Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales
Departamento de Computación

Un algoritmo para collares perfectos lexicográficamente máximos

Tesis de Licenciatura en Ciencias de la Computación

Tomás Tropea
tomastropeaa@gmail.com

Directora: Verónica Becher

Buenos Aires, 13 de septiembre de 2023

UN ALGORITMO PARA COLLARES PERFECTOS LEXICOGRÁFICAMENTE MÁXIMOS

Resumen. Un collar es una secuencia circular de símbolos. Los collares perfectos son variantes de las secuencias de De Bruijn: un collar es (n, k)-perfecto si todas las palabras de longitud n aparecen en el collar exactamente k veces, en posiciones distintas modulo k, para cualquier convención de la posición inicial. En esta tesis presentamos un algoritmo para generar los collares (n, k)-perfectos lexicográficamente máximos, cuando k divide a n. Nuestro algoritmo es una adaptación del algoritmo clásico de Fredericksen y Majorana basado en la concatenación de palabras de Lyndon. Como subproducto obtuvimos una demostración de la correctitud del algoritmo de Fredericksen y Majorana mucho más clara que la original.

Palabras clave: collares perfectos, secuencias de De Bruijn, palabras de Lyndon.

Índice general

1. INTRODUCCIÓN

Consideremos un alfabeto finito, y sea s la cantidad de símbolos de este alfabeto. Los collares —o palabras circulares— de De Bruijn de orden n son aquellos en los que todas las palabras de longitud n aparecen exactamente una vez. Los collares de de Bruijn de orden n tienen longitud s^n. Nicolaas Govert de Bruijn [1] dio esta definición al mismo tiempo que los caracterizó como circuitos eulerianos en los llamados grafos de De Bruijn, y contó cuántos hay. Una hermosa presentación de las secuencias de De Bruijn es el artículo de Jan Berstel [6].

Álvarez, Becher, Ferrari y Yuhjtman [7] presentaron una generalización de los collares de de Bruijn, que se llaman *collares perfectos*. Un collar es *perfecto* de orden (n, k) si todas las palabras de longitud n ocurren exactamente k veces, y todas ellas en distintas posiciones módulo k, para cualquier convención de la posición inicial. Los collares (n, k)-perfectos tienen longitud ks^n. En [7] los collares perfectos están caracterizados como circuitos eulerianos en los llamados grafos astutos, dándose una fórmula para la cantidad de collares (n, k)-perfectos.

Entre todos los collares de Bruijn de orden n hay uno que es el lexicográficamente máximo. Fredericksen y Maiorana [4] dieron un algoritmo para generarlo. Este algoritmo se basa en la concatenación de las llamadas palabras de Lyndon. Una palabra de Lyndon es una cadena no vacía que es estrictamente mayor en el orden lexicográfico que todas sus rotaciones. Existen varias definiciones equivalentes. Una palabra de Lyndon es el único elemento máximo en el orden lexicográfico del multiconjunto de todas sus rotaciones. Ser la rotación singularmente más grande implica que una palabra de Lyndon difiere de cualquiera de sus rotaciones no triviales y, por lo tanto, es aperiódica.

En esta tesis nos proponemos extender el algoritmo de Fredericksen y Maiorana para producir collares (n, k)-perfectos, cuando k divide a n. Como subproducto obtenemos una demostración de correctitud del algoritmo original de Fredericksen y Maiorana más simple de comprender.

El collar de Bruijn de orden n lexicográficamente máximo fue estudiado originalmente por Ford [2]. Más tarde Fredericksen [5, Algorithm 2] dio un algoritmo goloso que lo construye eligiendo un símbolo en cada paso, hasta completar el collar. También Fredericksen caracterizó esta secuencia por medio de los llamados *shift register* [5]. En [8] Ezequiel Zimenspitz extendió el algoritmo goloso de Fredricksen adaptándolo para construir collares (n, n)-perfectos lexicográficamente máximos. Sin embargo, dicho algoritmo no resultó ser el más eficiente ya que los collares (n, n)-perfectos son la concatenación de todas las palabras de longitud n en orden lexicográfico. Queda pendiente saber si hay un algoritmo para collares (n, k)-perfectos lexicográficamente máximos más eficiente que el que presentamos en esta tesis.

2. COLLARES

2.1. Collares donde k divide a n

Sea Σ un alfabeto de s símbolos, $\Sigma \coloneqq \{a_1, a_2, \dots a_s\}$. Sin pérdida de generalidad representamos estos símbolos mediante números $0, 1, \dots, s-1$. Asumimos el orden natural usual $0 < 1 < \dots < s-1$. Denotamos el conjunto de todas las palabras con Σ^*.

Para denotar palabras usaremos letras mayúsculas, A, U, V o bien directamente una secuencia de símbolos $a_1 \dots a_n$. Si A es una palabra, $|A|$ es su longitud, y escribimos A_i para el prefijo de los primeros i símbolos de A. Fijado un alfabeto Σ, una palabra es una secuencia finita de símbolos $a_1 \cdots a_n$ del alfabeto Σ. Usamos el símbolo $<$ para indicar la relación de ser menor en orden lexicográfico, y definimos $a_1 \dots a_n < b_1 \dots b_m$ si existe k tal que $a_i = b_i$, para $i = 0, \dots k-1$ y $a_k < b_k$.

Dadas dos palabras $W \coloneqq w_1 \cdots w_n$ y $X \coloneqq x_1 \cdots x_n$, la concatenación de estas es $WX \coloneqq w_1 \cdots w_n x_1 \cdots x_n$. La potencia W^n es la concatenación $W^{n-1}W$, donde W^0 es la palabra vacía.

Trabajaremos con pares formados por una palabra y un entero no negativo en $\mathbb{Z}/k\mathbb{Z}$, para un valor entero positivo k prefijado. Para referirnos a un par usaremos $\langle A, m\rangle$ y lo nombramos con una letras caligráficas $\mathcal{A}, \mathcal{B}, \mathcal{C}$, etc. Definimos la longitud de un par $\mathcal{A} = \langle A, m\rangle$ como la longitud de la palabra A, es decir $|\mathcal{A}| = |A|$.

Ejemplo. *Si consideramos el tamaño de alfabeto $s = 3$ y fijamos el módulo $k = 5$, estos son pares validos:*

- $\langle W, m\rangle = \langle 201011, 4\rangle$
- $\langle W, m\rangle = \langle 112201, 1\rangle$

Y estos otros no:

- $\langle W, m\rangle = \langle 201011, 7\rangle$*: El $m = 7$ no está en el rango del modulo $k = 5$.*
- $\langle W, m\rangle = \langle 115201, 1\rangle$*: La palabra tiene el símbolo 5, que se encuentra fuera del rango del alfabeto $s = 3$.*

Definición (Orden lexicográfico entre pares)**.** *Dados pares (U, m_U) y (V, m_V), el orden lexicográfico se define de la siguiente manera:*

- *Si $m_U < m_V$ entonces $\langle U, m_U\rangle > (V, m_V)$.*
- *Si $m_U > m_V$ entonces $(V, m_V) > (U, m_U)$.*
- *Si $m_U = m_V$ el orden se define a nivel palabras entre U y V.*

Ejemplo. *Los siguientes son ejemplos de los distintos casos que pueden darse con este orden lexicográfico:*

- $\langle 435, 3\rangle > \langle 123, 4\rangle$*: El caso donde los módulos difieren.*

- $\langle 125, 0\rangle > \langle 125, 0\rangle$*: El caso donde ambos módulos son iguales y se define a nivel palabras.*

Definición (Concatenación de pares)**.** *Definimos concatenación de pares* $\langle U, m\rangle$ *y* $\langle V, m\rangle$ *como* $\langle UV, m\rangle$*, siempre que* $|U|$ *y* $|V|$ *sean múltiplos de* k*.*

Ejemplo. *Dado* $k = 2$*, para los pares* $\langle 4357, 1\rangle$ *y* $\langle 24, 1\rangle$*, su concatenación seria* $\langle 435724, 1\rangle$*. Notar sin embargo que no podemos concatenar* $\langle 4357, 1\rangle$ *y* $\langle 24, 0\rangle$*, ya que difieren en su módulo.*

Definición (Rotación de un par)**.** *Sea el par* $\langle a_1 \cdots a_n, m\rangle \in \Sigma^* \times \mathbb{Z}/k\mathbb{Z}$*. Definamos las rotaciones a izquierda y derecha en una posición como* $\langle a_2 \cdots a_n a_1, m \oplus 1\rangle$ *y* $\langle a_n a_1 \cdots a_{n-1}, m \ominus 1\rangle$ *respectivamente. Las operaciones* $\oplus$ *y* $\ominus$ *son la suma y resta módulo* k *respectivamente.*

Ejemplo. *Si tenemos* $s = 3$*,* $k = 5$ *y el par* $\langle 13212, 4\rangle$*, la rotación a derecha es* $\langle 21321, 0\rangle$*, y a izquierda* $\langle 32121, 3\rangle$*.*

La clausura por rotación de un par $\langle a_1 \cdots a_n, m\rangle$, es el conjunto de pares.

$$\{\langle a_{i+1} \cdots a_n a_1 \cdots a_i, m \oplus i\rangle : 0 \leq i < n\}$$

Ejemplo. *Si* $s = 2$*,* $n = 3$ *y* $k = 3$ *las rotaciones de* $\langle 000, 0\rangle$ *son* $\langle 000, 1\rangle$ *y* $\langle 000, 2\rangle$*,* 3 *en total. Si* $s = 2$*,* $n = 4$ *y* $k = 2$*, el par* $\langle 0101, 0\rangle$ *tiene una sola rotación que es* $\langle 1010, 1\rangle$*, por lo que son en total* 2 *rotaciones. Si* $s = 2$*,* $n = 2$ *y* $k = 3$*, el par* $\langle 0101, 0\rangle$ *tiene estas otras rotaciones,* $\langle 1010, 1\rangle$*,* $\langle 0101, 2\rangle$*,* $\langle 1010, 0\rangle$*,* $\langle 0101, 1\rangle$ *y* $\langle 1010, 2\rangle$*, por lo que son en total* 6 *rotaciones distintas.*

La función de rotación induce una relación entre pares: un par está relacionado con otro si sucesivas aplicaciones de rotación sobre el primer par da como resultado el segundo par. Claramente esta relación es reflexiva y transitiva. En pares $\Sigma^n \times \mathbb{Z}/k\mathbb{Z}$, cuando k divide a n la función de rotación tiene inversa, (dada por rotaciones sucesivas), entonces la relación es también simétrica, y por lo tanto es una relación de equivalencia. Consideraremos las clases de equivalencia.

La siguiente es la nomenclatura de Fredericksen y Maiorana en [4].

Definición (Collar de un par)**.** *Dado* n *y* k *enteros positivos con* $k \mid n$ *ó* $n \mid k$ *, y un* m *en* $\mathbb{Z}/k\mathbb{Z}$*, llamamos collar al par lexicográficamente máximo de la clase de equivalencia dada por las rotaciones de un par* $\langle A, m\rangle$*, con* $|A| = n$*.*

Si bien la definición impone que el modulo sea 0, podríamos haber definido a los collares fijando cualquier otro valor entre 0 y $k - 1$, y el desarrollo hubiera sido idéntico. Es interesante notar que existen collares que tienen varias rotaciones iguales, y que todas ellas pueden ser lexicográficamente mayores que las demás. También hay otros collares donde todas las rotaciones son distintas.

