研究紹介

Expander 符号

安永憲司

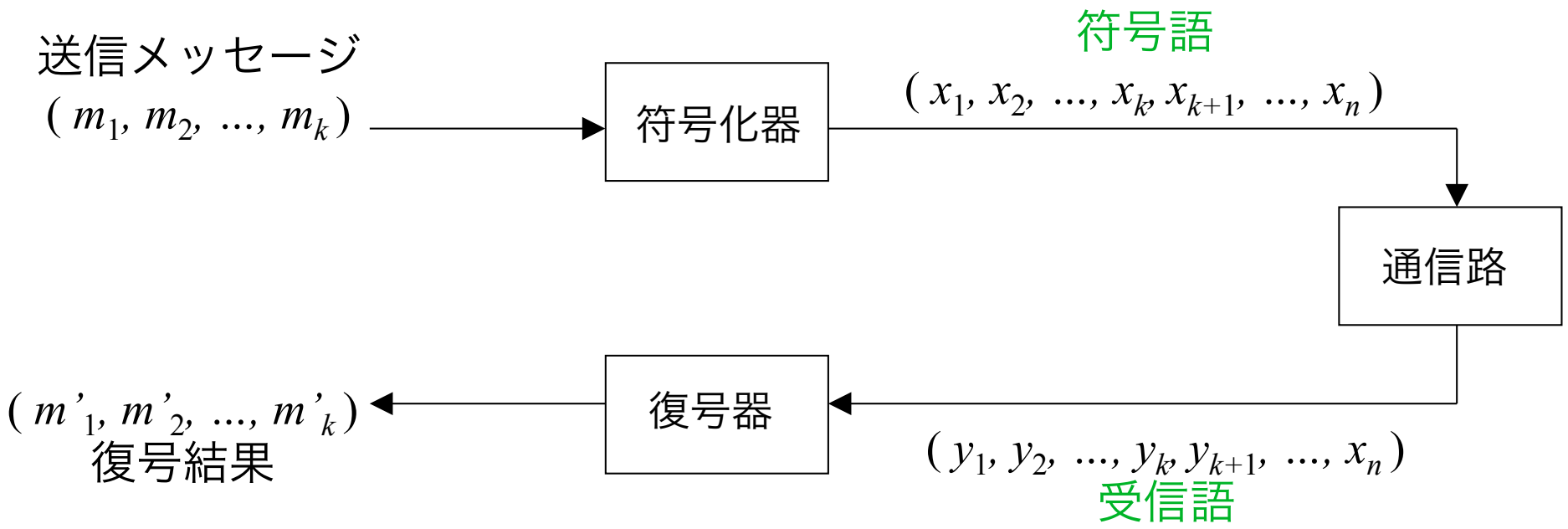
2008年4月某日

目次

- 線形符号
  - 用語の定義
  - 生成行列、パリティ検査行列、Tanner グラフ
- expander 符号
  - expander グラフ
  - [Sipser-Spieleman ‘96]
    - expander 符号の構成
    - bit-flipping 復号法
- まとめ
- 参考文献

誤り訂正符号

- 送信メッセージに冗長性をもたせることで
  通信路における誤りを訂正することができる



線形符号

- 符号 C とは
  - 符号化関数 $Enc : \{0, 1\}^k \to \{0, 1\}^n$
    - n : 符号長、k : 情報記号数（メッセージ長）
    - Enc は単射
  - ベクトルの集合 $S \subseteq \{0, 1\}^n$
    - $k = \log_2 |S|$

- C が線形符号 ⇔ Enc が線形関数 ⇔ S が線形空間

- C の最小距離 $d = \min_{\substack{c_1, c_2 \in C \\ c_1 \neq c_2}} d_H(c_1, c_2) = \underbrace{\min_{c \in C} w_H(c)}_{\text{線形符号の場合}}$
  - C の相対最小距離 $\delta = d/n$

