混合整数2次錐計画法による回帰式の変数選択

宮代　隆平，高野　祐一

混合整数 2 次錐計画問題 (MISOCP) は，整数変数を含む 2 次錐計画問題 (SOCP) であり，2 次錐制約と整数変数の高い表現能力により多種多様な問題を定式化することができる．本稿ではその一例として，回帰式の変数選択問題を取り上げ，赤池情報量規準 (AIC) やベイズ情報量規準 (BIC) など，各種の情報量規準の最小化が混合整数 2 次錐計画問題として定式化できることを示す．

キーワード：混合整数 2 次錐計画法，変数選択，情報量規準，回帰分析

1. はじめに

1.1 本稿の目的

2 次錐計画問題 (Second-Order Cone Program; SOCP) は内点法を用いて効率的に解けることと，その高い表現能力から，近年注目を集めている．混合整数 2 次錐計画問題 (Mixed-Integer Second-Order Cone Program; MISOCP) は，SOCP に変数の整数性が付加された問題である．MISOCP の魅力は，SOCP の表現能力に加えて，整数変数を用いた各種の定式化テクニックが使えることである．本稿では，MISOCP が多彩なクラスの問題を表せることの例として，統計学の回帰分析における変数選択問題を取り上げ，各種の情報量規準の最小化問題が MISOCP として定式化できることを述べる．

1.2 2 次錐計画問題と混合整数 2 次錐計画問題

2 次錐計画問題 (SOCP) の一般形は，以下のように表される（ただしここで $\|\boldsymbol{v}\| = (\boldsymbol{v}^\top \boldsymbol{v})^{1/2}$）：

$$\begin{aligned} &\underset{\boldsymbol{x}}{\text{minimize}} \quad \boldsymbol{c}_0^\top \boldsymbol{x} \\ &\text{subject to} \quad \|A_i \boldsymbol{x} + \boldsymbol{b}_i\| \le \boldsymbol{c}_i^\top \boldsymbol{x} + d_i \\ &\qquad\qquad\qquad\qquad (i = 1, 2, \ldots, m). \end{aligned}$$

SOCP は内点法を用いて多項式時間で解を求めることができ，実際に近年の数理計画ソルバーではかなり大規模な問題まで解ける．SOCP についてより詳しくは，[1, 2] などを参照されたい．

SOCP の制約式 $\|A_i \boldsymbol{x} + \boldsymbol{b}_i\| \le \boldsymbol{c}_i^\top \boldsymbol{x} + \boldsymbol{d}_i$ は 2 次錐制約と呼ばれるが，$\boldsymbol{x}^\top \boldsymbol{x} \le f \cdot g$（$f$，$g$ は非負のスカラー変数）という形の制約式（hyperbolic 制約式）は，以下のように 2 次錐制約として表現することができる：

$$\boldsymbol{x}^\top \boldsymbol{x} \le f \cdot g \quad \Longleftrightarrow \quad \left\| \begin{pmatrix} 2\boldsymbol{x} \\ f - g \end{pmatrix} \right\| \le f + g\,.$$

これにより，線形計画問題などでは直接扱うことの難しかった双線形項 $f \cdot g$ を，限定された形式ではあるがダイレクトに記述できる．このことからも，SOCP が高い表現能力を持つことが見てとれる．

混合整数 2 次錐計画問題 (MISOCP) とは，SOCP の一部または全ての変数に整数制約が課された問題である．MISOCP 自体は NP 困難であるが，整数制約を連続緩和した問題は SOCP となり，内点法を用いて解くことができる．最近では，内点法ベースの分枝限定法が実装された，MISOCP を扱える数理計画ソルバーが複数存在する．MISOCP についてより詳しくは，[3] などを参照されたい．

2. 回帰式の変数選択

2.1 変数選択とは

統計学における回帰分析では，説明変数と被説明変数の間の関係式（回帰式）を求めることが主要な目的となる．例えば，マンションの家賃を予測する場合には，その説明変数として，総面積，築年数，立地，最寄り駅までの距離などが考えられるだろう．しかし，推定に使用するサンプル数に対して説明変数が非常に多い場合や，予測に有効ではない説明変数が含まれる場

