



UNA LONGITUD EN SECUENCIAS DE NUCLEÓTIDOS Y EL NÚMERO NULA

L. Rodríguez, F. Novoa

*Departamento de Matemáticas, Facultad de Ciencias. Pontificia Universidad Javeriana,
Carrera 7ª No. 43-82, Bogotá, D.C.
lrodrigu@javeriana.edu.co,
fernando.novoa@javeriana.edu.co*

RESUMEN

Se define sobre secuencias de nucleótidos una longitud similar a la longitud de permutaciones para grupos de simetría. Esta longitud presenta algunas propiedades y relaciones similares a las propiedades analizadas en permutaciones y además permite establecer propiedades compatibles con la representación tridimensional de cadenas de nucleótidos como se definió en [1, 2].

Palabras clave: número nula, permutaciones, longitud de permutación, grupo simétrico.

ABSTRACT

A length over sequences of nucleotides is defined in similar way as is defined over permutations of the symmetric group. This length shows similar properties as the length for permutations and also shows some compatible properties with the three dimension representation defined in [1,2].

Key words: Nula number, permutations, length of a permutation, symmetric group

INTRODUCCIÓN

Para las cadenas de nucleótidos se tienen varias representaciones espaciales. En particular Lareo y Acevedo [2] definen una representación por medio de números de Gödel, sin embargo, esta representación no es uno a uno. Por lo tanto, es posible que a una terna $(x,y,z) \in R^3$ le correspondan varias cadenas que se identifican con esta terna según esta representación.

Una de las preguntas matemáticas que se intenta responder es si es posible determi-

nar el número de cadenas que se identifican con una terna, en términos de algún parámetro que posean estas cadenas de nucleótidos. Estas y otras preguntas sobre el número Nula ya han sido planteadas en [1].

Observando las cadenas correspondientes a una terna $(x,y,z) \in R^3$, se busca hallar relaciones para establecer los parámetros a estudiar bajo los cuales con esta representación tridimensional dos o más cadenas tienen la misma representación, y de esta forma poder contarlas en forma efi-

ciente. Considerando las cadenas como permutaciones generalizadas, se define una longitud sobre ellas y se observa que aquellas que tienen la misma representación tridimensional poseen longitudes similares. Es decir, el parámetro longitud parece tener alguna relación con el hecho de tener la misma representación tridimensional. A continuación se dan las definiciones básicas y posteriormente se presentan algunas pruebas de las conjeturas planteadas en este trabajo.

DEFINICIONES BÁSICAS

Definición 1. Una permutación de n elementos es una función biyectiva del conjunto $\{1, \dots, n\}$ sobre sí mismo. El conjunto de todas estas funciones biyectivas se denota como S_n y con la operación de composición usual de funciones forma un grupo, cuyo orden es $n!$.

Usualmente una permutación $w \in S_n$ se escribe como $w = w_1 \dots w_n$ en donde $w_i = w(i)$.

Definición 2. La longitud de una permutación $w \in S_n$ es el número de elementos del conjunto:

$$I(w) = \{(i, j) : i < j \text{ y } w_i > w_j\}.$$

El conjunto $I(w)$ se llama el conjunto de los descendentes de w , y la longitud de w se denota por $l(w)$.

Si $w = w_1 \dots w_n \in S_n$, denotamos por $m_i = \#\{j : j > i, w_i > w_j\}$ para todo $1 \leq i \leq n$.

Entonces $l(w) = \sum m_i$.

Ejemplo 1. Sea $w = [6, 2, 3, 1, 5, 4] \in S_6$. El conjunto de descendentes de w es:

$$I(w) = \{(1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 4), (3, 4), (5, 6)\}$$

por lo tanto su longitud es $l(w) = 8$.

Sea Q un conjunto finito de cardinalidad k , y el conjunto de todas las palabras (secuencias) de largo n que se pueden construir con los elementos (letras) de Q . Estas palabras de largo n se denotan por Q_n^* . Supongamos que Q está totalmente ordenado, es decir, $q_1 < q_2 < \dots < q_k$.

Definición 3. Sea $p \in Q_n^*$. La longitud de p , $l(p)$ es el número total de descendentes de p , es decir, el número de parejas (i, j) con $i < j$ y tal que $p_i > p_j$.

En este trabajo se utilizará $Q = \{A, C, G, T\}$ que representa el conjunto de nucleótidos. Por facilidad para algunos cálculos también se usará $Q = \{2, 3, 5, 7\}$, e identificará A, C, G y T con $2, 3, 5$ y 7 respectivamente.

Bajo esta identificación, es claro que para calcular la longitud de una secuencia se tiene $A < C < G < T$.

Ejemplo 2. La palabra $p = ACTCTATTAC$ es de largo 10 y su longitud es 15.

