Coursework 1 Mathematics for Machine Learning (CO-496)

CID:01366977 Name: Fanbo Meng

*Solution a)*

The three parameters for completed square form for f1(**x**) are:

**c** = **C** = c0=-1/6

*Solution b)*

When **x=** , function f1(**x**) achieves the minimum value, which is -1/6.

Explanation: we can write f1(**x**) in this form(assume that **x**=):

f(x1,x2)=4() - 2- ()

= =

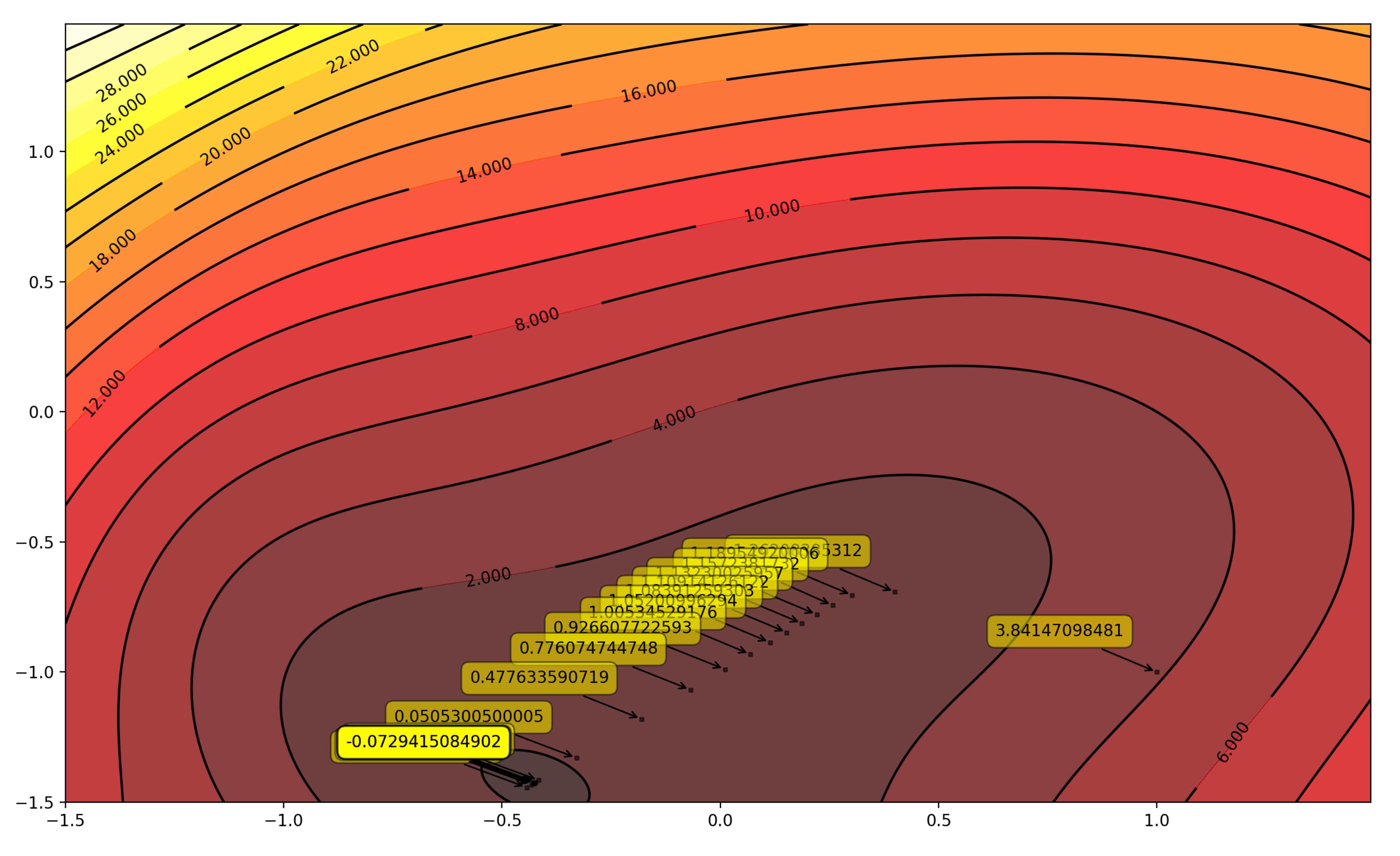
we get:

x1 = 1/6 x2 = 1/6

then we compute the second-order gradient:

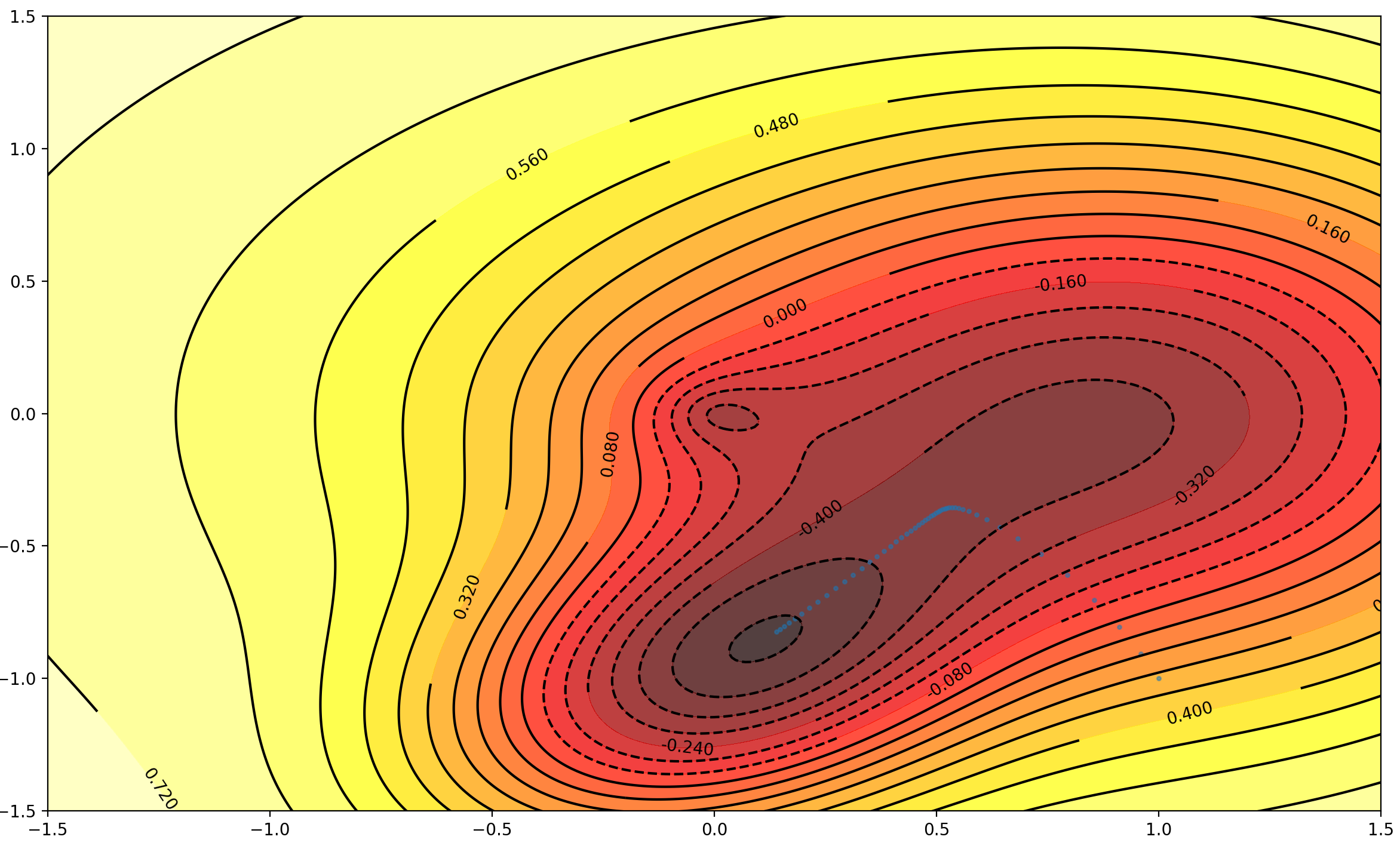
= 8 C: = 8 B:-2 -2

easily we get :AC-B2>0 then we can determine that **x**= is minimum point of function f1 and the minimum value is -1/6.

