交互近接勾配法とその応用

小野寺優希也　松下慎也　徐粒

2021 年 12 月 8 日

発表の流れ

目次

背景

本研究では以下の最適化問題を考える：

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

- $A\in\mathbb{R}^{m\times n}$，$b\in\mathbb{R}^m$，$\|\cdot\|_1$ は ℓ_1 ノルム [†]，$\|\cdot\|_2$ は ℓ_2 ノルム [‡]，$\lambda>0$
- 式 (1) の最適化問題はラッソ回帰と呼ばれている
- ラッソ回帰の応用例
  - 曲線フィッティング [5]
  - 画像信号処理 [2]

[†] $x=(x_1,\ldots,x_n)$ のとき，$\|x\|_1:=|x_1|+|x_2|+\cdots+|x_n|$

[‡] $x=(x_1,\ldots,x_n)$ のとき，$\|x\|_2:=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$

[5] M. Nagahara, *Sparsity methods for systems and control*, Now Publishers, 2020

[2] A. Chambolle and T. Pock, J. Math. Imaging Vision, 40 (2011), pp.120-145

ラッソ回帰

アメリカの都市における犯罪に関するデータ[†]に対し，ラッソ回帰を適用する．

Table 1: 米国犯罪データ

項目番号	データの内容
1	人口 100 万人あたりの犯罪率
2	警察への年間資金
3	25 歳以上で高校を卒業した人の割合
4	16-19 歳で高校に通っていない人の割合
5	18-24 歳で大学生の割合
6	25 歳以上で 4 年制大学を卒業した人の割合

- $b \in \mathbb{R}^{50}$（目的変数）：項目番号 1 のデータ
- $A \in \mathbb{R}^{50\times 5}$（説明変数）：項目番号 2 から 6 のデータ

[†] https://hastie.su.domains/StatLearnSparsity/data.html

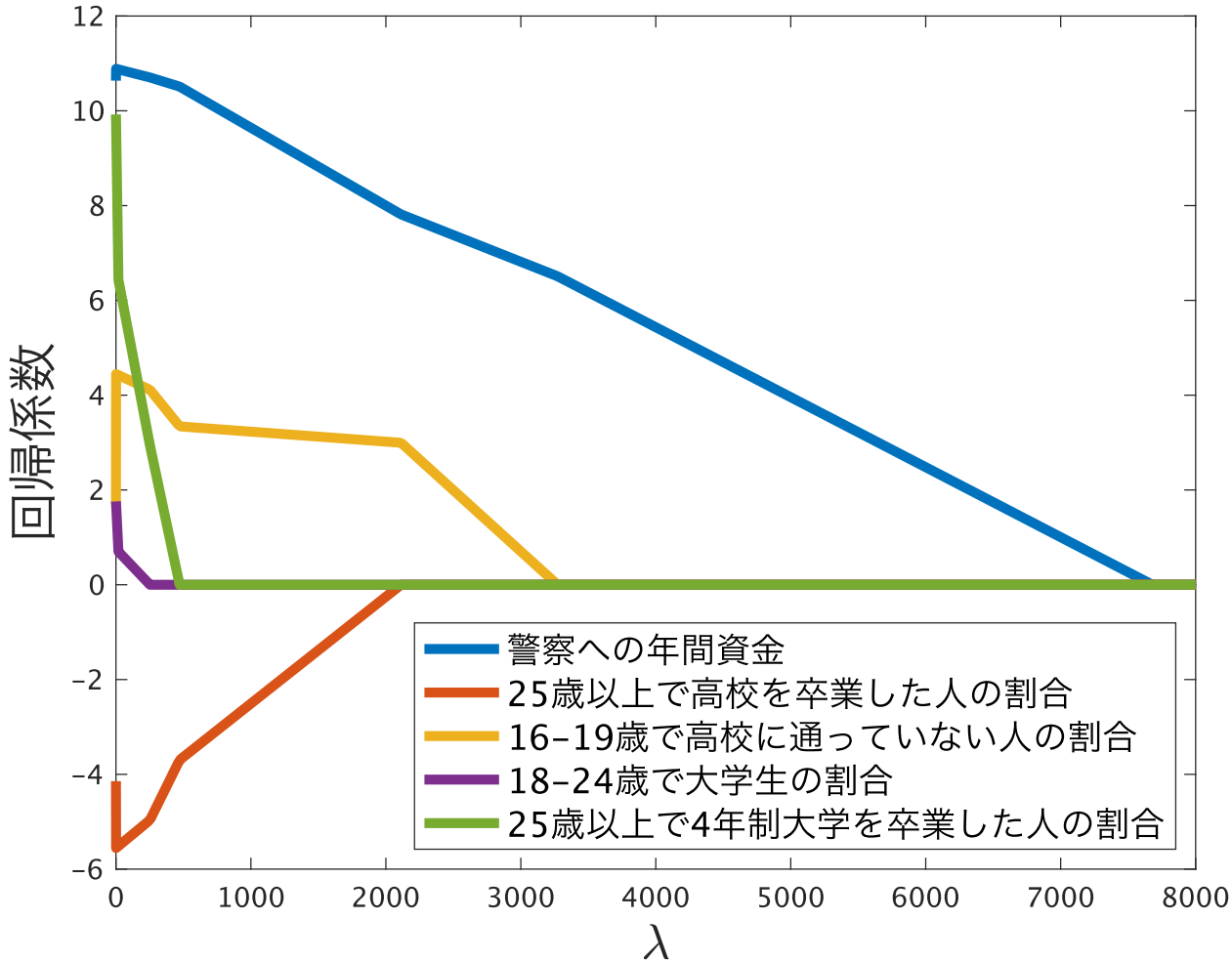

Figure 1: λ と回帰係数の関係

- 回帰係数が 0 となるタイミングは変数ごとに異なる.
- 目的変数 (b) に影響を与える変数と影響を与えない変数を選択することができる.

背景

- ラッソ回帰の解法：近接勾配法
  - 生成された点列は問題 (1) の最適解 x^* に単調に収束
- 近接勾配法の加速化：FISTA [1]
  - FISTA の問題点：振動しながら解に到達する [4]
- 振動を抑える方法として交互近接勾配法が提案された [3]
  - 収束の速さは FISTA に比べて劣っている

[1] H. Attouch and J. Peypouquet, SIAM J. Optim., 26 (2016), pp. 1824-1834.
[4] Z. Mu and Y. Peng, Stat. Optim. Inf. Comput., 3 (2015), pp. 241-248.
[3] F. Iutzeler and J. Malick, J. Optim. Theory Appl., 176 (2018), pp. 688-710.

