Projet n°4 (codes de Hamming concaténés)

L'objectif de ce projet est de comprendre les codes de Hamming d'ordre quelconque et les possibilités offertes par leur concaténation.

1. On définit le code de Hamming d'ordre M ($M \geqslant 2$) de la façon suivante. Soit H la matrice binaire à M lignes et $N = 2^M - 1$ colonnes, dont les vecteurs colonnes sont constitués des entiers 1 à N codés en binaire. Pour $M = 3$ par exemple, on a
$$H = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$
Le dictionnaire C (ensemble des mots de codes) associé est l'ensemble des vecteurs binaires x tels que $Hx = 0 \mod (2)$. Comme la matrice H est de rang plein et de taille $M \times N$, son noyau est de dimension $K = N - M$ et il y a donc exactement 2^K mots dans le dictionnaire. Le code de Hamming d'ordre M code donc des mots de K bits en N bits. Quel code retrouve-t-on pour pour $M = 3$? Et pour $M = 2$? Quel est le débit $R(M)$ du code de Hamming d'ordre M ? Montrer qu'avec un tel code on peut corriger parfaitement toute erreur qui ne modifie qu'un bit (sur les N) de chaque mot transmis, puis implémenter le décodage de Hamming d'ordre M (on pourra s'inspirer de la fonction decode_hamming74 donnée en complément sur le site web du cours).

2. On se place dans le cas du canal binaire symétrique de paramètre p (chaque bit passant dans le canal est donc altéré avec probabilité p, et transmis correctement avec la probabilité $1-p$). Dans toute la suite on prendra comme valeur numérique $p = 0,03$. Quelle est la capacité de ce canal ? On souhaite calculer $p_b(M)$, la probabilité d'erreur par bit (et non par bloc) obtenue après codage de Hamming, passage dans le canal, et décodage. Montrer que $p_b(M)$ est l'espérance (nombre moyen) de bits à 1 obtenus après décodage d'une séquence binaire aléatoire où chaque bit est tiré indépendamment à 1 avec la probabilité p et à 0 avec la probabilité $1-p$. Estimer numériquement $p_b(M)$ pour $M = 2, 3, 4, 5, 6, 7$ (on utilisera la propriété précédente avec une séquence aléatoire très longue, i.e. plusieurs centaines de milliers de bits). On prendra soin de donner un ordre de grandeur empirique de la précision avec laquelle $p_b(M)$ est estimé (refaire l'estimation plusieurs fois et examiner les valeurs obtenues). Que penser de la performance obtenue pour $M = 6$ et $M = 7$? Représenter dans le plan débit/distortion

(R, p_b) les performances des codes de Hamming d'ordre 2 à 7. On pourra utiliser une représentation avec des barres d'erreurs pour $p_b(M)$.

3. On s'intéresse maintenant à la concaténation de deux codes de Hamming d'ordre M_1 et M_2, notée $M_1 \circ M_2$. Un message x est encodé en y à l'aide du code de Hamming d'ordre M_2, puis y est à son tour encodé en z à l'aide du code de Hamming d'ordre M_1. Le message z passe alors dans le canal bruité, ressort en z', qui est décodé en y' par M_1, avant que y' ne soit décodé en x' par M_2. Calculer le débit du code concaténé en fonction de M_1 et M_2. Du point de vue de la distortion (p_b), a-t-on intérêt à décoder d'abord avec le code le plus robuste $(M_1 < M_2)$ ou avec le moins robuste $(M_1 > M_2)$? Valider la réponse numériquement en calculant la distortion p_b pour $(M_1, M_2) = (2, 3)$ puis $(M_1, M_2) = (3, 2)$.

4. On considère le cas $M_1 = M_2 = 2$. Quel est le codage associé à cette concaténation ? Quel est le décodage optimal pour ce code ? Montrer que le décodage séquentiel proposé à la question précédente n'est pas optimal, si possible sans calcul dans un premier temps, puis en calculant explicitement p_b dans le cas du décodage séquentiel puis dans le cas du décodage optimal.

5. On considère la concaténation successive de codes de Hamming d'ordres M_k, $k = 2, 3, \ldots$. Exhiber une suite $(M_k)_{k \in \mathbb{N}^*}$ pour laquelle le débit limite (i.e. obtenu avec une concaténation infinie) n'est pas nul (calculer numériquement ce débit limite).

6. Explorer numériquement la combinatoire des codes de Hamming concaténés, et représenter chaque combinaison qui vous semble pertinente dans le plan (R, p_b). On pourra éventuellement utiliser une échelle logarithmique pour p_b.