数理解析研究所講究録 1465 巻 2006 年 98-106

Error-Correcting Maximal 2-Consecutive Positive Detectable Matrix

名古屋大学　情報科学研究科　籾原　幸二 (Koji Momihara)
Graduate School of Imfomation Science,
Nagoya University

1. Introduction

$C=\{c_1,\ldots,c_n\}$ を item の集合とし、$\sigma:C\rightarrow\{0,1\}$ を各 item の状態を表す写像であるとする。もし $\sigma(c_i)=1$ ならば、c_i は positive であるといい、さもなければ negative であるという。DNA library screening において（この場合、各 item は clone に対応する）、C のすべての positive な item を決定することが必要とされる。グループテストの数を減らすために、group もしくは pool と呼ばれる C の部分集合 P を選び、pool P に対してテストを行う。もし pool P が少なくとも一つの positive item を含んでいるとき、pool P は positive であるといい、さもなければ negative であるという。positive な item を決定するためにさまざまな pool を選び、それらを test するプロセスとして定義される手法は group testing として知られている [1]。遂次 pool を作ってテストを行いながら、その結果を見て次の pool を作る方法を adaptive と呼び、あらかじめすべての pool を作っておく方法を nonadaptive と呼ぶ。DNA library screening では実験者と解析を行う人が異なるため adaptive な方法は難しい。このとき、n 個の item と m 個の pool をもつ nonadative group testing は $\{0,1\}$ 上の $m\times n$ 行列 $H=(h_{ji})$ として表現され、それをその group testing の incidence matrix という。ここで、H の各列はその item に対応し、各行はその pool に対応する。また、$h_{ji}=1(1\leq j\leq m, 1\leq i\leq n)$ は j 番目の pool が i 番目の item c_i を含んでいることを意味している。

Colbourn [1] は集合 C にある線形順序 $c_i\prec c_{i+1}(1\leq i\leq n)$ が定義され、positive item の集合は順序 $\prec$ の上で連続した集合であり、高々 s 個の positive item を含んでいるという性質 (s-consecutive positive property) をもっている状況を考えた。特に、Jimbo,Mueller [2] は $s=2$ の場合に対して、H のすべての列ベクトルと H の二つの連続した列の論理和によって得られるベクトルがすべて異なれば、任意の二つの連続した positive item を発見できるというアイデアからそのような行列 H を 2-consecutive positive detectable matrix (単に 2CPDM) と定義した。また、各 item は同じ再帰数を持っていることが望ましいことから、各列の重みが一定である 2CPDM を構成した。

DNA library screening において group testing の検査結果には誤りがあることが多く、そのために誤り訂正能力を持つ group testing が必要である [1][2][3][4]。それゆえ、Jimbo,Mueller は 2CPDM の概念を誤り訂正符号に拡張し、H の列ベクトルと論理和ベクトルが誤り訂正符号を成す 2CPDM を定義した。特に、各 item の再帰数が一定であり、符号の最小距離が 2 である列数最大の 2CPDM を構成した [3]。

ここでは、各列の重みが 3 でかつ非線形符号としての最小距離が 3 である、列数が最大の 2CPD-matrix の存在を示す。

2. Error-Correcting Maximal 2CPD-Matrix の存在

まず、導入で挙げた 2-consecutive detectable matrix の定義を行う。ここで、$\vee$ は論理和を意味する。つまり、$0 \vee 0 = 0, 0 \vee 1 = 1 \vee 0 = 1, 1 \vee 1 = 1$ である。

定義 1　$H = [x_1, x_2, \ldots, x_n]$ を列ベクトル x_i をもつ $GF(2)$ 上の $m \times n$ 行列とする。また、y_i を $y_i = x_i \vee x_{i+1}, 1 \leq i \leq n-1$ と定義する。

