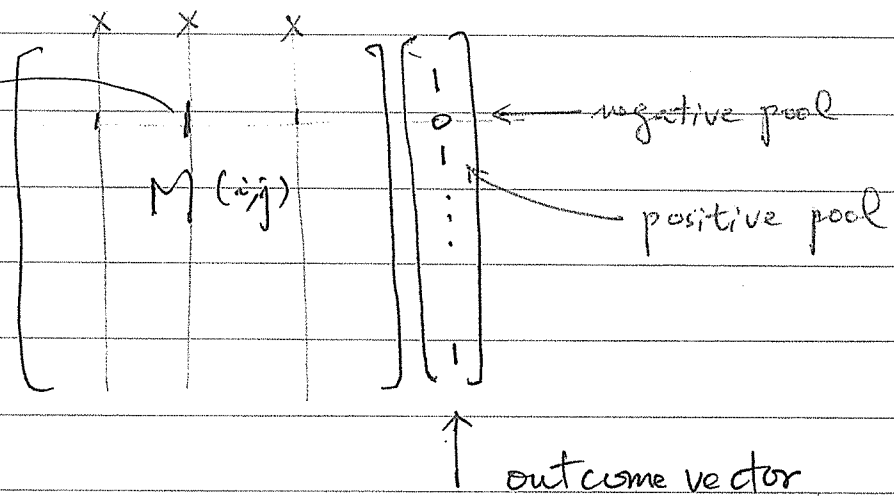


5, 4, 1, 2
①

C_j negative
↓

$M(i, j) = 1$
≡ the i th row
of M intersects the
 j th column.



- A test is ^(positive)negative if the corresponding component in the outcome vector is "0".
(1)
- negative (positive) test is also called negative (positive) pool.

觀察: Negative pool 所包含的 clones 全部是 "Negative"!

結果: 先利用 "negative" pools 把一些 clones 消去, 再從剩下的 clones 中找 d 個即可。

Hitting Set Problem

Given a set X and a collection $\mathcal{C}$ of subsets of X , find a minimum-cardinality subset Y of X such that $Y \cap C \neq \emptyset \forall C \in \mathcal{C}$.

Let M_D be a $k_D \times n_D$ matrix obtained from M by keeping only the set T_D of the positive pools and the set N_D of n_D clones not appearing in any negative pool. Then, the decoding is associated

Outcome Vector : Can be recognized as a set, V . (2)

Denote by $t_o^V(C) = |C \setminus V|$ where C is a column.

e.g.

	1	2	3	4	5	6	V		$t_1^V(C)$
t_1	1	1	1	0	0	0	1	$\{1, 2\}$	$= C \cap V $
t_2	1	0	0	1	1	0	0	$\{1, 4\}$	
t_3	0	1	0	1	0	1	1	$\{2, 3\}$	
t_4	0	0	1	0	1	1	1	$\{2, 6\}$	
		$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$	$\uparrow$		$\{3, 4\}$	
		C_1	C_2	C_3	C_4	C_5	C_6		

$V = C_2 \cup C_3$

Note that $t_o^V(C) = |C \setminus V|$, $t_1^V(C) = |C \cap V|$.
 $|C \setminus V| > 0 \Rightarrow C$ is a negative item.

$t_o^V(C_1) = 1$	$t_1^V(C_1) = 1$
$t_o^V(C_2) = 0$	$t_1^V(C_2) = 2$
$t_o^V(C_3) = 0$	$t_1^V(C_3) = 2$
$t_o^V(C_4) = 1$	$t_1^V(C_4) = 1$
$t_o^V(C_5) = 1$	$t_1^V(C_5) = 1$
$t_o^V(C_6) = 0$	$t_1^V(C_6) = 2$

$\uparrow$
 Negatives
 C_1, C_4, C_5

可解是 $\begin{array}{c} \textcircled{1} \textcircled{3} \\ \textcircled{2} \textcircled{6} \\ \textcircled{3} \textcircled{6} \end{array}$

如果是 2-disjunct 就不会留下 C_2, C_3, C_6 等待判断。

(•) We can use a d -disjunct matrix to identify up-to- d positive clones. ③

Theorem If a set of n clones which contains at most d positive then a d -disjunct matrix can be applied to identify the positive. Moreover, the decoding algorithm is $\hat{O}(n \log d)$ linear complexity.

