析取矩阵的包含变体的渐近速率界与构造

Yuto Mizunuma千叶大学信息学研究生院日本千叶县稻毛区弥生町 1-33，邮编 263-8522电子邮箱：mizunuma@chiba-u.jp

Yuichiro Fujiwara千叶大学信息学部日本千叶县稻毛区弥生町 1-33，邮编 263-8522电子邮箱：yuichiro.fujiwara@chiba-u.jp

arXiv:2601.09362v1 [cs.IT] 14 Jan 2026

摘要——**析取矩阵也称为覆盖自由族和叠加码，是广泛应用于群体测试的组合数组。在其各种变体中，满足一种称为包含性的额外组合性质的变体构成了一个特殊类别，适用于一般抑制剂复合体模型下计算高效且容错能力强的群体测试。该模型是一个广泛的框架，涵盖了 DNA 筛查等实际应用场景。尽管具有这一重要意义，析取矩阵包含变体的渐近行为仍鲜有研究。特别是，此前尚不清楚该变体能否实现渐近正速率，而这是可扩展群体测试设计的必要条件。本文首次建立了该包含变体最大可达率的非平凡渐近下界，与已知最强上界仅相差一个对数因子。我们的证明基于概率方法，并给出一种简单高效的随机化构造。此外，我们对该构造进行去随机化，得到一种确定性的多项式时间构造。这些结果阐明了一般抑制剂复合体模型下稳健且可扩展的群体测试所具有的渐近潜力。**

索引词—**去随机化、不相交矩阵、容错群体测试、一般抑制剂复合体模型、概率方法。**

I. 引言

群体测试是一种框架，可在大量项目中高效识别少量目标项目，称为缺陷项。其核心思想是将项目放入少量池中，并对池进行测试，而不是逐个测试项目。群体测试由 Dorfman 于 1943 年提出，旨在以经济的方式大规模筛查应征入伍者的血液是否感染梅毒 [1],此后，群体测试已应用于诸多领域，例如生物学 [2], [3], 神经科学 [4], 药物发现 [5],网络 [6],压缩感知 [7],以及近来的 COVID-19筛查 [8]。

本文聚焦于非自适应群体测试 [3], [9],，与其自适应对应方案 [10], [11],不同，在这种测试中，池会预先固定，且无法根据先前的测试结果进行调整。非自适应群体测试在每个池的测试耗时较长、但可以并行进行时尤其有用。

本研究得到 JST SPRING资助项目 JPMJSP2109（Y.M.）和日本学术振兴会科研费 JP24K14815（Y.F.）的支持。

在非自适应分组测试中，称为 不相交矩阵 的特殊组合数组被广泛用于高效且稳健的池化设计 [9]。事实上，在经典设置下，合适的$t \times n$ 不相交矩阵能够帮助我们设计 t 池，并根据 t 测试结果，在时间 $O(tn)$内推断出 n 项目中所有缺陷项的位置，即使部分测试结果存在错误 [9], [12], [13]。

群体测试的一个核心目标是用少量测试处理大量项目。理想情况下，测试次数应与总体规模的对数成正比。因此，借用编码理论中的术语，我们将码率 $\log_t^2 n$ 视为 $t \times n$ 不相交矩阵测试效率的关键参数，并希望构造具有较高速率的不相交矩阵。如第二节所述，传统分离矩阵能够达到渐近正的速率 [14], [15], 这意味着，在经典群体测试模型中，我们只需进行与总体规模的对数成正比的测试次数，就能识别出所有缺陷项。

文献中提出了各种群体测试模型 [3], [9]。经典群体测试模型假设，每个阳性结果都由单个项目触发；在不存在测试错误的情况下，当且仅当一个池中至少包含一个有缺陷项目时，测试结果才为阳性。然而，许多实际应用无法完全纳入这一经典框架。例如，在关注具有特定属性的项目组合的应用中，目标是在一组复合体中识别目标，即从总体的非空子集中进行识别，而不是从单个项目中识别 [16]–[18]。这正是将经典群体测试模型推广为复合体模型 [19]的动机。

经典设置无法处理的另一个例子，是存在被称为抑制剂的棘手项目或复合体；它们在一个池中的存在可能会掩盖或抑制原本由缺陷项导致的阳性结果 [9, 第 8] 章。已报告此类现象的情境包括血液检测 [20], 药物发现 [5], 以及 DNA文库筛选 [21]。

群体测试中的抑制剂最早由 [2], 系统研究，该研究为某些分子生物学物质对群体测试产生的负面影响提出了一个简单模型。在涉及抑制剂的众多群体测试方法中，有一种尤为通用的框架，称为一般抑制剂

模型，由 [22], 提出。该模型假设可能存在抑制剂，但不提供关于抑制剂何时或如何导致假阴性的更多可利用信息。随后， [23]提出了一般抑制剂复合体模型，将复合体模型与一般抑制剂模型结合起来。在该模型中，缺陷项和抑制剂都被视为复合体；与一般抑制剂模型一样，该模型不对抑制剂的具体行为作出假设。由于一般抑制剂复合体模型能够处理广泛的实际应用，研究该模型的高效推断方法和理论局限性具有重要价值。

由于在这种一般设定下，抑制剂可能以任意对抗方式行动，因此设计有效的群体测试方案颇具挑战性。尽管如此，原则上，由于假阴性仍属于错误，我们可以采用专为无抑制剂群体测试设计的高容错方案来应对对抗性抑制剂。遗憾的是，到目前为止，这种方法只能得到计算开销极大的推断算法。事实上，假设复合体中的抑制剂至多有 h 个 [23]，传统容错可分离矩阵需要耗时 $O(tn^h)$ 来判定每个复合体是否有缺陷 $t\times n$。

为降低这一时间复杂度， [24] 为不相交矩阵引入了一种称为包含性的附加性质，该性质保证每个有缺陷对象仅能在数量有限的池中被抑制剂掩盖。借助不相交矩阵的包含式变体（本文称之为包含式不相交矩阵），判定每个复合体是否有缺陷所需的时间可降至 $O(t)$，比耗时 $O(tn^h)$ 的方案实用得多。因此，要高效应对行为未知的抑制剂，只需设计码率较高的包含式不相交矩阵。

尽管这一问题具有重要意义，包含式变体的渐近行为仍鲜有研究。尤其是，虽然应用通常需要较大的正速率，但此前尚不清楚包含式不相交矩阵这一类能否实现渐近正速率。

本文首次给出了包含式不相交矩阵最大可达率的非平凡渐近下界，该下界与目前已知的最强上界仅相差一个对数因子。我们的证明基于概率方法 [25] ，并给出一种简单高效的随机化构造。此外，我们通过条件期望方法 [26] 对该构造进行去随机化，从而得到一种确定性多项式时间构造。这些结果表明，包含式不相交矩阵这一类能够以高效构造实现渐近正速率，并阐明了一般抑制剂复合模型下稳健且可扩展的群体测试的渐近潜力。

在下一节中，我们将正式定义包含式不相交矩阵，并回顾已有结果。我们证明了包含式不相交矩阵最大可达率的首个非平凡渐近下界，并在第三节给出一种高效的随机化构造，随后在第四节对其进行去随机化。最后，第五节总结全文并给出一些评论。

II. 包含式不相交矩阵

在此，我们正式定义包含式不相交矩阵，并简要回顾已有结果。关于更全面的介绍，包括非自适应群体检测的基础理论以及生物信息学背景，请读者参阅 [3], [9], [15] 及其中的参考文献。