みやしろ　りゅうへい
東京農工大学大学院工学研究院
〒 184–8588 東京都小金井市中町 2–24–16
たかの　ゆういち
専修大学ネットワーク情報学部
〒 214–8580 神奈川県川崎市多摩区東三田 2–1–1

合には，手元のデータには良く当てはまるが，未知のデータに対しては予測性能が低い回帰式が得られてしまうことがある．この現象は過剰適合と呼ばれる．

このようなときには，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ことにより，過剰適合を軽減して回帰式の予測性能を向上させることができる．また，この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ほかにも，回帰分析の結果の解釈が容易になることや，推定に必要な計算量を削減できることなどがある [4].

回帰式の変数選択は，統計分野では古くから重要な課題として知られている [5〜7]．また，機械学習やデータマイニングの分野では特徴選択とも呼ばれ，扱うデータ量の増加を背景に近年大きな注目を集めている [8].

2.2 変数選択と情報量規準

以下本稿では，線形回帰モデルに対する変数選択問題を考える．

n 個の観測値 $(y_i; x_{i1}, x_{i2}, \ldots, x_{ik})$ $(i = 1, 2, \ldots, n)$ が与えられ，y_i を被説明変数，x_{ij} $(j = 1, 2, \ldots, k)$ を k 種類の説明変数とする．このとき，被説明変数の値を予測する回帰モデルは以下のように表される：

$$y_i = b + a_1 x_{i1} + a_2 x_{i2} + \cdots + a_k x_{ik} + \varepsilon_i \,, \qquad (1)$$

ただし ε_i は i 番目の観測値の予測残差．

このとき，候補となる p 種類の変数の中から最適な説明変数集合を選択する変数選択問題を考える．

一般に，説明変数を多く含むモデルは与えられたデータに対しては当てはまりが良いものの，未知のデータに対しては予測能力が低くなる．モデルの複雑さ（選択した説明変数の個数）とモデルの当てはまりの良さ（所与のデータに対する予測誤差の小ささ）とのバランスを考慮した適合度指標の代表的なものとして，赤池情報量規準 (Akaike Information Criterion; AIC [9]) やベイズ情報量規準 (Bayesian Information Criterion; BIC [10]) がある．回帰モデル (1) に対するいくつかの自然な仮定の下で，定数項を削除した AIC 値は

$$n \log \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + 2k \,, \qquad (2)$$

定数項を削除した BIC 値は

$$n \log \left(\frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^2 \right) + k \log n \qquad (3)$$

と表わされ，これらの値が小さいほど良いモデルとされる．以下では主として AIC 値の最小化を例にとり説明を行うが，AIC，BIC 以外にも多数の適合度指標が提案されている（[11] などを参照のこと）.

2.3 変数選択に関する先行研究

変数選択のための手法は，これまでに多くのものが提案されているが [4, 8, 12, 13]，それら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，最も一般的な手法はステップワイズ法 [14] 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 個ずつ追加／削除することを繰り返す方法である．ステップワイズ法は R [15] や MATLAB [16] などのソフトウェアに実装されており，計算も高速なため広く用いられているが，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必ずしも出力しない．

一方で厳密解法としては，枝刈り操作を組み合わせた総当り法 [17] や整数計画法 [18] がある．これらの手法は 1970 年代から提案されていたが，当時は計算機が非力だったために小さなサイズの問題しか扱うことができず，あまり注目を集めなかったと考えられる．しかし，計算機や数理計画ソルバーの性能が飛躍的に向上した現在では，これらの厳密解法を研究するための機は熟したと考え，筆者らは特に整数計画法を用いた解法に注目している．

