le cnam

Codes et Automates finis

Cours 5 Matrice de contrôle

- Introduction

On se donne un code linéaire φ de dimension k et de longueur $n \geq k$: $u = \varphi(a)$, $a \in (\mathbb{F}_2)^k$, $u \in (\mathbb{F}_2)^n$. Le rendement de ce code est défini par le rapport $\frac{k}{n}$.
La matrice génératrice G s'obtient en calculant les vecteurs $c_j = \varphi(e_j)$ pour $1 \leq j \leq k$ avec $e_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ vecteur de base de l'espace $(\mathbb{F}_2)^k$. Les vecteurs $c_j \in (\mathbb{F}_2)^n$ forment les lignes de la matrice génératrice. Sous une forme standard, on écrit $G = (\mathrm{I}_k \; P)$ où I_k est la matrice identité à k lignes et k colonnes et P est la matrice de parité qui comporte k lignes et $(n-k)$ colonnes.

- Matrice de contrôle

La matrice de contrôle, notée le plus souvent H, comporte $(n-k)$ lignes et n colonnes. Elle s'écrit à l'aide de la matrice de parité P: $H = (-P^{\mathrm{t}} \; \mathrm{I}_{n-k})$.
Nous remarquons que dans le corps $\mathbb{F}_2$, on a bien sûr $-P^{\mathrm{t}} = P^{\mathrm{t}}$. Mais nous gardons le signe "moins" dans la définition générale car il permet de mettre en place une propriété qui sera énoncée plus loin.

- Exemples de matrices de contrôle

Nous calculons la matrice de contrôle H pour les exemples introduits jusqu'ici dans le cours.
Exemple 1. Duplication d'un bit : $k = 1$ et $n = 2$.
On a $G = (1 \; 1)$, $P = (1)$ et $H = (1 \; 1)$.
Exemple 2. Double répétition : $k = 1$ et $n = 3$.
On a $G = (1 \; 1 \; 1)$, $P = (1 \; 1)$ et $H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$.
Exemple 3. Duplication de deux bits : $k = 2$ et $n = 4$.
On a $G = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $H = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$.
Exemple 4. Contrôle de parité : $k = 3$ et $n = 4$.
On a $G = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ et $H = (1 \; 1 \; 1 \; 1)$.
Exemple 5. Code Hamming H7 : $k = 4$ et $n = 7$.
On a $G = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$, $P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ et $H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$.

- Propriété fondamentale de la matrice de contrôle

On a toujours $G\,H^{\rm t} = 0$.
On vérifie sans difficulté cette propriété pour les cinq cas particuliers proposés ci-dessus.
Dans le cas général, il suffit d'effectuer le produit par blocs de $G = (\,{\rm I}_k \;\; P\,)$ par $H^{\rm t} = \begin{pmatrix} -P \\ {\rm I}_{n-k} \end{pmatrix}$.
Le calcul d'un produit de deux matrices par blocs est identique au calcul d'un produit de deux matrices dans le cas où les éléments de matrice sont des nombres. Mais il faut au préalable s'assurer que chacun des produits matriciels a bien un sens. C'est le cas ici : ${\rm I}_k\,P$ est bien défini et $P\,{\rm I}_{n-k}$ l'est aussi car la matrice de parité P est à k lignes et $(n-k)$ colonnes. Par contre, si $n-k \neq k$, les produits de matrice $P\,{\rm I}_k$ et ${\rm I}_{n-k}\,P$ ne sont pas définis !
Le résultat du produit de la matrice G à k lignes et n colonnes par la matrice $H^{\rm t}$ à n lignes et k colonnes est une matrice à k lignes et k colonnes et on a :
$G\,H^{\rm t} = {\rm I}_k\,(-P) + P\,{\rm I}_{n-k} = -P + P = 0$.

- Syndrome

Si $v \in (\mathbb{F}_2)^n$, le syndrome $s(v)$ est défini par $s(v) = v\,H^{\rm t}$. C'est une application de $(\mathbb{F}_2)^n$ dans $(\mathbb{F}_2)^{n-k}$.
Le syndrome d'un mot du code $v \in \mathscr{C}$ est toujours nul.
Par exemple pour la duplication de deux bits ($k=2$, $n=4$), si $v = (1\;0\;0\;1)$, alors

$$s(v) = (1\;0\;0\;1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} = (1\;1).$$

Autre exemple avec le contrôle de parité ($k=3$, $n=4$). Si $v = (1\;1\;1\;0)$, alors

$$s(v) = (1\;1\;1\;0) \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = (1).$$

- Linéarité

L'application syndrome $(\mathbb{F}_2)^{n-k} \ni v \longmapsto s(v) = v\,H^{\rm t}$ est linéaire.

