数理解析研究所講究録
第1849巻 2013年 109-115
2012年度冬のLAシンポジウム [S16]

A Fourier analytic approach to list-decoding for sparse random linear codes

山根一航* 河内亮周*

平成25年3月15日

概要

疎なランダム線形符号のリスト復号問題はこれまで時間効率的なアルゴリズムが知られていない重要な問題である．Kopparty と Saraf はこの問題に対して受信語へのクエリー回数が効率的なリスト復号器を構成した [3]．彼らの手法は疎なランダム線形符号のリスト復号問題を時間効率的なリスト復号が可能な Hadamard 符号のリスト復号問題への帰着を与えるというものである．本論文では Kopparty と Saraf と同様にクエリー回数が効率的な疎なランダム線形符号の別のリスト復号器を構成する．Kopparty と Saraf が組合せ論的な操作により他の符号への帰着を与えたのに対して我々のアルゴリズムは Fourier 解析的な手法を用いることでより直接的なリスト復号器となっている．

1 はじめに

誤り訂正符号は計算機科学の中でも重要な研究対象であり，とりわけ**リスト復号**と呼ばれる復号の問題は，理論計算機科学における他の問題との深い関連性を持つことが知られている．リスト復号とは，任意のメッセージを，固定された符号で符号化して得た符号語が，ノイズの付与する通信路を通して送られてきたという設定の下で，通信路のエラー率の上界と送られてきた受信語をもらって，その情報から考えられる元のメッセージの候補をもれなく列挙する問題である．リスト復号の問題は，元のメッセージを一意に推測しなければならない一意復号の問題とは異なり，高いエラー率にも対処することができ，広い応用を持つ．例えば疑似乱数の理論では疑似乱数生成器などの疑似乱数性証明にリスト復号器を用いたり [1]，学習理論でも与えられた学習対象のサンプルに誤りが付与された場合の学習アルゴリズム [6] に応用できることが知られている．このように理論計算機科学の多くの分野で考えられる問題は単純なリスト復号の問題として統一的に解釈でき，広い応用を持つ．より詳しくはリスト復号およびその応用に関するサーベイ論文 [5] を参照せよ．

そのようなリスト復号問題について，本論文では特にランダム線形符号を議論する．**ランダム線形符号**とは，生成行列をランダムに選ぶような線形符号のことである．ランダム線形符号の復号は歴史が長く，本質的に難しい問題であると思われてきた．しかし2007年に Kaufman と Sudan は，低エラー率の場合に限り，符号語数が符号語長の多項式で抑えられる**疎な**ランダム線形符号に対してクエリー効率的に一意復号が可能であることを示した [2]．ここでクエリー効率的とは，復号器が受信語 w の全ビットを読むのではなく，w のビットごとのアクセスが可能なモデルにおいて，読み取る w のビット数が $\mathrm{poly}(n, 1/\varepsilon)$ で抑えられる場合（つまり受信語 w がオラクル $i \mapsto w_i$ として与えられた場合にこのオラクルへのクエリー回数が $\mathrm{poly}(n, 1/\varepsilon)$）をいう．さらに2010年には Kopparty と Saraf が，疎なランダム線

*Dept. Math. & Comp. Sci., Tokyo Instute of Technology

形符号のリスト復号をHadamard符号のリスト復号へ帰着することによって，高エラー率の下での，クエリー効率的リスト復号器を与えた[3]．Hadamard符号のリスト復号は，いくつかのアルゴリズムによって効率的に解くことが出来ることが知られている[1, 4]．ここで効率的という場合にはクエリー効率的かつ復号器の時間計算量も $\mathrm{poly}(n, 1/\varepsilon)$ で抑えられる場合をいう．

このようにKopparty と Sarafのアプローチはランダム線形符号に対するリスト復号器を設計したというよりもランダム線形符号とHadamard符号の関連性を組合せ論的な方法から明らかにした結果であり，直接的にランダム線形符号のリスト復号器を設計したものではない．本論文ではKushilevitz と Mansourの，Fourier解析的性質を利用した学習アルゴリズム[4]（以下KMアルゴリズムと呼ぶ）に着目し，それを拡張することによって疎なランダム線形符号に対するクエリー効率的リスト復号器を直接的に構成した．この拡張KMアルゴリズムはKopparty と Sarafのものと同様時間効率的ではないが，よりアルゴリズム的なアプローチであり，時間効率的なランダム線形符号のリスト復号器に向けての新たな知見を与え，またそのようなFourier解析的手法が疎なランダム線形符号に対しても有用であることを示すものである．

2 準備

本論文ではアルファベットとして二元体 $\mathbb{F}_2 := \{0, 1\}$ のみを対象とする．長さ $k \in \mathbb{N}$ の**語**または**文字列** w とは，各成分が $\mathbb{F}_2$ の元であるような k 次元の行ベクトルである．$k \geq 1$ のときは w の $i \in [k]$ 文字目を w_i と書く．つまり $w := (w_1, \ldots, w_k)$ である．また w の長さを $|w|$ で表し，文字列 v, w の連結を vw で表す．0のみからなる長さ k の文字列を 0^k と書き，文脈で長さが判断できる場合は単に $\mathbf{0}$ と書く．w の**正規化ハミング重み**を $\mathrm{wt}(w) := |\{i \in [k] \mid w_i \neq 0\}|/|w|$ で定義し，v と w の**正規化ハミング距離**を $\Delta(v, w) := \mathrm{wt}(v - w)$ で定義する．

