Homework 8

1

(a)

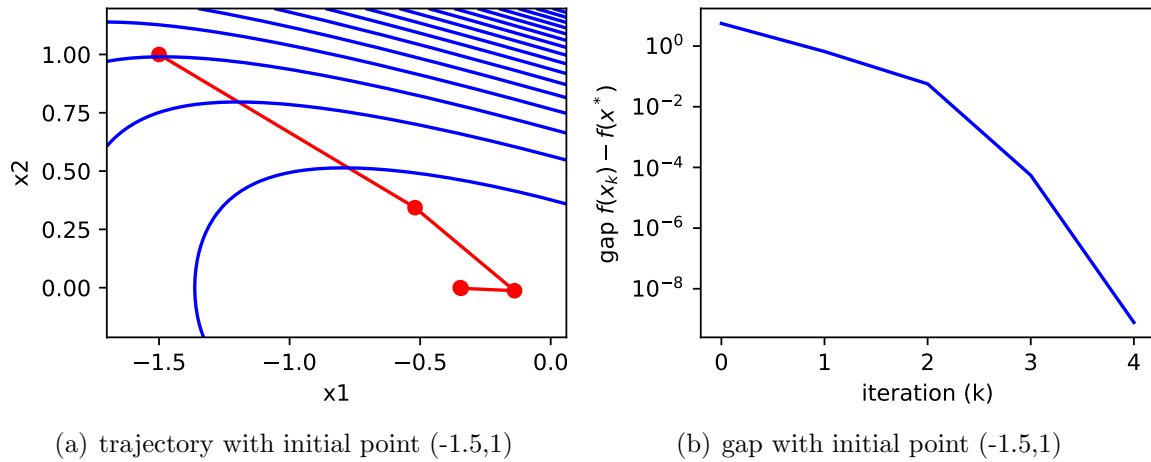(a) trajectory with initial point (-1.5,1)　　　(b) gap with initial point (-1.5,1)

Figure 1: trajectories and gaps for part (a)

The solution is $(x_1, x_2) = (-0.3466, 0)$ and the optimal value is 2.56. Number of iterations is only 4.

(b)

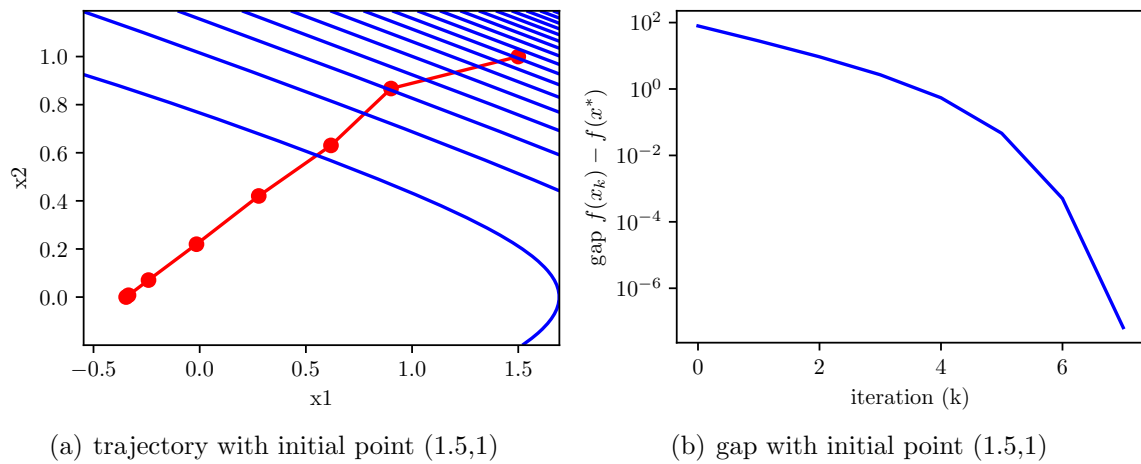(a) trajectory with initial point (1.5,1)　　　(b) gap with initial point (1.5,1)

Figure 2: trajectories and gaps for part (b)

With another initial point the solution is $(x_1, x_2) = (-0.3464, 0)$ and the optimal value is 2.56. Number of iterations is 7 this time.