- C のレート $R = k/n$

どんな符号を作りたいか

- 最小距離が大きい
  - 誤り訂正能力が高くなる
- レートが大きい
  - なるべく冗長性を小さく
- 符号化の計算量が小さい
  - 線形符号なら $O(n^2)$
- 復号の計算量が小さい
- いろんな n, k をとることができる
  - 符号族 $C = \{C_1, C_2, \cdots\}$;
    - C_i は符号長 n_i の符号であり $n_i < n_j$ for $i < j$

線形符号の例

例1. $C_1 = \{00000, 11100, 00111, 11011\}$

- $\forall \boldsymbol{x}_1, \boldsymbol{x}_2 \in C_1, \boldsymbol{x}_1 + \boldsymbol{x}_2 \in C_1$ なので C_1 は線形符号
- $n = 5,\ k = \log_2|C_1| = 2,\ R = 2/5,\ d = 3$

例2. $C_2 = \{00..0, 11..1\}$

- $k = \log_2|C_2| = 1,\ R = 1/n,\ d = n$

例3. $E_1(\boldsymbol{x}) = \boldsymbol{x} \circ \boldsymbol{x} \circ \boldsymbol{x}$; ∘：ベクトルの連接

- $R = 1/3,\ d = 3$

生成行列・パリティ検査行列

- 線形符号 C は生成行列 G もしくはパリティ検査行列 H で定義される

- G : $k \times n$ 行列;
  $\boldsymbol{x}$ が C の符号語 $\Leftrightarrow \boldsymbol{x} = \boldsymbol{m}G$ for some $\boldsymbol{m} \in \{0,1\}^k$

- H : $(n-k) \times n$ 行列;
  $\boldsymbol{x}$ が C の符号語 $\Leftrightarrow H\boldsymbol{x}^T = 0$

生成行列・パリティ検査行列の例

$C_1 = \{00000, 11100, 00111, 11011\}$

- $G_1 = \begin{pmatrix} 11100 \\ 00111 \end{pmatrix}$
- $H_1 = \begin{pmatrix} 10101 \\ 00011 \\ 11000 \end{pmatrix}$
  - $\boldsymbol{x} = (x_1, x_2, .., x_5) \in C_1 \Leftrightarrow$ $x_1+x_3+x_5=0$, $x_4+x_5=0$, $x_1+x_2=0$ を満たす

