

「非凸最適化アルゴリズムとその計算量解析」

演習問題の解答例

丸茂 直貴

2024/8/24–26

問題 1. 基礎的問題

(a) 微分積分学の基本定理より,

$$\nabla f(x) - \nabla f(y) = \int_0^1 \nabla^2 f(y + t(x - y))(x - y) dt.$$

両辺のノルムをとり, $\|\nabla^2 f(x)\| \leq L$ を使うと,

$$\|\nabla f(x) - \nabla f(y)\| \leq \int_0^1 \|\nabla^2 f(y + t(x - y))(x - y)\| dt \leq \int_0^1 L\|x - y\| dt = L\|x - y\|$$

を得る.

(b) 方向微分と偏微分の関係より,

$$\nabla^2 f(x)v = \lim_{t \rightarrow 0} \frac{\nabla f(x + tv) - \nabla f(x)}{t}.$$

両辺のノルムをとり, ∇f の Lipschitz 連続性を使うと,

$$\|\nabla^2 f(x)v\| = \lim_{t \rightarrow 0} \frac{\|\nabla f(x + tv) - \nabla f(x)\|}{|t|} \leq L\|v\|$$

を得る. これが任意の $v \in \mathbb{R}^2$ に対し成り立つため, $\|\nabla^2 f(x)\| \leq L$ である.

(c) 例えば $f(x) = 100x^2$ とすると, f' の Lipschitz 定数は 200, f'' の Lipschitz 定数は 0 である.

(d) 例えば $f(x) = \frac{\sin(100x)}{10000}$ とすると, $f''(x) = -\sin(100x)$, $f'''(x) = -100 \cos(100x)$ であるため, f' の Lipschitz 定数は 1, f'' の Lipschitz 定数は 100 である.

問題 2. 正則化無しの Newton 法が収束しない強凸平滑関数の例

本問題は [Doikov, 2021, Example 1.4.3] を参考に作成しました.

(a)

$$g'(x) = \frac{x}{\sqrt{1+x^2}}, \quad g''(x) = \frac{1}{(1+x^2)^{3/2}}$$

である. $0 \leq g''(x) \leq 1$ より g は凸関数であり, g' は 1-Lipschitz 連続である.

(b)

$$g'''(x) = -\frac{3x}{(1+x^2)^{5/2}}$$

であり,

$$\begin{aligned} |g'''(x)| &= \frac{3|x|}{(1+x^2)^{5/2}} = 3 \left(|x|^{-2/5} + |x|^{8/5} \right)^{-5/2} \\ &= 3 \left(4 \cdot \frac{|x|^{-2/5}}{4} + |x|^{8/5} \right)^{-5/2} \\ &\leq 3 \left(5 \cdot 2^{-8/5} \right)^{-5/2} \end{aligned}$$

が成り立つ. 最後の不等号には AM-GM 不等式を用いた. $|g'''(x)|$ が有界であるため, g'' は Lipschitz 連続である.

(c)

$$h(x) := x - \frac{f'(x)}{f''(x)}$$

とおくと,

$$h(x) = x - \frac{g'(x) + \mu x}{g''(x) + \mu} = -\frac{g'(x) - xg''(x)}{g''(x) + \mu} = -\frac{x^3}{1 + \mu(1+x^2)^{3/2}}$$

である. 若干唐突だが, $h(x) = -x$ なる $x > 0$ について考える (これは $h(x)$ を小さい μ に対してプロットすると思いつける). $x > 0$ の下では $h(x) = -x$ は

$$\mu = \frac{x^2 - 1}{(x^2 + 1)^{3/2}}$$

と等価である. よって, $\mu > 0$ が集合

$$\left\{ \frac{x^2 - 1}{(x^2 + 1)^{3/2}} \mid x > 0 \right\}$$

に含まれるほど小さければ, $h(\alpha) = -\alpha$ なる $\alpha > 0$ が存在する. $x_0 = \alpha$ とすると, $x_1 = -\alpha$, $x_2 = \alpha$, $x_3 = -\alpha, \dots$ となり, Newton 法は収束しない. なお,

$$\max_{x>0} \left\{ \frac{x^2 - 1}{(x^2 + 1)^{3/2}} \right\} = \frac{\sqrt{6}}{9}$$

であるため, $\mu \leq \frac{\sqrt{6}}{9}$ のときに上記の α が存在する.

