知能システム論第 5 回課題

37186305
航空宇宙工学専攻修士一年
荒居秀尚

2018 年 11 月 3 日

1　宿題 1

$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ のとき、$\alpha = 1/2$ のアルミホ規準は

$$\begin{aligned}
g(\epsilon_k) &= \frac{1}{2}(\boldsymbol{x} - \epsilon_k\boldsymbol{A}\boldsymbol{x})^T\boldsymbol{A}(\boldsymbol{x} - \epsilon_k\boldsymbol{A}\boldsymbol{x}) - \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} && (1.1)\\
&= \frac{1}{2}\boldsymbol{x}^T(\boldsymbol{I} - \epsilon_k\boldsymbol{A})^T\boldsymbol{A}(\boldsymbol{I} - \epsilon_k\boldsymbol{A})\boldsymbol{x} - \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} && (1.2)\\
&= -\frac{1}{2}\epsilon_k\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} - \frac{1}{2}\epsilon_k(\boldsymbol{x}^T\boldsymbol{A}^2\boldsymbol{x} - \epsilon_k\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}^2\boldsymbol{x}) && (1.3)\\
&\le \alpha\epsilon_k g'(0) = -\frac{1}{2}\epsilon_k\boldsymbol{x}^T\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} && (1.4)
\end{aligned}$$

これを満たす最大の ϵ_k は等号が成り立つときに得られるため、$\boldsymbol{A}$ が正定値対称行列であるとすれば、

$$\epsilon_k = \frac{\boldsymbol{x}^T\boldsymbol{A}^2\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{A}^3\boldsymbol{x}} \tag{1.5}$$

2　宿題 2

$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ に対する厳密直線探索を用いた最急降下法において

$$\begin{aligned}
\boldsymbol{x}_{k+1} &= \boldsymbol{x} - \epsilon_k\nabla f(\boldsymbol{x}_k) && (2.1)\\
\nabla f(\boldsymbol{x}) &= \boldsymbol{A}\boldsymbol{x} && (2.2)\\
\epsilon_k &= \frac{\|\nabla f(\boldsymbol{x}_k)\|^2}{\nabla f(\boldsymbol{x}_k)^T\boldsymbol{A}\nabla f(\boldsymbol{x}_k)} && (2.3)
\end{aligned}$$

となる。簡単のため $\nabla f(\boldsymbol{x}_k) = \boldsymbol{g_k}$ と表記すると

$$\begin{aligned}
f(\boldsymbol{x}_{k+1}) &= \frac{1}{2}\left(\boldsymbol{x}_k - \frac{\boldsymbol{g}_k{}^T\boldsymbol{g}_k}{\boldsymbol{g}_k{}^T\boldsymbol{A}\boldsymbol{g}_k}\boldsymbol{g}_k\right)^T\boldsymbol{A}\left(\boldsymbol{x}_k - \frac{\boldsymbol{g}_k{}^T\boldsymbol{g}_k}{\boldsymbol{g}_k{}^T\boldsymbol{A}\boldsymbol{g}_k}\boldsymbol{g}_k\right) && (2.4)\\
&= \frac{1}{2}\boldsymbol{x}_k\boldsymbol{A}\boldsymbol{x}_k - \frac{2\left(\boldsymbol{g}_k{}^T\boldsymbol{g}_k\right)\left(\boldsymbol{g}_k{}^T\boldsymbol{A}\boldsymbol{x}_k\right) - \left(\boldsymbol{g}_k{}^T\boldsymbol{g}_k\right)^2}{2\boldsymbol{g}_k{}^T\boldsymbol{A}\boldsymbol{g}_k} && (2.5)
\end{aligned}$$

ここで、$\boldsymbol{g}_k = \boldsymbol{A}\boldsymbol{x}_k$ であることから、上式の第二項は

$$\frac{2\left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)\left({\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{x}_k\right) - \left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)^2}{2{\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{g}_k} = \frac{2\left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)^2 - \left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)^2}{2{\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{g}_k} = \frac{\left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)^2}{2{\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{g}_k} \tag{2.6}$$

したがって

$$f(\boldsymbol{x}_{k+1}) = \left(1 - \frac{\left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)^2}{\left({\boldsymbol{x}_k}^T\boldsymbol{A}\boldsymbol{x}_k\right)\left({\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{g}_k\right)}\right) f(\boldsymbol{x}_k) \tag{2.7}$$

$$= \left(1 - \frac{\left({\boldsymbol{g}_k}^T\boldsymbol{g}_k\right)^2}{\left({\boldsymbol{g}_k}^T\boldsymbol{A}^{-1}\boldsymbol{g}_k\right)\left({\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{g}_k\right)}\right) f(\boldsymbol{x}_k) \tag{2.8}$$

