ollowing were the results for the three methods:
[N: size of data, d: no. of dimensions]

Method 1 : take $\eta = 10^{-5}$

No. of iterations required = 873

Time complexity per iteration:

→ time to calculate derivative = $O(N \cdot d)$

~~Space~~ Total = $O(N \cdot d)$ time

Space complexity per iteration:

→ To store the derivative : $O(N \cdot d)$

Total = $O(N \cdot d)$ space.

Method 2 : optimal $\eta = \dfrac{\|\nabla J\|^2}{\nabla J^T H \nabla J}$

No. of iterations required = 265

Time complexity per iteration:

→ time to calculate derivative $O(N \cdot d)$

→ time to calculate H $O(Nd^2)$.

→ time to calculate denominator $O(Nd^2)$

Total = $O(Nd^2)$

Space complexity per iteration

→ to store. ∇J $O(Nd)$

→ to store H $O(d^2)$

→ to store $\nabla J^T H$ (intermediate), $O(d^2)$

Total complexity = $O(Nd + d^2)$

Method 3: $S = -H^{-1} \nabla J$

~~Time complexity~~

No. of iterations required $= 2$

Time complexity $\Rightarrow$

- for calculating $H^{-1} = O(d^3)$
- for calculating $H = O(nd^2)$
- for calculating $\nabla J = O(nd)$

Total $O(nd^2 + d^3)$ ~~com~~ time

Space complexity

- for $\nabla J = O(nd)$
- for $H^{-1} = O(d^2)$

Total $O(nd + d^2)$.