整数計画法を用いた変数選択では，選択する説明変数の個数を事前に指定するタイプの定式化が多い [18]〜[21]．選択する説明変数の個数 k を定数と見なすと，関数 (2) および関数 (3) の最小化は $\sum_{i=1}^{n} \varepsilon_i^2$ の最小化と等価になり，これは「p 個の説明変数候補から，残差 2 乗和を最小化するように k 個の説明変数を選ぶ問題」になる．この問題は以下の混合整数 2 次計画問題として定式化できることが知られている（例えば [19]）：

$$\underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, \boldsymbol{z}}{\text{minimize}} \quad \sum_{i=1}^{n} \varepsilon_i^2 \qquad (4)$$

subject to

$$\varepsilon_i = y_i - \left(b + \sum_{j=1}^{p} a_j x_{ij} \right) \quad (i = 1, 2, \ldots, n), \qquad (5)$$

$$-Mz_j \le a_j \le Mz_j \quad (j = 1, 2, \ldots, p), \qquad (6)$$

$$\sum_{j=1}^{p} z_j = k \,, \qquad (7)$$

$$z_j \in \{0, 1\} \quad (j = 1, 2, \ldots, p), \qquad (8)$$

ただし M は十分に大きい正の定数．

ここで連続変数 a_j および b はそれぞれ推定すべき偏回帰係数および切片であり，ε_i は残差である．また変数 z_j は 0-1 変数であり，p 個中の j 番目の説明変数候補をモデルに含めるとき 1，そうでないときに 0 をとるものと定義する．目的関数 (4) は残差 2 乗和の最小化を表しており，制約式 (5) は残差の定義そのものである．制約式 (6) は，$z_j = 0$ の場合に j 番目の説明変数はモデルから取り除かれることを意味しており，制約式 (7) でモデルに含まれる説明変数の個数が k 個であることを定めている．

上記の問題は，凸 2 次関数を線形制約と整数制約の下で最小化する混合整数 2 次計画問題であり，既存の数理計画ソルバーなどで扱える．しかしながら，この定式化では k を所与の定数として扱っており，現実には AIC 値や BIC 値を最小にする説明変数の個数は事前にわからないという問題点がある．

そこで次節では，AIC，BIC などの情報量規準に基づいて，説明変数の個数 k も同時に最適化する変数選択問題が MISOCP として定式化できることを示す．ほかにも筆者らのグループでは，自由度調整済み決定係数を指標とした定式化 [22]，Mallows の C_p を指標とした定式化 [23]，多重共線性を考慮した定式化 [24] などを提案しており，興味がある読者は各文献を参照していただきたい．

3. MISOCP による情報量規準の最小化

AIC 最小化問題は，関数 (2) を最小化すればよいことから，直接的には以下の非線形混合整数計画問題として定式化できる（k も変数であることに注意）：

$$\begin{array}{ll} \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} & n \log\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \varepsilon_i^2\right) + 2k \\ \text{subject to} & \text{制約式 (5)，(6)，(7)，(8).} \end{array} \tag{9}$$

しかし，この定式化の目的関数 (9) は非凸な非線形関数であり，0-1 変数 z_j の整数性を緩和したうえでもこの問題を解くことは困難である．整数計画問題に対して分枝限定法がうまく働くには，その連続緩和問題が効率的に解ける形になっている必要がある．以下ではそれを念頭におき，問題を等価変形する．

まず，目的関数 (9) を次のように変換する．

$$\begin{array}{ll} & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad n \log\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \varepsilon_i^2\right) + 2k \\ \iff & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad \log\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \varepsilon_i^2\right) + \dfrac{2k}{n} \\ \iff & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad \exp\left(\log\left(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \varepsilon_i^2\right) + \dfrac{2k}{n}\right) \\ \iff & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad \exp\left(\dfrac{2k}{n}\right) \cdot \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \varepsilon_i^2 \\ \iff & \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, k,\, \boldsymbol{z}}{\text{minimize}} \quad \exp\left(\dfrac{2k}{n}\right) \cdot \displaystyle\sum_{i=1}^{n} \varepsilon_i^2 \end{array}$$