カントロビッチの不等式において $c = \frac{1-\lambda_{min}/\lambda_{max}}{1+\lambda_{min}/\lambda_{max}}$ とおくと

$$\frac{4\lambda_{min}\lambda_{max}}{(\lambda_{max}+\lambda_{min})^2} = 1 - c^2 \leq \frac{\|\boldsymbol{g}_k\|^4}{{\boldsymbol{g}_k}^T\boldsymbol{A}^{-1}\boldsymbol{g}_k \cdot {\boldsymbol{g}_k}^T\boldsymbol{A}\boldsymbol{g}_k} \tag{2.9}$$

これにより

$$f(\boldsymbol{x}_{k+1}) \leq c^2 f(\boldsymbol{x}_k) \tag{2.10}$$

が成り立つ。

3 宿題 3

Julia 1.0.0 を用いて実装した。

```
using Plots

# Functions
f(x, y) = 10x^2 + y^2
df(x) = [20x[1], 2x[2]]

function backtrack(x, epsilon, alpha, beta)
    x_new = x - epsilon .* df(x)
    while f(x_new[1], x_new[2]) - f(x[1], x[2]) > -alpha * epsilon * sum(df(x
        ).^2)
        epsilon *= beta
        x_new = x - epsilon .* df(x)
    end
    x_new
end

# Variables
x = y = range(-5, stop=5, length=100)
p = plot(x, y, f, st = [:contourf])

niter = 5
pos = 10 * rand(2) .- 5
```

```
epsilon = 1.0
alpha = 0.5
beta = 0.8
previous = [0.0, 0.0]
for i in 1:niter
    previous = copy(pos)
    pos = backtrack(pos, epsilon, alpha, beta)
    plot!(p[1], [previous[1], pos[1]], [previous[2], pos[2]], line = (:white, 1))
end
```

