日本オペレーションズ・リサーチ学会
2021年 春季研究発表会

Frank-Wolfe 法における適応的なステップ幅の選択

05000192 日本大学 *伊藤勝 ITO Masaru
University of Minnesota LU Zhaosong
University of Minnesota HE Chuan

1. はじめに

本研究では以下の最適化問題を考える.

$$\varphi^* = \min_{x\in\mathbb{E}} \varphi(x), \quad \varphi(x) := f(x) + g(x) \tag{1}$$

ここで $\mathbb{E}$ は $\langle\cdot,\cdot\rangle : \mathbb{E}^2 \to \mathbb{R}$ を内積にもつ実ヒルベルト空間とし $f, g : \mathbb{E} \to \mathbb{R}\cup\{+\infty\}$ は下半連続な関数とする. また, g は凸関数で $\operatorname{dom} g = \{x \in \mathbb{E} \mid g(x) < +\infty\}$ は有界であるとする. 関数 g は凸集合に対する標示関数であったり, 正則化項としての役割を持つ.

最適化問題 (1) に対する Frank-Wolfe 法は, 以下の更新により点列 $\{x_t\}_{t\geq 0}$ を構成する.

$$\begin{aligned} &v_t \in \operatorname{Argmin}_{x\in\mathbb{E}}\{\langle\nabla f(x_t), x\rangle + g(x)\}, \\ &x_{t+1} = (1-\tau_t)x_t + \tau_t v_t. \end{aligned} \tag{2}$$

ただし $\tau_t \in [0,1]$ はステップ幅と呼ばれるパラメータである. Frank-Wolfe 法は各反復で ∇f の計算と $\min_{x\in\mathbb{E}}\{\langle c, x\rangle + g(x)\}$ という形の補助問題を解く必要がある. 問題 (1) に対する他のよく知られたアルゴリズムのひとつである近接勾配法と比較すると, 補助問題の計算量を小さく済ますことができ, 統計や機械学習などに現れる大規模な最適化問題に対する一次法の有用な選択肢として活発に研究されている.

Frank-Wolfe 法において以下の値を定義する.

$$\delta_t := \langle\nabla f(x_t), x_t - v_t\rangle + g(x_t) - g(v_t) \geq 0. \tag{3}$$

条件 $\delta_t = 0$ は x_t の局所最適性の必要条件になっている. さらに f が凸であれば $\delta_t \geq \varphi(x_t) - \varphi^*$ が成り立つ. そこで条件 $\delta_t \leq \varepsilon$ をアルゴリズムの停止条件として用いることにする.

ステップ幅 τ_t の選択はアルゴリズムの収束率に大きく影響する. 本研究では関数 f と g に関して以下の仮定のもとで有用なステップ幅の選択を考察する.

仮定 1 (ヘルダー条件)**.** f は $\operatorname{dom} g$ 上連続的微分可能で, ある $\nu \in (0,1]$, $L_\nu > 0$ および $\mathbb{E}$ 上のノルム $\|\cdot\|$ に対して

$$\|\nabla f(x) - \nabla f(y)\|_* \leq L_\nu \|x-y\|^\nu, \forall x, y \in \operatorname{dom} g.$$

ここで $\|\cdot\|_*$ は $\|\cdot\|$ の双対ノルムである.

$\nu = 1$ の場合, これは ∇f のリプシッツ条件でありよく調べられている.

仮定 2. ある $\kappa \geq 0$ と $\rho \geq 2$ が存在して, 任意の $x, v \in \mathbb{E}$ と $v^* \in \operatorname{Argmin}_w\{\langle\nabla f(x), w\rangle + g(w)\}$ に対して

$$\begin{aligned} &\langle\nabla f(x), v\rangle + g(v) - \langle\nabla f(x), v^*\rangle - g(v^*) \\ &\geq \frac{\kappa}{\rho}\|v - v^*\|^\rho \end{aligned}$$

この仮定は g が一様凸 [1] な集合の標示関数であったり, 一様凸な関数である場合に成り立つ. $\kappa = 0$ のとき仮定 2 は自明に成り立つ.

上記の仮定にはパラメータ ν, L_ν, κ, ρ が現れるが, これらのパラメータに自動に適応して高速化された収束率を保証するステップ幅の選択規則を考察する.

2. ステップ幅の選択規則

ステップ幅の選択規則はこれまで様々な提案がなされてきた. 既存の選択規則をいくつか挙げよう.

- 正確な直線探索による選択規則

$$\tau_t \in \operatorname*{Argmin}_{\tau\in[0,1]} \varphi((1-\tau)x_t + \tau v_t). \tag{4}$$

後に示すように, 上記の仮定のもとで理論的に有意義な結果が得られる.

- 目的関数の近似に対する直線探索 [3]

$$\tau_t = \min\left\{1, \left(\frac{\delta_t}{L_\nu\|x_t - v_t\|^{1+\nu}}\right)^{\frac{1}{\nu}}\right\}. \tag{5}$$

(4) と同じ理論保証を維持しつつ計算コストが小さいがパラメータ ν および L_ν が必要である.

