小样本条件下基于矩阵乘法和秩分析的LDPC参数估计方法

刘 倩[1],张 昊[1],宋莹炯[2],王 刚[1]
(1. 信息工程大学信息系统工程学院,河南郑州 450001;2. 信息工程大学密码工程学院,河南郑州 450001)

摘 要: 在非合作通信背景下,利用传统的盲识别算法获取有用信息往往需要大量的截获数据.本文利用少量的截获数据,基于码字空间与其对偶空间的正交性、完整码字比特间的线性相关性和矩阵乘积秩的性质,提出了矩乘秩减算法,在无误码和低误码率情形下恢复了LDPC (Low-Density Parity-Check)长码的码长和起点.仿真实验表明,与传统算法相比,达到同样的识别效果本文算法能够节省至少20%的数据量,且运算量没有明显增加.

关键词: 盲识别; 编码参数; LDPC码; 高斯列消元; 矩阵的秩; 方阵的乘积
中图分类号: TN911.6 **文献标识码:** A **文章编号:** 0372-2112(2022)05-1075-08
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20210485

LDPC Parameter Estimation by Matrices Product and Rank Analysis Under the Condition of Small Sampling

LIU Qian[1],ZHANG Hao[1],SONG Ying-jiong[2],WANG Gang[1]
(1. *School of Information System Engineering, Information Engineering University, Zhengzhou, Henan* 450001, *China*;
2. *School of Cryptographic Engineering, Information Engineering University, Zhengzhou, Henan* 450001, *China*)

Abstract: In the scenario of non-cooperative communications, usually it takes a large amount of intercepted data for blind identification to obtain useful information with the traditional methods. This paper presents an approach called rank reduction with matrices production to estimate block length and synchronization of long LDPC(Low-Density Parity-Check) codes under the condition of small sampling with noise-free or lower bit-error rate data. Our method is based on the orthogonality of codeword space and its dual space, the linear correlations among the bits in a whole codeword, and the property of rank reduction of matrices. Experimental results show that, our method can save 25% data at least to reach the same identification probability compared with the traditional methods, and the computation has no obvious increasement.

Key words: blind identification; encoder parameters; low-density parity-check codes(LDPC); Gaussian column elimination; rank of matrix; product of square matrices

1 引言

信道编码类型众多,其中的LDPC(Low-Density Parity-Check)码[1,2]由于其逼近香农限的传输性能和强纠错能力,以及适用于硬件并行运算的置信度传播解码算法被广泛应用于各种通信标准中. LDPC码不同于其他线性码,(如Hamming码、循环码、卷积码和Turbo码等),这类编码由其稀疏校验矩阵来确定,一般具有较长的码长. 由于计算复杂度或空间复杂度的原因,原本适用于短码的编码参数识别算法[3]很难再对LDPC码起作用. 因此,对LDPC码的识别分析方法自成一派.

目前对LDPC码的参数识别算法主要集中于闭集识别[4~8],即信号接收方手里已有若干种LDPC编码方式构成的集合,且发送方选取的编码方式为集合中的一种,只需要利用接收码字序列将编码方式挑出即可. 开集识别时,由于信号截获方没有任何先验知识,识别难度更大导致研究成果较少[9~16]. 已有工作中有基于对偶码的方法,通过寻找具有一定的汉明重量(或小汉明重量)的对偶码获得校验向量,与此同时恢复码长、码

收稿日期: 2021-04-15; 修回日期:2021-06-29; 责任编辑: 覃怀银
基金项目: 国家自然科学基金(No.61802430, No.62072057); 博士后科学基金(No.2016M603035)

率和起点[9]. 也有假设编码长度、码率及起点均已知，在此基础上恢复稀疏校验矩阵[10~13]. 或者假设截获信号序列长度足够长，在此基础上利用秩分析法（分析构造码字矩阵的秩率[14]或近似秩率[15,16]）估计码长和起点，这两种方法均是基于高斯列消元方法. 近两年还出现了利用秩的概率分布获取正确的编码长度和起点的算法[17,18]. 算法中需要多次从编码矩阵中随机获取方阵，计算它们的秩并得到其经验分布规律. 如果其与已有的同阶随机方阵的秩的分布规律相同，则被认为是随机方阵，此时码长和起点不正确. 否则便被认为具有某种代数结构，此时码长和起点正确. 关于已有工作[14~18]的详细介绍和对其原理及缺点的具体分析将在节2.4节中给出.

以上方法均要求截获比特流的长度是足够供信息截获者进行分析的，但是，在信息对抗过程中，信息对抗方是被动接收信息，并不能保证截获所得数据量充足，此时已有的方法全部失效. 本文在截获比特数量远少于现有方法所需的比特数量的情况下（不含或含少量错误比特），基于码字空间与其对偶空间的正交性、完整码字比特间的线性相关性、矩阵乘积的秩不大于每个因子矩阵的秩等性质，提出了一种估计LDPC码的码长、起点等参数的矩乘秩减算法，并对算法的计算复杂度和正确识别概率进行了分析.

2 数学模型和理论基础

2.1 符号说明

接下来在二元域F_2上展开问题的研究. 用C表示所采用的纠错码，n,k分别表示纠错码的码长和维数. 信息向量和码字向量等矢量分别用粗体$\boldsymbol{u}$和$\boldsymbol{c}$表示，变量用小写斜体字母表示. 大写斜粗体字母$\boldsymbol{A}$和$\boldsymbol{A}^{\mathrm{T}}$分别表示码字矩阵及其转置，矩阵$\boldsymbol{A}$的零空间记为$\mathrm{Kel}(\boldsymbol{A})$，$\boldsymbol{A}(i,:)$和$\boldsymbol{A}(:,i)$分别表示矩阵$\boldsymbol{A}$的第$i$行和第$i$列. $\boldsymbol{W}$和$\boldsymbol{W}^{\perp}$分别代表纠错码码字空间和其对偶空间. d表示正确的起点位置，l_0和t_0分别表示利用我们提出的算法得出的码字长度和起点位置，l和t是在一定区间内变化的变量，将截获比特流在不同的起点t下按列数l排列成的码字矩阵记为$\boldsymbol{D}_{l,t}$.