2.1 （リスト）復号問題

本論文では誤り訂正符号の復号問題を考える．通信路のモデルとしては予め決まった数の任意のビットが反転するエラーを与えるものを考える．つまり符号語 $c(x)$ とその受信語 w に対してエラーベクトル $e := w - c(x)$ が**優位度パラメータ** $\varepsilon \in (0, 1)$ で表される条件 $\mathrm{wt}(e) \leq (1 - \varepsilon)/2$ を満たすことだけが保証される．ここで $\mathrm{wt}(e)$ を w の**エラー率**と呼ぶ．符号 $C = \{c(x) \in \mathbb{F}_2^N \mid x \in \mathbb{F}_2^n\}$ の復号問題はさまざまな定式化が考えられるが，本論文で対象とするのは，通信路のエラー率の上界と送られてきた受信語から考えうるメッセージ候補をもれなく列挙する**リスト復号**の問題である．

問題 2.1 (C のリスト復号)．
任意の受信語 $w \in \mathbb{F}_2^n$ と優位度パラメータ $\varepsilon \in (0, 1)$ を受け取って，$\{x \in \mathbb{F}_2^n \mid \Delta(c(x), w) \leq (1 - \varepsilon)/2\} \subseteq L$ を満たす $L \subseteq \mathbb{F}_2^n$ を求めよ．

本論文を始め，多くの論文ではリスト復号器の性能を次の2つの観点から評価を行っている．1つ目は**クエリー数**であり，2つ目は**リストサイズ**である．リストサイズはそのまま出力リスト L の要素数 $|L|$ のことを指すが，クエリー数については以下に説明する．

標準的な誤り訂正符号の理論では受信者は受信語の全ビットを読むことを前提に復号を行う場合が多いが，理論計算機科学の文脈ではメッセージ長 n に対して非常に長い符号語 N に対する復号を考えることが多い．このような場合，復号器は符号語に対して局所的なアクセスを行うことで復号を行う．つまり，受信語 $w \in \mathbb{F}_2^N$ を関数（オラクル）$w : [N] \to \mathbb{F}_2$ としてみなし，復号器はクエリー $i \in [N]$ に対して $w_i \in \mathbb{F}_2$ をオラクル $w : i \mapsto w_i$ から受け取ることができる．復号のために必要なこのクエリー数を性能評価に用いる．

我々の目標はこの2つの量を $n, 1/\varepsilon$ の多項式で抑えることである．これを達成するリスト復号器を**クエリー効率的**なリスト復号器と呼ぶことにする．

2.2 ランダム線形符号

本論文を通して**ランダム線形符号**と呼ばれる符号が議論の中心となる．ランダム線形符号は，各要素を独立に1/2の確率で0か1に決めて作られた**ランダム行列**を生成行列とする線形符号である．$2^n \leq \mathrm{poly}(N)$ となるような大きな符号語長をもつランダム線形符号を**疎な**ランダム線形符号という．

符号 $\mathcal{C}$ の**偏り**は $\mathrm{bias}(\mathcal{C}) := \max_{c\in\mathcal{C}-\{0\}}|1/2 - \mathrm{wt}(c)|$ で定義され，ある定数 $\gamma > 0$ が存在して，$\mathrm{bias}(\mathcal{C}) \leq N^{-\gamma}$ を満たすとき，$\mathcal{C}$ は**不偏**であるという．ランダム線形符号の偏りについて次の定理が成り立つ．(この定理は簡単であるため証明は省略する.)

定理 2.2. 任意の $\beta \in (0,1), n, N \in \mathbb{N}$ に対して，$1 - O(2^n \exp\{-\beta^2 N/6\})$ 以上の確率で $\mathrm{bias}(\mathcal{C}) \leq \beta$.