La representación tridimensional de una cadena de nucleótidos K fue definida en [2] y está dada por

$$\eta(K) = (2^\alpha 3^\beta 5^\gamma 7^\delta, 2^a 3^c 5^g 7^t, n)$$

en donde α, β, γ y δ se definen como la suma de las posiciones en K de las A, C, G y T respectivamente y a, c, g y t son el número de A 's, C 's, G 's y T 's respectivamente que hay en K . Podemos identificar esta representación tridimensional por $\eta(K) = (\alpha, \beta, \gamma, \delta, a, c, g, t)$. Por lo tanto, las siguientes relaciones se tienen trivialmente:

$$a + c + g + t = n$$

$$\alpha + \beta + \gamma + \delta = \frac{(n+1)n}{2} \tag{1.1}$$

Ejemplo 3. La cadena $K = ACGTTACCTAG = 23577233725$ tiene como representación tridimensional

$$\eta(K) = (2^{17}3^{17}5^{14}7^{18}, 2^33^35^27^3, 11),$$

Las siguientes secuencias también tienen la misma representación tridimensional:

[2, 7, 3, 3, 2, 5, 7, 5, 7, 3, 2], [2, 3, 7, 5, 2, 3, 7, 7, 3, 5, 2], [2, 3, 7, 5, 2, 7, 3, 3, 7, 5, 2], [2, 7, 3, 5, 2, 3, 7, 3, 7, 5, 2], [3, 2, 7, 2, 5, 3, 7, 7, 5, 3, 2], [3, 2, 7, 2, 7, 5, 3, 5, 3, 7, 2], [7, 2, 3, 2, 3, 5, 7, 5, 3, 7, 2], [7, 2, 3, 2, 5, 3, 7, 3, 5, 7, 2], [2, 3, 7, 3, 5, 2, 7, 7, 5, 2, 3], [2, 3, 5, 7, 7, 2, 3, 3, 7, 2, 5], [2, 3, 7, 7, 5, 2, 3, 3, 5, 2, 7], [2, 7, 5, 3, 3, 2, 7, 3, 7, 2, 5], [3, 2, 5, 7, 2, 7, 3, 7, 3, 2, 5], [3, 2, 7, 7, 2, 5, 3, 5, 3, 2, 7], [7, 2, 5, 3, 2, 3, 3, 7, 7, 2, 5], [3, 7, 2, 2, 3, 5, 7, 5, 7, 2, 3], [3, 7, 2, 2, 7, 5, 3, 5, 3, 2, 7], [7, 3, 2, 2, 5, 7, 3, 3, 5, 2, 7], [2, 3, 7, 3, 7, 5, 2, 5, 2, 7, 3], [2, 3, 5, 7, 3, 7, 2, 7, 2, 3, 5], [2, 3, 7, 7, 3, 5, 2, 5, 2, 3, 7], [2, 7, 3, 3, 7, 5, 2, 5, 2, 3, 7], [2, 7, 3, 5, 7, 3, 2, 3, 2, 5, 7], [2, 7, 5, 3, 3, 7, 2, 3, 2, 7, 5], [3, 2, 7, 5, 3, 2, 7, 7, 2, 5, 3], [7, 2, 5, 3, 3, 2, 7, 3, 2, 7, 5], [7, 3, 2, 3, 2, 5, 7, 5, 2, 7, 3], [7, 3, 2, 5, 2, 7, 3, 3, 2, 5, 7], [3, 2, 7, 5, 3, 7, 2, 2, 7, 5, 3], [3, 2, 5, 7, 7, 3, 2, 2, 7, 3, 5], [3, 2, 7, 7, 5, 3, 2, 2, 5, 3, 7], [7, 2, 3, 3, 5, 7, 2, 2, 5, 3, 7], [7, 2, 3, 5, 3, 7, 2, 2, 3, 5, 7], [3, 7, 2, 5, 3, 2, 7, 2, 7, 5, 3], [7, 3, 2, 3, 5, 2, 7, 2, 5, 7, 3], [3, 7, 2, 5, 7, 2, 3, 2, 3, 5, 7], [3, 7, 5, 2, 2, 3, 7, 2, 7, 3, 5], [3, 7, 5, 2, 2, 7, 3, 2, 3, 7, 5], [7, 3, 5, 2, 2, 3, 7, 2, 3, 7, 5], [7, 3, 5, 2, 3, 2, 2, 7, 7, 3, 5],

Las longitudes de todas estas secuencias que tienen la misma representación están entre 20 y 22.

En general, el número de secuencias de nucleótidos que tienen la misma representación tridimensional que una secuencia K está dada por el número de soluciones del sistema

$$\begin{aligned} \chi_1 + \dots + \chi_a &= \alpha \\ \chi_{a+1} + \dots + \chi_{a+c} &= \beta \end{aligned}$$

$$\begin{aligned} \chi_{a+c+1} + \dots + \chi_{a+c+g} &= \gamma \\ \chi_{a+c+g+1} + \dots + \chi_n &= \delta \end{aligned}$$

bajo las condiciones $1 \leq x_i \leq n$, y además $x_i \neq x_j$.

Esto es equivalente al número de distintas particiones de $\alpha, \beta, \gamma, \delta$ en a, c, g, t partes respectivamente que sean estrictas y disyuntas y que satisfacen las condiciones de la ecuación (1.1).

Para las siguientes definiciones recordemos que los nucleótidos A y T lo mismo que C y G se dicen complementarios.