2.2 码字空间的封闭性及与其对偶空间的正交性

一般的线性分组码的编码方式是由生成矩阵定义：给定生成矩阵$\boldsymbol{G}_{k\times n}$，输入由$k$个比特组成的信息块$\boldsymbol{u}=(u_1,u_2,\cdots,u_k)$，得到一个完整的码字$\boldsymbol{c}=(c_1,c_2,\cdots,c_n)=\boldsymbol{u}\cdot\boldsymbol{G}_{k\times n}$. 已知$\boldsymbol{G}_{k\times n}$的行向量线性无关，分别记之为$\boldsymbol{g}_1,\boldsymbol{g}_2,\cdots,\boldsymbol{g}_k$，则

$$\begin{aligned}\boldsymbol{c}=(c_1,c_2,\cdots,c_n)&=(u_1,u_2,\cdots,u_k)\cdot\begin{pmatrix}\boldsymbol{g}_1\\ \boldsymbol{g}_2\\ \vdots\\ \boldsymbol{g}_k\end{pmatrix}\\&=u_1\boldsymbol{g}_1+u_2\boldsymbol{g}_2+\cdots+u_k\boldsymbol{g}_k\end{aligned}\tag{1}$$

可知所有码字$\boldsymbol{c}$均表示为生成矩阵行向量的线性组合，因此生成矩阵行向量张成一个k维线性空间，称为码字空间，记为$\boldsymbol{W}$. 当$\boldsymbol{G}_{k\times n}$为系统形式$(\boldsymbol{I}_{k\times k},\boldsymbol{P}_{k\times(n-k)})$时，则生成系统码.

若有两个码字$\boldsymbol{c}_i=(c_{i1},c_{i2},\cdots,c_{in})=\boldsymbol{u}_i\cdot\boldsymbol{G}_{k\times n}$和$\boldsymbol{c}_j=(c_{j1},c_{j2},\cdots,c_{jn})=\boldsymbol{u}_j\cdot\boldsymbol{G}_{k\times n}$，则

$$\boldsymbol{c}_i\oplus\boldsymbol{c}_j=(c_{i1}\oplus c_{j1},c_{i2}\oplus c_{j2},\cdots,c_{in}\oplus c_{jn})=(\boldsymbol{u}_i\oplus\boldsymbol{u}_j)\cdot\boldsymbol{G}_{k\times n}$$

仍然是一个合法码字，因此码字空间关于F_2上的加法运算$\oplus$封闭.

容易知道，在二元域上有$(\boldsymbol{I}_{k\times k},\boldsymbol{P}_{k\times(n-k)})\cdot(\boldsymbol{P}^{\mathrm{T}}_{k\times(n-k)},\boldsymbol{I}_{(n-k)\times(n-k)})^{\mathrm{T}}=0$成立，因此，$\boldsymbol{H}=(\boldsymbol{P}^{\mathrm{T}}_{k\times(n-k)},\boldsymbol{I}_{(n-k)\times(n-k)})$的行向量与生成矩阵$\boldsymbol{G}_{k\times n}=(\boldsymbol{I}_{k\times k},\boldsymbol{P}_{k\times(n-k)})$的行向量正交. 可知校验矩阵$\boldsymbol{H}_{(n-k)\times n}$的行向量线性无关且张成了一个$n-k$维的线性空间，并且与码字空间$\boldsymbol{W}$正交，称之为$\boldsymbol{W}$的对偶空间，记为$\boldsymbol{W}^{\perp}$.

LDPC码是一种特殊的线性分组码，仅由一个包含很少个1的校验矩阵$\boldsymbol{H}_{(n-k)\times n}$的零空间所定义. 因此当给定$k$个信息比特$(u_1,u_2,\cdots,u_k)$时，需要利用校验矩阵求解$n-k$个校验比特$(u_{k+1},u_{k+2},\cdots,u_n)$，即系统码字$\boldsymbol{c}=(c_1,c_2,\cdots,c_n)=(u_1,u_2,\cdots,u_k,u_{k+1},u_{k+2},\cdots,u_n)$满足$\boldsymbol{c}\cdot\boldsymbol{H}^{\mathrm{T}}=0$，因此LDPC码的码字空间不仅关于$\oplus$封闭，并且与其对偶空间具有正交性. 而LDPC码的识别过程也正是基于这种关系.

2.3 高斯若当列消元法

高斯若当列消元法GJETP（Gauss-Jordan Elimination Through Pivoting），通过行置换和列变换将一个矩阵变换为下三角矩阵，文献[14~16]均是在此基础上给出各自的算法. 先给出F_2上GJETP法实施的具体过程.

对于一个$s\times l$阶的矩阵$\boldsymbol{A}$，i的变化范围是从1到$\min\{s,l\}$.

（1）如果$\boldsymbol{A}$的第i列的第i个元素为0，并且存在最小的$j(j>i)$使得第j列的第i个元素为1，则交换$\boldsymbol{A}$的第i列和第j列；

（2）如果$\boldsymbol{A}$的第i列的第i个元素为0，并且不存在第$j(j>i)$列其第i个元素为1，但有最小的$j(j>i)$使得$\boldsymbol{A}$的第j行中第i个元素为1，则交换$\boldsymbol{A}$的第i行和第j行；

（3）如果$\boldsymbol{A}$的第i列的第i个元素为1，则将其加到

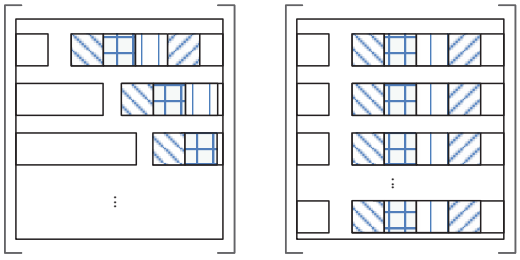
第i行中所有列数大于i的位置上为1的列上.