*Solution f)*

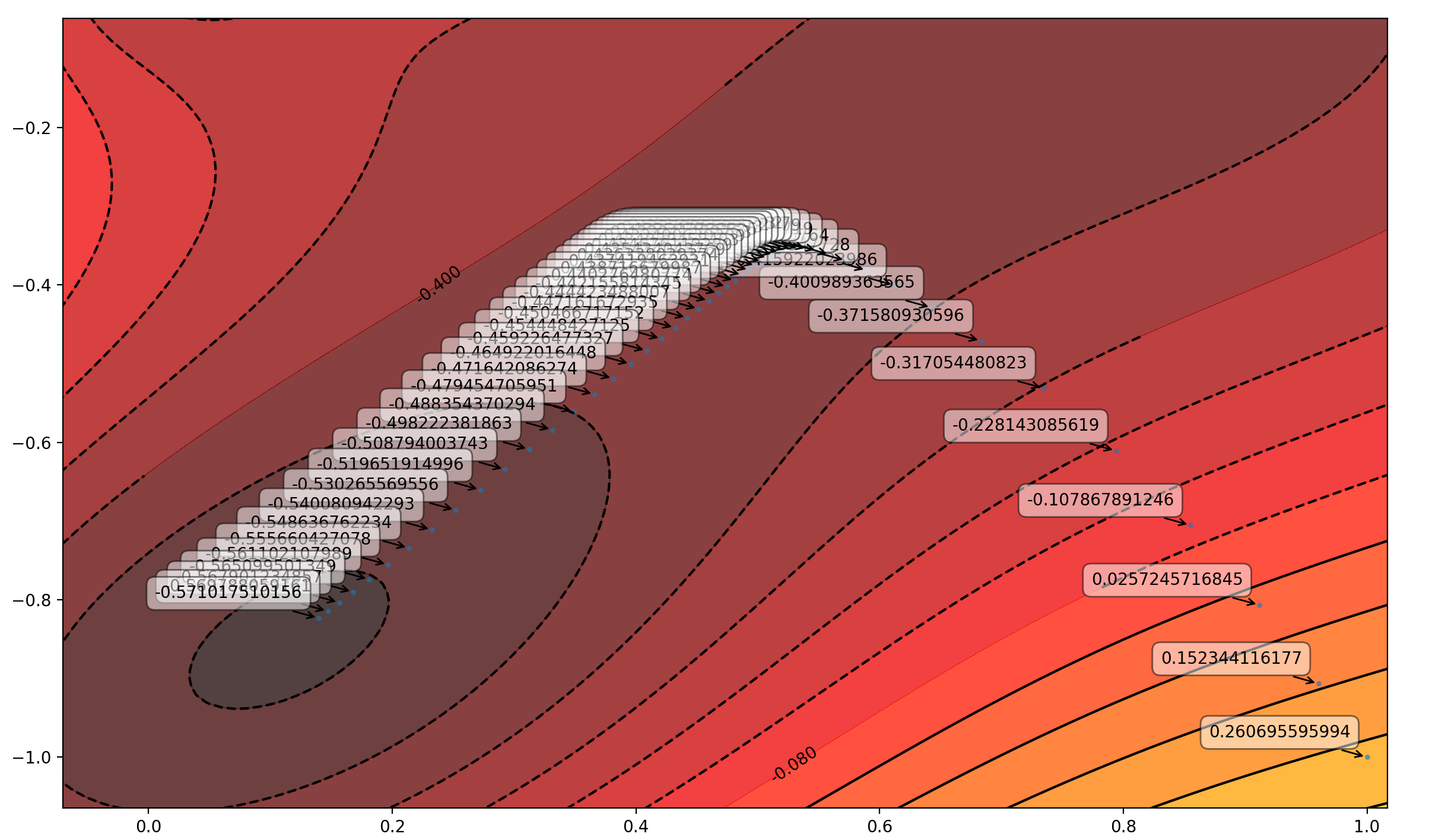
1.1

1.1:Gradient decent graph for function f2 with a pace 0.01



1.2

1.2:Gradient decent graph for function f3 with a pace 0.01