- Caractérisation des mots du code

Rappelons que l'ensemble $\mathscr{C}$ des mots du code est défini par
$\mathscr{C} = \{\varphi(a), a \in (\mathbb{F}_2)^k\} = \{a\,G, a \in (\mathbb{F}_2)^k\}$. C'est un sous-ensemble de $(\mathbb{F}_2)^n$. Si l'ensemble $\mathscr{C}$ des mots du code résulte d'un codage linéaire, on a la propriété suivante.
Un mot $v \in (\mathbb{F}_2)^n$ est un mot du code ($v \in \mathscr{C}$) si et seulement si son syndrome est nul ($s(v) = 0$) ; on a l'équivalence $(v \in \mathscr{C}) \Longleftrightarrow (v\,H^{\rm t} = 0)$.
Le syndrome permet de s'intéresser au vecteur d'erreur γ tel que $v = u + \gamma$ avec $u \in \mathscr{C}$.
Par linéarité, on a $s(v) = s(\gamma)$ puisque $u \in \mathscr{C}$.
Attention ! La décomposition $v = u + \gamma$ avec $u \in \mathscr{C}$ n'est pas unique en général. Dans l'exemple suivant, pour la duplication de deux bits, on a d'une part
$v = (1\;0\;1\;1) = (1\;0\;1\;0) + (0\;0\;0\;1)$ et d'autre part $v = (1\;0\;1\;1) = (1\;1\;1\;1) + (0\;1\;0\;0)$.
Les vecteurs d'erreurs $\gamma_1 = (0\;0\;0\;1)$ et $\gamma_2 = (0\;1\;0\;0)$ sont tous deux de poids égal à un.

Si $v = u + \gamma$, alors la distance de Hamming $d(u,v)$ est égale au poids $|\gamma|$ de l'erreur γ: c'est le nombre de bits où u et v diffèrent.

- Corriger un mot reçu

Afin de corriger un mot reçu $v \notin \mathscr{C}$, on a besoin d'utiliser la distance minimale d du code linéaire φ. Deux mots différents du code $\mathscr{C}$ different d'au moins d bits. On a vu lors de la leçon précédente que l'on a $d = \min\{|u|, u \in \mathscr{C} \text{ et } u \neq 0\}$ pour un code linéaire. La distance minimale est le plus petit des poids $|u|$ des mots du code non nuls.

- Une inégalité entre dimension, longueur et distance minimale

Pour un code linéaire de dimension k, de longueur n et de distance minimale d, on a l'inégalité suivante $d \leq n+1-k$.

Pour les cinq codes déja introduits dans ce cours, on peut vérifier cette inégalité.

Exemple 1. Ajout d'un bit de contrôle : $k = 1$ et $n = 2$. On a $\mathscr{C} = \{00, 11\}$ et $d = 2$. On a bien $2 \leq 2+1-1$.

Exemple 2. Double répétition : $k = 1$ et $n = 3$. Dans ce cas, $\mathscr{C} = \{000, 111\}$ et $d = 3$. On vérifie que $3 \leq 3+1-1$.

Exemple 3. Duplication de deux bits : $k = 2$ et $n = 4$. Le code $\mathscr{C}$ comporte 2^2 mots : $\mathscr{C} = \{0000, 0101, 1010, 1111\}$ et $d = 2$. On a bien la relation $2 \leq 4+1-2$.

Exemple 4. Contrôle de parité : $k = 3$ et $n = 4$. Le code $\mathscr{C}$ est composé de 2^3 mots : $\mathscr{C} = \{0000, 1001, 0101, 0011, 1100, 1010, 0110, 1111\}$. Il est clair que $d = 2$ et $2 \leq 4+1-3$.

Exemple 5. Code Hamming H7 : $k = 4$ et $n = 7$. On a vu lors de la leçon précédente que $d = 3$. De plus, $3 \leq 7+1-4$.

- Code détecteur

Un code $\mathscr{C} \subset (\mathbb{F}_2)^n$ de distance minimale d peut détecter à coup sûr au plus $(d-1)$ erreurs de transmission. En d'autres termes, si $d \geq 2$ et si $v \in (\mathbb{F}_2)^n$ peut s'écrire sous la forme $v = u + \gamma$ avec $u \in \mathscr{C}$ et $1 \leq |\gamma| \leq d-1$, on est certain que $v \notin \mathscr{C}$.

La chaîne de bits reçue n'appartient pas à l'ensemble des mots du code et elle est assez proche de l'ensemble $\mathscr{C}$. Alors on est cartain qu'il y a eu au moins une erreur de transmission.

- Une notation

On rapproche les trois nombres entiers de longueur, dimension et distance minimale sous la forme $[n, k, d]$.

- Code correcteur ?

