対数的な幅を持つ疎行列圧縮の NP 完全性

坂内 英夫 [*]　　後藤 啓介 [†]　　神田 峻介 [†]　　クップル ドミニク [*] [‡]

概要

疎行列圧縮は，行列の列を一本の文字列に直線化し，列の配置をオフセットで識別する問題である．1977 年に Ziegler によって提案されたこの疎行列圧縮問題は，オフセットの長さが 2 以下の場合でも NP 困難であることが知られているが，この問題を難しくするパラメータについてはほとんどわかっていない．本研究の貢献は二つある．まず，オフセットの長さが 2 以下の場合について証明を表示する．次に，制限付きの有向ハミルトンパス問題からの帰着により，対数的な行列幅においても NP 困難であることを示す．

1　序論

Ziegler [12] によって提案された行列圧縮問題は，現在の文献では Sparse Matrix Compression (smc) または Compressed Transition Matrix 問題として言及されている．これは次のように表すことができる．

問題 1 (smc, [8, 課 A4.2, 問題 SR13])**.** $n \times \ell$ の二進行列 $A[1..n][1..\ell]$ である．ただし，n と ℓ はそれぞれに行の個数と列の個数を示す．すべての $i \in [1..n]$, $j \in [1..\ell]$ について，$A[i][j] \in \{0,1\}$ である．整数 $k \in [0..\ell \cdot (n-1)]$ が与えられたとき，以下の 2 つのことの存在を判定する．

1. 整数配列 $C[1..\ell+k]$，任意 $i \in [1..\ell+k]$ に対して $C[i] \in [0..n]$,
2. シフト関数 $s : [1..n] \to [0..k]$.

ただし，C と s は以下の条件を満たす．すべての $i \in [1..n]$，$j \in [1..\ell]$ に対して，$A[i][j] = 1 \Leftrightarrow C[s(i)+j] = i$ が成り立つ．ボーダーケースとして，すべて j に対して，$A[0][j] = 0$ とする．そして，$C[i] = 0$ とする可能になるで，$C[i] = 0$ が設定されていないことをモデル化する．

例 1. 二進行列 $A \in \{0,1\}^{3\times 5}$ が与えられたとする．

$$A = \begin{pmatrix} 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \end{pmatrix} \quad \text{と} \quad k = 2.$$

そして $n = 3$ と $\ell = 5$ である．問題 *1* の解を与えるために，$C[1..7]$ の値とシフト関数 s を選択して，行列 A を C と s で表現する必要がある．この例では，$s(1) = 0, s(2) = 1$, および $s(3) = 2$ を選択することで，行 *1* はそのままに，行 *2* を *1* つ，行 *3* を *2* つシフトすることを意味する．

$$\begin{pmatrix} 1 & 0 & 1 & 0 & 0 & & \\ & 1 & 0 & 0 & 1 & 0 & \\ & & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

このシフトされた行列 A の各列には *1* つ以上の ‘1′ が含まれていないことに観測できる．したがって，各列を，その列に ‘1′ を持つ行のインデックス（またはその列にすべての行が ‘0′ を格納している場合は値 0）で表すことができる．
左から列ごとにこの表現を読むと，シーケンス $C = (1,2,1,3,2,3,3)$ が得られる．　□

Even et al. [6] は，未発表の原稿で，問題 1 が NP 完全であることを示した．さらに，シフトを 2 に制限

[*] 東京医科歯科大学
[†] 株式会社リーガルオンテクノロジーズ
[‡] 山梨大学 大学院　総合研究部　工学域 電気電子情報工学系

した場合，つまり，すべての i に対して $s(i) \in [0..2]$ であるときでも NP 完全であることを示した．しかし，彼らの帰着は，行列幅 $\ell \in O(n)$ まで要求するように見える．多くの参考文献で引用されているが，私たちはその原稿を入手できなかった．しかし，Garey and Johnson [8] の教科書にあるノートにある，グラフ 3 彩色からの帰着に従って，おそらく証明を再構築する．したがって，問題が行列幅 ℓ に対していまだ NP 困難であるかどうかは未解決の問題であった．本論文では，行列幅 $\ell \in \Omega(\lg n)$ に対して NP 困難であることを証明することで，この問いに一部答える．これにより，ℓ の範囲を $\omega(\lg\lg n)$[9] から $o(\lg n)$ に縮小し，その問いが未解決のままである範囲を残す．