Ejemplo. *Si* $s = 2$*,*$n = 3$ $k = 3$*, la clase de equivalencia de* $\langle 010, 1\rangle$ *tiene los siguientes pares:*

- $\langle 010, 1\rangle$
- $\langle 100, 2\rangle$
- $\langle 001, 0\rangle$

donde el representante del collar es el par $\langle 100, 0\rangle$*. Notemos que en este caso todas las rotaciones de la palabra son distintas, si bien se repiten para distintos módulos.*

Por otro lado, si tenemos $s = 2$*,* $n = 4$*, y* $k = 2$*, la clase de equivalencia de* $\langle 1010, 1\rangle$ *tiene los siguientes pares:*

- $\langle 1010, 1\rangle$
- $\langle 0101, 0\rangle$

donde el representante del collar es el par $\langle 0101, 0\rangle$*. En este caso no todas las rotaciones de la palabra son distintas ya que al rotar 2 veces volvemos a la misma.*

Definición (Potencia de par)**.** *Dado un par* $\langle A, m\rangle$*, con* $|A|$ *múltiplo de* k*, y un entero positivo* n*, definimos* $\langle A, m\rangle^n$ *como la operación* $\langle A^n, m\rangle$*.*

Ejemplo. *Para el par* $\langle 456, 2\rangle$*, la potencia* $\langle 456, 2\rangle^3 = \langle 456456456, 2\rangle$*.*

Observación 1. *Sean* n *y* k *enteros positivos, tal que* k *divide a* n *y* $\mathcal{A} = \langle A, 0\rangle$*, con* $|A| = n$*.* $\mathcal{A}$ *es collar si y solo si* $\mathcal{B} = \mathcal{A}^p$ *es collar, donde* $p \geq 1$*.*

Demostración. $\implies$ Si $\mathcal{B}$ no fuera collar, entonces por definición existe otra rotación más grande, llamémosla $\mathcal{R}$, y tiene la forma $\mathcal{R} = \langle (a_{j+1} \cdots a_n a_1 \cdots a_j)^p, 0\rangle$, con j múltiplo de k. Dado que $\langle a_{j+1} \cdots a_n a_1 \cdots a_j, 0\rangle^p > \langle a_1 \cdots a_n, 0\rangle^p$, necesariamente ocurre que

$$\langle a_{j+1} \cdots a_n a_1 \cdots a_j, 0\rangle > \langle a_1 \cdots a_n, 0\rangle.$$

Pero esto implica que hay otra rotación de $\mathcal{A}$ que es mayor, y $\mathcal{A}$ era collar. Esto es absurdo y vino de suponer que $\mathcal{B} = \mathcal{A}^p$ no era collar. Concluimos que $\mathcal{B}$ sí lo es.

$\impliedby$ Si $\mathcal{A}$ no fuera collar, por el mismo argumento que en el otro caso existe una rotación $\mathcal{R} = \langle R, 0\rangle$ mayor. Luego puedo construirme $\mathcal{R}^p$, que a su vez va a ser mayor que $\mathcal{B} = \mathcal{A}^p$, porque ambos tienen modulo 0, y si tomo el prefijo de tamaño n de ambos pares, ocurre que $R > A$ porque $\mathcal{R} > \mathcal{A}$. Esto es absurdo porque $\mathcal{B}$ era collar por hipótesis, y vino de suponer que $\mathcal{A}$ no era collar. Concluimos que $\mathcal{A}$ sí lo es. □

Lema 1. *Sean* n *y* k *enteros positivos tales que* k *divide a* n*, y un collar* $\mathcal{A} = \langle a_1 \cdots a_n, 0\rangle$ *distinto de* $\langle 0^n, 0\rangle$*. Si* $a_i > 0$*, entonces* $\mathcal{B} = \langle A_{i-1}(a_i - 1)(s-1)^{j-i}, 0\rangle$ *también es collar, donde* j *es el mínimo múltiplo de* k *mayor igual que* i*, es decir,* $j = i + ((n - i) \mod k)$*.*

Demostración. Si $\mathcal{B} = \langle A_{i-1}(a_i - 1)(s-1)^{j-i}, 0\rangle$ no es un collar, entonces para algún ℓ:

$$\langle a_{\ell+1} \cdots a_{i-1}(a_i - 1)(s-1)^{j-i} A_\ell, 0 \oplus \ell\rangle > \langle A_{i-1}(a_i - 1)(s-1)^{j-i}, 0\rangle$$

Tenemos dos posibilidades:

- ℓ no es múltiplo de k.
- ℓ es múltiplo de k.

El primer caso lo descartamos porque sabemos que ℓ tiene que cumplir que $0 \oplus \ell = 0$, porque sino $\mathcal{B}$ sería mayor por tener módulo 0. Vamos por el segundo caso, donde ocurre necesariamente:

$$a_{\ell+1} \cdots a_{i-1} \geq a_1 \cdots a_{(i-1)-(\ell+1)+1}$$

Pero dado que $\mathcal{A}$ es un collar y usando que ℓ es múltiplo de k sabemos que:

$$a_1 \cdots a_{(i-1)-(\ell+1)+1} \geq a_{\ell+1} \cdots a_{i-1}.$$

Por lo tanto, $a_{\ell+1} = a_1$, $a_{\ell+2} = a_2$, $\ldots$, $a_{i-1} = a_{(i-1)-(\ell+1)+1}$, y deducimos que $a_i - 1 \geq a_{(i-1)-(\ell+1)+2}$. Pero si esto pasa tenemos que:

$$a_{\ell+1} \cdots a_{i-1} a_i > a_1 \cdots a_{(i-1)-(\ell+1)+2}$$

contradiciendo que $\mathcal{A}$ es un collar. □

Ejemplo. *Para $n = 6$, $k = 3$ y $s = 7$ podemos construir los siguientes collares a partir del collar $\langle 456123, 0\rangle$,*

- *$\langle 455, 0\rangle$: En este caso se tomo el prefijo con $i = 3$, y como este es múltiplo de k no fue necesario rellenar con el símbolo $s - 1$.*
- *$\langle 366, 0\rangle$: En este otro caso se tomo el prefijo con $i = 1$, y sí fue necesario rellenar dos veces con el símbolo $s - 1$ a diferencia del anterior.*

A continuación definimos el operador θ, que dado un par con modulo 0, pero distinto de $\langle 0^n, 0\rangle$, da otro par también módulo 0.

Definición (Operador θ)**.** *Sean n y k dos enteros positivos tales que k divide a n. Dado $\mathcal{A} = \langle A, 0\rangle = \langle a_1 \cdots a_n, 0\rangle$, tal que $a_i > a_{i+1} = \cdots = a_n = 0$, definimos*

$$\langle A, 0\rangle\theta = \langle [A_{i-1}(a_i - 1)(s-1)^{j-i}]^q A_{n-qj}, 0\rangle,$$

donde

- *j es el mínimo múltiplo de k mayor o igual a i. Es decir, $j = i + ((n - i) \bmod k)$,*
- *q es el máximo tal que $0 \leq n - qj < j$.*

Ejemplo. *Supongamos $s = 2$, $n = 6$ y $k = 2$.*
Si $\mathcal{A} = (010000, 0)$ entonces $\mathcal{A}\theta = (000000, 0)$.
Si $\mathcal{A} = (011000, 0)$, entonces $\mathcal{A}\theta = (010101, 0)$.
Si $\mathcal{A} = (011101, 0)$ entonces $\mathcal{A}\theta = (011100, 0)$.

El operador θ es una función de pares de longitud n en pares de longitud n, que no es inyectiva ni suryectiva. Por ejemplo, si tomamos $s = 2$, $n = 4$ y $k = 2$, podemos observar que:

- $\langle 0100, 0\rangle\theta = \langle 0000, 0\rangle$
- $\langle 0001, 0\rangle\theta = \langle 0000, 0\rangle$

Y por lo tanto θ no es inyectiva, ya que tenemos 2 pares que terminan en el mismo par al aplicarles θ. También si analizamos el par $\mathcal{B} = \langle 1011, 0\rangle$, no existe un par $\mathcal{A}$ tal que $\mathcal{A}\theta = \mathcal{B}$, y por lo tanto θ no es suryectiva.

2.2. Collares donde n divide a k

A diferencia del caso k divide a n, ahora *todo* par de la forma $\langle A, 0\rangle$ es un collar. Es sencillo ver esto porque sólo existe una rotación de la forma $\langle B, 0\rangle$ a partir de $\langle A, 0\rangle$, y es con $A = B$, porque k es un múltiplo de n.

Notemos que cuando $n = k$, tenemos $k|n$ y $n|k$, entonces podría ser tratado en cualquiera de los dos casos. Simplemente por claridad expositiva consideraremos los casos disjuntos tomando el primer caso como $k|n$, y el otro caso como $n|k$ con $n < k$.

Definición (Operador θ)**.** *Sean n y k dos enteros positivos tales que n divide a k y $n < k$. Dado $\mathcal{A} = \langle A, 0\rangle = \langle a_1 \cdots a_n, 0\rangle$, tal que $a_i > a_{i+1} = \cdots = a_n = 0$, definimos*

$$\langle A, 0\rangle\theta = \langle A_{i-1}(a_i - 1)(s-1)^{n-i}, 0\rangle.$$

Observación 2. *El operador θ es biyectivo cuando k es múltiplo de n. Para ver esto notemos primero que este pasa por pares de la forma $\langle A, 0\rangle$, dentro de los cuales están todos los collares. Dado que todo par de la forma $\langle A, 0\rangle$ es collar, el θ recorre todos los pares posibles. Esto garantiza que para cualquier par por el que pase θ, podemos encontrar la preimagen que lo generó, que es el par inmediato anterior en el orden lexicográfico.*

2.3. Lista de collares

Observación 3. *Sea $\mathcal{A} = \langle A, 0\rangle$. El operador $\mathcal{A}\theta$ se puede aplicar si y solo si $\mathcal{A} \neq \langle 0^n, 0\rangle$.*

Demostración. $\implies$ Si $\mathcal{A}\theta$ se puede aplicar entonces existe un $a_i > 0$ en A. Luego, necesariamente $A \neq 0^n$.

$\impliedby$ Si $\mathcal{A} \neq \langle 0^n, 0\rangle$, existe un $a_i > 0$, y podemos aplicar $\mathcal{A}\theta$. □

Definición. *Sean n y k enteros positivos tal que k divide a n ó k divide a n. Sea $\mathcal{A} = \langle (s-1)^n, 0\rangle$. Definimos la lista $\mathcal{L}$ de pares dada por la sucesión $(\theta^i\mathcal{A})_{i\geq 0}$.*

Lema 2. *Sean n y k enteros positivos tal que k divide a n. La lista $\mathcal{L}$ de pares es estrictamente decreciente en el orden lexicográfico.*

Demostración. Observemos que para cualquier par $\mathcal{A}$, se cumple que:

$$\mathcal{A} > \mathcal{A}\theta$$

Asumamos que k divide a n. Reemplazando por la definición de θ y de $\mathcal{A}$ tenemos:

$$\langle A, 0\rangle > \langle [A_{i-1}(a_i - 1)(s-1)^{j-i}]^q A_{n-qj}, 0\rangle.$$

Ambos lados tienen modulo 0. La desigualdad es estricta porque si tomamos el prefijo de longitud i de ambas palabras:

$$a_1 \cdots a_i = A_{i-1}a_i > A_{i-1}(a_i - 1).$$

vemos que coinciden en los primeros $i-1$ símbolos, pero el i-ésimo símbolo de $\mathcal{A}$ es mayor. Concluimos que la sucesión $(\theta^i A)_{i\geq 0}$ es estricta decreciente.