項 $\exp(2k/n) \cdot \sum_{i=1}^{n} \varepsilon_i^2$ の上界を表す連続変数 f を導入することにより，問題は以下の形に変形できる：

$$\begin{array}{ll} \underset{\boldsymbol{a},\, b,\, \boldsymbol{\varepsilon},\, f,\, k,\, \boldsymbol{z}}{\text{minimize}} & f \\ \text{subject to} & \displaystyle\sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot \exp\left(-\dfrac{2k}{n}\right), \\ & \text{制約式 (5)，(6)，(7)，(8).} \end{array} \tag{10}$$

次に，制約式 (10) の非線形性を解消する．新たな 0-1 変数 w_j $(j = 0, 1, \ldots, p)$ を定義し，w_j に関して制約式

$$\sum_{j=0}^{p} (w_j \cdot j) = k, \quad \sum_{j=0}^{p} w_j = 1$$

を付け加える．すると，制約式 (7) と (8) より k は常に整数値を取るため，w_j は「$j = k$ の時，またその時のみ $w_j = 1$」を満たす 0-1 変数となる．この w_j を用いることにより，非線形の制約式 (10) は以下の線形制約式で表現できる：

$$\begin{array}{ll} & \displaystyle\sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot \exp\left(-\dfrac{2k}{n}\right) \\ \iff & \displaystyle\sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot g, \quad g = \exp\left(-\dfrac{2k}{n}\right) \\ \iff & \begin{cases} \displaystyle\sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot g, \quad g = \sum_{j=0}^{p} \left(w_j \cdot \exp\left(-\dfrac{2j}{n}\right)\right), \\ k = \displaystyle\sum_{j=0}^{p} (w_j \cdot j), \quad \sum_{j=0}^{p} w_j = 1. \end{cases} \end{array}$$

ここで制約式 $g = \sum_{j=0}^{p} (w_j \cdot \exp(-2j/n))$ は，変数 w_j に関して線形であることに留意されたい（この線形化手法は SOS type 1 [25, 26] として整数計画の分野で古くから知られている）．制約式 $\sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot g$ は非線形であるが，これは 1.2 節で説明したように 2 次錐制約として表現可能である．

これらをまとめ，AIC 最小化問題の定式化として以

下を得る：

$$
\begin{aligned}
&\underset{\substack{\boldsymbol{a},\,b,\,\boldsymbol{\varepsilon},\,f\\ g,\,k,\,\boldsymbol{w},\,\boldsymbol{z}}}{\text{minimize}} \quad f\\
&\text{subject to}\\
&\varepsilon_i = y_i - \left(b + \sum_{j=1}^{p} a_j x_{ij}\right) \quad (i = 1, 2, \ldots, n),\\
&\sum_{i=1}^{n} \varepsilon_i^2 \le f \cdot g, \quad g = \sum_{j=0}^{p} \left(w_j \cdot \exp\left(-\frac{2j}{n}\right)\right),\\
&\sum_{j=0}^{p} (w_j \cdot j) = k, \quad \sum_{j=0}^{p} w_j = 1, \quad \sum_{j=1}^{p} z_j = k,\\
&-Mz_j \le a_j \le Mz_j \quad (j = 1, 2, \ldots, p),\\
&w_j \in \{0, 1\} \quad (j = 0, 1, \ldots, p),\\
&z_j \in \{0, 1\} \quad (j = 1, 2, \ldots, p).
\end{aligned}
$$

この定式化で表される問題は，線形制約，2 次錐制約，整数制約の下で線形目的関数を最小化する MISOCP である．

また上記の定式化のうち，制約式

$$
g = \sum_{j=0}^{p} \left(w_j \cdot \exp\left(-\frac{2j}{n}\right)\right)
$$

を制約式

$$
g = \sum_{j=0}^{p} \left(w_j \cdot \exp\left(-\frac{j \log n}{n}\right)\right) = \sum_{j=0}^{p} \left(w_j \cdot n^{-j/n}\right)
$$

