誤り訂正符号

通信途中でのノイズ (雑音) による誤りに
どう対処するか

$\longrightarrow$ 多少の誤りなら検出・訂正できる仕組みを
数理技術により実現 (**誤り訂正符号**)

情報通信中に誤り (らしきこと) に出遭ったとき

- 聞き直す (より安全なプロトコルの採用)
- 見当を付ける (誤りの自動訂正)

$\longrightarrow$ **冗長性**を利用して**安全性**を確保

誤り訂正符号 (お話)

「誤りの自動訂正」が出来るように
予め適切に冗長性を持たせて通信する

- **誤り訂正性能は高く**
- **とは言えなるべく効率的に**

$\longrightarrow$ 有限体上の線型代数・代数幾何などの
数理構造を利用

誤り訂正符号

誤り検出:

符号 $\mathcal{C}^* : S^* \longrightarrow T^*$ **で、**
受信した語 $y \in T^*$ **が** $y \notin \mathrm{Im}\mathcal{C}^*$ **なら誤り**

誤り訂正:

正しくは y **に “一番近い”** $x \in \mathrm{Im}\mathcal{C}^*$ **だろう**

誤り訂正符号

受信語 $y \notin \mathrm{Im}\mathcal{C}^* \subset T^*$ で誤り検出

$\longrightarrow$ T^* 全部は使わない

$\longrightarrow$ 冗長度を持たせて誤り検出・訂正

とは言え

- **より効率良く (冗長度少なめ)**
- **より高い誤り対処性能を持つ**
  (誤りが沢山あっても大丈夫)
  ものが望ましい

誤り訂正符号

以下では、

- 生起確率の違いを考慮しない
- 等長符号のみを考える
  (全ての符号語が同じ長さ)

効率良い情報源符号で符号化された文字列を
一定の個数毎に切って再符号化する
と想定 (通信路符号)

誤り訂正符号

符号

$$\mathcal{C} : S \longrightarrow V := T^n \quad (n\text{：符号語長})$$

の像 $\mathrm{Im}\mathcal{C} =: U \subset V$ **のみが大事**

$\longrightarrow$ **寧ろ、像** U **をも単に** $\mathcal{C}$ **と書き、**
これを符号と呼ぶ： $\mathcal{C} \subset V$

誤り訂正符号の性能を表すには

- **符号長 n (当座固定)**
- **符号語数 $M := \#\mathcal{C}$ (大きいほど良い)**
- **訂正出来る誤りの数 t (大きいほど良い)**

しかし一般に、

「$M \longrightarrow$ 大」と「$t \longrightarrow$ 大」とは

相反する要求!!

誤りの訂正

符号 $\mathcal{C} \subset V = T^n$ で、

- **受信語 $y \notin \mathcal{C}$ によって誤り検出**
- **y に “一番近い” $x \in \mathcal{C}$ が正しい、として誤り訂正**

$\longrightarrow$ “一番近い” とは？

$\longrightarrow$ V に “距離” を導入
(通常 Hamming 距離 を用いる)

距離の公理

X：集合

$d : X \times X \longrightarrow \boldsymbol{R}_{\geq 0}$：$X$ 上の**距離**
(metric, distance)

であるとは、

- $d(x, y) = 0 \Longleftrightarrow x = y$
- $d(x, y) = d(y, x)$
- $d(x, y) + d(y, z) \geq d(x, z)$
  ：**三角不等式 (triangle inequality)**

Hamming 距離

$V = T^n$ 上に次で定まる距離

$$d : V \times V \longrightarrow \boldsymbol{R}_{\geq 0}$$

を V 上の **Hamming 距離** (distance) と呼ぶ：

$\boldsymbol{x} = (x_1, \ldots, x_n), \boldsymbol{y} = (y_1, \ldots, y_n) \in V$ に対し、

$$d(\boldsymbol{x}, \boldsymbol{y}) := \#\{i | x_i \neq y_i\}$$

誤り訂正の性能

$\mathcal{C} \subset V = T^n$

n 箇所のうち何箇所違っても訂正できるか？

$=$ 距離が幾ら以内なら訂正できるか？

t 箇所違っても一意に訂正できる

$\Longleftrightarrow \forall \boldsymbol{y} \in V$ に対し

$d(\boldsymbol{x}, \boldsymbol{y}) \leq t$ なる $\boldsymbol{x} \in \mathcal{C}$ は高々 1 つ

