Sul numero delle parole senza quadrati parzialmente abeliani per un alfabeto di tre lettere(*)

Maria Rosaria Formisano

Dipartimento di Matematica e Applicazioni R. Caccioppoli

Università di Napoli "Federico II"

Riassunto. *Si espone il contenuto di alcuni lavori dell' autore e in collaborazione con R. Cori, a proposito della nozione di "parole senza quadrati parzialmente abeliani" ; in particolare ci si sofferma sul risultato ottenuto sulla crescita polinomiale di tali parole quando l'alfabeto ha tre lettere.*

1. Parole infinite senza quadrati parzialmente abeliani.

Sia A^* il monoide libero generato dall'alfabeto A , Λ la parola vuota di A^* . Una parola w di A^* è un quadrato se esiste una parola $u \neq \Lambda$ tale che $w = u.u = u^2$. La parola w è " senza quadrati" se non esiste una fattorizzazione $w = w_1u^2w_2$, con $u \neq \Lambda$. Si estende tale definizione alle parole infinite su A, cioè alle applicazioni $\mathbf{w} : N \rightarrow A$. Una parola infinita $\mathbf{w}$ è "senza quadrati" se ogni suo prefisso finito $w_n = \mathbf{w}(1)\mathbf{w}(2)...\mathbf{w}(n)$ lo è .

E' noto che la minima cardinalità dell' alfabeto A per costruire una parola infinita senza quadrati è 3. Parole infinite senza quadrati su tre lettere sono state costruite con vari metodi. Per esempio la parola **m** di Thue-Morse ottenuta mediante iterazione del morfismo $\varphi : \{a,b,c\}^* \rightarrow \{a,b,c\}^*$, definito da $\varphi(a) = abc$, $\varphi(b) = ac$, $\varphi(c) = b$ è una parola infinita senza quadrati ([8], [14]), mentre la parola di Arson è un altro esempio, che però non è generabile mediante l'iterazione di un morfismo, come ha mostrato J.Berstel ([1] , [4]).

(*) Testo di un intervento su invito al " Séminaire Lotharingien de Combinatoire", settembre 1990, Salzburg, Austria.

Nell'insieme delle regolarità "evitabili" (1) è nota la nozione di quadrato abeliano, cioè di una parola che si fattorizza come u.p(u), essendo p(u) una qualsiasi permutazione di u. Una parola senza quadrati abeliani è anche senza quadrati, mentre ovviamente il viceversa non è vero. Per esempio cabcacba è senza quadrati ma ha il quadrato abeliano abc.acb . La cardinalità minima di un alfabeto A perchè esista una parola infinita su A senza quadrati abeliani è 5, come ha dimostrato P.Pleasants, costruendone una mediante l'iterazione di un morfismo "abelian square-free", che ha, cioè, la proprietà di conservare l'assenza di quadrati abeliani ([16]).

Una congettura di P. Erdos a tal proposito asserisce che è possibile costruire una parola infinita senza quadrati abeliani su un alfabeto di 4 lettere ([10]), mentre P. Pleasants in [16] propone di utilizzare, anche in questo caso, l' iterazione di un morfismo. Risultati connessi a questa congettura di P. Erdos sono contenuti in [6], [7], [11], dove viene introdotta la nozione di quadrato parzialmente abeliano, intermedia tra la nozione di quadrato e quella di quadrato abeliano, e che si mostra essere un' altra regolarità evitabile, quando l' alfabeto ha cardinalità $|A| \geq 3$.

La nozione di quadrato parzialmente abeliano si pone naturalmente nell'ambito della teoria dei monoidi liberi parzialmente commutativi, strutture algebriche particolarmente interessanti dal punto di vista combinatorio, estensivamente studiate da Cartier e Foata [5] . In seguito ai lavori di molti studiosi di modelli matematici della computazione parallela, è cresciuta la popolarità di tale teoria, anche tra gli informatici teorici : [2], [8], [9], [13], [15] e molti altri. Più diffuso è il nome di traccia, per il generico elemento di un monoide libero parzialmente commutativo, e suoi sottoinsiemi sono i cosiddetti linguaggi di tracce .

(1) Una parola $u \in A^*$ si dice una regolarità "evitabile", se il linguaggio $A^* \setminus A^* u A^*$ è infinito ([14]).