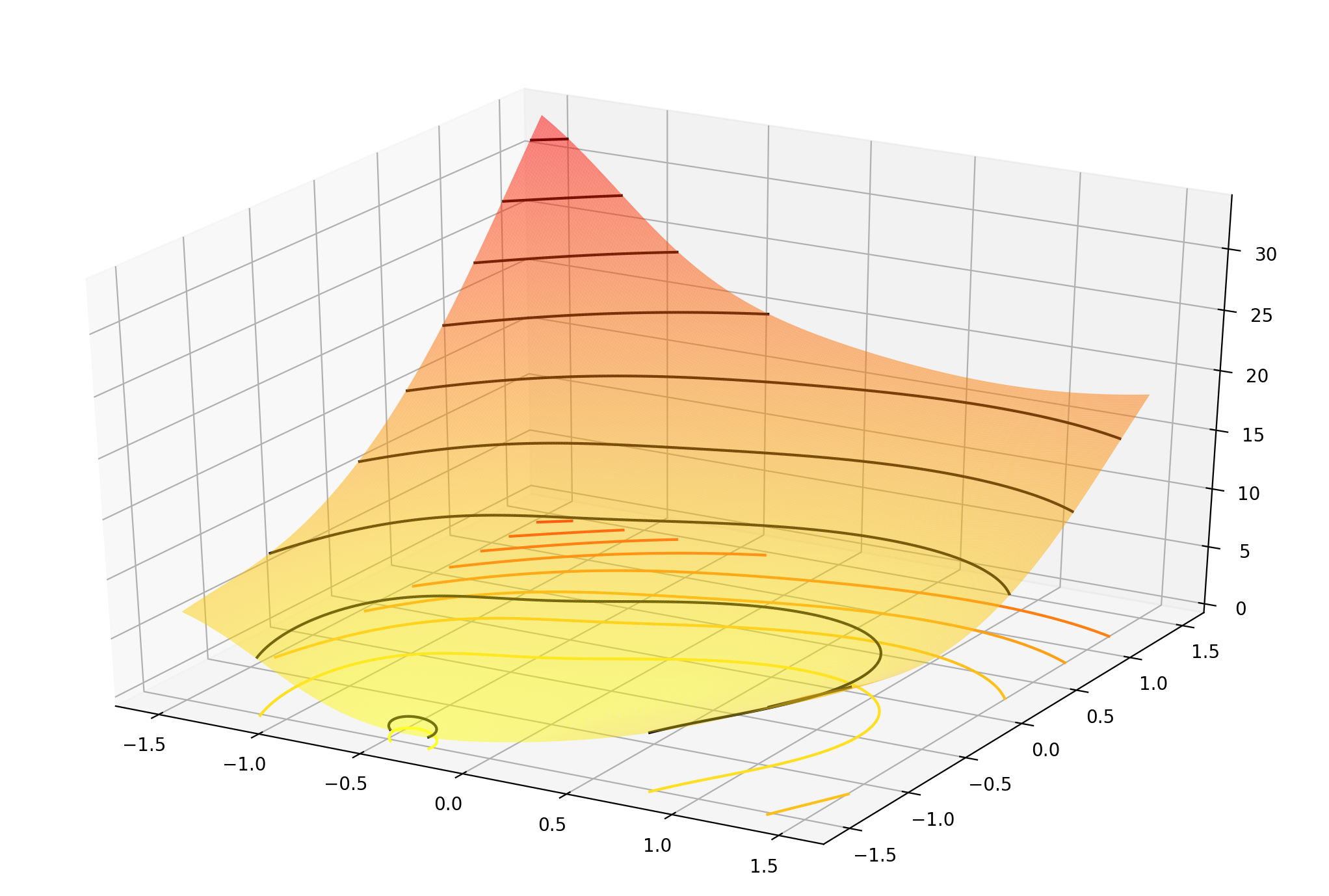


1.3

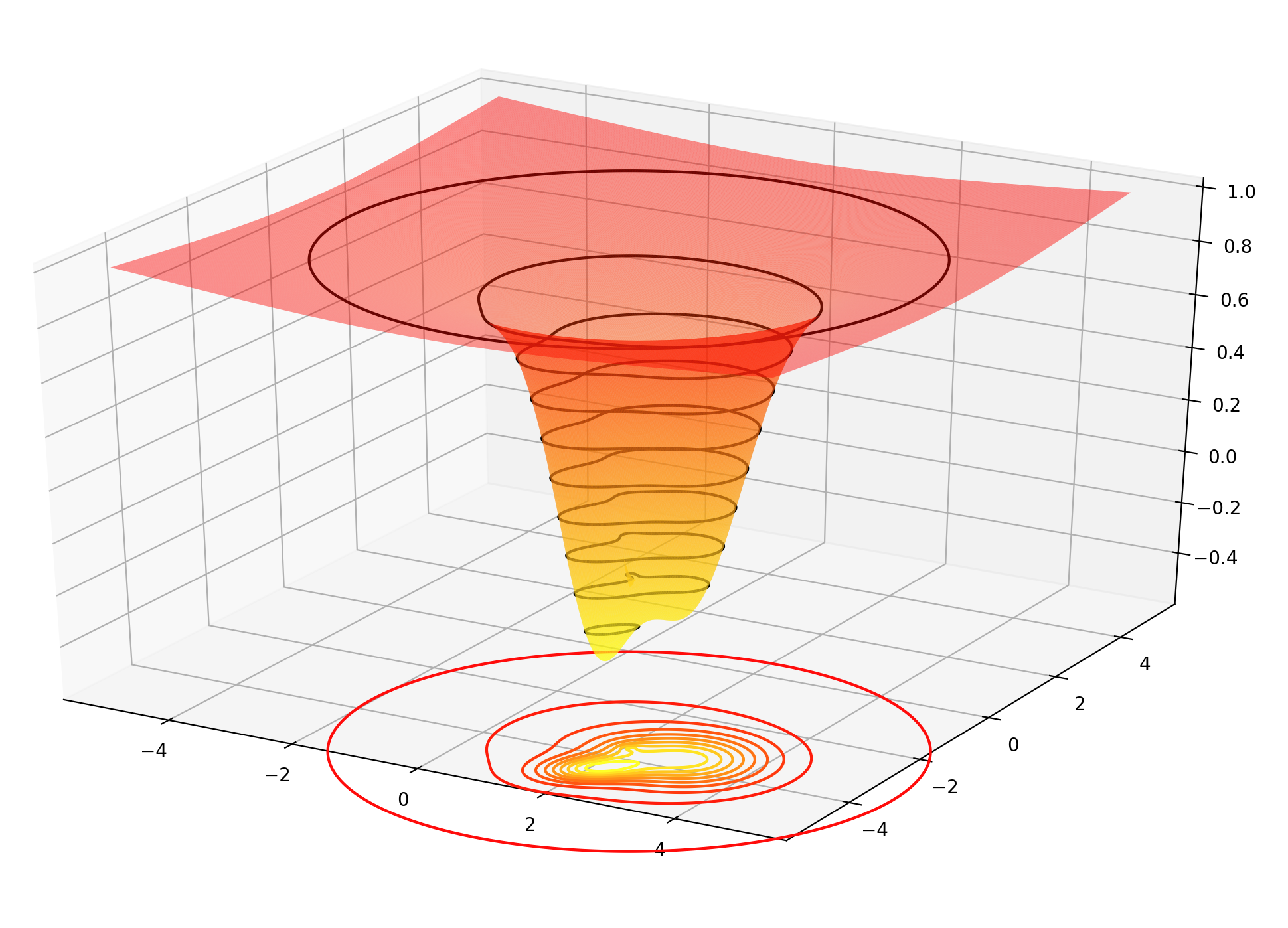
1.3: A labeled zoomed version for 1.2

*Solution g):*

To make my illustration more understandable, here is some graphs showing different paces on gradient descend algorithm on f2 and f3 and 3D graph of f2 and f3:



1.4 (3D graph of f2)

