Course Project Report: Advanced Math Analysis with Matlab

KeZheng Xiong
22920202204622

December 21, 2021

Abstract

The report for the end-of-term project of Advanced Math Analysis with MATLAB fall 2021 course. All the source code is open-sourced on the Github repository `https://github.com/SmartPolarBear/matlab-math-analysis-csxmu-2021` under **GPLv3** license

Contents

1 Problem 1

1.1 Problem Description

Given function $F(x, y) = 0.2x^2 + 0.1y^2 + sin(x + y)$, please work out its gradient. Based on the gradient, please find out the local extreme of function F(x,y) when both x and y are in the range of $[-2 * \pi, 2 * \pi]$. The 2D and 3D views of the function is given in Fig. 1.

1.2 Solution

1.2.1 The gradient of the function

I get the gradient of the function using the following code

```
syms x y;
f=0.2*x^2+0.1*y^2+sin(x+y);
diff(f,x)
diff(f,y)
```

Listing 1: Gradient Calculation

Based on the result, the gradient is

$$\nabla \cdot f(x, y) = (\frac{2 * x}{5} + cos(x + y), \frac{y}{5} + cos(x + y)) \tag{1}$$

1.2.2 Find the extreme values

To find the extreme values of $F(x, y)$ with gradient decent method, we walk little steps towards the direction of the gradient. To formalize this idea, the algorithm is shown as follows.

Algorithm 1 Gradient Descent

Input: Initial point x_0, a constant α, $k = 0$

while termiation condition does not hold **do**

$k = k + 1$

$x_{k+1} = x_k - \alpha \nabla \cdot f(x_k)$

Various problems occurs if this brute-force algorithm is implemented directly. The speed of convergence is annoyingly slow if parameters are not chosen right. In fact, I never succeeded finding a set of parameters that works. A well-known optimization called Stochastic gradient descent, or SGD, improves the performance significantly.

In the brute-force algorithm, The given parameter α controls how long a step is from the point $\mathbf{x_0}$ towards the direction of the gradient. It is a fixed value. On the contrary, in SGD gradient descent algorithm, the length of a step , m , which is also referred to as "learning rate", will be changed each round according to the current position. To be more exact, SGD algorithm tries to find a learning rate m, so that it can minimize the function

$$h(x, y, m) = \mathbf{x_0} + \nabla \cdot f(x, y) \tag{2}$$

So, in the implementation of the algorithm, the following equation is solved each round in the while-loop to get the expected m value

$$\frac{\partial h}{\partial m} = 0 \tag{3}$$

The implementation is shown in Code 2. There are four results of the function. *endp* is the final point, or the result of gradient descent algorithm. *num* is the number of steps taken. *hist* is a vector which records all the coordinates of each step, and *list* contains the corresponding function values. The three parameters are the function, the initial point and the precision, respectively.

```
function [endp,num,hist,list]=gradient_descent(f,x0,eps)
    syms x y m;
    d=-[diff(f,x);diff(f,y)];

    nd=subs(d,x,x0(1));
    nd=subs(nd,y,x0(2));
    nrm=double(norm(nd));

    list=[];

    k=0;
    while(nrm>=eps)
    nx0=x0+m*nd;
    nf=subs(f,x,nx0(1));
    nf=subs(nf,y,nx0(2));
    h=diff(nf,m);
    mm=solve(h);

    x0=x0+mm*nd;
    k=k+1;

    hist(k,:)=[double(x0(1));double(x0(2))];
    vf=subs(f,x,x0(1));
    vf=subs(vf,y,x0(2));
    list = [list double(eval(vf))];

    nd=subs(d,x,x0(1));
    nd=subs(nd,y,x0(2));
    nrm=double(norm(nd));
    end

    num=k;
    endp=x0;
end
```

Listing 2: SGD Gradient Descent

To test the implementation, Each point of the steps taken in the procedure of gradient descent algorithm is plotted on Figure 1, in red scattered lines, and the result is in Table 1. In Table 1, the 4th column shows the value from the builtin $fminsearch$ function, which reveals that the implementation of SGD gradient descent algorithm gives the right answers.

The initial points for gradient descent chosen are $(-3,-4)$, $(0,-1)$, and $(-2,-2)$. The choice of initial points are significant. It can **influence the speed of convergence dramatically**. Given a bad initial point, the algorithm may not reach the point of convergence at all, or take a unbearable long time to reach there. It worth mentioning that the second choice of $(0,-1)$ is especially difficult. It takes a great many attempts to find this point which contributes to a high speed of convergence.

