

Lipschitz 連続的平滑化を用いた平滑化加速近接勾配法

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 導入

本研究では次の最適化問題を効率よく解くアルゴリズムを考える:

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + h(x) + g(x). \quad (1)$$

- $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$: L 平滑な凸関数.
- $g: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$: 下半連続な真凸関数で prox_g の計算が容易.
- $h: \mathbb{R}^n \rightarrow \mathbb{R}$: h は平滑でなく, prox_{h+g} の計算も重い.
- 最適解集合 X^* は非空.

この問題の ϵ 近似解は, 非平滑な h を平滑な h_μ で近似 (平滑化) した問題

$$\min_{x \in \mathbb{R}^n} F_\mu(x) := f(x) + h_\mu(x) + g(x)$$

に対して加速近接勾配法 (FISTA) を適用することで得ることができる (Beck and Teboulle, 2012, Theorem 3.1). ただし, h_μ は h の $1/\mu$ 平滑近似関数 (定義 1) で, μ の値と必要な反復回数は ϵ に依存して決まる.

定義 1. 凸関数 $h: \mathbb{R}^n \rightarrow \mathbb{R}$ が (α, β) 平滑化可能であるとは, 任意の $\mu > 0$ に対して微分可能な凸関数 $h_\mu: \mathbb{R}^n \rightarrow \mathbb{R}$ が存在して, 次が成り立つことをいう:

1. 任意の $x \in \mathbb{R}^n$ に対して $h_\mu(x) \leq h(x) \leq h_\mu(x) + \beta\mu$.
2. h_μ は α/μ 平滑.

h_μ を h に対する $1/\mu$ 平滑近似関数とよぶ.

複数の ϵ から定まる μ で平滑化した問題に FISTA を適用した結果 (図 1) から, μ を徐々に小さくするアルゴリズムの有効性が示唆される.

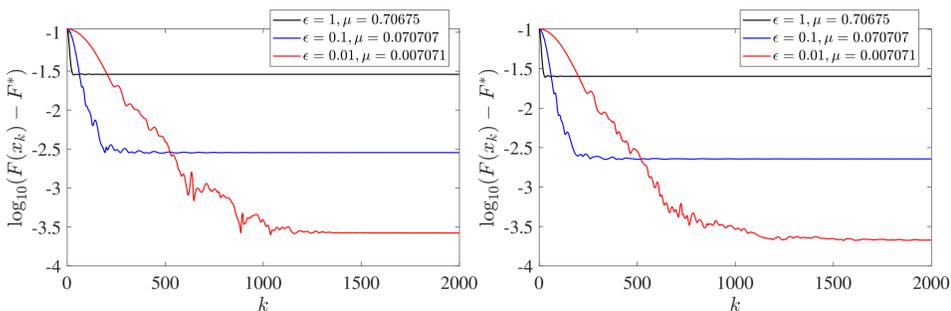


図 1: $F(x_k) - F^*$ の対数プロット. (左: f が非強凸. 右: f が強凸.)

2 平滑化近接勾配法

前述の μ を徐々に小さくするアルゴリズムとして次のようなもの考える.

アルゴリズム 1 問題 (1) に対する平滑化近接勾配法 (S-PGM)

- 1: 広義単調減少して 0 に収束する数列 $\{\mu_k\}$ を適当にとる.
- 2: 適当な初期点 $x_0 \in \text{dom } g$ をとる.
- 3: **for** $k := 0, 1, \dots$ ▷ 適当な終了条件を満たしたとき停止.
- 4: h に対する $1/\mu_{k+1}$ 平滑近似関数 $h_{\mu_{k+1}}$ を構成.
- 5: $t_{k+1} := 1/(L + \alpha/\mu_{k+1})$ とする.
- 6: $x_{k+1} := \text{prox}_{t_{k+1}g}(x_k - t_{k+1}\nabla(f + h_{\mu_{k+1}})(x_k))$ と更新.

このアルゴリズムの収束解析では $F_{\mu_{k+1}}(x_{k+1}) - F_{\mu_k}(x_k)$ を評価する. 平滑近似関数の定義から直ちに従う $h_{\mu_{k+1}}(x_{k+1}) - h_{\mu_k}(x_k) \leq \beta\mu_k$ は緩い. 以下では平滑近似関数に課す条件を強めて解析を行なう.

定義 2. (α, β) 平滑化可能な関数 $h: \mathbb{R}^n \rightarrow \mathbb{R}$ が (α, β) Lipschitz 連続的平滑化可能であるとは, 任意の広義単調減少して 0 に収束する平滑化パラメータの列 $\{\mu_k\}$ が与えられたとき, h に対するパラメータ (α, β) の $1/\mu_k$ 平滑近似関数 h_{μ_k} が存在して, 次が成り立つことをいう:

$$h_{\mu_{k+1}}(x) - h_{\mu_k}(x) \leq \beta(\mu_k - \mu_{k+1}) \quad (x \in \mathbb{R}^n). \quad (2)$$

h_{μ_k} を h に対する $1/\mu_k$ Lipschitz 連続的平滑近似関数とよぶ.

