Coursework 1Mathematics for Machine Learning (CO-496)

CID:01366977 Name: Fanbo Meng

Solution a)

The three parameters for completed square form for $f_1(\mathbf{x})$ are:

$\mathbf{c} = \begin{pmatrix} 1/6 \\ 1/6 \end{pmatrix}$ $\mathbf{C} = \begin{pmatrix} 4 & -1 \\ -1 & 4 \end{pmatrix}$ c_0=-1/6

Solution b)

When $\mathbf{x}=$ $\begin{pmatrix} 1/6 \\ 1/6 \end{pmatrix}$, function $f_1(\mathbf{x})$ achieves the minimum value, which is -1/6.

Explanation: we can write $f_1(\mathbf{x})$ in this form(assume that $\mathbf{x}=\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$):

$f(x_1,x_2)=4(x_1^2 + x_2^2) - 2x_1x_2 - (x_1+x_2)$

$\frac{\partial f}{\partial x_1}=8x_1 - 2x_2 - 1$ $\frac{\partial f}{\partial x_2}=8x_2 - 2x_1 - 1$

$make\ \frac{\partial f}{\partial x_1} = 0$ $\frac{\partial f}{\partial x_2} = 0$ we get:

$x_1 = 1/6$ $x_2 = 1/6$

then we compute the second-order gradient:

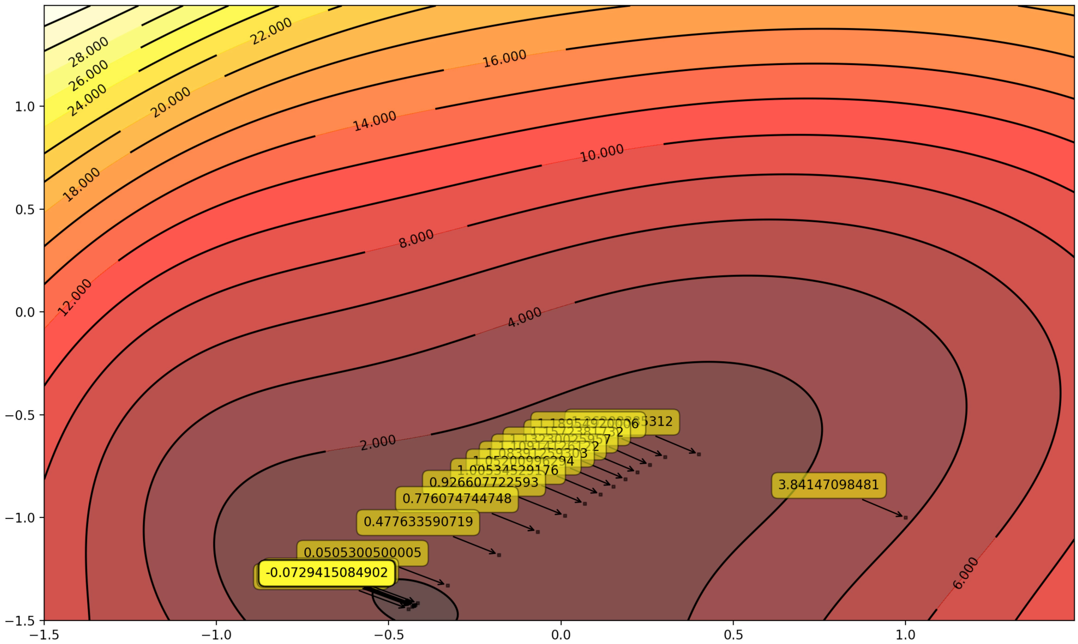
A: $\frac{\partial^2 f}{\partial x_1^2} = 8$ C: $\frac{\partial^2 f}{\partial x_2^2} = 8$ B: $\frac{\partial^2 f}{\partial x_1 x_2} = -2$