問題 3. ステップサイズを大きくすることも許す場合の最急降下法の計算量

アルゴリズムが ε -停留点を見つけるまでに, 5 行目が N_{suc} 回, 7 行目が N_{unsuc} 回実行されるとする.
総反復数は $N_{\text{suc}} + N_{\text{unsuc}}$ である.

アルゴリズム中の ℓ が ℓ_{max} を超えないことから,

$$\ell_{\text{init}} \beta^{N_{\text{suc}}} \alpha^{N_{\text{unsuc}}} \leq \ell_{\text{max}}$$

が成り立つ. 両辺の対数を取り,

$$N_{\text{unsuc}} \leq N_{\text{suc}} \log_{\alpha} \left(\frac{1}{\beta} \right) + \log_{\alpha} \left(\frac{\ell_{\text{max}}}{\ell_{\text{init}}} \right)$$

を得る. 一方, 講義中に示した $\beta = 1$ の場合と同様に

$$N_{\text{suc}} \leq \left\lceil \frac{2\ell_{\text{max}}\Delta}{\varepsilon^2} \right\rceil$$

が成り立つ. これらを合わせると,

$$\begin{aligned} N_{\text{suc}} + N_{\text{unsuc}} &\leq N_{\text{suc}} \log_{\alpha} \left(\frac{\alpha}{\beta} \right) + \log_{\alpha} \left(\frac{\ell_{\text{max}}}{\ell_{\text{init}}} \right) \\ &\leq \left\lceil \frac{2\ell_{\text{max}}\Delta}{\varepsilon^2} \right\rceil \log_{\alpha} \left(\frac{\alpha}{\beta} \right) + \log_{\alpha} \left(\frac{\ell_{\text{max}}}{\ell_{\text{init}}} \right) \end{aligned}$$

を得る.

問題 4. 最急降下法の収束が遅くなる関数の例

本問題は [Cartis et al., 2010] を参考に作成しました.

(a) 2 次関数

$$g(x) = \int_0^x \left(\left(1 - \frac{r_1}{r_0} \right) t - r_0 \right) dt = \left(1 - \frac{r_1}{r_0} \right) \frac{x^2}{2} - r_0 x$$

は条件を全て満たす. 例えば, $g(r_0) = -\frac{1}{2}r_0(r_0 + r_1) \geq -r_0^2$ である.

(b) 数列 $(r_k)_{k \in \mathbb{N}}$ を

$$r_k = (k+1)^{-\frac{1}{2-\delta}}$$

と定め, (a) と同様に関数の列 $(g_k)_{k \in \mathbb{N}}$ を

$$g_k(x) = \left(1 - \frac{r_{k+1}}{r_k} \right) \frac{x^2}{2} - r_k x$$

と定める. これを用いて関数 $f: \mathbb{R} \rightarrow \mathbb{R}$ を

$$f(x) := \begin{cases} -r_0 x & \text{for } x \leq 0, \\ g_k \left(x - \sum_{i=0}^{k-1} r_i \right) + \sum_{i=0}^{k-1} g_i(r_i) & \text{for } \sum_{i=0}^{k-1} r_i \leq x \leq \sum_{i=0}^k r_i \quad (k = 0, 1, \dots) \end{cases}$$

と定める. すると f は連続的微分可能である. この f と

$$L = 1, \quad \Delta = \frac{2}{\delta}$$

が条件 (i)–(iii) を満たすことを示す.

まず, (a) と f の定義から f' は 1-Lipschitz 連続であるため, 条件 (i) を満たす.

次に, f が $\mathbb{R}$ 上で単調減少であることと,

$$\lim_{x \rightarrow +\infty} f(x) = \sum_{i=0}^{\infty} g_i(r_i) \geq -\sum_{i=0}^{\infty} r_i^2 = -\sum_{k=1}^{\infty} k^{-\frac{2}{2-\delta}} \geq -\frac{2}{\delta}$$

を合わせると,

$$f(0) - \inf_{x \in \mathbb{R}} f(x) \leq \frac{2}{\delta} = \Delta$$

が得られ, 条件 (ii) が示された.