式 (2) といわゆる prox-grad 不等式を用いると次の補題を得る.

補題 3. h_{μ_k} を h の $1/\mu_k$ Lipschitz 連続的平滑近似関数とし, 平滑化近接勾配法によって生成された点列を $\{x_k\}$ とする. このとき, 任意の $x^* \in X^*$ と $k = 1, 2, \dots$ に対して次が成り立つ:

$$F(x_k) - F^* \leq \frac{(1/2)\|x_0 - x^*\|_2^2 + \beta \sum_{l=1}^k t_l \mu_l}{\sum_{l=1}^k t_l}. \quad (3)$$

本研究では $\mu_k = \mu_1 k^{-\nu}$ ($0 < \nu \leq 1$) の場合について式 (3) の右辺を評価した. オーダの意味で最良だった $\nu = 1/2$ の場合の結果を以下に示す.

定理 4. 補題 3 において $\mu_k = \mu_1 k^{-1/2}$ とする. このとき, 任意の $x^* \in X^*$ と $k = 1, 2, \dots$ に対して次が成り立つ:

$$F(x_k) - F^* \leq O(k^{-1/2} \log k).$$

3 平滑化加速近接勾配法

詳細は割愛するが, 本研究では FGM (Nesterov, 1983; Florea and Vorobyov, 2019) と AMGS (Nesterov, 2013) という 2 種類の加速近接勾配法に基づく平滑化加速近接勾配法の収束率の解析を行なった. 結果を表 1 に示す.

表 1: 収束率のオーダの比較.

アルゴリズム	f の強凸性	$\mu_k =$	$F(x_k) - F^* \leq$
S-PGM	利用しない	$\mu_1 k^{-1/2}$	$O(k^{-1/2} \log k)$
S-FGM	利用しない	$\mu_1 k^{-1}$	$O(k^{-1} \log k)$
S-FGM	利用する	$(4\alpha/\sigma)k^{-2}$	$O(k^{-2} \log k)$
S-AMGS	利用しない	$\mu_1 k^{-1}$	$O(k^{-1} \log k)$
S-AMGS	利用する	$(2\alpha/\sigma)k^{-2}$	$O(k^{-2} \log k)$

4 計算機実験

あるスパース制御問題に対して適用した結果を図 2-3 に示す. 平滑化加速近接勾配法 (特に強凸性を利用したもの) の有効性が示唆される.

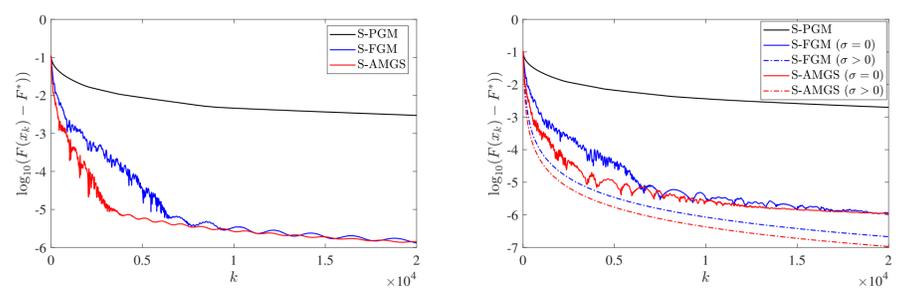


図 2: $F(x_k) - F^*$ の対数プロット. (左: f が非強凸. 右: f が強凸.)

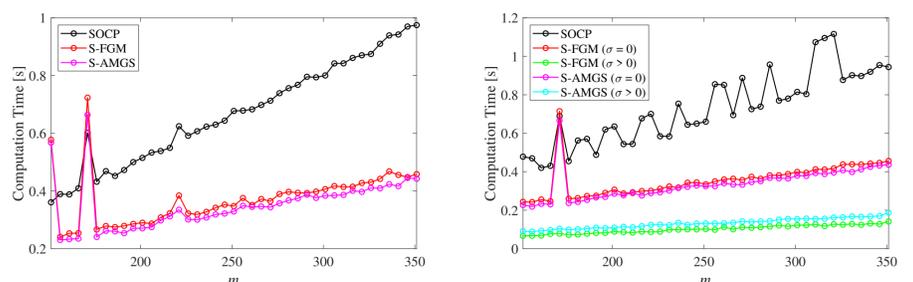


図 3: 計算時間. (左: f が非強凸. 右: f が強凸. SOCP: YALMIP-SeDuMi.)

謝辞: 本研究は豊田充氏 (都立大) との共同研究です.