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

Secuencias de De Bruijn con discrepancia mínima

Tesis de Licenciatura en Ciencias de la Computación

Ivo Pajor

Directora: Verónica Becher

Buenos Aires, 2024

SECUENCIAS DE DE BRUIJN CON DISCREPANCIA MÍNIMA

La discrepancia de una cadena binaria es la máxima diferencia (absoluta) entre el número de unos y el número de ceros sobre todas las subcadenas posibles de la cadena binaria inicial. En este trabajo determinamos la discrepancia mínima que puede lograr una secuencia binaria de de Bruijn de orden n, la cual es n. Este era un problema abierto hasta ahora. Damos un algoritmo que construye una secuencia binaria de de Bruijn con una discrepancia mínima. Una ligera modificación de este algoritmo lidia con alfabetos arbitrarios y produce secuencias de de Bruijn de orden n con una discrepancia de como máximo 1 por encima del límite inferior trivial n.

Palabras claves: secuencias de de Bruijn, discrepancia

DE BRUIJN SEQUENCES WITH MINIMUM DISCREPANCY

The discrepancy of a binary string is the maximum (absolute) difference between the number of ones and the number of zeroes over all possible substrings of the given binary string. In this work we determine the minimal discrepancy that a binary de Bruijn sequence of order n can achieve, which is n. This was an open problem until now. We give an algorithm that constructs a binary de Bruijn sequence with minimal discrepancy. A slight modification of this algorithm deals with arbitrary alphabets and yields de Bruijn sequences of order n with discrepancy at most 1 above the trivial lower bound n.

Keywords: de Bruijn sequence, discrepancy.

Índice general

1. INTRODUCCIÓN

Las secuencias de De Bruijn, llamadas así en honor al matemático neerlandés Nicolaas Govert De Bruijn, son secuencias circulares de símbolos que tienen la propiedad de contener todas las posibles secuencias de un cierto tamaño como subsecuencias continuas [**?**]. Estas secuencias tienen aplicaciones en diversas áreas como biología, informática y teoría de la información.

Debido a esto, hay numerosos trabajos sobre cómo construirlas. En general, estos se enfocan en la complejidad algorítmica y en garantizar propiedades específicas para una determinada aplicación. Entre estos, se destaca el algoritmo propuesto por Huang [**?**]. En dicho trabajo, Huang propone un algoritmo para generar secuencias binarias de De Bruijn uniendo los ciclos producidos por el *complemented cycling register* (CCR) definido por Golomb en [**?**]. El uso de los CCR confiere, según Huang, un buen balance local entre ceros y unos.

Gabric y Sawada en [**?**] formalizan la definición de balance de Huang. Además, demuestran que el algoritmo de Huang construye secuencias con discrepancia asintóticamente mínima.

Al final de su trabajo Gabric y Sawada dejan dos problemas abiertas:

- la construcción hecha por Huang, ¿tiene la menor discrepancia entre todas las secuencias de De Bruijn?
- generalizar la investigación de la discrepancia en las secuencias de De Bruijn de alfabeto arbitrario.

En el presente trabajo vamos a responder a la primera pregunta de manera negativa y presentar una construcción computacionalmente eficiente que sí alcanza la mínima discrepancia entre todas las secuencias binarias de De Bruijn. Además, dicha construcción se puede generalizar a alfabetos arbitrarios con una discrepancia casi óptima. Demostramos el siguiente teorema:

Teorema 1. *Existe un algoritmo que produce una secuencia De Bruijn de orden n con discrepancia n en caso de que el alfabeto tenga dos símbolos, y con discrepancia a lo sumo $n+1$ en caso de que el alfabeto tenga más de dos símbolos. El algoritmo utiliza $O(n)$ memoria y genera cada símbolo en $O(n)$ tiempo.*

2. NOTACIÓN

A continuación presentamos la notación que sera utilizada en el trabajo. Dado Σ un alfabeto:

- Llamamos cadenas a las secuencias finitas de símbolos de Σ.
- Denotamos Σ^n al conjunto de cadenas de longitud n.
- Utilizamos una numeración *0-based*.
- Para denotar el i-ésimo símbolo de la cadena s usamos $s[i]$.
- Si s es una cadena, llamamos $|s|$ a su longitud.

3. MARCO TEÓRICO

3.1. Secuencias de De Bruijn

Una secuencia de De Bruijn de orden n sobre un alfabeto de tamaño b es una cadena circular de longitud b^n tal que cada cadena de longitud n aparece exactamente una vez en ella como subsecuencia continua.

Ejemplo:

- Para $b = 2$ y $n = 2$ una posible secuencia de De Bruijn es 0011, ya que:
  - 00 aparece de la posición 1 a la 2
  - 01 aparece de la posición 2 a la 3.
  - 11 aparece de la posición 3 a la 4.
  - 10 aparece de la posición 4 a la 1, recordando que la secuencia es circular.

3.2. Discrepancia

La discrepancia de una cadena es un número entero no negativo que indica la diferencia máxima, en cualquier subcadena, entre el número de apariciones del símbolo que más aparece y el símbolo que menos aparece en esa subcadena.

Formalmente: sea Σ un alfabeto y w una cadena sobre Σ. Denotemos $|w|_a$ al número de apariciones del símbolo a en w. Sea $S(w)$ el conjunto de todas las subcadenas de w donde interpretamos w como una cadena circular. La discrepancia de una cadena w sobre el alfabeto Σ se define como

$$\text{discrepancia}(w) = \max_{s \in S(w)} \left(\max_{a \in \Sigma} |s|_a - \min_{c \in \Sigma} |s|_c \right)$$

Ejemplo

- discrepancia($\underline{2012}2101\underline{22}$) es 4 si tomamos $\Sigma = \{0, 1, 2\}$, la subcadena subrayada es la que logra dicho valor.

Notemos que en una secuencia de De Bruijn de orden n la discrepancia es al menos n ya que la subcadena $\underbrace{000\ldots0}_{\text{n ceros}}$ tiene que aparecer en la secuencia.

Esta definición es una versión generalizada de la provista en [**?**] para alfabetos arbitrarios.

3.2.1. Análisis de discrepancia

Existen unos pocos trabajos sobre la discrepancia de las secuencias de De Bruijn. La discrepancia máxima y mínima asintótica alcanzable por una secuencia de De Bruijn de orden n son $\Theta(2^n/\sqrt{n})$ y $\Theta(n)$ respectivamente. Esto fue demostrado por Gabric y Sawada en [**?**, Teorema 5.3] y [**?**, Teorema 2.2].

¿Cuál es la discrepancia mínima que puede lograr una secuencia de De Bruijn? Claramente, en cada alfabeto, la discrepancia de una secuencia de De Bruijn de orden n es al menos n, porque debe existir una racha de al menos n símbolos iguales consecutivos.