最後に条件 (iii) を考える. 帰納的に $x_k = \sum_{i=0}^{k-1} r_i$ と $f'(x_k) = -r_k = -(k+1)^{-\frac{1}{2-\delta}}$ がいえる. よって $|f'(x_k)| \leq \varepsilon$ を満たす最小の k は

$$\frac{1}{\varepsilon^{2-\delta}} - 1 = L \cdot \frac{\delta \Delta}{2} \cdot \frac{1}{\varepsilon^{2-\delta}} - 1$$

以上であり, 条件 (iii) も成り立つ.

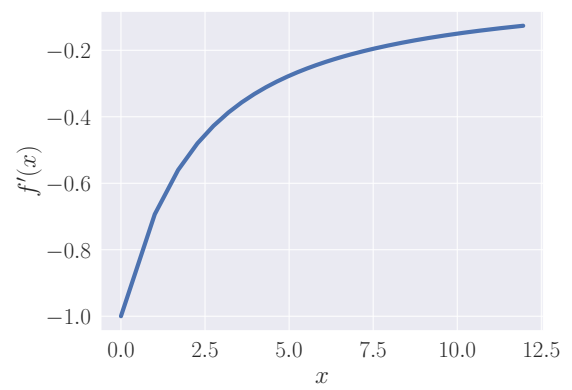
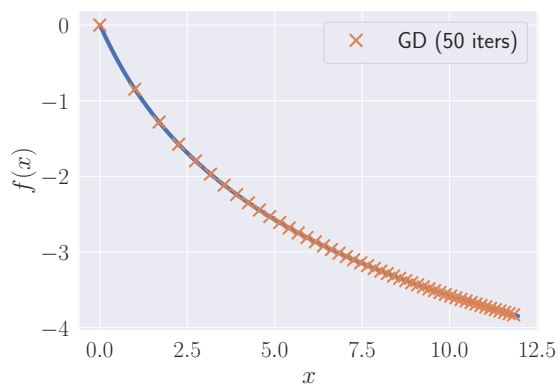


図 1: 問題 4(b) で構成した関数 f に最急降下法を適用した結果 ($\delta = 0.1$)

問題 5. 3 次正則化 Newton 法の子問題の性質

本問題は [Cartis et al., 2011, Theorem 3.1] を参考に作成しました.

- (a) 1 次の最適性条件より $\nabla F(u^*) = \mathbf{0}$ が成り立つ. これと

$$\begin{aligned}\nabla f(u^*) &= g + Hu^* + \frac{M}{2}\|u^*\|u^* \\ &= g + (H + \sigma I)u^*\end{aligned}$$

より示された.

- (b) 2 次の最適性条件より $\nabla^2 f(u^*) \succeq O$ が成り立つ. これと

$$\begin{aligned}\nabla^2 f(u^*) &= H + \frac{M}{2}\left(\|u^*\|I + \frac{1}{\|u^*\|}u^*(u^*)^\top\right) \\ &= H + \sigma I + \sigma\left(\frac{u^*}{\|u^*\|}\right)\left(\frac{u^*}{\|u^*\|}\right)^\top\end{aligned}$$

より示された.

- (c) $\|u^* + \alpha v\| = \|u^*\|$ かつ $\alpha \neq 0$ となるように

$$\alpha = -2\frac{\langle u^*, v \rangle}{\|v\|^2}$$

ととってみる. このとき,

$$\begin{aligned}f(u^* + \alpha v) - f(u^*) &= \left(\langle g, u^* + \alpha v \rangle + \frac{1}{2}\|u^* + \alpha v\|_H^2\right) - \left(\langle g, u^* \rangle + \frac{1}{2}\|u^*\|_H^2\right) \\ &= \alpha\langle g + Hu^*, v \rangle + \frac{\alpha^2}{2}\|v\|_H^2 \\ &= -\alpha\sigma\langle u^*, v \rangle + \frac{\alpha^2}{2}\|v\|_H^2 && ((a) \text{ より}) \\ &= \frac{\alpha^2\sigma}{2}\|v\|^2 + \frac{\alpha^2}{2}\|v\|_H^2 && (\alpha \text{ の定義}) \\ &= \frac{\alpha^2}{2}\|v\|_{H+\sigma I}^2\end{aligned}$$

が成り立つ.

