最適化手法の実装（1）：Github 公開用

最終更新：2021 年 9 月 20 日

バックトラック直線探索を用いた勾配法を実装し，以下の 2 変数関数を最小化することを考える。

$$f(\boldsymbol{x}) = \frac{1}{2}{x_1}^4 - 2{x_1}^2 x_2 + 4{x_2}^2 + 8x_1 + 8x_2 \tag{1}$$

大学 3 年生の時に受講した授業で、バックトラック直線探索を用いた勾配法を実装する課題が出された。同じ問題を使用することはできないので、ここでは、その問題と類似する問題である、文献 [1] の例 2.5 の問題を例に実装方法を説明する。

最適化問題を解くシンプルな方法に、**勾配降下法（gradient descent method）**である。この方法は深層学習でも利用されている。現在の点 $\boldsymbol{x}_k$ を更新公式

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \alpha_k \boldsymbol{d}_k \tag{2}$$

に従って更新する。この際、$\boldsymbol{d}_k$ として、

$$\boldsymbol{d}_k = -\nabla f(\boldsymbol{x}_k) \tag{3}$$

を選択する方法は**最急降下法（steepest descent method）**と呼ばれる。ここで、注意することとしては、目的関数 $f(\cdot)$ に対して、

$$f(\boldsymbol{x}_k) > f(\boldsymbol{x}_{k+1}) \tag{4}$$

が成立することが求められる。なぜなら、α_k が大きすぎると、$\boldsymbol{x}_k$ において、降下方向（f が減少する方向）をとっても、式 (4) が成立しない可能性があるからである。そこで、定数 $c \in (0, 1)$ に対して、

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) + c\langle \nabla f_k, \alpha \boldsymbol{d}_k \rangle \tag{5}$$

を満たす $\alpha > 0$ を α_k として使用する。この際、用意した α が式 (5) を満たさない場合に、α の代わりに $\rho\alpha$ $(0 < \rho < 1)$ に置き換えて、再び式 (5) が成立するかどうかを確認することを繰り返しながら、適切な α_k を見つける方法を**バックトラック法**という。

本問題では、このバックトラック法により、適切な α_k を決めてから、最急降下方向に点を更新することを繰り返して、最適解を求めることを試みた。

式 (1) より、

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} 2{x_1}^3 - 4x_1 x_2 + 8 \\ -2{x_1}^2 + 8x_2 + 8 \end{pmatrix} \tag{6}$$

である。また、最急降下法を使用する際、式 (3) を式 (5) に代入することで、適切な α_k であるかどうかを判定する式は

$$f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) \leq f(\boldsymbol{x}_k) - c\alpha \|\nabla f(\boldsymbol{x}_k)\|^2 \tag{7}$$

となる。

ここまでの情報をもとに Python で、バックトラック直線探索を用いた勾配法を実装した．その際のソースコードが `backtrack.py` である。今回は初期点を $\boldsymbol{x}_0 = (3, 1)^\top$ とし、上記の説明に出てきたパラメータ

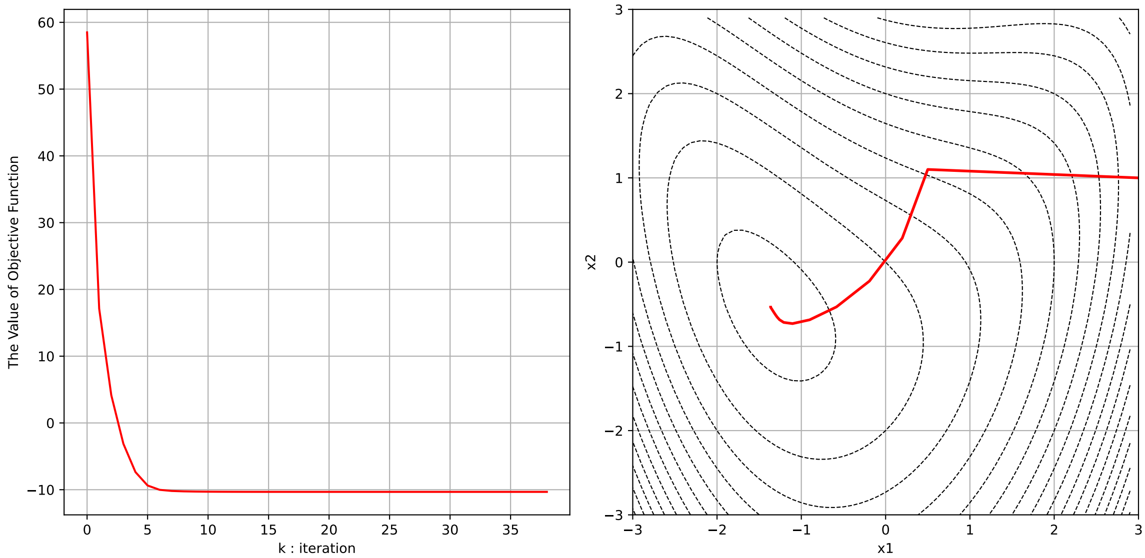

図 1: 実行結果：左図は目的関数 $f(\boldsymbol{x})$ が $\boldsymbol{x}_k$ の反復更新とともにどのように減少するかを表す。一方、右図は反復更新とともに $\boldsymbol{x}_k$ がどのように移動するかを表す。

を $\alpha = 0.05,\ c = 0.01,\ \rho = 0.8$ とした。また、アルゴリズムの収束判定条件を $f(\boldsymbol{x_{k+1}}) - f(\boldsymbol{x_k}) < 10^{-8}$ とした(1)。

実行結果は図 1 のようになった。収束値は $\boldsymbol{x}^* = (-1, 3646, -0.5345)^\top$ で、そのときの目的関数の値は $f(\boldsymbol{x}^*) = -10.3256$ であった。この結果は文献 [1] の例 2.5 に記されている値と一致している。

参考文献

[1] 寒野善博，土谷隆．東京大学工学教程 基礎系 数学 最適化と変分法（第 5 刷）．丸善出版，2018

[2] 金森敬文，鈴木大慈，竹内一郎，佐藤一誠．機械学習のための連続最適化（第 2 刷）．講談社, 2017

[3] 寒野善博．最適化手法入門．講談社, 2019

(1) 文献 [2, 3] では、アルゴリズムの反復を終了する条件として,「十分小さい ε に対して，$\|\nabla f(\boldsymbol{x_k})\| \leq \varepsilon$」とすることが紹介される。ただ、直線探索では，$\boldsymbol{x_{k+1}} = \boldsymbol{x_k} - \varepsilon_k \nabla f(\boldsymbol{x_k})$ (ε_k はステップ幅) により，点列 $\{\boldsymbol{x_k}\}$ を更新するため,

$$\begin{aligned} \|\nabla f(\boldsymbol{x_k})\| \text{ が十分小さい} \quad &\Longleftrightarrow \quad \boldsymbol{x_{k+1}} \approx \boldsymbol{x_k} \\ &\Longleftrightarrow \quad f(\boldsymbol{x_{k+1}}) - f(\boldsymbol{x_k}) \text{ が十分小さい} \end{aligned}$$

という関係が成り立つと考え，$f(\boldsymbol{x_{k+1}}) - f(\boldsymbol{x_k}) < 10^{-8}$ を反復を終了させるかどうかの判定条件とすることは問題ないと考えている。