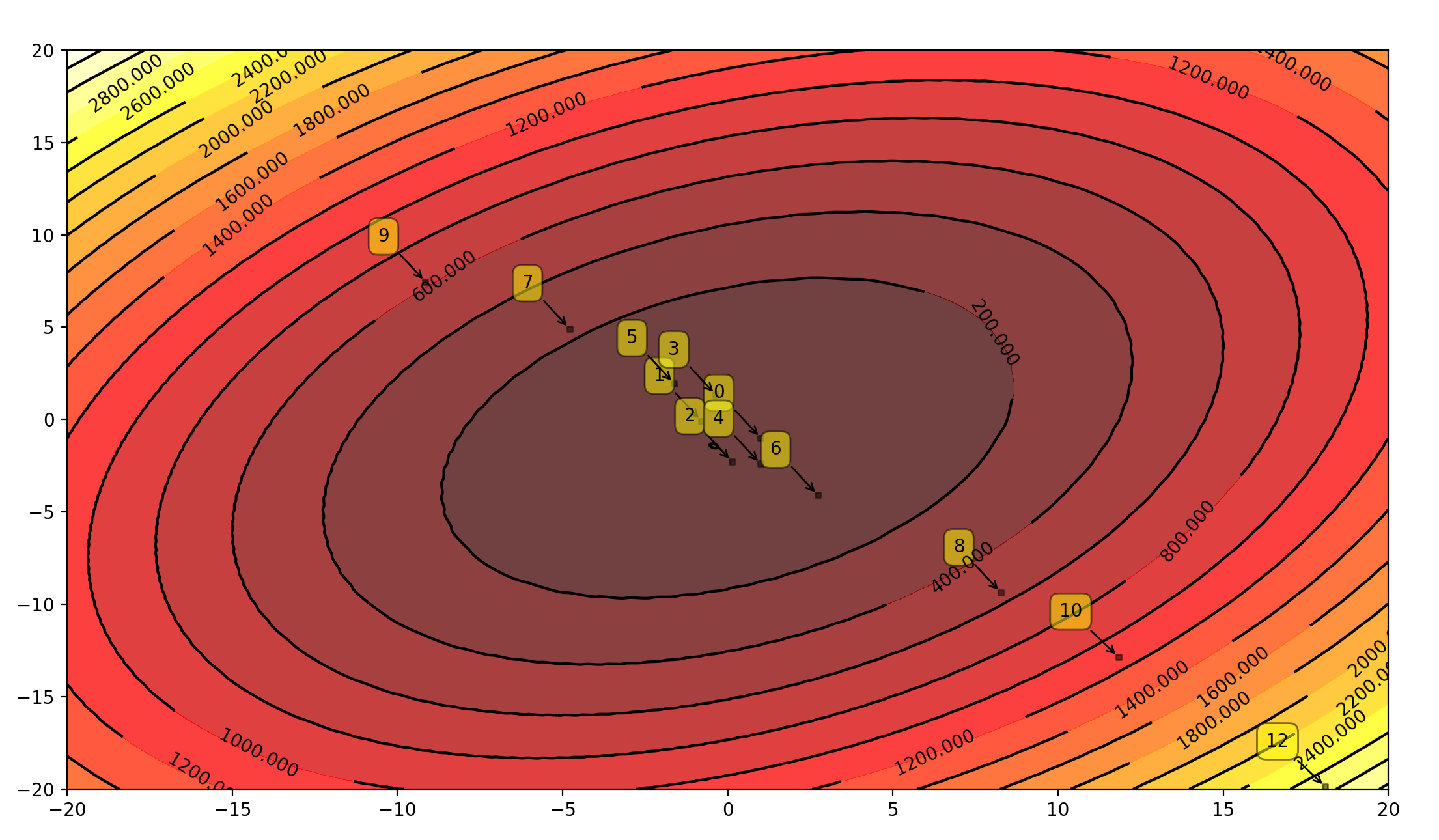


1.5(3D graph of f3)