Figure 1: The visualization of the steps of the SGD gradient descent algorithm

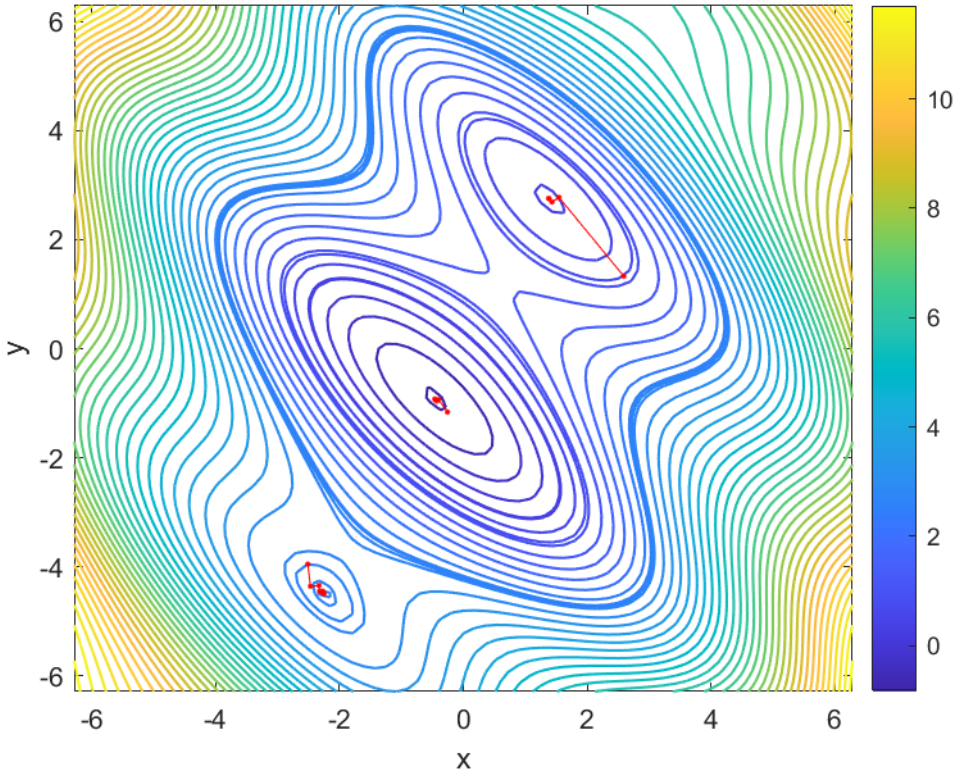

Table 1: The Extreme Values of $F(x, y)$

No.	X-Y Coordinates	$F(x, y)$	Result of $fminsearch()$
0	$(-2.2469, -4.4902)$	2.5874	2.5874
1	$(-0.4611, -0.9241)$	-0.8549	-0.8549
2	$(1.3768, 2.7529)$	0.3020	0.3020

1.3 Analysis and Conclusion

Gradient descent is a widely-used way of finding the extreme values of any function. There are many optimizations for this method, which will improve the speed of convergence. SGD is taken in this project, which yields satisfying results.

SGD gradient descent has a major flaw that it is known for "easy to be trapped in a local minimum", which may account for the difficulties in finding a good initial point for the second extreme value. More sophisticated optimization such as Momentum gradient descent and AdaGrad gradient descent can be applied if the speed of convergence is too slow.

2 Problem 2

2.1 Problem Description

A factory supplies engine for its customer. According to the contract, the factory should deliver its product to the customer in the end of the 1st season: 40 engines; in the end of the 2nd season: 60 engines; in the end of the 3rd season: 80 engines. Due to limited productivity, to its most, the factory is only able to produce 100 machines in a single season. The production cost is given as $f(x) = 50x + 0.2x^2$, where x is the

number of machines produced in one season. It is possible to produce machines more than the required quota in each season. In this case, the storage cost is 4 dollars for one machine/season. Assuming there is no reservations at the beginning for the first season, please work out a production plan that minimizes the production cost

3 Problem 3

3.1 Problem Description

Prove: Among the top 32 World Cup finals, the information entropy of the event that a team winning is 5.

4 Acknowledgment

Thanks to (TODO)