情報理論第二（10）

ハミング符号
秘密鍵暗号

人間コミュニケーション学科
梶本　裕之
kajimoto@hc.uec.ac.jp

授業の流れ（予定）

第１週（10/ 8） 情報量と情報エントロピー
第２週（10/15） 相互情報量
第３週（10/22） 体育祭のため休講
第４週（10/29） 情報源符号化とデータ圧縮
第５週（11/ 5） ハフマン符号とデータ圧縮
第６週（11/12） 情報源符号化定理
第７週（11/19） 調布祭のため休講
第８週（11/26） 出張のため休講
第９週（12/ 3） マルコフ情報源モデル
第10週（12/10） 通信路のモデル化
第11週（12/17） 誤り検出・誤り訂正符号
第12週（12/24） 出張のため休講
第13週（ 1/ 7） 線形符号
第14週（ 1/14） ハミング符号
第15週（ 1/21） 秘密鍵暗号
第16週（ 1/28） 公開鍵暗号
第17週（ 2/ 4） 出張のため休講

復習：線形符号

- 元の語ｘ：ｋビット
- 符号語ｙ：ｎビット

ある適当な生成行列G（ｎ行ｋ列）を用いて

$$y = Gx$$

と符号化する方法を線形符号という

復習：生成行列の標準形

- 生成行列は，適当な変形により，
  以下のような**標準形**に変形できる

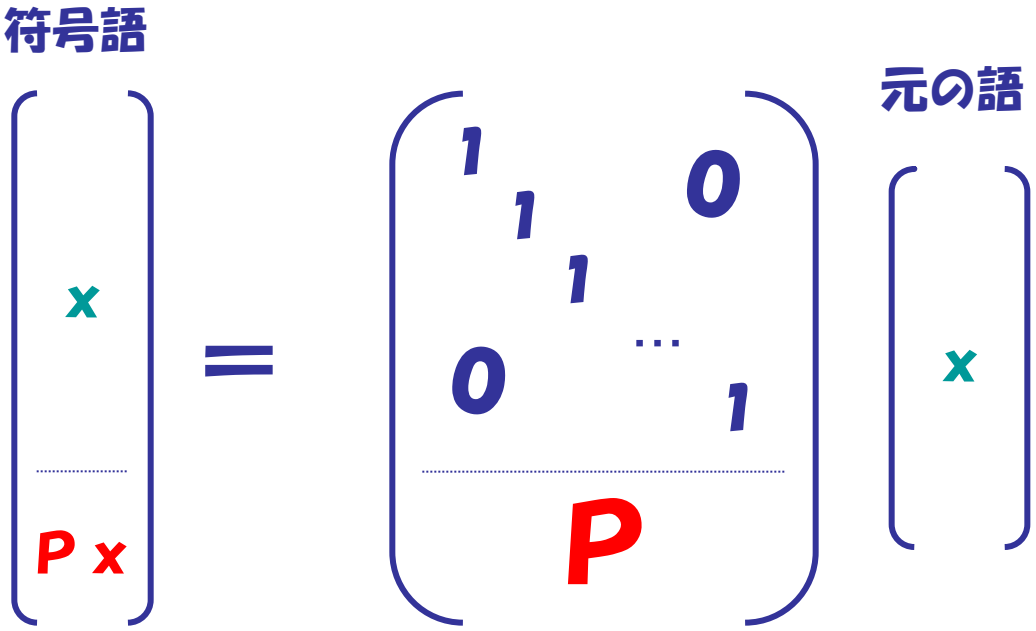
$$G = \left(\begin{array}{c|c} \begin{matrix} 1 & & & 0 \\ & 1 & & \\ & & 1 & \\ 0 & & \cdots & 1 \end{matrix} & \\ \hline & P \end{array}\right)$$

