5.3-4 Newton 法，準 Newton 法

森田研輪読

髙橋 優輝

2023.11.22, 28

大阪大学大学院情報科学研究科

目次

5.3 Newton 法

導入

目的関数 f について，$f \in C^2(\mathbb{R}^d)$ とする．

本節では，局所的 2 次収束性を有する Newton 法について見る．

Newton 法は以下の設定 (S1)-(S3) を用いる勾配法である．

Definition (Newton 法の設定)

(S1) **$x^\star$ に十分近い初期解** $\boldsymbol{x}_0 \in \mathbb{R}^d$

(S2) **ステップサイズ** $\alpha_k := 1 \ (k \in \mathbb{N})$

(S3) **探索方向** $\boldsymbol{d}_k := -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k) \ (k \in \mathbb{N})$

探索方向の気持ち (1/2)

Taylor 展開を用いると，目的関数 f は $\boldsymbol{x}_k$ のまわりで次のように近似される．

$$f(\boldsymbol{x}) \approx f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{x} - \boldsymbol{x}_k \rangle_2 + \frac{1}{2}\langle \boldsymbol{x} - \boldsymbol{x}_k, \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x} - \boldsymbol{x}_k) \rangle$$

ここで，$\boldsymbol{d} := \boldsymbol{x} - \boldsymbol{x}_k$ として，任意の $\boldsymbol{d} \in \mathbb{R}^d$ に対して，関数 $q_k \colon \mathbb{R}^d \to \mathbb{R}$ を次のように定義する．（q_k は f の近似関数となっている．）

$$q_k(\boldsymbol{d}) := f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d} \rangle_2 + \frac{1}{2}\langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d} \rangle \tag{5.19}$$

f の最小化が目標であることを踏まえ，f の近似関数 q_k を $\boldsymbol{d}$ について最小化することを考える．

探索方向の気持ち (2/2)

式 (5.19)[1] について，$\nabla^2 f(\boldsymbol{x}_k) \in \mathbb{S}^d_{++}$ と仮定すると，q_k は凸関数となる．
命題 3.2.1(2) より，凸関数 q_k の大域的最適解 $\boldsymbol{d}_k^\star$ は $\nabla q_k(\boldsymbol{d}_k^\star) = \boldsymbol{0}$，すなわち次式を満たすことが分かる．（$\nabla^2 f(\boldsymbol{x}_k) \in \mathbb{S}^d_{++}$ より逆行列が存在）

$$\begin{aligned} \nabla q_k(\boldsymbol{d}_k^\star) &= \nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d}_k^\star = \boldsymbol{0} \\ \iff \boldsymbol{d}_k^\star &= -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k) \end{aligned} \tag{5.20}$$

反復回数 k において $\nabla f(\boldsymbol{x}_k) \neq \boldsymbol{0}$ とすると，仮定 $\nabla^2 f(\boldsymbol{x}_k) \in \mathbb{S}^d_{++}$ の下では，

$$\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k^\star \rangle_2 = \langle \nabla f(\boldsymbol{x}_k), -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k) \rangle_2 < 0 \tag{5.21}$$

となるので，探索方向 $\boldsymbol{d}_k := -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$ は降下方向である．

[1] $q_k(\boldsymbol{d}) := f(\boldsymbol{x}_k) + \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d} \rangle_2 + \frac{1}{2} \langle \boldsymbol{d}, \nabla^2 f(\boldsymbol{x}_k) \boldsymbol{d} \rangle \quad (5.19)$

アルゴリズムとそのイメージ

Require: $\boldsymbol{x}_0 \in \mathbb{R}^d$（初期点の設定）

Ensure: $\boldsymbol{x}_K$（停止条件を満たすベクトル）

1: $k \leftarrow 0$

2: **repeat**

3: $\quad \boldsymbol{d}_k := -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$

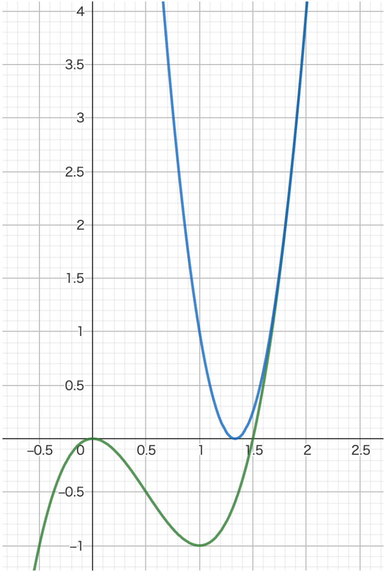
4: $\quad \boldsymbol{x}_{k+1} := \boldsymbol{x}_k + \boldsymbol{d}_k$

5: $\quad k \leftarrow k + 1$

6: **until** 停止条件を満たす

右図は $f(x) = 2x^3 - 3x^2$ に対して，

$x_0 = 2$ としたときの Taylor 展開



Theorem (定理 5.3.1: Newton 法の局所的２次収束性)

$f \in C^2(\mathbb{R}^d)$ として，$\nabla^2 f$ は Lipscitz 連続で，$\nabla^2 f$ の Lipscitz 定数を $L(>0)$ とする．停留点 $\boldsymbol{x}^\star$ に対して，$\nabla^2 f(\boldsymbol{x}^\star) \in \mathbb{S}^d_{++}$ として，ε を (5.22) を満たす正定数とする．

$$\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 L\varepsilon \leq \frac{1}{2} \tag{5.22}$$

このとき，$\boldsymbol{x}_0 \in N_2(\boldsymbol{x}^\star;\varepsilon)$ ならば，Newton 法で生成される点列 $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ は，

$$\begin{cases} \lim_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2 = 0 \\ \lim_{k\to\infty} \|\nabla f(\boldsymbol{x}_k)\|_2 = 0 \\ \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 \leq L\left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2 \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \qquad \longrightarrow \text{２次収束} \\ \|\nabla f(\boldsymbol{x}_{k+1})\|_2 \leq 2L\left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2^2 \|\nabla f(\boldsymbol{x}_k)\|_2^2 \end{cases}$$

を満たす．

まず，$\forall k \in \mathbb{N}\ (\boldsymbol{x}_k \in N_2(\boldsymbol{x}^\star; \varepsilon))$ を数学的帰納法で示す．

仮定より，$\boldsymbol{x}_0 \in N_2(\boldsymbol{x}^\star; \varepsilon)$ である．

ある番号 $k \in \mathbb{N}$ について，$\boldsymbol{x}_k \in N_2(\boldsymbol{x}^\star; \varepsilon)$ とする．このとき，次の不等式が成立する．

$$
\begin{aligned}
&\left\| \left(\nabla^2 f(\boldsymbol{x}^\star)\right)^{-1} \left(\nabla^2 f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}^\star)\right) \right\|_2 \\
&\le \left\| \left(\nabla^2 f(\boldsymbol{x}^\star)\right)^{-1} \right\|_2 \left\| \nabla^2 f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}^\star) \right\|_2 \quad (\because \|\boldsymbol{A}\boldsymbol{B}\|_2 \le \|\boldsymbol{A}\|_2 \|\boldsymbol{B}\|_2 \quad (2.12)) \\
&\le \left\| \left(\nabla^2 f(\boldsymbol{x}^\star)\right)^{-1} \right\|_2 L \left\| \boldsymbol{x}_k - \boldsymbol{x}^\star \right\|_2 \quad (\because L \text{ の定義}) \\
&< \left\| \left(\nabla^2 f(\boldsymbol{x}^\star)\right)^{-1} \right\|_2 L\varepsilon \quad (\because \boldsymbol{x}_k \in N_2(\boldsymbol{x}^\star; \varepsilon)) \\
&\le \frac{1}{2} \quad \left(\because \|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 L\varepsilon \le \frac{1}{2} \quad (5.22)\right)
\end{aligned}
\tag{5.24}
$$

（前スライドの最後の不等式：$\| \left(\nabla^2 f(\boldsymbol{x}^\star)\right)^{-1} \left(\nabla^2 f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}^\star)\right) \|_2 < \frac{1}{2} \quad (5.24)$）

よって，Banach の摂動定理[2] から，$\nabla^2 f(\boldsymbol{x}_k)$ は正則となり，
$\boldsymbol{x}_{k+1} := \boldsymbol{x}_k - \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1} \nabla f(\boldsymbol{x}_k)$ を定義できる．また，次式が成立する．

$$\begin{aligned}\|(\nabla^2 f(\boldsymbol{x}_k))^{-1}\|_2 &\leq \frac{\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2}{1 - \|(\nabla^2 f(\boldsymbol{x}^\star))^{-1} \left(\nabla^2 f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}^\star)\right) \|_2} \\ &< 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \quad (\because (5.24))\end{aligned} \tag{5.25}$$

[2] $\boldsymbol{A} \in \mathbb{R}^{d\times d}$ が正則で，$\boldsymbol{B} \in \mathbb{R}^{d\times d}$ が $\|\boldsymbol{A}^{-1}(\boldsymbol{B} - \boldsymbol{A})\|_2 < 1$ を満たすならば，$\boldsymbol{B}$ は正則となり，
$\|\boldsymbol{B}^{-1}\|_2 \leq \dfrac{\|\boldsymbol{A}^{-1}\|_2}{1 - \|\boldsymbol{A}^{-1}(\boldsymbol{B} - \boldsymbol{A})\|_2}$ が成り立つ．

（前スライドの最後の不等式：$\|(\nabla^2 f(\boldsymbol{x}_k))^{-1}\|_2 < 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \quad (5.25)$）

さらに，(5.25) から次式が成立する．

$$
\begin{aligned}
&\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 \\
&= \left\| \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1} \nabla f(\boldsymbol{x}_k) - (\boldsymbol{x}_k - \boldsymbol{x}^\star) \right\|_2 \quad (\because \boldsymbol{x}_{k+1}\text{の定義}) \\
&= \left\| \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1} \left\{\nabla f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\} \right\|_2 \\
&\le \left\| \left(\nabla^2 f(\boldsymbol{x}_k)\right)^{-1} \right\|_2 \left\|\nabla f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2 \quad (\because \|\boldsymbol{A}\boldsymbol{v}\|_2 \le \|\boldsymbol{A}\|_2\|\boldsymbol{v}\|_2 \quad (2.10)) \\
&< 2\left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2 \left\|\nabla f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2 \quad (\because (5.25)) \\
&= 2\left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2 N_k(\boldsymbol{x}^\star) \qquad (5.25.1)
\end{aligned}
$$

ここで，今後の議論のため，

$$
N_k(\boldsymbol{x}^\star) := \left\|\nabla f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2
$$

とした．

ここで，$N_k(\boldsymbol{x}^\star)$ について，次式が成立する．

$$
\begin{aligned}
N_k(\boldsymbol{x}^\star) &:= \left\|\nabla f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2 \\
&= \left\|\nabla f(\boldsymbol{x}_k) - \nabla f(\boldsymbol{x}^\star) - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2 \quad (\because \nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}) \\
&= \left\|\int_0^1 \nabla^2 f(t\boldsymbol{x}_k + (1-t)\boldsymbol{x}^\star)(\boldsymbol{x}_k - \boldsymbol{x}^\star)dt - \nabla^2 f(\boldsymbol{x}_k)(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2 \\
&\quad \left(\because \nabla f(\boldsymbol{y}) = \nabla f(\boldsymbol{x}) + \int_0^1 \nabla^2 f(t\boldsymbol{y} + (1-t)\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})dt \quad (\text{命題 } 2.2.4)\right) \\
&= \left\|\int_0^1 \left\{\nabla^2 f(t\boldsymbol{x}_k + (1-t)\boldsymbol{x}^\star) - \nabla^2 f(\boldsymbol{x}_k)\right\}(\boldsymbol{x}_k - \boldsymbol{x}^\star)dt\right\|_2 \\
&\leq \int_0^1 \left\|\left\{\nabla^2 f(t\boldsymbol{x}_k + (1-t)\boldsymbol{x}^\star) - \nabla^2 f(\boldsymbol{x}_k)\right\}(\boldsymbol{x}_k - \boldsymbol{x}^\star)\right\|_2 dt \\
&\leq \int_0^1 \left\|\nabla^2 f(t\boldsymbol{x}_k + (1-t)\boldsymbol{x}^\star) - \nabla^2 f(\boldsymbol{x}_k)\right\|_2 \left\|\boldsymbol{x}_k - \boldsymbol{x}^\star\right\|_2 dt \quad (\because \|\boldsymbol{A}\boldsymbol{v}\|_2 \leq \|\boldsymbol{A}\|_2\|\boldsymbol{v}\|_2 \ (2.10)) \\
&\leq \int_0^1 L \left\|t\boldsymbol{x}_k + (1-t)\boldsymbol{x}^\star - \boldsymbol{x}_k\right\|_2 \left\|\boldsymbol{x}_k - \boldsymbol{x}^\star\right\|_2 dt \quad (\because L \text{ の定義}) \\
&= L\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \int_0^1 (1-t)dt = \frac{L}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2
\end{aligned}
\tag{5.25.2}
$$