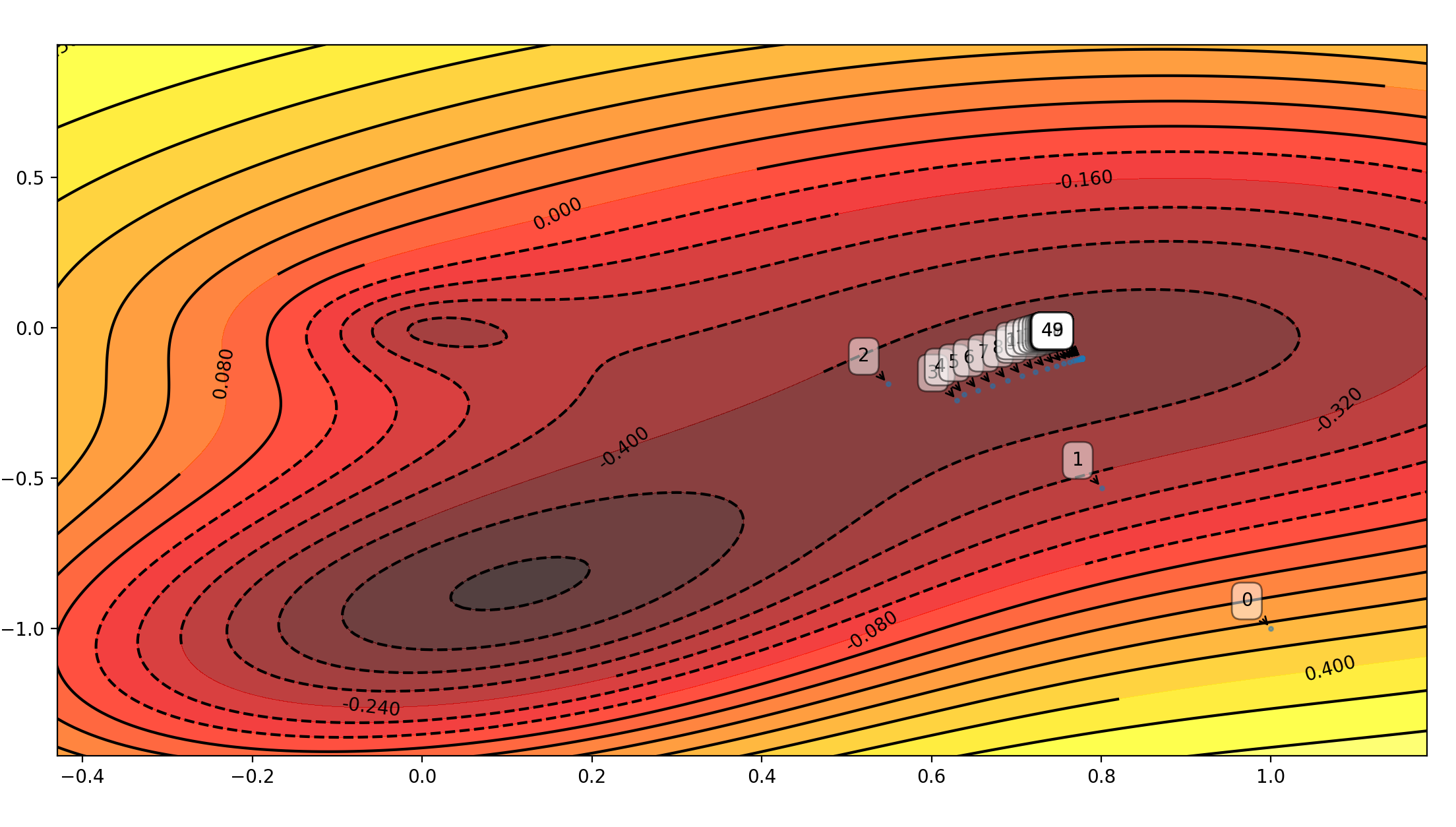
12.Due to the computing with small sizes of steps, one minima point can be found near the true minima point of the function, which can be sensibly found in graph 1.4. When using large step sizes to search minima of f2, it can be seen from graph 1.6 that each iteration point fluctuates with a central point near minima point, moreover, the fluctuation pace is getting longer with the increasing times of iteration.

345.When computing minima point of f3, one minima point can be found near the true minima point of the function, which can be sensibly found in graph 1.5. When changing the step size to 0.5, it can be seen that the value goes towards a local but not global minima. When the step size is 1, it can be seen that the minima point searched fluctuates with a central point near local minima point. When the step size is 2, it can be seen that the points searched randomly distributed near the local minima.

6.Explanation: It is because the attribute of step size in gradient descend algorithm. Step size is a parameter determining the granularity of searching. For function using gradient descend algorithm to find a minima value, if the step size is small, it means that minima value is searched "carefully", and the alteration of each iteration is small. So when enlarging the step size, the process of searching minima via gradient descend algorithm will be "careless", in this case, f2 will finally fluctuate and f3 may converge to a local minima or also fluctuate.



