

• 通信与信息工程 •



基于 RBIBD 的最优局部修复码构造

王 静*, 李静辉, 杨佳蓉, 王 娥

(长安大学信息工程学院 西安 710064)

【摘要】随着数据量的迅速增长,对存储海量数据的分布式存储系统的可靠性和有效性的要求日益增加。局部修复码(LRCs)具有良好的修复局部性,能够有效实现海量数据在分布式存储系统中的可靠高效存储,构造具有 (r,t) 局部性的局部修复码已经成为当前研究的热点。为此,提出了一种基于可分解均衡不完全区组设计(RBIBD)的最优局部修复码的构造方法,构造信息位具有 (r,t) 局部性的二元最优单校验 LRCs。性能分析表明,构造的 LRCs 达到了最小距离最优边界,且在码率上表现得更优。

关 键 词 分布式存储系统; 局部修复码; 最小距离; 均衡不完全区组设计

中图分类号 TN911.2 **文献标志码** A **doi**:10.12178/1001-0548.2022158

Construction of Optimal Locally Repairable Codes Based on RBIBD

WANG Jing*, LI Jinghui, YANG Jiarong, and WANG E

(School of Information Engineering, Chang'an University Xi'an 710064)

Abstract With the rapid growth of data, the requirements of the reliability and effectiveness for distributed storage systems are increasing. Locally repairable codes (LRCs) have better locality, which can effectively realize the reliable and efficient storage of massive data in distributed storage system. It has become a research hotspot to construct locally repairable codes with (r, t) locality. Therefore, this paper proposes a construction method of optimal locally repairable codes based on resolvable balanced incomplete block design (RBIBD), which can construct binary optimal single parity LRCs with (r, t) locality of information symbols. The performance analyses show that, the constructed LRCs reach the minimum distance bound, and compared with the LRCs proposed by Xia *et al.*, the LRCs proposed in this paper perform better in code rate. In particular, when the availability $t=2$, the LRCs is also the code with the optimal code rate.

Key words distributed storage system; locally repairable code; minimum distance; resolvable balanced incomplete block design

随着互联网的快速发展,各种数据出现爆炸式增长,分布式存储系统得到了广泛应用^[1]。在分布式存储系统中,通常采取数据冗余技术,最简单的方式是复制^[2],这需要非常高的存储开销。纠删码技术可降低存储开销^[3],但是当存储节点故障时,修复成本远高于复制。为了解决上述问题,文献[4]提出了局部修复码(locally repairable codes, LRCs)的概念。LRCs 具有很好的局部性,在修复过程中通过访问少量的节点来降低修复成本。目前对 LRCs 的研究主要集中在参数研究和其构造方面。

已有不少文献对 LRCs 的参数边界进行了研究。文献[5]提出的修复局部性 r 是衡量修复效率的重

要指标,对于信息位具有局部性 r 的 LRCs 给出了码的最小距离的上界。文献[6-9]给出了最小距离达到此上界的 LRCs 的相关构造。为了在多个节点故障的情况下实现局部恢复,文献[10]提出了 (r,t) 局部性的概念,并且证明了信息位具有 (r,t) 局部性的单校验 LRCs 最小距离的上界,如果码的最小距离达到该上界,则该码是最小距离最优的码。文献[11]提出并证明了具有 t 个大小为 r 的不相交修复集的 (r,t) -LRCs 的码率上界。文献[12]通过将文献[5]提出的修复局部性进行推广,构造了可以同时恢复两个 Erasure(删余)的 LRCs,证明了可用性 $t=2$ 时的最优码率。

收稿日期: 2022-05-24; 修回日期: 2022-10-09

基金项目: 国家自然科学基金(62001059); 陕西省重点研发计划(2021GY-019)

