

极小值很近的一点。因此，相比于共轭梯度，BFGS的优点是其花费更少的时间改进每个线性搜索。在另一方面，BFGS算法必须存储Hessian逆矩阵  $\mathbf{M}$ ，需要  $O(n^2)$  的存储空间，使BFGS不适用于大多数具有百万级参数的现代深度学习模型。

