P-8

Frank-Wolfe 法におけるヘルダー条件のもとでのステップ幅選択

On the step size rules for Frank-Wolfe method under Hölder condition

○伊藤勝 [1]
*Masaru Ito[1]

Abstract: We propose a new step size rule for Frank-Wolfe method solving optimization problems over compact convex sets. We prove its convergence rate under the Hölder continuity for the gradient of the objective function. Although our method does not require the parameters in the Hölder condition, the method provides an adaptive performance to these parameters.

1. 最適化問題

与えられた C^1 級関数 $f:\mathbb{R}^n\to\mathbb{R}$ と閉集合 $X\subset\mathbb{R}^n$ に対して, 下限値

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

と最小点 $x\in X^*:=\{x^*\in X\mid f(x^*)=\inf_{z\in X}f(z)\}$ を求める問題を最適化問題と呼び, 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 $X^*\neq\emptyset$）. この最適化問題を最適解を厳密に求めること一般に困難であり, 代わりにこの問題の停留点を, アルゴリズムを用いて近似的に計算することは一般的な. ここで $x^*\in X$ がこの問題の停留点であるとは, ある内積 $\langle\cdot,\cdot\rangle$ に対して次が成り立つことである.

$$-\nabla f(x^*)\in N_X(x^*)=\{y\in\mathbb{R}^n:\langle y,x-x^*\rangle\le 0,\ \forall x\in X\}.$$

$x^*\in X$ が停留点であることは, x^* が最適解であるための必要条件である.

本研究では, 以下の仮定のもとでのアルゴリズムの構築を考える.

仮定 1. (A1) X は有界閉凸集合であり, 任意の線形関数 $p:\mathbb{R}^n\to\mathbb{R}$ の X 上での最小点が計算可能である.
(A2) ∇f はヘルダー条件を満たす. すなわち, ある $\nu\in(0,1]$ と $M_\nu\ge 0$ に対して

$$\|\nabla f(x)-\nabla f(y)\|\le M_\nu\|x-y\|^\nu,\quad \forall x,y\in X.$$

ただし $\|\cdot\|$ は内積 $\langle\cdot,\cdot\rangle$ が誘導するノルムである.

Frank-Wolfe 法 [1] は以下に記述される反復に従って点列 $\{x_k\}\subset X$ を生成し, $\{x_k\}$ が停留点に収束することを目指す.

アルゴリズム 2 (Frank-Wolfe 法)**.** $x_0\in X$ を任意に選ぶ.
各 $t=0,1,2,\ldots,$ に対して以下の手順 1 から 4 を反復する:
1. $v_t\in\mathrm{Argmin}_{x\in X}\langle\nabla f(x_t),x\rangle:=\{v\in X:\langle\nabla f(x_t),v\rangle=\min_{x\in X}\langle\nabla f(x_t),x\rangle\}$ により v_t を計算.
2. $\delta_t=\langle\nabla f(x_t),x_t-v_t\rangle$ が十分小さければ終了.
3. ステップ幅 $\tau_t\in[0,1]$ を選ぶ.
4. $x_{t+1}=(1-\tau_t)x_t+\tau_t v_t$ により x_{t+1} を計算.

Frank-Wolfe 法は, もとの最適化問題 $\min_{x\in X}f(x)$ の代わりに, それを x_t において一次近似した $\min_{x\in X}\langle\nabla f(x_t),x\rangle$ を繰り返し解くことによって最適解の近似を試みる手法とみることができる. 単純な近似を用いるため, 各反復が容易に実行できる場面では有用であり, 特にスパース推定や低ランク行列補完といった大規模な最適化問題への応用がある.

Frank-Wolfe 法の終了条件には δ_t という量が判定に使われており, δ_t は Frank-Wolfe ギャップと呼ばれる. 次の補題は, Frank-Wolfe ギャップの性質を述べている.

1: 日大理工・教員・数学

補題 3. Frank-Wolfe 法において $\delta_t \geq 0$ が成り立ち, $\delta_t = 0$ であることと x_t がもとの最適化問題の停留点であることは同値である.

Frank-Wolfe 法ではステップ幅 τ_t の選び方によって, アルゴリズムの効率が大きく変わるため, τ_t の選択方法について様々な研究がある. 先行研究におけるステップ幅 τ_t の選び方のいくつかを以下に挙げる.