我们首先正式定义不相交矩阵。对于正整数 m，令 $[m]$ 表示集合 $\{1,2,\dots,m\}$。设$T=(t_{ij})$为定义在 $\{0,1\}$上的一个 $t\times n$ 二元矩阵。对于每个$j\in[n]$，令 $C_j:=\{i\in[t]:t_{ij}=1\}$ 表示第 j列向量的支撑集。对于不相交集合 $A,\ B\subseteq[n]$，定义

$$Z_T(A,B):=\left|\left(\bigcap_{j\in B}C_j\right)\setminus\left(\bigcup_{j\in A}C_j\right)\right|$$

为满足以下条件的行数：以 A 中的元素为索引的列中，所有 $|A|$矩阵元素均为 0，且以 B 中的元素为索引的列中，所有 $|B|$矩阵元素均为 1。对于正整数 d、 r和 z，若对于任意一对不相交集合 $A,\ B\subseteq[n]$ 满足 $|A|=d$ 和$|B|=r$，矩阵 T 称为（$d,\ r;z$)-不相交矩阵，且有 $Z(T,A,B)\geq z$。

不相交矩阵对应着各种情境中最受欢迎且最基础的群体测试方案 [3]–[9], [11], [12], [24], [27]。在编码理论中，它们也称为叠加码 [28]–[35] ，在极值组合学中则称为覆盖自由族 [14], [15], [36]–[41] 。关于其在群体测试以外的应用以及已知的存在性结果，感兴趣的读者可参阅近期的综述 [15]。

在无抑制剂的标准复合模型中，分别以测试和项目为行、列索引的$t\times n$($d,\ r;\ z$)-不相交矩阵，作为混合设计的关联矩阵，可用于判定每个复合项（非空项目集）是否有缺陷；假设最多有 d 个缺陷项、复合项大小至多为 r，且错误结果至多为 $\lfloor\frac{z-1}{2}\rfloor$ 个，则判定时间与 t成线性关系 [9], [12], [13]。

如上一节所述，不相交矩阵的码率越高越好。因此，群体测试的核心问题之一是理解最大可能列数的渐近行为

$$N(t,d,r;z]:=\max\left\{n:\ \begin{array}{l}\text{there is a } t\times n\\ (d,r;z]\text{-disjunct matrix}\end{array}\right\}$$

对于恰有 t 行的 ($d,r;\ z$)-不相交矩阵，以及相应的最大可达率

$$R(t,d,r;z]:=\frac{\log_2 N(t,d,r;z]}{t}$$

对于给定的正整数 t、 d、 r 和 z。

$R(t,\ d,\ r;\ z]$ 的渐近行为已得到广泛研究 [14], [27], [34], [35], [38], [41]–[52]。 [14,定理2.19、定理3.9] 给出的目前已知最佳渐近上界和下界表明，对于任意正整数$z\leq\frac{c_0}{d^r}t$，我们有

$$\frac{c_1+o(1)}{d^{r+1}}\leq R(t,d,r;z]\leq\frac{(c_2+o(1))\log_2 d}{d^{r+1}}\qquad(1)$$

当 $t \to \infty$时，其中 c_0、 c_1和 c_2 是与 t、 d和 z无关的正常数。这意味着，在无抑制剂的标准复合模型下，即使存在多达 $\lfloor \frac{z-1}{2} \rfloor$ 个错误，也只需进行 $t = \Theta(\log_2 n)$次测试即可识别所有缺陷项。

与理论发展成熟的传统分离矩阵相比，其包含式变体的渐近行为尚未得到充分理解。对于不相交集合 A, $B \subseteq [n]$，定义

$$Y_T(A,B) \coloneqq \left| \left(\bigcap_{j \in B} C_j \right) \cap \left(\bigcup_{j \in A} C_j \right) \right|$$

为：在这些行中，按集合 A 中元素索引的列里至少有一个 $|A|$矩阵元素为 1，且按集合 B 中元素索引的列里所有 $|B|$ 矩阵元素均为 1 的行数。对于正整数 r 以及非负整数 h和 y，二元矩阵 T 称为 $(h,r;y)$-包含式的，如果对于任意一对不相交集合 $A,B \subseteq [n]$，满足 $|A| = h$ 和 $|B| = r$，都有 $Y_T(A,B) \leq y$。对于正整数 d、 r 和 x，以及非负整数h，若存在整数 $z > 0$ 和 $y \geq 0$，满足 $z - y \geq x$，使得T同时为 $(d,r;z)$-分离的和 $(h,r;y)$-包含式的，则称 T为 $(d,h,r;x)$-包含式分离的。文献 [24] 表明，一个 $t \times n$ $(d,h,r;x)$-包含式分离矩阵能够在一般抑制剂复合体模型下高效识别所有缺陷项，其中缺陷项至多 d 个、抑制剂至多 h 个、复合项大小至多 r，且非由抑制剂引起的错误结果至多为$\lfloor \frac{x-1}{2} \rfloor$ 个。文献 [24] 中的推断算法能够在线性于 t 的时间内判定每个复合项是否有缺陷，相较于文献 [23], 的方法有显著改进；后者使用一个 $t \times n$ $(d+h,r;z)$-分离矩阵，所需时间为 $O(tn^h)$。

定义

$$N(t,d,h,r;x] \coloneqq \max \left\{ n : \begin{array}{l} \text{there is a } t \times n \ (d,h,r;x] \\ \text{-inclusively disjunct matrix} \end{array} \right\}.$$

与传统分离矩阵一样，可达到的最大码率

$$R(t,d,h,r;x] \coloneqq \frac{\log_2 N(t,d,h,r;x]}{t}$$

也是决定包含式变体效率的关键参数。以下引理源自 [24]，给出了 $R(t, d, h, r; x]$的上界。

引理 *2.1*（*[24,* 引理 *5.1]*）：每个（d, h, r; x]-包含式分离矩阵都是（$d+h$, r; x]-分离的。

将该引理与（1）结合，可得

$$\begin{aligned} R(t,d,h,r;x] &\leq R(t,d+h,r;x] \\ &\leq \frac{(c_2 + o(1))\log_2(d+h)}{(d+h)^{r+1}} \end{aligned} \tag{2}$$

对于 $x \leq c_0 (d+h)^r t$, 有 $t \to \infty$，其中 c_0 和 c_2 是与 t、 d、 h 和 x 无关的正常数。

尽管 [24] 的推断算法效率很高，但大多数已知的包含式分离矩阵都需要大量的行。对于 $r = 1$, 在 [53] 中首次构造为标准不相交矩阵的特殊集合系统，后来被证明是包含式分离的，见 [24, 引理 3.3、定理 3.4], ，由此得到目前已知的最佳下界

$$R(t,d,h,1;x] \geq \frac{\log_2 3}{16(d+h+x-1)^2} - \frac{\log_2 3 - 1}{t}.$$

若错误容忍参数 x 固定，则此下界意味着

$$R(t,d,h,1;x] \geq \frac{c + o(1)}{(d+h)^2}$$

对于一个与 t、 d、 h和 x无关的正常数 c，这在相差一个对数因子的意义下与上界（2）相匹配。然而，若 x 是关于 t的递增函数，则该下界渐近趋于零。

遗憾的是，对于一般的 $r \geq 1$, 大多数（d, r; z]）-不相交矩阵的构造都无法保证包含性。事实上，尽管 [24] 指出，若干（d, r; z]）-不相交矩阵的构造会生成（d, h, r; x]）-包含式不相交矩阵，但即使 x 为常数，其输出的速率仍渐近趋于零。难点在于，增强分离性往往会增加包含性旨在限制的重叠；因此，构造必须在这些相互竞争的约束之间取得平衡。这引出了以下自然问题：对于一般的 $r \geq 1$, （d, h, r; x]）-包含式不相交矩阵能否达到渐近正速率？若能，是否存在这类矩阵的高效构造？在下一节中，我们将给出肯定的解答。