Proof. The proof follows by evaluating $t_0^V(C)$ for each clone C . (V is the outcome vector.) We remark here that

" $t_0^V(C) = 0$ if C is positive and $t_0^V(C) \geq 1$ if C is negative. "

$t_0^V(C) = 0$ if C is positive and $t_0^V(C) \geq 1$ if C is negative.
 ($t_0^V(C) = |C \setminus V|$)
 Consider $C_0, \overset{\text{positive}}{C_1, \dots, C_d}, \parallel \overset{\text{positive } C_0}{\text{high - } 2\text{-disjunct matrix}}$
 (说明) $V = \bigcup_{C \text{ is positive}} C$, 如果存在 C' 是 negative, 而 $|C' \setminus V| = 0$
 则 $C' \subseteq V$, 这 $\Rightarrow$ d-disjunct 的假设不符。

Algorithm (Clone Selection ($N, V, D(e)$)) 指定 $e \geq 0$

1. for each clone $C \in N$, ($D \leftarrow \emptyset$)
2. compute $\pi_o^V(C)$
3. if $\pi_o^V(C) \leq e$
4. then $D \leftarrow D \cup \{C\}$
5. return

$$V = UC$$

C positive

(*) 上面定理用 clone selection (N, V, D, ϕ) 。

✓ Key point If C is positive, then $f_0^V(C) = 0$; and
(V is the outcome vector.)
 C is negative, then $f_0^V(C) \geq 1$.

(4)

Definition $(d; z)$ -disjunct matrix.

A matrix is said to be $(d; z)$ -disjunct if for any $d+1$ columns $C_0, C_1, \dots, C_d$,

$$\left| C_0 \setminus \bigcup_{i=1}^d C_i \right| \geq z.$$

Note $(d; 1)$ -disjunct $\approx$ d -disjunct.

(*) There exist at least z rows in each of which C_0 has a "1"-entry and every C_i , $1 \leq i \leq d$, has a "0"-entry.

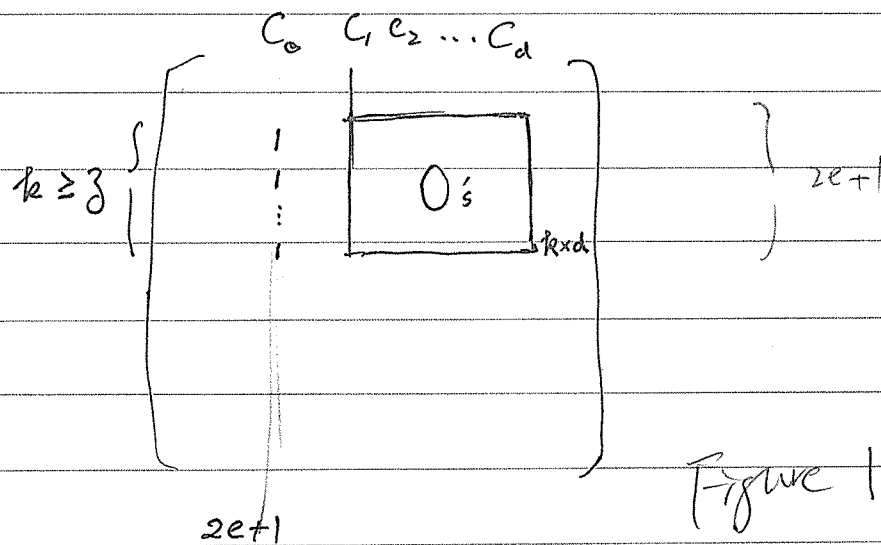


Figure 1

Theorem A $(d; z)$ -disjunct matrix can be applied to identify up-to- d positives with at most e errors.

Proof We remark that the errors occur on the conclusion of tests, i.e., the outcome vector (as a codeword) has at most e errors. (Due the fact that errors are possible, we shall have a set of possible outcomes, instead of just one outcome vector.) Now, assume the number of errors is e . For each negative clone $\bar{C}$, by definition, there exist at least $ze+1$ rows intersecting $\bar{C}$ but none of D (the set ^{of} positives). This implies that for each outcome

⑤

说明:

由於 Outcome vector 是所有 Positive clones 的
 交集; 如果用 $(d; 2e+1)$ -disjunct 矩陣來測試的
 話, 對於所有 positive clone C_i^+ , $t_1^V(C_i^+) \geq 2e+1$ (沒
 有任何 Errors 產生的情況)。這可以從上頁 Figure 1
 看出來: 把 C_0 當成 $C(\text{positive})$, $|V| \geq 2e+1$ 。

