跳频循环码和光正交码的改进吉尔伯特–瓦尔沙莫夫界

张晨阳、上官冲和葛根年

arXiv:2301.02042v1 [cs.IT] 5 Jan 2023

摘要

跳频循环码（HCC）是（非线性）循环码，具有这样一项附加性质：给定码字的 n 个循环移位彼此不同，其中 n 为码长。恒重二元跳频循环码也称为光正交码（OOC）。HCC 和 OOC 在实际中有多种应用，多年来一直受到广泛研究。

本文主要关注为这类码给出改进的吉尔伯特–瓦尔沙莫夫型下界，适用于最小距离至少为码长的线性因子的情形。对于 HCC，我们将 Niu、Xing 和 Yuan 目前已知的最佳下界提高了一个码长的线性因子。对于 OOC，我们将 Chung 、 Salehi 和 Wei，以及杨和富贾目前已知的最佳下界提高了一个码长的二次因子。此外，我们还给出了跳频序列集和纠错弱互不相关码的改进下界。我们的证明基于概率论和图论中的工具，特别是关于 Lipschitz 函数集中性的 McDiarmid 不等式，以及局部稀疏图的独立数。

关键词。吉尔伯特–瓦尔沙莫夫界；非线性循环码；跳频循环码；光正交码；跳频序列集；纠错弱互不相关码

一、引言

给定整数 q、 n 和 d，估计长度为 n 、最小距离为 d 的 q 元码的最大规模，是编码理论中的一个基本问题。吉尔伯特–瓦尔沙莫夫界（简称 GV 界）是码的规模的经典下界。对于固定的 q、 $n\to\infty$，以及下界为 n 的线性因子的 d，要显著改进 GV 界是一项众所周知的难题。本文将给出若干类非线性循环码的改进 GV 型界。

为了进一步讨论，我们先介绍一些必要的定义。设 q 为正整数，令 $[q]=\{0,1,\ldots,q-1\}$。向量 $\boldsymbol{x}\in[q]^n$ 记作 $\boldsymbol{x}=(x_1,\ldots,x_n)$。对于两个向量 $\boldsymbol{x}$、 $\boldsymbol{y}\in[q]^n$，汉明距离 $d(\boldsymbol{x},\boldsymbol{y})$ 是它们不同的坐标数，即 $d(\boldsymbol{x},\boldsymbol{y})=|\{1\le i\le n: x_i\neq y_i\}|$。长度为 n 、字母表大小为 q 的码是 $[q]^n$ 的一个子集，其元素称为码字。码 $C\subseteq[q]^n$ 的最小距离 $d(C)$ 定义为 $d(C):=\min\{d(\boldsymbol{x},\boldsymbol{y}): \boldsymbol{x},\boldsymbol{y}\in C,\boldsymbol{x}\neq\boldsymbol{y}\}$。长度为 n 、规模为 M 、最小距离为 d 的 q 元码记作 $(n,M,d)_q$- 码。

接下来，我们简要回顾一些已知的 GV 型界。

A. GV 型界

a) 一般码：对于整数 $1\le t\le n$ 和向量 $\boldsymbol{x}\in[q]^n$，令 $B(\boldsymbol{x},t):=\{\boldsymbol{y}\in[q]^n: d(\boldsymbol{x},\boldsymbol{y})\le t\}$ 表示以 $\boldsymbol{x}$ 为中心、半径为 t 的汉明球。给定半径 t，容易看出，对于每个 $\boldsymbol{x}\in[q]^n$，汉明球 $B(\boldsymbol{x},t)$ 的体积与 $\boldsymbol{x}$ 无关，并且满足

$$|B(\boldsymbol{x},t)|=\sum_{i=0}^{t}\binom{n}{i}(q-1)^i=:Vol_q(n,t).$$

本项目得到国家重点研发计划（项目编号：2020YFA0712100 和 2018YFA0704703）、国家自然科学基金（项目编号：11971325、12231014 和 12101364）、山东省自然科学基金（项目编号：ZR2021QA005）以及北京学者计划的资助。

C. Zhang 任职于山东大学数学与交叉科学研究中心，中国青岛 266237（电子邮箱：chener@mail.sdu.edu.cn）。C. Shangguan 任职于山东大学数学与交叉科学研究中心，中国青岛 266237，同时也任职于教育部非线性期望前沿科学中心，中国青岛 266237（电子邮箱：theoreming@163.com）。G. Ge 任职于首都师范大学数学科学学院，中国北京 100048（电子邮箱：gnge@zju.edu.cn）。

由 Gilbert [7] 和 Varshamov [22], 分别独立证明的 GV 界表明，存在具有如下性质的 $(n, M, d)_q$-码

$$M \geq \frac{q^n}{Vol_q(n, d-1)}. \tag{1}$$

对于二元码和 $d/n \leq 0.499$，Jiang 和 Vardy [11] 将 (1) 中的下界提高了 n 这一线性因子。基于他们的工作，Vu 和 Wu [23] 证明，对于每个 $q \geq 2$ 和 $\tau \leq d/n \leq 1 - 1/q - \epsilon$，其中 τ、ϵ 是一些绝对实数，存在 $(n, M, d)_q$-码，使得

$$M = \Omega\left(\frac{nq^n}{Vol_q(n, d-1)}\right). \tag{2}$$

b) 非线性循环码。：对于每个 $\boldsymbol{x} = (x_1, \ldots, x_n) \in [q]^n$ 和 $0 \leq i \leq n-1$，我们称向量$\pi_i(\boldsymbol{x}) := (x_{i+1}, x_{i+2}, \ldots, x_{i+n})$[1] 为 $\boldsymbol{x}$ 的第 i 次循环移位。令 $C(\boldsymbol{x}) := \{\pi_i(\boldsymbol{x}) : 0 \leq i \leq n-1\}$ 表示由 $\boldsymbol{x}$ 的所有循环移位组成的集合。注意，$C(\boldsymbol{x})$ 被视为一个多重集。若对于每个 $\boldsymbol{x} \in C$，都有 $C(\boldsymbol{x}) \subseteq C$，则称码 $C \subseteq [q]^n$ 为循环码。文献中有许多研究致力于构造最小距离从常数到 n 的次线性函数（例如，$O(n/\ln n)$）不等的线性循环码（参见例如 [19,第 8] 章）。然而，当最小距离以下界为 n 的线性因子时，是否存在渐近良好的线性循环码，仍是编码理论中一个长期悬而未决的问题（参见 [6], [8]）。

对于非线性循环码，上述问题直到最近才得到解决。Haviv、Langberg、Schwartz 和 Yaakobi [8] 证明，对于每个素数码长，都存在渐近达到 GV 界的二元循环码。随后，Niu、Xing 和 Yuan [16] 证明，对于每个 q 和 n，都存在 $(n, M, d)_q$-循环码，满足

$$M \geq \frac{q^n(1 - n^2 e^{-\frac{\epsilon^2(\sqrt{n}-2)}{2}})}{Vol_q(n, d-1) - 1}. \tag{3}$$

c) 恒重码：对于向量 $\boldsymbol{x} \in \{0,1\}^n$，重量 $wt(\boldsymbol{x})$ 定义为其非零坐标的数量，即 $wt(\boldsymbol{x}) = |\{1 \leq i \leq n : x_i \neq 0\}|$。如果一个码的所有码字具有相同的重量，则称其为恒重码。具有恒定重量 w 的 $(n, M, d)_2$-码记为 $(n, M, d; w)$-码，其中我们省略了下标中的 $q = 2$。对于 $\boldsymbol{x} \in \{0,1\}^n, wt(\boldsymbol{x}) = w$ 和 $0 \leq t \leq 2w$，令 $B(\boldsymbol{x}, t; w) := \{\boldsymbol{y} \in \{0,1\}^n : d(\boldsymbol{x}, \boldsymbol{y}) \leq t, wt(\boldsymbol{y}) = w\}$ 表示以 $\boldsymbol{x}$ 为中心、半径为 t 的恒重汉明球。给定 t 和 w，已知对于每个满足 $wt(\boldsymbol{x}) = w$ 的 $\boldsymbol{x} \in \{0,1\}^n$，$B(\boldsymbol{x}, t; w)$ 的体积与 $\boldsymbol{x}$ 无关，并且满足

$$|B(\boldsymbol{x}, t; w)| = \sum_{i=0}^{\lfloor t/2 \rfloor} \binom{w}{i}\binom{n-w}{i} =: Vol(n, t; w).$$

