中国科学技术大学

博士学位论文

关于数据安全中若干编码问题的研究

作者姓名： 陈婷婷

学科专业： 网络空间安全

导师姓名： 张先得 特任教授

完成时间： 二〇二二年五月二十七日

University of Science and Technology of China

A dissertation for doctor's degree

Research on some coding problems in data security

Author: Tingting Chen

Speciality: Cyberspace Security

Supervisor: Prof. Xiande Zhang

Finished time: May 27, 2022

中国科学技术大学学位论文原创性声明

本人声明所呈交的学位论文，是本人在导师指导下进行研究工作所取得的成果。除已特别加以标注和致谢的地方外，论文中不包含任何他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的贡献均已在论文中作了明确的说明。

作者签名：陈婷婷　　签字日期：2022.5.27

中国科学技术大学学位论文授权使用声明

作为申请学位的条件之一，学位论文著作权拥有者授权中国科学技术大学拥有学位论文的部分使用权，即：学校有权按有关规定向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅，可以将学位论文编入《中国学位论文全文数据库》等有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。本人提交的电子文档的内容和纸质论文的内容相一致。

控阅的学位论文在解密后也遵守此规定。

☑ 公开　☐ 控阅（____年）

作者签名：陈婷婷　　导师签名：

签字日期：2022.5.27　　签字日期：2022.5.27

摘　要

近年来，随着互联网的蓬勃发展，数字化和网络化的时代逐步到来。而由此产生的海量数据，使得对大数据存储以及信息安全传输的需求日益增加，特别是社交网络、自媒体和短视频以及视频通话等软件，要求对大量数据进行实时存储、访问、传输和安全保护。为了保障数据安全，提高存储系统的数据可靠性以及访问和更新数据的效率，避免网络拥堵，研究存储系统的可靠性技术、负载均衡对于构造大规模的存储系统具有重要意义。另一方面，为了保障信息安全，防止信息泄露，研究在信息传输过程中，如何变换使其不被窃取或攻击破坏，具有重要价值。因此，本文主要从编码的角度出发，分别对经典存储系统下计算负载均衡及其更新问题、DNA 存储系统中纠正串联复制错误的纠错码以及可应用于 McEliece 密码系统的（广义）扭 Reed-Solomon（RS）码三方面进行了研究。特别地，前两点主要是针对于信息在存储时可靠性的研究，后一点是考虑信息在传输过程中为保证安全而所需的防御措施的研究。本文的主要研究工作和贡献陈述如下。

1. 针对数据计算负载均衡以及更新问题，我们研究了稀疏平衡的 MDS 码。为了能够有效地应用到实际系统中，我们考虑在较小的有限域上构造稀疏平衡的 MDS 码。虽然目前关于小域上稀疏平衡的 $[n,k]_q$ MDS 码存在性的研究较多，但是 q 被要求至少为 $n+\left\lceil \frac{k(k-1)}{n} \right\rceil$。本文中，当 $n \leqslant 2k$ 时，我们将其改进到了 $q \geqslant n-1$。更具体地，首先，当有限域 $\mathbb{F}_q$ 的大小满足 $q \geqslant n$ 时，我们给出了一个由生成矩阵零模式刻画的稀疏 $[n,k]_q$ MDS 码存在的充分条件。这个充分条件，将构造 MDS 码这一代数问题转化为了一个组合问题（即构造满足条件 $(P_1)-(P_3)$ 的集族，具体定义见第2.4.1小节）。基于这个条件，我们通过设计几个多项式时间算法，找到了满足要求的集族对应的二元矩阵，从而构造出了码长满足 $n \leqslant 2k$ 的所有稀疏平衡的 $[n,k]_q$ MDS 码。进一步地，通过扩展坐标，我们将域的大小改进到 $q \geqslant n-1$。而当码长 $n > 2k$ 时，对任意整数 e,s,m，满足 $e \leqslant s-2$ 且 $m \leqslant p-1$，或者 $e = s-1$ 且 $m < \frac{p}{2}$，我们利用平衡和集 $A+B$，其中 $|A| = k$ 以及 $|B| = k-1$，构造出了所有稀疏平衡的 $[n = q = p^s, k = p^e m]_q$ MDS 码。
2. 在 DNA 存储系统中，由于 DNA 分子复制时容易发生串联复制突变，导致数据丢失或出错，为了恢复原始信息，我们研究了能够纠正串联复制错误的纠错码。特别地，这类纠错码的构造可以转化为构造一类 ℓ_1 度量下非负整数集合 $\mathbb{Z}_{\geqslant 0}$ 以及 $I_q = \{0,1,\cdots,q-1\}$ 上的常重码。但目前关于 ℓ_1 度量下的常重码问题，相关的结果比较少，特别是非负整数 $\mathbb{Z}_{\geqslant 0}$ 上的最优码，其

上下界都比较粗糙。本文中，给定一个常重码，我们利用其码字支集，一方面，刻画出了一个通用的必要条件（称为 UNC 条件），它表明了 ℓ_1 度量下的码与填充集族之间的关系；另一方面，给出了一个距离公式，利用该公式，可得到相应最大码字个数的上界。进一步地，根据 UNC 条件以及距离公式，我们将构造常重码问题转化为找到一个合适的填充集族，并在其每个区组上合理分配码字元素问题。由于受到重量 w 的限制，我们分别针对 $\mathbb{Z}_{\geqslant 0}$ 和 I_3 这两种字母集，确定了重量 $w \leqslant 4$ 的所有最优常重码。而对于一般的 w，当码长 n 充分大且满足 $n \equiv 1, w, -w+2, -2w+3 \pmod{w(w-1)}$ 时，我们确定了权重为 w 和距离为 $2w-2$ 的三元常重码最大码字个数。

3. 针对（广义）扭 RS 码问题，我们研究了（广义）扭 RS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 的性质以及相关构造问题。虽然关于扭 RS 码的构造性结果较多，但大多针对于添加一个扭结的情况，即 $\ell = 1$，且对于其对偶封闭性的研究较少。本文中，我们具体刻画了这一点，特别是当其所有估值点构成某个多项式根集合的时候，利用该多项式系数分布情况，我们给出了码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 对偶封闭的充分条件，并得到了相应的校验矩阵。基于这一结果，我们构造了相应的自对偶码。特别是当 $\ell = 1$ 时，所得自对偶码是 MDS 或近 MDS 码。而当 $\ell = 3$ 时，所得自对偶码的最小距离在 $n-k-2$ 和 $n-k+1$ 之间。

关键词：MDS 码；Reed-Solomon 码；DNA 存储；串联复制错误；常重码；扭 Reed-Solomon 码

ABSTRACT

In recent years, with the vigorous development of the Internet, the digital and networked world has gradually arrived. The resulting massive data has increased the demand for big data storage and information transmission security, especially software such as social networks, self-media, short videos, and video calls, which require real-time storage, access, transmission, and security protection of large amounts of data. On the one hand, in order to ensure data security, improve the data reliability of the storage system and the efficiency of accessing and updating data, and avoid network congestion, it is of great significance to study the reliability technology of storage systems and access data balance for constructing large-scale storage systems. On the other hand, in order to ensure information security and prevent information leakage, it is of great value to study how to transform it so that it will not be stolen or attacked during the information transmission process. Therefore, this dissertation mainly carries out an investigation starts from the point of view of coding theory in three aspects: computational load balancing and update problems in classic storage systems, error-correcting codes for correcting tandem-duplication errors in DNA storage systems, and (generalized) twisted Reed-Solomon codes that can be applied to McEliece cryptosystems. In particular, the first two points are mainly aimed at the research on the reliability of information during storage, and the latter point is the research on the defense measures required to consider the security of information during transmission. The main research work and contributions of this dissertation are listed as follows.

1. For the problems of updating and load balance on data computation, we study the sparse and balanced MDS codes. In order to be effectively applied to practical systems, we consider constructing sparse and balanced MDS codes over small finite fields. There are many studies on the existence of sparse and balanced $[n, k]_q$ MDS codes, but the size of q needs to be at least $n + \left\lceil \frac{k(k-1)}{n} \right\rceil$. In this dissertation, we improve it to $q \geqslant n - 1$. Firstly, given a finite field $\mathbb{F}_q$ which satisfies $q \geqslant n$, a sufficient condition for the existence of a sparse $[n, k]_q$ MDS code over $\mathbb{F}_q$ characterized by the zero pattern of the generator matrix is provided. This sufficient condition transforms the algebraic problem of constructing an MDS code into a combinatorial problem (that is, constructing a set system that satisfies the conditions $(P_1)-(P_3)$, see Section 2.4.1 for the specific definition). Based on this condition, we find the binary matrices corresponding to the required set system

by designing several algorithms with complexity running in polynomial time in n and k. And then using these matrices, we construct all sparse and balanced $[n,k]_q$ MDS codes provided that $n \leqslant 2k$. Further, by extending the coordinates, the demand for field size can be relaxed to $q \geqslant n-1$. For the case of $n > 2k$, we give some constructions for $q = n = p^s$ and $k = p^e m$ based on sumsets, when $e \leqslant s-2$ and $m \leqslant p-1$, or $e = s-1$ and $m < \frac{p}{2}$.

2. In the DNA-based storage system, data stored in this medium are subject to errors such as tandem duplication arising from various mutations, which need to be corrected to maintain data integrity. In order to restore the original information, we study the error-correcting codes for errors caused by tandem duplications. Constructing this type of error-correcting codes are equivalent to building a set of constant weight codes with ℓ_1 metric over non-negative integers $\mathbb{Z}_{\geqslant 0}$ or $I_q = \{0, 1, \cdots, q-1\}$. However, there are few results on the constant weight code problem under ℓ_1 metric, especially for the optimal code over $\mathbb{Z}_{\geqslant 0}$, the upper and lower bounds for its optimal code size are relatively rough. Thus we have considered them in this dissertation. More specifically, given a constant weight code, we establish a universal necessary condition (the so called UNC condition) and a distance formula for the collection of supports of all codewords, which reveals a connection between packing set systems and ℓ_1 metric codes. Therefore, by using group divisible designs and packings in combinatorial design theory, we give constructions of optimal codes over non-negative integers and I_3 with ℓ_1 weight $w \leqslant 4$ for all possible distances. In general, we also derive the size of the largest ternary code with constant weight w and distance $2w-2$ for sufficiently large length n satisfying $n \equiv 1, w, -w+2, -2w+3 \pmod{w(w-1)}$.
3. For the problem of (generalized) twisted RS code , we study the properties of (generalized) twisted RS code $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ and give some constructions. There are many results for twisted RS codes, but most of them concentrate on one twist, that is $\ell = 1$, and only a fraction of them are related to their structural properties. In this dissertation, we focus on it, especially when all its evaluation points form the root set of a polynomial, and then prove that the code $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ is closed under duality if the polynomial coefficients have certain distribution, and the corresponding parity check matrix is given. Using this result, we construct the corresponding self-dual code. Especially when $\ell = 1$, the resulting self-dual codes are either MDS codes or near MDS codes. And when $\ell = 3$, the minimum distance of the dual code is between $n-k-2$ and $n-k+1$.

Key Words: MDS code; Reed-Solomon code; DNA storage; Tandem-duplication error; Constant-weight code; Twisted Reed-Solomon code

目　录

第 1 章　绪　　论

1.1　研究背景

随着互联网的快速发展与信息技术的广泛应用，网络化、数字化以及智能化时代逐渐来临。根据中国互联网络信息中心（CNNIC）2022 年 2 月发布的《第 49 次中国互联网络发展状况统计报告》表明 [1]，截至 2021 年 12 月，中国网民规模已达 10 亿以上，且 20-49 岁网民占比高达 55.6%，而 50 岁及以上网民占比也比 2020 年有所增加，互联网正逐步融入各个群体的生活。随之而来的，各类互联网应用用户规模也呈增长态势。据 CNNIC 统计，即时通信、短视频、网络购物以及搜索引擎等应用软件，其网民使用率均高达 80% 以上，如图 1.1所示。这些软件在使用的同时，要求对海量数据进行实时存储、访问、传输和安全保护。由此，对网络空间实施安全保护，是非常重要和紧迫的。

	2020.12		2021.12		
应用	用户规模（万）	网民使用率	用户规模（万）	网民使用率	增长率
即时通信	98111	99.2%	100666	97.5%	2.6%
网络视频（含短视频）	92677	93.7%	97471	94.5%	5.2%
短视频	87335	88.3%	93415	90.5%	7.0%
网络支付	85434	86.4%	90363	87.6%	5.8%
网络购物	78241	79.1%	84210	81.6%	7.6%
搜索引擎	76977	77.8%	82884	80.3%	7.7%
网络新闻	74274	75.1%	77109	74.7%	3.8%
网络音乐	65825	66.6%	72946	70.7%	10.8%
网络直播	61685	62.4%	70337	68.2%	14.0%
网络游戏	51793	52.4%	55354	53.6%	6.9%
网络文学	46013	46.5%	50159	48.6%	9.0%
网上外卖	41883	42.3%	54416	52.7%	29.9%
网约车	36528	36.9%	45261	43.9%	23.9%
在线办公	34560	34.9%	46884	45.4%	35.7%
在线旅行预订	34244	34.6%	39710	38.5%	16.0%
在线医疗	21480	21.7%	29788	28.9%	38.7%
互联网理财	16988	17.2%	19427	18.8%	14.4%

图 1.1　2020.12-2021.12 中国各类互联网软件用户规模和网民使用率

网络空间安全的核心是信息安全，而对于信息安全，根据国际标准化组织

ISO 的定义：为数据处理系统建立和采取的技术及管理的安全保护，即保护计算机软硬件和数据不因偶然或恶意的原因而遭到破坏、更改和泄漏。总的来说，信息安全是确保数据存储或传输中，在未经授权时不被他人有意或无意地窃取和破坏。据 CNNIC 统计，我国网络安全问题主要集中于个人信息泄露、设备中病毒或木马等，如图1.2所示。因此，一方面，随着网络数据海量增加，为防止各类存储设备由于病毒或环境等因素而损坏，以及负载不均衡致使的网络拥堵，从而导致的数据丢失，保障数据安全，研究存储系统的可靠性技术、负载均衡对于构造大规模的存储系统具有重要意义。另一方面，为了防止信息泄露或被篡改，研究在信息传输过程中，如何变换使其不被泄露或攻击破坏，具有重要价值。

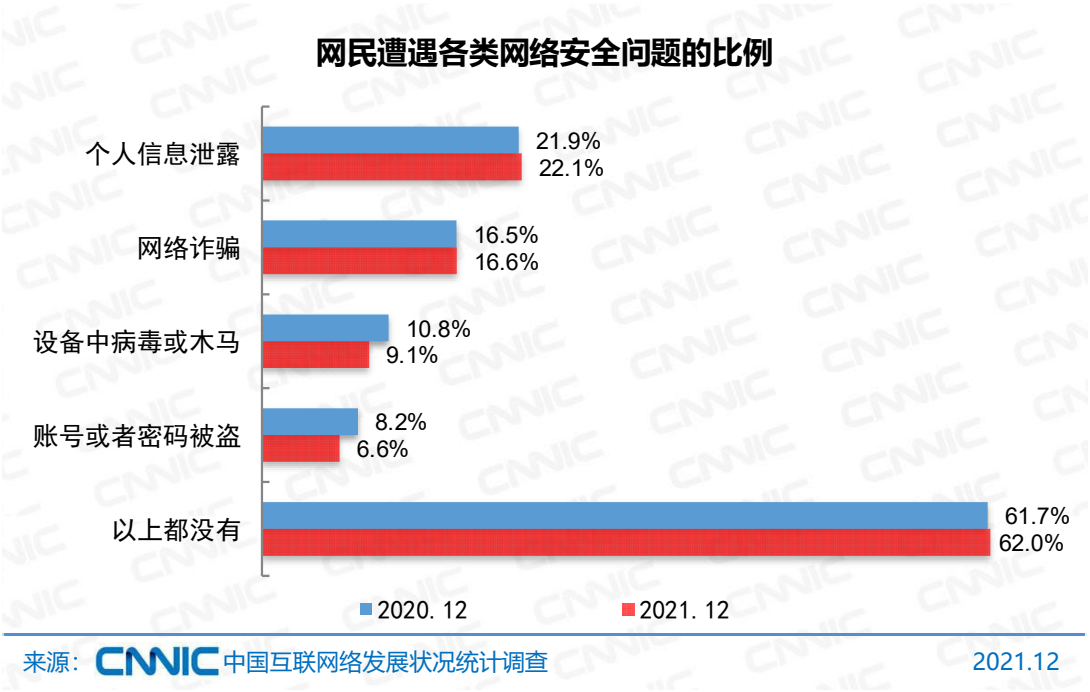


图 1.2　网络安全问题占比

基于上述问题，本文将从以下两个环节入手：数据存储和传输过程。在数据存储过程中，当某个存储节点发生故障时，为了防止数据丢失，需要在存储系统中引入冗余，从而提高节点故障时的可靠性。在经典的存储系统中，比较常见且简单的添加冗余的方式是多副本策略，即将数据复制多份进行存储。而纠删码是一种新型的添加冗余的存储策略，相对于多副本策略而言，其主要优点是在相同的冗余下，能够实现更高的可靠性，即存储开销相对较小。因此，本文主要考虑将编码技术应用于存储系统，即，采用纠删（错）码策略。另外一点，在数据传输过程中，为了防止他人攻击而导致数据丢失或破坏，我们考虑将编码技术应用其中，研究能够抵抗某些代数攻击的码。下面，我们将具体介绍应用于这两个课题的编码技术研究背景与研究现状，简要说明本文的研究内容和创新点。

1.1.1　经典存储系统下的纠删码策略和稀疏平衡码

在经典的存储系统中，以分布式存储系统为例，信息通常是按某个固定大小的块形式存储的，称其为信息块。为了提高存储系统可靠性，通常会对信息块采取一定的策略，添加一些与其相关的冗余块，亦或称之为校验块。信息块和校验块，统称为数据块。存放这些数据块的节点统称为存储节点，而存储开销，指的是所有数据块的个数与信息块的个数之比。以一个长度为 n，维数为 k 的线性纠删码 $\mathcal{C}$ 为例，我们首先给出它的具体定义。

定义 1.1　给定两个正整数 $n \geqslant k$，有限域 $\mathbb{F}_q$ 上一个码长为 n，维数为 k 的线性码 $\mathcal{C}$ 指的是 n 维向量空间 $\mathbb{F}_q^n$ 的一个 k 维子空间。对任意 $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}_q^n$，定义向量 $\boldsymbol{x} = (x_1, \cdots, x_n), \boldsymbol{y} = (y_1, \cdots, y_n)$ 之间的汉明距离（Hamming distance）为

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = |\{i \in [n] : x_i \neq y_i\}|.$$

则 $\mathcal{C}$ 的最小汉明距离为 $d = \min_{\boldsymbol{x} \neq \boldsymbol{y} \in \mathcal{C}} d_H(\boldsymbol{x}, \boldsymbol{y})$，称 $\mathcal{C}$ 为 $[n, k, d]_q$ 线性码。对于一个 $k \times n$ 阶矩阵 $\boldsymbol{G}$，若其行向量是 $\mathcal{C}$ 的一组基，则称 $\boldsymbol{G}$ 是 $\mathcal{C}$ 的一个生成矩阵。对于一个 $(n-k) \times n$ 阶矩阵 $\boldsymbol{H}$，若满足

$$\mathcal{C} = \{\boldsymbol{x} \in \mathbb{F}_q^n : \boldsymbol{H}\boldsymbol{x}^T = 0\},$$

则称 $\boldsymbol{H}$ 是 $\mathcal{C}$ 的一个校验矩阵。

那么，在数据存储时，可将其过程抽象为以下模型。首先将信息划分为 k 个固定大小的信息块，对应一个 k 长的向量 $\boldsymbol{c} = (c_1, \cdots, c_k)$，通过 $\mathcal{C}$ 对其进行编码，产生 $n-k$ 个校验块，得到一个 n 长的向量 $\boldsymbol{m} = (m_1, \cdots, m_n)$。再将这 n 个数据块分别存储到 n 个存储节点上。记 $\boldsymbol{G} = (g_1, \cdots, g_n)$，其中每一个 g_i 是一个 k 长的列向量。那么具体的编码过程如下：

$$\boldsymbol{c}\boldsymbol{G} = (c_1, \cdots, c_k)(g_1, \cdots, g_n) = (m_1, \cdots, m_n) = \boldsymbol{m}. \tag{1.1}$$

容易看出，利用 $\mathcal{C}$ 进行编码，其存储开销为 n/k 倍。而对于 $n-k+1$ 副本策略，在其冗余块个数也为 $n-k$ 的情况下，其存储开销为 $n-k+1$ 倍（大于 n/k 倍）。由此可看出，相对于多副本策略而言，在相同的冗余下，纠删码的存储开销更小。特别地，在上述编码策略中，若在编码后，保持信息块不变，即，$\boldsymbol{c}$ 是编码后的向量 $\boldsymbol{m}$ 的一个 k 长子向量，则称 $\mathcal{C}$ 是一个系统码。

给定一个 $[n, k, d]_q$ 线性码 $\mathcal{C}$，Singleton 界给出了其最小汉明距离的上界，即：$d \leqslant n-k+1$。当 $d = n-k+1$ 时，称 $\mathcal{C}$ 是一个*最大距离可分码*（Maximum Distance Separable code（MDS 码））。由定义可以看出，MDS 码在给定数据容量下，其码字间的距离达到了最大，故其容错能力也相应达到最大。除此以外，MDS 码的生成矩阵 $\boldsymbol{G}$ 具有任意 k 列线性无关的性质，这在实际存储场景中，具有很大优

势。具体是指，在某些存储节点发生故障导致存储在上面的数据无法获取时，只要发生故障节点数量不超过 $n-k$ 个，就可以通过任意 k 个幸存节点中的数据将这些故障节点处的数据恢复出来。而（Generalized）Reed-Solomon 码（（G）RS 码）作为一种特殊的 MDS 码，因其结构简单、易于实现，且具有高效的解码算法等特性，得到了广泛应用。

在衡量一个线性码作为一种存储策略时，其性能的优劣，除去上述所说的存储开销、编解码复杂度，纠错能力之外，其编码速度也是一个值得考虑的指标。特别是当有大量数据需要被存储时，编码速度的高低对其影响就显得尤为重要。在某些特殊存储场景下，例如数据存档服务，为了提高编码速度，有时可以牺牲其它性能来优化编码过程。而稀疏平衡的 MDS 码，能很好地权衡这些性能。以其编码过程为例。给定一个稀疏平衡的 $[n,k,d]_q$ MDS 码 C，其生成矩阵为 G。C 的“稀疏性”指的是 G 的每一行具有最少的非零元素，即 $n-k+1$ 个，而“平衡性”表示任意两列中非零元素的个数最多相差一个，即 $k-\left\lceil \frac{k(k-1)}{n} \right\rceil$ 或 $k-\left\lfloor \frac{k(k-1)}{n} \right\rfloor$ 个。当系统中出现数据更新需求时，以式 (1.1) 为例，假设需要更新数据 c_1，由于系统存储的是编码后的数据 $(m_1,\cdots,m_n)$，需要根据 c_1 的变动更新后续存储的数据。在进行编码时，c_1 只与 G 的第一行元素进行运算，故第一行向量中的非零元个数即为需要更新的数据块个数。当 C 是稀疏码时，G 的每行向量重量都达到最小，即为 $n-k+1$，此时只需更新对应行非零元素所在位置上的 $n-k+1$ 个 m_i 即可。因此，C 的稀疏性保证了在更新单个数据时，影响的存储节点最少。另一方面，从式 (1.1) 可以看出，对任意 $i\in[n]$，有 $m_i=cg_i$。因此，在编码时，生成第 i 个数据块的时间与 g_i 的重量成正比。当 C 是一个平衡码时，由于 G 中每个列向量重量几乎相同，这保证了生成每个数据块时间大致相同。因此，C 的平衡性确保了整个存储系统的计算负载平衡，即没有存储节点成为瓶颈，降低了网络拥堵的概率，使得因此而丢失数据的可能性降低。

本文研究的基于小域 $\mathbb{F}_q$ 上的稀疏平衡的 MDS 码，由于 q 足够小，且提供了有效的算法去构造相应码字，使得在实际应用中得以实现，且开销更小。

1.1.2　DNA 存储系统下的编码以及串联复制纠错码

在经典的存储系统中，数据存储在一些固态介质中，例如磁盘，SSD，闪存等。这些介质由于自身硬件条件所限制，数据密度低且需要消耗电能，且随着使用时长的增加，磨损加剧，从而导致数据损坏或丢失，使用寿命较短。最近，采用 DNA 分子作为存储介质的存储系统在文献 [2] 中被提出。微软宣布，开始进行在 DNA 中存储约 100 兆字节数据的大规模实验，美国情报高级研究计划局（IARPA）也表达了对这一领域的兴趣 [3]。

与传统介质相比，以活体 DNA 分子作为存储介质能够带来许多优点，如

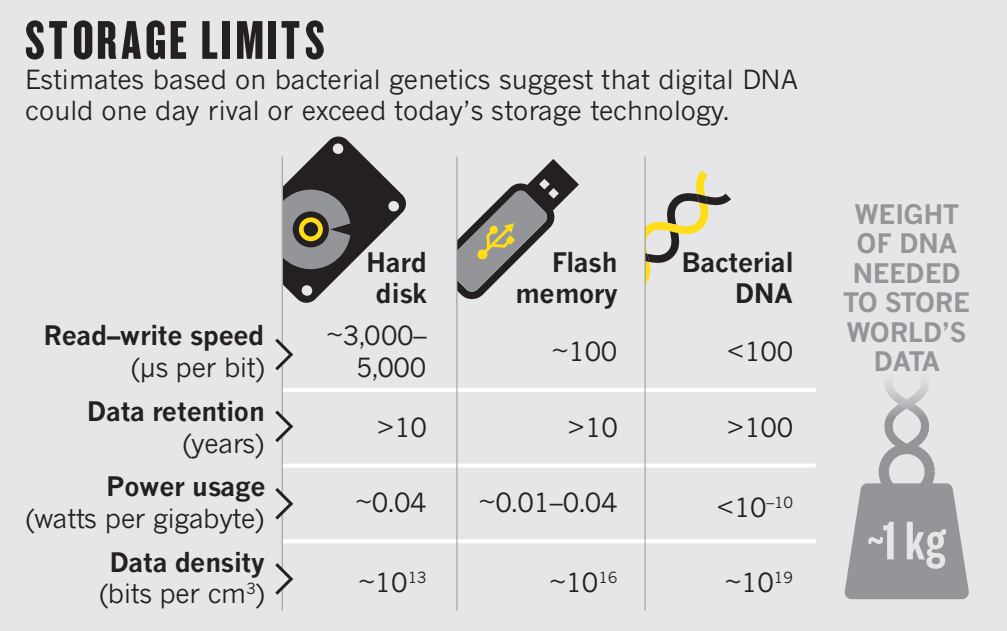


图 1.3　传统存储介质与 DNA 存储介质优缺点 [3]

图1.3所示。可以看出，DNA 分子数据密度高，存活时间久且几乎无电能消耗。近年来，伴随着 DNA 合成与测序成本的降低，使得以 DNA 为存储介质的系统，其读写速度也有所提高，从而受到了人们广泛关注。

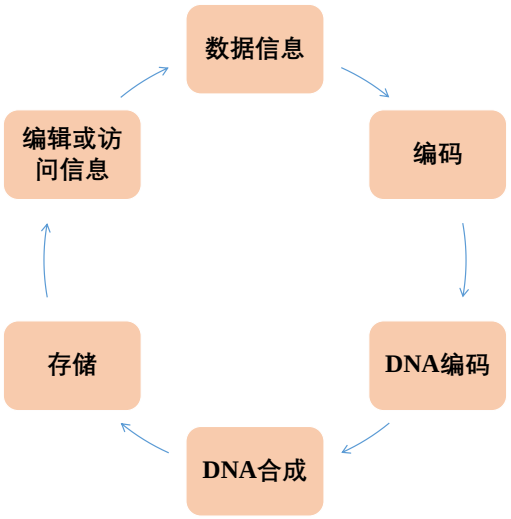


图 1.4　基于 DNA 存储系统原型框图 [4]

图1.4给出了基于 DNA 存储系统的基本流程。在 DNA 存储系统中，首先对经典源信息（数据信息）进行编码，将其转化为 ASCII 或者某些专门的字格式，这期间可能会进行压缩，使得其可以在四元字母集上被表示成字符串。这四元字母集可以对应于构成 DNA 分子的 A、T、C、G 四种脱氧核苷酸。其次，将所得字符串按照标准 DNA 规则并以一定策略添加冗余进行编码，合成相应的 DNA 码字。这里添加冗余的方式可以使用一些纠错码来进行，主要是为了克服 DNA 分子在合成复制过程中所发生的突变，例如串联复制 (tandem duplication) 错误、

点突变、插入、删除等。再将所得 DNA 分子复制扩充，并存储起来。而编辑模块指的是在存储的 DNA 字符串中创建突变的过程（通过删除一个或多个子字符串，并可能插入其它字符串），对于访问信息，实际上是对 DNA 分子进行测序从而确定存储内容的过程。

为了保证以活体 DNA 为存储介质的存储系统可靠性，考虑能纠正上述突变错误的纠错码尤为重要。与纠删码应用场景不同，这里并不确定发生错误的位置，故考虑使用纠错码。在经典的存储系统中，由于存储介质的损坏往往导致擦除错误，即明确错误的位置，因此，在上一小节中讨论的是纠删码。本工作比较关心的是串联复制错误，这种突变指的是将一个 DNA 片段复制至少一次并插入到原始位置的过程。例如，对于一个序列 $AGCTCT$，$CTCT$ 是 CT 上长度为 2 的 2-串联复制错误，也称字符串 $AGCTCT$ 可由 $AGCT$ 发生 2 次 2-串联复制错误得到。串联复制错误约占人类基因组的 3% [5]，并可能导致一些疾病的产生，例如染色体脆性（chromosome fragility）、扩张疾病（expansion diseases）、基因沉寂（silencing genes）[6] 和快速形态变异（rapid morphological variation）[7] 等。因此，考虑能够纠正这种串联复制错误的纠错码，对基于 DNA 存储的系统而言是比较关键的。

定义 1.2　给定字母集 Σ，若存在一个集合 $\mathcal{C} \subset \Sigma^n$ 且满足 $|\mathcal{C}| = M$，使得对任意两个不同的向量 $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ 有，

$$D_k^t(\boldsymbol{x}) \cap D_k^t(\boldsymbol{y}) = \varnothing,$$

则称码 $\mathcal{C}$ 是一个可纠正 t 次 k-串联复制错误的纠错码，记为 $(n, M; t)_k$ 码。这里 $D_k^t(\boldsymbol{x}) = \{\boldsymbol{y} : \boldsymbol{y}$ 可由 $\boldsymbol{x}$ 发生 t次 k-串联复制错误得到$\}$，t 为任意非负整数，或不限定复制次数。

2017 年，Jain 等人 [8] 提出了一个基于 ℓ_1 度量下非负整数上的常重码的编码方案用以克服串联复制错误。更具体地，他们刻画了一个等价条件：ℓ_1 度量下的非负整数上的最优常重码可以用来构造最优的串联复制纠错码。这使得我们去研究对应的常重码的性质和构造。另外，在 DNA 分子发生串联复制时，如果限制复制次数 $t \leqslant q-1$ 的话，那么考虑 ℓ_1 度量下 $I_q = \{0, 1, \cdots, q-1\}$ 上常重码的构造也是有意义的，这是因为在 I_q 上，码字的每个分量最多为 $q-1$，恰好与 t 一一对应。特别地，当 q 充分大时，这类码字的性质可以作为研究非负整数上的码的参考。基于此，本文第三部分致力于研究正整数和 I_q 上常重码的性质以及相应的构造。

1.1.3 扭 Reed-Solomon 码

扭（Twisted）Reed-Solomon（TRS）码是由 Beelen，Puchinger 和 Rosenkilde né Nielsen[9] 在 2017 年首次提出的，它是 RS 码的一种推广，主要是受到扭 Gabidulin 码 [10] 的启发。RS 码和 TRS 码都是多项式码，即其中的每个码字都可以看成一个多项式。更具体地说，给定有限域 $\mathbb{F}_q$ 上的一个 $[n,k]$ RS 码 $\mathcal{C}_1$，其定义如下，

$$\mathcal{C}_1 = \{(f(a_1), \cdots, (f(a_n)) : f(x) \in \mathbb{F}_q[x], \deg(f(x)) < k\},$$

其中 $a_1, \cdots, a_n$ 是 $\mathbb{F}_q$ 上 n 个互不相同的元素，称其为 $\mathcal{C}_1$ 的*估值点*（evaluation point）。也就是说 $\mathcal{C}_1$ 可以看成是 $\mathbb{F}_q$ 上所有度不超过 $k-1$ 的多项式集合。而一个 $[n,k]$ TRS 码 $\mathcal{C}_2$ 可以看作是 $\mathcal{C}_1$ 中多项式的延伸，与 RS 码不同的是，它的码字中允许出现次数大于 $k-1$ 的多项式。更具体地，给定 $\mathcal{C}_1$ 中一个多项式 $f_1(x) = b_0 + b_1 x + \cdots + b_{k-1}x^{k-1}$，可唯一对应一个 $\mathcal{C}_2$ 中的一个多项式

$$f_2(x) = b_0 + b_1 x + \cdots + b_{k-1}x^{k-1} + \eta b_h x^{k-1+t},$$

这里 h 和 t 为两个给定的整数，满足 $0 \leqslant h \leqslant k-1, 1 \leqslant t \leqslant n-k$，$\eta \in \mathbb{F}_q^*$。注意到，$b_h$ 是多项式 $f_1(x)$ 中的第 h 次项系数，当 $b_h = 0$ 时，$f_1(x) = f_2(x)$，即 $\mathcal{C}_1$ 和 $\mathcal{C}_2$ 中的码字存在重叠部分。特别地，我们称 h 为*钩*（*hook*），t 为*扭结*（*twist*）。事实上，在上述多项式的基础之上，可以进一步延拓，通过添加 $\ell \geqslant 1$ 组 (h_i, t_i, η_i)，其中 $i \in [\ell]$ 且 $0 \leqslant h_i \leqslant k-1, 1 \leqslant t_i \leqslant n-k, \eta_i \in \mathbb{F}_q^*$，则称这种添加多个钩以及扭结的码为广义 TRS（TGRS）码 [11]。

RS 码由于其良好的代数性质，可以被用来构造基于编码理论的密码方案。第一个基于广义 RS 码的密码方案是由 Niederreiter [12] 于 1986 年提出的，但被文献 [13] 中提出的攻击方案恢复了底层的 RS 码，由此证明了此密码方案是不安全的。在过去的几年里，有很多学者对这个方案进行了改良 [14-15]，使得其可以避免文献 [16] 中的针对 RS 码的攻击。但随之，针对这几种改良方案的攻击也陆续产生 [17]。而 TGRS 码作为 RS 码的一种推广，它既具有一些与 RS 码类似的好的性质，例如解码算法快、结构清晰（在给定条件下，对偶码也可确定）等，又具有一些 RS 码不具备的特性，例如可通过控制 t, ℓ 来达到某些特定的结构等。利用这一特点，最近在文献 [11] 中，TGRS 码被提议作为 Goppa 码的替代方案，应用于 McEliece 密码系统 [18]，从而可能减少密钥大小。McEliece 密码系统是一种公钥密码系统，应用于其中的码需要配备一个高效的解码算法。TGRS 码恰好能够满足这一点，在具有高效解码算法的同时，又能够抵抗针对 RS 类型码的攻击。因此，本文第四章旨在研究 TGRS 码的相关构造，并利用其构造一些编码理论中比较重要的自对偶码。

1.2　国内外研究现状

近年来，为保障数据安全，关于将编码技术应用于数据存储和传输过程，有很多研究工作。我们主要就稀疏平衡的 MDS 码、ℓ_1 度量下的最优常重码以及扭 RS 码三个方面的工作来作归纳与总结。

（1）稀疏平衡的 MDS 码

稀疏平衡的 MDS 码，除去能在分布式存储系统编码过程保证计算负载均衡以及更新数据时影响最少存储节点之外，在弱安全协作数据交换 [19-21]、多址网络 [22-23]、无线传感器网络 [24] 等方面也有着很强的应用。因此受到了广泛的关注。如何在较小的有限域上构造稀疏平衡的 MDS 码，是大家比较关心的问题。这个问题最初是在 [24] 中考虑的，他们通过概率方法证明了在任何大小为 $q > \binom{n-1}{k-1}$ 的有限域上，总是存在一个稀疏平衡的 MDS 码。在文献 [25-26] 中，对于任何素数幂 $q = n+1$ 和任何正整数 k 满足 $1 \leqslant k \leqslant n$，他们构造了一个稀疏平衡的循环 $[n,k]_q$ RS 码。但是，他们的方法仅适用于 $q = n+1$，而 $q \geqslant n-1$ 和 $q \neq n+1$ 的情况尚未解决。Song 和 Cai [27] 进一步扩展了这一结果，给出了几个算法，利用类似于 Schwartz-Zippel 定理的方法，证明了对于任何正整数 n 和 k，以及 $q \geqslant n + \left\lceil \frac{k(k-1)}{n} \right\rceil$，满足 $1 \leqslant k \leqslant n$，都存在 $[n,k]_q$ 广义 RS 码，且该码是稀疏平衡的。但是 $k \geqslant 3$ 时，码长和域的大小 q 还是有差距的。故在 $q < n + \left\lceil \frac{k(k-1)}{n} \right\rceil$ 的较小域上，稀疏平衡的 $[n,k]_q$MDS 码的存在性值得关注。

（2）ℓ_1 度量下的最优常重码

常重码是编码理论中的经典研究对象，其任一码字，在给定度量下，都具有相同的重量。特别是对于汉明度量下的常重码，其在有效带宽信道（bandwidth-efficient channels）上的编码 [28] 和 DNA 计算中寡核苷酸序列（oligonucleotide sequence）的设计 [29-30] 等领域具有广泛应用，从而受到了学者们的大量关注和研究。而常重码研究的核心问题之一，是根据它们与组合设计理论的密切关系，从而确定其码字大小以及相应的最优码，参见文献 [31-40]。尽管在编码理论中，学者们已经考虑了几种不同度量下的常重码，但据我们所知，除了汉明距离之外，其他度量下的结果并不多。

而关于 ℓ_1 度量下的码，除去能够用来构造串联复制纠错码之外，它在闪存的秩调制方案（rank modulation scheme）等领域 [41-46] 也有着广泛的应用。然而，大多数工作都集中在置换码或多重置换码上。2018 年，Kovačević 和 Tan [47] 研究了多重集码（multiset codes），它们适用于传输多组信息符号并可能伴随有插入和删除等错误的信道。事实上，他们研究的多重集码本质上是一个 ℓ_1 度量下基于非负整数上的常重码，其重量为 $w = \Theta(n)$。借助 Sidon 集和格理论（lattice）

等工具，他们给出了在给定最小距离的情况下，码长要么不变，要么随重量成比例变化时一系列关于最大码字个数的渐近结果。特别是对于小距离或者小码长时，这些界是渐近最优的。但是，他们的方法主要应用于码长和重量可比的情况，若权重是固定值，结论还有较大的改进空间。Jinushi 和 Sakaniwa [48] 提出了一种基于广义 Hadamard 矩阵 [49] 性质的 ℓ_1 度量下纠错码的构造。他们使用的术语 绝对值和距离（absolute summation distance），我们认为其本质是 ℓ_1 距离。

综上，关于 ℓ_1 度量下码的研究相对较少，因此，具有广阔的开拓空间，特别是构造最优码问题，值得我们进一步去探索。

（3）扭 Reed-Solomon 码以及自对偶码

扭（Twisted）Reed-Solomon（TRS）码是由 Beelen，Puchinger 和 Rosenkilde né Nielsen[9] 在 2017 年首次提出的，并提供了相应的丰富的构造。当 $t = 1$ 时，该码字的最小距离极其接近于 Singleton 界，因此，他们给出了 TRS 码达到 Singleton 界，即是 MDS 码的充分必要条件，利用这一条件，他们证明了由 TRS 码构成的集合中，存在大量不同构于 (G)RS 码的码。进一步地，在文献 [11] 中，他们将 TRS 码推广到了 TGRS 码，并给出了相应 TGRS 码对偶封闭的条件。结合这两点，陆续有学者利用其构造自对偶的 MDS 码和线性互补对偶（LCD）码。且容易知道，由这种方法构造出来的码，可以得到一些不与 RS 码同构的码类。更具体地，在 [50] 中，作者给出了一些 $h = k - 1$ 和 $t = 1$ 时的 TGRS 码，然后借助它们构造了几类自对偶 MDS 或自对偶 NMDS 码。在 [51] 中，他们给出了 $(h, t) = (0, 1)$ 以及 $(h, t) = (k - 1, 1)$ 时的 TGRS 码，并用这些码来构造 LCD 码。同时，在 [52] 中，作者利用不同的代数方法，也给出了一些 $(h, t) = (0, 1)$ 时的 TGRS 码，并进一步用这些码构造了 MDS 或 NMDS LCD 码。这两个 LCD 码都是基于欧几里得内积的。在 [53] 中，作者基于 Beelen 等人 [9] 给出的 TGRS 码，构造了不等价于 RS 码的欧几里得 LCD MDS 码和 Hermitian LCD MDS 码。注意到，在上述构造中，使用的都是添加一个扭结的 TGRS 码，即 $\ell = 1$。而关于多个扭结的 TGRS 码，相关的结果还不多，有待进一步研究。

1.3　本文的主要研究内容和贡献

本文主要从数据存储和传输过程入手，将编码技术应用其中，以保障数据的安全性。主要研究了以下三种纠删（错）码。

1.3.1　基于小域上的稀疏平衡的 MDS 码

在第1.2节 **(1)** 中，我们知道，对于任何正整数 n 和 k，以及 $q \geqslant n + \left\lceil \frac{k(k-1)}{n} \right\rceil$，满足 $1 \leqslant k \leqslant n$，都存在稀疏平衡的 $[n, k]_q$MDS 码。这与 MDS 猜想中的 q 与 n 的

约束还有很大改进空间。因此，对于稀疏平衡的 MDS 码，我们把对有限域 q 的限制改进到了 $q \geqslant n-1$。更具体地，我们在一个大小满足 $q \geqslant n-1$ 的有限域 $\mathbb{F}_q$ 上，且当 $3 \leqslant k \leqslant n \leqslant 2k$ 时，构造了一个稀疏平衡的 $[n,k]_q$ MDS 码。事实上，构造这样的码，等价于寻找它的生成矩阵 G，即在 $\mathbb{F}_q$ 上找到一个具有稀疏平衡零模式的 $k \times n$ 阶矩阵 G，满足 G 的所有 k 阶子式非零。这里矩阵 G 的零模式定义为一个集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$，其中 $S_i = \{j \leqslant n : G_{i,j} = 0\} \subset \{1, 2, \cdots, n\}$。我们首先给出一个由集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 刻画的稀疏生成矩阵 G 存在的充分条件，参见定理 2.4，它推广了 [54] 中的定理 II.5。然后我们证明了只有当 $n \leqslant 2k$ 时，满足定理 2.4 充分条件的集族 $\mathcal{S}$ 才是平衡的。最后，我们通过设计几个算法证明了只要 $n \leqslant 2k$，对应于 $\mathcal{S}$ 的二元矩阵就存在，即 $\mathcal{S}$ 和 G 存在。主要结果如下。

定理 1.1　对任何正整数 $k \geqslant 3$，若 k 是偶数，则令 $n \leqslant 2k$，若 k 是奇数，则令 $n \leqslant 2k-1$。那么在任何满足 $q \geqslant n-1$ 的有限域 $\mathbb{F}_q$ 上，都存在一个具有稀疏平衡生成矩阵的 $[n,k]_q$ MDS 码。

事实上，我们首先是在 $q \geqslant n$ 的情况下证明了定理 1.1，这是通过设计几个复杂度关于 k 和 n 是多项式时间的算法来完成的[①]。这些算法均输出一个稀疏且平衡的二元矩阵，它与满足定理 2.4 的集族 $\mathcal{S}$ 一一对应。然后我们再将这种构造方法扩展到 $q = n-1$。

为了克服 $n \leqslant 2k$ 的限制，我们利用和集构造了当码长 $n = q = p^s$ 时的 MDS 码。具体来说，我们需要找到 $\mathbb{Z}_p^s$ 的 k-子集 A 和 $(k-1)$-子集 B，使得它们的和集 $A + B = \{a + b : a \in A, b \in B\}$ 满足，$\mathbb{Z}_p^s$ 的每个元素在其中出现的次数几乎相同，以确保码的稀疏性和平衡性。我们将结果总结为以下定理。

定理 1.2　对于任何正整数满足 $n = q = p^s$, $k = p^e m$ 与 $1 \leqslant m \leqslant p-1$ 和 $0 \leqslant e \leqslant s-1$ 除了 $e = \frac{s-1}{2}$，在这种情况下 $m = 1$ 并且 s 必须是奇数，存在稀疏平衡的 $[n,k]_q$ MDS 码。

注意到，在定理 1.2 中，当 $e \leqslant s-2$ 或 $e = s-1$ 且 $m < \frac{p}{2}$ 时，有 $n > 2k$，此时上述结果不包含在定理 1.1 中。

1.3.2　基于 ℓ_1 度量下的最优常重码

在 DNA 存储系统下，为了能够纠正 DNA 分子复制过程中发生的串联复制错误，我们研究了能够用于构造串联复制纠错码的常重码，即 ℓ_1 度量下非负整数以及 $I_3 = \{0, 1, 2\}$ 上的最优常重码。

更具体地，我们通过使用填充集族（packing）和可分组设计（group divisible design（GDD））来构造 ℓ_1 度量下，固定重量 w 和距离 d 的常重码，并确定了所有重量 $w \leqslant 4$ 下，任意距离 d 和码长 n 的最大码字个数。注意到，当 $w = 1$ 或 2

[①]开源代码链接见：https://github.com/ttchenday/Sparse_and_Balanced_MDS_Codes_over_Small_Fields。

时，相应的最优码是平凡的。我们的主要贡献如下。

1. 考虑重量 $w \in \{3,4\}$ 的非负整数上的常重码，在给定任意最小距离 d 和码长 n 时，我们完全确定了其最大码字个数 $A(n,d,w)$，并给出相应的最优码构造。主要用到的工具是区组大小为 3 的最优 2-填充 [55]，和区组大小为 4 的最优 3-填充。特别地，相比于文献 [47] 中给出的特殊参数下 $A(n,d,w)$ 的上下界：$\frac{n^2}{6} \lesssim A(n,4,3) \lesssim \frac{n^3}{6}$，$\frac{n^3}{24} \lesssim A(n,4,4) \lesssim \frac{n^4}{24}$，以及 $A(n,6,4) \lesssim \frac{n^4}{24}$[①]，我们确定了它们具体阶数：$A(n,4,3) \sim \frac{n^2}{6}, A(n,4,4) \sim \frac{n^3}{24}$ 以及 $A(n,6,4) \sim \frac{n^2}{12}$。
2. 考虑集合 $\{0,1,2\}$ 上的常重码，即三元常重码，通过 Steiner 三元系 [56] 以及带有特殊剩余图的填充，我们构造出了所有重量 $w = 3$ 的最优常重码。而当 $w = 4$ 时，利用 GDD，除去少数小码长的码外，我们给出了其余所有参数最优码构造，而针对这些小码长的码，我们提供了相应的上下界。
3. 对于 $d = 2w - 2$ 的三元常重码，我们使用 Alon 等人在文献 [57] 中给出的图填充结果，得到了一个一般构造，且当码长 n 充分大并满足 $n \equiv 1, w, -w+2, -2w+3 \pmod{w(w-1)}$ 时，确定了最大码字个数。

此外，对于 ℓ_1 度量下一般的 q 元常重码，我们用其码字支集刻画出了一个通用必要条件（称其为 UNC 条件），它展现了 ℓ_1 度量下的码与填充集族之间的联系。这种联系为构造 ℓ_1 度量下的固定重量 w 的最优常重码提供了一些启示。

1.3.3　扭 Reed-Solomon 码和自对偶码

针对（广义）扭 RS 码问题，我们研究了（广义）扭 RS 码 $\boldsymbol{C}^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 的性质以及相关构造问题。在 [11] 中，Beelen 等人证明了，如果 TGRS 码的估值点集合是一个 $\mathbb{F}_q^*$ 的乘法子群，则 TGRS 码在对偶下是封闭的。我们的结果放宽了这个条件，并推广了这个结论，只需要其所有估值点构成某个多项式的根集合即可。利用该多项式系数分布情况，我们刻画出了码 $\boldsymbol{C}^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 对偶封闭的充分条件，并给出了相应的校验矩阵。基于这一结果，我们构造了相应的自对偶码。特别是当 $\ell = 1$ 时，所得自对偶码是 MDS 或近 MDS 码。而当 $\ell = 3$ 时，所得自对偶码的最小距离在 $n-k-2$ 和 $n-k+1$ 之间。

1.4　本文的组织结构

本文的组织结构以及各章节之间的联系如图1.5所示。

第一章绪论。介绍了本文的研究背景、研究现状以及研究内容，分析了本文的创新点。

第二章基于小域上的稀疏平衡的 MDS 码。本章介绍了稀疏平衡的 MDS 码

[①] $f(n) \sim g(n)$ 指的是 $\lim_{n\to\infty} f(n)/g(n) = 1$，而 $f(n) \lesssim g(n)$ 指的是 $\liminf_{n\to\infty} g(n)/f(n) \geqslant 1$。

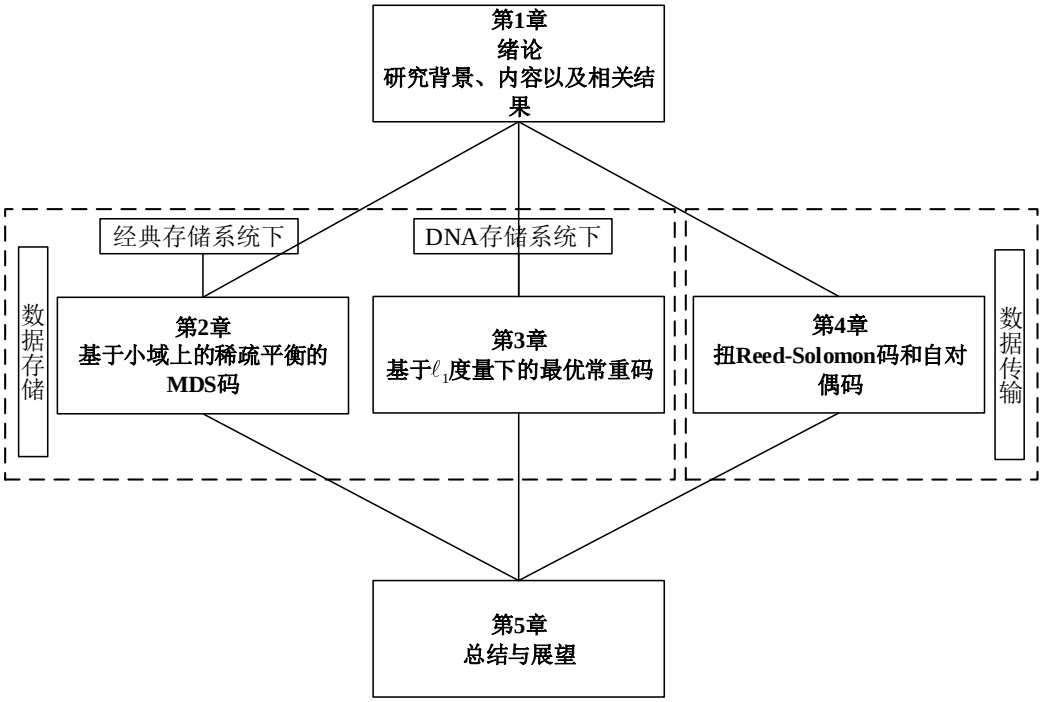


图 1.5　本学位论文组织结构

在数据存储中的一些应用，给出了其在任意小域 $\mathbb{F}_q$（$q \geqslant n-1$，这里 n 为码长）上的构造。

第三章基于 ℓ_1 度量下的最优常重码。本章利用了组合设计里面的填充集族，当重量 $w \leqslant 4$ 时，分别给出了非负整数上以及 $I_3 = \{0,1,2\}$ 上所有参数下的最优常重码的构造；而针对一般的重量 w 和距离 $2w-2$，利用图分解方法，给出了相应最优码的渐近性结果。

第四章扭 Reed-Solomon 码和自对偶码。本章给出了部分对偶封闭的扭 Reed-Solomon 码的构造，并利用该码构造出了编码理论中比较重要的自对偶码。

