計算が困難な目的関数をもつ凸最適化問題に対する Frank–Wolfe 法

田中未来　数理・推論研究系 数理最適化グループ, 統計的機械学習研究センター 准教授

1　計算が困難な目的関数をもつ凸最適化問題

- 凸関数 $f\colon \mathbb{R}^n \to \mathbb{R}$, 凸集合 $S \subseteq \mathbb{R}^n$ で定まる次の凸最適化問題を考える:

$$f^* := \min_{x \in S} f(x). \tag{1}$$

- 多くの凸最適化アルゴリズムは次のような f や S に関する計算が容易であることが前提:
  - f に関する計算: $f(x)$ や $\nabla f(x)$ の計算など.
  - S に関する計算: $x \in S$? の判定や $\mathrm{proj}_S(x)$ の計算など.
- 実際の応用ではこれらの計算はしばしば困難.
- 本研究では S に関するある計算は容易だが f に関する計算が困難な設定で問題 (1) を解くアルゴリズムを考える.

2　目的関数に関する設定: なにが計算/利用可能?

- 0 次と 1 次の神託あり: $f(x)$, $\nabla f(x)$ の正確な値を計算可能.
  - 多くの凸最適化アルゴリズムで仮定. 本研究では仮定しない.
- 雑音入り 0 次の神託あり: $\nabla f(x)$ は計算不可能だが, $f(x)$ は雑音入りで計算可能.
  - Bayes 最適化や進化計算などと関係がありそう.
- 神託なし雑音入りデータあり: $f(x)$ は雑音入りでも計算不可能だが, 雑音入りデータ $\{(x_i, y_i)\}_{i=1}^m$ が利用可能.
  - あまり考えられていない設定のようだが, 実際の応用でもありそう.

3　素朴な方法: 2 段階法

アルゴリズム 1 問題 (1) に対する雑音入りデータを用いた 2 段階法

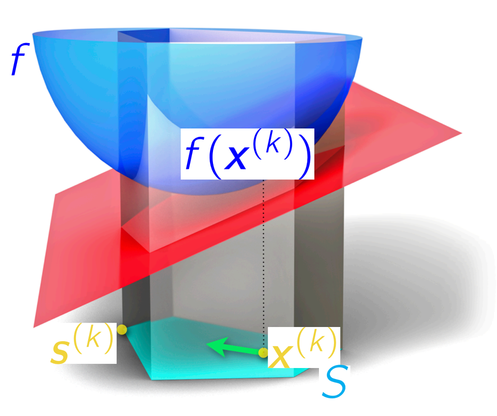
モデル $\hat{f}(\cdot, \boldsymbol{\theta})$ を用いて目的関数を推定: $\hat{\boldsymbol{\theta}} \in \operatorname*{argmin}_{\boldsymbol{\theta} \in \Theta} \sum_{i=1}^m (y_i - \hat{f}(\boldsymbol{x}_i, \boldsymbol{\theta}))^2$.
推定した目的関数 $\hat{f}(\cdot, \hat{\boldsymbol{\theta}})$ を用いて最適化: $\hat{x} \in \operatorname*{argmin}_{x \in S} \hat{f}(\boldsymbol{x}, \hat{\boldsymbol{\theta}})$.

- 問題点: 複雑な $\hat{f}$ を用いると予測精度は上がるが, 最適化が困難になる.

4　Frank–Wolfe 法 (Frank and Wolfe, 1956)

アルゴリズム 2 問題 (1) に対する Frank–Wolfe 法

$\boldsymbol{x}^{(0)} \in S$ を適当に定める.
for $k = 0, 1, \ldots$:
　子問題の求解: $\boldsymbol{s}^{(k)} \in \operatorname*{argmin}_{\boldsymbol{s} \in S} \langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{s} \rangle$.
　解の更新: $\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} + \gamma^{(k)}(\boldsymbol{s}^{(k)} - \boldsymbol{x}^{(k)})$.



- 実用上は子問題が容易に解ける必要がある (S に関する神託).
  - 近似解を s_k としてもよい (Jaggi (2013), S に関する弱い神託):

$$\boldsymbol{s}^{(k)} \in \left\{ \boldsymbol{s} \in S \colon \langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{s} \rangle \le \min_{\boldsymbol{s}' \in S} \langle \nabla f(\boldsymbol{x}^{(k)}), \boldsymbol{s}' \rangle + \delta_k \right\}.$$

- ステップ幅 $\gamma^{(k)}$ の決め方:
  - $\gamma^{(k)} \in \operatorname*{argmin}_{\gamma \in [0,1]} f(\boldsymbol{x}^{(k)} + \gamma(\boldsymbol{s}^{(k)} - \boldsymbol{x}^{(k)}))$ とするのが自然.
  - $\gamma^{(k)} := \dfrac{2}{k+2}$ としても理論的な収束の速さは変わらない.

