

HMLasso に対する最適化手法に関する研究 *

佐々木大夢¹, 松下慎也²

¹ 秋田県立大学大学院 システム科学技術研究科

Graduate School of Systems Science and Technology

Akita Prefectural University

² 秋田県立大学 システム科学技術学部

Faculty of Systems Science and Technology

Akita Prefectural University

1 背景

複数の項目を持つデータ (ここで扱うデータは $p + 1$ 個の項目を持ち, 各項目は n 個の成分を持つとする) に対して, 興味のある項目のデータを目的変数 $\mathbf{y} \in \mathbb{R}^n$, それ以外の複数項目のデータを説明変数 $X \in \mathbb{R}^{n \times p}$ に設定し, 次の連立一次方程式系を考える.

$$\mathbf{y} = X\boldsymbol{\beta} \quad (1.1)$$

方程式 (1.1) の解を推定する問題は, X が $\mathbf{y}$ に与える影響度を理解するために有用である. 方程式 (1.1) は解が存在しない場合があるが, 残差を最小にする問題 (つまり, $\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \|\mathbf{y} - X\boldsymbol{\beta}\|_2$) を解くことでその近似解を見つける事ができる.

Lasso (least absolute shrinkage and selection operator) は, 残差をなるべく小さくし, 求める解の非零成分が少ない (スパースである) 解を推定する事が期待できる. Lasso では次の問題を考える.

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \frac{1}{2n} \|\mathbf{y} - X\boldsymbol{\beta}\|_2^2 + \alpha \|\boldsymbol{\beta}\|_1 \quad (1.2)$$

ここで, $\|\cdot\|_1$ は ℓ^1 ノルム, $\alpha > 0$ は正則化パラメータと呼ばれ, 交差検証などを用いて決定する. 問題 (1.2) を解くことで, スパースな解を見つけることができる. 一方, 現実の問題で扱われるデータは高次元でかつ欠損を多く含むという特徴を持つが (例えば, [6]), 説明変数 X に欠損値が含まれていると Lasso は解くことができないため, 欠損を平均値や中央値で補完する方法などが利用される.

* This work was supported by the Research Institute for Mathematical Sciences, an International Joint Usage/Research Center located in Kyoto University

最近, Takada 等 [5] によって欠損を含むデータから高精度な解を推定できる HMLasso (Lasso with high missing rate) が提案された. HMLasso では, 以下の問題 (1.3) 及び問題 (1.4) を解くことで所望の解を推定する.

$$\min_{\Sigma \succeq 0} \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2 \quad (1.3)$$

$$\min_{\beta} \frac{1}{2} \beta^\top \hat{\Sigma} \beta - \rho^{\text{pair}^\top} \beta + \alpha \|\beta\|_1 \quad (1.4)$$

ここで, $S^{\text{pair}}, W \in \mathbb{R}^{p \times p}$, $\rho^{\text{pair}} \in \mathbb{R}^p$ (詳細は 2 章で解説する), $\|\cdot\|_F$ は Frobenius ノルム, $\odot$ はアダマール積を表す. X が欠損を含む場合, 部分問題 (1.3) を解いて得られた解を主問題 (1.4) の $\hat{\Sigma}$ に代入して (1.4) を解く. HMLasso は二つの最適化問題を解く必要があるが, 主問題 (1.4) には座標降下法, 近接勾配法やその加速化法など, 解を高速に近似する最適化手法がこれまでに提案されている. 一方, 先行研究 [5] では部分問題 (1.3) を解くために ADMM (Alternating Direction Method of Multipliers [3]) が適用されていた.

本論文では, 部分問題 (1.3) に対する最適化手法について検討する. 具体的には, (1.3) の目的関数が持つ性質について検討し, 得られた知見を駆使することで, ADMM とは異なる最適化手法を適用し, その有効性の検証を行う.

