Unconstrained optimization

The problem (P):

$$\min f(x), \quad x \in R^n.$$

1. **Gradient method**

Algorithm:

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k).$$

Choosing α_k:

(i). Choose α_k so that

$$f(x^k - \alpha_k \nabla f(x^k)) = \min_{\alpha \geq 0}(f(x^k - \alpha \nabla f(x^k))).$$

(ii).
(1) Choose $\alpha > 0$, $0 < \varepsilon < 1$, $0 < \lambda < 1$.
(2) If

$$f(x^k - \alpha \nabla f(x^k)) - f(x^k) < -\varepsilon \alpha \|\nabla f(x^k)\|^2, \tag{1}$$

then $\alpha_k = \alpha$, otherwise multiple α by λ until (1) will be satisfied.

Theorem 1.1. Let f be a continuously differentiable function, the set

$$\{x : f(x) \leq f(x^0)\}$$

is compact and α_k is chosen in accordance with (i) or (ii). Then

$$\lim \|\nabla f(x_k)\| = 0.$$

Theorem 1.2. Let f be a twice continuously differentiable function, its Hessian matrix $H(x)$ satisfies the following condition:

$$m\|y\|^2 \leq (H(x)y, y) \leq M\|y\|^2, \; M > m > 0,$$

for each $x, \; y \in R^n$, x^0 ia an arbitrary initial point and α_k is chosen in accordance with (i) or (ii. Then

$$x^k \to x^*, \; f(x^k) \to f(x^*),$$

where x^* is a solution of the problem (P). Moreover the following estimate is valid:

$$f(x^k) - f(x^*) \le q^k(f(x^0) - f(x^*)),\ 0 < q < 1.$$

2. **The Newton method**

Algorithm.
1. $p^k = -H^{-1}(x^k)\nabla f(x^k)$.
2. $x^{k+1} = x^k + \alpha_k p^k$.

Choosing α_k:

(i) $\alpha_k = 1$ for all k (Basic Newton method).

(ii). Choose α_k so that

$$f(x^k + \alpha_k p^k) = \min_{\alpha \ge 0} f(x^k + \alpha p^k).$$

(iii).
(1) Choose $\alpha = 1$, $0 < \varepsilon < 1/2$, $0 < \lambda < 1$.
(2) If

$$f(x^k + \alpha p^k) - f(x^k) < \varepsilon\alpha(\nabla f(x^k), p^k) \tag{2}$$

then $\alpha_k = \alpha$, otherwise multiple α by λ until (2) will be satisfied.

Theorem 2.1. Let f be three times continuously differentiable in a neighbourhood of $x^* \in R^n$, and x^* be a non-degenerate local minimizer of f, that is, $\nabla f(x^*) = 0$ and $H(x^*)$ is positive definite. Then the Basic Newton method, starting close enough to x^*, converges to x^* quadratically.

Theorem 2.2. Let f be a strictly convex twice continuously differentiable function with the Hessian matrix satisfying the condition:

$$m\|y\|^2 \le (H(x)y, y) \le M\|y\|^2,\ M > m > 0,$$

for each $x,\ y \in R^n$. If α_k are chosen with accordance with (ii) or (iii), then $x^k \to x^*$ for each x^0 and

$$\|x^{N+l} - x^*\| \le C\lambda_N \dots \lambda_{N+l}$$

for some $N, C > 0$. $\lambda_{N+l} < 1$ and $\lim_{l\to\infty} \lambda_{N+l} = 0$.

3. **Self-concordant functions**

Let f be a three times continuously differentiable convex function defined on an open convex set Q, and $\varphi_{x,h}(t) = f(x+th),\ x \in Q, h \in R^n, t \in R$. We say that f is a *self-concordant function* if

1. $|\varphi'''_{x,h}(0)| \leq 2(\varphi''_{x,h}(0))^{3/2}$, for each $x \in Q, h \in R^n$.

2. If a sequence $\{x^k\} \subset Q$ converges to a boundary point of Q then $f(x^k) \to \infty$.

The second condition is irrelevant if $Q = R^n$

Examples

1. The convex quadratic function $f(x) = \frac{1}{2}(x, Ax) + (b, x) + c$ (A is a symmetric positive semi-definite $n \times n$ matrix) is self-concordant on R^n;

2. Function $-\ln(x)$ is self-concordant on $Q = \{x \in R : x > 0\}$.

3. Let $Q = \{x \in R^n :\ g_i(x) < 0,\ i = 1, 2, \ldots, m\}$, where $g_i(x)$ are linear or strict convex quadratic functions. Then

$$f(x) = -\sum_{i=1}^{m} \ln(-g_i(x))$$

is self-concordant on Q.

Theorem 3.1. Let f be a strictly convex self-concordant below bounded function on Q. Suppose α_k in the Newton method is chosen as

$$\alpha_k = \frac{1}{1+\lambda_k},$$

where

$$\lambda_k = \sqrt{(\nabla f(x^k), H^{-1}(x^k)\nabla f(x^k))}.$$

Then the Newton method has the following properties for each initial point $x^0 \in Q$:

1. $x^k \in Q$, for each k;

2. $\{x^k\}$ converges to the unique minimizer x^* of f on Q;

3. the number of steps required to get an x such that $f(x) - f(x^*) < \varepsilon$ for $\varepsilon < 0.1$ is no more than

$$N(\varepsilon) = A(f(x^0) - f(x^*) + \ln\ln 1/\varepsilon),$$

where A is an absolute constant.