5　提案手法: 近似勾配を用いた Frank–Wolfe 法

アルゴリズム 3 問題 (1) に対する近似勾配を用いた Frank–Wolfe 法

$\boldsymbol{x}^{(0)} \in S$ を適当に定める.
for $k = 0, 1, \ldots$:
　近似勾配の計算: $\boldsymbol{g}^{(k)} :\approx \nabla f(\boldsymbol{x}^{(k)})$.
　子問題の求解: $\boldsymbol{s}^{(k)} \in \operatorname*{argmin}_{\boldsymbol{s} \in S} \langle \boldsymbol{g}^{(k)}, \boldsymbol{s} \rangle$.
　解の更新: $\boldsymbol{x}^{(k+1)} := \boldsymbol{x}^{(k)} + \dfrac{2}{k+2}(\boldsymbol{s}^{(k)} - \boldsymbol{x}^{(k)})$.

- 1 次の神託ありの場合, $\boldsymbol{g}^{(k)} := \nabla f(\boldsymbol{x}^{(k)})$ としてアルゴリズム 2 に一致.
- 雑音入り 0 次の神託ありの場合, 勾配の差分近似を用いる.
- 神託なし雑音入りデータありの場合, 局所線形回帰の回帰係数を用いる:

$$(g_0^{(k)}, \boldsymbol{g}^{(k)}) \in \operatorname*{argmin}_{(g_0, \boldsymbol{g}) \in \mathbb{R} \times \mathbb{R}^n} \sum_{i=1}^m \kappa(\boldsymbol{x}_i, \boldsymbol{x}^{(k)})(y_i - g_0 - \langle \boldsymbol{g}, \boldsymbol{x}_i \rangle)^2. \tag{2}$$

6　提案手法の理論解析

6.1　アルゴリズムの収束解析

定理 1. ∇f の Lipschitz 定数を L, S の直径を D とし, アルゴリズム 3 によって生成された点列を $\{\boldsymbol{x}^{(k)}\}$ とする. このとき, 次が成り立つ:

$$f(\boldsymbol{x}^{(k)}) - f^* \le \frac{2LD^2}{k+2} + \frac{2}{k+1} \sum_{l=0}^{k-1} \delta_l \quad (k = 0, 1, \ldots).$$

ここで δ_k は $\delta_k \ge \|\nabla f(\boldsymbol{x}^{(k)}) - \boldsymbol{g}^{(k)}\|$ を満たす定数.

- S に関する弱い神託を用いたアルゴリズム 2 の収束解析 (Jaggi, 2013) に基づく.
- 右辺第 1 項はアルゴリズム 2 の収束率に一致: $\mathrm{O}(k^{-1})$.
- 右辺第 2 項は近似勾配の誤差に由来.
  - $\delta_k = \mathrm{O}(k^{-1})$ のとき, $\mathrm{O}(k^{-1})$ の収束率で $\lim_{k \to \infty} f(\boldsymbol{x}^{(k)}) = f^*$.
  - $\delta_k \le \delta$ (定数) のとき, $\mathrm{O}(k^{-1})$ の収束率で $\limsup_{k \to \infty} f(\boldsymbol{x}^{(k)}) \le f^* + 2\delta$.
- 雑音入り 0 次の神託ありの場合で勾配を差分近似したとき, $\mathrm{E}(\|\nabla f(\boldsymbol{x}^{(k)}) - \boldsymbol{g}^{(k)}\|)$ の上界が評価できる (Brekelmans et al., 2005).

6.2　局所線形回帰を用いた近似勾配の誤差解析

定理 2. 式 (2) で $\boldsymbol{g}^{(k)}$ を計算するとき, $\{x_i\}_{i=1}^m$ のうち $\boldsymbol{x}^{(k)}$ の m' 最近傍の添字集合を $N_{m'}(\boldsymbol{x}^{(k)})$ として, $i \in N_{m'}(\boldsymbol{x}^{(k)})$ のとき $\kappa(\boldsymbol{x}_i, \boldsymbol{x}^{(k)}) = 1$, そうでないとき 0 とする κ を用いる. $\{x_i\}_{i \in N_{m'}(\boldsymbol{x}^{(k)})}$ を並べた行列の最小特異値 $\rho_{\min}^{(k)}$ が正であることを仮定し, $\{\epsilon_i := y_i - f(x_i)\}_{i=1}^m$ は独立で, 各 $i = 1, \ldots, m$ について $\mathrm{E}(\epsilon_i) = 0$, $\mathrm{E}(\epsilon_i^2) = \sigma^2$ であることを仮定する. このとき, 次が成り立つ:

$$\mathrm{E}(\|\nabla f(\boldsymbol{x}^{(k)}) - \boldsymbol{g}^{(k)}\|) \le \frac{1}{\rho_{\min}^{(k)}} \sqrt{n\sigma^2 + \frac{m' L^2 D^4}{4}} \quad (k = 0, 1, \ldots).$$

本研究は山根大輝氏, 小林健氏, 中田和秀氏 (東工大) との共同研究である.