Exercise Sheet #12

Submission due on Jan 26th, 2023 at 2.15pm on Ilias

1. **EXAMple Question**
   Let $q : \mathbb{R}^D \to \mathbb{R}$, $q(\boldsymbol{\theta}) := \mathcal{L}(\boldsymbol{\theta}^{(k)}) + (\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\intercal \boldsymbol{g} + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})^\intercal \boldsymbol{H}(\boldsymbol{\theta} - \boldsymbol{\theta}^{(k)})$ denote the quadratic approximation of the loss $\mathcal{L}$ around $\boldsymbol{\theta}^{(k)} \in \mathbb{R}^D$, with $\boldsymbol{g} := \nabla\mathcal{L}(\boldsymbol{\theta}^{(k)}) \in \mathbb{R}^D$ and $\boldsymbol{H} := \nabla^2\mathcal{L}(\boldsymbol{\theta}^{(k)}) \in \mathbb{R}^{D\times D}$.

   (a) Show that a cut $h : \mathbb{R} \to \mathbb{R}$, $h(\tau) := q(\boldsymbol{\theta}^{(k)} + \tau\boldsymbol{d})$ through the quadratic in some direction $\boldsymbol{d} \in \mathbb{R}^D$ is a parabola with curvature $h''(\tau) = \boldsymbol{d}^\intercal\boldsymbol{H}\boldsymbol{d}$.

   (b) Show that, if the direction is set to a normalized eigenvector $\boldsymbol{e}$ of the Hessian (i.e. $\|e\| = 1$), the directional curvature coincides with the corresponding eigenvalue.

2. **Coding exercise**
   This week's coding exercise is concerned with the Hessian-free method and the role of damping in stochastic second-order optimization. You can find all instructions in the Jupyter notebook `Exercise_12.ipynb`.