B: $\frac{\partial^2 f}{\partial x_2 x_1} = -2$

easily we get :$AC-B^2>0$ then we can determine that

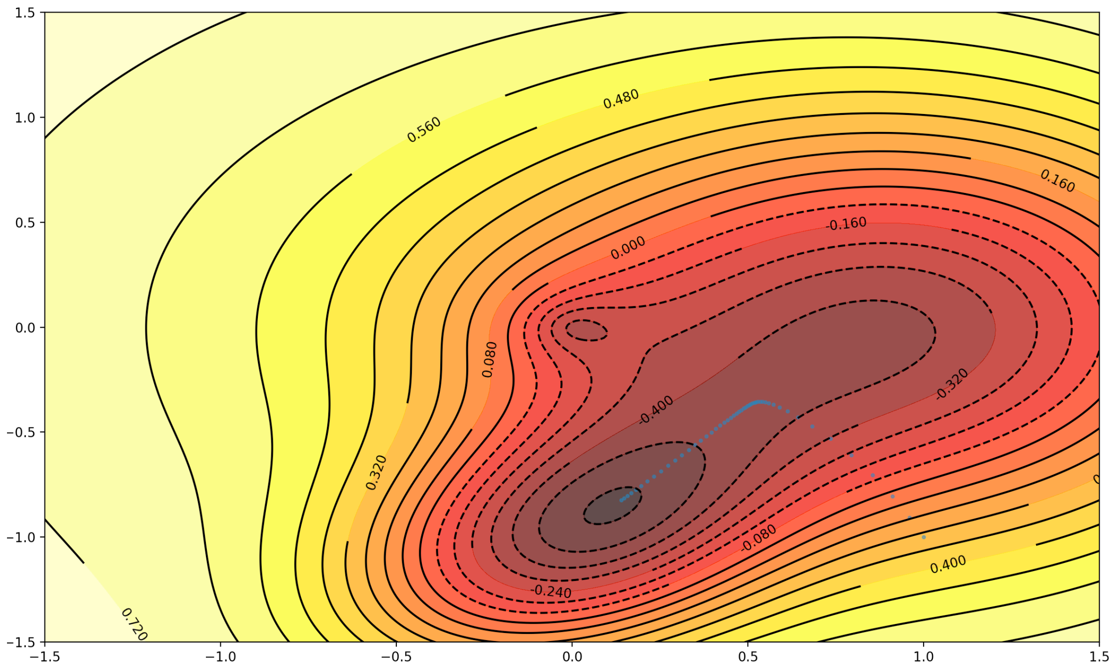
$\mathbf{x}=\begin{pmatrix}1/6\\1/6\end{pmatrix}$ is minimum point of function f_1 and the minimum value is -1/6.

