CO496 CW1
Chengyi Zhang
01604256

(a). $f_1(x) = x^T x + x^T B x - a^T x + b^T x$
$= x^T(I + B)x - (a - b)^T x$
$= (x - c)^T C(x - c) + c_0$

Get $\mathrm{C} = \mathrm{I} + \mathrm{B} = \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}$. Note that C is symmetric

$f_1(x) = x^T C x - 2c^T C x + c^T C c + c_0$

Get $Cc = \frac{(a-b)}{2} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$, solve and get $\mathrm{c} = \begin{bmatrix} 1/6 \\ 1/6 \end{bmatrix}$, $c_0 = -c^T C c = -1/6$

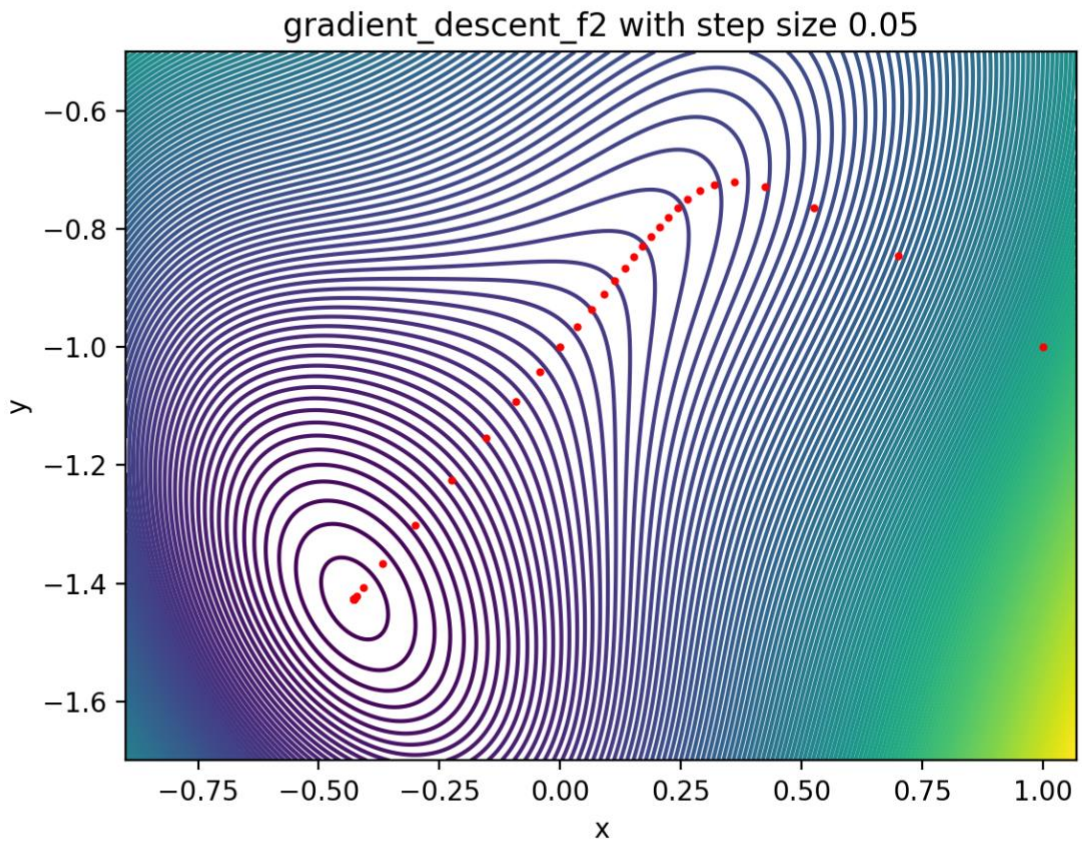
$f_1(x) = \left(x - \begin{bmatrix} 1/6 \\ 1/6 \end{bmatrix}\right)^T \begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix} \left(x - \begin{bmatrix} 1/6 \\ 1/6 \end{bmatrix}\right) - 1/6$

