Steepest Descent Method

Let us consider multivariate non-linear programming problem

$$\text{Min } f(\underline{x}) \quad , \text{ where } \underline{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$$

- Necessary Condition :- $\left.\frac{\partial f(\underline{x})}{\partial x_i}\right|_{X = x^*} = 0$

- Sufficient Condition :- $\left(\frac{\partial f(x)}{\partial x_i \partial x_j}\right) = J|_{x = x^*} > 0$

i.e. Jacobian Matrix is positive definite.

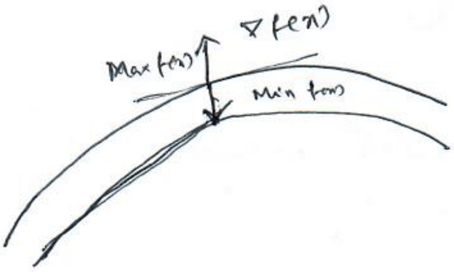
Now, ~~when the~~ When the function $f(x)$ is not differentiable then the above Necessary and sufficient condition will not work. For example,

$$f(x) = \begin{cases} x & , x \geq 0 \\ -x & , x < 0 \end{cases}$$

This function $f(x)$ is not differentiable at $x = 0$ (minima point)

Then we go via gradient method (indirect search Method)

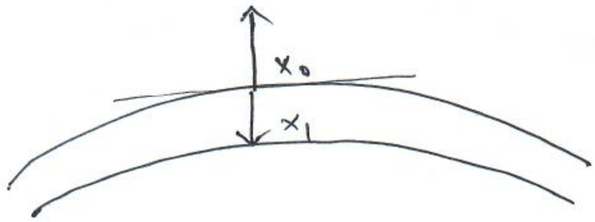
Gradient of $f(x) = \nabla f(x) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$



$\nabla f(x)$ vector – magnitude
> direction.

If the problem is minimization one, it is very clear that we just move in the negative to the gradient direction. The next question comes how far should we walk? Within one iteration it is not possible to reach optimal solution. Therefore, there is a concept of step length, there is a concept of direction in the direct search method and we will discuss the step length and direction for each and every iteration.

Since the problem is minimization we will use steepest descent method, and we will ~~go~~ move through the negative gradient direction. And we will use steepest ascent method for maximization problem, And we will move through the positive gradient direction.



Min $f(x)$
s.t. $x \in [a,b]$

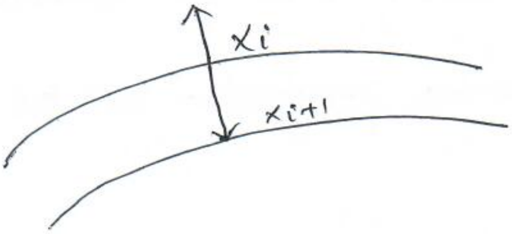
$x_1 = x_0 + \lambda_0 s_0$

$\lambda_0 \rightarrow$ step-length.

$s_0 \rightarrow$ direction (negative)

$x_2 = x_1 + \lambda_1 s_1, \ldots, x_{i+1} = x_i + \lambda_i s_i$

How to find optimal step-length λ_i?



s_i = negative gradient direction at x_i

$= -\nabla f|_{x=x_i}$

$$\frac{d f(x_i + \lambda_i s_i)}{d \lambda_i} = 0$$

(finding the λ_i which gives minimum value at x_{i+1}.

$$\sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial \lambda_i}\bigg|_{x=x_{i+1}} = 0$$

Now, $$\frac{\partial x_{i+1}}{\partial x_i} = \frac{\partial(x_i + \lambda_i s_i)}{\partial \lambda_i} = s_i$$

f $\frac{\partial f}{\partial x_j} \rightarrow$ gradient of f.

or, $\nabla f \Big|_{x=x_{i+1}}^{T} s_i = 0$

In particular, for quadratic function

$$f(x) = \frac{1}{2} X^T A X + B^T X + c$$

We may consider A as positive definite (symmetric) matrix. Then the function is convex, And the optimal solution we get via steepest decent method will be the global optimal solution,

Now, $\nabla f = AX + B$.

$$s_i^T \nabla f \Big|_{X=X_{i+1}} = 0 \qquad [\, X_{i+1} = X_i + \lambda_i S_i \,]$$

or, $s_i^T (A X_{i+1} + B) = 0$

or, $s_i^T (A (X_i + \lambda_i s_i) + B) = 0$

or, $s_i^T ((A X_i + B) + \lambda_i A S_i) = 0$

For steepest decent method $S = -\nabla f = -(AX+B)$

or, $-S_i^T S_i + \lambda_i S_i^T A S_i = 0$

$$\Rightarrow \lambda_i = \frac{s_i^T s}{s_i^T A S_i}$$

$$\boxed{\lambda_i^* = \frac{s_i^T s_i}{s_i^T A s_i}}$$ Optimal step-length.

Algorithm for Steepest Descent Method :-

Step-1 : Start with initial starting point $x_0 \in \mathbb{R}^n$

Step-2 : Set $i = 1$

Find search direction

$s_i = -\nabla f|_{x=x_i}$ for Minimization problem

$s_i = \nabla f|_{x=x_i}$ for Maximization problem.

Step-3 : $\lambda_i^* = \dfrac{s_i^T s_i}{s_i^T A s_i}$

Step-4 : Formulate new approximation for optimal solution

$x_{i+1} = x_i + \lambda_i^* s_i$

Step-5 : Check for optimality, stop otherwith, go to step-2 and consider $i = i+1$.