2.3 Fourier解析

本論文では問題の定式化や証明に Fourier 解析の道具を用いる．ここでは Fourier 解析についての基本的な定義や事実について簡単にまとめる．関数の空間として $\mathcal{F}_n := \{f : \mathbb{F}_2^n \to \mathbb{R}\}$ を考える．これは $\mathbb{R}$ 上の 2^n 次元ベクトル空間とみなすことができる．任意の $f, g \in \mathcal{F}_n$ に対して次のように内積を，$\langle f, g\rangle := \mathbf{E}_{X\sim U_n}[f(X)g(X)]$ と定義する．ここで U_n は $\mathbb{F}_2^n$ 上の一様分布である．また，長さ $k \in \mathbb{N}$ の文字列 a でラベル付けられた関数 $\chi_a \in \mathcal{F}_k$ を $\chi_a(x) := (-1)^{a\cdot x}$ で定義する．ここで $a\cdot x = a_1x_1 + \cdots + a_kx_k$ である．$f \in \mathcal{F}_j, g \in \mathcal{F}_k$ に対し，$f \bullet g \in \mathcal{F}_{j+k}$ を $f \bullet g : xy \mapsto f(x)g(y)$ で定義する．任意の $a, b \in \mathbb{F}_2^*$ について $\chi_{ab} = \chi_a \bullet \chi_b$ であること，$\sum_{v\in\mathbb{F}_2^k}\chi_v(a) = 2^k \mathbf{E}_{X\sim U_k}[\chi_a(X)] = 2^k\delta_k(a)$，および $|a| = |b|$ のとき $\langle\chi_a, \chi_b\rangle = \delta_k(a+b)$ が成り立つことなどは簡単に確かめられる．ここで $\delta_k \in \mathcal{F}_k$ は $x = 0^k$ ならば $\delta_k(x) = 1$ それ以外は $\delta_k(x) = 0$ となる関数である．$\{\chi_a\}_{a\in\mathbb{F}_2^n}$ は $\mathcal{F}_n$ の正規直交基底となるが．これを Fourier 基底と呼び，Fourier 基底の各要素を Fourier 基底ベクトルと呼ぶ．基底の性質から，$f(x) = \sum_{a\in\mathbb{F}_2}\langle f, \chi_a\rangle\chi_a(x)$ が成り立つ．また，f の ℓ_2 ノルム $\|f\|_2$ を $\langle f, f\rangle$ の平方根で定義すると，次の等式が成り立つことが知られている．

命題 2.3 (Parseval の等式). $\|f\|_2^2 = \sum_{a\in\mathbb{F}_2}\langle f, \chi_a\rangle^2$.

本論文では，内積の拡張にあたる μ 半内積と，合成積の拡張にあたる μ 合成積を用いる．

定義 2.4 (μ 半内積，μ 合成積). μ を $\mathbb{F}_2^n$ 上の分布とする．$f \in \mathcal{F}_n, l \in \mathbb{N}$ に対して μ **半内積** $\langle f, g\rangle_\mu$ を $\mathbf{E}_{Y\sim\mu}[f(Y)g(Y)]$ で定義する．また，$Y^{(1)}, \ldots, Y^{(l)} \sim \mu$ とし，$Y^{(1)} + \cdots + Y^{(l)}$ が従う分布を $\mu^{(l)}$ と書くとき，f の l **階** μ **合成積** $f^{[\mu,l]}$ を $\mu^{(l)}(y) > 0$ となる $y \in \mathbb{F}_2^n$ で $f^{[\mu,l]}(y) := \mathbf{E}\left[\prod_{i=1}^l f(Y^{(i)}) \mid \sum_{i=1}^l Y^{(i)} = y\right]$，$\mu^{(l)}(y) = 0$ となる $y \in \mathbb{F}_2^n$ では $f^{[\mu,l]}(y) := 0$ と定義する．

命題 2.5. 任意の $f \in \mathcal{F}_n, a \in \mathbb{F}_2^n$ と $\mathbb{F}_2^n$ 上の分布について $\langle f^{[\mu,l]}, \chi_a\rangle_{\mu^{(l)}} = \langle f, \chi_a\rangle_\mu^l$.

2.4 KMアルゴリズム

KM アルゴリズム [4] は関数の学習などに用いられるが，正確には次で定義される **Fourier 基底ベクトル列挙問題**を効率的に解くアルゴリズムである．

問題 2.6 (Fourier 基底ベクトル列挙問題 通常版). 関数 $f : \mathbb{F}_2^n \to [-1, +1]$ がオラクル $x \mapsto f(x)$ として，しきい値 $\varepsilon \in (0,1)$ が入力として与えられたとき，$\{a \in \mathbb{F}_2^n \mid \varepsilon \leq |\langle f, \chi_a\rangle|\} \subseteq L$ を満たすリスト $L \subseteq \mathbb{F}_2^n$ を求めよ．

KM アルゴリズムを簡単に説明をするために次のような完全2分木を考える．各節点は $\mathbb{F}_2^{\leq n}$ の元であり，根は空列である．節点 $\alpha \in \mathbb{F}_2^{<n}$ は節点 $\alpha 0$ と節点 $\alpha 1$ を子に持つ．そして葉として $\mathbb{F}_2^n$ の元，つまりメッセージが並ぶ．$\varepsilon \leq \langle f, \chi_a\rangle$ を満たす $a \in \mathbb{F}_2^n$ を**必要な葉**と呼ぶ．KM アルゴリズムはこの木の節点を根から順に深さ優先的に訪問する．その際，訪問した節点が，必要な葉を子孫として持ち得るかどうかを後述する方法により評価し，持ち得ないことがわ

かるとその頂点を根とする部分木を枝刈りする．最後まで枝刈りされることなく残った葉の中には，必要な葉がすべて含まれていることになる．

枝刈りの基準を定めるために，$\alpha \in \mathbb{F}_2^k$ に対して**見積もり関数** $S_2(\alpha) := \sum_{\alpha' \in \mathbb{F}_2^{n-k}} \langle f, \chi_a \rangle^2$ を定義する．$S_2(\alpha) < \varepsilon^2$ ならば節点 α は必要な葉を子孫として持たないので，α を根とする部分木は枝刈りすることができる．この枝刈りの操作は，訪問する節点数を $\mathrm{poly}(n, 1/\varepsilon)$ 個に抑え，また，必要でない葉をふるい落として出力リストサイズを絞る効果がある．