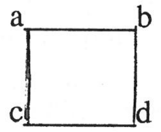
Se A è un alfabeto e θ una relazione binaria simmetrica e irriflessiva (l' insieme delle commutazioni "parziali"), il monoide libero parzialmente commutativo M(A, θ) è il monoide quoziente $A^*/_{\sim\theta}$, essendo $\sim\theta$ la congruenza minima in A* generata da θ . Se f e g sono due parole di A* che individuano la stessa traccia in M (A, θ), $[f]_{\sim\theta} = [g]_{\sim\theta}$, la parola fg di A* si dirà un <u>quadrato parzialmente abeliano (θ)</u> o <u>θ– quadrato</u>, mentre il prodotto $[f]_{\sim\theta}[g]_{\sim\theta}$ è un quadrato in M(A, θ). Sia Λ la parola vuota di A* ; allora è possibile dare le seguenti definizioni ([6], [11]) :

DEFINIZIONE 1. Un θ– quadrato in A* è una parola fg tale che $f \sim\theta\ g$ e $f \neq \Lambda \neq g$.

DEFINIZIONE 2. Una parola f di A* è senza θ – quadrati se, per ogni fattorizzazione $f = gf'f''h$, con $f' \sim\theta\ f''$, risulta $f' = \Lambda = f''$.

Ad esempio, sia A = {a, b, c}, θ = {(a,c), (c,a)}. La parola <u>bca</u>.<u>bac</u> è un θ–quadrato, ma non è un quadrato, mentre è un quadrato abeliano. E' ovvio che una parola quadrato è un θ–quadrato, per ogni commutazione parziale θ, ed è anche un quadrato abeliano. La sequenza di P.Pleasants è una parola infinita, su di un alfabeto di 5 lettere, senza θ–quadrati, per ogni θ, e quindi senza quadrati . E' dunque naturale chiedersi quale sia la cardinalità minima di un alfabeto A perchè esista una parola infinita su A senza θ–quadrati , con $\theta \neq \emptyset$.

A questa domanda si è risposto dimostrando che la parola **m** di Thue-Morse fornisce un esempio di parola infinita sull' alfabeto A ={a, b, c} senza θ–quadrati, dove θ ={(a,c),(c,a)},per cui la risposta è 3 ([6], [11]). Se θ' è una qualsiasi commutazione parziale contenente propriamente θ, una parola senza θ'- quadrati ha al più lunghezza 15 [6], sempre per $|A| = 3$. Nel caso $|A| = 4$ si è dimostrato in [6] che esistono parole infinite su A senza θ–quadrati, dove θ è la commutazione parziale rappresentata dal seguente grafo :



In esso i vertici sono gli elementi di $A = \{a, b, c, d\}$, ed i lati sono le coppie di lettere che commutano, cioè le relazioni θ nella presentazione $M(A, \theta)$; questo, finora è il migliore risultato noto sulla citata congettura di P. Erdos.

2. Numero delle parole senza θ – quadrati per $|A| = 3$

Sia $L_2(\theta)$ il linguaggio delle parole in $\{a,b,c\}^*$ senza θ– quadrati. Nel caso in cui $\theta = \emptyset$ esso coincide con il linguaggio delle parole senza quadrati ordinari, L_2, che è un insieme infinito, dal momento che esistono parole infinite su tre lettere senza quadrati. Nel caso $\theta = \{(a,c),(c,a)\}$, $L_2(\theta)$ è ancora un insieme infinito, ma la sua funzione di crescita è polinomiale ([7]), mentre quella di L_2 è esponenziale, da un noto risultato di F. Branderburg [3], il quale dimostra che il numero delle parole di $\{a,b,c\}^*$ di lunghezza n è una funzione esponenziale di n.

Diamo ora una caratterizzazione di un θ– quadrato, utilizzando la nozione di "proiezione" di un elemento f di A^* sul sottoalfabeto B di A : $\pi_B(f)$ è l' elemento di B^* che si ottiene da f cancellando ogni lettera x che non appartiene a B. Se f è un elemento di $(A \setminus B)^*$, risulta $\pi_B(f) = \Lambda$. Le seguenti proposizioni sono utili conseguenze della 1.1 in [8].