2.4 已有工作和数学模型

在一个完整码字中,校验比特是某些信息比特的线性组合. 如果接收比特足够多,且由无误码比特序列构建的矩阵$\boldsymbol{A}$的每一行恰好为一个完整的码字,则其列与列之间的比特服从相同的线性约束关系,因此至多只有k列线性无关,$\boldsymbol{A}$的秩至多为k. 在通过GJETP将$\boldsymbol{A}$变换为有$n-k$列全零列的矩阵的过程中,设列变换矩阵为$\boldsymbol{Q}$. 若编码是系统码,则存在$\boldsymbol{Q}=\left(\boldsymbol{Q}_1\ \boldsymbol{Q}_2\right)$使得$\boldsymbol{AQ}_1=\boldsymbol{V}_{M\times k}$,$\boldsymbol{AQ}_2=\boldsymbol{0}_{M\times(n-k)}$,记作

$$\boldsymbol{AQ}=\boldsymbol{A}\left(\boldsymbol{Q}_1\middle|\boldsymbol{Q}_2\right)=\left(\boldsymbol{V}_{M\times k}\middle|\boldsymbol{0}_{M\times(n-k)}\right)$$

可知,$\boldsymbol{Q}_2$的列向量与全部码字正交,因此,有$\boldsymbol{Q}_2^{\mathrm{T}}\subset\boldsymbol{W}^{\perp}$,而$\mathrm{rank}\left(\boldsymbol{Q}_2^{\mathrm{T}}\right)=n-k$,可知,由$\boldsymbol{Q}_2^{\mathrm{T}}$的行向量张成的线性空间$R\left(\boldsymbol{Q}_2^{\mathrm{T}}\right)=\boldsymbol{W}^{\perp}$. 将$\boldsymbol{Q}_2^{\mathrm{T}}$通过初等行变换化成系统形式,则得到系统校验矩阵$\boldsymbol{H}_{(n-k)\times n}$. 但如果由无误码码字序列构建的$s\times l$阶矩阵$\boldsymbol{A}$的列数$l\neq\beta n$($\beta$为正整数),则上下对应关系被打乱. 列与列的比特之间服从的线性约束关系不再相同,此时$\boldsymbol{A}$可被视为随机矩阵,其秩远大于k. 两种情形的对比关系可见图1(a)和1(b).



(a) $l\neq n$　　(b) $l=n$

图1　当$l\neq\beta n$和$l=\beta n$时,矩阵$\boldsymbol{A}$在形式上的差别

为了更好的做出对比,将矩阵的秩除以l得到归一化秩率[14,15],便可利用这两种情况下归一化秩率的不同表现来区分预设的编码长度和起点是否正确. 为了进一步提高算法的容错能力,降低错误比特对码字矩阵秩的影响,文献[16]改为寻找矩阵的"近似秩". 其做法是利用GJETP将码字矩阵化成下三角阶梯型矩阵后,寻找矩阵中列重大于某一阈值的列,其数目便为矩阵的近似秩. 同样地,近似秩归一化后得到近似秩的归一化秩率,将使得其取得最小值的矩阵的列数和起点视为正确的参数. 这里把文献[14,15]中的方法叫做秩准则,把文献[16]中的方法叫做近似秩准则,这两种方法都为基于GJETP的秩分析法.

但在利用上述秩分析法获取码字长度和起点等参数的过程中,一般要求矩阵$\boldsymbol{A}$的行数s大于其列数l. 如果截获码字个数较少使得构建的码字矩阵的行数远远小于列数,则秩分析法失效. 以秩准则为例,设码字为(n,k)线性分组码,截获的码字序列为Z,其中含有M个比特. 假定码字长度和起点分别为l和t, 构造码字矩阵$\boldsymbol{D}_{l,t}$,其中$\boldsymbol{D}_{l,t}$的第i个行向量为

$$\boldsymbol{D}_{l,t}(i,:)=Z\left(t+(i-1)\cdot l+1:t+i\cdot l\right),1\leq i\leq\lfloor M/l\rfloor \quad (2)$$

在$\lfloor M/l\rfloor\leq k$的情况下分析问题. 此时,如果$l\neq\beta n$,相应的矩阵$\boldsymbol{D}_{l,t}$为随机矩阵,由于其行数远小于列数,因此有$\mathrm{rank}(\boldsymbol{D}_{l,t})\leq k$. 而当$l=\beta n$时,矩阵$\boldsymbol{D}_{l,t}$是结构矩阵,其行向量是至多$k$个线性无关的行向量(生成矩阵$\boldsymbol{G}$的行向量)的线性组合,因此也有$\mathrm{rank}(\boldsymbol{D}_{l,t})\leq k$. 此时无法由秩准则区分假定的码字长度和起点是否正确.

近似秩准则[16]在估计编码参数问题上表现突出. 但是近似秩准则有一个要求,那就是截获比特数量必须非常大,使得构造的码字矩阵为"高瘦型". 矩阵越高,近似秩才越准确,从而得出正确参数的概率也越大,并可以利用近似秩进一步确定校验关系. 而当码字矩阵为方阵或者"矮胖型"矩阵时,则无法计算近似秩. 因此近似秩准则在截获比特数量较少时完全失效(高瘦型矩阵指的是其行数远远大于列数,矮胖型矩阵指的是其行数远远小于其列数).

在利用随机方阵秩的分布获得编码参数的做法[17,18]中,需要在构造的码字矩阵$\boldsymbol{D}_{l,t}$中随机选取行向量构造出许多个方阵. 计算方阵的秩得出其经验分布函数,然后将其与同阶随机方阵的秩的分布规律进行比较. 如果分布规律相同,则判定预设码字长度和起点不正确,如果不同则认为码字长度和起点正确. 但在构造码字矩阵为"矮胖型"的情况下,其行向量不可能构造出方阵,因此算法[17,18]也失效. 另一方面,算法需要做大量实验统计不同设定码长和起点下方阵的秩的分布规律,因此其计算量也是一个天文数字,在码长较大时并不实用.

3 本文算法、原理及复杂度分析

在给出本文具体算法之前,先以定理的形式给出一个乘积矩阵的秩的性质[19].

