

$$h_k := \frac{h}{\sqrt{k+1}}.$$

This is the simplest strategy.

2. Exact Line Search (Cauchy Step-Size)

The sequence $\{h_k\}_{k=0}^{\infty}$ is chosen such that

$$h_k := \arg \min_{h \geq 0} f(\mathbf{x}_k - hf'(\mathbf{x}_k)).$$

This choice is only theoretical since even for the one dimensional case, it is very difficult and expensive.

3. Goldstein-Armijo Rule

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

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

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

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

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

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

4. Barzilai-Borwein Step-Size¹

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

$$\begin{aligned} h_k^1 &:= \frac{\|\mathbf{s}_{k-1}\|_2^2}{\langle \mathbf{s}_{k-1}, \mathbf{y}_{k-1} \rangle}, \\ h_k^2 &:= \frac{\langle \mathbf{s}_{k-1}, \mathbf{y}_{k-1} \rangle}{\|\mathbf{y}_{k-1}\|_2^2}. \end{aligned}$$

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

Now, consider the problem

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

where $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$, and $f(\mathbf{x})$ is bounded from below.

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

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

$$\begin{aligned}
f(\mathbf{y}) &\leq f(\mathbf{x}) + \langle f'(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 \\
&= f(\mathbf{x}) - h\|f'(\mathbf{x})\|_2^2 + \frac{h^2 L}{2} \|f'(\mathbf{x})\|_2^2 \\
&= f(\mathbf{x}) - h \left(1 - \frac{h}{2} L\right) \|f'(\mathbf{x})\|_2^2.
\end{aligned} \tag{5}$$

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

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

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

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

and from (5)

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

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

Also, substituting in

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

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

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

for some positive constant ω .

Summing up the above inequality we have:

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

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

As a simple consequence we have

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

Finally,

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

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

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:	<ol style="list-style-type: none"> 1. $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$ 2. $f \in \mathcal{C}_L^{1,1}(\mathbb{R}^n)$ 3. $f(\mathbf{x})$ is bounded from below
Oracle:	Only function values are available
Approximate solution:	Find $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that $f(\bar{\mathbf{x}}) \leq f(\mathbf{x}_0)$ and $\ f'(\bar{\mathbf{x}})\ _2 < \epsilon$

From (6), we have

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

Remark 5.10 This is much better than the result of Theorem 5.6, since *it does not depend on n* .

Finally, consider the following problem under Assumption 5.11.

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$$

Assumption 5.11

1. $f \in \mathcal{C}_M^{2,2}(\mathbb{R}^n)$;
2. There is a local minimum \mathbf{x}^* of the function $f(\mathbf{x})$;
3. We know some bound $0 < \ell \leq L < \infty$ for the Hessian at \mathbf{x}^* :

$$\ell \mathbf{I} \preceq f''(\mathbf{x}^*) \preceq L \mathbf{I};$$

4. Our starting point \mathbf{x}_0 is close enough to \mathbf{x}^* .

Theorem 5.12 Let $f(\mathbf{x})$ satisfy our assumptions above and let the starting point \mathbf{x}_0 be close enough to a local minimum:

$$r_0 = \|\mathbf{x}_0 - \mathbf{x}^*\|_2 < \bar{r} := \frac{2\ell}{M}.$$

Then, the steepest descent method with step-size $h^* = 2/(L + \ell)$ converges as follows:

$$\|\mathbf{x}_k - \mathbf{x}^*\|_2 \leq \frac{\bar{r} r_0}{\bar{r} - r_0} \left(1 - \frac{2\ell}{L + 3\ell} \right)^k.$$

This rate of convergence is called (R-)linear.

Proof:

In the steepest descent method, the iterates are $\mathbf{x}_{k+1} = \mathbf{x}_k - h_k f'(\mathbf{x}_k)$.

Since $f'(\mathbf{x}^*) = 0$,

$$f'(\mathbf{x}_k) = f'(\mathbf{x}_k) - f'(\mathbf{x}^*) = \int_0^1 f''(\mathbf{x}^* + \tau(\mathbf{x}_k - \mathbf{x}^*))(\mathbf{x}_k - \mathbf{x}^*) d\tau = \mathbf{G}_k(\mathbf{x}_k - \mathbf{x}^*),$$

and therefore,

$$\mathbf{x}_{k+1} - \mathbf{x}^* = \mathbf{x}_k - \mathbf{x}^* - h_k \mathbf{G}_k(\mathbf{x}_k - \mathbf{x}^*) = (\mathbf{I} - h_k \mathbf{G}_k)(\mathbf{x}_k - \mathbf{x}^*).$$