Definición 3. Sea K una cadena de largo n . La cadena complementaria de K es la cadena

$$K' = k'_1 \dots k'_n$$

en donde k'_i es el nucleótido complementario de k_i .

Ejemplo 4. La cadena complementaria de $K = ACGTTACCTAG$ es la cadena $K' = TGCAATGGATC$.

Definición 4. Sea K una cadena de largo n . La opuesta a K es la cadena \bar{K} , en donde $k_i = \bar{k}_{n+1-i}$.

Ejemplo 5. La cadena opuesta a $K = ACGTTACCTAG$ es $\bar{K} = GATCCATTGCA$.

Ejemplo 6. Entre todas las cadenas de largo 11 con 3 A 's, C 's, y T 's, 2 G 's, la de mayor longitud es la secuencia $K_0 = TTTGGCCCAA$ cuya longitud es

$$l(K_0) = 3(8) + 2(6) + 3(3) = 45.$$

Esta cadena se define como la cadena descendente cuando el número de cada uno de los nucleótidos sea fijo.

Ejemplo 7. Las cadenas de los ejemplos 4 y 5, es decir, $K' = TGCAATGGATC$ y $\bar{K} = GATCCATTGCA$ tienen longitud 25. La cadena $K = ACGTTACCTAG$ tiene longitud 20. De esto, se plantea la siguiente conjetura:

$$l(K) = l(\bar{K}) = l(K_0)$$

$$l(K) = l(K') = l(K_0) = l(K'_0)$$

Observación 1. La cadena K_0 depende únicamente del número de cada uno de los nucleótidos en la cadena K y por lo tanto, existen cadenas K, Q distintas con $K_0 = Q_0$.

Proposición 1. Entre todas las cadenas con exactamente el mismo número de nucleótidos, la secuencia decreciente es la de mayor longitud.

Demostración:

Sea n la suma total de los nucleótidos. Sea K una de tales cadenas, y supongamos que existe i tal que $k_i < k_{i+1}$. Entonces, la cadena Ks_i formada al permutar estas dos posiciones en la cadena K tiene al menos un descendente más que la cadena K y por lo tanto $l(Ks_i) > l(K)$. Siguiendo de esta forma, la cadena de mayor longitud es la cadena descendente.

Adicionalmente, se observa como se comporta la representación espacial o número Nula, respecto a los operadores inversión y complemento de cadenas.

Definición 5. Sea K una cadena de nucleótidos. Se define la imagen inversa del número Nula de K , como el conjunto de todas las cadenas Q , tales que $\eta(Q) = \eta(K)$. Es decir,

$$\{\eta^{-1}(K) = Q: \eta(Q) = \eta(K)\}.$$

Ejemplo 8. Sea $K = ACGTTACCTAG$. Se calcula su imagen inversa y luego a cada uno de sus elementos se le halla su cadena opuesta. Es decir, se calcula el conjunto

$\eta^{-1}(K)$ y luego $\eta^{-1}(\bar{K})$. Este cómputo se lleva a cabo con el programa Nula.pkg [2] desarrollado en CoCoA¹:

U = Nula.SeqSecuenciasOpuestas (Nula.ImagenInversa ([2,3,5,7,7,2,3,3,7,2,5]));

U;

[[2, 3, 7, 5, 7, 5, 2, 3, 3, 7, 2], [2, 5, 3, 7, 7, 3, 2, 5, 7, 3, 2], [2, 5, 7, 3, 7, 3, 2, 5, 3, 7, 2], [2, 3, 5, 7, 7, 3, 5, 2, 7, 2, 3], [2, 7, 3, 5, 3, 5, 7, 2, 7, 2, 3], [2, 7, 3, 5, 7, 5, 3, 2, 3, 2, 7], [2, 7, 5, 3, 7, 3, 5, 2, 3, 2, 7], [3, 2, 5, 7, 7, 2, 5, 3, 7, 3, 2], [5, 2, 7, 3, 3, 2, 7, 7, 5, 3, 2], [7, 2, 5, 3, 3, 2, 5, 7, 7, 3, 2], [5, 2, 7, 3, 7, 2, 3, 3, 5, 7, 2], [5, 2, 3, 7, 3, 7, 2, 7, 5, 2, 3], [7, 2, 3, 5, 3, 5, 2, 7, 7, 2, 3], [5, 2, 7, 7, 3, 3, 2, 3, 5, 2, 7], [3, 2, 7, 5, 7, 5, 3, 2, 2, 7, 3], [7, 2, 3, 5, 3, 5, 7, 2, 2, 7, 3], [7, 2, 5, 3, 3, 7, 5, 2, 2, 3, 7], [3, 7, 2, 5, 2, 5, 7, 3, 7, 3, 2], [5, 3, 2, 7, 2, 7, 3, 7, 5, 3, 2], [7, 3, 2, 5, 2, 5, 3, 7, 7, 3, 2], [7, 3, 2, 5, 2, 5, 7, 3, 3, 7, 2], [7, 5, 2, 3, 2, 3, 7, 5, 3, 7, 2], [5, 7, 2, 3, 2, 7, 3, 3, 5, 7, 2], [3, 5, 2, 7, 7, 2, 3, 5, 7, 2, 3], [5, 7, 2, 3, 7, 2, 3, 5, 2, 7], [3, 7, 2, 5, 7, 5, 2, 3, 7], [7, 5, 2, 3, 3, 7, 2, 5, 2, 3, 7], [3, 5, 7, 2, 2, 7, 3, 5, 7, 2, 3], [5, 3, 7, 2, 2, 3, 7, 7, 5, 2, 3], [7, 3, 5, 2, 2, 3, 5, 7, 7, 2, 3], [7, 3, 5, 2, 2, 7, 5, 3, 3, 2, 7], [7, 5, 3, 2, 2, 7, 3, 5, 3, 2, 7], [3, 5, 7, 2, 7, 2, 3, 5, 2, 7, 3], [3, 7, 5, 2, 7, 2, 5, 3, 2, 3, 7], [7, 5, 3, 2, 3, 2, 7, 5, 2, 7, 3], [5, 3, 7, 2, 7, 3, 2, 2, 5, 7, 3], [5, 7, 3, 2, 3, 7, 2, 2, 5, 7, 3], [5, 7, 3, 2, 7, 3, 2, 2, 5, 3, 7], [5, 3, 7, 7, 2, 2, 3, 2, 5, 3, 7]]