近接勾配法，FISTA，交互近接勾配法の比較

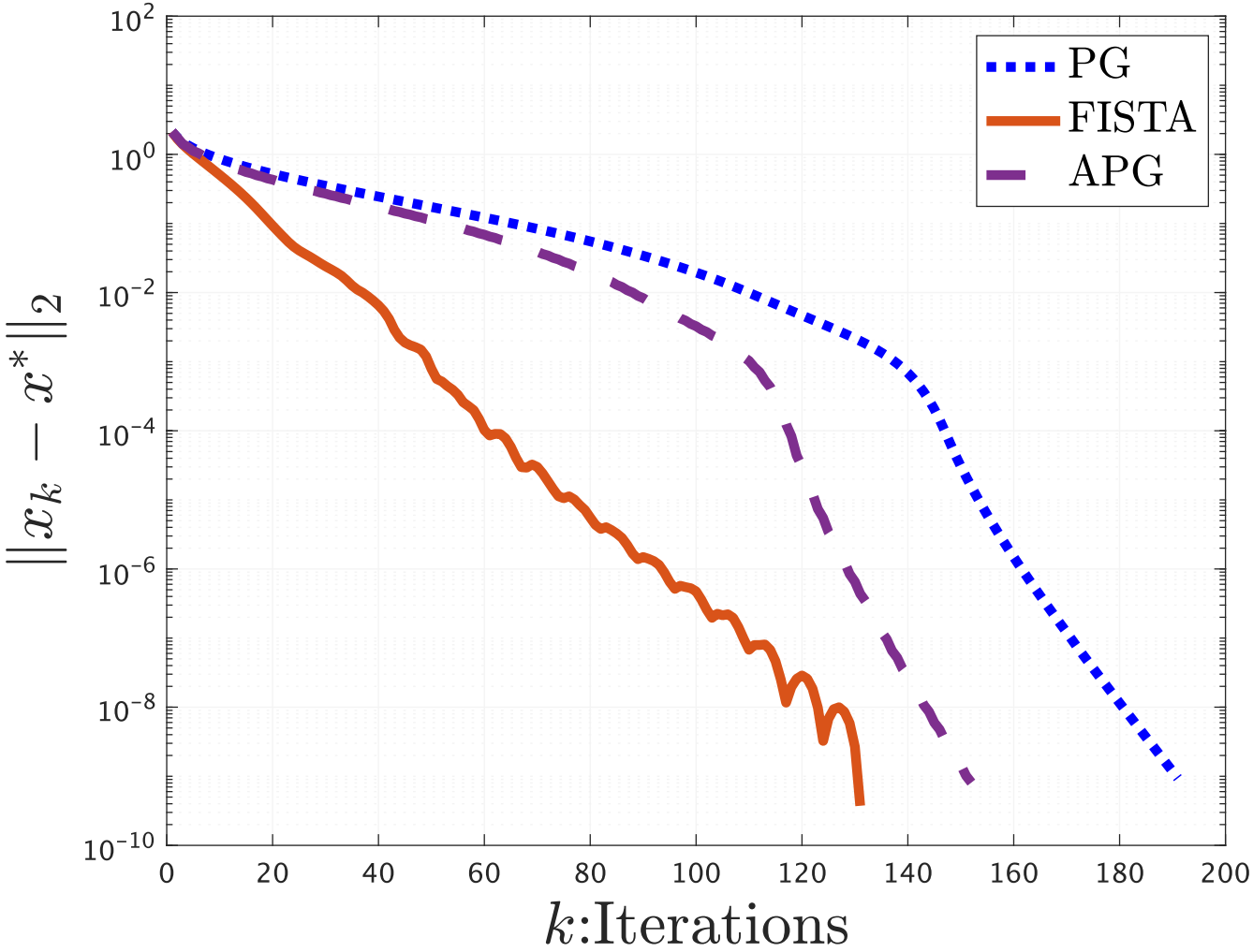

Figure 2: ラッソ回帰において近接勾配法 (PG)，FISTA，交互近接勾配法 (APG) を適用した結果

目的

Table 2: 終了条件 $\|x_k - x^*\|_2 < 10^{-8}$ に対する繰り返し回数と実行時間の平均の比較

	Iteration	CPU time[s]
PG	159.4	0.01542
FISTA	115.6	0.01030
APG	126.2	0.01266

目的

- 先行研究に基づく新たなアルゴリズムを提案する.
- 提案手法をラッソ回帰に対して適用し，数値実験によって収束の速さを比較する.

