

日本 OR 学会 2025 年春季研究発表会

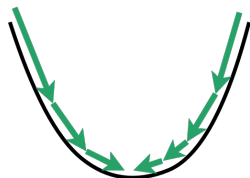
非有界な凸関数に対する加速勾配法の収束性

坂部圭哉 (東京大学)

成蹊大学

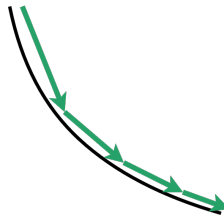
2025 年 3 月 6 日

(通常の) 凸最適化



最適解に収束

“非有界凸最適化”



無限遠に発散

このような状況における最適化手法の振る舞いは？ 発散の速度や方向は？

動機： 有界性が明らかでない凸最適化問題 (幾何計画問題・行列スケーリング)

→ 通常： 有界性判定 → 最適化手法

本研究： 非有界な場合でも最適化手法を適用

$f: \mathbb{R}^n \rightarrow \mathbb{R}$: 凸関数

凸共役関数 (Legendre–Fenchel 変換)

- 凸共役関数 $f^*(p) := \sup_{x \in \mathbb{R}^n} \langle x, p \rangle - f(x) \quad (p \in \mathbb{R}^n)$
- 共役空間 $\text{dom } f^* := \{p \in \mathbb{R}^n \mid f^*(p) < \infty\}$

性質:

- f^* は凸関数
- $\text{dom } f^*$ は凸集合
- f の有界性: $\inf_{x \in \mathbb{R}^n} f(x) > -\infty \iff 0 \in \text{dom } f^*$
- f の平滑性のもと, $\nabla f(\mathbb{R}^n) \subseteq \text{dom } f^* \subseteq \overline{\nabla f(\mathbb{R}^n)}$ ($\overline{}$: 閉包)

$\langle \cdot, \cdot \rangle$: Euclid 内積, $\nabla f(\mathbb{R}^n) := \{\nabla f(x) \mid x \in \mathbb{R}^n\}$

既存の結果：最急降下法

3/13

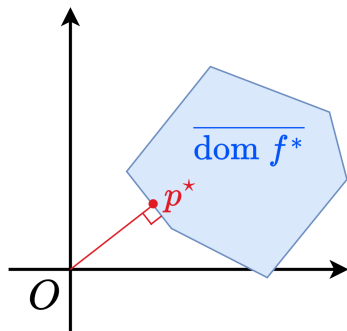
命題 [Hirai and Sakabe, 2024]

$f: \mathbb{R}^n \rightarrow \mathbb{R}$: L -平滑凸

$p^* : \overline{\text{dom } f^*}$ の最小ノルム点

$p^* \neq 0$ ならば,
最急降下法 $x^{(k+1)} := x^{(k)} - \nabla f(x^{(k)})/L$ について,

$$\lim_{k \rightarrow \infty} \nabla f(x^{(k)}) = p^*.$$



疑問：

- 収束レートが得られないか？
- 実用上有用な加速勾配法ではどうか？

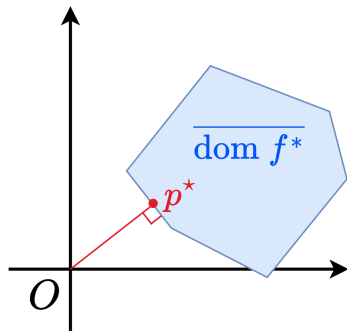
命題 [Hirai and Sakabe, 2024]

$f: \mathbb{R}^n \rightarrow \mathbb{R}$: L -平滑凸

$p^* : \overline{\text{dom } f^*}$ の最小ノルム点

$p^* \neq 0$ ならば,
最急降下法 $x^{(k+1)} := x^{(k)} - \nabla f(x^{(k)})/L$ について,

$$\lim_{k \rightarrow \infty} \nabla f(x^{(k)}) = p^*.$$



疑問：

- 収束レートが得られないか？
- 実用上有用な加速勾配法ではどうか？

本研究： 条件 $p^* \in \text{dom } f^*$ のもと、加速勾配法 [Ushiyama et al., 2023] が同様の収束性を示すことを、収束レート付きで示した。