(b). From (a), can get $\nabla f_1(x) = 2\begin{bmatrix} 4 & -1 \\ -1 & 4 \end{bmatrix}\left(x - \begin{bmatrix} 1/6 \\ 1/6 \end{bmatrix}\right), H = \begin{bmatrix} 8 & -2 \\ -2 & 8 \end{bmatrix}$,
Solve $(8 - \lambda)^2 - 4 = 0$, get $\lambda_1 = 6, \lambda_2 = 10$, and by all eigenvectors being positive, hessian matrix H is positive definite and f_1 is convex. Thus f_1 achieves minimum at x if $\nabla f_1(x) = 0$, where $\mathrm{x} = \begin{bmatrix} 1/6 \\ 1/6 \end{bmatrix}$, and $f_1\left(\begin{bmatrix} 1/6 \\ 1/6 \end{bmatrix}\right) = -1/6$

CO496 CW1
Chengyi Zhang
01604256

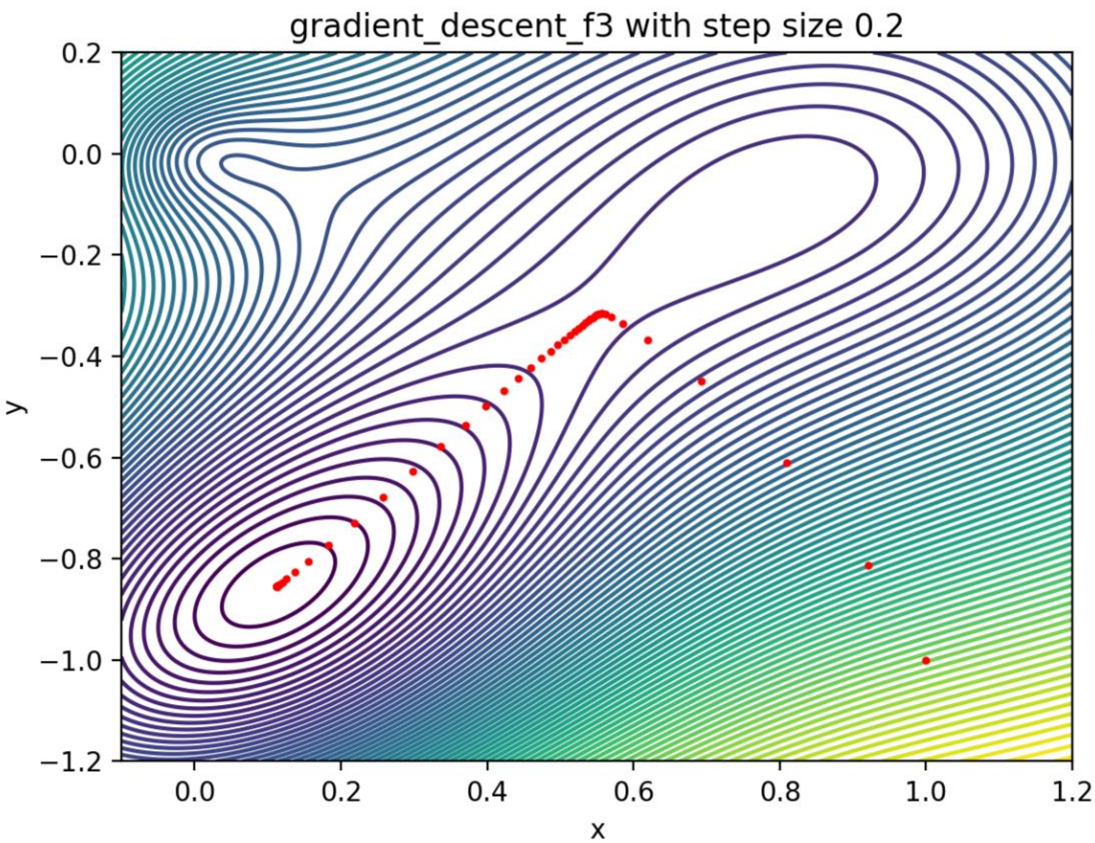
(f). For $f_2(x) = \sin\left(\left(x - \begin{bmatrix}1\\0\end{bmatrix}\right)^T \left(x - \begin{bmatrix}1\\0\end{bmatrix}\right)\right) + (x - \begin{bmatrix}0\\-1\end{bmatrix})^T \begin{bmatrix}4 & -1\\-1 & 4\end{bmatrix} (x - \begin{bmatrix}0\\-1\end{bmatrix})$,