定理1　设矩阵$\boldsymbol{A}$和$\boldsymbol{B}$都是l阶方阵,则有$\mathrm{rank}(\boldsymbol{AB})\leq\min\left\{\mathrm{rank}(\boldsymbol{A}),\mathrm{rank}(\boldsymbol{B})\right\}$成立. 容易将此结论推广到$n$个方阵相乘的情形. 即对$n$个方阵$\boldsymbol{A}_1,\boldsymbol{A}_2,\cdots,\boldsymbol{A}_n$,有

$$\mathrm{rank}(\boldsymbol{A}_1\cdots\boldsymbol{A}_n)\leq\min\left\{\mathrm{rank}(\boldsymbol{A}_1),\cdots,\mathrm{rank}(\boldsymbol{A}_n)\right\} \quad (3)$$

3.1 本文算法

为了解决在截获比特数量较少的情况下编码参数的识别问题,本文改变研究矩阵的秩的方式. 注意到不管码字个数有多么少,一个完整码字中的比特间的线性约束关系总是存在的. 因此,我们不去关心全部的校验关系,而将注意力放在寻找满足同一个校验关系的比特上. 在

$l=\beta n$的情形下，设$\boldsymbol{h}=\left(h_1,h_2,\cdots,h_n\right)\in\boldsymbol{W}^{\perp}$，令集合

$$\tau(\boldsymbol{h})=\left\{j\,\middle|\,h_j\neq 0,\boldsymbol{h}=\left(h_1,h_2,\cdots,h_n\right)\in\boldsymbol{W}^{\perp}\right\}.$$

从矩阵$\boldsymbol{D}_{l,t}$中随机选取$\lfloor M/l\rfloor$列构建一个$\lfloor M/l\rfloor$阶的方阵$\boldsymbol{S}_l$，设选取列的列标集合为θ. 若有$\tau(\boldsymbol{h})\subset\theta$，则$\boldsymbol{S}_l$不满秩. 由于LDPC码的校验向量的稀疏性，可知满足同一个校验的比特数目较少，只要这几个比特所在列被选入$\boldsymbol{S}_l$，$\boldsymbol{S}_l$就不满秩. 若是几个校验关系中的比特所在列均被选入$\boldsymbol{S}_l$，则$\boldsymbol{S}_l$秩亏的更多，因此使得$\boldsymbol{S}_l$秩亏的条件较容易达到. 当从$\boldsymbol{D}_l$中随机选取p个方阵，分别记为$\boldsymbol{S}_{l1},\boldsymbol{S}_{l2},\cdots,\boldsymbol{S}_{lp}$，作乘积得$\boldsymbol{S}_{l1}\boldsymbol{S}_{l2}\cdots\boldsymbol{S}_{lp}$，由式(3)可得

$$\operatorname{rank}\left(\boldsymbol{S}_{l1}\cdots\boldsymbol{S}_{lp}\right)\leq\min\left\{\operatorname{rank}\left(\boldsymbol{S}_{l1}\right),\cdots,\operatorname{rank}\left(\boldsymbol{S}_{lp}\right)\right\}\tag{4}$$

这使得乘积矩阵$\boldsymbol{S}_{l1}\boldsymbol{S}_{l2}\cdots\boldsymbol{S}_{lp}$的秩较小的概率大大提升. 若$l\neq\beta n$，则$\boldsymbol{D}_{l,t}$为随机矩阵，其列之间并不存在线性结构. 随机从中选取$\lfloor M/l\rfloor$列组成的方阵$\boldsymbol{S}_l$仍然是随机矩阵，因此乘积矩阵$\boldsymbol{S}_{l1}\boldsymbol{S}_{l2}\cdots\boldsymbol{S}_{lp}$也是随机矩阵，其秩在一般情况下不会较小. 在此分析基础上，为了更好地区分两种情形，定义归一化秩率函数

$$f_l=\frac{\operatorname{rank}\left(\boldsymbol{S}_{l1}\boldsymbol{S}_{l2}\cdots\boldsymbol{S}_{lp}\right)}{\lfloor M/l\rfloor}\tag{5}$$

在具体实施过程中，先估计l_0，再将t_0确定，提出了下面的矩乘秩减算法，如算法1所示.

算法1 矩乘秩减算法

输入：截获比特流序列Z，比特流的长度M，预先设定码长l的变化范围$\left[l_{\min},l_{\max}\right]$，随机取方阵的个数$p$，rank_ratio=0；

输出：码字长度$l_0=\underset{l\in\left[l_{\min},l_{\max}\right]}{\arg\min}\sum_{t=0}^{5}\frac{\operatorname{rank}\left(\boldsymbol{S}_{l1}\boldsymbol{S}_{l2}\cdots\boldsymbol{S}_{lp}\right)}{\lfloor M/l\rfloor}$，起点$t_0$.

for $l=l_{\min}:l_{\max}$
  for $t=0:5$
    for $i=1:\left\lfloor(M-t)/l\right\rfloor$
      $\boldsymbol{D}_{l,t}(i,:)=Z\left((i-1)\cdot l+t+1:i\cdot l+t\right)$;
    end
    for $j=1:p$
      从$\boldsymbol{D}_{l,t}$中随机选取$\lfloor M/l\rfloor$列构造矩阵$\boldsymbol{S}_{lj}$;
    end
    rank_ratio =rank_ratio +f_l;
  end
将rank_ratio的值记录在表格Δ_1中；
end
选取Δ_1中的最小值，其对应的l即被认为是正确的码长，记为l_0；
for $t=0:l_0-1$
    for $i=1:\left\lfloor(M-t)/l_0\right\rfloor$
      $\boldsymbol{D}_{l_0,t}(i,:)=Z\left((i-1)\cdot l_0+t+1:i\cdot l_0+t\right)$;
    end
对$\boldsymbol{D}_{l_0,t}$实施与寻找l_0的过程同样的操作步骤，将所得归一化秩率的数值记录在表格Δ_2中；
end
选取Δ_2中的最小值，其对应的t即被认为是正确的起点，记为t_0.