目次

ℓ_1 ノルムとスパース性

minimize $\|x\|_1$
subject to $Ax = b$

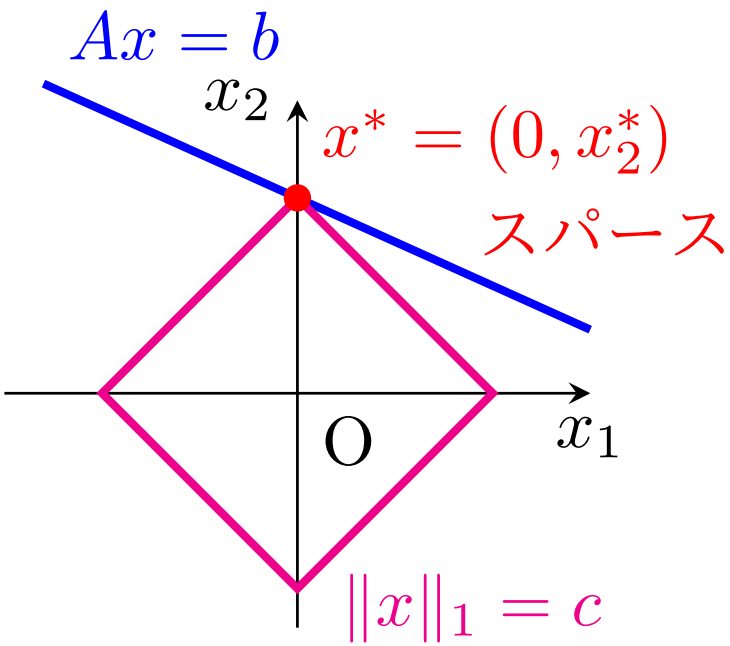

Figure 3: ℓ_1 ノルム

minimize $\|x\|_2$
subject to $Ax = b$

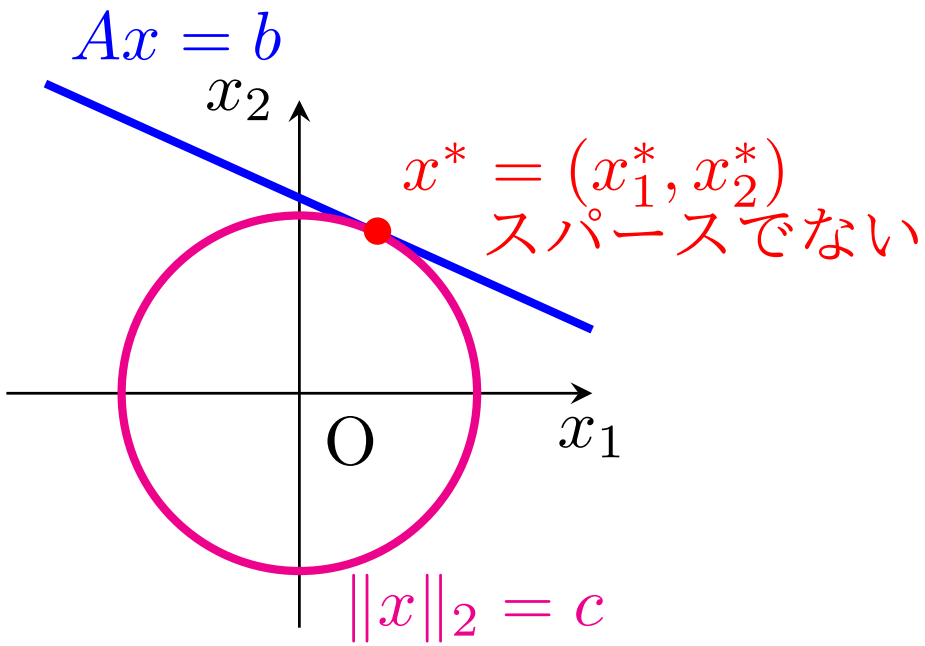

Figure 4: ℓ_2 ノルム

近接写像

アルゴリズムにおいて ℓ_1 ノルム

$$\|x\|_1 := |x_1| + |x_2| + \cdots + |x_n|$$

の近接写像を利用する.

Definition 1 (近接写像)

$\gamma > 0$, $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ を凸関数とする. f の近接写像 $\mathrm{prox}_{\gamma f} : \mathbb{R}^n \to \mathbb{R}^n$ を以下で定義する.

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

ここで, $\arg\min g(x)$ で $g(x)$ を最小とする解を表す.

近接写像の例

Example 1 (ℓ_1 ノルムの近接写像)

ℓ_1 ノルムの近接写像 $\mathrm{prox}_{\gamma\|\cdot\|_1}$ は以下で与えられる.

$$[\mathrm{prox}_{\gamma\|\cdot\|_1}(x)]_i = \begin{cases} x_i - \gamma & (x_i > \gamma) \\ 0 & (-\gamma \le x_i \le \gamma) \\ x_i + \gamma & (x_i < -\gamma) \end{cases} \quad (i = 1, 2, \ldots, n) \tag{3}$$

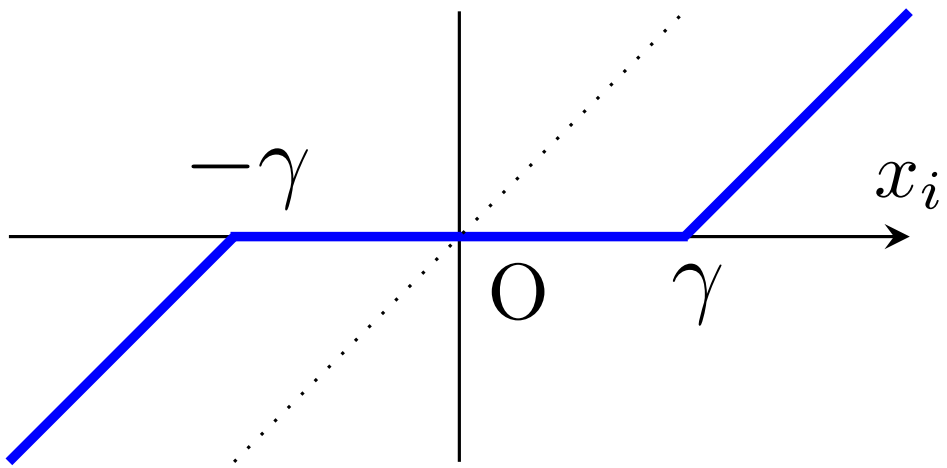

Figure 5: $[\mathrm{prox}_{\gamma\|\cdot\|_1}(x)]_i$

目次

近接勾配法

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

問題 (1) に対して次の近接勾配法が知られている.

近接勾配法

$x_0\in\mathbb{R}^n$, $\gamma=1/\|A^TA\|$, $\lambda>0$, $k=0,1,2,\ldots$ に対して

$$x_{k+1}=\mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_k-\gamma A^T(Ax_k-b)). \tag{4}$$

- $\{x_k\}$ は問題 (1) の最適解 x^* に収束する
- $\sqrt{k}\|x_{k+1}-x_k\|_2\to 0\ (k\to\infty)\Leftrightarrow\|x_{k+1}-x_k\|_2=o(1/\sqrt{k})$

FISTA[1]

近接勾配法の加速化法として FISTA が知られている.

FISTA

$x_0, x_1 \in \mathbb{R}^n$, $\gamma = 1/\|A^TA\|$, $\lambda > 0$, $k = 1, 2, \ldots$ に対して

$$\begin{cases} y_k = x_k + \alpha_k(x_k - x_{k-1}) \\ x_{k+1} = \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(y_k - \gamma A^T(Ay_k - b)), \end{cases} \tag{5}$$

ただし,

$$\alpha_k = \frac{t_{k-1} - 1}{t_k}, \quad t_k = \frac{k+3}{3}, \quad t_0 = 1. \tag{6}$$

- $\{x_k\}$ は問題 (1) の最適解 x^* に収束する.
- $k\|x_{k+1} - x_k\|_2 \to 0$ $(k \to \infty) \Leftrightarrow \|x_{k+1} - x_k\|_2 = o(1/k)$

[1] H. Attouch and J. Peypouquet, SIAM J. Optim., 26 (2016), pp. 1824-1834.

