1 More on Steepest Descent (July 9)

1.1 Convergence of Steepest Descent

Theorem 1.1. *Suppose f is a C^1 function on an open set $\Omega \subseteq \mathbb{R}^n$. Let $x_0 \in \Omega$, and let $\{x_k\}_{k=0}^{\infty}$ be the sequence generated by the method of steepest descent. If there is a compact $K \subseteq \Omega$ containing all x_k, then every convergent subsequence of $\{x_k\}_{k=0}^{\infty}$ in K will converge to a critical point x_* of f.*

Proof. Choose a convergent subsequence $\{x_{k_i}\}$ converging to a point $x_* \in K$. Note that $\{f(x_{k_i})\}$ decreases and converges to $f(x_*)$. Since $\{f(x_k)\}$ is a decreasing sequence, it also converges to $f(x_*)$.

Suppose for the sake of contradiction that $\nabla f(x_*) \neq 0$. Since f is C^1, $\nabla f(x_{k_i})$ converges to $\nabla f(x_*)$. Define $y_{k_i} = x_{k_i} - \alpha_{k_i} \nabla f(x_{k_i})$ (i.e. $y_{k_i} = x_{k_1+1}$). We may therefore assume without loss of generality that y_{k_i} converges to some $y_* \in K$. Since $\nabla f(x_*) \neq 0$, we may write

$$\alpha_{k_i} = \frac{|y_{k_i} - x_{k_i}|}{|\nabla f(x_{k_i})|}.$$

Taking the limit as $i \to \infty$, we have

$$\alpha_* := \lim_{i \to \infty} \alpha_{k_i} = \frac{|y_* - x_*|}{|\nabla f(x_*)|}$$

Taking the same limit in the definition of y_{k_i} we have

$$y_* = x_* - \alpha_* \nabla f(x_*).$$

Note that

$$f(y_{k_i}) = f(x_{k_i+1}) = \min_{\alpha \geq 0} f(x_{k_i} - \alpha \nabla f(x_{k_i})).$$

Thus $f(y_{k_i}) \leq f(x_{k_i} - \alpha \nabla f(x_{k_i}))$ for all $\alpha \geq 0$. For any fixed $\alpha \geq 0$, taking the limit $i \to \infty$ gives us

$$f(y_*) \leq f(x_* - \alpha \nabla f(x_*)),$$

implying

$$f(y_*) \leq \min_{\alpha \geq 0} f(x_* - \alpha \nabla f(x_*)) < f(x_*),$$

since the function f decreases in the direction of $-\nabla f(x_*) \neq 0$.

We can also argue the following: $f(x_{k_i+1}) \to f(x_*)$. But since $x_{k_i+1} = y_{k_i}$, we have $f(y_{k_i}) \to f(y_*)$, implying $f(x_*) = f(y_*)$, a contradiction. □

1.2 Steepest Descent in the Quadratic Case

Consider a function f of the form $f(x) = \frac{1}{2} x^T Q x - b^T x$ for $b, x \in \mathbb{R}^n$ and Q an $n \times n$ symmetric positive definite matrix. Let $\lambda = \lambda_1 \leq \cdots \leq \lambda_n = \Lambda$ be the eigenvalues of Q. (Note that they are all strictly positive.) Note that $\nabla^2 f(x) = Q$ for any x, so f is strictly convex. There therefore exists a unique global minimizer x_* of f in $\mathbb{R}^n$ such that $Qx_* = b$.

Let

$$q(x) = \frac{1}{2}(x - x_*)^T Q(x - x_*) = f(x) + \frac{1}{2}x_*^T Q x_*.$$

So q and f differ by a constant. Therefore it suffices to find the minimizer of q, rather than f. Note that $q(x) \geq 0$ for all x, since Q is positive definite. So we shall study the minimizer x_* of q.

Note that $\nabla f(x) = \nabla q(x) = Qx - b$; let $g(x) = Qx - b$. The method of steepest descent may therefore be written as

$$x_{k+1} = x_k - \alpha_k g(x_k).$$

We would like a formula for the optimal step α_k. Recall that α_k is defined to be the minimizer of the function $f(x_k - \alpha g(x_k))$ over $\alpha \geq 0$. Thus

$$0 = \left.\frac{d}{d\alpha}\right|_{\alpha=\alpha_k} f(x_k - \alpha g(x_k)) = \nabla f(x_k - \alpha_k g(x_k)) \cdot (-g(x_k)).$$

This simplifies to

$$0 = (Q(x_k - \alpha_k g(x_k)) - b) \cdot (-g(x_k)) = -(\underbrace{Qx_k - b}_{=g(x_k)} - \alpha_k Q g(x_k)) \cdot g(x_k)$$

giving

$$0 = -|g(x_k)|^2 + \alpha_k g(x_k)^T Q g(x_k).$$

Therefore

$$\alpha_k = \frac{|g(x_k)|^2}{g(x_k)^T Q g(x_k)}. \qquad (*)$$

Theorem 1.2.

$$q(x_{k+1}) = \left(1 - \frac{|g(x_k)|^4}{(g(x_k)^T Q g(x_k))(g(x_k)^T Q^{-1} g(x_k))}\right) q(x_k)$$

Proof.

$$\begin{aligned}
q(x_{k+1}) &= q(x_k - \alpha_k g(x_k)) \\
&= \frac{1}{2}(x_k - \alpha_k g(x_k) - x_*)^T Q(x_k - \alpha_k g(x_k) - x_*) \\
&= \frac{1}{2}(x_k - x_* - \alpha_k g(x_k))^T Q(x_k - x_* - \alpha_k g(x_k)) \\
&= \frac{1}{2}(x_k - x_*)^T Q(x_k - x_*) - \alpha_k g(x_k)^T Q(x_k - x_*) + \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k) \\
&= q(x_k) - \alpha_k g(x_k)^T Q(x_k - x_*) + \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k),
\end{aligned}$$

implying

$$q(x_k) - q(x_{k+1}) = \alpha_k g(x_k)^T Q(x_k - x_*) - \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k).$$

Dividing by $q(x_k)$ gives

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{\alpha_k g(x_k)^T Q(x_k - x_*) - \frac{1}{2}\alpha_k^2 g(x_k)^T Q g(x_k)}{\frac{1}{2}(x_k - x_*)^T Q(x_k - x_*)}.$$

Let $g_k = g(x_k)$ and $y_k - x_k - x_*$. Then

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{\alpha_k g_k^T Q y_k - \frac{1}{2}\alpha_k^2 g_k^T Q g_k}{\frac{1}{2} y_k^T Q y_k}.$$

Note that $g_k = Qx_k - b = Q(x - x_*) = Qy_k$, so $y_k = Q^{-1} g_k$. The above formula therefore simplifies to

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{2\alpha_k |g_k|^2 - \alpha_k^2 g_k^T Q g_k}{g_k^T Q^{-1} g_k}.$$

Now recall the formula

$$\alpha_k = \frac{|g_k|^2}{g_k^T Q g_k}. \tag{*}$$

This implies that

$$\frac{q(x_k) - q(x_{k+1})}{q(x_k)} = \frac{2\frac{|g_k|^4}{g_k^T Q g_k} - \frac{|g_k|^4}{g_k^T Q g_k}}{g_k^T Q^{-1} g_k} = \frac{|g_k|^4}{(g_k^T Q g_k)(g_K^T Q^{-1} g_k)},$$

proving the theorem. □