文章编号:0253-2778(2020)02-0140-06

多节点修复的代数几何码

胡万宝,胡 帅,陈雯雯,崔良武

(安庆师范大学数理学院,安徽安庆 246133)

摘要:最小存储再生码的每个节点具有最小的数据存储,因而是最大距离可分码,这样其节点数的上界为 2^b,其中 b 是存储在每个节点中的数据的比特数.从理论和实践的角度来看,我们很自然地会去考虑这样的再生码:其具有接近最小的数据存储并且节点数不受此界的限制.针对这一问题,Jin 等用代数几何码构造再生码,推广了 Wotters 和 Guruswami 的 Reed-Solomon 修复算法.本文在此基础上进行了扩展,给出了多节点修复的代数几何再生码.这推广和改进了最近一些关于再生码的结果,例如多失效节点的 Reed-Solomon 码和 scalar MDS 码.

关键词:分布式存储系统;再生码;Reed-Solomon 码;代数几何码;带宽

中图分类号:O157.4;TN911.22 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2020.02.009

2010 Mathematics Subject Classification:94B05

引用格式:胡万宝,胡帅,陈雯雯,等. 多节点修复的代数几何码[J]. 中国科学技术大学学报,2020,50(2):140-145.

HU Wanbao, HU Shuai, CHEN Wenwen, et al. Repairing multiple failures for algebraic geometry codes[J]. Journal of University of Science and Technology of China, 2020,50(2):140-145.

Repairing multiple failures for algebraic geometry codes

HU Wanbao, HU Shuai, CHEN Wenwen, CUI Liangwu

(*School of Mathematics and Physics, Anqing Normal University, Anqing* 246133, *China*)

Abstract: Minimum storage regenerating codes have minimum storage of data in each node and therefore are maximal distance separable codes. Thus, the number of nodes is upper-bounded by 2^b, where b is the bits of data stored in each node. From both theoretical and practical points of view, it is natural to consider regenerating codes that nearly have minimum storage of data, and meanwhile, the number of nodes is unbounded. Aiming at the problem, Jin et al. constructed the regenerating codes by algebraic geometry codes, which generalized the repairing algorithm of Reed-Solomon codes by Guruswami and Wotters. This paper mainly gives a construction to repair multiple failures for algebraic geometry codes, which extends the framework of repairing one failure for the regenerating codes. The results generalize some quite recent results in which regenerating codes, for instance, Reed-Solomon codes and scalar codes with multiple erasures.

Key words: distributed storage system; regenerating codes; Reed-Solomon codes; algebraic geometry codes; bandwidth

收稿日期:2018-09-17;**修回日期**:2019-04-10

基金项目:国家自然科学基金(11601009)资助.

作者简介:胡万宝(通讯作者),男,1963年生,博士/教授.研究方向:代数编码. E-mail:huwanb@126.com

0　引言

在分布式存储系统中，一个大文件被编码并分布存储在多个节点. 当有少数节点失效时，我们希望可以利用剩余的存活节点有效地来重构这些失效节点. 在通信过程中，精确修复最具实践意义，精确修复是指能精确地修复失效节点. 在传统方案中我们利用最大距离可分(maximum-distance-separable，MDS)码来解决精确修复问题，例如 Reed-Solomon 码，但是利用 Reed-Solomon 码去解决精确修复问题并不理想[3]. 最小存储再生码[1]改进了这一问题，文献[2]给出了这方面很好的综述. 最近，Wotters 和 Guruswami[3]构造出一种 Reed-Solomon 码的线性精确修复方案，其带宽小于用传统 Reed-Solomon 码修复方案的带宽.

在文献[3]的修复方案中，码长受到字符集的限制，Jin 等[10]利用代数几何码的修复方案解决了这一问题. 但是他们仅研究修复单个失效节点. 事实上在通信过程中时常会出现多个节点失效[4-5,12]. Mardia 等[12]通过对文献[3]的修复方案进行推广，给出了多重修复的 scalar MDS 码的线性修复方案. 但是码长受字符集限制这一问题仍未得到解决. 受文献[12]的启发，我们对文献[10]的单节点的代数几何修复码进行扩展，得到多节点修复的代数几何码. 特别的，文中的代数函数域为有理函数域时，其结果推广了文献[4-5,12]的结果.

本文专注于中心模型(centralized model) 下的多重修复. 在这个模型中，修复中心负责对所有的节点进行修复. 将这个修复中心下载的信息比特数用来计为带宽，而不考虑中心与任何替换节点之间的信息交流. 此外，本文中的修复也是指精确修复.

1　准备工作

本节介绍一些将要用到的记号及定义.

1.1　一些记号

记 $[n]$ 为整数集合 $\{1,2,\cdots,n\}$. 对于向量 ν，$\omega\in\mathbb{F}_p^n$，用 $\langle\nu,\omega\rangle=\sum_{i\in[n]}\nu_i\omega_i$ 表示标准内积.

