Mathématique pour l'informatique 1
INFOB125

Chapitre 4 : Codage – Décodage

Marie-Ange Remiche

Cours donné par Martine De Vleeschouwer

Objectifs – à partir d'une illustration

Étude de cas

On distingue essentiellement **deux moments clé pour le *codage d'un message*** :

1. Le premier, appelé *codage-source* a pour objectif de **coder** selon un code source particulier un message exprimé dans un langage particulier en un message binaire. Le code ASCII est un exemple de codage-source.
2. Le second, appelé *codage-transmission* a pour objectif de **préparer** le message binaire en vue d'**une transmission efficace**, sans erreur et rapide. En effet, toute transmission de message est sujette à des problèmes de transmission et donc toute transmission est susceptible de comporter des erreurs de transmission.

Objectifs – à partir d'une illustration

Étude de cas – suite

2. **Le codage-transmission** part du principe d'ajouter au message original des bits d'information pour limiter les conséquences des erreurs de transmission.
   Le message exprimé grâce à un ensemble de bits, ce qu'on appelle un **mot**, est complété par des bits supplémentaires, ou encore complètement transformé et ensuite complété par des bits supplémentaires. On parle de **codage par blocs**.
   Exemple : Dans une adresse postale, le nom de la ville est complété du code postal afin de minimiser les erreurs de lecture du nom de la ville.

Objectifs – à partir d'une illustration

Étude de cas – suite

Ce *codage par blocs* est donc une méthode de présentation des mots à transmettre en vue de permettre au récepteur du mot de détecter la présence éventuelle d'erreurs de transmission , et le cas échéant de les corriger.

Objectifs

Être capable...

- de préciser les propriétés d'un codage, tels que son rendement, sa distance minimale,...
- de déterminer si un codage est linéaire
- de manipuler les codages linéaires : construire le tableau standard, la matrice génératrice, la matrice de contrôle, la table des syndromes à partir des renseignements fournis,
- corriger en fonction tout message détecté comme erroné,
- déterminer si un code est un code de Hamming.

Ce que nous allons voir dans ce chapitre :

Table des matières

1. Quelques **définitions** préalables sont nécessaires
2. pour **détecter, corriger** des erreurs de transmission.
3. On se concentre sur le **codage linéaire**.

Pré-requis : Calcul matriciel...

Définition

Un mot **b** est un élément de $\mathbb{B}^n$, avec $\mathbb{B} = \{0, 1\}$. On le désigne par

$$\mathbf{b} = b_1 b_2 \dots b_n \ ,$$

où $b_i \in \mathbb{B}$.

Exemple

Dans $\mathbb{B}^3$, nous avons le mot $\mathbf{b} = 011$, composé des bits $b_1 = 0,\ b_2 = 1,\ b_3 = 1$.

Définitions

1. L'addition modulo 2 définie sur $\mathbb{B}$ est :

$$\begin{aligned} 1 \oplus 1 &\overset{\text{déf.}}{=} 0 \\ 1 \oplus 0 &\overset{\text{déf.}}{=} 1 \\ 0 \oplus 1 &\overset{\text{déf.}}{=} 1 \\ 0 \oplus 0 &\overset{\text{déf.}}{=} 0 \end{aligned}$$

2. Sur $\mathbb{B}^n$, l'addition modulo 2 de deux mots binaires de longueur n est obtenue en sommant modulo 2 chaque bit de même position.

Exercice

Que vaut $1001001 \oplus 1011101$?

Remarque : ces 2 mots appartiennent à $\mathbb{B}^7$.

Réponse : 0010100

Propriété

$$a \oplus b = c \quad \Leftrightarrow \quad b \oplus c = a \quad \Leftrightarrow \quad c \oplus a = b$$

Exemple – Utilité

Imaginons que nous devons enregistrer la séquence suivante de seize bits :

$$1100011010111001$$

sur trois disques.
Pour créer de la redondance, on peut placer tous les bits en position impaire sur le premier disque et les autres sur le second disque. Le troisième disque contient ensuite le résultat de l'opération $\oplus$ appliquée aux bits situés en même position sur les deux disques.

Si sur le premier disque, on trouve le bit 1 et sur le troisième disque, le bit 0, quel est le bit qui se trouve sur le second disque ?

Le bit 1.

Définitions

- Le **poids d'un mot** $\mathbf{b}$, noté $w(\mathbf{b})$, est le nombre de bits égaux à 1 dans le mot $\mathbf{b}$.
- La **distance de Hamming** entre $\mathbf{a}$ et $\mathbf{b}$, tous deux issus de $\mathbb{B}^n$, distance notée $\delta(\mathbf{a}, \mathbf{b})$ est le nombre de fois qu'un bit de même position est distinct dans $\mathbf{a}$ et $\mathbf{b}$.