列文斯坦 [14] 证明了恒重码的 GV 型界，表明存在满足以下条件的 $(n, M, d; w)$-码

$$M \geq \frac{\binom{n}{w}}{Vol(n, d-1; w)}. \tag{4}$$

类似于 (2) 对 (1) 的改进，Kim、Liu 和 Tran [12] 通过线性因子 n改进了 (4)，表明对于 $\tau \leq d/n \leq (1-\epsilon)p(1-p)$，其中 τ, ϵ, p 是一些绝对实数，存在满足以下条件的 $(n, M, d; pn)$-码

$$M = \Omega\left(\frac{n\binom{n}{pn}}{Vol(n, d-1; pn)}\right). \tag{5}$$

d) 恒定重量非线性循环码：文献中也存在恒重非线性循环码的 GV 型下界。更准确地说，钟、萨利希和魏 [1], [3] 以及杨和富贾 [24] 隐式地表明，存在满足以下条件的 $(n, M, d; w)$-循环码

$$M \geq \frac{\binom{n}{w} - f(n, w, d)}{n \cdot Vol(n, d-1; w)}, \tag{6}$$

其中 $f(n, w, d)$ 是 n, w, d 的某个函数（详见 [1, 定理 2] 和 [24, 定理 3]）。

[1]本文中，下标中的加法 x 均按模 n 计算。

鉴于上述讨论，自然会想到，能否像从 (1) 和 (4) 改进得到 (2) 和 (5) 那样，分别将 (3) 和 (6) 改进一个与 n 成线性或多项式因子的幅度？这是本文的动机问题之一。需要注意的是，证明 (2) 和 (5) 的方法不能直接用于改进 (3) 和 (6)，因为必须考虑循环码的性质。借助一些概率论和图论工具，以及来自 [11], [12], [16], [23]，的一些思想，我们将在下文的定理 II.1 和 II.2 中以有力的方式回答上述问题。

本文主要致力于为两类特殊的非线性循环码，即跳频循环码及其恒定重量对应类——光正交码，给出改进的 GV 型界。作为附带结果，我们还为跳频序列集和纠错弱互不相关码给出了改进的下界。

接下来，我们将详细介绍这些码及我们的结果。

II. 主要结果

A. 跳频循环码

跳频循环码是一类具有额外性质的（非线性）循环码：每个给定码字的 n 个循环移位彼此不同。形式而言，若对每个 $\boldsymbol{x} \in C$、$C(\boldsymbol{x}) \subseteq C$，并且 $C(\boldsymbol{x})$ 由 n 个互不相同的元素组成，则称码 $C \subseteq [q]^n$ 为跳频循环码（简称 HCC）。若跳频循环码本身是一个 $(n, M, d)_q$ 码，我们称其为 $(n, M, d)_q$-HCC。HCC 最初是为构造跳频序列集而设计的（见 [4], [16] 及下文第三节 A），但作为一类引人关注的码，它们本身也具有研究价值。有关 HCC 的更多背景，请参阅 [16] 。

对于任意绝对常数 $\epsilon > 0$ 和 $d < (n - 2\sqrt{n})(1 - 1/q - \epsilon)$，Niu、Xing 和 Yuan（见 [16]中的定理 III.5）证明了存在 $(n, M, d)_q$-HCC，其 M 下界为 (3)。

我们的第一个主要结果将上述 HCC 的 GV 型界（因而也将非线性循环码的 GV 型界）渐近地提高了 n这一线性因子。

定理 II.1. 对于满足 $\tau \le d/n \le 1 - 1/q - \epsilon$的正整数 n、q、d和绝对实数 τ、$\epsilon \in (0, 1 - 1/q)$，存在 $(n, M, d)_q$-HCC，其

$$M \ge \frac{cnq^n}{Vol_q(n, d-1)},$$

其中，c是仅取决于 τ、ϵ、q的正实数。

B. 光正交码

恒重二元跳频循环码也称为光正交码（简称 OOC）。光正交码使大量异步用户能够高效且可靠地传输信息。它们已广泛应用于码分多址系统、扩频通信等各种实际场景。已有大量论文致力于光正交码的构造，例如参见 [1], [2], [9], [13], [17]–[24]。

我们将恒定重量为 w 的 $(n, M, d)_2$-HCC 记为 $(n, M, d; w)$-光正交码，其中省略了下标中的 $q = 2$。钟、萨利希和魏 [1], [3] 以及杨和富贾 [24] 证明，存在满足 M 下界为 (6) 的 $(n, M, d; w)$-光正交码。

我们的第二个主要结果将上述光正交码（进而也包括恒定重量非线性循环码）的 GV 型界渐近地提高了一个 n 的二次因子。

定理 II.2. 对于正整数 n, q, d 和绝对实数 $p \in (0, 1)$ $\epsilon \in (0, 1/10)$ $\tau \in (0, p(1-p))$，满足 $pn \in \mathbb{Z}$，$\tau \le d/n \le (1-\epsilon)p(1-p)$，存在 $(n, M, d; pn)$-光正交码，使得

$$M \ge \frac{cn\binom{n}{pn}}{Vol(n, d-1; pn)},$$

其中 c 是一个正实数，仅取决于 p、ϵ、τ。

C. 证明概要及与相关工作的比较

自 Jiang 和 Vardy 的工作 [11]，以来，众所周知，可以通过近似定义的局部稀疏图的独立数下界，给出码大小的下界。粗略地说，设 G 是一个顶点集为 $[q]^n$ 的图，当且仅当 $d(\boldsymbol{x},\boldsymbol{y}) \le d-1$ 时，两个顶点（向量） $\boldsymbol{x}$、$\boldsymbol{y} \in [q]^n$ 之间存在一条边。那么， G 的一个独立集 I 对应于一个 $(n, |I|, d)_q$- 码。事实上， G 的独立集与 $[q]^n$ 中最小距离至少为 d 的码一一对应。因此，给定最小距离 d，在 $[q]^n$ 中寻找一个大码，等价于在 G 中寻找一个大独立集。

由于对于每个 $\boldsymbol{x} \in [q]^n$，恰有 $Vol_q(n, d-1)$ 个 -1 向量 $\boldsymbol{y} \in [q]^n \setminus \{\boldsymbol{x}\}$ 满足 $d(\boldsymbol{x},\boldsymbol{y}) \le d-1$，因此 G 中的每个顶点都恰好与 $[q]^n$ 中的 $Vol_q(n, d-1)$ 个 -1 顶点相连。所以，可以通过贪心地选取顶点并移除其邻点，在 G 中构造一个大小为 $\frac{q^n}{Vol_q(n,\ d-1)}$ 的独立集，或者等价地，构造一个 $(n, \frac{q^n}{Vol_q(n,\ d-1)}, d)_q$-码。这给出了 GV 界 (1)。

Jiang 和 Vardy [11] 证明，上述图 G 实际上是“局部稀疏”的（这一点将在 IV-A 节中进一步说明）。因此，寻找 G 中一个较大独立集的方法比贪心算法更为经济（见下文引理 IV.1）。这使他们将一般码的 GV 界 (1) 提高了一个 n 的线性因子。Kim、Liu 和 Tran [12] 在证明 (5) 时采用了类似的思路，将恒重码的 GV 界 (4) 提高了一个 n 的线性因子。

定理 II.1 和 II.2 的证明总体思路将遵循上述框架。然而，由于本文所研究的码不仅需要具有较大的最小距离，还必须是跳跃循环的，因此相应的图并不像上文所述的图 G 那样规整。

定理 II.1 和 II.2 的证明可概括如下：
步骤 1. 定义一个恰当的图 G_{HCC}（分别为 G_{OOC}），使其独立集对应于 HCC（分别对应于光正交码）；步骤 2. 证明 G_{HCC}（分别为 G_{OOC}）具有足够多的顶点；步骤 3. 证明 G_{HCC}（分别为 G_{OOC}）实际上是局部稀疏的；步骤 4. 利用关于局部稀疏图的独立数的已知下界，证明 G_{HCC}（分别为 G_{OOC}）具有足够大的独立集，因此存在足够大的 HCC（分别为光正交码）。

接下来，我们想将本文工作与一些先前的论文 [11], [12], [16], [23] 进行比较。一方面，在[11]、[12]、[23]、G 中，G 及其恒重对应版本分别被用于证明一般码和恒重码的改进 GV 型界。本文定义的 G_{HCC}（分别为 G_{OOC}）与 G（分别为其恒重对应版本）有很大不同。更具体地说，对于 $\boldsymbol{x} \in [q]^n$，令

$$d(\boldsymbol{x}) = \min\{d_H(\pi_i(\boldsymbol{x}), \pi_j(\boldsymbol{x})) : 0 \le i < j \le n-1\} = \min\{d_H(\boldsymbol{x}, \pi_i(\boldsymbol{x})) : 1 \le i \le n-1\} \tag{7}$$