El caso de n divide a k es análogo, con la definición correspondiente del θ. □

Es posible construir la lista $\mathcal{L}$ en orden inverso, es decir yendo desde el par lexicográficamente mínimo $\langle 0^n, 0\rangle$ al lexicográficamente máximo $\langle (s-1)^n, 0\rangle$. La única dificultad es que, cuando k divide a n el operador θ no es inyectivo entonces hay pares que tienen más de una preimagen por θ. Sin embargo, en la lista $\mathcal{L}$ cada elemento (salvo el primero) tiene un antecesor, que es exactamente una de las posibles preimagenes por θ.

Lema 3. *Supongamos k divide a n. Sea $\mathcal{A} = \langle A, 0\rangle$ un elemento de la lista $\mathcal{L}$, pero no el primero. Entonces el antecesor de $\mathcal{A}$ en $\mathcal{L}$ es la preimagen por θ con la siguiente forma:*

$$\langle A_{u-1}(a_u + 1))0^{n-u}, 0\rangle$$

donde u proviene de la factorización

$$A = (A_r)^w A_v \text{ con } A_r = A_u(s-1)^{r-u},$$

$a_u < (s-1)$ y r es el mínimo múltiplo de k tal que $v < r$ y $r - k \leq u \leq r$.

Demostración. Primero notemos que esta factorización

$$A = (A_r)^w A_v = (A_u(s-1)^{r-u})^w A_v$$

siempre existe. En el peor caso $r = n$, $v = 0$ y $A = A_n = A_r = A_u(s-1)^{r-u}$. Para encontrar la preimagen buscada de $\mathcal{A}$ por θ, llamémosla $\mathcal{B}$, simplemente deshacemos lo que hace el operador, sabiendo que $\mathcal{A}$ y $\mathcal{B}$ están en $\mathcal{L}$. Buscamos $\mathcal{B} = \langle B, 0\rangle$ tal que

$$\mathcal{B}\theta = \langle [B_{i-1}(b_i - 1)(s-1)^{j-i}]^q B_{n-qj}, 0\rangle = \mathcal{A}$$

donde i es tal que $b_{i+1} = \ldots b_n = 0$, y j es el mínimo múltiplo de k mayor o igual que i. Es decir, buscamos B tal que

$$[B_{i-1}(b_i - 1)(s-1)^{j-i}]^q B_{n-qj} = (A_u(s-1)^{r-u})^w A_v.$$

La palabra B queda definida al identificar

$$\begin{aligned} r &= j \\ u &= i \\ w &= q \\ v &= n - qj \\ A_{i-1}(a_i + 1) &= B_i. \end{aligned}$$

Luego,

$$\begin{aligned} \langle B_{i-1} b_i 0^{n-j}, 0\rangle &= \langle A_{u-1}(a_u + 1))0^{n-u}, 0\rangle, \\ \langle [B_{i-1}(b_i - 1)(s-1)^{j-i}]^q B_{n-qj}, 0\rangle &= [A_{u-1} a_u (s-1)^{r-u}]^w A_v. \end{aligned}$$

□

Lema 4. *Si $\mathcal{A} > \mathcal{B} > \mathcal{C}$, con $\mathcal{C} = \mathcal{A}\theta$, entonces $\mathcal{B}$ no es un collar.*

Demostración. Sean $\mathcal{A} := \langle A, 0\rangle$, $\mathcal{B} := \langle B, 0\rangle$. y sea $\mathcal{C} := \mathcal{A}\theta = \langle C, 0\rangle$ con

$$\begin{aligned} A &:= a_1 \cdots a_i 0^{n-i}, \text{con } a_i > 0, \\ B &:= b_1 \cdots b_n, \\ C &:= c_1 \cdots c_n = a_1 \cdots a_{i-1}(a_i - 1)(s-1)^{j-i}]^q a_1 \cdots a_{n-qj}. \end{aligned}$$

Asumamos $\mathcal{A} > \mathcal{B} > \mathcal{C}$ y supongamos $\mathcal{B}$ es un collar. Necesariamente,

$$b_1 = a_1, \quad b_2 = a_2, \quad \ldots, \quad b_{i-1} = a_{i-1},$$

porque $\mathcal{A} > \mathcal{B}$. Y debe ocurrir que $a_i > b_i$, pero al mismo tiempo como $\mathcal{B} > \mathcal{C}$ tiene que pasar que $b_i \geq c_i = a_i - 1$, lo cual implica que $b_i = a_i - 1$.

Hasta ahora sabemos que $B_i = C_i$. Notemos también que tiene que pasar que $b_{i+\ell} = s-1$ para $\ell = 1, \ldots j-i$. Esto último vale porque si $\mathcal{B} \geq \mathcal{C}$ y $c_{i+1} \cdots c_{i+\ell} = (s-1)^{j-i}$, entonces $b_{i+1} \cdots b_{i+\ell} = (s-1)^{j-i}$ porque no hay un símbolo mas grande que $s-1$. Veamos ahora que el resto de los símbolos de $\mathcal{B}$ y $\mathcal{C}$ también coinciden. Como $\mathcal{B}$ es collar, vale que $B \geq b_{j+1} \cdots b_n b_1 \cdots b_j$ y sabemos que $a_1 = b_1 \geq b_{j+1}$. Como $\mathcal{B} \geq \mathcal{C}$ tenemos que $b_{j+1} \geq c_{j+1} = a_1$, y luego $b_{j+1} = a_1$. A partir de esto se repite el mismo argumento que antes, y luego sucede que para todo $1 \leq m \leq q$, $1 \leq p < i$:

- $b_{mj+p} = a_p$ si $1 \leq p < i$ y $mj + p \leq n$.
- $b_{mj+i} = a_i - 1$.
- $b_{mj+i+\ell} = s - 1$ con $1 \leq \ell \leq j - i$,

Esto resulta en que $\mathcal{B} = \mathcal{C} = A\theta$, y por lo tanto no existe un $\mathcal{B}$ que cumpla $\mathcal{A} > \mathcal{B} > \mathcal{C}$ y sea collar. □

Definición. *Sean n y k enteros positivos tales que k divide a a n o n divide a k. Definimos la lista $\mathcal{C}$ de collares de longitud n y módulo k.*

En el caso de n divide a k y $n < k$ la lista $\mathcal{C}$ es directamente la lista $\mathcal{L}$. En el caso k divide a n debemos construir $\mathcal{C}$ quitando de la lista $\mathcal{L}$ aquellos pares que no son collares. El primer collar en la lista $\mathcal{C}$ de collares es $((s-1)^n, 0)$. Y para cada elemento A de la lista $\mathcal{L}$ excepto para $\mathcal{A} = \langle 0^n, 0\rangle$, el sucesor de $\mathcal{A}$ en $\mathcal{C}$ es $\mathcal{A}\theta^h$, donde h es el mínimo número de aplicaciones del operador θ tal que $\mathcal{A}\theta^h$ es un collar. El último elemento de la lista $\mathcal{C}$ es $\langle 0^n, 0\rangle$.

Ejemplo. *Con $s = 2$, $n = 6$ y $k = 2$, el algoritmo genera la siguiente lista (leerla de arriba hacia abajo, y luego sigue la siguiente columna en orden).*

- $\langle 111111, 0\rangle$
- $\langle 111110, 0\rangle$
- $\langle 111101, 0\rangle$
- $\langle 111100, 0\rangle$
- $\langle 111010, 0\rangle$
- $\langle 111001, 0\rangle$
- $\langle 111000, 0\rangle$
- $\langle 110110, 0\rangle$
- $\langle 110101, 0\rangle$
- $\langle 110100, 0\rangle$
- $\langle 110010, 0\rangle$
- $\langle 110001, 0\rangle$
- $\langle 110000, 0\rangle$
- $\langle 101010, 0\rangle$
- $\langle 101001, 0\rangle$
- $\langle 101000, 0\rangle$
- $\langle 100101, 0\rangle$
- $\langle 100100, 0\rangle$
- $\langle 100001, 0\rangle$
- $\langle 100000, 0\rangle$
- $\langle 010101, 0\rangle$
- $\langle 010100, 0\rangle$
- $\langle 010000, 0\rangle$
- $\langle 000000, 0\rangle$

Teorema 1. *Sean n y k enteros positivos tales que k divide a n ó n divide a k. La lista $\mathcal{C}$ tiene todos los collares de longitud n y módulo k ordenados decrecientemente.*

Demostración. Caso $k \mid n$:
Para probar esto basta ver que :

1. *La lista $\mathcal{C}$ empieza con el par mas grande, $\langle (s-1)^n, 0\rangle$.*
2. *Agrega collares de manera ordenada decreciente.*
3. *No se pierde ningún collar.*
4. *Termina en el par lexicográficamente mínimo, $\langle 0^n, 0\rangle$.*

Veamos cada uno de estos puntos.

1. *Esto es trivial, porque lo definimos así.*
2. *La lista $\mathcal{C}$ proviene de aplicar sucesivas veces el operador θ, y por el Lema 2 sabemos que este es estrictamente decreciente.*
3. *Dado un collar $\mathcal{A}$ en $\mathcal{C}$, su sucesor es $\mathcal{A}\theta^h$. Si existiese un $\mathcal{B}$ collar tal que está en el medio, necesariamente cae en uno de estos 3 casos:*

   - $\mathcal{A} > \mathcal{B} > \mathcal{A}\theta$
   - $\mathcal{A}\theta^{h-1} > \mathcal{B} > \mathcal{A}\theta^h$
   - $\mathcal{A} > \mathcal{A}\theta > \ldots > \mathcal{A}\theta^i > \mathcal{B} > \mathcal{A}\theta^{i+1} > \ldots > \mathcal{A}\theta^{h-1} > \mathcal{A}\theta^h$.

   En cualquiera de estos tres casos, el $\mathcal{B}$ siempre se encuentra entre un $\mathcal{D}$ y su sucesor $\mathcal{D}\theta$. Pero sabemos que esto es imposible por el Lema 4.
4. *Si la lista $\mathcal{C}$ no terminara en el collar $\langle 0^n, 0\rangle$, entonces ocurriría alguna de estas:*

   - *Pasa por $\langle 0^n, 0\rangle$ y sigue agregando otros collares. Esto es imposible por el punto (2) y que $\langle 0^n, 0\rangle$ es el mínimo collar.*
   - *Pasa de un $\mathcal{D} > \langle 0^n, 0\rangle$ a otro $\mathcal{D}\theta < \langle 0^n, 0\rangle$. Esto no pasa por el Lema 4.*
   - *Nunca llega a $\langle 0^n, 0\rangle$ porque frena antes. Sabemos por el punto (2) que el θ es monótono estricto decreciente, y por el Lema 3 siempre puede aplicarse, a menos que llegue a este collar $\langle 0^n, 0\rangle$. Por lo tanto no frena antes.*

Luego, dado que la lista $\mathcal{C}$ empieza por el collar lexicográficamente máximo, termina en el mínimo, agrega collares de manera decreciente, y no omite ninguno, queda demostrado el teorema.