2 準備

2.1 ノルム

$A, B \in \mathbb{R}^{n \times p}$ とする. A の Frobenius ノルム $\|A\|_F$ 及び A と B の Hadamard 積 $A \odot B$ をそれぞれ次のように定義する.

$$\|A\|_F := \sqrt{\sum_{j,k} A_{jk}^2}, \quad (A \odot B)_{jk} := A_{jk} B_{jk} \quad (2.1)$$

$\mathbb{S}^p$ を p 次対称行列の集合とし, $C \in \mathbb{S}^p$ とする. 任意の $\mathbf{x} \in \mathbb{R}^p$ に対して $\mathbf{x}^\top C \mathbf{x} \geq 0$ が成り立つとき C を半正定値といい, $C \succeq 0$ と書く. また, 半正定値対称行列の集合を $\mathbb{S}_+^p$ と書く.

2.2 関数

$\mathcal{D} \subset \mathbb{S}^p$ とする. 集合 $\mathcal{D}$ に対する指示関数を以下で定義する.

$$I_{\mathcal{D}}(\Sigma) := \begin{cases} 0 & (\Sigma \in \mathcal{D}); \\ \infty & (\Sigma \notin \mathcal{D}). \end{cases} \quad (2.2)$$

$f: \mathbb{S}^p \rightarrow (-\infty, \infty]$ を下半連続な真凸関数とする. f に対する近接写像 prox_f を次のように定義する.

$$\text{prox}_f(\Sigma) := \underset{W \in \mathbb{S}^p}{\operatorname{argmin}} \left\{ f(W) + \frac{1}{2} \|W - \Sigma\|_F^2 \right\}.$$

集合 $\mathbb{S}_+^p$ の指示関数の近接写像について, 以下が成り立つ.

命題 2.1. ([4])

$$\text{prox}_{I_{\mathbb{S}_+^p}}(\Sigma) = P_{\mathbb{S}_+^p}(\Sigma) \quad (2.3)$$

但し, $P_{\mathbb{S}_+^p}$ は集合 $\mathbb{S}_+^p$ の上への距離射影であり, 次のように定義する.

$$P_{\mathbb{S}_+^p}(\Sigma) := \begin{cases} \Sigma & (\Sigma \in \mathbb{S}_+^p) \\ Z\hat{\Lambda}Z^\top & (\Sigma \notin \mathbb{S}_+^p) \end{cases}$$

ここで, $F := \frac{(\Sigma + \Sigma^\top)}{2}$ とし, Z は F の固有値分解によって得られる直交行列, Λ は F の固有値を成分に持つ対角行列とする. $\hat{\Lambda}$ の各成分は Λ を用いて以下のように定義する.

$$\hat{\lambda}_{ii} := \begin{cases} \lambda_{ii} & (\lambda_{ii} \geq 0) \\ 0 & (\lambda_{ii} < 0) \end{cases} \quad (1 \leq i \leq p)$$

2.3 HMLasso

I_{jk} は行列 X の第 j, k 列のうち, x_{ij} と x_{ik} が欠損していない行番号の添字全体の集合, つまり

$$I_{jk} := \{i \mid x_{ij} \text{ と } x_{ik} \text{ は欠損していない}\}.$$

また, n_{jk} は I_{jk} の要素の個数とする. I_{jk} と n_{jk} を用いて, S^{pair} と W の各成分をそれぞれ以下のように定義する.

$$(S^{\text{pair}})_{jk} := \begin{cases} \frac{1}{n_{jk}} \sum_{i \in I_{jk}} x_{ij}x_{ik} & (I_{jk} \neq \emptyset) \\ 0 & (I_{jk} = \emptyset) \end{cases}, \quad (W)_{jk} := \frac{n_{jk}}{n}$$

3 HMLasso の部分問題について

本論文では, 次の問題を考える.

$$\min_{\Sigma \succeq 0} \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2. \quad (3.1)$$

先行研究 [5] では, 問題 (3.1) を解くための最適化手法として ADMM [3] が適用されていた.

問題 (3.1) に対する ADMM の有効性を具体例を使って確認する. X の欠損率を変化させた部分問題 (3.1) に対する ADMM の収束の様子を図 1 に示す. 図 1 の横軸は繰り返し回数, 縦軸は CVX*¹ で求めた部分問題の解と ADMM で生成した近似列との距離である. 図 1 より X の欠損率が高いほど近似列が解に収束する速さが遅くなる. 先行研究 [5] では ADMM 以外の最適化手法を用いた検討は行われていなかった. 本研究では部分問題 (3.1) の目的関数の持つ性質を検討し, 部分問題 (3.1) を高速に求解するための最適化手法について検討する.

