

有限平面 LDPC 码的停止集

夏树涛 胡懋智

(清华大学深圳研究生院网络工程研究中心 深圳 518055)

摘 要: 有限平面 LDPC 码是一类重要的有结构的 LDPC 码, 在利用和积算法(SPA)等迭代译码方法进行译码时表现出卓越的纠错性能。众所周知, 次优的迭代译码不是最大似然译码, 因而如何对迭代译码的性能进行理论分析一直是 LDPC 码的核心问题之一。近几年来, Tanner 图上的停止集(stopping set)和停止距离(stopping distance)由于其在迭代译码性能分析中的重要作用而引起人们的重视。该文通过分析有限平面 LDPC 码的停止集和停止距离, 从理论上证明了有限平面 LDPC 码的最小停止集一定是最小重量码字的支撑, 从而对有限平面 LDPC 码在迭代译码下的良好性能给出了理论解释。

关键词: 低密度校验(LDPC)码; 有限几何; 迭代译码; 停止集; 停止距离

中图分类号: TN911.22

文献标识码: A

文章编号: 1009-5896(2007)06-1365-04

On the Stopping Sets of Finite Plane LDPC Codes

Xia Shu-tao Hu Mao-zhi

(Graduate School at Shenzhen of Tsinghua University, Shenzhen 518055, China)

Abstract: Finite plane LDPC codes are important structured LDPC codes, which have excellent performance under iterative decoding algorithm. It is a key problem that to evaluate the performance of LDPC codes under iterative decoding. Recently, the stopping sets and stopping distance of Tanner graph are of interests in performance evaluation. In this paper, the smallest sets of finite plane LDPC codes are studied. It shows that for finite plane LDPC codes, a smallest stopping set is the support of a codeword. These results give positive consequences for the good performance of finite plane LDPC codes under iterative decoding.

Key words: Low-Density Parity-Check (LDPC) codes; Finite geometry; Iterative decoding; Stopping set; Stopping distance

1 引言

低密度校验(LDPC)码是一类线性(分组)码, LDPC码由于在许多信道中具有接近香农限的纠错性能并有可实现的译码而受到极大的关注。LDPC码是通过低密度校验矩阵定义的一类线性码。设 C 为二元 $[n, k, d]$ 线性码, 其中 n 为码长, k 为维数, d 为极小距离, 设 C 的校验矩阵 H 是一个 $m \times n$ 矩阵, H 的行可以是相关的, 但必须满足 $\text{rank}(H) = n - k$ 。一个二元向量的支撑是指该向量非零分量位置的集合。对于线性码 C 的每个校验矩阵 H , 都有一个一一对应的图表示, 称为 Tanner 图, 记为 G_H , girth 定义为 Tanner 图中的最小圈长, girth 一定是不小于 4 的偶数, 通常认为 girth 越大, 迭代译码的性能越好。容易看出, G_H 的 girth 等于 4 的充要条件是 H 存在两行在某两个位置都为 1。给定 LDPC 码的校验矩阵 H , 其译码可在 G_H 上通过迭代的消息传递算法进行, 和积算法(Sum Product Algorithm, SPA)^[1]是其中最著名的一种。二元码的纠错性能是用“译码错误概率”来衡量的, 对于同一

个二元码, 不同的译码算法得到的译码错误概率是不同的, 在所有码字等概率发送的条件下, 最大似然译码是使得译码错误概率最低的最优译码, 但对于线性码, 最大似然译码一般是不可行的 NP 完全问题, 和积算法在校验矩阵中 1 的“密度”很低时是一种可行的次优译码, 是目前最好的可行译码算法之一。对于固定的线性码, 如果校验矩阵(或其 Tanner 图)选择的好, 那么和积算法可能会很卓越, 甚至与最大似然译码相差无几, 否则和积算法可能会很差。因此从迭代译码的角度来看, 构造好的 LDPC 码的主要问题是构造一个好的校验矩阵。