これが線形制約なので
線形符号であるともいえる

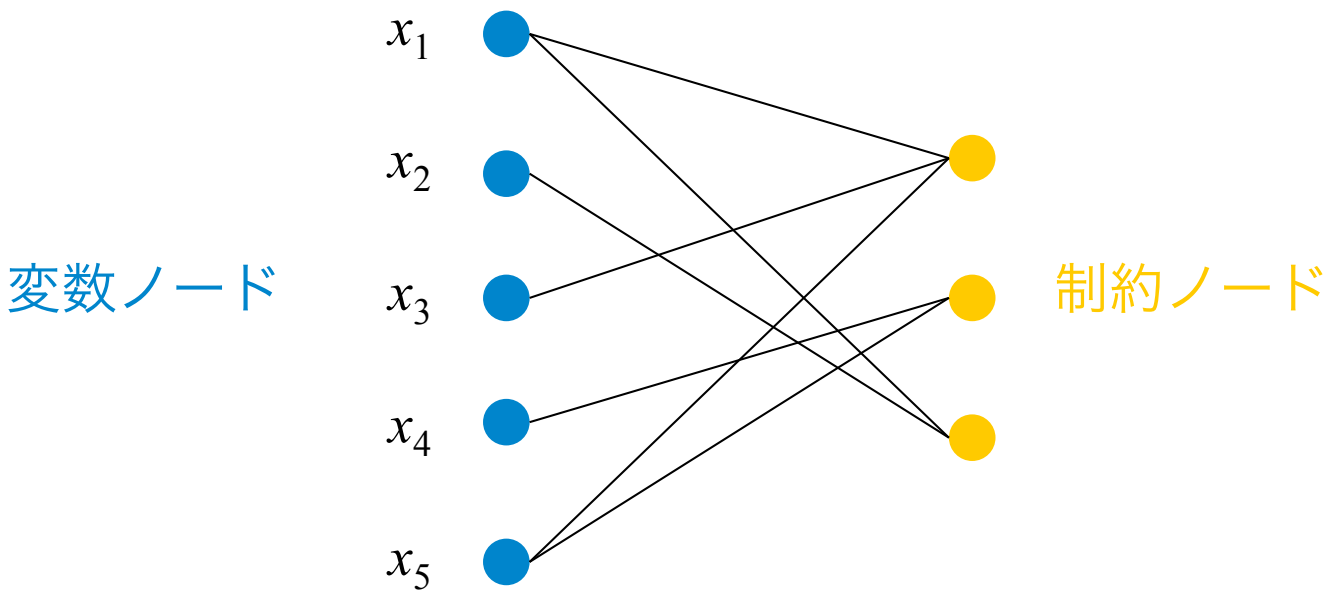
Tanner グラフ

- パリティ検査行列を２部グラフで表現したもの[Tanner ‘84]

- Tanner グラフ $G := (L, R, E)$,
  - L : 左頂点集合, 符号語の各ビットに対応, $|L| = n$
  - R : 右頂点集合, 線形制約に対応, $|R| = n - k$
  - E : 辺集合, 検査行列の１の要素に対応
  - 符号語
    ⇔ すべての右頂点において、値１をもつ隣接する
    左頂点の数が偶数

Tanner グラフの例

- $C_1 = \{00000, 11100, 00111, 11011\}$
- $H_1 = \begin{pmatrix} 10101 \\ 00011 \\ 11000 \end{pmatrix}$



LDPC 符号

- 低密度パリティ検査(low density parity check)符号
- パリティ検査行列で１が少ない符号
  ＝ Tanner グラフで枝数が少ない符号
  - Tanner グラフで左頂点の次数が定数のものなど
- [Gallager ‘63] で導入
- 線形時間復号法（belief propagation, sum product）で非常にすぐれた誤り特性を示すことから、ここ10年非常に研究が活発

Tanner グラフで最小距離を大きくするには？

- 最小距離が d
  ⇔ Tanner グラフにおいて、サイズ $d-1$ 以下の任意の左頂点集合 S に対し、S と奇数本接続している右頂点が存在

- ここで、各 S は１つの右頂点に３本以上接続しないとする

- このとき、最小距離を大きくするには、各 S に対し１本しか接続しない右頂点(unique neighbor)が存在すればよい

  ⇒　あるサイズ以下のすべての左頂点集合 S が、多くの右頂点に接続しているグラフであればよい（左頂点次数が低い場合）

  ⇒　expander グラフの利用 [Sipser-Spielman ‘96]

expander グラフ

- グラフ中の枝数が少ないにもかかわらず、各ノードが非常によい連結性を持つグラフ
  - expander の定義の仕方は何通りもある
  - ランダムに構成したグラフは expander の性質を満たす

- グラフ $G = (V, E)$ が (α, β) expander
  $\Leftrightarrow$ すべての $S \subseteq V$ s.t. $|S| \leqq \alpha |V|$ に対して $|\Gamma(S)| \geqq \beta |S|$
  - $\Gamma(S)$: 頂点集合 S 内の頂点と隣接する頂点の集合

- ２部グラフ $G = (L, R, E)$ が (α, β) expander
  $\Leftrightarrow$ すべての $S \subseteq L$ s.t. $|S| \leqq \alpha |L|$ に対して $|\Gamma(S)| \geqq \beta |S|$