に置きかえることにより，BIC 最小化問題を MISOCP として定式化できる．さらに，AIC の有限修正 (corrected AIC; [27])，Hannan-Quinn 情報量規準 (HQ; [28])，残差情報量規準 (RIC; [29, 30]) などの様々な情報量規準の最小化問題，あるいは自由度調整済み決定係数 $\bar{R}^2$ [31] の最大化問題も同様に MISOCP として定式化できるが，ここでは省略する．詳細は [22] を参照されたい．

なお，より一般的に線形制約の下で「$\log f_1(\boldsymbol{x}) + f_2(k)$; ただし f_1 は凸 2 次関数，k は整数変数（f_2 の凸性は不要）」の形の目的関数を最小化する問題も，同様の変形により MISOCP として定式化できることを述べておく．

4. 計算機実験および考察

本節では，AIC および BIC の最小化を目的とした変数選択問題に関する計算機実験の結果を紹介する．実験には，UCI Machine Learning Repository [32] で公開されている 10 種類の回帰分析用のデータセットを使用した（表 1）．なお，MISOCP を解く際に生じる数値的不安定性を軽減するために，量的変数は全て平均 0，標準偏差 1 に正規化した．

表 1　データのリスト

略称	n	p	データセット [32]
`Housing`	506	13	Housing
`Servo`	167	19	Servo
`AutoMPG`	392	25	Auto MPG
`SolarFlareC`	1066	26	Solar Flare (C-class)
`SolarFlareM`	1066	26	Solar Flare (M-class)
`SolarFlareX`	1066	26	Solar Flare (X-class)
`BreastCancer`	194	32	Breast Cancer Wisconsin
`ForestFires`	517	63	Forest Fires
`Automobile`	159	65	Automobile
`Crime`	1993	100	Communities and Crime

MISOCP の計算には数理計画ソルバー IBM ILOG CPLEX 12.5 [33] を使用した．計算機環境は，CPU: Intel Xeon W5590 3.33 GHz×2; RAM: 24 GB; OS: 64bit Windows 7 Ultimate SP1; チップセット：Intel 5520 のマシンを使用し，各 MISOCP に対して並列分枝限定法（8 スレッド）を最大 12,000 秒実行した．また比較のため，MATLAB R2012b [16] の `LinearModel.stepwise` 関数を用いて，ステップワイズ法による変数選択を行った．こちらの計算機環境は CPU: Intel Core i7-2600S 2.80 GHz; RAM: 8 GB; OS: 64bit Windows 7 Professional SP1; チップセット：Intel Q67 Express である．

表 2 に AIC 最小化の，表 3 に BIC 最小化の計算機実験の結果を示す．「手法」はそれぞれ

- $\mathrm{SW}_{\mathrm{const}}$：切片のみのモデルから開始するステップワイズ変数選択
- $\mathrm{SW}_{\mathrm{all}}$：全ての候補変数を含むモデルから開始するステップワイズ変数選択
- MISOCP：MISOCP を用いた変数選択

を意味し，AIC 値または BIC 値が最も良かった手法の値を太字で示してある．また k は選択された説明変数の個数である．なお計算時間については，MISOCP の計算が 12,000 秒を超えた場合は分枝限定法を打ち切った．そのため，その場合は得られた説明変数集合は最適解とは限らない．

計算結果を見ると，まず 2 種類のステップワイズ法によって得られた解は，選択された変数集合がかなり異なっていることがわかる．$\mathrm{SW}_{\mathrm{const}}$ と $\mathrm{SW}_{\mathrm{all}}$ では k の値がかなり異なるほか，特に `Automobile` では AIC