表示最小距离 $d(C(\boldsymbol{x}))$。根据定义，只有当对于每个 $\boldsymbol{x} \in C$，$d(\boldsymbol{x}) \ge 1$ 时，$C \subseteq [q]^n$ 才是一个 HCC；此外，只有当对于每个 $\boldsymbol{x} \in C$，$d(\boldsymbol{x}) \ge d$ 时，C 才是一个 $(n, M, d)_q$-HCC。因此，与顶点集为 $[q]^n$ 的 G 相比，在 G_{HCC} 的顶点集中，我们必须排除所有满足 $d(\boldsymbol{x}) < d$ 的向量 $\boldsymbol{x}$。事实上，G_{HCC} 的顶点集并非 $[q]^n$ 中的向量，而是 $[q]^n$ 中向量的一些定义明确的子集。因此，为了执行步骤 1 和 2，我们必须证明，对于许多向量 $\boldsymbol{x} \in [q]^n$，$d(\boldsymbol{x})$ 相当大。对 G_{OOC} 也有类似的讨论。详情见下文第 V 节和第 VI 节。

另一方面，为了证明 G_{HCC} 和 G_{OOC} 是局部稀疏的，我们将使用一些经过充分发展的工具；这些工具在 [11], [12], [23] 中不断得到改进。具体而言，我们将使用汉明球交集体积的两个上界，这两个上界在 [12] 中得到证明，用于证明 G 及其恒定重量对应形式是局部稀疏的。

最后，我们将本文工作与 [16] 进行比较。尽管 [16] 并非图论的，但他们在证明 (3) 时的一个重要步骤，同样是证明对于 $q^n - o(q^n)$，向量 $\boldsymbol{x} \in [q]^n$, $d(\boldsymbol{x})$ 的数量相当大（见引理 III.4] [16, ）。为此，他们使用了鞅的某种标准集中不等式。本文则改用 McDiarmid 不等式，这使我们的证明更简单，并且能更好地控制低阶项 $o(q^n)$（见下文注 V.2）。

表 I 总结了前文提到的 GV 型界。

本文其余部分安排如下。第 III 节将介绍定理 II.1 的两个应用，该定理为跳频序列集和纠错弱互不相关码给出了新的下界。第 IV 节将汇总步骤 1–4 中使用的工具。第 IV-A 节将正式定义局部稀疏图（将在步骤 1 中使用），并给出其独立数的一个下界（将在步骤 4 中使用，见下文引理 IV.1）。第 IV-B 节将介绍 McDiarmid 不等式（将在步骤 2 中使用，见下文引理 IV.2）。第 IV-C 节将给出汉明球交集体积的两个上界

表 I：GV 型界汇总

	M 的 GV 型下界	M 的改进 GV 型下界
通用（n, M, d）q-码	$\frac{q^n}{Vol_q(n,d-1)}$ (1)（参见 [7], [22]）	$\Omega\left(\frac{nq^n}{Vol_q(n,d-1)}\right)$ (2)（见 [11], [23]）
非线性（n, M, d）q-循环码	$\frac{(1-o(1))\,q^n}{Vol_q(n,\,d-1)}$ （3）（见 [9], [16]）	$\frac{cnq^n}{Vol_q(n,\,d-1)}$ 定理 II.1（本文）
常重量（n, M, d; pn）码	$\frac{\binom{n}{pn}}{Vol(n,d-1;pn)}$ (4)（见 [14]）	$\Omega\left(\frac{n\binom{n}{pn}}{Vol_{n\,d-1}(,\,;\,pn)}\right)$ (5)（见 [12]）
常重量非线性（$n, M, d; pn$）循环码	$\frac{\binom{n}{w}-f(n,d,w)}{n\cdot Vol(n,d-1;w)}$ (6)（见 [1], [3], [24]）	$\frac{cn\binom{n}{pn}}{Vol(n,d-1;pn)}$ 定理 II.2（本文）

（将在步骤 3 中使用，见下文引理 IV.3 和 IV.4）。定理 II.1 和 II.2 的证明将分别在第 V 节和第 VI 节给出。本文将在第 VII 节作结。

III. 主要结果的应用

A. 跳频序列

在本小节中，我们将介绍定理 II.1 在跳频序列中的一个应用。我们需要下面这个简单的引理。

引理 III.1. (i) 对于每个 $\boldsymbol{x}, \boldsymbol{y} \in [q]^n$, $\boldsymbol{x} \in C(\boldsymbol{y})$，当且仅当 $\boldsymbol{y} \in C(\boldsymbol{x})$；此外，$C(\boldsymbol{x}) = C(\boldsymbol{y})$ 当且仅当$\boldsymbol{x} \in C(\boldsymbol{y})$。(ii) 设 C是一个 $(n, M, d)_q$-HCC。那么 $n \mid M$，且 C可划分为若干个两两不相交的 M/n $C(\boldsymbol{x})$'的并，其中 $\boldsymbol{x} \in C$。

证明。注意，(i) 可由 $C(\boldsymbol{x})$ 的定义直接得到。为证明 (ii)，不难验证，在集合 C 上由以下条件定义的关系 $\sim$ 是一个等价关系：若 $C(\boldsymbol{x}) = C(\boldsymbol{y})$，则 $\boldsymbol{x} \sim \boldsymbol{y}$。由 (i) 和 HCC 的定义可知，每个等价类都等于某个 $C(\boldsymbol{x})$，其中 $\boldsymbol{x} \in C$，因此由 n 个不同向量组成。因此，(ii) 可由以下事实得出：$\sim$ 给出的等价类构成了 C 的一个划分。 □

设 C为一个 $(n, M, d)_q$-HCC。根据上述引理，考虑由 $\sim$给出的等价类对 C进行的划分。从每个等价类中选取一个代表元，则这些代表元构成一个跳频序列集（简称 FHS 集）。通过上述方式由一个 $(n, M, d)_q$-HCC 得到的 FHS 集记为 $(n, M/n, n-d)_q$-FHS，它是一个满足以下条件的集合 $F \subseteq [q]^n$，其中 $|F| = M/n$

- 对每个 $\boldsymbol{x} \in F$和 $0 < i \le n-1$，$H_{\boldsymbol{x},\,\boldsymbol{x}}(i) := n - d_H(\boldsymbol{x}, \pi_i(\boldsymbol{x})) \le n-d$;
- 对每个互不相同的 $\boldsymbol{x}$、$\boldsymbol{y} \in F$和 $0 \le i \le n-1$，$H_{\boldsymbol{x},\,\boldsymbol{y}}(i) := n - d_H(\boldsymbol{x}, \pi_i(\boldsymbol{y})) \le n-d$;

其中，$H_{\boldsymbol{x},\boldsymbol{x}}(i)$ 和 $H_{\boldsymbol{x},\boldsymbol{y}}(i)$ 称为时延 i 下的自汉明相关函数和互汉明相关函数（见 [4]–[5], [16]）。跳频序列最初是为传输无线电信号而设计的，并已得到广泛研究。事实上，Niu 等人 [16] 的工作正是受到跳频序列研究的启发。

请注意，也可以直接利用汉明相关函数来定义跳频序列。为节省篇幅，我们不正式给出这样的定义。[4], [16] 中指出，跳频序列的两种定义是等价的，因为 HCC 与跳频序列之间存在一一对应关系。基于这种对应关系和（3），Niu 等人（见 [16, 推论 III.6]）证明，对于任意常数 $\epsilon > 0$ 和 $\lambda \ge n - (n - 2\sqrt{n})(1 - 1/q - \epsilon)$，都存在 $(n, M, \lambda)_q$-跳频序列，其

$$M \ge \frac{q^n(1 - n^2 e^{-\frac{\epsilon^2(\sqrt{n}-2)}{2}})}{n(Vol_q(n, n-\lambda-1)-1)}. \tag{8}$$

以下结果将（8）提高了 n 的线性因子。

命题 III.2. 对于正整数 n, q, λ 和绝对实数 $\tau, \epsilon \in (0, 1-1/q)$，满足 $1/q + \epsilon \le \lambda/n \le 1 - \tau$，存在 $(n, M, \lambda)_q$-跳频序列，使得

$$M \ge \frac{cq^n}{Vol_q(n, n-\lambda-1)},$$

其中， c是一个正实数，仅取决于 τ、 ϵ、 q。

证明。这是引理 III.1 (ii)、定理 II.1 以及 FHS 定义的直接推论 s. □

B. 纠错弱互不相关码

对于 $1 \leq \kappa \leq n$，若一个码 $C \in [q]^n$ 满足以下条件，则称其为 κ-弱互不相关码（简称 WMUC）：对于所有 $\kappa \leq \ell \leq n-1$，码 C 中任一码字长度为 ℓ 的真前缀，都不会作为码 C中另一码字（包括其自身）的后缀出现。形式化地说，若对于每个 $\boldsymbol{x}$、 $\boldsymbol{y} \in C$（二者可以相同）以及每个 $\kappa \leq \ell \leq n-1$，均有$(x_1, \ldots, x_\ell) \neq (y_{n-\ell+1}, \ldots, y_n)$，则称 C 为 κ-WMUC。若一个 WMUC 具有足够大的最小距离，则称其为纠错码。