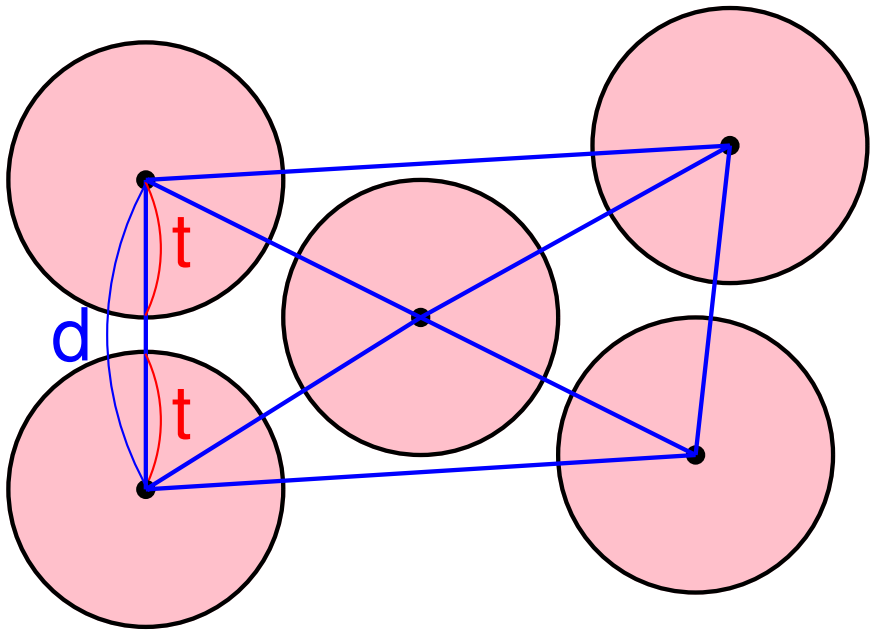
$\left(\#\{\boldsymbol{x} \in \mathcal{C} \middle| d(\boldsymbol{x}, \boldsymbol{y}) \leq t\} \leq 1\right)$

誤り訂正の性能

$d := \min\{d(\boldsymbol{x}, \boldsymbol{x}') | \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{C}, \boldsymbol{x} \neq \boldsymbol{x}'\}$

: $\mathcal{C}$ の**最小距離 (minimum distance)**

$d \geq 2t+1$ **なら一意に訂正可能** $\longrightarrow t = \left\lfloor \dfrac{d-1}{2} \right\rfloor$



誤り訂正符号の性能を表すには

- **符号長 n (当座固定)**
- **符号語数 $M := \#\mathcal{C}$ (大きいほど良い)**
- **訂正出来る誤りの数 t (大きいほど良い)**

$\longrightarrow$ **訂正性能 t は最小距離 d で計れる!!**

誤り訂正符号の性能を表すには

- **符号長** n **(当座固定)**
- **符号語数** $M := \#\mathcal{C}$ **(大きいほど良い)**
- **最小距離** d **(大きいほど良い)**

$$\longrightarrow (n, M, d)\text{-}\textbf{符号}$$

誤り訂正符号の性能を表すには

- **符号長 n (当座固定)**
- **符号語数 $M := \#\mathcal{C}$ (大きいほど良い)**
- **最小距離 d (大きいほど良い)**

「$M \longrightarrow$ 大」と「$d \longrightarrow$ 大」とは
相反する要求!!

問題: n, d を固定した時の M の上限は ?

符号語数の評価

問題: n, d を固定した時の M の上限は ?

d : **設計距離**

$$A_q(n,d) := \max\{M \,|\, \exists \mathcal{C} : q \text{ 元 } (n, M, d)\text{-符号 }\}$$

$$\longrightarrow n, d \text{ と } q := \#T \text{ とで評価する}$$

符号語数の評価

Hamming 距離で “半径 t の球” の元の個数は?

$$B(\boldsymbol{x}, t) := \{\boldsymbol{y} \in V \,|\, d(\boldsymbol{x}, \boldsymbol{y}) \leq t\}$$

: 中心 x, 半径 t の球体

$$\#B(\boldsymbol{x}, t) = \sum_{k=0}^{t} \binom{n}{k} (q-1)^k$$

Hamming の球充填上界 (sphere-packing bound)

$q := \#T$ のとき、

$(n, M, 2t+1)$-符号が存在

$$\Longrightarrow M \left(\sum_{k=0}^{t} \binom{n}{k} (q-1)^k \right) \leq q^n$$

従って、

$$A_q(n,d) \left(\sum_{k=0}^{t} \binom{n}{k} (q-1)^k \right) \leq q^n \quad (t = \left\lfloor \tfrac{d-1}{2} \right\rfloor)$$

Gilbert-Varshamov の下界

$q := \#T$ **のとき、**

$$A_q(n,d)\left(\sum_{k=0}^{d-1}\binom{n}{k}(q-1)^k\right) \geq q^n$$

即ち、

$$M\left(\sum_{k=0}^{d-1}\binom{n}{k}(q-1)^k\right) \leq q^n$$

$$\Longrightarrow (n,M,d)\text{-}\textbf{符号が存在}$$

問題: n, d を固定した時の M の上限は ?

符号語数 M の大きい符号を
組織的に構成するには、

$\mathcal{C}$ の元がなるべく "均等に" 分布する
のが望ましい

$\longrightarrow V = T^n$ の持つ数理構造 (対称性) を利用
$\cdots$ 線型符号