

多重共線性を考慮した回帰式の変数選択 — 混合整数半正定値計画法を用いた解法 —

小林健 (Ken Kobayashi) * 高野祐一 (Yuichi Takano) †

宮代隆平 (Ryuhei Miyashiro) ‡ 中田和秀 (Kazuhide Nakata) *

概要

本論文では回帰分析における変数選択問題を扱う。回帰分析において、説明変数間に一次従属の関係がある場合、モデル推定の結果の信頼性が損なわれる。この現象を多重共線性といい、回帰分析では多重共線性が生じないように適切に変数を選択することが重要となる。そこで本論文では、多重共線性が生じないという条件のもとモデル推定に最適な変数を選択する問題を考え、この問題を混合整数半正定値計画問題として定式化できることを示す。さらに、本論文では定式化した問題に対して切除平面法による解法を提案する。

1 はじめに

本論文では、線形回帰モデルの説明変数を選択する問題を扱う。説明変数間に強い相関があったり、一次従属な関係がある場合、数値誤差の影響を強く受けて推定量の信頼度が低下することがある。この現象を多重共線性といい、回帰分析では多重共線性を回避するよう適切に説明変数を選択してモデル推定を行うことが望まれる。

回帰分析において候補となる変数の中から有用な説明変数集合を選び出す問題を変数選択問題という。変数選択を行う利点としては、上で述べたような多重共線性の問題を回避できることのみならず、過剰適合を抑制して予測精度の向上につながることや、回帰分析の結果の解釈が容易になること、モデル推定に要する計算時間を削減できることなどがあげられる [7]。このように様々な利点があるため、統計分野において回帰式の変数選択問題は古くから重要な課題として知られている。

変数選択のための手法は、これまで数多くのもものが提案されているが、その大部分はヒューリスティクスとみなすことができる。その中でも最も一般的な方法はステップワイズ法 [5] であり、回帰式の適合度指標や説明変数の有意性に基づいて変数を 1 つずつ追加/削除することを繰り返す方法である。この手法は計算が高速であるという利点はあるものの、局所探索型のヒューリスティクスであるため、適合度指標の意味で最適なモデルを出力するとは限らないと

* 東京工業大学 大学院社会理工研究科 (Graduate School of Decision Science and Technology, Tokyo Institute of Technology)

† 専修大学 ネットワーク情報学部 (School of Network and Information, Senshu University)

‡ 東京農工大学 大学院工学研究院 (Institute of Engineering, Tokyo University of Agriculture and Technology)

いう欠点がある。

一方で厳密解法としては、総当り法 [6] や整数計画法による解法 [3, 10, 11] が提案されている。特に近年では計算機や整数計画ソルバーの性能が飛躍的に向上したことにともない、整数計画法を用いた解法が注目を集めている。整数計画法を用いた近年の研究としては、AIC などの情報量基準や, Mallows の C_p 基準に基づく変数選択問題が混合整数計画問題として定式化されている [12, 13]。しかしこれらの適合度指標は予測モデルの汎化性能を評価する指標であるため、これらの指標に基づく変数選択を行ったとしても、必ずしも多重共線性を除去できるとは限らない。

そこで本論文では、多重共線性が生じないという条件のもとモデル推定に最適な変数を選択する問題に対して、混合整数計画法を用いた解法を提案する。多重共線性を検出する指標の 1 つとして、説明変数の相関係数行列の条件数がある [4]。本論文ではこの指標に着目し、条件数の制約のもと最小 2 乗基準に基づいて最適な説明変数を選択する問題を混合整数半正定値計画問題 (Mixed Integer Semidefinite Program; MISDP) として定式化する。定式化した MISDP は直接解くことが困難な問題であるため、本論文では Konno, Gotoh, Uno and Yuki [9] が提案した SDP を解くための切除平面法を拡張し、混合整数 2 次計画問題 (Mixed Integer Quadratic Program; MIQP) を繰り返し解くことによって MISDP を解く切除平面法を提案する。

本論文の構成は以下のとおりである: 2 節では本論文で扱う線形回帰モデルの最小 2 乗推定と多重共線性の問題について説明する。3 節では多重共線性を考慮した変数選択問題を MISDP として定式化し、4 節では定式化した MISDP を解く切除平面法のアルゴリズムを示す。その後、5 節では計算機実験によって提案手法の有用性を確認し、6 節で結論及び今後の課題について述べる。

2 最小 2 乗推定と多重共線性

2.1 最小 2 乗推定

n 個のサンプル $\{(y_i; x_{i1}, x_{i2}, \dots, x_{ip}) \mid i = 1, 2, \dots, n\}$ が与えられ、 y_i を被説明変数、 x_{ij} ($j = 1, 2, \dots, p$) を p 種類の説明変数とする。ただし、ここで被説明変数と各説明変数は、平均 0、分散 1 に正規化してあるものとする。本論文では、以下のような切片を含まない線形モデル*1で被説明変数を予測することを考える:

$$y_i = a_1 x_{i1} + a_2 x_{i2} + \dots + a_p x_{ip} + \varepsilon_i \quad (i = 1, 2, \dots, n). \quad (1)$$

*1 被説明変数および説明変数を正規化して、切片を含む線形回帰モデルの最小 2 乗推定を行った場合、切片は 0 になることが知られている。したがって本論文において切片を含まない線形回帰モデルを対象とすることは一般性を失わない。

ここで, a_j は j 番目の説明変数の偏回帰係数, ε_i は i 番目のサンプルに対する予測残差とする.
いま,

$$\mathbf{y} := (y_1, y_2, \dots, y_n)^\top, \mathbf{a} := (a_1, a_2, \dots, a_p)^\top, \boldsymbol{\varepsilon} := (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top,$$

$$\mathbf{X} := \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}$$

と記号を定義すると, 線形回帰モデル (1) は以下のように書き直せる:

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \boldsymbol{\varepsilon}. \quad (2)$$

このとき残差 2 乗和は

$$\sum_{i=1}^n \varepsilon_i^2 = \boldsymbol{\varepsilon}^\top \boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\mathbf{a})^\top (\mathbf{y} - \mathbf{X}\mathbf{a}) = \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\mathbf{a} + \mathbf{a}^\top \mathbf{X}^\top \mathbf{X}\mathbf{a} \quad (3)$$

と表せる. したがって残差 2 乗和を最小化する最小 2 乗推定量を $\hat{\mathbf{a}}$ とすると, $\hat{\mathbf{a}}$ は

$$\nabla_{\mathbf{a}} (\mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\mathbf{a} + \mathbf{a}^\top \mathbf{X}^\top \mathbf{X}\mathbf{a}) \Big|_{\mathbf{a}=\hat{\mathbf{a}}} = -2\mathbf{X}^\top \mathbf{y} + 2\mathbf{X}^\top \mathbf{X}\hat{\mathbf{a}} = \mathbf{0}$$

を満たす. よって最小 2 乗推定量 $\hat{\mathbf{a}}$ を求めるためには, 以下の線形方程式を解けばよい:

$$\mathbf{X}^\top \mathbf{X}\hat{\mathbf{a}} = \mathbf{X}^\top \mathbf{y}. \quad (4)$$

なお, ここでは説明変数を事前に正規化しているため, 説明変数の相関係数行列を

$$\mathbf{R} := \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{pmatrix} \quad (r_{j\ell} \text{ は } j \text{ 番目の説明変数と } \ell \text{ 番目の説明変数の相関係数})$$

と表すと,

$$\mathbf{R} = \frac{1}{n} \mathbf{X}^\top \mathbf{X} \quad (5)$$

が成り立つことに注意する.

2.2 多重共線性と条件数

最小 2 乗推定量 $\hat{\mathbf{a}}$ を求めるためには, 線形方程式 (4) を解けばよい. しかし, 説明変数間に一次従属の関係が存在する場合, この線形方程式が悪条件となり, 数値誤差の影響を強く受け推定量の信頼性が低下する. この現象を多重共線性という.

Chatterjee and Hadi [4] は多重共線性を検出する指標の 1 つとして、相関係数行列 \mathbf{R} の条件数をあげている。相関係数行列 \mathbf{R} の条件数 $\text{cond}(\mathbf{R})$ は、 \mathbf{R} の最大固有値と最小固有値の比 $\lambda_{\max}(\mathbf{R})/\lambda_{\min}(\mathbf{R})$ で表される。条件数は行列の数値的扱いやすさを表す標準的な指標であり、条件数が小さく 1 に近い値であるときその行列は良条件であるといい、条件数が非常に大きいとき悪条件であるという。式 (5) から $\text{cond}(\mathbf{R}) = \text{cond}(\mathbf{X}^\top \mathbf{X})$ が成り立つことに注意すると、相関係数行列の条件数 $\text{cond}(\mathbf{R})$ が大きいほど線形方程式 (4) の係数行列が悪条件となる。したがって信頼度の高い最小 2 乗推定量を求めるためには、相関係数行列の条件数が大きくなるないように説明変数を選ぶことが望ましい。

そこで次節では、選択した変数の相関係数行列の条件数に上限を設け、残差 2 乗和を最小化する変数選択問題を混合整数半正定値計画問題として定式化する。

3 条件数制約つき変数選択問題

ここでは、相関係数行列の条件数に上限を課した変数選択問題 (条件数制約つき変数選択問題) を定式化する。まず、 j 番目の説明変数を選択しない/するを表す 0-1 決定変数

$$z_j \in \{0, 1\} \quad (j = 1, 2, \dots, p) \quad (6)$$

を導入する。そして、 $z_j = 0$ の場合、以下の制約により説明変数を回帰式から取り除く：

$$z_j = 0 \Rightarrow a_j = 0 \quad (j = 1, 2, \dots, p). \quad (7)$$

この制約は、big- M 法や SOS 制約などにより表現できる。

続いて、選択した説明変数の相関係数行列について、条件数の制約を半正定値制約として表すことを考える。以下では、実対称行列 \mathbf{Q} が半正定値行列であることを $\mathbf{Q} \succeq \mathbf{O}$ と表し、 $\mathbf{P} - \mathbf{Q} \succeq \mathbf{O}$ であるとき $\mathbf{P} \succeq \mathbf{Q}$ と表す。実対称行列 \mathbf{Q} の最大固有値が α 以下かつ最小固有値が β 以上であることは、以下と同値であると知られている [14, 15]:

$$\alpha \mathbf{I} \succeq \mathbf{Q} \succeq \beta \mathbf{I}. \quad (8)$$

ただし、 \mathbf{I} は適当なサイズの単位行列とする。相関係数行列 \mathbf{R} に対し、 $\mathbf{z} = (z_1, z_2, \dots, z_p)^\top \in \{0, 1\}^p$ で選択された説明変数の相関係数行列 (\mathbf{R} の主小行列) を $\mathbf{R}(\mathbf{z})$ と表すことにすると、ある定数 $\kappa > 1$ に対して、

$$\begin{aligned} \kappa \geq \text{cond}(\mathbf{R}(\mathbf{z})) &\iff \kappa \lambda_{\min}(\mathbf{R}(\mathbf{z})) \geq \lambda_{\max}(\mathbf{R}(\mathbf{z})), \\ &\iff \exists \lambda \geq 0, \kappa \lambda \mathbf{I} \succeq \mathbf{R}(\mathbf{z}) \succeq \lambda \mathbf{I} \end{aligned}$$