このとき、$x_1, \ldots, x_n, y_1, \ldots, y_{n-1}$ がすべて異なるベクトルならば、H は 2-consecutive positive detectable matrix (2CPD-matrix) といい、2CPDM(n) で記す。また、2CPD-matrix H の各列の重み k が一定であるとき、H を 2CPDM(n, k) と記す。

さらに、$C' = \{x_1, \ldots, x_n, y_1, \ldots, y_{n-1}\}$ が最小距離 d の誤り訂正非線形符号を成すとき、H は Error-correcting 2-consecutive detectable matrix といい、2CPDM($n \mid d$) と記す。また、2CPDM($n \mid d$) の各列の重み k が一定であるとき、H は 2CPDM($n, k \mid d$) と記す。

H を 2CPD-matrix とする。そのとき、$m \times (2n-1)$ 行列 $H^{\vee}$ を $H^{\vee} = [x_1, y_1, x_2, y_2, \ldots, x_{n-1}, y_{n-1}, x_n]$ と定義する。また、与えられた行のサイズ m と各列の重み k(一定) に対し 2CPD-matrix の列数が最大となるとき、その 2CPD-matrix は maximal であるといい、M2CPDM(n, k) と記す。

導入で挙げたように Jimbo,Mueller は M2CPDM($n, k \mid 2$) の存在について以下を示した [3]。特に $d = 2$ の場合、2CPDM($n, k \mid 2$) は列数が $\binom{n}{k}$ のとき maximal となる。

命題 1　次のパラメータ k と n_k に対して、$n \geq n_k$ を満たすすべての n に対して、M2CPDM($n, k \mid 2$) が存在する。

k	2	3	4	5	6	7
n_k	6	8	11	12	17	19

表 1：M2CPDM$(n,k \mid 2)$

この節では、M2CPDM$(n,3 \mid 3)$ の存在について述べる。このとき、$d = 3, k = 3$ であるから 2CPDM$(n,3 \mid 3)$ の最大列数は明らかに、一定重み 3 でかつ最小距離 3 の符号の最大符号語数となる。この存在を示すためにいくつかの概念の導入をする。

定義 2　$m > k > 1$ とする。X を m 個の元 (点という) からなる集合とし、$\mathcal{B}$ を X の k-部分集合からなる集合族ですべての t-部分集合が高々1 つの block に生起するものとする。このとき、$(X, \mathcal{B})$ を t-$(m,k,1)$ packing という。

また、$D(m,k,t)$ を任意の t-$(m,k,1)$ packing の最大 block 数とし t-$(m,k,1)$ packing が $|\mathcal{B}| = D(m,k,t)$ を満たすとき、その packing は optimal であるという。optimal 2-$(m,3,1)$ packing の存在は表 2 のように知られている [8]。

$m \equiv$	$D(m,3,2)$
1, 3 (mod 6)	$(m^2 - m)/6$
0, 2 (mod 6)	$(m^2 - 2m)/6$
4 (mod 6)	$(m^2 - 2m - 2)/6$
5 (mod 6)	$(m^2 - m - 8)/6$

表 2：optimal 2-$(m,3,1)$ packing

optimal 2-$(m,3,1)$ packing の点と block のなす生起行列の列ベクトルの集合 C は明らかに一定重み 3 で最小距離 4 の符号語数最大の符号である。さらに、C の各符号語の重みの偶奇性より C は最小重み 3 の符号でもある。このことから、2CPDM$(m,3 \mid 3)$ の最大列数は optimal 2-$(m,3,1)$ packing の block 数に一致する。よって、2CPDM$(m,3 \mid 3)$ は列数が optimal 2-$(m,3,1)$ packing の block 数に一致するときに maximal である。

次に packing における block 0-intersection graph を定義する。

定義 3　G を packing$D = (X, \mathcal{B})$ の block に対応する頂点をもち、二つの block $B_1, B_2 \in \mathcal{B}$ が共通に点を含んでいないときに辺 $\{B_1, B_2\}$ をもつグラフとする。このとき、グラフ G を D の block 0-intersection graph という。

