

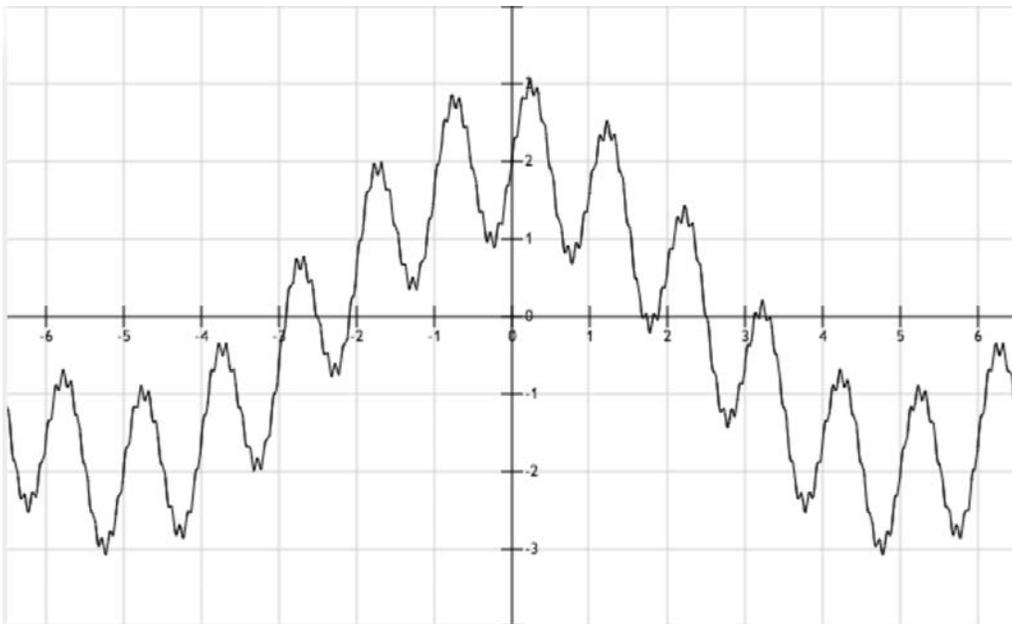
M2.L4: Série d'exercices sur la compression de données

1 Filtre Passe-Bande

Nous définissons le *filtre passe bande idéal* de **fréquence de coupure haute f_{ch}** et **fréquence de coupure basse f_{cb}** -- telles que $f_{ch} > f_{cb}$ --, le filtre qui ne conserve que les signaux de fréquence compris entre f_{cb} et f_{ch} .

1. Comment le filtre passe-bande s'exprime-t'il en fonction d'un passe haut et d'un passe bas?

Soit le signal périodique $f(x)$ suivant:



$$f(x) = 2 \cdot \cos(0.1 \cdot 2\pi \cdot x) + \sin(1 \cdot 2\pi \cdot x) + 0.1 \cdot \sin(10 \cdot 2\pi \cdot x)$$

2. Comment conserver seulement les 2 basses fréquences ? A quoi ressemble le signal résultant ? Le dessiner.

3. Comment conserver seulement les 2 hautes fréquences ? A quoi ressemble le signal résultant ? Le dessiner.

4. Comment conserver seulement la fréquence intermédiaire ? A quoi ressemble le signal résultant ? Le dessiner.

2 Algorithme de Shannon-Fano

- a) En utilisant l'algorithme de Shannon-Fano, représentez la séquence suivante (sans tenir compte des espaces) par une séquence de bits:

INFORMATION CALCUL ET COMMUNICATION

- b) Combien de bits par lettre en moyenne sont-ils nécessaires pour représenter cette séquence?
c) Calculez l'entropie de la séquence. Vérifie-t-on les inégalités vues au cours?

Indication: Si le calcul de l'entropie vous fatigue, essayez <http://www.shannonentropy.netmark.pl/> (attention à supprimer les espaces!)

- d) Si vous vous restreignez à utiliser un code qui ne tient pas compte des probabilités d'apparition et qui utilise exactement le même nombre de bits pour chaque lettre, de combien de bits aurez-vous besoin pour représenter la séquence?

Voici une autre séquence de lettres (où on oublie à nouveau les espaces, les accents, les traits d'union, les virgules et les apostrophes!):

DIDON DINA, DIT-ON, DU DOS D'UN DODU DINDON

- e) A priori, pouvez-vous deviner laquelle des deux séquences ci-dessus a la plus faible entropie?
f) Représentez à nouveau la séquence par une séquence de bits en utilisant l'algorithme de Shannon-Fano.
g) Répondez à nouveau aux questions c) et d) pour cette deuxième séquence de lettres.