第五章总结与展望。本章总结了全文的主要研究工作和成果，并提出了可能改进的方向和进一步的研究工作。

第 2 章　基于小域上的稀疏平衡的 MDS 码

本章旨在满足 $q \geqslant n-1$ 的有限域上，构造稀疏平衡的 $[n,k]_q$ MDS 码。其出发点是为了提高存储系统可靠性，更具体地，主要是为了避免由于负载不均衡而导致的网络拥堵现象，以防止数据丢失。在第2.1节中，我们简单的介绍了稀疏平衡码的研究背景和进展，以及本章贡献。在第2.2节中，我们首先给出必要的记号和定义，然后在第2.3节中给出了 $q \geqslant n$ 时，根据生成矩阵零模式刻画的稀疏的 $[n,k]_q$ MDS 码存在的充分条件。当 $n \leqslant 2k$ 时，满足充分条件的平衡零模式的构造细节在第2.4节中给出，主要用到的方法是基于矩阵上的元素置换操作。在第2.5节中，当 $q = n = p^s$ 且 k 的形式为 $p^e m$ 时，我们构造了稀疏平衡的 $[n,k]_q$ MDS 码，这一构造提供了部分 $n > 2k$ 时的稀疏平衡码。最后在第2.6节中对本章进行了简单的总结。

2.1　介绍

稀疏平衡的 MDS 码问题，其本质上是寻找一个具有约束生成矩阵的 MDS 码问题，这里的约束指的是生成矩阵要求是稀疏平衡的。这个问题最初是在 [24] 中考虑的，其出发点是在编码过程中可以使得计算负载平衡和更新数据时，影响最少存储节点 [25, 58]。在分布式存储系统中，考虑稀疏平衡的 MDS 码策略时，一方面，由于计算每个码字符号所需时间与生成矩阵对应列中非零元素数量成正比，根据生成矩阵的平衡性可知，每个码字符号的计算时间大致相同。因此可以确保计算负载的均衡。另一方面，其码字的稀疏性，可以保证更新单个码字符号只会影响存储系统中的 $n-k+1$ 个存储节点，即所需更新的数据量最小。

事实上，具有约束生成矩阵的 MDS 码问题，指的是，确定 MDS 码是否存在一个具有给定零模式的生成矩阵问题。这种问题已经分别在文献 [54, 59-60] 中针对具有汉明度量的 MDS 码和文献 [61-62] 中针对具有秩度量（Gabidulin 码）的 MDS 码进行了研究。令 G 为一个 $[n,k]_q$ MDS 码的 $k \times n$ 阶生成矩阵，其中 $k \leqslant n$，它的零模式为集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$。而利用这个集族刻画出的一个条件被称为 *MDS* 条件：对于任何非空子集 $I \subseteq \{1, 2, \cdots, k\}$，有 $|I| + |\cap_{i \in I} S_i| \leqslant k$。Dau 等人在文献 [59] 中猜想：当 $q \geqslant n+k-1$ 时，如果 MDS 条件成立，那么具有给定零模式生成矩阵的 MDS 码是存在的。这个猜想被称为 *GM-MDS* 猜想，并吸引了很多学者的关注，参见文献 [63-65]。最近 GM-MDS 猜想分别被 Lovett [66] 和 Yildiz 以及 Hassibi [60] 独立证明是正确的，我们将其重述如下。

定理 2.1 ([59-60, 66]$^{\text{GM-MDS 定理}}$)　设 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 是一个集族，其中

$S_i \subseteq \{1,2,\cdots,n\}$, $1 \leqslant i \leqslant k$。那么当 $q \geqslant n+k-1$ 时，存在一个 $\mathbb{F}_q$ 上的 $[n,k]_q$ MDS 码，使得其生成矩阵 G 满足每当 $j \in S_i$ 时有 $G_{i,j}=0$ 成立，当且仅当 $\mathcal{S}$ 满足 MDS 条件。

在文献 [54] 中，Greaves 和 Syatriadi 进一步考虑了具有约束生成矩阵的 MDS 码存在性问题，其生成矩阵的支集约束比 MDS 条件稍强，但有限域大小可以放松到 $q \geqslant n$ 或 $q \geqslant n+1$。但是，他们的结果不能用来构造稀疏平衡的 MDS 码。更具体地，他们给出了一些特殊类型的 $[n,k]_q$ RS 码的两种构造，其生成矩阵具有特定支集约束。其中之一是在任何满足 $q \geqslant n$ 的有限域 $\mathbb{F}_q$ 上，如果生成矩阵的零模式 $\mathcal{S}=\{S_1,\cdots,S_k\}$ 满足 MDS 条件，且对任意 $1 \leqslant i \leqslant k$，满足 $|\cap_{j=1}^{i} S_j| = k-i$，那么存在一个 $[n,k]_q$ RS 码。注意到，当 $i=k-1$ 时，有 $|\cap_{j=1}^{k-1} S_j| = 1$，这意味着生成矩阵 G 中至少有一列包含 $k-1$ 个 0。所以只有当 $\left\lceil \frac{k(k-1)}{n} \right\rceil \geqslant k-1$，即 $n=k$ 或 $k+1$ 时，G 才有可能平衡。他们给出的第二个构造是在任何满足 $q \geqslant n+1$ 的有限域 $\mathbb{F}_q$ 上，如果对于所有 $i=1,\cdots,k$，有 $|S_i| \leqslant i-1$，那么存在一个 $[n,k]_q$ RS 码。注意当 $i \leqslant k-1$ 时，有 $|S_i| \leqslant k-2$，由此构造得到的生成矩阵 G 不是稀疏的。因此，综上所述，我们需要重新刻画集族 $\mathcal{S}=\{S_1,\cdots,S_k\}$，使得其可以生成一个稀疏平衡的 MDS 码。

受到上述问题的启发，本章主要研究在比较小的有限域上构造稀疏平衡的 MDS 码问题。首先对于有限域 $\mathbb{F}_q$ 满足 $q \geqslant n$ 时，给出一个由生成矩阵零模式刻画的稀疏的 $[n,k]_q$ MDS 码存在的充分条件。这一充分条件，推广了 Greaves 和 Syatriadi 的结果，将构造 MDS 码这一代数问题转化为了一个构造特定集族 $\mathcal{S}=\{S_1,\cdots,S_k\}$ 的组合问题。基于这个条件，我们通过设计几个多项式时间算法，找到了满足要求的集族对应的二元矩阵，从而构造出了码长满足 $n \leqslant 2k$ 的所有稀疏平衡的 $[n,k]_q$ MDS 码。进一步地，通过扩展坐标，我们将域的大小改进到 $q \geqslant n-1$。而当码长 $n > 2k$ 时，对任意整数 e,s,m，满足 $e \leqslant s-2$ 且 $m \leqslant p-1$，或者 $e=s-1$ 且 $m < \frac{p}{2}$，我们利用平衡和集 $A+B$，其中 $|A|=k$ 以及 $|B|=k-1$，构造了所有稀疏平衡的 $[n=q=p^s, k=p^e m]_q$ MDS 码。

2.2　预备知识

本节中，我们将介绍一些必要的记号、定义，并给出二元矩阵和码生成矩阵的一些联系。

对于任意两个整数 $a<b$，令 $[a,b]$ 表示整数集合 $\{a,a+1,\cdots,b\}$。我们进一步将 $[1,b]$ 缩写为 $[b]$。令 $a \bmod^+ n$ 表示唯一的正整数 $r \in [n]$，使得 n 整除 $a-r$，而 $[a,b] \bmod^+ n$ 表示集合 $\{x \bmod^+ n : x \in [a,b]\}$。

对于任何素数幂 q，我们用 $\mathbb{F}_q$ 来表示具有 q 个元素的有限域。我们记 $[n,k,d]_q$

为 $\mathbb{F}_q$ 上长度为 n、维数为 k 以及最小汉明距离为 d 的线性码 $\mathcal{C}$。当 $\mathcal{C}$ 是 MDS 码时，即 $d = n - k + 1$，有时会省略 d 而写成 $[n, k]_q$。

2.2.1　稀疏平衡的 MDS 码

码 $\mathcal{C}$ 的一个生成矩阵 G 若满足下面两个条件：

(1) 稀疏条件：重量，即 G 的每一行的非零元素恰好有 $n - k + 1$ 个；

(2) 平衡条件：G 的每一列的重量要么是 $\left\lceil \frac{k(n-k+1)}{n} \right\rceil$，要么是 $\left\lfloor \frac{k(n-k+1)}{n} \right\rfloor$。

则称矩阵 G 是稀疏平衡的。具有稀疏平衡的生成矩阵的 MDS 码，则称为稀疏平衡的 MDS 码。在本章中，我们关注稀疏平衡的 RS 码的构造。回顾一下，一个 $[n, k]_q$ RS 码是一种特殊的 MDS 码，定义为 $\{(f(a_1), \cdots, (f(a_n)) : f(x) \in \mathbb{F}_q[x], \deg(f) < k\}$，其中 $a_1, \cdots, a_n$ 是 $\mathbb{F}_q$ 上 n 个互不相同的元素，称其为估值点。

令 $\mathcal{P}$ 为 $\mathbb{F}_q[x]$ 中 k 个度不超过 $t - 1$ 的多项式 $f_1(x), f_2(x), \cdots, f_k(x)$ 的集合，则 $\mathcal{P}$ 的系数矩阵，记为 $C(\mathcal{P})$ 或 $C(f_1, f_2, \cdots, f_k)$，是一个 $k \times t$ 阶矩阵，其中第 (i, j) 个位置上的元素为 f_i 中 x^{t-j} 项的系数，即 $[x^{t-j}]f_i$。如果 $\mathcal{P}$ 只包含一个多项式 f，我们用 $C(f)$ 表示为记录 f 所有系数的行向量。

给定一个集族 $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$，其中对于每个正整数 $i \in [k]$，有 $S_i \subset [n]$ 且 $|S_i| \leqslant k - 1$。令 $a_1, \cdots, a_n$ 是 $\mathbb{F}_q$ 上任意固定的 n 个不同元素。对于所有 $i \in [k]$，定义 $P_{S_i}(x) \triangleq \prod_{j \in S_i}(x - a_j) = p_{i,0}x^{k-1} + p_{i,1}x^{k-2} + \cdots + p_{i,k-1} \in \mathbb{F}_q[x]$。在本章的其余部分，我们统一用 $\mathcal{P}$ 表示多项式序列 $P_{S_1}, P_{S_2}, \cdots, P_{S_k}$，则对任意 $i \in [k]$ 和 $j \in [0, k-1]$，有 $C(\mathcal{P}) = (p_{i,j})_{k \times k}$。令 $G = (g_{i,j})$ 是 $\mathbb{F}_q$ 上的 $k \times n$ 阶矩阵，其中对于 $i \in [k]$ 和 $j \in [n]$，$g_{i,j} = P_{S_i}(a_j)$。则有

$$
\begin{aligned}
G &= \begin{pmatrix} p_{1,0} & p_{1,1} & \cdots & p_{1,k-1} \\ p_{2,0} & p_{2,1} & \cdots & p_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,0} & p_{k,1} & \cdots & p_{k,k-1} \end{pmatrix} \begin{pmatrix} a_1^{k-1} & a_2^{k-1} & \cdots & a_n^{k-1} \\ a_1^{k-2} & a_2^{k-2} & \cdots & a_n^{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ a_1^0 & a_2^0 & \cdots & a_n^0 \end{pmatrix} \\
&= C(\mathcal{P}) \cdot V,
\end{aligned}
$$

其中 V 是由 $a_1, \cdots, a_n$ 定义的 $k \times n$ 阶 Vandermonde 矩阵。容易验证若 $P_{S_1}, P_{S_2}, \cdots, P_{S_k}$ 在 $\mathbb{F}_q$ 上线性无关，则 $\det(C(\mathcal{P})) \neq 0$，因此 G 的任意 k 列都是线性无关的，可将其可以看作是一个 $[n, k]_q$ RS 码的生成矩阵，其估值点为 $a_1, \cdots, a_n$。换句话说，要构造一个 $[n, k]_q$ MDS 码，等价于构造一个集族 $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$，使得由 $\mathcal{S}$ 定义的多项式 $P_{S_1}, P_{S_2}, \cdots, P_{S_k}$ 在 $\mathbb{F}_q$ 上是线性无关的。

2.2.2　二元支撑矩阵

基于上一小节给出的记号，令 $M_S = (m_{i,j})$ 是一个 $k \times n$ 阶二元矩阵，其中 $m_{i,j} = 0$ 当且仅当 $j \in S_i$。于是有 $m_{i,j} = 0$ 当且仅当 $g_{i,j} = 0$，因此我们称 M_S 为 G 的*支撑矩阵*。给定一个 $[n,k]_q$ MDS 码的生成矩阵 G，我们可以确定它的支撑矩阵，然后得到一个集族 $S = \{S_1, S_2, \cdots, S_k\}$，其中对于每个 $i \in [k]$，有 $S_i = \{j \in [n] : g_{i,j} = 0\} \subset [n]$，且由最小距离为 $d = n - k + 1$ 可知，每个 S_i 的大小最多为 $k-1$。如果 G 是稀疏的，则所有 $|S_i| = k - 1$。

2.3　稀疏的 MDS 码存在性的支集约束条件

本节中，我们将证明具有特定性质的集族 $S = \{S_1, S_2, \cdots, S_k\}$，会在满足 $q \geqslant n$ 的有限域 $\mathbb{F}_q$ 上生成 k 个线性无关的多项式，记其构成的多项式序列为 $\mathcal{P}$，即 $\det(C(\mathcal{P})) \neq 0$。因此，根据上一节，由 $\mathcal{P}$ 定义的矩阵 G 可作为一个 $[n,k]_q$ RS 码的生成矩阵。

2.3.1　稀疏码的支集约束条件

在本小节中，令 $n \geqslant k > 1$。设 $S = \{S_1, S_2, \cdots, S_k\}$ 为 $[n]$ 上的 $(k-1)$-*一致*的集族，即对于每个 $i \in [k]$，有 $|S_i| = k - 1$。如果存在 $i \in [k-1]$，使得 $|S_1 \cap S_2 \cap \cdots \cap S_i| = k - i$ 以及 $|S_{i+1} \cap S_{i+2} \cap \cdots \cap S_k| = i$，那么称 S 在 i 处*可分*，i 是*分离索引*。如果 $\cap_{i \in [k]} S_i = \varnothing$，则称 S 是*不交的*。

给定一个不交且 $(k-1)$-一致的集族 $S = \{S_1, S_2, \cdots, S_k\}$，它在某个分离索引 $i \in [k-1]$ 处可分，假设交集 $A = S_1 \cap S_2 \cap \cdots \cap S_i$，其大小为 $k-i$，并且交集 $B = S_{i+1} \cap S_{i+2} \cap \cdots \cap S_k$，其大小为 i。通过不交的性质可知，$A \cap B = \varnothing$。对于任意 $j \in [i]$，令 $S'_j = S_j \setminus A$，且对于任意 $j \in [i+1, k]$，令 $S'_j = S_j \setminus B$。我们说 $\mathcal{A} = \{S'_1, \cdots, S'_i\}$ 和 $\mathcal{B} = \{S'_{i+1}, \cdots, S'_k\}$ 是 S 在分隔索引 i 处的两个*剩余集族*。注意到，$\mathcal{A}$ 是 $(i-1)$- 一致的，$\mathcal{B}$ 是 $(k-i-1)$-一致的。特别是当 $i = 1$ 或 $i = k-1$ 时，$\mathcal{A}$ 或 $\mathcal{B}$ 将退化为 $\{\varnothing\}$。令 $\mathcal{P}_1$ 为由 $\mathcal{A}$ 定义的多项式集合 $P_{S'_1}, P_{S'_2}, \cdots, P_{S'_i}$，$\mathcal{P}_2$ 为由 $\mathcal{B}$ 定义的多项式集合 $P_{S'_{i+1}}, \cdots, P_{S'_k}$。在这里，如果集合 $S' = \varnothing$，则令 $P_{S'}(x) = 1$。给定如下两个多项式：$f_0(x) = \prod_{u \in A}(x - a_u)$ 和 $g_0(x) = \prod_{v \in B}(x - a_v)$。最后，令 $\mathcal{Q}_1$ 为多项式集合 $\{x^{i-1} f_0, x^{i-2} f_0, \cdots, f_0\}$，$\mathcal{Q}_2$ 为多项式集合 $\{x^{k-i-1} g_0, x^{k-i-2} g_0, \cdots, g_0\}$。注意到，$\mathcal{Q}_1$ 和 $\mathcal{Q}_2$ 中每个多项式的度数最多为 $k-1$。在这些符号下，我们给出以下引理 2.2–2.3。为方便起见，我们将 $[M_1; M_2]$ 表示为相同列数的两个矩阵 M_1 和 M_2 的垂直串联。

引理 2.2　设 $(k-1)$-一致的集族 $S = \{S_1, S_2, \cdots, S_k\}$ 是不交的且在 $i \in [k-1]$

处可分，则

$$C(\mathcal{P}) = \begin{pmatrix} C(\mathcal{P}_1) & 0 \\ 0 & C(\mathcal{P}_2) \end{pmatrix} \begin{pmatrix} C(\mathcal{Q}_1) \\ C(\mathcal{Q}_2) \end{pmatrix}.$$

证明　对任意 $j \in [i]$，令 $P_{S'_j}(x) = \prod_{\ell \in S'_j}(x - a_\ell) = c_{j,0}x^{i-1} + c_{j,1}x^{i-2} + \cdots + c_{j,i-1}$。则

$$P_{S_j} = f_0 \cdot P_{S'_j} = c_{j,0}x^{i-1}f_0 + c_{j,1}x^{i-2}f_0 + \cdots + c_{j,i-1}f_0.$$

因此在 P_{S_j} 中，对任意 $e \in [0, k-1]$，x^e 项系数为 $[x^e]P_{S_j} = \sum_{\ell=0}^{i-1} c_{j,\ell} \times [x^e](x^{i-\ell-1}f_0)$。由于 $C(P_{S_1}, \cdots, P_{S_i})$ 第 $(j, k-e)$ 个位置上的元素为 $[x^e]P_{S_j}$，故

$$\begin{aligned} & C(P_{S_1}, \cdots, P_{S_i}) \\ =& \begin{pmatrix} c_{1,0} & \cdots & c_{1,i-1} \\ c_{2,0} & \cdots & c_{2,i-1} \\ \vdots & \ddots & \vdots \\ c_{i,0} & \cdots & c_{i,i-1} \end{pmatrix} \begin{pmatrix} [x^{k-1}](x^{i-1}f_0) & \cdots & [x^0](x^{i-1}f_0) \\ [x^{k-1}](x^{i-2}f_0) & \cdots & [x^0](x^{i-2}f_0) \\ \vdots & \ddots & \vdots \\ [x^{k-1}]f_0 & \cdots & [x^0]f_0 \end{pmatrix} \\ =& C(\mathcal{P}_1)C(\mathcal{Q}_1). \end{aligned}$$

$C(P_{S_{i+1}}, \cdots, P_{S_k}) = C(\mathcal{P}_2)C(\mathcal{Q}_2)$ 可以用类似方法得到。注意到 $C(\mathcal{P}) = \left[C\left(P_{S_1}, \cdots, P_{S_i}\right); C\left(P_{S_{i+1}}, \cdots, P_{S_k}\right)\right]$，故证明完成。 ∎

引理 2.3　矩阵 $[C(\mathcal{Q}_1); C(\mathcal{Q}_2)]$ 的行列式非零。特别地，$\det[C(\mathcal{Q}_1); C(\mathcal{Q}_2)] = \prod_{u \in A, v \in B}(a_u - a_v)$。

证明　令 $A = \{u_1, \cdots, u_{k-i}\}$ 和 $B = \{v_1, \cdots, v_i\}$。对任意 $s \in [k-i], t \in [i]$，设 $f_s(x) = (x - a_{u_{s+1}}) \cdots (x - a_{u_{k-i}}) = c_{s,0}x^{k-i-s} + \cdots + c_{s,k-i-s}$ 以及 $g_t(x) = (x - a_{v_{t+1}}) \cdots (x - a_{v_i}) = d_{t,0}x^{i-t} + \cdots + d_{t,i-t}$。注意到，$f_s$ 和 g_t 可以分别通过删除 f_0 和 g_0 的一些线性因子得到。此外，对于所有 $s \in [k-i]$ 和 $t \in [i]$，$c_{s,0} = d_{t,0} = 1$，并且 $f_{k-i} = g_i = 1$。

令 $M_0 = [C(\mathcal{Q}_1); C(\mathcal{Q}_2)] = \left[C(x^{i-1}f_0); \cdots; C(f_0); C(x^{k-i-1}g_0); \cdots; C(g_0)\right]$。我们将通过进行初等行变换来计算矩阵 M_0 的行列式。由于 M_0 的每一行都对应一个多项式，我们使用多项式运算来表示行变换。为了方便起见，对任意 $\ell \in [i]$，$\ell' \in [k-i]$，我们用 R^ℓ，表示行 $C(x^{\ell-1}f_0)$，以及 $R_{\ell'}$ 表示行 $C(x^{\ell'-1}g_0)$。接下来，我们将按步骤进行以下行变换。

第 1 步。将 R_{k-i} 变换为 $c_{1,0}R_{k-i} + \cdots + c_{1,k-i-1}R_1 - (d_{1,0}R^i + \cdots + d_{1,i-1}R^1)$。

注意到 $c_{1,0}=1$。故此时这一行对应的多项式就变成了

$$\begin{aligned}&\sum_{j=0}^{k-i-1} c_{1,j}x^{k-i-1-j}g_0-\sum_{j=0}^{i-1} d_{1,j}x^{i-1-j}f_0\\=&f_1g_0-f_0g_1=(x-a_{v_1})f_1g_1-(x-a_{u_1})f_1g_1\\=&(a_{u_1}-a_{v_1})f_1g_1.\end{aligned}$$

对新得到的 $\boldsymbol{R}_{k-i}$ 继续进行行操作，具体来说，从中减去项 $(a_{u_1}-a_{v_1})(d_{2,0}\boldsymbol{R}^{i-1}+\cdots+d_{2,i-2}\boldsymbol{R}^1)$，此时可得到

$$\begin{aligned}&f_1g_0-f_0g_1-(a_{u_1}-a_{v_1})f_0g_2\\=&(a_{u_1}-a_{v_1})(x-a_{v_2})f_1g_2-(a_{u_1}-a_{v_1})(x-a_{u_1})f_1g_2\\=&(a_{u_1}-a_{v_1})(a_{u_1}-a_{v_2})f_1g_2.\end{aligned}$$

继续对 $\boldsymbol{R}_{k-i}$ 进行类似行变换，我们可得到多项式

$$f_1g_0-f_0g_1-\sum_{t=1}^{i-1}\left(\prod_{s=1}^{t}(a_{u_1}-a_{v_s})\right)f_0g_{t+1}=f_1\prod_{t=1}^{i}(a_{u_1}-a_{v_t}).$$

因此，我们将行 $C(x^{k-i-1}g_0)$ 变换为 $C(f_1\prod_{t=1}^{i}(a_{u_1}-a_{v_t}))$，且行列式不变。将非零因子 $\prod_{t=1}^{i}(a_{u_1}-a_{v_t})$ 提取出来并假设新行是 $C(f_1)$。观察到对于每个 $j=0,1,\cdots,i-1$，有 $x^jf_0+x^ja_{u_1}f_1=x^j(x-a_{u_1})f_1+x^ja_{u_1}f_1=x^{j+1}f_1$。然后我们可以做一系列的行操作：对任意 $\ell=1,\cdots,i-1$，将 $a_{u_1}\times\boldsymbol{R}_{k-i}$ 加到 $\boldsymbol{R}^1$，将 $a_{u_1}\times\boldsymbol{R}^{\ell}$ 加到 $\boldsymbol{R}^{\ell+1}$。经过这些操作后，可以得到一个新的矩阵 $\boldsymbol{M}_1=\prod_{t=1}^{i}(a_{u_1}-a_{v_t})\left[C(x^if_1);\cdots;C(f_1);C(x^{k-i-2}g_0);\cdots;C(g_0)\right]$。

第 2 步。注意到矩阵 $\boldsymbol{M}_1$ 的结构与初始矩阵 $\boldsymbol{M}_0$ 类似。故我们将更新行记号如下：对任意 $\ell\in[i+1]$，$\ell'\in[k-i-1]$，我们用 $\boldsymbol{R}^{\ell}$，表示行 $C(x^{\ell-1}f_0)$，以及 $\boldsymbol{R}_{\ell'}$ 表示行 $C(x^{\ell'-1}g_0)$。基于这些新记号下，可以对 $\boldsymbol{R}_{k-i-1}$ 进行类似于第一步中的行变换，由此得到新矩阵 $\boldsymbol{M}_2=\prod_{t\in[i],j\in[2]}(a_{u_j}-a_{v_t})\left[C(x^{i+1}f_2);\cdots;C(f_2);C(x^{k-i-3}g_0);\cdots;C(g_0)\right]$。我们将在第 r 步中详细说明这些变换。

第 $r\leqslant k-i$ 步。此时我们得到矩阵 $\boldsymbol{M}_{r-1}=\prod_{t\in[i],j\in[r-1]}(a_{u_j}-a_{v_t})[\,C(x^{i+r-2}f_{r-1});\cdots;C(f_{r-1});C(x^{k-i-r}g_0);\cdots;C(g_0)\,]$。更新记号如下：对任意 $\ell\in[i+r-1]$，$\ell'\in[k-i-r+1]$，我们用 $\boldsymbol{R}^{\ell}$，表示行 $C(x^{\ell-1}f_0)$，以及 $\boldsymbol{R}_{\ell'}$ 表示行 $C(x^{\ell'-1}g_0)$。考虑 $\boldsymbol{R}_{k-i-r+1}$，即行 $C(x^{k-i-r}g_0)$，首先做行变换 $c_{r,0}\boldsymbol{R}_{k-i-r+1}+\cdots+c_{r,k-i-r}\boldsymbol{R}_1-(d_{1,0}\boldsymbol{R}^i+\cdots+d_{1,i-1}\boldsymbol{R}^1)$ 得到 $f_rg_0-f_{r-1}g_1$，然后对新得

到的 $R_{k-i-r+1}$ 继续进行行变换 $\sum_{t=1}^{i-1}(\prod_{s=1}^{t}(a_{u_r}-a_{v_s}))(d_{t+1,0}R^{i-t}+\cdots+d_{t+1,i-t-1}R^1)$，得到

$$f_r g_0 - f_{r-1}g_1 - \sum_{t=1}^{i-1}\left(\prod_{s=1}^{t}(a_{u_r}-a_{v_s})\right) f_{r-1}g_{t+1} = f_r\prod_{t=1}^{i}\left(a_{u_r}-a_{v_t}\right).$$

因此我们将行 $C(x^{k-i-r}g_0)$ 更新为 $C(f_r\prod_{t=1}^{i}(a_{u_r}-a_{v_t}))$ 且不改变初始矩阵行列式。观察到对于每个 j 满足 $0\leqslant j\leqslant i+r-2$，都有 $x^{j+1}f_r = x^j f_{r-1} + a_{u_r}x^j f_r$。然后我们可以做一系列的行操作：对任意 $\ell = 1,\cdots,i+r-2$，将 $a_{u_r}\times R_{k-i-r+1}$ 加到 R^1 以及 $a_{u_r}\times R^{\ell}$ 加到 $R^{\ell+1}$。经过第 r 步操作后，我们可以将第 $r-1$ 步中的矩阵 M_{r-1} 更新到如下新矩阵

$$\prod_{t\in[i],j\in[r]}(a_{u_j}-a_{v_t})\left[C(x^{i+r-1}f_r);\cdots;C(f_r);C(x^{k-i-r-1}g_0);\cdots;C(g_0)\right].$$

经过 $(k-i)$ 步操作后，我们将初始矩阵 M_0 更新为如下形式：

$$\prod_{t\in[i],j\in[k-i]}(a_{u_j}-a_{v_t})\left[C(x^{k-1});C(x^{k-2});\cdots;C(1)\right].$$

由于上述步骤中的所有行操作都没有改变行列式，故证明完成。 ∎

2.3.2　良好支集的刻画

在第2.3.1小节中，由引理 2.2 和 2.3可知，当 $\mathcal{S}$ 是不交的且可分时，我们有

$$\det(C(\mathcal{P})) = \det(C(\mathcal{P}_1))\det(C(\mathcal{P}_2))\prod_{u\in A,v\in B}(a_u-a_v).$$

为了确保 $\det(C(\mathcal{P}))\neq 0$，我们需要 $\det(C(\mathcal{P}_1))$ 和 $\det(C(\mathcal{P}_2))$ 都是非零的。这促使我们通过以下方式从集族 $\mathcal{S}$ 出发，定义二叉树。令集族 $\mathcal{S}=\{S_1,S_2,\cdots S_k\}$ 为根节点。如果 $\mathcal{S}$ 不是不交的或可分的，则停止。否则，令 $\mathcal{S}$ 的两个剩余集族 $\mathcal{A}$ 和 $\mathcal{B}$ 为 $\mathcal{S}$ 的左右子节点。然后考虑 $\mathcal{A}$ 和 $\mathcal{B}$，以 $\mathcal{A}$ 为例。如果 $\mathcal{A}$ 不是不交的或可分的，或者 $\mathcal{A}=\{\varnothing\}$，则停止。否则，我们可以通过它的两个剩余集族来扩展 $\mathcal{A}$。继续执行这样的操作，直到不能再扩展为止。如果生成的二叉树，其所有叶子节点都是 $\{\varnothing\}$，那么我们称它是一个*良好的* 二叉树。注意到，由 $\mathcal{S}$ 构造的二叉树可能不是唯一的。如果 $\mathcal{S}$ 至少可以产生一棵良好的二叉树，则称 $\mathcal{S}$ 是 *良好的*。

例 2.1　设 $S_1=\{5,6,7,8\}, S_2=\{1,6,7,8\}, S_3=\{1,2,7,8\}, S_4=\{1,2,3,4\}$ 以及 $S_5=\{2,3,4,5\}$。则 $\mathcal{S}=\{S_1,S_2,S_3,S_4,S_5\}$ 是一个 4-一致的集族。根据上述定义，$\mathcal{S}$ 可以构造出一棵良好的二叉树，故 $\mathcal{S}$ 是良好的。

事实上，$\mathcal{S}$ 是不交的且在 $i=3$ 处可分，这是因为 $|S_1\cap S_2\cap S_3| = |\{7,8\}| = 2$ 以及 $|S_4\cap S_5| = |\{2,3,4\}| = 3$。因此我们有 $\mathcal{S}$ 在分隔索引 3 处的两个剩余集族 $\mathcal{A}=$

$\{S'_1, S'_2, S'_3\}$ 和 $\mathcal{B} = \{S'_4, S'_5\}$，其中 $S'_1 = \{5,6\}, S'_2 = \{1,6\}, S'_3 = \{1,2\}, S'_4 = \{1\}$ 以及 $S'_5 = \{5\}$。由于 $S'_1 \cap S'_2 = \{6\}$，集族 $\mathcal{A}$ 是不交的且可分。令 $S''_1 = \{5\}, S''_2 = \{1\}$ 和 $\mathcal{C} = \{S''_1, S''_2\}$。那么 $\mathcal{B}$ 和 $\mathcal{C}$ 都是平凡的不交且可分的集族。综上所述，$\mathcal{S}$ 可以生成一个如图 2.1所示的二叉树，可以看到该二叉树的所有叶子节点都是 $\{\varnothing\}$，故是良好的。

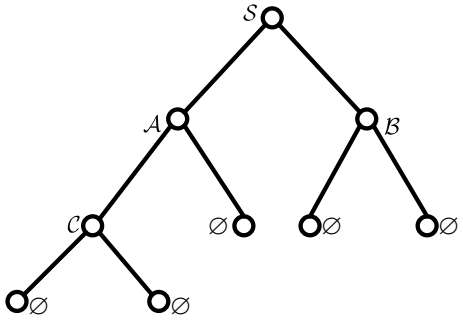


图 2.1　例2.1中集族 $\mathcal{S}$ 生成的二叉树

为了更加清楚的说明由集族 $\mathcal{S}$ 生成的良好二叉树的结构，我们从上到下逐层绘制二叉树，并且在每一层中，节点以自然的方式从左到右列出。然后我们使用从左到右和从上到下的非叶子节点的分离索引构成的序列，来表示特定的二叉树，其中不同层的分离索引用分号隔开。例如图 2.1中的二叉树可以用一个序列 $(3;2,4;1)$ 表示，其中 3 是 $\mathcal{S}$ 在第一层的分离索引，2 和 4 分别是第二层中 $\mathcal{A}$ 和 $\mathcal{B}$ 的分离索引，1 是第三层中唯一的非叶子节点 $\mathcal{C}$ 的分离索引。注意到，对于表示剩余集族的每个非叶子节点，例如节点 $\mathcal{B}$，我们使用根据 $\mathcal{S}$ 定义的原始分离索引 4 而不是其自身的分离索引 1。

注 1　对于一个良好的 $(k-1)$-一致的集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$，或由它生成的良好的二叉树，我们有以下事实。

1. 二叉树中恰好有 k 个叶子节点，每个叶子节点对应于 $\mathcal{S}$ 中的一个集合。
2. 分离索引序列包含每个分离索引 $i \in [k-1]$ 恰好一次。
3. 对 $\mathcal{S}$ 的支撑矩阵的列进行置换，可以得到一个新的良好集族的支撑矩阵。

特别地，如果一个集族只有一个元素，且是空集，即 $\{\varnothing\}$，那么该集族只定义了一个常数多项式。因此，相应的系数矩阵是一个 1×1 阶矩阵且其中元素为 1，即 $C(\{\varnothing\}) = (1)$，其行列式为 1。

2.3.3　稀疏的 MDS 码存在性的刻画

结合引理 2.2和2.3，以及之前的所有分析可知，如果集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 是良好的，那么可从 $\mathbb{F}_q$ 中任意选取 n 个不同的元素 $a_1, \cdots, a_n \in \mathbb{F}_q$ 来定义 P_{S_i}，从而生成一个可逆系数矩阵 $C(\mathcal{P})$，最后得到一个矩阵 G，G 可作为一个 $[n,k]_q$ RS 码的生成矩阵，且由此得到的 RS 码的支撑矩阵为 $M_{\mathcal{S}}$。我们将这个结果总结为

以下定理。

定理 2.4　设 $\mathcal{S}=\{S_1,\cdots,S_k\}$ 是 $[n]$ 上的良好的 $(k-1)$-一致集族。那么对于任何满足 $q\geqslant n$ 的有限域 $\mathbb{F}_q$ 上，均存在一个稀疏的 $[n,k]_q$ RS 码，其生成矩阵 G，满足 $G_{ij}=0$ 当且仅当 $j\in S_i$，即 $M_{\mathcal{S}}$ 是 G 的支撑矩阵。

注 2　在文献 [54] 的定理 II.5 中，作者也给出了当 $q\geqslant n$ 时，$[n,k]_q$ MDS 码存在的充分条件，即对任意 $1\leqslant i\leqslant k$，其生成矩阵的零模式 $\mathcal{S}=\{S_1,\cdots,S_k\}$ 满足 $|\cap_{j=1}^{i}S_j|=k-i$。注意到，当 $\mathcal{S}$ 是 $(k-1)$-一致的集族时，这个 $\mathcal{S}$ 是良好的，因为它可以生成一个良好的二叉树 $(k-1;k-2;\cdots;1)$。故定理 2.4 可以看作是文献 [54] 的定理 II.5 对一致的情况的扩展。

接下来，我们将通过下面的例子来说明引理 2.2和2.3，以及定理 2.4。

例 2.2　设 $\mathcal{S}=\{S_1,S_2,S_3,S_4,S_5\}$ 是例 2.1中给定的集族，则 $\mathcal{S}$ 是一个 $[8]$ 上良好的 4-一致集族，且它对应一棵良好的二叉树 $(3;2,4;1)$，见图 2.1。我们将在 $\mathbb{F}_8$ 上构造一个 5×8 阶矩阵 G，且 G 中元素满足 $G_{ij}=0$ 当且仅当 $j\in S_i$。令 $\mathcal{P}$ 为 $\mathbb{F}_8[x]$ 中的多项式序列 $P_{S_1},\cdots,P_{S_5}$，其中

$$
\begin{aligned}
P_{S_1}(x)&=(x-a_5)(x-a_6)(x-a_7)(x-a_8),\\
P_{S_2}(x)&=(x-a_1)(x-a_6)(x-a_7)(x-a_8),\\
P_{S_3}(x)&=(x-a_1)(x-a_2)(x-a_7)(x-a_8),\\
P_{S_4}(x)&=(x-a_1)(x-a_2)(x-a_3)(x-a_4),\\
P_{S_5}(x)&=(x-a_2)(x-a_3)(x-a_4)(x-a_5).
\end{aligned}
$$

由引理 2.2 和 2.3可知，矩阵 $C(\mathcal{P})$ 的行列式为

$$
\det(C(\mathcal{P}))=\left(\prod_{u\in\{7,8\},v\in\{2,3,4\}}(a_u-a_v)\right)\times\left(\prod_{v'\in\{1,2\}}(a_6-a_{v'})\right)\times(a_1-a_5)\times(a_5-a_1).
$$

故若 $a_1,\cdots,a_8$ 是 $\mathbb{F}_8$ 上互不相同的 8 个元素，则 $\det(C(\mathcal{P}))$ 在 $\mathbb{F}_8$ 上非零。令 ζ 是 $\mathbb{F}_8$ 的一个本原元，且满足 $\zeta^3+\zeta+1=0$。设 $a_1=0$ 且对任意 $i\in[2,8]$，$a_i=\zeta^{i-2}$。则有 $\det(C(P))=1\neq 0$，且可得到一个 $\mathbb{F}_8$ 上的 5×8 阶矩阵 G，

$$
G=\begin{pmatrix}
\zeta^4 & \zeta^5 & \zeta^6 & \zeta^2 & 0 & 0 & 0 & 0\\
0 & \zeta^4 & 1 & \zeta^6 & \zeta & 0 & 0 & 0\\
0 & 0 & \zeta & \zeta^4 & \zeta^3 & \zeta^5 & 0 & 0\\
0 & 0 & 0 & 0 & \zeta^2 & \zeta^5 & \zeta^4 & \zeta^6\\
\zeta^6 & 0 & 0 & 0 & 0 & 1 & \zeta & \zeta^4
\end{pmatrix}.
$$

由定理 2.4可知，G 是一个 $\mathbb{F}_8$ 上的 $[8,5]$ RS 码的稀疏生成矩阵。且容易看出，G 也是平衡的。

事实上，满足定理 2.4 条件的良好的 $(k-1)$-一致集族 $\mathcal{S}$ 是平凡存在的。例如，对任意 $i \in [k]$，令 $S_i = [i, i+k-2] \bmod^+ n$，则 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 是一个良好的集族且其可生成一棵良好的二叉树 $(1; 2; \cdots; k-1)$，其中非叶子节点在分离索引 1，2，$\cdots$，$k-1$ 处按顺序分隔。因此，对于任何 $q \geqslant n$，都存在一个稀疏的 $[n,k]_q$ RS 码。然而，一般情况下，这种构造方法无法得到一个平衡的 RS 码。为了得到一个稀疏且平衡的 RS 码，我们需要找到一个良好的 $(k-1)$-一致集族 $\mathcal{S}$，使得矩阵 $M_{\mathcal{S}}$ 每列的重量几乎相同。为了方便起见，如果矩阵 G 是稀疏且平衡的，我们也称 G 所对应的集族 $\mathcal{S}$ 或矩阵 $M_{\mathcal{S}}$ 是稀疏且平衡的，如果 $\mathcal{S}$ 是良好的，则称 $M_{\mathcal{S}}$ 也是良好的。

2.4　码长 $n \leqslant 2k$ 下稀疏平衡的 MDS 码的构造

在本节中，我们证明主要结果，即定理 1.1，它指出了，在任意满足 $q \geqslant n-1$ 的小域 $\mathbb{F}_q$ 上，均存在稀疏平衡的 MDS 码。首先，我们证明 $q \geqslant n$ 的情况，即定理 2.5。对于 $q = n-1$ 的情况，可以从定理 2.5所得到的集族 $\mathcal{S}$ 中，有选择的删除一个恰当的坐标，然后再延伸到长度 n 得到，从而完成定理 1.1。

定理 2.5　对任何正整数 $k \geqslant 3$，若 k 是偶数，则令 $n \leqslant 2k$，若 k 是奇数，则令 $n \leqslant 2k-1$。那么在任何满足 $q \geqslant n$ 的有限域 $\mathbb{F}_q$ 上，都存在一个具有稀疏平衡生成矩阵的 $[n,k]_q$ MDS 码。

2.4.1　小域（$q \geqslant n$）上稀疏平衡的 MDS 码存在性的等价刻画

在本小节中，为了证明定理 2.5，我们需要通过定理 2.4 构造一个良好且平衡的 $(k-1)$-一致集族 $\mathcal{S}$。换句话说，我们需要构造一个满足以下性质的 $k \times n$ 阶二元矩阵 $M = (m_{i,j}) = M_{\mathcal{S}}$：

(P_1) 稀疏条件：M 的每一行都恰好有 $k-1$ 个 0；

(P_2) 平衡条件：M 中有 θ 列均含有 $\left\lceil \frac{k(k-1)}{n} \right\rceil$ 个 0，剩余 $n-\theta$ 列均含有 $\left\lfloor \frac{k(k-1)}{n} \right\rfloor$ 个 0，其中 $\theta \triangleq k(k-1) \bmod^+ n$；

(P_3) 良好条件：对任意 $i \in [k]$，设 $S_i = \{j : m_{i,j} = 0\} \subseteq [n]$，则集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 是良好的。

在本章剩余部分，我们不区分 $\mathcal{S}$ 和 $M_{\mathcal{S}}$。注意到，对任意 $i \in [k]$，令 $S_i = [i, i+k-2] \bmod^+ n$，则此集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 可以给出一个同时满足 (P_1) 和 (P_3) 的矩阵 $M_{\mathcal{S}}$。而构造一个同时满足 (P_1) 和 (P_2) 的矩阵也很容易。然而，构造一个满足上述所有三个条件的矩阵并不简单。我们首先证明，为了满足 (P_1)-(P_3)，定理 2.5 中对于 n, k 的限制是必要的。

引理 2.6　给定两个正整数 n, k 且满足 $n \geqslant k \geqslant 3$，如果存在一个稀疏、良好

且平衡的 $k \times n$ 阶二元矩阵，则 $n \leqslant 2k$。此外，如果 $n = 2k$，则 k 必须是偶数。

证明　设 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 和 $M_{\mathcal{S}}$ 是稀疏、良好且平衡的。根据良好这一性质知，$\mathcal{S}$ 在某个分离索引处可分，即存在 $i \in [k-1]$，使得 $|S_1 \cap S_2 \cap \cdots \cap S_i| = k - i$ 以及 $|S_{i+1} \cap S_{i+2} \cap \cdots \cap S_k| = i$。根据平衡条件可知，$i$ 和 $k-i$ 均不超过 $\left\lceil \frac{k(k-1)}{n} \right\rceil \leqslant \left\lceil \frac{k(k-1)}{2k} \right\rceil = \left\lceil \frac{k-1}{2} \right\rceil$。若 $i < \left\lceil \frac{k(k-1)}{n} \right\rceil \leqslant \left\lceil \frac{k-1}{2} \right\rceil$，则

$$\begin{aligned} k - i &\geqslant k - \left\lceil \frac{k-1}{2} \right\rceil + 1 = k - 1 - \left\lceil \frac{k-1}{2} \right\rceil + 2 \\ &= \left\lfloor \frac{k-1}{2} \right\rfloor + 2 \geqslant \left\lceil \frac{k-1}{2} \right\rceil + 1. \end{aligned}$$

这与 $k - i \leqslant \left\lceil \frac{k-1}{2} \right\rceil$ 相矛盾。同理 $k - i < \left\lceil \frac{k(k-1)}{n} \right\rceil$ 的情形也是不成立的。因此 $i = k - i = \frac{k}{2} = \left\lceil \frac{k(k-1)}{n} \right\rceil$。故 k 必须是偶数。由于 $\mathcal{S}$ 在 $i = \frac{k}{2}$ 处是可分的，那么 $M_{\mathcal{S}}$ 中至少有 k 列，它们每列至少包含 $\frac{k}{2}$ 个 0。这不包括所有 $k \geqslant 4$ 且 $n = 2k+1, 2k+2$，以及 $k = 4$ 且 $n = 2k+3$ 的情况，因为对于这些情况，$\theta = k(k-1) \bmod^+ n$ 严格小于 k。更具体地，

a) 若 $n = 2k+1$，当 $k \geqslant 4$ 时，有 $\theta = k(k-1) \bmod^+ n \equiv k(k+\frac{1}{2}) - \frac{3k}{2} + 2k + 1 \pmod n = \frac{k}{2} + 1 < k$；

b) 若 $n = 2k+2$，当 $k \geqslant 4$ 时，有 $\theta = k(k-1) \bmod^+ n \equiv k(k+1) - 2k + 2k + 2 \pmod n = 2 < k$；

c) 若 $n = 2k+3$，当 $k = 4$ 时，即 $n = 11$，此时 $\theta = k(k-1) \bmod^+ n = 12 \bmod^+ 11 = 1 < 4$。

下面，可以根据 $\frac{k}{2} = \left\lceil \frac{k(k-1)}{n} \right\rceil$ 排除其它所有情况。设 $t \geqslant 4$ 且 $n = 2k + t$，或者 $t = 3$ 且 $k \geqslant 6$，则有

$$\left\lceil \frac{k(k-1)}{n} \right\rceil = \left\lceil \frac{k(k-1)}{2k+t} \right\rceil = \left\lceil \frac{k}{2} - \frac{(t+2)k}{4k+2t} \right\rceil \leqslant \frac{k}{2} - 1,$$

因此矛盾。

故由假设可知，要么 $n = 2k$ 且 k 是偶数，要么 $n < 2k$。 ∎

由引理 2.6可知，下面我们只需要考虑 $n < 2k$，或者 $n = 2k$ 且 k 是偶数的情形。对于后一种情况，我们有 $\theta = k$，也就是说，$M_{\mathcal{S}}$ 中恰好有 k 列，均包含 $\frac{k}{2}$ 个 0，而剩余 k 列，每列均包含 $\frac{k}{2} - 1$ 个 0。这样的 $k \times n$ 阶矩阵可以通过循环向右移动向量 $(\underbrace{0, \cdots, 0}_{k-1}, 1, \cdots, 1)$ 和 $(\underbrace{1, \cdots, 1}_{k}, \underbrace{0, \cdots, 0}_{k-1}, 1)$ 各 $\frac{k}{2}$ 次得到。正式地构造将在第2.4.3小节给出，见构造 2.13。

2.4.2　一般码长的稀疏平衡的 MDS 码的构造

在本小节中，设 $n < 2k$。我们将通过给出几个算法来保证稀疏平衡 MDS 码的存在性。具体想法是，首先给定一个初始矩阵 $M_{\mathcal{S}}$，该矩阵满足性质 (P_1) 和

(P_3)，但不满足 (P_2)，即该矩阵是稀疏且良好的，但不平衡。然后按照特定规则逐步地更新 $\boldsymbol{M}_{\mathcal{S}}$，且每次更新都保证矩阵的稀疏性和良好性不变，但更新后所得矩阵变得更加平衡。直到更新后的新矩阵 $\boldsymbol{M}_{\mathcal{S}}$ 完全平衡为止。这里的特定规则由引理 2.7给出。

由第2.4.1小节的分析可知，当 $n = 2k$ 时，可以通过循环两个特定的向量来构造 $\boldsymbol{M}_{\mathcal{S}}$。但是当 $n < 2k$ 时，这种方法可能会失效，这是因为对二元矩阵 $\boldsymbol{M}_{\mathcal{S}}$ 而言，零的数量始终保持不变，即 $k(k-1)$ 个，但是 n 在不断减小，如果继续使用这种构造方法，将会使得 $\boldsymbol{M}_{\mathcal{S}}$ 中间列中包含的零不足。

事实上，二元矩阵的稀疏平衡性很容易保证，针对良好性，观察到，一个由长度为 n 且权重为 $n-k+1$ 的二元向量生成的循环矩阵，它自然的满足良好条件，且同时也是稀疏的。但是，一般情况下这样的矩阵是不平衡的。接下来，我们证明，如果 $\boldsymbol{M}_{\mathcal{S}}$ 是二元循环矩阵，那么我们可以在每一行中调整零的位置，但保持每行零的数量不变，使得更新后的 $\boldsymbol{M}_{\mathcal{S}}$ 仍然保持稀疏性和良好性，但和初始矩阵相比，变得更加平衡。具体可以参见例 2.3。

例 2.3　设 $k = 6$ 以及 $n = 10$。令 $\boldsymbol{M}_{\mathcal{S}}$ 是通过将向量 $(0,0,0,0,0,1,1,1,1,1)$ 循环右移六次得到的 6×10 阶二元循环矩阵。注意到，$\mathcal{S} = \{S_1, \cdots, S_6\}$ 可以生成一棵良好的二叉树 $(3;1,5;2,4)$。令 $i = 3$，如下所示，我们通在第 $(i-1)$ 和第 $(i+1)$ 行，以及第 $(i-1)$ 和第 $(i+k-1)$ 列处，将 $\boldsymbol{M}_{\mathcal{S}}$ 分成几个块。注意到，$i = 3$ 是第一个分离索引，即 $\mathcal{S}$ 在 $i = 3$ 处可分。

$$
\boldsymbol{M}_{\mathcal{S}} = \left(\begin{array}{cc|cccccc|cc}
0 & 0 & 0 & 0 & 0 & \textcircled{1} & 1 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & \textcircled{1} & 1 & 1 & 1 \\
\hline
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
\hline
1 & 1 & 1 & \textcircled{1} & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 1 & 1 & 1 & \textcircled{1} & 0 & 0 & 0 & 0 & 0
\end{array}\right).
$$

接下来，我们所有的调整都将在同一行进行。观察到，左上角和右下角的 0 可以与它们所在行中的任何 1 交换，圆圈中的 1 除外。否则，它们可能会产生重复的行。对于所有其他 0，保持原处不动。这种交换不会破坏 $\mathcal{S}$ 的良好性，且新得到的集族也与 $\mathcal{S}$ 具有相同的二叉树。例如我们可以将 $\boldsymbol{M}_{\mathcal{S}}$ 更新为下面的 $\boldsymbol{M}_{\mathcal{S}'}$，新的集族 $\mathcal{S}' = \{S_1', \cdots, S_6'\}$ 仍然是良好的，因为它可以生成相同的二叉树 $(3;1,5;2,4)$。

$$M_{\mathcal{S}'}=\left(\begin{array}{cc|ccccccc|cc}
1&0&0&0&0&1&1&1&1&0\\
1&1&0&0&0&0&1&1&0&1\\
\hline
1&1&0&0&0&0&0&1&1&1\\
1&1&1&0&0&0&0&0&1&1\\
\hline
0&1&1&1&0&0&0&0&1&1\\
1&0&0&1&1&0&0&0&1&1
\end{array}\right).$$

给定 $\mathcal{S}$ 的一棵良好的二叉树 $\mathcal{T}$，我们将每个子节点 $\mathcal{A}$ 与对应的交集 A 相关联, 见第 2.3.1小节开头给出的记号。容易观察到例 2.3中的两个集族 $\mathcal{S}$ 和 $\mathcal{S}'$ 不仅有相同的二叉树,而且对于所有的非叶子节点,其交集也相同。例如,对于第一个分离索引 $3, S_1\cap S_2\cap S_3 = S'_1\cap S'_2\cap S'_3 = \{3,4,5\}, S_4\cap S_5\cap S_6 = S'_4\cap S'_5\cap S'_6 = \{6,7,8\}$。事实上，在每次操作中保持这些交集不变，是使得更新后的 $\mathcal{S}'$ 拥有一棵与 $\mathcal{S}$ 相同的良好二叉树的主要原因。基于这一观察，我们将例 2.3 推广到更一般的情况。

引理 2.7 (关键操作)　给定正整数 $n \geqslant k$ 和 $\alpha \in [n]$。$\mathcal{S} = \{S_1, \cdots, S_k\}$ 是一个 $(k-1)$-一致的良好集族，且生成的良好二叉树的第一个分离索引是 β。对任意 $j \in [\beta]$，设 $S_j = [\alpha + j - 1, \alpha + j + k - 3] \bmod^+ n$。则对任意 $s \in [\beta - 1]$，可在 $\bmod^+ n$ 意义下定义如下集合:

(1) 对任意 $t \in [s]$，$S'_t = [\Gamma_1, \alpha + k + t - 3] \cup X_t$，其中 $X_t \subseteq [n] \setminus [\Gamma_1, \alpha + k + t - 2]$ 且 $|X_t| = \Gamma_2$;

