P-16

問題構造に適応した勾配法の構築について
On the construction of adaptive first-order methods

○伊藤勝 [1]
*Masaru Ito[1]

Abstract: For an optimization problem with a smooth convex objective function, we discuss a restart technique of first-order methods. Under the assumption of an error-bound inequality, we show an upper bound for of the complexity of the restart scheme. It turns out that this upper bound is tight up to a logarithmic factor, in view of the lower bound of the complexity.

1. 凸最適化問題

凸最適化問題とは, 与えられた C^1 級の凸関数 $f:\mathbb{R}^n\to\mathbb{R}$ と閉凸集合 $X\subset\mathbb{R}^n$ に対して, 下限値

$$f^* := \inf_{x\in X} f(x)$$

と最小点 $x\in X^* := \{x^*\in X \mid f(x^*)=\inf_{z\in X} f(z)\}$ を求める問題であり, f を目的関数, f^* を最適値, X^* の元を最適解と呼ぶ. ここでは最適解が少なくともひとつ存在すると仮定する（すなわち $X^*\neq\emptyset$）. この凸最適化問題を厳密に解くことが難しい状況を考え, 代わりにこの問題を近似的に解くためのアルゴリズムの構築を試みる. 本研究では, 勾配法というアルゴリズムに着目する.

勾配法は, 目的関数 f に対して関数値 $f(x)$ と勾配 $\nabla f(x)$ が各点 $x\in X$ で計算可能であると仮定し, これらの情報を使って f^* の近似を試みる. さらに X への直交射影 $\pi_X(x)$, すなわち $\|z-x\|$ を最小にする $z\in X$ が計算可能であると仮定したとき, これを用いた勾配法を射影勾配法 (projected gradient method) という. この種のアルゴリズムは, 大規模な凸最適化問題に対して有用な手段として活発に研究されている.

次のアルゴリズム（M_{pg} と呼ぶことにする）は最もよく知られた射影勾配法である:

$$x_0\in X,\quad x_{k+1} := \pi_X(x_k-\lambda_k\nabla f(x_k)),\quad k=0,1,2,\ldots.$$

ここで $\lambda_k>0$ はステップ幅と呼ばれ, x_k がなるべく「最適に近く」なるようにこのパラメータを調整する. 本研究では, この最適性の指標として $\|g_\lambda(x)\|$ を用いる. ただし $\|\cdot\|$ はユークリッドノルムを表し,

$$T_\lambda(x)=\pi_X(x-\lambda\nabla f(x)),\quad g_\lambda(x)=\frac{x-T_\lambda(x)}{\lambda}.$$

任意の $\lambda>0$ に対して, $x\in X^* \iff \|g_\lambda(x)\|=0$ が成り立つ. そこで我々は $\|g_{\lambda_k}(x_k)\|$ がなるべく速く 0 に収束するように $\{x_k\}$（および $\{\lambda_k\}$）を構築することを目指す.

ここで「収束の速さ」を計るために, 反復計算量を定義する. 与えられたアルゴリズム M を上記の問題に当てはめて点列 $\{x_k\}$ を生成するとき, このアルゴリズムの ε-反復計算量とは, $\|g_{\lambda_k}(x_k)\|\le\varepsilon$ となる最小の k のことを言い, これを $C(M,f,\varepsilon)$ と表す. すなわち,

$$C(M,f,\varepsilon)=\min\{k\in\mathbb{Z}_{\ge 0}\mid \|g_{\lambda_k}(x_k)\|\le\varepsilon\}.$$

$C(M,f,\varepsilon)$ がなるべく小さくなるようなアルゴリズム M を見つけることが望ましい.

2. 反復計算量の解析

本研究では以下の仮定のもとで, 反復計算量の評価を行う.

仮定 1. $\inf_{x\in X} f(x)$ を求める最適化問題に対して, 以下を仮定する.

1: 日大理工・教員・数学

(A) ∇f は定数 $L > 0$ に対して X 上でリプシッツ連続である. すなわち,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \quad \forall x, y \in X.$$

(B) ある定数 $\kappa > 0$ と $\rho > 0$ が存在して

$$\mathrm{dist}(T_{1/L}(x), X^*) \leq \kappa \|g_{1/L}(x)\|^\rho, \quad \forall x \in X_0.$$

ただし $X_0 = \{x \in X \mid f(x) \leq f(x_0),\ \|g_{1/L}(x)\| \leq \|g_{1/L}(x_0)\|,\ T_{1/L}(x) \notin X^*\}$ であり, $\mathrm{dist}(x, S) := \inf_{y \in S} \|y - x\|$ とする.