近接勾配法と FISTA の比較

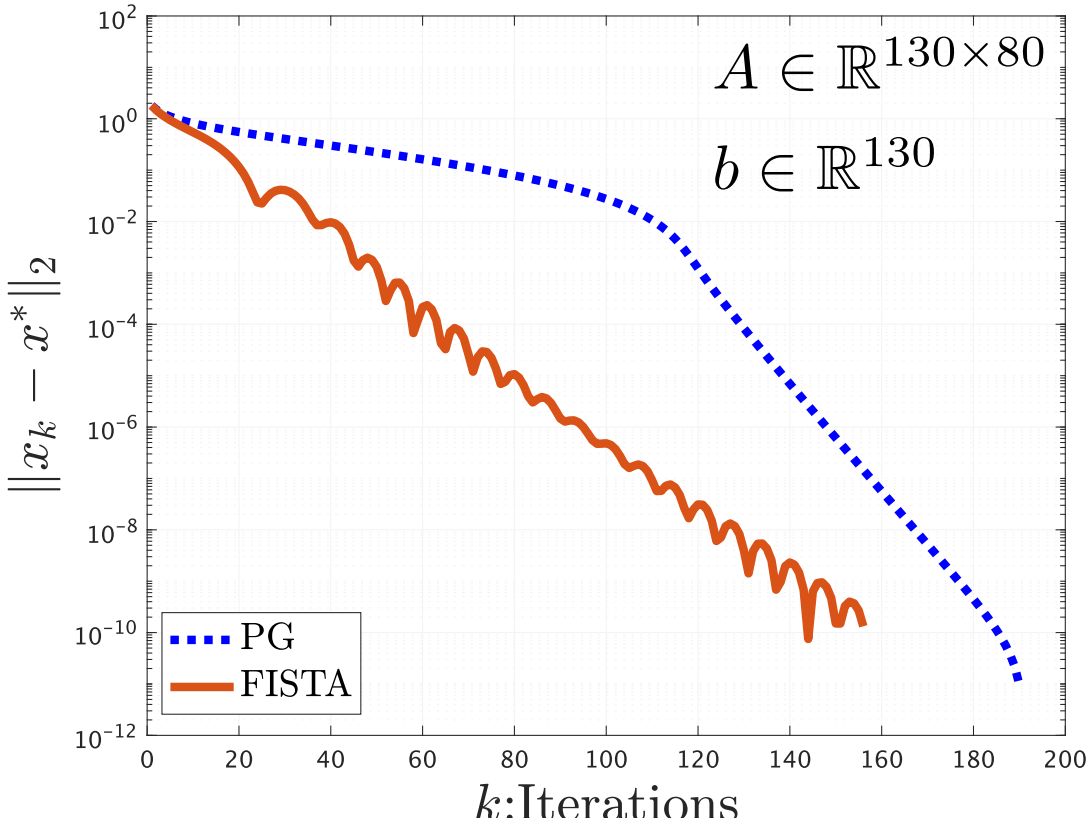

Figure 6: ラッソ回帰に対して近接勾配法 (PG) と FISTA を適用して点列 $\{x_k\}$ と最適解 x^* の差 $\|x_k - x^*\|_2$ を比較した結果

FISTA：振動しながら解に近づいている

FISTA の問題点 [4]

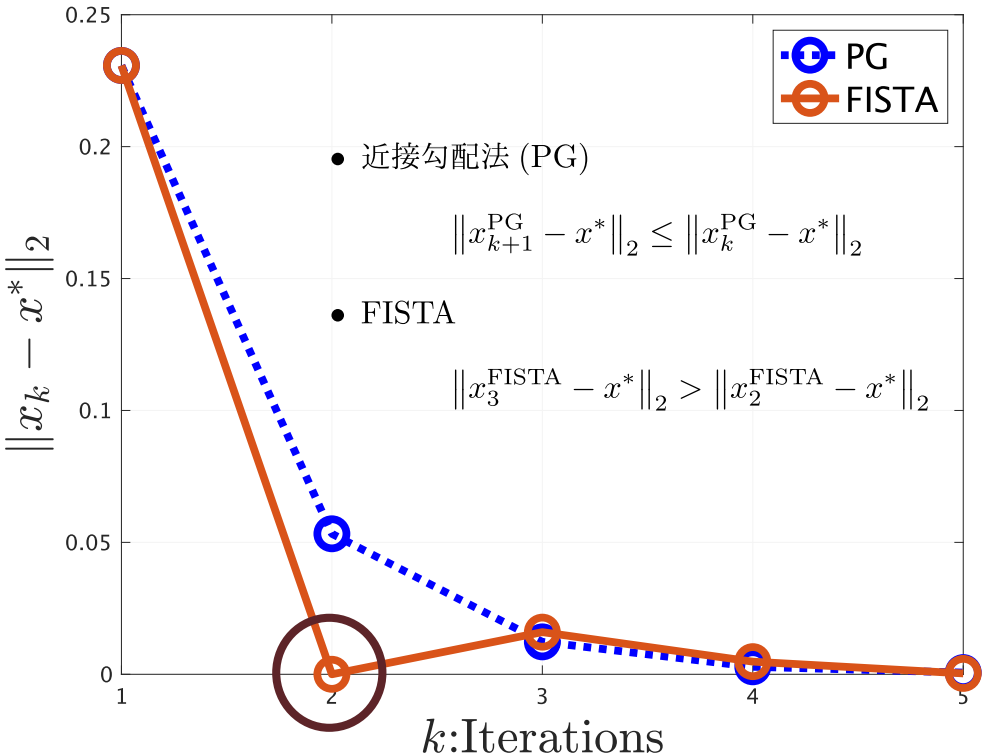

Figure 7: $\min \dfrac{x^2}{2}$ に対して近接勾配法 (PG) と FISTA を適用した結果

- FISTA：x_2 で解に到達しているが，x_3 では解から離れる

[4] Z. Mu and Y. Peng, Stat. Optim. Inf. Comput., 3 (2015), pp. 241-248.

交互近接勾配法 [3]

振動を抑える方法として交互近接勾配法が知られている.

交互近接勾配法

$x_0, x_1 \in \mathbb{R}^n$, $\gamma = 1/\|A^T A\|$, $\beta_k = 0.5$, $k = 1, 2, \ldots,$ に対して,

$$x_{k+1} = \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(y_k - \gamma A^T(Ay_k - b)) \tag{7}$$

ただし,

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

$\{x_k\}$ は問題 (1) の最適解 x^* に収束する.