有限几何 LDPC 码是一类重要的有结构的 LDPC 码^[2,3], 在利用和积算法(SPA)等迭代译码方法进行译码时表现出卓越的纠错性能。众所周知, 次优的迭代译码不是最大似然译码, 因而如何对迭代译码的性能进行理论分析一直是 LDPC 码的核心问题之一。近几年来, 关于 LDPC 码的性能分析有了较大进展, 许多新的概念在译码的框架下被提了出来, 停止集(stopping set)和停止距离(stopping distance)是其中重要的两个。本文通过分析有限平面 LDPC 码的停止集和停止距离, 从理论上证明了有限平面 LDPC 码的最小停止集一定

2005-10-24 收到, 2006-04-19 改回

国家自然科学基金(60402031)和国家重点基础研究发展计划 973 (2003CB314805)资助课题

是最小重量码字的支撑,从而对有限平面LDPC码在迭代译码下的良好性能给出了理论解释。

2 有限几何 LDPC 码和停止集

Kou, Lin, Fossorier^[2]在2001年利用有限几何的方法首次构造了有结构的LDPC码,引起极大关注。其中有限平面LDPC码至今都是最好的LDPC码之一,有限平面LDPC码的girth为6。有限几何分为射影几何和仿射几何两大类,仿射几何又称为欧氏几何。有限几何可以通过有限域 F_q 来构造,设 $PG(m, q)$ 表示 F_q 上的 m 次射影几何, $AG(m, q)$ 表示 F_q 上的 m 次仿射几何。 $m=2$ 时,有限几何称为有限平面。

射影平面 $PG(2, 2^s)$ 有以下性质^[2]: $PG(2, 2^s)$ 包含 n 个点, n 条线, $n = 4^s + 2^s + 1$, H_{PG} 为 $n \times n$ 点线关联矩阵(incident matrix); 每条线有 $2^s + 1$ 个点, H_{PG} 的行重 $\rho = 2^s + 1$; 每个点在 $2^s + 1$ 条线上, H_{PG} 的列重 $\gamma = 2^s + 1$; 任意两条线交于唯一一点, $\text{girth} > 4$ (实际上 $\text{girth} = 6$); 以 H_{PG} 为校验矩阵的LDPC码称为射影平面LDPC码 C_{PG} , 极小距离 $d = 2^s + 2$, 维数 $k = 4^s - 3^s + 2^s$ 。

仿射平面 $AG(2, 2^s)$ ^[2]包含 4^s 个点, $4^s + 2^s$ 条线, 每条线有 2^s 个点, 每个点在 $2^s + 1$ 条线上。在这里本文做一点小的改动。去掉 $AG(2, 2^s)$ 中任意一点及经过该点的 $2^s + 1$ 条线, 剩余部分的点线关联矩阵记为 H_{AG} , 满足以下性质^[2]: H_{AG} 为 $n \times n$ 矩阵, $n = 4^s - 1$; 每条线有 2^s 个点, H_{AG} 的行重 $\rho = 2^s$; 每个点在 2^s 条线上的列重 $\gamma = 2^s$; 任意两条线要么交于唯一的一点, 要么平行, $\text{girth} > 4$ (实际上 $\text{girth} = 6$); 以 H_{AG} 为校验矩阵的LDPC码称为仿射平面LDPC码 C_{AG} , 极小距离 $d = 2^s + 1$, 维数 $k = 4^s - 3^s$ 。

设 C 为二元 $[n, k, d]$ 线性码, 其中 n 为码长, k 为维数, d 为极小距离。设 H 为 C 的校验矩阵, H 为 C 中最小码字(即重量为 d 的码字)的个数。众所周知, 当信噪比(SNR)较大时, 线性码 C 最大似然译码的译码错误概率主要依赖于 d 和 A_d 。当 d 较小时, C 的错误概率曲线将会出现所谓的“地板效应”(error floor), 增加线性码的极小距离是消除或减少地板效应的有效手段。次优的迭代译码不是最大似然译码, 其性能不适宜这样来刻画。文献[5-7]定义了校验矩阵 H (或其Tanner图)的停止集, 停止集 S 是位置集合 $\{1, 2, 3, \dots, n\}$ 的子集, 使得 H 在 S 上的投影不包含重量为1的行, 停止集元素个数的最小值称为停止距离, 记为 $s(H)$, 设 $B_s(H)$ 为最小停止集(即包含 $s(H)$ 个元的停止集)的个数。从Tanner图的观点来看, 停止集 S 是变量节点的子集使得 S 的邻居 $N(S)$ 中的每一个校验节点与 S 相连的边数大于等于2。文献[5]指出在二元删除信道(BEC)中的迭代译码完全依赖于停止集。在其他信道中, 停止集和停止距离也起着重要的作用, 反映了迭代译码的效率和性能^[7]。 $s(H)$ 与 $B_s(H)$ 在迭代译码中起着类似于 d 和 A_d 在最大似然译码中的作用。所以我们应该选择适当的校验矩阵 H , 使得 $s(H)$ 尽量大而且 $B_s(H)$ 尽量小。容易知道, 由于码字一定

