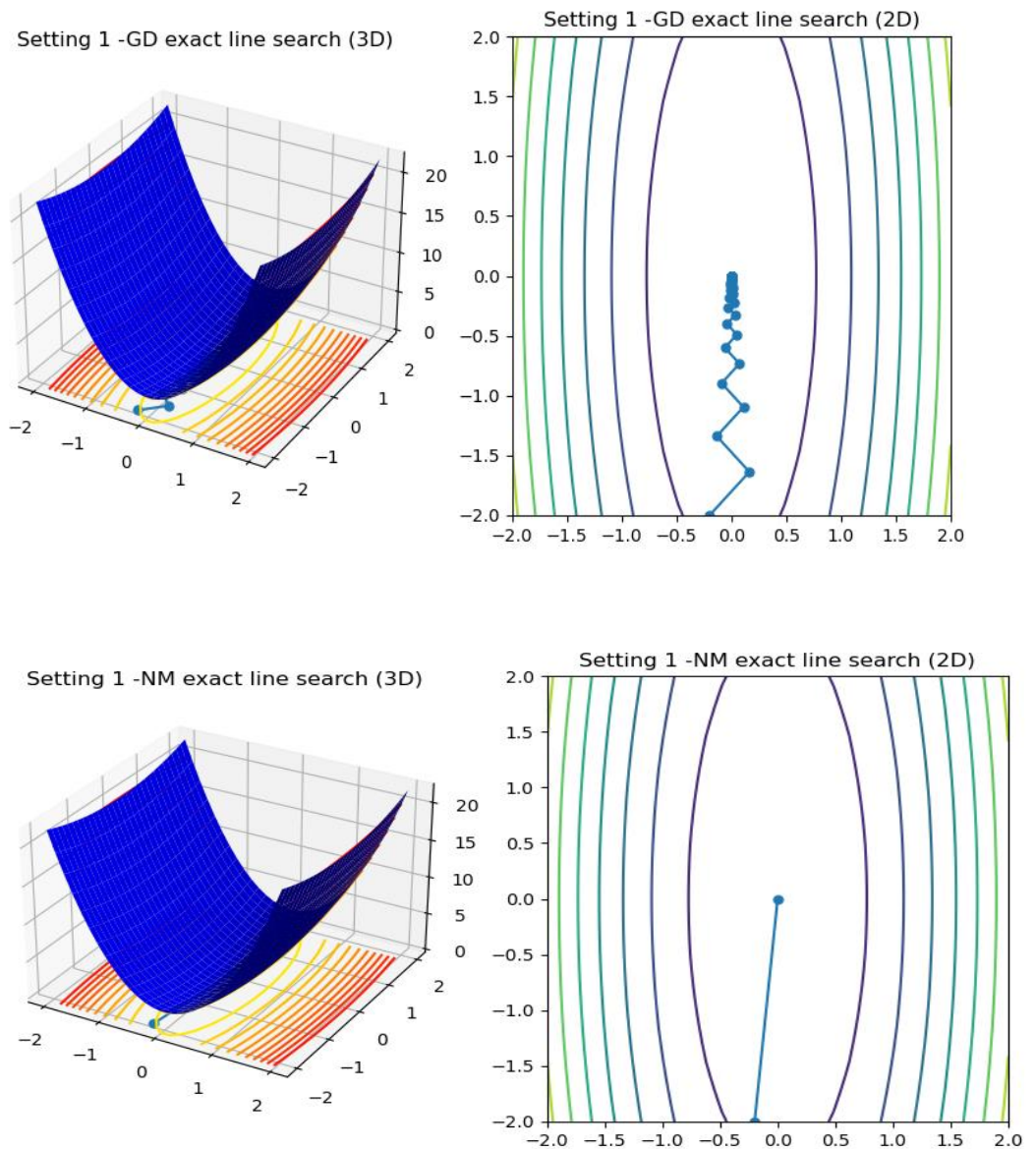


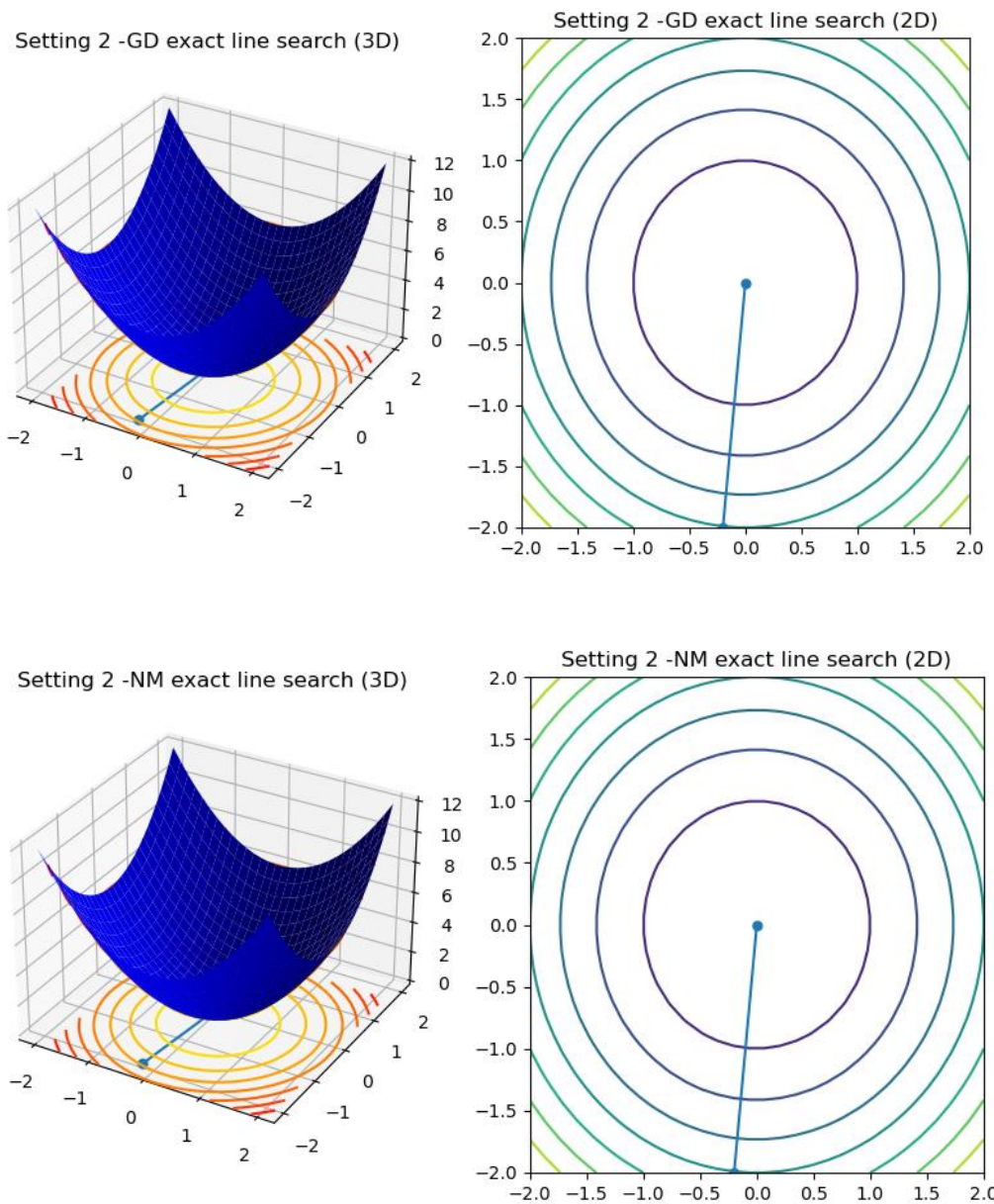
Setting 1:



We see the function is R^2 parabola, the gradient at each point is vertical to the function value, thus gradient descent takes more steps than Newton methods which uses the curvature of the function, the step size is optimal using exact line search.

The ratio between the largest and smallest eigenvalues of Q determines the convergence rate and it is unfortunately $10/1$.