[3] F. Iutzeler and J. Malick, J. Optim. Theory Appl., 176 (2018), pp. 688-710.

近接勾配法，FISTA，交互近接勾配法の比較

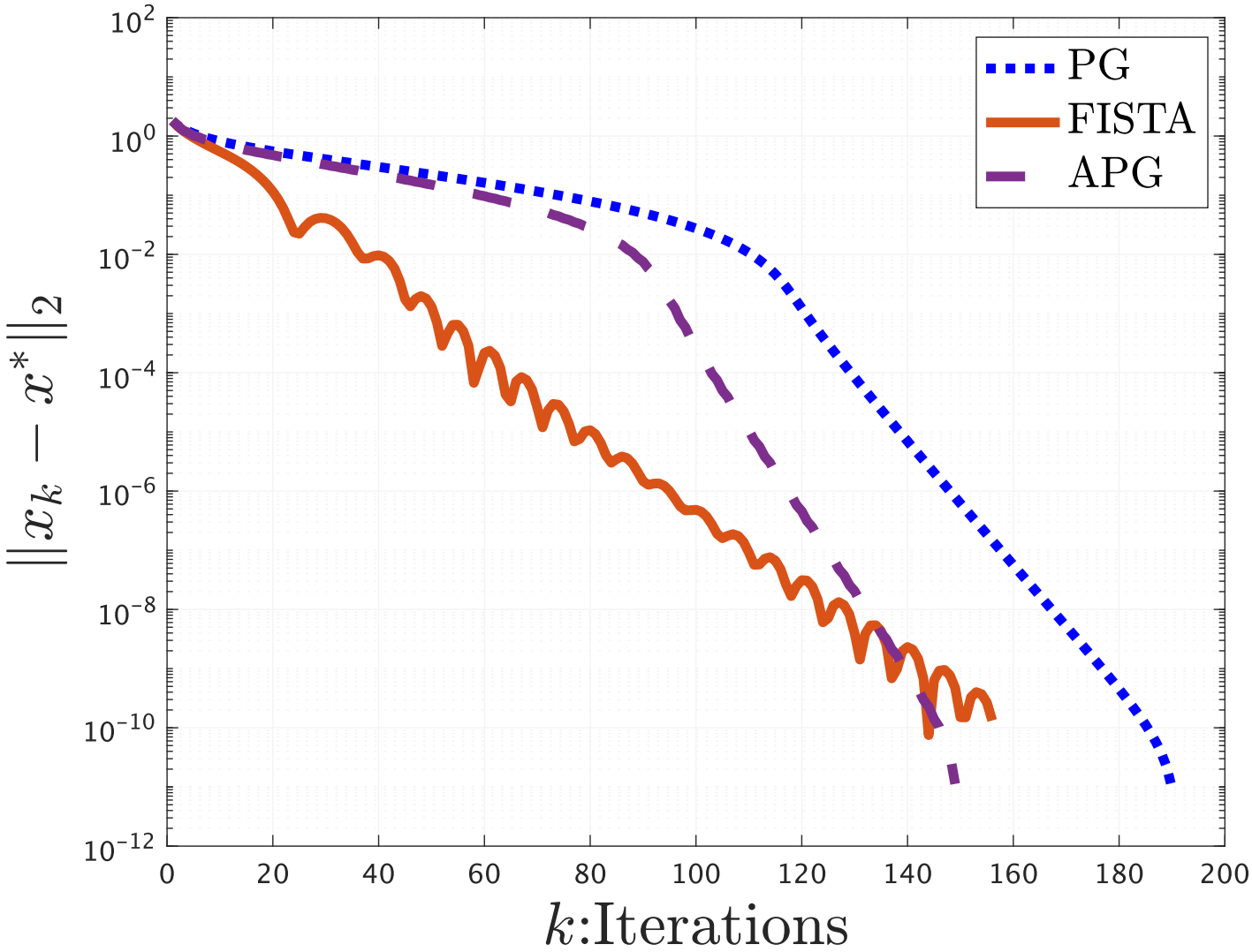

Figure 8: 近接勾配法 (PG)，FISTA，交互近接勾配法 (APG) の比較

FISTA と交互近接勾配法の比較

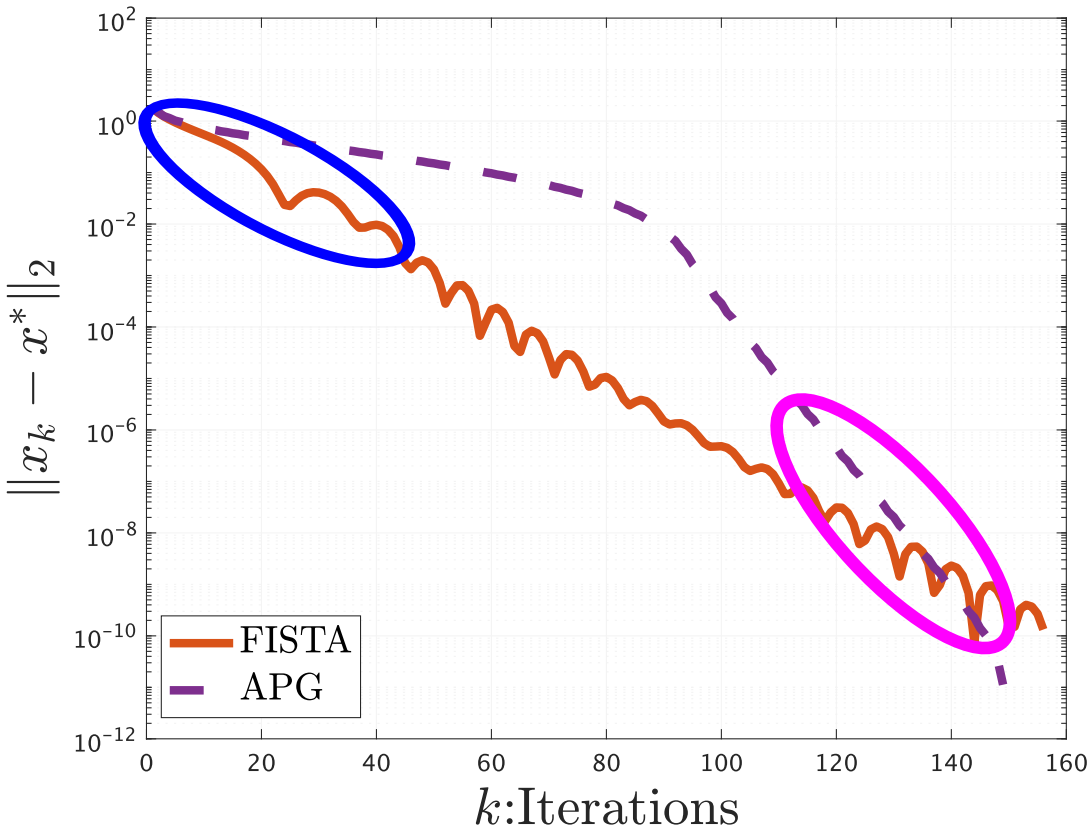

Figure 9: FISTA と交互近接勾配法 (APG) の比較

- 解から遠い：FISTA の方が速く解に近づく
- 解から近い：交互近接勾配法の方が速く解に近づく

考察

Figure 9 からわかること

- 解から遠い：FISTA で点列を生成
- 解から近い：交互近接勾配法で点列を生成

とするとより速く解に収束するアルゴリズムができるのではないか？

(FISTA ⇒ 交互近接勾配法) アルゴリズムを切り替える必要がある

- $\|x_k - x^*\|_2$：解がわかっている場合のみしか使用できない
- $\|x_k - \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_k - \gamma A^T(Ax_k - b))\|_2$：解に到達した場合にアルゴリズムが切り替わる

切り替えの評価式

アルゴリズムの切り替える評価式として
$\|x_k - \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_k - \gamma A^T(Ax_k - b))\|_2$ を採用する

提案手法

提案手法

$x_0, x_1 \in \mathbb{R}^n$，$\gamma = 1/\|A^TA\|$，$k = 1, 2, \ldots$ に対して

$$x_{k+1} = \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(y_k - \gamma A^T(Ay_k - b)). \tag{9}$$

$\|x_k - \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_k - \gamma A^T(Ax_k - b))\|_2 > \varepsilon'$ のとき，

$$y_k = x_k + \alpha_k(x_k - x_{k-1}) \tag{10}$$

k_0 が $\|x_{k_0} - \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_{k_0} - \gamma A^T(Ax_{k_0} - b))\|_2 \leq \varepsilon'$ を満たす最小の値のとき，$k \geq k_0$ に対して

