GRADIENT METHOD

1 Introduction

Consider an unconstrained optimization in the case of a function $f(x_1, x_2, \cdots, x_n)$. Let $x = (x_1, x_2, \cdots, x_n)$ and $f(x) = f(x_1, x_2, \cdots, x_n)$. We know that f has a minimum at a point $x = x_0$ in a region E(where f is defined) if $f(x) \geq f(x_0), \forall x \in R$.(Similarly for maximum,$f(x) \leq f(x_0), \forall x \in R$).

f is said to have a local minimum at y if $f(x) \geq f(y), \forall x$ in the neighborhood of y that is for all x satisfying $\|x - y\| = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots (x_n - y_n)^2]^{\frac{1}{2}} < r$ where $y = (y_1, y_2, \cdots, y_n)$ and $r > 0$ is sufficiently small.

Similarly, f is said to have a local maximum at y if $f(x) \leq f(y), \forall x$ in the neighborhood of y satisfying $\|x - y\| < r$.

If f is differentiable and has an extremum(maxima or minima) at a point x_0 in the interior of a region R(that is not on the boundary), then the partial derivatives $\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \cdots, \frac{\partial f}{\partial x_n}$, must be zero at x_0. These components of a vector are called the gradient of f and is denoted by *gradf* or ∇f. Thus

$$\nabla f(x_0) = 0. \tag{1.1}$$

A point x_0 at which (1.1) holds is called a stationary point of f.

Condition (1.1) is necessary for an extremum of f at x_0 in the interior of R, but is not sufficient.

For example $F(x) = x^3$ or $f(x_1 x_2)$ has no extremum, however the conditions further $f''(x_0) > 0$ may guarantee a local minima at x_0. However, in practice even solving (1.1) will often be difficult. So, generally the solution is approximated by iteration that is by a search process that starts at some point and moves stepwise to points at which f is smaller (if a minimum of f is wanted) or larger (in case of a maximum).

One such iterative method is called gradient method or method od steepest descent. That is to find a $f(x)$ by repeatedly computing minima of a function $g(t)$ of a single variable t as follows:

Suppose that f has a minimum at y and we start at a point x. Then we look for a minimum of f closest to x along the straight line in the direction of $-\nabla f(x)$, which is the direction of steepest descent(=direction of maximum decrease) of f at x.

Procedure: We determine the value of t and the corresponding point $z(t) = x - t\nabla f(x)$ at which the function $g(t) = f(z(t))$ has a minimum. Hence, the $z(t)$ is the next approximation to y and we repeat this procedure untill we rae close to y starting with an initial approximation.

2 Illustration

Example 1 *Determine a minimum of $f(x) = x_1^2+3x_2^2$ starting from x_0=(6,3).*

(Note that, by inspection, f(x) has a minimum at x=0 and we keep this information beforehand to experiment the steepest descent method for the first time)

Solution: $x = x_1 i + x_2 j$

$$\nabla f(x) = 2x_1 i + 6x_2 j$$

$$z(t) = x - t\nabla f(x) = (1-2t)x_1 i + (1-6t)x_2 j$$

$$g(t) = f(z(t)) = (1-2t)^2 x_1^2 + 3(1-6t)^2 x_2^2$$

Set $g'(t) = 0$ *and solve for* t*, thus we have* $t = \frac{x_1^2+9x_2}{2x_1^2+54x_2}$. *With initial approximation x_0=(6,3), compute x and the following table gives the values of successive iterations and we can see, getting closer t0 actual extremum at the origin. Also sketch the points and see it is descending towards the origin.*

n	x
0	(6,3)
1	(3.484, -0.774)
2	(1.327, 0.664)
3	(0.771, -0.171)
4	(0.294, 0.147)
5	(0.170, -0.038)
6	(0.065, 0.032)

Table of approximations

Example 2 *Iterate 3 steps for extremum of* $f(x) = x_1^2 + 2x_2^2 - x_1 - 6x_2$ *starting from x_0=(0,0) using steepest descent method.*

Example 3 *Iterate 3 steps for extremum of* $f(x) = x_1^2 + 0.1_2^2 + 8x_1 + x_2 + 22.5$ *starting from x_0=(2,-1) using steepest descent method.*

Example 4 *Iterate 3 steps for extremum of* $f(x) = 3x_1^2 + 2x_2^2 - 12x_1 + 16x_2$ *starting from x_0=(1,1) using steepest descent method.*