(2) 对任意 $t \in [s+1, \beta]$，$S'_t = [\alpha + t - 1, \Gamma_3] \cup X_t$，其中 $X_t \subseteq [n] \setminus [\alpha + t - 2, \Gamma_3]$ 且 $|X_t| = \Gamma_4$;

这里，当 $s \in [2, \beta - 2]$ 时，$\Gamma_1 = \alpha + s - 1, \Gamma_2 = s - t, \Gamma_3 = \alpha + k + s - 2$ 以及 $\Gamma_4 = t - s - 1$；否则，$\Gamma_1 = \alpha + s, \Gamma_2 = s - t + 1, \Gamma_3 = \alpha + k + s - 3$ 且 $\Gamma_4 = t - s$。则所得到的新集族 $\mathcal{S}' = \{S'_1, \cdots, S'_\beta, S_{\beta+1}, \cdots, S_k\}$ 仍然是良好的，且可以与 $\mathcal{S}$ 生成一棵相同的良好二叉树。

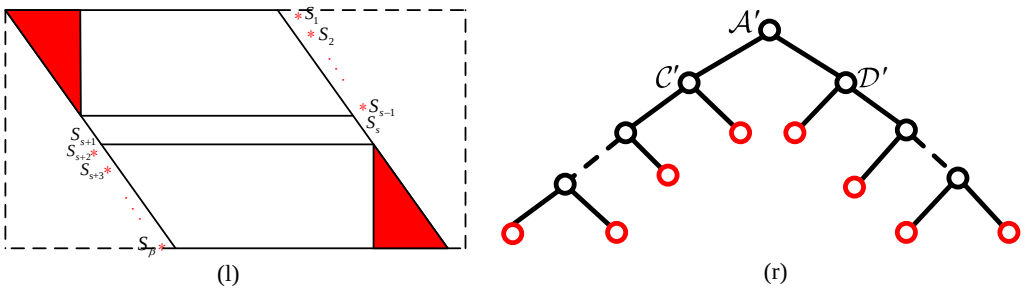
在证明引理 2.7之前，我们首先从几个方面阐述上述操作的具体意义。为了方便起见，设 $\bar{\mathcal{S}} = \{S_1, \cdots, S_\beta\}$ 以及 $\bar{\mathcal{S}}' = \{S'_1, \cdots, S'_\beta\}$。由于矩阵 $M_{\bar{\mathcal{S}}}$ 是一个循环矩阵，且由注 1知，不妨设 $\alpha = 1$。

(i) 在引理 2.7中，我们假设 $M_{\bar{\mathcal{S}}}$ 是循环的，且只对集族 $\bar{\mathcal{S}}$ 中的集合进行了更新。如果其余集合 $S_{\beta+1}, \cdots, S_k$ 也构成循环结构，则引理 2.7 可以同时应用于 $\mathcal{S} \setminus \bar{\mathcal{S}}$。取 $\alpha = 1, \beta = 3$，此时例 2.3中更新后的矩阵可以通过引理 2.7得到，且 $M_{\bar{\mathcal{S}}}$ 和 $M_{\mathcal{S}\setminus\bar{\mathcal{S}}}$ 都是循环矩阵。更具体地，例如在更新 $M_{\bar{\mathcal{S}}}$ 时，可令 $s = 2$。

(ii) 由于我们只需要考虑集族 $\bar{\mathcal{S}}$ 的变化，我们将 $M_{\bar{\mathcal{S}}}$ 中 0 的位置画在图 2.2 (l)

中，其中 0 被实线包围。$M_{\bar{S}}$ 的每一行代表一个集合 S_i，其中 $i \in [\beta]$。这里我们为了简单起见，假设 $k+\beta-2 \leqslant n$，在这种情况下，集族 $\bar{S}$ 中每个集合的最大元素最多为 n。

(iii) 引理 2.7 (1)-(3) 表明对于每一行 $t \in [s-1] \cup [s+2, \beta]$，白色梯形中的 0 不会被移动，并且红色三角形中的 0 可以与除红色星号外的同一行中的任何 1 位置交换。这里，当 $t \in [s-1]$ 时，红色星号代表位置 $\alpha+k+t-2$，而当 $t \in [s+2, \beta]$ 时，代表 $\alpha+t-2$，即紧邻梯形右侧（左侧）的那个位置。特别地，当 $s=1, \beta-1$ 时，矩阵 $M_{\bar{S}}$ 中第 $(\alpha+s-1)$ 列以及第 $(\alpha+k+s-2)$ 列中的零元素可以被更新。



(l)　　(r)

图 2.2　图 (l) 刻画了 $M_{\bar{S}}$ 中零的位置，而图 (r) 是引理 2.7证明中 $\mathcal{A}'$ 的二叉树。

引理 2.7的证明　继续沿用上述给定的符号。注意到 $A = S'_1 \cap \cdots \cap S'_\beta = [\alpha+\beta-1, \alpha+k-2] = S_1 \cap \cdots \cap S_\beta$，其大小为 $k-\beta$，$B = S_{\beta+1} \cap \cdots \cap S_k$，其大小为 β，则 S' 是不交的且在 β 处可分。因此可知对任意集合 $X \in \bar{S}'$ 以及 $Y \in S' \setminus \bar{S}'$，均有 $X \neq Y$，否则 $X = Y \supset (A \cup B)$ 的大小至少为 $k-\beta+\beta = k$，矛盾。令 $\mathcal{A}$ 和 $\mathcal{A}'$ 分别为由 $\bar{S}$ 和 $\bar{S}'$ 生成的 S 和 S' 的剩余集族。通过构造可知，$\mathcal{A}$ 是良好的，且所生成的良好二叉树不唯一。事实上，对任意 $s \in [2, \beta-2]$，$\mathcal{A}$ 可以生成一棵二叉树 $(s; 1, \beta-1; 2, \beta-2; \cdots)$，而当 $s=1$ 或 $\beta-1$ 时，$\mathcal{A}$ 可对应生成二叉树 $(\beta-1; \beta-2; \cdots; 1)$ 或 $(1; 2; \cdots; \beta-1)$，记 $\mathcal{A}$ 生成的二叉树为 $\mathcal{T}$。故只需要证明，对任意 $s \in [\beta-1]$，$\mathcal{A}'$ 可以对应生成与 $\mathcal{A}$ 相同的二叉树即可。图 2.2中刻画了 $s \in [2, \beta-2]$ 的情形。因此，只要说明下面两点：

(a) 集族 $\mathcal{A}'$ 中的各个集合互不相同；或者等价地，集族 $\bar{S}'$ 中的各个集合也互不相同。

(b) 对于二叉树 $\mathcal{T}$ 中的每个非叶子节点，其所对应的交集对于集族 $\mathcal{A}$ 和 $\mathcal{A}'$ 来说是相同的。

容易看出 (a) 是成立的，这是因为对于集族 $\bar{S}'$ 中的任何两个集合，总是可以找到包含在一个集合中但不包含于另外一个集合中的元素。例如，当 $i < s$ 且 $i < j \leqslant \beta$ 时，元素 $\alpha+k+i-2$ 在集合 S'_j 但不在 S'_i 中，故 $S'_i \neq S'_j$。

考虑 (b)。首先我们假设二叉树 $\mathcal{T}$ 可以由 $\mathcal{A}'$ 生成，然后计算 $\mathcal{T}$ 的每个非叶子节点对应的交集，可以验证所有这些交集都与 $\mathcal{A}$ 生成的二叉树 $\mathcal{T}$ 对应节点的交集相同，这里我们省略不再细述。事实上，从图 2.2(l) 以及 $\mathcal{T}$ 的索引序列 $(s;1,\beta-1;2,\beta-2;\cdots)$ 也可以看出，这里我们取 $s\in[2,\beta-2]$。例如，考虑这张图中最上面的梯形，这里梯形的每一行代表一个集合。在 $\mathcal{T}$ 的每一层上，经过更新后，最上面的一行总是与下面的行是分开的，即每次总是将第一个集合与下面的集合分开，且下面的集合交集保持不变。 ∎

引理 2.7说明了，给定一个良好的，且稀疏的循环初始矩阵 $\boldsymbol{M}_S$，我们可以在保持稀疏且良好特性的同时，按照引理中的关键操作，将其更新成一个更加平衡的矩阵。这里的更加平衡指的是，相比于初始矩阵 $\boldsymbol{M}_S$，更新后的矩阵，其任意两列之间零的数量之差更接近于 0 或 1。在接下来的小节中，我们将在后面的算法中多次应用引理 2.7，直到输出一个稀疏、良好且平衡的二元矩阵为止，该矩阵可以作为一个 RS 码生成矩阵的支撑矩阵。

令 $n=2k-t$，其中 $t\in[k]$。下面，我们来构造稀疏、良好且平衡的 $k\times n$ 阶二元矩阵。为了方便起见，假设 k 是偶数。k 为奇数时构造方法与偶数类似，因此我们将其简述在文献 [67] 中。对于 $t=1,2$ 时的矩阵，我们将在第2.4.3小节中给出具体的构造。因此在本小节中，我们考虑 $t\in[3,k]$ 的情形，且通过应用引理 2.7，给出三个算法，再由算法输出目标矩阵。

令 $a\triangleq\left\lceil\frac{k(k-1)}{n}\right\rceil$，$b\triangleq\left\lfloor\frac{k(k-1)}{n}\right\rfloor$，即目标矩阵中每列要么含有 a 个 0，要么含有 b 个 0。回顾一下，目标矩阵中含有 a 个 0 的有 $\theta\triangleq k(k-1)\bmod^{+}n$ 列。当 $t\in[3,k]$ 时，令 $t=4m+u$，其中 $u=0,1,2,3$。则有

$$\frac{k(k-1)}{2k-t}=\frac{k(k-\frac{t}{2})+(\frac{t}{2}-1)k}{2(k-\frac{t}{2})}=\frac{k}{2}+\frac{(t-2)k}{4(k-\frac{t}{2})}=\frac{k}{2}+\frac{t-2}{4}+\frac{t^2-2t}{8k-4t}.$$

当 t 满足 $3\leqslant t<\frac{1+\sqrt{8k+1}}{2}$ 时，我们有 $\frac{t^2-2t}{8k-4t}<\frac{1}{4}$。此时 $a=\left\lceil\frac{k(k-1)}{n}\right\rceil=\frac{k+r}{2}$，$b=\left\lfloor\frac{k(k-1)}{n}\right\rfloor=\frac{k+r-2}{2}$，其中 r 的值在表 2.1 中给出。

表 2.1　r 值分布。

	$u=0$	$u=1$	$u=2$	$u=3$
r	$\frac{t}{2}$	$\frac{t-1}{2}$	$\frac{t+2}{2}$	$\frac{t+1}{2}$

1) 当 $\mathbf{t}\in\left[3,\left\lceil\frac{1+\sqrt{8k+1}}{2}\right\rceil-1\right]$

现在我们假设 t 是一个满足 $3\leqslant t<\frac{1+\sqrt{8k+1}}{2}$ 的整数，并以 $u=1$ 为例来说明我们的算法。其余三种情况都是类似的。在这种情况下，$t=4m+1$ 且 $m<\frac{\sqrt{8k+1}-1}{8}$，

$a=\frac{k+2m}{2}, b=\frac{k+2m-2}{2}$ 以及 $\theta=\frac{3k}{2}+4m^2-3m-1<n$。

给定初始矩阵 $M_S=[M_1;M_2]$, 其中 M_1 和 M_2 都是 $\frac{k}{2}\times n$ 阶循环矩阵。M_1 的第一行对应集合 $[k-1]$,M_2 的第一行对应集合 $[k-3m+1, 2k-3m-1] \bmod^+ n=[k-3m+1,n]\cup[m]$。图 2.3刻画了 M_S 中 0 的位置, 其中 0 都在实线封闭区域内。容易看出 M_S 是稀疏良好的,且可以生成一棵二叉树 $(\frac{k}{2};\frac{k}{2}-1,\frac{k}{2}+1;\frac{k}{2}-2,\frac{k}{2}+2;\cdots;1,k-1)$, 但并不平衡。我们将应用引理 2.7 中的关键操作来调整 M_S 中 0,1 的位置，使其成为一个平衡矩阵。具体构造在算法 2.1中给出，且在图 2.4中对算法执行过程进行了说明。

算法 2.1　参数为 $t=4m+1$，且 k 是偶数的矩阵 M_S 的构造。

Input: 正整数 k,m，满足 $1\leqslant m<\frac{\sqrt{8k+1}-1}{8}$ 且 k 是偶数;

Output: 一个稀疏、良好且平衡的二元矩阵 M_S。

1 计算 $t=4m+1, n=2k-t$。构造初始矩阵 $M_S=(m_{i,j})$ 如下：当 $i\in[\frac{k}{2}]$ 且 $j\in[i,i+k-2]$ 时，$m_{i,j}=0$；当 $i\in[\frac{k}{2}+1,k], j\in[m-1+i-\frac{k}{2}]\cup[k-3m+i-\frac{k}{2},n]$ 时，$m_{i,j}=0$；其余情况，$m_{i,j}=1$。

2 更新 0 的位置。

3 当 $i\in[\frac{k}{2}-m+2,\frac{k}{2}], j\in[i-(\frac{k}{2}-m+1)]$ 以及 $j'\in[\frac{3k}{2}-2m,\frac{3k}{2}-2m+i-(\frac{k}{2}-m+2)]$ 时，令 $m_{i,j}=0$，$m_{i,j'}=1$。

4 当 $i\in[\frac{k}{2}-2m+2,\frac{k}{2}-m], j\in[\frac{3k}{2}-3m,\frac{3k}{2}-3m+i-(\frac{k}{2}-2m+2)]$ 以及 $j'\in[\frac{3k}{2}-2-(i-(\frac{k}{2}-2m+2)),\frac{3k}{2}-2]$ 时，令 $m_{i,j}=1$，$m_{i,j'}=0$。

5 当 $i\in[\frac{k}{2}-m+1,\frac{k}{2}], j\in[\frac{3k}{2}-4m^2+m-1,\frac{3k}{2}-2m-1]$ 以及 $j'\in[\frac{3k}{2}-1,\frac{3k}{2}+4m^2-3m-1]$ 时，令 $m_{i,j}=1$，$m_{i,j'}=0$。

6 当 $i\in[\frac{k}{2}-m+2,\frac{k}{2}], j\in[k+2m-1,\frac{3k}{2}-4m^2+m-2]$ 以及 $j'\in[\frac{3k}{2}+4m^2-3m,n]$ 时，令 $m_{i,j}=1$，$m_{i,j'}=0$。

7 当 $i\in[\frac{k}{2}+1,\frac{k}{2}+m-1], j\in[i,\frac{k}{2}+m-1]$ 以及 $j'\in[\frac{3k}{2}-4m+(i-\frac{k}{2}),\frac{3k}{2}-3m-1]$ 时，令 $m_{i,j}=0$，$m_{i,j'}=1$。

8 当 $i\in[\frac{k}{2}+2,\frac{k}{2}+m], j\in[k-3m+1,k-3m+(i-\frac{k}{2}-1)]$ 以及 $j'\in[\frac{3k}{2}-4m,\frac{3k}{2}-4m-1+(i-\frac{k}{2}-1)]$ 时，令 $m_{i,j}=0$，$m_{i,j'}=1$。

9 对任意两列 j,j' 满足 $j\in[\frac{k}{2}+m,k-3m]$ 以及 $j'\in[k-2m+1,\frac{3k}{2}-4m-1]$。找出一行 $i\geqslant\frac{k}{2}+1$ 满足 $m_{i,j}=1$ 以及 $m_{i,j'}=0$，将这两个位置上的 0,1 互换，即: $m_{i,j}=0$ 以及 $m_{i,j'}=1$。

10 重复第 9步，直至所有在 $[\frac{k}{2}+m,k-3m]\cup[k-2m+1,\frac{3k}{2}-4m-1]$ 中的每一列均恰好含有 $\frac{k+2m}{2}$ 个 0。

11 Return M_S;

下面，沿用图 2.4中的记号，我们来详细阐述算法 2.1的核心思想。在初始矩阵 M_S 中，$[A,B]$ 中的每一列已经恰好有 a 个 0。在算法 2.1执行完第3-6步后，所有在 $[\mathbf{B}]\cup[\mathbf{J},\mathbf{O}]$ 中的列都含有 a 个 0，而在 $[\mathbf{O}+1,n]$ 中的列均含有 b 个 0。注意到 $|[\mathbf{O}+1,n]|=\frac{k}{2}-4m^2-m=n-\theta$，故这些是包含 b 个 0 的所有列，接下来我们只需要使剩余所有列都包含 a 个 0 即可。因此，对于第 j 列，其中 j 位于 $[\mathbf{E}+2,\mathbf{F}]$ 或 $[\mathbf{F}+1,\mathbf{G}-1]$ 或 $[\mathbf{G},\mathbf{H}]$，或 $[\mathbf{H}+1,\mathbf{J}-1]$，我们需要从该列中分别删除 $j-(k-2m),2m-1,m,m-1$ 个 0。而对于第 $(\mathbf{E}+1)$ 列，则不需要调整，因为

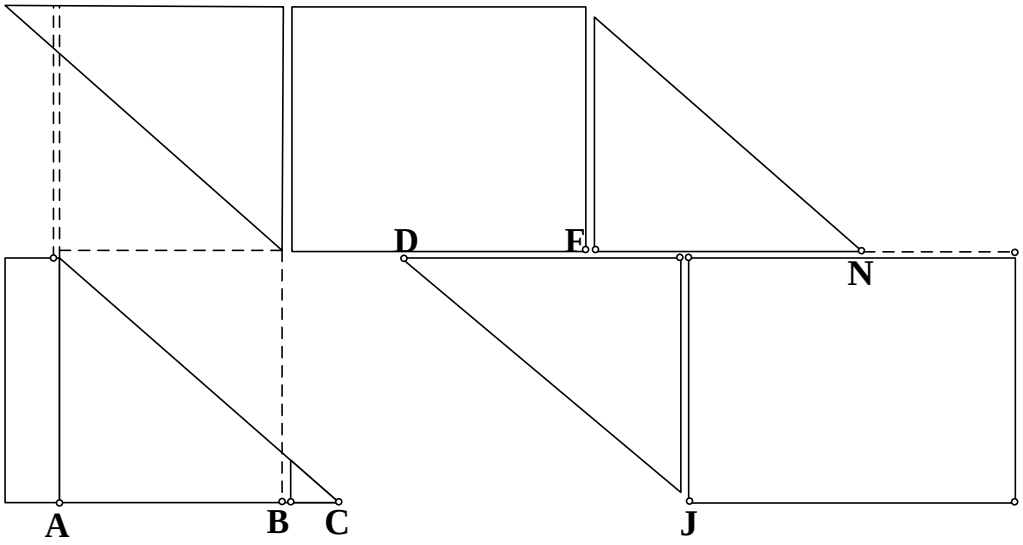


图 2.3　初始矩阵 M_S。其中各列顶点的坐标如下：$\mathbf{A}=m$, $\mathbf{B}=\frac{k}{2}$, $\mathbf{C}=\frac{k}{2}+m-1$, $\mathbf{D}=k-3m+1$, $\mathbf{F}=k-1$，$\mathbf{J}=\frac{3k}{2}-3m$，$\mathbf{N}=\frac{3k}{2}-2$。此外，两个相邻点的坐标相差 1，本小节中，下面所有图都是如此。

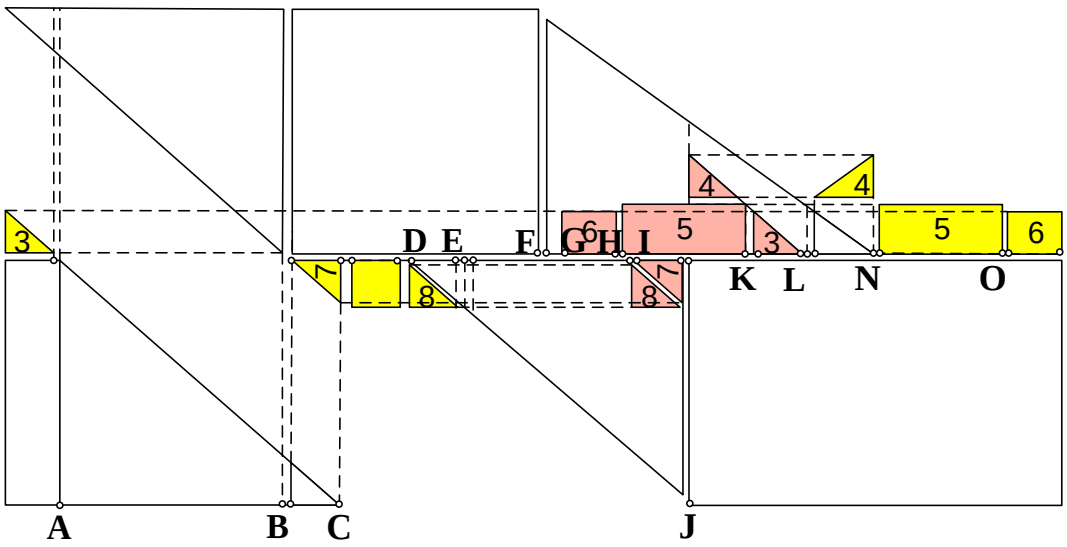


图 2.4　算法2.1执行过程图示。其中，数字对应于算法中的具体步骤。特别地，对于标有相同数字的区域，红色表示该步骤进行前矩阵中 0 的位置，而黄色表示该步骤完成后 0 的位置。例如，在算法2.1的第3步中，标有数字 3 的红色区域内的 0 将会被移至标有数字 3 的黄色区域。其中各列顶点的坐标如下：$\mathbf{A}=m$, $\mathbf{B}=\frac{k}{2}$, $\mathbf{C}=\frac{k}{2}+m-1$, $\mathbf{D}=k-3m+1$, $\mathbf{E}=k-2m-1$, $\mathbf{F}=k-1$, $\mathbf{G}=k+2m-1$, $\mathbf{H}=\frac{3k}{2}-4m^2+m-2$, $\mathbf{I}=\frac{3k}{2}-4m+1$, $\mathbf{J}=\frac{3k}{2}-3m$, $\mathbf{K}=\frac{3k}{2}-2m-1$, $\mathbf{L}=\frac{3k}{2}-m-2$，$\mathbf{N}=\frac{3k}{2}-2$，$\mathbf{O}=\frac{3k}{2}+4m^2-3m-1$。

它已经有 a 个 0。而第7 和 8步进一步使位于 $[\mathbf{B}+1,\mathbf{C}]\cup[\mathbf{D},\mathbf{E}]\cup[\mathbf{I}-1,\mathbf{J}-1]$ 中的列含有 0 的数量达到 a。

步骤 3-8 是具体的, 即给出了 0, 1 调整前和调整后的确切位置。而算法 2.1中第9和10步是不确定性操作, 0, 1 最终位置并不唯一。观察到, 在 $[\mathbf{C}+1,\mathbf{D}-1]$ 中的每一列都需要有 m 个 0 要移入, 而对于在 $[\mathbf{E}+2,\mathbf{F}]$ 或 $[\mathbf{F}+1,\mathbf{G}-1]$ 或 $[\mathbf{G},\mathbf{H}]$, 以及 $[\mathbf{H}+1,\mathbf{I}-2]$ 中的每一列, 记为第 j 列, 分别需要移出 $j-(k-2m),2m-1,m,m-1$ 个 0。因此根据目标矩阵的性质，我们只需要检查这些移入的 0 的数量是否与移出的 0 的数量相同，即可说明第9和10步的可行性。注意到需要移入的 0 个数是 $(\frac{k}{2}-4m+1)m=\frac{k}{2}m-4m^2+m$，而需要移出 0 的数量是

$$m(2m-1)+(2m-1)(2m-1)+(\frac{k}{2}-4m^2-m)m+(4m^2-4m+1-m)(m-1)=\frac{k}{2}m-4m^2+m.$$

因此算法 2.1中的第 9和10步是可行的, 最终可输出一个稀疏平衡的二元矩阵。由于初始矩阵 M_S 是良好的，并且算法 2.1中的所有步骤都满足引理 2.7的关键操作，所以最终的新矩阵 M_S 仍然对应一个良好的二叉树 $(\frac{k}{2};\frac{k}{2}-1,\frac{k}{2}+1;\frac{k}{2}-2,\frac{k}{2}+2;\cdots;1,k-1)$。

接下来，我们考虑 $t\geqslant\frac{1+\sqrt{8k+1}}{2}$，并给出两种算法来构造不同范围 t 的 M_S。事实上，这两种算法是算法 2.1 的推广，但其中一种不再是确定性算法，原因是 θ 和 a 的准确值无法确定。

2) 当 $\mathbf{t}\in\left[\left\lceil\frac{\mathbf{1+\sqrt{8k+1}}}{\mathbf{2}}\right\rceil,\left\lfloor\mathbf{k}-\sqrt{\mathbf{k}}\right\rfloor\right]$

在这种情况下, 如果 $t\leqslant\frac{k}{2}+2$, 则 $a-1<\frac{k}{2}+a-t+2$, 否则 $a-1\geqslant\frac{k}{2}+a-t+2$。且当 $t=k-\sqrt{k}$ 是一个整数时，有 $a=t$ 以及 $\theta=n$；其余情况下，$a-t\geqslant1$。给定初始矩阵 $M_S=[M_1;M_2]$，如图 2.5所示，其中 M_1,M_2 都是 $\frac{k}{2}\times n$ 阶矩阵，定义如下。为方便起见，我们遵循图 2.5 中的符号。矩阵 M_1 是从一个第一行对应于集合 $[k-1]$ 的循环矩阵中，将该矩阵从第 $\mathbf{K}$ 列开始的右下角区域的 0 平移到矩阵的最左边而得到的。矩阵 M_2 是一个循环矩阵，更具体地，如果 $\theta\geqslant\frac{n}{2}$，则其第一行对应于集合 $[\frac{k}{2}+a-t+2,\frac{3k}{2}+at]\bmod^+n=[\mathbf{D},n]\cup[\mathbf{A}+1]$；如果 $\theta<\frac{n}{2}$，则对应于集合 $[\frac{k}{2}+a-t+1,\frac{3k}{2}+at-1]\bmod^+n=[\mathbf{D}-1,n]\cup[\mathbf{A}]$。这里，图 2.5 是针对 $\theta\geqslant\frac{n}{2}$ 以及 $t\leqslant\frac{k}{2}+2$ 的情况。

本小节中, 我们只考虑 $\theta\geqslant\frac{n}{2}$ 的情况, 其余情况类似。容易验证初始矩阵 M_S 是稀疏且良好的, 并且可以生成一棵二叉树 $(\frac{k}{2};\frac{k}{2}-1,\frac{k}{2}+1;\frac{k}{2}-2,\frac{k}{2}+2;\cdots;1,k-1)$。令 II、III、IV、VI、VII 表示图 2.5中相应的由 0 组成的灰色区域，令 I、V 表示由 1 组成的区域。通过一些计算，我们知道在初始矩阵 M_S 中，$[\mathbf{B}+1,\mathbf{E}]\cup[\mathbf{K},n]$ 中的每一列最多有 b 个 0，而 $[\mathbf{B}]\cup[\mathbf{E}+1,\mathbf{J}]$ 中的每一列至少有 a 个 0。故算法 2.2 旨在仔细地将多余的 0 从区域 II、III 以及可能从 IV 移至区域 I（第 3步），并从区

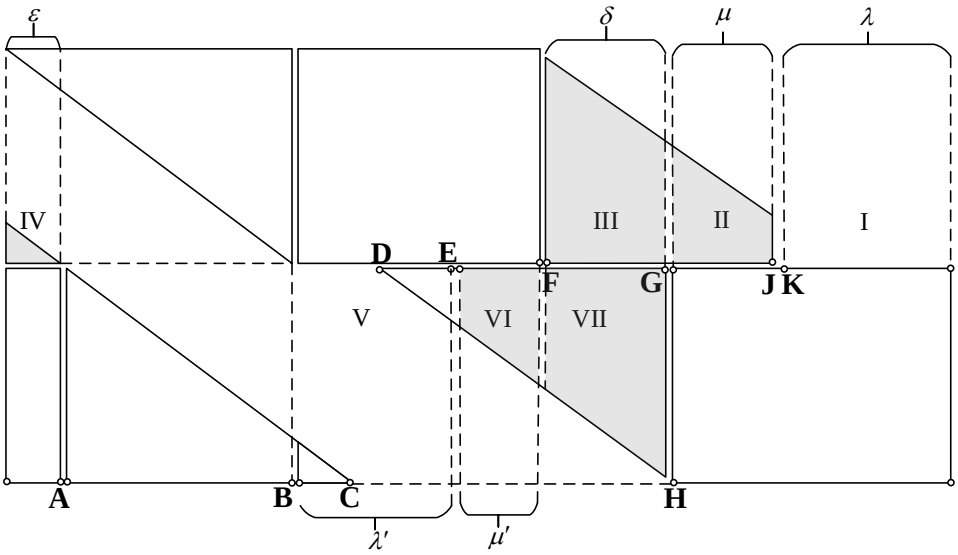


图 2.5　算法2.2中当 $\theta \geqslant \frac{n}{2}$ 以及 $t \leqslant \frac{k}{2}+2$ 时的初始矩阵，其中 λ, λ'，以及 $\varepsilon, \mu, \mu', \delta$ 分别表示相应列中缺失的 0 和多出的 0 的总数。图中各点的列坐标如下：$\mathbf{A} = a - \frac{k}{2} - 1$，$\mathbf{B} = \frac{k}{2}$，$\mathbf{C} = a-1$，$\mathbf{D} = \frac{k}{2}+a-t+2$，$\mathbf{E} = 2a-t$，$\mathbf{F} = k$，$\mathbf{G} = k+a-t$，$\mathbf{H} = k+a-t+1$，$\mathbf{J} = 2k-a-1$，$\mathbf{K} = 2k-a$。

域 VI、VII 中将多余的 0 移到区域 V(第5步)，从而使新得到的矩阵达到平衡。注意到，第3步是应用了引理 2.7 当 $\beta = \frac{k}{2}, s = 1$ 时来更新 $\{S_1, \cdots, S_\beta\}$ 的，而第 5步是取 $s = k-1$ 来更新 $\{S_{\beta+1}, \cdots, S_k\}$ 的。算法 2.2中的第 6至8步是针对 $\theta < \frac{n}{2}$ 的情形。

在算法2.2中，有两种情况需要缜密的处理。第一种是，在迭代第3 步时，可能会出现 $[\mathbf{B}] \cup [\mathbf{K}, n]$ 中的每一列已经有 a 或 b 个 0，但 $[\mathbf{H}, \mathbf{J}]$ 中仍有一些列有多余 0 的情况；这些多余的 0 不能直接移到集合 $[\frac{k}{2}]$ 所有行中的任何其他列，否则会使得 $[\mathbf{B}] \cup [\mathbf{K}, n]$ 中的列含有多余的 0；并且它们也不能移到集合 $[\frac{k}{2}+1, k]$ 所有行中的任何其他列，否则可能会破坏底部块的良好性，因为这种操作未遵循引理 2.7。另一种情况是 $[\mathbf{A}] \cup [\mathbf{F}, \mathbf{J}]$ 列中多出 0 的数量可能不足以补偿在 $[\mathbf{K}, n]$ 这些列里所缺少的 0。幸运的是，我们可以证明这两种情况都不会发生，因此算法2.2 中的第3–4步可行，且在执行完毕之后移至第5步。

引理 2.8　算法 2.2 是可行的，且将在有限次迭代后输出结果。

证明　我们只考虑 $\theta \geqslant \frac{n}{2}$ 的情况，此时 $k-2a+t-3 \geqslant 0$，其余情况类似。第 2步是初始化矩阵 M_S。可以看出，如果第 3–4步可以迭代结束，那么第 5步也可以终止。因此只需要说明第3–4 步可以迭代结束即可。注意到，在第 3–4步中，我们希望将区域 IV、III 和 II 中多出的 0 移至区域 I。

根据目标矩阵以及初始矩阵 M_S 的性质可知，在不考虑良好性的前提下，M_S 可以通过在同一行内调整 0, 1 的位置，得到一个平衡矩阵，记为 M'_S，它在

算法 2.2　参数为 $t \in \left[\left\lceil \frac{1+\sqrt{8k+1}}{2} \right\rceil, \left\lfloor k-\sqrt{k} \right\rfloor\right]$，且 k 是偶数的矩阵 M_S 的构造。

Input: 正整数 k,t，满足 $t \in \left[\left\lceil \frac{1+\sqrt{8k+1}}{2} \right\rceil, \left\lfloor k-\sqrt{k} \right\rfloor\right]$ 且 k 是偶数；

Output: 一个稀疏、良好且平衡的二元矩阵 M_S。

1 计算 $n=2k-t$，$a=\left\lceil \frac{k(k-1)}{2k-t} \right\rceil, b=\left\lfloor \frac{k(k-1)}{2k-t} \right\rfloor$，以及 $\theta \triangleq k(k-1) \bmod^{+} n$。

2 构造初始矩阵 $M_S=(m_{i,j})$ 如下：若 $\theta \geqslant \frac{n}{2}$，则当 $i \in [k-a+1]$ 且 $j \in [i, i+k-2]$ 时，有 $m_{i,j}=0$，当 $i \in [k-a+2, \frac{k}{2}]$ 且 $j \in [i-(k-a+1)] \cup [i, 2k-a-1]$ 时，有 $m_{i,j}=0$，当 $i \in [\frac{k}{2}+1, k], j \in [a-\frac{k}{2}-1+i-\frac{k}{2}] \cup [\frac{k}{2}+a-t+1+i-\frac{k}{2}, n]$ 时，有 $m_{i,j}=0$，其余情况，$m_{i,j}=1$；完成初始化后，转至第3步。若 $\theta < \frac{n}{2}$，则当 $i \in [k-a+2]$ 且 $j \in [i, i+k-2]$ 时，有 $m_{i,j}=0$，当 $i \in [k-a+3, \frac{k}{2}]$ 且 $j \in [i-(k-a+2)] \cup [i, 2k-a]$ 时，有 $m_{i,j}=0$，当 $i \in [\frac{k}{2}+1, k], j \in [a-\frac{k}{2}-2+i-\frac{k}{2}] \cup [\frac{k}{2}+a-t+i-\frac{k}{2}, n]$ 时，有 $m_{i,j}=0$，其余情况，$m_{i,j}=1$；完成初始化后，转至第 6 步。

3 对于任意两列 j_1, j_2 满足 $j_1 \in [a-\frac{k}{2}-1] \cup [k, 2k-a-1]$ 以及 $j_2 \in [2k-a, n]$，选择尽可能大的 j_1，在 $[2, \frac{k}{2}]$ 中找到一行 i 满足 $m_{i,j_1}=0$ 和 $m_{i,j_2}=1$，且若 $j_1 \in [a-\frac{k}{2}-1]$，则 $i \geqslant k-a+1+j_1$，将这两个位置上的 0, 1 互换，即：$m_{i,j_1}=1$ 以及 $m_{i,j_2}=0$。

4 重复第 3步直到满足以下条件：集合 $[\frac{k}{2}] \cup [k+a-t+1, n]$ 中的所有列都有 a 或 b 个 0；这些列中包含 a 个 0 的列数尽可能大，但不超过 θ；集合 $[k, k+a-t]$ 中的所有列都至少有 b 个 0；集合 $[\frac{k}{2}] \cup [k, n]$ 中包含 b 个 0 的列数不超过 $n-\theta$。

5 对于集合 $[\frac{k}{2}+1, k+a-t]$ 的列和集合 $[\frac{k}{2}+1, k]$ 的行，不断调整这些区域中 0 的位置直到矩阵中所有列都有 a 或 b 个 0，然后转至第9步。

6 对于任意两列 j_1, j_2 满足 $j_1 \in [k, 2k-a]$ 和 $j_2 \in [\frac{k}{2}-2] \cup [2k-a+1, n]$，选择尽可能大的 j_1，在 $[2, \frac{k}{2}]$ 中找到一行 i 满足 $m_{i,j_1}=0$ 和 $m_{i,j_2}=1$，且若 $j_2 \in [\frac{k}{2}-2]$，则 $i \geqslant j_2+2$，将这两个位置上的 0, 1 互换，即：$m_{i,j_1}=1$ 和 $m_{i,j_2}=0$。

7 重复第6步直到满足以下条件：集合 $[k+a-t, n]$ 中的所有列都有 a 或 b 个 0；集合 $[k, k+a-t-1]$ 中的列都至少有 b 个 0；集合 $[k, n]$ 中包含 b 个 0 的列数尽可能多但不超过 $n-\theta-\frac{k}{2}$。

8 对于集合 $[\frac{k}{2}+1, k+a-t-1]$ 的列和集合 $[\frac{k}{2}+1, k]$ 的行，不断调整这些区域中 0 的位置直到矩阵中所有列都有 a 或 b 个 0，然后转至第9步。

9 Return M_S；

$[\mathbf{A}+1, \mathbf{B}]$ 的每一列中都含有 a 个 0。注意到 M'_S 并不唯一，因为具体哪些列含有 a 个 0 是随机的。通过与矩阵 M'_S 比较，可定义 M_S 的每一列中多出（缺失）0 的数量。记 M_S 中集合 $[\mathbf{A}], [\mathbf{E}+1, \mathbf{F}-1], [\mathbf{F}, \mathbf{G}]$ 和 $[\mathbf{H}, \mathbf{J}]$ 中的列总共多出 0 的数量分别为 $\varepsilon, \mu', \delta$ 以及 μ，而集合 $[\mathbf{B}+1, \mathbf{E}]$ 和 $[\mathbf{K}, n]$ 中的列总共缺失 0 的数量分别为 λ', λ，这些 0 的分布具体可见图 2.5。由于最终的平衡矩阵 M'_S 不唯一，这些 $\varepsilon, \mu', \delta, \mu$ 以及 λ', λ 的值也不唯一。但是有 $\varepsilon+\mu'+\delta+\mu=\lambda'+\lambda$ 成立。因此，通过考虑最后目标矩阵具有 a 或 b 个 0 的相应列，可以得到 $\varepsilon, \delta, \lambda, \lambda', \mu'$ 和 μ 的上下界，如下。更具体地，以 δ 为例，可假设 M'_S 在集合 $[\mathbf{F}, \mathbf{G}]$ 中的列均含有 b（或 a）个 0，则可得到相应的 δ 的上界（或下界）。

$$0 \leqslant \varepsilon \leqslant a-\frac{k}{2}-1,$$

$$w_1 := (k-2a+t-2)(a-t+1) \leqslant \delta \leqslant (k-2a+t-1)(a-t+1) := w_2,$$

$$x_1 := (a-\frac{k}{2}-1)(a-t+1) \leqslant \lambda \leqslant (a-\frac{k}{2})(a-t+1) := x_2,$$

$$y_1 := (a-\frac{k}{2}-1)(a-t) \leqslant \lambda' \leqslant (a-\frac{k}{2})(a-t+1) := y_2,$$

$$z_1 := \frac{(k-2a+t-2)(k-2a+t-1)}{2} \leqslant \mu, \mu' \leqslant \frac{(k-2a+t-1)(k-2a+t)}{2} := z_2.$$

断言 2.9　$\mu \leqslant \lambda$。

断言2.9的证明　反证法。假设 $\mu > \lambda$, 由 $\varepsilon+\mu'+\delta+\mu = \lambda'+\lambda$ 可知, $\lambda' > \mu'$。进一步地，$\mu > \lambda$ 意味着 $z_2 > x_1$，则有

$$\begin{aligned}\lambda'-\mu' &\leqslant y_2 - z_1 = (a-\frac{k}{2})(a-t+1) - \frac{(k-2a+t-2)(k-2a+t-1)}{2}\\ &= x_1 - z_2 + a - t + 1 + k - 2a + t - 1\\ &\leqslant a - t + k - 2a + t - 1 = (a-t+1) + (k-2a+t-2).\end{aligned}$$

若 $a = t$，则有 $\theta = n$，这意味着 M'_S 中的每一列都有 a 个 0。因此有 $z_1 = \mu = \mu' < \lambda' = \lambda = x_2 = y_2$, 矛盾。故可设 $a \neq t$, 此时有 $a-t+1 \geqslant 2$、$k-2a+t-2 \geqslant 1$ 以及 $\delta+\mu-\lambda \geqslant w_1+1 = (a-t+1)(k-2a+t-2)+1$。结合 $\lambda'-\mu'$ 的上界以及 $\lambda'-\mu' = \varepsilon+\delta+\mu-\lambda$, 可知 $k-2a+t-2 = 1$。在这种情况下有 $\varepsilon = 0, \lambda'-\mu' = a-t+2$, $\delta = a-t+1$ 以及 $\mu-\lambda = 1$，这意味着 $\lambda' = y_2, \mu' = z_1, \mu = z_2, \lambda = x_1$。基于此，我们可以将 M_S 更新至一个新的稀疏平衡矩阵 M''_S，它在集合 $[\mathbf{B}+1, \mathbf{G}]$ 中存在一列含有 b 个 0，除此以外，在集合 $[\mathbf{G}]$ 的每一列中都有 a 个 0；并且在集合 $[\mathbf{H}, n]$ 中只有一列含有 a 个 0, 其余均含有 b 个 0。此时，可将 M'_S 替换为 M''_S，则这种情况可以推导出 $\mu = \lambda$。 ∎

断言 2.10　$\delta+\mu+a-\frac{k}{2}-1 \geqslant \lambda$。

断言2.10的证明　$a = t$ 的情况可类似于断言2.9中的证明导出矛盾。因此我们只考虑 $a \neq t$。反证法。假设 $\delta+\mu+a-\frac{k}{2}-1 < \lambda$，则由 $\delta+\mu+\varepsilon-\lambda = \lambda'-\mu'$ 以及 $\varepsilon \leqslant a-\frac{k}{2}-1$ 可知，$\lambda' < \mu'$。因此，$z_2 \geqslant y_1+1$，即

$$z_1 + k - 2a + t - 1 = z_2 \geqslant y_1 + 1.$$

则有

$$\begin{aligned}&\delta+\mu+a-\frac{k}{2}-1 \geqslant w_1+z_1+a-\frac{k}{2}-1\\ \geqslant& w_1+y_1+1-(k-2a+t-1)+a-\frac{k}{2}-1\\ =&(k-2a+t-2)(a-t+1)-(k-2a+t-2)+(a-\frac{k}{2}-1)(a-t+1)\\ =&x_2+(k-2a+t-2)(a-t+1)-(k-2a+t-2)-(a-t+1).\end{aligned}$$

这意味着

$$\begin{aligned}&x_2+(k-2a+t-2)(a-t+1)-(k-2a+t-2)-(a-t+1)\\ \leqslant&\delta+\mu+a-\frac{k}{2}-1<\lambda\leqslant x_2.\end{aligned}$$

则 $k-2a+t-3=0$, 此时有 $\delta=a-t+1, \mu'=z_2=y_1+1, \mu=z_1, \lambda'=y_1$, $\lambda=x_2$ 以及 $\delta+\mu+a-\frac{k}{2}-1=x_2-1$。因此我们可将 $M_{\mathcal{S}}$ 更新至一个稀疏平衡的新矩阵 $M_{\mathcal{S}}'''$, 它在集合 $[\mathbf{F},n]$ 中存在一列含有 b 个 0, 除此以外, 在集合 $[\mathbf{A}+1,\mathbf{B}]\cup[\mathbf{F},n]$ 中的每一列均含有 a 个 0; 并且在集合 $[\mathbf{A}]\cup[\mathbf{B}+1,\mathbf{F}-1]$ 中只有一列含有 a 个 0, 其余列均含有 b 个 0。此时, 可将 $M_{\mathcal{S}}'$ 替换为 $M_{\mathcal{S}}'''$, 则这种情况可以推导出 $\delta+\mu+a-\frac{k}{2}-1=\lambda$。 ∎

断言 2.11　第3–4步是可行的。

断言2.11的证明　首先, 根据断言2.9知 $\mu\leqslant\lambda$, 故我们可以将区域 II 中所有多出的 0 移到区域 I, 以便集合 $[\mathbf{H},\mathbf{J}]$ 中的每一列都是平衡的（即, 具有 a 或 b 个 0)。由于 μ 和 λ 的确切值是基于某个平衡矩阵得到的, 这保证了在这些移动之后, 整个矩阵中包含 a 个 0 的列数不超过 θ。

接下来, 有两种情形需要考虑。在每种情况下, 调整 0 的位置的同时, 都要确保矩阵中已经调整过的列中, 含有 a 个 0 的列数不会超过 θ。

如果 $\delta+\mu\geqslant\lambda$, 那么进一步地, 可将区域 III 中多出的 0 移动到 I 中, 使得在集合 $[\mathbf{H},n]$ 中的所有列达到平衡, 而无需从区域 IV 中移动 0, 并且所有在集合 $[\mathbf{F},\mathbf{G}]$ 中的列, 在调整后, 都至少有 b 个 0。通过控制从区域 III 移动到 I 的 0 的数量, 集合 $[\mathbf{F},n]$ 中包含 b 个 0 的列数可以很容易地限制为不超过 $n-\theta$。

如果 $\delta+\mu<\lambda$, 那么由断言2.10知 $\delta+\mu+a-\frac{k}{2}-1\geqslant\lambda$。此时, 区域 II 和 III 中多出的 0 不足以补偿区域 I 中缺少的 0。故我们需要从 IV 中移出 ε 个 0 到区域 I。注意到, 由于区域 IV 中的每列恰好含有 a 个 0, 我们每次至多能从一列中移出 1 个 0。反证法。假设第3–4 步不可行, 即在我们将 ε 个 0 从区域 IV 移动到 I 之后, 集合 $[\mathbf{A}]\cup[\mathbf{F},n]$ 中已经存在 $n-\theta$ 列恰好包含 b 个 0, 但集合 $[\mathbf{K},n]$ 中的某些列仍然不平衡。即 $\delta+\mu+\varepsilon<\lambda$, 因此 $\lambda'<\mu'$。进一步可知, 集合 $[\mathbf{B}+1,\mathbf{F}-1]$ 中的所有列最后都必须有 a 个 0, 故有 $\lambda'=y_2=x_2$ 和 $\mu'=z_1$。所以 $z_1>y_2=x_2$, 即 $\mu>\lambda$, 这与断言2.9相矛盾。 ∎

由于第3步是单向移动, 因此第3–4步可以在有限多次迭代后完成。故证明完成。 ∎

3) 当 $\mathbf{t}\in\left[\left\lfloor\mathbf{k}-\sqrt{\mathbf{k}}\right\rfloor,\mathbf{k}\right]$

在这种情况下 $a = t$，我们给出一个比算法2.2更具体的构造，见算法 2.3。当 $\sqrt{k}$ 是整数时，此时 $t = k - \sqrt{k}$ 的情况已经包含在 **2)** 中。因此，在本小节剩余部分，我们假设 $\sqrt{k}$ 不是整数。当 $t = k$ 时，矩阵 M_S 可以取单位矩阵。令 $u = k - t$，则有 $1 \leqslant u \leqslant \left\lfloor \sqrt{k} \right\rfloor$ 以及 $n = k + u$。由于

$$\frac{k(k-1)}{k+u} = k - \frac{k(u+1)}{k+u} = k - u - 1 + \frac{u(u+1)}{k+u},$$

我们有 $a = k - u$，$b = k - u - 1$ 以及 $\theta = u(u+1)$。与前面的算法不同，算法 2.3 的初始矩阵是一个循环的 $k \times n$ 阶矩阵，其第一行对应向量 $(\underbrace{0, \cdots, 0}_{k-1}, 1, \cdots, 1)$。当 $u = 1$ 时，初始矩阵本身满足稀疏平衡性，故设 $u \geqslant 2$。

算法 2.3　参数为 $\left\lceil k - \sqrt{k} \right\rceil \leqslant t \leqslant k - 1$ 的矩阵 M_S 的构造

Input: 正整数 k, u，满足 $2 \leqslant u \leqslant \left\lfloor \sqrt{k} \right\rfloor$，且 $\sqrt{k}$ 不是整数;

Output: 一个稀疏、良好且平衡的二元矩阵 M_S。

1 计算 $n = k + u, a = k - u, b = k - u - 1$ 以及 $\nu = \frac{u(u-1)}{2}$。

2 构造初始矩阵 $M_S = (m_{i,j})$ 如下：当 $i \in [k]$ 且 $j \in [i, i + k - 2] \bmod^+ n$ 时，$m_{i,j} = 0$；其余情况，$m_{i,j} = 1$。

3 对任意两列 s_1, s_2 满足 $s_1 \in [\nu]$ 以及 $s_2 \in [k, n-2]$，在集合 $[3, \nu + 2]$ 中找到一行 i，使得 i 的取值尽可能大且满足 $s_1 \leqslant i - 2$，以及 $m_{i,s_1} = 1$ 和 $m_{i,s_2} = 0$，将这两个位置上的 0, 1 互换，即：$m_{i,s_1} = 0$ 以及 $m_{i,s_2} = 1$。

4 重复第3步直到集合 $[\nu] \cup [k, n-2]$ 中所有列均恰好含有 a 个 0。

5 对任意两列 s_1, s_2 满足 $s_1 \in [\nu + 1, 2\nu]$ 以及 $s_2 \in [a + 1, k - 1]$，在集合 $[\nu + 3, k]$ 中找到一行 i，满足 $s_1 \geqslant i - u$，以及 $m_{i,s_1} = 1$ 和 $m_{i,s_2} = 0$，将这两个位置上的 0, 1 互换，即：$m_{i,s_1} = 0$ 以及 $m_{i,s_2} = 1$。

6 重复第5步直到矩阵中所有列都有 a 或 b 个 0。

7 Return M_S;

引理 2.12　算法2.3 返回一个良好的矩阵 M_S。

证明　算法 2.3 是按照使集合 $[u(u-1)] \cup [a, n-1]$ 中的所有列都含有 a 个 0 的规则运行的，这是可行的，因为 $\theta = u(u+1) = u(u-1) + n - a$。而其初始矩阵 M_S 是一个循环矩阵，第一行对应集合 $[k-1]$，它可以生成一棵良好的二叉树 $(\nu + 2; \nu + 1, \nu + 3; \nu, \nu + 4; \cdots)$，其中 $\nu = \frac{u(u-1)}{2}$ 在算法 2.3 的第1步中定义。观察矩阵 M_S 可知，当 j 分别在集合 $[a-1], [a, k-1]$ 以及集合 $[k, n]$ 时，M_S 的第 j 列分别含有 b, j 以及 $2k - j - 1$ 个 0。注意到，集合 $[a + 1, k - 1]$ 中的列中总共有 ν 个多出的 0，集合 $[k, n-2]$ 中的列也是如此。

在第3–6步中，我们将集合 $[a + 1, n - 2]$ 中的列多出的 0 移至集合 $[2\nu]$ 中的列。注意到，每一列最多可以被移进一个 0。由于 $2\nu \leqslant a$，第4步和第 6步将会在 ν 次迭代后完成。由引理 2.7可知，最终得到的新矩阵 M_S 仍然可以生成一棵良好的二叉树 $(\nu + 2; \nu + 1, \nu + 3; \nu, \nu + 4; \cdots)$。 ∎

在算法2.1–2.3中，每个初始矩阵 M_S 都是稀疏的且满足良好条件，且每一步

更新都满足引理 2.7，因此在算法中，我们不需要花额外步骤来验证所得的平衡矩阵是否良好。且根据引理 2.8和2.12的证明，以及对算法2.1的分析，可以看出算法 2.1 –2.3 均是关于 k 和 n 是多项式时间的算法。

2.4.3　特殊码长的稀疏平衡的 MDS 码的构造

在上一小节中，我们针对码长 $n = 2k - t$ 且 t 满足 $t \in [3, k]$ 的所有码，给出了稀疏平衡矩阵 $M_{\mathcal{S}}$ 的相应构造。本小节，我们考虑 $t = 0, 1, 2$ 的情况。根据第2.4.1小节的分析可知，当 $n = 2k$ 时，$M_{\mathcal{S}}$ 可通过循环两个特定的向量来生成，具体如下。

构造 2.13　对任意 $n = 2k$ 且 $k \geqslant 4$ 是一个偶数，构造集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 如下：当 $i \in [\frac{k}{2}]$ 时，令 $S_i = [i, i + k - 2]$；当 $i \in [\frac{k}{2} + 1, k]$ 时，令 $S_i = [\frac{k}{2} + i, \frac{3k}{2} + i - 2] \bmod^{+} n$。

