Codes correcteurs d'erreurs

09 mars, 2023



1/29

Commençons par un petit jeu

Règles du jeu

- Quelqu'un choisit un nombre au hasard entre 0 et 15 sans me le dire.
- Je lui pose 7 questions où il faut répondre par oui ou non.
- La personne a le droit de mentir une seule fois.
- Il n'est pas forcé de le faire.
- Je vais retrouver le nombre.

2/29

Questions

- Le nombre choisi est-il > 8?
- Le nombre apparaît-il dans la liste 4 5 6 7 12 13 14 15?
- Le nombre apparaît-il dans la liste 2 3 6 7 10 11 14 15?
- 4 Le nombre choisi est-il impair?
- Le nombre apparaît-il dans la liste 1 3 4 6 8 10 13 15?
- Le nombre apparaît-il dans la liste 1 2 5 6 8 11 12 15?
- Le nombre apparaît-il dans la liste 2 3 4 5 8 9 14 15?

Questions

- Le nombre choisi est-il ≥ 8 ?
- Le nombre apparaît-il dans la liste 4 5 6 7 12 13 14 15?
- Le nombre apparaît-il dans la liste 2 3 6 7 10 11 14 15?
- 4 Le nombre choisi est-il impair?
- Le nombre apparaît-il dans la liste 1 3 4 6 8 10 13 15?
- Le nombre apparaît-il dans la liste 1 2 5 6 8 11 12 15?
- Le nombre apparaît-il dans la liste 2 3 4 5 8 9 14 15?

Suspens

Questions

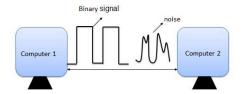
- **1** Le nombre choisi est-il ≥ 8 ?
- 2 Le nombre apparaît-il dans la liste 4 5 6 7 12 13 14 15?
- Le nombre apparaît-il dans la liste 2 3 6 7 10 11 14 15?
- 4 Le nombre choisi est-il impair?
- Le nombre apparaît-il dans la liste 1 3 4 6 8 10 13 15?
- Le nombre apparaît-il dans la liste 1 2 5 6 8 11 12 15?
- Le nombre apparaît-il dans la liste 2 3 4 5 8 9 14 15?

Suspens

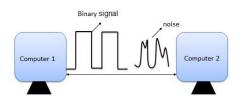
Le truc

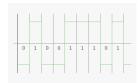
J'ai utilisé un code de Hamming qui est un exemple de code correcteur d'erreurs.

Code correcteurs d'erreurs, pourquoi?

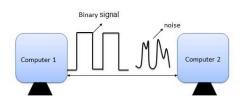


Code correcteurs d'erreurs, pourquoi?





Code correcteurs d'erreurs, pourquoi?







Applications

Communications

Internet

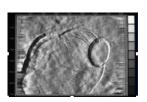


- Communications sans fil (téléphones, wifi...).
- QR-code

Applications

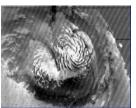
Communications

• Missions Mariner 2 et 9 sur Mars

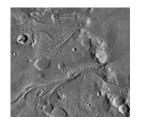


Mont Olympus

Voyager



Pôle nord de Mars





Saturne

Applications

Stockage



- Un CD a facilement plus de 500 000 erreurs!
- Un code correcteur de Reed-Solomon permet de corriger environ 4000 bits consécutifs soit une rayure de plus d'un millimètre de large.

Clés dans les numéros

- Comptes bancaires,
- Sécurité sociale,
- ISBN des livres.

Avant l'informatique



Α	Alpha	N	November			
В	Bravo	0	Oscar			
С	Charlie	Р	Papa			
D	Delta	Q	Quebec			
Ε	Echo	R	Romeo			
F	Foxtrot	S	Sierra			
G	Golf	Т	Tango			
Н	Hotel	U	Uniform			
1	India	V	Victor			
J	Juliet	W	Whisky			
K	Kilo	X	X-ray			
L	Lima	Υ	Yankee			
M	Mike	Z	Zulu			

- Messages codés plus longs
- Messages codés qui ont du sens éloignés deux à deux

Principe fondamental

Rajouter de la redondance

- → Détecter s'il y a eu quelques erreurs
- → (éventuellement) les corriger.