Yazdi、Kiah、Gabrys 和 Milenkovic [20] 将 WMUC 和纠错 WMUC 引入作为一种技术，使得在基于DNA的存储系统中可以随机访问编码后的 DNA链。关于更多背景信息，请参阅 [20] 。

下一个结果表明，HCC 还具有同时成为弱互不相关码的优点。

命题 III.3. 对于正整数 n, q, κ 和满足 $1/q + \epsilon \leq \kappa/n \leq 1 - \tau$ 的绝对实数 $\tau, \epsilon \in (0, 1 - 1/q)$，存在 κ-WMU $(n, M, n - \kappa + 1)_q$-码，使得

$$M \geq \frac{cq^n}{Vol_q(n, n-\kappa)},$$

其中， c 是仅依赖于 τ, ϵ, q 的正实数。

证明. 根据定理 II.1，在命题的假设下，存在一个 $(n, M', n-\kappa+1)_q$-HCC $C' \subseteq [q]^n$，满足 $M' \geq \frac{cnq^n}{Vol_q(n,(n-\kappa)}$，其中 c 是仅~~依赖于 τ~~, ϵ, q 的正实数。根据引理 III.1 (ii)， C' 可划分为若干个两两不相交的 M'/n $C(\boldsymbol{x})$ 的并。与命题 III.2 的证明相同，令 C 为 C' 的一个子码，其构造方法是从划分中的每个 $C(\boldsymbol{x})$ 中恰好选取一个码字。于是， C 是一个 $(n, M, n-\kappa+1)_q$-码，满足 $M \geq \frac{cq^n}{Vol_q(n,(n-\kappa)}$。

还需证明 C 也是 κ-WMU。反设存在 $\boldsymbol{x}$、 $\boldsymbol{y} \in C$（可能相同），使得对于某个 $\kappa \leq \ell \leq n-1$，前缀（$x_1, \ldots, x_\ell$）与后缀（$y_{n-\ell+1}, \ldots, y_n$）相同。注意到 $\{\boldsymbol{x}, \boldsymbol{y}\} \subseteq C \subseteq C'$。由于 C 是一个 HCC， $\pi_{n-\ell}$（y） = （$y_{n-\ell+1}, \ldots, y_n, y_1, \ldots, y_{n-\ell}$） $\in C'$。由于 d_H（$x, \pi_{n-\ell}$（y）） $\leq n - \ell \leq n - \kappa$，其严格小于 C' 的最小距离，因此我们必有 $x = \pi_{n-\ell}$（y）。这意味着 $\{\boldsymbol{x}, \boldsymbol{y}\} \subseteq C$（$\boldsymbol{x}$）且 $|C \cap C$（$\boldsymbol{x}$）$| \geq 2$，矛盾。 □

IV. 一些有用引理的汇集 S

本节将汇集本文中用到的工具。

A. 局部稀疏图及其独立集

本小节将介绍局部稀疏图及其独立数的一个下界。图 $G = (V, E)$由顶点集 V和边集 E组成，其中 V是一个有限集， E是 V的一个由二元子集构成的族。若 $\{u, v\} \in E$，则称两个顶点 $u, v \in V$由一条边连接。两个相连的顶点称为邻点。图 G 中顶点 u 的邻点集合记作 $N_G(u)$。图 G中顶点 $u \in V$的度记作 $\deg_G(u)$，其值为该顶点邻域的大小，即 $\deg_G(u) = |N_G(u)|$。注意，当上下文已明确底层图 G 时，我们会省略 $N_G(u)$和 $\deg_G(u)$中的下标 G。图的最大度是其所有顶点的度中的最大值。对于子集 $S \subseteq V$，由 S 诱导的边集包含所有两个端点都在 S中的边，即边 $\{\{u, v\} \in E : u, v \in S\}$。如果每个顶点的邻域所诱导的边数都不超过某个常数，则称该图为局部稀疏图。

给定一个图 $G = (V, E)$，若子集 $I \subseteq V$ 中的任意两个顶点在 I 中均不相连，则称 $I \subseteq V$ 为 G 的独立集。G的独立数记为 $\alpha(G)$，即 G的最大独立集的大小。利用第 II-C 节中提到的贪心算法，不难看出，任何最大度为 D的图都至少有一个大小为 $\frac{|V|}{D+1}$的独立集。下面的引理 IV.1 是 [18, 的直接推论，它为局部稀疏图的独立数给出了一个更好的下界。

引理 IV.1. *[18, 推论 1]* 设 G 是一个有 n 个顶点的图，其最大度不超过 D，其中 $D \to \infty$ 如 $n \to \infty$。假设对于 G的每个顶点，其邻域至多诱导出 $\frac{D^2}{K}$ 条边，其中 $1 \le K \le D^2+1$。那么

$$\alpha(G) \ge (1-o(1)) \cdot \frac{|V|}{D} \cdot \ln(\min\{D, K\}),$$

其中 $o(1) \to 0$ 当 $K \to \infty$。

B. McDiarmid 不等式

我们将使用强大的 McDiarmid 不等式。粗略地说，该不等式指出，随机变量的 Lipschitz 函数会集中在其期望附近。

引理 IV.2. *[15]* (McDiarmid 不等式)。给定集合 $\mathcal{X}_1$、……、$\mathcal{X}_n$，设 X_1、……、X_n 是相互独立的随机变量，分别取值于 $\mathcal{X}_1$、……、$\mathcal{X}_n$。设 $f\colon \mathcal{X}_1 \times \cdots \times \mathcal{X}_n \to \mathbb{R}$ 是一个映射。若存在常数 c_1、……、$c_n \in \mathbb{R}$，使得对每个 $1 \le i \le n$,

$$\sup_{x_i' \in \mathcal{X}_i, x_1 \in \mathcal{X}_1, \ldots, x_n \in \mathcal{X}_n} |f(x_1, \ldots, x_i, \ldots, x_n) - f(x_1, \ldots, x_i', \ldots, x_n)| \le c_i,$$

（即，将 i 个坐标的值改为 x_i 时，f 的绝对值变化至多为 c_i），则对任意 $t > 0$,

$$\Pr[f(X_1, X_2, \ldots, X_n) - \mathbb{E}[f(X_1, X_2, \ldots, X_n)] \le -t] \le \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right).$$

C. 两个汉明球的交集体积

我们需要以下关于两个汉明球交集体积的估计。

引理 IV.3. *[12, 引理 4.2]* 对于正整数 n, q, d 以及两个满足$\tau \le d/n \le 1 - 1/q - \epsilon$的绝对实数 $\tau, \epsilon \in (0, 1-1/q)$，有

$$\frac{Vol_q(n, d-\tau n/2)}{Vol_q(n, d-1)} \le e^{-\Omega_{\tau,q}(1)\cdot n}.$$

此外，对于任意 $\boldsymbol{x}, \boldsymbol{y} \in [q]^n$,

$$\frac{|B(\boldsymbol{x}, d-1) \cap B(\boldsymbol{y}, d-1)|}{Vol_q(n, d-1)} = e^{-\Omega_{\tau,q}(1)\cdot d(\boldsymbol{x},\boldsymbol{y})}.$$

对于两个恒重汉明球的交集体积，也可以证明类似的上界。

引理 IV.4. *[12, 引理 4.3]* 给定正整数 n, d以及满足 $\tau \le d/n \le (1-\epsilon)p(1-p)$ 的实数 $p \in (0, 1)$ $\epsilon \in (0, 1/10)$ $\tau \in (0, p(1-p))$。则

$$\frac{Vol(n, d-\tau n/2; pn)}{Vol(n, d-1; pn)} \le 2e^{-\Omega_{\epsilon,\tau}(1)n}.$$

此外，对于任意 Hamming 重量为 pn 的 $\boldsymbol{x}, \boldsymbol{y} \in \{0,1\}^n$，

$$\frac{|B(\boldsymbol{x}, d-1; pn) \cap B(\boldsymbol{y}, d-1; pn)|}{Vol(n, d-1; pn)} = 2e^{-\Omega_\epsilon(1)\cdot(d(\boldsymbol{x},\boldsymbol{y}) + d(\boldsymbol{x},\boldsymbol{y})^2/(d-1))}.$$