前々スライド最後の式：$\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^\star\|_2 < 2\left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2 N_k(\boldsymbol{x}^\star) \quad (5.25.1)$

前スライド最後の式：$N_k(\boldsymbol{x}^\star) \leq \frac{L}{2}\|\boldsymbol{x}_k-\boldsymbol{x}^\star\|_2^2 \quad (5.25.2)$

(5.25.1), (5.25.2) を組み合わせることで，次式が得られる．

$$\begin{aligned}
\|\boldsymbol{x}_{k+1}-\boldsymbol{x}^\star\|_2 &< L\left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2 \|\boldsymbol{x}_k-\boldsymbol{x}^\star\|_2^2 \\
&\leq \frac{1}{2\varepsilon}\|\boldsymbol{x}_k-\boldsymbol{x}^\star\|_2^2 \quad \left(\because \|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 L\varepsilon \leq \frac{1}{2} \quad (5.22)\right) \\
&< \frac{1}{2}\|\boldsymbol{x}_k-\boldsymbol{x}^\star\|_2 \quad (\because \boldsymbol{x}_k \in N_2(\boldsymbol{x}^\star;\varepsilon)) \\
&< \frac{\varepsilon}{2} \quad (\because \boldsymbol{x}_k \in N_2(\boldsymbol{x}^\star;\varepsilon)) \\
&< \varepsilon
\end{aligned}$$

よって，数学的帰納法より．

$$\forall k \in \mathbb{N}\,(\boldsymbol{x}_k \in N_2(\boldsymbol{x}^\star;\varepsilon)) \tag{5.23}$$

が示せた．

また，証明の過程で次式が得られた．

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 L \left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2 \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \qquad \text{(結果 1)}$$

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 < \frac{1}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2$$

この２つ目の式から，次式が得られる．

$$\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 < \frac{1}{2^{k+1}}\|\boldsymbol{x}_0 - \boldsymbol{x}^\star\|_2 \to 0 \quad (\text{as } k \to \infty)$$

$$\implies \lim_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\| = 0 \qquad \text{(結果 2; 5.26)}$$

次に，$\|\nabla f(\boldsymbol{x}_k)\|_2 \to 0 \quad (\text{as } k \to \infty)$ を示す．任意の $k \in \mathbb{N}$ について，次式が成立する．

$$\begin{aligned}
\nabla f(\boldsymbol{x}_{k+1}) &= \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d}_k \quad (\because \boldsymbol{d}_k := -(\nabla^2 f(\boldsymbol{x}_k))^{-1}\nabla f(\boldsymbol{x}_k)) \\
&= \int_0^1 \nabla^2 f(t\boldsymbol{x}_{k+1} + (1-t)\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)dt - \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d}_k \\
&\quad \left(\because \nabla f(\boldsymbol{y}) = \nabla f(\boldsymbol{x}) + \int_0^1 \nabla^2 f(t\boldsymbol{y} + (1-t)\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})dt \quad (\text{命題 } 2.2.4)\right)
\end{aligned}$$

前スライド最後の式：

$\nabla f(\boldsymbol{x}_{k+1}) = \int_0^1 \nabla^2 f(t\boldsymbol{x}_{k+1} + (1-t)\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)dt - \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d}_k$

よって，次式が成立する．（先程と同様の計算）

$$\begin{aligned}
\|\nabla f(\boldsymbol{x}_{k+1})\|_2 &= \left\| \int_0^1 \nabla^2 f(t\boldsymbol{x}_{k+1} + (1-t)\boldsymbol{x}_k)\boldsymbol{d}_k dt - \nabla^2 f(\boldsymbol{x}_k)\boldsymbol{d}_k \right\|_2 \\
&= \left\| \int_0^1 \left\{ \nabla^2 f(t\boldsymbol{x}_{k+1} + (1-t)\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k) \right\} \boldsymbol{d}_k dt \right\|_2 \\
&\leq \int_0^1 \left\| \nabla^2 f(t\boldsymbol{x}_{k+1} + (1-t)\boldsymbol{x}_k) - \nabla^2 f(\boldsymbol{x}_k) \right\|_2 \|\boldsymbol{d}_k\|_2 \, dt \\
&\leq L\|\boldsymbol{d}_k\|_2^2 \int_0^1 t dt = \frac{L}{2}\|\boldsymbol{d}_k\|_2^2 \qquad (5.27)
\end{aligned}$$

ここで，ステップサイズ $\alpha = 1$，すなわち $\boldsymbol{d}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$ より，任意の $k \in \mathbb{N}$ に対して，次式が成立する．（三角不等式を用いる）

$$\|\boldsymbol{d}_k\|_2 = \|(\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star) + (\boldsymbol{x}^\star - \boldsymbol{x}_k)\|_2 \leq \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 + \|\boldsymbol{x}^\star - \boldsymbol{x}_k\|_2$$

前スライドの式: $\begin{cases} \|\nabla f(\boldsymbol{x}_{k+1})\|_2 \le \frac{L}{2}\|\boldsymbol{d}_k\|_2^2 \quad (5.27) \\ \|\boldsymbol{d}_k\|_2 \le \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 + \|\boldsymbol{x}^\star - \boldsymbol{x}_k\|_2 \end{cases}$

これと (5.26)[3] を用いると，次式が得られる．

$$\lim_{k\to\infty} \|\boldsymbol{d}_k\|_2 = 0 \implies \lim_{k\to\infty} \|\nabla f(\boldsymbol{x}_k)\|_2 = 0 \qquad \text{(結果 3)}$$

また，任意の $k \in \mathbb{N}$ に対して次式が成立することが分かる．

$$\begin{aligned} &\|\nabla f(\boldsymbol{x}_{k+1})\|_2 \\ &\le \frac{L}{2} \left\|(\nabla^2 f(\boldsymbol{x}_k))^{-1}\right\|_2^2 \|\nabla f(\boldsymbol{x}_k)\|_2^2 \quad (\because (5.27) \text{ に } \boldsymbol{d}_k \text{の定義を代入}) \\ &\le 2L \left\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\right\|_2^2 \|\nabla f(\boldsymbol{x}_k)\|_2^2 \quad (\because \|(\nabla^2 f(\boldsymbol{x}_k))^{-1}\|_2 \le 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \quad (5.25)) \end{aligned}$$

(結果 4)

□

[3] $\lim_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\| = 0$

局所的収束性

$$f(x) := 2x^3 - 3x^2$$

$$\nabla f(x) = 6x^2 - 6x$$

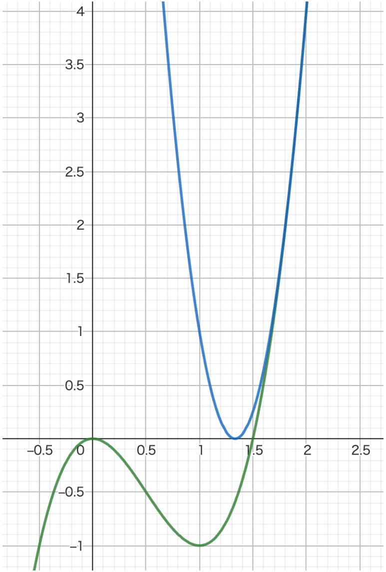
$$\nabla^2 f(x) = 12x - 6$$

$$L = 12$$

$$x^\star = 1$$

$$\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 L\varepsilon \leq \frac{1}{2}$$

$$\iff \varepsilon \leq \frac{1}{2}$$



5.4 準 Newton 法

導入

目的関数 f について，$f \in C^2(\mathbb{R}^d)$ とする．

本節では，大域的超 1 次収束性を有する準 Newton 法について見る．

準 Newton は以下の設定 (S1)-(S3) を用いる勾配法である．

Definition (準 Newton 法の設定)

(S1) 任意初期点	$\boldsymbol{x}_0 \in \mathbb{R}^d$	
(S2) ステップサイズ	$\alpha_k > 0 \ (k \in \mathbb{N})$	
(S3) 探索方向	$\boldsymbol{d}_k := -\boldsymbol{B}_k^{-1} \nabla f(\boldsymbol{x}_k) \ (k \in \mathbb{N})$	（$\boldsymbol{B}_k$： 次ページ以降）

Definition (超 1 次収束)

$(\boldsymbol{x}_k)_{k \in \mathbb{N}}$ が $\boldsymbol{x}^\star$ に超 1 次収束するとは，次式が成り立つことをいう．

$$\lim_{k \to \infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|} = 0$$

探索方向の気持ち (1)

前節で見た Newton 法では，$\nabla^2 f(\boldsymbol{x}_k) \in \mathbb{S}^d_{++}$ という仮定の下で，探索方向として次のものを使用していた．

$$\boldsymbol{d}_k^{\mathsf{Newton}} := -(\nabla^2 f(\boldsymbol{x}_k))^{-1} \nabla f(\boldsymbol{x}_k)$$

しかし，一般の目的関数 f に対する Hesse 行列 $\nabla^2 f(\boldsymbol{x})$ の正定値性は保証されない．その一方で，Newton 法の高速収束性に見られるように，目的関数の二階微分の情報を取り入れることで高速収束性が期待される．準 Newton 法では $\boldsymbol{B}_k$ を用いて，探索方向を次式で定義する．

$$\boldsymbol{d}_k^{\mathsf{quasi\text{-}Newton}} := -\boldsymbol{B}_k^{-1} \nabla f(\boldsymbol{x}_k)$$

ここで，この $\boldsymbol{B}_k$ は任意の $k \in \mathbb{N}$ に対して次式を満たすものである．

$$\boldsymbol{B}_k \in \mathbb{S}^d_{++}, \quad \boldsymbol{B}_k \approx \nabla^2 f(\boldsymbol{x}_k)$$

このとき，$\boldsymbol{d}_k^{\mathsf{quasi\text{-}Newton}}$ は Newton 法のときと同じ議論により，降下方向となる．

探索方向の気持ち (2)

$\nabla f(\boldsymbol{x}_k)$ の $\boldsymbol{x}_{k+1}$ におけるテイラー展開から次式が成り立つ．

$$\begin{aligned}&\nabla f(\boldsymbol{x}_{k+1}) \approx \nabla f(\boldsymbol{x}_k) + \nabla^2 f(\boldsymbol{x}_{k+1})(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)\\ \iff\ &\nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k) \approx \nabla^2 f(\boldsymbol{x}_{k+1})(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)\end{aligned}$$

このことから，$\boldsymbol{B}_{k+1}$ に対して，次の条件を課す．この条件をセカント条件という．

$$\boldsymbol{B}_{k+1}(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k) = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k) \tag{5.30}$$

セカント条件を満たすための $\boldsymbol{B}_k$ の更新式で有名なものの１つが BFGS 公式[4] である．

$$\boldsymbol{B}_{k+1} := \boldsymbol{B}_k - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \quad (\boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \boldsymbol{y}_k := \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)) \tag{5.31}$$

[4] the Broyden-Fletcher-Goldfarb-Shanno formula

BFGS 公式の性質

Theorem (BFGS 公式の正定値対称性)

$\boldsymbol{B}_k \in \mathbb{S}^d_{++}$ とし，$\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 \neq 0$ とする[5]．ただし，
$\boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \boldsymbol{y}_k := \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$ である．このとき，*BFGS* 公式

$$\boldsymbol{B}_{k+1} := \boldsymbol{B}_k - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \tag{5.31}$$