命题 2.14　给定一个由构造2.13得到的集族 $\mathcal{S}$，则矩阵 $M_{\mathcal{S}}$ 是稀疏、良好且平衡的。

证明　容易看出集族 $\mathcal{S}$ 是 $(k-1)$-一致的。且当 $j \in [\frac{k}{2} - 1] \cup [k, \frac{3k}{2} - 1] \cup \{2k\}$ 时，$M_{\mathcal{S}}$ 的第 j 列中含有 $\frac{k}{2} - 1$ 个 0，而所有剩余列中，均含有 $\frac{k}{2}$ 个 0。因此，$M_{\mathcal{S}}$ 是稀疏平衡的。

为了说明 $\mathcal{S}$ 是良好的，注意到 $A = S_1 \cap S_2 \cap \cdots \cap S_{k/2} = [\frac{k}{2}, k - 1]$，以及 $B = S_{k/2+1} \cap S_{k/2+2} \cap \cdots \cap S_k = [\frac{3k}{2}, 2k - 1]$，且 A, B 的大小均为 $\frac{k}{2}$。故 $\mathcal{S}$ 是不交的且在 $\frac{k}{2}$ 处可分，它的两个剩余集族分别为 $\mathcal{A} = \{S'_j = S_j \setminus A : j \in [\frac{k}{2}]\}$ 和 $\mathcal{B} = \{S'_j = S_j \setminus B : j \in [\frac{k}{2} + 1, k]\}$。沿用来自 $\mathcal{S}$ 的原始索引，可知 $\mathcal{A}$ 和 $\mathcal{B}$ 可以分别生成两棵良好的二叉树 $(1; 2; \cdots; \frac{k}{2} - 1)$、$(\frac{k}{2} + 1; \frac{k}{2} + 2; \cdots; k - 1)$，其中它们的每个非叶子节点按顺序由第一个分离索引分隔开。结合这两个子树，可知 $\mathcal{S}$ 是良好的，且所对应生成的二叉树为 $(\frac{k}{2}; 1, \frac{k}{2} + 1; 2, \frac{k}{2} + 2; \cdots; \frac{k}{2} - 1, k - 1)$。 ∎

类似的，可根据上述方法，给出相应码长为 $n = 2k - 1, 2k - 2$ 时的构造。

构造 2.15　考虑 $n = 2k - 1$，此时 $\theta = \frac{3k}{2} - 1$ 以及 $a = \frac{k}{2}$。而当 $n = 2k - 2$ 时，有 $\theta = n$ 以及 $a = \frac{k}{2}$。对这两种情况，构造集族 $\mathcal{S} = \{S_1, \cdots, S_k\}$ 如下：当 $i \in [\frac{k}{2}]$ 时，令 $S_i = [i, i + k - 2]$；当 $i \in [\frac{k}{2} + 1, k]$ 时，令 $S_i = [\frac{k}{2} + i - 1, \frac{3k}{2} + i - 3] \bmod^{+} n$。

容易验证构造2.15 中的 $\mathcal{S}$ 是良好的且可以生成二叉树 $(\frac{k}{2}; 1, \frac{k}{2} + 1; 2, \frac{k}{2} + 2; \cdots; \frac{k}{2} - 1, k - 1)$，证明方法类似命题2.14。因此，我们有以下命题。

命题 2.16　给定一个由构造2.15得到的集族 $\mathcal{S}$，则矩阵 $M_{\mathcal{S}}$ 是稀疏、良好且平衡的。

至此，我们构造出了所有 $n \leqslant 2k$ 下的稀疏平衡且良好的二元矩阵，再由定理 2.4可知，当 $q \geqslant n$ 时，相应的稀疏平衡 RS 码存在。即定理 2.5完成。

2.4.4　推广到小域 ($q \geqslant n-1$) 上的 MDS 码

本小节中，我们将着手处理 $q = n-1$ 的情况。令 $a_1, \cdots, a_{n-1}$ 是 $\mathbb{F}_q$ 中 n 个互不相同的元素。回顾一下，$a = \left\lceil \frac{k(k-1)}{n} \right\rceil$ 和 $b = \left\lfloor \frac{k(k-1)}{n} \right\rfloor$ 是目标矩阵每列中 0 的数量。沿用第2.2节中给出的符号，我们将在集合 $[n-1]$ 上找到一个集族 $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$，使得其中存在 a 或 b 个大小为 $k-2$ 的集合，其余的大小均为 $k-1$。根据 $\mathcal{S}$，可以定义一个由多项式 $P_{S_1}, P_{S_2}, \cdots, P_{S_k}$ 组成的序列 $\mathcal{P}$，然后构造矩阵 G 如下，

$$G = \begin{pmatrix} p_{1,0} & p_{1,1} & \cdots & p_{1,k-1} \\ p_{2,0} & p_{2,1} & \cdots & p_{2,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,0} & p_{k,1} & \cdots & p_{k,k-1} \end{pmatrix} \begin{pmatrix} a_1^{k-1} & a_2^{k-1} & \cdots & a_{n-1}^{k-1} & 1 \\ a_1^{k-2} & a_2^{k-2} & \cdots & a_{n-1}^{k-2} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_1^0 & a_2^0 & \cdots & a_{n-1}^0 & 0 \end{pmatrix} \tag{2.1}$$

$$\triangleq C(\mathcal{P}) \cdot V' = [C(\mathcal{P})A \ \ \boldsymbol{w}], \tag{2.2}$$

其中矩阵 A 由矩阵 V' 的前 $n-1$ 列组成，$\boldsymbol{w} = (p_{1,0}, p_{2,0}, \cdots, p_{k,0})^T$。首先，我们证明 G 总是稀疏的，即 G 的每一行都有 $k-1$ 个 0。当 $|S_i| = k-1$ 时，第 i 行对应有 $k-1$ 个 0。而当 $|S_i| = k-2$ 时，多项式 P_{S_i} 的次数为 $k-2$，因此 $p_{i,0} = 0$，这意味着 G 的第 i 行也有 $k-1$ 个 0，故证明完毕。其次，我们需要 G 是平衡的，即 G 的每一列都有 a 或 b 个 0。这对于最后一列 $\boldsymbol{w}$ 显然是成立的，因为 $\mathcal{P}$ 中存在 a 或 b 个度为 $k-2$ 的多项式。故我们只需使得矩阵 $C(\mathcal{P})A$ 平衡即可，或者等价地，$k \times n$ 阶矩阵 $M_{\mathcal{S}}$ 平衡。最后，为了使 G 可以作为一个 $[n,k]_q$ MDS 码的生成矩阵，即 G 的任何 k 列都线性无关，类似地，只需矩阵 $C(\mathcal{P})$ 可逆即可。

综上所述，我们需要在集合 $[n-1]$ 上找到一个集族 $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$，使得其中存在 a 或 b 个大小为 $k-2$ 的集合，其余的大小均为 $k-1$，且满足 $M_{\mathcal{S}}$ 是平衡的以及 $C(\mathcal{P})$ 是可逆的。事实上，这样的集族 $\mathcal{S}$，可以通过从一个良好且平衡的集合 $[n]$ 上 $(k-1)$-一致的集族 $\mathcal{S}' = \{S_1', \cdots, S_k'\}$ 中删除一个恰当的元素来得到，而这种集族在前面的小节中已经有相应的构造。具体的操作如下，这里我们省略证明。

注 3　(1) 给定一个集族 $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$，其任意一个集合 S_i 的大小不超过 $k-1$。我们将 $\mathcal{S}$ 可分的定义推广如下：若存在一个 $i \in [k-1]$，有两个非负整数 λ 和 μ，满足 $|S_1 \cap S_2 \cap \cdots \cap S_i| = k-i-\lambda \geqslant 1$ 以及 $|S_{i+1} \cap S_{i+2} \cap \cdots \cap S_k| = i-\mu \geqslant 1$，则称$\mathcal{S}$ 在 i 处可分，相应的 $\mathcal{S}$ 在分离索引 i 处的两个剩余集族也可以自然的定义出来。基于这些定义，引理 2.2仍然成立，且若 $0 \in \{\lambda, \mu\}$，则引理 2.3也成立，此时行列式相差 ± 1 倍。类似地，可以推广二叉树的定义，注意此时只有当节点的集族是不交的，可分的，且 $0 \in \{\lambda, \mu\}$ 时，才可以通过它的两个剩余集族来扩展子节点。类似

有定理 2.4成立。故式 (2.1) 中由一个良好的集族定义出的矩阵 G，可以作为一个扩展的 $[n,k]_{n-1}$ RS 码的生成矩阵。

(2) 下面，我们解释如何从给定的一个在集合 $[n]$ 上的 $(k-1)$-一致且良好的集族 $\mathcal{S}'$ 中，选择要删除的元素 i。假设 $\mathcal{S}'$ 可生成一个良好的二叉树 $\mathcal{T}$，正如引理 2.7的证明中提到的，$\mathcal{T}$ 的每个节点都自然的与一个交集相关联。则元素 i 的选择如下：首先，元素 i 属于与第二层节点关联的大小至少为 2 的交集中；接下来，我们转向 $\mathcal{T}$ 的另一半子树部分，如果 i 出现在其某个节点的交集中，则该节点必须是叶子节点。通过从 $\mathcal{S}'$ 的每个集合中删除 i，我们可以得到一个好的集族 $\mathcal{S}$，它可以生成一个与 $\mathcal{T}$ 结构完全相同的二叉树。

观察到，我们总是可以从算法 2.1–2.3 得到的集族中选择元素 $i=n$，其中 i 满足注 3(2)，因此定理 1.1 完成。

例 2.4　假设集族 $\mathcal{S}'$ 是例 2.1–2.2中给出的稀疏平衡集族。此时，取元素 $i=8$，将其从 $\mathcal{S}'$ 中删除，得到一个新的集族 $\mathcal{S}=\{S_1,\cdots,S_5\}$，其中 $S_1=\{5,6,7\}$，$S_2=\{1,6,7\}$，$S_3=\{1,2,7\}$，$S_4=\{1,2,3,4\}$，$S_5=\{2,3,4,5\}$。故若 $a_1,\cdots,a_7$ 是 $\mathbb{F}_7$ 上互不相同的 7 个元素，则 $\det(C(\mathcal{P}))$ 在 $\mathbb{F}_7$ 上非零，这里 $\mathcal{P}=\{P_{S_1},\cdots,P_{S_5}\}$，且

$$
\begin{aligned}
P_{S_1}(x)&=(x-a_5)(x-a_6)(x-a_7),\\
P_{S_2}(x)&=(x-a_1)(x-a_6)(x-a_7),\\
P_{S_3}(x)&=(x-a_1)(x-a_2)(x-a_7),\\
P_{S_4}(x)&=(x-a_1)(x-a_2)(x-a_3)(x-a_4),\\
P_{S_5}(x)&=(x-a_2)(x-a_3)(x-a_4)(x-a_5).
\end{aligned}
$$

设 $a_1=0$ 且对任意 $i\in[2,7]$，$a_i=i-1$。则有 $\det(C(P))=6\neq 0$，且可得到一个 $\mathbb{F}_7$ 上的 5×8 阶矩阵 G，

$$
G=\begin{pmatrix}
6 & 3 & 4 & 1 & 0 & 0 & 0 & 0\\
0 & 6 & 3 & 4 & 1 & 0 & 0 & 0\\
0 & 0 & 6 & 3 & 4 & 1 & 0 & 0\\
0 & 0 & 0 & 0 & 3 & 1 & 3 & 1\\
3 & 0 & 0 & 0 & 0 & 3 & 1 & 1
\end{pmatrix}.
$$

观察到 G 是 $\mathbb{F}_7$ 上稀疏平衡的矩阵，且可作为一个 $[8,5]_7$ MDS 码的生成矩阵。

2.5　基于和集的稀疏平衡的 MDS 码构造

在第2.4节中，我们在 $n\leqslant 2k$ 时构造出了稀疏平衡的 $[n,k]_q$ MDS 码。这些构造都是基于一个事实，即一个稀疏、良好且平衡的二元矩阵是存在的。但当

$n > 2k$ 时，由引理 2.6可知，这样的矩阵并不存在。在本节中，我们给出了当 $q = n = p^s$ 且 $n > 2k$ 时的一些构造，其中 p 是素数。该构造思想基于以下引理。

引理 2.17 ([68]引理 9)　令 $p(x) \in \mathbb{F}_q[x]$ 是一个首项系数为 L 且度为 $k-1$ 的多项式。对任意 $i, j \in [k]$，$x_i, y_j \in \mathbb{F}_q$，设 $\det_{1\leqslant i,j\leqslant k}(p(x_i + y_j))$ 是一个 $k \times k$ 阶矩阵的行列式，其第 (i, j) 位置上的元素是 $p(x_i + y_j)$。则有

$$\det_{1\leqslant i,j\leqslant k}(p(x_i + y_j)) = L^k \prod_{i=1}^{k-1} \binom{k-1}{i} \prod_{1\leqslant i<j\leqslant k} (x_i - x_j)(y_j - y_i).$$

考虑有限域 $\mathbb{F}_q$，满足 $q = n = p^s$。假设我们有一个大小为 k 的子集 A 和一个大小为 $k-1$ 的子集 B，使得 $\mathbb{F}_q$ 的每个元素在它们的和集 $A + B$ 中的重数为 $\left\lceil \frac{k(k-1)}{n} \right\rceil$ 或 $\left\lfloor \frac{k(k-1)}{n} \right\rfloor$。对每个元素 $a \in A$，定义集合 $S_a = \{-(a + b) : b \in B\}$。则 $\mathcal{S} = \{S_a : a \in A\}$ 是 $\mathbb{F}_q$ 上的平衡 $(k-1)$-一致集族。令 $P(x) = \prod_{b\in B}(x + b)$，且对于每个 $a \in A$，定义 $P_{S_a} = P(x + a)$。令 $\mathcal{P} = \{P_{S_a} : a \in A\}$，并定义矩阵 $G = C(\mathcal{P}) \cdot V(a_1, \cdots, a_n)$，其中 $V(a_1, \cdots, a_n)$ 是由 $\mathbb{F}_q$ 中的 n 个互不相同的元素 $a_1, \cdots, a_n$ 定义的 $k \times n$ 阶 Vandermonde 矩阵。通过 $\mathcal{S}$ 的平衡性，我们知道 G 是稀疏且平衡的。故只需验证 G 的任一 k 阶子式是否为 0 即可。不失一般性，我们计算由 G 的前 k 列形成的子式，则由引理 2.17可得

$$\begin{aligned}\det(C(\mathcal{P}) \cdot V(a_1, \cdots, a_k)) &= \det_{a\in A, j\leqslant k}(P(a + a_j)) \\ &= \prod_{i=1}^{k-1} \binom{k-1}{i} \prod_{1\leqslant s<t\leqslant k, a\neq a'\in A} (a - a')(a_t - a_s).\end{aligned}$$

因此，若 $p \nmid \prod_{i=1}^{k-1} \binom{k-1}{i}$，则 G 的任一 k 阶子式均不为 0，故 G 可看作是一个 MDS 码的稀疏平衡的生成矩阵。

引理 2.18　设 k 是一个正整数，p 是素数。则 $p \nmid \prod_{i=1}^{k-1} \binom{k-1}{i}$ 当且仅当 $k = p^e m$，其中 $m \in [p-1]$。

证明　充分性是显而易见的。而对于必要性，如果 $k \leqslant p$ 则证明完成。故下面我们考虑 $k \geqslant p+1$，令 $k-1 = m_s p^s + m_{s-1}p^{s-1} + \cdots + m_1 p + m_0$，其中 $s \geqslant 1$，且对任意 $t \in [0, s-1]$ 有 $0 \leqslant m_t \leqslant p-1$，以及 $1 \leqslant m_s \leqslant p-1$。如果存在 $t \in [0, s-1]$，使得 $m_t \leqslant p-2$，则由 Lucas 定理 [69] 可知，当 $i = (p-1)p^t < k$ 时，有 $\binom{k-1}{i} \equiv 0 \pmod{p}$，矛盾。因此，对于所有 $t \in [0, s-1]$，有 $m_t = p-1$，且若 $m_s \leqslant p-2$，则有 $k = (m_s + 1)p^s$；若 $m_s = p-1$，则 $k = p^{s+1}$。至此证明完成。 ∎

下面，我们构造满足上述性质的 $\mathbb{F}_q$ 子集 A 和 B，即 $\mathbb{F}_q$ 中的每个元素在 $A + B$ 中的重数为 $\left\lceil \frac{k(k-1)}{n} \right\rceil$ 或 $\left\lfloor \frac{k(k-1)}{n} \right\rfloor$。为方便起见，我们将 $\mathbb{F}_q$ 的加法群视为加法群 $\mathbb{Z}_p$ 的 s 个拷贝的直积，即 $\mathbb{Z}_p^s$。

构造 2.19　给定正整数 s, e，满足 $\lceil \frac{s}{2} \rceil \leqslant e < s$。令 $n = p^s, k = p^e m$，其中 $m \in [p-1]$。取 $\mathbb{Z}_p^s$ 的一个大小为 p^e 的子空间 U，以及它的补空间 $\overline{U}$，即 $\overline{U}$ 满足 $U + \overline{U} = \mathbb{Z}_p^s$ 且其大小为 p^{s-e}。在 $\overline{U}$ 中选取 m 个互不相同的元素 $x_1, \cdots, x_m$，在 U 中选取 $p^{2e-s}m$ 个互不相同的元素 $y_1 = \mathbf{0}, y_2, \cdots, y_{p^{2e-s}m}$。则可令目标子集 $A = \bigcup_{i\in[m]}(x_i + U)$ 以及 $B = \bigcup_{j\in[p^{2e-s}m]}(y_j + \overline{U}) \setminus \{\mathbf{0}\}$。

在构造2.19中，$|A| = k$ 且 $|B| = k - 1$。由于 $m < p < p^{s-e}$，则 $p^{2e-s}m < p^e$，故 $p^{2e-s}m$ 个互不相同的 U 中元素 $y_1 = \mathbf{0}, y_2, \cdots, y_{p^{2e-s}m}$ 存在。则有

$$\begin{aligned} A + B &= \bigcup_{i\in[m], j\in[p^{2e-s}m]} (x_i + y_j + U + \overline{U}) - A \\ &= \bigcup_{i\in[m], j\in[p^{2e-s}m]} (x_i + y_j + \mathbb{Z}_p^s) - A \\ &= p^{2e-s}m^2\mathbb{Z}_p^s - A. \end{aligned}$$

因此，$\mathbb{Z}_p^s \setminus A$ 中每个元素在 $A + B$ 中的重数为 $p^{2e-s}m^2$，而 A 中的元素重数为 $p^{2e-s}m^2 - 1$。事实上，构造 2.19 也适用于 $k = n$，在这种情况下取 $A = \mathbb{Z}_p^s$ 以及 $B = A \setminus \{\mathbf{0}\}$ 即可。

构造 2.20　给定正整数 s, e，满足 $0 \leqslant e \leqslant \lfloor \frac{s}{2} \rfloor - 1$。令 $n = p^s, k = p^e m$，其中 $m \in [p-1]$。取 $\mathbb{Z}_p^s$ 的一个大小为 p^{e+1} 的子空间 U，以及它的补空间 $\overline{U}$，其大小为 p^{s-e-1}；取 U 的一个大小为 p^e 的子空间 W，以及它在 U 中的补空间 $\overline{W}$，即 $\overline{W}$ 满足 $W + \overline{W} = U$。在 $\overline{W}$ 中选取 m 个互不相同的元素 $x_1 = \mathbf{0}, \cdots, x_m$。则目标子集 A 可以是 $\overline{U}$ 的任何一个大小为 $p^e m$ 的子集，$B = \bigcup_{i\in[m]}(x_i + W) \setminus \{\mathbf{0}\}$。

在构造2.20中，$|A| = k$ 且 $|B| = k - 1$。由于 $0 \leqslant e \leqslant \frac{s}{2} - 1$，$p^e m < p^{s-e-1}$，$\overline{U}$ 中存在一个大小为 $p^e m$ 的子集，即 A 存在。而 $B \subset U$，故对任意 $a_1 \neq a_2 \in A$，有 $S_{a_1} \cap S_{a_2} = \emptyset$。因此可知 $A + B \subseteq \mathbb{Z}_p^s$ 且 $\mathbb{Z}_p^s$ 中每个元素在 $A + B$ 中至多出现一次。事实上，容易验证，当 s 是奇数且 $k = p^{(s-1)/2}$ 时，构造2.20也成立。

结合构造2.19和2.20，可知定理 1.2成立，为了方便阅读，我们重述如下。

定理 (定理 1.2)　对于任何正整数满足 $n = q = p^s$，$k = p^e m$ 与 $1 \leqslant m \leqslant p - 1$ 和 $0 \leqslant e \leqslant s - 1$ 除了 $e = \frac{s-1}{2}$，在这种情况下 $m = 1$ 并且 s 必须是奇数，存在稀疏平衡的 $[n, k]_q$ MDS 码。

在定理 1.2中, 当 $e \leqslant s-2$ 或 $e = s-1$ 且 $m < \frac{p}{2}$ 时, 有 $n > 2k$。因此定理 1.2给出了很多不能由定理 2.5得到的稀疏平衡的 $[n, k]_q$ MDS 码。但码长 $n = q = p^s$ 对域的大小限制太苛刻，且当 $m \in [2, p-1]$，s 是奇数，$k = p^{(s-1)/2}m$ 时，上述两种构造均失效。

2.6　本章总结

在本章中，我们首先针对 $q \geqslant n$，给出一个由生成矩阵零模式刻画的稀疏的 $[n,k]_q$ MDS 码存在的充分条件。基于这个条件，在有限域满足 $q \geqslant n$ 时，我们构造出了码长满足 $n \leqslant 2k$ 的稀疏平衡的 $[n,k]_q$ MDS 码，主要是通过设计几个关于 k 和 n 是多项式时间的算法来完成的。通过扩展坐标，进一步地降低了对有限域大小的限制，即只需要 $q \geqslant n-1$ 即可。为了克服 $n \leqslant 2k$ 的限制，我们提出了一些基于平衡和集 $A+B$ 的构造，其中 $|A|=k$ 以及 $|B|=k-1$。这种方法可以给出部分 $n > 2k$ 时稀疏平衡的 MDS 码。而完全解决 $n > 2k$ 时的构造问题仍具有挑战性，我们将其留作将来研究。

第 3 章　基于 ℓ_1 度量下的最优常重码

在本章中，我们从数据存储在 DNA 分子中的纠错问题出发，研究了 ℓ_1 度量下的常重码。在第3.1节中，我们简单的介绍了 ℓ_1 度量下常重码的研究背景，并简要说明本章贡献。在第3.2节中，我们给出了必要的定义、记号、组合设计理论的相关结果，以及 ℓ_1 度量下的码和填充集族的联系。在第3.3 节中，我们给出了非负整数上重量为 3 和 4 最优常重码的构造。在第3.4 节中，我们考虑了三元常重码，并分别给出了重量为 3 和 4 最优常重码的组合构造。在第3.5节中，我们通过图填充这一工具，给出了一般重量 w 和最小距离 $2w-2$ 的三元常重码的一些结果。最后在第3.7节中对本章进行了简单的总结。

3.1　介绍

为了证明信息存储在活体 DNA 分子中的可靠性，能够纠正各种由于 DNA 分子复制过程中突变引起的串联复制（tandem duplication）、点突变、插入和删除等错误的纠错码随之被人们研究。特别是对于串联复制这种突变，最近，一些学者们研究了串联复制纠错码，参见文献 [8, 70-73]。串联复制，这是一个将 DNA 片段的副本插入到紧邻其原始位置的过程。例如，对于序列 $\boldsymbol{AGCTCT}$，$\boldsymbol{CTCT}$ 是 $\boldsymbol{CT}$ 上长度为 2 的 2-串联复制错误。Jain 等人 [8] 提出了一种编码方案来对抗串联复制错误，该方案是基于非负整数上的 ℓ_1 度量下的常重码存在性构造的。更具体地说，由文献 [8] 的构造 B 可知，给定一个 ℓ_1 度量下的，长度为 n，ℓ_1 重量为 w 和最小 ℓ_1 距离为 $2(t+1)$ 的常重码集合，并给定正整数 m,k，满足 $1 \leqslant n \leqslant m-k+1$ 和 $0 \leqslant w \leqslant \lfloor \frac{mk}{k} \rfloor$，则可以用这些常重码构造一个长度为 m，且能够纠正 t 次 k-串联复制错误的纠错码。因此，选择一组恰当的具有特定重量和长度的最优 ℓ_1 度量常重码，将可以得到一个纠正串联复制错误的最优码。

本章中，我们主要关注 ℓ_1 度量下的最优常重码构造问题。在文献 [34-35] 中，作者使用可分组码（group divisible code）这一与组合设计理论中的可分组设计（group divisible designs (GDD)）[74] 类似的工具，研究了汉明度量下最优三元常重码的构造问题。在他们的方法启发下，我们通过使用填充（packing）和 GDD 来构造 ℓ_1 度量下，固定重量 w 和距离 d 的常重码，并确定了所有重量 $w \leqslant 4$ 下，任意距离 d 和码长 n 的最大码字个数。而对于一般的 w，利用图填充理论，我们得到了权重为 w 和距离为 $2w-2$ 的三元码最优码字大小的渐近结果。

3.2　预备知识

本节中，首先我们将给出一些必要的记号。其次，将正式介绍 ℓ_1 度量下的常重码，以及一些后续有用的性质。并回顾组合设计中一些经典的结果和图的相关概念，以便于后面构造最优码。特别地，我们将利用码字支集刻画出两个码字之间的距离公式，由此可以看出 ℓ_1 度量下的码与填充集族之间的联系，为接下来推导最优码上界做基础。

令 $\mathbb{Z}_{\geqslant 0}$ 表示非负整数的集合，对整数 $q \geqslant 2$，$\mathbb{Z}_q$ 表示模 q 整数环。对于两个集合 A 和 B，定义它们的对称差 为在 A 或 B 中而不在它们交集中的元素集合，记为 $A \vartriangle B$。

3.2.1　ℓ_1 度量下的常重码

给定任意整数 $q \geqslant 2$，令 $I_q := \{0, 1, \cdots, q-1\} \subset \mathbb{Z}_{\geqslant 0}$。定义 I_q^n 为 I_q 上所有 n 长向量的集合。一个长度为 n 的 q 元码 $\mathcal{C}$ 是集合 I_q^n 中的一组向量。$\mathcal{C}$ 的元素称为码字。任意给定两个码字 $\mathbf{u} = (\mathbf{u}_1, \cdots, \mathbf{u}_n), \mathbf{v} = (\mathbf{v}_1, \cdots, \mathbf{v}_n) \in \mathcal{C}$，定义 $\mathbf{u}$ 的支集为 $supp(\mathbf{u}) = \{x \in [n] \mid \mathbf{u}_x \neq 0\}$，$\mathbf{u}$ 和 $\mathbf{v}$ 之间的ℓ_1 距离定义为 $d_{\ell_1}(\mathbf{u}, \mathbf{v}) = \sum_{x\in[n]} |\mathbf{u}_x - \mathbf{v}_x|$（这里的计算基于整数环上）。$\mathbf{u}$ 的ℓ_1 重量指的是 $\mathbf{u}$ 和零向量之间的 ℓ_1 距离，即 $\mathrm{wt}_{\ell_1}(\mathbf{u}) = \sum_{x\in[n]} |\mathbf{u}_x|$，有时也称为$\ell_1$ 权重。若对于码 $\mathcal{C}$ 中的任意码字 $\mathbf{u}$，都有 $\mathrm{wt}_{\ell_1}(\mathbf{u}) = w$，则称码 $\mathcal{C}$ 具有常重量 w；若对于码 $\mathcal{C}$ 中的任意两个不同码字 $\mathbf{u}, \mathbf{v}$，有 $d_{\ell_1}(\mathbf{u}, \mathbf{v}) \geqslant d$，则称码 $\mathcal{C}$ 具有最小 ℓ_1 距离 d；若这两条性质均满足，则称码 $\mathcal{C}$ 是一个 ℓ_1 度量下最小距离为 d 的常重码，且如果 $\mathcal{C}$ 是 q 元的，则记为 $(n, d, w)_q$ 码，如果 $\mathcal{C}$ 是非负整数上的码字，则记为 (n, d, w) 码。为方便起见，对于任意元素 $i \in [q-1]$，若一个码字 $\mathbf{u}$ 的非零元中恰好有 e_i 个 i，则称 $\mathbf{u}$ 的 型为 $1^{e_1}2^{e_2}\cdots(q-1)^{e_{q-1}}$。

在文献 [8] 中，Jain 等人在能够纠正串联复制的纠错码和 ℓ_1 度量下的常重码之间建立了联系。他们表明，若一个纠错码能够纠正 t 个 k-串联复制错误，当且仅当所有 k-共轭码字的 z-部分的零签名（zero signatures），可以构成一个 ℓ_1 度量下非负整数上的最小距离为 $2(t+1)$ 的常重码（详见文献 [8] 的定理 20）。更重要的是，由文献 [8] 的构造 B 可知，为某些权重和长度选择一个最优 ℓ_1 度量下常重码，将可用来构造最优的串联复制纠错码。且给定长度 n，可能需要所有重量从 1 遍历到 n 的任何常重码。此外，如果我们假设每个串联复制的片段连续出现次数不超过 $q-1$ 次，则此时上述提到的常重码可转换为 ℓ_1 度量下的 q 元常重码。特别是当 q 充分大时，这类码字的性质也可以作为研究非负整数上常重码的参考。

受这种联系的启发，本章中，我们考虑在 ℓ_1 度量下常重码的最大码字大小问题。由于我们在本章节中只考虑 ℓ_1 度量，我们将省略下标 ℓ_1 或术语 ℓ_1 度量，

除非另有说明。一个 $(n,d,w)_q$ 码能达到的最大码字个数记为 $A_q(n,d,w)$，达到极值情况的 $(n,d,w)_q$ 码称为是最优的。类似地，对于非负整数 $\mathbb{Z}_{\geqslant 0}$ 上的码，我们用 $A(n,d,w)$ 来表示最大码字个数。

在本章的剩余部分，我们主要着力于，通过构造最优常重码来确定 $A_q(n,d,w)$ 和 $A(n,d,w)$ 的值。下面，我们来看一些简单的性质。

定理 3.1　(a) $A_q(n,2\delta-1,w)=A_q(n,2\delta,w)$；$A(n,2\delta-1,w)=A(n,2\delta,w)$。

(b) 若 $w<\delta$，则 $A_q(n,2\delta,w)=A(n,2\delta,w)=1$。

(c) $A_q(n,2w,w)=\left\lfloor \frac{n}{\lceil \frac{w}{q-1} \rceil} \right\rfloor$；$A(n,2w,w)=n$。

(d) 若 $w\leqslant q-1$，则 $A_q(n,2,w)=A(n,2,w)=\binom{n+w-1}{w}$；若 $w>q-1$，则 $A_q(n,2,w)=\sum_{j=0}^{t}(-1)^j\binom{n}{j}\binom{n-1+w-jq}{w-jq}$，其中 $t=\left\lfloor \frac{w}{q} \right\rfloor$。

证明　对于 (a)，这是因为任何两个权重相等的码字之间的 ℓ_1 距离均是偶数。(b) 中的结果是显而易见的。

对于 (c)，由定义可知，若任何两个权重为 w 的码字，它们之间的距离为 $2w$，当且仅当它们的支集是不交的。而对于集合 I_q 和 $\mathbb{Z}_{\geqslant 0}$ 上的码字，它们的支集大小分别至少为 $\left\lceil \frac{w}{q-1} \right\rceil$ 和 1，故可得结论。

对于 (d)，可以看出，$A(n,2,w)$ 的值等于所有长度为 n、权重为 w 的不同向量的数量，即方程 $x_1+\cdots+x_n=w$ 的非负整数解的个数，其恰好有 $\binom{n+w-1}{w}$ 个。对于 q 元码，我们将分为以下两种情况考虑。如果 $w\leqslant q-1$，则码字的每个位置上的元素最多为 w，这与非负整数上的码相同，故 $A_q(n,2,w)=A(n,2,w)=\binom{n+w-1}{w}$。如果 $w>q-1$，则码字的每个位置上的元素最多为 $q-1$。令 a_j 是长度为 n、权重为 j 的不同 q 元向量的个数，即 a_j 为方程 $x_1+\cdots+x_n=j$ 的非负整数解的个数，其中 $0\leqslant x_i\leqslant q-1$。那么序列 a_j 的生成函数为

$$(1+x+\cdots+x^{q-1})^n=\sum_{j=0}^{kn}a_jx^j.$$

将上述等式展开，可得 $A_q(n,2,w)=a_w=\sum_{j=0}^{t}(-1)^j\binom{n}{j}\binom{n-1+w-jq}{w-jq}$，其中 $t=\left\lfloor \frac{w}{q} \right\rfloor$。∎

由定理 3.1可知，对于任何重量为 w 的常重码，我们只需要考虑其在 4 和 $2w-2$ 之间的最小偶距离。

3.2.2　组合设计

给定一个二元组 $\mathcal{S}=(X,\mathcal{B})$，其中 X 是一个有限点集，$\mathcal{B}$ 是一些 X 的子集集合，$\mathcal{B}$ 中的元素称为区组，则称 $\mathcal{S}$ 是一个阶为 $|X|$、大小为 $|\mathcal{B}|$ 的集族。

一个图 G 是一个集族 (V, E)，其中 E 的所有区组都是 V 的 2-子集。V 和 E 中的元素分别称为图 G 的*顶点* 和*边*。顶点 $v \in V$ 的*度* 指的是包含 v 的边数，记为 $d_G(v)$。令 $\delta(G)$ 表示 G 的最小顶点度数。如果一个图中，每对顶点都有一条边连接，则称该图为*完全图*或者*团*，且若 $|V| = n$，则用 K_n 表示。给定一个由 m 个互不相同的顶点构成的序列 $(v_1, v_2, \cdots, v_m)$，如果对于所有 $i \in [m-1]$，有 $\{v_i, v_{i+1}\} \in E$ 和 $\{v_m, v_1\} \in E$，则称该序列是一个长度为 m 的*圈*。

令 K 是一些正整数的集合。给定一个阶为 n 的集族 $(X, \mathcal{B})$，若对每个区组 $B \in \mathcal{B}$，都有 $|B| \in K$，且 X 的任一 t-子集最多包含于 $\mathcal{B}$ 的一个区组中，则称集族 $(X, \mathcal{B})$ 是一个 t-$(n, K, 1)$-填充。当 $K = \{k\}$ 时，我们只写 k 而不是 $\{k\}$。一个 t-$(n, k, 1)$-填充所能达到的最大区组个数称为*填充数*，记为 $D(n, k, t)$。如果 $|\mathcal{B}| = D(n, k, t)$，则称 t-$(n, k, 1)$-填充 $(X, \mathcal{B})$ 是*最优的*。进一步地，如果 X 的任一 t-子集恰好出现在一个区组中，则称其为t-$(n, k, 1)$- *设计*，或简称为 t-设计。当 $t = 2$ 且 $k = 3$ 时，这样的 2-设计又被叫作 n 阶的*斯坦纳三元系*，记为 STS(n)。t-$(n, k, 1)$-填充的*剩余图*（leave graph）是一个集族 (X, E)，其中 E 由 X 的所有未出现在任何区组的 t-子集构成。对于 $t = 2$ 以及 $k = 3, 4$，或 $t = 3$ 以及 $k = 4$，相应的填充数已经完全确定，见文献 [55]，且对应的剩余图也被完全刻画。我们在下面列出它们以供后续使用。

引理 3.2 ([55])　对任意正整数 n, 若 $n \not\equiv 5 \pmod 6$, 有 $D(n, 3, 2) = \left\lfloor \frac{n}{3} \left\lfloor \frac{n-1}{2} \right\rfloor \right\rfloor$; 若 $n \equiv 5 \pmod 6$，有 $D(n, 3, 2) = \left\lfloor \frac{n}{3} \left\lfloor \frac{n-1}{2} \right\rfloor \right\rfloor - 1$，且对应最优填充的剩余图由一个长度为 4 的圈和 $n-4$ 个孤立点组成。

引理 3.3 ([55])　对任意正整数 n，若 $n \notin \{8, 9, 10, 11, 17, 19\}$，则

$$D(n, 4, 2) = \begin{cases} \left\lfloor \frac{n}{4} \left\lfloor \frac{n-1}{3} \right\rfloor \right\rfloor - 1, & n \equiv 7, 10 \pmod{12}, \\ \left\lfloor \frac{n}{4} \left\lfloor \frac{n-1}{3} \right\rfloor \right\rfloor, & \text{其它。} \end{cases}$$

若 $n = 8, 9, 10, 11, 17, 19$，相应的填充数 $D(n, 4, 2)$ 分别为 $2, 3, 5, 6, 20, 25$。

引理 3.4 ([75])　对任意正整数 n，有

$$D(n, 4, 3) = \begin{cases} \left\lfloor \frac{n}{4} \left\lfloor \frac{n-1}{3} \left\lfloor \frac{n-2}{2} \right\rfloor \right\rfloor \right\rfloor, & \text{若} n \not\equiv 0 \pmod 6, \\ \left\lfloor \frac{n}{4} \left(\left\lfloor \frac{n-1}{3} \left\lfloor \frac{n-2}{2} \right\rfloor \right\rfloor - 1 \right) \right\rfloor, & \text{若} n \equiv 0 \pmod 6. \end{cases}$$

给定一个三元组 $(X, \mathcal{G}, \mathcal{B})$，其中 $(X, \mathcal{B})$ 是一个阶为 n 的集族，$\mathcal{G}$ 是 X 的一些子集的集合，且这些子集恰好构成 X 的一个划分，称 $\mathcal{G}$ 中的元素为*组*。若 $(X, \mathcal{G}, \mathcal{B})$ 满足以下三个条件：(1) 对每个区组 $B \in \mathcal{B}$，都有 $|B| \in K$；(2)X 中的每对未包含于 $\mathcal{G}$ 中组的点对，一定恰好出现在一个区组中；(3) X 中的每对包含于某个组的点对一定不出现在任何区组中。那么称三元组 $(X, \mathcal{G}, \mathcal{B})$ 是一个阶为 n 的可分组设计（K-GDD）。如果对任意 $i = 1, 2, \cdots, s$，$\mathcal{G}$ 中恰好有 a_i 个大小为 g_i

的组，则记该 GDD 的型为 $g_1^{a_1}g_2^{a_2}\cdots g_s^{a_s}$。称一个型为 1^n 且阶为 n 的 K-GDD 为成对平衡设计，记为 (n,K)-PBD。事实上，一个 K-GDD 也是一个 2-$(n,K,1)$-填充，而一个 (n,K)-PBD 是一个满足 X 中的任意一个点对恰好出现在一个区组中的 2-$(n,K,1)$-设计。

引理 3.5 ([74])　对任一正整数 $u\geqslant 3$，一个型为 3^u 的 3-GDD 存在的充分必要条件是 $u\equiv 1 \pmod 2$。

引理 3.6 ([76])　给定正整数 $u\geqslant 4$ 和 $m>0$。对每一个正整数 $g\in\{2,6,7,9,12,15,24,27,36\}$，除去 $(g,u,m)=(2,6,5)$，以及可能的三元组 $(g,u,m)\in\{(2,33,23),(2,33,29),(2,39,35),(6,13,27),(6,13,33),(6,17,39),(6,19,45),(6,19,51),(6,23,63)\}$ 和当 $g\in\{6,12,24,36\}$ 时 $m\geqslant 0$ 外，一个型为 g^um^1 的 4-GDD 存在当且仅当 $m\leqslant g(u-1)/2, gu\equiv 0 \pmod 3, g(u-1)+m\equiv 0 \pmod 3$ 以及 $\binom{gu+m}{2}-u\binom{g}{2}-\binom{m}{2}\equiv 0 \pmod 6$。

为了方便以后使用，除了型为 7^4、9^4、9^5、27^4 和 39^46^1 的 4-GDD 外，这些已被证明存在，见文献 [74]，我们针对几类具体的 GDD，根据引理 3.6推导出它们存在的充分必要条件如下。

推论 3.7　对于所有非负整数 u,m，有

(1) 当 $u\geqslant 6, u\equiv 0 \pmod 3$ 以及 $m\equiv 2 \pmod 3$ 且 $2\leqslant m\leqslant u-1$ 时，除去 $(u,m)=(6,5)$ 以及可能的 $(u,m)\in\{(33,23),(33,29),(39,35)\}$ 这些情况外，对其余所有满足条件的 u,m，存在一个型为 2^um^1 的 4-GDD。

(2) 一个型为 12^um^1 的 4-GDD 存在当且仅当要么 $u=3$ 且 $m=12$，要么 $u\geqslant 4$ 以及 $m\equiv 0 \pmod 3$ 且 $0\leqslant m\leqslant 6(u-1)$。

(3) 一个型为 15^um^1 的 4-GDD 存在当且仅当要么 $u\equiv 0 \pmod 4$ 且 $m\equiv 0 \pmod 3$，$0\leqslant m\leqslant (15u-18)/2$；要么 $u\equiv 1 \pmod 4$ 且 $m\equiv 0 \pmod 6$，$0\leqslant m\leqslant (15u-15)/2$; 要么 $u\equiv 3 \pmod 4$ 且 $m\equiv 3 \pmod 6, 0<m\leqslant (15u-15)/2$。

(4) 对任意 $u\geqslant 4$，一个型为 24^um^1 的 4-GDD 存在当且仅当 $m\equiv 0 \pmod 3$ 且 $0\leqslant m\leqslant 12(u-1)$。

(5) 对任意 $u\geqslant 4$，一个型为 36^um^1 的 4-GDD 存在当且仅当 $m\equiv 0 \pmod 3$ 且 $0\leqslant m\leqslant 18(u-1)$。

(6) 型为 6^7，6^{15}，$6^{11}30^1$，$6^{12}30^1$，7^4，$7^{12}10^1$，9^4，9^46^1，9^5，9^56^1，27^4，27^49^1，27^5 和 39^46^1 的 4-GDD 均存在。

在本章的后续部分中，如果没有特别说明，则所有涉及到的 4-GDD 都可以在推论3.7 中找到。

3.2.3　填充集族与 ℓ_1 度量下常重码的联系

我们首先考虑二元码的情形，即 I_2^n 中的一组向量。I_2^n 中的所有向量 $\mathbf{u}$ 和 $[n]$ 的所有子集 $supp(\mathbf{u})$ 之间存在自然的一一对应关系，因此一个二元码 $C \subset I_2^n$ 自然对应于一个集族 $(X, \{supp(\mathbf{u}) : \mathbf{u} \in C\})$，这里 $X = [n]$。观察到，对于任意两个不同的码字 $\mathbf{u}, \mathbf{v} \in I_2^n$，它们之间的距离为 $|supp(\mathbf{u}) \vartriangle supp(\mathbf{v})|$。假设 C 是一个 $(n, 2t, w)_2$ 码，那么对任意两个码字 $\mathbf{u}, \mathbf{v}$，有 $d(\mathbf{u}, \mathbf{v}) \geqslant 2t$，即 $|supp(\mathbf{u}) \vartriangle supp(\mathbf{v})| \geqslant 2t$，这表明它们的支集交集大小最多为 $w - t$。也就是说，X 的任何 $(w - t + 1)$-子集最多出现在 $\{supp(\mathbf{u}) : \mathbf{u} \in C\}$ 的一个区组中，因此 $(X, \{supp(\mathbf{u}) : \mathbf{u} \in C\})$ 是一个 $(w - t + 1)$-$(n, w, 1)$-填充。

下面，我们针对 q 元码考虑类似的码字支集与填充之间的联系。注意到，任意两个 q 元码字 $\mathbf{u}$ 和 $\mathbf{v}$ 之间的距离可以表示为

$$d(\mathbf{u}, \mathbf{v}) = 2w - 2 \times \sum_{x \in supp(\mathbf{u}) \cap supp(\mathbf{v})} \min\{\mathbf{u}_x, \mathbf{v}_x\}. \tag{3.1}$$

观察到，当 $q = 2$ 时，等式右边退化为 $|supp(\mathbf{u}) \vartriangle supp(\mathbf{v})|$。与二元码情况类似，要达到相同的距离 $2t$，则 $supp(\mathbf{u})$ 和 $supp(\mathbf{v})$ 的交集大小不能超过 $w - t$。所以，我们有一个普适的必要条件，陈述如下并称之为 **UNC** 条件。

***UNC**: 如果 C 是一个 $(n, 2t, w)_q$ 常重码，则集族 $(X, \{supp(\boldsymbol{u}) : \boldsymbol{u} \in C, |supp(\boldsymbol{u})| \geqslant \tau\})$ 是一个 τ-$(n, \{w, w-1, \cdots, \tau\}, 1)$-填充，其中 $\tau \triangleq w - t + 1$。*

注意到，两个不同的码字，它们的支集也有可能相同。不过，只有当两个支集的大小都小于 $w - t + 1$ 时才会发生这种情况，而这种码字将不会包含在填充中。例如重量 $w = 3$，距离为 $d = 2$ 的两个码字：$(0, 1, 2, 0), (0, 2, 1, 0)$，它们的支集均为 $\{2, 3\}$，此时 $t = 1$，故 $w - t + 1 = 3$。

此外，根据式 (3.1) 可知，如果交集大小 $|supp(\mathbf{u}) \cap supp(\mathbf{v})|$ 足够接近 $w - t$，则 $\mathbf{u}_x$ 和 $\mathbf{v}_x$ 的非零元中，最小的元素不能太大。例如，如果交集大小正好是 $w - t$，那么式 (3.1) 的求和项中的每个最小值都必须为 1。如果发生另一种极端情况，即交集大小仅为 1，则最小值可以是以 $w - t$ 为上限的任何值。这也是我们推导出具有特定距离的码存在性的必要条件时，非常关键的一点。为了区别出码字不同位置上的非零元，我们将 q 元码的码字 $\mathbf{u}$ 与一个子集 $\phi(\mathbf{u}) := \{(x, i) : x \in supp(\mathbf{u}) \text{且} \mathbf{u}_x = i\} \subset X \times [q-1]$ 对应起来，这里 $X = [n]$。为了简化符号，我们将在 $\phi(\mathbf{u})$ 中把 (x, i) 写成 x_i。事实上，通过集族的语言刻画一些一般条件，从而使得 $(X \times [q-1], \{\phi(\mathbf{u}) : \mathbf{u} \in C\})$ 是一个 $(n, 2t, w)_q$ 码是非常困难的，因为具有不同 q、t 或 w 的码可能会导致非常不同的要求。为了叙述方便，在不造成混淆的情况下，我们有时不区分 $\mathbf{u}$ 和 $\phi(\mathbf{u})$，因为它们指的是相同的对象。例如，我们

有一个 $(6,6,4)_3$ 常重码 $C \subset I_3^n$，其中包含四个码字 210100、021010、002101 和 100012。等价地，我们可以将它们描述为 $\{1_2,2_1,4_1\}$，$\{2_2,3_1,5_1\}$，$\{3_2,4_1,6_1\}$ 和 $\{6_2,5_1,1_1\}$，它们是 $[6]\times[2]$ 的子集。在本文的后续部分，除了经典的集合 $[n]$，我们有时还假设 X 是一个不为 $[n]$ 的有序集合，例如 $\mathbb{Z}_n$ 等。

在本章剩余部分中，在构造某些码时，我们通常会利用某些群对码字坐标作用，从而生成所有码字。所以一般情况下，我们将使用加法群 $\mathbb{Z}_n$ 来表示码字坐标，而不是经典的集合 $[n]$，并且给 $\mathbb{Z}_n$ 的元素一个自然的顺序。例如，我们有一个 $(4,4,3)_3$ 码 $C \subset I_3^4$，其中有四个码字 1200、0120、0012 和 2001，或者等价地，我们有一个 $(4,4,3)_3$ 码 $C \subset \mathbb{Z}_4\times[2]$，其中码字为 $\{0_1,1_2\}$，$\{1_1,2_2\}$，$\{2_1,3_2\}$ 和 $\{3_1,0_2\}$。事实上，C 可表示为 $C=\{\{(0+i)_1,(1+i)_2\}:i\in\mathbb{Z}_4\}$，即 C 可以通过群 $\mathbb{Z}_4$ 作用于码字 $\{0_1,1_2\}$ 的支集上得到。我们称这样的码字 $\{0_1,1_2\}$ 为基码字（或基区组）。受这个例子的启发，当我们试图找到某种码时，我们只需要找到一个带有额外条件的基码字（或一组基码字），就可以通过群作用得到所有码字。这是组合设计理论中很常见的一种通过计算机搜索找到小设计的方法，对我们找到小码很有帮助。为了节省空间，在表达码时，我们只列出所有基码字和相应的群作用，而不是列出所有码字。注意到，群作用通常作用于码字的支集，但不改变对应位置上的非零元。

3.3　非负整数上的常重码

在本节中，我们考虑在 $\mathbb{Z}_{\geqslant 0}$ 上重量为 $w\leqslant 4$ 的码，并对最小距离满足 $4\leqslant d=2t\leqslant 2w-2$ 以及 $w=3,4$ 时，确定 $A(n,d,w)$ 的确切值。事实上，我们可以选择一个比较小的整数 $q\geqslant w+1$，那么 $\mathbb{Z}_{\geqslant 0}$ 上的重量为 w 的常重码也可以看作是一个 q 元码。正如我们在第 3.2.3小节中观察到的，集族 $(X,\{supp(\mathbf{u}):\mathbf{u}\in C,|supp(\mathbf{u})|\geqslant\tau\})$ 是一个 τ-$(|X|,\{w,w-1,\cdots,\tau\},1)$-填充，且若两个码字，它们支集的交集大小非常接近于 $w-t$，则它们每个相同位置上的两个非零元间的最小值不能太大。

3.3.1　$w=3$ 的常重码上界和最优构造

对于重量为 3 的码字，存在三种类型：1^3，1^12^1，以及 3^1。由于 $w=3$，故只需要考虑最小距离 $d=4$ 的情形，此时 $t=2$ 且 $\tau=2$。下面的定理可通过观察对距离的限制得到。

引理 3.8　码 $C\subset\mathbb{Z}_n\times\mathbb{Z}_{\geqslant 0}$ 是一个 $(n,4,3)$ 常重码当且仅当下面的两个条件成立：

(1) 对于 C 中所有型为 1^3 或 1^12^1 的码字，其支集集合构成一个 2-$(n,\{2,3\},1)$-填充。

(2) 对于 C 中任意两个码字 $\mathbf{u},\mathbf{v}$, 若 $i \in supp(\mathbf{u}) \cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i, \mathbf{v}_i\} = 1$。

证明　必要性。(1) 可由 **UNC** 条件得到，(2) 可由式 (3.1) 中的距离公式给出。

充分性。由 (1) 可知，任意两个码字，它们的支集最多交一个元素。若它们支集不交的话，则其距离为 6；若交的话，由式 (3.1) 和条件 (2) 可知，因为最小值为 1，则其距离为 4。 ∎

由引理 3.8可推出 $A(n,4,3)$ 的上界，再利用最优 2-$(n,3,1)$-填充即可得到相应的最优码，具体如下。

定理 3.9　$A(n,4,3) = D(n,3,2) + n$。

证明　令 x, y 和 z 分别表示型为 $1^3, 1^1 2^1$ 和 3^1 的码字个数。由引理 3.8的条件 (1) 可知,

$$x \leqslant D(n,3,2),$$

这是因为型为 1^3 的码字构成一个 2-$(n,3,1)$-填充。由引理 3.8的条件 (2) 可知，不存在两个码字 $\mathbf{u},\mathbf{v}$，使得对某一个位置 i，$\mathbf{u}_i \geqslant 2$ 和 $\mathbf{v}_i \geqslant 2$ 同时成立。因此，通过计算所有码字中元素 2 和 3 出现的次数，可得

$$y + z \leqslant n.$$

故 $A(n,4,3) = x + y + z \leqslant D(n,3,2) + n$。另一方面，我们可以构造一个 $(n,4,3)$ 常重码 C 如下：从一个最优 2-$(n,3,1)$-填充得到所有型为 1^3 的码字，再取 n 个支集不交的型为 3^1 的码字。容易验证 $|C| = D(n,3,2) + n$。 ∎

3.3.2　$w = 4$ 的常重码上界和最优构造

对于重量为 3 的码字，存在五种类型：1^4，$1^2 2^1$，$1^1 3^1$，2^2 和 4^1。我们首先考虑 $d = 4$ 的情形，此时 $t = 2$ 以及 $\tau = 3$。类似于引理 3.8，码 $C \subset \mathbb{Z}_n \times \mathbb{Z}_{\geqslant 0}$ 是一个 $(n,4,4)$ 常重码当且仅当下面三个条件成立：

(i) 对于 C 中所有型为 1^4 或 $1^2 2^1$ 的码字, 其支集集合构成一个 3-$(n,\{3,4\},1)$-填充。

(ii) 对于 C 中任意两个码字 $\mathbf{u},\mathbf{v}$，若 $i \in supp(\mathbf{u}) \cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i, \mathbf{v}_i\} \leqslant 2$。