Principe fondamental

Rajouter de la redondance

- → Détecter s'il y a eu quelques erreurs
- → (éventuellement) les corriger.

But

- Ajouter le moins de redondance possible,
- Pouvoir corriger le plus d'erreurs possibles,
- En pratique, codage et décodage rapide!

Construire un code

On traduit notre information à transmettre en mots dans un certain alphabet.

Exemple du jeu: écriture binaire

Information = un nombre entre 0 et 15, Alphabet= $\{0,1\}$, Mot à 4 lettres.

Construire un code

On traduit notre information à transmettre en mots dans un certain alphabet.

Exemple du jeu: écriture binaire

Information = un nombre entre 0 et 15, Alphabet= $\{0,1\}$, Mot à 4 lettres.

0	0000	
1	0001	
2	0010	
3	0011	
4	0100	
5	0101	
6	0110	
7	0111	

8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111

Construire un code

On traduit notre information à transmettre en mots dans un certain alphabet.

Exemple du jeu: écriture binaire

Information = un nombre entre 0 et 15, Alphabet= $\{0,1\}$, Mot à 4 lettres.

0	0000	
1	0001	
2	0010	
3	0011	
4	0100	
5	0101	
6	0110	
7	0111	

8	1000
9	1001
10	1010
11	1011
12	1100
13	1101
14	1110
15	1111

Remarque

Ce sont nos quatre premières questions!

Détecter une erreur

Détecter une erreur

Idée: envoyer deux fois le même message (faire sur l'exemple)

Détection d'une unique erreur, mais on ne sait pas où.

Détecter une erreur

Idée: envoyer deux fois le même message (faire sur l'exemple)

Détection d'une unique erreur, mais on ne sait pas où.

Inconvénient:

la taille du message est doublée pour la même quantité d'informations.

Taux d'information = 1/2.

Rajouter une lettre à un mot pour qu'il ait un nombre pair de 1.

1000	1000
1001	1001
1101	1101

Rajouter une lettre à un mot pour qu'il ait un nombre pair de 1.

1000	1000 <mark>1</mark>
1001	1001
1101	1101

Rajouter une lettre à un mot pour qu'il ait un nombre pair de 1.

1000	1000 <mark>1</mark>
1001	1001 <mark>0</mark>
1101	1101

Rajouter une lettre à un mot pour qu'il ait un nombre pair de 1.

1000	1000 <mark>1</mark>
1001	1001 <mark>0</mark>
1101	1101 <mark>1</mark>

Rajouter une lettre à un mot pour qu'il ait un nombre pair de 1.

Exemples:

1000	10001
1001	1001 <mark>0</mark>
1101	1101 <mark>1</mark>

Une erreur rend le nombre de 1 impair

Taux d'information: 4/5

Rajouter une lettre à un mot pour qu'il ait un nombre pair de 1.

Exemples:

1000	10001
1001	1001 <mark>0</mark>
1101	1101 <mark>1</mark>

Une erreur rend le nombre de 1 impair

Taux d'information: 4/5

Applications

Un bit de parité est systématiquement rajouté aux octets (mots de 8 lettres) au sein d'un ordinateur ou pour les transmissions sur internet en raison de sa simplicité. En cas d'erreur, on redemande l'information.

Les clés RIB, SS, ISBN fonctionnent sur le même principe (avec des formules plus compliquées).

Corriger une erreur

Corriger une erreur

Idée: envoyer trois fois le même message. (Faire l'exemple)

Corriger une erreur

Idée: envoyer trois fois le même message. (Faire l'exemple)

Inconvénient: la taille du message a triplé pour la même quantité d'informations!

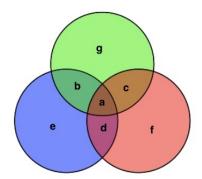
Taux d'information = 1/3.

