J. L. Lassez

Messages parasites et codes synchronisants

Groupe d'étude d'algèbre, tome 1 (1975-1976), exp. n° 17, p. 1-4
<http://www.numdam.org/item?id=GEA_1975-1976__1__A17_0>

Groupe d'étude d'ALGÈBRE
(Marie-Paule MALLIAVIN)
1re année, 1975/76, n° 17, 4 p.

28 mai 1976

MESSAGES PARASITES ET CODES SYNCHRONISANTS

par J. L. LASSEZ

Résumé. - On démontre, ou retrouve, diverses propriétés des codes synchronisants réguliers en utilisant la notion de message parasite.

Introduction.

La notion très naturelle de sous-message parasite nous permet de classer les codes en trois catégories. Les codes à parasitisme borné, les codes à parasitisme étendu et les codes à parasitisme concentré. On établit facilement l'identité entre la famille des codes synchronisants à délai borné et la famille des codes à parasitisme borné. RESTIVO (cf. [5], [6]), résolvant un problème posé par McNANGHTON et PAPERT, a établi des liens entre les codes synchronisants et les codes très purs (que nous avons aussi considérer dans le cadre des mots circulaires). Etablissant le lemme qu'un code régulier à parasitisme étendu ne peut être très pur, nous retrouvons un résultat de RESTIVO, et améliorons la borne qu'il propose. Une autre conséquence simple de ce lemme est qu'on peut décider si un code régulier est synchronisant à délai borné. Certains des résultats que nous présentons ici, ainsi que leurs démonstrations, ont été publiés dans [1].

Notations et définitions.

Soient X un ensemble fini non vide, et X^* le monoïde libre engendré par X, c'est-à-dire l'ensemble des mots formés avec les éléments (lettres) de l'alphabet X. Pour une partie A de X^*, nous notons A^* le sous-monoïde qu'elle engendre. Un sous-ensemble C de X^* est un code si, et seulement si, tout mot de C^* se décompose de manière unique en éléments de C. Un code est donc le système minimal de générateurs d'un sous-monoïde libre. Soit W un mot de C^*, les séparations de W sont les positions où l'on peut décomposer W comme produit (unique) de mots de C. Un code C est appelé très pur [5] si, et seulement si, $\forall\, u , v \in X^*$, uv et $vu \in C^* \Longrightarrow u$ et $v \in C^*$. Soit C un code ; une paire (u , v) d'éléments de C^* est dite synchronisante si, et seulement si, $\forall\, f , f'$ X^*, $fuvf' \in C^*$ implique fu et $vf' \in C^*$. C est synchronisant à délai borné si, et seulement si, il existe un nombre p tel que toute paire d'éléments de C^p est synchronisante, le délai de C est alors le plus petit nombre ayant cette propriété. A cette définition classique [7], nous rajoutons la précision suivante qui nous sera utile par la suite. C a un délai de synchronisation égal à 0 si, et seulement si, $\forall\, u \in C$, $\forall\, f , f' \in X^*$, $fuf' \in C^* \Longrightarrow f$ et $f' \in C^*$. Soit C un code, p et q des entiers positifs, $a = a_1 a_2 \dots a_p$ $(a_i \in C)$, $b = b_1 b_2 \dots b_q$ $(b_i \in C)$: b est un sous-message parasite de a si, et seulement si, $\exists\, h , g \in X^*$, tel que $a = hbg$, et si

$a_1 a_2 \ldots a_i = hb_1 b_2 \ldots b_j$, alors $i = p$ et $j = q$ (en d'autres termes, les séparations de b ne coïncident pas avec les séparations de a sauf éventuellement la première ou la dernière). C est à parasitisme borné si, et seulement si, il existe un entier d tel qu'aucun mot de C^n , $n > d$, ne peut être sous-message parasite d'un mot de C^* . Le plus petit des entiers d ayant cette propriété est alors le degré de parasitisme borné. Un code C , qui n'est pas à parasitisme borné, est à parasitisme étendu si, et seulement si, les mots de C ne peuvent admettre de sous-messages parasites dans C^q avec q arbitairement grand. Un code C est à parasitisme concentré si, et seulement si, il existe des mots de C ayant des sous-messages parasites appartenant à C^q avec q arbitrairement grand.
Il est clair que tout code possède une, et une seule, de ces trois propriétés.

Résultats préliminaires.

Une conséquence directe de ces définitions, partiellemnt formulée dans [1] et [6], est la proposition suivante.

PROPOSITION 1. - Un code C est synchronisant à délai borné q si, et seulement s, il est à parasitisme borné de degré d . On a alors $q \leqslant d \leqslant 2q$.

En effet, si un code C est synchronisant à délai borné q , il est immédiat qu'il ne peut y avoir de sous-messages parasites ayant plus de 2q mots de C . Réciproquement, soit un code C à parasitisme borné de degré d . Considérons un sous-message formé de 2d mots de C , les d + 1 mots de gauche vont admettre une séparation commune avec le message principal ainsi que les d + 1 mots de droite, et comme C est un code, toutes les séparations situées entre ces deux séparations seront communes au message et au sous-message, en particulier celle du milieu du sous-message, ce qui exprime que C est synchronisant à délai borné inférieur ou égal à d .
Dans [1] nous avons montré que si, à partir d'un code C , on peut construire des mots circulaires admettant des doubles décompositions, alors on peut construire de tels mots utilisant un nombre arbitrairement grand de mots de C . Ce qui signifie que si un code est non circulaire, il est à parasitisme non borné. Exprimé différement, nous retrouvons le résultat de RESTIVO [6].

PROPOSITION 2. - Tout code synchronisant à délai borné est très pur.