Caso $n \mid k$, $k > n$: *Es la misma demostración que para el caso anterior, pero más simple ya que el operador θ recorre exactamente todos los collares en orden lexicográfico.* □

3. ALGORITMO PARA COLLARES PERFECTOS LEXICOGRÁFICAMENTE MÁXIMOS

Definición (collar (n,k)-perfecto)**.** *Dado un alfabeto Σ de s símbolos y enteros positivos n y k, un (n,k)-collar perfecto es un collar de longitud ks^n, tal que todas las palabras de longitud n aparecen exactamente k veces, en posiciones que son todas diferentes en modulo k.*

Ejemplo. *Con $s = 2$, $n = 2$ y $k = 2$, el (n,k)-collar perfecto es* 11100100*, donde están presenten los siguientes pares:*

- $\langle 11, 0\rangle$
- $\langle 10, 0\rangle$
- $\langle 01, 0\rangle$
- $\langle 00, 0\rangle$
- $\langle 11, 1\rangle$
- $\langle 00, 1\rangle$
- $\langle 10, 1\rangle$
- $\langle 01, 1\rangle$

3.1. Caso k divide a n

Definición (reducción periódica)**.** *Dados dos enteros positivos n y k tales que k divide a n, la reducción periódica de la palabra $A = a_1 \cdots a_n$, a la que denotamos $\overline{A}$ es la palabra $a_1 \cdots a_p$ con el mínimo p tal que $k \mid p$, $p \mid n$ y $a_1 \cdots a_n = (a_1 \cdots a_p)^{n/p}$. Definimos la reducción del par $\mathcal{A} = \langle A, m\rangle$, a la que denotamos $\overline{\mathcal{A}}$, como el par $\langle \overline{A}, m\rangle$.*

Notar que la reducción periódica siempre se puede aplicar porque p como máximo es n, que cumple $k \mid n$, $n \mid n$ y $a_1 \cdots a_n = (a_1 \cdots a_n)^1$.

Ejemplo. *Para $s = 10$, $n = 8$ y $k = 2$, podemos aplicar estas reducciones:*

- $\overline{\langle 45454545, 0\rangle} \coloneqq \langle 45, 0\rangle$
- $\overline{\langle 12341234, 0\rangle} \coloneqq \langle 1234, 0\rangle$
- $\overline{\langle 12345678, 0\rangle} \coloneqq \langle 12345678, 0\rangle$

Observación 4. *Si A es collar, entonces no tiene ninguna rotación mayor, pero puede tener una rotación igual. Sin embargo, si $\mathcal{A}$ es un collar reducido entonces no tiene ninguna rotación mayor ni igual. Esto es porque todas las rotaciones de un collar reducido son distintas entre si.*

Algoritmo 1. *Dados n y k enteros positivos tales que k divide a n, podemos construir un (n,k)-collar perfecto lexicográficamente máximo concatenando en orden las palabras de los pares que surgen de las reducciones periódicas de los collares la lista $\mathcal{C}$ generada por la Definicion 2.3.*

Si tenemos $s = 2$, $n = 6$ y $k = 2$, podemos construir el (n,k)-collar perfecto:

11 | 111110 | 111101 | 111100 | 111010 | 111001 | 111000 | 110110 | 110101 | 110100 | 110010 | 110001 | 110000 | 10 | 101001 | 101000 | 100101 | 100100 | 100001 | 100000 | 01 | 010100 | 010000 | 00

Usamos | para simbolizar la concatenación entre palabras.

Lema 5. *Si $\mathcal{A}$ y $\mathcal{B}$ son dos collares vecinos en la lista de collares $\mathcal{C}$ entonces $\mathcal{A}$ es el prefijo de $\overline{\mathcal{AB}}$.*

Demostración. Dado que $\mathcal{A}$ y $\mathcal{B}$ son collares vecinos en la lista, $\mathcal{B} = \mathcal{A}\theta^h$. Podemos reescribir $\mathcal{A}$ como $\overline{\mathcal{A}}^{n/p} = \langle (A_i 0^{p-i})^q, 0\rangle$, donde $q = n/p$, $a_i > 0$. Luego dividimos en casos según el q:

- Si $q = 1$, $\overline{\mathcal{A}} = \mathcal{A}$ y $\overline{\mathcal{AB}} = A\overline{\mathcal{B}}$.
- Si $q > 1$ entonces $\mathcal{B} = \langle \overline{A}^{q-1} A_{i-1}(a_i - 1)(s-1)^{j-i} C, 0\rangle$ para alguna palabra C y j el mínimo tal que $i \leq j \leq p$ y $k \mid j$. Finalmente,

$$\overline{\mathcal{AB}} = \overline{\mathcal{A}}\langle \overline{A}^{q-1} A_{i-1}(a_i - 1)(s-1)^{j-i} C, 0\rangle = \mathcal{A}\langle \mathcal{A}_{i-1}(a_i - 1)(s-1)^{j-i} C, 0\rangle.$$

En todos los casos $\mathcal{A}$ es prefijo de $\overline{\mathcal{AB}}$, por lo tanto el lema quedó demostrado. □

3.2. Caso n divide a k

La reducción periódica del caso k divide a n pasa ahora a ser expansión periódica.

Definición (expansión periódica)**.** *Dados dos enteros positivos n y k tales que n divide a k y $n < k$, la expansión periódica de la palabra $A = a_1 \cdots a_n$, a la que denotamos $\tilde{A}$, es la palabra $A^{k/n}$. Definimos la expansión del par $\mathcal{A} = \langle A, m\rangle$, a la que denotamos $\tilde{\mathcal{A}}$, como el par $\langle \tilde{A}, m\rangle$.*

Notar que la reducción periódica siempre se puede aplicar porque k es múltiplo de n.

Ejemplo. *Para $s = 10$, $n = 2$ y $k = 8$, podemos aplicar estas reducciones:*

- $\overline{\langle 45, 0\rangle} \coloneqq \langle 45454545, 0\rangle$
- $\overline{\langle 12, 0\rangle} \coloneqq \langle 12121212, 0\rangle$

Algoritmo 2. *El algoritmo 1 en este caso es exactamente el mismo que antes, pero utilizando la definición de expansión periódica en lugar de reducción periódica.*

3.3. El resultado principal

Teorema 2. *Sean n y k enteros positivos. Si k divide a n, el Algoritmo 1 genera un (n,k)-collar perfecto lexicográficamente máximo. En cambio, si n divide a k y $n < k$, la versión más simple dada por el Algoritmo 2 lo genera.*

3.3.1. Demostración Teorema 2 del caso k divide a n

En total la longitud de la palabra construida es

$$\sum_{\overline{\mathcal{A}} \text{ es collar}} |\overline{\mathcal{A}}|.$$

Para cada collar $\mathcal{A}$, la longitud de su reducción $\overline{\mathcal{A}}$ es la cantidad de rotaciones de la clase de equivalencia del collar $\mathcal{A}$, lo cual se deduce de la Observación 4. Entonces, $\overline{\mathcal{A}}$ da origen a tantos pares $\langle a_1 \cdots a_n, m\rangle$ como su longitud. Dado que concatenamos todos los collares reducidos exactamente una vez, concluimos que en la palabra construida hay exactamente una posición para cada uno de todos los pares $\langle a_1 \cdots a_n, m\rangle$, con $a_i \in \Sigma$, $m \in \mathbb{Z}/k\mathbb{Z}$. Concluimos que la longitud de la palabra construida es

$$ks^n.$$

Para demostrar que la secuencia construida es un (n,k)-collar perfecto necesitamos ver que cada palabra de longitud n ocurre exactamente k veces, en posiciones que son distintas módulo k. Equivalentemente, debemos ver que todas las rotaciones de cada collar concatenado por el Algoritmo 1 aparecen en el (n,k)-collar perfecto. Es decir, para cada collar $\overline{\mathcal{A}}$ que concatenó el Algoritmo 1, tenemos que identificar todas las rotaciones $\mathcal{B} = \mathcal{A} \ominus i$, para $i = 0, \ldots, p-1$, donde $p = |\overline{\mathcal{A}}|$. Notemos que p es múltiplo de k, entonces siempre se llega hasta $k-1$ o más.

Consideremos los pares $\mathcal{N} = \langle A, 0\rangle$.

Caso $A = 0^n$. Recordemos que k divide a n, entonces el collar reducido es $\overline{\mathcal{N}} = \langle 0^k, 0\rangle$. Consideremos los pares de la forma

$$\mathcal{M} = \langle M, 0\rangle = \langle 0^i 10^{k-1-i} 0^{n-k}, 0\rangle.$$

para $i < k$. Todos estos son collares, porque cualquier otra rotación que lo deja en modulo 0 es lexicográficamente menor que la original que tenia el único 1 entre los primeros k símbolos. El siguiente collar a $\mathcal{M}$ en el orden lexicográfico es:

$$\mathcal{M}\theta = \mathcal{Q} = \langle Q, 0\rangle = \langle (0^{i+1}(s-1)^{k-1-i})^{n/k}, 0\rangle.$$

Es claro que $\mathcal{Q}$ es collar, porque todas las rotaciones que tienen módulo 0 siempre van a ser idénticas a $\mathcal{Q}$, ya que $\overline{\mathcal{Q}} = \langle 0^{i+1}(s-1)^{k-1-i}, 0\rangle$.

Concluimos que el Algoritmo 1, pone $\overline{\mathcal{M}}$ e inmediatamente después pone $\overline{\mathcal{Q}}$. Notemos que en

$$\overline{\mathcal{M}\mathcal{Q}} = \langle 0^i 10^{k-1-i} 0^{n-k} 0^{i+1} (s-1)^{k-i-1}, 0\rangle = \langle 0^i 10^n (s-1)^{k-i-1}, 0\rangle$$

aparece el par $\langle 0^n, (i+1) \mod k\rangle$ que corresponde al segmento 0^n. Como esto pasa para cada posible i, tal que $0 \leq i < k$, tenemos todas las rotaciones de $\mathcal{N} = \langle 0^n, 0\rangle$ que buscábamos, desde $\langle 0^n, 0\rangle$ hasta $\langle 0^n, k-1\rangle$.

Caso $A = (s-1)^n$**.** Recordemos una vez mas que como k divide a n, el collar reducido es $\overline{\mathcal{N}} = \langle (s-1)^k, 0\rangle$. El sucesor de $\mathcal{N}$ es

$$\mathcal{N}\theta = \mathcal{M} = \langle M, 0\rangle = \langle (s-1)^{n-1}(s-2), 0\rangle.$$

Al ser concatenados por el Algoritmo 1 queda:

$$\overline{\mathcal{N}\mathcal{M}} = \langle (s-1)^k, 0\rangle\langle (s-1)^{n-1}(s-2), 0\rangle = \langle (s-1)^{n+k-1}(s-2), 0\rangle.$$

Esto implica que tenemos todas las rotaciones de $\mathcal{N}$, desde $\langle (s-1)^n, 0\rangle$ hasta $\langle (s-1)^n, k-1\rangle$.