$$y_k = \begin{cases} x_k & (k \text{ は偶数}) \\ x_k + \beta_k(x_k - x_{k-1}) & (k \text{ は奇数}) \end{cases}. \tag{11}$$

目次

数値実験の概要

ラッソ回帰

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

- 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
- 終了条件：(相対双対ギャップ) $< 10^{-8}$

Table 2: 数値実験で用いた計算環境

OS	macOS Big Sur
メモリ	8 GB 1600 MHz DDR3
CPU	1.8 GHz Intel Core i5
使用ソフト	MATLAB R2021a

数値実験の問題設定

ラッソ回帰

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

- $A \in \mathbb{R}^{m \times n}$：各成分が標準正規分布に従うように生成
- $b \in \mathbb{R}^m$：$b = A\hat{x} + e$
  - $\hat{x} \in \mathbb{R}^n$：標準正規分布から得られる 10 % スパースベクトル
  - $e \in \mathbb{R}^m$：平均 0，分散 0.001^2 の正規分布に従うように生成
- 問題のサイズ：$(m, n) = (130, 80), (650, 400), (1300, 800)$
- 正則化パラメータ：$\lambda = 0.1$
- ε'(提案手法)：$\varepsilon' = 10^{-3}$

ε' の決め方

提案手法において FISTA と交互近接勾配法の切り替える評価式の $\|x_k - \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_k - \gamma A^T(Ax_k - b))\|_2 < \varepsilon'$ の ε' を変更させて実験を行ない，繰り返し回数を比較した．

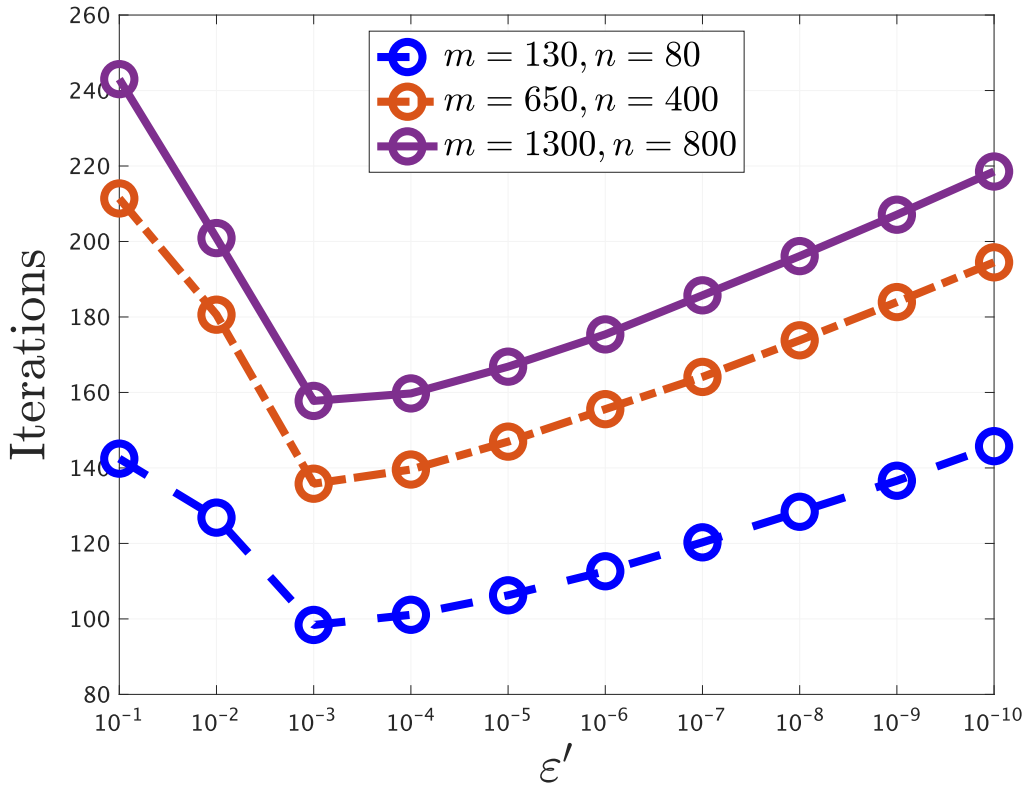

Figure 10: A，b を 200 回変更し，初期点を固定した際の実験結果

数値実験の結果

- A，b を先程のように生成し，初期点を原点とし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比較した．