<u>Solution f)</u>



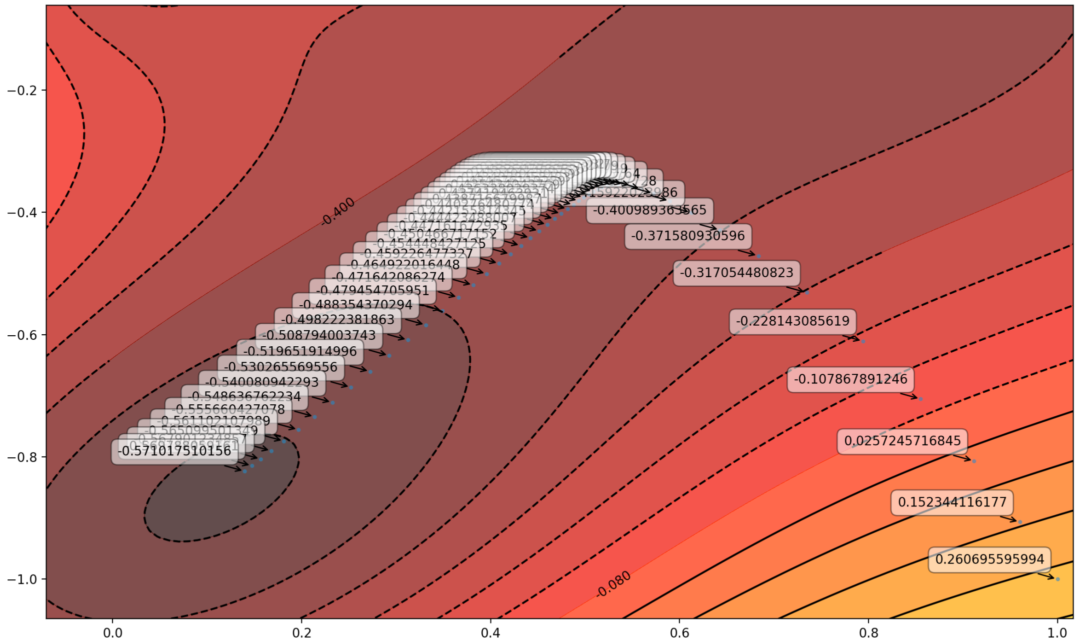
1.1

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

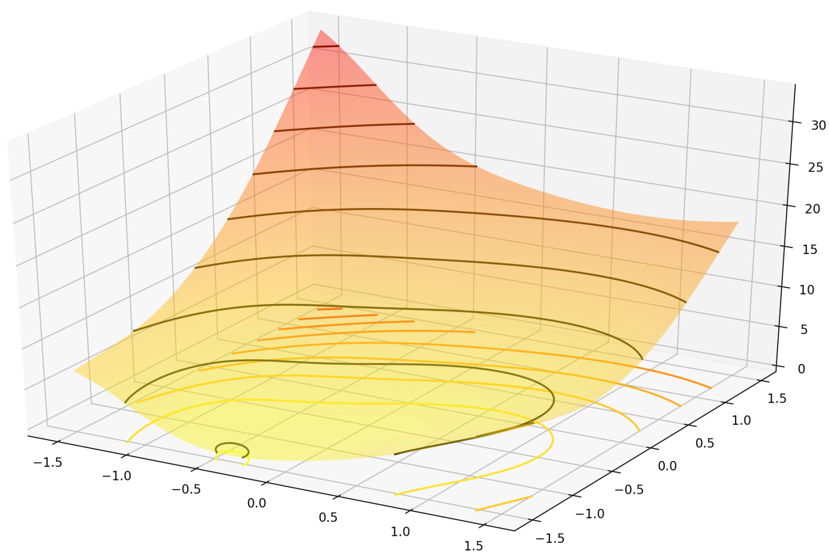


1.3

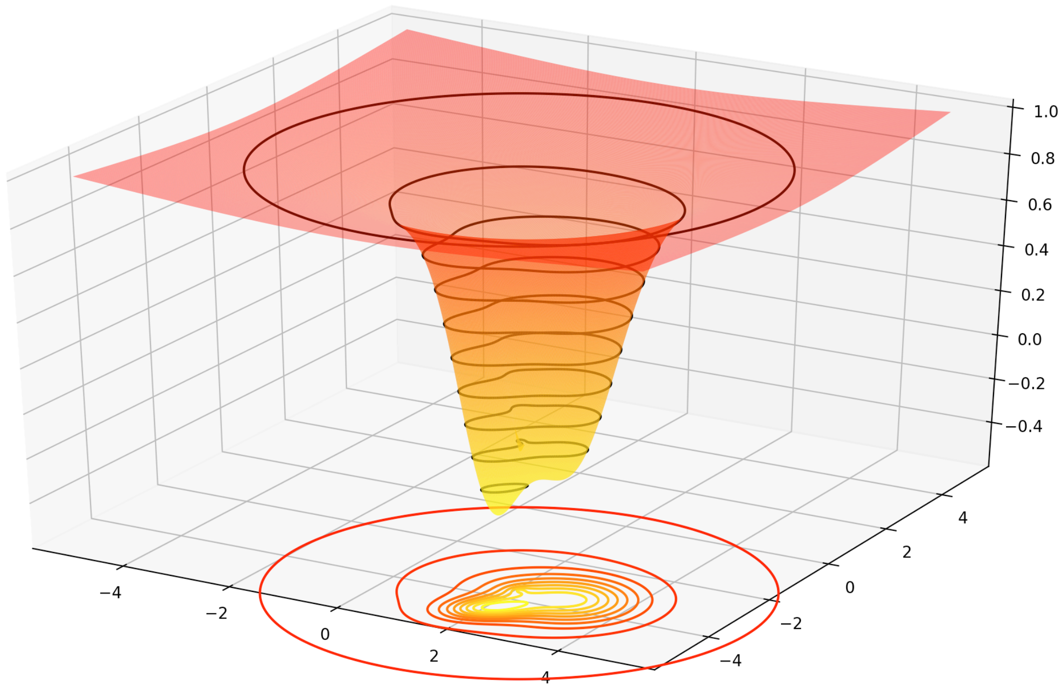
1.3: A labeled zoomed version for 1.2

__Solution g):__

To make my illustration more understandable, here is some graphs showing different paces on gradient descend algorithm on f_2 and f_3 and 3D graph of f_2 and f_3:



1.4 (3D graph of f_2)

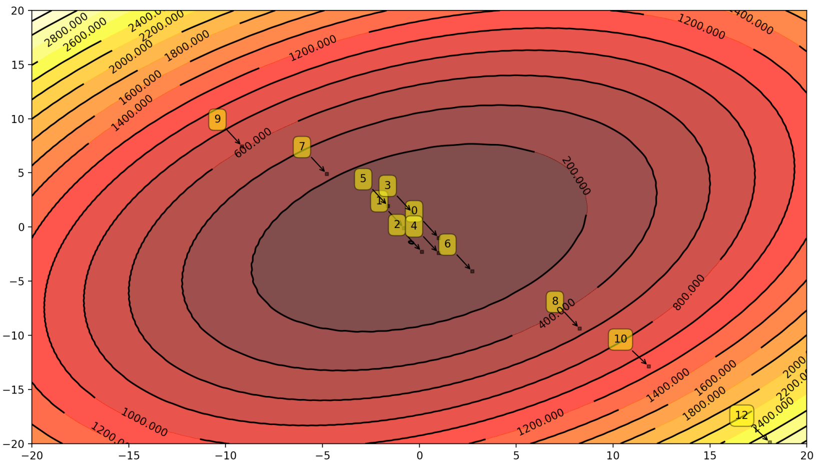