向量和矩阵记号：把 $\mathbb{F}_p$ 上的向量 v 的第 i 个分量记作 v_i，v^{T} 表示为向量 v 的转置. 对于向量 $\nu\in\mathbb{F}_q^n$ 和集合 $[n]\supseteq I=\{i_1,i_2,\cdots,i_r:i_1<i_2<\cdots<i_n\}$，记向量 $v_I=(\nu_{i_1},\nu_{i_2},\cdots,\nu_{i_r})$. 对于矩阵 M，用 $M[:,i]$（$M[i,:]$）表示矩阵 M 的第 i 列(行).

有限域记号：本文 $\mathbb{F}_q$ 为偶特征的有限域. $\mathbb{F}_q/\mathbb{F}_p$ 是域的扩张且扩张次数 $[\mathbb{F}_q:\mathbb{F}_p]=t$. 令 $\{\zeta_1,\zeta_2,\cdots,\zeta_t\}$ 是 $\mathbb{F}_q$ 在 $\mathbb{F}_p$ 上的一组基. 对于任意的 $\alpha\in\mathbb{F}_q$，记 α 在 $\mathbb{F}_p$ 上的 trace 函数 $\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(\alpha)$ 为

$$\mathrm{Tr}_{\mathbb{F}_q/\mathbb{F}_p}(\alpha)=\alpha+\alpha^q+\cdots+\alpha^{q^{t-1}},$$

简记为 $\mathrm{Tr}(\alpha)$.

对于 $\mathbb{F}_q$ 的子集 $\{\beta_1,\beta_2,\cdots,\beta_v\}$，把由 $\{\beta_1,\beta_2,\cdots,\beta_v\}$ 生成的 $\mathbb{F}_p$ 向量空间记作 $\mathrm{Span}_{\mathbb{F}_p}\{\beta_1,\beta_2,\cdots,\beta_v\}$. 令 V 是 $\mathbb{F}_q$ 的一个 $\mathbb{F}_p$ 线性子空间. 定义一个 p 线性化多项式 $L_V(x):=\prod_{\alpha\in V}(x-\alpha)$. 则 $L_V(x)$ 通过 $\beta\mapsto L_V(x)$ 给出了一个 $\mathbb{F}_q$ 到 $\mathbb{F}_p$ 的 $\mathbb{F}_p$ 映射. 显然，$L_V(x)$ 的核就是 V，因此象 $\mathrm{Im}(L_V)$ 为 $\mathbb{F}_p$ 上的一个维数为 $\log_p q-\dim_{\mathbb{F}_p}(V)$ 的线性空间.

以下两节中的名词及记号均来自于文献[9].

1.2　代数函数域背景

对于域 K，F 为 K 的扩域，若 $x\in F$ 为 K 上的超越元，且 F 是 $K(x)$ 的有限代数扩张，则称 F/K 为 K 上的代数函数域. 简称 F/K 为函数域.

令 F 为 $\mathbb{F}_q$ 上亏格为 g 的代数函数域. 这里考虑 $\mathbb{F}_q$ 为 F 的满常域(full constant field)，即 F_q 在 F 中的代数闭域为 $\mathbb{F}_q$. $\mathcal{O}$ 为函数域 $F/\mathbb{F}_q$ 的赋值环(valuation ring)，位 P(place) 为对应赋值环 $\mathcal{O}_P$ 的唯一极大理想，因此 $\mathcal{O}_P/P$ 为域. 我们通过嵌入映射将 K 看作 $\mathcal{O}_P\backslash P$ 的子域，并记位 P 的次数 $\deg(P)=[\mathcal{O}_P/P:K]$. 记 v_P 为 F 中关于位 P 的离散赋值映射，并规定 F 中次数为 1 的位 P 为有理位(rational place). F 中所有位的集合记为 $\mathbb{P}_F$. F 中的除子(divisor) $G=\sum_{P\in\mathbb{P}_F}m_P P$（仅有有限个非零系数 m_P），并定义：G 的次数

$$\deg(G)=\sum_{P\in\mathbb{P}_F}m_P\deg(P),$$

G 的支撑集 $\mathrm{supp}(G)=\{P\in\mathbb{P}_F:m_P\neq 0\}$. 对任意非零函数 f 我们定义 f 的主除子(principal divisor)为 $(f):=\sum_{P\in\mathbb{P}_F}v_P(f)P$，其次数为零.

记 F 的微分集合为 Ω_F. F 的所有微分构成 F 上一维向量空间. 也就是说如果 $t\in F$ 满足 $\mathrm{d}t\neq 0$，则有 $\Omega_F=F\mathrm{d}t$，$\deg(\mathrm{d}t)=2g-2$. 因此，对于 $f\in F\backslash\{0\}$ 和非零元 $f\mathrm{d}t$ 有 $\deg(f\mathrm{d}t)=\deg(f)+\deg(\mathrm{d}t)=2g-2$. 对于 $f\in F\backslash\{0\}$，我们称除子 $(f\mathrm{d}t)$ 为 F 的标准除子(canonical divisor)，次数为 $2g-2$.