さらに任意のグラフにおけるハミルトンサイクルの存在は、Ore [6] によって命題 2 のように特徴付けされている。また、2-連結正則グラフにおけるハミルトンサイクルの存在は、Bondy,Kouider [5] によって命題 3 のように特徴付けされている。

命題 2　グラフ G が $n \geq 3$ 個の頂点をもち、任意の非隣接な二頂点の次数の和が少なくとも n 以上ならば、G はハミルトンサイクルをもつ。

命題 3　2-連結正則グラフ G が $n \geq 3$ 個の頂点を持ち、頂点の次数が少なくとも $(n-1)/3$ 以上ならば G はハミルトンサイクルをもつ。

さて、M2CPDM$(m, 3 \mid 3)$ の存在について述べる準備ができた。
ここで、二つのベクトルが共通の座標位置に成分 1 が生起しないとき、その二つのベクトルは disjoint であると定義する。

補題　D を optimal 2-$(m, 3, 1)$ packing とし、C を D の生起行列 N の列ベクトルの集合とする。ここで、C の連続した二つのベクトル x_i, x_{i+1} が disjoint であるように並び替えられるとき、$C' = \{x_1, \ldots, x_n, y_1 = x_1 \vee x_2, \ldots, y_{n-1} = x_{n-1} \vee x_n\}$ は最小距離 3 をもつ符号である。

証明：　D の任意の二つの block は高々一つの点で交差する。よって、C は最小重み 4 で重み一定の符号になる。さらに、もし C の連続した二つのベクトル x_i, x_{i+1} が disjoint であるように並び替えられるとき、x_i と $y_i = x_i \vee x_{i+1}$ はちょうど三つの座標位置で交差し、x_i と $y_j = x_j \vee x_{j+1} (j \neq i)$ に対しても明らかに高々三つの座標位置で交差する。最後に、$i \neq j$ なる y_i と y_j に対しても同様に高々四つの座標位置で交差する。このようにして C' は最小距離 3 をもつ。//

例 1　Maximal 2CPD-matrix$(10, 3 \mid 3)$

$$H = \begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1
\end{bmatrix}$$

例 2　Maximal 2CPD-matrix$(11, 3 \mid 3)$

$$
H = \begin{bmatrix}
0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}
$$

例 3　Maximal 2CPD-matrix$(12, 3 \mid 3)$

$$
H = \begin{bmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1
\end{bmatrix}
$$

定理　任意の $m \geq 10$ に対して maximal 2CPD-matrix$(m, 3 \mid 3)$ が存在する。

略証：　C を optimal 2-$(m, 3, 1)$ packing$D$ の生起行列の列ベクトルすべての集合とする。C が符号語数最大の最小距離 3 の符号であることと補題より、C のベクトルを連続した二つのベクトルが disjoint であるように並び替えられればよい。これは、C の block 0-intersection graph のハミルトンパスを見つける問題に置き換えられる。ハミルトンサイクルがあれば、明らかにハミルトンパスは存在するので、命題 2,3 を適用することができる。実際に、D の任意の block に対して点が共通に生起しない block の数を数え上げて命題

2,3 を適用すると、任意の $m \geq 13$ に対してハミルトンサイクルが存在することがわかる。$m = 10, 11, 12$ に対しては例 1,2,3 によって与えられる。//

このようにして、任意の $m \geq 10$ に対して maximal 2CPD-matrix$(m, 3 \mid 3)$ の存在が示された。ここで、$m \leq 9$ に対しては明らかに maximal 2CPD-matrix$(m, 3 \mid 3)$ は存在しないことに注意する。

3. Concluding Remark

Error-correcting 2CPD-matrix を item が s-consecutive positive property をもつ group testing における pooling design の incidence matrix として用いることで、pool の判定誤りを訂正し、かつ正しい positive item の位置を発見することができる。

