

Autor: Daniel Sebastián San Martín

Director: Luis Miguel Pardo Vasallo

Tesis Doctoral

**CONJUNTOS CONSTRUCTIBLES Y
CONJUNTOS CUESTORES:**

**TEORÍA DE LA INTERSECCIÓN Y PRIMERAS
INTERACCIONES CON LA TEORÍA DE
APRENDIZAJE COMPUTACIONAL**

PhD Thesis

**CONSTRUCTIBLE SETS AND CORRECT
TEST SEQUENCES:**

**INTERSECTION THEORY AND INITIAL
CONNECTIONS WITH COMPUTATIONAL
LEARNING THEORY**



Escuela de Doctorado

A mis padres, Antonio y Violeta, por haber estado siempre a mi lado.

Agradecimientos

Quiero comenzar expresando mi más sincero agradecimiento a mi director de tesis, **Luis Miguel Pardo**. A lo largo de estos años de doctorado, no solo ha sabido sacar lo mejor de mí, sino que me ha acompañado y apoyado en cada etapa del camino. Tras innumerables pizarras y conversaciones, me ha transmitido una forma única de entender las matemáticas que siempre llevaré conmigo. Es, sin duda, una inspiración y un modelo a seguir como científico. Solo puedo decir: gracias por todo, Luis Miguel.

También quiero dar las gracias a **Joos Heintz**, por ser una fuente inagotable de inspiración matemática y por las enriquecedoras conversaciones matemáticas que compartimos durante sus visitas a Santander en sus últimos años. Del mismo modo, agradezco a **Marc Giusti** las interesantes discusiones que mantuvimos al inicio de mi doctorado.

Mi gratitud se extiende a todos **los miembros del Departamento de Matemáticas, Estadística y Computación de la Universidad de Cantabria**, quienes durante varios años me acogieron como uno más. En especial, a **Javier**, por aceptar ser mi tutor de tesis y acompañarme en mi primera experiencia docente. Por otro lado, quisiera dar las gracias a **Fátima** por todos los cafés, comidas y charlas compartidas a lo largo de estos años y, sobre todo, por su amistad incondicional.

Finalmente, quiero expresar mi más profundo agradecimiento a **mi familia**. Estos últimos años han sido especialmente difíciles para todos, pero siempre hemos estado unidos y nos hemos apoyado mutuamente. En primer lugar, me gustaría dar las gracias a mi madre, **Violeta**, por darme todo sin pedir nunca nada a cambio. También quiero dedicar unas palabras a mi padre, **Antonio**, y a mi abuelo, **Jaime**, quienes, aunque ya no están con nosotros, permanecen vivos en nuestra memoria y en nuestro corazón. **Fueron el mejor padre y el mejor abuelo que uno podría desear**, y estoy seguro de que estarían muy orgullosos de esta tesis. Igualmente, quiero dar las gracias a mi abuela **Luisa**, a mis tíos **Javi, Madi, Luisina, Chechu, Jorge e Isabel**, y a mis primos **Diego, David, Mateo y Sara**, por su cariño, su apoyo constante y por estar siempre presentes.

Contenidos

Capítulo 0. Contexto y Resumen de Contenidos de la Memoria	i
0.1. El contexto de las preguntas que motivan esta memoria	i
0.2. Los conjuntos cuestores y una primera aproximación a los principales resultados	v
0.3. Resumen pormenorizado de los contenidos	ix
0.3.1. Capítulo 2: Desigualdad de Bézout para conjuntos constructibles	x
0.3.2. Capítulo 3: Conjuntos Cuestores	xii
0.3.3. Capítulo 4: Erzeugungsgrad y dimensión de Vapnik-Chervonenkis	xxi
0.3.4. Capítulo 5: Redes Neuronales con función de activación racional y conjuntos constructibles de parámetros	xx
0.3.5. Capítulo 6: Un enfoque alternativo a la relación entre aprendizaje y polinomios multivariados: Reescritura y Aprendizaje multiclase. Grado de salida promedio, dimensiones, shifting y relaciones con bases monomiales de K -álgebras de Artin	xxii
Capítulo 1. Nociones y notaciones básicas. Resultados preliminares	1
1.1. Conjuntos localmente cerrados y constructibles en la topología de Zariski	1
1.2. Grado de conjuntos localmente cerrados de acuerdo a [He, 83] y su Desigualdad de Bézout	5
1.3. Teoría de Vapnik-Chervonenkis	7
Capítulo 2. Desigualdad de Bézout para conjuntos constructibles	9
2.1. Dos nociones de grado y dos Desigualdades de Bézout para conjuntos constructibles	9
2.2. Cotas superiores para la intersección de varios conjuntos constructibles	17
2.2.1. Una extensión directa de la Proposición 2.3 de [HS, 82] a conjuntos localmente cerrados	18
2.2.2. Demostración del Ítem <i>i</i>) del Teorema 2.2.1	22
2.2.3. Demostración del Ítem <i>ii</i>) del Teorema 2.2.1	25
2.2.4. Dos aplicaciones inmediatas de las cotas superiores previas	28
2.3. Una observación sobre las cotas de las imágenes de conjuntos constructibles por aplicaciones lineales	30
2.3.1. Una primera cota para el defecto	30
2.3.2. Una cota extrínseca para el grado de una proyección basada en el Nullstellensatz Efectivo	34
Capítulo 3. Conjuntos Cuestores	39
3.1. La noción, algunas nociones equivalentes y primeras propiedades	39
3.1.1. El caso de polinomios y listas de polinomios	44
3.2. Alta probabilidad de existencia de Conjuntos Cuestores de longitud asintóticamente óptima en conjuntos localmente cerrados equidimensionales para conjuntos constructibles de listas de polinomios	45
3.2.1. Demostración del Teorema 3.2.1	46
3.2.2. Algunos Corolarios útiles	49
3.2.3. Conjuntos constructibles unirracionales	53
3.2.4. El conjunto de muestreo de DeMillo–Lipton–Schwartz–Zippel	54
3.3. Algunos Conjuntos Cuestores en la literatura matemática	54
3.3.1. Conjuntos Cuestores y variedades determinantaes	54
3.3.2. Conjuntos Cuestores y Conjuntos de Kakeya	55
3.3.3. Conjuntos Cuestores, Nullstellensatz Combinatorio de Alon y dualidad	59
3.4. Un algoritmo probabilista para detectar la dimensión genérica: “Suite Sécante”	64
3.4.1. Algunos aspectos básicos sobre las “Suites Sécantes”	64
3.4.2. El Algoritmo	68
3.4.3. Clase de todas las listas de polinomios evaluables por un esquema de evaluación	71

Capítulo 4. Erzeugungsgrad y dimensión de Vapnik-Chervonenkis	75
4.1. Erzeugungsgrad: revisando la Sección 3 de [He, 83] combinada con conjuntos constructibles	76
4.1.1. Definiciones y propiedades básicas	76
4.1.2. Cotas combinatorias	84
4.2. Teoría de Vapnik-Chervonenkis sobre cuerpos algebraicamente cerrados	91
4.2.1. Dimensión VC y clasificadores constructibles	91
4.2.2. Dimensión VC de abiertos distinguidos definidos por una clase de polinomios	93
4.2.3. Variedades evasivas	94
4.2.4. Los conjuntos cuestores están densamente distribuidos en variedades evasivas de dimensión positiva	98
Capítulo 5. Redes Neuronales con función de activación racional y conjuntos constructibles de parámetros	101
5.1. Sintaxis de una Red Neuronal	102
5.2. Semántica de una Red Neuronal: instanciación de parámetros	103
5.3. Ejemplos de funciones de activación: ReLUs, esquemas de evaluación, polinomiales, racionales, de Nash y de Pfaff	104
5.4. Propiedades del conjunto constructible $\mathcal{W}(\mathcal{N}, \Lambda)$ en el caso de funciones de activación polinómicas	106
5.5. Redes neuronales con función de activación racional que evalúan funciones racionales	108
Capítulo 6. Un enfoque alternativo a la relación entre aprendizaje y polinomios multivariados: Reescritura y Aprendizaje multiclasa. Grado de salida promedio, dimensiones, shifting y relaciones con bases monomiales de K -álgebras de Artin	115
6.1. Nociones básicas: Pre-orden de Dickson y bases monomiales	115
6.1.1. Pre-orden de Dickson y clases de conceptos cerradas hacia abajo: escalones y escaleras	116
6.1.2. Órdenes monomiales y Bases de Gröbner	118
6.2. El hiper-grafo de una inclusión: grado promedio, grado de salida promedio y orientaciones	122
6.2.1. Grado promedio, grado de salida promedio y densidad del hiper-grafo de una inclusión	124
6.2.2. Grado promedio, grado de salida promedio y conjuntos cerrados hacia abajo	137
6.3. Pseudo-cubos, dimensión DS y dimensión exponencial	139
6.3.1. Pseudo-cubos: definición, caracterización y propiedades básicas	139
6.3.2. Dimensiones	141
6.3.3. Pseudo-dimensión DS	144
6.3.4. El escalón \mathcal{E}_h y sus dimensiones	145
6.3.5. Pseudo-cubos cerrados hacia abajo	146
6.4. La técnica del “shifting” en el Aprendizaje multiclasa	150
6.4.1. La técnica del shifting	150
6.4.2. Shifting, clases cerradas hacia abajo y bases monomiales	152
6.4.3. Shifting e invariantes de clases de conceptos finitas	159
Bibliografía	171

Contexto y Resumen de Contenidos de la Memoria

Índice

0.1. El contexto de las preguntas que motivan esta memoria	i
0.2. Los conjuntos cuestores y una primera aproximación a los principales resultados	v
0.3. Resumen pormenorizado de los contenidos	ix
0.3.1. Capítulo 2: Desigualdad de Bézout para conjuntos constructibles	x
0.3.2. Capítulo 3: Conjuntos Cuestores	xii
0.3.3. Capítulo 4: Erzeugungsgrad y dimensión de Vapnik-Chervonenkis	xvi
0.3.4. Capítulo 5: Redes Neuronales con función de activación racional y conjuntos constructibles de parámetros	xx
0.3.5. Capítulo 6: Un enfoque alternativo a la relación entre aprendizaje y polinomios multivariados: Reescritura y Aprendizaje multiclasa. Grado de salida promedio, dimensiones, shifting y relaciones con bases monomiales de K -álgebras de Artin	xxii

0.1. El contexto de las preguntas que motivan esta memoria

Esta memoria es un primer sondeo preliminar de las potenciales interacciones entre la Geometría Algebraica Computacional (llamada, desde el siglo XIX, Teoría de la Eliminación) y el Aprendizaje Computacional. Para comprender las motivaciones de este estudio hay que remontarse a la última década del siglo XX y a los desarrollos de algoritmos eficientes intrínsecos para la resolución de sistemas de ecuaciones polinomiales multivariadas dentro de la corriente TERA (Turbo Evaluation and Rapid Algorithms).

El problema esencial de la Geometría Algebraica Computacional es su falta de eficacia, es decir, los excesivos requerimientos de complejidad computacional (tanto en tiempo como en espacio) exigidos por todos los algoritmos conocidos en el tratamiento de sus problemas fundacionales. Para explicar esta dificultad, fijemos dos problemas centrales:

PROBLEMA 1 (Nullstellensatz de Hilbert). *Dada una familia de polinomios $f_1, \dots, f_s, g \in \kappa[X_1, \dots, X_n]$, donde κ es un cuerpo primo con clausura algebraica K , sea $V_{\mathbb{A}} := V(f_1, \dots, f_s) \subseteq \mathbb{A}^n(K)$ la variedad algebraica de los ceros comunes de los primeros s polinomios, que supondremos no vacía, i.e.*

$$V := V_{\mathbb{A}}(f_1, \dots, f_s) := \{x \in \mathbb{A}^n(K) : f_1(x) = \dots = f_s(x) = 0\} \neq \emptyset.$$

Nos preguntamos si g se anula en algún punto de V , es decir, tratamos de decidir (algorítmicamente) si:

$$\exists x \in \mathbb{A}^n(K), x \in V \wedge g(x) = 0.$$

PROBLEMA 2 (Polinomio Eliminante). *Con las mismas notaciones para κ y K , dada una familia de polinomios $f_1, \dots, f_s, g \in \kappa[X_1, \dots, X_n]$, consideramos la variedad algebraica $V := V_{\mathbb{A}}(f_1, \dots, f_s) \subseteq \mathbb{A}^n(K)$ de los ceros comunes de los s primeros polinomios. Supongamos que V es no vacía, equidimensional y de dimensión de Krull r . Supongamos, además, que las variables están en posición de Noether con respecto a V , es decir, que la siguiente es una extensión entera de anillos:*

$$\kappa[X_1, \dots, X_n] \hookrightarrow \kappa[V] := \kappa[X_1, \dots, X_n]/I(V),$$

donde $I(V) = \{h \in \kappa[X_1, \dots, X_n] : h|_V \equiv 0\} = \sqrt{(f_1, \dots, f_s)}$. Consideramos la siguiente aplicación polinomial:

$$\begin{aligned} \pi_g : \mathbb{A}^n(K) &\longrightarrow \mathbb{A}^{r+1}(K) \\ (x_1, \dots, x_n) &\longmapsto (x_1, \dots, x_r, g(x_1, \dots, x_n)) \end{aligned}$$

Entonces, la imagen $\pi_g(V) \subseteq \mathbb{A}^{r+1}(K)$ es una hipersuperficie definida por un único polinomio:

$$\chi_g \in \kappa[X_1, \dots, X_r, T],$$

es decir,

$$\pi_g(V) = V_{\mathbb{A}}(\chi_g).$$

Hallar el polinomio χ_g (llamado *polinomio eliminante*).

Ambos problemas presentan enormes dificultades en términos de complejidad computacional. El Nullstellensatz de Hilbert (**HN**), llamado así a partir de [BSS, 89], es un problema \mathbf{NP}_C -completo en la Teoría de las Máquinas BSS (cf. [BSS, 89]). Además, como se observa en [Pa, 95] y [Pa, 12], todos los problemas \mathbf{NP} -completos de la Teoría Clásica de Complejidad son instancias particulares del problema **HN**. Aunque las (discutibles) máquinas BSS solo sirven para dar un contexto teórico a problemas sobre alfabetos continuos, las dos consideraciones precedentes indican la dificultad intrínseca de encontrar un algoritmo eficiente para **HN**: sería tanto como resolver afirmativamente la *Conjetura de Cook*, uno de los Millenium Problems del Instituto Clay y también uno de los problemas de la Lista de Steve Smale para el siglo XXI.

En cuanto al segundo problema, la dificultad para encontrar un algoritmo eficiente también es enorme. Para comenzar, utilizando un poco de la Teoría de Intersección (en el sentido de [He, 83]), se observa que el grado del polinomio χ_g es muy alto. Aunque pueden existir casos defectivos, el grado usual de χ_g verifica:

$$\deg(\chi_g) \leq \deg(V) \cdot \deg(g).$$

Por lo tanto, escribir este polinomio en la forma densa usual (i.e. expresar χ_g como la suma de todos sus términos, tengan o no coeficiente nulo) da lugar a un objeto de tamaño:

$$\binom{\deg(V) \cdot \deg(g) + r + 1}{r + 1} \approx (\deg(V) \cdot \deg(g))^{r+1}.$$

Como, además, $\deg(V)$ tiende a ser una cantidad próxima a exponencial en el número de variables, representar χ_g llevaría a una expresión de tamaño exponencial y, por consiguiente, intratable en términos de complejidad.

Ambos problemas están, de hecho, muy relacionados entre sí. Por ejemplo, si $V := V_{\mathbb{A}}(f_1, \dots, f_s)$ es una variedad cero-dimensional (i.e. un conjunto finito de puntos), g se anula en algún punto de V (pregunta de **HN**) si y solo si el polinomio eliminante $\chi_g \in \kappa[T]$ se anula en $t = 0$ (i.e. tiene término independiente nulo).

Sin embargo, el propósito de esta memoria no es estudiar la interacción entre estos dos problemas, que hasta la fecha siguen siendo intratables.

Podemos decir que, en la última década del siglo XX, la corriente TERA se propuso llevar al límite la complejidad de estos problemas (y otros relacionados) utilizando varios elementos:

- i) En primer lugar, se discutió la idea de *solving*, esto es, se buscaron formas alternativas de representar la información relativa a la variedad $V = V_{\mathbb{A}}(f_1, \dots, f_s)$. A partir de esta discusión surgen varias ideas que, esencialmente, toman como concepto de *solución* una descripción de V mediante un isomorfismo birracional con alguna hipersuperficie (por ejemplo, una solución “à la Kronecker”).
- ii) Para representar la hipersuperficie mencionada en el punto anterior (mediante su polinomio eliminante), es necesario poner las variables en posición de Noether y representar todos los polinomios multivariados involucrados mediante una estructura de datos denominada *straight-line program*.
- iii) Jugar de forma adecuada con la dualidad entre *solving* y *eliminating*: la tradición de Teoría de la Eliminación supone “resolver primero” y “eliminar después” (lo que está implícito, por ejemplo, en la Conjetura de Cook). Sin embargo, la corriente TERA propuso un cambio de paradigma basado en los Puntos *i*) y *ii*) anteriores: “eliminar para resolver, para volver a eliminar” de manera iterativa.

Esta serie de ideas se inicia con [GH, 91], se continúa en [GHS, 93] y se culmina en [KP, 94] y [KP, 96]. Posteriores resultados extenderán estas ideas (como [GHSSS, 95]). Conviene destacar especialmente las reflexiones sobre el Punto *ii*) precedente realizadas en la Sección 3 de [KP, 96]. En dicho trabajo se desarrolla una variante de los *straight-line programs*, con especial énfasis en el grafo subyacente y en el paralelismo. Esta variante es una estructura de datos que recibe el nombre de *esquema de evaluación* (o *evaluation scheme*). A la sazón, las nociones de redes neuronales no eran ni tan populares ni tan conocidas fuera de ámbitos muy restringidos dentro del Aprendizaje Computacional. Sin embargo, los esquemas de evaluación definidos en [KP, 96] resultan ser, en esencia, las “*neural networks with activation function* $\varphi(t) = t^2$ ”, que se discutirán en detalle en el Capítulo 5 de esta memoria. Entre otros elementos innovadores, encontramos en [KP, 96] una primera conexión entre Geometría Algebraica Computacional y Teoría del Aprendizaje:

COROLARIO 0.1.1 ([KP, 96]). *Con las notaciones e hipótesis del Problema 2, concerniente al Polinomio Eliminante, y suponiendo que d es una cota superior de los grados de los polinomios $f_1, \dots, f_s, g \in \kappa[X_1, \dots, X_n]$ de la lista de entrada, tenemos que:*

- i) $\deg(\chi_g) \leq d^{n+1}$.
- ii) *Existe una red neuronal con función de activación $\varphi(t) = t^2$ que codifica χ_g con:*
 - *talla acotada por $d^{O(n)}$,*
 - *profundidad (número de multicapas) en $O(n \log_2(d))$.*

En otras palabras, los polinomios eliminantes admiten una representación como redes neuronales (con función de activación $\varphi(t) = t^2$) de talla polinomial en su grado y de profundidad (número de multicapas) logarítmica en dicho grado. Esto constituye la primera aproximación entre Geometría Algebraica Computacional y ciertas nociones típicas de Aprendizaje Computacional. Desde un punto de vista conceptual, el resultado anterior puede considerarse el punto de partida de esta memoria.

Posteriormente, la corriente TERA siguió profundizando en los Puntos *i*), *ii*) y *iii*) mencionados previamente. De este enfoque surgen los algoritmos de tipo intrínseco que se desarrollan en la serie de trabajos [Pa, 95], [GHMP, 95], [GHMP, 97] y [GHHMMP, 97]. El uso de estructuras de datos tipo esquema de evaluación (i.e. redes neuronales con función de activación $\varphi(t) = t^2$) continuará hasta [HMPS, 00]. Con el tiempo, la corriente TERA pasó a denominarse TERA-Kronecker o, simplemente, Kronecker, con desarrollos ulteriores que evitaron el uso de esta estructura de datos con técnicas de eliminación restringidas a dimensión uno (esto es, curvas). Estos desarrollos corresponden a [GLS, 01] y [HMW, 01] (véase también [GHMPPP, 25] y sus referencias).

En la secuencia de trabajos de la corriente TERA se produce un salto cualitativo muy relevante entre la primera serie de trabajos, que va de [GH, 91] a [KP, 96], y la segunda, que abarca desde [Pa, 95] hasta [GHMMP, 98]. Entre ambas series de trabajos se introducen numerosas variaciones conceptuales y elementos técnicos que cambian tanto la perspectiva como el diseño de los algoritmos, introduciendo una dependencia de elementos semánticos intrínsecos. Entre las novedades de la segunda serie de trabajos tiene un papel relevante el uso de una técnica clásica: *el operador de Newton no Arquimediano*. A partir de una variedad cero-dimensional construida ad-hoc (denominada *lifting fiber* en [GHHMMP, 97]), se reconstruye información que describe variedades de intersección completa de dimensión positiva. No entraremos en el detalle de la descripción de este proceso, pero sí diremos que interviene un esquema de evaluación con divisiones con una profundidad extremadamente pequeña. En términos de la Teoría del Aprendizaje, los algoritmos de la serie que va de [Pa, 95] a [GHMMP, 98] emplean redes neuronales con función de activación racional inevitable cuya talla es $d^{O(n)}$ y que acumula un número de multicapas del orden $O(n \log_2(d))$. Estas redes neuronales se utilizan para modelizar el operador de Newton-Hensel en el anillo de series de potencias formales y constituyen un ejemplo del tipo de redes neuronales que se analizan en el Capítulo 5 de esta memoria. Cabe destacar que este tipo de redes neuronales sigue siendo de uso común en posteriores evoluciones de la corriente TERA-Kronecker como en [GLS, 01] y [HMW, 01], así como en desarrollos más recientes (cf. [GHMPPP, 25] y sus referencias).

Todo el esfuerzo intelectual que suponen estos avances lleva la algorítmica hasta una barrera de tipo semántico: *el número de Bézout*. Así, suponiendo en el Problema 2, relativo al Polinomio Eliminante, que $s = n$, $\deg(f_i) = d_i$, $1 \leq i \leq n$, y que $V = V_{\mathbb{A}}(f_1, \dots, f_n)$ es una variedad cero-dimensional, el grado de V satisface, genéricamente, la siguiente igualdad:

$$\deg(V) = \prod_{i=1}^n d_i = \mathcal{D}.$$

La cantidad \mathcal{D} anterior es el llamado número de Bézout del sistema $\{f_1, \dots, f_n\}$. Los resultados de la corriente TERA y TERA-Kronecker establecen que tanto **HN** como el cálculo del polinomio eliminante χ_g pueden hacerse en tiempo polinomial en \mathcal{D} . Como en la conferencia de L. M. Pardo *Symbolic solving: Getting closer to the barrier?* (MSRI, 1998), nos preguntamos si podemos superar esa barrera.

En [CGHMP, 03] se demuestra que la barrera del número de Bézout es infranqueable si se pretenden obtener algoritmos de resolución *universales*. Un algoritmo de resolución es universal si permite responder de manera continua en los coeficientes de los polinomios de entrada a cualquier pregunta de eliminación. De hecho, en [CGHMP, 03] se prueba que existen polinomios eliminantes tales que, si se codifican por cualquier medio que dependa continuamente de los coeficientes de los polinomios de entrada, su representación requiere una dimensión exponencial. En otras palabras, manteniendo la forma usual de diseñar algoritmos en Geometría Algebraica Computacional, el número de Bézout será una barrera infranqueable de la complejidad computacional.

Con posterioridad, surgen algunos intentos de búsqueda de algoritmos no universales que puedan evitar la barrera del número de Bézout. Un buen ejemplo es el trabajo [PSM, 04] para Sistemas de Pham Generalizados. Sin embargo, la buena filosofía para los algoritmos no universales no viene del ámbito simbólico-algebraico sino del ámbito numérico. La pregunta conocida como *Smale's 17th Problem*, de la Lista de Steve Smale para el siglo XXI, plantea precisamente la cuestión sobre la existencia de algoritmos numéricos eficientes (i.e. polinomiales en tiempo) para resolver sistemas de ecuaciones polinomiales multivariadas. En [BP, 09] y [BP, 11] se resolvió afirmativamente esta pregunta: se mostraron dos algoritmos tipo Las Vegas que, en tiempo polinomial en codificación densa de los polinomios dados como entrada, permiten obtener una buena aproximación de una de sus soluciones. Más aún, el algoritmo [BP, 11] optimiza la entropía de tal modo que cualquiera de las soluciones del sistema puede ser alcanzada. Esta propiedad no será posible en los algoritmos deterministas (dada una lista

de polinomios de entrada, el algoritmo siempre aproxima la misma solución entre las \mathcal{D} soluciones del sistema genérico dado).

En cualquier caso, tanto con algoritmos probabilistas como con algoritmos deterministas, si los métodos numéricos pretenden aproximar *toda la variedad* V , también necesitarán una complejidad mayor al número de Bézout \mathcal{D} . En otras palabras, los algoritmos no universales pueden sortear la barrera impuesta por el número de Bézout, pero si se pretende obtener una información completa (universal) sobre todas las soluciones, el número de Bézout \mathcal{D} reaparece (véase [Pa, 00] para una discusión).

Con todo lo expuesto hasta aquí, el lector podría llegar a la conclusión de que esta línea de investigación había llegado, en el primer decenio del presente siglo, a un callejón sin salida. Sin embargo, aún no está todo dicho. Ya en [Pa, 00] y [CGHMP, 03] se introduce una posibilidad aparentemente remota pero no imposible. Al fin y al cabo, para el científico lo imposible no existe, salvo que se pruebe de manera irrefutable. En esos trabajos se prueba que existe una forma alternativa de codificar los polinomios eliminantes de forma muy sucinta. Retomando las notaciones del Problema 2, relativo al Polinomio Eliminante, y suponiendo, por ejemplo, que todos los polinomios involucrados $f_1, \dots, f_s, g \in K[X_1, \dots, X_n]$ tienen grado acotado por 3 (i.e. ecuaciones cúbicas), se obtiene que el polinomio eliminante χ_g tendría grado $3^{O(n)}$. Además, las redes neuronales con función de activación polinomial que lo codifican también tendrán talla $3^{O(n)}$. Sin embargo, existe una representación biunívoca (i.e. mediante una biyección) que permite describir χ_g usando información explícita acotada por $O(n^4)$. Esto es, los polinomios eliminantes admiten representaciones cortas y biunívocas, pero, seguramente, no permitirán responder a preguntas mediante mecanismos continuos. Esta nueva codificación se basa en los valores del polinomio eliminante en un conjunto de puntos especial, que recibe el nombre de *conjunto cuestor* (o *correct test sequence*).

La clave de lo descrito en [CGHMP, 03] pasa por disponer de χ_g mediante sus valores en un *conjunto cuestor* cuya longitud es lineal en la dimensión de la variedad de polinomios eliminantes y no en su grado. En lo que respecta a la posibilidad de utilizar representaciones más habituales, como redes neuronales de profundidad acotada, en [HOPRS, 23] se demuestra que dichas redes deben tener una talla comparable al número de Bézout, incluso cuando se admite el uso de funciones de activación de clase \mathcal{C}^1 .

Por otra parte, las reflexiones realizadas en [CGHMP, 03] nos llevan a preguntarnos cómo obtener información (posiblemente incompleta) a partir de la codificación de χ_g sobre un conjunto cuestor. Esta cuestión parece estar ligada de forma natural con el Aprendizaje Computacional: dada la lista $(x_1, y_1), \dots, (x_L, y_L)$ formada por un conjunto cuestor $\{x_1, \dots, x_L\}$ y los valores $y_i = \chi_g(x_i)$, ¿qué clase de preguntas sobre el polinomio eliminante χ_g puedo responder con métodos de Aprendizaje Computacional?

Esta última pregunta marca el punto de partida de esta memoria. Sin embargo, el propósito de esta memoria no es cerrar la respuesta. El punto crítico es que *hay una enorme tarea preliminar necesaria antes de afrontar la pregunta precedente con garantías de pisar el terreno apropiado*. Esa “enorme tarea preliminar” comienza con esta memoria y se continuará en desarrollos futuros de la investigación que aquí se inicia.

Para afrontar esa “enorme tarea preliminar” hemos tratado de centrarnos en algunos aspectos esenciales que se describen a continuación:

- i) La Teoría de la Intersección (i.e. la teoría del grado y la desigualdad de Bézout) para *conjuntos constructibles* está incompleta desde [He, 83] y [He, 85]. En esta memoria se cierra dicha teoría, aportando no solo una, sino dos nociones de grado para conjuntos constructibles que satisfacen la Desigualdad de Bézout (cf. Capítulo 2).
- ii) La idea de conjunto cuestor nace en [HS, 82] y se profundiza en la Sección 3 de [KP, 96] (véase también [Pa, 95]). El Capítulo 3 está destinado a llevar al límite las ideas y la metodología de [HS, 82] para obtener cotas finas de la presencia de conjuntos cuestores a partir de la Teoría de la Intersección de conjuntos constructibles desarrollada en el Capítulo 2.
- iii) En el Capítulo 4 comenzamos a establecer algunas conexiones entre familias de conjuntos constructibles, conjuntos cuestores y los rudimentos de la Teoría del Aprendizaje en el contexto de la Teoría de Vapnik-Chervonenkis (i.e. el aprendizaje en el caso binario). Para ello, hacemos un desarrollo de la noción de Erzeugungsgrad (o “grado de generación”) de familias de conjuntos constructibles, introducida en [He, 83] para familias de conjuntos localmente cerrados y olvidada en el tiempo. Usando la noción previa, probamos que la dimensión de Vapnik-Chervonenkis de una familia de conjuntos constructibles está linealmente acotada por la dimensión de Krull del espacio que parametriza dicha familia. Por otro lado, en este capítulo, también estudiamos la probabilidad de encontrar conjuntos cuestores de longitud apropiada en variedades evasivas de dimensión positiva.
- iv) Por último, el Capítulo 6 presenta una primera aproximación al *Aprendizaje multiclase*. Curiosamente, este tipo de aprendizaje está estrechamente vinculado al estudio de polinomios presentados en la forma

de *fewnomials*, al estilo de Khovanskii (cf. [Kh, 91]), dados por sus términos con coeficiente no nulo. Aunque los resultados descritos en este capítulo aún tienen consecuencias por explorar, destacamos la introducción de una nueva noción: el *grado de salida promedio*. En este capítulo probamos que dicha noción es un invariante central, ligado a todas las nociones de dimensión en Aprendizaje multiclase descritas en [BCDMY, 22]. Además, llevamos a cabo el primer estudio sistemático de la noción de pseudo-cubo definida en [BCDMY, 22].

Por supuesto, esta memoria contiene muchos más aspectos y desarrollos relevantes que requieren explicaciones (y notaciones) más precisas. Por esta razón, incluimos a continuación la Sección 0.2, en la que ofrecemos una visión general de los principales resultados, así como la Sección 0.3, que proporciona una revisión detallada de los distintos capítulos.

0.2. Los conjuntos cuestores y una primera aproximación a los principales resultados

En [HS, 82], para paliar las dificultades de los *Tests de Nulidad probabilísticos de DeMillo–Lipton–Schwartz–Zippel* (cf. [DML, 78], [Zp, 79] y [Sch, 80]), se definen los *conjuntos cuestores*. Este concepto se inspira en la demostración realizada por L. Adleman (cf. [Ad, 78]) de la inclusión $\mathbf{BPP} \subseteq \mathbf{P/poly}$.

DEFINICIÓN 1 (Conjuntos cuestores en el sentido de [HS, 82]). *Sea K un cuerpo algebraicamente cerrado. Sea Ω una variedad algebraica de polinomios en $K[X_1, \dots, X_n]$ de dimensión de Krull finita. Decimos que un conjunto de puntos $\{x_1, \dots, x_L\} \subseteq \mathbb{A}^n(K)$ es un conjunto cuestor para Ω si:*

$$\forall f \in \Omega, f(x_1) = \dots = f(x_L) = 0 \Rightarrow f \equiv 0,$$

Un ejemplo básico de conjuntos cuestores es el siguiente: si $\Omega := \{f \in K[X] : \deg(f) \leq d\}$, entonces, por el Teorema Fundamental del Álgebra, cualquier conjunto de $d + 1$ puntos distintos es un conjunto cuestor para Ω . Presentamos, a continuación, sin entrar en demasiados detalles, el resultado principal de [HS, 82]:

TEOREMA 0.2.1 (Teorema principal de [HS, 82]). *Sea K un cuerpo algebraicamente cerrado. Supongamos que $\text{char}(K) = 0$. Sea Ω la clausura Zariski de un conjunto constructible parametrizable de polinomios (por ejemplo, un esquema de evaluación o una red neuronal con función de activación polinomial) de grado acotado por d . Consideramos $\{1, \dots, u\} \subseteq \mathbb{Z}$ un conjunto de números enteros y $\Lambda := \{1, \dots, u\}^n \subseteq \mathbb{A}^n(K)$ un retículo, con $u \geq 2d^2$. Sea $L \geq 6 \dim(\Omega)$, donde $\dim(\Omega)$ es la dimensión de Krull de Ω . Entonces:*

- **Existencia:** existen conjuntos cuestores $\mathbf{Q} := (x_1, \dots, x_L) \in \Lambda^L$ para Ω .
- **Densidad:**

$$\text{Prob}_{(x_1, \dots, x_L) \in \Lambda^L} [(x_1, \dots, x_L) \text{ es un conjunto cuestor para } \Omega] \geq 1 - \frac{1}{u^{L/6}}.$$

donde la probabilidad considerada en Λ^L es la uniforme.

Las herramientas principales que utilizaron estos autores para demostrar la existencia y la densidad de los conjuntos cuestores son la Teoría de la Intersección, la noción de grado y la Desigualdad de Bézout para conjuntos localmente cerrados de [He, 83]. Es evidente que el teorema precedente presenta varias limitaciones: $\text{char}(K) = 0$, Ω es la clausura Zariski de un conjunto constructible parametrizable o Λ es un retículo.

Por otro lado, los conjuntos cuestores pueden verse como una técnica “similar” a los *Tests de Nulidad probabilísticos de DeMillo–Lipton–Schwartz–Zippel* (cf. [DML, 78], [Zp, 79] y [Sch, 80]). Sin embargo, estos dos enfoques presentan algunas sutiles diferencias que describimos a continuación:

- Los conjuntos cuestores son, por definición, “correctos” para toda la clase de entradas considerada: la probabilidad de error en sus respuestas es siempre cero, independientemente de la instancia concreta de entrada.
- Los Tests probabilísticos de DeMillo–Lipton–Schwartz–Zippel se basan en un conjunto de muestreo fácil de construir, que en sí mismo es un conjunto cuestor. En esta clase tan amplia, cada instancia de entrada posee alguna muestra precisa donde no se anula, pero diferentes entradas requieren distintos puntos de muestreo adecuados. Acotando la probabilidad de error de una elección aleatoria dentro del conjunto de muestreo, obtenemos el algoritmo probabilista.
- Los conjuntos cuestores tienen una longitud del mismo orden asintótico que la dimensión de Krull del conjunto de entrada considerado, mientras que los Tests probabilísticos de DeMillo–Lipton–Schwartz–Zippel requieren de un conjunto de muestreo de tamaño exponencial en el número de variables para obtener una probabilidad de error nula.
- No se conoce ningún algoritmo probabilista “eficiente” para generar conjuntos cuestores (ver [HS, 82], [Sax, 09], [Sax, 14] y sus referencias). Por lo tanto, su existencia (y generación) se basa en un resultado que demuestra que son altamente densos en términos probabilísticos.

- Bajo algunas hipótesis no muy complicadas, podemos ver que los conjuntos cuestores son altamente densos dentro de los conjuntos de muestreo (de tamaño exponencial) de DeMillo–Lipton–Schwartz–Zippel.

Estas sutiles diferencias requieren explicaciones adicionales y nos conducen a la siguiente pregunta: *¿Qué son los conjuntos cuestores y qué puedo hacer con ellos?* Por lo tanto, el primer objetivo de esta memoria es comprender el significado, la potencia y las limitaciones de los *conjuntos cuestores*. En particular, nos centramos en generalizaciones de la noción en sí misma: buscamos resultados fundacionales, tratamos de ver cómo esta noción subyace a otras presentes en la literatura matemática existente y, al igual que en el manuscrito original [HS, 82], demostramos no solo su existencia sino también resultados probabilísticos.

Desafortunadamente, para llegar a nuestro objetivo nos encontramos con un problema en el camino. Para trabajar con los conjuntos cuestores en el caso de cualquier conjunto constructible de polinomios necesitamos definir y saber cómo utilizar la noción de *grado para conjuntos constructibles*. Hasta donde sabemos, no existen antecedentes de una *Desigualdad de Bézout*. Los conjuntos constructibles son objetos geométricos que surgen de manera natural como proyecciones de variedades algebraicas (ver, por ejemplo, [Ch, 64-65]) y, por lo tanto, son objetos naturales en la Teoría de la Eliminación o en la Geometría Algebraica Computacional. Sin embargo, la noción de grado de estos objetos geométricos no ha sido correctamente formulada hasta el presente trabajo. El estudio de esta noción fue iniciado en [He, 83], donde J. Heintz definió una primera noción de grado de un conjunto constructible $C \subseteq \mathbb{A}^n(K)$ como el grado de su clausura Zariski (que en esta memoria denotaremos por $\deg_z(C)$). Lamentablemente, la noción de grado de J. Heintz no satisface una Desigualdad de Bézout (ver La Croix de Berny en el Ejemplo 2.1.1). En [He, 85], el autor ya observó que sus resultados en [He, 83] solo eran válidos para subconjuntos localmente cerrados de un espacio afín. Dedicamos el Capítulo 2 de esta memoria a este tema. Definimos dos nociones originales y distintas de grado para un conjunto constructible, que denotamos por $\deg_{\text{Lci}}(C)$ (grado como descomposición en conjuntos localmente cerrados irreducibles) y $\deg_{\pi}(C)$ (grado como proyección), y demostramos que cumplen la Desigualdad de Bézout. Además, estudiamos otras propiedades relevantes de ambas nociones.

Una vez que hemos obtenido nociones “correctas” para el grado de conjuntos constructibles, pasamos a estudiar los conjuntos cuestores en el Capítulo 3. En primer lugar, generalizamos los conjuntos cuestores a aplicaciones a valores vectoriales y enfatizamos que no están definidos de forma natural para distinguir funciones nulas y no nulas, sino para detectar, mediante evaluaciones en puntos, si una función pertenece o no a algún subconjunto “discriminante”. Además, observamos que los conjuntos cuestores aparecen bajo muchas formas distintas en la literatura matemática (cf. Sección 3.1 para una discusión de diversas interpretaciones).

En esta memoria, extendemos los resultados de [HS, 82] al caso de listas de polinomios. Uno de nuestros primeros resultados en este contexto es que los conjuntos cuestores satisfacen la *maldición de la dimensionalidad*: la dimensión de Krull del conjunto constructible de listas de polinomios que estamos considerando es una cota inferior para la longitud de cualquier conjunto cuestor. Usando la teoría del grado para conjuntos constructibles que hemos desarrollado, obtenemos resultados que aseguran que los conjuntos cuestores existen para cualquier conjunto constructible (y no únicamente para clausuras Zariski de conjuntos constructibles parametrizables como en [HS, 82]). Además, probamos que podemos encontrar conjuntos cuestores en cualquier conjunto constructible de co-dimensión y grado apropiados (y no solo en conjuntos con forma de “retículo” como se establece en la literatura existente). Nuestros resultados no son únicamente existenciales, sino que también probamos un resultado probabilístico que asegura que los conjuntos cuestores de longitud asintóticamente óptima (i.e. de longitud del orden de la dimensión de Krull del conjunto constructible de listas de polinomios) están densamente distribuidos en cualquier conjunto constructible de co-dimensión y grado adecuados. Por otro lado, mostramos que en el caso de listas de polinomios, el “asunto” de los conjuntos cuestores no son los Tests de Nulidad, sino la dimensión.

Por otra parte, mostramos la ubicuidad de los conjuntos cuestores en la literatura matemática a través de nuestras aplicaciones a la mejora de dos resultados famosos en el ámbito de la Combinatoria, fundamentales para el *Método Polinomial* (cf. [Tao, 14]): *las cotas inferiores de Dvir para los conjuntos de Kakeya sobre cuerpos finitos* (cf. [Dv, 09]) y el *Nullstellensatz Combinatorio de Alon* (cf. [Al, 99]).

Uno de nuestros principales resultados es un algoritmo probabilista “poco común”, basado en conjuntos cuestores, que decide en tiempo polinomial probabilístico si una lista de polinomios multivariados define o no una variedad de intersección completa (este problema es conocido como “Suite Sécante”). La principal característica de este algoritmo es que, a diferencia de los métodos clásicos de la Geometría Algebraica Computacional, no realiza ninguna manipulación sobre la lista de polinomios de entrada. Se trata de un algoritmo de “muestreo y evaluación”: simplemente evalúa la lista de polinomios de entrada en unos pocos puntos seleccionados aleatoriamente, que, con alta probabilidad, permiten obtener una respuesta correcta. La probabilidad de éxito dependerá de si los puntos seleccionados forman un conjunto cuestor o no. Este enfoque probabilista permite mostrar tanto

la potencia como las limitaciones de los conjuntos cuestores en algoritmos de “muestreo y evaluación”. Por un lado, los conjuntos cuestores son potentes, ya que permiten obtener respuestas rápidas con una probabilidad de error muy baja. Por otro lado, cuando se considera más de una ecuación, nuestro algoritmo deja de ser de tipo Montecarlo, lo que implica que no se puede garantizar la corrección de la respuesta en ninguno de los dos posibles sentidos.

En los párrafos anteriores, hemos visto que, en el caso de polinomios, los conjuntos cuestores nos permiten, de alguna forma, “aprender” si un polinomio es nulo o si una lista de polinomios define una variedad de intersección completa. Durante los últimos años, se ha producido un auge de las técnicas de Aprendizaje Computacional en muchos ámbitos de la algorítmica y la ingeniería con resultados bastante prometedores. Nuestra pregunta es si estas técnicas se pueden aplicar a nuestro contexto para *responder preguntas sobre variedades algebraicas*. Casualmente, diez años antes de los resultados de J. Heintz y C. P. Schnorr sobre conjuntos cuestores surgió la Teoría de Vapnik-Chervonenkis (cf. [VC, 71]). El resultado central de esta teoría es el Teorema de Vapnik-Chervonenkis (cf. Teorema 1.3.2) que, bajo ciertas condiciones, garantiza el “aprendizaje” de familias de funciones características (también llamadas clasificadores en el ámbito del Aprendizaje binario). Cuando dichas funciones están asociadas a variedades algebraicas, este teorema se puede interpretar como un método Montecarlo para Tests de Nulidad de Polinomios (ver Teorema 4.2.9 y su demostración para los detalles de esta interpretación).

La cantidad clave involucrada en el Teorema de Vapnik-Chervonenkis es la función de crecimiento. El exponente que aparece en la función de crecimiento es una *dimensión* que indica el número de puntos que se necesitan para obtener ciertas garantías estadísticas. En el caso de polinomios, los conjuntos cuestores son los objetos de *menor longitud teórica demostrable* que siguen las pautas de la Teoría de Vapnik-Chervonenkis. Puesto que existen conjuntos cuestores cuya longitud es del orden de la dimensión de Krull, L. M. Pardo y el autor de esta memoria propusieron la siguiente conjetura durante una conversación con J. Heintz en su última visita a Santander:

CONJETURA 1. *Existe una fuerte conexión entre la dimensión de Krull del espacio de parámetros y alguna de las nociones de dimensión en la Teoría de Aprendizaje Computacional, al menos en el caso de familias constructibles de clasificadores binarios (sobre cuerpos algebraicamente cerrados).*

La respuesta de J. Heintz fue un desafío: “*¡Suena interesante!... pero no tengo ni idea. Probadlo, si podéis*”. Dedicamos el Capítulo 4 a responder positivamente al desafío propuesto por J. Heintz.

En el caso de la Geometría Algebraica Real y la Teoría de Vapnik-Chervonenkis, existen conexiones entre sus invariantes (cf. [GJ, 95], [CKKLW,95], [KM, 97], [MPC, 08], [MP, 09] y sus referencias). Dado que los conjuntos constructibles sobre cuerpos algebraicamente cerrados de característica 0 pueden verse como conjuntos semi-algebraicos de algún espacio afín real, una primera idea sería emplear los resultados anteriores para establecer algunas conexiones en nuestro contexto. Sin embargo, estas conexiones solo serían válidas en el caso de característica 0 y se basan en cotas superiores para el número de componentes conexas de conjuntos semi-algebraicos (cf. [Mi, 64], [Th, 65], [Ol, 51], [OP, 62] y [Wa, 68]), y no en cantidades naturales de la teoría de conjuntos constructibles. Nuestro objetivo es obtener resultados que sean válidos para cualquier característica y que usen nociones naturales de la teoría de conjuntos constructibles. Por lo tanto, hemos tenido que buscar otras técnicas para dar respuesta a la Conjetura 1 y al desafío de J. Heintz.

Sorprendentemente, la noción clave que nos va a permitir conectar la Geometría Algebraica sobre cuerpos algebraicamente cerrados y la Teoría de Vapnik-Chervonenkis es la noción de Erzeugungsgrad introducida por J. Heintz en [He, 83]. Esta noción fue utilizada por J. Heintz para acotar el número de celdas no vacías que aparecían tras un proceso de eliminación de cuantificadores. El resultado principal sobre el Erzeugungsgrad de J. Heintz son unas cotas combinatorias que permiten acotar el número de celdas no vacías. En cierto sentido, estas cotas combinatorias recuerdan a *las funciones de crecimiento* usadas en la Teoría de Aprendizaje Computacional con un propósito distinto. En este caso, la cantidad que aparece en el exponente es la dimensión de Krull de un conjunto constructible. Esta coincidencia es la que nos permitirá resolver positivamente la Conjetura 1 y el desafío de J. Heintz.

Antes de demostrar nuestros resultados, tenemos que resolver un problema técnico que aparece en la formulación original del Erzeugungsgrad en [He, 83]. J. Heintz utilizó el \deg_z como punto de partida para definir el Erzeugungsgrad. Como ya hemos comentado, esta noción de grado no verifica una Desigualdad de Bézout para conjuntos constructibles y, por lo tanto, su uso afecta tanto a los argumentos como a las cotas que utilizan el Erzeugungsgrad. Por ello, en la Sección 4.1.1, adaptamos el Erzeugungsgrad y las cotas combinatorias del Teorema 2 de [He, 83] al caso de conjuntos constructibles.

Una vez resueltas las dificultades del Erzeugungsgrad, pasamos en la Sección 4.2 a establecer conexiones entre la Geometría Algebraica y la Teoría de Vapnik-Chervonenkis. Nuestros resultados principales, que utilizan como

ingrediente clave el Erzeugungsgrad, nos permiten responder de forma positiva a la Conjetura 1 y al desafío de J. Heintz: *en el caso de clasificadores binarios de clases parametrizadas por familias de conjuntos constructibles, la dimensión de Vapnik-Chervonenkis y la dimensión de Krull están relacionadas linealmente, salvo por factores logarítmicos basados en la Teoría de la Intersección*. Usando este resultado, demostramos que los conjuntos cuestores están densamente distribuidos en variedades evasivas de dimensión positiva, extendiendo al caso de dimensión positiva el tratamiento habitual de dimensión de Krull cero.

Por otro lado, el Capítulo 5 está dedicado al estudio de redes neuronales con función de activación polinómica o racional. En el caso racional, consideramos únicamente redes neuronales que no admiten “Vermeidung von Divisionen”, es decir, redes neuronales que calculan funciones racionales que no son polinomios. Comenzamos el capítulo introduciendo el formalismo y recordando sus conexiones con la Geometría Algebraica Computacional a través del algoritmo clásico TERA. Posteriormente, aplicamos los resultados anteriores para, por un lado, estudiar las funciones de crecimiento de la clase de clasificadores binarios asociada a una red neuronal con función de activación polinómica o racional mientras que, por otro lado, discutimos la existencia y la densidad de conjuntos cuestores para clases de polinomios y funciones racionales definidas por familias parametrizadas de redes neuronales.

A finales de los años 80, N. Natarajan introdujo en [Na, 88] y [Na, 89] una generalización del Aprendizaje binario de L. G. Valiant (cf. [Va, 84]) y de la Teoría de Vapnik-Chervonenkis (cf. [VC, 71]). La noción de N. Natarajan fue denominada *Aprendizaje multiclase*. El Aprendizaje multiclase difiere del Aprendizaje binario de Vapnik y Chervonenkis en varios aspectos. En el Aprendizaje multiclase se considera una clase de entradas \mathcal{X} y un conjunto de etiquetas que extiende al caso binario (aunque sigue siendo finito). Denotamos a este último conjunto por $\mathcal{Y} := \mathbb{Z}/p\mathbb{Z} = \{0, \dots, p-1\}$, que son los representantes canónicos de las clases de los restos módulo p , donde $p \geq 2$, no necesariamente primo. Nótese que el caso $p = 2$ del Aprendizaje multiclase es precisamente el Aprendizaje binario. Denotamos por $\mathcal{Y}^{\mathcal{X}}$ al conjunto de todas las funciones de la forma $f : \mathcal{X} \rightarrow \mathcal{Y}$. En el caso finito, tenemos que $\mathcal{X} := [n] := \{1, \dots, n\}$, por lo que las funciones serán elementos de $\mathcal{Y}^{[n]} = \mathcal{Y}^n$. Además, como en todos los casos de aprendizaje, disponemos del conjunto $\mathcal{Z} := \mathcal{X} \times \mathcal{Y}$, dotado con una σ -álgebra $\mathcal{B}(\mathcal{Z})$ y una distribución de probabilidad “desconocida” $\nu : \mathcal{B}(\mathcal{Z}) \rightarrow [0, 1]$ (esto significa que los resultados deben probarse independientemente de la distribución de probabilidad escogida).

Inspirándonos en la relación entre los conjuntos cuestores y la teoría de Vapnik-Chervonenkis (i.e Aprendizaje binario), surge de forma natural la siguiente pregunta: *¿existe alguna conexión entre el Aprendizaje multiclase, los polinomios y los conjuntos cuestores?* El Capítulo 6 constituye un primer acercamiento a esta cuestión y el inicio de una línea de investigación orientada a explorar estas posibles conexiones. En este capítulo, nos centramos en el siguiente objetivo preliminar: *el estudio de los elementos implicados en el algoritmo OIG (One Inclusion Graph) en el contexto del Aprendizaje multiclase*. Nuestro objetivo no es estudiar en algoritmo en sí, sino analizar los elementos ligados a este algoritmo y a su estructura de datos subyacente: el *hiper-grafo de una inclusión*. El hiper-grafo de una inclusión y el algoritmo OIG fueron empleados por primera vez en [Ha, 95, HLW, 94] para el caso binario. En [RBR, 06], sus autores extendieron tanto la noción como el algoritmo al caso multiclase. Posteriormente, en [DS, 14] y [BCDMY, 22] se vuelve a analizar el algoritmo en el caso multiclase.

Aunque nuestro objetivo no es estudiar el algoritmo OIG, lo expondremos para ayudar al lector a comprender la motivación de nuestra investigación. Comenzamos explicando en qué consiste el “problema de aprendizaje” en el contexto multiclase. Partimos de una clase de funciones $\mathcal{H} \subseteq \mathcal{Y}^{\mathcal{X}}$, habitualmente denominada clase de conceptos, que representa el conjunto de funciones que se desea aprender. El objetivo es determinar una función a partir de una lista finita de muestras (también llamada *dataset*) de longitud m :

$$\mathcal{S} := \{(x_1, y_1), \dots, (x_m, y_m)\} \in (\mathcal{X} \times \mathcal{Y})^m.$$

Nuestro objetivo no es interpolar una función a partir del dataset, sino “calcular algorítmicamente” una función $h^* \in \mathcal{H}$ que minimice cierta función de riesgo. Más aún, en esta memoria no aspiramos a obtener la función h^* , sino únicamente a predecir su valor en un punto dado $x \in \mathcal{X}$.

En el algoritmo OIG no trabajaremos directamente con la clase de conceptos \mathcal{H} , sino con su restricción al dataset junto al punto que queremos evaluar (supondremos que esta restricción puede “calcularse”):

$$H := \mathcal{H}|_{(x_1, \dots, x_m, x)} := \{(h(x_1), \dots, h(x_m), h(x)) \in \mathcal{Y}^{m+1} : h \in \mathcal{H}\}.$$

Antes de presentar el algoritmo, necesitamos introducir la noción de hiper-grafo de una inclusión. Todo el material que exponemos a continuación se detalla en la Sección 6.2. Consideramos una clase de conceptos finita $H \subseteq \mathcal{Y}^{m+1}$, donde, al igual que antes, $\mathcal{Y} := \{0, \dots, p-1\}$. Sea $\pi_i : \mathcal{Y}^{m+1} \rightarrow \mathcal{Y}^m$ la proyección que “olvida” la coordenada i -ésima. Para cada $y \in \mathcal{Y}^m$, definimos la fibra sobre y de π_i en H como:

$$e_{y,i} := \pi_i^{-1}(\{y\}) \cap H.$$

Si esta fibra es no vacía, definimos la arista: $e_{y,i} \times \{i\} \subseteq H \times [m+1]$ y consideramos todas las aristas asociadas a la proyección π_i :

$$E_i(H) := \{e_{y,i} \times \{i\} : y \in \pi_i(H)\}.$$

Definimos el conjunto de aristas $E(H)$ del hiper-grafo de una inclusión como la siguiente unión disjunta:

$$E(H) := \bigsqcup_{i=1}^{m+1} E_i(H).$$

Para una arista $e := e_{y,i} \times \{i\} \in E(H)$ y un elemento $h \in H$, decimos que $h \in e$ si y solo si $h \in e_{y,i}$. El hiper-grafo de una inclusión asociado a H viene dado por $G(H) = (H, E(H))$. Una noción central en nuestra discusión es la de orientación de $G(H)$. Una orientación (cf. Definición 38) es una aplicación $\sigma : E(H) \rightarrow H$ que satisface la siguiente propiedad:

$$\forall e \in E(H), \sigma(e) \in e.$$

Una vez introducidas las nociones necesarias, estamos en condiciones de presentar el algoritmo OIG (revisado), siguiendo el esquema propuesto en [DS, 14] (ya que la presentación que realizan en [BCDMY, 22] no nos parece correcta):

ALGORITMO 0.2.2 (Algoritmo OIG, revisado, formato Toy Example).

- INPUT:
 - Un punto $(x_1, \dots, x_{m+1}) \in \mathcal{X}^{m+1}$
 - La clase de conceptos $H := \mathcal{H}|_{(x_1, \dots, x_{m+1})} \subseteq \mathcal{Y}^{m+1}$.
 - Un “dataset” $y := (y_1, \dots, y_m) \in \mathcal{Y}^m$.
- guess un índice $i \in [m+1]$
- verify que $y \in \pi_i(H)$
- compute una orientación $\sigma : E(H) \rightarrow H$ que minimice $\text{outdeg}_{\max}(\sigma)$
- OUTPUT: $h := \sigma(e_{y,i} \times \{i\}) \in H$ % Es decir, $h(x_i) = h_i$, $h(x_j) = y_j$ si $1 \leq j \leq i-1$ y $h(x_j) = y_{j-1}$ si $i+1 \leq j \leq m+1$

La inclusión de la instrucción **guess** (con distribución de probabilidad uniforme en $[m+1]$) es un requisito necesario para que la Proposición 15 de [BCDMY, 22] (clave para que el algoritmo anterior sea un algoritmo de aprendizaje) sea cierta. En otro caso, la Proposición 15 de [BCDMY, 22] es falsa. Por otro lado, resulta evidente que la cantidad clave del algoritmo anterior es:

$$\min - \text{outdeg}_{\max}(H) = \min\{\text{outdeg}_{\max}(\sigma) : \sigma : E(H) \rightarrow H \text{ es una orientación}\}.$$

La finalidad de nuestro estudio es controlar, de algún modo, la cantidad anterior. Esto se hace en el Teorema 6.4.27, mediante el uso de una noción original de este capítulo: *el grado de salida promedio*. Para llegar a este resultado, ha sido necesario llevar a cabo un análisis detallado de las distintas nociones de grado (cf. Sección 6.2) y dimensión (cf. Sección 6.3), así como de la técnica de shifting (cf. Sección 6.4).

0.3. Resumen pormenorizado de los contenidos

Los resultados de esta memoria están basados en los siguientes trabajos:

- **Capítulos 2 y 3:** [PS, 22] L. M. Pardo, D. Sebastián, *A promenade through correct test sequences I: Degree of constructible sets, Bézout’s Inequality and density*. J. of Complexity **68** (2022), 101588.
- **Capítulos 4 y 5:** [PS, 25] L. M. Pardo, D. Sebastián, *Erzeugungsgrad, VC-Dimension and Neural Networks with rational activation function*. En arXiv:2504.11345 [cs.LG]. Sometido a Applicable Algebra in Engineering, Communication and Computing.
- **Capítulo 6:** [PSZ, 25] L. M. Pardo, D. Sebastián, J. Zhang, *Some comments on the expected out-degree and the average degree of the one-inclusion hyper-graph in multiclass learning*. Manuscrito en preparación.

En las siguientes secciones hemos resumido los principales resultados de cada capítulo. Hemos omitido el Capítulo 1 por tratarse de un capítulo de material preliminar.

0.3.1. Capítulo 2: Desigualdad de Bézout para conjuntos constructibles. El objetivo de este capítulo es generalizar la teoría del grado para conjuntos localmente cerrados de [He, 83] al caso de conjuntos constructibles. En primer lugar, un *conjunto constructible* es un subconjunto $C \subseteq \mathbb{A}^n(K)$ que puede obtenerse como la proyección de una variedad algebraica de un espacio afín de dimensión mayor (ver, por ejemplo, [Ch, 64-65]). Uno de los problemas de la Geometría Algebraica afín es la falta de una teoría consistente para el grado de conjuntos constructibles que cumpla, al menos, una Desigualdad de Bézout (el grado de la intersección está acotado por el producto de los grados). En [He, 83], J. Heintz definió el grado de un conjunto constructible C como el grado de su clausura Zariski, que denotamos por $\deg_z(C)$. Como ya observó el propio J. Heintz en [He, 85], la Desigualdad de Bézout de [He, 83] es únicamente válida en el caso de conjuntos localmente cerrados. En el Ejemplo 2.1.1, mostramos un ejemplo original de un conjunto constructible C , conocido como *La Croix de Berny*, y una variedad algebraica V tal que $\deg_z(C \cap V)$ no está acotado por el producto $\deg_z(C) \deg_z(V)$. Por lo tanto, \deg_z no satisface ni la Desigualdad de Bézout ni otros resultados de [He, 83].

Inspirados por el espíritu implícito en [He, 83], nuestro trabajo crea un *paradigma* de lo que cualquier noción de grado para conjuntos constructibles debe satisfacer. Un grado para conjuntos constructibles ha de ser una cantidad (que actúa como un volumen o una probabilidad) que verifica las siguientes propiedades:

- Es siempre finita y coincide con la noción habitual en el caso de variedades algebraicas y conjuntos localmente cerrados.
- Es sub-aditiva.
- Presenta un “buen comportamiento” con respecto a las intersecciones con variedades lineales afines y con respecto al producto cartesiano.
- Satisface una Desigualdad de Bézout.
- Está “controlada” bajo proyecciones e imágenes de aplicaciones lineales.
- Satisface alguna variación de la Proposición 2.3 de [HS, 82].

De las propiedades anteriores, únicamente la Proposición 2.3 de [HS, 82] requiere algunas explicaciones adicionales. Nótese que una aplicación directa de la Desigualdad de Bézout sitúa la co-dimensión en el exponente de las cotas para el grado de la intersección de varios objetos geométricos. Es decir, dados $f_1, \dots, f_m \in K[X_1, \dots, X_n]$ tales que $V := V_{\mathbb{A}}(f_1, \dots, f_m)$ tiene dimensión $n - m$ y una variedad algebraica $W \subseteq \mathbb{A}^n(K)$, la Desigualdad de Bézout nos proporciona la siguiente cota superior:

$$\deg(W \cap V_{\mathbb{A}}(f_1, \dots, f_m)) \leq \deg(W) \prod_{i=1}^m \deg(f_i) \leq \deg(W) (\max\{\deg(f_1), \dots, \deg(f_m)\})^{\text{codim}(V)}.$$

Sin embargo, la Proposición 2.3 de [HS, 82] reemplaza la co-dimensión de V por la dimensión de W , destacando el papel de la dimensión de W en este tipo de cotas superiores. Esto es, aplicando la Proposición 2.3 de [HS, 82], obtenemos:

$$\deg(W \cap V_{\mathbb{A}}(f_1, \dots, f_m)) \leq \deg(W) (\max\{\deg(f_1), \dots, \deg(f_m)\})^{\dim(W)}.$$

El “trade-off” entre estas dos cotas superiores es el ingrediente clave para demostrar en el Capítulo 3 la existencia y la alta probabilidad de conjuntos cuestores de longitud asintóticamente óptima en cualquier conjunto constructible.

Teniendo en cuenta los requisitos previos, hemos definido dos nociones originales y distintas de grado para conjuntos constructibles. La primera noción de grado para conjuntos constructibles $C \subseteq \mathbb{A}^n(K)$ está basada en el hecho de que los conjuntos constructibles pueden representarse como uniones finitas de conjuntos localmente cerrados irreducibles (cf. Lema 1.1.2 para más detalles). Denotaremos a esta noción por $\deg_{\text{lci}}(C)$ (ver Definición 13). Por otro lado, la segunda noción de grado se basa en que los conjuntos constructibles pueden representarse como proyecciones $C = \pi(W)$ de conjuntos localmente cerrados $W \subseteq \mathbb{A}^m(K)$, donde $m \geq n$. Para esta segunda noción, usaremos la notación $\deg_{\pi}(C)$ (ver Definición 14). A continuación, veremos cómo las diferentes propiedades del paradigma descrito previamente se aplican de forma distinta al \deg_{lci} y \deg_{π} . No hemos tomado una decisión sobre cuál de las nociones es “mejor” o “más apropiada”. En los siguientes párrafos, resumimos algunas de las propiedades de estos grados que se demuestran en esta memoria.

En el Ejemplo 2.1.3, mostramos un conjunto constructible $C \subseteq \mathbb{A}^2(\mathbb{C})$ (una variación de La Croix de Berny) en el que las tres nociones de grado toman valores distintos:

$$\deg_z(C) < \deg_{\pi}(C) < \deg_{\text{lci}}(C).$$

Hemos descartado \deg_z como noción de grado ya que no verifica la Desigualdad de Bézout. Por lo tanto, nos centraremos en el estudio de \deg_{lci} y \deg_{π} . Demostramos que ambas nociones de grado verifican las propiedades esperadas: son sub-aditivas (cf. Proposición 2.1.2), generalizan la noción de grado para conjuntos localmente cerrados de [He, 83] (cf. Proposición 2.1.9) y tienen un buen comportamiento para intersecciones

con variedades afines lineales y el producto cartesiano (cf. Proposición 2.1.13). Finalmente, probamos que ambas nociones de grado satisfacen una Desigualdad de Bézout para conjuntos constructibles (cf. Teorema 2.1.14). Este resultado afirma simplemente que, dados dos conjuntos constructibles $C, D \subseteq \mathbb{A}^n(K)$, se verifican las siguientes desigualdades:

$$\deg_\pi(C \cap D) \leq \deg_\pi(C) \deg_\pi(D), \quad \deg_{\text{lci}}(C \cap D) \leq \deg_{\text{lci}}(C) \deg_{\text{lci}}(D).$$

Desafortunadamente, no hemos podido esclarecer completamente el papel de ambas nociones con respecto a las dos propiedades restantes: *la Proposición 2.3 de [HS, 82] y el grado de las imágenes bajo aplicaciones lineales*. Obviamente, \deg_π es “perfecto” para acotar el grado de las imágenes bajo aplicaciones lineales. Por otro lado, se puede extender la Proposición 2.3 de [HS, 82] para la intersección de un conjunto constructible con varios conjuntos localmente cerrados tanto para \deg_{lci} (cf. Proposición 2.2.2) como para \deg_π (bajo la hipótesis de presentable equidimensionalmente, cf. Proposición 2.2.3).

En la Sección 2.2, demostramos que \deg_{lci} satisface una estimación similar a la Proposición 2.3 de [HS, 82] para conjuntos constructibles de cualquier tipo (cf. Teorema 2.2.1), intercambiando la co-dimensión por la dimensión en el exponente de la cota superior. A continuación, reproducimos dicho teorema para facilitar al lector el acceso a los resultados principales:

TEOREMA (DESTACABLE) 0.3.1 (Cotas superiores para el grado LCI de la intersección de varios conjuntos constructibles). *Sea K un cuerpo algebraicamente cerrado y sea $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ una familia de conjuntos constructibles. Sea $r := \dim(C_1) \geq 0$ la dimensión del conjunto constructible C_1 . Se verifican las siguientes desigualdades:*

i) Primera cota superior:

$$\deg_\pi \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \binom{s+r-1}{r} \deg_{\text{lci}}(C_1) (\max\{\deg_{\text{lci}}(C_j) : 2 \leq j \leq s\})^r.$$

ii) Segunda cota superior:

$$\deg_\pi \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{lci}}(C_1) \left(1 + \sum_{i=2}^s \deg_{\text{lci}}(C_i) \right)^r.$$

iii) Cota superior en términos del grado promedio: *Dada una familia de conjuntos constructibles*

$$C_1, \dots, C_s \subseteq \mathbb{A}^n(K),$$

definimos su grado promedio como:

$$\deg_{\text{av}}(C_1, \dots, C_s) := \frac{1}{s} \sum_{i=1}^s \deg_{\text{lci}}(C_i).$$

Entonces, también tenemos que:

$$\deg_\pi \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{lci}}(C_1) s^r (\deg_{\text{av}}(C_1, \dots, C_s))^r.$$

No hemos logrado establecer un resultado similar con \deg_π en el lado derecho de alguna de las desigualdades anteriores. Tampoco hemos encontrado un contra-ejemplo que demuestre que \deg_π no es adecuado para este tipo de resultados.

Por otro lado, también hemos estudiado cómo se comporta el grado de las imágenes de conjuntos constructibles bajo aplicaciones lineales (como en el Lema 2 de [He, 83]). Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ una aplicación lineal. En el Ejemplo 2.1.3, demostramos que existen variedades algebraicas $W \subseteq \mathbb{A}^3(\mathbb{C})$ que verifican la siguiente desigualdad:

$$\deg_{\text{lci}}(\ell(W)) > \deg_{\text{lci}}(W) = \deg(W).$$

Por lo tanto, el Lema 2 de [He, 83] es falso para conjuntos constructibles y \deg_{lci} . Esto es, el grado de las imágenes de conjuntos constructibles por aplicaciones lineales puede aumentar. En el Ítem i) de la Proposición 2.1.11, probamos que: $\deg_z(\ell(C)) \leq \deg_{\text{lci}}(C)$. Como ya hemos comentado antes, el \deg_π se comporta de forma adecuada respecto a las aplicaciones lineales (cf. Proposición 2.1.12). En concreto, demostramos que se cumple lo siguiente:

$$\deg_z(\ell(C)) \leq \deg_\pi(\ell(C)) \leq \deg_\pi(C) \leq \deg_{\text{lci}}(C),$$

para cualquier conjunto constructible C y cualquier transformación lineal ℓ . Estas desigualdades dan lugar a la noción de *defecto* de la imagen $\ell(C)$ de un conjunto constructible, definida como la siguiente diferencia:

$$\deg_{\text{lci}}(\ell(C)) - \deg_z(\ell(C)).$$

Este capítulo estaría, de alguna forma, incompleto si no intentásemos obtener algunas cotas superiores para este defecto. Esto se hace en la Subsección 2.3.1, donde obtenemos cotas superiores para el defecto en el Lema 2.3.2. En la Observación 2.3.3, mostramos un ejemplo en el que la cota superior obtenida en el Lema 2.3.2 es óptima.

Las desigualdades anteriores muestran un control del grado de la imagen de un conjunto constructible obtenido mediante aplicaciones lineales. Son cotas del tipo intrínseco-semántico en la medida en que los únicos elementos que intervienen dependen del conjunto constructible C . Sin embargo, en muchas ocasiones nuestro objeto geométrico viene dado mediante expresiones sintácticas que involucran ecuaciones polinomiales. Los polinomios particulares que definen un objeto geométrico no están unívocamente determinados por el objeto y pueden tener variaciones con impacto en la Teoría de la Intsección (cf. [HMPP, 21], por ejemplo). En este sentido, podemos considerar extrínsecas las acotaciones de grados de objetos geométricos a partir de las ecuaciones que los definen. Más aún, en muchas aplicaciones solo disponemos de la información sintáctica-extrínseca para poder controlar los grados de los objetos geométricos y sus transformaciones. Por ello, en la Subsección 2.3.2, hemos obtenido una cota superior extrínseca para $\deg_{\text{lci}}(\ell(C))$. Para que el lector tenga una idea más clara de a qué nos referimos, reproducimos a continuación el Teorema 2.3.4 de la Subsección 2.3.2. En este resultado, demostramos que $\deg_{\text{lci}}(\ell(C))$ está acotado (y, por lo tanto, “controlado”) en términos de algunas cantidades sintácticas que dependen únicamente de C :

TEOREMA (DESTACABLE) 0.3.2. *Sea K un cuerpo algebraicamente cerrado. Sea $V \subseteq \mathbb{A}^n(K)$ una variedad algebraica irreducible de dimensión r . Sea m un entero positivo tal que $m \leq n$. Supongamos que existen polinomios $g_1, \dots, g_s \in K[X_1, \dots, X_n]$ de grados $d_i := \deg(g_i)$, $1 \leq i \leq s$, tales que $V = V_{\mathbb{A}}(g_1, \dots, g_s)$ y se verifican las siguientes desigualdades:*

$$d_1 \geq d_2 \geq \dots \geq d_s.$$

Consideremos la siguiente cantidad:

$$N := N(d_1, \dots, d_s, n, r) := \begin{cases} \prod_{i=1}^s d_i & \text{si } s \leq n - m \\ 2d_s \left(\prod_{i=1}^{n-m-1} d_i \right) - 1 & \text{si } s > n - m \end{cases}$$

Definamos las siguientes cantidades:

$$\tilde{N} := \binom{N + (n - m)}{n - m},$$

$$M := \sum_{i=1}^s \binom{N - d_i + (n - m)}{(n - m)},$$

y

$$\mathcal{N}' := \min\{\tilde{N}, M + 1\}.$$

Sea $\pi : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ la proyección canónica que “olvida” las últimas $n - m$ variables y sea $W := \overline{\pi(V)}^z$ la clausura Zariski de $\pi(V)$ en $\mathbb{A}^m(K)$. Entonces, tenemos que:

$$\deg_{\text{lci}}(\pi(V)) \leq \deg(V) (2d_1)^{\dim(W)} (\mathcal{N}')^{\dim(W)+1}.$$

0.3.2. Capítulo 3: Conjuntos Cuestores. En el Capítulo 3 revisamos y generalizamos la noción de conjunto cuestor y sus resultados. Comenzamos la Sección 3.1 generalizando la definición de conjunto cuestor introducida en [HS, 82] para Tests de Nulidad de Polinomios al caso de aplicaciones a valores vectoriales (ver Definición 16). Esto es, consideramos un conjunto X y una clase de funciones $\mathcal{F}(X) \subseteq K^X$ con valores en un cuerpo K . Dado $\Omega \subseteq \mathcal{F}(X)^m$, una clase distinguida de funciones en $\mathcal{F}(X)^m$ (i.e. aplicaciones e X a K^n), y un subconjunto $\Sigma \subsetneq \Omega$, un *conjunto cuestor de longitud L para Ω con discriminante Σ* es un conjunto finito de L elementos $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq X$ tal que se verifica la siguiente fórmula:

$$(0.3.1) \quad \forall f \in \Omega, \quad f(x_1) = \dots = f(x_L) = 0 \in K^m \implies f \in \Sigma.$$

En el caso $\Sigma = \{0\} \subsetneq \Omega$, decimos simplemente que \mathbf{Q} es un conjunto cuestor para Ω . El objetivo de esta sección es mostrar varias nociones equivalentes a los conjuntos cuestores según el contexto: como puntos para Tests de Identidad (ver Proposición 3.1.1), como *conjuntos normantes* finitos en el caso de que K sea un cuerpo con algún valor absoluto (ver Proposición 3.1.5), en el anillo de funciones continuas $\mathcal{C}(X)$ sobre un espacio topológico (ver Proposición 3.1.7) o como muestras en Espacios de Hilbert con Núcleo Reprodutor (ver Proposición 3.1.10).

En la Subsección 3.1.1, nos centramos en el caso de los polinomios multivariados. Sea K un cuerpo algebraicamente cerrado y $d \in \mathbb{N}$ un entero positivo. Consideramos el espacio $P_d^K := P_d^K(X_1, \dots, X_n)$ de todos los polinomios en $K[X_1, \dots, X_n]$ de grado a lo sumo d . Para una lista de grados $(d) := (d_1, \dots, d_m) \in \mathbb{N}^m$, consideramos la clase $\mathcal{P}_{(d)}^K$ de todas las listas $f := (f_1, \dots, f_m)$ de m polinomios tales que $f_i \in P_{d_i}^K$ para cada i , $1 \leq i \leq m$. Estudiaremos los conjuntos cuestores \mathbf{Q} asociados a subconjuntos constructibles $\Omega \subseteq \mathcal{P}_{(d)}$ con respecto a algún subconjunto discriminante $\Sigma \subsetneq \Omega$. Nuestro primer resultado en este contexto es la *maldición de la dimensionalidad de los conjuntos cuestores* (cf. Proposición 3.1.12): si \mathbf{Q} es un conjunto cuestor de longitud L para un conjunto constructible $\Omega \subseteq \mathcal{P}_{(d)}$ con respecto a $\Sigma \subseteq \Omega$, entonces:

$$L \geq \dim(\Omega) - \dim(\Sigma) = \text{codim}_\Omega(\Sigma).$$

La cota anterior establece un límite inferior para la longitud de los conjuntos cuestores. Por este motivo, *decimos que un conjunto cuestor tiene longitud asintóticamente óptima si su longitud está en $O(\dim(\Omega))$* .

A continuación, en la Sección 3.2, abordamos la cuestión de la existencia conjuntos cuestores de longitud asintóticamente óptima. Nuestro resultado principal (cf. Teorema 3.2.1) demuestra, en primer lugar, que existen conjuntos cuestores para cualquier conjunto constructible y no únicamente para conjuntos constructibles parametrizados como en [HS, 82]. En segundo lugar, probamos que los conjuntos cuestores pueden encontrarse en cualquier conjunto localmente cerrado de co-dimensión y grado adecuados, ampliando así la clase de conjuntos “reticulares” (de la forma $(Q_1 \times \dots \times Q_n)^L \subseteq (\mathbb{A}^n)^L$, donde $Q_i \subseteq K$ es un conjunto finito) considerada en [HS, 82]. Finalmente, y no menos importante, demostramos que los conjuntos cuestores de longitud asintóticamente óptima son altamente densos en términos probabilísticos en cualquier conjunto de la forma C^L , donde C es un localmente cerrado y L está en $O(\dim(\Omega))$.

Dado un conjunto localmente cerrado $C \subseteq \mathbb{A}^n(K)$, un entero positivo $L \in \mathbb{N}$ y Ω y Σ como antes, denotamos por $R(\Omega, \Sigma, C, L)$ al conjunto constructible de todas las listas $\mathbf{Q} \in C^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . El Teorema 3.2.1 demuestra que, bajo ciertas hipótesis técnicas sobre el grado y la co-dimensión de C , para $L \geq 6 \dim(\Omega)$, el conjunto $R(\Omega, \Sigma, C, L)$ es altamente denso en C^L . En el Corolario 3.2.4, extendemos el teorema anterior al caso en el que C es también un conjunto constructible de grado y co-dimensión adecuados.

Una de las dificultades que hemos tenido para demostrar este tipo de resultados es que no hay una forma “inmediata” de definir una distribución de probabilidad en $R(\Omega, \Sigma, C, L)$. Al comienzo de la Sección 3.2, incluimos varias referencias que explican cómo lograr lo anterior a través del número esperado de puntos en intersecciones con variedades lineales afines “aleatorias” $A \subseteq \mathbb{A}^n(K)$ en el Grassmaniano $\mathbb{G}(n, n-r)$, donde r es la co-dimensión de C . Esto es, con el fin de estudiar la distribución de probabilidad de los conjuntos cuestores de longitud L para Ω con respecto a Σ en un conjunto constructible C , estudiamos la siguiente probabilidad:

$$\text{Prob}_{(C \cap A)^L} [R(\Omega, \Sigma, C, L)] = \frac{\#(R(\Omega, \Sigma, C, L) \cap A^L)}{\#((C \cap A)^L)}$$

donde $(C \cap A)^L$ está dotado de su distribución de probabilidad uniforme y A es una variedad afín lineal que pertenece a un subconjunto abierto y denso con respecto a la topología final (de Zariski) en $\mathbb{G}(n, n-r)$.

Discutimos el caso particular en el que C es una variedad de intersección completa en varios corolarios. Reproducimos aquí el Corolario 3.2.5, que muestra la alta densidad de conjuntos cuestores cuando C es una variedad de intersección completa de co-dimensión y grado adecuados:

TEOREMA (DESTACABLE) 0.3.3. *Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}(X_1, \dots, X_n)$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:*

$$(0.3.2) \quad \forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $C := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ una variedad algebraica de intersección completa de co-dimensión $r \geq (n-m) + m/2 + 1/2$ tal que $\deg(C) \geq \delta^r$, donde $\delta := \min\{\deg(h_1), \dots, \deg(h_r)\}$. Sea $L \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $L \geq 6 \dim(\Omega)$,
- ii) $\log(\delta) \geq \max\{2(1 + \log(d+1)), \frac{2 \log(\deg_{\text{ici}}(\Omega))}{\dim(\Omega)}\}$ y,
- iii) $\max\{\deg(h_1), \dots, \deg(h_r)\} \leq (1 + \frac{1}{n-m})\delta$.

Sea $R := R(\Omega, \Sigma, C, L)$ el conjunto constructible de todas las listas $\mathbf{Q} \in C^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, n-r)$, con

respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(C)$, la probabilidad de que una lista $\mathbf{Q} \in (V \cap A)^L$ escogida aleatoriamente esté en R satisface:

$$\text{Prob}_{(C \cap A)^L}[R] \geq 1 - \frac{1}{\text{deg}_{\text{lci}}(\Omega)e^{\dim(\Omega)+(m-1)L}},$$

donde $(V \cap A)^L$ está dotado de su distribución de probabilidad uniforme.

Nótese que si dotamos a C de una distribución de probabilidad basada en una fórmula del tipo Poincaré para cuerpos algebraicamente cerrados (ver la Sección 2 de [BP, 07] para el caso complejo) involucrando $\sharp(C \cap A)$ con $A \in \mathbb{G}(n, n-r)$, el resultado anterior simplemente establece que la medida de $R(\Omega, \Sigma, C, L)$ en C^L es cercana a 1.

Con el objetivo de enfatizar algunas de las características en las que se diferencian los Tests de DeMillo–Lipton–Schwartz–Zippel y los conjuntos cuestores, demostramos que los conjuntos cuestores de longitud asintóticamente óptima son altamente densos en el conjunto de nuestro empleado por estos autores en [DML, 78], [Zp, 79] and [Sch, 80]. Lo anterior se recoge en el Corolario 3.2.9, que reproducimos a continuación:

COROLARIO (DESTACABLE) 0.3.4. *Sean $n, d \in \mathbb{N}$ dos enteros positivos. Sea κ un cuerpo y K su clausura algebraica y sea Ω un subconjunto constructible de P_d^K . Supongamos que $\Omega \setminus \{0\} \neq \emptyset$. Sea $Q \subseteq \kappa$ un conjunto finito. Sea $L \in \mathbb{N}$ un entero positivo tal que se verifican las siguientes propiedades:*

- i) $L \geq 6 \dim(\Omega)$ y,
- ii) $\sharp(Q) \geq \max\{(e(d+1))^2, \text{deg}_{\text{lci}}(\Omega)^{\frac{2}{\dim(\Omega)}}\}$.

Entonces, la probabilidad de que una lista $\mathbf{Q} \in V^L = (Q^n)^L$ escogida aleatoriamente sea un conjunto cuestor para Ω es mayor que:

$$1 - \frac{1}{\text{deg}_{\text{lci}}(\Omega)e^{\dim(\Omega)}}.$$

En el caso de entrada denso (i.e. cuando $\Omega = P_d^K$), para cualquier $L \geq 6 \dim(P_d^K)$ y $Q \subseteq \kappa$ tal que:

$$\sharp(Q) \geq (e(d+1))^2,$$

la probabilidad de que una lista $\mathbf{Q} \in (Q^n)^L$ escogida aleatoriamente sea un conjunto cuestor para P_d^K es mayor que:

$$1 - \frac{1}{e^{\binom{d+n}{n}}}.$$

En la Sección 3.3, reformulamos dos enunciados del *Método Polinomial* (terminología debida a T. Tao en [Tao, 14]) en términos de conjuntos cuestores. Por un lado, discutimos, en la Subsección 3.3.2, la cota inferior exponencial obtenida por Z. Dvir para los conjuntos de Kakeya (cf. [Dv, 09]) en términos de conjuntos cuestores. Básicamente, un conjunto $E \subseteq \mathbb{A}^n(\mathbb{F}_q)$, donde \mathbb{F}_q es un cuerpo finito, se denomina conjunto de Kakeya si contiene una recta en cada dirección. Demostramos que los conjuntos de Kakeya son conjuntos cuestores para ciertos conjuntos constructibles y, por lo tanto, la cota inferior de Z. Dvir es una consecuencia inmediata de la *maldición de la dimensionalidad* discutida previamente (cf. Corolario 3.3.2). Además, también demostramos que la mayoría de los conjuntos cuestores no son conjuntos de Kakeya cuando el grado de los polinomios involucrados está acotado por $q^{1-\varepsilon} - 1$ para valores pequeños de $\varepsilon > 0$ (cf. Corolario 3.3.3). A continuación, reproducimos el Corolario 3.3.3 con el objetivo de facilitar al lector el acceso a los resultados principales:

COROLARIO (DESTACABLE) 0.3.5. *Sea \mathbb{F}_q un cuerpo finito con q elementos y sea $\overline{\mathbb{F}_q}$ su clausura algebraica. Sea $d \in \mathbb{N}$ un grado y $k \in \mathbb{N}$ un entero positivo tal que:*

$$(0.3.3) \quad d < q^{1-\frac{1}{k}} - 1.$$

Entonces, para $s := k \binom{d+n}{n}$, la probabilidad de que una lista de puntos $\mathbf{Q} \in \mathbb{A}^n(\mathbb{F}_q)^s$ sea un conjunto cuestor para $P_d(X_1, \dots, X_n)$ con respecto a $\{0\}$ es al menos:

$$1 - \frac{1}{q^{\dim(\Omega)}}.$$

Si, además, d, k, q y n verifican la siguiente desigualdad:

$$k \binom{d+n}{n} < \frac{q^n}{2^n},$$

entonces existen conjuntos cuestores para $P_d(X_1, \dots, X_n)$ con respecto a $\{0\}$ que no son conjuntos de Kakeya.

Por otro lado, en la Subsección 3.3.3, demostramos también que el Nullstellensatz Combinatorio de Alon (cf. [Al, 99]) está relacionado con los conjuntos cuestores. En particular, mostraremos cómo la dualidad de álgebras reducidas cero-dimensionales (véase también [Pa, 24]) está relacionada con el resultado principal de Alon en [Al, 99].

En la última sección del capítulo, esto es, en la Sección 3.4, presentamos un algoritmo probabilista que muestra tanto la potencia como las limitaciones de los conjuntos cuestores. Empleamos conjuntos cuestores de longitud asintóticamente óptima para diseñar un algoritmo que decide si una lista de ecuaciones es una “Suite Sécante”. Este problema puede enunciarse de la siguiente forma:

PROBLEMA 3 (Problema de la “Suite Sécante”). *Dado $f := (f_1, \dots, f_m) \in \mathcal{P}_{(d)}$, decidir si f es una “Suite Sécante”. Es decir, decidir si se verifica la siguiente propiedad:*

La variedad algebraica $V_{\mathbb{A}}(f_1, \dots, f_m) \subseteq \mathbb{A}^n(K)$ es una variedad no vacía de dimensión $n - m$.

Nuestro algoritmo probabilista es algo “extraño” si lo comparamos con las técnicas que se utilizan habitualmente en Geometría Algebraica Computacional. En este ámbito, la entrada de los algoritmos suele ser una lista de polinomios y el objetivo es determinar, entre otras cosas, la dimensión del conjunto afín de sus ceros comunes. Esto se lleva a cabo manipulando la lista de entrada de diferentes formas (calculando una base de Gröbner del ideal, utilizando variaciones de los Teoremas de Bertini para aproximar el conjunto de ceros mediante una variedad de intersección completa, etc...).

La característica principal de nuestro algoritmo es que responde al problema de la “Suite Sécante” realizando un número polinomial de operaciones aritméticas sin manipular la lista de los polinomios de entrada: *simplemente evalúa la lista de ecuaciones polinomiales f en unos puntos seleccionados aleatoriamente, es decir, calcula $f(x_1), \dots, f(x_L) \in K^m$ para algunos x_1, \dots, x_L* . Obviamente, la corrección de la respuesta no está garantizada: *depende de una probabilidad de error acotada.*

Por otro lado, nos restringimos al caso de *cuerpos que son adecuados para muestrear conjuntos cuestores*. Decimos que un cuerpo K es adecuado para muestrear conjuntos cuestores en variedades cero-dimensionales si verifica la siguiente propiedad: para cada $R \in \mathbb{N}$ y para cada entero positivo $n \in \mathbb{N}$, existe una variedad cero-dimensional $Z_R \subseteq \mathbb{A}^n(K)$ de grado R^n , dada por ecuaciones polinomiales de grado a lo sumo R , y de manera que la siguiente tarea pueda realizarse con a lo sumo $O(n \log_2(R))$ operaciones aritméticas:

muestrea aleatoriamente $x \in Z_R$,

con respecto a la distribución uniforme en Z_R .

El Teorema 3.4.2 muestra al mismo tiempo la potencia y las limitaciones de los conjuntos cuestores. Por un lado, los conjuntos cuestores son potentes ya que ayudan a obtener respuestas rápidas con probabilidad de error pequeña. Sin embargo, al mismo tiempo, para $m \geq 2$, no podemos garantizar que la respuesta sea correcta, ni siquiera en uno de los lados. Es decir, nuestro algoritmo probabilista no es del tipo Montecarlo. La necesidad de muestrear puntos de forma aleatoria se debe a que no se conocen algoritmos deterministas eficientes para calcular conjuntos cuestores (ver [Sax, 14] y sus referencias). A continuación, reproducimos el Teorema 3.4.2 para facilitar al lector el acceso a los resultados principales:

TEOREMA (DESTACABLE) 0.3.6. *Sea K un cuerpo adecuado para muestrear conjuntos cuestores. Sea $(d) := (d_1, \dots, d_m)$ una lista de grados y $\Omega \subseteq \mathcal{P}_{(d)}^K$ un subconjunto constructible de listas de polinomios. Sea $U \subseteq \mathcal{P}_{(d)}$ el subconjunto abierto Zariski de sistemas de ecuaciones polinomiales $f = (f_1, \dots, f_m)$ tales que:*

$$\dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Supongamos que $\Omega \setminus U$ tiene dimensión a lo sumo $\dim(\Omega) - 1$. Entonces, existe un algoritmo probabilista para el problema de la “Suite Sécante” con entradas restringidas a Ω que satisface las siguientes propiedades:

- i) Dado como entrada $f \in \Omega$ y las cantidades $\dim(\Omega)$ y $\deg_{\text{lci}}(\Omega)$:*
- *Si $f \in \Omega \cap U$, el algoritmo responde “ f es una Suite Sécante” con probabilidad mayor que:*

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{m \dim(\Omega)}}.$$

- *Si el algoritmo responde “ f no es una Suite Sécante”, entonces $f \in \Omega \setminus U$ (i.e. f no es “Suite Sécante”) con probabilidad mayor que:*

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{m \dim(\Omega)}}.$$

ii) Si $\mathcal{B} \subseteq 2^\Omega$ es una σ -álgebra que contiene a los subconjuntos de Borel de Ω con respecto a la topología de Zariski y $\mu : \mathcal{B} \rightarrow [0, 1]$ es una distribución de probabilidad sobre \mathcal{B} , entonces:

$$\mu[f \in \Omega : \text{el algoritmo responde de forma incorrecta sobre } f]$$

es a lo sumo:

$$\frac{1}{\deg_{\text{ici}}(\Omega)e^{m \dim(\Omega)}} + \mu[\Omega \setminus U].$$

iii) El tiempo de ejecución del algoritmo en términos de operaciones aritméticas es a lo sumo:

$$O(\dim(\Omega)n((T + \log(\dim(\Omega)) + \log(d)) + Tn \log(\deg_{\text{ici}}(\Omega))),$$

donde T es el número máximo de operaciones aritméticas requeridas para evaluar cualquier lista $f := (f_1, \dots, f_m) \in \Omega$ en un punto y $d := \max\{d_1, \dots, d_m\}$.

En el caso denso (i.e. $\Omega = \mathcal{P}_{(d)}^K$), el número total de operaciones aritméticas es del orden:

$$O(nN_{(d)}^2),$$

donde $N_{(d)} = \dim_K(\mathcal{P}_{(d)})$ y

$$\mu[f \in \mathcal{P}_{(d)}^K : \text{el algoritmo responde de forma incorrecta sobre } f]$$

está acotado por:

$$\frac{1}{e^{mN_{(d)}}} + \mu[\mathcal{P}_{(d)}^K \setminus U],$$

donde μ es una distribución de probabilidad en $\mathcal{P}_{(d)}^K$.

Por otra parte, mostramos en el Corolario 3.4.7 una aplicación del teorema anterior al caso en que Ω es un conjunto constructible de listas de ecuaciones polinómicas dadas por un esquema de evaluación de talla y profundidad acotada.

En el caso $m = 1$, el algoritmo está en la clase \mathbf{RP}_K . Con algunas pequeñas modificaciones, podemos transformar de manera sencilla nuestro algoritmo en otro que demuestre que el problema de la “Suite Sécante” está en la clase $\mathbf{RP}_K^{\text{HN}}$. Esto es, el problema de la “Suite Sécante” es decidible por un algoritmo probabilista del tipo Montecarlo usando oráculos que deciden si una lista de ecuaciones polinómicas define o no la variedad vacía (Nullstellensatz de Hilbert).

La complejidad de nuestro algoritmo está determinada por dos factores: el coste de muestrear una lista aleatoria de puntos $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq \mathbb{A}^n(K)$ y el coste de evaluar la lista de polinomios de entrada $f := (f_1, \dots, f_m)$ en los puntos de la lista muestreada. El segundo factor, esto es, el número de operaciones aritméticas requeridas para evaluar los polinomios en la lista de puntos, está relacionado con el problema de la evaluación multivariada en múltiples puntos (en inglés “multivariate multi-point evaluation”). Varios autores, en sucesivos artículos, han demostrado que la complejidad asintótica de la evaluación multivariada en múltiples puntos es arbitrariamente cercana a la lineal (ver el increíble trabajo realizado en [KU, 08], [KU, 11], [HL, 20], [NRS, 20], [HL, 21] y sus referencias). Dentro de algunos contextos, como el caso de cuerpos finitos “grandes” K , el tiempo de ejecución de nuestro algoritmo probabilista es casi lineal gracias a los avances de estos autores.

Nótese que, nuestro enfoque, es de alguna forma “inverso” a la evaluación multivariada en múltiples puntos, ya que nuestros conjuntos cuestores son una especie de “múltiples puntos” donde la interpolación está garantizada aunque no sea deseable (ver Proposición 3.1.2).

0.3.3. Capítulo 4: Erzeugungsgrad y dimensión de Vapnik-Chervonenkis. En este capítulo, establecemos conexiones entre la Teoría de la Intersección afín sobre cuerpos algebraicamente cerrados y la Teoría de Vapnik-Chervonenkis en el caso de familias de clasificadores binarios asociadas a clases parametrizadas de conjuntos constructibles. Estas conexiones nos permitirán responder afirmativamente a la Conjetura 1 (y al desafío de J. Heintz).

En el caso de la Geometría Algebraica Real y la Teoría de Vapnik-Chervonenkis, existen ciertas conexiones (cf. [GJ, 95], [CKKLW, 95], [KM, 97], [MPC, 08], [MP, 09] y sus referencias) que no son válidas para cuerpos de característica positiva y se basan en cotas para el número de componentes conexas de conjuntos semi-algebraicos (cf. [Mi, 64], [Th, 65], [Ol, 51], [OP, 62] y [Wa, 68]). En este capítulo, estableceremos conexiones válidas para cualquier característica. Para ello, emplearemos el Erzeugungsgrad, una noción de la Teoría de la Intersección definida por J. Heintz en [He, 83].

Antes de estudiar estas conexiones, debemos corregir un defecto que aparece en la formulación original del Erzeugungsgrad en [He, 83]. J. Heintz utilizó el \deg_z para conjuntos constructibles para definir esta noción. En el Ejemplo 2.1.1, mostramos que el \deg_z no satisface la Desigualdad de Bézout y, en consecuencia, su uso afecta tanto a los argumentos como a las cotas que involucran al Erzeugungsgrad. Dedicamos la Sección 4.1

a adaptar el Erzeugungsgrad y sus cotas combinatorias al caso de conjuntos constructibles usando nociones adecuadas para el grado (en particular, emplearemos el \deg_{LCI} que hemos introducido en la Definición 13).

Comenzamos la Subsección 4.1.1 generalizando la noción de grado LCI de un conjunto constructible a una familia finita de conjuntos constructibles:

DEFINICIÓN 2 (Grado de una familia finita de conjuntos constructibles). Sea \mathcal{F} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Sea \mathcal{C} una aplicación que asocia a cada subconjunto constructible $X \in \mathcal{F}$ una descomposición en subconjuntos localmente cerrados irreducibles de grado LCI mínimo de acuerdo a la Definición 13. Entonces, definimos la clase finita de variedades irreducibles asociadas a \mathcal{C} mediante la siguiente igualdad:

$$\mathcal{C}(\mathcal{F}, \mathcal{C}) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{F}, \exists W \in \mathcal{C}(X), V = \overline{W^z}\} = \bigcup_{X \in \mathcal{F}} \{\overline{W^z} : W \in \mathcal{C}(X)\}.$$

Definimos el grado de la familia \mathcal{F} con respecto a la aplicación \mathcal{C} como la suma de los grados de todas las variedades irreducibles $V \in \mathcal{C}(\mathcal{F}, \mathcal{C})$:

$$\deg(\mathcal{F}, \mathcal{C}) := \sum_{V \in \mathcal{C}(\mathcal{F}, \mathcal{C})} \deg(V).$$

Finalmente, definimos el grado de la familia \mathcal{F} como el mínimo de estos grados:

$$\deg(\mathcal{F}) := \min\{\deg(\mathcal{F}, \mathcal{C}) : \mathcal{C} \text{ asocia a cada } X \in \mathcal{F} \text{ una descomposición de grado LCI mínimo}\}.$$

Siguiendo las ideas de J. Heintz en [He, 83], también consideramos el Z -grado de una familia finita de conjuntos constructibles:

DEFINICIÓN 3 (Grado de las clausuras Zariski de una familia finita de conjuntos constructibles). Sea \mathcal{F} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Definimos:

$$\mathcal{D}(\mathcal{F}) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{F}, V \text{ es una componente irreducible de } \overline{X^z}\}.$$

El grado de las clausuras Zariski de una familia \mathcal{F} se define como la siguiente cantidad:

$$\deg_z(\mathcal{F}) := \sum_{V \in \mathcal{D}(\mathcal{F})} \deg(V).$$

Nótese que si denotamos por $\overline{\mathcal{F}^z} := \{\overline{X^z} : X \in \mathcal{F}\}$, entonces:

$$\deg_z(\mathcal{F}) = \deg(\overline{\mathcal{F}^z}),$$

que se corresponde con la noción original de [He, 83]. Las dos nociones de grado que acabamos de definir son sub-aditivas (cf. Lema 4.1.4) y coinciden en el caso de familias de conjuntos localmente cerrados (cf. Proposición 4.1.5).

A continuación, generalizamos la noción de celda definida en la Sección 3 de [He, 83] como sigue:

DEFINICIÓN 4 (\mathcal{H} -definible, \mathcal{H} -celda). Sea \mathcal{H} una clase finita de subconjuntos constructibles de $\mathbb{A}^n(K)$ y sea $C \subseteq \mathbb{A}^n(K)$ otro conjunto constructible.

- Sea $\mathcal{B}(\mathcal{H}, C)$ el álgebra booleana de los subconjuntos de C generados por \mathcal{H} . Un subconjunto X de C recibe el nombre de \mathcal{H} -definible en C si pertenece a esta álgebra booleana, i.e. $X \in \mathcal{B}(\mathcal{H}, C)$.
- Llamamos \mathcal{H} -celda en C a cualquier subconjunto $X \subseteq C$ de modo que existe $M \subseteq \mathcal{H}$ tal que:

$$X = C \cap \left(\bigcap_{Y \in M} Y \right) \cap \left(\bigcap_{Y \notin M} (C \setminus Y) \right).$$

Denotamos por $\mathcal{L}(\mathcal{H}, C)$ a la clase de todas las \mathcal{H} -celdas no vacías en C .

- Finalmente, definimos la clase $\text{Gen}(\mathcal{B}(\mathcal{H}, C))$ como la clase de todas las familias finitas de variedades algebraicas $\mathcal{W} := \{W_1, \dots, W_t\}$ tales que $\mathcal{B}(\mathcal{H}, C) \subseteq \mathcal{B}(\mathcal{W}, C)$.

Nótese que $\mathcal{L}(\mathcal{H}, C)$ define una partición de C y que cada elemento en $\mathcal{B}(\mathcal{H}, C)$ puede escribirse como la unión finita de \mathcal{H} -celdas en C . Finalmente, definimos la noción clave del capítulo, el Erzeugungsgrad de una familia de conjuntos constructibles, de la siguiente forma:

DEFINICIÓN 5 (Erzeugungsgrad). Sea \mathcal{H} una familia finita de conjuntos constructibles de $\mathbb{A}^n(K)$. Definimos el Erzeugungsgrad o “grado de generación” de \mathcal{H} como:

$$\text{grad}(\mathcal{H}) := \min\{\deg(\mathcal{W}) : \mathcal{W} \in \text{Gen}(\mathcal{B}(\mathcal{H}, \mathbb{A}^n(K))), \mathcal{W} \text{ es una familia finita de conjuntos cerrados}\}.$$

El lector debe observar que, en el caso de variedades algebraicas, el grado (en el sentido de [He, 83]) y el Erzeugungsgrad, que acabamos de introducir, son cantidades muy diferentes. El Erzeugungsgrad siempre es menor o igual que el grado, i.e. $\text{grad}(\mathcal{H}) \leq \text{deg}_z(\mathcal{H})$, donde \mathcal{H} es una clase formada por una sola variedad algebraica. En algunos casos, el Erzeugungsgrad puede ser extremadamente menor. A modo de ejemplo, consideremos el caso de una clase de conjuntos constructibles $\mathcal{H} := \{V\}$ formada por una sola variedad algebraica V de dimensión $n - r$. Supongamos, además, que esta variedad es de intersección completa, es decir, existen polinomios $f_1, \dots, f_r \in K[X_1, \dots, X_n]$ de grados respectivos d_1, \dots, d_r , tales que $V := V(f_1, \dots, f_r)$. Genéricamente, el grado de V (y, por tanto, el grado de la clase \mathcal{H}) es el número de Bézout:

$$\mathcal{D} := \prod_{i=1}^r d_i,$$

que es una cantidad exponencial en la co-dimensión. Sin embargo, $\text{grad}(\mathcal{H})$ está acotado por la suma:

$$\sum_{i=1}^r d_i,$$

que es lineal en la co-dimensión. Esta diferencia hace que el uso del Erzeugungsgrad para cotas combinatorias (como se hace en el Teorema 0.3.7 más abajo) permita obtener cotas mucho más finas que las que se obtendrían usando únicamente el grado como invariante. Nuestro resultado principal en este contexto es la siguiente generalización del Teorema 2 de [He, 83]. En nuestro resultado evitamos los problemas causados por el Z -grado y extendemos las cotas para celdas definidas dentro un conjunto constructible fijo $C \subseteq \mathbb{A}^n(K)$. Es importante señalar que este teorema conecta aspectos combinatorios con nociones de la Teoría de Intersección afín mediante cotas superiores (que no son necesariamente óptimas).

