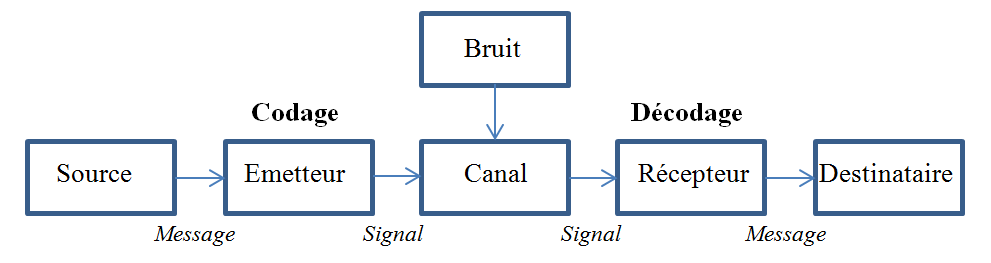
# Théorie des codes

## Introduction

Ce cours est dédié à l’étude des techniques qui permettent de coder sans ambiguïté et efficacement, par rapport aux objectifs que l’on désire atteindre, les messages émis par une source d’information.

Ces messages sont codés sur un alphabet différent de celui utilisé par la source, de manière telle à pouvoir être envoyés dans un canal de transmission.

Schéma de Shannon (Théorie de l’information)



Source => source d’information qui émet des messages

Codage => procédé qui permet de coder un message émis par la source

Canal => canal de transmission au travers duquel passe le message

Lors du passage dans le canal, le message peut subir des modifications.

On dit que le canal est soumis à un bruit ou que le canal est bruité.

Décalage => procédé qui permet de retrouver le message émis par la source pour être envoyé au Récepteur.

Le couple codage/décodage doit être adapté au canal et aux objectifs que l’on désire atteindre. Par exemple,

En cryptographie, on cherche à coder un message de manière telle qu’une personne non autorisée ne puisse pas avoir accès au message.

En correction d’erreurs, on cherche à coder un message de manière telle que l’o, puisse repérer et corriger les erreurs qui peuvent se produire lors du passage dans le canal.

En compression des données, on cherche à coder un message de manière telle à pouvoir réduire la taille du message.

Dans le codage à contraintes spectrales le canal de transmission est considéré comme un support physique que l’on modélise par l’ensemble des contraintes qui doivent satisfaire les messages pour pouvoir être transmis.

Un exemple est un canal de transmission qui ne suspecte pas les alternances trop rapide entre 0 et 1 (alphabet de source {0,1}) et qui ne fait pas la différence entre une longue suite de 0 et une logique suite de 1. Le codage, dans ce cas, consiste à transformer une suite quelconque de 0 et de 1 en une suite de 0 et de 1 qui n’a pas comme facteur 010101, 000000, et 111111. (Systèmes sofiques)

L’ordre de ces différents codages est important.

Par exemple, le codage de compression des données doit s’effectuer dans un canal sans bruit. Comme, en général, canaux de transmission sont bruités, on effectue un codage de correction d’erreur pour rendre le canal sans bruit.

Codage 🡪 Codage 🡪 Codage 🡪 Canal 🡪 Décodage 🡪 Décodage🡪 Décodage 🡪 Réception

Compression

Correction d’erreur

à contraintes spectrales

à corr. Spectrales

Correction d’erreur

Compression

Codage de source

A={a1,a2,…,an}=alphabet

A\* ensemble des mots sur A,y compris le mot vide, noté E.

U appartient à A\*, |u| = longeur de u = nombre de lettre qui composent u

E est le seul mot sans lettre, donc de longueur zéro, |E| = 0

A+=A\*/{E}

Définition

Un code C sur l’alphabet A est un ensemble de mots sur A qui ne contient pas le mot vide, CcA+

Les éléments de C sont appelées les mots de code.

Exemples :

C={0,10,100} est un code sur A{0,1}

C={ab,bba,bbaba,bc} est un code sur A={a,bc}

Un procédé simple de codage, appelé codage par blocs à longueur variable ou tout simplement codage à longueur variable, consiste à substituer chaque lettre de l’alphabet de source par un mot sur un autre alphabet. Un exemple trio-connu de ce type de codage est le codage de Morse où chaque lettre de l’alphabet latin est remplacée par un mot de longueur 3 sur l’alphabet composé par 3 éléments : point, trait, blanc.