Exemple

Dans $\mathbb{B}^5$, le poids du mot 11010 est 3.

La distance entre 10101 et 00111 est 2.

Propriété

Nous pouvons démontrer que

$$\forall \mathbf{a}, \mathbf{b} \in \mathbb{B}^n \ : \quad \boxed{\delta(\mathbf{a}, \mathbf{b}) = w(\mathbf{a} \oplus \mathbf{b})}$$

Propriété

La distance de Hamming est *une distance* à proprement parler :

Soient $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{B}^n$.

Elle est donc :

- symétrique : $\delta(\mathbf{a}, \mathbf{b}) = \delta(\mathbf{b}, \mathbf{a})$
- positive : $\delta(\mathbf{a}, \mathbf{b}) \geq 0$
- elle respecte l'inégalité triangulaire :
  $\delta(\mathbf{a}, \mathbf{b}) + \delta(\mathbf{b}, \mathbf{c}) \geq \delta(\mathbf{a}, \mathbf{c})$
- $(\delta(\mathbf{a}, \mathbf{b}) = 0) \Leftrightarrow \mathbf{a} = \mathbf{b}$

Enfin, elle est invariante par translation :

$$\text{pour tout mot } \mathbf{x} \in \mathbb{B}^n : \quad \delta(\mathbf{a} \oplus \mathbf{x}, \mathbf{b} \oplus \mathbf{x}) = \delta(\mathbf{a}, \mathbf{b}) \ .$$

Définitions

- Le **codage** est la transformation Φ d'un mot de $\mathbb{B}^n$ dans un mot de $\mathbb{B}^p$ (avec $p > n$). $\Phi : \mathbb{B}^n \to \mathbb{B}^p$
- On parle de **bits de contrôle** ou **bits de parité** pour désigner ces bits supplémentaires.
- Lorsque ces bits de contrôle sont placés à la fin du mot, on dit que le codage est **systématique**.

Exemple : codage avec contrôle de parité

Le principe d'un **codage avec contrôle de parité** est le suivant : chaque mot $\mathbf{b} = b_1 b_2 \dots b_n$ est codé par $\Phi(\mathbf{b}) = b_1 b_2 \dots b_n b_{n+1}$ où le bit supplémentaire b_{n+1} est tel que

$$b_{n+1} = \begin{cases} 1 & \text{si } w(\mathbf{b}) \text{ est impair} \\ 0 & \text{dans le cas contraire.} \end{cases}$$

Définitions

- Dans le cas d'une transformation Φ de mot de longueur n vers des mots de longueur p, on parle alors de $(n, p)-$**codage**.
- **L'image $\Phi(\mathbb{B}^n)$ d'un codage Φ** est *l'ensemble des mots-codes* obtenus par le codage Φ.
- L'ensemble $\Phi(\mathbb{B}^n)$ est **le code** obtenu par Φ.
- Le **rendement** ou le **taux d'un codage** est le rapport τ défini comme
$$\tau \overset{\text{déf.}}{=} \frac{n}{p}.$$
- Le mot transmis est appelé **message**.
On ne l'appelera **mot-code** que si le codage a été correctement réalisé.

Exemple : codage avec contrôle de parité

Soit $\Phi : \mathbb{B}^3 \to \mathbb{B}^4$ un codage *avec contrôle de parité* tel que défini précédemment.

- Le mot 110 sera donc codé en le **mot-code** 1100.
- Le **rendement** de ce $(3,4)-$codage est $3/4$.
- Quel mot est réprésenté par **le message** : 1101 ?
  et par **le message** 1000 ?

On détecte certainement une erreur si le nombre de bits marqués à 1 dans un message est un nombre impair.

Autre exemple : le codage par répétition

Le mot $\mathbf{b} = b_1 b_2 \ldots b_n$ sera transmis sous la forme :

$$\Phi(\mathbf{b}) = \underbrace{b_1 \ldots b_1}_{k \text{ fois}} \underbrace{b_2 \ldots b_2}_{k \text{ fois}} \ldots \underbrace{b_n \ldots b_n}_{k \text{ fois}}$$

Le rendement de ce $(n - kn)-$codage est $\frac{1}{k}$.

On détecte certainement une erreur si un mot transmis ne peut être divisé en blocs de k bits tous égaux.

(a) La détection

Principe de détection

Une fois le message désigné comme n'étant pas un mot-code, c-à-d une fois l'erreur détectée, il faut proposer une retransmission du message ou proposer une correction de l'erreur.

On corrige le message erroné selon le **principe du maximum de vraisemblance**, c-à-d :

L'erreur commise est vraisemblablement l'erreur la plus probable

(c-à-d celle comprenant le moins de bits erronés)

Exemple : codage avec contrôle de parité

Si le message 1101 est réceptionné, comment le corriger ?

