

Chapter 09.04

Multidimensional Gradient Method

After reading this chapter, you should be able to:

1. *Understand how multi-dimensional gradient methods are different from direct search methods*
2. *Understand the use of first and second derivatives in multi-dimensions*
3. *Understand how the steepest ascent/descent method works*
4. *Solve multi-dimensional optimization problems using the steepest ascent/descent method*

How do gradient methods differ from direct search methods in multi-dimensional optimization?

The difference between gradient and direct search methods in multi-dimensional optimization is similar to the difference between these approaches in one-dimensional optimization. Direct search methods are useful when the derivative of the optimization function is not available to effectively guide the search for the optimum. While direct search methods explore the parameter space in a systematic manner, they are not computationally very efficient. On the other hand, gradient methods use information from the derivatives of the optimization function to more effectively guide the search and find optimum solutions much quicker.

Newton's Method

When Newton's Method

(http://numericalmethods.eng.usf.edu/topics/opt_newtons_method.html) was introduced as a one-dimensional optimization method, we discussed the use of the first and second derivative of the optimization function as sources of information to determine if we have reached an optimal point (where the value of the first derivative is zero). If that optimal point is a maximum, the second derivative is negative. If the point is a minimum, the second derivative is positive.

What are Gradients and Hessians and how are Gradients and Hessians used in multi-dimensional optimization?

Gradients and Hessians describe the first and second derivatives of functions, respectively in multiple dimensions and are used frequently in various gradient methods for multi-dimensional optimization. We describe these two concepts of gradients and Hessians next.

Gradient:

The gradient is a vector operator denoted by ∇ (referred to as “del”) which, when applied to a function f , represents its directional derivatives. For example, consider a two dimensional function $f(x, y)$ which shows elevation above sea level at points x and y . If you wanted to move in the direction that would gain you the most elevation, this direction could be defined along a direction h which forms an angle θ with the x -axis. For an illustration of this, see Figure 1. The elevation along this new axis can be described by a new function $g(h)$ where your current location is the origin of the new coordinate axis or $h=0$. The slope in this direction can be calculated by taking the derivative of the new function $g(h)$ at this point, namely $g'(0)$. The slope is then calculated by

$$g'(0) = \frac{\partial f}{\partial x} \cos \theta + \frac{\partial f}{\partial y} \sin \theta$$

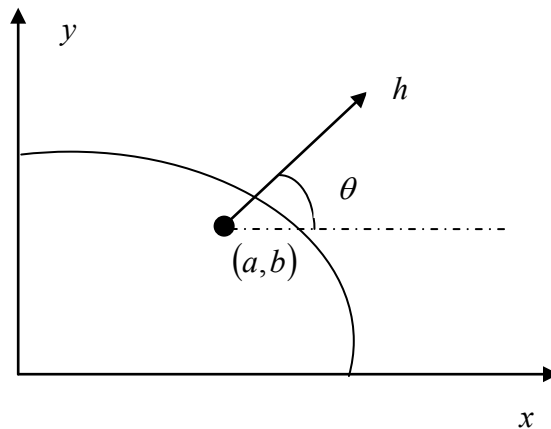


Figure 1. Determining elevation along a new axis

The gradient is a special case where the direction of the vector gains the most elevation, or has the steepest ascent. If the goal was to decrease elevation, then this would be termed as the steepest descent.

The gradient of $f(x, y)$ or ∇f is the vector pointing in the direction of the steepest slope at that point. The gradient is calculated by

$$\nabla f = \frac{\partial f}{\partial x} \mathbf{i} + \frac{\partial f}{\partial y} \mathbf{j} \quad (1)$$

Example 1

Calculate the gradient to determine the direction of the steepest slope at point (2, 1) for the function $f(x, y) = x^2 y^2$.

Solution

To calculate the gradient; the partial derivatives must be evaluated as

$$\begin{aligned}\frac{\partial f}{\partial x} &= 2xy^2 \\ &= 2(2)(1)^2 \\ &= 4 \\ \frac{\partial f}{\partial y} &= 2x^2 y \\ &= 2(2)^2(1) \\ &= 8\end{aligned}$$

which are used to determine the gradient at point (2,1) as

$$\nabla f = 4i + 8j$$

Traveling along this direction, we would gain elevation equal to the magnitude of the gradient which is $\nabla f = \sqrt{4^2 + 8^2} = 8.94$. Note that there is no other direction along which we can move to increase the slope.

Hessians:

The Hessian matrix or just the Hessian, is the Jacobian Matrix of the second-order partial derivatives of a function. For example, in a two dimensional function the Hessian matrix is simply

$$H = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (2)$$

The determinant of the Hessian matrix is also referred to as the Hessian.

Example 2

Calculate the Hessian at point (2, 1) for the function $f(x, y) = x^2 y^2$.