对于一个除子 G，我们可以定义如下两种空间：

$$\mathscr{L}(G)=\{f\in F\backslash\{0\}:(f)+G\geqslant 0\}\cup\{0\},$$

$$\Omega(G)=\{\omega\in\Omega_F\backslash\{0\}:(\Omega)\geqslant G\}\cup\{0\}.$$

$\mathscr{L}(G)$ 和 $\Omega(G)$ 都是 $\mathbb{F}_q$ 上有限维空间，并分别记 $\mathscr{L}(G)$ 和 $\Omega(G)$ 的维数为 $l(G)$ 和 $i(G)$. 令 K 为标准除子，有 $i(G)=l(K-G)$. 则 Rieman-Roch 定理可改写为

$$l(G)=\deg(G)-g+1+i(G)=\deg(G)-g+1+l(K-G).$$

相应的有 $l(G)\geqslant\deg(G)-g+1$. 并且当 $\deg(G)>2g-2$ 时，$l(G)=\deg(G)-g+1$.

1.3 代数几何码

令 $F/\mathbb{F}_q$ 为代数函数域. 假设 F 有 n 个有理位 $\{P_1,P_2,\cdots,P_n\}$，记 $\mathscr{P}=\{P_1,\cdots,P_n\}$. G 为 F 中除子并满足 $\mathrm{supp}(G)\cap P=\varnothing$. 考虑由 Goppa 定义的代数几何码 $C_L(G,\mathscr{P})$：

$$\{(f(P_1),f(P_2),\cdots,f(P_n)):f\in\mathscr{L}(G)\}.$$

则称码 $C_L(G,\mathscr{P})$ 为函数代数几何码.

下面定义另一种码 $C_\Omega(G,\mathscr{P})$：

$$\{(\mathrm{res}_{P_1}(\omega),\mathrm{res}_{P_1}(\omega),\cdots,\mathrm{res}_{P_n}(\omega)):\omega\in\Omega(G-\sum_{i=1}^{n}P_i)\},$$

其中，$\mathrm{res}_P(\omega)$ 表示 ω 在 P_i 处的留数. 我们称码 $C_\Omega(G,\mathscr{P})$ 为微分代数几何码，并有 $C_\Omega(G,\mathscr{P})$ 是 $C_L(G,\mathscr{P})$ 的欧几里德对偶码[7]. 相应的有以下结论：

命题 1.1 $C_L(G,\mathscr{P})$ 的对偶码为 $C_\Omega(G,\mathscr{P})$. 进一步的有

①码 $C_L(G,\mathscr{P})$ 的维数

$$k=l(G)-l(G-\sum_{i=1}^{n}P_i),$$

且码 $C_\Omega(G,\mathscr{P})$ 的维数 $k^\perp=i(G-\sum_{i=1}^{n}P_i)-i(G)$.

②如果 $\deg(G)<n$，则 $k=l(G)\geqslant\deg(G)-g+1$.

③如果 $\deg(G)>2g-2$，则

$$k^\perp=i(G-\sum_{i=1}^{n}P_i)\geqslant n-\deg(G)+g-1.$$

④如果 $2g-2<\deg(G)<n$. 则 $k=\deg(G)-g+1$ 且 $k^\perp=n-\deg(G)+g-1$.

1.4 代数几何码的对偶码

由节 1.3 我们知道码 $C_L(G,\mathscr{P})$ 的对偶码为 $C_\Omega(G,\mathscr{P})$. 本小节将找出码 $C_\Omega(G,\mathscr{P})$ 中码字在某些特殊位置有零或非零的分量.

命题 1.2 令 G 是亏格为 g 的函数域 $F/\mathbb{F}_q$ 的一个除子，满足 $\deg(G)<d$ 且 $\mathrm{supp}(G)\cap\mathscr{P}=\varnothing$. 对任意 $I\subseteq[n]:|I|=r$ 和 $S\subseteq[n]\backslash I:|S|=d$. 则存在码字 $(\mathrm{res}_{P_1}(\omega),\cdots,\mathrm{res}_{P_n}(\omega))\in C_\Omega(G,\mathscr{P})$，对于所有的 $i\in I$ 有 $\mathrm{res}_{P_i}\neq 0$，对于所有的 $j\in[n]\backslash(S\cup I)$ 有 $\mathrm{res}_{P_j}(\omega)=0$.