※最急降下法についても $p^* \in \text{dom } f^*$ のもと収束レートが得られるが、本日は割愛。

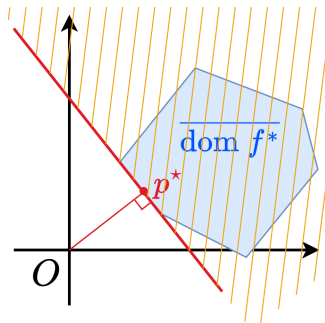
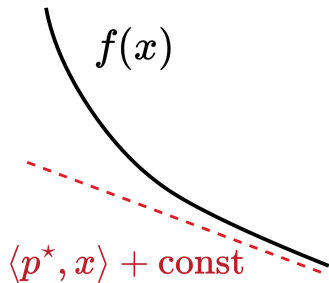
条件 $p^* \in \text{dom } f^*$

5/13

凸関数 $g(x) := f(x) - \langle p^*, x \rangle$ は以下の 2 つの性質を持つ：

- $\inf_{x \in \mathbb{R}^n} g(x) > -\infty$
($\iff f^*(p^*) < \infty \iff p^* \in \text{dom } f^*$)

- $\langle \nabla g(x), p^* \rangle \geq 0 \quad (\forall x \in \mathbb{R}^n)$
(p^* が $\overline{\nabla f(\mathbb{R}^n)} = \overline{\text{dom } f^*}$ の
最小ノルム点より)



$f: \mathbb{R}^n \rightarrow \mathbb{R}$: L -平滑凸

加速勾配法 [Ushiyama et al., 2023]

$$\begin{cases} z^{(k)} := \frac{k}{k+2}x^{(k)} + \frac{2}{k+2}v^{(k)} \\ v^{(k+1)} := v^{(k)} - \frac{k+1}{2L}\nabla f(z^{(k)}) \\ x^{(k+1)} := \frac{k}{k+2}x^{(k)} + \frac{2}{k+2}v^{(k+1)} \\ x^{(0)} := v^{(0)} := x_0 \text{ (初期点)} \end{cases}$$

連続極限



離散化

cf. Nesterov の加速勾配流

[Su et al., 2016, Wilson et al., 2021]

$$\begin{cases} \dot{v}(t) = -\frac{t}{2}\nabla f(x(t)) \\ \dot{x}(t) = \frac{2}{t}(v(t) - x(t)) \\ x(0) := v(0) := x_0 \text{ (初期点)} \end{cases}$$

- 加速勾配流の Lyapunov 関数の散逸性を自然に保つ離散化 (後述)
- f に最適解 x^* が存在するならば, $f(x^{(k)}) - f(x^*) = \mathcal{O}(k^{-2})$
→一次法の理論的な最速レート (e.g., [Nesterov, 2018])

※ 元論文 [Ushiyama et al., 2023] のパラメタに具体的な値を代入している.

$f: \mathbb{R}^n \rightarrow \mathbb{R}$: L -平滑凸

p^* : $\overline{\text{dom } f^*}$ の最小ノルム点

定理 (本研究)

$(x^{(k)})_k$: 前頁の加速勾配法による点列.

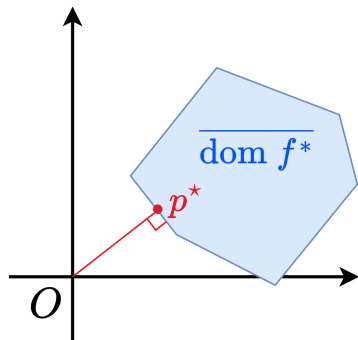
$p^* \neq 0$, $p^* \in \text{dom } f^*$ ならば,

$p^{(k)} := -\frac{12L}{3k+2} (x^{(k)} - x^{(k-1)})$ について,

- $p^{(k)} \in \text{dom } f^*$
- $\|p^{(k)} - p^*\| = \mathcal{O}(k^{-1})$
- $\|p^{(k)}\| - \|p^*\| = \mathcal{O}(k^{-2})$

★ 最急降下法の $-(x^{(k+1)} - x^{(k)}) \propto \nabla f(x^{(k)}) \rightarrow p^*$ と自然に対応

※ $\nabla f(x) \simeq p^*$ なる x を見つけるわけではない



以下の Lyapunov 関数 E_y が加速勾配法の収束性解析に有用 ($y \in \mathbb{R}^n$: 基準点) :

Lyapunov 関数 (下に有界な場合)

$$E_y^{(k)} := \frac{k(k+1)}{L} \left(f(x^{(k)}) - f(y) \right) + 2 \|v^{(k)} - y\|^2$$