- (d) (b) より, $\langle u^*, v \rangle = 0$ なる任意の $v \in \mathbb{R}^d$ に対し $\|v\|_{H+\sigma I}^2 \geq 0$ である. また (c) より, $\langle u^*, v \rangle \neq 0$ なる任意の $v \in \mathbb{R}^d$ に対しても $\|v\|_{H+\sigma I}^2 \geq 0$ である. 以上で示された.

問題 6. Hesse 行列の Lipschitz 連続性から従う Hessian-free 不等式

本問題は [Marumo and Takeda, 2024a, Lemmas 3.1 and 3.2] をもとに作成しました.

(a) 微分積分学の基本定理より

$$\nabla f(\bar{x}) - \nabla f(x_i) = \int_0^1 \nabla^2 f(\bar{x} + t(x_i - \bar{x}))(\bar{x} - x_i) dt.$$

これに λ_i をかけ、 i について足すと、

$$\begin{aligned} \nabla f(\bar{x}) - \sum_{i=1}^n \lambda_i \nabla f(x_i) &= \sum_{i=1}^n \lambda_i \int_0^1 \nabla^2 f(\bar{x} + t(x_i - \bar{x}))(\bar{x} - x_i) dt \\ &= \sum_{i=1}^n \lambda_i \int_0^1 (\nabla^2 f(\bar{x} + t(x_i - \bar{x})) - \nabla^2 f(\bar{x}))(\bar{x} - x_i) dt \end{aligned}$$

を得る. ノルムをとり、 $\nabla^2 f$ の M -Lipschitz 連続性を使うと、

$$\begin{aligned} \left\| \nabla f(\bar{x}) - \sum_{i=1}^n \lambda_i \nabla f(x_i) \right\| &\leq \sum_{i=1}^n \lambda_i \int_0^1 \left\| \nabla^2 f(\bar{x} + t(x_i - \bar{x})) - \nabla^2 f(\bar{x}) \right\| \|\bar{x} - x_i\| dt \\ &\leq \sum_{i=1}^n \lambda_i \int_0^1 M t \|x_i - \bar{x}\|^2 dt = \frac{M}{2} \sum_{i=1}^n \lambda_i \|x_i - \bar{x}\|^2 \end{aligned}$$

を得る. これで不等号が示された. 等号は以下のように示される:

$$\begin{aligned} \sum_{1 \leq i < j \leq n} \lambda_i \lambda_j \|x_i - x_j\|^2 &= \frac{1}{2} \sum_{i,j=1}^n \lambda_i \lambda_j \|x_i - x_j\|^2 \\ &= \frac{1}{2} \sum_{i,j=1}^n \lambda_i \lambda_j \|x_i\|^2 + \frac{1}{2} \sum_{i,j=1}^n \lambda_i \lambda_j \|x_j\|^2 - \sum_{i,j=1}^n \langle \lambda_i x_i, \lambda_j x_j \rangle \\ &= \sum_{i=1}^n \lambda_i \|x_i\|^2 - \|\bar{x}\|^2 = \sum_{i=1}^n \lambda_i \|x_i - \bar{x}\|^2. \end{aligned}$$

(b)

$$\begin{aligned} f(x) - f(y) &= \int_0^1 \langle \nabla f(tx + (1-t)y), x - y \rangle dt, \\ \frac{1}{2} \langle \nabla f(x) + \nabla f(y), x - y \rangle &= \int_0^1 \langle t \nabla f(x) + (1-t) \nabla f(y), x - y \rangle dt \end{aligned}$$