3.3. Construcción de secuencias de De Bruijn

3.3.1. Grafo de De Bruijn

Dado un número entero n y un alfabeto $\Sigma = \{0, 1, \ldots, b-1\}$, donde b es un entero positivo, el grafo de De Bruijn $B_n = (V_n, A_n)$ es un grafo con un conjunto de nodos $V_n = \Sigma^n$, el conjunto de todas las cadenas de longitud n, y un arco $(s, t) \in A_n$ si y solo si el sufijo de s de longitud $n-1$ es igual al prefijo de t de longitud $n-1$. Un ejemplo de un grafo de De Bruijn se puede ver en la Figura 3.1.

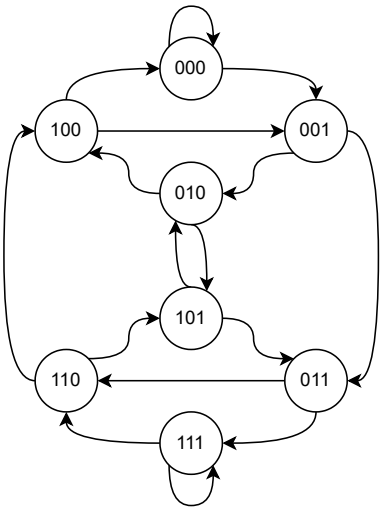

Fig. 3.1: Grafo de De Bruijn para $n = 3$ y $\Sigma = \{0, 1\}$

3.3.2. Ciclo Hamiltoniano

Sea $G = (V, E)$ una grafo dirigido, decimos que G que es *hamiltoniano* si existe un ciclo simple en nodos que recorre todos los nodos de G. A dicho ciclo lo llamaremos *ciclo hamiltoniano*. Un ejemplo de un grafo hamiltoniano se puede ver en la Figura 3.2. Un resultado que nos va a ser útil es que todos los grafos de De Bruijn son hamiltonianos.

3.3.3. Equivalencia de ciclos hamiltonianos con secuencias de De Bruijn

Los ciclos hamiltonianos en el grafo de De Bruijn B_n corresponden a secuencias de De Bruijn de orden n: si $s_0, s_1, \ldots, s_{k-1}$ es un ciclo hamiltoniano en B_n, entonces la secuencia

$$s_0[0], s_1[0], \ldots, s_{k-1}[0]$$

que consiste del primer símbolo de cada cadena, es una secuencia de De Bruijn.

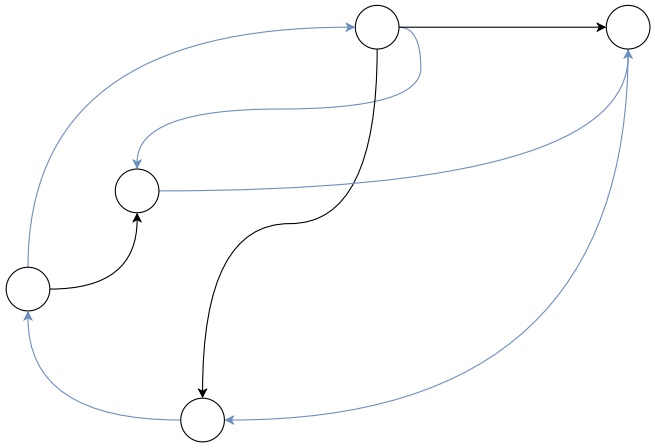

Fig. 3.2: Ejemplo de un grafo hamiltoniano. Se muestra en azul las aristas pertenecientes al ciclo.

Por ejemplo, tomemos el grafo de De Bruijn de la Figura 3.1, un ciclo hamiltoniano posible en este grafo es el siguiente:

$$000 \to 001 \to 011 \to 111 \to 110 \to 101 \to 010 \to 100$$

Luego, si tomamos el primer carácter de cada cadena y los concatenamos obtenemos:

$$00011101$$

que es un secuencia de De Bruijn.

4. DESARROLLO: DEMOSTRACIÓN DEL TEOREMA PRINCIPAL

4.1. Estructura de la demostración

La estructura del algoritmo y su prueba de corrección es la siguiente:

- Definimos un subgrafo del grafo de De Bruijn de orden n, al que llamamos subgrafo válido, donde todos los arcos tienen una determinada forma y cada camino en ese grafo corresponde a una cadena con discrepancia acotada.
- Descomponemos el grafo de De Bruijn en ciclos disjuntos en nodos usando la regla de CCR, como lo hace Huang [**?**]. Luego, probamos que todos estos ciclos están contenidos dentro del subgrafo válido.
- Elegimos una forma de conectar estos ciclos en una estructura similar a un árbol para formar una ciclo hamiltoniano, de modo que los arcos entre los ciclos también estén contenidos en el subgrafo válido.
- Finalmente, desarrollamos un algoritmo para recorrer este árbol, recorriendo completamente todos los ciclos y produciendo así un ciclo hamiltoniano. Dado que está contenido dentro del subgrafo válido, esto garantiza que el ciclo hamiltoniano resultante tenga una discrepancia acotada.

Antes de avanzar, necesitamos algunas definiciones.

Definición 1 (Diferencia)**.** *La discrepancia de una cadena w se define como, discrepancia* : $\Sigma^* \to \mathbb{Z}$*,*

$$\textit{discrepancia}(w) = \max_{s \in S(w)} \left(\max_{a \in \Sigma} |s|_a - \min_{c \in \Sigma} |s|_c \right).$$

donde el valor $\max_{a \in \Sigma} |s|_a - \min_{c \in \Sigma} |s|_c$ *es la* diferencia *de la subcadena s, denotado $D(s)$.*

Para acotar la discrepancia de las secuencias de De Bruijn, necesitamos acotar la diferencia de todas sus subcadenas. Una subcadena de una secuencia de De Bruijn es una secuencia de la forma

$$s_i[0], s_{i+1}[0], \ldots, s_j[0]$$

donde $s_i, s_{i+1}, \ldots, s_j$ es una subsecuencia contigua del ciclo hamiltoniano. Por lo tanto, las subcadenas de una cadena de De Bruijn corresponden a caminos en el grafo de De Bruijn.

4.2. Subgrafo válido

4.2.1. Histogramas

Definimos un concepto que nos ayudará a calcular la diferencia de una cadena.

Definición 2 (Histograma)**.** *Dada una cadena $s \in \Sigma^*$, definamos $H(s) : \Sigma \to \mathbb{Z}$ como $H(s)(c) = |s|_c$.*

La diferencia de un histograma se define de la siguiente manera.

Definición 3 (Diferencia de un histograma)**.** *La diferencia $D(H)$ de un histograma $H : \Sigma \to \mathbb{Z}$ está dada por*

$$D(H) = \max_{i,j \in \Sigma} H(i) - H(j).$$

Como era de esperar, $\mathrm{D}(\mathrm{H}(s)) = \mathrm{D}(s)$. Usamos el mismo símbolo para indicar la diferencia de una cadena y de un histograma. Sin embargo, el significado del símbolo quedará claro a partir del contexto en el que se utiliza. Otra operación que necesitamos para los histogramas es la *suma parcial*.