命題： $E_y^{(k+1)} - E_y^{(k)} \leq 0$ [Ushiyama et al., 2023]

従って、もし f の最適解 x^* が存在するならば、 $y = x^*$ とすることで

$$\begin{aligned} \frac{k(k+1)}{L} \left(f(x^{(k)}) - f(x^*) \right) &\leq E_{x^*}^{(k)} \leq E_{x^*}^{(0)} \\ \implies f(x^{(k)}) - f(x^*) &\leq \frac{LE_{x^*}^{(0)}}{k(k+1)} = \mathcal{O}(k^{-2}) \end{aligned}$$

下に非有界な場合の Lyapunov 関数 (本研究)

9/13

f が下に非有界なときは、以下のように修正した Lyapunov 関数 $\mathcal{E}_y$ を提案：

Lyapunov 関数 (下に非有界な場合)

$$\mathcal{E}_y^{(k)} := \frac{k(k+1)}{L} \left(g(x^{(k)}) - g(y) \right) + 2 \left\| v^{(k)} + \frac{k(k+1)}{4L} p^* - y \right\|^2$$

再掲： $p^* : \overline{\text{dom } f^*}$ の最小ノルム点， $g(x) := f(x) - \langle p^*, x \rangle$

※ $p^* = 0$ のとき (下に有界なとき) は，元の Lyapunov 関数と一致

補題 (本研究)： $\mathcal{E}_y^{(k+1)} - \mathcal{E}_y^{(k)} \leq \frac{-k(k+1)^2}{2} \langle \nabla g(z^{(k)}), p^* \rangle \leq 0$

後半の不等号は，一般に $\langle \nabla g(x), p^* \rangle \geq 0$ より成立．前半の不等号の証明は省略．

主定理の証明に， $y = x_0$ (初期点) の場合の結果を用いる (次頁)

$$\boxed{\|p^{(k)} - p^*\| = \mathcal{O}(k^{-1})}$$

加速勾配法より $p^{(k)} = \sum_{i=1}^k \alpha_i v^{(i)}$ という形の表示を得て、各 $v^{(i)}$ に

$$\begin{aligned} \left\| v^{(k)} - x_0 + \frac{k(k+1)}{4L} p^* \right\|^2 &= \mathcal{E}_{x_0}^{(k)} - \frac{k(k+1)}{L} (g(x^{(k)}) - g(x_0)) \\ &\leq \mathcal{E}_{x_0}^{(0)} + \frac{k(k+1)}{L} \left(g(x_0) - \inf_{x \in \mathbb{R}^n} g(x) \right) \\ &= \frac{k(k+1)}{L} \left(g(x_0) - \inf_{x \in \mathbb{R}^n} g(x) \right) \end{aligned}$$

による評価を適用することで、 $\|p^{(k)} - p^*\| = \mathcal{O}(k^{-1})$ を得る.

$$p^{(k)} \in \text{dom } f^*$$

加速勾配法より $\left(\sum_{i=0}^{k-1} i(i+1)^2 \right) p^{(k)} = \sum_{i=0}^{k-1} i(i+1)^2 \nabla f(z^{(i)})$ という表示 (☆) を得る。
 $\implies p^{(k)}$ は $\{\nabla f(z^{(i)})\}_{i=0, \dots, k-1}$ の凸結合。

$$\|p^{(k)}\| - \|p^*\| = \mathcal{O}(k^{-2})$$

既に $\|p^{(k)} - p^*\| = \mathcal{O}(k^{-1})$ であるので, $p^{(k)}$ の p^* 方向成分 $\langle p^{(k)}, p^* \rangle$ が $\mathcal{O}(k^{-2})$ で収束することを示せばよい. また $p^{(k)} \in \text{dom } f^*$ より $\langle p^{(k)} - p^*, p^* \rangle \geq 0$. (☆) より

$$\begin{aligned} \left(\sum_{i=0}^{k-1} i(i+1)^2 \right) \langle p^{(k)} - p^*, p^* \rangle &= \sum_{i=0}^{k-1} i(i+1)^2 \langle \nabla g(z^{(i)}), p^* \rangle \leq 2 \left(\mathcal{E}_{x_0}^{(0)} - \mathcal{E}_{x_0}^{(k)} \right) = -2\mathcal{E}_{x_0}^{(k)} \\ &\leq -\frac{k(k+1)}{L} (g(x^{(k)}) - g(x_0)) \leq \frac{k(k+1)}{L} \left(g(x_0) - \inf_{x \in \mathbb{R}^n} g(x) \right) \end{aligned}$$

$$\implies \langle p^{(k)} - p^*, p^* \rangle \leq \mathcal{O}(k^{-2}).$$

$$\text{補題 (再掲): } \mathcal{E}_y^{(k+1)} - \mathcal{E}_y^{(k)} \leq \frac{-k(k+1)^2}{2} \langle \nabla g(z^{(k)}), p^* \rangle \leq 0$$