より、

$$\begin{aligned} f(x) - f(y) - \frac{1}{2} \langle \nabla f(x) + \nabla f(y), x - y \rangle &= \int_0^1 \langle \nabla f(tx + (1-t)y) - t \nabla f(x) - (1-t) \nabla f(y), x - y \rangle dt \\ &\leq \int_0^1 \left\| \nabla f(tx + (1-t)y) - t \nabla f(x) - (1-t) \nabla f(y) \right\| \|x - y\| dt \end{aligned}$$

が成り立つ. 1 つ目のノルムは、(a) の不等式で $n = 2$ としたものを使うと

$$\left\| \nabla f(tx + (1-t)y) - t \nabla f(x) - (1-t) \nabla f(y) \right\| \leq \frac{M}{2} t(1-t) \|x - y\|^2$$

と評価できる. あとは積分計算 $\int_0^1 t(1-t) dt = \frac{1}{6}$ をすれば示すべき式を得る.

問題 7. Heavy-ball 法における関数値減少量の評価

本問題は [Marumo and Takeda, 2024b] をもとに作成しました。

(a) 更新式は

$$u_k = u_{k-1} - \frac{1}{L} \nabla f(x_k)$$

と書き直せる。これを使って $\nabla^2 f$ の Lipschitz 連続性から従う不等式を変形すると、

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \frac{1}{2} \langle \nabla f(x_k) + \nabla f(x_{k+1}), u_k \rangle + \frac{M}{12} \|u_k\|^3 \\ &= \frac{L}{2} \langle (u_{k-1} - u_k) + (u_k - u_{k+1}), u_k \rangle + \frac{M}{12} \|u_k\|^3 \\ &= \frac{L}{2} \langle u_{k-1}, u_k \rangle - \frac{L}{2} \langle u_k, u_{k+1} \rangle + \frac{M}{12} \|u_k\|^3 \end{aligned}$$

を得る。

(b) (a) の不等式の添字 k を i に取り替え、 $0 \leq i < k$ について足し合わせると

$$f(x_k) - f(x_0) \leq -\frac{L}{2} \langle u_{k-1}, u_k \rangle + \frac{M}{12} \sum_{i=0}^{k-1} \|u_i\|^3$$

を得る（これは $k=0$ でも成立する）。一方、 ∇f の Lipschitz 連続性から

$$\begin{aligned} f(x_{k+1}) - f(x_k) &\leq \langle \nabla f(x_k), u_k \rangle + \frac{L}{2} \|u_k\|^2 = L \langle u_{k-1} - u_k, u_k \rangle + \frac{L}{2} \|u_k\|^2 \\ &= L \langle u_{k-1}, u_k \rangle - \frac{L}{2} \|u_k\|^2 \end{aligned}$$

も成り立つ。これを $1/2$ 倍して先の不等式に足すと示すべき式を得る。

(c) (b) の不等式の添字 k を i に取り替え、 $0 \leq i < k$ について足し合わせると

$$\begin{aligned} \left(\frac{1}{2} f(x_k) + \sum_{i=1}^{k-1} f(x_i) \right) - \left(k - \frac{1}{2} \right) f(x_0) &\leq -\frac{LS_k}{4} + \frac{M}{12} \sum_{j=0}^{k-1} \sum_{i=0}^{j-1} \|u_i\|^3 \\ &\leq -\frac{LS_k}{4} + \frac{M}{12} (k-1) \sum_{i=0}^{k-2} \|u_i\|^3 \end{aligned}$$

を得る。左辺第 1 項を $(k - \frac{1}{2}) \min_{1 \leq i \leq k} f(x_i)$ で下から押さえ、両辺を $k - \frac{1}{2} = \frac{2k-1}{2}$ で割ると、示すべき式を得る。