結果は以下の図 3.1 のようになった。

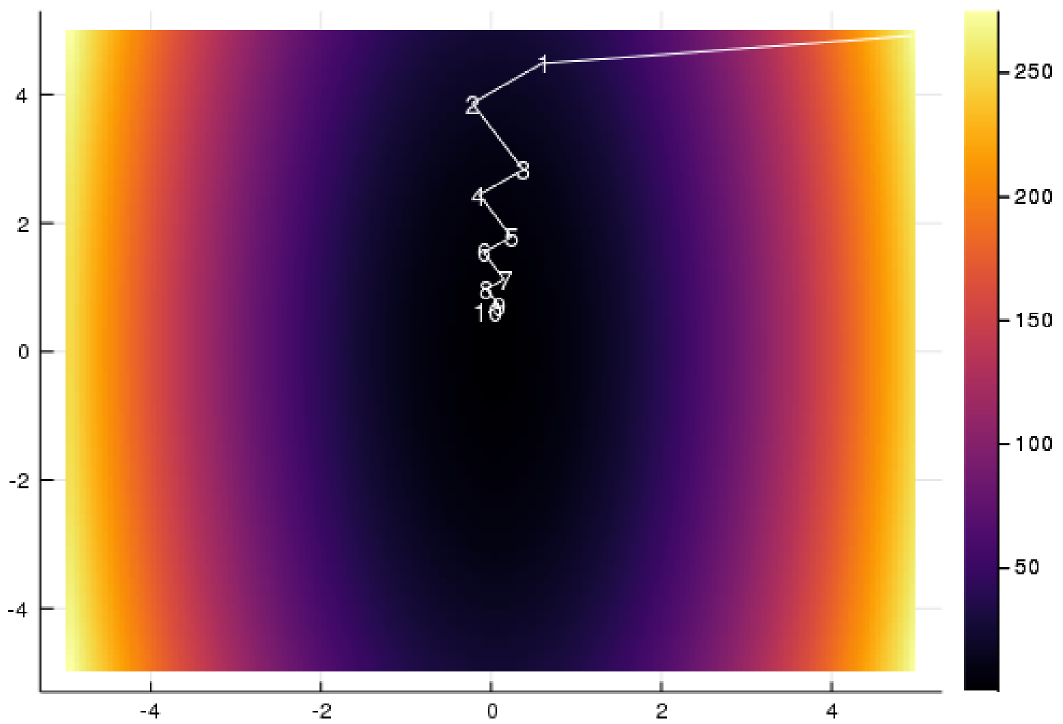

図 3.1: バックトラック直線探索を用いた最急降下法

4　宿題 4

4.1　A

BFGS アルゴリズムのヘッセ行列の更新式において

$$\boldsymbol{H}_{k+1} = \boldsymbol{H}_k + \frac{\boldsymbol{t}_k \boldsymbol{t}_k{}^T}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} - \frac{\boldsymbol{H}_k \boldsymbol{s}_k \boldsymbol{s}_k{}^T \boldsymbol{H}_k}{\boldsymbol{s}_k{}^T \boldsymbol{H}_k \boldsymbol{s}_k} \tag{4.1}$$

$$= \boldsymbol{B} - \frac{\boldsymbol{H}_k \boldsymbol{s}_k \boldsymbol{s}_k{}^T \boldsymbol{H}_k}{\boldsymbol{s}_k{}^T \boldsymbol{H}_k \boldsymbol{s}_k} \tag{4.2}$$

のように第 1 項と第 2 項をまとめて $\boldsymbol{B}$ と呼ぶことにする。また、スカラー量 $\boldsymbol{s}_k{}^T\boldsymbol{t}_k$ を a と表記することにする。まず、$\boldsymbol{B}$ の逆行列についてシャーマン・モリソン公式を用いると

$$\boldsymbol{B} = \boldsymbol{H}_k + \frac{\boldsymbol{t}_k\boldsymbol{t}_k{}^T}{a} \tag{4.3}$$

なので

$$\boldsymbol{B}^{-1} = \boldsymbol{H}_k{}^{-1} - \frac{\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{a + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k} \tag{4.4}$$

これを用いて

$$\begin{aligned}
\boldsymbol{H}_{k+1}{}^{-1} &= \boldsymbol{B}^{-1} + \frac{\boldsymbol{B}^{-1}\boldsymbol{H}_k\boldsymbol{s}_k\boldsymbol{s}_k{}^T\boldsymbol{H}_k{}^T\boldsymbol{B}^{-1}}{\boldsymbol{s}_k{}^T\boldsymbol{H}_k\boldsymbol{s}_k - \boldsymbol{s}_k{}^T\boldsymbol{H}_k{}^T\boldsymbol{B}^{-1}\boldsymbol{H}_k\boldsymbol{s}_k} && (4.5)\\
&= \boldsymbol{B}^{-1} + \frac{\boldsymbol{B}^{-1}\boldsymbol{H}_k\boldsymbol{s}_k\boldsymbol{s}_k{}^T\boldsymbol{H}_k{}^T\boldsymbol{B}^{-1}}{\boldsymbol{s}_k{}^T\boldsymbol{H}_k\boldsymbol{s}_k - \boldsymbol{s}_k{}^T\boldsymbol{H}_k{}^T\left(\boldsymbol{H}_k{}^{-1} - \frac{\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{a+\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k}\right)\boldsymbol{H}_k\boldsymbol{s}_k} && (4.6)\\
&= \boldsymbol{B}^{-1} + \frac{\boldsymbol{B}^{-1}\boldsymbol{H}_k\boldsymbol{s}_k\boldsymbol{s}_k{}^T\boldsymbol{H}_k{}^T\boldsymbol{B}^{-1}}{\boldsymbol{s}_k{}^T\boldsymbol{H}_k\boldsymbol{s}_k - \boldsymbol{s}_k{}^T\boldsymbol{H}_k\boldsymbol{s}_k + \frac{a^2}{a+\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k}} && (4.7)\\
&= \boldsymbol{B}^{-1} + \frac{\left(a + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\right)\left(\boldsymbol{H}_k{}^{-1} - \frac{\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{a+\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k}\right)\boldsymbol{H}_k\boldsymbol{s}_k\boldsymbol{s}_k{}^T\boldsymbol{H}_k{}^T\left(\boldsymbol{H}_k{}^{-1} - \frac{\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{a+\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k}\right)}{a^2} && (4.8)\\
&= \boldsymbol{B}^{-1} + \frac{\left(a + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\right)\boldsymbol{s}_k\boldsymbol{s}_k{}^T - \left(\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{s}_k\boldsymbol{s}_k{}^T + \boldsymbol{s}_k\boldsymbol{s}_k{}^T\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\right)}{a^2} + \frac{\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{s}_k\boldsymbol{s}_k{}^T\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{a^2\left(a + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\right)} && (4.9)
\end{aligned}$$