k 行

n-k 行

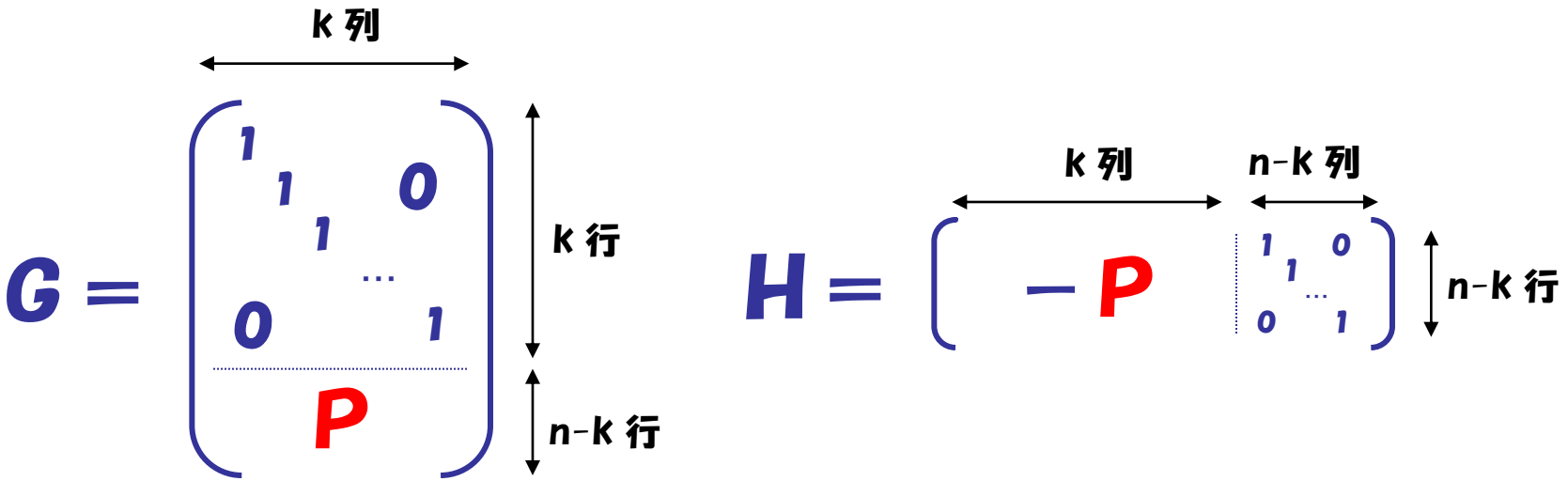
復習：標準形の生成行列による符号化

- 標準形をした生成行列:
- 元の語と付加されたパリティ部分が
  きれいに分かれた符号語を生成できる！



復習：検査行列

- 生成行列Gに対して，
  検査行列Hをつくることができる



ただし2進数の場合 −P=P.

復習：検査行列の性質

- 生成行列Gによって作られた符号語をHに適用すると，0になる！！

$$H(Gx) = (HG)x = 0x = 0$$

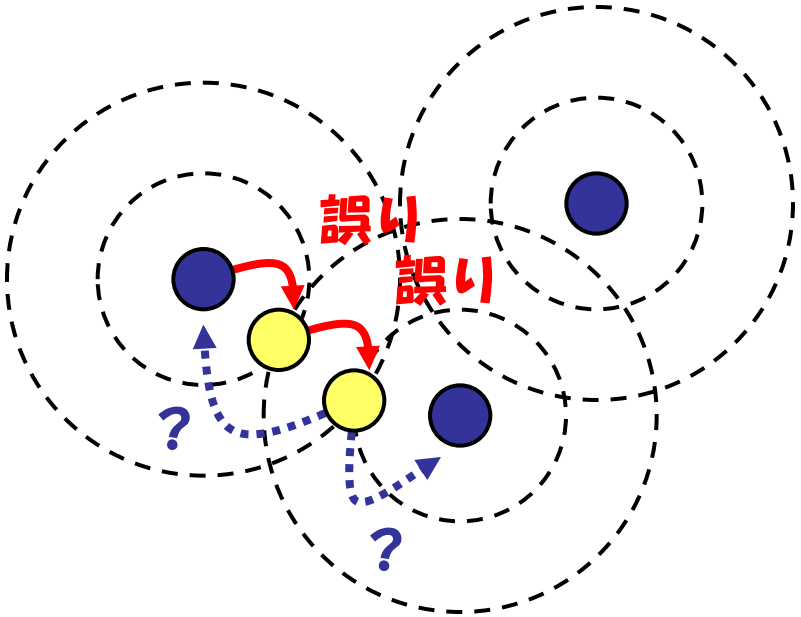
$$\begin{pmatrix} -P & \begin{matrix} 1 & 0 \\ 1 & \\ 0 & 1 \end{matrix} \end{pmatrix} \begin{pmatrix} 1 & & & \\ & 1 & & 0 \\ & & 1 & \\ 0 & & & 1 \\ \hline & & P & \end{pmatrix} \begin{pmatrix} \\ x \\ \\ \end{pmatrix} = (-P+P)x$$

