

# EPISTASIS SUPERIOR

por

**Vicente Suárez Peñaranda**

Memoria realizada bajo la dirección de la profesora Dra. D<sup>a</sup> M<sup>a</sup> Teresa Iglesias Otero y el profesor Dr. D. Alain Verschoren, para obtener el grado de Doctor por la Universidade da Coruña.

A Coruña, noviembre de 2005



A Ana, María y Martín



# Agradecimientos

En primer lugar debo, y deseo, mostrar mi agradecimiento a la profesora Dra. D<sup>a</sup> M<sup>a</sup> Teresa Iglesias Otero, directora de esta memoria, por animarme a reanudar mi trabajo de investigación. En todo momento ha estado dispuesta a prestarme su atención, a aconsejarme y a enmendar mis errores. Al mismo tiempo, he contado con su ánimo, siempre acompañado de una vitalidad aparentemente inagotable. También quiero agradecer la ayuda del profesor Dr. D. Alain Verschoren, codirector de este trabajo. Él ha sido un faro por el que guiarme en esta larga travesía. Asimismo debo agradecer la colaboración de la profesora Dra. D<sup>a</sup> Concepción Vidal Martín. En el campo informático ha sido muy valiosa la ayuda de Joaquín Argibay Villanueva, becario del Departamento de Matemáticas, y de Carlos Díaz Pérez, técnico informático de la Escuela Universitaria Politécnica de Ferrol. Amplío los agradecimientos para alcanzar a todos mis compañeros del Departamento de *Matemáticas* de la Universidade da Coruña, en especial a los más cercanos, Marcelino Ladra González, Ignacio López de Ullibarri Galparsoro y Araceli Torres Miño.

Al Departamento de *Electrónica y Sistemas* de la Universidade da Coruña, responsable del programa de Doctorado: “Arquitecturas e Tecnoloxías Multimedia”

Con todo mi cariño, deseo dar las gracias, además de dedicarles esta memoria, a sus damnificados, que han soportado los malos ratos que el cansancio provoca, han notado mi ausencia en muchos momentos y, a pesar de ello, me han animado y confortado desde el primer al último momento, mi esposa Ana y mis hijos María y Martín.

Por último, señalar que este trabajo ha sido financiado, parcialmente, por los proyectos: *La epistasis generalizada y las funciones difíciles de optimizar* (PGIDIT02PXIA10501AF, Xunta de Galicia) y *La epistasis generalizada en la optimización funcional* (PGIDIT03PXIA10502PR, Xunta de Galicia), de los que su investigadora principal es la profesora Dra. D<sup>a</sup> M<sup>a</sup> Teresa Iglesias Otero.



# Índice General

<b>Introducción</b>	<b>1</b>
<b>1 Algoritmos Genéticos</b>	<b>13</b>
1.1 Aspectos generales sobre optimización . . . . .	13
1.1.1 Métodos clásicos de optimización . . . . .	14
1.1.2 Inteligencia Artificial . . . . .	15
1.2 Descripción general del Algoritmo Genético Clásico . . . . .	19
1.3 ¿Cómo actúan los Algoritmos Genéticos y por qué?	
Teorema de los esquemas . . . . .	25
1.4 Funciones decepcionantes. Funciones difíciles. Resultados sorprendentes . . .	28
<b>2 Epistasis en el caso binario</b>	<b>35</b>
2.1 Varianza epistática . . . . .	35
2.2 Descripción algebraica de la epistasis . . . . .	39
2.2.1 Epistasis normalizada . . . . .	39
2.2.2 Valores extremos de la epistasis normalizada . . . . .	41
2.2.3 Un ejemplo de cálculo directo de la epistasis normalizada $\epsilon^*$ . . . . .	43
2.3 Epistasis y transformadas de Walsh . . . . .	46
2.3.1 Coeficientes de partición y transformadas de Walsh . . . . .	46
2.3.2 Funciones Template: cálculo de su epistasis mediante los coeficientes de Walsh . . . . .	51
2.4 Relación entre orden y dificultad . . . . .	53
2.5 Orden y decepción . . . . .	57
2.6 Coeficientes de Walsh de las funciones de orden $k$ . . . . .	63
2.6.1 Caracterización de las funciones de segundo orden en términos de sus coeficientes de Walsh . . . . .	63
2.6.2 Coeficientes de Walsh de las funciones de cualquier orden . . . . .	64

<b>3</b>	<b>2-epistasis sobre alfabetos binarios</b>	<b>73</b>
3.1	2-epistasis y su representación matricial . . . . .	74
3.1.1	2-epistasis global de una función . . . . .	74
3.1.2	Representación matricial . . . . .	84
3.2	2-epistasis normalizada de algunas funciones interesantes . . . . .	92
3.2.1	La función de Dirac . . . . .	92
3.2.2	Funciones Camello . . . . .	93
3.2.3	Funciones de Ponderación . . . . .	94
3.2.4	Funciones Royal Road generalizadas . . . . .	96
3.2.5	Funciones Template . . . . .	101
3.3	Invarianza de la 2-epistasis normalizada . . . . .	109
3.4	Valores extremos . . . . .	112
3.4.1	Valor mínimo . . . . .	113
3.4.2	Valor máximo . . . . .	114
<b>4</b>	<b>Epistasis superior</b>	<b>121</b>
4.1	Epistasis de orden superior . . . . .	121
4.1.1	Las matrices $\mathbf{G}_{\ell,k}$ . . . . .	121
4.1.2	Propiedades algebraicas . . . . .	125
4.1.3	$k$ -epistasis . . . . .	130
4.2	$k$ -epistasis normalizada de algunas funciones . . . . .	131
4.2.1	Función de Dirac . . . . .	131
4.2.2	Funciones Camello . . . . .	132
4.2.3	Funciones cuasi-camello . . . . .	134
4.2.4	Funciones de Ponderación . . . . .	135
4.2.5	Funciones Royal Road generalizadas . . . . .	138
4.2.6	Funciones Template . . . . .	139
4.3	Valores extremos de la $k$ -epistasis . . . . .	145
4.3.1	Valores extremos de la 0-epistasis . . . . .	146
4.3.2	Valor mínimo . . . . .	147
4.3.3	Valor máximo . . . . .	148
4.4	Relación con la epistasis en el sentido de Davidor . . . . .	154
<b>A</b>	<b>El Teorema de los Esquemas</b>	<b>159</b>
<b>B</b>	<b>Resultados auxiliares</b>	<b>163</b>
B.1	El producto tensor de matrices . . . . .	163
B.2	Resultados auxiliares del capítulo 2 . . . . .	164
B.3	Resultados auxiliares del capítulo 4 . . . . .	165

**C Resultados experimentales**

**173**

**Bibliografía**

**179**



# Introducción

---

La teoría de la evolución fue desarrollada de forma independiente por Charles Darwin (1809–1882) y Alfred Wallace (1823–1913), y presentada en 1858. Darwin es el autor del libro más famoso sobre el tema: *“Sobre el Origen de las Especies mediante la Selección Natural”*. La hipótesis que ambos científicos defendían es que pequeños cambios heredables en los seres vivos y la selección natural son los dos hechos que provocan el cambio en la Naturaleza y la evolución de las especies. Sin embargo, ninguno de los dos conocía la base de la herencia. Pensaban que los rasgos de un ser vivo eran como un “fluido” y que los “fluidos” de los dos padres se mezclaban en la descendencia.

Gregor Mendel (1822–1884), con sus trabajos sobre la transmisión de los caracteres de las plantas a través de sucesivas generaciones, descubrió los mecanismos de la herencia: los caracteres se heredan de forma discreta y se toman del padre o de la madre. A estos caracteres, que pueden tomar diferentes valores, se les llamó posteriormente genes y a los valores que toman, alelos.

Mendel, actualmente considerado el fundador de la Genética moderna, presentó sus resultados en 1865 ante la Sociedad de Ciencias Naturales de Brünn, que los publicó al año siguiente con el título *“Experimentos de hibridación de plantas”*. Las teorías de Mendel, quien trabajó en total aislamiento, se olvidaron y no se volvieron a redescubrir hasta principios del siglo XX.

En la misma época en que Darwin y Mendel publicaban sus descubrimientos, otros investigadores estudiaban la estructura del núcleo celular. Walther Flemming (1843–1905) observó estructuras en forma de cinta en el interior del núcleo, a las que denominó “cromatina”. De hecho, observó que tales cintas (en realidad los cromosomas) se dividen longitudinalmente en

dos mitades idénticas, es decir, apreció el comportamiento paralelo de la división cromosómica y la segregación de los “caracteres” de Mendel. No fue hasta 1900 cuando tres investigadores, Hugo de Vries (1848–1935), Carl Correns (1864–1933) y Hugo Tschermak (1871–1962), redescubrieron los mismos principios que Mendel y “fundieron” en un único marco explicativo general tanto la teoría celular como la genética mendeliana y la teoría darwinista de la evolución.

Con el paso de los años se descubrió que las células de cada especie viviente tienen un número fijo y característico de cromosomas, que la base molecular de los genes está en el ADN (ácido desoxirribonucleico), que los cromosomas están compuestos de ADN y que, por tanto, los genes están en los cromosomas.

La macromolécula de ADN está compuesta por bases púricas y pirimidínicas: adenina, citosina, guanina y timina. La secuencia de estas bases (*genotipo*) es única para cada ser vivo y es ella la que a su vez determina la secuencia de aminoácidos característica de las proteínas que constituyen cada organismo (*fenotipo*).

Todos estos hechos forman hoy en día la teoría del neodarwinismo, que afirma que la historia de la mayoría de la vida está causada por una serie de procesos que actúan en y dentro de las poblaciones: reproducción, mutación, competencia y selección. La evolución se puede definir entonces como los cambios en el conjunto genético de una población.

Son varios los mecanismos que provocan la evolución de una determinada población:

- **Mecanismos que disminuyen la variabilidad:**

**Selección natural:** los individuos menos aptos no llegarán a reproducirse y, por lo tanto, su patrimonio genético desaparecerá de la población.

**Deriva génica:** en poblaciones pequeñas pueden desaparecer variantes. El hecho de que un alelo sea más común en la población que otro, provocará que su proporción aumente en dicha población.

- **Mecanismos que aumentan la diversidad:**

**Mutación:** es una alteración del genotipo. Las mutaciones, que se producen con baja frecuencia, son totalmente aleatorias y constituyen un mecanismo fundamental de generación de variabilidad genética.

**Poliploidía:** mientras que las células somáticas poseen dos copias de cada cromosoma (diploides), las células reproductivas sólo tienen una (haploides). Si por accidente alguna célula reproductiva tuviera dos copias y se lograra combinar con otra célula

daría lugar a un ser vivo con varias copias de cada cromosoma (poliploide). La mayoría de las veces la poliploidía da lugar a individuos con algún defecto, pero en algunos casos se crean individuos viables.

**Recombinación:** cuando las dos células sexuales (una masculina y otra femenina) se combinan, los cromosomas de cada una también lo hacen, intercambiándose fragmentos de ADN, que a partir de ese momento pertenecerán a un cromosoma diferente.

**Flujo genético:** es el intercambio de material genético entre seres vivos de diferentes especies.

En resumen, la selección natural actúa sobre el fenotipo y suele disminuir la diversidad, haciendo que sobrevivan sólo los individuos más aptos. Los mecanismos que generan diversidad y que combinan características actúan habitualmente sobre el genotipo.

En la naturaleza, los individuos de una población compiten entre sí en la búsqueda de recursos tales como comida, agua o refugio. Incluso los miembros de una misma especie compiten a menudo en la búsqueda de una pareja. Aquellos individuos que tienen más éxito en sobrevivir y en atraer parejas, tienen mayor probabilidad de generar un gran número de descendientes. Por el contrario, individuos poco adaptados producirán un menor número de descendientes. Esto significa que los genes de los individuos mejor adaptados se propagarán en sucesivas generaciones más que los de los menos adaptados. La combinación de buenas características provenientes de diferentes ancestros puede a veces producir descendientes cuya adaptación es mucho mayor que la de cualquiera de sus ancestros (“superindividuos”). De esta manera, las especies evolucionan logrando unas características cada vez mejor adaptadas al entorno en que viven.

Así, la evolución puede ser considerada como un método de búsqueda entre una enorme cantidad de posibles soluciones. En Biología, esas soluciones son las posibles secuencias genéticas y la solución deseada, el individuo que mejor se adapta: un organismo capaz de sobrevivir y reproducirse en el ambiente en que vive. Visto de esta forma, los mecanismos de la evolución pueden inspirar métodos computacionales de búsqueda. Los criterios que hacen a un individuo óptimo son variables, por lo que la evolución “elige” en un conjunto de posibilidades que está en permanente cambio. La búsqueda de soluciones en un ambiente variable es precisamente lo requerido para programas de computación adaptativos. Además, la evolución es un método de búsqueda paralelo: no sólo trabaja sobre una especie, sino que evalúa y cambia millones de especies a la vez. Finalmente, diremos que las reglas que rigen la evolución parecen claras: las especies evolucionan mediante mecanismos aleatorios (mutación, recombinación y otros operadores) seguidos de la selección natural. Los individuos óptimos

tienden a reproducirse y sobrevivir, propagando así su material genético a las generaciones venideras.

Los Algoritmos Genéticos (abreviadamente AG) son métodos adaptativos basados en el proceso genético de los organismos vivos, que pueden usarse para resolver problemas de búsqueda y optimización. Imitando el proceso de la evolución, los Algoritmos Genéticos son capaces de ir creando soluciones para problemas del mundo real.

Un Algoritmo Genético consiste en una función matemática o una rutina de software que toma como entradas a los individuos y retorna como salidas cuáles de ellos deben generar descendencia para la nueva generación. Una definición bastante completa de un Algoritmo Genético es la propuesta por John Koza ([29]):

*“Es un algoritmo matemático altamente paralelo que transforma un conjunto de objetos matemáticos individuales con respecto al tiempo, usando operaciones modeladas de acuerdo al principio darwiniano de reproducción y supervivencia del más apto, y tras haberse presentado de forma natural una serie de operaciones genéticas de entre las que destaca la recombinación sexual. Cada uno de estos objetos matemáticos suele ser una cadena de caracteres (letras o números) de longitud fija que se ajusta al modelo de las cadenas de cromosomas, y se les asocia con una cierta función matemática que refleja su aptitud.”*

El considerado “padre” de los Algoritmos Genéticos, John Holland (1929–), se preguntaba cómo logra la naturaleza crear seres cada vez más perfectos. Lo curioso es que todo se lleva a cabo a base de interacciones locales entre individuos, y entre éstos y lo que les rodea. Desconocía la respuesta, pero tenía una cierta idea de cómo hallarla: tratando de hacer pequeños modelos de la naturaleza, que tuvieran alguna de sus características y viendo cómo funcionaban, para luego extrapolar sus conclusiones a la totalidad.

A principios de los años 60, en la Universidad de Michigan, encontró un ambiente intelectual fértil que permitió que sus ideas comenzaran a dar frutos. Fue además, leyendo un libro escrito por el matemático y biólogo evolucionista R. A. Fisher (1890–1962), titulado “*La teoría genética de la selección natural*”, como comenzó a descubrir los medios para llevar a cabo sus propósitos de comprensión de la naturaleza. De ese libro aprendió que la evolución es una forma de adaptación más potente que el simple aprendizaje y tomó la decisión de aplicar estas ideas para desarrollar programas bien adaptados para un fin determinado.

En esa universidad, Holland impartía un curso titulado “*Teoría de sistemas adaptativos*”. Dentro de este curso, y con una participación activa por parte de sus estudiantes, fue donde

surgieron las ideas que más tarde se convertirían en los AG. Cuando Holland comenzó a diseñar Algoritmos Genéticos, los objetivos de su investigación eran dos:

- imitar los procesos adaptativos de los sistemas naturales, y
- diseñar sistemas artificiales (normalmente programas) que imiten los mecanismos importantes de los sistemas naturales.

Unos 15 años más tarde, David E. Goldberg (1953–) conoció a Holland y se convirtió en alumno suyo. Goldberg, ingeniero industrial, fue uno de los primeros que trató de aplicar los Algoritmos Genéticos a problemas industriales. Aunque Holland trató de disuadirle, porque pensaba que el problema era excesivamente complicado como para aplicarle AG, Goldberg consiguió lo que quería, escribiendo un Algoritmo Genético en un ordenador personal Apple II. Éstas y otras aplicaciones creadas por estudiantes de Holland convirtieron a los Algoritmos Genéticos en un campo con base suficiente para celebrar el primer congreso sobre el tema en 1985: International Conference on Genetic Algorithms. Tal congreso se sigue celebrando bianualmente.

Para resolver un problema específico, los Algoritmos Genéticos usan una analogía directa con el comportamiento de la naturaleza. La entrada del AG es un conjunto de soluciones potenciales de ese problema, codificadas en vectores de símbolos (cadenas) y una función, llamada *fitness* o función de idoneidad, que permite evaluar cuantitativamente a cada candidato. En la naturaleza esto equivaldría al grado de efectividad de un organismo para competir por unos determinados recursos. Luego el AG evalúa cada posible solución de acuerdo con la función de idoneidad. Cuanto mayor sea la idoneidad de una cadena, mayor será la probabilidad de que sea seleccionada para reproducirse, cruzando su material genético con otra cadena seleccionada. Esta recombinación genética producirá nuevos individuos descendientes de los anteriores, que comparten algunas características con sus padres. Los individuos que no han mejorado con los cambios en su genotipo son eliminados y los que han mejorado se seleccionan para formar una nueva generación.

De esta manera se produce una nueva población de posibles soluciones, que reemplaza a la anterior y verifica la interesante propiedad de que contiene una proporción mayor de buenas características que la población anterior. Así, a lo largo de las generaciones, las buenas características se propagan a través de la población. Favoreciendo el cruce de los individuos mejor adaptados van siendo exploradas las áreas más prometedoras del espacio de búsqueda.

Los principios básicos de los Algoritmos Genéticos fueron establecidos por Holland [17] y se encuentran bien descritos en varios textos, por ejemplo, Goldberg [13], Davis [5] o Michalewicz [30].

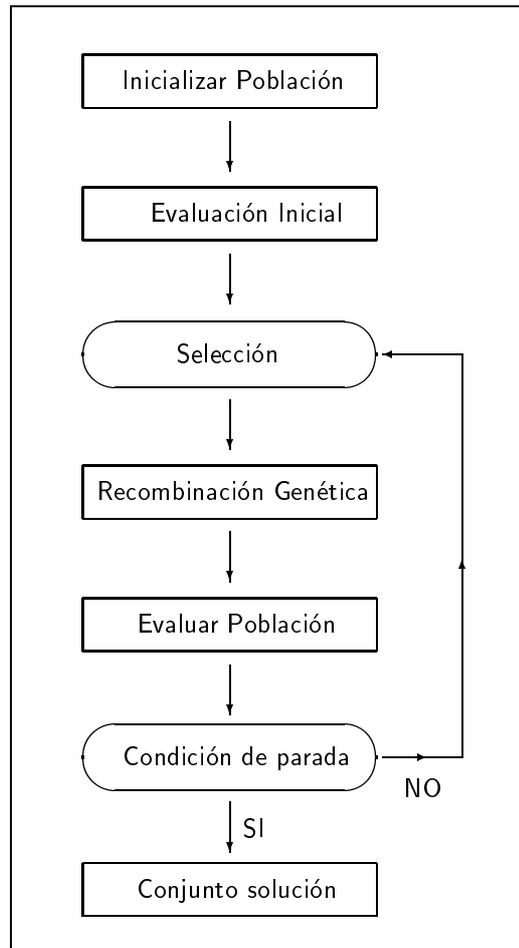


Figura 1: Esquema de un Algoritmo Genético.

Las principales propiedades que caracterizan a los AG son:

- No necesitan conocimientos específicos sobre el problema que intentan resolver.
- Operan de forma simultánea con varias soluciones, en vez de trabajar de forma secuencial como las técnicas tradicionales.
- Cuando se usan para problemas de optimización (maximizar una función objetivo) resultan menos afectados por los máximos locales (falsas soluciones) que las técnicas tradicionales.
- Resulta sumamente fácil ejecutarlos con las modernas arquitecturas masivamente paralelas.
- Usan operadores probabilísticos en vez de los típicos operadores determinísticos de las otras técnicas.

- Pueden tardar mucho en converger o no converger en absoluto, dependiendo en cierta medida de los parámetros que se utilicen: tamaño de la población, número de generaciones, etc.
- Pueden converger prematuramente.

El poder de los Algoritmos Genéticos proviene del hecho de que constituye una técnica robusta y que puede tratar con éxito una gran variedad de problemas provenientes de diferentes áreas, incluyendo aquellos en los que otros métodos encuentran dificultades. Si bien no se garantiza que un Algoritmo Genético encuentre la solución óptima de un problema, existe evidencia empírica de que se encuentran soluciones de un nivel aceptable en un tiempo competitivo con el resto de algoritmos de optimización.

El gran campo de aplicación de los Algoritmos Genéticos se halla en aquellos problemas para los que no existen técnicas especializadas. Incluso en el caso en que dichas técnicas existan y funcionen bien, pueden efectuarse mejoras de las mismas hibridándolas con los AG.

La aplicación más común de los Algoritmos Genéticos ha consistido en la solución de problemas de optimización, en donde han mostrado ser muy eficientes y fiables. Sin embargo, no todos los problemas son apropiados para esta técnica y, en general, se recomienda tener en cuenta las siguientes características del problema antes de intentar su aplicación:

- Su espacio de búsqueda (i.e., sus posibles soluciones) debe estar delimitado dentro de un cierto rango.
- Debe poderse definir una función de aptitud que nos indique cuan buena o mala es una cierta respuesta.
- Las soluciones deben codificarse de una forma que resulte relativamente fácil de implementar en la computadora.

El primer punto es muy importante y lo más recomendable es intentar resolver problemas que tengan espacios de búsqueda discretos, aunque éstos sean muy grandes.

La función de aptitud no es más que la función objetivo de nuestro problema de optimización. El Algoritmo Genético se diseña habitualmente para maximizar, pero la minimización puede realizarse fácilmente utilizando la opuesta de la función maximizante. Una característica que debe tener esta función es que tiene que ser capaz de “castigar” a las malas soluciones y “premiar” a las buenas, de forma que sean estas últimas las que se propaguen con mayor rapidez.

La codificación más común de las soluciones es la que se realiza a través de cadenas binarias, aunque se han utilizado también números reales y letras. El primero de estos esquemas es el que se propuso originalmente, por ser muy sencillo su procesamiento en un ordenador.

A pesar de la gran aplicabilidad que han demostrado tener los AG en la optimización funcional, existen ocasiones en las que esta tarea resulta extremadamente difícil. Cuando un AG no puede encontrar el óptimo de una función en una cantidad razonable de tiempo, es decir, cuando el número de generaciones que el AG necesita crear hasta verificarse la condición de parada es alto, se dice que la función es “difícil” para el algoritmo. No existe un criterio definitivo que permita caracterizar a los “problemas difíciles” para un Algoritmo Genético, aunque sí se han determinado algunos de los factores que provocan esta “dificultad”.

La primera explicación sobre la “dificultad” condujo al concepto de *engaño* (*deception*) –al que nosotros nos referiremos como *decepción*–, que tiene que ver con la existencia de datos engañosos que conducen al algoritmo lejos del óptimo. Ejemplos de funciones difíciles, pero no decepcionantes, como las funciones Royal Road (que se introducen en la sección 1.4) y de funciones decepcionantes fáciles para el algoritmo, llevaron a los investigadores a concluir que la existencia de decepción no es necesaria ni suficiente para justificar por sí sola la dificultad de una función. Otra de las causas de la dificultad de optimizar un función es la *epistasis*, un concepto procedente de la Genética, que tiene que ver con la existencia de interacciones entre los alelos de las distintas posiciones de las cadenas que codifican los individuos de la población. En Genética, se dice que un gen es epistático si enmascara el efecto fenotípico de otro gen.

En esta memoria se profundiza en el análisis del concepto de epistasis. Se muestra como –conjuntamente con el orden de una función (otra característica a tener en cuenta)– la epistasis permite explicar el comportamiento de, entre otras, dos clases de funciones, cuyos extremos son conocidos (y su dificultad también) y que habitualmente sirven como “funciones de laboratorio” en el estudio de la dificultad.

Esta adecuada correlación en funciones de orden bajo parece resultar incompleta como explicación en el caso de funciones de órdenes altos. Ello nos ha llevado a generalizar la noción de epistasis, mostrando cómo esta generalización complementa la información que proporciona la epistasis “clásica”. La complejidad de la formulación matemática que conlleva este estudio hace necesario el uso de la herramienta algebraica junto con las transformadas de Walsh, que nos han permitido caracterizar distintas propiedades, de las funciones de idoneidad, de forma sencilla y elegante. El desarrollo de este estudio en esta memoria se estructura de la forma siguiente: el primer capítulo se dedica a la introducción de los Algoritmos Genéticos, la descripción de su funcionamiento y una comparativa con los métodos clásicos de optimización.

Se introducen los conceptos de *esquema*, *orden* y *longitud de definición* de un esquema y se relacionan con el proceso de selección y los operadores genéticos, conduciendo al teorema fundamental de los AG: el *Teorema de los Esquemas*, establecido por Holland en 1975 ([17]), cuya demostración constituye el apéndice A.

El capítulo 2 comienza con un resumen de los resultados ya conocidos sobre epistasis, entendida como estimador de la dificultad. En el campo de los AG, la idea inicial de epistasis la introduce de forma intuitiva Rawlins en [35]. Rawlins observa dos comportamientos extremos de las funciones de idoneidad: aquellas para las que el valor de una cadena se obtiene como combinación lineal de funciones cada una de las cuales depende de un sólo bit (epistasis mínima) y el caso contrario en el que ningún subconjunto de bits es independiente de los demás en el cálculo de la idoneidad de una cadena (epistasis máxima).

Davidor ([4]) formula matemáticamente la definición de epistasis y, posteriormente, Van Hove ([41]) la reescribe de forma más compacta. Por último, el hecho de que para cualquier número real no nulo  $\lambda$ , las funciones  $f$  y  $\lambda f$  deban tener la misma epistasis, conduce a la normalización del concepto, definiéndose la epistasis normalizada de una función ([39]). Además, el estudio algebraico de los valores extremos de este indicador, muestran que la epistasis normalizada es mínima cuando la función  $f$  es de orden 1, es decir  $f = \sum_i g_i$ , siendo  $g_i$  una función que depende únicamente del  $i$ -ésimo bit. Éstas son las funciones lineales, sencillas de optimizar por los AG. Para funciones no negativas, se comprueba que el valor máximo se alcanza para un tipo de funciones cuyo valor sobre una cadena depende de todos y cada uno de los bits de la misma.

El cálculo práctico de la epistasis normalizada resulta en ocasiones muy complicado. Esta ardua tarea se simplifica con la aplicación de las transformadas de Walsh. Éstas pueden considerarse una versión binaria de la transformada discreta de Fourier, y han demostrado una gran eficiencia en el cálculo de este estimador. Los llamados *coeficientes de Walsh* caracterizan completamente a cualquier función, y su cálculo se realiza de una forma cómoda mediante el *Teorema del Valor Medio para Hiperplanos*. Para resaltar la simplificación alcanzada en el cálculo de la epistasis normalizada con el uso de este teorema, frente al cálculo directo de la epistasis a partir de la definición, incluimos en el apartado 2.2.3 el desarrollo realizado en [20] para calcular la epistasis normalizada de las funciones *Template* a partir de la definición y en el 2.3.2 el que realizamos nosotros con el uso del Teorema del Valor Medio para Hiperplanos. El contenido de esta subsección es parte de una comunicación presentada en el congreso *Encontro de Álgebra Linear e Aplicações*, celebrado en Lisboa en septiembre de 2003 (ver [21]).

En los apartados 2.4 y 2.5 establecemos conclusiones acerca de otro de los estimadores de la dificultad: el orden de la función de idoneidad. Este último apartado constituye el

contenido de la publicación *Order and Deceptivity* (remitido para su publicación a Bulletin of the Belgian Mathematical Society Simon Stevin).

El capítulo 2 concluye con la caracterización del orden de una función mediante sus coeficientes de Walsh, problema que estaba completamente resuelto para funciones de orden uno y parcialmente para las de orden dos. Gran parte de la sección 2.6 constituye el contenido de la publicación *Higher order functions and Walsh coefficients* (aceptada en: Bulletin of the Belgian Mathematical Society Simon Stevin).

En el capítulo 3 se introduce un nuevo estimador de la dificultad de una función, la *2-epistasis normalizada*. Siguiendo las ideas de Davidor que condujeron a la epistasis normalizada, nosotros proponemos establecer una medida de distancia entre cualquier función de idoneidad y el espacio de funciones de orden 2. Cabe preguntarse por qué introducir este estimador cuando la epistasis normalizada proporciona una buena caracterización de las funciones difíciles. La respuesta a esta pregunta se encuentra en la sección 2.4, en la que se compara la ejecución de un Algoritmo Genético sobre dos funciones de laboratorio: las funciones *Template* y las funciones *Royal Road generalizadas*, definidas sobre cadenas de longitud 64. La epistasis normalizada crece a la vez que lo hace el número de generaciones necesarias para la convergencia del algoritmo. Sin embargo, para valores altos de la epistasis normalizada, resulta sorprendente la gran diferencia de generaciones alcanzada para estas dos funciones. Por el contrario, los valores de la epistasis normalizada no reflejan esa significativa diferencia en el comportamiento de ambas funciones. Parece que la comparación de una función de ajuste con las funciones de orden 1 podría ser demasiado fuerte. Cabe plantearse si la epistasis normalizada es o no capaz de distinguir bien entre los diferentes “grados” de dificultad que presentan dos funciones cuando los valores de este estimador son altos. La generalización de este concepto, introducido en el capítulo 3, ayuda a entender mejor estas cuestiones.

En la subsección 3.4.1 se relaciona satisfactoriamente la 2-epistasis normalizada con las funciones de segundo orden: son precisamente aquellas funciones que alcanzan el valor mínimo (cero) para este estimador. Recordemos que la epistasis normalizada clásica verifica que se anula si y sólo si se calcula sobre las funciones de primer orden. Asimismo, el cálculo de la 2-epistasis normalizada para algunas funciones interesantes, en particular las funciones *Template* y las *Royal Road generalizadas*, permite entender un poco mejor el diferente comportamiento del Algoritmo Genético frente a estas funciones.

Gran parte del contenido de este tercer capítulo constituye la publicación *2-epistasis over binary alphabets* (remitido para su publicación a Computers and Informatics). Asimismo, los resultados expuestos en la subsección 3.2.4 con las funciones *Royal Road generalizadas*

completan los ya obtenidos en *Computing epistasis of Generalized Royal Road functions using Walsh transforms and the Hyperplane Averaging Theorem*, comunicación presentada en el congreso *International Symposium on Symbolic and Algebraic Computation*, celebrado en Santander en julio de 2004 (ver [22]).

En el último capítulo se introduce el concepto de epistasis de orden superior, desde un punto de vista puramente algebraico. La formulación matemática del concepto de epistasis normalizada desarrollado por Van Hove y la obtenida en el capítulo 2 para la 2-epistasis, permiten realizar un desarrollo similar para cualquier valor de  $k$ ,  $0 \leq k \leq \ell$ . Se introduce el concepto de *k-epistasis normalizada* y se calcula su valor para las funciones consideradas en el capítulo 3. En el último apartado de este capítulo se relaciona el entramado algebraico desarrollado con las ideas iniciales de epistasis descritas por Davidor.

Como ya se ha comentado al principio de la introducción, en el apéndice A se detalla la demostración del Teorema de los Esquemas.

En el apéndice B se exponen algunos resultados utilizados en los capítulos 2 y 4. Aunque podrían encontrarse en textos de Combinatoria, dado que sus demostraciones no son inmediatas, hemos considerado conveniente incluirlos en un apéndice. Los hemos enunciado como lemas, pues constituyen resultados intermedios en las demostraciones de algunas de las proposiciones de esos dos capítulos.

El último apéndice contiene un resumen de resultados experimentales. Comprobamos la valía de los estimadores definidos a lo largo de este trabajo. La conclusión obtenida es que la *k-epistasis normalizada*, definida en el capítulo 4, es un buen complemento para los estimadores de la dificultad de una función hasta ahora conocidos.



# Capítulo 1

## Algoritmos Genéticos

---

*“La vehemente búsqueda de la perfección por parte del hombre se expresa en la teoría de la optimización. Ésta estudia cómo describir y conseguir lo que es mejor, una vez se conoce cómo medir y alterar lo que es bueno o malo... La teoría de la optimización abarca el estudio cuantitativo de los óptimos y los métodos para encontrarlos.”*

BEIGHTLER, C.S. ET AL ([1])

### 1.1 Aspectos generales sobre optimización

La optimización funcional pretende la obtención y mejora de métodos que permiten llegar a un punto o puntos óptimos. Existe una clara distinción entre el proceso de acercamiento al óptimo y la obtención del mismo. De hecho, cuando se juzgan procesos de optimización, se insiste en la convergencia del método, y se tiende a olvidar la ejecución intermedia. Este énfasis, que procede de los orígenes del cálculo, es un hecho natural.

### 1.1.1 Métodos clásicos de optimización

Dentro del campo de la optimización funcional, los métodos clásicos se pueden agrupar en los siguientes tipos:

#### 1. Métodos analíticos

Los métodos basados en el cálculo se subdividen a su vez en indirectos e directos. Los métodos indirectos realizan la búsqueda de extremos locales mediante la resolución de sistemas de ecuaciones, generalmente no lineales, resultantes de igualar a cero el gradiente de la función a optimizar. Por otro lado, los métodos directos buscan óptimos locales moviéndose sobre el espacio de búsqueda en la dirección que marca el máximo valor del gradiente local. Éste es el comportamiento del método analítico directo más sencillo, denominado *escalamiento de la colina* (*hillclimbing*), que ante la búsqueda de un máximo local en una superficie escalará a través de la misma hasta que ya no obtenga mejora.

Los métodos analíticos carecen de robustez, pues, en primer lugar, se necesita conocer explícitamente la expresión de la función a optimizar, obteniéndose además, en muchos casos, óptimos locales. Los indirectos necesitan para su aplicación algunas condiciones sobre la función a optimizar, pueden quedar atrapados en el óptimo local más cercano a la solución inicial y, además, presentan un mal comportamiento en zonas donde la función es casi constante. Por último, no son válidos para problemas multimodales, en los cuales la función a optimizar tiene varios óptimos.

#### 2. Métodos exhaustivos o enumerativos

La técnica que emplean estos algoritmos es bastante clara: recorren todo el espacio de búsqueda y evalúan la función objetivo en cada punto del espacio, quedándose con la mejor solución. Aunque la simplicidad de este algoritmo lo hace atractivo, sufren de una gran carencia –la eficiencia– ya que, en general, los espacios de trabajo suelen resultar demasiado grandes para realizar esta búsqueda elemento a elemento.

Por ejemplo, algunos de los problemas a los que no se le puede aplicar este tipo de algoritmos son los llamados *NP-completos*, que son un subconjunto de problemas especialmente difíciles entre aquellos cuya complejidad crece exponencialmente con el tamaño del problema. Uno de ellos es el *Problema del Viajante* (*Travelling Salesman Problem*), en el cual un viajante debe recorrer un número determinado de ciudades y volver al punto de partida minimizando la distancia recorrida.

### 3. Métodos aleatorios

En los métodos aleatorios se va muestreando el espacio de búsqueda acotando las zonas que no han sido exploradas; se escoge la mejor solución y además se proporciona el intervalo de confianza de la solución encontrada.

Un algoritmo de este tipo es el conocido como *Templado Simulado* (*Simulated Annealing* o SA) (ver [28]), nombre que procede de la forma en que se consiguen ciertas aleaciones en forja: una vez fundido el metal, se va enfriando poco a poco, para conseguir finalmente la estructura cristalina correcta, que haga que la aleación sea dura y resistente.

El SA es una generalización del método de Montecarlo que ha demostrado ser útil en numerosos problemas de optimización combinatoria, y es particularmente eficaz en problemas de diseños de circuitos.

Este algoritmo se podría calificar como un escalador estocástico, y su principal objetivo es evitar los mínimos locales en los que suelen caer los métodos escaladores. Para ello, no siempre acepta la solución óptima, sino que a veces puede escoger una solución subóptima, siempre que ésta supere un nivel umbral prefijado, que depende de un parámetro denominado “temperatura”. La estrategia del método es comenzar con una temperatura inicial alta, lo cual proporciona una probabilidad también alta de aceptar un movimiento de no mejora. En cada iteración se va reduciendo la temperatura y por lo tanto la probabilidad de no mejorar es cada vez más pequeña conforme avanza el procedimiento y nos acercamos a la solución óptima. De este modo, inicialmente se realiza una diversificación de la búsqueda sin controlar demasiado el coste de las soluciones visitadas. En iteraciones posteriores resulta cada vez más difícil aceptar malos movimientos y, por lo tanto, se produce un descenso del coste.

#### 1.1.2 Inteligencia Artificial

Se llama *Inteligencia Artificial* (IA) a la rama científica que se dedica a la creación de sistemas que exhiben cierto poder de aprendizaje que imite, de algún modo, la inteligencia humana. Para ello, la IA no sólo consiste en idear algoritmos y estructuras de datos para resolver problemas, también trata acerca de la inteligencia humana y, por extensión, de la vida.

Dentro de la IA se encuentra un apartado denominado *Computación Natural*, que se caracteriza por simular las técnicas que emplean los seres vivos para evolucionar en la naturaleza. Algunos de los campos que abarca son:

#### Lógica Difusa

Tiene como objetivo crear un modelo aproximado al del raciocinio humano, pretendiendo desarrollar sistemas computacionales capaces de tomar decisiones racionales en un ambiente de incertidumbre e imprecisión. La Lógica Difusa ofrece un mecanismo para manipular informaciones imprecisas, tales como los conceptos de mucho, poco, alto, bueno, caliente, tibio, frío, ... suministrando una respuesta para una cuestión basada en un conocimiento que es inexacto, incompleto, o no totalmente fiable.

#### Sistemas Expertos

Es una rama de la Inteligencia Artificial que ejecuta, a partir de una base de conocimientos o de reglas, algunas tareas que normalmente requieren cierto nivel de experiencia por parte del ser humano.

Los expertos en un determinado campo solucionan los problemas utilizando una combinación de conocimientos basados en hechos constatados y en su capacidad de razonamiento. En los sistemas expertos, estos dos elementos básicos están contenidos en dos componentes separados, aunque relacionados: una base de conocimientos y una máquina de deducción, o de inferencia. La base de conocimientos proporciona hechos objetivos y reglas sobre el tema, mientras que la máquina de deducción proporciona la capacidad de razonamiento que permite al sistema experto extraer conclusiones.

#### Redes Neuronales Artificiales (RNA)

Son modelos computacionales inspirados en la estructura y modo de operación del cerebro humano, cuyo objetivo es conseguir que las máquinas reproduzcan características del hombre, tales como: aprendizaje, percepción, raciocinio, evolución y adaptación.

La programación informática basada en RNA parte de un conjunto de datos de entrada suficientemente significativo y tiene como objetivo conseguir que la red *aprenda* automáticamente las propiedades deseadas. En este sentido, el diseño de la red tiene que ver menos con flujos de datos y detección de condiciones, y más con cuestiones tales como selección del modelo de red, de las variables a incorporar y del procesamiento de la información que formará el *conjunto de entrenamiento*. Asimismo, el proceso de creación de la red no se denomina genéricamente programación, sino *entrenamiento*.

Por ejemplo, en una red que se va a aplicar al diagnóstico de imágenes médicas, la fase de entrenamiento consiste en la recepción de imágenes de tejidos que se saben cancerígenos y otros que se saben sanos, así como las respectivas clasificaciones de dichas imágenes. Una vez concluido el entrenamiento y si éste es adecuado, el sistema prodrá clasificar imágenes de tejidos (*sano/no sano*) con un buen grado de fiabilidad.

### Computación Evolutiva

En la naturaleza todos los seres vivos se enfrentan a problemas que deben resolver con éxito, como conseguir el calor de la luz del sol o alimentarse con los productos de su entorno. La Computación Evolutiva (CE) interpreta la naturaleza como una inmensa máquina de resolver problemas y trata de encontrar el origen de dicha potencialidad para utilizarla en programas informáticos.

La Computación Evolutiva retoma los conceptos de la Evolución y la Genética para resolver principalmente problemas de optimización. La principal aportación de la CE a la metodología de resolución de problemas consiste en el uso de mecanismos de selección de soluciones potenciales y de construcción de nuevos candidatos por recombinación de características de otros ya presentes, de modo parecido a como ocurre en la evolución de los organismos naturales. La CE no trata tanto de reproducir ciertos fenómenos que suceden en la naturaleza, como de aprovechar las ideas genéricas que hay detrás de ellos.

Esta rama de la IA tiene sus raíces en tres técnicas relacionadas pero independientes entre sí: *Algoritmos Genéticos*, *Programación Evolutiva* y *Estrategias Evolutivas*.

- Algoritmos Genéticos

Son algoritmos que hacen evolucionar a una población someténdola a acciones aleatorias semejantes a las que actúan en la evolución biológica (recombinación genética); así como también a una selección de acuerdo a un criterio previamente establecido, en función del cual se decide cuáles son los individuos mejor adaptados, que sobreviven, y los menos aptos que, en general, son descartados.

- Programación Evolutiva

Es una técnica en la cual los candidatos a ser solución para una tarea determinada se representan por máquinas de estados finitos, cuyos diagramas de estados de transición evolucionan mediante mutación aleatoria.

- Estrategias Evolutivas

Trabajan con una población de individuos que pertenecen al dominio de los números reales, que mediante procesos de recombinación, evolucionan para alcanzar el óptimo de la función objetivo.

Cada individuo de la población es un posible óptimo de la función objetivo. La representación de cada individuo de la población consta de dos tipos de variables: las variables objeto y las variables estratégicas. Las variables objeto son los posibles valores que hacen que la función objetivo alcance el óptimo global y las variables estratégicas son los parámetros mediante los que se gobierna el proceso evolutivo.

### Vida artificial

Puede ser definida como el intento de crear vida, o algo parecido a la vida, mediante la combinación de símbolos (datos) y procesos de símbolos (programas) independientemente del soporte físico de estos símbolos y procesos.

Por una parte están los intentos “hardware” de emulación de vida. Por ejemplo, es posible construir un pequeño robot con aspecto de ratón capaz de encontrar la salida de un laberinto.

Por otra parte están las simulaciones “software”, que tienen la ventaja de permitir la construcción de un gran número de seres vivos y entornos en los que éstos existen, de manera que es más fácil estudiar comportamientos sociales. Podemos construir los seres artificiales con el objetivo de solucionar los problemas de interés y que aprendan o colaboren entre ellos hasta conseguir el resultado deseado.

Nosotros nos centraremos, a partir de ahora, en los Algoritmos Genéticos. Como ya se ha comentado anteriormente, son una herramienta matemática inspirada en los mecanismos de la evolución natural y recombinación genética que se vienen empleando con éxito en campos tan diversos como la Economía (evolución del mercado bursátil), la Medicina (análisis de electrocardiogramas), la Ingeniería Aeronáutica (diseño de turbinas de aviones) y, por supuesto, en Matemáticas en la optimización funcional, nuestro campo de trabajo.

## 1.2 Descripción general del Algoritmo Genético Clásico

John Holland, matemático de la Universidad de Michigan, fue quien, durante las décadas de 1960 y 1970, sentó las bases actuales de los Algoritmos Genéticos. En contraste con las estrategias evolutivas y la programación evolutiva, el propósito original de Holland no era diseñar algoritmos para resolver problemas concretos, sino estudiar, de un modo formal, el fenómeno de la adaptación, tal y como ocurre en la naturaleza, y desarrollar vías para extrapolar estos mecanismos de adaptación natural a los sistemas computacionales.

La técnica de los AG ofrece un mecanismo de búsqueda adaptativa basada en el principio Darwiniano de reproducción y supervivencia de las especies. De acuerdo con la teoría de Darwin, el principio de selección privilegia a los individuos más aptos con una mayor longevidad y, por lo tanto, con mayor probabilidad de reproducción. Los individuos con más descendientes tienen más oportunidades de transmitir sus códigos genéticos a las generaciones venideras.

En [17], Holland presenta el Algoritmo Genético como una abstracción de la evolución biológica, y proporciona su entramado teórico. El algoritmo de Holland es un método para desplazarse de una población de individuos a una nueva población, utilizando un sistema similar a la “selección natural”, junto con operadores de cruces y mutaciones inspirados en la Genética. En esa monografía, publicada en 1975, se sientan las bases teóricas que fundamentan el desarrollo, desde el punto de vista computacional, de la teoría de los Algoritmos Genéticos; abstrayendo los conceptos de la genética natural y aplicándolos a la Economía, al Reconocimiento de Patrones y a la Teoría de Juegos.

Frente a las limitaciones de los métodos clásicos de optimización, los Algoritmos Genéticos pueden ser usados para la resolución de cualquier tipo de problema, sin ninguna exigencia previa respecto de la función a optimizar, ni al espacio de posibles soluciones.

A pesar de que existe la evidencia empírica de que los AG encuentran la solución óptima de muchos problemas, no son una panacea. Existen ocasiones en que fracasan, debido a diversos factores.

Con todas estas consideraciones, se puede decir que estos algoritmos son una buena herramienta de búsqueda y optimización donde los métodos de cálculo tradicional no son viables. En definitiva, son una herramienta complementaria a las ya existentes y no una técnica sustitutoria de las mismas.

Se puede caracterizar un AG mediante las siguientes componentes:

### 1. Representación de las soluciones del problema

Los AG no trabajan directamente con las posibles soluciones del problema (*individuos*), sino que emplean como elementos de la población (*cromosomas*) codificaciones de las mismas. Estas codificaciones son vectores de longitud fija (*cadena*). Cada una de las posiciones de la cadena se conoce como *gen*. Cualquier carácter de un individuo puede tener diferentes manifestaciones, es decir, el gen puede tomar diferentes valores (*alelos*) y ocupará una posición determinada de la cadena (*locus*).

Si denotamos por  $\mathcal{D}$  al conjunto de elementos del espacio de búsqueda y por  $\Sigma$  al alfabeto utilizado (conjunto de todos los alelos), la codificación genética de los individuos se consigue con una aplicación inyectiva  $C : \mathcal{D} \hookrightarrow \Sigma^\ell$  que representa, de forma unívoca, cada elemento de  $\mathcal{D}$ .

En los Algoritmos Genéticos clásicos los cromosomas son cadenas binarias, por ser fáciles de manipular a través de operadores genéticos y fácilmente transformables en un número entero o real. A lo largo de esta memoria trabajaremos con este tipo de codificación.

Por ejemplo, si se desea obtener un valor óptimo para una función  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$  en el dominio  $(x, y) \in [-3.0, 12.1] \times [4.1, 5.8]$  y se supone que la precisión requerida para la solución es de 4 dígitos en la parte decimal de cada variable, entonces, una codificación binaria exigirá que el dominio de la primera variable, que tiene longitud 15.1, sea dividido en al menos  $15.1 \times 10000$  intervalos de igual longitud. Dado que  $2^{17} < 151000 < 2^{18}$ , se necesitarán 18 bits para la primera parte del cromosoma. Por una razón análoga, se necesitarán 15 bits para la segunda parte (pues  $2^{14} < 17000 < 2^{15}$ ). La longitud total del cromosoma es 33. La decodificación de un cromosoma, por ejemplo el (010001001011010000111110010100010) se realiza como sigue:

$$x = -3.0 + decimal(010001001011010000_2) \times \frac{12.1 - (-3.0)}{2^{18} - 1} = 1.052426,$$

$$y = 4.1 + decimal(111110010100010_2) \times \frac{5.8 - (4.1)}{2^{15} - 1} = 5.755330.$$

### 2. Inicialización de la población

Se forma una población inicial  $\mathcal{P}$ , con individuos elegidos, generalmente, de forma aleatoria entre todos los elementos del espacio de búsqueda.

### 3. Evaluación

Se realiza a través de una *función de ajuste o idoneidad (fitness function)*, que represente de forma adecuada el problema y tenga como objetivo suministrar una medida de

aptitud de cada posible solución, es decir, mide la calidad de cada individuo del espacio de búsqueda. La función de ajuste se define sobre las codificaciones de las posibles soluciones del problema, y no sobre el espacio de éstas.

#### 4. Selección

Este proceso elige una parte de la población  $\mathcal{P}$  para la recombinación. Esta selección se basa, de nuevo, en la calidad de los individuos: los individuos más aptos tienen mayor probabilidad de ser escogidos para la siguiente etapa.

Algunos de los criterios de selección son: el *Torneo*, la *Regla de la Ruleta* y el *Ranking Lineal*.

- **Torneo:** Se baraja la población y se hace competir a los individuos que la integran en grupos de tamaño predefinido. Si la competición se realiza por parejas, entonces la población se debe barajar dos veces. Nótese que esta técnica garantiza la obtención de múltiples copias del mejor individuo en la próxima generación.
- **Regla de la Ruleta (Selección proporcional):** Intuitivamente el proceso construye una ruleta o diagrama de sectores en el que cada sector representa un individuo. El sector que corresponde a cada individuo es proporcional a su valor de ajuste. A continuación, y al igual que en un casino, se hace girar la ruleta y se elige el individuo al que corresponde el sector en el que se ha detenido la bola.
- **Ranking Lineal:** La población se ordena de acuerdo con sus valores en la función objetivo. El nivel de ajuste asignado a cada individuo depende de su posición en el rango y no de su valor en la función objetivo. Con este criterio de selección se evita una convergencia prematura hacia un individuo no óptimo con valor de ajuste alto, ya que el rango de reproducción es limitado: ningún individuo genera un número excesivo de hijos.

#### 5. Recombinación

A los individuos elegidos se les aplican los operadores genéticos para crear descendientes y construir una nueva y mejor población (que constituirá una nueva *generación*). Los operadores genéticos más usuales son:

- **Cruce:** recombina dos individuos de la población inicial (los padres) para formar nuevos individuos (los descendientes). Este proceso se aplica con una cierta probabilidad fija llamada *probabilidad de cruce*.

La forma tradicional de cruce es la de *cruce basado en un punto*, en el que se selecciona un punto entre dos posiciones previamente determinadas, se rompen las cadenas y se intercambian las partes, como se refleja en la figura 1.1.

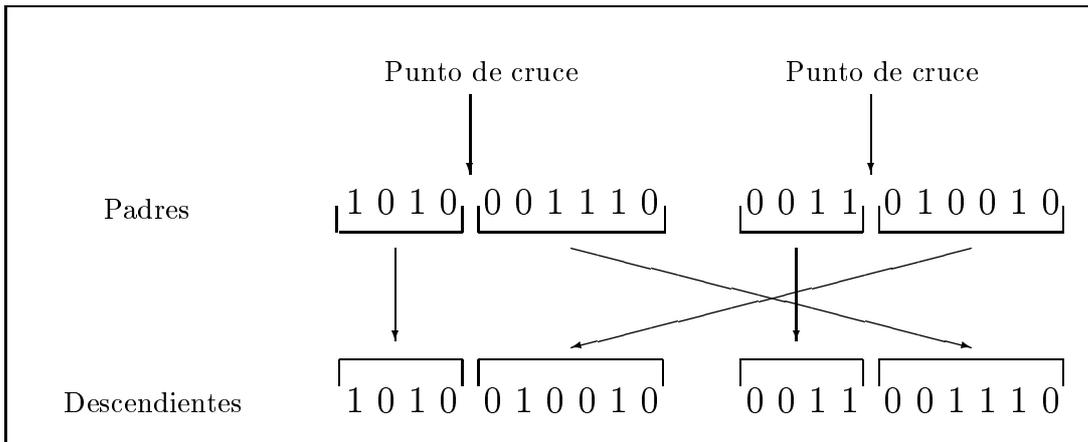


Figura 1.1: Cruce basado en un punto.

Existen variantes del método de cruce anterior. Una de las más conocidas es el *cruce basado en dos puntos*, que selecciona dos puntos entre posiciones previamente determinadas, rompiéndose las cadenas en tres trozos e intercambiándose las partes centrales (ver figura 1.2).

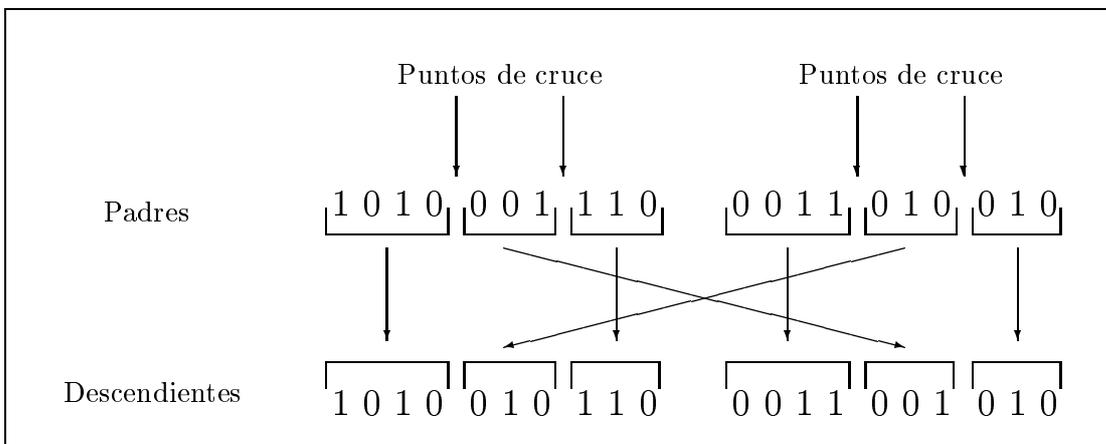


Figura 1.2: Cruce basado en dos puntos.

Una generalización de los cruces basados en puntos es el *cruce uniforme*, donde para cada bit en el primer descendiente se decide, con una probabilidad  $p$ , qué progenitor contribuye a su valor en esa posición. El segundo descendiente recibe el bit del otro progenitor (ver figura 1.3).

- *Mutación*: Actúa sobre individuos, cambiando el valor de un gen. La mutación se produce con una probabilidad usualmente baja, manteniendo la idea de que éste

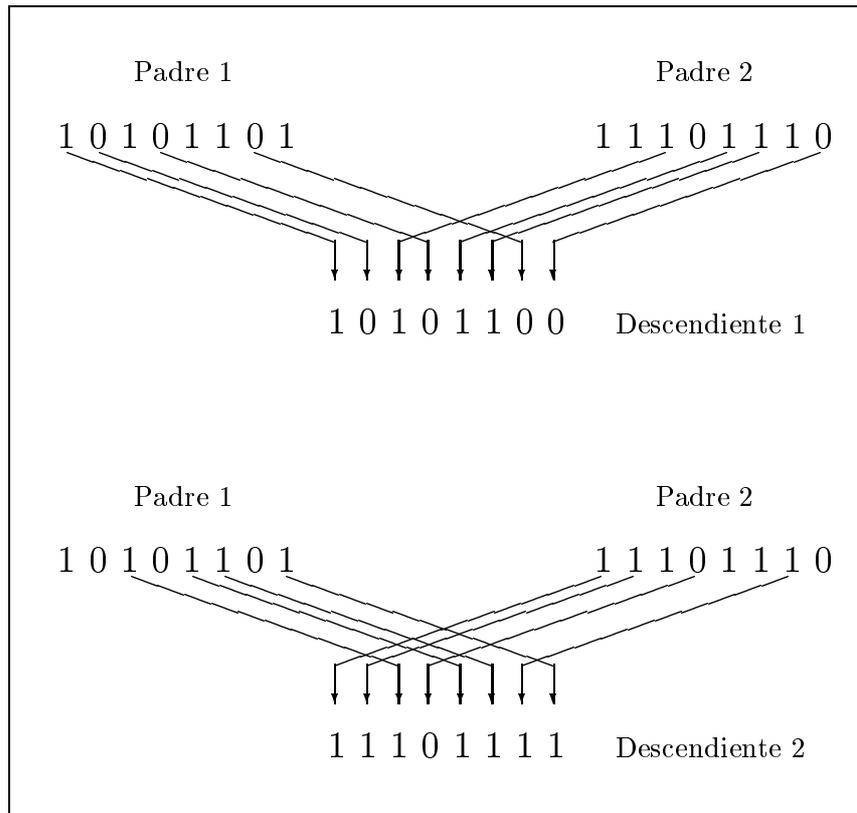


Figura 1.3: Cruce uniforme.

no es un hecho corriente en la naturaleza (ver figura 1.4).

- *Inversión*: Selecciona dos puntos entre posiciones previamente fijadas e invierte el orden de los alelos entre estas posiciones (ver figura 1.5).

## 6. Criterio de parada

Decide si el algoritmo continúa o se para. Ésta es una etapa crucial, pues el algoritmo debe descubrir si contiene al óptimo o está suficientemente cerca de él. Existen varios criterios: se puede detener el algoritmo tras un número preestablecido de iteraciones (o generaciones) –considerando como óptimo el óptimo de la población en ese momento–, o también se puede parar el proceso cuando una parte suficientemente grande de la población alcanza un valor umbral prefijado.

En el caso en el que el criterio decida continuar con el algoritmo, se vuelve a la etapa de evaluación con la población obtenida tras la recombinación.

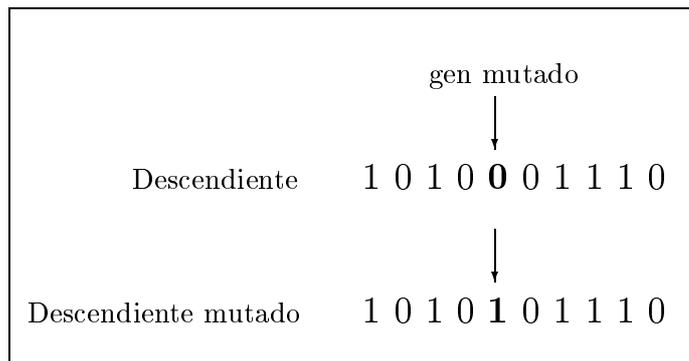


Figura 1.4: Mutación.

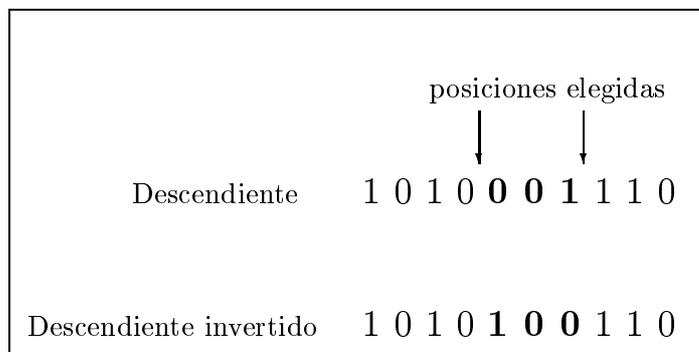


Figura 1.5: Inversión.

Las principales diferencias entre los métodos de optimización clásicos y los Algoritmos Genéticos, son que éstos:

- i) trabajan con codificaciones de las posibles soluciones del problema, no con ellas directamente,
- ii) no se limitan a buscar en un entorno de un punto, sino que trabajan con conjuntos de puntos,
- iii) utilizan únicamente la información que les proporciona la función objetivo,
- iv) utilizan reglas de transición probabilísticas, y no deterministas, para realizar la búsqueda.

En cuanto a cuándo se deben usar los Algoritmos Genéticos, De Jong proporciona en [6] el siguiente criterio:

*“El punto clave para decidir si usar o no un Algoritmo Genético para un problema concreto se centra en la cuestión: ¿Cuál es el espacio en el que se realiza la búsqueda? Si este espacio se entiende bien y posee una estructura que puede ser explotada para usar técnicas de búsqueda específicas, el uso de Algoritmos Genéticos es, en general, computacionalmente menos eficiente. Si el espacio de búsqueda no se entiende bien y está relativamente desestructurado, y se puede obtener una representación efectiva de este espacio, entonces los Algoritmos Genéticos proporcionan una sorprendentemente poderosa búsqueda heurística para espacios grandes y complejos.”*

Por supuesto, si se desea estudiar teóricamente el comportamiento de un Algoritmo Genético, se deben utilizar funciones cuyo óptimo sea conocido de antemano.