证明 因为

$$i(G-\sum_{l\in S}P_l)=|S|-\deg(G)+2g-2+1=d-\deg(G)+g-1.$$

且

$$i(G-\sum_{l\in S}P_l-\sum_{i\in I}P_i)=|S|+|I|-\deg(G)+2g-2-g+1=d-\deg(G)+r+g-1.$$

则存在微分

$$\omega\in\Omega(G-\sum_{l\in S}P_l-\sum_{i\in I}P_i)\backslash\Omega(G-\sum_{l\in S}P_l).$$

因而对 $i\in I$ 有 $\mathrm{res}_{P_i}(\omega)\neq 0$，对 $j\in[n]\backslash(S\cup I)$，$\mathrm{res}_{P_j}(\omega)=0$. 易知

$$(\mathrm{res}_{P_1}(\omega),\cdots,\mathrm{res}_{P_n}(\omega))\in C_\Omega(G,\mathscr{P}).$$

1.5 再生码

定义 1.1 一个局部参数为 m、带宽为 B 的 k 维子空间 $C\subseteq\mathbb{F}_q^n$ 是一个修复 r 个节点的 q 元 (n,k,d) 再生码，如果

①若 $c=(c_1,c_2,\cdots,c_n)\in C$，对于每一个子集合 $I\subseteq[n]:|I|\leqslant r$ 和任意的子集合 $J\subseteq[n]\backslash I:|J|=m$，$c_I$ 可以用 c_J 来修复.

②对任意的 $I\subseteq[n]:|I|\leqslant r$ 和任意的 $S\subseteq[n]\backslash I:|S|=d$，可以通过从 c_S 下载至多 B 比特数来决定 c_I.

注 1.1 ①自然的，我们有 $r\leqslant n-d$.

②这里定义的局部参数 $m=|J|$ 与局部修复码(locality repairable code)中定义的略有不同，这里的集合 J 具有任意性.

2 多节点修复的代数几何码

在本节，我们将给出多节点修复的代数几何码的构造. 在节 1.3 考虑除子 $G=(m-1)P_\infty$，相应的代数几何码 $C_L((m-1)P_\infty,\mathscr{P})$（通常称其为一点代数几何码(one point algebraic geometry code)），令 $Z=\{\zeta_1,\zeta_2,\cdots,\zeta_t\}$ 是 $\mathbb{F}_q$ 在 $\mathbb{F}_p$ 上的一组基，V 为

$\mathbb{F}_p$ 上 l 维的 $\mathbb{F}_p$ 子空间，并考虑节 1.2 中定义的 $\mathbb{F}_p$ 线性映射 $L_V(x)=\prod_{a\in V}(x-\alpha)$.

为了将文献[10]中的单节点修复的代数几何码推广至多节点，关键是要找到一个合适的函数 $h_{(a,u)}$. 令 $F/\mathbb{F}_q$ 为 $\mathbb{F}_q$ 上的函数域. 对于 $I\subseteq[n]$: $|I|=r$，$u\in[t]$，$a\in[r]$，定义

$$h_{(a,u)}=\frac{L_V(\zeta_u\cdot\prod_{i\in I}h_i\cdot h^{a-1})}{\prod_{i\in I}h_i} \tag{1}$$

式中，$h_i\in\mathscr{L}((g+1)P_\infty-P_i)$，$i\in I$；$h\in\mathscr{L}((g+1)P_\infty)$，并且对 $i,j\in I, i\neq j$ 满足 $h(P_i)\neq h(P_j)$. 现在作矩阵：

$$M_I=\begin{pmatrix} h_{(1,1)}(P_1) & \cdots & h_{(1,t)}(P_1) & \cdots & h_{(r,1)}(P_1) & \cdots & h_{(r,t)}(P_1)\\ h_{(1,1)}(P_2) & \cdots & h_{(1,t)}(P_2) & \cdots & h_{(r,1)}(P_2) & \cdots & h_{(r,t)}(P_2)\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots\\ h_{(1,1)}(P_r) & \cdots & h_{(1,t)}(P_r) & \cdots & h_{(r,1)}(P_r) & \cdots & h_{(r,t)}(P_r)\end{pmatrix}.$$

对于矩阵 M_I，我们有如下的两个引理.

引理 2.1　令 $F/\mathbb{F}_q$ 是亏格为 g 的代数函数域. F 有 $n+1$ 个有理位 $P_\infty,P_1,\cdots,P_n$. 记集合 $\mathscr{P}=\{P_1,P_2,\cdots,P_n\}$，$h_{(a,u)}$ 和 Z 的定义如上. 则矩阵 M_I 是 $\mathbb{F}_p$ 上的一个满秩矩阵，其意义是：对任意 $0\neq\vec{x}\in\mathbb{F}_p^{rt}$ 有 $M_I\cdot\vec{x}^{\mathrm{T}}\neq\mathbf{0}$，这里 $\mathbf{0}$ 表示零矩阵.