the gradient descent of 50 iterations is



CO496 CW1
Chengyi Zhang
01604256

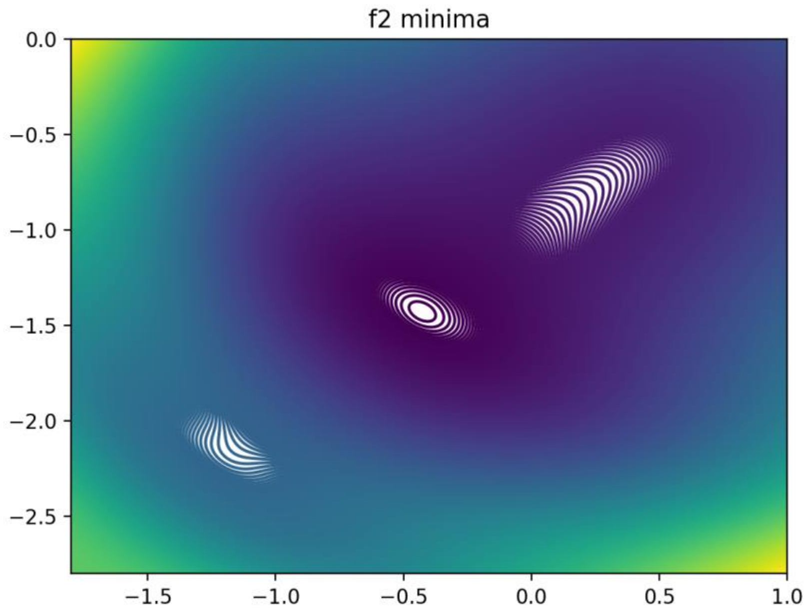
For $f_3(x) = 1 - (\exp\left(-\left(x - \begin{bmatrix}1\\0\end{bmatrix}\right)^T \left(x - \begin{bmatrix}1\\0\end{bmatrix}\right)\right) +$
$\exp\left(-\left(x - \begin{bmatrix}0\\-1\end{bmatrix}\right)^T \begin{bmatrix}4 & -1\\-1 & 4\end{bmatrix} \left(x - \begin{bmatrix}0\\-1\end{bmatrix}\right)\right) -$
$\frac{1}{10}\log\left|\frac{1}{100}I + xx^T\right|)$,
the gradient descent of 50 iterations is



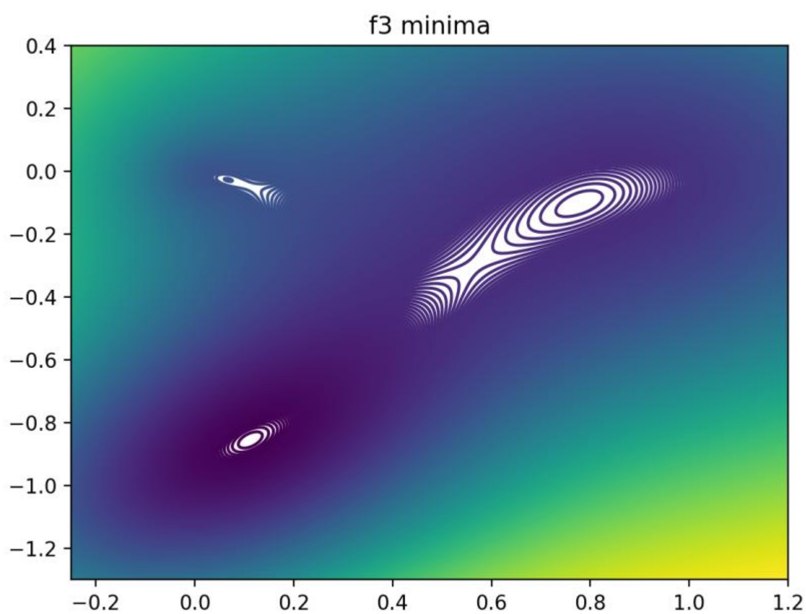
CO496 CW1
Chengyi Zhang
01604256

(g). With a larger scale, we can see the minima of both $f_2 and\ f_3$.

