Method of Steepest Descent and its Applications

Xu Wang[1]

[1]*Department of Engineering, University of Tennessee, Knoxville, TN 37996*

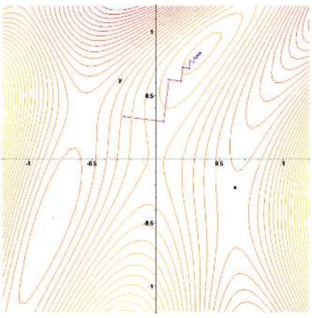
(Dated: November 25, 2008)

The method of steepest descent is also known as The Gradient Descent, which is basically an optimization algorithm to find the local minimum of a function. It is a method that's widely popular among mathematicians and physicists due to its easy concept and relatively small work steps.This paper introduces the basic concept of the method of steepest descent, the advantage and disadvantage of using such method, and some of its applications.

I. THE METHOD

The method of steepest descent is the simplest of the gradient methods. Imagine that there's a function $F(x)$, which can be defined and differentiable within a given boundary, so the direction it decreases the fastest would be the negative gradient of $F(x)$. To find the local minimum of $F(x)$, The Method of The Steepest Descent is employed, where it uses a zig-zag like path from an arbitrary point X_0 and gradually slide down the gradient, until it converges to the actual point of minimum.

FIG. 1: The Method of Steepest Descent Approaches the Local Minimum in a zig-zag path, and the next search direction would be orthogonal to the next



To put this step into a function, one can get:

$$x_{k+1} = x_k - \lambda_k \nabla F(x_k) = x_k - \lambda_k g(x_k) \tag{1}$$

In the above iterative form, the term $g(x_k)$ is the gradient at a given point. It is obvious that in order to find the point where $F(x)$ is a minimum, the directional derivative at that point would be zero, and in this case, the directional derivative is given by:

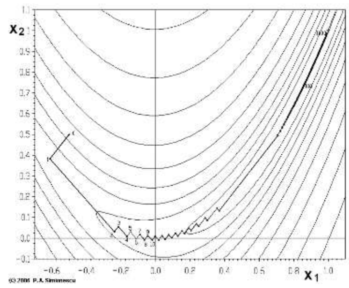
$$d/d\lambda_k F(x_{k+1}) = \nabla F x_{k+1} T d/d\lambda_k x_{k+1} = -\nabla F(x_{k+1}^T) g(x_k) \tag{2}$$

By setting the above equation equal to zero, it is clear that the term λ_k should be used as the step taken in the gradient direction so that $\nabla F(x_{k+1})$ and $g(x_k)$ become orthogonal. By taking steps in this direction of the negative gradient, this essentially is a minimization problem along a line for different values of λ_k.

It is not hard to see why the method of steepest descent is so popular among many mathematicians: it is very simple, easy to use, and each repetition is fast. But the biggest advantage of this method lies in the fact that it is guaranteed to find the minimum through numerous times of iterations as long as it exists.

However, this method also has some big flaws:If it is used on a badly scaled system, it will end up going through an infinite number of iterations before locating the minimum, and since each of steps taken during iterations are extremely small, thus the convergence speed is pretty slow, this process can literally take forever! Although a larger step size will increase the convergence speed, but it could also result in an estimate with large error.

FIG. 2: In a case of quadratic function with a long, narrow valley, each step size decreases as it keeps crossing and recrossing the valley to locate the minimum



II. APPLICATIONS

Since the method is very easy to use, it has various applications in mathematics and physics. One of most frequent employment of the method of steepest descent is to use it in order to solve complex integrals, for example: suppose that an integral is defined as:

$$I = \int_a^b e^{Nf(x)} dx \tag{3}$$

Where $f(x)$ is a function and N is a number of large value. It is obvious that we can't evaluate this integral exactly, however, since N is a large number, we can obtain a very accurate approximate value. Between a to b, the integral is dominated by the range of x around the maximum of $f(x)$. This is because even if $f(x)$ is only a little bigger at its maximum than at other values of x, $e^{Nf(x)}$ would be much larger at its maximum than at any other point since N is a large number. So to get an accurate estimate for this integral, we need to approximate the function by its form near the values of the maximum.

