Une application intéressante des matrices : le chiffre de Hill

Didier Müller, Lycée Cantonal de Porrentruy

Article paru dans le bulletin No 90 de la SSPMP (www.vsmp.ch), octobre 2002

Résumé

Dans cet article, nous présentons le **chiffre de Hill**. C'est une méthode de chiffrement qui utilise des matrices carrées. Nous pensons que c'est une manière attrayante pour les élèves de se familiariser au calcul matriciel, au calcul modulo *n* et à la notion d'algorithme.

Introduction

Les élèves posent souvent la même question : « Mais à quoi ça sert ? ». D'où le souci constant pour l'enseignant de trouver des applications pratiques à la théorie qu'il expose. Le sujet que je propose ici regroupe deux thèmes principaux : le calcul matriciel et le calcul modulo 26. Le contexte est celui des messages secrets. C'est un sujet qui est très riche en applications mathématiques et qui intéresse beaucoup les élèves.
Le chiffre que nous allons étudier a été publié par *Lester S. Hill* en 1929 (cf. réf. [2]). C'est un chiffre polygraphique, c'est-à-dire qu'on ne (dé)chiffre pas les lettres les unes après les autres, mais par paquets. Nous étudierons la version bigraphique, c'est-à-dire que nous grouperons les lettres deux par deux, mais on peut imaginer des paquets plus grands.

Chiffrement de Hill

Les lettres sont d'abord remplacées par leur rang dans l'alphabet :

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

Tableau 1 (ou 0)

Les lettres P_k et P_{k+1} du texte clair seront chiffrées C_k et C_{k+1} avec la formule ci-dessous :

$$\begin{pmatrix} C_k \\ C_{k+1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} P_k \\ P_{k+1} \end{pmatrix} \pmod{26}$$

Ce qui signifie, pour fixer les idées, que les deux premières lettres du message clair (P_1 et P_2) seront chiffrées (C_1 et C_2) selon les deux équations suivantes :

$$C_1 = a \cdot P_1 + b \cdot P_2 \pmod{26}$$
$$C_2 = c \cdot P_1 + d \cdot P_2 \pmod{26}$$

Exemple de chiffrement

Alice prend comme clef de chiffrement la matrice $\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}$ pour chiffrer le message « je vous aime ».

D'après le tableau 1, P_1 = « j » = 10 et P_2 = « e » = 5. Les deux premières lettres du message seront donc cryptées ainsi :

$$C_1 = 9 \cdot 10 + 4 \cdot 5 \pmod{26} = 110 \pmod{26} = 6$$
$$C_2 = 5 \cdot 10 + 7 \cdot 5 \pmod{26} = 85 \pmod{26} = 7$$

En procédant de même avec les paires de lettres suivantes, elle obtiendra finalement :

Lettres	j	e	v	o	u	s	a	i	m	e
Rangs (P_k)	10	5	22	15	21	19	1	9	13	5
Rangs chiffrés (C_k)	6	7	24	7	5	4	19	16	7	22
Lettres chiffrées	F	G	X	G	E	D	S	P	G	V

Tableau 2

Le message chiffré sera donc « FGXGE DSPGV » (on a l'habitude d'écrire le message chiffré en lettres majuscules groupées par cinq).

Remarque : si le nombre de lettres du message clair avait été impair, Alice aurait simplement ajouté une lettre arbitraire à la fin du message original.

Remarques sur la matrice de chiffrement

On ne peut pas prendre n'importe quoi comme matrice de chiffrement ! Ses composantes doivent tout d'abord être des **nombres entiers positifs**. Il faut aussi qu'elle ait une matrice inverse dans Z_{26}. Cette **matrice inverse existe si *(ad–bc)*$^{-1}$ (mod 26) existe**, ce qui est le cas quand *(ad–bc)* et 26 sont premiers entre eux (le lecteur intéressé trouvera une preuve de ce théorème dans [4], p. 15). Il faut donc contrôler que ***(ad–bc)* est impair et n'est pas multiple de 13**.

Déchiffrement

Pour déchiffrer, le principe est le même que pour le chiffrement : on prend les lettres deux par deux, puis on les multiplie par la matrice de déchiffrement.

$$\begin{pmatrix} P_k \\ P_{k+1} \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} \begin{pmatrix} C_k \\ C_{k+1} \end{pmatrix} \pmod{26}$$

Ordinairement, l'inverse de la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est : $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Mais qu'en est-il dans Z_{26} ? Reprenons notre exemple.