で生成される $\boldsymbol{B}_{k+1}$ は次の性質 *(P1)-(P3)* を満たす．

(P1) セカント条件 $\boldsymbol{B}_{k+1} \boldsymbol{s}_k = \boldsymbol{y}_k$

(P2) 対称性の保証 $\boldsymbol{B}_{k+1} \in \mathbb{S}^d$

(P3) 正定値の保証 *Wolfe* 条件に基づいた直線探索を行うならば，$\boldsymbol{B}_{k+1} \in \mathbb{S}^d_{++}$

[5] Wolfe 条件に基づいた直線探索を行うとき，$\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 > 0$ が成立するので，この仮定は不要となる．

(P1) の証明

(P1) $\boldsymbol{B}_{k+1}\boldsymbol{s}_k = \boldsymbol{y}_k$ を示す．

BFGS 公式に右から $\boldsymbol{s}_k$ をかけると得られる．

$$\begin{aligned}
\boldsymbol{B}_{k+1}\boldsymbol{s}_k &= \boldsymbol{B}_k\boldsymbol{s}_k - \frac{\boldsymbol{B}_k\boldsymbol{s}_k\boldsymbol{s}_k^\top\boldsymbol{B}_k\boldsymbol{s}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{s}_k\rangle_2} + \frac{\boldsymbol{y}_k\boldsymbol{y}_k^\top\boldsymbol{s}_k}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle_2} \\
&= \boldsymbol{B}_k\boldsymbol{s}_k - \frac{\boldsymbol{B}_k\boldsymbol{s}_k\langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{s}_k\rangle_2}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{s}_k\rangle_2} + \frac{\boldsymbol{y}_k\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle_2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k\rangle_2} \\
&= \boldsymbol{B}_k\boldsymbol{s}_k - \boldsymbol{B}_k\boldsymbol{s}_k + \boldsymbol{y}_k \\
&= \boldsymbol{y}_k
\end{aligned}$$

$\square$

(P2) の証明

(P2) $\boldsymbol{B}_{k+1} \in \mathbb{S}^d$ ($\overset{\text{def}}{\Longleftrightarrow}$ $\boldsymbol{B}_{k+1}^\top = \boldsymbol{B}_{k+1}$) を示す.

$$
\begin{aligned}
\boldsymbol{B}_{k+1}^\top &= \boldsymbol{B}_k^\top - \frac{\left(\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k\right)^\top}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\left(\boldsymbol{y}_k \boldsymbol{y}_k^\top\right)^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \\
&= \boldsymbol{B}_k - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \quad (\because \boldsymbol{B}_k \in \mathbb{S}^d) \\
&= \boldsymbol{B}_{k+1}
\end{aligned}
$$

$\square$

(P3) の証明

(i) $\to$ (ii) という手順で示す．

(i) Wolfe 条件に基づいた直線探索を行うならば，$\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 > 0$

(ii) $\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 > 0 \implies \boldsymbol{B}_{k+1} \in \mathbb{S}^d_{++}$

(i). Wolfe 条件の２つ目の式から，次式が成り立つ．$(c_2 \in (0,1))$

$$
\begin{aligned}
&\langle \nabla f(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k), \boldsymbol{d}_k \rangle_2 \geq c_2 \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \rangle_2 \\
\iff &\left\langle \nabla f(\boldsymbol{x}_{k+1}), \frac{\boldsymbol{s}_k}{\alpha_k} \right\rangle_2 \geq c_2 \left\langle \nabla f(\boldsymbol{x}_k), \frac{\boldsymbol{s}_k}{\alpha_k} \right\rangle_2 \quad (\because \boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k \boldsymbol{d}_k) \\
\iff &\langle \nabla f(\boldsymbol{x}_{k+1}), \boldsymbol{s}_k \rangle_2 \geq c_2 \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{s}_k \rangle_2 \quad (\because \alpha_k > 0)
\end{aligned}
$$

よって，$\boldsymbol{y}_k = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$ から，次式が成り立つ．

$$
\begin{aligned}
\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 &= \langle \nabla f(\boldsymbol{x}_{k+1}), \boldsymbol{s}_k \rangle_2 - \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{s}_k \rangle_2 \\
&\geq (c_2 - 1)\alpha_k \langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \rangle_2 \quad (\because \boldsymbol{s}_k = \alpha_k \boldsymbol{d}_k) \\
&> 0 \quad (\because c_2 \in (0,1), \alpha_k > 0, \boldsymbol{d}_k : \text{降下方向})
\end{aligned}
$$

$\square$

(ii). $\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 > 0 \implies \boldsymbol{B}_{k+1} \in \mathbb{S}^d_{++}$ を示す．

$\boldsymbol{x} \neq \boldsymbol{0}$ を満たす任意の $\boldsymbol{x} \in \mathbb{R}^d$ について，次式が成立する．

$$
\begin{aligned}
\langle \boldsymbol{x}, \boldsymbol{B}_{k+1}\boldsymbol{x} \rangle_2 &= \langle \boldsymbol{x}\boldsymbol{B}_k\boldsymbol{x} \rangle_2 - \frac{\langle \boldsymbol{x}, \boldsymbol{B}_k\boldsymbol{s}_k\boldsymbol{s}_k^\top \boldsymbol{B}_k\boldsymbol{x} \rangle_2}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{s}_k \rangle_2} + \frac{\langle \boldsymbol{x}, \boldsymbol{y}_k\boldsymbol{y}_k^\top \boldsymbol{x} \rangle_2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \\
&= \frac{\langle \boldsymbol{x}, \boldsymbol{B}_k\boldsymbol{x} \rangle_2 \langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{s}_k \rangle_2 - \langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{x} \rangle_2^2}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k\boldsymbol{s}_k \rangle_2} + \frac{\langle \boldsymbol{y}_k, \boldsymbol{x} \rangle_2^2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \\
&= \frac{\left\| \boldsymbol{B}_k^{1/2}\boldsymbol{x} \right\|_2^2 \left\| \boldsymbol{B}_k^{1/2}\boldsymbol{s}_k \right\|_2^2 - \left\langle \boldsymbol{B}_k^{1/2}\boldsymbol{s}_k, \boldsymbol{B}_k^{1/2}\boldsymbol{x} \right\rangle_2^2}{\left\| \boldsymbol{B}_k^{1/2}\boldsymbol{s}_k \right\|_2^2} + \frac{\langle \boldsymbol{y}_k, \boldsymbol{x} \rangle_2^2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \\
&\quad \left(\because \boldsymbol{B}_k \in \mathbb{S}^d_{++} \text{ より}, \boldsymbol{B}_k = \left(\boldsymbol{B}_k^{1/2} \right)^2 \text{ を満たす } \boldsymbol{B}_k^{1/2} \in \mathbb{S}^d_{++} \text{ が存在} \right)
\end{aligned}
$$

前スライドの情報：$\boldsymbol{x} \neq \mathbf{0}$ を満たす任意の $\boldsymbol{x} \in \mathbb{R}^d$ について，次式が成立．

$$\langle \boldsymbol{x}, \boldsymbol{B}_{k+1}\boldsymbol{x} \rangle_2 = \frac{\left\| \boldsymbol{B}_k^{1/2}\boldsymbol{x} \right\|_2^2 \left\| \boldsymbol{B}_k^{1/2}\boldsymbol{s}_k \right\|_2^2 - \left\langle \boldsymbol{B}_k^{1/2}\boldsymbol{s}_k, \boldsymbol{B}_k^{1/2}\boldsymbol{x} \right\rangle_2^2}{\left\| \boldsymbol{B}_k^{1/2}\boldsymbol{s}_k \right\|_2^2} + \frac{\langle \boldsymbol{y}_k, \boldsymbol{x} \rangle_2^2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}$$

ここで，Cauchy-Schwartz の不等式から，右辺第一項は 0 以上となる．

上記の式について，右辺第一項が正のとき，仮定より $\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2 > 0$ であるので，式全体が正となる．

一方，右辺第一項が 0 であるとき，等号成立の条件から，$\boldsymbol{x} = c\boldsymbol{s}_k$ を満たす $c \in \mathbb{R}$ が存在する．いま，$\boldsymbol{s}_k \coloneqq \boldsymbol{x}_{k+1} - \boldsymbol{x}_k \neq \mathbf{0}$ かつ $\boldsymbol{x} \neq \mathbf{0}$ なので，そのような $c \neq 0$ である．$\boldsymbol{x} = c\boldsymbol{s}_k$ を右辺第二項に代入すると，

$$\frac{\langle \boldsymbol{y}_k, \boldsymbol{x} \rangle_2^2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} = c^2 \langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle > 0$$

となる．よって，$\boldsymbol{x} \neq \mathbf{0}$ を満たす任意の $\boldsymbol{x} \in \mathbb{R}^d$ について，$\langle \boldsymbol{x}, \boldsymbol{B}_{k+1}\boldsymbol{x} \rangle_2 > 0$ となるので，$\boldsymbol{B}_{k+1} \in \mathbb{S}_{++}^d$ である． $\square$

アルゴリズム

先程の定理から，以降では Wolfe 条件を用いた直線探索を利用する準 Newton 法について見る．

Require: $c_1, c_2 : 0 < c_1 < c_2 < 1, \boldsymbol{x}_0 \in \mathbb{R}^d, \boldsymbol{B}_0 \in \mathbb{S}^d_{++}$

Ensure: $\boldsymbol{x}_K$

$k \leftarrow 0$

repeat

    $\boldsymbol{d}_k = -\boldsymbol{B}_k^{-1} \nabla f(\boldsymbol{x}_k)$

    $\alpha_k > 0$: Wolfe 条件を満たす

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

    $\boldsymbol{B}_{k+1}$を BGFS 公式から定義

    $k \leftarrow k + 1$

until 停止条件を満たす

Theorem (定理 5.4.1(1): 準 Newton 法の大域的収束性)

$f \in C^2(\mathbb{R}^d)$ とする．

このとき，ある $m > 0$ と $M > 0$ が存在して，任意の $\boldsymbol{x} \in \mathbb{R}^d$ に対して，

$$m\boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{x}) \preceq M\boldsymbol{I} \tag{5.32}$$

ならば，準 Newton 法は f に関する制約なし凸最適化問題の一意解 $\boldsymbol{x}^\star$ に大域的収束する．

命題 2.3.4(3)[6] から，f が強凸であるときに，準 Newton 法の大域的収束性が保証されることを言っている．

[6] $f \in C^2(\mathbb{R}^d)$ が c-強凸であるための必要十分条件は $\forall \boldsymbol{x} \in \mathbb{R}^d (\nabla^2 f(\boldsymbol{x}) \succeq c\boldsymbol{I})$

まず，ある番号 $k_0 \in \mathbb{N}$ に対して，$\boldsymbol{d}_{k_0} = \mathbf{0}$ とする．このとき，準 Newton 法の手続きから，$\nabla f(\boldsymbol{x}_{k_0}) = -\boldsymbol{B}_{k_0}\mathbf{0} = \mathbf{0}$ となる．ところで，f が凸関数であるとき，$\boldsymbol{x}^\star$ が大域的最適解であることの必要十分条件は $\nabla f(\boldsymbol{x}^\star) = \mathbf{0}$ であるので，$\boldsymbol{x}_{k_0}$ は大域的最適解である．

このような理由から，以降，任意の $k \in \mathbb{N}$ に対して，$\boldsymbol{d}_k \neq \mathbf{0}$ とする．

定理 5.4.1(1) の証明は長いので，$\spadesuit 1 \sim 6$ に分割する．

Theorem (♠1)

$$m_k := \frac{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2}$$

$$M_k := \frac{\|\boldsymbol{y}_k\|_2^2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}$$

このとき，任意の $k \in \mathbb{N}$ に対して，次の不等式が成立する．

$$m \le m_k, \quad M_k \le M$$

$\boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \boldsymbol{y}_k := \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$