1.5(3D graph of f_3)

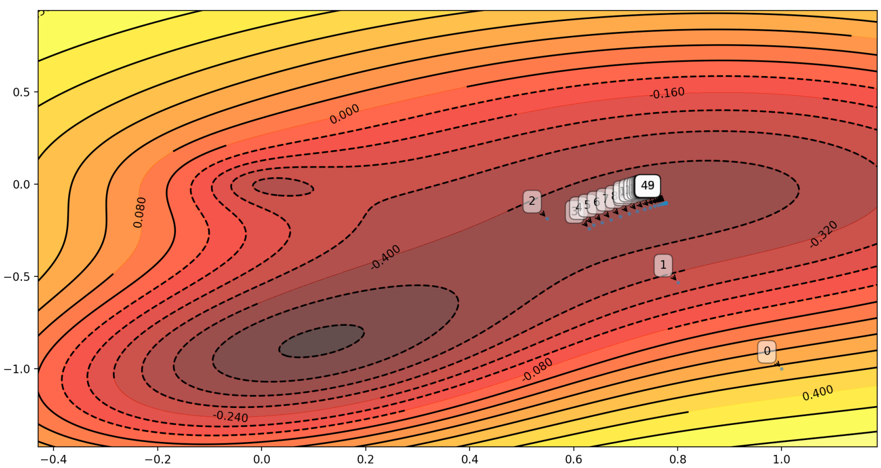
12.Due to the computing with small sizes of steps, one minima point can be found near the true minima point of the function, which can be sensibly found in graph 1.4. When using large step sizes to search minima of f_2, it can be seen from graph 1.6 that each iteration point fluctuates with a central point near minima point, moreover, the fluctuation pace is getting longer with the increasing times of iteration.