### 1.3 ¿Cómo actúan los Algoritmos Genéticos y por qué? Teorema de los esquemas

Para explicar el proceso de actuación de los Algoritmos Genéticos consideremos, por ejemplo, las siguientes cadenas binarias y sus correspondientes valores de ajuste en un determinado problema: calcular el máximo de la función  $f$  dada por  $f(x) = x^2$ .

Cadena	Ajuste
01101	169
11000	576
01000	64
10011	361

Tabla 1.1: Algunos valores de  $f(x) = x^2$ .

Observando en la tabla anterior las cadenas consideradas y sus valores de idoneidad, podemos notar que un 1 en la primera posición de la izquierda parece ser una buena característica para conseguir un valor alto de la función. La propia dinámica del AG explotará esta característica generando muchas cadenas con un 1 en ese primer bit, debido a que éstas tienen, en general, un valor de idoneidad alto, y los operadores de cruce y mutación tienden a mantener ese 1 en la descendencia. Si la calidad de los cromosomas que tienen un 1 esa posición es superior al valor de ajuste medio de todos los cromosomas del espacio de búsqueda, diremos

que esas cadenas son un buen patrón o *esquema*. El hecho de que existan “buenos esquemas” es determinante para la convergencia de los AG.

Formalmente, un esquema describe un conjunto de cadenas con similitudes en ciertas posiciones. Si se considera el alfabeto binario  $\Sigma = \{0, 1\}$  y añadimos un símbolo  $\#$  para crear un alfabeto extendido  $\bar{\Sigma} = \{0, 1, \#\}$ , entonces un esquema en el alfabeto binario es un elemento de  $\bar{\Sigma}^\ell$ ,  $\mathcal{H} = h_{\ell-1} \dots h_0 \in \{0, 1, \#\}^\ell$ , donde cada vez que en una posición aparezca  $\#$  significa que esa posición puede estar ocupada por cualquier alelo del alfabeto binario. Así pues,  $\mathcal{H}$  es el subconjunto del conjunto de todas las cadenas binarias de longitud  $\ell$ , que denotamos por  $\Omega_\ell = \{0, 1\}^\ell$ ,

$$\mathcal{H} = h_{\ell-1} \dots h_0 \leftrightarrow \{s_{\ell-1} \dots s_0 \in \Omega_\ell : \forall j; h_j \neq \# \implies s_j = h_j\}.$$

Por ejemplo, si  $\ell = 3$ ,

$$\mathcal{H} = 1\#\# \leftrightarrow \{100, 101, 110, 111\}.$$

Cualquier conjunto de cadenas  $\mathcal{C}$  define un esquema  $\mathcal{H}$ , dado por

$$h_j = \begin{cases} \# & \text{si } \pi_j(\mathcal{C}) = \{0, 1\} \\ 1 & \text{si } \pi_j(\mathcal{C}) = \{1\} \\ 0 & \text{si } \pi_j(\mathcal{C}) = \{0\} \end{cases}$$

donde  $\pi_j : \Omega_\ell \rightarrow \{0, 1\}$  denota la proyección sobre la  $j$ -ésima componente.

Nótese que el conjunto de cadenas que genera el cruce sobre los elementos de  $\mathcal{C}$  es exactamente el esquema  $\mathcal{H}$  determinado por  $\mathcal{C}$ . Es decir, los esquemas son los conjuntos explorados por el cruce.

Por otra parte, obsérvese la importancia del operador mutación. Su introducción permite al AG salir de un rango concreto en el espacio de búsqueda y explorar nuevas zonas en las que el operador cruce, por sí sólo, no consigue entrar.

El cruce es tradicionalmente considerado como el principal mecanismo de exploración de los AG, por lo que la tendencia ha sido etiquetar a los algoritmos que sólo usan mutación con nombres diferentes al de “Algoritmo Genético”. Si el cruce es el operador que define a los AG, y si los esquemas caracterizan a los conjuntos a los que el cruce explora, se entiende la importancia del estudio de los esquemas.

Fijada una longitud  $\ell$  para las cadenas, existen  $3^\ell$  esquemas sobre el alfabeto binario. Además, cada cadena de longitud  $\ell$  representará a  $2^\ell$  esquemas. En consecuencia, en una

población de  $n$  individuos el número de esquemas representados estará comprendido entre  $2^\ell$  y  $n \cdot 2^\ell$ . Esto da una idea de la magnitud de información que procesa un Algoritmo Genético.

Obviamente, diferentes esquemas poseen características diferentes. Existen dos parámetros importantes en los esquemas: *orden* y *longitud de definición*.

- El *orden* de un esquema  $\mathcal{H}$  –que se denota por  $o(\mathcal{H})$ – es el número de posiciones fijas del esquema, es decir, aquellas ocupadas por un 0 o un 1. Por ejemplo, los esquemas

$$\mathcal{H}_1 = 0110\#1\#\#,$$

$$\mathcal{H}_2 = 0\#\#\#\#\#\#,$$

tienen órdenes  $o(\mathcal{H}_1) = 5$  y  $o(\mathcal{H}_2) = 1$ , respectivamente.

La noción de orden de un esquema es útil para el cálculo de la probabilidad de supervivencia<sup>1</sup> de un esquema tras una mutación: cuanto mayor sea el orden, mayor será la probabilidad de sufrir mutación en uno de sus alelos fijos.

- La *longitud de definición* de un esquema  $\mathcal{H}$  –que se denota por  $\delta(\mathcal{H})$ – es la distancia entre la primera y la última de sus posiciones fijas. De esta forma, para los esquemas definidos anteriormente tenemos:  $\delta(\mathcal{H}_1) = 5$  y  $\delta(\mathcal{H}_2) = 0$ .

La noción de longitud de definición de un esquema está relacionada con su probabilidad de supervivencia tras un cruce: a mayor longitud de definición, mayor probabilidad de que un cruce lo destruya.

El efecto combinado de la selección, el cruce y la mutación conduce a la formulación definitiva del *Teorema de los Esquemas* o *Teorema Fundamental de los Algoritmos Genéticos*, que establece que esquemas con idoneidad por encima de la media de la población, con longitud de definición pequeña y con orden bajo –conocidos como *bloques constructivos* (*building blocks*)– aumentarán su presencia en la siguiente generación (para una exposición detallada de este teorema véase el apéndice A).

Parece razonable esperar que la yuxtaposición de bloques constructivos conduzca al algoritmo a alcanzar un comportamiento óptimo. Esta suposición, basada en la conclusión del

---

<sup>1</sup>Se dice que un esquema sobrevive a un operador cuando la estructura del esquema pervive en, al menos, uno de los descendientes.

teorema de los esquemas, se conoce como *Hipótesis de los Bloques Constructivos* (*Building Block Hypothesis* o BBH).

Sin embargo, aunque en muchos casos los AG se comportan de acuerdo con la BBH, esto no siempre es así. Ello se debe a que la explicación que proporciona el teorema de los esquemas sobre el modo de actuación de los AG es puramente cualitativa. El teorema describe el modo de propagación de un esquema “presente en la población”, de una generación a la siguiente, pero no informa sobre la evolución futura de esquemas que todavía no están representados y que pueden surgir como consecuencia de cruces y/o mutaciones.

Tampoco explica el comportamiento del Algoritmo Genético si la población inicial es de baja calidad. Podría suceder que bloques constructivos desplacen a bloques mejores dirigiendo el algoritmo hacia soluciones subóptimas. En este caso se dice que *el AG ha sufrido un engaño* (*deception*). A lo largo de esta memoria, nosotros nos referiremos a este hecho como al *fenómeno de la decepción* y a las *funciones decepcionantes*.

## 1.4 Funciones decepcionantes. Funciones difíciles. Resultados sorprendentes

Aunque los Algoritmos Genéticos se vienen aplicando con éxito a la resolución de muchos problemas, existen también resultados menos afortunados. En ambas circunstancias, éxito o fracaso, existe todavía un escaso conocimiento acerca de por qué uno u otro se han producido. Dado el creciente interés en aplicar este tipo de algoritmos a un rango de problemas cada vez más amplio es esencial, en su estudio teórico, entender de una forma más completa cómo funcionan y determinar “a priori” cuándo lo harán satisfactoriamente.

Por ello, quizá más que en cualquier otra cuestión, los investigadores teóricos de los Algoritmos Genéticos se han centrado en identificar factores que influyan en la capacidad de un AG para resolver problemas.

Como se ha dicho, una de las causas que puede conducir a un mal funcionamiento de los Algoritmos Genéticos es el engaño sufrido cuando esquemas de bajo orden proporcionan información engañosa al algoritmo acerca de esquemas de mayor orden que refinan a los anteriores (esquemas de orden alto contenidos en los de bajo orden).

Existen varias propuestas de definición formal del concepto de decepción que se basan en la noción de *competición entre esquemas*: dos esquemas son competitivos cuando los bits

definidos de ambos ocupan las mismas posiciones, pero al menos uno de los valores de esos bits es diferente. Por ejemplo, en cadenas binarias de longitud cuatro, con la estructura  $\#dd\#$  (donde “ $d$ ” denota un alelo definido) se pueden representar los cuatro esquemas del conjunto  $\mathcal{P} = \{\#00\#, \#01\#, \#10\#, \#11\#\}$ . Nótese que  $\mathcal{P}$  constituye una partición del espacio de búsqueda  $\Omega_4 = \{0, 1\}^4$ . Se dice que el *orden de la partición* es dos por ser ese el número de posiciones definidas en la estructura  $\#dd\#$ . El *ganador de la competición* entre los hiperplanos de  $\mathcal{P}$  es el esquema con valor de idoneidad media más alta.

Por otra parte, se puede establecer una relación de orden entre las particiones de esquemas: la subsumisión. Se dice que una partición  $\mathcal{P}$  *subsume* a una partición  $\mathcal{P}'$  si cada esquema de  $\mathcal{P}'$  es un subconjunto de un esquema de  $\mathcal{P}$ . Por ejemplo, la anterior partición  $\mathcal{P}$  representada por la estructura  $\#dd\#$ , subsume a la partición  $\mathcal{P}'$  de orden tres, compuesta por los ocho elementos con estructura  $\#ddd$ .

A continuación enunciamos algunas de las propuestas de definición de decepción (establecidas por Whitley en [46]), que son consistentes con la mayoría de las definiciones que de este concepto aparecen en la literatura.

**Definición 1.4.1.** *Un problema contiene decepción si el ganador de una partición  $\mathcal{P}$  posee un bit definido con diferente valor al del correspondiente ganador de otra partición  $\mathcal{Q}$  de mayor orden subsumida por  $\mathcal{P}$ .*

**Definición 1.4.2.** *Un problema es completamente decepcionante de orden  $n$  si, dada una partición  $\mathcal{P}$  de orden  $n$ , los ganadores de todas las particiones de menor orden que subsumen a  $\mathcal{P}$ , conducen a un esquema en  $\mathcal{P}$  distinto del ganador.*

Algunos autores han intentado cuantificar la cantidad de engaño que requiere un problema para hacer errar a un Algoritmo Genético. Con ese fin, Goldberg ([11]) construye el problema más sencillo que puede provocar la divergencia del AG del óptimo global. Para hacer esto, se viola la hipótesis de los bloques constructivos, es decir, se desea que los bloques constructivos cortos y de bajo orden conduzcan a otros bloques constructivos largos y de orden alto que proporcione un subóptimo. El problema más sencillo que puede contener engaño, conocido como *Problema Decepcionante Mínimo* (*Minimal Deceptive Problem* o MDP), es un problema de dos bits.

Se considera el espacio  $\Omega_2$ , formado por las cuatro cadenas 00, 01, 10 y 11, y supongamos que 11 es la cadena en la que una función  $f : \Omega_2 \rightarrow \mathbb{R}$  alcanza el máximo, es decir:

$$f(11) > f(00), f(11) > f(01), f(11) > f(10).$$

Para introducir el engaño se obliga a que el esquema ganador de orden 1 difiera en su alelo fijo del correspondiente alelo de la cadena 11. Esto se consigue exigiendo una de las dos siguientes condiciones:

$$f(0\#) > f(1\#), \quad f(\#0) > f(\#1), \quad (1.1)$$

donde  $f(\mathcal{H})$  denota la idoneidad media del esquema  $\mathcal{H}$ , i.e.,

$$f(\mathcal{H}) = \frac{1}{|\mathcal{H}|} \sum_{s \in \mathcal{H}} f(s),$$

con  $|\mathcal{H}|$  el cardinal de  $\mathcal{H}$ .

Las condiciones (1.1) son, respectivamente, equivalentes a:

$$\frac{f(00) + f(01)}{2} > \frac{f(10) + f(11)}{2}, \quad \frac{f(00) + f(10)}{2} > \frac{f(01) + f(11)}{2}.$$

El cumplimiento simultáneo de las dos inecuaciones conduciría a que  $f(00) > f(11)$ , contradiciendo<sup>2</sup> la hipótesis inicial. Por ello, y sin que suponga ninguna restricción, se considerará cierta la primera. Así, el MDP queda caracterizado por las dos siguientes condiciones:

- i) 11 es la cadena óptima,
- ii)  $f(0\#) > f(1\#)$ .

Estas dos condiciones permiten clasificar el problema en dos tipos, tipo I:  $f(01) > f(00)$  y tipo II:  $f(00) \geq f(01)$ .

Los resultados experimentales demuestran que el MDP de tipo I no resulta difícil para el algoritmo. Aunque inicialmente la cadena subóptima 01 tiene mayor número de representantes que 11, finalmente el algoritmo converge hacia el óptimo global en un tiempo razonable (ver figura 1.6)

En cuanto al MDP de tipo II, no siempre se puede asegurar la convergencia al óptimo 11. La evolución de las poblaciones dependerá del valor medio de los esquemas y del número de representantes de estos esquemas en la población inicial. De hecho, en la figura 1.7 se observa que con un 25% de copias de cadenas de cada uno de los cuatro esquemas de la partición, el algoritmo converge hacia el óptimo global. Sin embargo, esta situación en la práctica no siempre se tiene. En general, en la población inicial la proporción de representantes de cada uno de los esquemas no es la misma. En la figura 1.8 se observa que en el caso de partir de un 75% de cadenas con valor 00, el algoritmo es engañado y converge al subóptimo.

---

<sup>2</sup>Esta contradicción significa que el problema no es completamente decepcionante de orden uno.

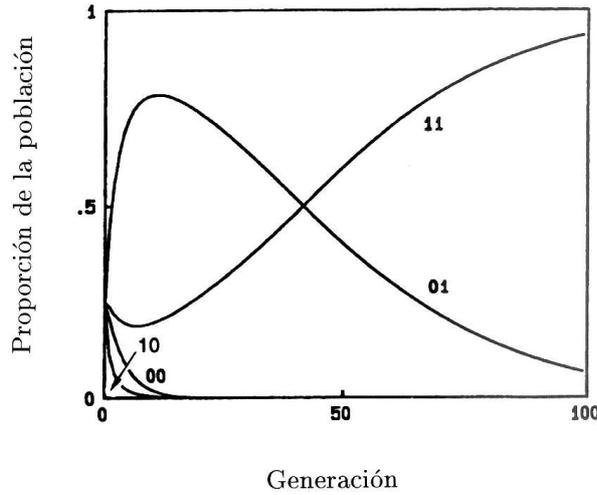


Figura 1.6: Ejecución de un AG para el MDP de tipo I (tomada de [11]).

Del estudio del MDP se deduce que la existencia del fenómeno de decepción no es suficiente para garantizar la dificultad en la búsqueda del óptimo. Tampoco es una condición necesaria. Existen diversos ejemplos de problemas difíciles que no contienen decepción<sup>3</sup>. Nosotros a continuación resumimos los resultados obtenidos por Forrest y Mitchell en [8]. En este trabajo, las autoras construyen funciones aparentemente fáciles (*Royal Road functions*), aunque resultaron ser difíciles. Para su construcción se seleccionó una cadena óptima, rompiéndola posteriormente en bloques constructivos.

Por comodidad en la notación, denotaremos por  $\alpha^{(n)}$  los esquemas formados por  $n$  copias de  $\alpha$ , con  $\alpha \in \{0, 1, \#\}$ ; i.e.,  $\alpha^{(n)} = \alpha \overbrace{\dots}^n \alpha$ .

La cadena óptima es  $1^{(64)}$ , y los bloques constructivos que caracterizan a la primera de las funciones Royal Road ( $\mathfrak{R}_1$ ) son:

$$\begin{aligned} \sigma_1 &= 1^{(8)} \#^{(56)} \\ \sigma_2 &= \#^{(8)} 1^{(8)} \#^{(48)} \\ \sigma_3 &= \#^{(16)} 1^{(8)} \#^{(40)} \\ \sigma_4 &= \#^{(24)} 1^{(8)} \#^{(32)} \\ \sigma_5 &= \#^{(32)} 1^{(8)} \#^{(24)} \\ \sigma_6 &= \#^{(40)} 1^{(8)} \#^{(16)} \\ \sigma_7 &= \#^{(48)} 1^{(8)} \#^{(8)} \end{aligned}$$

---

<sup>3</sup>Para un estudio más detallado ver, por ejemplo, [15].

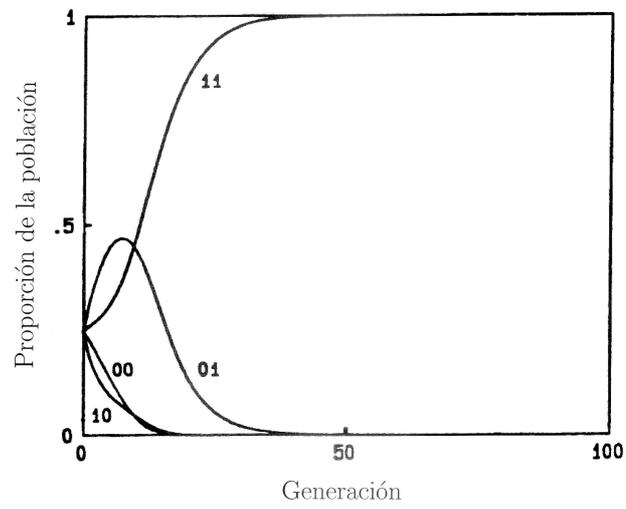


Figura 1.7: Ejecución de un AG para el MDP de tipo II, caso convergente (tomada de [11]).

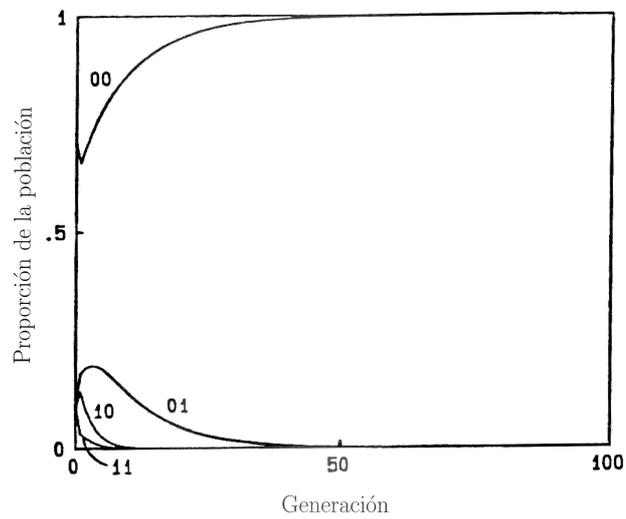


Figura 1.8: Ejecución de un AG para el MDP de tipo II, caso divergente (tomada de [11]).

$$\sigma_8 = \#^{(56)} 1^{(8)}.$$

A cada bloque se le asigna un valor  $c_i = 8$  ( $i = 1, \dots, 8$ ), coincidente con el número de bits definidos. La función  $\mathfrak{R}_1 : \Omega = \{0, 1\}^{64} \rightarrow \mathbb{R}$  se define como sigue: una cadena  $s \in \Omega$  añade 8 puntos a su valor de ajuste por cada esquema  $\sigma_i$  al que representa. Por ejemplo, si  $s$  pertenece exactamente a dos de los bloques constructivos de orden 8, entonces  $\mathfrak{R}_1(s) = 16$ . En general,  $\mathfrak{R}_1(s)$  es la suma de los coeficientes  $c_i$  correspondientes a cada uno de los esquemas dados de los cuales es un ejemplo. La contribución al valor de ajuste de una escalón intermedio en el camino hacia el óptimo (un bloque constructivo de tamaño un múltiplo de 8, resultante de la composición de  $\sigma_i$  y  $\sigma_j$ ,  $i \neq j$ ) es, por tanto, una combinación lineal de las contribuciones de los bloques constructivos de bajo orden.

La función  $\mathfrak{R}_2$  se construye de forma similar a  $\mathfrak{R}_1$ , añadiendo a los bloques constructivos anteriores, los seis siguientes:

$$\begin{aligned} \sigma_9 &= 1^{(16)} \#^{(48)} \\ \sigma_{10} &= \#^{(16)} 1^{(16)} \#^{(32)} \\ \sigma_{11} &= \#^{(32)} 1^{(16)} \#^{(16)} \\ \sigma_{12} &= \#^{(48)} 1^{(16)} \\ \sigma_{13} &= 1^{(32)} \#^{(32)} \\ \sigma_{14} &= \#^{(32)} 1^{(32)}, \end{aligned}$$

siendo ahora  $c_i = 8$  ( $i = 1, \dots, 8$ ),  $c_i = 16$  ( $i = 9, \dots, 12$ ),  $c_i = 32$  ( $i = 13, 14$ ).  $\mathfrak{R}_2(s)$  se obtiene sumando los coeficientes  $c_i$  ( $i = 1, \dots, 14$ ) correspondientes a los esquemas  $\sigma_i$  de los cuales  $s$  es un representante. Por ejemplo,  $\mathfrak{R}_2(1^{(8)} \#^{(48)} 1^{(8)}) = 16$ , ya que la cadena representa a  $\sigma_1$  y  $\sigma_8$ , mientras que  $\mathfrak{R}_2(1^{(16)} \#^{(48)}) = 32$ , ya que la cadena representa a  $\sigma_1$ ,  $\sigma_2$  y  $\sigma_9$ . Por tanto, el valor de ajuste de una cadena no sólo depende del número de bloques constructivos a los que representa, sino también de sus posiciones en la cadena.

De antemano, cabe esperar que el algoritmo encuentre más rápidamente el óptimo en  $\mathfrak{R}_2$  que en  $\mathfrak{R}_1$ . En  $\mathfrak{R}_2$  existe un claro camino que muestra cómo obtener, mediante el operador cruce, los bloques constructivos de orden 16 a partir de los de orden 8, a continuación los de orden 32 y, finalmente, el óptimo (ver figura 1.9). Parece razonable esperar que la existencia de esta vía acelere la obtención del óptimo por parte del algoritmo. Sin embargo, los resultados experimentales mostraron lo contrario: ambas funciones resultaron ser difíciles siendo  $\mathfrak{R}_2$  más difícil que  $\mathfrak{R}_1$ .

Analizando las ejecuciones del algoritmo se observa cómo la aparición de cadenas representantes de bloques constructivos de orden alto hace que desaparezcan de la población, en

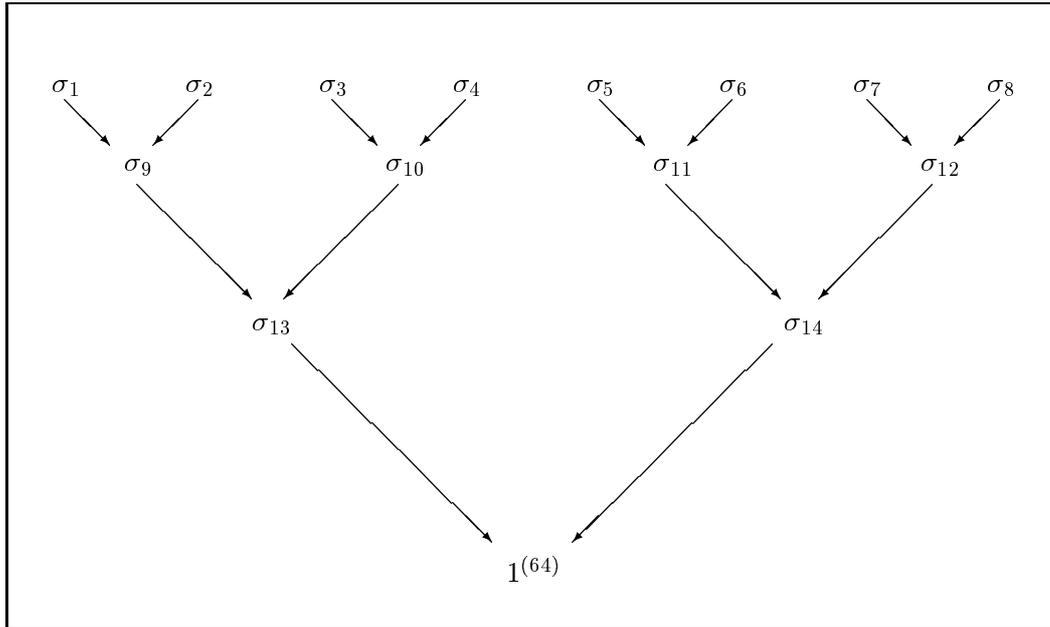


Figura 1.9: Evolución de los bloques constructivos en la función  $\mathfrak{R}_2$ .

la siguiente generación, los representantes de bloques constructivos de orden bajo adyacentes al anterior, que son necesarios para la construcción del óptimo (para más detalles ver [8]).

En  $\mathfrak{R}_1$ , que carece de la información extra que proporcionan los bloques constructivos intermedios, la aparición de representantes de bloques constructivos de orden alto no produce un efecto tan dramático en el descubrimiento y persistencia de los bloques constructivos de orden más bajo. Tengamos en cuenta, por ejemplo, que un elemento  $s$  de  $\sigma_{11}$  tiene valores de ajuste  $\mathfrak{R}_1(s) = 16$  y  $\mathfrak{R}_2(s) = 32$ , mientras que si  $s'$  es un representante de  $\sigma_4$ , se tiene  $\mathfrak{R}_1(s') = 8$  y  $\mathfrak{R}_2(s') = 8$ . El refuerzo no lineal (*epistasis*) usado para definir  $\mathfrak{R}_2$  parece ser uno de los causantes de la lenta convergencia hacia el óptimo, en comparación con  $\mathfrak{R}_1$ . En el capítulo siguiente se analiza esta característica.

## Capítulo 2

# Epistasis en el caso binario

---

### 2.1 Varianza epistática

Como se ha expuesto en el capítulo anterior, existen muchos factores que provocan que una función sea difícil de optimizar mediante un AG. Uno de ellos es la existencia de vínculos entre los alelos de las cadenas en las que se codifican las posibles soluciones del problema—como sucede en las funciones Royal Road—. En [35] Rawlins compara este fenómeno con una situación similar en Genética: cuando un gen situado en un locus enmascara el efecto de otro gen situado en diferente locus. Cuando este fenómeno ocurre, se dice que el primer gen es *epistático* con el segundo.

Adaptando esta idea al campo de los Algoritmos Genéticos, Rawlins habla de *epistasis mínima* cuando cada gen de una cadena es independiente de cualquier otro; es decir, cuando la función de idoneidad puede considerarse básicamente como una combinación lineal de funciones, cada una de las cuales depende únicamente de un gen. En el otro extremo, se tiene *epistasis máxima* si ningún subconjunto propio de genes es independiente de ningún otro gen, y esta situación se corresponde esencialmente con una función de idoneidad aleatoria.

En [4], para una codificación binaria, Davidor propone una medida que predice el valor de la función a partir de los bits, como estimación de la no linealidad que existe en la representación. Esta medida, llamada *varianza epistática*, presenta una perspectiva sobre la

dificultad de optimización mediante un AG, diferente a la caracterizada con el fenómeno de la decepción.

Para la definición formal de la varianza epistática, en [4], se parte del espacio  $\Omega_\ell = \{0, 1\}^\ell$  de cadenas binarias de longitud  $\ell$ , de una función  $f : \Omega_\ell \rightarrow \mathbb{R}$  y, dada una muestra  $\mathcal{P}$  de  $\Omega_\ell$  de tamaño  $N_{\mathcal{P}} = |\mathcal{P}|$ , se considera la idoneidad media de la población  $\mathcal{P}$ :

$$\bar{f} = \frac{1}{N_{\mathcal{P}}} \sum_{s \in \mathcal{P}} f(s),$$

donde  $f(s)$  denota el valor de ajuste de la cadena  $s$ . Para un alelo  $a \in \{0, 1\}$ , situado en la  $i$ -ésima posición de las cadenas de  $\mathcal{P}$ , Davidor define el ajuste medio de  $a$  como:

$$f_i(a) = \frac{1}{N_i(a)} \sum_{s \in \mathcal{P}_i(a)} f(s),$$

siendo  $\mathcal{P}_i(a)$  el conjunto formado por las cadenas de  $\mathcal{P}$  cuya posición  $i$ -ésima está ocupada por el alelo  $a$ , y  $N_i(a) = |\mathcal{P}_i(a)|$  es el cardinal de  $\mathcal{P}_i(a)$ . El *valor de exceso del alelo* se define como:

$$E_i(a) = f_i(a) - \bar{f}.$$

Para una cadena  $s \in \Omega_\ell$ , el *valor de exceso génico* es:

$$EG(s) = \sum_{i=0}^{\ell-1} E_i(s_i)$$

y su *valor génico previsto*:

$$\tilde{f}(s) = EG(s) + \bar{f}.$$

La diferencia  $\varepsilon(s) = f(s) - \tilde{f}(s)$  es la *epistasis de la cadena  $s$* . Finalmente define la *varianza epistática* como:

$$\sigma^2 = \frac{1}{N_{\Omega_\ell}} \sum_{s \in \Omega_\ell} \varepsilon(s)^2.$$

A continuación mostramos el cálculo de la varianza epistática de tres funciones. La primera ( $\mathcal{L}$ ) es una función lineal. La segunda ( $c$ ) se anula en todas las cadenas salvo dos de ellas. La elección de las cadenas con imagen no nula no es arbitraria, deben ser una cadena y su complemento a uno. A las funciones de este tipo se las conoce como “funciones camello” –debido a sus dos picos– y se analizan en la subsecciones 2.2.2, 3.2.2 y 4.2.2. La tercera función ( $\delta$ ) se anula en todas las cadenas excepto en una de ellas. La normalización de esta función proporciona la función de Dirac  $\delta_0$ , que se estudia en las subsecciones 3.2.1, 4.2.1 y 4.3.1.

Los valores que toman las funciones  $\mathcal{L}$ ,  $c$  y  $\delta$  sobre cadenas de longitud 3 se reflejan en la tabla 2.1.

cadenas	$\mathcal{L}$	$c$	$\delta$
000	0	14	28
001	1	0	0
010	2	0	0
011	3	0	0
100	4	0	0
101	5	0	0
110	6	0	0
111	7	14	0

$$\bar{\mathcal{L}} = \bar{c} = \bar{\delta} = 3.5$$

Tabla 2.1: Valores de las funciones  $\mathcal{L}$ ,  $c$  y  $\delta$  sobre cadenas de longitud 3.

$i$	$a$	$\mathcal{L}_i(a)$	$E_i(a)$
0	0	3	-0.5
	1	4	0.5
1	0	2.5	-1
	1	4.5	1
2	0	1.5	-2
	1	5.5	2

$s$	$\mathcal{L}(s)$	$EG(s)$	$\tilde{\mathcal{L}}(s)$	$\varepsilon(s)$
000	0	-3.5	0	0
001	1	-2.5	1	0
010	2	-1.5	2	0
011	3	-0.5	3	0
100	4	0.5	4	0
101	5	1.5	5	0
110	6	2.5	6	0
111	7	3.5	7	0

Tabla 2.2: Cálculo de la epistasis de las cadenas de  $\Omega_3$  para una función lineal.

La función  $\mathcal{L}$  tiene varianza epistática nula, como se deduce de los cálculos de la tabla 2.2 (tomada de [4]).

Por otra parte, la varianza epistática de  $c$  es  $\sigma^2 = 36.75$ . Este valor se obtiene fácilmente de la tabla 2.3.

Finalmente, la varianza epistática de  $\delta$ ,  $\sigma^2 = 67.375$ , se calcula con los datos de tabla 2.4.

$i$	$a$	$c_i(a)$	$E_i(a)$
0	0	3.5	0
	1	3.5	0
1	0	3.5	0
	1	3.5	0
2	0	3.5	0
	1	3.5	0

$s$	$c(s)$	$EG(s)$	$\tilde{c}(s)$	$\varepsilon(s)$
000	14	0	3.5	10.5
001	0	0	3.5	-3.5
010	0	0	3.5	-3.5
011	0	0	3.5	-3.5
100	0	0	3.5	-3.5
101	0	0	3.5	-3.5
110	0	0	3.5	-3.5
111	14	0	3.5	10.5

Tabla 2.3: Cálculo de la epistasis de las cadenas de  $\Omega_3$  para una función camello.

$i$	$a$	$\delta_i(a)$	$E_i(a)$
0	0	7	3.5
	1	0	-3.5
1	0	7	3.5
	1	0	-3.5
2	0	7	3.5
	1	0	-3.5

$s$	$\delta(s)$	$EG(s)$	$\tilde{\delta}(s)$	$\varepsilon(s)$
000	28	10.5	14	14
001	0	3.5	7	-7
010	0	3.5	7	-7
011	0	-3.5	0	0
100	0	3.5	7	-7
101	0	-3.5	0	0
110	0	-3.5	0	0
111	0	-10.5	-14	14

Tabla 2.4: Cálculo de la epistasis de las cadenas de  $\Omega_3$  para una función  $\delta$ .

Como se puede observar de estos tres sencillos ejemplos, el cálculo de la varianza epistática de una función (a través del proceso definitorio de Davidor) es un trabajo laborioso, por lo que se hace necesario un método alternativo menos costoso. Esto se consigue con la ayuda del Álgebra Lineal.

## 2.2 Descripción algebraica de la epistasis

Con vistas a obtener la descripción algebraica de la epistasis, introducimos en esta sección alguna terminología y notaciones básicas. Asimismo se exponen, sin demostración, algunos resultados conocidos que nos serán útiles posteriormente.

### 2.2.1 Epistasis normalizada

Como en el epígrafe anterior, tomaremos como alfabeto  $\Sigma = \{0, 1\}$ , así los cromosomas son todas las cadenas binarias de longitud  $\ell$ . Enumeraremos las posiciones de las cadenas –bits– de derecha a izquierda y de 0 a  $\ell - 1$ .

Asimismo, denotaremos por  $\mathcal{M}_n(\mathcal{C})$  al conjunto de las matrices cuadradas de orden  $n$  con elementos en  $\mathcal{C}$ . En nuestro caso,  $\mathcal{C}$  será  $\mathbb{Z}$  o  $\mathbb{Q}$ .

Considerando como población  $\mathcal{P} = \Omega_\ell$  el conjunto de cadenas binarias de longitud  $\ell$ , en [41] se reescribe la definición de Davidor en una única fórmula. Así, para una cadena  $s = s_{\ell-1} \dots s_1 s_0$ <sup>1</sup>, su epistasis es:

$$\varepsilon_\ell(s) = f(s) - \sum_{i=0}^{\ell-1} \frac{1}{2^{\ell-1}} \sum_{t \in \Omega_i(s_i)} f(t) + \frac{\ell-1}{2^\ell} \sum_{t \in \Omega_\ell} f(t),$$

donde  $\Omega_i(s_i)$  denota el conjunto de cadenas en  $t \in \Omega_\ell$  tales que  $t_i = s_i$ .

Los autores definen la *epistasis global* de  $f$  como:

$$\varepsilon_\ell(f) = \sqrt{\sum_{s \in \Omega_\ell} \varepsilon_\ell^2(s)}.$$

En [42], esta definición se reformula con el uso del Álgebra Lineal al introducir los vectores:

$$\mathbf{e} = \begin{pmatrix} \varepsilon(00 \dots 0) \\ \varepsilon(00 \dots 1) \\ \vdots \\ \varepsilon(11 \dots 1) \end{pmatrix},$$

$$\mathbf{f} = \begin{pmatrix} f(00 \dots 0) \\ f(00 \dots 1) \\ \vdots \\ f(11 \dots 1) \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{2^\ell-1} \end{pmatrix},$$

---

<sup>1</sup>En ocasiones, identificaremos una cadena  $s = s_{\ell-1} \dots s_1 s_0 \in \Omega_\ell$  con el valor numérico  $\sum_{0 \leq i < \ell} s_i 2^i$ .

y la matriz real simétrica  $2^\ell$ -dimensional  $\mathbf{E}_\ell = (e_{st}^\ell) \in \mathcal{M}_{2^\ell}(\mathbb{Q})$ , dada por:

$$e_{st}^\ell = \frac{1}{2^\ell}(\ell + 1 - 2d_{st}), \quad 0 \leq s, t \leq 2^\ell - 1,$$

donde  $d_{st}$  denota la distancia Hamming entre  $s$  y  $t$ , i.e., el número de bits en los cuales las representaciones binarias de  $s$  y  $t$  difieren<sup>2</sup>. Para valores pequeños de  $\ell$  se tiene:

$$\begin{aligned} \mathbf{E}_0 &= (1), \\ \mathbf{E}_1 &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \\ \mathbf{E}_2 &= \frac{1}{4} \begin{pmatrix} 3 & 1 & 1 & -1 \\ 1 & 3 & -1 & 1 \\ 1 & -1 & 3 & 1 \\ -1 & 1 & 1 & 3 \end{pmatrix} \end{aligned}$$

y

$$\mathbf{E}_3 = \frac{1}{8} \begin{pmatrix} 4 & 2 & 2 & 0 & 2 & 0 & 0 & -2 \\ 2 & 4 & 0 & 2 & 0 & 2 & -2 & 0 \\ 2 & 0 & 4 & 2 & 0 & -2 & 2 & 0 \\ 0 & 2 & 2 & 4 & -2 & 0 & 0 & 2 \\ 2 & 0 & 0 & -2 & 4 & 2 & 2 & 0 \\ 0 & 2 & -2 & 0 & 2 & 4 & 0 & 2 \\ 0 & -2 & 2 & 0 & 2 & 0 & 4 & 2 \\ -2 & 0 & 0 & 2 & 0 & 2 & 2 & 4 \end{pmatrix},$$

por ejemplo.

En [42] se demuestra que  $\mathbf{e} = \mathbf{f} - \mathbf{E}_\ell \mathbf{f}$ . Nótese que  $\varepsilon_\ell(f) = \|\mathbf{e}\|$ .

La intuición dice que dada una función  $f$ , ella y cualquier múltiplo suyo  $\alpha f$  ( $\alpha \neq 0$ ) deben tener la misma epistasis. En [39], los autores normalizan este concepto definiendo la *epistasis normalizada* de una función de ajuste  $f$  como

$$\varepsilon_\ell^*(f) = \varepsilon_\ell^2 \left( \frac{f}{\|\mathbf{f}\|} \right) = \frac{{}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{E}_\ell) \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}} = 1 - \frac{{}^t \mathbf{f} \mathbf{E}_\ell \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}}, \quad (2.1)$$

donde  $\mathbf{I}_\ell$  es la matriz identidad de orden  $2^\ell$ .

---

<sup>2</sup>Cuando sea necesario, se trabajará con cadenas de distintas longitudes. En ese caso, si  $s \in \Omega_\ell$  y  $t \in \Omega_j$ , su distancia Hamming se calculará con las subcadenas  $s_{m-1} \dots s_0$  y  $t_{m-1} \dots t_0$ , siendo  $m = \text{mínimo}\{\ell, j\}$ , y se denotará  $d_{st}^m$ .

En la práctica es más fácil trabajar con la matriz  $\mathbf{G}_\ell = 2^\ell \mathbf{E}_\ell \in \mathcal{M}_{2^\ell}(\mathbb{Z})$ . Esto permite escribir la epistasis normalizada de  $f$  como

$$\varepsilon_\ell^*(f) = 1 - \frac{1}{2^\ell} \frac{\gamma_\ell(f)}{\|\mathbf{f}\|^2}, \quad (2.2)$$

donde  $\gamma_\ell(f) = {}^t \mathbf{f} \mathbf{G}_\ell \mathbf{f}$ .

En [39] se demuestra que, para todo  $\ell \geq 0$ ,  $\mathbf{G}_\ell$  puede construirse recursivamente mediante:

$$\mathbf{G}_{\ell+1} = \begin{pmatrix} \mathbf{G}_\ell + \mathbf{U}_\ell & \mathbf{G}_\ell - \mathbf{U}_\ell \\ \mathbf{G}_\ell - \mathbf{U}_\ell & \mathbf{G}_\ell + \mathbf{U}_\ell \end{pmatrix}, \quad (2.3)$$

siendo  $\mathbf{U}_\ell = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \in \mathcal{M}_{2^\ell}(\mathbb{Z})$  y  $\mathbf{G}_0 = (1)$ .

Además, se verifica que  $\text{Rg}(\mathbf{G}_\ell) = \ell + 1$ , y sus autovalores son 0 y  $2^\ell$  (ver [39] para detalles). En consecuencia, los autovalores de  $\mathbf{E}_\ell$  son 0 y 1, y  $\mathbb{R}^{2^\ell} = \text{Ker}(\mathbf{E}_\ell) \oplus \text{Im}(\mathbf{E}_\ell)$ .

## 2.2.2 Valores extremos de la epistasis normalizada

La definición original de epistasis (debida a Davidor) está basada en la idea de que si una representación de los datos tiene baja epistasis debería ser procesada de forma eficiente por un AG, mientras que si contiene alta epistasis el proceso de búsqueda probablemente acabará dirigiendo el algoritmo a un óptimo local. En esta subsección resumimos la caracterización de las funciones con epistasis mínima y máxima obtenida en [39].

En [41] se ha demostrado que  $\varepsilon_\ell^*(f) = 0$  si y sólo si  $f$  tiene epistasis mínima en el sentido de Rawlins; es decir, cuando  $f(s_{\ell-1}, \dots, s_0) = \sum_{i=0}^{\ell-1} g_i(s_i)$ , donde  $g_i$  depende sólo del  $i$ -ésimo bit  $s_i$  de  $s$ . En [39], esta caracterización se obtiene de una forma sencilla y elegante. Para ello, los autores consideran, para  $0 \leq i \leq \ell - 1$ , las funciones  $h_i^\ell : \{0, 1\} \rightarrow \mathbb{R}$ :

$$h_i^\ell(s_i) = \begin{cases} 1 & \text{si } s_i = 1 \\ 0 & \text{en otro caso,} \end{cases}$$

cuyos vectores asociados denotan por  $\mathbf{h}_i^\ell$ . Asimismo consideran  $u_\ell$ , la función constante con valor 1, cuyo vector asociado es  $\mathbf{u}_\ell$ . Entonces, un argumento de inducción les permite demostrar que  $\{h_0^\ell, \dots, h_{\ell-1}^\ell, u_\ell\}$  es una base para el subespacio  $\text{Im}(\mathbf{G}_\ell)$ .

Por otra parte, una función  $g_i : \{0, 1\} \rightarrow \mathbb{R}$  que depende sólo del  $i$ -ésimo bit se puede describir como:

$$g_i(s_i) = \begin{cases} a_i & \text{si } s_i = 1 \\ b_i & \text{si } s_i = 0, \end{cases}$$

es decir,  $g_i = a_i h_i^\ell + b_i(u_\ell - h_i^\ell)$ .

Claramente, si  $f$  tiene epistasis mínima en el sentido de Rawlins, entonces  $f \in \text{Im}(\mathbf{G}_\ell)$ , es decir,  $\varepsilon_\ell^*(f) = 0$ . Recíprocamente, si  $\varepsilon_\ell^*(f) = 0$  se tiene

$$f = \sum_{i=0}^{\ell-1} \alpha_i h_i^\ell + \beta u_\ell = \sum_{i=0}^{\ell-1} g_i,$$

donde  $g_0 = (\alpha_0 + \beta)h_0^\ell + \beta(u_\ell - h_0^\ell)$  y  $g_i = \alpha_i h_i^\ell$ , para  $i = 1, \dots, \ell - 1$ .

En efecto, el valor mínimo teórico de la epistasis normalizada  $\varepsilon_\ell^*(f) = 0$  se alcanza para cualquier valor de  $\ell$ . De hecho, si  $\ell = 1$  entonces  ${}^t \mathbf{f} \mathbf{G}_1 \mathbf{f} = {}^t \mathbf{f} 2 \mathbf{I}_2 \mathbf{f} = 2 \|\mathbf{f}\|^2$ , por lo que

$$\varepsilon_1^*(f) = 1 - \frac{1}{2} \frac{{}^t \mathbf{f} \mathbf{G}_1 \mathbf{f}}{\|\mathbf{f}\|^2} = 0.$$

Por otra parte, si  $\ell > 1$  basta considerar, por ejemplo,  $\mathbf{f} = 2^{-\frac{\ell}{2}} \mathbf{u}_\ell$ , siendo  $\mathbf{u}_\ell = {}^t(1, \dots, 1) \in \mathbb{R}^{2^\ell}$ . Entonces  $\|\mathbf{f}\| = 1$  y

$${}^t \mathbf{f} \mathbf{G}_\ell \mathbf{f} = \sum_{i,j=0}^{2^\ell-1} g_{ij} f_i f_j = 2^{-\ell} \sum_{i,j=0}^{2^\ell-1} g_{ij} = 2^{-\ell} \sum_{i=0}^{2^\ell-1} \sum_{j=0}^{2^{\ell-1}-1} (g_{ij} + g_{i\hat{j}}) = 2.$$

Aquí se ha utilizado que  $g_{ij} + g_{i\hat{j}} = 2$ , ( $0 \leq i, j \leq 2^\ell - 1$ ), donde  $\hat{j} = 2^\ell - 1 - j$  (ver [39] para más detalles).

Con respecto al valor máximo de  $\varepsilon_\ell^*$ , en [39] se demuestra:

**Proposición 2.2.1.** ([39]) *Para cualquier entero positivo  $\ell$  y cualquier función  $f : \Omega_\ell \rightarrow \mathbb{R}$  no negativa, con  $\|\mathbf{f}\| = 1$ , se verifica:*

$$\varepsilon_\ell^*(f) \leq 1 - \frac{1}{2^{\ell-1}}.$$

Este valor teórico máximo para funciones positivas se alcanza. Para ello basta considerar

$$\mathbf{f} = \begin{pmatrix} \frac{\sqrt{2}}{2} \\ 0 \\ \vdots \\ 0 \\ \frac{\sqrt{2}}{2} \end{pmatrix}.$$

Este ejemplo (funciones camello) es, esencialmente, el único. De hecho,

**Proposición 2.2.2.** ([39]) *Para cualquier entero positivo  $\ell \geq 2$  y cualquier función no negativa  $f$ , con  $\|f\| = 1$ , son equivalentes:*

1.  $\varepsilon_\ell^*(f) = 1 - \frac{1}{2^{\ell-1}}$ ,
2. existe algún  $0 \leq i \leq 2^{\ell-1}$  tal que  $f_i = f_{2^{\ell-1}-i} = 1$ ,  $f_j = 0$  en otro caso.

### 2.2.3 Un ejemplo de cálculo directo de la epistasis normalizada $\varepsilon^*$

A pesar de que, desde el punto de vista teórico, esta formulación algebraica del cálculo de la epistasis normalizada es sencilla y elegante, a nivel práctico los cálculos que se deben efectuar son de una gran complejidad. Como ejemplo de ello resumimos aquí el desarrollo realizado en [20] para un tipo de funciones –las llamadas *funciones Template*– que se utilizan como una interesante clase de funciones de laboratorio con las que chequear ciertas propiedades de los AG.

Las funciones Template (*Funciones Patrón*) se caracterizan porque el valor de ajuste de las cadenas depende de dos parámetros: la longitud de las cadenas ( $\ell$ , si  $\Omega = \{0, 1\}^\ell$ ) y la longitud  $n$  de una subcadena  $t$ , fijada de antemano, que actúa como un “patrón” cuya presencia (o ausencia) en la cadena determina su valor de idoneidad. En concreto, el valor de ajuste de una cadena  $s$  es el número de veces que el patrón  $t$  aparece en  $s$ . Si el patrón es la cadena  $t = 1^n = \overbrace{1 \cdots 1}^n$  y  $T_\ell^n$  denota la función Template definida en  $\Omega = \{0, 1\}^\ell$ , entonces se tiene, por ejemplo,

$$T_\ell^2(1^\ell) = T_\ell^2(11 \dots 11) = \ell - 1,$$

y

$$T_\ell^3(01110 \dots 011) = 1.$$

Denotando por  $\mathbf{T}_\ell^n \in \mathbb{R}^{2^\ell}$  el vector que determina la función  $T_\ell^n$ , i.e.,

$$\mathbf{T}_\ell^n = \begin{pmatrix} T_\ell^n(00 \dots 0) \\ \vdots \\ T_\ell^n(11 \dots 1) \end{pmatrix},$$

se sigue que  $\mathbf{T}_\ell^n$  es el vector nulo si  $\ell < n$  y, para cualquier  $\ell \geq n$ , se tiene:

$$\mathbf{T}_\ell^n = \begin{pmatrix} \mathbf{T}_{\ell-1}^n \\ \mathbf{T}_{\ell-1}^n + \mathbf{D}_{\ell-1}^n \end{pmatrix},$$

con

$$\mathbf{D}_{\ell-1}^n = \begin{pmatrix} \mathbf{0}_{\ell-2} \\ \vdots \\ \mathbf{0}_{\ell-n} \\ \mathbf{u}_{\ell-n} \end{pmatrix} \in \mathbb{R}^{2^{\ell-1}},$$

donde, para todo  $2 \leq i \leq n$ , se denota por  $\mathbf{0}_{\ell-i}$  el vector nulo en  $\mathbb{R}^{2^{\ell-i}}$  y  $\mathbf{u}_{\ell-n} = {}^t(1, \dots, 1) \in \mathbb{R}^{2^{\ell-n}}$ .

Para calcular la norma de  $\mathbf{T}_\ell^n$ , un argumento de inducción muestra que:

**Lema 2.2.3.** ([20]) *Para cualquier par de enteros  $n \leq \ell$  y para cualquier  $0 \leq i \leq n$ , se tiene:*

1.

$$\text{Tr}(\mathbf{T}_\ell^n) = 2^{\ell-n}(\ell - n + 1), \quad (2.4)$$

2.

$$\|\mathbf{T}_{n+i}^n\|^2 = 2^i(3i - 1) + 2,$$

donde  $\text{Tr}(\mathbf{T}_\ell^n)$  denota la traza del vector  $\mathbf{T}_\ell^n$ .

De este resultado se deduce que:

**Proposición 2.2.4.** ([20]) *Para cualquier par de enteros  $\ell \geq n$ , se verifica:*

$$\|\mathbf{T}_\ell^n\|^2 = \begin{cases} 2^{\ell-n}(3(\ell - n) - 1) + 2 & \text{si } \ell \leq 2n, \\ 2^{\ell-n}(3(\ell - n) - 1) + 2^{\ell-2n}(2 + (\ell - 2n)(\ell - 2n - 1)) & \text{si } \ell \geq 2n. \end{cases}$$

*Demostración.* El primer caso ( $\ell \leq 2n$ ) es exactamente la segunda parte del lema previo. Para  $\ell \geq 2n$ , se aplica inducción en  $\ell$  y se usa  $\mathbf{T}_{n+i}^n \mathbf{D}_{n+i}^n = 2^{i+1} - 1$  y la aserción (2.4) del lema 2.2.3.  $\square$

Por la fórmula dada en (2.2), para obtener la expresión de la epistasia normalizada de  $T_\ell^n$ , se necesita calcular  $\gamma(T_\ell^n) = {}^t\mathbf{T}_\ell^n \mathbf{G}_\ell \mathbf{T}_\ell^n$ .

En [20] se comprueba que, para todo  $\ell \geq n$ ,

$$\begin{aligned} \gamma(T_\ell^n) &= {}^t\mathbf{T}_\ell^n \mathbf{G}_\ell \mathbf{T}_\ell^n \\ &= ({}^t\mathbf{T}_{\ell-1}^n, {}^t\mathbf{T}_{\ell-1}^n + {}^t\mathbf{D}_{\ell-1}^n) \begin{pmatrix} \mathbf{G}_{\ell-1} + \mathbf{U}_{\ell-1} & \mathbf{G}_{\ell-1} - \mathbf{U}_{\ell-1} \\ \mathbf{G}_{\ell-1} - \mathbf{U}_{\ell-1} & \mathbf{G}_{\ell-1} + \mathbf{U}_{\ell-1} \end{pmatrix} \begin{pmatrix} \mathbf{T}_{\ell-1}^n \\ \mathbf{T}_{\ell-1}^n + \mathbf{D}_{\ell-1}^n \end{pmatrix} \\ &= 4 {}^t\mathbf{T}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{T}_{\ell-1}^n + 4 {}^t\mathbf{T}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{D}_{\ell-1}^n + {}^t\mathbf{D}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{D}_{\ell-1}^n + {}^t\mathbf{D}_{\ell-1}^n \mathbf{U}_{\ell-1}^n \mathbf{D}_{\ell-1}^n \end{aligned}$$

donde:

$${}^t\mathbf{D}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{D}_{\ell-1}^n = n \cdot 4^{\ell-n}$$

y

$${}^t\mathbf{D}_{\ell-1}^n \mathbf{U}_{\ell-1}^n \mathbf{D}_{\ell-1}^n = 2^{\ell-n} \|\mathbf{u}_{\ell-n}\|^2 = 4^{\ell-n}.$$

Asimismo, usando la simetría de  $\mathbf{G}_\ell$  y un argumento recursivo, los autores obtienen:

$$\begin{aligned} {}^t\mathbf{T}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{D}_{\ell-1}^n &= 2^{n-2} (2 {}^t\mathbf{u}_{\ell-n} \mathbf{G}_{\ell-n} \mathbf{T}_{\ell-n}^n + {}^t\mathbf{u}_{\ell-n} \mathbf{G}_{\ell-n} \mathbf{D}_{\ell-n} + {}^t\mathbf{u}_{\ell-n} \mathbf{U}_{\ell-n} \mathbf{D}_{\ell-n}) \\ &\quad + 4^{\ell-n-1} (n-2)(n+3) \end{aligned}$$

y, como  ${}^t\mathbf{u}_{\ell-n} \mathbf{G}_{\ell-n} = {}^t\mathbf{u}_{\ell-n} \mathbf{U}_{\ell-n} = 2^{\ell-n} {}^t\mathbf{u}_{\ell-n}$  y  ${}^t\mathbf{T}_{n+i}^n \mathbf{D}_{n+i}^n = 2^{i+1} - 1$ , entonces

$$\begin{aligned} {}^t\mathbf{T}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{D}_{\ell-1}^n &= 2^{\ell-1} (Tr(\mathbf{T}_{\ell-n}^n) + Tr(\mathbf{D}_{\ell-n}^n)) + 4^{\ell-n-1} (n-2)(n+3) \\ &= 2^{\ell-1} (2^{\ell-2n} (\ell - 2n + 1) + 2^{\ell-2n+1}) + 4^{\ell-n-1} (n-2)(n+3) \\ &= 4^{\ell-n-1} (n^2 - 3n + 2\ell). \end{aligned}$$

Combinando todos estos hechos, se verifica para  $\ell \geq 2n$ ,

$$\begin{aligned} \gamma(T_\ell^n) &= 4 {}^t\mathbf{T}_{\ell-1}^n \mathbf{G}_{\ell-1} \mathbf{T}_{\ell-1}^n + 4^{\ell-n} (n^2 - 3n + 2\ell) + n 4^{\ell-n} + 4^{\ell-n} \\ &= 4^{\ell-n} \left( (\ell - 2n)(n^2 + \ell + 2) + \frac{n}{3} (2n^2 + 7) + 2n^2 + 1 \right). \end{aligned}$$

La demostración del caso  $n \leq \ell \leq 2n$  es similar (ver [20] para detalles), y se obtiene:

$$\gamma(T_{n+i}^n) = 4^i \left( n(i+1)^2 + 1 + \frac{i}{3} (4 - i^2) \right),$$

para  $0 \leq i \leq n$ .

En consecuencia,

**Teorema 2.2.5.** ([20]) *La epistasis normalizada de la función Template  $T_\ell^n$  es:*

$$\varepsilon^*(T_\ell^n) = \begin{cases} 1 - \frac{1+n(\ell-n+1)^2 + \frac{(\ell-n)}{3}(4-(\ell-n)^2)}{2^n(3(\ell-n)-1+2^{n-\ell+1})} & \text{si } n \leq \ell \leq 2n \\ 1 - \frac{(\ell-2n)(n^2+\ell+2) + \frac{n}{3}(2n^2+7) + 2n^2+1}{2^n(3(\ell-n)-1+(\ell-2n)^2+2(n+1)-\ell)} & \text{si } \ell \geq 2n. \end{cases} \quad (2.5)$$

## 2.3 Epistasis y transformadas de Walsh

### 2.3.1 Coeficientes de partición y transformadas de Walsh

El ejemplo de las funciones Template muestra que, aunque el uso de la notación algebraica simplifica el cálculo de la epistasis de Davidor, la recursividad de la matriz  $\mathbf{G}_\ell$  y su dimensión, cuando  $\ell$  crece, impiden realizar los cálculos en una cantidad razonable de tiempo. Con el objeto de simplificar estos cálculos y teniendo en cuenta que, como se afirma en [24],

*“Las transformadas de Walsh parecen desempeñar un papel fundamental en el análisis de los Algoritmos Genéticos. Tanto ellas como sus coeficientes de Walsh describen esencialmente la idoneidad media de la función respecto a ciertos esquemas elementales bien determinados, lo que proporciona una descripción muy práctica de su epistasis”*

Naudts et al utilizan en [33] las *Funciones de Walsh* (analizadas en [12]) que introducimos a continuación.

Para cualquier cadena  $t \in \Omega_\ell = \{0, 1\}^\ell$ , la *función de Walsh*  $\psi_t$  es:

$$\psi_t(s) = (-1)^{s \cdot t} = \prod_{i=0}^{\ell-1} (-1)^{s_i t_i},$$

donde  $s \cdot t$  denota el producto escalar de  $s$  y  $t$ , i.e.,  $s \cdot t = \sum_{i=0}^{\ell-1} s_i t_i$ .

Por ejemplo,  $\psi_{01101}(10111) = 1$  y  $\psi_{01101}(10110) = -1$ .

Es sabido ([12]) que las funciones de Walsh constituyen una base del espacio vectorial de funciones sobre  $\Omega_\ell$  con valores reales, ya que son independientes y cada función  $f : \Omega_\ell \rightarrow \mathbb{R}$  se puede escribir como:

$$f(s) = \sum_{t \in \Omega_\ell} v_t \psi_t(s) \quad \left( \equiv \sum_{j=0}^{2^\ell - 1} v_j \psi_j(s) \right)$$

para cualquier  $s \in \Omega_\ell$ , donde cada coeficiente  $v_t \in \mathbb{R}$  está dado por:

$$v_t = \frac{1}{2^\ell} \sum_{s \in \Omega_\ell} f(s) \psi_t(s).$$

Como  $\psi_0 \equiv 1$ ,  $v_0$  es el valor medio de  $f$  en  $\Omega_\ell$ .