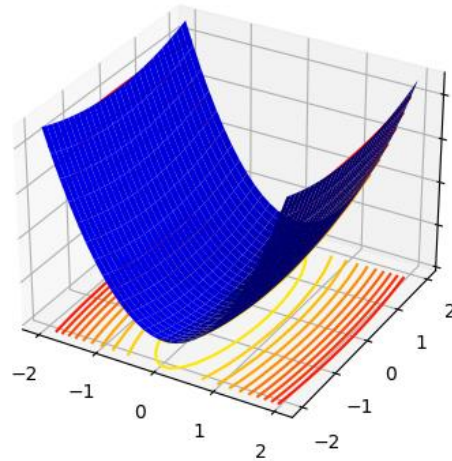
Setting 2:



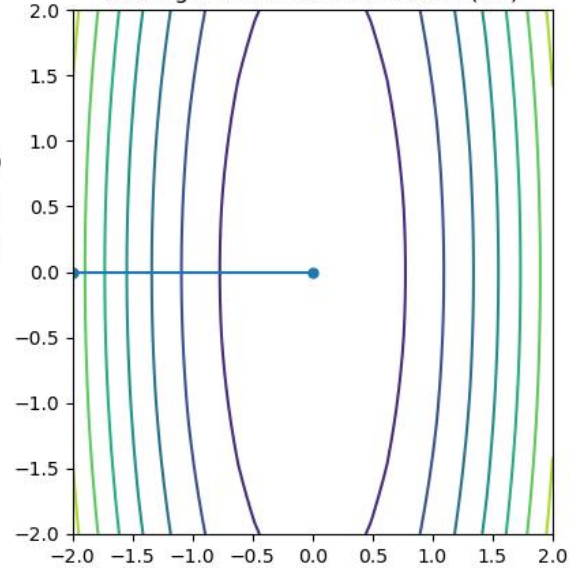
This time the ratio between the eigenvalues of Q is exactly 1, thus gradient descent has quick convergence, newton method is not beneficial this time (note: we are still using exact line search here and in setting 3)

Setting 3:

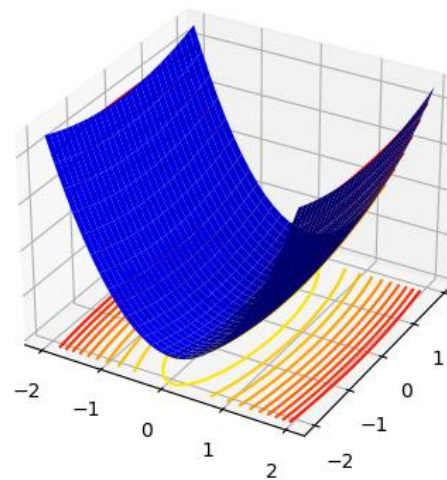
Setting 3 -GD exact line search (3D)



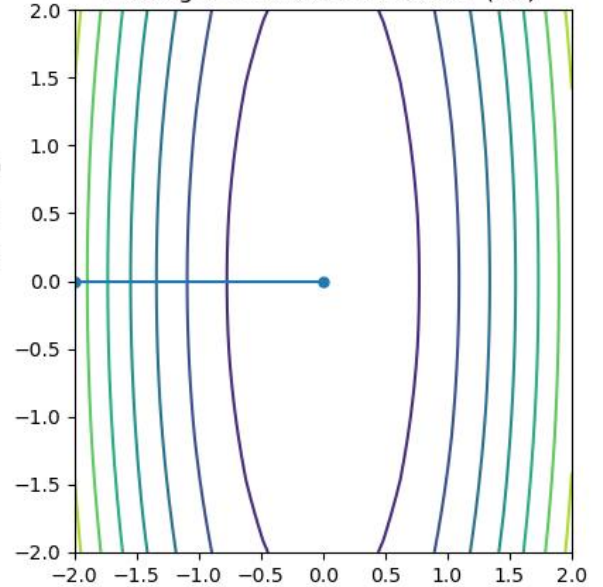
Setting 3 -GD exact line search (2D)



Setting 3 -NM exact line search (3D)