表 2　AIC 最小化の実験結果

データ	n	p	手法	AIC 値	k	計算時間（秒）
`Housing`	506	13	$\mathrm{SW_{const}}$	**776.36**	11	1.31
			$\mathrm{SW_{all}}$	**776.36**	11	0.51
			MISOCP	**776.36**	11	10.62
`Servo`	167	19	$\mathrm{SW_{const}}$	**258.66**	9	1.85
			$\mathrm{SW_{all}}$	266.36	14	0.75
			MISOCP	**258.66**	9	8.41
`AutoMPG`	392	25	$\mathrm{SW_{const}}$	**333.22**	15	3.96
			$\mathrm{SW_{all}}$	339.44	19	1.59
			MISOCP	**333.22**	15	51.23
`SolarFlareC`	1066	26	$\mathrm{SW_{const}}$	**2816.34**	9	3.09
			$\mathrm{SW_{all}}$	2819.73	13	4.02
			MISOCP	**2816.34**	9	227.25
`SolarFlareM`	1066	26	$\mathrm{SW_{const}}$	**2926.93**	7	2.38
			$\mathrm{SW_{all}}$	**2926.93**	7	5.99
			MISOCP	**2926.93**	7	92.18
`SolarFlareX`	1066	26	$\mathrm{SW_{const}}$	**2882.81**	3	1.20
			$\mathrm{SW_{all}}$	**2882.81**	3	7.59
			MISOCP	**2882.81**	3	10.73
`BreastCancer`	194	32	$\mathrm{SW_{const}}$	509.72	8	3.07
			$\mathrm{SW_{all}}$	510.58	14	8.13
			MISOCP	**508.73**	10	10637.66
`ForestFires`	517	63	$\mathrm{SW_{const}}$	**1429.81**	12	9.56
			$\mathrm{SW_{all}}$	**1429.81**	12	71.19
			MISOCP	1430.25	13	> 12000
`Automobile`	159	65	$\mathrm{SW_{const}}$	-26.87	21	14.20
			$\mathrm{SW_{all}}$	-47.50	38	24.75
			MISOCP	**−58.49**	32	> 12000
`Crime`	1993	100	$\mathrm{SW_{const}}$	3424.26	41	84.94
			$\mathrm{SW_{all}}$	**3410.92**	50	312.82
			MISOCP	3419.65	51	> 12000

値で 20 以上の差，BIC 値で 10 以上の差があり，これは無視できない差と言える．またこれらの結果から，ステップワイズ法を用いる場合には，$\mathrm{SW_{const}}$ と $\mathrm{SW_{all}}$ の両方を試すべきだという知見が得られる．

次に，$p < 30$ であるような小さなデータセットに対しては，MISOCP は数分以内に最適解を計算できていることがわかる．また，$p > 60$ であるような大きなデータセットに対しては，12,000 秒以内に最適解を求めることはできていないが，その場合でも 2 種類のステップワイズ法とほぼ同等またはそれ以上の解を出力している．特に `Automobile` では，$\mathrm{SW_{const}}$ と $\mathrm{SW_{all}}$ の良い方と比較しても AIC 値，BIC 値ともに 10 以上小さな変数集合を得ている．もちろん，もとよりステップワイズ法は解の最適性を保証するものではないが，変数選択の結果が重要な意味を持つ事例においては MISOCP による手法を試す価値があると思われる．

計算時間に関しては，MISOCP はステップワイズ法よりかなり長い時間を要している．これもヒューリスティクスと厳密解法との性質の違いではあるが，特に MISOCP では分枝限定法の実行中に解くべき連続緩和問題が SOCP になり内点法が必要となるため，通常の線形整数計画問題のように双対単体法を利用したホットスタートが働かない．これが計算時間増大の理由であると考えられる．また，概して AIC 最小化より BIC 最小化の方が MISOCP による計算時間が短いが，これは AIC より BIC の方が説明変数の個数にかかるペナルティが大きいことによるものと思われる．

5. おわりに

本稿では，回帰式の変数選択における情報量規準の最小化問題について，MISOCP による定式化を紹介した．SOCP は hyperbolic 制約式のような非線形関数を直接扱うことができ，さらに MISOCP では整数計画法の様々な定式化テクニックも利用することがで