5, 3, 7, 2, 7, 3, 2, 2, 5, 7, 3], [5, 7, 3, 2, 3, 7, 2, 2, 5, 7, 3], [5, 7, 3, 2, 7, 3, 2, 2, 5, 3, 7], [5, 3, 7, 7, 2, 2, 3, 2, 5, 3, 7]]

V = Nula.Imagen Inversa (Nula.Secuencia Opuesta ([2,3,5,7,7,2,3,3,7,2,5]));

V;

[[2, 5, 3, 7, 7, 3, 2, 5, 7, 3, 2], [2, 5, 7, 3, 3, 7, 2, 5, 7, 3, 2], [2, 3, 7, 5, 7, 5, 2, 3, 3, 7, 2], [2, 5, 7, 3, 7, 3, 2, 5, 3, 7, 2], [3, 2, 5, 7, 7, 2, 5, 3, 7, 3, 2], [5, 2, 7, 3, 3, 2, 7, 7, 5, 3, 2], [7, 2, 5,

¹ CoCoA es un programa de Álgebra Conmutativa.

3, 3, 2, 5, 7, 7, 3, 2], [5, 2, 7, 3, 7, 2, 3, 3, 5, 7, 2], [3, 7, 2, 5, 2, 5, 7, 3, 7, 3, 2], [5, 3, 2, 7, 2, 7, 3, 7, 5, 3, 2], [7, 3, 2, 5, 2, 5, 3, 7, 7, 3, 2], [7, 3, 2, 5, 2, 5, 7, 3, 3, 7, 2], [7, 5, 2, 3, 2, 3, 7, 5, 3, 7, 2], [5, 7, 2, 3, 2, 7, 3, 3, 5, 7, 2], [2, 3, 5, 7, 7, 3, 5, 2, 7, 2, 3], [2, 7, 3, 5, 3, 5, 7, 2, 7, 2, 3], [2, 7, 3, 5, 7, 5, 3, 2, 3, 2, 7], [2, 7, 5, 3, 7, 3, 5, 2, 3, 2, 7], [5, 2, 3, 7, 3, 7, 2, 7, 5, 2, 3], [7, 2, 3, 5, 3, 5, 2, 7, 7, 2, 3], [5, 2, 7, 7, 3, 3, 2, 3, 5, 2, 7], [3, 5, 2, 7, 7, 2, 3, 5, 7, 2, 3], [5, 7, 2, 3, 7, 2, 3, 3, 5, 2, 7], [3, 5, 7, 2, 2, 7, 3, 5, 7, 2, 3], [5, 3, 7, 2, 2, 3, 7, 7, 5, 2, 3], [7, 3, 5, 2, 2, 3, 5, 7, 7, 2, 3], [7, 3, 5, 2, 2, 7, 5, 3, 3, 2, 7], [7, 5, 3, 2, 2, 7, 3, 5, 3, 2, 7], [3, 2, 7, 5, 7, 5, 3, 2, 2, 7, 3], [7, 2, 3, 5, 3, 5, 7, 2, 2, 7, 3], [7, 2, 5, 3, 3, 7, 5, 2, 2, 3, 7], [3, 7, 2, 5, 7, 5, 2, 3, 2, 3, 7], [7, 5, 2, 3, 3, 7, 2, 5, 2, 3, 7], [3, 5, 7, 2, 7, 2, 3, 5, 2, 7, 3], [7, 5, 3, 2, 3, 2, 7, 5, 2, 7, 3], [3, 7, 5, 2, 7, 2, 5, 3, 2, 3, 7], [5, 3, 7, 2, 7, 3, 2, 2, 5, 7, 3], [5, 7, 3, 2, 3, 7, 2, 2, 5, 7, 3], [5, 7, 3, 2, 7, 3, 2, 2, 5, 3, 7], [5, 3, 7, 2, 2, 3, 2, 5, 3, 7]]

Con una simple rutina se puede verificar que estos dos conjuntos son iguales. Por lo tanto se puede conjeturar que

$$\overleftarrow{\eta^{-1}(K)} = \overleftarrow{\eta^{-1}(\overline{K})} \quad (1.2)$$

y de manera similar:

$$(\eta^{-1}(K))' = \eta^{-1}(K') \quad (1.3)$$

que se demuestra a partir de los siguientes resultados.