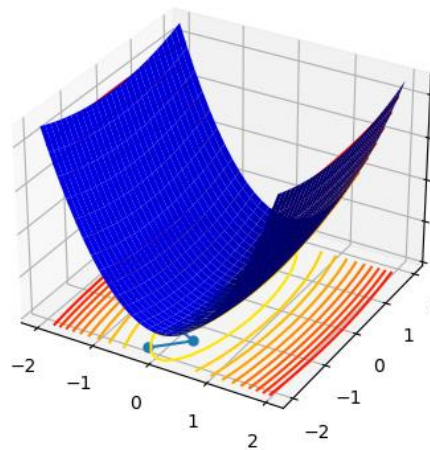
Setting 3 -NM exact line search (2D)



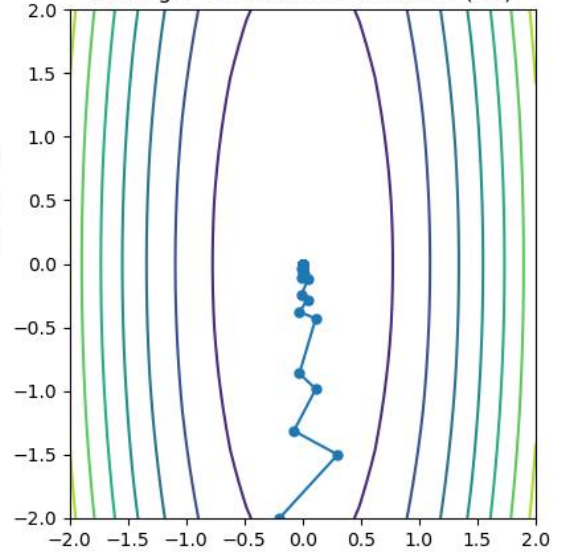
This time we got lucky, the gradient at the starting point allows us to get to the optimum using one iteration, the line search gets us to optimum immediately.

Setting 4:

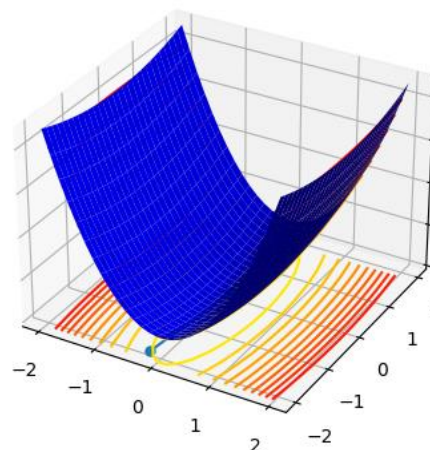
Setting 4 -GD inexact line search (3D)



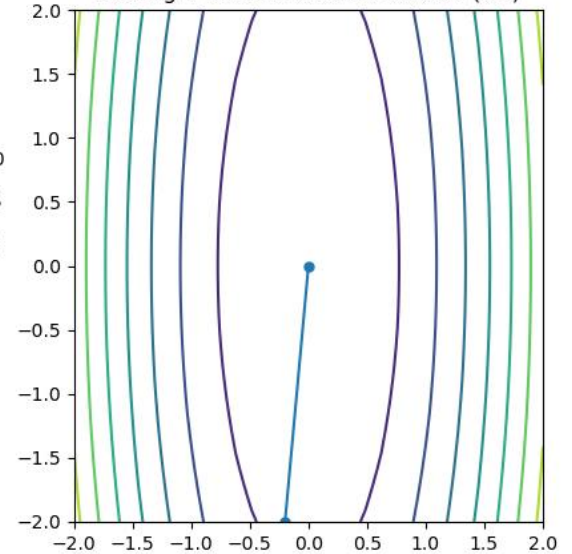
Setting 4 -GD inexact line search (2D)



Setting 4 -NM inexact line search (3D)