PROPOSIZIONE 1. La parola f.g è un θ– quadrato se, e soltanto se, le seguenti condizioni (i) e (ii) sono soddisfatte :

(i) $|f|_x = |g|_x \quad \forall\, x \in A.$

(ii) $\pi_{x,y}(f) = \pi_{x,y}(g) \;\; \forall\, (x,y) \in \theta$

PROPOSIZIONE 2. Le parole f e $g \in \{a,b,c\}^*$ sono tali che $f \sim_\theta g$ se, e soltanto se, esistono fattorizzazioni di f e g :

$$f = f_1\, b\, f_2\, b \ldots f_p\, b\, f_{p+1}, \qquad g = g_1\, b\, g_2\, b \ldots g_p\, b\, g_{p+1}$$

soddisfacenti

$$\forall i = 1 \ldots p+1 \qquad f_i, g_i \in \{a,c\}^* \quad e \quad |f_i|_a = |g_i|_a, |f_i|_c = |g_i|_c .$$

Con l' ausilio delle precedenti proposizioni si prova in [11] che una parola senza θ–quadrati non può contenere come fattori "interni" i fattori bacb né bcab, la cui presenza darebbe immediatamente luogo ad un quadrato ordinario oppure ad un θ–quadrato. Analizzando ancora la struttura di una parola w su tre lettere senza θ– quadrati è possibile facilmente riconoscere altri fattori "proibiti", come le parole aba e cbc , la cui assenza ci permetterà di collegare tali w alle parole su due lettere senza "overlapping factors", e di utilizzare così un risultato di A. Restivo e di S. Salemi del 1983, nonché il più recente di Y. Kobayashi ([7], [17], [12]).

Sia M_2 il sottoinsieme di A^* costituito dalle parole senza quadrati che non hanno aba né cbc come fattori, cioè $M_2 = L_2 \setminus A^* \{aba, cbc\} A^*$. Siano α_n e β_n il numero delle parole di lunghezza n rispettivamente di $L_2(\theta)$ e di M_2. Le relazioni tra queste due cardinalità contenute nella proposizione seguente sono il punto cruciale per la dimostrazione del risultato principale .

PROPOSIZIONE 3. I numeri α_n e β_n verificano :

(i) $\beta_n \geq 4\,\alpha_{n-2} \qquad \forall\, n \geq 3$

(ii) $\alpha_n \geq 9\beta_{n-4} \qquad \forall\, n \geq 7$

La dimostrazione si trova in [7], ma qui ne diamo lo schema, che si articola sui Lemmi 1 e 2 che seguono.

LEMMA 1. Se f è una parola di M_2, di lunghezza $|f| \geq 2$, e se g è tale che $f = xgy$, con x, $y \in A$, allora $g \in L_2(\theta)$.

LEMMA 2. Se f è una parola di $L_2(\theta)$, di lunghezza $|f| \geq 7$, e se g è tale che $f = ugv$, con $u,v \in A^2$, allora $g \in M_2$.

Entrambi questi due lemmi esprimono sostanzialmente il fatto, conseguenza della Proposizione 2, che parole senza quadrati su $\{a,b,c\}$ e che non contengono i fattori aba né cbc, sono niente altro, a meno di "bordi" piccoli, che parole senza θ– quadrati ; viceversa, parole senza θ– quadrati abbastanza lunghe non contengono, come fattori "interni", le parole aba né cbc.

DIMOSTRAZIONE della Proposizione 3 .

(i) : sia $xgy \in A^2A^*$, con $x,y \in A$. Se $xgy \in M_2$, dal Lemma 1 è $g \in L_2(\theta)$ e ci sono 9 diverse parole xgy, al variare dei bordi x, y in A, che hanno lo stesso fattore "centrale" g in A^* ; se però $xgy \in M_2$, x non può essere la prima lettera di g, ed anche y non può essere l' ultima . Allora le parole distinte $xgy \in M_2 \cap A^n$ tali che g è in $L_2(\theta) \cap A^{n-2}$ sono al più 4.

(ii) : sia $ugv \in A^4A^*$, con $u,v \in A^2$. Se $g \in M_2$ ed è di lunghezza $|g| \geq 3$, le possibili coppie (u,v) tali che $ugv \in L_2(\theta)$ sono al più 9 : ciò si dimostra osservando che $L_2(\theta) \subset L_2$, e che per ogni parola f senza quadrati, di lunghezza $|f| > 2$ esistono al più tre parole $u \in A_2$ tali che uf è ancora in L_2. Infatti, se $f = xyg$, con x, $y \in A$, e se z è una lettera di A distinta sia da x che da y, allora le relazioni $u', u'' \in A$, $u'u''f \in L_2$ implicano $u'u'' \in \{xz, yz, zy\}$. Dunque ogni parola g di M_2 di lunghezza $n-4$ si può prolungare in al più 9 parole distinte $ugv \in L_2$ di lunghezza n, e si perviene cosi' alla diseguaglianza contenuta nella (ii).