TEOREMA (DESTACABLE) 0.3.7 (Erzeugungsgrad y cotas combinatorias). *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y \mathcal{H} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Entonces, tenemos que:*

- i) $\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \text{deg}_{\text{lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$.
- ii) $\text{deg}_z(\mathcal{B}(\mathcal{H}, C)) \leq \text{deg}_{\text{lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$.
- iii) *Cualquier subconjunto finito \mathcal{H} -definible de C contiene a lo sumo $\text{deg}_{\text{lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$ puntos.*
- iv) $\sharp(\mathcal{B}(\mathcal{H}, C)) \leq 2^{\text{deg}_{\text{lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}}$.

A continuación, en la Sección 4.2 mostramos cómo el Erzeugungsgrad nos permite relacionar la Teoría de Intersección afín y la Teoría de Vapnik-Chervonenkis así como algunas aplicaciones.

Exponemos ahora algunos términos básicos de la Teoría de Vapnik-Chervonenkis para facilitar la lectura. La noción más importante de esta teoría es la dimensión de Vapnik-Chervonenkis:

DEFINICIÓN 6 (Desmenuzar, Dimensión de Vapnik-Chervonenkis). *Sea Y un conjunto y $\mathcal{H} \subseteq \{0, 1\}^Y$ una familia de funciones características de subconjuntos de Y , llamadas comúnmente clasificadores (binarios) sobre Y . Dado un subconjunto finito $X \subseteq Y$, decimos que \mathcal{H} desmenuza X si para cada $F \subseteq X$, existe un clasificador $\chi \in \mathcal{H}$ tal que la restricción a X de χ es igual a la restricción a X de χ_F . Es decir, \mathcal{H} desmenuza X si para cada $F \subseteq X$ existe un $\chi \in \mathcal{H}$ tal que:*

$$\chi|_X = \chi_F|_X.$$

Definimos la dimensión de Vapnik-Chervonenkis de una familia de clasificadores \mathcal{H} como:

$$\text{dim}_{VC}(\mathcal{H}) := \max\{\sharp(X) : \mathcal{H} \text{ desmenuza } X\}.$$

Por otro lado, la noción clave que usaremos en nuestra discusión es la función de crecimiento:

DEFINICIÓN 7 (Función de crecimiento). *Sea Y un conjunto y $\mathcal{H} \subseteq \{0, 1\}^Y$ una familia de clasificadores. Para cada subconjunto $X \subseteq Y$, sea $\mathcal{H}|_X$ la clase de restricciones a X de las funciones en \mathcal{H} . Es decir,*

$$\mathcal{H}|_X := \{f|_X : f \in \mathcal{H}\}.$$

La función de crecimiento $G(\mathcal{H}, \cdot)$ de una familia de clasificadores \mathcal{H} es la función:

$$G(\mathcal{H}, \cdot) : \mathbb{N} \rightarrow \mathbb{N},$$

dada por la siguiente identidad para cada entero positivo $m \in \mathbb{N}$:

$$G(\mathcal{H}, m) := \sup\{\sharp(\mathcal{H}|_X) : X \subseteq Y, \sharp(X) = m\}.$$

En la Subsección 4.2.1, consideramos familias de clasificadores binarios definidas por clases parametrizadas de conjuntos constructibles. Además, tratamos el caso en el que el conjunto de los parámetros también es un conjunto constructible. La idea de parámetro hace referencia a la existencia de alguna “función” φ que parte de una clase de parámetros Λ y para cada $\lambda \in \Lambda$ asocia un conjunto constructible $\varphi(\lambda) \subseteq \mathbb{A}^n(K)$. Esto lleva, de manera natural, a pensar en una variedad de incidencia $V(\varphi)$, asociada al “grafo” de la parametrización, de tal modo que $V(\varphi) \subseteq \Lambda \times \mathbb{A}^n(K)$, donde para cada $\lambda \in \Lambda$, los elementos de la forma $(\{\lambda\} \times \mathbb{A}^n(K)) \cap V(\varphi)$ identifican a los elementos del conjunto constructible $\varphi(\lambda)$. Hemos escogido esta interpretación por ser más general y ajustarse a la idea de parametrización. Por ello, procedemos del siguiente modo para definir familias de constructibles parametrizadas por otro constructible:

Sean $N, n \in \mathbb{N}$ dos enteros positivos y $V \subseteq \mathbb{A}^N(K) \times \mathbb{A}^n(K)$ un conjunto constructible. Por ejemplo, V puede ser el grafo de una aplicación constructible. Consideramos también un subconjunto constructible $\Lambda \subseteq \mathbb{A}^N(K)$, al que llamamos *conjunto constructible formado por los parámetros de nuestra familia de clasificadores*. Además, definimos también las dos proyecciones canónicas restringidas a V :

$$\begin{array}{ccc} & V & \\ \pi_1 \swarrow & & \searrow \pi_2 \\ \mathbb{A}^N(K) & & \mathbb{A}^n(K) \end{array}$$

Con estas notaciones, definimos la siguiente clase de conjuntos constructibles:

$$\mathcal{C}(V, \Lambda) := \{\pi_2(\pi_1^{-1}(\{a\})) : a \in \Lambda\}.$$

Definimos también $\text{grad}(V) = \text{grad}(\{V\})$ como el Erzeugungsgrad de la clase definida por V . Finalmente, definimos la clase de clasificadores dada por las funciones características definidas por los subconjuntos de $\mathcal{C}(V, \Lambda)$:

$$\mathcal{H}(V, \Lambda) := \{\chi_U : U \in \mathcal{C}(V, \Lambda)\}.$$

Nuestra principal contribución en la Subsección 4.2.1 es el Teorema 4.2.1, que reproducimos a continuación:

TEOREMA (DESTACABLE) 0.3.8. *Con las mismas notaciones e hipótesis que antes, se cumple la siguiente desigualdad:*

$$\#(\mathcal{H}(V, \Lambda)|_X) \leq \text{deg}_{\text{lci}}(\Lambda)(1 + \#(X) \text{grad}(V))^{\dim(\Lambda)}.$$

Además, tenemos que:

$$G(\mathcal{H}(V, \Lambda), m) \leq \text{deg}_{\text{lci}}(\Lambda)(1 + m \text{grad}(V))^{\dim(\Lambda)}$$

Sin embargo, la conclusión más descriptiva se presenta en el Corolario 4.2.2, que resuelve de forma afirmativa la Conjetura 1 (y el desafío de J. Heintz): *salvo por ciertas cantidades logarítmicas basadas en la Teoría de la Intersección, la dimensión de Vapnik-Chervonenkis de una familia de clasificadores binarios asociada a una clase parametrizada de conjuntos constructibles está acotada por la dimensión del espacio de parámetros*. Es decir, tenemos que:

COROLARIO (DESTACABLE) 0.3.9. *Con las notaciones e hipótesis precedentes, se verifica la siguiente desigualdad:*

$$\frac{s}{\log_2(s) + k} - \frac{\log_2(\text{deg}_{\text{lci}}(\Lambda))}{\log_2(s) + k} \leq \dim(\Lambda),$$

donde $s = \dim_{VC}(\mathcal{H}(V, \Lambda))$ y $k = 1 + \log_2(\text{grad}(V))$.

El resto de la Subsección 4.2 está dedicada a exponer aplicaciones de los resultados previos. En la Subsección 4.2.2, aplicamos estos resultados al estudio de la función de crecimiento y de la dimensión de Vapnik-Chervonenkis de clases parametrizadas de polinomios.

En la Subsección 4.2.3 estudiamos la noción de variedad evasiva. Esencialmente, una variedad evasiva (cf. Definición 26) es una variedad algebraica que no está completamente embebida en ninguna variedad algebraica de intersección completa definida por polinomios de un conjunto constructible. Posteriormente, en la Subsección 4.2.4, presentamos nuestra principal aplicación de los resultados anteriores. Demostramos que los conjuntos cuestores están densamente distribuidos en cualquier *variedad evasiva* irreducible de *dimensión positiva* con respecto a cualquier distribución de probabilidad con “buen comportamiento” con respecto a subvariedades propias. El ingrediente clave en la demostración es la conexión entre la dimensión de Krull y la dimensión de Vapnik-Chervonenkis en el caso de clases parametrizadas por polinomios. Cabe destacar, que los estudios clásicos sobre conjuntos cuestores se han realizado para conjuntos cero-dimensionales (como en [HS, 82], [KP, 96] y sus referencias). La novedad es que extendemos estos resultados a dimensión positiva y a cualquier distribución de probabilidad sobre la variedad con “buen comportamiento” con respecto a subvariedades propias.

TEOREMA (DESTACABLE) 0.3.10. *Sea $\Omega \subseteq P_d^K(X_1, \dots, X_n)$ un conjunto constructible de polinomios de grado a lo sumo d . Sea $V \subseteq \mathbb{A}^n(K)$ una variedad irreducible de dimensión positiva que es evasiva para hipersuperficies en Ω . Sea $\mathcal{B} \subseteq 2^V$ una σ -álgebra que contenga a los subconjuntos de Borel de V con respecto a la topología de Zariski. Sea $\mu : \mathcal{B} \rightarrow [0, 1]$ una distribución de probabilidad sobre \mathcal{B} que verifica la siguiente propiedad: para cada cerrado Zariski $A \subseteq V$, si $\dim(A) < \dim(V)$, entonces $\mu(A) = 0$. Sea $L \in \mathbb{N}$ un entero positivo, $\mathcal{B}^{\otimes L}$ la σ -álgebra en el producto V^L inducida por (V, \mathcal{B}) y $\mu^{\otimes L}$ la distribución de probabilidad definida sobre $\mathcal{B}^{\otimes L}$ por $\mu : \mathcal{B} \rightarrow [0, 1]$. Sea $R(\Omega, V, L)$ la clase de todas las listas $\mathbf{Q} \in V^L$ que son conjuntos cuestores para Ω con respecto a $\{0\}$. Supongamos que se satisface la siguiente desigualdad:*

$$(0.3.4) \quad 64 \left(1 + \frac{1 + \log(\deg_{\text{lci}}(\Omega))}{\dim(\Omega)} + \log(L(d+1)) \right) < \frac{L}{\dim(\Omega)}$$

donde \log denota al logaritmo neperiano. Entonces, se verifica la siguiente desigualdad:

$$\text{Prob}_{x \in V^L} [x \in R(\Omega, V, L)] \geq - \frac{1}{\deg_{\text{lci}}(\Omega) e^{\dim(\Omega)}}.$$

Es importante destacar que, en el teorema precedente, se obtiene el mismo nivel de error de probabilidad que en el caso cero-dimensional (cf. Corolario 3.2.6).

0.3.4. Capítulo 5: Redes Neuronales con función de activación racional y conjuntos constructibles de parámetros. En este capítulo, aplicamos las ideas introducidas y trabajadas en los capítulos precedentes para estudiar familias parametrizadas de redes neuronales con funciones de activación polinómicas y racionales.

Las redes neuronales son la estructura de datos más exitosa en la Teoría del Aprendizaje Computacional. En la Sección 5.1 hemos tratado de fijar formalmente la noción de red neuronal. La razón es que no hemos encontrado una definición precisa y formalmente matemática, adaptada a toda la casuística conocida, en la literatura usual de la Teoría del Aprendizaje. Creemos que nuestra definición recoge todos los casos particulares discutidos en la vastísima literatura del tema.

Una red neuronal con variables de entrada $\{X_1, \dots, X_n\}$ es una terna $\mathcal{N} := (G, \mathcal{A}, \Phi)$, donde:

- G es un grafo dirigido acíclico. Sus vértices y aristas son dados por:

$$V := \{(i, j) : 0 \leq i \leq \ell, 1 \leq j \leq L_i\}, \quad E \subseteq \{(\nu_1, \nu_2) \in V^2 : d(\nu_1) \leq d(\nu_2) - 1\},$$

donde $d(\nu)$ indica la profundidad del nodo ν .

- \mathcal{A} es la clase de funciones de activación.
- Φ es una aplicación que asocia a cada nodo ν del grafo G , de profundidad mayor que 1, un par formado por una función de activación y una lista de parámetros asociada a su abanico de entrada $\text{Fan} - \text{In}(\nu)$. Los nodos de profundidad 0 están asociados a las variables en $\{X_1, \dots, X_n\}$ y a la constante 1. Esto es,

$$\Phi(\nu) := \begin{cases} \left(f_\nu, \left\{ A_\nu^{(i,j)} : (i,j) \in \text{Fan} - \text{In}(\nu) \right\} \right) & \text{si } d(\nu) \geq 1 \\ 1, & \text{si } \nu = (0, 0) \\ X_j, & \text{si } \nu = (0, j), 1 \leq j \leq n \end{cases}.$$

Las variables paramétricas $A_\nu^{(i,j)}$ reciben habitualmente el nombre de “pesos de la red neuronal”.

Por otro lado, tratamos la semántica de una red neuronal en la Sección 5.2. En este caso, la idea clave es la *instanciación de parámetros*. A grandes rasgos, instanciar una red neuronal en el vector de parámetros:

$$\underline{a} := (a_\nu^{(i,j)} : \nu \in V, (i,j) \in \text{Fan} - \text{In}(\nu)),$$

consiste en evaluar las variables $A_\nu^{(i,j)}$ en los valores $a_\nu^{(i,j)}$. Además, dado un nodo $\nu \in V$, le asociamos una función instanciada $f_\nu(\underline{a}, X_1, \dots, X_n)$ definida en términos de la profundidad como sigue:

- Si ν es un nodo de profundidad 0: $f_{(0,0)}(\underline{a}, X_1, \dots, X_n) := 1$, $f_{(0,j)}(\underline{a}, X_1, \dots, X_n) := X_j$, $1 \leq j \leq n$.
- En otro caso,

$$f_\nu(\underline{a}, X_1, \dots, X_n) := f_\nu \left(\sum_{(i,j) \in \text{Fan} - \text{In}(\nu)} a_\nu^{(i,j)} \cdot f_{(i,j)}(\underline{a}, X_1, \dots, X_n) \right).$$

Sea $O_{\mathcal{N}}$ el conjunto de nodos de salida de la red neuronal \mathcal{N} , con $m = \#(O_{\mathcal{N}})$, y \underline{a} un vector de parámetros. La red neuronal \mathcal{N} instanciada en \underline{a} evalúa la siguiente función:

$$\begin{aligned} f_{\mathcal{N}}(\underline{a}, \cdot) : \text{Dom}(f_{\mathcal{N}}(\underline{a}, \cdot)) \subseteq \mathbb{A}^n(K) &\longrightarrow K^m \\ x &\longmapsto (f_\nu(\underline{a}, x_1, \dots, x_n) : \nu \in O_{\mathcal{N}}) \end{aligned}$$

Dado un conjunto Λ de parámetros, denotamos por $\mathcal{W}(\mathcal{N}, \Lambda)$ a la clase de todas las funciones evaluadas por \mathcal{N} con parámetros en Λ . Uno de los problemas centrales de la Teoría del Aprendizaje Computacional es el estudio de las propiedades de aprendizaje y aproximación de las clases $\mathcal{W}(\mathcal{N}, \Lambda)$.

Una vez que hemos establecido las notaciones y el formalismo matemático de las redes neuronales, nos centramos en la clase de *funciones de activación* empleadas por una familia de redes neuronales. En la Sección 5.3, presentamos varios ejemplos de redes neuronales, usando nuestro formalismo, habituales en la literatura matemática. En particular, nos enfocamos en las redes neuronales “algebraicas” (i.e. su función de activación es un polinomio o una función racional), ya que son las más relacionadas con nuestro estudio previo. En el Ejemplo 5.3.2 tratamos las redes neuronales con función de activación $\varphi(t) = t^2$. Este tipo de redes neuronales coinciden con la clase de *esquemas de evaluación* definidos en la Sección 3 de [KP, 96]. Estos esquemas de evaluación son centrales en la Teoría de la Eliminación, ya que todos los *polinomios de eliminación* (cf. [GH, 91] y [KP, 96]) admiten una codificación como redes neuronales bien paralelizables con función de activación $\varphi(t) = t^2$. Esta es la idea hay que hay detrás del algoritmo clásico TERA (cf. [GH, 91], [KP, 96], [Pa, 95], [GHMP, 95], [GHMP, 97], [GHHMMP, 97], [GHMMP, 98] o [HMPS, 00]).

Posteriormente, en la Sección 5.4 estudiamos en detalle las propiedades del conjunto constructible $\mathcal{W}(\mathcal{N}, \Lambda)$ en el caso redes neuronales con función de activación polinómica. Además, discutimos la función del crecimiento de la clase de clasificadores binarios asociada a $\mathcal{W}(\mathcal{N}, \Lambda)$ así como la longitud y la densidad de conjuntos cuestores para este tipo de redes neuronales. En el caso polinomial, es importante destacar su relación con el clásico Teorema de Stone-Weierstrass. Si consideramos un espacio compacto (por ejemplo, la esfera unidad compleja) y el conjunto separante (es decir, funciones que distinguen puntos) formado por las proyecciones X_1, \dots, X_n , entonces el Teorema de Stone-Weierstrass nos asegura que el álgebra generada por estas funciones es densa en el espacio de funciones continuas sobre X . Esto implica que podemos aproximar funciones continuas mediante polinomios. En particular, la función de activación t^2 permite expresar multiplicaciones $X \cdot Y$ (ver Ejemplo 5.3.2). Esto quiere decir, que las redes neuronales con función de activación $\varphi(t) = t^2$ constituyen el ejemplo “más sencillo” de redes neuronales que “satisfacen” el Teorema de Stone-Weierstrass y, por lo tanto, permiten aproximar funciones continuas.

Finalmente, en la Sección 5.5, nos centramos en la clase de redes neuronales con función de activación racional (esto es, la función de activación es $\varphi = p/q$, donde p y q son dos polinomios univariados en $K[T]$). En algunos casos, las funciones de activación racionales son evitables y podemos aplicar la técnica “Vermeidung von Divisionen” de [St, 73] (véase su adaptación al caso de redes neuronales en [KP, 96]). En estos casos, podemos evitar la presencia de divisiones y reducir el problema al caso de función de activación cuadrática (i.e. $\varphi(t) = t^2$). Por lo tanto, la novedad de esta sección es considerar redes neuronales con funciones de activación racionales *no evitables*. Este tipo de redes neuronales es la estructura de datos natural para representar funciones racionales multivariadas (es decir, funciones en $K(X_1, \dots, X_n)$). En el Lema 5.5.3 demostramos que dada una red neuronal \mathcal{N} con función de activación racional, existe otra red neuronal “doble” con función de activación $\varphi(t) = t^2$ que calcula los numeradores y denominadores de las funciones evaluadas por \mathcal{N} . Usaremos esta construcción para demostrar nuestros resultados principales. En el Corolario 5.5.4 estudiamos la función de crecimiento de la clase de clasificadores binarios asociados a una red neuronal con función de activación racional:

COROLARIO (DESTACABLE) 0.3.11. *Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y un único nodo de salida. Supongamos que \mathcal{N} tiene N aristas y que $\mathcal{A} = \{\varphi\}$, donde $\varphi := \frac{p}{q}$ es una función racional univariada, dada por el cociente de dos polinomios univariados coprimos, de grado d (i.e. $\deg(\varphi) = \max\{\deg(p), \deg(q)\} = d$). Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de los parámetros de \mathcal{N} y sea $\mathcal{W}(\mathcal{N}, \Lambda)$ la clase de todas las funciones racionales evaluables por \mathcal{N} con parámetros en Λ . Sea \mathcal{H} la clase de clasificadores dada por las funciones características de los subconjuntos abiertos distinguidos de $\mathbb{A}^n(K)$ definidos por funciones racionales en $\mathcal{W}(\mathcal{N}, \Lambda)$:*

$$\mathcal{H} = \{\chi_{D(h)} : h \in \mathcal{W}(\mathcal{N}, \Lambda)\},$$

Entonces,

$$G(\mathcal{H}, m) \leq \deg_{\text{ici}}(\Lambda) \left((2(dS)^{c\ell} - 2) \cdot (1 + 2m((dS)^\ell + 1)) \right)^{LS},$$

donde $c > 0$ es una constante universal.

Por otro lado, también discutimos la existencia y la densidad de conjuntos cuestores para familias parametrizadas de redes neuronales con función de activación racional en el Corolario 5.5.5. En el Corolario 5.5.6 empleamos los conjuntos cuestores como Test de Identidad de funciones racionales dadas por redes neuronales. En el contexto de la Teoría de Aprendizaje Computacional, este último resultado nos permite comprobar si dos redes neuronales con función de activación racional han “aprendido” la misma función.

0.3.5. Capítulo 6: Un enfoque alternativo a la relación entre aprendizaje y polinomios multivariados: Reescritura y Aprendizaje multiclase. Grado de salida promedio, dimensiones, shifting y relaciones con bases monomiales de K -álgebras de Artin. Este capítulo difiere metodológicamente de los anteriores. Mientras que en los capítulos previos se empleaban la Teoría de la Intersección y los conjuntos constructibles como substrato matemático, aquí se inicia la conexión entre el Aprendizaje Computacional y el contexto de la manipulación de representaciones monomiales de polinomios: lo que, conforme al pensamiento iniciado en [MM, 82], conduce a los “*word problems for commutative semigroups*” o, en palabras de varios autores del ámbito del álgebra computacional, a las “*técnicas de reescritura*” (o “*rewriting techniques*”).

El capítulo difiere asimismo en sus objetivos respecto a los anteriores. En los Capítulos del 1 al 5 se construyen puentes entre la Teoría de la Intersección, la búsqueda de conjuntos cuestores y el Aprendizaje Computacional binario. En cambio, este capítulo aborda una refundación del Aprendizaje multiclase a partir de una noción novedosa (*el grado de salida promedio*) y sus relaciones con las diversas nociones de dimensión conocidas en el ámbito, fundamentalmente la dimensión de Daniely-Shalev-Schwartz y la dimensión exponencial.

Finalmente, este capítulo difiere de los anteriores en las fuentes bibliográficas que lo inspiran y en el potencial impacto de los resultados originales que aquí se muestran. Recientemente, ha habido un fuerte impulso en la fundamentación del Aprendizaje Computacional multiclase, concretamente en los trabajos [DS, 14] y [BCDMY, 22], donde los autores reanalizan las propuestas clásicas de B. K. Natarayan (cf. [Na, 88] y [Na, 89]) utilizando nuevas ideas combinatorias con el propósito de refundar esta parte del Aprendizaje Computacional. Por otra parte, aunque los resultados originales que se exponen son prometedores, aún no tenemos claridad sobre su impacto futuro: son las bases para el desarrollo de *Futuras Líneas de Investigación*. En particular, pretendemos explorar en futuras investigaciones la conexión profunda entre el Aprendizaje multiclase y los polinomios dados en codificación *fewnomial* de A. Khovanskii (cf. [Kh, 91]). Nuestro objetivo es realizar una publicación que hemos incluido en la bibliografía como [PSZ, 25].

En el transcurso de la investigación presentada en este capítulo, hemos identificado y desarrollado nuevas nociones, caracterizado diversos conceptos y obtenido resultados originales en el contexto del Aprendizaje multiclase. Además, hemos logrado mejoras significativas con respecto a la literatura existente (cf. [DS, 14] y [BCDMY, 22]), que es muy reciente.

En este capítulo se discuten las propiedades elementales de las nociones que surgen en el análisis del *hiper-grafo de una inclusión* (la estructura de datos subyacente al Algoritmo 0.2.2), asociado a una clase de conceptos, en el Aprendizaje multiclase. La introducción del “hiper-grafo de una inclusión” asociado a una clase de conceptos binarios se remonta a [Ha, 95, HLW, 94]. Posteriormente, esta noción fue extendida al contexto del Aprendizaje multiclase en [RBR, 06] y discutida en [DS, 14] y [BCDMY, 22]. Seguimos esencialmente la terminología de [BCDMY, 22] con algunas variaciones notacionales.

En particular, nos centraremos en el estudio de dos nociones ligadas al hiper-grafo de una inclusión del Aprendizaje multiclase. La primera de las nociones está bien establecida en la literatura matemática y es simplemente el *promedio de los grados* del hiper-grafo de una inclusión (que denotamos por $\text{avd}(h)$). La segunda noción, en cambio, es original y se trata *grado de salida promedio* de cualquier orientación del hiper-grafo de una inclusión (que denotamos por $E_H[\text{outdeg}(\sigma)]$). Esta segunda noción coincide con la noción auxiliar avd' definida en [BCDMY, 22], aunque estos autores no la relacionan con las orientaciones ni con la densidad del hiper-grafo, y tampoco la estudian o la consideran una noción central.

Antes de presentar nuestros resultados, es necesario que introduzcamos las notaciones básicas que vamos a emplear. En primer lugar, fijamos el rango de las funciones que estudiamos en el Aprendizaje multiclase. Consideramos el conjunto $\mathcal{Y} := \{0, \dots, p-1\} = \mathbb{Z}/p\mathbb{Z}$, donde $p \in \mathbb{N}$ es un entero positivo (no necesariamente primo). Habitualmente, la mayoría de los autores consideran el conjunto $\mathcal{Y} = [p] := \{1, \dots, p\}$. Sin embargo, hemos decidido modificar el rango para resaltar la relación de esta terminología con el estudio de bases monomiales.

A continuación, consideramos una clase de conceptos finita $H \subseteq \mathcal{Y}^n := \mathcal{Y}^{[n]}$, con $n \geq 2$, formada por funciones $h : [n] \rightarrow \mathcal{Y}$. Obviamente, cada una de estas funciones se puede identificar con un vector $(h(1), \dots, h(n)) \in (\mathbb{Z}/p\mathbb{Z})^n$. En esta memoria, usaremos ambas interpretaciones indistintamente. Como ya hemos comentado antes, asociado a la clase de conceptos H existe un *hiper-grafo de una inclusión* $G(H) = (H, E(H))$. Los vértices de este hiper-grafo son los elementos de H mientras que sus aristas pueden definirse de la siguiente forma. Sea $\pi_i : \mathcal{Y}^{[n]} \rightarrow \mathcal{Y}^{n-1}$ la proyección que “olvida” la coordenada i -ésima. Para cada $f \in \mathcal{Y}^{n-1}$, sea $e_{f,i} := \pi_i^{-1}(\{f\}) \cap H \neq \emptyset$ la fibra sobre f en H , siempre que esta fibra sea no vacía (o, equivalentemente, dado que $f \in \pi_i(H)$). El conjunto de aristas $E(H)$ se define mediante la siguiente igualdad:

$$E(H) := \{e_{f,i} \times \{i\} : f \in \pi_i(H), i \in [n]\}.$$

Para todo $e := e_{f,i} \times \{i\} \in E(H)$ y todo $h \in H$, decimos que $h \in e$ si y solo si $h \in e_{f,i}$ (o, equivalentemente, si y solo si $\pi_i(h) = f$). Denotamos por $\sharp(e)$ a la cardinalidad del subconjunto subyacente $e_{f,i} \subseteq H$, donde $e = e_{f,i} \times \{i\}$. Una arista $e \in E(H)$ recibe el nombre de *unaria* si verifica $\sharp(e) = 1$. Denotamos por $E^{(1)}(H)$ a la clase de aristas en $E(H)$ que son unarias. La estructura de hiper-grafo que acabamos de definir no es propiamente un hiper-grafo en sentido formal, ya que las aristas unarias pueden ser “contadas” varias veces (cf. Observación 6.2.1).

Una orientación del hiper-grafo de una inclusión $G(H)$ es una aplicación $\sigma : E(H) \rightarrow H$ que satisface la siguiente propiedad:

$$\forall e \in E(H), \sigma(e) \in e.$$

El objetivo de los algoritmos de aprendizaje discutidos en [DS, 14] y [BCDMY, 22] (cf. Algoritmo 0.2.2) es encontrar una orientación que verifique ciertas propiedades. A continuación, pasamos a introducir el resto de nociones asociadas a nuestro hiper-grafo de una inclusión:

- El *grado de un vértice* $h \in H$ se define como:

$$\deg(h) := \sharp\{e \in E : h \in e, \sharp(e) \geq 2\}.$$

- El *grado promedio* (con respecto a la probabilidad uniforme en H) del hiper-grafo $G(H)$ es:

$$\text{avd}(H) := \frac{1}{\sharp(H)} \sum_{h \in H} \deg(h).$$

- El *grado de salida* de una orientación σ en un vértice $h \in H$ se define de la siguiente forma:

$$\text{outdeg}(\sigma, h) := \sharp(\{e \in E(H) : h \in e, \sigma(e) \neq h\}).$$

- El *grado de salida promedio* (con respecto a la probabilidad uniforme en H) de una orientación σ viene dada por:

$$E_H[\text{outdeg}(\sigma)] := \frac{1}{\sharp(H)} \sum_{h \in H} \text{outdeg}(\sigma, h).$$

- La *densidad del hiper-grafo* $G(H)$ se define como el siguiente número racional:

$$\text{gd}(H) := \frac{\sharp(E(H))}{\sharp(H)}.$$

El grado de salida máximo de una orientación es también relevante, especialmente para diseñar algoritmos voraces:

$$\text{outdeg}_{\max}(\sigma) = \max\{\text{outdeg}(\sigma, h) : h \in H\}.$$

Presentamos ahora el primer resultado original relevante (cf. Proposición 6.2.5) del capítulo. Observamos no solo que $E_H[\text{outdeg}]$ es independiente de la orientación σ sino que está ligado a la densidad del hiper-grafo (introducida en [RBR, 06]).

PROPOSICIÓN (DESTACABLE) 0.3.12 (Principales caracterizaciones de $\text{avd}(H)$, $E_H[\text{outdeg}]$ y $\text{gd}(H)$). *Con las notaciones anteriores y suponiendo que $n \geq 2$, tenemos que:*

- i) *El grado promedio del hiper-grafo de una inclusión $G(H)$ verifica:*

$$\text{avd}(H) = n - \frac{\sharp(E^{(1)}(H))}{\sharp(H)}.$$

- ii) *El grado de salida promedio de cualquier orientación $\sigma : E(H) \rightarrow H$ satisface:*

$$E_H[\text{outdeg}(\sigma)] = n - \text{gd}(H) = n - \frac{\sharp(E(H))}{\sharp(H)}.$$

En particular, el grado de salida promedio de una orientación es independiente de la orientación escogida y solo depende de la dimensión n del espacio ambiente \mathcal{V}^n y de la densidad del hiper-grafo $\text{gd}(H)$. Por lo tanto, a partir de ahora, escribiremos simplemente $E_H[\text{outdeg}]$ para denotar al grado de salida promedio de cualquier orientación del hiper-grafo. En particular, tenemos que:

$$\text{avd}(H) - E_H[\text{outdeg}] - \text{gd}(H) = -\frac{\sharp(E^{(1)}(H))}{\sharp(H)} \leq 0.$$

- iii) *También se cumplen las siguientes desigualdades:*

$$E_H[\text{outdeg}] \leq \text{avd}(H) \leq 2E_H[\text{outdeg}].$$

iv) Para cualquier orientación σ , tenemos que:

$$\text{gd}(H) - \frac{\#(E^{(1)}(H))}{\#(H)} \leq E_H[\text{outdeg}].$$

v) Sea $G \subseteq H$ una subclase de H y sea $\sigma : E(H) \rightarrow H$ una orientación. Entonces, existe una orientación $\sigma' : E(G) \rightarrow G$ de modo que se verifica:

$$\forall g \in G, \text{outdeg}(\sigma', g) \leq \text{outdeg}(\sigma, g).$$

En particular, para toda orientación $\sigma : E(H) \rightarrow H$, existe una orientación $\sigma' : E(G) \rightarrow G$ tal que:

$$\text{outdeg}_{\max}(\sigma') \leq \text{outdeg}_{\max}(\sigma).$$

Finalmente, el grado de salida promedio raramente coincide con el grado promedio, ya que se cumple que:

- O bien

$$E_H[\text{outdeg}] < \text{avd}(H),$$

- o bien $E(H) = E^{(1)}(H)$ (i.e. $G(H)$ tiene solo aristas unarias), en cuyo caso:

$$E_H[\text{outdeg}] = \text{avd}(H) = 0.$$

Cabe destacar que los valores del grado promedio y del grado de salida promedio dependen fuertemente de la distribución uniforme escogida sobre H . En la Observación 6.2.8 mostramos que ni el grado promedio ni el grado de salida promedio son independientes de la distribución de probabilidad definida sobre H . En el Lema 6.2.9 y el Ejemplo 6.2.10 demostramos que el Ítem ii) de la Proposición 6.2.5 es independiente de la distribución de probabilidad definida en H . Por otro lado, la primera desigualdad del Ítem iii) también sigue siendo cierta si cambiamos la distribución de probabilidad, pero esto no sucede con la segunda, como se muestra en el Ejemplo 6.2.13.

Otra noción de interés para nuestro estudio es la de *conjunto cerrado hacia abajo* (o escalera). En [Di, 1913], el autor definió un pre-orden sobre \mathbb{N}^n , usualmente denominado pre-orden de Dickson y denotado por \preceq . Este pre-orden ha sido útil para demostrar propiedades de finitud de los exponentes de ideales de polinomios multivariados (véase la serie de cuatro volúmenes que culmina con [Mo, 16] o algún texto introductorio como [CLO, 15]). Una clase finita de conceptos $H \subseteq \mathcal{Y}^{[n]}$ se dice cerrada hacia abajo si y solo si es cerrada con respecto al pre-orden de Dickson. Las clases cerradas hacia abajo finitas $H \subseteq \mathcal{Y}^{[n]}$ son exactamente los complementarios de los conjuntos de exponentes de ciertos ideales en $K[X_1, \dots, X_n]$, donde K es un cuerpo. Entre los conjuntos cerrados hacia abajo más relevantes, consideramos los *escalones* \mathcal{E}_h . Dado $h \in \mathbb{N}^n$, definimos el escalón \mathcal{E}_h como el conjunto de todos los elementos en \mathbb{N}^n menores o iguales que h con respecto al pre-orden de Dickson. Los escalones son relevantes ya que cualquier conjunto cerrado hacia abajo $H \subseteq \mathcal{Y}^{[n]}$ tiene una descomposición única como unión finita de escalones determinados por los elementos maximales de H con respecto al pre-orden de Dickson, también llamados componentes maximales de la escalera H (cf. Proposición 6.1.2). Discutimos el grado promedio y el grado de salida promedio de los escalones en la Proposición 6.2.18.

En lo que sigue, consideramos *funciones monótonas con respecto a la inclusión de conjuntos*. Una función $\varphi : 2^H \rightarrow \mathbb{R}$ recibe el nombre de monótona (con respecto a la inclusión de conjuntos) si se verifica la siguiente propiedad:

$$(0.3.5) \quad \forall F, G \in 2^H, \text{ si } F \subseteq G, \text{ entonces } \varphi(F) \leq \varphi(G).$$

Desafortunadamente, ni el grado promedio ni el grado de salida promedio son funciones monótonas con respecto a la inclusión de conjuntos. Esto se discute en detalle en la Proposición 6.2.16, mostrando condiciones suficientes que incrementan o disminuyen estas cantidades al eliminar un elemento $h \in H$. En este contexto, las funciones monótonas naturales son las *distintas nociones de dimensión*, las cuales se tratan en la Sección 6.3.

Antes de abordar las diversas nociones de dimensión, discutimos la noción de pseudo-cubo. Esta noción fue empleada implícitamente en [DS, 14] sin definirla de forma explícita. Será [BCDMY, 22] quien, retomando las ideas de [DS, 14], la defina por primera vez. La noción es bastante inusual en Matemática Discreta y nunca ha sido estudiada de manera sistemática. Como es una noción nueva en Aprendizaje multiclase, realizamos un estudio sistemático en este capítulo. En la Proposición 6.3.5 presentamos las siguientes caracterizaciones de la noción de pseudo-cubo (y, a partir del Lema 12 de [BCDMY, 22], deducimos la última desigualdad):

PROPOSICIÓN (DESTACABLE) 0.3.13 (Caracterización de los pseudo-cubos de dimensión n). Con las notaciones precedentes, sea $n \in \mathbb{N}$, $n \geq 1$, $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita, con $\#(H) \geq 2$, y $G(H) = (H, E(H))$ el hiper-grafo de una inclusión asociado a H . Los siguientes ítems son equivalentes:

- El hiper-grafo de una inclusión $G(H)$ es un pseudo-cubo de dimensión n .
- Para cada arista $e \in E(H)$, $\#(e) \geq 2$.

- iii) El grado promedio verifica $\text{avd}(H) = n$.
- iv) Para todo $h \in H$, $\text{deg}(h) = n$.
- v) El grado promedio, el grado de salida promedio y la densidad del hiper-grafo cumplen la siguiente igualdad:

$$\text{avd}(H) = E_H[\text{outdeg}] + \text{gd}(H).$$

En particular, si H es un pseudo-cubo de dimensión n y $n \geq 2$, tenemos que:

$$(0.3.6) \quad \frac{n}{2} \leq E_H[\text{outdeg}] < n.$$

Posteriormente, introducimos en la Definición 43 la noción de dimensión DS (por Daniely-Shalev-Shwartz en [DS, 14]), que denotamos por \dim_{DS} , empleando términos similares a los usados en [BCDMY, 22], esto es, basándonos en los pseudo-cubos.

Una pregunta natural es si la primera desigualdad en la cadena de Desigualdades (0.3.6) (con $n = \dim_{DS}(H)$), sigue siendo cierta si H no es un pseudo-cubo. Esto no es cierto en general y el Lema 12 de [BCDMY, 22] no puede extenderse para el grado de salida promedio. En el Ejemplo 6.3.11 mostramos que, para cada $p \in \mathbb{N}$, dado $\mathcal{Y}_p = \{0, \dots, p-1\}$, existe una clase de conceptos asociada $H_p \subseteq \mathcal{Y}_p^{[2]}$ de modo que se verifica lo siguiente:

$$\lim_{p \rightarrow \infty} E_{H_p}[\text{outdeg}] = \lim_{p \rightarrow \infty} \text{avd}(H_p) = 0 < 2 = \dim_{DS}(H_p).$$

La segunda noción de dimensión, llamada *dimensión exponencial* (cf. Definición 44), se trata siguiendo la definición introducida en [BCDMY, 22]. Denotamos por $\dim_E(H)$ a la dimensión exponencial de H . Como ejemplo, mostramos los valores de la dimensión DS y de la dimensión exponencial en el caso de escalones en la Proposición 6.3.15.

PROPOSICIÓN (DESTACABLE) 0.3.14. *Sea $n \in \mathbb{N}$ un entero positivo, $n \geq 2$, $h \in \mathbb{N}^n$ un exponente monomial y \mathcal{E}_h la clase de conceptos definida como el escalón determinado por h . Entonces, se verifican las siguientes propiedades:*

- i) El grado promedio de \mathcal{E}_h , el grado de h en $G(\mathcal{E}_h)$ y la dimensión DS de \mathcal{E}_h coinciden. Es decir, tenemos que:

$$\text{deg}(h) = \sharp(\text{DS}(h)) = \text{avd}(\mathcal{E}_h) = \dim_{DS}(\mathcal{E}_h).$$

- ii) La dimensión exponencial de \mathcal{E}_h verifica:

$$\dim_E(\mathcal{E}_h) := \min\left\{n, \left\lfloor \left(\sum_{i \in \text{DS}(h)} \log_2(1 + h_i) \right) \right\rfloor\right\}.$$

En particular, se cumplen las siguientes desigualdades:

$$E_{\mathcal{E}_h}[\text{outdeg}] < \text{avd}(\mathcal{E}_h) = \dim_{DS}(\mathcal{E}_h) \leq \dim_E(\mathcal{E}_h).$$

La última desigualdad se convierte en una igualdad si y solo si \mathcal{E}_h es un pseudo-cubo de dimensión n (o, equivalentemente, si $h \in (\mathbb{Z}^{\times})^n$, donde $\mathbb{Z}^{\times} := \mathbb{Z} \setminus \{0\}$, o, también, si $\text{DS}(h) = [n]$).

Por otro lado, también estudiamos las caracterizaciones de los pseudo-cubos cerrados hacia abajo en el Corolario 6.3.17.

COROLARIO (DESTACABLE) 0.3.15. *Si $H \subseteq \mathcal{Y}^{[n]}$, $n \geq 1$, es un conjunto cerrado hacia abajo, las siguientes propiedades son equivalentes:*

- i) H es un pseudo-cubo de dimensión n .
- ii) Se verifica lo siguiente:

$$E_H[\text{outdeg}] < \text{avd}(H) = \dim_{DS}(H) = \dim_E(H) = n \leq 2E_H[\text{outdeg}].$$

- iii) Dada una descomposición H en sus componentes (escalones) maximales:

$$H =: \bigcup_{i=1}^s \mathcal{E}_{h_i},$$

entonces, se verifica lo siguiente:

$$\text{avd}(H) = \frac{1}{s} \sum_{i=1}^s \text{deg}(h_i) = \max\{\text{deg}(h_1), \dots, \text{deg}(h_s)\} = \dim_{DS}(H) = \dim_E(H) = n.$$

iv) Dada una descomposición de H en sus componentes (escalones) maximales:

$$H =: \bigcup_{i=1}^s \mathcal{E}_{h_i},$$

entonces, para cada i , $1 \leq i \leq s$, se cumple que:

$$\text{avd}(H) = \text{deg}(h_i) = \dim_{\text{DS}}(\mathcal{E}_{h_i}) = \dim_{\text{E}}(\mathcal{E}_{h_i}) = \dim_{\text{DS}}(H) = \dim_{\text{E}}(H) = n.$$

Estas desigualdades no siguen siendo ciertas si eliminamos la hipótesis de pseudo-cubo. En el Ejemplo 6.3.18, mostramos una sucesión de clases $\{F_p \subseteq \mathcal{Y}^{[2]} : p \in \mathbb{N}\}$, donde $\mathcal{Y} := \{0, 1, \dots, p-1\}$, tal que, para cada $p \in \mathbb{N}$, $p \geq 2$, F_p es cerrado hacia abajo pero no es un pseudo-cubo. Estas clases F_p verifican:

$$1 = E_{F_p}[\text{outdeg}] = \lim_{p \rightarrow \infty} \text{avd}(F_p) < \dim_{\text{DS}}(F_p) = \dim_{\text{E}}(F_p) \leq 2E_{F_p}[\text{outdeg}] < 2\text{avd}(F_p).$$

Para subconjuntos cerrados hacia abajo demostramos el siguiente resultado (ver Proposición 6.3.21):

PROPOSICIÓN (DESTACABLE) 0.3.16. *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita cerrada hacia abajo. Entonces, tenemos que:*

i) Si $\dim_{\text{DS}}(H) \geq 1$, entonces:

$$E_H[\text{outdeg}] < \text{avd}(H) \leq \dim_{\text{DS}}(H).$$

ii) Si $\dim_{\text{E}}(H) \geq 1$, entonces:

$$E_H[\text{outdeg}] < \text{avd}(H) \leq \dim_{\text{E}}(H).$$

Nótese que esta última proposición mejora la cota superior de la Proposición 27 de [BCDMY, 22] en el caso de subconjuntos cerrados hacia abajo.

Posteriormente, en la Sección 6.4, realizamos un estudio exhaustivo de la técnica del shifting en el caso multiclase (que definimos en detalle en la Sección 6.4.1). En [Ha, 95], D. Haussler utilizó esta técnica para producir una demostración del Lema de Sauer-Shelah-Perles y otros resultados en el contexto del Aprendizaje binario. Esta técnica apareció por primera vez en [Fr, 83] e, independientemente, en [Al, 83] (véase también el papel del shifting en relación con la traza y la dualidad en [Pa, 24]). En [RBR, 06], los autores extendieron esta técnica al caso del Aprendizaje multiclase. Posteriormente, en [BCDMY, 22] vuelven a analizar esta técnica. Para nuestra presentación, seguiremos esta última referencia. Definimos el shift $\mathbb{S}_i : 2^{\mathcal{Y}^n} \rightarrow 2^{\mathcal{Y}^n}$ de una arista $e_{f,i}$ en la dirección de π_i como el siguiente operador:

$$\mathbb{S}_i(e_{f,i}) := \{g \in \mathcal{Y}^n : \pi_i(g) = f, g(i) \in \{0, \dots, \#(e_{f,i}) - 1\}\} \subseteq \mathcal{Y}^n.$$

Por otro lado, el shift $\mathbb{S}_i : 2^{\mathcal{Y}^n} \rightarrow 2^{\mathcal{Y}^n}$ de una clase de conceptos H en la dirección de π_i viene dado por:

$$\mathbb{S}_i(H) := \bigcup_{f \in \pi_i(H)} \mathbb{S}_i(e_{f,i}).$$

La primera observación, ya realizada en [BCDMY, 22], es que la dimensión exponencial se preserva tras un shift. Esto es, dada una clase de conceptos finita $H \subseteq \mathcal{Y}^{[n]}$ y dado $i \in [n]$, el shift \mathbb{S}_i en la dirección de π_i satisface:

$$\dim_{\text{E}}(\mathbb{S}_i(H)) \leq \dim_{\text{E}}(H).$$

Sin embargo, la dimensión DS puede incrementarse después de un shift (como ya se observó en el Ejemplo 19 de [BCDMY, 22]). En el Ejemplo 6.4.24 mostramos una clase de conceptos $H \subseteq \{0, 1, 2, 3\}^{[2]}$ de modo que si consideramos $H' := \mathbb{S}_2(H)$, entonces $\text{avd}(H') < \text{avd}(H)$. Sin embargo, si consideramos $H'' := \mathbb{S}_1(H)$, que es cerrada hacia abajo, se cumple $\text{avd}(H'') > \text{avd}(H) > \text{avd}(H')$. En el Lema 6.4.25 y la Proposición 6.4.26 explicamos, parcialmente, el comportamiento del grado promedio después de un shift.

En [Ha, 95], se muestra que, dada una clase de conceptos $H \subseteq \mathcal{Y}^{[n]}$, existe una palabra $\omega \in [n]^*$ que define una composición de una secuencia de shifts:

$$\mathbb{S}_\omega := \mathbb{S}_{\omega_1} \circ \dots \circ \mathbb{S}_{\omega_r} : \mathcal{Y}^{[n]} \rightarrow \mathcal{Y}^{[n]},$$

tal que la clase:

$$H_* := \mathbb{S}_\omega(H),$$

es cerrada hacia abajo con respecto al pre-orden de Dickson. El Ejemplo 6.4.24 también muestra que H_* no es único y algunos invariantes pueden cambiar de forma caótica.

La técnica del shifting nos permite obtener una caracterización de los conjuntos cerrados hacia abajo (o escaleras) en términos de ideales “especiales”: ideales de polinomios que se anulan en los puntos de la escalera. Esta caracterización es consecuencia de las ideas de [Me, 19] y su predecesor [FRR, 06]. Antes de exponer estos

resultados, es necesario fijar algunas notaciones. Sea $\leq_{1\text{ex}}$ el orden monomial lexicográfico canónico sobre el monoide $(\mathbb{N}^n, +)$ (cf. Definición 36) y sea $\mathfrak{a} \subseteq K[X_1, \dots, X_n]$ un ideal de polinomios. Sea $\text{exp}_{1\text{ex}}(\mathfrak{a}) \subseteq \mathbb{N}^n$ el conjunto de los exponentes de los términos directores de los polinomios del ideal \mathfrak{a} con respecto al orden $\leq_{1\text{ex}}$ y $\text{Nexp}_{1\text{ex}}(\mathfrak{a}) = \mathbb{N}^n \setminus \text{exp}_{1\text{ex}}(\mathfrak{a})$. El resultado principal de [Me, 19] se apoya fuertemente en un resultado de [FRR, 06], quien, a su vez, está altamente inspirado en elementos de la Teoría de Juegos (en particular, del *lex game*). En este capítulo damos una demostración original del resultado de [Me, 19] usando únicamente elementos propios de la División de Weierstarss-Hironaka y las Bases de Gröbner (cf. Teorema 6.4.11):

TEOREMA (DESTACABLE) 0.3.17 (**Teorema 2 de [Me, 19]**). *Con las notaciones precedentes, tenemos que:*

$$\text{Nexp}_{1\text{ex}}(I(H)) = \mathbb{S}_{n \dots 1}(H).$$

Usando el resultado anterior y nuestras discusiones de la Subsección 6.1.2, obtenemos la siguiente caracterización de los conjuntos cerrados hacia abajo (cf. Corolario 6.4.12):

COROLARIO (DESTACABLE) 0.3.18 (**H es un conjunto cerrado hacia abajo si y solo si $H = \text{Nexp}_{1\text{ex}}(I(H))$**). *Con las notaciones precedentes, un subconjunto $H \subseteq \mathcal{Y}^n$ es cerrado hacia abajo si y solo si:*

$$H = \text{Nexp}_{1\text{ex}}(I(H)),$$

donde $I(H)$ es el ideal de todos los polinomios $p \in K[X_1, \dots, X_n]$ que se anulan en H .

Por otra parte, en la Subsubsección 6.4.2 estudiamos el comportamiento del grado de salida promedio tras aplicar un shift. La Proposición 6.4.18 prueba que el grado de salida promedio no disminuye después de un shift:

PROPOSICIÓN (DESTACABLE) 0.3.19. *Con las notaciones precedentes, dados $H \subseteq \mathcal{Y}^{[n]}$ e $i \in [n]$, el número de aristas y la densidad del hiper-grafo de una inclusión $G(H)$ no crecen tras realizar un shift, i.e.*

$$\begin{aligned} \sharp(E(\mathbb{S}_i(H))) &\leq \sharp(E(H)), \\ \text{gd}(\mathbb{S}_i(H)) &\leq \text{gd}(H). \end{aligned}$$

En particular, el grado de salida promedio no decrece después de un shift, i.e.

$$E_{\mathbb{S}_i(H)}[\text{outdeg}] \geq E_H[\text{outdeg}].$$

A partir de nuestras discusiones, y utilizando una secuencia de shifts (como en [Ha, 95]), podemos concluir el Teorema 6.4.20, que reproducimos a continuación:

TEOREMA (DESTACABLE) 0.3.20 (**El grado de salida promedio es estrictamente menor que la dimensión exponencial**). *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Entonces, tenemos que:*

$$E_H[\text{outdeg}] < \dim_E(H).$$

El resultado anterior implica el siguiente corolario:

COROLARIO (DESTACABLE) 0.3.21 (**Mejora de la Proposición 27 de [BCDMY, 22]**). *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Entonces, tenemos que:*

i) *Las dimensiones satisfacen la siguiente desigualdad:*

$$\dim_{\text{DS}}(H) < 2 \dim_E(H).$$

ii) *El grado promedio y la dimensión exponencial verifican:*

$$\text{avd}(H) < 2 \dim_E(H).$$

El Ítem ii) del resultado anterior mejora la Proposición 27 de [BCDMY, 22] por un factor estrictamente mayor que 2. El corolario anterior nos permite demostrar una mejora estricta del Corolario 28 de [BCDMY, 22] sobre la existencia de orientaciones con grado de salida máximo controlado. Antes de enunciar el resultado con esta mejora, introducimos una serie de nociones necesarias para su formulación. En primer lugar, definimos el grado mínimo de una clase de conceptos $H \subseteq \mathcal{Y}^{[n]}$ como el mínimo de los grados de sus elementos en el hiper-grafo de una inclusión $G(H)$, i.e.

$$\text{deg}_{\min}(H) := \min\{\text{deg}(h) : h \in H\}.$$

En [DS, 14], los autores consideran también el grado promedio “maximal”:

$$\text{avd}_{\max}(H) := \max\{\text{avd}(F) : F \subseteq H\},$$

que es obviamente una función monótona para la inclusión de conjuntos. Análogamente, consideramos el grado de salida promedio “maximal” definido mediante la siguiente expresión:

$$E_H^{(\max)}[\text{outdeg}] := \max\{E_F[\text{outdeg}] : F \subseteq H\},$$

que es, de nuevo, una función monótona con respecto a la inclusión de conjuntos. Obviamente, como $\text{avd}(F) \leq 2E_F[\text{outdeg}]$ para todo $F \subseteq H$, también concluimos que:

$$(0.3.7) \quad \text{avd}_{\max}(H) \leq 2E_H^{(\max)}[\text{outdeg}].$$

Finalmente, introducimos la siguiente cantidad:

$$\min - \text{outdeg}_{\max}(H) = \min\{\text{outdeg}_{\max}(\sigma) : \sigma : E(H) \longrightarrow H \text{ es una orientación}\}.$$

TEOREMA (DESTACABLE) 0.3.22 (Mejora estricta del Corolario 28 de [BCDMY, 22]). *Con las mismas notaciones que antes, sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Sea $\varphi : 2^H \longrightarrow \mathbb{R}$ una función monótona con respecto a la inclusión en 2^H . Supongamos que la función φ también verifica la siguiente hipótesis:*

$$(0.3.8) \quad \forall F \in 2^H \quad \text{deg}_{\min}(F) \leq \varphi(F).$$

Entonces, existe una orientación $\sigma : E(H) \longrightarrow H$ que cumple la siguiente desigualdad:

$$(0.3.9) \quad \text{outdeg}_{\max}(\sigma) \leq \varphi(H).$$

Además, tenemos que:

i) La Hipótesis (0.3.8) se verifica siempre que la siguiente condición sea cierta:

$$(0.3.10) \quad \text{avd}(F) \leq \varphi(F), \quad \forall F \subseteq H.$$

En particular, la función monótona $\varphi(F) := 2 \dim_E(F)$ satisface la Desigualdad (0.3.10) y, por lo tanto, existe una orientación $\sigma : E(H) \longrightarrow H$ tal que:

$$(0.3.11) \quad \text{outdeg}_{\max}(\sigma) \leq 2 \dim_E(H).$$

ii) El grado promedio maximal también verifica la Hipótesis (0.3.10) y, por consiguiente, existe una orientación $\sigma : E(H) \longrightarrow H$ de modo que:

$$\text{outdeg}_{\max}(\sigma) \leq \text{avd}_{\max}(H) \leq 2E_H^{(\max)}[\text{outdeg}].$$

iii) El $E_H[\text{outdeg}]$ es la cantidad que controla la complejidad del algoritmo OIG (cf. Algoritmo 0.2.2) ya que:

$$E_H[\text{outdeg}] \leq \min - \text{outdeg}_{\max}(H) \leq 2E_H^{(\max)}[\text{outdeg}].$$

El Ítem *i)* de este último teorema mejora el Corolario 28 de [BCDMY, 22] por un factor estrictamente mayor que 2. Por otra parte, el Ítem *ii)* extiende el Lema 3 de [DS, 14] incluyendo dos veces el grado de salida promedio maximal. Finalmente, el Ítem *iii)* relaciona nuestros resultados con la complejidad del Algoritmo 0.2.2.

El Ejemplo 6.4.28 muestra que la Hipótesis (0.3.10) no es satisfecha por la dimensión DS . Este ejemplo prueba también que $\text{avd}_{\max}(H)$ puede ser mayor que $\dim_{DS}(H)$. Por lo tanto, el Lema 3 de [DS, 14] no puede utilizarse para producir orientaciones cuyo grado de salida máximo esté acotado por $\dim_{DS}(H)$. En consecuencia, parece que:

$$2E_H^{(\max)}[\text{outdeg}]$$

es la cantidad “más fina” que podemos encontrar de modo que sea estable por shifting y que permita controlar el grado de salida máximo de alguna orientación en el hiper-grafo de una inclusión $G(H)$ asociado a H .

Nociones y notaciones básicas. Resultados preliminares

Índice

1.1.	Conjuntos localmente cerrados y constructibles en la topología de Zariski	1
1.2.	Grado de conjuntos localmente cerrados de acuerdo a [He, 83] y su Desigualdad de Bézout	5
1.3.	Teoría de Vapnik-Chervonenkis	7

El objetivo de este primer capítulo es exponer las notaciones, las nociones y los resultados básicos de la Geometría Algebraica, la Teoría de la Intersección y la Teoría del Aprendizaje Computacional (en particular, de la Teoría de Vapnik-Chervonenkis) que utilizaremos en esta memoria

En la Sección 1.1 introducimos los conjuntos localmente cerrados y constructibles en la topología de Zariski, así como sus propiedades fundamentales. En particular, en el Lema 1.1.2 mostramos las descomposiciones de conjuntos constructibles en subconjuntos localmente cerrados irreducibles que vamos a considerar repetidamente a lo largo de esta memoria. Además, presentamos el conjunto constructible *La Croix de Berny* (ver Ejemplo 1.1.3), que jugará un papel fundamental en las discusiones del Capítulo 2.

Por otra parte, en la Sección 1.2 discutimos la teoría del grado para conjuntos localmente cerrados desarrollada por J. Heintz (cf. [He, 83]). Los resultados de esta teoría se presentan de forma gradual, culminando con la famosa Desigualdad de Bézout para conjuntos localmente cerrados, un resultado clave en la Teoría de la Intersección afín. En el Capítulo 2 generalizamos esta teoría al caso de conjuntos constructibles (demostrando dos Desigualdades de Bézout para dos nociones de grado originales y distintas).

Finalmente, en la Sección 1.3, fijamos las notaciones básicas de la Teoría del Aprendizaje Computacional y presentamos los resultados principales de la Teoría de Vapnik-Chervonenkis (cf. [VC, 71]). Utilizaremos esta terminología y estos resultados en los Capítulos 4 y 5 de la memoria.

1.1. Conjuntos localmente cerrados y constructibles en la topología de Zariski

En esta sección recopilamos las notaciones, las nociones y los resultados básicos de la Geometría Algebraica que utilizaremos a lo largo de la memoria. En particular, nos centraremos en las propiedades de los conjuntos localmente cerrados y constructibles.

Sea K un cuerpo algebraicamente cerrado y $K[X_1, \dots, X_n]$ el anillo de polinomios en el conjunto de variables $\{X_1, \dots, X_n\}$ con coeficientes en K . Denotamos por $\mathbb{A}^n(K)$ (o simplemente \mathbb{A}^n cuando no haya riesgo de confusión) al espacio afín de dimensión n sobre K . Dado un conjunto finito de polinomios $\{f_1, \dots, f_s\} \subseteq K[X_1, \dots, X_n]$, denotamos por $V_{\mathbb{A}}(f_1, \dots, f_s) \subseteq \mathbb{A}^n(K)$ a la variedad algebraica de los ceros comunes de estos polinomios en $\mathbb{A}^n(K)$. Es decir,

$$V_{\mathbb{A}}(f_1, \dots, f_s) := \{x \in \mathbb{A}^n(K) : f_1(x) = f_2(x) = \dots = f_s(x) = 0\}.$$

Si $\mathfrak{a} \subseteq K[X_1, \dots, X_n]$ es un ideal, también denotamos por $V_{\mathbb{A}}(\mathfrak{a}) \subseteq \mathbb{A}^n(K)$ al conjunto de los ceros comunes de todos los polinomios en \mathfrak{a} . Obviamente, si \mathfrak{a} es el ideal generado por el conjunto finito $\{f_1, \dots, f_s\}$ (i.e. $\mathfrak{a} = (f_1, \dots, f_s)$), entonces tenemos que $V_{\mathbb{A}}(\mathfrak{a}) = V_{\mathbb{A}}(f_1, \dots, f_s)$.

Existe una única topología en $\mathbb{A}^n(K)$ cuyos conjuntos cerrados son las variedades algebraicas afines, la cual es usualmente conocida como *la topología de Zariski* en $\mathbb{A}^n(K)$. Para cada subconjunto $S \subseteq \mathbb{A}^n(K)$, denotamos por \overline{S}^z a la clausura de S respecto a la topología de Zariski en $\mathbb{A}^n(K)$. Para todo subconjunto $X \subseteq \mathbb{A}^n(K)$, la topología inducida en X por la topología de Zariski de $\mathbb{A}^n(K)$ se denominará topología de Zariski de X . Usaremos los términos *abierto Zariski en X* o *cerrado Zariski en X* para describir los conjuntos abiertos y cerrados en X con respecto a su topología de Zariski. Nótese que todos los resultados de esta memoria también son válidos para la topología de Zariski de los conjuntos κ -definibles en $\mathbb{A}^n(K)$, donde κ es un cuerpo y K su clausura algebraica.

Sea $f \in K[X_1, \dots, X_n]$ un polinomio, definimos el *abierto distinguido* definido por f como el subconjunto abierto Zariski dado por:

$$(1.1.1) \quad D(f) := \{x \in \mathbb{A}^n(K) : f(x) \neq 0\} = \mathbb{A}^n(K) \setminus V_{\mathbb{A}}(f).$$

La familia de los abiertos distinguidos forma una base de abiertos para la topología de Zariski y, como consecuencia del Teorema de la Base de Hilbert, todo abierto $U \subseteq \mathbb{A}^n(K)$ con respecto a la topología de Zariski puede escribirse como la unión finita de abiertos distinguidos, i.e.

$$U := D(f_1) \cup \dots \cup D(f_s),$$

donde $f_1, \dots, f_s \in K[X_1, \dots, X_n]$. A continuación, introducimos los conjuntos localmente cerrados y constructibles, que desempeñarán un papel central en esta memoria:

DEFINICIÓN 8 (Conjuntos localmente cerrados y constructibles en la topología de Zariski de $\mathbb{A}^n(K)$). *La intersección de un subconjunto abierto y uno cerrado en la topología de Zariski de $\mathbb{A}^n(K)$ se denominará subconjunto localmente cerrado. Se dice que un conjunto localmente cerrado $V \subseteq \mathbb{A}^n(K)$ es irreducible si V es un subconjunto abierto en una variedad algebraica cerrada irreducible de $\mathbb{A}^n(K)$. Las uniones finitas de conjuntos localmente cerrados reciben el nombre de conjuntos constructibles (en la tradición más clásica de la Teoría de la Eliminación, ver [Ch, 64-65]).*

Algunos autores prefieren usar el término *variedad cuasi-proyectiva* en lugar de conjunto localmente cerrado (ver, por ejemplo, [Sha, 77]). Dado que el contexto de esta memoria es esencialmente afín, evitaremos discutir la terminología proyectiva, excepto cuando sea estrictamente necesario.

Como se muestra en el Ejemplo 1.1.3, existen conjuntos constructibles que no son localmente cerrados. Por otro lado, la clase de los conjuntos constructibles es cerrada para uniones finitas, intersecciones finitas y complementación.

El Teorema clásico de Chevalley explica la necesidad de utilizar conjuntos constructibles en la Teoría de la Eliminación: estos son las proyecciones de conjuntos cerrados.

TEOREMA 1.1.1 ([Ch, 64-65]). *Un subconjunto $C \subseteq \mathbb{A}^n(K)$ es constructible si y sólo si existe algún $m \in \mathbb{N}$ y alguna variedad algebraica $V \subseteq \mathbb{A}^{n+m}(K)$ tal que $C := \pi(V)$, donde $\pi : \mathbb{A}^{n+m}(K) \rightarrow \mathbb{A}^n(K)$ es la proyección canónica.*

Obviamente, los grafos de las aplicaciones polinomiales definidas sobre conjuntos constructibles son constructibles. Por lo tanto, las imágenes de conjuntos constructibles mediante aplicaciones polinomiales son también constructibles. Recíprocamente, se puede demostrar de forma sencilla que un subconjunto $C \subseteq \mathbb{A}^n(K)$ es constructible si y solo si existen $m \in \mathbb{N}$, una variedad algebraica $V \subseteq \mathbb{A}^m(K)$ y una aplicación polinomial $\varphi : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K)$ tal que $C = \varphi(V)$. En otras palabras, *los conjuntos constructibles son simplemente las imágenes de variedades algebraicas afines mediante aplicaciones polinomiales.*

De nuevo, por el Teorema de la Base de Hilbert, tenemos que los conjuntos localmente cerrados admiten una descomposición minimal como unión finita de conjuntos localmente cerrados irreducibles. La unicidad, salvo permutación, de estas descomposiciones minimales de conjuntos localmente cerrados nos permite emplear el término *componentes localmente cerradas irreducibles*. Para un conjunto localmente cerrado $V \subseteq \mathbb{A}^n(K)$, sus componentes localmente cerradas irreducibles están determinadas de manera única ya que se encuentran en biyección con las componentes irreducibles de su clausura Zariski \overline{V}^z . Este hecho es fácil de demostrar, ya que todo conjunto localmente cerrado irreducible es denso en su clausura Zariski. Más adelante, veremos que esta unicidad no puede extenderse de manera inmediata a los conjuntos constructibles.

Para cada subconjunto localmente cerrado $V \subseteq \mathbb{A}^n(K)$, denotamos por $K[V]$ al anillo de funciones polinomiales sobre V . Una función polinomial entre dos conjuntos localmente cerrados $\varphi : V \rightarrow W$ se denomina *dominante* si la imagen $\varphi(V)$ es densa en W respecto a la topología de Zariski (lo cual es equivalente al hecho de que $K[W]$ está inmerso como subanillo en $K[V]$). Además, V es irreducible si y solo si $K[V]$ es un dominio de integridad. Por último, denotamos por $K(V)$ al cuerpo de funciones racionales definidas sobre V (i.e. al cuerpo de fracciones de $K[V]$).

La *dimensión* (de Krull) de un conjunto constructible $C \subseteq \mathbb{A}^n(K)$ será su dimensión de Krull como espacio topológico. Es decir, la longitud de la cadena más larga de subconjuntos localmente cerrados irreducibles de $\mathbb{A}^n(K)$ contenidos en C . Esta dimensión es sub-aditiva y no aumenta bajo aplicaciones polinomiales (es decir, dado un conjunto constructible $C \subseteq \mathbb{A}^n(K)$ y una aplicación polinomial $\varphi : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$, tenemos que $\dim(\varphi(C)) \leq \dim(C)$).

En [Ch, 64-65], C. Chevalley definió las componentes irreducibles de un conjunto constructible como las componentes irreducibles de su clausura Zariski. Como veremos en el Capítulo 2, este enfoque no es el adecuado para

nuestros propósitos. Por ello, en esta memoria, consideraremos las siguientes descomposiciones de conjuntos constructibles en “componentes” localmente cerradas irreducibles:

LEMA 1.1.2. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible. Entonces, existe un conjunto finito*

$$\mathcal{C} := \{U_1 \cap V_1, \dots, U_s \cap V_s\}$$

de conjuntos localmente cerrados irreducibles que verifica las siguientes propiedades:

$$(1.1.2) \quad C := (U_1 \cap V_1) \cup \dots \cup (U_s \cap V_s),$$

y

- i) V_i es una variedad algebraica irreducible en $\mathbb{A}^n(K)$,
- ii) U_i es el máximo de los conjuntos abiertos Zariski $O_i \subseteq \mathbb{A}^n(K)$ tales que $O_i \cap V_i \subseteq C$,
- iii) $V_i \neq V_j$.

DEMOSTRACIÓN. Como C es una unión finita de conjuntos localmente cerrados:

$$C = W_1 \cup \dots \cup W_s,$$

al tomar las componentes irreducibles de cada W_i , concluimos que C es una unión finita de conjuntos localmente cerrados irreducibles. Dado que los subconjuntos abiertos no vacíos de variedades algebraicas irreducibles son densos en la topología de Zariski, simplemente reordenando estos conjuntos localmente cerrados irreducibles, podemos asumir una descomposición como la siguiente:

$$(1.1.3) \quad C := (O_1 \cap V_1) \cup \dots \cup (O_r \cap V_r),$$

donde para cada i , $1 \leq i \leq r$, $O_i \subseteq \mathbb{A}^n(K)$ es un subconjunto abierto Zariski y $V_i \subseteq \mathbb{A}^n(K)$ es una variedad algebraica irreducible tal que $V_i \neq V_j$ para cada $i \neq j$. Si existieran i y j tales que $V_i = V_j$, simplemente reemplazaríamos $(O_i \cap V_i) \cup (O_j \cap V_j)$ por $(O_i \cup O_j) \cap V_i$. Por lo tanto, podemos asumir que la Propiedad iii) se verifica. Finalmente, basta con tomar el máximo de los conjuntos abiertos Zariski $U_i \subseteq \mathbb{A}^n(K)$ tales que $U_i \cap V_i \subseteq C$ para finalizar la demostración. La existencia de este máximo está garantizada por el hecho de que la unión de conjuntos abiertos es abierta en cualquier topología. \square

EJEMPLO 1.1.3 (**La Croix de Berny**). En primer lugar, observamos que la descomposición de conjuntos constructibles en subconjuntos localmente cerrados irreducibles no es única. Además, veremos que, aunque las imágenes de variedades algebraicas por aplicaciones polinomiales son siempre conjuntos constructibles, no son siempre localmente cerradas. Consideremos la siguiente hipersuperficie cúbica de $\mathbb{A}^3(\mathbb{C})$:

$$(1.1.4) \quad W := \{(x, y, z) \in \mathbb{A}^3(\mathbb{C}) : zxy + (x^2 + y^2 - 1) = 0\}.$$

Nótese que la anterior hipersuperficie es irreducible ya que el polinomio:

$$h(X, Y, Z) = ZXY + (X^2 + Y^2 - 1) \in \mathbb{C}[X, Y][Z]$$

es un polinomio primitivo (basta observar que $\text{mcd}(XY, X^2 + Y^2 - 1) = 1$ en $\mathbb{C}[X, Y]$) de grado 1. Por otro lado, el grado de esta hipersuperficie es 3 de acuerdo a la noción de grado de [He, 83] (ver Sección 1.2 para más detalles).

Sea $\pi : \mathbb{A}^3(\mathbb{C}) \rightarrow \mathbb{A}^2(\mathbb{C})$ la proyección que “olvida” la coordenada z (i.e. $\pi(x, y, z) := (x, y)$, para todo $(x, y, z) \in \mathbb{A}^3(\mathbb{C})$). Entonces, tenemos que $\pi(W)$ es el siguiente conjunto constructible, al que denominaremos *La Croix de Berny*:

$$C := \pi(W) = \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : xy \neq 0\} \cup \{(0, 1), (0, -1), (1, 0), (-1, 0)\}.$$

Esta es una descomposición de C en cinco conjuntos localmente cerrados irreducibles que satisface las propiedades del Lema 1.1.2 anterior:

$$C := C_1 \cup \{(1, 0)\} \cup \{(-1, 0)\} \cup \{(0, 1)\} \cup \{(0, -1)\},$$

donde $C_1 \subseteq \mathbb{A}^2(\mathbb{C})$ es el siguiente subconjunto abierto Zariski:

$$C_1 := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : xy \neq 0\}.$$

También tenemos otra descomposición de $C = \pi(W)$ como la unión de dos subconjuntos localmente cerrados irreducibles:

$$C := C_1 \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x^2 + y^2 - 1 = 0\}.$$

Esto sucede porque C no es un conjunto localmente cerrado. Además, las componentes irreducibles de la clausura Zariski \overline{C}^z de C , que, en este caso, es únicamente la variedad irreducible $\mathbb{A}^2(\mathbb{C})$, no están en biyección con las clausuras Zariski de los conjuntos irreducibles que aparecen en las descomposiciones anteriores. En particular, las componentes irreducibles de C como espacio topológico Noetheriano (en el sentido de la Definición 9, la Proposición 9 y las páginas 22 – 23 de [Ch, 64-65]) no son siempre conjuntos localmente cerrados.

Veamos a continuación un problema básico relacionado con la dimensión de los conjuntos constructibles:

EJEMPLO 1.1.4 (Dimensión local y dimensión global en conjuntos constructibles). Reconsideremos el ejemplo de La Croix de Berny introducido en el Ejemplo 1.1.3 anterior. Allí, consideramos el siguiente conjunto constructible:

$$C := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : xy \neq 0\} \cup \{(0, 1), (0, -1), (1, 0), (-1, 0)\}.$$

Sea C_a el germen local de C en el punto $a = (1, 0)$ con respecto a la topología de Zariski (los argumentos también son válidos para la topología Euclídea en $\mathbb{A}^2(\mathbb{C})$). Es obvio que la dimensión de C en a es 2: cualquier vecindad de a contiene una pieza de dimensión 2 de C . En particular, C tiene dimensión *local* y *global* (en cualquier punto $b \in C$) igual a 2, sin embargo, no puede representarse como unión finita de subconjuntos localmente cerrados irreducibles de dimensión 2. Si C fuese unión de varios subconjuntos localmente cerrados irreducibles de dimensión 2, entonces C sería un subconjunto abierto Zariski de $\mathbb{A}^2(\mathbb{C})$, que no es el caso como ya hemos visto en el Ejemplo 1.1.3.

La unicidad de las “componentes” localmente cerradas irreducibles solo está garantizada para las componentes de dimensión más alta.

PROPOSICIÓN 1.1.5 (Componentes de dimensión más alta de un conjunto constructible). *Con las notaciones y las hipótesis precedentes, si $C \subseteq \mathbb{A}^n(K)$ es un conjunto constructible, entonces C admite una descomposición que satisface las Propiedades i), ii) y iii) del Lema 1.1.2 anterior:*

$$C := (U_1 \cap V_1) \cup \dots \cup (U_s \cap V_s),$$

tal que existe $r \leq s$, verificando:

$$\dim(C) = \dim(V_i) \iff 1 \leq i \leq r.$$

Las variedades V_1, \dots, V_r están determinadas de manera única y son las componentes irreducibles de dimensión más alta de la clausura Zariski de C . Las variedades V_{r+1}, \dots, V_s reciben el nombre de componentes inmersas de C .

DEFINICIÓN 9 (Conjuntos constructibles presentables equidimensionalmente). *Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Decimos que C es presentable equidimensionalmente si existe una descomposición de C como unión finita de subconjuntos localmente cerrados irreducibles:*

$$C := (U_1 \cap V_1) \cup \dots \cup (U_s \cap V_s),$$

que satisface las propiedades del Lema 1.1.2 y tal que su dimensión también verifica:

$$\dim(C) = \dim(U_i \cap V_i) = \dim(V_i), \quad 1 \leq i \leq s.$$

Como ya hemos visto, el conjunto constructible descrito en el Ejemplo 1.1.3 (i.e. La Croix de Berny) no es presentable equidimensionalmente, aunque es localmente equidimensional y es la proyección de una hipersuperficie. En particular, los subconjuntos localmente cerrados cuya clausura Zariski es equidimensional son presentables equidimensionalmente.

OBSERVACIÓN 1.1.6 (Descomposición en piezas presentables equidimensionalmente). Dada una descomposición de un conjunto constructible $C = W_1 \cup \dots \cup W_s \subseteq \mathbb{A}^n(K)$, con $W_i = U_i \cap V_i$, como en el Lema 1.1.2, también podemos descomponer C como unión finita de conjuntos constructibles presentables equidimensionalmente:

$$C := C_r \cup \dots \cup C_0,$$

donde $r = \dim(C)$ y

$$C_k := \bigcup_{\dim(W_i)=k} W_i.$$

Diremos que la descomposición anterior es una *descomposición en piezas presentables equidimensionalmente de C* . Como ya hemos visto en el Ejemplo 1.1.3, las descomposiciones en piezas presentables equidimensionalmente no son únicas. En concreto, La Croix de Berny admite varias descomposiciones en piezas presentables equidimensionalmente:

$$C := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : xy \neq 0\} \cup \{(\pm 1, 0), (0, \pm 1)\},$$

y

$$C := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : xy \neq 0\} \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x^2 + y^2 - 1 = 0\}.$$

1.2. Grado de conjuntos localmente cerrados de acuerdo a [He, 83] y su Desigualdad de Bézout

El objetivo de esta sección es exponer la teoría del grado para conjuntos localmente cerrados de J. Heintz y su Desigualdad de Bézout (cf. [He, 83]). En el Capítulo 2 generalizamos esta teoría al caso de conjuntos constructibles.

Sea $V \subseteq \mathbb{A}^n(K)$ un subconjunto localmente cerrado irreducible de dimensión (de Krull) r . Sea $\mathbb{A}^{rn} := \mathbb{A}^{rn}(K) = \mathcal{M}_{r \times n}(K)$ el espacio de las matrices de dimensión $r \times n$. Sea $\mathbb{A}^r := \mathbb{A}^r(K)$ el espacio afín de dimensión r sobre K . Al igual que en [He, 83], consideramos la siguiente función polinomial:

$$(1.2.1) \quad \begin{aligned} \Phi : \mathbb{A}^{rn} \times V &\longrightarrow \mathbb{A}^{rn} \times \mathbb{A}^r \\ (M, x) &\longmapsto (M, Mx) \end{aligned}$$

El siguiente resultado fue demostrado en [He, 83]:

PROPOSICIÓN 1.2.1. *Con las notaciones anteriores, tenemos que:*

i) *El morfismo Φ es un morfismo dominante y la siguiente es una extensión de cuerpos finita y separable:*

$$\Phi^* : K(\mathbb{A}^{rn} \times \mathbb{A}^r) \hookrightarrow K(\mathbb{A}^{rn} \times V).$$

ii) *Para cada punto $(M, b) \in \mathbb{A}^{rn} \times \mathbb{A}^r$, si la fibra $\Phi^{-1}(\{(M, b)\})$ es finita, entonces tenemos que:*

$$\sharp(\Phi^{-1}(\{(M, b)\})) \leq [K(\mathbb{A}^{rn} \times V) : K(\mathbb{A}^{rn} \times \mathbb{A}^r)].$$

iii) *Existe un subconjunto abierto Zariski $\mathcal{U} \subseteq \mathbb{A}^{rn} \times \mathbb{A}^r$ tal que para cada $(M, b) \in \mathcal{U}$, la fibra $\Phi^{-1}(\{(M, b)\})$ es finita y satisface:*

$$\sharp(\Phi^{-1}(\{(M, b)\})) = [K(\mathbb{A}^{rn} \times V) : K(\mathbb{A}^{rn} \times \mathbb{A}^r)].$$

Dado un par $(M, b) \in \mathbb{A}^{r(n+1)}$, consideramos la variedad afín lineal $\mathbb{G}(M, b)$ dada por la siguiente identidad:

$$\mathbb{G}(M, b) := \{x \in \mathbb{A}^n : Mx^t - b = 0\},$$

donde x^t es la traspuesta de $x = (x_1, \dots, x_n)$. Sea $G(n, r) \subseteq \mathcal{M}_{r \times n}(K) \times \mathbb{A}^r(K) = \mathbb{A}^{rn}(K) \times \mathbb{A}^r(K)$ el subconjunto constructible denso con respecto a la topología de Zariski de todos los pares (M, b) tales que la correspondiente variedad afín lineal es no vacía y $\text{codim}(\mathbb{G}(M, b)) = r$. Por otro lado, denotamos por $\mathbb{G}(n, r)$ al “Grassmaniano” abstracto de todas las subvariedades afines lineales de codimensión r . Obviamente, tenemos la siguiente aplicación suprayectiva:

$$(1.2.2) \quad \begin{aligned} \mathcal{G} : G(n, r) &\longrightarrow \mathbb{G}(n, r) \\ (M, b) &\longmapsto \mathbb{G}(M, b). \end{aligned}$$

Entonces, podemos dotar a $\mathbb{G}(n, r)$ con la topología final inducida por esta aplicación suprayectiva. Esta topología final recibirá el nombre de *topología de Zariski en $\mathbb{G}(n, r)$* . En consecuencia, llamamos subconjuntos abiertos, cerrados o localmente cerrados de $\mathbb{G}(n, r)$ a los subconjuntos abiertos, cerrados o localmente cerrados de $\mathbb{G}(n, r)$ con respecto a esta topología.

Sea $V \subseteq \mathbb{A}^n(K)$ un subconjunto localmente cerrado irreducible de dimensión r . Introducimos el siguiente subconjunto de $\mathbb{G}(n, r)$:

$$\mathbb{G}(V) := \{A \in \mathbb{G}(n, r) : \sharp(V \cap A) < \infty\}.$$

Las siguientes propiedades se deducen de [He, 83]:

PROPOSICIÓN 1.2.2. *Con las notaciones y las hipótesis anteriores, tenemos que:*

i) *La clase $\mathbb{G}(V)$ es no vacía y contiene un subconjunto abierto de $\mathbb{G}(n, r)$.*

ii) *El máximo $\max\{\sharp(A \cap V) : A \in \mathbb{G}(V)\}$ es finito.*

iii) *Existe un subconjunto abierto denso $\mathcal{U} \subseteq \mathbb{G}(n, r)$, con respecto a la topología definida previamente, tal que para todo $A \in \mathcal{U}$ se cumple:*

$$\sharp(A \cap V) = \max\{\sharp(T \cap V) : T \in \mathbb{G}(V)\}.$$

DEFINICIÓN 10 (**Grado de un subconjunto localmente cerrado**). *Sea $V \subseteq \mathbb{A}^n(K)$ un subconjunto localmente cerrado irreducible. Definimos el grado de V como la siguiente cantidad:*

$$\text{deg}(V) := \max\{\sharp(A \cap V) : A \in \mathbb{G}(n, r), \sharp(A \cap V) < \infty\}.$$

Sea $W \subseteq \mathbb{A}^n(K)$ cualquier subconjunto localmente cerrado y sean C_1, \dots, C_s sus componentes localmente cerradas irreducibles. El grado de W se define como:

$$\text{deg}(W) := \sum_{i=1}^s \text{deg}(C_i).$$

A continuación, mostramos algunas de las propiedades del grado obtenidas en [He, 83].

PROPOSICIÓN 1.2.3. *Con las mismas notaciones e hipótesis de las páginas anteriores, tenemos que:*

i) *Para cualquier subconjunto localmente cerrado $V \subseteq \mathbb{A}^n(K)$, su grado coincide con el grado de su clausura Zariski, i.e.*

$$\deg(V) = \deg(\overline{V}^z).$$

ii) *El grado de un conjunto finito $C \subseteq \mathbb{A}^n(K)$ es igual a su cardinalidad:*

$$\deg(C) = \#(C).$$

iii) *El número de componentes irreducibles de un conjunto localmente cerrado está acotado superiormente por su grado.*

iv) *El grado de cualquier variedad afín lineal es 1.*

v) *El grado de los conjuntos localmente cerrados es invariante por isomorfismos lineales o afines.*

vi) *Para cualquier polinomio no constante $f \in K[X_1, \dots, X_n]$, el grado de la hipersuperficie $V_{\mathbb{A}}(f)$ es a lo sumo el grado del polinomio $\deg(f)$ y $\deg(V_{\mathbb{A}}(f)) = \deg(f)$ si y sólo si f es libre de cuadrados.*

El grado presenta un buen comportamiento para la intersección con variedades lineales afines:

LEMA 1.2.4. *Sea $V \subseteq \mathbb{A}^n(K)$ un conjunto localmente cerrado y $A \subseteq \mathbb{A}^n(K)$ una variedad afín lineal. Entonces, se cumple:*

$$\deg(V \cap A) \leq \deg(V).$$

LEMA 1.2.5. *Sea $V \subseteq \mathbb{A}^n(K)$ un subconjunto localmente cerrado equidimensional de dimensión r . Entonces, tenemos que:*

$$\deg(V) = \max\{\#(A \cap V) : A \in \mathbb{G}(n, r), \#(A \cap V) < \infty\}.$$

Además, existe un subconjunto abierto Zariski no vacío $\mathcal{U} \subseteq \mathbb{A}^{rn+r}(K)$ tal que para todo $(M, b) \in \mathcal{U}$ se verifica:

$$\#(\mathbb{G}(M, b) \cap V) = \deg(V),$$

donde, como antes, $\mathbb{G}(M, b) := \{x \in \mathbb{A}^n : Mx^t = b\}$.

Por otro lado, se conserva el grado de la clausura Zariski al aplicar transformaciones lineales:

PROPOSICIÓN 1.2.6. *Sea $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ una aplicación lineal y $V \subseteq \mathbb{A}^n(K)$ un subconjunto localmente cerrado. Entonces, se verifica:*

$$\deg(\overline{\ell(V)}^z) \leq \deg(V).$$

Además, si $\ell(V)$ es localmente cerrado, se obtiene:

$$\deg(\ell(V)) \leq \deg(V).$$

La clave de los resultados obtenidos en [He, 83] radica en el hecho de que el grado de los conjuntos localmente cerrados se comporta de “forma adecuada” con respecto al producto cartesiano:

TEOREMA 1.2.7. *Sean $V \subseteq \mathbb{A}^n(K)$ y $W \subseteq \mathbb{A}^m(K)$ dos conjuntos localmente cerrados. Sea $V \times W \subseteq \mathbb{A}^{n+m}(K)$ su producto cartesiano, que es también localmente cerrado. Entonces, tenemos que:*

$$\deg(V \times W) = \deg(V) \deg(W).$$

Los resultados anteriores nos permiten establecer la *Desigualdad de Bézout para conjuntos localmente cerrados* de [He, 83]. El elemento principal de la demostración de J. Heintz consiste en considerar $V \cap W$ como la proyección de $(V \times W) \cap \Delta^{(2n)}$, donde $\Delta^{(2n)} \subseteq \mathbb{A}^n(K) \times \mathbb{A}^n(K)$ es la subvariedad diagonal en el espacio del producto cartesiano. Dado que la imagen $V \cap W$ es localmente cerrada, al aplicar la Proposición 1.2.6 junto con el Lema 1.2.4 y el Teorema 1.2.7, se obtiene el siguiente resultado:

TEOREMA 1.2.8 (**Desigualdad de Bézout para conjuntos localmente cerrados**, [He, 83]). *Sean $V, W \subseteq \mathbb{A}^n(K)$ dos conjuntos localmente cerrados. Entonces, tenemos que:*

$$\deg(V \cap W) \leq \deg(V) \deg(W).$$

1.3. Teoría de Vapnik-Chervonenkis

En esta última sección del capítulo, introducimos las notaciones, las nociones y los resultados básicos de la Teoría del Aprendizaje Computacional para clasificadores binarios y de la Teoría de Vapnik-Chervonenkis (cf. [VC, 71]) que utilizaremos en los Capítulos 4 y 5 de esta memoria.

Sea Y un conjunto cualquiera y sea $\{0, 1\}^Y$ la clase de todas las funciones definidas en Y con valores en $\{0, 1\}$. Diremos que $\{0, 1\}^Y$ es la clase de *funciones características* definidas sobre Y . Para cada subconjunto $F \subseteq Y$, denotamos por $\chi_F \in \{0, 1\}^Y$ a la función característica (a veces llamada función indicatriz) asociada a F como subconjunto de Y . Si $X \subseteq Y$ es un subconjunto de Y , para cada $\chi \in \{0, 1\}^Y$, denotamos por $\chi|_X \in \{0, 1\}^X$ a la restricción de χ al subconjunto X .

En el contexto de la Teoría del Aprendizaje Computacional, las funciones que asignan *datos de entrada* al rango $\{0, 1\}$ suelen recibir el nombre de *clasificadores (binarios)*. Estos clasificadores son funciones características que determinan si una entrada pertenece a una clase particular (etiquetada como 1) o no (etiquetada como 0). La *dimensión de Vapnik-Chervonenkis* (o simplemente *dimensión VC*), introducida en [VC, 71], es una “medida” de capacidad de un conjunto de clasificadores. La importancia de la dimensión VC radica en el hecho de que una clase de clasificadores es “aprendible” si y solo si su dimensión VC es finita (cf. [Va, 84], [BEHW, 89] y otras referencias clásicas en el campo).

DEFINICIÓN 11 (Desmenuzar, Dimensión de Vapnik-Chervonenkis). *Sea Y un conjunto y $\mathcal{H} \subseteq \{0, 1\}^Y$ una familia de clasificadores. Dado un subconjunto finito $X \subseteq Y$, decimos que \mathcal{H} desmenuza X si para cada $F \subseteq X$, existe un clasificador $\chi \in \mathcal{H}$ tal que la restricción a X de χ es igual a la restricción a X de χ_F . Es decir, \mathcal{H} desmenuza X si para cada $F \subseteq X$ existe un $\chi \in \mathcal{H}$ tal que:*

$$\chi|_X = \chi_F|_X.$$

Definimos la dimensión de Vapnik-Chervonenkis de una familia de clasificadores \mathcal{H} como:

$$\dim_{VC}(\mathcal{H}) := \max\{\#(X) : \mathcal{H} \text{ desmenuza } X\}.$$

A continuación, introducimos la noción de *función de crecimiento*, que es ampliamente utilizada en el contexto de la Teoría del Aprendizaje Computacional. Este concepto jugará un papel central en nuestra discusión del Capítulo 4.

DEFINICIÓN 12 (Función de crecimiento). *Sea Y un conjunto y $\mathcal{H} \subseteq \{0, 1\}^Y$ una familia de clasificadores. Para cada subconjunto $X \subseteq Y$, sea $\mathcal{H}|_X$ la clase de restricciones a X de las funciones en \mathcal{H} . Es decir,*

$$\mathcal{H}|_X := \{f|_X : f \in \mathcal{H}\}.$$

La función de crecimiento $G(\mathcal{H}, \cdot)$ de una familia de clasificadores \mathcal{H} es la función:

$$G(\mathcal{H}, \cdot) : \mathbb{N} \rightarrow \mathbb{N},$$

dada por la siguiente identidad para cada entero positivo $m \in \mathbb{N}$:

$$G(\mathcal{H}, m) := \sup\{\#(\mathcal{H}|_X) : X \subseteq Y, \#(X) = m\}.$$

Si $d := \dim_{VC}(\mathcal{H})$, entonces para todo $m \leq d$, tenemos que $G(\mathcal{H}, m) = 2^m$. El famoso Lema de Sauer-Shelah-Perles (cf. [Sau, 72], [She, 72], [Ha, 95], [Pa, 24] y sus referencias) demuestra que si \mathcal{H} es una familia de clasificadores, entonces la función de crecimiento $G(\mathcal{H}, m)$ siempre toma valores finitos y proporciona una cota superior.

LEMA 1.3.1 (Lema de Sauer-Shelah-Perles). *Sea Y un conjunto y $\mathcal{H} \subseteq \{0, 1\}^Y$ una familia de clasificadores. Sea $d := \dim_{VC}(\mathcal{H})$. Si $m > d$, tenemos la siguiente desigualdad:*

$$G(\mathcal{H}, m) \leq \sum_{i=0}^d \binom{m}{i} \leq \left(\frac{em}{d}\right)^d.$$

Nótese que el exponente que aparece en la cota superior del lema anterior es la dimensión VC.

Finalmente, enunciaremos el famoso Teorema de Vapnik-Chervonenkis, que utilizaremos en la Subsección 4.2.4 para demostrar que los conjuntos cuestores están densamente distribuidos en variedades evasivas de dimensión positiva:

TEOREMA 1.3.2 (Teorema de Vapnik-Chervonenkis, [VC, 71]). *Sea Y un conjunto y $Z = Y \times \{0, 1\}$. Sea (Z, \mathcal{B}, μ) un espacio probabilístico, donde $\mathcal{B} \subseteq 2^Z$ es una σ -álgebra y $\mu : \mathcal{B} \rightarrow [0, 1]$ una distribución de*

probabilidad definida sobre \mathcal{B} . Sea $\mathcal{H} \subseteq \{0, 1\}^Y$ una familia de clasificadores. Sea $m \in \mathbb{N}$ un entero positivo y $\varepsilon > 0$ un número real positivo. Supongamos que:

$$\frac{4 \log(2)}{m} \leq \varepsilon^2.$$

Sea $\mathcal{B}^{\otimes m}$ la σ -álgebra producto en Z^m inducida por (Z, \mathcal{B}) y sea $\mu^{\otimes m}$ la distribución de probabilidad definida sobre $\mathcal{B}^{\otimes m}$ por $\mu : \mathcal{B} \rightarrow [0, 1]$. Sea $\ell : \{0, 1\}^2 \rightarrow [0, 1]$ cualquier función no negativa (normalmente llamada "función de error"). Entonces, tenemos que:

$$(1.3.1) \quad \text{Prob}_{z \in Z^m} \left[\sup_{f \in \mathcal{H}} \left| \frac{1}{m} \sum_{i=1}^m \ell(f(x_i), y_i) - \mathbb{E}_{(x,y) \in Z} [\ell(f(x), y)] \right| > \varepsilon \right] \leq 4G(\mathcal{H}, m) e^{-\frac{m\varepsilon^2}{32}}.$$

OBSERVACIÓN 1.3.3. En la literatura actual sobre el Teorema de Vapnik-Chervonenkis, la constante que aparece en el lado derecho de la Desigualdad (1.3.1) en el teorema anterior es 8. Una revisión cuidadosa de la prueba del Teorema de Vapnik-Chervonenkis nos permite afirmar que la cota superior en la Desigualdad (1.3.1) se cumple.

Desigualdad de Bézout para conjuntos constructibles

Índice

2.1. Dos nociones de grado y dos Desigualdades de Bézout para conjuntos constructibles	9
2.2. Cotas superiores para la intersección de varios conjuntos constructibles	17
2.2.1. Una extensión directa de la Proposición 2.3 de [HS, 82] a conjuntos localmente cerrados	18
2.2.2. Demostración del Ítem <i>i</i>) del Teorema 2.2.1	22
2.2.3. Demostración del Ítem <i>ii</i>) del Teorema 2.2.1	25
2.2.3.1. Resultados preliminares	25
2.2.3.2. Demostración de la segunda cota superior del grado en el Teorema 2.2.1.	27
2.2.4. Dos aplicaciones inmediatas de las cotas superiores previas	28
2.3. Una observación sobre las cotas de las imágenes de conjuntos constructibles por aplicaciones lineales	30
2.3.1. Una primera cota para el defecto	30
2.3.2. Una cota extrínseca para el grado de una proyección basada en el Nullstellensatz Efectivo	34

A comienzos de la década de 1980, tres autores llegaron de forma independiente a tres demostraciones diferentes de ciertas Desigualdades de Bézout. Abordaron la noción de *grado de una variedad algebraica* desde diferentes enfoques y encontraron distintas técnicas para obtener sus resultados. Estos tres autores fueron W. Vogel (cf. [Vo, 84]), J. Heintz (cf. [He, 83]) y W. Fulton (cf. [Fu, 83]). Ni Fulton ni Vogel trataron con conjuntos constructibles, mientras que J. Heintz desarrolló en [He, 83] una versión afín de la noción que se llamó sistemáticamente *grado de conjuntos constructibles*. Sin embargo, como ya observó el propio J. Heintz en [He, 85], sus resultados no son válidos para conjuntos constructibles y únicamente son ciertos para conjuntos localmente cerrados de un espacio afín. El lector podrá encontrar los resultados de J. Heintz para conjuntos localmente cerrados en la Sección 1.2.

La cuestión que quedó abierta fue la de introducir una noción “correcta” de grado para conjuntos constructibles y demostrar que esta satisface una Desigualdad de Bézout. Al comienzo de la Subsección 0.3.1, presentamos un paradigma con las propiedades que debe satisfacer una noción de grado “correcta”. En este capítulo definimos dos nociones de grado “correctas” que verifican una Desigualdad de Bézout y analizamos su comportamiento con respecto al resto de propiedades establecidas en nuestro paradigma.

2.1. Dos nociones de grado y dos Desigualdades de Bézout para conjuntos constructibles

Una primera noción de grado para conjuntos constructibles fue introducida en [He, 83]. En este trabajo, J. Heintz definió el “grado” de un conjunto constructible $C \subseteq \mathbb{A}^n(K)$ como el grado de su clausura Zariski. A esta noción de grado la llamaremos *Z-grado de C* y usaremos la notación $\deg_z(C) := \deg(\overline{C}^z)$. Esta noción de grado no satisface una Desigualdad de Bézout para conjuntos constructibles. El siguiente ejemplo original demuestra este hecho.

EJEMPLO 2.1.1 (El Z-grado no satisface una Desigualdad de Bézout para conjuntos constructibles). Consideremos de nuevo el Ejemplo 1.1.3 anterior. Tenemos la hipersuperficie cúbica irreducible:

$$W := \{(x, y, z) \in \mathbb{A}(\mathbb{C})^3 : zxy + (x^2 + y^2 - 1) = 0\}$$

y su imagen $C := \pi(W)$ (i.e La Croix de Berny) bajo la proyección canónica $\pi : \mathbb{A}^3(\mathbb{C}) \rightarrow \mathbb{A}^2(\mathbb{C})$ que “olvida” la última coordenada. Tenemos que $\overline{C}^z = \mathbb{A}^2(\mathbb{C})$ y, por tanto, el Z-grado de C satisface $\deg_z(C) = 1$. Consideramos también el siguiente par de rectas:

$$V := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : xy = 0\},$$

que es una variedad algebraica de grado 2. La intersección $C \cap V$ está formada por los siguientes cuatro puntos del plano:

$$C \cap V = \{(0, 1), (0, -1), (1, 0), (-1, 0)\}.$$

Por lo tanto, $\deg_z(C \cap V) = \deg(C \cap V) = 4$. Sin embargo, tenemos que:

$$4 \not\leq \deg_z(C) \deg_z(V) = 2.$$

Este ejemplo muestra que las afirmaciones de [He, 83] referentes a conjuntos constructibles son *erróneas o incompletas al utilizar el Z -grado de un conjunto constructible como noción de grado*. Este hecho ya fue observado en [He, 85], donde el autor limitó la corrección de todas las afirmaciones de [He, 83] al caso de conjuntos localmente cerrados (como hicimos en la Sección 1.2 precedente). No hemos encontrado en la literatura matemática una noción de grado para conjuntos constructibles que generalice la noción de grado de conjuntos localmente cerrados y que, además, satisfaga una Desigualdad de Bézout. Esta situación nos llevó a buscar una noción “correcta” de grado para conjuntos constructibles que cumpla una Desigualdad de Bézout.

A continuación, definimos dos nociones originales de grado para un conjunto constructible: el grado mínimo de una presentación como unión de conjuntos localmente cerrados irreducibles (*grado LCI*) y el grado mínimo de una presentación como proyección de un conjunto localmente cerrado (*π -grado*).

DEFINICIÓN 13 (Grado LCI de un conjunto constructible). *Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Sea $\mathcal{C} := \{U_1 \cap V_1, \dots, U_r \cap V_r\}$ una descomposición de C como unión finita de conjuntos localmente cerrados irreducibles que satisface el Lema 1.1.2. Definimos el grado de C relativo a esta descomposición como:*

$$\deg(C, \mathcal{C}) := \sum_{i=1}^r \deg(V_i).$$

Finalmente, definimos el grado LCI (del inglés “locally closed irreducible”) de C como el mínimo de todos los grados de todas las descomposiciones de este tipo:

$$\deg_{\text{LCI}}(C) := \min\{\deg(C, \mathcal{C}) : \mathcal{C} \text{ es una descomposición de } C \text{ que satisface el Lema 1.1.2}\}.$$

Decimos que una descomposición $\mathcal{C} := \{U_1 \cap V_1, \dots, U_r \cap V_r\}$ de un conjunto constructible $C \subseteq \mathbb{A}^n(K)$ es una descomposición de grado LCI mínimo de C si \mathcal{C} satisface el Lema 1.1.2 y además:

$$\deg_{\text{LCI}}(C) = \sum_{i=1}^r \deg(V_i).$$

DEFINICIÓN 14 (π -grado de un conjunto constructible). *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible. Para cada $m \in \mathbb{N}$, $m \geq n$, definimos:*

$$\Pi_m(C) := \{V \subseteq \mathbb{A}^m(K) : V \text{ es localmente cerrado y } \pi(V) = C\},$$

donde $\pi : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K)$ es la proyección canónica que “olvida” las últimas $m - n$ coordenadas. Definimos entonces la clase de todos los subconjuntos localmente cerrados que se proyectan sobre C mediante:

$$\Pi(C) := \bigcup_{m \geq n} \Pi_m(C),$$

Finalmente, definimos el grado de proyección (o π -grado) de C como el siguiente mínimo:

$$\deg_{\pi}(C) := \min\{\deg_z(V) : V \in \Pi(C)\}.$$

Obviamente, tomando $m = n$, si C es localmente cerrado, al usar la “proyección” $\pi := Id_n : \mathbb{A}^n(K) \rightarrow \mathbb{A}^n(K)$, obtenemos:

$$(2.1.1) \quad \deg_{\pi}(C) \leq \deg(C) = \deg_z(C).$$

A continuación, demostramos que las tres nociones de grado son funciones sub-aditivas:

PROPOSICIÓN 2.1.2. *Las tres nociones de grado \deg_z , \deg_{π} y \deg_{LCI} son funciones sub-aditivas. Es decir, dados dos subconjuntos constructibles $C, C' \subseteq \mathbb{A}^n(K)$, tenemos que:*

$$\deg_z(C \cup C') \leq \deg_z(C) + \deg_z(C'),$$

$$\deg_{\pi}(C \cup C') \leq \deg_{\pi}(C) + \deg_{\pi}(C')$$

y,

$$\deg_{\text{LCI}}(C \cup C') \leq \deg_{\text{LCI}}(C) + \deg_{\text{LCI}}(C').$$

DEMOSTRACIÓN. Como la clausura de una unión finita es la unión de las clausuras, la primera desigualdad se verifica y \deg_z es sub-aditivo.