III. 渐近下界与构造

在本节中，我们为 $R(t, d, h, r; x]$ 确立第一个非平凡下界，并给出一种高效的随机化构造，构造出可达到该界的 $(d, h, r; x]$-包含式不相交矩阵。我们通过概率方法证明以下定理。

定理 *3.1*：对于任意固定的正整数 r，存在仅依赖于 r 的正常数 c_0 和 c，使得以下结论成立。对于任意固定的整数 $d \geq 1$ 和 $h \geq 0$，以及任意正整数 t 和 $x \leq c_0(\max\{d,h\})^r t$,

$$R(t,d,h,r;x] \geq \frac{c + o(1)}{(\max\{d,h\})^{r+1}},$$

其中 $o(1)$ 是 t 的一个函数，满足当 $t \to \infty$时， $o(1) \to 0$。

该渐近下界与上界 (2) 仅相差一个对数因子。此外，我们的保证允许错误容忍参数 x 随 t线性增长；对于经典参数 z，这与 (1) 中关于 t 的线性增长区间仅相差一个常数因子。因此，我们的界是对 (1) 中给出的 $R(t, d, r; z]$ 渐近下界的自然推广。

为证明定理 3.1，我们使用两种标准的概率工具：马尔可夫不等式和*Chernoff*界 [26]。

引理 *3.2*（马尔可夫不等式）：设 X 是一个非负随机变量。那么，对于任意 $a > 0$,

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

引理 *3.3*（*Chernoff*界）：设 X_1、 X_2、.....、 X_n 是相互独立且取值于 $\{0,1\}$ 的随机变量，

并定义 $X \coloneqq X_1 + X_2 + \cdots + X_n$。则有以下结论。

1) For any $\delta \in (0,1)$,

$$\Pr[X \le (1-\delta)\mathbb{E}[X]] \le \exp\left(-\frac{\delta^2 \cdot \mathbb{E}[X]}{2}\right).$$

2）对于任意 $\delta > 0$,

$$\Pr[X \ge (1+\delta)\mathbb{E}[X]] \le \exp\left(-\frac{\delta^2 \cdot \mathbb{E}[X]}{2+\delta}\right).$$

为方便起见，我们引入违例集合的概念。设 $z > 0$、$y \ge 0$ 和 $x > 0$ 为满足 $z - y \ge x$ 的整数。对于定义在 $\{0,1\}$ 上的 $t \times n$ 二元矩阵 T，如果 $B \subseteq [n]$ 是一个 r-集，且存在一个满足 $Z_T(A,B) \le z-1$ 的 d-子集 $A \subseteq [n] \setminus B$，或者存在一个满足 $Y_T(A',B) \ge y+1$ 的 h-子集 $A' \subseteq [n] \setminus B$，则称其为 $(d,h,r;\ z,y)$-违例。如果 T 没有大小为 $[n]$ 的 $(d,h,r;\ z,y)$-违例子集，那么 T 同时是 $(d,r;\ z)$-分离矩阵和 $(h,r;\ y)$-包含式矩阵，因此它是 $(d,h,r;\ x)$-包含式分离矩阵，因为 $z - y \ge x$。

我们现在用一种称为修改法 [25]的强大技术来证明定理 3.1。这种概率组合学方法通过修改随机对象，获得所需结构。在下面的构造中，我们首先生成一个含有额外列的随机二进制矩阵。然后从每个违反条件的集合中选取一个元素，并删除由这些元素索引的所有列，从而得到所需的包含型不相交矩阵。

定理 *3.1* 的证明：固定绝对常数 $\varepsilon \in (0,\ 1)$、$\delta \in (0,\ 2^{\varepsilon} - 1)$ 和 $\sigma > 0$。令 $D \coloneqq \max\{d,h\}$。由于任何 $(D,\ D,\ r;\ x)$-包含式分离矩阵也是 $(d,\ h,\ r;\ x)$-包含式分离矩阵，因此只需构造一个 $(D,\ D,\ r;\ x)$-包含式分离矩阵。

设 n 为一个稍后确定的正整数。令

$$p \coloneqq 1 - 2^{-\frac{1-\varepsilon}{D}}.$$

由于对于任意 $u \ge 0$，$\exp(-u) \le 1 - u + \frac{u^2}{2}$，因此我们得到

$$\begin{aligned} p &= 1 - \exp\left(-\frac{(1-\varepsilon)\ln 2}{D}\right) \\ &\ge 1 - \left(1 - \frac{(1-\varepsilon)\ln 2}{D} + \frac{((1-\varepsilon)\ln 2)^2}{2D^2}\right) \\ &\ge \frac{(1-\varepsilon)\ln 2}{2D}. \end{aligned}$$

令 T 为一个 $t \times (1+\sigma)n$ 随机矩阵，其定义在 $\{0,1\}$ 上，各矩阵元素独立地以概率 p 取值 1，以概率 $1-p$ 取值 0。

根据期望的线性性，对于任意一对不相交集合$A,\ B \subseteq [(1+\sigma)n]$，满足 $|A| = D$ 和 $|B| = r$，我们有

$$\mathbb{E}[Z_T(A,B)] = p^r(1-p)^D \cdot t$$

and

$$\mathbb{E}[Y_T(A,B)] = p^r(1-(1-p)^D) \cdot t.$$

Set

$$z \coloneqq \lfloor (1-\delta)p^r(1-p)^D \cdot t \rfloor + 1$$

and

$$y \coloneqq \lceil (1+\delta)p^r(1-(1-p)^D) \cdot t \rceil - 1.$$

Then，我们有

$$\begin{aligned} z - y &\ge p^r(2(1-p)^D - 1 - \delta) \cdot t \\ &\ge \left(\frac{(1-\varepsilon)\ln 2}{2D}\right)^r (2^{\varepsilon} - 1 - \delta) \cdot t. \end{aligned}$$

令

$$c_0 \coloneqq \left(\frac{(1-\varepsilon)\ln 2}{2}\right)^r (2^{\varepsilon} - 1 - \delta)$$

并选取满足 $x \le \frac{c_0}{D^r} t$ 的正整数 x，即可得到 $z - y \ge x$。

根据 Chernoff界，我们有

$$\begin{aligned} &\Pr[Z_T(A,B) \le z-1] \\ \le\ &\Pr[Z_T(A,B) \le (1-\delta) \cdot \mathbb{E}[Z_T(A,B)]] \\ \le\ &\exp\left(-\frac{\delta^2 \cdot \mathbb{E}[Z_T(A,B)]}{2}\right) \\ =\ &\exp\left(-\frac{\delta^2 p^r (1-p)^D}{2} t\right) \end{aligned}$$

and

$$\begin{aligned} &\Pr[Y_T(A,B) \ge y+1] \\ \le\ &\Pr[Y_T(A,B) \ge (1+\delta) \cdot \mathbb{E}[Y_T(A,B)]] \\ \le\ &\exp\left(-\frac{\delta^2 \cdot \mathbb{E}[Y_T(A,B)]}{2+\delta}\right) \\ =\ &\exp\left(-\frac{\delta^2 p^r (1-(1-p)^D)}{2+\delta} t\right). \end{aligned}$$