*1 <http://cvxr.com/cvx/>

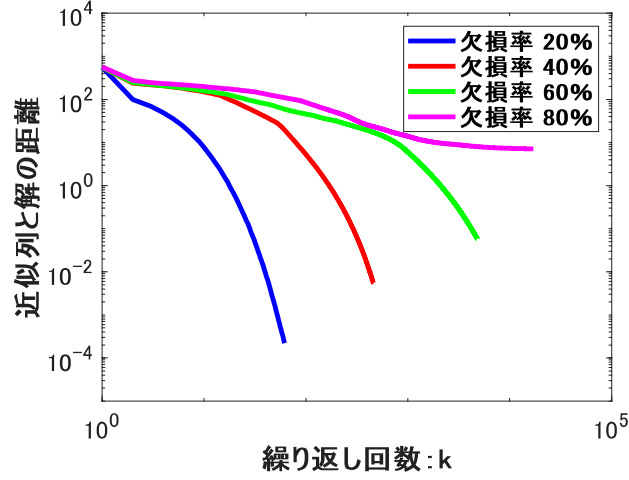


図1 X の欠損率を変化させた時の ADMM の収束

3.1 部分問題の再定式化

部分問題 (3.1) は指示関数を使うことで以下の制約無しの最適化問題となる.

$$\min_{\Sigma \in \mathbb{S}^p} \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2 + I_{\mathbb{S}_+^p}(\Sigma) \quad (3.2)$$

ここで,

$$f(\Sigma) := \frac{1}{2} \|W \odot (\Sigma - S^{\text{pair}})\|_F^2 \quad (3.3)$$

$$g(\Sigma) := I_{\mathbb{S}_+^p}(\Sigma) \quad (3.4)$$

とおくことで, 問題 (3.2) は2つの凸関数 f と g の和の最小化問題となる. 関数 f は次の性質を持つ.

命題 3.1. (3.3) で定義された関数 f は微分可能で, その勾配 ∇f は

$$\nabla f(\Sigma) = W \odot W \odot (\Sigma - S^{\text{pair}}) \quad (3.5)$$

となる. また, ∇f は以下の性質を持つ.

- (リプシッツ連続性) 任意の $\Sigma_i \in \mathbb{S}^p$ ($i = 1, 2$) に対して次が成り立つ.

$$\|\nabla f(\Sigma_1) - \nabla f(\Sigma_2)\|_F \leq \|W \odot W\|_F \|\Sigma_1 - \Sigma_2\|_F \quad (3.6)$$

このとき, ∇f は $\|W \odot W\|_F$ -リプシッツ連続であるという.

- (強凸性) 任意の $\Sigma_i \in \mathbb{S}^p$ ($i = 1, 2$) に対して次が成り立つ.

$$\langle \nabla f(\Sigma_1) - \nabla f(\Sigma_2), \Sigma_1 - \Sigma_2 \rangle \geq \min_{j,k} (W \odot W)_{jk} \|\Sigma_1 - \Sigma_2\|_F^2 \quad (3.7)$$

このとき, f は $\min_{j,k} (W \odot W)_{jk}$ -強凸関数という.

3.2 最適化手法について

2つの凸関数の和の最小化問題について, これまでにさまざまな最適化手法が開発されている [1]. 次の問題を考える.

$$\min_{\Sigma \in \mathbb{S}^p} f(\Sigma) + g(\Sigma) \quad (3.8)$$

ここで, 関数 $f: \mathbb{S}^p \rightarrow \mathbb{R}$ は μ -強凸関数でその勾配 ∇f は L -リプシッツ連続, 関数 $g: \mathbb{S}^p \rightarrow \mathbb{R}$ は凸関数とする. 最適化問題 (3.8) に対する次の最適化手法について考える.