注1 关于随机选取方阵的个数p的选择问题. 由式(4)可知，随着p的增大，乘积矩阵的秩是单调不增的. 只要出现一个乘因子矩阵的秩比较小，那么乘积矩阵的秩就会很小. 但若$\boldsymbol{S}_{l1},\boldsymbol{S}_{l2},\cdots,\boldsymbol{S}_{lp}$都是随机矩阵，它们的乘积也是随机矩阵，出现秩较小的概率较低. 因此，随着p的增大，在预设码长正确和不正确两种情况下，乘积矩阵的秩的差距进一步拉大，码长和起点的正确识别概率也会提升(在理论分析部分详细介绍). 但如果p的取值过大，则会使得计算矩阵乘法的次数较多，计算量显著增加. 所以，需要在识别概率和计算复杂度之间做一个权衡，选择合适的p值.

注2 关于起点所在位置对算法的识别概率的影响问题. 在估计码字长度的过程中，若预先设定的码字长度l_0正确，t_0未知，则码字矩阵$\boldsymbol{D}_{l_0}$的每一行由上一个码字的l_0-t_0个比特和下一个码字的t_0个比特构成. 当t_0的值靠近0或者l_0时，由于大部分比特在同一个码字中，随机选取的方阵中出现具有线性相关关系的列的可能性较大，此时更容易得到秩较小的矩阵，识别概率较高. 当t_0的值靠近$l_0/2$时，每一个行向量中约一半的比特分别位于在上一个码字和下一个码字中，而两个码字中的比特之间一般不具有线性相关关系，因此随机选取的方阵秩较大的概率较高. 此时，与非正确起点下乘积矩阵的秩区分度不大，识别概率较低.

3.2 算法理论分析

设LDPC码稀疏校验矩阵为$\boldsymbol{H}$，其行向量记为$\boldsymbol{h}_1,\boldsymbol{h}_2,\cdots,\boldsymbol{h}_r$，记$w(\boldsymbol{h})=\min\limits_{1\leq i\leq r}\left\{w\left(\boldsymbol{h}_i\right)\right\}$，$w(\cdot)$代表向量的Hamming重量. 我们想要在下面两个假设检验中做判决

$$\mathrm{H}_0:l_0=n\tag{6}$$

$$\mathrm{H}_1:l_0\neq n\tag{7}$$

下面来考虑正确识别概率，这需要二元域上随机方阵的秩的分布规律. 由于二元域上随机方阵秩的具体分布规律目前还未知，已有的仅仅是当方阵阶数趋于无穷大时随机方阵秩的分布规律[20]. 因此，利用多次蒙特-卡罗实验将有限阶随机方阵的秩的经验分布函数得出. 在10000次实验中，为了映照后面实验部分选择的编码类型，图2(a)中给出了280阶随机方阵的秩的经验分布函数. 图2(b)为阶数从200变化到280的随机方阵的秩取不同值时的频率. 从图中可知不管方阵阶数如何变化，l阶随机方阵的秩极大概率都落在区间$[l-$

$2,l]$上,其中以$l-1$比例最大.

考察正确检测概率

$$P_{cd}=\Pr\{l_0=n\}=\Pr\left\{\frac{\operatorname{rank}(\prod_{i=1}^{p}\boldsymbol{S}_{l_0 i})}{\lfloor M/l_0\rfloor}<\min_{l\neq l_0}\frac{\operatorname{rank}(\prod_{i=1}^{p}\boldsymbol{S}_{li})}{\lfloor M/l\rfloor}\right\} \tag{8}$$

假定随机选取的方阵之间的相关性不大,则可认为其是相互独立的,那么正确检测概率可表示为

$$P_{cd}\approx\prod_{l\neq l_0}\Pr\left\{\frac{\operatorname{rank}(\prod_{i=1}^{p}\boldsymbol{S}_{l_0 i})}{\lfloor M/l_0\rfloor}<\frac{\operatorname{rank}(\prod_{i=1}^{p}\boldsymbol{S}_{li})}{\lfloor M/l\rfloor}\right\} \tag{9}$$

由于只要有某两个$\tau(\boldsymbol{h}_i)\subset\theta(i=1,2,\cdots,r)$,便有

$$\operatorname{rank}\left(\prod_{i=1}^{p}\boldsymbol{S}_{l_0 i}\right)\leq\lfloor M/l_0\rfloor-2$$

若再有多个,则$\operatorname{rank}\left(\prod_{i=1}^{p}\boldsymbol{S}_{l_0 i}\right)$与$\lfloor M/l_0\rfloor$之间的差距进一步拉大. 结合近似计算,可得$P_{cd}$的一个下界为

$$P_{cd}\geq\left[\left(\begin{matrix}l_0-2\cdot w(\boldsymbol{h})\\ \lfloor M/l_0\rfloor-2\cdot w(\boldsymbol{h})\end{matrix}\right)\Big/\left(\begin{matrix}l_0\\ 2\lfloor M/l_0\rfloor\end{matrix}\right)\right]^{l_{\max}-l_{\min}} \tag{10}$$

式(10)中$w(\boldsymbol{h})$为稀疏校验向量中重量的最大值. 可知随着$w(\boldsymbol{h})$变小,P_{cd}变大,随着M变小,P_{cd}变小. LDPC码的校验矩阵为稀疏的且没有4环,校验矩阵的行重一般来说远远小于码长,因此提出的算法特别适用于LDPC码的参数识别.