V. 定理 II.1 的证明

给定两个实数 $\tau, \epsilon \in (0, 1-1/q)$，定理 II.1 的目标是证明，只要 $\tau \le d/n \le 1 - 1/q - \epsilon$，就存在规模较大的 $(n, M, d)_q$-HCC。定理 II.1 的证明将按照第 II-C 节概述的步骤进行。首先注意，$(n, M, d)_q$-码 C只有在对每个 $\boldsymbol{x} \in C$, $d(\boldsymbol{x}) \ge d$ 时，才是一个 $(n, M, d)_q$-HCC（回顾式 (7)）。因此，要构造规模较大的 HCC，我们需要证明存在一个较大的集合 $A \subseteq [q]^n$，使得对每个 $\boldsymbol{x} \in A$, $d(\boldsymbol{x}) \ge d$。正如下一个引理所示，这样的集合 A 确实存在。

引理 V.1. 设 n、q 为正整数，且 $\epsilon \in (0, 1-1/q)$ 为实数。设

$$A = \{\boldsymbol{x} \in [q]^n : d(\boldsymbol{x}) > n(1 - 1/q - \epsilon)\}.$$

Then

$$|A| \geq q^n \left(1 - (n-1)\exp(-\frac{\epsilon^2 n}{2})\right).$$

证明. 设 $X = (X_1, \ldots, X_n)$ 是从 $[q]^n$ 中均匀随机选取的元素。要证明该引理，只需证明

$$\Pr[d(X) \leq n(1-1/q-\epsilon)] \leq (n-1)\exp(-\frac{\epsilon^2 n}{2}). \tag{9}$$

回忆一下， $d(X) = \min\{d(X, \pi_i(X)) : 1 \leq i \leq n-1\}$，其中 $\pi_i(X) = (X_{i+1}, \ldots, X_{i+n})$。由并集界可得，

$$\begin{aligned}\Pr[d(X) \leq n(1-1/q-\epsilon)] &= \Pr[\exists\ 1 \leq i \leq n-1 \text{ s.t. } d(X, \pi_i(X)) \leq n(1-1/q-\epsilon)] \\ &\leq \sum_{i=1}^{n-1} \Pr[d(X, \pi_i(X)) \leq n(1-1/q-\epsilon)].\end{aligned}$$

根据 上面的讨论，要证明 (9)，只需证明对每个固定的 $1 \leq i \leq n-1$ ，

$$\Pr[d(X, \pi_i(X)) \leq n(1-1/q-\epsilon)] \leq \exp(-\frac{\epsilon^2 n}{2}).$$

注意，均匀随机地选择 $X = (X_1, \ldots, X_n) \in [q]^n$，等价于独立且均匀随机地选择 $X_1, \ldots, X_n \in [q]$。令

$$\delta : [q] \times [q] \to \{0, 1\}$$

表示克罗内克函数，使得当 $a, b \in [q]$, $\delta(a, b) = 1$ 时取值为 $a = b$，而当 $\delta(a, b) = 0$ 时取值为 $a \neq b$。根据定义，很容易验证对每个 $\boldsymbol{x} \in [q]^n$,

$$d(\boldsymbol{x}, \pi_i(\boldsymbol{x})) = n - \sum_{j=1}^{n} \delta(x_j, x_{j+i}). \tag{10}$$

我们将对 $f(x_1, \ldots, x_n) := d(\boldsymbol{x}, \pi_i(\boldsymbol{x}))$ 应用引理 IV.2。注意，对每个 $1 \leq j \leq n$，改变 x_j 的值，只会改变 (10) 右侧的 $\delta(x_{j-i}, x_j)$ 和 $\delta(x_j, x_{j+i})$ 的值，进而使 $d(\boldsymbol{x}, \pi_i(\boldsymbol{x}))$ 的绝对值以及 f 至多改变 2。换言之，对每个 $1 \leq j \leq n$,

$$\sup_{x_j', x_1, \ldots, x_n \in [q]} |f(x_1, \ldots, x_j, \ldots, x_n) - f(x_1, \ldots, x_j', \ldots, x_n)| \leq 2. \tag{11}$$

此外，根据期望的线性性，容易看出，对于每个 $1 \leq i \leq n-1$ 和 $1 \leq j \leq n$， $\mathbb{E}(\delta(X_j, X_{j+i})) = 1/q$，这意味着 $\mathbb{E}[f(X_1, \ldots, X_n) = n(1-1/q)$。将引理 IV.2 应用于 f，可得

$$\Pr\left[d(X, \pi_i(X)) - n(1-1/q) \leq -\epsilon n\right] \leq \exp(-\frac{\epsilon^2 n}{2}).$$

将上述不等式代入前述并集界，即可证明 (9)，进而证明引理 V.1。 □

注 V.2. 利用鞅的标准集中不等式， *[16,* 引理 III.*4]* 证明了一个稍弱的结果

$$|\{\boldsymbol{x} \in [q]^n : d(\boldsymbol{x}) > (n - 2\sqrt{n})(1-1/q-\epsilon)\}| \geq q^n \left(1 - n^2 e^{-\frac{\epsilon^2(\sqrt{n}-2)}{2}}\right).$$

注 V.3. 根据定义不难验证， A 是一个最小距离为 1 的 HCC。因此，类似于引理 III.1 (ii) 的证明，可以证明，在等价关系 $\sim$下， A 可以划分为一个等价类族V，其中每个等价类恰好包含 n 个来自 $[q]^n$的不同向量。由引理 V.1 可知

$$|V| = \frac{|A|}{n} \geq \frac{1}{n} \cdot q^n \left(1 - (n-1)\exp(-\frac{\epsilon^2 n}{2})\right). \tag{12}$$

此外，根据 A 的定义，对于每个 $C(\boldsymbol{x}) \in V$,

$$d(\boldsymbol{x}) > n(1-1/q-\epsilon), \tag{13}$$

对于每个互不相同的 $C(\boldsymbol{x})$ $C(\boldsymbol{y}) \in V$，

$$C(\boldsymbol{x}) \cap C(\boldsymbol{y}) = \emptyset. \tag{14}$$

为证明定理 II.1，我们将在顶点集 V 上构造一个辅助图，然后对该图应用引理 IV.1，具体如下。

引理 V.4. 采用定理 II.1 中的记号，固定某个 $d \in [\tau n, n(1-1/q-\epsilon)]$,，令 $G_{HCC} = (V, E)$ 为如下构造的图，其中 V 是将 A 划分为若干等价类的集合，并且两个不同的顶点（或等价类） $C(\boldsymbol{x})$ $C(\boldsymbol{y}) \in V$ 在 E 中构成一条边，当且仅当

$$d(C(\boldsymbol{x}), C(\boldsymbol{y})) = \min\{d(\boldsymbol{x}_i, \boldsymbol{y}_j) : 0 \le i, j \le n-1\} \le d-1.$$

则有如下结论：
（i）对于每个独立集 $I \subseteq V$，向量集合 $C := \bigcup_{C(\boldsymbol{x}) \in I} C(\boldsymbol{x})$ 构成一个 $(n, n|I|, d)_q$-HCC；（ii） G_{HCC} 的最大度至多为 $D := Vol_q(n, d-1)$；（iii）对于 G_{HCC} 的每个顶点，其邻域诱导的边数至多为 $\frac{D^2}{K}$，其中 $K = e^{\Omega_{\tau,q}(1)\cdot n}$。

证明。为证明（i），首先注意，对于每个 $C(\boldsymbol{x}) \in V$， $C(\boldsymbol{x})$ 由 n 个互不相同的向量组成，这意味着 C 确实是一个 HCC。此外，由 (14) 可得 $|C| = n|I|$。最后，为证明 $d(C) \ge d$，设 $\boldsymbol{c}^1$、 $\boldsymbol{c}^2 \in C$ 是两个不同的码字。一方面，若对于某个 $C(\boldsymbol{x}) \in I \subseteq V$，有 $\{\boldsymbol{c}^1, \boldsymbol{c}^2\} \subseteq C(\boldsymbol{x})$，则由 (13) 可得

$$d(\boldsymbol{c}^1, \boldsymbol{c}^2) \ge d(\boldsymbol{x}) > n(1-1/q-\epsilon) \ge d.$$