が成り立つ。したがって、説明変数の相関係数行列の条件数が κ 以下となる制約は、次のよう

に表せることがわかる:

$$\underbrace{\kappa\lambda \begin{pmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_p \end{pmatrix}}_{\kappa\lambda \text{Diag}(\mathbf{z})} \succeq \underbrace{\begin{pmatrix} r_{11}z_1z_1 & r_{12}z_1z_2 & \cdots & r_{1p}z_1z_p \\ r_{21}z_2z_1 & r_{22}z_2z_2 & \cdots & r_{2p}z_2z_p \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1}z_pz_1 & r_{p2}z_pz_2 & \cdots & r_{pp}z_pz_p \end{pmatrix}}_{\mathbf{R} \circ \mathbf{z}\mathbf{z}^\top} \succeq \underbrace{\lambda \begin{pmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_p \end{pmatrix}}_{\lambda \text{Diag}(\mathbf{z})}. \quad (9)$$

ここで, $\text{Diag}(\mathbf{z})$ は \mathbf{z} の各成分を対角成分に並べた行列を表し, $\mathbf{R} \circ \mathbf{z}\mathbf{z}^\top$ は 2 つの行列 $\mathbf{R}, \mathbf{z}\mathbf{z}^\top$ の Hadamard 積 (要素ごとの積) を表す. この Hadamard 積によって, 選択されていない説明変数に対応する相関係数行列の成分 ($\mathbf{R}(\mathbf{z})$ 以外の成分) が 0 になる. また, λ は相関係数行列の最小固有値を表す決定変数とし,

$$\lambda \geq 0 \quad (10)$$

を満たすものとする.

ここまでの議論から, 相関係数行列の条件数が κ 以下という制約のもとで残差 2 乗和を最小化する変数選択問題は, 次のように定式化できる:

$$\left| \begin{array}{l} \text{minimize}_{\mathbf{a}, \mathbf{z}, \lambda} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p a_j x_{ij} \right)^2 \\ \text{subject to} \quad (6), (7), (9), (10). \end{array} \right. \quad (11)$$

制約 (9) には, $z_j z_\ell$ や, λz_j といった決定変数の双線形項が現れている. しかしこれらの双線形項は, 線形不等式制約を用いて $w_{j\ell} = z_j z_\ell, u_j = \lambda z_j$ と線形項に置き換えられる [1]. したがって条件数制約つき変数選択問題 (11) は, 以下のような 0-1 整数制約, 線形不等式制約, 半正定値制約のもとで凸 2 次関数を最小化する MISDP として定式化できる:

$$\left| \begin{array}{l} \text{minimize}_{\substack{\mathbf{a}, \mathbf{z}, \lambda, \\ \mathbf{u}, \mathbf{W}}} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p a_j x_{ij} \right)^2 \\ \text{subject to} \quad z_j = 0 \Rightarrow a_j = 0 \quad (j = 1, 2, \dots, p), \\ \quad \kappa \text{Diag}(\mathbf{u}) \succeq \mathbf{R} \circ \mathbf{W} \succeq \text{Diag}(\mathbf{u}) \\ \quad w_{j\ell} \geq 0, w_{j\ell} \geq z_j + z_\ell - 1, z_j \geq w_{j\ell}, z_\ell \geq w_{j\ell} \quad (j, \ell = 1, 2, \dots, p), \\ \quad u_j \geq 0, u_j \geq Lz_j + \lambda - L, Lz_j \geq u_j, \lambda \geq u_j \quad (j = 1, 2, \dots, p), \\ \quad L \geq \lambda \geq 0, \\ \quad z_j \in \{0, 1\} \quad (j = 1, 2, \dots, p). \end{array} \right. \quad (12)$$

ただし,

$$\mathbf{u} := (u_1, u_2, \dots, u_p)^\top,$$

$$\mathbf{W} := \begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1p} \\ w_{21} & w_{22} & \cdots & w_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ w_{p1} & w_{p2} & \cdots & w_{pp} \end{pmatrix}$$

であり, L は相関係数行列の最小固有値の上限を表す定数である. 一般に相関係数行列の最小固有値は 1 以下であるので, 実際に問題を解く際は $L = 1$ としておけばよい.

4 切除平面法を用いた解法

4.1 切除平面法のアルゴリズム

問題 (12) は MISDP であり, この問題を直接解くことは難しい. そこで本論文では, Konno *et al.* [9] が提案した SDP を解くための切除平面法を拡張し, MIQP を繰り返し解くことによって MISDP を解く切除平面法を提案する.

まず問題 (12) に対し, 半正定値制約

$$\kappa \text{Diag}(\mathbf{u}) \succeq \mathbf{R} \circ \mathbf{W} \succeq \text{Diag}(\mathbf{u}) \quad (13)$$

を取り除いた以下の緩和問題を考える:

$$\left\{ \begin{array}{ll} \underset{\substack{\mathbf{a}, \bar{\mathbf{z}}, \bar{\lambda}, \\ \bar{\mathbf{u}}, \bar{\mathbf{W}}}}{\text{minimize}} & \sum_{i=1}^n \left(y_i - \sum_{j=1}^p a_j x_{ij} \right)^2 \\ \text{subject to} & z_j = 0 \Rightarrow a_j = 0 \quad (j = 1, 2, \dots, p), \\ & w_{j\ell} \geq 0, w_{j\ell} \geq z_j + z_\ell - 1, z_j \geq w_{j\ell}, z_\ell \geq w_{j\ell} \quad (j, \ell = 1, 2, \dots, p), \\ & u_j \geq 0, u_j \geq Lz_j + \lambda - L, Lz_j \geq u_j, \lambda \geq u_j \quad (j = 1, 2, \dots, p), \\ & L \geq \lambda \geq 0, \\ & z_j \in \{0, 1\} \quad (j = 1, 2, \dots, p). \end{array} \right. \quad (14)$$