满足校验方程, 故任何码字的支撑都是一个停止集, 所以对任意校验矩阵 H 都有 $s(H) \leq d$, 另一方面, Schwartz和Vardy^[7]证明只要仔细选择校验矩阵 H , 上述上界是一定可以达到的, 即 $s(H) = d$, 此时我们记 $B_s(H)$ 为 $B_d(H)$ 。由于最小重量码字的支撑同样是停止集, 故对任何使得 $s(H) = d$ 的校验矩阵 H 都有 $B_d(H) \geq A_d$ 。我们能使 $B_d(H)$ 达到最小值 A_d 吗? 在后面的部分, 我们将证明 H_{PG} 和 H_{AG} 正是使得有限平面LDPC码达到 $s(H) = d$ 且 $B_d(H) = A_d$ 的校验矩阵!

3 主要结果

设 C 为二元 $[n, k, d]$ 线性码, 其中 n 为码长, k 为维数, d 为极小距离, 设 H 为 C 的一个 $m \times n$ 校验矩阵。

定义 1^[7] 停止集 S 是位置集合 $\{1, 2, 3, \dots, n\}$ 的子集, 使得 H 在 S 上的投影不包含重量为1的行。停止集元素个数的最小值称为停止距离, 记为 $s(H)$, 换句话说, 停止距离是最大整数 $s(H)$, 使得 H 中任意 $s(H)-1$ 或更少的列包含至少一个重量为1的行。

由于 H 中任意 $s(H)-1$ 或更少的列包含一个重量为1的行, 故 H 的任意 $s(H)-1$ 列线性无关, 所以 $d \geq s(H)$ 。从另一个角度看, H 的每一行表示一个校验方程, 由于每一个码字一定满足所有 m 个校验方程, 故 H 的每一行在码字支撑上一定具有偶数个1, 所以码字支撑必然构成停止集, 因此 $s(H) \leq d$ 对任意校验矩阵 H 都成立。对于任何信道, 大的停止距离对于迭代译码的性能至关重要^[5]。文献[7]给出了以下重要结果: 任给校验矩阵 H_0 , 一定可以通过增加 H_0 的行得到新的校验矩阵 H 使得停止距离 $s(H)$ 达到上界 d 。由于最小重量码字的支撑同样是停止集, 故立刻有以下引理。

引理 1 设 C 为二元 $[n, k, d]$ 线性码, 其中 n 为码长, k 为维数, d 为极小距离。设 H 为使得 $s(H) = d$ 成立的校验矩阵, 则 $A_d \leq B_d$, 其中 A_d 为 C 中最小码字的个数, $B_d(H)$ 为 H 中最小停止集的个数。

有限平面LDPC码的非凡之处在于标准构造中的校验矩阵 H_{PG} 和 H_{AG} 恰好达到最优的情形, 即 $s(H) = d$ 且 $A_d = B_d(H)$, 我们认为这是有限平面LDPC码在迭代译码下表现出色的原因之一。为证明这个结论, 我们需要以下定理。

定理 1 设 C 为二元 $[n, k, d]$ 线性码, 其中 n 为码长, k 为维数, d 为极小距离。设 H 为 C 的校验矩阵, H 满足以下正交性质: 对每一个位置 $i, 1 \leq i \leq n$, H 中存在 γ 行($\gamma \geq 1$)在第 i 位为1, 在其他位置该 γ 行至多有一行为1, 则 H 的任意 u 列($1 \leq u \leq \gamma$)至少包含 $u(\gamma + 1 - u)$ 个重量为1的行, 故有 $s(H) \geq \gamma + 1$ 。