応用的側面から見て 2CPD-matrix の具体的構成法が必要とされる。$m \equiv 0, 1, 2, 3 \pmod 6$ のときは cyclic triple system を使って簡単に構成できる。これらは、Appendix に提示する。また、$m \equiv 4, 5 \pmod 6$ のときの構成法については、考察中である。

Appendix

Case 1. $(m = 6k+1)$

point を群 $Z_m = \{0, 1, \ldots, m-1\}$ の元とし、また、$\{a, b, c\} + i, i \in Z_m$ を $\{a+i, b+i, c+i\}$ で定義する。ここで演算は Z_m 上で行うものとする。

$m = 18s+1, s \geq 3$ のとき

1. $\{0, 1, 4s+2\} + 2i(i = 0, 1, \ldots, 18s)$
2. $\{0, 3r+1, 4s+r+2\} + i(i = 0, 1, \ldots, 18s), r = 1, \ldots, s-1$
3. $\{0, 2, 8s+2\} + 3i(i = 0, 1, \ldots, 18s)$
4. $\{0, 3r+2, 8s+2r+2\} + i(i = 0, 1, \ldots, 18s), r = 1, \ldots, s-1$
5. $\{0, 3r+3, 6s+r+2\} + i(i = 0, 1, \ldots, 18s), r = 0, 1, \ldots, s-2$
6. $\{1, 3s+1, 6s+2\} + i(i = 0, 1, \ldots, 18s)$

$m = 18s+7, s \geq 3$ のとき

1. $\{0, 1, 8s+2\} + 2i(i = 0, 1, \ldots, 18s+6)$
2. $\{0, 3r+1, 8s+2r+4\} + i(i = 0, 1, \ldots, 18s+6), r = 1, \ldots, s-1$

3. $\{0, 2, 6s+3\} + 3i (i = 0, 1, \ldots, 18s+6)$
4. $\{0, 3r+2, 6s+r+3\} + i (i = 0, 1, \ldots, 18s+6), r = 1, \ldots, s-1$
5. $\{0, 3r+3, 4s+2r+4\} + i (i = 0, 1, \ldots, 18s+6), r = 0, 1, \ldots, s-1$
6. $\{0, 3s+1, 7s+3\} + i (i = 0, 1, \ldots, 18s+6)$

$m = 18s + 13, s \geq 3$ のとき

1. $\{0, 1, 4s+4\} + 2i (i = 0, 1, \ldots, 18s+12)$
2. $\{0, 3r+1, 4s+2r+4\} + i (i = 0, 1, \ldots, 18s+12), r = 1, \ldots, s$
3. $\{0, 2, 6s+5\} + 3i (i = 0, 1, \ldots, 18s+12)$
4. $\{0, 3r+2, 6s+r+5\} + i (i = 0, 1, \ldots, 18s+12), r = 1, \ldots, s-1$
5. $\{0, 3r+3, 8s+2r+8\} + i (i = 0, 1, \ldots, 18s+12), r = 0, 1, \ldots, s-1$
6. $\{0, 3s+2, 8s+6\} + i (i = 0, 1, \ldots, 18s+12)$

Case 2. $(m = 6k)$

Case 1 で得られた順序付けされた block 集合に対して、point0 と point0 を含んだ block の除去することで得られる。

Case 3. $(m = 6k+3)$

point を群 $Z_m = \{0, 1, \ldots, m-1\}$ の元とし、また、$\{a, b, c\} + i, i \in Z_m$ を $\{a+i, b+i, c+i\}$ で定義する。ここで演算は Z_m 上で行うものとする。

$m = 18s + 3, s \geq 4$ のとき

1. $\{0, 1, 8s+2\} + 2i (i = 0, 1, \ldots, 18s+2)$
2. $\{0, 3r+1, 8s+2r+2\} + i (i = 0, 1, \ldots, 18s+2), r = 1, \ldots, s-1$
3. $\{0, 2, 4s+2\} + 4i (i = 0, 1, \ldots, 18s+2)$
4. $\{0, 3r+2, 4s+2r+2\} + i (i = 0, 1, \ldots, 18s+2), r = 1, \ldots, s-1$
5. $\{0, 3r+3, 6s+r+2\} + i (i = 0, 1, \ldots, 18s+2), r = 0, 1, \ldots, s-1$
6. $\{0, 6s+1, 12s+2\} + i (i = 0, 1, \ldots, 6s)$