作者简介: 王静(1982-),女,博士,教授,主要从事网络编码及分布式存储编码等方面的研究。

*通信作者: 王静, E-mail: 342573224@qq.com

除以上研究文献外,目前也有大量文献对LRCs的构造进行研究。文献[11]基于 t 个二进制 $(r+1, r)$ 奇偶校验码的直积构造了具有 (r, t) 局部性的二元局部修复码(binary locally repairable codes, BLRCs),构造出的BLRCs虽然能够实现任意的 (r, t) 局部性,但是其码率较小。文献[13]利用有限域上的迹函数构造了一类循环 (r, t) LRCs,构造出的BLRCs达到了最优最小距离,但是此构造方法只能构造特定参数的BLRCs且码率较小。文献[14]采用有限平面LDPC码和阵列LDPC码构造了具有 (r, t) 局部性的BLRCs,构造出的BLRCs虽然达到了最优最小距离,但是参数条件限制较多且码率较小。文献[15]对文献[14]中的构造进行了扩展,进一步构造了最小距离更大的最优单校验 (r, t) LRCs,但其码率仍然不高。文献[16]基于 (r, t) 超图构造的BLRCs的最小距离达到了最优最小距离界,但是构造出的BLRCs具有很大的参数限制。文献[17]采用矩阵迭代的方法构造了具有 (r, t) 局部性的BLRCs,但是这种构造方法只能构造可用性 $t=2$ 的BLRCs,限制了可用性的参数选择。文献[18]基于随机函数提出了一种新的局部性 $r=2$ 且具有不均匀可用性的BLRCs,但是构造出的BLRCs码率较低。

为解决上述LRCs构造中存在的参数限制较多且码率较低的问题,本文基于可分解均衡不完全区组设计(resolvable balanced incomplete block design, RBIBD),提出了信息位具有 (r, t) 局部性的单校验最优BLRCs的构造方法。首先利用方阵循环移位进行RBIBD设计,运用RBIBD若干平行类中的区组与有限集中元素的对应关系构造关联矩阵,基于关联矩阵构造BLRCs。构造的BLRCs达到了最优最小距离界,是最小距离最优的LRCs;特别地,当可用性 $t=2$ 时,该BLRCs达到了文献[12]提出的最优码率边界,同时也是码率最优的LRCs。与文献[15]提出的单校验BLRCs相比,构造的BLRCs在码率上表现得更优并且参数选择更加灵活,和文献[11]提出的直积码相比,该构造在码率上表现得也更优。

1 基础知识

1.1 局部修复码

设 C 是有限域 F_q 上的 (n, k) 线性码, C 的生成矩阵 $G = [g_1, g_2, \dots, g_n]$, 如果 C 中码字的第 i 位可以由其余最多 r 位线性表示出来, 即 g_i 是 G 中 r 列的线性

组合, 那么称 C 的第 i 位具有局部性 r 。这样的码 C 称之为局部修复码。

定义 1^[10] 有限域 F_q 中的 (n, k, d) 线性分组码, 给定 $[n] = \{1, 2, \dots, n\}$, $c = (c_1, c_2, \dots, c_n)$ 是码字, 如果编码符号 c_i 具有 (r, t) 局部性, 那么需要满足下列条件:

1) 存在 t 个子集, 满足 $\varphi_1(i), \varphi_2(i), \dots, \varphi_t(i) \subset [n] \setminus \{i\}$, 即 c_i 能够从 $\varphi_j(i) (j \in [t])$ 中恢复出来, $\varphi_j(i) (j \in [t])$ 称为 c_i 的修复集;

2) $|\varphi_j(i)| \leq r, j \in [t]$;

3) $\varphi_j(i) \cap \varphi_l(i) = \emptyset, j \neq l \in [t]$ 。

若每个信息位码元的每个修复集只有一个校验码元, 则称此码为单校验 $(n, k, r, t)_i$ LRCs。如果只有信息符号具有 (r, t) 局部性, 则此LRCs具有信息 (r, t) 局部性。

定义 2^[19] 在 (n, k) 分组码中, 任两个码字之间距离的最小值, 称为该分组码的最小汉明距离, 简称最小距离。最小距离是 (n, k) 分组码的重要参数, 它表明了分组码抗干扰能力的大小。对于局部修复码来说, 码 C 的最小距离越大, 可容忍的故障节点数量越多。