仮定 (A) は応用にもよく現れ, 勾配法の解析において標準的に用いられる. 仮定 (B) は先行研究 [1, 2] で扱われたいくつかの「エラーバウンド」と呼ばれる条件の一般化になっている. 広い応用問題で (B) が成り立ちかつ (A) だけを仮定するよりも収束が速いアルゴリズムを構築できる. 例えば, f が C^2 級で, 任意の $x \in X$ に対して $\nabla^2 f(x)$ の最小固有値が $\lambda (> 0)$ 以上であるとき, $\kappa = 1/\lambda$, $\rho = 1$ として仮定 (B) が成り立つ. 仮定 (B) に現れるパラメータ κ と ρ はアルゴリズムの構築に有用であるが, これらがわかっていない場合にも, 仮定 (B) に自動的に適応した収束の速度を発揮するアルゴリズムを開発することは応用上重要である. まずは射影勾配法 M_{pg} はこの性質を持つことを以下に示す.

命題 2. ステップ幅 $\lambda_k \equiv 1/L$ を用いた射影勾配法 M_{pg} は, 仮定 (A) のもとで $C(M_{\mathrm{pg}}, f, \varepsilon) = O(L\,\mathrm{dist}(x_0, X^*)\varepsilon^{-1})$ を満たす. ここで $O(\cdot)$ はランダウの記号を表す. さらに (B) を仮定すると, 以下の反復計算量の上界を得る:

$$C(M_{\mathrm{pg}}, f, \varepsilon) = \begin{cases} O(L\kappa\varepsilon^{-(1-\rho)}) & (\rho < 1), \\ O(L\kappa \log \frac{\|g_{1/L}(x_0)\|}{\varepsilon}) & (\rho = 1). \end{cases}$$

また, $\rho > 1$ のとき, $\|g_{1/L}(x_k)\|$ は 0 に超一次収束する.

本研究では, 射影勾配法 M_{pg} よりも小さな反復計算量の上界を保証するアルゴリズム M_* を提案する. アルゴリズムは [1] のものと同様の手法を用いる.

アルゴリズム 3 (提案手法 M_*). $x_0 \in X$, L, $\sigma > 0$ を取る. 各 $k = 0, 1, 2, \ldots,$ に対して以下の反復を繰り返す.

1. 補助最適化問題 $f(x) + \frac{\sigma}{2}\|x - x_k\|^2$ を考え, これを Nesterov の射影勾配法 [3] によって解く. 点列 $\{x_{k,j}\}_{j \in \mathbb{N}}$ とステップ幅 $\{\lambda_{k,j}\}_{j \in \mathbb{N}}$ が得られる. ある j に対して次のステップ 2 またはステップ 3 が必ず起こる.
2. ある j に対して $\|g_{1/L}(x_{k,j})\| \leq \|g_{1/L}(x_{k,0})\|/2$ となったら $x_{k+1} = x_{k,j}$ として第 $(k+1)$ 反復へ.
3. ある j に対して $\sum_{j=0}^{k} \lambda_{k,j} \geq L/\sigma^2$ となったら $\sigma := \sigma/2$ として, 第 k 反復をやり直す（手順 1 に戻る）.

提案手法 M_* は仮定 (B) のパラメータ κ や ρ に依存せず, 以下に示すように射影勾配法 M_{pg} よりも小さな反復計算量の上界を保証する.

定理 4 (主結果). $\{x_k\}$ を提案手法 M_* によって生成したとする. 仮定 (A), (B) のもとで, 以下の反復計算量の上界が成り立つ.

$$C(M_{\mathrm{pg}}, f, \varepsilon) = \begin{cases} O(\sqrt{L\kappa\varepsilon^{-(1-\rho)}} \log(L\varepsilon^{-(1-\rho)})) & (\rho < 1), \\ O(\sqrt{L\kappa} \log(L\kappa) \log \frac{\|g_{1/L}(x_0)\|}{\varepsilon}) & (\rho = 1). \end{cases}$$

また, $\rho > 1$ のとき, $\|g_{1/L}(x_k)\|$ は 0 に超一次収束する.

仮定 (A) および (B) のもとでは, 射影勾配法の反復計算量の下界が知られている [1]. 定理 4 の上界はこの下界と比べて, 対数倍の違いで一致する.

3. 参考文献

[1] Masaru Ito and Mituhiro Fukuda, Nearly optimal first-order methods for convex optimization under gradient norm measure: An adaptive regularization approach, *arXiv preprint*, arXiv:1912.12004v2, 2020.

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

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