Si se considera:

$$\mathbf{v} = \begin{pmatrix} v_0 \\ \vdots \\ v_{2^\ell - 1} \end{pmatrix}$$

y la matriz simétrica,  $2^\ell$ -dimensional,  $\mathbf{V}_\ell = (\psi_t(s))_{s,t \in \Omega_\ell}$ , se sigue que:

$$\mathbf{f} = \mathbf{V}_\ell \mathbf{v}.$$

En la práctica, es habitual trabajar con la matriz (simétrica)  $\mathbf{W}_\ell = 2^{-\ell/2} \mathbf{V}_\ell$ , llamada *matriz de Walsh*, que verifica  $\mathbf{W}_\ell^2 = \mathbf{I}_\ell$ . Para valores pequeños de  $\ell$ ,  $\mathbf{W}_\ell$  se calcula fácilmente:

$$\begin{aligned} \mathbf{W}_0 &= (1) && (= \mathbf{V}_0), \\ \mathbf{W}_1 &= 2^{-1/2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, && \mathbf{W}_2 = 2^{-1} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}. \end{aligned}$$

Además puede obtenerse por recursión ya que satisface, para  $\ell \geq 0$  (ver [33]):

$$\mathbf{W}_{\ell+1} = 2^{-\frac{1}{2}} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix}. \quad (2.6)$$

Se denominan *coeficientes de Walsh* de  $f$  a las componentes  $w_t = w_t(f) = 2^{\ell/2} v_t(f)$  del vector  $\mathbf{w}$  tal que  $\mathbf{f} = \mathbf{W}_\ell \mathbf{w}$ , lo que conduce a:

$$f(s) = 2^{-\ell/2} \sum_{t \in \Omega_\ell} w_t \psi_t(s).$$

Nótese que, al ser la matriz  $\mathbf{W}_\ell$  idempotente, se verifica:

$$\|\mathbf{f}\| = \|\mathbf{w}\|. \quad (2.7)$$

Por otra parte, si se considera la matriz diagonal  $\mathbf{D}_\ell$ , cuyos únicos elementos no nulos  $d_{ii}$  tienen valor 1 y están situados en  $i = 0$  y  $i = 2^j$ ,  $0 \leq j \leq \ell - 1$  entonces:

**Lema 2.3.1.** ([33]) *Para todo  $\ell \geq 0$ ,*

$$\mathbf{W}_\ell \mathbf{E}_\ell \mathbf{W}_\ell = \mathbf{D}_\ell.$$

Este resultado, junto con la idempotencia de  $\mathbf{W}_\ell$  y las expresiones (2.1) (página 40) y (2.7) proporciona el valor de  $\varepsilon_\ell^*(f)$  en términos de los coeficientes de Walsh:

**Proposición 2.3.2.** ([33]) *Si  $w_0, \dots, w_{2^\ell-1}$  son los coeficientes de Walsh de una función de ajuste  $f$ , entonces la epistasis normalizada de  $f$  está dada por*

$$\varepsilon_\ell^*(f) = 1 - \frac{w_0^2 + \sum_{i=0}^{\ell-1} w_{2^i}^2}{\sum_{j=0}^{2^\ell-1} w_j^2}. \quad (2.8)$$

La expresión (2.8) caracteriza, de forma sencilla, las funciones con epistasis normalizada mínima. Claramente, el valor mínimo,  $\varepsilon_\ell^*(f) = 0$ , se alcanza para aquellas funciones tales que

$$w_0^2 + \sum_{i=0}^{\ell-1} w_{2^i}^2 = \sum_{j=0}^{2^\ell-1} w_j^2,$$

es decir, aquellas funciones cuyos únicos coeficientes de Walsh no nulos son  $w_0$  y  $w_{2^i}$  ( $0 \leq i \leq \ell - 1$ ).

Estas funciones que, como no podía ser de otra manera, coinciden exactamente con las caracterizadas en la subsección 2.2.2, se denominan *funciones de primer orden*: funciones con la propiedad de que existen

$$g_i : \{0, 1\} \rightarrow \mathbb{R}, \quad (0 \leq i \leq \ell - 1)$$

tales que

$$f(s) = \sum_{i=0}^{\ell-1} g_i(s_i).$$

Hemos indicado al principio de esta sección que los coeficientes de Walsh describen el valor medio de la función sobre ciertos esquemas elementales. En efecto, se ha demostrado en [12] el llamado *Teorema del Valor Medio para Hiperplanos* que establece:

$$f(\mathcal{H}) = 2^{-\ell/2} \sum_{t \in J(\mathcal{H})} w_t \psi_t(\beta(\mathcal{H})),$$

donde

- $f(\mathcal{H})$  es el valor medio de  $f$  sobre el esquema  $\mathcal{H}$ , i.e.,

$$f(\mathcal{H}) = \frac{1}{|\mathcal{H}|} \sum_{s \in \mathcal{H}} f(s),$$

- $\beta : \{0, 1, \#\}^\ell \rightarrow \Omega_\ell$  es la función dada por:

$$\beta(\mathcal{H})_i = \begin{cases} 1 & \text{si } h_i = 1, \\ 0 & \text{en otro caso,} \end{cases}$$

y

- $J(\mathcal{H}) = \{t \in \Omega_\ell; \forall 0 \leq i < \ell, h_i = \# \Rightarrow t_i = 0\}$ , es decir, el conjunto de todas las cadenas que tienen un 0 en las posiciones en que  $\mathcal{H}$  tiene un #.

Por ejemplo, si  $\mathcal{H} = 1\#00\#1$  entonces  $\beta(\mathcal{H}) = 100001$  y  $J(\mathcal{H}) = \#0\#\#0\#$ .

Esto hace que, en particular,

- I) si  $\mathcal{H} = \Omega_\ell = \#\dots\#$  entonces  $J(\mathcal{H}) = \{0\dots0\}$ ,  $\beta(\mathcal{H}) = 0\dots0$  y, en consecuencia,

$$f(\Omega_\ell) = f(\#\dots\#) = 2^{-\ell/2}w_0,$$

- II) si  $\mathcal{H} = \#\dots\#\overset{i}{\downarrow}a\#\dots\#$ , con  $a \in \{0, 1\}$ , está claro que  $\beta(\mathcal{H}) = 0\dots\overset{i}{\downarrow}a\dots0$  y  $t \in J(\mathcal{H})$ , si y sólo si,  $t = 0\dots0t_i0\dots0$ , entonces:

$$\psi_i(\beta(\mathcal{H})) = \begin{cases} 1 & \text{si } t_i = 0 \\ (-1)^a & \text{si } t_i = 1 \end{cases}$$

y

$$f(\mathcal{H}) = f(\#\dots\#\overset{i}{\downarrow}a\#\dots\#) = 2^{-\ell/2} (w_0 + (-1)^a w_{2^i}).$$

Por otra parte, también en [12] se define el valor  $\epsilon(\mathcal{H})$  –llamado *coeficiente de partición del esquema  $\mathcal{H}$* – como

$$f(\mathcal{H}) = \sum_{\mathcal{H}' \supset \mathcal{H}} \epsilon(\mathcal{H}'),$$

que permite escribir el valor medio de un esquema más general en función de los coeficientes de Walsh. De hecho, usando la definición anterior, se sigue que:

$$\epsilon(\mathcal{H}) = f(\mathcal{H}) - \sum_{\mathcal{H}' \supsetneq \mathcal{H}} \epsilon(\mathcal{H}'),$$

lo que permite calcular  $\epsilon(\mathcal{H})$  por recursión.

En particular, si se consideran  $\Omega_\ell = \#\dots\#$  y  $\mathcal{H}_i = \#\dots\#\overset{i}{\downarrow}0\#\dots\#$ , entonces:

$$\epsilon(\#\dots\#) = f(\Omega_\ell) = 2^{-\ell/2}w_0, \tag{2.9}$$

$$\epsilon(\# \dots \# \overset{i}{\downarrow} 0 \# \dots \#) = f(\mathcal{H}_i) - f(\Omega_\ell) = 2^{-\ell/2} w_{2^i},$$

por lo que

$$w_{2^i} = 2^{\ell/2} f(\mathcal{H}_i) - w_0 \quad (2.10)$$

y, si ahora se considera  $\mathcal{H}_{ij} = \# \dots \# \overset{j}{\downarrow} 0 \# \dots \# \overset{i}{\downarrow} 0 \# \dots \#$ , entonces:

$$\begin{aligned} \epsilon(\# \dots \# \overset{j}{\downarrow} 0 \# \dots \# \overset{i}{\downarrow} 0 \# \dots \#) &= f(\mathcal{H}_{ij}) - \epsilon(\# \dots \# \overset{j}{\downarrow} 0 \# \dots \# \# \# \dots \#) - \epsilon(\# \dots \# \# \# \dots \# \overset{j}{\downarrow} 0 \# \dots \#) \\ &\quad - \epsilon(\# \dots \#) \\ &= 2^{-\ell/2} w_{2^{i+2j}}, \end{aligned}$$

y, en consecuencia,

$$w_{2^{i+2j}} = 2^{\ell/2} f(\mathcal{H}_{ij}) - w_{2^i} - w_{2^j} - w_0. \quad (2.11)$$

En general, si  $\mathcal{H}_{i_1 \dots i_k} = \# \dots \# \overset{i_k}{\downarrow} 0 \# \dots \# \overset{i_1}{\downarrow} 0 \# \dots \#$ , entonces

$$\begin{aligned} \epsilon(\# \dots \# \overset{i_k}{\downarrow} 0 \# \dots \# \overset{i_1}{\downarrow} 0 \# \dots \#) &= f(\mathcal{H}_{i_1 \dots i_k}) \\ &\quad - \epsilon(\# \dots \#) - \sum_{p=1}^{k-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_p \leq k} \epsilon(\# \dots \# \overset{i_{\lambda_p}}{\downarrow} 0 \# \dots \# \overset{i_{\lambda_1}}{\downarrow} 0 \# \dots \#) \\ &= 2^{-\ell/2} w_{2^{i_1 + \dots + i_k}}, \end{aligned}$$

y

$$w_{2^{i_1 + \dots + i_k}} = 2^{\ell/2} f(\mathcal{H}_{i_1 \dots i_k}) - w_0 - \sum_{p=1}^{k-1} \left( \sum_{1 \leq \lambda_1 < \dots < \lambda_p \leq k} w_{2^{i_{\lambda_1} + \dots + i_{\lambda_p}}} \right). \quad (2.12)$$

A continuación –y con el objeto de comparar el coste computacional del cálculo de la epistasis normalizada directamente de la definición y a través de las transformadas de Walsh–nosotros calcularemos nuevamente, mediante un método más simple, la epistasis normalizada de las funciones Template ya obtenida en la subsección 2.2.3.

### 2.3.2 Funciones Template: cálculo de su epistasis mediante los coeficientes de Walsh

Para calcular la epistasis normalizada de  $T_\ell^n$ , teniendo en cuenta (2.7) y las proposiciones 2.2.4 y 2.3.2, basta determinar el valor de los coeficientes de Walsh de orden cero,  $w_0$ , y los de orden uno,  $w_{2^i}$ , para  $i = 0, \dots, \ell - 1$ .

Para ello, en primer lugar nótese que:

**Lema 2.3.3.** *Para cualquier  $0 \leq i < \ell$ , el valor de ajuste del esquema  $\mathcal{H}_i = \# \cdots \# \overset{i}{0} \# \cdots \# \in \{0, 1, \#\}^\ell$  está dado por*

$$T_\ell^n(\mathcal{H}_i) = 2^{i-\ell+1} Tr(\mathbf{T}_{\ell-i-1}^n) + 2^{-i} Tr(\mathbf{T}_i^n).$$

*Demostración.* Como cualquier cadena  $s \in \mathcal{H}_i$  se puede escribir en la forma  $s = s'0s''$  con  $s' = s_{\ell-1} \dots s_{i+1} \in \Omega_{\ell-i-1}$  y  $s'' = s_{i-1} \dots s_0 \in \Omega_i$  entonces, por la definición de  $T_\ell^n$ , directamente se sigue

$$\begin{aligned} T_\ell^n(s) &= T_\ell^n(s_{\ell-1}, \dots, s_{i+1}, 0, \dots, 0) + T_\ell^n(0, \dots, 0, s_{i-1}, \dots, s_0) \\ &= T_{\ell-i-1}^n(s') + T_i^n(s''), \end{aligned}$$

para todo  $s \in \mathcal{H}_i$ . Por tanto,

$$\begin{aligned} T_\ell^n(\mathcal{H}_i) &= \frac{1}{2^{\ell-1}} \sum_{s \in \mathcal{H}_i} T_\ell^n(s) \\ &= \frac{1}{2^{\ell-1}} \left\{ 2^i \sum_{s' \in \Omega_{\ell-i-1}} T_{\ell-i-1}^n(s') + 2^{\ell-i-1} \sum_{s'' \in \Omega_i} T_i^n(s'') \right\} \\ &= 2^{i-\ell+1} Tr(\mathbf{T}_{\ell-i-1}^n) + 2^{-i} Tr(\mathbf{T}_i^n). \end{aligned}$$

□

Ahora tenemos:

**Proposición 2.3.4.** *Para cualquier par de enteros  $\ell \geq n$ , el valor de  $\Gamma = w_0^2 + \sum_{i=0}^{\ell-1} w_{2^i}^2$  es:*

$$\Gamma = \begin{cases} 2^{\ell-2n} \left\{ (2n - \ell + 1)(\ell - n + 1)^2 + \frac{(\ell-n)(\ell-n+1)(2(\ell-n)+1)}{3} \right\} & \text{si } \ell \leq 2n, \\ 2^{\ell-2n} \left\{ (\ell - n + 1)^2 + n^2(\ell - 2n) + \frac{n(n+1)(2n+1)}{3} \right\} & \text{si } \ell \geq 2n. \end{cases}$$

*Demostración.* Con los resultados previos y (2.10), nosotros directamente obtenemos los coeficientes de Walsh  $w_{2^i}$ , ( $0 \leq i < \ell$ ), que están reflejados en las tablas 2.5 y 2.6:

$i$	$Tr(\mathbf{T}_{\ell-i-1}^n)$	$Tr(\mathbf{T}_i^n)$	$T_\ell^n(\mathcal{H}_i)$	$w_{2^i}$
$0 \leq i < \ell - n$	$2^{\ell-n-i-1}(\ell - n - i)$	0	$2^{-n}(\ell - n - i)$	$-2^{\frac{\ell-2n}{2}}(i + 1)$
$\ell - n \leq i < n$	0	0	0	$-w_0$
$n \leq i < \ell$	0	$2^{i-n}(i - n + 1)$	$2^{-n}(i - n + 1)$	$-2^{\frac{\ell-2n}{2}}(\ell - i)$

Tabla 2.5: Coeficientes de Walsh de primer orden de  $T_\ell^n$  ( $\ell \leq 2n$ ).

$i$	$Tr(\mathbf{T}_{\ell-i-1}^n)$	$Tr(\mathbf{T}_i^n)$	$T_\ell^n(\mathcal{H}_i)$	$w_{2^i}$
$0 \leq i < n$	$2^{\ell-n-i-1}(\ell - n - i)$	0	$2^{-n}(\ell - n - i)$	$-2^{\frac{\ell-2n}{2}}(i + 1)$
$n \leq i < \ell - n$	$2^{\ell-n-i-1}(\ell - n - i)$	$2^{i-n}(i - n + 1)$	$2^{-n}(\ell - 2n + 1)$	$-2^{\frac{\ell-2n}{2}}n$
$\ell - n \leq i < \ell$	0	$2^{i-n}(i - n + 1)$	$2^{-n}(i - n + 1)$	$-2^{\frac{\ell-2n}{2}}(\ell - i)$

Tabla 2.6: Coeficientes de Walsh de primer orden de  $T_\ell^n$  ( $\ell \geq 2n$ ).

Además por (2.4),

$$w_0 = 2^{\ell/2} T_\ell^n(\Omega_\ell) = 2^{-\ell/2} Tr(\mathbf{T}_\ell^n) = 2^{\frac{\ell-2n}{2}} (\ell - n + 1).$$

Por tanto, para  $\ell \leq 2n$ , tenemos que

$$\begin{aligned} w_0^2 + \sum_{i=0}^{\ell-1} w_{2^i}^2 &= w_0^2 + \sum_{i=0}^{\ell-n-1} 2^{\ell-2n} (i + 1)^2 + \sum_{i=\ell-n}^{n-1} w_0^2 + \sum_{i=n}^{\ell-1} 2^{\ell-2n} (\ell - i)^2 \\ &= (2n - \ell + 1) w_0^2 + 2 \sum_{j=1}^{\ell-n} 2^{\ell-2n} j^2 \\ &= (2n - \ell + 1) w_0^2 + 2^{\ell-2n+1} \binom{\ell - n + 1}{2} \frac{2(\ell - n) + 1}{3} \end{aligned}$$

$$= 2^{\ell-2n} \left\{ (2n - \ell + 1)(\ell - n + 1)^2 + \frac{(\ell - n)(\ell - n + 1)(2(\ell - n) + 1)}{3} \right\},$$

ya que  $\sum_{j=1}^{\ell-n} j^2 = \binom{\ell-n+1}{2} \frac{2(\ell-n)+1}{3}$ .

Similarmente, si  $\ell \geq 2n$ , se tiene

$$\begin{aligned} w_0^2 + \sum_{i=0}^{\ell-1} w_{2^i}^2 &= w_0^2 + 2^{\ell-2n} \left\{ \sum_{i=0}^{n-1} (i+1)^2 + \sum_{i=n}^{\ell-n-1} n^2 + \sum_{i=\ell-n}^{\ell-1} (\ell-i)^2 \right\} \\ &= 2^{\ell-2n} \left\{ (\ell-n+1)^2 + n^2(\ell-2n) + 2 \sum_{j=1}^n j^2 \right\} \\ &= 2^{\ell-2n} \left\{ (\ell-n+1)^2 + n^2(\ell-2n) + \frac{n(n+1)(2n+1)}{3} \right\}. \end{aligned}$$

□

Ahora un cálculo directo nos conduce a la expresión de  $\varepsilon_\ell^*(T_\ell^n) = 1 - \frac{\Gamma}{\|\mathbf{T}_\ell^n\|^2}$  dada en (2.5), en la página 45.

## 2.4 Relación entre orden y dificultad

Se ha mostrado ya en el capítulo anterior, que existen diversos factores (v.g. fenómeno de la decepción, epistasis) que pueden contribuir a que un problema sea difícil de optimizar por un AG pero que, por sí solos, no son suficientes para estimar esa dificultad. De hecho, en [9] se asegura:

*“Cualquier esfuerzo investigador que tenga éxito en la teoría de la evolución de los AG debe tener en cuenta esta multiplicidad, en vez de concentrarse sólo en un factor.”*

El aserto anterior está avalado por numerosos resultados experimentales que muestran que, aún siendo independientes, estos factores parecen reforzarse mutuamente. Se ha demostrado (por ejemplo en [34]) que existe una fuerte correlación entre la alta epistasis y la dificultad para algunas clases de funciones, en particular para aquellas que se pueden caracterizar por un número limitado de parámetros. Las funciones Template, las funciones Royal Road y sus generalizadas son ejemplos de tales funciones.

Por otra parte, las denominadas funciones de primer orden (ver página 48) –funciones con epistasis nula– no presentan ninguna dificultad para ser optimizadas.

Las funciones de primer orden se llaman así porque sólo contienen componentes de primer orden; es decir, funciones que dependen exactamente de un único bit, i.e.,

$$f : \Omega_\ell \rightarrow \mathbb{R}$$

$$s = s_{\ell-1} \dots s_0 \mapsto f(s) = \sum_{i=0}^{\ell-1} g_i(s_i)$$

donde  $g_i : \{0, 1\} \rightarrow \mathbb{R}$ .

Este concepto se puede generalizar definiendo una función  $f$  de orden  $k$  ( $k > 1$ ) como

$$f(s) = \sum_{0 \leq i < \ell} g_i(s_i) + \sum_{0 \leq i_1 < i_2 < \ell} g_{i_1 i_2}(s_{i_1}, s_{i_2}) + \dots + \sum_{0 \leq i_1 < \dots < i_k < \ell} g_{i_1 \dots i_k}(s_{i_1}, \dots, s_{i_k}), \quad (2.13)$$

donde  $g_{i_1 \dots i_r}(s_{i_1}, \dots, s_{i_r})$  describe la interacción entre los  $r$  alelos situados en las posiciones  $i_1, i_2, \dots, i_r$ .

Obsérvese que, de acuerdo con la definición anterior, las difíciles funciones camello –funciones de epistasis normalizada máxima– son funciones de orden alto, en concreto orden  $\ell$  si  $\ell$  es impar y orden  $\ell - 1$  si  $\ell$  es par (como se demostrará en la subsección 4.2.2). El orden, pues, parece ser un factor importante en el estudio de la dificultad.

En esta sección realizamos un estudio comparativo entre las funciones Template y las funciones Royal Road. Ambas presentan un comportamiento difícil frente a los Algoritmos Genéticos. Calculamos su orden y mostramos cómo su dificultad se puede explicar a través del análisis de su orden junto con su epistasis. Nosotros tenemos:

**Lema 2.4.1.** *La función  $T_\ell^n$  es de orden  $n$ .*

*Demostración.* En efecto,  $T_\ell^n(s) = \sum_{k=0}^{\ell-n} \tau_k(s_k, \dots, s_{k+n-1})$ , donde, para  $k = 0, \dots, \ell - n$ ,

$$\tau_k : \Omega_n \rightarrow \mathbb{R},$$

$$\tau_k(s_k, \dots, s_{k+n-1}) = \begin{cases} 1 & \text{si } s_k = \dots = s_{k+n-1} = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

□

En vista de los lazos existentes entre los valores de los alelos, parece razonable esperar que, fijado  $\ell$ , al aumentar  $n$  también lo haga el valor de la epistasis normalizada de  $T_\ell^n$ , así como la dificultad de optimización. Este hecho se demuestra en [20], en donde se comprueba experimentalmente la fuerte correlación existente entre la epistasis normalizada y el número de generaciones necesarias por un *Algoritmo Genético Simple* para alcanzar el óptimo, y se calcula explícitamente (de forma laboriosa) el valor de  $\varepsilon_\ell^*(T_\ell^n)$  dada por (2.5) como se ha indicado en la subsección 2.2.3.

Imitando las construcciones realizadas en [9] (detalladas en la sección 1.4), en [34] se introducen, para cada par de números enteros no negativos  $m \leq n$ , las funciones Royal Road generalizadas  $\mathfrak{R}_m^n$  de tipo I a través de los esquemas

$$\sigma_p^{n,m} = \#^{(2^m p)} \mathbf{1}^{(2^m)} \#^{(2^n - 2^m(p+1))}$$

donde  $0 \leq p < 2^{n-m}$  ( $p \in \mathbb{Z}$ ). El valor de  $\mathfrak{R}_m^n$  sobre una cadena  $s$  de longitud  $2^{2^n}$ , i.e.,  $s \in \Omega_{2^n} = \{0, 1\}^{2^n}$ , es:

$$\mathfrak{R}_m^n(s) = \sum_{s \in \sigma_p^{n,m}} 2^m.$$

Por ejemplo, si  $n = 4$  y  $m = 2$ , entonces se tiene  $\mathfrak{R}_2^4(0\dots00011) = 0$ ,  $\mathfrak{R}_2^4(1\dots11100) = 12$  y  $\mathfrak{R}_2^4(0\dots01111) = 4$ .

Obviamente,  $\mathfrak{R}_3^6$  es la función Royal Road  $\mathfrak{R}_1$  originalmente definida por Forrest y Mitchell en [9] (ver sección 1.4).

Además, en [34] (por un procedimiento largo y complicado) se obtiene

$$\varepsilon_{2^n}^*(\mathfrak{R}_m^n) = \frac{2^{2^m} - 2^m - 1}{2^{2^m} + 2^{n-m} - 1}.$$

Nosotros ahora tenemos:

**Lema 2.4.2.** *La función  $\mathfrak{R}_m^n$  es de orden  $2^m$ .*

*Demostración.*  $\mathfrak{R}_m^n(s) = \sum_{j=0}^{2^{n-m}-1} \rho_j(s_{j \cdot 2^m}, \dots, s_{(j+1) \cdot 2^m - 1})$ , siendo

$$\rho_j(s_{j \cdot 2^m}, \dots, s_{(j+1) \cdot 2^m - 1}) = \begin{cases} 2^m & \text{si } s_{j \cdot 2^m} = \dots = s_{(j+1) \cdot 2^m - 1} = 1, \\ 0 & \text{en otro caso,} \end{cases}$$

con  $j = 0, \dots, 2^{n-m} - 1$ . □

En [34] se comprueba la buena correlación entre la (alta) epistasis y la dificultad de las funciones  $\mathfrak{R}_m^n$  (en cadenas de longitud 64).

Nosotros, en la tabla 2.7, realizamos un estudio comparativo entre los resultados obtenidos por Naudts et al en [34] y los que nosotros hemos obtenido al ejecutar un AG sobre funciones Template definidas en  $\Omega_{64}$ . Para que el estudio sea válido, el AG considerado tiene las mismas características que el fijado por los autores para las funciones Royal Road: selección ranking lineal, cruce en un punto con probabilidad 0.7 y mutación con probabilidad 0.001. Se para el AG cuando el 50% de la población está constituido por copias de la cadena de valor máximo, y como medida de la dificultad se utiliza el número de generaciones ( $NG$ ) necesarias para alcanzar el óptimo.

<i>orden</i>	<i>función</i>	$\varepsilon_{64}^*$	$NG$
1	$\mathfrak{R}_0^6$	0	37
	$T_{64}^1$	0	41
2	$\mathfrak{R}_1^6$	0.0285	68
	$T_{64}^2$	0.0147	46
4	$\mathfrak{R}_2^6$	0.3548	164
	$T_{64}^4$	0.2134	63
8	$\mathfrak{R}_3^6$	0.9391	> 1200
	$T_{64}^8$	0.8505	99
16	$\mathfrak{R}_4^6$	0.9997	> 1200
	$T_{64}^{16}$	0.9985	> 1200

Tabla 2.7: Comparativa entre las funciones Template y Royal Road generalizadas.

Se observa que, fijado un mismo orden para ambas funciones, a medida que éste aumenta, crecen el valor de la epistasis normalizada y el número de generaciones necesarias para alcanzar el máximo, con valores coherentes con la dificultad de las funciones (mayor  $NG$  en las funciones Royal Road que en las funciones Template para un orden fijo, salvo si  $orden = 1$  en el que ambas funciones son la misma y la diferencia en el número de generaciones se debe al componente probabilístico de este tipo de algoritmos). Esto parece indicar que el orden refuerza la utilidad de la epistasis como estimador de la dificultad.

A pesar de esta satisfactoria relación entre orden y epistasis, se pueden apreciar grandes diferencias entre el número generaciones necesitados por las funciones anteriores para alcanzar

el valor máximo. El caso más evidente es aquél en que se consideran funciones de orden 8, donde se obtiene un valor de 99 generaciones de poblaciones para la función Template, frente a las más de 1200 que necesitaron las Royal Road.

Desde el punto de vista práctico, queda por resolver todavía cómo caracterizar de una forma simple el orden de una función. El caso de las funciones de primer orden se resolvió satisfactoriamente en [33] con el uso de transformadas de Walsh. En la sección 2.6, nosotros completamos este estudio para funciones de orden superior a uno. Lo haremos inspirándonos en las ideas de Goldberg ([12]), quien aplicó con éxito el análisis de Walsh para calcular idoneidades medias de esquemas. Esta caracterización a nosotros nos ha permitido hallar una forma muy eficiente de calcular su epistasis.

## 2.5 Orden y decepción

Dos de las medidas de dificultad analizadas en esta memoria, decepción y epistasis, han sido relacionadas ya en [33]. En ese trabajo se calcula la epistasis normalizada de cualquier función de ajuste  $f : \Omega_2 \rightarrow \mathbb{R}$ , con el objetivo de relacionar epistasis y decepción en las funciones con las que Goldberg describió el problema decepcionante mínimo. Se comprueba que el valor de la epistasis normalizada es proporcional al cuadrado de  $w_3$ , el único coeficiente de Walsh de orden dos de  $f$ . Concretamente, si  ${}^t\mathbf{f} = (f_{00}, f_{01}, f_{10}, f_{11})$ , entonces  $\varepsilon^*(f) = \frac{1}{4}(f_{00} - f_{01} - f_{10} + f_{11})^2$ . Suponiendo que  $f$  alcanza el valor máximo en la cadena 11 y en función de las relaciones entre  $f_{00}$ ,  $f_{01}$  y  $f_{10}$ , los autores clasifican la epistasis normalizada de  $f$  en tres tipos: baja, media y alta, comprobando que si la epistasis es baja la función no es decepcionante; si la epistasis tiene valor medio la función es decepcionante de tipo I, mientras que si la epistasis es alta la función es decepcionante de tipo II.

En dicho trabajo también se comprobó que las funciones definidas en  $\Omega = \{0, 1\}^2$  con valor mínimo de la epistasis normalizada, es decir, las funciones de orden 1, no pueden ser completamente decepcionantes de orden 1. A continuación, nosotros generalizaremos este resultado, concluyendo que no puede ocurrir ningún tipo de decepción en las funciones de primer orden.

**Proposición 2.5.1.** *Si  $f : \Omega_\ell \rightarrow \mathbb{R}$  es función de primer orden, entonces  $f$  no contiene decepción.*

*Demostración.* Si  $f$  es función de primer orden, se puede escribir como suma de funciones simples de orden 1, es decir:

$$f(s) = \sum_{i=0}^{\ell-1} g_i(s_i), \text{ con } g_i(s_i) = \begin{cases} a_i & \text{si } s_i = 0, \\ b_i & \text{si } s_i = 1. \end{cases} \quad (2.14)$$

Sin pérdida de generalidad, se puede suponer que el máximo de  $f$  es único y se alcanza para la cadena  $1^{(\ell)} = \overbrace{1\dots 1}^{\ell}$ . Así,  $f(1^{(\ell)}) > f(s)$  para toda cadena  $s \in \Omega_\ell$ ,  $s \neq 1^{(\ell)}$ , en particular para las cadenas  $(1\dots 101\dots 1)$ . Por tanto, para todo  $k = 0, \dots, \ell - 1$ , se cumple:

$$\sum_{i=0}^{\ell-1} b_i = f(1^{(\ell)}) > f(1\dots 1\underset{\downarrow k}{0}1\dots 1) = a_k + \sum_{\substack{i=0 \\ i \neq k}}^{\ell-1} b_i$$

y, en consecuencia:  $b_k > a_k$ .

Consideremos ahora una partición  $\mathcal{P}$  formada por esquemas competitivos, de orden  $n$ , tales que las posiciones definidas de los esquemas de  $\mathcal{P}$  son los elementos del conjunto  $I = \{i_1, \dots, i_n\}$ . Entonces, si para cada  $J \subset I$  denotamos por  $\mathcal{H}_J^{I-J}$  el conjunto de cadenas

$$\mathcal{H}_J^{I-J} = \{s \in \Omega_\ell; s_j = 0 \text{ para } j \in J, s_k = 1 \text{ para } k \in I - J\},$$

es fácil ver que  $\mathcal{P} = \{\mathcal{H}_J^{I-J}; J \subset I\}$ . Por ejemplo, si  $\ell = 3$ , para  $n = 2$  y  $\mathcal{P} = \{\#00, \#01, \#10, \#11\}$  ( $I = \{i_1, i_2\} = \{0, 1\}$ ) entonces  $\mathcal{H}^I = \mathcal{H}_{\emptyset}^{\{0,1\}} = \#11$ ,  $\mathcal{H}_{\{0\}}^{\{1\}} = \#10$ ,  $\mathcal{H}_{\{1\}}^{\{0\}} = \#01$  y  $\mathcal{H}_{\{0,1\}}^{\emptyset} = \#00$ .

Si el esquema ganador de  $\mathcal{P}$  es  $\mathcal{H}^I = \# \dots \# \underset{\downarrow i_n}{1} \# \dots \# \underset{\downarrow i_{n-1}}{1} \# \dots \# \underset{\downarrow i_1}{1} \# \dots \#$ , la función  $f$  no puede contener decepción de orden  $n$ . En efecto, dado cualquier esquema  $\mathcal{H}_J^{I-J}$  de  $\mathcal{P}$  con  $J \neq \emptyset$ , se verifica:

$$\begin{aligned} f(\mathcal{H}_J^{I-J}) &= \frac{1}{2^{\ell-n}} \sum_{s \in \mathcal{H}_J^{I-J}} f(s) = \frac{1}{2^{\ell-n}} \sum_{s \in \mathcal{H}_J^{I-J}} \sum_{i=0}^{\ell-1} g_i(s_i) = \frac{1}{2^{\ell-n}} \sum_{i=0}^{\ell-1} \sum_{s \in \mathcal{H}_J^{I-J}} g_i(s_i) \\ &= \frac{1}{2^{\ell-n}} \left( \sum_{j \in J} \sum_{s \in \mathcal{H}_J^{I-J}} g_j(s_j) + \sum_{k \in I-J} \sum_{s \in \mathcal{H}_J^{I-J}} g_k(s_k) + \sum_{k \notin I} \sum_{s \in \mathcal{H}_J^{I-J}} g_k(s_k) \right) \\ &= \frac{1}{2^{\ell-n}} \left[ 2^{\ell-n} \sum_{j \in J} a_j + 2^{\ell-n} \sum_{k \in I-J} b_k + 2^{\ell-n-1} \sum_{k \notin I} (a_k + b_k) \right] \\ &= \sum_{j \in J} a_j + \sum_{k \in I-J} b_k + \frac{1}{2} \sum_{k \notin I} (a_k + b_k) \end{aligned}$$

$$\begin{aligned} &< \sum_{j \in J} b_j + \sum_{k \in I-J} b_k + \frac{1}{2} \sum_{k \notin I} (a_k + b_k) \\ &= f(\mathcal{H}^I). \end{aligned}$$

□

Las funciones de segundo orden no presentan un comportamiento tan satisfactorio como las de orden 1, en cuanto a la decepción. Es fácil construir funciones de segundo orden que contengan decepción –de hecho, el MDP de tipo II propuesto por Goldberg es un ejemplo de ello–. Sin embargo, sí podemos afirmar que las funciones de segundo orden no pueden ser completamente decepcionantes.

**Proposición 2.5.2.** *Si  $f : \Omega_\ell \rightarrow \mathbb{R}$  es una función de segundo orden, entonces no puede ser completamente decepcionante de primer orden.*

*Demostración.* Si  $f$  es función de segundo orden, existen funciones básicas de primer orden  $g_i$ ,  $0 \leq i < \ell$ , y de segundo orden  $g_{ij}$ ,  $0 \leq i < j < \ell$ , tales que  $f$  se puede escribir como combinación lineal de ellas, es decir:

$$f(s) = \sum_{0 \leq i < \ell} g_i(s_i) + \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j)$$

con  $g_i$  como en (2.14) y

$$g_{ij}(s_i, s_j) = \begin{cases} a_{ij} & \text{si } s_j = 0, s_i = 0 \\ b_{ij} & \text{si } s_j = 0, s_i = 1 \\ c_{ij} & \text{si } s_j = 1, s_i = 0 \\ d_{ij} & \text{si } s_j = 1, s_i = 1. \end{cases}$$

Supongamos que  $f(1^{(\ell)}) > f(s)$  para toda cadena  $s \in \Omega_\ell$ ,  $s \neq 1^{(\ell)}$ . Si la función es completamente decepcionante de primer orden, entonces debe suceder que, para todo  $k \in \{0, \dots, \ell - 1\}$ ,  $f(\mathcal{H}_k) > f(\mathcal{H}^k)$ , siendo  $\mathcal{H}_k = \# \dots \overset{k}{\#0} \# \dots \#$  y  $\mathcal{H}^k = \# \dots \overset{k}{\#1} \# \dots \#$ . Así, teniendo en cuenta que

$$\begin{aligned} f(\mathcal{H}_k) &= \frac{1}{2^{\ell-1}} \sum_{s \in \mathcal{H}_k} f(s) = \frac{1}{2^{\ell-1}} \sum_{s \in \mathcal{H}_k} \left[ \sum_{i=0}^{\ell-1} g_i(s_i) + \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j) \right] \\ &= \frac{1}{2^{\ell-1}} \left[ \sum_{i=0}^{\ell-1} \sum_{s \in \mathcal{H}_k} g_i(s_i) + \sum_{0 \leq i < j < \ell} \sum_{s \in \mathcal{H}_k} g_{ij}(s_i, s_j) \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2^{\ell-1}} \left[ 2^{\ell-1} a_k + 2^{\ell-2} \sum_{\substack{i=0 \\ i \neq k}}^{\ell-1} (a_i + b_i) + 2^{\ell-3} \sum_{\substack{0 \leq i < j < \ell \\ i, j \neq k}} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) \right. \\
&\quad \left. + 2^{\ell-2} \sum_{\substack{i=0 \\ i < k}}^{\ell-1} (a_{ik} + b_{ik}) + 2^{\ell-2} \sum_{\substack{j=0 \\ k < j}}^{\ell-1} (a_{kj} + c_{kj}) \right] \\
&= a_k + \frac{1}{2} \sum_{\substack{i=0 \\ i \neq k}}^{\ell-1} (a_i + b_i) + \frac{1}{4} \sum_{\substack{0 \leq i < j < \ell \\ i, j \neq k}} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) \\
&\quad + \frac{1}{2} \sum_{\substack{i=0 \\ i < k}}^{\ell-1} (a_{ik} + b_{ik}) + \frac{1}{2} \sum_{\substack{j=0 \\ k < j}}^{\ell-1} (a_{kj} + c_{kj})
\end{aligned}$$

y que, de forma análoga,

$$\begin{aligned}
f(\mathcal{H}^k) &= b_k + \frac{1}{2} \sum_{\substack{i=0 \\ i \neq k}}^{\ell-1} (a_i + b_i) + \frac{1}{4} \sum_{\substack{0 \leq i < j < \ell \\ i, j \neq k}} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) \\
&\quad + \frac{1}{2} \sum_{\substack{i=0 \\ i < k}}^{\ell-1} (c_{ik} + d_{ik}) + \frac{1}{2} \sum_{\substack{j=0 \\ k < j}}^{\ell-1} (b_{kj} + d_{kj}),
\end{aligned}$$

resulta evidente que  $f(\mathcal{H}_k) > f(\mathcal{H}^k)$  si y sólo si

$$a_k + \frac{1}{2} \left[ \sum_{\substack{i=0 \\ i < k}}^{\ell-1} (a_{ik} + b_{ik}) + \sum_{\substack{i=0 \\ k < j}}^{\ell-1} (a_{kj} + c_{kj}) \right] > b_k + \frac{1}{2} \left[ \sum_{\substack{i=0 \\ i < k}}^{\ell-1} (c_{ik} + d_{ik}) + \sum_{\substack{i=0 \\ k < j}}^{\ell-1} (b_{kj} + d_{kj}) \right].$$

Entonces:

$$\sum_{k=0}^{\ell-1} \left( a_k + \sum_{\substack{i=0 \\ i < k}}^{\ell-1} \frac{a_{ik} + b_{ik}}{2} + \sum_{\substack{i=0 \\ k < j}}^{\ell-1} \frac{a_{kj} + c_{kj}}{2} \right) > \sum_{k=0}^{\ell-1} \left( b_k + \sum_{\substack{i=0 \\ i < k}}^{\ell-1} \frac{c_{ik} + d_{ik}}{2} + \sum_{\substack{i=0 \\ k < j}}^{\ell-1} \frac{b_{kj} + d_{kj}}{2} \right),$$

o, equivalentemente,

$$\sum_{k=0}^{\ell-1} a_k + \sum_{0 \leq i < j < \ell} a_{ij} + \frac{1}{2} \sum_{0 \leq i < j < \ell} (b_{ij} + c_{ij}) > \sum_{k=0}^{\ell-1} b_k + \sum_{0 \leq i < j < \ell} d_{ij} + \frac{1}{2} \sum_{0 \leq i < j < \ell} (b_{ij} + c_{ij}).$$

En consecuencia,

$$\sum_{k=0}^{\ell-1} a_k + \sum_{0 \leq i < j < \ell} a_{ij} > \sum_{k=0}^{\ell-1} b_k + \sum_{0 \leq i < j < \ell} d_{ij},$$

lo que equivale a  $f(0^{(\ell)}) > f(1^{(\ell)})$ , contradiciendo la hipótesis inicial sobre el máximo de  $f$ .  $\square$

El resultado anterior nos permite demostrar ahora:

**Proposición 2.5.3.** *Una función  $f : \Omega_\ell \rightarrow \mathbb{R}$  de segundo orden no puede ser completamente decepcionante de orden  $n$ , para ningún  $n$ .*

*Demostración.* Sea  $f$  función de segundo orden y sea  $1 \leq n \leq \ell$ . En lo que sigue se considerará

$$f(s) = \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j), \text{ con } g_{ij}(s_i, s_j) = \begin{cases} a_{ij} & \text{si } s_j = 0, s_i = 0 \\ b_{ij} & \text{si } s_j = 0, s_i = 1 \\ c_{ij} & \text{si } s_j = 1, s_i = 0 \\ d_{ij} & \text{si } s_j = 1, s_i = 1. \end{cases} \quad (2.15)$$

que es una función que denominaremos *función simple* de orden 2. Esto supone una gran simplificación en los cálculos y no resta generalidad al resultado.

Si  $f$  es completamente decepcionante de orden  $n$ , entonces los ganadores de todas las particiones de orden menor que  $n$  tienen al menos un bit diferente al ganador de la correspondiente partición de orden  $n$ , en particular los de orden  $n - 1$ .

Sea  $\mathcal{P}$  una partición formada por esquemas competitivos de orden  $n$  en la que, sin pérdida de generalidad, suponemos definidas las posiciones del conjunto  $I = \{0, \dots, n - 1\}$ . Sin que ello suponga ninguna restricción, consideraremos que el esquema ganador de  $\mathcal{P}$  es  $\mathcal{H}^I = \# \dots \# \overbrace{1 \dots 1}^n = \{s \in \Omega, s_i = 1, 0 \leq i \leq n - 1\}$ . Entonces se verifica:

$$\begin{aligned} f(\mathcal{H}^I) &= \frac{1}{2^{\ell-n}} \sum_{s \in \mathcal{H}^I} f(s) = \frac{1}{2^{\ell-n}} \sum_{s \in \mathcal{H}^I} \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j) = \frac{1}{2^{\ell-n}} \sum_{0 \leq i < j < \ell} \sum_{s \in \mathcal{H}^I} g_{ij}(s_i, s_j) \\ &= \frac{1}{2^{\ell-n}} \left[ 2^{\ell-n-2} \sum_{j, i \geq n} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) + 2^{\ell-n-1} \sum_{i < n \leq j} (b_{ij} + d_{ij}) \right. \\ &\quad \left. + 2^{\ell-n} \sum_{i, j < n} d_{ij} \right] \\ &= \frac{1}{4} \sum_{j, i \geq n} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) + \frac{1}{2} \sum_{i < n \leq j} (b_{ij} + d_{ij}) + \sum_{i, j < n} d_{ij}. \end{aligned}$$

Cualquier esquema de  $\mathcal{P}$  diferente de  $\mathcal{H}^I$  está caracterizado por un subconjunto  $J$  de  $I$ . Así, estos representantes de  $\mathcal{P}$  se pueden escribir en la forma  $\mathcal{H}_J^{I-J} = \{s \in \Omega_\ell, s_j = 0, j \in$

$J, s_i = 1, i \in I - J\}$ , para algún subconjunto  $I$  de índices. Además,

$$\begin{aligned}
f(\mathcal{H}_J^{I-J}) &= \frac{1}{2^{\ell-n}} \sum_{s \in \mathcal{H}_J^{I-J}} f(s) = \frac{1}{2^{\ell-n}} \sum_{s \in \mathcal{H}_J^{I-J}} \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j) \\
&= \frac{1}{2^{\ell-n}} \sum_{0 \leq i < j < \ell} \sum_{s \in \mathcal{H}_J^{I-J}} g_{ij}(s_i, s_j) \\
&= \frac{1}{2^{\ell-n}} \left[ 2^{\ell-n-2} \sum_{j, i \geq n} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) \right. \\
&\quad \left. + 2^{\ell-n-1} \sum_{\substack{j \geq n \\ i \in J}} (a_{ij} + c_{ij}) + 2^{\ell-n-1} \sum_{\substack{j \geq n \\ i \in I-J}} (b_{ij} + d_{ij}) \right. \\
&\quad \left. + 2^{\ell-n} \sum_{i, j \in J} a_{ij} + 2^{\ell-n} \sum_{\substack{i \in J \\ j \in I-J}} c_{ij} + 2^{\ell-n} \sum_{\substack{i \in I-J \\ j \in J}} b_{ij} + 2^{\ell-n} \sum_{i, j \in I-J} d_{ij} \right] \\
&= \frac{1}{4} \sum_{j, i \geq n} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) + \frac{1}{2} \sum_{\substack{j \geq n \\ i \in J}} (a_{ij} + c_{ij}) + \frac{1}{2} \sum_{\substack{j \geq n \\ i \in I-J}} (b_{ij} + d_{ij}) \\
&\quad + \sum_{i, j \in J} a_{ij} + \sum_{\substack{i \in I-J \\ j \in J}} b_{ij} + \sum_{\substack{i \in J \\ j \in I-J}} c_{ij} + \sum_{i, j \in I-J} d_{ij}.
\end{aligned}$$

Con los coeficientes  $a_{ij}$ ,  $b_{ij}$ ,  $c_{ij}$  y  $d_{ij}$  que se usan en (2.15) para definir  $f$ , se construye una nueva función de segundo orden  $h : \Omega_n \rightarrow \mathbb{R}$ ,  $h(s_{n-1}, \dots, s_0) = \sum_{0 \leq i < j < n} h_{ij}(s_i, s_j) + \sum_{0 \leq i < n} h_i(s_i)$ , dada por:

$$h_{ij}(s_i, s_j) = \begin{cases} a_{ij} & \text{si } s_j = 0, s_i = 0 \\ b_{ij} & \text{si } s_j = 0, s_i = 1 \\ c_{ij} & \text{si } s_j = 1, s_i = 0 \\ d_{ij} & \text{si } s_j = 1, s_i = 1 \end{cases}$$

$$h_i(s_i) = \begin{cases} \frac{1}{2} \sum_{j \geq n} (a_{ij} + c_{ij}) & \text{si } s_i = 0 \\ \frac{1}{2} \sum_{j \geq n} (b_{ij} + d_{ij}) & \text{si } s_i = 1. \end{cases}$$

Para cualquier esquema de la partición  $\mathcal{P}$  se verifica entonces:

$$f(\# \dots \# s_{n-1} \dots s_0) = \frac{1}{4} \sum_{j, i \geq n} (a_{ij} + b_{ij} + c_{ij} + d_{ij}) + h(s_{n-1} \dots s_0),$$

por lo que si  $f$  es completamente decepcionante de orden  $n$ , entonces  $h$  sería completamente decepcionante de orden 1, lo que, por la proposición 2.5.2 no es posible al ser  $h$  función de segundo orden.  $\square$

## 2.6 Coeficientes de Walsh de las funciones de orden $k$

### 2.6.1 Caracterización de las funciones de segundo orden en términos de sus coeficientes de Walsh

Las funciones de segundo orden son aquellas que, como ya se ha mencionado, se pueden escribir en la forma:

$$f(s_{\ell-1}, \dots, s_0) = \sum_{0 \leq i < \ell} g_i(s_i) + \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j),$$

para ciertas funciones  $g_i$  y  $g_{ij}$  con valores reales, que sólo dependen, respectivamente, de una y dos posiciones.

Debido a la simplicidad de estas funciones, no es difícil calcular sus coeficientes de Walsh. De hecho, en [40] se demuestra que los coeficientes  $w_t$  no nulos son aquellos que verifican  $t = 0, 2^i, 2^i + 2^j$ , para  $0 \leq i < j < \ell$ .

Ahora nosotros podemos demostrar:

**Proposición 2.6.1.** *Si  $f$  es una función cuyos coeficientes de Walsh  $w_t$  son nulos para  $t \neq 0, 2^i, 2^i + 2^j$ , con  $0 \leq i < j < \ell$ , entonces  $f$  es una función de segundo orden.*

*Demostración.* Consideremos los vectores  $\mathbf{w}^{(1)}$  y  $\mathbf{w}^{(2)}$  cuyas  $k$ -ésimas componentes  $\mathbf{w}_k^{(1)}$ , resp.  $\mathbf{w}_k^{(2)}$ , son:

$$\mathbf{w}_k^{(1)} = \begin{cases} \mathbf{w}_k & \text{para } k = 0, 2^i \quad (0 \leq i < \ell) \\ 0 & \text{en otro caso} \end{cases}$$

y

$$\mathbf{w}_k^{(2)} = \begin{cases} \mathbf{w}_k & \text{para } k = 2^i + 2^j \quad (0 \leq i < j < \ell) \\ 0 & \text{en otro caso,} \end{cases}$$

respectivamente.

Entonces  $\mathbf{w} = \mathbf{w}^{(1)} + \mathbf{w}^{(2)}$  y, como  $\mathbf{f} = \mathbf{W}_\ell \mathbf{w}$ , obtenemos:

$$\begin{aligned} f(s) &= (\mathbf{W}_\ell \mathbf{w})_s = 2^{-\ell/2} (\mathbf{V}_\ell \mathbf{w})_s \\ &= 2^{-\ell/2} \left( \mathbf{V}_\ell \mathbf{w}^{(1)} \right)_s + 2^{-\ell/2} \left( \mathbf{V}_\ell \mathbf{w}^{(2)} \right)_s \end{aligned}$$

$$\begin{aligned}
&= 2^{-\ell/2} \sum_{k \in \Omega_\ell} (-1)^{sk} \mathbf{w}_k^{(1)} + 2^{-\ell/2} \sum_{k \in \Omega_\ell} (-1)^{sk} \mathbf{w}_k^{(2)} \\
&= 2^{-\ell/2} w_0 + 2^{-\ell/2} \sum_{0 \leq i < \ell} (-1)^{s_i} w_{2^i} + 2^{-\ell/2} \sum_{0 \leq i < j < \ell} (-1)^{s_i + s_j} w_{2^i + 2^j}.
\end{aligned}$$

Ahora, basta definir las funciones  $g_i : \{0, 1\} \rightarrow \mathbb{R}$ , respectivamente  $g_{ij} : \{0, 1\} \times \{0, 1\} \rightarrow \mathbb{R}$ ,

$$g_i(s_i) = 2^{-\ell/2} \left[ \frac{1}{\ell} w_0 + (-1)^{s_i} w_{2^i} \right], \quad (0 \leq i < \ell)$$

y

$$g_{ij}(s_i, s_j) = 2^{-\ell/2} [(-1)^{s_i + s_j} w_{2^i + 2^j}], \quad (0 \leq i < j < \ell)$$

entonces,

$$f(s) = \sum_{0 \leq i < \ell} g_i(s_i) + \sum_{0 \leq i < j < \ell} g_{ij}(s_i, s_j),$$

como se quería. □

## 2.6.2 Coeficientes de Walsh de las funciones de cualquier orden

Dado que las transformadas de Walsh respetan la separación formal entre las partes de orden 1, 2, ...,  $k$  de la expresión dada en (2.13), para una función  $f$  arbitraria de orden  $k$ , basta considerar cada una de esas componentes separadamente.

Denotando por  $\mathcal{G}$  una función definida en  $\Omega_\ell$ , de orden  $k$  ( $k \leq \ell$ ) en la que no intervienen componentes de orden inferior a  $k$ ; es decir,  $\mathcal{G}(s) = \mathcal{G}_{i_1 \dots i_k}(s_{i_1}, \dots, s_{i_k})$  (a la que nos referiremos como *función simple de orden  $k$* ), demostraremos que los coeficientes de Walsh de  $\mathcal{G}$  de orden mayor que  $k$  son nulos y de ello se deducirá el resultado para una función arbitraria  $f$  de orden  $k$  dada por (2.13). Para ello necesitamos introducir alguna notación y hacer uso de algunos lemas, dos de los cuales se demuestran en el apéndice B.

En lo que sigue, dado  $1 \leq p \leq \ell$ , y fijado un conjunto de índices  $J = \{j_1, \dots, j_p\} \subset \{0, \dots, \ell - 1\}$  de cardinal  $p$ , para cada  $0 \leq n \leq p$ , consideraremos el conjunto

$$\mathbf{P}_J = \{(\alpha_n, \beta_{p-n}); \{\alpha_n, \beta_{p-n}\} \text{ es una partición de } J\} \subset \mathcal{P}(J) \times \mathcal{P}(J)$$

y denotaremos por  $\mathcal{H}_{\alpha_n}^{\beta_{p-n}} \in \{0, 1, \#\}^\ell$  el esquema de orden  $p$

$$\mathcal{H}_{\alpha_n}^{\beta_{p-n}} = \{t \in \Omega_\ell; t_i = 0 \text{ si } i \in \alpha_n, t_j = 1 \text{ si } j \in \beta_{p-n}\},$$

donde  $(\alpha_n, \beta_{p-n}) \in \mathbf{P}_J$ .

**Nota:** Los subíndices de los elementos de  $\mathcal{P}(J)$  indican los cardinales de los respectivos elementos de la partición.

En ocasiones será conveniente escribir explícitamente las posiciones fijas del esquema, en cuyo caso denotaremos por  $\mathcal{H}_{i_1 \dots i_n}^{j_1 \dots j_{p-n}}$  el esquema

$$\mathcal{H}_{i_1 \dots i_n}^{j_1 \dots j_{p-n}} = \{s \in \Omega_\ell; s_{i_1} = \dots = s_{i_n} = 0, s_{j_1} = \dots = s_{j_{p-n}} = 1\}.$$

Por otra parte,  $\mathcal{H}_{i_1 \dots i_p}$ , respectivamente  $\mathcal{H}^{j_1 \dots j_p}$ , corresponde al esquema

$$\mathcal{H}_{i_1 \dots i_p} = \{s \in \Omega_\ell; s_{i_1} = \dots = s_{i_p} = 0\}$$

respectivamente

$$\mathcal{H}^{j_1 \dots j_p} = \{s \in \Omega_\ell; s_{j_1} = \dots = s_{j_p} = 1\}.$$

Es fácil ver que  $\Omega_\ell$  (esquema de orden cero) se puede escribir como unión disjunta de dos esquemas de orden 1. Por ejemplo, fijado  $i$  y considerando los esquemas  $\mathcal{H}_i = \{s \in \Omega_\ell; s_i = 0\}$  y  $\mathcal{H}^i = \{s \in \Omega_\ell; s_i = 1\}$ , entonces  $\{\mathcal{H}_i, \mathcal{H}^i\}$  define una partición de  $\Omega_\ell$ . Además, para cualquier función de ajuste  $f : \Omega_\ell \rightarrow \mathbb{R}$ , si  $f(\Omega_\ell)$ ,  $f(\mathcal{H}_i)$  y  $f(\mathcal{H}^i)$  son los valores medios de  $\Omega_\ell$ ,  $\mathcal{H}_i$  y de  $\mathcal{H}^i$ , respectivamente, entonces:

$$f(\Omega_\ell) = \frac{1}{2^\ell} \sum_{s \in \Omega_\ell} f(s) = \frac{1}{2^\ell} \left[ \sum_{s \in \mathcal{H}_i} f(s) + \sum_{s \in \mathcal{H}^i} f(s) \right] = \frac{1}{2} [f(\mathcal{H}_i) + f(\mathcal{H}^i)].$$

En general, cualquier esquema de orden  $p$  se puede escribir como unión disjunta de dos esquemas de orden  $p+1$ . Entonces, en particular,  $f(\Omega_\ell) = \frac{1}{4} [f(\mathcal{H}_{ij}) + f(\mathcal{H}_i^j) + f(\mathcal{H}_j^i) + f(\mathcal{H}^{ij})]$ , por ejemplo. Iterando el proceso, tenemos:

**Lema 2.6.2.**  $\Omega_\ell$  se puede escribir como la unión de  $2^p$  esquemas de orden  $p$  disjuntos dos a dos.

*Demostración.* Sea  $E_n^{p-n}(J) = \bigcup_{\mathbf{P}_J} \mathcal{H}_{\alpha_n}^{\beta_{p-n}}$  ( $\binom{p}{n}$  esquemas disjuntos dos a dos), entonces:

$$\Omega_\ell = \bigcup_{0 \leq n \leq p} E_n^{p-n}(J) = \bigcup_{0 \leq n \leq p} \left( \bigcup_{\mathbf{P}_J} \mathcal{H}_{\alpha_n}^{\beta_{p-n}} \right).$$

□

Como consecuencia directa, la idoneidad del espacio de  $\Omega_\ell$  es la media aritmética de las idoneidades de dichos esquemas, pues

$$f(E_n^{p-n}(J)) = \frac{1}{2^{\ell-p} \binom{p}{n}} \sum_{s \in \bigcup_{\mathbf{P}_J} \mathcal{H}_{\alpha_n}^{\beta_{p-n}}} f(s)$$

$$= \frac{1}{2^{\ell-p} \binom{p}{n}} \sum_{\mathbf{P}_J} \sum_{s \in \mathcal{H}_{\alpha_n}^{\beta_{p-n}}} f(s) = \frac{1}{\binom{p}{n}} \sum_{\mathbf{P}_J} f(\mathcal{H}_{\alpha_n}^{\beta_{p-n}}) \quad (2.16)$$

y, entonces, la media  $f(\Omega_\ell)$  es:

$$f(\Omega_\ell) = \frac{1}{2^\ell} \sum_{n=0}^p \sum_{s \in E_n^{p-n}(J)} f(s) = \frac{1}{2^\ell} \sum_{n=0}^p 2^{\ell-p} \binom{p}{n} f(E_n^{p-n}(J)) = \frac{1}{2^p} \sum_{n=0}^p \sum_{\mathbf{P}_J} f(\mathcal{H}_{\alpha_n}^{\beta_{p-n}}). \quad (2.17)$$

Consideraremos ahora para el conjunto de índices  $J = \{j_1, \dots, j_p\} \subset \{0, \dots, \ell - 1\}$  y para una partición de  $J$  formada por dos elementos,  $Q = \{j_{\lambda_1}, \dots, j_{\lambda_q}\}$  y  $J - Q$ , el esquema de orden  $p$ ,  $\mathcal{H}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}} \in \bar{\Sigma}^\ell$ , dado por:

$$\mathcal{H}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}} = \{t \in \Omega_\ell; t_j = 0, j \in \alpha_m \cup \tilde{\alpha}_{n-m}, t_j = 1, j \in \beta_{q-m} \cup \tilde{\beta}_{p-n-(q-m)}\} \quad (2.18)$$

donde  $(\alpha_m, \beta_{q-m}) \in \mathbf{P}_Q$  y  $(\tilde{\alpha}_{n-m}, \tilde{\beta}_{p-n-(q-m)}) \in \mathbf{P}_{J-Q}$ . Como antes, los subíndices en (2.18) denotan los cardinales de los respectivos elementos de las particiones y en las particiones impropias se suprime el correspondiente índice.

**Ejemplo 2.6.3.** Si  $\ell = 4$ ,  $J = \{0, 2, 3\}$  y  $Q = \{0, 2\}$ , entonces tenemos:  $\mathcal{H}_{023}$ ,  $\mathcal{H}_{02}^3$  si  $m = 2$ ;  $\mathcal{H}_{03}^2$ ,  $\mathcal{H}_0^{23}$ ,  $\mathcal{H}_2^{03}$ ,  $\mathcal{H}_{23}^0$  si  $m = 1$  y  $\mathcal{H}_3^{02}$ ,  $\mathcal{H}^{023}$  para  $m = 0$ .

Sea ahora el conjunto de índices  $I = \{i_1, \dots, i_k\} \subset \{0, \dots, \ell - 1\}$ , para un valor  $k$  fijo ( $k \geq 1$ ). Si denotamos por  $\tilde{\mathcal{H}}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}}$  el esquema en  $\bar{\Sigma}^k$  obtenido de  $\mathcal{H}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}}$ , al considerar sólo los bits  $t_{i_1}, \dots, t_{i_k}$ , i.e.,  $\tilde{\mathcal{H}}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}} \in \Omega_k$ , de forma análoga al lema anterior, se tiene:

**Lema 2.6.4.** Cada esquema de orden  $q$  se puede escribir como unión de  $2^{p-q}$  esquemas de orden  $p$  disjuntos dos a dos.