2

(a)

We have shown that

$$\frac{\partial f}{\partial \boldsymbol{w}} = (\sum_{i=1}^{m} \frac{-y_i \boldsymbol{x_i} e^{-y_i \boldsymbol{x_i^T w}}}{1 + e^{-y_i \boldsymbol{x_i^T w}}})^T$$

Or equivalently we can write that component-wise

$$\frac{\partial f}{\partial w_k} = \sum_{i=1}^{m} \frac{-y_i x_{ik} e^{-y_i \boldsymbol{x_i^T w}}}{1 + e^{-y_i \boldsymbol{x_i^T w}}}$$

Then we take second derivative directly and get

$$\frac{\partial^2 f}{\partial w_k \partial w_j} = \sum_{i=1}^{m} \frac{y_i^2 x_{i_k} x_{i_j} e^{-y_i \boldsymbol{x_i^T w}}(1 + e^{-y_i \boldsymbol{x_i^T w}}) - y_i^2 x_{i_k} x_{i_j} e^{-y_i \boldsymbol{x_i^T w}} e^{-y_i \boldsymbol{x_i^T w}}}{(1 + e^{-y_i \boldsymbol{x_i^T w}})^2}$$

$$\frac{\partial^2 f}{\partial w_k \partial w_j} = \sum_{i=1}^{m} \frac{e^{-y_i \boldsymbol{x_i^T w}}}{(1 + e^{-y_i \boldsymbol{x_i^T w}})^2} x_{i_k} x_{i_j}$$

$$\frac{\partial^2 f}{\partial w_k \partial w_j} = \sum_{i=1}^{m} \sigma'(y_i \boldsymbol{x_i^T w}) x_{i_k} x_{i_j}$$

And that is the element of the k-th row and j-th column in the Hessian Matrix. If we write that in the matrix form we can get the Hessian Matrix. Which is

$$\nabla^2 f(\boldsymbol{w}) = \sum_{i=1}^{m} \sigma'(y_i \boldsymbol{x_i^T w}) \boldsymbol{x_i} \boldsymbol{x_i^T}$$

(b)

The optimal $\boldsymbol{w}^*$ given by damped Newton's method is $(-1.47, 4.44, -4.38)$ and the optimal value is 2.88. The number of iterations in the outer loop and the total number of iterations in the inner loop are both 9. The gaps and stepsizes are below.
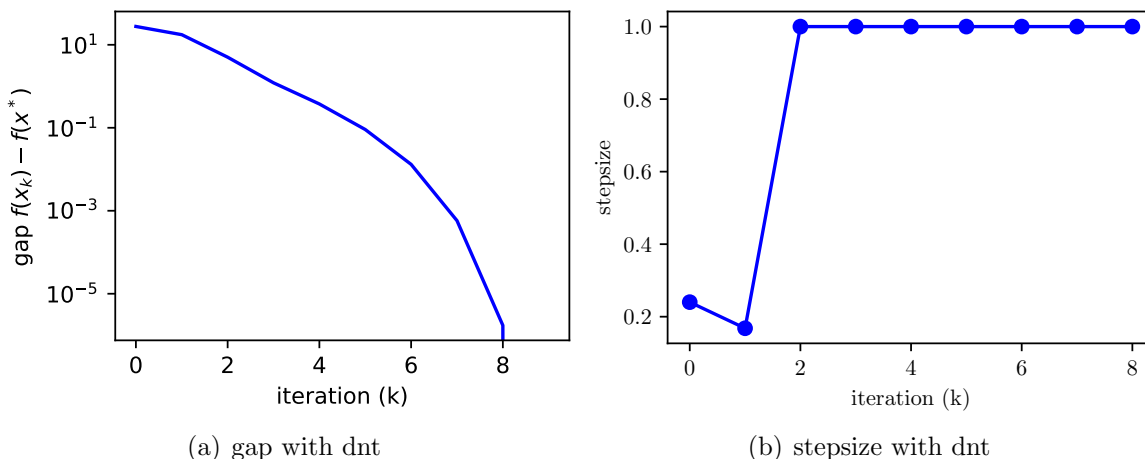
(a) gap with dnt

(b) stepsize with dnt

Figure 3: gap and stepsize with damped Newton's method for part (b)

(c)

If we use pure Newton's method with the same initial point we cannot do gradient descent because the Hessian Matrix at some point is nearly not invertible where the Newton step goes to infinity, consequently we cannot do the iteration.