N.3 Dai θ – quadrati ai fattori overlapping.

In questo numero si collegheranno le parole senza θ– quadrati su tre lettere alle parole senza "overlapping" factors su due lettere, e lo faremo utilizzando il morfismo $\delta : \{a,b,c\}^* \rightarrow \{0,1\}^*$, cosi' definito: $\delta(a) = 011$, $\delta(b) = 01$, $\delta(c) = 0$. Allora si ha la seguente Proposizione:

PROPOSIZIONE 4. Se f è una parola di M_2, $\delta(f)$ è una parola senza "overlapping" factors. Inversamente, se $w \in \{0,1\}^*$, $|w| > 6$, è una parola che inizia con 0, senza "overlapping" factors, allora esiste un unico elemento g di M_2 e lettere $x, y \in \{a,b,c\}$ tali che $\delta(xgy) = w$.

La dimostrazione della Proposizione 4 è contenuta in [7], ed è ottenuta con una "reductio ad absurdum" utilizzando la definizione di δ e i risultati sulla struttura delle parole di M_2.

Finalmente,come corollario della 4, la seguente Proposizione 5 stabilisce due utili disuguaglianze che permettono, come già detto, di applicare il seguente risultato di Y. Kobayashi *" Esistono costanti c_1 e c_2, tali che i numeri γ_n di parole senza "overlaps" su due lettere verificano:*

$$c_1 n^{1.155} \leq \gamma_n \leq c_2 n^{1.587} \text{"} \quad ([\ 12\ ],\ [\ 17\ ]).$$

PROPOSIZIONE 5. I numeri β_n di parole di M_2 di lunghezza n, e γ_n di parole senza "overlaps" di lunghezza n, verificano :

(i) $\forall n \ \exists k \in \{2n-1, 2n, 2n+1\}$, tale che $\beta_n \leq 3/2\, \gamma_k$

(ii) $\gamma_{2n} \leq 2\, \beta_{n-2}$

Conseguenze delle precedenti proposizioni sono essenzialmente due risultati che sembrano non avere alcuna relazione tra di loro, ma che invece scaturiscono entrambi dall' analisi dell' albero delle parole su tre lettere senza quadrati e del suo sottoalbero delle parole senza θ– quadrati, i quali hanno due funzioni di crescita diverse, esponenziale e polinomiale, rispettivamente.Essi sono:

$\underline{1}$) Un risultato enumerativo sul numero delle parole su tre lettere senza θ– quadrati.

$\underline{2}$) Un teorema di non esistenza per morfismi non elementari liberi da θ– quadrati.

TEOREMA 1. Il numero α_n delle parole senza θ– quadrati di lunghezza n verifica :

$$c_1 n^{1.155} \leq \alpha_n \leq c_2 n^{1.587} .$$

Per la dimostrazione basta applicare il risultato di Kobayashi per γ_n, ed inoltre le Prop. 3 e 5.

TEOREMA 2. Ogni morfismo senza θ– quadrati di $\{a, b, c\}^*$ in sé è elementare.

Riportiamo qui la definizione di morfismo senza θ– quadrati, che è un morfismo che, applicato ad una parola senza θ– quadrati, dà per risultato ancora una parola senza θ– quadrati. Esempi sono l'identità e il morfismo φ così definito : $\varphi(b) = b$, $\varphi(a) = c$, $\varphi(c) = a$, che chiameremo <u>elementari</u>.

DIMOSTRAZIONE del Teorema 2 .