*Demostración.* Para demostrar el resultado basta tener en cuenta que, dado un esquema  $\mathcal{H}_{\alpha_m}^{\beta_{q-m}}$  de orden  $q$  y dado un conjunto de índices  $J = \{j_1, \dots, j_{p-q}\}$  tales que  $J \cap \{\alpha_m \cup \beta_{q-m}\} = \emptyset$ , se puede escribir:

$$\mathcal{H}_{\alpha_m}^{\beta_{q-m}} = \bigcup_{\mathbf{P}_J} \mathcal{H}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}},$$

donde  $(\tilde{\alpha}_{n-m}, \tilde{\beta}_{p-n-(q-m)})$  es una partición de  $J$ . □

Consecuencia de este resultado es que:

$$f(\mathcal{H}_{\alpha_m}^{\beta_{q-m}}) = \frac{1}{2^{\ell-q}} \sum_{s \in \mathcal{H}_{\alpha_m}^{\beta_{q-m}}} f(s)$$

$$\begin{aligned}
 &= \frac{1}{2^{\ell-q}} \sum_{0 \leq n \leq p-q} \sum_{\mathbf{P}_J} \sum_{s \in \mathcal{H}_{\alpha_m}^{\beta_{q-m}} \tilde{\beta}_{p-n-(q-m)}} f(s) \\
 &= \frac{1}{2^{p-q}} \sum_{0 \leq n \leq p-q} \sum_{\mathbf{P}_J} f\left(\mathcal{H}_{\alpha_m}^{\beta_{q-m}} \tilde{\beta}_{p-n-(q-m)}\right).
 \end{aligned}$$

**Proposición 2.6.5.** Para cualquier conjunto  $J = \{j_1, \dots, j_p\} \subset \{0, \dots, \ell-1\}$ , el valor del coeficiente  $w_{2^{j_1} + \dots + 2^{j_p}}$  de una función simple de orden  $k$ ,  $\mathcal{G}$ , es:

$$w_{2^{j_1} + \dots + 2^{j_p}} = \begin{cases} \frac{2^{\ell/2}}{2^p} \sum_{n=0}^p (-1)^{p-n} \binom{p}{n} \mathcal{G}(E_n^{p-n}(J)) & \text{si } J \subset I, \\ 0 & \text{si } J \not\subset I. \end{cases}$$

*Demostración.* (Por inducción en  $p$ ). Por comodidad calculamos los valores  $v_{2^j} (= 2^{-\ell/2} w_{2^j})$ .

Para  $p = 1$ ,  $J = \{j\}$ . Entonces, si  $j \in I$ ,

$$\begin{aligned}
 v_{2^j} &= \mathcal{G}(\mathcal{H}_j) - v_0 = \mathcal{G}(\mathcal{H}_j) - \mathcal{G}(\Omega_\ell) = \mathcal{G}(\mathcal{H}_j) - \frac{1}{2} [\mathcal{G}(\mathcal{H}_j) + \mathcal{G}(\mathcal{H}^j)] \\
 &= \frac{1}{2} [\mathcal{G}(\mathcal{H}_j) - \mathcal{G}(\mathcal{H}^j)] = \frac{1}{2} [-\mathcal{G}(E_0^1(J)) + \mathcal{G}(E_1^0(J))].
 \end{aligned}$$

Si  $j \notin I$  entonces, como  $\mathcal{G}$  es una función simple de orden  $k$ , claramente  $\mathcal{G}(\mathcal{H}_j) = \mathcal{G}(\mathcal{H}^j)$  y, en consecuencia  $v_{2^j} = 0$ .

Supuesto ahora el resultado válido hasta  $p-1$ , para demostrarlo para  $p$  distinguimos dos casos:

**Caso 1.**  $J \subset I$ .

En primer lugar, no es difícil demostrar que –por ser  $\mathcal{G}$  una función simple de orden  $k$ –, dado un conjunto de índices  $Q \subset J$ , de cardinal  $q$ , entonces

$$\mathcal{G}(\mathcal{H}_{\alpha_m}^{\beta_{q-m}}) = \mathcal{G}(\tilde{\mathcal{H}}_{\alpha_m}^{\beta_{q-m}}), \tag{2.19}$$

para todo par  $(\alpha_m, \beta_{q-m}) \in \mathbf{P}_Q$ . En efecto,

$$\begin{aligned}
 \mathcal{G}(\mathcal{H}_{\alpha_m}^{\beta_{q-m}}) &= \frac{1}{2^{\ell-p}} \sum_{s \in \mathcal{H}_{\alpha_m}^{\beta_{q-m}}} \mathcal{G}(s) = \frac{1}{2^{\ell-p}} \sum_{s \in \tilde{\mathcal{H}}_{\alpha_m}^{\beta_{q-m}}} 2^{\ell-k} \mathcal{G}(s) \\
 &= \frac{1}{2^{k-p}} \sum_{s \in \tilde{\mathcal{H}}_{\alpha_m}^{\beta_{q-m}}} \mathcal{G}(s) = \mathcal{G}(\tilde{\mathcal{H}}_{\alpha_m}^{\beta_{q-m}}).
 \end{aligned}$$

En particular,

$$\mathcal{G}(\mathcal{H}_{j_1 \dots j_p}) = \mathcal{G}(\tilde{\mathcal{H}}_{j_1 \dots j_p}).$$

Además, si  $q < p$ , por la hipótesis de inducción y por (2.16) y (2.19),

$$\begin{aligned} v_{2^{j\lambda_1 + \dots + 2^{j\lambda_q}}} &= \frac{1}{2^q} \sum_{m=0}^q (-1)^{q-m} \binom{q}{m} \mathcal{G}(E_m^{q-m}(Q)) \\ &= \frac{1}{2^q} \sum_{m=0}^q (-1)^{q-m} \sum_{\mathbf{P}_Q} \mathcal{G}(\mathcal{H}_{\alpha_m}^{\beta_{q-m}}) \\ &= \frac{1}{2^q} \sum_{m=0}^q (-1)^{q-m} \sum_{\mathbf{P}_Q} \mathcal{G}(\tilde{\mathcal{H}}_{\alpha_m}^{\beta_{q-m}}) \tag{2.20} \\ &= \frac{1}{2^p} \sum_{n=0}^p \sum_{\mathbf{P}_Q \times \mathbf{P}_{J-Q}} (-1)^{q-m} \mathcal{G}(\tilde{\mathcal{H}}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{p-n-(q-m)}}) \\ &= \frac{1}{2^p} \sum_{n=0}^p \sum_{\mathbf{P}_J} (-1)^{q-m} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}), \end{aligned}$$

donde  $m$  es el cardinal de  $\mu_n \cap Q$  y  $q - m$  el de  $\tau_{p-n} \cap Q$ .

Además, por (2.17) y (2.20)

$$v_0 = \mathcal{G}(\Omega_\ell) = \frac{1}{2^p} \sum_{n=0}^p \sum_{\mathbf{P}_J} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}) = \frac{1}{2^p} \left( \mathcal{G}(\tilde{\mathcal{H}}_{j_1 \dots j_p}) + \sum_{n=0}^{p-1} \sum_{\mathbf{P}_J} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}) \right). \tag{2.21}$$

Ahora, para calcular  $v_{2^{j_1 + \dots + 2^{j_p}}}$  -utilizando (2.12) de la página 50- tenemos en cuenta previamente el valor de  $\Gamma = v_0 + \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} v_{2^{j\lambda_1 + \dots + 2^{j\lambda_q}}}$  que, por la hipótesis de inducción y por (2.20) y (2.21), es:

$$\begin{aligned} \Gamma &= \frac{1}{2^p} \left( \mathcal{G}(\tilde{\mathcal{H}}_{j_1 \dots j_p}) + \sum_{n=0}^{p-1} \sum_{\mathbf{P}_J} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}) \right) \\ &\quad + \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} \frac{1}{2^p} \sum_{n=0}^p \sum_{\mathbf{P}_J} (-1)^{q-m} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}) \\ &= \frac{1}{2^p} \left( \mathcal{G}(\tilde{\mathcal{H}}_{j_1 \dots j_p}) + \sum_{n=0}^{p-1} \sum_{\mathbf{P}_J} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}) \right) \\ &\quad + \frac{1}{2^p} \sum_{n=0}^p \sum_{\mathbf{P}_J} \left\{ \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} (-1)^{q-m} \mathcal{G}(\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}}) \right\}. \end{aligned}$$

Si  $n = p$ , en el último sumando de la expresión anterior  $q - m = 0$  y  $\tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} = \tilde{\mathcal{H}}_{j_1 \dots j_p}$ , por lo que,

$$\begin{aligned} \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right) &= \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right) \\ &= \sum_{q=1}^{p-1} \binom{p}{q} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right) \\ &= (2^p - 2) \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right). \end{aligned}$$

Por otra parte, si  $n < p$ , fijado un elemento  $(\mu_n, \tau_{p-n}) \in \mathbf{P}_J$  y un valor de  $q$  ( $q < p$ ), entonces

$$\sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right) = \sum_{m=0}^q \binom{n}{m} \binom{p-n}{q-m} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right), \quad (2.22)$$

donde, como antes,  $m$  denota el cardinal de  $\mu_n \cap Q$  y  $q - m$  el cardinal de  $\tau_{p-n} \cap Q$ , para todas las posibles elecciones de subconjuntos  $Q \subset J$  de cardinal  $q$  ( $Q = \{j_{\lambda_1}, \dots, j_{\lambda_q}\}$ )<sup>3</sup>. Así,

$$\sum_{\mathbf{P}_J} \left\{ \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right) \right\} = \sum_{\mathbf{P}_J} \gamma_{p,n} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right)$$

con

$$\gamma_{p,n} = \sum_{q=1}^{p-1} \sum_{m=0}^q (-1)^{q-m} \binom{n}{m} \binom{p-n}{q-m}.$$

Ahora, con la ayuda del lema B.2.1 (página 164), se obtiene de forma directa:

$$\begin{aligned} v_{2j_1 + \dots + 2j_p} &= \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right) - \Gamma \\ &= \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right) - \frac{2^p - 1}{2^p} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right) - \frac{1}{2^p} \sum_{n=0}^{p-1} \sum_{\mathbf{P}_J} (1 + \gamma_{p,n}) \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right) \\ &= \frac{1}{2^p} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_p} \right) - \frac{1}{2^p} \sum_{n=0}^{p-1} \sum_{\mathbf{P}_J} (-1)^{p-n+1} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right) \\ &= \frac{1}{2^p} \sum_{n=0}^p (-1)^{p-n} \sum_{\mathbf{P}_J} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{p-n}} \right) \end{aligned}$$

---

<sup>3</sup>En (2.22) hemos utilizado que, fijado un elemento de  $\mathbf{P}_J$ , existen  $\binom{p}{q}$  sumandos en la primera parte de la expresión (2.22) cada uno de los cuales va precedido del signo  $(-1)^{q-m}$ , determinado por la cardinalidad de  $Q \cap \tau_{p-n}$ .

$$= \frac{1}{2^p} \sum_{n=0}^p (-1)^{p-n} \binom{p}{n} \mathcal{G} (E_n^{p-n}(J)).$$

**Caso 2.**  $J = \{j_1, \dots, j_p\} \not\subset I$ .

Consideremos, sin que ello suponga restricción, que  $R = \{j_1, \dots, j_r\} \subset I$ ,  $\{j_{r+1}, \dots, j_p\} \cap I = \emptyset$ . Un razonamiento análogo al del caso 1 muestra ahora que  $\mathcal{G} (\mathcal{H}_{j_1 \dots j_p}) = \mathcal{G} (\tilde{\mathcal{H}}_{j_1 \dots j_r})$ . En efecto:

$$\begin{aligned} \mathcal{G} (\mathcal{H}_{j_1 \dots j_p}) &= \frac{1}{2^{\ell-p}} \sum_{s \in \mathcal{H}_{j_1 \dots j_p}} \mathcal{G}(s) \\ &= \frac{1}{2^{\ell-p}} \sum_{s \in \tilde{\mathcal{H}}_{j_1 \dots j_r}} 2^{\ell-k-(p-r)} \mathcal{G}(s) \\ &= \frac{1}{2^{k-r}} \sum_{s \in \tilde{\mathcal{H}}_{j_1 \dots j_r}} \mathcal{G}(s) \\ &= \mathcal{G} (\tilde{\mathcal{H}}_{j_1 \dots j_r}). \end{aligned}$$

Por otra parte, por la hipótesis de inducción:

$$\begin{aligned} v_{2^{j_1} + \dots + 2^{j_p}} &= \mathcal{G}(\mathcal{H}_{j_1 \dots j_p}) - \left( v_0 + \sum_{q=1}^{p-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq p} v_{2^{j_{\lambda_1}} + \dots + 2^{j_{\lambda_q}}} \right) \\ &= \mathcal{G}(\mathcal{H}_{j_1 \dots j_p}) - \left( v_0 + \sum_{q=1}^r \sum_{\substack{1 \leq \lambda_1 < \dots < \lambda_q \leq r \\ j_{\lambda_i} \in R}} v_{2^{j_{\lambda_1}} + \dots + 2^{j_{\lambda_q}}} \right) \\ &= \mathcal{G}(\mathcal{H}_{j_1 \dots j_p}) - \Gamma. \end{aligned}$$

Para un conjunto  $Q = \{j_{\lambda_1}, \dots, j_{\lambda_q}\} \subset R$ , de cardinal  $q$ , la hipótesis de inducción junto con un razonamiento análogo al del caso 1 nos conduce ahora a que:

$$\begin{aligned} v_{2^{j_{\lambda_1}} + \dots + 2^{j_{\lambda_q}}} &= \frac{1}{2^r} \sum_{n=0}^r \sum_{\mathbf{P}_Q \times \mathbf{P}_{R-Q}} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\alpha_m \tilde{\alpha}_{n-m}}^{\beta_{q-m} \tilde{\beta}_{r-n-(q-m)}} \right) \\ &= \frac{1}{2^r} \sum_{n=0}^r \sum_{\mathbf{P}_R} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right), \end{aligned}$$

siendo  $m$  el cardinal de  $\mu_n \cap Q$  y  $q-m$  el de  $\tau_{r-n} \cap Q$ . Así,

$$\begin{aligned} \Gamma &= \frac{1}{2^r} \sum_{n=0}^r \sum_{\mathbf{P}_R} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) + \frac{1}{2^r} \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} \left( \sum_{n=0}^r \sum_{\mathbf{P}_R} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) \right) \\ &= \frac{1}{2^r} \left\{ \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) + \sum_{n=0}^{r-1} \sum_{\mathbf{P}_R} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) \right\} \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2^r} \sum_{n=0}^r \sum_{\mathbf{P}_R} \left\{ \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) \right\} \\
 = & \frac{1}{2^r} \left\{ \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) + \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) \right\} \\
 & + \frac{1}{2^r} \left\{ \sum_{n=0}^{r-1} \sum_{\mathbf{P}_R} \left( 1 + \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} (-1)^{q-m} \right) \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) \right\} \\
 = & \mathcal{A} + \mathcal{B}
 \end{aligned}$$

donde,

$$\begin{aligned}
 \mathcal{A} & = \frac{1}{2^r} \left\{ \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) + \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) \right\} \\
 & = \frac{1}{2^r} \left\{ \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) + \sum_{q=1}^r \binom{q}{r} \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) \right\} \\
 & = \frac{1}{2^r} \sum_{q=0}^r \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) = \mathcal{G} \left( \tilde{\mathcal{H}}_{j_1 \dots j_r} \right) = \mathcal{G} \left( \mathcal{H}_{j_1 \dots j_r} \right),
 \end{aligned}$$

(ya que si  $n = r$  entonces  $\tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} = \tilde{\mathcal{H}}_{j_1 \dots j_r}$  y, en consecuencia,  $q - m = 0$ )

y

$$\mathcal{B} = \frac{1}{2^r} \left\{ \sum_{n=0}^{r-1} \sum_{\mathbf{P}_R} \left( 1 + \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} (-1)^{q-m} \right) \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) \right\}. \quad (2.23)$$

Para finalizar la demostración basta tener en cuenta que en los sumandos de la expresión  $\mathcal{B}$  se cumple  $n < r$  y entonces el razonamiento empleado en el caso 1 nos conduce a que:

$$\begin{aligned}
 \sum_{\mathbf{P}_R} \sum_{q=1}^r \sum_{1 \leq \lambda_1 < \dots < \lambda_q \leq r} (-1)^{q-m} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) & = \sum_{\mathbf{P}_R} \sum_{q=1}^r \left( \sum_{m=0}^q (-1)^{q-m} \binom{n}{m} \binom{r-n}{q-m} \right) \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right) \\
 & = \sum_{\mathbf{P}_R} \zeta_{r,n} \mathcal{G} \left( \tilde{\mathcal{H}}_{\mu_n}^{\tau_{r-n}} \right)
 \end{aligned}$$

con

$$\begin{aligned}
 \zeta_{r,n} & = \sum_{q=1}^r \sum_{m=0}^q (-1)^{q-m} \binom{n}{m} \binom{r-n}{q-m} \\
 & = \gamma_{r,n} + \sum_{m=0}^r (-1)^{r-m} \binom{n}{m} \binom{r-n}{r-m} \\
 & = [(-1)^{r-n+1} - 1] + (-1)^{r-n} = -1.
 \end{aligned} \quad (2.24)$$

Sustituyendo (2.24) en (2.23) se obtiene directamente que  $\mathcal{B} = 0$  y entonces  $\Gamma = \mathcal{A} = \mathcal{G}(\mathcal{H}_{j_1 \dots j_p})$ . Ahora, de (2.12) claramente  $v_{2^{j_1} + \dots + 2^{j_p}} = \mathcal{G}(\mathcal{H}_{j_1 \dots j_p}) - \Gamma = 0$ .  $\square$

**Corolario 2.6.6.** *Si  $\mathcal{G} : \Omega_\ell \rightarrow \mathbb{R}$  es una función simple de orden  $k$ , entonces los coeficientes de Walsh de  $\mathcal{G}$  de orden  $p$  son nulos, para cada  $p > k$ .*

Finalmente, obtenemos:

**Teorema 2.6.7.** *Una función de ajuste  $f : \Omega_\ell \rightarrow \mathbb{R}$ , es de orden  $k$  si, y sólo si, sus coeficientes de Walsh  $w_t$  son nulos para  $t \notin \{0, 2^{i_1} + \dots + 2^{i_j}; 0 \leq i_1 < \dots < i_j < \ell, 1 \leq j \leq k\}$ .*

*Demostración.* Por los resultados anteriores, sólo falta demostrar el recíproco del corolario anterior. Ahora bien, como

$$f(s) = (\mathbf{W}_\ell \mathbf{w})_s = w_0 + \sum_{j=1}^k \sum_{0 \leq i_1 < \dots < i_j < \ell} (-1)^{(s_{i_1} + \dots + s_{i_j})} w_{2^{i_1} + \dots + 2^{i_j}},$$

basta definir para  $0 \leq i_1 < \dots < i_j < \ell, j \geq 1$ ,

$$g_{i_1 \dots i_j}(s_{i_1} \dots s_{i_j}) = 2^{-\ell/2} \left( \frac{w_0}{k \binom{\ell}{j}} + (-1)^{(s_{i_1} + \dots + s_{i_j})} w_{2^{i_1} + \dots + 2^{i_j}} \right).$$

$$\text{Entonces, } f(s) = \sum_{j=1}^k \sum_{0 \leq i_1 < \dots < i_j < \ell} g_{i_1 \dots i_j}(s_{i_1} \dots s_{i_j}). \quad \square$$

## Capítulo 3

# 2-epistasia sobre alfabetos binarios

---

En este capítulo, y en el siguiente, extenderemos la teoría de la epistasia del capítulo 2. La motivación es doble: desde un punto de vista matemático la extensión del concepto es natural y no debería suponer demasiadas complicaciones. Por otra parte, la epistasia normalizada introducida por Van Hove en [42] puede distinguir funciones de primer orden de funciones de orden superior, pero no puede diferenciar de modo fiable entre grados de interacción de orden mayor que uno. No obstante, los experimentos muestran (ver tabla 2.7) que la epistasia de las funciones de orden  $n$  tiene un valor más bajo que la epistasia de funciones de orden  $n + 1$ . Además, en dos grandes clases de funciones –las funciones Template y las funciones Royal Road generalizadas– hemos comprobado que el orden de las funciones presenta una buena correlación con la dificultad de las mismas a ser optimizadas usando Algoritmos Genéticos.

Como se afirma en [24],

*“... parece demasiado ambicioso tratar de resumir toda la riqueza de la vasta cantidad de los diferentes problemas de búsqueda en un número que represente la distancia a una clase particular de problemas sencillos. Si los únicos problemas fáciles para un AG fuesen las funciones de primer orden, la idea de medir la distancia desde un conjunto parecería razonable ..., sin embargo, éste no es el caso...”*

Esto nos ha llevado a estudiar una nueva medida epistática que, generalizando a la epistasia normalizada, permita completar la información que ya se tiene sobre la dificultad de una función, sobre todo para funciones de orden superior a uno.

### 3.1 2-epistasia y su representaci3n matricial

#### 3.1.1 2-epistasia global de una funci3n

En este primer ep3grafe introducimos un nuevo estimador de la dificultad de optimizaci3n de una funci3n mediante un AG: la *2-epistasia normalizada*. Su definici3n sigue las ideas desarrolladas tanto en [4], sobre los c3lculos de valores medios esperados para ciertos esquemas, como en [42], para el posterior desarrollo algebraico.

Dada cualquier cadena binaria  $s = s_{\ell-1}\dots s_0 \in \Omega_\ell = \{0, 1\}^\ell$ , para definir su *2-epistasia* comencemos considerando el subconjunto de  $\Omega_\ell$ :

$$\Omega_{i,k}(a, b) = \{t \in \Omega_\ell; t_i = a, t_k = b\} = \#\dots\overset{k}{\downarrow}\#b\#\dots\overset{i}{\downarrow}\#a\#\dots\#,$$

para cualesquiera alelos  $a, b \in \{0, 1\}$  y posiciones  $i, k$  ( $i \neq k$ ); y tambi3n  $\Omega_i(a)$ , el conjunto de cadenas en  $\Omega_\ell$  cuyo  $i$ -3simo bit toma el valor  $a$ . Denotemos por  $f_{i,k}(a, b)$  el valor medio de la funci3n  $f$  en  $\Omega_{i,k}(a, b)$ , es decir:

$$f_{i,k}(a, b) = f(\Omega_{i,k}(a, b)) = \frac{1}{2^{\ell-2}} \sum_{t \in \Omega_{i,k}(a, b)} f(t).$$

Para un par de alelos  $(a, b)$ , y posiciones  $i, k \in \{0, \dots, \ell - 1\}$ , ( $i \neq k$ ) se define el *valor de exceso al3lico* del par  $(a, b)$  como:

$$E_{i,k}(a, b) = f_{i,k}(a, b) - \bar{f}.$$

donde  $\bar{f}$ , como siempre, denota el valor medio de la funci3n, i.e.,

$$\bar{f} = \frac{1}{2^\ell} \sum_{s \in \Omega_\ell} f(s).$$

Para  $s \in \Omega_\ell$ , definimos su *valor de exceso g3nico*:

$$EG(s) = \sum_{0 \leq i < k < \ell} E_{i,k}(s_i, s_k) - (\ell - 2) \sum_{0 \leq i < \ell} E_i(s_i),$$

donde, como en la secci3n 2.1,  $E_i(s_i)$  denota el valor de exceso al3lico de  $s_i$ . Entonces, el *valor g3nico previsto* es:

$$\tilde{f}(s) = EG(s) + \bar{f}.$$

**Definición 3.1.1.** La 2-epistasis de la cadena  $s \in \Omega_\ell$  es:

$$\tilde{\varepsilon}_\ell(s) = f(s) - \tilde{f}(s).$$

Como consecuencia de las definiciones anteriores, la 2-epistasis de la cadena  $s$  es:

$$\begin{aligned} \tilde{\varepsilon}_\ell(s) &= f(s) - \tilde{f}(s) \\ &= f(s) - [EG(s) + \bar{f}] \\ &= f(s) - \left[ \sum_{0 \leq i < k < \ell} E_{i,k}(s_i, s_k) - (\ell - 2) \sum_{0 \leq i < \ell} E_i(s_i) + \bar{f} \right] \\ &= f(s) - \left[ \sum_{0 \leq i < k < \ell} \{f_{i,k}(s_i, s_k) - \bar{f}\} - (\ell - 2) \sum_{0 \leq i < \ell} \{f_i(s_i) - \bar{f}\} + \bar{f} \right] \\ &= f(s) - \left[ \sum_{0 \leq i < k < \ell} \frac{1}{2^{\ell-2}} \sum_{t \in \Omega_{i,k}(s_i, s_k)} f(t) - \binom{\ell}{2} \bar{f} \right. \\ &\quad \left. - (\ell - 2) \sum_{0 \leq i < \ell} \frac{1}{2^{\ell-1}} \sum_{t \in \Omega_i(s_i)} f(t) + (\ell - 2) \binom{\ell}{1} \bar{f} + \bar{f} \right] \\ &= f(s) - \left[ \sum_{0 \leq i < k < \ell} \frac{1}{2^{\ell-2}} \sum_{t \in \Omega_{i,k}(s_i, s_k)} f(t) - (\ell - 2) \sum_{0 \leq i < \ell} \frac{1}{2^{\ell-1}} \sum_{t \in \Omega_i(s_i)} f(t) \right. \\ &\quad \left. - \left\{ \frac{\ell(\ell-1)}{2} - (\ell-2)\ell - 1 \right\} \frac{1}{2^\ell} \sum_{t \in \Omega_\ell} f(t) \right] \\ &= f(s) - \frac{1}{2^\ell} \left[ \sum_{0 \leq i < k < \ell} 4 \sum_{t \in \Omega_{i,k}(s_i, s_k)} f(t) \right. \\ &\quad \left. - (\ell - 2) \sum_{0 \leq i < \ell} 2 \sum_{t \in \Omega_i(s_i)} f(t) + \frac{(\ell-1)(\ell-2)}{2} \sum_{t \in \Omega_\ell} f(t) \right] \\ &= f(s) - \frac{1}{2^\ell} \{C_1 - C_2 + C_3\}, \end{aligned}$$

siendo, respectivamente,

$$C_1 = 4 \sum_{0 \leq i < k < \ell} \sum_{t \in \Omega_{i,k}(s_i, s_k)} f(t),$$

$$C_2 = 2(\ell - 2) \sum_{0 \leq i < \ell} \sum_{t \in \Omega_i(s_i)} f(t)$$

y

$$C_3 = \binom{\ell-1}{2} \sum_{t \in \Omega_\ell} f(t).$$

Dado que la suma  $\mathcal{C} = \mathcal{C}_1 - \mathcal{C}_2 + \mathcal{C}_3$  es una combinación lineal de los valores  $f(t)$ ,  $t \in \Omega_\ell$ ,

$$\mathcal{C} = \sum_{t \in \Omega_\ell} \lambda_t f(t),$$

para obtener cada coeficiente  $\lambda_t$  basta conocer, para cada  $t \in \Omega_\ell$ , el número de veces que aparece  $f(t)$  en  $\mathcal{C}_i$ , ( $i = 1, 2, 3$ ). Para ello tengamos en cuenta que, fijado  $s \in \Omega_\ell$ ,

- una cadena  $t$  del espacio de búsqueda pertenece al conjunto  $\Omega_{i,k}(s_i, s_k)$  si y sólo si  $d_{st}^\ell \leq \ell - 2$ . Ahora, como una tal cadena  $t$  pertenece a un total de  $\binom{\ell - d_{st}^\ell}{2}$  conjuntos  $\Omega_{i,k}(s_i, s_k)$ , con  $0 \leq i < k < \ell$ , tenemos directamente que el coeficiente de  $\mathcal{C}_1$  es  $4 \binom{\ell - d_{st}^\ell}{2}$ ,
- una condición necesaria y suficiente para que una cadena  $t \in \Omega_\ell$  sea un elemento del conjunto  $\Omega_i(s_i)$  es que  $d_{st}^\ell \leq \ell - 1$ . En ese caso,  $t$  pertenece a tantos conjuntos  $\Omega_i(s_i)$  como alelos comunes tiene con  $s$ ; es decir,  $t$  es elemento de  $\ell - d_{st}^\ell$  conjuntos  $\Omega_i(s_i)$  y, en consecuencia, el coeficiente de  $\mathcal{C}_2$  es  $2(\ell - 2) \binom{\ell - d_{st}^\ell}{1}$ .

Por todo ello,

$$\lambda_t = 4 \binom{\ell - d_{st}^\ell}{2} - 2(\ell - 2) \binom{\ell - d_{st}^\ell}{1} + \binom{\ell - 1}{2}$$

y entonces:

$$\tilde{\varepsilon}_\ell(s) = f(s) - \frac{1}{2^\ell} \sum_{t \in \Omega_\ell} \left[ 4 \binom{\ell - d_{st}^\ell}{2} - 2(\ell - 2) \binom{\ell - d_{st}^\ell}{1} + \binom{\ell - 1}{2} \right] f(t).$$

Podemos emplear ahora álgebra matricial para reescribir la 2-epistasis de una función  $f$  definida sobre  $\Omega_\ell$  y caracterizada por el vector  $\mathbf{f}$ . Denotando el vector epistático por  $\mathbf{b}_\ell$ :

$$\mathbf{b}_\ell = \begin{pmatrix} \tilde{\varepsilon}_\ell(00 \dots 0) \\ \tilde{\varepsilon}_\ell(00 \dots 1) \\ \vdots \\ \tilde{\varepsilon}_\ell(11 \dots 1) \end{pmatrix}$$

y por  $\mathbf{B}_\ell \in \mathcal{M}_{2^\ell}(\mathbb{Q})$  la matriz (simétrica) cuyo elemento genérico  $b_{st}^\ell$  es:

$$b_{st}^\ell = \frac{1}{2^\ell} \left[ 4 \binom{\ell - d_{st}^\ell}{2} - 2(\ell - 2) \binom{\ell - d_{st}^\ell}{1} + \binom{\ell - 1}{2} \right],$$

para cualesquiera  $0 \leq s, t < 2^\ell$ , se puede entonces escribir:

$$\mathbf{b}_\ell = \mathbf{f} - \mathbf{B}_\ell \mathbf{f} = (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f},$$

siendo  $\mathbf{I}_\ell$  la matriz identidad de orden  $2^\ell$ .

**Definición 3.1.2.** La 2-epistasis global de una función  $f$  se define como:

$$\tilde{\varepsilon}_\ell(f) = \|\mathbf{b}_\ell\|.$$

A continuación procedemos al cálculo de la 2-epistasis global de algunas funciones definidas sobre cadenas de longitud 4. Éstas son:  $\mathcal{L}$ , una función lineal,  $c$ , función camello y  $\delta$ , aquella cuya normalización proporciona la función de Dirac  $\delta_0$ .

Los valores que toman las funciones  $\mathcal{L}$ ,  $c$  y  $\delta$  sobre cadenas de longitud 4 se reflejan en la tabla 3.1.

cadenas	$\mathcal{L}$	$c$	$\delta$
0000	0	60	120
0001	1	0	0
0010	2	0	0
0011	3	0	0
0100	4	0	0
0101	5	0	0
0110	6	0	0
0111	7	0	0
1000	8	0	0
1001	9	0	0
1010	10	0	0
1011	11	0	0
1100	12	0	0
1101	13	0	0
1110	14	0	0
1111	15	60	0

$$\bar{\mathcal{L}} = \bar{c} = \bar{\delta} = 7.5$$

Tabla 3.1: Valores de las funciones  $\mathcal{L}$ ,  $c$  y  $\delta$  sobre cadenas de longitud 4.

En las tablas 3.2, 3.4 y 3.6 se recogen los valores de exceso de cada alelo y cada par de alelos, necesarios para obtener la 2-epistasis global de las funciones consideradas.

$i$	$a$	$\mathcal{L}_i(a)$	$E_i(a)$
0	0	7	-0.5
	1	8	0.5
1	0	6.5	-1
	1	8.5	1
2	0	5.5	-2
	1	9.5	2
3	0	3.5	-4
	1	11.5	4

$i, k$	$a, b$	$\mathcal{L}_{i,k}(a, b)$	$E_{i,k}(a, b)$
1, 0	0, 0	6	-1.5
	0, 1	7	-0.5
	1, 0	8	0.5
	1, 1	9	1.5
2, 0	0, 0	5	-2.5
	0, 1	6	-1.5
	1, 0	9	1.5
	1, 1	10	2.5
3, 0	0, 0	3	-4.5
	0, 1	4	-3.5
	1, 0	11	3.5
	1, 1	12	4.5
2, 1	0, 0	4.5	-3
	0, 1	6.5	-1
	1, 0	8.5	1
	1, 1	10.5	3
3, 1	0, 0	2.5	-5
	0, 1	4.5	-3
	1, 0	10.5	3
	1, 1	12.5	5
3, 2	0, 0	1.5	-6
	0, 1	5.5	-2
	1, 0	9.5	2
	1, 1	13.5	6

Tabla 3.2: Cálculo de los valores de exceso de alelos y pares de alelos de una función lineal.

La función  $\mathcal{L}$  tiene 2-epistasis nula, como se deduce de los cálculos de la tabla 3.3.

$s$	$\mathcal{L}(s)$	$EG(s)$	$\mathcal{L}(s)$	$\bar{\varepsilon}(s)$
0000	0	-7.5	0	0
0001	1	-6.5	1	0
0010	2	-5.5	2	0
0011	3	-4.5	3	0
0100	4	-3.5	4	0
0101	5	-2.5	5	0
0110	6	-1.5	6	0
0111	7	-0.5	7	0
1000	8	0.5	8	0
1001	9	1.5	9	0
1010	10	2.5	10	0
1011	11	3.5	11	0
1100	12	4.5	12	0
1101	13	5.5	13	0
1110	14	6.5	14	0
1111	15	7.5	15	0

Tabla 3.3: Cálculo de la 2-epistasis de las cadenas de  $\Omega_4$  para una función lineal.

$i$	$a$	$c_i(a)$	$E_i(a)$
0	0	7.5	0
	1	7.5	0
1	0	7.5	0
	1	7.5	0
2	0	7.5	0
	1	7.5	0
3	0	7.5	0
	1	7.5	0

$i, k$	$a, b$	$c_{i,k}(a, b)$	$E_{i,k}(a, b)$
1, 0	0, 0	15	7.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	15	7.5
2, 0	0, 0	15	7.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	15	7.5
3, 0	0, 0	15	7.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	15	7.5
2, 1	0, 0	15	7.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	15	7.5
3, 1	0, 0	15	7.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	15	7.5
3, 2	0, 0	15	7.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	15	7.5

Tabla 3.4: Cálculo de los valores de exceso de alelos y pares de alelos de una función camello.

Por otra parte, la 2-epistasia global de  $c$  es  $\tilde{\varepsilon}(c) = 42.43$ . Este valor se obtiene fácilmente de la tabla 3.5.

$s$	$c(s)$	$EG(s)$	$\tilde{c}(s)$	$\tilde{\varepsilon}(s)$
0000	60	30	37.5	22.5
0001	0	0	7.5	-7.5
0010	0	0	7.5	-7.5
0011	0	-15	-7.5	7.5
0100	0	0	7.5	-7.5
0101	0	-15	-7.5	7.5
0110	0	-15	-7.5	7.5
0111	0	0	7.5	-7.5
1000	0	0	7.5	-7.5
1001	0	-15	-7.5	7.5
1010	0	-15	-7.5	7.5
1011	0	0	7.5	-7.5
1100	0	-15	-7.5	7.5
1101	0	0	7.5	-7.5
1110	0	0	7.5	-7.5
1111	60	30	37.5	22.5

Tabla 3.5: Cálculo de la 2-epistasia de las cadenas de  $\Omega_4$  para una función camello.

$i$	$a$	$\delta_i(a)$	$E_i(a)$
0	0	15	7.5
	1	0	-7.5
1	0	15	7.5
	1	0	-7.5
2	0	15	7.5
	1	0	-7.5
3	0	15	7.5
	1	0	-7.5

$i, k$	$a, b$	$\delta_{i,k}(a, b)$	$E_{i,k}(a, b)$
1, 0	0, 0	30	22.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	0	-7.5
2, 0	0, 0	30	22.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	0	-7.5
3, 0	0, 0	30	22.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	0	-7.5
2, 1	0, 0	30	22.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	0	-7.5
3, 1	0, 0	30	22.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	0	-7.5
3, 2	0, 0	30	22.5
	0, 1	0	-7.5
	1, 0	0	-7.5
	1, 1	0	-7.5

Tabla 3.6: Cálculo de los valores de exceso de alelos y pares de alelos de la función  $\delta$ .

Finalmente, la 2-epistasis de  $\delta$  es 187.35. Este valor se obtiene fácilmente de la tabla 3.7.

$s$	$\delta(s)$	$EG(s)$	$\tilde{\delta}(s)$	$\tilde{\varepsilon}(s)$
0000	120	30	37.5	82.5
0001	0	75	82.5	-82.5
0010	0	75	82.5	-82.5
0011	0	-15	-7.5	7.5
0100	0	75	82.5	-82.5
0101	0	-15	-7.5	7.5
0110	0	-15	-7.5	7.5
0111	0	-15	-7.5	7.5
1000	0	75	82.5	-82.5
1001	0	-15	-7.5	7.5
1010	0	-15	-7.5	7.5
1011	0	-15	-7.5	7.5
1100	0	-15	-7.5	7.5
1101	0	-15	-7.5	7.5
1110	0	-15	-7.5	7.5
1111	0	15	22.5	-22.5

Tabla 3.7: Cálculo de la 2-epistasis de las cadenas de  $\Omega_4$  para la función  $\delta$ .

### 3.1.2 Representación matricial

Al igual que en el capítulo 2, y por comodidad en los cálculos posteriores, de ahora en adelante se considerará la matriz  $\mathbf{A}_\ell = 2^\ell \mathbf{B}_\ell \in \mathcal{M}_{2^\ell}(\mathbb{Z})$ . Para valores pequeños de  $\ell$ , las matrices correspondientes son:

$$\begin{aligned} \mathbf{A}_0 &= (1) \quad (= \mathbf{G}_0) \\ \mathbf{A}_1 &= \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \quad (= \mathbf{G}_1) \\ \mathbf{A}_2 &= \begin{pmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix} \quad (= 4\mathbf{I}_2) \\ \mathbf{A}_3 &= \begin{pmatrix} 7 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 7 & -1 & 1 & -1 & 1 & 1 & -1 \\ 1 & -1 & 7 & 1 & -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & 7 & 1 & -1 & -1 & 1 \\ 1 & -1 & -1 & 1 & 7 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 & 1 & 7 & -1 & 1 \\ -1 & 1 & 1 & -1 & 1 & -1 & 7 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & 7 \end{pmatrix}. \end{aligned}$$

De la definición de distancia Hamming, directamente se deduce que  $d_{st}^{\ell+1} = d_{st}^\ell$  para  $0 \leq s, t < 2^\ell$  o  $2^\ell \leq s, t < 2^{\ell+1}$  y  $d_{st}^{\ell+1} = d_{st}^\ell + 1$  para  $s \in \{0, \dots, 2^\ell - 1\}$ ,  $t \in \{2^\ell, \dots, 2^{\ell+1} - 1\}$ .

En el primer caso, i.e., si  $d_{st}^{\ell+1} = d_{st}^\ell$ , tenemos que el elemento  $a_{st}^{\ell+1}$  situado en la posición  $(s, t)$  de la matriz  $\mathbf{A}_{\ell+1}$  está dado por:

$$\begin{aligned} a_{st}^{\ell+1} &= 4 \binom{\ell+1 - d_{st}^{\ell+1}}{2} - 2(\ell-1) \binom{\ell+1 - d_{st}^{\ell+1}}{1} + \binom{\ell}{2} \\ &= 4 \binom{\ell - d_{st}^\ell}{2} + 4(\ell - d_{st}^\ell) - 2(\ell-1)(\ell - d_{st}^\ell) - 2(\ell-1) + \binom{\ell-1}{2} + (\ell-1) \\ &= 4 \binom{\ell - d_{st}^\ell}{2} - 2(\ell-2) (\ell - d_{st}^\ell) + \binom{\ell-1}{2} + [\ell+1 - 2d_{st}^\ell] \end{aligned}$$

$$= a_{st}^\ell + g_{st}^\ell,$$

donde, con nuestra notación,  $g_{st}^\ell$  denota una componente genérica de la matriz  $\mathbf{G}_\ell$  del capítulo anterior.

Por otra parte, cuando  $s \in \{0, \dots, 2^\ell - 1\}$  y  $t \in \{2^\ell, \dots, 2^{\ell+1} - 1\}$  se tiene  $d_{st}^{\ell+1} = d_{st}^\ell + 1$ , por lo que:

$$\begin{aligned} a_{st}^{\ell+1} &= 4 \binom{\ell+1-d_{st}^{\ell+1}}{2} - 2(\ell-1) \binom{\ell+1-d_{st}^{\ell+1}}{1} + \binom{\ell}{2} \\ &= 4 \binom{\ell-d_{st}^\ell}{2} - 2(\ell-2) \binom{\ell-d_{st}^\ell}{1} - 2 \binom{\ell-d_{st}^\ell}{1} + \binom{\ell-1}{2} + (\ell-1) \\ &= 4 \binom{\ell-d_{st}^\ell}{2} - 2(\ell-2) \binom{\ell-d_{st}^\ell}{1} + \binom{\ell-1}{2} - (\ell+1-2d_{st}^\ell) \\ &= a_{st}^\ell - g_{st}^\ell. \end{aligned}$$

Entonces, un proceso de inducción conduce a:

**Proposición 3.1.3.** *Para todo  $\ell \geq 0$  se tiene:*

$$\mathbf{A}_{\ell+1} = \begin{pmatrix} \mathbf{A}_\ell + \mathbf{G}_\ell & \mathbf{A}_\ell - \mathbf{G}_\ell \\ \mathbf{A}_\ell - \mathbf{G}_\ell & \mathbf{A}_\ell + \mathbf{G}_\ell \end{pmatrix}, \quad (3.1)$$

siendo  $\mathbf{A}_0 = (1)$ .

Como consecuencia de los resultados previos se obtienen las siguientes propiedades de la matriz  $\mathbf{A}_\ell$ :

**Lema 3.1.4.** *Para cualquier entero positivo  $\ell$ , la suma de todos los elementos de  $\mathbf{A}_\ell$  está dada por:*

$$\sum_{i,j=0}^{2^\ell-1} a_{ij}^\ell = 2^{2^\ell}.$$

*Demostración.* Para  $\mathbf{A}_0 = (1)$  la afirmación es evidente. Supongámosla cierta hasta  $\ell - 1$  y usemos la fórmula de recursión para  $\mathbf{A}_\ell$ . Entonces,

$$\sum_{i,j=0}^{2^\ell-1} a_{ij}^\ell = 4 \sum_{i,j=0}^{2^{(\ell-1)}-1} a_{ij}^{\ell-1} = 4 \cdot 2^{2^{(\ell-1)}} = 2^{2^\ell},$$

como afirmábamos. □

**Proposición 3.1.5.** Para cualquier  $\ell \geq 0$ ,  $\text{Rg } \mathbf{A}_\ell = \binom{\ell}{2} + \ell + 1$ .

*Demostración.* Procedamos por inducción en  $\ell$ . El resultado es obviamente cierto para  $\ell = 0$ . Supongamos ahora que se cumple para  $0, \dots, \ell - 1$  y demostrémoslo para  $\ell$ . En primer lugar, nótese que, por la fórmula de recursión dada en la proposición 3.1.3, la expresión de  $\mathbf{A}_\ell$  puede transformarse en:

$$\begin{pmatrix} \mathbf{A}_{\ell-1} & \mathbf{O}_{\ell-1} \\ \mathbf{O}_{\ell-1} & \mathbf{G}_{\ell-1} \end{pmatrix},$$

mediante operaciones elementales de filas y columnas, donde  $\mathbf{O}_\ell$  denota la matriz nula de orden  $2^\ell$ . Si aplicamos ahora la hipótesis de inducción, tenemos:

$$\begin{aligned} \text{Rg } \mathbf{A}_\ell &= \text{Rg } \mathbf{A}_{\ell-1} + \text{Rg } \mathbf{G}_{\ell-1} \\ &= \binom{\ell-1}{2} + (\ell-1) + 1 + (\ell-1) + 1 \\ &= \binom{\ell-1}{2} + \binom{\ell-1}{1} + \ell + 1 \\ &= \binom{\ell}{2} + \ell + 1. \end{aligned}$$

□

**Proposición 3.1.6.** Para  $\ell \geq 0$ ,  $(\mathbf{A}_\ell)^2 = 2^\ell \mathbf{A}_\ell$ .

*Demostración.* Nuevamente actuamos mediante inducción en  $\ell$ . La afirmación obviamente es cierta para  $\ell = 0$ . Por otra parte, para  $\ell > 0$  las fórmulas de recursión para  $\mathbf{A}_\ell$  y  $\mathbf{G}_\ell$ , junto con la hipótesis de inducción, proporcionan directamente el resultado:

$$\begin{aligned} (\mathbf{A}_\ell)^2 &= 2 \begin{pmatrix} (\mathbf{A}_{\ell-1})^2 + (\mathbf{G}_{\ell-1})^2 & (\mathbf{A}_{\ell-1})^2 - (\mathbf{G}_{\ell-1})^2 \\ (\mathbf{A}_{\ell-1})^2 - (\mathbf{G}_{\ell-1})^2 & (\mathbf{A}_{\ell-1})^2 + (\mathbf{G}_{\ell-1})^2 \end{pmatrix} \\ &= 2 \cdot 2^{\ell-1} \begin{pmatrix} \mathbf{A}_{\ell-1} + \mathbf{G}_{\ell-1} & \mathbf{A}_{\ell-1} - \mathbf{G}_{\ell-1} \\ \mathbf{A}_{\ell-1} - \mathbf{G}_{\ell-1} & \mathbf{A}_{\ell-1} + \mathbf{G}_{\ell-1} \end{pmatrix} \\ &= 2^\ell \mathbf{A}_\ell. \end{aligned}$$

□

Así, los autovalores de  $\mathbf{A}_\ell$  son 0 y  $2^\ell$ . De hecho, si  $v \in \mathbb{R}^{2^\ell}$  es un autovector con autovalor  $\alpha$ , de  $\mathbf{A}_\ell v = \alpha v$ , se deduce que:

$$2^\ell \alpha v = 2^\ell \mathbf{A}_\ell v = (\mathbf{A}_\ell)^2 v = \alpha^2 v$$

y, por tanto,  $\alpha = 0$  o  $\alpha = 2^\ell$ , como se afirmó.

Como consecuencia directa de esto, tenemos una interesante y útil propiedad de  $\mathbf{B}_\ell$ :

**Corolario 3.1.7.** *Para cualquier entero positivo  $\ell$  la matriz  $\mathbf{B}_\ell$  es idempotente y sus autovalores son 0 y 1.*

Denotemos por  $V^0$  y  $V^1$  (o  $V_\ell^0$  resp.  $V_\ell^1$  si existiese ambigüedad) los espacios de autovectores en  $\mathbb{R}^{2^\ell}$  correspondientes a los autovalores 0 y 1, respectivamente, de  $\mathbf{B}_\ell$  (o, equivalentemente, a 0 resp.  $2^\ell$ , autovalores de  $\mathbf{A}_\ell$ ), entonces  $\mathbb{R}^{2^\ell} = V^0 \oplus V^1$  y como  $V^0 = \ker \mathbf{B}_\ell$  y  $V^1 = \text{Im } \mathbf{B}_\ell$ , se tiene:

$$\dim V^0 = 2^\ell - \binom{\ell}{2} - \ell - 1$$

y

$$\dim V^1 = \binom{\ell}{2} + \ell + 1.$$

Con el objeto de obtener explícitamente una base de  $V^1$  consideramos la matriz diagonal  $\tilde{\mathbf{D}}_\ell \in \mathcal{M}_{2^\ell}(\mathbb{Z})$ , cuyos únicos elementos no nulos  $\tilde{d}_{ss}$  toman el valor 1 y están situados en las filas  $s = 0$ ,  $s = 2^i$  y  $s = 2^i + 2^j$ , para  $0 \leq i < j < \ell$ . Por tanto,

$$\tilde{\mathbf{D}}_0 = (1), \tilde{\mathbf{D}}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tilde{\mathbf{D}}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

y

$$\tilde{\mathbf{D}}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

por ejemplo.

Un argumento directo de inducción demuestra que:

**Proposición 3.1.8.** *Para cualquier  $\ell \geq 1$ ,*

$$\tilde{\mathbf{D}}_\ell = \begin{pmatrix} \tilde{\mathbf{D}}_{\ell-1} & \mathbf{O}_{\ell-1} \\ \mathbf{O}_{\ell-1} & \mathbf{D}_{\ell-1} \end{pmatrix},$$

donde  $\mathbf{D}_{\ell-1}$  denota la matriz diagonal introducida en el capítulo anterior.

Ahora nosotros tenemos:

**Proposición 3.1.9.** *Para cualquier  $\ell \geq 0$ , se verifica:*

$$\mathbf{W}_\ell \mathbf{B}_\ell \mathbf{W}_\ell = \tilde{\mathbf{D}}_\ell, \quad (3.2)$$

donde  $\mathbf{W}_\ell$  es la matriz de Walsh que satisface la propiedad de recursividad (2.6).

*Demostración.* Usemos un argumento de inducción en  $\ell$ . Para  $\ell = 0$  la afirmación es evidente. Se supone cierto el resultado hasta  $\ell$  y lo demostraremos para  $\ell + 1$ .

En primer lugar, nótese que:

$$\mathbf{W}_{\ell+1} = 2^{-1/2} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} = 2^{-1/2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \mathbf{W}_\ell,$$

donde  $\otimes$  denota el producto de Kronecker de matrices (ver sección B.1 del apéndice). Por otra parte, la matriz  $\mathbf{A}_{\ell+1}$  se puede escribir:

$$\begin{aligned} \mathbf{A}_{\ell+1} &= \begin{pmatrix} \mathbf{A}_\ell + \mathbf{G}_\ell & \mathbf{A}_\ell - \mathbf{G}_\ell \\ \mathbf{A}_\ell - \mathbf{G}_\ell & \mathbf{A}_\ell + \mathbf{G}_\ell \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_\ell & \mathbf{A}_\ell \\ \mathbf{A}_\ell & \mathbf{A}_\ell \end{pmatrix} + \begin{pmatrix} \mathbf{G}_\ell & -\mathbf{G}_\ell \\ -\mathbf{G}_\ell & \mathbf{G}_\ell \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{A}_\ell + \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{G}_\ell. \end{aligned}$$

Entonces, claramente,

$$\begin{aligned} \mathbf{W}_{\ell+1} \mathbf{A}_{\ell+1} &= 2^{-1/2} \left[ \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \mathbf{W}_\ell \right] \left[ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{A}_\ell \right] \\ &\quad + 2^{-1/2} \left[ \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \mathbf{W}_\ell \right] \left[ \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \otimes \mathbf{G}_\ell \right] \\ &= 2^{-1/2} \left[ \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} \otimes (\mathbf{W}_\ell \mathbf{A}_\ell) + \begin{pmatrix} 0 & 0 \\ 2 & -2 \end{pmatrix} \otimes (\mathbf{W}_\ell \mathbf{G}_\ell) \right] \end{aligned}$$

y, por tanto,

$$\mathbf{W}_{\ell+1} \mathbf{A}_{\ell+1} \mathbf{W}_{\ell+1} = 2^{-1} \left[ \begin{pmatrix} 2 & 2 \\ 0 & 0 \end{pmatrix} \otimes (\mathbf{W}_\ell \mathbf{A}_\ell) \right] \left[ \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \mathbf{W}_\ell \right]$$

$$\begin{aligned}
& +2^{-1} \left[ \begin{pmatrix} 0 & 0 \\ 2 & -2 \end{pmatrix} \otimes (\mathbf{W}_\ell \mathbf{G}_\ell) \right] \left[ \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \otimes \mathbf{W}_\ell \right] \\
& = 2 \left[ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \otimes (\mathbf{W}_\ell \mathbf{A}_\ell \mathbf{W}_\ell) + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \otimes (\mathbf{W}_\ell \mathbf{G}_\ell \mathbf{W}_\ell) \right].
\end{aligned}$$

Ahora, teniendo en cuenta la hipótesis de inducción, que  $\mathbf{B}_{\ell+1} = 2^{-(\ell+1)} \mathbf{A}_{\ell+1}$ ,  $\mathbf{G}_\ell = 2^\ell \mathbf{E}_\ell$  y el lema 2.3.1 de la página 47, obtenemos finalmente:

$$\begin{aligned}
\mathbf{W}_{\ell+1} \mathbf{B}_{\ell+1} \mathbf{W}_{\ell+1} & = \frac{1}{2^{\ell+1}} \mathbf{W}_{\ell+1} \mathbf{A}_{\ell+1} \mathbf{W}_{\ell+1} \\
& = \begin{pmatrix} \tilde{\mathbf{D}}_\ell & \mathbf{O}_\ell \\ \mathbf{O}_\ell & \mathbf{O}_\ell \end{pmatrix} + \begin{pmatrix} \mathbf{O}_\ell & \mathbf{O}_\ell \\ \mathbf{O}_\ell & \mathbf{D}_\ell \end{pmatrix} \\
& = \tilde{\mathbf{D}}_{\ell+1}.
\end{aligned}$$

□

Una base para  $V^1$  se proporciona en el siguiente resultado:

**Proposición 3.1.10.** *Para cualquier entero positivo  $\ell$ , el conjunto formado por los vectores columna de la matriz de Walsh  $\mathbf{W}_\ell$ , situados en las posiciones  $i = 0, 2^j, 2^j + 2^k$ , con  $0 \leq j < k < \ell$ , es una base para  $V^1$ .*

*Demostración.* Dada la base canónica de  $\mathbb{R}^{2^\ell}$ ,  $\{\mathbf{e}_i; 0 \leq i < 2^\ell\}$ , consideramos el conjunto formado por los vectores  $\mathbf{w}_m = \mathbf{W}_\ell \mathbf{e}_m$ , con  $m = 0, 2^j, 2^j + 2^k$  y  $0 \leq j < k < \ell$ . Éste es claramente un conjunto de vectores linealmente independientes y como su cardinal es  $\binom{\ell}{2} + \ell + 1$  sólo falta demostrar que todos estos vectores pertenecen a  $V^1$ .

El hecho de que un vector  $\mathbf{v} \in \mathbb{R}^{2^\ell}$  pertenezca al espacio vectorial  $V^1$  si, y sólo si,  $\mathbf{B}_\ell \mathbf{v} = \mathbf{v}$ , junto con la idempotencia de la matriz de Walsh, la proposición 3.1.9 y el hecho de que, por la definición de  $\tilde{\mathbf{D}}_\ell$ , para todo  $m \in \{0, 2^j, 2^j + 2^k; 0 \leq j < k < \ell\}$ ,

$$\mathbf{e}_m = \tilde{\mathbf{D}}_\ell \mathbf{e}_m,$$

nos permite escribir:

$$\begin{aligned}
\mathbf{w}_m & = \mathbf{W}_\ell \mathbf{e}_m \\
& = \mathbf{W}_\ell \tilde{\mathbf{D}}_\ell \mathbf{e}_m = \mathbf{W}_\ell \tilde{\mathbf{D}}_\ell \mathbf{W}_\ell \mathbf{W}_\ell \mathbf{e}_m \\
& = \mathbf{W}_\ell \tilde{\mathbf{D}}_\ell \mathbf{W}_\ell \mathbf{w}_m = \mathbf{B}_\ell \mathbf{w}_m,
\end{aligned}$$

lo que concluye la demostración. □

Como ya se ha señalado en la introducción, la epistasia (y la 2-epistasia también) es una medida de interacciones entre genes. Por supuesto, estas interacciones no deben cambiar si se multiplica una función de ajuste por una constante. Puesto que  $\tilde{\varepsilon}(\lambda f) = \lambda \tilde{\varepsilon}(f)$ , es necesario remediar esto normalizando el concepto como sigue:

**Definición 3.1.11.** La 2-epistasia normalizada  $\tilde{\varepsilon}_\ell^*(f)$  de una función  $f : \Omega_\ell \rightarrow \mathbb{R}$  se define como:

$$\tilde{\varepsilon}_\ell^*(f) = \tilde{\varepsilon}_\ell^2 \left( \frac{f}{\|\mathbf{f}\|} \right).$$

Por ser la matriz  $\mathbf{B}_\ell$  idempotente y simétrica (es decir, una proyección ortogonal), también lo es  $\mathbf{I}_\ell - \mathbf{B}_\ell$ . De esta forma:

$$\begin{aligned} \tilde{\varepsilon}_\ell^*(f) &= \tilde{\varepsilon}_\ell^2 \left( \frac{f}{\|\mathbf{f}\|} \right) \\ &= \left\| (\mathbf{I}_\ell - \mathbf{B}_\ell) \left( \frac{\mathbf{f}}{\|\mathbf{f}\|} \right) \right\|^2 = \frac{1}{\|\mathbf{f}\|^2} \|(\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f}\|^2 \\ &= \frac{1}{\|\mathbf{f}\|^2} {}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{B}_\ell) {}^t (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f} = \frac{1}{\|\mathbf{f}\|^2} {}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f} \quad (3.3) \\ &= \frac{{}^t \mathbf{f} \mathbf{f} - {}^t \mathbf{f} \mathbf{B}_\ell \mathbf{f}}{\|\mathbf{f}\|^2} = 1 - \frac{{}^t \mathbf{f} \mathbf{B}_\ell \mathbf{f}}{\|\mathbf{f}\|^2}. \end{aligned}$$

**Nota:** En lugar de utilizar la norma del vector  $\mathbf{b}_\ell = \mathbf{f} - \mathbf{B}_\ell \mathbf{f}$ , para definir la epistasia de la función  $f$ , se podría usar  $\mathcal{R} = \|\cos \theta\|$ , donde  $\theta$  es el ángulo formado por  $\mathbf{f}$  y  $\mathbf{b}_\ell$ . De hecho,

$$\begin{aligned} \mathcal{R}^2 &= \cos^2(\mathbf{f}, (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f}) \\ &= \frac{\langle \mathbf{f}, (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f} \rangle^2}{\|\mathbf{f}\|^2 \|(\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f}\|^2} \\ &= \frac{{}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f} \, {}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f}}{{}^t \mathbf{f} \mathbf{f} \, {}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f}} \\ &= \frac{{}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{B}_\ell) \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}} = 1 - \frac{{}^t \mathbf{f} \mathbf{B}_\ell \mathbf{f}}{\|\mathbf{f}\|^2}. \end{aligned}$$

En Estadística,  $\mathcal{R}^2$  se conoce como la razón de correlación (o coeficiente de determinación) entre  $\mathbf{f}$  y  $\mathbf{b}_\ell$  y en los modelos lineales es el cuadrado del coeficiente de correlación. Así

$$\mathcal{R}^2 = \frac{\left( \sum_{i=0}^{2^\ell-1} f_i(\mathbf{b}_\ell)_i \right)^2}{\sum_{i=0}^{2^\ell-1} f_i^2 \sum_{i=0}^{2^\ell-1} (\mathbf{b}_\ell)_i^2} = \tilde{\varepsilon}_\ell^*(f).$$

Nótese que  $0 \leq \tilde{\varepsilon}_\ell^*(f) \leq 1$ . Claramente  $\tilde{\varepsilon}_\ell^*(f) = 0$ , (respectivamente  $\tilde{\varepsilon}_\ell^*(f) = 1$ ) si y sólo si,  $\mathbf{f} \in V_\ell^1$  (respectivamente  $\mathbf{f} \in V_\ell^0$ ).

Como la proyección ortogonal sobre subespacios es única, la 2-epistasia normalizada de una función,  $\tilde{\varepsilon}_\ell^*(f)$ , proporciona una medida de la distancia de  $f$  a la clase de funciones que pertenecen al espacio  $V_\ell^1$ . En la sección 3.4 demostraremos que estas funciones son las funciones de orden dos.