- $\tau_t \in \mathrm{Argmin}_{\tau\in[0,1]} f((1-\tau)x_t + \tau v_t)$ によりステップ幅を選ぶ手法を正確な直線探索と呼ぶ. 目的関数 f が複雑な場合, τ_t の計算は困難であるが, 理論解析ではよく用いられる.
- $\tau_t = \min\left\{1, \left(\frac{\delta_t}{M_\nu\|x_t - v_t\|^{1+\nu}}\right)^{\frac{1}{\nu}}\right\}$ によりスッテプ幅を選ぶ手法が [1] や [2] で提案された. M_ν と ν がわかっている場合に有効である.

本研究では M_ν や ν が不明である場合やゆるい見積もりしか得られていない場合に, 上記のステップ幅の選び方と同等の理論保証を実現するステップ幅選択規則を提案する.

アルゴリズム 4 (提案するステップ幅選択規則)**.** Frank-Wolfe 法の手順 3 において, ステップ幅を以下のように選択する (ただし $L_{-1} > 0$ は任意): 整数 $i = 0, 1, 2, \ldots$ に対して $L_t^{(i)} = 2^{i-1}L_{t-1}$, $\tau_t^{(i)} = \min\left\{1, \frac{\delta_t}{2L_t^{(i)}\|x_t - v_t\|^2}\right\}$ とおくとき,

$$f((1-\tau_t^{(i)})x_t + \tau_t^{(i)}v_t) \leq f(x_t) - \frac{1}{2}\tau_t^{(i)}\delta_t + \frac{1}{2}L_t^{(i)}(\tau_t^{(i)})^2\|x_t - v_t\|^2$$

が成り立つ最小の i に対して, ステップ幅を $\tau_t := \tau_t^{(i)}$ と選ぶ.

定理 5 (主結果)**.** 提案手法 (アルゴリズム 4) を用いた Frank-Wolfe 法について次の主張が成り立つ.

(i) $\{f(x_t)\}$ は単調減少し $f_* = \lim_{t\to\infty} f(x_t)$ が存在する. また, 以下が成り立つ.

$$\min\{\delta_0, \ldots, \delta_t\} \leq 4\max\left\{\frac{f(x_{\tilde{t}_0}) - f_*}{t+1-\tilde{t}_0}, \left(\frac{A(f(x_{\tilde{t}_0}) - f_*)}{t+1-\tilde{t}_0}\right)^{\frac{\nu}{1+\nu}}\right\}, \quad \forall t \geq \tilde{t}_0. \tag{1}$$

ただし $A = 2^{\frac{1+\nu}{2\nu}} M_\nu^{\frac{1}{\nu}} \left(\frac{1-\nu}{1+\nu}\right)^{\frac{1-\nu}{2\nu}} D_X^{\frac{1+\nu}{\nu}}$, $D_X = \sup_{x,y\in X}\|x-y\|^{\frac{1+\nu}{\nu}}$ であり $\tilde{t}_0$ は $\{L_{-1}, x_0, v_0, \nu, M_\nu\}$ により定まる整数.

(ii) さらに f が凸関数であると仮定する. $\overline{\gamma}_t := [\frac{4\nu A}{4\nu+t-t_0}]^\nu$ に対して

$$f(x_t) - f^* \leq \overline{\gamma}_{t-\tilde{t}_0}\ (\forall t \geq \tilde{t}_0 + t_0), \quad \min\{\delta_0, \ldots, \delta_t\} \leq e^{\frac{1}{e}}\overline{\gamma}_{\lfloor(t+1-\tilde{t}_0)/2\rfloor}\ (\forall t \geq \tilde{t}_0 + 2t_0)$$

が成り立つ. ただし t_0 は $\{L_{-1}, x_0, v_0, \nu, M_\nu, D_X\}$ により定まる整数.

以下に提案手法の利点を挙げる.

- 上記の既存手法と比べて, ステップ幅 τ_t の計算が容易である.
- より良い ν および M_ν に適応した収束速度を発揮することが期待される.
- 上記の既存手法においても定理 5 と同等の収束速度が保証される. すなわち, 提案手法はこれらと同等の理論保証をもちつつ, 効率性を兼ね備えている.

2. 参考文献

[1] M. Frank and P. Wolfe, An algorithm for quadratic programming, Nav. Res. Logist. Q., **3**, pp. 95–110, 1956.

[2] R. Zhao and R. M. Freund, Analysis of the Frank-Wolfe method for logarithmically-homogeneous barriers, with an extension, ArXiv preprint, arXiv:2010.08999v1, 2020.