任意の $k \in \mathbb{N}$ について，次式が成立する．

$$\begin{aligned}
\boldsymbol{y}_k &\coloneqq \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k) \\
&= \int_0^1 \nabla^2 f(t\boldsymbol{x}_{k+1} + (1-t)\boldsymbol{x}_k)(\boldsymbol{x}_{k+1} - \boldsymbol{x}_k)dt \\
&\quad \left(\because \nabla f(\boldsymbol{y}) = \nabla f(\boldsymbol{x}) + \int_0^1 \nabla^2 f(t\boldsymbol{y} + (1-t)\boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})dt \quad (\text{命題 } 2.2.4)\right) \\
&= \left\{\int_0^1 \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k)dt\right\}\alpha_k \boldsymbol{d}_k \quad (\because \boldsymbol{s}_k \coloneqq \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k \boldsymbol{d}_k) \\
&\coloneqq \boldsymbol{G}_k \alpha_k \boldsymbol{d}_k \quad \left(\because \boldsymbol{G}_k \coloneqq \int_0^1 \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k)dt\right) \\
&= \boldsymbol{G}_k s_k \quad (\because \boldsymbol{s}_k \coloneqq \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k \boldsymbol{d}_k) \qquad (5.34)
\end{aligned}$$

また，仮定（式 (5.32)[7]）より，

$$\forall \boldsymbol{x}, \boldsymbol{z} \in \mathbb{R}^d \ (m\|\boldsymbol{z}\|_2^2 \le \langle \boldsymbol{z}, \nabla^2 f(\boldsymbol{x})\boldsymbol{z}\rangle \le M\|\boldsymbol{z}\|_2^2)$$

が成り立つので，任意の $\boldsymbol{z} \in \mathbb{R}^d$ に対して，次の不等式が成立する．

$$\begin{aligned}
& m\|\boldsymbol{z}\|_2^2 \le \langle \boldsymbol{z}, \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k)\boldsymbol{z}\rangle_2 \le M\|\boldsymbol{z}\|_2^2 \\
\iff & m\|\boldsymbol{z}\|_2^2 \le \int_0^1 \langle \boldsymbol{z}, \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k)\boldsymbol{z}\rangle_2 dt \le M\|\boldsymbol{z}\|_2^2 \quad (\because \text{辺々 [0,1] で積分}) \\
\iff & m\|\boldsymbol{z}\|_2^2 \le \langle \boldsymbol{z}, \boldsymbol{G}_k \boldsymbol{z}\rangle_2 \le M\|\boldsymbol{z}\|_2^2 \quad (\because \boldsymbol{G}_k := \int_0^1 \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k) dt)
\end{aligned}$$

よって，

$$\boldsymbol{O} \preceq m\boldsymbol{I} \preceq \boldsymbol{G}_k \preceq M\boldsymbol{I}$$

[7] $\forall \boldsymbol{x} \in \mathbb{R}^d \ (m\boldsymbol{I} \preceq \nabla^2 f(\boldsymbol{x}) \preceq M\boldsymbol{I}) \quad (5.32)$

前々スライドの式：$\boldsymbol{y}_k = \boldsymbol{G}_k s_k \quad (\because \boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k \boldsymbol{d}_k) \quad (5.34)$
前スライドの式：$\boldsymbol{O} \preceq m\boldsymbol{I} \preceq \boldsymbol{G}_k \preceq M\boldsymbol{I}$

よって，任意の $k \in \mathbb{N}$ に対して，

$$m_k := \frac{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2} = \frac{\langle \boldsymbol{s}_k, \boldsymbol{G}_k \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2} \geq \frac{\langle \boldsymbol{s}_k, m\boldsymbol{I}\boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2} = m$$

$$M_k := \frac{\|\boldsymbol{y}_k\|_2^2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} = \frac{\langle \boldsymbol{G}_k \boldsymbol{s}_k, \boldsymbol{G}_k \boldsymbol{s}_k \rangle_2}{\langle \boldsymbol{s}_k, \boldsymbol{G}_k \boldsymbol{s}_k \rangle_2} = \frac{\left\langle \boldsymbol{G}_k^{\frac{1}{2}} \boldsymbol{s}_k, \boldsymbol{G}_k \boldsymbol{G}_k^{\frac{1}{2}} \boldsymbol{s}_k \right\rangle_2}{\left\langle \boldsymbol{G}_k^{\frac{1}{2}} \boldsymbol{s}_k, \boldsymbol{G}_k^{\frac{1}{2}} \boldsymbol{s}_k \right\rangle_2} = \frac{\langle \boldsymbol{z}_k, \boldsymbol{G}_k \boldsymbol{z}_k \rangle_2}{\|\boldsymbol{z}_k\|_2^2} \leq M$$

ここで，$\boldsymbol{z}_k := \boldsymbol{G}_k^{\frac{1}{2}} \boldsymbol{s}_k$ とした． $\square$

Theorem (♠2)

$$p_k := \frac{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2}$$

さらに，θ_k をベクトル $\boldsymbol{s}_k$ とベクトル $\boldsymbol{B}_k \boldsymbol{s}_k$ のなす角とする．

このとき，任意の $k \in \mathbb{N}$ に対して，次の等式が成立する．

$$\mathrm{Tr}(\boldsymbol{B}_{k+1}) = \mathrm{Tr}(\boldsymbol{B}_k) - \frac{p_k}{\cos^2 \theta_k} + M_k \tag{5.35}$$

$$\mathrm{Det}(\boldsymbol{B}_{k+1}) = \mathrm{Det}(\boldsymbol{B}_k) \frac{m_k}{p_k} \tag{5.36}$$

$$
\begin{aligned}
\operatorname{Tr}\left(\boldsymbol{B}_{k+1}\right) &= \operatorname{Tr}\left(\boldsymbol{B}_k - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}\right) \\
&= \operatorname{Tr}\left(\boldsymbol{B}_k\right) - \operatorname{Tr}\left(\frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\left\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \right\rangle_2}\right) + \operatorname{Tr}\left(\frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\left\langle \boldsymbol{y}_k, \boldsymbol{s}_k \right\rangle_2}\right) \\
&= \operatorname{Tr}\left(\boldsymbol{B}_k\right) - \frac{\left\| \boldsymbol{B}_k \boldsymbol{s}_k \right\|_2^2}{\left\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \right\rangle_2} + \frac{\left\| \boldsymbol{y}_k \right\|_2^2}{\left\langle \boldsymbol{y}_k, \boldsymbol{s}_k \right\rangle_2} \quad (\because \operatorname{Tr}(\boldsymbol{x}\boldsymbol{x}^\top) = \|\boldsymbol{x}\|_2^2) \\
&= \operatorname{Tr}\left(\boldsymbol{B}_k\right) - \frac{\left\| \boldsymbol{B}_k \boldsymbol{s}_k \right\|_2^2}{\left\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \right\rangle_2} + M_k \quad (\because M_k := \frac{\left\| \boldsymbol{y}_k \right\|_2^2}{\left\langle \boldsymbol{y}_k, \boldsymbol{s}_k \right\rangle_2}) \\
&= \operatorname{Tr}\left(\boldsymbol{B}_k\right) - \frac{p_k}{\cos^2 \theta_k} + M_k
\end{aligned}
$$

$$
(\because \frac{\left\| \boldsymbol{B}_k \boldsymbol{s}_k \right\|_2^2}{\left\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \right\rangle_2} = \frac{\left\| \boldsymbol{B}_k \boldsymbol{s}_k \right\|_2^2 \left\| \boldsymbol{s}_k \right\|_2^2}{\left\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \right\rangle_2^2} \frac{\left\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \right\rangle_2}{\left\| \boldsymbol{s}_k \right\|_2^2} = \frac{p_k}{\cos^2 \theta_k})
$$

$p_k := \frac{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2}$, θ_k はベクトル $\boldsymbol{s}_k$ とベクトル $\boldsymbol{B}_k \boldsymbol{s}_k$ のなす角

$$
\begin{aligned}
&\mathrm{Det}(\boldsymbol{B}_{k+1}) \\
&= \mathrm{Det}\left(\boldsymbol{B}_k - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}\right) \\
&= \mathrm{Det}\left(\boldsymbol{I} - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top \boldsymbol{B}_k^{-1}}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}\right) \mathrm{Det}(\boldsymbol{B}_k) \quad (\because \mathrm{Det}(\boldsymbol{A}\boldsymbol{B}) = \mathrm{Det}(\boldsymbol{A})\,\mathrm{Det}(\boldsymbol{B})) \\
&= \mathrm{Det}\left(I \underbrace{- \boldsymbol{B}_k \boldsymbol{s}_k}_{\boldsymbol{x}} \underbrace{\left(\frac{\boldsymbol{s}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}\right)^\top}_{\boldsymbol{y}^\top} + \underbrace{\frac{\boldsymbol{y}_k}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}}_{\boldsymbol{u}} \underbrace{\left(\boldsymbol{B}_k^{-1} \boldsymbol{y}_k\right)^\top}_{\boldsymbol{v}^\top} \right) \times \mathrm{Det}\left(\boldsymbol{B}_k\right) \\
&= \mathrm{Det}(\boldsymbol{B}_k) \left\{ \left(1 - \frac{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}\right) \left(1 + \frac{\langle \boldsymbol{B}_k^{-1} \boldsymbol{y}_k, \boldsymbol{y}_k \rangle_2}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}\right) \right. \\
&\qquad \left. + \langle \boldsymbol{B}_k \boldsymbol{s}_k, \boldsymbol{B}_k^{-1} \boldsymbol{y}_k \rangle_2 \left\langle \frac{\boldsymbol{s}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}, \frac{\boldsymbol{y}_k}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \right\rangle_2 \right\} \quad (\because \text{下の式}) \\
&= \mathrm{Det}(\boldsymbol{B}_k) \frac{\langle \boldsymbol{s}_k, \boldsymbol{y}_k \rangle_2}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} = \mathrm{Det}(\boldsymbol{B}_k) \frac{\|\boldsymbol{s}_k\|_2^2}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} \frac{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2^2} = \mathrm{Det}(\boldsymbol{B}_k) \frac{m_k}{p_k}
\end{aligned}
$$

$\mathrm{Det}(\boldsymbol{I} + \boldsymbol{x}\boldsymbol{y}^\top + \boldsymbol{u}\boldsymbol{v}^\top) = (1 + \langle \boldsymbol{x}, \boldsymbol{y} \rangle_2)(1 + \langle \boldsymbol{u}, \boldsymbol{v} \rangle_2) - \langle \boldsymbol{x}, \boldsymbol{v} \rangle_2 \langle \boldsymbol{y}, \boldsymbol{u} \rangle_2$（証明は p.319–）

Theorem (♠3)

$$C(\boldsymbol{B}_{k+1}) := \mathrm{Tr}(\boldsymbol{B}_{k+1}) - \log \mathrm{Det}(\boldsymbol{B}_{k+1})$$
$$c := M - \log m - 1$$

このとき，任意の $k \in \mathbb{N}$ に対して，次の不等式が成立する．

$$0 < C(\boldsymbol{B}_0) + c(k+1) + \sum_{j=0}^{k} \log \cos^2 \theta_j$$

♠2 の結果[8] より，任意の $k \in \mathbb{N}$ に対して，次式が成り立つ．

$$
\begin{aligned}
C\left(\boldsymbol{B}_{k+1}\right) &:= \operatorname{Tr}\left(\boldsymbol{B}_{k+1}\right) - \log \operatorname{Det}\left(\boldsymbol{B}_{k+1}\right) \\
&= \operatorname{Tr}\left(\boldsymbol{B}_{k}\right) - \frac{p_k}{\cos^2\theta_k} + M_k - \log \operatorname{Det}\left(\boldsymbol{B}_{k}\right) - \log m_k + \log p_k \quad (\because \text{♠2 の結果}) \\
&= C\left(\boldsymbol{B}_{k}\right) + M_k - \log m_k - \frac{p_k}{\cos^2\theta_k} + \log p_k \\
&= C\left(\boldsymbol{B}_{k}\right) + \left(M_k - \log m_k - 1\right) + \left(1 - \frac{p_k}{\cos^2\theta_k} + \log p_k\right) \\
&= C\left(\boldsymbol{B}_{k}\right) + \left(M_k - \log m_k - 1\right) \\
&\quad + \left(1 - \frac{p_k}{\cos^2\theta_k} + \log \frac{p_k}{\cos^2\theta_k}\right) + \log \cos^2\theta_k
\end{aligned}
$$