表 3　BIC 最小化の実験結果

データ	n	p	手法	BIC 値	k	計算時間（秒）
`Housing`	506	13	$\mathrm{SW_{const}}$	834.88	8	1.04
			$\mathrm{SW_{all}}$	**827.07**	11	0.47
			MISOCP	**827.07**	11	13.26
`Servo`	167	19	$\mathrm{SW_{const}}$	**288.93**	8	1.60
			$\mathrm{SW_{all}}$	303.87	11	1.50
			MISOCP	**288.93**	8	10.51
`AutoMPG`	392	25	$\mathrm{SW_{const}}$	**390.96**	11	3.20
			$\mathrm{SW_{all}}$	405.71	14	2.92
			MISOCP	**390.96**	11	59.92
`SolarFlareC`	1066	26	$\mathrm{SW_{const}}$	2855.93	6	2.04
			$\mathrm{SW_{all}}$	**2855.89**	6	6.51
			MISOCP	**2855.89**	6	73.73
`SolarFlareM`	1066	26	$\mathrm{SW_{const}}$	2956.02	4	1.52
			$\mathrm{SW_{all}}$	**2954.42**	4	6.96
			MISOCP	**2954.42**	4	20.59
`SolarFlareX`	1066	26	$\mathrm{SW_{const}}$	**2900.12**	2	0.93
			$\mathrm{SW_{all}}$	**2900.12**	2	7.83
			MISOCP	**2900.12**	2	5.52
`BreastCancer`	194	32	$\mathrm{SW_{const}}$	529.28	3	1.34
			$\mathrm{SW_{all}}$	528.90	3	11.70
			MISOCP	**527.86**	3	1198.73
`ForestFires`	517	63	$\mathrm{SW_{const}}$	**1463.81**	3	2.82
			$\mathrm{SW_{all}}$	**1463.81**	3	71.63
			MISOCP	**1463.81**	3	> 12000
`Automobile`	159	65	$\mathrm{SW_{const}}$	31.28	15	10.56
			$\mathrm{SW_{all}}$	42.59	27	35.19
			MISOCP	**20.81**	23	> 12000
`Crime`	1993	100	$\mathrm{SW_{const}}$	**3574.68**	13	24.92
			$\mathrm{SW_{all}}$	3594.84	22	390.06
			MISOCP	3591.94	16	> 12000

きる．特に本稿では変数選択を 0-1 変数によって表し，非線形関数の線形化手法を利用して非凸非線形な目的関数を持つ情報量基準最小化問題を MISOCP に等価変形する方法を示した．

このように，SOCP の表現能力と整数計画法の定式化手法を活用することで，非常に広いクラスの問題が MISOCP として定式化できる．一方で，本稿の計算機実験でも観察されたように，分枝限定法においてホットスタートが働かないことが足かせとなり，MISOCP は計算時間の面で苦しくなる場合があることは否めない．また，SOCP を解く際の数値的不安定性が原因となって分枝限定法の計算が破綻するようなケースもあり，現時点では MISOCP として定式化することが有効な問題かどうかを見極めることが必要であると言える．MISOCP がその真価を発揮するためにも，解法のさらなる安定化と高速化が期待されている．

謝辞　本研究の一部は JSPS 科研費 26560165 の助成を受けたものである．

参考文献

[1] M. S. Lobo, L. Vandenberghe, S. Boyd and H. Lebret, “Applications of second-order cone programming,” *Linear Algebra and its Applications*, **284**, 193–228, 1998.

[2] F. Alizadeh and D. Goldfarb, “Second-order cone programming,” *Mathematical Programming*, **95**, 3–51, 2003.

[3] H. Y. Benson and Ü. Sağram, “Mixed-integer second-order cone programming: A survey,” *2013 Tutorials in Operations Research: Theory Driven by Influential Applications*, H. Topaloglu and J. C. Smith (eds.), INFORMS, pp. 13–36, 2013.

[4] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, **3**, 1157–1182, 2003.

[5] K. P. Burnham and D. R. Anderson, *Model Selection and Multimodel Inference: A Practical Information Theoretic Approach*, 2nd ed., Springer, 2002.

