P−9

# 凸最適化問題に対する一次法による近接点法の近似
## Approximation of the proximal point method via first-order methods for convex optimization problems

○伊藤勝 [1]，　福田光浩 [2]
*Masaru Ito[1]，　Mituhiro Fukuda[2]

Abstract: For convex optimization problems with smooth objective functions, we propose a new first-order method which can be regarded as an approximation of the proximal point method. This algorithm equips a practical advantage that it is independent of the parameters in the problem structure such as the Hölderian growth condition. We show the nearly optimal performance of the proposed algorithm in the sense of the iteration complexity.

### 1. 研究の背景

$C^1$ 級凸関数 $f : \mathbb{R}^n \to \mathbb{R}$ および閉凸集合 $X \subset \mathbb{R}^n$ に対して

$$f^* = \inf_{x\in X} f(x), \quad X^* = \{x \in X \mid f(x) = f^*\}$$

と定め, $X^* \neq \emptyset$ と仮定する. 凸最適化問題は, $f^*$ の値および $X^*$ の元を求める問題である. この問題に対する反復法（ここでは単にアルゴリズムとも呼ぶ）とは, 計算可能な反復的手続きによって点列 $\{x_k\} \subset X$ を生成し,

$$\lim_{k\to\infty} f(x_k) = f^*$$

を目指すものである. $\{x_0, \ldots, x_k\}$ から $x_{k+1}$ を計算する手続きは, その反復法の第 $k$ 反復という. 一次法とは, 反復法において $f$ の一次までの情報 $(f(x), \nabla f(x))$ を利用したものである. 特に, 射影勾配法とは, 反復法のうちで以下の計算可能性を仮定して, それらを各反復で（たかだか定数回だけ）利用するものである:

$$(f(x), \nabla f(x)):\ f \text{ の値およびその勾配},\quad \pi_X(x):\ \text{閉凸集合 } X \text{ への射影}.$$

射影勾配法（を近接点法との組み合わせて一般化した近接勾配法）は, この十数年の間にデータ分析や機械学習などを中心とした応用への有用性が注目された. これらの応用に現れる問題は特殊な構造をしているが, 以下の性質を満たすものが多くある（cf. [1]）.

**仮定 1** (勾配のリプシッツ連続性)**.** $f$ の勾配は $L$-リプシッツ関数である. すなわち, 定数 $L > 0$ に対して

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

ただし $\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$.

**仮定 2** (Hölderian growth)**.** 与えられた $x_0 \in X$ に対して, ある定数 $\kappa > 0$ と $\rho \ge 1$ が存在して以下が成り立つ.

$$f(x) - f^* \ge \kappa \operatorname{dist}(x, X^*)^\rho, \quad \forall x \in X \text{ with } f(x) \le f(x_0).$$

ただし $\operatorname{dist}(x, C) = \inf_{y\in C}\|y - x\|$.

我々の目的は, これらの仮定のもとで収束 $f(x_k) \to f^*$ がなるべく速い射影勾配法を見つけることである. 与えられた反復法 $M$ の収束の速度を測る尺度のひとつとして, $\varepsilon$-反復計算量 $C(M, f, \varepsilon)$（$\varepsilon > 0$）がある:

$$C(M, f, \varepsilon) = \min\{k \in \mathbb{Z}_{\ge 0} \mid f(x_k) - f^* \le \varepsilon\},\ \text{ただし } \{x_k\} \text{ は反復法 } M \text{ の生成点列}$$

このとき, 収束 $f(x_k) \to f^*$ が速いことは $\varepsilon$-反復計算量 $C(M, f, \varepsilon)$ が小さいことに対応する. したがって, 我々の目標は

$$C_*(\varepsilon) = \inf_{M:\text{射影勾配法}} \sup_{f:\text{仮定 1,2}} C(M, f, \varepsilon)$$

になるべく近い $\varepsilon$-反復計算量を達成する射影勾配法 $M$ を見つけることである.

1: 日大理工・教員・数学　　2: 東京工業大学

2. 先行研究と提案手法

先行研究として Liu と Yang の射影勾配法 $M_\rho$ がある. これは $\rho$ に依存したアルゴリズムであり, 次の評価が成り立つ（[2] による）.