Par 1100 ou 1111 ou ...

Définition

La **distance minimale du code**, notée δ, est la plus petite distance de Hamming séparant deux mots-code distincts.

Exemple : codage avec contrôle de parité

Dans le codage avec contrôle de parité de $\mathbb{B}^3 \to \mathbb{B}^4$, il y a 8 mots-code, à savoir : 0000, 0011, 0101, 0110, 1001, 1100, 1111.

La **distance minimale du code** est donc. . .

	0000	0011	0101	0110	1001	1010	1100
0011	2						
0101	2	2					
0110	2	2	2				
1001	2	2	2	4			
1010	2	2	4	2	2		
1100	2	4	2	2	2	2	
1111	4	2	2	2	2	2	2

(a) La détection

Propriété et définition

Tous les messages faux comportant un nombre d'erreurs strictement plus petit que δ sont nécessairement détectés.
On dit alors que le codage Φ est $(\delta - 1)$**-détecteur**.

Exemple : codage avec contrôle de parité

Le codage avec contrôle de parité est 1-détecteur :
On ne peut détecter la présence d'erreur dans un message où deux bits auraient été modifiés.
Le message 1100 est-il correct ?

(b) La correction

Soit un codage $\Phi : \mathbb{B}^n \to \mathbb{B}^p$

Définition

On dit d'un **décodage** $\psi: \mathbb{B}^p \to \mathbb{B}^n$ associé à un codage Φ qu'il est ***k*-correcteur** si, chaque fois que le nombre d'erreurs de transmission ne dépasse pas k, le mot reçu est corrigé sans ambiguité.

Propriété

Lorsque le nombre N d'erreurs d'un message est tel que $N < \delta/2$, le message peut être corrigé.

Exemple : codage avec contrôle de parité

Le codage avec contrôle de parité est 0-correcteur.

(a) Définition

Définitions

- Un code est un **code linéaire** lorsque la somme modulo 2 de deux mots-code quelconques est un mot-code.
- Un codage est un **codage linéaire** lorsque la transformation Φ respecte l'égalité suivante :

$$\Phi(\mathbf{m}_1 \oplus \mathbf{m}_2 \ldots \oplus \mathbf{m}_k) = \Phi(\mathbf{m}_1) \oplus \Phi(\mathbf{m}_2) \ldots \oplus \Phi(\mathbf{m}_k)$$

et ce, pour tout k et tous mots $\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_k$.

Exemple : codage avec contrôle de parité

Le codage avec contrôle de parité est-il linéaire ?

Exemple : codage avec contrôle de parité

Le codage avec contrôle de parité est un codage linéaire.

Pour le démontrer, il faut vérifier que, **pour n'importe quel** couple $(\mathbf{m}_i, \mathbf{m}_j)$ de mots-code, on a :

$$\Phi(\mathbf{m}_i \oplus \mathbf{m}_j) = \Phi(\mathbf{m}_i) \oplus \Phi(\mathbf{m}_j).$$

Nous ne le faisons ici que pour $\mathbf{m}_1 = 000$ et $\mathbf{m}_8 = 111$:

- $\Phi(\mathbf{m}_1 \oplus \mathbf{m}_8) = \Phi(000 \oplus 111) = \Phi(111) = \underbrace{1111}$
- $\Phi(\mathbf{m}_1) \oplus \Phi(\mathbf{m}_8) = \Phi(000) \oplus \Phi(111) = 0000 \oplus 1111 = \underbrace{1111}$

Propriété

Un **codage linéaire** est tel qu'il est **complètement défini** lorsqu'on connaît les mots-code obtenus **par le codage des** n **mots suivants** :

$$\begin{aligned} \mathbf{b}_1 &= 100\ldots00 \\ \mathbf{b}_2 &= 010\ldots00 \\ &\ldots \\ \mathbf{b}_{n-1} &= 000\ldots10 \\ \mathbf{b}_n &= 000\ldots01\ . \end{aligned}$$

Définition

L'ensemble des $\mathbf{b}_i$ (pour $i = 1, \ldots n$) décrits ci-dessus constitue ce qu'on appelle la **base canonique de** $\mathbb{B}^n$.