(d) $k=1$ の場合は容易。以下 $k \geq 2$ を仮定する。 ℓ_p ノルムの単調性より、 $\sum_{i=0}^{k-2} \|u_i\|^3 \leq (S_{k-1})^{3/2}$ が成り立つ。これを (c) の不等式に適用し少々計算すると、以下のように示すべき不等式を得る：

$$\begin{aligned} \min_{1 \leq i \leq k} f(x_i) - f(x_0) &\leq \frac{1}{2k-1} \left(-\frac{LS_k}{2} + \frac{M}{6} (k-1) (S_{k-1})^{3/2} \right) \\ &\leq \frac{1}{2k-1} \left(-\frac{L}{2} + \frac{M}{6} (k-1) \sqrt{S_{k-1}} \right) S_k \quad (S_{k-1} \leq S_k \text{ より}) \\ &\leq \frac{1}{2k-1} \left(-\frac{L}{2} + \frac{L}{8(k-1)^{3/2}} \right) S_k \quad (\sqrt{(k-1)^5 S_{k-1}} \leq \frac{3L}{4M} \text{ より}) \\ &\leq \frac{1}{2k-1} \left(-\frac{L}{2} + \frac{L}{4k} \right) S_k \quad (k \geq 2 \text{ より}) \\ &= -\frac{LS_k}{4k}. \end{aligned}$$

問題 8. Heavy-ball 法における勾配ノルムの評価

本問題は [Marumo and Takeda, 2024b] をもとに作成しました。

(a) 三角不等式と Cauchy–Schwarz の不等式を使うと,

$$\|x_i - x_j\|^2 \leq \left(\sum_{l=i}^{j-1} \|u_l\| \right)^2 \leq (j-i) \sum_{l=i}^{j-1} \|u_l\|^2 \leq k \sum_{l=0}^{k-2} \|u_l\|^2 = kS_{k-1}$$

を得る。

(b) Heavy-ball 法の更新式を足し合わせると,

$$u_{k-1} = -\frac{1}{L} \sum_{i=0}^{k-1} \nabla f(x_i)$$

を得る。一方, $\nabla^2 f$ の Lipschitz 連続性から従う Hessian-free 不等式と (a) の不等式を使うと,

$$\left\| \nabla f(\bar{x}_k) - \frac{1}{k} \sum_{i=0}^{k-1} \nabla f(x_i) \right\| \leq \frac{M}{2k^2} \sum_{0 \leq i < j < k} \|x_i - x_j\|^2 \leq \frac{M}{2k} \sum_{0 \leq i < j < k} S_{k-1} = \frac{M}{4} (k-1) S_{k-1}$$

を得る。これらを組み合わせると示すべき不等式を得る。

(c) (b) の不等式の添字 k を i に取り替え, $1 \leq i \leq k$ について足し合わせると,

$$\sum_{i=1}^k i \|\nabla f(\bar{x}_i)\| \leq L \sum_{i=1}^k \|u_{i-1}\| + \frac{M}{4} \sum_{i=1}^k i(i-1) S_{i-1}$$

を得る。右辺の各項を

$$\begin{aligned} \sum_{i=1}^k \|u_{i-1}\| &\leq \sqrt{kS_k} \quad (\text{Cauchy–Schwarz の不等式より}), \\ \sum_{i=1}^k i(i-1) S_{i-1} &\leq S_{k-1} \sum_{i=1}^k i(i-1) = \frac{1}{3} k(k^2-1) S_{k-1} \end{aligned}$$

と押さえると, 示すべき不等式を得る。

(d) $k=1$ の場合は容易。以下 $k \geq 2$ を仮定する。示すべき不等式は以下のように得られる:

$$\begin{aligned} \min_{1 \leq i \leq k} \|\nabla f(\bar{x}_i)\| &\leq \frac{2}{k(k+1)} \sum_{i=1}^k i \|\nabla f(\bar{x}_i)\| \\ &\leq 2L \sqrt{\frac{S_k}{k(k+1)^2}} + \frac{M}{6} (k-1) S_{k-1} \quad ((c) \text{ の不等式より}) \\ &\leq 2L \sqrt{\frac{S_k}{k^3}} + \frac{M}{6} (k-1) \sqrt{S_{k-1}} \sqrt{S_k} \\ &\leq 2L \sqrt{\frac{S_k}{k^3}} + \frac{L}{8(k-1)^{3/2}} \sqrt{S_k} \quad (\sqrt{(k-1)^5 S_{k-1}} \leq \frac{3L}{4M} \text{ より}) \\ &\leq 2L \sqrt{\frac{S_k}{k^3}} + \frac{L}{8(k/2)^{3/2}} \sqrt{S_k} \quad (k \geq 2 \text{ より}) \\ &\leq 3L \sqrt{\frac{S_k}{k^3}}. \end{aligned}$$

