日本オペレーションズ・リサーチ学会
2023 年春季研究発表会

目的関数が未知の凸最適化問題に対する局所線形回帰を用いた非厳密 Frank–Wolfe 法

	東京工業大学	*山根 大輝	YAMANE Daiki
02303360	統計数理研究所・理化学研究所	田中 未来	TANAKA Mirai
05000410	東京工業大学	小林 健	KOBAYASHI Ken
01405430	東京工業大学	中田 和秀	NAKATA Kazuhide

1. はじめに

目的関数 $f\colon \mathbb{R}^n \to \mathbb{R}$ を C^1 級の凸関数, $\Omega \subseteq \mathbb{R}^n$ をコンパクトな凸集合として, 以下の凸最適化問題を解くことを考える:

$$\min_{\boldsymbol{x}\in\Omega} f(\boldsymbol{x}). \tag{1}$$

最適化の研究では, 任意の $\boldsymbol{x} \in \Omega$ において目的関数値 $f(\boldsymbol{x})$ や目的関数の勾配 $\nabla f(\boldsymbol{x})$ の計算が容易であることを仮定することが多いが, 現実ではこれらの計算が困難な場合がある.

このように目的関数値とその勾配の計算が困難な状況下で問題 (1) を解く素朴な手法としては, 変数 $\boldsymbol{x}$ と対応する目的関数値 $f(\boldsymbol{x})$ に関する標本集合から目的関数 f を推定し, 推定後の関数を Ω 上で最適化する手法が考えられる. しかし目的関数の推定に非線形なモデルを用いると, 得られる最適化問題は一般に非凸になり解きづらい.

そこで本研究では, 目的関数の勾配を現在の反復点のまわりでの局所線形回帰で得られる回帰係数で代替して, 近似的に Frank–Wolfe (FW) 法 [2] を実行する方法を提案し, その理論的性質を考察する.

以降では, $\|\cdot\|\colon \mathbb{R}^n \to \mathbb{R}$, $\langle\cdot,\cdot\rangle\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ をそれぞれ $\mathbb{R}^n$ 上のノルム, 内積とする.

2. 設定

本発表では以下の設定のもとで問題 (1) を解くことを考える:

1. $\boldsymbol{x}_i \in \Omega$ を変数に関する i 番目の標本, $y_i \in \mathbb{R}$ を変数 $\boldsymbol{x}_i$ に対応する誤差を含む目的関数値

$$y_i = f(\boldsymbol{x}_i) + \varepsilon_i \quad (i = 1, 2, \ldots, d)$$

として, 標本集合 $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^d$ が所与とする. ここで ε_i は i 番目の標本における目的関数値の観測誤差を表す.

2. 最適化計算の過程で目的関数値 $f(\boldsymbol{x})$ および目的関数の勾配 $\nabla f(\boldsymbol{x})$ の計算はできない.
3. 実行可能領域 Ω は扱いやすい構造をもつ. 具体的には, Ω 上で線形関数を最小化する最適化問題は容易に解ける.

3. 関連研究

3.1. ブラックボックス最適化

目的関数 f が未知の最適化問題を解く手法の多くはブラックボックス最適化とよばれる. ブラックボックス最適化の研究では, 目的関数 f を代理するモデルを推定して解を探索する手法や, 目的関数値 $f(\boldsymbol{x})$ の計算のみを用いて解を探索する手法など, 状況設定に応じて様々な手法が提案されている [1]. しかし, 既存の手法の多くは最適化計算の過程で目的関数値 $f(\boldsymbol{x})$ の計算が可能であることを仮定しており, 2 節で述べた状況下ではこれらの手法を直接適用することはできない.

3.2. Frank–Wolfe 法

目的関数 f が既知の場合における凸最適化問題 (1) のひとつの解法として, FW 法 [2] がある. FW 法では目的関数を線形近似した関数の最小化を用いて解を更新する. 基本的な FW 法の解析では, 目的関数値 $f(\boldsymbol{x})$ とその勾配 $\nabla f(\boldsymbol{x})$ が厳密に計算可能であることを仮定する. Jaggi [3] は勾配を厳密に計算しなくても, 近似を徐々によくすれば収束性には影響を及ぼさないことを示している.

4. 局所線形回帰を用いた非厳密 Frank–Wolfe 法

2 節で述べた状況下で問題 (1) を解く提案手法を述べる. この手法は, 現在の反復点の近傍に含まれる標本を用いた線形回帰で目的関数 f を線形近似する関数を構成して, その近似関数の最小化を繰り返して解を更新する. 提案手法は目的関数の

線形近似を勾配ではなく線形回帰で行う FW 法の変種とみなせる.

提案手法の k 反復目 $(k \geq 0)$ では, 現在の反復点 $\boldsymbol{x}^k$ における局所線形回帰を用いて f の近似勾配を求める. 具体的には, $\boldsymbol{x}^k$ に距離が近い m 個の標本 $(m \geq n)$ に含まれる標本の添字集合を $\mathcal{I}(\boldsymbol{x}^k) \subseteq \{1, 2, \ldots, d\}$ として, 以下の最小 2 乗推定問題の最適解 $(\boldsymbol{\theta}^k, \theta_0^k) \in \mathbb{R}^n \times \mathbb{R}$ を求める:

$$(\boldsymbol{\theta}^k, \theta_0^k) \leftarrow \operatorname*{argmin}_{(\boldsymbol{\theta}, \theta_0) \in \mathbb{R}^n \times \mathbb{R}} \sum_{i \in \mathcal{I}(\boldsymbol{x}^k)} (y_i - \langle \boldsymbol{\theta}, \boldsymbol{x}_i \rangle - \theta_0)^2. \tag{2}$$

ここで関数 $\langle \boldsymbol{\theta}^k, \boldsymbol{x} \rangle + \theta_0^k$ は目的関数 f を $\boldsymbol{x}^k$ のまわりで線形近似する関数であり, 問題 (2) の最適解 $\boldsymbol{\theta}^k$ は勾配 $\nabla f(\boldsymbol{x}^k)$ の近似とみなせる.