Definición 4 (Suma parcial de un histograma)**.** *Dado $H : \Sigma \to \mathbb{Z}$, definimos su suma parcial $P(H)$ como*

$$P(H)(i) = \sum_{j=0}^{i} H(j).$$

Notemos que P es lineal, es decir, sean dos histogramas $H_1, H_2 : \Sigma \to \mathbb{Z}$ entonces:

$$\mathrm{P}(H_1 + H_2) = \mathrm{P}(H_1) + \mathrm{P}(H_2)$$

Sea K el histograma constante. x—Es decir, $K : \Sigma \to \mathbb{Z}$ tal que $K(b) = 1$ para todo $b \in \Sigma$. Observemos que si sumamos K a un histograma, la diferencia no se ve afectada. Por tanto, podemos considerar los histogramas en el espacio cociente.

$$\frac{\Sigma \to \mathbb{Z}}{\langle K \rangle}$$

4.2.2. Regla del registro de ciclo incremental

Definición 5 (Incremented Cycle Register Rule)**.** *Dado un entero positivo k, definimos $ICR_k : \Sigma^n \to \Sigma^n$ como*

$$ICR_k(s) = s[1]s[2] \dots s[n-1](s[0] + k).$$

Entonces, $ICR_k(s)$ es la cadena que consiste en una rotación cíclica de s, pero donde el último símbolo se incrementa en k. A menos que se especifique lo contrario, $ICR = ICR_1$.

Consideremos el subgrafo de B_n dado por $(\Sigma^n, \{(s, ICR(s)) : s \in \Sigma^n\})$ que llamamos subgrafo ICR de B_n. Debido a que ICR es una función biyectiva, este subgrafo consiste de una colección de ciclos disjuntos en nodos.

Consideremos un camino $s_0, s_1, \dots, s_k$ en el subgrafo ICR. ¿Podemos reconstruir el histograma de los primeros símbolos de estas cadenas sabiendo solo s_0 y s_k?

Lema 1. *Sea $s \in \Sigma^n$ y sea $b = s[0]$ el primer símbolo de s, entonces*

$$P(H(s)) - P(H(ICR(s))) \equiv e_b \mod K,$$

donde e_b la función indicadora en b.

Demostración. Observemos que, por definición, $\mathrm{H}(s) - \mathrm{H}(\mathrm{ICR}(s)) = e_b - e_{b+1}$. Si $b+1 \neq 0$, entonces $\mathrm{P}(e_b - e_{b+1}) = e_b$ y por linealidad de P hemos terminado. Si $b + 1 = 0$, entonces $\mathrm{P}(e_b - e_{b+1}) = e_b - K$. □

Lema 2. *Sea $s_0, \ldots, s_k$ una secuencia de cadenas tal que $s_{i+1} = ICR(s_i)$ para todo i. Entonces se cumple la siguiente igualdad*

$$P(H(s_0)) - P(H(s_k)) \equiv \sum_{i=0}^{k-1} e_{s_i[0]} \mod K.$$

Demostración. Aplicar Lema 1 y usar suma telescópica. □

4.2.3. Subgrafo válido

Dado que el subgrafo ICR de B_n no es suficiente para construir un ciclo hamiltoniano, introducimos otro grado de libertad. Consideremos que $d : \Sigma^n \to \Sigma$ es una asignación de profundidad, y para cada $s \in \Sigma^n$, d_s es su profundidad.

Definición 6 (Subgrafo válido)**.** *Dada una asignación de profundidad $d : \Sigma^n \to \Sigma$, un arco (s, t) en el grafo de De Bruijn B_n se llama* válido *si, para $b = s[0]$ y $c = t[n-1]$,*

- $b + 1 = c$ *y* $d_s = d_t$, *o*
- $b + 1 = d_t$ *y* $c = d_s$

El conjunto de arcos válidos en B_n forman el subgrafo válido.

Lema 3. *Dada una asignación de profundidad $d : \Sigma^n \to \Sigma$, y $s, t \in \Sigma^n$ tal que (s, t) es un arco válido en el grafo de De Bruijn, entonces:*

$$e_{s[0]} \equiv P(H(s) + e_{d_s}) - P(H(t) + e_{d_t}) \mod K.$$

Demostración. Sea $b = s[0]$ y $c = t[n-1]$. Entonces tenemos

$$\begin{aligned}
e_b &\equiv \mathrm{P}(\mathrm{H}(s) + e_{d_s}) - \mathrm{P}(\mathrm{H}(t) + e_{d_t}) && \mod K \Leftrightarrow \\
e_b &\equiv \mathrm{P}(\mathrm{H}(s) - \mathrm{H}(t)) + \mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t}) && \mod K \Leftrightarrow \\
e_b &\equiv \mathrm{P}(e_b - e_c) + \mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t}) && \mod K \Leftrightarrow \\
e_b &\equiv \mathrm{P}(e_b) - \mathrm{P}(e_c) + \mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t}) && \mod K \Leftrightarrow \\
\mathrm{P}(e_b) - \mathrm{P}(e_{b+1}) &\equiv \mathrm{P}(e_b) - \mathrm{P}(e_c) + \mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t}) && \mod K \Leftrightarrow \\
0 &\equiv \mathrm{P}(e_{b+1}) - \mathrm{P}(e_c) + \mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t}) && \mod K
\end{aligned}$$

Los casos en los que esto se cumple son

- $b+1 = c$ y $d_s = d_t$, que es el caso donde las profundidades son iguales y $t = ICR(s)$
- $b + 1 = d_t$ y $c = d_s$.

□

Notemos que si la asignación de profundidad es constante en los ciclos producidos por la regla ICR, entonces el subgrafo ICR es un subgrafo del subgrafo válido.

Además, si tenemos algún camino en el subgrafo válido, entonces se cumple una igualdad telescópica.

Lema 4. *Sea $d : \Sigma^n \to \Sigma$ una asignación de profundidad y sea $s_0, \ldots, s_k$ un camino en el subgrafo válido, entonces*

$$\sum_{i=0}^{k-1} e_{s_i[0]} \equiv P\left(H(s_0) + e_{d_{s_0}}\right) - P\left(H(s_k) + e_{d_{s_k}}\right) \quad \text{mód } K.$$

El histograma de la secuencia de los primeros símbolos de un camino en el subgrafo válido de B_n depende sólo de la cadena inicial y final. Observemos que en la fórmula del Lema 4, el histograma ignora el primer símbolo de la última cadena (la suma solo llega hasta $k-1$). Esto no es un problema para nuestra aplicación: supongamos que, para cada par de cadenas $s, t \in \Sigma^n$ logramos demostrar una cota

$$\mathrm{D}\left(\mathrm{P}\left(\mathrm{H}(s) + e_{d_s}\right) - \mathrm{P}\left(\mathrm{H}(t) + e_{d_t}\right)\right) \leq B \qquad B \in \mathbb{N}$$

Entonces, cada secuencia de De Bruijn que surge de un ciclo hamiltoniano $s_0, \ldots, s_k$ en el subgrafo válido, tiene una discrepancia acotada por B. De hecho, si consideramos la subcadena $s_i[0], s_{i+1}[0] \ldots, s_j[0]$, entonces podemos acotar

$$\mathrm{D}(\mathrm{H}(s_i[0], s_{i+1}[0] \ldots, s_j[0]))$$

aplicando Lema 4 en el camino $s_i, s_{i+1}, \ldots, s_j, s_{j+1}$ con un elemento adicional después del final.