- 左頂点次数が小さく（ある定数 l ）、β が大きい（l に近い）expander グラフの構成が難しい問題

expander 符号のレートと最小距離

- 定理 1 : $G = (L, R, E)$ s.t. $|L| = n$, $|R| = m$ が左頂点 l 正則の (α, β) expander であり, $\beta > l/2$ ならば, それを Tanner グラフとする符号は, レート $R \geqq 1 - m/n$, 最小距離 $d \geqq 2\alpha\beta n/l$ $(> \alpha n)$
  - $R \geqq 1 - m/n$ は明らか
  - $U(S)$: $S \subseteq L$ と 1 本だけ接続している右頂点集合
  - すべての $S \subseteq L$ s.t. $|S| \leqq \alpha n$ に対し $|U(S)| > 0$ である
    - $V(S)$: S と 2 本以上接続している右頂点集合
    - $|U(S)| + |V(S)| = |\Gamma(S)| \geqq \beta|S|$ であり $|U(S)| + 2|V(S)| \leqq l|S|$ であるので $\Rightarrow (2\beta - l)|S| \leqq |U(S)|$
  - すべての $T \subseteq L$ s.t. $|T| < 2\alpha\beta n/l$ に対し $|U(T)| > 0$ である
    - $|T| > \alpha n$ とする
    - $S \subseteq T$ s.t. $|S| = \alpha n$ とすると、$|U(S)| \geqq (2\beta - l)|S| = (2\beta - l)\alpha n$
    - $|T \setminus S| < 2\alpha\beta n/l - \alpha n = (2\beta/l - 1)\alpha n$ なので $|\Gamma(T \setminus S)| < (2\beta - l)\alpha n$
    - したがって $|U(S)| > |\Gamma(T \setminus S)| > 0$ となり $|U(T)| > 0$

bit-flipping 復号法

- 誤りベクトルに対応する左頂点集合を S_e とする
- expander グラフなので $|U(S_e)|$ は大きいと予想される
- $U(S_e)$ に含まれる右頂点は、充足されてない（奇数本の 1 と接続）
- アルゴリズム : 充足している右頂点より充足していない右頂点と多く接続している、そんな左頂点が存在するとき、その頂点の値の 0 と 1 をひっくり返す(bit-flipping)

復号法の分析

- 定理 2 : $\beta > (3/4)l$ のとき bit-flipping 復号法は $\alpha n/2$ 個未満の誤りを線形時間で訂正する

  - 以下を証明する
    補題 1 : $|S_e| < \alpha n$ ならば flip すべき左頂点が存在
    補題 2 : $|S_e| < (2\beta - l)\alpha n/l$ ならば、復号を繰り返す間、誤りビットの数は $(2\beta - l)\alpha n/l$ 以上にならない

  - 復号の 1 ステップで、充足している右頂点の数は必ず増えるので、正しい符号語へ復号する

  - 右頂点数は $n - k$ なので $O(n)$ ステップで終了、各ステップは $O(1)$ で可能なので、$O(n)$ 時間で復号可能

- 補題 1 : $|S_e| < \alpha n$ ならば充足している右頂点より充足していない右頂点と多く接続している左頂点が存在
  - $|S_e| < \alpha n$ なので $|U(S_e)| \geqq (2\beta - l)|S_e|$
  - S_e 中の 1 つの頂点あたり $2\beta - l$ 個が $U(S_e)$ に含まれる
  - $2\beta - l > l/2$ なので少なくとも一つの左頂点で、充足していない右頂点の方が多く接続している