(iii) 对于 C 中任意两个码字 $\mathbf{u},\mathbf{v}$，若 $\{i, i'\} \subset supp(\mathbf{u}) \cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i, \mathbf{v}_i\} = \min\{\mathbf{u}_{i'}, \mathbf{v}_{i'}\} = 1$。

充分性可类似引理 3.8去证明。而对于必要性，(i) 可由 **UNC** 条件给出，(ii) 和 (iii) 可由式 (3.1) 得到。更具体地，若 $|supp(\mathbf{u}) \cap supp(\mathbf{v})| = 1$，则最小值至多 2。但是，如果 $|supp(\mathbf{u}) \cap supp(\mathbf{v})| = 2$，则这两个位置上的最小值都必须是 1。由式 (3.1) 可知，对于任意两个码字，它们的交集大小最多为 2，故可得 (i)。

根据上述的充分必要条件，我们可以确定 $(n,4,4)$ 常重码的最大个数。

定理 3.10　$A(n,4,4) = D(n,4,3) + \frac{n(n-1)}{2} + n$。

证明　令 x, y, z, a 和 b 分别表示型为 1^4，$1^2 2^1$，2^2，$1^1 3^1$ 和 4^1 的码字个数。由条件 (i) 和 (ii) 可知，

$$x \leqslant D(n,4,3), \text{ 以及} a + b \leqslant n.$$

由条件 (iii) 知，对于任一码字 $\mathbf{u} \in C$，所有满足 $\mathbf{u}_i \geqslant 1$ 和 $\mathbf{u}_{i'} \geqslant 2$ 的有序对 (i, i') 互不相同。通过对这种对子进行计数，可得

$$2y + 2z + a \leqslant n(n-1).$$

结合上述不等式，我们有

$$A(n,4,4) = x + y + z + a + b \leqslant D(n,4,3) + \frac{n(n-1)}{2} + n.$$

注意到，每个型为 4^1 或 2^2 的码字，与型既不是 $1^2 2^1$ 也不是 $1^1 3^1$ 的码字距离至少为 4，这样的码字一共有 $\binom{n}{2} + n$ 个。故我们可以构造一个 $(n,4,4)$ 常重码 C 如下：从一个最优 3-$(n,4,1)$-填充得到所有型为 1^4 的码字，再取所有型为 2^2 和 4^1 的码字。容易验证 $|C| = D(n,4,3) + \frac{n(n-1)}{2} + n$。 ∎

当距离 $d = 6$ 时，$t = 3$ 以及 $\tau = 2$。由 **UNC** 条件和式 (3.1) 中的距离公式可知，码 $C \subset \mathbb{Z}_n \times \mathbb{Z}_{\geqslant 0}$ 是一个 $(n,6,4)$ 常重码当且仅当下面的两个条件成立：

(a) 对于 C 中所有型为 1^4，$1^2 2^1$，$1^1 3^1$ 和 2^2 的码字，其支集集合构成一个 2-$(n, \{2,3,4\}, 1)$-填充。

(b) 对于 C 中任意两个码字 $\mathbf{u}, \mathbf{v}$，若 $i \in supp(\mathbf{u}) \cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i, \mathbf{v}_i\} = 1$。

定理 3.11　$A(n,6,4) = D(n,4,2) + n$。

证明　沿用定理 3.10中定义的符号，根据条件 (a) 和 (b)，我们有

$$x \leqslant D(n,4,2), \text{ 以及} y + 2z + a + b \leqslant n.$$

则有

$$A(n,6,4) = x + y + z + a + b \leqslant D(n,4,2) + n.$$

故我们可以构造一个 $(n,6,4)$ 常重码 C 如下：从一个最优 2-$(n,4,1)$-填充得到所有型为 1^4 的码字，再取 n 个型为 4^1 的码字。容易验证 $|C| = D(n,4,2) + n$。 ∎

在本章的剩余部分，我们考虑三元常重码，主要思想也是利用最优填充去构造最优码。

3.4　三元常重码

在本节中，我们考虑重量 $w=3,4$ 时的三元常重码，并确定所有距离下码字最大个数。当 $w=3$ 时，只需考虑 $d=4$ 的情形，我们给出了相应最优码的构造。当 $w=4$ 时，最小距离 $d=4$ 或 6，其中 $d=4$ 的情形与非负整数上的码类似。而 $d=6$ 时，我们利用可分组设计给出了相应最优码的构造。

3.4.1　基于填充集族的 $w=3$ 的最优三元常重码构造

本小节中，我们考虑重量为 3 的三元常重码，此时码字有两种类型：1^3 和 1^12^1。当 $d=4$ 时，有 $t=2$ 以及 $\tau=2$。根据 **UNC** 条件和式 (3.1) 中的距离公式可知，码 $C\subset\mathbb{Z}_n\times[2]$ 是一个 $(n,4,3)_3$ 常重码当且仅当下面的两个条件成立：

(1′) 对于 C 中所有码字，其支集集合构成一个 2-$(n,\{2,3\},1)$-填充。

(2′) 对于 C 中任意两个码字 $\mathbf{u},\mathbf{v}$，若 $i\in supp(\mathbf{u})\cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i,\mathbf{v}_i\}=1$。

引理 3.12　$A_3(n,4,3)\leqslant\left\lfloor\frac{n^2+3n}{6}\right\rfloor$。

证明　令 x 和 y 分别表示型为 1^3 和 1^12^1 的码字个数。由条件 (1′) 和 (2′) 可知，

$$3x+y\leqslant\binom{n}{2},\ \text{以及}y\leqslant n.$$

前一个不等式由填充的定义可得，即所有区组中包含的点对数不能超过 $\binom{n}{2}$。后一个不等式是计算元素 2 出现的次数。故 $A_3(n,4,3)=x+y\leqslant\left\lfloor\frac{n^2+3n}{6}\right\rfloor$。 ∎

从引理 3.12 的证明中可以看出，当 $y=n$ 或 $n-1$ 时，码字有可能达到上界。假设我们已经找到了 n 或 $n-1$ 个型为 1^12^1 且满足条件 (2′) 的码字，那么我们只需要找到一个合适的 2-$(n,3,1)$-填充，从中得到型为 1^3 的码字，使得条件 (1′) 满足。如果这个 2-$(n,3,1)$-填充的大小是 $\left\lfloor\frac{n^2+3n}{6}\right\rfloor-n$ 或 $\left\lfloor\frac{n^2+3n}{6}\right\rfloor-n+1$，即可达到引理 3.12中的上界。

定理 3.13　$A_3(n,4,3)=\left\lfloor\frac{n^2+3n}{6}\right\rfloor$。

证明　上界由引理 3.12可知。下面我们只需构造出一个达到上界的最优码即可。容易验证，当 $n\leqslant3$ 时，显然成立；而对于 $n\geqslant4$，我们将构造一个达到上界的 $(n,4,3)_3$ 常重码 C 如下。

当 $n\equiv1,5\pmod6$ 时，存在一个大小为 $\frac{n^2-3n+2}{6}$ 的 2-$(n,3,1)$-填充 $(X,\mathcal{B})$，它的剩余图由一个长为 $n-1$ 的圈和一个孤立点组成 [77]。令 $X=\mathbb{Z}_{n-1}\cup\{\infty\}$，且该圈为 $(0,1,2,\cdots,n-2)$。取 C 的坐标集为 $\mathbb{Z}_{n-1}\cup\{\infty\}$，则 C 可由如下码字组成：对每一个区组 $\{a,b,c\}\in\mathcal{B}$，构造型为 1^3 的码字 $\{a_1,b_1,c_1\}$；由长为 $n-1$ 的圈，构造 $(n-1)$ 个型为 1^12^1 的码字：$\{0_1,1_2\},\{1_1,2_2\},\cdots,\{(n-3)_1,(n-2)_2\},\{(n-2)_1,0_2\}$。

当 $n \equiv 2,4 \pmod 6$ 时，存在一个大小为 $\frac{(n-1)(n-2)}{6}$ 的 2-$(n-1,3,1)$-设计 $(\mathbb{Z}_{n-1}, \mathcal{B})$，事实上，由引理 3.2可知，$(\mathbb{Z}_{n-1}, \mathcal{B})$ 是一个 STS$(n-1)$。取 $\mathcal{C}$ 的坐标集为 $\mathbb{Z}_{n-1} \cup \{\infty\}$，则 $\mathcal{C}$ 可由如下码字组成：对每一个区组 $\{a,b,c\} \in \mathcal{B}$，构造型为 1^3 的码字 $\{a_1, b_1, c_1\}$；由点 $\{\infty\}$，构造 $(n-1)$ 个型为 1^12^1 的码字：$\{0_2, \infty_1\}, \{1_2, \infty_1\}, \cdots, \{(n-3)_2, \infty_1\}, \{(n-2)_2, \infty_1\}$。

当 $n \equiv 3 \pmod 6$ 时，Colbourn 和 Rosa[77] (以及 Colbourn 和 Ling[78]) 证明了存在一个大小为 $\frac{n^2-3n}{6}$ 的 2-$(n,3,1)$-填充 $(\mathbb{Z}_n, \mathcal{B})$，其剩余图由所有满足 $i-j \equiv \pm 1 \pmod n$ 这样的对子构成。取 $\mathcal{C}$ 的坐标集为 $\mathbb{Z}_n$，则 $\mathcal{C}$ 可由如下码字组成：对每一个区组 $\{a,b,c\} \in \mathcal{B}$，构造型为 1^3 的码字 $\{a_1, b_1, c_1\}$；由剩余图，构造 n 个型为 1^12^1 的码字：$\{0_1, 1_2\}, \{1_1, 2_2\}, \cdots, \{(n-2)_1, (n-1)_2\}, \{(n-1)_1, 0_2\}$。

当 $n \equiv 0 \pmod 6$ 时，由引理 3.2可知，存在一个大小为 $\frac{n^2-3n-6}{6}$ 的 2-$(n-1,3,1)$-填充 $(\mathbb{Z}_{n-1}, \mathcal{B})$，其剩余图由一个长度为 4 的圈和 $n-5$ 个孤立点构成。不妨设该圈为 $(0,1,2,3)$。取 $\mathcal{C}$ 的坐标集为 $\mathbb{Z}_{n-1} \cup \{\infty\}$，则 $\mathcal{C}$ 可由如下码字组成：对每一个区组 $\{a,b,c\} \in \mathcal{B}$，构造型为 1^3 的码字 $\{a_1, b_1, c_1\}$；由点 $\{\infty\}$，构造一个型为 1^3 的码字 $\{1_1, 2_1, \infty_1\}$ 和 $(n-5)$ 个型为 1^12^1 的码字：$\{4_2, \infty_1\}, \{5_2, \infty_1\}, \cdots, \{(n-3)_2, \infty_1\}, \{(n-2)_2, \infty_1\}$；由长为 4 的圈，构造五个型为 1^12^1 的码字：$\{0_2, \infty_1\}$，$\{0_1, 1_2\}$，$\{2_2, 3_1\}$，$\{0_1, 3_2\}$ 和 $\{3_1, \infty_2\}$。

下面只需验证码 $\mathcal{C}$ 的大小是否满足要求即可，这里我们省略。而对于距离，由填充及其剩余图的性质，容易验证。 ∎

为了使得定理 3.13证明中的构造更加直观，我们针对每种码长 n，给一些具体的例子。

例 3.1　当 $n=6$ 时，$A_3(6,4,3)=9$。故存在一个点集为 $\mathbb{Z}_5$ 且大小为 2 的 2-$(5,3,1)$-填充，其区组集合为 $024, 134$，剩余图为一个 4 长的圈：$(0,1,2,3)$。取码字坐标为 $\mathbb{Z}_5 \cup \{\infty\}$，则可得到相应最优码如下：

$$\begin{array}{lll} \{0_1,2_1,4_1\} & \{1_1,3_1,4_1\} & \{1_1,2_1,\infty_1\} \\ \{0_2,\infty_1\} & \{0_1,1_2\} & \{2_2,3_1\} \\ \{0_1,3_2\} & \{3_1,\infty_2\} & \{4_2,\infty_1\}. \end{array}$$

当 $n=7$ 时，$A_3(7,4,3)=11$。取一个点集为 $\mathbb{Z}_6 \cup \{\infty\}$ 且大小为 5 的 2-$(7,3,1)$-填充，其区组集合为 $14\infty, 25\infty, 03\infty, 135, 024$，剩余图是一个 6 长的圈：$(0,1,2,3,4,5)$。取码字坐标为 $\mathbb{Z}_6 \cup \{\infty\}$，则可得到相应最优码如下：

$$\begin{array}{llll} \{1_1,4_1,\infty_1\} & \{2_1,5_1,\infty_1\} & \{0_1,3_1,\infty_1\} & \{1_1,3_1,5_1\} \\ \{0_1,2_1,4_1\} & \{0_1,1_2\} & \{1_1,2_2\} & \{2_1,3_2\} \\ \{3_1,4_2\} & \{4_1,5_2\} & \{5_1,0_2\}. & \end{array}$$

当 $n=8$ 时，$A_3(8,4,3)=14$。取一个点集为 $\mathbb{Z}_7$ 且大小为 7 的 STS(7)，其区组集合为 $124,235,346,045,156,026,013$。取码字坐标为 $\mathbb{Z}_7$，则可得到相应最优码如下：

$$\begin{array}{llll}\{1_1,2_1,4_1\} & \{2_1,3_1,5_1\} & \{3_1,4_1,6_1\} & \{0_1,4_1,5_1\}\\ \{1_1,5_1,6_1\} & \{0_1,2_1,6_1\} & \{0_1,1_1,3_1\} & \{0_2,\infty_1\}\\ \{1_2,\infty_1\} & \{2_2,\infty_1\} & \{3_2,\infty_1\} & \{4_2,\infty_1\}\\ \{5_2,\infty_1\} & \{6_2,\infty_1\}. & & \end{array}$$

当 $n=9$ 时，$A_3(9,4,3)=18$。取一个点集为 $\mathbb{Z}_9$ 且大小为 9 的 2-$(9,3,1)$-填充，其区组集合由基区组 035 通过群 $\mathbb{Z}_9$ 作用生成，剩余图为一个长为 9 的圈。取码字坐标为 $\mathbb{Z}_9$，则可得到相应最优码如下：

$$\begin{array}{llll}\{0_1,3_1,5_1\} & \{1_1,4_1,6_1\} & \{2_1,5_1,7_1\} & \{3_1,6_1,8_1\}\\ \{4_1,7_1,0_1\} & \{5_1,8_1,1_1\} & \{6_1,0_1,2_1\} & \{7_1,1_1,3_1\}\\ \{8_1,2_1,4_1\} & \{0_1,1_2\} & \{1_1,2_2\} & \{2_1,3_2\}\\ \{3_1,4_2\} & \{4_1,5_2\} & \{5_1,6_2\} & \{6_1,7_2\}\\ \{7_1,8_2\} & \{8_1,0_2\}. & & \end{array}$$

注 4　当 $n\equiv 3 \pmod 6$ 时，我们给出另一种最优码的构造如下。令 $u=n/3\geqslant 3$ 是一个正整数，则由引理 3.5可知，存在一个型为 3^u 的 3-GDD，记为 $(X,\mathcal{G},\mathcal{B})$。对于每一个 $\mathcal{G}$ 中的组 $G=\{a,b,c\}$，我们可以得到三个型为 1^12^1 的码字：$\{a_1,b_2\}$，$\{b_1,c_2\}$ 和 $\{c_1,a_2\}$，故这种码字有 n 个。容易验证，所有这些型为 1^12^1 的 n 个码字，再结合从 $\mathcal{B}$ 获得的所有型为 1^3 的码字，构成一个最优 $(n,4,3)_3$ 常重码。

事实上，上述提到的型为 3^u 的 3-GDD 也是一个大小为 $\frac{n^2-3n}{6}$ 的 2-$(n,3,1)$-填充，它的剩余图是 u 个长度为 3 的圈 (a,b,c) 的并，其中 $\{a,b,c\}$ 是 $\mathcal{G}$ 中的一个组。再加上定理 3.13 中当 $n\equiv 3 \pmod 6$ 时给出的 2-$(n,3,1)$-填充，我们得到了两个大小相同，但剩余图不同的不同构填充，我们可以使用其中任何一个来构造最优码。

注意到，GDD 的剩余图是一些不相交团的并集，这对我们来说是很方便的，因为我们可以在每个团上输入一个最优的短码，而不会破坏填充集族的性质。例如，在上面的构造中，我们在每个大小为 3 的组上输入了一个最优的 $(3,4,3)_3$ 常重码。这一发现是非常关键的，因为在引理 3.6 中给出了丰富的 GDD 存在性结果，剩下的只需要在每个组上找到一个最优的短码即可。我们将在下一小节中使用该方法，通过 4-GDD 来构造最优 $(n,6,4)_3$ 常重码。

3.4.2　基于可分组设计的 $w=4$ 的最优三元常重码构造

本小节中，我们考虑重量为 4 的三元常重码，此时码字有三种类型：1^4，1^22^1 和 2^2。由于 $w=4$，故只需考虑 $d=4$ 和 6 的情形。

当 $d=4$ 时，重量为 4 的常重码在 $\mathbb{Z}_{\geqslant 0}$ 和 I_3 上的唯一区别是，后者没有型为 $1^1 3^1$ 和 4^1 的码字。利用同定理 3.10证明中类似的思想，我们有如下结果。

定理 3.14　$A_3(n,4,4)=D(n,4,3)+\frac{n(n-1)}{2}$。

当 $d=6$ 时，有 $t=3$ 以及 $\tau=2$。根据 **UNC** 条件和式 (3.1) 中的距离公式可知，码 C 是一个 $(n,6,4)_3$ 常重码当且仅当下面的两个条件成立：

(a') 对于 C 中所有码字，其支集集合构成一个 2-$(n,\{2,3,4\},1)$-填充。

(b') 对于 C 中任意两个码字 $\mathbf{u},\mathbf{v}$，若 $i\in supp(\mathbf{u})\cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i,\mathbf{v}_i\}=1$。

由这两个条件，我们可以给出 $A_3(n,6,4)$ 的上界。

引理 3.15　$A_3(n,6,4)\leqslant\left\lfloor\frac{n(n+5)}{12}\right\rfloor=:U(n)$。

证明　令 x,y,z 分别表示型为 1^4，$1^2 2^1$ 和 2^2 的码字个数。由条件 (a') 以及 2-$(n,\{2,3,4\},1)$-填充的定义可知，区组中包含的点对数不能超过 $\binom{n}{2}$，故有

$$6x+3y+z\leqslant\binom{n}{2}.$$

条件 (b') 表明，在任何一个坐标上，元素 2 最多只能出现在一个码字中，因此

$$y+2z\leqslant n.$$

综上可得 $A_3(n,6,4)=x+y+z\leqslant\frac{1}{6}(\binom{n}{2}+3n-z)\leqslant\left\lfloor\frac{n(n+5)}{12}\right\rfloor$。 ∎

注 5　若 $A_3(n,6,4)=U(n)$，则对 z，即型为 2^2 的码字个数有如下性质：当 $n\equiv 0,3,4,7\pmod{12}$ 时，有 $z=0$；当 $n\equiv 2,5\pmod{12}$ 时，有 $z\leqslant 1$；当 $n\equiv 1,6,9,10\pmod{12}$ 时，有 $z\leqslant 3$；以及当 $n\equiv 8,11\pmod{12}$ 时，有 $z\leqslant 4$。

事实上，由引理 3.15的证明，我们可以得到如下上界

$$x+y+z\leqslant\left\lfloor\frac{n(n+5)-2z}{12}\right\rfloor.$$

当 $n\equiv 1,6,9,10\pmod{12}$ 时，我们有 $U(n)=\frac{n(n+5)-6}{12}$。若 $z>3$，则 $x+y+z\leqslant U(n)-1$，矛盾。因此当 $n\equiv 1,6,9,10\pmod{12}$ 时，有 $z\leqslant 3$。其余的情况也可类似推导。

在本小节的其余部分，我们将致力于构造达到引理 3.15中上界的 $(n,6,4)_3$ 最优码。根据注 4给出的思想，我们将利用已知存在的 4-GDD，并输入最优短码到相应 GDD 的每一组来构造长度为 n 的最优码。故，特定长度且码字个数为 $U(n)$ 的最优短码是非常重要的。因此，我们首先构造最优短码。

1) 最优短码的构造

容易看出 $A_3(1,6,4)=0=U(1)$，$A_3(2,6,4)=1=U(2)$，$A_3(3,6,4)=1=U(3)-1$ 以及 $A_3(4,6,4)=2=U(4)-1$。当 $n=5$ 时，通过穷举法可知，不存在

码字个数为 4 的 $(5,6,4)_3$ 常重码，但 21100，10012，02020 这三个码字可构成一个 $(5,6,4)_3$ 码，故 $A_3(5,6,4)=3=U(5)-1$。当 $n\in[6,11]$ 时，$A_3(n,6,4)=U(n)$，相应的最优码构造如下。

引理 3.16　对任意 $n\in[6,11]$，有 $A_3(n,6,4)=U(n)$。

证明　对任意 $n\in[6,11]$，构造最优码 $\mathcal{C}$ 如下。

当 $n=6$ 时，$A_3(6,6,4)=5$。给定一个点集为 $\mathbb{Z}_7$ 的 STS(7)，其区组集合为 $124,235,346,450,561,602,013$。从点集 $\mathbb{Z}_7$ 中删去点 6 以及所有包含点 6 的区组。利用剩下的四个区组，通过将元素 2 分别放置在 $\{0,1,2,5\}$ 这四个坐标上，即可得到四个型为 1^22^1 的码字。再利用点对 $\{3,4\}$，可得到一个型为 2^2 的码字。综上，可得最优码 $\mathcal{C}$ 如下。

$$\{0_2,1_1,3_1\}\quad\{1_2,2_1,4_1\}\quad\{2_2,3_1,5_1\}$$
$$\{4_1,5_2,0_1\}\quad\{3_2,4_2\}.$$

当 $n=7$ 时，$A_3(7,6,4)=7$。考虑上述相同的斯坦纳三元系 STS(7)，注意到，该三元系可由一个基区组 013 通过 $\mathbb{Z}_7$ 群作用得到所有区组。故最优码 $\mathcal{C}$ 可由基码字 $\{0_2,1_1,3_1\}$ 通过 $\mathbb{Z}_7$ 在其支集上的群作用得到。

当 $n=8$ 时，$A_3(8,6,4)=8$。最优码 $\mathcal{C}$ 可由基码字 $\{0_2,1_1,3_1\}$ 通过 $\mathbb{Z}_8$ 在其支集上的群作用得到。

当 $n=9$ 时，$A_3(9,6,4)=10$。根据注5 以及条件 (b') 可知，型为 2^2 的码字个数最多为 3，且元素 2 在同一个坐标上至多只能出现在一个码字中。通过控制型为 2^2 的码字个数，我们可得最优码 $\mathcal{C}$ 如下。

$$\{0_1,1_1,2_1,3_1\}\quad\{0_1,4_1,5_1,6_1\}\quad\{7_1,8_1,0_2\}\quad\{3_1,6_1,7_2\}$$
$$\{5_1,7_1,2_2\}\qquad\{4_1,7_1,1_2\}\qquad\{1_1,8_1,6_2\}\quad\{3_1,8_1,5_2\}$$
$$\{2_1,4_1,8_2\}\qquad\{3_2,4_2\}.$$

当 $n=10$ 时，$A_3(10,6,4)=12$。此时最优码 $\mathcal{C}$ 可由一个 $(10,\{3,4\})$-PBD 通过合理分配元素 2 的位置得到，如下。

$$\{0_1,1_1,2_1,3_1\}\quad\{3_1,5_1,9_2\}\quad\{2_1,6_1,8_2\}\quad\{2_1,4_2,9_1\}$$
$$\{0_1,4_1,5_1,6_1\}\quad\{1_1,6_2,9_1\}\quad\{1_2,4_1,7_1\}\quad\{2_2,5_1,7_1\}$$
$$\{0_1,7_1,8_1,9_1\}\quad\{3_2,4_1,8_1\}\quad\{1_1,5_2,8_1\}\quad\{3_1,6_1,7_2\}.$$

当 $n=11$ 时，$A_3(11,6,4)=14$。此时最优码 $\mathcal{C}$ 可通过计算机搜索得到，如下。

$$\{3_1,6_1,8_1,10_1\}\quad\{1_1,9_2,10_1\}\quad\{0_1,3_1,5_2\}\quad\{0_2,8_1,9_1\}$$
$$\{2_1,3_1,7_1,9_1\}\quad\{2_2,5_1,10_1\}\quad\{4_1,7_1,10_2\}\quad\{4_2,5_1,9_1\}$$
$$\{0_1,2_1,4_1,6_1\}\quad\{1_1,5_1,6_2\}\quad\{0_1,1_1,7_2\}\quad\{5_1,7_1,8_2\}$$
$$\{1_2,2_1,8_1\}\qquad\{1_1,3_2,4_1\}.$$

∎

当 $n = 12$ 时，其最优码个数无法达到上界 $U(12)$，具体如下。

引理 3.17　$A_3(12, 6, 4) = U(12) - 1 = 16$。

证明　当 $n = 12$ 时，$U(n) = 17$。令最优 $(12, 6, 4)_3$ 常重码为 $\mathcal{C}$。根据引理 3.15的证明以及注 5可知, 若 $|\mathcal{C}| = U(12) = 17$, 则相应的 $z = 0$, $x = 5$, $y = 12$, 以及对 $\mathcal{C}$ 中所有型为 1^4 和 $1^2 2^1$ 的码字，其支集构成一个 2-$(12, \{3, 4\}, 1)$-填充 $(X = \mathbb{Z}_{12}, \mathcal{B})$。由于每个型为 1^4 和 $1^2 2^1$ 的码字，其支集中分别包含 6 和 4 个点对，而 $5 \times 6 + 12 \times 3 = 66 = \binom{12}{2}$，即 $\mathbb{Z}_{12}$ 的每个点对恰好包含于一个支集中，因此该填充的剩余图为空图。反证法，假设 $|\mathcal{C}| = U(12) = 17$。对任意 $i \in \mathbb{Z}_{12}$，令 x_i 表示型为 1^4，且在坐标 i 上的元素非零的码字个数。注意到，每个这样的码字，其支集均包含了三个含有 i 的点对，由条件 (a') 知，所有这样的点对必须互不相同。由于包含 i 的点对共有 11 个，我们有 $x_i \leqslant \left\lfloor \frac{11}{3} \right\rfloor = 3$。通过计算所有型为 1^4 的 5 个码字中非零元的个数，可得

$$x_0 + x_1 + \cdots + x_{11} = 20.$$

由于 $\mathcal{C}$ 中的所有码字, 其支集集合构成一个剩余图为空图的 2-$(12, \{3, 4\}, 1)$-填充，则包含 i 的点对数 $11 = 3x_i + 2y_i$，这里 y_i 指的是型为 $1^2 2^1$，且在坐标 i 上的元素非零的码字个数。由此可知，x_i 必须为奇数，即为 1 或 3。取 $j = 1, 3$，令 d_j 为满足 $x_i = j$ 的 i 的个数，则有

$$\begin{cases} d_1 + 3d_3 = 20, \\ d_1 + d_3 = 12. \end{cases}$$

因此 $d_1 = 8, d_3 = 4$。不失一般性, 可设 $x_0 = 3$, 且 $\mathcal{C}$ 中存在三个码字 $\{0_1, 1_1, 2_1, 3_1\}$, $\{0_1, 4_1, 5_1, 6_1\}$ 以及 $\{0_1, 7_1, 8_1, 9_1\}$。由于 $d_1 = 8$ 和 $d_3 = 4$，则一定存在一个点 $i \in [9]$，使得 $x_i = 3$，进一步地，不妨设 $x_1 = 3$。若一个包含 $\{10_1, 11_1\}$ 为子集的型为 1^4 的码字含有 1_1, 在这种情况下, 我们无法再找到一个型为 1^4 的码字添加进 $\mathcal{C}$ 中, 使得 $x_1 = 3$, 矛盾。因此可设 $\mathcal{C}$ 中含有码字 $\{1_1, 4_1, 7_1, 10_1\}, \{1_1, 5_1, 8_1, 11_1\}$。为了确保 x_5 是奇数，我们仍需再找一个型为 1^4 的码字，而此时 $\mathcal{C}$ 中已经有 5 个型为 1^4 的码字了，矛盾。故 $A_3(12, 6, 4) \leqslant 16$。

下面，我们给出一个大小为 16 的最优码 $\mathcal{C}$。

$$\begin{array}{llll} \{3_1, 7_1, 8_1, 11_1\} & \{4_1, 11_1, 0_2\} & \{1_1, 5_1, 7_2\} & \{7_1, 10_1, 4_2\} \\ \{2_1, 6_1, 8_1, 10_1\} & \{1_1, 2_1, 11_2\} & \{0_1, 5_1, 6_2\} & \{3_1, 5_1, 2_2\} \\ \{1_1, 3_1, 4_1, 6_1\} & \{6_1, 11_1, 9_2\} & \{5_1, 9_1, 8_2\} & \{1_1, 9_1, 10_2\} \\ \{0_1, 2_1, 7_1, 9_1\} & \{0_1, 10_1, 3_2\} & \{0_1, 8_1, 1_2\} & \{10_1, 11_1, 5_2\}. \end{array}$$

∎

2) 基于 4-GDD 的递归构造

接下来，受到注4的启发，我们将利用 4-GDD 和特定的最优短码来构造码字个数达到 $U(n)$ 的最优 $(n,6,4)_3$ 常重码。主要的思想如下：给定一个阶为 n 的 4-GDD，由于其每个区组都是其点集的 4-子集，可将其看作是某个码字的支集，故从该 GDD 的每个区组出发，都可得到一个型为 1^4 且长度 n 的码字；对于 GDD 的每个组，假设其大小为 g，则我们可取一个坐标集为该组内元素的最优 $(g,6,4)_3$ 码，然后通过把零分配给其余的坐标，从而将其扩展成长度为 n 的码；由此所有这样的码字构成的集合即目标码。在这种方法中，通过条件 (a') 知，可以将一个 $(n,6,4)_3$ 码看成一个 2-$(n,\{2,3,4\},1)$ 填充，然后由条件 (b') 知，需给元素 2 分配恰当的坐标，使得每个位置上最多有一个码字含有元素 2。

尽管上述构造方法看起来很简单，但要保证最终生成的码的大小达到上界 $U(n)$，并且码长 n 遍历所有正整数，并不容易。为了完成这一点，我们的方法在很大程度上取决于不同型的 4-GDD 和特殊码长的最优短码存在性，具体要求如下。更具体地实现，可见定理 3.18–3.20。

(1) **良好的 4-GDD**。给定正整数 g 和非负整数 m，使得对几乎所有正整数 u，型为 $g^u m^1$ 的 4-GDD 均存在。故相应的最优 $(g,6,4)_3$ 和 $(m,6,4)_3$ 短码的存在性，极大可能可以保证，码长为 $n=gu+m$ 的最优码的存在性。由推论3.7可知，当 $g\equiv 0 \pmod{12}$ 和 $m\equiv 0 \pmod 3$ 时，符合要求的 4-GDD 存在。如果上述方法有效，则可以得到所有码长 n 满足 $n\equiv 0 \pmod 3$ 的最优码。

(2) **添加额外的点**。当 $g\equiv 0 \pmod{12}$ 和 $m\equiv 0 \pmod 3$ 时，尽管型为 $g^u m^1$ 的 4-GDD 是良好的，但此时只能构造出码长 n 满足 $n\equiv 0 \pmod 3$ 的码。为了能够使用相同的 4-GDD 构造出覆盖所有长度的最优码，我们在输入短码时需要添加一个或两个额外的点。即我们需要在 4-GDD 的点集中添加 t 个额外的点，其中 $t=1$ 或 2，使得 4-GDD 的每个组共享这 t 个点，再向这些组中对应输入一个合适的 $(g+t,6,4)_3$ 或 $(m+t,6,4)_3$ 短码，从而得到码长为 $n=gu+m+t\equiv t \pmod 3$ 的最优码。

(3) **良好的最优短码**。给定上述 $t\geqslant 0$，我们应当非常谨慎地选择 $(g+t,6,4)_3$ 短码，特别是当 $t>0$ 时，因为它将用于每个组，或每个组与额外点的并集，且这些出现在每一组的额外点（如果有的话）都是相同的 t 个点。因此，首先，当 $t=1,2$ 时，由条件 (b') 可知，此码不能有码字在坐标是额外点上含有元素 2。其次，由条件 (a') 可知，当 $t=2$ 时，对应填充集族的区组中，不能包含由额外两个点构成的点对。最后，对应填充的剩余图应该是空图（除了 $t=2$，此时剩余图是一个只有一条由两个额外的点连接的边和孤立点构成。），否则，随着 u 的增大，最终得到的码，其剩余图中包含无限数量的边，根据引理 3.15的证明可知，其码字大小将远小于 $U(n)$。

基于以上分析，我们将分成三种情况去构造最优码：即 $t = 0, 1, 2$，对应 $n \equiv 0, 1, 2 \pmod 3$。递归构造的具体细节和良好短码的定义将根据构造的不同，而随之发生改变，我们将在用到时，再给出详细的阐述。

a) 当 $t = 0$

在这种情况下，我们将使用型为 $g^u m^1$ 的 4-GDD 为所有 n 满足 $n \equiv 0 \pmod 3$，构造最优 $(n, 6, 4)$ 码，除去极个别的 n 比较小的时候。

定理 3.18　设存在一个型为 $g^u m^1$ 的 4-GDD，其中 $g \equiv 0, 3, 4, 7 \pmod{12}$。如果 $A_3(g, 6, 4) = U(g)$ 且 $A_3(m, 6, 4) = U(m)$，则 $A_3(gu + m, 6, 4) = U(gu + m)$。

证明　令 $n = gu + m$。给定一个型为 $g^u m^1$ 的 4-GDD $(X, \mathcal{G}, \mathcal{B})$，其中 $|X| = n$，我们构造一个 $(n, 6, 4)_3$ 码 $\mathcal{C} \subset X \times [2]$ 如下。对每个组 $G \in \mathcal{G}$，构造一个最优 $(|G|, 6, 4)_3$ 码 $\mathcal{C}_G \subset G \times [2]$，由假设可知这样的码是存在的，注意到，我们可将 $\mathcal{C}_G$ 看成是 $X \times [2]$ 的一个子集，即，通过将零分配给其余的坐标，从而将其扩展为一个 $(n, 6, 4)_3$ 码。令 $\mathcal{C}_0$ 为由 $\mathcal{B}$ 中的每个区组所生成的型为 1^4 的所有码字集合，即 $\mathcal{C}_0 = \{\{a_1, b_1, c_1, d_1\} : \{a, b, c, d\} \in \mathcal{B}\}$。则容易验证 $\mathcal{C} = \mathcal{C}_0 \bigcup (\cup_{G \in \mathcal{G}} \mathcal{C}_G)$ 是一个 $(n, 6, 4)_3$ 码。进一步地，由于当 $g \equiv 0, 3, 4, 7 \pmod{12}$ 时，$\frac{g(g+5)}{12}$ 是一个整数，可计算出 $\mathcal{C}$ 的大小如下。

$$\begin{aligned}
|\mathcal{C}| &= |\mathcal{B}| + u \cdot U(g) + U(m) \\
&= \frac{(g(u-1)+m)gu + gum}{12} + u \cdot \frac{g(g+5)}{12} + \left\lfloor \frac{m(m+5)}{12} \right\rfloor \\
&= \frac{gu(gu + 2m + 5)}{12} + \left\lfloor \frac{m(m+5)}{12} \right\rfloor \\
&= \left\lfloor \frac{(gu+m)(gu+m+5)}{12} \right\rfloor = U(gu + m).
\end{aligned}$$

故证明完成。 ∎

注意到，在定理 3.18中，对最优短码 $(g, 6, 4)_3$ 的大小有限制。事实上，当 $g \equiv 0, 3, 4, 7 \pmod{12}$ 时，由注5可知，一个大小达到上界 $U(g)$ 的 $(g, 6, 4)_3$ 码中，不存在型为 2^2 的码字。更具体地，这种码由 g 个型为 $1^2 2^1$，和 $U(g) - g$ 个型为 1^4 的码字构成。因此通过简单的计算可知，该码对应的填充集族的剩余图是一个空图。下面的例子给出了一个这种码的具体构造。

例 3.2　一个大小为 $U(36)$ 的最优 $(36, 6, 4)_3$ 码可由以下基码字，通过在其支集上加 6 模 36 生成。其中在短轨道中的码字以粗体显示。

$\{\mathbf{2_1}, \mathbf{11_1}, \mathbf{20_1}, \mathbf{29_1}\}$ $\{19_1, 22_1, 26_1, 27_1\}$ $\{0_1, 19_1, 23_1, 29_1\}$
$\{\mathbf{1_1}, \mathbf{10_1}, \mathbf{19_1}, \mathbf{28_1}\}$ $\{3_1, 10_1, 13_1, 33_1\}$ $\{8_1, 15_1, 16_1, 30_1\}$
$\{\mathbf{0_1}, \mathbf{9_1}, \mathbf{18_1}, \mathbf{27_1}\}$ $\{6_1, 16_1, 18_1, 22_1\}$ $\{9_1, 14_1, 20_1, 31_1\}$
$\{4_1, 15_1, 17_1, 19_1\}$ $\{2_1, 4_1, 23_1, 35_1\}$ $\{2_1, 7_1, 22_1, 30_1\}$
$\{2_1, 16_1, 17_1, 28_1\}$ $\{2_1, 5_1, 6_1, 21_1\}$ $\{1_1, 2_1, 12_1, 14_1\}$
$\{6_1, 7_1, 9_1, 23_1\}$ $\{21_1, 24_1, 8_2\}$ $\{24_1, 29_1, 18_2\}$
$\{3_1, 35_1, 27_2\}$ $\{1_1, 31_1, 23_2\}$ $\{9_1, 35_1, 28_2\}$
$\{0_1, 31_1, 7_2\}$.

根据推论 3.7，当 $u \geqslant 4$ 和 $m \equiv 0 \pmod 3$ 以及 $0 \leqslant m \leqslant 18(u-1)$ 时，存在型为 $36^u m^1$ 的 4-GDD。取 $g = 36$，以及例3.2 中的最优 $(36, 6, 4)_3$ 码，如果存在最优 $(m, 6, 4)_3$ 码的话，应用定理 3.18，即可得到长度为 $n = 36u + m$ 的最优码。此外，当 m 遍历 $0, 3, 6, \cdots, 33 \pmod{36}$ 的所有代表元时，$n = 36u + m$ 也同时遍历了几乎所有 $n \equiv 0 \pmod 3$ 的正整数，我们由此可以得到所有长度满足 $n \equiv 0 \pmod 3$ 的最优码，除了个别由于当 $u < 4$ 时，型为 $36^u m^1$ 的 4-GDD 不存在而导致的小码长无法被覆盖。

b) 当 $t = 1$

在这种情况下，我们将利用 4-GDD，通过给其点集中添加一个额外的点，使得它的每个组都共享这个点，来构造最优码。由良好短码的要求知，我们需要一个满足*性质 (A)* 的最优 $(n, 6, 4)_3$ 码 $\mathcal{C}$。这里性质 (A) 指的是：$|\mathcal{C}| = U(n)$，且 $\mathcal{C}$ 中恰好有 $n-1$ 个型为 $1^2 2^1$ 的码字，没有型为 2^2 的码字，其余码字的型均为 1^4。事实上，只有当 $n \equiv 1, 6, 9, 10 \pmod{12}$ 时，码 $\mathcal{C}$ 才具有性质 (A) 且其对应填充的剩余图为空图。为了证明这一点，只需要考虑当下面这个等式成立时，n 所需要满足的条件即可：$(U(n) - (n-1)) \times 6 + (n-1) \times 3 = \binom{n}{2}$，即所有码字中包含的点对总数为 $\binom{n}{2}$。注意到，引理 3.16中给出的 $(10, 6, 4)_3$ 码具有性质 (A)。为了方便后续使用，下面的例子中，给出了另一个具有性质 (A) 的 $(13, 6, 4)_3$ 常重码。

例 3.3　具有性质 (A) 的最优 $(13, 6, 4)_3$ 码可通过计算机搜索得到，列举如下。

$\{9_1, 10_1, 11_1, 12_1\}$ $\{6_1, 7_1, 8_1, 12_1\}$ $\{2_1, 3_1, 8_1, 11_1\}$
$\{3_1, 4_1, 5_1, 12_1\}$ $\{1_1, 5_1, 7_1, 9_1\}$ $\{0_1, 4_1, 6_1, 10_1\}$
$\{0_1, 1_1, 2_1, 12_1\}$ $\{5_1, 11_1, 6_2\}$ $\{0_1, 11_1, 7_2\}$
$\{7_1, 10_1, 3_2\}$ $\{1_1, 4_1, 11_2\}$ $\{2_1, 5_1, 10_2\}$
$\{3_1, 6_1, 1_2\}$ $\{1_1, 10_1, 8_2\}$ $\{2_1, 6_1, 9_2\}$
$\{0_1, 8_1, 5_2\}$ $\{4_1, 7_1, 2_2\}$ $\{3_1, 9_1, 0_2\}$
$\{8_1, 9_1, 4_2\}$.

下面，我们将给出具体的构造细节。核心思想主要是给 4-GDD 的点集中添加一个额外的点，并在每组中输入一个带有性质 (A) 的最优短码。推论 3.7说明了，当 $u \geqslant 4$、$m \equiv 0 \pmod 3$ 以及 $0 \leqslant m \leqslant 6(u-1)$ 时，型为 $12^u m^1$ 的良好 4-GDD 是存在的。取 $g = 12$，以及例3.3中具有性质 (A) 的最优 $(13, 6, 4)_3$ 码，如果存在最优 $(m+1, 6, 4)_3$ 码的话，应用定理 3.19，即可得到所有长为 $n = 12u + m + 1$ 的最优码。此外，当 $m+1$ 遍历 $1, 4, 7, 10 \pmod{12}$ 的所有代表元时，除去个别 n 非常小的情况，我们便可以得到所有长度满足 $n \equiv 1 \pmod 3$ 的最优码。

定理 3.19　设存在一个型为 $g^u m^1$ 的 4-GDD，其中 $g \equiv 0, 5, 8, 9 \pmod{12}$。如果存在一个具有性质 (A) 的最优 $(g+1, 6, 4)_3$ 码且 $A_3(m+1, 6, 4) = U(m+1)$，则 $A_3(gu+m+1, 6, 4) = U(gu+m+1)$。

证明　令 $(X', \mathcal{G}, \mathcal{B})$ 是一个型为 $g^u m^1$ 的 4-GDD，其中 $X = X' \cup \{\infty\}$。我们构造一个长度为 $gu+m+1$ 的最优码 $\mathcal{C} \subset X \times [2]$ 如下。对每个长为 g 的组 $G \in \mathcal{G}$，构造一个具有性质 (A) 的最优 $(g+1, 6, 4)_3$ 码 $\mathcal{C}_G \subset (G \cup \{\infty\}) \times [2]$，使得 $\mathcal{C}_G$ 中 g 个型为 $1^2 2^1$ 的码字中元素 2 均匀分布在坐标集 G 上，而不在点 $\{\infty\}$ 上；对每个长为 m 的组 $G \in \mathcal{G}$，令 $\mathcal{C}_G \subset (G \cup \{\infty\}) \times [2]$ 是一个最优 $(m+1, 6, 4)_3$ 码；对每个组 $G \in \mathcal{G}$，可自然的将码 $\mathcal{C}_G$ 看成是 $X \times [2]$ 的一个子集；令 $\mathcal{C}_0$ 为由 $\mathcal{B}$ 中的每个区组所生成的型为 1^4 的所有码字集合。则 $\mathcal{C} = \mathcal{C}_0 \bigcup (\cup_{G \in \mathcal{G}} \mathcal{C}_G) \subset X \times [2]$ 是一个 $(gu+m+1, 6, 4)_3$ 码，其大小为

$$
\begin{aligned}
|\mathcal{C}| =& |\mathcal{B}| + u \cdot U(g+1) + U(m+1) \\
=& \frac{(g(u-1)+m)gu + gum}{12} + u \cdot \frac{(g+1)(g+6)-6}{12} \\
&+ \left\lfloor \frac{(m+1)(m+6)}{12} \right\rfloor \\
=& \frac{gu(gu+2m+7)}{12} + \left\lfloor \frac{(m+1)(m+6)}{12} \right\rfloor \\
=& \left\lfloor \frac{(gu+m+1)(gu+m+6)}{12} \right\rfloor \\
=& U(gu+m+1).
\end{aligned}
$$

其中第二个等式，可由 $U(g+1) = \frac{(g+1)(g+6)-6}{12}$ 得到，这里 $g \equiv 0, 5, 8, 9 \pmod{12}$。 ∎

c) 当 $t = 2$

在这种情况下，我们将利用 4-GDD，通过给其点集中添加两个额外的点，使得它的每个组都共享这一点对，来构造最优码。类似地，我们需要一个满足*性质 (B)* 的最优 $(n, 6, 4)_3$ 码 $\mathcal{C}$。这里性质 (B) 指的是：$|\mathcal{C}| = U(n)$，且 $\mathcal{C}$ 中恰好有 $n-2$ 个型为 $1^2 2^1$ 的码字，一个型为 2^2 的码字，其余码字的型均为 1^4。和性质 (A) 类

似，只有当 $n \equiv 2,5 \pmod{12}$ 时，码 $\mathcal{C}$ 才具有性质 (B) 且其对应填充的剩余图为空图。将型为 2^2 的码字删去，则此时剩余图中恰好含有一条边。下面的例子中，给出了一个具有性质 (B) 的 $(26,6,4)_3$ 常重码。

例 3.4　给定如下基码字，则一个具有性质 (B) 的最优 $(26,6,4)_3$ 码，可由自同构 (0 6 12 18)(1 7 13 19)(2 8 14 20)(3 9 15 21)(4 10 16 22)(5 11 17 23)(24 25) 作用在这些基码字支集上生成。其中在短轨道中的码字以粗体显示。注意到，在该码中，有且仅有一个型为 2^2 的码字 $\{24_2, 25_2\}$。

$$
\begin{array}{lll}
\{\mathbf{3_1, 9_1, 15_1, 21_1}\} & \{\mathbf{4_1, 10_1, 16_1, 22_1}\} & \{\mathbf{1_1, 7_1, 13_1, 19_1}\} \\
\{\mathbf{0_1, 6_1, 12_1, 18_1}\} & \{\mathbf{5_1, 11_1, 17_1, 23_1}\} & \{\mathbf{2_1, 8_1, 14_1, 20_1}\} \\
\{4_1, 12_1, 19_1, 25_1\} & \{0_1, 10_1, 19_1, 20_1\} & \{9_1, 12_1, 16_1, 17_1\} \\
\{8_1, 17_1, 18_1, 25_1\} & \{4_1, 6_1, 8_1, 9_1\} & \{3_1, 4_1, 17_1, 24_1\} \\
\{7_1, 9_1, 20_1, 24_1\} & \{2_1, 5_1, 9_1, 18_1\} & \{2_1, 7_1, 10_1, 21_1\} \\
\{1_1, 21_1, 12_2\} & \{20_1, 22_1, 17_2\} & \{1_1, 5_1, 22_2\} \\
\{5_1, 7_1, 14_2\} & \{5_1, 19_1, 3_2\} & \{0_1, 17_1, 1_2\} \\
\{\mathbf{24_2, 25_2}\}. & &
\end{array}
$$

根据推论3.7可知，对任意正整数 u、非负整数 m，满足 $u \geqslant 4$、$m \equiv 0 \pmod{3}$ 以及 $0 \leqslant m \leqslant 12(u-1)$，都存在型为 $24^u m^1$ 的良好 4-GDD。则与定理 3.19类似，我们可通过添加两个额外点来构造目标码字。具体地，取 $g = 24$，以及例3.4中具有性质 (B) 的最优 $(26,6,4)_3$ 码，如果存在最优 $(m+2,6,4)_3$ 码的话，应用定理 3.20，即可得到所有长为 $n = 24u + m + 2$ 的最优码。此外，当 $m+2$ 遍历 $2,5,8,\cdots,23 \pmod{24}$ 的所有代表元时，除去个别 n 非常小的情况，我们便可以得到所有长度满足 $n \equiv 2 \pmod{3}$ 的最优码。

定理 3.20　设存在一个型为 $g^u m^1$ 的 4-GDD，其中 $g \equiv 0,3 \pmod{12}$。如果存在一个具有性质 (B) 的最优 $(g+2,6,4)_3$ 码且 $A_3(m+2,6,4) = U(m+2)$，则 $A_3(gu+m+2,6,4) = U(gu+m+2)$。

证明　令 $(X', \mathcal{G}, \mathcal{B})$ 是一个型为 $g^u m^1$ 的 4-GDD，其中 $G_0 \in \mathcal{G}$ 是唯一一个大小为 m 的组。设 $X = X' \cup \{\iota, \jmath\}$，且 $\iota, \jmath \notin X'$。我们构造一个长度为 $gu+m+2$ 的最优码 $\mathcal{C} \subset X \times [2]$ 如下。对每个长为 g 的组 $G \in \mathcal{G}$，构造一个具有性质 (B) 的最优 $(g+2,6,4)_3$ 码 $\mathcal{C}'_G \subset (G \cup \{\iota, \jmath\}) \times [2]$，使得 $\mathcal{C}'_G$ 中 g 个型为 $1^2 2^1$ 的码字中元素 2 均匀分布在坐标集 G 上，而型为 2^2 的码字为 $\{\iota_2, \jmath_2\}$，然后取 $\mathcal{C}_G = \mathcal{C}'_G \setminus \{\{\iota_2, \jmath_2\}\}$。对组 G_0，令 $\mathcal{C}_{G_0} \subset (G_0 \cup \{\iota, \jmath\}) \times [2]$ 是一个最优 $(m+2,6,4)_3$ 码。然后，对每个组 $G \in \mathcal{G}$，将码 $\mathcal{C}_G$ 自然的看成是 $X \times [2]$ 的一个子集。最后，取 $\mathcal{C}_0$ 为由 $\mathcal{B}$ 中的每个区组所生成的型为 1^4 的所有码字集合。则 $\mathcal{C} = \mathcal{C}_0 \cup (\cup_{G \in \mathcal{G}} \mathcal{C}_G) \subset X \times [2]$ 是

一个 $(gu+m+2,6,4)_3$ 码，其大小为

$$\begin{aligned}|\mathcal{C}| =&|\mathcal{B}| + u\cdot(U(g+2)-1)+U(m+2)\\ =&\frac{(g(u-1)+m)gu+gum}{12}\\ &+u\cdot\left(\frac{(g+2)(g+7)-2}{12}-1\right)+\left\lfloor\frac{(m+2)(m+7)}{12}\right\rfloor\\ =&\frac{gu(gu+2m+9)}{12}+\left\lfloor\frac{(m+2)(m+7)}{12}\right\rfloor\\ =&\left\lfloor\frac{(gu+m+2)(gu+m+7)}{12}\right\rfloor\\ =&U(gu+m+2).\end{aligned}$$

第二个等号可由 $U(g+2)=\frac{(g+2)(g+7)-2}{12}$ 得到，其中 $g\equiv 0,3 \pmod{12}$。 ∎

3) 码的分类处理表格

在表 3.1 中，我们给出了利用一些良好的 4-GDD 和最优短码，再通过应用定理 3.18–3.20 来构造最优 $(n,6,4)_3$ 码的一般框架。注意到，当 g 给定时，m 的值遍历了所有模 g 代表元。对于码长为 $m+t$ 的小码，除了码长为 54 和 60，它们在表 3.2 给出外，其余都可以在引理 3.16 和附录中找到。由于大部分 4-GDD 的存在性需要 $u\geqslant 4$（除了 $m+t=41,60$ 的情况，此时需要 $u\geqslant 5$），这导致某些小码长的码无法依赖于 4-GDD 去构造，此时这些参数的码不能被表 3.1 覆盖。我们在表 3.2 的第二列中，列举了所有这些码的码长，并一一处理。码长 n 为斜体的码是通过应用定理 3.18–3.20 和表 3.2 的第三列中列出的其它型的 4-GDD 构造的。对于那些以粗体显示的码长，相应码字在附录中单独给出。除此以外，剩下的 22 个码长，它们的最优码个数尚未确定，其上下界在表 3.3 中给出。这里，表 3.3 中给出的下界可由附录 A.4 中的引理 A.1 得到。