4.2.4. Cotas de diferencia para caminos en el grafo válido

Empecemos ignorando las profundidades.

Lema 5 (Diferencia de sumas parciales de histogramas)**.** *Si $s, t \in \Sigma^n$ entonces*

$$D(P(H(s) - H(t))) \leq n.$$

Demostración. El histograma $\mathrm{H}(s)$ se puede escribir como una suma de todos los símbolos de s,

$$\mathrm{H}(s) = \sum_{i=0}^{n-1} e_{s[i]}.$$

Similarmente,

$$\mathrm{H}(t) = \sum_{i=0}^{n-1} e_{t[i]}.$$

Por lo tanto,

$$\mathrm{P}(\mathrm{H}(s) - \mathrm{H}(t)) = \sum_{i=0}^{n-1} \mathrm{P}(e_{s[i]} - e_{t[i]}).$$

Observemos que $\mathrm{P}(e_{s[i]} - e_{t[i]})$ tiene una diferencia de a lo sumo 1 ya que sus valores son todos 0s y 1s, o todos 0s y -1s. Por lo tanto, debido a la subaditividad obtenemos la cota deseada. □

Lema 6 (Diferencia de sumas parciales de histogramas con símbolo compartido)**.** *Si $s, t \in \Sigma^n$ y existe un símbolo que aparece en ambas cadenas, entonces*

$$D(P(H(s) - H(t))) \leq n - 1.$$

Demostración. Como en Lema 5 transformar la fórmula

$$\mathrm{P}(\mathrm{H}(s) - \mathrm{H}(t)) = \sum_{i=0}^{n-1} \mathrm{P}(e_{s[i]} - e_{t[i]}).$$

Dado que los histogramas no dependen del orden de los símbolos, podemos asumir sin pérdida de generalidad que $s[0] = t[0]$, y por lo tanto el primer sumando de la suma desaparece. Como cada uno de los $n-1$ restantes puede contribuir a lo sumo 1 a la diferencia, obtenemos el resultado. □

Ahora consideremos la diferencia de sumas parciales pero incluyendo profundidades.

Lema 7. *Si $s, t \in \Sigma^n$ son cadenas y $d : \Sigma^n \to \Sigma$ es una asignación de profundidad, entonces*

$$D\left(P\left(H(s) + e_{d_s}\right) - P\left(H(t) + e_{d_t}\right)\right) \leq n + 1.$$

Demostración. La fórmula se desprende del Lema 5 usando la subaditividad de D, la linealidad de P y el hecho de que $\mathrm{D}(\mathrm{P}(e_{d_s} - e_{d_t})) \leq 1$. □

Lema 8. *Si $s, t \in \Sigma^n$ son cadenas y $d : \Sigma^n \to \Sigma$ es una asignación de profundidad, tal que*

- *$d_s = d_t$, o*
- *s y t comparten al menos un símbolo*

Entonces

$$D\left(P\left(H(s) + e_{d_s}\right) - P\left(H(t) + e_{d_t}\right)\right) \leq n.$$

Demostración. En el caso donde $d_s = d_t$, el término $\mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t})$ desaparece y contribuye 0 a la discrepancia total, por lo que podemos aplicar la cota de diferencia de n para los términos sin profundidad.

En el caso de que s y t compartan un símbolo, se puede aplicar la cota de $n-1$ en los términos sin profundidad, y como el término $\mathrm{P}(e_{d_s}) - \mathrm{P}(e_{d_t})$ puede aportar a lo suma una unidad adicional de diferencia, obtenemos el resultado deseado. □

Los resultados anteriores nos llevan a los siguientes teoremas.

Teorema 2. *Si existe una asignación de profundidad $d : \Sigma^n \to \Sigma$ tal que el subgrafo válido de B_n tenga un ciclo hamiltoniano, entonces dicho ciclo corresponde a una secuencia de De Bruijn de orden n con discrepancia a lo sumo $n + 1$.*

Teorema 3. *Si existe una asignación de profundidad $d : \Sigma^n \to \Sigma$ tal que*

- *cada par de cadenas con diferente profundidad comparten un símbolo, y*
- *el subgrafo válido de B_n tiene un ciclo hamiltoniano,*

entonces ese ciclo corresponde a una secuencia de De Bruijn de orden n con discrepancia a lo sumo n.

Para $|\Sigma| = 2$, la primera condición en el teorema 3 se cumple si y sólo si a la cadena con todos ceros y a la cadena con todos unos se les asigna la misma profundidad.

4.3. Árbol de ciclos ICR

Dado que $\mathrm{ICR}_1 : \Sigma^n \to \Sigma^n$ es una función biyectiva, divide el espacio Σ^n en ciclos. Sea $\mathbf{N}$ el conjunto de ciclos dado por la regla ICR_1. Usamos este conjunto para construir una secuencia de De Bruijn, para ello buscamos unir los ciclos para obtener un ciclo hamiltoniano.

Definición 7 (Órbita de un nodo)**.** *Definimos* $orbit : \Sigma^n \to \mathcal{P}(\Sigma^n)$ *como*

$$orbit(s) = \{s, ICR(s), ICR^2(s), \dots\}.$$

Alternativamente, se puede definir como la clase de equivalencia de la relación definida por:

$$R(s,t) \iff t = \mathrm{ICR}(s)^k \text{ para algún } k \in \mathbb{Z}.$$

Definición 8 (Grafo de ciclos)**.** *Definimos un grafo dirigido* $\mathbf{G}$ *con conjunto de nodos* $\mathbf{N}$ *y un arco* $(\mathcal{U}, \mathcal{V})$ *entre ciclos* $\mathcal{U}, \mathcal{V} \in \mathbf{N}$ *si solo si existe un* $s \in \mathcal{U}$ *tal que* $ICR_0(s) \in \mathcal{V}$*.*

Supongamos que existe un árbol dirigido $\mathbf{T}$ con raíz en $\mathcal{R} \in \mathbf{N}$ que abarca $\mathbf{G}$ (el árbol está dirigido desde la raíz hasta las hojas). En la siguiente sección demostramos la existencia de $\mathbf{T}$. El árbol $\mathbf{T}$ define una relación de parentesco $\mathrm{parent} : \mathbf{N} \setminus \{\mathcal{R}\} \to \mathbf{N}$.