Sia φ un morfismo senza θ– quadrati e sia $f \in L_2 \cap A^5$. Ma le parole di lunghezza 5 senza quadrati sono anche senza θ– quadrati, quindi $f \in L_2(\theta) \cap A^5$, e per l' ipotesi su φ, $\varphi(f) \in L_2(\theta)$ il che assicura, per una caratterizzazione dovuta a Crochemore, che φ è un morfismo senza quadrati

OSSERVAZIONE 1. Sia φ tale che due lettere in $\{a, b, c\}$ hanno immagini di lunghezza maggiore di 1. Allora, se $f \in L_2(\theta)$, né aba né cbc possono essere fattori di $\varphi(f)$. Infatti, sia f una parola di lunghezza 2 tale che $\varphi(f)$ ha aba oppure cbc come fattore: $f = xy$, con $x \neq y$, $z \in \{a,b,c\}$, $z \neq x$, $z \neq y$. La parola $g = yzxyzy$ è senza quadrati, per ogni scelta siffatta di x, y, $z \in \{a, b, c\}$. Risulta : $u = \varphi(g) = \varphi(y)\varphi(z)\varphi(x)\varphi(y)\varphi(z)\varphi(y) = u'aba\,u''$, con $|u'| \geq 3$, $|u''| \geq 3$, che, per il Lemma 2, contiene un θ– quadrato. Contraddizione.

OSSERVAZIONE 2. Se due lettere x, y hanno immagini di lunghezza 1, x' e y', allora φ non è un morfismo senza quadrati, oppure φ è elementare. Siano infatti $\varphi(x) = x'$, $\varphi(y) = y'$; allora, essendo φ non elementare, $|\varphi(z)| > 1$, il che implica che $\varphi(z) = z'u'z'$, con $u' \in (x',y'\}^*$. Consideriamo allora la parola senza quadrati uz, $u \in \{x,y\}$, tale che $\varphi(u) = u'$. Risulta $\varphi(uz) =$

u'z'u'z', che è un quadrato. Dunque, un morfismo senza $\theta-$ quadrati non elementare porta L_2 in M_2, e ciò è in contraddizione con il risultato di Brandenburg ([3]) e con il Teorema 1 ([7]).

BIBLIOGRAFIA

[1] S. E. ARSON, Dém.de l' existence de suites asym. infinies Mat.Sib. 44 (1937)19 - 38.

[2] A. BERTONI, M. BRAMBILLA, N. SABADINI, G. MAURI, An application of the theory of free partially commutative monoids: asymptotic densities of trace languages, Lecture Notes in Comp. Sci. 118 (1981) 205 - 215.

[3] F. BRANDENBURG, Uniformely growing k-th power-free homomorphisms, Theoretical Comp. Sci. 23 (1983) 145 - 149.

[4] J. BERSTEL, Mots sans carré et morphisms itérés, Discr. Math. 29 (1980)

[5] P. CARTIER, D. FOATA, Propriétés Combinatoires de Commutations et de Réarrangements, Lecture Notes in Mathematics 5 (1969), Springer Verlag.

[6] R. CORI, M.R. FORMISANO, Partially abelian square-free words, RAIRO Informathique Théorique, to appear.

[7] R. CORI, M.R. FORMISANO, On the number of partially abelian square-free words on a three letter alphabet, Theoretical Comp. Sci., to appear.

[8] R. CORI, D. PERRIN, Automates et Commutations Partielles, RAIRO Inf.Th.(1985) 232.

[9] C. DUBOC, Commutations dans les monoides libres : un cadre théorique pour l' étude du parallelisme, Doctoral Thesis, University of Rouen, (March 1986).

[10] P. ERDOS, Some unsolved problems, Magiar Tud. Akad. Mat. 6, (1961) 221 - 254.

[11] M.R. FORMISANO, Mots infinis sans carrés partiellement abéliens sur trois lettres. Rapport scientifique n. 84.19 du CNRS, Equipe du laboratoire associé 226, Bordeaux1.

[12] Y. KOBAYASHI, Enumeration of irreducible binary words. To appear.

[13] R. KONIG, Graphs and free partially commutative monoids. To appear in T.C.S.

[14] M. LOTHAIRE, Combinatorics on Words, Addison Wesley (1983).

[15] A. MAZURKIEWICZ, Concurrent Program Schemes and their Interpretations, DIAMI-PB 76, Aarhus University, Aarhus (1977).

[16] P. A. B. PLEASANTS, Non repetitive sequences, Proc.Cambridge Phil. Soc. (1970).

[17] A. RESTIVO, A. SALEMI, On weakely square-free words, Bull (1983) 49 - 56.

[18] A. THUE, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norshe vid. Selsk. Skr. I. Mat - Nat. Kl., Christiana. 1 (1912) 1 - 67 .

[19] Per una bibliografia abbastanza completa sulla teoria delle tracce, si rimanda al testo "Combinatorics on traces " di Volker Diekert, 454 di Lecture Notes in Computer Science.