3. **Goldstein-Armijo Rule**

   Find a sequence $\{h_k\}_{k=0}^{\infty}$ such that

$$\begin{array}{rcl}\alpha\langle f'(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1}\rangle & \leq & f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}), \\ \beta\langle f'(\boldsymbol{x}_k), \boldsymbol{x}_k - \boldsymbol{x}_{k+1}\rangle & \geq & f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}),\end{array}$$

   where $0 < \alpha < \beta < 1$ are fixed parameters.

   Since $f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k - h_k f'(\boldsymbol{x}_k))$,

$$f(\boldsymbol{x}_k) - \beta h_k \|f'(\boldsymbol{x}_k)\|_2^2 \leq f(\boldsymbol{x}_{k+1}) \leq f(\boldsymbol{x}_k) - \alpha h_k \|f'(\boldsymbol{x}_k)\|_2^2.$$

   The acceptable steps exist unless $f(\boldsymbol{x}_{k+1}) = f(\boldsymbol{x}_k - hf'(\boldsymbol{x}_k))$ is not bounded from below.

4. **Barzilai-Borwein Step-Size**[1]

   Let us define $\boldsymbol{s}_{k-1} := \boldsymbol{x}_k - \boldsymbol{x}_{k-1}$ and $\boldsymbol{y}_{k-1} := f'(\boldsymbol{x}_k) - f'(\boldsymbol{x}_{k-1})$. Then, we can define the Barzilai-Borwein (BB) step sizes $\{h_k^1\}_{k=1}^{\infty}$ and $\{h_k^2\}_{k=1}^{\infty}$:

$$h_k^1 := \frac{\|\boldsymbol{s}_{k-1}\|_2^2}{\langle \boldsymbol{s}_{k-1}, \boldsymbol{y}_{k-1}\rangle},$$

$$h_k^2 := \frac{\langle \boldsymbol{s}_{k-1}, \boldsymbol{y}_{k-1}\rangle}{\|\boldsymbol{y}_{k-1}\|_2^2}.$$

   The first step-size is the one which minimizes the following secant condition $\|\frac{1}{h}\boldsymbol{s}_{k-1} - \boldsymbol{y}_{k-1}\|_2^2$ while the second one minimizes $\|\boldsymbol{s}_{k-1} - h\boldsymbol{y}_{k-1}\|_2^2$.

Now, consider the problem

$$\boxed{\begin{array}{c}\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \\ \text{where } f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n), \text{ and } f(\boldsymbol{x}) \text{ is bounded from below.}\end{array}}$$

Let us evaluate the result of one step of the steepest descent method.
Consider $\boldsymbol{y} = \boldsymbol{x} - hf'(\boldsymbol{x})$. From Lemma 3.4,

$$\begin{array}{rcl}f(\boldsymbol{y}) & \leq & f(\boldsymbol{x}) + \langle f'(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x}\rangle + \dfrac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \\ & = & f(\boldsymbol{x}) - h\|f'(\boldsymbol{x})\|_2^2 + \dfrac{h^2 L}{2}\|f'(\boldsymbol{x})\|_2^2 \\ & & \\ & = & f(\boldsymbol{x}) - h\left(1 - \dfrac{h}{2}L\right)\|f'(\boldsymbol{x})\|_2^2. \end{array} \qquad (5)$$

Thus, one step of the steepest descent method decreases the value of the objective function at least as follows for $h^* = 1/L$.

$$f(\boldsymbol{y}) \leq f(\boldsymbol{x}) - \frac{1}{2L}\|f'(\boldsymbol{x})\|_2^2.$$

[1] J. Barzilai and J. M. Borwein, "Two-point step size gradient methods," *IMA Journal of Numerical Analysis*, **8** (1988), pp. 141–148.

Now, for the Goldstein-Armijo Rule, since $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - h_k f'(\boldsymbol{x}_k)$, we have:

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \leq \beta h_k \|f'(\boldsymbol{x}_k)\|_2^2,$$

and from (5)

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \geq h_k \left(1 - \frac{h_k}{2} L\right) \|f'(\boldsymbol{x}_k)\|_2^2.$$

Therefore, $h_k \geq 2(1-\beta)/L$.

Also, substituting in

$$f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \geq \alpha h_k \|f'(\boldsymbol{x}_k)\|_2^2 \geq \frac{2}{L}\alpha(1-\beta)\|f'(\boldsymbol{x}_k)\|_2^2.$$

Thus, in the three step-size strategies excepting the BB step size considered here, we can say that

$$\boxed{f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k+1}) \geq \frac{\omega}{L}\|f'(\boldsymbol{x}_k)\|_2^2}$$

for some positive constant ω.

Summing up the above inequality we have:

$$\frac{\omega}{L}\sum_{k=0}^{N} \|f'(\boldsymbol{x}_k)\|_2^2 \leq f(\boldsymbol{x}_0) - f(\boldsymbol{x}_{N+1}) \leq f(\boldsymbol{x}_0) - f^*$$

where f^* is the optimal value of the problem.

As a simple consequence we have

$$\|f'(\boldsymbol{x}_k)\|_2 \to 0 \quad \text{as} \quad k \to \infty.$$

Finally,

$$g_N^* \equiv \min_{0 \leq k \leq N} \|f'(\boldsymbol{x}_k)\|_2 \leq \frac{1}{\sqrt{N+1}}\left[\frac{1}{\omega}L(f(\boldsymbol{x}_0) - f^*)\right]^{1/2}. \tag{6}$$

Remark 5.8 $g_N^* \to 0$, but we cannot say anything about the rate of convergence of the sequence $\{f(\boldsymbol{x}_k)\}$ or $\{\boldsymbol{x}_k\}$.

Example 5.9 Consider the function $f(x,y) = \frac{1}{2}x^2 + \frac{1}{4}y^4 - \frac{1}{2}y^2$. $(0,-1)^T$ and $(0,1)^T$ are local minimal solutions, but $(0,0)^T$ is a stationary point.

If we start the steepest descent method from $(1,0)^T$, we will only converge to the stationary point.

We focus now on the following problem class:

Model:	1. $\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x})$ 2. $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$ 3. $f(\boldsymbol{x})$ is bounded from below
Oracle:	Only function values are available
Approximate solution:	Find $\bar{\boldsymbol{x}} \in \mathbb{R}^n$ such that $f(\bar{\boldsymbol{x}}) \leq f(\boldsymbol{x}_0)$ and $\|f'(\bar{\boldsymbol{x}})\|_2 < \epsilon$

From (6), we have

$$g_N^* < \varepsilon \quad \text{if} \quad N+1 > \frac{L}{\omega\varepsilon^2}(f(\boldsymbol{x}_0) - f^*).$$