Proposición 2. Sea K una cadena de largo n con número espacial (número nula) determinado por $(\alpha, \beta, \gamma, \delta, a, c, g, t, n)$. Entonces el número espacial de la cadena opuesta \overline{K} está determinado por

$$(a(n+1) - \alpha, c(n+1) - \beta, g(n+1) - \gamma, t(n+1) - \delta, a, c, g, t, n)$$

Demostración:

Inicialmente se puede notar que el número de cada nucleótido es el mismo en K y en \overline{K} .

Se supone que la cadena K tiene en las posiciones i_1, \dots, i_a los nucleótidos A . Por lo tanto, la cadena \overline{K} tiene en las posiciones $n+1-i_1, \dots, n+1-i_a$ los nucleótidos A .

Como además,

$$\sum_{j=1}^a i_j = \alpha$$

entonces,

$$\sum_{j=1}^a (n+1-i_j) = (n+1) - \alpha$$

De manera similar se calculan las otras coordenadas del número nula de \overline{K} .

Proposición 3. Sea K una cadena de largo n con número espacial determinado por $(\alpha, \beta, \gamma, \delta, a, c, g, t, n)$. Entonces el número espacial de K' está determinado por $(\delta, \gamma, \beta, \alpha, t, g, c, a, n)$.

Demostración:

Inmediata a partir de la definición de K' .

De lo anterior se deduce la siguiente proposición.

Proposición 4. Dada una cadena K , se tiene que:

$$\overleftarrow{\eta^{-1}(K)} = \overleftarrow{\eta^{-1}(\overline{K})}$$

y

$$(\eta^{-1}(K))' = \eta^{-1}(K')$$

Demostración:

Si $\eta(K) = (\alpha, \beta, \gamma, \delta, a, c, g, t, n)$. Sea $Q \in \eta^{-1}(\overline{K})$. Entonces existe $Q_1 \in \eta^{-1}(K)$ tal que $Q = \overline{Q_1}$. Por lo tanto,

$$\eta(Q_1) = (\alpha, \beta, \gamma, \delta, a, c, g, t, n).$$

De esta forma,

$$\eta(Q) = a(n+1) - \alpha, c(n+1) - \beta, g(n+1) - \gamma, t(n+1) - \delta, a, c, g, t, n = \eta(\overline{K})$$

por lo tanto, $Q \in \eta^{-1}(\bar{K})$. De manera similar se obtiene la otra contención.

Ahora bien, si $Q \in (\eta^{-1}(K))'$, entonces existe $Q_1 \in \eta^{-1}(K)$ tal que $Q = Q_1'$. Por lo tanto, $\eta(Q_1) = (\alpha, \beta, \gamma, \delta, a, c, g, t, n)$ y entonces $\eta(Q) = (\delta, \gamma, \beta, \alpha, t, g, c, a, n)$ que es el número nula de K' , es decir, $\eta(Q) = \eta(K')$ y por lo tanto $Q_1 \in \eta^{-1}(K')$. La otra contención se obtiene similarmente.

Definición 6. Una partición de un entero positivo es una secuencia finita de enteros positivos $\lambda = (\lambda_1, \dots, \lambda_k)$ tales que

$$n = \sum_{i=1}^k \lambda_i$$

con

$$\lambda_1 \geq \dots \geq \lambda_k > 0.$$

Dada una partición $\lambda = (\lambda_1, \dots, \lambda_k)$ los elementos λ_i se llaman las partes de λ y k es el número de partes de λ . En lo sucesivo, consideraremos particiones con a lo más cuatro partes.

Ejemplo 9. Sea $n = 11$ y $\lambda = (3, 3, 3, 2)$. Considérense las secuencias $K_0 = TTTGGCCCAA$ y $Q_0 = TTTGGCCCAA$. Estas secuencias determinan la misma partición, de 11, $\lambda = (3, 3, 3, 2)$. Sus longitudes son:

$$l(K_0) = 3(8) + 2(6) + 3(3) = 45$$

$$l(Q_0) = 3(8) + 3(5) + 3(2) = 45.$$

De lo anterior, se puede conjeturar que dos cadenas descendentes del mismo largo cuyos nucleótidos forman la misma partición, tienen la misma longitud.

En efecto, se tienen los siguientes resultados.

Proposición 5. Sea S_0 una cadena descendente y sean a, c, g, t el número de nucleótidos A, C, G, T respectivamente, entonces

$$l(S_0) = t(a + c + g) + g(a + c) = ca.$$

Demostración:

Inmediata.