Leur mot-code correspondant est $\Phi(\mathbf{b}_1), \Phi(\mathbf{b}_2), \ldots, \Phi(\mathbf{b}_n)$

Preuve de la propriété

Ainsi, prenons $\mathbf{b}$ qui s'écrit comme une combinaison linéaire modulo 2 de $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_n$, c-à-d :

$$\mathbf{b} = \bigoplus_{i=1}^{n} a_i \mathbf{b}_i$$

avec $a_i = 0$ ou 1, pour tout $i \in \{1, \ldots, n\}$. Alors :

$$\begin{aligned} \Phi(\mathbf{b}) &= \Phi\left(\bigoplus_{i=1}^{n} a_i \mathbf{b}_i\right) \\ &= \bigoplus_{i=1}^{n} a_i \Phi\left(\mathbf{b}_i\right) \end{aligned}$$

par les propriétés d'un *codage linéaire*.

Remarque : à la place du symbole $\bigoplus$, on peut utiliser le symbole $\sum$ en précisant qu'il s'agit de la somme modulo 2.

Illustration : Exemple 4.5, page 50 du syllabus

Reprenons le codage avec contrôle de parité où $n = 3$ et $p = 4$.

La base canonique est :
$$\begin{aligned} \mathbf{b}_1 &= 100 \\ \mathbf{b}_2 &= 010 \\ \mathbf{b}_3 &= 001 \end{aligned}$$

et les mots-code correspondant sont :
$$\begin{aligned} \Phi(\mathbf{b}_1) &= 1001 \\ \Phi(\mathbf{b}_2) &= 0101 \\ \Phi(\mathbf{b}_3) &= 0011\ . \end{aligned}$$

Ainsi, le mot $\mathbf{b} = 011$ tel que $\mathbf{b} = (0 \cdot \mathbf{b}_1) \oplus (1 \cdot \mathbf{b}_2) \oplus (1 \cdot \mathbf{b}_3)$ est tel que son mot-code correspondant est :

$$\begin{aligned} \Phi(\mathbf{b}) &= \Phi(0 \cdot \mathbf{b}_1 \oplus 1 \cdot \mathbf{b}_2 \oplus 1 \cdot \mathbf{b}_3) \\ \Phi(\mathbf{b}) &= 0 \cdot \Phi(\mathbf{b}_1) \oplus 1 \cdot \Phi(\mathbf{b}_2) \oplus 1 \cdot \Phi(\mathbf{b}_3) \\ &= 0101 \oplus 0011 \\ &= 0110\ . \end{aligned}$$

La matrice génératrice

Définition

La **matrice génératrice du codage** *linéaire systématique*, notée **G**, est obtenue par la juxtaposition de deux matrices :

- une matrice identité de taille n, soit les mots $\mathbf{b}_i$ écrits sous forme de vecteur ligne,
- une **matrice de** *parité*, contenant les bits de contrôle des mots-code $\Phi(\mathbf{b}_i)$.

Exemple : Codage avec contrôle de parité

Le codage avec contrôle de parité est *linéaire et systématique*.

Sa matrice génératrice, dans le cas où $n = 3$, est : $\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

Obtenir un mot-code par la matrice génératrice

Soit un codage $\Phi : \mathbb{B}^n \to \mathbb{B}^p$, avec $p > n$.

Grâce à **G**, sa matrice génératrice $n \times p$, on obtient directement le codage de n'importe quel mot **b** de $\mathbb{B}^n$ en procédant comme suit :

$$\mathbf{b}\,\mathbf{G}$$

Attention : dans le calcul matriciel, les sommes sont des sommes modulo 2 ! ! !

Exemple : Codage avec contrôle de parité

Le mot 110 peut être codé en faisant :

$$(1 \quad 1 \quad 0) \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} = (1 \quad 1 \quad 0 \quad 0)$$

Le produit matriciel

Le produit matriciel tel que manipulé ici se définit de la manière suivante. Soit les matrices $\mathbf{A} = (a_{ij})$ (avec $1 \leq i \leq n$ et $1 \leq j \leq p$) et $\mathbf{B} = (b_{ij})$ (avec $1 \leq i \leq p$ et $1 \leq j \leq m$), la matrice $\mathbf{C} = \mathbf{AB}$ est une matrice $\mathbf{C} = (c_{ij})$ telle que

$$\begin{aligned} c_{ij} &= \sum_{k=1}^{p} (a_{ik}\, b_{kj}) \\ &= a_{i1}\, b_{1j} \oplus a_{i2}\, b_{2j} \oplus \ldots \oplus a_{ip}\, b_{pj}, \end{aligned}$$

pour $1 \leq i \leq n$ et $1 \leq j \leq m$.

Dans le calcul matriciel, les sommes sont des sommes modulo 2 ! ! !

Le symbole $\sum$ est donc un $\bigoplus$, comme détaillé dans la seconde égalité.

Exercice

Soit le (2,4)-code linéaire de matrice génératrice

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

C'est la matrice génératrice associée au codage $\Phi : \mathbb{B}^2 \to \mathbb{B}^4$.

dont le code est $\{0000, 1011, 0101, 1110\}$.