枝刈り条件 $S_2(\alpha) < \varepsilon^2$ は定義どおり判定すると多くの時間を必要とするが，Fourier 基底の内積 $\langle \cdot, \cdot \rangle$ における直交性を利用すると，$S_2(\alpha)$ は次のような期待値の形に変形でき，サンプル平均により効率的に近似できる：$S_2(\alpha) = \mathbf{E}_{X^{(1)}, X^{(2)} \overset{\text{i.i.d}}{\sim} U_n}[f(X^{(1)}) f(X^{(2)}) \chi_\alpha(X^{(1)}{}_{\leq k} + X^{(2)}{}_{\leq k}) \mid X^{(1)}{}_{>k} + X^{(2)}{}_{>k} = \mathbf{0}]$．ここで $X^{(1)}, X^{(2)} \overset{\text{i.i.d}}{\sim} U_n$ である．これにより一回の枝刈り判定も効率的に行え，全体としても効率的に動作する．

3 Fourier 解析的アプローチによるリスト復号

本節の目標は次の定理を示すことである．

定理 3.1 (疎なランダム線形符号のクエリー効率的なリスト復号)．疎なランダム線形符号 C を選んだとき確率 $1 - o(1)$ で以下が成り立つ．C に対して問題 2.1 はリストサイズ $\mathrm{poly}(1/\varepsilon)$，クエリー数 $\varepsilon^{-O(1)} \cdot O(n \ln(n/\varepsilon))$ で定数確率以上で解ける．

ここで $1/\varepsilon = 2^{\Omega(n)}$ となる極端な状況では，リストサイズやクエリー数が 2^n となる自明なアルゴリズムによって問題を解くことが出来るため，以下では $1/\varepsilon = 2^{o(n)}$ となる場合のみを考える．

定理 3.1 を示すために 3.1 節でまず問題 2.1 を後述の問題 3.2 に帰着し，その問題 3.2 を解くための拡張 KM アルゴリズムを 3.2 節で構成する．また 3.3 節でそのアルゴリズムの解析を行う．

3.1 Fourier 基底ベクトル列挙問題への帰着

本節では線形符号のリスト復号の問題が Fourier 基底ベクトル列挙問題に帰着出来ることを示す．ただしそのためには Fourier 基底ベクトル列挙の問題を問題 2.6 よりも一般的な形にしておく必要がある．μ を $\mathbb{F}_2^n$ 上の分布とする．定義 2.4 で定義した μ 半内積を用いて問題 2.6 を一般化した次の問題を考える．

問題 3.2 (Fourier 基底ベクトル列挙問題 拡張版)．関数 $f : \mathbb{F}_2^n \to [-1, 1]$ がオラクル $x \mapsto f(x)$ として，しきい値 $\varepsilon \in (0, 1)$ が入力として与えられたときに，$\{a \in \mathbb{F}_2^n \mid \varepsilon \leq |\langle f, \chi_a \rangle_\mu|\} \subseteq L$ を満たすリスト $L \subseteq \mathbb{F}_2^n$ を求めよ．

任意の線形符号 C に対して，問題 2.1 は問題 3.2 に帰着出来ることが知られている．それを示すためにいくつか記号を準備する．線形符号 C の生成行列を $G = (g_1, \ldots, g_N)$ とし，$V := \{g_1{}^{\mathsf{T}}, \ldots, g_N{}^{\mathsf{T}}\}$ とする．簡単のため $g_1, \ldots, g_N$ はすべて異なるベクトルであるとする．ランダム線形符号は高い確率でこの仮定を満たすこと，この後の議論は一般の場合に容易に拡張可能であることを補足する．受信語 w に対して関数 $f_w : \mathbb{F}_2^n \to [-1, 1]$ を，$x = g_i{}^{\mathsf{T}}$ ならば $f_w(x) = (-1)^{w_i}$ それ以外は $\delta_k(x) = 0$ として定義する．また C **に関連した分布** μ を V の上での一様分布として定義する．次の補題が成り立つ．

補題 3.3. 問題 2.1 は問題 3.2 に帰着可能であり，後者がリストサイズ $|L|$，クエリー数 Q で解けるならば前者はリストサイズ $O(|L|)$，クエリー数 $O(Q)$ で解ける．

証明. 受信語 $w \in \mathbb{F}_2^N$ と優位度パラメータ $\varepsilon \in (0, 1)$ に対して符号 C についての問題 2.1 を解きたいとしよう．関数 f_w へのオラクル $x \mapsto f_w(x)$ は w へのオラクル $i \mapsto w_i$ で模倣することができ，この計算における w へのクエリー数は $O(1)$ である．C に関連した分布 μ をとり，関数 f_w としきい値 ε に対して μ についての問題 3.2 を解いて得られた解を L とすると，定義からわかるように，$\Delta(w, c(a)) \leq (1 - \varepsilon)/2$