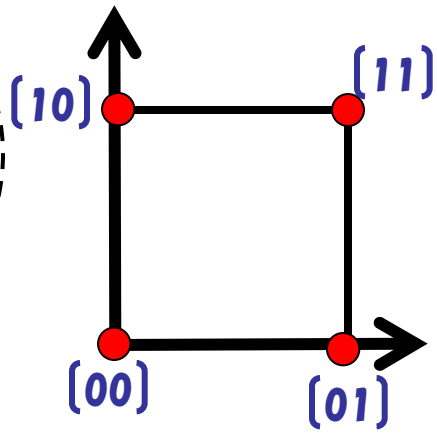
受信語にHをかけて0になるかどうかで
伝送中に誤りが生じたかどうか検査できる

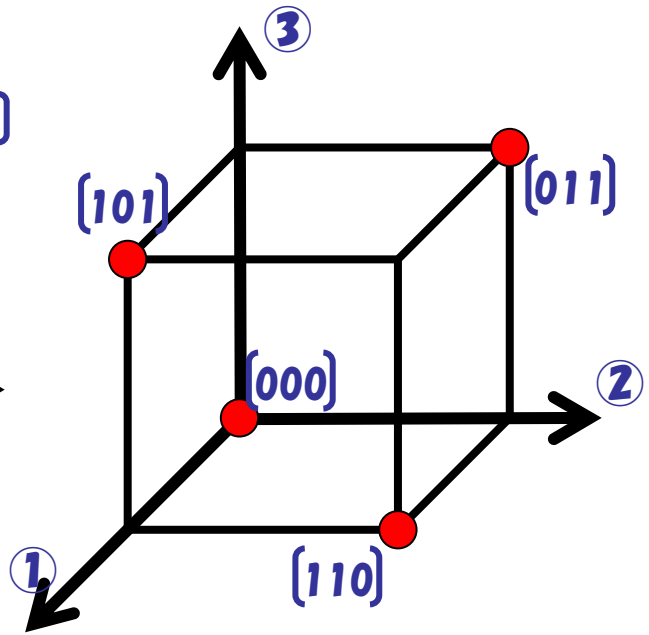
線形符号の符号語間最小距離

復習：何がしたかったのか

元の語：2bit(2次元)の情報．４通り
符号語：3bit(3次元)．8つの格子点の中の４点
⇒できた「すき間」＝誤り検出，誤り訂正のための冗長性
つまり，この「すき間」の最小の間隔が重要！！

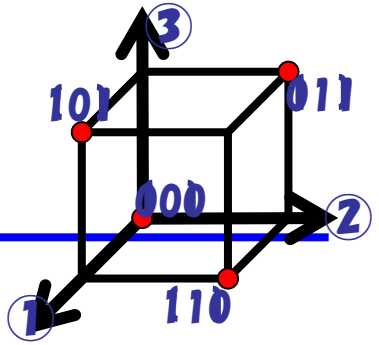






この「最小すき間」はどうしたら計算可能か？

符号語間距離＝すき間



符号語y_1とy_2の間のハミング距離 $d_H(y_1, y_2)$ を考える

- $\{0,1\}$ の有限体上では以下のようになる

$$d_H(y_1, y_2) = d_H(y_1 - y_2, 0)$$

∵各桁について考えると，

$(y_1, y_2)=(0,0), (1,1) \rightarrow d_H(0-0, 0)= d_H(1-1, 0)= d_H(0, 0)= 0$

$(y_1, y_2)=(0,1), (1,0) \rightarrow d_H(1-0, 0)= d_H(0-1, 0)= d_H(1, 0)= 1$

ハミング距離は，各桁同士で比較した時の「違うbitの個数」だから，各桁で上式が成り立っていれば当然成り立つ．

線形符号における符号語間距離

- $d_H(y_1, y_2) = d_H(y_1 - y_2, 0)$

ここで，$y_1 = Gx_1$，$y_2 = Gx_2$ とすると，

$y_1 - y_2 = G(x_1 - x_2)$　但し $x_1 - x_2 \in F_2^k$

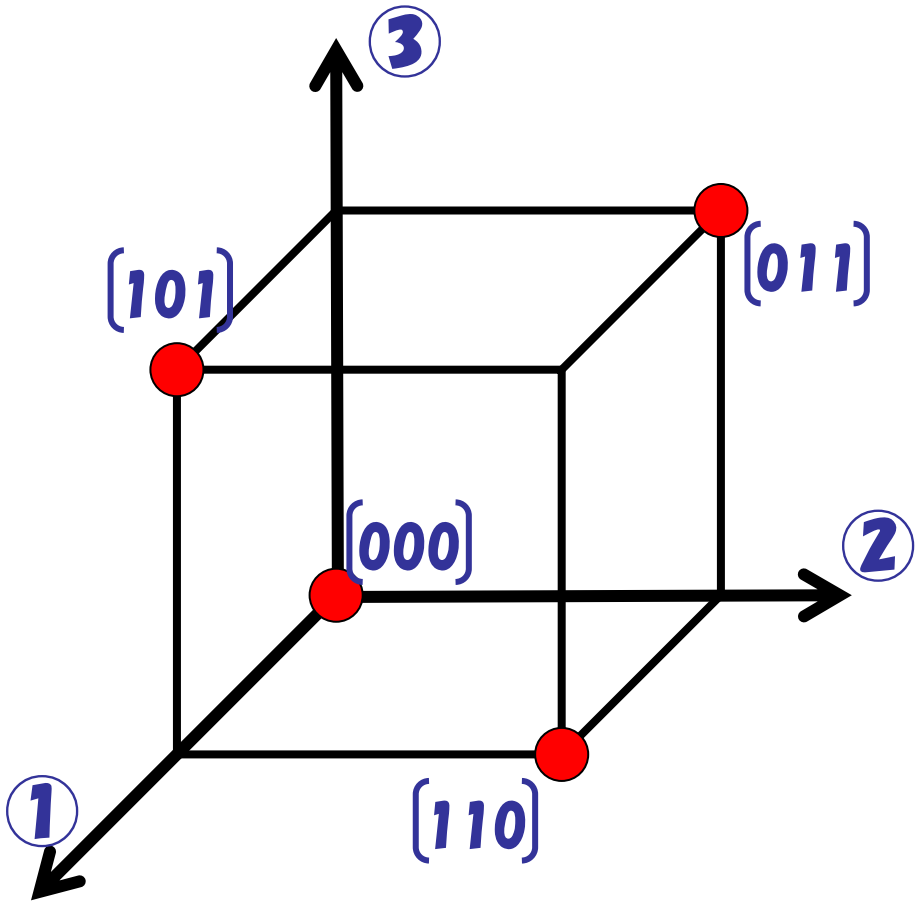
したがって $y_1 - y_2$ も，ある元の語$(x_1 - x_2)$から生成された符号語に必ずなっている．