回顾一下，r-子集 $B \subseteq [(1+\sigma)n]$ 是 $(D,D,r;\ z,y)$-违例，当且仅当存在一个满足以下任一条件的 D-子集 $A \subseteq [(1+\sigma)n] \setminus B$：$Z_T(A,B) \le z-1$ 或 $Y_T(A,B) \ge y+1$。定义 X 为 $[(1+\sigma)n]$ 的 $(D,D,r;\ z,y)$-违例子集的数量。接着，从 $[(1+\sigma)n]$ 的每个 $(D,D,r;z,y)$-违例子集中选取一个元素，并删除所有由这些元素索引的列，便得到一个 $t \times n'$ $(D,D,r;x)$-包含型不相交矩阵 T'，其中 n' 是满足 $(1+\sigma)n - X \le n' \le (1+\sigma)n$ 的整数。

根据期望的线性性和并集界，我们有

$$\begin{aligned} \mathbb{E}[X] &= \sum_{\substack{B \subseteq [(1+\sigma)n], \\ |B|=r}} \Pr[B \text{ is } (D,D,r;z,y)\text{-violated}] \\ &\le \sum_{\substack{B \subseteq [(1+\sigma)n], \\ |B|=r}} \sum_{\substack{A \subseteq [(1+\sigma)n] \setminus B, \\ |A|=D}} (\Pr[Z_T(A,B) \le z-1] \\ &\quad + \Pr[Y_T(A,B) \ge y+1]) \qquad (3) \\ &< 2(1+\sigma)^{D+r} n^{D+r} \exp\left(-\frac{\delta^2 p^r (1-(1-p)^D)}{2+\delta} t\right). \end{aligned}$$

根据马尔可夫不等式，

$$\begin{aligned} &\Pr[X \ge \sigma n] \\ \le\ &\frac{\mathbb{E}[X]}{\sigma n} \end{aligned}$$

$$< \frac{2(1+\sigma)^{D+r}n^{D+r-1}}{\sigma}\exp\left(-\frac{\delta^2 p^r(1-(1-p)^D)}{2+\delta}t\right).$$

因此，只要满足以下条件， X（D， D， r； z， y]-违例子集的数量$[(1+\sigma)n]$ 以正概率小于 σn

$$\frac{2(1+\sigma)^{D+r}n^{D+r-1}}{\sigma}\exp\left(-\frac{\delta^2 p^r(1-(1-p)^D)}{2+\delta}t\right)\le 1. \tag{4}$$

因此，当 n满足（4）时， T' 以正概率是一个 $t\times n'$(D， D， r； x]-包含型不相交矩阵，且 $n+1\le n'\le(1+\sigma)n$ 。

Fix

$$n=\left\lfloor\left(\frac{\sigma}{2(1+\sigma)^{D+r}}\right)^{\frac{1}{D+r-1}}\exp\left(\frac{\delta^2 p^r(1-(1-p)^D)}{(2+\delta)(D+r-1)}t\right)\right\rfloor$$

使得（4）成立。于是，存在一个 $t\times n'$(D， D， r； x]-包含型不相交矩阵，使得 $n+1\le n'\le(1+\sigma)n$。这意味着

$$\begin{aligned}
R(t,d,h,r;x] &\ge R(t,D,D,r;x]\\
&\ge \frac{\log_2 n'}{t}\\
&\ge \frac{\log_2(n+1)}{t}\\
&\ge \frac{\log_2\sigma-1-(D+r)\log_2(1+\sigma)}{(D+r-1)\cdot t}\\
&\quad+\frac{\delta^2 p^r(1-(1-p)^D)}{(2+\delta)(D+r-1)\ln 2}\\
&\ge \frac{(1+o(1))\delta^2(1-\varepsilon)^r(\ln 2)^{r-1}(1-2^{\varepsilon-1})}{2^r(2+\delta)(D+r-1)D^r}\\
&\ge \frac{(1+o(1))\delta^2(1-\varepsilon)^r(\ln 2)^{r-1}(1-2^{\varepsilon-1})}{2^r(2+\delta)rD^{r+1}}\\
&= \frac{c+o(1)}{D^{r+1}}
\end{aligned}$$

当 $t\to\infty$时，其中 c 是由下式定义的正常数：

$$c:=\frac{\delta^2(1-\varepsilon)^r(\ln 2)^{r-1}(1-2^{\varepsilon-1})}{2^r(2+\delta)r}.$$

证明完毕。 ■

上述证明给出了一种简单的包含式不相交矩阵蒙特卡罗构造。该构造具有较高效率：当 d、 h和 r固定时，其运行时间关于矩阵规模 tn为多项式时间 $O(tn^{\max\{d,h\}+r})$。

在该构造中，我们取满足（4）的 n ，使得 $\Pr[X\ge\sigma n]<1$。然而，这会导致列数以一定概率变得过小。为了降低这一失败概率，例如，应取满足下式的 n

$$\frac{2(1+\sigma)^{\max\{d,h\}+r}n^{\max\{d,h\}+r-1}}{\sigma}\cdot\exp\left(-\frac{\delta^2 p^r(1-(1-p)^{\max\{d,h\}})}{2+\delta}t\right)\le\frac{1}{t^s} \tag{5}$$

而不是式 (4)，其中 s 是任意固定的正常数。在这种情况下，由于失败概率 $\Pr[X\ge\sigma n]$ 小于 $\frac{1}{t^s}=o(1)$，蒙特卡罗构造以概率 $1-o(1)$ 成功。此外，所得矩阵达到了证明中的相同渐近速率保证。该构造的高概率版本如图1所示。

最后，我们指出，利用概率组合学中的另一种强大工具——*Lov´asz* 局部引理 [41], [54]，也可以推导出相同的下界。利用算法化 Lov´asz 局部引理 [55]，，可以在期望 $O(tn^{\max\{d,h\}+r+1})$ 时间内构造拉斯维加斯构造的包含式不相交矩阵，其码率达到该下界。

IV. 去随机化

这里，我们对定理 3.1 证明中的蒙特卡罗构造进行去随机化。

我们的去随机化基于一种称为条件期望方法 [26], [56],（也称为条件概率法 [25]，）的标准技术，该技术广泛用于高效构造组合对象，包括（超）图着色 [25], [56], 不相交矩阵 [49], 和完美哈希族 [57]–[59]。回顾一下，在定理 3.1 的证明中，我们将 X定义为随机矩阵 T 中 $[(1+\sigma)n]$的 $(D,D,r;z,y]$-违例子集数量，并选择 n 使得$\mathbb{E}[X]<\sigma n$。在去随机化过程中，我们贪心地固定 T的矩阵元素，力求使 X 的条件期望低于 σn。由于所得确定性矩阵中的 $(D,D,r;z,y]$-违例子集少于 σn (D)，按照定理 3.1 证明中的做法删除一些列，即可得到所需的矩阵。