Caso $0^n < A < (s-1)^n$**.** Sabemos que todo collar $\mathcal{N} = \langle A, 0\rangle$ distinto de $\langle 0^n, 0\rangle$ tiene la forma $\mathcal{N} = \langle (A_p)^{q+1}, 0\rangle$, donde A_p es la palabra reducida de A, con p el mínimo entero tal que $A = (A_p)^{q+1}$, y $q+1 = (n/p)$.

Consideremos primero $q > 0$. El collar $\mathcal{N}$, que es distinto al $\langle 0^n, 0\rangle$, siempre tiene un collar sucesor $\mathcal{M}$ de la forma:

$$\mathcal{M} = \mathcal{N}\theta = \langle (A_p)^q A_{i-1}(a_i - 1)(s-1)^{j-i} b_{i+1}\cdots b_p, 0\rangle,$$

donde j es el mínimo múltiplo de k con $j \geq i$. Notar que $\mathcal{M}$ tiene esta forma porque $A_p \neq 0^p$. Como $\overline{\mathcal{M}} = \mathcal{M}$

$$\overline{\mathcal{N}\mathcal{M}} = \mathcal{N}\langle A_{i-1}(a_i - 1)(s-1)^{j-i} b_{i+1}\cdots b_p, 0\rangle,$$

y luego las primeras i rotaciones a izquierda de $\mathcal{N}$ ocurren en esta parte. Estas son $\langle a_{r+1}\cdots a_n A_r, 0 \oplus r\rangle$, con $0 \leq r < i$.

Resta encontrar las $p-i$ rotaciones a derecha. Notemos que

$$\mathcal{N} = \langle (A_p)^{q+1}, 0\rangle = \mathcal{Q}\theta$$

donde

$$\mathcal{Q} = \langle A_{p-1}(a_p + 1)0^{pq}, 0\rangle.$$

Para ver que $\mathcal{Q}$ es collar analicemos primero $\mathcal{R} = \langle A_p, 0\rangle$, que es collar por la Observación 1. A_p no tiene ningún sufijo propio $S = a_i \cdots a_p$ que coincida con su prefijo $P = A_{p-i+1}$. Si lo tuviera podríamos construir $\mathcal{D} = \langle SA_{i-1}, 0\rangle$ como la rotación de $\mathcal{R}$ que deja a S como prefijo, y pasaría alguno de estos:

- $\mathcal{D} = \mathcal{R}$: La Observación 4 dice que no existe tal rotación.
- $\mathcal{D} < \mathcal{R}$: Como $S = A_{p-i+1}$, necesariamente $A_{i-1} < a_{p-i+2}\cdots a_p$, y luego $\mathcal{R}$ no sería un collar.
- $\mathcal{D} > \mathcal{R}$: Encontré una rotación mayor que $\mathcal{R}$, contradiciendo que $\mathcal{R}$ es collar.

Como $\mathcal{D}$ no puede existir, concluimos que todos los sufijos $S = a_i \cdots a_p$ de $\mathcal{R}$ son menores que su prefijo $P = A_{p-i}$. Luego cualquier sufijo propio de $\langle A_{p-1}(a_p+1), 0\rangle$, $S = a_i \cdots a_p$, es a lo sumo igual a su prefijo $P = a_i \cdots a_{p-1}(a_p + 1)$. Finalmente $\mathcal{Q}$ es collar, porque toda otra rotación va a ser menor que esta:

- Si es de la forma $\langle a_i \cdots a_{p-1}(a_p+1)0^{pq}a_1 \cdots a_{i-1}, 0\rangle$, el sufijo $a_i \cdots a_{p-1}(a_p+1)$ es a lo sumo igual al prefijo de $\mathcal{Q}$ de longitud $p-i$. Luego de esto aparece 0^{pq}, entonces esta rotación es menor porque $\mathcal{Q}$ tiene el símbolo $a_p+1>0$.
- Si empieza con un sufijo de 0^{pq} de longitud m, necesariamente el prefijo $Q_m > 0^m$, porque sino $\mathcal{N}$ no seria collar.

Llamemos $\mathcal{Q} = \langle Q, 0\rangle$ y $\mathcal{N} = \langle N, 0\rangle$. Observemos que $\overline{Q} = Q$ y

$$\overline{QNM} = Q\overline{N}M$$

contiene a

$$\langle 0^{pq}A^{q+1}A_{i-1}, 0\rangle,$$

donde el modulo 0 se da por construcción de Q, al ser p múltiplo de k, si le sacamos los primeros p símbolos seguimos en modulo 0. Finalmente, dentro de la secuencia $Q\overline{N}M$ se encuentran las primeras $p-i$ rotaciones a derecha, que son de la forma

$$\langle 0^r A^q A_i 0^{p-i-r}, 0 \ominus r\rangle,$$

con $1 \leq r < p-i$. Como obtuvimos las primeras i rotaciones a izquierda y $p-i$ a derecha, obtuvimos todas las p rotaciones.

Consideremos ahora $q=0$. Debemos ver que para los collares

$$\mathcal{N} = \langle N, 0\rangle = \langle A^{q+1}, 0\rangle$$

con $q=0$ podemos encontrar todas las rotaciones.

Si $N = 0^{k-1}10^{n-k}$, entonces $\mathcal{M} = \langle M, 0\rangle = \mathcal{N}\theta = \langle 0^n, 0\rangle$, es el último collar de la lista $\mathcal{C}$. Luego

$$\overline{NM} = 0^{k-1}10^{n-k}0^k = 0^{k-1}10^n,$$

y podemos encontrar las primeras k rotaciones a izquierda de la forma

$$\langle 0^{k-1-r}10^{n-k+r}, 0 \oplus r\rangle, \text{ con } 0 \leq r \leq k-1.$$

Si $N \neq 0^{k-1}10^{n-k}$, entonces $N = \langle A_j 0^{n-j}, 0\rangle$, con j el mínimo múltiplo de k tal que $a_{j+1}...a_n = 0^{n-j}$, $\mathcal{M} = \langle M, 0\rangle = \mathcal{N}\theta$ y $\mathcal{P} = \langle P, 0\rangle = \mathcal{M}\theta = \mathcal{N}\theta^2$. Por el Lema 5 tenemos

$$\overline{NMP} = NMP$$

y además que

$$M = A_{i-1}(a_i - 1)(s-1)^{j-i}b_{i+1}\cdots b_p,$$

con i tal que $a_i > a_{i+1} = \ldots = a_n = 0$, y j el mínimo múltiplo de k con $j \geq i$. Concluimos que aparecen las primeras i rotaciones a izquierda de $\mathcal{N}$, que son de la forma

$$\langle a_{r+1}\cdots a_n A_r, 0 \oplus r\rangle, \text{ con } 0 \leq r \leq i-1.$$

Debemos hallar las rotaciones a derecha de $\mathcal{N}$ que faltan, que son de la forma $0^{n-j-h}A_j 0^h$. Si $h=0$ y $A_j = (s-1)^j$, el par $\langle 0^n(s-1)^n, 0\rangle$, que aparece en la concatenación de los últimos 2 collares de la lista $\mathcal{C}$ con los primeros 2, contiene al

par $\langle 0^{n-j}(s-1)^j, 0 \ominus (n-j)\rangle$, y la prueba para el caso $h = 0$ y $A_j = (s-1)^j$ queda concluida.

Para los demás casos, donde $A_j \neq (s-1)^j$ hallaremos tales rotaciones a derecha de $\mathcal{N}$ que faltan entre tres collares reducidos concatenados por el Algoritmo 1, a los que llamaremos $\overline{\mathcal{Q}_t\mathcal{P}\mathcal{R}}$. A continuación los definimos.

Recordemos que $\mathcal{N} = \langle N, 0\rangle = \langle A_j 0^{n-j}, 0\rangle$ es collar. Sea $\mathcal{Q}$ un par en la lista $\mathcal{L}$ de la forma

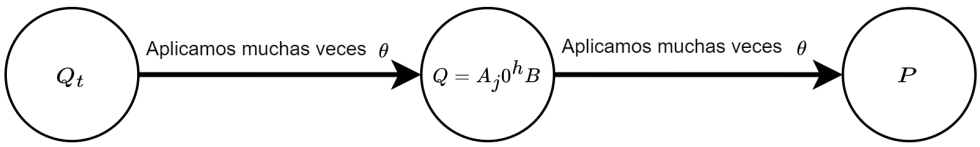
$$\mathcal{Q} = \langle A_j 0^h B, 0\rangle,$$

con $|B| = n - (j+h)$ no nulo. Notar que para cada valor de h hay exactamente una palabra B. El par $\mathcal{Q}$ existe en $\mathcal{L}$ porque $\mathcal{Q}$ es lexicográficamente mayor o igual que $\mathcal{N}$, pero nunca menor dado que $\mathcal{N}$ es $\langle A_j 0^{n-j}, 0\rangle$. Si empezamos en $\mathcal{Q}$ y aplicamos el operador θ sucesivas veces obtenemos el collar $\mathcal{N}$.

Para $i = 1, 2, 3, ...$ sea $\mathcal{Q}_i$ el par de la lista $\mathcal{L}$ tal que $\mathcal{Q}_i\theta^i = \mathcal{Q}$. Entonces

$$\mathcal{Q}_1 = \langle Q_1, 0\rangle = \langle A_{u_1-1}(a_{u_1}+1)0^{n-u_1}, 0\rangle,$$

donde $r - k \leq u_1 \leq r$, $a_{u_1} < (s-1)$ y $a_{u_1+1} \dots a_r = (s-1)^{r-u_1}$. Visualmente podemos verlo así:



Si nos paramos en $A_j 0^h B$, tanto cuando nos movemos para atrás como para adelante usando el operador θ, nos mantenemos siempre en los pares de $\mathcal{L}$. Entonces

$$\mathcal{Q}_1\theta = \langle (A_{u_1}(s-1)^{r_1-u_1})_1^w A_{v_1} B, 0\rangle = \langle (A_{r_1})_1^w A_{v_1}, 0\rangle = \langle A_j 0^h B, 0\rangle = \mathcal{Q},$$

con r_1 el mínimo múltiplo de k tal que $v_1 < r_1$, $a_{u_1} < (s-1)$, $A_{r_1} = A_{u_1}(s-1)^{r_1-u_1}$.