緩和問題 (14) は MIQP であり, 一般的な整数計画ソルバーを用いて解くことができる. 続いて, 得られた緩和問題の最適解 $(\bar{\mathbf{a}}, \bar{\mathbf{z}}, \bar{\lambda}, \bar{\mathbf{u}}, \bar{\mathbf{W}})$ に対し, $\text{cond}(\mathbf{R}(\bar{\mathbf{z}})) \leq \kappa$ であるか判定する. $\text{cond}(\mathbf{R}(\bar{\mathbf{z}})) \leq \kappa$ である場合, $\bar{\mathbf{z}}$ は元の問題 (12) の最適解となるので, アルゴリズムを終了する. 一方, $\text{cond}(\mathbf{R}(\bar{\mathbf{z}})) > \kappa$ である場合,

$$\mathbf{d}^\top (\kappa \text{Diag}(\bar{\mathbf{u}}) - \mathbf{R} \circ \bar{\mathbf{W}}) \mathbf{d} < 0 \quad (15)$$

$$\mathbf{e}^\top (\mathbf{R} \circ \bar{\mathbf{W}} - \text{Diag}(\bar{\mathbf{u}})) \mathbf{e} < 0 \quad (16)$$

なるベクトル $\mathbf{d}, \mathbf{e} \neq \mathbf{0}$ のどちらか一方が必ず存在する. 続いて, その \mathbf{d}, \mathbf{e} を用いて,

$$\mathbf{d}^\top (\kappa \text{Diag}(\mathbf{u}) - \mathbf{R} \circ \mathbf{W}) \mathbf{d} \geq 0 \quad (17)$$

$$\mathbf{e}^\top (\mathbf{R} \circ \mathbf{W} - \text{Diag}(\mathbf{u})) \mathbf{e} \geq 0 \quad (18)$$

という妥当不等式を, 緩和問題 (14) の制約に加えて, MIQP を解きなおす. このように不等式制約を追加しながら元の問題の実行可能解が得られるまで MIQP を解くことを繰り返すのが, 本論文で提案する切除平面法の概要である. 切除平面法のアルゴリズムを Algorithm 1 にまとめる. なお, 以下ではベクトル $\mathbf{v} \in \mathbb{R}^p$ と $\mathbf{z} \in \{0, 1\}^p$ に対し, \mathbf{v} の各成分のうち $z_j = 1$ である添字 j に対応する成分を取り出した部分ベクトルを $\mathbf{v}(\mathbf{z})$ と表記する.

Algorithm 1 問題 (12) に対する切除平面法

- 1: (初期化) $k = 0, \mathcal{F}_k = (\text{問題 (14) の実行可能領域})$ とする.
 - 2: **while** $k \leq$ 反復回数の上限 **do**
 - 3: (緩和問題を解く)

緩和問題 $\min \left\{ \sum_{i=1}^n \left(y_i - \sum_{j=1}^p a_j x_{ij} \right)^2 : (\mathbf{a}, \mathbf{z}, \lambda, \mathbf{u}, \mathbf{W}) \in \mathcal{F}_k \right\}$ を解き,
最適解 $(\mathbf{a}_k, \mathbf{z}_k, \lambda_k, \mathbf{u}_k, \mathbf{W}_k)$ を得る.
 - 4: **if** $\text{cond}(\mathbf{R}(\mathbf{z}_k)) \leq \kappa$ **then**
 - 5: **while** 文を抜ける.
 - 6: **else**
 - 7: (相関係数行列の固有ベクトルの計算)

$\mathbf{d}_k, \mathbf{e}_k \in \mathbb{R}^p$ を, $\mathbf{d}_k(\mathbf{z}_k), \mathbf{e}_k(\mathbf{z}_k)$ がそれぞれ $\mathbf{R}(\mathbf{z}_k)$ の最大固有値, 最小固有値に対応する固有ベクトルで, その他の成分は 0 のベクトルとする.
(ただし, $\|\mathbf{d}_k(\mathbf{z}_k)\| = 1, \|\mathbf{e}_k(\mathbf{z}_k)\| = 1$ とする)
 - 8: (切除平面の追加)

以下のとおり実行可能領域を更新:

$$\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k \cap \left\{ (\mathbf{a}, \mathbf{z}, \lambda, \mathbf{u}, \mathbf{W}) \left| \begin{array}{l} \mathbf{d}_k^\top (\kappa \text{Diag}(\mathbf{u}) - \mathbf{R} \circ \mathbf{W}) \mathbf{d}_k \geq 0 \\ \mathbf{e}_k^\top (\mathbf{R} \circ \mathbf{W} - \text{Diag}(\mathbf{u})) \mathbf{e}_k \geq 0 \end{array} \right. \right\}$$
 - 9: **end if**
 - 10: $k \leftarrow k + 1$
 - 11: **end while**
 - 12: **return** \mathbf{z}_k
-

4.2 アルゴリズムの性質

Algorithm 1 では, $\text{cond}(\mathbf{R}(\mathbf{z}_k)) > \kappa$ なる $\mathbf{z}_k \in \{0, 1\}^p$ に対して, 以下の不等式制約を追加する:

$$\mathbf{d}_k^\top (\kappa \text{Diag}(\mathbf{u}) - \mathbf{R} \circ \mathbf{W}) \mathbf{d}_k = \mathbf{d}_k^\top (\kappa \lambda \text{Diag}(\mathbf{z}) - \mathbf{R} \circ \mathbf{z} \mathbf{z}^\top) \mathbf{d}_k \geq 0, \quad (19)$$

$$\mathbf{e}_k^\top (\mathbf{R} \circ \mathbf{W} - \text{Diag}(\mathbf{u})) \mathbf{e}_k = \mathbf{e}_k^\top (\mathbf{R} \circ \mathbf{z} \mathbf{z}^\top - \lambda \text{Diag}(\mathbf{z})) \mathbf{e}_k \geq 0. \quad (20)$$

ただし, $\mathbf{d}_k, \mathbf{e}_k \in \mathbb{R}^p$ は, $\mathbf{d}_k(\mathbf{z}_k), \mathbf{e}_k(\mathbf{z}_k)$ がそれぞれ $\mathbf{R}(\mathbf{z}_k)$ の最大固有値, 最小固有値に対応する固有ベクトルで, その他の成分は 0 のベクトルとする.