证明 对固定的 $u, 1 \leq u \leq \gamma$, 令 $H(u)$ 为由 H 的任意 u 列组成的 $m \times u$ 矩阵。由正交性, $H(u)$ 存在 γ 行其第1个分量为1, 而其他分量至多有一个为1。现在, 我们通过两种方法计算 $H(u)$ 的这 γ 行中1的总数。从列来数, 由正交性

知 1 的总数至多为 $\gamma + (u-1)$ 。另一方面,从行来数,假设 x 是这 γ 行中重量为 1 的行数,则有 $\gamma - x$ 行重量大于等于 2,故 1 的总数至少为 $x + 2(\gamma - x)$ 。因此, $x + 2(\gamma - x) \leq \gamma + (u-1)$, 即 $x \geq \gamma + 1 - u$, 换句话说, $\mathbf{H}(u)$ 至少包含 $\gamma + 1 - u$ 个重量为 1 的行, 而且这些行的第 1 个分量为 1。同理, $\mathbf{H}(u)$ 至少包含 $\gamma + 1 - u$ 个重量为 1 的行, 而且这些行的第 u 个分量为 1。显然, $\mathbf{H}(u)$ 的这些重量为 1 的行是两两不同的, 所以 $\mathbf{H}(u)$ 包含至少 $u(\gamma + 1 - u)$ 个重量为 1 的行。由 $1 \leq u \leq \gamma, u(\gamma + 1 - u) \geq \gamma \geq 1$, 所以 $\mathbf{H}$ 中的任意 γ 或更少的列至少包含一个重量为 1 的行, 由定义 1 知 $s(\mathbf{H}) \geq \gamma + 1$ 。证毕

注 1 从文献[8]可以导出 $s(\mathbf{H})$ 大于等于“最小伪重量”。LDPC 码的伪码字和伪重量涉及到线性规划译码^[8,9], 定义比较复杂, 限于篇幅本文不做详细解释。从文献[10,11]可知, 若 $\mathbf{H}$ 的 girth 大于 4, 则最小伪重量大于等于 $\mathbf{H}$ 中列的最小重量加 1。因此, 对于满足这些条件的 LDPC 码, 例如有限平面 LDPC 码, 同样可以得到定理 1 中 $s(\mathbf{H}) \geq \gamma + 1$ 的结论。

注 2 满足定理 1 的线性码实际上是一步大数逻辑可译码, 所以必然有 $d \geq \gamma + 1$ ^[4]。这说明所有的一步大数逻辑可译码都具有较大的停止距离, 因此可用来构造 LDPC 码。由于 $s(\mathbf{H}) \leq d$, 故对一步大数逻辑可译码来说 $d = \gamma + 1$ 隐含 $s(\mathbf{H}) = d$ 。

定理 2 设 $\mathbf{C}$ 为二元 $[n, k, d]$ 线性码, 其中 n 为码长, k 为维数, d 为极小距离。设 $\mathbf{H}$ 为 $\mathbf{C}$ 的校验矩阵, $\mathbf{H}$ 的 girth 大于 4, 且最小列重量为 γ 。如果 $d = \gamma + 1$, 则 $B_d(\mathbf{H}) = A_d$, 其中 A_d 为 $\mathbf{C}$ 中最小码字的个数, $B_d(\mathbf{H})$ 为 $\mathbf{H}$ 中最小停止集的个数。

证明 由于 $\mathbf{H}$ 的 girth 大于 4, 且最小列重量为 γ , 易知 $\mathbf{H}$ 满足由定理 1 中的正交条件, 故 $s(\mathbf{H}) \geq \gamma + 1$ 。又由 $s(\mathbf{H}) \leq d$ 和 $d = \gamma + 1$ 知 $s(\mathbf{H}) = \gamma + 1$ 。令 S 为任意一个最小停止集, 下面我们证明 S 一定是最小码字的支撑, 即 S 的关联向量 $\chi(S)$ ($\chi(S)$ 在 S 中的位置为 1, 在其他位置为 0) 一定是一个码字。令 $\mathbf{H}(S)$ 为 $\mathbf{H}$ 在 S 上的投影。我们将证明 $\mathbf{H}(S)$ 的每一行重量要么是 0 要么是 2, 从而得出 $\chi(S)$ 一定是一个码字。