Definición 9 (Representante)**.** *Sea* $\mathcal{U} \in \mathbf{N} \setminus \{\mathcal{R}\}$*. El representante* s *de* $\mathcal{U}$ *es cualquier nodo (fijo)* $s \in \mathcal{U}$ *tal que* $ICR_0^{-1}(s) \in parent(\mathcal{U})$ *y* $s[n-1] \equiv d \mod |\Sigma|$*, donde* d *es la profundidad de* $\mathcal{U}$ *en el árbol. Lo vamos a denotar como* $rep(\mathcal{U}) = s$*, y también* $rep(u) = s$ *para cualquier* $u \in \mathcal{U}$*. El conjunto de representantes de todos los ciclos se denota con* **Reps**.

Observemos que para todos los $s \in \mathbf{Reps}$, $\mathrm{ICR}_0^{-1}(s) \in \mathrm{parent}(\mathrm{orbit}(s))$. Ahora, definamos la siguiente regla,

$$\mathrm{R}(s) = \begin{cases} \mathrm{ICR}_0(s) & \text{si } \mathrm{ICR}_0(s) \in \mathbf{Reps} \\ \mathrm{ICR}_2(s) & \text{si } \mathrm{ICR}_1(s) \in \mathbf{Reps} \\ \mathrm{ICR}_1(s) & \text{en caso contrario.} \end{cases}$$

Lema 9. *R es una función.*

Demostración. Tenemos que demostrar que no hay $s \in \Sigma^n$ tal que $\mathrm{ICR}_0(s), \mathrm{ICR}_1(s) \in \mathbf{Reps}$. Sólo necesitamos probar el caso en donde $|\Sigma| > 2$, ya que para $|\Sigma| = 2$, ICR_0 y ICR_2 coinciden.

Supongamos que $\mathrm{ICR}_0(s), \mathrm{ICR}_1(s) \in \mathbf{Reps}$. Entonces, como por definición $\mathrm{orbit}(s) = \mathrm{orbit}(\mathrm{ICR}_1(s))$, y dado que $\mathrm{ICR}_0(s) \in \mathbf{Reps}$, necesariamente $\mathrm{orbit}(s)$ es el padre de $\mathrm{orbit}(\mathrm{ICR}_0(s))$. Esto implica que la profundidad de $\mathrm{orbit}(\mathrm{ICR}_0(s))$ es uno más que la profundidad de $\mathrm{orbit}(\mathrm{ICR}_1(s))$. Por otro lado, el último símbolo de $\mathrm{ICR}_1(s)$ es uno más el último símbolo de $\mathrm{ICR}_0(s)$, y por ser representantes, estos coinciden con las profundidades de sus respectivas órbitas módulo $|\Sigma|$. Como $|\Sigma| > 2$, esto es una contradicción. □

Probemos los siguientes lemas sobre esta regla.

Lema 10 (Biyectividad)**.** *La función R es biyectiva.*

Demostración. Supongamos que $\mathrm{R}(s) = \mathrm{R}(t)$, entonces necesariamente los sufijos propios máximos de s y t coinciden. Ahora bien, es fácil ver que si la regla aplicada a s y t es la misma, entonces $s = t$, porque todos los ICR_k son inyectivos. Entonces, tenemos tres casos sin simetrías:

- $\mathrm{R}(s) = \mathrm{ICR}_0(s) = \mathrm{ICR}_1(t) = \mathrm{R}(t)$: como $\mathrm{ICR}_0(s) \in \mathbf{Reps}$, ya que aplicamos el primer caso, entonces $\mathrm{ICR}_1(t) \in \mathbf{Reps}$, pero esto no es posible ya que aplicamos la tercera regla en t.
- $\mathrm{R}(s) = \mathrm{ICR}_0(s) = \mathrm{ICR}_2(t) = \mathrm{R}(t)$: podemos ver que $\mathrm{ICR}_1(\mathrm{ICR}_0^{-1}(\mathrm{ICR}_1(t))) = \mathrm{ICR}_2(t)$ para todo t. Entonces $\mathrm{ICR}_0^{-1}(\mathrm{ICR}_1(t)) = \mathrm{ICR}_1^{-1}(\mathrm{ICR}_0(s))$, llamemos a esta cadena u. Ahora bien, sabemos que $\mathrm{ICR}_0(s), \mathrm{ICR}_1(t) \in \mathbf{Reps}$ por lo que $\mathrm{ICR}_0(u), \mathrm{ICR}_1(u) \in \mathbf{Reps}$. Si $|\Sigma| > 2$ esto no es posible, lo demostramos en el lema anterior. Para $|\Sigma| = 2$, ya que $\mathrm{ICR}_0 = \mathrm{ICR}_2$ y son inyectivas, $s = t$.
- $\mathrm{R}(s) = \mathrm{ICR}_1(s) = \mathrm{ICR}_2(t) = \mathrm{R}(t)$: esto significa que $\mathrm{ICR}_1(t) \in \mathbf{Reps}$, pero esto no es posible ya que $\mathrm{ICR}_0(s) = \mathrm{ICR}_1(t)$ y entonces $\mathrm{ICR}_0(s) \in \mathbf{Reps}$.

Como $\mathrm{R} : \Sigma^n \to \Sigma^n$ y es inyectiva, entonces es biyectiva. □

Lema 11 (Transitividad)**.** *La biyección R es transitiva. Es decir, para cualquier $s, t \in \Sigma^n$ existe un número entero k tal que $R^k(s) = t$.*

Demostración. Dado que R es una biyección, divide Σ^n en ciclos de R disjuntos. Supongamos que hay un ciclo ICR $\mathcal{U} \in \mathbf{N}$ tal que $\mathcal{U}$ no está completamente contenido en ningún ciclo de R. Eso significa que $\mathcal{U}$ contiene dos nodos que pertenecen a diferentes ciclos de R. En particular, hay al menos dos cadenas $s \in \mathcal{U}$ tales que s y $\mathrm{ICR}_1(s)$ pertenecen a diferentes ciclos R. Por lo tanto, sin pérdida de generalidad, podemos suponer que $\mathrm{ICR}_1(s) \notin \mathbf{Reps}$ (ya que $\mathcal{U} \cap \mathbf{Reps}$ tiene un tamaño como máximo uno), y también que $\mathcal{U}$ es el nodo más profundo del árbol que intersecta dos ciclos R diferentes.

Claramente, $\mathrm{R}(s) \neq \mathrm{ICR}_1(s)$, de lo contrario s y $\mathrm{ICR}_1(s)$ pertenecerían al mismo ciclo de R. Dado que $\mathrm{ICR}_1(s) \notin \mathbf{Reps}$, la posibilidad restante (de la definición de R) es que $\mathrm{ICR}_0(s) \in \mathbf{Reps}$. De la definición de **Reps**, obtenemos que la órbita de $\mathrm{ICR}_0(s)$ es hija de la órbita de s.