Table 3: 各アルゴリズムに対する繰り返し回数と実行時間の平均の比較 $(m = 130, n = 80)$

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

交互近接勾配法と比較して繰り返し回数が 30.9 %改善

数値実験の結果

Table 4: 各アルゴリズムに対する繰り返し回数と実行時間の平均の比較 $(m = 650, n = 400)$

	Iteration	CPU time[s]
PG	269.14	1.31309
FISTA	215.19	1.04852
APG	212.87	1.04364
Proposed	135.41	0.86179

繰り返し回数が 36.4% 改善

Table 5: 各アルゴリズムに対する繰り返し回数と実行時間の平均の比較 $(m = 1300, n = 800)$

	Iteration	CPU time[s]
PG	309.90	10.97978
FISTA	241.51	8.53192
APG	245.41	9.11189
Proposed	157.16	7.02321

繰り返し回数が 34.9% 改善

数値実験の結果

A，b を 10 回変更させて数値実験を行ない，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示した．

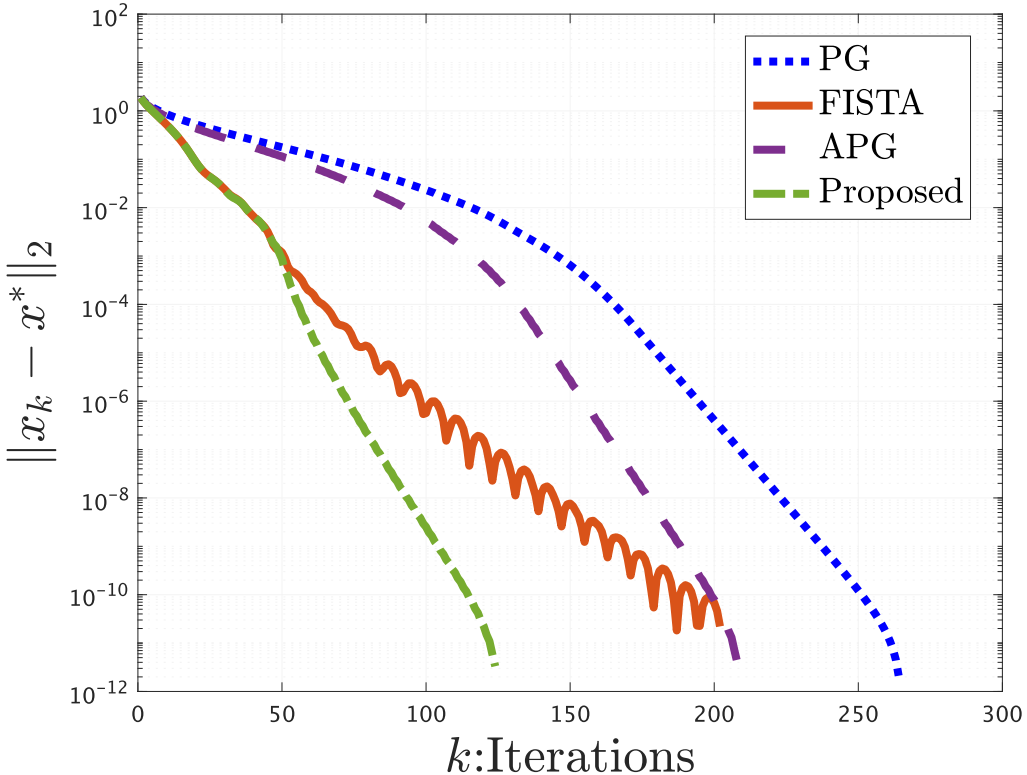

Figure 11: 各繰り返し回数における点列 $\{x_k\}$ と近似解 x^* の差 $\|x_k - x^*\|_2$ の関係 $(m = 130, n = 80)$