如果把 positives 都放入 $\{C_1, C_2, \dots, C_d\}$ 而 C_0 為
 negative, 則 $t_0^V(C) \geq 2e+1$ 。

於是在有 e 個錯誤的情況下 $t_0^V(C^-) \geq e+1$;
 另外 $t_0^V(C^+) = 0$ (沒有錯誤), 在有 e 個錯誤
 $\parallel$
 $|C^+ \setminus V| = |\emptyset|$

的情況下, $t_0^V(C^+) \leq e$ 。

上述的說明也提供了 Linear Time Decoding Algorithm.

$$\forall \text{ } t_0^V(C^+) \geq e+1 \quad (*)$$

(negative)

⑥

On the other hand, for each positive clone C^+ , $t_0^V(C^+) \leq e$.
 ($t_0^V(C) = 0$ if no errors occurred.) So, the proof follows by checking $n(=|N|)$ clones by using the function $t_0^V(C)$ where V is the outcome vector (in V). Note, there is exactly one outcome vector after the experiment is done. ■

Definition (Inhibitor)

A clone is an inhibitor if the clone neutralize positive clones, that is, the presence of an inhibitor in a pool dictates a negative outcome, regardless of the presence of positive clone in the pool.

(註) Inhibitor 出現時, 該 pool 是陰性反應. 加入 ⑥'

Theorem (Error-tolerant inhibitor model)

A $(d+h; 2e+1)$ -disjunct matrix can be applied to identify all positive clones if there are at most d positives and h inhibitor.

Proof. Observe that a clone C which appears in at most e positive pools, i.e., $t_1^V(C) \leq e$, cannot be positive due to the $(d+h; 2e+1)$ -disjunctness property. (If C is positive, then $t_1^V(C) \geq 2e+1$ if there are no errors.) For the inhibitors, each inhibitor can occur in at most e positive pools (with e errors), i.e., $t_1^V(I) \leq e$. $t_1^V(C) \leq e$.

Now, let O be the set of columns C such that $t_1^V(C) \leq e$. This implies that O contains all inhibitors but no positives. Hence, D , the set of positive clones is a subset of $N \setminus O$.

⑥'

在加入 Inhibitors 之後, 我們並不需要去找到那些 Clones 是 Inhibitors.

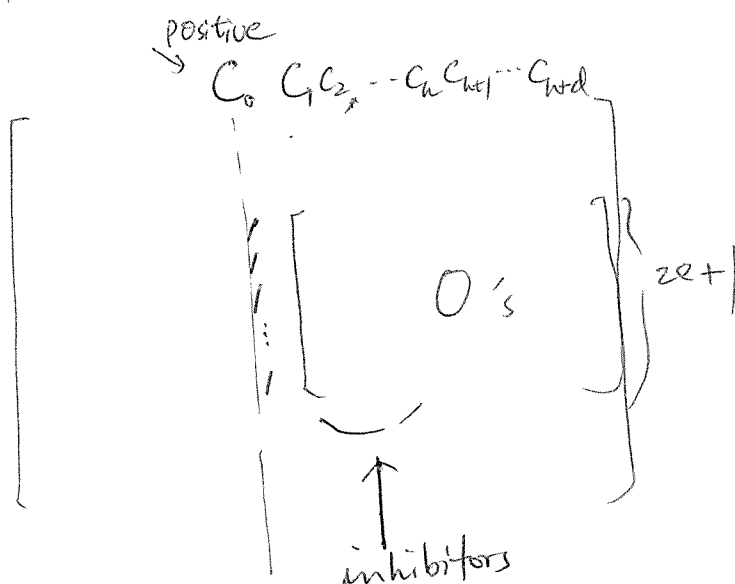
概念: 先把不可能是 positive clones 的 clones 找出來, 形成集合 O . 這個集合中可能有一些 negative clones.

這集合 O 可以由 positive pools 組成。

Reason

由於用了 $(d+h; 2e+1)$ -disjunct 矩陣,

$$t_1^V(c) \geq 2e+1 \quad (\text{沒有 Errors}).$$



由於 Inhibitors 不加入這些 pools, 所以 $t_1^V(c) \geq 2e+1$.

$\Rightarrow$ 如果有 e 個錯, $t_1^V(c) \geq e+1$.