ここでは, 以下の補題を用いて Algorithm 1 が有限回の反復で必ず最適解を得られることを示す:

補題 1. $\text{cond}(\mathbf{R}(\mathbf{z}_k)) > \kappa$ なる $\mathbf{z}_k \in \{0, 1\}^p$ と任意の $\lambda \geq 0$ について, \mathbf{z}_k, λ は制約 (19), (20) のどちらか一方を必ず満たさない.

証明. λ に関して場合分けをして証明をする.

1. $0 \leq \lambda < \frac{1}{\kappa} \lambda_{\max}(\mathbf{R}(\mathbf{z}_k))$ の場合

$$\begin{aligned} \mathbf{d}_k^\top (\kappa \lambda \text{Diag}(\mathbf{z}_k) - \mathbf{R} \circ \mathbf{z}_k \mathbf{z}_k^\top) \mathbf{d}_k &= \kappa \lambda - \lambda_{\max}(\mathbf{R}(\mathbf{z}_k)) \\ &< \lambda_{\max}(\mathbf{R}(\mathbf{z}_k)) - \lambda_{\max}(\mathbf{R}(\mathbf{z}_k)) \\ &= 0. \end{aligned}$$

2. $\lambda \geq \frac{1}{\kappa} \lambda_{\max}(\mathbf{R}(\mathbf{z}_k))$ の場合

$$\begin{aligned} \mathbf{e}_k^\top (\mathbf{R} \circ \mathbf{z}_k \mathbf{z}_k^\top - \lambda \text{Diag}(\mathbf{z}_k)) \mathbf{e}_k &= \lambda_{\min}(\mathbf{R}(\mathbf{z}_k)) - \lambda \\ &\leq \lambda_{\min}(\mathbf{R}(\mathbf{z}_k)) - \frac{1}{\kappa} \lambda_{\max}(\mathbf{R}(\mathbf{z}_k)) \\ &< 0 \quad (\because \text{cond}(\mathbf{R}(\mathbf{z}_k)) > \kappa). \end{aligned}$$

よって, \mathbf{z}_k, λ は制約 (19), (20) のどちらか一方を必ず満たさない. □

この補題より, 以下の定理が示される:

定理 1. Algorithm 1 は有限回の反復で終了し, 問題 (12) の最適解を出力する.

証明. 補題 1 の結果から, Algorithm 1 では, 各反復で必ず $\text{cond}(\mathbf{R}(\mathbf{z}_k)) > \kappa$ である解 \mathbf{z}_k を取り除くように実行可能領域を更新していく. したがって, たかだか 2^p 回の反復をすれば, 元の問題の最適解を緩和問題の解として得ることができる. □

4.3 変数減少法を用いた強い妥当不等式の構成法

今回の切除平面法では、各反復で緩和問題として MIQP を解く。しかし候補となる説明変数の数が多いデータの場合では、緩和問題を解きなおす回数が増大しアルゴリズム全体の計算時間も増大すると懸念される。この問題を解消するためには、各反復でなるべく強い妥当不等式を制約に追加することによって、反復回数を削減することが必要である。そこでここでは各反復において強い妥当不等式を構成する方法を考える。

いま、 $\text{cond}(\mathbf{R}(z_k)) > \kappa$ を満たす $z_k \in \{0, 1\}^p$ に対して、 $z \geq z_k$ である $z \in \{0, 1\}^p$ を考える。このとき、例えば Cauchy のインタレース定理を利用すると

$$z \geq z_k \Rightarrow \text{cond}(\mathbf{R}(z)) \geq \text{cond}(\mathbf{R}(z_k)) \Rightarrow \text{cond}(\mathbf{R}(z)) > \kappa \quad (21)$$

が示される。インタレース定理の詳細は Horn and Johnson [8, Theorem 4.3.8]などを参照されたい。式 (21) で成り立つ性質は、多重共線性が生じる変数集合を含む任意の変数集合では多重共線性が生じていることとして解釈できる。このような変数集合の包含関係について成り立つ性質を利用してなるべく強い妥当不等式を構成することを考える。

まず、補題 1 を拡張させた以下の命題が成り立つ:

命題 1. $\text{cond}(\mathbf{R}(z_k)) > \kappa$ なる $z_k \in \{0, 1\}^p$ に対し、 $z \geq z_k$ である任意の $z \in \{0, 1\}^p$, $\lambda \geq 0$ について考える。このとき z, λ は制約 (19), (20) のどちらか一方を必ず満たさない。

証明. $(z_k)_j = 0$ である添字 j に対し、 $(d_k)_j = 0, (e_k)_j = 0$ であることに注意すると、 $z \geq z_k$ である任意の $z \in \{0, 1\}^p, \lambda \geq 0$ に対し、

$$\begin{aligned} d_k^\top (\kappa \lambda \text{Diag}(z) - \mathbf{R} \circ z z^\top) d_k &= d_k^\top (\kappa \lambda \text{Diag}(z_k) - \mathbf{R} \circ z_k z_k^\top) d_k, \\ e_k^\top (\mathbf{R} \circ z z^\top - \lambda \text{Diag}(z)) e_k &= e_k^\top (\mathbf{R} \circ z_k z_k^\top - \lambda \text{Diag}(z_k)) e_k \end{aligned}$$