1.1 関連研究

Ziegler [12] が問題 1 のために提案した貪欲法は，行のうちに最も多くの ‘1’ を持つ行から順番に処理するというものであった．Chang and Buehrer [2] は，行列の行の循環回転を許可する変種を検討し，最終的な行列表現のスペースの境界を改善するためのヒューリスティックを提案した．Sadayappan and Visvanathan [10] は，次の行が収まる最初の位置に次の行を収めようとする別の貪欲アルゴリズムを検討したが，これも実用的な貢献しかしていない．行列圧縮以外にも，Bloom フィルタの圧縮 [3–5] や，トライの圧縮 [11] の問題が発見されており，後者は Ziegler の方法を適用して，各トライノードの子ポインタ配列を単一の配列に折りたたむことができる．

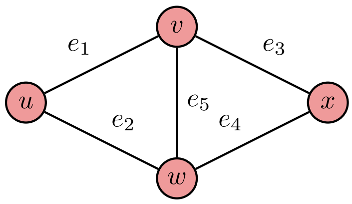
2 3 彩色問題からの帰着

以上は，Garey and Johnson [8] の NP 完全の仮定を詳細に証明する．

定理 1. SMC は *NP* 困難であり，最大シフトが $k = 3$ で上限されている場合でも *NP* 困難である．

証明. 頂点 $V := \{v_1, \ldots, v_n\}$ と m 個の辺 $E := \{e_1, \ldots, e_m\}$ を持つグラフ (V, E) が与えられたとする．以上は隣接行列のような行列 M を構築する．M は $n \times 3m$ 行列になる．M の行と列はそれぞれに V と E を表す．列を長さ 3 のグループに分割し，各グループに辺を割り当てる．サイズ 3 は，対応する辺で接続された頂点に割り当てられた色を符号化する．詳細は以下の通りである（例 2 は具体的な問題を表す）．まず，M をゼロで初期化する．次に，各行が頂点の色を符号化するようにしたい．このために，i 番目の行の $3m$ 列をサイズ 3 のブロックに分割し，j 番目のブロックの最初のマスに，v_i が e_j と隣接している場合には「1」と書く．意味は，ブロックの 3 つのマスを赤，青，緑の色の割り当てと考えると，最初はすべての頂点を赤で塗る．i 番目の行を 1 または 2 だけシフトすることで，v_i の色を緑または青に変更する．このようなシフトは，v_i のすべてのブロックの「1」を同時にシフトするため，色の割り当てが一意に定まる．問題 1 を解決するためには，各行を $\{0, 1, 2\}$ のオフセットでシフトする必要があり，各列の「1」マスが 1 以下であると確認する必要．このような解が存在するのは，グラフが 3 彩色可能である場合に限る．3 つの列のグループは，v_i と v_j が異なる色を持つ必要があるという制約は辺 $e = (v_i, v_j)$ に割り当てられる． □

例 2. 頂点 $V = \{u, v, w, x\}$ と辺 $E = \{e_1, \ldots, e_4\}$ を持つグラフが与えられたとする．



頂点 V の数を $n = |V|$，辺 E の数を $m = |E|$ とすると，n 行 $3m$ 列の行列 A を構築する．A は表 *1* に示されている．*3* つの列のグループごとに辺に割り当てられ，各行が頂点に割り当てられる．

表 *1* の A に対して，u と x の行を *1* つずつ，v の行を *2* つシフトすることで，各列には最大で *1* つの ‘1’ を持つ行列 B が得られる．B は表 *2* に示されて

$A =$

	e_1			e_2			e_3			e_4			e_5		
u	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0
v	1	0	0	0	0	0	1	0	0	0	0	0	1	0	0
w	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0
x	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0