何が言いたいか？

y_1, y_2 が符号語なら，　$y_1 - y_2$も符号語である．

線形符号における符号語間距離

y_1, y_2 が符号語なら，　$y_1 - y_2$も符号語である．



[1 0 1]-[0 1 1]=

[1 0 1]-[1 1 0]=

[1 1 0]-[0 1 1]=

線形符号における符号語間距離

- 誤り検出・誤り訂正という観点から，
  符号語間の**最小**距離 D を知りたい！！

$$D = \min d_H(y_1, y_2) = \min d_H(y_1 - y_2, 0)$$

$y_1 - y_2$ が符号語全体がつくる部分空間の中に必ず入っていることを考えれば，
$y = y_1 - y_2$ とおけて，

$$D = \min d_H(y, 0) \quad (y \neq 0)$$

線形符号における符号語間距離

$D = \min d_H(y, 0) \quad (y \neq 0)$

なにが言いたいか？

線形符号における**符号語間**最小距離は，

0ベクトルを除いた任意の符号語に含まれる要素1の個数の最小値となる

ことがわかる！

（符号語**同士**のハミング距離を求めなくても，
符号語と**原点**のハミング距離を求めれば充分）

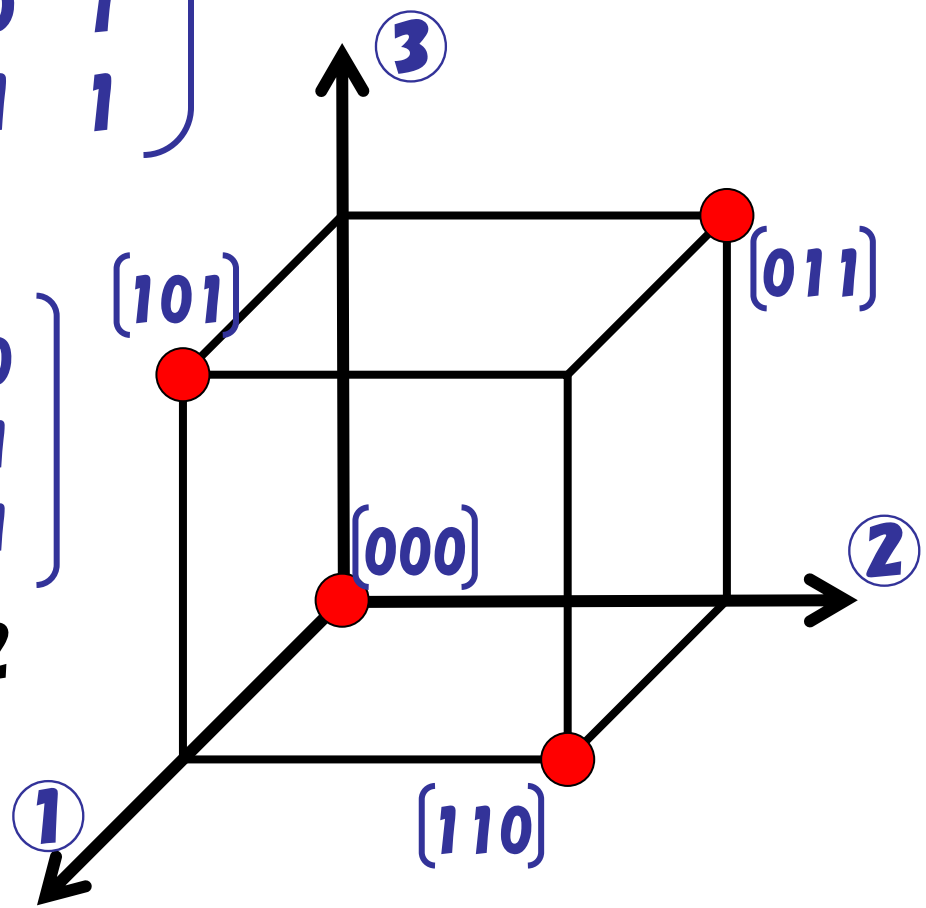
例

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$$

- 符号語は下記，

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \quad \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \quad \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