But we must note that this does not happen for all initial points. If we choose some other initial point, pure Newton's method actually works. For example if we choose initial point $\boldsymbol{w}_0 = (-1.0, 3.0, -3.0)$, pure Newton's method gives us exactly the same optima as damped Newton's method with only 6 iterations.(Can check the result by running q2.py)

3

(a)

Since this is a univariate funtion, we take the derivatives. $f'(x) = 4(x - a)^3$ and $f''(x) = 12(x - a)^2$. Then the Newton step is simply $x_{k+1} = \frac{2}{3}x_k + \frac{1}{3}a$.

(b)

$y_{k+1} = |x_{k+1} - a| = |\frac{2}{3}x_k + \frac{1}{3}a - a| = |\frac{2}{3}x_k - \frac{2}{3}a| = \frac{2}{3}y_k$

(c)

From $y_{k+1} = \frac{2}{3}y_k$ we can easily get $y_k = (\frac{2}{3})^k y_0$ or equivalently $|x_k - a| = (\frac{2}{3})^k |x_0 - a|$. So x_k goes to a exponentially fast as k goes to infinity.
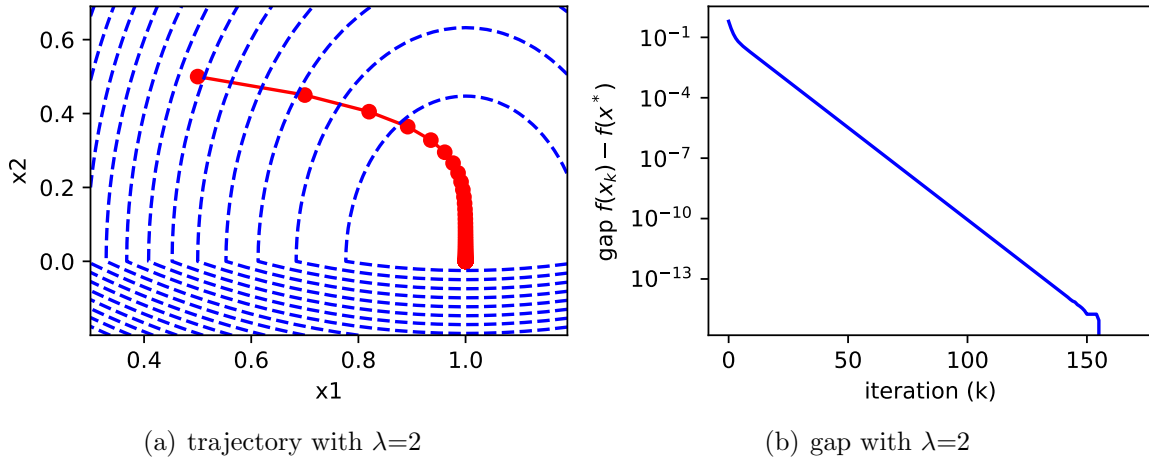
4

(a)

(a) trajectory with $\lambda=2$ (b) gap with $\lambda=2$

Figure 4: ISTA trajectory and gap for part (a)

The optima is $\boldsymbol{w}^* = (1, 0)$ with function value 6.5. The number of iterations is 169.

(b)

The optima is $\boldsymbol{w}^* = (1.25, 1)$ with function value 4.875. The number of iterations is 169. We can see that with a smaller λ there are no zero components in $\boldsymbol{w}^*$.

(c)

The optima is $\boldsymbol{w}^* = (0, 0)$ with function value 8.5. The number of iterations is 38. We can see that both components of $\boldsymbol{w}^*$ are zero.