条件期望方法通常用于对基于初等概率方法的算法进行去随机化 [25], [49], [56]–[59],而我们的蒙特卡罗构造基于修改法，该方法引入了一个额外的删除步骤。在许多情况下，删除步骤移除的列远少于 σn 列，从而得到的输出矩阵包含明显更多的列，超过了最坏情况保证；这只会在低阶项上改善码率，但对于实际的有限规模而言可能具有重要意义。

高效实现的主要障碍在于，在每次迭代步骤中直接计算 X 的条件期望十分困难。为克服这一点，我们维护该条件期望的一个可高效计算的上界，该上界通过像式 (3) 中那样应用并集界得到；随后，我们选择每个元素，使这一上界在整个过程中始终低于 σn 。这种方法称为悲观估计量法 [25], [60], [61]。

我们证明如下定理。

定理 ***4.1***： 对于任意固定的正整数 r，存在仅依赖于 r 的正实数常数 c_0 和 c，使得以下结论成立。对于任意固定的整数 $d\ge 1$ 和 $h\ge 0$，以及任意满足$_{(\max\{d,h\}),_r}$ t的正整数 t和 $x\le$ ^{c_0}，存在一个确定性算法，可在时间 $O(tn^{\max\{d,h\}+r+1})$内构造一个 $t\times n$(d， h， r； x]-包含式分离矩阵，其中 n是满足以下条件的某个正整数

$$\frac{\log_2 n}{t}\ge\frac{c+o(1)}{(\max\{d,h\})^{r+1}},$$

输入：绝对常数 $\varepsilon \in (0,1)$、 $\delta \in (0, 2^{\varepsilon}-1)$、 $\sigma > 0$和 $s > 0$，正整数 t、 d和 r，非负整数 h，以及满足 $x \leq c_0 (\max\{d,h\})^{r} t$的正整数 x，其中 c_0是定理3.1的证明中给出的正常数。输出：一个 $t \times n$ $(d, h, r; x)$-包含式分离矩阵，满足$\log_{2} n > c+o(1)\,(\max\{d,h\})^{r+1}$，其中 c是定理 3.1的证明申给出的正常数。1： $D \leftarrow \max\{d, h\}$，并按照定理 3.1 的证明设置 p、 z和 y。2：令 n 为满足(5)的最大正整数。3：取一个 $t \times (1+\sigma)n$ 矩阵 $T = (t_{ij})$，其中每个元素 t_{ij} 独立地以概率 p 设为1，以概率 $1-p$设为0。4：如果 $[(1+\sigma)n]$ 中$(D, D, r; z, y)$-违例子集的数量至少为 σn，则返回 FAIL。5：从 $[(1+\sigma)n]$ 的每个$(D, D, r; z, y)$-违例子集中选取一个元素，并从 T中删除所有由这些元素索引的列。6：返回 T

图1. 一种成功概率为 $1 - o(1)$ 的多项式时间蒙特卡洛构造。

其中 $o(1)$ 是满足 $o(1) \to 0$ as $t \to \infty$ 的 t 函数。

证明：固定绝对常数 $\varepsilon \in (0,1)$、 $\delta \in (0, 2^{\varepsilon}-1)$ 和 $\sigma > 0$，并按照定理 3.1 的证明设置 p、 z、 y、 n、 c_0 和 c。设 $D := \max\{d, h\}$。现在通过逐个固定矩阵元素，在 $\{0,1\}$ 上构造一个 $t \times (1+\sigma)n$ 矩阵。

假设我们已经确定了前 $i-1$ 行的所有矩阵元素，以及第 i 行中的 $j-1$ 个矩阵元素，现在试图确定第 i 行的第 j 个元素 t_{ij} 的值。我们将此称为 (i, j)- 步。令 T 为一个 $t \times (1+\sigma)n$ 部分随机矩阵，其中所有已确定的矩阵元素均固定，其余元素独立地以概率 p 取值 1，以概率 $1-p$ 取值 0。用 X 表示 $[(1+\sigma)n]$ 中 $(D, D, r;\ z, y]$- 违例子集的数量，其中 T。定义

$$P_{\text{prev}} := \sum_{\substack{B \subseteq [(1+\sigma)n],\\ |B|=r}} \sum_{\substack{A \subseteq [(1+\sigma)n] \setminus B,\\ |A|=D}} (\Pr[Z_T(A,B) \leq z-1] + \Pr[Y_T(A,B) \geq y+1]).$$

根据期望的线性性和并集界，我们有$\mathbb{E}[X] \leq P_{\text{prev}}$。对于 $k \in \{0,1\}$，定义

$$P_{A,B}(k) := \Pr[Z_T(A,B) \leq z-1 \mid t_{ij} = k] + \Pr[Y_T(A,B) \geq y+1 \mid t_{ij} = k]$$

and

$$P_{\text{curr}}(k) := \sum_{\substack{B \subseteq [(1+\sigma)n],\\ |B|=r}} \sum_{\substack{A \subseteq [(1+\sigma)n] \setminus B,\\ |A|=D}} P_{A,B}(k).$$

那么， $\mathbb{E}[X \mid t_{ij} = k] \leq P_{\text{curr}}(k)$。根据全概率公式，我们有

$$\begin{aligned} P_{\text{prev}} &= \sum_{\substack{B \subseteq [(1+\sigma)n],\\ |B|=r}} \sum_{\substack{A \subseteq [(1+\sigma)n] \setminus B,\\ |A|=D}} ((1-p)P_{A,B}(0) + pP_{A,B}(1)) \\ &= (1-p) \cdot P_{\text{curr}}(0) + p \cdot P_{\text{curr}}(1) \\ &\geq \min_{k \in \{0,1\}} P_{\text{curr}}(k). \end{aligned}$$

因此，如果

$$P_{\text{prev}} < \sigma n, \tag{6}$$

那么我们可以取 $k \in \{0,1\}$，使得 $P_{\text{curr}}(k) \leq P_{\text{prev}}$，从而

$$\begin{aligned} \mathbb{E}[X \mid t_{ij} = k] &\leq P_{\text{curr}}(k) \\ &\leq P_{\text{prev}} \\ &< \sigma n. \end{aligned}$$

如定理 3.1 的证明所示，式 (6) 在 (1,1) 步中成立。因此，通过归纳法，我们可以贪心地确定所有矩阵元素，使得所得确定性矩阵的违例子集少于σn $(D, D, r;\ z, y]$-个。由定理 3.1 的证明可知，从每个 $(D, D, r; z, y]$-违例子集中选取一个元素，并删除由这些元素索引的所有列，即可得到所需的矩阵。

接下来分析运行时间。定义 $q_1 := p^r(1-p)^D$ 和 $q_2 := p^r(1-(1-p)^D)$。为缩短运行时间，我们预先计算

$$s_1(z_0, i) := \sum_{l=0}^{z-z_0} \binom{t-i}{l} q_1^l (1-q_1)^{t-i-l}$$

and

$$s_2(y_0, i) := \sum_{l=0}^{y-y_0} \binom{t-i}{l} q_2^l (1-q_2)^{t-i-l}$$

对每个 $z_0 \in [z]$、 $y_0 \in \{0, 1, \ldots, y\}$ 和 $i \in \{0, 1, \ldots, t\}$，在关于 t 的多项式时间内计算，并将这些值存储在查找表中。为方便起见，对于任意 $z_0 > z$ 和 $y_0 > y$，我们定义 $s_1(z_0,\ i) = 0$ 和 $s_2(y_0,\ i) = 0$。