が成り立つ。したがって補題 1 と同様にして証明できる。 \square

Algorithm 1 では各反復で緩和問題の解 z_k を直接用いて妥当不等式 (19), (20) を制約に追加している。しかし命題 1 の結果から、 z_k ではなく $\bar{z} \leq z_k$, かつ $\text{cond}(\mathbf{R}(\bar{z})) > \kappa$ となる \bar{z} をみつけ、 \bar{z} を取り除く妥当不等式を制約に追加すれば、1 回の反復で z_k を含めより多くの実行不能解を取り除ける。また、強い妥当不等式を構成するためには、なるべく選択する変数の数が少ないような \bar{z} を構成することが望ましい。そこで本論文では、Algorithm 1 に変数減少法を組み合わせ、各反復で強い妥当不等式を追加する方法を考案する。

変数減少法を用いた \bar{z} の構成法について説明する。まず緩和問題の解 z_k を変数減少法の初期点 z_k^1 とする。次に、 z_k^1 において選択された変数の中で、被説明変数との偏相関が最も小さい変数を 1 つ取り除いた解 z_k^2 を構成する。このとき $z_k^1 \geq z_k^2$ であり

$\text{cond}(\mathbf{R}(z_k^1)) \geq \text{cond}(\mathbf{R}(z_k^2))$ となる. 続いて, z_k^2 において選択された変数の中で, 被説明変数との偏相関が最も小さい変数を 1 つ取り除いた解 z_k^3 を構成する. このように 1 つずつ変数を取り除くことを繰り返していくと, 相関係数行列の条件数は単調に減少していき, いずれかの段階で $\text{cond}(\mathbf{R}(z_k^{q-1})) > \kappa$, $\text{cond}(\mathbf{R}(z_k^q)) \leq \kappa$ なる解 z_k^{q-1}, z_k^q を得て, 問題 (12) の実行可能領域に帰還できる. ここで, 実行可能領域に到達する直前の実行不能解 z_k^{q-1} に注目し, $\bar{z} = z_k^{q-1}$ とする. そして \bar{z} を取り除く妥当不等式として, 以下の制約を緩和問題に追加する:

$$\mathbf{d}_k^\top (\kappa \text{Diag}(\mathbf{u}) - \mathbf{R} \circ \mathbf{W}) \mathbf{d}_k \geq 0, \quad (22)$$

$$\mathbf{e}_k^\top (\mathbf{R} \circ \mathbf{W} - \text{Diag}(\mathbf{u})) \mathbf{e}_k \geq 0. \quad (23)$$

ただし, $\mathbf{d}_k, \mathbf{e}_k \in \mathbb{R}^p$ は, $\mathbf{d}_k(\bar{z}), \mathbf{e}_k(\bar{z})$ がそれぞれ $\mathbf{R}(\bar{z})$ の最大固有値, 最小固有値に対応する固有ベクトルで, その他の成分は 0 のベクトルとする. このとき命題 1 の結果から, 制約 (22), (23) は, 問題 (12) の実行不能解 $z_k^1, z_k^2, \dots, z_k^{q-1}$ をまとめて取り除く制約となっている.

また $\bar{z} = z_k^q$ とすると, \bar{z} は問題 (12) の実行可能解である. したがって, \bar{z} で選択している変数で推定した回帰式の残差 2 乗和を計算すると, その値は問題 (12) の目的関数の上界値となる.

以上が変数減少法を用いた強い妥当不等式の構成法である. 変数減少法を組み合わせた切除平面法を Algorithm 2 にまとめる. ただし $\text{RSS}(\mathbf{z})$ は, $\mathbf{z} \in \{0, 1\}^p$ で選択した変数で推定した回帰式の残差 2 乗和を表す. なお Algorithm 2 では, 途中でアルゴリズムを打ち切った場合でも, 実行可能解 \hat{z} を出力することに注意する.

5 計算機実験

提案手法の有用性を確認するために行った計算機実験の結果を示す. 計算機実験においては, 変数増加法, 変数減少法の 2 つの変数選択の手法と比較して, 提案手法である Algorithm 2 の有用性を検証した. 実験には UCI Machine Learning Repository [2] で公開されている 5 つのデータを利用した (表 1). また相関係数行列の条件数の上限は $\kappa = 100, 225$ *2 とした.

今回の計算機実験では, 各手法で選択した変数集合で推定した回帰式に対して, 決定係数 (R^2 値) に着目して分析を行う. 決定係数とは回帰式が被説明変数をどの程度説明できているかを表す指標であり, 残差 2 乗和が小さい回帰式ほど決定係数は 1 に近い大きい値をとることに注意する.

計算機実験において Algorithm 2 は Python 2.7.3 で実装し, 緩和問題の求解には整数計画ソルバー Gurobi 5.6.0 を使用した. 比較手法である変数増加法, 変数減少法のアルゴリズムは MATLAB R2013a を用いて実装した. 計算機実験においては, CPU: Xeon 2.66GHz,

*2 Chatterjee and Hadi [4] は, 相関係数行列の条件数が 225 を超えた場合に多重共線性の兆候が現れると述べている.