$\mathbf{H}(S)$ 有 $\gamma + 1$ 列, 设 $\mathbf{b}$ 是 $\mathbf{H}(S)$ 中任意一个重量非零行。设 $b_j = 1$, 其中 $1 \leq j \leq \gamma + 1$ 。由于 $\mathbf{H}(S)$ 的第 j 列重量不小于 γ , 故存在不同于 $\mathbf{b}$ 的 $\gamma - 1$ 行, 其第 j 个分量为 1, 取该 $\gamma - 1$ 行与 $\mathbf{b}$ 一起构成矩阵 $\mathbf{H}(S, j)$, 显然 $\mathbf{H}(S, j)$ 为 $\gamma \times (\gamma + 1)$ 矩阵, girth 大于 4, 第 j 列为全 1 列, 其它列至多有一个 1。现在我们计算 $\mathbf{H}(S, j)$ 中 1 的总数。从列来数, $\mathbf{H}(S, j)$ 的第 j 列有 γ 个 1, 其它列至多有一个 1, 故 1 的总数至多为 2γ 。另一方面, 由于 S 是一个停止集, 而且 $\mathbf{H}(S, j)$ 的每一行重量非零, 故 $\mathbf{H}(S, j)$ 的每一行至少有 2 个 1, 这样 1 的总数至少为 2γ 。因此, $\mathbf{H}(S, j)$ 一定满足下面的性质: 第 j 列有 γ 个

1, 其它每一列恰有 1 个 1, 而且每一行恰好有两个 1。所以 $\mathbf{b}$ 恰有 2 个 1, 即 $\mathbf{H}(S)$ 中任意一个重量非零行都恰有两个 1。因此, $\mathbf{H}(S)$ 的每一行重量要么是 0 要么是 2, 从而满足 $\mathbf{H}$ 中所有的校验方程, 即 $\chi(S)$ 一定是一个码字。证毕

容易验证, 有限平面 LDPC 码满足定理 2 的条件, 所以我们有以下推论:

推论 1 设 $\text{PG}(2, 2^s)$ 是一个射影平面, 则以 $\mathbf{H}_{\text{PG}}$ 为校验矩阵的射影平面 LDPC 码 $\mathbf{C}_{\text{PG}}$ 一定满足: $s(\mathbf{H}_{\text{PG}}) = d = 2^s + 2, A_d = B_d(\mathbf{H}_{\text{PG}})$ 。

推论 2 设 $\text{AG}(2, 2^s)$ 是一个仿射平面, 则以 $\mathbf{H}_{\text{AG}}$ 为校验矩阵的仿射平面 LDPC 码 $\mathbf{C}_{\text{AG}}$ 一定满足: $s(\mathbf{H}_{\text{AG}}) = d = 2^s + 1, A_d = B_d(\mathbf{H}_{\text{AG}})$ 。

注 3 文献[6]讨论了组合设计 LDPC 码的停止集, 并对于有限平面 LDPC 码得到了停止距离等于极小距离的结果, 但没有进一步研究 A_d 是否等于 $B_d(\mathbf{H})$ 的问题。

例 1 考虑 $\text{PG}(2, 2)$, $\mathbf{C}_{\text{PG}}$ 是一个 $[7, 3, 4]$ 线性码, 校验矩阵 $\mathbf{H}$ 为一个循环 7×7 的矩阵, 首行为 $(1, 1, 0, 1, 0, 0, 0)$ 。停止集分布为 $1 + 7x^4 + 7x^6 + x^7$, 重量分布为 $1 + 7x^4$, 满足 $s(\mathbf{H}) = d = 4, A_d = B_d(\mathbf{H}) = 7$ 。

例 2 考虑 $\text{AG}(2, 4)$, $\mathbf{C}_{\text{AG}}$ 是一个 $[15, 7, 5]$ 线性码, 校验矩阵 $\mathbf{H}$ 为一个循环 15×15 的矩阵, 首行为 $(1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0)$ 。停止集分布为

$$1 + 18x^5 + 30x^6 + 15x^7 + 105x^8 + 325x^9 + 513x^{10} + 705x^{11} + 395x^{12} + 105x^{13} + 15x^{14} + x^{15},$$