Que faire si le message réceptionné est 1101 ?

Correction par le tableau standard

Un outil...

Le **tableau standard** est un tableau qui permet de repérer rapidement le mot-code le plus proche d'un message afin de le corriger en respectant le *principe du maximum de vraisemblance*.

Expliquons-en le principe : codage linéaire $\Phi : \mathbb{B}^n \to \mathbb{B}^p$

- 1ère ligne du **tableau** : les mots-code ;
  les autres lignes : les autres mots de $\mathbb{B}^p$.
- tout **message erroné** est corrigé en utilisant le **mot-code** (1ère ligne) de la même colonne où se trouve le **mot erroné**.

Par définition, le tableau standard d'un codage $\Phi : \mathbb{B}^n \to \mathbb{B}^p$ comportera 2^n colonnes et $2^{(p-n)}$ lignes

Définition

Soit un codage linéaire $\Phi : \mathbb{B}^n \to \mathbb{B}^p$ **Illustration :** $n = 3, p = 4$

Le **tableau standard** s'obtient en réalisant les étapes suivantes :

1. On classe **tous les mots** de $\mathbb{B}^p$ selon le poids de ceux-ci. Si deux mots ont même poids, on les classe par ordre lexicographique avec 0 avant 1. On obtient une liste $\mathcal{B}$.
2. Par définition, le tableau comportera 2^n colonnes et 2^{p-n} lignes.
3. Première ligne : mettre les **mots-code** tels qu'ils apparaissent dans $\mathcal{B}$ (ils remplissent la 1ère ligne car il y a 2^n mots-code).
4. Dans la ligne suivante, on place le premier **mot m** de la liste $\mathcal{B}$ qui n'est pas encore utilisé dans le remplissage du tableau.
5. On remplit ensuite cette ligne en inscrivant $\boxed{\mathbf{m} \oplus \mathbf{x}}$ dans la colonne ayant au sommet le mot du code **x**. On réitère les étapes 4 et 5 tant que des mots de $\mathbb{B}^p$ doivent encore être placés.

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

Étape 1

On classe **tous** les mots de $\mathbb{B}^4$ selon le poids de ceux-ci. Si deux mots ont le même poids, on les classe par ordre lexicographique avec 0 avant 1, c-à-d :

$\mathbb{B}^4$	Poids
0000	0
0001	1
0010	1
0011	2
0100	1
0101	2
0110	2
0111	3
1000	1
1001	2
1010	2
1011	3
1100	2
1101	3
1110	3
1111	4

On obtient alors une liste $\mathcal{B}$, qui correspond au classement suivant :

$$\mathcal{B} = 0000 \quad 0001 \quad 0010 \quad 0100 \quad 1000 \quad 0011 \quad 0101 \quad 0110 \quad 1001$$
$$1010 \quad 1100 \quad 0111 \quad 1011 \quad 1101 \quad 1110 \quad 1111$$

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

Étape 2

Par définition, le *tableau standard* comportera 2^n colonnes et $2^{(p-n)}$ lignes. Ici : $2^3 = 8$ colonnes et $2^{(4-3)} = 2$ lignes.

$\mathcal{B} =$ 0000 0001 0010 0100 1000 0011 0101 0110 1001
1010 1100 0111 1011 1101 1110 1111

Étape 3

Première ligne : mettre les **mots-code** tels qu'ils apparaissent dans $\mathcal{B}$ (ils remplissent la 1ère ligne car il y a 2^n mots-code), c-à-d ici :

0000	0011	0101	0110	1001	1010	1100	1111

ce qui donne la première ligne du tableau standard.

$\mathcal{B} =$ ~~0000~~ 0001 0010 0100 1000 ~~0011~~ ~~0101~~ ~~0110~~ ~~1001~~
~~1010~~ ~~1100~~ 0111 1011 1101 1110 ~~1111~~

$\mathcal{B} =$ ~~0000~~ 0001 0010 0100 1000 ~~0011~~ ~~0101~~ ~~0110~~ ~~1001~~
~~1010~~ ~~1100~~ 0111 1011 1101 1110 ~~1111~~

Étape 4

Dans la ligne suivante, on place le premier **mot m** de la liste $\mathcal{B}$ qui n'est pas encore utilisé dans le remplissage du tableau. On obtient :

0000	0011	0101	0110	1001	1010	1100	1111
0001							

$\mathcal{B} =$ ~~0000~~ ~~0001~~ 0010 0100 1000 ~~0011~~ ~~0101~~ ~~0110~~ ~~1001~~
~~1010~~ ~~1100~~ 0111 1011 1101 1110 ~~1111~~

0000	0011	0101	0110	1001	1010	1100	1111
0001							

Étape 5

On remplit alors la ligne en inscrivant $\boxed{\mathbf{m} \oplus \mathbf{x}}$ dans la colonne ayant au sommet le mot du code **x**.