Considérons une source qui émet un message sur un alphabet B (alphabet de source).

Soit C un ensemble de même cardinal que B, |C| = |B| et soit f :B🡪C une fonction bijective (appelé aussi fonction d’encodage). On appelle (B,C,f) un schéma de codage. On étend f à un morphisme f :B\*🡪C\*. Ainsi, le message u=b1b2…bk, bi appartient à B, pour tout i<i<=k.

Remarque

La bijectivité de f ne garantit pas que le décodage d’un message codé par le schéma (B,C,f) puisse se faire sans ambigüité.

Exemple : B = {a,b,c,…,z}, C={0,1,…,25}

Fonction d’encodage f : B🡪C

a|🡪0

b|🡪1

….

z|🡪x

Le mot reçu 1209 possède plusieurs décodages

(1)(2)(0)(9) 🡪bcaj

1209 (12)(0)(9) 🡪 maj ambiguité

(1)(20)(9) 🡪 buj

Définition

Un code CcA+ est uniquement déchiffrable ou non ambigu ssi pour tout couple de suites de mots de C, {xi}n,i=0, xi appartient à C, pour tout i<=i<=n, {yj}m, j=0, yj appartient à C, pour tout i<=j<= m, vérifiont x1,x2,…xn = y1y2…yn

On a n = m et xi = yi, pour tout 1<=i<=n

Exemples :

- C1 = {00,01,10,11} est uniquement déchiffrable.

Comme tous les mots de C1 ont la même longueur, toute concaténation C produit A de mots de C1 admet une seule décomposition comme produit de mots de C1.

11|00|10|11|01 appartient à C+,1

- C2 = {0,01,10} est ambigu (qui n’est pas uniquement déchiffrable.

010 0|10

01|0

Problèmes

1. Peut-on décider si un code est-uniquement déchiffrable ? oui
2. Si oui, comment ? algorithme

L’algorithme de Sardinas et Patterson (1950)

Soit CcA+ u code fini. On associe à C une suite d’ensemble de mots sur A, {Ri}i=0, que l’on appelle suite des restes droits.

Cette suite est définie de manière réccurente par :

R0 = C(-1)C\{E}

Ri+1 = C(-1)RiUR(-1,i)C, pour tout i=0

Rappel X,YcA+

X(-1)Y={ w appartient à A\*|il existe x appartient à X : xw appartient à Y }

X(-1)Y = Ensemble des restes droits de Y par X

Exemple : Y = {ab,babc,cab,ccb}, X={a,bc,ca,ccb}

X(-1)Y={b,E}

Exemples :

1. C={a,ba,bb}

R0 = C(-1)C\{E} = vide 🡪 Ri = vide, pour tout i >= 0

On appelle C un code préfixe.

Définition

C est un code préfixe ssi aucun mot de C ne commence par un autre mot de C.

Autrement dit, aucun mot de C n’est préfixe d’un autre mot de C.

Ramarque : Si C est préfixe, la suite des restes droits associée à C est composée par le seul ensemble vide.

1. C = {a,ab,bb}

{C n’est pas préfixe car ab commence par a}

R0 = C(-1)C\{E}={b}

R1 = C(-1)R0 U R(-1,0)C = vide U {b} = {b}

🡪 Ri = {b}, pour tout i>=0

La suite des restes droits associée à C est composée par le seul ensemble {b}

1. C={ab,abcd,cdab}

R0=C(-1)C\{E}={cd}

R1=C(-1)R0 U R0(-1)C = vide U {ab} = {ab}

R2 = C(-1)R1 U R1(-1)C = {E}U{E,cd}={E,cd}

R3 = C(-1)R2UR2(-1)C= vide U C U {ab} = C

CU{ab} = R(-1,2)C

R4 =C(-1)C U C(-1)C = {E,cd}

R4=R2 🡪 la suite est périodique