表 3.1　通过应用定理 3.18–3.20 和适当型的 4-GDD 以及最优短码来构造最优 $(n,6,4)_3$ 码的一般框架。对于码长为 $m+t$ 的小码，除了码长为 54 和 60，它们在表 3.2给出外，其余都可以在引理 3.16 和附录中找到。而当 $m+t=41$ 或 60 时，由推论3.7知，对应的 u 需要满足 $u\geqslant 5$。

n	t	4-GDD 的型	短码码长 $m+t$	来源
$n\equiv 0 \pmod 3$	$t=0$	$36^u m^1, u\geqslant 4$	0, 6, 9, 15, 21, 27, 30, 33, 39, 48, 54, 60	定理 3.18
$n\equiv 1 \pmod 3$	$t=1$	$12^u m^1, u\geqslant 4$	1, 7, 10, 16	定理 3.19
$n\equiv 2 \pmod 3$	$t=2$	$24^u m^1, u\geqslant 4$	8, 11, 20, 23, 26, 29, 38, 41	定理 3.20

我们将本节的主要结果总结如下。

定理 3.21　令 $M=\{14,17,18,24,35,42,44,47,56,59,68,71,72,$

表 3.2　表 3.1 中由于 $u \leqslant 3$，或者 $m+t=41,60$ 时 $u \leqslant 4$ 的 4-GDD 不存在，导致一些较小的码长未能被覆盖，这些码长将列举在下面第二列。其中，码长 n 为斜体的码是通过应用定理 3.18–3.20 和下面第三列中列出的 4-GDD 构造的。而对于以粗体显示的码长，相应码字在附录中单独给出。除此以外，剩下的 22 个码长，它们的最优码个数尚未确定，其上下界在表 3.3 中给出。

n	小码码长	4-GDD 的型	构造
$n \equiv 0 \pmod 3$	18, 24, 42, **45**, **51**, *54*, **57**, *60*, **63**, *66*, *69*, 72, *75*, 78, *81*, 84, **87**, 90, **93**, 96, **99**, 102, *105*, *108*, **111**, *114*, *117*, *120*, **123**, *126*, *129*, *132*, *135*, *138*, *141*, *147*, *156*, *162*, *168*, *204*	$15^3 9^1$, 15^4, $15^4 6^1$, $15^4 9^1$, 15^5, $15^4 21^1$, $15^5 30^1$, 27^4, $15^7 9^1$, $27^4 9^1$, 15^8, $15^8 6^1$, $15^8 9^1$, $15^7 27^1$, 15^9, $15^7 33^1$, $15^8 21^1$, $15^8 27^1$, $15^8 36^1$, $39^4 6^1$, $15^8 48^1$, $15^{11} 39^1$	定理 3.18
$n \equiv 1 \pmod 3$	**19**, **22**, **25**, *28*, **31**, **34**, *37*, **40**, *43*, *46*, *52*	7^4, 9^4, $9^4 6^1$, 9^5,$9^5 6^1$	定理 3.19 (这里 $n=28$ 的码字应用定理 3.18)
$n \equiv 2 \pmod 3$	14, 17, **32**, 35, 44, 47, **50**, **53**, 56, 59, **62**, **65**, 68, 71, **74**, **77**, 80, 83, **86**, **89**, 92, 95, **98**, **101**, *110*, **113**, *137*	27^4, 27^5	定理 3.20 (具有性质 (B) 的最优 $(29,6,4)_3$ 码在表 A.3中给出)

表 3.3　码长 n 较小时，$A_3(n,6,4)$ 的上下界

n	14	17	18	24	35	42	44	47	56	59	68
下界	21	30	33	55	114	161	176	200	280	310	409
上界	22	31	34	58	116	164	179	203	284	314	413
n	71	72	78	80	83	84	90	92	95	96	102
下界	445	461	538	562	603	616	705	738	786	803	901
上界	449	462	539	566	608	623	712	743	791	808	909

78，80，83，84，90，92，95，96，102}。则对任意正整数 n，有

$$A_3(n,6,4)=\begin{cases} U(n)-1, & \text{若} n=3,4,5,12, \\ U(n), & \text{若} n \notin M \cup \{3,4,5,12\}. \end{cases}$$

当 $n \in M$ 时，$A_3(n,6,4)$ 的上下界在表 3.3 中给出。

3.5　距离 $d=2w-2$ 的三元常重码

在本节中，我们考虑权重为 w 和距离 $2w-2$ 的三元常重码，这里 w 为任意正整数。并在 n 足够大且满足特定条件时，基于图分解这一工具，我们确定了 $A_3(n,2w-2,w)$ 的值。和上一节不同，与 $w=3,4$ 的显式构造相比，我们没有给出当 $A_3(n,2w-2,w)$ 确定时，长度 n 的下界。值得一提的是，尽管当 w 增加时，码字型的数量也随之增加，但我们最终得到的最优码中，只有两种不同类型的码字：1^w 和 $1^{w-2}2^1$。

3.5.1　$(n, 2w-2, w)_3$ 常重码上界

下面, 我们将给出一些图分解的概念, 并利用这一工具, 给出 $A_3(n, 2w-2, w)$ 的上界。

给定一个没有孤立点的图 H，$\gcd(H)$ 表示 H 中所有顶点度的最大公约数。对一个图 G，若存在一个正整数 d，使得 $\gcd(G)$ 可以被 d 整除，则称图 G 是d-可除的；若 G 中不存在任何一个顶点，它的度为 d 所整除，则称图 G 是d-不可除的。设 $\mathcal{P} = \{G_1, \cdots, G_s\}$ 是图 G 的某些边不交的子图集合，若其中每个子图都同构于 H，则称 $\mathcal{P}$ 是 G 的一个H-填充。进一步地，若 $G = G_1 \cup \cdots \cup G_s$，则称 $\mathcal{P}$ 是 G 的一个H-分解。图 G 的H-填充数，指的是图 G 的 H-填充的最大基数，记为 $P(H, G)$。特别地，如果 $H = K_k$ 且 $G = K_n$，即 H 和 G 都是完全图时，$P(H, G)$ 和第 3.2.2 小节中介绍的填充数 $D(n, k, 2)$ 相同。在后续的构造中，我们主要用到的工具是 Alon 等人在文献 [57] 中给出的结果。

定理 3.22 ([57])　给定一个图 H，它有 h 条边，令 $\gcd(H) = e$。则存在实数 $N = N(H)$，和 $\varepsilon = \varepsilon(H)$，使得对任意 e-可除的或者 e-不可除的图 $G = (V, E)$，满足 $|V| = n > N(H)$ 以及 $\delta(G) > (1 - \varepsilon(H))n$，有，

$$P(H, G) = \left\lfloor \frac{\sum_{v \in V} \alpha_v}{2h} \right\rfloor,$$

除去当 G 是 e-可除的且 $0 < |E| \pmod h \leqslant \frac{e^2}{2}$，此时

$$P(H, G) = \left\lfloor \frac{\sum_{v \in V} \alpha_v}{2h} \right\rfloor - 1.$$

这里，对任意 $v \in V$，若 $d_G(v) = ae + b$，其中 a, b 为非负整数且 $0 \leqslant b < e$，则 $\alpha_v = ae$，即 α_v 指的是顶点 v 的度数向下取舍到最大的 e 的倍数。

考虑一个重量为 w 的三元常重码 $\mathcal{C} \subset I_3^n$，则一定存在某个 $y \in \left[0, \left\lfloor \frac{w}{2} \right\rfloor\right]$ 和非负整数 x，满足 $x + 2y = w$，使得 $\mathcal{C}$ 中的任一码字的型为 $1^x 2^y$。根据 **UNC** 条件和式 (3.1) 中的距离公式可知，码 $\mathcal{C}$ 是一个 $(n, 2w-2, w)_3$ 常重码当且仅当下面的两个条件成立:

($1''$) 对于 $\mathcal{C}$ 中所有码字, 其支集集合构成一个 2-$(n, \{\left\lceil \frac{w}{2} \right\rceil, \left\lceil \frac{w}{2} \right\rceil + 1, \cdots, w\}, 1)$-填充。

($2''$) 对于 $\mathcal{C}$ 中任意两个码字 $\mathbf{u}, \mathbf{v}$，若 $i \in supp(\mathbf{u}) \cap supp(\mathbf{v})$，则 $\min\{\mathbf{u}_i, \mathbf{v}_i\} = 1$。

引理 3.23　$A_3(n, 2w-2, w) \leqslant \left\lfloor \frac{n(n-1-(w-1)(w-2))}{w(w-1)} \right\rfloor + n$。

证明　对任一 $y = 0, 1, \cdots, \left\lfloor \frac{w}{2} \right\rfloor$，令 β_y 表示型为 $1^x 2^y$ 的码字个数。由条件 ($1''$) 和 ($2''$) 可得,

$$\binom{w}{2}\beta_0 + \binom{w-1}{2}\beta_1 + \cdots + \binom{w - \left\lfloor \frac{w}{2} \right\rfloor}{2}\beta_{\left\lfloor \frac{w}{2} \right\rfloor} \leqslant \binom{n}{2} \tag{3.2}$$

和

$$\beta_1 + 2\beta_2 + \cdots + \left\lfloor \frac{w}{2} \right\rfloor \beta_{\left\lfloor \frac{w}{2} \right\rfloor} \leqslant n. \tag{3.3}$$

注意到 $\binom{w-t}{2} + t(w-1) = \binom{w}{2} + \binom{t}{2}$。计算 (3.2) + $(w-1)$(3.3) 可得,

$$\beta_0 + \beta_1 + \cdots + \beta_{\left\lfloor \frac{w}{2} \right\rfloor} \leqslant \left\lfloor \frac{n\,(n-1-(w-1)(w-2))}{w(w-1)} \right\rfloor + n,$$

故证明完成。 ∎

当 $w = 3$ 和 4 时，引理 3.23 中给出的上界与第3.4节中的相同。为方便起见，令

$$B(n) := \left\lfloor \frac{n\,(n-1-(w-1)(w-2))}{w(w-1)} \right\rfloor.$$

我们将利用定理 3.22 证明，对于某些特定的 n，上界 $B(n)+n$ 是可以达到的。且达到上界的码 $\mathcal{C}$ 中只有两种类型的码字：1^w 和 $1^{w-2}2^1$。考虑顶点集为 $\mathbb{Z}_n$ 的完全图 K_n。对于每个码字 $\mathbf{u} \in \mathcal{C}$，可将其视为顶点集为 $supp(\mathbf{u})$ 的完全图，故它是 K_n 的一个子图。事实上，这种对应是一个单射，因为我们不确定在支集 $supp(\mathbf{u})$ 上的元素是 1 还是 2。由条件 (1″) 可知，所有这些子图两两都是边不相交的，因此形成了 K_n 的一个填充。需要注意的是所有包含元素 2 的码字，由条件 (2″) 知，这些 2 的坐标应该互不相同。在下一小节中，我们将借助一个新的组合结构 Golomb 尺 [79]，去寻找这样的码字。

3.5.2　基于图分解的 $(n, 2w-2, w)_3$ 常重码构造

本小节中，我们将借助 Golomb 尺这一组合结构，找到型为 $1^{w-2}2^1$ 的码字。然后再利用定理 3.22，构造出目标码。换句话说，我们首先找到一些符合要求的阶为 $w-1$ 的完全子图，不妨设为 n' 个，从 K_n 中将这些子图删去，记剩下的图为 G'，再对图 G' 找到一个最大的 K_w-填充即可。这个填充中的每一个子图都对应一个型为 1^w 的码字，共有 $P(K_w, G')$ 个。综上，最终目标码字个数为 $n' + P(K_w, G')$。而选择合适的参数 n'，可使得 $n' + P(K_w, G') = B(n) + n$。

一个 (n, w) 模 n *Golomb* 尺（modular Golomb ruler）指的是一个由 w 个整数构成的集合 $\{a_1, a_2, \cdots, a_w\}$，使得其两两之间的差 $\{a_i - a_j \mid 1 \leqslant i \neq j \leqslant w\}$ 在模 n 的意义下非零且互不相同。设我们有一个 $(n, w-1)$ 模 n 哥伦布尺 $\{a_1, a_2, \cdots, a_{w-1}\}$，可将其看成一个基码字。则由其可生成 n 个型为 $1^{w-2}2^1$ 的码字 $\{(a_1+i)_2, (a_2+i)_1, \cdots, (a_{w-1}+i)_1\}, i \in \mathbb{Z}_n$。不难验证，这 n 个码字两两之间的距离至少为 $2w-2$，具体可用反证法，找到两个不同的点对，它们的差相同，导出矛盾。将这 n 个码字看成 n 个完全图 K_{w-1}，每一个对应点集 $\{a_1+i, a_2+i, \cdots, a_{w-1}+i\}$，$i \in \mathbb{Z}_n$，可知它们是 K_n 的边不交的子图，这里取 K_n 的顶点集为 $\mathbb{Z}_n$。令 $S = nK_{w-1}$，

即 S 是这些子图的并集。可以看出，S 是 K_n 的一个正则子图，即 S 中每个顶点的度均相同，为 $(w-1)(w-2)$。记 $G = K_n \setminus S$。我们将应用定理 3.22 找到 G 的一个 K_w-填充，从而得到所有型为 1^w 的码字。

定理 3.24　令 $w \geqslant 3$ 是一个给定的正整数。则对任意充分大的正整数 n，若满足 $n \equiv 1 \pmod{w-1}$，则 $A_3(n, 2w-2, w) \geqslant B(n) + n - 1$。进一步地，若 $n \equiv 1, w, -w+2, -2w+3 \pmod{w(w-1)}$，则 $A_3(n, 2w-2, w) = B(n) + n$。

证明　由文献 [79] 知，对任意 $n = \Omega(w^2)$，存在一个 $(n, w-1)$ 模 n 哥伦布尺。因此，根据上面的分析，我们可以得到 n 个型为 $1^{w-2}2^1$ 的码字，和一个正则图 $G = (V, E)$，其中顶点集 $V = \mathbb{Z}_n$，度为 $d = n - 1 - (w-1)(w-2)$。令 K_w 为定理 3.22 中的 H，则 $e = \gcd(H) = w - 1$ 和 $h = \frac{w(w-1)}{2}$。由于 n 充分大，可知 $d > (1-\varepsilon)n$，这里 $\varepsilon = \varepsilon(K_w)$ 在定理 3.22 中给出。进一步地，$n \equiv 1 \pmod{w-1}$ 可推出 $d \equiv 0 \pmod{w-1}$，即 G 是 $(w-1)$-可除的，且对任意 $v \in V$，$\alpha_v = d = n - 1 - (w-1)(w-2)$。应用定理 3.22，我们可以得到 G 的一个最大 K_w-填充，其相应的填充数满足

$$\begin{aligned} P(K_w, G) &\geqslant \left\lfloor \frac{\sum_{v \in V} \alpha_v}{2\binom{w}{2}} \right\rfloor - 1 \\ &= \left\lfloor \frac{n\,(n - 1 - (w-1)(w-2))}{w(w-1)} \right\rfloor - 1 \\ &= B(n) - 1. \end{aligned}$$

该填充里的每一个 K_w 都可以自然的看成一个型为 1^w 的码字。因此，我们得到了至少 $B(n)-1$ 个码字。结合由模 n 哥伦布尺生成的 n 个型为 $1^{w-2}2^1$ 的码字，我们得到了一个码字大小至少为 $B(n) + n - 1$ 的 $(n, 2w-2, w)_3$ 常重码。

当 $n \equiv w, -2w+3 \pmod{w(w-1)}$ 时，容易验证 $|E| = nd/2 \equiv 0 \pmod{\frac{w(w-1)}{2}}$。事实上，可以证明这是唯一满足 $|E| \pmod{\frac{w(w-1)}{2}} \notin [1, \frac{(w-1)^2}{2}]$ 的两个同余类。由定理 3.22 可知，此时 $P(K_w, G) = B(n)$，故 $A_3(n, 2w-2, w) = B(n) + n$。

当 $n \equiv 1, -w+2 \pmod{w(w-1)}$ 时，我们可为 K_n 取不同的点集，从而构造出另一种达到上界的最优码。令 S' 是一个顶点集为 $\mathbb{Z}_{n-1}$、度为 $(w-1)(w-2)$ 的正则图，即 S' 和 S 类似，是由一个 $(n-1, w-1)$ 模 $n-1$ 哥伦布尺生成的 $n-1$ 个子图 K_{w-1} 的并集。换句话说，S' 可生成 $n-1$ 个型为 $1^{w-2}2^1$ 的码字。令 $G' = K_n \setminus S' = (V', E')$，其顶点集 $V' = \mathbb{Z}_{n-1} \cup \{\infty\}$。注意到，对任意顶点 $v \in \mathbb{Z}_{n-1}$，$d_G(v) = n-1-(w-1)(w-2)$，而 $d_G(\infty) = n-1$。因此图 G' 是 $(w-1)$-可除的。进一步地，容易验证

$$
\begin{aligned}
|E'| &= \frac{(n-1)\,(n-1-(w-1)(w-2))+n-1}{2} \\
&\equiv 0 \pmod{\frac{w(w-1)}{2}}.
\end{aligned}
$$

由定理 3.22，我们有

$$
\begin{aligned}
P(K_w, G') &= \left\lfloor \frac{\sum_{v\in V}\alpha_v}{2\binom{w}{2}} \right\rfloor \\
&= \frac{n\,(n-1-(w-1)(w-2))+(w-1)(w-2)}{w(w-1)} \\
&= B(n)+1.
\end{aligned}
$$

最后一个等式成立是因为在这种情况下，$B(n) = \frac{n(n-1-(w-1)(w-2))-2(w-1)}{w(w-1)}$。因此，当 $n \equiv 1, -w+2 \pmod{w(w-1)}$ 时，我们有 $A_3(n, 2w-2, w) = n-1+P(K_w, G') = B(n)+n$。∎

定理 3.24 表明，当 n 充分大且满足 $n \equiv 1, w, -w+2, -2w+3 \pmod{w(w-1)}$ 时，$A_3(n, 2w-2, w) = B(n)+n$。特别地，当 $w = 4$ 且 $n \equiv 1 \pmod 3$ 时，$A_3(n, 6, 4) = B(n)+n = U(n)$；而当 $w = 3$ 且 $n \equiv 1 \pmod 2$ 时，$A_3(n, 4, 3) = B(n)+n = \left\lfloor \frac{n^2+3n}{6} \right\rfloor$。这表明，当 $w \leqslant 4, n \equiv 1 \pmod{w-1}$ 且 $d = 2w-2$ 时，本小节给出的结果与前面的一致，但整体来说，还是弱一些，这是因为前面我们证明了除去少数 $w = 4$ 的 n 外，$A_3(n, 2w-2, w) = B(n)+n$ 对所有 n 以及 $w \leqslant 4$ 均成立。

3.6　$(n, 2w-2, w)_3$ 常重码的进一步优化

在文献 [80] 中，我们进一步的考虑了权重为 w 和距离为 $2w-2$ 的三元常重码，并确定了，对任意充分大的 n，有 $A_3(n, 4, 3) = B(n)+n$ 成立。主要方法是，将码长分成 $w-1$ 个同余类，即 $n \equiv s \pmod{w-1}$，其中 $s = 0, 1, \cdots, w-2$。当 $n \equiv 0, 1 \pmod{w-1}$ 时，构造方法与第 3.5.2 小节类似，在选择图 S 的时候，其顶点集为 $\mathbb{Z}_{n'}$，而 $n = n'+k$，这里 k 远小于 n。例如在定理 3.24 给出的第二个构造中，$k = 1$。而当 $n \equiv s \pmod{w-1}$，其中 $s \neq 0, 1$ 时，我们主要是找到了一个有效的算法证明了 $A_3(n, 4, 3) = B(n)+n$ 成立。值得一提的是，该算法是基于一个码长满足 $n \equiv s \pmod{w-1}$ 的最优码运行的，这里 $\lceil \frac{w}{2} \rceil \leqslant s \leqslant w-2$。而这种码在 $w = 4$ 时，可以用来构造第 3.4.2 小节中，表 3.2 中提到的码。

3.7　本章总结

在本章中，我们确定了 ℓ_1 度量下在非负整数或集合 $I_3 = \{0,1,2\}$ 上的，所有重量 $w \leqslant 4$ 的常重码最优码字大小。对于一般的 w，我们还提供了权重为 w 和距离为 $2w-2$ 的三元码的构造，且当 n 充分大时，若 $n \equiv 1 \pmod{w-1}$，所得码字个数至少为 $B(n)+n-1$，进一步地若 $n \equiv 1, w, -w+2, -2w+3 \pmod{w(w-1)}$，我们确定了 $A_3(n,4,3) = B(n)+n$。此外，在非负整数上、或集合 I_q（$q \geqslant 4$）上针对一般的重量 w 构造最优码是一个更具挑战性的问题，值得进一步地研究。

第 4 章　扭 Reed-Solomon 码和自对偶码

本章我们主要研究扭 Reed-Solomon 码的构造性问题，这类码字在密码系统中有着重要应用。在第 4.1 节中，我们简单的介绍了扭 RS 码和相关自对偶码的研究背景，以及本章贡献。在第 4.2 节中，我们给出了 TRS 码、TGRS 码、MDS 码、NMDS 码以及自对偶码等一些码的定义，并提供了关于这些码现有的一些结果。在第 4.3 节中，我们对具有对偶封闭性的 TGRS 码进行了刻画。在第 4.4 节中，我们给出了具有对偶封闭性的 TGRS 码的一些构造，并借助它们构造了一些自对偶码。最后在第 4.5 节中对本章进行了简单的总结。

4.1　介绍

扭（Twisted）Reed-Solomon（TRS）码是由 Beelen，Puchinger 和 Rosenkilde né Nielsen[9] 在 2017 年首次提出的，它是 RS 码的一种推广，主要是受到扭 Gabidulin 码 [10] 的启发。RS 码和 TRS 码都是多项式码。更具体地说，有限域 $\mathbb{F}_q$ 上的一个 $[n,k]$ RS 码可以看作是 $\mathbb{F}_q$ 上的所有度至多为 $k-1$ 的多项式集合。与 RS 码不同的是，一个 $[n,k]$ TRS 码，它的每个多项式中允许出现一个次数大于 $k-1$ 的单项式 $a_h x^{k-1+t}$，其中 $h \in [0, k-1], t > 0$ 是两个给定的整数，a_h 是该多项式的 h 次项系数，称 h 为钩，t 为扭结。在文献 [11] 中，作者进一步给出了 TRS 码的一种推广，从添加一个扭结推广至 $\ell \geqslant 1$ 个扭结。类似地，通过在广义 RS（GRS）码中加入扭结，可以得到广义 TRS（TGRS）码。选择适当的扭结和估值点，可使 TGRS 码达到 Singleton 界，即使其成为 MDS 码。但一般情况下，TGRS 码并不一定是 MDS 码。

TGRS 码作为 RS 码的一种推广，它既具有一些与 RS 码类似的良好性质，例如解码算法快、结构清晰（在给定条件下，对偶码也可确定）等，又具有一些 RS 码不具备的特性，例如可通过控制 t, ℓ 来达到某些指定的结构。这使得它可以应用于一些特殊场景。其中比较著名的是，可以将其应用于 McEliece 密码系统 [18]。这是一种公钥密码系统，应用于其中的码需要配备一个高效的解码算法。在文献 [9] 的定理 18 中，作者说明了大多数 TGRS 码与 RS 码并不等价（尽管该定理证明有误，但在 [53] 的引理 2.7 中给出了正确的版本）。进一步地，他们在 $\mathbb{F}_q$ 上构造出了两种不等价于 RS 码的 TGRS MDS 码，这里 $q \geqslant 11$, 其码长 n 满足 $n \geqslant \frac{q}{2}$。这两种 TGRS 码可用于 McEliece 密码系统，因为它们不仅具有高效的解码算法，而且还能够抵抗一些针对 (G)RS 类型码 [11] 的代数攻击 [13]。尽管在文献 [81] 中，作者提出了对 McEliece 密码系统中使用的 TGRS 变体码的有

效密钥恢复攻击，但已经被文献 [11] 的作者们修复。因此，构造不等价于 RS 码的 TGRS MDS 码，且能够抵抗针对 (G)RS 类型码以及一些 TGRS 变体码的攻击是非常有必要的。

基于其代数结构，TGRS 码也可以用来构造编码理论中其它比较重要的码。注意到，一个 $[n,k]$ TGRS 码，它的最小距离不超过 $n-k+1$，且如果 $t=1$，此时所有多项式的度数最多为 k，那么最小距离至少为 $n-k$。因此这种类型的 TGRS 码要么是几乎（almost）MDS（AMDS）码，要么是 MDS 码。

由于 GRS 码在对偶下是封闭的，即一个 GRS 码的对偶码仍是 GRS 码，我们想知道 TGRS 码是否也具有相同的性质。在 [11] 中，Beelen 等人证明了，如果 TGRS 码的估值点集合是一个 $\mathbb{F}_q^*$ 的乘法子群，则 TGRS 码在对偶下是封闭的。本章中，我们的结果放宽了这个条件，只要估值点集合是某个特殊多项式的根集合即可，并给出了相应的校验矩阵，且利用这一结果，构造了一些自对偶码。从这个角度来看，Beelen 等人的结果中，可以将估值点集合看成是多项式 x^t-a 的根集合。因此，构造自对偶 MDS 或 NMDS 码可转为找寻某些特殊多项式。值得一提的是，文献 [50, 52] 中所用的方法，本质上也是与此类似。

4.2　预备知识

本节中，将给出一些必要的定义以及 TGRS 码的一些相关结果。我们记 $\mathbb{F}_q^* = \mathbb{F}_q \setminus \{0\}$。

4.2.1　扭 Reed-Solomon 码

定义 4.1　给定有限域 $\mathbb{F}_q$，以及正整数 n,k,ℓ。令 $\boldsymbol{t} = (t_1,\cdots,t_\ell), \boldsymbol{h} = (h_1,\cdots,h_\ell) \in \mathbb{Z}_{>0}^{\ell}$，且满足以下条件：

- $t_1,\cdots,t_\ell$ 互不相同且对任意 $i\in[\ell]$，有 $1\leqslant t_i \leqslant n-k$；
- $h_1,\cdots,h_\ell$ 互不相同且对任意 $i\in[\ell]$，有 $0\leqslant h_i < k$。

令 $\boldsymbol{\eta} = (\eta_1,\cdots,\eta_\ell) \in (\mathbb{F}_q^*)^\ell$。则 $(\boldsymbol{t},\boldsymbol{h},\boldsymbol{\eta})$-扭多项式集合定义如下

$$\mathcal{P}_{\boldsymbol{t},\boldsymbol{h},\boldsymbol{\eta}}^{n,k} = \left\{ \sum_{i=0}^{k-1} f_i x^i + \sum_{j=1}^{\ell} \eta_j f_{h_j} x^{k-1+t_j} \mid f_i \in \mathbb{F}_q \right\}.$$

容易看出 $\mathcal{P}_{\boldsymbol{t},\boldsymbol{h},\boldsymbol{\eta}}^{n,k} \subseteq \mathbb{F}_q[x]$ 是一个 $\mathbb{F}_q$ 上的 k 维线性空间，其中的一组基为 $\{x^{h_i} + \eta_i x^{k-1+t_i} : i\in[\ell]\} \cup \{x^j : j\in\{0,1,\cdots,k-1\}\setminus\{h_1,\cdots,h_\ell\}\}$。我们称 $h_1,\cdots,h_\ell$ 为钩，$t_1,\cdots,t_\ell$ 为扭结。

令 $\alpha_1,\cdots,\alpha_n$ 为 $\mathbb{F}_q$ 中 n 个互不相同的元素，称其为*估值点*，记为 $\boldsymbol{\alpha} =$

$(\alpha_1, \cdots, \alpha_n)$。一个估值点为 $\boldsymbol{\alpha}$ 的$(\boldsymbol{t}, \boldsymbol{h}, \boldsymbol{\eta})$-*扭 RS 码*（TRS 码）定义如下：

$$C^{n,k}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta}) = \{(f(\alpha_1), \cdots, f(\alpha_n)) : f \in \mathcal{P}^{n,k}_{\boldsymbol{t},\boldsymbol{h},\boldsymbol{\eta}}\}.$$

一个估值点为 $\boldsymbol{\alpha}$ 的广义 TRS 码（TGRS 码）指的是

$$C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta}) = \{(v_1 f(\alpha_1), \cdots, v_n f(\alpha_n)) : f \in \mathcal{P}^{n,k}_{\boldsymbol{t},\boldsymbol{h},\boldsymbol{\eta}}\},$$

其中，$\boldsymbol{v} = (v_1, \cdots, v_n)$ 且对任意 $i \in [n]$，满足 $v_i \in \mathbb{F}_q^*$。

为了方便叙述，我们有时不区分码字 $(f(\alpha_1), \cdots, f(\alpha_n))$ 或 $(v_1 f(\alpha_1), \cdots, v_n f(\alpha_n))$ 和多项式 f，它们两两本质上分别指代的是同一个对象。实际上，当 $\ell = 0$ 时，$C^{n,k}(\boldsymbol{\alpha}; 0; 0; \boldsymbol{\eta})$ 是一个 $[n, k]$RS 码，$C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 0; 0; \boldsymbol{\eta})$ 是一个 $[n, k]$GRS 码。且当 $\boldsymbol{v} = (1, 1, \cdots, 1)$ 时，$C^{n,k}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta}) = C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$，故在本章剩余部分中，我们有时只考虑 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$。

注意到，码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 是定义在多项式集合 $\mathcal{P}^{n,k}_{\boldsymbol{t},\boldsymbol{h},\boldsymbol{\eta}}$ 上的，由该集合的一组基，可以自然的给出码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 的生成矩阵，为 $G = [\boldsymbol{I}|\boldsymbol{L}] \cdot \boldsymbol{V} \cdot \mathrm{diag}(\boldsymbol{v})$。其中 $\boldsymbol{I}$ 是一个单位阵，且在本章不同位置，具有恰当的阶数。而矩阵 $\boldsymbol{L}$ 满足

$$L_{i,j} = \begin{cases} \eta_\mu, & (i, j) = (h_\mu + 1, t_\mu), \\ 0, & \text{其它}. \end{cases}$$

$\boldsymbol{V}$ 是一个由 $\boldsymbol{\alpha}$ 定义的 $k \times n$ 阶 Vandermonde 矩阵。本章剩余部分，我们将继续沿用这些符号。

4.2.2　MDS 或近 MDS 码及对偶码

给定一个 $[n, k, d]$ 线性码 $\mathcal{C}$，以及它的对偶码 $[n, n-k, d^\perp]$，记为 $\mathcal{C}^\perp$。本章中，如果不特地指代的话，对偶码指的是基于欧几里得内积下的。回顾一下，当 $d = n - k + 1$ 时，$\mathcal{C}$ 被称为 MDS 码；当 $d = n - k$ 时，$\mathcal{C}$ 被称为*几乎 MDS 码*（AMDS 码）；而当 $d + d^\perp = n$ 时，称 $\mathcal{C}$ 为*近 MDS 码*（NMDS 码）。事实上，由 Singleton 界可知，当 $d + d^\perp = n$ 时，有 $d = n - k$ 以及 $d^\perp = k$，因此一个 NMDS 码也是一个 AMDS 码。如果 $\mathcal{C} = \mathcal{C}^\perp$，则称 $\mathcal{C}$ 是*自对偶码*。

给定一个 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$，注意到，当 $\ell = 1$ 且 $t = 1$ 时，$C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; h; \eta)$ 中的每个多项式度至多为 k。若 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; h; \eta)$ 中存在一个重量至多为 $n - k - 1$ 的码字，换句话说，其对应的多项式在 $\mathbb{F}_q$ 上至少有 $k + 1$ 个根，矛盾。因此，$C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; h; \eta)$ 的最小距离在 $n - k$ 和 $n - k + 1$ 之间，即，它要么是一个 MDS 码，要么是一个 AMDS 码。特别地，如果 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; h; \eta)$ 对偶封闭的话，则前面的“AMDS”可以换成“NMDS”。

为了方便后续使用，下面提及一些关于自对偶的 TGRS 码的结果。

引理 4.1 ([50]引理 2.4)　给定集合 $S_k = \{\sum_{i\in I}\alpha_i : \text{对任意} I \subset [n] \text{且} |I| = k\}$，则有

- $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; k-1; \eta)$ 是一个 MDS 码当且仅当 $-\eta^{-1} \notin S_k$。
- $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; k-1; \eta)$ 是一个 NMDS 码当且仅当 $-\eta^{-1} \in S_k$。

定理 4.2 ([50]定理 2.5)　给定一个正整数 k。令 $n = 2k$，$\boldsymbol{G} = [\boldsymbol{I}|\boldsymbol{L}] \cdot \boldsymbol{V} \cdot \mathrm{diag}(\boldsymbol{v})$ 是码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; k-1; \eta)$ 的一个生成矩阵。则 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; 1; k-1; \eta)$ 是一个自对偶码当且仅当下面的条件成立：

1) 存在一个非零元素 $\lambda \in \mathbb{F}_q^*$，使得对任意 $i \in [n]$，有 $v_i^2 = \lambda P_i(\alpha_i)^{-1}$，这里 $P_i(\alpha_i) = \prod_{j\in[n], j\neq i}(\alpha_i - \alpha_j)$。
2) $\eta = -\frac{\eta}{1+\eta b_1}$，其中 $b_1 = \sum_{i=1}^n \alpha_i \neq 0$ 以及 $\eta \neq -b_1^{-1}$。

4.3　扭 Reed-Solomon 码的结构特性

在本节中，我们考虑对偶封闭的 TGRS 码。众所周知，（广义）RS 码在对偶下是封闭的，即其对偶码仍然是（广义）RS 码。而一般情况下，TGRS 码可能不是封闭的。

4.3.1　估值点是陪集的扭 Reed-Solomon 码

在文献 [11] 的定理 2 中，有如下结论：如果一个 TGRS 码，它的所有估值点形成一个乘法群，则其对偶码也仍然是一个 TGRS 码。事实上，本小节中，我们将证明，可将乘法群这个条件放宽至某个乘法群的陪集。在证明该结论之前，我们有以下引理，它是 [82] 的一个简单推广。

引理 4.3　给定一个有限域 $\mathbb{F}_q$ 上的多项式 $x^n - a^n$，其中 $a \neq 0$。如果 $(n, q) = 1$，且 $\alpha_1, \cdots, \alpha_n$ 是其所有的根，则有

$$\begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \cdots & \alpha_n^{n-1} \end{bmatrix}^{-1} = \frac{1}{n}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1^{-1} & \alpha_2^{-1} & \cdots & \alpha_n^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{-(n-1)} & \alpha_2^{-(n-1)} & \cdots & \alpha_n^{-(n-1)} \end{bmatrix}.$$

证明简要　由于 $\frac{\alpha_i^n}{\alpha_j^n} = \frac{\alpha_i^n/a^n}{\alpha_j^n/a^n} = 1$，故有

$$\frac{1}{n}\left(1 + \frac{\alpha_i}{\alpha_j} + \cdots + \frac{\alpha_i^{n-1}}{\alpha_j^{n-1}}\right) = \frac{1}{n}\left(\frac{\alpha_i^n/\alpha_j^n - 1}{\alpha_i/\alpha_j - 1}\right) = \delta_{i,j},$$

这里 $\delta_{i,j}$ 是一个示性函数。 ∎

令 $\alpha_1, \cdots, \alpha_n$ 构成的集合是 $\mathbb{F}_q$ 对于某一个阶为 n 的乘法子群的陪集，其陪集首为 a，即对任意 $i \in [n]$，有 $\alpha_i^n = a^n$。则由引理 4.3，有

$$\begin{bmatrix} 1 & \alpha_1 & \cdots & \alpha_1^{n-1} \\ 1 & \alpha_2 & \cdots & \alpha_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \alpha_n & \cdots & \alpha_n^{n-1} \end{bmatrix}^{-1} = \frac{1}{n}\boldsymbol{J} \cdot \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_n^{n-1} \end{bmatrix} \cdot \mathrm{diag}(\boldsymbol{\alpha}/a^n),$$

其中，$\boldsymbol{J}$ 是一个反对角线上元素全为 1、其它所有位置上的元素为 0 的 $n \times n$ 阶矩阵。在本章中，$\boldsymbol{J}$ 在不同的情形下，具有适当的阶数。类似地，利用文献 [11] 定理 2 中的方法，应用引理 4.3，即可得如下定理。

定理 4.4　令 $\alpha_1, \cdots, \alpha_n$ 构成的集合是 $\mathbb{F}_q$ 对于某一个阶为 n 的乘法子群的陪集，$C^{n,k}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 是一个 TRS 码，满足 $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_n)$。则，$C^{n,k}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 的对偶码是一个 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; k - \boldsymbol{h}; n - k - \boldsymbol{t}; -\boldsymbol{\eta})$，这里，$\boldsymbol{v} = (v_1, \cdots, v_n)$ 且对任意 $i \in [n]$，$v_i \in \mathbb{F}_q^*$ 为某个固定元素。

在定理 4.4中，$C^{n,k}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 自然的有一个生成矩阵 $\boldsymbol{G} = [\boldsymbol{I}|\boldsymbol{L}] \cdot \boldsymbol{V}$，不妨设对任意 $i \in [n]$，有 $\alpha_i^n = a^n$。那么容易验证，$C^{n,k}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 有一个校验矩阵 $\boldsymbol{H} = [\boldsymbol{I}| - \boldsymbol{J}\boldsymbol{L}^T\boldsymbol{J}] \cdot \boldsymbol{V} \cdot \mathrm{diag}(\boldsymbol{\alpha}/na^n)$，即 $\boldsymbol{G} \cdot \boldsymbol{H}^T = 0$，我们将在下一小节，详细阐述这一点。

4.3.2　估值点是多项式根集合的扭 Reed-Solomon 码

本小节中，我们将定理 4.4中的结果，由所有估值点构成一个陪集推广至构成某个多项式的根集合。

事实上，当 $\alpha_1, \cdots, \alpha_n$ 构成一个陪集时，它也是多项式 $x^n - a^n$ 的根集合，这里元素 a 是陪集首。注意到，该多项式中，除去最高项和常数项系数不为 0 之外，其余均为 0。这促使我们去考虑一个更一般的多项式，它的根集合由所有估值点构成，且除去最高项和常数项系数不为 0 之外，其余项系数几乎都为 0。

对任意 $i \in [n]$，令 $P_i(x) = \prod_{j \in [n], j \neq i}(x - \alpha_j)$。考虑矩阵 $\boldsymbol{E} = (E_{i,j})_{n \times n}, \boldsymbol{P} = \mathrm{diag}(P(\boldsymbol{\alpha}))$，其中 $E_{i,j} = (-1)^{n-j} e_{n-j}(\{\alpha_1, \cdots, \alpha_n\} \setminus \{\alpha_i\})$ 且 $P(\boldsymbol{\alpha}) = (P_1(\alpha_1), \cdots, P_n(\alpha_n))$。这里 $e_k(x)$ 是初等对称函数。因此，我们有 $\boldsymbol{P} = \boldsymbol{E} \cdot \boldsymbol{V}$，以及 $\boldsymbol{V}^{-1} = \boldsymbol{P}^{-1}\boldsymbol{E}$。进一步地，有

$$\boldsymbol{V}\cdot\boldsymbol{V}^{-1}=\begin{bmatrix}1 & 1 & \cdots & 1\\ \alpha_1 & \alpha_2 & \cdots & \alpha_n\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_n^{n-1}\end{bmatrix}\begin{bmatrix}P_1(\alpha_1)^{-1}E_{1,1} & P_1(\alpha_1)^{-1}E_{1,2} & \cdots & P_1(\alpha_1)^{-1}E_{1,n}\\ P_2(\alpha_2)^{-1}E_{2,1} & P_2(\alpha_2)^{-1}E_{2,2} & \cdots & P_2(\alpha_2)^{-1}E_{2,n}\\ \vdots & \vdots & \ddots & \vdots\\ P_n(\alpha_n)^{-1}E_{n,1} & P_n(\alpha_n)^{-1}E_{n,2} & \cdots & P_n(\alpha_n)^{-1}E_{n,n}\end{bmatrix} \tag{4.1}$$

$$=\boldsymbol{I}, \tag{4.2}$$

在后续的一些计算中，我们将多次使用式 (4.1) 的最后一列。

考虑多项式

$$m(x)=\prod_{i\in[n]}(x-\alpha_i)=x^n+\sum_{j=0}^{n-1}m_jx^j.$$

则有 $m(\alpha_i)=0=\alpha_i^n+\sum_{j=0}^{n-1}m_j\alpha_i^j$，进一步可得 $\alpha_i^n=-\sum_{j=0}^{n-1}m_j\alpha_i^j$ 以及 $\alpha_i^{n+k}=-\sum_{j=0}^{n-1}m_j\alpha_i^{j+k}$。定义

$$b_k=\begin{cases}\sum_{i=1}^{n}P_i(\alpha_i)^{-1}\alpha_i^{n+k-1}, & k\in[n-1],\\ 0, & \text{如果}k<1.\end{cases}$$

因此我们有

$$\boldsymbol{V}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{V}^T=\begin{bmatrix}1 & 1 & \cdots & 1\\ \alpha_1 & \alpha_2 & \cdots & \alpha_n\\ \vdots & \vdots & \ddots & \vdots\\ \alpha_1^{n-1} & \alpha_2^{n-1} & \cdots & \alpha_n^{n-1}\end{bmatrix}\begin{bmatrix}P_1(\alpha_1)^{-1} & P_1(\alpha_1)^{-1}\alpha_1 & \cdots & P_1(\alpha_1)^{-1}\alpha_1^{n-1}\\ P_2(\alpha_2)^{-1} & P_2(\alpha_2)^{-1}\alpha_2 & \cdots & P_2(\alpha_2)^{-1}\alpha_2^{n-1}\\ \vdots & \vdots & \ddots & \vdots\\ P_n(\alpha_n)^{-1} & P_n(\alpha_n)^{-1}\alpha_n & \cdots & P_n(\alpha_n)^{-1}\alpha_n^{n-1}\end{bmatrix}$$

$$=\begin{bmatrix}0 & 0 & \cdots & 0 & 0 & 1\\ 0 & 0 & \cdots & 0 & 1 & b_1\\ 0 & 0 & \cdots & 1 & b_1 & b_2\\ \vdots & \vdots & \iddots & \vdots & \vdots & \vdots\\ 0 & 1 & \cdots & b_{n-4} & b_{n-3} & b_{n-2}\\ 1 & b_1 & \cdots & b_{n-3} & b_{n-2} & b_{n-1}\end{bmatrix}. \tag{4.3}$$

注意到，对任意 $i\in[n]$，都有 $E_{i,n}=1$。由式 (4.1) 的最后一列，可得

$$b_1=\sum_{i=1}^{n}P_i(\alpha_i)^{-1}\alpha_i^n=-\sum_{i=1}^{n}P_i(\alpha_i)^{-1}\sum_{j=0}^{n-1}m_j\alpha_i^j$$

$$=-\sum_{j=0}^{n-1}m_j\sum_{i=1}^{n}P_i(\alpha_i)^{-1}\alpha_i^j=-m_{n-1}.$$

更一般地，

$$\begin{aligned} b_{k+1} &= \sum_{i=1}^{n} P_i(\alpha_i)^{-1}\alpha_i^{n+k} \\ &= -\sum_{j=0}^{n-1} m_j \sum_{i=1}^{n} P_i(\alpha_i)^{-1}\alpha_i^{j+k} \\ &= -\{m_{n-k-1} + m_{n-k}(\sum_{i=1}^{n} P_i(\alpha_i)^{-1}\alpha_i^{n}) + \cdots + m_{n-1}(\sum_{i=1}^{n} P_i(\alpha_i)^{-1}\alpha_i^{n+k-1})\} \\ &= -(m_{n-k-1} + m_{n-k}b_1 + m_{n-k+1}b_2 + \cdots + m_{n-1}b_k). \end{aligned}$$

继续沿用上述记号，我们有如下推论。

推论 4.5　设 $m(x) = \prod_{i\in[n]}(x-\alpha_i) = x^n + \sum_{j=0}^{n-1} m_j x^j$。如果 $m_{n-1} = m_{n-2} = \cdots = m_{n-k-1} = 0$ 成立，则有 $b_1 = b_2 = \cdots = b_{k+1} = 0$。

注 6　在定理 4.4中，可令 $m(x) = x^n - a^n$。由推论4.5我们知，$b_1 = b_2 = \cdots = b_{n-1} = 0$。则式 (4.3) 中的最后一个矩阵为 $\boldsymbol{J}$，即 $\boldsymbol{V}\boldsymbol{P}^{-1}\boldsymbol{V}^T = \boldsymbol{J}$。故容易验证码 $\boldsymbol{C}^{n,k}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 有一个校验矩阵 $H = [\boldsymbol{I}|-\boldsymbol{J}\boldsymbol{L}^T\boldsymbol{J}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}$，更具体地，

$$\boldsymbol{G}\cdot\boldsymbol{H}^T = [\boldsymbol{I}|\boldsymbol{L}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{V}^T\boldsymbol{J}\begin{bmatrix}-\boldsymbol{L}\\ \boldsymbol{I}\end{bmatrix}\boldsymbol{J} = 0.$$

因此，这给出了定理 4.4的另一种证明方法。事实上，$P_i(x) = \prod_{j\in[n],j\neq i}(x-\alpha_j) = \frac{m(x)}{x-\alpha_i} = x^{n-1} + \sum_{j=0}^{n-2}\alpha_i^{n-1-j}x^j$，$P_i(\alpha_i) = n\alpha_i^{n-1}$。故 $\boldsymbol{P}^{-1} = \mathrm{diag}(\boldsymbol{\alpha}/na^n)$。

在多项式 $m(x) = x^n + \sum_{j=0}^{n-1} m_j x^j$ 中，如果存在一个 $i\in[n-1]$，使得 $m_i \neq 0$，即，所有估值点集合不是一个陪集，此时，可能会存在某个 $b_j \neq 0$，则式 (4.3) 中的最后一个矩阵中，存在一些非零元不在其反对角线上。下面的定理告诉我们，只要非零的 m_i 不是很多，那么校验矩阵 H 的结构不受影响。

定理 4.6　给定一个 TRS 码 $\boldsymbol{C}^{n,k}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$，其生成矩阵为 $\boldsymbol{G} = [\boldsymbol{I}|\boldsymbol{L}]\cdot\boldsymbol{V}$，校验矩阵为 $\boldsymbol{H}$。令 $m(x) = \prod_{i\in[n]}(x-\alpha_i) = x^n + \sum_{j=0}^{n-1} m_j x^j$ 且 $\boldsymbol{P} = \mathrm{diag}(P(\boldsymbol{\alpha}))$，$\boldsymbol{v} = P(\boldsymbol{\alpha})^{-1} = (P_1(\alpha_1)^{-1},\cdots,P_n(\alpha_n)^{-1})$。则有

(1) 当 $\ell \geqslant 1$ 时，令 $\tau := \max\{k + t_\mu - 1 : \mu\in[\ell]\}$。如果 $m_{n-1} = m_{n-2} = \cdots = m_{n-\tau} = 0$，则 $H = [\boldsymbol{I}|-\boldsymbol{J}\boldsymbol{L}^T\boldsymbol{J}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}$。此时 $\boldsymbol{C}^{n,k}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 的对偶码是一个 TGRS 码 $\boldsymbol{C}^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};k-\boldsymbol{h};n-k-\boldsymbol{t};-\boldsymbol{\eta})$。

(2) 当 $\ell = 1$ 时，令 $h = h_1, t = t_1$ 以及 $\eta = \eta_1$。定义 $\theta := \max\{k+t-h-2, \{\arg\max_{i\neq h+1, i\in[k]}\{i-h\geqslant 2\}\}-1\}$。如果 $m_{n-1} = m_{n-2} = \cdots = m_{n-\theta} = 0$，则有

(i) 如果 $1+\eta b_{k+t-h-1} \neq 0$，则 $H = [\boldsymbol{I}|-\boldsymbol{J}\boldsymbol{A}^T\boldsymbol{J}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}$，其中，矩阵 $\boldsymbol{A}$

满足

$$\boldsymbol{A}_{i,j}=\begin{cases}\frac{\eta}{1+\eta b_{k+t-h-1}}, & (i,j)=(h+1,t),\\ 0, & \text{否则.}\end{cases}$$

此时 $\boldsymbol{C}^{n,k}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 的对偶码为 $\boldsymbol{C}^{n,k,\boldsymbol{\nu}}(\boldsymbol{\alpha};k-h;n-k-t;-\frac{\eta}{1+\eta b_{\theta+1}})$。

(ii) 如果 $1+\eta b_{k+t-h-1}=0$，则 $\boldsymbol{H}=[\boldsymbol{I}'|-\boldsymbol{J}\boldsymbol{B}^T\boldsymbol{J}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}$，其中矩阵 $\boldsymbol{B}$ 满足

$$\boldsymbol{B}_{i,j}=\begin{cases}\beta, & (i,j)=(h+1,t),\\ 0, & \text{否则,}\end{cases}$$

这里，β 是 $\mathbb{F}_q$ 中任意一个非零元。矩阵 $\boldsymbol{I}'$ 满足

$$\boldsymbol{I}'_{i,j}=\begin{cases}1, & i=j\neq n-k-t+1,\\ 0, & \text{否则.}\end{cases}$$

证明　注意到，定理中所有情况下，$\boldsymbol{H}$ 都满足 $\boldsymbol{H}=[\boldsymbol{I}|-\boldsymbol{J}\boldsymbol{M}^T\boldsymbol{J}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}$，其中矩阵 $\boldsymbol{M}$ 可能为 $\boldsymbol{L},\boldsymbol{A}$ 或 $\boldsymbol{B}$ 中的某一个。下面只需证明

$$\boldsymbol{G}\cdot\boldsymbol{H}^T=[\boldsymbol{I}|\boldsymbol{L}]\cdot\boldsymbol{V}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{V}^T\begin{bmatrix}\boldsymbol{I}\\ -\boldsymbol{J}\boldsymbol{M}\boldsymbol{J}\end{bmatrix}=0.$$

可以看出矩阵 $\boldsymbol{D}=\boldsymbol{V}\cdot\boldsymbol{P}^{-1}\cdot\boldsymbol{V}^T$ 以及 $\begin{bmatrix}\boldsymbol{I}\\ -\boldsymbol{J}\boldsymbol{M}\boldsymbol{J}\end{bmatrix}$ 的结构对于上述等式是否成立具有关键性作用。令 $\boldsymbol{E}=[\boldsymbol{I}|\boldsymbol{L}]$。对于一个矩阵 $\boldsymbol{Q}$，令 $\boldsymbol{R}_i^{\boldsymbol{Q}}$ 表示 $\boldsymbol{Q}$ 的第 i 行向量。则对于任意 $\mu\in[\ell]$，矩阵 $\boldsymbol{E}$ 的第 $(h_\mu+1)$ 行为

$$\boldsymbol{R}^{\boldsymbol{E}}_{(h_\mu+1)}=(\underbrace{0,\cdots,0}_{h_\mu},1,\ \underbrace{0,\cdots,0}_{k+t_\mu-h_\mu-2},\eta_\mu,0,\cdots,0).$$

而当 $i\neq h_\mu+1$ 时，矩阵 $\boldsymbol{E}$ 的第 i 行为

$$\boldsymbol{R}^{\boldsymbol{E}}_{i}=(\underbrace{0,\cdots,0}_{i-1},1,0,\cdots,0).$$

(1) 在这种情况下, 根据推论4.5可知 $b_1=b_2=\cdots=b_\tau=0$。由于对任意 $\mu\in[\ell]$，$k+t_\mu-1\leqslant\tau$，则有

$$\begin{aligned}\boldsymbol{R}^{\boldsymbol{D}}_{(k+t_\mu)}&=(\underbrace{0,\cdots,0}_{n-k-t_\mu},1,b_1,\cdots,b_{k+t_\mu-h_\mu-1},\cdots,b_{k+t_\mu-1})\\ &=(\underbrace{0,\cdots,0}_{n-k-t_\mu},1,0,\cdots,0).\end{aligned}$$

而对于矩阵 $\boldsymbol{ED}$ 的第 $(h_\mu+1)$ 行，

$$\begin{aligned}\boldsymbol{R}^{ED}_{(h_\mu+1)} &= (\underbrace{0,\cdots,0}_{n-k-t_\mu},\eta_\mu,0,\cdots,0,1,b_1+\eta_\mu b_{k+t_\mu-h_\mu},\cdots,b_{h_\mu}+\eta_\mu b_{k+t_\mu-1})\\ &= (\underbrace{0,\cdots,0}_{n-k-t_\mu},\eta_\mu,0,\cdots,0,1,\underbrace{0,\cdots,0}_{h_\mu}).\end{aligned}$$

当 $i\neq h_\mu+1$ 时，由于 $i\leqslant k\leqslant\tau$，我们有

$$\boldsymbol{R}^{ED}_i=(\underbrace{0,\cdots,0}_{n-i},1,b_1,\cdots,b_{i-1})=(\underbrace{0,\cdots,0}_{n-i},1,0,\cdots,0).$$