定理 1^[20] 对于信息位具有 (r, t) 局部性的 (n, k, d) LRCs, 最小距离应满足:

$$d \geq t + 1 \quad (1)$$

定理 2^[10] 若LRCs的所有信息位码元的每个修复集中只包含一个校验位, 那么该单校验 $(n, k, r, t)_i$ LRCs的最小距离满足:

$$d \leq n - k - \left\lceil \frac{kt}{r} \right\rceil + t + 1 \quad (2)$$

称达到边界(2)的LRCs是最小距离最优的单校验 $(n, k, r, t)_i$ LRCs。

定理 3^[11] 若 (n, k, r, t) LRCs具有 (r, t) 局部性, 那么该码码率满足:

$$R \leq \frac{1}{\prod_{i=1}^t (1 + \frac{1}{ir})} \quad (3)$$

称达到边界(3)的LRCs是码率最优 (n, k, r, t) LRCs。

定理 4 特别地, 文献[12]进一步地提出了可用性 $t=2$ 的最优码率的 $(n, k, r, t=2)$ LRCs, 该码码率满足:

$$R \leq \frac{r}{r+2} \quad (4)$$

1.2 区组设计

设 $X = \{x_1, x_2, \dots, x_v\}$ 是一个有限集, $B_1, B_2, \dots, B_b$

为 X 的 b 个子集, 则 $B = \{B_1, B_2, \dots, B_b\}$ 称为 X 上的一个构造。如果 B 中没有空集, 也没有重复的子集, 则 B 称为 X 上的一个组合设计, 其中 X 称为该设计的基集, $B_i (1 \leq i \leq b)$ 称为该设计的区组。

定义 3^[21] 设 $X = \{x_1, x_2, \dots, x_v\}$ 是一个有限集, $B = \{B_1, B_2, \dots, B_b\}$ 为 X 的子集的集合, 如果各区组 $B_i \in B$ 包含的元素数目相同, X 中各元素在 B 中出现的数目也相同, 则 (X, B) 构成区组设计。

定义 4^[22] 设 $X = \{x_1, x_2, \dots, x_v\}$ 是一个有限集, $B = \{B_1, B_2, \dots, B_b\}$ 为 X 的 k -子集的集合, r 为包含 X 中任意元素的 k -子集数, 若对 X 中任意一对元素 $x_i, x_j (i, j = 1, 2, \dots, v, i \neq j)$, 有 λ 个区组同时包括它们, 则 (X, B) 构成均衡不完全区组设计 (balanced incomplete block design, BIBD), 简记为 $\text{BIBD}(v, b, r, k, \lambda)$ 。

在均衡不完全区组设计中, 当 B 中部分区组构成基集 X 的一个划分时, 称这部分区组为一个平行类, 若 B 可以划分为若干个平行类, 则称该区组设计是可分解的, 或记为 $\text{RBIBD}(v, b, r, k, \lambda)$ 。

2 基于 RBIBD 的最优 BLRCs 构造

本文提出一种基于 RBIBD 的信息位具有 (r, t) 局部性的最优单校验 BLRCs 构造方法。具体地, 运用 RBIBD 若干平行类中的区组与有限集中元素的对应关系构造关联矩阵, 关联矩阵与单位矩阵级联得到 BLRCs 的生成矩阵, 构造出最小距离最优的 BLRCs。具体构造步骤如下。

1) 令 p 为素数, 构造方阵:

$$\Phi^{(0)} = \begin{pmatrix} \varphi_0^{(0)} \\ \varphi_1^{(0)} \\ \vdots \\ \varphi_{p-1}^{(0)} \end{pmatrix} = \begin{pmatrix} \varphi_{0,0} & \varphi_{0,1} & \cdots & \varphi_{0,p-1} \\ \varphi_{1,0} & \varphi_{1,1} & \cdots & \varphi_{1,p-1} \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_{p-1,0} & \varphi_{p-1,1} & \cdots & \varphi_{p-1,p-1} \end{pmatrix}$$

方阵 $\Phi^{(0)}$ 中包含 p^2 个互不相同的元素, 用 $\varphi_{i,j}$ 表示, 其中 $0 \leq i, j \leq p-1$; $\varphi_i^{(0)}$ 为 p 维行向量。

2) 由 $\Phi^{(0)}$ 逐步构造出 $\Phi^{(1)}, \Phi^{(2)}, \dots, \Phi^{(p-1)}, \Phi^{(p)}$, 其中:

$$\Phi^{(k)} = \begin{pmatrix} \varphi_0^{(k)} \\ \varphi_1^{(k)} \\ \vdots \\ \varphi_{p-1}^{(k)} \end{pmatrix} = \begin{pmatrix} \text{SH}_l^0(\varphi_0^{(k-1)}) \\ \text{SH}_l^1(\varphi_1^{(k-1)}) \\ \vdots \\ \text{SH}_l^{p-1}(\varphi_{p-1}^{(k-1)}) \end{pmatrix} \quad 1 \leq k \leq p-1$$