Let $r_k = \|\mathbf{x}_k - \mathbf{x}^*\|_2$. From Lemma 3.6,

$$f''(\mathbf{x}^*) - \tau M r_k \mathbf{I} \preceq f''(\mathbf{x}^* + \tau(\mathbf{x}_k - \mathbf{x}^*)) \preceq f''(\mathbf{x}^*) + \tau M r_k \mathbf{I}.$$

Integrating all parts from 0 to 1 and using our hypothesis,

$$(\ell - \frac{r_k}{2}M)\mathbf{I} \preceq \mathbf{G}_k \preceq (L + \frac{r_k}{2}M)\mathbf{I}.$$

Therefore,

$$\left(1 - h_k(L + \frac{r_k}{2}M)\right)\mathbf{I} \preceq \mathbf{I} - h_k \mathbf{G}_k \preceq \left(1 - h_k(\ell - \frac{r_k}{2}M)\right)\mathbf{I}.$$

We arrive at

$$\|\mathbf{I} - h_k \mathbf{G}_k\|_2 \leq \max\{|a_k(h_k)|, |b_k(h_k)|\}$$

where $a_k(h) = 1 - h(\ell - \frac{r_k}{2}M)$ and $b_k(h) = h(L + \frac{r_k}{2}M) - 1$.

Notice that $a_k(0) = 1$ and $b_k(0) = -1$.

Now, let us use our hypothesis that $r_0 < \bar{r}$.

When $a_k(h) = b_k(h)$, we have $1 - h(\ell - \frac{r_k}{2}M) = h(L + \frac{r_k}{2}M) - 1$, and therefore

$$h_k^* = \frac{2}{L + \ell}.$$

(Surprisingly, it does not depend neither on M nor r_k). Finally,

$$r_{k+1} = \|\mathbf{x}_{k+1} - \mathbf{x}^*\|_2 \leq \left(1 - \frac{2}{L + \ell} \left(\ell - \frac{r_k}{2}M\right)\right) \|\mathbf{x}_k - \mathbf{x}^*\|_2.$$

That is,

$$r_{k+1} \leq \left(\frac{L - \ell}{L + \ell} + \frac{r_k M}{L + \ell}\right) r_k.$$

and $r_{k+1} < r_k < \bar{r}$.

Now, let us analyze the rate of convergence. Multiplying the above inequality by $M/(L + \ell)$,

$$\frac{M r_{k+1}}{L + \ell} \leq \frac{M(L - \ell)}{(L + \ell)^2} r_k + \frac{M^2 r_k^2}{(L + \ell)^2}.$$

Calling $\alpha_k = \frac{M r_k}{L + \ell}$ and $q = \frac{2\ell}{L + \ell}$, we have

$$\alpha_{k+1} \leq (1 - q)\alpha_k + \alpha_k^2 = \alpha_k(1 + \alpha_k - q) = \frac{\alpha_k(1 - (\alpha_k - q)^2)}{1 - (\alpha_k - q)}. \quad (7)$$

Now, since $r_k < \frac{2\ell}{M}$, $\alpha_k - q = \frac{M r_k}{L + \ell} - \frac{2\ell}{L + \ell} < 0$, and $1 + (\alpha_k - q) = \frac{L - \ell}{L + \ell} + \frac{M r_k}{L + \ell} > 0$. Therefore, $-1 < \alpha_k - q < 0$, and (7) becomes $\leq \frac{\alpha_k}{1 + q - \alpha_k}$.

$$\frac{1}{\alpha_{k+1}} \geq \frac{1 + q}{\alpha_k} - 1.$$

$$\frac{q}{\alpha_{k+1}} - 1 \geq \frac{q(1 + q)}{\alpha_k} - q - 1 = (1 + q) \left(\frac{q}{\alpha_k} - 1\right).$$

and then,

$$\frac{q}{\alpha_k} - 1 \geq (1 + q)^k \left(\frac{q}{\alpha_0} - 1\right) = (1 + q)^k \left(\frac{2\ell}{L + \ell} \frac{L + \ell}{M r_0} - 1\right) = (1 + q)^k \left(\frac{\bar{r}}{r_0} - 1\right).$$

Finally, we arrive at

$$r_k = \|\mathbf{x}_k - \mathbf{x}^*\|_2 \leq \frac{\bar{r} r_0}{\bar{r} - r_0} \left(1 - \frac{2\ell}{L + 3\ell}\right)^k.$$

■