^2_{\mathbf{\Theta}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)\frac{\partial\mathbf{\Theta}}{\partial\mathbf{h}} = 0 \Rightarrow \\ \Rightarrow &\frac{\partial\mathbf{\Theta}}{\partial\mathbf{h}} = -\left(\nabla^2_{\mathbf{\Theta}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)\right)^{-1}\nabla_{\mathbf{\Theta},\mathbf{h}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right). \end{aligned} \tag{2}$$

$$\nabla_{\mathbf{h}}\mathcal{Q}\left(\mathsf{T}\left(\mathbf{\Theta}_0, \mathbf{h}\right)\right) = \nabla_{\mathbf{h}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) + \nabla_{\mathbf{\Theta}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)^{\mathsf{T}}\frac{\partial\mathbf{\Theta}}{\partial\mathbf{h}}. \tag{3}$$

Подставляя (2) в (3) получаем:

$$\begin{aligned} \nabla_{\mathbf{h}}\mathcal{Q}\left(\mathsf{T}\left(\mathbf{\Theta}_0, \mathbf{h}\right)\right) &= \nabla_{\mathbf{h}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) - \nabla_{\mathbf{\Theta}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)^{\mathsf{T}}\left(\nabla^2_{\mathbf{\Theta}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)\right)^{-1}\nabla_{\mathbf{\Theta},\mathbf{h}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) = \\ &= \nabla_{\mathbf{h}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) - \nabla_{\mathbf{\Theta},\mathbf{h}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)^{\mathsf{T}}\left(\nabla^2_{\mathbf{\Theta}}\mathcal{L}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right)\right)^{-1}\nabla_{\mathbf{\Theta}}\mathcal{Q}\left(\mathbf{\Theta}^{\eta}, \mathbf{h}\right) \end{aligned} \tag{4}$$

□