Solution

To calculate the Hessian, we would need to calculate

$$\begin{aligned}\frac{\partial^2 f}{\partial^2 x^2} &= 2y^2 \\ &= 2(1)^2 \\ &= 2\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 f}{\partial y^2} &= 2x^2 \\
&= 2(2)^2 \\
&= 8 \\
\frac{\partial^2 f}{\partial x \partial y} &= \frac{\partial^2 f}{\partial y \partial x} \\
&= 4xy \\
&= 4(2)(1) \\
&= 8
\end{aligned}$$

and the resulting Hessian matrix is

$$H = \begin{bmatrix} 2 & 8 \\ 8 & 8 \end{bmatrix}$$

Based on your knowledge of how second derivatives are used in one-dimensional optimization, you may guess that the value of the second derivatives in multi-dimensional optimization will tell us if we are at a maxima or minima. In multi-dimensional optimization, the Hessian of the optimization function contains the information of the second derivatives of the function, and is used to make such a determination. The determinant of the Hessian matrix denoted by $|H|$ can have three cases:

1. If $|H| > 0$ and $\partial^2 f / \partial x^2 > 0$ then $f(x, y)$ is a local minimum.
2. If $|H| > 0$ and $\partial^2 f / \partial x^2 < 0$ then $f(x, y)$ is a local maximum.
3. If $|H| < 0$ then $f(x, y)$ is a saddle point.

Referring to Example 2, since $|H| = -48 < 0$, then point (2, 1) is a saddle point.

Hessians are also used in determining search trajectories in more advanced multi-dimensional gradient search techniques such as the Marquardt Method which is beyond the scope of this module.

What is the Steepest Ascent (or Descent) method and how does it work?

In any multi-dimensional optimization algorithm, there are two key questions to be asked when searching for an optimal solution. The first one is about the direction of travel. The concept of gradients provides the answers to this question, at least in the short term. The second question is how long one should pursue a solution in this direction before reevaluating an alternative direction. If a re-evaluation strategy is executed too often, it increases computational costs. If it is executed too infrequently, one may end up pursuing a direction that may take the search away from the optimal solution. The steepest ascent algorithm proposes a simple solution to the second question by arbitrarily choosing a step size h . The special case where $h = h^*$ is referred to as the optimal steepest descent where h^* brings us to the local maximum along the direction of the gradient. Consider the following

example which illustrates the application of the optimal steepest ascent method to a multi-dimensional optimization problem.

Example 3

Determine the minimum of the function $f(x, y) = x^2 + y^2 + 2x + 4$. Use the point $(2, 1)$ as the initial estimate of the optimal solution.

Solution

Iteration 1:

To calculate the gradient; the partial derivatives must be evaluated as

$$\frac{\partial f}{\partial x} = 2x + 2 = 2(2) + 2 = 6$$

$$\frac{\partial f}{\partial y} = 2y = 2(1) = 2$$

which are used to determine the gradient at point $(2, 1)$ as

$$\nabla f = 6i + 2j$$

Now the function $f(x, y)$ can be expressed along the direction of gradient as

$$f\left(x_0 + \frac{\partial f}{\partial x} h, y_0 + \frac{\partial f}{\partial y} h\right) = f(2 + 6h, 1 + 2h) = (2 + 6h)^2 + (1 + 2h)^2 + 2(2 + 6h) + 4$$

Multiplying out the terms we obtain the one dimensional function along the gradient as

$$g(h) = 40h^2 + 40h + 13$$

This is a simple function and it is easy to determine $h^* = -0.5$ by taking the first derivative and solving for its roots. This means that traveling a step size of $h = -0.5$ along the gradient reaches a minimum value for the function in this direction. These values are substituted back to calculate a new value for x and y as follows:

$$x = 2 + 6(-0.5) = -1$$

$$y = 1 + 2(-0.5) = 0$$

Calculating the new values of x and y concludes the first iteration. Note that $f(-1, 0) = 3$ is less than $f(2, 1) = 13$ which indicates a move in the right direction.

Iteration 2:

The new initial point is $(-1, 0)$. We calculate the gradient at this point as

$$\frac{\partial f}{\partial x} = 2x + 2 = 2(-1) + 2 = 0$$

$$\frac{\partial f}{\partial y} = 2y = 2(0) = 0$$

which are used to determine the gradient at point $(-1, 0)$ as

$$\nabla f = 0i + 0j$$

This indicates that the current location is a local optimum and no improvement can be gained by moving in any direction. To ensure that we have reached a minimum, we can calculate the

Hessian of the function. To determine the Hessian, the second partial derivatives are determined and evaluated as follows

$$\frac{\partial^2 f}{\partial x^2} = 2$$

$$\frac{\partial^2 f}{\partial y^2} = 2$$

$$\frac{\partial^2 f}{\partial y \partial x} = 0$$

The resulting Hessian matrix and its determinant are

$$H = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \quad |H| = 4 - 0^2 = 4$$

Since $|H| > 0$ and $\partial^2 f / \partial^2 x^2 > 0$ then $f(-1,0)$ is a local minimum.

OPTIMIZATION

Topic	Multidimensional Gradient Method
Summary	Textbook notes for the multidimensional gradient method
Major	All engineering majors
Authors	Ali Yalcin
Date	December 22, 2012
Web Site	http://numericalmethods.eng.usf.edu