Si $\mathcal{Q}_1$ es collar entonces fijamos $t = 1$ y terminamos la búsqueda. Sino consideramos al antecesor de $\mathcal{Q}_1$ en $\mathcal{L}$,

$$\mathcal{Q}_2 = \langle Q_2, 0\rangle = \langle A_{u_2-1}(a_{u_2}+1)0^{n-u_2}, 0\rangle$$

tal que

$$\mathcal{Q}_2\theta = \langle (A_{u_2}(s-1)^{r_2-u_2})^{w_2} A_{v_2}, 0\rangle = \langle (A_{r_2})^{w_2} A_{v_2}, 0\rangle = \mathcal{Q}_1.$$

donde r_2 es el mínimo múltiplo de k, $v_2 < r_2$, y $a_{u_2} < (s-1)$ tal que $A_{r_2} = A_{u_2}(s-1)^{r_2-u_2}$. Sabemos que este $a_{u_2} < s-1$ existe porque $\mathcal{Q}_1 \neq \langle ((s-1)^n, 0\rangle$ (porque sino $\mathcal{Q}_1$ seria collar), y luego existe un $a_i < s-1$ para $1 \leq i \leq r_2$. Notemos que $u_2 < u_1$. Si $\mathcal{Q}_2$ es un collar entonces fijamos $t = 2$ y terminamos la búsqueda. Sino, seguimos repitiendo este procedimiento. De esta forma el antecesor de $\mathcal{Q}_{i-1}$ es

$$\mathcal{Q}_i = \langle Q_i, 0\rangle = \langle A_{u_i-1}(a_{u_i}+1)0^{n-u_i}, 0\rangle,$$

tal que

$$\mathcal{Q}_i\theta = \langle (A_{u_i}(s-1)^{r_i-u_i})^{w_i} A_{v_i}, 0\rangle = \langle (A_{r_i})^{w_i} A_{v_i}, 0\rangle = \mathcal{Q}_{i-1},$$

donde r_i es múltiplo de k, $v_i < r_i$, $u_i < u_{i-1}$ y $a_{u_i} < (s-1)$ tal que

$$A_{r_i} = A_{u_i}(s-1)^{r_i - u_i}.$$

Eventualmente hallamos t tal que Q_t es un collar. El collar $\mathcal{Q}_t$ existe porque en cada paso consideramos un par lexicográficamente mayor que el anterior en la lista $\mathcal{L}$, entonces a lo sumo llegaremos al $\langle (s-1)^n, 0\rangle$, que es el par lexicográficamente máximo y es collar.

Consideremos ahora los tres collares sucesivos de la lista $\mathcal{C}$: $\mathcal{Q}_t$, el sucesor de $\mathcal{Q}_t$, y el sucesor de este sucesor. Llamémoslos $\mathcal{Q}_t, \mathcal{P}$ y $\mathcal{R}$ respectivamente. Consideremos ahora los correspondientes collares reducidos, $\overline{\mathcal{Q}_t}, \overline{\mathcal{P}}$ y $\overline{\mathcal{R}}$. Notemos que $\overline{\mathcal{Q}_t}$ termina con 0^{n-u_t} y, por Lema 5, $\overline{\mathcal{P}\mathcal{R}}$ comienza con $\mathcal{P}$. Por lo tanto, $\overline{\mathcal{Q}_t\mathcal{P}\mathcal{R}}$ contiene a $0^{n-u_t}P = 0^{n-u_t}A_j0^hC$. Vale aclarar que $\mathcal{P}$ va a ser un collar mayor o igual que $\mathcal{N}$, y por lo tanto podemos afirmar que tiene este prefijo A_j0^h.

Finalmente, sabemos que vale

$$u_t \leq u_1 \leq j + h,$$

porque

$$u_t < u_{t-1} < \ldots < u_1,$$

y

$$u_1 \leq j + h$$

porque u_1 era la posición de un símbolo dentro de A_j0^h, que tiene longitud $j + h$. Entonces,

$$\begin{aligned} u_t \leq u_1 \leq j+h &\iff -u_t > -u_1 > -j-h \\ &\iff n-u_t > n-u_1 > n-j-h \\ &\iff n-u_t > n-j-h. \end{aligned}$$

Concluimos que $\overline{\mathcal{Q}_t\mathcal{P}\mathcal{R}}$ contiene a

$$\langle 0^{n-j-h}A_j0^h, (|\overline{Q_h}| - (n-j-h)) \mod k\rangle.$$

Esto último también se puede ver como el par

$$\langle 0^{n-j-h}A_j0^h, -(n-j-h) \mod k\rangle,$$

dado que $|\overline{Q_t}|$ es múltiplo de k, y por lo tanto tiene modulo 0.

A continuación argumentamos que el Algoritmo 1. produce el collar (n,k) perfecto lexicográficamente máximo. Por definición un collar es el lexicográficamente máximo entre sus rotaciones. Sabemos que en cada paso el algoritmo selecciona, entre todos los collares aún no usados, el lexicográficamente máximo, y agrega al final su versión reducida. A pesar de que se agrega la versión reducida, podemos razonar acerca del collar sin reducir, ya que por el Lema 5, la concatenación de los collares reducidos dan origen al collar sin reducir. La construcción determina $a_1, a_2, \ldots$ tal que cada uno de los pares

$(a_1, \ldots a_n, 0)$,

$(a_2, \ldots a_{n+1}, 1)$,
$(a_3, \ldots a_{n+2}, 2)$,
...
$(a_{n+1}, \ldots a_{2n}, 0)$,
$(a_{2n+1}, \ldots a_{3n}, 1)$
...
$(a_{ks^n-n+1}, \ldots a_{ks^n}, 1)$

es lexicográficamente máximo entre los aun no usados. Nuestra construcción implícitamente construye una función $f : \Sigma^n \to \Sigma$, tal que

$$a_{i+n} = f(a_i, a_{i+1}, ..., a_{i+n-1}).$$

La secuencia $a_1, a_2, \ldots a_{ks^n}$ es la lexicográficamente máxima entre todos los collares perfectos porque para cada posición $i + n$ el símbolo a_{i+n} es el símbolo lexicográficamente máximo entre todos los símbolos que podrían ponerse para definir un collar perfecto. Por lo tanto, si $b_1, b_2, \ldots, b_{ks^n}$ es otro collar (n, k)-perfecto entonces necesariamente hay una posición ℓ tal que $b_1 = a_1, b_2 = a_2, \ldots b_\ell = a_\ell$ y $b_{\ell+1} < a_{\ell+1}$.

□

3.3.2. Demostración del Teorema 2 del caso n divide a k

Es la misma idea que en el caso k divide a n, pero utilizando las definiciones correspondientes del θ, la lista $\mathcal{L}$ y la lista $\mathcal{C}$. Es más simple, porque requiere menos casos, y cada caso es más sencillo.

Encontrar rotaciones de la forma 0^k es igual al caso k divide a n, solo que ahora $k \geq n$. Encontrar rotaciones a izquierda para palabras A que satisfacen $0^n < A < (s-1)^n$ también es igual al caso k divide a n, ya que la cantidad de posiciones i a considerar en ambas definiciones del θ coincide.

Sin embargo, encontrar rotaciones a derecha para palabras A que satisfacen $0^n < A < (s-1)^n$ es ahora más simple porque el θ es biyectivo. Y dado que todos los pares obtenidos por el operador son collares, simplemente debemos encontrar el par anterior. En el caso k divide a n era necesario justificar cómo encontramos el último collar Q_t del θ, porque no es biyectivo.

□

4. EJEMPLOS DE COLLARES PERFECTOS LEXICOGRÁFICAMENTE MÁXIMOS

A continuación exhibimos collares (n, k)-perfectos lexicográficamente máximos para distintos alfabetos y distintos valores de n y k.

Se distingue dentro del (n,k)-collar perfecto a los collares reducidos utilizados mediante los colores rojo y azul.

Alfabeto $\{0, 1\}$

$n = 2, k = 2$

11100100

$n = 4, k = 2,$

1111101101110010100110001010000

$n = 5,\ k = 1$

1111101110011010110001010010000 0

$n = 6,\ k = 3$

1111111101111011111001110111110101110011110001101101011101001100111100101100011100001011011001011010100110100010010001110001010000110000001101101001100101100001001000100001010000001001000000

$n = 2, k = 4$

0000010110101111

$n = 2, k = 6$

000000010101101010111111

$n = 3, k = 6$

000000001001010010011011100100101101110110111111

Alfabeto $\{0, 1, 2\}$

$n = 2, k = 2$

222120121110020100

$n = 4, k = 2$

2222212220221222112210220222012200212120211221112110210210121002020122011201020022001200012121112101202120112001111101102110111001010021001100002020102000101010000

$n = 5,\ k = 1$

2222212222022211222102220122200221212212022111221102210122100220212202022011220102200122000212112121021201212002112021111211102110121100210202101121010021001210002020120200201112011020101201002001120010200001200001111101110011010110001010010000 0

$n = 6, k = 3$

222222221222220222212222211222210222202222201222200222122222121222120222112
221112221102221022221012221002220222220212220202220122201122201022200222200
122200022122122022122121122121102212022121102211222211212211202211122
211112211102211022211012211002210222102122102022101222101122101022100222100
122100022022022122022112202021022020222020122020022012222012122012022011222011122
201102201022201012201002200222200212200202200122200112200102200022200012200 0
021221221121221021220221220121220021212222121212121202121122121112121102121022
121012121002120222120212120202120122120112120102120022120012120002112112102 11
202211201211200211122211121211120211112111112111102111022111012111002110222 1
102121102021102121101121101021100221100121100021020202210201210200210122210 1
212101202101122101112101102101022101012101002100222100212100202100122100112 10
010210002210001210000202202201202200202122202121202120202112202111202110202 10
220210120210020202222020212020202020122020112020102020022020012020002012012002
011222011212011202011122011111201110201102201101201100201022201021201020201 01
220101120101020100220100120100020020012220012120012020011220011120011020010222
001012001002000222000212000202000122000112000102000022000012000001221221211222
120122112122111122110122102122101122100122022122021122020122012122011122010 12
200212200112200012112112012111212111112111012110212110112110012102212102112 10
201210121210111210101210021210011210001201201121201111201101201021201011201 0
012002212002112002012001212001112001012000212000112000011211211112110112102 1
121011121001120221120211120201120121120111120101120021120011120001111111101 11
102111101111100111022111021111020111012111011111010111002111001111000110110 10
211010111010011002211002111002011001211001111001011000211000111000010210210 11
021001020221020211020201020121020111020101020021020011020001011011001010221 01
021101020101012101011101010101002101001101000100100022100021100020100012100 01
110001010000210000110000002202202102202002201202201102201002200202200102200 00
210210200210120210110210100210020210010210000200200120200110200010200020200 0
102000001201201101201001200201200101200001101101010011002011001011000010010002 0
100010100000020020010020000010010000 00

$n = 2, k = 8$

0000000001010101020202021010101011111111121212122020202021212121222222222

$n = 4, k = 12$

000000000000001000100010002000200020010001000100011001100110012001200120020 0
020002000210021002100220022002201000100010001010101010101020102010201100110 01
100111011101110112011201120120012001200121012101210122012201220200020002000 20
102010201020202020202020210021002100211021102110212021202120220022002200220 10
022102220222022210001000100010011001100110021002100210101010101010111011101 11
012101210121020102010201021102110211022102210221100110011001101110110111011 0211
021102111011101110111111111111111121112111211201120112011211121112111221122112
212001200120012011201120112021202120212101210121012111211121112121212121220
122012201221122112211222122122200020002000200120012001200220022002201020102
010201120112011201220122012202020202020202120212021202220222022210021002100 21
012101210121022102210221102110211021112111211121112211221122121202120212021 2
121212122212221222200220022002201220122012202220222022102210221022112211221
122122212221222202220222022212221222122222222222222

Apéndice

A. COLLARES Y GRAFOS

A.1. Collares Perfectos y Grafos Astutos

Definición (Grafo de de Bruijn $B(n)$)**.** *El grafo de de Bruijn $B(n)$ es un grafo dirigido con conjunto de nodos Σ^n y hay una arista entre w y w' exactamente cuando los últimos $n-1$ símbolos de w coinciden con los primeros $n-1$ símbolos de w'.*

Estos grafos son eulerianos, porque están fuertemente conectados y cada nodo tiene el mismo grado de entrada y salida.