In the graph of f_2, we can see by the change of color that there is one minimum.

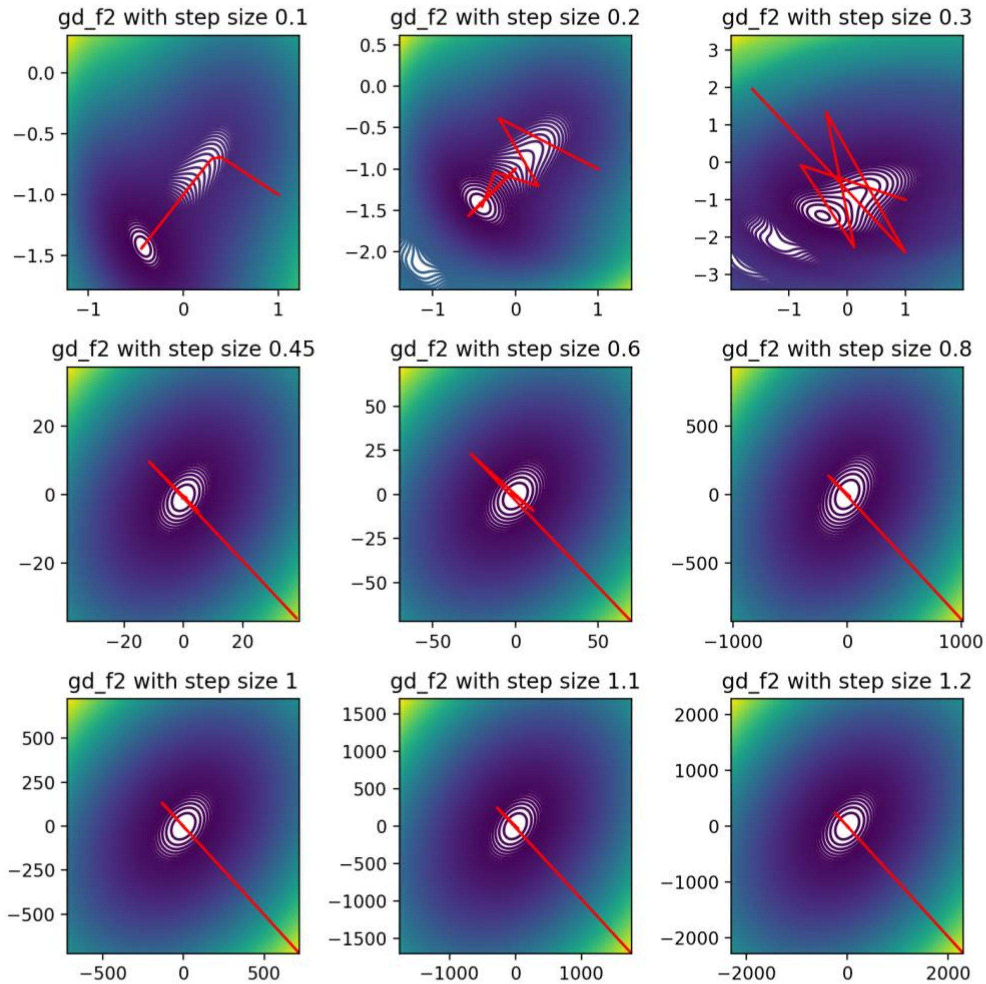


In the graph of f_3, we can see that there are three minima.