De la definición de R, obtenemos que

$$\mathrm{R}(\mathrm{ICR}_1^{-1}(\mathrm{ICR}_0(s))) = \mathrm{ICR}_2(\mathrm{ICR}_1^{-1}(\mathrm{ICR}_0(s))) = \mathrm{ICR}_1(s),$$

y por lo tanto $\mathrm{ICR}_1^{-1}(\mathrm{ICR}_0(s))$ y $\mathrm{ICR}_1(s)$ pertenecen al mismo ciclo R. Pero $\mathrm{ICR}_0(s)$ pertenece al mismo ciclo de R que s, por lo que la órbita de $\mathrm{ICR}_0(s)$ *también* contiene dos ciclos de R diferentes, lo que contradice la suposición de que $\mathcal{U}$ es el ciclo más profundo.

Ahora supongamos que hay dos órbitas adyacentes $\mathcal{U}, \mathcal{V} \in \mathbf{N}$ en el árbol que están contenidas en diferentes ciclos de R. Sin pérdida de generalidad, supongamos que $\mathcal{V}$ es hijo de $\mathcal{U}$ y sea $s = \mathrm{rep}(\mathcal{V})$. Dado que $s \in \mathbf{Reps}$, sabemos que $\mathrm{R}(\mathrm{ICR}_0^{-1}(s)) = s$ y que $\mathrm{ICR}_0^{-1}(s)$ está en la órbita del padre de s (es decir, $\mathcal{U}$). Esto es una contradicción, ya que asumimos que $\mathcal{U}$ y $\mathcal{V}$ no pertenecen al mismo ciclo de R. □

Como R es una función biyectiva y transitiva, y también satisface la propiedad de que el arco $(s, \mathrm{R}(s))$ está en el grafo de De Bruijn, podemos construir un ciclo hamiltoniano tomando una nodo arbitrario como inicio, llamémoslo s, y luego aplicamos repetidamente la función R hasta llegar a s nuevamente.

4.3.1. Ciclo hamiltoniano en el subgrafo válido

Ahora que hemos demostrado que R genera un ciclo hamiltoniano en el grafo de De Bruijn, demostremos que todos sus arcos pertenecen al subgrafo válido. Para definir el subgrafo válido es necesario dar una asignación de profundidad.

Definición 10 (Asignación de profundidad)**.** *Para un árbol fijo de ciclos ICR, definimos*

$$d : \Sigma^n \to \Sigma \qquad d(s) = p + 1,$$

donde p es la profundidad de la órbita de s en el árbol, módulo $|\Sigma|$.

Lema 12. *El arco $(s, R(s))$ es válido.*

Demostración. Sea b el primer símbolo de s y c sea el último símbolo de $\mathrm{R}(s)$. Necesitamos considerar los tres casos en la definición de $\mathrm{R}(s)$:

- Caso $\mathrm{ICR}_0(s) \in$ **Reps**: En este caso $\mathrm{R}(s) = \mathrm{ICR}_0(s)$ y por lo tanto $c = b$ y $c = d_{\mathrm{R}(s)} - 1$. Además, por definición de **Reps**, la órbita de $\mathrm{R}(s)$ es hija de la órbita de s, por lo que tenemos $d_s + 1 = d_{\mathrm{R}(s)}$. Al juntarlos obtenemos que $b + 1 = d_{\mathrm{R}(s)}$ y $c = d_s$, por lo que el arco es válido.

- Caso $\mathrm{ICR}_1(s) \in$ **Reps**: En este caso $\mathrm{R}(s) = \mathrm{ICR}_2(s)$ y por lo tanto $c = b + 2$, y ya que $\mathrm{ICR}_1(s) \in$ **Reps** , $b + 1 = d_s - 1$. Dado que $\mathrm{ICR}_2(s) = \mathrm{ICR}_1(\mathrm{ICR}_0^{-1}(\mathrm{ICR}_1(s)))$, la órbita de $\mathrm{ICR}_2(s)$ es el padre de la órbita de s, y por lo tanto $d_{\mathrm{R}(s)} = d_s - 1$. Al juntarlos obtenemos $b + 1 = d_{\mathrm{R}(s)}$ y $c = d_s$, por lo que el arco es válido.

- Caso restante: Tenemos $\mathrm{R}(s) = \mathrm{ICR}_1(s)$, entonces $c = b + 1$ y $d_s = d_{\mathrm{R}(s)}$, por lo que el arco es una vez más válido.

□

4.3.2. Discrepancia acotada para el caso general

Debido al Teorema 2, tenemos que la secuencia de De Bruijn asociada con R tiene discrepancia como máximo $n + 1$.

4.3.3. Discrepancia acotada para el caso binario

Para el caso binario, podemos usar el teorema 3: dado que las cadenas $00\ldots0$ y $11\ldots1$ pertenecen al mismo ciclo ICR, se les asigna el mismo valor de profundidad, y por lo tanto las hipótesis para el teorema se cumplen y tenemos que la secuencia de De Bruijn asociada con R tiene discrepancia como máximo n.

4.4. Árbol explícito

Para concluir la prueba, necesitamos construir un árbol dirigido en **G** y elegir los representantes para cada órbita. Para realizar el algoritmo de manera eficiente, no podemos construir explícitamente el árbol durante la ejecución del algoritmo. En lugar de ello, definimos explícitamente el árbol y damos un algoritmo eficaz para calcular la pertenencia al conjunto **Reps**.

Definición 11 (Arreglo de diferencias)**.** *Dado $s \in \Sigma^n$, definimos su arreglo de diferencias $\Delta(s) \in \Sigma^n$,*

$$\Delta(s)[i] = \begin{cases} s[i-1] - s[i] & si\ 0 < i < n \\ s[n-1] - s[0] - 1 & si\ i = 0. \end{cases}$$

El -1 adicional en el caso $i = 0$ garantiza la siguiente propiedad:

Lema 13. *Sea $s \in \Sigma^n$, entonces $\Delta(ICR_1(s)) = ICR_0(\Delta(s))$.*

Demostración. Separamos los casos $i = 0$, $0 < i < n-1$, $i = n-1$. Para el caso $i = 0$, tenemos que

$$\begin{aligned} \Delta(\mathrm{ICR}_1(s))[0] &= \mathrm{ICR}_1(s)[n-1] - \mathrm{ICR}_1(s)[0] - 1 \\ &= s[0] + 1 - s[1] - 1 \\ &= s[0] - s[1] \\ &= \mathrm{ICR}_0(\Delta(s))[0]. \end{aligned}$$

Para el caso $0 < i < n-1$

$$\begin{aligned} \Delta(\mathrm{ICR}_1(s))[i] &= \mathrm{ICR}_1(s)[i-1] - \mathrm{ICR}_1(s)[i] \\ &= s[i] - s[i+1] \\ &= \mathrm{ICR}_0(\Delta(s))[i]. \end{aligned}$$

Finalmente, para el caso $i = n-1 \wedge n > 1$

$$\begin{aligned} \Delta(\mathrm{ICR}_1(s))[n-1] &= \mathrm{ICR}_1(s)[n-2] - \mathrm{ICR}_1(s)[n-1] \\ &= s[n-1] - s[0] - 1 \\ &= \mathrm{ICR}_0(\Delta(s))[n-1]. \end{aligned}$$

Esto concluye la prueba. □

El lema 13 implica que para cada $\mathcal{U} \in \mathbf{N}$ sus elementos tienen el mismo arreglo de diferencia módulo rotaciones. Por convención, decimos que $\Delta(\mathcal{U})$ es la rotación lexicográficamente mínima de $\Delta(s)$ para cualquier $s \in \mathcal{U}$.

