# Bounding the error in approximations

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a strictly convex function and assume that we can bound its Hessian $H$ from below as

$$H \succeq H_0 \succeq 0$$

(where $H \succeq H_0$ means that $H - H_0$ is positive semidefinite).

Now let $\vec{x}^*$ denote the minimizer of the function $f$ and $\vec{x}^0$ an approximation and write $\vec{y} = \vec{x}^0 - \vec{x}^*$. Using Taylor's theorem in the Lagrange form, we find

$$f(\underbrace{\vec{x}^* + \vec{y}}_{\vec{x}^0}) = f(\vec{x}^*) + \frac{1}{2} \vec{y}^T H(\vec{\xi}^*) \vec{y}$$

where $\vec{\xi}$ is a vector somewhere between $\vec{x}^*$ and $\vec{x}^0$. Note that the linear order vanishes because $\vec{x}^*$ is a minimizer.

Similar we can do a Taylor expansion around $\vec{x}^0$ and find

$$f(\underbrace{\vec{x}^0 - \vec{y}}_{\vec{x}^*}) = f(\vec{x}^0) - \underset{\substack{\uparrow \\ \text{gradient}}}{\vec{g}}(\vec{x}^0)^T \vec{y} + \frac{1}{2} \vec{y}^T H(\vec{\xi}^0) \vec{y}.$$

Substituting the first equation into the second one yields

$$f(\vec{x}^*) = f(\vec{x}^*) + \frac{1}{2} \vec{y}^T H(\vec{\xi}^*) y - g(\vec{x}_0)^T \vec{y}$$
$$+ \frac{1}{2} \vec{y}^T H(\vec{\xi}^0) \vec{y}$$

$$\Rightarrow \quad 0 = -g^T(\vec{x}^0)^T \vec{y} + \frac{1}{2} \vec{y}^T \Big( \underbrace{H(\vec{\xi}^y) + H(\vec{\xi}^0)}_{\succcurlyeq H_0} \Big) \vec{y}$$

$$\Rightarrow \quad 0 \succcurlyeq -g(\vec{x}^0)^T \vec{y} + \vec{y}^T H_0 \vec{y}$$

$$\Leftrightarrow \quad \vec{y}^T H_0 \vec{y} \leq g(\vec{x}^0)^T \vec{y}$$

Now we can use the Cauchy – Schwarz inequality to get

$$|g^T \vec{y}| = (H_0^{-1/2} g)^T (H_0^{\frac{1}{2}} \vec{y}) \leq \| H_0^{-1/2} g \| \cdot \sqrt{\vec{y}^T H_0 \vec{y}}$$

Then we obtain a bound for the deviation in the form

$$\Rightarrow \quad \vec{y}^T H_0 \vec{y} \leq \vec{g}(\vec{x}_0)^T H_0^{-1/2} \vec{g}(\vec{x}_0)$$

~~Note~~
Three remarks:

→ The bound does <u>not</u> depend on $\vec{x}^*$, we just need the gradient $\vec{g}$ at $\vec{x}^0$ and a lower bound for the Hessian.

→ The success of this method crucially depends on whether we find a good bound $H_0$...

→ If we include constraints, we typically have to work with the Lagrangian instead of the objective function itself.