Ahora, de la proposición 3.1.9 se deduce el siguiente resultado:

**Proposición 3.1.12.** *Si  $w_i$ ,  $0 \leq i < 2^\ell$ , son los coeficientes de Walsh de una función  $f : \Omega_\ell \rightarrow \mathbb{R}$  entonces, la 2-epistasia normalizada está dada por:*

$$\tilde{\varepsilon}_\ell^*(f) = 1 - \frac{w_0^2 + \sum_{0 \leq i < \ell} w_{2^i}^2 + \sum_{0 \leq i < j < \ell} w_{2^i+2^j}^2}{\sum_{0 \leq i < 2^\ell} w_i^2}. \quad (3.4)$$

*Demostración.* Omitiremos el subíndice  $\ell$  en esta demostración. Al ser  $\mathbf{f} = \mathbf{W}\mathbf{w}$  y  $\mathbf{W}$  matriz simétrica e idempotente se tiene que:

$${}^t\mathbf{f}\mathbf{f} = {}^t(\mathbf{W}\mathbf{w})(\mathbf{W}\mathbf{w}) = {}^t\mathbf{w}{}^t\mathbf{W}\mathbf{W}\mathbf{w} = {}^t\mathbf{w}\mathbf{w}.$$

Por otra parte,

$${}^t\mathbf{f}\mathbf{B}\mathbf{f} = {}^t\mathbf{w}{}^t\mathbf{W}\mathbf{B}\mathbf{W}\mathbf{w} = {}^t\mathbf{w}\tilde{\mathbf{D}}\mathbf{w},$$

por lo que,

$$\tilde{\varepsilon}^*(f) = 1 - \frac{{}^t\mathbf{f}\mathbf{B}\mathbf{f}}{{}^t\mathbf{f}\mathbf{f}} = 1 - \frac{{}^t\mathbf{w}\tilde{\mathbf{D}}\mathbf{w}}{{}^t\mathbf{w}\mathbf{w}},$$

como queríamos. □

### 3.2 2-epistasis normalizada de algunas funciones interesantes

Presentamos en este apartado el cálculo de la 2-epistasis normalizada de algunas funciones que consideramos interesantes en la investigación sobre el comportamiento de los Algoritmos Genéticos, en la optimización funcional. Las funciones consideradas aquí responden a la pregunta formulada en [9] en donde se especula sobre la posibilidad de que sea la alta no linealidad presente en algunas representaciones binarias, la causante de las dificultades que los AG encuentran para optimizar estas representaciones.

#### 3.2.1 La función de Dirac

La función de Dirac se define como aquella en la que sólo uno de los  $2^\ell$  individuos tiene asignado un valor de idoneidad no nulo. Sin pérdida de generalidad, nosotros consideramos la función de Dirac  $f = \delta_0$ , i.e.,  $f(t) = \delta_{t0}$  cuya representación vectorial es:

$$\delta_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Cuando un AG busca el máximo de esta función, se enfrenta a un problema esencialmente aleatorio, por lo que ésta resulta ser una función “difícil” para el algoritmo.

Al ser la función de Walsh  $\psi_0$  tal que  $\psi_0(t) = 1$ , para todo  $t \in \Omega_\ell$ , obviamente se sigue que:

$$\mathbf{w} = \mathbf{W}_\ell \delta_0 = 2^{-\ell/2} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

y por tanto:

$$\begin{aligned} \tilde{\varepsilon}_\ell^*(\delta_0) &= 1 - \frac{w_0^2 + \sum_{0 \leq i < \ell} w_{2^i}^2 + \sum_{0 \leq i < j < \ell} w_{2^i+2^j}^2}{\|\mathbf{w}\|^2} \\ &= 1 - \frac{\left\{ 1 + \binom{\ell}{1} + \binom{\ell}{2} \right\} (2^{-\ell/2})^2}{2^\ell (2^{-\ell/2})^2} \\ &= 1 - \frac{1 + \ell + \binom{\ell}{2}}{2^\ell}, \end{aligned}$$

valor que tiende a uno a medida que  $\ell$  crece. Esto se corresponde con el hecho de que si la localización del máximo es desconocida, es tan ineficiente como un algoritmo de búsqueda exhaustiva. De hecho, los resultados experimentales han mostrado que el tiempo esperado para alcanzar la solución crece exponencialmente con la longitud de las cadenas. El problema que encierra la optimización de esta función –conocido en la literatura como *onemax problem*– así como la optimización de las funciones que consideraremos a continuación, las funciones camello (*twomax* o *twin peaks problem*) han sido extensamente estudiados (ver [7]).

### 3.2.2 Funciones Camello

Una función  $c : \Omega_\ell \rightarrow \mathbb{R}$  se dice de tipo camello (*Camel function*), si existe una única cadena  $s \in \Omega_\ell$  tal que  $c(s) = c(\hat{s}) = 1$ , con  $d(s, \hat{s}) = \ell$ , y  $c(t) = 0$  en otro caso. Estas funciones son las que alcanzan el máximo valor de la epistasis normalizada citada en el capítulo 3 (ver [39]). Son funciones “difíciles” para un Algoritmo Genético pues el operador cruce rompe, con probabilidad alta, las cadenas más idóneas formando descendientes de peor calidad. Se podría decir que el extremo de la función de Dirac es más estable que los dos extremos (con igual valor de idoneidad) de la función camello y esto le convierte en más difícil de optimizar que el problema de un sólo pico.

Consideremos, por ejemplo, la función  $c : \Omega_\ell \rightarrow \mathbb{R}$  definida por  $c(0\dots 0) = c(1\dots 1) = 1$  y  $c(t) = 0$  para las restantes cadenas  $t \in \Omega_\ell$ . Claramente,  $c = \delta_0 + \delta_{2^\ell - 1}$  y, por tanto, su representación vectorial es:

$$\mathbf{c} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Entonces, para cualquier cadena  $s \in \Omega_\ell$ :

$$\begin{aligned} w_s &= (\mathbf{w})_s = (\mathbf{W}\mathbf{c})_s \\ &= 2^{-\ell/2} (\psi_{0\dots 0}(s) + \psi_{1\dots 1}(s)) \\ &= 2^{-\ell/2} \left( (-1)^{0 \cdot s} + (-1)^{(2^\ell - 1) \cdot s} \right) \\ &= 2^{-\ell/2} \left( 1 + (-1)^{u(s)} \right), \end{aligned} \tag{3.5}$$

siendo  $u(s)$  el número de unos de la cadena  $s$ .

En particular,  $w_0 = 2^{-\ell/2} (1 + (-1)^0) = 2 \cdot 2^{-\ell/2}$  y, para cualquier par  $0 \leq i < j < \ell$ ,  $w_{2^i} = 2^{-\ell/2} (1 + (-1)^1) = 0$  y  $w_{2^i+2^j} = 2^{-\ell/2} (1 + (-1)^2) = 2 \cdot 2^{-\ell/2} = w_0$ .

Por tanto,

$$\begin{aligned} \tilde{\varepsilon}_\ell^*(c) &= 1 - \frac{w_0^2 + \sum_{0 \leq i < \ell} w_{2^i}^2 + \sum_{0 \leq i < j < \ell} w_{2^i+2^j}^2}{\|\mathbf{w}\|^2} \\ &= 1 - \frac{\left(1 + \binom{\ell}{2}\right) w_0^2}{1^2 + 1^2} \\ &= 1 - \frac{\left(1 + \binom{\ell}{2}\right) (2 \cdot 2^{-\ell/2})^2}{2} \\ &= 1 - \frac{1 + \binom{\ell}{2}}{2^{\ell-1}}, \end{aligned}$$

cuyo valor tiende a 1 cuando  $\ell \rightarrow \infty$ .

Nótese que, para un  $\ell$  fijo,  $\tilde{\varepsilon}_\ell^*(c) > \tilde{\varepsilon}_\ell^*(\delta_0)$ , lo que se corresponde con el comportamiento de los AG al tratar estas dos clases de problemas.

### 3.2.3 Funciones de Ponderación

Las llamadas funciones de ponderación (*unitation functions*) son funciones  $f$  definidas en términos del número de unos de una cadena. Formalmente, tienen la propiedad de que existe una función  $h : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $f(s) = h(u(s))$ , para todo  $s \in \Omega_\ell$  (ver [18]). Este tipo de funciones puede tomar, como máximo,  $\ell + 1$  valores diferentes, en contraste con los  $2^\ell$  valores que pueden llegar a tomar las funciones arbitrarias. En este caso, la idoneidad de una cadena depende del número de unos, sin importar el orden de colocación de los mismos. Por ejemplo,  $f(0011) = f(0101) = f(0110) = f(1001) = f(1010) = f(1100) = h(2)$ .

Como para cada  $0 \leq u \leq \ell$  hay  $\binom{\ell}{u}$  cadenas en  $\Omega_\ell$  con valor de ajuste  $h(u)$ , es fácil ver ([33]) que

$$w_0 = 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \binom{\ell}{u} h(u)$$

y

$$w_{2^i} = 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \left[ \binom{\ell-1}{u} - \binom{\ell-1}{u-1} \right] h(u). \quad (3.6)$$

Nótese que  $w_{2^i}$  no depende de  $i$ .

Con el objeto de obtener el valor de la 2-epistasia normalizada, es necesario calcular los coeficientes de Walsh de segundo orden:  $w_{2^{i+2j}}$  con  $0 \leq i < j < \ell$ . De (2.11) se deduce directamente que, para las funciones de ponderación,

$$w_{2^{i+2j}} = 2^{\frac{\ell}{2}} f(\mathcal{H}_{ij}) - (w_0 + 2 w_{2^i}),$$

donde la idoneidad media del esquema  $\mathcal{H}_{ij} = \# \dots \overset{j}{\downarrow} \# 0 \# \dots \overset{i}{\downarrow} \# 0 \# \dots \#$  está dada por:

$$\begin{aligned} f(\mathcal{H}_{ij}) &= \frac{1}{2^{\ell-2}} \sum_{t \in \mathcal{H}_{ij}} f(t) = \frac{1}{2^{\ell-2}} \sum_{t \in \mathcal{H}_{ij}} h(u(t)) \\ &= \frac{1}{2^{\ell-2}} \sum_{u=0}^{\ell-2} \binom{\ell-2}{u} h(u) = 2^{-(\ell-2)} \sum_{u=0}^{\ell} \binom{\ell-2}{u} h(u). \end{aligned}$$

Entonces, claramente,

$$\begin{aligned} w_{2^{i+2j}} &= 2^{\frac{\ell}{2}} f(\mathcal{H}_{ij}) - (w_0 + 2 w_{2^i}) \\ &= 2^{\frac{\ell}{2}} 2^{-(\ell-2)} \sum_{u=0}^{\ell} \binom{\ell-2}{u} h(u) \\ &\quad - 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \binom{\ell}{u} h(u) - 2 \cdot 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \left[ \binom{\ell-1}{u} - \binom{\ell-1}{u-1} \right] h(u) \\ &= 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \left[ 4 \binom{\ell-2}{u} - 2 \binom{\ell-1}{u} + 2 \binom{\ell-1}{u-1} - \binom{\ell}{u} \right] h(u) \\ &= 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \left[ \binom{\ell-2}{u} - 2 \binom{\ell-2}{u-1} + \binom{\ell-2}{u-2} \right] h(u). \end{aligned}$$

De la expresión anterior se obtiene que, como en el caso de los coeficientes de Walsh de primer orden, los de segundo orden tampoco dependen de  $i$  ni de  $j$ .

Finalmente, teniendo en cuenta que, para cualquier  $0 \leq i < j < \ell$ ,  $w_{2^i} = w_{2^0}$  y  $w_{2^{i+2j}} = w_{2^0+2^1}$ , el valor de la 2-epistasia normalizada de las funciones de ponderación es:

$$\tilde{\varepsilon}_{\ell}^*(f) = 1 - \frac{w_0^2 + \binom{\ell}{1} w_1^2 + \binom{\ell}{2} w_3^2}{\sum_{u=0}^{\ell} \binom{\ell}{u} h(u)^2}.$$

Nótese que la función de Dirac  $\delta_0$  es un caso particular de una función de ponderación. En este caso,  $w_0 = w_1 = w_3 = 2^{-\ell/2}$  y la función  $h : \{0, \dots, \ell\} \rightarrow \mathbb{R}$  asociada a  $f$  (i.e.,  $f(s) = h(u(s))$ ) es

$$h(u) = \begin{cases} 1 & \text{si } u = 0 \\ 0 & \text{en otro caso.} \end{cases}$$

### 3.2.4 Funciones Royal Road generalizadas

Las funciones Royal Road generalizadas  $\mathfrak{R}_m^n$  de tipo I (definidas en [31]) se han introducido en la sección 2.4 a través de los esquemas

$$\sigma_p^{n,m} = \#^{(2^m p)} 1^{(2^m)} \#^{(2^n - 2^m(p+1))},$$

donde  $0 \leq p < 2^{n-m}$  ( $p \in \mathbb{Z}$ ). El valor de  $\mathfrak{R}_m^n$  sobre una cadena  $s \in \Omega_{2^n} = \{0, 1\}^{2^n}$  es:

$$\mathfrak{R}_m^n(s) = \sum_{s \in \sigma_p^{n,m}} 2^m.$$

Si se denota por  $|\sigma_p^{n,m}|$  el cardinal de  $\sigma_p^{n,m}$  (el número de cadenas “en” o “satisfaciendo”  $\sigma_p^{n,m}$ ) entonces  $|\sigma_p^{n,m}| = 2^{2^n - 2^m}$ , para cualquier  $0 \leq p < 2^{n-m}$ .

Para calcular la 2-epistasis normalizada de las funciones  $\mathfrak{R}_m^n$  se necesita en primer lugar obtener sus coeficientes de Walsh. Obsérvese que el valor medio de  $\Omega_{2^n} = \# \dots \#$  está dado por:

$$\begin{aligned} \mathfrak{R}_m^n(\Omega) &= \frac{1}{2^{2^n}} \sum_{s \in \Omega} \mathfrak{R}_m^n(s) = \frac{1}{2^{2^n}} \sum_{s \in \Omega} \left( \sum_{p; s \in \sigma_p^{n,m}} 2^m \right) \\ &= \frac{1}{2^{2^n}} \sum_{p=0}^{2^{n-m}-1} \left( \sum_{s \in \sigma_p^{n,m}} 2^m \right) = \frac{1}{2^{2^n}} \sum_{p=0}^{2^{n-m}-1} 2^m |\sigma_p^{n,m}| \\ &= \frac{1}{2^{2^n}} 2^{n-m} 2^m 2^{2^n - 2^m} = 2^{n-2^m}. \end{aligned}$$

Así, teniendo en cuenta que, por (2.9),  $w_0 = 2^{\frac{2^n}{2}} \mathfrak{R}_m^n(\Omega_{2^n})$ , entonces tenemos

$$w_0 = 2^{\frac{2^n}{2}} 2^{n-2^m} = 2^{n+2^{n-1}-2^m} = 2^{n-m} \omega, \quad (3.7)$$

siendo  $\omega = 2^{m+2^{n-1}-2^m}$ .

En lo que sigue denominamos “bloque”,  $\mathbf{B}_j$ , de longitud  $2^m$  a la estructura compuesta por los  $2^m$  loci consecutivos situados a partir de la posición  $(j - 1) \cdot 2^m$  para algún  $j$  con  $1 \leq j \leq 2^{n-m}$ . Es decir,

$$\mathbf{B}_j = \square_{j \cdot 2^m - 1} \dots \square_{(j-1) \cdot 2^m + 1} \square_{(j-1) \cdot 2^m}.$$

Así, dada una cadena  $s \in \Omega_{2^n}$ , diremos que la subcadena  $s_{i+2^m-1} \dots s_{i+1} s_i$  es un representante del bloque  $\mathbf{B}_j$  (o que lo determina) si  $i = (j - 1) \cdot 2^m$ , para algún  $j$  ( $1 \leq j \leq 2^{n-m}$ ). Por ejemplo, si  $n = 4$  y  $m = 2$ , existen cuatro bloques determinados, respectivamente, por las subcadenas genéricas

$$\mathbf{B}_1 = s_3 s_2 s_1 s_0, \mathbf{B}_2 = s_7 s_6 s_5 s_4, \mathbf{B}_3 = s_{11} s_{10} s_9 s_8 \text{ y } \mathbf{B}_4 = s_{15} s_{14} s_{13} s_{12}.$$

Nótese que un bloque  $\mathbf{B}_j$  no es más que el esquema de orden 0 en  $\Omega_{2^m}$  considerado como subestructura en  $\Omega_{2^n}$  colocada a partir de ciertas posiciones.

Se puede demostrar:

**Proposición 3.2.1.** *Para cualesquiera  $i_1, \dots, i_k \in \{0, \dots, 2^n - 1\}$ , se verifica:*

$$w_{2^{i_1} + \dots + 2^{i_k}} = \begin{cases} (-1)^k \omega & \text{si } (j - 1) \cdot 2^m \leq i_1, \dots, i_k < j \cdot 2^m, \text{ para algún } j \in \{1, \dots, 2^{n-m}\} \\ 0 & \text{en otro caso.} \end{cases}$$

*Demostración.* Procedamos por inducción en  $k$ . Para  $k = 1$  consideremos el esquema  $\mathcal{H}_i = \# \dots \# \overset{i}{0} \# \dots \#$  y sea  $j$  el entero tal que  $(j - 1) \cdot 2^m \leq i < j \cdot 2^m$ . Entonces obviamente,  $\sigma_{2^{n-m-j}}^{n,m} \cap \mathcal{H}_i = \emptyset$  y  $|\sigma_p^{n,m} \cap \mathcal{H}_i| = 2^{2^n - 2^m - 1}$ , para cualquier  $0 \leq p < 2^{n-m}$  con  $p \neq 2^{n-m} - j$ . Por tanto, para todo  $0 \leq i < 2^n$ , el valor medio del esquema  $\mathcal{H}_i$  es:

$$\begin{aligned} \mathfrak{R}_m^n(\mathcal{H}_i) &= \frac{1}{|\mathcal{H}_i|} \sum_{s \in \mathcal{H}_i} \mathfrak{R}_m^n(s) = \frac{2}{2^{2^n}} \sum_{s \in \mathcal{H}_i} \left( \sum_{p; s \in \sigma_p^{n,m}} 2^m \right) \\ &= \frac{2}{2^{2^n}} \sum_{p; \sigma_p^{n,m} \cap \mathcal{H}_i \neq \emptyset} \left( \sum_{s \in \mathcal{H}_i \cap \sigma_p^{n,m}} 2^m \right) \\ &= \frac{2}{2^{2^n}} \sum_{p; p \neq 2^{n-m-j}} 2^m |\sigma_p^{n,m} \cap \mathcal{H}_i| \\ &= \frac{2}{2^{2^n}} 2^m 2^{2^n - 2^m - 1} (2^{n-m} - 1) \\ &= 2^{-2^m} (2^n - 2^m) \end{aligned}$$

y por (2.10),

$$\begin{aligned}
w_{2^i} &= 2^{\frac{2^n}{2}} \mathfrak{R}_m^n(\mathcal{H}_i) - w_0 \\
&= 2^{\frac{2^n}{2}} 2^{-2^m} (2^n - 2^m) - 2^{n+2^{n-1}-2^m} \\
&= -2^{m+2^{n-1}-2^m} \\
&= -\omega.
\end{aligned}$$

Ahora, supongamos que la afirmación es cierta hasta  $k-1$  y demostrémosla para  $k$ .

Para calcular  $w_{2^{i_1+\dots+2^{i_k}}}$  consideremos primero el caso en que  $k \leq 2^m$  y también consideremos el esquema:

$$\mathcal{H}_{i_1\dots i_k} = \# \dots \# \overset{i_k}{\downarrow} 0 \# \dots \# \overset{i_1}{\downarrow} 0 \# \dots \#$$

y supongamos que existe  $j$  tal que  $(j-1) \cdot 2^m \leq i_1, \dots, i_k < j \cdot 2^m$ . En este caso  $\mathcal{H}_{i_1\dots i_k} \cap \sigma_{2^{n-m-j}}^{n,m} = \emptyset$  y  $|\sigma_p^{n,m} \cap \mathcal{H}_{i_1\dots i_k}| = 2^{2^n-2^m-k}$ , para  $p \in \{0, \dots, 2^{n-m} - 1\}$ ,  $p \neq 2^{n-m} - j$ . Entonces:

$$\begin{aligned}
\mathfrak{R}_m^n(\mathcal{H}_{i_1\dots i_k}) &= \frac{1}{|\mathcal{H}_{i_1\dots i_k}|} \sum_{s \in \mathcal{H}_{i_1\dots i_k}} \mathfrak{R}_m^n(s) \\
&= \frac{2^k}{2^{2^n}} 2^m 2^{2^n-2^m-k} (2^{n-m} - 1) \\
&= 2^{n-2^m} - 2^{m-2^m} \\
&= 2^{-2^m} (2^n - 2^m)
\end{aligned} \tag{3.8}$$

y por (2.12) y la hipótesis de inducción:

$$\begin{aligned}
w_{2^{i_1+\dots+2^{i_k}}} &= 2^{2^{n-1}} \mathfrak{R}_m^n(\mathcal{H}_{i_1\dots i_k}) - w_0 \\
&\quad - \left\{ \sum_{p=1}^k w_{2^p} + \sum_{1 \leq p_1 < p_2 \leq k} w_{2^{p_1+2^{p_2}}} + \dots + \sum_{1 \leq p_1 < \dots < p_{k-1} \leq k} w_{2^{p_1+\dots+2^{p_{k-1}}}} \right\} \\
&= 2^{2^{n-1}} \mathfrak{R}_m^n(\mathcal{H}_{i_1\dots i_k}) - w_0 \\
&\quad - \left\{ \sum_{p=1}^k (-1)\omega + \sum_{1 \leq p_1 < p_2 \leq k} (-1)^2\omega + \dots + \sum_{1 \leq p_1 < \dots < p_{k-1} \leq k} (-1)^{k-1}\omega \right\} \\
&= 2^{2^{n-1}} 2^{-2^m} (2^n - 2^m) - w_0 - \omega \sum_{p=1}^{k-1} \binom{k}{p} (-1)^p
\end{aligned}$$

$$\begin{aligned}
 &= w_0 - \omega - w_0 - \omega \sum_{p=1}^{k-1} \binom{k}{p} (-1)^p \\
 &= -\omega \sum_{p=0}^{k-1} \binom{k}{p} (-1)^p \\
 &= (-1)^k \omega,
 \end{aligned}$$

como se quería.

Supongamos ahora que no existe  $j$  tal que  $(j-1) \cdot 2^m \leq i_1, \dots, i_k < j \cdot 2^m$ . Esto significa que no existe un bloque que contenga a todas las posiciones fijas del esquema  $\mathcal{H}_{i_1 \dots i_k}$ . Es decir, esas posiciones se encuentran distribuidas en  $b$  bloques, con  $b > 1$ . Sea  $\eta_r$  el número de posiciones con alelos nulos que se encuentran en el bloque  $\mathbf{B}_r$  ( $r = 1, \dots, b$ ). Entonces, por (2.12) y la hipótesis de inducción,

$$\begin{aligned}
 w_{2^{i_1} + \dots + 2^{i_k}} &= 2^{2^n/2} \mathfrak{R}_m^n(\mathcal{H}_{i_1 \dots i_k}) - w_0 - \sum_{p=1}^{k-1} \sum_{1 \leq \lambda_1 < \dots < \lambda_p \leq k} w_{2^{i_{\lambda_1}} + \dots + 2^{i_{\lambda_p}}} \\
 &= 2^{2^n-1} \mathfrak{R}_m^n(\mathcal{H}_{i_1 \dots i_k}) - w_0 - \Lambda,
 \end{aligned} \tag{3.9}$$

con

$$\begin{aligned}
 \Lambda &= \sum_{\substack{1 \leq \lambda_1 \leq k \\ \exists j / \lambda_1 \in \mathbf{B}_j}} w_{2^{i_{\lambda_1}}} + \sum_{\substack{1 \leq \lambda_1 < \lambda_2 \leq k \\ \exists j / \{\lambda_1, \lambda_2\} \subset \mathbf{B}_j}} w_{2^{i_{\lambda_1}} + 2^{i_{\lambda_2}}} + \dots + \sum_{\substack{1 \leq \lambda_1 < \dots < \lambda_p \leq k \\ \exists j / \{\lambda_1, \dots, \lambda_p\} \subset \mathbf{B}_j}} w_{2^{i_{\lambda_1}} + \dots + 2^{i_{\lambda_p}}} \\
 &= \sum_{j=1}^b \left\{ \sum_{\substack{1 \leq \lambda_1 \leq k \\ \lambda_1 \in \mathbf{B}_j}} w_{2^{i_{\lambda_1}}} + \sum_{\substack{1 \leq \lambda_1 < \lambda_2 \leq k \\ \{\lambda_1, \lambda_2\} \subset \mathbf{B}_j}} w_{2^{i_{\lambda_1}} + 2^{i_{\lambda_2}}} + \dots + \sum_{\substack{1 \leq \lambda_1 < \dots < \lambda_p \leq k \\ \{\lambda_1, \dots, \lambda_p\} \subset \mathbf{B}_j}} w_{2^{i_{\lambda_1}} + \dots + 2^{i_{\lambda_p}}} \right\} \\
 &= \sum_{j=1}^b \left\{ \binom{\eta_j}{1} (-1) \omega + \binom{\eta_j}{2} (-1)^2 \omega + \dots + \binom{\eta_j}{\eta_j} (-1)^{\eta_j} \omega \right\} \\
 &= \sum_{j=1}^b \sum_{\mu=1}^{\eta_j} \binom{\eta_j}{\mu} (-1)^\mu \omega = \sum_{j=1}^b -\omega = -b\omega.
 \end{aligned} \tag{3.10}$$

Por otra parte, un razonamiento similar al empleado en (3.8) conduce aquí a que

$$\mathfrak{R}_m^n(\mathcal{H}_{i_1 \dots i_k}) = \frac{2^k}{2^{2^n}} 2^m 2^{2^n - 2^m - k} (2^{n-m} - b) = 2^{-2^m} (2^n - b 2^m). \tag{3.11}$$

Sustituyendo (3.7), (3.10) y (3.11) en (3.9) se tiene directamente que:

$$w_{2^{i_1+\dots+2^{i_k}}} = 2^{2^n-1} 2^{-2^m} (2^n - b 2^m) - 2^{\frac{2^n}{2}} 2^{n-2^m} + b\omega = 0.$$

Para finalizar la demostración, si  $k > 2^m$ , no existe  $j$  tal que  $(j-1) \cdot 2^m \leq i_1 < \dots < i_k < j \cdot 2^m$ , por lo que la demostración se reduce al caso anterior y, en consecuencia,  $w_{2^{i_1+\dots+2^{i_k}}} = 0$ .  $\square$

Ahora, teniendo en cuenta que los únicos coeficientes de Walsh no nulos son aquellos que se construyen con índices correspondientes a un mismo bloque y que el número de bloques es  $2^{n-m}$ , se tiene:

**Corolario 3.2.2.** Para  $1 \leq k \leq 2^n$  se verifica:

$$\sum_{0 \leq i_1 < \dots < i_k < 2^n} w_{2^{i_1+\dots+2^{i_k}}}^2 = \binom{2^m}{k} 2^{n-m} \omega^2.$$

Ya estamos en condiciones de calcular la norma del vector  $\mathfrak{R}_m^n$ .

**Proposición 3.2.3.** Para cualquier  $m \leq n$  se tiene:

$$\|\mathfrak{R}_m^n\|^2 = 2^{n-m} (2^{n-m} + 2^{2^m} - 1) \omega^2.$$

*Demostración.* Puesto que  $\|\mathfrak{R}_m^n\| = \|\mathbf{w}\|$ , el corolario previo y (3.7) conducen a:

$$\begin{aligned} \|\mathfrak{R}_m^n\|^2 &= \sum_{i=0}^{2^{2^n}-1} w_i^2 \\ &= w_0^2 + \sum_{k=1}^{2^n} \left[ \sum_{0 \leq i_1 < \dots < i_k < 2^n} w_{2^{i_1+\dots+2^{i_k}}}^2 \right] \\ &= w_0^2 + \sum_{k=1}^{2^m} \left[ \sum_{0 \leq i_1 < \dots < i_k < 2^n} w_{2^{i_1+\dots+2^{i_k}}}^2 \right] \\ &= 2^{2n-2m} \omega^2 + \sum_{k=1}^{2^m} \binom{2^m}{k} 2^{n-m} \omega^2 \\ &= 2^{2n-2m} \omega^2 + 2^{n-m} \omega^2 (2^{2^m} - 1) \end{aligned}$$

$$= 2^{n-m} (2^{n-m} + 2^{2^m} - 1) \omega^2.$$

□

**Nota:** En [34] este valor se obtiene de una forma más técnica y por un camino extremadamente más complicado. Finalmente, obtenemos:

**Proposición 3.2.4.** *La 2-epistasis normalizada de las funciones Royal Road generalizadas de tipo I está dada por:*

$$\tilde{\varepsilon}_{2^n}^*(\mathfrak{R}_m^n) = \frac{2^{2^m} - 2^{2^m-1} - 2^{m-1} - 1}{2^{n-m} + 2^{2^m} - 1}.$$

*Demostración.* Es consecuencia inmediata de (3.7), las proposiciones 3.1.12, 3.2.1, 3.2.3 y el corolario 3.2.2. De hecho:

$$\begin{aligned} \tilde{\varepsilon}_{2^n}^*(\mathfrak{R}_m^n) &= 1 - \frac{w_0^2 + \sum_{0 \leq i < 2^n} w_{2^i}^2 + \sum_{0 \leq i < j < 2^n} w_{2^i+2^j}^2}{\sum_{0 \leq i < 2^{2^n}} w_i^2} \\ &= 1 - \frac{(2^{n-m}\omega)^2 + \sum_{0 \leq i < 2^n} (-\omega)^2 + \sum_{0 \leq i < j < 2^n} \omega^2}{\sum_{0 \leq i < 2^{2^n}} w_i^2} \\ &= 1 - \frac{(2^{n-m}\omega)^2 + 2^n \omega^2 + \binom{2^m}{2} 2^{n-m} \omega^2}{2^{n-m} (2^{n-m} + 2^{2^m} - 1) \omega^2} \\ &= 1 - \frac{2^{n-m} + 2^m + \binom{2^m}{2}}{2^{n-m} + 2^{2^m} - 1} \\ &= 1 - \frac{2^{n-m} + 2^{2^m-1} + 2^{m-1}}{2^{n-m} + 2^{2^m} - 1} \\ &= \frac{2^{2^m} - 2^{2^m-1} - 2^{m-1} - 1}{2^{n-m} + 2^{2^m} - 1}. \end{aligned}$$

□

### 3.2.5 Funciones Template

Nuestro último ejemplo lo constituyen las funciones Template introducidas en la subsección 2.2.3 y reconsideradas en la subsección 2.3.2 y en la sección 2.4. Para el cálculo de la 2-epistasis normalizada de este tipo de funciones se necesita conocer sus coeficientes de Walsh

de orden 0, 1 y 2. Aunque los coeficientes de orden 0 y 1 ya se han calculado en la proposición 2.3.4 y por lo tanto sólo restaría calcular los de orden 2, dado que éstos se van a obtener por un procedimiento distinto al usado para calcular  $w_0$  y  $w_{2^i}$  ( $i = 0, \dots, \ell - 1$ ), recalcularemos también los ya conocidos. La combinación del uso de las transformadas de Walsh, junto con el hecho de que la función Template  $T_\ell^n$  es de orden  $n$  (como se estableció en el lema 2.4.1 de la sección 2.4), pone de manifiesto la clara ventaja del uso de las transformadas de Walsh para obtener de una forma mucho más sencilla la 2-epistasia.

De acuerdo con el lema 2.4.1, el vector  $\mathbf{w} = \mathbf{W}_\ell \mathbf{T}_\ell^n$  se puede escribir como suma de los vectores de Walsh asociados a las funciones  $\tau_k$

$$\mathbf{w} = \mathbf{W}_\ell \mathbf{T}_\ell^n = \mathbf{W}_\ell \left( \sum_{k=0}^{\ell-n} \tau_k \right) = \sum_{k=0}^{\ell-n} \mathbf{W}_\ell \tau_k, \quad (3.12)$$

donde  $\tau_k$  denota el vector asociado a la función  $\tau_k$ , para todo  $k$ . Si denotamos por  $\mathbf{w}_{(\tau_k)} = \mathbf{W}_\ell \tau_k = (w_{(\tau_k)0}, w_{(\tau_k)1}, \dots, w_{(\tau_k)2^{\ell-1}})$  al vector de Walsh asociado a  $\tau_k$ , entonces, por (2.9),

$$w_{(\tau_k)0} = 2^{\frac{\ell}{2}} \tau_k(\Omega_\ell) = 2^{\frac{\ell}{2}} \frac{1}{2^\ell} \sum_{s \in \Omega_\ell} \tau_k(s) = 2^{-\frac{\ell}{2}} 2^{\ell-n} = 2^{\frac{\ell-2n}{2}}.$$

Así, el coeficiente de orden cero asociado a  $T_\ell^n$  es

$$w_0 = (\mathbf{W}_\ell \mathbf{T}_\ell^n)_0 = \sum_{k=0}^{\ell-n} (\mathbf{W}_\ell \tau_k)_0 = \sum_{k=0}^{\ell-n} w_{(\tau_k)0} = \sum_{k=0}^{\ell-n} 2^{\frac{\ell-2n}{2}} = (\ell - n + 1) 2^{\frac{\ell-2n}{2}}.$$

Para calcular los coeficientes de Walsh de primer orden asociados a  $\tau_k$  ( $0 \leq k \leq \ell - n$ ) denotemos por  $\Delta_k$  el conjunto de  $n$  índices consecutivos, comenzando en  $k$ , i.e.,  $\Delta_k = \{k, k + 1, \dots, k + n - 1\}$ .

Fijado un índice  $i$  ( $0 \leq i < \ell$ ), consideramos el esquema  $\mathcal{H}_i = \# \dots \overset{i}{\downarrow} \# 0 \# \dots \#$ . Si  $i \in \Delta_k$  (para algún  $k$ ) entonces, de la definición de  $\tau_k$  y de la expresión (2.10) de la página 50, se tiene:

$$\begin{aligned} \tau_k(\mathcal{H}_i) &= 0, \\ w_{(\tau_k)2^i} &= 2^{\frac{\ell}{2}} \tau_k(\mathcal{H}_i) - w_{(\tau_k)0} = 0 - 2^{\frac{\ell-2n}{2}} = -2^{\frac{\ell-2n}{2}}, \end{aligned} \quad (3.13)$$

mientras que si  $i \notin \Delta_k$ ,

$$\begin{aligned} \tau_k(\mathcal{H}_i) &= \frac{1}{2^{\ell-1}} \sum_{s \in \mathcal{H}_i} \tau_k(s) = \frac{1}{2^{\ell-1}} 2^{\ell-n-1} = 2^{-n}, \\ w_{(\tau_k)2^i} &= 2^{\frac{\ell}{2}} \tau_k(\mathcal{H}_i) - w_{(\tau_k)0} = 2^{\frac{\ell}{2}} 2^{-n} - 2^{\frac{\ell-2n}{2}} = 0. \end{aligned}$$

Entonces, de (3.12), el coeficiente de Walsh de orden uno  $w_{2^i}$  asociado a  $T_\ell^n$  es:

$$w_{2^i} = \sum_{k=0}^{\ell-n} w_{(\tau_k)2^i} = \sum_{\substack{k=0 \\ k; i \in \Delta_k}}^{\ell-n} -2^{\frac{\ell-2n}{2}}.$$

Debemos pues, determinar a cuántos conjuntos de índices  $\Delta_k$  pertenece  $i$ . Este número queda reflejado en las tablas 3.8 y 3.9 (Nótese que es necesario distinguir dos casos, dependiendo de la longitud  $n$  del patrón en la cadena  $s = s_{\ell-1} \dots s_0$ ).

Posición del índice	Número de conjuntos
$i \in \{0, \dots, \ell - n - 1\}$	$i + 1$
$i \in \{\ell - n, \dots, n - 1\}$	$\ell - n + 1$
$i \in \{n, \dots, \ell - 1\}$	$\ell - i$

Tabla 3.8: Conjuntos de índices  $\Delta_k$  a los que pertenece  $i$  en el caso  $\ell \leq 2n$ .

Posición del índice	Número de conjuntos
$i \in \{0, \dots, n - 1\}$	$i + 1$
$i \in \{n, \dots, \ell - n - 1\}$	$n$
$i \in \{\ell - n, \dots, \ell - 1\}$	$\ell - i$

Tabla 3.9: Conjuntos de índices  $\Delta_k$  a los que pertenece  $i$  en el caso  $\ell \geq 2n$ .

De esta forma, se puede asegurar que, para  $\ell \leq 2n$ :

$$w_{2^i} = \begin{cases} -2^{\frac{\ell-2n}{2}} (i + 1) & \text{si } 0 \leq i < \ell - n \\ -2^{\frac{\ell-2n}{2}} (\ell - n + 1) & \text{si } \ell - n \leq i < n \\ -2^{\frac{\ell-2n}{2}} (\ell - i) & \text{si } n \leq i < \ell \end{cases}$$

y, para  $\ell > 2n$ ,

$$w_{2^i} = \begin{cases} -2^{\frac{\ell-2n}{2}} (i + 1) & \text{si } 0 \leq i < n \\ -2^{\frac{\ell-2n}{2}} n & \text{si } n \leq i < \ell - n \\ -2^{\frac{\ell-2n}{2}} (\ell - i) & \text{si } \ell - n \leq i < \ell. \end{cases}$$

Como era de esperar, estos valores coinciden con los obtenidos en las últimas columnas de las tablas 2.5 y 2.6 de la proposición 2.3.4.

Los coeficientes de orden dos asociados a  $\tau_k$  se calcularán a partir de la fórmula (2.11) de la página 50:

$$w_{(\tau_k)2^i+2^j} = 2^{\frac{\ell}{2}} \tau_k(\mathcal{H}_{ij}) - w_{(\tau_k)2^i} - w_{(\tau_k)2^j} - w_{(\tau_k)0},$$

siendo  $\mathcal{H}_{ij} = \# \dots \overset{j}{\downarrow} \# 0 \# \dots \overset{i}{\downarrow} \# 0 \# \dots \#$ .

Si  $\{i, j\} \subset \Delta_k$  entonces  $\tau_k(\mathcal{H}_{ij}) = 0$  y, por (3.13),

$$w_{(\tau_k)2^i} = w_{(\tau_k)2^j} = -2^{\frac{\ell-2n}{2}}.$$

Así,

$$w_{(\tau_k)2^i+2^j} = 0 + 2^{\frac{\ell-2n}{2}} + 2^{\frac{\ell-2n}{2}} - 2^{\frac{\ell-2n}{2}} = 2^{\frac{\ell-2n}{2}}.$$

Si  $i \notin \Delta_k$ ,  $j \in \Delta_k$  entonces  $\tau_k(\mathcal{H}_{ij}) = 0$ ,  $w_{(\tau_k)2^i} = -2^{\frac{\ell-2n}{2}}$  y  $w_{(\tau_k)2^j} = 0$ . Así,

$$w_{(\tau_k)2^i+2^j} = 0 + 2^{\frac{\ell-2n}{2}} + 0 - 2^{\frac{\ell-2n}{2}} = 0.$$

Por último, si  $\{i, j\} \cap \Delta_k = \emptyset$  entonces  $\tau_k(\mathcal{H}_{ij}) = \frac{1}{2^{\ell-2}} 2^{\ell-n-2} = 2^{-n}$ ,

$$w_{(\tau_k)2^i} = w_{(\tau_k)2^j} = 0.$$

Por ello,

$$w_{(\tau_k)2^i+2^j} = 2^{\frac{\ell}{2}} 2^{-n} - 0 - 0 - 2^{\frac{\ell-2n}{2}} = 0.$$

Entonces,

$$w_{2^i+2^j} = \sum_{k=0}^{\ell-n} w_{(\tau_k)2^i+2^j} = \sum_{\substack{k=0 \\ k; \{i,j\} \subset \Delta_k}}^{\ell-n} 2^{\frac{\ell-2n}{2}}.$$

Finalmente, para calcular el coeficiente de Walsh de orden dos asociado a  $T_\ell^n$ ,  $w_{2^i+2^j}$  ( $i < j$ ), se debe determinar cuántos conjuntos  $\Delta_k$  contienen a  $\{i, j\}$ . Al igual que antes, este número queda reflejado en las tablas 3.10 y 3.11.

**Proposición 3.2.5.** *Dada la función Template  $T_\ell^n$ , el valor de  $\Gamma_2 = \sum_{0 \leq i < j < \ell} w_{2^i+2^j}^2$  viene dado por:*

$$\Gamma_2 = \begin{cases} 2^{\ell-2n} \left\{ \binom{\ell-n+1}{2} \left[ \binom{\ell-n}{2} + \frac{2}{3} (2n-\ell)(2(\ell-n)+1) \right] + \binom{2n-\ell}{2} (\ell-n+1)^2 \right\} & \text{si } \ell \leq 2n, \\ 2^{\ell-2n} \binom{n}{2} \left[ \binom{n+1}{2} + \frac{1}{3} (2n-1)(\ell-2n) \right] & \text{si } \ell \geq 2n. \end{cases}$$

Posición de los índices	Número de conjuntos
$i \in \{0, \dots, \ell - n - 1\}, j \in \{i + 1, \dots, \ell - n - 1\}$	$i + 1$
$i \in \{0, \dots, \ell - n - 1\}, j \in \{\ell - n, \dots, n - 1\}$	$i + 1$
$i \in \{0, \dots, \ell - n - 1\}, j \in \{n, \dots, i + n - 1\}$	$n - j + i$
$i \in \{\ell - n, \dots, n - 1\}, j \in \{i + 1, \dots, n - 1\}$	$\ell - n + 1$
$i \in \{\ell - n, \dots, n - 1\}, j \in \{n, \dots, \ell - 1\}$	$\ell - j$
$i \in \{n, \dots, \ell - 2\}, j \in \{i + 1, \dots, \ell - 1\}$	$\ell - j$

Tabla 3.10: Coeficientes de Walsh de orden 2 de  $T_\ell^n$  ( $\ell \leq 2n$ ).

Posición de los índices	Número de conjuntos
$i \in \{0, \dots, n - 1\}, j \in \{i + 1, \dots, n - 1\}$	$i + 1$
$i \in \{0, \dots, n - 1\}, j \in \{n, \dots, i + n - 1\}$	$n - j + i$
$i \in \{n, \dots, \ell - n - 1\}, j \in \{i + 1, \dots, i + n - 1\}$	$n - j + i$
$i \in \{\ell - n, \dots, \ell - 2\}, j \in \{i + 1, \dots, \ell - 1\}$	$\ell - j$

Tabla 3.11: Coeficientes de Walsh de orden 2 de  $T_\ell^n$  ( $\ell \geq 2n$ ).

*Demostración.* Para  $\ell \leq 2n$ ,

$$\begin{aligned}
 \Gamma_2 &= \sum_{0 \leq i < j < \ell} \sum_{\substack{k=0 \\ k; \{i,j\} \subset \Delta_k}}^{\ell-n} \left(2^{\frac{\ell-2n}{2}}\right)^2 \\
 &= 2^{\ell-2n} \left[ \sum_{i=0}^{\ell-n-1} \sum_{j=i+1}^{\ell-n-1} (i+1)^2 + \sum_{i=0}^{\ell-n-1} \sum_{j=\ell-n}^{n-1} (i+1)^2 + \sum_{i=0}^{\ell-n-1} \sum_{j=n}^{i+n-1} (n-j+i)^2 \right. \\
 &\quad \left. + \sum_{i=\ell-n}^{n-1} \sum_{j=i+1}^{n-1} (\ell-n+1)^2 + \sum_{i=\ell-n}^{n-1} \sum_{j=n}^{\ell-1} (\ell-j)^2 + \sum_{i=n}^{\ell-2} \sum_{j=i+1}^{\ell-1} (\ell-j)^2 \right].
 \end{aligned}$$

No resulta difícil comprobar que los sumandos primero, tercero y sexto de la expresión anterior coinciden, ya que:

$$\sum_{i=0}^{\ell-n-1} \sum_{j=i+1}^{\ell-n-1} (i+1)^2 = \sum_{i=0}^{\ell-n-1} (i+1)^2 (\ell-n-(i+1))$$

$$\begin{aligned}
&= (\ell - n) \sum_{i=0}^{\ell-n-1} (i+1)^2 - \sum_{i=0}^{\ell-n-1} (i+1)^3 \\
&= (\ell - n) \sum_{k=1}^{\ell-n} k^2 - \sum_{k=1}^{\ell-n} k^3,
\end{aligned}$$

$$\begin{aligned}
\sum_{i=0}^{\ell-n-1} \sum_{j=n}^{i+n-1} (n-j+i)^2 &= \sum_{i=0}^{\ell-n-1} \sum_{k=0}^{i-1} (i-k)^2 = \sum_{k=0}^{\ell-n-1} (\ell-n-k) k^2 \\
&= \sum_{k=1}^{\ell-n} (\ell-n-k) k^2 = (\ell-n) \sum_{k=1}^{\ell-n} k^2 - \sum_{k=1}^{\ell-n} k^3
\end{aligned}$$

y

$$\begin{aligned}
\sum_{i=n}^{\ell-2} \sum_{j=i+1}^{\ell-1} (\ell-j)^2 &= \sum_{i=n}^{\ell-2} \sum_{k=1}^{\ell-i-1} k^2 = \sum_{k=1}^{\ell-n-1} (\ell-n-k) k^2 \\
&= \sum_{k=1}^{\ell-n} (\ell-n-k) k^2 = (\ell-n) \sum_{k=1}^{\ell-n} k^2 - \sum_{k=1}^{\ell-n} k^3.
\end{aligned}$$

Ahora, como

$$\sum_{k=1}^{\ell-n} k^2 = \frac{2(\ell-n)+1}{3} \binom{\ell-n+1}{2}$$

y

$$\sum_{k=1}^{\ell-n} k^3 = (1+2+\dots+(\ell-n))^2 = \binom{\ell-n+1}{2}^2,$$

tenemos que el valor de cada una de esas tres sumas dobles es:

$$\begin{aligned}
(\ell-n) \sum_{k=1}^{\ell-n} k^2 - \sum_{k=1}^{\ell-n} k^3 &= \binom{\ell-n+1}{2} \left[ \frac{(\ell-n)(2(\ell-n)+1)}{3} - \binom{\ell-n+1}{2} \right] \\
&= \frac{1}{3} \binom{\ell-n+1}{2} \binom{\ell-n}{2}.
\end{aligned}$$

Por otra parte,

$$\sum_{i=0}^{\ell-n-1} \sum_{j=\ell-n}^{n-1} (i+1)^2 = (2n-\ell) \sum_{i=0}^{\ell-n-1} (i+1)^2 = (2n-\ell) \sum_{k=1}^{\ell-n} k^2,$$

$$\sum_{i=\ell-n}^{n-1} \sum_{j=n}^{\ell-1} (\ell-j)^2 = \sum_{i=\ell-n}^{n-1} \sum_{k=1}^{\ell-n} k^2 = (2n-\ell) \sum_{k=1}^{\ell-n} k^2$$

y

$$\begin{aligned}
\sum_{i=\ell-n}^{n-1} \sum_{j=i+1}^{n-1} (\ell-n+1)^2 &= (\ell-n+1)^2 \sum_{i=\ell-n}^{n-1} (n-(i+1)) \\
&= (\ell-n+1)^2 \left\{ (n-1)(2n-\ell) - \sum_{i=\ell-n}^{n-1} i \right\} \\
&= (\ell-n+1)^2 \left\{ (n-1)(2n-\ell) - \frac{\ell-1}{2}(2n-\ell) \right\} \\
&= (\ell-n+1)^2 \binom{2\ell-n}{2}.
\end{aligned}$$

Entonces,

$$\begin{aligned}
\Gamma_2 &= 2^{\ell-2n} \left\{ \binom{\ell-n+1}{2} \binom{\ell-n}{2} + \frac{2}{3} (2n-\ell) \binom{\ell-n+1}{2} (2(\ell-n)+1) \right. \\
&\quad \left. + \binom{2n-\ell}{2} (\ell-n+1)^2 \right\} \\
&= 2^{\ell-2n} \left\{ \binom{\ell-n+1}{2} \left[ \binom{\ell-n}{2} + \frac{2}{3} (2n-\ell)(2(\ell-n)+1) \right] \right. \\
&\quad \left. + \binom{2n-\ell}{2} (\ell-n+1)^2 \right\}.
\end{aligned}$$

Por otra parte, si  $\ell \geq 2n$ ,

$$\begin{aligned}
\Gamma_2 &= 2^{\ell-2n} \left[ \sum_{i=0}^{n-1} \sum_{j=i+1}^{n-1} (i+1)^2 + \sum_{i=0}^{n-1} \sum_{j=n}^{i+n-1} (n-j+i)^2 + \sum_{i=n}^{\ell-n-1} \sum_{j=i+1}^{i+n-1} (n-j+i)^2 \right. \\
&\quad \left. + \sum_{i=\ell-n}^{\ell-2} \sum_{j=i+1}^{\ell-1} (\ell-j)^2 \right]
\end{aligned}$$

Ahora, teniendo en cuenta que, por un razonamiento análogo al del caso  $\ell \leq 2n$ ,

$$\begin{aligned}
\sum_{i=0}^{n-1} \sum_{j=i+1}^{n-1} (i+1)^2 &= \sum_{i=0}^{n-1} \sum_{j=n}^{i+n-1} (n-j+i)^2 = n \sum_{k=1}^n k^2 - \sum_{k=1}^n k^3 \\
&= \frac{1}{3} \binom{n+1}{2} \binom{n}{2},
\end{aligned}$$

$$\sum_{i=n}^{\ell-n-1} \sum_{j=i+1}^{i+n-1} (n-j+i)^2 = \sum_{i=n}^{\ell-n-1} \sum_{k=1}^{n-1} k^2 = (\ell-2n) \sum_{k=1}^{n-1} k^2 = (\ell-2n) \frac{2n-1}{3} \binom{n}{2}$$

y

$$\begin{aligned} \sum_{i=\ell-n}^{\ell-2} \sum_{j=i+1}^{\ell-1} (\ell-j)^2 &= \sum_{i=\ell-n}^{\ell-2} \sum_{k=1}^{\ell-i-1} k^2 = \sum_{k=1}^{n-1} (n-k) k^2 \\ &= \sum_{k=1}^n (n-k) k^2 = n \sum_{k=1}^n k^2 - \sum_{k=1}^n k^3, \end{aligned}$$

un cálculo directo conduce a que

$$\begin{aligned} \Gamma_2 &= 2^{\ell-2n} \left[ \frac{1}{3} \binom{n+1}{2} \binom{n}{2} + \frac{1}{3} \binom{n+1}{2} \binom{n}{2} + \frac{1}{3} \binom{n}{2} (2n-1)(\ell-2n) \right. \\ &\quad \left. + \frac{1}{3} \binom{n+1}{2} \binom{n}{2} \right] \\ &= 2^{\ell-2n} \binom{n}{2} \left[ \binom{n+1}{2} + \frac{1}{3} (2n-1)(\ell-2n) \right]. \end{aligned}$$

□

Teniendo en cuenta todos los cálculos anteriores y el valor de la norma del vector  $\mathbf{T}_\ell^n$  se verifica:

**Proposición 3.2.6.** *El valor de la 2-epistasis normalizada de la función Template  $T_\ell^n$ ,  $\tilde{\varepsilon}_\ell^*(T_\ell^n)$ , viene dado por*

$$\begin{aligned} a) & 1 - 2^{\ell-2n} \frac{[(2n-\ell+1) + \binom{2n-\ell}{2}](\ell-n+1)^2 + \binom{\ell-n+1}{2} [\frac{2}{3}(2(\ell-n)+1) + \binom{\ell-n}{2} + \frac{2}{3}(2(\ell-n)+1)(2n-\ell)]}{2^{\ell-n}(3(\ell-n)-1)+2} \text{ si } \ell \leq 2n, \\ b) & 1 - 2^{\ell-2n} \frac{(\ell-n+1)^2 + n^2(\ell-2n) + \frac{2}{3}(2n+1)\binom{n+1}{2} + \binom{n}{2} [\binom{n+1}{2} + \frac{1}{3}(2n-1)(\ell-2n)]}{2^{\ell-n}(3(\ell-n)-1) + 2^{\ell-2n}(2 + (\ell-2n)(\ell-2n-1))} \text{ si } \ell \geq 2n. \end{aligned}$$

*Demostración.* El valor de la 2-epistasis normalizada de la función  $T_\ell^n$  se obtiene mediante la fórmula

$$\varepsilon_\ell^*(T_\ell^n) = 1 - \frac{\Gamma_0 + \Gamma_1 + \Gamma_2}{\|\mathbf{T}_\ell^n\|^2},$$

donde  $\Gamma_0 = w_0^2$ ,  $\Gamma_1 = \sum_{0 \leq i < \ell} w_{2^i}^2$  y  $\Gamma_2 = \sum_{0 \leq i < j < \ell} w_{2^i+2^j}^2$ .

Ahora, una sustitución directa del valor  $\Gamma_0 + \Gamma_1$  obtenido en la proposición 2.3.4 y  $\Gamma_2$  de la proposición anterior, junto con las relaciones (2.7) y (3.4) y la proposición 2.2.4, concluyen la demostración. □

### 3.3 Invarianza de la 2–epistasis normalizada

En el estudio de la 2–epistasis normalizada se observa que existen distintas funciones para las que el valor de este estimador coincide. Casos particulares son las funciones de segundo orden, para las que la 2–epistasis normalizada se anula. Ahora, nosotros expondremos un criterio que permitirá determinar en qué casos permanece invariante la 2–epistasis normalizada de una función.

En lo que sigue, denotamos por  $T$  una transformación ortogonal sobre funciones  $f : \Omega_\ell \rightarrow \mathbb{R}$ , y sea  $\mathcal{O}_n(\mathbb{R})$  el grupo de matrices ortogonales de orden  $n$ .

Se verifica:

**Proposición 3.3.1.** *El conjunto de transformaciones ortogonales que dejan invariante la 2–epistasis normalizada es isomorfo al grupo  $\mathcal{O}_{\binom{\ell}{2}+\ell+1}(\mathbb{R}) \times \mathcal{O}_{2^\ell - \binom{\ell}{2} - \ell - 1}(\mathbb{R})$ .*

*Demostración.* En tanto no haya confusión, se suprimirán los subíndices en la demostración.

Sea  $\mathbf{T} \in \mathcal{M}_{2^\ell}(\mathbb{R})$  la matriz asociada a una transformación ortogonal  $T$ . La caracterización de la 2–epistasis normalizada de una función  $f : \Omega_\ell \rightarrow \mathbb{R}$ , dada por (3.3) permite asegurar que

$$\tilde{\varepsilon}_\ell^*(T(f)) = \tilde{\varepsilon}_\ell^*(f) \Leftrightarrow {}^t\mathbf{T}\mathbf{B}\mathbf{T} = \mathbf{B} \Leftrightarrow \mathbf{B}\mathbf{T} = \mathbf{T}\mathbf{B}.$$

Consideremos la matriz diagonal  $\tilde{\mathbf{D}}$  definida en la subsección 3.1.2 (ver página 87), y sea  $\mathbf{P}$  ( $=\mathbf{P}_\ell$ ) la matriz  $2^\ell$ –dimensional que verifica

$$\mathbf{P}\tilde{\mathbf{D}}\mathbf{P} = \begin{pmatrix} \tilde{\mathbf{I}}_{\binom{\ell}{2}+\ell+1} & \tilde{\mathbf{O}} \\ \tilde{\mathbf{O}} & \hat{\mathbf{O}} \end{pmatrix},$$

donde aquí,  $\tilde{\mathbf{I}}_n$  denota la matriz identidad de orden  $n$ ,  $\tilde{\mathbf{O}}$  la matriz nula de orden  $\left(\binom{\ell}{2} + \ell + 1\right) \times \left(2^\ell - \binom{\ell}{2} - \ell - 1\right)$  y  $\hat{\mathbf{O}}$  la matriz cuadrada nula de orden  $2^\ell - \binom{\ell}{2} - \ell - 1$ .

La idempotencia de las matrices  $\mathbf{W}$  y  $\mathbf{P}$  permite afirmar que la relación  $\mathbf{B}\mathbf{T} = \mathbf{T}\mathbf{B}$  es equivalente a

$$\mathbf{P}(\mathbf{W}\mathbf{B}\mathbf{W})\mathbf{P}\mathbf{P}\mathbf{W}\mathbf{T}\mathbf{W}\mathbf{P} = \mathbf{P}\mathbf{W}\mathbf{T}\mathbf{W}\mathbf{P}\mathbf{P}(\mathbf{W}\mathbf{B}\mathbf{W})\mathbf{P},$$

es decir, por la relación (3.2),

$$\mathbf{P}\tilde{\mathbf{D}}\mathbf{P}\mathbf{P}\mathbf{W}\mathbf{T}\mathbf{W}\mathbf{P} = \mathbf{P}\mathbf{W}\mathbf{T}\mathbf{W}\mathbf{P}\mathbf{P}\tilde{\mathbf{D}}\mathbf{P}. \tag{3.14}$$

Sean

$$\mathbf{J} = \mathbf{P}\tilde{\mathbf{D}}\mathbf{P}$$

y

$$\mathbf{S} = \mathbf{P}\mathbf{W}\mathbf{T}\mathbf{W}\mathbf{P}.$$

Esta matriz  $\mathbf{S}$  se puede escribir por bloques de la siguiente forma:

$$\mathbf{S} = \begin{pmatrix} \mathbf{S}_{00} & \mathbf{S}_{01} \\ \mathbf{S}_{10} & \mathbf{S}_{11} \end{pmatrix}$$

con  $\mathbf{S}_{00}$  matriz cuadrada de orden  $\binom{\ell}{2} + \ell + 1$ ,  $\mathbf{S}_{01}$  de orden  $\left(\binom{\ell}{2} + \ell + 1\right) \times \left(2^\ell - \binom{\ell}{2} - \ell - 1\right)$ ,  $\mathbf{S}_{10}$  de orden  $\left(2^\ell - \binom{\ell}{2} - \ell - 1\right) \times \left(\binom{\ell}{2} + \ell + 1\right)$  y  $\mathbf{S}_{11}$  matriz cuadrada de orden  $2^\ell - \binom{\ell}{2} - \ell - 1$ .

Entonces, la relación (3.14) es equivalente a  $\mathbf{J}\mathbf{S} = \mathbf{S}\mathbf{J}$ , es decir

$$\begin{pmatrix} \tilde{\mathbf{I}}_{\binom{\ell}{2} + \ell + 1} & \tilde{\mathbf{O}} \\ {}^t\tilde{\mathbf{O}} & \hat{\mathbf{O}} \end{pmatrix} \begin{pmatrix} \mathbf{S}_{00} & \mathbf{S}_{01} \\ \mathbf{S}_{10} & \mathbf{S}_{11} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{00} & \mathbf{S}_{01} \\ \mathbf{S}_{10} & \mathbf{S}_{11} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{I}}_{\binom{\ell}{2} + \ell + 1} & \tilde{\mathbf{O}} \\ {}^t\tilde{\mathbf{O}} & \hat{\mathbf{O}} \end{pmatrix}$$

i.e.,

$$\begin{pmatrix} \mathbf{S}_{00} & \mathbf{S}_{01} \\ {}^t\tilde{\mathbf{O}} & \hat{\mathbf{O}} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_{00} & \tilde{\mathbf{O}} \\ \mathbf{S}_{10} & \hat{\mathbf{O}} \end{pmatrix}.$$

Así, la transformación ortogonal  $T$  deja invariante la 2-epistasis normalizada si y sólo si su matriz asociada  $\mathbf{T}$  verifica la relación:

$$\mathbf{T} = \mathbf{W}\mathbf{P}\mathbf{S}\mathbf{P}\mathbf{W},$$

con

$$\mathbf{S} = \begin{pmatrix} \mathbf{Q} & \tilde{\mathbf{O}} \\ {}^t\tilde{\mathbf{O}} & \bar{\mathbf{Q}} \end{pmatrix},$$

donde  $\mathbf{Q} \in \mathcal{O}_{\binom{\ell}{2} + \ell + 1}(\mathbb{R})$  y  $\bar{\mathbf{Q}} \in \mathcal{O}_{2^\ell - \binom{\ell}{2} - \ell - 1}(\mathbb{R})$ .  $\square$

### Ejemplo 3.3.2.

Si  $\ell = 3$  entonces  $\binom{\ell}{2} + \ell + 1 = 7$ , por lo que las matrices  $\mathbf{S}$  obtenidas en el resultado anterior son de la forma

$$\mathbf{S} = \begin{pmatrix} \mathbf{Q} & \tilde{\mathbf{O}} \\ {}^t\tilde{\mathbf{O}} & \epsilon \end{pmatrix}$$

con  $\mathbf{Q} \in \mathcal{O}_7(\mathbb{R})$  y  $\epsilon \in \mathcal{O}_1(\mathbb{R})$ , es decir,  $\epsilon = \pm 1$ .