[8] $\operatorname{Tr}(\boldsymbol{B}_{k+1}) = \operatorname{Tr}(\boldsymbol{B}_k) - \frac{p_k}{\cos^2\theta_k} + M_k, \quad \operatorname{Det}(\boldsymbol{B}_{k+1}) = \operatorname{Det}(\boldsymbol{B}_k)\frac{m_k}{p_k}$

前スライドの式：

$$C\left(\boldsymbol{B}_{k+1}\right) = C\left(\boldsymbol{B}_k\right) + \left(M_k - \log m_k - 1\right) + \left(1 - \tfrac{p_k}{\cos^2\theta_k} + \log\tfrac{p_k}{\cos^2\theta_k}\right) + \log\cos^2\theta_k$$

ここで，$\boldsymbol{B}_k$ の i 番目の固有値を λ_i とすると，次式が成り立つ．

$$\begin{aligned} C(\boldsymbol{B}_k) &= \operatorname{Tr}\left(\boldsymbol{B}_{k+1}\right) - \log\operatorname{Det}\left(\boldsymbol{B}_{k+1}\right) \\ &= \sum_{i=1}^{d}\lambda_i + \log\prod_{i=1}^{d}\lambda_i = \sum_{i=1}^{d}(\lambda_i - \log\lambda_i) > 0 \quad (\because x > \log x\ (x > 0)) \end{aligned}$$

また，$p(x) := 1 - x + \log x$ は $x > 0$ のとき，$p(x) \le 0$ であり，$\frac{p_k}{\cos^2\theta_k} > 0$ であるので，次式が成り立つ．

$$1 - \frac{p_k}{\cos^2\theta_k} + \log\frac{p_k}{\cos^2\theta_k} \le 0$$

また，♠1 の結果から，$m \le m_k, M_k \le M$ であるので，次の不等式が成り立つ．

$$\begin{aligned} 0 < C\left(\boldsymbol{B}_{k+1}\right) &\le C\left(\boldsymbol{B}_k\right) + \left(M - \log m - 1\right) + \log\cos^2\theta_k \\ &\le C\left(\boldsymbol{B}_0\right) + c(k+1) + \sum_{j=0}^{k}\log\cos^2\theta_j \quad (\because c := M - \log m - 1) \end{aligned}$$

Theorem ($\spadesuit$4)

次の命題が成立する．

$$\exists \varepsilon > 0 \, \forall j_0 \in \mathbb{N} \, \exists j \in \mathbb{N} \, (j \geq j_0 \wedge \cos^2 \theta_j > \varepsilon)$$

すなわち，$(\theta_k)_{k \in \mathbb{N}}$ の部分列 $(\theta_{k_j})_{j \in \mathbb{N}}$ が存在して，任意の $j \in \mathbb{N}$ に対して，$\cos^2 \theta_{k_j} > \varepsilon$ が成立する．

また，そのような添字 k_j について，任意の $j \in \mathbb{N}$ に対して，次の不等式が成立する．

$$\varepsilon \|\nabla f(\boldsymbol{x}_{k_j})\|_2^2 < \left(\frac{\langle \nabla f(\boldsymbol{x}_{\boldsymbol{k_j}}), \boldsymbol{d}_{k_j} \rangle_2}{\|\boldsymbol{d}_{k_j}\|_2} \right)^2 \tag{5.41}$$

ここで， $\lim_{k\to+\infty}\cos^2\theta_k = 0$, つまり，

$$\forall\varepsilon > 0\,\exists j_0 \in \mathbb{N}\,\forall j \in \mathbb{N}\,\left(j \geq j_0 \Longrightarrow \cos^2\theta_j \leq \varepsilon\right) \tag{5.38}$$

を仮定する． $\varepsilon = e^{-2c}$ とすると，任意の $j \geq j_0$ に対して $\log\cos^2\theta_j \leq -2c$ が成り立つので,

$$0 < C\left(\boldsymbol{B}_{k+1}\right) \leq C\left(\boldsymbol{B}_0\right) + c(k+1) + \sum_{j=0}^{j_0-1}\log\cos^2\theta_j - 2c\left(k - j_0 + 1\right)$$

が任意の $k \geq j_0$ に対して成立する． k を発散させると，上の不等式の右辺は $-\infty$ となり，0 より大きいことに矛盾する．よって，式 (5.38) の否定命題

$$\exists\varepsilon > 0\,\forall j_0 \in \mathbb{N}\,\exists j\ \in \mathbb{N}\left(j \geq j_0 \wedge \cos^2\theta_j > \varepsilon\right)$$

が成立することが分かる．すなわち，$(\theta_k)_{k\in\mathbb{N}}$ の部分列 $(\theta_{k_j})_{j\in\mathbb{N}}$ が存在して，任意の $j \in \mathbb{N}$ に対して，次式が成立する．

$$\cos^2\theta_{k_j} > \varepsilon \tag{5.39}$$

前スライドの最後の式：$\cos^2\theta_{k_j} > \varepsilon \quad (5.39)$

ここで，$\boldsymbol{s}_k = \alpha_k \boldsymbol{d}_k$ と $\boldsymbol{B}_k \boldsymbol{s}_k = \alpha_k \boldsymbol{B}_k \boldsymbol{d}_k = -\alpha_k \nabla f(\boldsymbol{x}_k)$ から，次式が成り立つ．

$$\cos^2\theta_k = \left(\frac{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2}{\|\boldsymbol{s}_k\|_2 \|\boldsymbol{B}_k \boldsymbol{s}_k\|_2}\right)^2 = \left(\frac{\langle \boldsymbol{d}_k, \nabla f(\boldsymbol{x}_k) \rangle_2}{\|\boldsymbol{d}_k\|_2}\right)^2 \frac{1}{\|\nabla f(\boldsymbol{x}_k)\|_2^2} \tag{5.40}$$

よって，式 (5.39),(5.40) から，任意の $j \in \mathbb{N}$ に対して，次式が成立する．

$$\varepsilon \left\|\nabla f\left(\boldsymbol{x}_{k_j}\right)\right\|_2^2 < \left(\frac{\left\langle \nabla f\left(\boldsymbol{x}_{k_j}\right), \boldsymbol{d}_{k_j} \right\rangle_2}{\left\|\boldsymbol{d}_{k_j}\right\|_2}\right)^2 \tag{5.41}$$

□

Theorem (♠5)

♠4 で存在が証明された $(\theta_k)_{k\in\mathbb{N}}$ の部分列 $(\theta_{k_j})_{j\in\mathbb{N}}$ に対して次式が成立する．

$$\lim_{j\to\infty} \boldsymbol{x}_{k_j} = \boldsymbol{x}^\star$$

f の m-強凸性より，命題 3.3.3(2)[9] から，f は下に有界であり，命題 2.2.5[10] と $\boldsymbol{O} \prec \nabla^2 f(\boldsymbol{x}) \preceq M\boldsymbol{I}$ より，∇f は M-Lipscitz 連続となる．よって，Zoutendijk の定理（定理 4.3.1）[11] より，次式が成立する．

$$\lim_{k\to\infty} \left(\frac{\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \rangle_2}{\|\boldsymbol{d}_k\|_2} \right)^2 = 0 \tag{5.42}$$

[9] $f \colon \mathbb{R}^d \to \mathbb{R}$ が強凸ならば，f はただ一つの大域的最適解を持つ

[10] $f \in C^2(\mathbb{R}^d)$ について，∇f がユークリッドノルムの意味で L-Lipschitz 連続である $\iff \forall \boldsymbol{x} \in \mathbb{R}^d \ (\|\nabla^2 f(\boldsymbol{x})\|_2 \leq L)$

[11] 下に有界な関数 $f \in C^1_L(\mathbb{R}^d)$ を最小化するための勾配法について，探索方向 $\boldsymbol{d}_k (\neq \boldsymbol{0})$ が降下方向であり，ステップサイズ α_k が Wolfe 条件を満たすとする．このとき，$\sum_{k=0}^{\infty} \left(\frac{\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k \rangle}{\|\boldsymbol{d}_k\|} \right)^2 < +\infty$

よって，式 (5.41), (5.42)[12] から，次式が成立する．

$$\lim_{j\to\infty} \|\nabla f(\boldsymbol{x}_{k_j})\|_2 = 0 \tag{5.43}$$

また，f の m-強凸性から，任意の $j \in \mathbb{N}$ に対して，次の不等式が成り立つ．

$$\begin{aligned} m\|\boldsymbol{x}_{k_j} - \boldsymbol{x}^\star\|_2^2 &\le \langle \boldsymbol{x}_{k_j} - \boldsymbol{x}^\star, \nabla f(\boldsymbol{x}_{k_j})\rangle_2 \quad (\because \text{命題 2.3.4(2)}^{13} \text{ と } \nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}) \\ &\le \|\boldsymbol{x}_{k_j} - \boldsymbol{x}^\star\|_2 \|\nabla f(\boldsymbol{x}_{k_j})\|_2 \quad (\because \text{Cauchy-Schwartz の不等式（命題 2.1.3）}) \\ \iff m\|\boldsymbol{x}_{k_j} - \boldsymbol{x}^\star\|_2 &\le \|\nabla f(\boldsymbol{x}_{k_j})\|_2 \end{aligned}$$

これと，式 (5.43) から，$\lim_{j\to\infty} \boldsymbol{x}_{k_j} = \boldsymbol{x}^\star$ が成り立つ． □

---

[12] $\varepsilon \left\|\nabla f\left(\boldsymbol{x}_{k_j}\right)\right\|_2^2 < \left(\frac{\left\langle \nabla f\left(\boldsymbol{x}_{k_j}, \boldsymbol{d}_{k_j}\right)\right\rangle_2}{\left\|\boldsymbol{d}_{k_j}\right\|_2}\right)^2 \quad (5.41), \lim_{k\to\infty}\left(\frac{\langle \nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k\rangle_2}{\|\boldsymbol{d}_k\|_2}\right)^2 = 0 \quad (5.42)$

[13] $f\colon \mathbb{R}^d \to \mathbb{R}$ が c-強凸 $\iff \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \forall \boldsymbol{g}_x \in \partial f(\boldsymbol{x}) \forall \boldsymbol{g}_y \in \partial f(\boldsymbol{y}) (\langle \boldsymbol{g}_x - \boldsymbol{g}_y, \boldsymbol{x} - \boldsymbol{y}\rangle \ge c\|\boldsymbol{x} - \boldsymbol{y}\|^2)$

Theorem (♠6)

$$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^\star$$

任意の $k \in \mathbb{N}$ に対して，次の不等式が成立する．

$$\begin{aligned} f(\boldsymbol{x}^\star) &\le f(\boldsymbol{x}_{k+1}) \quad (\because \boldsymbol{x}^\star \text{は大域的最適解}) \\ &\le f(\boldsymbol{x}_k) + c_1\alpha_k\langle\nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k\rangle_2 \quad (\because f(\boldsymbol{x}_k + \alpha\boldsymbol{d}_k) \le f(\boldsymbol{x}_k) + c_1\alpha_k\langle\nabla f(\boldsymbol{x}_k), \boldsymbol{d}_k\rangle\ (4.17)) \\ &\le f(\boldsymbol{x}_k) \quad (\because \boldsymbol{d}_k \text{は降下方向}, c_1 \in (0,1), \alpha_k > 0) \end{aligned}$$

よって，$(f(\boldsymbol{x}_k))_{k\in\mathbb{N}}$ は下に有界な単調減少数列であるので，$(f(\boldsymbol{x}_k))_{k\in\mathbb{N}}$ は収束する．
さらに，f の連続性と $\boldsymbol{x}_{k_j} \to \boldsymbol{x}^\star$ から，次式が成立する．

$$\lim_{k\to\infty} f(\boldsymbol{x}_k) = \lim_{j\to\infty} f(\boldsymbol{x}_{k_j}) = f\left(\lim_{j\to\infty} \boldsymbol{x}_{k_j}\right) = f(\boldsymbol{x}^\star) \tag{5.44}$$

前スライドの結果：$\lim_{k\to\infty} f(\boldsymbol{x}_k) = f(\boldsymbol{x}^\star) \quad (5.44)$