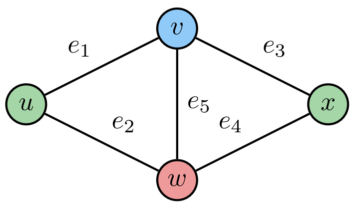
表 1: 例 2 の行列 A

$B =$

	e_1			e_2			e_3			e_4			e_5		
u	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
v	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
w	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0
x	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0

表 2: 例 2 の行列 B

いる．すべての行の最大シフトが 2 以内であるため，グラフ (V, E) は 3 彩色可能である．B は色の割り当てを符号化しており，グラフ上で次のように描写できる．



□

残念ながら，この帰着は幅 ℓ に制限を設けておらず，入力グラフの辺の数の 3 倍の幅に設定されている。したがって，最大で $3m$ になり得る．3 彩色問題は，出次数が最大 4 のグラフに対しても NP 困難であることが証明されている [7] ので、幅 $\ell \leq 12n$ を対応する必要がある。使用ケースでは，幅が対数的なサイズ（例えば，[11] で研究されているトライのような対数的なアルファベットサイズ）になることがあるため，これは不十分である．しかしながら，3 彩色に基づく，辺を配置して非ゼロのマスの領域を制限できる方法を見つけることができれば，この問題を解決できるかもしれない．我々の主な結果は 3 彩色問題の代わりに，制限付きの有向ハミルトンパス問題からの帰着により，対数的な行列幅においても NP 困難であることを示すことである [1].

定理 2. $\ell \in \Omega(\lg n)$ に対して，SMC は *NP* 困難である．

参考文献

[1] Hideo Bannai, Keisuke Goto, Shunsuke Kanda, and Dominik Köppl. NP-completeness for the space-optimality of double-array tries. *CoRR*, abs/2403.04951, 2024. doi: 10.48550/ARXIV.2403.04951.

[2] Chin-Chen Chang and Daniel J. Buehrer. An improvement to Ziegler's sparse matrix compression algorithm. *J. Syst. Softw.*, 35(1):67–71, 1996. doi: 10.1016/0164-1212(95)00086-0.

[3] Chin-Chen Chang and Tzong-Chen Wu. A letter-oriented perfect hashing scheme based upon sparse table compression. *Softw. Pract. Exp.*, 21(1):35–49, 1991. doi: 10.1002/SPE.4380210104.

[4] Chin-Chen Chang, Huey-Cheue Kowng, and Tzong-Chen Wu. A refinement of a compression-oriented addressing scheme. *BIT Numerical Mathematics*, 33(4):529–535, 1993. doi: 10.1007/BF01990533.

[5] Peter C. Dillinger, Lorenz Hübschle-Schneider, Peter Sanders, and Stefan Walzer. Fast succinct retrieval and approximate membership using ribbon. In *Proc. SEA*, volume 233 of *LIPIcs*, pages 4:1–4:20, 2022. doi: 10.4230/LIPICS.SEA.2022.4.

[6] S. Even, D.I. Lichtenstein, and Y. Shiloah. Remarks on Ziegler's method for matrix compression. unpublished, 1977.

[7] M. R. Garey, David S. Johnson, and Larry J. Stockmeyer. Some simplified NP-complete graph problems. *Theor. Comput. Sci.*, 1(3):237–267, 1976. doi: 10.1016/0304-3975(76)90059-1.

[8] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-completeness*. A Series of books in the mathematical sciences. Bell Laboratories, 1979.

[9] Roberto Grossi, Dominik Koeppl, Vincent Limouzy, Giulia Punzi, Takeaki Uno, and Kunihiro Wasa. Sparse matrix compression with small widths, private communication. unpublished, 2024.

[10] P. Sadayappan and V. Visvanathan. Efficient sparse matrix factorization for circuit simulation on vector supercomputers. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, 8(12):1276–1285, 1989. doi: 10.1109/43.44508.

[11] Robert Endre Tarjan and Andrew Chi-Chih Yao. Storing a sparse table. *Commun. ACM*, 22(11):606–611, 1979. doi: 10.1145/359168.359175.

[12] S. F. Ziegler. Small faster table driven parser. Technical report, Madison Academic Computing Center, University of Wisconsin, 1977. unpublished.