CO496 CW1
Chengyi Zhang
01604256

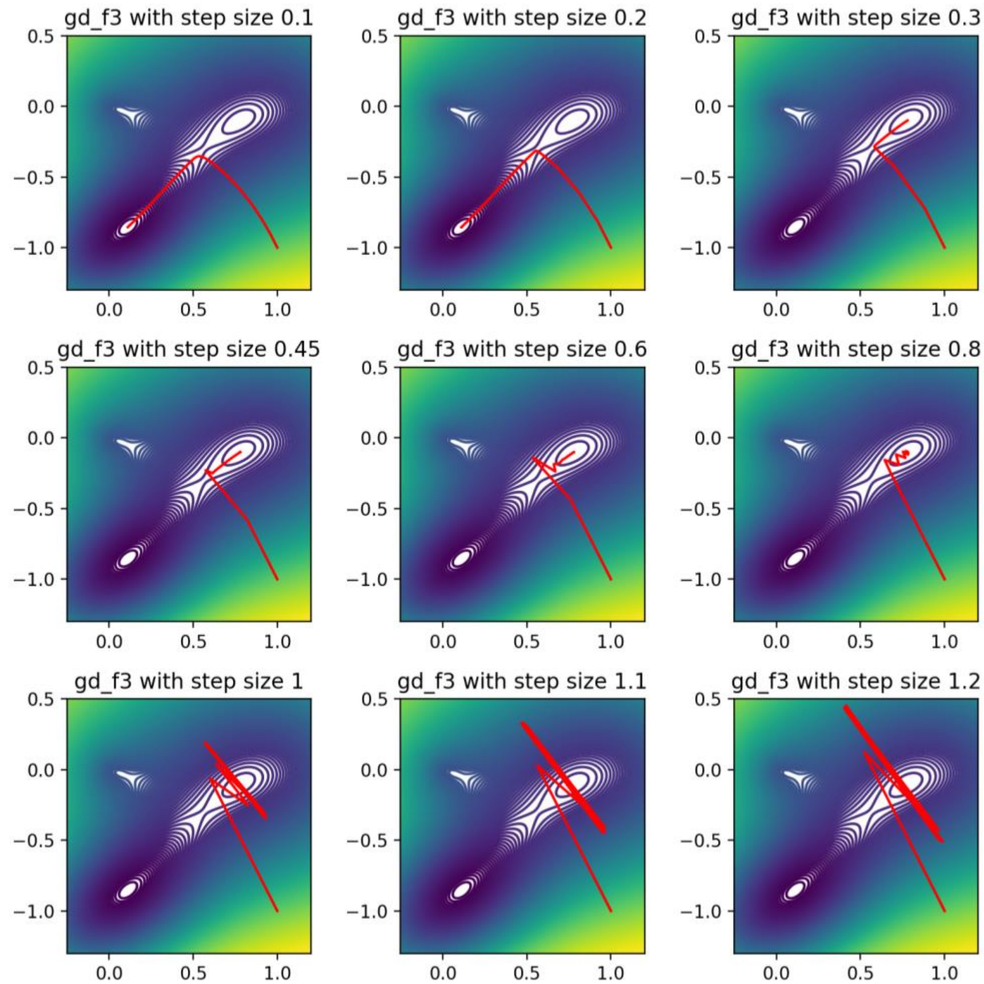
The following graph shows how step size effects the algorithm on f_2.



When the step size grows from 0.1 to 0.2, the algorithm starts to oscillate. And when step size becomes even larger, the algorithm quickly diverges.

CO496 CW1
Chengyi Zhang
01604256

The following graph shows how step size effects the algorithm on f_3.



When the step size grows to 0.3, the algorithm converges to another minimum. When the step size grows larger, it starts to oscillate around the second local minimum.

With a negative sign, gradient points out the direction in which the function decreases the steepest. However, gradient does not necessarily point to a local minimum. The gradient descent algorithm theoretically tends to approach the optima after each iteration. When step size is moderate, although different step sizes may change the directions of following steps, the function decreases after each iteration and the algorithm converges eventually to one of the minima. For all step sizes, the initial point and direction are the same, but different step sizes result in various following points and directions, and then the algorithm may find different minima, like subplot 1 and 3 of f_3. But when step is too large, the algorithm will go 'too far' along a direction and 'go back too far' during the next iteration, which causes oscillation. In some cases, when the oscillation is too severe, the function may even increase after some iterations and the algorithm eventually diverges (e.g. f_2 with step size ≥ 0.3).