另一方面，若对于互不相同的 $C(\boldsymbol{x})$ $C(\boldsymbol{y}) \in I$，有 $\boldsymbol{c}^1 \in C(\boldsymbol{x})$ 且 $\boldsymbol{c}^2 \in C(\boldsymbol{y})$，则由于 I 是 G 中的一个独立集，我们有

$$d(\boldsymbol{c}^1, \boldsymbol{c}^2) \ge d(C(\boldsymbol{x}), C(\boldsymbol{y})) \ge d,$$

至此，（i）的证明完成。

为证明（ii），注意对于每个 $C(\boldsymbol{x}) \in V$，

$$\begin{aligned}
\deg(C(\boldsymbol{x})) &= |\{C(\boldsymbol{y}) \in V \setminus \{C(\boldsymbol{x})\} : d(C(\boldsymbol{x}), C(\boldsymbol{y})) \le d-1\}| \\
&= |\{C(\boldsymbol{y}) \in V \setminus \{C(\boldsymbol{x})\} : \min_{0 \le i \le n-1} d(\boldsymbol{x}, \pi_i(\boldsymbol{y})) \le d-1\}| \\
&= |\{C(\boldsymbol{y}) \in V \setminus \{C(\boldsymbol{x})\} : \exists\ i \in [n] \text{ s.t. } \pi_i(\boldsymbol{y}) \in B(\boldsymbol{x}, d-1)\}| \\
&\le |B(\boldsymbol{x}, d-1)| = D,
\end{aligned}$$

正如所需。

为证明（iii），固定一个 顶点 $C(\boldsymbol{x}) \in V$，令 Γ 表示由 $C(\boldsymbol{x})$ 的邻域诱导的子图。划分 $V(\Gamma) = S \cup T$，其中

$$S = \{C(\boldsymbol{y}) \in V(\Gamma) : d(C(\boldsymbol{x}), C(\boldsymbol{y})) \le d - \tau n/2\}$$

and

$$T = \{C(\boldsymbol{y}) \in V(\Gamma) : d - \tau n/2 < d(C(\boldsymbol{x}), C(\boldsymbol{y})) \le d-1\}.$$

我们有如下断言。

断言 V.5. 存在某个 $K = e^{\Omega_{\tau,\ q}(1)\cdot n}$ 使得 $|S| \le D/K$，并且对于每个顶点 $C(\boldsymbol{y}) \in T$，$\deg_\Gamma(C(\boldsymbol{y})) \le D/K$。

鉴于该断言的正确性，可以推出

$$\begin{aligned}
|E(\Gamma)| &= \frac{1}{2}\left(\sum_{C(\boldsymbol{y}) \in S} \deg_\Gamma(C(\boldsymbol{y})) + \sum_{C(\boldsymbol{y}) \in T} \deg_\Gamma(C(\boldsymbol{y}))\right) \\
&\le \frac{1}{2}\left(|S| \cdot D + |T| \cdot \frac{D}{K}\right) \le \frac{D^2}{K},
\end{aligned}$$

从而完成（iii）的证明。 □

尚需证明断言 V.5。

断言 V.5 的证明。为证明第一个不等式，注意到

$$\begin{aligned}|S| &= |\{C(\boldsymbol{y}) \in V(\Gamma) : d(C(\boldsymbol{x}), C(\boldsymbol{y})) \le d - \tau n/2\}| \\ &= |\{C(\boldsymbol{y}) \in V(\Gamma) : \min_{0\le i\le n-1} d(\boldsymbol{x}, \pi_i(\boldsymbol{y})) \le d - \tau n/2\}| \\ &= |\{C(\boldsymbol{y}) \in V(\Gamma) : \exists\ i \in [n] \text{ s.t. } \pi_i(\boldsymbol{y}) \in B(\boldsymbol{x}, d - \tau n/2)\}| \\ &\le |B(\boldsymbol{x}, d - \tau n/2)| = Vol_q(n, d - \tau n/2).\end{aligned}$$

因此，

$$\frac{|S|}{D} \le \frac{Vol_q(n, d - \tau n/2)}{Vol_q(n, d-1)} \le e^{-\Omega_{\tau,q}(1)\cdot n}, \tag{15}$$

其中，最后一个不等式由引理 IV.3 的第一个等式得出。

为证明第二个不等式，注意对于每个 $C(\boldsymbol{y}) \in T$，我们有

$$\begin{aligned}N_\Gamma(C(\boldsymbol{y})) &= N(C(\boldsymbol{x})) \cap N(C(\boldsymbol{y})) \\ &= \{C(\boldsymbol{v}) \in V \setminus \{C(\boldsymbol{x}), C(\boldsymbol{y})\} : d(C(\boldsymbol{v}), C(\boldsymbol{x})) \le d-1,\ d(C(\boldsymbol{v}), C(\boldsymbol{y})) \le d-1\} \\ &= \{C(\boldsymbol{v}) \in V \setminus \{C(\boldsymbol{x}), C(\boldsymbol{y})\} : \min_{0\le i\le n-1} d(\boldsymbol{v}, \pi_i(\boldsymbol{x})) \le d-1, \min_{0\le j\le n-1} d(\boldsymbol{v}, \pi_j(\boldsymbol{y})) \le d-1\} \\ &= \{C(\boldsymbol{v}) \in V \setminus \{C(\boldsymbol{x}), C(\boldsymbol{y})\} : \exists\ 0 \le i,j \le n-1 \text{ s.t. } \boldsymbol{v} \in B(\pi_i(\boldsymbol{x}), d-1) \cap B(\pi_j(\boldsymbol{y}), d-1)\}, \\ &\subseteq \big\{C(\boldsymbol{v}) \in V : \boldsymbol{v} \in \bigcup_{0\le i,j\le n-1} B(\pi_i(\boldsymbol{x}), d-1) \cap B(\pi_j(\boldsymbol{y}), d-1)\big\},\end{aligned}$$

这意味着

$$|N_\Gamma(C(\boldsymbol{y}))| \le \sum_{0\le i,j\le n-1} |B(\pi_i(\boldsymbol{x}), d-1) \cap B(\pi_j(\boldsymbol{y}), d-1)|.$$

因此可得

$$\begin{aligned}\frac{\deg_\Gamma(C(\boldsymbol{y}))}{D} = \frac{|N_\Gamma(C(\boldsymbol{y}))|}{Vol_q(n,d-1)} &\le \frac{\sum_{0\le i,j\le n-1} |B(\pi_i(\boldsymbol{x}), d-1) \cap B(\pi_j(\boldsymbol{y}), d-1)|}{Vol_q(n,d-1)} \\ &= \sum_{0\le i,j\le n-1} \frac{B(\pi_i(\boldsymbol{x}), d-1) \cap B(\pi_j(\boldsymbol{y}), d-1)}{Vol_q(n,d-1)} \\ &\le \sum_{0\le i,j\le n-1} e^{-\Omega_{\tau,q}(1)\cdot d(\pi_i(\boldsymbol{x}), \pi_j(\boldsymbol{y}))} \\ &\le n^2 \cdot e^{-\Omega_{\tau,q}(1)\cdot n} = e^{-\Omega_{\tau,q}(1)\cdot n},\end{aligned} \tag{16}$$

其中，最后两个不等式由引理 IV.3 的第二个等式以及假设 $d(\pi_i(\boldsymbol{x}), \pi_j(\boldsymbol{y})) \ge d(C(\boldsymbol{x}), C(\boldsymbol{y})) \ge d \ge \tau n$ 得出。

结合式（15）和式（16）即可直接得到该结论。 □

下面给出定理 II.1 的证明。

定理 II.1 的证明。令 $G_{HCC} = (V, E)$ 为引理 V.4 中定义的图。结合式（12）以及引理 V.4 的（ii）、（iii），应用引理 IV.1 可得

$$\begin{aligned}\alpha(G_{HCC}) &\ge (1 - o(1)) \cdot \frac{|V|}{D} \cdot \ln(\min\{D, K\}) \\ &\ge (1 - o(1)) \cdot \frac{q^n}{n} \cdot \frac{1}{Vol_q(n, d-1)} \cdot \ln(\min\{Vol_q(n, d-1), e^{\Omega_{\tau,q}(n)}\}) \\ &\ge \frac{cq^n}{Vol_q(n, d-1)},\end{aligned}$$

其中，c 是一个与 n无关的常数。因此，根据引理 V.4 (i)，存在具有$M \ge n \cdot \alpha(G_{HCC})$的$(n, M, d)_q$-HCC，从而完成定理的证明。 □