Los collares de de Bruijn corresponden a ciclos hamiltonianos en el grafo de de Bruijn. Dado que $B(n)$ es el grafo de líneas de $B(n-1)$, los collares de de Bruijn de orden n corresponden exactamente a ciclos eulerianos en $B(n-1)$. Por lo tanto, construir un collar de de Bruijn se puede hacer obteniendo un ciclo euleriano.

Definición (Grafo astuto $G(n,k)$)**.** *El grafo astuto $G(n,k)$ es un grafo dirigido con conjunto de vértices $\Sigma^n \times \{0,..,k-1\}$ y hay una arista entre (w,m) y (w',m') exactamente cuando los últimos $n-1$ símbolos de w coinciden con los primeros $n-1$ símbolos de w' y $m' = m+1 \mod k$.*

El grafo de de Bruijn de orden n es exactamente el grafo astuto $G(n,1)$.

Como ocurre en el caso de de Bruijn, el grafo de líneas del grafo astuto $G(n,k)$ es $G(n+1,k)$. Por lo tanto, para construir collares (n,k)-perfectos se construyen ciclos eulerianos en el grafo astuto $G(n-1,k)$. Debido a la posibilidad de períodos en un collar (n,k)-perfecto, muchos ciclos eulerianos pueden generar el mismo collar (n,k)-perfecto. El número de collares (n,k)-perfectos se presenta en [7, Teorema 2].

Los collares (n,k)-perfectos corresponden a ciclos eulerianos en el llamado grafo *astuto*

Los collares (n,k)-perfectos corresponden a ciclos hamiltonianos en el llamado grafo $G(n,k)$ *astuto*, ver [7]. Estos grafos son el producto tensorial del grafo de de Bruijn de orden $n-1$ y un ciclo simple de longitud k.

Ejemplo. *El grafo $G(2,2)$ para el alfabeto $\Sigma = \{0,1\}$ tiene como conjunto de nodos V a $\Sigma^2 \times \{0,1\}$, que equivale a:*

$$\{\langle 11,0\rangle, \langle 11,1\rangle, \langle 10,0\rangle, \langle 10,1\rangle, \langle 01,0\rangle, \langle 01,1\rangle, \langle 00,0\rangle, \langle 00,1\rangle\}$$

Y el conjunto de aristas es:

$$\{(\langle u,m\rangle, \langle v, m+1 \mod k\rangle) : \textit{donde } u[1.,3] = v[0.,2]\}$$

que equivale a:

$$\{(\langle 11,0\rangle, \langle 11,1\rangle), (\langle 11,1\rangle, \langle 10,0\rangle), (\langle 11,1\rangle, \langle 11,0\rangle), (\langle 11,0\rangle, \langle 10,1\rangle),$$
$$(\langle 10,0\rangle, \langle 01,1\rangle), (\langle 10,1\rangle, \langle 00,0\rangle), (\langle 10,1\rangle, \langle 01,0\rangle), (\langle 10,0\rangle, \langle 00,1\rangle),$$
$$(\langle 01,0\rangle, \langle 11,1\rangle), (\langle 01,1\rangle, \langle 10,0\rangle), (\langle 01,1\rangle, \langle 11,0\rangle), (\langle 01,0\rangle, \langle 10,1\rangle),$$
$$(\langle 00,0\rangle, \langle 01,1\rangle), (\langle 00,1\rangle, \langle 00,0\rangle), (\langle 00,1\rangle, \langle 01,0\rangle), (\langle 00,0\rangle, \langle 00,1\rangle)\}$$

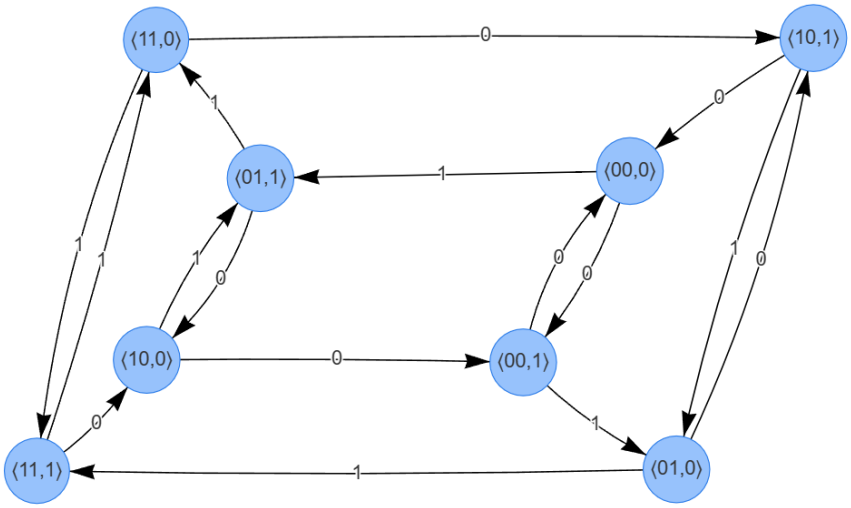


Fig. A.1: Grafo astuto $G(2,2)$

Y su representación visual es la Figura A.1, donde las aristas tienen como etiqueta el símbolo que se concatena al final de la palabra actual y que da lugar a la del nodo que se transiciona.

El grafo $G(3,2)$ para el alfabeto $\Sigma = \{0,1\}$ tiene como conjunto de nodos V a $\Sigma^3 \times \{0,1\}$, que equivale a:

$$\{\langle 111,0\rangle, \langle 111,1\rangle, \langle 110,0\rangle, \langle 110,1\rangle, \langle 101,0\rangle, \langle 101,1\rangle, \langle 100,0\rangle, \langle 100,1\rangle,$$
$$\langle 011,0\rangle, \langle 011,1\rangle, \langle 010,0\rangle, \langle 010,1\rangle, \langle 001,0\rangle, \langle 001,1\rangle, \langle 000,0\rangle, \langle 000,1\rangle\}$$

Y el conjunto de aristas es:

$$\{(\langle u,m\rangle, \langle v, m+1 \mod k\rangle) : \text{donde } u[1.,3] = v[0.,2]\},$$

que equivale a:

$$\{(\langle 111,0\rangle, \langle 111,1\rangle), (\langle 111,0\rangle, \langle 110,1\rangle), (\langle 111,1\rangle, \langle 111,0\rangle), (\langle 111,1\rangle, \langle 110,0\rangle),$$
$$(\langle 110,0\rangle, \langle 101,1\rangle), (\langle 110,0\rangle, \langle 100,1\rangle), (\langle 110,1\rangle, \langle 101,0\rangle), (\langle 110,1\rangle, \langle 100,0\rangle),$$
$$(\langle 101,0\rangle, \langle 011,1\rangle), (\langle 101,0\rangle, \langle 010,1\rangle), (\langle 101,1\rangle, \langle 011,0\rangle), (\langle 101,1\rangle, \langle 010,0\rangle),$$
$$(\langle 100,0\rangle, \langle 001,1\rangle), (\langle 100,0\rangle, \langle 000,1\rangle), (\langle 100,1\rangle, \langle 001,0\rangle), (\langle 100,1\rangle, \langle 000,0\rangle),$$
$$(\langle 011,0\rangle, \langle 111,1\rangle), (\langle 011,0\rangle, \langle 110,1\rangle), (\langle 011,1\rangle, \langle 111,0\rangle), (\langle 011,1\rangle, \langle 110,0\rangle),$$
$$(\langle 010,0\rangle, \langle 101,1\rangle), (\langle 010,0\rangle, \langle 100,1\rangle), (\langle 010,1\rangle, \langle 101,0\rangle), (\langle 010,1\rangle, \langle 100,0\rangle),$$
$$(\langle 001,0\rangle, \langle 011,1\rangle), (\langle 001,0\rangle, \langle 010,1\rangle), (\langle 001,1\rangle, \langle 011,0\rangle), (\langle 001,1\rangle, \langle 010,0\rangle),$$
$$(\langle 000,0\rangle, \langle 001,1\rangle), (\langle 000,0\rangle, \langle 000,1\rangle), (\langle 000,1\rangle, \langle 001,0\rangle), (\langle 000,1\rangle, \langle 000,0\rangle)\}$$

Y su representación visual es la Figura A.2:

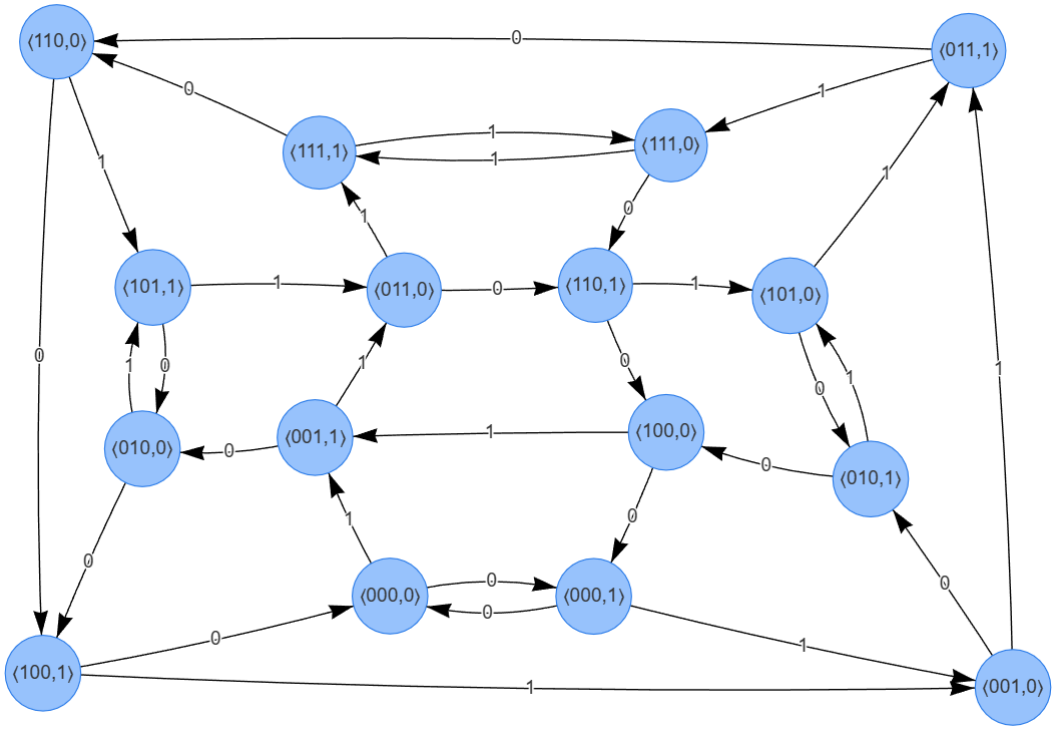


Fig. A.2: Grafo astuto $G(3,2)$

A.2. Palabras de Lyndon y su relación con grafos

Definición (Palabra de Lyndon)**.** *Sea n entero positivo. Una palabra A con $n = |A|$ se dice de Lyndon, si todas las demás rotaciones de esta son estrictamente menores.*

Ejemplo. *Dado $n = 3$ y $s = 2$, la palabra $A = 101$ es de Lyndon, porque sus rotaciones son:*

$$101, 011, 110$$

Y de estas 101 *es mayor estricto lexicográficamente que el resto.*

Definición (Par de Lyndon)**.** *Sean n y k enteros positivos, tal que k divide a n. Un par $\mathcal{A}$ se dice de Lyndon, si todas las demás rotaciones de este son estrictamente menores.*

Ejemplo. *Siguiendo el ejemplo anterior, con $n = 3$, $k = 2$ y $s = 2$, el par $\mathcal{A} = \langle 101, 0\rangle$ es de Lyndon, porque sus rotaciones son:*

$$\langle 101, 0\rangle, \langle 011, 1\rangle, \langle 110, 0\rangle, \langle 101, 1\rangle, \langle 011, 0\rangle, \langle 110, 1\rangle,$$

Y de estas $\langle 101, 0\rangle$ es mayor estricto lexicográficamente que el resto.