So let's say that maximum is at a point x_0, then we can write:

$$f(x) = f(x_0) - 1/2|f''(x_0)|\delta x^2 + ... \tag{4}$$

Where $\delta x = x - x_0$. If $f''(x_0) < 0$ at a maximum, then we can write the integral in terms of $|f''(x_0)|$ because it is positive and thus more convenient. We can also replace $deltax$ with variable t for simplicity. The integral then simplified to:

$$I = \int_{-\infty}^{\infty} e^{-N|f''(x_0)|t^2/2} dx \tag{5}$$

Notice that the range of integration is also changed to between ∞ to $-\infty$ so that it will take more iteration steps to converge and thus making the error value as small as possible. Moreover, it is very easy to determine the value of a simple integral such as $\int_{-\infty}^{\infty} e^{-at^2/2} dt$, which is simply $\sqrt{2\pi/a}$. So with the stated integral value, we can see that the approximation for the original integral for a large N is:

$$I = \sqrt{2\pi/N|f''(x_0)|} e^{Nf(x_0)} \tag{6}$$

This method can also refer as the saddle point method, because the plot of the function over the complex plane is called analytic landscape, which only has saddle points and troughs but it never has peaks. Moreover, the troughs reach down all the way to the complex plane, and if there are no poles, then the saddle points are next in line to dominate the original integral.

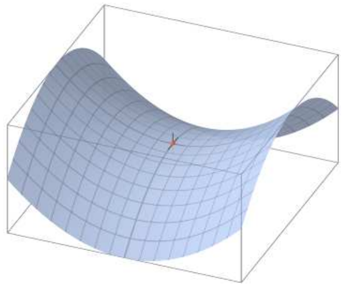


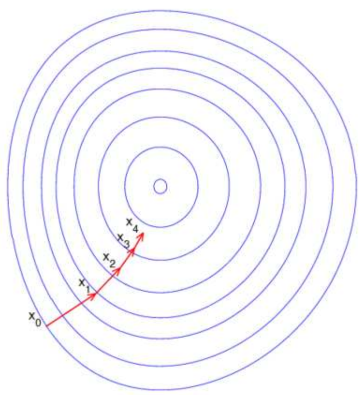
FIG. 3: For a saddle point, in general, the surface resembles a saddle that curves up in one direction, and curves down in a different direction (like a mountain pass). In terms of contour lines, a saddle point can be recognized, in general, by a contour that appears to intersect itself.

III. CONCLUSIONS

In conclusion, The Method of The Steepest Descent, also known as The Gradient Descent,is the simplest of the gradient methods. By using simple optimization algorithm, this popular method can find the local minimum of a function. It's concept is very easy to understand: we start by simply picking an arbitrary point x_0 that is within a function's range and take small steps towards the direction of greatest slope changes, which is the direction of the gradient, and eventually, after many iterations, we can find the minimum of the function. It is popular because of its conceptual simplicity, easy to use, with fast iterations and it can always locate an existing minimum. The only draw back is if it is applied to some badly scaled system, then its slow convergence will cause it to run numerous iterations process that will take forever before the minimum is located.

There are many useful applications of the method of steepest descent, the most common would be using it for a complex integral in order to find the saddle points. This is a truly diverse function that no personal with math background should overlook!

FIG. 4: The Method of Steepest Descent finds the local minimum through iterations, as the figure shows, it starts with an arbitrary point x_0 and taking small steps toward the direction of Gradient since it is the direction of fastest changes, and stops at the minimum



IV. REFERENCES

[1] On the Identification Property of a Projected Gradient Method. P. L. De Angelis; G. Toraldo.

[2] An Efficient Method to Solve the Minimax Problem Directly. C. Charalambous; A. R. Conn. SIAM Journal on Numerical Analysis, Vol. 15, No. 1. (Feb., 1978), pp. 162-187, Jstor.

[3] Iterative image restoration by a method of steepest descent. Y Ichioka et al 1981 J. Opt. 12 35-41 doi: 10.1088/0150-536X/12/1/002.

[4] Two Algorithms Related to the Method of Steepest Descent. T. M. Whitney, R. K. Meany. SIAM Journal on Numerical Analysis, Vol. 4, No. 1. (Mar., 1967), pp. 109-118, Jstor.

[5] An Historical Survey of Computational Methods in Optimal Control. E. Polak. SIAM Review, Vol. 15, No. 2, Part 2: Anniversary Supplement. (Apr., 1973), pp. 553-584, Jstor.

[6] Steepest Descent for Singular Linear Operator Equations. M. Z. Nashed. SIAM Journal on Numerical Analysis, Vol. 7, No. 3. (Sep., 1970), pp. 358-362, Jstor.