证明　对于 $i\in I$，$h_i\in\mathscr{L}((g+1)P_\infty-P_i)$ 是一个非零函数（这是合理的，因为 $\ell((g+1)P_\infty-P_i)\geqslant g-g+1=1$）. 则非零函数 $h_{(a,u)}\in\mathscr{L}(RP_\infty)$，这里 $R=(g+1)(2r-1)(p^l-1)$. 则有

$$h_{(a,u)}(P_i)=\frac{L_V(\zeta_u\cdot\prod_{j\in I}h_j\cdot h^{a-1})}{\prod_{j\in I}h_j}(P_i)=\zeta_u h^{a-1}(P_i).$$

为了说明矩阵 M_I 是满秩的，需要证明对于任意的非零向量 $\vec{x}\in\mathbb{F}_p^{rt}$，$M_I\cdot\vec{x}^{\mathrm{T}}\neq\mathbf{0}$. 为此，任取 $\mathbf{0}\neq\vec{x}\in\mathbb{F}_p^{rt}$，记 $\vec{x}=(x^{(1)},x^{(2)},\cdots,x^{(r)})$，这里的 $x^{(i)}\in\mathbb{F}_p^t(i=1,2,\cdots,r)$，则

$$\langle M_I[i,:],\vec{x}\rangle=\sum_{a=1}^{r}\langle\vec{\zeta},x^{(a)}\rangle\cdot h^{a-1}(P_i),$$

这里 $\vec{\zeta}=\langle\zeta_1,\zeta_2,\cdots,\zeta_t\rangle\in\mathbb{F}_q^t$. 对于 $i\in I$，令 $g(P_i)=\langle M_I[i,:],\vec{x}\rangle$，则有线性方程组

$$\langle\vec{\zeta},x^{(1)}\rangle+\langle\vec{\zeta},x^{(2)}\rangle h(P_i)+\cdots+\langle\vec{\zeta},x^{(r)}\rangle h^{r-1}(P_i)=g(P_i),i\in I \tag{2}$$

假设对于所有的 $i\in I$，$g(P_i)=0$. 观察方程组的系数行列式

$$\begin{vmatrix}1 & h(P_1) & \cdots & h^{r-1}(P_1)\\ 1 & h(P_2) & \cdots & h^{r-1}(P_2)\\ \vdots & \vdots & \ddots & \vdots\\ 1 & h(P_r) & \cdots & h^{r-1}(P_r)\end{vmatrix},$$

熟知其是一个 Vandermonde 行列式. 由于对 $i,j\in I$，$i\neq j$ 有 $h(P_i)\neq h(P_j)$，则该行列式不为零. 因此我们知道线性方程组(2) 只有零解，即对所有的 $a\in[r]$，有 $\langle\vec{\zeta},x^{(a)}\rangle=0$. 由于向量 $\vec{x}\neq\mathbf{0}$，存在 $a\in[r]$ 其对应的子向量 $x^{(a)}\neq\mathbf{0}$，而 $\{\zeta_i\}_{i\in[t]}$ 是 $\mathbb{F}_q$ 在 $\mathbb{F}_p$ 上的一组基，则 $\langle\vec{\zeta},x^{(a)}\rangle\neq0$. 矛盾！因而存在 $i\in I$ 有 $g(P_i)\neq0$. 得证.

引理 2.2　若 $\mathbb{F}_q^r$ 到 $\mathbb{F}_p^{rt}$ 的映射 φ 规定为 $\varphi(\vec{x})=\mathrm{Tr}(\vec{x}\cdot M_I)$，这里 $\mathrm{Tr}(\vec{x}\cdot M_I)$ 表示为

$$\varphi(\vec{x})=(\mathrm{Tr}(\langle\vec{x},M_I[:,1]\rangle),\cdots,\mathrm{Tr}(\langle\vec{x},M_I[:,rt]\rangle)),$$

则 φ 为可逆映射.

证明　为了说明 φ 为可逆映射，考虑映射 ψ: $\mathbb{F}_p^{rt}\to\mathbb{F}_q^r$，规定为 $\psi(\vec{y})=M_I\cdot\vec{y}$，这里 $\vec{y}=(y_1,y_2,\cdots,y_{rt})^{\mathrm{T}}\in\mathbb{F}_p^{rt}$. 显然 ψ 是一个 $\mathbb{F}_p$ 线性映射. 由引理 2.1 知 M_I 为满秩矩阵，对于 $0\neq\vec{y}\in\mathbb{F}_p^{rt}$，$M_I\cdot\vec{y}\neq0$，因而 ψ 为单射. 再通过对比维数可知 ψ 为满射，所以 ψ 为可逆映射. 又

$$\langle\varphi(\vec{x}),\vec{y}\rangle=\sum_{j\in[rt]}y_j\cdot\mathrm{Tr}(\langle\vec{x},M[:,j]\rangle)=\mathrm{Tr}\Big(\sum_{j\in[rt]}y_j\cdot\langle\vec{x},M[:,j]\rangle\Big)=\mathrm{Tr}(\langle\vec{x},M_I\cdot\vec{y}\rangle)=\mathrm{Tr}(\langle\vec{x},\psi(\vec{y})\rangle).$$

因此 ψ 为 φ 的伴随映射. 下面我们来证明 φ 为可逆映射.