は $\varepsilon \le \langle f_w, \chi_a \rangle_\mu$ と同値であり，これらが成り立つとき $a \in L$ が成り立つ．すなわち L は問題 2.1 の解でもある． □

3.2 拡張KMアルゴリズムの構成

前節で示した帰着より以下では問題 3.2 について議論する．疎なランダム線形符号のに関連した分布を含む $\mu = U_n$ でない場合の問題 3.2 は KM アルゴリズムによってクエリー効率的に解くことが出来ない．そこで我々は，KM アルゴリズムを，より一般的な分布 μ に対する問題 3.2 を解くものに拡張する．考えられる単純な拡張は見積もり関数を $S_2(\alpha) = \sum_{\alpha' \in \mathbb{F}_2^{n-k}} \langle f, \chi_{\alpha\alpha'} \rangle_\mu^2$ と変更することであるが，このようにすると $S_2(\alpha) < \varepsilon^2$ という条件判定をクエリー効率的に近似することが出来ない．それは μ 半内積の下では $\{\chi_a\}_{a \in \mathbb{F}_2^n}$ の異なる二つのベクトルの直交性が弱く，f と近い Fourier 基底ベクトルと遠いものとの差が μ 半内積で現れづらいことに起因している．この問題を解決するために我々は，**l 次見積もり関数** $S_l(\alpha) = \sum_{\alpha' \in \mathbb{F}_2^{n-k}} \langle f, \chi_{\alpha\alpha'} \rangle_\mu^l$ を新たに導入した．ここで l は偶数のパラメータである．$l = 2$ の場合は S_2 と一致する．l 次見積もり関数においてパラメータ l は，絶対値の大きな半内積と小さな半内積の差をどのぐらい強調するかを制御する．l を大きくすることはアルゴリズムの性能を下げる向きに働くが，疎なランダム線形符号に対しては l は高々定数で十分であることを示すことができる．すぐに分かるように，$a \in \mathbb{F}_2^n$ が $\varepsilon \le |\langle f, \chi_a \rangle_\mu|$ を満たすならばそのすべての接頭辞 α について $\varepsilon^l \le S_l(\alpha)$ が成り立つ．ゆえに $S_l(\alpha) < \varepsilon^l$ を枝刈り条件として用いることが出来る．

拡張 KM アルゴリズム EKM を Algorithm 1 に示す．ここで，与えるオラクルを右肩につけて表している．KM のときと同じように，枝刈り評価をクエリー効率的に行うために，EKM は S_l の代わりに近似を用いる，この近似値は Algorithm 2 で定義される $\bar{S}_l^f(\alpha)$ が計算する．アルゴリズム中の分布 $\mu|_{\le k}^l$ の定義は後述する．

Algorithm 1 $\mathrm{EKM}^f(\alpha, \varepsilon)$

1: **if** $\bar{S}_l^f(\alpha, \varepsilon) < \varepsilon^l/2$ **then return** .（枝刈り）
2: **else if** $|\alpha| = n$ **then** α を出力.
3: **else** $\mathrm{EKM}^f(\alpha 0, \varepsilon)$, $\mathrm{EKM}^f(\alpha 1, \varepsilon)$.
4: **end if**

Algorithm 2 $\bar{S}_l^f(\alpha, \varepsilon)$

1: $m \leftarrow O(\varepsilon^{-2l} \ln(n/\varepsilon)), k \leftarrow |\alpha|$
2: $(z^{(i,1)}, \ldots, z^{(i,l)}) \leftarrow \mu|_{\le k}^l$ for $i \in [m]$
3: **return** $\frac{1}{m} \sum_i \prod_{j=1}^l f(z^{(i,j)}) \chi_\alpha(\sum_{j=1}^l z^{(i,j)} \le k)$

3.3 拡張KMアルゴリズムの正当性

本節では次の定理を示す．

定理 3.4. $\rho(\mu) = 2^{-\Omega(n)}$ となる分布 μ について，任意の $f : \mathbb{F}_2^n \to [-1, 1], \varepsilon \in (0, 1)$ に対して $\mathrm{EKM}^f(\text{空列}, \varepsilon)$ を動作させると定数確率以上でクエリー数 $O(n\varepsilon^{-3l} \ln(n/\varepsilon))$，リストサイズ $O(1/\varepsilon^l)$ で問題 3.2 の解を出力する．

ここで $\rho(\mu) := \max_{a \ne b} |\langle \chi_a, \chi_b \rangle_\mu| \quad (a, b \in \mathbb{F}_2^n)$ であり，この量が小さいほど，μ 半内積において異なる Fourier 基底ベクトル同士の直交性が強いということができる．$\rho(\mu) = 2^{-\Omega(n)}$ という条件は次の命題に由来するものである．

命題 3.5. 疎なランダム線形符号に関連した分布 μ について，$1 - 2^{-\Omega(n)}$ 以上の確率で $\rho(\mu) = 2^{-\Omega(n)}$ が成り立つ．