問題 9. Levenberg–Marquardt 法の計算量 (難)

本問題は [Marumo et al., 2022, 2023] をもとに作成しました．以下では，表記の簡略化のために $F_k := F(x_k)$, $J_k := J(x_k)$ と書きます．

(a) $\bar{f}_k$ の λ_k -強凸性と $\nabla \bar{f}_k(x_{k+1}) = \mathbf{0}$ より，

$$\begin{aligned}\bar{f}_k(x_k) - \bar{f}_k(x_{k+1}) &\geq \langle \nabla \bar{f}_k(x_{k+1}), x_{k+1} - x_k \rangle + \frac{\lambda_k}{2} \|x_{k+1} - x_k\|^2 \\ &= \frac{\lambda_k}{2} \|x_{k+1} - x_k\|^2\end{aligned}$$

が成り立つ．これと $\bar{f}_k(x_k) = f(x_k)$ より示された．

(b) J の Lipschitz 連続性から従う不等式

$$\|F_{k+1} - F_k - J_k u_k\| \leq \frac{L}{2} \|u_k\|^2 \quad (1)$$

を使うと，

$$\begin{aligned}f(x_{k+1}) - \bar{f}_k(x_{k+1}) &= \frac{1}{2} \|F_{k+1}\|^2 - \frac{1}{2} \|F_k + J_k u_k\|^2 - \frac{\lambda_k}{2} \|u_k\|^2 \\ &\leq \frac{1}{2} \left(\|F_k + J_k u_k\| + \frac{L}{2} \|u_k\|^2 \right)^2 - \frac{1}{2} \|F_k + J_k u_k\|^2 - \frac{\lambda_k}{2} \|u_k\|^2 \\ &= \frac{L}{2} \|F_k + J_k u_k\| \|u_k\|^2 + \frac{L^2}{8} \|u_k\|^4 - \frac{\lambda_k}{2} \|u_k\|^2\end{aligned}$$

を得る．この右辺第一項に

$$\begin{aligned}\|F_k + J_k u_k\| &= \sqrt{2\bar{f}_k(x_{k+1}) - \lambda_k \|u_k\|^2} \\ &\leq \sqrt{2f(x_k) - 2\lambda_k \|u_k\|^2} \quad ((a) \text{ より}) \\ &= \sqrt{\frac{\lambda_k^2}{L^2} - 2\lambda_k \|u_k\|^2} \quad (\lambda_k = L\|F_k\| \text{ より}) \\ &\leq \frac{\lambda_k}{L} \left(1 - \frac{L^2}{\lambda_k} \|u_k\|^2 \right) \quad (\sqrt{1-2x} \leq 1-x \text{ より})\end{aligned}$$

を使うと，

$$\begin{aligned}f(x_{k+1}) - \bar{f}_k(x_{k+1}) &\leq \frac{\lambda_k}{2} \left(1 - \frac{L^2}{\lambda_k} \|u_k\|^2 \right) \|u_k\|^2 + \frac{L^2}{8} \|u_k\|^4 - \frac{\lambda_k}{2} \|u_k\|^2 \\ &= -\frac{3L^2}{8} \|u_k\|^4 \leq 0\end{aligned}$$

を得る．

(c) (a), (b) を合わせると，

$$f(x_{k+1}) \leq f(x_k) - \frac{\lambda_k}{2} \|u_k\|^2 \quad (2)$$

を得る．すなわち， $f(x_k)$ や $\lambda_k = L\|F_k\|$ は単調減少である．この事実はこれ以降の小問でも使う．

$\nabla f(x_{k+1})$ を

$$\begin{aligned}\nabla f(x_{k+1}) &= \nabla f(x_{k+1}) - \nabla \bar{f}_k(x_{k+1}) \\ &= J_{k+1}^\top F_{k+1} - J_k^\top (F_k + J_k u_k) - \lambda_k u_k \\ &= (J_{k+1} - J_k)^\top F_{k+1} + J_k^\top (F_{k+1} - F_k - J_k u_k) - \lambda_k u_k\end{aligned}$$