Lema 14. *Sea $s, t \in \Sigma^n$. Si $\Delta(s)$ y $\Delta(t)$ son iguales módulo rotaciones, entonces orbit$(s) =$ orbit(t).*

Demostración. Considere la cadena $u = \mathrm{ICR}_1^k(s)$. Podemos variar k de modo que $\Delta(u) = \Delta(t)$, de tal forma las cadenas u y t solo difieren en una constante agregada. Dado que las cadenas que difieren por constantes agregadas pertenecen a la misma órbita, y tanto s como u también pertenecen a la misma órbita, queda demostrado el resultado. □

Lema 15. *Sea $s \in \Sigma^n$. Entonces*

$$\sum_{i=0}^{n-1} \Delta(s)[i] = -1.$$

Demostración. Si consideramos la definición de Δ, cada símbolo de s aparece dos veces en la suma, una con signo positivo y otra con signo negativo. El único valor restante a tener en cuenta es el -1 en la definición de $\Delta(s)[0]$, que da el resultado. □

Lema 16. *Sea $t \in \Sigma^n$ tal que*

$$\sum_{i=0}^{n-1} t[i] = -1.$$

Entonces existe un $s \in \Sigma^n$ tal que $\Delta(s) = t$. Este s es único módulo constantes agregadas.

Demostración. Si elegimos $s[0]$ arbitrariamente, entonces una vez que se calcula $s[i]$, el valor de $s[i+1]$ se desprende directamente de la ecuación

$$\Delta(s)[i+1] = t[i+1]$$

La única ecuación que queda por satisfacer es $\Delta(s)[0] = t[0]$. Sin embargo, observemos que $\Delta(s)$ y t coinciden en $n-1$ lugares y tienen la misma suma, -1, por lo que deben ser idénticos. □

Equipados con el concepto de arreglo de diferencias, podemos definir el árbol $\mathbf{T}$ explícitamente de la siguiente manera.

Definición 12 (Árbol explícito)**.** *La raíz del árbol $\mathcal{R}$ se define como la órbita de $00\ldots0$. Para cualquier otro $\mathcal{U} \in \mathbf{N} \setminus \mathcal{R}$, definimos su padre $\mathcal{V} \in \mathbf{N}$ de la siguiente manera: Sea i el primer entero no negativo tal que $\Delta(\mathcal{U})[i] \neq 0$. Considere el arreglo a obtenido al disminuir el símbolo i-ésimo de $\Delta(\mathcal{U})$ e incrementar el símbolo $(i+1)$-ésimo. $\mathcal{V}$ es la única órbita tal que $\Delta(\mathcal{V})$ es igual a a módulo rotaciones.*

Para demostrar que esta definición es buena, necesitamos demostrar que tal i siempre existe, que $(\mathcal{V},\mathcal{U})$ es un arco en $\mathbf{G}$ para todo $\mathcal{U}$, y que la relación de parentesco no tiene ciclos.

La existencia de i se deriva del hecho de que la suma de los elementos del arreglo de diferencias es -1, por lo que no puede consistir únicamente en ceros. De hecho, $i \leq n-2$, ya que de lo contrario en $\Delta(\mathcal{U})$ habría $n-1$ ceros y, necesariamente, el símbolo restante sería -1, lo cual genera el arreglo de diferencias de $\mathcal{R}$.

Para demostrar que la relación de parentesco no tiene ciclos, notemos que a es lexicográficamente menor que $\Delta(\mathcal{U})$, por lo que

$$\Delta(\mathcal{V}) \leq a < \Delta(\mathcal{U})$$

y esto asegura que no haya ciclos.

Queda por demostrar que $\mathcal{U}$ y $\mathcal{V}$ son adyacentes. Para probar esto, basta con encontrar un $s \in \mathcal{U}$ tal que $\mathrm{ICR}_0^{-1}(s) \in \mathcal{V}$, que también puede servir como representante de $\mathcal{U}$. Sea i el primer entero no negativo tal que $\Delta(\mathcal{U})[i] \neq 0$, y considere el arreglo $k = \mathrm{ICR}_0^{i+1}(\Delta(\mathcal{U}))$. Es decir, el arreglo $\Delta(\mathcal{U})$ gira de manera que el primer elemento distinto de cero esté en la última posición. Sea $s \in \Sigma^n$ una cadena con $\Delta(s) = k$. Podemos elegir la constante agregada de s según sea necesario para satisfacer la restricción del último símbolo en la definición de representante.

Queremos demostrar que $\mathrm{ICR}_0^{-1}(s) \in \mathcal{V}$. Para ello, vamos a ver que $\Delta(\mathrm{ICR}_1(\mathrm{ICR}_0^{-1}(s)))$ es una rotación de a, ya que si demostramos esto $\mathrm{ICR}_1(\mathrm{ICR}_0^{-1}(s)) \in \mathcal{V}$ y como está en la misma órbita que $\mathrm{ICR}_0^{-1}(s)$, implicaría que este ultimo también está en $\mathcal{V}$.

Observemos que $\mathrm{ICR}_1(\mathrm{ICR}_0^{-1}(s))$ es lo mismo que s pero con el último símbolo aumentado en uno. El efecto que tiene aumentar el último símbolo de s sobre $\Delta(s)$ es el de disminuir el último símbolo y aumentar el primer símbolo. Dado que $\Delta(s) = \mathrm{ICR}_0^{i+1}(\Delta(\mathcal{U}))$, tenemos que $\Delta(\mathrm{ICR}_1(\mathrm{ICR}_0^{-1}(s)))$ es lo mismo que $\mathrm{ICR}_0^{i+1}(\Delta(\mathcal{U}))$ pero el último símbolo disminuyó y el primero aumentó, que es precisamente lo mismo que $\mathrm{ICR}_0^{i+1}(a)$, es decir, una rotación de a. Esto concluye la prueba.