3 Codage par plages (run-length encoding)

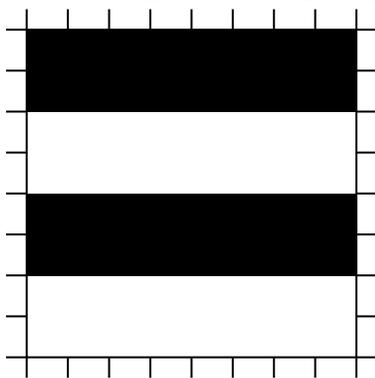


Image 1

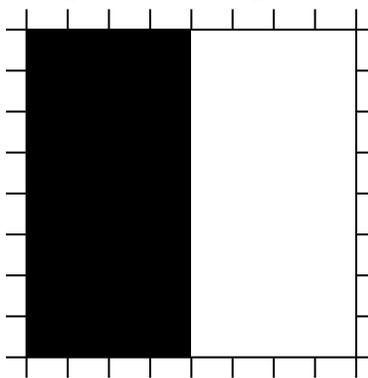


Image 2

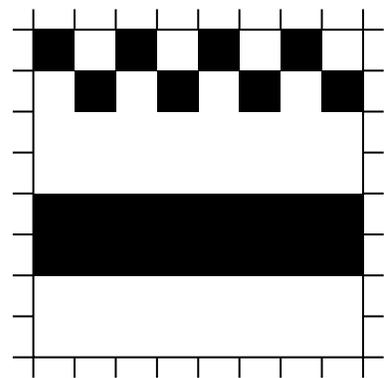


Image 3

Dans cet exercice, on considère un autre type de compression, utilisé principalement pour les images en noir et blanc comme celles présentées ci-dessus. Il s'agit du *codage par plages* ou run length encoding

L'algorithme procède alors itérativement comme suit :

1. Trouver les 2 lettres les moins fréquentes et regrouper les sous une question qui les distingue. En cas d'égalité, en choisir 2 au hasard (ça ne change pas la longueur moyenne du code).

Par exemple, si les 2 lettres les moins fréquentes sont A et F, la question serait (par exemple) est-ce A ?

2. Dans le tableau des nombres d'apparitions des lettres, supprimer les 2 dernières lettres considérées et les regrouper comme une seule nouvelle lettre, avec comme nombre d'apparitions la somme des deux nombres.

Pour continuer sur l'exemple précédent, si A apparaissait 1 fois et F 2 fois, alors on supprime A et F du tableau et on introduit la nouvelle lettre A|F (qui veut dire A ou F) avec comme nombre d'apparitions 3.

3. Recommencer en 1 tant qu'il y a des lettres .

Une fois l'arbre des questions construit, procéder comme pour le code de Shannon-Fano en décidant d'affecter le symbole 0 à toutes les réponses non et le symbole 1 à toutes les réponses oui .

4.2 Exemple

Considérons les lettres A,B,C,D et E avec les probabilités respectives de $\frac{1}{3}$, $\frac{1}{8}$, $\frac{1}{8}$, $\frac{1}{4}$ et $\frac{1}{6}$; par exemple dans (sans les espaces) :

A A AD AD ADE ABCDE ABCDE ABCDE

On commence donc avec le tableau

	A	B	C	D	E
probabilité	$\frac{1}{3}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{6}$

On regroupe donc B et C (les deux moins probables) sous une question est-ce C ? . Puis on les regroupe dans le tableau pour recommencer :

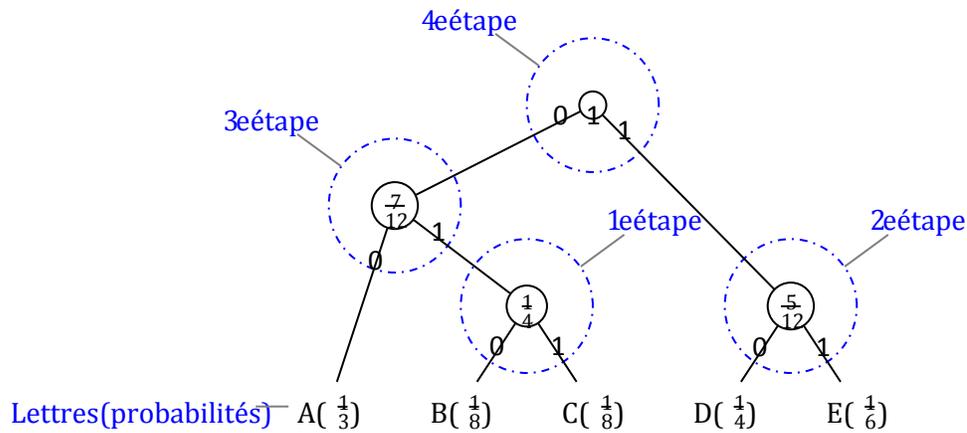


Figure 1: Exemple de code de Huffman.

	A	B—C	D	E
probabilité	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$

On recommence donc avec les 2 lettres les moins probables, donc E ($\frac{1}{6}$) et, par exemple, D ($\frac{1}{4}$). A la place de D, on aurait aussi pu choisir ici la nouvelle lettre B—C qui a aussi une probabilité de $\frac{1}{4}$. Il existe plusieurs codes de Huffman équivalents pour une même séquence. Il suffit d'en choisir un (et bien sûr d'en convenir avec son destinataire, comme pour tout code !).

Et on continue ainsi de suite. L'arbre des questions correspondant est donné en Figure 1, où au lieu d'écrire les questions elles-mêmes nous avons représenté les probabilités sommées de toutes les alternatives couvertes par chaque question.

Les mots de codes correspondants sont alors (pour A, B, C, D et E respectivement) : 00, 010, 011, 10 et 11.

4.3 Mise en pratique

Supposons que nous ayons un texte avec les nombres d'apparitions suivants :

	A	B	C	D	E
nb. app.	39	17	16	15	13

1. Construire le code de Shannon-Fano correspondant.
2. Construire le code de Huffman correspondant.
3. Quelle est la longueur moyenne de chaque code ? Comparer à la limite théorique donnée par le théorème de Shannon.