令 $\mathbf{0}\neq\vec{x}\in\mathbb{F}_q^r$，则存在 $\vec{z}\in\mathbb{F}_q^r$ 有 $\mathrm{Tr}(\langle\vec{x},\vec{z}\rangle)\neq0$. 又 ψ 为满射，存在某个 $\vec{y}\in\mathbb{F}_p^{rt}$，有 $\varphi(\vec{y})=\vec{z}$. 又 $\langle\varphi(\vec{x}),\vec{y}\rangle=\mathrm{Tr}(\langle\vec{x},\psi(\vec{y})\rangle)=\mathrm{Tr}(\langle\vec{x},\vec{z}\rangle)\neq0$.

因此有 $\varphi(\vec{x})\neq 0$. 即知 φ 为单射. 同样通过维数相等可知 φ 为满射,则 φ 为可逆映射.

由引理 2.1 和引理 2.2,我们得到下面的再生码.

定理 2.1 令 $F/\mathbb{F}_q$ 是亏格为 g 的代数函数域. F 有 $n+1$ 个有理位 $P_\infty,P_1,\cdots,P_n$. 记集合 $\mathscr{P}=\{P_1,P_2,\cdots,P_n\}$. 如果

$$2g\leqslant m\leqslant d-(g+1)(2r-1)(p^l-1),$$

则代数几何码 $C_L((m-1)P_\infty)$ 是一个局部参数为 m 、带宽 $B=rg\log q+(d-rg)(\log q-l\log p)$ 的修复 r 个节点的 q 元 $(n,m-g,d)$ 再生码.

证明 首先,确定码的维数. 由 $m-1\geqslant 2g-1$,从命题 1.1 得到 $k=l((m-1)P_\infty)=m-g$.

对于码字 $c=(f(P_1),f(P_2),\cdots,f(P_n))\in C_L((m-1)P_\infty,\mathscr{P})$,这里 $f\in\mathscr{L}((m-1)P_\infty)$. 对于失效节点集 $I\subseteq[n]$,下面要证明如何通过其他存活节点去决定 $c_I=\{f(P_i):i\in I\}$. 令 $S\subseteq[n]\backslash I$,$|S|=d$. 同样记 $R=(g+1)(2r-1)(p^l-1)$. 由命题 1.2 知,存在码字 $(\mathrm{res}_{P_1}(\omega),\mathrm{res}_{P_2}(\omega),\cdots,\mathrm{res}_{P_n}(\omega))\in C_\Omega((R+m-1)P_\infty,\mathscr{P})$,对于所有的 $i\in I,j\in[n]\backslash(S\cup I)$ 有

$$\mathrm{res}_{P_i}(\omega)\neq 0,\ \mathrm{res}_{P_j}(\omega)=0.$$

易知 $(h_{(a,u)}(P_1)f(P_1),h_{(a,u)}(P_2)f(P_2),\cdots,h_{(a,u)}(P_n)f(P_n))\in C_L((R+m-1)P_\infty,\mathscr{P})$. 由于 $C_L((R+m-1)P_\infty,\mathscr{P})$ 是 $C_\Omega((R+m-1)P_\infty,\mathscr{P})$ 的对偶码,因而有

$$0=\sum_{i=1}^{n}\mathrm{res}_{P_i}(\omega)h_{(a,u)}(P_i)f(P_i)=\sum_{j\in S\cup I}\mathrm{res}_{P_j}(\omega)h_{(a,u)}(P_j)f(P_j),$$

即

$$\sum_{i\in I}\mathrm{res}_{P_i}(\omega)h_{(a,u)}(P_i)f(P_i)=-\sum_{j\in S}\mathrm{res}_{P_j}(\omega)h_{(a,u)}(P_j)f(P_j).$$

因而对于 $a\in[r],u\in[t]$,有下列等式:

$$\sum_{i\in I}\mathrm{Tr}(\mathrm{res}_{P_i}(\omega)h_{(a,u)}(P_i)f(P_i))=-\sum_{j\in S}\mathrm{Tr}(\mathrm{res}_{P_j}(\omega)h_{(a,u)}(P_j)f(P_j))=-\sum_{j\in S}\mathrm{Tr}(\mathrm{res}_{P_j}(\omega)f(P_j)\cdot M_I[j,(a,u)]).$$

为了决定 $c_I=\{f(P_i):i\in I\}$,对于 $i\in[n]$,令 $e_i=\mathrm{res}_{P_i}(\omega)f(P_i)$. 考虑映射 $\varphi:\mathbb{F}_q^r\to\mathbb{F}_p^{rt}$ 定义为 $\varphi(\vec{x})=\mathrm{Tr}(\vec{x}\cdot M_I)$. 由引理 2.1 和引理 2.2 知矩阵 M_I 是 $\mathbb{F}_p$ 上的一个满秩矩阵,φ 为可逆映射. 观察下式中的 rt 个值

$$\sum_{j\in S}\mathrm{Tr}(e_j\cdot M_I[j,(a,u)]),a\in[r],u\in[t],$$

即给出了 $\varphi(e_I)$. 又 φ 为可逆映射,于是决定了 $e_I=\{e_i;i\in I\}$,又 $i\in I$,$\mathrm{res}_{P_i}(\omega)\neq 0$,因而得之 $c_I=\{f(P_i);i\in I\}$.