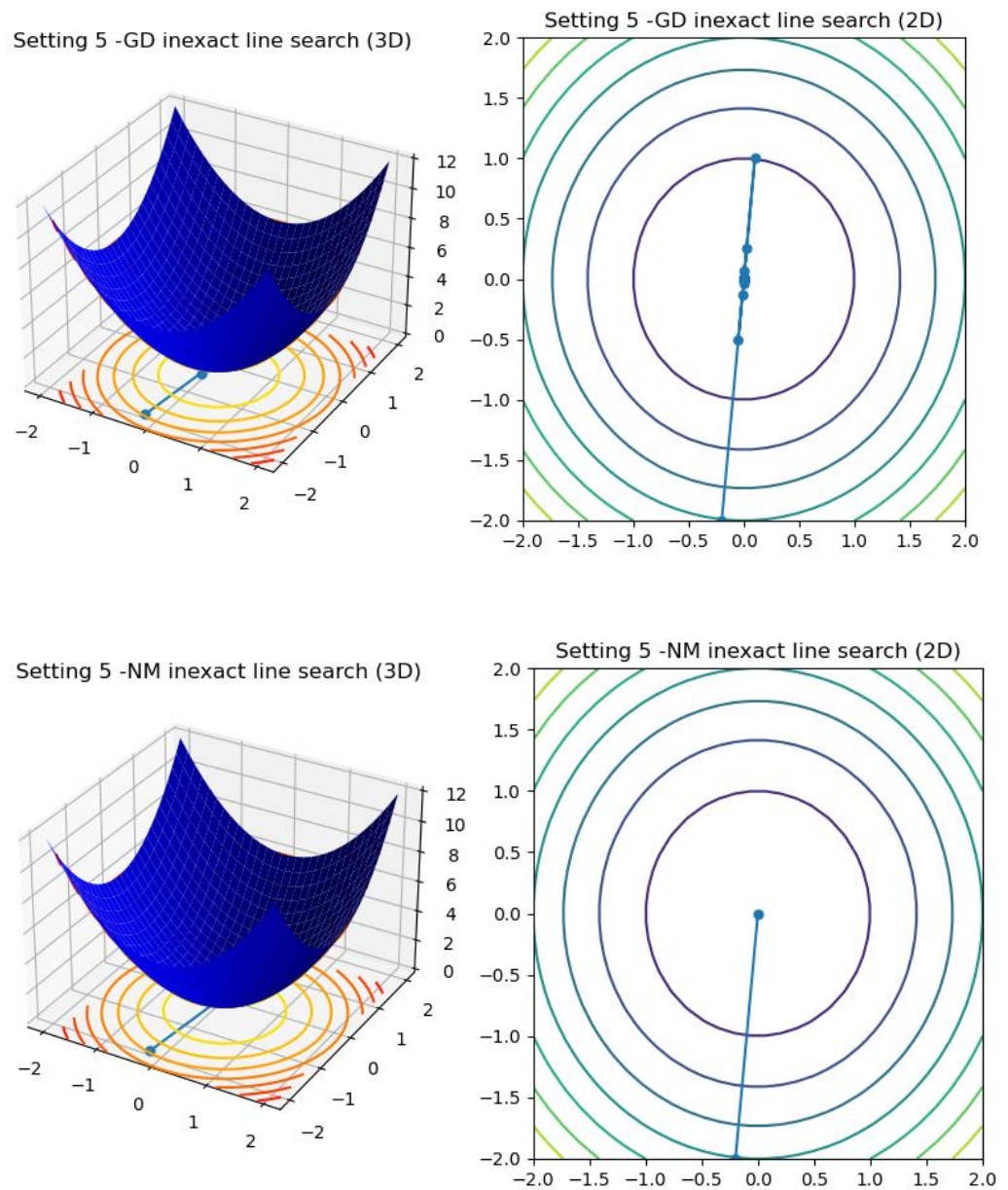


Setting 4 -NM inexact line search (2D)



This time we use inexact line search, thus the learning rate is not optimal for GD , thus it takes even more time. Newton method finds optimum with a single step because of its concern to the function curvature.