3.3　算法计算复杂度分析

当l在$[l_{\min},l_{\max}]$中变化时,对于给定的l,计算p个$\lfloor M/l\rfloor$阶方阵的乘积需要的乘法数量为

$$(p-1)(\lfloor M/l\rfloor)^3$$

加法数量为

$$(p-1)(\lfloor M/l\rfloor-1)(\lfloor M/l\rfloor)^2$$

因此,要估计出编码长度l_0,一共需要的乘法数量为

$$6\sum_{l=l_{\min}}^{l_{\max}}(p-1)(\lfloor M/l\rfloor)^3$$

加法数量为

$$6\sum_{l=l_{\min}}^{l_{\max}}(p-1)(\lfloor M/l\rfloor-1)(\lfloor M/l\rfloor)^2$$

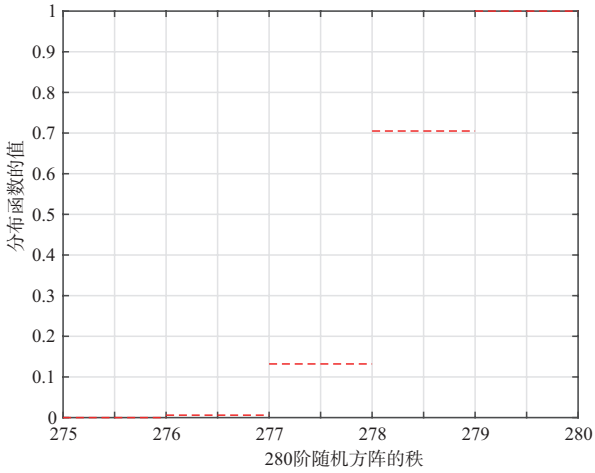
比较运算的数量为$(l_{\max}-l_{\min})$次. 固定编码长度,让起点t在$[1,l_0]$中变化. 要得出起点t_0,类似的操作下至多需要的乘法数量为

$$6l_0(p-1)(\lfloor M/l_0\rfloor)^3$$

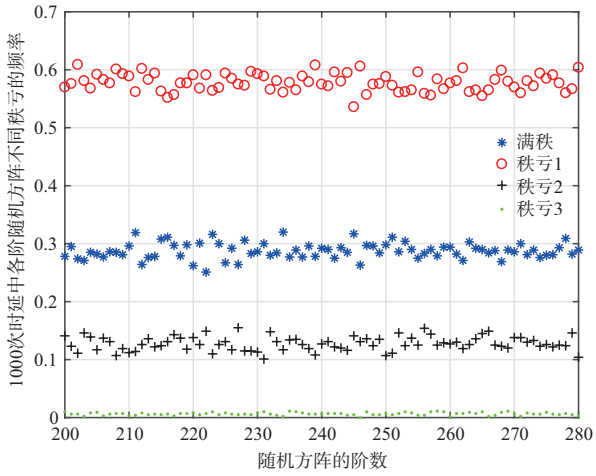
加法数量为

$$6l_0(p-1)(\lfloor M/l_0\rfloor-1)(\lfloor M/l_0\rfloor)^2$$

比较运算的数量为l_0-1次.



(a) 280阶随机方阵的秩的经验分布函数



(b) 各阶随机方阵不同秩亏的频率

图2　随机方阵的秩的分布规律

4　模拟仿真实验

4.1　实验设置

在模拟仿真实验中,均以IEEE802.16e标准中的(576,288)LDPC码为例. 在截获比特流长度M固定的情况下,在截获数据含少量误码和无误码两种情形下,我们分别考察了选取乘积矩阵的个数p对识别成功概率的影响. 令$M=132480$,分别在误比特率P_e为0和0.0001时令乘积矩阵的个数从1连续变化到10.

由于有这样的结论:若长向量线性相关,则从中截

取的短向量一定线性相关. 而短向量线性相关时,其延长得到的长向量未必线性相关,因此,M对识别概率也是有影响的. 一般情况下,由上面结论可知当比特流长度越长时,码字长度和起点的正确识别概率也会越高. 因此在p固定时,我们考察在有误码和无误码情况下M的变化对识别成功率的影响. 在比特流长度$M=q\cdot n$时,令q在180和320之间变化,将本文方法与文献[14~17]在无误码和P_e=0.0001两种情况下进行对比.

为了考察本文算法在不同的误比特率下的表现,我们在截获比特数量$M=132480$,$p=10$的情况下,令误比特率P_e从0.0001按间隔为10^{-4}变化到0.0008,考察P_e的变化对识别成功率的影响. 还研究了在码字长度确定的情况下,起点t_0的位置对归一化秩率的影响,印证了我们在第3节中的分析.

4.2 实验结果分析

从图3中可以看出,码字长度和起点的正确识别概率确实随着p的增大而增大. 原因是列向量的选取具有随机性,如果满足几个校验关系的列都被选中,则方阵的秩较小. 但若都未被选中,则方阵的秩可能比随机矩阵的秩大,也可能比随机矩阵的秩小. 因此,当p较小时不确定性较大,随着p的增大这种不确定性的影响渐渐被减弱. 在图3中还可以看出当$p=1$时,有误码情况下的识别概率稍大于无误码情况下的识别概率,正是这种原因的直观体现.

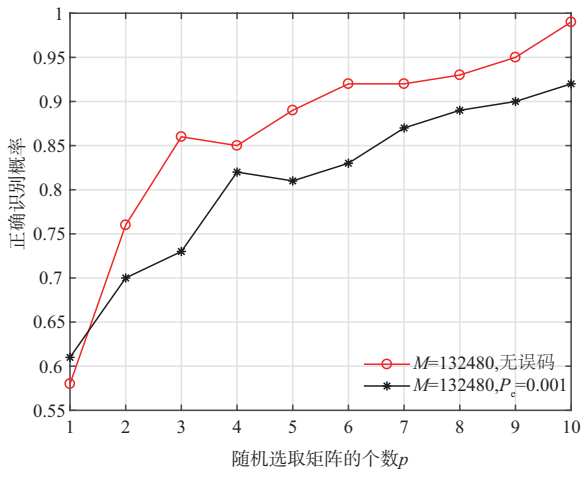


图3 选择作乘积的矩阵个数p对识别概率的影响

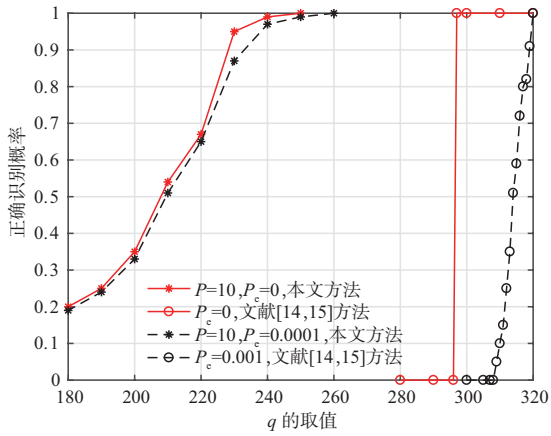
由于识别概率直接受到乘积矩阵的秩的影响,为了更加深入考察p的取值对乘积矩阵的秩的影响,在$l=n$和$l\neq n$两种情况下按照p取值从1到10的顺序分别做了十次实验. 将实验所得的秩求平均值,称其为平均秩. 表1给出了两种情况下平均秩在不同的p值下的取值. 可以看出,不管预设码长正确不正确,平均秩都随着p的增大而减小,并且码长正确时的平均秩整体上要比码长不正确时的平均秩小,这与我们提出算法的思想相符.