- $S_e(i)$ を i ステップ目における誤りの頂点集合とする
- 補題2 : $|S_e(0)| < (2\beta - l)\alpha n/l$ ならばすべての i において $|S_e(i)| < (2\beta - l)\alpha n/l$ である
  - $|S_e(0)| < \alpha n/2 < (2\beta - l)\alpha n/l$ である
  - 初めの段階で充足されてない右頂点は $l|S_e(0)| < (2\beta - l)\alpha n$ 個以下
  - ある t において $|S_e(t)| \geqq (2\beta - l)\alpha n/l$ だと仮定すると、$|S_e(t)| > \alpha n/2$
  - $S_e(i)$ は 1 つずつ変化するので、ある t' ステップ目において $|S_e(t')| = \alpha n/2$ である
  - このとき $|U(S_e(t'))| \geqq (2\beta - l)|S_e(t')| = (2\beta - l)\alpha n$ となり、これは充足されてない右頂点数が増えたことになり、矛盾

expander グラフの構成法

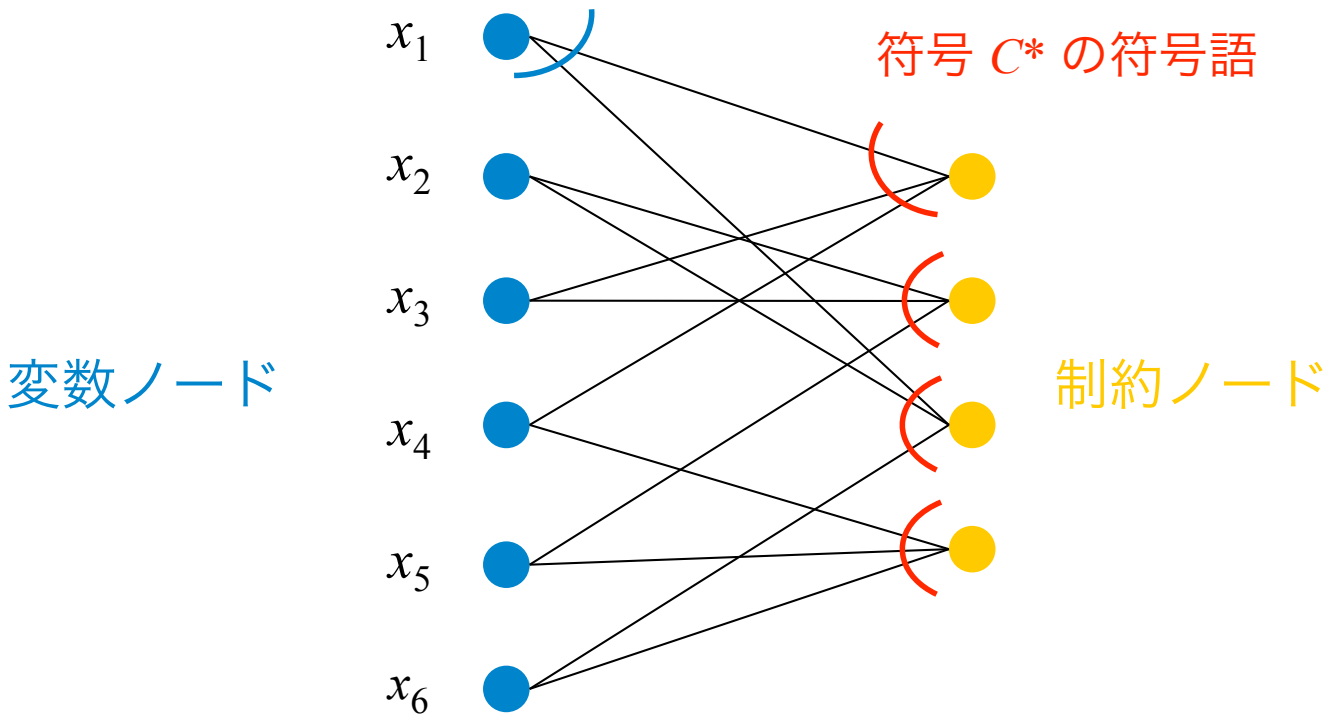
$G = (L, R, E)$ with $|L| = n$, $|R| = m$