VI. 定理 II.2 的证明

给定实数 $p \in (0,1)$ $\epsilon \in (0, 1/10)$ $\tau \in (0, p(1-p))$，定理 II.2 的目标是证明：只要 $\tau \le d/n \le (1-\epsilon)p(1-p)$，就存在规模较大的$(n, M, d;\ pn)$-光正交码。对于 $pn \in \mathbb{Z}_+$，令$\binom{[n]}{pn} = \{\boldsymbol{x} \in \{0,1\}^n :\ wt(\boldsymbol{x}) = pn\}$。注意，只有当对每个 $\boldsymbol{x} \in C$， $d(\boldsymbol{x}) \ge d$时，$(n, M, d;\ pn)$-码 C才是$(n, M, d;\ pn)$-光正交码。与第 V 节开头的讨论类似，要构造规模较大的光正交码，我们需要证明存在一个较大的集合 $B \subseteq \binom{[n]}{pn}$，使得对每个 $\boldsymbol{x} \in B$， $d(\boldsymbol{x}) \ge d$。这一结果的证明思路与引理 V.1 的证明相似，但技术上稍微复杂一些。

引理 VI.1. 设 $\epsilon > 0,\ 0 < p < 1$ 为实数， n 为正整数。设

$$B = \{\boldsymbol{x} \in \binom{[n]}{pn} : d(\boldsymbol{x}) > (1-\epsilon)np(1-p)\}.$$

Then

$$|B| \ge \binom{n}{pn}\left(1 - n^{3/2}\exp\big(-\frac{\Omega_{p,\epsilon}(1)\cdot n}{2}\big)\right).$$

证明. 设 $X = (X_1, \ldots, X_n) \in \{0,1\}^n$ 为一个随机向量，其中 $X_1, \ldots, X_n$ 是独立伯努利随机变量，满足对于 $1 \le i \le n$， $\Pr[X_i = 1] = p$ 且 $\Pr[X_i = 0] = 1-p$。为方便记号，令 $\mu = np(1-p)$。根据条件概率定律，

$$\Pr[d(X) \le (1-\epsilon)\mu | wt(\boldsymbol{x}) = pn] = \frac{\Pr[d(X) \le (1-\epsilon)\mu \wedge wt(\boldsymbol{x}) = pn]}{\Pr[wt(\boldsymbol{x}) = pn]} = \frac{\binom{n}{pn} - |B|}{\binom{n}{pn}}.$$

要证明该引理，只需证明

$$\Pr[d(X) \le (1-\epsilon)\mu | wt(\boldsymbol{x}) = pn] \le n^{3/2}\exp\big(-\frac{\Omega_{p,\epsilon}(1)\cdot n}{2}\big).$$

回忆 $d(X) = \min\{1 \le i \le n-1 :\ d(X, \pi_i(X))\}$，其中 $\pi_i(X) = (X_{i+1}, .., X_i)$。于是，

$$\begin{aligned}
&\Pr[d(X) \le (1-\epsilon)\mu | wt(X) = pn] \\
=&\Pr[\exists\ 1 \le i \le n-1 \text{ s.t. } d(X, \pi_i(X)) \le (1-\epsilon)\mu | wt(X) = pn] \\
\le&\sum_{i=1}^{n-1} \Pr[d(X, \pi_i(X)) \le (1-\epsilon)\mu | wt(X) = pn] \\
=&\sum_{i=1}^{n-1} \frac{\Pr[d(X, \pi_i(X)) \le (1-\epsilon)\mu \wedge wt(\boldsymbol{x}) = pn]}{\Pr[wt(\boldsymbol{x}) = pn]} \\
\le&\sum_{i=1}^{n-1} \frac{\Pr[d(X, \pi_i(X)) \le (1-\epsilon)\mu]}{\Pr[wt(X) = pn]} \\
\le&\left(\sum_{i=1}^{n-1} \Pr[d(X, \pi_i(X)) \le (1-\epsilon)\mu]\right) \cdot \sqrt{2\pi np(1-p)} \cdot \ell(n),
\end{aligned}$$

其中 $\ell(n) = \exp(-\frac{1}{12n+1} + \frac{1}{12pn} + \frac{1}{12(1-p)n})$，第一个不等式由并集界得出，最后一个不等式则由以下事实得出

$$\Pr[wt(X) = pn] = \binom{n}{pn} p^{pn}(1-p)^{(1-p)n}$$

以及斯特林近似

$$\binom{n}{pn} p^{pn}(1-p)^{(1-p)n} \ge \frac{1}{\sqrt{2\pi np(1-p)}} \cdot \exp\big(\frac{1}{12n+1} - \frac{1}{12pn} - \frac{1}{12(1-p)n}\big).$$

根据上述讨论，要证明该引理，只需证明对于每个固定的 $1 \le i \le n-1$,

$$\Pr[d(X, \pi_i(X)) \le (1-\epsilon)\mu] \le \exp\big(-\frac{\Omega_{p,\epsilon}(1)\cdot n}{2}\big).$$

与引理 V.1 的证明类似，我们将对 $f(x_1,\ldots,x_n) := d(\boldsymbol{x},\pi_i(\boldsymbol{x})) = n-\sum_{j=1}^{n}\delta(x_j,x_{j+i})$ 应用引理 IV.2。由于对每个 $i\neq j$, $\mathbb{E}[\delta(X_i,X_j)] = p^2+(1-p)^2,$，根据期望的线性性，我们有 $\mathbb{E}[f(X_1,X_2,\ldots,X_n)] = 2np(1-p) = 2\mu$。由式 (11)，对 f 应用引理 IV.2 可得

$$\Pr\left[d(X,\pi_i(X)) - 2\mu \leq -(1+\epsilon)\mu\right] \leq \exp\Big(-\frac{(1+\epsilon)^2p^2(1-p)^2n}{2}\Big) = \exp\Big(-\frac{\Omega_{p,\epsilon}(1)\cdot n}{2}\Big),$$

于是我们得到了所需的结论。 □

与注 V.3 中的讨论类似，由引理 VI.1 得到的集合 B 可划分为一族两两不相交的等价类 V，其中

$$|V| = \frac{|B|}{n} \geq \frac{1}{n}\cdot\binom{n}{pn}\left(1-n^{3/2}\exp(-\frac{\Omega_{p,\epsilon}(1)n}{2})\right). \tag{17}$$

为了证明定理 II.2，我们将在顶点集 V 上构造一个辅助图，具体如下。

引理 VI.2. 给定 $d\in[\tau n,(1-\epsilon)np(1-p)],$，设 $G_{OOC}=(V,E)$ 是如下构造的图，其中 V 是划分 B 的等价类族，且两个不同的顶点 $C(\boldsymbol{x})$ $C(\boldsymbol{y})$ $\in V$ 相连，当且仅当 $d(C(\boldsymbol{x})$ $C(\boldsymbol{y}))$ $\leq d-1$。则有：

(i) 对每个独立集 $I\subseteq V$，向量集 $C := \bigcup_{C(\boldsymbol{x})\in I} C(\boldsymbol{x})$ 构成一个 $(n,n|I|,d;\ pn)$-光正交码；(ii) G_{OOC} 的最大度至多为 $D := Vol(n,d-1;\ pn)$；(iii) 对于 G_{OOC} 的每个顶点，其邻域所诱导的边数至多为 $\frac{D^2}{K}$，其中 $K = e^{\Omega_{\tau,\ \epsilon}(1)\cdot n}$。

由于引理 VI.2 的证明与引理 V.4 的证明非常相似，我们将其放在附录中。接下来给出定理 II.2 的证明。

定理 II.2 的证明。设 $G_{OOC}=(V,E)$ 为引理 VI.2 中定义的图。于是，结合式 (17) 应用引理 IV.1，并利用引理 VI.2 的 (ii)、（iii），可得

$$\begin{aligned}\alpha(G_{OOC}) &\geq (1-o(1))\cdot\frac{|V|}{D}\cdot\ln(\min\{D,K\})\\ &\geq (1-o(1))\cdot\frac{1}{n}\cdot\binom{n}{pn}\cdot\frac{1}{Vol(n,d-1;pn)}\cdot\ln(\min\{Vol(n,d-1;pn),e^{\Omega_{\tau,\epsilon}(n)}\})\\ &\geq \frac{c\binom{n}{pn}}{Vol(n,d-1;pn)},\end{aligned}$$

其中 c 是一个与 n 无关的常数。因此，根据引理 VI.2 的 (i)，存在 $(n,M,d;\ pn)$-光正交码，其 $M\geq n\cdot\alpha(G_{OOC})$，定理得证。 □

VII. 结论

本文给出了跳频循环码和光正交码的改进 GV 型界，进而分别得到了非线性循环码和非线性定重循环码的改进 GV 型界。仍有若干有趣的问题尚待解决。

问题 VII.1. 已知对于素数幂 $q\geq 49$，存在一类称为代数几何码的码，能够通过指数因子 $\exp(\Omega(n))$ 显著改进 GV 界（1）（见 *[10], [21]*）。这种改进是否也适用于（非线性）循环码？定重循环码又如何？

问题 VII.2. 本文提及的所有结果都是非构造性的。我们能否显式构造一类渐近良好的（非线性）循环码？或者更准确地说，我们能否给出适用于（非线性）循环码的 Zyablov 型界？定重循环码又如何？

致谢

前两位作者　　谨此感谢 Yanxun Chang 教授关于定重循环码的精彩报告　　.