345.When computing minima point of f_3, 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.

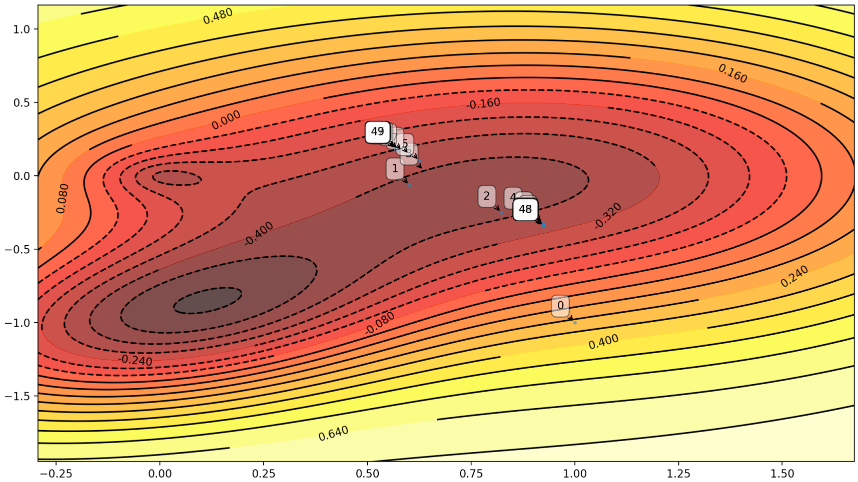
6.Explanation: It is because the attribute of step size in gradient descend algorithm. Step size is a parameter determining the granularity of searching. For function using gradient descend algorithm to find a minima value, if the step size is small, it means that minima value is searched "carefully", 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, f_2 will finally fluctuate and f_3 may converge to a local minima or also fluctuate.



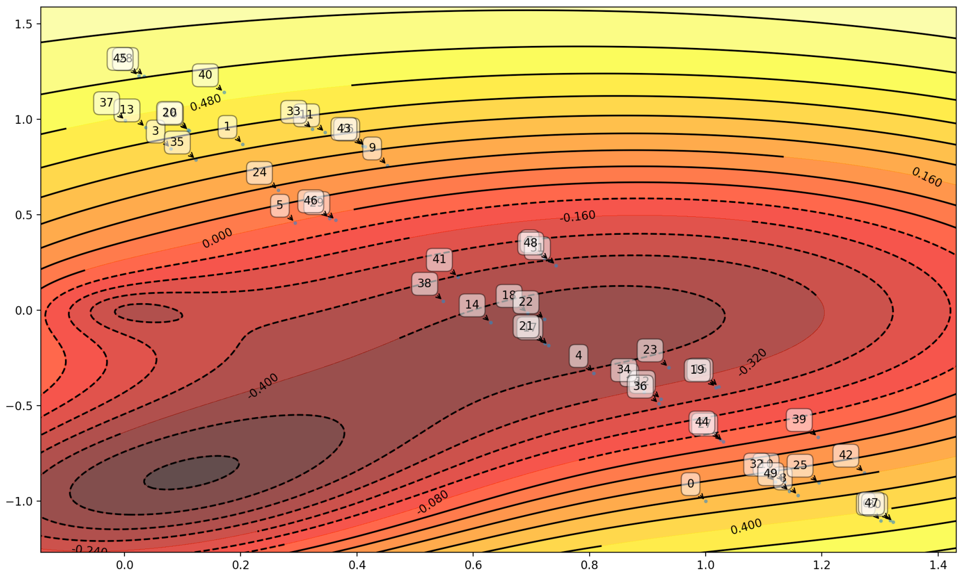
1.6(f_2 gradient descend with pace 0.3)



1.7(f_3 gradient descend with pace 0.5)



1.8(f_3 gradient descend with pace 1)



1.9(f_3 gradient descend with pace 2)