On obtient alors :

0000	0011	0101	0110	1001	1010	1100	1111
0001	0010	0100	0111	1000	1011	1101	1110

$$\mathcal{B} = \cancel{0000}\ \cancel{0001}\ \cancel{0010}\ \cancel{0100}\ \cancel{1000}\ \cancel{0011}\ \cancel{0101}\ \cancel{0110}\ \cancel{1001}\ \cancel{1010}\ \cancel{1100}\ \cancel{0111}\ \cancel{1011}\ \cancel{1101}\ \cancel{1110}\ \cancel{1111}$$

$$\mathcal{B} = \cancel{0000}\ \ \cancel{0001}\ \ \cancel{0010}\ \ \cancel{0100}\ \ \cancel{1000}\ \ \cancel{0011}\ \ \cancel{0101}\ \ \cancel{0110}\ \ \cancel{1001}$$
$$\cancel{1010}\ \ \cancel{1100}\ \ \cancel{0111}\ \ \cancel{1011}\ \ \cancel{1101}\ \ \cancel{1110}\ \ \cancel{1111}$$

Étape 5 – Suite

Ré-itérations des étapes 4 et 5 jusqu'à épuiser $\mathcal{B}$.

Dans ce cas-ci, on a donc terminé !

On a donc obtenu :

0000	0011	0101	0110	1001	1010	1100	1111
0001	0010	0100	0111	1000	1011	1101	1110

qui est le tableau standard associé au codage $\Phi : \mathbb{B}^3 \to \mathbb{B}^4$

Comment utiliser le tableau standard...

Grâce au tableau, tout **message erroné** est corrigé en utilisant le **mot-code** (situé sur la première ligne) de la même colonne où se trouve le **mot erroné**.

Codage avec contrôle de parité

Tableau standard associé au codage avec contrôle de parité

$\Phi : \mathbb{B}^3 \to \mathbb{B}^4$

0000	0011	0101	0110	1001	1010	1100	1111
0001	0010	0100	0111	1000	1011	1101	1110

Ainsi, si le message réceptionné est 0111, il est facile de détecter qu'il est faux. Le principe du maximum de vraisemblance préconise de le corriger par 0110.

On a donc modifié le bit de poids le plus faible (le plus à droite).

Un autre exercice de construction du tableau standard :

Exercice de construction du tableau standard

Reprenons le (2,4)-code linéaire de matrice génératrice

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \quad \text{c-à-d le codage } \Phi : \mathbb{B}^2 \to \mathbb{B}^4.$$

dont le code est $\{0000, 1011, 0101, 1110\}$.

Son tableau standard est :

0000	0101	1011	1110
0001	0100	1010	1111
0010	0111	1001	1100
1000	1101	0011	0110

Que faire si le message réceptionné est 1001 ?

Correction par le tableau standard

Inconvénients de la méthode...

La méthode de correction par le tableau standard présente **plusieurs inconvénients** :

- Le tableau est long à construire.
- Dès que la taille des blocs est relativement importante ($\geq$ 30 environ), le tableau devient beaucoup trop gros pour être utilisable.
- La recherche du message reçu dans le tableau est lente.

Nous allons donc voir une deuxième méthode, plus algébrique, pour la correction des erreurs : la correction par **liste de** *syndromes*.

Motivation : un exemple de détection d'erreur...

On a utilisé un codage *linéaire systématique* donné par la matrice

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

On reçoit le message 011111.

Comment savoir s'il est correct ?

Codage *systématique*, donc mot d'origine : 011

qui "aurait dû être codé" : 011110

Comparons les bits de contrôle du message reçu et du message qui "aurait dû être codé" : on a 110 à la place de 111.

L'erreur entre les deux est $110 \oplus 111 = 001 \neq 000$

Conclusion : le message n'est pas correct !

Correction par **liste de syndromes**

Le syndrome correspond à la somme des bits de contrôle du message reçu et des bits de contrôle *recalculés*.

Un outil. . .

La liste de syndromes est un tableau qui permet de repérer rapidement le mot-code le plus proche d'un message afin de le corriger en respectant le *principe du maximum de vraisemblance*.

La liste des syndromes est plus petit que le tableau standard.

Principe

On calcule le **syndrome** du message obtenu par le récepteur.
Si ce syndrome est nul, alors le message reçu est bien un mot-code.
Dans le cas contraire, la **liste des syndromes** nous indiquera la correction à appliquer au message erroné.

Le syndrome s'obtient grâce à la **matrice de contrôle**.