表1 十次实验中,p对乘积矩阵平均秩的影响

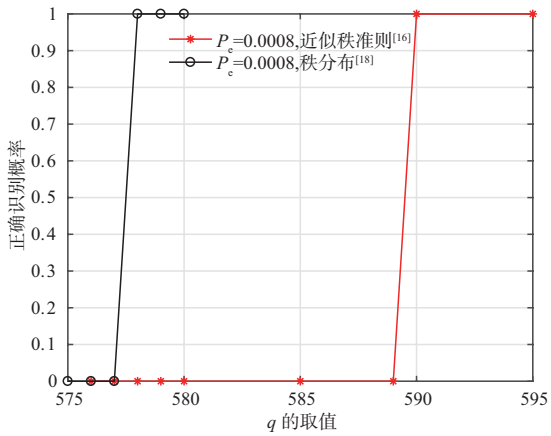
p	$l=n$时的平均秩	$l\neq n$时的平均秩
1	226.3	227.2
2	225.9	226.2
3	225.5	226
4	225.5	225.9
5	225.2	225.7
6	224.3	224.7
7	223.5	224.6
8	223.1	223.7
9	223.3	223.8
10	222.8	223.4

从图4(a)中可以看出,在无误码且$q\leq 300$时,秩准则方法[14,15]失效,而本文算法在$p=10$且$q=250$时正确识别率达到100%. 在误比特率P_e=0.0001且$q\leq 309$时,秩准则失效,而本文算法在$p=10$且$q=260$时正确识别率达到100%. 4(b)中展示了利用近似秩准则[16]以及秩分布方法[18]在P_e=0.0008时的结果,4(b)可作为4(a)的延展. 可以看出虽然它们的容错能力强于本文方法,但它们需要码字最少个数时的q也要远大于本文方法. 图5中显示,在这个变化过程中码长和起点的正确识别概率从89%下降到20%. 虽然本文算法容错能力仅在10^{-4}量级,但是在截获比特数量较少而其他现有算法均失效的情况下,这样的性能也是难能可贵的. 图6给出了在P_e=0.0001且预先设定码长l变化时,从相应的分析矩阵中随机选取$p=10$个方阵,其乘积矩阵的归一化秩率的值. 为了更清楚的显示归一化秩率随l变化的取值情况,我们截取了正确码长$l_0=576$所在的一段区间[558,594]. 可以看出当$l=576$时归一化秩率的值最小.

在确定码字长度之后,分析矩阵中的列都已经固定,差别仅仅是列的位置的不同. 此时,正确起点与不正确起点所对应的平均归一化秩率之间的差距进一步缩小,正确起点t_0的寻找也不是一件容易的事. 为了进一步增加确定性,确保正确起点被选出,可以让起点在$\left[t+1,t+l_0+1\right]$中变化. 如果某两个位置相差l_0,且对应的平均归一化秩率相较于其他值均较小,则可认定此位置为正确起点. 表2给出了(576,288)LDPC码在第1位即为正确起始位,无误码且$M=132480$,$p=10$的情况下,t变化时的平均归一化秩率的部分取值. 可以看到,标黑的第1位和第577位的值相对其他值较小,同时也



(a) 本文方法和秩准则的比较



(b) 近似秩准则和秩分布方法

图4　截获比特流长度$M=q\cdot n$的变化对识别概率的影响

可以看出当假定起始位置接近$l_0/2$时的平均归一化秩率明显要大. 这也印证了在第3节中关于t_0位置对平均归一化秩率的影响的分析.

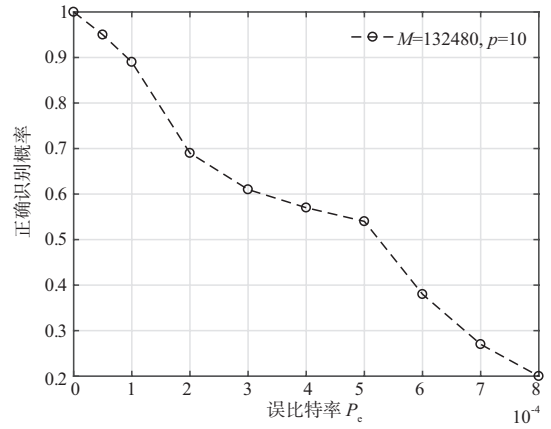


图5　误比特率较低时,误比特率P_e对识别概率的影响

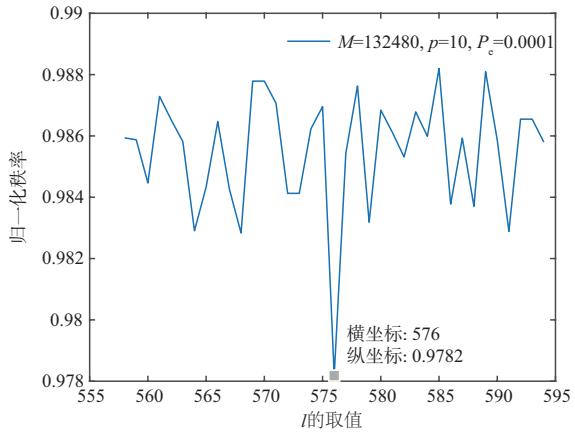


图6　乘积矩阵的归一化秩率随着l的变化的取值情况

表2　l_0已确定,t变化时,归一化秩率的取值情况

1	51	101	151	201	251	301
0.9788	0.9823	0.9838	0.9846	0.9853	0.9846	0.9838
351	401	451	501	551	**577**	578
0.9867	0.9860	0.9823	0.9831	0.9823	**0.9816**	0.9830