**命題 3** (Liu and Yang [1]). $X = \mathbb{R}^n$ のとき, Liu と Yang の射影勾配法 $M_\rho$ は, 以下を満たす.

$$(C_*(\varepsilon) \leq) \sup_{f:\text{仮定 } 1,2} C(M_\rho, f, \varepsilon) \leq \gamma \cdot C_*(\varepsilon) \log \frac{1}{\varepsilon}, \tag{1}$$

ただし $\gamma > 0$ は $\varepsilon$ に依存しない計算可能な定数.

すなわち射影勾配法 $M_\rho$ の反復計算量の $C_*(\varepsilon)$ との差は $\log \frac{1}{\varepsilon}$ の倍数で抑えられる. このような射影勾配法は準最適であると呼ばれる. $M_\rho$ は準最適性を達成する一方で, $\rho$ に依存するという欠点を持つ. 本研究で提案する射影勾配法 $M_*$ はどのパラメータにも依存せずに反復計算量の評価 (1) を達成する. 提案手法の要は近接点法という以下のアルゴリズムである:

$$x_0 \in X, \quad x_{k+1} := p(\sigma, f, x_k) \overset{\text{def}}{=} \operatorname*{argmin}_{x \in X} \left\{ f(x) + \frac{\sigma}{2}\|x - x_k\|^2 \right\}.$$

ただし, 連続な狭義凸関数 $g$ に対する $\operatorname{argmin}_{x\in X} g(x)$ は $\{x \in X \mid g(x) = \inf g(X)\}$ の唯一の元とする. 提案手法は, Nesterov の一次法を用いて $x_{k+1} \approx p(\sigma, f, x_k)$ を近似的に計算する点および $\sigma$ を問題構造（仮定 1,2）に応じて自動的に更新する点が特徴であり, これによって $\rho$ に依存しない準最適なアルゴリズムを実現している.

**アルゴリズム 4** (提案手法 $M_*$). $x_0 \in X$, $L$, $\sigma > 0$ を任意にとる. 各 $k = 0, 1, 2, \ldots,$ に対して以下の反復を繰り返す.

1. $p(\sigma, f, x_k)$ を近似的に計算する. 具体的には, 最適化問題 $\min_{x\in X} \left\{ f(x) + \frac{\sigma}{2}\|x - x_k\|^2 \right\}$ を Nesterov の射影勾配法 [3] によって解く. このとき, 以下が得られる.
    - 近似点列 $\{x_{k,j}\}_{j\geq 0} \subset X$（$p(\sigma, f, x_k)$ に収束する）
    - パラメータ $\{L_{k,j}\}_{j\geq 0}$ ($L_{k,j} > 0$): リプシッツ定数 $L$ の推定値の列
    - ステップ幅 $\{\lambda_{k,j}\}_{j\geq 0}$ ($\lambda_{k,j} > 0$).
2. ある $j$ において $\|x_{k,j} - \pi_X(x_{k,j} - \nabla f(x_{k,j})/L_{k,j})\| \leq \|x_{k,0} - \pi_X(x_{k,0} - \nabla f(x_{k,0})/L)\|/2$ となるならば, $x_{k+1} := x_{k,j}$, $L := L_{k,j}$ として第 $(k+1)$ 反復へ.
3. そのような $j$ が存在しなければ, ある $j$ に対して $\sum_{j=0}^{k} \lambda_{k,j} \geq L_{k,j}/\sigma^2$ が必ず成り立つ. そのような $j$ に到達したら $\sigma := \sigma/2$ として, 第 $k$ 反復をやり直す（手順 1 に戻る）.

**定理 5.** $X = \mathbb{R}^n$ のとき, 提案手法 $M_*$ は以下を満たす（ただし $\gamma > 0$ は $\varepsilon$ に依存しない計算可能な定数）.

$$(C_*(\varepsilon) \leq) \sup_{f:\text{仮定 } 1,2} C(M_*, f, \varepsilon) \leq \gamma \cdot C_*(\varepsilon) \log \frac{1}{\varepsilon}, \tag{2}$$

3. 参考文献


[1] Mingrui Liu and Tianbao Yang, Adaptive accelerated gradient converging methods under Hölderian error bound condition, *Advances in Neural Information Processing Systems 30*, 2017.

[2] Arkadi Nemirovsky and Yurii Nesterov, Optimal methods for smooth convex minimization, *Zh. Vychishl. Mat. i Mat. Fiz.*, **25**, pp. 356–369, 1985 (in Russian); English translation: *USSR Computational Mathematics and Mathematical Physics*, **24**, pp. 80–82, 1984.

[3] Yurii Nesterov, Gradient methods for minimizing composite functions, *Mathematical Programming*, **140**, pp. 125–161, 2013.