En cuanto a la segunda desigualdad, supongamos que $V \subseteq \mathbb{A}^{m_1}(K)$ y $V' \subseteq \mathbb{A}^{m_2}(K)$ son dos subconjuntos localmente cerrados tales que $m_1 \geq m_2 \geq n$ y $C := \pi_1(V)$, y $C' := \pi_2(V')$, donde $\pi_1 : \mathbb{A}^{m_1}(K) \rightarrow \mathbb{A}^n(K)$ y $\pi_2 : \mathbb{A}^{m_2}(K) \rightarrow \mathbb{A}^n(K)$ son las proyecciones canónicas. Supongamos también que $\deg_{\pi}(C) = \deg_z(V)$

y $\deg_\pi(C') = \deg_z(V')$. Como $m_1 \geq m_2$, podemos definir $V'' := V' \times \mathbb{A}^{m_1 - m_2}(K)$ y por el Teorema 1.2.7 sabemos que $\deg(V'') = \deg(V')$. Por otro lado, es también evidente que $\pi_1(V'') = \pi_2(V') = C'$ y que $\pi_1(V \cup V'') = C \cup C'$. Por lo tanto, tenemos que:

$$\deg_\pi(C \cup C') \leq \deg_z(V \cup V'') \leq \deg_z(V) + \deg_z(V'') = \deg_\pi(C) + \deg_\pi(C').$$

La tercera desigualdad es una consecuencia inmediata de la Definición 13. A partir de dos descomposiciones de grado *LCI* mínimo de C y C' , podemos reconstruir una descomposición de $C \cup C'$ que satisface las propiedades del Lema 1.1.2. El grado de tal descomposición es obviamente mayor o igual que el grado de la unión de $C \cup C'$. \square

EJEMPLO 2.1.3 (Tres “grados” distintos). Las tres nociones de grado \deg_z , \deg_π y \deg_{lci} son distintas. Con el objetivo de simplificar los argumentos, modificamos ligeramente el conjunto constructible que aparece en el Ejemplo 1.1.3 (i.e. La Croix de Berny). En primer lugar, consideramos la siguiente hipersuperficie cuadrática:

$$(2.1.2) \quad W' := \{(x, y, z) \in \mathbb{A}^3(\mathbb{C}) : xz + (y^2 - 1) = 0\},$$

y el subconjunto constructible $C := \pi(W') \subseteq \mathbb{A}^2(\mathbb{C})$, donde $\pi : \mathbb{A}^3(\mathbb{C}) \rightarrow \mathbb{A}^2(\mathbb{C})$ es la proyección canónica que “olvida” la variable z . Dado que C es un conjunto denso Zariski en $\mathbb{A}^2(\mathbb{C})$, obviamente tenemos que $\deg_z(C) = 1$. Como C es la proyección de W' y $\deg(W') = 2$, tenemos que $\deg_\pi(C) \leq 2$. Además, $\deg_\pi(C) \neq 1$ ya que, en otro caso, C sería la proyección de un abierto Zariski no vacío de una variedad afín lineal, lo cual es imposible ya que C no es un subconjunto abierto de $\mathbb{A}^2(\mathbb{C})$. Por otro lado, vamos a demostrar que $\deg_{\text{lci}}(C) = 3$. Tenemos la siguiente descomposición de C :

$$(2.1.3) \quad C := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x \neq 0\} \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x^2 + y^2 - 1 = 0\}.$$

Por lo tanto, concluimos que $\deg_{\text{lci}}(C) \leq 3$. A continuación, veamos que $2 < \deg_{\text{lci}}(C)$. Si suponemos que $\deg_{\text{lci}}(C) = 2$, como el conjunto abierto $C_1 = \mathbb{A}^2(\mathbb{C}) \setminus V_{\mathbb{A}}(X)$ es máximo (ya que C no es un conjunto localmente cerrado), entonces C podría descomponerse como unión de a lo sumo dos conjuntos localmente cerrados irreducibles:

$$C = \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x \neq 0\} \cup B,$$

donde B es un conjunto localmente cerrado irreducible de grado 1 (es decir, B es un subconjunto abierto Zariski de una subvariedad afín lineal de $\mathbb{A}^2(\mathbb{C})$). Consideremos ahora la siguiente intersección:

$$C \cap \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\} = B \cap \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\}.$$

Como

$$C_1 \cap \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\} = \emptyset,$$

$$C \cap \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\} = \{(0, 1), (0, -1)\}$$

y la Desigualdad de Bézout para conjuntos localmente cerrados (cf. Teorema 1.2.8) se verifica, concluimos que:

$$2 = \#(B \cap \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\}) \leq \deg(B) \deg(\{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\}) = 1 \cdot 1 = 1,$$

lo cual es falso. Por lo tanto, hemos probado que $\deg(B) \geq 2$ y, así, $3 \geq \deg_{\text{lci}}(C) = 1 + \deg(B) \geq 3$ como queríamos demostrar. Por consiguiente, existen conjuntos constructibles $C \subseteq \mathbb{A}^2(\mathbb{C})$ para los cuales las tres nociones de grado mencionadas previamente toman valores distintos:

$$1 = \deg_z(C) < 2 = \deg_\pi(C) < \deg_{\text{lci}}(C) = 3.$$

Por otro lado, no hay unicidad en las componentes inmersas de una descomposición de conjuntos constructibles que minimice el grado *LCI*:

OBSERVACIÓN 2.1.4 (No hay unicidad en las componentes inmersas que minimizan el grado *LCI*). Al igual que en el Ejemplo 2.1.3 anterior, consideramos un conjunto constructible $C \subseteq \mathbb{A}^2(\mathbb{C})$, denso en $\mathbb{A}^2(\mathbb{C})$ con respecto a la topología de Zariski y de grado *LCL* igual a 3, que viene dado por la siguiente descomposición como unión finita de subconjuntos localmente cerrados irreducibles que minimiza el grado:

$$C = \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x \neq 0\} \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x^2 + y^2 - 1 = 0\}.$$

Para cada $a \in \mathbb{C}$, definimos el polinomio:

$$h_a(X, Y) := X^2 + Y^2 + aXY - 1 \in \mathbb{C}[X, Y].$$

Entonces, hay infinitas descomposiciones de C como unión finita de conjuntos localmente cerrados que minimizan el grado *LCI* como se muestra en la siguiente igualdad:

$$C = \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x \neq 0\} \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : h_a(x, y) = 0\}.$$

Esto se debe a que los ideales $(X, X^2 + Y^2 - 1)$ y $(X, h_a(X, Y))$ son el mismo ideal en $\mathbb{C}[X, Y]$ para cada $a \in \mathbb{C}$.

Nuestro grado LCI generaliza la noción de grado para conjuntos localmente cerrados introducida en [He, 83]. Resumimos este hecho en la siguiente proposición:

PROPOSICIÓN 2.1.5. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible que sea localmente cerrado o presentable equidimensionalmente. Entonces, $\deg_z(C)$ y $\deg_{\text{lci}}(C)$ coinciden. Es decir,*

$$\deg_z(C) = \deg(\overline{C}^z) = \deg_{\text{lci}}(C).$$

DEMOSTRACIÓN. Si C es localmente cerrado, entonces existe una biyección entre sus componentes irreducibles y las de su clausura Zariski. Esta correspondencia permite concluir la igualdad del enunciado. Por otro lado, si C es presentable equidimensionalmente, el resultado se deduce de la Proposición 1.1.5. \square

OBSERVACIÓN 2.1.6 (Grado LCI y descomposición en piezas presentables equidimensionalmente). También podemos considerar una descomposición en piezas presentables equidimensionalmente de un conjunto constructible $C \subseteq \mathbb{A}^n(K)$ de dimensión r como en la Observación 1.1.6. Al igual que en dicha observación, sea $\mathcal{C} := \{C_r, \dots, C_0\}$ una descomposición de C como unión finita de conjuntos constructibles presentables equidimensionalmente. Consideramos la siguiente cantidad asociada a esta descomposición:

$$\deg_z(C, \mathcal{C}) = \sum_{i=0}^r \deg_z(C_i).$$

Dado que \deg_{lci} es una función sub-aditiva (cf. Proposición 2.1.2) y gracias a la Proposición 2.1.5, tenemos que:

$$\deg_{\text{lci}}(C) \leq \deg_z(C, \mathcal{C}).$$

Además, tomando una descomposición de C en conjuntos localmente cerrados irreducibles que minimiza el grado LCI :

$$C := W_1 \cup \dots \cup W_s,$$

tal que:

$$\deg_{\text{lci}}(C) := \sum_{i=1}^s \deg(W_i),$$

y reorganizando esta descomposición en piezas de la misma dimensión, obtenemos una descomposición $\mathcal{W} := \{D_r, \dots, D_0\}$ que verifica:

$$\deg_{\text{lci}}(C) := \sum_{i=0}^r \deg_{\text{lci}}(D_i) = \sum_{i=0}^r \deg_z(D_i).$$

Por lo tanto, el grado LCI podría haberse definido en términos de descomposiciones en piezas presentables equidimensionalmente de la siguiente manera:

$$\deg_{\text{lci}}(C) = \min\{\deg_z(C, \mathcal{C}) : \mathcal{C} \text{ es una descomposición de } C \text{ en piezas presentables equidimensionalmente como en la Observación 1.1.6}\}.$$

PROPOSICIÓN 2.1.7. *Para todo subconjunto constructible $C \subseteq \mathbb{A}^n(K)$ de dimensión r , se verifican las siguientes desigualdades:*

$$\deg_z(C) \leq \deg_\pi(C) \leq \deg_{\text{lci}}(C).$$

Además, dada una descomposición de C en conjuntos constructibles presentables equidimensionalmente:

$$C := C_r \cup \dots \cup C_0,$$

tenemos que:

$$\deg_{\text{lci}}(C) \leq (\dim(C) + 1) \max\{\deg_z(C_i) : 0 \leq i \leq r\}.$$

DEMOSTRACIÓN. Sea $V \in \Pi_m(C)$ un subconjunto localmente cerrado de $\mathbb{A}^m(K)$ tal que $\pi(V) := C$, donde $\pi : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K)$ es la proyección canónica, y $\deg_\pi(C) = \deg_z(V)$. Gracias a la Proposición 1.2.6, tenemos que:

$$\deg_z(C) = \deg(\overline{\pi(V)}^z) \leq \deg(V) = \deg_\pi(C),$$

lo que prueba la primera desigualdad.

Para la segunda desigualdad, supongamos una descomposición de C como unión finita de subconjuntos localmente cerrados irreducibles:

$$C := W_1 \cup \dots \cup W_r,$$

tal que:

$$\deg_{\text{lci}}(C) = \sum_{i=1}^r \deg(W_i).$$

Como \deg_π es una función sub-aditiva, obtenemos:

$$\deg_\pi(C) \leq \sum_{i=1}^r \deg_\pi(W_i).$$

Por lo tanto, como W_i es localmente cerrado, de la Desigualdad (2.1.1), concluimos $\deg_\pi(W_i) \leq \deg(W_i)$ y la segunda desigualdad es cierta. La última desigualdad se sigue de forma sencilla a partir de la Observación 2.1.6 anterior. \square

OBSERVACIÓN 2.1.8. En general, no es cierto que:

$$\deg_{\text{lci}}(C) \leq (\dim(C) + 1) \deg_z(C).$$

Basta considerar el siguiente ejemplo, que extiende al conjunto constructible introducido en el Ejemplo 2.1.3:

$$C' := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x \neq 0\} \cup \{(0, \pm 1)\} \cup \{(0, \pm 2)\}.$$

Usando argumentos similares a los empleados en el Ejemplo 2.1.3, se puede demostrar que $\deg_{\text{lci}}(C') = 5$. Por lo tanto, este ejemplo prueba que:

$$5 = \deg_{\text{lci}}(C') \not\leq (\dim(C') + 1) \deg_z(C') = 3 \cdot 1 = 3.$$

Nuestro π -grado también generaliza la noción de grado para conjuntos localmente cerrados (y conjuntos constructibles presentables equidimensionalmente):

COROLARIO 2.1.9. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible. Si C es localmente cerrado o presentable equidimensionalmente, tenemos que:*

$$\deg_z(C) = \deg(\overline{C}^z) = \deg_\pi(C) = \deg_{\text{lci}}(C).$$

Además, si $C \subseteq \mathbb{A}^n(K)$ es un conjunto constructible presentable equidimensionalmente de dimensión r , entonces existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que se cumple lo siguiente para cada $A \in \mathbb{G}(C)$:

$$\sharp(A \cap C) = \deg_z(C) = \deg_\pi(C) = \deg_{\text{lci}}(C).$$

DEMOSTRACIÓN. Combinando la Proposición 2.1.7 y la Proposición 2.1.5, obtenemos la primera igualdad de forma inmediata. Por otro lado, la segunda afirmación se deduce también directamente a partir las definiciones y de la Proposición 1.2.2. \square

En el caso general, lo único que tenemos para intersecciones con variedades afines lineales es el siguiente resultado:

COROLARIO 2.1.10. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible de dimensión r y sean V_1, \dots, V_s las componentes irreducibles de dimensión más alta de la clausura Zariski de C (como en la Proposición 1.1.5). Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que se cumple lo siguiente para cada $A \in \mathbb{G}(C)$:*

$$\sharp(A \cap C) = \sum_{i=1}^s \deg(V_i) \leq \deg_z(C) \leq \deg_\pi(C) \leq \deg_{\text{lci}}(C).$$

Los siguientes resultados corrigen y explican el Lema 2 de [He, 83] sobre el grado de las imágenes por aplicaciones lineales de conjuntos constructibles. De nuevo, comenzamos con el grado LCI:

PROPOSICIÓN 2.1.11. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ una aplicación lineal. Tenemos que:*

i) *El grado de la clausura Zariski de la imagen está acotado superiormente de la siguiente manera:*

$$\deg_z(\ell(C)) = \deg(\overline{\ell(C)}^z) \leq \deg_{\text{lci}}(C).$$

ii) *Si $\ell(C)$ es localmente cerrado o presentable equidimensionalmente, también se verifica que:*

$$\deg_z(\ell(C)) = \deg_{\text{lci}}(\ell(C)) \leq \deg_{\text{lci}}(C).$$

iii) *En general, no es cierto que dada una variedad afín lineal $A \subseteq \mathbb{A}^m(K)$ se cumpla lo siguiente:*

$$\sharp(A \cap \ell(C)) \leq \deg_z(C).$$

iv) *En general, no es cierto que $\deg_{\text{lci}}(\ell(C)) \leq \deg_{\text{lci}}(C)$.*

DEMOSTRACIÓN. El primer ítem es una consecuencia inmediata de la Proposición 1.2.6 combinada con la sub-aditividad del grado LCI (cf. Proposición 2.1.2 previa).

El Ítem *ii*) se deduce del hecho de que $\deg_{\text{Ici}}(C) = \deg_z(C)$ para todo subconjunto constructible $C \subseteq \mathbb{A}^n(K)$ que es localmente cerrado o presentable equidimensionalmente y la Proposición 1.2.6.

Para el Ítem *iii*), consideramos el conjunto constructible $C \subseteq \mathbb{A}^2(\mathbb{C})$ descrito en el Ejemplo 2.1.3 anterior, la Identidad $Id : C \rightarrow \mathbb{A}^2(\mathbb{C})$ como aplicación lineal y la recta $A := \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x = 0\}$ del plano. Entonces, tenemos que la intersección $A \cap Id(C)$ está formada por dos puntos, mientras que $\deg_z(C) = 1$. Por lo tanto,

$$2 = \#(A \cap Id(C)) \not\leq \deg_z(C) = 1.$$

Para el Ítem *iv*), empleamos de nuevo el conjunto constructible $C \subseteq \mathbb{A}^2(\mathbb{C})$ introducido en el Ejemplo 2.1.3 anterior. En dicho ejemplo, demostramos que C es un conjunto constructible tal que $\deg_{\text{Ici}}(C) = 3$ y, al mismo tiempo, es la proyección de una variedad algebraica $W' \subseteq \mathbb{A}^3(\mathbb{C})$ de grado 2. Por consiguiente, tenemos que:

$$3 = \deg_{\text{Ici}}(\pi(W')) \not\leq \deg_{\text{Ici}}(W') = \deg(W') = 2,$$

lo que prueba el Ítem *iv*). □

Por otro lado, tenemos que el π -grado se comporta de manera “adecuada” para las imágenes de conjuntos constructibles por aplicaciones lineales. Esto se muestra en el siguiente resultado:

PROPOSICIÓN 2.1.12. *Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Para todo $m \geq n$, consideramos la clase $\mathcal{L}_m(C)$ de todos los pares (V, ℓ) , donde $V \subseteq \mathbb{A}^m(K)$ es un conjunto localmente cerrado, $\ell : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K)$ es una aplicación lineal y $\ell(V) = C$. Finalmente, consideramos la clase:*

$$\mathcal{L}(C) := \bigcup_{m \geq n} \mathcal{L}_m(C).$$

Entonces, tenemos que:

i) El π -grado satisface:

$$\deg_{\pi}(C) = \min\{\deg_z(V) : \exists \ell : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K), (V, \ell) \in \mathcal{L}(C)\}.$$

ii) Para toda aplicación lineal $\lambda : \mathbb{A}^n(K) \rightarrow \mathbb{A}^r(K)$, se cumple:

$$\deg_z(\lambda(C)) \leq \deg_{\pi}(\lambda(C)) \leq \deg_{\pi}(C) \leq \deg_{\text{Ici}}(C).$$

DEMOSTRACIÓN. Demostramos cada ítem por separado:

- *Ítem i)*: Sea $(V, \ell) \in \mathcal{L}(C)$ uno de los pares descritos en el enunciado. Sea $G \subseteq \mathbb{A}^{m+n}(K)$ el grafo de la restricción $\ell|_V$. En primer lugar, nótese que G es también un conjunto localmente cerrado, ya que:

$$G := \{(z, y) \in \mathbb{A}^m(K) \times \mathbb{A}^n(K) : z \in V, \ell(z) - y = 0\}.$$

Obviamente, G es la intersección de $V \times \mathbb{A}^n(K)$ con la variedad afín lineal:

$$A := \mathbb{A}^m(K) \times \{(z, y) \in \mathbb{A}^{m+n}(K) : \ell(z) - y = 0\}.$$

Por lo tanto, aplicando la Desigualdad de Bézout para conjuntos localmente cerrados (Teorema 1.2.8 anterior), obtenemos:

$$\deg_z(G) = \deg(G) \leq \deg(V \times \mathbb{A}^n(K)) \deg(A) = \deg(V) \cdot 1 = \deg(V).$$

Además, sea $\pi : \mathbb{A}^{m+n}(K) \rightarrow \mathbb{A}^n(K)$ la proyección canónica dada por:

$$\pi(z, y) := y, \forall (z, y) \in \mathbb{A}^{m+n}(K).$$

Es evidente que $C := \pi(G)$ y, por lo tanto, concluimos que:

$$\deg_{\pi}(C) \leq \deg_z(V), \forall (V, \ell) \in \mathcal{L}(C).$$

La otra desigualdad es evidente ya que las proyecciones son casos particulares de aplicaciones lineales. Por lo tanto, hemos demostrado el primer ítem.

- *Ítem ii)*: Si $\lambda : \mathbb{A}^n(K) \rightarrow \mathbb{A}^r(K)$ es una aplicación lineal, entonces tenemos que:

$$\mathcal{L}(\lambda(C)) \supseteq \{(V, \lambda \circ \ell) : (V, \ell) \in \mathcal{L}(C)\}.$$

Por lo tanto, como:

$$\deg_{\pi}(\lambda(C)) = \min\{\deg_z(V) : \exists \rho, (V, \rho) \in \mathcal{L}(\lambda(C))\},$$

inmediatamente concluimos que:

$$\deg_{\pi}(\lambda(C)) \leq \min\{\deg_z(V) : \exists \ell, (V, \ell) \in \mathcal{L}(C), (V, \lambda \circ \ell) \in \mathcal{L}(\lambda(C))\} = \deg_{\pi}(C),$$

y el Ítem *ii*) se verifica. □

En lo que respecta al resto de operaciones naturales sobre conjuntos constructibles, tanto el grado *LCI* como el π -grado cumplen las condiciones establecidas en el paradigma introducido en la Subsección 0.3.1.

PROPOSICIÓN 2.1.13. Sean $C \subseteq \mathbb{A}^n(K)$ y $D \subseteq \mathbb{A}^m(K)$ dos conjuntos constructibles.

i) Para toda variedad afín lineal $A \subseteq \mathbb{A}^n(K)$, se verifican las siguientes desigualdades:

$$\begin{aligned} \deg_{\text{LCI}}(A \cap C) &\leq \deg_{\text{LCI}}(C), \\ \deg_{\pi}(A \cap C) &\leq \deg_{\pi}(C). \end{aligned}$$

ii) Los grados del producto cartesiano satisfacen las siguientes desigualdades:

$$\begin{aligned} \deg_{\text{LCI}}(C \times D) &\leq \deg_{\text{LCI}}(C) \deg_{\text{LCI}}(D), \\ \deg_{\pi}(C \times D) &\leq \deg_{\pi}(C) \deg_{\pi}(D). \end{aligned}$$

iii) Para todo subconjunto abierto Zariski $U \subseteq \mathbb{A}^n(K)$, tenemos que:

$$\begin{aligned} \deg_{\text{LCI}}(C \cap U) &\leq \deg_{\text{LCI}}(C), \\ \deg_{\pi}(C \cap U) &\leq \deg_{\pi}(C). \end{aligned}$$

DEMOSTRACIÓN. Demostramos las desigualdades primero para \deg_{LCI} y después para \deg_{π} :

- *Desigualdades para \deg_{LCI}* : Los Ítems *i*) y *ii*) se deducen de forma sencilla de la sub-aditividad de \deg_{LCI} (cf. Proposición 2.1.2) y de las correspondientes propiedades para conjuntos localmente cerrados. Para el Ítem *iii*), supongamos que C se descompone como unión finita de conjuntos localmente cerrados irreducibles:

$$C = W_1 \cup \dots \cup W_s,$$

tales que

$$\deg_{\text{LCI}}(C) = \sum_{i=1}^s \deg(W_i).$$

Como U es abierto Zariski, entonces para cada i , $1 \leq i \leq s$, tenemos únicamente dos opciones: o bien $U \cap W_i = \emptyset$, o bien $U \cap W_i \neq \emptyset$, en cuyo caso $U \cap W_i$ es localmente cerrado y denso Zariski en $\overline{W_i}^z$. Por lo tanto, obtenemos:

$$\deg_{\text{LCI}}(C \cap U) \leq \sum_{i=1}^s \deg(U \cap W_i) \leq \sum_{i=1}^s \deg(W_i) = \deg_{\text{LCI}}(C).$$

- *Desigualdades para \deg_{π}* : Para el Ítem *i*), supongamos que existe un subconjunto localmente cerrado $V \subseteq \mathbb{A}^m(K)$ tal que $\pi(V) := C$ y $\deg_{\pi}(C) = \deg_z(V)$, donde $\pi : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K)$ es la proyección canónica. Consideramos la subvariedad afín lineal $B := \pi^{-1}(A) \subseteq \mathbb{A}^m(K)$. Tenemos que $\pi(V \cap B) = C \cap A$ y, por consiguiente:

$$\deg_{\pi}(C \cap A) \leq \deg_z(V \cap B) \leq \deg_z(V).$$

Para el Ítem *ii*), supongamos que existen dos conjuntos localmente cerrados $V \subseteq \mathbb{A}^{m_1}(K)$ y $W \subseteq \mathbb{A}^{m_2}(K)$ tales que $m_1 \geq n$ y $m_2 \geq m$, y $\pi_1(V) := C$, $\pi_2(W) := D$, donde $\pi_1 : \mathbb{A}^{m_1}(K) \rightarrow \mathbb{A}^n(K)$ y $\pi_2 : \mathbb{A}^{m_2}(K) \rightarrow \mathbb{A}^m(K)$ son las proyecciones canónicas, $\deg_{\pi}(C) = \deg_z(V)$ y $\deg_{\pi}(D) = \deg_z(W)$. Consideramos el espacio afín $\mathbb{A}^{m_1+m_2}(K) := \mathbb{A}^{m_1}(K) \times \mathbb{A}^{m_2}(K)$ y la proyección producto:

$$\begin{aligned} \bar{\pi} := \pi_1 \times \pi_2 : \mathbb{A}^{m_1}(K) \times \mathbb{A}^{m_2}(K) &\longrightarrow \mathbb{A}^n(K) \times \mathbb{A}^m(K) \\ (z_1, z_2) &\longmapsto (\pi_1(z_1), \pi_2(z_2)). \end{aligned}$$

Entonces, $\bar{\pi}(V \times W) = C \times D$ y, como $V \times W$ es localmente cerrado, concluimos:

$$\deg_{\pi}(C \times D) \leq \deg(V \times W) = \deg_z(V) \deg_z(W) = \deg_{\pi}(C) \deg_{\pi}(D).$$

Finalmente, para el Ítem *iii*), tenemos que existe $V \subseteq \mathbb{A}^m(K)$ localmente cerrado tal que $\pi(V) := C$ y $\deg(V) = \deg_{\pi}(C)$, donde $\pi : \mathbb{A}^m(K) \rightarrow \mathbb{A}^n(K)$ es la proyección canónica. Tomando el subconjunto abierto $\tilde{U} := \pi^{-1}(U) \subseteq \mathbb{A}^m(K)$, tenemos que $C \cap U = \pi(V \cap \tilde{U})$ y las siguientes desigualdades se verifican ya que V es localmente cerrado:

$$\deg_{\pi}(C \cap U) \leq \deg_z(V \cap \tilde{U}) \leq \deg_z(V) = \deg_{\pi}(C).$$

□

Finalmente, demostramos que tanto el grado *LCI* como el π -grado satisfacen una Desigualdad de Bézout.

TEOREMA 2.1.14 (Desigualdades de Bézout para conjuntos constructibles). Sean $C, D \subseteq \mathbb{A}^n(K)$ dos conjuntos constructibles. Tenemos que:

$$\deg_{\text{lci}}(C \cap D) \leq \deg_{\text{lci}}(C) \deg_{\text{lci}}(D), \quad \deg_{\pi}(C \cap D) \leq \deg_{\pi}(C) \deg_{\pi}(D).$$

DEMOSTRACIÓN. Primero demostramos la desigualdad para \deg_{π} y, posteriormente, para \deg_{lci} . En el caso de \deg_{π} , el resultado se sigue de las definiciones del modo esperable. Consideramos en primer lugar el producto cartesiano $C \times D \subseteq \mathbb{A}^{2n}(K)$. A continuación, sea $\Delta^{(2n)} \subseteq \mathbb{A}^{2n}(K)$ la subvariedad afín lineal (llamada habitualmente subvariedad diagonal) dada por la siguiente identidad:

$$\Delta^{(2n)} := \{(x_1, \dots, x_n, y_1, \dots, y_n) \in \mathbb{A}^{2n}(K) : x_i - y_i = 0, 1 \leq i \leq n\}.$$

Sea $\pi : \mathbb{A}^{2n}(K) \rightarrow \mathbb{A}^n(K)$ la proyección canónica sobre las primeras n variables. Tenemos que:

$$\pi\left((C \times D) \cap \Delta^{(2n)}\right) = C \cap D.$$

Aplicando el Ítem *ii*) de la Proposición 2.1.12, obtenemos:

$$\deg_{\pi}(C \cap D) \leq \deg_{\pi}\left((C \times D) \cap \Delta^{(2n)}\right).$$

Como $\Delta^{(2n)}$ es una subvariedad afín lineal, deducimos del Ítem *i*) de la Proposición 2.1.13 que:

$$\deg_{\pi}(C \cap D) \leq \deg_{\pi}(C \times D).$$

Finalmente, el Ítem *ii*) de la Proposición 2.1.13 nos proporciona la desigualdad final para el π -grado:

$$\deg_{\pi}(C \cap D) \leq \deg_{\pi}(C \times D) \leq \deg_{\pi}(C) \deg_{\pi}(D).$$

En el caso de \deg_{lci} , la dificultad radica en que el grado no se conserva mediante transformaciones lineales (como se observó en el Ítem *iii*) de la Proposición 2.1.11). Por lo tanto, debemos realizar algunos ajustes sutiles en el esquema de demostración que hemos utilizado para \deg_{π} . Como antes, consideramos la subvariedad diagonal $\Delta^{(2n)} \subseteq \mathbb{A}^{2n}(K)$ y su intersección con el producto cartesiano:

$$C' := (C \times D) \cap \Delta^{(2n)}.$$

Por la Proposición 2.1.13, concluimos que:

$$\deg_{\text{lci}}(C') \leq \deg_{\text{lci}}(C \times D) \leq \deg_{\text{lci}}(C) \deg_{\text{lci}}(D).$$

A continuación, consideramos una descomposición de grado LCI mínimo de C en subconjuntos localmente cerrados irreducibles como en la Definición 13. Es decir, una descomposición:

$$C' = W_1 \cup \dots \cup W_s,$$

donde $W_i = U_i \cap V_i \neq \emptyset$, $U_i \subseteq \Delta^{(2n)}$ es un conjunto abierto Zariski en $\Delta^{(2n)}$, $V_i \subseteq \Delta^{(2n)}$ es una subvariedad algebraica irreducible para la topología de Zariski y tal que:

$$\deg_{\text{lci}}(C') = \sum_{i=1}^s \deg(W_i) = \sum_{i=1}^s \deg(V_i).$$

Para cada i , $1 \leq i \leq s$, consideramos:

- $O_i \subseteq \mathbb{A}^n(K)$ dado por:

$$O_i := \{\underline{x} \in \mathbb{A}^n(K) : (\underline{x}, \underline{x}) \in U_i\},$$

- $Q_i \subseteq \mathbb{A}^n(K)$ dado por:

$$Q_i := \{\underline{x} \in \mathbb{A}^n(K) : (\underline{x}, \underline{x}) \in V_i\}.$$

Es inmediato que, si $\pi : \mathbb{A}^{2n}(K) \rightarrow \mathbb{A}^n(K)$ es la proyección sobre las n primeras coordenadas, entonces la restricción $\pi|_{\Delta^{(2n)}} : \Delta^{(2n)} \rightarrow \mathbb{A}^n(K)$ es un isomorfismo birreglar. Por lo tanto, concluimos que O_i es un conjunto abierto Zariski en $\mathbb{A}^n(K)$, Q_i es un subconjunto cerrado Zariski de $\mathbb{A}^n(K)$ y tenemos que:

$$\pi(W_i) = O_i \cap Q_i.$$

Como $\pi(W_i)$ es localmente cerrado, por la Proposición 1.2.6 obtenemos:

$$\deg(\pi(W_i)) = \deg(\overline{\pi(W_i)}^z) = \deg(Q_i) = \deg(O_i \cap Q_i) \leq \deg(W_i).$$

Además, tenemos que:

$$C \cap D = \pi((C \times D) \cap \Delta^{(2n)}) = (O_1 \cap Q_1) \cup \dots \cup (O_s \cap Q_s).$$

Finalmente, aplicando la Proposición 2.1.2, concluimos la Desigualdad de Bézout para el grado LCI :

$$\deg_{\text{Lci}}(C \cap D) \leq \sum_{i=1}^s \deg(O_i \cap Q_i) \leq \sum_{i=1}^s \deg(W_i) = \deg_{\text{Lci}}(C') \leq \deg_{\text{Lci}}(C) \deg_{\text{Lci}}(D).$$

□

2.2. Cotas superiores para la intersección de varios conjuntos constructibles

De acuerdo con el paradigma que hemos introducido en la Subsección 0.3.1, una noción “correcta” de grado para conjuntos constructibles debe verificar alguna variante de la Proposición 2.3 de [HS, 82]. Dicha proposición establece una elegante cota superior para la intersección de varias variedades algebraicas. El objetivo de esta sección es generalizar la Proposición 2.3 de [HS, 82] de diversas formas.

En primer lugar, nótese que una aplicación literal de la Desigualdad de Bézout coloca la co-dimensión en el exponente de las cotas para el grado. Es decir, sean $f_1, \dots, f_m \in K[X_1, \dots, X_n]$ tales que $V := V_{\mathbb{A}}(f_1, \dots, f_m)$ tiene dimensión $n - m$, y sea $W \subseteq \mathbb{A}^n(K)$ una variedad algebraica. La Desigualdad de Bézout nos da la siguiente cota superior:

$$\deg(W \cap V_{\mathbb{A}}(f_1, \dots, f_m)) \leq \deg(W) \prod_{i=1}^m \deg(f_i) \leq \deg(W) (\max\{\deg(f_1), \dots, \deg(f_m)\})^{\text{codim}(V)}.$$

Sin embargo, la Proposición 2.3 de [HS, 82] reemplaza la co-dimensión de V por la dimensión de W , enfatizando el papel de la dimensión de W en este tipo de cotas superiores. Es decir, la Proposición 2.3 de [HS, 82] nos proporciona el siguiente resultado:

$$\deg(W \cap V_{\mathbb{A}}(f_1, \dots, f_m)) \leq \deg(W) (\max\{\deg(f_1), \dots, \deg(f_m)\})^{\dim(W)}.$$

El “trade-off” entre estas dos cotas superiores es el ingrediente clave para demostrar la existencia y la alta probabilidad de conjuntos cuestores de longitud asintóticamente óptima (ver Teorema 3.2.1 de la Sección 3.2). Nuestro objetivo es extender este tipo de cotas para la intersección de varios conjuntos constructibles. Descartamos \deg_z ya que no satisface una Desigualdad de Bézout. Por otro lado, no sabemos si este tipo de resultados es válido para \deg_{π} (y no hemos encontrado ningún contra-ejemplo). Por lo tanto, esta sección se centrará en demostrar una generalización de la Proposición 2.3 de [HS, 82] para familias finitas de conjuntos constructibles y \deg_{Lci} .

Nuestro resultado principal es el siguiente:

TEOREMA 2.2.1 (Cotas superiores para el grado LCI de la intersección de varios conjuntos constructibles). *Sea $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ una familia de conjuntos constructibles. Sea $r := \dim(C_1) \geq 0$ la dimensión del conjunto constructible C_1 . Se verifican las siguientes desigualdades:*

i) Primera cota superior:

$$\deg_{\pi} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{Lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \binom{s+r-1}{r} \deg_{\text{Lci}}(C_1) (\max\{\deg_{\text{Lci}}(C_j) : 2 \leq j \leq s\})^r.$$

ii) Segunda cota superior:

$$\deg_{\pi} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{Lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{Lci}}(C_1) \left(1 + \sum_{i=2}^s \deg_{\text{Lci}}(C_i) \right)^r.$$

iii) Cota superior en términos del grado promedio: *dada una familia de conjuntos constructibles*

$$C_1, \dots, C_s \subseteq \mathbb{A}^n(K),$$

definimos su grado promedio como:

$$\deg_{\text{av}}(C_1, \dots, C_s) := \frac{1}{s} \sum_{i=1}^s \deg_{\text{Lci}}(C_i).$$

Entonces, también tenemos que:

$$\deg_{\pi} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{Lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \deg_{\text{Lci}}(C_1) s^r (\deg_{\text{av}}(C_1, \dots, C_s))^r.$$

La primera cota superior del teorema anterior generaliza la cota principal de la Proposición 2.3 de [HS, 82]. Sin embargo, el coeficiente constante es exponencial en $s - 1$ ya que:

$$\left(\frac{r+s-1}{s-1}\right)^{s-1} \leq \binom{r+s-1}{r} \leq \left(\frac{e(r+s-1)}{s-1}\right)^{s-1},$$

donde e es la base del logaritmo neperiano. El lector también puede pensar que la constante es también exponencial en r dado que:

$$\left(\frac{r+s-1}{r}\right)^r \leq \binom{r+s-1}{r} \leq \left(\frac{e(r+s-1)}{r}\right)^r.$$

En cualquier caso, no es la constante 1 que esperaríamos de acuerdo a la Proposición 2.3 de [HS, 82]. La tercera cota superior se obtiene simplemente reescribiendo nuestra segunda cota superior, ya que siempre se cumple que $\deg_{\text{lci}}(C_1) \geq 1$. Por lo tanto, la segunda y la tercera cotas superiores son exponenciales en r , pero polinómicas en s , lo cual puede ser útil en algunas aplicaciones. Nótese que las desigualdades del lado izquierdo para el π -grado son consecuencia inmediata de la Proposición 2.1.7.

Esta sección está estructurada de la siguiente forma. En primer lugar, en la Subsección 2.2.1, demostraremos que la Proposición 2.3 de [HS, 82] se puede extender fácilmente al caso de la intersección de varios conjuntos localmente cerrados (ver Proposición 2.2.2). La demostración emplea un argumento inductivo elemental que se beneficia de la siguiente idea:

Dado un conjunto localmente cerrado irreducible Ω y otro conjunto localmente cerrado C , la intersección $\Omega \cap C$ debe verificar:

- O bien $\dim(\Omega \cap C) = \dim(\Omega)$, en cuyo caso $\Omega \cap C$ es un subconjunto abierto Zariski en Ω y, por lo tanto, $\deg(\Omega \cap C) = \deg(\Omega)$.
- O $\dim(\Omega \cap C) < \dim(\Omega)$ y podemos redirigir nuestros argumentos a los mismos pasos inductivos.

Esta dicotomía no es cierta cuando C es cualquier conjunto constructible. Puede ocurrir que $\dim(\Omega \cap C) = \dim(\Omega)$, pero $\deg_{\text{lci}}(\Omega \cap C) > \deg_{\text{lci}}(\Omega)$. Ejemplos de este comportamiento contraintuitivo de los conjuntos constructibles se derivan de forma sencilla del Ejemplo 1.1.3 (i.e. La Croix de Berny).

Por lo tanto, para demostrar el Teorema 2.2.1, necesitaremos realizar esfuerzos adicionales. Con este fin, en la Subsección 2.2.2, procedemos mediante otro argumento inductivo basado en la dimensión. Como la dicotomía previa no se mantiene, aparece un número combinatorio, lo que da lugar a la primera cota superior del Teorema 2.2.1.

No sabemos si este número combinatorio es imprescindible. Parece que surge debido al tipo de argumento que hemos escogido. Por este motivo, hemos buscado alternativas y esta es la segunda cota superior del Teorema 2.2.1. En la Subsección 2.2.3, probamos la segunda cota superior y su consecuencia obvia, la cota superior en términos del grado promedio. Para demostrar la segunda cota superior, seguimos una estrategia de prueba basada en “*Dividir y conquistar*”. En lugar de usar un argumento inductivo basado en la dimensión, estudiamos el número de celdas no vacías obtenidas como intersecciones finitas de varias familias finitas de conjuntos localmente cerrados. Esto se puede llevar a cabo con la ayuda de la técnica de conteo que subyace en el Lema 2.2.5.

Estas subsecciones dan lugar a algunos corolarios inmediatos. Por ello, hemos añadido la Subsección 2.2.4 donde presentamos un par de aplicaciones directas de nuestras cotas superiores para la intersección de varios conjuntos constructibles.

2.2.1. Una extensión directa de la Proposición 2.3 de [HS, 82] a conjuntos localmente cerrados.

En la siguiente proposición, extendemos el argumento inductivo de la Proposición 2.3 de [HS, 82] al caso de la intersección de un conjunto constructible con varios conjuntos localmente cerrados.

PROPOSICIÓN 2.2.2. *Sea $\Omega \subseteq \mathbb{A}^n(K)$ un conjunto constructible y sea $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ una familia finita de conjuntos localmente cerrados. Entonces, tenemos que:*

$$\deg_{\text{lci}} \left(\Omega \cap \left(\bigcap_{i=1}^s C_i \right) \right) \leq \deg_{\text{lci}}(\Omega) (\max\{\deg_{\text{lci}}(C_i) : 1 \leq i \leq s\})^{\dim(\Omega)}.$$

DEMOSTRACIÓN. El argumento es por inducción en s como en [HS, 82]. Comenzamos discutiendo el caso $s = 1$. Si $s = 1$ y $\dim(\Omega) \geq 1$, se trata de la Desigualdad de Bézout para conjuntos constructibles (ver Teorema 2.1.14). En el caso $\dim(\Omega) = 0$, es evidente que la intersección de un conjunto finito de puntos con cualquier otro conjunto C_1 está contenido en Ω . Por consiguiente, en este último caso, se cumple de manera obvia que: $\deg(\Omega \cap C_1) \leq \deg(\Omega)$. Por todo lo anterior, queda demostrado el caso $s = 1$.

Supongamos ahora que $s \geq 2$. En primer lugar, supongamos también que Ω es un conjunto localmente cerrado irreducible (i.e. un conjunto abierto no vacío en alguna variedad algebraica irreducible). Descomponemos la intersección de la siguiente forma:

$$\Omega \cap \left(\bigcap_{i=1}^s C_i \right) = (\Omega \cap C_1) \cap \left(\bigcap_{i=2}^s C_i \right).$$

Nótese que $\Omega \cap C_1$ es localmente cerrado. Por lo tanto, podemos considerar los dos casos siguientes:

- *Caso 1:* $\dim(\Omega \cap C_1) = \dim(\Omega)$. En este caso, como Ω es localmente cerrado irreducible, entonces $\Omega \cap C_1$ es un conjunto localmente cerrado denso en $\overline{\Omega}^z$ y, por consiguiente, se cumple que:

$$\deg(\Omega \cap C_1) = \deg(\overline{(\Omega \cap C_1)}^z) = \deg(\overline{\Omega}^z) = \deg(\Omega),$$

donde la última igualdad es cierta ya que Ω es localmente cerrado. Aplicando la hipótesis de inducción, concluimos:

$$\deg \left(\Omega \cap \left(\bigcap_{j=1}^s C_j \right) \right) \leq \deg(\Omega \cap C_1) (\max\{\deg(C_i) : 2 \leq i \leq s\})^{\dim(\Omega)},$$

lo que implica el resultado buscado.

- *Caso 2:* $\dim(\Omega \cap C_1) < \dim(\Omega)$. En este caso, utilizando simplemente la hipótesis de inducción, tenemos que:

$$\deg \left(\Omega \cap \left(\bigcap_{j=1}^s C_j \right) \right) \leq \deg(\Omega \cap C_1) (\max\{\deg(C_i) : 2 \leq i \leq s\})^{\dim(\Omega \cap C_1)}.$$

Empleando la Desigualdad de Bézout para conjuntos constructibles (nuestro Teorema 2.1.14), obtenemos:

$$\deg \left(\Omega \cap \left(\bigcap_{j=1}^s C_j \right) \right) \leq \deg(\Omega) \deg(C_1) (\max\{\deg(C_i) : 2 \leq i \leq s\})^{\dim(\Omega)-1},$$

y esto también nos proporciona la desigualdad buscada.

Para el caso general, consideramos una descomposición de Ω en subconjuntos localmente cerrados irreducibles de grado *LCI* mínimo. Es decir, consideramos una descomposición:

$$\Omega := (U_1 \cap V_1) \cup \dots \cup (U_t \cap V_t),$$

donde $U_i \cap V_i \neq \emptyset$, $U_i \subseteq \mathbb{A}^n(K)$ es un conjunto abierto Zariski, $V_i \subseteq \mathbb{A}^n(K)$ es un subconjunto cerrado irreducible con respecto a la topología de Zariski y tal que se verifica:

$$\deg_{\text{lci}}(\Omega) = \sum_{i=1}^t \deg(V_i).$$

Como el grado es sub-aditivo, tenemos que:

$$\deg_{\text{lci}} \left(\Omega \cap \left(\bigcap_{j=1}^s C_j \right) \right) \leq \sum_{i=1}^t \deg \left((U_i \cap V_i) \cap \left(\bigcap_{j=1}^s C_j \right) \right).$$

Dado que $(U_i \cap V_i)$ es localmente cerrado irreducible, al aplicar los argumentos previos obtenemos:

$$\deg_{\text{lci}} \left(\Omega \cap \left(\bigcap_{j=1}^s C_j \right) \right) \leq \sum_{i=1}^t \deg(V_i) (\max\{\deg(C_j) : 1 \leq j \leq s\})^{\dim(U_i \cap V_i)}.$$

Como $\dim(\Omega) = \max\{\dim(U_i \cap V_i) : 1 \leq i \leq t\}$, se concluye la prueba de la proposición. \square

Para el π -grado, tenemos una cota superior similar para conjuntos constructibles presentables equidimensionalmente:

COROLARIO 2.2.3. *Sea $\Omega \subseteq \mathbb{A}^n(K)$ un conjunto constructible presentable equidimensionalmente y sea $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ una familia finita de conjuntos localmente cerrados. Entonces, tenemos que:*

$$\deg_{\pi} \left(\Omega \cap \left(\bigcap_{i=1}^s C_i \right) \right) \leq \deg_{\pi}(\Omega) (\max\{\deg_{\pi}(C_i) : 1 \leq i \leq s\})^{\dim(\Omega)}.$$

DEMOSTRACIÓN. Esencialmente, se trata de la misma demostración que la de la proposición anterior. Únicamente, cabe señalar que, como Ω es presentable equidimensionalmente, entonces admite una descomposición en conjuntos localmente cerrados irreducibles $\Omega = W_1 \cup \dots \cup W_s$ de modo que se verifica la siguiente propiedad:

$$\deg_\pi(\Omega) = \deg_{\text{gli}}(\Omega) = \sum_{i=1}^s \deg(W_i).$$

□

En el siguiente corolario obtenemos, empleando la Proposición 2.2.2 anterior, estimaciones del grado de las imágenes de conjuntos constructibles por funciones polinomiales.

COROLARIO 2.2.4. *Sea $\varphi := (\varphi_1, \dots, \varphi_m) : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ una función polinomial. Supongamos que para cada i , $1 \leq i \leq m$, $\varphi_i \in K[X_1, \dots, X_n]$ es un polinomio de grado a lo sumo d , donde $d \geq 1$. Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Entonces, tenemos que:*

$$\deg_z(\varphi(C)) \leq \deg_{\text{gli}}(C)d^{\dim(C)},$$

$$\deg_z(\varphi(C)) \leq \deg_\pi(C)d^m \leq \deg_{\text{gli}}(C)d^m.$$

Las cotas no son siempre ciertas si reemplazamos \deg_{gli} por \deg_z en la parte derecha de las desigualdades.

Además, si $\varphi(C)$ es localmente cerrado o presentable equidimensionalmente, entonces también se verifica que:

$$\deg_z(\varphi(C)) = \deg_\pi(\varphi(C)) = \deg_{\text{gli}}(\varphi(C)) \leq \deg_{\text{gli}}(C)d^{\dim(C)},$$

$$\deg_z(\varphi(C)) = \deg_\pi(\varphi(C)) = \deg_{\text{gli}}(\varphi(C)) \leq \deg_\pi(C)d^m \leq \deg_{\text{gli}}(C)d^m.$$

DEMOSTRACIÓN.

- *Desigualdad $\deg_z(\varphi(C)) \leq \deg_{\text{gli}}(C)d^{\dim(C)}$:* Supongamos, en primer lugar, que C es un conjunto localmente cerrado irreducible. Por tanto, la clausura Zariski $W := \overline{\varphi(C)}^z \subseteq \mathbb{A}^m(K)$ de $\varphi(C)$ es una variedad irreducible. Sea $D := \deg(W)$ el grado de W y sea $s := \dim(W)$. Entonces, existe una subvariedad afín lineal $A \subseteq \mathbb{A}^m(K)$ de co-dimensión s tal que:

$$\#(A \cap W) = \#(A \cap \varphi(C)) = \deg(W) = D.$$

Consideremos ahora la variedad algebraica $\varphi^{-1}(A) \subseteq \mathbb{A}^n(K)$. Como A está dada por s ecuaciones lineales, entonces $\varphi^{-1}(A)$ viene también dada por s ecuaciones polinomiales de grado a lo sumo d (combinando las coordenadas de $\varphi_1, \dots, \varphi_m$ de φ con las ecuaciones que definen A en $\mathbb{A}^m(K)$). Por la Proposición 2.2.2, concluimos:

$$\deg_{\text{gli}}(C \cap \varphi^{-1}(A)) \leq \deg_{\text{gli}}(C)d^{\dim(C)}.$$

Sea \mathcal{C} la clase de componentes localmente cerradas irreducibles de $C \cap \varphi^{-1}(A)$. Como el número de componentes irreducibles de un conjunto localmente cerrado está acotado por su grado (ver Ítem *iii*) de la Proposición 1.2.3), concluimos:

$$\#(\mathcal{C}) \leq \deg_{\text{gli}}(C \cap \varphi^{-1}(A)) \leq \deg_{\text{gli}}(C)d^{\dim(C)}.$$

Sea $V \in \mathcal{C}$ una de las componentes irreducibles de $C \cap \varphi^{-1}(A)$. Sabemos que su clausura Zariski $\overline{\varphi(V)}^z$ es irreducible y, al mismo tiempo, tenemos que:

$$\varphi(V) \subseteq A \cap \varphi(C).$$

Como $A \cap \varphi(C)$ es un número finito de puntos, $\varphi(V)$ debe ser un punto en $\mathbb{A}^m(K)$. Además, como $\varphi(C \cap \varphi^{-1}(A)) = \varphi(C) \cap A$, existe una aplicación suprayectiva entre los siguientes conjuntos finitos:

$$\begin{aligned} \Phi : \mathcal{C} &\longrightarrow \varphi(C) \cap A \\ V &\longmapsto \Phi(V) := \varphi(V) \end{aligned}$$

Entonces, concluimos que si C es localmente cerrado irreducible, tenemos que:

$$\deg(\overline{\varphi(C)}^z) = \deg(W) = \#(\varphi(C) \cap A) \leq \#(\mathcal{C}) \leq \deg(C)d^{\dim(C)}.$$

Para cualquier conjunto constructible $C \subseteq \mathbb{A}^n(K)$, basta considerar una descomposición de grado LCI mínimo de C como en la Definición 13, i.e. una descomposición de C como unión finita de conjuntos localmente cerrados irreducibles que satisface el Lema 1.1.2:

$$C := (U_1 \cap V_1) \cup \dots \cup (U_s \cap V_s),$$

donde $U_i \cap V_i \neq \emptyset$, $U_i \subseteq \mathbb{A}^n(K)$ es abierto Zariski, $V_i \subseteq \mathbb{A}^n(K)$ es cerrado irreducible con respecto a la topología de Zariski y satisface:

$$(2.2.1) \quad \deg(C) = \sum_{i=1}^s \deg(V_i).$$

Entonces, tenemos que:

$$\overline{\varphi(C)}^z = \overline{\left(\bigcup_{i=1}^s \varphi(U_i \cap V_i) \right)}^z = \bigcup_{i=1}^s \overline{\varphi(U_i \cap V_i)}^z.$$

Como \deg es sub-aditivo cuando se aplica a variedades algebraicas, obtenemos:

$$\deg(\overline{\varphi(C)}^z) \leq \sum_{i=1}^s \deg(\overline{\varphi(U_i \cap V_i)}^z) \leq \sum_{i=1}^s \deg_{\text{lci}}(V_i \cap U_i) d^{\dim(V_i)} = \sum_{i=1}^s \deg(V_i) d^{\dim(V_i)}.$$

Como $\dim(C) := \max\{\dim(V_1), \dots, \dim(V_s)\}$, de la Identidad (2.2.1) concluimos:

$$\deg_z(\varphi(C)) \leq \left(\sum_{i=1}^s \deg(V_i) \right) d^{\dim(C)} = \deg_{\text{lci}}(C) d^{\dim(C)}.$$

- *Desigualdad* $\deg_z(\varphi(C)) \leq \deg_\pi(C) d^m \leq \deg_{\text{lci}}(C) d^m$: En primer lugar, consideramos el grafo de la función polinomial φ :

$$\begin{aligned} Gr(\varphi) &:= \{(x, y) \in C \times \mathbb{A}^m(K) : y_i - \varphi_i(x) = 0, i = 1, \dots, m\} = \\ &= (C \times \mathbb{A}^m(K)) \cap \left(\bigcap_{i=1}^m \{Y_i - \varphi_i(X) = 0\} \right). \end{aligned}$$

Entonces, sea Φ la siguiente función polinomial:

$$\begin{aligned} \Phi : C &\longrightarrow Gr(\varphi) \\ x &\longmapsto (x, \varphi(x)) \end{aligned}$$

Obviamente, tenemos que $\Phi(C) = Gr(\varphi)$ y, por lo tanto, $Gr(\varphi)$ es un conjunto constructible. Sea $\pi : \mathbb{A}^n(K) \times \mathbb{A}^m(K) \rightarrow \mathbb{A}^m(K)$ la proyección canónica que “olvida” las primeras n variables. Por consiguiente, tenemos que $\varphi(C) = \pi(Gr(\varphi))$ y $\deg_z(\varphi(C)) = \deg_z(\pi(Gr(\varphi)))$. Gracias a la Proposición 2.1.12, concluimos:

$$\deg_z(\pi(Gr(\varphi))) \leq \deg_\pi(\pi(Gr(\varphi))) \leq \deg_\pi(Gr(\varphi)).$$

Aplicando la Desigualdad de Bézout (para el π -grado, ver Teorema 2.1.14), deducimos que:

$$\begin{aligned} \deg_\pi(Gr(\varphi)) &= \deg_\pi \left((C \times \mathbb{A}^m(K)) \cap \left(\bigcap_{i=1}^m \{Y_i - \varphi_i(X) = 0\} \right) \right) \leq \\ &\leq \deg_\pi(C \times \mathbb{A}^m(K)) \prod_{i=1}^m \deg_\pi(\{Y_i - \varphi_i(X) = 0\}). \end{aligned}$$

Dado que $\deg_\pi(C \times \mathbb{A}^m(K)) \leq \deg_\pi(C) \deg_\pi(\mathbb{A}^m(K)) = \deg_\pi(C)$, y para todo i , $1 \leq i \leq m$, tenemos que $\deg_\pi(\{Y_i - \varphi_i(X) = 0\}) = \deg(Y_i - \varphi_i(X)) = \max\{\deg(\varphi_i), 1\} \leq d$, se obtiene:

$$\deg_\pi(C \times \mathbb{A}^m(K)) \prod_{i=1}^m \deg_\pi(\{Y_i - \varphi_i(X) = 0\}) \leq \deg_\pi(C) d^m.$$

Finalmente, aplicando la Proposición 2.1.7, concluimos:

$$\deg_z(\varphi(C)) \leq \deg_\pi(C) d^m \leq \deg_{\text{lci}}(C) d^m.$$

En cuanto al contra-ejemplo, basta tomar el que se empleó en el Ítem *iii*) de la Proposición 2.1.11.

Las dos últimas afirmaciones son evidentes por ser $\varphi(C)$ un conjunto localmente cerrado o presentable equidimensionalmente. \square

2.2.2. Demostración del Ítem i del Teorema 2.2.1. Procedemos por inducción. Comenzamos considerando el caso $s = 2$. En este caso, aplicando la Desigualdad de Bézout para conjuntos constructibles (cf. Teorema 2.1.14), obtenemos:

$$\deg_{\text{lci}}(C_1 \cap C_2) \leq \deg_{\text{lci}}(C_1) \deg_{\text{lci}}(C_2) \leq \binom{r+1}{r} \deg_{\text{lci}}(C_1) \deg_{\text{lci}}(C_2).$$

A continuación, supongamos que $s > 2$ y que $C_1 = W$ es un subconjunto localmente cerrado irreducible. Entonces, tenemos que:

$$W \cap C_2 \cap \left(\bigcap_{i=3}^s C_i \right) = W \cap \left(\bigcap_{i=2}^s C_i \right).$$

Distinguimos los dos casos siguientes:

- *Caso 1:* $\dim(W \cap C_2) < \dim(W)$. Aplicando la hipótesis de inducción obtenemos:

$$\deg_{\text{lci}} \left(W \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \binom{\dim(W \cap C_2) + s - 2}{\dim(W \cap C_2)} \deg_{\text{lci}}(W \cap C_2) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^{\dim(W \cap C_2)}.$$

Nótese que si $a \leq b$ son dos enteros positivos, se cumple:

$$\binom{a + (s-2)}{a} \leq \binom{b + (s-2)}{b} \leq \binom{b + (s-1)}{b}.$$

Como $\dim(W \cap C_2) \leq \dim(W) - 1 \leq \dim(W)$, tenemos que:

$$\deg_{\text{lci}} \left(W \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \binom{\dim(W) + s - 2}{\dim(W)} \deg_{\text{lci}}(W \cap C_2) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^{\dim(W)-1}.$$

Además, gracias a la Desigualdad de Bézout para conjuntos constructibles (cf. Teorema 2.1.14), obtenemos:

$$\deg_{\text{lci}} \left(W \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \binom{\dim(W) + s - 2}{\dim(W)} \deg(W) \deg_{\text{lci}}(C_2) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^{\dim(W)-1}.$$

Finalmente, agrupando y acotando algunas cantidades, obtenemos la desigualdad buscada:

$$\deg_{\text{lci}} \left(W \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \binom{\dim(W) + s - 1}{\dim(W)} \deg(W) (\max\{\deg_{\text{lci}}(C_j) : 2 \leq j \leq s\})^{\dim(W)}.$$

- *Caso 2:* $\dim(W \cap C_2) = \dim(W)$. Consideramos una descomposición de C_2 como unión finita de subconjuntos localmente cerrados irreducibles:

$$C_2 = W_1 \cup W_2 \cup \cdots \cup W_m,$$

tal que el grado LCI de C_2 verifica la siguiente igualdad:

$$\deg_{\text{lci}}(C_2) = \sum_{i=1}^m \deg(W_i).$$

Salvo reordenación de los W_i 's, dado que $\dim(W \cap C_2) = \dim(W)$, existe un entero k , $1 \leq k \leq m$, que verifica lo siguiente:

- Para todo i , $1 \leq i \leq k$, $\dim(W \cap W_i) = \dim(W)$.
- Para todo i , $k+1 \leq i \leq m$, $\dim(W \cap W_i) < \dim(W)$.

A continuación, vamos a demostrar que el siguiente conjunto es un subconjunto abierto Zariski en W y, por lo tanto, un conjunto localmente cerrado irreducible:

$$W \cap \left(\bigcup_{i=1}^k W_i \right).$$

Comencemos recordando que tanto W como W_i son localmente cerrados irreducibles. Por lo tanto, existen subconjuntos abiertos Zariski $U_1, U_2 \subseteq \mathbb{A}^n(K)$ y variedades algebraicas irreducibles $V, V_i \subseteq \mathbb{A}^n(K)$ tales que:

$$W = U_1 \cap V, \quad W_i = U_2 \cap V_i.$$

Como $\dim(W) = \dim(W \cap W_i)$, entonces $W \cap W_i \neq \emptyset$ y, por consiguiente,

$$W \cap W_i = U_1 \cap U_2 \cap V \cap V_i.$$

En particular, $U_1 \cap U_2 \cap V$ es un subconjunto abierto Zariski no vacío en V de dimensión igual a $\dim(V)$. Además, la dimensión de la intersección $U_1 \cap U_2 \cap V \cap V_i$ es también $\dim(V)$. Por lo tanto,

$$\dim(W \cap W_i) = \dim(U_1 \cap U_2 \cap V \cap V_i) = \dim(U_1 \cap U_2 \cap V) = \dim(V).$$

Entonces, la clausura Zariski de $W \cap W_i$ es una subvariedad cerrada de V de la misma dimensión que V . Dado que V es irreducible, se verifica que:

$$\overline{W \cap W_i}^z = V.$$

Por otro lado, es evidente que $\overline{W \cap W_i}^z \subseteq V_i$ y, por lo tanto, concluimos que $V \subseteq V_i$. Por consiguiente, tenemos que:

$$W \cap W_i = U_i \cap V,$$

para algún subconjunto abierto $U_i \subseteq \mathbb{A}^n(K)$. En conclusión, obtenemos:

$$W \cap \left(\bigcup_{i=1}^k W_i \right) = \left(\bigcup_{i=1}^k U_i \right) \cap V,$$

y se trata, como queríamos probar, de un subconjunto localmente cerrado irreducible.

Definimos ahora los dos siguientes subconjuntos constructibles:

$$C'_2 := \bigcup_{i=1}^k W_i, \quad \widetilde{C}_2 := \bigcup_{j=k+1}^s W_j.$$

Como $C'_2 \cap W$ es un subconjunto abierto Zariski de la clausura Zariski de W , concluimos que:

$$\deg(W \cap C'_2) = \deg(W).$$

Por otro lado, tenemos que:

$$\deg_{\text{lci}}(\widetilde{C}_2) = \deg_{\text{lci}}(C_2) - \sum_{i=1}^k \deg(W_i) \leq \deg_{\text{lci}}(C_2).$$

Por lo tanto, aplicando la Desigualdad de Bézout para conjuntos constructibles (cf. Teorema 2.1.14), obtenemos:

$$\deg_{\text{lci}}(W \cap \widetilde{C}_2) \leq \deg(W) (\deg_{\text{lci}}(C_2) - \sum_{i=1}^k \deg(W_i)) \leq \deg(W) \deg_{\text{lci}}(C_2).$$

A continuación, nótese que:

$$W \cap \left(\bigcap_{i=2}^s C_i \right) = \left((W \cap C'_2) \cap \left(\bigcap_{i=3}^s C_i \right) \right) \cup \left((W \cap \widetilde{C}_2) \cap \left(\bigcap_{i=3}^s C_i \right) \right).$$

Entonces, como \deg_{lci} es sub-aditivo para conjuntos constructibles (cf. Proposición 2.1.2), concluimos:

$$\deg_{\text{lci}} \left(W \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq I_1 + I_2,$$

donde

$$I_1 := \deg_{\text{lci}} \left((W \cap C'_2) \cap \left(\bigcap_{i=3}^s C_i \right) \right),$$

y

$$I_2 := \deg_{\text{lci}} \left((W \cap \widetilde{C}_2) \cap \left(\bigcap_{i=3}^s C_i \right) \right).$$

A continuación, acotamos ambas cantidades por separado:

- Tomando $r := \dim(W) = \dim(W \cap C'_2)$ y, sabiendo que $\deg(W) = \deg(W \cap C'_2)$, la aplicación de la hipótesis de inducción nos proporciona el siguiente resultado:

$$I_1 \leq \binom{r+s-2}{r} \deg(W) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^r.$$

- Por otro lado, consideramos $t := \dim(W \cap \widetilde{C}_2) \leq r - 1$. Sabiendo que $\deg_{\text{lci}}(W \cap \widetilde{C}_2) \leq \deg(W) \deg_{\text{lci}}(C_2)$, la aplicación de la hipótesis de inducción nos proporciona el siguiente resultado:

$$I_2 \leq \binom{t+s-2}{t} \deg(W) \deg_{\text{lci}}(C_2) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^t.$$

Como $t \leq r - 1$, también tenemos que:

$$\binom{t+s-2}{t} \leq \binom{r-1+s-2}{r-1} \leq \binom{r-1+s-1}{r-1},$$

y, por lo tanto, concluimos:

$$I_2 \leq \binom{r+s-2}{r-1} \deg(W) \deg_{\text{lci}}(C_2) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^{r-1}.$$

Combinando las dos desigualdades anteriores en una sola, tenemos que:

$$\begin{aligned} I_1 + I_2 &\leq \deg(W) (\max\{\deg(C_j) : 3 \leq j \leq s\})^{r-1} \\ &\quad \left(\binom{r+s-2}{r} \max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\} + \binom{r+s-2}{r-1} \deg_{\text{lci}}(C_2) \right). \end{aligned}$$

Por consiguiente,

$$\begin{aligned} I_1 + I_2 &\leq \deg(W) (\max\{\deg_{\text{lci}}(C_j) : 3 \leq j \leq s\})^{r-1} \\ &\quad \max\{\deg_{\text{lci}}(C_i) : 2 \leq i \leq s\} \left(\binom{r+s-2}{r} + \binom{r+s-2}{r-1} \right). \end{aligned}$$

Usando la igualdad del Triángulo de Pascal obtenemos:

$$I_1 + I_2 \leq \deg(W) (\max\{\deg_{\text{lci}}(C_j) : 2 \leq j \leq s\})^r \binom{r+s-1}{r},$$

que es la desigualdad buscada.

Para obtener la desigualdad general, supongamos que C_1 admite una descomposición en subconjuntos localmente cerrados irreducibles:

$$C_1 := V_1 \cup V_2 \cup \cdots \cup V_t,$$

tales que:

$$(2.2.2) \quad \deg_{\text{lci}}(C_1) = \sum_{j=1}^t \deg(V_j).$$

Dado que \deg_{lci} es sub-aditivo (cf. Proposición 2.1.2), obtenemos:

$$\deg_{\text{lci}} \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \sum_{j=1}^t \deg_{\text{lci}} \left(V_j \cap \left(\bigcap_{i=2}^s C_i \right) \right).$$

Aplicando lo discutido previamente y tomando $r_j = \dim(V_j) \leq \dim(V)$, concluimos:

$$\deg_{\text{lci}} \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \sum_{j=1}^t \binom{r_j+s-1}{r_j} \deg(V_j) (\max\{\deg_{\text{lci}}(C_i) : 2 \leq i \leq s\})^{r_j}.$$

Como $r_j \leq r$ para cada j , $1 \leq j \leq t$, tenemos que:

$$\binom{r_j+s-1}{r_j} \leq \binom{r+s-1}{r},$$

y, por lo tanto, deducimos que:

$$\deg_{\text{lci}} \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \binom{r+s-1}{r} \left(\sum_{j=1}^t \deg(V_j) \right) (\max\{\deg_{\text{lci}}(C_i) : 2 \leq i \leq s\})^r.$$

Usando la Identidad (2.2.2) obtenemos el resultado buscado.

2.2.3. Demostración del Ítem *ii*) del Teorema 2.2.1. En esta subsección adaptamos algunas de las estrategias de demostración de la Sección 3 de [He, 83] para obtener una demostración de una cota superior para el grado de la intersección de varios conjuntos constructibles. Antes de comenzar con esta demostración, necesitamos discutir algunos resultados preliminares.

2.2.3.1. *Resultados preliminares.*

DEFINICIÓN 15. Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y $V \subseteq \mathbb{A}^n(K)$ una variedad irreducible. Decimos que V es una componente irreducible de C con respecto al grado LCI de C si existe alguna descomposición de grado LCI mínimo de C en conjuntos localmente cerrados irreducibles:

$$C := W_1 \cup \cdots \cup W_s,$$

de modo que existe i con $V = \overline{W_i}^z$.

Nótese que si C es localmente cerrado o presentable equidimensionalmente, entonces la clase $\mathcal{D}(C)$ de todas las componentes irreducibles de C con respecto al grado LCI de C es un conjunto finito y está completamente determinado por C . Esto no es cierto en el caso general, como se mostró en la Observación 2.1.4.

LEMA 2.2.5. Sea $D \subseteq \mathbb{A}^n(K)$ un subconjunto localmente cerrado. Sean $\mathcal{V}_1, \dots, \mathcal{V}_s$ varias familias finitas de conjuntos localmente cerrados. Sea \mathcal{W} la clase de todos los subconjuntos localmente cerrados de $\mathbb{A}^n(K)$ que pueden definirse como intersecciones con D de algunos conjuntos localmente cerrados escogidos de acuerdo con la lista $\mathcal{V}_1, \dots, \mathcal{V}_s$. Es decir, definimos \mathcal{W} como la clase de conjuntos localmente cerrados dada por la siguiente identidad:

$$\mathcal{W} := \left\{ D \cap \left(\bigcap_{i \in S} V_i \right) : S \subseteq \{1, \dots, s\}, V_i \in \mathcal{V}_i \right\}.$$

Sea \mathcal{C} el conjunto de variedades algebraicas definidas por la siguiente identidad:

$$\mathcal{C} := \{V : V \text{ es una variedad irreducible y } \exists W \in \mathcal{W} \text{ tal que un subconjunto abierto no vacío de } V \text{ es una componente de } W\}.$$

Entonces, tenemos que:

$$(2.2.3) \quad \sum_{V \in \mathcal{C}} \deg(V) \leq \deg(D) \left(1 + \sum_{V \in \bigcup_{i=1}^s \mathcal{V}_i} \deg(V) \right)^{\dim(D)} \leq \deg(D) \left(1 + \sum_{i=1}^s \sum_{V \in \mathcal{V}_i} \deg(V) \right)^{\dim(D)}.$$

Además, existe una biyección que preserva el grado entre \mathcal{C} y el siguiente conjunto:

$$(2.2.4) \quad \mathcal{D} := \{C : \exists W \in \mathcal{W}, C \text{ es una componente irreducible de } \overline{W}^z\}.$$

y se verifica la misma cota superior:

$$(2.2.5) \quad \sum_{C \in \mathcal{D}} \deg(C) \leq \deg(D) \left(1 + \sum_{V \in \bigcup_{i=1}^s \mathcal{V}_i} \deg(V) \right)^{\dim(D)} \leq \deg(D) \left(1 + \sum_{i=1}^s \sum_{V \in \mathcal{V}_i} \deg(V) \right)^{\dim(D)}.$$

DEMOSTRACIÓN. En primer lugar, dado que el grado es una función sub-aditiva (cf. Proposición 2.1.2), bastará con probar el lema para D localmente cerrado irreducible. A partir de ahora, asumimos este hecho.

En segundo lugar, nótese que todos los elementos de \mathcal{W} son conjuntos localmente cerrados. Por lo tanto, para cada $W \in \mathcal{W}$, sus componentes localmente cerradas irreducibles están en biyección con las componentes irreducibles de \overline{W}^z . Esta biyección se produce por el hecho de que las componentes localmente cerradas irreducibles son simplemente subconjuntos abiertos Zariski en las componentes irreducibles de \overline{W}^z y se preserva el grado. Recíprocamente, las componentes irreducibles de \overline{W}^z son simplemente las clausuras Zariski de las componentes localmente cerradas irreducibles de W .

A continuación, pasamos a demostrar la cota superior de la Identidad (2.2.5). Sea $d = \dim(D)$ la dimensión de D . Nótese que cada $W \in \mathcal{W}$ y cada $C \in \mathcal{D}$ tiene dimensión menor o igual que d . Para cada k , $0 \leq k \leq n$, definimos $\mathcal{D}(k)$ como el conjunto de los elementos en \mathcal{D} de dimensión k , i.e.

$$\mathcal{D}(k) := \{C \in \mathcal{D} : \dim(C) = k\},$$

donde $\mathcal{D}(r) = \emptyset$ para $d+1 \leq r \leq n$. Como D es localmente cerrado irreducible, tenemos que $\mathcal{D}(d) = \{\overline{D}^z\}$.

Para $k < d$, tenemos la siguiente afirmación:

AFIRMACIÓN 2.2.5.1. Para cada $C \in \mathcal{D}(k)$ existen $C^* \in \mathcal{D}$ y $W \in \mathcal{V}_1 \cup \dots \cup \mathcal{V}_s$ tales que se verifican las siguientes propiedades:

- i) La dimensión de C^* es mayor que $k+1$ (i.e. $C^* \in \bigcup_{r=k+1}^d \mathcal{D}(r)$).
- ii) La variedad irreducible C es una componente irreducible de la clausura Zariski de la intersección $C^* \cap W$.

Demostración de la Afirmación. Sea $C \in \mathcal{D}(k)$ una componente irreducible y sea $S \subseteq \{1, \dots, s\}$ un subconjunto de cardinal mínimo de modo que existe un subconjunto abierto Zariski $U \cap C$ de C que es una componente localmente cerrada irreducible de alguna intersección:

$$D \cap \left(\bigcap_{i \in S} V_i \right),$$

donde $V_i \in \mathcal{V}_i$. Como $k < d$, S no puede ser el conjunto vacío. Definimos $S' := S \setminus \{j\}$ para algún $j \in S$. Entonces, $U \cap C$ debe ser una componente localmente cerrada irreducible de

$$(D \cap W') \cap V_j,$$

donde $W' := \left(\bigcap_{i \in S'} V_i \right)$.

Por lo tanto, existe una componente localmente cerrada irreducible C^* de W' de modo que algún subconjunto abierto Zariski de C es una componente localmente cerrada irreducible de $C^* \cap V_j$. Además, la dimensión de C^* debe ser mayor que $k+1$ ya que, en otro caso, C^* y C coincidirían en un abierto Zariski y, por lo tanto, S no tendría cardinalidad mínima. Esto prueba la afirmación. ■

La afirmación anterior nos permite definir la siguiente aplicación:

$$\begin{aligned} \Phi_k : \mathcal{D}(k) &\longrightarrow \left(\bigcup_{r=k+1}^d \mathcal{D}(r) \right) \times \left(\bigcup_{i=1}^s \mathcal{V}_i \right), \\ C &\longmapsto (C^*, V) \end{aligned}$$

dada por la siguiente regla: a cada $C \in \mathcal{D}(k)$ le asociamos un par (C^*, V) de modo que un subconjunto abierto Zariski no vacío de C es una componente localmente irreducible de $C^* \cap V$.

Definamos:

$$D(k) := \sum_{r=k}^d \sum_{C \in \mathcal{D}(r)} \deg(C),$$

y, a continuación, vamos a demostrar por inducción en $m = d - k$ la siguiente desigualdad:

$$(2.2.6) \quad D(k) = D(d - m) \leq \deg(D) \left(1 + \sum_{V \in \bigcup_{i=1}^s \mathcal{V}_i} \deg(V) \right)^m.$$

El caso $m = 0$ es obviamente cierto ya que $\mathcal{D}(d) = \{\overline{D^z}\}$ y D es localmente cerrado. Tenemos que $D(d) = \deg(D)$.

Supongamos ahora que $m \geq 1$. Tenemos la siguiente desigualdad:

$$\sum_{C \in \mathcal{D}(k)} \deg(C) \leq \sum_{(C^*, V) \in \left(\bigcup_{r=k+1}^d \mathcal{D}(r) \right) \times \left(\bigcup_{i=1}^s \mathcal{V}_i \right)} \sum_{C \in \Phi_k^{-1}(C^*, V)} \deg(C).$$

Dado que todos los subconjuntos involucrados son localmente cerrados, para cada $C^* \in \bigcup_{r=k+1}^d \mathcal{D}(r)$ y para cada $V \in \bigcup_{i=1}^s \mathcal{V}_i$, se obtiene del Teorema 1.2.8 que:

$$\sum_{C \in \Phi_k^{-1}(C^*, V)} \deg(C) \leq \deg(C^*) \deg(V).$$

Por lo tanto, tenemos que:

$$\sum_{C \in \mathcal{D}(k)} \deg(C) \leq \sum_{(C^*, V) \in \left(\bigcup_{r=k+1}^d \mathcal{D}(r) \right) \times \left(\bigcup_{i=1}^s \mathcal{V}_i \right)} \deg(C^*) \deg(V).$$

Reordenando las sumas, obtenemos:

$$\sum_{C \in \mathcal{D}(k)} \deg(C) \leq \left(\sum_{C^* \in \bigcup_{r=k+1}^d \mathcal{D}(r)} \deg(C^*) \right) \left(\sum_{V \in \bigcup_{i=1}^s \mathcal{V}_i} \deg(V) \right) = D(k+1)R,$$

donde

$$R = \sum_{V \in \bigcup_{i=1}^s \mathcal{V}_i} \deg(V).$$

Como

$$D(k) = D(k+1) + \sum_{C \in \mathcal{D}(k)} \deg(C),$$

llegamos a que:

$$D(k) = D(d-m) \leq D(d-(m-1))R + D(d-(m-1)) = D(k+1)(R+1).$$

Aplicando la hipótesis de inducción concluimos la Igualdad (2.2.6) y, por lo tanto, la demostración del lema. \square

2.2.3.2. *Demostración de la segunda cota superior del grado en el Teorema 2.2.1.* Supongamos que C_1 es un conjunto constructible. Si el resultado es cierto en el caso de conjuntos localmente cerrados irreducibles, podemos considerar una descomposición de C_1 como unión finita de conjuntos localmente cerrados irreducibles de acuerdo con el Lema 1.1.2:

$$C_1 := W_1 \cup \dots \cup W_{s(1)},$$

que minimice el grado LCI , es decir:

$$\deg_{\text{lci}}(C_1) := \sum_{j=1}^{s(1)} \deg(W_j).$$

Como el grado es sub-aditivo (ver Proposición 2.1.2) y

$$\dim(C_1) = \max\{\dim(W_j) : 1 \leq j \leq s(1)\},$$

concluiríamos la segunda cota superior a partir del caso de conjuntos localmente cerrados irreducibles mediante la siguiente cadena de desigualdades:

$$\begin{aligned} \deg_{\text{lci}} \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right) &\leq \sum_{j=1}^{s(1)} \deg_{\text{lci}} \left(W_j \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \sum_{j=1}^{s(1)} \deg(W_j) \left(1 + \sum_{i=1}^s \deg_{\text{lci}}(C_i) \right)^{\dim(W_j)} \leq \\ &\leq \left(\sum_{j=1}^{s(1)} \deg(W_j) \right) \left(1 + \sum_{i=1}^s \deg_{\text{lci}}(C_i) \right)^{\dim(C_1)} = \deg_{\text{lci}}(C_1) \left(1 + \sum_{i=1}^s \deg_{\text{lci}}(C_i) \right)^{\dim(C_1)}, \end{aligned}$$

lo que prueba la proposición para cualquier conjunto constructible C_1 .

Denotamos por $d := \dim(C_1)$ a la dimensión de C_1 y supongamos que C_1 es localmente cerrado irreducible. Consideramos también para cada conjunto constructible C_i , $2 \leq i \leq s$, una descomposición de acuerdo al Lema 1.1.2:

$$C_i := W_{i,1} \cup \dots \cup W_{i,s(i)},$$

donde cada $W_{i,j}$ es un subconjunto abierto Zariski en una variedad irreducible que denotamos por $V_{i,j}$. Supongamos, además, que estas descomposiciones minimizan el grado LCI de los C_i , i.e.

$$\deg_{\text{lci}}(C_i) := \sum_{j=1}^{s(i)} \deg(W_{i,j}) = \sum_{j=1}^{s(i)} \deg(V_{i,j}), \quad 2 \leq i \leq s.$$

Introducimos ahora las clases de conjuntos localmente cerrados irreducibles $\mathcal{V}_1, \dots, \mathcal{V}_s$ dadas por las siguientes identidades:

$$\mathcal{V}_i := \{W_{i,1}, \dots, W_{i,s(i)}\}, \quad 2 \leq i \leq s.$$

Definimos la clase \mathcal{V} de conjuntos localmente cerrados:

$$\mathcal{V} := \left\{ W : \forall i, 2 \leq i \leq s, \exists W_i \in \mathcal{V}_i, W = C_1 \cap \left(\bigcap_{i=2}^s W_i \right) \right\}.$$

Y, finalmente, definimos la clase \mathcal{C} de las variedades algebraicas irreducibles dada por la siguiente igualdad:

$$\mathcal{C} := \{C : C \text{ es irreducible y } \exists W \in \mathcal{V} \text{ tal que un subconjunto abierto no vacío de } C \text{ es una componente de } W\}.$$

Para cada $C \in \mathcal{C}$, sea $U_C \subseteq \mathbb{A}^n(K)$ un subconjunto abierto Zariski, máximo con la propiedad:

$$\emptyset \neq U_C \cap C \subseteq \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right).$$

AFIRMACIÓN 2.2.5.2. *Con las notaciones anteriores, tenemos que:*

$$C_1 \cap \left(\bigcap_{i=2}^s C_i \right) = \bigcup_{C \in \mathcal{C}} U_C \cap C.$$

Demostración de la Afirmación. Para cada $C \in \mathcal{C}$, existe un subconjunto abierto Zariski $U_0 \subseteq \mathbb{A}^n(K)$ tal que $U_0 \cap C \neq \emptyset$ y $U_0 \cap C$ es una componente localmente cerrada irreducible de un subconjunto localmente cerrado de la siguiente forma:

$$C_1 \cap (W_2 \cap \cdots \cap W_s),$$

donde $W_i \in \mathcal{V}_i$. Por lo tanto, $U_0 \cap C \subseteq (C_1 \cap (\bigcap_{i=2}^s C_i))$ y, por la definición de U_C , concluimos:

$$U_C \cap C \subseteq \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right).$$

Por otro lado, fijemos un punto $x \in (C_1 \cap (\bigcap_{i=2}^s C_i))$. Dado que se verifica la igualdad:

$$C_1 \cap \left(\bigcap_{i=2}^s C_i \right) = \bigcup_{(W_2, \dots, W_s) \in \mathcal{V}_2 \times \cdots \times \mathcal{V}_s} C_1 \cap \left(\bigcap_{i=2}^s W_i \right),$$

concluimos que existen $W_2 \in \mathcal{V}_2, \dots, W_s \in \mathcal{V}_s$ tales que $x \in (C_1 \cap (\bigcap_{i=2}^s W_i))$. En particular, x debe pertenecer a alguna componente localmente cerrada irreducible de $(C_1 \cap (\bigcap_{i=1}^s W_i))$ y, por lo tanto, debe existir algún $C \in \mathcal{C}$ tal que $x \in U_C \cap C$. ■

Por la Proposición 2.1.2 (el grado *LCI* es sub-aditivo), concluimos:

$$(2.2.7) \quad \deg_{\text{Lci}} \left(C_1 \cap \left(\bigcap_{i=2}^s C_i \right) \right) \leq \sum_{C \in \mathcal{C}} \deg(C).$$

Consideramos ahora la familia \mathcal{W} de conjuntos localmente cerrados (como en el Lema 2.2.5 anterior) dada por la siguiente identidad:

$$\mathcal{W} := \left\{ W : \exists S \subseteq \{2, \dots, s\}, \forall i \in S, \exists W_i \in \mathcal{V}_i, W = C_1 \cap \left(\bigcap_{i \in S} W_i \right) \right\}.$$

Definimos también una nueva familia de variedades irreducibles:

$$\mathcal{D} := \{C : C \text{ es irreducible y } \exists W \in \mathcal{W} \text{ tal que un subconjunto abierto no vacío de } C \text{ es una componente de } W\}.$$

Obviamente $\mathcal{C} \subseteq \mathcal{D}$ y, usando el Lema 2.2.5, concluimos:

$$\sum_{C \in \mathcal{C}} \deg(C) \leq \sum_{C \in \mathcal{D}} \deg(C) \leq \deg(C_1) \left(1 + \sum_{i=2}^s \sum_{V \in \mathcal{V}_i} \deg(V) \right)^{\dim(C_1)}.$$

Como tenemos que:

$$\deg_{\text{Lci}}(C_i) := \sum_{V \in \mathcal{V}_i} \deg(V),$$

de acuerdo con la Desigualdad (2.2.7), concluimos finalmente la segunda cota superior del Teorema 2.2.1:

$$\deg_{\text{Lci}} \left(\bigcap_{i=1}^s C_i \right) \leq \deg(C_1) \left(1 + \sum_{i=2}^s \deg_{\text{Lci}}(C_i) \right)^{\dim(C_1)}.$$

2.2.4. Dos aplicaciones inmediatas de las cotas superiores previas.

EJEMPLO 2.2.6 (Contando puntos \mathbb{F}_q -racionales en conjuntos constructibles). Sea \mathbb{F}_q un cuerpo finito y $\overline{\mathbb{F}_q}$ su clausura algebraica. Para cada subconjunto constructible $C \subseteq \mathbb{A}^n(\overline{\mathbb{F}_q})$, denotamos por $C_{\mathbb{F}_q} := C \cap \mathbb{F}_q^n$ al conjunto de sus puntos \mathbb{F}_q -racionales. El siguiente resultado generaliza a conjuntos constructibles un resultado clásico para hipersuperficies de \emptyset . Ore (cf. [Ore, 22]).

COROLARIO 2.2.7. *Con las notaciones precedentes, para cada conjunto constructible $C \subseteq \mathbb{A}^n(\overline{\mathbb{F}}_q)$, el número de puntos \mathbb{F}_q -racionales verifica:*

$$\#(C_{\mathbb{F}_q}) = \#(C \cap \mathbb{F}_q^n) \leq \deg_{\text{lci}}(C)q^{\dim(C)}.$$

En particular, para cada polinomio no nulo $f \in \overline{\mathbb{F}}_q[X_1, \dots, X_n]$, cuyo grado satisface $\deg(f) \leq q - 1$, existe algún punto $x \in \mathbb{F}_q^n$ tal que $f(x) \neq 0$. Además, el número de valores no nulos de f en \mathbb{F}_q^n satisface:

$$\#\{x \in \mathbb{F}_q^n : f(x) \neq 0\} \geq (q - \deg(f))q^{n-1}.$$

DEMOSTRACIÓN. En primer lugar, consideramos para cada j , $1 \leq j \leq n$, las siguientes variedades algebraicas:

$$W_j := \overline{\mathbb{F}}_q^{j-1} \times \{x_j \in \overline{\mathbb{F}}_q : x_j^q - x_j = 0\} \times \overline{\mathbb{F}}_q^{n-j} \subseteq \mathbb{A}^n(\overline{\mathbb{F}}_q).$$

Cada $W_j \subseteq \mathbb{A}^n(\overline{\mathbb{F}}_q)$ es una hipersuperficie de grado q y obviamente tenemos que:

$$\mathbb{F}_q^n = \bigcap_{j=1}^n W_j.$$

Como W_1, \dots, W_n son variedades algebraicas de grado a lo sumo q , por la Proposición 2.2.2, concluimos:

$$\deg_{\text{lci}}(C \cap \mathbb{F}_q^n) = \deg_{\text{lci}}\left(C \cap \left(\bigcap_{j=1}^n W_j\right)\right) \leq \deg_{\text{lci}}(C)q^{\dim(C)}.$$

Dado que $C \cap \mathbb{F}_q^n$ es una variedad cero-dimensional, tenemos que (cf. Proposición 1.2.3):

$$\#(C_{\mathbb{F}_q}) = \#(C \cap \mathbb{F}_q^n) = \deg_{\text{lci}}(C \cap \mathbb{F}_q^n).$$

Respecto a la segunda afirmación, como $V(f)$ es una hipersuperficie no vacía (de dimensión $n - 1$) en $\mathbb{A}^n(\overline{\mathbb{F}}_q)$, la Proposición 1.2.3 implica que el grado de $V(f)$ es a lo sumo $\deg(f)$. Por lo tanto, tenemos que:

$$\#(V(f) \cap \mathbb{F}_q^n) \leq \deg(V(f))q^{n-1} \leq \deg(f)q^{n-1} \leq (q - 1)q^{n-1}.$$

Finalmente, concluimos que:

$$\#\{x \in \mathbb{F}_q^n : f(x) \neq 0\} \geq (q - \deg(f))q^{n-1}.$$

□

Los resultados de Weil, Lang y Stepanov (cf. [We,49], [LW, 54] y [St, 69]) muestran que la cota del corolario previo es esencialmente óptima en el caso de que C sea una variedad algebraica.

EJEMPLO 2.2.8 (Variedades evasivas cero-dimensional de conjuntos constructibles).

COROLARIO 2.2.9. *Sea κ un cuerpo, $Q \subseteq \kappa$ un subconjunto finito y K la clausura algebraica de κ . Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Entonces, el número de puntos de intersección de C y Q^n satisface:*

$$\#(C \cap Q^n) \leq \deg_{\text{lci}}(C)\#(Q)^{\dim(C)}.$$

En otras palabras, la probabilidad de que un punto $x \in Q^n$ escogido al azar verifique $x \notin C$ es mayor que:

$$1 - \frac{\deg_{\text{lci}}(C)}{\#(Q)^{\text{codim}(C)}},$$

donde $\text{codim}(C) = n - \dim(C)$ es la co-dimensión de C en $\mathbb{A}^n(K)$.

DEMOSTRACIÓN. La demostración es similar a la del corolario precedente, reemplazando \mathbb{F}_q por Q , y sustituyendo la ecuación de grado q que define a \mathbb{F}_q en $\overline{\mathbb{F}}_q$ por un polinomio univariado de grado $\#(Q)$ cuyos ceros en K son exactamente los puntos del conjunto Q .

En lo que respecta a la cota inferior para la probabilidad, esta es consecuencia de la siguiente cadena de desigualdades:

$$\text{Prob}_{Q^n}[x \in Q^n : x \notin C] \geq 1 - \frac{\deg_{\text{lci}}(C)\#(Q)^{\dim(C)}}{\#(Q)^n} = 1 - \frac{\deg_{\text{lci}}(C)}{\#(Q)^{\text{codim}(C)}}.$$

□

Nótese que, en el caso $C = V(f)$, donde C es una hipersuperficie definida por un polinomio no nulo, este último corolario es la base del Lema de DeMillo–Lipton–Schwartz–Zippel (cf. [DML, 78], [Zp, 79] y [Sch, 80]).

2.3. Una observación sobre las cotas de las imágenes de conjuntos constructibles por aplicaciones lineales

Como ya observamos en las Proposiciones 2.1.11 y 2.1.12, una de las principales diferencias entre el “grado de proyección” \deg_π y el “grado como descomposición” \deg_{Ici} es que presentan un comportamiento distinto respecto a imágenes lineales de variedades algebraicas y conjuntos constructibles. Dada una aplicación lineal $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ y cualquier subconjunto constructible $C \subseteq \mathbb{A}^n(K)$, se verifican las siguientes desigualdades (ver Proposición 2.1.12):

$$\deg_z(\ell(C)) \leq \deg_\pi(\ell(C)) \leq \deg_\pi(C) \leq \deg_{\text{Ici}}(C).$$

Además, hemos visto un ejemplo en el que $\deg_{\text{Ici}}(\ell(C)) > \deg_{\text{Ici}}(C)$ en la demostración del Ítem *iv*) de la Proposición 2.1.11.

Para ser completos, en esta sección presentamos cotas superiores para el *defecto*:

$$\deg_{\text{Ici}}(\ell(C)) - \deg_z(\ell(C))$$

y, en términos de cotas extrínsecas, cotas superiores para el $\deg_{\text{Ici}}(\ell(C))$ basadas en el Nullstellensatz Efectivo de [Je, 05].

2.3.1. Una primera cota para el defecto. En primer lugar, veamos que el defecto puede controlarse por el grado del conjunto constructible de aquellos puntos cuyas fibras no tienen la dimensión genérica. Usaremos las mismas notaciones que en las secciones anteriores.

LEMA 2.3.1. *Sea $V \subseteq \mathbb{A}^n(K)$ una variedad algebraica irreducible y $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ una aplicación lineal. Sea $W := \overline{\ell(V)}^z$ la clausura Zariski de la imagen. Sea $U \subseteq \mathbb{A}^m(K)$ cualquier subconjunto abierto Zariski tal que $U \cap W \neq \emptyset$ y:*

$$(2.3.1) \quad \forall y \in W \cap U, \dim(\ell|_V^{-1}(\{y\})) = \dim(V) - \dim(W).$$

Por lo tanto, $U \cap W \subseteq \ell(V)$. Además, sea $U^c := \mathbb{A}^m(K) \setminus U$ la variedad algebraica de aquellos puntos de V para los cuales no se garantiza que las fibras de ℓ restringidas a V tengan la dimensión apropiada. Entonces, se cumple que:

$$\deg_{\text{Ici}}(\ell(V)) - \deg_z(\ell(V)) \leq \deg_{\text{Ici}}(\ell(V) \cap U^c),$$

donde $\dim(\ell(V) \cap U^c) \leq \dim(\ell(V)) - 1$. También tenemos que:

$$\deg_{\text{Ici}}(\ell(V)) - \deg_z(\ell(V)) \leq \deg_{\text{Ici}}(\ell(V \cap \ell^{-1}(U^c))).$$

DEMOSTRACIÓN. En primer lugar, tenemos que W es irreducible ya que V es irreducible. Por otro lado, la existencia del conjunto abierto Zariski U está garantizada por el Teorema de la Dimensión en la Fibra (cf. [Sha, 77], Teorema 7, p. 60).

Como $\dim(V) - \dim(W) \geq 0$, de la Identidad (2.3.1) se deduce que, para cada $y \in W \cap U$, se verifica que $\ell^{-1}(\{y\}) \neq \emptyset$. Por lo tanto, $\ell(V) \cap U = W \cap U$ es un subconjunto localmente cerrado. Entonces, tenemos que:

$$\ell(V) = (\ell(V) \cap U) \cup (\ell(V) \cap U^c) = (W \cap U) \cup (\ell(V) \cap U^c).$$

Por la sub-aditividad de \deg_{Ici} (cf. Proposición 2.1.2), obtenemos:

$$\deg_{\text{Ici}}(\ell(V)) \leq \deg_{\text{Ici}}(W \cap U) + \deg_{\text{Ici}}(\ell(V) \cap U^c).$$

Como $W \cap U$ es localmente cerrado y W es irreducible,

$$\deg_{\text{Ici}}(W \cap U) = \deg(W) = \deg_z(\ell(V)),$$

y, por lo tanto, tenemos que:

$$\deg_{\text{Ici}}(\ell(V)) - \deg_z(\ell(V)) \leq \deg_{\text{Ici}}(\ell(V) \cap U^c).$$

Dado que $W \cap U$ es denso Zariski en W , W es irreducible y $\ell(V) \cap U^c$ no interseca a U , deducimos que:

$$\overline{(\ell(V) \cap U^c)}^z \neq W,$$

y, por lo tanto, es una subvariedad cerrada propia de la variedad irreducible W . Por consiguiente,

$$\dim(\ell(V) \cap U^c) \leq \dim(\ell(V)) - 1.$$

Finalmente, vamos a demostrar que se cumple la siguiente igualdad:

$$\ell(V \cap \ell^{-1}(U^c)) = \ell(V) \cap U^c.$$

Si $x \in \ell(V \cap \ell^{-1}(U^c))$, es evidente que $x \in \ell(V) \cap U^c$. Por otro lado, si $x \in \ell(V) \cap U^c$, e $y \in V$ es tal que $\ell(y) = x$, obviamente tenemos que $y \in \ell^{-1}(U^c)$ y, por lo tanto, $x \in \ell(V \cap \ell^{-1}(U^c))$. \square

LEMA 2.3.2. Sea $V \subseteq \mathbb{A}^n(K)$ una variedad algebraica irreducible de dimensión r y grado D . Sea $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ una aplicación lineal. Supongamos que la restricción $\ell|_V : V \rightarrow \mathbb{A}^m(K)$ es dominante (y, por lo tanto, $m \leq r$). Sean $\{X_1, \dots, X_n\}$ las variables asociadas a $\mathbb{A}^n(K)$ y, del mismo modo, $\{Y_1, \dots, Y_m\}$ las variables asociadas a $\mathbb{A}^m(K)$. Entonces, existe un polinomio no nulo $q \in K[Y_1, \dots, Y_m]$ de modo que se verifican las siguientes propiedades:

i) El polinomio $p := q \circ \ell \in K[X_1, \dots, X_n]$ es también no nulo y sus respectivos grados satisfacen:

$$\deg(p) \leq \deg(q) \leq (n - r)(\deg(V) - 1).$$

ii) Para cada $y \in \mathbb{A}^m(K)$ tal que $q(y) \neq 0$, la dimensión en la fibra verifica:

$$\dim \ell|_V^{-1}(\{y\}) = \dim(V \cap \ell^{-1}(\{y\})) = r - m.$$

iii) O bien $V \cap V_{\mathbb{A}}(q) = \emptyset$, o la dimensión satisface $\dim(V \cap V_{\mathbb{A}}(q)) = \dim(V) - 1$.

iv) El defecto de $\ell(V)$ cumple:

$$\deg_{\text{Ici}}(\ell(V)) - \deg_z(\ell(V)) \leq \deg_{\text{Ici}}(\ell(V \cap V_{\mathbb{A}}(q))),$$

donde $\deg_z(\ell(V)) = 1$ ya que $\ell|_V$ es dominante sobre $\mathbb{A}^m(K)$.

DEMOSTRACIÓN. En primer lugar, dado que $\ell|_V$ es dominante, tenemos que $\ell(V)$ es denso en $\mathbb{A}^m(K)$ con respecto a la topología de Zariski, i.e. $\overline{\ell(V)}^z = \mathbb{A}^m(K)$. Como $\ell(V) \subseteq \ell(\mathbb{A}^n(K))$ y $\ell(\mathbb{A}^n(K))$ es un subespacio, tenemos que $\ell(\mathbb{A}^n(K)) = \mathbb{A}^m(K)$ y, por lo tanto, la aplicación $\ell : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ es suprayectiva.

Sean ℓ_1, \dots, ℓ_m las funciones coordenadas de ℓ . En particular, el anillo $K[\ell_1, \dots, \ell_m]$ es un anillo de polinomios en m variables con coeficientes en K . Salvo por un cambio lineal de coordenadas en $\mathbb{A}^n(K)$, supondremos que $\ell_1 = X_1, \dots, \ell_m = X_m$. Entonces, procedemos en nuestra demostración como si ℓ fuese la proyección canónica $\ell := \pi_m : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ en las primeras m coordenadas de un punto en $\mathbb{A}^n(K)$.

Sea $I(V) \in \text{Spec}(K[X_1, \dots, X_n])$ el ideal primo asociado a V . Como $\ell|_V$ es dominante, tenemos la siguiente extensión de anillos:

$$\ell^* : K[X_1, \dots, X_m] \hookrightarrow K[V] := K[X_1, \dots, X_n]/I(V).$$

Consideremos el sistema multiplicativo $S := K[X_1, \dots, X_m] \setminus \{0\}$ y sea $F := K(X_1, \dots, X_m)$ el cuerpo de cocientes de $K[X_1, \dots, X_m]$. Como

$$I(V) \cap K[X_1, \dots, X_m] = (0),$$

también tenemos la siguiente extensión de anillos:

$$S^{-1}\ell^* : F \hookrightarrow F[V] := S^{-1}K[V] = F[X_{m+1}, \dots, X_n]/\mathfrak{p},$$

donde $\mathfrak{p} = S^{-1}I(V)$.