⑦

Let $\bar{C} \in N \setminus O$. By disjointness property, we have $t_o^V(\bar{C}) \geq (2e+1) - e = e+1$. (?) On the other hand, if C^+ is a positive clone in $N \setminus O$, then $t_o^V(C^+) \leq e$. (?) Therefore, by calculating $t_o^V(C)$ where $C \in N \setminus O$, we conclude the proof. ■

说明：在这个定理中我们只负责把 positives 找出来，基本上不知道那几个 clones 是 inhibitors。

说明'：(d+h, z)-disjointness 是用来把 Inhibitors 分出来；如果没有 inhibitors (d, z)-disjointness 就够用了。

Group Testing on Complexes

In the complex model, we consider a set N of n clones and the positives are in an unknown family $P = \{P_i\}$ of subsets of N .
(known)

(有 positive outcome 要由一些 clones 所组合而成，不再是单一 clone 而已。)

(•) Each member of P is called a ^{positive} complex; (each subset of N is a candidate of a member of P .)

✓ (•) An experiment is applied to an arbitrary subset $S \subseteq N$ with two possible outcomes; a negative outcome implies S does not contain any $P_i \in P$, and a positive outcome implies otherwise

Complex Model

- 我们必需假设当 X, X' 皆为 complexes 时, 彼此没有互相包含的关系, 亦即 $X \not\subseteq X'$ 且 $X' \not\subseteq X$.

- Or the model must be monotone on positives, i.e., if X is positive and $X \subseteq X'$, then X' is also positive. But, this is not quite true in applications.

★ If not, and X contains a proper subset X^+ which is positive then X can only appear in positive pools no matter it is positive or negative. Thus, X cannot be identified.

- Let H denote the given set of complexes. Then H can be viewed as a hypergraph with clones as vertices and complexes as edges. Accordingly, complex model is dealing with searching a hidden graph (subgraph) P in a given hypergraph H . Clearly, P is induced by the set of positive edges.

- The formidable complexes of eukaryotic DNA transcription and RNA translation could involve hundred of molecules.

- Application: protein-to-protein interaction (Many) (Lappe and Holm, 2004).

(*) (Creating a protein-protein interaction network for

9

- A complex contains at most r clones! (Assumption)

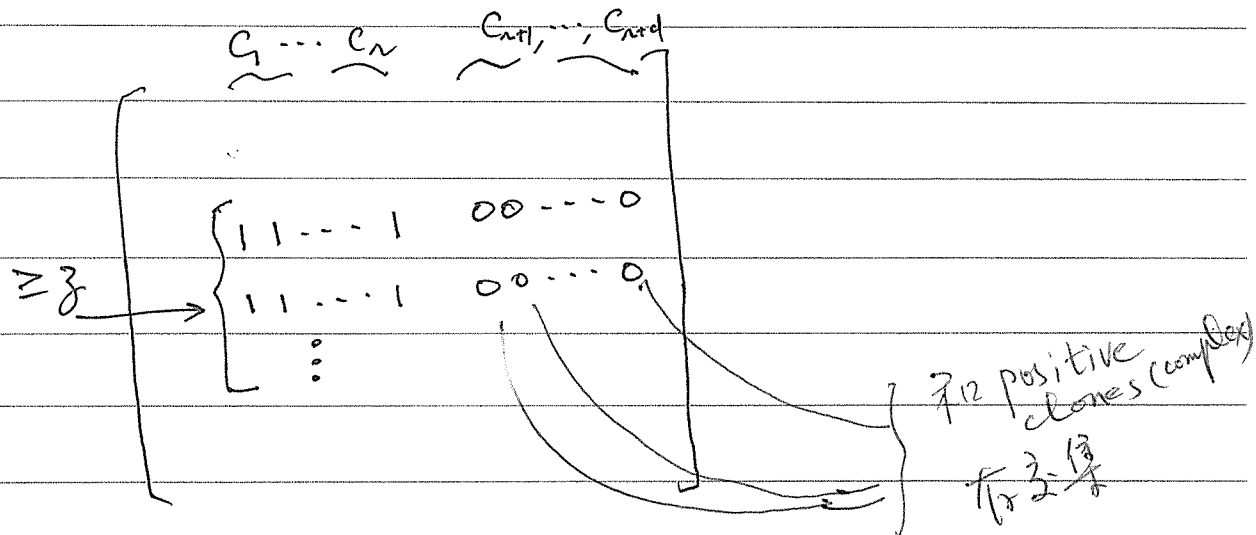
(Generalized cover-f

Definition $(d, r; z]$ -disjunct) Stinson and Wei, 2009 family

A matrix is $(d, r; z]$ -disjunct if for any $r+d$ columns

$C_1, C_2, \dots, C_{r+d}$,

$$\left| \bigcap_{i=1}^r C_i \setminus \bigcup_{i=r+1}^{r+d} C_i \right| \geq z.$$



Note that if $r=1$, then we use $(d; z]$ -disjunct for short.