注意到

$$(-\boldsymbol{JLJ})_{i,j}=\begin{cases}-\eta_\mu, & (i,j)=(k-h_\mu,n-k-t_\mu+1),\\ 0, & \text{否则.}\end{cases}$$

因此

$$\boldsymbol{G}\cdot\boldsymbol{H}^T=\boldsymbol{E}\cdot\boldsymbol{D}\cdot\begin{bmatrix}\boldsymbol{I}\\ -\boldsymbol{JLJ}\end{bmatrix}=0.$$

(2) 当 $\ell=1$ 时，由 θ 的定义可知，矩阵 $\boldsymbol{ED}$ 的所有行，都满足如下形式

$$\boldsymbol{R}^{ED}_i=(\underbrace{0,\cdots,0}_{n-i},1,b_1,\cdots,b_{i-1})=(\underbrace{0,\cdots,0}_{n-i},1,0,\cdots,0),$$

除去第 $(h+1)$ 行外，此时

$$\begin{aligned}\boldsymbol{R}^{ED}_{(h+1)}=&(\underbrace{0,\cdots,0}_{n-k-t},\eta,\eta b_1,\cdots,\eta b_{k+t-h-2},1+\eta b_{k+t-h-1},b_1+\eta b_{k+t-h},\cdots,\\ &b_h+\eta b_{k+t-1})\\ =&(\underbrace{0,\cdots,0}_{n-k-t},\eta,0,\cdots,0,1+\eta b_{k+t-h-1},b_1+\eta b_{k+t-h},\cdots,b_h+\eta b_{k+t-1}).\end{aligned}$$

注意到，如果 $\theta>k+t-h-2$，则 $b_{k+t-h-1}=0$。根据 θ 的定义可知，对任意 $i\in[k]$ 且 $i\neq h+1$，如果 $i-h<2$，则 $n-i+1>n-h$。由于当 $\boldsymbol{M}=\boldsymbol{A}$ 或 $\boldsymbol{B}$ 时，矩阵 $\boldsymbol{M}$ 中只有位置 $(h+1,t)$ 上的元素非零，这意味着矩阵 $\begin{bmatrix}\boldsymbol{I}\\ -\boldsymbol{JMJ}\end{bmatrix}$ 的后 i 行上的元素均为 0。因此矩阵 $\boldsymbol{G}\cdot\boldsymbol{H}^T$ 的第 i 行为零向量。下面，只需验证当 $i=h+1$ 或 $i-h\geqslant 2$ 时，$\boldsymbol{G}\cdot\boldsymbol{H}^T$ 的第 i 行是否为 0 即可。对于后一种情况，由于 $n-i+1\geqslant n-k+1$ 且 $n-i+1\neq n-h$，显然为 0。

当 $1+\eta b_{k+t-h-1}\neq 0$ 时，注意到

$$(-\boldsymbol{JAJ})_{i,j}=\begin{cases}-\frac{\eta}{1+\eta b_{k+t-h-1}}, & (i,j)=(k-h,n-k-t+1),\\ 0, & \text{否则.}\end{cases}$$

考虑矩阵 $\begin{bmatrix} \boldsymbol{I} \\ -\boldsymbol{JAJ} \end{bmatrix}$ 的第 $(n-k-t+1)$ 列, 它只有两个非零元 $1, -\frac{\eta}{1+\eta b_{k+t-h-1}}$, 分别位于第 $(n-k-t+1)$ 和 $(n-h)$ 行。容易验证 $\boldsymbol{G}\cdot\boldsymbol{H}^T$ 的第 $(h+1)$ 行为 0。因此 (i) 完成。当 $1+\eta b_{k+t-h-1}=0$, 即 $\eta=-b_{k+t-h-1}^{-1}$, 此时

$$\boldsymbol{R}_{(h+1)}^{ED}=(\underbrace{0,\cdots,0}_{n-k-t},\eta,0,\cdots,0,0,b_1+\eta b_{k+t-h},\cdots,b_h+\eta b_{k+t-1}),$$

且矩阵 $\begin{bmatrix} \boldsymbol{I}' \\ -\boldsymbol{JBJ} \end{bmatrix}$ 的第 $(n-k-t+1)$ 列只有一个非零元 $-\beta$, 位于第 $(n-h)$ 行, 故 (ii) 完成。

∎

注 7　(a) 定理 4.6 证明的核心思想, 在于观察式 (4.3) 中最后一行矩阵非零元的分布情况。

(b) 在定理 4.6 中, 若 $\ell=1$ 且不存在 i 使得 $i-h\geqslant 2$, 此时 $h=k-1$ 且 $\theta=k+t-h-2$。故有 $b_1=b_2=\cdots=b_{k+t-h-2}=0$ 以及

$$\boldsymbol{R}_{(h+1)}^{ED}=(\underbrace{0,\cdots,0}_{n-k-t},\eta,0,\cdots,0,1+\eta b_{\theta+1},b_1+\eta b_{k+t-h},\cdots,b_h+\eta b_{k+t-1}).$$

特别地, 当 $t=1,\theta=0$ 时,

$$\boldsymbol{R}_{(h+1)}^{ED}=(\underbrace{0,\cdots,0}_{n-k-1},\eta,1+\eta b_1,b_1+\eta b_2,\cdots,b_h+\eta b_k),$$

且非零元 $\boldsymbol{A}_{k,1}=\frac{\eta}{1+\eta b_1}$, 这里 $b_1=-m_{n-1}=\sum_{i=1}^{n}\alpha_i$。由此, 可推出文献 [50] 中的定理 2.2。

(c) 在 (ii) 中, β 不能为 0, 否则的话, $\boldsymbol{H}$ 的秩将小于 $n-k$。

(d) 事实上, 当 $\ell>1$ 时, 也存在和 (2) 中类似的结果, 此时对 $m(x)$ 的系数约束会放松, 即可能不需要令 $n-\tau$ 个系数为 0, 但是这取决于具体的 $\boldsymbol{t},\boldsymbol{h}$ 中值的分布。特别地, 当 $\boldsymbol{t},\boldsymbol{h}$ 中的值单调增或单调减时, 相应码的对偶封闭性也更容易刻画。我们将在下一节中, 构造这种类型的码字。

4.4　自对偶的 TGRS 码

本节中, 我们首先给出一些 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 的构造, 再借助它去构造自对偶码。特别地, 当 $\ell=1$ 时, 构造出来的码是自对偶 MDS 码或 NMDS 码。

4.4.1　$\ell=1$ 的扭 Reed-Solomon 自对偶码

本小节中, 我们考虑 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};1;k-1;\eta)$, 即 $\ell=t=1$ 以及 $h=k-1$。对于自对偶码而言, 一个很关键的地方在于, 它的生成矩阵也可以作为其校验矩

阵。故，如果 $\boldsymbol{C}^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};1;h;\eta)$ 是一个自对偶码，那么它一定对偶封闭。而当 $\ell=1$ 时，对偶封闭的 TGRS 码，其校验矩阵在定理 4.6中有明确的刻画。事实上，根据定理 4.6，我们只需要在 $\mathbb{F}_q$ 中，找个 n 个互不相同的点 $\alpha_1,\cdots,\alpha_n$，使得它们是某个多项式 $m(x)$ 的所有根，即 $m(x)=(x-\alpha_1)\cdots(x-\alpha_n)$，且 $m(x)$ 的系数满足一些特定性质即可。注意到，若 $(t,h)=(1,k-1)$，$\theta=0$，此时我们对 $m(x)$ 的系数不做限制。

定理 4.7　给定一个正整数 n，使得其满足 $q-1=t(n+1)$，其中，q 是一个奇素数幂，t 是一个正整数。令 $\boldsymbol{H}$ 是一个阶为 $n+1$ 的 $\mathbb{F}_q$ 的乘法子群。则 $\mathbb{F}_q^*$ 有一个陪集分解 $\mathbb{F}_q^*=\boldsymbol{H}\sqcup\beta_1\boldsymbol{H}\sqcup\cdots\sqcup\beta_{t-1}\boldsymbol{H}$，且满足 $1+\beta_1+\cdots+\beta_{t-1}\neq 0$。取一个正整数 $s\leqslant t$，使得 sn 为偶数。令 $m(x)=\frac{x^{n+1}-1}{x-1}\cdot\frac{x^{n+1}-\beta_1^{n+1}}{x-\beta_1}\cdots\frac{x^{n+1}-\beta_{s-1}^{n+1}}{x-\beta_{s-1}}$，$\alpha_1,\cdots,\alpha_{sn}$ 是 $m(x)$ 所有的根。设 $\eta=-2b_1^{-1}$。且对任意 $i\in[sn]$，存在 $v_i\in\mathbb{F}_{q^2}$，使得 $\boldsymbol{C}^{sn,\frac{sn}{2},\boldsymbol{v}}(\boldsymbol{\alpha};1;\frac{sn}{2}-1;\eta)$ 是一个 $\mathbb{F}_{q^2}$ 上的自对偶 TGRS 码。

证明　如果 $1+\beta_1=0$，则可从陪集 $\beta_1\boldsymbol{H}$ 中，选取另外一个元素 β 做陪集首，使得 $1+\beta\neq 0$ 成立，再将 β_1 更新为 β，即令 $\beta_1=\beta$。重复这一操作，即可找到所有 $\beta_1,\cdots,\beta_{t-1}$，使得 $1+\beta_1+\cdots+\beta_{t-1}\neq 0$ 成立。故定理的第一部分证明完成。不妨设 $m(x)=x^{sn}+\sum_{j=0}^{sn-1}m_jx^j$。令 γ 是群 $\boldsymbol{H}$ 的一个生成元，则 $m_{sn-1}=-\sum_{i=0}^{s-1}\sum_{j=1}^{n}\beta_i\gamma^j=-b_1\neq 0$。

对任意 $i=0,1,\cdots,s-1$，令 $f_i(x)=\frac{x^{n+1}-\beta_i^{n+1}}{x-\beta_i}$。注意到

$$\begin{aligned}f_i(x)&=\frac{x^{n+1}-\beta_i^{n+1}}{x-\beta_i}\\&=\frac{\beta_i^{n+1}\boldsymbol{\Pi}_{j=0}^{n}(\beta_i^{-1}x-\gamma^j)}{x-\beta_i}\\&=\boldsymbol{\Pi}_{j=1}^{n}(x-\beta_i\gamma^j).\end{aligned}$$

因此 $m(x)=f_0(x)f_1(x)\cdots f_{s-1}(x)=\boldsymbol{\Pi}_{i=0}^{s-1}\boldsymbol{\Pi}_{j=1}^{n}(x-\beta_i\gamma^j)$，故所有的 α_i 互不相同，$f_i'(x)=\frac{(n+1)x^n-f_i(x)}{x-\beta_i}$ 且 $m'(x)=f_0'f_1\cdots f_{s-1}+\cdots+f_0f_1\cdots f_{s-2}f_{s-1}'$。如果 $f_i(\alpha_k)=0$，则对任意 $j\neq i$，有 $f_i'(\alpha_k)=\frac{(n+1)\alpha_k^n}{\alpha_k-\beta_i}\neq 0$ 和 $f_j(\alpha_k)\neq 0$，以及

$$\begin{aligned}m'(\alpha_k)&=f_0(\alpha_k)\cdots f_{i-1}(\alpha_k)f_i'(\alpha_k)f_{i+1}(\alpha_k)\cdots f_{s-1}(\alpha_k)\\&=f_i'(\alpha_k)\boldsymbol{\Pi}_{j\neq i}f_j(\alpha_k)\\&=(n+1)g_k\in\mathbb{F}_q^*,\end{aligned}$$

其中 $g_k^{-1}\in\mathbb{F}_q$ 是 $\mathbb{F}_{q^2}$ 的一个平方元。令 $v_k^2=g_k^{-1}$，则有 $v_k^2=g_k^{-1}=(n+1)m'(\alpha_k)^{-1}=(n+1)\boldsymbol{P}_k(\alpha_k)^{-1}$。因此，定理 4.2中的条件 1) 和 2) 满足。证明完成。∎

事实上，当取 $s=1$ 且 n 为偶数时，定理 4.7可退化为文献 [50] 中的定理 3.8。注意到，根据引理 4.1可知，如果对任意 $I\subset[sn]$ 且 $|I|=\frac{sn}{2}$，有 $-\eta^{-1}\neq\sum_{i\in I}\alpha_i$，则 $C^{sn,\frac{sn}{2},\boldsymbol{v}}(\boldsymbol{\alpha};1;\frac{sn}{2}-1;\eta)$ 是一个 MDS 码；反之，则为 NMDS 码。

4.4.2　$\ell>1$ 的扭 Reed-Solomon 自对偶码

本小节中，我们考虑 $\ell=3$ 时的 TGRS 码，更具体地，取 $\boldsymbol{t}=(1,2,3),\boldsymbol{h}=(k-3,k-2,k-1)$ 以及 $\boldsymbol{\eta}=(\eta_1,\eta_2,\eta_3)$，即码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$。此时，它的生成矩阵如下。

$$G=\begin{bmatrix}
v_1 & v_2 & \cdots & v_n\\
v_1\alpha_1 & v_2\alpha_2 & \cdots & v_n\alpha_n\\
\vdots & \vdots & \ddots & \vdots\\
v_1\alpha_1^{k-4} & v_2\alpha_2^{k-4} & \cdots & v_n\alpha_n^{k-4}\\
v_1(\alpha_1^{k-3}+\eta_1\alpha_1^{k}) & v_2(\alpha_2^{k-3}+\eta_1\alpha_2^{k}) & \cdots & v_n(\alpha_n^{k-3}+\eta_1\alpha_n^{k})\\
v_1(\alpha_1^{k-2}+\eta_2\alpha_1^{k+1}) & v_2(\alpha_2^{k-2}+\eta_2\alpha_2^{k+1}) & \cdots & v_n(\alpha_n^{k-2}+\eta_2\alpha_n^{k+1})\\
v_1(\alpha_1^{k-1}+\eta_3\alpha_1^{k+2}) & v_2(\alpha_2^{k-1}+\eta_3\alpha_2^{k+2}) & \cdots & v_n(\alpha_n^{k-1}+\eta_3\alpha_n^{k+2})
\end{bmatrix}.$$

注意到，码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 中的任意一个多项式，其度最多为 $k+2$，故它的最小距离在 $n-k-2$ 和 $n-k+1$ 之间。令 $n=2k$，故码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 是一个自对偶码当且仅当 $G\cdot G^{\top}=0$。因此，我们有以下引理。

引理 4.8　如果 $v_i^2=\lambda P_i(\alpha_i)^{-1}$，其中 $\lambda\in\mathbb{F}_q^*$，且有 $b_1=b_2=b_4=b_5=0,2\eta_2+\eta_2^2b_3=0$ 以及 $\eta_1+\eta_3+\eta_1\eta_3b_3=0$，则 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha};\boldsymbol{t};\boldsymbol{h};\boldsymbol{\eta})$ 是一个 $\mathbb{F}_q$ 上的自对偶码。

证明　记 $G=(\boldsymbol{g}_0;\cdots;\boldsymbol{g}_{k-3};\boldsymbol{g}_{k-2};\boldsymbol{g}_{k-1})$，其中 $\boldsymbol{g}_i$ 是矩阵 G 的第 $i+1$ 行向量，这里采用的记号源自于第2.3.1小节。当 $0\leqslant m\leqslant 2k-2$ 时，有 $v_1^2\alpha_1^m+\cdots+v_n^2\alpha_n^m=0$。这意味着 $G_{k-3}\cdot G_{k-3}^{\top}=0$，其中 $G_{k-3}=(\boldsymbol{g}_0;\cdots;\boldsymbol{g}_{k-4})$。而当 $0\leqslant j\leqslant k-4$ 时，我们有

$$\boldsymbol{g}_{k-3}\boldsymbol{g}_j^{\top}=\sum_{i=1}^{n}v_i^2(\alpha_i^s+\eta_1\alpha_i^t),\text{对任意}k-3\leqslant s\leqslant 2k-7\text{ 以及}k\leqslant t\leqslant 2k-4,$$

$$\boldsymbol{g}_{k-2}\boldsymbol{g}_j^{\top}=\sum_{i=1}^{n}v_i^2(\alpha_i^s+\eta_1\alpha_i^t),\text{对任意}k-2\leqslant s\leqslant 2k-6\text{ 以及}k+1\leqslant t\leqslant 2k-3,$$

$$\boldsymbol{g}_{k-1}\boldsymbol{g}_j^{\top}=\sum_{i=1}^{n}v_i^2(\alpha_i^s+\eta_1\alpha_i^t),\text{对任意}k-1\leqslant s\leqslant 2k-5\text{以及}k+2\leqslant t\leqslant 2k-2.$$

注意到 α_i 在上述每个等式中, 其幂次均不超过 $2k-2$, 因此当 $i \in [k-3, k-1]$ 时, 我们有 $\boldsymbol{g}_i \cdot G_{k-3}^{\top} = 0$。下面, 只需验证对任意 $i, j \in [k-3, k-1]$, 有 $\boldsymbol{g}_i \cdot \boldsymbol{g}_j^{\top} = 0$。容易得到

$$
\begin{aligned}
\boldsymbol{g}_{k-3} \cdot \boldsymbol{g}_{k-3}^{\top} &= \sum_{i=1}^{n} v_i^2(\alpha_i^{2k-6} + 2\eta_1\alpha_i^{2k-3} + \eta_1^2\alpha_i^{2k}) = \sum_{i=1}^{n} v_i^2\eta_1^2\alpha_i^{2k} = \lambda\eta_1^2 b_1 = 0,\\
\boldsymbol{g}_{k-2} \cdot \boldsymbol{g}_{k-2}^{\top} &= \sum_{i=1}^{n} v_i^2(\alpha_i^{2k-4} + 2\eta_2\alpha_i^{2k-1} + \eta_2^2\alpha_i^{2k+2}) = \sum_{i=1}^{n} v_i^2(2\eta_2\alpha_i^{2k-1} + \eta_2^2\alpha_i^{2k+2})\\
&= \lambda(2\eta_2 + \eta_2^2 b_3) = 0,\\
\boldsymbol{g}_{k-1} \cdot \boldsymbol{g}_{k-1}^{\top} &= \sum_{i=1}^{n} v_i^2(\alpha_i^{2k-2} + 2\eta_3\alpha_i^{2k+1} + \eta_3^2\alpha_i^{2k+4}) = \sum_{i=1}^{n} v_i^2(2\eta_3\alpha_i^{2k+1} + \eta_3^2\alpha_i^{2k+4})\\
&= \lambda(2\eta_3 b_2 + \eta_3^2 b_5) = 0,\\
\boldsymbol{g}_{k-3} \cdot \boldsymbol{g}_{k-2}^{\top} &= \sum_{i=1}^{n} v_i^2(\alpha_i^{2k-5} + \eta_1\alpha_i^{2k-2} + \eta_2\alpha_i^{2k-2} + \eta_1\eta_2\alpha_i^{2k+1}) = \sum_{i=1}^{n} v_i^2\eta_1\eta_2\alpha_i^{2k+1}\\
&= \lambda\eta_1\eta_2 b_2 = 0,\\
\boldsymbol{g}_{k-3} \cdot \boldsymbol{g}_{k-1}^{\top} &= \sum_{i=1}^{n} v_i^2(\alpha_i^{2k-4} + \eta_1\alpha_i^{2k-1} + \eta_3\alpha_i^{2k-1} + \eta_1\eta_3\alpha_i^{2k+2})\\
&= \sum_{i=1}^{n} v_i^2(\eta_1\alpha_i^{2k-1} + \eta_3\alpha_i^{2k-1} + \eta_1\eta_3\alpha_i^{2k+2})\\
&= \lambda(\eta_1 + \eta_3 + \eta_1\eta_3 b_3) = 0,\\
\boldsymbol{g}_{k-2} \cdot \boldsymbol{g}_{k-1}^{\top} &= \sum_{i=1}^{n} v_i^2(\alpha_i^{2k-3} + \eta_2\alpha_i^{2k} + \eta_3\alpha_i^{2k} + \eta_2\eta_3\alpha_i^{2k+3})\\
&= \sum_{i=1}^{n} v_i^2(\eta_2\alpha_i^{2k} + \eta_3\alpha_i^{2k} + \eta_2\eta_3\alpha_i^{2k+3})\\
&= \lambda(\eta_2 b_1 + \eta_3 b_1 + \eta_2\eta_3 b_4) = 0.
\end{aligned}
$$

因此, 我们有 $G \cdot G^{\top} = 0$, 即, $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 是一个 $\mathbb{F}_q$ 上的自对偶码。 ∎

定理4.7以及引理4.8的证明, 隐含了一个构造自对偶的 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 的关键点: 找到一个多项式 $m(x) = x^n + \sum_{j=0}^{n} m_j x^j$, 它的 n 个根分别为 $\alpha_1, \cdots, \alpha_n$, 使得所有的 $m'(\alpha_i)$ 要么都是平方元, 要么都不是平方元。此外, $b_1 = b_2 = b_4 = b_5 = 0$ 意味着 $m_{n-1} = m_{n-2} = m_{n-4} = m_{n-5} = 0$, 因此我们有以下定理。

定理 4.9　令 $p > 3$ 是一个奇素数, n 是一个正偶数。设 $m(x) = x^n + m_{n-3}x^{n-3} + m_0$, 其中 $m_0, m_{n-3} \in \mathbb{F}_{q'}^{*} \subseteq \mathbb{F}_q$, $\mathbb{F}_q$ 是 $m(x)$ 在 $\mathbb{F}_{q'}$ 上的分裂域, p 是其特征。取

$m(x)$ 的 n 个根 $\alpha_1, \cdots, \alpha_n$。如果 $p \mid n$，那么一定存在 $\lambda, \eta_1, \eta_2, \eta_3 \in \mathbb{F}_q \setminus \{0\}$，使得对任意 $i \in [n]$，存在 $v_i \in \mathbb{F}_q^*$ 并满足 $v_i^2 = \lambda P_i(\alpha_i)^{-1}$，且 $2\eta_2 + \eta_2^2 b_3 = 0$ 以及 $\eta_1 + \eta_3 + \eta_1\eta_3 b_3 = 0$，则 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 是一个 $\mathbb{F}_q$ 上的自对偶码。

证明　根据定义可知，$b_3 = -m_{n-3}$，因此，满足条件的 η_1, η_2, η_3 存在。由于 $m'(x) = nx^{n-1} + m_{n-3}(n-3)x^{n-4} = -3m_{n-3}x^{n-4}$，$(m(x), m'(x)) = 1$，可知 $\alpha_1, \cdots, \alpha_n$ 互不相同。此外，$P_i(\alpha_i)^{-1} = m'(\alpha_i)^{-1} = (-3m_{n-3}\alpha_i^{n-4})^{-1}$。定义 $\lambda = -3m_{n-3}$ 以及 $v_i = \alpha_i^{\frac{4-n}{2}}$。由于 $p > 3$，可知 $\lambda \neq 0$。根据引理 4.8可推出，$C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 是一个 $\mathbb{F}_q$ 上的自对偶码。 ∎

在上述定理中，n 是偶数且 $p \mid n$，即 $2p \mid n$。那么码长 n 和 [50]定理 3.1 一样，但最小距离 d 可能不一样，因为 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 的最小距离满足 $n-k-2 \leqslant d \leqslant n-k+1$，即 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 可能既不是 MDS 也不是 NMDS 码。尽管利用 $\ell > 1$ 时的 TGRS 码构造自对偶码，目前看来没有丰富自对偶 MDS 或 NMDS 码的种类，但我们相信，这种码字在后续研究中，会在这方面有所突破。

4.5　本章总结

在本章中，我们考虑了 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$，特别是当其所有估值点构成某个多项式根集合的时候，刻画出了其校验矩阵的结构。即，TGRS 码在给定条件下，是对偶封闭的。利用这一结果，我们构造了相应的自对偶码。特别是当 $\ell = 1$ 时，所得自对偶码是 MDS 或 NMDS 码。而当 $\ell = 3$ 时，所得自对偶码的最小距离在 $n-k-2$ 和 $n-k+1$ 之间。在后续工作中，我们将尝试用 $\ell > 1$ 时对偶封闭的 TGRS 码 $C^{n,k,\boldsymbol{v}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 构造自对偶 MDS 或 NMDS 码。

第 5 章　总结与展望

本章，我们将对全文进行总结，指出一些研究工作的不足之处，并对进一步研究方向做出简要阐述。

5.1　本文的主要成果

本文主要从数据存储和传输过程入手，将编码技术应用其中，以保障数据的安全性。更具体地，我们从编码的角度出发，分别对经典存储系统下计算负载均衡问题、DNA 存储系统中串联复制纠错码以及可应用于 McEliece 密码系统的（广义）扭 RS 码三方面进行了研究。本文的主要研究工作和成果陈述如下。

1. 针对数据计算负载均衡以及更新问题，我们研究了小域上稀疏平衡的 $[n,k]_q$ MDS 码。已有的结果要求 $q \geqslant n + \left\lceil \frac{k(k-1)}{n} \right\rceil$，而我们在 $n \leqslant 2k$ 时，将其改进到了 $q \geqslant n-1$。更具体地，我们将构造 MDS 码这一代数问题转化为一个组合问题。通过构造满足条件的一致平衡集族，在 $q \geqslant n$ 且 $n \leqslant 2k$ 时，给出了稀疏平衡的 $[n,k]_q$ MDS 码。进一步，通过扩展坐标，有选择的更新已有的平衡集族，我们将对有限域的要求改进到了 $q \geqslant n-1$。当码长 $n > 2k$ 时，我们将构造 MDS 码转化为构造平衡和集 $A+B$，其中 $|A| = k$ 以及 $|B| = k-1$，得到了稀疏平衡的 $[n = q = p^s, k = p^e m]_q$ MDS 码，这里 $e \leqslant s-2$ 且 $m \leqslant p-1$，或者 $e = s-1$ 且 $m < \frac{p}{2}$。
2. 针对 DNA 存储系统中，数据在存储时由于 DNA 分子复制而引起的串联复制错误，导致数据丢失或破损问题，为了恢复原始信息，我们主要研究了能够纠正串联复制错误的纠错码。事实上，这类纠错码可以由 ℓ_1 度量下非负整数上的最优常重码或 q 元常重码得到。基于此，我们考虑了 ℓ_1 度量下非负整数以及 $I_3 = \{0,1,2\}$ 上的常重码，并利用码字支集刻画出了一个必要条件（称为 UNC 条件），它表明了 ℓ_1 度量下的码与填充集族之间的关系。根据 UNC 条件，我们将构造码字问题转化为找到一个合适的填充集族问题。由于受到重量 w 的限制，我们分别针对 $\mathbb{Z}_{\geqslant 0}$ 和 I_3 这两种字母集，确定了重量 $w \leqslant 4$ 的所有常重码最优码字大小。而对于一般的 w，我们还提供了权重为 w 和距离为 $2w-2$ 的三元码最优码字大小的渐近结果。
3. 针对（广义）扭 RS 码问题，我们研究了 TGRS 码 $C^{n,k,\boldsymbol{\nu}}(\boldsymbol{\alpha}; \boldsymbol{t}; \boldsymbol{h}; \boldsymbol{\eta})$ 的性质和构造。特别是考虑扭 RS 码对偶封闭性问题，已有的研究要求其所有估值点构成一个子群。我们的结论推广了这一点，当所有估值点构成某个多项式根集合的时候，我们分别针对 $\ell = 1$ 和 $\ell > 1$，刻画出了其校验矩阵的结

构。根据其校验矩阵可以看出，它是对偶封闭的。利用这一结果，我们构造了相应的自对偶码。特别是当 $\ell = 1$ 时，所得自对偶码是 MDS 或 NMDS 码。而当 $\ell = 3$ 时，所得自对偶码的最小距离在 $n-k-2$ 和 $n-k+1$ 之间。

5.2　本文不足与研究展望

本文致力于研究如何防御网络空间中的信息在存储和传输这两个环节中所面临的威胁，主要考虑的方式是引入特定性质的纠错码或纠删码。尽管我们做了大量研究，但仍存在不足之处，以及纠错码应用于其他场景下的工作及其相关结果，值得进一步研究和优化。

5.2.1　稀疏平衡码

本文第二章中，主要考虑了当有限域满足 $q \geqslant n-1$ 且码长 $n \leqslant 2k$ 时的稀疏平衡的 $[n,k]_q$ MDS 码。而当 $n > 2k$ 时，我们只得到了满足 $q = n$ 的部分码结果。因此，考虑当 $n > 2k$ 时进一步缩小 q 的大小，从而构造稀疏平衡 MDS 码是非常有意义的。值得一提的是，我们在定义稀疏平衡码的时候，是针对生成矩阵而言的，而校验矩阵的稀疏平衡性也值得进一步考虑。事实上，利用文献 [66] 中的方法，可以简单的推出当 $q = \Omega(n)$ 时，任意给定两个满足 MDS 条件的二元矩阵，存在一个 MDS 码，使得它的生成矩阵和校验矩阵的支撑矩阵分别为这两个二元矩阵。这一结果，与 MDS 猜想类似。我们由此提出一个问题：能否在小域上构造一个，具有稀疏平衡的生成矩阵和校验矩阵的 MDS 码。

另一方面，随着存储系统中所采用的纠删码的多样性，研究不同类型的稀疏平衡码也是非常有意义的。例如，在文献 [58] 中，作者研究了稀疏平衡的 Tamo-Barg 码。因此我们考虑能否将第二章中的方法推广到其它类型的码上，得到相应的稀疏平衡码。事实上，在考虑稀疏平衡的 NMDS 码时，可以利用第2.5节中的和集方法，得到部分结果。

5.2.2　ℓ_1 度量下的常重码

在考虑纠正串联复制错误的纠错码时，文献 [8] 给出：ℓ_1 度量下非负整数上的常重码可以用来构造串联复制错误纠错码。我们在考虑这一类型的常重码时，主要考虑了 $w \leqslant 4$ 的情况，而针对一般的 w，构造这种常重码相对来说比较困难，值得进一步研究。另外，当 q 充分大时，基于 I_q 上的常重码可以用来作为参考。特别地，在本文中，针对这一情况，且在 w 为任意固定正整数时，我们给出了部分距离为 $2w-2$ 的最优常重码，但在文献 [80] 中，我们完整地给出了这一参数最优码的构造。主要用到的工具是组合设计中的填充集族以及图填充相关结果。

而当距离 $4 \leqslant d \leqslant 2w-4$ 时，构造这种最优码仍需进一步研究。

事实上，考虑给定最小距离下，码字重量是码长的一个线性函数时，这样的常重码，在构造串联复制错误纠错码时也能给予很大的帮助。在文献 [47] 中，作者研究了这种类型的常重码。但对于最优码字的构造问题，仍没有完全解决，值得进一步探讨。

5.2.3　树上的编码

在图论学科中，一个具有 n 个标号顶点（节点）的标号树（labeled tree）指的是由 $n-1$ 条边构成的连通图。树是一种在计算机科学以及相关领域有广泛应用的一种图。例如，在信号处理（signal processing）中，图被用来表示波形；在编程语言中，树通常被用作结构来描述语言中的限制，等等。利用类似于编码理论里面的方法，若给定两棵树之间的距离，则可以研究在给定距离下的编码 [83]。这种码字可以用来纠正边擦除（edge erasures）错误，且有着广泛的应用场景。更具体地，在数据结构中，树是一种广泛使用的抽象模拟分层树结构的数据类型 [84]。这样的树数据结构，将信息存储在节点中，并使用节点之间的边作为它们之间的指针。当然也有一些数据结构是将信息存储在边上而不是节点上，例如后缀树（suffix trees）。通过添加冗余的边和节点，树编码可以纠正一些不匹配的指针错误；又或者当某一条边（存储在其上的数据）擦除时，可以通过树编码进行恢复，从而得到存储在该边上的信息。

由 Cayley 公式 [85] 知，n 个标号顶点上的所有不同标号树的数目是 n^{n-2}。将这个集合记为全空间 $\mathbf{T}$，在这个空间上定义两棵树 T_1, T_2 之间的距离为

$$d_{\mathcal{T}}(T_1, T_2) = n - 1 - |E_1 \cap E_2|,$$

其中 E_1, E_2 分别为 T_1, T_2 的边集。

定义 5.1　树上的一个码 $\mathcal{C}_{\mathcal{T}}$，记为 $\mathcal{T}$-(n, M, d)，指的是顶点集为 $[n]$ 的 M 棵最小距离为 $d = d_{\mathcal{T}}(\mathcal{C}_{\mathcal{T}}) = \min_{T_1 \neq T_2, T_1, T_2 \in \mathcal{C}_{\mathcal{T}}} \{d_{\mathcal{T}}(T_1, T_2)\}$ 的树的集合。$\mathcal{C}_{\mathcal{T}}$ 中的每棵树称为一个码树。

我们记 $A(n, d)$ 为一个 $\mathcal{T}$-(n, M, d) 码所能达到的最大码字个数。给定一棵 n 个顶点的标号树，可以为其自然的定义一个长度为 $\binom{n}{2}$ 的特征向量，该向量的坐标集由所有可能的 $\binom{n}{2}$ 条边构成，且在某个位置上为 1 当且仅当对应的边在这棵树中，其余位置为 0。由此可知，任意两棵树之间的距离可以看作是它们特征向量之间汉明距离的一半。因此，一个 $\mathcal{T}$-(n, M, d) 码是一个 $(\binom{n}{2}, 2d, n-1)_2$ 二元常重码，这里的度量是汉明距离。因此，我们有一个自然的上界。

定理 5.1　对任意 $n \geqslant 1$ 以及 $1 \leqslant d \leqslant n$，我们有 $A(n, d) = \mathcal{O}(n^{n-d})$。

事实上，在证明定理 5.1 时，我们并没有使用全空间 $\mathbf{T}$，而是取了所有重量

为 $n-1$、长度为 $\binom{n}{2}$ 的二元向量，集合大小为 $\binom{\binom{n}{2}}{n-1} > n^{n-2}$。注意到，一棵树可以唯一对应一个特征向量，反之则不然。因此，上述的上界是比较粗糙的。而如何用向量来描述一棵树的结构，从而减少空间的大小，使得可以借助处理二元常重码的一些方法来处理树上的码，值得进一步研究。

在文献 [83] 中,作者利用了球填充方法,改进了定理 5.1中的上界,即 $A(n,d) = \mathcal{O}(n^{n-1-d})$。而对于下界,他们给出了一个具体的码的构造,得到 $A(n,d) = \Omega(n^{n-2d})$。因此，如何给 $A(n,d)$ 一个恰当估计，需要进一步的研究。当 $d = n-1, n-2$ 时，$A(n,d)$ 的确切值已被确定。而当 $d \leqslant n-3$ 时，均未解决。特别地，当 $n \geqslant 9$ 且 $d = n-3$ 时，$A(n, n-3) \leqslant n^2$，达到这一上界的最优码，目前仍在研究当中。

参 考 文 献

[1] 中国互联网络信息中心（CNNIC）. 第 49 次《中国互联网络发展状况统计报告》[EB/OL]. 2022. https://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/202202/t20220225_71727.htm.

[2] HECKEL R, SHOMORONY I, RAMCHANDRAN K, et al. Fundamental limits of DNA storage systems[C]//Proceedings of the 2017 IEEE International Symposium on Information Theory. IEEE, 2017: 3130-3134.

[3] EXTANCE A. How DNA could store all the world's data[J]. Nature, 2016, 537(7618): 1-1.

[4] YAZDI S H T, KIAH H M, GARCIA-RUIZ E, et al. DNA-based storage: Trends and methods [J]. IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 2015, 1 (3): 230-248.

[5] LANDER E, LINTON L, BIRREN B, et al. Initial sequencing and analysis of the human genome[J]. Nature, 2001, 409(6822): 860-921.

[6] USDIN K. The biological effects of simple tandem repeats: lessons from the repeat expansion diseases[J]. Genome research, 2008, 18(7): 1011-1019.

[7] FONDON J W, GARNER H R. Molecular origins of rapid and continuous morphological evolution[J]. Proceedings of the National Academy of Sciences, 2004, 101(52): 18058-18063.

[8] JAIN S, HASSANZADEH F F, SCHWARTZ M, et al. Duplication-correcting codes for data storage in the DNA of living organisms[J]. IEEE Transactions on Information Theory, 2017, 63(8): 4996-5010.

[9] BEELEN P, PUCHINGER S, NÉ NIELSEN J R. Twisted Reed-Solomon codes[C]// Proceedings of the 2017 IEEE International Symposium on Information Theory. IEEE, 2017: 336-340.

[10] PUCHINGER S, ROSENKILDE J, SHEEKEY J. Further Generalisations of Twisted Gabidulin Codes[C]//Proceedings of International Workshop on Coding and Cryptography. 2017.

[11] BEELEN P, BOSSERT M, PUCHINGER S, et al. Structural properties of twisted Reed-Solomon codes with applications to cryptography[C]//Proceedings of the 2018 IEEE International Symposium on Information Theory. IEEE, 2018: 946-950.

[12] NIEDERREITER H. Knapsack-type cryptosystems and algebraic coding theory[J]. Problems of Control and Information Theory, 1986, 15(2): 157-166.

[13] SIDELNIKOV V M, SHESTAKOV S O. On insecurity of cryptosystems based on generalized Reed-Solomon codes[M]. Walter de Gruyter, Berlin/New York Berlin, New York, 1992.

[14] WIESCHEBRINK C. Two NP-complete Problems in Coding Theory with an Application

in Code Based Cryptography[C]//Proceedings of the 2006 IEEE International Symposium on Information Theory. 2006: 1733-1737.

[15] BALDI M, BIANCHI M, CHIARALUCE F, et al. Enhanced public key security for the McEliece cryptosystem[J]. Journal of Cryptology, 2016, 29(1): 1-27.

[16] SIDELNIKOV V M, SHESTAKOV S O. On insecurity of cryptosystems based on generalized Reed-Solomon codes[J]. Discrete Mathematics and Applications, 1992, 2(4): 439-444.

[17] COUVREUR A, GABORIT P, GAUTHIER-UMAÑA V, et al. Distinguisher-based attacks on public-key cryptosystems using Reed–Solomon codes[J]. Designs, Codes and Cryptography, 2014, 73(2): 641-666.

[18] MCELIECE R J. A public-key cryptosystem based on algebraic[J]. Coding Thv, 1978, 4244: 114-116.

[19] YAN M, SPRINTSON A. Algorithms for weakly secure data exchange[C]//Proceedings of the 2013 International Symposium on Network Coding. IEEE, 2013: 1-6.

[20] YAN M, SPRINTSON A, ZELENKO I. Weakly secure data exchange with generalized Reed Solomon codes[C]//Proceedings of the 2014 IEEE International Symposium on Information Theory. IEEE, 2014: 1366-1370.

[21] LI S, GASTPAR M. Cooperative data exchange based on MDS codes[C]//Proceedings of the 2017 IEEE International Symposium on Information Theory. IEEE, 2017: 1411-1415.

[22] HALBAWI W, HO T, YAO H, et al. Distributed Reed-Solomon codes for simple multiple access networks[C]//Proceedings of the 2014 IEEE International Symposium on Information Theory. IEEE, 2014: 651-655.

[23] DAU S H, SONG W, YUEN C. On Simple Multiple Access Networks[J]. IEEE Journal on Selected Areas in Communications, 2015, 2(33): 236-249.

[24] DAU S H, SONG W, DONG Z, et al. Balanced Sparsest generator matrices for MDS codes [C]//Proceedings of the 2013 IEEE International Symposium on Information Theory. IEEE, 2013: 1889-1893.

[25] HALBAWI W, LIU Z, HASSIBI B. Balanced Reed-Solomon codes for all parameters[C]// Proceedings of the 2016 IEEE Information Theory Workshop. 2016: 409-413.

[26] HALBAWI W, LIU Z, HASSIBI B. Balanced reed-solomon codes[C]//Proceedings of the 2016 IEEE International Symposium on Information Theory. IEEE, 2016: 935-939.

[27] SONG W, CAI K. Generalized Reed-Solomon codes with sparsest and balanced generator matrices[C]//Proceedings of the 2018 IEEE International Symposium on Information Theory. IEEE, 2018: 1-5.

[28] COSTELLO D J, FORNEY G D. Channel coding: The road to channel capacity[J]. Proceedings of the IEEE, 2007, 95(6): 1150-1177.

[29] KING O D. Bounds for DNA codes with constant GC-content[J]. The Electronic journal of combinatorics, 2003, 10(1): 33.

[30] MILENKOVIC O, KASHYAP N. On the design of codes for DNA computing[C]//Proceedings of the International Workshop on Coding and Cryptography. Springer, 2005: 100-119.

[31] GRAHAM R, SLOANE N. Lower bounds for constant weight codes[J]. IEEE Transactions on Information Theory, 1980, 26(1): 37-43.

[32] BROUWER A, SHEARER J, SLOANE N, et al. A new table of constant weight codes[J]. IEEE Transactions on Information Theory, 1990, 36(6): 1334-1380.

[33] AGRELL E, VARDY A, ZEGER K. Upper bounds for constant-weight codes[J]. IEEE Transactions on Information Theory, 2000, 46(7): 2373-2395.

[34] Zhang H, Ge G. Optimal ternary constant-weight codes of weight four and distance six[J]. IEEE Transactions on Information Theory, 2010, 56(5): 2188-2203.

[35] ZHANG H, ZHANG X, GE G. Optimal ternary constant-weight codes with weight 4 and distance 5[J]. IEEE Transactions on Information Theory, 2012, 58(5): 2706-2718.

[36] CHEE Y M, ZHANG H, ZHANG X. Complexity of dependences in bounded domains, armstrong Codes, and generalizations[J]. IEEE Transactions on Information Theory, 2014, 61(2): 812-819.

[37] CHEE Y M, GE G, ZHANG H, et al. Hanani triple packings and optimal q-ary codes of constant weight three[J]. Designs, Codes and Cryptography, 2015, 75(3): 387-403.

[38] CHEE Y M, ZHANG X. Linear size constant-composition codes meeting the Johnson bound [J]. IEEE Transactions on Information Theory, 2017, 64(2): 909-917.

[39] CHEE Y M, KIAH H M, ZHANG H, et al. Constructions of optimal and near-optimal multiply constant-weight codes[J]. IEEE Transactions on Information Theory, 2017, 63(6): 3621-3629.

[40] CHEE Y M, GAO F, KIAH H M, et al. Decompositions of edge-colored digraphs: A new technique in the construction of constant-weight codes and related families[J]. SIAM Journal on Discrete Mathematics, 2019, 33(1): 209-229.

[41] BARG A, MAZUMDAR A. Codes in permutations and error correction for rank modulation [J]. IEEE Transactions on Information Theory, 2010, 56(7): 3158-3165.

[42] JIANG A, LI H, WANG Y. Error scrubbing codes for flash memories[C]//Proceedings of the 11th Canadian Workshop on Information Theory. IEEE, 2009: 32-35.

[43] TALLINI L G, BOSE B. On L_1-distance error control codes[C]//Proceedings of the 2011 IEEE International Symposium on Information Theory. IEEE, 2011: 1061-1065.

[44] ZHOU H, SCHWARTZ M, JIANG A A, et al. Systematic error-correcting codes for rank modulation[J]. IEEE Transactions on Information Theory, 2014, 61(1): 17-32.

[45] FARNOUD F, SKACHEK V, MILENKOVIC O. Error-correction in flash memories via codes

in the Ulam metric[J]. IEEE Transactions on Information Theory, 2013, 59(5): 3003-3020.

[46] KABATIANSKY G, KRUGLIK S. On codes correcting constant number of errors in l_1 metric[C]//Proc. Collection Works 39-Te Dis-Tsiplinarno Cleavage Conf. IPI RAS eInformation Technol. Syst. 2015: 152-157.

[47] Kovačević M, Tan V Y F. Codes in the Space of multisets-coding for permutation channels with impairments[J]. IEEE Transactions on Information Theory, 2018, 64(7): 5156-5169.

[48] JINUSHI H, SAKANIWA K. A construction method for multilevel error-correcting codes based on absolute summation weight[C]//Proceedings of the 1990 IEEE International Symposium on Information Theory. IEEE, 1990: 87.

[49] BUTSON A. Generalized Hadamard matrices[J]. Proceedings of the American Mathematical Society, 1962, 13(6): 894-898.

[50] HUANG D, YUE Q, NIU Y, et al. MDS or NMDS self-dual codes from twisted generalized Reed–Solomon codes[J]. Designs, Codes and Cryptography, 2021, 89(9): 2195-2209.

[51] LIU H, LIU S. New Constructions of MDS twisted Reed-Solomon codes and LCD MDS codes. arXiv preprint arXiv:2008.03708[A]. 2020.

[52] HUANG D, YUE Q, NIU Y. MDS or NMDS LCD codes from twisted Reed–Solomon codes [Z]. 2020.

[53] WU Y, HYUN J Y, LEE Y. New LCD MDS codes of non-Reed-Solomon type[J]. IEEE Transactions on Information Theory, 2021, 67(8): 5069-5078.

[54] GREAVES G R W, SYATRIADI J. Reed-Solomon codes over small fields with constrained generator matrices[J]. IEEE Transactions on Information Theory, 2019, 65(8): 4764-4770.

[55] STINSON D R, WEI R, YIN J. Packings[M]//Handbook of Combinatorial Designs. Chapman and Hall/CRC, 2006: 576-582.

[56] COLBOURN C J, DINITZ J H. Handbook of combinatorial designs[M]. CRC press, 2006.

[57] ALON N, CARO Y, YUSTER R. Packing and covering dense graphs[J]. Journal of Combinatorial Designs, 1998, 6(6): 451-472.

[58] HALBAWI W, LIU Z, DUURSMA I M, et al. Sparse and balanced Reed–Solomon and Tamo–Barg codes[J]. IEEE Transactions on Information Theory, 2018, 65(1): 118-130.

[59] DAU S H, SONG W, YUEN C. On the existence of MDS codes over small fields with constrained generator matrices[C]//Proceedings of the 2014 IEEE International Symposium on Information Theory. IEEE, 2014: 1787-1791.

[60] YILDIZ H, HASSIBI B. Optimum linear codes with support-constrained generator matrices over small fields[J]. IEEE Transactions on Information Theory, 2019, 65(12): 7868-7875.

[61] Yildiz H, Hassibi B. Gabidulin Codes With Support Constrained Generator Matrices[J]. IEEE Transactions on Information Theory, 2020, 66(6): 3638-3649.

[62] Yildiz H, Raviv N, Hassibi B. Support Constrained Generator Matrices of Gabidulin Codes in Characteristic Zero[C]//Proceedings of the 2020 IEEE International Symposium on Information Theory. IEEE, 2020: 60-65.

[63] EFFROS M, KSCHISCHANG F, LANGBERG M. Between Shannon and Hamming: Network Information Theory and Combinatorics[Z]. 2015.

[64] HEIDARZADEH A, SPRINTSON A. An algebraic-combinatorial proof technique for the GM-MDS conjecture[C]//Proceedings of the 2017 IEEE International Symposium on Information Theory. IEEE, 2017: 11-15.

[65] YILDIZ H, HASSII B. Further progress on the GM-MDS conjecture for reed-solomon codes [C]//Proceedings of the 2018 IEEE International Symposium on Information Theory. IEEE, 2018: 16-20.

[66] LOVETT S. MDS matrices over small fields: A proof of the GM-MDS conjecture[C]// Proceedings of the 2018 IEEE Annual Symposium on Foundations of Computer Science. IEEE, 2018: 194-199.

[67] CHEN T, ZHANG X. Code of sparse and balanced MDS codes over small fields[EB/OL]. 2021. https://github.com/ttchenday/Sparse_and_Balanced_MDS_Codes_over_Small_Fields.

[68] KRATTENTHALER C. Advanced determinant calculus: A complement[J]. Linear Algebra and its Applications, 2005, 411: 68-166.

[69] DICKSON L E. History of the Theory of Numbers: Diophantine Analysis/by Leonard Eugene Dickson[M]. Chelsea., 1952.

[70] KOVAČEVIĆ M, TAN V Y F. Asymptotically optimal codes correcting fixed-length duplication errors in DNA storage systems[J]. IEEE Communications Letters, 2018, 22(11): 2194-2197.

[71] LENZ A, JÜNGER N, WACHTER-ZEH A. Bounds and Constructions for Multi-Symbol Duplication Error Correcting Codes[C]//Proceedings of the Sixteenth International Workshop on Algebraic and Combinatorial Coding Theory. 2018.

[72] Tang Y, Yehezkeally Y, Schwartz M, et al. Single-Error Detection and Correction for Duplication and Substitution Channels[J]. IEEE Transactions on Information Theory, 2020, 66(11): 6908-6919.

[73] YEHEZKEALLY Y, SCHWARTZ M. Reconstruction codes for DNA sequences with uniform tandem-duplication errors[J]. IEEE Transactions on Information Theory, 2019, 66(5): 2658-2668.

[74] GE G. Group divisible designs[M]//Handbook of Combinatorial Designs. Chapman and Hall/CRC Press, 2007: 255-260.

[75] BAO J, JI L. The completion determination of optimal (3,4)-packings[J]. Designs, Codes and

Cryptography, 2015, 77(1): 217-229.

[76] WEI H, GE G. Group divisible designs with block sizes from $K_{1(3)}$ and Kirkman frames of type $h^u m^1$[J]. Discrete Mathematics, 2014, 329: 42-68.

[77] COLBOURN C J, ROSA A. Quadratic leaves of maximal partial triple systems[J]. Graphs and Combinatorics, 1986, 2(1): 317-337.

[78] COLBOURN C J, LING A C. A class of partial triple systems with applications in survey sampling[J]. Communications in Statistics-Theory and Methods, 1998, 27(4): 1009-1018.

[79] SHEARER J B. Difference Triangle Sets[M]//Handbook of Combinatorial Designs. Chapman and Hall/CRC, 2006: 436-440.

[80] WEI X, CHEN T, ZHANG X. Optimal ternary codes with weight w and distance $2w-2$ in ℓ_1-metric[J]. IEEE Transactions on Information Theory, 2021, 67(11): 7221-7231.

[81] LAVAUZELLE J, RENNER J. Cryptanalysis of a system based on twisted Reed–Solomon codes[J]. Designs, Codes and Cryptography, 2020, 88(7): 1285-1300.

[82] ALTHAUS H, LEAKE R. Inverse of a finite-field Vandermonde matrix (Corresp.)[J]. IEEE Transactions on Information Theory, 1969, 15(1): 173-173.

[83] YOHANANOV L, YAAKOBI E. Codes Over Trees[J]. IEEE Transactions on Information Theory, 2021, 67(6): 3599-3622.

[84] CORMEN T H, LEISERSON C, RIVEST R L, et al. Introduction to Algorithms Cambridge, Massachusetts[M]. The MIT Press, 1990.

[85] CASAROTTO C. Graph theory and Cayley’ s formula[J]. University of Chicago, 2006.

附录 A　表3.1–3.2中的码

本章中，我们将通过计算机搜索，给出表3.1和3.2提到的递归构造中所需要的或未被前面构造方法覆盖的最优码。对于本章中的所有表格，短轨道中的码字均以粗体显示。为了方便验证这里列出的码字，我们在 GitHub 网站上开源了代码，以便有兴趣的读者可以检查正确性，代码链接见: https://github.com/llxu2020/verification.