数値実験の結果

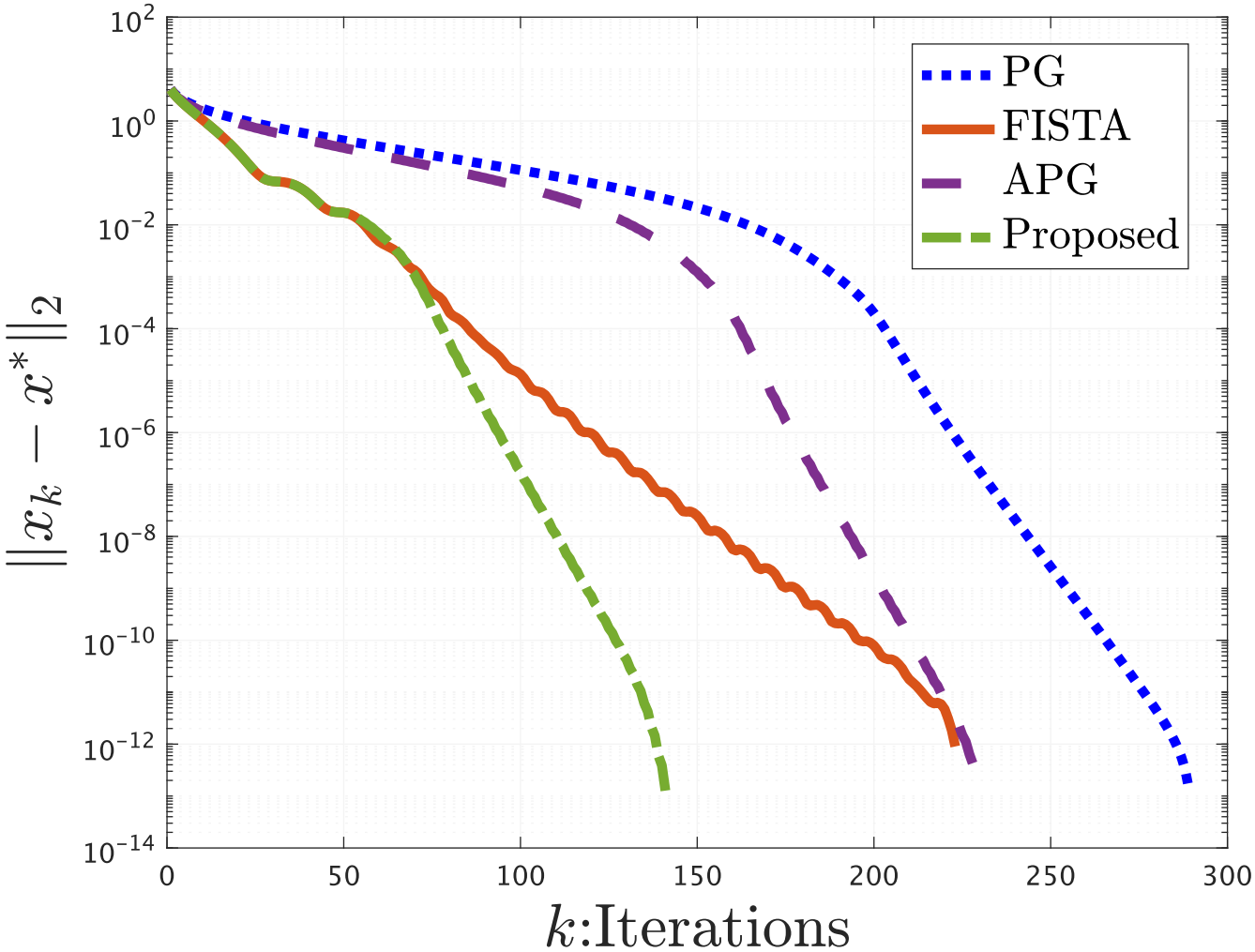

Figure 12: 各繰り返し回数における点列 $\{x_k\}$ と近似解 x^* の差 $\|x_k - x^*\|_2$ の関係 $(m = 650, n = 400)$

数値実験の結果

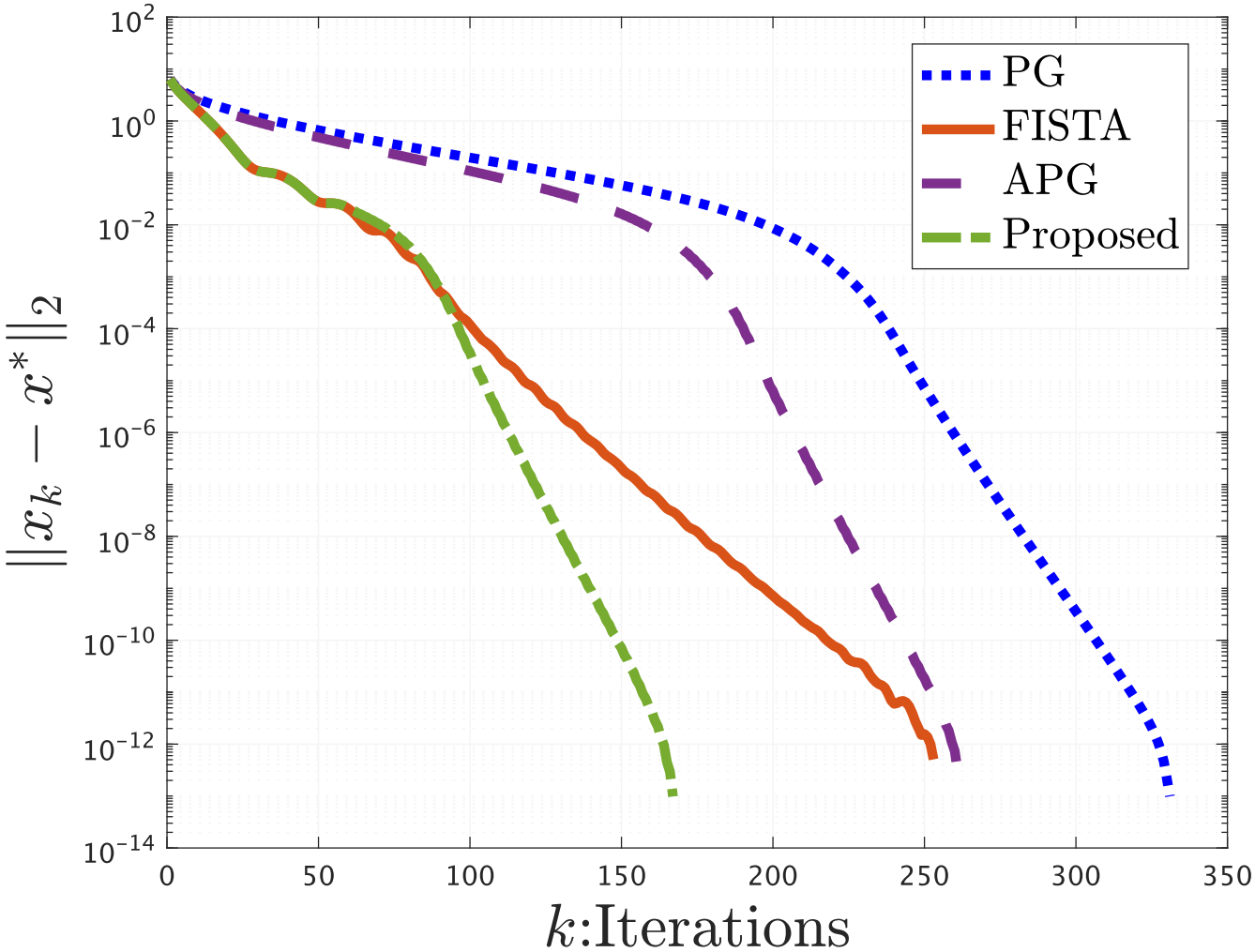

Figure 13: 各繰り返し回数における点列 $\{x_k\}$ と近似解 x^* の差 $\|x_k - x^*\|_2$ の関係 $(m = 1300, n = 800)$

目次

結論と今後の課題

結論

- ラッソ回帰における新たなアルゴリズムを提案した.
- 数値実験の結果，既存のアルゴリズムと比較して，繰り返し回数を以下のように改善することができた.
  - $(m = 130, n = 80)$: 30.9 %
  - $(m = 650, n = 400)$: 36.4 %
  - $(m = 1300, n = 800)$: 34.9 %

今後の課題

- アルゴリズムの切り替えの評価式である $\|x_k - \mathrm{prox}_{\gamma\lambda\|\cdot\|_1}(x_k - \gamma A^T(Ax_k - b))\|_2 < \varepsilon'$ の ε' の効率のよい決定方法を検討すること.

参考文献

[1] H. Attouch and J. Peypouquet, The 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.