Puede notarse que desarrollando la ecuación anterior se tiene,

$$l(S_0) = t(a + c + g) + g(a + c) = ca = ac + ag + at + cg + ct + gt.$$

Definición 7. El segundo polinomio elemental simétrico en cuatro variables x_1, \dots, x_4 se define como

$$e_2^4(x_1, \dots, x_4) = \sum_{1 \leq i < j \leq 4} x_i x_j = x_1 x_2 + x_1 x_3 + x_1 x_4 + x_2 x_3 + x_2 x_4 + x_3 x_4.$$

Para toda $\sigma \in S_4$ se tiene que $e_2^4(x_1, \dots, x_4) = e_2^4(x_{\sigma(1)}, \dots, x_{\sigma(4)})$.

También se tiene el siguiente resultado:

Proposición 6. Con la notación de la proposición anterior, $l(S_0) = e_2^4(a, c, g, t)$. Más aún, para toda cadena decreciente T_0 que determine la misma partición determinada por a, c, g, t fijos, se tiene $l(T_0) = l(S_0) = e_2^4(a, c, g, t)$.

Demostración:

Inmediata a partir de las anteriores observaciones.

De esta forma, para un n fijo, se pueden restringir sus particiones a aquellas a lo más con cuatro partes, para determinar la cadena de longitud maximal.

Ejemplo 10. Sea $n = 11$. Las particiones de n con cuatro partes a lo más están almacenadas en el vector Z calculado abajo:

U:=Sym.Parti2(11);

Z:=[];

For I:=1 To Len(U) Do

If Len(U[I])<=4 Then Append(Z,U[I]) EndIf

EndFor;

Z;

[[11], [10, 1], [9, 2], [9, 1, 1], [8, 3], [8, 2, 1], [8, 1, 1, 1], [7, 4], [7, 3, 1], [7, 2, 2], [7, 2, 1, 1], [6, 5], [6, 4, 1], [6, 3, 2], [6, 3, 1, 1], [6, 2, 2, 1], [5, 5, 1], [5, 4, 2], [5, 4, 1, 1], [5, 3, 3], [5, 3, 2, 1], [5, 2, 2, 2], [4, 4, 3], [4, 4, 2, 1], [4, 3, 3, 1], [4, 3, 2, 2], [3, 3, 3, 2]]

La evaluación² del segundo polinomio elemental simétrico en estas particiones es:

El segundo polinomio simétrico evaluado en [11, 0, 0, 0] tiene el valor 0

El segundo polinomio simétrico evaluado en [10, 1, 0, 0] tiene el valor 10

El segundo polinomio simétrico evaluado en [9, 2, 0, 0] tiene el valor 18

El segundo polinomio simétrico evaluado en [9, 1, 1, 0] tiene el valor 19

El segundo polinomio simétrico evaluado en [8, 3, 0, 0] tiene el valor 24

El segundo polinomio simétrico evaluado en [8, 2, 1, 0] tiene el valor 26

El segundo polinomio simétrico evaluado en [8, 1, 1, 1] tiene el valor 27

El segundo polinomio simétrico evaluado en [7, 4, 0, 0] tiene el valor 28

El segundo polinomio simétrico evaluado en [7, 3, 1, 0] tiene el valor 31

El segundo polinomio simétrico evaluado en [7, 2, 2, 0] tiene el valor 32

El segundo polinomio simétrico evaluado en [7, 2, 1, 1] tiene el valor 33

El segundo polinomio simétrico evaluado en [6, 5, 0, 0] tiene el valor 30

El segundo polinomio simétrico evaluado en [6, 4, 1, 0] tiene el valor 34

El segundo polinomio simétrico evaluado en [6, 3, 2, 0] tiene el valor 36

El segundo polinomio simétrico evaluado en [6, 3, 1, 1] tiene el valor 37

El segundo polinomio simétrico evaluado en [6, 2, 2, 1] tiene el valor 38

El segundo polinomio simétrico evaluado en [5, 5, 1, 0] tiene el valor 35

El segundo polinomio simétrico evaluado en [5, 4, 2, 0] tiene el valor 38

El segundo polinomio simétrico evaluado en [5, 4, 1, 1] tiene el valor 39

El segundo polinomio simétrico evaluado en [5, 3, 3, 0] tiene el valor 39

El segundo polinomio simétrico evaluado en [5, 3, 2, 1] tiene el valor 41

El segundo polinomio simétrico evaluado en [5, 2, 2, 2] tiene el valor 42

El segundo polinomio simétrico evaluado en [4, 4, 3, 0] tiene el valor 40

El segundo polinomio simétrico evaluado en [4, 4, 2, 1] tiene el valor 42

El segundo polinomio simétrico evaluado en [4, 3, 3, 1] tiene el valor 43

El segundo polinomio simétrico evaluado en [4, 3, 2, 2] tiene el valor 44

El segundo polinomio simétrico evaluado en [3, 3, 3, 2] tiene el valor 45

El ejemplo anterior vuelve a brindar otra conjetura acerca de cual de las particiones de n , con las cuales se está trabajando, proporciona la de mayor longitud.

