Algorithms in Optimization

Zheng Li

March 19, 2017

1 Basic

1.1 Computing Gradient

The numerical method for computing the gradient of a specific function at x is as follow:

$$f^{(1)}(x) = \frac{f(x+h) - f(x-h)}{2h}$$

where h is the step length.

now, we use Taylor series to validate it, and therefore

$$\begin{aligned}\frac{f(x+h) - f(x-h)}{2h} &= \frac{f(x) + f^{(1)}(x)h - f(x) + f^{(1)}(x)h + o(h^2)}{2h} \\ &= f^{(1)}(x) + o(h^2)\end{aligned}$$

1.2 Computing Hessian Matrix

Hessian matrix is widely used in optimization algorithms, this will be revealed later. In fact, computing the 2nd order partial derivatives is crucial for obtaining Hessian matrix. For a 2-dimensional function, its Hessian matrix is defined as:

$$\bigtriangledown^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} \end{bmatrix} \tag{1}$$

It's obvious that $\frac{\partial^2 f}{\partial x_1^2}$ is not difficult to compute, since

$$\frac{\partial^2 f}{\partial x_1^2} = \frac{f(x_1+h, x_2) + f(x_1-h, x_2) - 2f(x_1, x_2)}{h^2}$$

This equation can be generated by recursively using method aforementioned in section 1.1, and here we briefly use Taylor series to prove it.

$$f(x+h, y+k) = \sum_{m=0}^{n} \frac{1}{m!}(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y})^m f(x,y) + R_n$$

where,

$$R_n = \frac{1}{(n+1)!}(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y})^{n+1} f(x+\theta h, y+\theta k) \quad (0 < \theta < 1)$$

and $h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}$ is a partial derivative operator. Thus,

$$\begin{aligned}f(x_1+h, x_2) + f(x_1-h, x_2) - 2f(x_1, x_2) &= f(x_1, x_2) + h\frac{\partial}{\partial x_1} f(x_1, x_2) + \frac{1}{2}(h\frac{\partial}{\partial x_1})^2 f(x_1, x_2) \\ &\quad + f(x_1, x_2) - h\frac{\partial}{\partial x_1} f(x_1, x_2) + \frac{1}{2}(h\frac{\partial}{\partial x_1})^2 f(x_1, x_2) - 2f(x_1, x_2) \\ &= h^2 \frac{\partial^2 f}{\partial x_1^2}\end{aligned}$$

Divide both sides by h^2, we can get:

$$\frac{f(x_1+h,x_2)+f(x_1-h,x_2)-2f(x_1,x_2)}{h^2}=\frac{\partial^2 f}{\partial x_1^2}$$

Now, the problem is: how to calculate $\frac{\partial^2 f}{\partial x_1 \partial x_2}$ for a specific function at (x_1,x_2). Through the observation, we can sum up the following two equation side by side,

$$f(x_1+h,x_2+k)=f(x_1,x_2)+(h\frac{\partial}{\partial x_1}+k\frac{\partial}{\partial x_2})f(x_1,x_2)+\frac{1}{2}(h^2\frac{\partial^2}{\partial x_1^2}+2kh\frac{\partial^2}{\partial x_1 \partial x_2}+k^2\frac{\partial^2}{\partial x_2^2})f(x_1,x_2)$$

$$f(x_1-h,x_2-k)=f(x_1,x_2)-(h\frac{\partial}{\partial x_1}+k\frac{\partial}{\partial x_2})f(x_1,x_2)+\frac{1}{2}(h^2\frac{\partial^2}{\partial x_1^2}+2kh\frac{\partial^2}{\partial x_1 \partial x_2}+k^2\frac{\partial^2}{\partial x_2^2})f(x_1,x_2)$$

hence,

$$\begin{aligned}f(x_1+h,x_2+k)+f(x_1-h,x_2-k)&=2f(x_1,x_2)+(h^2\frac{\partial^2}{\partial x_1^2}+2kh\frac{\partial^2}{\partial x_1 \partial x_2}+k^2\frac{\partial^2}{\partial x_2^2})f(x_1,x_2)\\&=2f(x_1,x_2)+h^2\frac{\partial^2 f}{\partial x_1^2}+2kh\frac{\partial^2 f}{\partial x_1 \partial x_2}+k^2\frac{\partial^2 f}{\partial x_2^2}\end{aligned}$$