Algorithm 1 問題 (3.8) に対する最適化手法 [1]

Require: $\Sigma_1, \Phi_1 \in \mathbb{R}^{p \times p}, t_1 = 0, q = \frac{\mu}{L} (L > 0, \mu \geq 0)$

```

1: for  $k = 1, 2, \dots$  do
2:    $t_{k+1} = \frac{2t_k + 1 + \sqrt{4t_k + 4qt_k^2 + 1}}{2(1 - q)}$ 
3:    $\tau_k = \frac{(t_{k+1} - t_k)(1 + qt_k)}{t_{k+1} + 2qt_k t_{k+1} - qt_k^2}$ 
4:    $\delta_k = \frac{t_{k+1} - t_k}{1 + qt_{k+1}}$ 
5:    $\Psi_k = \Sigma_k + \tau_k(\Phi_k - \Sigma_k)$ 
6:    $\Sigma_{k+1} = \text{prox}_{(1/L)g}(\Psi_k - \frac{1}{L}\nabla f(\Psi_k))$ 
7:    $\Phi_{k+1} = (1 - q\delta_k)\Phi_k + q\delta_k\Psi_k + \delta_k(\Sigma_{k+1} - \Psi_k)$ 
8: end for
```

Ensure: Σ_{k+1}

注意 3.1.

- Algorithm 1 は Beck と Teboulle [2] によって開発された近接勾配法の加速化法に基づく最適化手法である.
- $F(\Sigma) := f(\Sigma) + g(\Sigma)$, $F^* := \argmin_{\Sigma \in \mathbb{S}^p} F(\Sigma)$ とおく. また Algorithm1 によって構成された近似列を $\{\Sigma_k\}$ とする. このとき, $F(\Sigma_k) \rightarrow F^* (k \rightarrow \infty)$ が成り立つ [1].

ここで特に, 問題 (3.8) の関数 f と g をそれぞれ (3.3) と (3.4) とおく. 命題 3.1 から ∇f のリプシッツ定数 L と関数 f の強凸性の定数 μ はそれぞれ次のようになる.

$$L = \|W \odot W\|_F, \quad \mu = \min_{j,k} (W \odot W)_{jk}. \quad (3.9)$$

また, Algorithm 1 で使用する近接写像 prox_g は $\mathcal{S}_+^p$ の上への距離射影 ((2.3) 参照), f の勾配 ∇f は (3.5) を使用する.

4 数値実験

問題 (3.8) に対して, Algorithm 1 を適用してその有効性を検証する. 実験環境は以下の通りである.

表 1 実験環境

OS	Windows11
CPU	12th Gen Intel(R) Core(TM) i5-12400
メモリ	8.00GB
プログラム言語	MATLAB(R2023b)

数値実験では文献 [5] で使用されていた建築データセット*²を利用する. $X \in \mathbb{R}^{300 \times 27}$ であり, X を 20%, 40%, 60%, 80% の割合で欠損させて数値実験を行った. 初期点は標準正規分布に従った乱数 10 点を使用した. グラフの横軸を繰り返し回数, 縦軸を近似列 $\{\Sigma_k\}$ と近似解 Σ_{CVX} の距離 $\|\Sigma_k - \Sigma_{\text{CVX}}\|_F$ とした. 但し, グラフの縦軸は異なる初期点で構成した近似列と近似解との距離の平均値とした. 繰り返しの終了条件は $\|\Sigma_{k+1} - \Sigma_k\|_F \leq 10^{-6}$ とした. また, 近似解 Σ_{CVX} は CVX*³を使用して求めた解を利用した. Algorithm1 の結果を”SFPG”, ADMM の結果を”先行研究”と表記する. 以下に実験結果を示す.