Definición 8: Sean λ, μ particiones de n . Decimos que $\lambda < \mu$ si existe j tal que $\lambda_j < \mu_j$ y $\lambda_i = \mu_i = y$ para todo $1 \leq i < j$. Este orden se suele llamar el orden lexicográfico.

Ejemplo 11. Las particiones de 11, ordenadas lexicográficamente son:

$(3,3,3,2) < (4,3,2,2) < (4,3,3,1) < \dots(10,1) < (11)$.

Por lo tanto, como resultado de los anteriores ejemplos (y muchos otros), se plantea la siguiente conjetura:

² Cálculo realizado con CoCoA.

Conjetura 1. Las cadenas de largo n de mayor longitud son las cadenas descendentes que determinan la partición mínima de n (en el orden lexicográfico). Su longitud máxima está dada por la evaluación del segundo polinomio elemental simétrico en cuatro variables, en dicha partición.

Se observa que si $n = 4k$ para algún k , la primera parte del enunciado de la conjetura se tiene, pues la partición minimal es (k,k,k,k) . Por otro lado, el polinomio simétrico $s(x, y, z, w) = xy + xz + xw + yz + yw + zw$ tiene su máximo en dicha partición sujeto a la condición que $x + y + z + w = n$.

Las siguientes observaciones tienen la finalidad de establecer algunas propiedades similares que presentan estas cadenas de nucleótidos si les definimos un “orden” y las permutaciones de S_n con el llamado orden de Bruhat.

En la definición 2, se introdujeron los elementos m_i para establecer la longitud de una permutación. Sea ahora $K = k_1 \dots k_n$ una cadena con n_i veces el nucleótido i , en donde $i = \{A, C, G, T\}$ se distinguen los siguientes casos, para todo $j, 1 \leq j \leq n$. Si, $k_j = A$ entonces $m_j = 0$. Si $k_j = C$ entonces el número de A's a la derecha de la posición j , es decir m_j sumado con el número de A's a la izquierda de la posición j , y denotado por M_j es igual a n_A . Entonces para la cadena opuesta, \overleftarrow{K} se tiene que $\overleftarrow{m}_{n+i-j} = M_j$. Por lo tanto se tiene que

$$m_j + M_j + n_A.$$

Si $k_j = G$ entonces la suma de $m_j + M_j$ es la suma de todos los nucleótidos menores a G que se encuentran a derecha e izquierda de la posición j , es decir,

$$m_j + M_j = n_A + n_C$$

De manera similar si $k_j = T$ se tiene que $m_j + M_j = n_A + n_C + n_G$ y por lo tanto se tiene que:

$$\begin{aligned} l(K) + l(\overleftarrow{K}) &= \sum_{j=1}^n m_j + \sum_{j=1}^n M_j \\ &= \sum_{j=1}^n m_j + \sum_{j=1}^n \overleftarrow{m}_{n+1-j} + \sum_{j=1}^n m_j + \sum_{j=1}^n M_j \\ &= \sum_{\{jk_j=A\}} m_j + M_j + \sum_{\{jk_j=C\}} m_j + M_j + \\ &\quad \sum_{\{jk_j=G\}} m_j + M_j + \sum_{\{jk_j=T\}} m_j + M_j \\ &= 0 + n_C(n_A) + n_G(n_A + n_C) + \\ &\quad n_T(n_A + n_C + n_G) = l(K_0), \end{aligned}$$

donde, K_0 es la cadena descendente con n_i veces el nucleótido i . De esta forma se tiene el siguiente resultado:

Proposición 7. Con la notación anterior, si K es una cadena de nucleótidos, entonces

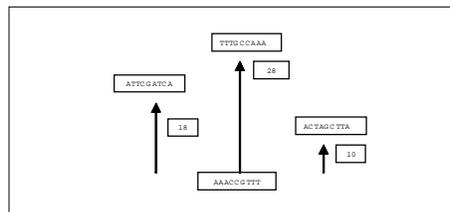
$$l(K) = l(\overleftarrow{K}) = l(K_0)$$

Ejemplo 12. Sea $K = ATTGCATCA$. Entonces $n_A = 3, n_C = 2, n_G = 1, n_T = 3$, y $n = 9$. La longitud de K es 18. La cadena opuesta es $\overleftarrow{K}_0 = ACTAGCTTA$. Notar que

$$m_8 + M_2 = 1 + 2 = n_A$$

La longitud de la cadena opuesta es 10 y la longitud de $K_0 = TTTGCCAAA$ es $l(K_0) = 3(6) + 1(5) + 2(3) = 28 = 18 + 10$

Lo observado anteriormente se puede representar gráficamente así:



Esta situación se presenta en permutaciones cuando están ordenadas por el orden de

Bruhat, (véase Macdonald, 1991) el cual está determinado básicamente por longitud.

Por ejemplo, en S_4 $w = 2431$, $w = 1342$, $w_0 = 4321$, se tiene

$$l(W_0) = 6 = 4 + 2 = l(W) = l(\overleftarrow{W}).$$

Además que en S_n $ww_0 = \overleftarrow{w}$.