重量分布为

$$1 + 18x^5 + 30x^6 + 15x^7 + 15x^8 + 30x^9 + 18x^{10} + x^{15},$$

满足 $s(\mathbf{H}) = d = 5, A_d = B_d(\mathbf{H}) = 18$ 。

4 结束语

本文研究了有限平面 LDPC 码的停止集和停止距离, 从理论上证明了有限平面 LDPC 码的最小停止集一定是最小重量码字的支撑, 从而对有限平面 LDPC 码在迭代译码下的良好性能给出了理论解释。有限平面 LDPC 码的非凡之处在于标准构造中的校验矩阵 $\mathbf{H}_{\text{PG}}$ 和 $\mathbf{H}_{\text{AG}}$ 都恰好达到最优的情形, 即停止距离等于极小距离同时最小停止集的个数恰好等于最小码字的个数, 我们认为这是有限平面 LDPC 码在迭代译码下表现出色原因之一。对于一般的有限几何 LDPC 码^[3], 确定其维数、极小距离和停止距离是一件困难的工作。伪码字集合是码字集合的超集。对于二元删除信道, 伪码字与停止集是等价的^[9]。对于加性白高斯噪声(AWGN)信道, 伪码字是一个比停止集更为“精细”的概念。Vontobel 等人^[12]讨论了有限平面 LDPC 码的极小伪码字(minimal pseudo-codeword)并给出了一些模拟结果, 但未能从理论上证明他们的想法, 即具有低伪重量的极小伪码字一定是一个极小码字(minimal codeword)。研究 LDPC 码的伪码字结构是比研究停止集结构更为困难和有趣的工作。

参 考 文 献

- [1] Kschischang F R, Frey B J, and Loeliger H A. Factor graphs and the sum-product algorithm. *IEEE Trans. on Inform. Theory*, 2001, 47(2): 498–519.
- [2] Kou Y, Lin S, and Fosshier M P C. Low-density parity-check codes based on finite geometries: a rediscovery and new results. *IEEE Trans. on Inform. Theory*, 2001, 47(7): 2711–2736.
- [3] Tang H, Xu J, Lin S, and Abdel-Ghaffar K A S. Codes on finite geometries. *IEEE Trans. on Inform. Theory*, 2005, 51(2): 572–569.
- [4] Lin S and Costello DJ. Error Control Coding: Fundamentals and Applications, 2nd Ed, Upper Saddle River, NJ: Prentice-Hall, 2004, Chap. 8: 273–282.
- [5] Di C, Proietti D, Telatar I E, Richardson T J, and Urbanke R L. Finite-length analysis of low-density parity-check codes on the binary erasure channel. *IEEE Trans. on Inform. Theory*, 2002, 48(6): 1570–1579.
- [6] Kshyap N and Vardy A. Stopping sets in codes from designs. in Proc. IEEE Int. Sym. Inform. Theory, Yokohama, Japan, Jul. 2003: 122.
- [7] Schwartz M and Vardy A. On the stopping distance and the stopping redundancy of codes. *IEEE Trans. on Inform. Theory*, 2006, 52(3): 922–932.
- [8] Koetter R and Vontobel P O. Graph covers and iterative decoding of finite-length codes. In Proc. 3rd Int. Symp. Turbo Codes and Related Topics, Brest, France, Sept. 2003: 75–82.
- [9] Feldman J, Wainwright M J, and Karger D R. Using linear programming to decode binary linear codes. *IEEE Trans. on Inform. Theory*, 2005, 51(3): 954–972.
- [10] Chaichanavong P and Siegel P H. Relaxation bounds on the minimum pseudo-weight of linear codes. In Proc. IEEE Int. Symp. Inform. Theory, Sept. Adelaide, Australia 2005: 805–809.
- [11] Vontobel P O and Koetter R. Lower bounds on the minimum pseudo-weight of linear codes. In Proc. IEEE Int. Symp. Inform. Theory, Chicago, U.S.A., 2004: 70.
- [12] Vontobel P O, Smarandache R, and Kiyavash N, *et al.* On the minimal pseudo-codewords of codes from finite geometries. In Proc. IEEE Int. Symp. Inform. Theory, Adelaide, Australia, Sept. 2005: 980–984.

夏树涛：男，1972 年生，博士，副教授，主要研究方向为信道编码和网络安全。

胡懋智：男，1984 年生，硕士生，研究方向为纠错码及计算机网络。