$m = 18s + 9, s \geq 5$ のとき

1. $\{0, 1, 4s+2\} + 2i (i = 0, 1, \ldots, 18s+8)$
2. $\{0, 3r+1, 4s+2r+4\} + i (i = 0, 1, \ldots, 18s+8), r = 1, \ldots, s$
3. $\{0, 2, 8s+5\} + 4i (i = 0, 1, \ldots, 18s+8)$
4. $\{0, 3r+2, 8s+2r+4\} + i (i = 0, 1, \ldots, 18s+8), r = 2, \ldots, s-2$

5. $\{0, 3r+3, 6s+r+4\} + i(i = 0, 1, \ldots, 18s+8), r = 1, \ldots, s-2$
6. $\{0, 5, 8s+7\} + i(i = 0, 1, \ldots, 18s+8)$
7. $\{0, 3s-1, 6s+1\} + i(i = 0, 1, \ldots, 18s+8)$
8. $\{0, 3, 8s+4\} + i(i = 0, 1, \ldots, 18s+8)$
9. $\{0, 3s, 8s+6\} + i(i = 0, 1, \ldots, 18s+8)$
10. $\{0, 6s+3, 12s+6\} + i(i = 0, 1, \ldots, 6s+2)$

$m = 18s + 15, s \geq 3$ のとき

1. $\{0, 1, 4s+4\} + 2i(i = 0, 1, \ldots, 18s+14)$
2. $\{0, 3r+1, 4s+2r+4\} + i(i = 0, 1, \ldots, 18s+14), r = 1, \ldots, s$
3. $\{0, 2, 8s+8\} + 4i(i = 0, 1, \ldots, 18s+14)$
4. $\{0, 3r+2, 8s+2r+8\} + i(i = 0, 1, \ldots, 18s+14), r = 1, \ldots, s$
5. $\{0, 3r+3, 6s+r+6\} + i(i = 0, 1, \ldots, 18s+14), r = 0, 1, \ldots, s-1$
6. $\{18s+12, 6s+2, 12s+7\} + i(i = 0, 1, \ldots, 6s+4)$

Case 4. $\quad$ **(m= 6k+2)**

Case 3 で得られた順序付けされた block の集合に対して、point0 と point0 を含んだ block の除去することで得られる。

References

[1] Colbourn, C. J.: Group testing for consecutive positives. Annals of Combinatorics 3 (1999), 37-41.

[2] Jimbo, M., Mueller, M.: Consecutive positive detectable matrices and group testing for consecutive positives. Descrete Math. 279, No. 1-3 (2004), 369-381.

[3] Jimbo, M., Mueller, M.: Cyclic sequences of k-subsets with distinct consecutive unions. submited (2004)

[4] Ngo, H. Q., Du, D. Z.: A survey on combinatorial group testing algorithms with applications to DNA library screening. DIMACS Series in Discrete Mathematics and Theoretical Computer Science/DU2, Providence, RI, 2000, Amer. Math. Soc.

[5] Bondy, J. A., Kouider, M.: Hamilton cycles in regular 2-connected graph. J. Combin (1988)

[6] Bose. R. C., Manvel. B.: *Introduction to Combinatorial Theory.* John Wiley & Sons. (1984)

[7] Colbourn, C. J.,Dinitz, J. H.: *The CRC handbook of Combinatorial Designs, eds.*, CRC Press, Boca Raton, USA. (1996)

[8] Lindner, C. C., Rodger, C. A.: *Design theory.*, CRC Press LLC, USA. (1997)