CMI - 2020-2021: DS-Analysis

Calculus -11

Some remarks

We begin with an interpretation of the gradient of a function.

For $x,\ y \in \mathbb{R}^n$, it is known that $|x \cdot y| \leq ||x||\ ||y||$; also, equality holds if and only if $y = cx$ for some $c \in \mathbb{R}$.

In particular, if $x,\ y$ are both unit vectors, then $|x \cdot y| \leq 1$, and equality holds if and only if $y = x$, or $y = -x$. So we have $-1 \leq x \cdot y \leq 1$. Also, if $y = x$, then $x \cdot y = 1$, and if $y = -x$, then $x \cdot y = -1$.

Let $f : U \to \mathbb{R}$ be a function with continuous partial derivatives, where $U \subseteq \mathbb{R}^n$ is an open set.

Let $x \in U$. Suppose $\bigtriangledown f(x) \neq 0$. Then $\xi = \bigtriangledown f(x)/(|| \bigtriangledown f(x)||)$ is the unit vector in the direction of the vector $\bigtriangledown f(x)$. Clearly $\bigtriangledown f(x) = || \bigtriangledown f(x)||\ \xi$.

Let $z \in \mathbb{R}^n$ be a unit vector. We know that $D_z f(x) = \bigtriangledown f(x) \cdot z$ is the directional derivative of f in the direction of z at the point x. Therefore we have, as $||z|| = 1$,

$$D_z f(x) = \bigtriangledown f(x) \cdot z = || \bigtriangledown f(x)||\ \xi \cdot z.$$

As $\xi,\ z$ are both unit vectors we get $||D_z f(x)|| \leq || \bigtriangledown f(x)||$ for any unit vector z. Recall that $D_z f(x)$ indicates the rate of change of f in the direction of z. Thus we have the following.

Theorem 1 *Assumptions and notations as above. Let $\bigtriangledown f(x) \neq 0$. Then*

(i) The direction of the vector $\bigtriangledown f(x)$ is the direction of maximal increase of the function f at the point x. Moreover, $|| \bigtriangledown f(x)||$ is the rate of increase in the direction of maximal increase.

(ii) The direction of the vector $- \bigtriangledown f(x)$ is the direction of maximal decrease of the function f at the point x. Moreover, $-|| \bigtriangledown f(x)||$ is the rate of decrease in the direction of maximal decrease.

Next we review some properties of convex functions.

A set $C \subseteq \mathbb{R}^n$ is said to be *convex* if for every $x,\ y \in C$, and every real number $\alpha \in [0,1]$, the point $\alpha x + (1-\alpha)y \in C$.
Examples: Open balls, closed balls, open rectangles, closed rectangles, quadrants, ...

A function $f : C \to \mathbb{R}$, where $C \subseteq \mathbb{R}^n$ is a convex set, is said to be *convex,* if for every $x,\ y \in C$, and every $0 \leq \alpha \leq 1$, we have

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha)f(y).$$

If for every $x,\ y \in C$, and every $0 < \alpha < 1$, we have

$$f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha)f(y),$$

then f is said be *strictly convex.*

A real valued function g defined on a convex set $C \subseteq \mathbb{R}^n$ is said to be *concave* if the function $f = -g$ is convex. The function g is *strictly concave* if $(-g)$ is strictly convex.

Three results on convex functions are given below.

Theorem 2 *Let C be an open convex set. Let $f : C \to \mathbb{R}$ have continuous first order partial derivatives. Then f is convex over C if and only if*

$$f(y) \geq f(x) + \bigtriangledown f(x) \cdot (y-x),$$

for all $x,\ y \in C$.

Theorem 3 *Let C be an open convex set. Let $f : C \to \mathbb{R}$ have all continuous first and second order partial derivatives. Let $H(x)$ denote the Hessian of f at x. Then f is convex over C if and only if for each $x \in C$, all the eigenvalues of $H(x)$ are ≥ 0, that is, if and only if for each $x \in C$, the matrix $H(x)$ is* non-negative definite.