Quand on reçoit le mot de quatre bits $v = (1\ 0\ 1\ 1)$ pour le codage *via* la duplication de deux bits, nous avons vu plus haut que l'on a à la fois $(1\ 0\ 1\ 1) = (1\ 0\ 1\ 0) + (0\ 0\ 0\ 1)$, avec $(1\ 0\ 1\ 0) \in \mathscr{C}$ et $\gamma_1 = (0\ 0\ 0\ 1)$ et $(1\ 0\ 1\ 1) = (1\ 1\ 1\ 1) + (0\ 1\ 0\ 0)$, avec $(1\ 1\ 1\ 1) \in \mathscr{C}$ et $\gamma_2 = (0\ 1\ 0\ 0)$.

On ne sais pas dire que l'une des erreurs γ_1 ou γ_2 est plus petite que l'autre car elles sont toutes deux de poids égal à un.

Dans le cas contraire, on peut projeter le mots $v \in (\mathbb{F}_2)^n$ sur les mots du code, c'est à dire trouver $u \in \mathscr{C}$ unique de distance minimale. Alors $u \in \mathscr{C}$, $d(u,v) \leq d(\widetilde{u}, v)$, $\forall \widetilde{u} \in \mathscr{C}$ et de plus si $\widetilde{u} \in \mathscr{C}$ est différent de u, alors $d(\widetilde{u}, v) \geq d(u,v) + 1$. Dans ce cas, l'erreur $\gamma = v - u$

comporte moins de bits égaux a 1 que toutes les autres erreurs $\widetilde{\gamma} = v - \widetilde{u}$. C'est l'erreur la plus probable. On corrige en général le mot reçu v en le remplaçant par le mot du code $u \in \mathscr{C}$ le plus proche. On parle dans ce cas d'un code correcteur.

- Distance minimale nécessaire d'un code correcteur

On se donne un entier $t \geq 0$ et un code de distance minimale d. Ce code corrige t erreurs si et seulement si $d \geq 2t+1$.

Parmi les cinq exemples de codes linéaires considérés dans ce chapitre, seuls la double répétition ([3, 1 , 3]) et le code de Hamming H7 ([7, 4 , 3]) satisfont à $d \geq 3$ et peuvent corriger au plus une erreur sans ambiguité.

- Un critère pratique

Un code linéaire corrige une erreur ($t = 1$) si et seulement si les colonnes de sa matrice de contrôle H sont distinctes et non nulles.

On vérifie que ce critère est effectivement satisfait pour les cinq codes introduits jusqu'ici.

Exemple 1. Ajout d'un bit de contrôle : $d = 2$ et $H = (1\ 1)$. Cette matrice a deux colonnes égales. Ce code est détecteur et détecte au plus une erreur.

Exemple 2. Double répétition : $d = 3$. On a maintenant $H = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$. Les trois colonnes de la matrice de contrôle H sont différentes. Ce code corrige une erreur et en détecte deux. Par contre, son rendement (1/3) est faible.

Exemple 3. Duplication de deux bits : $d = 2$. On a cette fois $H = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$. Cette matrice a deux paires de colonnes égales. Ce code n'est pas correcteur et détecte a coup sûr au plus une erreur.

Exemple 4. Contrôle de parité : $d = 2$. On a vu que $H = (1\ 1\ 1\ 1)$. Toutes les colonnes sont égales. Ce code détecte au plus une erreur.

Exemple 5. Code Hamming H7 : $d = 3$ et $H = \begin{pmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$. Cette matrice a toutes ses colonnes différentes. Ce code détecte de façon certaine au plus deux erreurs. Il en corrige au plus une en choisissant l'erreur la plus probable. Son rendement (3/7) est supérieur à celui du codage par double répétition.

Exercice

- Étude d'un code linéaire

On code des blocs de trois bits de la façon suivante : le bloc $b_1\,b_2\,b_3$ est codé $b_1\,b_2\,b_3\,a_1\,a_2\,a_3$ ou les trois derniers bits sont donnés par les relations $a_1 = b_1 + b_2$, $a_2 = b_1 + b_3$ et $a_3 = b_1 + b_2 + b_3$.

a) Donner la dimension k et la longueur n de ce code.

b) Montrer que ce code est linéaire.

c) Déterminer sa matrice génératrice G.

d) Déterminer sa matrice de contrôle H. $\left[H = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix} \right]$

e) Ecrire la liste $\mathscr{C}$ des mots du code.

f) Quelle est la distance minimale d de ce code ? [3]

g) Pour $\varepsilon_j = (0, \ldots, 0, 1, 0, \ldots 0)$ vecteur de base de $(\mathbb{F}_2)^6$, calculer le syndrome $s(\varepsilon_j)$.

On reçoit les messages suivants : $m_1 = 110110$, $m_2 = 011001$ et $m_3 = 111111$.

h) Calculer leurs syndromes et, le cas échéant, les corriger.

Le canal de transmission est supposé symétrique, sans mémoire et on note p la probabilité d'erreur dans la transmission de un bit.

i) Quelle est la probabilité de se tromper en corrigeant le mot m_2 ?

j) Donner l'ordre de grandeur de cette probabilité si $p = \frac{1}{100}$. [$\simeq 2\%$]