A.1　小码码长满足 $n \equiv 0 (\mathrm{mod}\ 3)$ 的码字

对每个码长 $n \in \{15, 21, 27, 30, 33, 39, 45, 48, 51, 57, 63, 87, 93, 99, 111, 123\}$，其相应最优码的基码字在表A.1列出，且不同码对应不同的群作用。当 $n = 21$、33 和 45 时，用来作用的自同构为 $(0\ 4\ 8 \cdots n-5)(1\ 5\ 9 \cdots n-4)(2\ 6\ 10 \cdots n-3)(3\ 7\ 11 \cdots n-2)(n-1)$；当 $n = 30$ 时，用来作用的自同构为 $(0\ 4\ 8 \cdots 20)(1\ 5\ 9 \cdots 21)(2\ 6\ 10 \cdots 22)(3\ 7\ 11 \cdots 23)(24\ 25\ 26 \cdots 29)$；而当 $n = 57$ 和 93 时，用来作用的自同构为 $(0\ 2\ 4 \cdots n-3)(1\ 3\ 5 \cdots n-2)(n-1)$；对于剩下的，除了 $n = 48$ 时，最优码可通过在基码字支集集合上加 6 模 48 生成外，其余的均为加 3 $(\mathrm{mod}\ n)$ 生成。

A.2　小码码长满足 $n \equiv 1 (\mathrm{mod}\ 3)$ 的码字

对每个码长 $n \in \{16, 19, 22, 25, 31, 34, 40\}$，其相应最优码的基码字在表A.2列出，且不同码对应不同的群作用。当 $n = 16$ 和 40 时，用来作用的自同构为 $(0\ 4\ 8 \cdots n-4)(1\ 5\ 9 \cdots n-3)(2\ 6\ 10 \cdots n-2)(3\ 7\ 11 \cdots n-1)$；当 $n = 19$ 和 31 时，用来作用的自同构为 $(0\ 1\ 2 \cdots n-1)$；当 $n = 22$ 和 34 时，用来作用的自同构为 $(0\ 3\ 6 \cdots n-4)(1\ 4\ 7 \cdots n-3)(2\ 5\ 8 \cdots n-2)(n-1)$；以及当 $n = 25$ 时，用来作用的自同构为 $(0\ 6\ 12\ 18)(1\ 7\ 13\ 19)(2\ 8\ 14\ 20)(3\ 9\ 15\ 21)(4\ 10\ 16\ 22)(5\ 11\ 17\ 23)(24)$。

A.3　小码码长满足 $n \equiv 2 (\mathrm{mod}\ 3)$ 的码字

对每个码长 $n \in \{20, 23, 26, 29, 32, 38, 41, 50, 53, 62, 65, 74, 77, 86, 89, 98, 101, 113\}$，其相应最优码的基码字在表A.3列出，且不同码对应不同的群作用。当 $n = 20$ 和 23 时，用来作用的自同构为 $(0\ \frac{n-8}{3}\ \frac{2(n-8)}{3})(1\ \frac{n-8}{3}+1\ \frac{2(n-8)}{3}+1) \cdots (\frac{n-8}{3}-1\ \frac{2(n-8)}{3}-1\ n-9)(n-8\ n-6\ n-4)(n-7\ n-5\ n-3)(n-2)(n-1)$；当 $n = 32$ 时，用来作用的自同构为 $(0\ 3\ 6 \cdots 21)(1\ 4\ 7 \cdots 22)(2\ 5\ 8 \cdots 23)(24\ 26\ 28$

30 25 27 29 31)；当 $n=38$、50、62、74、86 以及 98 时，用来作用的自同构为 $(0\,2\,4\,\cdots\,n-4)(1\,3\,5\,\cdots\,n-3)(n-2\,n-1)$；而当 $n=29$、41、53、65、77、89、101 以及 113 时，用来作用的自同构为 $(0\,3\,6\,\cdots\,n-5)(1\,4\,7\,\cdots\,n-4)(2\,5\,8\,\cdots\,n-3)(n-2)(n-1)$。

A.4　表3.3中列出的 22 个特殊的 n 对应的 $A_3(n,6,4)$ 的上下界

引理 A.1　给定一个正整数集合 $M=\{14,\ 17,\ 18,\ 24,\ 35,\ 42,\ 44,\ 47,\ 56,\ 59,\ 68,\ 71,\ 72,\ 78,\ 80,\ 83,\ 84,\ 90,\ 92,\ 95,\ 96,\ 102\}$。当 $n\in M$ 时，$A_3(n,6,4)$ 的上下界如表3.3所示。

证明　上界可由引理3.15给出，因此我们只需考虑下界。首先，考虑 $n\equiv 0 \pmod 3$ 的情况。当 $n=18$ 和 24 时，可通过计算机搜索知 $A_3(n,6,4)\geqslant 33$ 和 55。当 $n=42,72,78,84,90,102$ 时，相应的码可分别通过型为 6^7，$15^4 12^1$，$15^4 18^1$，12^7，6^{15}，$6^{12}30^1$ 的 4-GDD 和符合要求的最优短码构造得到。当 $n=96$ 时，该码可利用型为 $7^{12}10^1$ 的 4-GDD 和一个具有性质 (B) 的 $(9,6,4)_3$ 码，通过应用定理3.20得到。

下面，我们考虑 $n\equiv 2 \pmod 3$ 的情形。$A_3(17,6,4)\geqslant 30$ 可通过计算机搜索得到。当 $n=14,35,44,47,56,59,68,71,80,83,92,95$ 时，相应的码可分别通过型为 2^7，$2^{12}11^1$，2^{22}，$2^{18}11^1$，$2^{24}8^1$，$2^{24}11^1$，$2^{24}20^1$，$2^{24}23^1$，$2^{27}26^1$，$2^{30}23^1$，$2^{33}26^1$，$2^{33}29^1$ 的 4-GDD 和符合要求的最优短码构造得到。 ∎

A.5　表格

表 A.1　码长满足 $n\equiv 0 \pmod 3$ 的小码基码字。

n	基码字			
15	$\{0_1,5_1,8_1,9_1\}$ $\{1_1,7_1,3_2\}$	$\{1_1,4_1,8_1,14_1\}$	$\{0_1,1_1,2_2\}$	$\{0_1,3_1,10_2\}$
21	$\{7_1,9_1,14_1,19_1\}$ $\{0_1,5_1,9_1,11_1\}$ $\{8_1,14_1,6_2\}$	$\{3_1,4_1,6_1,7_1\}$ $\{8_1,18_1,4_2\}$	$\{5_1,14_1,17_1,18_1\}$ $\{4_1,20_1,1_2\}$	$\{0_1,7_1,12_1,13_1\}$ $\{10_1,20_1,19_2\}$
27	$\{3_1,7_1,19_1,21_1\}$ $\{2_1,3_1,15_1,20_1\}$	$\{2_1,14_1,18_1,21_1\}$ $\{12_1,19_1,18_2\}$	$\{5_1,8_1,16_1,22_1\}$ $\{7_1,15_1,25_2\}$	$\{1_1,2_1,23_1,25_1\}$ $\{1_1,6_1,8_2\}$
30	$\{16_1,23_1,28_1,29_1\}$ $\{6_1,10_1,15_1,17_1\}$ $\{5_1,8_1,13_1,25_1\}$ $\{12_1,23_1,15_2\}$	$\{\mathbf{0_1},\mathbf{6_1},\mathbf{12_1},\mathbf{18_1}\}$ $\{4_1,12_1,14_1,29_1\}$ $\{1_1,6_1,8_1,22_1\}$ $\{15_1,20_1,0_2\}$	$\{17_1,18_1,27_1,29_1\}$ $\{\mathbf{1_1},\mathbf{7_1},\mathbf{13_1},\mathbf{19_1}\}$ $\{2_1,3_1,23_1,27_1\}$ $\{3_1,25_1,14_2\}$	$\{1_1,11_1,12_1,21_1\}$ $\{1_1,3_1,10_1,26_1\}$ $\{16_1,25_1,17_2\}$ $\{\mathbf{24_2},\mathbf{27_2}\}$
33	$\{\mathbf{2_1},\mathbf{10_1},\mathbf{18_1},\mathbf{26_1}\}$ $\{16_1,17_1,18_1,27_1\}$ $\{\mathbf{0_1},\mathbf{8_1},\mathbf{16_1},\mathbf{24_1}\}$ $\{14_1,32_1,31_2\}$	$\{\mathbf{7_1},\mathbf{15_1},\mathbf{23_1},\mathbf{31_1}\}$ $\{\mathbf{1_1},\mathbf{9_1},\mathbf{17_1},\mathbf{25_1}\}$ $\{5_1,10_1,20_1,24_1\}$ $\{5_1,18_1,1_2\}$	$\{9_1,27_1,30_1,31_1\}$ $\{0_1,10_1,23_1,30_1\}$ $\{0_1,3_1,9_1,15_1\}$ $\{18_1,25_1,22_2\}$	$\{0_1,12_1,26_1,31_1\}$ $\{1_1,4_1,10_1,31_1\}$ $\{0_1,5_1,7_1,25_1\}$ $\{21_1,32_1,0_2\}$
39	$\{23_1,27_1,30_1,34_1\}$ $\{10_1,24_1,33_1,35_1\}$	$\{15_1,28_1,30_1,36_1\}$ $\{8_1,11_1,13_1,33_1\}$	$\{13_1,14_1,23_1,31_1\}$ $\{0_1,5_1,23_1,29_1\}$	$\{0_1,22_1,26_1,38_1\}$ $\{18_1,19_1,30_1,38_1\}$

表 A.1　码长满足 $n \equiv 0 \pmod 3$ 的小码基码字。

n	基码字			
	$\{10_1, 19_1, 7_2\}$	$\{19_1, 34_1, 24_2\}$	$\{7_1, 13_1, 20_2\}$	
45	$\{16_1, 21_1, 22_1, 33_1\}$ $\{9_1, 20_1, 22_1, 34_1\}$ $\{2_1, 4_1, 19_1, 25_1\}$ $\{3_1, 31_1, 34_1, 43_1\}$ $\{36_1, 37_1, 24_2\}$	$\{8_1, 19_1, 37_1, 43_1\}$ $\{6_1, 34_1, 39_1, 40_1\}$ $\{0_1, 16_1, 20_1, 23_1\}$ $\{17_1, 44_1, 22_2\}$	$\{10_1, 34_1, 35_1, 37_1\}$ $\{5_1, 15_1, 25_1, 41_1\}$ $\{0_1, 25_1, 27_1, 34_1\}$ $\{16_1, 44_1, 35_2\}$	$\{15_1, 23_1, 28_1, 37_1\}$ $\{7_1, 22_1, 30_1, 37_1\}$ $\{0_1, 8_1, 26_1, 30_1\}$ $\{25_1, 28_1, 21_2\}$
48	$\{\mathbf{5_1, 17_1, 29_1, 41_1}\}$ $\{\mathbf{3_1, 15_1, 27_1, 39_1}\}$ $\{26_1, 28_1, 30_1, 35_1\}$ $\{19_1, 25_1, 42_1, 44_1\}$ $\{11_1, 12_1, 15_1, 38_1\}$ $\{0_1, 27_1, 30_1, 41_1\}$ $\{8_1, 18_1, 26_1, 39_1\}$ $\{34_1, 39_1, 0_2\}$	$\{\mathbf{0_1, 12_1, 24_1, 36_1}\}$ $\{\mathbf{4_1, 16_1, 28_1, 40_1}\}$ $\{10_1, 20_1, 27_1, 47_1\}$ $\{13_1, 15_1, 44_1, 47_1\}$ $\{1_1, 14_1, 29_1, 34_1\}$ $\{6_1, 15_1, 16_1, 43_1\}$ $\{9_1, 14_1, 15_2\}$ $\{11_1, 26_1, 20_2\}$	$\{\mathbf{1_1, 13_1, 25_1, 37_1}\}$ $\{13_1, 17_1, 23_1, 42_1\}$ $\{19_1, 27_1, 35_1, 45_1\}$ $\{4_1, 15_1, 25_1, 26_1\}$ $\{8_1, 16_1, 19_1, 24_1\}$ $\{1_1, 23_1, 31_1, 45_1\}$ $\{10_1, 41_1, 43_2\}$ $\{10_1, 26_1, 23_2\}$	$\{\mathbf{2_1, 14_1, 26_1, 38_1}\}$ $\{10_1, 14_1, 19_1, 28_1\}$ $\{0_1, 15_1, 22_1, 28_1\}$ $\{10_1, 11_1, 29_1, 42_1\}$ $\{0_1, 6_1, 13_1, 20_1\}$ $\{6_1, 7_1, 10_1, 39_1\}$ $\{15_1, 17_1, 40_2\}$
51	$\{23_1, 27_1, 48_1, 50_1\}$ $\{9_1, 26_1, 31_1, 48_1\}$ $\{0_1, 1_1, 7_1, 48_1\}$ $\{5_1, 23_1, 22_2\}$	$\{32_1, 34_1, 44_1, 47_1\}$ $\{4_1, 7_1, 25_1, 27_1\}$ $\{0_1, 16_1, 20_1, 40_1\}$ $\{4_1, 16_1, 42_2\}$	$\{6_1, 39_1, 44_1, 50_1\}$ $\{2_1, 25_1, 34_1, 39_1\}$ $\{8_1, 22_1, 29_1, 37_1\}$	$\{26_1, 27_1, 36_1, 42_1\}$ $\{7_1, 15_1, 39_1, 47_1\}$ $\{20_1, 46_1, 11_2\}$
57	$\{39_1, 40_1, 46_1, 48_1\}$ $\{20_1, 24_1, 39_1, 49_1\}$ $\{10_1, 20_1, 50_1, 55_1\}$	$\{0_1, 22_1, 23_1, 53_1\}$ $\{5_1, 10_1, 13_1, 28_1\}$ $\{4_1, 16_1, 36_1, 43_1\}$	$\{\mathbf{0_1, 14_1, 28_1, 42_1}\}$ $\{0_1, 17_1, 21_1, 37_1\}$ $\{7_1, 25_1, 13_2\}$	$\{\mathbf{1_1, 15_1, 29_1, 43_1}\}$ $\{9_1, 22_1, 31_1, 33_1\}$ $\{17_1, 56_1, 4_2\}$
63	$\{14_1, 22_1, 48_1, 59_1\}$ $\{18_1, 37_1, 42_1, 44_1\}$ $\{8_1, 37_1, 47_1, 57_1\}$ $\{0_1, 9_1, 59_1, 61_1\}$ $\{26_1, 32_1, 9_2\}$	$\{8_1, 19_1, 27_1, 59_1\}$ $\{11_1, 12_1, 18_1, 33_1\}$ $\{1_1, 16_1, 54_1, 55_1\}$ $\{6_1, 9_1, 37_1, 55_1\}$	$\{1_1, 13_1, 44_1, 60_1\}$ $\{18_1, 45_1, 52_1, 58_1\}$ $\{21_1, 26_1, 43_1, 56_1\}$ $\{10_1, 29_1, 32_2\}$	$\{26_1, 47_1, 61_1, 62_1\}$ $\{4_1, 8_1, 46_1, 62_1\}$ $\{12_1, 24_1, 32_1, 57_1\}$ $\{52_1, 55_1, 19_2\}$
87	$\{50_1, 69_1, 76_1, 82_1\}$ $\{10_1, 15_1, 56_1, 67_1\}$ $\{7_1, 29_1, 36_1, 79_1\}$ $\{0_1, 38_1, 51_1, 73_1\}$ $\{3_1, 4_1, 14_1, 81_1\}$ $\{40_1, 66_1, 13_2\}$	$\{31_1, 49_1, 73_1, 74_1\}$ $\{29_1, 30_1, 46_1, 69_1\}$ $\{6_1, 10_1, 14_1, 38_1\}$ $\{19_1, 40_1, 53_1, 57_1\}$ $\{6_1, 8_1, 31_1, 81_1\}$ $\{27_1, 30_1, 86_2\}$	$\{41_1, 43_1, 74_1, 83_1\}$ $\{19_1, 55_1, 58_1, 66_1\}$ $\{44_1, 47_1, 52_1, 59_1\}$ $\{2_1, 12_1, 31_1, 40_1\}$ $\{0_1, 15_1, 33_1, 57_1\}$ $\{40_1, 52_1, 72_2\}$	$\{15_1, 38_1, 44_1, 65_1\}$ $\{17_1, 35_1, 69_1, 74_1\}$ $\{23_1, 43_1, 59_1, 84_1\}$ $\{9_1, 15_1, 36_1, 80_1\}$ $\{4_1, 58_1, 60_1, 77_1\}$
93	$\{\mathbf{0_1, 23_1, 46_1, 69_1}\}$ $\{8_1, 12_1, 14_1, 26_1\}$ $\{27_1, 62_1, 79_1, 84_1\}$ $\{4_1, 7_1, 31_1, 57_1\}$ $\{19_1, 92_1, 30_2\}$	$\{2_1, 38_1, 51_1, 78_1\}$ $\{41_1, 54_1, 55_1, 88_1\}$ $\{7_1, 29_1, 58_1, 88_1\}$ $\{0_1, 20_1, 44_1, 91_1\}$	$\{21_1, 33_1, 49_1, 58_1\}$ $\{1_1, 49_1, 64_1, 85_1\}$ $\{8_1, 40_1, 59_1, 90_1\}$ $\{3_1, 7_1, 37_1, 56_1\}$	$\{20_1, 25_1, 28_1, 35_1\}$ $\{7_1, 13_1, 67_1, 74_1\}$ $\{20_1, 48_1, 57_1, 74_1\}$ $\{13_1, 85_1, 11_2\}$
99	$\{36_1, 60_1, 68_1, 98_1\}$ $\{56_1, 66_1, 79_1, 96_1\}$ $\{27_1, 79_1, 81_1, 82_1\}$ $\{49_1, 62_1, 79_1, 88_1\}$ $\{20_1, 25_1, 36_1, 77_1\}$ $\{3_1, 31_1, 37_1, 90_1\}$ $\{58_1, 73_1, 66_2\}$	$\{31_1, 41_1, 77_1, 89_1\}$ $\{14_1, 47_1, 69_1, 90_1\}$ $\{9_1, 38_1, 42_1, 90_1\}$ $\{40_1, 47_1, 56_1, 74_1\}$ $\{0_1, 49_1, 72_1, 86_1\}$ $\{0_1, 53_1, 61_1, 73_1\}$ $\{5_1, 7_1, 43_2\}$	$\{14_1, 15_1, 25_1, 46_1\}$ $\{11_1, 30_1, 39_1, 45_1\}$ $\{12_1, 23_1, 54_1, 79_1\}$ $\{10_1, 30_1, 52_1, 93_1\}$ $\{72_1, 75_1, 77_1, 92_1\}$ $\{10_1, 32_1, 39_1, 82_1\}$	$\{14_1, 28_1, 73_1, 92_1\}$ $\{19_1, 20_1, 23_1, 94_1\}$ $\{5_1, 29_1, 73_1, 78_1\}$ $\{3_1, 38_1, 63_1, 67_1\}$ $\{19_1, 33_1, 52_1, 70_1\}$ $\{68_1, 74_1, 29_2\}$
111	$\{12_1, 55_1, 57_1, 93_1\}$ $\{5_1, 78_1, 99_1, 107_1\}$ $\{11_1, 48_1, 95_1, 99_1\}$ $\{23_1, 59_1, 61_1, 71_1\}$ $\{29_1, 34_1, 42_1, 69_1\}$ $\{4_1, 16_1, 58_1, 59_1\}$ $\{9_1, 13_1, 27_1, 58_1\}$ $\{97_1, 101_1, 42_2\}$.	$\{46_1, 67_1, 97_1, 104_1\}$ $\{33_1, 72_1, 75_1, 100_1\}$ $\{42_1, 79_1, 90_1, 99_1\}$ $\{14_1, 33_1, 48_1, 55_1\}$ $\{0_1, 11_1, 32_1, 82_1\}$ $\{6_1, 7_1, 16_1, 39_1\}$ $\{5_1, 28_1, 74_1, 75_1\}$	$\{55_1, 70_1, 86_1, 103_1\}$ $\{17_1, 20_1, 63_1, 79_1\}$ $\{42_1, 76_1, 82_1, 95_1\}$ $\{2_1, 13_1, 18_1, 31_1\}$ $\{9_1, 53_1, 98_1, 108_1\}$ $\{0_1, 35_1, 50_1, 70_1\}$ $\{10_1, 95_1, 35_2\}$	$\{2_1, 16_1, 91_1, 108_1\}$ $\{53_1, 71_1, 78_1, 102_1\}$ $\{34_1, 62_1, 68_1, 101_1\}$ $\{1_1, 27_1, 85_1, 88_1\}$ $\{11_1, 68_1, 96_1, 98_1\}$ $\{0_1, 6_1, 20_1, 52_1\}$ $\{41_1, 88_1, 49_2\}$

表 A.1　码长满足 $n \equiv 0 \pmod 3$ 的小码基码字。

n	基码字			
123	$\{32_1, 34_1, 115_1, 118_1\}$	$\{40_1, 60_1, 85_1, 111_1\}$	$\{12_1, 40_1, 61_1, 117_1\}$	$\{0_1, 68_1, 119_1, 122_1\}$
	$\{56_1, 81_1, 116_1, 121_1\}$	$\{25_1, 55_1, 104_1, 119_1\}$	$\{52_1, 59_1, 111_1, 115_1\}$	$\{17_1, 26_1, 93_1, 107_1\}$
	$\{30_1, 32_1, 80_1, 111_1\}$	$\{23_1, 30_1, 87_1, 119_1\}$	$\{22_1, 97_1, 102_1, 114_1\}$	$\{9_1, 43_1, 105_1, 116_1\}$
	$\{37_1, 56_1, 84_1, 103_1\}$	$\{13_1, 22_1, 46_1, 56_1\}$	$\{10_1, 16_1, 28_1, 41_1\}$	$\{50_1, 80_1, 86_1, 99_1\}$
	$\{13_1, 17_1, 54_1, 119_1\}$	$\{70_1, 72_1, 78_1, 85_1\}$	$\{36_1, 56_1, 74_1, 109_1\}$	$\{3_1, 20_1, 32_1, 58_1\}$
	$\{31_1, 59_1, 82_1, 83_1\}$	$\{6_1, 42_1, 90_1, 100_1\}$	$\{20_1, 28_1, 66_1, 82_1\}$	$\{17_1, 57_1, 62_1, 94_1\}$
	$\{29_1, 70_1, 86_1, 97_1\}$	$\{4_1, 15_1, 36_1, 39_1\}$	$\{3_1, 33_1, 56_1, 66_1\}$	$\{6_1, 14_1, 28_1, 115_1\}$
	$\{9_1, 10_1, 54_1, 63_1\}$	$\{21_1, 47_1, 6_2\}$	$\{41_1, 84_1, 61_2\}$	$\{21_1, 83_1, 122_2\}$

表 A.2　码长满足 $n \equiv 1 \pmod 3$ 的小码基码字。

n	基码字			
16	$\{\mathbf{2_1, 6_1, 10_1, 14_1}\}$	$\{\mathbf{3_1, 7_1, 11_1, 15_1}\}$	$\{\mathbf{0_1, 4_1, 8_1, 12_1}\}$	$\{\mathbf{1_1, 5_1, 9_1, 13_1}\}$
	$\{0_1, 7_1, 10_1, 13_1\}$	$\{0_1, 9_1, 14_1, 15_1\}$	$\{2_1, 7_1, 12_2\}$	$\{1_1, 10_1, 3_2\}$
	$\{0_1, 3_1, 5_2\}$	$\{0_1, 1_1, 2_2\}$		
19	$\{0_1, 2_1, 5_1, 15_1\}$	$\{0_1, 1_1, 8_2\}$		
22	$\{1_1, 3_1, 9_1, 10_1\}$	$\{3_1, 5_1, 6_1, 14_1\}$	$\{1_1, 11_1, 18_1, 21_1\}$	$\{1_1, 2_1, 4_1, 17_1\}$
	$\{0_1, 16_1, 10_2\}$	$\{1_1, 5_1, 8_2\}$	$\{3_1, 8_1, 12_2\}$	
25	$\{\mathbf{4_1, 10_1, 16_1, 22_1}\}$	$\{\mathbf{5_1, 11_1, 17_1, 23_1}\}$	$\{\mathbf{0_1, 6_1, 12_1, 18_1}\}$	$\{\mathbf{1_1, 7_1, 13_1, 19_1}\}$
	$\{\mathbf{2_1, 8_1, 14_1, 20_1}\}$	$\{\mathbf{3_1, 9_1, 15_1, 21_1}\}$	$\{1_1, 8_1, 10_1, 23_1\}$	$\{0_1, 8_1, 15_1, 16_1\}$
	$\{7_1, 10_1, 15_1, 17_1\}$	$\{3_1, 13_1, 17_1, 18_1\}$	$\{5_1, 9_1, 20_1, 24_1\}$	$\{0_1, 4_1, 7_1, 24_1\}$
	$\{8_1, 9_1, 12_1, 22_1\}$	$\{6_1, 8_1, 11_1, 19_1\}$	$\{0_1, 17_1, 22_2\}$	$\{2_1, 23_1, 12_2\}$
	$\{3_1, 8_1, 7_2\}$	$\{4_1, 21_1, 5_2\}$	$\{12_1, 13_1, 15_2\}$	$\{10_1, 19_1, 14_2\}$
31	$\{0_1, 2_1, 8_1, 20_1\}$	$\{0_1, 5_1, 14_1, 21_1\}$	$\{0_1, 1_1, 4_2\}$	
34	$\{19_1, 24_1, 27_1, 31_1\}$	$\{4_1, 22_1, 24_1, 29_1\}$	$\{2_1, 17_1, 27_1, 29_1\}$	$\{1_1, 14_1, 18_1, 33_1\}$
	$\{2_1, 3_1, 15_1, 21_1\}$	$\{0_1, 1_1, 10_1, 11_1\}$	$\{1_1, 5_1, 7_1, 29_1\}$	$\{5_1, 12_1, 21_2\}$
	$\{6_1, 28_1, 25_2\}$	$\{11_1, 28_1, 14_2\}$		
40	$\{4_1, 12_1, 15_1, 27_1\}$	$\{22_1, 25_1, 33_1, 37_1\}$	$\{3_1, 11_1, 12_1, 17_1\}$	$\{4_1, 16_1, 17_1, 30_1\}$
	$\{\mathbf{0_1, 10_1, 20_1, 30_1}\}$	$\{\mathbf{1_1, 11_1, 21_1, 31_1}\}$	$\{4_1, 20_1, 38_1, 39_1\}$	$\{2_1, 6_1, 18_1, 39_1\}$
	$\{0_1, 6_1, 29_1, 38_1\}$	$\{1_1, 2_1, 15_1, 19_1\}$	$\{8_1, 15_1, 17_1, 33_1\}$	$\{13_1, 19_1, 32_1, 34_1\}$
	$\{20_1, 37_1, 2_2\}$	$\{6_1, 35_1, 11_2\}$	$\{26_1, 35_1, 33_2\}$	$\{9_1, 16_1, 12_2\}$

表 A.3　码长满足 $n \equiv 2 \pmod 3$ 的小码基码字。

n	基码字			
20	$\{\mathbf{0_1, 6_1, 12_1, 18_1}\}$	$\{\mathbf{1_1, 7_1, 13_1, 19_1}\}$	$\{\mathbf{2_1, 8_1, 14_1, 20_1}\}$	$\{13_1, 18_1, 22_1, 30_1\}$
	$\{6_1, 11_1, 29_1, 30_1\}$	$\{2_1, 5_1, 10_1, 29_1\}$	$\{12_1, 13_1, 25_1, 28_1\}$	$\{5_1, 16_1, 20_1, 28_1\}$
	$\{2_1, 3_1, 16_1, 19_1\}$	$\{0_1, 2_1, 9_1, 28_1\}$	$\{0_1, 3_1, 10_1, 11_1\}$	$\{5_1, 24_1, 15_2\}$
	$\{27_1, 29_1, 13_2\}$	$\{0_1, 22_1, 20_2\}$	$\{\mathbf{24_2, 25_2}\}$	$\{\mathbf{30_2, 31_2}\}$
23	$\{10_1, 13_1, 20_1, 21_1\}$	$\{6_1, 11_1, 12_1, 18_1\}$	$\{\mathbf{15_1, 17_1, 19_1, 21_1}\}$	$\{6_1, 9_1, 10_1, 19_1\}$
	$\{3_1, 14_1, 18_1, 19_1\}$	$\{1_1, 9_1, 15_1, 18_1\}$	$\{3_1, 9_1, 20_1, 22_1\}$	$\{0_1, 7_1, 18_1, 20_1\}$
	$\{1_1, 3_1, 8_1, 10_1\}$	$\{5_1, 7_1, 10_1, 15_1\}$	$\{2_1, 7_1, 13_1, 14_1\}$	$\{2_1, 4_1, 6_1, 21_1\}$
	$\{15_1, 22_1, 2_2\}$	$\{3_1, 15_1, 6_2\}$	$\{10_1, 14_1, 4_2\}$	$\{1_1, 22_1, 0_2\}$
	$\{7_1, 19_1, 8_2\}$	$\{\mathbf{15_2, 16_2}\}$	$\{\mathbf{21_2, 22_2}\}$	
26	$\{\mathbf{3_1, 9_1, 15_1, 21_1}\}$	$\{\mathbf{4_1, 10_1, 16_1, 22_1}\}$	$\{\mathbf{5_1, 11_1, 17_1, 23_1}\}$	$\{\mathbf{0_1, 6_1, 12_1, 18_1}\}$
	$\{\mathbf{1_1, 7_1, 13_1, 19_1}\}$	$\{\mathbf{2_1, 8_1, 14_1, 20_1}\}$	$\{4_1, 12_1, 19_1, 25_1\}$	$\{0_1, 10_1, 19_1, 20_1\}$
	$\{9_1, 12_1, 16_1, 17_1\}$	$\{8_1, 17_1, 18_1, 25_1\}$	$\{4_1, 6_1, 8_1, 9_1\}$	$\{3_1, 4_1, 17_1, 24_1\}$
	$\{7_1, 9_1, 20_1, 24_1\}$	$\{2_1, 5_1, 9_1, 18_1\}$	$\{2_1, 7_1, 10_1, 21_1\}$	$\{1_1, 21_1, 12_2\}$
	$\{0_1, 17_1, 1_2\}$	$\{1_1, 5_1, 22_2\}$	$\{5_1, 7_1, 14_2\}$	$\{5_1, 19_1, 3_2\}$
	$\{20_1, 22_1, 17_2\}$	$\{\mathbf{24_2, 25_2}\}$		
29	$\{6_1, 14_1, 16_1, 28_1\}$	$\{2_1, 8_1, 17_1, 24_1\}$	$\{3_1, 16_1, 19_1, 25_1\}$	$\{2_1, 3_1, 6_1, 7_1\}$

表 A.3　码长满足 $n \equiv 2 \pmod 3$ 的小码基码字。

n	基码字			
	$\{0_1, 2_1, 19_1, 27_1\}$	$\{1_1, 13_1, 14_1, 17_1\}$	$\{12_1, 18_1, 0_2\}$	$\{1_1, 21_1, 8_2\}$
	$\{9_1, 26_1, 7_2\}$	$\{\mathbf{27_2}, \mathbf{28_2}\}$		
32	$\{\mathbf{0_1}, \mathbf{6_1}, \mathbf{12_1}, \mathbf{18_1}\}$	$\{\mathbf{1_1}, \mathbf{7_1}, \mathbf{13_1}, \mathbf{19_1}\}$	$\{\mathbf{2_1}, \mathbf{8_1}, \mathbf{14_1}, \mathbf{20_1}\}$	$\{13_1, 18_1, 22_1, 30_1\}$
	$\{6_1, 11_1, 29_1, 30_1\}$	$\{2_1, 5_1, 10_1, 29_1\}$	$\{12_1, 13_1, 25_1, 28_1\}$	$\{5_1, 16_1, 20_1, 28_1\}$
	$\{2_1, 3_1, 16_1, 19_1\}$	$\{0_1, 2_1, 9_1, 28_1\}$	$\{0_1, 3_1, 10_1, 11_1\}$	$\{5_1, 24_1, 15_2\}$
	$\{27_1, 29_1, 13_2\}$	$\{0_1, 22_1, 20_2\}$	$\{\mathbf{24_2}, \mathbf{25_2}\}$	$\{\mathbf{30_2}, \mathbf{31_2}\}$
38	$\{12_1, 15_1, 16_1, 28_1\}$	$\{10_1, 16_1, 17_1, 24_1\}$	$\{7_1, 12_1, 22_1, 27_1\}$	$\{5_1, 9_1, 17_1, 19_1\}$
	$\{0_1, 2_1, 13_1, 19_1\}$	$\{\mathbf{0_1}, \mathbf{9_1}, \mathbf{18_1}, \mathbf{27_1}\}$	$\{5_1, 37_1, 16_2\}$	$\{32_1, 36_1, 29_2\}$
	$\{\mathbf{36_2}, \mathbf{37_2}\}$			
41	$\{6_1, 25_1, 29_1, 39_1\}$	$\{11_1, 15_1, 20_1, 25_1\}$	$\{10_1, 17_1, 36_1, 40_1\}$	$\{17_1, 32_1, 34_1, 35_1\}$
	$\{0_1, 2_1, 8_1, 9_1\}$	$\{10_1, 19_1, 27_1, 38_1\}$	$\{2_1, 14_1, 24_1, 27_1\}$	$\{0_1, 4_1, 6_1, 27_1\}$
	$\{4_1, 10_1, 15_1, 22_1\}$	$\{7_1, 21_1, 6_2\}$	$\{10_1, 13_1, 23_2\}$	$\{20_1, 28_1, 4_2\}$
	$\{\mathbf{39_2}, \mathbf{40_2}\}$			
50	$\{\mathbf{0_1}, \mathbf{12_1}, \mathbf{24_1}, \mathbf{36_1}\}$	$\{\mathbf{1_1}, \mathbf{13_1}, \mathbf{25_1}, \mathbf{37_1}\}$	$\{14_1, 20_1, 25_1, 36_1\}$	$\{2_1, 17_1, 31_1, 47_1\}$
	$\{28_1, 35_1, 37_1, 45_1\}$	$\{17_1, 24_1, 37_1, 43_1\}$	$\{23_1, 27_1, 36_1, 44_1\}$	$\{0_1, 2_1, 20_1, 23_1\}$
	$\{5_1, 6_1, 10_1, 20_1\}$	$\{23_1, 48_1, 22_2\}$	$\{26_1, 49_1, 3_2\}$	$\{\mathbf{48_2}, \mathbf{49_2}\}$
53	$\{16_1, 23_1, 24_1, 50_1\}$	$\{28_1, 42_1, 43_1, 46_1\}$	$\{12_1, 15_1, 31_1, 50_1\}$	$\{5_1, 11_1, 25_1, 41_1\}$
	$\{0_1, 12_1, 22_1, 46_1\}$	$\{12_1, 30_1, 39_1, 44_1\}$	$\{7_1, 9_1, 32_1, 50_1\}$	$\{18_1, 20_1, 43_1, 49_1\}$
	$\{0_1, 28_1, 29_1, 51_1\}$	$\{0_1, 6_1, 17_1, 21_1\}$	$\{0_1, 8_1, 13_1, 52_1\}$	$\{5_1, 7_1, 16_1, 46_1\}$
	$\{43_1, 47_1, 3_2\}$	$\{12_1, 32_1, 19_2\}$	$\{2_1, 44_1, 5_2\}$	$\{\mathbf{51_2}, \mathbf{52_2}\}$
62	$\{3_1, 15_1, 28_1, 37_1\}$	$\{0_1, 16_1, 48_1, 50_1\}$	$\{13_1, 31_1, 33_1, 52_1\}$	$\{\mathbf{0_1}, \mathbf{15_1}, \mathbf{30_1}, \mathbf{45_1}\}$
	$\{23_1, 30_1, 31_1, 55_1\}$	$\{0_1, 18_1, 29_1, 40_1\}$	$\{5_1, 9_1, 52_1, 55_1\}$	$\{0_1, 4_1, 31_1, 37_1\}$
	$\{9_1, 10_1, 46_1, 53_1\}$	$\{0_1, 5_1, 8_1, 14_1\}$	$\{32_1, 60_1, 51_2\}$	$\{33_1, 60_1, 38_2\}$
	$\{\mathbf{60_2}, \mathbf{61_2}\}$			
65	$\{16_1, 21_1, 22_1, 62_1\}$	$\{18_1, 24_1, 28_1, 39_1\}$	$\{11_1, 25_1, 27_1, 64_1\}$	$\{16_1, 39_1, 47_1, 63_1\}$
	$\{12_1, 36_1, 49_1, 61_1\}$	$\{10_1, 27_1, 43_1, 47_1\}$	$\{9_1, 26_1, 31_1, 41_1\}$	$\{12_1, 17_1, 21_1, 55_1\}$
	$\{7_1, 20_1, 59_1, 62_1\}$	$\{1_1, 9_1, 16_1, 23_1\}$	$\{2_1, 11_1, 12_1, 31_1\}$	$\{28_1, 37_1, 55_1, 56_1\}$
	$\{0_1, 11_1, 23_1, 36_1\}$	$\{20_1, 26_1, 54_1, 56_1\}$	$\{0_1, 18_1, 44_1, 51_1\}$	$\{7_1, 31_1, 28_2\}$
	$\{6_1, 37_1, 9_2\}$	$\{13_1, 29_1, 11_2\}$	$\{\mathbf{63_2}, \mathbf{64_2}\}$	
74	$\{\mathbf{1_1}, \mathbf{19_1}, \mathbf{37_1}, \mathbf{55_1}\}$	$\{12_1, 14_1, 15_1, 26_1\}$	$\{4_1, 15_1, 37_1, 61_1\}$	$\{29_1, 43_1, 66_1, 71_1\}$
	$\{30_1, 38_1, 67_1, 68_1\}$	$\{\mathbf{0_1}, \mathbf{18_1}, \mathbf{36_1}, \mathbf{54_1}\}$	$\{44_1, 57_1, 61_1, 69_1\}$	$\{6_1, 26_1, 47_1, 57_1\}$
	$\{4_1, 8_1, 31_1, 48_1\}$	$\{4_1, 23_1, 28_1, 43_1\}$	$\{0_1, 6_1, 16_1, 69_1\}$	$\{3_1, 16_1, 59_1, 66_1\}$
	$\{10_1, 17_1, 55_1, 57_1\}$	$\{36_1, 73_1, 62_2\}$	$\{9_1, 72_1, 15_2\}$	$\{\mathbf{72_2}, \mathbf{73_2}\}$
77	$\{33_1, 51_1, 53_1, 72_1\}$	$\{23_1, 35_1, 62_1, 72_1\}$	$\{18_1, 24_1, 28_1, 43_1\}$	$\{25_1, 30_1, 65_1, 76_1\}$
	$\{23_1, 30_1, 46_1, 57_1\}$	$\{20_1, 24_1, 48_1, 70_1\}$	$\{12_1, 62_1, 67_1, 75_1\}$	$\{31_1, 32_1, 33_1, 65_1\}$
	$\{6_1, 18_1, 19_1, 67_1\}$	$\{1_1, 14_1, 38_1, 68_1\}$	$\{1_1, 47_1, 50_1, 65_1\}$	$\{1_1, 11_1, 31_1, 43_1\}$
	$\{10_1, 19_1, 42_1, 57_1\}$	$\{3_1, 10_1, 14_1, 45_1\}$	$\{6_1, 29_1, 35_1, 51_1\}$	$\{5_1, 7_1, 58_1, 64_1\}$
	$\{1_1, 4_1, 32_1, 45_1\}$	$\{8_1, 66_1, 69_1, 74_1\}$	$\{37_1, 54_1, 45_2\}$	$\{43_1, 64_1, 50_2\}$
	$\{50_1, 67_1, 31_2\}$	$\{\mathbf{75_2}, \mathbf{76_2}\}$		
86	$\{42_1, 53_1, 73_1, 75_1\}$	$\{\mathbf{0_1}, \mathbf{21_1}, \mathbf{42_1}, \mathbf{63_1}\}$	$\{0_1, 4_1, 49_1, 73_1\}$	$\{6_1, 22_1, 31_1, 81_1\}$
	$\{21_1, 59_1, 69_1, 77_1\}$	$\{20_1, 23_1, 56_1, 75_1\}$	$\{0_1, 52_1, 53_1, 57_1\}$	$\{24_1, 36_1, 44_1, 71_1\}$
	$\{1_1, 6_1, 13_1, 20_1\}$	$\{5_1, 28_1, 50_1, 52_1\}$	$\{4_1, 32_1, 38_1, 78_1\}$	$\{49_1, 50_1, 65_1, 79_1\}$
	$\{0_1, 13_1, 54_1, 71_1\}$	$\{6_1, 29_1, 32_1, 73_1\}$	$\{20_1, 84_1, 2_2\}$	$\{49_1, 85_1, 55_2\}$
	$\{\mathbf{84_2}, \mathbf{85_2}\}$			
89	$\{55_1, 62_1, 64_1, 81_1\}$	$\{27_1, 32_1, 37_1, 88_1\}$	$\{37_1, 51_1, 65_1, 87_1\}$	$\{17_1, 40_1, 55_1, 86_1\}$
	$\{31_1, 58_1, 70_1, 76_1\}$	$\{19_1, 20_1, 49_1, 85_1\}$	$\{37_1, 42_1, 61_1, 84_1\}$	$\{30_1, 54_1, 60_1, 76_1\}$
	$\{30_1, 63_1, 64_1, 67_1\}$	$\{16_1, 49_1, 51_1, 86_1\}$	$\{20_1, 34_1, 53_1, 59_1\}$	$\{4_1, 12_1, 38_1, 48_1\}$
	$\{16_1, 26_1, 72_1, 75_1\}$	$\{12_1, 27_1, 56_1, 86_1\}$	$\{12_1, 35_1, 44_1, 79_1\}$	$\{14_1, 65_1, 77_1, 85_1\}$
	$\{8_1, 29_1, 30_1, 55_1\}$	$\{5_1, 9_1, 36_1, 84_1\}$	$\{14_1, 39_1, 46_1, 59_1\}$	$\{5_1, 21_1, 32_1, 39_1\}$
	$\{0_1, 2_1, 13_1, 17_1\}$	$\{6_1, 56_1, 53_2\}$	$\{18_1, 27_1, 76_2\}$	$\{3_1, 79_1, 24_2\}$
	$\{\mathbf{87_2}, \mathbf{88_2}\}$			
	$\{13_1, 71_1, 72_1, 75_1\}$	$\{28_1, 80_1, 85_1, 95_1\}$	$\{20_1, 42_1, 46_1, 73_1\}$	$\{\mathbf{0_1}, \mathbf{24_1}, \mathbf{48_1}, \mathbf{72_1}\}$
98				

表 A.3　码长满足 $n \equiv 2 \pmod 3$ 的小码基码字。

n	基码字			
	$\{\mathbf{1_1}, \mathbf{25_1}, \mathbf{49_1}, \mathbf{73_1}\}$	$\{22_1, 41_1, 71_1, 90_1\}$	$\{6_1, 24_1, 82_1, 89_1\}$	$\{5_1, 21_1, 26_1, 62_1\}$
	$\{3_1, 17_1, 53_1, 59_1\}$	$\{26_1, 40_1, 72_1, 82_1\}$	$\{28_1, 45_1, 62_1, 63_1\}$	$\{35_1, 46_1, 67_1, 87_1\}$
	$\{4_1, 6_1, 12_1, 75_1\}$	$\{0_1, 9_1, 80_1, 93_1\}$	$\{1_1, 9_1, 24_1, 54_1\}$	$\{8_1, 67_1, 69_1, 95_1\}$
	$\{1_1, 52_1, 64_1, 75_1\}$	$\{37_1, 96_1, 8_2\}$	$\{68_1, 97_1, 61_2\}$	$\{\mathbf{96_2}, \mathbf{97_2}\}$
101	$\{24_1, 38_1, 43_1, 50_1\}$	$\{26_1, 50_1, 72_1, 84_1\}$	$\{32_1, 49_1, 93_1, 100_1\}$	$\{51_1, 64_1, 75_1, 84_1\}$
	$\{31_1, 44_1, 53_1, 93_1\}$	$\{22_1, 59_1, 77_1, 78_1\}$	$\{17_1, 52_1, 64_1, 73_1\}$	$\{14_1, 25_1, 55_1, 84_1\}$
	$\{31_1, 33_1, 34_1, 48_1\}$	$\{34_1, 35_1, 72_1, 80_1\}$	$\{29_1, 36_1, 42_1, 72_1\}$	$\{22_1, 58_1, 63_1, 83_1\}$
	$\{0_1, 21_1, 72_1, 89_1\}$	$\{15_1, 17_1, 43_1, 47_1\}$	$\{10_1, 42_1, 60_1, 64_1\}$	$\{54_1, 59_1, 65_1, 79_1\}$
	$\{20_1, 43_1, 59_1, 92_1\}$	$\{0_1, 16_1, 73_1, 83_1\}$	$\{6_1, 13_1, 41_1, 66_1\}$	$\{8_1, 37_1, 56_1, 84_1\}$
	$\{1_1, 16_1, 40_1, 67_1\}$	$\{5_1, 7_1, 13_1, 47_1\}$	$\{3_1, 6_1, 37_1, 60_1\}$	$\{2_1, 6_1, 70_1, 99_1\}$
	$\{44_1, 47_1, 0_2\}$	$\{29_1, 44_1, 8_2\}$	$\{76_1, 84_1, 94_2\}$	$\{\mathbf{99_2}, \mathbf{100_2}\}$
113	$\{58_1, 91_1, 95_1, 101_1\}$	$\{10_1, 29_1, 45_1, 100_1\}$	$\{12_1, 34_1, 81_1, 87_1\}$	$\{57_1, 72_1, 101_1, 109_1\}$
	$\{60_1, 64_1, 71_1, 106_1\}$	$\{22_1, 49_1, 107_1, 109_1\}$	$\{18_1, 42_1, 65_1, 101_1\}$	$\{0_1, 62_1, 94_1, 112_1\}$
	$\{32_1, 46_1, 52_1, 77_1\}$	$\{59_1, 60_1, 74_1, 105_1\}$	$\{26_1, 43_1, 81_1, 111_1\}$	$\{2_1, 41_1, 83_1, 104_1\}$
	$\{13_1, 27_1, 47_1, 94_1\}$	$\{27_1, 60_1, 65_1, 77_1\}$	$\{28_1, 30_1, 48_1, 60_1\}$	$\{11_1, 22_1, 38_1, 45_1\}$
	$\{27_1, 43_1, 48_1, 61_1\}$	$\{38_1, 67_1, 82_1, 95_1\}$	$\{12_1, 22_1, 31_1, 94_1\}$	$\{16_1, 19_1, 27_1, 87_1\}$
	$\{7_1, 29_1, 62_1, 87_1\}$	$\{9_1, 35_1, 58_1, 94_1\}$	$\{17_1, 60_1, 68_1, 87_1\}$	$\{5_1, 10_1, 55_1, 67_1\}$
	$\{17_1, 27_1, 30_1, 55_1\}$	$\{3_1, 12_1, 66_1, 73_1\}$	$\{5_1, 8_1, 45_1, 84_1\}$	$\{17_1, 110_1, 21_2\}$
	$\{93_1, 94_1, 95_2\}$	$\{47_1, 71_1, 1_2\}$	$\{\mathbf{111_2}, \mathbf{112_2}\}$	

致　谢

时光荏苒，研究生入学那日仿佛还在昨天。在科大的五年里，我收获了很多，也成长了很多。每一步对生活的探索，都有来自身边朋友们给予的勇气。尽管博士生涯已接近尾声，但这里的一切，让我对未来充满信心。至此，我想向我的老师、同学、朋友和家人们，表达最真挚的感谢。

感谢我的导师张先得教授。在科大的这段时间里，张老师给了我很多指导和关心，不管是在生活中，还是在学习上。这里的学生勤奋，但老师们更勤奋。在学习上，不论何时，你遇到了什么困难，总能在办公室找到张老师。为了让我提升对科研领域的眼界，她总是积极的推荐我参与各种学术会议和暑期班，让我能够接触到很多优秀的团队，见识到很多优秀的工作。在生活中，张老师是一个很鲜活立体的形象，她会打羽毛球、游泳等，还会参加大合唱。不论是对科研，还是生活，总能感觉到她对身边一切的热爱。在这里感受到的美好，让我对未来报以期待。

感谢首都师范大学的葛根年教授。葛老师严谨治学的态度、高瞻远瞩的学术视野，让我敬佩也收益良多。每一次在北京团队举办的寒暑期班的学习中，都能够感受到葛老师对学术的热情。而我们参与其中的小伙伴，不仅能够认识一些优秀的同行们，也能学习到很多前沿的知识和新的方法，在面对以后的学术问题上，能够有更深的理解。

感谢安徽大学的施敏加教授。施老师于我而言，有知遇之恩。正是他引领着我进入了组合数学与信息交叉科学研究团队，同时在进入研究生生活之前，我也有幸跟随施老师的科研团队一起，学习编码理论知识，在他的团队中提前感受学术氛围和科研气息。

感谢实验室一起学习的小伙伴们。是你们的陪伴和鼓励，让我的科研生活更加丰富多彩。首先感谢师兄们：余文俊、叶左。记得每次去北京学习的时候，面对陌生的城市，自己总是惶惶不安，余师兄总会去车站接待。每次学习上遇到一些困惑，总会向师兄们求助，不论是面对面，还是隔着网线，他们总是能给出细致的解答。感谢与我一届的小伙伴：石飞。我们不仅同为一个导师，也一起在网空学院转博，进入网空党支部，共同面对新学科和新环境的挑战，感谢一直以来的鼓励与陪伴。无论是在生活中、学习中，还是在支部工作中，你积极乐观的人生态度、孜孜不倦敢于提问勇于坚持的学习态度以及对待工作的热情，都影响并激励着我，感谢能够一起学习与成长。感谢与我一起合作的师弟师妹：马一鸣、李言智、魏歆。通过与他们的合作和讨论，我对研究的课题有了更深更全面的认识。感谢一起学习的同门们：吴荣胜、张树亮、王国平、章宛晨、于克凡、何奕

昀、任年新、王琛、王飞、朱玮奇等。感谢北京团队毕业的优秀师兄们：张一炜、张韬、汪馨、丁报昆、马景学等。感谢你们在学习与生活中分享的经验与建议。感谢在北京团队的小伙伴们：孔祥梁、韩雪娇、钱昺辰、徐子翔、奚元霄、兰昭君等。感谢在北京学习期间，你们在学习上的帮助与生活上的照顾。

感谢 1202 的小伙伴们：杨倩倩、谢天颖、曹梦月、祖春蕾、火清翼、杨天驰、高峻、徐淼。1202 实验室的场地虽不大，但却能够让我们在这里学习、成长并收获快乐。

感谢我的室友们：程小雨、成霄翔、王艺。我们四个自硕士起就住在一起，彼此默契，相互鼓励，一起约饭。转博后，幸运的是仍能跟雨哥一起，搬到如今宿舍，每天一起挣扎早起，相互鼓励互相“伤害”，雨哥无所不能，上知天文下知地理引经据典，是真正生活中的小太阳。感谢有你们的陪伴，让这五年的蜗壳生活充满欢声笑语。

感谢我的朋友们：周思雨、汪正发、盛长浩、徐璇、杨柳青、许小燕、李江红、陈亮等。我们在人生的不同阶段相遇，也走向各自不同的方向。感谢你们一直以来的鼓励和陪伴，你们带给我的不仅是快乐和勇气，还有对未来的期待。

感谢亲爱的徐亮亮一直以来对我的支持和陪伴。我们在大学相遇，一起迈向研究生生活，一起面对学习和生活中的困难，彼此分享喜悦，见证成长。在面临选择的时候，相互鼓励。不管我走到哪里，你都与我一同前进。你会在我每一次受挫时，肯定我鼓励我，帮我重塑信心。你还是一个万能代码库，在我每次因写代码而焦灼时，帮我理清思路。感谢有你，让我无惧于任何挑战！相信我们，一定能够在更加美好的未来中相知相守。

感谢一直以来默默支持我、鼓励我的父亲和母亲。感谢你们一直以来毫无保留的爱与支持，使我能够有动力坚持走下去。你们永远是我最坚强的后盾，在我面临各种选择的时候，给我竭尽所能的帮助和鼓励。在今后的工作和生活中，我会努力快乐的成长，不让你们担心，不辜负每一份付出。感谢我的弟弟，尽管他比我小，却时常像个长辈一样，给我输出鸡汤，告诉我后面有他，无所畏惧的前进即可，很幸运能够拥有这样可爱的他。

最后，感谢本文的审稿以及答辩专家，在百忙之中对我的毕业论文把好最后一道关。再次感谢生命中遇到的老师、同学、朋友以及亲人们，祝福您们！

谨以此文，献给所有我爱的人，期待更加美好的未来。

在读期间发表的学术论文与取得的研究成果

已发表论文

1. **Tingting Chen**, Yiming Ma and Xiande Zhang,"Optimal Codes With Small Constant Weight in ℓ_1-Metric", in IEEE Transactions on Information Theory (IEEE TIT 2021), vol. 67, no. 7, pp. 4239-4254, July 2021, doi: 10.1109/TIT.2021.3052191.（中国计算机协会 CCF 推荐 A 类期刊）
2. **Tingting Chen** and Xiande Zhang, "Sparse and Balanced MDS Codes over Small Fields", published online in IEEE Transactions on Information Theory (IEEE TIT 2022), doi: 10.1109/TIT.2022.3162524.（中国计算机协会 CCF 推荐 A 类期刊）
3. Xin Wei, **Tingting Chen** and Xiande Zhang, "Optimal Ternary Codes With Weight w and Distance $2w-2$ in ℓ_1-Metric", in IEEE Transactions on Information Theory (IEEE TIT 2021), vol. 67, no. 11, pp. 7221-7231, Nov. 2021, doi: 10.1109/TIT.2021.3105688.（中国计算机协会 CCF 推荐 A 类期刊）

荣誉奖项

1. 苏州育才奖学金，2021 年。