Se puede interpretar la longitud de las cadenas de la siguiente manera: Siguiendo con el ejemplo 12, se forma inicialmente una cadena K_1 de longitud 0, donde $K_1 = AAACCGTTT$ y se define como cadena ascendente. La cadena $K = ATTCGATCA$ tiene los mismos nucleótidos que K_1 pero en distinto orden. ¿Cuál será el número mínimo de permutaciones de elementos consecutivos para transformar K_1 en K ? Si se enumeran de izquierda a derecha los nucleótidos A en K de 1 a 3, los nucleótidos C de 4 a 5 y al nucleótido G se le asigna el 6 y a las T se enumeran del 7 a 9, se obtiene entonces:

$$K = ATTCGATCA = 178462953$$

$$\text{y } K_1 = AAACCGTTT = 123456789$$

y con estas dos representaciones como permutaciones en S_9 , sus longitudes son $l(K_1) = 0$ y $l(K) = 18$, es decir, conservan la longitud y el número mínimo de permutaciones de elementos consecutivos para ir de K_1 a K o viceversa que son 18. Estas 18 permutaciones son:

178462953 à 178426953 à 178246953 à
 172846953 à 127846953 à 127846935 à
 127846395 à 127843695 à 127834695 à
 127384695 à 123784695 à 123748695 à
 123478695 à 123478695 à 123478569 à
 123475869 à 123457869 à 123457689 à
 123456789

Este tipo de relaciones son las que se observan en las permutaciones de S_n . Sin em-

bargo, estas longitudes que se calcularon dependen de un orden escogido, el cual fue numerar los nucleótidos de una cierta clase de izquierda a derecha. ¿Qué pasa si se toma otro orden? Como es de esperarse la longitud ya no coincide.

Ejemplo 13. Con la misma $K = ATTCGATCA$ del ejemplo 12, si se toma otro orden se obtiene por ejemplo $K = ATTCGATCA = 179563842$ tenemos que la longitud es ahora $l(K) = 0 + 5 + 6 + 3 + 3 + 1 + 2 + 1 + 0 = 21$. El orden anterior es [1,3,2], [5,4], [6], [7,9,8]. La diferencia en la longitud se debe precisamente a los tres "saltos dados" en la enumeración de los nucleótidos y por lo tanto, la longitud es única salvo saltos en las enumeraciones, en la forma como se ha presentado.

El interés por estudiar la longitud de las cadenas se debe únicamente al hecho que si dos cadenas tienen la misma representación espacial, entonces sus longitudes son muy similares. De esta forma es posible definir una distancia entre cadenas.

Definición 9. Si K, Q son dos cadenas con el mismo número de nucleótidos de cada clase, $d(K, Q)$ representa el número mínimo de permutaciones elementales entre elementos consecutivos para transformar K en Q (o viceversa).

Proposición 8. La función d satisface las siguientes propiedades:

1. $d(K, Q) = 0$ si y solamente si $K = Q$.
2. $d(K, Q) = d(Q, K)$
3. $d(K, Q) = d(K, T) + d(T, Q)$ para todo T que tenga los mismos nucleótidos que K y Q .

Por lo tanto d es una distancia.

Aun cuando la siguiente función no es distancia, es de interés para el estudio.

Se define $d_1(K, Q) = |l(K) - l(Q)|$. Hasta el momento se tiene que si $\eta(K) = \eta(Q)$ entonces tanto $d(K, Q)$ como $d_1(K, Q)$ son pequeños.

CONCLUSIONES

Se han definido sobre cadenas de nucleótidos algunos parámetros como longitud y distancia, las que se comportan como se suelen comportar la longitud y la distancia en permutaciones o puntos en el espacio. Con estas nociones se ha logrado determinar que la imagen inversa del número nula de una cadena está formada por cadenas que tienen longitudes similares, es decir, que están relativamente cerca unas de otra. Se desea determinar en próximos trabajos si estas dos nociones pueden determinar completamente la imagen inversa del número nula. Es de particular interés encontrar cotas máximas para las cantidades $d(K, Q)$ y $d_1(K, Q)$ si las cadenas K, Q satisfacen $\eta(K) = \eta(Q)$.

AGRADECIMIENTOS

Especial mención de gratitud para L. Lareo, por enriquecedoras reuniones para discutir

estas y otras muchas ideas sobre el número Nula. Igualmente, para Martha Alvarado por la cooperación y soporte que el departamento nos ha brindado durante este año.

LITERATURA CITADA

- ALVARADO M., NOVOA F., y RODRÍGUEZ L. 2003. Una nota sobre el espacio Nula. *Universitas Scientiarum*, vol. 8 Investigaciones Matemáticas, 7-12.
- LAREO L. and ACEVEDO O. 1999. Sequence mapping in three dimensions space by a numeric method and some of its implications. *Acta Biotheoretica* (47) 123-128.
- MACDONALD I.G. 1991. *Notes on Schubert Polynomials*. Publications du Laboratoire de Combinatoire et d'Informatique Mathématique No. 6. Département de mathématiques et d'informatique, Université du Québec à Montréal.

Recibido: 11-11-2003

Aceptado: 12-08-2004