$$\Phi^{(k)} = (\Phi^{(0)})^T \quad k = p$$

式中, l 表示移位的方向, 代表向左移位, a 表示移位的偏移量, $\text{SH}_l^a(\varphi_i^{(k)})$ 表示将向量 $\varphi_i^{(k)}$ 循环左移 a 位, 如 $\text{SH}_l^1(\varphi_1^{(0)}) = \text{SH}_l^1(\varphi_{1,0} \ \varphi_{1,1} \ \cdots \ \varphi_{1,p-1}) = (\varphi_{1,1} \ \varphi_{1,2} \ \cdots \ \varphi_{1,0})$ 。

3) 用 $x_1, x_2, \dots, x_{p^2}$ 分别表示方阵 $\Phi^{(0)}$ 中的 p^2 个元素, 即令 $x_{pi+j+1} = \varphi_{i,j}$, 其中 $0 \leq i, j \leq p-1$ 。设 $X = \{x_1, x_2, \dots, x_{p^2}\}$ 为一有限集, $\Phi^{(0)}, \Phi^{(1)}, \dots, \Phi^{(p)}$ 的各行构成 $p(p+1)$ 个区组, 则可得 $\text{RBIBD}(p^2, p(p+1), p+1, p, 1)$ 设计。取此区组设计的 q 个平行类中的区组, 即取 $\Phi^{(0)}, \Phi^{(1)}, \dots, \Phi^{(p)}$ 中任意 q 个方阵的各行构成的 pq 个区组, 分别用 $B_1, B_2, \dots, B_{pq}$ 表示, 其中 $1 \leq q \leq p$ 。

4) 将 X 中所有元素与矩阵的行关联, 将区组 $B_1, B_2, \dots, B_{pq}$ 与矩阵的列关联, 那么该区组设计可用对应的关联矩阵 $M = (m_{ij})_{p^2 \times (pq)} (1 \leq i \leq p^2, 1 \leq j \leq pq)$ 表示, 其中:

$$m_{ij} = \begin{cases} 1 & \text{若 } x_i \in B_j \\ 0 & \text{若 } x_i \notin B_j \end{cases}$$

则该设计对应的关联矩阵 M 可表示为 $p^2 \times pq$ 阶的二元稀疏 $(r = p, t = q)$ -正则矩阵。

5) 关联矩阵 M 与 p^2 阶单位矩阵级联, 得到 BLRCs 的生成矩阵 $G = [I_{p^2} | M]$ 。由生成矩阵 $G = [I_{p^2} | M]$ 得到信息位具有局部性 r 和可用性 t 的单校验 (n, k, r, t) BLRCs, 其中参数满足条件 $n = p^2 + pq, k = p^2, r = p, t = q$ 。

例 1 假设 $p = 3, q = 2$, 由上述构造方法可得 $\Phi^{(0)}, \Phi^{(1)}, \Phi^{(2)}, \Phi^{(3)}$ 分别为:

$$\Phi^{(0)} = \begin{pmatrix} \varphi_0^{(0)} \\ \varphi_1^{(0)} \\ \varphi_2^{(0)} \end{pmatrix} = \begin{pmatrix} \varphi_{0,0} & \varphi_{0,1} & \varphi_{0,2} \\ \varphi_{1,0} & \varphi_{1,1} & \varphi_{1,2} \\ \varphi_{2,0} & \varphi_{2,1} & \varphi_{2,2} \end{pmatrix}$$

$$\Phi^{(1)} = \begin{pmatrix} \varphi_0^{(1)} \\ \varphi_1^{(1)} \\ \varphi_2^{(1)} \end{pmatrix} = \begin{pmatrix} \varphi_{0,0} & \varphi_{0,1} & \varphi_{0,2} \\ \varphi_{1,1} & \varphi_{1,2} & \varphi_{1,0} \\ \varphi_{2,2} & \varphi_{2,0} & \varphi_{2,1} \end{pmatrix}$$

$$\Phi^{(2)} = \begin{pmatrix} \varphi_0^{(2)} \\ \varphi_1^{(2)} \\ \varphi_2^{(2)} \end{pmatrix} = \begin{pmatrix} \varphi_{0,0} & \varphi_{0,1} & \varphi_{0,2} \\ \varphi_{1,2} & \varphi_{1,0} & \varphi_{1,1} \\ \varphi_{2,1} & \varphi_{2,2} & \varphi_{2,0} \end{pmatrix}$$