Dado que $\ell|_V$ es dominante, el Teorema de la Dimensión en la Fibra implica que la dimensión genérica de una fibra $(\ell|_V)^{-1}(\{x\})$ es $r - m$. Por lo tanto, la dimensión de Krull satisface:

$$\dim(F[V]) = \dim(F[X_{m+1}, \dots, X_n]/\mathfrak{p}) = r - m.$$

Como K es un cuerpo infinito, también lo es F , y existen genéricamente matrices $A \in \mathcal{M}_{n-m}(K)$ tales que el siguiente cambio lineal de coordenadas:

$$\begin{pmatrix} Y_{m+1} \\ \vdots \\ Y_n \end{pmatrix} = A \begin{pmatrix} X_{m+1} \\ \vdots \\ X_n \end{pmatrix},$$

coloca a las variables $\{Y_{m+1}, \dots, Y_n\}$ en posición de Noether respecto a \mathfrak{p} , esto es, la siguiente extensión es una extensión entera de anillos:

$$F[Y_{m+1}, \dots, Y_n] \hookrightarrow F[V] := F[Y_{m+1}, \dots, Y_n]/\mathfrak{p},$$

donde también llamamos \mathfrak{p} a la transformación de \mathfrak{p} bajo dicho cambio lineal de coordenadas. Entonces, para cada j , $r + 1 \leq j \leq n$, existe un polinomio mónico $h_j \in F[Y_{m+1}, \dots, Y_n][T]$ de grado positivo respecto a la variable T (i.e. $\delta_j = \deg_T(h_j) \geq 1$), tal que:

$$h_j(Y_j) \in \mathfrak{p}, \quad r + 1 \leq j \leq n.$$

Como \mathfrak{p} es un ideal primo, podemos suponer que h_j es irreducible. Para cada j , $r + 1 \leq j \leq n$, existe un polinomio primitivo:

$$H_j \in K[X_1, \dots, X_m, Y_{m+1}, \dots, Y_n][T],$$

tal que h_j y H_j son asociados en el anillo $F[Y_{m+1}, \dots, Y_r][T]$. Es decir, para cada j , $r+1 \leq j \leq n$, existe un polinomio de la siguiente forma:

$$H_j := e_{\delta_j}^{(j)}(X_1, \dots, X_m)T^{\delta_j} + \sum_{k=0}^{\delta_j-1} a_k^{(j)}(X_1, \dots, X_m, Y_{m+1}, \dots, Y_r)T^k,$$

donde $e_{\delta_j}^{(j)} \in K[X_1, \dots, X_m] \setminus \{0\}$, $a_k^{(j)} \in K[X_1, \dots, X_m, Y_{m+1}, \dots, Y_r]$ y se cumplen las siguientes propiedades:

- Cada H_j es un polinomio primitivo.
- Para cada j , $r+1 \leq j \leq n$, el polinomio H_j es irreducible en $K[X_1, \dots, X_m, Y_{m+1}, \dots, Y_r][T]$ y se verifica la siguiente igualdad:

$$h_j := \left(e_{\delta_j}^{(j)}\right)^{-1} H_j.$$

- Para cada j , $r+1 \leq j \leq n$, $H_j(Y_j) \in \mathfrak{p}$, $e_{\delta_j}^{(j)} \notin \mathfrak{p}$.

La última propiedad es consecuencia del hecho de que $e_{\delta_j}^{(j)} \neq 0$ y \mathfrak{p} es un ideal primo. Definimos entonces el siguiente polinomio no nulo:

$$q := \prod_{j=r+1}^n e_{\delta_j}^{(j)} \in K[X_1, \dots, X_m] \setminus \{0\}.$$

Nótese que, como $I(V)$ es un ideal primo, entonces $q \notin I(V)$. A continuación, definimos los siguientes conjuntos localmente cerrados y anillos:

- i) El subconjunto abierto distinguido de $\mathbb{A}^m(K)$ dado por:

$$D(q) := \{x \in \mathbb{A}^m : q(x) \neq 0\}.$$

- ii) El subconjunto abierto Zariski no vacío V_q de V dado por la siguiente identidad:

$$V_q := \{\zeta := (x, y, u) \in \mathbb{A}^m(K) \times \mathbb{A}^{r-m}(K) \times \mathbb{A}^{n-r}(K) : \zeta \in V, q(x) \neq 0\}.$$

- iii) El anillo de funciones regulares definidas sobre $D(q)$, dado como la localización en el sistema multiplicativo definido por q :

$$K[D(q)] := K[X_1, \dots, X_m]_q.$$

- iv) El anillo de funciones regulares definidas sobre V_q , dado por:

$$K[V_q] := K[X_1, \dots, X_m]_q[Y_{m+1}, \dots, Y_n]/\mathfrak{q},$$

donde \mathfrak{q} es, al mismo tiempo, la extensión de $I(V)$ a

$$K[X_1, \dots, X_m]_q[Y_{m+1}, \dots, Y_n]$$

y la contracción de \mathfrak{p} al mismo anillo.

Como para cada j , $r+1 \leq j \leq n$, $h_j \in K[X_1, \dots, X_m]_q[Y_{m+1}, \dots, Y_r][T]$ y $h_j(Y_j) \in \mathfrak{p}$, la siguiente extensión es una extensión entera de anillos:

$$(2.3.2) \quad K[D(q)][Y_{m+1}, \dots, Y_r] \hookrightarrow K[V_q] := K[X_1, \dots, X_m]_q[Y_{m+1}, \dots, Y_n]/\mathfrak{q}.$$

De esta extensión entera, se deduce inmediatamente que:

$$(2.3.3) \quad \forall x \in D(q), \ell|_V^{-1}(\{x\}) \neq \emptyset.$$

De hecho, recordando que $\ell = \pi_m$ es una proyección, la extensión entera de anillos de la Identidad (2.3.2) implica que la siguiente aplicación es suprayectiva:

$$\begin{aligned} \pi : \quad V_q &\longrightarrow D(q) \times \mathbb{A}^{r-m}(K) \\ \zeta = (x, y, u) &\longmapsto (x, y). \end{aligned}$$

Consideremos ahora las siguientes proyecciones:

$$\begin{aligned} \Pi_j : \quad V_q &\longrightarrow \mathbb{A}^{r+1}(K) \\ \zeta := (x, y, u) &\longmapsto (x, y, u_j), \end{aligned}$$

donde $x \in D(q)$, $y \in \mathbb{A}^{r-m}(K)$ y $u := (u_{r+1}, \dots, u_n) \in \mathbb{A}^{n-r}(K)$. A continuación, demostramos la siguiente afirmación:

AFIRMACIÓN 2.3.2.1. *Con las notaciones y las hipótesis anteriores, $\Pi_j(V_q)$ es una hipersuperficie en $D(q) \times \mathbb{A}^{r-m+1}(K)$ de grado a lo sumo $\deg(V)$ y su polinomio mínimo es el polinomio H_j introducido previamente, cuyo grado total verifica $\deg(H_j) \leq \deg(V)$.*

Demostración de la Afirmación. Es claro que $\Pi_j(V_q) \subseteq D(q) \times \mathbb{A}^{r-m+1}(K)$. Además, si $\lambda : \mathbb{A}^{r+1}(K) \rightarrow \mathbb{A}^r(K)$ es la proyección que “olvida” la última coordenada, tenemos que:

$$\lambda \circ \Pi_j = \pi,$$

donde π es la aplicación suprayectiva que hemos definido previamente. Por consiguiente, concluimos que:

$$\dim(\Pi_j(V_q)) \geq \dim(\lambda(\Pi_j(V_q))) \geq \dim(\pi(V_q)) = \dim(D(q) \times \mathbb{A}^{r-m}(K)) = r.$$

Por otro lado, el polinomio $H_j \in K[X_1, \dots, X_m, Y_{m+1}, \dots, Y_r][T]$ es un polinomio no nulo y tenemos que:

$$\Pi_j(V_q) \subseteq \{(x, y, u_j) \in \mathbb{A}^{r+1}(K) : q(x) \neq 0, H_j(x, y, u_j) = 0\}.$$

Por el Hauptidealsatz de Krull, la co-dimensión de $\Pi_j(V_q)$ es al menos 1 y hemos demostrado la siguiente igualdad:

$$\dim(\Pi_j(V_q)) = r.$$

Por lo tanto, $\overline{\Pi_j(V_q)}^z$ es una hipersuperficie en $\mathbb{A}^{r+1}(K)$. Además, es irreducible ya que V era irreducible y, por consiguiente, existe un polinomio irreducible $g_j \in K[X_1, \dots, X_m, Y_{m+1}, \dots, Y_r, Y_j]$ tal que $I(\overline{\Pi_j(V_q)}^z) = (g_j)$. Como H_j se anula en $\Pi_j(V_q)$, entonces tenemos que $g_j \mid H_j$ y, como H_j es primitivo e irreducible, concluimos que $g_j = H_j$ salvo por alguna constante en K .

Finalmente, los grados verifican las siguientes igualdades y desigualdades:

$$\deg(V) = \deg(V_q) \geq \deg_z(\Pi_j(V_q)) = \deg(\overline{\Pi_j(V_q)}^z) = \deg(g_j) = \deg(H_j).$$

■

Usando la afirmación previa, concluimos que para cada j , $r+1 \leq j \leq n$, tenemos que:

$$\deg(e_j^{(\delta_j)}) + \delta_j \leq \deg(H_j) \leq \deg(V).$$

Por lo tanto, como $\delta_j \geq 1$ para cada j , obtenemos:

$$(2.3.4) \quad \deg(q) = \sum_{j=r+1}^n \deg(e_j^{(\delta_j)}) \leq \sum_{j=r+1}^n (\deg(V) - 1) \leq (n-r)(\deg(V) - 1).$$

Combinando esta última desigualdad con el resultado que aparece en la Ecuación (2.3.3) y el Lema 2.3.1 precedente, concluimos que:

$$q \in K[X_1, \dots, X_m] \setminus \{0\}$$

es el polinomio que estamos buscando. □

OBSERVACIÓN 2.3.3. Este último lema explica la existencia del defecto en el ejemplo utilizado para demostrar el Ítem *iv*) de la Proposición 2.1.11. En dicho ejemplo, empleábamos la hipersuperficie cuadrática W' de $\mathbb{A}^3(\mathbb{C})$ de grado 2 dada por la siguiente identidad:

$$W' := \{(x, y, z) \in \mathbb{A}^3(\mathbb{C}) : xz + (y^2 - 1) = 0\}.$$

Consideramos ahora la proyección canónica $\ell = \pi_2 : \mathbb{A}^3(\mathbb{C}) \rightarrow \mathbb{A}^2(\mathbb{C})$ que “olvida” la última coordenada. El polinomio q citado en el Lema 2.3.2 previo es el coeficiente director del polinomio cuadrático que define a W' , esto es, $q := Z$. Su grado total verifica:

$$\deg(q) = 1 \leq (\text{codim}(W'))(\deg(W') - 1) = (3-2)(2-1) = 1.$$

Por consiguiente, el defecto está acotado por:

$$\deg_{\text{lci}}(\pi_2(W')) - \deg_z(\pi_2(W')) \leq \deg_{\text{lci}}(\pi_2(W' \cap V_{\mathbb{A}}(q))).$$

Tenemos que $W' \cap V_{\mathbb{A}}(q)$ es un par de rectas dadas por:

$$W' \cap V_{\mathbb{A}}(q) = \{(x, y, z) \in W' : z = 0\} = \{(x, y, 0) \in \mathbb{A}^3(\mathbb{C}) : y^2 - 1 = 0\}.$$

y, por lo tanto,

$$\pi_2(W' \cap V_{\mathbb{A}}(q)) = \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : y \in \{\pm 1\}\}$$

es una variedad algebraica. Entonces, se cumple que:

$$\deg_{\text{lci}}(\pi_2(W' \cap V_{\mathbb{A}}(q))) = \deg(\pi_2(W' \cap V_{\mathbb{A}}(q))) = 2.$$

Por otro lado, tenemos que $\deg_{\text{lci}}(\pi_2(W')) = 3$ y $\deg_z(\pi_2(W')) = 1$ (ver Ejemplo 2.1.3). En particular, la cota del Lema 2.3.2 anterior es óptima en este ejemplo ya que:

$$\deg_{\text{lci}}(\pi_2(W')) - \deg_z(\pi_2(W')) = 3 - 1 = 2 = \deg_{\text{lci}}(\pi_2(W' \cap V_{\mathbb{A}}(q))).$$

Esto no significa que la cota superior del Ítem *iv*) del Lema 2.3.2 precedente sea siempre óptima, sino que es óptima al menos en algunos ejemplos.

2.3.2. Una cota extrínseca para el grado de una proyección basada en el Nullstellensatz Efectivo. En esta subsección revisamos la Proposición 2.1.11. Obtendremos una cota superior sintáctica para el grado LCI de la imagen $\pi(V)$ de una variedad algebraica irreducible V con respecto a la proyección canónica π . La cota es sintáctica ya que únicamente tenemos en cuenta la descripción sintáctica de V . El método de demostración que hemos escogido se basa en el Nullstellensatz Efectivo. Esto simplemente significa que el grado LCI es “controlado” bajo imágenes lineales en términos de descripciones sintácticas del dominio. Aunque podríamos haber usado las cotas de J. Kollár (cf. [Ko, 88]) o M. Sombra (cf. [So, 99]), hemos preferido emplear las cotas del Nullstellensatz Efectivo de J. Jelonek (cf. [Je, 05]). También hemos utilizado los resultados de [Mu, 87] para las cotas de los grados, aunque podríamos haber usado otras referencias.

Al igual que antes, K es un cuerpo algebraicamente cerrado.

TEOREMA 2.3.4. *Sea $V \subseteq \mathbb{A}^n(K)$ una variedad algebraica irreducible de dimensión r . Sea $m \in \mathbb{N}$ un entero positivo tal que $m \leq n$. Supongamos que existen polinomios $g_1, \dots, g_s \in K[X_1, \dots, X_n]$ de grados $d_i := \deg(g_i)$, $1 \leq i \leq s$, tales que $V = V_{\mathbb{A}}(g_1, \dots, g_s)$ y se verifican las siguientes desigualdades:*

$$d_1 \geq d_2 \geq \dots \geq d_s.$$

Consideramos la siguiente cantidad:

$$N := N(d_1, \dots, d_s, n, r) := \begin{cases} \prod_{i=1}^s d_i & \text{si } s \leq n - m \\ 2d_s \left(\prod_{i=1}^{n-m-1} d_i \right) - 1 & \text{si } s > n - m \end{cases},$$

y definimos las siguientes cantidades:

$$\begin{aligned} \tilde{N} &:= \binom{N + (n - m)}{n - m}, \\ M &:= \sum_{i=1}^s \binom{N - d_i + (n - m)}{(n - m)}, \end{aligned}$$

y

$$\mathcal{N}' := \min\{\tilde{N}, M + 1\}.$$

Sea $\pi : \mathbb{A}^n(K) \rightarrow \mathbb{A}^m(K)$ la proyección canónica que “olvida” las últimas $n - m$ variables y sea $W := \overline{\pi(V)}^z$ la clausura Zariski de $\pi(V)$ en $\mathbb{A}^m(K)$. Entonces, tenemos que:

$$\deg_{\text{Lci}}(\pi(V)) \leq \deg(V) (2d_1)^{\dim(W)} (\mathcal{N}')^{\dim(W)+1}.$$

DEMOSTRACIÓN. En primer lugar, para cada exponente monomial:

$$\nu = (\nu_1, \dots, \nu_k) \in \mathbb{N}^k,$$

denotamos su grado por:

$$|\nu| := \nu_1 + \dots + \nu_k \in \mathbb{N}.$$

A continuación, para cada i , $1 \leq i \leq s$, vamos a introducir un nuevo conjunto de variables:

$$\underline{Z}^{(i)} := \{Z_{\mu}^{(i)} : \mu \in \mathbb{N}^{n-m}, |\mu| \leq N - d_i\},$$

ordenadas por el orden monomial “grado+lexicográfico”, y el polinomio con coeficientes genéricos:

$$F_i := \sum_{\mu \in \mathbb{N}^{n-m}, |\mu| \leq N - d_i} Z_{\mu}^{(i)} X_{m+1}^{\mu_{m+1}} \dots X_n^{\mu_n}.$$

Nótese que el número de variables involucradas en $\underline{Z}^{(i)}$ (y también el número de coeficientes genéricos de F_i) está dado por:

$$M_i := \binom{N - d_i + (n - m)}{(n - m)}.$$

Además, para cada i , $1 \leq i \leq s$, reescribimos los polinomios g_1, \dots, g_s como:

$$g_i := \sum_{\nu \in \mathbb{N}^{n-m}, |\nu| \leq d_i} h_{\nu}^{(i)}(X_1, \dots, X_m) X_{m+1}^{\nu_{m+1}} \dots X_n^{\nu_n}.$$

Finalmente, consideramos el vector columna:

$$\mathbf{1} := \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in K^{\tilde{N}},$$

que codifica los coeficientes de $\mathbf{1} \in K[X_{m+1}, \dots, X_n]$ usando un polinomio de grado a lo sumo \tilde{N} , con respecto al orden monomial “grado+lexicográfico” en $K[X_{m+1}, \dots, X_n]$. El número de coordenadas de $\mathbf{1}$ viene dado por la siguiente cantidad:

$$\tilde{N} := \binom{N + (n - m)}{n - m}.$$

A continuación, consideramos la siguiente suma:

$$\sum_{i=1}^s F_i g_i \in K[\underline{Z}^{(1)}, \dots, \underline{Z}^{(s)}, X_1, \dots, X_m][X_{m+1}, \dots, X_n].$$

Tomando $R := K[\underline{Z}^{(1)}, \dots, \underline{Z}^{(s)}, X_1, \dots, X_m]$, los coeficientes de esta suma como polinomios en $R[X_{m+1}, \dots, X_n]$ pueden representarse como un producto de matrices:

$$A(X_1, \dots, X_m) \begin{pmatrix} \underline{Z}^{(1)} \\ \vdots \\ \underline{Z}^{(s)} \end{pmatrix}.$$

Los coeficientes de $A(X_1, \dots, X_m)$, que son los elementos de la lista:

$$(2.3.5) \quad \mathcal{H} := \{h_\nu^{(i)} : \nu \in \mathbb{N}^{n-m}, |\nu| \leq d_i, 1 \leq i \leq s\},$$

tienen todos grado a lo sumo $D := d_1$. Además, la matriz $A(X_1, \dots, X_m)$ está en:

$$\mathcal{M}_{\tilde{N} \times M}(K[X_1, \dots, X_m]),$$

donde el número de filas y columnas están determinados por las siguientes reglas:

- El número de filas es la cantidad:

$$\tilde{N} := \binom{N + (n - m)}{n - m}.$$

- El número de columnas viene dado por la cantidad:

$$M := \sum_{i=1}^s \binom{N - d_i + (n - m)}{n - m}.$$

Finalmente, consideramos el sistema de ecuaciones lineales:

$$(2.3.6) \quad \mathcal{S}(X_1, \dots, X_m) \equiv \left\{ A(X_1, \dots, X_m) \begin{pmatrix} \underline{Z}^{(1)} \\ \underline{Z}^{(2)} \\ \vdots \\ \underline{Z}^{(s)} \end{pmatrix} = \mathbf{1} \right\}.$$

Nótese que las siguientes dos afirmaciones son equivalentes para cada $x = (x_1, \dots, x_m) \in \mathbb{A}^m(K)$:

- La fibra $\pi^{-1}(\{x\}) \cap V \neq \emptyset$.
- El sistema de ecuaciones lineales $\mathcal{S}(x_1, \dots, x_m)$, obtenido al especializar las variables X_i en las coordenadas x_i , es *incompatible*.

Seguimos ahora la idea principal de [Mu, 87], que es válida sobre cualquier cuerpo y en cualquier característica. El truco de Mulmuley funciona de la siguiente forma. En primer lugar, sea $W := \overline{\pi(V)}^z$ la clausura Zariski de $\pi(V)$ en $\mathbb{A}^m(K)$. Definimos las dos matrices siguientes:

$$A(X_1, \dots, X_m) \in \mathcal{M}_{\tilde{N} \times M}(K[W]),$$

$$B(X_1, \dots, X_m) := (A(X_1, \dots, X_m) \mid \mathbf{1}) \in \mathcal{M}_{\tilde{N} \times (M+1)}(K[W]).$$

Entonces, consideramos las dos siguientes matrices cuadradas y simétricas:

$$A^*(X_1, \dots, X_m) := \begin{pmatrix} A(X_1, \dots, X_m) & 0 \\ 0 & A(X_1, \dots, X_m)^T \end{pmatrix} \in \mathcal{M}_{\tilde{N}+M}(K[W]),$$

$$B^*(X_1, \dots, X_m) := \begin{pmatrix} B(X_1, \dots, X_m) & 0 \\ 0 & B(X_1, \dots, X_m)^T \end{pmatrix} \in \mathcal{M}_{\tilde{N}+M+1}(K[W]).$$

Además, tenemos que los rangos verifican las siguientes igualdades para todo $x \in \mathbb{A}^m(K)$:

$$\text{rango}(A^*(x)) = 2 \text{rango}(A(x)), \quad \text{rango}(B^*(x)) = 2 \text{rango}(B(x)).$$

Para cada $x \in W$, el sistema de ecuaciones lineales $\mathcal{S}(x)$, descrito en la Identidad (2.3.6), es incompatible si y solo si $\text{rango}(A^*(x)) \neq \text{rango}(B^*(x))$.

Como $A^*(X_1, \dots, X_m)$ y $B^*(X_1, \dots, X_m)$ son matrices simétricas, sus rangos dependen del orden de 0 como raíz de sus respectivos polinomios característicos. Por consiguiente, consideramos sus polinomios característicos módulo $I(W)$:

- El polinomio característico de la matriz $A^*(X_1, \dots, X_m)$ viene dado por:

$$\chi_A(T) := T^{\tilde{N}+M} + a_{\tilde{N}+M-1}T^{\tilde{N}+M-1} + \dots + a_1T + a_0 \in K[W][T],$$

donde los coeficientes a_i son de la forma $A_i + I(W)$, con $A_i \in K[X_1, \dots, X_m]$.

- Análogamente, consideramos el polinomio característico de la matriz $B^*(X_1, \dots, X_m)$:

$$\chi_B(T) := T^{\tilde{N}+M+1} + b_{\tilde{N}+M}T^{\tilde{N}+M} + \dots + b_1T + b_0 \in K[W][T].$$

También concluimos que los coeficientes $b_i \in K[W]$ son de la forma $B_i + I(W)$, con $B_i \in K[X_1, \dots, X_m]$.

Sean $\mathcal{N} := \min\{\tilde{N}, M\}$ y $\mathcal{N}' := \min\{\tilde{N}, M+1\}$ el mínimo entre el número de filas y columnas de $A(X_1, \dots, X_m)$ y $B(X_1, \dots, X_m)$ respectivamente. Para cada $x \in W$, tenemos que:

$$\text{rango}(A^*(x)) \leq 2\mathcal{N}, \quad \text{rango}(B^*(x)) \leq 2\mathcal{N}'.$$

Por lo tanto, el orden de 0 en $\chi_A(T)$ y $\chi_B(T)$ verifica las siguientes desigualdades en $K[W]$:

$$\text{ord}_0(\chi_A) \geq \tilde{N} + M - 2\mathcal{N}, \quad \text{ord}_0(\chi_B) \geq \tilde{N} + M + 1 - 2\mathcal{N}'.$$

En particular, se cumplen las igualdades:

$$\chi_A(T) := T^{\tilde{N}+M} + a_{\tilde{N}+M-1}T^{\tilde{N}+M-1} + \dots + a_{\tilde{N}+M-2\mathcal{N}}T^{\tilde{N}+M-2\mathcal{N}} \in K[W][T],$$

y

$$\chi_B(T) := T^{\tilde{N}+M+1} + b_{\tilde{N}+M}T^{\tilde{N}+M} + \dots + b_{\tilde{N}+M+1-2\mathcal{N}'}T^{\tilde{N}+M+1-2\mathcal{N}'} \in K[W][T].$$

Como las coordenadas de las matrices $A^*(X_1, \dots, X_m)$ y $B^*(X_1, \dots, X_m)$ están en la clase $\mathcal{H} \cup \{0, 1\}$, donde \mathcal{H} es la lista de polinomios introducida en la Identidad (2.3.5), y los polinomios en \mathcal{H} tienen grado a lo sumo $D := d_1$, concluimos:

- Para cada i , $1 \leq i \leq 2\mathcal{N}$, existe un polinomio $A_{\tilde{N}+M-i}(X_1, \dots, X_m) \in K[X_1, \dots, X_m]$ de grado a lo sumo $d_1 i$ tal que $A_{\tilde{N}+M-i} + I(W) = a_{\tilde{N}+M-i}$.
- Para cada j , $1 \leq j \leq 2\mathcal{N}'$, existe un polinomio $B_{\tilde{N}+M+1-j}(X_1, \dots, X_m) \in K[X_1, \dots, X_m]$ de grado a lo sumo $d_1 j$ tal que $B_{\tilde{N}+M+1-j} + I(W) = b_{\tilde{N}+M+1-j}$.

Para cada $x \in W$, como:

$$\text{rango}(B^*(x)) = 2 \text{rango}(B(x)) \geq 2 \text{rango}(A(x)) = \text{rango}(A^*(x)),$$

el sistema de ecuaciones lineales $\mathcal{S}(x)$ de la Identidad (2.3.6) es incompatible si y solo si existe $k \in \{0, \dots, \mathcal{N}\}$ tal que:

$$\text{rango}(A^*(x)) = 2k, \quad \text{rango}(B^*(x)) = 2(k+1).$$

Transformando estas igualdades de rango en el orden de 0 como raíz de χ_A y χ_B , para cada $x \in W$, el sistema de ecuaciones lineales $\mathcal{S}(x)$ descrito en la Identidad (2.3.6) es incompatible si y solo si existe k , $0 \leq k \leq \mathcal{N}$, tal que $x \in R_k$, donde R_k es la clase dada por la siguiente intersección:

$$R_k := \{x \in W : \text{ord}_0(\chi_{A(x)}(T)) = \tilde{N} + M - 2k\} \cap \{x \in W : \text{ord}_0(\chi_{B(x)}(T)) = \tilde{N} + M + 1 - 2(k+1)\}.$$

En términos de ecuaciones polinomiales e igualdades, lo anterior puede escribirse de la siguiente forma:

$$R_k := \mathcal{A}_k \cap \mathcal{B}_k,$$

donde

$$\mathcal{A}_k := \{x \in W : A_i(x) = 0, \tilde{N} + M - 2\mathcal{N} \leq i \leq \tilde{N} + M - 2k - 1, A_{\tilde{N}+M-2k}(x) \neq 0\},$$

y

$$\mathcal{B}_k := \{x \in W : B_j(x) = 0, \tilde{N} + M + 1 - 2\mathcal{N}' \leq j \leq \tilde{N} + M + 1 - 2(k+1) - 1, B_{\tilde{N}+M+1-2(k+1)}(x) \neq 0\}.$$

En particular, se verifica la siguiente igualdad:

$$\pi(V) := \bigcup_{k=0}^{\mathcal{N}-1} R_k.$$

Además, cada R_k es un conjunto localmente cerrado dado como la intersección de W con conjuntos abiertos e hipersuperficies dadas por ecuaciones polinómicas de grado a lo sumo:

$$\max \left(\{\deg(A_i) : \tilde{N} + M - 2\mathcal{N} \leq i\} \cup \{\deg(B_j) : \tilde{N} + M + 1 - 2\mathcal{N}' \leq j\} \right).$$

En particular, R_k es un conjunto localmente cerrado dado como la intersección de un subconjunto abierto Zariski de W con hipersuperficies de grado a lo sumo $2\mathcal{N}'$. Gracias a la Proposición 2.2.2, obtenemos:

$$\deg(R_k) \leq \deg(W) (2\mathcal{N}'d_1)^{\dim(W)}.$$

Como el grado de conjuntos constructibles es sub-aditivo (cf. Proposición 2.1.2), concluimos:

$$\deg_{\text{lci}}(\pi(V)) \leq \deg(W) \mathcal{N} (2\mathcal{N}'d_1)^{\dim(W)}.$$

Aplicando la Proposición 1.2.6 a la desigualdad previa obtenemos la desigualdad buscada. \square

Conjuntos Cuestores

Índice

3.1. La noción, algunas nociones equivalentes y primeras propiedades	39
3.1.1. El caso de polinomios y listas de polinomios	44
3.2. Alta probabilidad de existencia de Conjuntos Cuestores de longitud asintóticamente óptima en conjuntos localmente cerrados equidimensionales para conjuntos constructibles de listas de polinomios	45
3.2.1. Demostración del Teorema 3.2.1	46
3.2.2. Algunos Corolarios útiles	49
3.2.3. Conjuntos constructibles unirracionales	53
3.2.4. El conjunto de muestreo de DeMillo–Lipton–Schwartz–Zippel	54
3.3. Algunos Conjuntos Cuestores en la literatura matemática	54
3.3.1. Conjuntos Cuestores y variedades determinantaes	54
3.3.2. Conjuntos Cuestores y Conjuntos de Makeya	55
3.3.3. Conjuntos Cuestores, Nullstellensatz Combinatorio de Alon y dualidad	59
3.4. Un algoritmo probabilista para detectar la dimensión genérica: “Suite Sécante”	64
3.4.1. Algunos aspectos básicos sobre las “Suites Sécantes”	64
3.4.2. El Algoritmo	68
3.4.3. Clase de todas las listas de polinomios evaluables por un esquema de evaluación	71

En este capítulo revisamos la noción de conjunto cuestor o “correct test sequence” (introducida en [HS, 82]) y mostramos que esta noción es casi omnipresente en la literatura matemática (a través de varios ejemplos como los conjuntos normantes, los conjuntos de Makeya o el Nullstellensatz Combinatorio de Alon). Además, enfatizamos que la aplicación natural de los conjuntos cuestores no son los Tests de Nulidad, sino que están diseñados para detectar, mediante evaluación de “muestras”, si una función $f \in \Omega$ pertenece o no a un subconjunto “discriminante” $\Sigma \subsetneq \Omega$.

Por otro lado, estudiamos en detalle los conjuntos cuestores para listas de polinomios definidas en un conjunto constructible Ω . En primer lugar, probamos que los conjuntos cuestores verifican la “maldición de la dimensionalidad”: la dimensión de Krull de Ω es una cota inferior para la longitud de cualquier conjunto cuestor para Ω con respecto a un discriminante Σ . Además, usando la noción de grado *LCI* para conjuntos constructibles (ver Definición 13) y la Proposición 2.2.2, introducidas en el capítulo anterior, demostramos que los conjuntos cuestores existen y son altamente densos en cualquier conjunto constructible de co-dimensión y grado apropiados (no únicamente en conjuntos “reticulares” como en [HS, 82]).

En la última sección del capítulo, presentamos un algoritmo probabilista para el Problema de la “Suite Sécante” que no manipula la lista de polinomios de entrada: simplemente evalúa los polinomios dados en puntos de una “muestra” aleatoria.

3.1. La noción, algunas nociones equivalentes y primeras propiedades

Comenzamos esta sección generalizando al contexto de aplicaciones a valores vectoriales la noción de conjunto cuestor introducida en [HS, 82] para Tests de Nulidad de Polinomios:

DEFINICIÓN 16 (Conjuntos Cuestores para aplicaciones a valores vectoriales). *Sea X un conjunto, K un cuerpo y sea $\mathcal{F}(X) \subseteq K^X$ un subgrupo del grupo abeliano $(K^X, +)$. Sea $m \in \mathbb{N}$ un entero positivo y sea $\Omega \subseteq \mathcal{F}(X)^m$ un conjunto de listas de tales funciones. Sea $\Sigma \subsetneq \Omega$ un subconjunto propio de Ω , al que llamaremos discriminante. Un conjunto cuestor o “correct test sequence” (abreviado habitualmente como *CTS*) de longitud L para Ω con discriminante Σ es un conjunto finito de L elementos $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq X$ tal que se verifica la siguiente fórmula:*

$$(3.1.1) \quad \forall f \in \Omega, \quad f(x_1) = \dots = f(x_L) = 0 \in K^m \implies f \in \Sigma.$$

En el caso $\Sigma = \{0\} \subsetneq \Omega$, decimos simplemente que \mathbf{Q} es un conjunto cuestor para Ω .

Dado $\Omega \subseteq \mathcal{F}(X)^m$, denotamos por $\text{Dif}(\Omega)$ al subconjunto de $(\mathcal{F}(X)^m, +)$ dado por las diferencias entre los elementos de Ω . Es decir,

$$\text{Dif}(\Omega) := \{f - g : f, g \in \Omega\} \subseteq \mathcal{F}(X)^m.$$

Nótese que si Ω es un semigrupo de $(\mathcal{F}(X)^m, +)$, entonces $\text{Dif}(\Omega)$ es simplemente el subgrupo abeliano de $(\mathcal{F}(X)^m, +)$ generado por Ω .

Para una lista de puntos $\mathbf{Q} := \{x_1, \dots, x_L\}$, definimos la función de evaluación en los puntos de \mathbf{Q} como la siguiente aplicación:

$$\text{ev}_{\mathbf{Q}} : \begin{array}{ccc} \mathcal{F}(X)^m & \longrightarrow & K^{mL} \\ (f_1, \dots, f_m) & \longmapsto & (f_1(x_1), \dots, f_m(x_1), \dots, f_1(x_L), \dots, f_m(x_L)) \end{array}$$

que es una aplicación K -lineal cuando $\mathcal{F}(X)^m$ es un subespacio vectorial de $(K^m)^X$.

PROPOSICIÓN 3.1.1 (Conjuntos Cuestores y Test de Identidad de Funciones). *Con las notaciones de la Definición 16 precedente, sea Ω un subconjunto de $\mathcal{F}(X)^m$. Entonces, las siguientes propiedades son equivalentes para cada subconjunto finito $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq X$:*

- i) \mathbf{Q} es un conjunto cuestor para $\text{Dif}(\Omega)$ (con discriminante $\Sigma := \{0\}$).
- ii) Se verifica la siguiente fórmula:

$$\forall f, g \in \Omega, \quad f(x_1) = g(x_1), \dots, f(x_L) = g(x_L) \implies f = g.$$

Si, además, $\mathcal{F}(X)^m$ es un subespacio vectorial de $(K^m)^X$ y $0 \in \Omega$, las dos propiedades anteriores son equivalentes a las siguientes:

- iii) La función de evaluación $\text{ev}_{\mathbf{Q}}$ en los puntos de \mathbf{Q} es una aplicación K -lineal cuya restricción a Ω es inyectiva.
- iv) Si $L := \sharp(\mathbf{Q})$, existe una aplicación lineal $\Lambda : K^{mL} \longrightarrow \mathcal{F}(X)^m$ que cumple las dos siguientes propiedades:
 - Para todo $f \in \mathcal{F}(X)^m$, $f - \Lambda(\text{ev}_{\mathbf{Q}}(f)) \in \ker(\text{ev}_{\mathbf{Q}})$ y,
 - $\text{ev}_{\mathbf{Q}}(\Omega) \cap \ker(\Lambda) = \{0\}$.

DEMOSTRACIÓN. Las equivalencias entre los Ítems i), ii) y, en el caso de que $\mathcal{F}(X)$ sea un espacio vectorial, iii), son inmediatas. Respecto a la equivalencia entre los Ítems iii) y iv), basta asumir que Λ es una extensión a K^{mL} de la inversa por la derecha del epimorfismo entre $\mathcal{F}(X)^m$ y $\text{ev}_{\mathbf{Q}}(\mathcal{F}(X)^m) \subseteq K^{mL}$. \square

La interpretación más común de los conjuntos cuestores en la literatura matemática es la de Test de Identidad de Funciones. Cuando $\mathcal{F}(X)$ es un grupo de polinomios definidos sobre una variedad algebraica X , los conjuntos cuestores reciben el nombre de *Hitting Sets* (como en [Sax, 09], [Sax, 14] y sus referencias). Los Tests probabilísticos de Identidad de Polinomios han sido empleados en Teoría de la Eliminación durante años para calcular normalizaciones de Noether o descripciones birracionalmente de Kronecker de variedades equidimensionales, así como para otros muchos propósitos (ver, por ejemplo, [Pa, 95], [GHHMMP, 97], [GHMP, 97] y sus referencias). Retomamos el caso polinómico en la siguiente sección.

Con las notaciones precedentes, supongamos que $\mathcal{F}(X)$ es una K -álgebra. Entonces, para cada $x \in X$, el siguiente ideal es maximal en $\mathcal{F}(X)$:

$$\mathfrak{m}_x := \{f \in \mathcal{F}(X) : f(x) = 0\} \in \text{MaxSpec}(\mathcal{F}(X)).$$

Es más, \mathfrak{m}_x es el núcleo de la aplicación suprayectiva $\text{ev}_x : \mathcal{F}(X) \longrightarrow K$ dada por $\text{ev}_x(f) := f(x)$, para todo $f \in \mathcal{F}(X)$. Para cada $\mathbf{Q} \subseteq X$, denotamos por $I(\mathbf{Q})$ al ideal radical dado por la siguiente igualdad:

$$I(\mathbf{Q}) := \bigcap_{x \in \mathbf{Q}} \mathfrak{m}_x.$$

PROPOSICIÓN 3.1.2. *Con las notaciones anteriores, supongamos que $\mathcal{F}(X)$ es una K -álgebra y $\Omega \subseteq \mathcal{F}(X)$ un subconjunto. Las siguientes propiedades son equivalentes:*

- i) Un subconjunto finito $\mathbf{Q} \subseteq X$ es un conjunto cuestor para $\text{Dif}(\Omega)$.
- ii) La proyección canónica $\pi : \mathcal{F}(X) \longrightarrow \mathcal{F}(\mathbf{Q}) := \mathcal{F}(X)/I(\mathbf{Q})$ satisface que su restricción a Ω es inyectiva.
- iii) Existe un K -subespacio vectorial $W(\mathbf{Q})$ de $\mathcal{F}(X)$ de dimensión $\sharp(\mathbf{Q})$ que verifica las siguientes propiedades:
 - Para cada $f \in \Omega$, existe un único $g \in W(\mathbf{Q})$ tal que $f - g \in I(\mathbf{Q})$.
 - Para cada $g \in W(\mathbf{Q})$, existe a lo sumo un $f \in \Omega$ tal que $f - g \in I(\mathbf{Q})$.

DEMOSTRACIÓN. Por el Teorema Chino de los Restos, tenemos el siguiente isomorfismo de anillos:

$$\varphi : \mathcal{F}(X)/I(\mathbf{Q}) \longrightarrow \prod_{i=1}^L \mathcal{F}(X)/\mathfrak{m}_{x_i} \cong K^L,$$

donde $\mathbf{Q} := \{x_1, \dots, x_L\}$. Observando que:

$$\varphi(f + I(\mathbf{Q})) := (f + \mathfrak{m}_{x_1}, \dots, f + \mathfrak{m}_{x_L}),$$

concluimos la equivalencia entre los Ítems *i*) y *ii*). Además, el Teorema Chino de los Restos nos asegura que hay interpolación en $\mathcal{F}(X)$. Es decir, existen funciones $\chi_1, \dots, \chi_L \in \mathcal{F}(X)$ tales que $\chi_i(x_j) = \delta_{i,j}$, para cada i, j , $1 \leq i, j \leq L$, donde $\delta_{i,j}$ es la delta de Kronecker. Entonces, tomando $W(\mathbf{Q}) := \text{Span}(\{\chi_1, \dots, \chi_L\})$, el K -espacio vectorial generado por $\{\chi_1, \dots, \chi_L\}$, es claro que para cada $f \in \Omega$, la función:

$$g := \sum_{i=1}^L f(x_i)\chi_i \in W(\mathbf{Q}),$$

verifica los requisitos de nuestra afirmación. Por lo tanto, dado cualquier $g \in W(\mathbf{Q})$, existe a lo sumo un elemento $f \in \Omega$ tal que $f - g \in I(\mathbf{Q})$. \square

Nótese también que la condición de ser un conjunto cuestor es hereditaria, tal y como enuncia la siguiente proposición de demostración inmediata:

PROPOSICIÓN 3.1.3. Sean $\Omega \subseteq \tilde{\Omega} \subseteq \mathcal{F}(X)^m$ dos subconjuntos y sea $\Sigma \subsetneq \tilde{\Omega}$ un discriminante. Sea $\mathbf{Q} \subseteq X$ un conjunto cuestor para $\tilde{\Omega}$ con respecto a Σ . Entonces, \mathbf{Q} es un conjunto cuestor para Ω con respecto a $\Sigma \cap \Omega$.

EJEMPLO 3.1.4 (**Conjuntos Normantes finitos**). Con las notaciones precedentes, supongamos que $(K, |\cdot|)$ es un cuerpo con algún valor absoluto y, para cualquier entero positivo $m \in \mathbb{N}$, sea $\|\cdot\|$ cualquier norma en K^m inducida por $|\cdot|$. Supongamos que $\mathcal{F}(X)$ es un K -espacio vectorial y $\Omega \subseteq \mathcal{F}(X)$ un subconjunto. Un *conjunto normante finito* (o “finite norming set”) para Ω es un subconjunto finito $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq X$ de modo que la siguiente expresión es una norma en Ω :

$$\|f\|_{\infty}^{(\mathbf{Q})} := \max\{|f(x_i)| : 1 \leq i \leq L\}.$$

Como Ω no es siempre un subespacio vectorial, empleamos en este contexto el término “norma” en un sentido más amplio. Una norma para Ω es una función $\|\cdot\| : \Omega \rightarrow \mathbb{R}$ que verifica las siguientes propiedades:

- $\|f\| \geq 0$, $\forall f \in \Omega$.
- Para todo $f \in \Omega$, $\|f\| = 0$ si y solo si $f = 0$.
- Dados $f, g \in \Omega$, si $f + g \in \Omega$, entonces tenemos que:

$$\|f + g\| \leq \|f\| + \|g\|.$$

- Dados $f \in \Omega$ y $\lambda \in K$, si $\lambda f \in \Omega$, entonces $\|\lambda f\| \leq |\lambda| \|f\|$.

Obviamente, si Ω es un subespacio vectorial, esta es la noción habitual de norma. Ver [AK, 06] y sus referencias para el término *conjunto normante* (o “norming set”).

PROPOSICIÓN 3.1.5. Con las mismas notaciones e hipótesis que en el Ejemplo 3.1.4 precedente, dados $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq X$ un conjunto finito y $\Omega \subseteq \mathcal{F}(X)$ un subgrupo del grupo aditivo $(\mathcal{F}(X), +)$, tenemos que los siguientes ítems son equivalentes:

- i) El conjunto \mathbf{Q} es un conjunto cuestor para Ω .
- ii) El conjunto \mathbf{Q} es un conjunto normante finito para Ω .
- iii) La restricción de la función de evaluación $\text{ev}_{\mathbf{Q}}$ a Ω es un monomorfismo de grupos.
- iv) Existe p , $1 \leq p \leq \infty$, tal que:

$$\|f\|_p^{(\mathbf{Q})} := \left(\sum_{i=1}^L |f(x_i)|^p \right)^{1/p}$$

es una norma en Ω .

- v) Para cada p , $1 \leq p \leq \infty$:

$$\|f\|_p^{(\mathbf{Q})} := \left(\sum_{i=1}^L |f(x_i)|^p \right)^{1/p}$$

es una norma en Ω .

- vi) Ω es un espacio métrico con función distancia:

$$d_{\infty}^{(\mathbf{Q})}(f, g) := \|f - g\|_{\infty}^{(\mathbf{Q})}.$$

vii) Para cada p , $1 \leq p \leq \infty$, Ω es un espacio métrico con función distancia:

$$d_p^{(\mathbf{Q})}(f, g) := \|f - g\|_p^{(\mathbf{Q})}.$$

En particular, Ω tiene un conjunto cuestor de longitud finita si y solo si admite un conjunto normante finito. Si, además, Ω fuese un subespacio lineal de $\mathcal{F}(X)$, entonces Ω tiene un conjunto cuestor de longitud finita si y solo si Ω es un espacio lineal normado de dimensión finita.

DEMOSTRACIÓN. La equivalencia entre los ítems es obvia a partir de las definiciones. \square

Los casos interesantes se restringen, entonces, a subconjuntos (no necesariamente lineales) de algunos espacios vectoriales de dimensión finita:

EJEMPLO 3.1.6 (Conjuntos Cuestores y el anillo de funciones continuas). Supongamos que X es un espacio topológico compacto, $\kappa = \mathbb{R}$ y $\mathcal{F}(X) \subseteq \mathcal{C}(X)$ es un subespacio vectorial de la \mathbb{R} -álgebra de funciones continuas definidas en X con valores en \mathbb{R} . Supongamos que $\mathcal{C}(X)$ está dotado de la topología definida por la norma del máximo, lo que le confiere estructura de \mathbb{R} -álgebra de Banach. Por el Teorema de Banach–Stone–Čech–Gel’fand–Kolmogorov (cf. [GHJ, 54]), sabemos que existe una biyección entre los ideales maximales en el espectro de $\mathcal{C}(X)$ (que denotamos por $\text{MaxSpec}(\mathcal{C}(X))$) y el conjunto de puntos de X . Esta biyección viene dada por la siguiente identificación:

$$x \longmapsto \mathfrak{m}_x := \{f \in \mathcal{C}(X) : f(x) = 0\}.$$

Con estas notaciones e hipótesis, un subconjunto finito $\mathbf{Q} := \{x_1, \dots, x_L\} \subseteq X$ de cardinalidad L es un conjunto cuestor para $\Omega \subseteq \mathcal{F}(X)$ con respecto al discriminante $\Sigma \subsetneq \Omega$ si y solo si se verifica la siguiente expresión:

$$\Omega \cap \left(\bigcap_{i=1}^L \mathfrak{m}_{x_i} \right) \subseteq \Sigma,$$

o, equivalentemente, si y solo si se cumple la siguiente inclusión:

$$\Omega \setminus \Sigma \subseteq \left(\bigcup_{i=1}^L \mathfrak{m}_{x_i}^c \right),$$

donde $\mathfrak{m}_{x_i}^c = \mathcal{C}(X) \setminus \mathfrak{m}_{x_i}$.

PROPOSICIÓN 3.1.7. Con las notaciones precedentes, supongamos que X es un espacio topológico de Hausdorff compacto, $\mathcal{F}(X) = \mathcal{C}(X)$, $\Omega \subseteq \mathcal{C}(X)$ y $\Sigma \subsetneq \Omega$. Sea (Σ) el ideal en $\mathcal{C}(X)$ generado por Σ y supongamos que:

$$(3.1.2) \quad \sqrt[\mathcal{J}]{(\Sigma)} \cap \Omega = \Sigma,$$

donde $\sqrt[\mathcal{J}]{\cdot}$ denota al radical de Jacobson. Por lo tanto, si $\Omega \setminus \Sigma$ es casi-compacto en $\mathcal{C}(X)$ (i.e. todo recubrimiento abierto admite un subrecubrimiento finito), entonces existe un conjunto cuestor para Ω con respecto a Σ .

En particular, si $\Sigma = \{0\}$ y $\Omega \setminus \{0\}$ es casi-compacto, entonces existe un conjunto cuestor de longitud finita para Ω o, equivalentemente, Ω tiene un conjunto normante finito.

DEMOSTRACIÓN. Por el Teorema de Banach–Stone–Čech–Gel’fand–Kolmogorov, dado que X es un espacio topológico de Hausdorff compacto, se tiene la siguiente identificación:

$$X \cong \text{MaxSpec}(\mathcal{C}(X)).$$

Consideramos también el siguiente subconjunto cerrado de X :

$$V(\Sigma) := \{x \in X : f(x) = 0, \forall f \in \Sigma\} = \{x \in X : f(x) = 0, \forall f \in (\Sigma)\}.$$

Entonces, se verifican las siguientes igualdades:

$$\sqrt[\mathcal{J}]{(\Sigma)} = \bigcap_{\substack{\mathfrak{m} \in \text{MaxSpec}(\mathcal{C}(X)) \\ \mathfrak{m} \supseteq \Sigma}} \mathfrak{m} = \bigcap_{x \in V(\Sigma)} \mathfrak{m}_x.$$

La primera igualdad es simplemente la definición del radical de Jacobson de un ideal. Para la segunda igualdad, obviamente tenemos que:

$$\bigcap_{\substack{\mathfrak{m} \in \text{MaxSpec}(\mathcal{C}(X)) \\ \mathfrak{m} \supseteq \Sigma}} \mathfrak{m} \subseteq \bigcap_{x \in V(\Sigma)} \mathfrak{m}_x.$$

Recíprocamente, sea \mathfrak{m} un ideal maximal en $\mathcal{C}(X)$ que contiene a Σ . Entonces, existe $x \in X$ tal que $\mathfrak{m} = \mathfrak{m}_x$. Como $\mathfrak{m}_x \supseteq \Sigma$, entonces $f(x) = 0$ para todo $f \in \Sigma$ y, por lo tanto, $x \in V(\Sigma)$. Esto implica la inclusión inversa. Por consiguiente, obtenemos que:

$$\sqrt[\mathcal{C}]{\Sigma} \cap \Omega = \left(\bigcap_{x \in V(\Sigma)} \mathfrak{m}_x \right) \cap \Omega = \Sigma.$$

Por lo tanto,

$$\Omega \setminus \Sigma \subseteq \bigcup_{x \in V(\Sigma)} \mathfrak{m}_x^c.$$

Como \mathfrak{m}_x^c es abierto en $\mathcal{C}(X)$, tenemos un recubrimiento abierto de $\Omega \setminus \Sigma$, que es casi-compacto por hipótesis. Entonces, existe un subrecubrimiento finito:

$$\Omega \setminus \Sigma \subseteq \bigcup_{i=1}^L \mathfrak{m}_{x_i}^c,$$

para algún subconjunto finito $\mathbf{Q} := \{x_1, \dots, x_L\}$. Nótese que esta última inclusión simplemente significa que \mathbf{Q} es un conjunto cuestor de longitud L para Ω con respecto a Σ y, por lo tanto, la primera afirmación de nuestra proposición es cierta. El caso particular se verifica ya que $V(\{0\}) = X$ y $\sqrt[\mathcal{C}]{\{0\}} = \{0\}$. \square

Análogamente, tenemos que:

COROLARIO 3.1.8. *Sea $r \in \mathbb{N} \cup \{\infty\}$ un entero positivo o ∞ y sea X una \mathcal{C}^r -variedad compacta. Para cada $\Omega \subseteq \mathcal{C}^r(X)$ tal que $0 \in \Omega$, si $\Omega \setminus \{0\}$ es casi-compacto, entonces existe un conjunto cuestor para Ω de longitud finita o, equivalentemente, Ω tiene un conjunto normante finito.*

DEMOSTRACIÓN. Esencialmente la misma demostración que en la proposición anterior tras observar que:

$$\text{MaxSpec}(\mathcal{C}^r(X)) \cong X.$$

\square

Nótese que la hipótesis descrita en la Identidad (3.1.2) puede reescribirse de la siguiente forma:

$$\overline{(\Sigma)} \cap \Omega \subseteq \Sigma,$$

donde $\overline{(\Sigma)} = \sqrt[\mathcal{C}]{(\Sigma)}$ es la clausura del ideal (Σ) en $\mathcal{C}(X)$ con respecto a la m -topología de Hewitt (cf. [Hew, 48] y [GHJ, 54]).

Un ejemplo típico en el que se verifica la hipótesis “ $\Omega \setminus \Sigma$ es casi-compacto” es el caso en que Σ es cerrado en Ω , para la topología inducida por la de $\mathcal{F}(X)$, y Ω es un espacio topológico Noetheriano.

Inspirados por el caso de funciones continuas, introducimos la siguiente notación:

DEFINICIÓN 17 (Número de recubrimiento “en el contexto de Conjuntos Cuestores”). *Con las notaciones precedentes, el número de recubrimiento para $\Omega \subseteq \mathcal{F}(X)$ con respecto a Σ es la longitud mínima de un conjunto cuestor finito para Ω con respecto a Σ . Denotamos a esta cantidad por:*

$$\mathcal{N}_{\text{cts}}(\Omega, \Sigma) := \min\{L \in \mathbb{N} : \exists \text{ un conjunto cuestor } \mathbf{Q} \text{ de longitud } L \text{ para } \Omega \text{ con respecto a } \Sigma\}.$$

Por simplicidad, usaremos la siguiente notación:

$$\mathcal{N}_{\text{cts}}(\Omega) := \mathcal{N}_{\text{cts}}(\Omega, \{0\}).$$

EJEMPLO 3.1.9 (Conjuntos Cuestores en Espacios de Hilbert con Núcleo Reprodutor). Sea $K : X \times X \rightarrow \mathbb{C}$ un núcleo y sea $(\mathcal{H}_K, \langle \cdot, \cdot \rangle_K)$ su espacio de Hilbert asociado (cf. [Ar, 50] y [CZ, 07]). Recordemos que $\mathcal{H}_K \subseteq \mathbb{C}^X$ y, en el caso en el que X es un espacio métrico y K un núcleo de Mercer, $\mathcal{H}_K \subseteq \mathcal{C}(X, \mathbb{C}) \cong \mathcal{C}(X)[i] := \mathcal{C}(X)[T]/(T^2 + 1)$ está formado por funciones continuas con valores complejos. Para cada $x \in X$, sea $K_x : X \rightarrow \mathbb{C}$ el elemento en \mathcal{H}_K dado por:

$$K_x(y) := K(x, y), \quad \forall y \in X.$$

Entonces, es conocido que para cada $f \in \mathcal{H}_K$ y para cada $x \in X$, se cumple la siguiente igualdad:

$$f(x) = \langle f, K_x \rangle.$$

PROPOSICIÓN 3.1.10. *Con las notaciones anteriores, sea $\Omega \subseteq \mathcal{H}_K$ un subconjunto y sea $\Sigma \subsetneq \Omega$ un discriminante. Un subconjunto finito de puntos $\mathbf{Q} = \{x_1, \dots, x_L\} \subseteq X$ es un conjunto cuestor de longitud L para Ω con respecto a Σ si y solo si:*

$$\Omega \cap (\text{Span}(\{K_{x_1}, \dots, K_{x_L}\}))^\perp \subseteq \Sigma,$$

donde $\text{Span}(\{K_{x_1}, \dots, K_{x_L}\})$ es el subespacio vectorial de \mathcal{H}_K generado por $\{K_{x_1}, \dots, K_{x_L}\}$ y $^\perp$ denota el complemento ortogonal en \mathcal{H}_K .

DEMOSTRACIÓN. Basta tener en cuenta que este resultado es simplemente otra manera de expresar la Identidad (3.1.1). \square

3.1.1. El caso de polinomios y listas de polinomios. Para cada entero positivo $d \in \mathbb{N}$, denotamos por $P_d^K(X_1, \dots, X_n)$ al conjunto de todos los polinomios de grado a lo sumo d con coeficientes en K en el conjunto de variables $\{X_1, \dots, X_n\}$. Cuando el conjunto de variables sea claro por el contexto, escribiremos simplemente P_d^K . Si, además, se sobrentiende el cuerpo K , escribiremos simplemente P_d . El conjunto P_d^K es un K -espacio vectorial generado por una base de monomios de cardinal dado por siguiente el número combinatorio:

$$N_d := \binom{d+n}{n}$$

En esta memoria, utilizaremos la identificación $P_d^K(X_1, \dots, X_n) \cong \mathbb{A}^{N_d}(K)$.

Sea m un entero positivo y $(d) := (d_1, \dots, d_m)$ una lista de grados. Usualmente se considera $m \leq n$, aunque en algunas ocasiones también se consideran los sistemas sobredeterminados (i.e. con $m \geq n+1$) que son los que genéricamente definen \emptyset . Introducimos la clase $\mathcal{P}_{(d)}^K(X_1, \dots, X_n)$ de todas las listas de polinomios $f := (f_1, \dots, f_m)$ tales que cada $f_i \in P_{d_i}^K$. Es decir, $\mathcal{P}_{(d)}^K$ es el siguiente producto cartesiano:

$$\mathcal{P}_{(d)}^K(X_1, \dots, X_n) := \prod_{i=1}^m P_{d_i}^K(X_1, \dots, X_n).$$

Si el conjunto de variables está fijado por el contexto, escribiremos simplemente $\mathcal{P}_{(d)}^K$. Si, además, el cuerpo K está también fijado, escribiremos simplemente $\mathcal{P}_{(d)}$. Como $\mathcal{P}_{(d)}$ es el producto cartesiano de un número finito de K -espacios vectoriales finitos, su dimensión también será finita y viene dada por:

$$\dim_K(\mathcal{P}_{(d)}^K(X_1, \dots, X_n)) = \sum_{i=1}^m \dim_K(P_{d_i}^K(X_1, \dots, X_n)) = \sum_{i=1}^m N_{d_i}.$$

Denotaremos a esta cantidad por $N_{(d)}$, la cual verifica la siguiente igualdad:

$$N_{(d)} := \sum_{i=1}^m \binom{d_i+n}{n}.$$

Por otro lado, a lo largo de estas páginas emplearemos la identificación $\mathcal{P}_{(d)}^K(X_1, \dots, X_n) \cong \mathbb{A}^{N_{(d)}}(K)$.

En [HS, 82], los autores consideran conjuntos cuestores para (la clausura Zariski de) un subconjunto constructible $\Omega \subseteq P_d$ con respecto a $\{0\}$ como discriminante. En esta memoria también consideraremos conjuntos constructibles $\Omega \subseteq \mathcal{P}_{(d)}$ y discriminantes $\Sigma \subseteq \Omega$.

Para cada conjunto constructible $\Omega \subseteq \mathcal{P}_{(d)}$ y para cada punto $x \in \mathbb{A}^{N_{(d)}}(K)$, denotamos por $\Omega_x \subseteq \mathbb{A}^{N_{(d)}}(K)$ al siguiente conjunto constructible:

$$\Omega_x := \Omega \cap \{f \in \mathcal{P}_{(d)} : f(x) = 0\}.$$

Por lo tanto, podemos reescribir la noción de conjunto cuestor como una lista $\mathbf{Q} := (x_1, \dots, x_L) \in V^L$, donde $V \subseteq \mathbb{A}^{N_{(d)}}(K)$, de puntos tales que:

$$\Omega_{x_1} \cap \dots \cap \Omega_{x_L} \subseteq \Sigma.$$

PROPOSICIÓN 3.1.11. *Sea $\mathbf{Q} = \{x_1, \dots, x_L\}$ un conjunto cuestor para Ω con respecto a Σ . Entonces, tenemos que:*

$$\deg_{\text{Ici}}(\Omega_{x_1} \cap \dots \cap \Omega_{x_L}) \leq \deg_{\text{Ici}}(\Omega).$$

Además, si Σ es cero-dimensional, la intersección $\Omega_{x_1} \cap \dots \cap \Omega_{x_L}$ es un conjunto finito y tenemos que:

$$\#(\Omega_{x_1} \cap \dots \cap \Omega_{x_L}) \leq \deg_{\text{Ici}}(\Omega).$$

En particular, si $\deg_{\text{Ici}}(\Sigma) > \deg_{\text{Ici}}(\Omega)$, tenemos la inclusión estricta:

$$\Omega_{x_1} \cap \dots \cap \Omega_{x_L} \subsetneq \Sigma.$$

DEMOSTRACIÓN. Nótese que para cada punto $x \in \mathbb{A}^n(K)$, el conjunto Ω_x es la intersección de Ω con una variedad afín lineal. Por consiguiente, todas las afirmaciones son consecuencias inmediatas de la Proposición 2.1.13. \square

La siguiente proposición muestra cotas inferiores para la longitud de conjuntos cuestores en términos de la dimensión de Krull del conjunto constructible dado:

PROPOSICIÓN 3.1.12 (**Maldición de la dimensionalidad**). *Con las notaciones precedentes, sea $\Omega \subseteq \mathcal{P}_{(d)}$ un subconjunto constructible, $\Sigma \subseteq \Omega$ un discriminante en Ω y $\mathbf{Q} \in V^L$ un conjunto cuestor de longitud L para Ω con respecto a Σ . Entonces, tenemos que:*

$$\sharp(\mathbf{Q}) = L \geq \text{codim}_{\Omega}(\Sigma) = \dim(\Omega) - \dim(\Sigma).$$

En particular, el número de recubrimiento verifica:

$$\mathcal{N}_{\text{cts}}(\Omega, \Sigma) \geq \dim(\Omega) - \dim(\Sigma).$$

DEMOSTRACIÓN. La intersección $\Omega_{x_1} \cap \cdots \cap \Omega_{x_L}$ es la intersección de Ω con una familia de L ecuaciones lineales cuyas variables son los coeficientes de los elementos en $\mathcal{P}_{(d)}$. Entonces, por el Hauptidealsatz de Krull, la dimensión de esta intersección satisface:

$$\dim(\Omega \cap \{f \in \mathcal{P}_{(d)} : f(x_1) = 0\} \cap \cdots \cap \{f \in \mathcal{P}_{(d)} : f(x_L) = 0\}) \geq \dim(\Omega) - L.$$

Por otro lado, si $\mathbf{Q} := (x_1, \dots, x_L)$ es un conjunto cuestor para Ω con respecto a Σ , tenemos que:

$$\Omega \cap \{f \in \mathcal{P}_{(d)} : f(x_1) = 0\} \cap \cdots \cap \{f \in \mathcal{P}_{(d)} : f(x_L) = 0\} \subseteq \Sigma.$$

Por lo tanto, concluimos la proposición dado que tenemos:

$$\dim(\Omega) - L \leq \dim(\Sigma).$$

\square

Teniendo en cuenta la Proposición 3.1.12 precedente, diremos que *un conjunto cuestor para Ω tiene longitud asintóticamente óptima* si su longitud está en $O(\dim(\Omega))$.

3.2. Alta probabilidad de existencia de Conjuntos Cuestores de longitud asintóticamente óptima en conjuntos localmente cerrados equidimensionales para conjuntos constructibles de listas de polinomios

Esta sección está dedicada a demostrar el Teorema 3.2.1. Dicho teorema no solo establece la existencia de conjuntos cuestores en el caso de funciones polinomiales, sino que también proporciona cotas inferiores para la probabilidad de su existencia en una gran cantidad de dominios donde es posible el muestreo.

Con las mismas notaciones que en las secciones precedentes, sea $\Omega \subseteq \mathcal{P}_{(d)}^K$ un conjunto constructible de polinomios con $(d) = (d_1, \dots, d_m)$, $m \leq n$. Nuestro objetivo es generalizar el resultado probabilístico principal de [HS, 82], mostrando que el requisito de “retículo” no es imprescindible y que podemos encontrar (con alta probabilidad) conjuntos cuestores en muchos conjuntos localmente cerrados de co-dimensión y grado adecuados. Además, veremos que los conjuntos cuestores son adecuados no solo para Tests de Nulidad, sino también para decidir la pertenencia a subvariedades propias de Ω .

Al igual que en la Subsección 1.2, dados $n, r \in \mathbb{N}$ dos enteros positivos tales que $r \leq n$, consideramos el “Grassmaniano” $\mathbb{G}(n, r)$ de todas las variedades afines lineales en $\mathbb{A}^n(K)$ de co-dimensión r . Además, supondremos que $\mathbb{G}(n, r)$ está dotado de la topología Zariski final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2).

Sea $\Sigma \subseteq \Omega$ un subconjunto constructible de co-dimensión al menos 1. Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible y $L \geq 0$ un entero positivo. Denotamos por $R(\Omega, \Sigma, C, L)$ al conjunto de todas las listas $\mathbf{Q} \in C^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Nótese que:

$$R(\Omega, \Sigma, C, L) := \{\mathbf{Q} := (x_1, \dots, x_L) \in C^L : (\forall f \in \Omega, f(x_1) = \dots = f(x_L) = 0 \Rightarrow f \in \Sigma)\}$$

es un conjunto constructible ya que se puede expresar usando lógica de primer orden.

Hemos añadido una hipótesis (Hipótesis (3.2.1)) sobre la dimensión de las variedades definidas por listas de polinomios en $\Omega \setminus \Sigma$. Nótese que esto simplemente generaliza el caso $\Sigma = \{0\}$ y explica que el asunto principal de los conjuntos cuestores es la *dimensión* y no solo los Tests de Nulidad. Para conjuntos localmente cerrados $C \subseteq \mathbb{A}^n(K)$, denotaremos simplemente por $\deg(C)$ al grado de C (ver Subsección 1.2).

Antes de enunciar el teorema principal de esta sección, incluimos unas pocas palabras para explicar cómo medimos las probabilidades en la clase $R(\Omega, \Sigma, C, L)$ de conjuntos cuestores de longitud L en C para Ω con respecto a Σ con el fin de ayudar al lector de esta memoria.

Las variedades algebraicas proyectivas equidimensionales complejas $V \subseteq \mathbb{P}_n(\mathbb{C})$ tienen una distribución de probabilidad natural que ha sido ampliamente estudiada y empleada en la literatura matemática (cf. [BP, 07], [BP, 08], [BP, 09] y sus referencias). La clase $Reg(V) \subseteq V$ de puntos regulares de $V \subseteq \mathbb{P}_n(\mathbb{C})$ es una subvariedad Riemanniana compleja de $\mathbb{P}_n(\mathbb{C})$ de m -volumen finito (donde $m = \dim(V) = \dim(Reg(V))$). Por lo tanto, $Reg(V)$ está dotada de forma natural de una distribución de probabilidad, inducida por la estructura de subvariedad Riemanniana compleja de $\mathbb{P}_n(\mathbb{C})$. Como $Reg(V)$ es abierto y denso Zariski en V , esto se extiende de forma natural a V . En cuanto a las variedades afines complejas equidimensionales $W \subseteq \mathbb{A}^n(\mathbb{C})$, existe también una distribución de probabilidad natural inducida por la de su clausura proyectiva $\overline{W} \subseteq \mathbb{P}_n(\mathbb{C})$ (cf. [BP, 07], [BP, 08], [BP, 09] y sus referencias).

Desafortunadamente, no existe una estructura Riemanniana en $\mathbb{P}_n(K)$ cuando K es cualquier cuerpo algebraicamente cerrado y, por lo tanto, no hay una traducción directa de la distribución de probabilidad de variedades complejas a conjuntos constructibles de $\mathbb{A}^n(K)$. Sin embargo, hay una forma de interpretar la distribución de probabilidad inducida por la estructura Riemanniana a través de fórmulas tipo Poincaré (gran parte de este enfoque fue estudiado en la Sección 2 de [BP, 07]). A grandes rasgos, la probabilidad inducida por la métrica Riemanniana es la esperanza del número de puntos de la intersección de la variedad algebraica compleja $V \subseteq \mathbb{P}_n(\mathbb{C})$ con variedades proyectivas lineales A escogidas aleatoriamente en el Grassmaniano de las subvariedades de $\mathbb{P}_n(\mathbb{C})$ de co-dimensión igual a $\dim(V)$. Este hecho nos inspiró para estudiar probabilidades con respecto a la distribución uniforme de la forma:

$$\text{Prob}_{(C \cap A)^L} [R(\Omega, \Sigma, C, L)],$$

donde A pertenece al subconjunto abierto $\mathbb{G}(C)$ (cf. Corolarios 2.1.9 y 2.1.10 para esta notación) del Grassmaniano $\mathbb{G}(n, n-r)$ con respecto a la topología Zariski final, siendo r la co-dimensión de C .

TEOREMA 3.2.1. *Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}^K$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:*

$$(3.2.1) \quad \forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto localmente cerrado equidimensional de co-dimensión $r \geq (n-m) + m/2 + 1/2$. Supongamos que existen subconjuntos localmente cerrados $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ tales que $C := C_1 \cap \dots \cap C_s$. Sea $D := \max\{\deg(C_1), \dots, \deg(C_s)\} \geq 1$ el máximo de sus grados. Sea $L \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $L \geq 6 \dim(\Omega)$,
- ii) $\log(\deg(C)) \geq r \max\{2(1 + \log(d+1)), \frac{2 \log(\deg_{\text{ici}}(\Omega))}{\dim(\Omega)}\}$ y,
- iii) $D \leq (1 + \frac{1}{n-m}) \deg(C)^{\frac{1}{\text{codim}(C)}}$,

donde \log denota al logaritmo neperiano. Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, n-r)$, con respecto a la topología final inducida por la aplicación suprayectiva introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(C)$, la probabilidad de que una lista $\mathbf{Q} \in (C \cap A)^L$ escogida aleatoriamente esté en $R := R(\Omega, \Sigma, C, L)$ satisface:

$$\text{Prob}_{(C \cap A)^L} [R] \geq 1 - \frac{1}{\deg_{\text{ici}}(\Omega) e^{\dim(\Omega) + (m-1)L}},$$

donde $(C \cap A)^L$ está dotado de su distribución de probabilidad uniforme.

3.2.1. Demostración del Teorema 3.2.1. En primer lugar, denotamos por δ a la siguiente cantidad:

$$\delta := \deg(C)^{\frac{1}{\text{codim}(C)}}.$$

Obviamente, tenemos que $\deg(C) := \delta^r$ y nuestras hipótesis nos proporcionan las siguientes desigualdades:

- $\log(\delta) \geq \max\{2(1 + \log(d+1)), \frac{2 \log(\deg_{\text{ici}}(\Omega))}{\dim(\Omega)}\}$ y,
- $D \leq (1 + \frac{1}{n-m})\delta$.

A continuación, introducimos el siguiente conjunto constructible de incidencia:

$$V(\Omega, L) := \{(f, x_1, \dots, x_L) \in \Omega \times (\mathbb{A}^n(K))^L : f = (f_1, \dots, f_m) \in \mathcal{P}_{(d)}^K, f_i(x_j) = 0, 1 \leq i \leq m, 1 \leq j \leq L\}.$$

Consideramos también las siguientes dos proyecciones canónicas:

- La proyección sobre el conjunto de listas de ecuaciones consistentes: $\pi_1 : V(\Omega, L) \rightarrow \Omega$.
- La proyección sobre los posibles ceros: $\pi_2 : V(\Omega, L) \rightarrow (\mathbb{A}^n(K))^L$.

Por la Desigualdad de Bézout para conjuntos constructibles (ver Teorema 2.1.14), tenemos que:

$$\deg_{\text{lci}}(V(\Omega, L)) \leq \deg_{\text{lci}}(\Omega) \left(\prod_{i=1}^m (d_i + 1) \right)^L.$$

Sea \mathcal{D} la clase formada por conjuntos localmente cerrados irreducibles W_1, \dots, W_M tales que:

$$V(\Omega, L) = W_1 \cup \dots \cup W_M,$$

y que son mínimos con respecto al grado L de $V(\Omega, L)$:

$$\deg_{\text{lci}}(V(\Omega, L)) = \sum_{i=1}^M \deg(W_i).$$

Sea $\mathcal{C} \subseteq \mathcal{D}$ la clase de aquellas componentes localmente irreducibles de $V(\Omega, L)$ cuya proyección por π_1 no está completamente incluida en Σ :

$$\mathcal{C} := \{W \in \mathcal{D} : \pi_1(W) \setminus \Sigma \neq \emptyset\}.$$

Como π_1 es la restricción de una proyección, para todo $W \in \mathcal{C}$, podemos considerar la siguiente aplicación:

$$\pi_1|_{\overline{W}^z} := \overline{W}^z \longrightarrow \overline{\pi_1(W)}^z.$$

Se trata de un morfismo dominante entre dos variedades algebraicas irreducibles. Entonces, por el Teorema de la Dimensión en la Fibra, para todo $f \in \pi_1(W)$ tenemos que:

$$(3.2.2) \quad \dim \left(\pi_1|_{\overline{W}^z}^{-1}(\{f\}) \right) \geq \dim(\overline{W}^z) - \dim(\overline{\pi_1(W)}^z) \geq \dim(W) - \dim(\Omega).$$

Nótese que para cada $f \in \pi_1(W)$, se verifica la siguiente igualdad:

$$\pi_1|_{\overline{W}^z}^{-1}(\{f\}) = \{f\} \times (V_{\mathbb{A}}(f))^L.$$

Luego, para cada $W \in \mathcal{C}$, existe algún sistema de ecuaciones $f \in \pi_1(W) \setminus \Sigma$ y, por lo tanto, tenemos que:

$$\dim \left(\pi_1|_{\overline{W}^z}^{-1}(\{f\}) \right) = (n - m)L.$$

Por consiguiente, la Desigualdad (3.2.2) se convierte para cada $W \in \mathcal{C}$ en:

$$(3.2.3) \quad \dim(W) \leq \dim(\Omega) + (n - m)L.$$

Definimos ahora el conjunto constructible:

$$B(\Omega, L) := \bigcup_{W \in \mathcal{C}} W \subseteq V(\Omega, L).$$

Tenemos que:

- $\dim(B(\Omega, L)) = \max\{\dim(W) : W \in \mathcal{C}\} \leq (n - m)L + \dim(\Omega)$ y,
- por la sub-aditividad del grado para conjuntos constructibles (cf. Proposición 2.1.2), también se verifica:

$$\deg_{\text{lci}}(B(\Omega, L)) \leq \deg_{\text{lci}}(\Omega) \left(\prod_{i=1}^m (d_i + 1) \right)^L \leq \deg_{\text{lci}}(\Omega) (d + 1)^{mL},$$

donde $d := \max\{d_1, \dots, d_m\}$.

Al igual que en los Corolarios 2.1.9 y 2.1.10, existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, n - r)$, respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para todo $A \in \mathbb{G}(C)$ tenemos que:

$$\sharp(C \cap A) = \deg(C).$$

A continuación, consideramos $A \in \mathbb{G}(C)$ y el siguiente conjunto constructible:

$$\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L.$$

Aplicando la Proposición 2.1.11, concluimos:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \right) \leq \deg_{\text{lci}}(B(\Omega, L)) \leq \deg_{\text{lci}}(\Omega) (d + 1)^{mL}.$$

Además, como $C = C_1 \cap \dots \cap C_s$, podemos definir los siguientes conjuntos localmente cerrados:

$$C_{i,j} := \mathbb{A}^{n(j-1)}(K) \times C_i \times \mathbb{A}^{n(L-j)}(K),$$

cuyo grado es igual a $\deg(C_i)$. Definimos también la clase de variedades afines lineales A_j , $1 \leq j \leq L$, de grado 1, dadas por:

$$A_j := \mathbb{A}^{n(j-1)}(K) \times A \times \mathbb{A}^{n(L-j)}(K).$$

Entonces, tenemos que:

$$\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L = \overline{\pi_2(B(\Omega, L))^z} \cap \left(\bigcap_{i,j} C_{i,j} \right) \cap \left(\bigcap_{j=1}^L A_j \right).$$

Por la Proposición 2.2.2, concluimos también:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(B(\Omega, L)) D^{\dim(\pi_2(B(\Omega, L)))},$$

donde, como $\deg(A_j) = 1$, tenemos que:

$$1 \leq \max\{\deg(C_{i,j}) : 1 \leq i \leq s, 1 \leq j \leq L\} = D = \max\{\deg(C_i) : 1 \leq i \leq s\}.$$

Juntando todo lo anterior, como $\dim(\pi_2(B(\Omega, L))) \leq \dim(B(\Omega, L))$, obtenemos:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(\Omega) (d+1)^{mL} D^{\dim(\Omega) + (n-m)L}.$$

Esto implica:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(\Omega) D^{\dim(\Omega)} D^{(n-m)L} (d+1)^{mL}.$$

Como $\delta \geq e^2(d+1)^2$, tenemos que:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(\Omega) D^{\dim(\Omega)} D^{(n-m)L} \left(\frac{\delta}{e^2} \right)^{\frac{m}{2}L}.$$

Entonces, obtenemos:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) D^{\dim(\Omega)} D^{(n-m)L} \delta^{\frac{m}{2}L}}{e^{mL} \delta^{\text{codim}(C)L}}.$$

Como $\text{codim}(C) \geq (n-m) + m/2 + 1/2$, también tenemos que:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) D^{\dim(\Omega)} D^{(n-m)L}}{e^{mL} \delta^{L/2}} \frac{D^{(n-m)L}}{\delta^{(n-m)L}}.$$

Como $D \leq (1 + \frac{1}{n-m})\delta$, obtenemos:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) D^{\dim(\Omega)}}{e^{mL} \delta^{L/2}} \left(1 + \frac{1}{n-m} \right)^{(n-m)L}.$$

Por lo tanto,

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) D^{\dim(\Omega)}}{e^{mL} \delta^{L/2}} \left(\left(1 + \frac{1}{n-m} \right)^{(n-m)} \right)^L.$$

Por consiguiente,

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) D^{\dim(\Omega)}}{\delta^{L/2}} \left(\frac{1}{e} \right)^{(m-1)L}.$$

De nuevo, como $D \leq (1 + \frac{1}{n-m})\delta$ y $L \geq 6 \dim(\Omega)$, también tenemos que:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega)}{\delta^{\dim(\Omega)}} \frac{1}{\delta^{\dim(\Omega)}} \left(\frac{D}{\delta} \right)^{\dim(\Omega)} \left(\frac{1}{e} \right)^{(m-1)L}.$$

Es decir,

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega)}{\delta^{\dim(\Omega)}} \frac{1}{\delta^{\dim(\Omega)}} e^{\frac{\dim(\Omega)}{n-m}} \left(\frac{1}{e} \right)^{(m-1)L}.$$

Como $\delta \geq e^2(d+1)^2$, esto implica:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega)}{\delta^{\dim(\Omega)}} \left(\frac{1}{e} \right)^{\dim(\Omega)} \left(\frac{1}{e} \right)^{(m-1)L}.$$

Finalmente, como $\log(\delta) \dim(\Omega) \geq 2 \log(\deg_{\text{lci}}(\Omega))$, también obtenemos:

$$(3.2.4) \quad \frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg(C)^L} \leq \frac{1}{\deg_{\text{lci}}(\Omega)} \left(\frac{1}{e} \right)^{\dim(\Omega)} \left(\frac{1}{e} \right)^{(m-1)L}.$$

A continuación, consideramos el siguiente conjunto constructible:

$$R^c(\Omega, \Sigma, C, L) = \{\mathbf{Q} := (x_1, \dots, x_L) \in C^L : \mathbf{Q} \text{ no es un conjunto cwestor para } \Omega \text{ con respecto a } \Sigma\}.$$

Es obvio que:

$$R^c(\Omega, \Sigma, C, L) = (C)^L \setminus R(\Omega, \Sigma, C, L).$$

AFIRMACIÓN 3.2.1.1. *Con las notaciones precedentes, si $A \in \mathbb{G}(C)$, tenemos que:*

$$\#(R^c(\Omega, \Sigma, C, L) \cap A^L) \leq \deg\left(\overline{\pi_2(B(\Omega, L))}^z \cap (C \cap A)^L\right).$$

Demostración de la Afirmación. En primer lugar, nótese que:

$$R^c(\overline{\Omega}^z, \Sigma, C, L) \subseteq \pi_2(B(\Omega, L)) \cap C^L.$$

Si $\mathbf{Q} := (x_1, \dots, x_L) \in R^c(\Omega, \Sigma, C, L)$, entonces existe una lista de ecuaciones polinomiales $f \in \Omega \setminus \Sigma$ tal que:

$$f(x_1) = \dots = f(x_L) = 0.$$

Por lo tanto, $(f, x_1, \dots, x_L) \in V(\Omega, L)$. Además, debe haber alguna componente localmente irreducible $W \in \mathcal{D}$ de $V(\Omega, L)$ que contiene a (f, x_1, \dots, x_L) . Sin embargo, $f := \pi_1(f, x_1, \dots, x_L) \in \Omega \setminus \Sigma$ y, por lo tanto, $W \in \mathcal{C}$. Entonces, $(f, x_1, \dots, x_L) \in B(\Omega, L)$ y

$$\mathbf{Q} := (x_1, \dots, x_L) \in \pi_2(B(\Omega, L)).$$

Por consiguiente, concluimos:

$$R^c(\overline{\Omega}^z, \Sigma, C, L) \subseteq \pi_2(B(\Omega, L)) \cap C^L \subseteq \overline{\pi_2(B(\Omega, L))}^z \cap C^L,$$

y

$$R^c(\overline{\Omega}^z, \Sigma, C, L) \cap A^L \subseteq \overline{\pi_2(B(\Omega, L))}^z \cap (C \cap A)^L.$$

Como $A \in \mathbb{G}(C)$, entonces $C \cap A$ es una variedad algebraica cero-dimensional (i.e. un conjunto finito) y, por lo tanto, $R^c(\overline{\Omega}^z, \Sigma, C, L) \cap A^L$ es también un conjunto finito. Finalmente, como en el caso cero-dimensional el grado es igual a la cardinalidad (cf. Proposición 1.2.3), tenemos que la afirmación del enunciado es cierta:

$$\#(R^c(\Omega, \Sigma, C, L) \cap A^L) \leq \deg\left(\overline{\pi_2(B(\Omega, L))}^z \cap (C \cap A)^L\right).$$

■

Para finalizar, como para cada $A \in \mathbb{G}(C)$, $\#(C \cap A) = \deg(C)$, entonces obtenemos:

$$\text{Prob}_{(C \cap A)^L}[R] = 1 - \text{Prob}_{(C \cap A)^L}[R^c],$$

donde $R^c := R^c(\Omega, \Sigma, C, L)$. Es decir, tenemos que:

$$\text{Prob}_{(C \cap A)^L}[R] = 1 - \frac{1}{\#((C \cap A)^L)} \sum_{\zeta \in (C \cap A)^L} \chi_{R^c}(\zeta) = 1 - \frac{\#(R^c \cap (C \cap A)^L)}{\#((C \cap A)^L)},$$

donde χ_{R^c} es la función característica de R^c . Entonces, por la afirmación previa y la Desigualdad (3.2.4), concluimos:

$$\text{Prob}_{(C \cap A)^L}[R] \geq 1 - \frac{1}{\deg_{\text{Ici}}(\Omega)e^{\dim(\Omega)+(m-1)L}},$$

y se deduce el teorema.

3.2.2. Algunos Corolarios útiles. Tal y como mencionamos al comienzo de esta sección, nuestros resultados sobre distribuciones de probabilidad tienen una forma simplificada en el caso complejo $K = \mathbb{C}$. Recordemos que el Grassmaniano complejo está dotado de una distribución de probabilidad natural μ . Para cada conjunto de Borel $B \subseteq \mathbb{G}(n, n-r)$ con respecto a su topología Zariski final, la medida $\mu[B]$ verifica:

$$(3.2.5) \quad \mu[B] := \begin{cases} 1, & B \text{ tiene interior no vacío en } \mathbb{G}(n, r) \\ 0, & \text{en otro caso} \end{cases}.$$

Entonces, con esta medida de probabilidad, el Teorema 3.2.1 anterior se convierte en:

COROLARIO 3.2.2. *Supongamos que $K = \mathbb{C}$ y que μ es una medida de probabilidad sobre los conjuntos de Borel $\mathcal{B}(\mathbb{G}(n, n-r))$ que satisface la Identidad (3.2.5). Con las mismas notaciones e hipótesis que en el Teorema 3.2.1, se verifica la siguiente desigualdad:*

$$\frac{1}{\deg_{\text{Ici}}(C)^L} E_{\mathbb{G}(n, n-r)}[\#_R^{(L)}] \geq 1 - \frac{1}{\deg_{\text{Ici}}(\Omega)e^{\dim(\Omega)+(m-1)L}},$$

donde $E_{\mathbb{G}(n,n-r)}$ denota a la esperanza y

$$\begin{aligned} \sharp_R^{(L)} : \mathbb{G}(n,n-r) &\longrightarrow \mathbb{R}_+ \cup \{\infty\} \\ A &\longmapsto \sharp(R(\Omega, \Sigma, C, L) \cap A^L). \end{aligned}$$

DEMOSTRACIÓN. En primer lugar, nótese que con la hipótesis de la Identidad (3.2.5) tenemos que:

$$E_{\mathbb{G}(n,n-r)}[\sharp_R^{(L)}] = E_{\mathbb{G}(C)}[\sharp_R^{(L)}],$$

donde $\mathbb{G}(C)$ es el subconjunto abierto Zariski de $\mathbb{G}(n,n-r)$ discutido en la demostración del Teorema 3.2.1 precedente. Basta aplicar el Teorema 3.2.1 al cociente:

$$\frac{\sharp_R^{(L)}(A)}{\deg_{\text{lci}}(C)^L} = \text{Prob}_{(A \cap C)^L}[R],$$

para obtener el resultado buscado. \square

Con una ligera modificación de las hipótesis del Teorema 3.2.1, obtenemos el siguiente resultado para el caso en que C es un conjunto constructible presentable equidimensionalmente:

COROLARIO 3.2.3. Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}^K$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:

$$(3.2.6) \quad \forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible presentable equidimensionalmente de co-dimensión $r \geq (n - m) + m/2 + 1/2$. Supongamos que existen subconjuntos constructibles $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ tales que $C := C_1 \cap \dots \cap C_s$. Sea $D := \max\{\deg_{\text{lci}}(C_1), \dots, \deg_{\text{lci}}(C_s)\} \geq 1$ el máximo de sus grados. Sea $L \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguiente propiedades:

- i) $L \geq 6 \dim(\Omega)$,
- ii) $\log(\deg_z(C)) \geq r \max\{2(1 + \log(d + 1)), \frac{2 \log(\deg_{\text{lci}}(\Omega)) + sL}{\dim(\Omega)}\}$ y,
- iii) $D \leq \frac{1}{e} (1 + \frac{1}{n-m}) \deg_z(C)^{\frac{1}{\text{codim}(C)}}$.

Sea $R := R(\Omega, \Sigma, C, L)$ el conjunto constructible de todas las listas $\mathbf{Q} \in C^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, n-r)$, con respecto a la topología final inducida por la aplicación suprayectiva introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(C)$, la probabilidad de que una lista $\mathbf{Q} \in (C \cap A)^L$ escogida aleatoriamente esté en R satisfice:

$$\text{Prob}_{(C \cap A)^L}[R] \geq 1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{\dim(\Omega) + (m-1)L}},$$

donde $(C \cap A)^L$ está dotado de su distribución de probabilidad uniforme.

DEMOSTRACIÓN. Con las mismas notaciones que en la demostración del Teorema 3.2.1, tenemos que:

$$\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L = \overline{\pi_2(B(\Omega, L))^z} \cap \left(\bigcap_{i,j} C_{i,j} \right) \cap \left(\bigcap_{j=1}^L A_j \right),$$

donde los $C_{i,j}$'s son conjuntos constructibles. Del Teorema 2.2.1, obtenemos:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \left((sL + 1) + \dim(\pi_2(B(\Omega, L))) - 1 \right) \deg_{\text{lci}}(B(\Omega, L)) D^{\dim(\pi_2(B(\Omega, L)))}.$$

donde, como $\deg(A_j) = 1$, tenemos que:

$$1 \leq \max\{\deg_{\text{lci}}(C_{i,j}) : 1 \leq i \leq s, 1 \leq j \leq L\} = D = \max\{\deg_{\text{lci}}(C_i) : 1 \leq i \leq s\}.$$

Nótese que si $a, b \in \mathbb{Z}^+$, se verifica:

$$\binom{a+b}{b} \leq e^{a+b}$$

Por lo tanto, juntando todo lo anterior, como:

$$\dim(\pi_2(B(\Omega, L))) \leq \dim(B(\Omega, L)) \leq (n-m)L + \dim(\Omega),$$

obtenemos:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(\Omega) (d+1)^{mL} D^{\dim(\Omega) + (n-m)L} e^{sL + \dim(\Omega) + (n-m)L}.$$

Esto implica:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)} (De)^{(n-m)L} (d+1)^{mL} e^{sL}.$$

Como $\delta \geq e^2(d+1)^2$, tenemos que:

$$\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right) \leq \deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)} (De)^{(n-m)L} \left(\frac{\delta}{e^2} \right)^{\frac{m}{2}L} e^{sL}.$$

Entonces, obtenemos:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)} (De)^{(n-m)L} \delta^{\frac{m}{2}L} e^{sL}}{e^{mL} \delta^{\text{codim}(C)L}}.$$

Como $\text{codim}(C) \geq (n-m) + m/2 + 1/2$, también tenemos que:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)} (De)^{(n-m)L}}{e^{mL} \delta^{L/2}} \frac{(De)^{(n-m)L}}{\delta^{(n-m)L}} e^{sL}.$$

Como $D \leq \frac{1}{e} \left(1 + \frac{1}{n-m}\right) \delta$, obtenemos:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)}}{e^{mL} \delta^{L/2}} \left(1 + \frac{1}{n-m}\right)^{(n-m)L} e^{sL}.$$

Por consiguiente,

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)}}{e^{mL} \delta^{L/2}} \left(\left(1 + \frac{1}{n-m}\right)^{(n-m)} \right)^L e^{sL}.$$

Y, por lo tanto,

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega) (De)^{\dim(\Omega)}}{\delta^{L/2}} \left(\frac{1}{e} \right)^{(m-1)L} e^{sL}.$$

De nuevo, como $D \leq \frac{1}{e} \left(1 + \frac{1}{n-m}\right) \delta$ y $L \geq 6 \dim(\Omega)$, también tenemos que:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega)}{\delta^{\dim(\Omega)}} \frac{1}{\delta^{\dim(\Omega)}} \left(\frac{De}{\delta} \right)^{\dim(\Omega)} \left(\frac{1}{e} \right)^{(m-1)L} e^{sL}.$$

Es decir,

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega)}{\delta^{\dim(\Omega)}} \frac{1}{\delta^{\dim(\Omega)}} e^{\frac{\dim(\Omega)}{n-m}} \left(\frac{1}{e} \right)^{(m-1)L} e^{sL}.$$

Como $\delta \geq e^2(d+1)^2$, esto implica:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{\deg_{\text{lci}}(\Omega)}{\delta^{\dim(\Omega)}} \left(\frac{1}{e} \right)^{\dim(\Omega)} \left(\frac{1}{e} \right)^{(m-1)L} e^{sL}.$$

Finalmente, como $\log(\delta) \dim(\Omega) \geq 2 \log(\deg_{\text{lci}}(\Omega)) + sL$, también obtenemos:

$$\frac{\deg \left(\overline{\pi_2(B(\Omega, L))^z} \cap (C \cap A)^L \right)}{\deg_z(C)^L} \leq \frac{1}{\deg_{\text{lci}}(\Omega)} \left(\frac{1}{e} \right)^{\dim(\Omega)} \left(\frac{1}{e} \right)^{(m-1)L}.$$

La parte final de la demostración es idéntica a la del Teorema 3.2.1. \square

Usando el Corolario 2.1.10, podemos generalizar el corolario previo a cualquier conjunto constructible de la siguiente forma:

COROLARIO 3.2.4. Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}^K$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:

$$(3.2.7) \quad \forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible de co-dimensión $r \geq (n-m) + m/2 + 1/2$ y sean V_1, \dots, V_k las componentes irreducibles de dimensión más alta de la clausura Zariski de C (como en la Proposición 1.1.5). Supongamos que existen subconjuntos constructibles $C_1, \dots, C_s \subseteq \mathbb{A}^n(K)$ tales que $C := C_1 \cap \dots \cap C_s$. Sea

$D := \max\{\deg_{\text{lci}}(C_1), \dots, \deg_{\text{lci}}(C_s)\} \geq 1$ el máximo de sus grados. Sea $L \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $L \geq 6 \dim(\Omega)$,
- ii) $\log(\sum_{i=1}^k \deg(V_i)) \geq r \max\{2(1 + \log(d+1)), \frac{2 \log(\deg_{\text{lci}}(\Omega)) + sL}{\dim(\Omega)}\}$ y,
- iii) $D \leq \frac{1}{e} (1 + \frac{1}{n-m}) (\sum_{i=1}^k \deg(V_i))^{\frac{1}{\text{codim}(C)}}$.

Sea $R := R(\Omega, \Sigma, C, L)$ el conjunto constructible de todas las listas $\mathbf{Q} \in C^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, n-r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(C)$, la probabilidad de que una lista $\mathbf{Q} \in (C \cap A)^L$ escogida aleatoriamente esté en R satisfice:

$$\text{Prob}_{(C \cap A)^L}[R] \geq 1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{\dim(\Omega) + (m-1)L}},$$

donde $(C \cap A)^L$ está dotado de su distribución de probabilidad uniforme.

DEMOSTRACIÓN. La demostración es similar a la del Corolario 3.2.3. En este caso, tomamos:

$$\delta := \left(\sum_{i=1}^k \deg(V_i) \right)^{\frac{1}{\text{codim}(C)}},$$

y, por el Corolario 2.1.10, tenemos que existe un subconjunto abierto no vacío $\mathbb{G}(C) \subseteq \mathbb{G}(n, n-r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para todo $A \in \mathbb{G}(C)$, tenemos que:

$$\sharp(C \cap A) = \sum_{i=1}^k \deg(V_i).$$

□

Si suponemos que C es una variedad algebraica de intersección completa, obtenemos el siguiente resultado:

COROLARIO 3.2.5. Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}(X_1, \dots, X_n)$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:

$$(3.2.8) \quad \forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $V := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ una variedad algebraica de intersección completa de co-dimensión $r \geq (n-m) + m/2 + 1/2$ tal que $\deg(V) \geq \delta^r$, donde $\delta := \min\{\deg(h_1), \dots, \deg(h_r)\}$. Sea $L \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $L \geq 6 \dim(\Omega)$,
- ii) $\log(\delta) \geq \max\{2(1 + \log(d+1)), \frac{2 \log(\deg_{\text{lci}}(\Omega))}{\dim(\Omega)}\}$ y,
- iii) $\max\{\deg(h_1), \dots, \deg(h_r)\} \leq (1 + \frac{1}{n-m})\delta$.

Sea $R := R(\Omega, \Sigma, V, L)$ el conjunto constructible de todas las listas $\mathbf{Q} \in V^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(V) \subseteq \mathbb{G}(n, n-r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(V)$, la probabilidad de que una lista $\mathbf{Q} \in (V \cap A)^L$ escogida aleatoriamente esté en R satisfice:

$$\text{Prob}_{(V \cap A)^L}[R] \geq 1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{\dim(\Omega) + (m-1)L}},$$

donde $(V \cap A)^L$ está dotado de su distribución de probabilidad uniforme.

DEMOSTRACIÓN. Es una aplicación directa del Teorema 3.2.1. Tomando $C = V$, $s = r$ y $C_i := V_{\mathbb{A}}(h_i)$ para cada i , $1 \leq i \leq s$, obtenemos el resultado. □

Podemos enunciar el corolario anterior para el caso en el que busquemos conjuntos cuestores en variedades algebraicas cero-dimensionales:

COROLARIO 3.2.6. Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}(X_1, \dots, X_n)$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:

$$\forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $V := V_{\mathbb{A}}(h_1, \dots, h_n) \subseteq \mathbb{A}^n(K)$ una variedad algebraica cero-dimensional dada por ecuaciones polinomiales del mismo grado $\delta := \deg(h_i)$, $1 \leq i \leq n$. Supongamos que $\deg(V) = \delta^n$. Sea L un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $L \geq 6 \dim(\Omega)$ y,
- ii) $\log(\delta) \geq \max\{2(1 + \log(d+1)), \frac{2 \log(\deg_{\text{lci}}(\Omega))}{\dim(\Omega)}\}$.

Sea $R := R(\Omega, \Sigma, V, L)$ el conjunto constructible de todas las listas $\mathbf{Q} \in V^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Supongamos que V está dotado de su distribución de probabilidad uniforme. Entonces, tenemos que:

$$\text{Prob}_{V^L}[R] \geq 1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{\dim(\Omega) + (m-1)L}}.$$

3.2.3. Conjuntos constructibles unirracionales. Un contexto donde se utilizan habitualmente los conjuntos cuestores es el de conjuntos constructibles unirracionales, esto es, conjuntos constructibles dados como imágenes polinomiales de algún espacio afín (i.e. el espacio de parámetros). A continuación, discutimos nuestros resultados para este contexto.

COROLARIO 3.2.7 (Familias unirracionales de listas de polinomios). Con las notaciones anteriores, sea $(d) := (d_1, \dots, d_m)$ una lista de grados, $d := \max\{d_1, \dots, d_m\}$ y $\Omega \subseteq \mathcal{P}_{(d)}$ un conjunto constructible unirracional, dado como la imagen de una función polinomial:

$$\varphi := (\varphi_1, \dots, \varphi_N) : \Lambda \subseteq \mathbb{A}^M \longrightarrow \Omega \subseteq \mathcal{P}_{(d)},$$

donde $N := N_{(d)} = \dim(\mathcal{P}_{(d)})$, Λ es un conjunto constructible de dimensión s y $\Omega = \varphi(\Lambda)$. Sea $t := s / \dim(\Omega) \geq 1$ el cociente entre las dimensiones de Λ y Ω y sea $D := \max\{\deg(\varphi_i) : 1 \leq i \leq N\}$ el máximo de los grados de las coordenadas de φ . Sea $\Sigma \subseteq \Omega$ un conjunto constructible de co-dimensión al menos 1 en Ω y supongamos que se satisface la siguiente propiedad:

$$\forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $V := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ una variedad algebraica de intersección completa de co-dimensión $r \geq (n - m) + m/2 + 1/2$ tal que $\deg(V) \geq \delta^r$, donde $\delta := \min\{\deg(h_1), \dots, \deg(h_r)\}$. Sea $L \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $L \geq 6s$,
- ii) $\log(\delta) \geq \max\{2(1 + \log(d+1)), 2t \left(\frac{\log(\deg_{\text{lci}}(\Lambda))}{s} + \log(D) \right)\}$ y,
- iii) $\max\{\deg(h_1), \dots, \deg(h_r)\} \leq (1 + \frac{1}{n-m})\delta$.

Sea $R := R(\Omega, \Sigma, V, L)$ el conjunto constructible de todas las listas $\mathbf{Q} \in V^L$ de longitud L que son conjuntos cuestores para Ω con respecto a Σ . Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(V) \subseteq \mathbb{G}(n, n - r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(V)$, la probabilidad de que una lista $\mathbf{Q} \in (V \cap A)^L$ escogida aleatoriamente esté en R satisfice:

$$\text{Prob}_{(V \cap A)^L}[R] \geq 1 - \frac{1}{\deg_{\text{lci}}(\Lambda) e^{s(\log(D)+1) + (m-1)L}},$$

donde $(V \cap A)^L$ está dotado de su distribución de probabilidad uniforme.

DEMOSTRACIÓN. La demostración es similar a la del Teorema 3.2.1. Basta notar que los conjuntos cuestores son hereditarios (cf. Proposición 3.1.3) y que, por el Corolario 2.2.4, se verifica la siguiente igualdad:

$$\deg_z(\Omega) = \deg(\overline{\varphi(\Lambda)^z}) \leq \deg_{\text{lci}}(\Lambda) D^s.$$

Por lo tanto, obtenemos:

$$\frac{\deg(\overline{\pi_2(B(\overline{\Omega}^z, L))^z} \cap (V \cap A)^L)}{\deg(V)^L} \leq \frac{1}{\deg_{\text{lci}}(\Lambda) e^{s(\log(D)+1) + (m-1)L}},$$

y se concluye el resultado. \square

OBSERVACIÓN 3.2.8. Nótese que si Λ es una subvariedad afín lineal y la dimensión no disminuye demasiado a través de φ (por ejemplo, $t := \frac{s}{\dim(\Omega)} \leq 2$), las hipótesis del Corolario 3.2.7 anterior se convierten en:

- i) $L \geq 6s$,
- ii) $\log(\delta) \geq \max\{2(1 + \log(d+1)), 4(\log(D))\}$ y,
- iii) $\max\{\deg(h_1), \dots, \deg(h_r)\} \leq (1 + \frac{1}{n-m})\delta$,

y obtenemos la misma cota inferior:

$$\text{Prob}_{(V \cap A)^L} [R] \geq 1 - \frac{1}{e^{s(\log(D)+1)+(m-1)L}}.$$

3.2.4. El conjunto de muestreo de DeMillo–Lipton–Schwartz–Zippel. Sea κ un cuerpo y K su clausura algebraica. El Corolario 3.2.6 nos permite demostrar que, bajo ciertas condiciones en la cardinalidad del conjunto $Q \subseteq \kappa$, los conjuntos cuestores de longitud asintóticamente óptima para P_d^K están altamente distribuidos en el conjunto de muestreo Q^n :

COROLARIO 3.2.9. Sean $n, d \in \mathbb{N}$ dos enteros positivos. Sea Ω un subconjunto constructible de P_d^K . Supongamos que $\Omega \setminus \{0\} \neq \emptyset$. Sea $Q \subseteq \kappa$ un conjunto finito. Sea $L \in \mathbb{N}$ un entero positivo tal que se verifican las siguientes propiedades:

- i) $L \geq 6 \dim(\Omega)$ y,
- ii) $\sharp(Q) \geq \max\{(e(d+1))^2, \deg_{\text{lci}}(\Omega)^{\frac{2}{\dim(\Omega)}}\}$.