*² <https://archive.ics.uci.edu/ml/datasets/Residential+Building+Data+Set>

*³ <http://cvxr.com/cvx/>

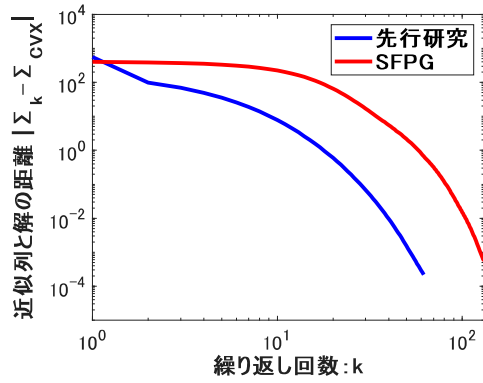


図2 欠損率 20% データに対する収束の様子

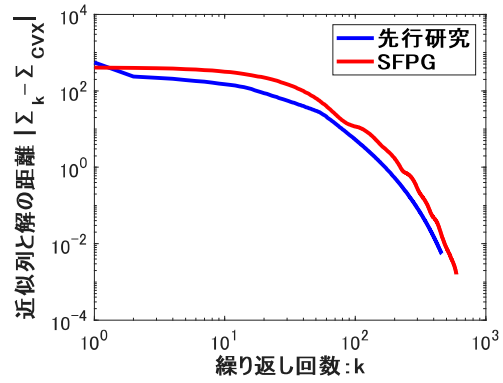


図3 欠損率 40% データに対する収束の様子

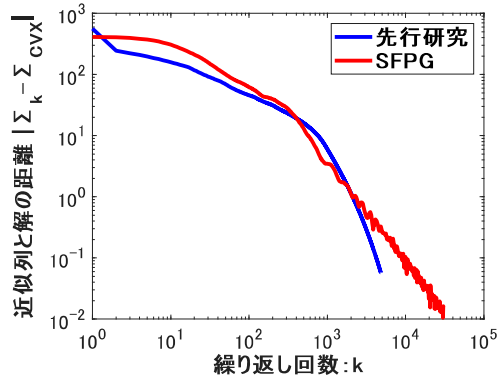


図4 欠損率 60% データに対する収束の様子

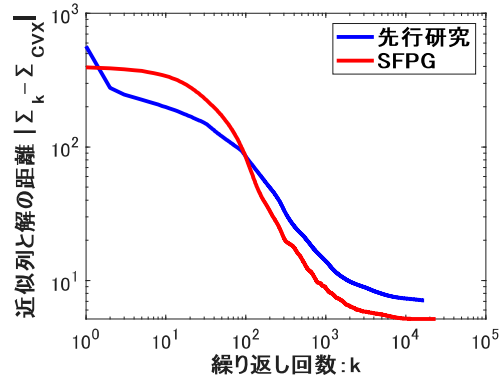


図5 欠損率 80% データに対する収束の様子

欠損率 20% では先行研究の方が速く解に収束し、その他の欠損率 (40%, 60%, 80%) において、Algorithm 1 と ADMM は同程度の速さで解に収束している。

5 結論

本研究では HMLasso の部分問題 (1.3) の目的関数について検討した。目的関数が強凸で、その勾配がリプシッツ連続であることを利用した Algorithm1 を適用し、数値実験によりその有効性を確認した。今回採用した Algorithm1 は先行研究の手法 (ADMM) と比較して収束の速さを改善するには至らなかった。高速に解を近似できる最適化手法についてさらに検討を行いたい。

また、本研究に取り組む過程で欠損値を含む実データが身近にあることを知った (例えば、学生の満足度調査、進学アンケート等)。欠損値を含む実データに HMLasso を適用して、データから有用な情報を抽出することで活用していきたい。

謝辞

HMLasso の研究を山口高康先生 (秋田県立大学) に紹介していただきました。本研究に取り組む機会を与えていただき、深く感謝申し上げます。

参考文献

- [1] A. d’Aspremont, D Scieur, and A Taylor, Acceleration Methods, Now Publishers 2021
- [2] A. Beck and M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems, SIAM J. Imag. Sci., 2 (2009) 183-202.
- [3] 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, 3, (2011) 1-122.
- [4] R. Escalante and M. Raydan, Alternating Projection Methods, SIAM (2011).
- [5] M. Takada, H. Fujisawa, and T. Nishikawa, HMLasso: Lasso with High Missing Rate, In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI2019), (2019) 3541-3547.
- [6] 高田正彬, 西川武一郎, 大量の欠損を含む製造データから不良要因を同定するスパースモデリング技術, 東芝レビュー, 76, (2021) 32-35.