1の数　0　2　2　2



符号語間の最小距離は2である⇒あっている

例

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

- 符号語は下記，

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

1の数	0	3	3	4	3	4	4	3

符号語間の最小距離は３である
（だから２誤り検出，１誤り訂正可能）

線形符号の符号語間最小距離

実は検査行列Hを観察するだけでわかる！！

定理：

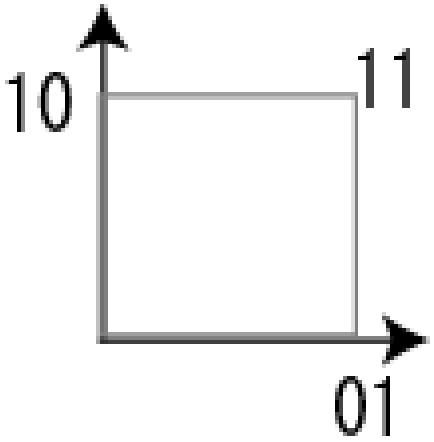
検査行列Hの列ベクトルのうち，任意のm−1個が1次独立ならば，Hによって検査される符号における符号語間最小距離はm以上である．

系：

上記に加え，更にあるm個について1次従属となるならば，最小距離はmである．

復習：一次独立

集合中の任意の要素が，**他の要素の和で書けない**こと．



- (0,1)と(1,0)は一次独立
- (0,1)と(1,0)と(1,1)は一次従属
  (0,1) + (1,0) = (1,1)

別の表現：要素の**線形和が0にならない**こと．

- a∗(0,1) + b∗(1,0)は，a,b≠0で0とならない．
- a∗(0,1) + b∗(1,0) + c∗(1,1)は，(a,b,c)=(1,1,−1)に対して0となる→一次従属．