また，任意の $k \in \mathbb{N}$ に対して，次の不等式が成立する．

$$\begin{aligned}
f(\boldsymbol{x}_k) &\geq f(\boldsymbol{x}^\star) + \langle \nabla f(\boldsymbol{x}^\star), \boldsymbol{x}_k - \boldsymbol{x}^\star \rangle_2 + \frac{m}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \quad (\because \text{命題 2.3.4(1)}) \\
&= f(\boldsymbol{x}^\star) + \frac{m}{2}\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \quad (\because \nabla f(\boldsymbol{x}^\star) = \boldsymbol{0}) \\
\iff \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 &\leq \frac{2}{m}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^\star)) \\
\implies \varlimsup_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 &\leq \varlimsup_{k\to\infty} \frac{2}{m}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^\star)) \quad (\because \text{命題 2.1.8(1)}) \qquad (\spadesuit\ 6.1)
\end{aligned}$$

さらに，次式が成立する．

$$\varlimsup_{k\to\infty} \frac{2}{m}(f(\boldsymbol{x}_k) - f(\boldsymbol{x}^\star)) = 0 \leq \varliminf_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \quad (\because (5.44)) \qquad (\spadesuit\ 6.2)$$

よって，式 ($\spadesuit$ 6.1), ($\spadesuit$ 6.2) と $\varliminf_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \leq \varlimsup_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2$ から，次式が成り立つ．

$$\varlimsup_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 = \varliminf_{k\to\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 \iff \lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^\star$$

$\square$

命題 2.3.4(1): $f\colon \mathbb{R}^d \to \mathbb{R}$ が c-強凸 $\iff \forall \boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d \forall \boldsymbol{g}_{\in \partial f(\boldsymbol{x})} (f(\boldsymbol{y}) \geq f(\boldsymbol{z}) + \langle \boldsymbol{g}, \boldsymbol{y} - \boldsymbol{x} \rangle + \frac{c}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2)$

命題 2.1.8(1): $(a_k)_{k\in\mathbb{N}}$ と $(b_k)_{k\in\mathbb{N}}$ は有界な実数列とする．このとき，$\forall k\ (a_k \leq b_k) \implies \varlimsup_{k\to\infty} a_k \leq \varlimsup_{k\to\infty} b_k$

Theorem (定理 5.4.1(2): 準 Newton 法の超 1 次収束性)

$f \in C^2(\mathbb{R}^d)$ とする．

$\nabla^2 f$ を Lipschitz 連続として，その Lipschitz 定数を L とする．
さらに，準 Newton 法で生成される点列 $(\boldsymbol{x}_k)_{k\in\mathbb{N}}$ が $\nabla^2 f(\boldsymbol{x}^\star) \in \mathbb{S}^d_{++}$ を満たす f の停留点 $\boldsymbol{x}^\star$ に収束し，さらに，次の不等式を満たすとする．

$$\sum_{k=0}^{\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2 < \infty \tag{5.33}$$

このとき，十分大きい k に対して，$\alpha_k = 1$ ならば，次式が成立する．

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2}{\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2} = 0 \qquad \longrightarrow \text{超 1 次収束}$$

この証明も長いので，♣1 ∼ 6 に分割する．

$$\boldsymbol{G}_\star := \nabla^2 f(\boldsymbol{x}^\star)$$
$$\tilde{\boldsymbol{s}}_k := \boldsymbol{G}_\star^{\frac{1}{2}} \boldsymbol{s}_k$$
$$\tilde{\boldsymbol{y}}_k := \boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{y}_k$$
$$\tilde{\boldsymbol{B}}_k := \boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{B}_k \boldsymbol{G}_\star^{-\frac{1}{2}}$$
$$\cos\tilde{\theta}_k := \frac{\left\langle \tilde{\boldsymbol{s}}_k \tilde{\boldsymbol{B}}_k \tilde{\boldsymbol{s}}_k \right\rangle_2}{\|\tilde{\boldsymbol{s}}_k\|_2 \left\| \tilde{\boldsymbol{B}}_k \tilde{\boldsymbol{s}}_k \right\|_2}$$

$$\tilde{p}_k := \frac{\left\langle \tilde{\boldsymbol{s}}_k, \tilde{\boldsymbol{B}}_k \tilde{\boldsymbol{s}}_k \right\rangle_2}{\|\tilde{\boldsymbol{s}}_k\|_2^2}$$
$$\tilde{M}_k := \frac{\|\tilde{\boldsymbol{y}}_k\|_2^2}{\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2}$$
$$\tilde{m}_k := \frac{\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2}{\|\tilde{\boldsymbol{s}}_k\|_2^2}$$
$$C\left(\tilde{\boldsymbol{B}}_{k+1}\right) := \mathrm{Tr}\left(\tilde{\boldsymbol{B}}_{k+1}\right) - \log \mathrm{Det}\left(\tilde{\boldsymbol{B}}_{k+1}\right)$$

とする．

Theorem (♣1)

$$
\begin{aligned}
0 &< C\left(\tilde{\boldsymbol{B}}_{k+1}\right) \\
&= C\left(\tilde{\boldsymbol{B}}_{k}\right) + \left(\tilde{M}_k - \log \tilde{m}_k - 1\right) + \left(1 - \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k} + \log \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k}\right) + \log \cos^2 \tilde{\theta}_k
\end{aligned}
$$

$$
\begin{aligned}
\tilde{\boldsymbol{B}}_{k+1} &= \boldsymbol{G}_\star^{-\frac{1}{2}} B_{k+1} \boldsymbol{G}_\star^{-\frac{1}{2}} \\
&= \boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{B}_k \boldsymbol{G}_\star^{-\frac{1}{2}} - \frac{\boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k \boldsymbol{G}_\star^{-\frac{1}{2}}}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{y}_k \boldsymbol{y}_k^\top \boldsymbol{G}_\star^{-\frac{1}{2}}}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \\
&= \tilde{\boldsymbol{B}}_k - \frac{\boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{B}_k \boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{G}_\star^{\frac{1}{2}} \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{G}_\star^{\frac{1}{2}} \boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{B}_k \boldsymbol{G}_\star^{-\frac{1}{2}}}{\left\langle \boldsymbol{G}_\star^{-\frac{1}{2}} \tilde{\boldsymbol{s}}_k, \boldsymbol{B}_k \boldsymbol{G}_\star^{-\frac{1}{2}} \tilde{\boldsymbol{s}}_k \right\rangle_2} + \frac{\tilde{\boldsymbol{y}}_k \tilde{\boldsymbol{y}}_k^\top}{\left\langle \boldsymbol{G}_\star^{\frac{1}{2}} \tilde{\boldsymbol{y}}_k, \boldsymbol{G}_\star^{-\frac{1}{2}} \tilde{\boldsymbol{s}}_k \right\rangle_2} \\
&= \tilde{\boldsymbol{B}}_k - \frac{\tilde{\boldsymbol{B}}_k \tilde{\boldsymbol{s}}_k \tilde{\boldsymbol{s}}_k^\top \tilde{\boldsymbol{B}}_k}{\left\langle \tilde{\boldsymbol{s}}_k, \tilde{\boldsymbol{B}}_k \tilde{\boldsymbol{s}}_k \right\rangle_2} + \frac{\tilde{\boldsymbol{y}}_k \tilde{\boldsymbol{y}}_k^\top}{\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2}
\end{aligned}
$$

前スライドの式は BFGS 公式と同様の形であるので，♠3 と同じ議論から，次式が成り立つ．

$$
\begin{aligned}
0 < C\left(\tilde{\boldsymbol{B}}_{k+1}\right) \coloneqq \operatorname{Tr}\left(\tilde{\boldsymbol{B}}_{k+1}\right) - \log \operatorname{Det}\left(\tilde{\boldsymbol{B}}_{k+1}\right) \\
= & C\left(\tilde{\boldsymbol{B}}_{k}\right) + \left(\tilde{M}_k - \log \tilde{m}_k - 1\right) \\
& + \left(1 - \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k} + \log \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k}\right) + \log \cos^2 \tilde{\theta}_k
\end{aligned}
$$

□

$$
\boldsymbol{B}_{k+1} \coloneqq \boldsymbol{B}_k - \frac{\boldsymbol{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^\top \boldsymbol{B}_k}{\langle \boldsymbol{s}_k, \boldsymbol{B}_k \boldsymbol{s}_k \rangle_2} + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^\top}{\langle \boldsymbol{y}_k, \boldsymbol{s}_k \rangle_2} \quad \left(\boldsymbol{s}_k \coloneqq \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \boldsymbol{y}_k \coloneqq \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)\right) \tag{5.31}
$$

$$
\begin{aligned}
C\left(\boldsymbol{B}_{k+1}\right) \coloneqq & \operatorname{Tr}\left(\boldsymbol{B}_{k+1}\right) - \log \operatorname{Det}\left(\boldsymbol{B}_{k+1}\right) \\
= & \operatorname{Tr}\left(\boldsymbol{B}_k\right) - \frac{p_k}{\cos^2 \theta_k} + M_k - \log \operatorname{Det}\left(\boldsymbol{B}_k\right) - \log m_k + \log p_k \\
= & C\left(\boldsymbol{B}_k\right) + \left(M_k - \log m_k - 1\right) \\
& + \left(1 - \frac{p_k}{\cos^2 \theta_k} + \log \frac{p_k}{\cos^2 \theta_k}\right) + \log \cos^2 \theta_k
\end{aligned}
$$

Theorem (♣2)

$$\varepsilon_k := \max\left\{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2, \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2\right\}$$

とする．このとき，次式が成り立つ．

$$\lim_{k\to\infty} \varepsilon_k = 0$$

$$\sum_{k=0}^{\infty} \varepsilon_k < \infty$$

仮定より，$\lim_{k\to\infty} \boldsymbol{x}_k = \boldsymbol{x}^\star$ であるので，一つ目の等式は OK.

仮定（式 5.33）より，$\sum_{k=0}^{\infty} \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2 < \infty$ なので，二つ目の等式は OK.

($\varepsilon_k \leq \|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 + \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2$ と考える.)

Theorem (♣3)

$$\|\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k\|_2 \le 3L \left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\right\|_2^2 \varepsilon_k \|\tilde{\boldsymbol{s}}_k\|_2$$

$\boldsymbol{y}_k = \boldsymbol{G}_k \boldsymbol{s}_k \quad (5.34)$ と，$\tilde{\boldsymbol{s}}_k := \boldsymbol{G}_\star^{\frac{1}{2}} \boldsymbol{s}_k, \quad \tilde{\boldsymbol{y}}_k := \boldsymbol{G}_\star^{-\frac{1}{2}} \boldsymbol{y}_k$ より，次式が成り立つ．

$$\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k = \boldsymbol{G}_\star^{-\frac{1}{2}} (\boldsymbol{y}_k - \boldsymbol{G}_\star \boldsymbol{s}_k) = \boldsymbol{G}_\star^{-\frac{1}{2}} (\boldsymbol{G}_k - \boldsymbol{G}_\star) \boldsymbol{G}_\star^{-\frac{1}{2}} \tilde{\boldsymbol{s}}_k \tag{♣ 3.1}$$

ここで，右辺の $\boldsymbol{G}_k - \boldsymbol{G}_\star$ について，次の不等式が成立する．

$$\begin{aligned}
\|\boldsymbol{G}_k - \boldsymbol{G}_\star\|_2 &= \left\| \int_0^1 \left\{ \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k) - \nabla^2 f(\boldsymbol{x}^\star) \right\} dt \right\|_2 \\
&\le \int_0^1 \left\| \nabla^2 f(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k) - \nabla^2 f(\boldsymbol{x}^\star) \right\|_2 dt \\
&\le L \int_0^1 \|(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k) - \boldsymbol{x}^\star\|_2 \, dt \quad (\because \nabla^2 f \text{ は } L\text{-Lipscitz 連続}) \qquad (♣\ 3.2)
\end{aligned}$$

前スライドの式：$\begin{cases} \tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k = \boldsymbol{G}_\star^{-\frac{1}{2}} \left(\boldsymbol{y}_k - \boldsymbol{G}_\star \boldsymbol{s}_k\right) = \boldsymbol{G}_\star^{-\frac{1}{2}} \left(\boldsymbol{G}_k - \boldsymbol{G}_\star\right) \boldsymbol{G}_\star^{-\frac{1}{2}} \tilde{\boldsymbol{s}}_k \quad (\clubsuit 3.1) \\ \left\|\boldsymbol{G}_k - \boldsymbol{G}_\star\right\|_2 \leq L \int_0^1 \left\|(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k) - \boldsymbol{x}^\star\right\|_2 dt \quad (\clubsuit 3.2) \end{cases}$