$\boldsymbol{t}_k{}^T\boldsymbol{s}_k = \boldsymbol{s}_k{}^T\boldsymbol{t}_k = a$ なので

$$\begin{aligned}
\boldsymbol{H}_{k+1}{}^{-1} &= \boldsymbol{B}^{-1} + \frac{\left(a + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\right)\boldsymbol{s}_k\boldsymbol{s}_k{}^T - a\left(\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{s}_k{}^T + \boldsymbol{s}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\right)}{a^2} + \frac{a^2\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{a^2\left(a + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\right)} && (4.10)\\
&= \boldsymbol{H}_k{}^{-1} + \frac{\left(\boldsymbol{s}_k{}^T\boldsymbol{t}_k + \boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\right)\boldsymbol{s}_k\boldsymbol{s}_k{}^T}{(\boldsymbol{s}_k{}^T\boldsymbol{t}_k)^2} - \frac{\boldsymbol{H}_k{}^{-1}\boldsymbol{t}_k\boldsymbol{s}_k{}^T + \boldsymbol{s}_k\boldsymbol{t}_k{}^T\boldsymbol{H}_k{}^{-1}}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k} && (4.11)
\end{aligned}$$

となる。

4.2 B

$\boldsymbol{H}_k$ が正定値かつ $\boldsymbol{s}_k{}^T\boldsymbol{t}_k > 0$ のとき $\boldsymbol{H}_{k+1}$ が正定値であることを、$\boldsymbol{H}_{k+1}{}^{-1}$ の正定値性から示す。

$$\boldsymbol{H}_{k+1}{}^{-1} = \left(\boldsymbol{I} - \frac{\boldsymbol{s}_k\boldsymbol{t}_k{}^T}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k}\right)\boldsymbol{H}_k{}^{-1}\left(\boldsymbol{I} - \frac{\boldsymbol{t}_k\boldsymbol{s}_k{}^T}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k}\right) + \frac{\boldsymbol{s}_k\boldsymbol{s}_k{}^T}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k} \tag{4.12}$$

と表現できることを用いると零ベクトル出ない任意のベクトル z を用いた二次形式

$$z^T\boldsymbol{H}_{k+1}{}^{-1}z = z^T\left(\boldsymbol{I} - \frac{\boldsymbol{s}_k\boldsymbol{t}_k{}^T}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k}\right)\boldsymbol{H}_k{}^{-1}\left(\boldsymbol{I} - \frac{\boldsymbol{t}_k\boldsymbol{s}_k{}^T}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k}\right)z + z^T\frac{\boldsymbol{s}_k\boldsymbol{s}_k{}^T}{\boldsymbol{s}_k{}^T\boldsymbol{t}_k}z \tag{4.13}$$

ここで

$$\boldsymbol{z}^T \left(\boldsymbol{I} - \frac{[s]_k \boldsymbol{t}_k{}^T}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} \right) = \boldsymbol{z}^T - \frac{\boldsymbol{z}^T \boldsymbol{s}_k}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} \boldsymbol{t}_k{}^T \tag{4.14}$$

$$= \left(\boldsymbol{z} - \frac{\boldsymbol{s}_k{}^T \boldsymbol{z}}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} \boldsymbol{t}_k \right)^T \tag{4.15}$$

$$= \left(\left(\boldsymbol{I} - \frac{\boldsymbol{t}_k \boldsymbol{s}_k{}^T}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} \right) \boldsymbol{z} \right)^T \tag{4.16}$$

よって $\left(\boldsymbol{I} - \frac{\boldsymbol{t}_k \boldsymbol{s}_k{}^T}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} \right) \boldsymbol{z}$ を $\boldsymbol{z}'$ と表記することにすると、

$$\boldsymbol{z}^T \boldsymbol{H}_{k+1}{}^{-1} \boldsymbol{z} = \boldsymbol{z}'^T \boldsymbol{H}_k{}^{-1} \boldsymbol{z}' + \frac{\left(\boldsymbol{z}^T \boldsymbol{s}_k \right)^2}{\boldsymbol{s}_k{}^T \boldsymbol{t}_k} \tag{4.17}$$

$\boldsymbol{H}_k$ が正定値かつ $\boldsymbol{s}_k{}^T \boldsymbol{t}_k > 0$ のときこの値は正であることがわかり、$\boldsymbol{H}_{k+1}{}^{-1}$ は正定値、よって $\boldsymbol{H}_{k+1}$ も正定値である。