

HMLasso に対する最適化手法の検討

Examination of optimization methods for HMLasso

○佐々木大夢*, 松下慎也*, 徐粒*

○Hiromu Sasaki*, Shin-ya Matsushita*, Li Xu*

*秋田県立大学システム科学技術研究科

*Akita Prefectural University

キーワード : HMLasso(Lasso with High Missing rate),

FISTA (fast iterative shrinkage-thresholding algorithm), 交互方向乗数法(ADMM)

連絡先 : 〒015-0055 秋田県由利本荘市土谷字海老ノ口 84-4 秋田県立大学 システム科学技術研究科
総合システム工学専攻 制御工学研究室 佐々木大夢 Tel.:(0184)27-2511,
E-Mail:m25n009@akita-pu.ac.jp

1. はじめに

$X \in \mathbb{R}^{m \times n}$, $y \in \mathbb{R}^m$ とする. 本研究ではラッソ回帰(Lasso)と呼ばれる最適化問題を考える.

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2n} \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \quad (1)$$

ここで, $\|\cdot\|_2$ は $\ell 2$ ノルム, $\|\cdot\|_1$ は $\ell 1$ ノルム, $\lambda > 0$ とする. 適切な λ を設定した(1)を解く事で, 成分のほとんどが0となる解(スパース解)を得ることができる. (1)は曲線フッティングやノイズ処理など広く応用されている^{1,2)}.

この問題でカギとなるのが, 第一項に含まれる行列 X である. 実測で得られたデータを X に代入するが, 測定ミスやセンサの故障などによって, 欠損値が多く含まれる場合, そのまま

(1)を解くことはできない.

この問題に対して先行研究では, (1)に部分問題を追加した HMLasso が提案されている³⁾. HMLasso の詳細は3節で説明する. 先行研究では HMLasso の部分問題を解くために交互方向乗数法(Alternating direction method of multipliers : ADMM)⁴⁾という最適化手法が利用されていた. 一方, ADMM の他にも Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)⁵⁾と呼ばれる最適化手法が知られている. FISTA は HMLasso の部分問題に対しても適応することができる為, ADMM と FISTA の有効性を数値実験によって比較する.

2. 準備

この節では、次節で紹介する HMLasso と最適化問題を解くアルゴリズムを説明するうえで必要な数学的な概念を定義する。本節を通して $x \in \mathbb{R}^n$, $A, B \in \mathbb{R}^{m \times n}$, $C \in \mathbb{R}^{n \times n}$ と仮定する。

2.1 本研究で使われている数学的な記号

Table1 本論文で使用する記号

記号	説明
\mathbb{R}	実数全体
\mathbb{R}^n	n 次元 Euclid 空間
$\mathbb{R}^{m \times n}$	n 行 m 列の行列の集合
A^T	行列 A の転置行列
$C \geq 0$	半正定値行列 C $v^T C v \geq 0$ ($\forall v \in \mathbb{R}^n$)
	ℓ_1 ノルム
$\ \cdot\ _1$	$\ x\ _1 = \sum_{i=1}^n x_i $
	フロベニウスノルム
$\ \cdot\ _F$	$\ A\ _F = \sqrt{\sum_{j,k} a_{jk}^2}$
	同じ要素同士の掛け算
$A \odot B$	$\begin{pmatrix} a_{11}b_{11} & \cdots & a_{1n}b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}b_{m1} & \cdots & a_{mn}b_{mn} \end{pmatrix}$
	同じ要素同士の割り算
$A \oslash B$	$\begin{pmatrix} \frac{a_{11}}{b_{11}} & \cdots & \frac{a_{1n}}{b_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{a_{m1}}{b_{m1}} & \cdots & \frac{a_{mn}}{b_{mn}} \end{pmatrix}$
PSD	半正定値行列の集合
P_{PSD}	PSD の上への射影

2.2 欠損値を含んだラッソ回帰問題

(1)は以下のように表現できる。

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \beta^T S \beta + \rho^T y + \lambda \|\beta\|_1 \quad (2)$$