与之前一样，考虑 (i, j) 步。假设我们已经计算出 P_{prev} 的值。用 T_{fixed} 表示由 T 的前 $i-1$ 行构成的 $(i-1) \times (1+\sigma)n$ 确定性矩阵，用 T_{random} 表示由其余各行构成的 $(t-i+1) \times (1+\sigma)n$ 随机矩阵。令 A 和 B 为 $[(1+\sigma)n]$ 的不相交子集，使得 $|A| = D$ 且 $|B| = r$。由于 T_{fixed} 中的所有矩阵元素都已确定，因此 $z_{A,B} := Z_{T_{\text{fixed}}}(A, B)$ 和 $y_{A,B} := Y_{T_{\text{fixed}}}(A, B)$ 都是确定性的。因此，对于每个 $k \in \{0,1\}$，有

$$P_{A,B}(k) = \Pr[Z_T(A,B) \leq z-1 \mid t_{ij} = k] + \Pr[Y_T(A,B) \geq y+1 \mid t_{ij} = k]$$

$$= \Pr[Z_{T_{\text{random}}}(A,B) \le z - z_{A,B} - 1 \mid t_{ij} = k] + 1 - \Pr[Y_{T_{\text{random}}}(A,B) \le y - y_{A,B} \mid t_{ij} = k].$$

定义

$$\omega_{A,B} \coloneqq \Pr\left[i \in \left(\bigcap_{l \in B} C_l \right) \setminus \left(\bigcup_{l \in A} C_l \right) \,\middle|\, t_{ij} = 0 \right] \tag{7}$$

and

$$\psi_{A,B} \coloneqq \Pr\left[i \in \left(\bigcap_{l \in B} C_l \right) \cap \left(\bigcup_{l \in A} C_l \right) \,\middle|\, t_{ij} = 0 \right]. \tag{8}$$

然后，只需检查由 $A \cup B$ 索引的 $D + r$ 个矩阵元素，即可在常数时间内（与 t 和 n 无关）计算 $\omega_{A,B}$ 和 $\psi_{A,B}$。根据全概率公式，我们有

$$\begin{aligned}
&\Pr[Z_{T_{\text{random}}}(A,B) \le z - z_{A,B} - 1 \mid t_{ij} = 0] \\
&= \omega_{A,B} \cdot \sum_{l=0}^{z - z_{A,B} - 2} \binom{t-i}{l} q_1^l (1-q_1)^{t-i-l} \\
&\quad + (1 - \omega_{A,B}) \cdot \sum_{l=0}^{z - z_{A,B} - 1} \binom{t-i}{l} q_1^l (1-q_1)^{t-i-l} \\
&= \omega_{A,B} \cdot s_1(z_{A,B} + 2, i) + (1 - \omega_{A,B}) \cdot s_1(z_{A,B} + 1, i)
\end{aligned}$$

and

$$\begin{aligned}
&\Pr[Y_{T_{\text{random}}}(A,B) \le y - y_{A,B} \mid t_{ij} = 0] \\
&= \psi_{A,B} \cdot \sum_{l=0}^{y - y_{A,B} - 1} \binom{t-i}{l} q_2^l (1-q_2)^{t-i-l} \\
&\quad + (1 - \psi_{A,B}) \cdot \sum_{l=0}^{y - y_{A,B}} \binom{t-i}{l} q_2^l (1-q_2)^{t-i-l} \\
&= \psi_{A,B} \cdot s_2(y_{A,B} + 1, i) + (1 - \psi_{A,B}) \cdot s_2(y_{A,B}, i).
\end{aligned}$$

因此，我们有

$$\begin{aligned}
P_{A,B}(0) = {} & \omega_{A,B} \cdot s_1(z_{A,B} + 2, i) \\
& + (1 - \omega_{A,B}) \cdot s_1(z_{A,B} + 1, i) \\
& + 1 - \psi_{A,B} \cdot s_2(y_{A,B} + 1, i) \\
& - (1 - \psi_{A,B}) \cdot s_2(y_{A,B}, i).
\end{aligned} \tag{9}$$

由于可以通过随机访问预先计算的查找表，在常数时间内（与 t 和 n 无关）计算每个 $P_{A,B}(0)$，因此我们可以在 $O(n^{D+r})$ 时间内计算 $P_{\text{curr}}(0)$。如果 $P_{\text{curr}}(0) \le P_{\text{prev}}$，则将 t_{ij} 的值设为 0；否则，将其设为 1。

完成每一行 i 后，我们会更新所有计数器 $z_{A,B}$ 和 $y_{A,B}$。每行需要 $O(n^{D+r})$ 时间，因为对于每对不相交集合 A, $B \subseteq [(1+\sigma)n]$，其中 $|A| = D$ 和 $|B| = r$，增量可通过仅检查该行中由 $A \cup B$ 索引的 $D + r$ 个矩阵元素来确定。与处理该行中矩阵元素所需的 $O(n^{D+r+1})$ 时间相比，这项更新的开销较小。

因此，总运行时间为 $O(tn^{D+r+1})$。证明完毕。 ■

所得的确定性构造总结见图2。

V. 总结

我们首次证明了包含式不相交矩阵最大可达率的非平凡渐近下界，并且该下界与目前已知的最强上界仅相差一个对数因子。我们的证明给出了一种可达到这一界限的简单随机化构造，并通过去随机化得到了一种确定性多项式时间构造。这些结果表明，即使面对最具对抗性的抑制剂，大规模容错群体测试仍可实现与经典情形大体相当的渐近效率。这进一步说明，在抑制剂行为难以预测的应用中，研究包含式不相交矩阵具有重要意义。

虽然我们的构造在矩阵规模 tn 上的运行时间为多项式时间，但在实际应用中，人们可能更倾向于采用速度更快的构造。对于传统分离矩阵，文献 [49] 和 [50], 分别给出了 $(d, \mathbf{1}; 1]$- 和 $(d, r;\ 1]$-分离矩阵的线性时间构造。自然会想到，能否将类似技术改用于包含式变体；然而，包含性约束妨碍了直接应用这些技术，而且由此得到的矩阵通常无法保证满足 $(d, h, r;\ x]$-包含式分离性。事实上，通过扩展文献 [49] 和 [50], 中的方法，我们可以针对 $r = 1$ 在线性时间内确定性地构造包含式分离矩阵；但对于 $r \ge 2$，由于需要同时控制所有 r-阶交集，实现类似的线性时间构造似乎遥不可及。在这一方向上的进展，将是推动包含式不相交矩阵在实践中得到广泛应用的重要一步。