Si consideramos, por ejemplo, la matriz

$$\mathbf{Q} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \in \mathcal{O}_7(\mathbb{R}),$$

Entonces, teniendo en cuenta que en este caso  $\mathbf{P} = \tilde{\mathbf{I}}_4$ , las transformaciones ortogonales  $T$  cuyas matrices asociadas, obtenidas a partir de  $\mathbf{S} = \begin{pmatrix} \mathbf{Q} & \tilde{\mathbf{O}} \\ {}^t\tilde{\mathbf{O}} & \epsilon \end{pmatrix}$ , conmutan con  $\mathbf{B}$  son dos y tienen asociadas las matrices

$$\frac{1}{2} \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & -1 & -1 \\ 0 & 1 & 1 & 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 0 & 0 & 1 & -1 \\ 0 & -1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & -1 & 0 & 1 \end{pmatrix} \text{ si } \epsilon = 1$$

y

$$\frac{1}{4} \begin{pmatrix} 3 & 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & -3 \\ 1 & 1 & 1 & 1 & -1 & -1 & 3 & -1 \\ -1 & -1 & 1 & 1 & 1 & -3 & -1 & -1 \\ 1 & -1 & -1 & 1 & 3 & 1 & 1 & -1 \\ -1 & 1 & -1 & -3 & 1 & -1 & 1 & -1 \\ -1 & -1 & 3 & -1 & 1 & 1 & 1 & 1 \\ 1 & -3 & -1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix} \text{ si } \epsilon = -1.$$

### 3.4 Valores extremos

En la sección 3.1 ya se ha demostrado que la 2-epistasis normalizada  $\tilde{\varepsilon}_\ell^*$  toma valores entre 0 y 1. Ahora calcularemos sus valores extremos. En primer lugar, observemos que los valores mínimo y máximo de  $\tilde{\varepsilon}_\ell^*(f)$  se corresponden, respectivamente, con el máximo y mínimo de  $\eta_\ell(f) = {}^t\mathbf{f} \mathbf{A}_\ell \mathbf{f}$ , con  $\mathbf{A}_\ell = 2^\ell \mathbf{B}_\ell$  y donde, por supuesto,  $0 \leq \eta_\ell(f) \leq 2^\ell \|\mathbf{f}\|^2$ . En particular, para funciones con vector asociado unitario, se tiene  $0 \leq \eta_\ell(f) \leq 2^\ell$ .

Para cualquier función  $f : \Omega_\ell \rightarrow \mathbb{R}$  con vector  ${}^t\mathbf{f} = (f_0, \dots, f_{2^\ell-1}) \in \mathbb{R}^{2^\ell}$ , denotamos por  $f^0, f^1 : \Omega_{\ell-1} \rightarrow \mathbb{R}$  las funciones cuyos respectivos vectores en  $\mathbb{R}^{2^{\ell-1}}$  son:

$$\mathbf{f}^0 = \begin{pmatrix} f_0 \\ \vdots \\ f_{2^{\ell-1}-1} \end{pmatrix}, \quad \mathbf{f}^1 = \begin{pmatrix} f_{2^{\ell-1}} \\ \vdots \\ f_{2^\ell-1} \end{pmatrix}. \quad (3.15)$$

Asimismo, consideremos las funciones:

$$g^+ = f^0 + f^1, \quad g^- = f^0 - f^1. \quad (3.16)$$

Si además de la forma  $\eta_\ell$  consideramos

$$\gamma_\ell(f) = {}^t\mathbf{f} \mathbf{G}_\ell \mathbf{f}$$

y se tiene en cuenta que las matrices  $\mathbf{G}_\ell$  y  $\mathbf{A}_\ell$  verifican las fórmulas de recurrencia (2.3) y (3.1), respectivamente, entonces:

$$\begin{aligned} \eta_\ell(f) &= {}^t\mathbf{f} \mathbf{A}_\ell \mathbf{f} \\ &= \begin{pmatrix} {}^t\mathbf{f}^0 & {}^t\mathbf{f}^1 \end{pmatrix} \begin{pmatrix} \mathbf{A}_{\ell-1} + \mathbf{G}_{\ell-1} & \mathbf{A}_{\ell-1} - \mathbf{G}_{\ell-1} \\ \mathbf{A}_{\ell-1} - \mathbf{G}_{\ell-1} & \mathbf{A}_{\ell-1} + \mathbf{G}_{\ell-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}^0 \\ \mathbf{f}^1 \end{pmatrix} \\ &= \begin{pmatrix} {}^t\mathbf{f}^0 & {}^t\mathbf{f}^1 \end{pmatrix} \begin{pmatrix} \mathbf{A}_{\ell-1} & \mathbf{A}_{\ell-1} \\ \mathbf{A}_{\ell-1} & \mathbf{A}_{\ell-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}^0 \\ \mathbf{f}^1 \end{pmatrix} \\ &\quad + \begin{pmatrix} {}^t\mathbf{f}^0 & {}^t\mathbf{f}^1 \end{pmatrix} \begin{pmatrix} \mathbf{G}_{\ell-1} & -\mathbf{G}_{\ell-1} \\ -\mathbf{G}_{\ell-1} & \mathbf{G}_{\ell-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}^0 \\ \mathbf{f}^1 \end{pmatrix} \\ &= {}^t(\mathbf{f}^0 + \mathbf{f}^1) \mathbf{A}_{\ell-1} (\mathbf{f}^0 + \mathbf{f}^1) + {}^t(\mathbf{f}^0 - \mathbf{f}^1) \mathbf{G}_{\ell-1} (\mathbf{f}^0 - \mathbf{f}^1) \\ &= \eta_{\ell-1}(g^+) + \gamma_{\ell-1}(g^-). \end{aligned} \quad (3.17)$$

Utilizando en (2.3) un razonamiento similar al empleado en (3.17) podemos afirmar que las formas  $\gamma_\ell$  y  $v_\ell$  verifican la relación:

$$\gamma_\ell(f) = \gamma_{\ell-1}(g^+) + v_{\ell-1}(g^-), \quad (3.18)$$

donde

$$v_\ell(f) = {}^t \mathbf{f} \mathbf{U}_\ell \mathbf{f} = \left( \sum_{k=0}^{2^\ell-1} f_k \right)^2.$$

Siguiendo la pauta establecida en (3.15), para cualquier función  $f : \Omega_\ell \rightarrow \mathbb{R}$  con vector  ${}^t \mathbf{f} = (f_0, \dots, f_{2^\ell-1}) \in \mathbb{R}^{2^\ell}$  denotamos por  $f^{00}, f^{01}, f^{10}, f^{11} : \Omega_{\ell-2} \rightarrow \mathbb{R}$  las funciones cuyos respectivos vectores en  $\mathbb{R}^{2^{\ell-2}}$  son:

$$\begin{aligned} \mathbf{f}^{00} &= \begin{pmatrix} f_0 \\ \vdots \\ f_{2^{\ell-2}-1} \end{pmatrix}, \quad \mathbf{f}^{01} = \begin{pmatrix} f_{2^{\ell-2}} \\ \vdots \\ f_{2^{\ell-1}-1} \end{pmatrix}, \\ \mathbf{f}^{10} &= \begin{pmatrix} f_{2^{\ell-1}} \\ \vdots \\ f_{2^{\ell-1}+2^{\ell-2}-1} \end{pmatrix}, \quad \mathbf{f}^{11} = \begin{pmatrix} f_{2^{\ell-1}+2^{\ell-2}} \\ \vdots \\ f_{2^\ell-1} \end{pmatrix} \end{aligned} \quad (3.19)$$

y, de forma similar a la establecida en (3.16), denotamos por  $g^{++}, g^{+-}, g^{-+}$  y  $g^{--}$  las funciones definidas en  $\Omega_{\ell-2}$  dadas por:

$$\begin{aligned} g^{++} &= (g^+)^0 + (g^+)^1 = f^{00} + f^{01} + f^{10} + f^{11}, \\ g^{+-} &= (g^+)^0 - (g^+)^1 = f^{00} - f^{01} + f^{10} - f^{11}, \\ g^{-+} &= (g^-)^0 + (g^-)^1 = f^{00} + f^{01} - f^{10} - f^{11}, \\ g^{--} &= (g^-)^0 - (g^-)^1 = f^{00} - f^{01} - f^{10} + f^{11}. \end{aligned}$$

### 3.4.1 Valor mínimo

Obsérvese en primer lugar, que el valor mínimo teórico  $\tilde{\varepsilon}_\ell^*(f) = 0$  –o, equivalentemente, el máximo valor de  $\eta_\ell(f)$ – se alcanza. De hecho, para  $\ell = 2$ ,  $\dim V_2^1 = 4$  y entonces  $V_2^1 = \mathbb{R}^4$ . Así, para todo  $\mathbf{f} \in \mathbb{R}^4$ , se tiene

$$\eta_2(f) = {}^t \mathbf{f} \mathbf{A}_2 \mathbf{f} = {}^t \mathbf{f} 4 \mathbf{I}_2 \mathbf{f} = 4 \|\mathbf{f}\|^2. \quad (3.20)$$

Por otra parte, para  $\ell > 2$ , consideremos el vector  $\mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^4$  y definamos

recursivamente los vectores  $\mathbf{v}_\ell = \begin{pmatrix} \mathbf{v}_{\ell-1} \\ \mathbf{v}_{\ell-1} \end{pmatrix}$ . Si  $f$  es la función de ajuste cuyo vector asociado es  $\mathbf{v}_\ell \in \mathbb{R}^{2^\ell}$  entonces –usando un argumento de inducción– es fácil demostrar que  $\eta_\ell(f) = 2^\ell \|\mathbf{v}_\ell\|^2$ . El caso  $\ell = 2$  se comprobó en (3.20). Supongamos que el resultado es cierto para  $\ell - 1$  y demostrémoslo para  $\ell$ .

$$\begin{aligned} \eta_\ell(f) &= {}^t \mathbf{v}_\ell \mathbf{A}_\ell \mathbf{v}_\ell \\ &= \begin{pmatrix} {}^t \mathbf{v}_{\ell-1} & {}^t \mathbf{v}_{\ell-1} \end{pmatrix} \begin{pmatrix} \mathbf{A}_{\ell-1} + \mathbf{G}_{\ell-1} & \mathbf{A}_{\ell-1} - \mathbf{G}_{\ell-1} \\ \mathbf{A}_{\ell-1} - \mathbf{G}_{\ell-1} & \mathbf{A}_{\ell-1} + \mathbf{G}_{\ell-1} \end{pmatrix} \begin{pmatrix} \mathbf{v}_{\ell-1} \\ \mathbf{v}_{\ell-1} \end{pmatrix} \\ &= 4 {}^t \mathbf{v}_{\ell-1} \mathbf{A}_{\ell-1} \mathbf{v}_{\ell-1} \\ &= 4 \cdot 2^{\ell-1} \|\mathbf{v}_{\ell-1}\|^2 \\ &= 2^\ell \|\mathbf{v}_\ell\|^2, \end{aligned}$$

para todo  $\ell > 2$ .

De las proposiciones 2.6.1 y 3.1.12 se sigue inmediatamente que:

**Teorema 3.4.1.** *Para cualquier función  $f : \Omega \rightarrow \mathbb{R}$ , las siguientes afirmaciones son equivalentes:*

1.  $f$  es una función de segundo orden,
2.  $\tilde{\varepsilon}_\ell^*(f) = 0$ .

### 3.4.2 Valor máximo

Nuestro próximo objetivo es encontrar el máximo valor de la 2-epistasia normalizada  $\tilde{\varepsilon}_\ell^*$ . Ya se ha mencionado que  $\tilde{\varepsilon}_\ell^* \leq 1$ . No obstante, si  $\ell > 2$  y nos restringimos a funciones no negativas con vector asociado unitario, entonces podemos asegurar que el valor máximo “teórico” no se alcanza. De hecho, el valor máximo de  $\tilde{\varepsilon}_\ell^*(f)$  es  $1 - 2^{2-\ell}$ , como se demuestra a continuación.

**Proposición 3.4.2.** Para cualquier  $\ell > 2$  y cualquier función de ajuste  $f$  que tome valores no negativos con  $\|\mathbf{f}\| = 1$ , se verifica:

$$\tilde{\varepsilon}_\ell^*(f) \leq 1 - \frac{1}{2^{\ell-2}}.$$

*Demostración.* Dado que la afirmación es equivalente a comprobar que  $\eta_\ell(f) \geq 4$  y, teniendo en cuenta que si  $\ell = 2$  y  $\|\mathbf{f}\| = 1$ , entonces  $\eta_2(f) = 4$ , bastará demostrar que  $\eta_\ell(f) \geq 4$  para  $\ell > 2$ . Para ello, observemos que, como  $f$  es no negativa, entonces  ${}^t\mathbf{f}^0 \cdot \mathbf{f}^1 \geq 0$ . Además,  $\|\mathbf{g}^+\| \geq 1$  ya que  $\|\mathbf{f}\| = 1$ .

Ahora supongamos que, para algún entero positivo  $\ell > 2$  y para alguna función  $f$  en las hipótesis de la proposición, se verifica  $\eta_\ell(f) < 4$ . Sea  $g = \frac{g^+}{\|\mathbf{g}^+\|}$  con vector asociado  $\mathbf{g}$  ( $\|\mathbf{g}\| = 1$ ). Entonces,

$$\begin{aligned} \eta_{\ell-1}(g) &= \eta_{\ell-1}\left(\frac{g^+}{\|\mathbf{g}^+\|}\right) = \frac{1}{\|\mathbf{g}^+\|^2} \eta_{\ell-1}(g^+) \leq \eta_{\ell-1}(g^+) \\ &\leq \eta_{\ell-1}(g^+) + \gamma_{\ell-1}(g^-) = \eta_\ell(f) < 4. \end{aligned}$$

(En la segunda inecuación se ha utilizado el hecho de que  $\mathbf{G}_\ell$  es una matriz semidefinida positiva y, en consecuencia,  $\gamma_{\ell-1}(g^-) \geq 0$ ).

Iterando este procedimiento se llegaría a que existen funciones de ajuste no negativas  $f$  definidas sobre  $\Omega = \{0, 1\}^2$  con  $\|\mathbf{f}\| = 1$  y  $\eta_2(f) < 4$ , lo cual es imposible, como ya sabemos.

Esta contradicción demuestra la afirmación. □

Por último, nos proponemos caracterizar a las funciones no negativas con vector asociado unitario y cuya 2-epistasis normalizada es máxima. Equivalentemente, aquellas funciones para las que  $\eta_\ell(f) = 4$  ( $\|\mathbf{f}\| = 1$ ).

Hemos demostrado ya en la proposición anterior que ese valor lo alcanza cualquier función de ajuste definida en  $\Omega_2 = \{0, 1\}^2$ . Sea ahora  $\ell > 2$  y definamos los vectores

$$\mathbf{w}^+ = \mathbf{W}_{\ell-1}\mathbf{g}^+ = {}^t(w_0^+, w_1^+, \dots, w_{2^{\ell-1}-1}^+)$$

y

$$\mathbf{w}^- = \mathbf{W}_{\ell-1}\mathbf{g}^- = {}^t(w_0^-, w_1^-, \dots, w_{2^{\ell-1}-1}^-).$$

Entonces, utilizando la idempotencia de la matriz  $\mathbf{W}_\ell$ , por la proposición 3.1.9 de la página 88 (respectivamente lema 2.3.1 de la página 47),  $\eta_{\ell-1}(g^+)$  (resp.  $\gamma_{\ell-1}(g^-)$ ) se puede reescribir como:

$$\eta_{\ell-1}(g^+) = {}^t\mathbf{g}^+ \mathbf{A}_{\ell-1} \mathbf{g}^+ = 2^{\ell-1} {}^t\mathbf{g}^+ \mathbf{B}_{\ell-1} \mathbf{g}^+ = 2^{\ell-1} {}^t\mathbf{w}^+ \tilde{\mathbf{D}}_{\ell-1} \mathbf{w}^+$$

$$= 2^{\ell-1} \left\{ (w_0^+)^2 + \sum_{0 \leq i < \ell-1} (w_{2^i}^+)^2 + \sum_{0 \leq i < j < \ell-1} (w_{2^i+2^j}^+)^2 \right\}$$

respectivamente,

$$\begin{aligned} \gamma_{\ell-1}(g^-) &= {}^t \mathbf{g}^- \mathbf{G}_{\ell-1} \mathbf{g}^- = 2^{\ell-1} {}^t \mathbf{g}^- \mathbf{E}_{\ell-1} \mathbf{g}^- = 2^{\ell-1} {}^t \mathbf{w}^- \mathbf{D}_{\ell-1} \mathbf{w}^- \\ &= 2^{\ell-1} \left\{ (w_0^-)^2 + \sum_{0 \leq i < \ell-1} (w_{2^i}^-)^2 \right\}. \end{aligned} \quad (3.21)$$

Ahora, teniendo en cuenta que  $\gamma_{\ell-1}(g^-) \geq 0$  y que

$$\eta_{\ell-1}(g^+) = \|\mathbf{g}^+\|^2 \eta_{\ell-1} \left( \frac{g^+}{\|\mathbf{g}^+\|} \right) \geq 4 \|\mathbf{g}^+\|^2 \geq 4,$$

se obtiene directamente que el valor mínimo de  $\eta_{\ell}(f) = \eta_{\ell-1}(g^+) + \gamma_{\ell-1}(g^-)$  se alcanza si  $\eta_{\ell-1}(g^+) = 4$  y  $\gamma_{\ell-1}(g^-) = 0$ , lo que es imposible si  $\|\mathbf{g}^+\| > 1$ .

Además, nosotros obtenemos:

**Proposición 3.4.3.** *Sea  $f$  una función de ajuste no negativa cuyo vector asociado  $\mathbf{f} \in \mathbb{R}^{2^\ell}$  es unitario y tiene la propiedad de que  $\|\mathbf{g}^+\| = 1$ . Entonces,*

$$1. \text{ Si } \ell = 3, \tilde{\varepsilon}_\ell^*(f) = 1 - \frac{1}{2^{\ell-2}} \text{ si, y sólo si, } \mathbf{f} = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \text{ o } \mathbf{f} = \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

$$2. \text{ Si } \ell \geq 4 \text{ entonces } \tilde{\varepsilon}_\ell^*(f) < 1 - \frac{1}{2^{\ell-2}}.$$

*Demostración.* Para demostrar la primera afirmación observamos que si  $\ell = 3$  entonces, de la expresión (3.21), se tiene que el único coeficiente de Walsh de  $\mathbf{w}^-$  no nulo debe ser  $\omega_3^-$ , ya que exigimos  $\gamma_{\ell-1}(g^-) = 0$ .

Por otra parte,

$${}^t \mathbf{f}^0 \mathbf{f}^1 = \sum_{k=0}^{2^{\ell-1}-1} \mathbf{f}_k^0 \mathbf{f}_k^1 = \sum_{k=0}^{2^{\ell-1}-1} \mathbf{f}_k \mathbf{f}_{2^{\ell-1}+k} = 0, \quad (3.22)$$

pues  $\|\mathbf{g}^+\| = \|\mathbf{f}\| = 1$ . Por supuesto, (3.22) es equivalente a

$$f_k f_{2^{\ell-1}+k} = 0, \quad \text{para todo } 0 \leq k < 2^{\ell-1}, \quad (3.23)$$

por ser  $f$  no negativa.

Además, para  $\ell = 3$  y  $\mathbf{w}^- = \mathbf{W}_2 \mathbf{g}^-$ , con  $\|\mathbf{g}^-\| = \|\mathbf{g}^+\| = 1$ , un cálculo sencillo demuestra que  $\omega_3^- = 1$  o  $\omega_3^- = -1$  y, por tanto,  $\mathbf{w}^- = {}^t(0, 0, 0, 1)$  o  $\mathbf{w}^- = {}^t(0, 0, 0, -1)$ , respectivamente. En consecuencia,

$${}^t\mathbf{g}^- = (0, 0, 0, 1)\mathbf{W}_2 = \left(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}\right),$$

respectivamente,

$${}^t\mathbf{g}^- = (0, 0, 0, -1)\mathbf{W}_2 = \left(-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right).$$

Observemos que, como todas las componentes de  $\mathbf{g}^- = \mathbf{f}^0 - \mathbf{f}^1$  son no nulas y  $f \geq 0$  entonces, por (3.23) directamente obtenemos que

$${}^t\mathbf{f} = \left(\frac{1}{2}, 0, 0, \frac{1}{2}, 0, \frac{1}{2}, \frac{1}{2}, 0\right)$$

o

$${}^t\mathbf{f} = \left(0, \frac{1}{2}, \frac{1}{2}, 0, \frac{1}{2}, 0, 0, \frac{1}{2}\right),$$

como asegurábamos.

Para demostrar la segunda afirmación, consideremos primero  $\ell = 4$  y, posteriormente, argumentaremos por inducción en  $\ell$ .

Utilizaremos las dos siguientes relaciones de recursión de  $\eta_\ell$  que se deducen directamente de (3.17) y (3.18).

$$\begin{aligned} \eta_\ell(f) &= \eta_{\ell-1}(g^+) + \gamma_{\ell-1}(g^-) \\ &= \eta_{\ell-1}(g^+) + \gamma_{\ell-2}(g^{-+}) + v_{\ell-2}(g^{--}) \end{aligned} \quad (3.24a)$$

y

$$\begin{aligned} \eta_\ell(f) &= \eta_{\ell-1}(g^+) + \gamma_{\ell-1}(g^-) \\ &= \eta_{\ell-2}(g^{++}) + \gamma_{\ell-2}(g^{+-}) + \gamma_{\ell-1}(g^-). \end{aligned} \quad (3.24b)$$

Estamos interesados en las funciones de ajuste con  $\eta_4(f) = 4$ . Usando la fórmula recurrente de  $\eta_\ell$  dada en (3.24a), el mismo argumento que el empleado en el caso  $\ell = 3$  nos conduce a que  $\eta_3(g^+) = 4$ , con  $\|\mathbf{g}^+\| = 1$ , y que  $\gamma_3(g^-) = 0$ , de donde se deduce que

$\gamma_2(g^{-+}) = v_2(g^{--}) = 0$ . Por tanto, obviamente,  $\mathbf{g}^+$  es uno de los vectores dados en la primera parte de la proposición. Es decir,

$${}^t\mathbf{g}^+ = \frac{1}{2}(1, 0, 0, 1, 0, 1, 1, 0)$$

o

$${}^t\mathbf{g}^+ = \frac{1}{2}(0, 1, 1, 0, 1, 0, 0, 1).$$

Comenzamos con el caso en que  ${}^t\mathbf{g}^+ = \frac{1}{2}(1, 0, 0, 1, 0, 1, 1, 0)$ .

Como  $\gamma_2(g^{-+}) = 0$  (lo que equivale a que  $g^{-+} \in V_2^0$ ) entonces  $\mathbf{g}^{-+}$  es un múltiplo del vector  $(1, -1, -1, 1)$ . Por ello, los únicos valores posibles para el vector  $\mathbf{g}^-$  son:

$$\mathbf{g}^- = \pm {}^t(1/2, 0, 0, 1/2, 0, -1/2, -1/2, 0).$$

Entonces, en el primer caso  $\mathbf{g}^{--} = {}^t(1/2, 1/2, 1/2, 1/2)$ , mientras que en el segundo  $\mathbf{g}^{--} = {}^t(-1/2, -1/2, -1/2, -1/2)$ . En ambos casos se verifica:

$$v_2(g^{--}) = {}^t\mathbf{g}^{--} \mathbf{U}_2 \mathbf{g}^{--} = \left( \sum_{k=0}^3 \mathbf{g}_k^{--} \right)^2 = 4$$

y, por tanto, el valor mínimo de  $\eta_4$  no se alcanza, i.e.,  $\eta_4 > 4$ .

De forma similar, si  ${}^t\mathbf{g}^+ = \frac{1}{2}(0, 1, 1, 0, 1, 0, 0, 1)$ , se obtienen los mismos valores para el vector  $\mathbf{g}^{--}$ .

Esto finaliza la demostración del caso  $\ell = 4$ .

Finalmente, para  $\ell \geq 4$ , iterando el proceso dado por (3.24b), y considerando las funciones

$$g^{\overbrace{+\dots+}^p} = \left( g^{\overbrace{+\dots+}^{p-1}} \right)^0 + \left( g^{\overbrace{+\dots+}^{p-1}} \right)^1,$$

y

$$g^{\overbrace{+\dots+}^{p-1}-} = \left( g^{\overbrace{+\dots+}^{p-1}} \right)^0 - \left( g^{\overbrace{+\dots+}^{p-1}} \right)^1$$

obtenemos que

$$\begin{aligned} \eta_\ell(f) &= \eta_{\ell-1}(g^+) + \gamma_{\ell-1}(g^-) \\ &= \eta_{\ell-2}(g^{++}) + \gamma_{\ell-2}(g^{+-}) + \gamma_{\ell-1}(g^-) \end{aligned}$$

$$\begin{aligned}
&= \eta_{\ell-3}(g^{+++}) + \gamma_{\ell-3}(g^{++-}) + \gamma_{\ell-2}(g^{+-}) + \gamma_{\ell-1}(g^{-}) \\
&\vdots \\
&= \eta_4(\overbrace{g^{+\cdots+}}^{\ell-4}) + \sum_{k=1}^{\ell-4} \gamma_{\ell-k}(\overbrace{g^{+\cdots+-}}^{k-1}) > 4,
\end{aligned}$$

ya que  $\eta_4 > 4$  y  $\gamma_{\ell-k} \geq 0$ , para todo  $k$ .

Esto finaliza la demostración.

□



# Capítulo 4

## Epistasis superior

---

### 4.1 Epistasis de orden superior

Como se ha comentado ya, en el análisis de la dificultad que presenta una función para ser optimizada por un Algoritmo Genético, se ha demostrado empíricamente que la epistasis normalizada es un buen estimador (ver [20] o [34], por ejemplo). Además, los resultados obtenidos en la sección 2.4 parecen indicar que el orden de la función objetivo es un buen complemento para medir esta dificultad.

En esta sección definiremos un nuevo estimador de la dificultad de una función que, a su vez, proporciona información sobre el orden de la misma. Una generalización del entramado algebraico de la epistasis normalizada conducirá al concepto de la *k-epistasis normalizada*.

#### 4.1.1 Las matrices $\mathbf{G}_{\ell,k}$

Este apartado se dedica a la construcción de un conjunto de matrices  $\mathbf{G}_{\ell,k}$ , siendo  $\ell$  y  $k$  enteros no negativos. Las introducimos de forma inductiva y, posteriormente, calculamos la expresión explícita de sus elementos.

Para  $\ell \geq 0$  y  $k = 0$ ,

$$\mathbf{G}_{\ell,0} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} = \mathbf{U}_\ell \in \mathcal{M}_{2^\ell}(\mathbb{Z}).$$

Cuando  $k \geq 1$  y  $\ell = 0$

$$\mathbf{G}_{0,k} = (1) \in \mathcal{M}_1(\mathbb{Z})$$

y, si  $\ell \geq 1$ , se define de forma recursiva:

$$\mathbf{G}_{\ell,k} = \begin{pmatrix} \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} \\ \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} \end{pmatrix} \in \mathcal{M}_{2^\ell}(\mathbb{Z}). \quad (4.1)$$

Por la construcción de la matriz  $\mathbf{G}_{\ell,k}$ , es fácil ver que, para  $k \geq \ell$ :

$$\mathbf{G}_{\ell,k} = 2^\ell \mathbf{I}_\ell, \quad (4.2)$$

donde  $\mathbf{I}_\ell$ , como siempre, denota la matriz identidad  $2^\ell$ -dimensional.

En el caso en que  $k < \ell$  se verifica:

**Proposición 4.1.1.** Para  $\ell > k$ , la matriz  $\mathbf{G}_{\ell,k} = \left( g_{st}^{\ell,k} \right)_{0 \leq s,t < 2^\ell}$  está dada por:

$$g_{st}^{\ell,k} = \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell-1-k+j}{j} \binom{\ell-d_{st}^\ell}{k-j}.$$

*Demostración.* Se realizará por inducción en el valor de  $\ell$ . Para  $\ell = 1$  el resultado es evidente. Se supone que la expresión es válida para  $\ell$ , y se demostrará para  $\ell + 1$ .

Por (4.1), si  $s, t \in \{0, \dots, 2^\ell - 1\}$  o  $s, t \in \{2^\ell, \dots, 2^{\ell+1} - 1\}$  se tiene:

$$g_{st}^{\ell+1,k} = g_{st}^{\ell,k} + g_{st}^{\ell,k-1},$$

siendo:

$$\begin{aligned} g_{st}^{\ell,k-1} &= \sum_{j=0}^{k-1} (-1)^j 2^{k-1-j} \binom{\ell-k+j}{j} \binom{\ell-d_{st}^\ell}{k-1-j} \\ &= 2 \sum_{j=0}^{k-1} (-1)^j 2^{k-1-j} \binom{\ell-k+j}{j} \binom{\ell-d_{st}^\ell}{k-1-j} \\ &\quad - \sum_{j=0}^{k-1} (-1)^j 2^{k-1-j} \binom{\ell-k+j}{j} \binom{\ell-d_{st}^\ell}{k-1-j} \end{aligned}$$

$$\begin{aligned}
&= \sum_{j=0}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-1-j} \\
&\quad + \sum_{j=0}^{k-1} (-1)^{j+1} 2^{k-1-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-1-j} \\
&= \sum_{j=0}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-1-j} \\
&\quad + \sum_{j=1}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j-1} \binom{\ell - d_{st}^\ell}{k-j}.
\end{aligned}$$

Entonces,

$$\begin{aligned}
g_{st}^{\ell+1,k} &= g_{st}^{\ell,k} + g_{st}^{\ell,k-1} \\
&= \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j} \binom{\ell - d_{st}^\ell}{k-j} \\
&\quad + \sum_{j=1}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j-1} \binom{\ell - d_{st}^\ell}{k-j} \\
&\quad + \sum_{j=0}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-1-j} \\
&= 2^k \binom{\ell - d_{st}^\ell}{k} + \sum_{j=1}^k (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-j} \\
&\quad + \sum_{j=0}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-1-j} \\
&= 2^k \binom{\ell - d_{st}^\ell}{k} + \sum_{j=1}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-j} \\
&\quad + (-1)^k \binom{\ell}{k} + 2^k \binom{\ell - d_{st}^\ell}{k-1} \\
&\quad + \sum_{j=1}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^\ell}{k-1-j} \\
&= 2^k \binom{\ell + 1 - d_{st}^\ell}{k} + \sum_{j=1}^{k-1} (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell + 1 - d_{st}^\ell}{k-j} + (-1)^k \binom{\ell}{k} \\
&= \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell + 1 - d_{st}^\ell}{k-j}.
\end{aligned}$$

Además, en este caso,  $d_{st}^{\ell+1} = d_{st}^{\ell}$ , por lo que:

$$g_{st}^{\ell+1,k} = \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell + 1 - d_{st}^{\ell+1}}{k - j}.$$

Por otra parte, si  $s \in \{0, \dots, 2^{\ell} - 1\}$ ,  $t \in \{2^{\ell}, \dots, 2^{\ell+1} - 1\}$ :

$$\begin{aligned} g_{st}^{\ell+1,k} &= g_{st}^{\ell,k} - g_{st}^{\ell,k-1} \\ &= \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j} \binom{\ell - d_{st}^{\ell}}{k - j} \\ &\quad - \sum_{j=0}^{k-1} (-1)^j 2^{k-1-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^{\ell}}{k - 1 - j} \\ &= \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j} \binom{\ell - d_{st}^{\ell}}{k - j} \\ &\quad - \sum_{j=1}^k (-1)^{j+1} 2^{k-j} \binom{\ell - 1 - k + j}{j - 1} \binom{\ell - d_{st}^{\ell}}{k - j} \\ &= \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j} \binom{\ell - d_{st}^{\ell}}{k - j} \\ &\quad + \sum_{j=1}^k (-1)^j 2^{k-j} \binom{\ell - 1 - k + j}{j - 1} \binom{\ell - d_{st}^{\ell}}{k - j} \\ &= 2^k \binom{\ell - d_{st}^{\ell}}{k} + \sum_{j=1}^k (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^{\ell}}{k - j} \\ &= \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell - d_{st}^{\ell}}{k - j}. \end{aligned}$$

Teniendo en cuenta que ahora  $d_{st}^{\ell} = d_{st}^{\ell+1} - 1$  se obtiene:

$$g_{st}^{\ell+1,k} = \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell - k + j}{j} \binom{\ell + 1 - d_{st}^{\ell+1}}{k - j}.$$

□

### 4.1.2 Propiedades algebraicas

**Lema 4.1.2.** Para  $0 \leq k \leq \ell$ , la matriz  $\mathbf{G}_{\ell,k}$  verifica:

$$\sum_{0 \leq s, t < 2^\ell} g_{st}^{\ell,k} = 2^{2^\ell}. \quad (4.3)$$

*Demostración.* La demostración se realizará por inducción en  $\ell$ . Si  $\ell = 0$ ,  $\mathbf{G}_{0,0} = (1)$ , por lo que el resultado es evidente. Supongamos que se verifica para  $\ell - 1$  y demostrémosla para  $\ell$ .

El uso de (4.1) permite escribir

$$\sum_{0 \leq s, t < 2^\ell} g_{st}^{\ell,k} = 4 \sum_{0 \leq s, t < 2^{\ell-1}} g_{st}^{\ell-1,k} = 4 \cdot 2^{2^{\ell-1}} = 2^{2^\ell}.$$

□

Para determinar el rango de  $\mathbf{G}_{\ell,k}$ , observemos que la relación (4.2) proporciona directamente que  $\text{Rg } \mathbf{G}_{\ell,k} = 2^\ell$  si  $\ell \leq k$ . En el caso  $\ell > k$  observemos, en primer lugar, que  $\text{Rg } \mathbf{G}_{\ell,0} = 1$  y, cuando  $k \geq 1$ , el uso de transformaciones elementales permite afirmar que  $\mathbf{G}_{\ell,k}$  es equivalente a

$$\begin{pmatrix} \mathbf{G}_{\ell-1,k} & \mathbf{O}_{\ell-1} \\ \mathbf{O}_{\ell-1} & \mathbf{G}_{\ell-1,k-1} \end{pmatrix}, \quad (4.4)$$

siendo  $\mathbf{O}_{\ell-1} = \begin{pmatrix} 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \in \mathcal{M}_{2^{\ell-1}}(\mathbb{Z})$ , de lo que se deduce:

$$\text{Rg } \mathbf{G}_{\ell,k} = \text{Rg } \mathbf{G}_{\ell-1,k} + \text{Rg } \mathbf{G}_{\ell-1,k-1}. \quad (4.5)$$

Entonces:

**Proposición 4.1.3.** Para  $\ell > k \geq 0$  se verifica:

$$\text{Rg } \mathbf{G}_{\ell,k} = \sum_{j=0}^k \binom{\ell}{j}.$$

*Demostración.* La construcción de  $\mathbf{G}_{\ell,k}$  permite crear un diagrama de tipo árbol que ayuda a calcular su rango en función del rango de las matrices  $\mathbf{G}_{j,i}$ , con  $i \leq k$  y  $j \leq \ell$ :

$$\begin{array}{cccc}
& & & \mathbf{G}_{\ell,k} \\
& & & \mathbf{G}_{\ell-1,k} \\
& & \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-2,k} \\
& & \mathbf{G}_{\ell-2,k-2} & \mathbf{G}_{\ell-2,k-1} \\
& \cdots & \vdots & \vdots \\
\mathbf{G}_{\ell-k,0} & \cdots & \mathbf{G}_{\ell-k,k-2} & \mathbf{G}_{\ell-k,k-1} & \mathbf{G}_{\ell-k,k} \\
\mathbf{G}_{\ell-k-1,0} & \cdots & \mathbf{G}_{\ell-k-1,k-2} & \mathbf{G}_{\ell-k-1,k-1} & \mathbf{G}_{\ell-k-1,k} \\
\vdots & & \vdots & \vdots & \vdots \\
\mathbf{G}_{0,0} & \cdots & \mathbf{G}_{0,k-2} & \mathbf{G}_{0,k-1} & \mathbf{G}_{0,k}
\end{array}$$

Usando recursivamente (4.5) tenemos

$$\begin{aligned}
\text{Rg } \mathbf{G}_{\ell,k} &= \text{Rg } \mathbf{G}_{\ell-1,k} + \text{Rg } \mathbf{G}_{\ell-1,k-1} \\
&= \text{Rg } \mathbf{G}_{\ell-2,k} + 2 \text{Rg } \mathbf{G}_{\ell-2,k-1} + \text{Rg } \mathbf{G}_{\ell-2,k-2} \\
&\vdots \\
&= \sum_{j=0}^k \binom{\ell}{\ell-j} \text{Rg } \mathbf{G}_{0,k-j} = \sum_{j=0}^k \binom{\ell}{j}.
\end{aligned}$$

□

**Proposición 4.1.4.** La matriz  $\mathbf{G}_{\ell,k}$  ( $\ell \geq k \geq 0$ ) verifica  $(\mathbf{G}_{\ell,k})^2 = 2^\ell \mathbf{G}_{\ell,k}$ .

*Demostración.* Realizaremos la demostración por inducción en el valor de  $k$ .

Si  $k = 0$ ,  $\mathbf{G}_{\ell,0} = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$ , por lo que es evidente que se cumple la relación para todo  $\ell \geq 0$ .

Supongamos que se verifica la propiedad para  $1, \dots, k-1$  y para cualquier  $\ell \geq 0$ . Demostraremos el resultado para  $k$ , y cualquier  $\ell \geq 0$ , por inducción en  $\ell$ .

Para  $\ell = 0$ ,  $\mathbf{G}_{0,0} = (1)$ , por lo que el resultado es obviamente cierto. Se supone ahora que  $(\mathbf{G}_{\ell-1,k})^2 = 2^{\ell-1} \mathbf{G}_{\ell-1,k}$ , y se demostrará para  $\mathbf{G}_{\ell,k}$ .

$$\begin{aligned}
\mathbf{G}_{\ell,k}^2 &= \begin{pmatrix} \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} \\ \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} \end{pmatrix}^2 \\
&= \begin{pmatrix} 2\mathbf{G}_{\ell-1,k}^2 + 2\mathbf{G}_{\ell-1,k-1}^2 & 2\mathbf{G}_{\ell-1,k}^2 - 2\mathbf{G}_{\ell-1,k-1}^2 \\ 2\mathbf{G}_{\ell-1,k}^2 - 2\mathbf{G}_{\ell-1,k-1}^2 & 2\mathbf{G}_{\ell-1,k}^2 + 2\mathbf{G}_{\ell-1,k-1}^2 \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
&= 2 \begin{pmatrix} \mathbf{G}_{\ell-1,k}^2 + \mathbf{G}_{\ell-1,k-1}^2 & \mathbf{G}_{\ell-1,k}^2 - \mathbf{G}_{\ell-1,k-1}^2 \\ \mathbf{G}_{\ell-1,k}^2 - \mathbf{G}_{\ell-1,k-1}^2 & \mathbf{G}_{\ell-1,k}^2 + \mathbf{G}_{\ell-1,k-1}^2 \end{pmatrix} \\
&= 2 \begin{pmatrix} 2^{\ell-1} [\mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1}] & 2^{\ell-1} [\mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1}] \\ 2^{\ell-1} [\mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1}] & 2^{\ell-1} [\mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1}] \end{pmatrix} \\
&= 2 \cdot 2^{\ell-1} \begin{pmatrix} \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} \\ \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} \end{pmatrix} \\
&= 2^\ell \mathbf{G}_{\ell,k}.
\end{aligned}$$

□

Al igual que sucedía con la 2-epistasis, de este resultado se sigue directamente:

**Corolario 4.1.5.** *Los autovalores de la matriz  $\mathbf{G}_{\ell,k}$  son 0 y  $2^\ell$ .*

Si para  $\ell$  y  $k$  enteros no negativos se considera la matriz  $2^\ell$ -dimensional con coeficientes racionales dada por  $\mathbf{E}_{\ell,k} = \frac{1}{2^\ell} \mathbf{G}_{\ell,k}$  entonces, de esta relación, se deduce inmediatamente que  $\mathbf{E}_{\ell,k}$  es simétrica e idempotente, lo que implica que sus autovalores son 0 y 1.

Análogamente a como se hacía en la página 87, dados  $0 \leq k \leq \ell$ , denotamos ahora por  $V_{\ell,k}^0$  y  $V_{\ell,k}^1$  los espacios de autovectores en  $\mathbb{R}^{2^\ell}$  correspondientes, respectivamente, a los autovalores 0 y 1 de  $\mathbf{E}_{\ell,k}$ .

Entonces  $\mathbb{R}^{2^\ell} = V_{\ell,k}^0 \oplus V_{\ell,k}^1$ , siendo  $V_{\ell,k}^0 = \ker(\mathbf{E}_{\ell,k})$  y  $V_{\ell,k}^1 = \text{Im}(\mathbf{E}_{\ell,k})$ . Además,

$$\dim V_{\ell,k}^1 = \sum_{j=0}^k \binom{\ell}{j}$$

y

$$\dim V_{\ell,k}^0 = 2^\ell - \sum_{j=0}^k \binom{\ell}{j} = \sum_{j=k+1}^{\ell} \binom{\ell}{j}.$$

Asimismo, para  $\ell, k \geq 0$ , denotamos por  $\mathbf{D}_{\ell,k}$  a la matriz diagonal en  $\mathcal{M}_{2^\ell}(\mathbb{Z})$  que definimos como sigue:

- $\mathbf{D}_{\ell,0}$  es la matriz cuyos elementos son todos nulos, excepto  $d_{00} = 1$ ,
- para  $\ell > k > 0$ , los únicos elementos no nulos de  $\mathbf{D}_{\ell,k}$ ,  $d_{ss} = 1$ , están situados en las filas  $s = 0, s = 2^{i_1}$ , con  $0 \leq i_1 < \ell$ ,  $s = 2^{i_1} + 2^{i_2}$ , con  $0 \leq i_1 < i_2 < \ell, \dots$ ,  $s = 2^{i_1} + 2^{i_2} + \dots + 2^{i_k}$ , con  $0 \leq i_1 < i_2 < \dots < i_k < \ell$ ,

- para  $k \geq \ell > 0$ ,  $\mathbf{D}_{\ell,k} = \mathbf{I}_\ell$ .

**Observación:** la matriz  $\tilde{\mathbf{D}}_\ell$  considerada en la página 87 es el caso particular con  $k = 2$  de las matrices  $\mathbf{D}_{\ell,k}$ .

Si  $\ell \geq k > 0$ , las matrices  $\mathbf{D}_{\ell,k}$  se pueden construir por recurrencia. En efecto,

**Proposición 4.1.6.** Para  $\ell \geq k > 0$  se verifica:

$$\mathbf{D}_{\ell,k} = \begin{pmatrix} \mathbf{D}_{\ell-1,k} & \mathbf{O}_{\ell-1} \\ \mathbf{O}_{\ell-1} & \mathbf{D}_{\ell-1,k-1} \end{pmatrix},$$

donde, como siempre,  $\mathbf{O}_{\ell-1}$  denota la matriz nula  $2^{\ell-1}$ -dimensional.

Ahora no es difícil comprobar que:

**Proposición 4.1.7.** Para cualquier  $\ell \geq 0$  se verifica:

$$\mathbf{W}_\ell \mathbf{E}_{\ell,k} \mathbf{W}_\ell = \mathbf{D}_{\ell,k}.$$

*Demostración.* Si  $k \geq \ell$ , el resultado es consecuencia inmediata de la idempotencia de  $\mathbf{W}_\ell$ .

Si  $\ell > k$ , el resultado se probará por inducción en  $k$ .

Para  $k = 0$

$$\begin{aligned} \mathbf{W}_\ell \mathbf{E}_{\ell,0} \mathbf{W}_\ell &= 2^{-\frac{\ell}{2}} \mathbf{V}_\ell 2^{-\ell} \mathbf{G}_{\ell,0} 2^{-\frac{\ell}{2}} \mathbf{V}_\ell = 2^{-2\ell} \mathbf{V}_\ell \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \mathbf{V}_\ell \\ &= 2^{-\ell} \begin{pmatrix} 1 & \dots & 1 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{pmatrix} \mathbf{V}_\ell = \mathbf{D}_{\ell,0}. \end{aligned}$$

Supongamos que hasta  $k$ , para todo  $\ell > k$ ,  $\mathbf{W}_\ell \mathbf{E}_{\ell,k} \mathbf{W}_\ell = \mathbf{D}_{\ell,k}$ . Demostraremos, por inducción en el valor de  $\ell$ , la igualdad para  $k + 1$ .

Si  $\ell = 1$  entonces  $k = 0$  y el resultado se reduce al caso anterior. Ahora, supuesto cierto hasta  $\ell$ , se demuestra para  $\ell + 1$ . Para ello, teniendo en cuenta las relaciones de recursividad de  $\mathbf{G}_{\ell,k}$  y  $\mathbf{W}_\ell$ , dadas por (4.1) y (2.6) (ver página 47) respectivamente, tenemos:

$$\begin{aligned}
\mathbf{W}_{\ell+1} \mathbf{E}_{\ell+1,k+1} \mathbf{W}_{\ell+1} &= 2^{-\frac{1}{2}} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} 2^{-(\ell+1)} \mathbf{G}_{\ell+1,k+1} 2^{-\frac{1}{2}} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \\
&= 2^{-(\ell+2)} \left[ \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \begin{pmatrix} \mathbf{G}_{\ell,k+1} & \mathbf{G}_{\ell,k+1} \\ \mathbf{G}_{\ell,k+1} & \mathbf{G}_{\ell,k+1} \end{pmatrix} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \right. \\
&\quad \left. + \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \begin{pmatrix} \mathbf{G}_{\ell,k} & -\mathbf{G}_{\ell,k} \\ -\mathbf{G}_{\ell,k} & \mathbf{G}_{\ell,k} \end{pmatrix} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \right] \\
&= 2^{-(\ell+2)} \left[ \begin{pmatrix} 2\mathbf{W}_\ell \mathbf{G}_{\ell,k+1} & 2\mathbf{W}_\ell \mathbf{G}_{\ell,k+1} \\ \mathbf{O}_\ell & \mathbf{O}_\ell \end{pmatrix} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \right. \\
&\quad \left. + \begin{pmatrix} \mathbf{O}_\ell & \mathbf{O}_\ell \\ 2\mathbf{W}_\ell \mathbf{G}_{\ell,k} & -2\mathbf{W}_\ell \mathbf{G}_{\ell,k} \end{pmatrix} \begin{pmatrix} \mathbf{W}_\ell & \mathbf{W}_\ell \\ \mathbf{W}_\ell & -\mathbf{W}_\ell \end{pmatrix} \right] \\
&= 2^{-(\ell+2)} \left[ \begin{pmatrix} 4\mathbf{W}_\ell \mathbf{G}_{\ell,k+1} \mathbf{W}_\ell & \mathbf{O}_\ell \\ \mathbf{O}_\ell & \mathbf{O}_\ell \end{pmatrix} + \begin{pmatrix} \mathbf{O}_\ell & \mathbf{O}_\ell \\ \mathbf{O}_\ell & 4\mathbf{W}_\ell \mathbf{G}_{\ell,k} \mathbf{W}_\ell \end{pmatrix} \right] \\
&= \begin{pmatrix} 2^{-\ell} \mathbf{W}_\ell \mathbf{G}_{\ell,k+1} \mathbf{W}_\ell & \mathbf{O}_\ell \\ \mathbf{O}_\ell & 2^{-\ell} \mathbf{W}_\ell \mathbf{G}_{\ell,k} \mathbf{W}_\ell \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{W}_\ell \mathbf{E}_{\ell,k+1} \mathbf{W}_\ell & \mathbf{O}_\ell \\ \mathbf{O}_\ell & \mathbf{W}_\ell \mathbf{E}_{\ell,k} \mathbf{W}_\ell \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{D}_{\ell,k+1} & \mathbf{O}_\ell \\ \mathbf{O}_\ell & \mathbf{D}_{\ell,k} \end{pmatrix} \\
&= \mathbf{D}_{\ell+1,k+1}.
\end{aligned}$$

□

El siguiente resultado proporciona una base para  $V_{\ell,k}^1$ .

**Proposición 4.1.8.** *Para cualquier entero positivo  $\ell$ , el conjunto formado por los vectores columna de la matriz  $\mathbf{W}_\ell$  situados en las posiciones  $i = 0, i = 2^{j_1} + \dots + 2^{j_p}, 0 \leq j_1 < \dots < j_p < \ell$ , con  $1 \leq p \leq k$ , es una base para  $V_{\ell,k}^1$ .*

*Demostración.* Dada la base canónica de  $\mathbb{R}^{2^\ell}$ ,  $\{\mathbf{e}_i; 0 \leq i < 2^\ell\}$ , consideramos el conjunto formado por los vectores  $\mathbf{w}_m = \mathbf{W}_\ell \mathbf{e}_m$ , con  $m = 0, m = 2^{j_1} + \dots + 2^{j_p}, 0 \leq j_1 < \dots < j_p < \ell$ , para cada  $p$  tal que  $1 \leq p \leq k$ . Éste es claramente un conjunto de vectores linealmente

independientes, y como su cardinal  $\sum_{i=0}^k \binom{\ell}{i}$  coincide con la dimensión de  $V_{\ell,k}^1$ , sólo falta demostrar que cada uno de esos vectores pertenece a  $V_{\ell,k}^1 = \text{Im}(\mathbf{E}_{\ell,k})$ .

El hecho de que, por la definición de  $\mathbf{D}_{\ell,k}$ , se verifique que  $\mathbf{e}_m = \mathbf{D}_{\ell,k} \mathbf{e}_m$ , para todo  $m \in \{0, 2^{j_1} + \dots + 2^{j_p}; 0 \leq j_1 < \dots < j_p < \ell, 1 \leq p \leq k\}$ , junto con la idempotencia de  $\mathbf{W}_\ell$ , conduce a:

$$\begin{aligned} \mathbf{w}_m &= \mathbf{W}_\ell \mathbf{e}_m = \mathbf{W}_\ell \mathbf{D}_{\ell,k} \mathbf{e}_m \\ &= \mathbf{W}_\ell \mathbf{D}_{\ell,k} \mathbf{W}_\ell \mathbf{W}_\ell \mathbf{e}_m \\ &= \mathbf{W}_\ell \mathbf{D}_{\ell,k} \mathbf{W}_\ell \mathbf{w}_m \\ &= \mathbf{E}_{\ell,k} \mathbf{w}_m, \end{aligned}$$

lo que concluye la demostración.  $\square$

### 4.1.3 $k$ -epistasis

**Definición 4.1.9.** Dada una función  $f : \Omega_\ell \rightarrow \mathbb{R}$ , la  $k$ -epistasis normalizada ( $k \leq \ell$ ) se define como:

$$\varepsilon_{\ell,k}^*(f) = 1 - \frac{{}^t \mathbf{f} \mathbf{E}_{\ell,k} \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}} = 1 - \frac{1}{{}^t \mathbf{f} \mathbf{f}} \frac{{}^t \mathbf{f} \mathbf{G}_{\ell,k} \mathbf{f}}{2^\ell}. \quad (4.6)$$

Como  $\mathbf{E}_{\ell,k}$  es una proyección ortogonal entonces, al igual que para la epistasis y la 2-epistasis normalizadas,  $0 \leq \varepsilon_{\ell,k}^*(f) \leq 1$ . De hecho, ambas son casos particulares de este nuevo estimador.

El valor mínimo teórico  $\varepsilon_{\ell,k}^*(f) = 0$  se alcanza cuando  $\mathbf{f} = \mathbf{E}_{\ell,k} \mathbf{f}$ , es decir, cuando  $f \in V_{\ell,k}^1$ . Por ejemplo, si para cualesquiera enteros  $\ell, k$  ( $0 \leq k \leq \ell$ ), se considera la función  $u$ , con vector asociado  $\mathbf{u} = {}^t(1, \dots, 1) \in \mathbb{R}^{2^\ell}$  y si se tiene en cuenta (4.3) (ver página 125), entonces se verifica

$$\varepsilon_{\ell,k}^*(u) = 1 - \frac{1}{{}^t \mathbf{u} \mathbf{u}} \frac{{}^t \mathbf{u} \mathbf{G}_{\ell,k} \mathbf{u}}{2^\ell} = 1 - \frac{1}{{}^t \mathbf{u} \mathbf{u}} \frac{\sum_{0 \leq s, t < 2^\ell} g_{st}^{\ell,k}}{2^\ell} = 1 - \frac{2^{2\ell}}{2^{2\ell}} = 0.$$

El cálculo práctico de la  $k$ -epistasis normalizada se puede realizar en función de los coeficientes de Walsh de la función:

**Proposición 4.1.10.** Sea  $f : \Omega_\ell \rightarrow \mathbb{R}$  una función con coeficientes de Walsh  $w_0, \dots, w_{2^\ell-1}$ , entonces la  $k$ -epistasis normalizada de  $f$  toma el valor:

$$\varepsilon_{\ell,k}^*(f) = 1 - \frac{w_0^2 + \sum_{0 \leq i < \ell} w_{2^i}^2 + \dots + \sum_{0 \leq i_1 < \dots < i_k < \ell} w_{2^{i_1} + \dots + 2^{i_k}}^2}{\sum_{0 \leq j < 2^\ell} w_j^2}. \quad (4.7)$$

*Demostración:* Omitiremos los subíndices en la demostración.

Se verifica, por ser  $\mathbf{W}$  simétrica e idempotente,

$${}^t\mathbf{f}\mathbf{f} = {}^t(\mathbf{W}\mathbf{w})(\mathbf{W}\mathbf{w}) = {}^t\mathbf{w}{}^t\mathbf{W}\mathbf{W}\mathbf{w} = {}^t\mathbf{w}\mathbf{w}.$$

Por otra parte, como consecuencia de la proposición 4.1.7, tenemos

$${}^t\mathbf{f}\mathbf{E}\mathbf{f} = ({}^t\mathbf{w}{}^t\mathbf{W})\mathbf{E}(\mathbf{W}\mathbf{w}) = {}^t\mathbf{w}\mathbf{D}\mathbf{w},$$

y, por tanto, por (4.6),

$$\varepsilon^*(f) = 1 - \frac{{}^t\mathbf{f}\mathbf{E}\mathbf{f}}{{}^t\mathbf{f}\mathbf{f}} = 1 - \frac{{}^t\mathbf{w}\mathbf{D}\mathbf{w}}{{}^t\mathbf{w}\mathbf{w}}.$$

□

## 4.2 $k$ -epistasis normalizada de algunas funciones

### 4.2.1 Función de Dirac

Al igual que en la subsección 3.2.1, consideramos la función de Dirac  $\delta_0$ , cuyo vector asociado es:

$$\delta_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

y su vector de Walsh es

$$\mathbf{w} = \mathbf{W}\delta_0 = 2^{-\ell/2} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Así,  $w_s = 2^{-\ell/2}$  para todo  $s \in \Omega_\ell$  y  $\|\mathbf{w}\| = 1$ . Entonces el valor de la  $k$ -epistasis normalizada es:

$$\begin{aligned} \varepsilon_{\ell,k}^*(\delta_0) &= 1 - \frac{w_0^2 + \sum_{j=1}^k \left( \sum_{0 \leq i_1 < \dots < i_j < \ell} w_{2^{i_1} + \dots + 2^{i_j}}^2 \right)}{\|\mathbf{w}\|^2} \\ &= 1 - \left( 1 + \sum_{j=1}^k \binom{\ell}{j} \right) (2^{-\ell/2})^2 \end{aligned}$$

$$= 1 - \frac{\sum_{j=0}^k \binom{\ell}{j}}{2^\ell}.$$

Claramente, fijado  $k$ , a medida que  $\ell$  aumenta,  $\varepsilon_{\ell,k}^*(\delta_0)$  tiende a 1, como era de esperar. Recíprocamente, fijado  $\ell$ , a medida que aumenta  $k$ , la  $k$ -epistasis normalizada sobre  $\Omega_\ell$  disminuye. En la tabla 4.1 se recogen, a modo de ejemplo, los valores obtenidos para las longitudes  $\ell = 10, 11, 12$  y  $13$  de las  $k$ -epistasis correspondientes.

$k$ -epistasis	$\ell = 10$	$\ell = 11$	$\ell = 12$	$\ell = 13$
0	0.9902	0.9995	0.9997	0.9998
1	0.9892	0.9941	0.9968	0.9982
2	0.9453	0.9672	0.9807	0.9887
3	0.8281	0.8867	0.927	0.9538
4	0.623	0.7255	0.8061	0.8665
5	0.3769	0.5	0.6127	0.7094
6	0.1718	0.2744	0.3872	0.5
7	0.0546	0.1132	0.1938	0.2905
8	0.0107	0.0327	0.0729	0.1334
9	0.0009	0.0058	0.0192	0.0461
10	0	0.0004	0.0031	0.0112
11	0	0	0.0002	0.0017
12	0	0	0	0.0001
13	0	0	0	0

Tabla 4.1:  $k$ -epistasis de la función de Dirac.

En las cuatro longitudes se observa que las epistasis de orden bajo toman valores muy altos, indicando así una gran dificultad de la función.

## 4.2.2 Funciones Camello

También ahora estudiamos las funciones camello consideradas ya en la subsección 3.2.2. Al igual que entonces (y sin pérdida de generalidad) consideramos la función camello  $c$  definida por  $c(0\dots 0) = c(1\dots 1) = 1$ , y  $c(t) = 0$  para los demás valores de  $t$  en  $\Omega_\ell$ . Así, la representación

vectorial de  $c$  ( $= \delta_0 + \delta_{2^\ell - 1}$ ) es:

$$\mathbf{c} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

y por (3.5), para cada  $s \in \Omega_\ell$ :

$$w_s = 2^{-\ell/2} \left( 1 + (-1)^{u(s)} \right).$$

donde, como siempre,  $u(s)$  es el número de bits no nulos en  $s$ . En particular:

$$w_0 = 2 \cdot 2^{-\ell/2}$$

y

$$w_{2^{i_1} + \dots + 2^{i_q}} = \begin{cases} 2 \cdot 2^{-\ell/2} & \text{si } q \text{ es par} \\ 0 & \text{si } q \text{ es impar.} \end{cases} \quad (4.8)$$

**Nota:** Como consecuencia de (4.8) se tiene que si  $\ell$  es par, la función  $c$  es de orden  $\ell$ , mientras que si es impar, su orden  $\ell - 1$ .

El valor de la  $k$ -epistasis normalizada ( $k \leq \ell$ ) es:

$$\begin{aligned} \varepsilon_{\ell,k}^*(c) &= 1 - \frac{w_0^2 + \sum_{j=1}^k \binom{\ell}{j} w_{2^{i_1} + \dots + 2^{i_j}}^2}{\|\mathbf{w}\|^2} \\ &= 1 - \frac{w_0^2 + \sum_{j=1}^m \binom{\ell}{2j} w_{2^{i_1} + \dots + 2^{i_{2j}}}^2}{\|\mathbf{w}\|^2} \\ &= 1 - \frac{(2 \cdot 2^{-\ell/2})^2 + \sum_{j=1}^m \binom{\ell}{2j} (2 \cdot 2^{-\ell/2})^2}{1^2 + 1^2} \\ &= 1 - \frac{\sum_{j=0}^m \binom{\ell}{2j}}{2^{\ell-1}}, \end{aligned}$$

donde  $m$  es el mayor entero tal que  $2m \leq k$ .