ここで、 $S = \frac{1}{n} X^T X$, $\rho = \frac{1}{n} X^T y$ とする。定義より S は半正定値となる。行列 X に欠損が含まれる場合、式(2)の S , ρ を、それぞれ以下で定義される S^{pair} , ρ^{pair} に変更する。

$$S^{pair} = \frac{1}{n_{jk}} \sum_{i \in I_{jk}} X_{ij} X_{ik} \quad (3)$$

$$\rho^{pair} = \frac{1}{n_j} \sum_{i \in I_j} X_{ij} y_i \quad (4)$$

ここで、 I_{jk} は、 X の各 j 列と k 列の両方が欠損していない行の添字集合を示しており、 n_{jk} は集合 I_{jk} の要素の個数を表している。しかし、 S^{pair} が半正定値行列とは限らない為、問題(2)の解の存在が保証されないという問題が生じる。

3. HMLasso

上記の問題を解決するため、先行研究では以下の HMLasso が提案された³⁾。

$$\min_{\beta \in \mathbb{R}^n} \frac{1}{2} \beta^T \hat{\sigma} \beta + \rho^{pair T} y + \lambda \|\beta\|_1 \quad (5)$$

$$\hat{\sigma} = \arg \min_{\sigma} \|W \odot (\sigma - S^{pair})\|_F^2 \quad s.t. \sigma \geq 0 \quad (6)$$

ここで、 $W_{jk} = \frac{n_{jk}}{n}$ である。HMLasso は欠損率が高いデータに対しても精度の高い解を求めることが出来る。先行研究では部分問題(6)に対して ADMM が適用されていたが、FISTA 等の他の最適化手法を適用した場合の有効性の検証が行われていなかった。

4. ADMM と FISTA

この節では、部分問題(6)の解を近似する2つの最適化手法を紹介する。

問題(6)に対する ADMM は Algorithm1 で与えられる。

$$\hat{D} = Z\lambda Z^T \quad (10)$$

Algorithm1 ADMM⁴⁾

Input : $S^{pair}, W, \mu, B_1, C_1$

for $k = 1, 2, \dots$ **do**

$$\begin{aligned} A_{k+1} &= P_{PSD}(B_k + S^{pair} - \mu C_k) \\ B_{k+1} &= (A_{k+1} - S^{pair} + \mu C_k) \\ &\quad \odot (\mu W \odot W + I) \\ C_{k+1} &= C_k + \frac{1}{\mu}(A_{k+1} - B_{k+1} - S^{pair}) \end{aligned}$$

end for

Output : A_k

FISTA は Algorithm2 で与えられる.

Algorithm2 FISTA⁵⁾

Input : $S^{pair}, W, \mu, A_1, B_1$

for $k = 1, 2, \dots$ **do**

$$\begin{aligned} A_{k+1} &= P_{PSD}(B_k - \mu W \odot W \odot (A_k - S^{pair})) \\ t_{k+1} &= 1 + \frac{1 + \sqrt{1 + 4t_k^2}}{2} \\ B_{k+1} &= A_{k+1} + \left(\frac{t_k - 1}{t_{k+1}}\right)(A_{k+1} - A_k) \end{aligned}$$

end for

Output : A_k

Algorithm1 及び Algorithm2 では, P_{PSD} (半正定値の集合(PSD)へ射影)が必要となる. $D \in \mathbb{R}^{n \times n}$ に対して, $\hat{D} = P_{PSD}(D)$ とおく. この時, \hat{D} は次の (7) から (10) を計算することで得られる⁶⁾.

$$F = \frac{(D + D^T)}{2} \quad (7)$$

$$F = Z\lambda Z^T \quad (8)$$

式(8)の右辺は行列 F の固有値分解を表しており, λ は行列 F の固有値を対角成分, それ以外に 0 を代入した行列, Z は直交行列を表している. また $\hat{\lambda}_i$ ($i = 1, 2, \dots, n$) を次のように設定する.

$$\hat{\lambda}_i = \begin{cases} \lambda_i & \lambda_i \geq 0 \\ 0 & \lambda_i < 0 \end{cases} \quad (9)$$

行列 $\hat{\lambda}$ は対角成分が $\hat{\lambda}_i$, それ以外に 0 を代入した行列とし, 以下の式(10)の右辺の行列を構成すると \hat{D} が得られる. 即ち

5. 数値実験

この説では, 問題(6)の部分問題に対して, 交互方向乗数法と FISTA を適用させ結果を比較する. 数値実験で用いた計算環境を Table2 に示す.