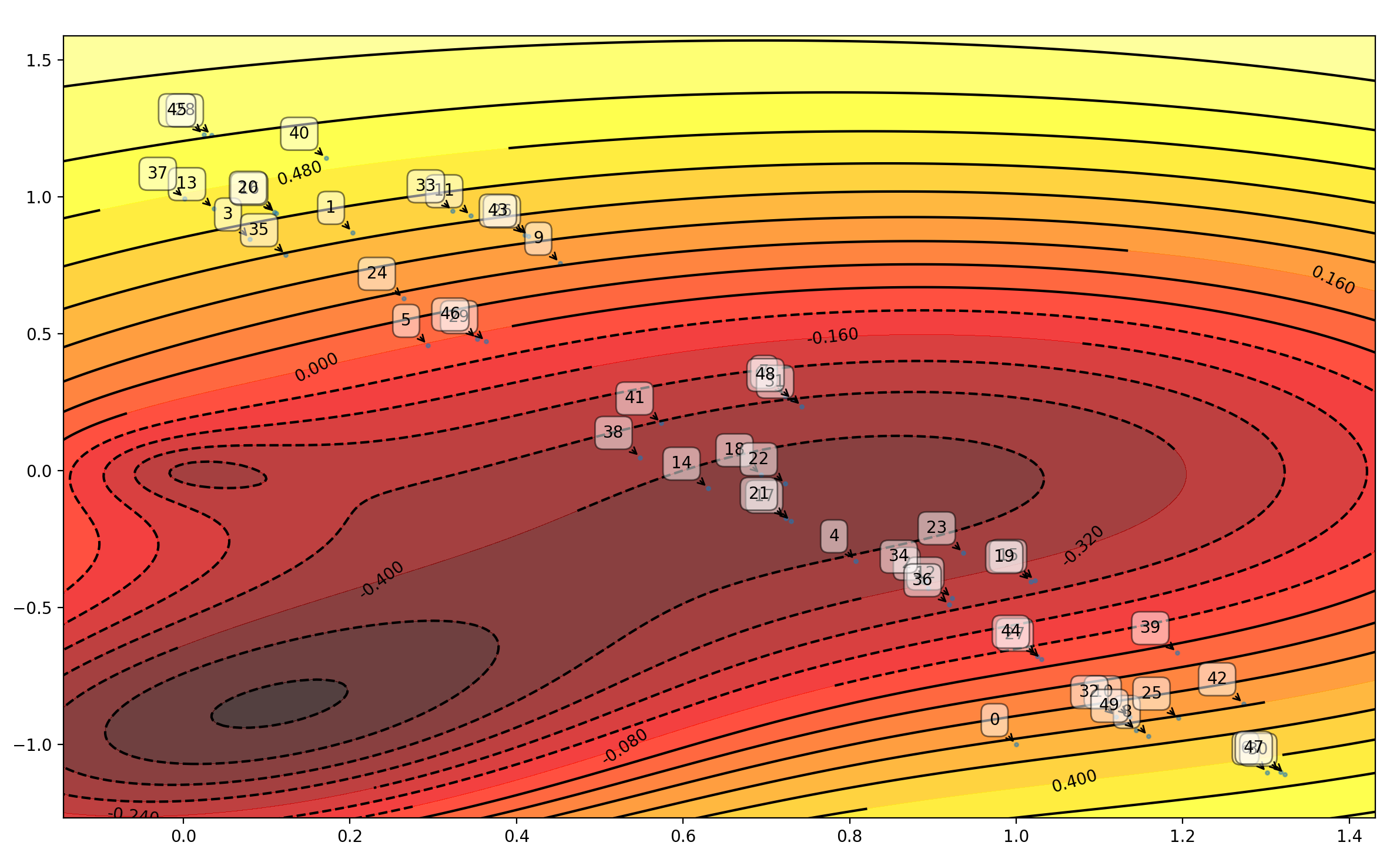
1.6(f2 gradient descend with pace 0.3)



1.7(f3 gradient descend with pace 0.5)



1.8(f3 gradient descend with pace 1)



1.9(f3 gradient descend with pace 2)