$$\Phi^{(3)} = \begin{pmatrix} \varphi_0^{(3)} \\ \varphi_1^{(3)} \\ \varphi_2^{(3)} \end{pmatrix} = \begin{pmatrix} \varphi_{0,0} & \varphi_{1,0} & \varphi_{2,0} \\ \varphi_{0,1} & \varphi_{1,1} & \varphi_{2,1} \\ \varphi_{0,2} & \varphi_{1,2} & \varphi_{2,2} \end{pmatrix}$$

令 $x_{3i+j+1} = \varphi_{i,j}$, 其中 $0 \leq i, j \leq 2$, 即用 $x_1, x_2, \dots, x_9$ 分别表示方阵中的 $p^2 = 9$ 个元素。设 $X = \{x_1, x_2, \dots, x_9\}$ 为一有限集, $B = \{B_1, B_2, \dots, B_{12}\}$ 为 X 的 3-子集的集合, 其中 $B_1, B_2, \dots, B_{12}$ 分别为方阵 $\Phi^{(0)}, \Phi^{(1)}, \Phi^{(2)}, \Phi^{(3)}$ 的各列构成 $p(p+1) = 12$ 个区组, 表示如下:

$$\begin{aligned} B_1 &= \{x_1 \ x_4 \ x_7\}, B_4 = \{x_1 \ x_5 \ x_9\} \\ B_2 &= \{x_2 \ x_5 \ x_8\}, B_5 = \{x_2 \ x_6 \ x_7\} \\ B_3 &= \{x_3 \ x_6 \ x_9\}, B_6 = \{x_3 \ x_4 \ x_8\} \\ B_7 &= \{x_1 \ x_6 \ x_8\}, B_{10} = \{x_1 \ x_2 \ x_3\} \\ B_8 &= \{x_2 \ x_4 \ x_9\}, B_{11} = \{x_4 \ x_5 \ x_6\} \\ B_9 &= \{x_3 \ x_5 \ x_7\}, B_{12} = \{x_7 \ x_8 \ x_9\} \end{aligned}$$

那么 (X, B) 构成了 RBIBD(9, 12, 4, 3, 1) 设计。由方阵 $\Phi^{(0)}, \Phi^{(1)}, \Phi^{(2)}$ 和 $\Phi^{(3)}$ 的各列构成的区组分别构成了此区组设计的 4 个平行类, 即 B_1, B_2, B_3 构成第 1 个平行类, B_4, B_5, B_6 构成第 2 个平行类, 等等。由于此例中 $q = 2$, 所以取此区组设计的 2 个平行类中的区组, 这里取 $B_1, B_2, \dots, B_6$ 这 6 个区组。

将元素 $x_1, x_2, \dots, x_9$ 与矩阵的行关联, 将区组 $B_1, B_2, \dots, B_6$ 与矩阵的列关联, 那么可知关联矩阵 M 是一个 9×6 的稀疏 (3, 2)-正则矩阵, 具体表示如下:

$$M_{9 \times 6} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

根据生成矩阵 $G = [I_9 | M_{9 \times 6}]$ 可以得到单校验 ($n = 15, k = 9, r = 3, t = 2$) BLRCs, 码的最小距离 $d = 3$ 。若信息位 c_1 发生故障, 则由 BLRCs 的生成矩阵 G 可知, 信息位 c_1 可用 $c_1 = c_{10} - c_4 - c_7 = c_{13} - c_5 - c_9$ 来恢复, 那么信息位 c_1 的修复集可表示为 $\varphi_1(1) = \{4, 7, 10\}$, $\varphi_2(1) = \{5, 9, 13\}$, 每个修复集均含有一个校验位符号。当多节点发生故障时, 分为 3 种情况讨论: 1) 假设信息位 c_1 和 c_2 发生故障, 此时, 信息位 c_1 的修复集同上, 信息位 c_2 的修复集为 $\varphi_1(2) = \{5, 8, 11\}$, $\varphi_2(2) = \{6, 7, 14\}$; 2) 假设信息位 c_1 和校验位 c_{10} 发生故障, 此时, 信息位 c_1 只能通过 $c_1 = c_{13} - c_5 - c_9$ 来恢复, 接着进一步恢复校验位 c_{10} ; 3) 假设两个校验位发生故障, 用对应的未故障的信息位恢复校验位。