提案手法では, 続いて近似関数を Ω 上で最小化する問題を解く:

$$\boldsymbol{s}^k \leftarrow \operatorname*{argmin}_{\boldsymbol{s} \in \Omega} \langle \boldsymbol{\theta}^k, \boldsymbol{s} \rangle.$$

そして $\alpha^k \in [0, 1]$ を適当なステップサイズとして,

$$\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k + \alpha^k (\boldsymbol{s}^k - \boldsymbol{x}^k)$$

と更新する. 以上を適当な終了条件を満たすまで繰り返す. 提案手法を Algorithm 1 にまとめる.

Algorithm 1 局所線形回帰を用いた非厳密 FW 法

Require: 標本集合 $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^d$, 初期点 $\boldsymbol{x}^0 \in \Omega$.
  $k \leftarrow 0$.
  for $k = 0, 1, \ldots$ **do**
    $(\boldsymbol{\theta}^k, \theta_0^k) \leftarrow \operatorname{argmin}_{(\boldsymbol{\theta}, \theta_0) \in \mathbb{R}^n \times \mathbb{R}} \sum_{i \in \mathcal{I}(\boldsymbol{x}^k)} (y_i - \langle \boldsymbol{\theta}, \boldsymbol{x}_i \rangle - \theta_0)^2$.
    $\boldsymbol{s}^k \leftarrow \operatorname{argmin}_{\boldsymbol{s} \in \Omega} \langle \boldsymbol{\theta}^k, \boldsymbol{s} \rangle$,
    $\boldsymbol{x}^{k+1} \leftarrow \boldsymbol{x}^k + \alpha^k (\boldsymbol{s}^k - \boldsymbol{x}^k)$.
  end for

5. 提案手法の理論的性質

提案手法の理論的性質について述べる. まず, 目的関数 f, 観測誤差 $\{\varepsilon_i\}_{i=1}^d$ それぞれに仮定をおく:

1. f は Ω 上で L-平滑 $(L \geq 0)$;
2. $\{\varepsilon_i\}_{i=1}^d$ は互いに独立して平均 0, 分散 σ^2 の正規分布にしたがう.

また, 問題 (1) の最適解を $\boldsymbol{x}^* \in \operatorname{argmin}_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x})$, 実行可能領域 Ω の直径を $D \coloneqq \max_{\boldsymbol{x}_1, \boldsymbol{x}_2 \in \Omega} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|$, k 反復目 $(k \geq 0)$ における勾配の近似誤差を $\delta^k \coloneqq \|\boldsymbol{\theta}^k - \nabla f(\boldsymbol{x}^k)\|$ $(k \geq 0)$ と表す. さらに, Algorithm 1 のステップサイズを $\alpha^k = 2/(k+2)$ $(k \geq 0)$ とする. このとき, 各反復における目的関数値と最適値との差は次のように抑えることができる:

命題 1. $k \geq 1$ で以下が成り立つ:

$$f(\boldsymbol{x}^k) - f(\boldsymbol{x}^*) \leq \frac{2LD^2}{k+2} + \frac{2D \sum_{j=0}^{k-1} \delta^j}{k+1}.$$

次に, 各反復における勾配の近似誤差 δ^k を評価する. $k \geq 0$ に対して $\boldsymbol{x}^k$ の m 最近傍 $\{\boldsymbol{x}_i\}_{i \in \mathcal{I}(\boldsymbol{x}^k)}$ の平均を $\boldsymbol{\mu}^k \in \mathbb{R}^n$ とし, 行ベクトル $(\boldsymbol{x}_i - \boldsymbol{\mu}^k)^\top$ $(i \in \mathcal{I}(\boldsymbol{x}^k))$ を縦に並べた行列 $\bar{\boldsymbol{X}}^k \in \mathbb{R}^{m \times n}$ の最小特異値を $\rho_{\min}^k$ とする. $\{\boldsymbol{x}_i\}_{i \in \mathcal{I}(\boldsymbol{x}^k)}$ の直径を D^k とする. このとき, 勾配の近似誤差 δ^k の期待値を次のように抑えることができる:

命題 2. $\rho_{\min}^k > 0$ のとき, 以下が成り立つ:

$$\mathbb{E}[\delta^k] \leq \frac{1}{\rho_{\min}^k} \sqrt{n\sigma^2 + \frac{mL^2(D^k)^4}{4}}.$$

6. 数値実験

本研究の有効性を確認するために人工的に生成した問題例を用いて数値実験を行った. 実験結果の詳細は当日報告する.

参考文献

[1] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization.* Springer International Publishing, 2017.

[2] M. Frank and P. Wolfe. An algorithm for quadratic programming. *Naval Research Logistics Quarterly*, 3(1-2):95–110, 1956.

[3] M. Jaggi. Revisiting Frank-Wolfe: Projection-free sparse convex optimization. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 427–435, 2013.