証明. 定義より $\rho(\mu)$ は $\max_{a \ne b} \left| \langle \chi_a, \chi_b \rangle_\mu \right|$ つまり $\max_{a \ne b} |1/2 - \Delta(c(a), c(b))| = \mathrm{bias}(C)$ に等しい．C が疎なランダム線形符号のときには，ある定数 $c > 0$ をが存在して十分大きなすべての n について符号語長 N は $N \ge 2^{n/c}$ を満たす．従って定理 2.2 より $1 - 2^{-\Omega(n)}$ 以上の確率で $\rho(\mu)$ は $\mathrm{bias}(C) \le 2^{-n/2c} n = 2^{-\Omega(n)}$ で抑えられる． □

定理 3.4 が成り立ったとすると補題 3.3 と主張 3.5 から定理 3.1 が成り立つことが示される．以降では定理 3.4 を示すための議論が続くが，ま

ずはじめに，$\bar{S}_l$ が S_l の近似値を計算することを示す．$Y^{(1)},\ldots,Y^{(l)} \overset{\text{i.i.d}}{\sim} \mu$ とする．また，これらと互いに独立な確率変数 $Y \sim \mu^{(l)}$ を用意する．$(\mathbb{F}_2^n)^l$ 上の分布 $\mu\,|_{\le k}^l$ を $\mu\,|_{\le k}^l\,(z^{(1)},\cdots,z^{(l)}) = \Pr[\bigwedge_{i=1}^l Y^{(i)} = z^{(i)} \mid \sum_{i=1}^l Y^{(i)}{}_{>k} = \mathbf{0}]$ で定め，$(Z^{(1)},\cdots,Z^{(l)}) \sim \mu\,|_{\le k}^l$ とする．そして $E_l(\alpha) := \mathbf{E}[\prod_{i=1}^l f(Z^{(i)})\chi_\alpha(\sum_{i=1}^l Z^{(i)}{}_{\le k})]$ と定義する．次の補題 3.6 は $S_l(\alpha)$ を $E_l(\alpha)$ に十分近くできることを主張するものである．

補題 3.6. $\rho(\mu) = 2^{-\Omega(n)}, 1/\varepsilon = 2^{o(n)}$ のとき，任意の $k \le n, \alpha \in \{0,1\}^k$ に対してある定数 $l > 0$ が存在して，$|S_l(\alpha) - E_l(\alpha)| \le \varepsilon^l/8$.

証明. $S_l(\alpha)$ すなわち $\sum_{\alpha' \in \mathbb{F}_2^{n-k}} \left\langle f^{[\mu,l]}, \chi_{\alpha\alpha'} \right\rangle_{\mu^{(l)}}$ は，$2^n \left\langle f^{[\mu,l]}, \chi_\alpha \bullet \delta_{n-k} \right\rangle_{\mu^{(l)}}$ に等しく，$2^n \Pr[Y_{>k} = \mathbf{0}]\, \mathbf{E}\left[f^{[\mu,l]}(Y)\chi_\alpha(Y_{\le k}) \mid Y_{>k} = \mathbf{0}\right]$ と変形できるので $S_l(\alpha) = 2^{n-k}\Pr[Y_{>k} = \mathbf{0}]E_l(\alpha)$ が成り立つ．したがって $|\Pr[Y_{>k} = \mathbf{0}] - 1/2^{n-k}| \le 2^{-n}\varepsilon^l/8$ を示せばよい．左辺の 2 乗は，$2^n\|\mu^{(l)} - U_n\|_2^2$ で上から抑えられる．Parseval の等式と $U_n = \chi_{0^n}/2^n$ を使うと，$\left\|\mu^{(l)} - U_n\right\|_2^2 = \sum_{a \in \{0,1\}^n} \left\langle \mu^{(l)} - \frac{\chi_{0^n}}{2^n}, \chi_a \right\rangle^2 = \sum_{a \ne 0^n} \left\langle \mu^{(l)}, \chi_a \right\rangle^2$．一方，$a \ne 0^n$ について $|\langle \mu^{(l)}, \chi_a \rangle| \le 2^{-n}\rho(\mu^{(l)})$ であり，さらに $\rho(\mu^{(l)}) = \max_{a \ne 0^n} \mathbf{E}[\chi_a(Y^{(1)})] \cdots \mathbf{E}[\chi_a(Y^{(1)})] \le \rho(\mu)^l$ が成り立つ．したがって $|\Pr[Y_{>k} = \mathbf{0}] - 1/2^{n-k}| \le \rho(\mu)^l$ である．ここで，$\rho(\mu) = 2^{\Omega(n)}$ よりある定数 $d > 0$ が存在して，十分大きなすべての n について，$\rho(\mu) \le 2^{-n/d}$ である．l を定数 $\lambda > 0$ で $l \ge d + \lambda$ を満たす値にとると，$2^n\rho(\mu)^l \le 2^{-\lambda n/d}$ であり，$\varepsilon = 2^{-o(n)}$ のとき，十分大きなすべての n について $\rho(\mu)^l \le 2^{-n}\varepsilon^l/8$ となる．□