Table2 数値実験で用いた計算環境

OS	Windows10
メモリ	8.00 GB
CPU	Intel(R) Core(TM) i5-7500
使用ソフト	MATLAB R2023a

今回考える問題設定は, 以下のとおりである. $X \in \mathbb{R}^{10000 \times 100}$ とする. ここで X は正規分布に従うように生成する. この行列 X をランダムに全体の 50% 欠損させて, 式(4)の S^{pair} を求める. 求めた S^{pair} を交互方向乗数法(ADMM)と FISTA を適用し, 実行回数ごとに近似解との差分の総和を求めた. これを, 各アルゴリズム 10 回初期点を変え同じ数値実験をし, 平均を求めグラフにする. 初期点は, ランダムな実数としている.

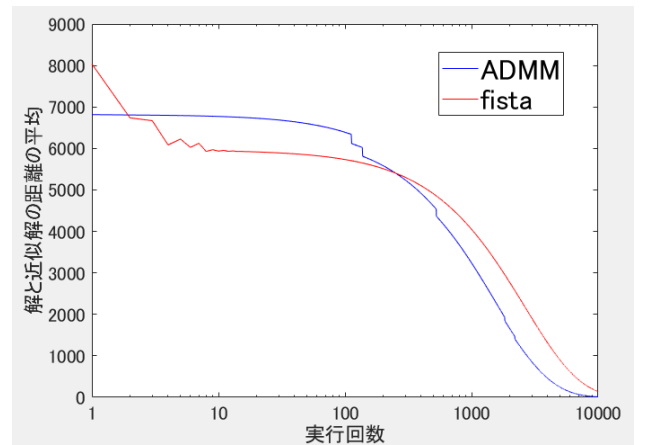


Fig.1 部分問題(13)に対して, 交互方向乗数法(ADMM), FISTA を実行, 解と近似解の距離の平均を比較した結果

近似解にどれだけ値を近づけることができるかを確認するために、試行回数は 10000 回としている。差分として用いる近似解は、CVX*という凸最適化問題を解くソフトウェアを用いて求めた値を使用している。結果を Fig.1 で示す。

Fig.1 から、10000 回 ADMM と FISTA を行ったところ、どちらも近似解に収束することが確認できる。収束の速さは、ADMM のほうが速く近似解に収束することがわかった。それぞれの手法を 10000 回実行した平均時間比較した結果を Table3 に示す。こちらも ADMM の方が速いことがわかった。

Table3 10000 回行った際の実行時間の平均

Algorithm	CPU time[s]
ADMM	28.09
FISTA	31.45

6. 結論

本実験では、数値実験でそれぞれのアルゴリズムの有効性を比較することで、交互方向乗数法(ADMM)の方が早く、近似解に収束していることがわかった。

今後の課題は 2 点ある。1 点目は、HMLasso は 2 つの問題を解くことで、解を得ることができるため、計算時間などを考慮するとそれぞれがもっと早く近似解にたどりつく必要がある。式(6)を解くアルゴリズムはほかにもあるため、他のアルゴリズムを検討し、さらに効率の良い手法を見つけることである。2 点目は、HMLasso を工場などの実問題に応用させ、有効性を検証することである。

参考文献

1) 永原正章:スパースモデリングー基礎から動的システムの応用ー, 15-71, コロナ社 (2017)

2) 鈴木讓:スパース推定 100 問 with Python, 6-13,共立出版 (2021)

3) M. Takada, H. Fujisawa and T. Nishikawa: HMLasso: Lasso with High Missing Rate, IJCAI, 3541-3547(2019)

4) S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein: Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 1-122(2011)

5) A. Beck and M. Teboulle: A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imaging Sciences, 183-202 (2009)

6) R. Escalante and M. Raydan: Alternating Projection Methods,78-79, Society for Industrial and Applied Mathematics (2011)