Code de Hamming (7,4,3)

• On part d'un mot de 4 lettres (0 ou 1) $a \quad b \quad c \quad d$

Code de Hamming (7,4,3)

- On lui rajoute 3 nouvelles lettres:
 - ▶ e bit de parité de a b d
 - ▶ f bit de parité de a c d
 - p g bit de parité de a b c



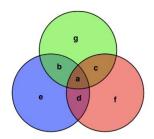
Code de Hamming (7,4,3)

0	0	0	0	0	0	0
•	-	-	U	U	U	٠ ا
0	0	0	1	1	1	0
0	0	1	0	0	1	1
0	0	1	1	1	0	1
0	1	0	0	1	0	1
0	1	0	1	0	1	1
0	1	1	0	1	1	0
0	1	1	1	0	0	0

1	0	0	0	1	1	1
1	0	0	1	0	0	1
1	0	1	0	1	0	0
1	0	1	1	0	1	0
1	1	0	0	0	1	0
1	1	0	1	1	0	0
1	1	1	0	0	0	1
1	1	1	1	1	1	1

16 mots de 7 lettres parmi 128

Deux mots ont au moins 3 lettres différentes.



0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	1
0	0	1	0	0	1	1	2
0	0	1	1	1	0	1	3
0	1	0	0	1	0	1	4
0	1	0	1	0	1	1	4 5
0	1	1	0	1	1	0	6
0	1	1	1	0	0	0	7

1	0	0	0	1	1	1	8
1	0	0	1	0	0	1	9
1	0	1	0	1	0	0	10
1	0	1	1	0	1	0	11
1	1	0	0	0	1	0	12
1	1	0	1	1	0	0	13
1	1	1	0	0	0	1	14
1	1	1	1	1	1	1	15

- Q5 demande la valeur de e
- Q6 demande la valeur de f
- Q7 demande la valeur de g

0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	1
0	0	1	0	0	1	1	2
0	0	1	1	1	0	1	3
0	1	0	0	1	0	1	4
0	1	0	1	0	1	1	5
0	1	1	0	1	1	0	6
0	1	1	1	0	0	0	7

Г	1	0	0	0	1	1	1	8
:	1	0	0	1	0	0	1	9
:	1	0	1	0	1	0	0	10
:	1	0	1	1	0	1	0	11
:	1	1	0	0	0	1	0	12
:	1	1	0	1	1	0	0	13
:	1	1	1	0	0	0	1	14
:	1	1	1	1	1	1	1	15

- Q5 demande la valeur de e
- Q6 demande la valeur de f
- Q7 demande la valeur de g

0	0	0	0	0	0	0	0
0	0	0	1 0	1	1	0	1
0	0	1		0	1	1	2
0	0	1	1 0	1	0	1	3
0	1	1 0	0	1	0	1	4
0	1	0	1	0	1	1	5
0	1	1	0	0 1 0 1 1 0 1	0 1 1 0 0 1 1	1 1 1 1 0 0	1 2 3 4 5 6 7
0	1	1	1	0	0	0	7

1	0	0	0	1	1	1	8
1	0	0	1	0	0	1	9
1	0	1	0	1	0	0	10
1	0	1	1	0	1	0	11
1	1	0	0	0	1	0	12
1	1	0	1	1	0	0	13
1	1	1	0	0	0	1	14
1	1	1	1	1	1	1	15

- Q5 demande la valeur de e
- Q6 demande la valeur de f
- Q7 demande la valeur de g

0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	1
0	0	1	0	0	1	1	2
0	0	1	1	1		1	3
0	1	0	0	1	0 1	1	4
0	1	0	1	0	1	1	5
0	1	1	0	1	1	0	6
0	1	1	1	0	0	0	7

1	0	0	0	1	1	1	8
1	0	0	1	0	0	1	9
1	0	1	0	1	0	0	10
1	0	1	1	0	1	0	11
1	1	0	0	0	1	0	12
1	1	0	1	1	0	0	13
1	1	1	0	0	0	1	14
1	1	1	1	1	1	1	15

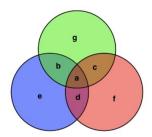
- Q5 demande la valeur de e
- Q6 demande la valeur de f
- Q7 demande la valeur de g