$f : \mathbb{R}^n \rightarrow \mathbb{R}$: L -平滑凸

加速勾配法 [Ushiyama et al., 2023]

$$\begin{cases} z^{(k)} := \frac{k}{k+2}x^{(k)} + \frac{2}{k+2}v^{(k)} \\ v^{(k+1)} := v^{(k)} - \frac{k+1}{2L}\nabla f(z^{(k)}) \quad , \quad x^{(0)} := v^{(0)} := x_0 \text{ (初期点)} \\ x^{(k+1)} := \frac{k}{k+2}x^{(k)} + \frac{2}{k+2}v^{(k+1)} \end{cases}$$

$\text{dom } f^*$ に最小ノルム点 p^* があるとき：

$$p^{(k)} := -\frac{12L}{3k+2} (x^{(k)} - x^{(k-1)}) \text{ について：}$$

$$\|p^{(k)}\| - \|p^*\| = O(k^{-2})$$

$$\|p^{(k)} - p^*\| = O(k^{-1})$$

Lyapunov 関数による証明：下に非有界な場合に適した定義・基準点 y の選び方

凸集合でのノルム最小化問題ソルバーとして：

- 我々のレート $\mathcal{O}(k^{-2})$ は, Frank–Wolfe 法 $\mathcal{O}(k^{-1})$ より高速
- 強凸集合の場合, Frank–Wolfe 法が線形収束 (e.g., [Levitin and Polyak, 1966])
- 多面体の場合, 射影勾配法により線形収束 [Necoara et al., 2019]
- 我々のレートが (現状) 最速であるような (良い) 問題があるか？
- 与えられた凸集合を $\text{dom } f^*$ とする f が構成可能か？ mirror descent との関連

- [Boyd et al., 2007] Boyd, S., Kim, S. J., Vandenberghe, L., and Hassibi, B. (2007).
A tutorial on geometric programming.
Optim. Eng., 8:67–127.
- [Hirai and Sakabe, 2024] Hirai, H. and Sakabe, K. (2024).
Gradient descent for unbounded convex functions on Hadamard manifolds and its applications to scaling problems.
In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS 2024)*, pages 2387–2402.
- [Levitin and Polyak, 1966] Levitin, E. S. and Polyak, B. T. (1966).
Constrained minimization methods.
USSR Comput. Math. Math. Phys., 6(5):1–50.
- [Necoara et al., 2019] Necoara, I., Nesterov, Y., and Glineur, F. (2019).
Linear convergence of first order methods for non-strongly convex optimization.
Math. Program., 175:69–107.
- [Nesterov, 2018] Nesterov, Y. (2018).
Lectures on Convex Optimization.
Springer Cham.
- [Rothblum and Schneider, 1989] Rothblum, U. G. and Schneider, H. (1989).
Scalings of matrices which have prespecified row sums and column sums via optimization.
Linear Algebra Appl., 114/115:737–764.
- [Sinkhorn, 1964] Sinkhorn, R. (1964).
A relationship between arbitrary positive matrices and doubly stochastic matrices.
Ann. Math. Statist., 35(2):876–879.
- [Su et al., 2016] Su, W., Boyd, S., and Candès, E. J. (2016).
A differential equation for modeling Nesterov's accelerated gradient method: theory and insights.
J. Mach. Learn. Res., 17(153):1–43.

[Ushiyama et al., 2023] Ushiyama, K., Sato, S., and Matsuo, T. (2023).

A unified discretization framework for differential equation approach with Lyapunov arguments for convex optimization.
In *Advances in Neural Information Processing Systems 36 (NeurIPS 2023)*, pages 26092–26120.

[Wilson et al., 2021] Wilson, A. C., Recht, B., and Jordan, M. I. (2021).

A Lyapunov analysis of momentum methods in optimization.
J. Mach. Learn. Res., 22(113):1–34.