[6] R. R. Hocking, "The analysis and selection of variables in linear regression," *Biometrics*, **32**, 1–49, 1976.
[7] A. Miller, *Subset Selection in Regression*, 2nd ed., Chapman and Hall/CRC, 2002.
[8] H. Liu and H. Motoda, *Computational Methods of Feature Selection*, Chapman and Hall/CRC, 2007.
[9] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, **19**, 716–723, 1974.
[10] G. Schwarz, "Estimating the dimension of a model," *Annals of Statistics*, **6**, 461–464, 1978.
[11] 小西貞則，北川源四郎，『情報量規準（シリーズ・予測と発見の科学 2）』，朝倉書店，2004.
[12] A. L. Blum and P. Langley, "Selection of relevant features and examples in machine learning," *Artificial Intelligence*, **97**, 245–271, 1997.
[13] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial Intelligence*, **97**, 273–324, 1997.
[14] M. A. Efroymson, "Multiple regression analysis," *Mathematical Methods for Digital Computers*, A. Ralston and H. S. Wilf (eds.), Wiley, pp. 191–203, 1960.
[15] R Core Team, "R: A language and environment for statistical computing," R Foundation for Statistical Computing, http://www.R-project.org（2014年9月1日閲覧）
[16] The MathWorks Inc., MATLAB R2012b, 2012.
[17] G. M. Furnival and R. W. Wilson Jr., "Regressions by leaps and bounds," *Technometrics*, **16**, 499–511, 1974.
[18] T. S. Arthanari and Y. Dodge, *Mathematical Programming in Statistics*, John Wiley & Sons, 1981.
[19] D. Bertsimas and R. Shioda, "Algorithm for cardinality-constrained quadratic optimization," *Computational Optimization and Applications*, **43**, 1–22, 2009.
[20] 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.
[21] H. Konno and R. Yamamoto, "Choosing the best set of variables in regression analysis using integer programming," *Journal of Global Optimization*, **44**, 272–282, 2009.
[22] R. Miyashiro and Y. Takano, "Mixed integer second-order cone programming formulations for variable selection," Technical Report 2013-7, Department of Industrial Engineering and Management, Tokyo Institute of Technology, 2013, http://www.me.titech.ac.jp/technicalreport/h25/2013-7.pdf
[23] R. Miyashiro and Y. Takano, "Subset selection by Mallows' C_p: A mixed integer programming approach," *Expert Systems with Applications*, **42**, 325–331, 2015.
[24] 小林健，高野祐一，宮代隆平，中田和秀，"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—，"日本オペレーションズ・リサーチ学会2014年秋季研究発表会アブストラクト集，164–165, 2014.
[25] E. M. L. Beale, "Two transportation problems," *Proceedings of the 3rd International Conference on Operational Research*, 780–788, 1963.
[26] E. M. L. Beale and J. A. Tomlin, "Special facilities in a general mathematical programming system for non-convex problems using ordered sets of variables," *Proceedings of the 5th International Conference on Operational Research*, 447–454, 1970.
[27] N. Sugiura, "Further analysts of the data by Akaike's information criterion and the finite corrections," *Communications in Statistics — Theory and Methods*, **7**, 13–26, 1978.
[28] E. J. Hannan and B. G. Quinn, "The determination of the order of an autoregression," *Journal of the Royal Statistical Society, Series B*, **41**, 190–195, 1979.
[29] C. Leng, "The residual information criterion, corrected," *arXiv*, 0711.1918v1, 2013, http://arxiv.org/pdf/0711.1918.pdf
[30] P. Shi and C.-L. Tsai, "Regression model selection: A residual likelihood approach," *Journal of the Royal Statistical Society, Series B*, **64**, 237–252, 2002.
[31] R. J. Wherry, "A new formula for predicting the shrinkage of the coefficient of multiple correlation," *The Annals of Mathematical Statistics*, **2**, 440–457, 1931.
[32] K. Bache and M. Lichman, *UCI Machine Learning Repository*, University of California, 2013, http://archive.ics.uci.edu/ml
[33] IBM ILOG, IBM ILOG CPLEX 12.5, 2012.

