1 Introduction to Gradient Descent II

1.1 Old Methods of Convex Optimisation

Assume that a convex function F is given, with the property that $F(x^N) - F_0 \leq \epsilon$, where N is the number of required iterations, corresponding to the number of steps required for the computation of $\partial F(x)$ or separation between the hyperplane and point Q, and suppose that a compact set generated by F is defined.

The problem is that finding the centre of mass is a computationally expensive operation. However, Lee Y.-Y., Sidford A., and Wong S.C-W.(2015) have shown that not everything is lost, and this method can still be promising.

1.2 Gradient Descent

We have already seen that the method of gradient descent is defined by the equation $x^{k+1} = x^k - h\nabla f(x^k)$.

Assuming that $F(x^k) - F_0 \leq \epsilon$, naive dimension analysis shows that $h = c\frac{\epsilon}{M^2}$.

Gradient descent is not the go-to method for problems requiring precision or if the dimensionality is low.

1.3 Key Words

- Lyapunov function
- Grunbaum-Kruger Theorem
- Restarts
- Tikhonov regularisation