Algorithm 2 問題 (12) に対する変数減少法を組み合わせた切除平面法

```

1: (初期化)  $k \leftarrow 0, \mathcal{F}_k \leftarrow$  (問題 (14) の実行可能領域),  $\hat{z} \leftarrow \mathbf{0}$  とする
2: while  $k \leq$  反復回数の上限 do
3:   (緩和問題を解く)
      緩和問題  $\min \left\{ \sum_{i=1}^n \left( y_i - \sum_{j=1}^p a_j x_{ij} \right)^2 : (\mathbf{a}, \mathbf{z}, \lambda, \mathbf{u}, \mathbf{W}) \in \mathcal{F}_k \right\}$  を解き,
      最適解  $(\mathbf{a}_k, \mathbf{z}_k, \lambda_k, \mathbf{u}_k, \mathbf{W}_k)$  を得る.
4:   if  $\text{cond}(\mathbf{R}(\mathbf{z}_k)) \leq \kappa$  then
5:     while 文を抜ける.
6:   else
7:     (変数減少法で実行可能領域へ帰還)
       $\mathbf{z}_k$  を初期点として変数減少法を適用し, 実行可能領域へ帰還.  $\tilde{z}$  を変数減少法で
      得た実行可能解,  $\tilde{z}$  を実行可能領域に達する直前の実行不能解とする.
8:     if  $\text{RSS}(\tilde{z}) < \text{RSS}(\hat{z})$  then
9:       (暫定解の更新)  $\hat{z} \leftarrow \tilde{z}$ 
10:    end if
11:    (相関係数行列の固有ベクトルの計算)
       $\mathbf{d}_k, \mathbf{e}_k \in \mathbb{R}^p$  を,  $\mathbf{d}_k(\tilde{z}), \mathbf{e}_k(\tilde{z})$  がそれぞれ  $\mathbf{R}(\tilde{z})$  の最大固有値, 最小固有値に
      対応する固有ベクトルで, その他の成分は 0 のベクトルとする.
      (ただし,  $\|\mathbf{d}_k(\tilde{z})\| = 1, \|\mathbf{e}_k(\tilde{z})\| = 1$  とする)
12:    (切除平面の追加)
      以下のとおり実行可能領域を更新:
      
$$\mathcal{F}_{k+1} \leftarrow \mathcal{F}_k \cap \left\{ (\mathbf{a}, \mathbf{z}, \lambda, \mathbf{u}, \mathbf{W}) \left| \begin{array}{l} \mathbf{d}_k^\top (\kappa \text{Diag}(\mathbf{u}) - \mathbf{R} \circ \mathbf{W}) \mathbf{d}_k \geq 0 \\ \mathbf{e}_k^\top (\mathbf{R} \circ \mathbf{W} - \text{Diag}(\mathbf{u})) \mathbf{e}_k \geq 0 \end{array} \right. \right\}$$

13:    end if
14:     $k \leftarrow k + 1$ 
15: end while
16: return  $\mathbf{z}_k, \hat{z}$ 

```

quad-core, 8threads; RAM: 24 GB; OS: Red Hat Enterprise Linux 5.3 Desktop (64bit version) の計算機を使用した。またアルゴリズムの最大実行時間は 10000 秒とした。

表 2 に条件数制約つき変数選択問題に対する計算機実験の結果を示す。「手法」はそれぞれ

- FWD: 初期変数集合を空集合とし, 条件数制約を満たす限り R^2 値にもとづいて変数を 1 つずつ追加していく変数増加法;
- BWD: 初期変数集合を全変数とし, 条件数制約を満たすまで R^2 値にもとづいて変数を

表1 実験に用いたデータ

| 略称 | サンプル数 | 説明変数の数 | データ名 |
|--------------|-------|--------|-------------------------|
| Servo | 167 | 19 | Servo |
| AutoMPG | 392 | 25 | Auto MPG |
| SolarFlare-C | 1066 | 26 | Solar Flare (C-class) |
| BreastCancer | 194 | 32 | Breast Cancer Wisconsin |
| Automobile | 159 | 65 | Automobile |

1 つずつ取り除いていく変数減少法;

- MISDP: MISDP を Algorithm 2 で解く提案手法;

を示している。また k は MISDP を解く切除平面法の反復回数, R^2 値 は決定係数, 「選択個数」は選択された説明変数の個数を示している。MISDP の計算時間が 10000 秒を超えた場合は, 切除平面法のアルゴリズムを打ち切っている。計算を打ち切った場合は, Algorithm 2 の各反復で変数減少法を用いて構成した実行可能解のなかで, 最も R^2 値の高い解 \hat{z} の結果を示している。

計算結果を分析する。まず変数増加法と変数減少法によって得られた解に注目すると, BreastCancer や Automobile では選択された変数の個数が大きく異なっていることがわかる。これは, 2 つの手法によって選択される変数集合が大きく異なることを示している。特に BreastCancer のデータで $\kappa = 225$ とした場合は, R^2 値も大きく異なっている。

次に, $p < 30$ であるデータに対しては, MISDP は反復回数が少なく, 高速に最適解を計算できていることがわかる。また $p > 30$ であるデータに対しては, 10000 秒以内で最適性の保証のある解を求めることはできていないが, その場合でも変数増加法, 変数減少法と比較して R^2 値が高い変数集合を選択できている。

$p > 30$ の場合に MISDP の計算時間が増大する原因としては, 緩和問題を 1 回解く計算時間そのものが長時間になっていることがあげられる。反復回数に関しては, BreastCancer のデータで $\kappa = 100$ とした場合に注目すると, MISDP では 90 回緩和問題を解き直してもアルゴリズムが終了しない。したがって, p が大きい場合は多くの反復回数が必要となり, これもアルゴリズム全体の計算時間を増大させる要因になっている。

6 おわりに

本論文では多重共線性を考慮した回帰式の変数選択問題を混合整数半正定値計画問題として定式化し, 切除平面法によって解くことを提案した。また提案手法の有用性を検証するため計