3 性能分析

3.1 最小距离

推论 1 由生成矩阵 $G = [I_{p^2} | M]$ 构造得到的单校验 $(n, k, r, t)_i$ BLRCs 是最小距离最优的 BLRCs, 且码的最小距离 $d = t + 1$ 。

证明: 由于构造的单校验 $(n, k, r, t)_i$ BLRCs 信息位具有局部性 r 和可用性 t , 根据式 (1), 最小距离应满足 $d \geq t + 1$ 。又因为构造的 BLRCs 的参数满足 $n = p^2 + pq, k = p^2, r = p, t = q$, 将这些参数代入式 (2), 可得:

$$\begin{aligned} d &\leq n - k - \left\lfloor \frac{kt}{r} \right\rfloor + t + 1 = \\ p^2 + pq - p^2 - \left\lfloor \frac{p^2 q}{p} \right\rfloor + q + 1 &= q + 1 \end{aligned}$$

因为 $q = t$, 所以构造出的 BLRCs 的最小距离 $d = t + 1$, 达到了式 (2) 中的最小距离边界, 则该码是最小距离最优的 BLRCs, 证毕。

3.2 码率

本文构造的 BLRCs 的码率为:

$$R = \frac{k}{n} = \frac{p^2}{p^2 + pq} = \frac{r}{r + t}$$

将本文构造的 BLRCs 与文献 [15] 基于阵列 LDPC 码构造的 BLRCs、文献 [11] 基于直积码构造的 BLRCs 以及基于矩阵迭代构造的 BLRCs^[17] 进行比较, 4 种 BLRCs 的构造参数如表 1 所示。

表 1 3 种 BLRCs 的构造参数

构造方法	n	k	R
本文构造的 BLRCs	$r^2 + rt$	r^2	$\frac{r}{r+t}$
基于阵列 LDPC 码构造的 BLRCs	$r^2 + rt + 1$	r^2	$\frac{r^2}{r^2 + rt + 1}$
基于直积码构造的 BLRCs	$(r+1)^t$	r^t	$\left(\frac{r}{r+1}\right)^t$
基于矩阵迭代构造的 BLRCs	$\frac{(r+1)(r+2)}{2}$	$\frac{(r-1)(r+2)}{2}$	$\frac{r-1}{r+1}$

由表 1 可知, 当这些码的局部性都取 (r, t) 时, 有 $\frac{r}{r+t} > \frac{r^2}{r^2 + rt + 1}$, 即本文构造的 BLRCs 的码率高于基于阵列 LDPC 码构造的 BLRCs。当 $t > 1$ 时, 可知 $\left(1 + \frac{1}{r}\right)^t > 1 + \frac{t}{r}$, 因此基于直积码构造的 BLRCs 的码率 $\left(\frac{r}{r+1}\right)^t = \frac{1}{\left(1 + \frac{1}{r}\right)^t} < \frac{1}{1 + \frac{t}{r}} = \frac{r}{r+t}$ 。经分析可

知,当局部性相同且 $t > 1$ 时,本文构造的 BLRCs 的码率总高于基于阵列 LDPC 码构造的 BLRCs 和基于直积码构造的 BLRCs。此外,当可用性 $t = 2$ 时,基于矩阵迭代构造的 BLRCs 的码率 $R = \frac{r-1}{r+1} < \frac{r}{r+2}$,即本文构造的 BLRCs 的码率高于基于矩阵迭代构造的 BLRCs。

图 1 给出了当局部性 $r = 11$ 时,码率 R 随可用性 t 的变化曲线。可以看出,当局部性一定,随着可用性 t 的增加,码率降低。当可用性 t 取相同值时,本文构造的 BLRCs 的码率高于基于阵列 LDPC 码构造的 BLRCs、基于直积码构造的 BLRCs 以及基于矩阵迭代构造的 BLRCs,尽管基于阵列 LDPC 码构造的 BLRCs 比本文构造的 BLRCs 在码率上小得不是很多,但是其参数限制较大,可用性 t 只能取偶数。这 3 种 BLRCs 的码率都没有达到文献 [11] 提出的码率上界。