と 3 項に分解する．第 1 項のノルムは

$$\begin{aligned}\|(J_{k+1} - J_k)^\top F_{k+1}\| &\leq \|J_{k+1} - J_k\| \|F_{k+1}\| \\ &\leq L \|u_k\| \|F_{k+1}\| \quad (J \text{ の Lipschitz 連続性より}) \\ &\leq \lambda_k \|u_k\| \quad (\lambda_k = L \|F_k\| \geq L \|F_{k+1}\| \text{ より})\end{aligned}$$

と，第 2 項のノルムは

$$\begin{aligned}\|J_k^\top (F_{k+1} - F_k - J_k u_k)\| &\leq \sigma \|F_{k+1} - F_k - J_k u_k\| \quad (\text{仮定 } \|J_k\| \leq \sigma \text{ より}) \\ &\leq \frac{L\sigma}{2} \|u_k\|^2 \quad ((1) \text{ より})\end{aligned}$$

と押さえられる．以上より示すべき式を得る．

(d) Cauchy–Schwarz の不等式を用いると，

$$\sum_{i=0}^{k-1} \lambda_i \|u_i\| \leq \sqrt{\left(\sum_{i=0}^{k-1} \lambda_i\right) \left(\sum_{i=0}^{k-1} \lambda_i \|u_i\|^2\right)} \leq \sqrt{(k\lambda_0) \cdot (2\Delta)} = \sqrt{2L\Delta \|F(x_0)\| k}$$

を得る．2 つ目の不等号には λ_k の単調減少性と (2) を用いた．

(e) (2) は λ_k を使って

$$\lambda_{k+1}^2 \leq \lambda_k^2 - L^2 \lambda_k \|u_k\|^2$$

と書き直せる．これと $\lambda_{k+1} \leq \lambda_k$ を使うと

$$\|u_k\|^2 \leq \frac{\lambda_k^2 - \lambda_{k+1}^2}{L^2 \lambda_k} = \frac{\lambda_k + \lambda_{k+1}}{\lambda_k} \frac{\lambda_k - \lambda_{k+1}}{L^2} \leq 2 \frac{\lambda_k - \lambda_{k+1}}{L^2}$$

を得る．

(f) (c) の不等式の添字 k を i に取り替えて $0 \leq i < k$ について足し合わせ，(d)，(e) を適用すると示すべき式を得る．

参考文献

- C. Cartis, N. I. M. Gould, and P. L. Toint. On the complexity of steepest descent, Newton’s and regularized Newton’s methods for nonconvex unconstrained optimization problems. *SIAM Journal on Optimization*, 20(6):2833–2852, 2010. URL <https://doi.org/10.1137/090774100>.
- C. Cartis, N. I. M. Gould, and P. L. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: Motivation, convergence and numerical results. *Mathematical Programming*, 127(2):245–295, 2011. URL <https://doi.org/10.1007/s10107-009-0286-5>.

- N. Doikov. *New second-order and tensor methods in convex optimization*. PhD thesis, Université catholique de Louvain, 2021.
- N. Marumo and A. Takeda. Parameter-free accelerated gradient descent for nonconvex minimization. *SIAM Journal on Optimization*, 34(2):2093–2120, 2024a. URL <https://doi.org/10.1137/22M1540934>.
- N. Marumo and A. Takeda. Universal heavy-ball method for nonconvex optimization under Hölder continuous Hessians. *Mathematical Programming*, 2024b. URL <https://doi.org/10.1007/s10107-024-02100-4>.
- N. Marumo, T. Okuno, and A. Takeda. Accelerated-gradient-based generalized Levenberg–Marquardt method with oracle complexity bound and local quadratic convergence. *arXiv preprint*, 2022. URL <https://arxiv.org/abs/2204.12016>.
- N. Marumo, T. Okuno, and A. Takeda. Majorization-minimization-based Levenberg–Marquardt method for constrained nonlinear least squares. *Computational Optimization and Applications*, 84(3):833–874, 2023. URL <https://doi.org/10.1007/s10589-022-00447-y>.