以降では μ を $\rho(\mu) = 2^{-\Omega(n)}$ である μ に限定し，l を $2^n\rho(\mu)^l < \varepsilon^l/8$ が成り立つ定数 l に固定して議論を進める．

補題 3.6 が示すように $\bar{S}_l(\alpha)$ は $E_l(\alpha)$ に近い．それに加えて $S_l^f(\alpha,\varepsilon)$ も $E_l(\alpha)$ に近くすることができれば，$\bar{S}_l^f(\alpha,\varepsilon)$ と $S_l(\alpha)$ も近いといえる．そこで EKM^f(空列, ε) の実行で呼び出された $\bar{S}_l^f(\alpha,\varepsilon)$ の結果が $|\bar{S}_l^f(\alpha,\varepsilon) - E(\alpha)| \le \varepsilon^l/8$ を満たしたとき，EKM^f(空列, ε) は「α で成功した」といい，文字列の集合 S に対して，S のすべての要素で成功したという意味で EKM^f(空列, ε) は「S で成功した」ということにする．EKM^f(空列, ε) の実行において $\mathrm{EKM}^f(\alpha,\varepsilon)$ が呼び出されたような α の集合を VISIT とおく．空列もこの集合に含めることとする．$\alpha \in$ VISIT のうち $\varepsilon^l/4 \le \bar{S}_l^f(\alpha,\varepsilon)$ が成り立ったものの集合を SURVIVE とおく．さらに $k = 0,1,\ldots,n$ に対して，$\mathrm{VISIT}_k := \mathrm{VISIT} \cap \mathbb{F}_2^k, \mathrm{SURVIVE}_k := \mathrm{SURVIVE} \cap \mathbb{F}_2^k$ とし，$\mathrm{VISIT}_{\le k} := \bigcup_{i \le k} \mathrm{VISIT}_i, \mathrm{SURVIVE}_{\le k} := \bigcup_{i \le k} \mathrm{SURVIVE}_i$ とする．

次の補題はアルゴリズムの成功確率を保証するものである．

補題 3.7. 定数確率以上で EKM^f(空列, ε) は VISIT で成功する．

この補題を示すための準備として，先に主張 3.8, 主張 3.9 および主張 3.10 を示す．

主張 3.8. α を EKM(空列, ε) が成功した任意の文字列とすると，$\varepsilon^l \le S_l(\alpha) \implies \varepsilon^l/2 \le \bar{S}_l(\alpha) \implies \varepsilon^l/4 \le S_l(\alpha)$.

証明. 補題 3.6 より $|\bar{S}_l^f(\alpha,\varepsilon) - S_l(\alpha)| \le |\bar{S}_l^f(\alpha,\varepsilon) - E_l(\alpha)| + |S_l(\alpha) - E_l(\alpha)| \le \varepsilon^l/4$ から成り立つ．□

また，$A_k(\theta) := \{\alpha \in \mathbb{F}_2^k \mid \theta \le S_l(\alpha)\}$ とおくと次の主張が成り立つ．

主張 3.9. 任意の $k \in [n]$ について，$\mathrm{VISIT}_{\le k}$ で成功したとき $A_k(\varepsilon^l) \subseteq \mathrm{SURVIVE}_k \subseteq A_k(\varepsilon^l/4)$.

証明. k についての帰納法で示す．$k = 0$ のとき，$\mathrm{VISIT}_{\le k} =$ { 空列 } で成功したとすると，主張 3.8 より主張の包含関係が成り立つ．次に $k = k' - 1 < n$ で主張が成り立つと仮定する．すると $k = k'$ について $\mathrm{VISIT}_{\le k}$ で成功するとき，$A_{k-1}(\varepsilon^l) \subseteq \mathrm{SURVIVE}_{k-1} \subseteq A_{k-1}(\varepsilon^l/4)$ が成り立つ．$\alpha \in A_k(\varepsilon^l)$ を任意に取ると，$S_l(\cdot)$ の単調性より α の長さ $|\alpha| - 1$ の接頭辞 α' について $\alpha' \in A_{k-1}(\varepsilon^l) \subseteq \mathrm{SURVIVE}_k \subseteq \mathrm{VISIT}_k$ となる．したがっ

て α で成功するので主張 3.8 より $\alpha \in \mathrm{SURVIVE}_k$ となる．つまり $A_k(\varepsilon^l) \subseteq \mathrm{SURVIVE}_k$ である．今度は $\alpha \in \mathrm{SURVIVE}_k \subseteq \mathrm{VISIT}_k$ を任意に取ると α で成功するので主張 3.8 より $\alpha \in A_k(\varepsilon^l/4)$ となる． □

主張 3.10. $|A_k(\theta)| = O(1/\theta)$.