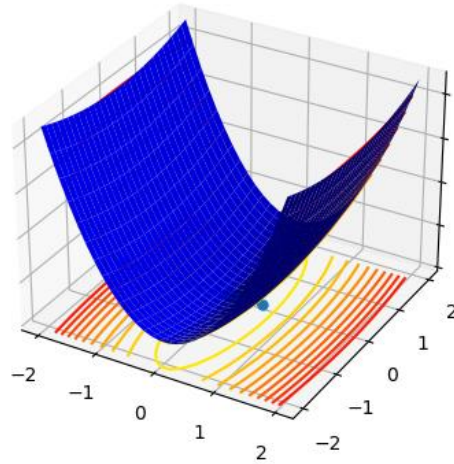
Setting 5:



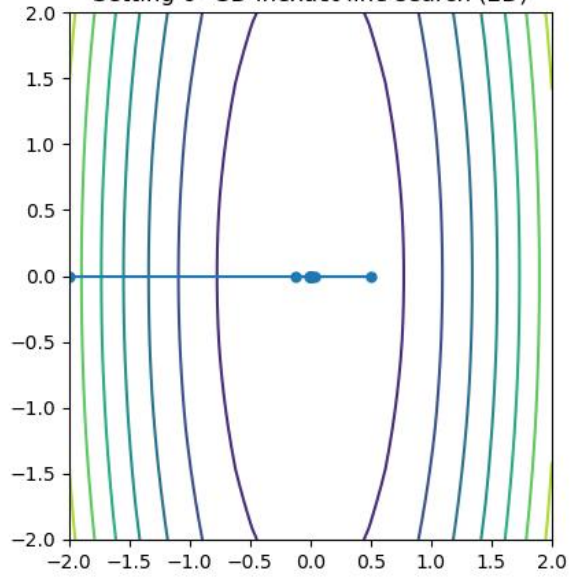
This time the inexact line search makes GD miss the optimum and takes more time to converge. NM finds optimum much easier as the gradient is optimal and line search finds the optimal learning rate easily.

Setting 6:

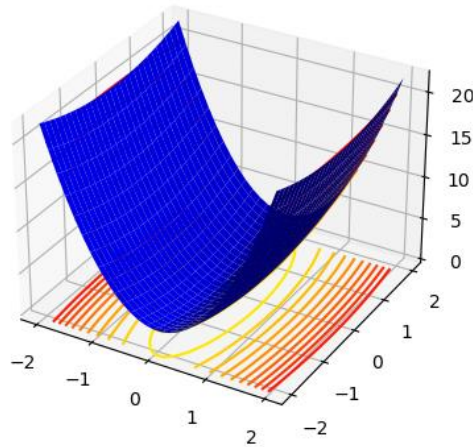
Setting 6 -GD inexact line search (3D)



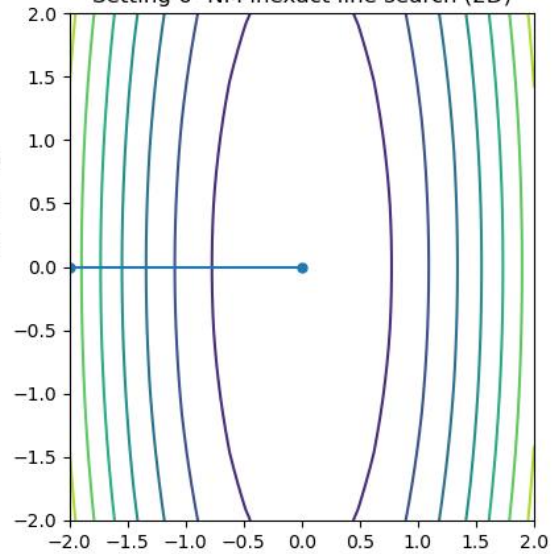
Setting 6 -GD inexact line search (2D)



Setting 6 -NM inexact line search (3D)