Définitions

- La **matrice de contrôle**, notée **H**, est une matrice à $p - n$ lignes et à p colonnes obtenue en juxtaposant la transposée de la matrice de parité **P** avec une matrice identité de dimension $p - n$.
  On a donc : $$\mathbf{H} = \left(\mathbf{P}^T \,\middle|\, I_{p-n} \right)$$
- On note $\mathbf{H}'$ la transposée $\left(\mathbf{H}^T\right)$ de la matrice de contrôle
  On a donc : $$\mathbf{H}' = \left(\begin{array}{c} \mathbf{P} \\ \hline I_{p-n} \end{array} \right)$$
- Le **syndrome** d'un message **m**, noté $\sigma(\mathbf{m})$, est

$$\boxed{\sigma(\mathbf{m}) \overset{\text{déf.}}{=} \mathbf{m}\,\mathbf{H}'}\,,$$

Le syndrome est donc un mot de longueur $p - n$.

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

La matrice génératrice était $\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

Dès lors, la matrice de parité est $\mathbf{P} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

Ainsi, la matrice de contrôle est :

$$\mathbf{H} = (1 \quad 1 \quad 1 \quad 1)$$

et la transposée de la matrice de contrôle est : $\mathbf{H'} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

La matrice génératrice était $\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

Ainsi, la matrice de contrôle est :

$$\mathbf{H} = (1 \quad 1 \quad 1 \quad 1)$$

et la transposée de la matrice de contrôle est : $\mathbf{H'} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

Le **syndrome** du mot $\mathbf{m} = 1100$ est

$$\sigma(\mathbf{m}) \overset{\text{déf.}}{=} \mathbf{m}\,\mathbf{H}' = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \end{pmatrix}$$

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

La matrice génératrice était $\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$

Ainsi, la matrice de contrôle est :

$$\mathbf{H} = (1 \quad 1 \quad 1 \quad 1)$$

et la transposée de la matrice de contrôle est : $\mathbf{H'} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$

Le **syndrome** du mot $\mathbf{m} = 1000$ est

$$\sigma(\mathbf{m}) \overset{\text{déf.}}{=} \mathbf{m}\,\mathbf{H}' = \begin{pmatrix} 1 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \end{pmatrix}$$

Autre exemple

On considère le codage *linéaire systématique* donné par la matrice

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{pmatrix}. \qquad \text{La matrice de parité est } \mathbf{P} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

La matrice de contrôle est $\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$

Sa transposée est $\mathbf{H}' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$

Autre exemple – Suite

On considère le codage *linéaire systématique* donné par la matrice

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

Matrice de contrôle $\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$, et $\mathbf{H}' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$

Le **syndrome** du mot $\mathbf{m} = 1100$ est

$$\sigma(\mathbf{m}) \overset{\text{déf.}}{=} \mathbf{m}\,\mathbf{H}' = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

Autre exemple – Suite

On considère le codage *linéaire systématique* donné par la matrice

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

Matrice de contrôle $\mathbf{H} = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{pmatrix}$, et $\mathbf{H}' = \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$

Le **syndrome** du mot $\mathbf{m} = 1110$ est

$$\sigma(\mathbf{m}) \overset{\text{déf.}}{=} \mathbf{m}\,\mathbf{H}' = \begin{pmatrix} 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

Dans le cas d'un codage linéaire systématique :

Propriété

Le syndrome de la somme modulo 2 de deux messages est la somme modulo 2 des syndromes de chaque message, c-à-d :

$$\sigma(\mathbf{m} \oplus \mathbf{n}) = \sigma(\mathbf{n}) \oplus \sigma(\mathbf{n}) \ .$$

On en déduit que...

1. Le syndrome du vecteur nul est nul.
   On peut l'observer par : $\sigma(\mathbf{0}) \overset{\text{déf.}}{=} \mathbf{0}\,\mathbf{H}'$,
2. Le syndrome d'un mot-code est nul.
   En effet, la somme de deux mots-code est un mot-code. Donc, le syndrome de ce mot-code est égal à la somme modulo 2 des syndromes des deux mots-code le composant. La seule opération possible $0 \oplus 0 = 0$.

On peut démontrer que...

...les messages (**mots) de** $\mathbb{B}^p$ peuvent être regroupés en sous-ensembles distincts (classes d'équivalence), **chacun correspondant à un syndrome particulier**.

Propriété

Il y a autant de syndromes différents que de lignes dans le tableau standard, soit 2^{p-n}

Définition

La **liste des syndromes** est un tableau de deux colonnes :

- dans la 1ère colonne se placent les syndromes (classés par ordre lexicographique) et
- dans la 2èmecolonne se trouve un message particulier de $\mathbb{B}^p$...