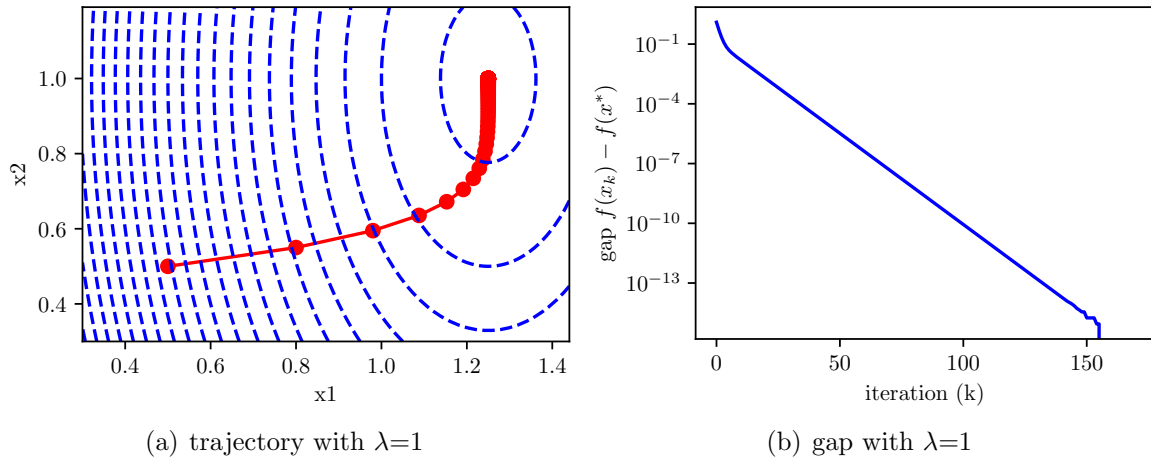From this problem we can see $\boldsymbol{w}^*$ gets more zero components as λ goes larger.

(a) trajectory with $\lambda=1$ (b) gap with $\lambda=1$

Figure 5: ISTA trajectory and gap for part (b)

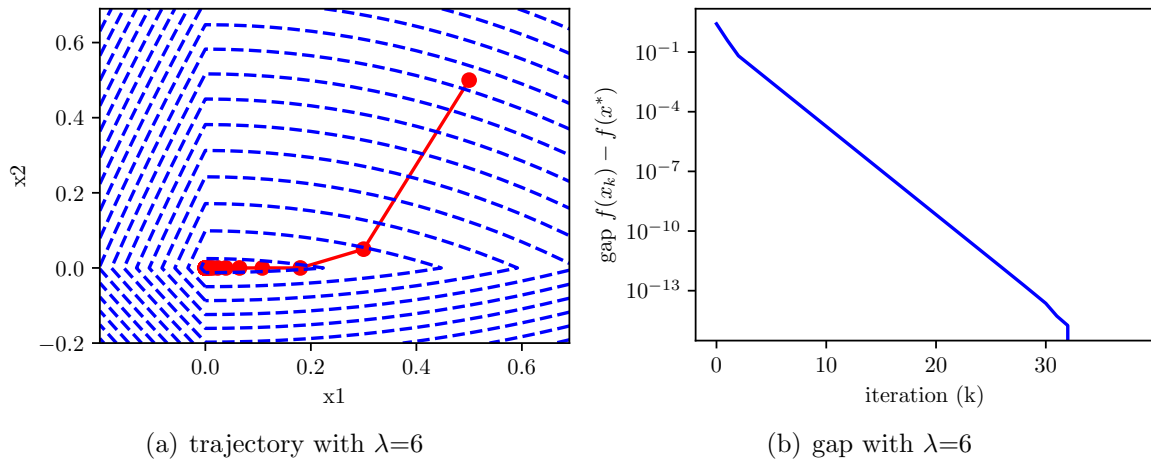

(a) trajectory with $\lambda=6$ (b) gap with $\lambda=6$

Figure 6: ISTA trajectory and gap for part (c)