Entonces, la probabilidad de que una lista $\mathbf{Q} \in V^L = (Q^n)^L$ escogida aleatoriamente sea un conjunto cuestor para Ω es mayor que:

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{\dim(\Omega)}}.$$

En el caso de entrada denso (i.e. cuando $\Omega = P_d^K$), para cualquier $L \geq 6 \dim(P_d^K)$ y $Q \subseteq \kappa$ tal que:

$$\sharp(Q) \geq (e(d+1))^2,$$

la probabilidad de que una lista $\mathbf{Q} \in (Q^n)^L$ escogida aleatoriamente sea un conjunto cuestor para P_d^K es mayor que:

$$1 - \frac{1}{e^{\frac{d+n}{n}}}.$$

Retomemos ahora el Corolario 2.2.9. Como aplicación concreta de este corolario, consideremos un subconjunto finito $Q \subseteq \kappa$, un polinomio no nulo $f \in P_d^K$ y la hipersuperficie $C := V_{\mathbb{A}}(f) \subseteq \mathbb{A}^n(K)$. Al igual que en [DML, 78], [Zp, 79] y [Sch, 80], hemos demostrado que la probabilidad de que f no se anule en un punto aleatorio $x \in Q^n$ es mayor que:

$$1 - \frac{d}{\sharp(Q)}.$$

Tomando Q con $\sharp(Q) \geq (e(d+1))^2$, el Corolario 3.2.9 anterior nos permite obtener un algoritmo probabilista para Tests de Nulidad de Polinomios en el caso de entrada denso. Este resultado es la base de los Tests probabilísticos de DeMillo–Lipton–Zippel–Schwartz, ya que muestra que la probabilidad de detectar la no nulidad al evaluar un polinomio en muestras aleatorias de Q^n es muy alta. La clave es que estos tests emplean conjuntos de muestreo lo suficientemente grandes como para contener, con alta densidad, conjuntos cuestores de longitud asintóticamente óptima.

3.3. Algunos Conjuntos Cuestores en la literatura matemática

En esta sección, presentamos y analizamos en detalle algunos ejemplos de conjuntos cuestores que aparecen en la literatura matemática.

3.3.1. Conjuntos Cuestores y variedades determinantes. Sean $m, n \in \mathbb{N}$ dos enteros positivos con $m \geq n$. Consideramos las variedades determinantes en el espacio $\mathcal{M}_{m \times n}(K)$ de matrices $m \times n$ con entradas en un cuerpo algebraicamente cerrado K :

$$\Sigma_r := \{M \in \mathcal{M}_{m \times n}(K) : \text{corango}(M) \geq r\}.$$

Las propiedades de las variedades determinantes pueden consultarse en [Ar, 74] o [Ha, 92]

Es bien conocido que Σ_r es una variedad algebraica definida por ecuaciones polinomiales homogéneas (y, por lo tanto, es también una variedad proyectiva), y su co-dimensión en $\Omega := \mathcal{M}_{m \times n}(K)$ viene dada por la siguiente identidad:

$$\text{codim}_{\Omega}(\Sigma_r) = \dim(\mathcal{M}_{m \times n}(K)) - \dim(\Sigma_r) = (m-r)(n-r).$$

Entonces, para cada conjunto cuestor $\mathbf{Q} := (x_1, \dots, x_L) \in \mathbb{A}^{nL}(K)$ para $\Omega := \mathcal{M}_{m \times n}(K)$ con respecto a Σ_r , tenemos que (ver Proposición 3.1.12):

$$L := \sharp(\mathbf{Q}) \geq (m-r)(n-r).$$

Además, en el Teorema 3.2.1 hemos demostrado que existen conjuntos cuestores para Ω con respecto a Σ_r de longitud $O(mn)$.

Análogamente, para $r < s$, tenemos que:

$$\text{codim}_{\Sigma_r}(\Sigma_s) = (r - s)(m + n - (r + s)).$$

Por lo tanto, para cada conjunto cuestor $\mathbf{Q} \subseteq \mathbb{A}^{nL}(K)$ para Σ_r con respecto a Σ_s , tenemos que:

$$\sharp(\mathbf{Q}) \geq (r - s)(m + n - (r + s)).$$

3.3.2. Conjuntos Cuestores y Conjuntos de Kakeya. La noción de conjuntos de Kakeya en el caso de cuerpos finitos se debe a [Wo, 99]. En [Dv, 09] se estableció una cota inferior exponencial para la cardinalidad de los conjuntos de Kakeya en el caso de cuerpos finitos. El lector podrá encontrar en [Tao, 14] otros resultados relacionados.

En [Tao, 14], T. Tao afirma: “*There is no known proof of the finite field Kakeya conjecture that does not go through the polynomial method*”. En esta subsección, explicamos este fenómeno observando que los *conjuntos de Kakeya son conjuntos cuestores para ciertos conjuntos constructibles de polinomios*. De hecho, la mayoría de los métodos utilizados en lo que se conoce como el “Método Polinomial” consisten simplemente en emplear un conjunto cuestor adaptado a alguna clase (típicamente una clase dada como un conjunto constructible) de polinomios. Sin embargo, probamos que el recíproco es falso: en el Corolario 3.3.3 demostramos que para un ε pequeño y positivo y grados menores que $q^{1-\varepsilon} - 1$, la mayoría de los conjuntos cuestores no son conjuntos de Kakeya. Por otro lado, no conocemos ningún conjunto cuestor para polinomios de grado $q - 1$ que no sea un conjunto de Kakeya.

En esta subsección, demostramos una generalización del resultado principal de [Dv, 09] usando la terminología de los conjuntos cuestores. Las notaciones son las mismas que en las secciones precedentes. Para un punto proyectivo $v \in \mathbb{P}_{n-1}(K)$, denotamos por $K\langle v \rangle \subseteq K^n$ al subespacio vectorial de dimensión 1 de K^n generado por cualquier representante no nulo $\tilde{v} \in K^n \setminus \{0\}$ del punto proyectivo $v \in \mathbb{P}_{n-1}(K)$. En primer lugar, “generalizamos” la noción de conjunto de Kakeya a su propio contexto natural.

DEFINICIÓN 18 (q -Conjunto de Kakeya con direcciones en una variedad proyectiva cero-dimensional). Con las notaciones anteriores, sea $V \subseteq \mathbb{P}_{n-1}(K)$ una variedad proyectiva cero-dimensional y $q \in \mathbb{N}$ un entero positivo. Un q -conjunto de Kakeya con direcciones en V es un subconjunto finito $E \subseteq \mathbb{A}^n(K)$ tal que se verifica la siguiente propiedad:

Para cada dirección $v \in V$, existe un punto $x \in E$ tal que la recta afín $\rho(x, v) := x + K\langle v \rangle$ satisface:

$$\sharp(E \cap \rho(x, v)) \geq q.$$

Cuando $\kappa = \mathbb{F}_q$ es un cuerpo finito con q elementos, denotamos por $\mathbb{F}_q\langle v \rangle$ al \mathbb{F}_q -espacio vectorial generado por v . Los conjuntos de Kakeya “usuales” son entonces definidos como:

DEFINICIÓN 19 (Conjuntos de Kakeya “usuales” sobre cuerpos finitos). Sea \mathbb{F}_q un cuerpo finito de cardinalidad q . Un subconjunto finito $E \subseteq \mathbb{A}^n(\mathbb{F}_q)$ es un conjunto de Kakeya si se verifica la siguiente propiedad: Para cada punto proyectivo $v \in \mathbb{P}_{n-1}(\mathbb{F}_q)$, existe un punto $x \in E$ tal que el conjunto de puntos \mathbb{F}_q -racionales de la recta afín $\rho(x, v) \subseteq \mathbb{A}^n(\mathbb{F}_q)$ determinada por x y v está totalmente contenido en E . Es decir, se cumple lo siguiente:

$$\rho(x, v) := x + \mathbb{F}_q\langle v \rangle := \{x + tv : t \in \mathbb{F}_q\} \subseteq E.$$

Recordemos que la función de Hilbert de una variedad proyectiva $V \subseteq \mathbb{P}_{n-1}(K)$ es una función $\chi_V : \mathbb{N} \rightarrow \mathbb{N}$ dada por la siguiente identidad para todo $m \in \mathbb{N}$:

$$\chi_V(m) := \dim_K(H_m(X_1, \dots, X_n)/I_m(V)) = \dim_K(H_m(X_1, \dots, X_n)) - \dim_K(I_m(V)),$$

donde:

- \dim_K es la dimensión como K -espacios vectoriales,
- $H_m(X_1, \dots, X_n)$ es el espacio vectorial generado por todos los polinomios homogéneos de grado m en el conjunto de variables $\{X_1, \dots, X_n\}$ con coeficientes en K y,
- $I_m(V)$ es el subespacio vectorial de los polinomios en $H_m(X_1, \dots, X_n)$ que se anulan en los puntos proyectivos de V .

La función de Hilbert es una función polinómica de grado igual a la dimensión de Krull de V como variedad proyectiva. Además, existe un único polinomio $h \in \mathbb{Z}[T]$ de grado igual a $\dim(V)$ y un entero R tal que:

$$\chi_V(m) = h(m), \quad \forall m \geq R.$$

El polinomio h se llama polinomio de Hilbert de V y el mínimo de esos R tal que χ_V y h coinciden recibe el nombre de *regularidad de la función de Hilbert de V* .

El siguiente resultado prueba que los q -conjuntos de Kakeya son conjuntos cuestores y, por lo tanto, tenemos una cota inferior de su cardinalidad basada en la maldición de la dimensionalidad de la Proposición 3.1.12.

PROPOSICIÓN 3.3.1. *Con las notaciones e hipótesis precedentes, dada una variedad proyectiva cero-dimensional $V \subseteq \mathbb{P}_{n-1}(K)$, para cada q -conjunto de Kakeya E con direcciones en V tenemos que:*

$$\sharp(E) \geq \chi_V(d), \quad \forall d \in \mathbb{N}, \quad d \leq q-1.$$

Además, si R es la regularidad de la función de Hilbert de V y $q-1 \geq R$, se cumple que:

$$\sharp(E) \geq \sharp(V) = \deg(V).$$

Por otro lado, los q -conjuntos de Kakeya con direcciones en $V \subseteq \mathbb{P}_{n-1}(K)$ con cardinalidad $(q-1)\deg(V) + 1$ existen.

DEMOSTRACIÓN. Sea $\mathbb{A}^N(K)$ el espacio afín sobre K dado como el K -espacio vectorial P_d de todos los polinomios de grado a lo sumo d con coeficientes en K , $\mathbb{A}^N(K) := P_d$, donde $N := N_d$. Para cada polinomio $f \in P_d(X_1, \dots, X_n)$, consideramos su descomposición en componentes homogéneas:

$$f = f_m + \dots + f_0,$$

donde $f_i \in H_i(X_1, \dots, X_n)$ y $f_m \neq 0$.

Denotamos por $LHC(f) \in P_d$ (“leading homogeneous component”) a la componente homogénea de mayor grado de f , $LHC(f) = f_m$. Para cada m , $0 \leq m \leq d$, definimos los siguientes subconjuntos constructibles de P_d :

$$\begin{aligned} \Omega_m &:= \{f \in P_d : \deg(f) = m\}, \\ \Sigma_m &:= \{f \in \Omega_m : LHC(f) \in I_m(V)\}, \end{aligned}$$

donde $I_m(V)$ es el espacio vectorial generado por todos los polinomios homogéneos de grado m que se anulan en V . Finalmente, definimos:

$$\Omega := \bigcup_{m=0}^d \Omega_m = P_d$$

y

$$\Sigma := \bigcup_{m=0}^d \Sigma_m \subseteq \Omega.$$

AFIRMACIÓN 3.3.1.1. *Con las notaciones anteriores, se verifican las siguientes propiedades:*

- i) Si E es un q -conjunto de Kakeya con direcciones en V , entonces E es un conjunto cuestor para Ω con respecto a Σ .
- ii) Definimos $d_0 := \max\{m \in \mathbb{N} : m \leq d, I_m(V) \neq \{0\}\}$. Entonces, o bien $d_0 = d$, o $\Sigma_m = \{0\}$ para todo $m \leq d$.
- iii) La dimensión de los conjuntos constructibles Ω y Σ viene dada por las siguientes identidades:

$$\begin{aligned} \dim(\Omega) &:= \sum_{i=0}^d \dim_K(H_i(X_1, \dots, X_n)) = \binom{d+n}{n}, \\ \dim(\Sigma) &:= \begin{cases} \dim_K(I_d(V)) + \sum_{i=0}^{d-1} \dim_K(H_i(X_1, \dots, X_n)), & \text{si } I_d(V) \neq \{0\} \\ 0, & \text{en otro caso} \end{cases}. \end{aligned}$$

Demostración de la Afirmación. Probamos cada ítem por separado:

- *Ítem i):* Sea $f \in \Omega$ un polinomio no nulo de grado m tal que $f|_E \equiv 0$. Entonces, para cada $v \in V$, existe $x \in E$ tal que f se anula en la intersección $E \cap \rho(x, v)$. Tomando las componentes homogéneas de f :

$$f = f_m + f_{m-1} + \dots + f_0,$$

donde $f_m = LHC(f) \neq 0$, existen polinomios en $2n$ variables h_i tales que para todo $t \in K$, se verifica la siguiente identidad:

$$f(x + tv) = f_m(v)t^m + \sum_{i=1}^{m-1} h_i(x, v)t^i + f_0(x).$$

Consideramos ahora el polinomio univariado:

$$F(T) := f(x + Tv) = f_m(v)T^m + \sum_{i=1}^{m-1} h_i(x, v)T^i + f_0(x) \in K[T].$$

Como E es un q -conjunto de Kakeya con direcciones en V , existen al menos q puntos distintos $t_1, \dots, t_q \in K$ tales que $x + t_i v \in E$, $1 \leq i \leq q$. Por lo tanto, deducimos que $F(t_i) = 0$, $1 \leq i \leq q$ y, por

argumentos elementales de interpolación, obtenemos que F es el polinomio nulo en $K[T]$. Entonces, concluimos que $LHC(f)(v) = f_m(v) = 0$ y, por consiguiente, $LHC(f) \in I_m(V)$ y hemos demostrado que:

$$\forall f \in \Omega, f|_E \equiv 0 \implies f \in \Sigma.$$

- *Ítem ii):* Nótese que si $d_0 > 0$, entonces para cada polinomio homogéneo no nulo $f \in I_{d_0}(V)$, el polinomio $X_1^{d-d_0}f \in I_d(V)$. Por lo tanto, o bien $d_0 = d$, o $\Sigma_m = \{0\}$ para todo $m \leq d$.
- *Ítem iii):* En primer lugar, observemos que para cada $m \leq d$, la dimensión de los conjuntos constructibles Ω_m y Σ_m satisface las siguientes identidades:

$$\dim(\Omega_m) := \sum_{i=0}^m \dim_K(H_i(X_1, \dots, X_n)),$$

mientras que:

$$\dim(\Sigma_m) := \begin{cases} \dim_K(I_m(V)) + \sum_{i=0}^{m-1} \dim_K(H_i(X_1, \dots, X_n)), & \text{si } I_m(V) \neq \{0\} \\ 0, & \text{en otro caso} \end{cases}.$$

Ambas identidades son inmediatas a partir de las definiciones de Ω_m y Σ_m combinadas con la definición de dimensión de Krull para conjuntos constructibles. A continuación, veamos que para $m \leq d$, se verifican las siguientes desigualdades:

$$\dim(\Omega_m) \leq \dim(\Omega_d), \quad \dim(\Sigma_m) \leq \dim(\Sigma_d).$$

La primera desigualdad es obvia. Para la segunda basta observar que si $m < d$ y $I_m(V) \neq \{0\}$, entonces $I_d(V) \neq \{0\}$. En este caso,

$$\dim_K(I_m(V)) + \sum_{i=0}^{m-1} \dim_K(H_i(X_1, \dots, X_n)) \leq \sum_{i=0}^m \dim_K(H_i(X_1, \dots, X_n)),$$

y, por lo tanto, tenemos que:

$$\dim(\Sigma_m) \leq \dim_K(I_d(V)) + \sum_{i=0}^{d-1} \dim_K(H_i(X_1, \dots, X_n)).$$

Finalmente, concluimos:

$$\dim(\Omega) = \max\{\dim(\Omega_m) : 0 \leq m \leq d\} = \dim(\Omega_d) = \binom{d+n}{n}.$$

Por otro lado, si $I_d(V) \neq \{0\}$, obtenemos:

$$\dim(\Sigma) = \max\{\dim(\Sigma_m) : 0 \leq m \leq d\} \leq \dim_K(I_d(V)) + \sum_{i=0}^{d-1} \dim_K(H_i(X_1, \dots, X_n)).$$

Obviamente, si $I_d(V) = \{0\}$, entonces $I_m(V) = \{0\}$ para todo $m \leq d$ y, por lo tanto, $\Sigma = \{0\}$.

■

Por consiguiente, aplicamos la Proposición 3.1.12 para concluir:

$$\sharp(E) \geq \dim(\Omega) - \dim(\Sigma) \geq \dim_K(H_d(X_1, \dots, X_n)) - \dim_K(I_d(V)) = \chi_V(d).$$

Como V es cero-dimensional, se verifica la segunda afirmación. Esto es, para cada $d \geq R$, se cumple que:

$$\chi_V(d) = \deg(V) = \sharp(V).$$

Finalmente, la última afirmación es obvia y no es óptima. Fijamos un conjunto Λ de q elementos en K incluyendo al 0 y cualquier aplicación $\varphi : V \rightarrow K^n$. Entonces, el siguiente conjunto es un q -conjunto de Kekeya con direcciones en V :

$$E := \bigcup_{v \in V} \{\varphi(v) + \lambda v : \lambda \in \Lambda\},$$

cuya cardinalidad es a lo sumo $q \deg(V) = q\sharp(V)$. Tomando φ como cualquier aplicación constante, obtenemos una cardinalidad acotada por:

$$(q-1) \deg(V) + 1.$$

□

A continuación, mostramos que el resultado principal de [Dv, 09] es una consecuencia de nuestra Proposición 3.3.1 anterior:

COROLARIO 3.3.2 ([Dv, 09]). Sea $\kappa := \mathbb{F}_q$ un cuerpo finito de cardinalidad q , y $K := \overline{\mathbb{F}_q}$ su clausura algebraica. Sea $P_d^K = P_d^K(X_1, \dots, X_n)$ el conjunto constructible de todos los polinomios en el conjunto de variables $\{X_1, \dots, X_n\}$ de grado a lo sumo d con coeficientes en $K = \overline{\mathbb{F}_q}$. Si $d \leq q-1$ y $E \subseteq \mathbb{A}^n(\mathbb{F}_q)$ es un conjunto de Kekeya, entonces E es un conjunto cuestor para P_d^K (con respecto a $\{0\}$). En particular, se verifica la siguiente desigualdad:

$$\sharp(E) \geq \dim(P_d^K) = \binom{d+n}{n}, \quad \forall d \leq q-1.$$

DEMOSTRACIÓN. En primer lugar, nótese que todo conjunto de Kekeya $E \subseteq \mathbb{A}^n(\mathbb{F}_q)$ es un q -conjunto de Kekeya con direcciones en $V = \mathbb{P}_{n-1}(\mathbb{F}_q)$. Por lo tanto, E es un conjunto cuestor para $\Omega := P_d^K$ con respecto a Σ , donde Σ es el subconjunto constructible de P_d^K definido en la demostración de la Proposición 3.3.1 precedente. Además, por la Proposición 3.3.1 concluimos que:

$$\sharp(E) \geq \chi_V(d) = \dim_K(H_d(X_1, \dots, X_n)) - \dim_K(I_d(V)), \quad \forall d \leq q-1,$$

donde $H_d(X_1, \dots, X_n)$ es el espacio vectorial sobre $\overline{\mathbb{F}_q}$ generado por todos los polinomios homogéneos de grado d con coeficientes en $\overline{\mathbb{F}_q}$ e $I_d(V)$ es el espacio vectorial generado por todos los polinomios $f \in H_d(X_1, \dots, X_n)$ que se anulan en V . El corolario se sigue demostrando que $\Sigma = \{0\}$ para cada $d \leq q-1$. Esto se consigue si probamos la siguiente afirmación:

$$\chi_{\mathbb{P}_{n-1}(\mathbb{F}_q)}(d) = \binom{d+n}{n}, \quad \forall d \leq q-1.$$

Equivalentemente, será suficiente demostrar que $I_m(\mathbb{P}_{n-1}(\mathbb{F}_q)) = \{0\}$ para todo $m \leq d \leq q-1$. Para probar esta última afirmación, sea $f \in H_m(X_1, \dots, X_n)$ un polinomio homogéneo no nulo de grado m . Como f es no nulo, al menos una de sus trazas afines debe ser no nula. Sin pérdida de generalidad, supongamos que ${}^a f(X_2, \dots, X_m) := f(1, X_2, \dots, X_m) \in \overline{\mathbb{F}_q}[X_2, \dots, X_m]$ es un polinomio no nulo de grado a lo sumo $m \leq q-1$. Como f se anula en todo el espacio $\mathbb{P}_{n-1}(\mathbb{F}_q)$, entonces ${}^a f$ se anula en los puntos afines de $\mathbb{A}^{n-1}(\mathbb{F}_q) \subseteq \mathbb{P}_{n-1}(\mathbb{F}_q)$, donde:

$$\mathbb{A}^{n-1}(\mathbb{F}_q) := \{(1 : x_2 : \dots : x_n) \in \mathbb{P}_{n-1}(\mathbb{F}_q) : x_2, \dots, x_n \in \mathbb{F}_q\}.$$

Si ${}^a f$ se anula en \mathbb{F}_q^{n-1} , tendríamos:

$$q^{n-1} \leq \sharp(V_{\mathbb{F}_q}({}^a f)),$$

donde $V_{\mathbb{F}_q}({}^a f)$ son los puntos \mathbb{F}_q -racionales de la hipersuperficie $V_{\mathbb{A}}({}^a f)$. Esto no puede ser cierto debido al Corolario 2.2.7: como ${}^a f$ es un polinomio no nulo de grado a lo sumo $m \leq d \leq q-1$, tendríamos que:

$$q^{n-1} \leq \sharp(V_{\mathbb{F}_q}({}^a f)) \leq \deg({}^a f)q^{n-2} \leq (q-1)q^{n-2},$$

lo cual es una contradicción y se deduce el corolario. \square

Sin embargo, la mayoría de los conjuntos cuestores no son conjuntos de Kekeya:

COROLARIO 3.3.3. Sea \mathbb{F}_q un cuerpo finito con q elementos y sea $\overline{\mathbb{F}_q}$ su clausura algebraica. Sea $d \in \mathbb{N}$ un grado y $k \in \mathbb{N}$ un entero positivo tal que:

$$(3.3.1) \quad d < q^{1-\frac{1}{k}} - 1.$$

Entonces, para $s := k \binom{d+n}{n}$, la probabilidad de que una lista de puntos $\mathbf{Q} \in \mathbb{A}^n(\mathbb{F}_q)^s$ sea un conjunto cuestor para $P_d(X_1, \dots, X_n)$ con respecto a $\{0\}$ es al menos:

$$1 - \frac{1}{q^{\dim(\Omega)}}.$$

Si, además, d, k, q y n verifican la siguiente desigualdad:

$$(3.3.2) \quad k \binom{d+n}{n} < \frac{q^n}{2^n},$$

entonces existen conjuntos cuestores para $P_d(X_1, \dots, X_n)$ con respecto a $\{0\}$ que no son conjuntos de Kekeya.

DEMOSTRACIÓN. Con las notaciones de la Subsección 3.3.2, consideramos el subconjunto constructible Ω dado por la siguiente identidad:

$$\Omega := P_d(X_1, \dots, X_n) := \{f \in \overline{\mathbb{F}_q}[X_1, \dots, X_n] : \deg(f) \leq d\}.$$

Reproducimos a continuación la demostración del Teorema 3.2.1. Dado $s := k \dim(\Omega) \in \mathbb{N}$ un entero positivo, consideramos la siguiente variedad de incidencia:

$$V(\Omega, s) := \{(f, x_1, \dots, x_s) \in \Omega \times (\mathbb{A}^n(\overline{\mathbb{F}_q}))^s : f(x_i) = 0, 1 \leq i \leq s\}.$$

Consideramos también las dos proyecciones canónicas $\pi_1 : V(\Omega, s) \rightarrow \Omega$ y $\pi_2 : V(\Omega, s) \rightarrow (\mathbb{A}^n(\overline{\mathbb{F}}_q))^s$, y la clase \mathcal{C} formada por todas las componentes irreducibles W de $V(\Omega, s)$ tales que $\pi_1(W) \setminus \{0\} \neq \emptyset$. Usando los mismos argumentos que en la demostración del Teorema 3.2.1, concluimos las siguientes propiedades:

- Para cada $W \in \mathcal{C}$, su dimensión satisface:

$$\dim(W) \leq (n-1)s + \dim(\Omega) = (n-1)s + \binom{d+n}{n}.$$

- Definiendo $B(\Omega, s) := \cup_{W \in \mathcal{C}} W$, por la Desigualdad de Bézout (cf. Teorema 2.1.14) obtenemos:

$$\deg_{\text{Ici}}(B(\Omega, s)) \leq \deg_{\text{Ici}}(\Omega) (d+1)^s.$$

Además, tenemos que:

$$\dim(B(\Omega, s)) \leq (n-1)s + \dim(\Omega) = (n-1)s + \binom{d+n}{n}.$$

Como Ω es lineal en P_d , se deduce que:

$$\deg_{\text{Ici}}(B(\Omega, s)) \leq (d+1)^s.$$

Gracias a la Proposición 2.1.11, tenemos que:

$$\deg(\overline{\pi_2(B(\Omega, s))^z}) \leq \deg_{\text{Ici}}(B(\Omega, s)) \leq (d+1)^s.$$

Tomando $C := \mathbb{F}_q^n$, deducimos de la Proposición 2.2.2 que:

$$\deg(\overline{\pi_2(B(\Omega, s))^z} \cap (\mathbb{F}_q^n)^s) \leq \deg_{\text{Ici}}(B(\Omega, s)) q^{\dim(B(\Omega, s))}.$$

Por lo tanto, de las cotas anteriores y de la última parte de la Demostración del Teorema 3.2.1, concluimos que la probabilidad de que una lista $\mathbf{Q} \in C^s$ no sea un conjunto cuestor para Ω con respecto a $\Sigma = \{0\}$ es a lo sumo:

$$\begin{aligned} \frac{\deg(\overline{\pi_2(B(\Omega, s))^z} \cap (\mathbb{F}_q^n)^s)}{\#((\mathbb{F}_q^n)^s)} &\leq \frac{\deg_{\text{Ici}}(B(\Omega, s)) q^{\dim(B(\Omega, s))}}{\#((\mathbb{F}_q^n)^s)} \leq \\ &\leq \frac{(d+1)^s q^{(n-1)s + \dim(\Omega)}}{q^{ns}} \leq \frac{(d+1)^s q^{\dim(\Omega)}}{q^s} \leq \frac{1}{q^{\dim(\Omega)\lambda}}, \end{aligned}$$

donde $\lambda := (k-1) - k \log_q(d+1)$. Por lo tanto, la probabilidad de que una lista escogida al azar de longitud s en \mathbb{F}_q^n sea un conjunto cuestor para Ω respecto a $\{0\}$ es al menos:

$$1 - \frac{1}{q^{\dim(\Omega)\lambda}}.$$

En particular, si $\log_q(d+1) < \frac{k-1}{k} = (1 - 1/k)$, existen conjuntos cuestores de longitud $k \dim(\Omega)$.

Por otro lado, de acuerdo con [DKSS, 09], se verifica la siguiente cota inferior para la cardinalidad de conjuntos de Kakeya en \mathbb{F}_q^n :

$$(3.3.3) \quad \#(E) \geq \frac{\#(\mathbb{F}_q^n)}{2^n} = \frac{q^n}{2^n}.$$

Entonces, si d, k, q y n satisfacen las Desigualdades (3.3.1) y (3.3.2), deducimos de la primera parte de la demostración y de la Desigualdad (3.3.3) que existen conjuntos cuestores para P_d con respecto a $\{0\}$ de longitud $k \binom{d+n}{n}$ que no son conjuntos de Kakeya. \square

Mientras que la Desigualdad (3.3.2) es natural para valores asintóticos de n , la Desigualdad (3.3.1) implica que nuestro método de demostración no nos permite exhibir conjuntos cuestores que no sean conjuntos de Kakeya para $d = q - 1$.

3.3.3. Conjuntos Cuestores, Nullstellensatz Combinatorio de Alon y dualidad. El resultado de [Al, 99], revisado en [Tao, 14], es otro método, al igual que los conjuntos de Kakeya, que muestra conjuntos cuestores de longitud exponencial, aunque comúnmente se le conoce como el Nullstellensatz Combinatorio. El resultado de Alon es el siguiente:

TEOREMA 3.3.4 (Nullstellensatz Combinatorio, de acuerdo a [Al, 99]). *Sea κ un cuerpo y K su clausura algebraica. Sea d_1, \dots, d_n una lista de enteros positivos y sea $P \in K[X_1, \dots, X_n]$ un polinomio de grado $d_1 + \dots + d_n$ tal que el coeficiente del monomio $X_1^{d_1} \dots X_n^{d_n}$ en P es no nulo. Entonces, para cada familia de subconjuntos $E_1, \dots, E_n \subseteq \kappa$, tales que $\#(E_i) > d_i$, existe un punto ζ en el producto cartesiano $E = E_1 \times \dots \times E_n$ tal que $P(\zeta) \neq 0$.*

Este resultado ha sido empleado en multitud de ocasiones por la comunidad científica de la Combinatoria, bajo el marco general del “Método Polinomial”. La Observación 5.3 de [Tao, 14] contiene información adicional sobre el uso del nombre “Nullstellensatz Combinatorio”.

En esta subsección, revisamos el resultado precedente bajo un formato diferente. Usaremos las mismas notaciones que en secciones anteriores. Para cada polinomio $f \in P_d(X_1, \dots, X_n)$ y para cada exponente $\underline{\mu} := (\mu_1, \dots, \mu_n) \in \mathbb{N}^n$, tal que $|\underline{\mu}| = \mu_1 + \dots + \mu_n \leq d$, denotamos por $f_{\underline{\mu}}$ al coeficiente del monomio $X_1^{\mu_1} \cdots X_n^{\mu_n}$ en la expansión monomial de f . Nuestro objetivo es demostrar el siguiente resultado:

TEOREMA 3.3.5 (Nullstellensatz Combinatorio). *Con las notaciones anteriores, sea $(d) := (d_1, \dots, d_n)$ una lista de grados, con $d_i \geq 1$. Sea $D := d_1 + \dots + d_n - n$ un entero no negativo. Sea $\Delta_{(d)} \subseteq \mathbb{N}^n$ el subconjunto dado por la siguiente igualdad:*

$$(3.3.4) \quad \Delta_{(d)} := \{\underline{\mu} \in \mathbb{N}^n : \underline{\mu} = (\mu_1, \dots, \mu_n), 0 \leq \mu_i \leq d_i - 1\}.$$

Sea $\Omega_{(d)} \subseteq P_D^K(X_1, \dots, X_n)$ el siguiente subconjunto constructible:

$$\Omega_{(d)} := \{f \in P_D^K(X_1, \dots, X_n) : \exists \underline{\mu} \in \Delta_{(d)}, f_{\underline{\mu}} \neq 0, |\underline{\mu}| = \deg(f)\} \cup \{0\}.$$

Entonces, para cada lista finita de subconjuntos $E_1, \dots, E_n \subseteq K$, tales que $\sharp(E_i) = d_i$, el producto cartesiano $E := E_1 \times \dots \times E_n$ es un conjunto cuestor para $\Omega_{(d)}$ con respecto a $\Sigma = \{0\}$. Además, si $\sharp(\kappa) \geq \max\{d_i + 1 : 1 \leq i \leq n\}$, podemos escoger $E_i \subseteq \kappa$ para cada i , $1 \leq i \leq n$.

Como $(d_1 - 1, \dots, d_n - 1) \in \Delta_{(d)}$, este último teorema incluye al Nullstellensatz Combinatorio de Alon. Presentaremos una nueva demostración de este resultado, inspirada en algunos aspectos discutidos en [Tao, 14], pero usando como principal ingrediente matemático la traza y la dualidad en álgebras de variedades cero-dimensionales.

Comenzamos con una construcción clásica que recordamos para fijar las notaciones. Con las mismas notaciones e hipótesis precedentes, sea κ un cuerpo y K su clausura algebraica. Para cada punto $z \in \mathbb{A}^n(K)$, denotamos por $\mathfrak{m}_z \subseteq K[X_1, \dots, X_n]$ al ideal maximal de todos los polinomios $K[X_1, \dots, X_n]$ que se anulan en z , i.e.

$$\mathfrak{m}_z := \{f \in K[X_1, \dots, X_n] : f(z) = 0\}.$$

Sea $E \subseteq \mathbb{A}^n(K)$ un conjunto finito y sea $D = \deg(E) = \sharp(E)$ su cardinalidad. Supongamos que $E = \{z_1, \dots, z_D\} \subseteq \mathbb{A}^n(K)$. Denotamos por $I(E) \subseteq K[X_1, \dots, X_n]$ al ideal de todos los polinomios en $K[X_1, \dots, X_n]$ que se anulan en todos los puntos de E , i.e.

$$I(E) = \{f \in K[X_1, \dots, X_n] : f(z) = 0, \forall z \in E\} = \bigcap_{i=1}^D \mathfrak{m}_{z_i}.$$

Sea $K[E] := K[X_1, \dots, X_n]/I(E)$ el anillo residual de $K[X_1, \dots, X_n]$ módulo $I(E)$, que es una K -álgebra de Artin cero-dimensional y un espacio vectorial de dimensión igual a $\sharp(E)$.

Si $E \subseteq \mathbb{A}^n(\kappa)$, consideramos también el anillo $\kappa[E]$ dado por:

$$\kappa[X_1, \dots, X_n]/I(E)^c,$$

donde $I(E)^c := I(E) \cap \kappa[X_1, \dots, X_n]$ es la contracción de $I(E)$. Entonces, $\kappa[E]$ es también un anillo de Artin cero-dimensional y tenemos que:

$$K \otimes_{\kappa} \kappa[E] = K[E].$$

En particular, si $E \subseteq \mathbb{A}^n(\kappa)$, las siguientes dimensiones (como espacios vectoriales) coinciden:

$$\dim_{\kappa}(\kappa[E]) = \dim_K(K[E]).$$

A continuación, recordamos una identificación clásica de clases residuales en $K[E]$ (o $\kappa[E]$) con homotecias sobre los mismos espacios vectoriales. Para cada $h \in K[X_1, \dots, X_n]$, denotamos por $\eta_h : K[E] \rightarrow K[E]$ al siguiente endomorfismo de K -espacios vectoriales:

$$\eta_h : \begin{array}{ccc} K[E] & \longrightarrow & K[E] \\ g + I(E) & \longmapsto & (hg) + I(E) \end{array}$$

Las trazas de estos endomorfismos nos permiten introducir la siguiente forma bilineal en $K[E]$:

$$\langle \cdot, \cdot \rangle_E : \begin{array}{ccc} K[E] \times K[E] & \longrightarrow & K \\ (g_1 + I(E), g_2 + I(E)) & \longmapsto & Tr(\eta_{g_1 g_2}) \end{array}$$

Los siguientes resultados son consecuencias conocidas del Teorema Chino de los Restos aplicado a $K[E]$:

Para cada $h \in K[X_1, \dots, X_n]$, η_h es un endomorfismo diagonalizable y su forma canónica de Jordan es la siguiente matriz diagonal:

$$\text{Diag}(h(z_1), \dots, h(z_D)).$$

En particular, la traza de η_h está dada por la siguiente identidad:

$$\mathrm{Tr}(\eta_h) = \sum_{i=1}^D h(z_i) = \sum_{z \in E} h(z).$$

Además, tenemos el siguiente resultado:

LEMA 3.3.6. *Con las notaciones precedentes, para cada base $B = \{v_1, \dots, v_D\}$, $v_i = f_i + I(E)$, de $K[E]$, como K -espacio vectorial, existe una base (llamada dual respecto a $\langle \cdot, \cdot \rangle_E$) $B^* = \{w_1, \dots, w_D\}$, $w_j = g_j + I(E)$, tal que para todo $i, j \in \{1, \dots, D\}$, tenemos que:*

$$\langle v_i, w_j \rangle_E = \sum_{z \in E} f_i(z)g_j(z) = \delta_{i,j},$$

donde $\delta_{i,j}$ es la delta de Kronecker.

DEMOSTRACIÓN. Sea $B := \{v_1, \dots, v_D\}$ una base ordenada de $K[E]$ como K -espacio vectorial. Por el Teorema Chino de los Restos, podemos identificar $K[E]$ y K^D mediante el siguiente isomorfismo:

$$\tilde{\varphi}: \begin{array}{ccc} K[E] & \longrightarrow & K^D \\ g + I(E) & \longrightarrow & (g(z_1), \dots, g(z_D)) \end{array}.$$

Entonces, B es una base de $K[E]$ si y solo si la siguiente matriz (de Vandermonde) es regular:

$$vdM(B) := \begin{pmatrix} v_1(z_1) & \cdots & v_1(z_D) \\ \vdots & \ddots & \vdots \\ v_D(z_1) & \cdots & v_D(z_D) \end{pmatrix}.$$

Para cada i , $1 \leq i \leq D$, sea e_k el k -ésimo vector de la base “canónica” de K^D y sea $\omega_i := (\omega_{i,1}, \dots, \omega_{i,D}) \in K^D$ la única solución del siguiente sistema de ecuaciones lineales:

$$vdM(B) \begin{pmatrix} \omega_{i,1} \\ \vdots \\ \omega_{i,D} \end{pmatrix} = e_i^t,$$

donde e_i^t es la matriz traspuesta del vector e_i . Entonces, por el isomorfismo $\tilde{\varphi}$ existe $g_i \in K[X_1, \dots, X_n]$ y $w_i = g_i + I(E) \in K[E]$ tal que:

$$\tilde{\varphi}(g_i + I(E)) = (g_i(z_1), \dots, g_i(z_D)) = (\omega_{i,1}, \dots, \omega_{i,D}).$$

La familia $B^* := \{w_1, \dots, w_D\}$ es una base “dual” de B . □

LEMA 3.3.7. *Sean $h_1, \dots, h_n \in \kappa[T]$ polinomios univariados con grados d_1, \dots, d_n respectivamente. Supongamos que h_1, \dots, h_n se escinden completamente en κ y que son libres de cuadrados (i.e. no tienen raíces múltiples). Definimos el ideal $\mathfrak{a} = (h_1(X_1), \dots, h_n(X_n)) \subseteq K[X_1, \dots, X_n]$ y, para cada i , $1 \leq i \leq n$, sea $E_i \subseteq \kappa$ el conjunto de ceros de h_i en κ . Consideramos también la variedad cero-dimensional dada por el producto cartesiano: $E = E_1 \times \cdots \times E_n \subseteq \mathbb{A}^n(K)$. Tenemos que:*

i) *El ideal $\mathfrak{a} = I(E)$ es radical (i.e. el anillo residual $K[X_1, \dots, X_n]/\mathfrak{a}$ no tiene elementos nilpotentes no nulos) y, por lo tanto:*

$$\mathfrak{a} = I(E) = \{f \in K[X_1, \dots, X_n] : f|_E \equiv 0\}.$$

ii) *La familia $\{h_1(X_1), \dots, h_n(X_n)\}$ es una base de Gröbner de $\mathfrak{a} = I(E)$ con respecto al orden monomial “grado+lexicográfico” con el siguiente orden de variables: $X_1 < X_2 < \dots < X_n$.*

iii) *El siguiente conjunto es una base monomial de $K[E]$:*

$$B = \{X_1^{\mu_1} \cdots X_n^{\mu_n} + \mathfrak{a} : 0 \leq \mu_i \leq d_i - 1, 1 \leq i \leq n\}.$$

iv) *Para cada i , $1 \leq i \leq n$, denotemos por B_i^* a la base dual en $K[E_i]$ respecto a la traza $\langle \cdot, \cdot \rangle_{E_i}$ de $B_i := \{T^{\mu_i} + (h_i) : 0 \leq \mu_i \leq d_i - 1\}$. Esta base dual viene dada por:*

$$B_i^* := \{g_k^{(i)}(T) + (h_i) : 0 \leq k \leq d_i - 1\} \subseteq K[E_i] := K[T]/(h_i).$$

Entonces, el siguiente conjunto es una base dual de B en $K[E]$ respecto a la forma bilineal $\langle \cdot, \cdot \rangle_E$:

$$B^* = \left\{ \left(\prod_{i=1}^n g_{\mu_i}^{(i)}(X_i) \right) + \mathfrak{a} : (\mu_1, \dots, \mu_n) \in \mathbb{N}^n, 0 \leq \mu_i \leq d_i - 1, 1 \leq i \leq n \right\}.$$

DEMOSTRACIÓN. Los Ítems *i*), *ii*) y *iii*) son conocidos y no necesitan demostración. Para el Ítem *iv*), denotemos por $(d) = (d_1, \dots, d_n)$ a la lista de grados y por $\Delta_{(d)}$ a la clase de exponentes monomiales introducida en la Identidad (3.3.4) del Teorema 3.3.5.

Nótese que para cada i , $1 \leq i \leq n$, tenemos que:

$$\sum_{z_i \in E_i} g_k^{(i)}(z_i) z_i^r := \delta_{k,r},$$

para $0 \leq k, r \leq d_i - 1$.

Para cada pareja de exponentes monomiales $\underline{\theta} := (\theta_1, \dots, \theta_n) \in \Delta_{(d)}$ y $\underline{\mu} := (\mu_1, \dots, \mu_n) \in \Delta_{(d)}$, definimos la siguiente cantidad:

$$D_{\underline{\mu}, \underline{\theta}} := \left\langle \prod_{i=1}^n g_{\mu_i}^{(i)}(X_i) + \mathfrak{a}, X_1^{\theta_1} \cdots X_n^{\theta_n} + \mathfrak{a} \right\rangle_E = \sum_{(z_1, \dots, z_n) \in E} \left(\left(\prod_{i=1}^n g_{\mu_i}^{(i)}(z_i) \right) z_1^{\theta_1} \cdots z_n^{\theta_n} \right).$$

Tenemos que:

$$D_{\underline{\mu}, \underline{\theta}} = \left(\sum_{z_1 \in E_1} g_{\mu_1}^{(1)}(z_1) z_1^{\theta_1} \right) \left(\sum_{(z_2, \dots, z_n) \in E_2 \times \cdots \times E_n} \left(\left(\prod_{i=2}^n g_{\mu_i}^{(i)}(z_i) \right) z_2^{\theta_2} \cdots z_n^{\theta_n} \right) \right).$$

Como B_1^* es una base dual de B_1 , concluimos que:

$$\left(\sum_{z_1 \in E_1} g_{\mu_1}^{(1)}(z_1) z_1^{\theta_1} \right) = \delta_{\mu_1, \theta_1},$$

y también tenemos que:

$$D_{\underline{\mu}, \underline{\theta}} = \delta_{\mu_1, \theta_1} \left(\sum_{(z_2, \dots, z_n) \in E_2 \times \cdots \times E_n} \left(\left(\prod_{i=2}^n g_{\mu_i}^{(i)}(z_i) \right) z_2^{\theta_2} \cdots z_n^{\theta_n} \right) \right).$$

Por inducción en n , finalmente concluimos:

$$D_{\underline{\mu}, \underline{\theta}} = \delta_{\mu_1, \theta_1} \delta_{\mu_2, \theta_2} \cdots \delta_{\mu_n, \theta_n} = \delta_{\underline{\mu}, \underline{\theta}},$$

y se demuestra el Ítem *iv*). □

El siguiente resultado es una variación de un lema que puede encontrarse en [Tao, 14]. Daremos una demostración basada en nuestros argumentos de dualidad:

LEMA 3.3.8. *Con las notaciones anteriores, sea $\underline{\theta} = (\theta_1, \dots, \theta_n) \in \Delta_{(d)}$ un exponente monomial en $\Delta_{(d)}$. Sea $\underline{\mu} = (\mu_1, \dots, \mu_n) \in \mathbb{N}^n$ otro exponente monomial tal que $|\underline{\mu}| = \mu_1 + \cdots + \mu_n = |\underline{\theta}|$. Si existe algún índice i tal que $\mu_i \geq d_i$, entonces se verifica la siguiente igualdad:*

$$D_{\underline{\mu}, \underline{\theta}} = \left\langle \left(\prod_{i=1}^n g_{\theta_i}^{(i)}(X_i) \right) + \mathfrak{a}, X_1^{\mu_1} \cdots X_n^{\mu_n} + \mathfrak{a} \right\rangle_E = 0.$$

DEMOSTRACIÓN. La demostración es similar a la del Lema 3.3.7 precedente. Sin pérdida de generalidad, supongamos que $\mu_n \geq d_n$. Dado que tenemos:

$$\mu_1 + \cdots + \mu_n = \theta_1 + \cdots + \theta_n, \quad 0 \leq \theta_i \leq d_i - 1,$$

entonces debe existir $j \neq n$ tal que $\mu_j \neq \theta_j$ y $0 \leq \mu_j \leq d_j - 1$. Para probar lo anterior, basta observar que si para todo $j \neq n$ tenemos que o bien $\mu_j = \theta_j$ o $\mu_j \geq d_j$, entonces obtendríamos:

$$|\underline{\mu}| = \mu_1 + \cdots + \mu_n \geq \theta_1 + \cdots + \theta_{n-1} + \mu_n \geq \theta_1 + \cdots + \theta_{n-1} + \theta_n + 1 > |\underline{\theta}| + 1,$$

que contradice la hipótesis $|\underline{\mu}| = |\underline{\theta}|$.

De nuevo, sin pérdida de generalidad, supongamos que $\mu_1 \neq \theta_1$ y $0 \leq \mu_1 \leq d_1 - 1$. Entonces, tenemos que:

$$\begin{aligned} D_{\underline{\mu}, \underline{\theta}} &:= \left\langle \left(\prod_{i=1}^n g_{\mu_i}^{(i)}(X_i) \right) + \mathfrak{a}, X_1^{\theta_1} \cdots X_n^{\theta_n} + \mathfrak{a} \right\rangle_E = \\ &= \sum_{(z_1, \dots, z_n) \in E} \left(\left(\prod_{i=1}^n g_{\mu_i}^{(i)}(z_i) \right) z_1^{\theta_1} \cdots z_n^{\theta_n} \right). \end{aligned}$$

Por lo tanto, obtenemos:

$$D_{\underline{\mu}, \underline{\theta}} = \left(\sum_{z_1 \in E_1} g_{\mu_1}^{(1)}(z_1) z_1^{\theta_1} \right) \left(\sum_{(z_2, \dots, z_n) \in E_2 \times \dots \times E_n} \left(\left(\prod_{i=2}^n g_{\mu_i}^{(i)}(z_i) \right) z_2^{\theta_2} \dots z_n^{\theta_n} \right) \right).$$

Y, como $\mu_1 \neq \theta_1$, concluimos:

$$D_{\underline{\mu}, \underline{\theta}} = 0 \cdot \left(\sum_{(z_2, \dots, z_n) \in E_2 \times \dots \times E_n} \left(\left(\prod_{i=2}^n g_{\mu_i}^{(i)}(z_i) \right) z_2^{\theta_2} \dots z_n^{\theta_n} \right) \right) = 0,$$

lo que finaliza la demostración del lema. \square

COROLARIO 3.3.9. *Con las notaciones precedentes, sean $\underline{\theta} \in \Delta_{(d)}$ y $\underline{\mu} \in \mathbb{N}^n$ dos exponentes monomiales. Si $|\underline{\mu}| \leq |\underline{\theta}|$, tenemos que:*

$$D_{\underline{\mu}, \underline{\mu}} = \left\langle \left(\prod_{i=1}^n g_{\theta_i}^{(i)}(X_i) \right) + \mathfrak{a}, X_1^{\mu_1} \dots X_n^{\mu_n} + \mathfrak{a} \right\rangle_E = \delta_{\underline{\mu}, \underline{\theta}},$$

donde $\delta_{\underline{\mu}, \underline{\theta}}$ es la delta de Kronecker.

DEMOSTRACIÓN. Si $\underline{\mu} \in \Delta_{(d)}$, el Lema 3.3.7 implica el resultado. En caso contrario, si $|\underline{\mu}| \leq |\underline{\theta}|$ y $\underline{\mu} \notin \Delta_{(d)}$, el resultado se verifica gracias al Lema 3.3.8. \square

Finalmente, estamos en condiciones de demostrar el Teorema 3.3.5:

DEMOSTRACIÓN. (DEL TEOREMA 3.3.5). Sean $d_1, \dots, d_n \in \mathbb{N}$ enteros positivos, $(d) = (d_1, \dots, d_n)$ la lista de los grados que definen y $\Delta_{(d)}$ la clase de los exponentes monomiales como en el enunciado del Teorema 3.3.5. Sea $D := d_1 + \dots + d_n - n$ la cantidad también definida en el enunciado de dicho teorema. Sea $\Omega_{(d)} \subseteq P_D(X_1, \dots, X_n)$ el siguiente subconjunto constructible formado por polinomios de grado a lo sumo D :

$$\Omega_{(d)} := \{f \in P_D^K(X_1, \dots, X_n) : \exists \underline{\mu} \in \Delta_{(d)}, f_{\underline{\mu}} \neq 0, |\underline{\mu}| = \deg(f)\} \cup \{0\}.$$

Sean $E_1, \dots, E_n \subseteq K$ subconjuntos finitos tales que $\sharp(E_i) = d_i$, $E := E_1 \times \dots \times E_n$ su producto cartesiano y $K[E]$ el anillo residual de $K[X_1, \dots, X_n]$ módulo $I(E)$. Sea $\langle \cdot, \cdot \rangle_E$ la forma bilineal asociada a la traza en $K[E]$. Para cada i , $1 \leq i \leq n$, sea $h_i \in K[T]$ el siguiente polinomio univariado:

$$h_i(T) := \prod_{\zeta \in E_i} (T - \zeta).$$

Sea $\mathfrak{a} := (h_1(X_1), \dots, h_n(X_n))$ el ideal que generan. De acuerdo con el Lema 3.3.7, tenemos que:

$$K[E] = K[X_1, \dots, X_n] / \mathfrak{a}.$$

Sea B la base monomial de $K[E]$ dada por la siguiente igualdad:

$$B = \{X_1^{\mu_1} \dots X_n^{\mu_n} + \mathfrak{a} : 0 \leq \mu_i \leq d_i - 1, 1 \leq i \leq n\},$$

y sea B^* la base dual de B que aparece en el Lema 3.3.7 anterior:

$$B^* = \left\{ \left(\prod_{i=1}^n g_{\mu_i}^{(i)}(X_i) \right) + \mathfrak{a} : (\mu_1, \dots, \mu_n) \in \mathbb{N}^n, 0 \leq \mu_i \leq d_i - 1, 1 \leq i \leq n \right\}.$$

Sea $f \in \Omega_{(d)}$ un polinomio y supongamos que $t := \deg(f)$. La expansión monomial de f tiene la siguiente forma:

$$f := \sum_{|\underline{\mu}| \leq t} a_{\underline{\mu}} X_1^{\mu_1} \dots X_n^{\mu_n}.$$

Dado que $f \in \Omega_{(d)}$, entonces existe $\underline{\theta} \in \Delta_{(d)}$ tal que $|\underline{\theta}| = \deg(f) = t$ y el coeficiente $f_{\underline{\theta}} = a_{\underline{\theta}} \neq 0$ es no nulo.

Sea $G_{\underline{\theta}} := \left(\prod_{i=1}^n g_{\theta_i}^{(i)}(X_i) \right) + \mathfrak{a} \in K[E]$. Por el Corolario 3.3.9 anterior, deducimos que para todo $\underline{\mu} \in \mathbb{N}^n$ tal que $\underline{\mu} \neq \underline{\theta}$, se verifica lo siguiente:

$$D_{\underline{\theta}, \underline{\mu}} = \langle G_{\underline{\theta}}, X_1^{\mu_1} \dots X_n^{\mu_n} + \mathfrak{a} \rangle_E = 0.$$

Como:

$$D_{\underline{\theta}, \underline{\theta}} = \langle G_{\underline{\theta}}, X_1^{\theta_1} \dots X_n^{\theta_n} + \mathfrak{a} \rangle_E = 1,$$

concluimos:

$$\langle G_{\underline{\theta}}, f + \mathfrak{a} \rangle_E = a_{\underline{\theta}} \neq 0.$$

Por otro lado, tenemos que:

$$a_\theta = \langle G_\theta, f + \mathbf{a} \rangle_E = \sum_{(z_1, \dots, z_n) \in E} \left(\prod_{i=1}^n g_{\theta_i}^{(i)}(z_i) \right) f(z_1, \dots, z_n).$$

Por lo tanto, concluimos que no es posible que f se anule en todos los puntos de E ya que, en otro caso, $a_\theta = 0$, lo cual no es posible. \square

3.4. Un algoritmo probabilista para detectar la dimensión genérica: “Suite Sécante”

En esta sección mostramos una aplicación algorítmica del uso de conjuntos cuestores que *difiere de los usuales Tests de Nulidad: el Problema de la “Suite Sécante”* (o Sucesión Secante). Usando las mismas notaciones que en las secciones precedentes, este problema puede formularse de la siguiente manera (para la notación $\mathcal{P}_{(d)}$, ver la Subsección 3.1.1):

PROBLEMA 4 (El Problema de la “Suite Sécante”). *Dado $f := (f_1, \dots, f_m) \in \mathcal{P}_{(d)}$, decidir si f es una “Suite Sécante”. Es decir, decidir si se verifica la siguiente propiedad:*

La variedad algebraica $V_{\mathbb{A}}(f_1, \dots, f_m) \subseteq \mathbb{A}^n(K)$ es una variedad no vacía de dimensión $n - m$.

No estudiamos únicamente el algoritmo que resuelve el problema anterior, sino que también analizamos su complejidad. Usaremos el modelo de computación de *Máquinas BSS sobre K* (cf. [BSS, 89], [BCSS, 98] y sus referencias). Distinguimos entre el *caso denso* y el *caso de entrada restringida*, donde las listas de polinomios de entrada pertenecerán a algún subconjunto constructible $\Omega \subseteq \mathcal{P}_{(d)}$. Tendremos en cuenta el número de operaciones aritméticas requeridas para evaluar los polinomios de la lista $f := (f_1, \dots, f_m)$. Esto se representa usando la codificación de *esquema de evaluación no escalar* de la lista de entrada f . Nótese que esta codificación es equivalente a *redes neuronales con funciones de activación $\varphi(t) = t^2$* (ver Ejemplo 5.3.2). Para más detalles sobre los esquemas de evaluación no escalares, dirigimos al lector a las Secciones 2 y 3 de [KP, 96].

Es importante señalar que en multitud de trabajos de *Geometría Algebraica Computacional* se ha tratado este problema (la bibliografía sobre este tema es muy extensa para citarlos a todos). *La principal novedad de nuestra contribución es que no realizamos ninguna operación (ni algebraica ni numérica) con los polinomios de entrada de la lista f : simplemente los evaluamos en algunos puntos seleccionados aleatoriamente* (ver Algoritmo 3.4.3 más abajo).

3.4.1. Algunos aspectos básicos sobre las “Suites Sécantes”. Comenzamos nuestra discusión exponiendo algunos resultados básicos sobre las “Suites Sécantes”. Incluimos las demostraciones para ser auto-contenidos.

En primer lugar, consideramos la siguiente variedad de incidencia $V_{(d)} \subseteq \mathcal{P}_{(d)} \times \mathbb{A}^n(K)$:

$$(3.4.1) \quad V_{(d)} := \{(f, x) \in \mathcal{P}_{(d)} \times \mathbb{A}^n(K) : f(x) = 0 \in \mathbb{A}^m(K)\}.$$

Podemos ver esta variedad de incidencia como la fibra del morfismo natural de evaluación. Es decir, consideramos el morfismo de evaluación definido como sigue:

$$(3.4.2) \quad \begin{array}{ccc} \text{ev}_{(d)} : \mathcal{P}_{(d)} \times \mathbb{A}^n(K) & \longrightarrow & \mathbb{A}^m(K) \\ (f, x) & \longmapsto & f(x) \end{array}.$$

Obviamente, $V_{(d)} := \text{ev}_{(d)}^{-1}(\{0\})$. Además, tenemos dos proyecciones canónicas que se definen de la siguiente forma:

- La proyección $\pi_1 : V_{(d)} \longrightarrow \mathcal{P}_{(d)}$ dada por $\pi_1(f, x) := f, \forall (f, x) \in V_{(d)}$.
- La proyección $\pi_2 : V_{(d)} \longrightarrow \mathbb{A}^n(K)$ dada por $\pi_2(f, x) := x, \forall (f, x) \in V_{(d)}$.

Nótese que para cada $f \in \mathcal{P}_{(d)}$, la fibra $\pi_1^{-1}(\{f\})$ puede identificarse con la variedad algebraica $V_{\mathbb{A}}(f)$ de ceros en $\mathbb{A}^n(K)$.

El siguiente resultado contiene las propiedades principales de esta variedad de incidencia.

TEOREMA 3.4.1. *Con las notaciones e hipótesis precedentes, tenemos que:*

- i) *La variedad de incidencia $V_{(d)}$ es una variedad lisa de intersección completa de dimensión $N_{(d)} + (n - m)$ y es pura (i.e. todas sus componentes irreducibles tienen la misma dimensión).*
- ii) *La proyección canónica π_2 es suprayectiva y la fibra $\pi_2^{-1}(\{x\})$ es una variedad afín lineal de dimensión $N_{(d)} - m$ contenida en $V_{(d)}$ e isomorfa a una variedad afín lineal de $\mathcal{P}_{(d)}$ de la misma dimensión.*
- iii) *Existe un subconjunto abierto Zariski $U_\emptyset \subseteq \mathcal{P}_{(d)}$ contenido en $\pi_1(V_{(d)})$ o, equivalentemente, existe una subvariedad algebraica propia cerrada $W_\emptyset \subseteq \mathcal{P}_{(d)}$, tal que para todo $f \in \mathcal{P}_{(d)} \setminus W_\emptyset = U_\emptyset$, $\pi_1^{-1}(\{f\}) = V_{\mathbb{A}}(f) \neq \emptyset$.*

iv) La proyección canónica π_1 es dominante en cada componente irreducible C de $V_{(d)}$ tal que $\pi_1(C)$ contiene una "Suite Sécante". Es más, para cada componente irreducible C de $V_{(d)}$ tenemos solo dos casos posibles:

- O bien existe $f \in \pi_1(C)$ tal que $\dim(\pi_1^{-1}(\{f\}) \cap C) = n - m$, en cuyo caso $\pi_1(C)$ es denso en $\mathcal{P}_{(d)}$ con respecto a la topología de Zariski.
- O para todo $f = (f_1, \dots, f_m) \in \pi_1(C)$, $\dim(\pi_1^{-1}(\{f\}) \cap C) \geq n - m + 1$ y la clausura Zariski de $\pi_1(C)$ está estrictamente contenida en $\mathcal{P}_{(d)}$. En este caso, f_1, \dots, f_m no es una "Suite Sécante".

En conclusión, existe un subconjunto abierto Zariski $U \subseteq \mathcal{P}_{(d)}$ tal que para cada $f = (f_1, \dots, f_m) \in U$, $V_{\mathbb{A}}(f) \neq \emptyset$ y es una variedad algebraica de dimensión $n - m$. Equivalentemente, para todo $f = (f_1, \dots, f_m) \in U$, la lista f_1, \dots, f_m es una "Suite Sécante".

DEMOSTRACIÓN. En primer lugar, consideramos el morfismo de evaluación $\text{ev}_{(d)}$. Resulta sencillo verificar que $\text{ev}_{(d)}$ es suprayectivo, ya que dados $v \in \mathbb{A}^m(K)$ y $f \in \pi_2^{-1}(\{0\})$, tenemos que $f + v \in \mathcal{P}_{(d)}$ y $\text{ev}_{(d)}(f + v, 0) = v$.

Por el Teorema de la Dimensión en la Fibra, como $\mathcal{P}_{(d)} \times \mathbb{A}^n(K)$ es irreducible, existe un subconjunto abierto Zariski $\mathcal{U} \subseteq \mathbb{A}^m(K)$ tal que para todo $v \in \mathcal{U}$, la fibra no vacía $\text{ev}_{(d)}^{-1}(\{v\})$ verifica:

$$\dim(\text{ev}_{(d)}^{-1}(\{v\})) = \dim(\mathcal{P}_{(d)} \times \mathbb{A}^n(K)) - m = N_{(d)} + (n - m).$$

Además, todas las fibras son identificables mediante la siguiente traslación:

$$\begin{aligned} \text{ev}_{(d)}^{-1}(\{v\}) &\longrightarrow \text{ev}_{(d)}^{-1}(\{0\}) \\ (f, x) &\longmapsto (f - v, x). \end{aligned}$$

Por lo tanto, concluimos que $V_{(d)} = \text{ev}_{(d)}^{-1}(\{0\})$ es una variedad algebraica de dimensión $N_{(d)} + (n - m)$. Consideramos dos conjuntos de variables para representar a los elementos de $\mathcal{P}_{(d)} \times \mathbb{A}^n(K)$. Por un lado, consideramos el conjunto de variables $\{X_1, \dots, X_n\}$ que representan las coordenadas de los puntos en $\mathbb{A}^n(K)$. Por otro lado, consideramos la familia de coeficientes genéricos de los polinomios en $\mathcal{P}_{(d)}$:

$$\{U_{\mu}^{(i)} : 1 \leq i \leq m, |\mu| \leq d_i\}.$$

Por consiguiente, los polinomios genéricos de P_{d_i} pueden describirse como sigue:

$$F_i := \sum_{|\mu| \leq d_i} U_{\mu}^{(i)} X_1^{\mu_1} \dots X_n^{\mu_n}.$$

Entonces, las ecuaciones que definen a $V_{(d)}$ pueden interpretarse como las siguientes m ecuaciones en estos dos conjuntos de variables:

$$F_i \left((U_{\mu}^{(i)} : 1 \leq i \leq m, |\mu| \leq d_i), X_1, \dots, X_n \right) = 0, \quad 1 \leq i \leq m,$$

donde los $F_i \in K[(U_{\mu}^{(i)} : 1 \leq i \leq m, |\mu| \leq d_i), X_1, \dots, X_n]$ son polinomios de grado $d_i + 1$ en estos dos conjuntos de variables. Consideramos también la matriz Jacobiana $D(F) := D(F_1, \dots, F_m)$ de estas ecuaciones con respecto a todas las variables involucradas. Esta matriz Jacobiana contiene una matriz identidad de tamaño $m \times m$:

$$\begin{pmatrix} \frac{\partial F_1}{\partial U_{(0)}^{(1)}} & \cdots & \frac{\partial F_1}{\partial U_{(0)}^{(m)}} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial U_{(0)}^{(1)}} & \cdots & \frac{\partial F_m}{\partial U_{(0)}^{(m)}} \end{pmatrix} = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix} = Id_m,$$

donde $(0) = (0, \dots, 0) \in \mathbb{N}^m$ es el exponente del término independiente. Por lo tanto, $\text{rango}(D(F)) = m$ y es independiente del punto $(f, x) \in V_{(d)}$ que estemos considerando. Por el Teorema del Criterio Jacobiano, concluimos que la variedad $V_{(d)}$ es lisa y que su espacio tangente $T_{(f,x)}V_{(d)}$ tiene dimensión $N_{(d)} + (n - m)$. Esto prueba el Ítem i).

Para el Ítem ii), basta observar que para cada $x \in \mathbb{A}^n(K)$, el subespacio $\{f \in P_{d_i} : f(x) = 0\}$ es un subespacio vectorial de co-dimensión 1 en P_{d_i} . El resto de la prueba de este ítem se sigue de forma inmediata.

En cuanto al Ítem iii), recordemos los *Sistemas Generalizados de Pham* como en [PSM, 04] y sus referencias. Sea $f := (f_1, \dots, f_m) \in \mathcal{P}_{(d)}$ una lista de polinomios. Trabajaremos en el subconjunto abierto Zariski $\mathcal{W}_{(d)} \subseteq \mathcal{P}_{(d)}$ de todas las listas de polinomios $f = (f_1, \dots, f_m) \in \mathcal{P}_{(d)}$ tales que $\deg(f_i) = d_i$ para cada i , $1 \leq i \leq m$.

Dado $f \in \mathcal{W}_{(d)}$, para cada i , $1 \leq i \leq m$, consideramos la componente homogénea de mayor grado de f_i . Es decir, si:

$$f_i := \sum_{|\mu| \leq d_i} a_\mu^{(i)} X_1^{\mu_1} \cdots X_n^{\mu_n},$$

entonces su parte homogénea de grado d_i viene dada por:

$$(f_i)_{d_i} := \sum_{|\mu|=d_i} a_\mu^{(i)} X_1^{\mu_1} \cdots X_n^{\mu_n}.$$

Consideramos ahora una matriz $M \in \mathcal{M}_{(n-m) \times n}(K)$ en el conjunto abierto Zariski $G((n-m) \times n, K)$ de todas las matrices en $\mathcal{M}_{(n-m) \times n}(K)$ de rango $n-m$ y un vector $b \in \mathbb{A}^{n-m}(K)$. Estos definen una lista de polinomios $\ell_1, \dots, \ell_m : \mathbb{A}^n(K) \rightarrow \mathbb{A}^{n-m}(K)$ de grado 1 mediante la siguiente identidad:

$$\begin{pmatrix} \ell_1(x) \\ \vdots \\ \ell_{n-m}(x) \end{pmatrix} := M \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} + b.$$

Entonces, tenemos un sistema ampliado $F := (f_1, \dots, f_m, \ell_1, \dots, \ell_{n-m})$ de ecuaciones polinomiales. Consideramos ahora las componentes homogéneas de mayor grado de todos ellos:

$$F_{\text{init}} := ((f_1)_{d_1}, \dots, (f_m)_{d_m}, (\ell_1)_1, \dots, (\ell_{n-m})_1).$$

Se trata de una familia de polinomios homogéneos de respectivos grados determinados por la lista $(\delta) := (d_1, \dots, d_m, 1, \dots, 1) \in \mathbb{N}^n$ en el conjunto de variables $\{X_1, \dots, X_n\}$. Al igual que en [PSM, 04], decimos que F es un *Sistema Generalizado de Pham* si el conjunto de ceros comunes en el espacio proyectivo $\mathbb{P}_{n-1}(K)$ de los polinomios en la lista F_{init} es vacío. Por lo tanto, gracias a la Teoría clásica de Resultantes Multivariados (cf. [Jo, 97], [CD, 05], [FP, 13], [BJ, 14] y sus referencias), sabemos que existe un polinomio no nulo en los coeficientes de $U_\mu^{(i)}$ de los polinomios y las coordenadas $T_{r,s}$ de las matrices $M \in \mathcal{M}_{(n-m) \times n}(K)$,

$$\text{Res}_{(\delta)} \in \mathbb{Z}[(U_\mu^{(i)} : 1 \leq i \leq m, |\mu| = d_i), (T_{r,s} : 1 \leq r \leq n-m, 1 \leq s \leq n)],$$

tal que:

$$\text{Res}_{(\delta)}((f_1)_{d_1}, \dots, (f_m)_{d_m}, (\ell_1)_1, \dots, (\ell_{n-m})_1) = 0 \iff V_{\mathbb{P}_{n-1}(K)}(F_{\text{init}}) \neq \emptyset,$$

donde $\text{Res}_{(\delta)}((f_1)_{d_1}, \dots, (f_m)_{d_m}, (\ell_1)_1, \dots, (\ell_{n-m})_1)$ representa el resultado de evaluar $\text{Res}_{(\delta)}$ en los coeficientes de los polinomios de la lista F_{init} y $V_{\mathbb{P}_{n-1}(K)}(F_{\text{init}})$ es el conjunto de ceros proyectivos en $\mathbb{P}_{n-1}(K)$ de la lista de ecuaciones polinómicas homogéneas en F_{init} . Esto implica que el siguiente conjunto (formado por Sistemas Generalizados de Pham) es un subconjunto abierto Zariski en $\mathcal{P}_{(d)} \times \mathcal{M}_{(n-m) \times n}(K) \times \mathbb{A}^{n-m}(K)$:

$$(3.4.3) \quad \begin{aligned} \mathcal{G}_{(d)} &:= \{F := (f_1, \dots, f_m, \ell_1, \dots, \ell_{n-m}) \in \mathcal{P}_{(d)} \times \\ &\times \mathcal{M}_{(n-m) \times n}(K) \times \mathbb{A}^{n-m}(K) : (f_1, \dots, f_m) \in \mathcal{W}_{(d)}, \\ &\text{y } F \text{ es un Sistema Generalizado de Pham}\}. \end{aligned}$$

Consideramos ahora la homogeneización con respecto a la nueva variable X_0 de una lista de ecuaciones polinómicas $F := (f_1, \dots, f_m, \ell_1, \dots, \ell_{n-m}) \in \mathcal{G}_{(d)}$. Denotamos por ${}^h F$ a esta nueva lista de ecuaciones polinómicas homogéneas en el conjunto de variables $\{X_0, \dots, X_n\}$. Como el número de ecuaciones homogéneas n y la dimensión de $\mathbb{P}_n(K)$ son iguales, la variedad de los ceros de ${}^h F$ en $\mathbb{P}_n(K)$ es no vacía y, como $F \in \mathcal{G}_{(d)}$, ningún cero de ${}^h F$ se encuentra en el “hiperplano del infinito”:

$$H_\infty := \{(x_0 : x_1 : \dots : x_n) \in \mathbb{P}_n(K) : x_0 = 0\}.$$

En otras palabras, existe un subconjunto abierto Zariski $\mathcal{G}_{(d)} \subseteq \mathcal{P}_{(d)} \times \mathcal{M}_{(n-m) \times n}(K) \times \mathbb{A}^{n-m}(K)$ tal que para todo $F \in \mathcal{G}_{(d)}$, el conjunto de ceros afines en $\mathbb{A}^n(K)$ de F coincide con el conjunto de ceros proyectivos de ${}^h F$ en $\mathbb{P}_n(K)$, el cual es no vacío:

$$V_{\mathbb{A}}(f_1, \dots, f_m, \ell_1, \dots, \ell_{n-m}) \cong V_{\mathbb{P}_n(K)}({}^h F) \neq \emptyset.$$

En particular, existe un polinomio no nulo en la lista de coeficientes $\mathcal{U} := (U_\mu^{(i)} : 1 \leq i \leq m, |\mu| \leq d_i)$ de las ecuaciones en $\mathcal{P}_{(d)}$, las coordenadas $\mathcal{M} := (T_{r,s} : 1 \leq r \leq n-m, 1 \leq s \leq n)$ de las matrices en $\mathcal{M}_{(n-m) \times n}(K)$ y las coordenadas $\mathcal{Z} := (Z_1, \dots, Z_{n-m})$ de los puntos en $\mathbb{A}^{n-m}(K)$:

$$P(\mathcal{U}, \mathcal{M}, \mathcal{Z}) \in K[\mathcal{U}, \mathcal{M}, \mathcal{Z}] \setminus \{0\},$$

de modo que si $P(f_1, \dots, f_m, \ell_1, \dots, \ell_{n-m}) \neq 0$, entonces la variedad:

$$V_{\mathbb{A}}(f_1, \dots, f_m, \ell_1, \dots, \ell_{n-m}) \neq \emptyset.$$

Como K es un cuerpo infinito, existe alguna matriz $M \in \mathcal{M}_{(n-m) \times n}(K)$ y algún punto $\zeta \in \mathbb{A}^{n-m}(K)$ tal que el polinomio:

$$Q(\mathcal{U}) := P(\mathcal{U}, M, \zeta) \in K[\mathcal{U}] \setminus \{0\}.$$

Por consiguiente, hemos demostrado que:

$$\forall f := (f_1, \dots, f_m) \in \mathcal{P}_{(d)}, Q(f_1, \dots, f_m) \neq 0 \implies V_{\mathbb{A}}(f_1, \dots, f_m) \neq \emptyset.$$

Por lo tanto, el siguiente subconjunto abierto Zariski está contenido en $\pi_1(V_{(d)})$ y el Ítem *iii*) se sigue de forma inmediata:

$$\pi_1(V_{(d)}) \supseteq \{f = (f_1, \dots, f_m) \in \mathcal{P}_{(d)} : Q(f_1, \dots, f_m) \neq 0\}.$$

Sea U_\emptyset un subconjunto abierto Zariski no vacío de $\mathcal{P}_{(d)}$ contenido en $\pi_1(V_{(d)})$. Entonces, para cada $f \in U_\emptyset$, $V_{\mathbb{A}}(f) \neq \emptyset$.

A continuación, para el Ítem *iv*), consideramos una componente irreducible C de $V_{(d)}$ y tomamos su proyección $\pi_1(C) \subseteq \mathcal{P}_{(d)}$. Consideramos también la siguiente restricción:

$$\pi_1|_C : C \longrightarrow \overline{\pi_1(C)}^z.$$

Se trata de un morfismo dominante y, por el Teorema de la Dimensión en la Fibra, concluimos que para cualquier $f \in \pi_1(C)$ se cumple que:

$$\dim(\pi_1^{-1}(\{f\}) \cap C) = \dim(\pi_1|_C^{-1}(\{f\})) \geq \dim(C) - \dim(\overline{\pi_1(C)}^z).$$

Si $\dim(\pi_1^{-1}(\{f\}) \cap C) = n - m$, entonces tenemos que:

$$\dim(\overline{\pi_1(C)}^z) \geq \dim(C) - (n - m) = N_{(d)} + (n - m) - (n - m) = N_{(d)},$$

y, por lo tanto, $\pi_1(C)$ es denso Zariski en $\mathcal{P}_{(d)}$.

En caso contrario, existe un subconjunto abierto Zariski $U_C \subseteq \pi_1(C) \subseteq \overline{\pi_1(C)}^z$ tal que para todo $g \in U_C$ tenemos que:

$$\dim(\pi_1^{-1}(\{g\}) \cap C) = \dim(\pi_1|_C^{-1}(\{g\})) = \dim(C) - \dim(\overline{\pi_1(C)}^z).$$

Si $\dim(\pi_1^{-1}(\{g\}) \cap C) \geq n - m + 1$, para todo $g \in \pi_2(C)$ concluimos:

$$\dim(\overline{\pi_1(C)}^z) = \dim(C) - \dim(\pi_1^{-1}(\{g\}) \cap C) \leq N_{(d)} + (n - m) - (n - m + 1) = N_{(d)} - 1,$$

y, por consiguiente, $\pi_1(C)$ no sería denso Zariski en $\mathcal{P}_{(d)}$.

En cuanto a la última afirmación, basta demostrar que la clase:

$$\mathcal{S} := \{f \in \mathcal{P}_{(d)} : f \text{ es una "Suite Sécante"}\},$$

contiene un subconjunto abierto Zariski no vacío. Sea \mathcal{C} la clase de todas las componentes irreducibles de $V_{(d)}$ tales que existe algún $f \in \pi_1(C)$ que satisface $\dim(\pi_1^{-1}(f) \cap C) = n - m$. Sea \mathcal{W} la clase de todas las componentes irreducibles de $V_{(d)}$ tales que $\pi_1(W)$ no es denso Zariski en $\mathcal{P}_{(d)}$.

Entonces, consideramos el subconjunto abierto Zariski no vacío de $\mathcal{P}_{(d)}$ dado por la siguiente identidad:

$$A := \mathcal{P}_{(d)} \setminus \left(\bigcup_{C \in \mathcal{W}} \overline{\pi_1(C)}^z \right).$$

Para cada $C \in \mathcal{C}$, tenemos que $\pi_1(C)$ es denso Zariski en $\mathcal{P}_{(d)}$. Entonces, por el Teorema de la Dimensión en la Fibra, para cada $C \in \mathcal{C}$ existe un abierto Zariski $U_C \subseteq \mathcal{P}_{(d)}$ tal que para todo $f \in U_C$, tenemos que:

$$\dim(\pi_1^{-1}(\{f\}) \cap C) = \dim(\pi_1|_C^{-1}(\{f\})) = \dim(C) - \dim(\mathcal{P}_{(d)}) = N_{(d)} + (n - m) - N_{(d)} = n - m.$$

Como $\mathcal{P}_{(d)}$ es irreducible, el siguiente subconjunto abierto $U \subseteq \mathcal{P}_{(d)}$ es no vacío:

$$U := A \cap U_\emptyset \cap \left(\bigcap_{C \in \mathcal{C}} U_C \right).$$

Además, para cada $f \in U$ tenemos que:

- Como $f \in U_\emptyset$, se cumple que $V_{\mathbb{A}}(f) \neq \emptyset$.

- Como $f \in A$, entonces $f \notin \pi_1(D)$ para cualquier $D \in \mathcal{W}$ o, equivalentemente, tenemos que $\pi_1^{-1}(\{f\}) \cap D = \emptyset$ para cada $D \in \mathcal{W}$. En particular, se verifican las siguientes igualdades:

$$\begin{aligned} V_{\mathbb{A}}(f) = \pi_1^{-1}(\{f\}) &= \left(\bigcup_{D \in \mathcal{W}} (\pi_1^{-1}(\{f\}) \cap D) \right) \cup \left(\bigcup_{C \in \mathcal{C}} (\pi_1^{-1}(\{f\}) \cap C) \right) = \\ &= \bigcup_{C \in \mathcal{C}} (\pi_1^{-1}(\{f\}) \cap C). \end{aligned}$$

- Además, como $f \in U_C$, para cada $C \in \mathcal{C}$, $\dim(\pi_1^{-1}(\{f\}) \cap C) = n - m$. Por lo tanto, para cada $f \in \mathcal{P}_{(d)}$ tenemos que:

$$\dim(V_{\mathbb{A}}(f)) = \dim(\pi_1^{-1}(\{f\})) = \max\{\dim(\pi_1^{-1}(\{f\}) \cap C) : C \in \mathcal{C}\} = n - m.$$

Es decir, para cada $f \in U$, $V_{\mathbb{A}}(f)$ es una variedad de intersección completa de dimensión $n - m$ y, por lo tanto, f es una “Suite Sécante” y se verifica la última afirmación. \square

3.4.2. El Algoritmo. Comenzamos esta subsección definiendo la clase de cuerpos que son adecuados para muestrear conjuntos cuestores.

DEFINICIÓN 20 (Cuerpos que son adecuados para muestrear Conjuntos Cuestores). *Decimos que un cuerpo K es adecuado para muestrear conjuntos cuestores en variedades cero-dimensionales si satisface la siguiente propiedad: para cada $R \in \mathbb{N}$ y para cada entero positivo $n \in \mathbb{N}$, existe una variedad cero-dimensional $Z_R \subseteq \mathbb{A}^n(K)$ de grado R^n , dada por ecuaciones polinómicas de grado a lo sumo R y de manera que la siguiente tarea pueda realizarse con a lo sumo $O(n \log_2(R))$ operaciones aritméticas:*

muestrea aleatoriamente $x \in Z_R$

con respecto a la distribución uniforme en Z_R .

Los cuerpos de característica cero son, evidentemente, adecuados para muestrear conjuntos cuestores. Basta con considerar el subconjunto:

$$\{1, \dots, R\} \subset \mathbb{Z} \subseteq K$$

y la variedad $Z_R := \{1, \dots, R\}^n$. Muchos otros cuerpos satisfacen una propiedad similar.

El poder restringido de los conjuntos cuestores nos permite probar resultados como el siguiente:

TEOREMA 3.4.2. *Sea K un cuerpo adecuado para muestrear conjuntos cuestores. Sea $(d) := (d_1, \dots, d_m)$ una lista de grados y sea $\Omega \subseteq \mathcal{P}_{(d)}^K$ un subconjunto constructible de listas de polinomios. Sea $U \subseteq \mathcal{P}_{(d)}$ el subconjunto abierto Zariski descrito en el Teorema 3.4.1 de listas que son “Suites Sécantes”. Supongamos que $\Omega \setminus U$ tiene dimensión a lo sumo $\dim(\Omega) - 1$. Entonces, existe un algoritmo probabilista para el problema “Suite Sécante” con entradas restringidas a Ω que satisface las siguientes propiedades:*

i) *Dados como entrada $f \in \Omega$ y las cantidades $\dim(\Omega)$ y $\deg_{\text{lci}}(\Omega)$:*

- *Si $f \in \Omega \cap U$, el algoritmo responde “ f es una Suite Sécante” con probabilidad mayor que:*

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{m \dim(\Omega)}}.$$

- *Si el algoritmo responde “ f no es una Suite Sécante”, entonces $f \in \Omega \setminus U$ (i.e. f no es “Suite Sécante”) con probabilidad mayor que:*

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega) e^{m \dim(\Omega)}}.$$

ii) *Si $\mathcal{B} \subseteq 2^\Omega$ es una σ -álgebra que contiene a los subconjuntos de Borel de Ω con respecto a la topología de Zariski y $\mu : \mathcal{B} \rightarrow [0, 1]$ es una distribución de probabilidad sobre \mathcal{B} , entonces:*

$$\mu[f \in \Omega : \text{el algoritmo responde de forma incorrecta sobre } f]$$

es a lo sumo:

$$\frac{1}{\deg_{\text{lci}}(\Omega) e^{m \dim(\Omega)}} + \mu[\Omega \setminus U].$$

iii) *El tiempo de ejecución del algoritmo en términos de operaciones aritméticas es a lo sumo:*

$$O(\dim(\Omega)n((T + \log(\dim(\Omega)) + \log(d)) + Tn \log(\deg_{\text{lci}}(\Omega))),$$

donde T es el número máximo de operaciones aritméticas requeridas para evaluar cualquier lista $f := (f_1, \dots, f_m) \in \Omega$ en un punto y $d := \max\{d_1, \dots, d_m\}$.

DEMOSTRACIÓN. El algoritmo se basa en la noción de conjunto cuestor y en los resultados descritos en la Sección 3.2 anterior (especialmente en el Corolario 3.2.6).

ALGORITMO 3.4.3 (“Suite Sécante” por evaluación de los polinomios de entrada en puntos de muestreo).

INPUT: Una lista de polinomios $f := (f_1, \dots, f_m) \in \Omega \subseteq \mathcal{P}_{(d)}^K$, y las cantidades $\dim(\Omega)$ y $\deg_{\text{ici}}(\Omega)$.

eval:

- $L := 6 \dim(\Omega)$ y,
- $R := \max\{[(e(d+1))^2], \lfloor \deg_{\text{ici}}(\Omega)^{\frac{2}{\dim(\Omega)}} \rfloor\}$.

Como K es adecuado para muestrear conjuntos cuestores, sea $Z_R \subseteq \mathbb{A}^n(K)$ la correspondiente variedad cero-dimensional de grado R^n .

muestrea aleatoriamente una lista $\mathbf{Q} := (x_1, \dots, x_L) \in Z_R^L$.

if $f(x_1) = f(x_2) = \dots = f(x_L) = 0 \in K^m$, then OUPUT: **No**.
else, OUTPUT: **Sí**.

fi

end

El tiempo de ejecución de este algoritmo (en términos de operaciones aritméticas) es del siguiente orden:

- Tiempo $O(Ln \log(\dim(\Omega)) + \log(d) + \frac{\log(\deg_{\text{ici}}(\Omega))}{\dim(\Omega)})$ para generar L , R y para realizar el muestreo de la lista $\mathbf{Q} := (x_1, \dots, x_L) \in Z_R^L$, dado que K es adecuado para muestrear conjuntos cuestores.
- Tiempo $O(LT)$ para evaluar los m polinomios de la lista de entrada $f := (f_1, \dots, f_m)$ en los L puntos de la lista \mathbf{Q} .

Como $L = 6 \dim(\Omega)$, el número total de operaciones aritméticas está acotado por:

$$O(\dim(\Omega)n((T + \log(\dim(\Omega)) + \log(d)) + Tn \log(\deg_{\text{ici}}(\Omega))).$$

A continuación, realizamos el análisis probabilístico del algoritmo. De acuerdo con el Corolario 3.2.6, bajo nuestras hipótesis, la probabilidad de que una lista escogida aleatoriamente $\mathbf{Q} = (x_1, \dots, x_L) \in Z_R^L$ sea un conjunto cuestor para Ω con respecto a $\Omega \setminus U$ es mayor que:

$$1 - \frac{1}{\deg_{\text{ici}}(\Omega)e^{\dim(\Omega)+(m-1)L}} = 1 - \frac{1}{\deg_{\text{ici}}(\Omega)e^{(6m-5)\dim(\Omega)}} \geq 1 - \frac{1}{\deg_{\text{ici}}(\Omega)e^{m \dim(\Omega)}}$$

Por lo tanto, tenemos los siguientes casos:

- Si $f \in \Omega \cap U$, entonces la probabilidad de que el algoritmo responda: “Sí, es una Suite Sécante” coincide con la probabilidad de que \mathbf{Q} sea un conjunto cuestor para Ω con respecto a $\Omega \setminus U$ y, por lo tanto, es mayor que:

$$1 - \frac{1}{\deg_{\text{ici}}(\Omega)e^{m \dim(\Omega)}}$$

- Si el algoritmo responde: “No, no es una Suite Sécante”, entonces la probabilidad de que $f \in \Omega \setminus U$ es mayor que la probabilidad de que \mathbf{Q} sea un conjunto cuestor y, por lo tanto, mayor que:

$$1 - \frac{1}{\deg_{\text{ici}}(\Omega)e^{m \dim(\Omega)}}$$

Finalmente, el algoritmo responde de forma incorrecta si $f \in \Omega \cap U$ y la respuesta es incorrecta o si $f \in \Omega \setminus U$ y la respuesta es incorrecta. La probabilidad de la unión de estos dos eventos es a lo sumo:

$$\frac{1}{\deg_{\text{ici}}(\Omega)e^{m \dim(\Omega)}} + \mu[\Omega \setminus U].$$

□

OBSERVACIÓN 3.4.4. Sea $\mathcal{B} \subseteq 2^\Omega$ una σ -álgebra que contiene a los subconjuntos de Borel de Ω con respecto a la topología de Zariski. Si K es un cuerpo algebraicamente cerrado adecuado para muestrear conjuntos cuestores, la clausura Zariski de Ω es irreducible y $\dim(\Omega \setminus U) < \dim(\Omega)$, entonces podemos introducir una distribución de probabilidad $\mu : \mathcal{B} \rightarrow [0, 1]$ “sencilla” en Ω como sigue:

$$\mu[B] := \begin{cases} 1, & \text{si } \overline{B^z} = \Omega \\ 0, & \text{en otro caso} \end{cases}.$$

Bajo estas condiciones, $\mu[\Omega \setminus U] = 0$. Por lo tanto, el algoritmo anterior satisface las siguientes propiedades:

- i) Dados como entrada $f \in \Omega$ y las cantidades $\dim(\Omega)$ y $\deg_{\text{ici}}(\Omega)$:

- Si $f \in \Omega \cap U$, el algoritmo responde “ f es una Suite Sécante” con probabilidad mayor que:

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega)e^{m \dim(\Omega)}}.$$

- Si el algoritmo responde “ f no es una Suite Sécante”, entonces $f \in \Omega \setminus U$ (i.e. f no es una “Suite Sécante”) con probabilidad mayor que:

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega)e^{m \dim(\Omega)}}.$$

- ii) Además, la probabilidad de que dado $f \in \Omega$, el algoritmo responda de forma incorrecta es a lo sumo:

$$\frac{1}{\deg_{\text{lci}}(\Omega)e^{m \dim(\Omega)}}.$$

- iii) El tiempo de ejecución del algoritmo en términos de las operaciones aritméticas es a lo sumo:

$$O(\dim(\Omega)n((T + \log(\dim(\Omega)) + \log(d)) + Tn \log(\deg_{\text{lci}}(\Omega))).$$

donde T es el número máximo de operaciones aritméticas requeridas para evaluar cualquier lista $f := (f_1, \dots, f_m) \in \Omega$ en un punto y $d := \max\{d_1, \dots, d_m\}$.

En el caso $m = 1$, el Hauptidealsatz de Krull implica que el Problema de la “Suite Sécante” se convierte en el “Test de Nulidad de Polinomios”. Por lo tanto, en el caso $m = 1$, el Teorema 3.4.2 anterior se convierte en el siguiente corolario:

COROLARIO 3.4.5 (Caso $m = 1$). *Sea K un cuerpo adecuado para muestrear conjuntos cuestores. Sea $d \in \mathbb{N}$ un entero positivo, $\Omega \subseteq P_d^K$ un subconjunto constructible y $\Sigma = \{0\}$. Supongamos que $\dim(\Omega) \geq 1$. Entonces, existe un algoritmo probabilista para el “Test de Nulidad de Polinomios” con entradas restringidas a Ω con respecto a $\{0\}$ que satisface las siguientes propiedades:*

- i) *Dados como entrada $f \in \Omega$ y las cantidades $\dim(\Omega)$ y $\deg_{\text{lci}}(\Omega)$:*

- *Si el algoritmo responde “ f no es cero”, entonces $f \in \Omega \setminus \{0\}$ (i.e. f es un polinomio no nulo).*
- *Si el algoritmo responde “ $f \equiv 0$ ”, entonces $f \equiv 0$ con probabilidad mayor que:*

$$1 - \frac{1}{\deg_{\text{lci}}(\Omega)e^{\dim(\Omega)}}.$$

- ii) *El tiempo de ejecución del algoritmo en términos de operaciones aritméticas es a lo sumo:*

$$O(\dim(\Omega)n((T + \log(\dim(\Omega)) + \log(d)) + Tn \log(\deg_{\text{lci}}(\Omega))),$$

donde T es el número máximo de operaciones aritméticas requeridas para evaluar cualquier $f \in \Omega$ en un punto.

Nótese que el mismo algoritmo tiene una “naturaleza” diferente en el caso $m = 1$ y que la corrección de la salida es mejor. Este algoritmo pone el “Test de Nulidad de Polinomios” en la clase \mathbf{RP}_K , lo cual no ocurre con el problema de la “Suite Sécante”.

El algoritmo anterior generaliza los resultados previos de [HS, 82]. En dicho trabajo, los conjuntos cuestores estaban restringidos a variedades cero-dimensionales de tipo “reticular” (de la forma $\{1, \dots, u\}^n$), mientras que nuestra formulación permite el muestreo en cualquier clase de variedades de grado y co-dimensión adecuados con alta probabilidad de éxito.

COROLARIO 3.4.6 (“Suite Sécante” con polinomios de entrada densos). *Sea K un cuerpo adecuado para muestrear conjuntos cuestores. Sea $(d) := (d_1, \dots, d_m)$ una lista de grados. Sea $\Omega = \mathcal{P}_{(d)}^K$ la clase de todas las listas de grados respectivos a lo sumo (d) y denotamos por $N_{(d)}$ a su dimensión como K -espacio vectorial. Sea $U \subseteq \mathcal{P}_{(d)}$ el subconjunto abierto Zariski descrito en el Teorema 3.4.1 de listas que son “Suites Sécantes”. Supongamos que $\Omega \setminus U$ tiene dimensión a lo sumo $\dim(\Omega) - 1$. Entonces, existe un algoritmo probabilista para el problema “Suite Sécante” con codificación densa de los polinomios de entrada que satisface las siguientes propiedades:*

- i) *Dado como entrada $f \in \mathcal{P}_{(d)}^K$:*

- *Si $f \in \mathcal{P}_{(d)}^K \cap U$, el algoritmo responde “ f es una Suite Sécante” con probabilidad mayor que:*

$$1 - \frac{1}{e^{mN_{(d)}}}.$$

- Si el algoritmo responde “ f no es una Suite Sécante”, entonces $f \in \mathcal{P}_{(d)}^K \setminus U$ (i.e. f no es una “Suite Sécante”) con probabilidad mayor que:

$$1 - \frac{1}{e^{mN_{(d)}}}.$$

- ii) Si $\mathcal{B} \subseteq 2^{\mathcal{P}_{(d)}^K}$ es una σ -álgebra que contiene a los subconjuntos de Borel de $\mathcal{P}_{(d)}^K$ con respecto a la topología de Zariski y $\mu : \mathcal{B} \rightarrow [0, 1]$ es una distribución de probabilidad sobre \mathcal{B} , entonces:

$$\mu[f \in \mathcal{P}_{(d)}^K : \text{el algoritmo responde de forma incorrecta sobre } f]$$

es a lo sumo:

$$\frac{1}{e^{mN_{(d)}}} + \mu[\mathcal{P}_{(d)}^K \setminus U].$$

- iii) El tiempo de ejecución del algoritmo en términos de operaciones aritméticas es a lo sumo:

$$O\left(nN_{(d)}^2\right).$$

Además, si K es un cuerpo algebraicamente cerrado adecuado para muestrear conjuntos cuestores, la Propiedad ii) de este algoritmo se convierte en:

- ii) La probabilidad de que dado $f \in \mathcal{P}_{(d)}^K$, el algoritmo responda de forma incorrecta es a lo sumo:

$$\frac{1}{e^{mN_{(d)}}}.$$

DEMOSTRACIÓN. El resultado se sigue de forma inmediata del Teorema 3.4.2 teniendo en cuenta que $\dim(\Omega) = N_{(d)}$ y $\deg_{\text{lc}}(\Omega) = 1$. \square

3.4.3. Clase de todas las listas de polinomios evaluables por un esquema de evaluación.

Consideramos ahora el caso en el que los polinomios de entrada vienen dados por algún esquema de evaluación (o una red neuronal con función de activación $\varphi(t) := t^2$). Es decir, consideramos el siguiente problema:

PROBLEMA 5 (“Suite Sécante” con entradas dadas por un esquema de evaluación).

INPUT:

- Un esquema de evaluación Γ de talla S y profundidad ℓ , con m nodos de salida en el conjunto de variables $\{X_1, \dots, X_n\}$.
- Un conjunto constructible de instanciaciones paramétricas $\Lambda \subseteq \mathbb{A}^N(K)$ (donde N denota el número de parámetros). Para cada $\underline{a} \in \Lambda$, consideramos la lista de polinomios $f_{\underline{a}} := (f_1, \dots, f_m)$ dada por:

$$f_i := f_i^\Gamma(\underline{a}, X_1, \dots, X_n) \in K[X_1, \dots, X_n], \quad 1 \leq i \leq m.$$

- Supongamos que la clase $\mathcal{W}(\Gamma, \Lambda) := \{f_{\underline{a}} : \underline{a} \in \Lambda\}$ verifica que, genéricamente, para cada lista $f = (f_1, \dots, f_m) \in \mathcal{W}(\Gamma, \Lambda)$, la dimensión de la variedad $V_{\mathbb{A}}(f_1, \dots, f_m)$ es $n - m$.

OUTPUT: Sí si y solo si $\dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m$. NO, en otro caso.

COROLARIO 3.4.7 (Interpretación en términos de esquemas de evaluación del Problema de la “Suite Sécante”). Sea K un cuerpo adecuado para muestrear conjuntos cuestores. Supongamos que nos proporcionan la siguiente información:

- Un esquema de evaluación Γ de talla S y profundidad ℓ , con m nodos de salida en el conjunto de variables $\{X_1, \dots, X_n\}$.
- Un conjunto constructible de instanciaciones paramétricas $\Lambda \subseteq \mathbb{A}^N(K)$ (donde N denota el número de parámetros). Para cada $\underline{a} \in \Lambda$, consideramos la lista de polinomios $f_{\underline{a}} := (f_1, \dots, f_m)$ dada por:

$$f_i := f_i^\Gamma(\underline{a}, X_1, \dots, X_n) \in K[X_1, \dots, X_n], \quad 1 \leq i \leq m.$$

- Sea $\mathcal{W}(\Gamma, \Lambda) := \{f_{\underline{a}} : \underline{a} \in \Lambda\}$ y sea $U \subseteq \mathcal{P}_{(2^\ell)}$ el subconjunto abierto Zariski descrito en el Teorema 3.4.1 de listas que son “Suites Sécantes”. Supongamos que $\mathcal{W}(\Gamma, \Lambda) \setminus U$ tiene dimensión a lo sumo $\dim(\mathcal{W}(\Gamma, \Lambda)) - 1$.

Entonces, existe un algoritmo probabilista para el problema de la “Suite Sécante” con entradas dadas por un esquema de evaluación que satisface las siguientes propiedades:

- i) Dados como entrada $f \in \mathcal{W}(\Gamma, \Lambda)$ y las cantidades $\dim(\Lambda)$ y $\deg_{\text{lc}}(\Lambda)$:
 - Si $f \in \mathcal{W}(\Gamma, \Lambda) \cap U$, el algoritmo responde “ f es una Suite Sécante” con probabilidad mayor que:

$$1 - \frac{1}{\deg_{\text{lc}}(\Lambda) e^{\dim(\Lambda)(\log(2^{\ell+1}-2)+m)}}.$$

- Si el algoritmo responde “*f* no es una Suite Sécante”, entonces $f \in \mathcal{W}(\Gamma, \Lambda) \setminus U$ (i.e. *f* no es una “Suite Sécante”) con probabilidad mayor que:

$$1 - \frac{1}{\deg_{\text{lci}}(\Lambda) e^{\dim(\Lambda)(\log(2^{\ell+1}-2)+m)}}.$$

- ii) Si $\mathcal{B} \subseteq 2^{\mathcal{W}(\Gamma, \Lambda)}$ es una σ -álgebra que contiene a los subconjuntos de Borel de $\mathcal{W}(\Gamma, \Lambda)$ con respecto a la topología de Zariski y $\mu : \mathcal{B} \rightarrow [0, 1]$ es una distribución de probabilidad sobre \mathcal{B} , entonces:

$$\mu[f \in \mathcal{W}(\Gamma, \Lambda) : \text{el algoritmo responde de forma incorrecta sobre } f]$$

es a lo sumo:

$$\frac{1}{\deg_{\text{lci}}(\Lambda) e^{\dim(\Lambda)(\log(2^{\ell+1}-2)+m)}} + \mu[\mathcal{W}(\Gamma, \Lambda) \setminus U].$$

- iii) El tiempo de ejecución del algoritmo en términos de operaciones aritméticas es a lo sumo:

$$O(\dim(\Lambda)n^2 S^2((\ell + 1) \dim(\Lambda) + \log(\deg_{\text{lci}}(\Lambda)))).$$

Además, si K es un cuerpo algebraicamente cerrado adecuado para muestrear conjuntos cuestores, la Propiedad ii) de este algoritmo se convierte en:

- ii) La probabilidad de que dado $f \in \mathcal{W}(\Gamma, \Lambda)$, el algoritmo responda de forma incorrecta es a lo sumo:

$$\frac{1}{\deg_{\text{lci}}(\Lambda) e^{\dim(\Lambda)(\log(2^{\ell+1}-2)+m)}}.$$

DEMOSTRACIÓN. La demostración combina el Corolario 3.2.7, el Teorema 3.4.2 y los resultados de las Secciones 2 y 3 de [KP, 96] (ver también [HS, 80] y [HS, 82]). En primer lugar, nótese que $\mathcal{W}(\Gamma, \Lambda)$ es un conjunto constructible unirracional en $\mathcal{P}_{(2^\ell)}^K$ dado como la imagen del conjunto constructible Λ por una lista de polinomios de grado a lo sumo $D = (2^{\ell+1} - 2)$. Al igual que en el Corolario 3.2.7, consideramos su clausura Zariski $\Omega := \overline{\mathcal{W}(\Gamma, \Lambda)}^z$ para estimar la probabilidad de error. Nótese que Ω es una variedad algebraica irreducible (para una demostración detallada de este hecho, ver la Sección 3.1 de [KP, 96]). Dado que los conjuntos cuestores son hereditarios (cf. Proposición 3.1.3), tenemos que cualquier conjunto cuestor para Ω es también un conjunto cuestor para $\mathcal{W}(\Gamma, \Lambda)$. Combinando el Corolario 2.2.4 con los Lemas 14 y 16 de [KP, 96], obtenemos las siguientes desigualdades:

$$\deg(\Omega) = \deg(\overline{\mathcal{W}(\Gamma, \Lambda)}^z) \leq \deg_{\text{lci}}(\Lambda) D^{\dim(\Lambda)} \text{ y}$$

$$\dim(\Omega) = \dim(\mathcal{W}(\Gamma, \Lambda)) \leq \dim(\Lambda).$$

Como en la demostración del Teorema 3.4.2, consideramos el siguiente algoritmo (basado en la noción de conjunto cuestor y en los resultados de la Sección 3.2 anterior, especialmente, en el Corolario 3.2.7):

ALGORITMO 3.4.8 (“Suite Sécante” mediante evaluación de polinomios dados por un esquema de evaluación en puntos de muestro).