例

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

- 下記の検査行列Hから，この符号における符号語間最小距離を求めよ．また，実際に符号語をいくつか生成し，求めた最小距離が正しいことを確認せよ．

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

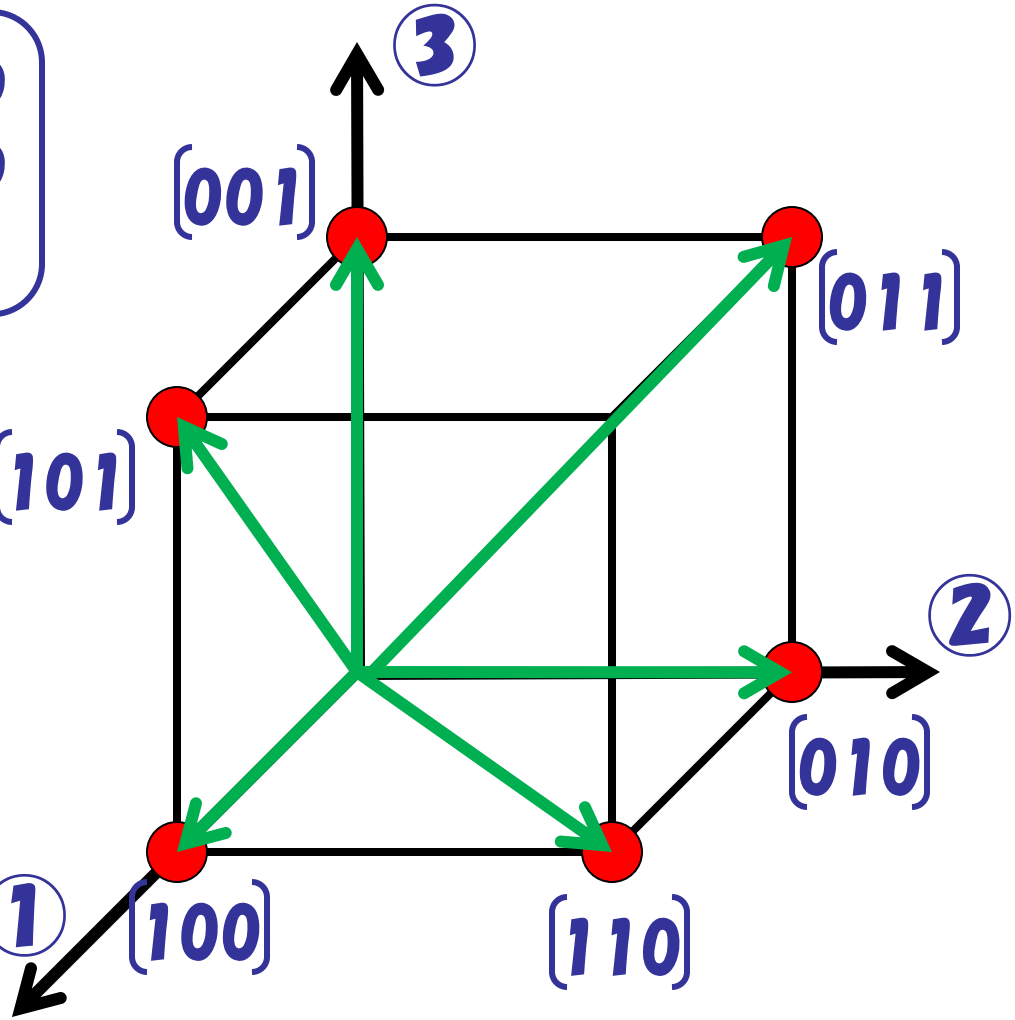
例

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

任意の二つのベクトルは
方向が重なっていないので
一次独立である.

３つだと例えば
(100)+(010)=(110)なので
一次従属

よってm=3



定理の証明

$H=(h_1\ h_2\ \ldots\ h_n)$（$h_i$ は列ベクトル）とする.
Hは検査行列だから，誤りのない符号語
$y=(y_1 y_2\ \ldots y_n)^T$について

$$h_1y_1 + h_2y_2 + \ldots + h_ny_n = 0$$

任意の $m-1$ 個の列ベクトルが1次独立
→ 0以外のyで，常に

$$h_{i1}y_{i1} + h_{i2}y_{i2} + \ldots + h_{i(m-1)}y_{i(m-1)} \neq 0$$

定理の証明

(1) $h_1 y_1 + h_2 y_2 + \cdots + h_n y_n = 0$
(2) $h_{i1} y_{i1} + h_{i2} y_{i2} + \cdots + h_{i(m-1)} y_{i(m-1)} \neq 0$

(2)の元で(1)が成り立つためには，残りのy_iのうち**少なくとも一つは1**である必要アリ

任意の $m-1$ 個のh_iに対してそうなるためには，
y_i は**m個以上1**である必要アリ．
→ 符号語間の最小距離は最低でもmとなる

線形符号における
誤り検出・誤り訂正

誤り訂正とシンドローム

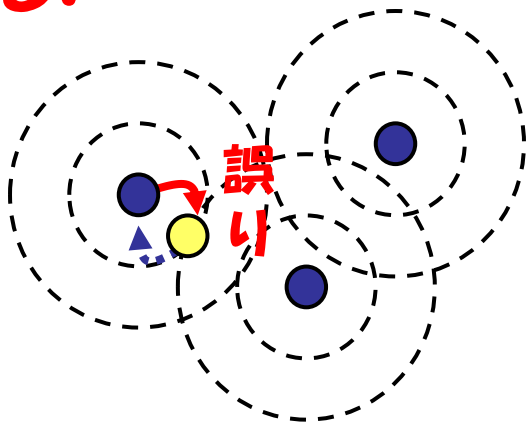
- 受信語ｙに検査行列Ｈをかけた結果Ｈｙを，受信語ｙのシンドロームという
- シンドローム＝０なら，受信語に誤りはない（Ｇによって生成された符号語が届いている）
- シンドローム≠０：どこかに誤りあり！

誤り訂正とシンドローム

- シンドローム≠0→どこかに誤りあり！
- 誤りベクトルをeとする．$y = y_{true} + e$
- $Hy = H(y_{true} + e)$
- $= Hy_{true} + He$
- $= He$
- より，シンドロームはエラーeに関する情報を，Heという形で間接的に与える．

誤り訂正とシンドローム

- シンドロームはエラーeに関する情報を，”He”という形で間接的に与える．



- ＜誤り訂正の手順＞
- (1) Ｈｅを与えるようなeを探す
- (2) そのようなeのうち，含まれる１の数が符号語間の最小距離の半分 よりも小さいものがあれば，そこに誤りが生じたとする
- (3)eが分かればyから$y_{\underline{true}}$を作れる(反転)

例題

- 受信語として $(1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 0)^T$ というベクトルを受け取った．この中に**1ビットの誤り**が含まれているものとする．下記の検査行列Hを用いて，この受信語の誤りを訂正せよ．
- （ヒント：1ビットの誤りということはeの候補は6通りだけ）

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

回答

$$He = Hy = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

エラーが1bitとすると，可能性は

$$e = \begin{pmatrix} \square \\ \square \\ \square \\ \square \\ \square \\ \square \end{pmatrix}$$

エラーの発生箇所を特定できた！！

の6通り．このときHeは

$$He = \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix} \begin{pmatrix} \square \\ \square \\ \square \end{pmatrix}$$