証明. $Y \sim \mu^{(l)}$ とする．$|A_k(\theta)|\,\theta$ は

$$\sum_{\alpha \in A_k(\theta)} S(\alpha) \le \sum_{\alpha \in \{0,1\}^k} \left\langle f^{[\mu,l]}, \sum_{\alpha'} \chi_{\alpha\alpha'} \right\rangle_{\mu^{(l)}}$$
$$= \left\langle f^{[\mu,l]}, \sum_{a \in \{0,1\}^n} \chi_a \right\rangle_{\mu^{(l)}} = 2^n \left\langle f^{[\mu,l]}, \delta_n \right\rangle_{\mu^{(l)}}$$
$$= 2^n \Pr[Y = 0] f^{[\mu,l]}(0) \le 2^n \Pr[Y = 0]$$

で抑えられ，右辺は補題 3.6 の証明の後半と同様の議論により $1 + \varepsilon^l/4 = O(1)$ で抑えられる． □

以上を用いて補題を示そう．

補題 3.7 の証明. Höffding の不等式を用いると，$\bar{S}_l^f(\cdot,\varepsilon)$ が失敗する確率は $2\exp\{-\varepsilon^{2l}m/32\} =: q$ で抑えられる．各 $k \in \{0,\dots,n\}$ について，VISIT_k で成功するのは $|\mathrm{VISIT}_k|$ 回実行される $\bar{S}_l^f(\cdot,\varepsilon)$ がすべて成功するときである．和集合上界を用いれば VISIT_k で失敗する確率は $|\mathrm{VISIT}_k|q$ で上から抑えられる．$\mathrm{VISIT}_{\le k-1}$ で成功したと仮定する．このとき主張 3.9 より $\mathrm{SURVIVE}_{k-1} \subseteq A_{k-1}(\varepsilon^l/4)$ である．$\mathrm{VISIT}_k = \{\alpha 0, \alpha 1 \mid \alpha \in \mathrm{SURVIVE}_{k-1}\}$ であることと主張 3.10 により $|\mathrm{VISIT}_k| = 2|\mathrm{SURVIVE}_{k-1}| \le 2|A_{k-1}(\varepsilon^l/4)| = O(1/\varepsilon^l)$ となる．以上から，$\Pr[\text{VISIT で成功}] = \Pr[\mathrm{VISIT}_0 \text{で成功}] \prod_{i=1}^n \Pr[\mathrm{VISIT}_i \text{で成功} \mid \mathrm{VISIT}_{i-1} \text{で成功}] \ge (1 - O(1/\varepsilon^l)q)^{n+1} \ge 1 - O(n/\varepsilon^l)q$ が成り立つ．右辺は $m = O(\varepsilon^{-2l}\ln(n/\varepsilon))$ を満たす十分大きな m により任意の定数確率以上にすることが出来る． □

定理 3.4 の証明. 補題 3.7 より定数確率以上で $\mathrm{EKM}^f(\text{空列},\varepsilon)$ は VISIT で成功する．以下では $\mathrm{EKM}^f(\text{空列},\varepsilon)$ が VISIT で成功したと仮定する．L は $\mathrm{SURVIVE}_n$ に等しい．主張 3.9 より $A_n(\varepsilon^l) \subseteq L$. $\varepsilon \le \langle f, \chi_a \rangle_\mu$ を満たすすべての $a \in \mathbb{F}_2^n$ について $a \in A_n(\varepsilon^l)$ であるから $a \in L$ である．リストサイズについては，主張 3.9 より，$L \subseteq A_n(\varepsilon^l/4)$ であり，主張 3.10 を用いれば $|L| \le |A_n(\varepsilon^l/4)| = O(1/\varepsilon^l)$. 最後にクエリー数であるが，主張 3.9 より，$k = 0,\dots,n$ について $\mathrm{SURVIVE}_k \subseteq A_k(\varepsilon^l/4)$ である．主張 3.10 から，$|\mathrm{VISIT}_k| = 2|\mathrm{SURVIVE}_{k-1}| \le 2\mathrm{SURVIVE}_{k-1} = O(1/\varepsilon^l)$ であり，$|\mathrm{VISIT}| \le (n+1)O(1/\varepsilon^l) = O(n/\varepsilon^l)$ となるから，$\bar{S}_l^f(\cdot,\varepsilon)$ が呼ばれる回数もこの量で上から抑えられる．$\bar{S}_l^f(\cdot,\varepsilon)$ 内部では $ml = O(\varepsilon^{-2l}\ln(n/\varepsilon))$ 回クエリーが行われるから，全体で行われる f へのクエリー数は $O(n\varepsilon^{-3l}\ln(n/\varepsilon))$ 回以下であることがわかる． □

参考文献

[1] O. Goldreich and L.A. Levin. A hard-core predicate for all one-way functions. In *Proc. STOC '89*, pp. 25–32, 1989.

[2] T. Kaufman and M. Sudan. Sparse random linear codes are locally decodable and testable. In *Proc. FOCS '07*, pp. 590–600, 2007.

[3] S. Kopparty and S. Saraf. Local list-decoding and testing of random linear codes from high error. In *Proc. STOC '10*, pp. 417–426, 2010.

[4] E. Kushilevitz and Y. Mansour. Learning decision trees using the Fourier spectrum. *SIAM J. Comput.* 22(6): 1331–1348, 1993.

[5] M. Sudan. List decoding: Algorithms and applications. *SIGACT News* 31:16–27, 2000.

[6] A. Blum, A. Kalai, and H. Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. In *Proc. of STOC '00*, pp. 435–440, 2000.