Even faster convergence is possible for "smooth" functions.

**DEF** ($\beta$-smoothness) $g:\mathbb{R}^d \to \mathbb{R}$ (differentiable) is $\beta$-smooth for $\beta > 0$ if

$$\|\nabla g(\theta) - \nabla g(\theta')\|_2 \le \beta \cdot \|\theta - \theta'\|_2^2,$$

for any $\theta, \theta' \in \mathrm{dom}(g)$.

**LEMMA 1** $g:\mathbb{R}^d \to \mathbb{R}$ is $\beta$-smooth

$$\Rightarrow \left| g(\phi) - g(\theta) - \nabla^T g(\theta)[\phi - \theta] \right| \le \frac{\beta}{2}\|\theta - \phi\|_2^2,$$

$\forall \theta, \phi \in \mathrm{dom}\, g$.

**Pf:** Given $\theta, \phi \in \mathrm{dom}(g)$, let $h(t) = g(\theta + t(\phi - \theta))$, $t \in [0,1]$.

Then, $h'(t) = \nabla^T g(\theta + t(\phi - \theta))[\phi - \theta]$.

By the fundamental theorem of calculus

$$h(1) - h(0) = g(\phi) - g(\theta) = \int_0^1 h'(t)\,dt$$

$$= \int_0^1 \nabla^T g(\theta + t(\phi - \theta))[\phi - \theta]\,dt$$

$$= \nabla^T g(\theta)[\phi - \theta] + \int_0^1 \left[\nabla g(\theta + t[\phi - \theta]) - \nabla g(\theta)\right]^T [\phi - \theta]\,dt$$

Then,

$$\left| g(\phi) - g(\theta) - \nabla^T g(\theta)[\phi - \theta] \right| \le \left| \int_0^1 \left[\nabla g(\theta + t(\phi - \theta)) - \nabla g(\theta)\right]^T [\phi - \theta]\,dt \right|$$

$$\le \int_0^1 \|\nabla g(\theta + t(\phi - \theta)) - \nabla g(\theta)\|_2 \cdot \|\phi - \theta\|_2\,dt$$

$$\le \int_0^1 \beta \cdot t \cdot \|\theta - \phi\|_2^2\,dt = \frac{\beta}{2}\|\theta - \phi\|_2^2.$$

**LEMMA 2 (Descent)** Let $g : \mathbb{R}^d \to \mathbb{R}$ be $\beta$-smooth. Given $\theta \in \mathrm{dom}\, g$, let $\varphi(\theta) = \theta - \frac{1}{\beta} \nabla g(\theta)$. Then,

$$g(\varphi(\theta)) - g(\theta) \le -\frac{1}{2\beta} \|\nabla g(\theta)\|_2^2 .$$

**Pf:** $\left| g(\varphi(\theta)) - g(\theta) + \frac{1}{\beta} \|\varphi(\theta) - \theta\|_2^2 \right| \le \frac{1}{2\beta} \|\nabla g(\theta)\|_2^2 .$

Thus,

$$g(\varphi(\theta)) - g(\theta) \le -\frac{1}{2\beta} \|\nabla g(\theta)\|_2^2 .$$

**Note:** If $g$ is $\beta$-smooth, then g.d with $\eta = \frac{1}{\beta}$ leads to descent in the function value.

**THEOREM** Assume that $g : \mathbb{R}^d \to \mathbb{R}$ is $\alpha$-sc., $\beta$-smooth. Then, GD with $\eta = \frac{1}{\beta}$ yields

$$g(\theta_T) - g(\theta^*) \le e^{-\frac{T}{\kappa}} \cdot \left( g(\theta_0) - g(\theta^*) \right),$$

for any $\theta_0 \in \mathrm{dom}\, g$, where $\kappa = \frac{\alpha}{\beta}$.

**Pf:**
$$\begin{aligned} g(\theta_{t+1}) - g(\theta^*) &\le g(\varphi(\theta_t)) - g(\theta^*) \\ &\le g(\theta_t) - g(\theta^*) - \frac{1}{2\beta} \|\nabla g(\theta_t)\|_2^2 \\ &\le g(\theta_t) - g(\theta^*) - \frac{\alpha}{\beta} \cdot \left( g(\theta_t) - g(\theta^*) \right) \\ &= \left[ 1 - \frac{\alpha}{\beta} \right] \left( g(\theta_t) - g(\theta^*) \right). \end{aligned}$$

By induction,

$$g(\theta_{t+1}) - g(\theta^*) \le \left[ 1 - \frac{\alpha}{\beta} \right]^{t+1} \left( g(\theta_0) - g(\theta^*) \right),$$

for any $t \ge 0$.