8.1. Gradient Methods

In this chapter we consider a class of search methods for real-valued functions on R^n. These methods use the gradient of the given function.

Recall that a level set of a function $f : R^n \to R$ is the set of points x satisfying $f(x) = c$ for some constant c. Thus, a point $x_0 \in R^n$ is on the level set corresponding to level c if $f(x_0) = c$. In the case of functions of two real variables, $f:\ R^2 \to R$.

The gradient of f at x_0, denoted $\nabla f(x_0)$, if it is not a zero vector, is orthogonal to the tangent vector to an arbitrary smooth curve passing through x_0 on the level set $f(x) = c$. Thus, the direction of maximum rate of increase of a real valued differentiable function at a point is orthogonal to the level set of a function through that point. In other words, the gradient acts in such a direction that for a given small displacement, the function f increases more in the direction of the gradient than in any other direction.

To prove this statement,recall that $\langle \nabla f(x), d \rangle$, $||d|| = 1$, is the rate of increase of f in the direction d at point x.

Notes from Vector Calculus

When f is differentiable, the partial derivatives determine the directional derivatives for all directions v.

$$let \quad F(t) = f(a + tv)$$

Let X be open in R^n; $f : X \subseteq R^n \to R$ is a scalar valued function, and $a \in X$. If $v \in R^n$, is any unit vector, then the directional derivative of f at a is the direction of v, denoted by $D_v f(a)$ *is*

$$D_v f(a) = \ lim_{h \to 0} \frac{f(a + hv) - f(a)}{h}$$

$$D_v f(a) = lim_{t \to 0} \frac{f(a + tv) - f(a)}{t - 0} = lim_{t \to 0} \frac{F(t) - F(0)}{t - 0} = F'(0)$$

That is

$$D_v f(a) \ = \ \frac{d}{dt} f(a + tv) \ \Big|_{t = 0} \tag{2}$$

The significance of equation(2) is that, when f is differentiable at a, we can apply the chain rule to the right-hand side . Let $x(t) \ = a + tv$. Then, by the chain rule,

$$\frac{d}{dt} f(a + tv) = \ Df(x)Dx(t) = Df(x).\, v$$

Evaluation at $t = 0,$ gives

$$D_v f(a) = Df(a).v = \nabla f(a).v \tag{3}$$

The purpose of equation (3) is to emphasize the geometry of the situation. The result above says that the directional derivatives is just the dot product of th gradient and the direction vector v.

Theorem : Let $X \subseteq R^n$ be open and suppose $f: X \to R$ is differentiable at $a \in X$. Then the directional derivatives $D_v f(a)$ exists for all directions (unit vector) $v \in R^n$ & moreover, we have

$$D_v f(a) = \nabla f(a).v$$

Suppose you are traveling in space near the planet Nilrebo and that one of your spaceships instruments measures the external atmospheric pressure on your ship as a function $f(x, y, z)$ of position. Assume, quite reasonably, that this function is differentiable . Then theorem applies and tells us that if you travel from point $a = (a, b, c)$ in the direction of the unit vector $u = ui + vj + kz$, the rate of change of pressure is given by :

$$D_u f(a) \ = \ \nabla f(a).u.$$

Now,we ask the following : In what direction is the pressure increasing the most?
If θ is the angle between u and the gradient vector, then we have :

$$D_u f(a) = ||\nabla f(a)|| \ ||u|| \ cos\theta \ = \ ||\nabla f(a)|| \ cos\theta$$

since u is a unit vector. Because $-1 \leq cos \ \theta \ \leq 1$. We have

$$-||\nabla f(a)|| \leq D_u f(a) \leq ||\nabla f(a)||$$

Moreover, $cos \ \theta = 1,$ when $\theta = 0,$ and $cos \ \theta = -1$ when $\theta = \pi$. Thus we have established the following :

Theorem : The directional derivatives $D_u f(a)$ is maximized, with respect to direction, when u points in the same direction as $\nabla f(a)$ and is minimum when u points in the opposite direction. Furthermore,the maximum and minimum values of

$$D_u f(a) \ are \ ||\nabla f(a)|| \ and \ ||\nabla f(a)||, \ respectively.$$

By the Cauchy-Schwarz inequality:

$$\langle \nabla f(x), d \rangle \leq \|\nabla f(x)\|$$

because $\|d\| = 1$. But if $d = \nabla f(x) / \|\nabla f(x)\|$, *then*

$$\langle \nabla f(x), \frac{\nabla f(x)}{\|\nabla f(x)\|} \rangle = \|\nabla f(x)\|.$$

Thus, the direction in which $\nabla f(x)$ points in the direction of maximum rate of increase of f at x. The direction in which $-\nabla f(x)$ points is the direction of maximum rate of decrease of f at x. Hence, the direction of negative gradient is a good direction to search if we want a function minimizer.

We proceed as follows. Let $x^{(0)}$ be a starting point , and consider the point $x^{(0)} - \alpha \nabla f\left(x^{(0)}\right)$. Then, by Taylor's theorem, we obtain

Taylor's Theorem in one variable : Let X be open in R and suppose $f : X \subseteq R \rightarrow R$ is differentiable up to (at least) order k.
Given $a \in X$, *let*

$$p_k(x) = f(a) + f'(a)(x-a) + \frac{(x-a)^2}{2!} f''(a) + \ldots + \frac{f^{(k)}(a)}{k!}(x-a)^k. \tag{1}$$

Then

$$f(x) = p_k(x) + R_k(x, a),$$

where the reminder term R_k is such that $R_k(x,a) / (x-a)^k \rightarrow 0$ and $x \rightarrow a$.

The polynomial defined in the formula (1) is called the $k-th$
order Taylor Polynomial of f at a.
The essence of Taylor's theorem is this : For x near a, the Taylor polynomial p_k approximates f in the sense that the error R_k involved in making this approximation tends to zero even faster than $(x-a)^k$ does.

$$f(x^{(0)} - \alpha f\left(x^{(0)}\right)$$
$$= f\left(x^{(0)}\right) +$$
$$f\left(x^{(0)}\right) - \alpha \|\nabla f(x^{(0)}\|^2 + o(\alpha),$$

Thus, if $\nabla f\left(x^{(0)}\right) \neq 0$, then for sufficiently small $\alpha > 0$, we have

$$f(x^{(0)} - \alpha \nabla f\left(x^{(0)}\right) < f\left(x^{(0)}\right).$$

This means that the point $x^{(0)} - \alpha \nabla f\left(x^{(0)}\right)$ is an improvement over the point $x^{(0)}$ if we are searching for a minimizer.