5　总结

本文在截获少量比特的背景下,此时其他算法由于所需比特数量众多而全部失效的情况下,对LDPC码的起点和码长进行了识别. 利用完整码字中比特间具有线性相关性和矩阵乘积的秩不大于因子矩阵的秩等性质,从构造的码字矩阵中随机选取列,得到一系列方阵并作乘积,再选出使得乘积矩阵的归一化秩率最小时的分析矩阵列数作为码长. 而后,利用类似的方法对列数固定、起点发生变化时构造的分析矩阵进行处理,并且改变分析矩阵的行数多次实验. 将使得平均归一化秩率最小时的位置确定为码字起点. 我们称之为矩乘秩减算法. 与传统算法相比,达到相同的正确识别率时本文算法所需数据量至少减少25%,但计算量没有明显增加.

参考文献

[1] GALLAGER R G. Low Density Parity Check Codes[M]. Cambridge, MA: MIT Press, 1963.

[2] MACKAY D. Good error-correcting codes based on very sparse matrices[J]. IEEE Transactions on Information Theory, 1999, 45(2): 399-431.

[3] 张永光, 楼才义. 信道编码及其识别分析[M]. 北京: 电子工业出版社, 2010: 1-159.

[4] MOOSAVI R, LARSSON E G. Fast blind recognition of channel codes[J]. IEEE Transactions on Communications, 2014, 62(5): 1393-1405.

[5] XIA T, WU H C. Novel blind identification of LDPC

codes using average LLR of syndrome a posterior probability[J]. IEEE Trans Signal Process, 2014, 62(3): 632-640.

[6] YU P, PENG H. On blind recognition of channel codes within a candidate set[J]. IEEE Communications Letters, 2016, 20(4): 736-739.

[7] WU Z, ZHANG L, ZHENG Z, et al. Blind recognition of LDPC codes over candidate set[J]. IEEE Communications Letters, 2020, 24(1): 11-14.

[8] LIU Q, ZHANG H, YU P, et al. An improved method for identification of LDPC codes within a candidate set[J]. IEEE Access, 2021, (9): 1896-1903.

[9] CLUZEAU M, FINIASZ M. Recovering a code's length and synchronization from a noisy intercepted bitstream[C]// ISIT 2009. Seoul, Korea (South): ISIT, 2009.

[10] 包昕,周磊硐,何可,等. LDPC码稀疏校验矩阵的重建方法[J]. 电子科技大学学报, 2016, 45(2): 192-196.
BAO Xin, ZHOU Lei-ke, HE Ke, et al. A method of restructuring LDPC parity-check matrix[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(2): 192-196. (in Chinese)

[11] 包昕, 周磊硐, 何可, 等. 误码条件下的LDPC码盲识别算法[J]. 西安交通大学学报, 2015, 49(12): 54-58.
BAO Xin, ZHOU Lei-ke, HE Ke, et al. A recognition algorithm for LDPC codes of blind in a noisy environment [J]. Journal of Xi'an Jiaotong University, 2015, 49(12): 54-58. (in Chinese)

[12] 陈泽亮, 彭华, 巩克现, 等. 误码条件下LDPC码参数的盲估计[J]. 电子学报, 2018, 46(3): 462-468.
CHEN Ze-liang, PENG Hua, GONG Ke-xian, et al. A method for blind recognition of LDPC codes in a noisy environment[J]. Acta Electronic Sinica, 2018, 46(3): 462-468. (in Chinese)

[13] WANG W, PENG H, LI J. Blind identification of LDPC codes based on decoding[C]//2017 International Conference on Computer Technology, Electronics and Communication. Dalian, China: ICCTEC, 2017: 998-1001.

[14] BUREL G, GAUTIER R. Blind estimation of encoder and interleaver characteristics in a non-cooperative context[C]//International Conference on Communications, Internet and Information Technology. Scottsdale, AZ, USA: CIIT, 2003.

[15] SWAMINATHAN R, MADHUKUMARA S. Classification of error correcting codes and estimation of interleaver parameters in a noisy transmission environment[J]. IEEE Transactions on Broadcasting, 2017, 63(3): 463-478.

[16] SICOT G, HOUCKE S, BARBIER J. Blind detection of interleaver parameters[J]. Signal Processing, 2009, (89): 450-462.

[17] CHOI C, YOON D. Enhanced blind interleaver parameters estimation algorithm for noisy environment[J]. IEEE Access, 2018, (6): 5910-5915.

[18] CHOI C, YOON D. Novel blind interleaver parameter estimation in a non-cooperative context[J]. IEEE Trans Aerosp Electro Sys, 2019, 55(4): 2079-2085.

[19] GOLUB G, LOAN C V. Matrix Computations[M]. Baltimore, MD, USA: The Johns Hopkins University Press, 1989.

[20] FERREIRA P J S G, JESUS B, VIEIRAJ, et al. The rank of random binary matrices and distributed storage applications[J]. IEEE Comm Lett, 2013, 17(1): 151-154.

作者简介

刘 倩 女,1984年10月出生于河北省任丘市. 在读博士,信息工程大学副教授. 主要研究方向为信道编码的识别分析.
E-mail: liuqian2006815@126.com

张 昊 男, 1984年3月出生于河南省郑州市. 博士. 现为信息工程大学讲师. 主要研究方向为傅里叶分析、信息隐藏和信道编码.
E-mail: haozhang78@126.com

宋莹炯 男, 1996年5月出生于浙江省绍兴市. 现为信息工程大学密码工程学院本科生.
E-mail: 2768730013@qq.com

王 刚 男, 1980年出生于安徽省滁州市. 博士. 现为信息工程大学讲师. 主要研究方向为信源编码.
E-mail: phzttyw@126.com