最后,计算 $C_L((m-1)P_\infty)$ 的带宽.

对于 $j\in S$,定义

$$b_j^{(I)}=\dim_{\mathbb{F}_p}\mathrm{Span}\{h_{(a,u)}(P_j):a=1,2,\cdots,r,u=1,2,\cdots,t\}.$$

令 $J\subseteq[r]\times[t]$,$|J|=b_j^{(I)}$,规定集合 $\{h_{(b,v)}(P_j):(b,v)\in J\}$ 是 $\mathrm{Span}\{h_{(a,u)}(P_j):a=1,2,\cdots,r,u=1,2,\cdots,t\}$ 在 $\mathbb{F}_p$ 上的张成基. 因此从存储 $f(P_j)$ 的节点只需下载如下的数据:

$$\mathrm{Tr}(\mathrm{res}_{P_j}(\omega)h_{(b,v)}f(P_j)),(b,v)\in J \quad (3)$$

这意味着整个修复过程一共只需下载 $\sum_{j\in S}b_j^{(I)}\log p$ 的比特数. 事实上,对于 $a\in[r],u\in[t]$,$h_{(a,u)}(P_j)$ 是 $\{h_{(b,v)}(P_j):(b,v)\in J\}$ 的一个 $\mathbb{F}_p$ 线性组合,即存在 $\mathbb{F}_p$ 的元素的集合 $\{\lambda_{(b,v)}\}_{(b,v)\in J}$ 使得 $h_{(a,u)}(P_j)=\sum_{(b,v)\in J}\lambda_{(b,v)}h_{(b,v)}(P_j)$. 于是有

$$\mathrm{Tr}(\mathrm{res}_{P_j}(\omega)h_{(a,u)}(P_j)f(P_j))=\sum_{(b,v)\in J}\lambda_{(b,v)}\mathrm{Tr}(\mathrm{res}_{P_j}(\omega)h_{(b,v)}(P_j)f(P_j)).$$

因而要计算 $\mathrm{Tr}(\mathrm{res}_{P_j}(\omega)h_{(a,u)}(P_j)f(P_j))$ $(a\in[r],u\in[t])$,只需要下载式(3)中的 $|J|=b_j^{(I)}$ 个数据. 因此带宽

$$B=\max_{I\subseteq[n],S\subseteq[n]\backslash I,|I|=r,|S|=d}\sum_{j\in S}b_j^{(I)}\log p \quad (4)$$

对于 $C_L((m-1)P_\infty)$,我们进一步可以给出显示的带宽 B.

令 $F_I=\prod_{i\in I}h_i$,$J_I=\{j\in S:F_I(P_j)=0\}$. 则 $F_I\in\mathscr{L}(r(g+1)P_\infty-\sum_{i\in I}P_i-\sum_{j\in J_I}P_j)$,因此有 $\ell(F_I)\geqslant -r(g+1)+r+|J_I|$,即 $|J_I|\leqslant rg$. 下面求带宽 B.

如果 $j\notin S\backslash J_I$,有

$$\mathrm{Span}_{\mathbb{F}_p}\{h_{(a,u)}(P_j):a\in[r],u\in[t]\}=\mathrm{Span}_{\mathbb{F}_p}\{\frac{L_V(\zeta_u\cdot\prod_{i\in I}h_i\cdot h^{a-1}(P_j))}{\prod_{i\in I}h_i(P_j)}:a\in[r],u\in[t]\}\subseteq\frac{1}{\prod_{i\in I}h_i(P_j)}L_V(\mathbb{F}_q).$$

因而对于 $j \notin S\backslash J_I$，

$$b_j^{(I)} \leqslant \dim_{\mathbf{F}_p}(L_V(\mathbb{F}_q)) = \log_p q - l.$$

如果 $j \in J_I$，则有平凡的界 $b_j^{(I)} \leqslant \log_p q$．最后由(4) 可以知道带宽的上界为

$$\sum_{j\in S} b_j^{(I)} \log\ p = \sum_{j\in J_I} b_j^{(I)} \log\ p + \sum_{j\in S\backslash J_I} b_j^{(I)} \log p \leqslant$$

$$rg\log\ q + (d - rg)(\log\ q - l\log\ p).$$

得证.

3　有理函数域上的再生码

如果 $F = \mathbb{F}_q(x)$ 是 $\mathbb{F}_q$ 上的有理函数域，则相应的代数几何码就是一个 Reed-Solomon 码(这里码长 $n \leqslant q$). 令 P_∞ 是 x 唯一极点.

定理 3.1　如果 $m \leqslant d - (2r-1)(p^l - 1)$，则 Reed-Solomon 码 $C_L((m-1)P_\infty, \mathscr{P})$ 是一个局部参数为 m 、带宽 $B = d\log\ q - dl\log\ p$ 的修复 r 个节点的 q 元 (n, m, d) 再生码.