Retour sur les questions

0	0	0	0	0	0	0	0
0	0	0	1	1	1	0	1
0	0	1	0	0	1	1	2
0	0	1	1	1	0	1	1 2 3
0	1	0	0	1	0	1	
0	1	0	1	0	1	1	4 5
0	1	1	0	1	1	0	6
0	1	1	1	0	0	0	7

- Q5 demande la valeur de e
- Q6 demande la valeur de f
- Q7 demande la valeur de g

1	0	0	0	1	1	1	8
1	0	0	1	0	0	1	9
1	0	1	0	1	0	0	10
1	0	1	1	0	1	0	11
1	1	0	0	0	1	0	12
1	1	0	1	1	0	0	13
1	1	1	0	0	0	1	14
1	1	1	1	1	1	1	15



Et si mon message fait plus de 4 bits? 010011101001010101?

Et si mon message fait plus de 4 bits? 010011101001010101? Le code de Hamming (comme de nombreux autres) est un code en blocs .

Et si mon message fait plus de 4 bits? 01001110100101011?

Le code de Hamming (comme de nombreux autres) est un code en blocs .

Et si mon message fait plus de 4 bits?

010011101001010101?

Le code de Hamming (comme de nombreux autres) est un code en blocs .

Et si mon message fait plus de 4 bits?

010011101001010101?

Le code de Hamming (comme de nombreux autres) est un code en blocs

```
010011101001010101 \rightarrow 0001 \qquad 0011 \qquad 1010 \qquad 0101 \qquad 0101 \\ \rightarrow \boxed{0001110 \quad 0011101 \quad 1010100 \quad 0101011 \quad 0101011}
```

Distance de Hamming

C'est le nombres de lettres différentes entre deux mots.

Exemples:

00100110 00100100 distance 1

Et si mon message fait plus de 4 bits?

010011101001010101?

Le code de Hamming (comme de nombreux autres) est un code en blocs

```
010011101001010101 \rightarrow 0001 \qquad 0011 \qquad 1010 \qquad 0101 \qquad 0101 \\ \rightarrow \qquad 0001110 \quad 0011101 \quad 1010100 \quad 0101011 \quad 0101011
```

Distance de Hamming

C'est le nombres de lettres différentes entre deux mots.

Exemples:

$$00100110 \\ 00110100$$
 distance 2

Et si mon message fait plus de 4 bits? 01001110100101011?

Le code de Hamming (comme de nombreux autres) est un code en blocs .

```
010011101001010101 \rightarrow 0001 \qquad 0011 \qquad 1010 \qquad 0101 \qquad 0101 \\ \rightarrow \boxed{0001110 \quad 0011101 \quad 1010100 \quad 0101011 \quad 0101011}
```

Distance de Hamming

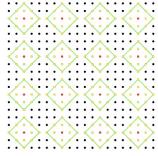
C'est le nombres de lettres différentes entre deux mots.

Exemples:

distance 2

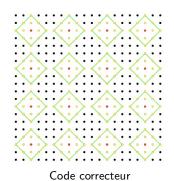
Paramètres du codes





Code correcteur

- En rouge: les mots du code ,
- En jaune: les mots que l'on peut corriger,
- En noir: les autres.



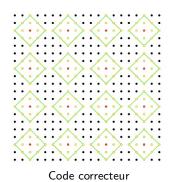
• En rouge: les mots du code ,

- En jaune: les mots que l'on peut corriger,
- En noir: les autres.



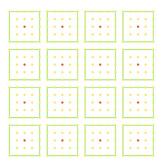
Les sphères s'intersectent...

Un code qui peut détecter les erreurs mais pas les corriger.



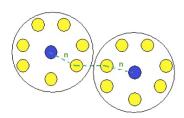
• En rouge: les mots du code ,

- En jaune: les mots que l'on peut corriger,
- En noir: les autres.

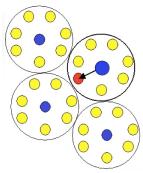


Les boules couvrent tout l'espace

Un tel code est dit parfait (pas de points noirs)



Si distance entre 2 mots du code est toujours $\geq 2n+1$ Alors, les boules de rayon n sont disjointes.