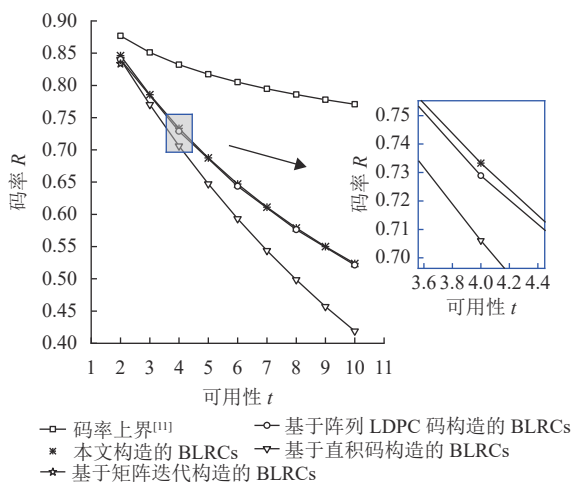


图 1 局部性 $r = 11$ 时,码率 R 随可用性 t 变化曲线

进一步,在可用性 $t = 2$ 时,本文构造的 BLRCs 的码率 $R = \frac{r}{r+2}$,达到了文献 [12] 提出的码率上界,是码率最优的码。图 2 给出了可用性 $t = 2$ 时,码率 R 随局部性 r 的变化曲线。可以看出,当可用性 $t = 2$,随着局部性 r 的增大,3 种码的码率同时呈增大的趋势;当局部性 r 取相同值时,本文构造的 BLRCs 的码率高于基于阵列 LDPC 码构造的 BLRCs、基于直积码构造的 BLRCs 以及基于矩阵迭代构造的 BLRCs。

此外,由不同的构造方法构造出的 BLRCs 的参数可取范围也不尽相同,灵活的参数选择可以为系统设计提供便利。文献 [13] 基于迹函数构造的 BLRCs 需满足 $t = r + 1$,其码率 $R = \frac{2^r - 1}{2^{r+1} - 1} < \frac{1}{2}$,小

于本文构造的 BLRCs 的码率。文献 [14] 基于仿射平面仅能构造局部性 r 与可用性 t 相等的 BLRCs,并且其码率 R 仅为 $\frac{1}{2}$ 。文献 [16] 基于 (r, t) 超图构造具有 (r, t) 局部性的 BLRCs,但是 (r, t) 超图存在的必要条件是超图的顶点数 $v \geq t(r-1) + 1$ 并且 $r|vt$,使得构造出的 BLRCs 具有极大的参数限制。文献 [18] 基于随机函数仅能构造局部性 $r = 2$ 的 BLRCs,构造出的 BLRCs 主要是针对特定参数的系统,且其码率 $R = \frac{1}{k-1}$,当信息位长度 k 越大时,码率 R 越低。本文基于 RBIBD 构造的 BLRCs 不仅参数选择更加灵活,而且在码率上表现得更优。

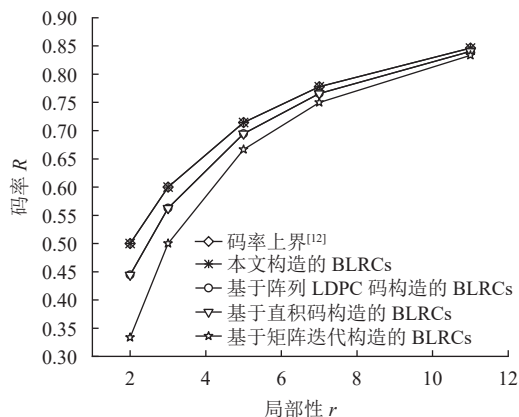


图 2 可用性 $t = 2$ 时,码率 R 随局部性 r 变化曲线

4 结束语

本文提出了一种基于 RBIBD 的最优单校验 BLRCs 构造方法,得到的 BLRCs 信息位具有 (r, t) 局部性。具体地,运用 RBIBD 若干平行类中的区组与有限集中元素的对应关系构造关联矩阵,基于关联矩阵得到生成矩阵来构造 BLRCs。构造的 BLRCs 达到了最优最小距离界,是最小距离最优的 BLRCs。

参 考 文 献

- [1] SIDDIQA A, KARIM A, GANI A. Big data storage technologies: A survey[J]. Frontiers of Information Technology & Electronic Engineering, 2017, 18(8): 1040-1070.
- [2] BSOUL M, ABDALLAH A E, ALMAKADMEH K, et al. A round-based data replication strategy[J]. IEEE Transactions on Parallel & Distributed Systems, 2015, 27(1): 31-39.
- [3] 罗象宏,舒继武. 存储系统中的纠错码研究综述[J]. 计算机研究与发展, 2012, 49(1): 1-11.

