

Compression de données - rattrapage

G. TOCHON

Janvier 2018

Durée totale : 2h

Tous les documents sont interdits. L'utilisation d'une calculatrice non programmable est autorisée.

Compression conservative

bzip2 est un algorithme de compression sans pertes de données, développé entre 1996 et 2000 par Julian Seward afin de concurrencer l'algorithme LZW et ses variantes (notamment l'algorithme **deflate** sur lequel repose **gzip**). Bien que plus complexe et moins rapide, **bzip2** offre des performances de compression supérieures à ses concurrents. Étant donnée une chaîne de caractères à compresser, les 3 étapes principales de **bzip2** sont :

1. La transformée de Burrows-Wheeler de la chaîne de caractères.
2. La transformée *Move-to-front* du résultat de l'étape 1.
3. L'encodage de Huffman du résultat de l'étape 2.

1.1 Rappelez, en quelques mots, l'objectif et l'intérêt de la transformée de Burrows-Wheeler.

1.2 Explicitiez les différentes étapes du calcul de la transformée de Burrows-Wheeler de la chaîne de caractères **ananas\$** (on supposera que le symbole de fin de ligne **\$** précède toutes les autres lettres pour l'ordre lexicographique).

On suppose que le résultat de la **question 1.2** est **s\$nnaaa**.

2.1 Calculez la transformée *Move-to-front* de **s\$nnaaa**, en supposant que la liste initiale est $\mathcal{L}_0 = [\$, a, n, s]$.

2.2 Quel est l'état de la liste \mathcal{L} à la fin de la transformée *Move-to-front* ?

On suppose que le résultat de la transformée *Move-to-front* est $F = \{5, 4, 4, 5, 0, 0, 4, 1, 5, 0, 0, 0, 0\}^1$.

3.1 Calculez l'encodage de Huffman des symboles contenus dans F (on suppose que la distribution de probabilité des symboles est donnée par leur fréquence relative d'apparition dans le fichier).

3.2 Combien de bits permet de gagner la compression du fichier F ?

Compression non conservative

1 Énoncez le théorème d'échantillonnage de Shannon. Que se passe-t-il si la condition de Nyquist n'est pas respectée ?

2 Expliquez en quelques mots (ou, encore mieux, avec un dessin) l'interpolation par plus proche voisin. Quel est le principal défaut d'une image après interpolation par plus proche voisin ?

3 Quel est l'intérêt majeur de passer dans le domaine fréquentiel pour compresser une image avec pertes ?

4 Rappelez les 4 étapes de la compression JPEG d'une image en niveaux de gris.

5 Quels sont les deux artefacts de la compression/décompression JPEG pour une image ?

6 Comment se nomment les composantes Y, U et V de l'espace colorimétrique YUV ? Quel est son intérêt (par rapport à l'espace RGB classique) pour la compression JPEG d'une image couleur ?

(bonus) Écrire la décomposition en série de Fourier d'une fonction f de période T .

1. C'est en fait le résultat de Burrows-Wheeler + MTF sur la chaîne de caractères **bananeananas\$**, donc pas de panique si vous n'avez pas trouvé ça à la **question 2.1**.