

交互近接勾配法の改良について

秋田県立大学大学院電子情報システム学専攻 小野寺優希也

2021 年 12 月 17 日 (金)

1 背景

本研究では以下の最適化問題を考える：

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 + \lambda \|x\|_1, \quad (1)$$

ここで、 $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $\|\cdot\|_1$ は ℓ_1 ノルム, $\|\cdot\|_2$ は ℓ_2 ノルム, $\lambda > 0$ である．式 (1) の最適化問題はラッソ回帰と呼ばれている．ラッソ回帰は曲線フィッティングや画像信号処理へ応用できる数理モデルとして知られている [2, 5]．

問題 (1) の最適解を求める方法として近接勾配法が知られており、近接勾配法で生成された点列は問題 (1) の最適解 x^* に単調に収束する．一方、Attouch と Peypouquet は近接勾配法の加速化法として FISTA を提案した [1]．しかし、FISTA によって生成された点列は単調に解に収束せず、振動しながら解に近づく問題点がある [4]．最近、FISTA の問題点を改善するために Iutzeler と Hendrickx によって交互近接勾配法が提案された [3]．しかし、収束の速さは FISTA に比べて劣っている．

2 目的

本研究では交互近接勾配法に基づく新たなアルゴリズムを提案する．具体的には先行研究で挙げられているアルゴリズムよりも繰り返し回数が少なくなるようなアルゴリズムを提案する．加えて、提案手法をラッソ回帰に対して適用し、数値実験により収束の速さを比較する．

3 準備

$\gamma > 0$, $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ を凸関数とする．このとき、 f の近接写像 $\text{prox}_{\gamma f}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ を以下のように定義する．

$$\text{prox}_{\gamma f}(x) := \arg \min_{y \in \mathbb{R}^n} \left(f(y) + \frac{1}{2\gamma} \|x - y\|_2^2 \right) \quad (2)$$

ここで、 $\arg \min g(x)$ で $g(x)$ を最小とする解を表す．本研究で用いるアルゴリズムにおいて ℓ_1 ノルムの近接写像を利用する． ℓ_1 ノルムの近接写像 $\text{prox}_{\gamma \|\cdot\|_1}$ は $i = 1, 2, \dots, n$ に対して

$$[\text{prox}_{\gamma \|\cdot\|_1}(x)]_i = \text{sgn}(x_i) \max\{|x_i| - \gamma, 0\} \quad (3)$$

で与えられる．ここで、 sgn は符号を表す．

4 交互近接勾配法と提案手法

Iutzeler と Hendrickx によって提案された交互近接勾配法を紹介する [3]．問題 (1) に対する交互近接勾配法は $k = 1, 2, \dots$ に対して

$$x_{k+1} = \text{prox}_{\gamma \lambda \|\cdot\|_1}(y_k - \gamma A^T(Ay_k - b)) \quad (4)$$

で与えられる．ただし、

$$y_k = \begin{cases} x_k & (k \text{ は偶数}) \\ x_k + \beta_k(x_k - x_{k-1}) & (k \text{ は奇数}) \end{cases} \quad (5)$$

であり、 $\gamma = 1/\|A^T A\|$, $\beta_k = 0.5$ である． $\{x_k\}$ を交互近接勾配法から生成された点列とし、 x^* を問題 (1) の最適解としたとき、 $\{x_k\}$ は x^* に収束する．

以下では交互近接勾配法を用いた新たなアルゴリズムを紹介する．問題 (1) に対する提案手法は $k = 1, 2, \dots$ に対して、

$$x_{k+1} = \text{prox}_{\gamma \lambda \|\cdot\|_1}(y_k - \gamma A^T(Ay_k - b)) \quad (6)$$

である．ただし、 $\|x_k - \text{prox}_{\gamma \lambda \|\cdot\|_1}(x_k - \gamma A^T(Ax_k - b))\|_2 > \varepsilon'$ のとき、点列 $\{y_k\}$ は

$$y_k = x_k + \alpha_k(x_k - x_{k-1}) \quad (7)$$

である．点列 $\{\alpha_k\}$ の詳細は発表で紹介する．また、 k_0 が $\|x_{k_0} - \text{prox}_{\gamma \lambda \|\cdot\|_1}(x_{k_0} - \gamma A^T(Ax_{k_0} - b))\|_2 \leq \varepsilon'$ を満たす最小の値のとき、 $k \geq k_0$ に対して、 $\{y_k\}$ は

$$y_k = \begin{cases} x_k & (k \text{ は偶数}) \\ x_k + \beta_k(x_k - x_{k-1}) & (k \text{ は奇数}) \end{cases} \quad (8)$$

である．

5 数値実験

ラッソ回帰に対して近接勾配法 (PG), FISTA, 交互近接勾配法 (APG), 提案手法 (Proposed) を適用し結果を比較した.