- 選択規則 $\tau_t = \frac{2}{t+2}$ はパラメータに依存しない単純なものである. 仮定 1 と f の凸性のもとでは正確な直線探索と同様の理論保証を持つ [2].

定理 1. ステップ幅 (4) または (5) による Frank-Wolfe 法について, $\delta_t \le \varepsilon$ となるための反復回数 T_ε は以下で抑えられる.

$$T_\varepsilon \le O(1)\left(\frac{L_\nu D_g^{1+\nu}}{\varepsilon}\right)^{\frac{1}{\nu}} \frac{\Delta_0}{\varepsilon}.$$

である. ただし $\Delta_0 = \varphi(x_0) - \varphi^*$, $D_g = \operatorname{diam}(\operatorname{dom} g)$ とし $O(1)$ はパラメータに依存しない定数である. さらに $\kappa > 0$ であれば以下が成り立つ.

$$T_\varepsilon \le O(1)\left(\frac{\rho^{1+\nu} L_\nu^\rho}{\kappa^{1+\nu}\varepsilon^{\rho-1-\nu}}\right)^{\frac{1}{\nu\rho}} \frac{\Delta_0}{\varepsilon}.$$

この定理は, 提案手法が仮定 1 と 2 に適応した収束率が得られることを示している. 以下に示すように, f に凸性をさらに仮定すると, 収束率が改善される.

定理 2 (凸の場合)**.** ステップ幅 (4) または (5) による Frank-Wolfe 法について, f が凸関数であると仮定する. このとき $\delta_t \le \varepsilon$ となるための反復回数 T_ε は以下で抑えられる.

$$T_\varepsilon \le O(1)\left(\frac{L_\nu D_g^{1+\nu}}{\varepsilon}\right)^{\frac{1}{\nu}}.$$

さらに $\kappa > 0$ であれば以下が成り立つ.

$$T_\varepsilon \le \begin{cases} O(1)\dfrac{L_1}{\kappa}\log\dfrac{\Delta_0}{\varepsilon} & (\rho = \nu + 1 = 2), \\ O(1)\left(\dfrac{\rho^{1+\nu} L_\nu^\rho}{\kappa^{1+\nu}\varepsilon^{\rho-1-\nu}}\right)^{\frac{1}{\nu\rho}} & (\text{otherwise}). \end{cases}$$

上記ふたつの定理において, $\kappa = 0$ のときの定理 2 の主張は, 選択規則 $\tau_t = \frac{2}{t+2}$ に対しても成り立つ.

3. 提案手法

本研究では以下のステップ幅の選択規則による Frank-Wolfe 法を提案する.

アルゴリズム 1.

Choose $x_0 \in \operatorname{dom} g$ and $L_{-1} > 0$.

For $t = 0, 1, 2, \ldots,$ do:

1. $v_t \in \operatorname{Argmin}_{x\in\mathbb{E}}\{\langle \nabla f(x_t), x\rangle + g(x)\}$
2. Compute δ_t by (3)
3. [Backtracking] Repeat for $i = 0, 1, 2, \ldots,$
   $L_t^{(i)} := 2^{i-1} L_{t-1}$
   $\tau_t^{(i)} := \min\left\{1, \frac{\delta_t}{2L_t^{(i)}\|x_t - v_t\|^2}\right\}$
   $x_{t+1}^{(i)} := (1 - \tau_t^{(i)})x_t + \tau_t^{(i)} v_t$
   Until $\varphi(x_{t+1}^{(i)}) \le \varphi(x_t) - \frac{1}{2}\tau_t^{(i)}\delta_t + \frac{1}{2}L_t^{(i)}(\tau_t^{(i)})^2\|x_t - v_t\|^2$ $\cdots\cdots$ $(*)$
4. Set $(x_{t+1}, L_t, \tau_t) \leftarrow (x_{t+1}^{(i)}, L_t^{(i)}, \tau_t^{(i)})$.

提案手法の利点は, パラメータに依存しないで実行でき, 正確な直線探索と同様の理論保証が成り立つ点である. 仮定 1 のもとで以下の不等式が成り立つことに着目すると, 条件 $(*)$ は有限回の内部反復の後に満たされる.

$$f(y) \le f(x) + \langle \nabla f(x), y - x\rangle + \frac{L(\varepsilon)}{2}\|y - x\|^2 + \varepsilon,$$

$$\text{where } L(\varepsilon) = \left(\frac{1-\nu}{1+\nu}\cdot\frac{1}{2\varepsilon}\right)^{\frac{1-\nu}{1+\nu}} L_\nu^{\frac{2}{1+\nu}}.$$

定理 3 (主結果)**.** アルゴリズム 1 において, $\delta_t \le \varepsilon$ となる反復回数を T_ε とおく. T_ε の上界について定理 1, 2 と同じ主張がそれぞれ成り立つ.

参考文献

[1] Thomas Kerdreux, Alexandre d'Aspremont, and Sebastian Pokutta, Projection-free optimization on uniformly convex sets, ArXiv preprint, arXiv:2004.11053v2, 2020.

[2] Yu. Nesterov, Complexity bounds for primal-dual methods minimizing the model of objective function, Math. Program., **171**:311–330, 2018.

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