Using a $(d, r; z]$ -disjunct matrix

Proposition For a negative complex X , there are at least z rows containing X but no positive complexes.

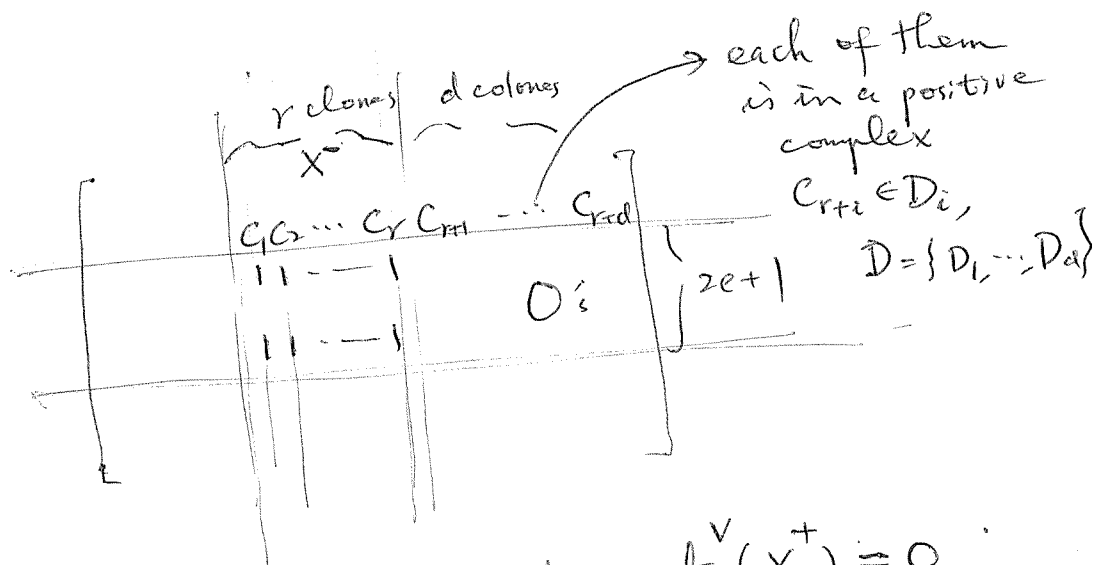
For a complex X ,

Proof. $t_o^V(X) =_{\text{def}} |\cap X \setminus V|$ and $t_1^V(X) =_{\text{def}} |(\cap X) \cap V|$.

- We can use $(d, r; 2e+1)$ -disjunct matrix to identify up-to- d positives with at most e errors.

Theorem

A $(d, r; 2e+1)$ -disjunct matrix can be applied to identify all d (at most) positive complexes with at most e errors.

Proof.

By the definition of the matrix $t_0^V(X^+) = 0$ (with no errors). So, if there are at most e errors,

$$t_0^V(X^+) \leq e.$$

On the other hand $t_0^V(X^-)$ will be at least $2e+1$ if we let $\{c_1, \dots, c_r\}$ be a negative complex (no errors) and $c_{r+i} \in D_i$ where D_i is a positive complex in D .

Now, even there are e errors, $t_0^V(X^-) \geq e+1$.

If there are h inhibitors, then we use a

$(d+h, r; 2e+1)$ -disjunct matrix.

(Ideal 和前面不是 complex model 和 100%)

14

Threshold Group Testing [Damarschke, 2005]

Let l and u be nonnegative integers $l < u$, called the lower and upper threshold. A group test (threshold model) for a subset S of clones is positive if S contains at least u positives, and negative if at most l positives are present. Furthermore, if the number of positives in S is between l and u , the outcome can give an arbitrary answer either positive or negative.

$g = u - l - 1$ is called the gap.

Note: Use $(d, u; z)$ -disjunct matrices to identify the positives with at most e errors. ($z \geq 2e+1$)

目前能够做的部分 $\rightarrow (d-u+1, u; 2e+1)$ -disjunct
 $\rightarrow (g=0) \quad l=u-1$

We have stop somewhere! Thank you for your patience. ^_^

补充: (d, c) -disjunct $\bigcup_{i=1}^c C_i \neq \bigcup_{i=c+1}^{d+c} C_i$

另外的用途?