Boules disjointes

⇒ décodage possible!

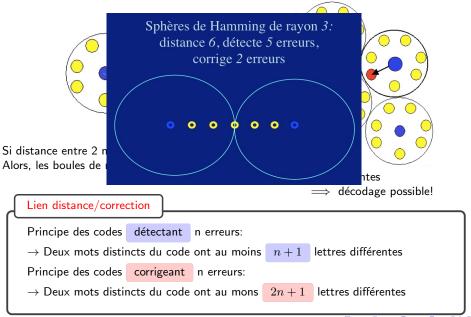
Lien distance/correction

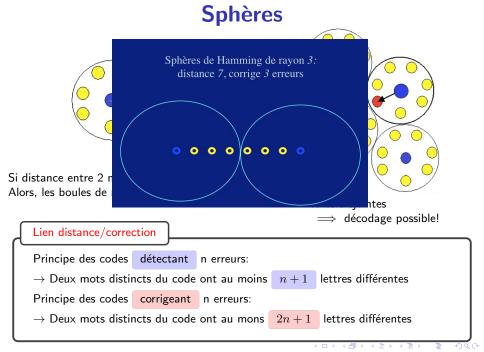
Principe des codes détectant n erreurs:

ightarrow Deux mots distincts du code ont au moins n+1 lettres différentes

Principe des codes corrigeant n erreurs:

 \rightarrow Deux mots distincts du code ont au mons 2n+1 lettres différentes





Hamming(7,4,3) sur un canal bruité

En pratique: canal bruité \rightarrow chaque bit a une probabilité p d'être perturbé.

Dès qu'il y a 2 erreurs dans un mot, le décodage échoue!

Probabilité que le mot est correct

р	sans chiffrage	avec chiffrage $(7,4,3)$
1/4	32%	44%
1/10	66%	85%
1/20	81%	96%
1/100	96%	99,8%

Parenthèse à propos du théorème de Shannon

Canal: chaque bit a une probabilité p d'être perturbé.

Théorème de Shannon (1948)

Soit
$$\rho = 1 - (p \log_2(1/p) + (1-p) \log_2(1/(1-p)))$$

- Soient $\delta > 0$ et $\varepsilon > 0$, il existe un code
 - de taux $> \rho \delta$
 - de probabilité d'erreur $< \varepsilon^n$.
- Si un code a un taux $\tau>\rho$, alors il existe un mot qui est décodé avec probabilité $\leq 1/2$.

En clair

Sur un tel canal, un code "idéal" a un taux égal à ρ .

Comment encoder et décoder rapidement?
Les codes de Hamming font partie de la famille des codes linéaires .

Comment encoder et décoder rapidement?

Les codes de Hamming font partie de la famille des codes linéaires.

Si on considère les 16 mots de 4 lettres $x \in \{0,1\}^4$,

les produits $x \cdot G$ donnent les 16 mots de Hamming (7,4,3).

Comment encoder et décoder rapidement?

Les codes de Hamming font partie de la famille des codes linéaires

Si on considère les 16 mots de 4 lettres $x \in \{0, 1\}^4$,

les produits $x \cdot G$ donnent les 16 mots de Hamming (7,4,3).

Calcul modulo 2 0+0=0 0+1=1 1+0=1 1+1=0

Comment encoder et décoder rapidement?

Les codes de Hamming font partie de la famille des codes linéaires

Si on considère les 16 mots de 4 lettres $x \in \{0, 1\}^4$,

les produits $x \cdot G$ donnent les 16 mots de Hamming (7,4,3).

Calcul modulo 2
$$0+0=0 \qquad 0+1=1 \\ 1+0=1 \qquad 1+1=0$$

Encodage très rapide

Codes correcteurs d'erreurs

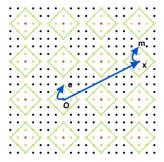
Le produit scalaire d'une ligne de G et d'une ligne de H égale 0.

