1 Утверждение

Теорема 1. *Пусть* $h \in \mathcal{F}^1$ *— дифференцируемая выпуклая функция. Пусть также гессиан* $\nabla_1^2 h$ *функции потерь обратимым на множестве оптимальных точек, тогда:*

$$\nabla_\lambda g\left(\mathsf{T}\left(\boldsymbol{X}_0, \lambda\right), \lambda\right) = \nabla_\lambda g(X^\nu(\lambda), \lambda) - \left(\nabla_{X,\lambda}^2 h\right)^T \left(\nabla_X^2 h\right)^{-1} \nabla_X g(X^\nu(\lambda), \lambda)$$

Доказательство. Запишем условие оптимальности первого порядка

$$\nabla_X h(X(\lambda), \lambda) = 0$$

продифференцируем это уравнение

$$\nabla_{X,\lambda}^2 h + \nabla_X^2 h \cdot \mathrm{D}X = 0$$

выразим производную X, используя обратимость гессиана.

$$\mathrm{D}X = -\left(\nabla_X^2 h\right)^{-1} \nabla_{X,\lambda}^2 h$$

Теперь запишем полный градиент

$$\nabla_\lambda g\left(\mathsf{T}\left(\boldsymbol{X}_0, \lambda\right), \lambda\right) = \nabla_\lambda g(X^\nu(\lambda), \lambda) + \nabla_X g(X^\nu(\lambda), \lambda)\mathrm{D}X$$

Используем знание от том, что $A^T b = b^T A$ и подставляем выражение для $(\mathrm{D}X)^T = -\left(\left(\nabla_X^2 h\right)^{-1} \nabla_{X,\lambda}^2 h\right)^T = \left(\nabla_{X,\lambda}^2 h\right)^T \left(\nabla_X^2 h\right)^{-T}$, поскольку гессиан симметричен, то $-T$ переходит в -1. Подставляем

$$\nabla_\lambda g\left(\mathsf{T}\left(\boldsymbol{X}_0, \lambda\right), \lambda\right) = \nabla_\lambda g(X^\nu(\lambda), \lambda) - \left(\nabla_{X,\lambda}^2 h\right)^T \left(\nabla_X^2 h\right)^{-1} \nabla_X g(X^\nu(\lambda), \lambda)$$

Что и требовалось доказать. □