Alternativamente, podemos definir un par de Lyndon dado $\mathcal{A} = \langle A, 0\rangle$ collar con $|A| = n$, como $\mathcal{B} = \overline{\mathcal{A}}$. Esto satisface las condiciones para ser de Lyndon por la Observación 4.

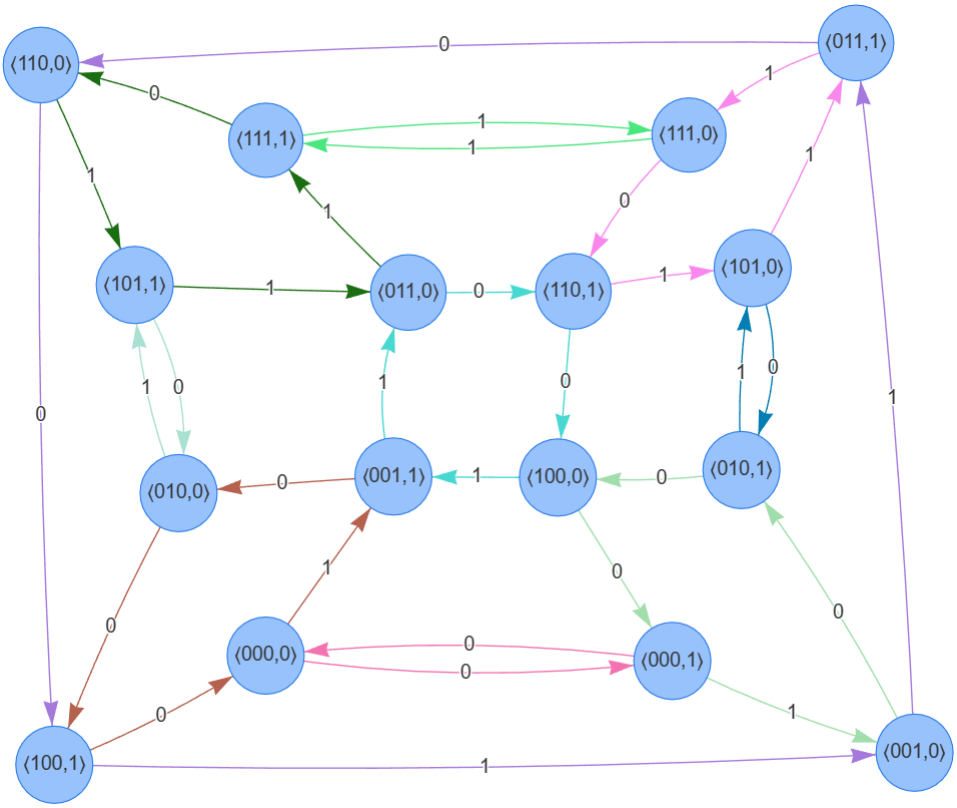


Fig. A.3: Grafo astuto $G(3,2)$ con los ciclos resaltados por cada par de Lyndon.

Esta definicion alternativa de los pares de Lyndon es util para ver como se interpretan en el grafo astuto $G(n-1,k)$.

Observación 5 (Pares de Lyndon como ciclos)**.** *Sean n y k enteros positivos, tal que k divide a n. Para cada collar $\mathcal{A}$ tal que $|A| = n$, el par de Lyndon que surge de este es un ciclo en el grafo astuto $G(n-1,k)$. Las aristas $(\mathcal{B},\mathcal{C})$ que forman parte del ciclo se definen a través de las rotaciones de $\mathcal{A} = \langle A, 0\rangle$, con:*

$$\mathcal{B} = \langle a_{r+1}\cdots a_n a_1 \cdots a_r, 0 \oplus r\rangle_{n-1}$$
$$\mathcal{C} = \langle a_{r+2}\cdots a_n a_1 \cdots a_{r+1}, 0 \oplus (r+1)\rangle_{n-1},$$

donde $0 \leq r < |\overline{\mathcal{A}}|$. Es decir,las aristas $(\mathcal{B},\mathcal{C})$ son las aristas que representan las transiciones entre dos pares $\mathcal{B} = \langle B, 0\rangle$ y $\mathcal{C} = \langle C, 0\rangle$ consecutivos en $\mathcal{A}$, cuyas palabras B y C tienen longitud $|B| = |C| = n-1$.

Ejemplo. *Dado $n = 4$, $k = 2$ y $s = 2$, para el par de Lyndon $\langle 1100, 0\rangle$, las aristas (f,t) que forman parte del ciclo son:*

$$(\langle 110,0\rangle, \langle 100,1\rangle), (\langle 100,1\rangle, \langle 001,0\rangle), (\langle 001,0\rangle, \langle 011,1\rangle), (\langle 011,1\rangle, \langle 110,0\rangle).$$

Aún más, no existen dos pares de Lyndon tal que sus ciclos comparten una arista, ya que eso implicaría que tienen una rotación en común. Básicamente no puede pasar porque el par de Lyndon provino de $\mathcal{A}$, que es collar y no comparte rotaciones con otro collar $\mathcal{B}$.

Ejemplo. *Dado $n = 4$, $k = 2$ y $s = 2$, estos son todos los posibles pares de Lyndon:*

$$\langle 11, 0\rangle, \langle 1110, 0\rangle, \langle 1101, 0\rangle, \langle 1100, 0\rangle, \langle 10, 0\rangle,$$
$$\langle 1001, 0\rangle, \langle 1000, 0\rangle, \langle 01, 0\rangle, \langle 0100, 0\rangle, \langle 00, 0\rangle$$

La Figura A.3 muestra la interpretación en el grafo astuto $G(3, 2)$ donde cada ciclo derivado de un par de Lyndon en base a lo dicho anteriormente tiene un color que lo distingue del resto.

El Algoritmo 1 construye un collar (n, k)-perfecto concatenando todas los pares de Lyndon en orden decreciente para ese alfabeto, n y k.

B. ALGORITMO *PREFER ONE*

Transcribimos aqui el algortimo de Fredericksen llamado Prefer One [5, Algorithm 2].

Input: número entero positivo n
Output: Collar de Bruijn de orden n lexicográficamente máximo.

1. Comenzamos con una palabra de n ceros.
2. Si la palabra tiene longitud 2^n, frenamos; de lo contrario,
3. Para el i-ésimo bit de la palabra, si $i > n$, escribimos un 1 si la palabra de tamaño n no ha aparecido antes en la palabra; de lo contrario, escribimos un cero. Finalmente aumentamos i y repetimos desde el paso 2;

Fredericksen [3] muestra que Prefer One es un algoritmo goloso que genera un collar de De Bruijn de orden n con alfabeto $\mathcal{A} = \{0, 1\}$ que tiene una complejidad temporal lineal con respecto al tamaño de la salida.

El siguiente algoritmo recorre un *grafo de De Bruijn* $B(n)$ formando un circuito euleriano.

1. Buscamos un árbol enraizado en 0^{n-1} en $B(n-1)$
2. Agregamos el bucle en el nodo 0^{n-1} a cualquier árbol enraizado en 0^{n-1} en $B(n-1)$, y colocamos una estrella en él.
3. Colocamos una estrella en todas las aristas del árbol.
4. Seleccionamos el nodo 0^{n-1} como el nodo inicial
5. Cuando estemos en el nodo i, saldremos de este por la arista no estrellada a menos que haya sido usado. De lo contrario, saldremos por la arista estrellada,
6. Podemos determinar el estado del proceso por la etiqueta de la arista recorrida.
7. Cuando llegamos a un nodo, si la arista estrellada que sale de este ha sido utilizada, frenamos. De lo contrario, volvemos al paso 5.

Fredricksen afirma y prueba que al usar el árbol enraizado en 0^{n-1} formado por las aristas con etiqueta 0, la palabra obtenida en el circuito euleriano es la misma que obtenemos con el algoritmo *Prefer One*.

Ejemplo (collares de De Bruijn)**.**
$n = 1 \to [01]$
$n = 2 \to [0011]$
$n = 3 \to [00011101]$

En cada iteración del bucle el algoritmo realiza un número constante de operaciones, y cada decisión es local con respecto al nodo que se visita.

Para tomar una decisión necesitamos verificar si la nueva palabra de tamaño n ya se ha escrito en la salida. Así que mantenemos un conjunto en el que, si la palabra existe, entonces se escribió en la salida, de lo contrario, no lo fue. En cada iteración del bucle:

1. el algoritmo hace una búsqueda en el conjunto y elige 1 o 0.
2. escribe la decisión en la salida.
3. escribe el conjunto con la palabraformada por los últimos n símbolos.

Sabemos que hay 2^n palabras posibles de longitud n con 2 letras, por lo que habrá a lo sumo 2^n iteraciones de bucle. Por lo tanto, el algoritmo tiene una complejidad temporal lineal con respecto al tamaño de la salida.

REFERENCIAS

[1] Nicolaas G. de Bruijn. «A combinatorial problem». En: *Nederl. Akad. Wetensch., Proc.* 49 (1946), 758-764 = Indagationes Math. 8, 461-467 (1946).

[2] Jr. L. R. Ford. «A cyclic arrangement of m-tuples». En: *report p-1071, Rand Corporation, Santa Monica, California* (1957).

[3] Harold Fredricksen. «The lexicographically least de Bruijn cycle». En: *Journal of Combinatorial Theory* 9.1 (1970), págs. 1-5.

[4] Harold Fredricksen y James Maiorana. «Necklaces of beads in k colors and k-ary de Bruijn sequences». En: *Discrete Mathematics* 23.3 (1978), págs. 207-210.

[5] Harold Fredricksen. «A Survey of Full Length Nonlinear Shift Register Cycle Algorithms». En: *SIAM Review* 24.2 (1982), págs. 195-221.

[6] Jean Berstel y Dominique Perrin. «The origins of combinatorics on words». En: *European Journal of Combinatorics* 28.3 (2007), págs. 996-1022.

[7] Nicolás Álvarez, Verónica Becher, Pablo Ferrari y Sergio Yuhjtman. «Perfect necklaces». En: *Advances in Applied Mathematics* 80 (2016), págs. 48-61.

[8] Ezequiel Zimenspitz. *Collares perfectos máximos*. Tesis de Licenciatura en Ciencias de la Computación,Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. 2020.