### 4.2.3 Funciones cuasi-camello

Introducimos a continuación una función que usaremos en el apéndice C, que denominamos función cuasi-camello,  $d$ , definida sobre cadenas de  $\Omega_\ell$  así:  $d(0\dots 00) = d(1\dots 10) = 1$ , y  $d(t) = 0$  para los demás valores de  $t$ . La representación vectorial de  $d$  es:

$$\mathbf{d} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Por la relación entre  $\mathbf{w}$  y  $\mathbf{d}$  se tiene, para cada  $s = (s_{\ell-1}, \dots, s_1, s_0) \in \Omega_\ell$

$$\begin{aligned} w_s &= (\mathbf{w})_s = (\mathbf{Wd})_s \\ &= 2^{-\ell/2} (\psi_{0\dots 00}(s) + \psi_{1\dots 10}(s)) \\ &= 2^{-\ell/2} \left( (-1)^{0 \cdot s} + (-1)^{(2^\ell - 2) \cdot s} \right) \\ &= 2^{-\ell/2} \left( 1 + (-1)^{u(s_{\ell-1}, \dots, s_1)} \right), \end{aligned}$$

donde  $u(s_{\ell-1}, \dots, s_1)$  denota el número de bits no nulos en la subcadena  $(s_{\ell-1}, \dots, s_1)$  de  $s$ . En particular:

$$w_0 = 2^{-\ell/2} (1 + (-1)^0) = 2 \cdot 2^{-\ell/2}$$

y los coeficientes de Walsh de orden 1 son

$$w_{2^0} = 2^{-\ell/2} (1 + (-1)^0) = 2 \cdot 2^{-\ell/2}$$

y, si  $i \neq 0$ ,

$$w_{2^i} = 2^{-\ell/2} (1 + (-1)^1) = 0.$$

En general, para  $2 \leq j \leq \ell$  y  $1 \leq i_1 < \dots < i_{j-1} < \ell$ ,

$$w_{2^{0+2^{i_1}+\dots+2^{i_{j-1}}}} = \begin{cases} 0 & \text{si } j \text{ es par,} \\ 2 \cdot 2^{-\ell/2} & \text{si } j \text{ es impar.} \end{cases}$$

y, si  $1 \leq i_1 < \dots < i_j < \ell$ ,

$$w_{2^{i_1+\dots+2^{i_j}}} = \begin{cases} 2 \cdot 2^{-\ell/2} & \text{si } j \text{ es par,} \\ 0 & \text{si } j \text{ es impar.} \end{cases}$$

De esta forma, se puede asegurar que existen  $\binom{\ell-1}{j}$  coeficientes de orden  $j$  no nulos si  $j$  es par y  $\binom{\ell-1}{j-1}$  coeficientes de orden  $j$  no nulos si  $j$  es impar.

Finalmente, el valor de la  $k$ -epistasia normalizada de la función cuasi-camello es:

$$\varepsilon_{\ell,k}^*(d) = \begin{cases} 1 - 2^{-\ell+1} \left\{ 2 \sum_{j=0}^{m-1} \binom{\ell-1}{2j} + \binom{\ell-1}{2m} \right\} & \text{si } k = 2m, \\ 1 - 2^{-\ell+2} \sum_{j=0}^m \binom{\ell-1}{2j} & \text{si } k = 2m + 1. \end{cases}$$

#### 4.2.4 Funciones de Ponderación

En este apartado se calcula la  $k$ -epistasia normalizada de las llamadas funciones de ponderación ya descritas en la subsección 3.2.3. Para ello se generaliza la expresión que ha permitido calcular los coeficientes de Walsh de segundo orden para este tipo de funciones (ver página 94).

En primer lugar recordemos que el valor del coeficiente de Walsh de orden 0 de una función de ponderación  $f$  (calculado en [33]) es

$$w_0 = 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \binom{\ell}{u} h(u). \quad (4.9)$$

donde  $h$  es la función asociada a  $f$ , i.e.,  $f(s) = h(u(s))$ ,  $s \in \Omega_{\ell}$  y  $u(s)$  es el número de unos de  $s$ .

Ahora nosotros tenemos.

**Proposición 4.2.1.** *Si  $f$  es una función de ponderación, sus coeficientes de Walsh  $w_t$ , con  $t = 2^{i_1} + \dots + 2^{i_p}$ ,  $0 \leq i_1 < \dots < i_p < \ell$  son:*

$$w_{2^{i_1} + \dots + 2^{i_p}} = 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \left[ \sum_{j=0}^p (-1)^j \binom{p}{j} \binom{\ell-p}{u-j} \right] h(u).$$

*Demostración.* La demostración para  $p = 1$  puede verse en [33] (nosotros hemos incluido la expresión de  $w_{2^i}$  en la fórmula (3.6) de la página 94).

Como hipótesis de inducción se considera cierto el resultado para el caso  $p$  y se comprobará para  $p + 1$ .

Para ello consideramos el esquema  $\mathcal{H} = \# \dots \# \overset{i_{p+1}}{\downarrow} 0 \# \dots \# \overset{i_p}{\downarrow} 0 \# \dots \# \overset{i_1}{\downarrow} 0 \# \dots \#$ , cuyo valor medio es:

$$\begin{aligned} f(\mathcal{H}) &= \frac{1}{2^{\ell-(p+1)}} \sum_{s \in \mathcal{H}} f(s) = 2^{-(\ell-(p+1))} \sum_{u=0}^{\ell-(p+1)} \binom{\ell-(p+1)}{u} h(u) \\ &= 2^{-(\ell-(p+1))} \sum_{u=0}^{\ell} \binom{\ell-(p+1)}{u} h(u). \end{aligned}$$

Por otra parte, de (2.12) (ver página 50) y de (4.9), utilizando la hipótesis de inducción, tenemos:

$$\begin{aligned} w_{2^{i_1} + \dots + 2^{i_{p+1}}} &= 2^{\frac{\ell}{2}} f(\mathcal{H}) - \left[ w_0 + \sum_{q=1}^p \sum_{1 \leq \lambda_1 < \dots < \lambda_q < p+1} w_{2^{i_{\lambda_1}} + \dots + 2^{i_{\lambda_q}}} \right] \\ &= 2^{\frac{\ell}{2}} f(\mathcal{H}) - \left[ w_0 + \sum_{q=1}^p \binom{p+1}{q} w_{2^{i_1} + \dots + 2^{i_q}} \right] \\ &= 2^{\frac{\ell}{2}} 2^{-\ell+(p+1)} \sum_{u=0}^{\ell} \binom{\ell-(p+1)}{u} h(u) - 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \binom{\ell}{u} h(u) \\ &\quad - 2^{-\frac{\ell}{2}} \sum_{q=1}^p \binom{p+1}{q} \sum_{u=0}^{\ell} \left[ \sum_{j=0}^q (-1)^j \binom{q}{j} \binom{\ell-q}{u-j} \right] h(u) \\ &= 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} 2^{p+1} \binom{\ell-(p+1)}{u} h(u) \\ &\quad - 2^{-\frac{\ell}{2}} \sum_{q=0}^p \binom{p+1}{q} \sum_{u=0}^{\ell} \left[ \sum_{j=0}^q (-1)^j \binom{q}{j} \binom{\ell-q}{u-j} \right] h(u) \\ &= 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \alpha(u) h(u), \end{aligned}$$

donde

$$\alpha(u) = 2^{p+1} \binom{\ell-(p+1)}{u} - \sum_{q=0}^p \binom{p+1}{q} \sum_{j=0}^q (-1)^j \binom{q}{j} \binom{\ell-q}{u-j}.$$

Puesto que

$$\binom{\ell-q}{u-j} = \sum_{i=0}^{p+1-q} \binom{p+1-q}{i} \binom{\ell-(p+1)}{u-j-i},$$

se tiene:

$$\begin{aligned}
 \alpha(u) &= 2^{p+1} \binom{\ell - (p+1)}{u} \\
 &\quad - \sum_{q=0}^p \binom{p+1}{q} \left[ \sum_{j=0}^q (-1)^j \binom{q}{j} \sum_{i=0}^{p+1-q} \binom{p+1-q}{i} \binom{\ell - (p+1)}{u-j-i} \right] \\
 &= 2^{p+1} \binom{\ell - (p+1)}{u} \\
 &\quad - \sum_{q=0}^p \binom{p+1}{q} \left[ \sum_{j=0}^q (-1)^j \binom{q}{j} \sum_{r=j}^{j+p+1-q} \binom{p+1-q}{r-j} \binom{\ell - (p+1)}{u-r} \right] \\
 &= \sum_{r=0}^{p+1} \lambda(u, r) \binom{\ell - (p+1)}{u-r},
 \end{aligned}$$

donde,

i) si  $r = 0$ ,

$$\lambda(u, 0) = 2^{p+1} - \sum_{q=0}^p \binom{p+1}{q} = \binom{p+1}{p+1} = 1,$$

ii) si  $r > 0$ , el lema B.3.2 permite afirmar que

$$\begin{aligned}
 \lambda(u, r) &= - \sum_{q=0}^p \binom{p+1}{q} \left[ \sum_{j=0}^q (-1)^j \binom{q}{j} \binom{p+1-q}{r-j} \right] \\
 &= (-1)^r \binom{p+1}{r}.
 \end{aligned}$$

Así

$$\alpha(u) = \sum_{r=0}^{p+1} \lambda(u, r) \binom{\ell - (p+1)}{u-r} = \sum_{r=0}^{p+1} (-1)^r \binom{p+1}{r} \binom{\ell - (p+1)}{u-r}$$

y, finalmente,

$$w_{2^{i_1} + \dots + 2^{i_p} + 2^{i_{p+1}}} = 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \alpha(u) h(u) = 2^{-\frac{\ell}{2}} \sum_{u=0}^{\ell} \left[ \sum_{r=0}^{p+1} (-1)^r \binom{p+1}{r} \binom{\ell - (p+1)}{u-r} \right] h(u),$$

como se quería demostrar. □

**Corolario 4.2.2.** Para cualesquiera  $0 \leq i_1 < i_2 < \dots < i_p < \ell$ , se verifica:

$$w_{2^{i_1}+2^{i_2}+\dots+2^{i_p}} = w_{2^0+2^1+\dots+2^{p-1}} = w_{2^p-1}.$$

**Corolario 4.2.3.** La  $k$ -epistasis normalizada de una función de ponderación  $f : \Omega_\ell \rightarrow \mathbb{R}$  está dada por:

$$\varepsilon_{\ell,k}^*(f) = \frac{\sum_{i=k+1}^{\ell} \binom{\ell}{i} w_{2^i-1}^2}{\sum_{i=0}^{\ell} \binom{\ell}{i} w_{2^i-1}^2}.$$

*Demostración.* Es consecuencia directa de la expresión (4.7) de la página 130 junto con el corolario 4.2.2 y el hecho de que  $\|\mathbf{w}\|^2 = \sum_{i=0}^{\ell} \binom{\ell}{i} w_{2^i-1}^2$ .

En efecto:

$$\varepsilon_{\ell,k}^*(f) = 1 - \frac{\sum_{0 \leq i_1 < \dots < i_k < \ell} w_{2^{i_1}+2^{i_2}+\dots+2^{i_k}}^2}{\sum_{0 \leq j < 2^\ell} w_j^2} = 1 - \frac{\sum_{i=0}^k \binom{\ell}{i} w_{2^i-1}^2}{\sum_{i=0}^{\ell} \binom{\ell}{i} w_{2^i-1}^2} = \frac{\sum_{i=k+1}^{\ell} \binom{\ell}{i} w_{2^i-1}^2}{\sum_{i=0}^{\ell} \binom{\ell}{i} w_{2^i-1}^2}.$$

□

#### 4.2.5 Funciones Royal Road generalizadas

El cálculo de la  $k$ -epistasis normalizada de las funciones Funciones Royal Road generalizadas resulta ahora sencillo, si tenemos en cuenta que en la sección 3.2.4 se calcularon todos los coeficientes de Walsh. Además, por el corolario 3.2.2 de la página 100 se verifica:

$$\sum_{0 \leq i_1 < \dots < i_j < 2^n} w_{2^{i_1}+\dots+2^{i_j}}^2 = \binom{2^m}{j} 2^{n-m} \omega^2,$$

donde  $\omega = 2^{m+2^{n-1}-2^m}$  (ver página 96). Así:

$$\begin{aligned} w_0^2 + \sum_{j=1}^k \left[ \sum_{0 \leq i_1 < \dots < i_j < 2^n} w_{2^{i_1}+\dots+2^{i_j}}^2 \right] &= 2^{2n-2m} w^2 + \sum_{j=1}^k \binom{2^m}{j} 2^{n-m} \omega^2 \\ &= 2^{n-m} \omega^2 \left[ 2^{n-m} + \sum_{j=1}^k \binom{2^m}{j} \right], \end{aligned}$$

en donde se ha utilizado el valor de  $w_0$  calculado en (3.7).

Finalmente, teniendo en cuenta el valor de  $\|\mathfrak{R}_m^n\|^2$  dado en la proposición 3.2.3, nosotros tenemos:

$$\varepsilon_{2^n, k}^*(\mathfrak{R}_m^n) = 1 - \frac{2^{n-m} + \sum_{j=1}^k \binom{2^m}{j}}{2^{n-m} + 2^{2^m} - 1} = \frac{2^{2^m} - \sum_{j=0}^k \binom{2^m}{j}}{2^{n-m} + 2^{2^m} - 1} = \frac{\sum_{j=k+1}^{2^m} \binom{2^m}{j}}{2^{n-m} + 2^{2^m} - 1}.$$

### 4.2.6 Funciones Template

Para la obtención de la *k*-epistasis normalizada de las funciones Template se procederá a calcular sus coeficientes de Walsh. Dado que los coeficientes de órdenes 0, 1 y 2 ya se han calculado (secciones 2.3.2 y 3.2.5), en esta apartado vamos a calcular los restantes.

Como ya se ha comentado en la sección 2.4, la función Template  $T_\ell^n : \Omega_\ell \rightarrow \mathbb{R}$  es de orden *n* y se puede obtener como suma de las funciones simples  $\tau_k$  de orden *n*, siendo:

$$\tau_k : \Omega_n \rightarrow \mathbb{R}, \quad \tau_k(s_k, \dots, s_{k+n-1}) = \begin{cases} 1 & \text{si } s_k = \dots = s_{k+n-1} = 1, \\ 0 & \text{en otro caso.} \end{cases}$$

Para calcular los coeficientes de Walsh asociados a  $T_\ell^n$  se obtendrán, en primer lugar, los asociados a las funciones  $\tau_k$ . Manteniendo las notaciones de la subsección 3.2.5, denotamos por  $w_{(\tau_k)2^{i_1+\dots+i_p}}$  a los citados coeficientes y por  $\Delta_k$  al conjunto de índices  $\{k, \dots, k+n-1\}$  ( $0 \leq k \leq \ell-n$ ). Se verifica:

**Proposición 4.2.4.** *Para  $1 \leq k \leq \ell$ , los coeficientes de Walsh de orden *k*, asociados a  $\tau_k$  toman el valor:*

$$w_{(\tau_k)2^{i_1+\dots+i_p}} = \begin{cases} (-1)^p 2^{\frac{\ell-2n}{2}} & \text{si } \{i_1, \dots, i_p\} \subset \Delta_k, \\ 0 & \text{en otro caso.} \end{cases}$$

*Demostración.* Se realizará por inducción en los valores de *p*.

El caso  $p = 1$  se demostró en la sección 3.2.5. Supongamos ahora el resultado cierto para  $p - 1$  y demostrémoslo para *p*. El coeficiente  $w_{(\tau_k)2^{i_1+\dots+i_p}}$  de orden *p* se calculará con la relación

$$w_{(\tau_k)2^{i_1+\dots+i_p}} = 2^{\frac{\ell}{2}} \tau_k(\mathcal{H}_{i_1 \dots i_p}) - \left( \sum_{q=1}^{p-1} \sum_{0 \leq \lambda_1 < \dots < \lambda_q \leq p} w_{(\tau_k)2^{i_{\lambda_1} + \dots + i_{\lambda_q}}} \right) - w_{(\tau_k)0},$$

siendo  $\mathcal{H}_{i_1 \dots i_p} = \# \dots \# \overset{i_p}{\downarrow} 0 \# \dots \# \overset{i_1}{\downarrow} 0 \# \dots \#$ .

Si  $\{i_1, \dots, i_p\} \subset \Delta_k$  entonces  $\tau_k(\mathcal{H}_{i_1 \dots i_p}) = 0$  y por la hipótesis de inducción el valor de  $w_{(\tau_k)2^{i_{\lambda_1} + \dots + 2^{i_{\lambda_q}}}}$  es  $(-1)^q 2^{\frac{\ell-2n}{2}}$ . Así,

$$\begin{aligned} w_{(\tau_k)2^{i_1 + \dots + 2^{i_p}}} &= 0 - \left( \sum_{q=1}^{p-1} \sum_{0 \leq \lambda_1 < \dots < \lambda_q \leq p} (-1)^q 2^{\frac{\ell-2n}{2}} \right) - 2^{\frac{\ell-2n}{2}} \\ &= - \left( \sum_{q=1}^{p-1} (-1)^q \binom{p}{q} 2^{\frac{\ell-2n}{2}} \right) - 2^{\frac{\ell-2n}{2}} \\ &= - \left( \sum_{q=0}^{p-1} (-1)^q \binom{p}{q} 2^{\frac{\ell-2n}{2}} \right) \\ &= (-1)^p 2^{\frac{\ell-2n}{2}}. \end{aligned}$$

Si  $\{i_1, \dots, i_p\} \not\subset \Delta_k$  y  $\{i_1, \dots, i_p\} \cap \Delta_k \neq \emptyset$ , entonces  $\tau_k(\mathcal{H}_{i_1 \dots i_p}) = 0$ . Sin ninguna restricción se puede suponer que  $\{i_1, \dots, i_r\} \subset \Delta_k$ ,  $\{i_{r+1}, \dots, i_p\} \cap \Delta_k = \emptyset$ . Por la hipótesis de inducción,  $w_{(\tau_k)2^{i_{\lambda_1} + \dots + 2^{i_{\lambda_q}}}} = (-1)^q 2^{\frac{\ell-2n}{2}}$  cuando  $\{i_{\lambda_1}, \dots, i_{\lambda_q}\} \subset \Delta_k$  y  $w_{2^{i_{\lambda_1} + \dots + 2^{i_{\lambda_q}}}} = 0$  en otro caso. Entonces

$$\begin{aligned} w_{(\tau_k)2^{i_1 + \dots + 2^{i_p}}} &= 0 - \left( \sum_{q=1}^r \sum_{0 \leq \lambda_1 < \dots < \lambda_q \leq r} (-1)^q 2^{\frac{\ell-2n}{2}} \right) - 2^{\frac{\ell-2n}{2}} \\ &= - \left( \sum_{q=1}^r (-1)^q \binom{r}{q} 2^{\frac{\ell-2n}{2}} \right) - 2^{\frac{\ell-2n}{2}} \\ &= -2^{\frac{\ell-2n}{2}} \sum_{q=0}^r (-1)^q \binom{r}{q} \\ &= 0. \end{aligned}$$

Si  $\{i_1, \dots, i_p\} \cap \Delta_k = \emptyset$  entonces  $\tau_k(\mathcal{H}_{i_1 \dots i_p}) = \frac{1}{2^{\ell-p}} 2^{\ell-n-p} = 2^{-n}$  y, por la hipótesis de inducción,  $w_{2^{i_{\lambda_1} + \dots + 2^{i_{\lambda_q}}}} = 0$  para todo  $\{\lambda_1, \dots, \lambda_q\} \subset \{1, \dots, p\}$ , por lo que

$$w_{(\tau_k)2^{i_1 + \dots + 2^{i_p}}} = 2^{\frac{\ell}{2}} 2^{-n} - 0 - 2^{\frac{\ell-2n}{2}} = 0.$$

□

Ahora observemos que, para el cálculo de un coeficiente  $w_{2^{i_1 + \dots + 2^{i_p}}}$  ( $0 \leq i_1 < \dots < i_p < \ell$ ), se debe realizar un recuento de los conjuntos  $\Delta_k$  que contienen a  $\{i_1, i_2, \dots, i_p\}$ . Es evidente que  $\{i_1, i_2, \dots, i_p\} \subset \Delta_k$  si y sólo si  $\{i_1, i_p\} \subset \Delta_k$ . Por lo que los números de conjuntos

Posición de los índices	Número de conjuntos
$i_1 \in \{0, \dots, \ell - n - 1\}, i_p \in \{i_1 + 1, \dots, \ell - n - 1\}$	$i_1 + 1$
$i_1 \in \{0, \dots, \ell - n - 1\}, i_p \in \{\ell - n, \dots, n - 1\}$	$i_1 + 1$
$i_1 \in \{0, \dots, \ell - n - 1\}, i_p \in \{n, \dots, i_1 + n - 1\}$	$n - i_p + i_1$
$i_1 \in \{\ell - n, \dots, n - 1\}, i_p \in \{i_1 + 1, \dots, n - 1\}$	$\ell - n + 1$
$i_1 \in \{\ell - n, \dots, n - 1\}, i_p \in \{n, \dots, \ell - 1\}$	$\ell - i_p$
$i_1 \in \{n, \dots, \ell - p\}, i_p \in \{i_1 + 1, \dots, \ell - 1\}$	$\ell - i_p$

Tabla 4.2: Recuento de los conjuntos  $\Delta_k$  que contienen a  $\{i_1, i_p\}$  si  $\ell \leq 2n$ .

Posición de los índices	Número de conjuntos
$i_1 \in \{0, \dots, n - 1\}, i_p \in \{i_1 + 1, \dots, n - 1\}$	$i_1 + 1$
$i_1 \in \{0, \dots, n - 1\}, i_p \in \{n, \dots, i_1 + n - 1\}$	$n - i_p + i_1$
$i_1 \in \{n, \dots, \ell - n - 1\}, i_p \in \{i_1 + 1, \dots, i_1 + n - 1\}$	$n - i_p + i_1$
$i_1 \in \{\ell - n, \dots, \ell - p\}, i_p \in \{i_1 + 1, \dots, \ell - 1\}$	$\ell - i_p$

Tabla 4.3: Recuento de los conjuntos  $\Delta_k$  que contienen a  $\{i_1, i_p\}$  si  $\ell \geq 2n$ .

buscados coinciden con los de las tablas 3.8 y 3.9 de la subsección 3.2.5 al considerar  $i = i_1$ ,  $j = i_p$  y que nosotros ahora recogemos en las tablas 4.2 y 4.3.

**Proposición 4.2.5.** El valor de la suma  $\Gamma_p = \sum_{0 \leq i_1 < \dots < i_p < \ell} w_{2^{i_1 + \dots + 2^{i_p}}}^2$  es:

1. Si  $\ell \leq 2n$  y  $1 \leq p \leq \ell$  entonces:

$$\Gamma_p = 2^{\ell - 2n} \left[ \binom{2n - \ell}{p} (\ell - n + 1)^2 + \sum_{j=1}^{\ell - n} j^2 \left\{ 2 \binom{n - j}{p - 1} + (\ell - n - j) \binom{n - j - 1}{p - 2} \right\} \right].$$

2. Si  $\ell \geq 2n$  y  $2 \leq p \leq \ell$ , entonces:

$$\Gamma_p = 2^{\ell-2n} \sum_{j=1}^{n-1} j^2 \left\{ 2 \binom{n-j}{p-1} + (\ell-n-j) \binom{n-j-1}{p-2} \right\}.$$

*Demostración.* Para  $\ell \leq 2n$  y  $p = 1$  los valores que proporciona el enunciado son los ya obtenidos en la proposición 2.3.4. Consideremos ahora  $p \geq 2$ . La demostración se basa en un recuento de los coeficientes de Walsh cuyos valores coinciden.

Los valores de los coeficientes de Walsh de orden  $p$  de  $T_\ell^n$  son:

$$w_{2^{i_1+\dots+2^{i_p}}} = \sum_{k=0}^{\ell-n} w_{(\tau_k)2^{i_1+\dots+2^{i_p}}} = \sum_{\substack{k=0 \\ k; \{i_1, i_p\} \subset \Delta_k}}^{\ell-n} (-1)^p 2^{\frac{\ell-2n}{2}}.$$

Haciendo uso de la tabla 4.2 se observa que los posibles valores de los coeficientes de Walsh de orden  $p$ , ( $p \geq 2$ ) son (salvo el signo):  $j \cdot 2^{\frac{\ell-2n}{2}}$ ,  $j \in \{1, \dots, \ell-n+1\}$ . Si  $j \in \{1, \dots, \ell-n\}$  el valor se alcanza cuando  $i_1 = j-1$ ,  $i_p - i_1 = n-j$  o  $i_p = \ell-j$ , mientras que para  $j = \ell-n+1$  se obtiene cuando  $\{i_1, i_p\} \subset \{\ell-n, \dots, n-1\}$ . Analizaremos por separado cada una de estas situaciones.

- (i) Si  $i_1 = j-1$ , entonces  $i_p \in \{j, \dots, n-1\}$ . Así, para obtener un coeficiente con valor absoluto  $j \cdot 2^{\frac{\ell-2n}{2}}$  bastará con situar los  $p-1$  índices  $i_2, \dots, i_p$  en  $\{j, \dots, n-1\}$ . De esta manera, existen  $\binom{n-j}{p-1}$  coeficientes de orden  $p$  con valor  $j \cdot 2^{\frac{\ell-2n}{2}}$ .
- (ii) Si  $i_p - i_1 = n-j$ , existen  $n-j-1$  posiciones intermedias entre las cuales, para obtener un coeficiente de orden  $p$ , se deben elegir  $p-2$  de ellas en las que situar  $i_2, \dots, i_{p-1}$ . Así, se obtienen  $\binom{n-j-1}{p-2}$  coeficientes de orden  $p$  con valor absoluto  $j \cdot 2^{\frac{\ell-2n}{2}}$ .
- (iii) Si  $i_p = \ell-j$ , estamos ante una situación análoga al caso (i).
- (iv) Si  $j = \ell-n+1$ , el valor de un coeficiente de Walsh es  $(\ell-n+1)2^{\frac{\ell-2n}{2}}$  si y sólo si  $\{i_1, i_p\} \subset \{\ell-n, \dots, n-1\}$ , por lo que el número de coeficientes de orden  $p$  con valor absoluto  $(\ell-n+1)2^{\frac{\ell-2n}{2}}$  es  $\binom{2n-\ell}{p}$ .

La demostración del caso  $\ell \geq 2n$  es similar al caso  $\ell \leq 2n$ , teniendo en cuenta que ahora los posibles valores (salvo el signo) de los coeficientes de Walsh de orden  $p$  ( $p \geq 2$ ) son:  $j \cdot 2^{\frac{\ell-2n}{2}}$ ,  $j \in \{1, \dots, n-1\}$  y, por ello, ningún coeficiente de orden mayor que 1 coincide con  $w_0 = 2^{\frac{\ell-2n}{2}}(\ell-n+1)$ .

□

Ahora, estamos en condiciones de calcular la norma del vector  $\mathbf{T}_\ell^n$ . A pesar de que este valor ha sido calculado ya en [20], el método es diferente al que seguimos nosotros, el uso de todos los coeficientes de Walsh asociados a  $T_\ell^n$ . Por esta razón, hemos decidido incluirlo en esta memoria.

**Proposición 4.2.6.** *El valor de  $\|\mathbf{T}_\ell^n\|^2$  es:*

$$\|\mathbf{T}_\ell^n\|^2 = \begin{cases} 2^{\ell-n}(3(\ell-n)-1)+2 & \text{si } \ell \leq 2n, \\ 2^{\ell-n}(3(\ell-n)-1)+2^{\ell-2n}(2+(\ell-2n)(\ell-2n-1)) & \text{si } \ell \geq 2n. \end{cases}$$

*Demostración.* Si  $\ell \leq 2n$

$$\begin{aligned} \|\mathbf{T}_\ell^n\|^2 &= w_0^2 + \sum_{k=1}^{\ell} \sum_{0 \leq \lambda_1 < \dots < \lambda_k < \ell} w_{2^{i\lambda_1 + \dots + i\lambda_k}}^2 = w_0^2 + \sum_{k=1}^{\ell} \Gamma_k \\ &= 2^{\ell-2n}(\ell-n+1)^2 + 2^{\ell-2n} \sum_{k=1}^{\ell} \binom{2n-\ell}{k} (\ell-n+1)^2 \\ &\quad + 2^{\ell-2n} \sum_{k=1}^{\ell} \sum_{j=1}^{\ell-n} j^2 \left\{ 2 \binom{n-j}{k-1} + (\ell-n-j) \binom{n-j-1}{k-2} \right\} \\ &= 2^{\ell-2n} \sum_{k=0}^{2n-\ell} \binom{2n-\ell}{k} (\ell-n+1)^2 \\ &\quad + 2^{\ell-2n} \sum_{k=1}^n \sum_{j=1}^{\ell-n} \left[ 2 \binom{n-j}{k-1} + (\ell-n-j) \binom{n-j-1}{k-2} \right] j^2 \\ &= (\ell-n+1)^2 + 2^{\ell-2n} \sum_{j=1}^{\ell-n} \sum_{k=1}^n \left[ 2 \binom{n-j}{k-1} + (\ell-n-j) \binom{n-j-1}{k-2} \right] j^2 \\ &= (\ell-n+1)^2 + 2^{\ell-2n} \sum_{j=1}^{\ell-n} [2 \cdot 2^{n-j} + (\ell-n-j) \cdot 2^{n-j-1}] j^2 \\ &= (\ell-n+1)^2 + 2^{\ell+1-n} \sum_{j=1}^{\ell-n} 2^{-j} j^2 + \sum_{j=1}^{\ell-n} (\ell-n-j) j^2 2^{\ell-n-j-1} \end{aligned}$$

Ahora, ayudándonos de los apartados a) y c) del lema B.3.6, se tiene

$$\|\mathbf{T}_\ell^n\|^2 = (\ell-n+1)^2 + 12 \cdot 2^{\ell-n} - 2(\ell^2 - 2\ell n + n^2 + 4\ell - 4n + 6)$$

$$\begin{aligned}
& +2^{\ell-n}(3\ell - 3n - 13) + \ell^2 - 2\ell n + n^2 + 6\ell - 6n + 13 \\
& = (\ell - n + 1)^2 + 2 - (\ell - n + 1)^2 + 2^{\ell-n}(3(\ell - n) - 1) \\
& = 2 + 2^{\ell-n}(3(\ell - n) - 1).
\end{aligned}$$

Si  $\ell \geq 2n$ , de nuevo el lema B.3.6, junto con el hecho de que  $\sum_{j=1}^{n-1} j^2 = \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n$ , proporciona:

$$\begin{aligned}
\|\mathbf{T}_\ell^n\|^2 & = w_0^2 + \sum_{k=1}^{\ell} \sum_{0 \leq \lambda_1 < \dots < \lambda_k < \ell} w_{2^{i_{\lambda_1} + \dots + 2^{i_{\lambda_k}}}}^2 \\
& = 2^{\ell-2n} \left[ (\ell - n + 1)^2 + \frac{2}{3} \binom{n+1}{2} (2n+1) + n^2(\ell - 2n) \right. \\
& \quad \left. + \sum_{k=2}^{\ell} \sum_{j=1}^{n-1} \left\{ 2 \binom{n-j}{k-1} + (\ell - n - j) \binom{n-j-1}{k-2} \right\} j^2 \right] \\
& = 2^{\ell-2n} \left[ (\ell - n + 1)^2 + \frac{2}{3} \binom{n+1}{2} (2n+1) + n^2(\ell - 2n) \right. \\
& \quad \left. + \sum_{j=1}^{n-1} \sum_{k=2}^{\ell} j^2 \left\{ 2 \binom{n-j}{k-1} + (\ell - n - j) \binom{n-j-1}{k-2} \right\} \right] \\
& = 2^{\ell-2n} \left[ (\ell - n + 1)^2 + \frac{2}{3} \binom{n+1}{2} (2n+1) + n^2(\ell - 2n) \right. \\
& \quad \left. + \sum_{j=1}^{n-1} \left\{ 2 \sum_{k=2}^{\ell} \binom{n-j}{k-1} j^2 \right\} + \sum_{j=1}^{n-1} \left\{ \sum_{k=2}^{\ell} (\ell - n - j) \binom{n-j-1}{k-2} j^2 \right\} \right] \\
& = 2^{\ell-2n} \left[ (\ell - n + 1)^2 + \frac{2}{3} \binom{n+1}{2} (2n+1) + n^2(\ell - 2n) \right. \\
& \quad \left. + 2 \sum_{j=1}^{n-1} (2^{n-j} - 1) j^2 + \sum_{j=1}^{n-1} (\ell - n - j) 2^{n-j-1} j^2 \right] \\
& = 2^{\ell-2n} \left[ (\ell - n + 1)^2 + \frac{2}{3} \binom{n+1}{2} (2n+1) + n^2(\ell - 2n) \right] \\
& \quad + 2^{\ell-n+1} \sum_{j=1}^{n-1} 2^{-j} j^2 - 2^{\ell-2n+1} \sum_{j=1}^{n-1} j^2
\end{aligned}$$

$$\begin{aligned}
& +2^{\ell-n-1}(\ell-n) \sum_{j=1}^{n-1} 2^{-j} j^2 - 2^{\ell-n-1} \sum_{j=1}^{n-1} 2^{-j} j^3 \\
= & 2^{\ell-2n} \left( \ell^2 - 2\ell n + 2\ell + 2n^2 - \frac{5}{3}n + 1 + n^2\ell - \frac{4}{3}n^3 \right) \\
& + 12 \cdot 2^{\ell-n} - 2^{\ell-2n}(4n^2 + 8n + 12) - 2^{\ell-2n} \left( \frac{2}{3}n^3 - n^2 + \frac{1}{3}n \right) \\
& + 2^{\ell-n} 3(\ell-n) - 2^{\ell-2n}(\ell n^2 + 2\ell n + 3\ell - n^3 - 2n^2 - 3n) \\
& - 13 \cdot 2^{\ell-n} + 2^{\ell-2n}(n^3 + 3n^2 + 9n + 13) \\
= & 2^{\ell-n} (3(\ell-n) - 1) + 2^{\ell-2n} (2 + (\ell-2n)(\ell-2n-1)).
\end{aligned}$$

□

**Corolario 4.2.7.** *El valor de la  $k$ -epistasia de la función Template  $T_\ell^n$  es:*

1. si  $\ell \leq 2n$  y  $1 \leq k \leq \ell$ ,

$$\varepsilon_{\ell,k}^*(T_\ell^n) = 1 - \frac{2^{\ell-2n} \left[ \binom{2n-\ell}{k} (\ell-n+1)^2 + \sum_{j=1}^{\ell-n} j^2 \left\{ 2 \binom{n-j}{k-1} + (\ell-n-j) \binom{n-j-1}{k-2} \right\} \right]}{2^{\ell-n}(3(\ell-n)-1) + 2},$$

2. si  $\ell \geq 2n$  y  $2 \leq k \leq \ell$ ,

$$\varepsilon_{\ell,k}^*(T_\ell^n) = 1 - \frac{\sum_{j=1}^{n-1} j^2 \left[ 2 \binom{n-j}{k-1} + (\ell-n-j) \binom{n-j-1}{k-2} \right]}{2^n(3(\ell-n)-1) + 2 + (\ell-2n)(\ell-2n-1)}.$$

### 4.3 Valores extremos de la $k$ -epistasia

En esta sección caracterizaremos a las funciones con  $k$ -epistasia mínima (nula). Veremos que son precisamente las funciones de orden  $k$ . Asimismo, calculamos el valor máximo de la  $k$ -epistasia y discutimos cuándo se alcanza ese valor.

El estudio realizado en la primera sección de este capítulo, muestra claramente como la denominada  $k$ -epistasia generaliza a la epistasia normalizada definida por Van Hove en

[42] y a nuestra 2-epistasis definida en el tercer capítulo de esta memoria. En concreto, la epistasis normalizada de Van Hove (denotada en el capítulo 2 como  $\varepsilon^*$ ) corresponde a la  $k$ -epistasis normalizada con  $k = 1$ . Al caso (degenerado) de la 0-epistasis se dedica la siguiente subsección.

### 4.3.1 Valores extremos de la 0-epistasis

Como la expresión de  $\varepsilon_{\ell,0}^*(f)$  en términos de los coeficientes de Walsh se reduce a:

$$\varepsilon_{\ell,0}^*(f) = 1 - \frac{w_0^2}{\|\mathbf{w}\|^2},$$

entonces, el valor mínimo se alcanza para aquellas funciones  $f$  tales que  $\mathbf{w} = \mathbf{W}_\ell \mathbf{f}$  verifica que  $w_j = 0$ , para  $j \neq 0$ . Por ello:

$$\mathbf{f} = \mathbf{W}_\ell \mathbf{w} = \mathbf{W}_\ell \begin{pmatrix} a \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} a \\ a \\ \vdots \\ a \end{pmatrix},$$

esto es, el valor mínimo se alcanza en las funciones constantes (*funciones de orden cero*).

Por otra parte,

**Proposición 4.3.1.** *Para una función de ajuste  $f : \Omega_\ell \rightarrow \mathbb{R}$  que toma valores no negativos, se verifica:*

$$\varepsilon_{\ell,0}^*(f) \leq 1 - \frac{1}{2^\ell}.$$

*Demostración.*

$$\varepsilon_{\ell,0}^*(f) = 1 - \frac{1}{2^\ell} \frac{{}^t \mathbf{f} \mathbf{G}_{\ell,0} \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}} = 1 - \frac{1}{2^\ell} \frac{{}^t \mathbf{f} \mathbf{U}_\ell \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}} = 1 - \frac{1}{2^\ell} \frac{\left( \sum_{0 \leq i < 2^\ell} f_i \right)^2}{\sum_{0 \leq i < 2^\ell} f_i^2}. \quad (4.10)$$

Como  $\left( \sum_{0 \leq i < 2^\ell} f_i \right)^2 \geq \sum_{0 \leq i < 2^\ell} f_i^2$  entonces  $\frac{\left( \sum_{0 \leq i < 2^\ell} f_i \right)^2}{\sum_{0 \leq i < 2^\ell} f_i^2} \geq 1$ , de lo se deduce el resultado.  $\square$

**Proposición 4.3.2.** Para  $f : \Omega_\ell \rightarrow \mathbb{R}^+ \cup \{0\}$ , con  $\|\mathbf{f}\| = 1$ , las siguientes condiciones son equivalentes:

1.  $\varepsilon_{\ell,0}^*(f) = 1 - \frac{1}{2^\ell}$ ,
2.  $f$  es una función de Dirac.

*Demostración.* Supongamos que  $\varepsilon_{\ell,0}^*(f) = 1 - \frac{1}{2^\ell}$ , entonces por (4.10)

$$\varepsilon_{\ell,0}^*(f) = 1 - \frac{1}{2^\ell} \frac{\left( \sum_{0 \leq i < 2^\ell} f_i \right)^2}{\sum_{0 \leq i < 2^\ell} f_i^2} = 1 - \frac{1}{2^\ell}$$

equivale a que  $\left( \sum_{0 \leq i < 2^\ell} f_i \right)^2 = \sum_{0 \leq i < 2^\ell} f_i^2$ , es decir  $\sum_{0 \leq i, j < 2^\ell} f_i f_j = 0$ . Ahora bien, como la función es no negativa, la condición  $\sum_{0 \leq i, j < 2^\ell} f_i f_j = 0$  equivale a que  $f_i f_j = 0$  ( $i \neq j$ ). Además, como  $f$  es no nula, debe existir al menos un índice  $i$  para el que  $f_i \neq 0$  y, en consecuencia,  $f_j = 0$ , para todo  $i \neq j$ . Finalmente, al ser  $\mathbf{f}$  un vector unitario y  $f_k \geq 0$ , para todo  $k$ , se tiene que  $f_i = 1$  y, en consecuencia,  $f = \delta_i$ .

Veamos que el recíproco también es cierto. Si  $f$  es función de Dirac, entonces existe  $i \in \{0, \dots, 2^\ell - 1\}$  con  $f_i = 1$ ,  $f_j = 0$  para  $i \neq j$ . Así:

$$\varepsilon_{\ell,0}^*(\delta_i) = 1 - \frac{1}{2^\ell} \frac{{}^t \boldsymbol{\delta}_i \mathbf{U}_\ell \boldsymbol{\delta}_i}{{}^t \boldsymbol{\delta}_i \boldsymbol{\delta}_i} = 1 - \frac{1}{2^\ell} u_{ii} = 1 - \frac{1}{2^\ell}.$$

□

### 4.3.2 Valor mínimo

Tras haber comprobado que la 0-epistasia, respectivamente la 1-epistasia (epistasia clásica) y la 2-epistasia, se anulan para funciones de orden cero, respectivamente funciones de orden uno (ver subsección 2.2.2) y de orden dos (ver subsección 3.4.1), a continuación completamos este estudio para la  $k$ -epistasia con  $k$  de orden superior. Para ello basta tener en cuenta la expresión (4.7) de la página 130 que proporciona el valor de la  $k$ -epistasia normalizada en términos de los coeficientes de Walsh. Esta expresión permite asegurar que:

$$\varepsilon_{\ell,k}^*(f) = 0 \Leftrightarrow w_0^2 + \sum_{0 \leq i < \ell} w_{2^i}^2 + \dots + \sum_{0 \leq i_1 < \dots < i_k < \ell} w_{2^{i_1 + \dots + i_k}}^2 = \sum_{0 \leq j < 2^\ell} w_j^2.$$

Es decir,

$$w_j = 0, \text{ para } j \notin \{0, 2^{i_1 + \dots + i_p}, 1 \leq p \leq k, 0 \leq i_1 < \dots < i_k < \ell\}.$$

En consecuencia, se tiene demostrado el siguiente resultado:

**Teorema 4.3.3.** *Para una función de ajuste  $f : \Omega_\ell \rightarrow \mathbb{R}$  las siguientes condiciones son equivalentes:*

1.  $\varepsilon_{\ell,k}^*(f) = 0$ ,
2.  $f$  es una función de orden  $k$ .

*Demostración.* Es consecuencia directa del teorema 2.6.7 de la página 72. □

### 4.3.3 Valor máximo

Una vez estudiados los casos  $k = 0, 1, 2$ , nos centramos en la búsqueda del valor máximo para la  $k$ -epistasis, con  $k \geq 3$ .

Ya se ha comentado que el valor máximo de la  $k$ -epistasis normalizada es 1. Sin embargo, si  $\ell \geq k$  y nos restringimos a funciones no negativas con vector asociado unitario, se puede asegurar que el máximo valor teórico no se alcanza.

Observemos, en primer lugar, que la relación recurrente (4.1), con la se construyeron las matrices  $\mathbf{G}_{\ell,k}$ , permite comprobar que las formas  $\gamma_{\ell,k}(f) = {}^t \mathbf{f} \mathbf{G}_{\ell,k} \mathbf{f}$  verifican:

$$\begin{aligned} \gamma_{\ell,k}(f) &= \begin{pmatrix} {}^t \mathbf{f}^0 & {}^t \mathbf{f}^1 \end{pmatrix} \begin{pmatrix} \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} \\ \mathbf{G}_{\ell-1,k} - \mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k} + \mathbf{G}_{\ell-1,k-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}^0 \\ \mathbf{f}^1 \end{pmatrix} \\ &= \begin{pmatrix} {}^t \mathbf{f}^0 & {}^t \mathbf{f}^1 \end{pmatrix} \begin{pmatrix} \mathbf{G}_{\ell-1,k} & \mathbf{G}_{\ell-1,k} \\ \mathbf{G}_{\ell-1,k} & \mathbf{G}_{\ell-1,k} \end{pmatrix} \begin{pmatrix} \mathbf{f}^0 \\ \mathbf{f}^1 \end{pmatrix} \\ &\quad + \begin{pmatrix} {}^t \mathbf{f}^0 & {}^t \mathbf{f}^1 \end{pmatrix} \begin{pmatrix} \mathbf{G}_{\ell-1,k-1} & -\mathbf{G}_{\ell-1,k-1} \\ -\mathbf{G}_{\ell-1,k-1} & \mathbf{G}_{\ell-1,k-1} \end{pmatrix} \begin{pmatrix} \mathbf{f}^0 \\ \mathbf{f}^1 \end{pmatrix} \\ &= {}^t (\mathbf{f}^0 + \mathbf{f}^1) \mathbf{G}_{\ell-1,k} (\mathbf{f}^0 + \mathbf{f}^1) + {}^t (\mathbf{f}^0 - \mathbf{f}^1) \mathbf{G}_{\ell-1,k-1} (\mathbf{f}^0 - \mathbf{f}^1) \\ &= \gamma_{\ell-1,k}(g^+) + \gamma_{\ell-1,k-1}(g^-), \end{aligned} \tag{4.11}$$

donde  $g^+$  y  $g^-$  son las funciones sobre  $\Omega_{\ell-1}$  definidas en (3.16), i.e., aquellas cuyos respectivos vectores asociados son  $\mathbf{g}^+ = \mathbf{f}^0 + \mathbf{f}^1$  y  $\mathbf{g}^- = \mathbf{f}^0 - \mathbf{f}^1$ , con  $\mathbf{f}^0$  y  $\mathbf{f}^1$  dados por la expresión (3.15) de la página 112.

**Proposición 4.3.4.** Para una función de ajuste  $f : \Omega_\ell \rightarrow \mathbb{R}$  que tome valores no negativos y con  $\|\mathbf{f}\| = 1$  se verifica, para  $\ell \geq k$ :

$$\varepsilon_{\ell,k}^*(f) \leq 1 - \frac{1}{2^{\ell-k}}.$$

*Demostración.* Comprobar que la afirmación del enunciado es cierta equivale a comprobar que  $\gamma_{\ell,k}(f) = {}^t \mathbf{f} \mathbf{G}_{\ell,k} \mathbf{f} \geq 2^k$ , para  $\ell \geq k$ . Observemos en primer lugar que, si  $\ell = k$ , entonces

$$\gamma_{k,k}(f) = {}^t \mathbf{f} 2^k \mathbf{I}_{2^k} \mathbf{f} = 2^k \|\mathbf{f}\|^2 = 2^k.$$

Supongamos ahora que, para algún entero positivo  $\ell > k$ , y para alguna función  $f$  en las condiciones del enunciado, se verifica  $\gamma_{\ell,k}(f) < 2^k$ . Manteniendo las notaciones de la subsección 3.4.2 consideramos la función  $g = \frac{g^+}{\|\mathbf{g}^+\|}$  cuyo vector asociado  $\mathbf{g}$  es unitario.

Entonces:

$$\begin{aligned} \gamma_{\ell-1,k}(g) &= \gamma_{\ell-1,k} \left( \frac{g^+}{\|\mathbf{g}^+\|} \right) = \frac{1}{\|\mathbf{g}^+\|^2} \gamma_{\ell-1,k}(g^+) \\ &\leq \gamma_{\ell-1,k}(g^+) \leq \gamma_{\ell-1,k}(g^+) + \gamma_{\ell-1,k-1}(g^-) \\ &= \gamma_{\ell,k}(f) < 2^k, \end{aligned}$$

ya que  $\mathbf{G}_{\ell-1,k-1}$  es una matriz semidefinida positiva.

Iterando este procedimiento se llegaría a que existen funciones de ajuste no negativas  $f$  definidas sobre  $\Omega = \{0, 1\}^k$  con  $\|\mathbf{f}\| = 1$  y  $\gamma_{k,k}(f) < 2^k$ , lo que es imposible.

Esta contradicción finaliza la demostración.  $\square$

A continuación, abordamos el estudio de las funciones no negativas con vector asociado unitario y cuya  $k$ -epistasis normalizada es máxima, o equivalentemente, aquellas funciones  $f$  para las que  $\gamma_{\ell,k}(f) = 2^k$ , siendo  $\|\mathbf{f}\| = 1$ . Veremos que, para  $\ell > k + 1$ , la cota establecida en la proposición 4.3.4 no se alcanza.

Consideremos los vectores  $\mathbf{g}^+ = \mathbf{f}^0 + \mathbf{f}^1$  y  $\mathbf{g}^- = \mathbf{f}^0 - \mathbf{f}^1$ , así como

$$\mathbf{w}^+ = \mathbf{W}_{\ell-1} \mathbf{g}^+ = {}^t (w_0^+, w_1^+, \dots, w_{2^{\ell-1}-1}^+)$$

y

$$\mathbf{w}^- = \mathbf{W}_{\ell-1} \mathbf{g}^- = {}^t (w_0^-, w_1^-, \dots, w_{2^{\ell-1}-1}^-).$$

Entonces, la relación entre  $\mathbf{G}_\ell$  y  $\mathbf{E}_\ell$ , la idempotencia de  $\mathbf{W}_\ell$ , junto con la proposición 4.1.7 (página 128) permiten escribir

$$\begin{aligned}
\gamma_{\ell-1,k}(g^+) &= {}^t \mathbf{g}^+ \mathbf{G}_{\ell-1,k} \mathbf{g}^+ = 2^{\ell-1} {}^t \mathbf{g}^+ \mathbf{E}_{\ell-1,k} \mathbf{g}^+ \\
&= 2^{\ell-1} {}^t (\mathbf{W}_{\ell-1} \mathbf{w}^+) \mathbf{E}_{\ell-1,k} \mathbf{W}_{\ell-1} \mathbf{w}^+ = 2^{\ell-1} {}^t \mathbf{w}^+ \mathbf{D}_{\ell-1,k} \mathbf{w}^+ \\
&= 2^{\ell-1} \left\{ (w_0^+)^2 + \sum_{0 \leq i < \ell-1} (w_{2^i}^+)^2 + \sum_{0 \leq i < j < \ell-1} (w_{2^i+2^j}^+)^2 + \dots \right. \\
&\quad \left. \dots + \sum_{0 \leq i_1 < \dots < i_k < \ell-1} (w_{2^{i_1} + \dots + 2^{i_k}}^+)^2 \right\} \quad (4.12a)
\end{aligned}$$

y, análogamente,

$$\begin{aligned}
\gamma_{\ell-1,k-1}(g^-) &= 2^{\ell-1} {}^t \mathbf{w}^- \mathbf{D}_{\ell-1,k-1} \mathbf{w}^- \\
&= 2^{\ell-1} \left\{ (w_0^-)^2 + \sum_{0 \leq i < \ell-1} (w_{2^i}^-)^2 + \sum_{0 \leq i < j < \ell-1} (w_{2^i+2^j}^-)^2 + \dots \right. \\
&\quad \left. \dots + \sum_{0 \leq i_1 < \dots < i_{k-1} < \ell-1} (w_{2^{i_1} + \dots + 2^{i_{k-1}}}^-)^2 \right\}. \quad (4.12b)
\end{aligned}$$

Ahora, teniendo en cuenta que, por (4.11),

$$\gamma_{\ell,k}(f) = \gamma_{\ell-1,k}(g^+) + \gamma_{\ell-1,k-1}(g^-) \geq \gamma_{\ell-1,k}(g^+)$$

(ya que  $\gamma_{\ell-1,k-1}(g^-) \geq 0$ , por ser  $\mathbf{G}_{\ell-1,k-1}$  semidefinida positiva) y que, por la proposición 4.3.4,

$$\gamma_{\ell-1,k}(g^+) = \|\mathbf{g}^+\|^2 \gamma_{\ell-1,k} \left( \frac{\mathbf{g}^+}{\|\mathbf{g}^+\|} \right) \geq 2^k \|\mathbf{g}^+\|^2,$$

se deduce que el valor mínimo de  $\gamma_{\ell,k}(f) = 2^k$  se alcanzará cuando se verifique  $\gamma_{\ell-1,k}(g^+) = 2^k$  y  $\gamma_{\ell-1,k-1}(g^-) = 0$ . Esto sólo puede suceder si  $\|\mathbf{g}^+\| = 1$ . (Nótese que, en general,  $\|\mathbf{g}^+\| \geq \|\mathbf{f}\| = 1$ ).

Nosotros ahora tenemos:

**Proposición 4.3.5.** *Sea  $f$  una función de ajuste no negativa, cuyo vector asociado  $\mathbf{f} \in \mathbb{R}^{2^\ell}$  es unitario y verifica  $\|\mathbf{g}^+\| = \|\mathbf{f}^0 + \mathbf{f}^1\| = 1$ , entonces:*

1. Si  $\ell = k + 1$ ,  $\varepsilon_{\ell,k}^*(f) = 1 - \frac{1}{2^{\ell-k}}$  si y sólo si:

$${}^t \mathbf{f} = 2^{-\frac{k+2}{2}} \left( \left( 1 + (-1)^{u(s)} \right)_{0 \leq s < 2^k}, \left( 1 + (-1)^{u(s)+1} \right)_{0 \leq s < 2^k} \right) \quad (4.13a)$$

$${}^o \quad {}^t \mathbf{f} = 2^{-\frac{k+2}{2}} \left( \left( 1 + (-1)^{u(s)+1} \right)_{0 \leq s < 2^k}, \left( 1 + (-1)^{u(s)} \right)_{0 \leq s < 2^k} \right), \quad (4.13b)$$

donde, como antes,  $u(s)$  denota el número de bits no nulos de la cadena  $s$ .

2. Si  $\ell \geq k + 2$ , entonces  $\varepsilon_{\ell,k}^*(f) < 1 - \frac{1}{2^{\ell-k}}$ .

*Demostración.*

1. Comenzamos la demostración con el caso  $\ell = k + 1$ .

Como es sabido,  $\varepsilon_{\ell,k}^*(f) = 1 - \frac{\gamma_{\ell,k}(f)}{2^\ell \|\mathbf{f}\|^2}$ . Por tanto, si  $\|\mathbf{f}\| = 1$ , para  $\ell = k + 1$  equivale  $\varepsilon_{k+1,k}^*(f) = 1 - \frac{1}{2}$  a que  $\gamma_{k+1,k}(f) = 2^k$ . Como se comentó en la página anterior, para que  $\gamma_{k+1,k}(f) = 2^k$  se necesita, en particular, que  $\gamma_{k,k-1}(g^-) = 0$ ; es decir, por (4.12b) los coeficientes de Walsh de  $g^-$  de orden menor que  $k$  deben ser nulos. Así, el único coeficiente de Walsh de  $g^-$  que puede tomar valor distinto de cero es  $w_{2^1+\dots+2^k}^-$ .

Por otra parte, como  $\mathbf{f}$  y  $\mathbf{g}^+ = \mathbf{f}^0 + \mathbf{f}^1$  son unitarios, necesariamente

$${}^t \mathbf{f}^0 \mathbf{f}^1 = \sum_{j=0}^{2^k-1} f_j f_{2^k+j} = 0, \quad (4.14)$$

y como  $f$  toma valores no negativos se cumple  $f_j f_{2^k+j} = 0$ , para todo  $0 \leq j < 2^k$ . Esto conduce a que  $\mathbf{g}^- = \mathbf{f}^0 - \mathbf{f}^1$  también es unitario.

Ahora, por la idempotencia de  $\mathbf{W}_\ell$ ,  $\|\mathbf{w}^-\| = \|\mathbf{g}^-\| = 1$  y como consecuencia de esto

$$\mathbf{w}^- = \pm(0, \dots, 0, 1).$$

Entonces

$${}^t \mathbf{g}^- = \mathbf{w}^- \mathbf{W}_k = (0, \dots, 0, 1) 2^{-k/2} \mathbf{V}_k = 2^{-\frac{k}{2}} \left( (-1)^{u(s \wedge 2^k - 1)} \right)_{0 \leq s < 2^k} = 2^{-\frac{k}{2}} \left( (-1)^{u(s)} \right)_{0 \leq s < 2^k}$$

o

$${}^t \mathbf{g}^- = (0, \dots, 0, -1) \mathbf{W}_k = -2^{-\frac{k}{2}} \left( (-1)^{u(s \wedge 2^k - 1)} \right)_{0 \leq s < 2^k} = 2^{-\frac{k}{2}} \left( (-1)^{u(s)+1} \right)_{0 \leq s < 2^k}.$$

De esta forma, si  $\mathbf{w}^- = (0, \dots, 0, 1)$ , se tiene

$$\begin{aligned} {}^t \mathbf{g}^- &= (0, \dots, 0, 1) \mathbf{W}_k = 2^{-\frac{k}{2}} \left( (-1)^{u(s)} \right)_{0 \leq s < 2^k} \\ &= 2^{-\frac{k}{2}} \left( (-1)^{u(s)} \frac{1 + (-1)^{u(s)}}{2} \right)_{0 \leq s < 2^k} - 2^{-\frac{k}{2}} \left( (-1)^{u(s)+1} \frac{1 + (-1)^{u(s)+1}}{2} \right)_{0 \leq s < 2^k} \end{aligned}$$

$$= 2^{-\frac{k+2}{2}} \left(1 + (-1)^{u(s)}\right)_{0 \leq s < 2^k} - 2^{-\frac{k+2}{2}} \left(1 + (-1)^{u(s)+1}\right)_{0 \leq s < 2^k}.$$

Esta separación de las componentes según su signo permite afirmar que:

$${}^t\mathbf{f}^0 = 2^{-\frac{k+2}{2}} \left(1 + (-1)^{u(s)}\right)_{0 \leq s < 2^k},$$

$${}^t\mathbf{f}^1 = 2^{-\frac{k+2}{2}} \left(1 + (-1)^{u(s)+1}\right)_{0 \leq s < 2^k},$$

y, en consecuencia

$${}^t\mathbf{f} = 2^{-\frac{k+2}{2}} \left( \left(1 + (-1)^{u(s)}\right)_{0 \leq s < 2^k}, \left(1 + (-1)^{u(s)+1}\right)_{0 \leq s < 2^k} \right).$$

Del mismo modo, para  $\mathbf{g}^- = \mathbf{W}_k {}^t(0, \dots, 0, -1)$ , se obtiene:

$${}^t\mathbf{f} = 2^{-\frac{k+2}{2}} \left( \left(1 + (-1)^{u(s)+1}\right)_{0 \leq s < 2^k}, \left(1 + (-1)^{u(s)}\right)_{0 \leq s < 2^k} \right).$$

2. En la segunda parte de la demostración ( $\ell > k + 1$ ) veremos que ningún vector alcanza el valor máximo de la  $k$ -epistasis normalizada. Comenzaremos demostrando que, para  $\ell = k + 2$ , no existen funciones no negativas  $f$ , con vector asociado  $\mathbf{f} \in \mathbb{R}^{2^{k+2}}$  unitario, tales que  $\gamma_{k+2,k}(f) = 2^k$ .

Teniendo en cuenta la relación (4.11) de la página 148, se verifica:

$$\begin{aligned} \gamma_{k+2,k}(f) &= \gamma_{k+1,k}(g^+) + \gamma_{k+1,k-1}(g^-) \\ &= \gamma_{k+1,k}(g^+) + \gamma_{k,k-1}(g^{-+}) + \gamma_{k,k-2}(g^{--}). \end{aligned}$$

Puesto que  $\gamma_{k+1,k}(g^+) \geq 2^k$ ,  $\gamma_{k,k-1}(g^{-+}) \geq 0$  y  $\gamma_{k,k-2}(g^{--}) \geq 0$ , para que  $\gamma_{k+2,k}(f)$  tome el valor  $2^k$ , se necesita  $\gamma_{k+1,k}(g^+) = 2^k$  y  $\gamma_{k,k-1}(g^{-+}) = \gamma_{k,k-2}(g^{--}) = 0$ .

Supondremos que  $\gamma_{k+1,k}(g^+) = 2^k$ ,  $\gamma_{k,k-1}(g^{-+}) = 0$  y demostraremos que, en estas condiciones,  $\gamma_{k,k-2}(g^{--}) > 0$ .