INPUT: Una lista de polinomios $f := (f_1, \dots, f_m) \in \mathcal{W}(\Gamma, \Lambda) \subseteq \mathcal{P}_{(2^\ell)}^K$, y las cantidades $\dim(\Lambda)$ y $\deg_{\text{lci}}(\Lambda)$.

eval:

- $L := 6 \dim(\Lambda)$ y,
- $R := \max\{\lfloor (e(d+1))^2 \rfloor, \lfloor \deg_{\text{lci}}(\Lambda) \frac{2}{\dim(\Lambda)} D^2 \rfloor\}$.

Como K es adecuado para muestrear conjuntos cuestores, sea $Z_R \subseteq \mathbb{A}^n(K)$ la correspondiente variedad cero-dimensional de grado R^n .

muestrea aleatoriamente una lista $\mathbf{Q} := (x_1, \dots, x_L) \in Z_R^L$.

if $f(x_1) = f(x_2) = \dots = f(x_L) = 0 \in K^m$, then OUPUT: No.

else, OUTPUT: Sí.

fi

end

El coste de evaluar un elemento de $\mathcal{W}(\Gamma, \Lambda)$ en un punto está acotado por $O(S^2)$. Según Teorema 3.4.2, el número total de operaciones aritméticas, reemplazando las estimaciones por nuestras desigualdades precisas, es:

$$O(\dim(\Lambda)n^2 S^2((\ell + 1) \dim(\Lambda) + \log(\deg_{\text{lci}}(\Lambda)))).$$

De acuerdo al Corolario 3.2.7, bajo nuestras hipótesis, la probabilidad de que una lista escogida aleatoriamente $\mathbf{Q} := (x_1, \dots, x_L)$ sea un conjunto cuestor para $\mathcal{W}(\Gamma, \Lambda)$ con respecto a $\mathcal{W}(\Gamma, \Lambda) \setminus U$ es mayor que:

$$1 - \frac{1}{\deg_{\text{lci}}(\Lambda) e^{\dim(\Lambda)(\log(D)+m)}}$$

Usando los mismos argumentos que en el Teorema 3.4.2, obtenemos fácilmente las cotas de las Propiedades *i*) y *ii*). Si K es un cuerpo algebraicamente cerrado adecuado para muestrear conjuntos cuestores, nuestras hipótesis verifican las condiciones de la Observación 3.4.4 y, por lo tanto, tenemos que $\mu[\mathcal{W}(\Gamma, \Lambda) \setminus U] = 0$. □

Erzeugungsgrad y dimensión de Vapnik-Chervonenkis

Índice

4.1. Erzeugungsgrad: revisando la Sección 3 de [He, 83] combinada con conjuntos constructibles	76
4.1.1. Definiciones y propiedades básicas	76
4.1.2. Cotas combinatorias	84
4.2. Teoría de Vapnik-Chervonenkis sobre cuerpos algebraicamente cerrados	91
4.2.1. Dimensión VC y clasificadores constructibles	91
4.2.2. Dimensión VC de abiertos distinguidos definidos por una clase de polinomios	93
4.2.3. Variedades evasivas	94
4.2.3.1. Mejora del Teorema 2.2 de [DKL, 14]	94
4.2.3.2. Variedades evasivas para un conjunto constructible de listas de polinomios	95
4.2.3.3. Variedades evasivas para la dimensión	97
4.2.4. Los conjuntos cuestosores están densamente distribuidos en variedades evasivas de dimensión positiva	98

El objetivo de este capítulo es responder afirmativamente a la Conjetura 1 (y al desafío de J. Heintz). Esta conjetura, propuesta por L. M. Pardo y el autor de esta memoria durante una conversación con J. Heintz en su última visita a Santander, sostiene que debe existir alguna conexión entre la dimensión de Krull del espacio de parámetros y alguna de las nociones de dimensión de la Teoría de Aprendizaje Computacional en el contexto de familias constructibles de clasificadores binarios (sobre cuerpos algebraicamente cerrados). En el caso de la Geometría Algebraica Real y la Teoría de Vapnik-Chervonenkis existen algunas conexiones, sin embargo, dichas conexiones no son válidas para característica positiva. Esto nos ha llevado a buscar otras técnicas. Curiosamente, la noción que nos va a permitir establecer las conexiones buscadas es el Erzeugungsgrad (o grado de generación), una noción introducida por J. Heintz en [He, 83].

El Erzeugungsgrad fue fundamental en [He, 83] para acotar el número de celdas no vacías que aparecen después de un proceso de eliminación de cuantificadores. Además, como veremos en este capítulo, el Erzeugungsgrad es también relevante porque, independientemente, proporciona información sobre la “función de crecimiento” de la Teoría del Aprendizaje Computacional, especialmente en el contexto de clasificadores binarios, de la dimensión de Vapnik-Chervonenkis y del Lema de Sauer-Shelah-Perles. En otras palabras, el Erzeugungsgrad está estrechamente relacionado con la noción de “desmenuzar” conjuntos finitos. El objetivo de este capítulo es explorar esta relación en detalle.

Desafortunadamente, antes de explorar dicha relación, debemos corregir un inconveniente que aparece en la formulación original del Erzeugungsgrad en [He, 83], el cual afecta a sus consecuencias. J. Heintz utilizó la noción de Z -grado para conjuntos constructibles como punto de partida para definir este concepto. Como ya hemos visto en el Ejemplo 2.1.1, el Z -grado no satisface la desigualdad de Bézout y, en consecuencia, su uso repercute tanto a los argumentos como a las cotas que involucran al Erzeugungsgrad. Por esta razón, dedicamos la Sección 4.1 a definir el Erzeugungsgrad utilizando nociones adecuadas para el grado de conjuntos constructibles (en particular, emplearemos la noción $\deg_{\text{Ici}}(C)$ que hemos introducido en la Definición 13) así como a corregir y generalizar las cotas combinatorias del Teorema 2 de [He, 83] basadas en el Erzeugungsgrad, proporcionando de esta forma un estudio más completo de esta noción.

Una vez restaurada la definición de Erzeugungsgrad y sus resultados principales, en la Sección 4.2 exploraremos algunas de las conexiones existentes entre el Erzeugungsgrad (una noción de la Teoría de la Intersección) y la función de crecimiento en la Teoría del Aprendizaje Computacional sobre cuerpos algebraicamente cerrados. En particular, mostraremos que, en el caso de familias de clasificadores binarios de clases parametrizadas de conjuntos constructibles, la dimensión de Vapnik-Chervonenkis y la dimensión de Krull están relacionadas linealmente, salvo por unos factores logarítmicos basados en la Teoría de la Intersección. Este resultado permite responder afirmativamente a la Conjetura 1 (y al desafío de J. Heintz).

4.1. Erzeugungsgrad: revisando la Sección 3 de [He, 83] combinada con conjuntos constructibles

4.1.1. Definiciones y propiedades básicas. Comenzamos esta subsección generalizando la noción de grado a familias finitas de conjuntos constructibles:

DEFINICIÓN 21 (Grado de una familia finita de conjuntos constructibles). Sea \mathcal{F} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Sea \mathcal{C} una aplicación que asocia a cada subconjunto constructible $X \in \mathcal{F}$ una descomposición en subconjuntos localmente cerrados irreducibles de grado LCI mínimo de acuerdo a la Definición 13. Entonces, definimos la clase finita de variedades irreducibles asociadas a \mathcal{C} mediante la siguiente igualdad:

$$\mathcal{C}(\mathcal{F}, \mathcal{C}) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{F}, \exists W \in \mathcal{C}(X), V = \overline{W^z}\} = \bigcup_{X \in \mathcal{F}} \{\overline{W^z} : W \in \mathcal{C}(X)\}.$$

Definimos el grado de la familia \mathcal{F} con respecto a la aplicación \mathcal{C} como la suma de los grados de todas las variedades irreducibles $V \in \mathcal{C}(\mathcal{F}, \mathcal{C})$:

$$\deg(\mathcal{F}, \mathcal{C}) := \sum_{V \in \mathcal{C}(\mathcal{F}, \mathcal{C})} \deg(V).$$

Finalmente, definimos el grado de la familia \mathcal{F} como el mínimo de estos grados:

$$\deg(\mathcal{F}) := \min\{\deg(\mathcal{F}, \mathcal{C}) : \mathcal{C} \text{ asocia a cada } X \in \mathcal{F} \text{ una descomposición de grado LCI mínimo}\}.$$

Definimos también $\text{Max-deg}(\mathcal{F})$ como el máximo de los grados LCI de cualquier $X \in \mathcal{F}$:

$$\text{Max-deg}(\mathcal{F}) := \max\{\deg_{\text{LCI}}(X) : X \in \mathcal{F}\}.$$

Además, sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Denotamos por $\mathcal{F}|_C$ a la restricción de los elementos de \mathcal{F} a C . Es decir,

$$\mathcal{F}|_C := \{C \cap X : X \in \mathcal{F}\}.$$

Usaremos la notación $\deg(\mathcal{F}, C)$ para referirnos al grado de $\mathcal{F}|_C$.

OBSERVACIÓN 4.1.1. Nótese que la aplicación $\mathcal{C}(X)$ se basa en una descomposición de grado LCI mínimo y no en las componentes irreducibles de la clausura Zariski $\overline{X^z}$ de X . La razón es que las componentes irreducibles de $\overline{X^z}$ están entre las componentes irreducibles de una descomposición de grado LCI mínimo. Sin embargo, una descomposición de grado LCI mínimo puede contener conjuntos localmente cerrados irreducibles relevantes (componentes inmersas) que desaparecen al tomar la clausura Zariski (ver Proposición 1.1.5).

EJEMPLO 4.1.2 (Dependencia de la aplicación \mathcal{C}). La cantidad $\deg(\mathcal{F}, \mathcal{C})$ depende de la aplicación \mathcal{C} , tal y como se muestra a continuación. Consideremos nuevamente el Ejemplo 2.1.3. Teníamos la siguiente hipersuperficie cuadrática:

$$(4.1.1) \quad W' := \{(x, y, z) \in \mathbb{A}^3(\mathbb{C}) : xz + (y^2 - 1) = 0\},$$

y el subconjunto constructible $C := \pi(W) \subseteq \mathbb{A}^2(\mathbb{C})$, donde $\pi : \mathbb{A}^3(\mathbb{C}) \rightarrow \mathbb{A}^2(\mathbb{C})$ es la proyección canónica que "olvida" la última coordenada. Sea D el abierto Zariski definido por:

$$D := \mathbb{A}^2(\mathbb{C}) \setminus V_{\mathbb{A}}(X) = \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x \neq 0\}.$$

Tenemos las siguientes dos descomposiciones distintas de C que minimizan su grado LCI:

$$C = D \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : x^2 + y^2 - 1 = 0\},$$

$$C = D \cup \{(x, y) \in \mathbb{A}^2(\mathbb{C}) : (x + 1) - y^2 = 0\}.$$

A continuación, definimos $\mathcal{F} := \{C, V\}$, donde $V := V_{\mathbb{A}}((X + 1) - Y^2)$ es la variedad algebraica irreducible dada como el conjunto de ceros en $\mathbb{A}^2(\mathbb{C})$ del polinomio $h(X, Y) = (X + 1) - Y^2 \in \mathbb{C}[X, Y]$. Consideramos ahora dos aplicaciones distintas, \mathcal{C}_1 y \mathcal{C}_2 , sobre \mathcal{F} , definidas de la siguiente forma:

$$\mathcal{C}_1(C) := \{D, V_{\mathbb{A}}(X^2 + Y^2 - 1)\}, \quad \mathcal{C}_1(V) = V, \quad \mathcal{C}_2(C) := \{D, V\}, \quad \mathcal{C}_2(V) = V.$$

Estas dos aplicaciones generan dos clases diferentes de componentes irreducibles:

$$\mathcal{C}(\mathcal{F}, \mathcal{C}_1) = \{\mathbb{A}^2(\mathbb{C}), V_{\mathbb{A}}(X^2 + Y^2 - 1), V\}, \quad \mathcal{C}(\mathcal{F}, \mathcal{C}_2) = \{\mathbb{A}^2(\mathbb{C}), V\}.$$

Estas clases dan resultado a dos grados distintos:

$$\deg(\mathcal{F}, \mathcal{C}_1) = \sum_{W \in \mathcal{C}(\mathcal{F}, \mathcal{C}_1)} \deg(W) = \deg(\mathbb{A}^2(\mathbb{C})) + \deg(V_{\mathbb{A}}(X^2 + Y^2 - 1)) + \deg(V) = 1 + 2 + 2 = 5,$$

$$\deg(\mathcal{F}, \mathcal{C}_2) = \sum_{W \in \mathcal{C}(\mathcal{F}, \mathcal{C}_2)} \deg(W) = \deg(\mathbb{A}^2(\mathbb{C})) + \deg(V) = 1 + 2 = 3.$$

OBSERVACIÓN 4.1.3. Nótese que también se cumplen las siguientes desigualdades y que la igualdad entre ellas no es siempre cierta:

$$\text{Max-deg}(\mathcal{F}) \leq \text{deg}(\mathcal{F}) \leq \sum_{D \in \mathcal{F}} \text{deg}_{\text{LCI}}(D) \leq \#(\mathcal{F}) \cdot \text{Max-deg}(\mathcal{F}).$$

DEFINICIÓN 22 (**Grado de las clausuras Zariski de una familia finita de conjuntos constructibles**).

Sea \mathcal{F} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Definimos:

$$\mathcal{D}(\mathcal{F}) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{F}, V \text{ es una componente irreducible de } \overline{X^z}\}.$$

El grado de las clausuras Zariski de una familia \mathcal{F} se define como la siguiente cantidad:

$$\text{deg}_z(\mathcal{F}) := \sum_{V \in \mathcal{D}(\mathcal{F})} \text{deg}(V).$$

Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible. Definimos:

$$\text{deg}_z(\mathcal{F}, C) := \text{deg}_z(\mathcal{F}|_C).$$

Nótese que si denotamos por $\overline{\mathcal{F}^z} := \{\overline{X^z} : X \in \mathcal{F}\}$, entonces:

$$\text{deg}_z(\mathcal{F}) = \text{deg}(\overline{\mathcal{F}^z}),$$

que se corresponde con la noción original introducida en [He, 83].

A continuación, demostramos que las dos nociones de grado que acabamos de definir para familias de conjuntos constructibles son funciones sub-aditivas:

LEMA 4.1.4. Sean \mathcal{F} y \mathcal{G} dos familias finitas de subconjuntos constructibles de $\mathbb{A}^n(K)$ tales que $\mathcal{F} \subseteq \mathcal{G}$. Entonces, tenemos que:

$$\text{deg}(\mathcal{F}) \leq \text{deg}(\mathcal{G}), \quad \text{deg}_z(\mathcal{F}) \leq \text{deg}_z(\mathcal{G}).$$

Esto es, ambos grados son funciones sub-aditivas.

DEMOSTRACIÓN. Sea \mathcal{C}_1 una aplicación que asocia a cada elemento de \mathcal{G} una descomposición en componentes localmente cerradas irreducibles que minimizan su grado LCI y que verifica la siguiente igualdad:

$$D = \text{deg}(\mathcal{G}) = \text{deg}(\mathcal{G}, \mathcal{C}_1).$$

A continuación, consideramos $\mathcal{C}_2 := \mathcal{C}_1|_{\mathcal{F}}$, la restricción de \mathcal{C}_1 a \mathcal{F} . Obviamente, si \mathcal{C}_1 asocia a cada elemento de \mathcal{G} una descomposición de grado LCI mínimo, \mathcal{C}_2 hace lo mismo para todos los elementos de \mathcal{F} . Tenemos las dos siguientes clases de variedades irreducibles:

$$\mathcal{C}(\mathcal{G}, \mathcal{C}_1) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{G}, \exists W \in \mathcal{C}_1(X), V = \overline{W^z}\},$$

$$\mathcal{C}(\mathcal{F}, \mathcal{C}_2) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{F}, \exists W \in \mathcal{C}_2(X), V = \overline{W^z}\}.$$

Dado que $\mathcal{C}_2(Y) = \mathcal{C}_1(Y)$ para cada $Y \in \mathcal{F}$, concluimos que $\mathcal{C}(\mathcal{F}, \mathcal{C}_2) \subseteq \mathcal{C}(\mathcal{G}, \mathcal{C}_1)$ y, por lo tanto:

$$\text{deg}(\mathcal{F}, \mathcal{C}_2) = \sum_{V \in \mathcal{C}(\mathcal{F}, \mathcal{C}_2)} \text{deg}(V) \leq \sum_{V \in \mathcal{C}(\mathcal{G}, \mathcal{C}_1)} \text{deg}(V) = \text{deg}(\mathcal{G}, \mathcal{C}_1) = \text{deg}(\mathcal{G}) = D.$$

Como $\text{deg}(\mathcal{F})$ es el mínimo de $\text{deg}(\mathcal{F}, \mathcal{C})$ para cualquier aplicación \mathcal{C} , concluimos que:

$$\text{deg}(\mathcal{F}) \leq \text{deg}(\mathcal{G}) = D.$$

La segunda desigualdad es una consecuencia inmediata de la primera ya que:

$$\text{deg}_z(\mathcal{F}) = \text{deg}(\overline{\mathcal{F}^z}), \quad \text{deg}_z(\mathcal{G}) = \text{deg}(\overline{\mathcal{G}^z})$$

y $\overline{\mathcal{F}^z} \subseteq \overline{\mathcal{G}^z}$. □

El siguiente lema demuestra que si \mathcal{W} es una familia de subconjuntos localmente cerrados de $\mathbb{A}^n(K)$, entonces $\text{deg}(\mathcal{W}) = \text{deg}_z(\mathcal{W})$.

LEMA 4.1.5. Sea \mathcal{W} una familia finita de subconjuntos localmente cerrados de $\mathbb{A}^n(K)$. Sea $\mathcal{D}(\mathcal{W})$ la clase finita de variedades algebraicas irreducibles introducida en la Definición 22. Es decir:

$$\mathcal{D}(\mathcal{W}) := \left\{ V \subseteq \mathbb{A}^n(K) : \exists W \in \mathcal{W}, V \text{ es una componente irreducible de } \overline{W^z} \right\}.$$

Entonces, tenemos que:

$$\text{deg}(\mathcal{W}) = \sum_{V \in \mathcal{D}(\mathcal{W})} \text{deg}(V) = \text{deg}_z(\mathcal{W}).$$

DEMOSTRACIÓN. Sea \mathcal{C} cualquier aplicación que asocia a cada $W \in \mathcal{W}$ una descomposición de grado LCI mínimo. Para cada $W \in \mathcal{W}$, usaremos la notación $\mathcal{C}(W, \mathcal{C}) := \mathcal{C}(\{W\}, \mathcal{C})$, de acuerdo con la Definición 21. Además, sea $\mathcal{D}(W) := \mathcal{D}(\{W\})$, de acuerdo con la Definición 22. Usando las notaciones precedentes, tenemos que:

$$\begin{aligned}\mathcal{C}(\mathcal{W}, \mathcal{C}) &= \bigcup_{W \in \mathcal{W}} \mathcal{C}(W, \mathcal{C}), \text{ y} \\ \mathcal{D}(\mathcal{W}) &= \bigcup_{W \in \mathcal{W}} \mathcal{D}(W).\end{aligned}$$

A continuación, vamos a demostrar que para cada $W \in \mathcal{W}$,

$$\mathcal{D}(W) \subseteq \mathcal{C}(W, \mathcal{C}).$$

Para probar lo anterior, observemos que W admite una descomposición de la forma $W = V \cap U$, donde $V \subseteq \mathbb{A}^n(K)$ es un cerrado Zariski y $U \subseteq \mathbb{A}^n(K)$ es un abierto Zariski. Nótese también que es posible elegir V de manera que admita una descomposición de la forma:

$$(4.1.2) \quad V = V_1 \cup \dots \cup V_m,$$

tal que $V_i \cap U \neq \emptyset$ para todo i , $1 \leq i \leq m$. Entonces, concluimos que la siguiente es una descomposición de \overline{W}^z en componentes irreducibles:

$$(4.1.3) \quad \overline{W}^z = V_1 \cup \dots \cup V_m.$$

A continuación, supongamos que $\mathcal{C}(W) = \{A_1, \dots, A_s\}$, donde A_i es un subconjunto localmente cerrado irreducible de $\mathbb{A}^n(K)$. Entonces, también tenemos que:

$$(4.1.4) \quad \overline{W}^z = \overline{A_1}^z \cup \dots \cup \overline{A_s}^z,$$

y $\mathcal{C}(W, \mathcal{C}) = \{\overline{A_1}^z, \dots, \overline{A_s}^z\}$. Por lo tanto, a partir de las Igualdades (4.1.3) y (4.1.4), y dado que la Igualdad (4.1.3) es una descomposición de \overline{W}^z es componentes irreducibles, concluimos:

$$\{V_1, \dots, V_m\} \subseteq \{\overline{A_1}^z, \dots, \overline{A_s}^z\}.$$

Como $\mathcal{D}(W) = \{V_1, \dots, V_m\}$, hemos demostrado que para cada $W \in \mathcal{W}$,

$$\mathcal{D}(W) \subseteq \mathcal{C}(W, \mathcal{C}).$$

Entonces, $\mathcal{D}(\mathcal{W}) \subseteq \mathcal{C}(\mathcal{W}, \mathcal{C})$ y, por lo tanto:

$$\sum_{V \in \mathcal{D}(\mathcal{W})} \deg(V) \leq \sum_{V \in \mathcal{C}(\mathcal{W}, \mathcal{C})} \deg(V) = \deg(\mathcal{W}, \mathcal{C}).$$

Por otro lado, dado que $W = V \cap U$ y teniendo en cuenta la descomposición de V descrita en la Igualdad (4.1.2), tenemos una aplicación \mathcal{C}_1 que asociada a $W \in \mathcal{W}$ la siguiente descomposición de grado LCI mínimo:

$$\mathcal{C}_1(W) = \{V_1 \cap U, \dots, V_m \cap U\}.$$

Entonces, $\mathcal{C}(W, \mathcal{C}_1) \subseteq \mathcal{D}(W)$ para cada $W \in \mathcal{W}$ y, por lo tanto,

$$\deg(\mathcal{W}, \mathcal{C}_1) = \sum_{V \in \mathcal{C}(W, \mathcal{C}_1)} \deg(V) \leq \sum_{V \in \mathcal{D}(W)} \deg(V).$$

Esto prueba que:

$$\deg(\mathcal{W}) \leq \sum_{V \in \mathcal{D}(\mathcal{W})} \deg(V) = \deg_z(\mathcal{W}),$$

lo que completa la demostración del lema. □

DEFINICIÓN 23 (\mathcal{H} -definible, \mathcal{H} -celda). Sea \mathcal{H} una clase finita de subconjuntos constructibles de $\mathbb{A}^n(K)$ y sea $C \subseteq \mathbb{A}^n(K)$ otro conjunto constructible.

- Sea $\mathcal{B}(\mathcal{H}, C)$ el álgebra booleana de los subconjuntos de C generados por \mathcal{H} . Un subconjunto X de C recibe el nombre de \mathcal{H} -definible en C si pertenece a esta álgebra booleana, i.e. $X \in \mathcal{B}(\mathcal{H}, C)$. En el caso $C = \mathbb{A}^n(K)$, escribiremos simplemente $\mathcal{B}(\mathcal{H}) := \mathcal{B}(\mathcal{H}, \mathbb{A}^n(K))$.
- Llamamos \mathcal{H} -celda en C a cualquier subconjunto $X \subseteq C$ de modo que existe $M \subseteq \mathcal{H}$ tal que:

$$X = C \cap \left(\bigcap_{Y \in M} Y \right) \cap \left(\bigcap_{Y \notin M} (C \setminus Y) \right).$$

Denotamos por $\mathcal{Z}(\mathcal{H}, C)$ a la clase de todas las \mathcal{H} -celdas no vacías en C . En el caso $C = \mathbb{A}^n(K)$, escribiremos simplemente $\mathcal{Z}(\mathcal{H}) := \mathcal{Z}(\mathcal{H}, \mathbb{A}^n(K))$.

- Finalmente, definimos la clase $\text{Gen}(\mathcal{B}(\mathcal{H}, C))$ como la clase de todas las familias finitas de variedades algebraicas $\mathcal{W} := \{W_1, \dots, W_t\}$ tales que $\mathcal{B}(\mathcal{H}, C) \subseteq \mathcal{B}(\mathcal{W}, C)$. En el caso $C = \mathbb{A}^n(K)$, escribiremos simplemente $\text{Gen}(\mathcal{B}(\mathcal{H})) := \text{Gen}(\mathcal{B}(\mathcal{H}, \mathbb{A}^n(K)))$.

Nótese que $\mathcal{L}(\mathcal{H}, C)$ define una partición finita de C y que cada elemento en $\mathcal{B}(\mathcal{H}, C)$ es la unión finita de \mathcal{H} -celdas en C .

DEFINICIÓN 24 (Erzeugungsgrad). Sea \mathcal{H} una familia finita de conjuntos constructibles. Definimos el Erzeugungsgrad (o grado de generación) de \mathcal{H} como:

$$\text{grad}(\mathcal{H}) := \min\{\text{deg}(\mathcal{W}) : \mathcal{W} \in \text{Gen}(\mathcal{B}(\mathcal{H})), \mathcal{W} \text{ es una familia finita de conjuntos cerrados}\}.$$

OBSERVACIÓN 4.1.6 (Grado y Erzeugungsgrad: variedades algebraicas). En el caso de variedades algebraicas, el grado (en el sentido de [He, 83]) y el Erzeugungsgrad (o grado de generación), que acabamos de definir, son cantidades muy diferentes. El Erzeugungsgrad siempre es menor o igual que el grado, i.e. $\text{grad}(\mathcal{H}) \leq \text{deg}_z(\mathcal{H})$, donde \mathcal{H} es una clase formada por una sola variedad algebraica.

En algunos casos, el Erzeugungsgrad puede ser extremadamente menor que el grado. A modo de ejemplo, consideremos el caso de una clase de conjuntos constructibles $\mathcal{H} := \{V\}$ formada por una sola variedad algebraica V de dimensión $n - r$. Supongamos, además, que esta variedad es de intersección completa, es decir, existen polinomios $f_1, \dots, f_r \in K[X_1, \dots, X_n]$ de grados respectivos d_1, \dots, d_r , tales que $V := V(f_1, \dots, f_r)$. Genéricamente, el grado de V (y, por tanto, el grado de la clase \mathcal{H}) es el número de Bézout:

$$\mathcal{D} := \prod_{i=1}^r d_i,$$

que es una cantidad exponencial en la co-dimensión. Sin embargo, $\text{grad}(\mathcal{H})$ está acotado por la siguiente suma:

$$\sum_{i=1}^r d_i,$$

que es lineal en la co-dimensión. Esta diferencia hace que el uso del Erzeugungsgrad para cotas combinatorias (como se hace en el Teorema 4.1.11) permita obtener cotas mucho más finas que las que se obtendrían usando únicamente el grado como invariante.

OBSERVACIÓN 4.1.7. Dada una clase finita de subconjuntos constructibles \mathcal{H} , existe una familia finita \mathcal{W} de variedades algebraicas irreducibles tales que $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W})$ y:

$$\text{grad}(\mathcal{H}) = \text{deg}(\mathcal{W}) = \sum_{W \in \mathcal{W}} \text{deg}(W).$$

Para ver esto, sea $\widetilde{\mathcal{W}}$ una familia finita de variedades algebraicas tales que $\mathcal{H} \subseteq \mathcal{B}(\widetilde{\mathcal{W}})$ y $\text{grad}(\mathcal{H}) = \text{deg}(\widetilde{\mathcal{W}})$. Al igual que en la Definición 22, sea $\mathcal{D}(\widetilde{\mathcal{W}})$ el conjunto de componentes irreducibles de $\widetilde{\mathcal{W}}$. Obviamente, tenemos que $\mathcal{H} \subseteq \mathcal{B}(\mathcal{D}(\widetilde{\mathcal{W}}))$, ya que $\widetilde{\mathcal{W}} \subseteq \mathcal{B}(\mathcal{D}(\widetilde{\mathcal{W}}))$ y $\mathcal{H} \subseteq \mathcal{B}(\widetilde{\mathcal{W}})$, y:

$$\text{deg}(\widetilde{\mathcal{W}}) = \sum_{W \in \mathcal{D}(\widetilde{\mathcal{W}})} \text{deg}(W)$$

Por lo tanto, $\mathcal{W} = \mathcal{D}(\widetilde{\mathcal{W}})$ es una familia finita de variedades algebraicas irreducibles tales que $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W})$ y:

$$\text{grad}(\mathcal{H}) = \text{deg}(\mathcal{W}) = \sum_{W \in \mathcal{W}} \text{deg}(W).$$

El siguiente lema es la herramienta técnica que nos permitirá demostrar los resultados principales de esta sección:

LEMA 4.1.8. Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y $\mathcal{W} = \{W_1, \dots, W_s\}$ una familia finita de subconjuntos localmente cerrados de $\mathbb{A}^n(K)$. Definimos el siguiente conjunto:

$$\mathcal{T}(\mathcal{W}, C) := \left\{ X \subseteq \mathbb{A}^n(K) : \exists S \subseteq \{1, \dots, s\}, X = C \cap \left(\bigcap_{i \in S} W_i \right) \right\}.$$

Sea $\mathcal{D} := \mathcal{D}(\mathcal{T}(\mathcal{W}, C))$ la clase de las variedades algebraicas irreducibles asociadas a $\mathcal{T}(\mathcal{W}, C)$ como en la Definición 22. Entonces, tenemos que:

$$\sum_{A \in \mathcal{D}} \text{deg}(A) \leq \text{deg}_{\text{Ici}}(C)(1 + \text{deg}(\mathcal{W}))^{\dim(C)}.$$

DEMOSTRACIÓN. En primer lugar, observamos que es suficiente probar este lema en el caso en que C es localmente cerrado. Para probar la afirmación anterior, consideramos una descomposición del conjunto constructible C en subconjuntos localmente cerrados irreducibles que minimice el grado LCI como en la Definición 13, i.e.

$$C := C_1 \cup \dots \cup C_m,$$

tal que:

$$\deg_{\text{LCI}}(C) = \sum_{i=1}^m \deg(C_i).$$

Con estas notaciones, se verifica lo siguiente:

AFIRMACIÓN 4.1.8.1.

$$\mathcal{D} \subseteq \bigcup_{j=1}^m \mathcal{D}(\mathcal{T}(\mathcal{W}, C_j)).$$

Demostración de la Afirmación. Sea $V \in \mathcal{D}$ una variedad algebraica irreducible tal que existe $X \in \mathcal{T}(\mathcal{W}, C)$ verificando que V es una componente irreducible de \overline{X}^z . Además, dado que $X \in \mathcal{T}(\mathcal{W}, C)$, existe $S \subseteq \{1, \dots, s\}$ tal que:

$$X = C \cap \left(\bigcap_{i \in S} W_i \right).$$

Por lo tanto,

$$X = \bigcup_{j=1}^m \left(C_j \cap \left(\bigcap_{i \in S} W_i \right) \right).$$

Entonces, tomando las clausuras Zariski, tenemos que:

$$\overline{X}^z = \bigcup_{j=1}^m \overline{C_j \cap \left(\bigcap_{i \in S} W_i \right)}^z.$$

A continuación, recordemos que cualquier descomposición de una variedad algebraica A como unión finita de variedades irreducibles puede refinarse para obtener una descomposición de A en sus componentes irreducibles. Por lo tanto, las componentes irreducibles de \overline{X}^z son elementos del conjunto:

$$\bigcup_{j=1}^m \overline{C_j \cap \left(\bigcap_{i \in S} W_i \right)}^z.$$

Como V es una componente irreducible de \overline{X}^z , tenemos que V debe ser una componente irreducible de $\overline{C_j \cap \left(\bigcap_{i \in S} W_i \right)}^z$ para algún $j \in \{1, \dots, m\}$. Por lo tanto, dado que:

$$C_j \cap \left(\bigcap_{i \in S} W_i \right) \in \mathcal{T}(\mathcal{W}, C_j),$$

concluimos que $V \in \mathcal{D}(\mathcal{T}(\mathcal{W}, C_j))$ y la afirmación queda demostrada. ■

Si el lema se cumple en el caso localmente cerrado, tendríamos la siguiente desigualdad para cada $j \in \{1, \dots, m\}$:

$$\sum_{A \in \mathcal{D}(\mathcal{T}(\mathcal{W}, C_j))} \deg(A) \leq \deg(C_j) (1 + \deg(\mathcal{W}))^{\dim(C_j)}.$$

Dado que $\dim(C) = \max\{\dim(C_j) : 1 \leq j \leq m\}$ y aplicando la Afirmación 4.1.8.1, concluimos:

$$\begin{aligned} \sum_{A \in \mathcal{D}} \deg(A) &\leq \sum_{j=1}^m \left(\sum_{A \in \mathcal{D}(\mathcal{T}(W, C_j))} \deg(A) \right) \leq \\ &\leq \sum_{j=1}^m \deg(C_j) (1 + \deg(W))^{\dim(C_j)} \leq \\ &\leq \sum_{j=1}^m \deg(C_j) (1 + \deg(W))^{\dim(C)} = \\ &= \left(\sum_{j=1}^m \deg(C_j) \right) (1 + \deg(W))^{\dim(C)} = \\ &= \deg_{\text{lci}}(C) (1 + \deg(W))^{\dim(C)}, \end{aligned}$$

que prueba el lema para cualquier conjunto constructible C .

A partir de ahora, supondremos, como se ha afirmado previamente, que C es un subconjunto localmente cerrado de $\mathbb{A}^n(K)$. Demostraremos el lema para el caso en que C es localmente cerrado (en cuyo caso $\deg(C) = \deg_z(C) = \deg_\pi(C) = \deg_{\text{lci}}(C)$).

Sea $d := \dim(C)$. Definimos las siguientes clases de componentes irreducibles:

- $\mathcal{D}(d) := \{V : V \text{ es una componente irreducible de } \overline{C}^z\}$.
- Para $0 \leq k \leq d-1$, definimos:

$$(4.1.5) \quad \mathcal{D}(k) := \{V \in \mathcal{D} : \dim(V) = k, V \notin \mathcal{D}(d)\}.$$

Por lo tanto, tenemos una partición de \mathcal{D} dada por la siguiente unión disjunta:

$$\mathcal{D} = \bigcup_{k=0}^d \mathcal{D}(k).$$

A continuación, probamos la siguiente afirmación:

AFIRMACIÓN 4.1.8.2. *Sea $k \in \mathbb{N}$ un entero no negativo tal que $0 \leq k \leq d-1$. Sea $A \in \mathcal{D}(k)$ un elemento de $\mathcal{D}(k)$ de acuerdo a la Identidad (4.1.5). Entonces, existe $A^* \in \mathcal{D}$ y $V \in \mathcal{D}(W)$ tal que:*

- i) $\dim(A^*) \geq \dim(A) + 1 = k + 1$.
- ii) A es una componente irreducible de $A^* \cap V$.

Demostración de la Afirmación. Como $A \in \mathcal{D}(k) \subseteq \mathcal{D}$, existe un subconjunto $S \subseteq \{1, \dots, s\}$ de cardinal mínimo que verifica las siguientes dos propiedades:

- El conjunto localmente cerrado B dado por:

$$B := C \cap \left(\bigcap_{i \in S} W_i \right)$$

es un elemento de $\mathcal{T}(W, C)$.

- A es una componente irreducible de \overline{B}^z .

Dado que $A \in \mathcal{D}(k)$ y $k < d$, sabemos que A no es una componente irreducible de \overline{C}^z . Por lo tanto, $S \neq \emptyset$ (en caso contrario $B = C$) y podemos elegir $i \in S$. Definimos $S' := S \setminus \{i\}$ y:

$$B' := C \cap \left(\bigcap_{j \in S'} W_j \right) \in \mathcal{T}(W, C).$$

Sea $W := W_i \in \mathcal{W}$ y observemos que $B = B' \cap W$. Dado que A es una componente irreducible de $\overline{B}^z \subseteq \overline{B'}^z$, siempre existe una componente irreducible A^* de $\overline{B'}^z$ tal que $A \subseteq A^*$. Además, debido a la cardinalidad mínima de S , A no puede ser una componente irreducible de $\overline{B'}^z$. En particular, para cada componente irreducible A^* de $\overline{B'}^z$ tal que $A \subseteq A^*$, tenemos que $A \subsetneq A^*$. Como ambos son irreducibles, también concluimos que para cada componente irreducible A^* de $\overline{B'}^z$ tal que $A \subseteq A^*$, tenemos que $\dim(A^*) \geq \dim(A) + 1$.

Dado que estamos suponiendo que C es localmente cerrado y que los elementos de la familia \mathcal{W} son también localmente cerrados, podemos concluir que B' es también un subconjunto localmente cerrado de $\mathbb{A}^n(K)$. Por

lo tanto, existe una descomposición de B' como unión finita de conjuntos localmente cerrados irreducibles no vacíos:

$$B' = (A_1^* \cap B_1) \cup \cdots \cup (A_t^* \cap B_t),$$

donde:

- A_j^* es un subconjunto cerrado irreducible de $\mathbb{A}^n(K)$,
- B_j es un subconjunto abierto de $\mathbb{A}^n(K)$ tal que $A_j^* \cap B_j \neq \emptyset$, $\forall j \in \{1, \dots, t\}$,
- $\overline{B'^z} = A_1^* \cup \cdots \cup A_t^*$ es la descomposición de $\overline{B'^z}$ en componentes irreducibles.

Nótese que W es también localmente cerrado y, por lo tanto, admite una descomposición en conjuntos localmente cerrados irreducibles no vacíos:

$$W = (V_1 \cap U_1) \cup \cdots \cup (V_r \cap U_r),$$

tal que:

- V_ℓ es una variedad algebraica irreducible de $\mathbb{A}^n(K)$,
- U_ℓ es un abierto Zariski de $\mathbb{A}^n(K)$ que verifica $V_\ell \cap U_\ell \neq \emptyset$, $\forall \ell \in \{1, \dots, r\}$,
- $\overline{W^z} = V_1 \cup \cdots \cup V_r$ es la descomposición de $\overline{W^z}$ en sus componentes irreducibles.

Nótese que, en particular, se verifica:

$$\{V_1, \dots, V_r\} \subseteq \mathcal{D}(W).$$

En consecuencia, concluimos que:

$$B = B' \cap W = \bigcup_{j=1}^t \bigcup_{\ell=1}^r (A_j^* \cap V_\ell) \cap (B_j \cap U_\ell).$$

Por lo tanto:

$$\overline{B^z} = \overline{B' \cap W^z} = \bigcup_{j=1}^t \bigcup_{\ell=1}^r \overline{(A_j^* \cap V_\ell) \cap (B_j \cap U_\ell)^z}.$$

Dado que la descomposición de la variedad algebraica $\overline{B^z}$ como unión finita de conjuntos irreducibles puede refinarse para obtener una descomposición de $\overline{B^z}$ en sus componentes irreducibles, concluimos que existen $j \in \{1, \dots, t\}$ y $\ell \in \{1, \dots, r\}$ tales que A es una componente irreducible de:

$$\overline{(A_j^* \cap V_\ell) \cap (B_j \cap U_\ell)^z}.$$

A continuación, nótese que $A_j^* \cap V_\ell \subseteq \mathbb{A}^n(K)$ es una variedad algebraica y $B_j \cap U_\ell \subseteq \mathbb{A}^n(K)$ un subconjunto abierto Zariski. Supongamos que la siguiente es una descomposición de $A_j^* \cap V_\ell$ en sus componentes irreducibles:

$$A_j^* \cap V_\ell = Z_1 \cup \cdots \cup Z_p.$$

Entonces, para todo q , $1 \leq q \leq p$, tenemos que $Z_q \cap (B_j \cap U_\ell)$ es o bien vacío, o bien denso en Z_q . Salvo reordenación de los índices, podemos suponer que u , con $1 < u \leq p$, es tal que:

$$Z_q \cap (B_j \cap U_\ell) \neq \emptyset \text{ si y solo si } 1 \leq q \leq u.$$

Entonces,

$$\overline{(A_j^* \cap V_\ell) \cap (B_j \cap U_\ell)^z} = Z_1 \cup \cdots \cup Z_u,$$

y, por lo tanto, $A \in \{Z_1, \dots, Z_u\}$. Tomando $A^* = A_j^*$ y $V = V_\ell$ obtenemos el resultado. ■

Debido a la afirmación anterior, estamos en condiciones de demostrar el lema. En primer lugar, para cada $k < d$, definimos la siguiente aplicación:

$$\Phi_k : \begin{array}{ccc} \mathcal{D}(k) & \longrightarrow & \left(\bigcup_{r=k+1}^d \mathcal{D}(r) \right) \times \mathcal{D}(W), \\ A & \longmapsto & (A^*, V) \end{array},$$

dada por la siguiente regla: a cada $A \in \mathcal{D}(k)$, con $k < d$, le asociamos un par (A^*, V) de forma que A es una componente irreducible de $A^* \cap V$.

Definimos la siguiente cantidad:

$$\mathcal{D}(k) := \sum_{r=k}^d \left(\sum_{A \in \mathcal{D}(r)} \deg(A) \right).$$

Demostramos ahora por inducción sobre $m = d - k$ que la siguiente igualdad se cumple:

$$(4.1.6) \quad \mathcal{D}(k) = \mathcal{D}(d - m) \leq \deg(C)(1 + R)^m,$$

donde:

$$R := \deg(W).$$

Nótese que:

$$(4.1.7) \quad \mathcal{D}(k) = \mathcal{D}(k+1) + \sum_{A \in \mathcal{D}(k)} \deg(A).$$

El caso $m = 0$ es evidente ya que $\mathcal{D}(d)$ es la clase de todas las componentes irreducibles del conjunto localmente cerrado C . Por lo tanto:

$$\sum_{A \in \mathcal{D}(d)} \deg(A) = \deg_z(C) = \deg(C).$$

Supongamos ahora que la Desigualdad (4.1.6) se cumple para $m-1$ y vamos a demostrar que también se cumple para $m = d - k \geq 1$. En primer lugar, nótese que se verifica la siguiente desigualdad:

$$\sum_{A \in \mathcal{D}(k)} \deg(A) \leq \sum_{(A^*, V) \in (\bigcup_{r=k+1}^d \mathcal{D}(r) \times \mathcal{D}(\mathcal{W}))} \sum_{A \in \Phi_k^{-1}(A^*, V)} \deg(A).$$

Usando la afirmación previa, tenemos que A es una componente irreducible de $A^* \cap V$ para cada $A \in \Phi_k^{-1}(A^*, V)$. Así, aplicando la Desigualdad de Bézout de [He, 83] (ver Teorema 1.2.8) y utilizando la definición de Φ_k , obtenemos:

$$\sum_{A \in \Phi_k^{-1}(A^*, V)} \deg(A) \leq \deg(A^* \cap V) \leq \deg(A^*) \cdot \deg(V).$$

Por lo tanto, tenemos que:

$$\sum_{A \in \mathcal{D}(k)} \deg(A) \leq \sum_{(A^*, V) \in (\bigcup_{r=k+1}^d \mathcal{D}(r) \times \mathcal{D}(\mathcal{W}))} \deg(A^*) \cdot \deg(V).$$

Por consiguiente,

$$(4.1.8) \quad \sum_{A \in \mathcal{D}(k)} \deg(A) \leq \left(\sum_{A^* \in (\bigcup_{r=k+1}^d \mathcal{D}(r))} \deg(A^*) \right) \cdot \left(\sum_{V \in \mathcal{D}(\mathcal{W})} \deg(V) \right).$$

Aplicando ahora el Lema 4.1.5, obtenemos:

$$\sum_{V \in \mathcal{D}(\mathcal{W})} \deg(V) = \deg(\mathcal{W}) = R.$$

Nótese que:

$$\sum_{A^* \in (\bigcup_{r=k+1}^d \mathcal{D}(r))} \deg(A^*) = \sum_{r=k+1}^d \left(\sum_{A^* \in \mathcal{D}(r)} \deg(A^*) \right) = \mathcal{D}(k+1).$$

Por lo tanto, la Desigualdad (4.1.8) se convierte en:

$$\sum_{A \in \mathcal{D}(k)} \deg(A) \leq \mathcal{D}(k+1) \cdot R.$$

En conclusión, combinando la Identidad (4.1.7) con la desigualdad previa, obtenemos:

$$\mathcal{D}(k) = \mathcal{D}(k+1) + \sum_{A \in \mathcal{D}(k)} \deg(A) \leq \mathcal{D}(k+1) + \mathcal{D}(k+1) \cdot R = \mathcal{D}(k+1) \cdot (1 + R).$$

Aplicando la hipótesis de inducción para $m-1 = d - (k+1)$, concluimos:

$$\mathcal{D}(k) \leq \deg(C)(1 + R)^{m-1} \cdot (1 + R) = \deg(C)(1 + R)^m.$$

Por lo tanto, el lema es cierto cuando C es localmente cerrado. De acuerdo con lo expuesto al comienzo de esta demostración, el lema también se verifica para todo conjunto constructible $C \subseteq \mathbb{A}^n(K)$. \square

COROLARIO 4.1.9. *Con las mismas notaciones e hipótesis que en el lema previo, tenemos que:*

$$\sum_{A \in \mathcal{D}(\mathcal{W}, C)} \deg(A) \leq \deg_{\text{lci}}(C) \left(1 + \sum_{i=1}^s \deg(W_i) \right)^{\dim(C)}.$$

DEMOSTRACIÓN. Basta observar que:

$$\deg(\mathcal{W}) \leq \sum_{i=1}^s \deg(W_i).$$

\square

4.1.2. Cotas combinatorias.

DEFINICIÓN 25. Sea $\mathcal{W} = \{W_1, \dots, W_s\}$ una familia finita de subconjuntos localmente cerrados de $\mathbb{A}^n(K)$. Definimos la cantidad $\widetilde{\deg}(\mathcal{W})$ como el mínimo del siguiente conjunto:

$$\{\deg(\widetilde{\mathcal{W}}) : \widetilde{\mathcal{W}} = \{A_1, \dots, A_s, B_1, \dots, B_s\}, A_i, B_i \subseteq \mathbb{A}^n(K) \text{ cerrados Zariski, } 1 \leq i \leq s, \\ \text{y } W_i = A_i \cap (\mathbb{A}^n(K) \setminus B_i), 1 \leq i \leq s\}.$$

Nótese que si $\mathcal{W} = \{W_1, \dots, W_s\}$ está formado únicamente por variedades algebraicas, tomando $A_i = W_i$ y $B_i = \emptyset$ para cada $1 \leq i \leq s$, concluimos:

$$\widetilde{\deg}(\mathcal{W}) = \deg(\mathcal{W}).$$

PROPOSICIÓN 4.1.10. Sea $C \subseteq \mathbb{A}^n(K)$ un subconjunto constructible. Sea $\mathcal{W} = \{W_1, \dots, W_s\}$ una familia finita de subconjuntos localmente cerrados de $\mathbb{A}^n(K)$. Sea \mathcal{H} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Supongamos que:

$$\mathcal{H}|_C := \{H \cap C : H \in \mathcal{H}\} \subseteq \mathcal{B}(\mathcal{W}, C).$$

Entonces, tenemos que:

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \deg_{\text{lci}}(C)(1 + \widetilde{\deg}(\mathcal{W}))^{\dim(C)}.$$

Si \mathcal{W} está formado únicamente por cerrados Zariski, se verifica que:

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \deg_{\text{lci}}(C)(1 + \deg(\mathcal{W}))^{\dim(C)}.$$

DEMOSTRACIÓN. En primer lugar, nótese que $\mathcal{Z}(\mathcal{H}, C)$ depende únicamente de sus intersecciones con C . Por lo tanto,

$$\mathcal{Z}(\mathcal{H}|_C, C) = \mathcal{Z}(\mathcal{H}, C),$$

y podemos suponer que los subconjuntos constructibles de \mathcal{H} son también subconjuntos constructibles de C .

Comenzamos demostrando la siguiente afirmación, que relaciona subconjuntos de índices con celdas:

AFIRMACIÓN 4.1.10.1. Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y $\mathcal{H} = \{H_1, \dots, H_t\}$ una familia finita de conjuntos constructibles. Entonces, la siguiente aplicación es una biyección:

$$(4.1.9) \quad \left\{ \begin{array}{l} T \subseteq [t] : Y_T \neq \emptyset \\ T \end{array} \right\} \begin{array}{l} \longrightarrow \\ \longmapsto \end{array} \begin{array}{l} \mathcal{Z}(\mathcal{H}, C) \\ Y_T \end{array},$$

donde:

$$Y_T := C \cap \left(\bigcap_{k \in T} H_k \right) \cap \left(\bigcap_{\ell \in [t] \setminus T} (\mathbb{A}^n(K) \setminus H_\ell) \right).$$

Demostración de la Afirmación. Basta observar que si T y T' son dos subconjuntos distintos de $[t]$, con $Y_T \neq \emptyset$ y $Y_{T'} \neq \emptyset$, entonces $Y_T \cap Y_{T'} = \emptyset$ ya que ambos son \mathcal{H} -celdas en C . Por lo tanto, si $T \neq T'$, entonces $Y_T \neq Y_{T'}$, y la aplicación de la Identidad (4.1.9) es inyectiva. Dado que la aplicación es obviamente suprayectiva, se trata de una biyección. ■

A continuación, acotamos el número de \mathcal{H} -celdas en C por el número de $\widetilde{\mathcal{W}}$ -celdas en C para cualquier clase de conjuntos constructibles tal que $\mathcal{H} \subseteq \mathcal{B}(\widetilde{\mathcal{W}}, C)$:

AFIRMACIÓN 4.1.10.2. Con las mismas notaciones e hipótesis que en el enunciado de la proposición, sea $\widetilde{\mathcal{W}} = \{V_1, \dots, V_r\}$ cualquier clase de conjuntos constructibles tal que $\mathcal{H} \subseteq \mathcal{B}(\widetilde{\mathcal{W}}, C)$. Entonces, tenemos que:

$$(4.1.10) \quad \sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C)).$$

Demostración de la Afirmación. Supongamos que $\mathcal{H} = \{H_1, \dots, H_t\}$. Para cada $T \subseteq [t]$, consideramos la \mathcal{H} -celda en C asociada con T :

$$Y_T := C \cap \left(\bigcap_{k \in T} H_k \right) \cap \left(\bigcap_{\ell \in [t] \setminus T} (\mathbb{A}^n(K) \setminus H_\ell) \right).$$

Como $\mathcal{H} \subseteq \mathcal{B}(\widetilde{\mathcal{W}}, C)$, para cada $T \subseteq [t]$, con $Y_T \neq \emptyset$, existe $L(T) \subseteq 2^{[r]}$ tal que:

$$Y_T = \bigcup_{J \in L(T)} C \cap \left(\bigcap_{i \in J} V_i \right) \cap \left(\bigcap_{j \in [r] \setminus J} (\mathbb{A}^n(K) \setminus V_j) \right).$$

Para cada $J \subseteq [r]$, denotamos por $X_J \in \mathcal{Z}(\widetilde{W}, C)$ al conjunto constructible:

$$X_J := C \cap \left(\bigcap_{i \in J} V_i \right) \cap \left(\bigcap_{j \in [r] \setminus J} (\mathbb{A}^n(K) \setminus V_j) \right).$$

Para cada $T \subseteq [t]$, con $Y_T \neq \emptyset$, existe al menos un $J(T) \in L(T)$ tal que $X_{J(T)} \neq \emptyset$. Esto nos permite construir la siguiente aplicación:

$$(4.1.11) \quad \begin{array}{ccc} \{T \subseteq [t] : Y_T \neq \emptyset\} & \longrightarrow & \{J \subseteq [r] : X_J \neq \emptyset\} \\ T & \longmapsto & J(T) \end{array}.$$

A continuación, demostramos que se trata de una aplicación inyectiva. Si T y T' son dos subconjuntos distintos de $[t]$ tales que $Y_T \neq \emptyset$, $Y_{T'} \neq \emptyset$ y $J(T) = J(T')$, tendríamos:

$$X_{J(T)} = X_{J(T')} \subseteq Y_T \cap Y_{T'} = \emptyset,$$

lo que contradice que $X_{J(T)} = X_{J(T')} \neq \emptyset$. Por lo tanto, concluimos la Desigualdad (4.1.10) aplicando simplemente la Afirmación 4.1.10.1.

■

Demostramos ahora la proposición para el caso en que C es un conjunto localmente cerrado y \mathcal{W} es una familia finita de variedades algebraicas:

AFIRMACIÓN 4.1.10.3. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto localmente cerrado, $\mathcal{W} = \{W_1, \dots, W_s\}$ una familia finita de variedades algebraicas de $\mathbb{A}^n(K)$ y \mathcal{H} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Supongamos que $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W}, C)$. Entonces tenemos que:*

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \deg(C)(1 + \deg(\mathcal{W}))^{\dim(C)},$$

donde $\deg(C) = \deg_\pi(C) = \deg_{\text{gici}}(C)$ y $\deg(\mathcal{W}) = \widetilde{\deg}(\mathcal{W})$.

Demostración de la Afirmación. Consideramos las siguientes clases de conjuntos:

$$\mathcal{D}(\mathcal{Z}(\mathcal{W}, C)) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{Z}(\mathcal{W}, C), V \text{ es una componente irreducible de } \overline{X}^z\},$$

$$\mathcal{T}(\mathcal{W}, C) := \left\{ X \subseteq C : \exists S \subseteq \mathcal{W}, X = C \cap \left(\bigcap_{W \in S} W \right) \right\}$$

y

$$\mathcal{D}(\mathcal{T}(\mathcal{W}, C)) := \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{T}(\mathcal{W}, C), V \text{ es una componente irreducible de } \overline{X}^z\}.$$

La siguiente afirmación demuestra que $\mathcal{D}(\mathcal{Z}(\mathcal{W}, C))$ es un subconjunto de $\mathcal{D}(\mathcal{T}(\mathcal{W}, C))$:

AFIRMACIÓN 4.1.10.3.1. *Con las notaciones anteriores, tenemos que:*

$$(4.1.12) \quad \mathcal{D}(\mathcal{Z}(\mathcal{W}, C)) \subseteq \mathcal{D}(\mathcal{T}(\mathcal{W}, C)).$$

Demostración de la Afirmación. Sea $V \in \mathcal{D}(\mathcal{Z}(\mathcal{W}, C))$ una componente irreducible de \overline{X}^z , donde $X \in \mathcal{Z}(\mathcal{W}, C)$. Queremos probar que V es una componente irreducible de algún \overline{Y}^z , con $Y \in \mathcal{T}(\mathcal{W}, C)$. Como $X \in \mathcal{Z}(\mathcal{W}, C)$, existe un subconjunto $S \subseteq [s]$ tal que:

$$X := C \cap \left(\bigcap_{i \in S} W_i \right) \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right) \neq \emptyset.$$

A continuación, consideramos el siguiente conjunto localmente cerrado y su descomposición en componentes localmente cerradas irreducibles:

$$Y := C \cap \left(\bigcap_{i \in S} W_i \right) := A_1 \cup \dots \cup A_k,$$

donde $A_i = U_i \cap V_i \subseteq \mathbb{A}^n(K)$ es localmente cerrado irreducible, con $U_i \subseteq \mathbb{A}^n(K)$ abierto Zariski y $V_i \subseteq \mathbb{A}^n(K)$ cerrado irreducible con respecto a la topología de Zariski. Nótese que $Y \in \mathcal{T}(\mathcal{W}, C)$. Como todo abierto Zariski no vacío de una variedad algebraica irreducible es denso en la variedad, tenemos que $\overline{A_i}^z = V_i$ para $1 \leq i \leq k$. Por lo tanto, tenemos la siguiente descomposición de \overline{Y}^z en componentes irreducibles:

$$\overline{Y}^z = V_1 \cup \dots \cup V_k.$$

Nótese que el conjunto X puede reescribirse, utilizando los conjuntos que acabamos de definir, de la siguiente manera:

$$\begin{aligned} X &= Y \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right) = (A_1 \cup \dots \cup A_k) \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right) = \\ &= \bigcup_{i=1}^k \left(A_i \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right) \right) = \bigcup_{i=1}^k B_i, \end{aligned}$$

donde:

$$B_i := A_i \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right).$$

Para cada i , con $1 \leq i \leq k$, tenemos únicamente dos casos posibles:

- *Caso 1:*

$$B_i := A_i \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right) \neq \emptyset,$$

entonces, B_i es un subconjunto abierto en $\overline{A_i}^z$ para la topología de Zariski.

- *Caso 2:*

$$B_i := A_i \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right) = \emptyset.$$

Salvo reordenación de los índices, supongamos que $r \leq k$ es tal que $B_i \neq \emptyset$ si y solo si $1 \leq i \leq r$. Entonces, tenemos que:

$$X = \bigcup_{i=1}^k B_i = \bigcup_{i=1}^r B_i.$$

Dado que la clausura de una unión finita es la unión de las clausuras, concluimos:

$$\overline{X}^z = \bigcup_{i=1}^r \overline{B_i}^z.$$

Puesto que todo subconjunto abierto Zariski no vacío de una variedad algebraica irreducible es denso en la variedad, obtenemos:

$$\overline{B_i}^z = \overline{A_i \cap \left(\bigcap_{j \notin S} (C \setminus W_j) \right)}^z = \overline{A_i}^z = V_i.$$

Por lo tanto, tenemos la siguiente descomposición de \overline{X}^z en componentes irreducibles:

$$\overline{X}^z = \bigcup_{i=1}^r \overline{B_i}^z = \bigcup_{i=1}^r V_i.$$

Por consiguiente, si V_ℓ es una componente irreducible de \overline{X}^z para algún $\ell \in [r]$, entonces V_ℓ es una componente irreducible de \overline{Y}^z . Así, tomando $V = V_\ell$, concluimos que $V \in \mathcal{D}(\mathcal{Z}(\mathcal{W}, C))$ y se verifica la inclusión de la Identidad (4.1.12). ■

A continuación, probamos que la cardinalidad de $\mathcal{Z}(\mathcal{W}, C)$ es menor o igual que la cardinalidad de $\mathcal{D}(\mathcal{Z}(\mathcal{W}, C))$:

AFIRMACIÓN 4.1.10.3.2. *Con las notaciones precedentes, tenemos que:*

$$(4.1.13) \quad \sharp(\mathcal{Z}(\mathcal{W}, C)) \leq \sharp(\mathcal{D}(\mathcal{Z}(\mathcal{W}, C))).$$

Demostración de la Afirmación. Recordemos, en primer lugar, que $\mathcal{Z}(\mathcal{W}, C)$ es una partición de C en conjuntos localmente cerrados. Vamos a demostrar que para cada $X \in \mathcal{Z}(\mathcal{W}, C)$ no vacío, existe una componente irreducible V_X de \overline{X}^z tal que V_X no es una componente irreducible de ningún otro $Y \in \mathcal{Z}(\mathcal{W}, C)$ con $Y \neq X$.

Argumentamos por reducción al absurdo. Sean $X, Y \in \mathcal{Z}(\mathcal{W}, C)$, con $Y \neq X$. Supongamos que existe una componente irreducible T de \overline{X}^z que es también una componente irreducible de \overline{Y}^z . Dado que T es una componente irreducible de \overline{X}^z , debe existir un subconjunto abierto Zariski $U \subseteq \mathbb{A}^n(K)$ tal que $A = T \cap U$ es una componente localmente cerrada irreducible de X . Por otro lado, existe otro subconjunto abierto Zariski

$U' \subseteq \mathbb{A}^n(K)$ tal que $A' = T \cap U'$ es una componente localmente cerrada irreducible de Y . Como T es irreducible, la intersección de dos conjuntos abiertos Zariski no vacíos en T es no vacía, y tenemos que:

$$\emptyset \neq U \cap U' \cap T \subseteq X \cap Y = \emptyset,$$

lo cual contradice que $X \cap Y = \emptyset$. Por lo tanto, no existe una componente irreducible de \overline{X}^z , con $X \in \mathcal{Z}(\mathcal{W}, C)$, que sea también una componente irreducible de algún otro \overline{Y}^z , con $Y \in \mathcal{Z}(\mathcal{W}, C)$ e $Y \neq X$. Esto demuestra la Desigualdad (4.1.13). ■

Combinando la Desigualdad (4.1.12) y la Inclusión (4.1.13), obtenemos:

$$(4.1.14) \quad \begin{aligned} \sharp(\mathcal{Z}(\mathcal{W}, C)) &\leq \sharp(\mathcal{D}(\mathcal{Z}(\mathcal{W}, C))) \leq \sharp(\mathcal{D}(\mathcal{T}(\mathcal{W}, C))) \leq \\ &\leq \deg(\mathcal{D}(\mathcal{T}(\mathcal{W}, C))) = \sum_{A \in \mathcal{D}(\mathcal{T}(\mathcal{W}, C))} \deg(A). \end{aligned}$$

Aplicando el Lema 4.1.8, concluimos:

$$(4.1.15) \quad \sum_{A \in \mathcal{D}(\mathcal{T}(\mathcal{W}, C))} \deg(A) \leq \deg(C) (1 + \deg(\mathcal{W}))^{\dim(C)}.$$

Como $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W}, C)$, la Afirmación 4.1.10.2 implica:

$$(4.1.16) \quad \sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \sharp(\mathcal{Z}(\mathcal{W}, C)).$$

Combinando las Desigualdades (4.1.14), (4.1.15) y (4.1.16), concluimos finalmente que:

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \deg(C) (1 + \deg(\mathcal{W}))^{\dim(C)}.$$

■

A continuación, acotamos el número de $\widetilde{\mathcal{W}}$ -celdas en un conjunto constructible C por la suma del número de $\widetilde{\mathcal{W}}$ -celdas de los conjuntos localmente cerrados irreducibles en una descomposición de C , donde $\widetilde{\mathcal{W}}$ es una familia finita de conjuntos constructibles.

AFIRMACIÓN 4.1.10.4. *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y $\widetilde{\mathcal{W}} = \{V_1, \dots, V_t\}$ una familia finita de conjuntos constructibles en $\mathbb{A}^n(K)$. Supongamos que la siguiente es una descomposición de C en subconjuntos localmente cerrados irreducibles que satisface las afirmaciones del Lema 1.1.2:*

$$C = C_1 \cup \dots \cup C_m.$$

Entonces, tenemos que:

$$\sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C)) \leq \sum_{i=1}^m \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C_i)).$$

Demostración de la Afirmación. Sea $S \subseteq [t]$ un subconjunto de índices. La $\widetilde{\mathcal{W}}$ -celda en C asociada con el subconjunto S viene dada por la siguiente igualdad:

$$X_S := C \cap \left(\bigcap_{i \in S} V_i \right) \cap \left(\bigcap_{j \in [t] \setminus S} (\mathbb{A}^n(K) \setminus V_j) \right).$$

Para cada i , $1 \leq i \leq m$, definimos el subconjunto constructible $X_S^{(i)} \subseteq C_i$ mediante la siguiente igualdad:

$$X_S^{(i)} := C_i \cap \left(\bigcap_{i \in S} V_i \right) \cap \left(\bigcap_{j \in [t] \setminus S} (\mathbb{A}^n(K) \setminus V_j) \right) \subseteq C_i.$$

Obviamente, $X_S^{(i)}$ es una $\widetilde{\mathcal{W}}$ -celda en C_i y

$$X_S = \bigcup_{i=1}^m X_S^{(i)}.$$

Por lo tanto, para cada $S \subseteq [t]$ tal que $X_S \neq \emptyset$, podemos elegir algún $i(S) \in [m]$ tal que $X_S^{(i(S))} \neq \emptyset$. Esto define la siguiente aplicación:

$$(4.1.17) \quad \begin{aligned} \{S \subseteq [t] : X_S \neq \emptyset\} &\longrightarrow \bigcup_{i=1}^m \mathcal{Z}(\widetilde{\mathcal{W}}, C_i) \\ S &\longmapsto X_S^{(i(S))} \end{aligned}$$

A continuación, vamos a demostrar que esta aplicación es inyectiva. Si S y S' son dos subconjuntos distintos de $[t]$ tales que $X_S \neq \emptyset$, $X_{S'} \neq \emptyset$ y $X_S^{i(S)} = X_{S'}^{i(S')}$, tendríamos:

$$X_S^{i(S)} = X_{S'}^{i(S')} \subseteq X_S \cap X_{S'} = \emptyset,$$

lo que contradice que $X_S^{i(S)} = X_{S'}^{i(S')} \neq \emptyset$. Usando la Afirmación 4.1.10.1, concluimos:

$$\sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C)) \leq \sharp\left(\bigcup_{i=1}^m \mathcal{Z}(\widetilde{\mathcal{W}}, C)\right) \leq \sum_{i=1}^m \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C)),$$

y se verifica la afirmación. ■

Finalmente, estamos en condiciones de demostrar el enunciado de nuestra proposición. Por la Afirmación 4.1.10.2, dado que $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W}, C)$, tenemos que:

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \sharp(\mathcal{Z}(\mathcal{W}, C)).$$

Sea $\widetilde{\mathcal{W}} = \{A_1, \dots, A_s, B_1, \dots, B_s\}$ una familia finita de variedades algebraicas tal que:

- $\widetilde{W}_i = A_i \cap (\mathbb{A}^n(K) \setminus B_i)$, $1 \leq i \leq s$,
- $\deg(\mathcal{W}) = \deg(\widetilde{\mathcal{W}})$.

Además, como $\mathcal{W} \subseteq \mathcal{B}(\widetilde{\mathcal{W}}, C)$, de la Afirmación 4.1.10.2 se deduce:

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \sharp(\mathcal{Z}(\mathcal{W}, C)) \leq \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C)).$$

Consideramos ahora una descomposición de C en subconjuntos localmente cerrados irreducibles que minimizan su grado LCI como en la Definición 13, i.e.

$$C = C_1 \cup \dots \cup C_m,$$

tal que:

$$\deg_{\text{Lci}}(C) = \sum_{i=1}^m \deg(C_i).$$

Gracias a la Afirmación 4.1.10.4, obtenemos:

$$(4.1.18) \quad \sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C)) \leq \sum_{i=1}^m \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C_i)).$$

A continuación, por la Afirmación 4.1.10.3, como C_i es localmente cerrado y los elementos de $\widetilde{\mathcal{W}}$ son subconjuntos cerrados Zariski de $\mathbb{A}^n(K)$, concluimos:

$$(4.1.19) \quad \sharp(\mathcal{Z}(\widetilde{\mathcal{W}}, C_i)) \leq \deg(C_i)(1 + \deg(\widetilde{\mathcal{W}}))^{\dim(C_i)} \leq \deg(C_i)(1 + \deg(\widetilde{\mathcal{W}}))^{\dim(C)}.$$

Para obtener la desigualdad del enunciado de la proposición, basta con combinar las Desigualdades (4.1.18) y (4.1.19):

$$\begin{aligned} \sharp(\mathcal{Z}(\mathcal{H}, C)) &\leq \left(\sum_{i=1}^m \deg(C_i)\right) (1 + \deg(\widetilde{\mathcal{W}}))^{\dim(C)} = \deg_{\text{Lci}}(C)(1 + \deg(\widetilde{\mathcal{W}}))^{\dim(C)} = \\ &= \deg_{\text{Lci}}(C)(1 + \deg(\mathcal{W}))^{\dim(C)}. \end{aligned}$$

□

El siguiente resultado corrige y generaliza el Teorema 2 de [He, 83]:

TEOREMA 4.1.11 (Erzeugungsgrad y cotas combinatorias). *Sea $C \subseteq \mathbb{A}^n(K)$ un conjunto constructible y \mathcal{H} una familia finita de subconjuntos constructibles de $\mathbb{A}^n(K)$. Entonces, tenemos que:*

- i) $\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \deg_{\text{Lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$.
- ii) $\deg_z(\mathcal{B}(\mathcal{H}, C)) \leq \deg_{\text{Lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$.
- iii) *Cualquier subconjunto finito \mathcal{H} -definible de C contiene a lo sumo $\deg_{\text{Lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$ puntos.*
- iv) $\sharp(\mathcal{B}(\mathcal{H}, C)) \leq 2^{\deg_{\text{Lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}}$.

DEMOSTRACIÓN. Sea $\mathcal{W} = \{W_1, \dots, W_s\}$ una familia finita de variedades algebraicas tal que $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W})$ y $\text{grad}(\mathcal{H}) = \text{grad}(\mathcal{W})$. Entonces, tenemos que:

$$\mathcal{H}|_C := \{H \cap C : H \in \mathcal{H}\} \subseteq \mathcal{B}(\mathcal{W}, C).$$

Por lo tanto, gracias a la Proposición 4.1.10, concluimos:

$$\sharp(\mathcal{Z}(\mathcal{H}, C)) \leq \deg_{\text{Lci}}(C)(1 + \text{grad}(\mathcal{W}))^{\dim(C)} = \deg_{\text{Lci}}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}$$

y se obtiene el Ítem *i*).

Para el Ítem *ii*), supongamos que C es localmente cerrado. En primer lugar, recordemos que:

$$\mathcal{D}(\mathcal{B}(\mathcal{H}, C)) = \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{B}(\mathcal{H}, C), V \text{ es una componente irreducible de } \overline{X^z}\},$$

y

$$\deg_z(\mathcal{B}(\mathcal{H}, C)) = \sum_{V \in \mathcal{D}(\mathcal{B}(\mathcal{H}, C))} \deg(V).$$

Dado que cada elemento $X \in \mathcal{B}(\mathcal{H}, C)$ es una unión finita de \mathcal{H} -celdas en C , las componentes irreducibles de $\overline{X^z}$ se encuentran entre las componentes irreducibles de los $\overline{Y^z}$'s, donde Y es una \mathcal{H} -celda en C . Es decir, tenemos que $\mathcal{D}(\mathcal{B}(\mathcal{H}, C)) \subseteq \mathcal{D}(\mathcal{Z}(\mathcal{H}, C))$ y concluimos:

$$(4.1.20) \quad \deg_z(\mathcal{B}(\mathcal{H}, C)) \leq \sum_{V \in \mathcal{D}(\mathcal{Z}(\mathcal{H}, C))} \deg(V).$$

Al igual que en la demostración del Ítem *i*), sea $\mathcal{W} = \{W_1, \dots, W_s\}$ una familia finita de variedades algebraicas en $\mathbb{A}^n(K)$ tal que $\text{grad}(\mathcal{H}) = \text{deg}(\mathcal{W})$ y $\mathcal{H} \subseteq \mathcal{B}(\mathcal{W}, C)$. Como en el Lema 4.1.8, definimos la siguiente clase de subconjuntos localmente cerrados:

$$\mathcal{F}(\mathcal{W}, C) = \left\{ X \subseteq \mathbb{A}^n(K) : \exists S \subseteq \{1, \dots, s\}, X = C \cap \left(\bigcap_{i \in S} W_i \right) \right\}$$

y la siguiente clase de variedades algebraicas irreducibles:

$$\mathcal{D}(\mathcal{F}(\mathcal{W}, C)) = \{V \subseteq \mathbb{A}^n(K) : \exists X \in \mathcal{F}(\mathcal{W}, C), V \text{ es una componente irreducible de } \overline{X^z}\}.$$

A continuación, vamos a demostrar que se cumple la siguiente inclusión:

$$(4.1.21) \quad \mathcal{D}(\mathcal{Z}(\mathcal{H}, C)) \subseteq \mathcal{D}(\mathcal{F}(\mathcal{W}, C)).$$

Para probar esta inclusión, dado que C es localmente cerrado, podemos suponer que $C = W_0 \cap U$, donde $W_0 \subseteq \mathbb{A}^n(K)$ es una variedad algebraica y $U \subseteq \mathbb{A}^n(K)$ es un subconjunto abierto Zariski. Sea $V \in \mathcal{D}(\mathcal{Z}(\mathcal{H}, C))$ una componente irreducible de $\overline{X^z}$, donde:

$$X = C \cap \left(\bigcap_{i \in S} W_i \right) \cap \left(\bigcap_{j \notin S} (\mathbb{A}^n(K) \setminus W_j) \right).$$

para cierto subconjunto $S \subseteq [s]$.

Denotamos por $\tilde{U} \subseteq \mathbb{A}^n(K)$ al subconjunto abierto Zariski dado por:

$$\tilde{U} := U \cap \left(\bigcap_{j \notin S} (\mathbb{A}^n(K) \setminus W_j) \right).$$

A continuación, consideramos la variedad algebraica $W' \subseteq \mathbb{A}^n(K)$ dada por:

$$W' := W_0 \cap \left(\bigcap_{i \in S} W_i \right).$$

Consideramos una descomposición de W' en componentes irreducibles:

$$(4.1.22) \quad W' = V_1 \cup \dots \cup V_m.$$

Salvo reordenación de los índices, existe r , $1 \leq r \leq m$, tal que $\tilde{U} \cap V_k \neq \emptyset$ si y solo si $1 \leq k \leq r$. Entonces, tenemos que:

$$X = W' \cap \tilde{U} = \bigcup_{k=1}^r (V_k \cap \tilde{U})$$

y, por lo tanto,

$$\overline{X^z} = \bigcup_{k=1}^r V_k.$$

En suma, $V \in \{V_1, \dots, V_r\}$.

Por otro lado, sea $Y \in \mathcal{F}(\mathcal{W}, C)$ el siguiente subconjunto localmente cerrado:

$$Y = C \cap \left(\bigcap_{i \in S} W_i \right) = W' \cap U.$$

Consideramos de nuevo la descomposición de W' en componentes irreducibles descrita en la Identidad (4.1.22). Nótese que para cada k , $1 \leq k \leq r$, como $\tilde{U} \cap V_k \neq \emptyset$, entonces $U \cap V_k \neq \emptyset$ (porque $\tilde{U} \subseteq U$). Por consiguiente, salvo reordenación de los índices, existe algún t tal que, $1 \leq r \leq t \leq m$ y $V_k \cap U \neq \emptyset$ si $1 \leq k \leq t$. Por lo tanto, concluimos que:

$$Y = \bigcup_{k=1}^t (V_k \cap U), \quad \bar{Y}^z = \bigcup_{k=1}^t V_k.$$

En particular, $V \in \{V_1, \dots, V_r\} \subseteq \{V_1, \dots, V_t\}$. Además, $Y \in \mathcal{T}(\mathcal{W}, C)$ y $\{V_1, \dots, V_t\} \subseteq \mathcal{D}(\mathcal{T}(\mathcal{W}, C))$. Por lo tanto, $V \in \mathcal{D}(\mathcal{T}(\mathcal{W}, C))$ y se cumple la inclusión de la Identidad (4.1.21).

Entonces, combinando las Identidades (4.1.20) y (4.1.21), y aplicando el Lema 4.1.8, concluimos que si C es localmente cerrado, tenemos que:

$$(4.1.23) \quad \deg_z(\mathcal{B}(\mathcal{H}, C)) \leq \sum_{V \in \mathcal{D}(\mathcal{T}(\mathcal{W}, C))} \deg(V) \leq \deg(C)(1 + \deg(\mathcal{W}))^{\dim(C)}.$$

En cuanto al caso general, consideramos una descomposición de grado LCI mínimo de C :

$$C = C_1 \cup \dots \cup C_m.$$

Nótese que cada elemento en $\mathcal{B}(\mathcal{H}, C)$ es una unión finita de elementos en:

$$\mathcal{B}(\mathcal{H}, C_1) \cup \dots \cup \mathcal{B}(\mathcal{H}, C_m).$$

Por lo tanto, si V es una componente irreducible de \bar{X}^z , donde $X \in \mathcal{B}(\mathcal{H}, C)$, existe algún $i \in [m]$ y algún $Y \in \mathcal{B}(\mathcal{H}, C_i)$ tales que V es una componente irreducible de \bar{Y}^z . En conclusión, tenemos que:

$$\mathcal{D}(\mathcal{B}(\mathcal{H}, C)) \subseteq \bigcup_{i=1}^m \mathcal{D}(\mathcal{B}(\mathcal{H}, C_i)).$$

Finalmente, usando la Identidad (4.1.23), concluimos:

$$\begin{aligned} \deg_z(\mathcal{B}(\mathcal{H}, C)) &= \sum_{V \in \mathcal{D}(\mathcal{B}(\mathcal{H}, C))} \deg(V) \leq \\ &\leq \sum_{i=1}^m \left(\sum_{V \in \mathcal{D}(\mathcal{B}(\mathcal{H}, C_i))} \deg(V) \right) \leq \\ &\leq \sum_{i=1}^m \deg(C_i) (1 + \deg(\mathcal{W}))^{\dim(C_i)} \leq \\ &\leq \deg_{lci}(C) (1 + \deg(\mathcal{W}))^{\dim(C)}. \end{aligned}$$

Como $\deg(W) = \text{grad}(\mathcal{H})$, obtenemos el Ítem *ii*).

Para el Ítem *iii*), nótese que cada punto x de un subconjunto finito \mathcal{H} -definible X es irreducible y, por lo tanto, pertenece a $\mathcal{D}(\mathcal{B}(\mathcal{H}, C))$. Por consiguiente, aplicando el Ítem *ii*) del enunciado, obtenemos que el número de tales puntos satisface:

$$\sharp(X) = \deg(X) \leq \deg_z(\mathcal{B}(\mathcal{H}, C)) \leq \deg_{lci}(C) (1 + \text{grad}(\mathcal{H}))^{\dim(C)},$$

y se verifica el Ítem *iii*) del teorema.

Finalmente, para el Ítem *iv*), basta observar que $\mathcal{Z}(\mathcal{H}, C)$ define una partición de C de modo que cada conjunto constructible en $\mathcal{B}(\mathcal{H}, C)$ se puede escribir como unión finita de algunos elementos de $\mathcal{Z}(\mathcal{H}, C)$. Por lo tanto, aplicando el Ítem *i*) del enunciado, obtenemos:

$$\sharp(\mathcal{B}(\mathcal{H}, C)) \leq 2^{\sharp(\mathcal{Z}(\mathcal{H}, C))} \leq 2^{\deg_{lci}(C)(1 + \text{grad}(\mathcal{H}))^{\dim(C)}},$$

y se completa la demostración del teorema. \square

4.2. Teoría de Vapnik-Chervonenkis sobre cuerpos algebraicamente cerrados

En Teoría del Aprendizaje Computacional, la *función de crecimiento* (ver Definición 12) se utiliza para acotar el número de subconjuntos finitos que pueden ser descritos por un número finito de restricciones de clasificadores binarios. El ingrediente clave en estas cotas es el exponente que aparece en ellas, al que se suele denominar *dimensión* (por ejemplo, en el Lema 1.3.1, la cantidad que aparece en el exponente de la cota superior es la dimensión de Vapnik-Chervonenkis). Los distintos contextos (casos binario, multiclase o continuo) de la Teoría del Aprendizaje Computacional generan diferentes nociones de dimensión. El objetivo de la función de crecimiento empleada en la Teoría del Aprendizaje Computacional es proporcionar una cota para la longitud mínima del *conjunto de datos* necesario en un problema de aprendizaje, de forma que se garantice una alta probabilidad de éxito (o una baja probabilidad de error) para cualquier algoritmo de aprendizaje que trabaje con muestras aleatorias.

Las cotas superiores de [He, 83] y nuestras generalizaciones en la Proposición 4.1.10 y el Teorema 4.1.11 proporcionan una especie de *función de crecimiento* para acotar $\sharp(\mathcal{Z}(\mathcal{H}, C))$ superiormente. En este caso, el exponente que aparece en dichas cotas es la dimensión de Krull. Este enfoque es parecido al descrito en la Subsección 1.3 en el contexto de la Teoría del Aprendizaje Computacional aunque con un propósito distinto. El objetivo de esta sección es hacer explícito lo que aparentemente es similar en ambos enfoques y mostrar algunas consecuencias.

4.2.1. Dimensión VC y clasificadores constructibles. Una gran cantidad de investigaciones en la literatura matemática han estudiado la dimensión de Vapnik-Chervonenkis de clasificadores definidos por conjuntos semi-algebraicos (cf. [GJ, 95], [CKKLW, 95], [KM, 97], [MPC, 08], [MP, 09] y sus referencias). La mayor parte de estas investigaciones se basan en las cotas superiores para el número de componentes conexas de conjuntos semi-algebraicos obtenidas por J. Milnor ([Mi, 64]), R. Thom ([Th, 65]), O. Oleinik e I. Petrovski ([Ol, 51], [OP, 62]) o H. E. Warren ([Wa, 68]). En nuestro contexto, estas cotas superiores topológicas solo son aplicables al estudio de la dimensión de Vapnik-Chervonenkis de clasificadores de $\mathbb{A}^n(K)$, donde K es un cuerpo algebraicamente cerrado de característica cero. Esto es, no pueden aplicarse en el caso de característica positiva.

Completamos este vacío considerando familias de clasificadores binarios definidas por clases parametrizadas de conjuntos constructibles. Además, abordamos el caso en el que el conjunto de los parámetros también es un conjunto constructible. La idea de parámetro hace referencia a la existencia de alguna “función” φ que parte de una clase de parámetros Λ y para cada $\lambda \in \Lambda$ asocia un conjunto constructible $\varphi(\lambda) \subseteq \mathbb{A}^n(K)$. Esto lleva, de manera natural, a pensar en una variedad de incidencia $V(\varphi)$, asociada al “grafo” de la parametrización, de tal modo que $V(\varphi) \subseteq \Lambda \times \mathbb{A}^n(K)$, donde para cada $\lambda \in \Lambda$, los elementos de la forma $(\{\lambda\} \times \mathbb{A}^n(K)) \cap V(\varphi)$ identifican a los elementos del conjunto constructible $\varphi(\lambda)$. Hemos escogido esta interpretación por ser más general y ajustarse a la idea de parametrización. Por ello, procedemos del siguiente modo para definir familias de conjuntos constructibles parametrizadas por otro constructible:

Consideramos $N, n \in \mathbb{N}$ dos enteros positivos y $V \subseteq \mathbb{A}^N(K) \times \mathbb{A}^n(K)$ un conjunto constructible. Por ejemplo, V puede ser el grafo de una aplicación constructible. También consideramos un subconjunto constructible $\Lambda \subseteq \mathbb{A}^N(K)$, denominado “el conjunto constructible formado por los parámetros de nuestra familia de clasificadores”. Además, consideramos las dos proyecciones canónicas restringidas a V :

$$\begin{array}{ccc} & V & \\ \pi_1 \swarrow & & \searrow \pi_2 \\ \mathbb{A}^N(K) & & \mathbb{A}^n(K) \end{array}$$

Con estas notaciones, definimos la siguiente clase de conjuntos constructibles:

$$\mathcal{C}(V, \Lambda) := \{\pi_2(\pi_1^{-1}(\{a\})) : a \in \Lambda\}.$$

Definimos también $\text{grad}(V) = \text{grad}(\{V\})$ como el Erzeugungsgrad de la clase formada únicamente por V . Finalmente, definimos la clase de clasificadores dada por las funciones características definidas por los subconjuntos en $\mathcal{C}(V, \Lambda)$:

$$\mathcal{H}(V, \Lambda) := \{\chi_U : U \in \mathcal{C}(V, \Lambda)\}.$$

Al igual que en la Subsección 1.3, para cada subconjunto finito $X \subseteq \mathbb{A}^n(K)$, definimos la clase de restricciones a X de los clasificadores en $\mathcal{H}(V, \Lambda)$ como sigue:

$$\mathcal{H}(V, \Lambda)|_X := \{\chi|_X : \chi \in \mathcal{H}(V, \Lambda)\}.$$

Entonces, se verifica el siguiente resultado:

TEOREMA 4.2.1 (**Cota superior de la función de crecimiento**). *Con las mismas notaciones e hipótesis que antes, se cumple la siguiente desigualdad:*

$$\sharp(\mathcal{H}(V, \Lambda)|_X) \leq \deg_{\text{Ici}}(\Lambda)(1 + \sharp(X) \text{grad}(V))^{\dim(\Lambda)}.$$

Además, tenemos que:

$$G(\mathcal{H}(V, \Lambda), m) \leq \deg_{\text{Ici}}(\Lambda)(1 + m \text{grad}(V))^{\dim(\Lambda)}$$

DEMOSTRACIÓN. Denotemos por Π_1 y Π_2 respectivamente a las proyecciones canónicas de $\mathbb{A}^N(K) \times \mathbb{A}^n(K)$ sobre $\mathbb{A}^N(K)$ y $\mathbb{A}^n(K)$. Obviamente, tenemos que:

$$\pi_1 = \Pi_1|_V \text{ y } \pi_2 = \Pi_2|_V.$$

Supongamos, como se indica en la Observación 4.1.7, que $\mathcal{W} = \{W_1, \dots, W_s\}$ es una familia finita de variedades algebraicas irreducibles $W_i \subseteq \mathbb{A}^N(K) \times \mathbb{A}^n(K)$ tales que $V \in \mathcal{B}(\mathcal{W})$ y se cumple la siguiente igualdad:

$$(4.2.1) \quad \text{grad}(V) = \deg(\mathcal{W}) = \sum_{j=1}^s \deg(W_j).$$

Supongamos también que $X = \{x_1, \dots, x_m\} \subseteq \mathbb{A}^n(K)$ con $m = \sharp(X)$. Para cada $i \in [m]$, sea $L_i \subseteq \mathbb{A}^N(K) \times \mathbb{A}^n(K)$ la variedad afin lineal dada por la siguiente identidad:

$$L_i := \Pi_2^{-1}(\{x_i\}) = \mathbb{A}^N(K) \times \{x_i\}.$$

A continuación, consideramos las secciones de V con L_i y, luego, proyectamos mediante Π_1 . Es decir, definimos:

- Los subconjuntos constructibles:

$$V_i := \Pi_1(V \cap L_i) = \pi_1(\pi_2^{-1}(\{x_i\})) \subseteq \mathbb{A}^N(K).$$

- Las proyecciones:

$$W_{j,i} := \Pi_1(W_j \cap L_i) \subseteq \mathbb{A}^N(K),$$

donde $j \in [s]$ y $i \in [m]$.

Además, tenemos que se verifican las siguientes igualdades:

- $V \cap L_i = V_i \times \{x_i\}$, para cada $i \in [m]$.
- $W_j \cap L_i = W_{j,i} \times \{x_i\}$, para cada $j \in [s]$, $i \in [m]$.

En particular, tenemos que, como $W_j \cap L_i$ es una variedad algebraica, $W_{j,i}$ es también una variedad algebraica de la misma dimensión: $W_j \cap L_i$ y $W_{j,i}$ son birregularmente isomorfas. Además, dado que el isomorfismo birregular entre W_j y $W_{j,i}$ es lineal (la proyección Π_1 restringida a $W_j \cap L_i$) y su inversa es también lineal, concluimos inmediatamente, a partir de la Desigualdad de Bézout, que:

$$(4.2.2) \quad \deg(W_{j,i}) = \deg(W_j \cap L_i) \leq \deg(W_j).$$

Finalmente, nótese que, como $V \in \mathcal{B}(\mathcal{W})$, la familia finita de conjuntos constructibles:

$$\mathcal{V} = \{V_1, \dots, V_m\} \subseteq \mathcal{B}(\widetilde{\mathcal{W}}),$$

donde $\widetilde{\mathcal{W}}$ es la familia finita de variedades algebraicas dadas por la siguiente identidad:

$$\widetilde{\mathcal{W}} := \{W_{j,i} : j \in [s], i \in [m]\}.$$

En particular, concluimos:

$$(4.2.3) \quad \text{grad}(\mathcal{V}) \leq \deg(\widetilde{\mathcal{W}}) \leq \sum_{i=1}^m \sum_{j=1}^s \deg(W_{j,i}).$$

Combinando las Desigualdades (4.2.2) y (4.2.3) y la Identidad (4.2.1), obtenemos:

$$(4.2.4) \quad \text{grad}(\mathcal{V}) \leq \sum_{i=1}^m \sum_{j=1}^s \deg(W_j) = m \text{grad}(V).$$

Por otro lado, vamos a demostrar que se verifica la siguiente igualdad:

$$(4.2.5) \quad \sharp(\mathcal{H}(V, \Lambda)|_X) = \sharp(\mathcal{Z}(\mathcal{V}, \Lambda)),$$

donde $\mathcal{Z}(\mathcal{V}, \Lambda)$ es el conjunto de \mathcal{V} -celdas no vacías en Λ . Para probar la Identidad (4.2.5), consideremos $S \subseteq [m]$ y la celda:

$$\Lambda_S = \Lambda \cap \left(\bigcap_{i \in S} V_i \right) \cap \left(\bigcap_{j \notin S} (\mathbb{A}^N(K) \setminus V_j) \right) \subseteq \Lambda.$$

Si Λ_S es una \mathcal{V} -celda no vacía, sea $a \in \Lambda_S$ cualquier punto. Entonces, consideramos:

$$V(a) = \pi_2(\pi_1^{-1}(\{a\})) \in \mathcal{C}(V, \Lambda),$$

y su función característica $\chi := \chi_{V(a)}$. Por lo tanto, tenemos que:

$$\chi|_X(x_k) = 1 \text{ si y solo si } k \in S,$$

y $\chi|_X \in \mathcal{H}(V, \Lambda)|_X$. Recíprocamente, sea $\chi \in \mathcal{H}(V, \Lambda)|_X$ y $U \in \mathcal{C}(V, \Lambda)$ tal que $\chi = \chi_U|_X$. Entonces, existe $a \in \Lambda$ tal que $U = \pi_2(\pi_1^{-1}(\{a\}))$. Consideramos $S = \{i \in [m] : \chi(x_i) = 1\}$ y:

$$\Lambda_S = \Lambda \cap \left(\bigcap_{i \in S} V_i \right) \cap \left(\bigcap_{j \notin S} (\mathbb{A}^N(K) \setminus V_j) \right) \subseteq \Lambda.$$

Claramente, $a \in \Lambda_S$ y, por consiguiente, Λ_S es una \mathcal{V} -celda no vacía en Λ . Por lo tanto, hemos obtenido una biyección entre las funciones en $\mathcal{H}(V, \Lambda)|_X$ y las celdas no vacías en $\mathcal{Z}(\mathcal{V}, \Lambda)$, lo que demuestra la Identidad (4.2.5).

Finalmente, concluimos a partir de la Identidad (4.2.5), el Teorema 4.1.11 y la Desigualdad (4.2.4) la siguiente cadena de igualdades y desigualdades:

$$\begin{aligned} \sharp(\mathcal{H}(V, \Lambda)|_X) &= \sharp(\mathcal{Z}(\mathcal{V}, \Lambda)) \leq \deg_{\text{lci}}(\Lambda) (1 + \text{grad}(\mathcal{V}))^{\dim(\Lambda)} \leq \\ &\leq \deg_{\text{lci}}(\Lambda) (1 + m \text{grad}(V))^{\dim(\Lambda)}. \end{aligned}$$

Para la cota sobre la función de crecimiento, basta con observar que:

$$\begin{aligned} G(\mathcal{H}(V, \Lambda), m) &= \sup\{\sharp(\mathcal{H}(V, \Lambda)|_X) : X \subseteq \mathbb{A}^n(K), \sharp(X) = m\} \leq \\ &\leq \deg_{\text{lci}}(\Lambda) (1 + m \text{grad}(V))^{\dim(\Lambda)}. \end{aligned}$$

□

Nuestra conclusión principal a partir de estas discusiones técnicas es que la dimensión VC de una familia de clasificadores está linealmente acotada, salvo por unas cantidades logarítmicas basadas en la Teoría de la Intersección, por la dimensión de Krull del espacio de parámetros. En concreto, tenemos que:

COROLARIO 4.2.2. *Con las mismas notaciones e hipótesis precedentes, se verifica la siguiente desigualdad:*

$$\frac{s}{\log_2(s) + k} - \frac{\log_2(\deg_{\text{lci}}(\Lambda))}{\log_2(s) + k} \leq \dim(\Lambda),$$

donde $s = \dim_{VC}(\mathcal{H}(V, \Lambda))$ y $k = 1 + \log_2(\text{grad}(V))$.

4.2.2. Dimensión VC de abiertos distinguidos definidos por una clase de polinomios. Dado un polinomio $f \in K[X_1, \dots, X_n]$, llamamos *abierto distinguido definido por f* al siguiente abierto Zariski:

$$D(f) := \{x \in \mathbb{A}^n(K) : f(x) \neq 0\} = \mathbb{A}^n(K) \setminus V_{\mathbb{A}}(f).$$

Consideramos el conjunto constructible $V \subseteq \mathbb{A}^N(K) \times \mathbb{A}^n(K)$ (el complementario de la variedad solución en [SS, 93]), donde $N := N_d$, dado por la siguiente igualdad:

$$V := \{(f, x) \in \mathbb{A}^N(K) \times \mathbb{A}^n(K) : f(x) \neq 0\}.$$

Tenemos las dos proyecciones canónicas $\pi_1 : V \rightarrow P_d^K(X_1, \dots, X_n)$ y $\pi_2 : V \rightarrow \mathbb{A}^n(K)$. Sea $\Omega \subseteq \mathbb{A}^N(K)$ un conjunto constructible que define una familia de polinomios a través de sus coeficientes. Nótese que V es el complementario de una hipersuperficie algebraica de grado $d + 1$. Por lo tanto, $\text{grad}(V) = d + 1$.

Además, Ω define una familia de abiertos distinguidos de $\mathbb{A}^n(K)$:

$$\mathcal{C}_d(V, \Omega) = \{\pi_2(\pi_1^{-1}(\{a\})) : a \in \Omega\}.$$

Definimos ahora la familia de clasificadores dada por las funciones características de los abiertos distinguidos de $\mathbb{A}^n(K)$ definidos por una ecuación $f \in \Omega$ de la siguiente manera:

$$\mathcal{H}_d(V, \Omega) := \{\chi_{D(f)} : f \in \Omega\}.$$

TEOREMA 4.2.3. *Con las notaciones e hipótesis previas, para todo conjunto finito $X \subseteq \mathbb{A}^n(K)$, se verifica la siguiente desigualdad:*

$$\sharp(\mathcal{H}_d(V, \Omega)|_X) \leq \deg_{\text{Ici}}(\Omega)(1 + \sharp(X)(d+1))^{\dim(\Omega)},$$

En particular, si $\sharp(X) = \dim_{VC}(\mathcal{H}_d(V, \Omega)) = s$, tenemos que:

$$\frac{s}{\log_2(s) + 2 + \log_2(d)} - \frac{\log_2(\deg_{\text{Ici}}(\Omega))}{\log_2(s) + 2 + \log_2(d)} \leq \dim(\Omega).$$

Además,

$$G(\mathcal{H}_d(V, \Omega), m) \leq \deg_{\text{Ici}}(\Omega)(1 + m(d+1))^{\dim(\Omega)}.$$

DEMOSTRACIÓN. Se trata de una aplicación inmediata del Teorema 4.2.1 de la subsección anterior. Basta observar que la clase $\mathcal{W} := \{\mathbb{A}^{N+n}(K) \setminus V\}$ es una familia finita de variedades algebraicas tal que $V \in \mathcal{B}(\mathcal{W})$ y $\deg(\mathbb{A}^{N+n}(K) \setminus V) = d+1$. \square

4.2.3. Variedades evasivas. En esta subsección estudiaremos las variedades evasivas. Básicamente, una variedad evasiva es una variedad algebraica que verifica “ciertas propiedades” respecto a su intersección con variedades algebraicas definidas por polinomios que pertenecen a un conjunto constructible. En la literatura matemática podemos encontrar varias nociones de “variedades evasivas” distintas y con propósitos diferentes pero que guardan cierta relación entre sí (ver, por ejemplo, las definiciones de [DKL, 14] y [Guo, 21]).

Hemos estructurado esta subsección de la siguiente manera. En primer lugar, mejoramos el resultado principal de [DKL, 14] usando la Proposición 2.3 de [HS, 82]. A continuación, presentamos nuestra definición de “variedad evasiva” y, bajo ciertas hipótesis, demostramos su existencia. Por último, partiendo del Corolario 3.2.5 y usando técnicas de conjuntos cuestos, probamos la existencia de “variedades evasivas para la dimensión”.

En la Subsección 4.2.4 demostraremos que, bajo ciertas condiciones, los conjuntos cuestos están densamente distribuidos en variedades evasivas de dimensión positiva para cualquier distribución de probabilidad razonable (cf. Teorema 4.2.9).

4.2.3.1. *Mejora del Teorema 2.2 de [DKL, 14].* En [DKL, 14], una “variety evasive set” es un conjunto $U \subseteq \mathbb{A}^n(K)$ tal que, para toda variedad algebraica $V \subseteq \mathbb{A}^n(K)$ de grado y dimensión dados, $U \cap V$ es una variedad cero-dimensional (i.e. un conjunto finito) con una cota superior en la cardinalidad de la intersección $\sharp(U \cap V)$. Su resultado principal es una familia de ecuaciones (esencialmente un Sistema de Pham) cuyos ceros definen un conjunto algebraico $U \subseteq \mathbb{A}^n(\overline{\mathbb{F}}_q)$ con “buenas propiedades”, donde \mathbb{F}_q es un cuerpo finito y $\overline{\mathbb{F}}_q$ es la clausura algebraica de \mathbb{F}_q .

Aplicando simplemente la Proposición 2.3 de [HS, 82] (o nuestra generalización en el Teorema 2.2.2), se puede mejorar la cota superior para $\sharp(U \cap V)$ en el Teorema 2.2 de [DKL, 14] de la siguiente forma:

PROPOSICIÓN 4.2.4 (Mejora de las cotas superiores del Teorema 2.2 de [DKL, 14]). *Sean $k, D \geq 1$ enteros positivos y sea $m > k$ un entero positivo tal que $m \mid n$. Sean $d_1 > d_2 > \dots > d_m > D$ enteros positivos primos entre sí. Sea $A := (a_{i,j}) \in \mathcal{M}_{k \times m}(\mathbb{F}_q)$ una matriz k -regular. Definimos los polinomios $f_1, \dots, f_k \in \mathbb{F}_q[X_1, \dots, X_m]$ de la siguiente manera:*

$$f_i(X_1, \dots, X_m) := \sum_{j=1}^m a_{i,j} X_j^{d_j}.$$

Sea $U := V(f_1, \dots, f_k) \subseteq \mathbb{A}^m(\overline{\mathbb{F}}_q)$ la variedad algebraica de sus ceros comunes. Finalmente, sea $U^{n/m} \subseteq \mathbb{A}^n(\overline{\mathbb{F}}_q)$ la variedad dada como el producto cartesiano de U consigo mismo n/m veces. Entonces, para toda variedad algebraica afín $V \subseteq \mathbb{A}^n(\overline{\mathbb{F}}_q)$ de grado D y dimensión k , la intersección $V \cap U^{n/m}$ es una variedad cero-dimensional (i.e. un conjunto finito). Además, su cardinalidad verifica:

$$(4.2.6) \quad \sharp(V \cap U^{n/m}) \leq \deg(V) \cdot d_1^k = D \cdot d_1^k.$$

DEMOSTRACIÓN. La construcción de U se debe a [DKL, 14]. A continuación, demostramos la cota superior para la cardinalidad de la intersección.

En primer lugar, nótese que U es la intersección de m hipersuperficies de grado a lo sumo d_1 . Por lo tanto, $U^{n/m}$ es la intersección de $n/m \cdot k$ hipersuperficies, cada una de ellas de grado a lo sumo d_1 . Para probar esto, definimos para cada i , $1 \leq i \leq k$, y para cada j , $1 \leq j \leq n/m$, los polinomios $f_{i,j} \in \mathbb{F}_q[X_1, \dots, X_n]$ de la siguiente forma:

$$f_{i,j} := f_i(X_{m(j-1)+1}, \dots, X_{mj}) \in \mathbb{F}_q[X_1, \dots, X_n].$$

Entonces, tenemos que:

$$U^{n/m} = \bigcap_{i=1}^k \bigcap_{j=1}^{n/m} V_{\mathbb{A}^n(\overline{\mathbb{F}}_q)}(f_{i,j}).$$

De lo anterior, se deduce que:

$$\sharp(V \cap U^{n/m}) = \deg(V \cap U^{n/m}) = \deg \left(V \cap \left(\bigcap_{i=1}^k \bigcap_{j=1}^{n/m} V_{\mathbb{A}^n(\overline{\mathbb{F}}_q)}(f_{i,j}) \right) \right).$$

Por consiguiente, aplicando la a Proposición 2.3 de [HS, 82], obtenemos:

$$\sharp(V \cap U^{n/m}) = \deg(V) \left(\max\{\deg(V_{\mathbb{A}^n(\overline{\mathbb{F}}_q)}(f_{i,j})) : 1 \leq i \leq k, 1 \leq j \leq n/m\} \right)^{\dim(V)}.$$

Finalmente, concluimos:

$$\sharp(V \cap U^{n/m}) \leq \deg(V) \cdot d_1^{\dim(V)} = D \cdot d_1^k,$$

como queríamos demostrar. \square

OBSERVACIÓN 4.2.5. Nuestra cota superior en la Desigualdad (4.2.6) mejora considerablemente la cota superior de [DKL, 14]. Como puede observar el lector, usando las notaciones precedentes, la cota superior principal presentada en [DKL, 14] es:

$$D^{k+1} \left(\prod_{i=1}^k d_i \right)^k,$$

la cual es mucho más débil que la que mostramos en la Desigualdad (4.2.6) anterior.