Pour obtenir la liste des syndromes :

La **liste des syndromes** s'obtient en réalisant les étapes suivantes :

1. dans la 1ère colonne, on place l'ensemble des syndromes possibles (tels que définis dans $\mathbb{B}^{p-n}$).
2. On classe les mots de $\mathbb{B}^p$ par poids croissant. À poids égal, on les classe par ordre lexicographique. On obtient $\mathcal{B}$.
3. Dans l'ordre ainsi déterminé, on calcule le syndrome de chaque mot jusqu'à obtenir tous les syndromes possibles. On retient dans le tableau le premier mot de $\mathcal{B}$ qui donne chaque syndrome.

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

Étape 1

L'ensemble des syndromes possible est 0 et 1 (élém. de $\mathbb{B}^{p-n}$).

Étape 2

Nous avons $\mathcal{B}$:

0000 0001 0010 0100 1000 0011 0101 0110

1001 1010 1100 0111 1011 1101 1110 1111

Étape 3

La liste des syndromes est ici :

Syndrome	Message
0	0000
1	0001

Correction par **liste de syndromes**

Correction...

Grâce à ce tableau, on peut procéder à la correction de message faux.

On procède comme décrit ci-dessous :

1. Calculer le syndrome du message reçu.
2. Corriger le message **en lui additionnant** le message répertorié dans la liste des syndromes pour l'entrée correspondant au syndrome obtenu.

Correction par **liste de syndromes**

Exemple : Codage avec contrôle de parité où $n = 3$ et $p = 4$

Dans le cas d'un codage avec contrôle de parité où $n = 3$ et $p = 4$, le récepteur reçoit le message suivant :

$$0010$$

Son syndrome est $\sigma(0010) = 1$.

Or, nous avions établi la liste des syndromes de ce codage :

Syndrome	Message
0	0000
1	0001

Dans cette liste, le message correspond au syndrome 1 est 0001.

Le message corrigé est donc 0011 , car $0010 \oplus 0001 = 0011$.

Exemple : Codage linéaire précisé par sa matrice génératrice

Soit le codage linéaire défini par la matrice génératrice suivante :

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix} \qquad \text{d'où } \mathbf{P} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

On en déduit directement que la matrice de contrôle est :

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

Exemple – suite

Les syndromes sont les éléments de $\mathbb{B}^{6-3}$, c-à-d :

000	001	010	100	011	101	110	111

Exemple – suite

On prend les mots de $\mathbb{B}^6$ par ordre de poids (et par ordre lexicographique), on calcule leur syndrome.

Dès qu'un syndrome non encore observé est obtenu, on place le mot dans la liste des syndromes.

La matrice de contrôle transposée est :

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

Exemple – suite

La matrice de contrôle transposée est :

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

La liste des syndromes est ici :

Syndrome	Message
000	000000
001	000001
010	000010
100	000100
011	100000
101	010000
110	001000
111	001001

Propriété

La condition nécessaire et suffisante **pour qu'un codage linéaire permette de corriger de façon certaine toutes les erreurs de poids 1** est que la matrice $\mathbf{H}'$ ait toutes ses lignes distinctes et non-nulles.

On en déduit...

Supposons qu'on travaille avec r **bits de contrôle**.

On a donc : $\Phi : \mathbb{B}^n \to \mathbb{B}^p$ avec $p - n = r$.

Le codage est entièrement déterminé par la matrice de parité **P** (de taille $n \times r$).

Il faut donc la choisir, pour que la matrice **H** respecte la condition énoncée, comme composée de n mots...

- de longueur r ,
- distincts entre eux,
- différents de la base canonique de $\mathbb{B}^r$,
- et distincts du vecteur nul.

Dans $\mathbb{B}^r$, reste le choix entre $2^r - r - 1$ mots.

Ces n mots distincts existent lorsque $n \leq 2^r - r - 1$.

Définition

Un **code de Hamming** est un code linéaire avec $n = 2^r - r - 1$ et $p = (n + r)$ dont la matrice $\mathbf{H}'$ est alors constituée de tous les mots binaires de longueur r non-nul.

Propriété

On peut démontrer que pour tout code de Hamming, nous avons $\delta = 3$.

Rappel : Table des matières :

Ce que noss avons vu. . .

1. Quelques **définitions** préalables nécessaires
2. pour **détecter, corriger** des erreurs de transmission.
3. On se concentre sur le **codage linéaire**.

Rappels des objectifs

Être capable...

- de préciser les propriétés d'un codage, tels que son rendement, sa distance minimale,...
- de déterminer si un codage est linéaire
- de manipuler les codages linéaires : construire le tableau standard, la matrice génératrice, la matrice de contrôle, la table des syndromes à partir des renseignements fournis,
- corriger en fonction tout message détecté comme erroné,
- déterminer si un code est un code de Hamming.