参考文献

[1] R. Dorfman，“大规模群体中有缺陷成员的检测”，*Ann. Math. Stat.*，第 14 卷，第 436–440 页，1943年。[2] M. Farach、S. Kannan、E. Knill 和 S. Muthukrishnan，“实验分子生物学中带序列的分组测试问题”，载于 *序列压缩与复杂性 1997*。IEEE，1997 年，第 357–367 页。[3] D.-Z. Du 和 F. K. Hwang，组合分组测试及其应用，第 2 版。World Scientific，2000 年。[4] T. V. Bui、M. Kuribayashi、M. Cheraghchi 和 I. Echizen，“含抑制剂的广义分组测试框架及其在神经科学中的潜在应用”，2019 年。 [在线]。可访问：https://arxiv.org/abs/ 1810 01086 2 。v [5] M. Xie、K. Tatsuoka、J. Sacks 和 S. S. Young，“含阻断剂和协同作用的分组测试”，*J. Am. Stat. Assoc.*，第 96 卷，第 92–102 页，2001 年。[6] A. D’yachkov、N. Polyanskii、V. Shchukin 和 I. Vorobyev，“对称多址信道的可分离码”，*IEEE Trans. Inf. Theory*，第 65 卷，第 3738–3750 页，2019 年。[7] G. K. Atia 和 V. Saligrama，“布尔压缩感知与含噪分组测试”，*IEEE Trans. Inf. Theory*，第 58 卷，第 1880–1901 页，2012 年。[8] N. Shental、S. Levy、V. Wuvshet、S. Skorniakov、B. Shalem、A. Ottolenghi、Y. Greenshpan、R. Steinberg、A. Edri、R. Gillis、M. Goldhirsh、K. Moscovici、S. Sachren、L. M. Friedman、L. Nesher、Y. Shemer-Avni、A. Porgador 和 T. Hertz，“高效高通量 SARS-CoV-2 检测，以发现无症状携带者”，*Sci. Adv.*，第 eabc 5961 页，2020 年，第 6 卷。[9] D.-Z. Du 和 F. K. Hwang，池化设计与非自适应分组测试。World Scientific，2006 年。[10] J. Scarlett，“含噪自适应分组测试：界与算法”，*IEEE Trans. Inf. Theory*，第 65 卷，第 3646–3661 页，2019 年。[11] M. Aldridge、O. Johnson 和 J. Scarlett，“分组测试：信息论视角”，*Found. Trends Commun. Inf. Theory*，第 15 卷，第 196–392 页，2019 年。[12] A. J. Macula，“使用 d^e-不相交矩阵的纠错非自适应分组测试”，*Discrete Appl. Math.*，第 80 卷，第 217–222 页，1997 年。

输入：绝对常数 $\varepsilon \in (0,1)$**、** $\delta \in (0, 2^{\varepsilon}-1)$ **和** $\sigma > 0$**，正整数** t**、** d **和** r**，非负整数** h**，以及满足** $x \leq c_0 (\max\{d,h\})_r t$ **的正整数** x**，其中** c_0 **是定理 3.1 证明中给出的正数。输出：一个** $t \times n$**(**d, h, r**;** x**)-包含型不相交矩阵，其** $\log_2 n > c+o(1)\,(\max\{d,h\})^{r+1}$**，其中** c **是定理 3.1 证明中给出的正数。1：** $D \leftarrow \max\{d,h\}$**，并按定理 3.1 的证明设置** p**、** z**、** y **和** n**。2：对每个，，和，计算**

$$s_1\ z_0\ i\ = \sum_{l=0}^{z-z_0} \binom{t-i}{l} q_1^{\,l}\ 1-q_1\ ^{t-i-l} \qquad s_2\ y_0\ i\ = \sum_{l=0}^{y-y_0} \binom{t-i}{l} q_2^{\,l}\ 1-q_2\ ^{t-i-l} \qquad z_0 \in [z]\ \ y_0 \in (\,,\,) \{0,1,\ldots,y\} \qquad i \in \{0,1,\ldots,t\} \qquad q_1 = p^r\ 1-p\ ^D \qquad q_2 = p^r\ 1-\ 1-p\ ^D$$

(,**) (**,**)，其中 (**,**) 和 (**,**)。3：** $P_{\text{prev}} \leftarrow \binom{(1+\sigma)n}{r}\binom{(1+\sigma)^{n-r}}{D}(s_1(1,0) + 1 - s_2(0,0))$**。4：取一个** $t \times (1+\sigma)n$ **矩阵** $T = (t_{ij})$**，并以符号** $*$**填充，该符号表示未确定值。5：对每一对不相交集合** A**、** $B \subseteq [(1+\sigma)n]$**，使得** $|A| = D$ **和** $|B| = r$**，设置** $z_{A,\,B} \leftarrow 0$ **以及** $y_{A,\,B} \leftarrow 0$**。6：对于** $i = 1$ **到** t**，执行。7：对** T **第** i **行中的每个元素** t_{ij}**，执行。8：对每一对不相交集合** A**、** $B \subseteq [(1+\sigma)n]$**，使得** $|A| = D \qquad |B| = r$ **且：(**,**)(**

$$n]\ \sum_{A\subseteq[,\ (1+\sigma\ n]\setminus B\ P_A)}$$

)、 $_{B}(0)$ **和** $P_{\text{curr}}(1)$**，分别按 (7)、(8) 和 (9) 计算** $\omega_{A,\,B}$**，** $\psi_{A,\,B}$**，以及** $P_{p\ A,\,B}(0)$**。9：**

$$|B|=r \qquad |A|=D \qquad {}^{9} \qquad P_{\text{curr}}\ 0\ \leftarrow \sum_{B\subseteq[\ 1+\sigma}$$

$$\leftarrow {}^{P_{\text{prev}}-}(1-p)P_{\text{curr}}(0)\ 10 \qquad t_{ij} \quad 0 \quad P_{\text{curr}}\ 0\ \leq P_{\text{prev}} \qquad \frac{\qquad\qquad 1 \qquad\qquad}{}$$

：若 (,**)，则将其固定为；否则，将其固定为。11：** $P_{\text{prev}} \leftarrow P_{\text{curr}}(t_{ij})$**。12：结束循环。13：对每一对不相交集合** A**、** $B \subseteq [(1+\sigma)n]$**，使得** $|A| = D$ **和** $|B| = r$**，如果第** i **行对** $Z_{T_{\text{fixed}}}(A,B)$ **和** $Y_{T_{\text{fixed}}}(A,B)$ **分别有贡献，则将** $z_{A,\,B}$ **以及** $y_{A,\,B}$ **各加 1，其中** T_{fixed} **是由** T **中前** i **个已确定行组成的** $i \times (1+\sigma)n$ **矩阵。14：结束循环。15：从** $[(1+\sigma)n]$ **的每个 (**D, D, r**;** z, y**)-违反子集中各选取一个元素，并删除** T **中由这些元素索引的所有列。16：返回** T