4.2.3.2. *Variedades evasivas para un conjunto constructible de listas de polinomios.* En nuestro caso, definimos las variedades evasivas como aquellas que no están completamente inmersas en las variedades definidas por una familia de polinomios en un conjunto constructible $\Omega \subseteq \mathcal{P}_{(d)}^K$.

DEFINICIÓN 26. Sea $(d) := (d_1, \dots, d_r)$ una lista de grados con $r \leq n - 1$. Sea $\Omega \subseteq \mathcal{P}_{(d)}^K(X_1, \dots, X_n)$ un conjunto constructible de listas de polinomios (ver Subsección 3.1.1 para esta notación) y $\Sigma \subseteq \Omega$ un subconjunto constructible de co-dimensión al menos 1 en Ω . Supongamos que para toda lista $f := (f_1, \dots, f_r) \in \Omega \setminus \Sigma$, la variedad $V_{\mathbb{A}}(f) := V_{\mathbb{A}}(f_1, \dots, f_r)$ tiene dimensión $n - r$. Una subvariedad equidimensional $V \subseteq \mathbb{A}^n(K)$ recibe el nombre de evasiva para intersecciones completas en Ω con respecto a Σ si se verifica la siguiente propiedad:

$$\forall f \in \Omega, f|_V \equiv 0 \Rightarrow f \in \Sigma.$$

A continuación, demostramos que este tipo de variedades evasivas existen:

PROPOSICIÓN 4.2.6. Sea $(d) := (d_1, \dots, d_r)$ una lista de grados con $r \leq n - 1$ y $\Omega \subseteq \mathcal{P}_{(d)}^K(X_1, \dots, X_n)$ un conjunto constructible de listas de polinomios. Sea $\Sigma \subseteq \Omega$ un conjunto constructible de co-dimensión al menos 1 tal que para todo $f := (f_1, \dots, f_r) \in \Omega \setminus \Sigma$, la variedad $V_{\mathbb{A}}(f) := V_{\mathbb{A}}(f_1, \dots, f_r)$ tiene dimensión $n - r$. Sea $m \in \mathbb{N}$ un entero positivo tal que $1 \leq m \leq n - r$. Sea $\Delta \in \mathbb{N}$ un entero positivo tal que $\Delta^r > \mathcal{D}_{(d)} := \prod_{i=1}^r d_i$, donde $\mathcal{D}_{(d)}$ es el número de Bézout de la lista de grados (d) . Si $V \subseteq \mathbb{A}^n(K)$ es una variedad equidimensional de dimensión m que satisface las siguientes propiedades:

- i) V está dada como la variedad solución de un conjunto de polinomios de grado a lo sumo Δ y,
- ii) $\deg(V) = \Delta^{n-m}$, donde $\deg(V)$ es el grado de V como en la Sección 1.2,

entonces, V es evasiva para intersecciones completas en Ω con respecto a Σ .

DEMOSTRACIÓN. Supongamos que existe una lista de polinomios $g_1, \dots, g_s \in K[X_1, \dots, X_n]$ de grado a lo sumo Δ , tal que la variedad algebraica de sus ceros comunes $V := V_{\mathbb{A}}(g_1, \dots, g_s)$ es equidimensional de grado Δ^{n-m} .

Para cada $f := (f_1, \dots, f_r) \in \Omega \setminus \Sigma$, denotamos por $V_{\mathbb{A}}(f)$ a la variedad algebraica $V_{\mathbb{A}}(f) := V_{\mathbb{A}}(f_1, \dots, f_r)$, la cual tiene dimensión $n - r$. Como $\deg(f_i) \leq d_i$, para cada i , $1 \leq i \leq r$, tenemos que:

$$\deg(V_{\mathbb{A}}(f_1, \dots, f_r)) \leq \prod_{i=1}^r \deg(f_i) \leq \prod_{i=1}^r d_i = \mathcal{D}_{(d)}.$$

Consideramos ahora el K -espacio vectorial de matrices $\mathcal{M}_{m \times (n+1)}(K)$. Para cada matriz $A \in \mathcal{M}_{m \times (n+1)}(K)$, tenemos la siguiente función afín:

$$A : \mathbb{A}^n(K) \longrightarrow \mathbb{A}^m(K) \\ x \longmapsto (A_1(x), \dots, A_m(x))'$$

donde cada $A_i(x) := a_{i,1}X_1 + \cdots + a_{i,n}X_n + a_{i,n+1}$ es una función afín lineal (i.e. un polinomio de grado 1). Para cada $f \in \Omega \setminus \Sigma$ y para cada $A \in \mathcal{M}_{m \times (n+1)}(K)$, denotamos por $W_A(f)$ a la siguiente intersección:

$$W_{\mathbb{A}}(f) := V_{\mathbb{A}}(f) \cap V_{\mathbb{A}}(A_1, \dots, A_m).$$

Como $1 \leq m \leq n - r = \dim(V_{\mathbb{A}}(f))$, por el Lema de Normalización de Noether, existe un abierto Zariski no vacío $U_f \subseteq \mathcal{M}_{m \times (n+1)}(K)$ de matrices A tal que se cumple la siguiente igualdad de dimensiones:

$$(4.2.7) \quad \dim(W_{\mathbb{A}}(f)) = \dim(V_{\mathbb{A}}(f) \cap (V_{\mathbb{A}}(A_1, \dots, A_m))) = \dim(V_{\mathbb{A}}(f)) - m = n - r - m.$$

El conjunto abierto U_f es el conjunto definido por submatrices de m filas de cualquier normalización de Noether genérica de $V_{\mathbb{A}}(f)$.

Por otro lado, como V es pura (i.e. todas sus componentes irreducibles tienen la misma dimensión), por la Proposición 1.2.2, existe también un abierto Zariski no vacío $U_1 \subseteq \mathcal{M}_{m \times (n+1)}(K)$ de matrices tal que, si $A \in U_1$, entonces $V \cap V_{\mathbb{A}}(A_1, \dots, A_m)$ es un conjunto finito y se verifica la siguiente igualdad:

$$(4.2.8) \quad \deg(V \cap V_{\mathbb{A}}(A_1, \dots, A_m)) = \sharp(V \cap V_{\mathbb{A}}(A_1, \dots, A_m)) = \deg(V) = \Delta^{n-m}.$$

Como $\mathcal{M}_{m \times (n+1)}(K)$ es irreducible, cualquier abierto no vacío es denso Zariski y, entonces, la intersección de dos abiertos no vacíos es no vacía. Por lo tanto, para cada matriz $A \in U := U_f \cap U_1 \neq \emptyset$, también se cumple lo siguiente:

$$(4.2.9) \quad \deg(W_{\mathbb{A}}(f)) = \deg(V_{\mathbb{A}}(f) \cap V_{\mathbb{A}}(A_1 \cap \dots \cap A_m)) \leq \deg(V_{\mathbb{A}}(f)) \leq \mathcal{D}_{(d)}.$$

Por consiguiente, aplicando la Proposición 2.3 de [HS, 82], obtenemos:

$$\begin{aligned} \deg(W_{\mathbb{A}}(f) \cap V) &= \deg(W_{\mathbb{A}}(f) \cap V_{\mathbb{A}}(g_1, \dots, g_s)) \leq \\ &\leq \deg(W_{\mathbb{A}}(f)) (\max\{\deg(g_i) : 1 \leq i \leq s\})^{\dim(W_{\mathbb{A}}(f))}. \end{aligned}$$

A partir de la Igualdad (4.2.7) y la Desigualdad (4.2.9), obtenemos:

$$\deg(W_{\mathbb{A}}(f) \cap V) \leq \mathcal{D}_{(d)} (\max\{\deg(g_i) : 1 \leq i \leq s\})^{n-r-m} \leq \mathcal{D}_{(d)} \Delta^{n-m-r}.$$

Como $\Delta^r > \mathcal{D}_{(d)} = \prod_{i=1}^r d_i$, de la Igualdad (4.2.8) concluimos:

$$\begin{aligned} \Delta^{n-m} = \deg(V \cap V_{\mathbb{A}}(A_1, \dots, A_m)) &= \Delta^r \Delta^{n-m-r} > \mathcal{D}_{(d)} \Delta^{n-m-r} \geq \\ &\geq \sharp((V \cap V_{\mathbb{A}}(A_1, \dots, A_m) \cap V_{\mathbb{A}}(f))). \end{aligned}$$

En particular, debe existir un punto $x \in V \cap V_{\mathbb{A}}(A_1, \dots, A_m)$ que no está en $V_{\mathbb{A}}(f)$. En otras palabras, debe existir un punto $x \in V$ tal que $f(x) \neq 0 \in K^r$. Por lo tanto, concluimos que V es evasiva para intersecciones completas en Ω con respecto a Σ . \square

Un caso particular es el de las hipersuperficies, que puede definirse de la siguiente forma:

DEFINICIÓN 27. Sea $d \in \mathbb{N}$ un entero positivo. Sea $\Omega \subseteq P_d^K(X_1, \dots, X_n)$ un conjunto constructible tal que $\Omega \neq \{0\}$. Una subvariedad equidimensional $V \subseteq \mathbb{A}^n(K)$ recibe el nombre de evasiva para hipersuperficies en Ω si se verifica la siguiente propiedad:

$$\forall f \in \Omega, f|_V \equiv 0 \Rightarrow f = 0.$$

Tomando $r = 1$ en la Proposición 4.2.6, concluimos fácilmente que existen las variedades evasivas para hipersuperficies:

COROLARIO 4.2.7. Sea $d \in \mathbb{N}$ un entero positivo y $\Omega \subseteq P_d^K(X_1, \dots, X_n)$ un conjunto constructible. Sea $m \in \mathbb{N}$ un entero positivo tal que $1 \leq m \leq n - 1$. Sea $\Delta \in \mathbb{N}$ otro entero positivo tal que $\Delta > d$. Si $V \subseteq \mathbb{A}^n(K)$ es una variedad equidimensional de dimensión m que verifica las siguientes propiedades:

- i) V está dada como la variedad solución de un conjunto de polinomios de grado a lo sumo Δ y,
- ii) $\deg(V) = \Delta^{n-m}$,

entonces, V es evasiva para hipersuperficies en Ω .

4.2.3.3. *Variedades evasivas para la dimensión.* El Corolario 3.2.5 nos proporciona condiciones suficientes para que existan variedades de intersección completa que sean “evasivas para la dimensión” (en [Guo, 21] aparece otra variante que también tiene que ver con la dimensión) respecto a cualquier sistema de ecuaciones de un conjunto constructible dado. Dada una lista de grados $(d) := (d_1, \dots, d_m)$ y un conjunto constructible $\tilde{\Omega} \subseteq \mathcal{P}_{(d)}$, decimos que una variedad $V \subseteq \mathbb{A}^n(K)$ es evasiva para la dimensión respecto a $\tilde{\Omega}$ si se verifica lo siguiente:

$$\forall f = (f_1, \dots, f_m) \in \tilde{\Omega}, \dim(V \setminus V_{\mathbb{A}}(f_1, \dots, f_m)) = \dim(V).$$

En el caso de que V sea equidimensional, la propiedad anterior significa simplemente que para cualquier sistema $f \in \tilde{\Omega}$, existe alguna componente irreducible de V que no está completamente inmersa en $V_{\mathbb{A}}(f_1, \dots, f_m)$ o, equivalentemente, que siempre existe una componente irreducible que “evade” cualquier variedad definida por ecuaciones de $\tilde{\Omega}$.

COROLARIO 4.2.8 (Variedades evasivas para la dimensión). Sean $m, n \in \mathbb{N}$ dos enteros positivos, con $m \leq n$, y sean $(d) := (d_1, \dots, d_m)$ una lista de grados y $d := \max\{d_1, \dots, d_m\}$. Sean $\Sigma \subseteq \Omega$ dos subconjuntos constructibles de $\mathcal{P}_{(d)}(X_1, \dots, X_n)$ tales que Σ tiene co-dimensión al menos 1 en Ω . Supongamos que $\Omega \setminus \Sigma$ satisface la siguiente propiedad:

$$\forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $V \subseteq \mathbb{A}^n(K)$ cualquier variedad de intersección completa tal que $V := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ de co-dimensión $r \geq (n - m) + m/2 + 1/2$ y $\deg(V) \geq \delta^r$, donde $\delta := \min\{\deg(h_1), \dots, \deg(h_r)\}$. Supongamos que se verifican las siguientes propiedades:

- i) $\log(\delta) \geq \max\{2(1 + \log(d + 1)), \frac{2 \log(\deg_{\text{lc}}(\Omega))}{\dim(\Omega)}\}$ y,
- ii) $\max\{\deg(h_1), \dots, \deg(h_r)\} \leq (1 + \frac{1}{n-m})\delta$.

Entonces, V evade a $\Omega \setminus \Sigma$ respecto a la dimensión.

DEMOSTRACIÓN. Supongamos que V no evade a $\Omega \setminus \Sigma$. Entonces, existe algún $f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma$ tal que:

$$(4.2.10) \quad \dim(V \setminus V_{\mathbb{A}}(f_1, \dots, f_m)) < \dim(V).$$

Como V es de intersección completa, todas sus componentes irreducibles tienen la misma dimensión $n - r$. Por lo tanto, la Desigualdad (4.2.10) implica que todas las componentes irreducibles de V están incluidas en $V_{\mathbb{A}}(f_1, \dots, f_m)$.

Por la misma razón, existe un subconjunto abierto no vacío $\mathbb{G}_1(V) \subseteq \mathbb{G}(n, n - r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para todo $A \in \mathbb{G}_1(V)$ y para cada componente irreducible W de V , tenemos que:

$$\sharp(W \cap A) = \deg(W).$$

Sea $\mathbb{G}(V) \subseteq \mathbb{G}(n, n - r)$ el subconjunto abierto no vacío descrito en el Corolario 3.2.5. Entonces, $\mathbb{G}_1(V) \cap \mathbb{G}(V)$ es también un subconjunto abierto respecto a su topología Zariski final. Finalmente, para cualquier entero positivo $L \in \mathbb{N}$, tal que $L \geq 6 \dim(\Omega)$, y cualquier $A \in \mathbb{G}_1(V) \cap \mathbb{G}(V)$, tenemos que se verifican las siguientes propiedades:

- Para cada $\mathbf{Q} \in (A \cap V)^L$, \mathbf{Q} no es un conjunto custer para Ω con respecto a Σ . Esto se debe a que f es idénticamente nula en \mathbf{Q} . En particular, con las mismas notaciones que antes, obtenemos:

$$\sharp(R(\Omega, \Sigma, V, L) \cap A^L) = 0.$$

- Por Corolario 3.2.5, dado que todas sus hipótesis se cumplen, también tenemos que:

$$\frac{\sharp(R(\Omega, \Sigma, V, L) \cap A^L)}{\deg(V)^L} \geq 1 - \frac{1}{\deg_{\text{lc}}(\Omega) e^{\dim(\Omega) + (m-1)L}} > 0.$$

Hemos llegado a una contradicción y, por lo tanto, no existe $f \in \Omega \setminus \Sigma$. □

En el caso $\Sigma = \{0\}$, la noción anterior coincide con la de variedades evasivas para hipersuperficies (ver Definición 27).

4.2.4. Los conjuntos cuestores están densamente distribuidos en variedades evasivas de dimensión positiva. En esta sección, demostraremos que los conjuntos cuestores están densamente distribuidos en cualquier variedad evasiva irreducible de dimensión positiva con respecto a cualquier distribución de probabilidad con “buen comportamiento”. Esto repara las imprecisiones sobre distribuciones de probabilidad discutidas en la Sección 3.2. A partir de ahora, podemos hablar de distribuciones de probabilidad con toda naturalidad, siempre que consideremos variedades evasivas irreducibles.

TEOREMA 4.2.9. *Sea $\Omega \subseteq P_d^K(X_1, \dots, X_n)$ un conjunto constructible de polinomios de grado a lo sumo d . Sea $V \subseteq \mathbb{A}^n(K)$ una variedad algebraica irreducible de dimensión positiva que es evasiva para hipersuperficies en Ω . Sea $\mathcal{B} \subseteq 2^V$ una σ -álgebra que contiene a los subconjuntos de Borel de V con respecto a la topología de Zariski. Sea $\mu : \mathcal{B} \rightarrow [0, 1]$ una distribución de probabilidad sobre \mathcal{B} que verifica la siguiente propiedad: para cada cerrado Zariski $A \subseteq V$, si $\dim(A) < \dim(V)$, entonces $\mu(A) = 0$. Sea $L \in \mathbb{N}$ un entero positivo, $\mathcal{B}^{\otimes L}$ la σ -álgebra en el producto V^L inducida por (V, \mathcal{B}) y $\mu^{\otimes L}$ la distribución de probabilidad definida sobre $\mathcal{B}^{\otimes L}$ por $\mu : \mathcal{B} \rightarrow [0, 1]$. Sea $R(\Omega, V, L)$ la clase de todas las listas $\mathbf{Q} \in V^L$ que son conjuntos cuestores para Ω con respecto a $\{0\}$. Supongamos que se satisface la siguiente desigualdad:*

$$(4.2.11) \quad 64 \left(1 + \frac{1 + \log(\deg_{\text{Ici}}(\Omega))}{\dim(\Omega)} + \log(L(d+1)) \right) < \frac{L}{\dim(\Omega)}$$

donde \log denota al logaritmo neperiano. Entonces, se verifica la siguiente desigualdad:

$$\text{Prob}_{x \in V^L} [x \in R(\Omega, V, L)] \geq 1 - \frac{1}{\deg_{\text{Ici}}(\Omega)e^{\dim(\Omega)}}.$$

DEMOSTRACIÓN. El clásico Teorema de Vapnik-Chervonenkis (ver Teorema 1.3.2) puede interpretarse como una estimación para métodos de Montecarlo en una variedad V dotada con cualquier distribución de probabilidad. Esta interpretación implica el siguiente resultado para el método de Montecarlo asociado a una familia \mathcal{F} de funciones características definidas en V . Para cada punto $x \in V^L$, denotamos por $x_i \in V$ a la i -ésima componente de $x := (x_1, \dots, x_L)$. Entonces, tenemos que:

$$(4.2.12) \quad \text{Prob}_{x \in V^L} \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{L} \sum_{i=1}^L f(x_i) - \int_{z \in V} f(z) d\mu(z) \right| > \varepsilon \right] \leq 4G(\mathcal{F}, L)e^{-\frac{L\varepsilon^2}{32}},$$

donde $\varepsilon > 0$ es un número real positivo y $G(\mathcal{F}, m)$ es la función de crecimiento asociada a la clase \mathcal{F} . Con las notaciones anteriores, supongamos que $\mathcal{F} := \mathcal{H}_d(V', \Omega)$, donde:

$$V' := \{(f, x) \in \mathbb{A}^N(K) \times \mathbb{A}^n(K) : f(x) \neq 0\}$$

y que V es una variedad irreducible evasiva para hipersuperficies en Ω . Por lo tanto, dado $f \in \Omega \setminus \{0\}$, tenemos que f no se anula idénticamente en V (ya que V es evasiva) y, por lo tanto, $V_{\mathbb{A}}(f) \cap V \subsetneq V$ es una subvariedad propia de dimensión menor. Debido a nuestras hipótesis sobre la distribución de probabilidad μ en V , tenemos que para todo $f \in \Omega \setminus \{0\}$, $\mu(V_{\mathbb{A}}(f) \cap V) = 0$. Por consiguiente, concluimos:

$$\forall f \in \Omega \setminus \{0\}, \int_{z \in V} \chi_{D(f)}(z) d\mu(z) = 1.$$

En el Teorema 4.2.3 anterior, hemos obtenido la siguiente cota superior para la función de crecimiento de la clase \mathcal{F} :

$$G(\mathcal{F}, L) \leq \deg_{\text{Ici}}(\Omega)(1 + L(d+1))^{\dim(\Omega)}.$$

Por lo tanto, la Ecuación 4.2.12 se convierte en:

$$(4.2.13) \quad \text{Prob}_{x \in V^L} \left[\sup_{f \in \mathcal{F}} \left| \frac{1}{L} \sum_{i=1}^L \chi_{D(f)}(x_i) - B(f) \right| > \varepsilon \right] \leq 4 \deg_{\text{Ici}}(\Omega)(1 + L(d+1))^{\dim(\Omega)} e^{-\frac{L\varepsilon^2}{32}},$$

donde:

$$B(f) := \begin{cases} 1, & \text{si } f \neq 0 \\ 0, & \text{en otro caso.} \end{cases}$$

Si f es la función nula, i.e. $f \equiv 0$, obviamente obtenemos:

$$\left| \frac{1}{L} \sum_{i=1}^L \chi_0(x_i) - B(f) \right| = 0.$$

Por lo tanto, también tenemos que:

$$(4.2.14) \quad \text{Prob}_{x \in V^L} \left[\sup_{f \in \mathcal{F} \setminus \{0\}} \left| \frac{1}{L} \sum_{i=1}^L \chi_{D(f)}(x_i) - 1 \right| > \varepsilon \right] \leq 4 \deg_{\text{Ici}}(\Omega)(1 + L(d+1))^{\dim(\Omega)} e^{-\frac{L\varepsilon^2}{32}}.$$

Para tener el mismo nivel de probabilidad de error que en el caso cero-dimensional (ver Corolario 3.2.6 con $m = 1$), supongamos que:

$$\varepsilon^2 = \frac{32}{L} (\log(4) + 2 \log(\deg_{\text{lci}}(\Omega)) + \dim(\Omega)(1 + \log(1 + L(d + 1)))).$$

Nótese que:

$$\frac{4 \log(2)}{L} \leq \varepsilon^2,$$

lo que significa que ε satisface las hipótesis del Teorema 1.3.2. Además, tenemos que:

$$\varepsilon^2 \leq \frac{64 \dim(\Omega)}{L} \left(1 + \frac{1 + \log(\deg_{\text{lci}}(\Omega))}{\dim(\Omega)} + \log(L(d + 1)) \right) < 1.$$

y, entonces se verifica $\varepsilon < 1$. De lo anterior, se deduce la siguiente igualdad:

$$(4.2.15) \quad 4 \deg_{\text{lci}}(\Omega)(1 + L(d + 1))^{\dim(\Omega)} e^{-\frac{L\varepsilon^2}{32}} = \frac{1}{\deg_{\text{lci}}(\Omega)e^{\dim(\Omega)}}.$$

Entonces, obtenemos:

$$\text{Prob}_{x \in V^L} [x \notin R(\Omega, V, L)] = \text{Prob}_{x \in V^L} \left[\sup_{f \in \mathcal{F} \setminus \{0\}} \left| \frac{1}{L} \sum_{i=1}^L \chi_{D(f)}(x_i) - 1 \right| = 1 \right].$$

Por lo tanto, también se cumple que:

$$\text{Prob}_{x \in V^L} [x \notin R(\Omega, V, L)] \leq \text{Prob}_{x \in V^L} \left[\sup_{f \in \mathcal{F} \setminus \{0\}} \left| \frac{1}{L} \sum_{i=1}^L \chi_{D(f)}(x_i) - 1 \right| > \varepsilon \right].$$

Aplicando la Identidad (4.2.15), obtenemos:

$$\text{Prob}_{x \in V^L} [x \notin R(\Omega, V, L)] \leq \frac{1}{\deg_{\text{lci}}(\Omega)e^{\dim(\Omega)}}.$$

Finalmente, concluimos:

$$\text{Prob}_{x \in V^L} [x \in R(\Omega, V, L)] \geq 1 - \frac{1}{\deg_{\text{lci}}(\Omega)e^{\dim(\Omega)}}.$$

□

Redes Neuronales con función de activación racional y conjuntos constructibles de parámetros

Índice

5.1. Sintaxis de una Red Neuronal	102
5.2. Semántica de una Red Neuronal: instanciación de parámetros	103
5.3. Ejemplos de funciones de activación: ReLUs, esquemas de evaluación, polinomiales, racionales, de Nash y de Pfaff	104
5.4. Propiedades del conjunto constructible $\mathcal{W}(\mathcal{N}, \Lambda)$ en el caso de funciones de activación polinómicas	106
5.5. Redes neuronales con función de activación racional que evalúan funciones racionales	108

La noción de *red neuronal* es omnipresente en el Aprendizaje Computacional actual. En numerosos trabajos dentro de este campo, los clasificadores y las funciones a aprender se representan mediante redes neuronales. Sin embargo, resulta difícil encontrar en la literatura una definición matemática precisa del concepto de red neuronal. Esto se debe a que los investigadores suelen emplear nociones de red neuronal adaptadas a los contextos específicos de sus estudios. Por esta razón, hemos optado por introducir una noción matemática general de red neuronal, aunque solo buscamos un estudio sistemático de un tipo particular dependiente del uso de funciones de activación definidas por funciones polinomiales o racionales.

Comenzamos el capítulo describiendo la sintaxis de las redes neuronales. Formalmente, una red neuronal es un grafo dotado de ciertas estructuras subyacentes sobre las que se evalúan funciones pertenecientes a una clase \mathcal{A} . Desde el punto de vista semántico, estas funciones, conocidas habitualmente como *funciones de activación*, operan sobre combinaciones lineales del conjunto (o abanico) de entrada de cada nodo. La elección de las funciones de activación desempeña un papel crucial en la determinación del conjunto de funciones que la red neuronal es capaz de evaluar.

Una vez establecida la sintaxis abstracta en la Sección 5.1, pasamos a describir la semántica de una red neuronal en la Sección 5.2. Fijada una sintaxis y una clase de funciones de activación \mathcal{A} , podemos instanciar los parámetros de la red neuronal. En particular, si \mathcal{N} es una red neuronal y Λ una clase de parámetros admisibles, denotamos por $\mathcal{W}(\mathcal{N}, \Lambda)$ a la clase de todas las funciones que \mathcal{N} puede evaluar utilizando parámetros en Λ . Gran parte de la literatura en Teoría del Aprendizaje Computacional se centra en el estudio del aprendizaje de la clase $\mathcal{W}(\mathcal{N}, \Lambda)$ y en su capacidad para aproximar funciones que se desea aprender.

Posteriormente, en la Subsección 5.3, presentamos varios ejemplos de redes neuronales habituales en la literatura científica. En particular, hemos incluido el Ejemplo 5.3.2, que se centra en las redes neuronales con función de activación $\varphi(t) = t^2$. Cabe destacar que este tipo de redes neuronales coincide con lo que los autores clásicos de la Teoría de la Complejidad Algebraica denominaban *straight-line programs* (ver [Os, 54], [Mo, 55], [He, 89], [St, 73] y otras referencias clásicas). Por lo tanto, no se trata de un concepto nuevo. Un estudio exhaustivo de las redes neuronales con función de activación $\varphi(t) = t^2$ puede encontrarse en la Sección 3 de [KP, 96]. Estos autores no se referían a tales estructuras como redes neuronales, sino como *esquemas de evaluación*. Su trabajo presenta una colección de resultados que explican en detalle diversas estrategias para trabajar con estos esquemas de evaluación (o, equivalentemente, redes neuronales con función de activación $\varphi(t) = t^2$), utilizando una sintaxis que tiene en cuenta no solo el tamaño del grafo subyacente, sino también su profundidad, entendida como una medida de la interacción entre las distintas capas de la red.

A simple vista, todo el trabajo de este capítulo puede parecer puramente abstracto. Sin embargo, las redes neuronales están involucradas en la práctica totalidad de los problemas de la Geometría Algebraica Computacional (llamada, desde el siglo XIX, Teoría de la Eliminación). En [GH, 91] y [KP, 96] se demuestra que *todos los polinomios de eliminación pueden representarse con una red neuronal con función de activación $\varphi(t) = t^2$* , de tamaño “admisibles” y baja profundidad. En concreto, si definimos un polinomio de eliminación como un polinomio $\chi_g \in K[U_1, \dots, U_m]$ que surge de la proyección de una intersección $V(f_1, \dots, f_s, g)$, donde

$f_1, \dots, f_s, g \in [X_1, \dots, X_n, U_1, \dots, U_m]$ son polinomios de grado a lo sumo d , entonces χ_g puede representarse mediante una red neuronal de talla $d^{O(n+m)}$ y profundidad $O(n+m(\log_2(d)))$, utilizando $\varphi(t) = t^2$ como única función de activación y usando únicamente los parámetros necesarios para codificar $\{f_1, \dots, f_s, g\}$. Mejoras de los resultados presentados en [GH, 91] y [KP, 96] pueden encontrarse en la serie de artículos del algoritmo TERA: [Pa, 95], [GHMP, 95], [GHMP, 97], [GHHMMP, 97], [GHMMP, 98] o [HMPS, 00]. Todos estos estudios dependen en gran medida del hecho de que todos los polinomios que aparecen en Geometría Algebraica Computacional admiten representaciones razonables como *esquemas de evaluación* (o, equivalentemente, redes neuronales con función de activación $\varphi(t) = t^2$).

En resumen, las redes neuronales están intrínsecamente involucradas en la Teoría de la Eliminación como codificaciones de los polinomios de eliminación. Esto establece una conexión “imprecisa” entre la Geometría Algebraica Computacional y la Teoría Computacional del Aprendizaje, cuyas consecuencias son aún desconocidas.

5.1. Sintaxis de una Red Neuronal

Comenzamos fijando las notaciones para redes neuronales:

DEFINICIÓN 28 (Red Neuronal). *Sea K un cuerpo. Una red neuronal \mathcal{N} sobre el cuerpo K , con variables de entrada en $\{X_1, \dots, X_n\}$, es una terna $\mathcal{N} := (G, \mathcal{A}, \Phi)$, donde:*

- i) $G := (V, E)$ es un grafo dirigido acíclico, cuyos vértices (que representan nodos o neuronas) están dados por:

$$V := \{(i, j) : 0 \leq i \leq \ell, 1 \leq j \leq L_i\}.$$

Usamos ℓ para denotar la profundidad de la red neuronal. En el par $\nu := (i, j)$, la primera coordenada i , $0 \leq i \leq \ell$, se denomina profundidad del nodo (denotada por $d(\nu)$), mientras que la segunda coordenada j , $1 \leq j \leq L_i$, hace referencia a una enumeración del conjunto $\mathcal{L}_i := \{(i, j) : 1 \leq j \leq L_i\}$, que contiene a todos los vértices de profundidad i , donde $\sharp(\mathcal{L}_i) = L_i$. Además, supondremos que el conjunto de aristas E verifica:

$$E \subseteq \{(\nu_1, \nu_2) \in V^2 : d(\nu_1) \leq d(\nu_2) - 1\}.$$

En otras palabras, el abanico de entrada de los nodos de profundidad $i \geq 1$ está formado por nodos de profundidad a lo sumo $i - 1$ ¹. Denotamos por $\text{Fan} - \text{In}(\nu)$ al abanico de entrada del nodo $\nu \in V$ en el grafo G . El grafo tiene L_ℓ nodos con abanico de salida vacío, los cuales reciben el nombre de nodos o neuronas de salida. El conjunto de los nodos de salida se denota por $O_{\mathcal{N}}$, donde $\sharp(O_{\mathcal{N}}) = L_\ell$.

- ii) \mathcal{A} es una clase de funciones univariadas parcialmente definidas $f : \text{Dom}(f) \subseteq K \rightarrow K$, y se denomina clase de funciones de activación.
- iii) El último elemento de la terna, Φ , es una asignación de etiquetas a cada nodo del grafo de acuerdo con las siguientes reglas. En primer lugar, existe un conjunto de variables, llamadas parámetros de la red neuronal (también conocidos como “pesos” de la red neuronal), asociado a cada arista de la red neuronal. Los denotamos de la siguiente forma:

$$P_{\mathcal{N}} := \{A_\nu^\mu : \nu \in V, \mu \in \text{Fan} - \text{In}(\nu)\}.$$

Denotamos por $\mathcal{P}(P_{\mathcal{N}})$ a la clase de subconjuntos de este espacio de parámetros. La asignación Φ es una aplicación:

$$\Phi : V \rightarrow (\mathcal{A} \times \mathcal{P}(P_{\mathcal{N}})) \cup \{1, X_1, \dots, X_n\},$$

definida mediante la siguiente identidad:

$$(5.1.1) \quad \Phi(\nu) := \begin{cases} \left(f_\nu, \left\{ A_\nu^{(i,j)} : (i,j) \in \text{Fan} - \text{In}(\nu) \right\} \right) \in \mathcal{A} \times \mathcal{P}(P_{\mathcal{N}}), & \text{si } d(\nu) \geq 1 \\ 1, & \text{si } \nu = (0, 0) \\ X_j, & \text{si } \nu = (0, j), 1 \leq j \leq n \end{cases}.$$

Es decir, esta aplicación asocia a cada nodo ν , con $d(\nu) \geq 1$, un par formado por una función de activación $f_\nu : \text{Dom}(f_\nu) \subseteq K \rightarrow K \in \mathcal{A}$, denominada función de activación del nodo ν , y la lista de parámetros asociada a su abanico de entrada $\text{Fan} - \text{In}(\nu)$. Los nodos de profundidad 0 están asociados a las variables en $\{X_1, \dots, X_n\}$ y a la constante 1.

¹Esta restricción elemental puede reemplazarse por la siguiente condición: “el abanico de entrada de los nodos de profundidad i está formado únicamente por nodos de profundidad $i - 1$ ”. Hemos decidido que el abanico de entrada esté formado por nodos de profundidad a lo sumo $i - 1$ en lugar de restringirlo a nodos de profundidad $i - 1$ para simplificar las demostraciones. Sin embargo, es evidente que ambas condiciones son equivalentes.

El número de nodos que no son de entrada, $L(\mathcal{N}) := \sum_{i=1}^{\ell} L_i$, se denomina talla de la red neuronal. Finalmente, el máximo de las cardinalidades de los abanicos de entrada de cualquier nodo de la red neuronal recibe el nombre de requerimientos de espacio de la red neuronal:

$$S(\mathcal{N}) := \max\{\#\text{Fan} - \text{In}(\nu) : \nu \in V\}.$$

5.2. Semántica de una Red Neuronal: instanciación de parámetros

Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal definida como en la sección precedente. Consideramos, en primer lugar, el K -espacio vectorial $K_{\circ}^{(n)}[\mathcal{A}]$ formado por todas las funciones parciales $f : \text{Dom}(f) \subseteq \mathbb{A}^n(K) \rightarrow K$ generadas a partir de la función constante, las proyecciones (representadas por las variables $\{X_1, \dots, X_n\}$) y las funciones de activación de \mathcal{A} , y que es estable por combinaciones lineales y composición. Decimos que $K_{\circ}^{(n)}[\mathcal{A}]$ es la clase de funciones parciales con variables de entrada $\{X_1, \dots, X_n\}$ potencialmente evaluables por una red neuronal que emplea funciones de la clase \mathcal{A} como funciones de activación.

Una red neuronal $\mathcal{N} = (G, \mathcal{A}, \Phi)$ es una representación sintáctica de una clase de funciones parciales $\mathcal{W}(\mathcal{N}) \subseteq K_{\circ}^{(n)}[\mathcal{A}]$ que pueden ser evaluadas por \mathcal{N} mediante *instanciación de parámetros*. Este proceso semántico se define de la siguiente manera:

DEFINICIÓN 29. Sea $\mathcal{N} = (G, \mathcal{A}, \Phi)$ una red neuronal. Instanciar \mathcal{N} consisten en fijar un vector de parámetros:

$$\underline{a} := (a_{\nu}^{\mu} : \nu \in V, \mu \in \text{Fan} - \text{In}(\nu)) \in K^N,$$

donde N es el número total de aristas en el grafo G . Una red neuronal \mathcal{N} instanciada en \underline{a} es una red neuronal $\mathcal{N}_{\underline{a}} := (G, \mathcal{A}, \Phi_{\underline{a}})$, donde G y \mathcal{A} son los mismos que en \mathcal{N} y $\Phi_{\underline{a}}$ es una asignación a cada nodo determinada por Φ y \underline{a} de la siguiente manera:

$$\Phi_{\underline{a}}(\nu) := \begin{cases} (f_{\nu}, (a_{\nu}^{\mu} : \mu \in \text{Fan} - \text{In}(\nu))) \in \mathcal{A} \times K^{\#\text{Fan} - \text{In}(\nu)}, & \text{si } d(\nu) \geq 1 \\ 1, & \text{si } \nu = (0, 0) \\ X_j, & \text{si } \nu = (0, j), 1 \leq j \leq n \end{cases}.$$

DEFINICIÓN 30. Dada una red neuronal $\mathcal{N} = (G, \mathcal{A}, \Phi)$ con N aristas y nodos de entrada asociados a las variables en $\{X_1, \dots, X_n\}$, una lista de parámetros $\underline{a} := (a_{\nu}^{\mu} : \nu \in V, \mu \in \text{Fan} - \text{In}(\nu)) \in K^N$, y la instanciación $\mathcal{N}_{\underline{a}}$, podemos asociar a cada nodo ν del grafo subyacente de \mathcal{N} una función instanciada $f_{\nu}^{\mathcal{N}}(\underline{a}, \cdot) \in K_{\circ}^{(n)}[\mathcal{A}]$ definida recursivamente en términos de la profundidad de la siguiente forma:

- Para los nodos de entrada, definimos las siguientes funciones para cada $x := (x_1, \dots, x_n) \in K^n$:

$$f_{(0,0)}^{\mathcal{N}}(\underline{a}, x) := 1, \quad f_{(0,j)}^{\mathcal{N}}(\underline{a}, x) = x_j, \quad 1 \leq j \leq n.$$

- Para los nodos ν de profundidad $d(\nu) \geq 1$, dado $x := (x_1, \dots, x_n) \in K^n$ en el dominio de definición de todas las funciones en la clase $\{f_{\mu}(\underline{a}, \cdot) : \mu \in \text{Fan} - \text{In}(\nu)\}$, definimos:

$$f_{\nu}^{\mathcal{N}}(\underline{a}, x) := f_{\nu} \left(\sum_{\mu \in \text{Fan} - \text{In}(\nu)} a_{\nu}^{\mu} f_{\mu}^{\mathcal{N}}(\underline{a}, x) \right),$$

dado que:

$$\sum_{\mu \in \text{Fan} - \text{In}(\nu)} a_{\nu}^{\mu} f_{\mu}^{\mathcal{N}}(\underline{a}, x) \in \text{Dom}(f_{\nu}^{\mathcal{N}}(\underline{a}, \cdot)) = \bigcap_{\mu \in \text{Fan} - \text{In}(\nu)} \text{Dom}(f_{\mu}^{\mathcal{N}}(\underline{a}, \cdot)).$$

En ocasiones, también decimos que la red neuronal \mathcal{N} , instanciada en \underline{a} , es un “dispositivo” que representa todas las funciones parciales asociadas a cada uno de sus nodos:

$$\{f_{\nu}^{\mathcal{N}}(\underline{a}, \cdot) : \text{Dom}(f_{\nu}^{\mathcal{N}}(\underline{a}, \cdot)) \subseteq \mathbb{A}^n(K) \rightarrow K : \nu \in V\} \subseteq K_{\circ}^{(n)}[\mathcal{A}].$$

DEFINICIÓN 31. Sea \mathcal{N} una red neuronal como en la definición anterior y sea $m := \#\text{Out}(\mathcal{N}) = L_{\ell}$ el número de nodos de salida de la red neuronal \mathcal{N} . Sea \underline{a} una instanciación de los parámetros. Decimos que $\mathcal{N}_{\underline{a}}$ evalúa la función parcial:

$$(5.2.1) \quad f_{\mathcal{N}}(\underline{a}, \cdot) : \underset{x}{\text{Dom}(f_{\mathcal{N}}(\underline{a}, \cdot)) \subseteq \mathbb{A}^n(K)} \longrightarrow \underset{\nu \in O_{\mathcal{N}}}{K^m} \quad \longmapsto \quad (f_{\nu}^{\mathcal{N}}(\underline{a}, x_1, \dots, x_n) : \nu \in O_{\mathcal{N}}),$$

donde $\text{Dom}(f_{\mathcal{N}}(\underline{a}, \cdot)) \subseteq \mathbb{A}^n(K)$ es el dominio (posiblemente vacío) en el que la función $f_{\mathcal{N}}(\underline{a}, \cdot)$ está definida, el cual depende tanto de la red neuronal \mathcal{N} como de la instancia de parámetros $\underline{a} \in K^N$.

Finalmente, definimos la clase de funciones evaluables por una red neuronal mediante instanciación de los parámetros de la siguiente manera:

DEFINICIÓN 32. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con N aristas, nodos de entrada asociados a las variables en $\{X_1, \dots, X_n\}$ y m nodos de salida en $O_{\mathcal{N}}$. Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un subconjunto. Definimos la clase de funciones (parciales) definidas por \mathcal{N} con parámetros instanciados en Λ como el siguiente subconjunto de $(K_{\circ}^{(n)}[\mathcal{A}])^m$:

$$\mathcal{W}(\mathcal{N}, \Lambda) := \{f_{\mathcal{N}}(\underline{a}, \cdot) \in (K_{\circ}^{(n)}[\mathcal{A}])^m : \underline{a} \in \Lambda\}.$$

5.3. Ejemplos de funciones de activación: ReLUs, esquemas de evaluación, polinomiales, racionales, de Nash y de Pfaff

EJEMPLO 5.3.1 (**Redes neuronales ReLU y funciones racionales algebraicas tropicales**). Es la clase más simple de redes neuronales, pero su estudio está fuera del alcance de esta memoria. El cuerpo base $K = \mathbb{R}$ es el cuerpo de los números reales. La clase \mathcal{A} de funciones de activación está formada por una única función, a veces denotada por $(\cdot)_+$ (cf. [PMRML, 17] y sus referencias):

$$\begin{aligned} (\cdot)_+ : \mathbb{R} &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto (x)_+ := \max\{x, 0\} \end{aligned}$$

La clase $\mathbb{R}_{\circ}^{(n)}[\mathcal{A}]$ de funciones potencialmente evaluables por una red neuronal ReLU con n nodos de entrada $\{X_1, \dots, X_n\}$ y un solo nodo de salida es una extensión de la clase de *funciones racionales algebraicas tropicales* (cf. [Mi, 06], [MS, 15], [Gr, 18] o [Gr, 20] y sus referencias para más detalles) que admite exponentes reales. Recordemos que un polinomio tropical sobre \mathbb{R} es una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ para la cual existen $a_1, \dots, a_m \in \mathbb{Z}^n$ y $b_1, \dots, b_m \in \mathbb{R}$ tales que:

$$f(x) := \min\{\langle a_1, x \rangle + b_1, \dots, \langle a_m, x \rangle + b_m\},$$

donde $\langle a, x \rangle := \sum_{i=1}^n a_i x_i$ para cada $a = (a_1, \dots, a_n) \in \mathbb{Z}^n$ y $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Una *función racional algebraica tropical* es la diferencia entre dos polinomios tropicales (cf. [Gr, 20] y sus referencias). También puede definirse como una función continua definida a trozos con pendientes enteras. Cambiando:

$$\min\{\langle a_1, x \rangle + b_1, \dots, \langle a_m, x \rangle + b_m\} = -\max\{-\langle a_1, x \rangle - b_1, \dots, -\langle a_m, x \rangle - b_m\},$$

se establece la relación entre ambas teorías. Dada una red neuronal ReLU $\mathcal{N} := (G, \mathcal{A}, \Phi)$ con un solo nodo de salida, la clase $\mathbb{R}_{\circ}^{(n)}[\mathcal{A}]$ de funciones potencialmente evaluables por \mathcal{N} se define de la siguiente manera (ver Proposición 5.6 de [ZNL, 18] para más detalles): una función parcial $f : \text{Dom}(f) \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$ verifica que $f \in \mathbb{R}_{\circ}^{(n)}[\mathcal{A}]$ si y solo si existen $a_1, \dots, a_m, c_1, \dots, c_r \in \mathbb{R}^n$ y $b_1, \dots, b_m, d_1, \dots, d_r \in \mathbb{R}$ tales que:

$$f(x) := \min\{\langle a_1, x \rangle + b_1, \dots, \langle a_m, x \rangle + b_m\} - \min\{\langle c_1, x \rangle + d_1, \dots, \langle c_r, x \rangle + d_r\}.$$

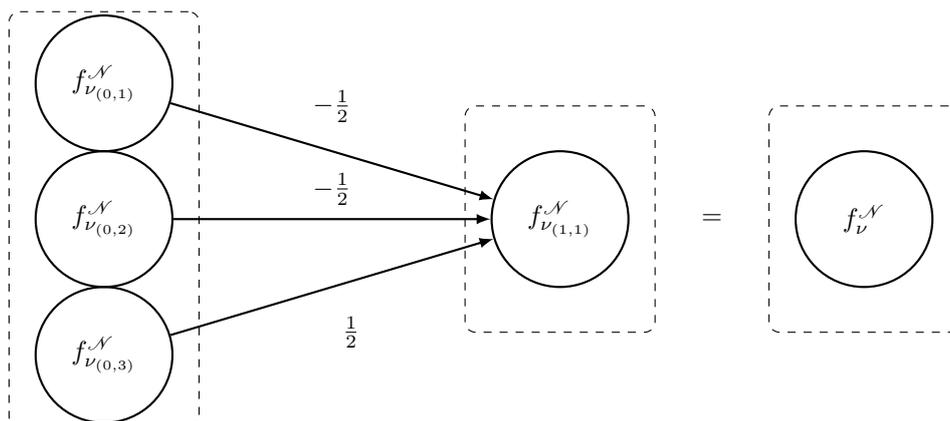
EJEMPLO 5.3.2 (**Esquemas de evaluación o redes neuronales con función de activación $\varphi(t) = t^2$**). Consideramos una red neuronal $\mathcal{N} := (G, \mathcal{A}, \Phi)$ sobre un cuerpo K y con variables de entrada en $\{X_1, \dots, X_n\}$. La clase de funciones de activación \mathcal{A} se reduce a una sola función $\varphi : K \rightarrow K$, dada por $\varphi(t) = t^2$. Esta clase de redes neuronales es exactamente la misma que la clase de programas descrita en las Secciones 2 y 3 de [KP, 96] como *esquemas de evaluación*. Con el objetivo de ser completos, hacemos explícita la equivalencia entre los esquemas de evaluación y las redes neuronales con función de activación $\varphi(t) = t^2$. Usando las notaciones previas, la función evaluada en un nodo ν (que no es de entrada) en un esquema de evaluación (según lo definido en [KP, 96]) viene dada por:

$$f_{\nu}(\underline{a}, x) = \left(\sum_{\mu \in \text{Fan-In}(\nu)} a_{\nu}^{\mu,1} f_{\mu}(\underline{a}, x) \right) \cdot \left(\sum_{\mu \in \text{Fan-In}(\nu)} a_{\nu}^{\mu,2} f_{\mu}(\underline{a}, x) \right).$$

Nótese que la siguiente igualdad relaciona el producto y la operación de elevar al cuadrado:

$$X \cdot Y = \frac{1}{2} ((X + Y)^2 - X^2 - Y^2).$$

Por lo tanto, si la característica de K es distinta de 2, la función evaluada en el nodo ν es equivalente a la siguiente red neuronal de 4 nodos y profundidad 2:



donde:

$$f_{\nu(0,1)}^{\mathcal{N}}(\underline{a}, x) := \left(\sum_{\mu \in \text{Fan-In}(\nu)} a_{\nu}^{\mu,1} f_{\mu}^{\mathcal{N}}(\underline{a}, x) \right)^2,$$

$$f_{\nu(0,2)}^{\mathcal{N}}(\underline{a}, x) := \left(\sum_{\mu \in \text{Fan-In}(\nu)} a_{\nu}^{\mu,2} f_{\mu}^{\mathcal{N}}(\underline{a}, x) \right)^2,$$

$$f_{\nu(0,3)}^{\mathcal{N}}(\underline{a}, x) := \left(\sum_{\mu \in \text{Fan-In}(\nu)} (a_{\nu}^{\mu,1} + a_{\nu}^{\mu,2}) f_{\mu}^{\mathcal{N}}(\underline{a}, x) \right)^2,$$

y

$$f_{\nu(1,1)}^{\mathcal{N}}(\underline{a}, x) := \frac{1}{2} f_{\nu(0,3)}^{\mathcal{N}}(\underline{a}, x) - \frac{1}{2} f_{\nu(0,1)}^{\mathcal{N}}(\underline{a}, x) - \frac{1}{2} f_{\nu(0,2)}^{\mathcal{N}}(\underline{a}, x).$$

Por lo tanto, reemplazado cada nodo del esquema de evaluación por la red neuronal que acabamos de describir, obtenemos una red neuronal con función de activación $\varphi(t) = t^2$ que es equivalente al esquema de evaluación. Nótese que también se verifica el recíproco. Dada una red neuronal con función de activación $\varphi(t) = t^2$, podemos representarla como un esquema de evaluación utilizando los mismos parámetros para ambos términos del producto en cada nodo.

En este contexto, la clase de funciones potencialmente evaluables $K_{\circ}^{(n)}[\mathcal{A}] = K[X_1, \dots, X_n]$ es la clase de todos los polinomios en n variables con coeficientes en K . Se puede demostrar fácilmente que, si \mathcal{N} tiene profundidad a lo sumo ℓ , las funciones evaluables por \mathcal{N} son polinomios de grado a lo sumo 2^ℓ . Si fijamos un conjunto de parámetros $\Lambda \subseteq K^N$, donde N es el número de aristas de \mathcal{N} , y si \mathcal{N} tiene m nodos de salida, la clase de todos los polinomios evaluables por \mathcal{N} con parámetros en Λ es un subconjunto constructible:

$$\mathcal{W}(\mathcal{N}, \Lambda) \subseteq \mathcal{P}_{(2^\ell)}^K(X_1, \dots, X_n),$$

donde $(2^\ell) = (2^\ell, \dots, 2^\ell) \in \mathbb{N}^m$ es una lista de grados. Nótese que esta clase de polinomios ya ha sido empleada en la Sección 3.4.3.

EJEMPLO 5.3.3 (Redes neuronales con función de activación polinómica). En lugar de utilizar $\varphi(t) = t^2$ como única función de activación, también consideramos redes neuronales cuya función de activación $\varphi = p \in K[T]$ es un polinomio. Discutiremos algunas propiedades de estas redes neuronales en la Sección 5.4.

EJEMPLO 5.3.4 (Redes neuronales con función de activación racional). Al igual que en los ejemplos previos, consideramos redes neuronales $\mathcal{N} := (G, \mathcal{A}, \Phi)$ sobre un cuerpo K y con variables de entrada en $\{X_1, \dots, X_n\}$ tales que la clase de funciones de activación está formada, de nuevo, por una única función $\mathcal{A} := \{\varphi\}$, donde $\varphi : \text{Dom}(\varphi) \subseteq K \rightarrow K$, es una función racional. Es decir, existen dos polinomios univariados $p, q \in K[T]$, tales que $\text{Dom}(\varphi) = \{t \in K : q(t) \neq 0\}$ y

$$\varphi(t) = \frac{p(t)}{q(t)}, \quad \forall t \in \text{Dom}(\varphi).$$

Definimos el *grado de la función racional* φ de la siguiente forma:

$$\text{deg}(\varphi) := \max\{\text{deg}(p), \text{deg}(q)\},$$

donde p y q son los polinomios anteriores. Esta clase extiende a las clases de redes neuronales definidas en los Ejemplos 5.3.2 y 5.3.3 ya que estamos admitiendo cualquier función racional $\varphi \in K(T)$ como función de

activación. En este caso, la clase de funciones potencialmente evaluables por \mathcal{N} con un número finito de funciones racionales en \mathcal{A} satisface $K_{\circ}^{(n)}[\mathcal{A}] \subseteq K(X_1, \dots, X_n)$. A continuación, distinguimos dos casos:

- i) Redes neuronales con función de activación que admiten el “Vermeidung von Divisionen” de Strassen (cf. [St, 73]). Este es el ejemplo discutido en detalle en la Sección 3 de [KP, 96], y utilizado en trabajos posteriores del proyecto TERA (como en [GHMP, 95], [GHMP, 97], [GHHMMP, 97] o [GHMMP, 98]). Este caso no será el tema principal de nuestra discusión en esta sección.
- ii) Redes neuronales con función de activación racional que evalúan funciones racionales que no son polinomios. Estudiaremos en detalle este caso en la Subsección 5.5. Nótese que este tipo de redes neuronales son las que se utilizan para modelizar el operador de Newton-Hensel en el anillo de series de potencias formales en la serie de trabajos de la corriente TERA que va de [Pa, 95] a [GHMMP, 98].

EJEMPLO 5.3.5 (Redes Neuronales con funciones de activación de Nash o de Pfaff). En algunos casos, se consideran funciones más sofisticadas que los polinomios o las funciones racionales. Esto ocurre, por ejemplo, cuando la función de activación es una función de Nash (función algebraica analítica) o una función de Pfaff (como se introducen en [Kh, 91]). Remitimos al lector a la amplia literatura sobre este tema, ya que en esta memoria no trataremos estos casos: [CKKLW, 95], [Ko, 99], [GJ, 95], [KM, 97], [MPC, 08], [MP, 09] y sus referencias.

5.4. Propiedades del conjunto constructible $\mathcal{W}(\mathcal{N}, \Lambda)$ en el caso de funciones de activación polinómicas

Consideramos el caso de redes neuronales con función de activación polinómica. Las siguientes estimaciones son útiles para estudiar la función de crecimiento de la clase de clasificadores binarios asociada, así como la longitud y la densidad de conjuntos cuestores para $\mathcal{W}(\mathcal{N}, \Lambda)$.

PROPOSICIÓN 5.4.1. *Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asociados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y m nodos de salida. Supongamos que \mathcal{N} tiene N aristas y que $\mathcal{A} = \{p\}$, donde p es un polinomio univariado de grado $d \geq 2$. Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de los parámetros de \mathcal{N} . Sea $V(G)$ el conjunto de nodos del grafo G . Sea $\{A_{\nu}^{\mu} : \nu \in V(G), \mu \in \text{Fan} - \text{In}(\nu)\}$ el conjunto de las variables de los parámetros de la red neuronal y sea:*

$$K[A_{\nu}^{\mu} : \nu \in V(G), \mu \in \text{Fan} - \text{In}(\nu)],$$

el anillo de polinomios con coeficientes en K en este conjunto de variables. Entonces, tenemos que:

- i) *Para cada nodo $\sigma \in V(G)$ de la red neuronal de profundidad i y para cada multi-índice $\theta = (\theta_1, \dots, \theta_n) \in \mathbb{N}^n$ de grado total $|\theta| = \theta_1 + \dots + \theta_n \leq d^i$, existe un polinomio:*

$$Q_{\sigma}^{(\theta)} \in K[A_{\nu}^{\mu} : \nu \in V(G), \mu \in \text{Fan} - \text{In}(\nu)],$$

de grado total a lo sumo $d^{i+1} - 2$ tal que se verifica la siguiente igualdad para todo $\underline{a} \in \Lambda$ y para todo $x := (x_1, \dots, x_n) \in \mathbb{A}^n(K)$:

$$f_{\sigma}^{\mathcal{N}}(\underline{a}, x) = \sum_{\theta \in \mathbb{N}^n, |\theta| \leq d^i} Q_{\nu}^{(\theta)}(\underline{a}) x_1^{\theta_1} \dots x_n^{\theta_n} \in K.$$

- ii) *El conjunto $\mathcal{W}(\mathcal{N}, \Lambda) \subseteq \mathcal{P}_{(d^{\ell})}^K$, donde $(d^{\ell}) = (d^{\ell}, \dots, d^{\ell}) \in \mathbb{N}^m$, es un conjunto constructible que satisface las siguientes propiedades:*

- (a) *La dimensión de Krull de $\mathcal{W}(\mathcal{N}, \Lambda)$ está acotada por la siguiente desigualdad:*

$$\dim(\mathcal{W}(\mathcal{N}, \Lambda)) \leq \dim(\Lambda) \leq N \leq LS.$$

- (b) *El grado de la clausura Zariski de $\mathcal{W}(\mathcal{N}, \Lambda)$ verifica la siguiente desigualdad:*

$$\deg_z(\mathcal{W}(\mathcal{N}, \Lambda)) \leq \deg_{\text{ici}}(\Lambda) (d^{\ell+1} - 2)^{\dim(\Lambda)} \leq \deg_{\text{ici}}(\Lambda) (d^{\ell+1} - 2)^{LS}.$$

DEMOSTRACIÓN. El Ítem *i*) se demuestra fácilmente usando los mismos argumentos inductivos en la profundidad empleados en la prueba del Lema 14 de [KP, 96]. En cuanto al segundo ítem, sea $\{\nu_1, \dots, \nu_m\}$ la lista de nodos de salida de \mathcal{N} . Nótese que las funciones $f_{\nu_k}^{\mathcal{N}}(\underline{a}, \cdot)$ evaluadas por \mathcal{N} son polinomios de grado a lo sumo d^{ℓ} . Por otro lado, los polinomios $Q_{\nu_k}^{(\theta)}$ definen una aplicación desde Λ que parametriza $\mathcal{W}(\mathcal{N}, \Lambda)$:

$$\begin{aligned} Q : \Lambda &\longrightarrow \mathcal{W}(\mathcal{N}, \Lambda) \subseteq \mathcal{P}_{(d^{\ell})}^K[X_1, \dots, X_n] \\ \underline{a} &\longmapsto (f_{\nu_1}^{\mathcal{N}}(\underline{a}, X_1, \dots, X_n), \dots, f_{\nu_m}^{\mathcal{N}}(\underline{a}, X_1, \dots, X_n))' \end{aligned}$$

donde para cada k , $1 \leq k \leq m$, tenemos que:

$$f_{\nu_k}^{\mathcal{N}}(\underline{a}, X_1, \dots, X_n) := \sum_{\theta \in \mathbb{N}^n, |\theta| \leq d^\ell} Q_{\nu_k}^{(\theta)}(\underline{a}) X_1^{\theta_1} \dots X_n^{\theta_n} \in K[X_1, \dots, X_n].$$

Por lo tanto, hemos demostrado que $\mathcal{W}(\mathcal{N}, \Lambda)$ es un subconjunto constructible dado como la imagen de un subconjunto constructible $\Lambda \subseteq \mathbb{A}^N(K)$ por una lista de polinomios de grado a lo sumo $d^{\ell+1} - 2$. Obviamente, la dimensión de $\mathcal{W}(\mathcal{N}, \Lambda)$ está acotada por la dimensión de Λ y, aplicando el Corolario 2.2.4, concluimos:

$$\deg_z(\mathcal{W}(\mathcal{N}, \Lambda)) \leq \deg_{\text{ici}}(\Lambda) (d^{\ell+1} - 2)^{\dim(\Lambda)} \leq \deg_{\text{ici}}(\Lambda) (d^{\ell+1} - 2)^N \leq \deg_{\text{ici}}(\Lambda) (d^{\ell+1} - 2)^{LS},$$

como queríamos demostrar. \square

OBSERVACIÓN 5.4.2. Nótese que si Λ es una abierto Zariski, la cota superior del Ítem *ii*) (b) se convierte en:

$$\deg_z(\mathcal{W}(\mathcal{N}, \Lambda)) \leq (d^{\ell+1} - 2)^{LS}.$$

COROLARIO 5.4.3. *Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y un único nodo de salida. Supongamos que $\mathcal{A} = \{p\}$, donde p es un polinomio univariado de grado d . Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de parámetros de \mathcal{N} y $\mathcal{W}(\mathcal{N}, \Lambda)$ la clase de todos los polinomios evaluables por \mathcal{N} con parámetros en Λ . Sea \mathcal{H} la clase de clasificadores dada por las funciones características de subconjuntos abiertos distinguidos de $\mathbb{A}^n(K)$ definidos por polinomios en $\mathcal{W}(\mathcal{N}, \Lambda)$:*

$$\mathcal{H} = \{\chi_{D(f)} : f \in \mathcal{W}(\mathcal{N}, \Lambda)\}.$$

Entonces,

$$G(\mathcal{H}, m) \leq \deg_{\text{ici}}(\Lambda) ((d^{\ell+1} - 2) \cdot (1 + m(d^\ell + 1)))^{LS}.$$

DEMOSTRACIÓN. Para obtener el resultado, basta notar que la función de crecimiento es monótona con respecto a la inclusión de familias de clasificadores y combinar el Teorema 4.2.3 con las estimaciones obtenidas en la Proposición 5.4.1. \square

COROLARIO 5.4.4. *Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y m nodos de salida. Supongamos que $\mathcal{A} = \{p\}$, donde p es un polinomio univariado de grado d . Sea $\Lambda \subseteq \mathbb{A}^N(K)$ el conjunto constructible en el espacio afín de los parámetros de \mathcal{N} y $\Omega = \mathcal{W}(\mathcal{N}, \Lambda)$ la clase de todas las listas de polinomios evaluables por \mathcal{N} con parámetros en Λ . Sea $s := \dim(\Lambda)$ la dimensión de Λ y $t := s / \dim(\Omega)$ el cociente entre las dimensiones de Λ y Ω . Sea $\Sigma \subseteq \Omega$ un subconjunto constructible de co-dimensión al menos 1 en Ω y supongamos que se satisface la siguiente propiedad:*

$$\forall f := (f_1, \dots, f_m) \in \Omega \setminus \Sigma, \dim(V_{\mathbb{A}}(f_1, \dots, f_m)) = n - m.$$

Sea $V := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ una variedad algebraica de intersección completa de co-dimensión $r \geq (n - m) + m/2 + 1/2$ tal que $\deg(V) \geq \delta^r$, donde $\delta := \min\{\deg(h_1), \dots, \deg(h_r)\}$. Sea M un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $M \geq 6LS$,*
- ii) $\log(\delta) \geq \max\left\{2(1 + \log(d^\ell + 1)), 2t \left(\frac{\log(\deg_{\text{ici}}(\Lambda))}{s} + \log(d^{\ell+1} - 2)\right)\right\}$ y,*
- iii) $\max\{\deg(h_1), \dots, \deg(h_r)\} \leq (1 + \frac{1}{n-m})\delta$.*

Sea $R := R(\Omega, \Sigma, V, M)$ el conjunto constructible de todas las listas $\mathbf{Q} \in V^M$ de longitud M que son conjuntos cuestores para Ω con respecto a Σ . Entonces, existe un subconjunto abierto no vacío $\mathbb{G}(V) \subseteq \mathbb{G}(n, n - r)$, con respecto a la topología final inducida por la aplicación suprayectiva \mathcal{G} introducida en la Ecuación (1.2.2), tal que para cada $A \in \mathbb{G}(V)$, la probabilidad de que una lista $\mathbf{Q} \in (V \cap A)^M$ esté en R satisfice:

$$\text{Prob}_{(V \cap A)^M}[R] \geq 1 - \frac{1}{\deg_{\text{ici}}(\Lambda) e^{s(\log(d^{\ell+1} - 2) + 1) + (m-1)M}},$$

donde $(A \cap V)^M$ está dotado de la distribución de probabilidad uniforme.

DEMOSTRACIÓN. Es una aplicación inmediata del Corolario 3.2.7 usando las cotas obtenidas en la Proposición 5.4.1. \square

5.5. Redes neuronales con función de activación racional que evalúan funciones racionales

En esta sección, reducimos las redes neuronales con función de activación racional a redes neuronales con función de activación $\varphi(t) = t^2$. Además, discutimos cotas superiores para la función de crecimiento de la clase de clasificadores binarios asociada así como la longitud y la densidad de los conjuntos cuestores para este tipo de redes neuronales.

LEMA 5.5.1. *Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y m nodos de salida. Supongamos que \mathcal{N} tiene N aristas y que $\mathcal{A} = \{\varphi\}$, donde $\varphi := \frac{p}{q}$ es una función racional univariada, dada como el cociente de dos polinomios univariados coprimos, de grado d (i.e. $\deg(\varphi) = \max\{\deg(p), \deg(q)\} = d$). Entonces, existe una red neuronal $\mathcal{N}' := (G', \mathcal{A}', \Phi')$ con variables de entrada en $\{X_1, \dots, X_n\}$ sobre el mismo cuerpo K que verifica las siguientes propiedades:*

- i) *La clase \mathcal{A}' de nuevas funciones de activación se reduce a una única función $\mathcal{A}' := \{t^2\}$.*
- ii) *La talla de \mathcal{N}' está en $O((d + S^2)L)$, la profundidad de \mathcal{N}' está en $O(\ell(\log_2(d) + \log_2(S)))$, los requerimientos de espacio de \mathcal{N}' están en $O(S(d + S^2))$ y el número de aristas de \mathcal{N}' es del orden $O(N(d + S^2))$.*
- iii) *El número de nodos de salida de \mathcal{N}' doblan al número m de nodos de salida de \mathcal{N} . Además, si $O_{\mathcal{N}} := \{\nu_1, \dots, \nu_m\}$ son los nodos de salida de \mathcal{N} , los nodos de salida de \mathcal{N}' están dados por la siguiente igualdad:*

$$O_{\mathcal{N}'} := \{(\nu_1, 1), (\nu_1, 2), \dots, (\nu_m, 1), (\nu_m, 2)\}.$$

- iv) *Las funciones evaluadas por \mathcal{N}' y las funciones evaluadas por \mathcal{N} están relacionadas de la siguiente forma:*

$$f_{\nu_i}^{\mathcal{N}}(\underline{A}, X_1, \dots, X_n) = \frac{f_{(\nu_i, 1)}^{\mathcal{N}'}(\underline{A}, X_1, \dots, X_n)}{f_{(\nu_i, 2)}^{\mathcal{N}'}(\underline{A}, X_1, \dots, X_n)},$$

para cada i , $1 \leq i \leq m$.

DEMOSTRACIÓN. La idea es construir una nueva red neuronal $\mathcal{N}' = (G', \mathcal{A}', \Phi')$ a partir de $\mathcal{N} = (G, \mathcal{A}, \Phi)$, donde cada nodo ν del grafo G es reemplazado por una red neuronal $\mathcal{N}_{\varphi}^{I\nu}$ que evalúa el numerador y el denominador de φ en el nodo ν . Para hacer esto, primero construimos una red neuronal que evalúe φ de la siguiente manera. Supongamos que los polinomios p y q que definen φ vienen dados por las siguientes identidades:

$$p(T) := b_d T^d + b_{d-1} T^{d-1} + \dots + b_0, \quad q(T) := c_d T^d + c_{d-1} T^{d-1} + \dots + c_0,$$

donde $b_i, c_j \in K$. Sea I un conjunto finito de cardinalidad $\sharp(I) = M$. Definimos los dos siguientes polinomios multivariados parametrizados basados en las homogeneizaciones de p y q :

$$P_I(\underline{A}, Z_1, \dots, Z_M, Y, X_1, \dots, X_M) := \sum_{k=0}^d b_k Y^{d-k} \left(\sum_{i \in I} A_i Z_i X_i \right)^k,$$

y

$$Q_I(\underline{A}, Z_1, \dots, Z_M, Y, X_1, \dots, X_M) := \sum_{k=0}^d c_k Y^{d-k} \left(\sum_{i \in I} A_i Z_i X_i \right)^k,$$

donde $\underline{A} := (A_1, \dots, A_M) := (A_i : i \in I)$ es una lista de variables paramétricas. Supongamos ahora que nos dan un par de elementos en K :

$$\{(\alpha_i, \beta_i) \in K^2 : i \in I\},$$

donde $\beta_i \neq 0$ para todo $i \in I$ y una lista de parámetros:

$$\underline{a} := (a_1, \dots, a_M) = (a_i : i \in I).$$

Definimos los siguientes elementos de K :

$$\theta := \prod_{i \in I} \beta_i,$$

y para cada $i \in I$, definimos:

$$\theta_i := \left(\prod_{j \in I, j \neq i} \beta_j \right) = \frac{\theta}{\beta_i}.$$

Nótese que se verifica la siguiente cadena de igualdades:

$$\varphi \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right) = \frac{p \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right)}{q \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right)} = \frac{(\prod_{i \in I} \beta_i)^d p \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right)}{(\prod_{i \in I} \beta_i)^d q \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right)},$$

y, por lo tanto,

$$\varphi \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right) = \frac{\sum_{k=0}^d b_k (\prod_{i \in I} \beta_i)^{d-k} \left(\sum_{i \in I} a_i \left(\prod_{j \in I, j \neq i} \beta_j \right) \alpha_i \right)^k}{\sum_{k=0}^d c_k (\prod_{i \in I} \beta_i)^{d-k} \left(\sum_{i \in I} a_i \left(\prod_{j \in I, j \neq i} \beta_j \right) \alpha_i \right)^k}.$$

Es decir, tenemos que:

$$\varphi \left(\sum_{i \in I} a_i \frac{\alpha_i}{\beta_i} \right) = \frac{P_I(\underline{a}, \theta_1, \dots, \theta_M, \theta, \alpha_1, \dots, \alpha_M)}{Q_I(\underline{a}, \theta_1, \dots, \theta_M, \theta, \alpha_1, \dots, \alpha_M)}.$$

Usando esta última identidad, concluimos que para cada conjunto finito I de cardinalidad $\sharp(I) = M$, existe una red neuronal \mathcal{N}_φ^I con variables de entrada en $\{X_1, \dots, X_M, Y_1, \dots, Y_M\}$ de talla y requerimientos de espacio en $O(d + M^2)$, profundidad en $O(\log_2(d) + \log_2(M))$, parámetros en $\{A_1, \dots, A_M\}$, función de activación $\phi(t) := t^2$ y dos nodos de salida: (1) y (2), que realiza las siguientes tareas:

- En primer lugar, la red neuronal calcula las siguientes cantidades para cada $i \in I$:

$$Z_i := \prod_{j \in I, j \neq i} Y_j,$$

utilizando $O(M^2) = O(\sharp(I)^2)$ nodos y profundidad $O(\log_2(M))$.

- Además, en $O(1)$ nodos y profundidad $O(1)$, la red neuronal calcula $Y := \prod_{i \in I} Y_i$ (simplemente multiplicando Z_1 por Y_1 , por ejemplo).
- En $O(M)$ nodos adicionales y profundidad $O(1)$, la red neuronal calcula la siguiente lista de productos:

$$Z_i X_i, \quad i \in I.$$

- Luego, la red neuronal evalúa para cada k , $0 \leq k \leq d$:

$$\left(\sum_{i \in I} A_i Z_i X_i \right)^k,$$

empleando $O(d)$ nodos en profundidad $O(\log_2(d))$.

- Usando $O(d)$ nodos adicionales en profundidad $O(\log_2(d))$, la red neuronal evalúa:

$$Y^{d-k}, \quad 0 \leq k \leq d.$$

- Con solo $O(d)$ nodos en profundidad $O(1)$, usando (b_0, \dots, b_d) y (c_0, \dots, c_d) como parámetros, la red neuronal calcula los nodos de salida (1) y (2) de la siguiente forma:

$$P_I(\underline{A}, Z_1, \dots, Z_M, Y, X_1, \dots, X_M),$$

y

$$Q_I(\underline{A}, Z_1, \dots, Z_M, Y, X_1, \dots, X_M).$$

Esta idea elemental, aplicada a la red neuronal $\mathcal{N} := (G, \mathcal{A}, \Phi)$, nos proporciona el resultado. Para cada nodo ν de \mathcal{N} , consideramos $I_\nu := \text{Fan} - \text{In}(\nu)$ el abanico de entrada de ν . Definimos entonces $\mathcal{N}' := (G', \mathcal{A}', \Phi')$ de la siguiente manera. En primer lugar, reemplazados la ocurrencia del nodo ν en el grafo G de \mathcal{N} por la red neuronal $\mathcal{N}_\varphi^{I_\nu}$ asociada al nodo ν , determinada por φ y la lista de índices I_ν , tal y como se describe en las identidades previas. Esto determina el nuevo grafo G' . La nueva red únicamente usa la función de activación $\mathcal{A}' := \{t^2\}$. Para cada nodo ν del grafo original G , el subgrafo asociado a $\mathcal{N}_\varphi^{I_\nu}$ tiene dos nodos de salida: $\{(\nu, 1), (\nu, 2)\}$. Estos dos nodos asociados al nodo original ν satisfacen la siguiente propiedad:

Supongamos que la función original evaluada en ν por \mathcal{N} es una función de la forma:

$$f_\nu^{\mathcal{N}}((A_\nu^\mu : \mu \in I_\nu), X_1, \dots, X_n).$$

Supongamos que las dos funciones asociadas al nodo ν en la nueva red neuronal \mathcal{N}' vienen dadas por:

$$f_{(\nu, 1)}^{\mathcal{N}'}((A_\nu^\mu : \mu \in I_\nu), X_1, \dots, X_n),$$

y

$$f_{(\nu, 2)}^{\mathcal{N}'}((A_\nu^\mu : \mu \in I_\nu), X_1, \dots, X_n),$$

donde $M = \sharp(I_\nu)$. Estas dos funciones están dadas recursivamente en términos de la profundidad mediante la siguiente regla:

$$\begin{aligned} f_{(\nu,1)}^{\mathcal{N}'} &:= P_{I_\nu} \left((A_\nu^\mu : \mu \in I_\nu), Z_1^\nu, \dots, Z_M^\nu, Y, f_{(\mu_1,1)}^{\mathcal{N}'}, \dots, f_{(\mu_M,1)}^{\mathcal{N}'} \right), \\ f_{(\nu,2)}^{\mathcal{N}'} &:= Q_{I_\nu} \left((A_\nu^\mu : \mu \in I_\nu), Z_1^\nu, \dots, Z_M^\nu, Y, f_{(\mu_1,1)}^{\mathcal{N}'}, \dots, f_{(\mu_M,1)}^{\mathcal{N}'} \right), \end{aligned}$$

donde $\text{Fan} - \text{In}(\nu) = \{\mu_1, \dots, \mu_M\}$, Z_i^ν viene dado por la siguiente identidad:

$$Z_i^\nu := \prod_{\mu \in I_\nu, \mu \neq \mu_i} f_{(\mu,2)}^{\mathcal{N}'},$$

e

$$Y := \prod_{\mu \in I_\nu} f_{(\mu,2)}^{\mathcal{N}'}$$

Por lo tanto, recursivamente, podemos probar fácilmente que para cada nodo ν de la red neuronal original \mathcal{N} , tenemos que:

$$f_\nu^{\mathcal{N}}((A_\nu^\mu : \mu \in I_\nu), X_1, \dots, X_n) = \frac{f_{(\nu,1)}^{\mathcal{N}'}((A_\nu^\mu : \mu \in I_\nu), X_1, \dots, X_n)}{f_{(\nu,2)}^{\mathcal{N}'}((A_\nu^\mu : \mu \in I_\nu), X_1, \dots, X_n)},$$

y se obtiene el resultado buscado. \square

COROLARIO 5.5.2. *Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , de talla L , requerimientos de espacio acotados por S y m nodos de salida. Supongamos que \mathcal{N} tiene N aristas y que $\mathcal{A} = \{\varphi\}$, donde $\varphi := \frac{p}{q}$ es una función racional univariada, dada por el cociente de dos polinomios univariados coprimos, de grado d (i.e. $\deg(\varphi) = \max\{\deg(p), \deg(q)\} = d$). Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de los parámetros de \mathcal{N} . Sea $V(G)$ el conjunto de nodos del grafo G . Sea $\{A_\nu^\mu : \nu \in V(G), \mu \in \text{Fan} - \text{In}(\nu)\}$ el conjunto de las variables de los parámetros de la red neuronal y sea:*

$$K[A_\nu^\mu : \nu \in V(G), \mu \in \text{Fan} - \text{In}(\nu)],$$

el anillo de polinomios con coeficientes en K en este conjunto de variables. Entonces, tenemos que:

i) Existen dos subconjuntos constructibles:

$$\mathcal{W}_1(\mathcal{N}, \Lambda), \mathcal{W}_2(\mathcal{N}, \Lambda) \subseteq \mathcal{P}_{((dS)^\ell)}^K(X_1, \dots, X_n),$$

donde $((dS)^\ell) = ((dS)^\ell, \dots, (dS)^\ell) \in \mathbb{N}^m$, tales que las funciones evaluadas en los m nodos de salida $\{\nu_1, \dots, \nu_m\}$ de \mathcal{N} , mediante instanciación de parámetros en $\underline{a} \in \Lambda$, pueden describirse como pares de listas $(f_{(1,\underline{a})}, \dots, f_{(m,\underline{a})}) \in \mathcal{W}_1(\mathcal{N}, \Lambda)$, $(g_{(1,\underline{a})}, \dots, g_{(m,\underline{a})}) \in \mathcal{W}_2(\mathcal{N}, \Lambda)$, tales que, para cada i , $1 \leq i \leq m$, tenemos que:

$$f_{\nu_i}^{\mathcal{N}}(\underline{a}, X_1, \dots, X_n) = \frac{f_{(i,\underline{a})}(X_1, \dots, X_n)}{g_{(i,\underline{a})}(X_1, \dots, X_n)} \in K(X_1, \dots, X_n).$$

ii) Las dimensiones verifican las siguientes desigualdades para cada $i \in \{1, 2\}$:

$$\dim(\mathcal{W}_i(\mathcal{N}, \Lambda)) \leq \dim(\Lambda) \leq LS.$$

iii) Existe una constante universal $c > 0$, independiente de \mathcal{N} y φ , tal que los grados de las clausuras Zariski de $\mathcal{W}_1(\mathcal{N}, \Lambda)$ y $\mathcal{W}_2(\mathcal{N}, \Lambda)$ verifican las siguientes desigualdades para cada $i \in \{1, 2\}$:

$$\deg_z(\mathcal{W}_i(\mathcal{N}, \Lambda)) \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{\dim(\Lambda)} \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{LS}.$$

DEMOSTRACIÓN. Usamos la red neuronal \mathcal{N}' del Lema 5.5.1 y las estimaciones de la Proposición 5.4.1. La red neuronal \mathcal{N}' , construida a partir de \mathcal{N} y φ , es una red neuronal de profundidad $O(\ell(\log_2(d) + \log_2(S)))$. Definimos los conjuntos constructibles $\mathcal{W}_1(\mathcal{N}, \Lambda)$ y $\mathcal{W}_2(\mathcal{N}, \Lambda)$ de la siguiente manera:

$$\begin{aligned} \mathcal{W}_1(\mathcal{N}, \Lambda) &= \{(f_{(\nu_1,1)}^{\mathcal{N}'}, \dots, f_{(\nu_m,1)}^{\mathcal{N}'}) : \underline{a} \in \Lambda\}, \\ \mathcal{W}_2(\mathcal{N}, \Lambda) &= \{(f_{(\nu_1,2)}^{\mathcal{N}'}, \dots, f_{(\nu_m,2)}^{\mathcal{N}'}) : \underline{a} \in \Lambda\}. \end{aligned}$$

Observando las operaciones realizadas, deducimos fácilmente que los grados de los polinomios en los nodos de salida de \mathcal{N}' , con respecto a las variables $\{X_1, \dots, X_n\}$, están acotados por $(dS)^\ell$. A continuación, dado un vector de parámetros $\underline{a} \in \Lambda$, consideramos para cada i , $1 \leq i \leq m$:

$$f_{(i,\underline{a})}(X_1, \dots, X_n) = f_{(\nu_i,1)}^{\mathcal{N}'}(\underline{a}, X_1, \dots, X_n),$$

y

$$g_{(i,\underline{a})}(X_1, \dots, X_n) = f_{(\nu_i,2)}^{\mathcal{N}'}(\underline{a}, X_1, \dots, X_n).$$

Obviamente, $(f_{(1,\underline{a})}, \dots, f_{(m,\underline{a})}) \in \mathcal{W}_1(\mathcal{N}, \Lambda)$, $(g_{(1,\underline{a})}, \dots, g_{(m,\underline{a})}) \in \mathcal{W}_2(\mathcal{N}, \Lambda)$ y se verifica el Ítem *i*).

Excepto por los coeficientes del numerador y el denominador de φ , que se utilizan instanciados, los parámetros de \mathcal{N}' son los mismos que los de \mathcal{N} . Por lo tanto, la dimensión de estos conjuntos constructibles está acotada por la dimensión del conjunto constructible de parámetros Λ de \mathcal{N} , y, por lo tanto, se cumple el Ítem *ii*).

Para el último ítem, podemos considerar \mathcal{N} como una red neuronal doble, una para los numeradores y otra para los denominadores. Aunque la talla de las redes neuronales aumenta, su profundidad está acotada por $c\ell(\log_2(d) + \log_2(S))$ para una constante universal $c > 0$. Por otro lado, la función de activación $\phi(t) = t^2$ utilizada por \mathcal{N}' tiene grado 2. Usando las cotas del Ítem *i* de la Proposición 5.4.1, concluimos que tanto $\mathcal{W}_1(\mathcal{N}, \Lambda)$ como $\mathcal{W}_2(\mathcal{N}, \Lambda)$ están parametrizados por polinomios de grado a lo sumo:

$$2^{c\ell(\log_2(d)+\log_2(S))+1} - 2 = 2(dS)^{c\ell} - 2,$$

con parámetros en el conjunto constructible Λ . Por lo tanto, aplicando el Corolario 2.2.4, concluimos que para cada $i \in \{1, 2\}$:

$$\deg_z(\mathcal{W}_i(\mathcal{N}, \Lambda)) \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{\dim(\Lambda)} \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{LS},$$

lo que prueba el Ítem *iii*) del corolario. \square

LEMA 5.5.3. *Con las mismas notaciones e hipótesis anteriores, existe un subconjunto constructible $\mathcal{C}(\mathcal{N}, \Lambda) \subseteq \mathcal{W}_1(\mathcal{N}, \Lambda) \times \mathcal{W}_2(\mathcal{N}, \Lambda) \subseteq \mathcal{P}_{((dS)^\ell)}^K(X_1, \dots, X_n)^2$, donde $((dS)^\ell) = ((dS)^\ell, \dots, (dS)^\ell) \in \mathbb{N}^m$, de modo que, a cada vector de parámetros $\underline{a} \in \Lambda$, le podemos asociar una lista de polinomios $(f_{(1,\underline{a})}, \dots, f_{(m,\underline{a})}, g_{(1,\underline{a})}, \dots, g_{(m,\underline{a})}) \in \mathcal{C}(\mathcal{N}, \Lambda)$ que satisface, para cada i , $1 \leq i \leq m$:*

$$f_{\nu_i}^{\mathcal{N}}(\underline{a}, X_1, \dots, X_n) = \frac{f_{(i,\underline{a})}(X_1, \dots, X_n)}{g_{(i,\underline{a})}(X_1, \dots, X_n)} \in K(X_1, \dots, X_n).$$

Además, tenemos que:

$$\dim(\mathcal{C}(\mathcal{N}, \Lambda)) \leq \dim(\Lambda) \leq LS,$$

$$\deg_z(\mathcal{C}(\mathcal{N}, \Lambda)) \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{\dim(\Lambda)} \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{LS},$$

donde $c > 0$ es una constante universal.

DEMOSTRACIÓN. Sea \underline{X} el conjunto de variables $\{X_1, \dots, X_n\}$. Al igual que en el corolario anterior, consideramos para cada i , $1 \leq i \leq n$:

$$f_{(i,\underline{a})}(\underline{X}) = f_{(\nu_i,1)}^{\mathcal{N}'}(\underline{a}, \underline{X}) \text{ y } g_{(i,\underline{a})}(\underline{X}) = f_{(\nu_i,2)}^{\mathcal{N}'}(\underline{a}, \underline{X}),$$

donde \mathcal{N}' es la red neuronal introducida en el Lema 5.5.1. A continuación, definimos la siguiente aplicación:

$$\begin{aligned} Q: \quad \Lambda &\longrightarrow \mathcal{W}_1(\mathcal{N}, \Lambda) \times \mathcal{W}_2(\mathcal{N}, \Lambda) \subseteq \mathcal{P}_{((dS)^\ell)}^K(X_1, \dots, X_n)^2 \\ \underline{a} &\longmapsto \left(f_{(\nu_1,1)}^{\mathcal{N}'}(\underline{a}, \underline{X}), \dots, f_{(\nu_m,1)}^{\mathcal{N}'}(\underline{a}, \underline{X}), f_{(\nu_1,2)}^{\mathcal{N}'}(\underline{a}, \underline{X}), \dots, f_{(\nu_m,2)}^{\mathcal{N}'}(\underline{a}, \underline{X}) \right) \end{aligned}$$

y el conjunto constructible $\mathcal{C}(\mathcal{N}, \Lambda) = \text{Im}(Q) = Q(\Lambda)$. Obviamente,

$$(f_{(1,\underline{a})}, \dots, f_{(m,\underline{a})}, g_{(1,\underline{a})}, \dots, g_{(m,\underline{a})}) \in \mathcal{C}(\mathcal{N}, \Lambda).$$

Gracias al Ítem *i*) de la Proposición 5.4.1, es inmediato concluir que $\mathcal{C}(\mathcal{N}, \Lambda)$ es la imagen del conjunto constructible Λ por una lista de polinomios de grado $2(dS)^{c\ell} - 2$. Por lo tanto, tenemos que:

$$\dim(\mathcal{C}(\mathcal{N}, \Lambda)) \leq \dim(\Lambda) \leq LS$$

y, aplicando el Corolario 2.2.4, obtenemos:

$$\deg_z(\mathcal{C}(\mathcal{N}, \Lambda)) \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{\dim(\Lambda)} \leq \deg_{\text{lci}}(\Lambda) (2(dS)^{c\ell} - 2)^{LS}.$$

\square

Sea $h \in K(X_1, \dots, X_n)$ una función racional, definimos el abierto distinguido definido por h como el subconjunto abierto Zariski dado por:

$$D(h) := \left\{ x \in \mathbb{A}^n(K) : f(x) \cdot g(x) \neq 0, h = \frac{f}{g}, f, g \in K[X_1, \dots, X_n] \right\}$$

A continuación, obtenemos estimaciones para la función de crecimiento de la clase de conjuntos abiertos distinguidos definida por una red neuronal con función de activación racional:

COROLARIO 5.5.4. Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y un único nodo de salida. Supongamos que \mathcal{N} tiene N aristas y que $\mathcal{A} = \{\varphi\}$, donde $\varphi := \frac{p}{q}$ es una función racional univariada, dada por el cociente de dos polinomios univariados coprimos, de grado d (i.e. $\deg(\varphi) = \max\{\deg(p), \deg(q)\} = d$). Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de los parámetros de \mathcal{N} y sea $\mathcal{W}(\mathcal{N}, \Lambda)$ la clase de todas las funciones racionales evaluables por \mathcal{N} con parámetros en Λ . Sea \mathcal{H} la clase de clasificadores dada por las funciones características de los subconjuntos abiertos distinguidos de $\mathbb{A}^n(K)$ definidos por funciones racionales en $\mathcal{W}(\mathcal{N}, \Lambda)$:

$$\mathcal{H} = \{\chi_{D(h)} : h \in \mathcal{W}(\mathcal{N}, \Lambda)\}.$$

Entonces,

$$G(\mathcal{H}, m) \leq \deg_{\text{lci}}(\Lambda) \left((2(dS)^{c\ell} - 2) \cdot (1 + 2m((dS)^\ell + 1)) \right)^{LS},$$

donde $c > 0$ es una constante universal.

DEMOSTRACIÓN. Del Lema 5.5.3, deducimos que, para toda función racional $h \in \mathcal{W}(\mathcal{N}, \Lambda)$, existe un par de polinomios $(f, g) \in \mathcal{C}(\mathcal{N}, \Lambda)$ tal que $h = f/g$. Por otro lado, se cumple que:

$$h(x) = \frac{f(x)}{g(x)} \neq 0 \text{ si y solo si } f(x) \cdot g(x) \neq 0.$$

De lo anterior, se deduce la siguiente igualdad:

$$\mathcal{H} = \{\chi_{D(h)} : h \in \mathcal{W}(\mathcal{N}, \Lambda)\} = \{\chi_{D(f \cdot g)} : (f, g) \in \mathcal{C}(\mathcal{N}, \Lambda)\}.$$

Definimos ahora el conjunto constructible:

$$V := \{((f, g), x) \in \mathbb{A}^{2D}(K) \times \mathbb{A}^n(K) : f(x) \cdot g(x) \neq 0\},$$

donde:

$$D = \binom{(dS)^\ell + n}{n}.$$

Nótese que V es un subconjunto abierto Zariski dado como el complementario de una hipersuperficie algebraica de grado $2((dS)^\ell + 1)$. Finalmente, para obtener la desigualdad del enunciado, basta tener en cuenta que la función de crecimiento es monótona con respecto a la inclusión de familias de clasificadores y combinar el Teorema 4.2.1 con las estimaciones obtenidas en el Lema 5.5.3:

$$G(\mathcal{H}, m) \leq \deg_{\text{lci}}(\Lambda) \left((2(dS)^{c\ell} - 2) \cdot (1 + 2m((dS)^\ell + 1)) \right)^{LS}.$$

□

Generalizamos la noción clásica de conjuntos cuestores para conjuntos constructibles de polinomios a clases de funciones racionales como sigue:

DEFINICIÓN 33. Sea $\mathcal{R} \subseteq K[X_1, \dots, X_n]$ una clase de funciones racionales. Un conjunto cuestor de longitud M para \mathcal{R} es un conjunto finito de M elementos $\mathbf{Q} := \{x_1, \dots, x_M\} \subseteq (K^n)^M$ tal que:

$$\forall h = \frac{f}{g} \in \mathcal{R}, f, g \in K[X_1, \dots, X_n], (f \cdot g)(x_1) = 0, \dots, (f \cdot g)(x_M) = 0 \Rightarrow h = 0.$$

COROLARIO 5.5.5. Sea K un cuerpo algebraicamente cerrado. Sea $\mathcal{N} := (G, \mathcal{A}, \Phi)$ una red neuronal con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, de profundidad ℓ , talla L , requerimientos de espacio acotados por S y un único nodo de salida. Supongamos que \mathcal{N} tiene N aristas y que $\mathcal{A} = \{\varphi\}$, donde $\varphi := \frac{p}{q}$ es una función racional univariada, dada como el cociente de dos polinomios univariados coprimos, de grado d (i.e. $\deg(\varphi) = \max\{\deg(p), \deg(q)\} = d$). Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de los parámetros de \mathcal{N} y sea $\mathcal{W}(\mathcal{N}, \Lambda)$ la clase de todas las funciones racionales evaluables por \mathcal{N} con parámetros en Λ . Sea $\Omega = \mathcal{W}_p(\mathcal{N}, \Lambda)$ el conjunto constructible dado por:

$$\mathcal{W}_p(\mathcal{N}, \Lambda) = \left\{ f \cdot g : h = \frac{f}{g} \in \mathcal{W}(\mathcal{N}, \Lambda) \right\}.$$

Sea $s := \dim(\Lambda)$ la dimensión de Λ y $t := s/\dim(\Omega)$ el cociente entre las dimensiones de Λ y Ω . Sea $V := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ una variedad algebraica cero-dimensional dada por ecuaciones polinomiales del mismo grado $\delta := \deg(h_i)$, $1 \leq i \leq r$. Supongamos que $\deg(V) = \delta^n$. Sea $M \in \mathbb{N}$ un entero positivo y supongamos que se verifican las siguientes propiedades:

- i) $M \geq 6LS$ y,
- ii) $\log(\delta) \geq \max \left\{ 2(1 + \log(2(dS)^\ell + 1)), 2t \left(\frac{\log(\deg_{\text{lci}}(\Lambda))}{s} + \log(4(dS)^{c\ell} - 4) \right) \right\},$

donde $c > 0$ es una constante universal. Sea $R := R(\mathcal{W}(\mathcal{N}, \Lambda), \Omega, V, M)$ el conjunto constructible de todas las listas $\mathbf{Q} \in V^M$ de longitud M que son conjuntos cuestores para $\mathcal{W}(\mathcal{N}, \Lambda)$. Supongamos que V está dotado de su distribución de probabilidad uniforme. Entonces, tenemos que:

$$\text{Prob}_{V^M}[R] \geq 1 - \frac{1}{\text{deg}_{\text{ici}}(\Lambda)e^{s(\log(4(dS)^{c\ell}-4)+1)}},$$

DEMOSTRACIÓN. Comenzamos aplicando el Lema 5.5.1 a la red neuronal \mathcal{N} para obtener la red neuronal \mathcal{N}' . A continuación, construimos una nueva red neuronal \mathcal{N}'' añadiendo $O(1)$ nodos a la red neuronal \mathcal{N}' , de manera que \mathcal{N}'' calcula el producto de los dos nodos de salida de la red neuronal \mathcal{N}' . Obviamente, el grado del nodo de salida de \mathcal{N}'' está acotado por $2(dS)^\ell$. De lo anterior, se deduce la siguiente igualdad:

$$\mathcal{W}_p(\mathcal{N}, \Lambda) = \mathcal{W}(\mathcal{N}'', \Lambda) = \{f_{(\nu,1)}^{\mathcal{N}''}(\underline{a}, X_1, \dots, X_n) \cdot f_{(\nu,2)}^{\mathcal{N}''}(\underline{a}, X_1, \dots, X_n) : \underline{a} \in \Lambda\}.$$

Por otro lado, señalar que los parámetros de \mathcal{N}'' son los mismos que los de \mathcal{N}' y, por lo tanto, tenemos la siguiente cota superior para la dimensión de $\mathcal{W}_p(\mathcal{N}, \Lambda)$:

$$\dim(\mathcal{W}_p(\mathcal{N}, \Lambda)) \leq \dim(\Lambda) \leq LS.$$

A partir de la Proposición 5.4.1 y del Corolario 5.5.2, concluimos que $\mathcal{W}_p(\mathcal{N}, \Lambda)$ es un conjunto constructible que está parametrizado por una lista de polinomios de grado a lo sumo $4(dS)^{c\ell} - 4$. A continuación, aplicando el Corolario 2.2.4, obtenemos:

$$\text{deg}_z(\mathcal{W}_p(\mathcal{N}, \Lambda)) \leq \text{deg}_{\text{ici}}(\Lambda)(4(dS)^{c\ell} - 4)^{\dim(\Lambda)} \leq \text{deg}_{\text{ici}}(\Lambda)(4(dS)^{c\ell} - 4)^{\dim(\Lambda)}.$$

Finalmente, el resultado del corolario se obtiene al aplicar los Corolarios 3.2.6 y 3.2.7 con las estimaciones previas para el conjunto constructible $\mathcal{W}_p(\mathcal{N}, \Lambda)$. \square

Por último, usamos los conjuntos cuestores para realizar Tests de Identidad de funciones racionales dadas por redes neuronales:

COROLARIO 5.5.6. Sea K un cuerpo algebraicamente cerrado. Sean $\mathcal{N}_1 := (G_1, \mathcal{A}_1, \Phi_1)$ y $\mathcal{N}_2 := (G_2, \mathcal{A}_2, \Phi_2)$ dos redes neuronales con nodos de entrada asignados a las variables en $\{X_1, \dots, X_n\}$, profundidad ℓ , talla L , requerimientos de espacio acotados por S , un único nodo de salida y N aristas. Supongamos que $\mathcal{A}_1 = \{\varphi\}$ y $\mathcal{A}_2 = \{\psi\}$, donde φ y ψ son funciones racionales univariadas, dadas como cociente de dos polinomios coprimos, de grado d . Sea $\Lambda \subseteq \mathbb{A}^N(K)$ un conjunto constructible del espacio afín de los parámetros de \mathcal{N}_1 y sea $\mathcal{W}(\mathcal{N}_1, \Lambda)$ la clase de todas las funciones racionales evaluables por \mathcal{N}_1 con parámetros en Λ . Análogamente, sea $\Gamma \subseteq \mathbb{A}^N(K)$ un conjunto constructible en el espacio afín de los parámetros de \mathcal{N}_2 y sea $\mathcal{W}(\mathcal{N}_2, \Gamma)$ la clase de todas las funciones racionales evaluables por \mathcal{N}_2 con parámetros en Γ . Sea:

$$\mathcal{W}(\mathcal{N}_1, \Lambda) - \mathcal{W}(\mathcal{N}_2, \Gamma) = \{h_1 - h_2 : h_1 \in \mathcal{W}(\mathcal{N}_1, \Lambda), h_2 \in \mathcal{W}(\mathcal{N}_2, \Gamma)\}$$

la clase de las diferencias entre los elementos de $\mathcal{W}(\mathcal{N}_1, \Lambda)$ y $\mathcal{W}(\mathcal{N}_2, \Gamma)$. Sea $\Omega = \mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma))$ el conjunto constructible definido por:

$$\mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma)) = \left\{ f \cdot g : h = \frac{f}{g} \in \mathcal{W}(\mathcal{N}_1, \Lambda) - \mathcal{W}(\mathcal{N}_2, \Gamma) \right\}.$$

Sea $s := \dim(\Lambda \times \Gamma)$ la dimensión de $\Lambda \times \Gamma$ y $t := s / \dim(\Omega)$ el cociente entre las dimensiones de $\Lambda \times \Gamma$ y Ω . Sea $V := V_{\mathbb{A}}(h_1, \dots, h_r) \subseteq \mathbb{A}^n(K)$ una variedad algebraica cero-dimensional dada por ecuaciones polinomiales del mismo grado $\delta := \deg(h_i)$, $1 \leq i \leq n$. Supongamos que $\deg(V) = \delta^n$. Sea $M \in \mathbb{N}$ un entero positivo y supongamos que se verifican la siguientes propiedades:

- i) $M \geq 12LS$ y,
- ii) $\log(\delta) \geq \max \left\{ 2(1 + \log(4(dS)^\ell + 1)), 2t \left(\frac{\log(\text{deg}_{\text{ici}}(\Lambda)\text{deg}_{\text{ici}}(\Gamma))}{s} + \log(8(dS)^{c\ell} - 8) \right) \right\},$

donde $c > 0$ es una constante universal. Sea

$$R := R(\mathcal{W}(\mathcal{N}_1, \Lambda) - \mathcal{W}(\mathcal{N}_2, \Gamma), \Omega, V, M)$$

el conjunto constructible de todas las listas $\mathbf{Q} \in V^M$ de longitud M que son conjuntos cuestores para $\mathcal{W}(\mathcal{N}_1, \Lambda) - \mathcal{W}(\mathcal{N}_2, \Gamma)$. Supongamos que V está dotado de su distribución de probabilidad uniforme. Entonces, tenemos que:

$$\text{Prob}_{V^M}[R] \geq 1 - \frac{1}{\text{deg}_{\text{ici}}(\Lambda)e^{s(\log(8(dS)^{c\ell}-8)+1)}},$$

DEMOSTRACIÓN. En primer lugar, aplicamos el Lema 5.5.1 a las redes neuronales \mathcal{N}_1 y \mathcal{N}_2 para obtener las redes neuronales \mathcal{N}'_1 y \mathcal{N}'_2 , respectivamente. A continuación, construimos una nueva red neuronal \mathcal{N}'' combinando \mathcal{N}'_1 y \mathcal{N}'_2 y añadiendo $O(1)$ nodos que calcula el producto del numerador y el denominador de la

diferencia entre las funciones racionales asociadas a \mathcal{N}'_1 y \mathcal{N}'_2 . Para ello, supongamos que las funciones calculadas por \mathcal{N}'_1 son $f_1^{\mathcal{N}'_1}$ (numerador) y $f_2^{\mathcal{N}'_1}$ (denominador) y que las calculadas por \mathcal{N}'_2 son $f_2^{\mathcal{N}'_2}$ (numerador) y $f_1^{\mathcal{N}'_2}$ (denominador). Podemos calcular la salida de la red neuronal \mathcal{N}'' de la siguiente manera:

$$(f_1^{\mathcal{N}'_1} \cdot f_2^{\mathcal{N}'_2} - f_2^{\mathcal{N}'_1} \cdot f_1^{\mathcal{N}'_2}) \cdot (f_2^{\mathcal{N}'_1} \cdot f_2^{\mathcal{N}'_2})$$

El grado del nodo de salida de la red neuronal \mathcal{N}'' está acotado por $4(dS)^\ell$. Sea \underline{X} el conjunto de las variables $\{X_1, \dots, X_n\}$. Tenemos que:

$$\begin{aligned} \mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma)) &= \mathcal{W}(\mathcal{N}'', \Lambda \times \Gamma) = \\ &= \left\{ (f_1^{\mathcal{N}'_1}(\underline{a}, \underline{X}) \cdot f_2^{\mathcal{N}'_2}(\underline{b}, \underline{X}) - f_2^{\mathcal{N}'_1}(\underline{a}, \underline{X}) \cdot f_1^{\mathcal{N}'_2}(\underline{b}, \underline{X})) \cdot (f_2^{\mathcal{N}'_1}(\underline{a}, \underline{X}) \cdot f_2^{\mathcal{N}'_2}(\underline{b}, \underline{X})) : \underline{a} \in \Lambda, \underline{b} \in \Gamma \right\}. \end{aligned}$$

Gracias a la Proposición 5.4.1 y al Corolario 5.5.2, podemos concluir que $\mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma))$ es la imagen del conjunto constructible $\Lambda \times \Gamma$ por una lista de polinomios de grado a lo sumo $8(dS)^{\ell} - 8$. Obviamente, tenemos que:

$$\dim(\mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma))) \leq \dim(\Lambda) + \dim(\Gamma) \leq 2LS.$$

Usando ahora el Corolario 2.2.4, obtenemos:

$$\begin{aligned} \deg_z(\mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma))) &\leq \deg_{\text{lci}}(\Lambda \times \Gamma)(8(dS)^{\ell} - 8)^{\dim(\Lambda \times \Gamma)} \leq \\ &\leq \deg_{\text{lci}}(\Lambda) \deg_{\text{lci}}(\Gamma)(8(dS)^{\ell} - 8)^{2LS}. \end{aligned}$$

Finalmente, aplicamos los Corolarios 3.2.6 y 3.2.7 usando las cotas anteriores para el conjunto constructible $\mathcal{W}_p((\mathcal{N}_1, \Lambda), (\mathcal{N}_2, \Gamma))$ para obtener el resultado. \square

En el contexto de la Teoría de Aprendizaje Computacional, este último resultado nos permite comprobar si dos redes neuronales con función de activación racional han “aprendido” la misma función.