La réciproque n'est pas vraie en général, les deux lemmes qui suivent vont nous permettre de trouver une réciproque dans le cas régulier [1].

LEMME 1. - Un code régulier C est à parasitisme concentré si, et seulement si, $\exists\, h , g \in X^*$, $\exists\, a \in CC^*$ tel que $ha^* g \subset C$.

LEMME 2. - Un code régulier à parasitisme étendu n'est pas très pur.

Ce dernier lemme est à la base des méthodes nous permettant d'obtenir les résultats qui suivent.

Résultats principaux.

Une conséquence directe des propositions 1 et 2 du lemme 2 est le résultat de RESTIVO [6], établi indépendament dans [1].

THÉORÈME 1. - *Soit* C *un code régulier ;* C *a un délai de synchronisation borné si, et seulement si,* C *est très pur sans parasitisme concentré.*

De ce théorème on déduit le théorème suivant.

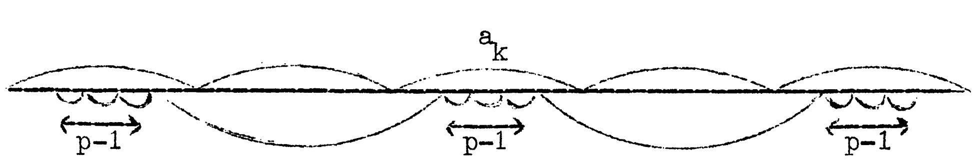
THÉORÈME 2. - *On peut décider si un code régulier* C *a un délai de synchronisation borné.*

En effet, on vérifie d'abord si le code C est très pur en considérant le monoïde syntactique associé à C^*. La notion de très pur pour C^* dans X^* s'étend à l'image de C^* dans le monoïde syntactique, et comme ce dernier est fini, la propriété est décidable. Si C est très pur, le lemme 1 nous permet de savoir si C est à parasitisme concentré, simple exercice sur des intersections finies de langages réguliers.

Soit maintenant un code régulier C à parasitisme non concentré, alors il existe un plus petit entier naturel p tel que $C \cap X^* C^p X^* = \emptyset$. RESTIVO a montré [6] que si un tel code a un degré de synchronisation borné q, alors $q < 2p(S + 1)$. Un tel code vérifie la propriété $F(p)$. En utilisant la notion de degré de parasitisme, nous pouvons établir diverses bornes, en particulier le cas suivant.

THÉORÈME 3. - *Soit* C *un code régulier vérifiant la propriété* $F(p)$. *Si* C *a un délai de synchronisation borné* q, *alors* $q \leq Sp - 1$.

Démonstration. - C ayant un degré de synchronisation borné q, nous savons qu'il est à parasitisme borné $d \geq q$. Soit donc $b = b_1 \dots b_d$ sous-message parasite de $a = a_1 \dots a_\ell = hbg$. Soit h_i le mot associé à b_i tel que $hb_1 \dots b_i$ admet h_i comme suffixe, et h_i est préfixe d'un certain a_k. A chaque h_i correspond ainsi un a_k. Ces a_k sont les seuls mots du message principal qui contient des séparations du message parasite. Dans la démonstration du lemme 2 (parue dans [1]), nous avons vu qu'il ne pouvait y avoir plus de S tels a_k, sinon C ne serait pas très pur, en contradiction avec les hypothèses, puisque un code synchronisant à délai borné est très pur. Chacun de ces a_k admet comme facteur au plus $p - 1$ mots du sous-message parasite. Nous avons donc au plus $S(p - 1)$ mots de b facteurs de mots de a. Pour compléter b, nous avons au plus $S - 1$ mots du sous-message reliant deux tels facteurs et donc $q \leq d \leq S(p - 1) + S - 1 = Sp - 1$.



D'autre bornes peuvent être trouvées où l'on majore p. Par exmemple on voit facilement que p est inférieur ou égal au nombre d'états de l'automate minimal acceptant C . Cependant si C est un code préfixe nous pouvons montrer [1] le résultat suivant.

THEOREME 4. - Soit C un code préfixe régulier. Si C est synchronisant à délai borné q alors $q < S$.

Ces bornes peuvent être généralisées ou améliorées, cependant on trouve des familles de codes ayant tous un délai de synchronisation égal à un, mais un automate minimal ayant un nombre arbitrairement grand d'états.

BIBLIOGRAPHIE

[1] LASSEZ (J. L.). - Circular codes and synchronization, Intern. J. Computer and Inf. Sc., t. 5, 1976, fasc. 2.

[2] McNAUGHTON (R.) et PAPERT (S.). - Counter-free automata. - Cambridge, M. I. T. Press, 1971 (Research Monograph, 65).

[3] NIVAT (M.). - Eléments de la théorie générale des codes, "Automata theory", p. 278-294. - New York, Academic Press, 1966.

[4] RESTIVO (A.). - Codes and aperiodic languages, "Fachtagung Automatentheorie formale Sprachen [1973. Bonn], Band 1, p. 175-181. - Berlin, Springer-Verlag, 1973 (lecture Notes in Computer Science, 2).

[5] RESTIVO (A.). - On a question of McNaughton and Papert, Inf. and Control, t. 25, 1974, p. 93-101.

[6] RESTIVO (A.). - A combinatorial property of codes having finite synchronization delay, Theoritical Computer Sc., t. 1, 1975, p. 95-101.

[7] SCHÜTZENBERGER (M. P.). - Codes à longueurs variable, Cours à l'école d'été de l'OTAN sur les méthodes combinatoires en théorie du codage, Royan 1965.

J. L. LASSEZ