回答

確認（エラービットを反転させてみる）

$$Hy = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

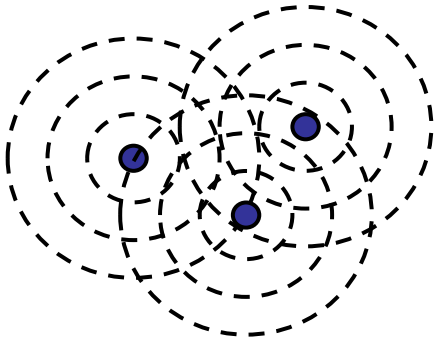
確かに検査行列によって0となった！

→正しい符号である！

ハミング符号

復習：符号語間距離

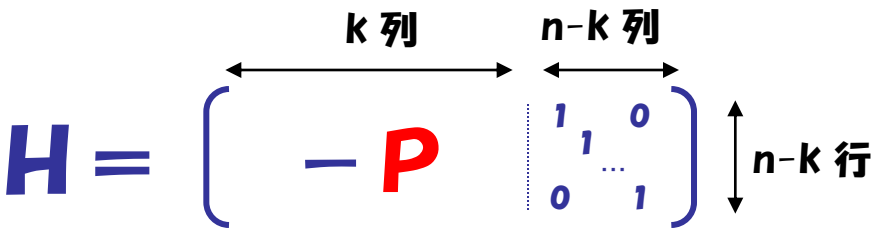
- 一般に，各符号語間のハミング距離が最低でもｋと保証されているならば，その符号は最低でも
  - ｋ より少ない誤りについて検出可能
  - k/2 より少ない誤りについて訂正可能



ある符号体系の誤り訂正能力＝符号語間の最小距離で決まる

復習：線形符号の符号語間最小距離

実は**検査行列H**をみるだけでわかる！！



定理：

検査行列Hの列ベクトルのうち，**任意の**m−1個が1次独立なら，符号語間最小距離はm以上.

系：

さらにあるm個について1次従属となるなら，符号語間最小距離はm.

（証明は前回を参照）

ハミング符号

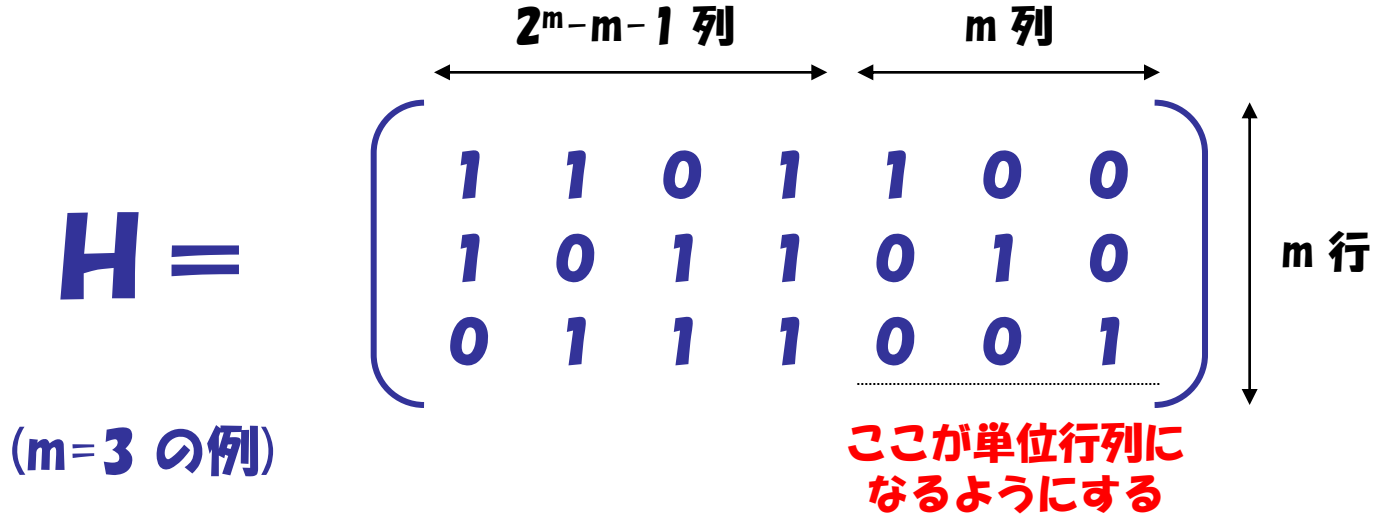
定理：検査行列Hの列ベクトルのうち，任意のm−1個が１次独立なら，符号語間最小距離はm以上.

系：さらにあるm個について１次従属となるなら，最小距離はm.