证明　注意到，有理函数域的亏格 $g = 0$，因此定理 2.1 证明过程中的 $J_I = \varnothing$．对于 $I \subseteq [n]$：$|I| = r$ 和任意集合 $S \subseteq [n]\backslash I$：$|S| = d$，$m \leqslant d - (2r-1)p^l + r$ 表明函数

$$h_{(a,u)} \in \mathscr{L}((2r-1)p^l - r)P_\infty).$$

重复定理 2.1 的证明过程，可得定理.

注 3.1　将文献[4-5]的结果进行对比. 选取 $n = q, d = n - r$，当 $r = 2$ 时定理 3.1 改进了文献[5]中的结果，当 $r = 3$ 时定理 3.1 推广了文献[4]的结果.

最后，为了对比文献[12]中的结果，我们在有理函数域 $\mathbb{F}_q(x)$ 中调整函数 $h_{(a,u)}$．取 $h = x \in \mathscr{L}((g+1)P_\infty)$，这是合理的，因为 $(x) + P_\infty = (x)_0 - (x)_\infty + P_\infty = (x)_0 \geqslant 0$. 同样对于 $i \in I$，$\alpha_i \in \mathbb{F}_q$，令 $h_i = x - \alpha_i$．令 $F_I(x) = \prod_{i\in I}(x - \alpha_i)$，则 $h_{(a,u)} = \dfrac{L_V(\zeta_u \cdot F_I(x) \cdot x^{a-1})}{F_I(x)}$ 与文献[12，定理 2]的函数 $P_{(\zeta,p)}(X)$ 一致. 因此有如下结果：

推论 3.1　如果

$$m \leqslant n - r - (2r-1)(p^l - 1).$$

则 Reed-Solomon 码 $C_L((m-1)P_\infty, \mathscr{P})$ 是一个局部参数为 m 、带宽

$$B \leqslant \left((n-r)\left(t - \lfloor \log_p\left(\frac{n-m+r-1}{2r-1}\right)\rfloor\right)\right)\log\ p$$

的修复 r 个节点的 q 元 $(n, m, n-r)$ 再生码，这里 $\log_p\left(\dfrac{n-m+r-1}{2r-1}\right)$ 为 l 的上界.

4　结论

我们在单节点修复代数几何码[10]的基础上，给出了多节点修复的修复方案，并推广和改进了文献[4-5，12]中一些关于 Reed-Solomon 码以及 MDS 码的结果. 本文用有理函数域作为例子，得到一些好的结果. 由于代数函数域的丰富性，若使用非有理函数域作为工具，如 Hermite 函数域，或许会得到更多的好结果，这留待将来继续研究.

参考文献(References)

[1] DIMAKIS A G, GODFREY P B, WU Y, et al. Network coding for distributed storage systems[J]. IEEE Trans Inf Theory, 2010, 56(9): 4539-4551.

[2] DIMAKIS A G, RAMCHANDRAN K, WU Y, et al. A survey on network codes for distributed storage[J]. Proc IEEE, 2011, 99(3): 476-489.

[3] GURUSWAMI V, WOOTTERS M. Repairing Reed-Solomon codes[J]. IEEE Trans Inf Theory, 2016, 63(9): 5684-5698.

[4] DAU H, DUURSMA I, KIAH H M, et al. Repairing Reed-Solomon codes with multiple erasures[DB/OL]. [2018-09-01]. https://arxiv.org/abs/1612.01361.

[5] DAU H, DUURSMA I, KIAH H M, et al. Repairing Reed-Solomon codes with two erasures[DB/OL]. [2018-09-01]. https://arxiv.org/abs/1701.07118.

[6] CADAMBE V R, HUANG C, LI J. Permutation code: Optimal exact-repair of a single failed node in MDS code based distributed storage systems[C]// 2011 IEEE International Symposium on Information Theory Proceedings. IEEE, 2011: 1225-1229.

[7] GOPPAV D. Codes on algebraic curves[J]. Dokl Akad Nauk SSSR, 1981, 24(1): 170-172. (in Russian)

[8] STICHTENOTH H. Algebraic Function Fields and Codes[M]. Berlin: Springer, 1993.

[9] LI J, TANG X, TIAN C. Enabling all-node-repair in minimum storage regenerating codes[DB/OL]. [2018-09-01]. https://arxiv.org/abs/1604.07671.

[10] JIN L, LUO Y, XING C. Repairing algebraic geometry codes[J]. IEEE Trans Inf Theory, 2018, 64(2): 900-908.

[11] LIDL R, NIEDERREITER H. Finite Fields[M]. Cambridge: Cambridge University Press, 2003.

[12] MARDIA J, BARTANY B, WOOTTERS M. Repairing multiple failures for scalar MDS codes[DB/OL]. [2018-09-01]. https://arxiv.org/abs/1707.02241.