参考文献

[1] F. R. K. Chung, J. A. Salehi 和 V. K. Wei，《光正交码：设计、分析与应用》，IEEE 信息论汇刊，第 35 卷，第 3 期，第 595–604 页，1989年。[2] J.-H. Chung 和 K. Yang，《具有新参数的渐近最优光正交码》，IEEE 信息论汇刊，第 59 卷，第 6 期，第 3999–4005 页，2013年。[3] Chung, Fan R. K. 和 Salehi, Jawad A. 和 Wei, Victor K.，《更正："光正交码：设计、分析与应用" [IEEE 信息论汇刊 35 (1989)，第 3 期，595–604；MR1022081 (91g:94032)]》，IEEE 信息论汇刊，第 38 卷，第 4 期，第 1429 页，1992年。[4] C. Ding、R. Fuji-Hara、Y. Fujiwara、M. Jimbo 和 M. Mishima，《跳频序列集：界与最优构造》，IEEE 信息论汇刊，第 55 卷，第 7 期，第 3297–3304 页，2009年。[5] C. Ding、Y. Yang 和 X. Tang，《由线性循环码构造的最优跳频序列集》，IEEE 信息论汇刊，第 56 卷，第 7 期，第 3605–3612 页，2010年。[6] S. Dougherty、J.-L. Kim 和 P. Sol´e，《编码理论中的开放问题》，载于《非交换环及其应用》，丛书：当代数学，美国数学学会，罗德岛州普罗维登斯，2015年，第 634 卷，第 79–99 页。[7] E. N. Gilbert，《信号字母表的比较》，《贝尔系统技术期刊》，第 31 卷，第 3 期，第 504–522 页，1952年。[8] I. Haviv、M. Langberg、M. Schwartz 和 E. Yaakobi，《达到吉尔伯特–瓦尔沙莫夫界的非线性循环码》，载于《2017年IEEE国际信息论研讨会（ISIT）》，2017年，第 586–588 页。[9] Haviv, Ishay 和 Langberg, Michael 和 Schwartz, Moshe 和 Yaakobi, Eitan，《达到吉尔伯特–瓦尔沙莫夫界的非线性循环码》，载于《2017年IEEE国际信息论研讨会（ISIT）》，2017年，第 586–588 页。[10] T. Høholdt、J. van Lint 和 G. Pellikaan，《代数几何码》。荷兰：爱思唯尔，1998年，第 871–961 页。[11] T. Jiang 和 A. Vardy，《二元码规模的吉尔伯特–瓦尔沙莫夫界的渐近改进》，IEEE 信息论汇刊，第 50 卷，第 8 期，第 1655–1664 页，2004年。[12] J. Kim、H. Liu 和 T. Tran，《交集体积的指数衰减及其在列表可译码性和吉尔伯特–瓦尔沙莫夫型界中的应用》，arXiv 预印本 arXiv:2112.11274，2021年。[13] L. Lan、Y. Chang 和 L. Wang，《循环定重码：上界与新的最优构造》，IEEE 信息论汇刊，第 62 卷，第 11 期，第 6328–6341 页，2016年。[14] V. I. 列文斯坦，《定重码的上界估计》，《信息传输问题》，第 7 卷，第 3–12 页，1971年。[15] C. McDiarmid，《有界差分法》，丛书：伦敦数学会讲义系列。剑桥大学出版社，1989年，第 148–188 页。[16] X. Niu、C. Xing 和 C. Yuan，《跳频序列的渐近吉尔伯特–瓦尔沙莫夫界》，IEEE 信息论汇刊，第 66 卷，第 2 期，第 1213–1218 页，2020年。[17] R. Pan 和 Y. Chang，《最大光正交签名图样码的组合构造》，《离散数学》，第 313 卷，第 24 期，第 2918–2931 页，2013年。[18] F. Pirot 和 E. Hurley，《用一阶矩法为局部稀疏图着色》，arXiv 预印本 arXiv:2109.15215v3，2021年。[19] R. Roth，《编码理论导论》。剑桥大学出版社，2006年。[20] S. M. H. Tabatabaei Yazdi、H. M. Kiah、R. Gabrys 和 O. Milenkovic，《用于基于DNA的数据存储的互不相关引物》，IEEE 信息论汇刊，第 64 卷，第 9 期，第 6283–6296 页，2018年。[21] M. A. Tsfasman、S. G. Vl˘adut, 和 T. Zink，《模曲线、志村曲线与优于瓦尔沙莫夫–吉尔伯特界的 Goppa 码》，《数学新闻》，第 109 卷，第 21–28 页，1982年。[22] R. Varshamov，《纠错码中信号数量的估计》，《科学院学报》，第 117 卷，第 739–741 页，1957年。[23] V. Vu 和 L. Wu，《改进 q元码的吉尔伯特–瓦尔沙莫夫界》，IEEE 信息论汇刊，第 51 卷，第 9 期，第 3200–3208 页，2005年。[24] G.-C. Yang 和 T. Fuja，《具有不等自相关与互相关约束的光正交码》，IEEE 信息论汇刊，第 41 卷，第 1 期，第 96–106 页，1995年。

VIII. 附录：引理 VI.2 的证明

引理 VI.2 的证明。注意，给定引理 V.4 前两项的证明后，很容易证明引理 VI.2 的（i）和（ii），因此为简洁起见，我们省略这两个证明。为证明（iii），令 Γ、S 和 T 如引理 V.4 的证明中所定义。于是，只需证明在引理 VI.2 的设定下，断言 V.5 仍然成立。

断言 VIII.1（断言 V.5 的重述）。存在某些 $K = e^{\Omega_{\tau,\epsilon,p}(1)\cdot n}$ 使得 $|S| \leq D/K$，并且对于每个顶点 $C(\boldsymbol{y}) \in T$，$\deg_\Gamma(C(\boldsymbol{y})) \leq D/K$。

为证明断言的第一部分，注意到

$$\frac{|S|}{D} = \frac{|\{C(\boldsymbol{y}) \in V(\Gamma) : d(C(\boldsymbol{x}), C(\boldsymbol{y})) \leq d - \tau n/2\}|}{D} \leq \frac{Vol(n, d-\tau n/2; pn)}{Vol(n, d-1; pn)} \leq 2e^{-\Omega_{\tau,\epsilon}(1)n}, \tag{18}$$

其中，第一个不等式由证明（15）中第一个不等式所用的相同技巧得出，第二个不等式则由引理 IV.4 的第一个等式得出。

至于第二部分　　对于该断言，采用证明 (16) 的相同策略，可以证明对于每个 $C(\boldsymbol{y}) \in T$　　，

$$\begin{aligned}\frac{\deg_\Gamma(C(\boldsymbol{y}))}{D} &= \frac{|N_\Gamma(C(\boldsymbol{y}))|}{Vol(n,d-1;pn)} \leq \frac{\sum_{0\leq i,j\leq n-1}|B(\pi_i(\boldsymbol{x}),d-1)\cap B(\pi_j(\boldsymbol{y}),d-1)|}{Vol(n,d-1;pn)} \\ &= \sum_{0\leq i,j\leq n-1}\frac{B(\pi_i(\boldsymbol{x}),d-1;pn)\cap B(\pi_j(\boldsymbol{y}),d-1)}{Vol_q(n,d-1;pn)} \\ &\leq \sum_{0\leq i,j\leq n-1} e^{-\Omega_\epsilon(1)\cdot(d(\pi_i(\boldsymbol{x}),\pi_j(\boldsymbol{y}))+d(\pi_i(\boldsymbol{x}),\pi_j(\boldsymbol{y}))^2/(d-1))} \\ &\leq n^2\cdot e^{-\Omega_{\tau,\epsilon,p}(1)\cdot n} = e^{-\Omega_{\tau,\epsilon,p}(1)\cdot n},\end{aligned} \tag{19}$$

其中，最后两个不等式由引理 IV.4 的第二个等式以及假设 $d(\pi_i(\boldsymbol{x}),\pi_j(\boldsymbol{y})) \geq d(C(\boldsymbol{x}),C(\boldsymbol{y})) \geq d \geq \tau n$ 得出。

将（18）和（19）结合起来，即可直接得出该结论。 □