Sufficient condition for local minimizer

Let f be a twice continuously differentiable function from $\mathbb{R}^n$ to $\mathbb{R}$. Suppose that $x^* \in \mathbb{R}^n$ satisfies

(a) $\nabla f(x^*) = 0$

(b) $H(x^*)$ is positive definite

Then x^* is a local minimizer of f.

Example Let $f(x) = x_1^2 + x_2^2$. Find points satisfying FONC. Check that these points are local minimum point or not.

Sol^n

$$\nabla f(x) = (2x_1, 2x_2)$$

now $\nabla f(x) = 0 \Rightarrow x_1 = x_2 = 0$

$\therefore$ $(0,0)$ satisfy the FONC.

$$H(x) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad \therefore H(0,0) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Now let $d \in \mathbb{R}^n$, then $d^T H(0,0)\, d$

$$\Rightarrow [d_1 \quad d_2] \begin{bmatrix} 2d_1 \\ 2d_2 \end{bmatrix} \Rightarrow 2d_1^2 + 2d_2^2 > 0 \text{ if } d \neq 0$$

$\therefore$ $H(0,0)$ is a positive definite matrix.

Hence by sufficient condition, $x^* = (0,0)$ is a local minimizer

Prob 2 $\quad \nabla f(x^*) = 0 \implies x^* = (1,1)$

From now onwards we consider the unconstrained optimization problem $f: \mathbb{R}^n \to \mathbb{R}$ and the problem is

$$\min_{x \in \mathbb{R}^n} f(x)$$

Descent direction Let $x \in \mathbb{R}^n$. If there exists a direction $d \in \mathbb{R}^n$ and $\varepsilon > 0$ such that

$$\boxed{f(x + \alpha d) < f(x)} \quad \forall \alpha \in (0, \varepsilon),$$

then d is said to be a descent direction.

Lemma Let $f: \mathbb{R}^n \to \mathbb{R}$ be differentiable and for any $d \in \mathbb{R}^n$, $\boxed{\nabla f(x) \cdot d < 0}$, then d is a $\boxed{\text{descent direction.}}$?

Note $\boxed{\nabla f(x)^T d}$ with $\boxed{\|d\| = 1}$, is the rate of increase of f along the direction d at the point x.

$\nabla f(x)^T d = \langle \nabla f(x), d \rangle \le \|\nabla f(x)\| \, \|d\|$ [By Cauchy-Schwarz inequ]

$\le \|\nabla f(x)\|$ [$\because \|d\| = 1$] — 1

But if we choose $\boxed{d = \dfrac{\nabla f(x)}{\|\nabla f(x)\|}}$ $\qquad -\nabla f(x)$

then $\nabla f(x)^T d = \langle \nabla f(x), d \rangle = \|\nabla f(x)\| \dfrac{\nabla f(x)}{\|\nabla f(x)\|} = \|\nabla f(x)\|$

Thus the direction in which $\boxed{\nabla f(x)}$ indicates is the direction of $\boxed{\text{maximum rate of increase.}}$

The direction in which $\boxed{-\nabla f(x)}$ indicates is the direction of $\boxed{\text{maximum rate}}$ of $\boxed{\text{decrease.}}$

Structure of optimization method

Usually an optimization method is an iterative method for finding the optimum point. The basic idea is given an initial point $x^{(0)} \in \mathbb{R}^n$, one have to generate the iterative seq^n $\{x_k\}_{k=0}^{\infty}$ by means of some iterative rule which converges to the minimum point.

Algorithm

1. Initialize $x^{(0)}$, $k=0$
2. stopping criteria
3. If the stopping criteria is not satisfied at x^k

   (a) Find $x^{(k+1)}$ such that $f(x^{k+1}) < f(x^k)$

   (b) $k = k+1$

end if output $x^* = x^k$ is the local minimum point.

Stopping criteria :- In general, the most useful stopping criteria is $\boxed{\|\nabla f(x^k)\| < \varepsilon}$ where $\varepsilon > 0$ is the tolerance

?