Exemple de déchiffrement

Pour déchiffrer le message d'Alice, Bob doit calculer :

$$\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}^{-1} \pmod{26} = \frac{1}{9\cdot 7 - 4\cdot 5}\begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix} \pmod{26} = 43^{-1}\begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix} \pmod{26} = \ldots ?$$

Le problème est maintenant de calculer l'inverse de 43 modulo 26. Il existe des algorithmes efficaces pour déterminer l'inverse de k (mod n), par exemple *l'algorithme d'Euclide étendu*. Mais quand n = 26, la méthode *force brute* est sans doute la manière la plus simple :

Algorithme pour trouver k^{-1} modulo 26 *(force brute)*

1. Multiplier successivement k par les entiers m de l'ensemble {1, 3, 5, 7, 9, 11, 15, 17, 19, 21, 23, 25}
2. Stopper quand le produit $k{\cdot}m$ est égal à 1 (mod 26) ; k^{-1} modulo 26 = m.

L'utilisation de cet algorithme nous dit que 43^{-1} (mod 26) = 23.

Bob peut maintenant terminer de calculer sa matrice de déchiffrement :

$$\begin{pmatrix} 9 & 4 \\ 5 & 7 \end{pmatrix}^{-1} \pmod{26} = 23 \begin{pmatrix} 7 & -4 \\ -5 & 9 \end{pmatrix} \pmod{26} = \begin{pmatrix} 161 & -92 \\ -115 & 207 \end{pmatrix} \pmod{26} = \begin{pmatrix} 5 & 12 \\ 15 & 25 \end{pmatrix}$$

Bob va donc utiliser la matrice $\begin{pmatrix} 5 & 12 \\ 15 & 25 \end{pmatrix}$ pour déchiffrer le message « FGXGE DSPGV ». Après avoir remplacé les lettres par leur rang, il calculera :

$$P_1 = 5 \cdot 6 + 12 \cdot 7 \pmod{26} = 114 \pmod{26} = 10$$
$$P_2 = 15 \cdot 6 + 25 \cdot 7 \pmod{26} = 265 \pmod{26} = 5$$

En procédant de même avec les paires de lettres suivantes, il obtiendra finalement :

Lettres chiffrées	F	G	X	G	E	D	S	P	G	V
Rangs chiffrés (C_k)	6	7	24	7	5	4	19	16	7	22
Rangs (P_k)	10	5	22	15	21	19	1	9	13	5
Lettres	j	e	v	o	u	s	a	i	m	e

Tableau 3

Le message déchiffré sera donc bien « je vous aime ».

Pour aller plus loin

Je viens d'exposer les grandes lignes de cet algorithme de chiffrement. Si vous voulez aller plus loin avec le chiffre de Hill, par exemple essayer de décrypter un message donc on ne connaît pas la matrice de chiffrement (ni celle de déchiffrement bien sûr), ou passer à la version trigraphique, vous pouvez vous reporter au site que j'ai consacré à la cryptologie :

http://www.apprendre-en-ligne.net/crypto

Vous choisirez ensuite dans l'index alphabétique « Hill ».
Certaines pages interactives du site vous permettront de calculer rapidement la matrice de déchiffrement à partir de la matrice de chiffrement.
On peut aussi adapter ce chiffre en ajoutant à l'alphabet du *tableau 1* les lettres accentuées et les signes de ponctuation. On travaillera alors modulo le nombre de caractères de cet alphabet.
Enfin, on peut montrer aux élèves *l'algorithme d'Euclide étendu*, qui leur permettra de calculer l'inverse de *b* modulo *n* s'il existe. C'est une bonne introduction à la notion d'algorithme. Voyez la page :

http://www.apprendre-en-ligne.net/crypto/rabin/euclide.html

Références

[1] Eastaway R. et Wyndham J., **Pourquoi les bus arrivent-il toujours par trois ?** Flammarion, 2001, pp. 95-107
[2] Hill Lester S., « Cryptography in an Algebraic Alphabet », *American Mathematical Monthly*, **36**, 1929, pp. 306-312
[3] Lewand Robert Edward, **Cryptological Mathematics**, published by The Mathematical Association of America, 2000, pp. 124-140
[4] Stinson Douglas, **Cryptographie, Théorie et pratique**, Vuibert, 2001, pp. 12-16