We can use the aforementioned equation to convert $\frac{\partial^2 f}{\partial x_1^2}$ and $\frac{\partial^2 f}{\partial x_2^2}$,

$$\begin{aligned}f(x_1+h,x_2+k)+f(x_1-h,x_2-k)&=2f(x_1,x_2)+h^2\frac{\partial^2 f}{\partial x_1^2}+2kh\frac{\partial^2 f}{\partial x_1 \partial x_2}+k^2\frac{\partial^2 f}{\partial x_2^2}\\&=2f(x_1,x_2)+f(x_1+h,x_2)+f(x_1-h,x_2)-2f(x_1,x_2)\\&\quad+2kh\frac{\partial^2 f}{\partial x_1 \partial x_2}+f(x_1,x_2+h)+f(x_1,x_2-h)-2f(x_1,x_2)\end{aligned}$$

Because what we want is $\frac{\partial^2 f}{\partial x_1 \partial x_2}$,

$$\begin{aligned}2kh\frac{\partial^2 f}{\partial x_1 \partial x_2}&=f(x_1+h,x_2+k)+f(x_1-h,x_2-k)-2f(x_1,x_2)\\&\quad-[f(x_1+h,x_2)+f(x_1-h,x_2)-2f(x_1,x_2)]\\&\quad-[f(x_1,x_2+h)+f(x_1,x_2-h)-2f(x_1,x_2)]\\&=f(x_1+h,x_2+k)+f(x_1-h,x_2-k)-f(x_1+h,x_2)-f(x_1-h,x_2)\\&\quad-f(x_1,x_2+h)-f(x_1,x_2-h)+2f(x_1,x_2)\end{aligned}$$

Finally,

$$\frac{\partial^2 f}{\partial x_1 \partial x_2}=\frac{F(x_1,x_2,h,k)}{2kh}$$

where

$$\begin{aligned}F(x_1,x_2,h,k)&=f(x_1+h,x_2+k)+f(x_1-h,x_2-k)-f(x_1+h,x_2)-f(x_1-h,x_2)\\&\quad-f(x_1,x_2+h)-f(x_1,x_2-h)+2f(x_1,x_2)\end{aligned}$$

Usually, we let $k=h$ when this equation is implemented.

From the above deduction we can know, it's possible to numerically compute the Hessian matrix of a specific function with two variables at point (x_1,x_2). However, this equation can be generalized into multi-dimensions easily, just let other irrelevant variables fixed.

Therefore, we can make the conclusion, for a function with multi-variables, i.e., $f(x_1,x_2,...,x_n)$. If we regard x_i and x_j $(i,j\in N)$ is the key variables for computing 2nd order partial derivative of function f , we denote $\tilde{x}$ is the set all of irrelevant variables. When $i=j$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}=\frac{f(x_i+h,\tilde{x})+f(x_i-h,\tilde{x})-2f(x_i,\tilde{x})}{h^2}$$

when $i \neq j$ $(i, j \in N)$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{F(x_i, x_j, \tilde{x}, h)}{2h^2}$$

where,

$$\begin{aligned} F(x_i, x_j, \tilde{x}, h) = & f(x_i + h, x_j + h, \tilde{x}) + f(x_i - h, x_j - h, \tilde{x}) - f(x_i + h, x_j, \tilde{x}) - f(x_i - h, x_j, \tilde{x}) \\ & - f(x_i, x_j + h, \tilde{x}) - f(x_i, x_j - h, \tilde{x}) + 2f(x_i, x_j, \tilde{x}) \end{aligned}$$

and h is the step length small enough.