The function f is strictly convex over C if and only if $H(x)$ is strictly positive definite for each $x \in C$.

Theorem 4 *Let $C \subseteq \mathbb{R}^n$ be an open convex set. Let $f : C \to \mathbb{R}$ be a convex function. Then any relative minimum of f is a global minimum of f. Also the set $\Gamma = \{x \in C : f$ attains global minimum $\}$ is a convex set.*

Example: *Quadratic function.* Let Q be an $(n \times n)$ real symmetric strictly positive definite matrix, $b \in \mathbb{R}^n$. Let

$$f(x) = \frac{1}{2} x Q x^t - b x^t, \; x \in \mathbb{R}^n.$$

Then it can be seen that f is a strictly convex function on $\mathbb{R}^n$.

Descent methods are used for iteratively solving unconstrained minimisation problems. Starting from an initial point x_0, one determines, according to a *fixed rule,* a direction of movement; and then moves in that direction to a relative minimum of the objective function on that line. At the new point, a new direction is determined, again by the same rule, and the procedure is repeated. In this way, one gets a sequence $\{x_k\}$ of points, possibly approaching the global minimum point x^*. The basic difference between the various descent methods is the 'rule' by which successive directions are determined.

We will now describe briefly the *method of steepest descent.* This is one of the simplest and theoretically well-studied method. Its technique has basically inspired many other methods. We assume $U = \mathbb{R}^n$ for convenience.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function satisfying:

(i) f is a strictly convex function;

(ii) f has continuous first order partial derivatives;

(iii) f has a global minimum.

By (i) and (iii), note that the global minimum is unique; denote it by x^*. Also, $\bigtriangledown f(x^*) = 0$ at the global minimum x^*.

Recall that we have been regarding $\bigtriangledown f(x)$ as a $n-$dimensional row vector. For certain notational convenience, we set $n-$dimensional column vector $g(x) = \bigtriangledown f(x)^t$. Also, when there is no ambiguity, we may suppress the argument x and, for example, write g_k for $g(x_k) = \bigtriangledown f(x_k)^t$. (Depending on the context, x may be regarded as a row vector or column vector.)

The method of steepest descent is defined by the iterative algorithm

$$x_{k+1} = x_k - \alpha_k g_k, \;\; k = 0, 1, 2, \cdots ,$$

where 'step-size' α_k is a non-negative number (possibly) minimising $f(x_k - \alpha g_k)$. In other words, from the point x_k, we search along the direction of the negative gradient $(-g_k)$ to a minimum point on this line; this minimum point is taken to be x_{k+1}.

Theorem 5 *Assumptions as above. In addition, assume that there is a constant $\beta > 0$ such that for any $x, \; y$,*

$$|| \bigtriangledown f(x) - \bigtriangledown f(y)|| \leq \beta||x - y||.$$

Then the method of steepest descent, taking the step-size as $\alpha_k = (1/\beta)$ for all k, generates a sequence $\{x_k\}$ such that

$$0 \;\leq\; f(x_k) - f(x^*) \;\leq\; \frac{\beta}{2(k+1)}||x_0 - x^*||^2.$$

So $f(x_k) \to f(x^*)$ as $k \to \infty$.

Suppose f is the quadratic function. As Q is a strictly positive definite matrix, there are $0 < a \leq A < \infty$, such that $a \leq \lambda \leq A$, where λ is any eigenvalue of Q.

In this case, α_k minimising $f(x_k - \alpha_k g_k)$ can be explicitly found. It can be proved that $x_k \to x^*$; hence $f(x_k) \to f(x^*)$.

Let $r = (A/a)$; clearly $r \geq 1$. If r is close to 1, the convergence can be very fast.

In general, if the initial point x_0 is close to x^* the convergence rate is faster.

For more information on steepest descent method, various other methods, comparison of methods, etc. there are many books. Two such books are mentioned below.

1. R. Fletcher: *Practical Methods of Optimization.* Second edition. John Wiley, 2004.

2. D. G. Luenberger: *Linear and Non-linear Programming.* Second edition. Springer (India), 2008.