4.4.1. Algoritmo

El núcleo del algoritmo es la regla de transición, que definimos de la siguiente manera

$$\mathrm{ER}(s) = \begin{cases} \mathrm{ICR}_0(s) & \text{si } \mathrm{ICR}_0(s) \in \mathbf{Reps} \\ \mathrm{ICR}_2(s) & \text{si } \mathrm{ICR}_1(s) \in \mathbf{Reps} \\ \mathrm{ICR}_1(s) & \text{en caso contrario} \end{cases}$$

El paso computacionalmente más difícil consiste en determinar si una cadena pertenece a **Reps** o no. Este tiene dos partes:

- El arreglo de diferencias de la cadena debe tener la forma correcta. Es decir, debería ser la rotación lexicográfica mínima, desplazada de modo que el primer elemento distinto de cero esté al final. Para ser un representante s tampoco debe estar en la órbita de la raíz, lo cual también podemos verificar con el arreglo de diferencias.
- El último símbolo de la cadena debe ser igual a la profundidad de la órbita módulo $|\Sigma|$.

La primera parte se implementa en el Algoritmo 1. La segunda parte se implementa junto con la función de transición en el Algoritmo 2. Notemos que ambas funciones tienen complejidad lineal tanto en tiempo como en memoria. Para determinar si d_2 es la rotación lexicográficamente mínima podemos utilizar el algoritmo de Booth [**?**], por ejemplo.

Algorithm 1: CorrectDifferenceArray

Data: $s \in \Sigma^n$
Result: Verdadero si s tiene el arreglo de diferencias correcto para ser un representante, Falso en caso contrario
$d \leftarrow \Delta(s)$
if $d[n-1] = 0$ **then**
  return *False*
end
$i \leftarrow n-1$
while $i > 0$ *y* $s[i-1] = 0$ **do**
  $i \leftarrow i-1$
end
if $i = 0$ **then**
  return *False*
end
$d_2 \leftarrow \mathrm{ICR}_0^i(d)$
return d_2 *es su rotación lexicográficamente mínima?*

4.4.2. Comportamiento del algoritmo para alfabeto arbitrario

Para $|\Sigma| = 2$, sabemos que nuestro algoritmo produce secuencias de De Bruijn de discrepancia exactamente n, porque esa es tanto la cota superior como la cota inferior trivial. Para el caso $n = 1$, también sabemos que nuestro algoritmo produce una secuencia de discrepancia exactamente n, porque todas las secuencias de De Bruijn para $n = 1$ tienen discrepancia n. Por tanto, en ambos casos nuestro algoritmo es óptimo.

Algorithm 2: Transición

Data: $s \in \Sigma^n$
profundidad $\in \Sigma$
Result: Un par $(s', \text{profundidad}')$ que corresponde al siguiente nodo en el ciclo
if *CorrectDifferenceArray*$(ICR_0(s))$ *y* $ICR_0(s)[n-1] = profundidad + 1$ **then**
  return $ICR_0(s), profundidad + 1$
end
if *CorrectDifferenceArray*$(ICR_1(s))$ *y* $ICR_1(s)[n-1] = profundidad$ **then**
  return $ICR_2(s), profundidad - 1$
end
return $ICR_1(s), profundidad$

Sin embargo, para el caso $|\Sigma| > 2$ y $n > 1$, el límite inferior y el límite superior tienen un espacio de tamaño uno: sabemos que nuestro algoritmo produce una secuencia de discrepancia de De Bruijn como máximo $n + 1$, y debe ser al menos n, y no sabemos cuál es. Para determinar esto, realizamos experimentos para $2 \leq n \leq 7$ y $3 \leq |\Sigma| \leq 9$, y encontramos que en todos los casos nuestro algoritmo tiene una discrepancia de $n + 1$.

4.4.3. Discrepancia mínima alcanzable

Para comparar el comportamiento del algoritmo con la discrepancia mínima real alcanzable, decidimos realizar una búsqueda exhaustiva de la discrepancia más pequeña que se puede obtener con ciertos parámetros de n y $|\Sigma|$, para decidir si nuestro algoritmo era óptimo o no. Los resultados que obtuvimos se pueden ver en la Tabla 4.1.

$\|\Sigma\|$**n**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
2	n	n	n	n	n	n	n
3	n	n+1	n	n			
4	n	n+1	n				
5	n	n					
6	n	n					
7	n	n					
8	n	n					

Tab. 4.1: Cada celda representa la discrepancia mínima alcanzable para una secuencia de De Bruijn con los parámetros correspondientes. Los casos que no se pudieron calcular debido al alto costo computacional se dejaron en blanco.

La experimentación está muy limitada debido a la naturaleza doblemente exponencial de la búsqueda. Sin embargo, con base en los resultados, planteamos la siguiente conjetura

Conjetura 1. *La discrepancia óptima de una secuencia de De Bruijn de orden n y alfabeto* Σ *es n, con la excepción de los casos* $n = 2$ *con* $|\Sigma| = 3$*; y* $n = 2$ *con* $|\Sigma| = 4$*.*

Esto se basa en la intuición de que

- La mayoría de las pruebas tuvieron secuencias de discrepancia n, excepto los dos casos antes mencionados.

- El grafo de De Bruijn se interconecta más al aumentar n y $|\Sigma|$, por lo que esperamos que sea más fácil lograr una secuencia de De Bruijn de baja discrepancia cuanto mayores sean estos parámetros.

Para $n = 1$ o $|\Sigma| = 2$, ya hemos demostrado que nuestro algoritmo produce la discrepancia mínima. En la Tabla 4.1, podemos ver que los únicos otros casos donde la discrepancia obtenida fue óptima fueron estas excepciones conjeturadas, $(n, |\Sigma|) \in \{(2, 3), (2, 4)\}$.

5. FUTURAS INVESTIGACIONES

Hay dos caminos de futuras investigaciones que se pueden realizar a partir de este trabajo:

- Estudiar la discrepancia mínima de las secuencias de De Bruijn para bloques de tamaño mayor a 1. La dificultad en esos casos es que bloques consecutivos no son independientes por lo cual no podemos regular la discrepancia tan fácilmente como lo hacíamos para símbolos.
- Estudiar la conjetura 1 y diseñar un algoritmo que logre la discrepancia óptima para alfabeto arbitrario. Pudimos esbozar una demostración de que utilizando ciclos ICR para construir la secuencia de De Bruijn no se podrá mejorar la discrepancia de $n + 1$. Por lo cual, se debe buscar otro camino.

Bibliografía

[1] Kellogg S. Booth. Lexicographically least circular substrings. *Information Processing Letters*, 10(4):240–242, 1980.

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

[3] Daniel Gabric and Joe Sawada. Investigating the discrepancy property of de bruijn sequences. *Discrete Mathematics*, 345(4):112780, 2022.

[4] Solomon W. Golomb. *Shift register sequences*. Holden-Day, Inc., San Francisco, Calif.-Cambridge-Amsterdam, 1967. With portions co-authored by Lloyd R. Welch, Richard M. Goldstein, and Alfred W. Hales.

[5] Yue Jiang Huang. A new algorithm for the generation of binary de Bruijn sequences. *Journal of Algorithms*, 11(1):44–51, 1990.