Attention

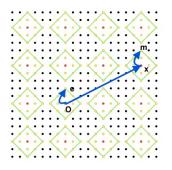
Ce n'est pas un vrai produit scalaire.

Codes correcteurs d'erreurs

Message à décoder m

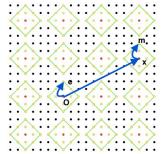


Message à décoder m



$$\begin{split} m &= x + e. \\ \text{Or } H \cdot x &= 0, \\ \text{donc } H \cdot m &= H \cdot x + H \cdot e = H \cdot e \\ \text{syndrome du message} \end{split}$$

Message à décoder m



$$\begin{split} m &= x + e. \\ \text{Or } H \cdot x &= 0, \\ \text{donc } H \cdot m &= H \cdot x + H \cdot e = H \cdot e \\ \text{syndrome du message} \end{split}$$

Si on connait tous les syndromes $H \cdot e$ pour $e \in B(O)$

H(e)	е
001	0000001
010	0000010

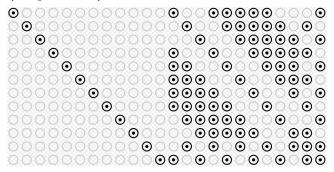
On retrouve e et donc x = m - e.

Autre exemple de code linéaire : Code de Golay

Mots de longueur 23 :

- 12 bits de données.
- 11 bits de contrôle.

Distance = 7 (corrige 3 erreurs).



Matrice génératrice

Empilement parfait

Utilisée pour les photos de Jupiter et Saturne par sondes Voyager 1 et 2

Borne de Singleton

Question?

Si on fixe n la taille des mots du code et d la distance, quel est la valeur maximale pour k?

Cas de base:

- d=1, on peut obtenir un code (n, n, 1).
- d=2, on peut obtenir un code (n, n-1, 2) (bit de parité).

Borne de Singleton

Question?

Si on fixe n la taille des mots du code et d la distance, quel est la valeur maximale pour k?

Cas de base:

- d=1, on peut obtenir un code (n, n, 1).
- d=2, on peut obtenir un code (n, n-1, 2) (bit de parité).

Singleton

Si C est un code (n, k, d), alors $k \leq n - d + 1$.

Remarque

Un code satisfaisant l'égalité est appelé MDS (Maximum Distance Separable) Cette notion d'optimalité est incomparable avec les codes parfaits.

Code de Reed-Solomon

Code de Reed-Solomon:

- Choisir n éléments α_i de \mathbb{F}_q
- ② Un message $m \in \mathbb{F}_q^k$ est une liste d'éléments c_0, \ldots, c_{k-1} . Représentons le message comme un polynôme $P(X) = \sum_{i=0}^{k-1} c_i X^i$.
- **1** L'encodage du message est le n-uplet $E(m) = (P(\alpha_1), \dots, P(\alpha_n))$.

Exemple

$$q = 5, k = 3, n = 3$$

Choisissons les points 1, 3 et 4 dans \mathbb{F}_5 .

Notre message est m = (4, 1, 2) $\rightarrow P(X) = 4 + X + 2X^2$

$$E(m) = (P(1), P(3), P(4)) = (2, 0, 0)$$

Code de Reed-Solomon / Reed-Muller

Propriétés des codes de Reed-Solomon

- Ces codes sont des MDS. Raison: deux polynômes distincts de degrés $\leq k-1$ coincident en au plus k-1 points.
- Intérêt: on peut choisir q grand.
 Donc résiste aux erreurs consécutives.
 Code utilisé pour les CDs, DVDs et les QR-codes.

Code de Reed-Solomon / Reed-Muller

Propriétés des codes de Reed-Solomon

- Ces codes sont des MDS.
 - Raison: deux polynômes distincts de degrés $\leq k-1$ coincident en au plus k-1 points.
- Intérêt: on peut choisir q grand.
 Donc résiste aux erreurs consécutives.
 Code utilisé pour les CDs, DVDs et les QR-codes.

Généralisation aux polynômes multivariés : codes de Reed-Muller

→ Code utilisé pour les sondes Voyager pour Uranus et Neptune.

Murci Merxi Merce