图2. 一种确定性多项式时间构造。

[13] H.-B. Chen、H.-L. Fu 和 F. K. Hwang，“容错复合模型中池化设计测试次数的上界”，《优化快报》，第2卷，第425–431页，2008年。[14] D. R. Stinson 和 R. Wei，“广义覆盖自由族”，《离散数学》，第279卷，第463–477页，2004年。[15] T. B. Idalino 和 L. Moura，“覆盖自由族综述：构造、应用与推广”，载于《设计、编码与密码学新进展》。施普林格，2024年，第195–239页。[16] M. Lappe 和 L. Holm，“以近乎最优的效率揭示蛋白质相互作用网络”，《自然·生物技术》，第22卷，第98–103页，2004年。[17] Y. Li、M. T. Thai、Z. Liu 和 W. Wu，“二分图中的蛋白质-蛋白质相互作用与群体测试”，《国际生物信息学研究与应用杂志》，第1卷，第414–419页，2005年。[18] H.-B. Chen、D.-Z. Du 和 F. K. Hwang，“四个看似无关的问题意外交汇：图测试、DNA复合筛查、叠加码与安全密钥分发”，《组合优化杂志》，第14卷，第121–129页，2007年。[19] D. C. Torney，“集合池化设计”，《组合年刊》，第3卷，101，1999年，95页。[20] R. M. Phatarfod 和 A. Sudbury，“方阵方案在血液检测中的应用”，《医学统计学》，第13卷，第2337–2343页，1994年。[21] E. Barillot、B. Lacroix 和 D. Cohen，“使用 N 维池化策略进行文库筛查的理论分析”，《核酸研究》，第19卷，第6241–6247页，1991年。[22] F. K. Hwang 和 F.-H. Chang，“一般抑制剂模型中阳性克隆的识别”，《计算机与系统科学杂志》，第73卷，1094，2007年，1090页。[23] F.-H. Chang、H. Chang 和 F. K. Hwang，“抑制剂复合模型下克隆文库筛查的池化设计”，《组合优化杂志》，第22卷，第145–152页，2011年。[24] H. Chang、H.-B. Chen 和 H.-L. Fu，“ 抑制剂模型池化设计中的识别与分类问题 ”，《计算生物学杂志》，第17卷，第927–941页，2010年。[25] N. Alon 和 J. H. Spencer，概率方法，第 4 版。威利，2016。[26] M. Mitzenmacher 和 E. Upfal，概率与计算：算法与数据分析中的随机化和概率技术，第 2 版。剑桥大学出版社，2017。[27] T. V. Bui、M. Cheraghchi 和 I. Echizen，“带间隔的阈值群组测试的改进非自适应算法”，*IEEE* 信息论汇刊，第 67 卷，第 7180–7196 页，2021。[28] W. Kautz 和 R. Singleton，“非随机二元叠加码”，*IEEE* 信息论汇刊，第 10 卷，第 363–377 页，1964。[29] T. Berger、N. Mehravari、D. Towsley 和 J. Wolf，“随机多址通信与群体测试”，*IEEE* 通信汇刊，第 32 卷，第 769–779 页，1984。[30] J. Wolf，“重生群体测试：多址通信”，*IEEE* 信息论汇刊，第 31 卷，第 185–191 页，1985。[31] A. G. D’yachkov、A. J. Macula 和 V. V. Rykov，“叠加码的新构造”，*IEEE* 信息论汇刊，第 46 卷，第 290 页，2000，284 页。[32] A. De Bonis 和 U. Vaccaro，“广义叠加码的构造及其在群体测试和多址信道冲突消解中的应用”，理论计算机科学，第 306 卷，第 243 页，2003，223 页。[33] Y. Fujiwara 和 C. J. Colbourn，“X 容错压缩电路的组合方法”，*IEEE* 信息论汇刊，第 56 卷，第 3206 页，2010，3196 页。[34] L. Gargano、A. A. Rescigno 和 U. Vaccaro，“低重量叠加码及相关组合结构：界与应用”，理论计算机科学，第 806 卷，第 655–672 页，2020。[35] A. A. Rescigno 和 U. Vaccaro，“广义叠加码的界与算法”，信息处理快报，第 182 卷，文章编号 106365，2023。[36] P. Erd˝os 、 P. Frankl 和 Z. F¨uredi，“不存在任何集合被另外两个集合的并集覆盖的有限集族”，组合理论杂志 *A* 辑，第 33 卷，第 158–166 页，1982。

[37] ——，“不包含任何集合可被其余 r 个集合的并集覆盖的有限集族”，《以色列数学杂志》，第 51 卷，第 79–89 页，1985 年。[38] M. Ruszink´o，“ r-覆盖自由族大小的上界”，《组合理论杂志 *A* 辑》，第 66 卷，第 302–310 页，1994 年。[39] A. D’yachkov、P. Vilenkin、D. Torney 和 A. Macula，“不包含任何 ℓ 个集合的交集可被其余 s 个集合的并集覆盖的有限集族”，《组合理论杂志 *A* 辑》，第 99 卷，第 195–218 页，2002 年。[40] H.-G. Yeh，“d-不相交矩阵：界与 Lov´asz 局部引理”，《离散数学》，第 253 卷，第 97–107 页，2002 年。[41] D. Deng、D. R. Stinson 和 R. Wei，“Lov´asz 局部引理及其在某些组合数组中的应用”，《设计、编码与密码学》，第 32 卷，第 121–134 页，2004 年。[42] A. G. D’yachkov 和 V. V. Rykov，“非相交码长度的界”，《信息传输问题》，第 18 卷，第 7–13 页，1982 年，俄文。[43] A. G. D’yachkov、V. V. Rykov 和 A. M. Rashad，“叠加距离码”，《控制与信息理论问题》，第 18 卷，第 237–250 页，1989 年。[44] M. Dyer、T. Fenner、A. Frieze 和 A. Thomason，“安全网络中的密钥存储”，《密码学杂志》，第 8 卷，第 189–200 页，1995 年。[45] K. Engel，“布尔格中的区间装箱与覆盖”，《组合、概率与计算》，第 5 卷，第 373–384 页，1996 年。[46] Z. F¨uredi，“关于 r-覆盖自由族”，《组合理论杂志 *A* 辑》，第 73 卷，第 172–173 页，1996 年。[47] D. R. Stinson、R. Wei 和 L. Zhu，“覆盖自由族的一些新界”，《组合理论杂志 *A* 辑》，第 90 卷，第 224–234 页，2000 年。[48] R. Wei，“关于覆盖自由族”，2006 年。 [在线]。网址：https://arxiv.org/abs/2303 17524 。[49] E. Porat 和 A. Rothschild，“显式非自适应组合群测试方案”，《*IEEE* 信息论汇刊》，第 57 卷，第 7982–7989 页，2011 年。[50] N. H. Bshouty，“某些 d-限制问题的线性时间构造——题目”，载于算法与复杂性（*CIAC 2015*）。施普林格，2015年，第74–88页。[51] A. Z. Abdi 和 N. H. Bshouty，“覆盖自由族的下界”，组合学电子期刊，第23卷，第1–7页，2016年。[52] Y. Tsunoda 和 Y. Fujiwara，“具有改进渐近速率的弱叠加码及其随机化构造”，载于*2022 IEEE*国际信息论研讨会（*ISIT*），2022年，第784–789页。[53] F. K. Hwang 和 V. T. S´os，“非自适应超几何群体检测”，匈牙利数学科学研究，第22卷，第257–263页，1987年。[54] P. Erd˝os 和 L. Lov´asz，“关于三色超图的问题与结果及一些相关问题”，载于无限集与有限集。北荷兰出版社，1973年，第609–627页。[55] R. A. Moser 和 G. Tardos，“一般Lov´asz局部引理的构造性证明”，*ACM*期刊，第57卷，第1–15页，2010年。[56] M. Molloy 和 B. Reed，图着色与概率方法，算法与组合学丛书。施普林格，2002年，第23卷。[57] C. J. Colbourn，“使用贪心算法构造完美哈希族”，载于编码与密码学。World Scientific，2008年，第109–118页。[58] R. Dougherty，“哈希族及其在 t-限制中的应用”，博士学位论文，亚利桑那州立大学，2019年。[59] R. E. Dougherty 和 C. J. Colbourn，“完美哈希族：推广至高指标”，载于离散数学与应用。施普林格，2020年，第177–197页。[60] P. Raghavan，“确定性算法的概率构造：逼包装整数规划”，《计算机与系统科学杂志》，第37卷，第130–143页，1988年。[61] A. Wigderson 和 D. Xiao，“使用悲观估计量去随机化Ahlswede-Winter矩阵值Chernoff界及其应用”，计算理论，第4卷，第53–76页，2008年。