Si  $\gamma_{k+1,k}(g^+) = 2^k$ , entonces  $\mathbf{g}^+$  es uno de los vectores dados en (4.13a) y (4.13b). Supondremos, sin que ello suponga ninguna restricción, que  $\mathbf{g}^+$  es el vector dado por la expresión (4.13a), i.e.,

$${}^t\mathbf{g}^+ = 2^{-\frac{k+2}{2}} \left( \left(1 + (-1)^{u(s)}\right)_{0 \leq s < 2^k}, \left(1 + (-1)^{u(s)+1}\right)_{0 \leq s < 2^k} \right).$$

De esta forma, se puede asegurar que:

$$(\mathbf{g}^+)_s = \begin{cases} 2^{-\frac{k}{2}} & \text{para } 0 \leq s < 2^k, \text{ y } u(s) \text{ par,} \\ 0 & \text{para } 0 \leq s < 2^k, \text{ y } u(s) \text{ impar,} \\ 0 & \text{para } 2^k \leq s < 2^{k+1}, \text{ y } u(s) \text{ par,} \\ 2^{-\frac{k}{2}} & \text{para } 2^k \leq s < 2^{k+1}, \text{ y } u(s) \text{ impar.} \end{cases}$$

Observemos que, al ser  $f$  no negativa, si  $(\mathbf{g}^+)_s = (\mathbf{f}^0)_s + (\mathbf{f}^1)_s = f_s + f_{s+2^k} = 0$ , entonces  $f_s = f_{s+2^k} = 0$  y, en consecuencia,  $(\mathbf{g}^-)_s = f_s - f_{s+2^k} = 0$ . Por ello, para  $0 \leq s < 2^k$

i) si  $u(s)$  es impar, entonces  $(\mathbf{g}^-)_s = 0$ ,

ii) si  $u(s)$  es par, entonces  $(\mathbf{g}^-)_{s+2^k} = 0$ .

De esto se deduce que el vector  $\mathbf{g}^{-+} = (\mathbf{g}^-)^0 + (\mathbf{g}^-)^1$  verifica, si  $0 \leq s < 2^k$ :

$$(\mathbf{g}^{-+})_s = (\mathbf{g}^-)_s + (\mathbf{g}^-)_{s+2^k} = \begin{cases} (\mathbf{g}^-)_s & \text{para } u(s) \text{ par,} \\ (\mathbf{g}^-)_{s+2^k} & \text{para } u(s) \text{ impar,} \end{cases}$$

y el vector  $\mathbf{g}^{--} = (\mathbf{g}^-)^0 - (\mathbf{g}^-)^1$  verifica para  $0 \leq s < 2^k$ :

$$(\mathbf{g}^{--})_s = (\mathbf{g}^-)_s - (\mathbf{g}^-)_{s+2^k} = \begin{cases} (\mathbf{g}^-)_s & \text{para } u(s) \text{ par,} \\ -(\mathbf{g}^-)_{s+2^k} & \text{para } u(s) \text{ impar.} \end{cases}$$

Como, por hipótesis,  $\gamma_{k,k-1}(g^{-+}) = 0$ , la proposición B.3.3 afirma que  $\mathbf{g}^{-+}$  es un múltiplo del vector formado por los elementos de la última columna de la matriz de Walsh  $\mathbf{W}_k$ , es decir:

$$(\mathbf{g}^{-+})_s = 2^{-\frac{k}{2}} (-1)^{u(s)} \begin{cases} 2^{-\frac{k}{2}} & \text{para } u(s) \text{ par,} \\ -2^{-\frac{k}{2}} & \text{para } u(s) \text{ impar,} \end{cases}$$

y, en consecuencia,

$$(\mathbf{g}^{--})_s = \begin{cases} 2^{-\frac{k}{2}} & \text{para } u(s) \text{ par,} \\ 2^{-\frac{k}{2}} & \text{para } u(s) \text{ impar,} \end{cases}$$

es decir,  $(\mathbf{g}^{--})_s = 2^{-\frac{k}{2}} {}^t(1, \dots, 1)$ .

**Nota:** La otra opción posible para el vector  $\mathbf{g}^+$  conduce al mismo vector  $\mathbf{g}^{--}$ .

En cualquier caso, teniendo en cuenta (4.3) (ver página 125), se verifica:

$$\gamma_{k,k-2}(g^{--}) = 2^{-k} {}^t(1, \dots, 1) \mathbf{G}_{k,k-2}(1, \dots, 1)$$

$$\begin{aligned}
&= 2^{-k} \sum_{0 \leq i, j < 2^k} g_{ij}^{k, k-2} \\
&= 2^{-k} 2^{2k} = 2^k > 0,
\end{aligned}$$

lo que concluye la demostración. □

#### 4.4 Relación con la epistasis en el sentido de Davidor

En el capítulo 3 se introdujo la noción de 2-epistasis siguiendo las ideas originalmente descritas por Davidor en [4]. Posteriormente, el uso del Álgebra Lineal nos ha permitido manejar el concepto de 2-epistasis normalizada con una caracterización que, al menos desde el punto de vista teórico y calculatorio, resulta cómoda. Sin embargo, en este capítulo, se ha introducido la  $k$ -epistasis normalizada desde un punto de vista puramente algebraico, sin tener en cuenta las ideas originales de Davidor.

En esta sección veremos como el desarrollo algebraico realizado en las secciones precedentes de este cuarto capítulo enlaza satisfactoriamente con el concepto de epistasis definido a través de los valores medios de ciertos esquemas.

Para una función  $f : \Omega_\ell \rightarrow \mathbb{R}$  se ha definido la  $k$ -epistasis normalizada como:

$$\varepsilon_{\ell, k}^*(f) = 1 - \frac{{}^t \mathbf{f} \mathbf{E}_{\ell, k} \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}} = \frac{{}^t \mathbf{f} (\mathbf{I}_\ell - \mathbf{E}_{\ell, k}) \mathbf{f}}{{}^t \mathbf{f} \mathbf{f}}.$$

En lo que sigue, dado que no existe posibilidad de confusión, suprimiremos los índices  $\ell$  y  $k$ .

En primer lugar, definimos el vector  $\mathbf{e} = \mathbf{f} - \mathbf{E} \mathbf{f}$  de componentes:

$$\mathbf{e}_s = f(s) - (\mathbf{E} \mathbf{f})_s,$$

siendo:

$$(\mathbf{E} \mathbf{f})_s = \frac{1}{2^\ell} (\mathbf{G} \mathbf{f})_s = \frac{1}{2^\ell} \sum_{t=0}^{2^\ell-1} g_{st} f(t)$$

$$\begin{aligned}
&= \frac{1}{2^\ell} \sum_{t=0}^{2^\ell-1} \left[ \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell-1-k+j}{j} \binom{\ell-d_{st}^\ell}{k-j} \right] f(t) \\
&= \frac{1}{2^\ell} \sum_{j=0}^k (-1)^j 2^{k-j} \binom{\ell-1-k+j}{j} \left[ \sum_{t=0}^{2^\ell-1} \binom{\ell-d_{st}^\ell}{k-j} f(t) \right] \\
&= \sum_{j=0}^{k-1} (-1)^j 2^{k-j-\ell} \binom{\ell-1-k+j}{j} \left[ \sum_{t=0}^{2^\ell-1} \binom{\ell-d_{st}^\ell}{k-j} f(t) \right] \\
&\quad + 2^{-\ell} (-1)^k \binom{\ell-1}{k} \sum_{t=0}^{2^\ell-1} f(t).
\end{aligned}$$

Para  $j \in \{0, \dots, k-1\}$ , la suma

$$\sum_{t=0}^{2^\ell-1} \binom{\ell-d_{st}^\ell}{k-j} f(t) \tag{4.16}$$

es una combinación lineal de los valores de  $f$  en aquellas cadenas  $t \in \Omega_\ell$  para las que  $\ell - d_{st}^\ell \geq k - j$ , es decir, las cadenas que coinciden con  $s$  en, al menos,  $k - j$  alelos.

Si denotamos por  $\Omega_{i_1 \dots i_{k-j}}(s)$  el conjunto de cadenas cuyos alelos  $i_1, \dots, i_{k-j}$  coinciden con los de  $s$ , es decir:

$$\Omega_{i_1 \dots i_{k-j}}(s) = \{t \in \Omega_\ell; t_{i_1} = s_{i_1}, \dots, t_{i_{k-j}} = s_{i_{k-j}}\}$$

y por  $f(\Omega_{i_1 \dots i_{k-j}}(s))$  su idoneidad media,

$$f(\Omega_{i_1 \dots i_{k-j}}(s)) = \frac{1}{2^{\ell-k+j}} \sum_{t \in \Omega_{i_1 \dots i_{k-j}}(s)} f(t),$$

entonces se verifica que las cadenas  $t$  para las que el coeficiente de  $f(t)$  es no nulo en la suma (4.16) son las que pertenecen a alguno de los conjuntos  $\Omega_{i_1 \dots i_{k-j}}(s)$  antes descritos. Además, el coeficiente  $\binom{\ell-d_{st}^\ell}{k-j}$  no es otra cosa que el número de conjuntos  $\Omega_{i_1 \dots i_{k-j}}(s)$  a los que pertenece la cadena  $t$ , es decir:

$$\sum_{t=0}^{2^\ell-1} \binom{\ell-d_{st}^\ell}{k-j} f(t) = \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} \sum_{t \in \Omega_{i_1, \dots, i_{k-j}}(s)} f(t).$$

Entonces:

$$(\mathbf{E} f)_s = \sum_{j=0}^{k-1} (-1)^j 2^{k-j-\ell} \binom{\ell-1-k+j}{j} \left( \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} \sum_{t \in \Omega_{i_1, \dots, i_{k-j}}(s)} f(t) \right)$$

$$\begin{aligned}
& +2^{-\ell}(-1)^k \binom{\ell-1}{k} \sum_{t \in \Omega_\ell} f(t) \\
= & \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} \left( \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} \frac{1}{2^{\ell-k+j}} \sum_{t \in \Omega_{i_1 \dots i_{k-j}}(s)} f(t) \right) \\
& + (-1)^k \binom{\ell-1}{k} \frac{1}{2^\ell} \sum_{t \in \Omega_\ell} f(t).
\end{aligned}$$

Sustituyendo en el último sumando de la expresión anterior el valor de  $\binom{\ell-1}{k}$  dado por el lema B.3.5 tenemos:

$$\begin{aligned}
(\mathbf{E}f)_s &= \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} f(\Omega_{i_1 \dots i_{k-j}}(s)) \\
& + (-1)^{2k} \left\{ 1 + \sum_{j=0}^{k-1} (-1)^{j+1} \binom{\ell}{k-j} \binom{\ell-1-k+j}{j} \right\} \bar{f} \\
= & \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} f(\Omega_{i_1 \dots i_{k-j}}(s)) \tag{4.17} \\
& - \left\{ \sum_{j=0}^{k-1} (-1)^j \binom{\ell}{k-j} \binom{\ell-1-k+j}{j} \right\} \bar{f} + \bar{f} \\
= & \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} \\
& \times \left\{ \left( \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} f(\Omega_{i_1 \dots i_{k-j}}(s)) \right) - \binom{\ell}{k-j} \bar{f} \right\} + \bar{f} \\
= & \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} \left[ \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} \{f(\Omega_{i_1 \dots i_{k-j}}(s)) - \bar{f}\} \right] + \bar{f}.
\end{aligned}$$

Con las notaciones usadas en la epistasis ordinaria y en la 2-epistasis, el término

$$f(\Omega_{i_1 \dots i_{k-j}}(s)) - \bar{f}$$

se define como el *valor de exceso alélico* de la  $(k-j)$ -tupla  $(s_{i_1}, \dots, s_{i_{k-j}})$ ,  $E_{i_1 \dots i_{k-j}}(s_{i_1} \dots s_{i_{k-j}})$ , y la suma

$$\begin{aligned}
EG(s) &= \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} \sum_{0 \leq i_1 < \dots < i_{k-j} < \ell} \{f(\Omega_{i_1 \dots i_{k-j}}(s)) - \bar{f}\} \\
&= \sum_{j=0}^{k-1} (-1)^j \binom{\ell-1-k+j}{j} E_{i_1 \dots i_{k-j}}(s_{i_1} \dots s_{i_{k-j}})
\end{aligned}$$

se define como el *exceso génico* de la cadena  $s \in \Omega_\ell$ . Por todo ello,

$$EG(s) + \bar{f} = (\mathbf{E} \mathbf{f})_s$$

es el *valor génico previsto* de  $s$  y, finalmente,  $\mathbf{e}_s$  es su  $k$ -*epistasis*, retornando así a la idea original de Davidor, ahora generalizada.



# Apéndice A

## El Teorema de los Esquemas

---

Este primer apéndice lo dedicamos a una explicación detallada del Teorema de los Esquemas, también conocido como teorema fundamental de los Algoritmos Genéticos. En él se analiza la evolución de los esquemas presentes en una población a la que se le aplica un Algoritmo Genético y se proporciona una estimación del número de individuos pertenecientes a estos esquemas que se espera tener en la generación siguiente.

Supongamos que en una determinada generación  $t$  en la población  $\mathcal{P}(t)$  existen  $m = m(\mathcal{H}, t)$  individuos pertenecientes a un esquema  $\mathcal{H}$  (nuestra notación refleja el hecho de que este número  $m$  depende tanto del esquema elegido  $\mathcal{H}$ , como de la generación  $t$  de la población) y consideremos el modelo de selección proporcional. Entonces, durante la selección, la probabilidad de que un individuo  $i$  sea seleccionado es proporcional a su calidad, i.e.,  $p_i = \frac{f(i)}{\sum_{j \in \mathcal{P}(t)} f(j)}$ . En consecuencia, la probabilidad de elegir un elemento del esquema  $\mathcal{H}$  es:

$$p(s \in \mathcal{H}) = \sum_{s \in \mathcal{H}} p_s = \sum_{s \in \mathcal{H}} \frac{f(s)}{\sum_{j \in \mathcal{P}(t)} f(j)} = p.$$

Si el tamaño de la población  $\mathcal{P}(t)$  es  $n$ , entonces el número esperado de representantes de  $\mathcal{H}$  en la generación siguiente será la esperanza de la variable aleatoria que sigue la distribución binomial  $\mathcal{B}(n, p)$ , es decir:

$$E(m(\mathcal{H}, t+1)) = np = n \sum_{s \in \mathcal{H}} \frac{f(s)}{\sum_{j \in \mathcal{P}(t)} f(j)} = m(\mathcal{H}, t) \frac{f(\mathcal{H})}{\bar{f}} \quad (\text{A.1})$$

donde,  $\bar{f}$  denota la idoneidad media de la población  $\mathcal{P}(t)$ ,

$$\bar{f} = \frac{1}{n} \sum_{j \in \mathcal{P}(t)} f(j)$$

y

$$f(\mathcal{H}) = \frac{1}{m(\mathcal{H}, t)} \sum_{s \in \mathcal{H}} f(s)$$

es el ajuste medio del esquema  $\mathcal{H}$  en la generación  $t$ -ésima.

De (A.1) se deduce que crecerá el número de individuos pertenecientes a esquemas con idoneidad por encima de la media  $\bar{f}$ , mientras que disminuirá ese número en aquellos esquemas cuya calidad es inferior a la calidad media.

Dado que la selección no introduce elementos nuevos en la población (no explora otras zonas del espacio de búsqueda ya que simplemente realiza copias de individuos presentes en la población) es necesario introducir el efecto de los operadores genéticos.

Con el objeto de analizar cómo afecta el cruce en los diferentes esquemas consideremos la cadena  $c = 0111000$  que representa (entre otros) a los esquemas  $\mathcal{H}_1 = \#1\#\#\#0$  y  $\mathcal{H}_2 = \#\#\#10\#\#$ .

Si consideramos el cruce clásico (en un punto) entre el tercer y cuarto alelo<sup>1</sup>, a no ser que el cromosoma que se cruce con  $c$  tenga un cero en su último alelo o un uno en el segundo, la estructura del esquema  $\mathcal{H}_1$  se destruirá con ese cruce. Por el contrario el esquema  $\mathcal{H}_2$  sobrevivirá al cruce, pues al menos uno de los descendientes conserva la estructura de  $\mathcal{H}_2$ .

Teniendo en cuenta que en el cruce clásico el punto de corte se elige aleatoriamente, entonces la probabilidad de que el esquema  $\mathcal{H}_1$  sea destruido es  $\frac{\delta(\mathcal{H}_1)}{6} = \frac{5}{6}$ ; mientras que la probabilidad de que lo sea  $\mathcal{H}_2$  es  $\frac{\delta(\mathcal{H}_2)}{6} = \frac{1}{6}$ . En general la probabilidad de supervivencia de un esquema  $\mathcal{H}$  tras el cruce será:  $p_{superv} = 1 - \frac{\delta(\mathcal{H})}{\ell-1}$ ; siendo  $\ell$  la longitud de las cadenas.

En realidad, el cruce se realiza con una probabilidad  $p_c$  –fijada por el analista– y, por tanto, la probabilidad de supervivencia de un esquema  $\mathcal{H}$  es:

$$p_{superv} \geq 1 - p_c \frac{\delta(\mathcal{H})}{\ell-1}. \quad (\text{A.2})$$

El símbolo  $\geq$  se debe a que aún en el supuesto de que el punto de cruce caiga entre dos posiciones fijas, es posible que el esquema sobreviva aunque, obviamente, esta probabilidad es muy pequeña.

---

<sup>1</sup>Las posiciones se consideran aquí contadas de izquierda a derecha.

Considerando el efecto combinado (e independiente) de la selección y el cruce, las fórmulas (A.1) y (A.2) proporcionan:

$$E(m(\mathcal{H}, t+1)) \geq m(\mathcal{H}, t) \frac{f(\mathcal{H})}{\bar{f}} \left[ 1 - p_c \frac{\delta(\mathcal{H})}{\ell - 1} \right].$$

Finalmente, denotemos por  $p_m$  la probabilidad de realizar una mutación en un alelo. Para que un esquema sobreviva a una mutación no debe cambiar ninguno de sus alelos fijos (aquellos distintos de  $\#$ ). Así, la probabilidad de que un esquema  $\mathcal{H}$  sobreviva a una mutación es  $(1 - p_m)^{o(\mathcal{H})}$  que, para valores muy pequeños de  $p_m$ , se aproxima por  $1 - o(\mathcal{H})p_m$ .

El efecto combinado de la selección, el cruce y la mutación conduce a:

$$E(m(\mathcal{H}, t+1)) \geq m(\mathcal{H}, t) \frac{f(\mathcal{H})}{\bar{f}} \left[ 1 - p_c \frac{\delta(\mathcal{H})}{\ell - 1} \right] [1 - p_m o(\mathcal{H})].$$

expresión que, para valores pequeños de  $p_m$ , se aproxima por

$$E(m(\mathcal{H}, t+1)) \geq m(\mathcal{H}, t) \frac{f(\mathcal{H})}{\bar{f}} \left[ 1 - p_c \frac{\delta(\mathcal{H})}{\ell - 1} - p_m o(\mathcal{H}) \right].$$

que constituye la formulación definitiva del *Teorema de los Esquemas* o *Teorema Fundamental de los Algoritmos Genéticos*:

El Teorema de los Esquemas describe la propagación de un esquema de una generación a otra. Sin embargo, si la composición de la población es desconocida, determinar el número de representantes de un esquema es imposible. Recíprocamente, si los esquemas se desconocen, también será desconocida la población. Por ello, frecuentes afirmaciones que aluden a la dirección en que los AG realizan la búsqueda no se ajustan fielmente a las conclusiones del teorema.

Como se afirma en [45],

*“El Teorema de los esquemas clásico no es suficientemente poderoso para hacer predicciones acerca de la dirección de la búsqueda genética. No dice nada acerca de cuándo o cómo un esquema puede ser representado por cadenas que no están en la población inicial. [...] El Teorema de los Esquemas es sólo una afirmación acerca de cómo la selección actúa sobre la población inicial.”*

Lo que se puede resumir diciendo que este teorema proporciona una explicación cualitativa de la actuación de los AG.



# Apéndice B

## Resultados auxiliares

---

En la primera sección de este apéndice se incluyen, de forma concisa, la definición y algunas de las propiedades del producto tensor de matrices –o producto de Kronecker– utilizadas a lo largo de capítulos anteriores.

Las secciones restantes del apéndice contienen una serie de resultados que se han utilizado a lo largo de la presente memoria. Aunque pueden resultar conocidos para el lector, se ha considerado conveniente incluirlos. Otros son novedosos, pero su gran extensión pueden distraer del argumento principal del resultado en el que se utilizan.

### B.1 El producto tensor de matrices

**Definición B.1.1.** *El producto tensor de dos matrices  $\mathbf{A} \in \mathcal{M}_{n \times m}(\mathbb{Q})$  y  $\mathbf{B} \in \mathcal{M}_{p \times q}(\mathbb{Q})$  es la matriz:*

$$\mathbf{A} \otimes \mathbf{B} = (a_{ij}\mathbf{B})_{ij} \in \mathcal{M}_{np \times mq}(\mathbb{Q}).$$

Por ejemplo, si  $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$  y  $\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$ , entonces

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{pmatrix}.$$

**Proposición B.1.2.** Para cualesquiera  $\mathbf{A}, \mathbf{C} \in \mathcal{M}_{n \times m}(\mathbb{Q})$  y  $\mathbf{B}, \mathbf{D} \in \mathcal{M}_{p \times q}(\mathbb{Q})$ , se verifica:

i)  $(\mathbf{A} \otimes \mathbf{B}) \cdot (\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A} \cdot \mathbf{C}) \otimes (\mathbf{B} \cdot \mathbf{D})$ .

ii)  ${}^t(\mathbf{A} \otimes \mathbf{B}) = {}^t\mathbf{A} \otimes {}^t\mathbf{B}$ .

iii)  $\mathbf{A} \otimes \mathbf{B} + \mathbf{A} \otimes \mathbf{D} = \mathbf{A} \otimes (\mathbf{B} + \mathbf{D})$ .

## B.2 Resultados auxiliares del capítulo 2

El siguiente lema se utiliza en la demostración de la proposición 2.6.5.

**Lema B.2.1.** Para cualesquiera números enteros  $p$  y  $n$  tales que  $0 \leq n \leq p$ , con  $p > 0$ , se verifica:

$$\sum_{q=0}^{p-1} \left( \sum_{j=0}^q (-1)^{q-j} \binom{n}{j} \binom{p-n}{q-j} \right) = (-1)^{p-n+1}.$$

*Demostración.*

$$\begin{aligned} \sum_{q=0}^{p-1} \left( \sum_{j=0}^q (-1)^{q-j} \binom{n}{j} \binom{p-n}{q-j} \right) &= \sum_{j=0}^{p-1} \left( \sum_{q=j}^{p-1} (-1)^{q-j} \binom{n}{j} \binom{p-n}{q-j} \right) \\ &= \sum_{j=0}^{p-1} \left( \sum_{z=0}^{p-j-1} (-1)^z \binom{n}{j} \binom{p-n}{z} \right) \\ &= \sum_{z=0}^{p-1} \left( \sum_{j=0}^{p-z-1} (-1)^z \binom{n}{j} \binom{p-n}{z} \right) \\ &= \sum_{z=0}^{p-n} (-1)^z \binom{p-n}{z} \sum_{j=0}^{p-z-1} \binom{n}{j}. \end{aligned}$$

Ahora, como

$$\sum_{j=0}^{p-z-1} \binom{n}{j} = \begin{cases} \sum_{j=0}^n \binom{n}{j} = 2^n & \text{si } p-z-1 \geq n \text{ (i.e., } z > p-n) \\ \sum_{j=0}^{n-1} \binom{n}{j} = 2^n - 1 & \text{si } z = p-n \end{cases}$$

entonces

$$\begin{aligned} \sum_{z=0}^{p-n} (-1)^z \binom{p-n}{z} \sum_{j=0}^{p-z-1} \binom{n}{j} &= \sum_{z=0}^{p-n-1} (-1)^z \binom{p-n}{z} \sum_{j=0}^{p-z-1} \binom{n}{j} + (-1)^{p-n} \sum_{j=0}^{n-1} \binom{n}{j} \\ &= 2^n \sum_{z=0}^{p-n-1} (-1)^z \binom{p-n}{z} + (-1)^{p-n} (2^n - 1) \\ &= 2^n \sum_{z=0}^{p-n} (-1)^z \binom{p-n}{z} + (-1)^{p-n+1} \\ &= (-1)^{p-n+1}, \end{aligned}$$

como se quería. □

### B.3 Resultados auxiliares del capítulo 4

El siguiente lema se utiliza en la demostración del lema B.3.2 de este apéndice.

**Lema B.3.1.** *Dados  $p, q, r, i \in \mathbb{N}$ , se verifica:*

$$\binom{p+1}{q} \binom{q}{q-i} \binom{p+1-q}{r-q+i} = \binom{p+1}{r} \binom{p+1-r}{i} \binom{r}{q-i}.$$

*Demostración.* Si denotamos por  $\Upsilon$  al producto  $\binom{p+1}{q} \binom{q}{q-i} \binom{p+1-q}{r-q+i}$ , se verifica

$$\begin{aligned} \Upsilon &= \frac{(p+1)!}{q! (p+1-q)!} \frac{q!}{(q-i)! i!} \frac{(p+1-q)!}{(r-q+i)! (p+1-r-i)!} \\ &= \frac{(p+1)!}{(q-i)! i! (r-q+i)! (p+1-r-i)!} \end{aligned}$$

$$\begin{aligned}
&= \frac{(p+1)!}{r! (p+1-r)!} \frac{(p+1-r)!}{i! (p+1-r-i)!} \frac{r!}{(q-i)! (r-q+i)!} \\
&= \binom{p+1}{r} \binom{p+1-r}{i} \binom{r}{q-i}.
\end{aligned}$$

□

El siguiente lema se utiliza en la demostración de la proposición 4.2.1.

**Lema B.3.2.** Para todo  $p \in \mathbb{N}$  y  $0 < r \leq p+1$  se verifica:

$$\sum_{q=0}^p \binom{p+1}{q} \left[ \sum_{j=0}^q (-1)^j \binom{q}{j} \binom{p+1-q}{r-j} \right] = (-1)^{r+1} \binom{p+1}{r}. \quad (\text{B.1})$$

*Demostración.* Denotemos por  $\mathcal{S}$  al primer miembro de (B.1), entonces

$$\begin{aligned}
\mathcal{S} &= \sum_{q=0}^p \binom{p+1}{q} \left[ \sum_{i=0}^q (-1)^{q-i} \binom{q}{q-i} \binom{p+1-q}{r-q+i} \right] \\
&= \sum_{q=0}^p \binom{p+1}{q} \left[ \sum_{i=0}^{\min\{q, p+1-r\}} (-1)^{q-i} \binom{q}{q-i} \binom{p+1-q}{r-q+i} \right] \\
&= \sum_{q=0}^p \sum_{i=0}^{\min\{q, p+1-r\}} (-1)^{q-i} \binom{p+1}{q} \binom{q}{q-i} \binom{p+1-q}{r-q+i} \\
&= \sum_{i=0}^{\min\{p, p+1-r\}} \left[ \sum_{q=i}^p (-1)^{q-i} \binom{p+1}{q} \binom{q}{q-i} \binom{p+1-q}{r-q+i} \right] \\
&= \sum_{i=0}^{p+1-r} \left[ \sum_{q=i}^p (-1)^{q-i} \binom{p+1}{q} \binom{q}{q-i} \binom{p+1-q}{r-q+i} \right].
\end{aligned}$$

Ahora, el lema B.3.1 conduce a

$$\begin{aligned}
\mathcal{S} &= \sum_{i=0}^{p+1-r} \left[ \sum_{q=i}^p (-1)^{q-i} \binom{p+1}{r} \binom{p+1-r}{i} \binom{r}{q-i} \right] \\
&= \sum_{i=0}^{p+1-r} \binom{p+1}{r} \binom{p+1-r}{i} \left[ \sum_{q=i}^p (-1)^{q-i} \binom{r}{q-i} \right].
\end{aligned}$$

Ahora bien, en la suma con índice  $i$ , todos los sumandos se anulan, excepto el último. En efecto, si  $i \leq p - r$ , entonces  $p - i \geq r$  y así:

$$\sum_{q=i}^p (-1)^{q-i} \binom{r}{q-i} = \sum_{q=i}^{i+r} (-1)^{q-i} \binom{r}{q-i} = \sum_{z=0}^r (-1)^z \binom{r}{z} = 0.$$

En consecuencia, el sumando correspondiente al índice  $i = p + 1 - r$  proporciona el valor de la suma  $\mathcal{S}$ :

$$\begin{aligned} \mathcal{S} &= \binom{p+1}{r} \sum_{q=p+1-r}^p (-1)^{q-(p+1-r)} \binom{r}{q-(p+1-r)} \\ &= \binom{p+1}{r} \sum_{z=0}^{r-1} (-1)^z \binom{r}{z} = \binom{p+1}{r} (-1)^{r+1}. \end{aligned}$$

□

El siguiente resultado se ha utilizado en la demostración de la proposición 4.3.5.

**Proposición B.3.3.** *Sea  $f : \Omega_\ell \rightarrow \mathbb{R}$  con vector asociado  $\mathbf{f}$ . Si  $\gamma_{\ell, \ell-1}(f) = 0$  entonces  $\mathbf{f}$  es proporcional al vector  $\mathbf{v}_\ell = ((-1)^{u(s)})_{0 \leq s < 2^\ell}$ .*

**Nota:** Las componentes de  $\mathbf{v}_\ell$  coinciden con los elementos del vector que constituye la última columna de la matriz de Walsh  $\mathbf{W}_\ell$ .

*Demostración.* La demostración se realiza por inducción en el valor de  $\ell$ .

Si  $\ell = 1$  el resultado es evidente. Supongamos que éste se verifica para  $\ell$  y lo comprobaremos para  $\ell + 1$ .

Si  $\gamma_{\ell+1, \ell}(f) = 0$ , entonces

$$\begin{aligned} \gamma_{\ell+1, \ell}(f) &= \gamma_{\ell, \ell}(f^0 + f^1) + \gamma_{\ell, \ell-1}(f^0 - f^1) \\ &= 2^\ell \|\mathbf{f}^0 + \mathbf{f}^1\|^2 + \gamma_{\ell, \ell-1}(f^0 - f^1). \end{aligned}$$

Por ser la matriz  $\mathbf{G}_{\ell, \ell-1}$  semidefinida positiva y la hipótesis de inducción, la igualdad  $\gamma_{\ell+1, \ell}(f) = 0$  implica  $\mathbf{f}^0 = -\mathbf{f}^1$  y  $\mathbf{f}^0 - \mathbf{f}^1 = \alpha \mathbf{v}_\ell$ , para algún  $\alpha \in \mathbb{R}$ . En consecuencia,

$$\mathbf{f}^0 = \frac{1}{2} (\mathbf{f}^0 - \mathbf{f}^1) = \frac{\alpha}{2} \mathbf{v}_\ell = \frac{\alpha}{2} \left( (-1)^{u(s)} \right)_{0 \leq s < 2^\ell},$$

$$\mathbf{f}^1 = -\frac{\alpha}{2} \left( (-1)^{u(s)} \right)_{0 \leq s < 2^\ell} = \frac{\alpha}{2} \left( (-1)^{u(s)+1} \right)_{0 \leq s < 2^\ell} = \frac{\alpha}{2} \left( (-1)^{u(s)} \right)_{2^\ell \leq s < 2^{\ell+1}}$$

y, finalmente,

$$\mathbf{f} = \frac{\alpha}{2} \left( (-1)^{u(s)} \right)_{0 \leq s < 2^{\ell+1}} = \frac{\alpha}{2} \mathbf{v}_{\ell+1}.$$

□

El siguiente lema es necesario para demostrar el lema B.3.5 de este apéndice.

**Lema B.3.4.** *Para cualesquiera  $\ell, k \in \mathbb{Z}$ , tales que  $0 \leq k \leq \ell - 2$ , se verifica:*

$$\sum_{i=0}^k \frac{(-1)^i}{\ell - k + i - 1} \binom{k}{i} = \frac{1}{(k+1) \binom{\ell-1}{k+1}}.$$

*Demostración.* Se realizará por inducción en el valor de  $\ell$ . Para  $k = 0$  el resultado es evidente para cualquier  $\ell$ , en particular para  $\ell = 2$  ( $k = 0$  es el único valor posible en ese caso). Supongamos ahora la afirmación válida hasta  $\ell - 1$  y demostrémosla para  $\ell$  (por inducción en los valores de  $k$ ).

Dado que para  $k = 0$  es cierta (para cualquier valor de  $\ell$ ), supondremos que para  $\ell$  es cierto el resultado hasta  $k$  y lo demostraremos para  $k + 1$ .

En efecto,

$$\begin{aligned} \sum_{i=0}^{k+1} \frac{(-1)^i}{\ell - (k+1) + i - 1} \binom{k+1}{i} &= \sum_{i=0}^{k+1} \frac{(-1)^i}{\ell - k + i - 2} \left[ \binom{k}{i-1} + \binom{k}{i} \right] \\ &= \frac{1}{\ell - k - 2} + \sum_{i=1}^k \frac{(-1)^i}{\ell - k + i - 2} \left[ \binom{k}{i-1} + \binom{k}{i} \right] \\ &\quad + \frac{(-1)^{k+1}}{\ell - 1} \\ &= \frac{1}{\ell - k - 2} + \frac{(-1)^{k+1}}{\ell - 1} + \sum_{i=1}^k \frac{(-1)^i}{\ell - k + (i-1) - 1} \binom{k}{i-1} \\ &\quad + \sum_{i=1}^k \frac{(-1)^i}{\ell - (k+1) + (i-1)} \binom{k}{i} \\ &= \frac{(-1)^{k+1}}{\ell - 1} + \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j - 1} \binom{k}{j} + \frac{1}{\ell - k - 2} \end{aligned}$$

$$\begin{aligned}
& + \sum_{i=1}^k \frac{(-1)^i}{\ell - (k+1) + (i-1)} \binom{k}{i} \\
& = - \sum_{j=0}^k \frac{(-1)^j}{\ell - k + j - 1} \binom{k}{j} \\
& \quad + \sum_{i=0}^k \frac{(-1)^i}{\ell - (k+1) + (i-1)} \binom{k}{i} \\
& = - \frac{1}{(k+1) \binom{\ell-1}{k+1}} + \frac{1}{(k+1) \binom{\ell-2}{k+1}} \\
& = \frac{1}{k+1} \frac{\binom{\ell-1}{k+1} - \binom{\ell-2}{k+1}}{\binom{\ell-2}{k+1} \binom{\ell-1}{k+1}} \\
& = \frac{1}{\ell - k - 2} \frac{1}{\binom{\ell-1}{k+1}} \\
& = \frac{1}{k+2} \frac{(k+2)! (\ell - k - 3)!}{(\ell - 1)!} \\
& = \frac{1}{k+2} \frac{1}{\binom{\ell-1}{k+2}}.
\end{aligned}$$

□

El siguiente lema se ha utilizado para obtener la expresión (4.17) de la página 156.

**Lema B.3.5.** Para cualesquiera  $k, \ell \in \mathbb{Z}$ ,  $0 \leq k \leq \ell - 1$  se verifica:

$$\binom{\ell-1}{k} = (-1)^k \left[ 1 + \sum_{j=0}^{k-1} (-1)^{j+1} \binom{\ell}{k-j} \binom{\ell-1-k+j}{j} \right].$$

*Demostración.* Teniendo en cuenta que:

$$\begin{aligned}
\binom{\ell}{k-j} \binom{\ell-1-k+j}{j} & = \frac{\ell (\ell-1)!}{(k-j)! (\ell-k+j) (\ell-k+j-1)!} \frac{(\ell-k+j-1)!}{j! (\ell-k-1)!} \\
& = \frac{\ell}{(\ell-k+j)} \frac{(\ell-1)!}{(\ell-k-1)!} \frac{1}{(k-j)! j!} \\
& = \frac{\ell}{(\ell-k+j)} \frac{(\ell-1)!}{(\ell-k-1)!} \frac{k!}{(k-j)! j!}
\end{aligned}$$

$$= \frac{\ell}{(\ell - k + j)} \binom{\ell - 1}{k} \binom{k}{j},$$

bastará demostrar que:

$$\binom{\ell - 1}{k} = (-1)^k \left[ 1 + \ell \binom{\ell - 1}{k} \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j} \right]. \quad (\text{B.2})$$

Esto lo haremos por inducción en los valores de  $\ell$ . Si  $\ell = 1$  entonces  $k = 0$  y el resultado es evidente.

Supongamos la propiedad cierta para  $\ell - 1$  y demostrémosla para  $\ell$  (por inducción en los valores de  $k$ ). Si  $k = 0$  de nuevo el resultado se obtiene de la sustitución directa en (B.2). Supongamos que la propiedad es cierta para  $k$  y la demostraremos para  $k + 1$ . Como

$$\begin{aligned} \binom{\ell - 1}{k + 1} &= \frac{(\ell - 1)!}{(k + 1)!(\ell - k - 2)!} & (\text{B.3}) \\ &= \frac{\ell - k - 1}{k + 1} \binom{\ell - 1}{k} \\ &= \left[ \frac{\ell - (k + 1)}{k + 1} \right] (-1)^k \left[ 1 + \ell \binom{\ell - 1}{k} \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j} \right] \\ &= (-1)^{k+1} \left[ 1 - \frac{\ell}{k + 1} \right] \left[ 1 + \ell \binom{\ell - 1}{k} \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j} \right] \\ &= (-1)^{k+1} \left[ 1 - \frac{\ell}{k + 1} + \ell \frac{k + 1 - \ell}{k + 1} \binom{\ell - 1}{k} \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j} \right] \\ &= (-1)^{k+1} \left[ 1 + \ell \left\{ -\frac{1}{k + 1} - \binom{\ell - 1}{k + 1} \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j} \right\} \right], \end{aligned}$$

comprobar (B.2) es equivalente a demostrar que

$$\binom{\ell - 1}{k + 1} \sum_{j=0}^k \frac{(-1)^{j+1}}{\ell - (k + 1) + j} \binom{k + 1}{j} = -\frac{1}{k + 1} - \binom{\ell - 1}{k + 1} \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j}$$

o, lo que es lo mismo,

$$\binom{\ell - 1}{k + 1} \left\{ \sum_{j=0}^k \frac{(-1)^{j+1}}{\ell - (k + 1) + j} \binom{k + 1}{j} + \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell - k + j} \binom{k}{j} \right\} = -\frac{1}{k + 1}.$$

En efecto, el valor de  $\mathcal{S} = \sum_{j=0}^k \frac{(-1)^{j+1}}{\ell - (k+1) + j} \binom{k+1}{j} + \sum_{j=0}^{k-1} \frac{(-1)^{j+1}}{\ell + j - k} \binom{k}{j}$  es:

$$\begin{aligned} \mathcal{S} &= -\frac{1}{\ell - (k+1)} + \sum_{j=1}^k \frac{(-1)^{j+1}}{\ell - (k+1) + j} \binom{k+1}{j} \\ &\quad - \sum_{j=1}^k \frac{(-1)^{j+1}}{\ell - (k+1) + j} \binom{k}{j-1} \\ &= -\frac{1}{\ell - (k+1)} + \sum_{j=1}^k \frac{(-1)^{j+1}}{\ell - (k+1) + j} \binom{k}{j} \\ &= \sum_{j=0}^k \frac{(-1)^{j+1}}{\ell - (k+1) + j} \binom{k}{j} \\ &= -\frac{1}{(k+1) \binom{\ell-1}{k+1}}, \end{aligned}$$

donde la última igualdad se debe al lema B.3.4.

Esto finaliza la demostración. □

El siguiente lema se ha usado en la demostración de la proposición 4.2.6.

**Lema B.3.6.** *Para todo número natural  $p$  se verifica:*

$$\begin{aligned} a) \quad & \sum_{j=1}^p \frac{j^2}{2^j} = 6 - \frac{p^2 + 4p + 6}{2^p}, \\ b) \quad & \sum_{j=1}^p \frac{j^3}{2^j} = 26 - \frac{p^3 + 6p^2 + 18p + 26}{2^p}, \\ c) \quad & \sum_{j=1}^p \frac{(p-j)j^2}{2^{j+1-p}} = 2^p(3p - 13) + p^2 + 6p + 13. \end{aligned}$$

*Demostración.* Se realiza por inducción en los valores de  $p$ .

a) El caso  $p = 1$  es evidente. Supongamos cierto el resultado para el caso  $p$  y demostrémoslo para  $p + 1$ .

$$\sum_{j=1}^{p+1} \frac{j^2}{2^j} = \sum_{j=1}^p \frac{j^2}{2^j} + \frac{(p+1)^2}{2^{p+1}} = 6 - \frac{p^2 + 4p + 6}{2^p} + \frac{(p+1)^2}{2^{p+1}}$$

$$= 6 - \frac{p^2 + 6p + 11}{2^{p+1}} = 6 - \frac{(p+1)^2 + 4(p+1) + 6}{2^{p+1}}.$$

b) La demostración de este apartado es análoga a la del anterior.

c) El caso  $p = 1$  evidente. Supuesto cierto el resultado para el caso  $p$ , usando el apartado a), lo demostraremos para  $p + 1$ .

$$\begin{aligned} \sum_{j=1}^{p+1} \frac{(p+1-j)j^2}{2^{j-p}} &= \sum_{j=1}^{p+1} \frac{(p-j)j^2}{2^{j-p}} + \sum_{j=1}^{p+1} \frac{j^2}{2^{j-p}} \\ &= 2 \sum_{j=1}^{p+1} \frac{(p-j)j^2}{2^{j+1-p}} + 2^p \sum_{j=1}^{p+1} \frac{j^2}{2^j} \\ &= 2 \{2^p(3p-13) + p^2 + 6p + 13 - 2^{-2}(p+1)^2\} \\ &\quad + 2^p \{6 - 2^{-(p+1)} [(p+1)^2 + 4(p+1) + 6]\} \\ &= 2^{p+1}(3p-10) + p^2 + 8p + 20 \\ &= 2^{p+1}(3(p+1) - 13) + (p+1)^2 + 6(p+1) + 13. \end{aligned}$$

□

# Apéndice C

## Resultados experimentales

---

La última sección de la memoria la dedicamos a recoger algunos resultados experimentales. Presentamos dos ejemplos con los que tratamos de mostrar la eficacia de los estimadores de la dificultad introducidos a lo largo de este trabajo. El algoritmo implementado es el construido en [37].

En el primer ejemplo se aplica un AG para optimizar funciones difíciles: una función nula en todas las cadenas salvo en una, una función camello y una cuasi-camello. El segundo trata de explicar, para dos tipos de funciones de laboratorio –funciones Template y funciones Royal Road generalizadas–, la diferencia en la dificultad de optimización de dichas funciones mediante un AG.

### **Ejemplo C.1:**

Las funciones elegidas en este primer ejemplo son: la función de Dirac  $\delta_0$ , que resulta difícil para el algoritmo por ser, esencialmente, una búsqueda aleatoria; la función camello  $c = \delta_0 + \delta_{2^\ell - 1}$ , una de las funciones para las que el valor de la epistasis normalizada es máximo y la función cuasi-camello  $d = \delta_0 + \delta_{2^\ell - 2}$  introducida en la subsección 4.2.3 que, al menos desde un punto de vista teórico, debe ser una función difícil –aunque menos que la camello, pues en ésta el cruce de dos cadenas óptimas diferentes conduce a la creación de una nueva

cadena no óptima-. Sin embargo, en las cuasi-camello existe una posibilidad entre  $\ell - 1$  de que esto no se produzca.

El algoritmo se ha ejecutado sobre cadenas de longitudes 10, 11, 12 y 13. Partimos de una población inicial de 50 individuos, siendo Ranking Lineal el tipo de selección elegido,  $p_c = 0.125$  la probabilidad de cruce y  $p_m = 0.01$  la de mutación. Se ejecuta el algoritmo en 20 ocasiones, deteniéndose si el número de generaciones es superior a 5000 o si el 90% de la población está formada por individuos en los que la función alcanza el valor máximo.

longitud de las cadenas	$\delta_0$	$c$	$d$
$\ell = 10$	932	514	503
$\ell = 11$	1152	789	735
$\ell = 12$	2475	1564	1314
$\ell = 13$	3101	2318	2119

Tabla C.1: Número de generaciones necesarias para obtener el máximo.

Los resultados, reflejados en la tabla C.1, pueden sorprender si se comparan con los valores de la epistasis normalizada de cada una de las funciones, pero se comprenden si, en lugar de considerar sólo este estimador, se consideran los valores de la  $k$ -epistasis para  $1 \leq k < \ell$ . Estos valores se muestran en las tablas C.2 y C.3.

Los valores de éstas tablas parecen mostrar que, si bien los valores de la  $k$ -epistasis (con  $k$  pequeño) clasifican bien las funciones difíciles, no es suficiente el valor de estos estimadores para “ordenar” las funciones atendiendo al grado de dificultad, entendiendo por grado de dificultad el número de generaciones que necesita un algoritmo genético para obtener el máximo.

$\ell = 10$			
$k$ -epistasis	$\delta_0$	$c$	$d$
0	0.9902	0.998	0.9980
1	0.9892	0.998	0.996
2	0.9453	0.9101	0.9257
3	0.8281	0.9101	0.8554
4	0.6230	0.5	0.6093
5	0.3769	0.5	0.3632
6	0.1718	0.0898	0.1992
7	0.0546	0.0898	0.0351
8	0.0107	0.0019	0.0175
9	0.0009	0.0019	0

$\ell = 11$			
$k$ -epistasis	$\delta_0$	$c$	$d$
0	0.9995	0.999	0.999
1	0.9941	0.999	0.998
2	0.9672	0.9453	0.9541
3	0.8867	0.9453	0.9101
4	0.7255	0.623	0.705
5	0.5	0.623	0.5
6	0.2744	0.1718	0.2949
7	0.1132	0.1718	0.0898
8	0.0327	0.0107	0.0458
9	0.0058	0.0107	0.0195
10	0.0004	0	0.0009

Tabla C.2: Ejecución de un AG para las funciones  $\delta_0$ ,  $c$  y  $d$  con  $\ell = 10, 11$ .

$\ell = 12$			
$k$ -epistasis	$\delta_0$	$c$	$d$
0	0.9997	0.9995	0.9995
1	0.9968	0.9995	0.999
2	0.9807	0.9672	0.9721
3	0.927	0.9672	0.9453
4	0.8061	0.7255	0.7841
5	0.6127	0.7255	0.623
6	0.3872	0.2744	0.3974
7	0.1938	0.2744	0.1718
8	0.0729	0.0327	0.0913
9	0.0192	0.0327	0.0174
10	0.0031	0.0004	0.0053
11	0.0002	0.0004	0

$\ell = 13$			
$k$ -epistasis	$\delta_0$	$c$	$d$
0	0.9998	0.9997	0.9997
1	0.9982	0.9997	0.9995
2	0.9887	0.9807	0.9834
3	0.9538	0.9807	0.9672
4	0.8665	0.8061	0.8464
5	0.7094	0.8061	0.7255
6	0.5	0.3872	0.5
7	0.2905	0.3872	0.2744
8	0.1334	0.0729	0.1535
9	0.0461	0.0729	0.0327
10	0.0112	0.0031	0.0166
11	0.0017	0.0031	0.0004
12	0.0001	0	0.0002

Tabla C.3: Ejecución de un AG para las funciones  $\delta_0$ ,  $c$  y  $d$  con  $\ell = 12, 13$ .

**Ejemplo C.2:**

Al igual que en el caso de las funciones anteriores, se ha aplicado un AG para comparar la dificultad de las funciones Royal Road generalizadas y las funciones Template. En la sección 2.4 se incluía una tabla con el número de generaciones que ha necesitado el algoritmo para alcanzar el máximo de cada función, así como la epistasis normalizada de cada una de ellas. En esa sección hacíamos notar la gran diferencia entre el número de generaciones necesitado para la convergencia en las funciones  $\mathfrak{R}_3^6$  y  $T_{64}^8$ , a pesar de que la diferencia entre los valores de la epistasis no parece tan grande. A nuestro juicio la razón está en que, en el cálculo de la epistasis, se mide la distancia de cualquier función  $f$  a un conjunto de funciones “muy selectivo”: las funciones de primer orden.

Si consultamos las tablas C.4, C.5 y C.6, en las que se muestran los valores de la  $k$ -epistasis de las funciones  $T_{64}^{2^i}$  y  $\mathfrak{R}_i^6$  ( $2 \leq i \leq 4$ ) para  $1 \leq k \leq 2^i - 1$ , se puede observar que los valores de  $k$  para los que las  $k$ -epistasis más se diferencian son  $k = 2, 3$  y  $4$ . Esto consolida nuestra afirmación de que los valores de estos nuevos estimadores para valores pequeños de  $k$  diferencian mejor la dificultad de las funciones que considerar únicamente la epistasis normalizada clásica.

$orden = 4$	$T_{64}^4$	$\mathfrak{R}_2^6$	Diferencia
$\varepsilon_1^*$	0.2134	0.3548	0.1414
$\varepsilon_2^*$	0.0714	0.1612	0.0898
$\varepsilon_3^*$	0.0102	0.0322	0.0220

Tabla C.4:  $k$ -epistasis de las funciones  $T_{64}^4$  y  $\mathfrak{R}_2^6$  ( $1 \leq k \leq 7$ ).

$orden = 8$	$T_{64}^8$	$\mathfrak{R}_3^6$	Diferencia
$\varepsilon_1^*$	0.8505	0.9391	0.0886
$\varepsilon_2^*$	0.6788	0.8327	0.1539
$\varepsilon_3^*$	0.4361	0.6197	0.1836
$\varepsilon_4^*$	0.2094	0.3536	0.1442
$\varepsilon_5^*$	0.0692	0.1406	0.0714
$\varepsilon_6^*$	0.0138	0.0342	0.0204
$\varepsilon_7^*$	0.0012	0.0038	0.0026

Tabla C.5:  $k$ -epistasis de las funciones  $T_{64}^8$  y  $\mathfrak{R}_3^6$  ( $1 \leq k \leq 7$ ).

$orden = 16$	$T_{64}^{16}$	$\mathfrak{R}_4^6$	diferencia
$\varepsilon_1^*$	0.9985	0.9996	0.0011
$\varepsilon_2^*$	0.9925	0.9978	0.0053
$\varepsilon_3^*$	0.9719	0.9893	0.0174
$\varepsilon_4^*$	0.9193	0.9615	0.0422
$\varepsilon_5^*$	0.8169	0.8949	0.0780
$\varepsilon_6^*$	0.6606	0.7727	0.1121
$\varepsilon_7^*$	0.4714	0.5981	0.1267
$\varepsilon_8^*$	0.2887	0.4017	0.1130
$\varepsilon_9^*$	0.1478	0.2272	0.0794
$\varepsilon_{10}^*$	$6.1626 \times 10^{-2}$	0.1050	$4.3374 \times 10^{-2}$
$\varepsilon_{11}^*$	$2.0269 \times 10^{-2}$	$3.8405 \times 10^{-2}$	$1.8136 \times 10^{-2}$
$\varepsilon_{12}^*$	$5.0435 \times 10^{-3}$	$1.0635 \times 10^{-2}$	$5.5915 \times 10^{-3}$
$\varepsilon_{13}^*$	$8.9014 \times 10^{-4}$	$2.0904 \times 10^{-3}$	$1.2003 \times 10^{-3}$
$\varepsilon_{14}^*$	$9.9118 \times 10^{-5}$	$2.5939 \times 10^{-4}$	$1.6027 \times 10^{-4}$
$\varepsilon_{15}^*$	$5.2280 \times 10^{-6}$	$1.5258 \times 10^{-5}$	$1.003 \times 10^{-5}$

Tabla C.6:  $k$ -epistasis de las funciones  $T_{64}^{16}$  y  $\mathfrak{R}_4^6$  ( $1 \leq k \leq 15$ ).



# Bibliografía

- [1] Beightler, C.S., Phillips, D.T. y Anderson, C.W. (1979), *Foundations of optimization*, 2<sup>a</sup> ed., Englewood Cliffs, NJ: Prentice Hall.
- [2] Bethke, A.D. (1980), *Genetic algorithms as function optimizers* (Tesis Doctoral, Universidad de Michigan) y *Dissertation Abstracts International*, **41** (9). University Microfilms No. 8106101.
- [3] Das, R. y Whitley, L.D. (1991), *The only challenging problems are deceptive: Global search by solving order-1 hyperplanes*, Ed. R.K. Belew y L.B. Booker, Morgan Kaufmann Publishers, *Proceedings of the Fourth International Conference on Genetic Algorithms*, 166–173, San Mateo, CA.
- [4] Davidor, Y. (1991), *Epistasis and Variance: A Viewpoint on GA-Hardness*, *Foundations of Genetic Algorithms 1*, 23–25. Ed. G.J.E. Rawlins, Morgan Kaufmann Publishers, San Mateo.
- [5] Davis, L. (1991), *The Handbook of Genetic Algorithms*, Van Nostrand Reinhold.
- [6] De Jong, K.A. (1990), *Introduction to the second special issue on genetic algorithms*, *Machine Learning*, **5** (4), 351–353.
- [7] Eshelman, L.J. y Schaffer, J.D. (1993), *Crossover' Niche*, S. Forrest, Ed., *Proceedings of the 5th International Conference on Genetic Algorithms*, 9–14. Morgan Kaufmann, San Francisco.
- [8] Forrest, S. y Mitchell, M. (1993), *Relative Building-Block Fitness and the Building-Block Hypothesis*, *Foundations of Genetic Algorithms*, **2**, Ed. L.D. Whitley, Morgan Kaufmann Publishers, San Mateo.
- [9] Forrest, S. y Mitchell, M. (1993), *What makes a problem hard for a Genetic Algorithm? Some anomalous results and their explanation*, *Machine Learning*, **13**, 285–319.
- [10] Freeman, J.A. y Skapura, D.M. (1993). *Redes Neuronales, Algoritmos, Aplicaciones y Técnicas de Programación*. Addison-Wesley.

- [11] Goldberg, D.E. (1987), *Simple Genetic Algorithms and The Minimal Deceptive Problem*, Genetic Algorithms and Simulated Annealing, Ed. L. Davis, Morgan Kaufmann Publishers.
- [12] Goldberg, D.E. (1989), *Genetic Algorithms and Walsh functions: Part I, A Gentle Introduction*, Complex Systems, **3**, 129–152.
- [13] Goldberg, D.E. (1989), *Genetic Algorithms in search, optimization and machine learning*. Reading, Mass., Addison–Wesley.
- [14] Goldberg, D.E. (1992), *Construction of High-order Deceptive Functions Using Low-order Walsh Coefficients*, Annals of Mathematics and Artificial Intelligence, **5**, 35–48.
- [15] Grefenstette, J.J. (1993), *Deception Considered Harmful*, Foundations of Genetic Algorithms, **2**, Ed. Whitley, L.D. Morgan Kaufmann Publishers 75–91.
- [16] Heckendorn, R.B. y Whitley, D. (1999), *Predicting epistasis from mathematical models*. Evolutionary Computation, **7** (1), 69–101.
- [17] Holland, J.H. (1998), *Adaptation in Natural and Artificial Systems*, 5<sup>a</sup> ed. Massachusetts Institute of Technology, MIT Press.
- [18] Iglesias, M.T. (1997), *Algoritmos Genéticos generalizados: variaciones sobre un tema*, Tesis Doctoral, Universidade da Coruña.
- [19] Iglesias, M.T., Suys, D., Vidal, C. y Verschoren, A. (1998), *Epistasis and Unitation*, Computers and Artificial Intelligence, **18** (5), 467–483.
- [20] Iglesias, M.T., Verschoren, A. y Vidal, C. (2000), *Template Functions and their Epistasis*, Proceedings of MS'2000 International Conference on Modelling and Simulation, 539–546, Las Palmas de Gran Canaria.
- [21] Iglesias, M.T., Peñaranda, V. S., Verschoren, A. y Vidal, C. (2003), *Computing epistasis of Template functions with the Hyperplane Averaging Theorem*, Encontro de Álgebra Linear e Aplicações (EALA'2003), Lisboa.
- [22] Iglesias, M.T., Peñaranda, V. S. y Verschoren, A. (2004), *Computing epistasis of Generalized Royal Road functions using Walsh transforms and the Hyperplane Averaging Theorem*, International Symposium on Symbolic and Algebraic Computation (ISAAC'2004), Santander.
- [23] Iglesias, M.T., Verschoren, A. y Vidal, C. (2005), *Computing Epistasis of Template Functions through Walsh Transforms*, Computing and Informatics, **3**.

- [24] Iglesias, M.T., Naudts, B., Verschoren, A. y Vidal, C. (2005), *A Combinatorial Approach to Epistasis*, Springer.
- [25] Iglesias, M.T., Peñaranda, V. S. y Verschoren, A. (2005), *Higher order functions and Walsh coefficients*, Aceptado para su publicación en: *Bulletin of the Belgian Mathematical Society Simon Stevin*.
- [26] Iglesias, M.T., Peñaranda, V. S. y Verschoren, A. (2005), *2-epistasis over binary alphabets*. Remitido a *Computers and Informatics*.
- [27] Iglesias, M.T., Peñaranda, V. S. y Verschoren, A. (2005), *Order and Deceptivity*. Remitido a *Bulletin of the Belgian Mathematical Society Simon Stevin*.
- [28] Kirkpatrick, S. Jr., Gelatt, C.D y Vecchi, M. (1983), *Optimization by simulated annealing*, *Science*, **220**, 671–680.
- [29] Koza, J. R. (1992), *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT press.
- [30] Michalewicz, Z. (1996), *Genetic Algorithms + Data Structures = Evolution Programs*, 3<sup>a</sup> ed. Springer-Verlag.
- [31] Mitchell, M., Forrest, S. y Holland, J. (1992), *The Royal Road for Genetic Algorithm: Fitness landscape and GA Performance*, F. J. Varela and P. Bourguine, editors, *Towards a Practice of Autonomous Systems: Proceedings of the First European Conference on Artificial Life*, 245–254, Cambridge, MA, USA. MIT Press. Santa Fe Institute. Working paper 91-10-046.
- [32] Mitchell, M. (1996), *An introduction to Genetic Algorithm*, MIT Press, Cambridge, MA, USA.
- [33] Naudts, B., Suys, D. y Verschoren, A. (1998), *Epistasis, Deceptivity and Walsh Transforms*, *Proceedings of the International ICSC Symposium on Engineering of Intelligent Systems (EIS'98)*, vol 1: *Fuzzy Logic/Genetic Algorithms*, ICSC Academic Press, 210–216.
- [34] Naudts, B., Suys, D. y Verschoren, A. (2000), *Generalized Royal Road functions and their Epistasis*, *Artificial Intelligence*, **19**, 317–334.
- [35] Rawlins, G.J.E. (1991), *Foundations of Genetic Algorithms*, Morgan Kaufmann Publishers.
- [36] Rich, E. y Knight, K. (1994), *Inteligencia Artificial*. McGraw Hill.

- [37] Rodríguez Pan, M<sup>a</sup> I. (1999), *Implementación del concepto de epistasis de una función*. Proyecto de fin de carrera de Ingeniería Técnica en Informática de Gestión. Departamento de Computación. Facultad de Informática. Universidade da Coruña.
- [38] Strickberger, M. W. (1988), *Genética*. Ediciones Omega S.A., 3<sup>a</sup> edición.
- [39] Suys, D. y Verschoren, A. (1996), *Extreme Epistasis*, International Conference on Intelligent Technologies in Human-Relates Sciences (ITHURS 96), **2**, 251–258, León.
- [40] Suys, D. (1999), *A Mathematical Approach to Epistasis*, Tesis doctoral, Universidad de Amberes.
- [41] Van Hove, H. y Verschoren, A. (1994), *On Epistasis*, Computers and AI, **14** (3), 271–277.
- [42] Van Hove, H. (1995), *Representational Issues in Genetic Algorithms*, Tesis doctoral, Universidad de Amberes.
- [43] Van Hove, H. y Verschoren, A. (1998), *What is Epistasis?*, Bull. Soc. Math. Belg., Simon Stevin, **5**, 69–77.
- [44] Van Lint, J.H. (1982), *Introduction to Coding Theory*, Springer-Verlag.
- [45] Vose, M.D. (1993), *A critical examination of the schema theorem*. Informe Técnico UT-CS-93212, Universidad de Tennessee, Computer Science Department, Knoxville, TN, USA.
- [46] Whitley, L.D. (1991), *Fundamental Principles of Deception in Genetic Search*, Foundations of Genetic Algorithms, **1**, Ed. G. J. E. Rawlins, Morgan Kaufmann Publishers, San Mateo.
- [47] Whitley, L.D. (1993), *Foundations of Genetic Algorithms*, **2**, Ed. G. J. E. Rawlins, Morgan Kaufmann Publishers, San Mateo.