ここでさらに，任意の $t \in [0,1]$ について，次の不等式が成り立つ．

$$
\begin{aligned}
&\left\|(\boldsymbol{x}_k + t\alpha_k \boldsymbol{d}_k) - \boldsymbol{x}^\star\right\|_2 \\
&\leq \left\|\boldsymbol{x}_k - \boldsymbol{x}^\star\right\|_2 + t\left\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\right\|_2 \quad (\because \text{三角不等式}, \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k \boldsymbol{d}_k) \\
&\leq \left\|\boldsymbol{x}_k - \boldsymbol{x}^\star\right\|_2 + t\left(\left\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\right\|_2 + \left\|\boldsymbol{x}^\star - \boldsymbol{x}_k\right\|_2\right) \quad (\because \text{三角不等式}) \\
&\leq \varepsilon_k + 2t\varepsilon_k \quad (\because \varepsilon_k := \max\left\{\left\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\right\|_2, \left\|\boldsymbol{x}_k - \boldsymbol{x}^\star\right\|_2\right\}) \\
&\leq 3\varepsilon_k \quad (\because t \in [0,1])
\end{aligned}
$$

よって，式 ($\clubsuit$3.2) と合わせて，次の不等式が成り立つ．

$$
\left\|\boldsymbol{G}_k - \boldsymbol{G}_\star\right\|_2 \leq 3L\varepsilon_k
$$

さらに，式 ($\clubsuit$3.1) と合わせて，次の不等式が成り立つ．

$$
\begin{aligned}
\left\|\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k\right\|_2 &= \left\|\boldsymbol{G}_\star^{-\frac{1}{2}} \left(\boldsymbol{G}_k - \boldsymbol{G}_\star\right) \boldsymbol{G}_\star^{-\frac{1}{2}} \tilde{\boldsymbol{s}}_k\right\|_2 \leq \left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\right\|_2^2 \left\|\boldsymbol{G}_k - \boldsymbol{G}_\star\right\|_2 \left\|\tilde{\boldsymbol{s}}_k\right\|_2 \\
&\leq 3L \left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\right\|_2^2 \varepsilon_k \left\|\tilde{\boldsymbol{s}}_k\right\|_2
\end{aligned}
$$

Theorem (♣4)

$$\lim_{k\to+\infty} \cos\tilde{\theta}_k = 1, \qquad \lim_{k\to+\infty} \tilde{p}_k = 1 \tag{A.19}$$

$c(>0)$ を次式で定義する．

$$c := 3L \left\| \boldsymbol{G}_\star^{-\frac{1}{2}} \right\|_2^2$$

すなわち，定理 ♣3 の結果について，次式が成り立つ．

$$\left\| \tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k \right\|_2 \leq 3L \left\| \boldsymbol{G}_\star^{-\frac{1}{2}} \right\|_2^2 \varepsilon_k \left\| \tilde{\boldsymbol{s}}_k \right\|_2 = c\varepsilon_k \left\| \tilde{\boldsymbol{s}}_k \right\|_2$$

このとき，三角不等式（$\tilde{\boldsymbol{s}_k} = (\tilde{\boldsymbol{s}}_k - \tilde{\boldsymbol{y}}_k) + \tilde{\boldsymbol{y}}_k$ と考える．）から，次の不等式が成り立つ．

$$\begin{aligned} & \left| \left\| \tilde{\boldsymbol{y}}_k \right\|_2 - \left\| \tilde{\boldsymbol{s}}_k \right\|_2 \right| \leq \left\| \tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k \right\|_2 \leq c\varepsilon_k \left\| \tilde{\boldsymbol{s}}_k \right\|_2 \\ \Longleftrightarrow\ & (1 - c\varepsilon_k) \left\| \tilde{\boldsymbol{s}}_k \right\|_2 \leq \left\| \tilde{\boldsymbol{y}}_k \right\|_2 \leq (1 + c\varepsilon_k) \left\| \tilde{\boldsymbol{s}}_k \right\|_2 \end{aligned}$$

前スライドの結果：$\|\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k\|_2 \le c\varepsilon_k \|\tilde{\boldsymbol{s}}_k\|_2, (1 - c\varepsilon_k)\|\tilde{\boldsymbol{s}}_k\|_2 \le \|\tilde{\boldsymbol{y}}_k\|_2$

$\varepsilon_k \to 0$ から，ある番号 $\bar{k}$ が存在して，任意の $k \ge \bar{k}$ に対して，
$\varepsilon_k < 1/c \iff 1 - c\varepsilon_k > 0$ を満たす．
これと，命題 2.1.1[14] と，前スライドの結果から，任意の $k \ge \bar{k}$ に対して，次式が成り立つ．

$$
\begin{aligned}
\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2 &= \frac{1}{2}\left(\|\tilde{\boldsymbol{y}}_k\|_2^2 + \|\tilde{\boldsymbol{s}}_k\|_2^2 - \|\tilde{\boldsymbol{y}}_k - \tilde{\boldsymbol{s}}_k\|_2^2 \right) \\
&\ge \frac{1}{2}\left\{ (1 - c\varepsilon_k)^2 + 1 - (c\varepsilon_k)^2 \right\} \|\tilde{\boldsymbol{s}}_k\|_2^2 = (1 - c\varepsilon_k)\|\tilde{\boldsymbol{s}}_k\|_2^2
\end{aligned}
$$

さらに，$\tilde{c}$ を $\tilde{c} > 3c$ を満たす正数とし，$\varepsilon (> 0)$ を次式で定義する．

$$
\varepsilon := \frac{\tilde{c} - 3c}{c(c + \tilde{c})}
$$

[14] $\|\boldsymbol{x} - \boldsymbol{y}\| = \|\boldsymbol{x}\|^2 - 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \|\boldsymbol{y}\|^2$

前スライドの式：$\forall k \geq \bar{k}\ (\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2 \geq (1 - c\varepsilon_k) \|\tilde{\boldsymbol{s}}_k\|_2^2)$

$\varepsilon_k \to 0$ から，ある番号 $\hat{k}$ が存在して，任意の $k \geq \hat{k} (\geq \bar{k})$ に対して，次式が成立する．

$$\varepsilon_k \leq \varepsilon := \frac{\tilde{c} - 3c}{c(c + \tilde{c})} < \frac{1}{c} \tag{A.18}$$

さらに，任意の $k \geq \hat{k}$ に対して，$0 < 1 - c\varepsilon_k < 1$ となることに注意すると，次式が成立する．

$$\begin{aligned} \tilde{M}_k := \frac{\|\tilde{\boldsymbol{y}}_k\|_2^2}{\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2} \leq \frac{(1 + c\varepsilon_k)^2 \|\tilde{\boldsymbol{s}}_k\|_2^2}{(1 - c\varepsilon_k) \|\tilde{\boldsymbol{s}}_k\|_2^2} &\leq \frac{(1 + c\varepsilon_k)^2}{1 - c\varepsilon_k} \quad (\because \|\tilde{\boldsymbol{y}}_k\|_2 \leq (1 + c\varepsilon_k) \|\tilde{\boldsymbol{s}}_k\|_2) \\ &= 1 + \frac{c\varepsilon_k (3 + c\varepsilon_k)}{1 - c\varepsilon_k} \leq 1 + \tilde{c}\varepsilon_k \quad (\because \text{(A.18) を } \tilde{c} \text{ について整理}) \end{aligned} \tag{♣4.1}$$

$$\tilde{m}_k := \frac{\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2}{\|\tilde{\boldsymbol{s}}_k\|_2^2} \geq \frac{(1 - c\varepsilon_k) \|\tilde{\boldsymbol{s}}_k\|_2^2}{\|\tilde{\boldsymbol{s}}_k\|_2^2} = 1 - c\varepsilon_k \quad (\because \forall k \geq \bar{k}\ (\langle \tilde{\boldsymbol{y}}_k, \tilde{\boldsymbol{s}}_k \rangle_2 \geq (1 - c\varepsilon_k) \|\tilde{\boldsymbol{s}}_k\|_2^2)) \tag{♣4.2}$$

$\exists \bar{k} \in \mathbb{N}\ \forall k \geq \bar{k}\ \left(\varepsilon_k < \frac{1}{c} \iff 1 - c\varepsilon_k > 0\right)$

式 ($\clubsuit$4.2)[15] について，両辺の対数を取ると，任意の $k \geq \hat{k}$ に対して，次式が成立する．

$$\log \tilde{m}_k \geq \log(1 - c\varepsilon_k) \geq 1 - \frac{1}{1 - c\varepsilon_k} = -\frac{c\varepsilon_k}{1 - c\varepsilon_k} > -\frac{c + \tilde{c}}{4}\varepsilon_k \qquad (\clubsuit 4.3)$$

ここで，二つ目の不等号について，$\forall x > 0 \ (1 - x + \log x \leq 0)$ と $\forall k \geq \hat{k} \ (\frac{1}{1 - c\varepsilon_k} > 0)$ を，最後の不等号については，式 (A.18)[16] から導かれる $1 - c\varepsilon_k > 1 - c\varepsilon = \frac{4c}{c + \tilde{c}}$ を用いた．

[15] $\forall k \geq \hat{k} \ (\tilde{m}_k \geq 1 - c\varepsilon_k) \quad (\clubsuit 4.2)$

[16] $\forall k \geq \hat{k} \ (\varepsilon_k \leq \varepsilon := \frac{\tilde{c} - 3c}{c(c + \tilde{c})} < \frac{1}{c}) \quad$ (A.18)

式 $(\clubsuit 4.2), (\clubsuit 4.3)$[17] を式 (A.18) に代入すると，任意の $k \geq \hat{k}$ に対して，次式が成り立つ．

$$\begin{aligned} 0 &< C\left(\tilde{\boldsymbol{B}}_{k+1}\right) \\ &= C\left(\tilde{\boldsymbol{B}}_k\right) + \left(\tilde{M}_k - \log \tilde{m}_k - 1\right) + \left(1 - \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k} + \log \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k}\right) + \log \cos^2 \tilde{\theta}_k \\ &\leq C\left(\tilde{\boldsymbol{B}}_k\right) + \left(\tilde{c} + \frac{c + \tilde{c}}{4}\right) \varepsilon_k + \left(1 - \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k} + \log \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k}\right) + \log \cos^2 \tilde{\theta}_k \end{aligned}$$

ここで，$\hat{c} := \tilde{c} + \frac{c+\tilde{c}}{4}$ として，$k = \hat{k}, \ldots, K$ で辺々足し合わせて，$K \to \infty$ とすると，次式が成り立つ．

$$\sum_{k=\hat{k}}^{+\infty} \left\{ \log \frac{1}{\cos^2 \tilde{\theta}_k} - \left(1 - \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k} + \log \frac{\tilde{p}_k}{\cos^2 \tilde{\theta}_k}\right) \right\} \leq C\left(\tilde{B}_{\hat{k}}\right) + \hat{c} \sum_{k=\hat{k}}^{+\infty} \varepsilon_k$$

[17] $\forall k \geq \hat{k}\ (\tilde{m}_k \geq 1 - c\varepsilon_k) \quad (\clubsuit 4.2), \log \tilde{m}_k > -\frac{c+\tilde{c}}{4} \varepsilon_k \quad (\clubsuit 4.3)$

前スライドの最後の式：

$$\sum_{k=\hat{k}}^{+\infty}\left\{\log\frac{1}{\cos^2\tilde{\theta}_k}-\left(1-\frac{\tilde{p}_k}{\cos^2\tilde{\theta}_k}+\log\frac{\tilde{p}_k}{\cos^2\tilde{\theta}_k}\right)\right\}\leq C\left(\tilde{B}_{\hat{k}}\right)+\hat{c}\sum_{k=\hat{k}}^{+\infty}\varepsilon_k$$

について，定理 ♣2 より，右辺はある値に収束する．

さらに，

$$0\leq\sum_{k=\hat{k}}^{+\infty}\log\frac{1}{\cos^2\tilde{\theta}_k}\quad(\because 0\leq\cos^2\theta\leq 1)$$

$$0\leq\sum_{k=\hat{k}}^{+\infty}\left\{-\left(1-\frac{\tilde{p}_k}{\cos^2\tilde{\theta}_k}+\log\frac{\tilde{p}_k}{\cos^2\tilde{\theta}_k}\right)\right\}\quad(\because 1-x+\log x\leq 0(x>0),\tilde{p}_k>0)$$