Gradient Descent method

Since $\boxed{-\nabla f(x)}$ gives the direction in which the rate of decrease for the function is $\boxed{\text{maximum}}$, hence the direction of negative gradient is a good direction for minimizing a function.

To formulate the algorithm, suppose we are given a point $x^{(k)}$. To find the next point x^{k+1}, we start moving by an amount $-\alpha_k \nabla f(x^k)$, where α_k is very small positive scalar. Thus the iterative algorithm is

?

$$\boxed{x^{k+1} = x^{(k)} - \alpha_k \nabla f(x^k)}$$

we refer this as a gradient descent method or gradient method.

The method of $\boxed{\text{steepest descent}}$

The method of steepest descent is a gradient method where the step size α_k is choosen to achieve the maximum amount of decrease of the cost function at each individual step. Specially, α_k is chosen to minimize $\boxed{\phi_k(\alpha) \triangleq f(x^k - \alpha \nabla f(x^k))}$.

In otherwords $\alpha_k = \underset{\alpha \geq 0}{\arg\min}\, f(x^k - \alpha \nabla f(x^k))$

<u>Example</u> Apply method of steepest descent to the function $f(x,y) = 4x^2 - 4xy + 2y^2$ with the initial point $x^{(0)} = (2,3)$

<u>Sol^n</u> $\nabla f(x,y) = (8x - 4y,\ 4y - 4x)$

$\nabla f(x^0) = (4, 4)$

$$\phi_0(\alpha) = f(x^0 - \alpha \nabla f(x^0))$$

$$= f(2 - 4\alpha,\ 3 - 4\alpha)$$

$$\Rightarrow \phi_0'(\alpha) = \nabla f(2-4\alpha, 3-4\alpha) \cdot (-4,-4)$$

$$= 64\alpha - 32$$

hence $\boxed{\phi_0'(\alpha) = 0} \Rightarrow \alpha = \frac{1}{2}$. Again, $\boxed{\phi_0''(\alpha) = 64 > 0}$

$\alpha = \frac{1}{2}$ giving the $\boxed{\text{minimum value of } \phi_0}$ over all $\boxed{\alpha \geq 0}$.

$$x^{(1)} = x^{(0)} - \frac{1}{2} \nabla f(x^0)$$

$$= (0, 1)$$

$\nabla f(x^{(1)}) = (-4, 4)$

$$\phi_1(\alpha) = f(x^{(1)} - \alpha \nabla f(x^1))$$

$$= f(4\alpha,\ 1 - 4\alpha)$$

$$\phi_1'(\alpha) = \nabla f(4\alpha, (1-4\alpha)) \cdot (4, -4)$$

$$= 320\alpha - 32$$

$$\phi_1'(\alpha) = 0 \Rightarrow \alpha = \frac{1}{10}$$

$$\therefore \quad x^{(2)} = x^{(1)} - \frac{1}{10} \nabla f(x^1)$$

$$= (0, 1) - \frac{1}{10}(-4, 4)$$

$$= \left(\frac{2}{5}, \frac{3}{5}\right)$$

Repeating this process, we get $x^{(3)} = (0, \frac{2}{10})$.
we can see that the method of steepest descent produces a seq^n that is converging towards the minimum point $\boxed{x^* = (0, 0)}$ of the function.

Steepest descent algorithm

Step 0 :- Let $0 < \varepsilon < 1$ be the tolerance on stopping threshold. Given initial point $x^{(0)}$. Let $k = 0$

Step 1 If $\|\nabla f(x^k)\| \leq \varepsilon$ then stop. Otherwise choose $\boxed{d^k = -\nabla f(x^k)}$

Step 2 Find the step size α_k such that

$$f(x^k + \alpha_k d^k) = \min_{\alpha \geq 0} f(x^k + \alpha d^k)$$

Step 3 $\boxed{x^{(k+1)} = x^k + \alpha_k d^k}$

Step 4 : $k = k+1$, return to step 1
Output $x^k = x^*$.