

## 4.3

1

ステップごとのパラメータの変化量はステップサイズによって決定される。そのため、ステップサイズが小さければ小さいほど一回のステップでの更新量が小さく、最適解付近に到達するまで時間がかかり、ステップサイズが大きいほど時間がかからないと考えられる。

2

GDでは、全データの二乗誤差と正則化の和で表される目的関数について勾配方向にパラメータを更新しているため、ステップごとに目的関数の値が小さくなることが保証される。一方で、SGDでは、ランダムに選んだ単一のデータについてのみ勾配を計算するため、長期的に見れば目的関数は減少することが期待されるが、各更新の方向は必ずしも目的関数を減少させる方向であるとは限らないため、単調減少しない。

3

SDGでは一回の更新で必ず目的関数が減少をすることはできないので、GDの収束判定法を用いるとアルゴリズムがすぐ停止してしまう。方法として、過去の最小の誤差を記憶しておき、目的関数が $n$ 回連続で改善されなければ収束したとしてアルゴリズムを停止する、というものを考えた。