表2 条件数制約付き変数選択問題を解いた結果

| データ名 | n | p | κ | 手法 | k | 計算時間 | 条件数 | R^2 値 | 選択個数 |
|--------------|------|-----|----------|-------|-----|---------|-------|---------|------|
| Servo | 167 | 19 | 100 | MISDP | 6 | 1.26 | 95.3 | 0.75877 | 15 |
| | | | | FWD | — | 0.09 | 63.2 | 0.75862 | 14 |
| | | | | BWD | — | 0.04 | 39.0 | 0.75877 | 15 |
| | | | 225 | MISDP | 5 | 0.96 | 102.5 | 0.75877 | 15 |
| | | | | FWD | — | 0.11 | 146.3 | 0.75877 | 15 |
| | | | | BWD | — | 0.04 | 39.0 | 0.75877 | 15 |
| AutoMPG | 392 | 25 | 100 | MISDP | 5 | 8.98 | 92.8 | 0.87430 | 21 |
| | | | | FWD | — | 0.17 | 87.6 | 0.87335 | 21 |
| | | | | BWD | — | 0.09 | 86.2 | 0.87429 | 19 |
| | | | 225 | MISDP | 4 | 1.60 | 182.0 | 0.87438 | 22 |
| | | | | FWD | — | 0.18 | 185.3 | 0.87438 | 22 |
| | | | | BWD | — | 0.05 | 181.1 | 0.87438 | 22 |
| SolarFlare-C | 1066 | 26 | 100 | MISDP | 14 | 10.92 | 21.3 | 0.19715 | 21 |
| | | | | FWD | — | 0.20 | 36.4 | 0.19713 | 19 |
| | | | | BWD | — | 0.12 | 26.9 | 0.19715 | 19 |
| | | | 225 | MISDP | 14 | 12.26 | 106.5 | 0.19715 | 19 |
| | | | | FWD | — | 0.21 | 36.4 | 0.19713 | 19 |
| | | | | BWD | — | 0.12 | 26.9 | 0.19715 | 19 |
| BreastCancer | 194 | 32 | 100 | MISDP | >90 | >10000 | 85.6 | 0.28827 | 20 |
| | | | | FWD | — | 0.17 | 91.8 | 0.27580 | 15 |
| | | | | BWD | — | 0.27 | 61.8 | 0.26572 | 9 |
| | | | 225 | MISDP | 56 | 5819.14 | 215.6 | 0.30513 | 20 |
| | | | | FWD | — | 0.21 | 217.9 | 0.30010 | 19 |
| | | | | BWD | — | 0.27 | 61.8 | 0.26572 | 9 |
| Automobile | 159 | 65 | 100 | MISDP | >26 | >10000 | 64.7 | 0.96882 | 24 |
| | | | | FWD | — | 0.76 | 99.9 | 0.96447 | 27 |
| | | | | BWD | — | 1.89 | 67.3 | 0.96571 | 24 |
| | | | 225 | MISDP | >22 | >10000 | 224.4 | 0.97391 | 36 |
| | | | | FWD | — | 1.14 | 224.4 | 0.97124 | 39 |
| | | | | BWD | — | 1.85 | 183.5 | 0.97153 | 29 |

算機実験を実施し、変数増加法、変数減少法との比較を行った。変数の数が小さいデータに対しては高速に最適解を得ることができ、変数の数が多いデータでは変数減少法、変数増加法より良い解を得られることを確認した。

本論文で提案した手法は、多重共線性を検出する指標として相関係数行列の条件数を用いている。多重共線性を検出する指標としては分散拡大要因や行列式など様々な指標がある。した

がってこれらの指標について制約を課した変数選択問題に対して、本論文で提案した解法を拡張させることが今後の課題としてあげられる。

参考文献

- [1] F. AL-KHAYYAL, Jointly constrained bilinear programs and related problems: An overview, *Computers & Mathematics with Applications*, **19**, 53–62 (1990).
- [2] K. BACHE AND M. LICHMAN, UCI Machine Learning Repository, <http://archive.ics.uci.edu/ml>.
- [3] D. BERTSIMAS AND R. SHIODA, Algorithm for cardinality-constrained quadratic optimization, *Computational Optimization and Applications*, **43**, 1–22 (2009).
- [4] S. CHATTERJEE AND A. S. HADI, *Regression Analysis by Example*, John Wiley & Sons, 2013.
- [5] M. EFROYMSON, Multiple regression analysis, *Mathematical Methods for Digital Computers*, **1**, 191–203 (1960).
- [6] G. M. FURNIVAL AND R. W. WILSON, Regressions by leaps and bounds, *Technometrics*, **16**, 499–511 (1974).
- [7] I. GUYON AND A. ELISSEEFF, An introduction to variable and feature selection, *The Journal of Machine Learning Research*, **3**, 1157–1182 (2003).
- [8] R. HORN AND C. JOHNSON, *Matrix Analysis*, Cambridge University Press, 1990.
- [9] H. KONNO, J. GOTOH, T. UNO, AND A. YUKI, A cutting plane algorithm for semi-definite programming problems with applications to failure discriminant analysis, *Journal of Computational and Applied Mathematics*, **146**, 141–154 (2002).
- [10] H. KONNO AND Y. TAKAYA, Multi-step methods for choosing the best set of variables in regression analysis, *Computational Optimization and Applications*, **46**, 417–426 (2010).
- [11] H. KONNO AND R. YAMAMOTO, Choosing the best set of variables in regression analysis using integer programming, *Journal of Global Optimization*, **44**, 273–282 (2009).
- [12] R. MIYASHIRO AND Y. TAKANO, Mixed integer second-order cone programming formulations for variable selection, Technical Report, Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2013.
- [13] R. MIYASHIRO AND Y. TAKANO, Subset selection by Mallows' Cp: a mixed integer programming approach, *Expert Systems with Applications*, **42**, 325–331 (2015).
- [14] M. J. TODD, Semidefinite optimization, *Acta Numerica* 2001, **10**, 515–560 (2001).

- [15] L. VANDENBERGHE AND S. BOYD, Semidefinite programming, *SIAM Review*, **38**, 49–95 (1996).