検査行列Hが持つ上記の性質を利用して，
符号語間の最小距離が３
（＝２誤り検出・１誤り訂正可能）
となる線形符号を生成する手法

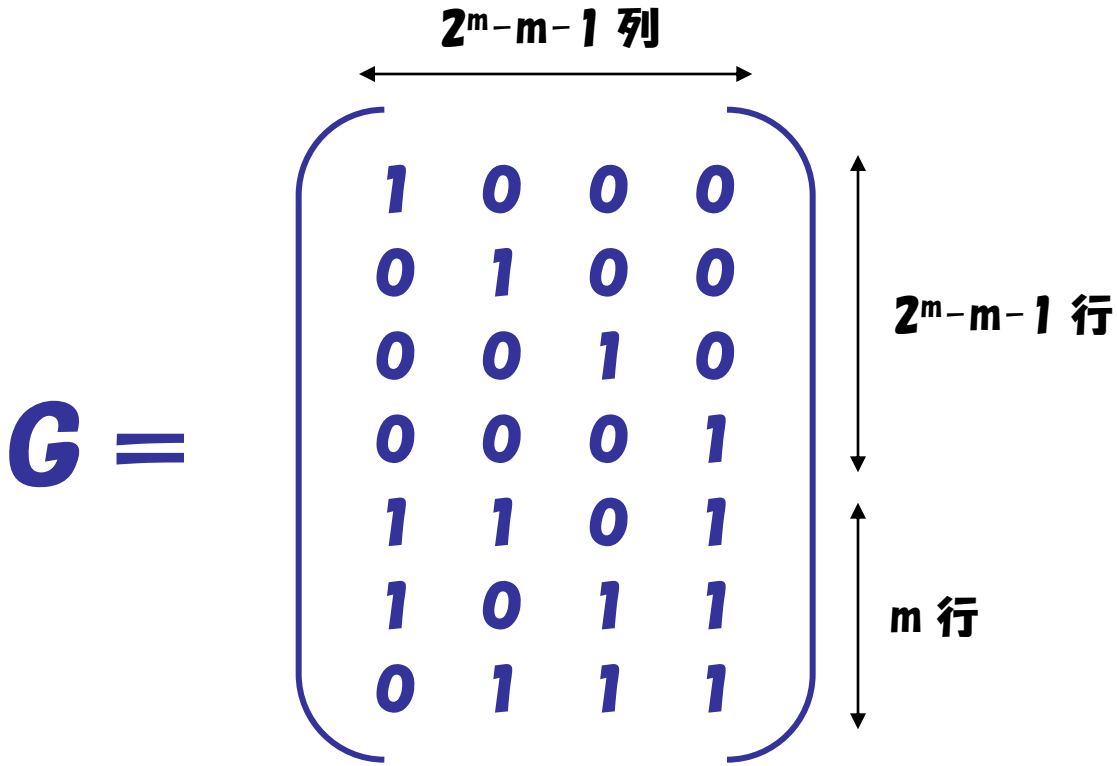
ハミング符号の作り方

- まず適当な整数 $m > 1$ を用意する
- mビットの列ベクトルを，0ベクトルを除いて全てつくり，それを横に適当に並べて検査行列Hを作る



ハミング符号の作り方

- 検査行列Hから，対応する生成行列Gをつくれば出来上がり！



ハミング符号の符号長

- 元の語： $2^m - m - 1$ ビット
- 符号語： $2^m - 1$ ビット
  （冗長性として付加される分が m ビット）
  となる
  - 1ビット→3ビット
  - 4ビット→7ビット
  - 11ビット→15ビット，etc…

ハミング符号の符号語間最小距離

$$H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

- 検査行列Hの列ベクトルはどの2つも互いに1次独立である → 最小距離≧3
- $(1\ 0\ 0\ \ldots\ 0)^T$, $(0\ 1\ 0\ \ldots\ 0)^T$, $(1\ 1\ 0\ \ldots\ 0)^T$ という3つの列ベクトルを取るとこれらは1次従属になっている → 最小距離＝3

ハミング符号を用いた誤り訂正

- ハミング符号の誤り訂正能力は 常に 1
- シンドロームが0でなかった場合，

$$e = \begin{pmatrix}1\\0\\0\\\vdots\\0\end{pmatrix} \text{ or } \begin{pmatrix}0\\1\\0\\\vdots\\0\end{pmatrix} \text{ or } \begin{pmatrix}0\\0\\1\\\vdots\\0\end{pmatrix} \cdots \begin{pmatrix}0\\0\\0\\\vdots\\1\end{pmatrix}$$

の $2^m - 1$ 個の誤りベクトルの候補から
$Hy = He$ となるものを探せばよい

（Hの列ベクトルからシンドロームに一致するものを探せばよい）

小レポート

- m＝２の場合のハミング符号は，これまでに出てきた

  0→000，1→111

  という３重化による誤り訂正符号になることを確かめよ．

- m＝４の場合についてハミング符号を作成せよ．また，その生成行列を用いて生成した任意の符号語に１ビットの誤りを加え，それを検査行列を用いて誤り訂正してみよ．