今回考えるラッソ回帰の問題設定は以下の通りである. 問題 (1) において, $A \in \mathbb{R}^{130 \times 80}$, $b \in \mathbb{R}^{130}$ とする. ここで, A の各成分は標準正規分布に従っており, b は $b = A\hat{x} + e$ である. ただし, $\hat{x} \in \mathbb{R}^{80}$ は標準正規分布から得られる 10% スパースベクトル, $e \in \mathbb{R}^{130}$ の各成分は平均 0, 分散 0.001^2 の正規分布に従っている. 本実験では, 正則化パラメータは $\lambda = 0.1$ としている. また, 提案手法で用いる ε' は $\varepsilon' = 10^{-3}$ としている.

A , b を上記のように生成し, 初期点を原点とする. また, 終了条件を (相対双対ギャップ) $< 10^{-8}$ とする. このとき, 終了条件を満たすまでにかかった繰り返し回数と実行時間を求めた. 各アルゴリズムに対して実験を 100 回行ない, 繰り返し回数と実行時間の平均を取ったものを表 1 に示す. また, 同様の実験を 10 回行ない, 各繰り返し回数における点列 $\{x_k\}$ と近似解 x^* の差 $\|x_k - x^*\|_2$ の平均を取ったものを図 1 に示す. 表 1 より提案手法をラッソ回帰へ適用した結果, 繰り返し回数を交互近接勾配法と比較して 30.9 % 改善することができた.

行列のサイズを $(m, n) = (650, 400), (1300, 800)$ に変更して数値実験を行なった. こちらの数値実験の結果に関しては発表で紹介する.

6 結論

本研究ではラッソ回帰における新たなアルゴリズムを提案した. 提案手法をラッソ回帰に適用した結果, 先行研究のアルゴリズムと比較して繰り返し回数を 30.9 % から 36.4 % 改善することができた.

今後の課題は提案手法で用いているアルゴリズムの切り替えの評価式の ε' の効率の良い決定方法を検討することである.

参考文献

[1] H. ATTOUCH AND J. PEYPOUQUET, *The*

表 1 各アルゴリズムに対する繰り返し回数 (Iterations) と実行時間 (CPU time) の平均の比較 ($m = 130, n = 80$)

	Iteration	CPU time[s]
PG	181.89	0.02090
FISTA	157.32	0.01782
APG	143.32	0.01609
Proposed	99.05	0.01375

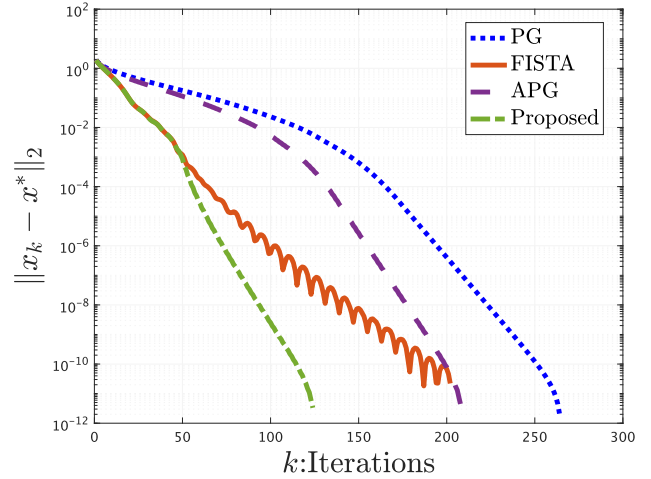


図 1 各繰り返し回数における点列 $\{x_k\}$ と近似解 x^* の差 $\|x_k - x^*\|_2$ の関係 ($m = 130, n = 80$)

rate of convergence of nesterov's accelerated forward-backward method is actually faster than $1/k^2$, SIAM J. Optim., 26 (2016), pp. 1824–1834.

- [2] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40 (2011), pp. 120–145.
- [3] F. IUTZELER AND J. MALICK, *On the proximal gradient algorithm with alternated inertia*, J. Optim. Theory Appl., 176 (2018), pp. 688–710.
- [4] Z. MU AND Y. PENG, *A note on the inertial proximal point method*, Stat. Optim. Inf. Comput., 3 (2015), pp. 241–248.
- [5] M. NAGAHARA, *Sparsity Methods for Systems and Control*, Now Publishers, 2020.