VinAI representation content

Hi all, welcome everyone to my talk. My talk today will be on the **Optimal learning rate on Local Descent**. Let me start with some introduction about myself. My name is Tran Van Khoa, and I am a third-year student at University of Engineering and Technology, belongs to Vietnam National University Hanoi. My major in school is Robotics Engineering. After learning about hardware as well as software, theory as well as practice, I eventually found my interest field that suits to me, that is Machine Learning theory, especially Optimization.

So, we now turn to my topic that I will discuss in this talk. In here, we try to minimize the objective function $f()$ with respect to parameter α - called **leanring rate** or **step size**. The parameter x denoted for a current point and we have to move this point follow the direction d by a distance αd to get new point such that the objective function $f()$ get a local minimum value. And next, I will introduce some approach that we can use to find the optimal parameter α.

The first one, I will talk about **Line search**. In this method, we assume that we have chosen a descent direction d, for example, we can use gradient method to determine d. Then, we use this **Line search** method to choose the α. This is a simple method, a native approach, we just search the α value in an interval and choose an α value that make the function $f()$ obtains the minimum value. I have a figure to make you clear. This is the graph of function $f()$ with the parameter α. We consider α in an interval, and we see that this function gets the minimum value at the position α^*. For another example, this is a visualization in 2D space where the red lines represent for direction d, and the black points are the optimal new points in each step. We can see that, after some iteration, our value of objective function is tend to local minimum. This is some algorithm we can use for search the optimal value α: **Fibonacci search**, **Golden section search** and **Shubert-Piyavskii method**. There are some algorithms can find local or global minimum in an interval. I just list it here because I have no time to discuss deeper.

The next one, I talk about **Approximate line search**. In this method, we use the **strong Wolfe conditions** to encourage faster convergence. The **strong Wolfe condition** contains two conditions. The first condition for sufficient decrease, requires that the step size cause a sufficient decrease in the objective value. I have an example in here. The blue parts of the line are satisfy the first Wolfe condition. In the second condition, as known as the strong curvature condition which requires the directional derivative at the next iterate to be shallower. In this figure, the green parts of the line are satisfy the second Wolfe condition.

In the last one, I will mention to the **Taylor series approximation**. First, we have Taylor series in second order in the formula (3) where g is the gradient and H is the Hessian matrix. The g in here correspond to $-d$ in the examples we discussed before. If we replace t by $x^{(k)} - \alpha g$, we obtain the formula (4) and return to our main problem. Because H is semi-positive defined matrix, so this function always exists a global maximum value. Therefore, we take the derivative of function $f()$ with respect to α and set it equals to 0, we have a solution in the formula (5). To analyze this result, we use a property of Hessian matrix that show in the formula (6), apply to our result, we can obtain the range that contains the optimal value of α.

So, that is all of my talk. Thanks for listening!