(i) $\left| \dfrac{f(x_{i+1}) - f(x_i)}{f(x_i)} \right| < \varepsilon$

(ii) $\|\nabla f\| \big|_{x=x_i} < \varepsilon$

(iii) $|x_{i+1} - x_i| < \varepsilon$

Example :- Minimize $f(x_1, x_2) = 2x_1^2 + 2x_1x_2 + x_2^2 + x_1 - x_2$

Consider the starting point $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$\Rightarrow$ Here, $X_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

$$f(x) = \frac{1}{2} x^T A x + B^T x + c$$

$$A = \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$\nabla f = \begin{cases} 4x_1 + 2x_2 + 1 = \frac{\partial f}{\partial x_1} \\ 2x_1 + 2x_2 - 1 = \frac{\partial f}{\partial x_2} \end{cases}$$

$$\nabla f \big|_{X = X_0} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

$$S_0 = -\nabla f \big|_{X = X_0} = \begin{pmatrix} -1 \\ +1 \end{pmatrix}$$

$$X_1 = X_0 + \lambda_0^* S_0$$

$$\lambda_0^* = \frac{S_0^T S_0}{S_0^T A S_0} = \frac{(-1 \ \ 1)\begin{pmatrix} -1 \\ 1 \end{pmatrix}}{(-1 \ \ 1)\begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix}}$$

$$= \frac{2}{(-2 \cdot 0)\begin{pmatrix} -1 \\ 1 \end{pmatrix}} = 1$$

$$X_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ +1 \end{pmatrix}$$

$\nabla f \big|_{X = X_0} \neq 0$ (Move further)

$\therefore X_2 = X_1 + \lambda_1^* S_1$

Now, $S_1 = -\nabla f \big|_{X = X_1} = -\begin{pmatrix} 4x_1 + 2x_2 + 1 \\ 2x_1 + 2x_2 - 1 \end{pmatrix}_{X = \begin{pmatrix} -1 \\ +1 \end{pmatrix}}$

$$= \begin{pmatrix} -1 \\ -1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$\lambda_1^* = \frac{S_1^T S_1}{S_1^T A S_1} = \frac{(+1 \; +1) \begin{pmatrix} +1 \\ 1 \end{pmatrix}}{(+1 \; +1) \begin{pmatrix} 4 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} +1 \\ +1 \end{pmatrix}} = \frac{2}{(6 \; 4) \begin{pmatrix} 1 \\ 1 \end{pmatrix}} = \frac{2}{10}$$

$$= \frac{1}{5}$$

Therefore,

$$X_2 = \begin{pmatrix} -1 \\ 1 \end{pmatrix} + \frac{1}{5} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} -4/5 \\ 6/5 \end{pmatrix}$$

Further,

$X_3 = X_2 + \lambda_2^* S_2 \; . \; . \; . \; . \; .$

Iteration NO	x_i^T	$\nabla f\|_{x=x_i}$	s_i^T	λ_i^*	$x_{i+1} = x_i + \lambda_i s_i$
0	(0, 0)	(1 −1)	(−1 1)	1	(−1 1)
1.	(−1 1)	(−1 −1)	(1 1)	0.2	(−0.8. 1.2)
2.	(−0.8. 1.2)	(0.2 −0.2)	(−0.2, 0.2)	1	(−1 1.4)
3.	(−1 1.4)	(−0.2 −0.2)	(0.2, 0.2)	0.2	(0.96 . 1.44)
4.	(0.96 . 1.44)	(7.72 3.8)	(−7.72 −3.8)	0.193	(−0.53 0.71)
5.	(−0.53 0.71)	(0.31 −0.64)	(−0.31 0.64)	1.24	(−0.9144 1.497)
6.	(−0.9144 1.497)	(0.337 0.166)	(−0.337 −0.166)	0.192	(−0.979 1.465)
7.	(−0.979 1.465)	(0.01366 −0.0277)	(−0.01366 0.0277)	1.24	(−0.996 1.4998)
8.	(−0.996 1.4998)	(0.0147 0.007)	(−0.0147 −0.007)	0.192	(−0.999 1.49)

Exercise:- Use the method of steepest ascent to approximate the solution to

$$\text{Max } z = -(x_1-3)^2 - (x_2-2)^2 \equiv f(x_1, x_2)$$

s.t. $(x_1, x_2) \in \mathbb{R}^2$, Initial point $x_0 = \binom{1}{1}$.

Suppose, X_0, ~~λ_0~~ given

then $x_1 = X_0 + \lambda_0^* s_0$ $\qquad S_0 = \nabla f|_{x=x_0}$

$= X_0 + \lambda_0^* \nabla f|_{x=x_0}$.

$\therefore$ $\text{Max } f(x)$
s.t. $x \geq 0$.

$$\Rightarrow \left.\begin{array}{l} \text{Max } f\left(X_0 + \lambda_0^* \nabla f(x_0)\right) \\ \text{s.t. } \lambda_0^* \geq 0 \end{array}\right\} \text{——— } (*)$$

Solve this (*) NLP by the method discussed earlier (Necessary-sufficient condition or Golden section method etc.)

- If $\|\nabla f(x_1)\|$ is small (say, <0.01) we may terminate the algorithm. ~~[illegible]~~ with the knowledge that x_1 is near a stationary point x^* having
$\nabla f(x^*) = 0$.