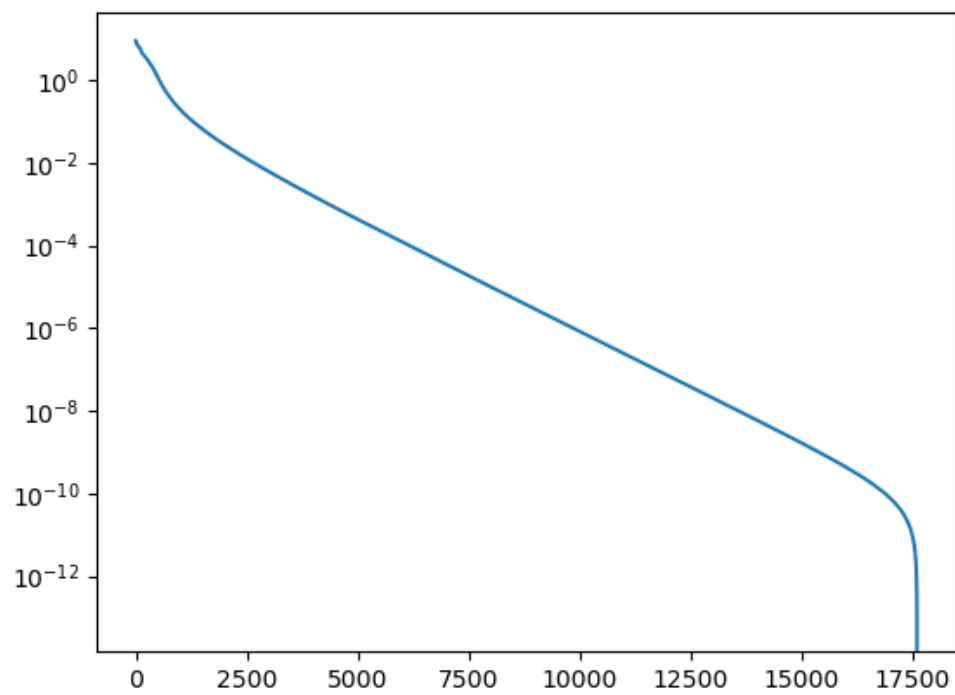


Setting 6 -NM inexact line search (2D)

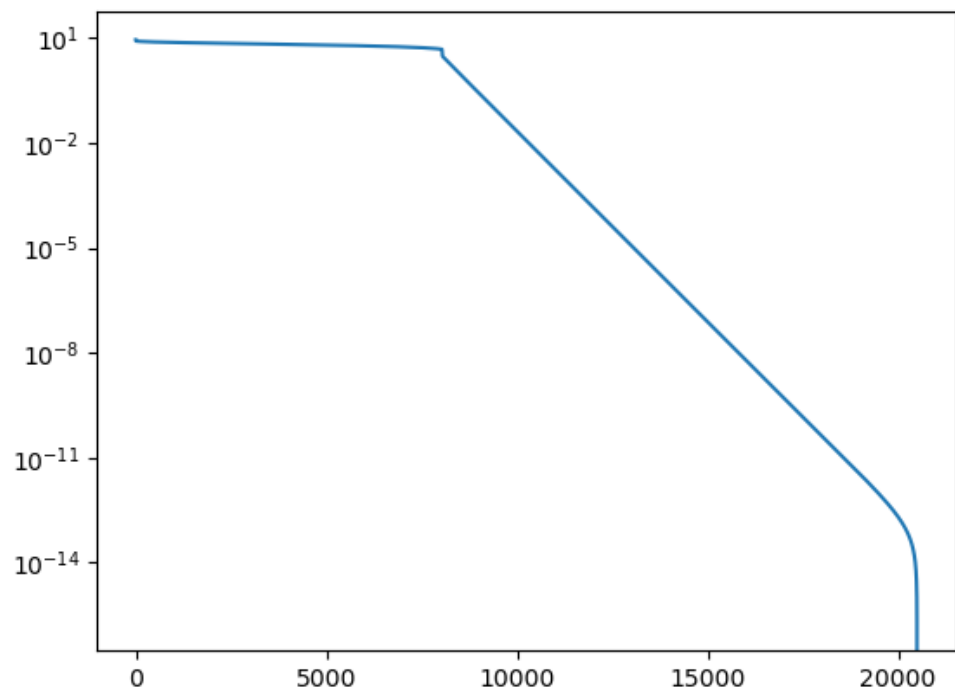


Again inexact line search makes our learning rate not optimal, NM fixes that again using better gradient.

Rosenbrock GD:



Rosenbrock NM:



The gradient descent searches for steepest descent immediately, thus it starts to converge fast, but the descent later although fast is slower than the descent of the Newton method, this method searches for good curvature, thus it starts by not descending the fastest, but searching for very steep descent which it finds thousands of iterations later, than it starts to converge very fast.

Overall NM takes more steps than GD, this is possible as rosenbrock function is not convex.