- [Capalbo et al. '02]
  - $n/m = O(1)$ のとき、任意の $\varepsilon > 0$ に対して、
    $l = O(1)$ の $(O(1), 1 - \varepsilon)$ expander グラフの構成法

- [Guruswami et al. '07]
  - 任意の $\gamma > 0$, n, $\alpha<1$, $\varepsilon > 0$ に対して,
    $l = O((\log n)(\log \alpha n)/\varepsilon)^{1+1/\gamma}$, $m \leqq l^2(\alpha n)^{1+\gamma}$ の
    $(\alpha, 1 - \varepsilon)$ expander グラフの構成法

Tanner 符号による最小距離の改善

- Tanner 符号 [Tanner '84]
  - 変数ノードも制約ノードも定数次数である Tanner グラフ
  - 制約ノードに入る辺がある符号 C^* の符号語になっている



Tanner 符号による最小距離の改善

- 定理 3 : $G = (L, R, E)$ s.t. $|L| = n$, $|R| = m$ が左頂点 l 正則・右頂点 r 正則の (α, β) expander であり、C^* が $(r, r - t, \Delta)$ 符号であるとする。このとき, $\beta > l/\Delta$ ならば, それを Tanner グラフとする符号は, レート $R \geqq 1 - (m/n)t$, 最小距離 $d > \alpha n$

  - 制約の数が mt 個あるので $R \geqq 1 - (m/n)t$
  - $U'(S)$: $S \subseteq L$ と Δ 本未満接続している右頂点集合
  - $V'(S)$: $S \subseteq L$ と Δ 本以上接続している右頂点集合
  - すべての $S \subseteq L$ s.t. $|S| \leqq \alpha n$ に対し $|U'(S)| > 0$ である
    - $|U'(S)| + |V'(S)| = |\Gamma(S)| \geqq \beta|S|$ であり $|U'(S)| + \Delta|V'(S)| \leqq l|S|$ なので
      $\Rightarrow (\Delta\beta - l)|S|/(\Delta - 1) \leqq |U(S)|$

まとめ

- Tanner グラフとして expander グラフを利用した符号 [Sipser-Spielman '96] の最小距離の分析・復号法を紹介
  - 最小距離の下界の証明には $\beta > l/2$ が必要
  - bit-flipping 復号法が最小距離の半分まで訂正できるには $\beta > (3/4)l$ が必要
  - その後の研究で、上記の β を満たす expander グラフの構成法が示された
  - Tanner 符号を考えるとレートは下がるが必要な β が小さくなる

参考文献

[Capalbo et al. ‘02] Capalbo, Reingold, Vadhan, Wigderson, “Randomness conductors and constant-degree lossless expanders,” in *Proc. the 34th ACM Symposium on Theory of Computing (STOC 2002)* .

[Gallager ‘63] Gallager, “*Low Density Parity Check Codes*,” MIT Press, 1963.

[Guruswami et al. ‘07] Guruswami, Uman, Vadhan, “Unbalanced expanders and randomness extractors from Parvaresh-Vardy codes,” in *Proc. IEEE Conference on Computational Complexity,* 2007.

[Sipser-Spielman ‘96] Sipser, Spielman, “Expander codes,” *IEEE Trans. on Information Theory*, vol. 42, no. 6, 1996.

[Tanner ‘84] Tanner, “A recursive approach to low complexity codes,” *IEEE Trans. on Information Theory*, vol. 27, no. 5, 1984.

- 今回の資料はおもに以下を参考にして作成
  - Madhu Sudan’s course notes of “Essential coding theory” at MIT, 2001, 2002, and 2004. Available from http://people.csail.mit.edu/madhu/teaching.html
  - Venkatesan Guruswami’s course notes of “Error-Correcting Codes: Constructions and Algorithms” at University of Washington, 2006, and “Codes and Pseudorandom Objects” at University of Washington, 2003.
    Available from http://www.cs.washington.edu/homes/venkat/#teaching