であるので，次式が成り立つ．

$$\lim_{k\to+\infty}\log\frac{1}{\cos^2\tilde{\theta}_k}=0,\lim_{k\to+\infty}\left(1-\frac{\tilde{p}_k}{\cos^2\tilde{\theta}_k}+\log\frac{\tilde{p}_k}{\cos^2\tilde{\theta}_k}\right)=0$$

$$\iff\lim_{k\to+\infty}\cos\tilde{\theta}_k=1,\quad\lim_{k\to+\infty}\tilde{p}_k=1$$

Theorem (♣5)

$$\lim_{k\to+\infty}\frac{\left\|\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2}{\left\|\boldsymbol{s}_k\right\|_2}=\lim_{k\to+\infty}\frac{\left\|\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{d}_k\right\|_2}{\left\|\boldsymbol{d}_k\right\|_2}=0 \tag{A.20}$$

$$\begin{aligned}\left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2 &= \left\langle\boldsymbol{G}_\star^{-\frac{1}{2}}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k,\boldsymbol{G}_\star^{-\frac{1}{2}}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\rangle_2\\ &= \left\langle\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k,\boldsymbol{G}_\star^{-1}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\rangle_2\\ &\geq \lambda_{\mathsf{min}}\left(\boldsymbol{G}_\star^{-1}\right)\left\|\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2 \quad \left(\because \lambda_{\min}(\boldsymbol{A})\|\boldsymbol{x}\|_2^2\leq\langle\boldsymbol{x},\boldsymbol{A}\boldsymbol{x}\rangle_2 \quad (2.14)\right)\end{aligned}$$

$$\iff \left\|\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2\leq\frac{\left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2}{\lambda_{\mathsf{min}}\left(\boldsymbol{G}_\star^{-1}\right)}$$

$$\iff \frac{\left\|\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2}{\left\|\boldsymbol{s}_k\right\|_2^2}\leq\frac{\left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2}{\lambda_{\mathsf{min}}\left(\boldsymbol{G}_\star^{-1}\right)\left\|\boldsymbol{s}_k\right\|_2^2}\leq\frac{\left\|\boldsymbol{G}_\star^{\frac{1}{2}}\right\|_2^2}{\lambda_{\mathsf{min}}\left(\boldsymbol{G}_\star^{-1}\right)}\frac{\left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\left(\boldsymbol{B}_k-\boldsymbol{G}_\star\right)\boldsymbol{s}_k\right\|_2^2}{\left\|\tilde{\boldsymbol{s}}_k\right\|_2^2}$$

$$\left(\because \left\|\tilde{\boldsymbol{s}}_k\right\|_2=\left\|\boldsymbol{G}_\star^{\frac{1}{2}}\boldsymbol{s}_{\boldsymbol{k}}\right\|_2\leq\left\|\boldsymbol{G}_\star^{\frac{1}{2}}\right\|_2\left\|\boldsymbol{s}_k\right\|_2 \iff \frac{1}{\left\|\boldsymbol{s}_k\right\|_2^2}\leq\frac{\left\|\boldsymbol{G}_\star^{\frac{1}{2}}\right\|_2^2}{\left\|\tilde{\boldsymbol{s}}_k\right\|_2^2}\right)$$

前スライド最後の式：$\dfrac{\left\|(\boldsymbol{B}_k - \boldsymbol{G}_\star)\,\boldsymbol{s}_k\right\|_2^2}{\|\boldsymbol{s}_k\|_2^2} \le \dfrac{\left\|\boldsymbol{G}_\star^{\frac{1}{2}}\right\|_2^2}{\lambda_{\min}\left(\boldsymbol{G}_\star^{-1}\right)} \dfrac{\left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\,(\boldsymbol{B}_k - \boldsymbol{G}_\star)\,\boldsymbol{s}_k\right\|_2^2}{\|\tilde{\boldsymbol{s}}_k\|_2^2}$

ここで，右辺の一部について，次式が成り立つ．

$$
\begin{aligned}
\frac{\left\|\boldsymbol{G}_\star^{-\frac{1}{2}}\,(\boldsymbol{B}_k - \boldsymbol{G}_\star)\,\boldsymbol{s}_k\right\|_2^2}{\|\tilde{\boldsymbol{s}}_k\|_2^2} &= \frac{\left\|\left(\tilde{\boldsymbol{B}}_k - \boldsymbol{I}\right)\tilde{\boldsymbol{s}}_k\right\|_2^2}{\|\tilde{\boldsymbol{s}}_k\|_2^2} \quad \left(\because\ \tilde{\boldsymbol{s}}_k := \boldsymbol{G}_\star^{\frac{1}{2}}\boldsymbol{s}_k\right) \\
&= \frac{1}{\|\tilde{\boldsymbol{s}}_k\|_2^2}\left(\left\|\tilde{\boldsymbol{B}}_k\tilde{\boldsymbol{s}}_k\right\|_2^2 - 2\left\langle \tilde{\boldsymbol{s}}_k, \tilde{\boldsymbol{B}}_k\tilde{\boldsymbol{s}}_k\right\rangle_2 + \|\tilde{\boldsymbol{s}}_k\|_2^2\right) \\
&= \frac{\tilde{p}_k^2}{\cos^2\tilde{\theta}_k} - 2\tilde{p}_k + 1 \to 0\ (\text{as } k \to \infty) \quad (\because\ \text{定理}\clubsuit 4)
\end{aligned}
$$

よって，次式が成り立つ．

$$
\lim_{k\to+\infty}\frac{\left\|(\boldsymbol{B}_k - \boldsymbol{G}_\star)\,\boldsymbol{s}_k\right\|_2}{\|\boldsymbol{s}_k\|_2} = \lim_{k\to+\infty}\frac{\left\|(\boldsymbol{B}_k - \boldsymbol{G}_\star)\,\boldsymbol{d}_k\right\|_2}{\|\boldsymbol{d}_k\|_2} = 0 \tag{A.20}
$$

□

$\cos\tilde{\theta}_k := \dfrac{\left\langle \tilde{\boldsymbol{s}}_k \tilde{\boldsymbol{B}}_k\tilde{\boldsymbol{s}}_k\right\rangle_2}{\|\tilde{\boldsymbol{s}}_k\|_2\left\|\tilde{\boldsymbol{B}}_k\tilde{\boldsymbol{s}}_k\right\|_2}, \quad \tilde{p}_k := \dfrac{\left\langle \tilde{\boldsymbol{s}}_k, \tilde{\boldsymbol{B}}_k\tilde{\boldsymbol{s}}_k\right\rangle_2}{\|\tilde{\boldsymbol{s}}_k\|_2^2}, \quad \boldsymbol{s}_k := \boldsymbol{x}_{k+1} - \boldsymbol{x}_k = \alpha_k\boldsymbol{d}_k$

Theorem (♣6)

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2}{\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2} = 0$$

$\lim_{k\to\infty} \boldsymbol{x}_k$ より，十分大きい反復回数 k の近似解 $\boldsymbol{x}_k$ における Hesse 行列は正定値となり，定理 5.3.1（Newton 法の収束性）から，$\boldsymbol{d}_k^{\mathrm{N}} := -\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\right)^{-1} \nabla f\left(\boldsymbol{x}_k\right)$ を探索方向に持つ Newton 法は 2 次収束する．

十分大きい k に対して，次式が成り立つ．

$$
\begin{aligned}
\left\|\boldsymbol{d}_k - \boldsymbol{d}_k^{\mathsf{N}}\right\|_2 &= \left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\right)^{-1}\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\boldsymbol{d}_k + \nabla f\left(\boldsymbol{x}_k\right)\right)\right\|_2 \quad \left(\because \boldsymbol{d}_k^{\mathsf{N}} := -\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\right)^{-1}\nabla f\left(\boldsymbol{x}_k\right)\right) \\
&= \left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\right)^{-1}\left(\nabla^2 f\left(\boldsymbol{x}_k\right) - \boldsymbol{B}_k\right)\boldsymbol{d}_k\right\|_2 \quad \left(\because \boldsymbol{d}_k := -\boldsymbol{B}_k^{-1}\nabla f(\boldsymbol{x}_k)\right) \\
&\le \left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\right)^{-1}\right\|_2 \left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right) - \boldsymbol{B}_k\right)\boldsymbol{d}_k\right\|_2 \\
&\le 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right) - \boldsymbol{B}_k\right)\boldsymbol{d}_k\right\|_2 \\
&\left(\because \left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right)\right)^{-1}\right\|_2 \le 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \quad (5.25)\right) \\
&= 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \left\|\left(\left(\nabla^2 f\left(\boldsymbol{x}_k\right) - \boldsymbol{G}_\star\right) + \left(\boldsymbol{G}_\star - \boldsymbol{B}_k\right)\right)\boldsymbol{d}_k\right\|_2 \\
&\le 2\|(\nabla^2 f(\boldsymbol{x}^\star))^{-1}\|_2 \left(\left\|\left(\nabla^2 f\left(\boldsymbol{x}_k\right) - \boldsymbol{G}_\star\right)\boldsymbol{d}_k\right\|_2 + \|\left(\left(\boldsymbol{G}_\star - \boldsymbol{B}_k\right)\right)\boldsymbol{d}_k\|_2\right)
\end{aligned}
$$

ここで，$\nabla^2 f\left(\boldsymbol{x}_k\right) \to \boldsymbol{G}_\star$ と式 (A.20)[18] より，次式が成り立つ．

$$
\left\|\boldsymbol{d}_k - \boldsymbol{d}_k^{\mathsf{N}}\right\|_2 = o\left(\|\boldsymbol{d}_k\|_2\right) \qquad (\clubsuit 6.1)
$$

[18] $\displaystyle\lim_{k\to+\infty}\frac{\|\left(\boldsymbol{B}_k - \boldsymbol{G}_\star\right)\boldsymbol{d}_k\|_2}{\|\boldsymbol{d}_k\|_2} = 0$

前スライドの式：$\left\|\boldsymbol{d}_k - \boldsymbol{d}_k^{\mathsf{N}}\right\|_2 = o\left(\|\boldsymbol{d}_k\|_2\right) \quad (\clubsuit 6.1)$

また，十分大きい k に対して，次式が成り立つ．

$$
\begin{aligned}
\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 &= \|\boldsymbol{x}_k + \boldsymbol{d}_k - \boldsymbol{x}^\star\|_2 \quad (\because \text{十分大きな } k \text{ に対して，} \alpha_k = 1) \\
&= \|(\boldsymbol{x}_k + \boldsymbol{d}_k^{\mathsf{N}} - \boldsymbol{x}^\star) + (\boldsymbol{d}_k - \boldsymbol{d}_k^{\mathsf{N}})\|_2 \\
&\leq \left\|\boldsymbol{x}_k + \boldsymbol{d}_k^{N} - \boldsymbol{x}^\star\right\|_2 + \left\|\boldsymbol{d}_k - \boldsymbol{d}_k^{N}\right\|_2 \quad (\because \text{三角不等式}) \\
&= O\left(\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2\right) + o\left(\|\boldsymbol{d}_k\|_2\right) \quad (\because \text{Newton 法の２次収束性と式 } (\clubsuit 6.1))
\end{aligned}
$$

よって，任意の $\varepsilon \in (0, 1/2)$ に対して，ある番号 k_1 が存在して，任意の $k \geq k_1$ に対して，$\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2 \leq \varepsilon$, かつ,

$$
\begin{aligned}
\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 &\leq M \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2^2 + \varepsilon\left(\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 + \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2\right) \\
&\leq M\varepsilon \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2 + \varepsilon\left(\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 + \|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2\right)
\end{aligned}
$$

つまり，

$$
\frac{1}{2}\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 \leq (1-\varepsilon)\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2 \leq \varepsilon(1+M)\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2
$$

が成立するので，（M : ε に依らない正定数）$\displaystyle\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^\star\|_2}{\|\boldsymbol{x}_k - \boldsymbol{x}^\star\|_2} = 0$