LUO X H, SHU J W. Summary of research for erasure code in storage system[J]. Journal of Computer Research and

- Development, 2012, 49(1): 1-11.
- [4] PAPALIOPOULOS D S, DIMAKIS A G. Locally repairable codes[J]. *IEEE Transactions on Information Theory*, 2014, 60(10): 5843-5855.
- [5] GOPALAN P, HUANG C, SIMITCI H, et al. On the locality of codeword symbols[J]. *IEEE Transactions on Information Theory*, 2012, 58(11): 6925-6934.
- [6] LI X D, MA L M, XING C P. Optimal locally repairable codes via elliptic curves[J]. *IEEE Transactions on Information Theory*, 2018, 65(1): 108-117.
- [7] JIN L F. Explicit construction of optimal locally recoverable codes of distance 5 and 6 via binary constant weight codes[J]. *IEEE Transactions on Information Theory*, 2019, 65(8): 4658-4663.
- [8] TAMO I, BARG A, GOPARAJU S, et al. Cyclic LRC codes, binary LRC codes, and upper bounds on the distance of cyclic codes[J]. *International Journal of Information and Coding Theory*, 2016, 3(4): 345-364.
- [9] LUO Y, XING C P, YUAN C. Optimal locally repairable codes of distance 3 and 4 via cyclic codes[J]. *IEEE Transactions on Information Theory*, 2018, 65(2): 1048-1053.
- [10] RAWAT A S, PAPALIOPOULOS D S, DIMAKIS A G, et al. Locality and availability in distributed storage[J]. *IEEE Transactions on Information Theory*, 2016, 62(8): 4481-4493.
- [11] TAMO I, BARG A. Bounds on locally recoverable codes with multiple recovering sets[C]//IEEE International Symposium on Information Theory. [S.l.]: IEEE, 2014: 691-695.
- [12] PRAKASH N, LALITHA V, BALAJI S B, et al. Codes with locality for two erasures[J]. *IEEE Transactions on Information Theory*, 2019, 65(12): 7771-7789.
- [13] WANG A Y, ZHANG Z F, LIN D D. Two classes of (r, t) -locally repairable codes[C]//2016 IEEE International Symposium on Information Theory (ISIT). [S.l.]: IEEE, 2016: 445-449.
- [14] HAO J, XIA S T. Constructions of optimal binary locally repairable codes with multiple repair groups[J]. *IEEE Communications Letters*, 2016, 20(6): 1060-1063.
- [15] HAO J, XIA S T, CHEN B. On the single-parity locally repairable codes with availability[C]//2016 IEEE/CIC International Conference on Communications in China (ICCC). Chengdu: IEEE, 2016: 1-4.
- [16] KIM J H, SONG H Y. Hypergraph-Based binary locally repairable codes with availability[J]. *IEEE Communications Letters*, 2017, 21(11): 2332-2335.
- [17] ZHANG M, LI R H. Two families of LRCs with availability based on iterative matrix[C]//2020 13th International Symposium on Computational Intelligence and Design (ISCID). [S.l.]: IEEE, 2020: 334-337.
- [18] LEE K S, PARK H, NO J S. New binary locally repairable codes with locality 2 and uneven availabilities for hot data[J]. *Entropy*, 2018, 20(9): 636.
- [19] 王新梅, 肖国镇. 纠错码: 原理与方法[M]. 西安: 西安电子科技大学出版社, 2001.
- WANG X M, XIAO G Z. Principle and method for error correcting codes[M]. Xi'an: Xidian University Press, 2001.
- [20] TAN P, ZHOU Z C, SIDORENKO V, et al. Two classes of optimal LRCs with information (r, t) -locality[J]. *Designs, Codes and Cryptography*, 2020, 88(9): 1741-1757.
- [21] ROBERTS F, TESMAN B. Applied combinatorics[M]. [S. l.]: CRC Press, 2009.
- [22] COLBOURN C J, DINITZ J H. The CRC handbook of combinatorial designs[M]. Boca Raton, FL: CRC Press, 2007.

编辑 税红