Un enfoque alternativo a la relación entre aprendizaje y polinomios multivariados: Reescritura y Aprendizaje multiclase. Grado de salida promedio, dimensiones, shifting y relaciones con bases monomiales de K -álgebras de Artin

Índice

6.1. Nociones básicas: Pre-orden de Dickson y bases monomiales	115
6.1.1. Pre-orden de Dickson y clases de conceptos cerradas hacia abajo: escalones y escaleras	116
6.1.2. Órdenes monomiales y Bases de Gröbner	118
6.2. El hiper-grafo de una inclusión: grado promedio, grado de salida promedio y orientaciones	122
6.2.1. Grado promedio, grado de salida promedio y densidad del hiper-grafo de una inclusión	124
6.2.2. Grado promedio, grado de salida promedio y conjuntos cerrados hacia abajo	137
6.2.2.1. Ejemplo: escalones \mathcal{E}_h	137
6.3. Pseudo-cubos, dimensión DS y dimensión exponencial	139
6.3.1. Pseudo-cubos: definición, caracterización y propiedades básicas	139
6.3.2. Dimensiones	141
6.3.3. Pseudo-dimensión DS	144
6.3.4. El escalón \mathcal{E}_h y sus dimensiones	145
6.3.5. Pseudo-cubos cerrados hacia abajo	146
6.4. La técnica del “shifting” en el Aprendizaje multiclase	150
6.4.1. La técnica del shifting	150
6.4.2. Shifting, clases cerradas hacia abajo y bases monomiales	152
6.4.2.1. Las ideas de [Me, 19]	157
6.4.3. Shifting e invariantes de clases de conceptos finitas	159
6.4.3.1. Shifting: grado de salida promedio	159
6.4.3.2. Shifting: grado promedio	164
6.4.3.3. Orientaciones con grado de salida máximo controlado	168

En este capítulo discutimos las propiedades básicas de las nociones comunes en el estudio del hiper-grafo de una inclusión, asociado a una clase de conceptos, en el contexto del Aprendizaje multiclase. Nos centramos principalmente en las nociones presentadas en [RBR, 06], [DS, 14] y [BCDMY, 22]. Además, introducimos y estudiamos tres nuevas nociones: el grado de salida promedio, la pseudo-dimensión DS y el grado de salida promedio maximal del hiper-grafo de una inclusión. Cabe destacar que en este capítulo se realiza el primer estudio sistemático de la noción de pseudo-cubo, que es clave para definir la dimensión DS del Aprendizaje multiclase.

Por otro lado, llevamos a cabo un análisis novedoso de la técnica del shifting, introducida en [RBR, 06] para el caso multiclase. En particular, empleando esta técnica, obtenemos relaciones entre los conjuntos cerrados hacia abajo y las bases monomiales de K -álgebras de Artin finitamente generadas (con una nueva demostración de un resultado obtenido en [Me, 19]). Nuestro análisis permite mejorar varios de los resultados obtenidos en [BCDMY, 22].

Parte del material presentado en este capítulo está ligado al artículo en preparación [PSZ, 25].

6.1. Nociones básicas: Pre-orden de Dickson y bases monomiales

El objetivo de esta primera sección del capítulo es presentar la terminología y algunos resultados básicos sobre el pre-orden de Dickson, los conjuntos cerrados hacia abajo (también conocidos como “escaleras”), los órdenes monomiales y las Bases de Gröbner que utilizaremos en el resto del capítulo. Además, estudiaremos la relación que existe entre los conjuntos cerrados hacia abajo y las bases monomiales de K -álgebras de Artin. En particular,

en el Corolario 6.1.11 identificamos los conjuntos cerrados hacia abajo con bases monomiales de ciertas K -álgebras de Artin, dadas como los cocientes $K[X_1, \dots, X_n]/\mathfrak{a}$, para un ideal monomial \mathfrak{a} construido ac hoc, con respecto a cualquier orden monomial. En la Subsección 6.4.2 probaremos una versión “más completa” para el caso del orden lexicográfico canónico (cf. Corolario 6.4.12).

Antes de pasar a discutir todo lo anterior, introducimos la noción básica de nuestro estudio: *las clases de conceptos (finitas)*. Al igual que en [BCDMY, 22], sea $[n] := \{1, \dots, n\}$ el conjunto de los primeros enteros positivos. Por otro lado, denotamos por $\mathcal{Y} := \mathbb{Z}/p\mathbb{Z} := \{0, \dots, p-1\}$ al conjunto de los representantes canónicos de las clases de restos módulo p (con $p \geq 2$). Cabe mencionar que, normalmente, en el contexto del Aprendizaje multiclase, los autores prefieren considerar $\mathcal{Y} = [p] := \{1, \dots, p\}$. Obviamente, ambos conjuntos son “equivalentes”. Nosotros hemos preferido utilizar $\mathcal{Y} = \{0, 1, \dots, p-1\}$ para reflejar la conexión con los monomios considerada en este capítulo. Finalmente, llamamos *clase de conceptos (finita)* $H \subseteq \mathcal{Y}^{[n]} = \mathcal{Y}^n$ a cualquier subconjunto de la clase de funciones $\mathcal{Y}^{[n]}$. Cada una de las funciones $h : [n] \rightarrow \mathcal{Y}$ de la clase de funciones $\mathcal{Y}^{[n]}$ puede identificarse de forma evidente con un vector $(h(1), \dots, h(n)) \in (\mathbb{Z}/p\mathbb{Z})^n$. Emplearemos ambas interpretaciones de los elementos de $\mathcal{Y}^{[n]}$ indistintamente. Para cada subconjunto $S \subseteq [n]$, denotamos por $H|_S \subseteq \mathcal{Y}^S$ a la clase de restricciones a S de las funciones en H , i.e.

$$(6.1.1) \quad \{h|_S : S \rightarrow \mathcal{Y} : h \in H\}.$$

En el caso $p = 2$, hablamos de Aprendizaje binario, el cual responde a las técnicas y propiedades descritas en las Secciones 1.3 y 4.2. Por lo tanto, el Aprendizaje binario es solo un caso particular del Aprendizaje multiclase que discutimos en este capítulo. Curiosamente, al generalizarlo, con $p \geq 3$, los resultados del caso binario resultan insuficientes. Por ello, muchos autores, comenzando en [Na, 88] y [Na, 89] hasta el más reciente [BCDMY, 22], han desarrollado elementos propios que permiten entender mejor las ideas del aprendizaje en el caso multiclase. Este capítulo pretende profundizar en algunas de las nociones que van surgiendo, ofreciendo nociones y resultados nuevos y aproximaciones originales con el propósito de cerrar esta teoría.

6.1.1. Pre-orden de Dickson y clases de conceptos cerradas hacia abajo: escalones y escaleras.

En esta subsección introducimos el pre-orden de Dickson y los conjuntos cerrados hacia abajo así como algunas de sus propiedades. La mayor parte de los resultados de esta subsección son elementales y se deducen de forma casi inmediata a partir de las definiciones, por lo que omitiremos sus demostraciones.

Sea $H \subseteq \mathcal{Y}^{[n]} = \mathcal{Y}^n$ una clase de conceptos. Dotamos a $\mathcal{Y}^{[n]}$ (y también a H) del pre-orden de Dickson. Es decir, consideramos la siguiente relación de orden parcial: dados $h := (h_1, \dots, h_n), g := (g_1, \dots, g_n) \in \mathcal{Y}^{[n]}$, decimos que h es menor que g con respecto al pre-orden de Dickson (y lo denotamos por $h \preceq g$) si y solo si se cumple lo siguiente:

$$h_i \leq g_i, \quad 1 \leq i \leq n.$$

Consideramos también el pre-orden de Dickson en toda la clase de listas de longitud n de enteros positivos, es decir, (\mathbb{N}^n, \preceq) es un conjunto pre-ordenado. El siguiente lema muestra el resultado fundamental de L. E. Dickson (cf. [Di, 1913]):

LEMA 6.1.1 (Lema de Dickson). *Con las notaciones anteriores, todo subconjunto no vacío $S \subseteq \mathbb{N}^n$ posee un número finito de elementos minimales con respecto a \preceq .*

A continuación, introducimos la noción de *conjunto cerrado hacia abajo*, que jugará un papel central en nuestra discusión:

DEFINICIÓN 34 (Conjuntos cerrados hacia abajo). *Decimos que una clase de conceptos $H \subseteq \mathcal{Y}^{[n]}$ es cerrada hacia abajo (con respecto al pre-orden de Dickson \preceq) si se verifica la siguiente propiedad:*

$$\forall h \in H, \forall g \in \mathcal{Y}^{[n]}, g \preceq h \implies g \in H.$$

El ejemplo más sencillo de una clase de conceptos cerrada hacia abajo es el *escalón* determinado por algún elemento $h \in \mathbb{N}^n$:

$$(6.1.2) \quad \mathcal{E}_h := \{g \in \mathbb{N}^n : g \preceq h\}.$$

Obviamente, si $h \in H$ y H es una clase de conceptos cerrada hacia abajo, entonces $\mathcal{E}_h \subseteq H$. El siguiente resultado muestra que los conjuntos finitos cerrados hacia abajo pueden expresarse como unión finita de escalones:

PROPOSICIÓN 6.1.2 (Escalera finita). *Una clase de conceptos finita es cerrada hacia abajo si y solo si H admite una descomposición del siguiente tipo: existen $h_1, \dots, h_s \in \mathbb{N}^n$ tales que se verifican las siguientes propiedades:*

i) *Cada elemento h_i no está en los escalones determinados por el resto de los elementos:*

$$h_i \notin \bigcup_{j \neq i} \mathcal{E}_{h_j}.$$

ii) La clase H es la unión finita de tales escalones (i.e. H es una escalera):

$$H := \bigcup_{i=1}^s \mathcal{E}_{h_i}.$$

Además, los escalones $\mathcal{E}_{h_1}, \dots, \mathcal{E}_{h_s}$ están determinados de manera única por H y los elementos h_1, \dots, h_s son elementos maximales de H con respecto al pre-orden de Dickson. Los escalones (únicamente determinados) $\mathcal{E}_{h_1}, \dots, \mathcal{E}_{h_s}$ reciben el nombre de componentes (o escalones) maximales del conjunto finito cerrado hacia abajo H .

DEMOSTRACIÓN. Si H es un conjunto finito cerrado hacia abajo, entonces tiene solo un número finito de elementos maximales con respecto a \preceq . Como \mathcal{E}_h es siempre cerrado hacia abajo, obviamente tenemos que si $h \in H$, el escalón que determina verifica que $\mathcal{E}_h \subseteq H$. Esto prueba que si H es cerrado hacia abajo, entonces es la unión finita de los escalones de sus elementos maximales con respecto a \preceq , lo que demuestra la Propiedad ii).

Nótese que los elementos maximales de H no son comparables respecto al pre-orden de Dickson. Para demostrarlo, supongamos que h_1 y h_2 son dos elementos maximales distintos de H y que $h_1 \preceq h_2$. Como $h_1 \neq h_2$, esto contradice el hecho de que h_1 sea maximal. Por lo tanto, concluimos que dos elementos maximales distintos no son comparables. De lo anterior, se deduce que para cada $i \in \{1, \dots, s\}$,

$$h_i \notin \bigcup_{j \neq i} \mathcal{E}_{h_j},$$

lo que prueba la Propiedad i).

Obviamente, las uniones finitas de escalones son conjuntos cerrados hacia abajo, lo que prueba el recíproco. \square

Por lo discutido en la proposición anterior, también llamaremos escaleras a los conjuntos cerrados hacia abajo.

OBSERVACIÓN 6.1.3. Dado un subconjunto finito $\mathcal{E} \subseteq \mathbb{N}^n$ como unión de escalones:

$$\mathcal{E} := \bigcup_{k=1}^t \mathcal{E}_{h_k},$$

es obvio que \mathcal{E} es cerrado hacia abajo. Además, esta descomposición puede refinarse para obtener una descomposición que cumpla las Propiedades i) y ii) de la proposición precedente. Esto es, tomando los índices de los elementos maximales $S \subseteq \{1, \dots, t\}$, tenemos que la descomposición:

$$\mathcal{E} = \bigcup_{k \in S} \mathcal{E}_{h_k},$$

satisface las Propiedades i) y ii) de la Proposición 6.1.2.

Por otro lado, dado $h \in \mathbb{N}^n$, definimos el *anti-escalón* de h como el siguiente subconjunto de \mathbb{N}^n :

$$h + \mathbb{N}^n := \{g \in \mathbb{N}^n : \exists m \in \mathbb{N}^n, h + m = g\} = \{g \in \mathbb{N}^n : h \preceq g\},$$

donde $h + m$ denota la suma estándar en el monoide $(\mathbb{N}^n, +)$.

PROPOSICIÓN 6.1.4. Con las mismas notaciones e hipótesis que antes, para cada $H \subseteq \mathcal{Y}^{[n]}$, H es cerrado hacia abajo si y solo existen $g_1, \dots, g_t \in \mathbb{N}^n$ no comparables entre sí con respecto al pre-orden de Dickson tales que:

$$H := \bigcap_{i=1}^t (\mathbb{N}^n - (g_i + \mathbb{N}^n)).$$

Además, los elementos g_1, \dots, g_t están determinados de manera única por H por ser los elementos minimales de $\mathbb{N}^n \setminus H$ con respecto al pre-orden de Dickson \preceq .

DEMOSTRACIÓN. Basta considerar el subconjunto $\mathbb{N}^n \setminus H$ y aplicar el Lema de Dickson (cf. Lema 6.1.1) para concluir que solo existe un número finito $\{g_1, \dots, g_t\}$ de elementos minimales de $\mathbb{N}^n \setminus H$. Estos son precisamente los elementos mencionados en el enunciado de la proposición. \square

Nuestro siguiente resultado prueba que la condición de ser cerrado hacia abajo es hereditaria por restricciones:

LEMA 6.1.5. Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita y $S \subseteq [n]$ un subconjunto. Entonces, H es cerrada hacia abajo si y solamente si lo son todas sus restricciones $H|_S$ para cualquier subconjunto $S \subseteq [n]$.

DEMOSTRACIÓN. Supongamos que H es una clase de conceptos finita cerrada hacia abajo. Sea $S \subseteq [n]$ un subconjunto. Sea $g \in H$ una función de la clase de conceptos H y consideremos su restricción $g|_S \in H|_S$ al subconjunto S . Consideramos también una función $f : S \rightarrow \mathcal{Y}$ tal que $f \preceq g|_S$. Definimos ahora la extensión $\tilde{f} : [n] \rightarrow \mathcal{Y}$ de f como sigue:

$$\tilde{f}(i) := \begin{cases} f(i), & \text{si } i \in S \\ g(i), & \text{si } i \notin S \end{cases}$$

Obviamente, $\tilde{f}|_S = f$ y $\tilde{f} \preceq g$. Por lo tanto, como H es cerrado hacia abajo, se cumple que $\tilde{f} \in H$. De lo anterior, se deduce inmediatamente que $f = \tilde{f}|_S \in H|_S$, lo que prueba que la clase $H|_S$ es cerrada hacia abajo. El recíproco es inmediato tomando $S = [n]$. \square

6.1.2. Órdenes monomiales y Bases de Gröbner. Esta subsección está dedicada a introducir la terminología y los resultados necesarios para demostrar el Corolario 6.1.11 así como los resultados de la Subsección 6.4.2. Comenzamos recordando los órdenes monomiales sobre el monoide $(\mathbb{N}^n, +)$.

DEFINICIÓN 35 (Orden monomial). Un orden monomial sobre el monoide $(\mathbb{N}^n, +)$ es un buen orden \leq sobre \mathbb{N}^n que satisface las siguientes propiedades:

- i) Dados $\mu, \nu, \tau \in \mathbb{N}^n$, si $\mu \leq \nu$, entonces $\mu + \tau \leq \nu + \tau$.
- ii) El elemento $\underline{0} := (0, \dots, 0) \in \mathbb{N}^n$ es el elemento mínimo de \mathbb{N}^n con respecto al orden \leq .

En el caso $n = 1$, cualquier orden monomial es el orden usual sobre \mathbb{N} .

OBSERVACIÓN 6.1.6. El pre-orden de Dickson introducido en la Subsección 6.1.1 no es un orden monomial para $n \geq 2$ (cf. [Pa, 25] para más detalles). Sin embargo, si $\mu, \nu \in \mathbb{N}^n$ y $\mu \preceq \nu$, entonces por la Propiedad i) de la Definición 35, tenemos que $\mu \leq \nu$ para cualquier orden monomial \leq .

A continuación, introducimos el orden monomial “lexicográfico canónico”, que emplearemos en nuestros resultados de la Subsección 6.4.2:

DEFINICIÓN 36 (Orden lexicográfico canónico). Con las notaciones anteriores, el orden lexicográfico canónico \leq_{lex} es el orden monomial definido por la siguiente regla: dados $\mu := (\mu_1, \dots, \mu_n), \nu := (\nu_1, \dots, \nu_n) \in \mathbb{N}^n$, decimos que $\mu \leq_{\text{lex}} \nu$ si existe $k \in \{0, \dots, n\}$ de modo que se verifican las dos propiedades siguientes:

- i) $\mu_i = \nu_i$, para $1 \leq i \leq k$,
- ii) $\mu_{k+1} < \nu_{k+1}$.

Usaremos la notación $\mu =_{\text{lex}} \nu$ si $k = n$, y $\mu <_{\text{lex}} \nu$ si $k < n$.

Por otro lado, es evidente que existe una biyección entre el monoide $(\mathbb{N}^n, +)$ y el monoide multiplicativo del conjunto de monomios en n variables. Sea K un cuerpo y $K[X_1, \dots, X_n]$ el anillo de polinomios en n variables con coeficientes en K . Consideramos el conjunto de sus monomios:

$$\mathcal{M}(X_1, \dots, X_n) := \{X_1^{\mu_1} \cdots X_n^{\mu_n} : (\mu_1, \dots, \mu_n) \in \mathbb{N}^n\},$$

que es una base de $K[X_1, \dots, X_n]$ como K -espacio vectorial. Tenemos el siguiente isomorfismo de monoïdes:

$$\begin{aligned} \exp : \quad (\mathbb{N}^n, +) &\longrightarrow (\mathcal{M}(X_1, \dots, X_n), \cdot) \\ \mu = (\mu_1, \dots, \mu_n) &\longmapsto X_1^{\mu_1} \cdots X_n^{\mu_n} \end{aligned}$$

donde \cdot es el producto del anillo $K[X_1, \dots, X_n]$ restringido a $\mathcal{M}(X_1, \dots, X_n)$. Obviamente, un orden monomial sobre \mathbb{N}^n puede transferirse a un orden sobre los monomios $\mathcal{M}(X_1, \dots, X_n)$ que respeta la operación natural de producto de monomios. Lo denotaremos usando los mismos símbolos, es decir, escribimos $X_1^{\mu_1} \cdots X_n^{\mu_n} \leq X_1^{\nu_1} \cdots X_n^{\nu_n}$ si y solo si $(\mu_1, \dots, \mu_n) \leq (\nu_1, \dots, \nu_n)$. Por esta razón, también llamaremos *exponentes monomiales* a los elementos de \mathbb{N}^n .

A partir de ahora, cuando el contexto esté claro, dado $\mu = (\mu_1, \dots, \mu_n)$, escribiremos \underline{X}^μ en lugar de $X_1^{\mu_1} \cdots X_n^{\mu_n}$. Consideramos el polinomio $f \in K[X_1, \dots, X_n]$. La expansión monomial de f viene dada por:

$$f = \sum_{\mu \in S} a_\mu \underline{X}^\mu,$$

donde $S \subseteq \mathbb{N}^n$ es un conjunto finito y $a_\mu \in K$. Este tipo de representación es posible ya que $\mathcal{M}(X_1, \dots, X_n)$ es una base de $K[X_1, \dots, X_n]$ como K -espacio vectorial. Además, podemos escoger un subconjunto finito $S \subseteq \mathbb{N}^n$ tal que $a_\mu \neq 0$ para todo $\mu \in S$. En este contexto, el conjunto S recibe el nombre de *soporte* de f y lo denotamos por $\text{supp}(f)$. Por lo tanto, podemos expresar f de la siguiente forma:

$$f = \sum_{\mu \in \text{supp}(f)} a_\mu \underline{X}^\mu.$$

Dado que $\text{supp}(f)$ es un conjunto finito, el orden inducido por \leq será un orden total y un buen orden. Por lo tanto, $\text{supp}(f)$ posee tanto elemento máximo como mínimo.

Por otro lado, llamamos monomio director de f (y lo denotamos por $\text{LM}_{\leq}(f)$) con respecto al orden monomial \leq sobre $\mathcal{M}(X_1, \dots, X_n)$ al máximo del conjunto $\text{supp}(f)$, i.e.

$$\text{LM}_{\leq}(f) = \underline{X}^{\mu},$$

donde $\mu = \max(\text{supp}(f))$. Además, usaremos la notación $\exp_{\leq}(f) = \mu$, donde $\mu = \max(\text{supp}(f))$. Evidentemente, el monomio director de f depende del orden monomial escogido.

Sea $\mathfrak{a} \subseteq K[X_1, \dots, X_n]$ un ideal y sea \leq un orden monomial, definimos el conjunto de los exponentes de \mathfrak{a} con respecto al orden \leq como:

$$\exp_{\leq}(\mathfrak{a}) := \{\mu \in \mathbb{N}^n : \exists f \in \mathfrak{a}, \underline{X}^{\mu} = \text{LM}_{\leq}(f)\} = \{\mu \in \mathbb{N}^n : \exists f \in \mathfrak{a}, \mu = \exp_{\leq}(f)\}.$$

Consideramos también su complementario:

$$\text{Nexp}_{\leq}(\mathfrak{a}) := \mathbb{N}^n \setminus \exp_{\leq}(\mathfrak{a}).$$

PROPOSICIÓN 6.1.7. *Sea \leq un orden monomial sobre $(\mathbb{N}^n, +)$, K un cuerpo y \mathfrak{a} un ideal del anillo de polinomios $K[X_1, \dots, X_n]$. Entonces, tenemos que:*

- i) *El conjunto $\exp_{\leq}(\mathfrak{a})$ es cerrado hacia arriba para el pre-orden de Dickson \preceq .*
- ii) *Su complementario $\text{Nexp}_{\leq}(\mathfrak{a})$ es cerrado hacia abajo para el pre-orden de Dickson \preceq .*

DEMOSTRACIÓN. Comenzamos demostrando el Ítem i). Sean $\mu, \nu \in \mathbb{N}^n$ dos exponentes monomiales. Si $\mu \preceq \nu$ y $\mu \in \exp_{\leq}(\mathfrak{a})$ es porque existen:

- $\gamma := (\gamma_1, \dots, \gamma_n) \in \mathbb{N}^n$ tal que $\nu = \mu + \gamma$ en el monoide $(\mathbb{N}^n, +)$.
- $f \in \mathfrak{a}$ tal que $\text{LM}_{\leq}(f) = \mu$.

Consideramos ahora el polinomio:

$$g := X_1^{\gamma_1} \cdots X_n^{\gamma_n} f \in \mathfrak{a}.$$

Como los órdenes monomiales respetan la estructura del monoide $(\mathbb{N}^n, +)$ (ver Propiedad i) de la Definición 35), concluimos:

$$\text{LM}_{\leq}(g) = \gamma + \text{LM}_{\leq}(f) = \gamma + \mu = \nu \in \exp_{\leq}(\mathfrak{a}).$$

Esto prueba que el conjunto $\exp_{\leq}(\mathfrak{a})$ es cerrado hacia arriba para el pre-orden de Dickson \preceq . Obviamente, su complementario será cerrado hacia abajo con respecto a \preceq , lo que prueba el Ítem ii). \square

El Lema de Dickson (ver Lema 6.1.1) nos permite concluir el siguiente resultado:

COROLARIO 6.1.8. *Si $\mathfrak{a} \subseteq K[X_1, \dots, X_n]$ es un ideal y \leq un orden monomial, entonces existe un conjunto finito de elementos $\{g_1, \dots, g_s\} \subseteq \mathfrak{a}$ tales que:*

$$\exp_{\leq}(\mathfrak{a}) = \bigcup_{i=1}^s \exp_{\leq}(g_i) + \mathbb{N}^n.$$

La división de Weierstrass-Hironaka garantiza el siguiente resultado:

TEOREMA 6.1.9 (Existencia de Bases de Gröbner). *Sea K un cuerpo, \mathfrak{a} un ideal del anillo de polinomios $K[X_1, \dots, X_n]$ y \leq un orden monomial sobre $(\mathbb{N}^n, +)$ (i.e. sobre $(\mathcal{M}(X_1, \dots, X_n), \cdot)$). Entonces, existen $g_1, \dots, g_s \in \mathfrak{a}$ tales que:*

- i) $\mathfrak{a} = (g_1, \dots, g_s)$,
- ii) $\exp_{\leq}(\mathfrak{a}) = \bigcup_{i=1}^s \exp_{\leq}(g_i) + \mathbb{N}^n$.

Además, el conjunto $\{\underline{X}^{\mu} + \mathfrak{a} : \mu \in \text{Nexp}_{\leq}(\mathfrak{a})\}$ es una base como K -espacio vectorial del anillo cociente $K[X_1, \dots, X_n]/\mathfrak{a}$ que recibe el nombre de base monomial y está en biyección con $\text{Nexp}_{\leq}(\mathfrak{a})$.

El conjunto de elementos $\{g_1, \dots, g_s\} \subseteq \mathfrak{a}$ que verifica las Propiedades i) y ii) del teorema anterior recibe el nombre de Base de Gröbner (o base estándar) del ideal \mathfrak{a} . Las Bases de Gröbner no son, en realidad, bases del ideal como K -espacio vectorial, sino sistemas finitos de generadores del ideal. El uso del término “base” en este contexto se debe a que el teorema anterior implica al Teorema de la Base de Hilbert para un anillo de polinomios $K[X_1, \dots, X_n]$ con coeficientes en un cuerpo K .

Por otra parte, tenemos el siguiente resultado:

TEOREMA 6.1.10. *Con las notaciones anteriores, tenemos que:*

- i) $\text{Nexp}_{\leq}(\mathfrak{a})$ es un conjunto finito.
- ii) El anillo $K[X_1, \dots, X_n]/\mathfrak{a}$ es un anillo de Artin.

- iii) El anillo $K[X_1, \dots, X_n]/\mathfrak{a}$ es un K -espacio vectorial de dimensión finita.
 iv) Los siguientes espectros coinciden:

$$\text{Spec}(K[X_1, \dots, X_n]/\mathfrak{a}) = \text{MaxSpec}(K[X_1, \dots, X_n]/\mathfrak{a}).$$

- v) Existe un número finito de ideales maximales: $\{\mathfrak{m}_1, \dots, \mathfrak{m}_s\} \subseteq \text{MaxSpec}(K[X_1, \dots, X_n])$, tales que $\mathfrak{a} \subseteq \mathfrak{m}_i$ y existen $n_1, \dots, n_s \in \mathbb{N}$, $n_i \geq 1$, tales que:

$$(0 + \mathfrak{a}) = (\mathfrak{m}_1 / \mathfrak{a})^{n_1} \cdots (\mathfrak{m}_s / \mathfrak{a})^{n_s}.$$

Si K es un cuerpo algebraicamente cerrado, las propiedades anteriores son equivalentes a:

- v) La variedad algebraica:

$$V_{\mathbb{A}}(\mathfrak{a}) := \{\zeta \in K^n : f(\zeta) = 0, \forall f \in \mathfrak{a}\}$$

es un conjunto finito de puntos.

Finalmente, estamos en condiciones de demostrar el resultado principal de esta subsección. En realidad, este resultado no debe ser desconocido. Lo enunciamos y lo demostramos por falta de una referencia apropiada.

COROLARIO 6.1.11 (Las escaleras finitas son equivalentes a las bases monomiales de K -álgebras de Artin). Sea K un cuerpo de característica distinta de 2. Sea $p \in \mathbb{N}$ un entero positivo con $p \geq 2$. Sea $\mathcal{Y} := \{0, \dots, p-1\}$ y $H \subseteq \mathcal{Y}^n$ un subconjunto finito. Entonces, los siguientes ítems son equivalentes:

- i) H es una escalera finita (i.e. un conjunto finito cerrado hacia abajo con respecto al pre-orden de Dickson \preceq).
 ii) Para cualquier orden monomial \leq sobre el monoide $(\mathbb{N}^n, +)$, existe un ideal \mathfrak{a} tal que:
- $H := \text{Nexp}_{\leq}(\mathfrak{a}) = \mathbb{N}^n \setminus \text{exp}_{\leq}(\mathfrak{a})$.
 - La K -álgebra $K[X_1, \dots, X_n]/\mathfrak{a}$ es una K -álgebra de Artin, que tiene la siguiente base monomial como K -espacio vectorial:

$$\mathcal{B}(H)/\mathfrak{a} := \{X_1^{\mu_1} \cdots X_n^{\mu_n} + \mathfrak{a} : \mu := (\mu_1, \dots, \mu_n) \in H\}.$$

Además, se verifica la siguiente igualdad:

$$\dim_K(K[X_1, \dots, X_n]/\mathfrak{a}) = \#(H).$$

Adicionalmente, en el Ítem ii), es posible escoger el ideal \mathfrak{a} de modo que sea un ideal monomial de $K[X_1, \dots, X_n]$, aunque no es imprescindible.

DEMOSTRACIÓN. Probamos cada implicación por separado:

- $i) \implies ii)$: Comenzamos considerando el ideal monomial $\mathfrak{b} := (X_1^p, \dots, X_n^p)$. Obviamente, el anillo residual $A := K[X_1, \dots, X_n]/\mathfrak{b}$ es un anillo de Artin y un K -espacio vectorial de dimensión p^n . A continuación, consideramos el conjunto $\mathcal{Y}^n \setminus H$ y el conjunto finito de monomios asociado:

$$\mathcal{M}(\mathcal{Y}^n \setminus H) := \{X_1^{\mu_1} \cdots X_n^{\mu_n} : \mu := (\mu_1, \dots, \mu_n) \in \mathcal{Y}^n \setminus H\}.$$

Definimos los exponentes $\{p \cdot e_1, \dots, p \cdot e_n\} \subseteq \mathbb{N}^n$, donde $\{e_1, \dots, e_n\} \subseteq \{0, 1\}^n$ es la base canónica usual de \mathbb{R}^n , y consideramos el conjunto complementario de H en \mathbb{N}^n , dado por:

$$H^c := \mathbb{N}^n \setminus H := (\mathcal{Y}^n \setminus H) \cup (\cup_{i=1}^n (p \cdot e_i + \mathbb{N}^n)).$$

Sea $\mathfrak{m}(H^c)$ el conjunto de elementos minimales de H^c con respecto al pre-orden de Dickson \preceq . Por la propia definición de elementos minimales de este conjunto y por ser este conjunto cerrado hacia arriba, tenemos que:

$$(6.1.3) \quad H^c = \bigcup_{\mu \in \mathfrak{m}} (\mu + \mathbb{N}^n).$$

Escribimos $\mathfrak{m}(H^c) := \{\mu_1, \dots, \mu_t\}$, con la condición de que son distintos dos a dos. Obviamente, tenemos la siguiente inclusión:

$$\mathfrak{m}(H^c) \subseteq H^c.$$

Finalmente, definimos la siguiente familia de monomios:

$$\mathcal{M}_0(H^c) := \{X_1^{\mu_{i,1}} \cdots X_n^{\mu_{i,n}} : \mu_i := (\mu_{i,1}, \dots, \mu_{i,n}) \in \mathfrak{m}(H^c)\}.$$

Consideramos ahora el ideal monomial \mathfrak{a} generado por $\mathcal{M}_0(H^c)$, esto es:

$$\mathfrak{a} := (\mathcal{M}_0(H^c)) = (\mathcal{M}_0(H^c)) + \mathfrak{b},$$

ya que los exponentes $p \cdot e_i$ de los monomios de la forma X_i^p deben pertenecer a algún conjunto $\mu_i + \mathbb{N}^n$, para algún $i \in [t]$ minimal en H^c con respecto al pre-orden de Dickson \preceq . Por lo tanto, $\mathfrak{b} \subseteq \mathfrak{a}$ y podemos considerar el ideal $\bar{\mathfrak{a}} := \mathfrak{a}/\mathfrak{b}$ de clases residuales de elementos en el ideal \mathfrak{a} módulo el ideal

b. Además, tenemos el epimorfismo natural de K -álgebras que, combinado con el Segundo Teorema de Isomorfía, nos proporciona:

$$\pi : A \longrightarrow A/\bar{\mathfrak{a}} \cong K[X_1, \dots, X_n]/\mathfrak{a}.$$

En particular, π es un epimorfismo de K -espacios vectoriales. Por el Teorema 6.1.10, concluimos que $K[X_1, \dots, X_n]/\mathfrak{a}$ es un K -espacio vectorial de dimensión finita y una K -álgebra de Artin. Además, nótese que el conjunto $\mathcal{M}_0(H^c)$ no es solamente un sistema generador del ideal monomial \mathfrak{a} de $K[X_1, \dots, X_n]$, sino que también es lo que en [CLO, 15] (Problema 8 de la Sección 4 del Capítulo 2) recibe el nombre de “*base minimal*” del ideal monomial \mathfrak{a} . La razón es que no puede haber relación de divisibilidad entre ellos ya que sus exponentes son los elementos minimales de H^c con respecto al pre-orden de Dickson y son distintos dos a dos. Finalmente, probamos que para cualquier orden monomial \leq , se cumple que:

$$(6.1.4) \quad \exp_{\leq}(\mathfrak{a}) = \bigcup_{i=1}^t (\mu_i + \mathbb{N}^n).$$

La inclusión \subseteq es obvia ya que \mathfrak{a} está generado por los monomios $\mathfrak{m}(H^c) := \{\mu_1, \dots, \mu_t\}$. La otra inclusión se sigue del hecho de que un orden monomial \leq induce una graduación sobre el anillo de polinomios $K[X_1, \dots, X_n]$ de la siguiente forma:

$$K[X_1, \dots, X_n] := \bigoplus_{\mu \in \mathbb{N}^n} K\langle X_1^{\mu_1} \cdots X_n^{\mu_n} \rangle,$$

donde $K\langle X_1^{\mu_1} \cdots X_n^{\mu_n} \rangle$ es el K -subespacio vectorial de $K[X_1, \dots, X_n]$ de dimensión 1 generado por el monomio $X_1^{\mu_1} \cdots X_n^{\mu_n}$. Los ideales monomiales son precisamente los ideales homogéneos con respecto a esta graduación y, por lo tanto, si $f \in \mathfrak{a}$, entonces todos sus términos están en \mathfrak{a} . En particular, $LM_{\leq}(f) \in \mathfrak{a}$ y, por consiguiente, debe ser divisible por alguno de los monomios en $\mathcal{M}_0(H^c)$ (que es lo que viene a decir el Lema 2 de la Sección 4 del Capítulo 2 de [CLO, 15]). Por lo tanto, para cada $f \in \mathfrak{a}$, tenemos que:

$$\exp_{\leq}(f) \in \bigcup_{i=1}^t (\mu_i + \mathbb{N}^n) = \bigcup_{\mu \in \mathfrak{m}(H^c)} (\mu + \mathbb{N}^n).$$

Esto prueba la Identidad (6.1.4) que, gracias a la Igualdad (6.1.3), podemos escribir como:

$$\exp_{\leq}(\mathfrak{a}) = \bigcup_{\mu \in \mathfrak{m}(H^c)} (\mu + \mathbb{N}^n) = H^c.$$

Finalmente, el teorema sobre la existencia de Bases de Gröbner (cf. Teorema 6.1.9) nos permite concluir que $\mathcal{B}(H)/\mathfrak{a}$ es una base de $K[X_1, \dots, X_n]/\mathfrak{a}$ como K -espacio vectorial. El resto de la demostración se deduce a partir de esta última conclusión.

- $ii) \implies i)$: Para demostrar esta implicación, basta probar que $\text{Nexp}_{\leq}(\mathfrak{a}) = \mathbb{N}^n \setminus \exp_{\leq}(\mathfrak{a})$ es siempre un conjunto finito cerrado hacia abajo, bajo la hipótesis del Ítem $ii)$. Consideramos $\mathfrak{b} = (X_1^p, \dots, X_n^p)$. En primer lugar, nótese que el anillo residual $A := K[X_1, \dots, X_n]/\mathfrak{b}$ es una K -álgebra de Artin, ya que es un K -espacio vectorial de dimensión p^n , cuya base puede expresarse de forma sencilla como clases residuales módulo \mathfrak{b} de elementos de \mathcal{Y}^n , y cumple el Teorema 6.1.10 precedente. Al igual que en la implicación anterior, definiendo $\bar{\mathfrak{a}} := \mathfrak{a}/\mathfrak{b}$ como las clases residuales de los elementos de $\mathfrak{a} \supseteq \mathfrak{b}$ módulo el ideal \mathfrak{b} , tenemos el siguiente epimorfismo de K -álgebras dado por Segundo Teorema de Isomorfía:

$$\pi : A \longrightarrow A/\bar{\mathfrak{a}} \cong K[X_1, \dots, X_n]/\mathfrak{a}.$$

Por lo tanto, usando el Teorema 6.1.10, concluimos que $\mathcal{B}(H)/\mathfrak{a}$ es base del anillo residual

$$K[X_1, \dots, X_n]/\mathfrak{a}$$

como K -espacio vectorial. Por consiguiente, $\text{Nexp}_{\leq}(\mathfrak{a})$ es un conjunto finito (biyectable con $\mathcal{B}(H)/\mathfrak{a}$ para cualquier orden monomial \leq). Por otro lado, gracias a Proposición 6.1.7, sabemos que el conjunto $\text{Nexp}_{\leq}(\mathfrak{a})$ es cerrado hacia abajo para el pre-orden de Dickson \preceq . En conclusión, tenemos que $\text{Nexp}_{\leq}(\mathfrak{a})$ es un conjunto finito cerrado hacia abajo o, equivalentemente, una escalera finita. \square

En conclusión, para cualquier escalera finita $H \subseteq \mathcal{Y}^n$ y para cualquier orden monomial \leq , existe un ideal \mathfrak{a} tal que:

$$H := \mathbb{N}^n \setminus \exp_{\leq}(\mathfrak{a}),$$

donde dicho ideal puede elegirse como un ideal monomial. Veremos más adelante (Subsección 6.4.2) una relación más intrínseca.

6.2. El hiper-grafo de una inclusión: grado promedio, grado de salida promedio y orientaciones

El objetivo de esta sección es presentar la noción de *hiper-grafo de una inclusión* (también conocido como *OIG: One Inclusion Graph*) y realizar un primer estudio de algunos de sus invariantes. Esta noción fue definida por primera vez en [Ha, 95, HLW, 94] para el Aprendizaje binario. Posteriormente, en [RBR, 06], los autores extendieron la noción al caso de Aprendizaje multiclase. Recientemente, esta noción ha sido discutida en [DS, 14] y [BCDMY, 22]. En nuestra exposición, seguiremos la terminología empleada en [BCDMY, 22], con pequeñas variaciones notacionales y corrigiendo algunas imprecisiones.

Con las mismas notaciones que en la sección precedente, sea $\pi_i : \mathcal{Y}^{[n]} \rightarrow \mathcal{Y}^{n-1}$ la proyección que “olvida” la i -ésima coordenada. Es decir,

$$(6.2.1) \quad \begin{array}{ccc} \mathcal{Y}^{[n]} & \longrightarrow & \mathcal{Y}^{[n] \setminus \{i\}} = \mathcal{Y}^{n-1} \\ (h_1, \dots, h_n) & \longmapsto & (h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_n). \end{array}$$

Para cada clase de conceptos finita $H \subseteq \mathcal{Y}^{[n]}$, denotamos por $\pi_i(H)$ a la imagen de H en la “dirección” de π_i . Obviamente, $\pi_i(H) = H|_{[n] \setminus \{i\}}$. Recíprocamente, para cada $f \in \mathcal{Y}^{n-1}$, consideramos la fibra sobre $\{f\}$ en la dirección de π_i :

$$\pi_i^{-1}(\{f\}) := \{h \in \mathcal{Y}^{[n]} : h|_{[n] \setminus \{i\}} = f\}.$$

Si intersecamos la fibra $\pi_i^{-1}(\{f\})$ con H , i.e.

$$\pi_i^{-1}(\{f\}) \cap H,$$

tenemos que esta intersección es no vacía si y solo si $f \in \pi_i(H)$. Por lo tanto, para cada $f \in \pi_i(H)$, definimos los siguientes subconjuntos de H , que constituyen las “aristas” de nuestro hiper-grafo de una inclusión:

$$(6.2.2) \quad e_{f,i} := \pi_i^{-1}(\{f\}) \cap H := \{h \in H : h|_{[n] \setminus \{i\}} = f\}.$$

En primer lugar, consideramos el conjunto de aristas en la dirección determinada por π_i :

$$E_i(H) := \{e_{f,i} : f \in \pi_i(H)\}.$$

Obviamente, las fibras en la dirección i -ésima determinan una partición de H . Además, tenemos que:

$$(6.2.3) \quad \#(E_i(H)) = \#(\pi_i(H)).$$

Finalmente, definimos el *conjunto de aristas* del hiper-grafo de una inclusión determinado por H (y las proyecciones) como la unión disjunta de las familias $E_i(H)$:

$$(6.2.4) \quad E(H) := \bigsqcup_{i \in [n]} E_i(H),$$

donde el símbolo \sqcup denota la unión disjunta. Una forma clásica de representar esta unión disjunta es la utilizada en [BCDMY, 22], que reproducimos a continuación. Para cada $i \in [n]$ y para cada $f \in \pi_i(H)$, consideramos el producto cartesiano:

$$e_{f,i} \times \{i\} \subseteq H \times [n].$$

Por lo tanto, podemos identificar:

$$E_i(H) \equiv \{e_{f,i} \times \{i\} : f \in \pi_i(H)\} \text{ y } E(H) \equiv \bigcup_{i \in [n]} \{e_{f,i} \times \{i\} : f \in \pi_i(H)\}.$$

Dado $e = e_{f,i} \times \{i\} \in E(H)$ y $h \in H$, escribimos $h \in e$ para indicar que $h \in e_{f,i}$ y definimos el cardinal de una arista como el cardinal de la fibra subyacente: $\#(e) := \#(e_{f,i})$. Una arista $e \in E(H)$ recibe el nombre de *unaria* si $\#(e) = 1$. Denotamos por $E^{(1)}(H)$ a la clase de aristas en $E(H)$ que son unarias.

DEFINICIÓN 37 (Hiper-grafo de una inclusión). Con las notaciones anteriores, definimos el hiper-grafo de una inclusión de la clase de conceptos $H \subseteq \mathcal{Y}^{[n]}$ como el hiper-grafo $G(H) := (H, E(H))$, donde $E(H)$ es el conjunto de aristas definido en la Identidad (6.2.4).

OBSERVACIÓN 6.2.1 (El hiper-grafo de una inclusión no es propiamente un hiper-grafo: las aristas unarias pueden ser “contadas” varias veces). Aunque hemos empleado el término hiper-grafo, la estructura $G(H)$, que acabamos de presentar en la Definición 37, no es propiamente un hiper-grafo. Algunas de las aristas pueden ser consideradas como objetos distintos dependiendo de la dirección tenida en cuenta y, por lo tanto, contadas varias veces. Dos aristas $e := e_{f,i} \times \{i\} \in E(H)$ y $e' := e_{g,j} \times \{j\} \in E(H)$ con $i \neq j$ y la misma fibra subyacente, contenida en H (esto es, $e_{f,i} = e_{g,j}$), pueden ser aristas unarias y, sin embargo, considerarse como aristas distintas. Esto es, puede existir $h \in H$ tal que $e = \{h\} \times \{i\}$ y $e' = \{h\} \times \{j\}$, pero e y e' se considerarán aristas distintas en nuestro hiper-grafo de una inclusión ya que $i \neq j$.

Resulta sencillo obtener ejemplos de esta situación. Basta tomar $H := \{(1, 1)\} \subseteq \{0, 1\}^2$ y considerar las dos aristas distintas $e := \{(1, 1)\} \times \{1\}$ y $e' := \{(1, 1)\} \times \{2\}$.

OBSERVACIÓN 6.2.2 (El hiper-grafo de una inclusión difiere de la estructura de grafo definida en [Ha, 95]). En [Ha, 95], D. Haussler asoció a las clases de conceptos finitas $H \subseteq \{0, 1\}^n$ (i.e. en el caso binario $p = 2$) una estructura de grafo que difiere de la que hemos presentado, de acuerdo a [BCDMY, 22], en la Definición 37. A continuación, vamos a comparar estas estructuras.

En [Ha, 95] se considera una función de distancia proporcional a la distancia de Hamming. Esto es, dados $f, g \in H$, D. Haussler define la siguiente función:

$$\rho(f, g) := \frac{1}{n} \sum_{i \in [n]} |f(i) - g(i)| = \frac{d_{\text{Ham}}(f, g)}{n},$$

donde $d_{\text{Ham}}(f, g)$ es la distancia de Hamming o, equivalentemente, donde:

$$|f(i) - g(i)| := \begin{cases} 1, & \text{si } f(i) \neq g(i) \\ 0, & \text{en otro caso} \end{cases}.$$

A continuación, reproducimos la misma función distancia para el caso general. Consideramos $H \subseteq Y^n$, donde $Y = \{0, \dots, p-1\}$, y la función distancia:

$$\begin{aligned} \rho : H \times H &\longrightarrow \mathbb{R}_+ \\ (f, g) &\longmapsto \frac{d_{\text{Ham}}(f, g)}{n}. \end{aligned}$$

Dado $h \in H$, consideramos también la bola cerrada de radio $1/n$ centrada en h con respecto a la distancia ρ :

$$\overline{B}_\rho(h, 1/n) := \{g \in H : \rho(h, g) \leq 1/n\}.$$

Resulta sencillo verificar que esta bola cerrada viene dada por:

$$\overline{B}_\rho(h, 1/n) = \bigcup_{i \in [n]} e_{\pi_i(h), i}.$$

Teniendo en cuenta que $h \in e_{\pi_i(h), i}$, $\forall i \in [n]$, obtenemos la siguiente igualdad:

$$\sharp(\overline{B}_\rho(h, 1/n)) = \sum_{i=1}^n \sharp(e_{\pi_i(h), i}) - (n-1).$$

En el caso $p = 2$ (i.e. caso binario), las aristas de la forma $e_{\pi_i(h), i}$ tienen a lo sumo dos elementos. En [Ha, 95], el autor considera una estructura de grafo sobre H definiendo las aristas de la siguiente forma:

$$\text{Haussl}(H) := \{(g, h) \in H^2 : \rho(g, h) = 1/n\}.$$

Nótese que un par $(h, g) \in \text{Haussl}(H)$ si y solo si $g \in \overline{B}_\rho(h, 1/n) \setminus \{h\}$. En particular, en el caso $p = 2$, tenemos una biyección entre el abanico de salida:

$$\text{Fan} - \text{Out}_{\text{Haussl}}(h) := \{g \in H : (h, g) \in \text{Haussl}(H)\},$$

y el conjunto de aristas que contienen a h y tienen exactamente dos elementos distintos, i.e.

$$\{e \in E(H) : h \in e, \sharp(e) = 2\}.$$

En particular, la cardinalidad del Fan-Out o abanico de salida de h en $\text{Haussl}(H)$ coincide con lo que llamaremos posteriormente el grado de h en $G(H)$ (cf. Definición 39):

$$\sharp(\text{Fan} - \text{Out}_{\text{Haussl}}(h)) = \deg(h).$$

Esto también significa que, en el caso $p = 2$ (i.e. caso binario), cada arista $e_{f,i}$ es contada dos veces si no es un elemento de $E^{(1)}(H)$. Por lo tanto, concluimos:

$$(6.2.5) \quad \sharp(\text{Haussl}(H)) = 2 \left(\sharp \left(E(H) \setminus E^{(1)}(H) \right) \right).$$

El siguiente lema establece las propiedades básicas del hiper-grafo de una inclusión:

LEMA 6.2.3. *Con las notaciones anteriores, se verifican las siguientes propiedades:*

i) Para cada $h \in H$, denotamos por $E(H)_h$ a la clase de todas las aristas que contienen a h . Es decir,

$$E(H)_h := \{e : h \in e\}.$$

Entonces, tenemos que:

$$\#(E(H)_h) = n.$$

ii) Se cumple la siguiente igualdad:

$$\#(E(H)) = \sum_{i \in [n]} \#(\pi_i(H)).$$

iii) Para cada $i \in [n]$, la siguiente descomposición describe una partición de H en la dirección de π_i :

$$H := \bigsqcup_{f \in \pi_i(H)} e_{f,i}.$$

Al tratarse de una unión disjunta, tenemos que:

$$\#(H) = \sum_{f \in \pi_i(H)} \#(e_{f,i}).$$

DEMOSTRACIÓN. Los tres ítems se deducen de forma inmediata a partir de las definiciones. \square

DEFINICIÓN 38. *Una orientación del hiper-grafo de una inclusión $G(H)$ es una aplicación:*

$$\sigma : E(H) \longrightarrow H,$$

que verifica la siguiente propiedad:

$$\forall e \in E(H), \sigma(e) \in e.$$

A continuación, presentamos una nueva partición del conjunto de aristas $E(H)$. Para cada $k \in \mathbb{N}$, definimos:

$$(6.2.6) \quad E^{(k)}(H) := \{e \in E(H) : \#(e) = k\}.$$

Dicho de otro modo, $E^{(k)}(H)$ es el conjunto de aristas cuyas fibras subyacentes tienen cardinal k . Como cada arista tiene a lo sumo p elementos (i.e. $\#(e_{f,i}) \leq \#(\mathcal{Y}) = p$), la siguiente descomposición es una partición de $E(H)$ en términos del número de elementos de las aristas:

$$E(H) := \bigsqcup_{k=1}^p E^{(k)}(H).$$

DEFINICIÓN 39 (**Grado, Grado de salida**). *Con las notaciones precedentes, definimos:*

- Para cada $h \in H$, el grado de h en $G(H)$ viene dado por:

$$\deg_H(h) := \#(\{e \in E(H) : h \in e, e \notin E^{(1)}(H)\}) = n - \#(\{e \in E^{(1)}(H) : h \in e\}).$$

- Para cada $h \in H$ y para cada orientación $\sigma : E(H) \longrightarrow H$, definimos el grado de salida de σ en h mediante la siguiente expresión:

$$\text{outdeg}_H(\sigma, h) := \#(\{e \in E(H) : h \in e, \sigma(e) \neq h\}).$$

NOTACIÓN 6.2.4. Como es habitual, omitiremos el sub-índice H de \deg_H y outdeg_H y escribiremos simplemente $\deg(h)$ y $\text{outdeg}(\sigma, h)$, siempre que H y $G(H)$ queden determinados por el contexto. No obstante, ambas cantidades dependen fuertemente de la clase de conceptos finita H y del hiper-grafo de una inclusión asociado $G(H)$.

6.2.1. Grado promedio, grado de salida promedio y densidad del hiper-grafo de una inclusión.

DEFINICIÓN 40 (**Grado promedio, Grado de salida promedio y Densidad del hiper-grafo**). *Con las notaciones anteriores, definimos nuestros primeros invariantes cuantitativos del hiper-grafo de una inclusión $G(H)$ (con respecto a la distribución uniforme en H):*

i) El grado promedio de los vértices $h \in H$ viene dado por la siguiente igualdad:

$$(6.2.7) \quad \text{avd}(H) := \frac{1}{\#(H)} \sum_{h \in H} \deg(h).$$

ii) El grado de salida promedio de una orientación $\sigma : E(H) \longrightarrow H$ se define como:

$$(6.2.8) \quad E_H[\text{outdeg}(\sigma)] := \frac{1}{\sharp(H)} \sum_{h \in H} \text{outdeg}(\sigma, h).$$

iii) La densidad de $G(H)$ viene dada por el siguiente número racional:

$$\text{gd}(H) := \frac{\sharp(E(H))}{\sharp(H)}.$$

iv) El grado de salida máximo de una orientación se define mediante la siguiente expresión:

$$\text{outdeg}_{\max}(\sigma) := \max\{\text{outdeg}(\sigma, h) : h \in H\}.$$

Nótese que $\text{avd}(H)$ depende únicamente de $G(H)$, mientras que $E_H[\text{outdeg}(\sigma)]$ parece depender también de la orientación $\sigma : E(H) \longrightarrow H$. Nuestra primera tarea será caracterizar estas nociones en términos del hiper-grafo de una inclusión $G(H)$. En particular, demostraremos que el *grado de salida promedio no depende de la orientación escogida* (ver Ítem ii) de la Proposición 6.2.5), sino que es un invariante de $G(H)$ que solo depende de la dimensión del espacio “ambiente” (i.e. n) y de la *densidad del hiper-grafo* $G(H)$ (término que hemos tomado de [RBR, 06]).

PROPOSICIÓN 6.2.5 (Principales caracterizaciones de $\text{avd}(H)$, $E_H[\text{outdeg}]$ y $\text{gd}(H)$). Con las notaciones anteriores y suponiendo que $n \geq 2$, tenemos que:

i) El grado promedio del hiper-grafo de una inclusión $G(H)$ verifica:

$$\text{avd}(H) = n - \frac{\sharp(E^{(1)}(H))}{\sharp(H)}.$$

ii) El grado de salida promedio de cualquier orientación $\sigma : E(H) \longrightarrow H$ satisface:

$$E_H[\text{outdeg}(\sigma)] = n - \text{gd}(H) = n - \frac{\sharp(E(H))}{\sharp(H)}.$$

En particular, el grado de salida promedio de una orientación es independiente de la orientación escogida y solo depende de la dimensión n del espacio ambiente \mathcal{Y}^n y de la densidad del hiper-grafo $\text{gd}(H)$. Por lo tanto, a partir de ahora, escribiremos simplemente $E_H[\text{outdeg}]$ para denotar al grado de salida promedio de cualquier orientación del hiper-grafo. En particular, tenemos que:

$$\text{avd}(H) - E_H[\text{outdeg}] - \text{gd}(H) = -\frac{\sharp(E^{(1)}(H))}{\sharp(H)} \leq 0.$$

iii) También se cumplen las siguientes desigualdades:

$$E_H[\text{outdeg}] \leq \text{avd}(H) \leq 2E_H[\text{outdeg}].$$

iv) Para cualquier orientación σ , tenemos que:

$$\text{gd}(H) - \frac{\sharp(E^{(1)}(H))}{\sharp(H)} \leq E_H[\text{outdeg}].$$

v) Sea $G \subseteq H$ una subclase de H y sea $\sigma : E(H) \longrightarrow H$ una orientación. Entonces, existe una orientación $\sigma' : E(G) \longrightarrow G$ de modo que se verifica:

$$\forall g \in G, \text{outdeg}(\sigma', g) \leq \text{outdeg}(\sigma, g).$$

En particular, para toda orientación $\sigma : E(H) \longrightarrow H$, existe una orientación $\sigma' : E(G) \longrightarrow G$ tal que:

$$\text{outdeg}_{\max}(\sigma') \leq \text{outdeg}_{\max}(\sigma).$$

Finalmente, el grado de salida promedio raramente coincide con el grado promedio, ya que se cumple que:

- O bien

$$E_H[\text{outdeg}] < \text{avd}(H),$$

- o bien $E(H) = E^{(1)}(H)$ (i.e. $G(H)$ tiene solo aristas unarias), en cuyo caso:

$$E_H[\text{outdeg}] = \text{avd}(H) = 0.$$

DEMOSTRACIÓN. Demostramos cada ítem por separado:

- *Ítem i)*: En primer lugar, dado $i \in [n]$, consideramos la clase $E_i(H)$. Tenemos que:

$$\#\{\{e \in E_i(H) : h \in e, \#(e) \geq 2\}\} = \#\{\{e \in E_i(H) : h \in e\}\} - \#\{\{e \in E_i(H) : e = \{h\}\}\}.$$

Como exactamente n aristas contienen al vértice $h \in H$, tenemos que:

$$\sum_{i=1}^n \#\{\{e \in E_i(H) : h \in e\}\} = n.$$

Por lo tanto, como $E_1(H), \dots, E_n(H)$ definen una partición de $E(H)$, obtenemos:

$$\sum_{i=1}^n \left(\sum_{h \in H} \#\{\{e \in E_i(H) : h \in e, \#(e) \geq 2\}\} \right) = \sum_{h \in H} \deg(h).$$

Intercambiando los sumatorios, tenemos que:

$$\sum_{h \in H} \deg(h) = \sum_{h \in H} \left(\sum_{i=1}^n \#\{\{e \in E_i(H) : h \in e, \#(e) \geq 2\}\} \right).$$

Por otro lado, se verifica la siguiente igualdad:

$$\sum_{i=1}^n \#\{\{e \in E_i(H) : h \in e, \#(e) \geq 2\}\} = n - \left(\sum_{i=1}^n \#\{\{e \in E_i(H) : e = \{h\}\}\} \right).$$

Por lo tanto, concluimos que:

$$(6.2.9) \quad \sum_{h \in H} \deg(h) = n\#(H) - \sum_{h \in H} \left(\sum_{i=1}^n \#\{\{e \in E_i(H) : e = \{h\}\}\} \right).$$

Para terminar, tenemos que:

$$E^{(1)}(H) = \bigcup_{h \in H} \left(\bigcup_{i=1}^n \{e \in E_i(H) : e = \{h\}\} \right)$$

es una descomposición como unión disjunta de aristas unarias. En términos de cardinalidad, se cumple que:

$$\#(E^{(1)}(H)) = \sum_{h \in H} \left(\sum_{i=1}^n \#\{\{e \in E_i(H) : e = \{h\}\}\} \right).$$

Por lo tanto, combinando la Identidad (6.2.9) con la igualdad anterior y dividiendo por $\#(H)$, concluimos:

$$\text{avd}(H) = \frac{1}{\#(H)} \sum_{h \in H} \deg(h) = n - \frac{\#(E^{(1)}(H))}{\#(H)}.$$

- *Ítem ii)*: Sea $\sigma : E(H) \rightarrow H$ una orientación. Por lo tanto, tenemos la siguiente partición de $E(H)$:

$$(6.2.10) \quad E(H) := \bigcup_{h \in H} \sigma^{-1}(\{h\}).$$

Además, para cada $h \in H$, consideramos el siguiente subconjunto:

$$O(h, \sigma) := \{e \in E(H) : h \in e, h \neq \sigma(e)\}.$$

Nótese que:

$$O(h, \sigma) := \{e \in E(H) : h \in e\} \setminus \{e \in E(H) : h \in e, h = \sigma(e)\}.$$

Esto es,

$$O(h, \sigma) := \{e \in E(H) : h \in e\} \setminus \sigma^{-1}(\{h\}).$$

Por lo tanto, sus cardinalidades verifican:

$$\text{outdeg}(\sigma, h) = \#(O(h, \sigma)) = n - \#(\sigma^{-1}(\{h\})).$$

Por consiguiente,

$$(6.2.11) \quad \sum_{h \in H} \text{outdeg}(\sigma, h) = n\#(H) - \left(\sum_{h \in H} \#(\sigma^{-1}(\{h\})) \right).$$

Como la Identidad (6.2.10) define una partición de $E(H)$, la última igualdad se convierte en:

$$\sum_{h \in H} \text{outdeg}(\sigma, h) = n\sharp(H) - \sharp(E(H)).$$

Dividiendo ambos lados por $\sharp(H)$, obtenemos la igualdad principal del Ítem *ii*). El resto se deduce al combinar el Ítem *i*) del enunciado con lo que acabamos de demostrar.

- *Ítem iii*): En primer lugar, vamos a demostrar que se cumple la siguiente igualdad:

$$(6.2.12) \quad \text{avd}(H) = \frac{1}{\sharp(H)} \sum_{\substack{e \in E(H) \\ \sharp(e) \geq 2}} \sharp(e).$$

Para probar esta primera igualdad, denotamos por $\chi_e : H \rightarrow \{0, 1\}$ a la función característica definida por la arista $e \in E(H)$, considerada como subconjunto de H . Por lo tanto, tenemos que:

$$\sum_{\substack{e \in E(H) \\ \sharp(e) \geq 2}} \sharp(e) = \sum_{\substack{e \in E(H) \\ \sharp(e) \geq 2}} \sum_{h \in H} \chi_e(h).$$

Por consiguiente, obtenemos:

$$\sum_{\substack{e \in E(H) \\ \sharp(e) \geq 2}} \sharp(e) = \sum_{h \in H} \left(\sum_{\substack{e \in E(H) \\ \sharp(e) \geq 2}} \chi_e(h) \right) = \sum_{h \in H} \text{deg}(h).$$

Dividiendo por $\sharp(H)$, obtenemos la Identidad (6.2.12). A continuación, pasamos a demostrar la siguiente igualdad:

$$(6.2.13) \quad E_H[\text{outdeg}] = \frac{1}{\sharp(H)} \sum_{e \in E(H)} (\sharp(e) - 1).$$

Para probar esta segunda igualdad, consideramos nuevamente la siguiente descomposición de $E(H)$ como unión disjunta (según se describe en la Igualdad (6.2.4)):

$$\sum_{e \in E(H)} (\sharp(e) - 1) = \sum_{i=1}^n \left(\sum_{e \in E_i(H)} (\sharp(e) - 1) \right).$$

Dado $i \in [n]$, tenemos que la siguiente descomposición es también una partición de H :

$$H = \bigsqcup_{e \in E_i(H)} e.$$

Entonces,

$$\sharp(H) = \sum_{e \in E_i(H)} \sharp(e).$$

Por lo tanto, tenemos que:

$$\sum_{e \in E_i(H)} (\sharp(e) - 1) = \left(\sum_{e \in E_i(H)} \sharp(e) \right) - \left(\sum_{e \in E_i(H)} 1 \right) = \sharp(H) - \sharp(E_i(H)).$$

Por consiguiente:

$$\sum_{e \in E(H)} (\sharp(e) - 1) = \left(\sum_{i=1}^n \sharp(H) \right) - \left(\sum_{i=1}^n \sharp(E_i(H)) \right) = n\sharp(H) - \sharp(E(H)).$$

Dividiendo por $\sharp(H)$ en ambos lados y aplicando el Ítem *ii*) del enunciado, concluimos la demostración de la Identidad (6.2.13). Finalmente, demostramos el Ítem *iii*) usando las Identidades (6.2.12) y (6.2.13). Tenemos que:

$$E_H[\text{outdeg}] = \frac{1}{\sharp(H)} \sum_{e \in E(H)} (\sharp(e) - 1) = \frac{1}{\sharp(H)} \sum_{\substack{e \in E(H) \\ \sharp(e) \geq 2}} (\sharp(e) - 1).$$

Por lo tanto:

$$E_H[\text{outdeg}] \leq \frac{1}{\#(H)} \sum_{\substack{e \in E(H) \\ \#(e) \geq 2}} \#(e) = \text{avd}(H).$$

Para la segunda desigualdad, nótese que:

$$E_H[\text{outdeg}] = \frac{1}{\#(H)} \sum_{\substack{e \in E(H) \\ \#(e) \geq 2}} (\#(e) - 1).$$

Pero, para cada arista $e \in E(H)$ tal que $\#(e) \geq 2$, tenemos que $\#(e) - 1 \geq \frac{\#(e)}{2}$. Por lo tanto, obtenemos:

$$E_H[\text{outdeg}] \geq \frac{1}{\#(H)} \sum_{\substack{e \in E(H) \\ \#(e) \geq 2}} \frac{\#(e)}{2} = \frac{\text{avd}(H)}{2},$$

lo que implica la segunda desigualdad del Ítem *iii*).

- *Ítem iv*) : A partir de los Ítems *i*) y *ii*) anteriores, se deduce inmediatamente que:

$$(6.2.14) \quad \text{avd}(H) = E_H[\text{outdeg}] + \frac{\#(E(H)) - \#(E^{(1)}(H))}{\#(H)}.$$

Aplicando ahora la segunda desigualdad del Ítem *iii*), concluimos:

$$E_H[\text{outdeg}] + \frac{\#(E(H)) - \#(E^{(1)}(H))}{\#(H)} \leq 2E_H[\text{outdeg}].$$

- *Ítem v*) : En primer lugar, conviene observar la relación existente entre los conjuntos de aristas $E(H)$ y $E(G)$. Dada una arista $e := e_{f,i} \times \{i\} \in E(H)$, donde $e_{f,i}$ es como en la Identidad (6.2.2), denotamos por $e \cap G := (e_{f,i} \cap G) \times \{i\} \in E(G)$, siempre que $f \in \pi_i(G) \subseteq \pi_i(H)$ (o, equivalentemente, siempre que $e_{f,i} \cap G \neq \emptyset$). Entonces, tenemos que:

Dado $e' \in E(G)$, existe una única arista $e \in E(H)$ tal que $e' = e \cap G \neq \emptyset$.

Se trata de una consecuencia inmediata de la construcción de los hiper-grafos de una inclusión: dada $e' \in E(G)$, existe $i \in [n]$, $f \in \pi_i(G) \subseteq \mathcal{Y}^{n-1}$ tal que $e' := e'_{f,i} \times \{i\}$, donde:

$$e'_{f,i} := \{g \in G : g|_{[n] \setminus \{i\}} = f\}.$$

Por lo tanto, es evidente que existe un $e \in E(H)$ tal que $e' = e \cap G$ y esta arista $e \in E(H)$ solo puede ser $e := e_{f,i} \times \{i\} \in E(H)$.

Por consiguiente, estamos en condiciones de definir $\sigma' : E(G) \rightarrow G$ en términos de σ . Distinguiamos los siguientes casos:

- Si $e' \in E(G)$ es tal que existe $e \in E(H)$ con $e' := e \cap G$ y $e \in \sigma^{-1}(G) \subseteq E(H)$, entonces definimos:

$$\sigma'(e') := \sigma(e).$$

Está bien definida ya que e es única con la propiedad $e' := e \cap G$.

- En otro caso, se cumple lo siguiente: $e' \in E(G)$ es tal que existe una única $e \in E(H) \setminus \sigma^{-1}(G)$ y $e' := e \cap G$. En este caso, $e = e_{f,i} \times \{i\}$, $\sigma(e) \notin G$ y $e' := (e_{f,i} \cap G) \times \{i\} \in E(G)$. Pero, entonces, $e_{f,i} \cap G \neq \emptyset$ y, por lo tanto, podemos tomar un punto $g_e \in e_{f,i} \cap G$. Como e es única para cada e' dado, podemos definir:

$$\sigma'(e') := g_e.$$

Finalmente, tenemos las notaciones necesarias para expresar nuestro argumento. Supongamos $g \in G$. Definimos:

$$\begin{aligned} O(\sigma, g) &:= \{e \in E(H) : g \in e, g \neq \sigma(e)\}, \\ O(\sigma', g) &:= \{e' \in E(G) : g \in e', g \neq \sigma'(e')\}. \end{aligned}$$

Por lo tanto, tenemos la siguiente descomposición:

$$O(\sigma', g) = O_1(\sigma', g) \sqcup O_2(\sigma', g),$$

donde:

$$\begin{aligned} O_1(\sigma', g) &= \{e' \in E(G) : \exists e \in \sigma^{-1}(G), g \in e' = e \cap G, g \neq \sigma'(e') = \sigma(e)\}, \\ O_2(\sigma', g) &= \{e' \in E(G) : \exists e \in E(H) \setminus \sigma^{-1}(G), g \in e' = e \cap G, g \neq g_e\}. \end{aligned}$$

Por consiguiente, tenemos la siguiente aplicación inyectiva:

$$\varphi : O(\sigma', g) \longrightarrow O(\sigma, g),$$

dada por:

$$\varphi(e') := e,$$

donde $e' := e \cap G$. Ya hemos probado que e es única para cada e' y φ es claramente inyectiva por construcción. Únicamente nos queda por ver que está bien definida (i.e. $\varphi(O(\sigma', g)) \subseteq O(\sigma, g)$):

- Si $e' \in O_1(\sigma, g)$, entonces la única arista $e \in E(H)$ tal que $e' = e \cap G$, satisface $\sigma(e) \neq g$ y, por lo tanto, $\varphi(e) \in O(\sigma, g)$.
- Si $e' \in O_2(\sigma, g)$, entonces, obviamente, la única arista $e \in E(H)$ tal que $e' = e \cap G$, verifica que $\sigma(e) \notin G$ y, en particular, $\sigma(e) \neq g$.

En definitiva, tenemos que:

$$\text{outdeg}(\sigma', g) = \sharp(O(\sigma', g)) \leq \sharp(O(\sigma, g)) = \text{outdeg}(\sigma, g),$$

y se deduce el Ítem *v*).

La dicotomía final se sigue de forma inmediata de los Ítems *i*) y *ii*) ya que $E_H[\text{outdeg}] = \text{avd}(H)$ si y solo si $E(H) = E^{(1)}(H)$ y, en este caso, $\sharp(E^{(1)}(H)) = n\sharp(H)$ (o, equivalentemente, $\text{avd}(H) = 0$). \square

OBSERVACIÓN 6.2.6 (Caso $n = 1$). En el caso $n = 1$ algunas de las nociones anteriores carecen de un significado apropiado. Si $n = 1$, no existen proyecciones que “olviden” la 1-ésima coordenada. Además, si $n = 1$ y $\sharp(H) = 1$, tenemos un cubo “canónico” de dimensión 1 y tanto el grado promedio como el grado de salida promedio son 0, mientras que la densidad del hiper-grafo es 1.

OBSERVACIÓN 6.2.7. Nótese que, en la demostración de la Proposición 6.2.5, hemos usado la Identidad (6.2.13). Esta identidad implica que el grado de salida promedio $E_H[\text{outdeg}]$ coincide con la cantidad avd' , denominada “*shifting average degree*”, utilizada implícitamente en una demostración en [BCDMY, 22]. Sin embargo, en dicho trabajo no se realiza ningún estudio ni análisis específico sobre ella.

OBSERVACIÓN 6.2.8 (Los “grados promedio” dependen de la distribución uniforme sobre H). Los valores del grado promedio y el grado de salida promedio dependen fuertemente de la distribución uniforme escogida en H . A continuación, incluimos un par de ejemplos elementales que muestran esta dependencia.

- *Ejemplo 1:* En este caso, $\text{avd}(H)$ no cambia al modificar la distribución de probabilidad definida sobre H . Por otro lado, el grado de salida promedio de las orientaciones varía cuando modificamos la distribución de probabilidad. Además, la independencia del grado de salida promedio no se mantiene cuando cambiamos la orientación. El ejemplo consiste en un cubo de dimensión 2 con $p = 2$. Es decir, consideramos:

$$H := \{(0, 0), (1, 0), (0, 1), (1, 1)\} = \{0, 1\}^2.$$

Las aristas en $E(H)$ son cuatro subconjuntos, todos ellos de cardinalidad 2:

$$E(H) = \{e_1, e_2, e_3, e_4\},$$

donde:

$$e_1 = \{(0, 1), (1, 1)\}, e_2 = \{(1, 0), (0, 0)\}, e_3 = \{(0, 1), (0, 0)\}, e_4 = \{(1, 0), (1, 1)\}.$$

El grado de cada $h \in H$ es 2 ya que cada vértice está exactamente en dos aristas del hiper-grafo de una inclusión. Por consiguiente, el promedio de los grados, con respecto a cualquier distribución de probabilidad sobre H , es igual a $\text{avd}(H) = 2$.

Por otro lado, consideramos las orientaciones $\sigma_1, \sigma_2 : E(H) \longrightarrow H$, dadas por las siguientes tablas:

$$\begin{array}{ccc} \sigma_1 : E(H) & \longrightarrow & H \\ e_1 & \longmapsto & (0, 1) \\ e_2 & \longmapsto & (0, 0) \\ e_3 & \longmapsto & (0, 0) \\ e_4 & \longmapsto & (1, 0) \end{array} \quad \begin{array}{ccc} \sigma_2 : E(H) & \longrightarrow & H \\ e_1 & \longmapsto & (1, 1) \\ e_2 & \longmapsto & (1, 0) \\ e_3 & \longmapsto & (0, 1) \\ e_4 & \longmapsto & (1, 1) \end{array}$$

Obviamente, tenemos que:

$$\begin{aligned} \text{outdeg}(\sigma_1, (0, 1)) &= \text{outdeg}(\sigma_1, (1, 0)) = 1, & \text{outdeg}(\sigma_1, (0, 0)) &= 0, & \text{outdeg}(\sigma_1, (1, 1)) &= 2, \\ \text{outdeg}(\sigma_2, (0, 1)) &= \text{outdeg}(\sigma_2, (1, 0)) = 1, & \text{outdeg}(\sigma_2, (0, 0)) &= 2, & \text{outdeg}(\sigma_2, (1, 1)) &= 0. \end{aligned}$$

Si consideramos la distribución uniforme sobre H , obtenemos:

$$E_H[\text{outdeg}(\sigma_1)] = E_H[\text{outdeg}(\sigma_2)] = E_H[\text{outdeg}] = 1.$$

Sin embargo, si elegimos cualquier otra distribución de probabilidad que difiera en $(0, 0)$ y $(1, 1)$, como por ejemplo:

$$\nu_{(0,0)} = \frac{1}{8}, \nu_{(1,0)} = \nu_{(0,1)} = \frac{1}{4}, \nu_{(1,1)} = \frac{3}{8},$$

entonces, tenemos que:

$$\frac{5}{4} = \sum_{h \in H} \text{outdeg}(\sigma_1, h) \cdot \nu_h \neq \sum_{h \in H} \text{outdeg}(\sigma_2, h) \cdot \nu_h = \frac{3}{4}.$$

Por lo tanto, el grado de salida promedio varía cuando modificamos la distribución de probabilidad y, en general, no es independiente de la orientación escogida.

- *Ejemplo 2:* El grado promedio tampoco es un invariante de la distribución de probabilidad elegida sobre H , como se muestra en el siguiente ejemplo. Supongamos que H viene dado por la siguiente igualdad:

$$H := \{(0, 0), (1, 0), (0, 1)\},$$

y, por lo tanto, sus aristas son:

$$E(H) = \{e_1, e_2\},$$

donde:

$$e_1 = \{(0, 0), (0, 1)\}, e_2 = \{(0, 0), (1, 0)\}.$$

Los grados de los vértices son:

$$\deg((0, 0)) = 2, \deg((0, 1)) = \deg((1, 0)) = 1.$$

Suponiendo la distribución uniforme sobre H , tenemos que:

$$\text{avd}(H) = \frac{4}{3}.$$

Si consideramos la distribución de probabilidad sobre H dada por:

$$\nu_{(0,0)} = \frac{1}{8}, \nu_{(1,0)} = \frac{3}{8}, \nu_{(0,1)} = \frac{4}{8},$$

tenemos que el grado promedio con respecto a esta distribución de probabilidad verifica:

$$\sum_{h \in H} \deg(h) \cdot \nu_h = \frac{9}{8} \neq \text{avd}(H).$$

En conclusión, ni el grado de salida promedio ni el grado promedio son independientes de la distribución de probabilidad definida sobre H . Además, la independencia de la orientación del grado de salida promedio no se preserva si cambiamos la distribución de probabilidad sobre H .

A continuación, mostramos que algunas de las propiedades demostradas en la Proposición 6.2.5 son independientes de la distribución de probabilidad elegida sobre H .

LEMA 6.2.9. *Con las notaciones anteriores, sea $\sigma : E(H) \rightarrow H$ una orientación y $\{\nu_h : h \in H\}$ cualquier distribución de probabilidad definida sobre H . Definimos:*

$$E_H^\nu[\text{outdeg}(\sigma)] := \sum_{h \in H} \text{outdeg}(\sigma, h) \cdot \nu_h,$$

$$E_H^\nu[\#\sigma^{-1}] := \sum_{h \in H} \#(\sigma^{-1}(\{h\})) \cdot \nu_h.$$

Entonces, tenemos que:

$$E_H^\nu[\text{outdeg}(\sigma)] + E_H^\nu[\#\sigma^{-1}] = n.$$

DEMOSTRACIÓN. Para cada $h \in H$, consideramos el siguiente conjunto:

$$O(h, \sigma) := \{e \in E(H) : h \in e, h \neq \sigma(e)\}.$$

Nótese que:

$$O(h, \sigma) := \{e \in E(H) : h \in e\} \setminus \{e \in E(H) : h \in e, h = \sigma(e)\}.$$

Esto es,

$$O(h, \sigma) := \{e \in E(H) : h \in e\} \setminus \sigma^{-1}(\{h\}).$$

Entonces, en términos de cardinalidades, tenemos que:

$$\text{outdeg}(\sigma, h) = \#(O(h, \sigma)) = n - \#(\sigma^{-1}(\{h\})).$$

Por lo tanto,

$$\begin{aligned} E_H^\nu[\text{outdeg}(\sigma)] &= \sum_{h \in H} \text{outdeg}(\sigma, h) \cdot \nu_h = \sum_{h \in H} (n - \sharp(\sigma^{-1}(\{h\}))) \cdot \nu_h = \\ &= \sum_{h \in H} n \cdot \nu_h - \sum_{h \in H} \sharp(\sigma^{-1}(\{h\})) \cdot \nu_h = n \sum_{h \in H} \nu_h - E_H^\nu[\sharp\sigma^{-1}]. \end{aligned}$$

Como

$$\sum_{h \in H} \nu_h = 1,$$

concluimos:

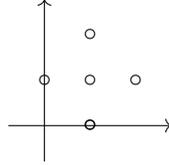
$$E_H^\nu[\text{outdeg}(\sigma)] = n - E_H^\nu[\sharp\sigma^{-1}].$$

□

EJEMPLO 6.2.10. El lema anterior prueba que el Ítem *ii*) de la Proposición 6.2.5 (donde se considera la distribución uniforme en H) es cierto para cualquier distribución de probabilidad sobre H . Sin embargo, las cantidades $E_H^\nu[\text{outdeg}(\sigma)]$ y $E_H^\nu[\sharp\sigma^{-1}]$ dependen de la orientación escogida como muestra el siguiente ejemplo. Consideramos $\mathcal{Y} = \{0, 1, 2\}$ y $H \subseteq \mathcal{Y}^{[2]}$ dado por las siguiente igualdad:

$$H = \{(1, 0), (0, 1), (1, 1), (2, 1), (1, 2)\},$$

cuya imagen es:



El hiper-grafo de una inclusión asociado tiene 6 aristas:

$$E(H) = \{e_1, e_2, e_3, e_4, e_5, e_6\},$$

donde:

$$\begin{aligned} e_1 &= \{(1, 0)\}, \quad e_2 = \{(0, 1), (1, 1), (2, 1)\}, \quad e_3 = \{(1, 2)\}, \\ e_4 &= \{(0, 1)\}, \quad e_5 = \{(1, 0), (1, 1), (1, 2)\}, \quad e_6 = \{(2, 1)\}. \end{aligned}$$

Consideramos ahora las siguientes orientaciones de $G(H)$:

$$\begin{array}{ccc} \varphi : E(H) & \longrightarrow & H & \quad \psi : E(H) & \longrightarrow & H \\ e_1 & \longmapsto & (1, 0) & e_1 & \longmapsto & (1, 0) \\ e_2 & \longmapsto & (1, 1) & e_2 & \longmapsto & (2, 1) \\ e_3 & \longmapsto & (1, 2) & e_3 & \longmapsto & (1, 2) \\ e_4 & \longmapsto & (0, 1) & e_4 & \longmapsto & (0, 1) \\ e_5 & \longmapsto & (1, 1) & e_5 & \longmapsto & (1, 0) \\ e_6 & \longmapsto & (2, 1) & e_6 & \longmapsto & (2, 1) \end{array}$$

A continuación, calculamos el grado de salida de cada vértice respecto a cada orientación:

$$\begin{aligned} \text{outdeg}(\varphi, (1, 0)) &= 1, \quad \text{outdeg}(\varphi, (0, 1)) = 1, \quad \text{outdeg}(\varphi, (1, 1)) = 0, \quad \text{outdeg}(\varphi, (2, 1)) = 1, \quad \text{outdeg}(\varphi, (1, 2)) = 1, \\ \text{outdeg}(\psi, (1, 0)) &= 0, \quad \text{outdeg}(\psi, (0, 1)) = 1, \quad \text{outdeg}(\psi, (1, 1)) = 2, \quad \text{outdeg}(\psi, (2, 1)) = 0, \quad \text{outdeg}(\psi, (1, 2)) = 1. \end{aligned}$$

Calculamos también el cardinal de las fibras de cada vértice respecto a cada orientación:

$$\begin{aligned} \sharp(\varphi^{-1}((1, 0))) &= 1, \quad \sharp(\varphi^{-1}((0, 1))) = 1, \quad \sharp(\varphi^{-1}((1, 1))) = 2, \quad \sharp(\varphi^{-1}((2, 1))) = 1, \quad \sharp(\varphi^{-1}((1, 2))) = 1, \\ \sharp(\psi^{-1}((1, 0))) &= 2, \quad \sharp(\psi^{-1}((0, 1))) = 1, \quad \sharp(\psi^{-1}((1, 1))) = 0, \quad \sharp(\psi^{-1}((2, 1))) = 2, \quad \sharp(\psi^{-1}((1, 2))) = 1. \end{aligned}$$

Consideramos la distribución de probabilidad $\{\nu_h : h \in H\}$ sobre H dada por:

$$\nu_{(1,0)} = \frac{1}{7}, \quad \nu_{(0,1)} = \frac{2}{7}, \quad \nu_{(1,1)} = \frac{2}{7}, \quad \nu_{(2,1)} = \frac{1}{7}, \quad \nu_{(1,2)} = \frac{1}{7},$$

Obtenemos:

$$E_H^\nu[\text{outdeg}(\varphi)] = \sum_{h \in H} \text{outdeg}(\varphi, h) \cdot \nu_h = 1 \cdot \frac{1}{7} + 1 \cdot \frac{2}{7} + 0 \cdot \frac{2}{7} + 1 \cdot \frac{1}{7} + 1 \cdot \frac{1}{7} = \frac{5}{7},$$

$$\begin{aligned}
E_H^\nu[\#\varphi^{-1}] &= \sum_{h \in H} \#(\varphi^{-1}(\{h\})) \cdot \nu_h = 1 \cdot \frac{1}{7} + 1 \cdot \frac{2}{7} + 2 \cdot \frac{2}{7} + 1 \cdot \frac{1}{7} + 1 \cdot \frac{1}{7} = \frac{9}{7}, \\
E_H^\nu[\text{outdeg}(\psi)] &= \sum_{h \in H} \text{outdeg}(\psi, h) \cdot \nu_h = 0 \cdot \frac{1}{7} + 1 \cdot \frac{2}{7} + 2 \cdot \frac{2}{7} + 0 \cdot \frac{1}{7} + 1 \cdot \frac{1}{7} = 1, \\
E_H^\nu[\#\psi^{-1}] &= \sum_{h \in H} \#(\psi^{-1}(\{h\})) \cdot \nu_h = 2 \cdot \frac{1}{7} + 1 \cdot \frac{2}{7} + 0 \cdot \frac{2}{7} + 2 \cdot \frac{1}{7} + 1 \cdot \frac{1}{7} = 1.
\end{aligned}$$

Finalmente, nótese que:

$$\begin{aligned}
\frac{5}{7} &= E_H^\nu[\text{outdeg}(\varphi)] \neq E_H^\nu[\text{outdeg}(\psi)] = 1, \\
\frac{9}{7} &= E_H^\nu[\#\varphi^{-1}] \neq E_H^\nu[\#\psi^{-1}] = 1, \\
E_H^\nu[\text{outdeg}(\varphi)] + E_H^\nu[\#\varphi^{-1}] &= \frac{5}{7} + \frac{9}{7} = 2, \\
E_H^\nu[\text{outdeg}(\psi)] + E_H^\nu[\#\psi^{-1}] &= 1 + 1 = 2.
\end{aligned}$$

LEMA 6.2.11. *Con las notaciones precedentes, para cada $h \in H$ y cada orientación $\sigma : E(H) \rightarrow H$, tenemos que:*

$$\text{outdeg}(h, \sigma) \leq \text{deg}(h).$$

DEMOSTRACIÓN. Comenzamos definiendo el siguiente conjunto:

$$A = \{e \in E(H) : h \in e, |e| \geq 2\}.$$

Obviamente, $\#(A) = \text{deg}(h)$. Por otro lado, nótese que podemos escribir el conjunto A como unión disjunta de los siguiente subconjuntos:

$$A_1 = \{e \in A : \sigma(e) = h\} \text{ y } A_2 = \{e \in A : \sigma(e) \neq h\}.$$

Tenemos que:

$$\#(A_2) = \text{outdeg}(\sigma, h).$$

Por lo tanto,

$$\text{outdeg}(\sigma, h) = \#(A_2) \leq \#(A_1) + \#(A_2) = \text{deg}(h).$$

□

OBSERVACIÓN 6.2.12. Del lema anterior, se deduce que la primera desigualdad del Ítem *iii*) de la Proposición 6.2.13 es cierta para cualquier distribución de probabilidad $\{\nu_h : h \in H\}$ definida sobre H . Es decir, se cumple que:

$$\sum_{h \in H} \text{outdeg}(h, \sigma) \cdot \nu_h \leq \sum_{h \in H} \text{deg}(h) \cdot \nu_h.$$

EJEMPLO 6.2.13. La segunda desigualdad del Ítem *iii*) de la Proposición 6.2.13 no es cierta si cambiamos la distribución de probabilidad en H como se muestra en el siguiente ejemplo. Sea $\mathcal{Y} := \{0, 1, 2\}$. Consideramos la clase de conceptos $H \subseteq \mathcal{Y}^{[2]}$ dada por la siguiente igualdad:

$$H := \{(1, 0), (0, 1), (1, 1), (2, 1), (1, 2)\}.$$

El hiper-grafo de una inclusión asociado tiene 6 aristas:

$$E(H) = \{e_1, e_2, e_3, e_4, e_5, e_6\},$$

donde:

$$\begin{aligned}
e_1 &= \{(1, 0)\}, \quad e_2 = \{(0, 1), (1, 1), (2, 1)\}, \quad e_3 = \{(1, 2)\}, \\
e_4 &= \{(0, 1)\}, \quad e_5 = \{(1, 0), (1, 1), (1, 2)\}, \quad e_6 = \{(2, 1)\}.
\end{aligned}$$

Los grados de los vértices son:

$$\text{deg}((1, 0)) = 1, \quad \text{deg}((0, 1)) = 1, \quad \text{deg}((1, 1)) = 2, \quad \text{deg}((2, 1)) = 1, \quad \text{deg}((1, 2)) = 1.$$

Consideramos la siguiente orientación de $G(H)$:

$$\begin{array}{rcl}
\sigma : E(H) & \longrightarrow & H \\
e_1 & \longmapsto & (1, 0) \\
e_2 & \longmapsto & (2, 1) \\
e_3 & \longmapsto & (1, 2) \\
e_4 & \longmapsto & (0, 1) \\
e_5 & \longmapsto & (1, 0) \\
e_6 & \longmapsto & (2, 1)
\end{array}$$

Obviamente, tenemos que:

$$\text{outdeg}(\sigma, (1, 0)) = 0, \text{outdeg}(\sigma, (0, 1)) = 1, \text{outdeg}(\sigma, (1, 1)) = 2, \text{outdeg}(\sigma, (2, 1)) = 0, \text{outdeg}(\sigma, (1, 2)) = 1.$$

Tomando la siguiente distribución de probabilidad sobre H :

$$\nu_{(1,0)} = \frac{3}{9}, \nu_{(0,1)} = \frac{1}{9}, \nu_{(1,1)} = \frac{1}{9}, \nu_{(2,1)} = \frac{3}{9}, \nu_{(1,2)} = \frac{1}{9},$$

obtenemos:

$$\sum_{h \in H} \text{deg}(h) \cdot \nu_h = \frac{10}{9} \text{ y } \sum_{h \in H} \text{outdeg}(h, \sigma) \cdot \nu_h = \frac{4}{9}.$$

Por lo tanto, concluimos:

$$\frac{8}{9} = 2 \sum_{h \in H} \text{outdeg}(h, \sigma) \cdot \nu_h < \sum_{h \in H} \text{deg}(h) \cdot \nu_h = \frac{10}{9},$$

lo que demuestra que la segunda desigualdad del Ítem *iii*) de la Proposición 6.2.13 varía si cambiamos la distribución de probabilidad definida sobre H .

Tras esta breve discusión sobre cómo varían las definiciones y los resultados al modificar la distribución de probabilidad sobre H , retomamos nuestra exposición considerando, de nuevo, la distribución de probabilidad uniforme en la clase de conceptos H .

OBSERVACIÓN 6.2.14. En el caso $p = 2$, tenemos que:

$$(6.2.15) \quad \text{avd}(H) = E_H[\text{outdeg}] + \frac{\#(\text{Haussl}(H))}{2\#(H)},$$

y

$$\frac{\#(\text{Haussl}(H))}{2\#(H)} \leq E_H[\text{outdeg}] \leq \text{outdeg}_{\max}(\sigma),$$

donde $\text{Haussl}(H)$ es la estructura de grafo definida en [Ha, 95]. En particular, en el caso $p = 2$, la diferencia entre el grado promedio de H y el grado de salida promedio de H es un medio de la densidad del grafo considerado por D. Haussler en [Ha, 95].

El siguiente lema muestra cómo varían el grado promedio y el grado de salida promedio al eliminar un elemento $h \in H$. En particular, demostramos que ninguno de estos invariantes es monótono con respecto a la inclusión de conjuntos.

LEMA 6.2.15. Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita tal que $\#(H) \geq 2$ y $n \geq 2$. Sea $h \in H$ un elemento de H . Definimos $H' := H \setminus \{h\}$ como la clase obtenida al eliminar el elemento h de H . Denotamos por $\text{deg}(g)$ al grado de un elemento $g \in H'$ en el hiper-grafo $G(H)$ y por $\text{deg}'(g)$ al grado correspondiente en el hiper-grafo $G(H')$. Entonces, tenemos que:

$$(6.2.16) \quad \text{deg}'(g) \leq \text{deg}(g), \forall g \in H'.$$

DEMOSTRACIÓN. Nótese que las aristas que contienen a $g \in H'$ en el hiper-grafo de una inclusión $G(H')$ son:

$$e'_{\pi_i(g), i} := \{t \in H' : t|_{[n] \setminus \{i\}} = \pi_i(g)\} \neq \emptyset.$$

Denotamos por $e_{\pi_i(g), i}$ a las aristas que contienen a $g \in H' \subseteq H$ en el hiper-grafo de una inclusión $G(H)$, i.e.

$$e_{\pi_i(g), i} := \{t \in H : t|_{[n] \setminus \{i\}} = \pi_i(g)\} \neq \emptyset.$$

Entonces, tenemos tres casos posibles:

- *Caso 1*: Si $h \in e_{\pi_i(g), i}$, y $e'_{\pi_i(g), i} \neq \emptyset$, entonces:

$$e'_{\pi_i(g), i} = e_{\pi_i(g), i} \setminus \{h\},$$

y, por lo tanto,

$$\#(e'_{\pi_i(g), i}) = \#(e_{\pi_i(g), i}) - 1.$$

En este caso, $\#(e_{\pi_i(g), i}) \geq 2$. Nótese también que, en este caso, puede ocurrir que $\#(e'_{\pi_i(g), i}) = 1$ dado que $\#(e_{\pi_i(g), i}) = 2$ o, equivalentemente, que $e_{\pi_i(g), i} = \{g, h\}$.

- *Caso 2:* Si $h \notin e_{\pi_i(g),i}$, y $\#(e'_{\pi_i(g),i}) \geq 2$, entonces:

$$e'_{\pi_i(g),i} = e_{\pi_i(g),i},$$

y, por lo tanto,

$$\#(e'_{\pi_i(g),i}) = \#(e_{\pi_i(g),i}).$$

En este caso, se cumple que:

$$\#(e_{\pi_i(g),i}) \geq 2.$$

- *Caso 3:* Si $h \notin e_{\pi_i(g),i}$, y $\#(e'_{\pi_i(g),i}) = 1$, entonces:

$$e'_{\pi_i(g),i} = e_{\pi_i(g),i} = \{g\},$$

y, por lo tanto,

$$\#(e'_{\pi_i(g),i}) = \#(e_{\pi_i(g),i}) = 1.$$

Por consiguiente, concluimos que únicamente las aristas de los dos primeros casos cuentan para $\deg'(g)$ y, por lo tanto, tenemos que:

$$\deg'(g) = \#(\{e \setminus \{h\} \in E(H') : e \in E(H), g \in e, h \in e, \#(e \setminus \{h\}) \geq 2\}) + \#(\{e \in E(H') : e \in E(H), g \in e, h \notin e, \#(e) \geq 2\}).$$

Esto es, se verifica la siguiente desigualdad:

$$\deg'(g) \leq \#(\{e \in E(H) : g \in e, h \in e, \#(e) \geq 2\}) + \#(\{e \in E(H) : g \in e, h \notin e, \#(e) \geq 2\}).$$

De lo anterior, se deduce que:

$$\deg'(g) \leq \deg(g),$$

lo que prueba la Desigualdad (6.2.16). \square

PROPOSICIÓN 6.2.16 (Los “grados promedio” no son funciones monótonas respecto a la inclusión de conjuntos). Con las notaciones precedentes, sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita, $\#(H) \geq 2$ y $h \in H$. Definimos $H' := H \setminus \{h\}$ y sea $G(H')$ el hiper-grafo de una inclusión asociado a H' . Tenemos que:

- i) Si $\deg(h) \geq \text{avd}(H)$, entonces:

$$\text{avd}(H') \leq \text{avd}(H).$$

- ii) Sin embargo, el grado promedio (avd) no es una función monótona con respecto a la inclusión de conjuntos. Supongamos que existe $h \in H$ tal que $\deg(h) + e_2(h, H) \leq \text{avd}(H)$, donde:

$$e_2(h, H) := \#(\{e \in E(H) : \#(e) = 2, h \in e\}).$$

Entonces, si $H' := H \setminus \{h\}$, tenemos que:

$$\text{avd}(H) \leq \text{avd}(H').$$

Además, $\deg(h) + e_2(h, H) < \text{avd}(H)$ implica $\text{avd}(H) < \text{avd}(H')$.

- iii) Si $\deg(h) \geq E_H[\text{outdeg}]$, entonces $E_{H'}[\text{outdeg}] \leq E_H[\text{outdeg}]$.
- iv) Si, por el contrario, $\deg(h) \leq E_H[\text{outdeg}]$, entonces $E_{H'}[\text{outdeg}] \geq E_H[\text{outdeg}]$.
- v) $E_H[\text{outdeg}]$ no es una función monótona con respecto a la inclusión de conjuntos. Además, si $\deg(h) < E_H[\text{outdeg}]$, entonces $E_H[\text{outdeg}] < E_{H'}[\text{outdeg}]$.

DEMOSTRACIÓN. Demostramos cada ítem por separado:

- *Ítem i):* Por la definición de $\text{avd}(H)$, sabemos que:

$$\text{avd}(H) = \frac{1}{\#(H)} \left(\sum_{g \in H} \deg(g) \right).$$

Por lo tanto, tenemos que:

$$\#(H) \text{avd}(H) = \left(\sum_{g \in H'} \deg(g) \right) + \deg(h).$$

Por consiguiente, obtenemos:

$$\#(H) \text{avd}(H) = (\#(H) - 1) \frac{\left(\sum_{g \in H'} \deg(g) \right)}{(\#(H) - 1)} + \deg(h).$$

Modificando la expresión anterior, llegamos a que:

$$(\#(H) - 1) \left(\text{avd}(H) - \frac{\left(\sum_{g \in H'} \text{deg}(g) \right)}{(\#(H) - 1)} \right) = \text{deg}(h) - \text{avd}(H).$$

Como, por hipótesis, el lado derecho de esta igualdad es positivo (i.e. $\text{deg}(h) - \text{avd}(H) \geq 0$), obtenemos:

$$\text{avd}(H) - \frac{\left(\sum_{g \in H'} \text{deg}(g) \right)}{(\#(H) - 1)} \geq 0,$$

o, equivalentemente:

$$\text{avd}(H) \geq \frac{\left(\sum_{g \in H'} \text{deg}(g) \right)}{(\#(H) - 1)}.$$

Por el Lema 6.2.15, sabemos que $\text{deg}'(g) \leq \text{deg}(g)$, para todo $g \in H'$ y, por lo tanto, concluimos:

$$\text{avd}(H) \geq \frac{\left(\sum_{g \in H'} \text{deg}'(g) \right)}{(\#(H) - 1)} = \text{avd}(H').$$

- *Ítem ii*): Mostramos a continuación un ejemplo en el que avd no es una función monótona. Consideramos $\mathcal{Y} := \{0, 1, 2\}$ y la clase de conceptos $H \subseteq Y^{[2]}$ dada por:

$$H := \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 2)\}.$$

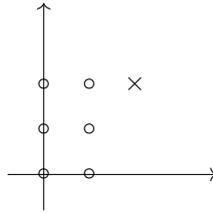
El hiper-grafo de una inclusión $G(H)$ tiene una única arista unaria: $\{(2, 2)\}$. Por lo tanto, aplicando el Ítem *i*) de la Proposición 6.2.5, obtenemos:

$$\text{avd}(H) = 2 - \frac{1}{7} = 13/7 < 2.$$

Si eliminamos el vértice $h := (2, 2) \in H$, obtenemos la clase $H' := H \setminus \{(2, 2)\}$. Nótese que la clase H' es un cubo con 6 vértices (y, por lo tanto, no tiene aristas unarias). Por consiguiente, es evidente que:

$$\text{avd}(H') = 2 > \text{avd}(H),$$

aunque $H' \subseteq H$. La siguiente imagen muestra todos los vértices de H' representados mediante \circ y el elemento $h := (2, 2) \in H$ que ha sido eliminado con \times :



Para el segundo resultado de este ítem, consideramos las siguientes particiones de las aristas unarias de H y H' , respectivamente:

$$E^{(1)}(H) = \{e \in E^{(1)}(H) : h \in e\} \sqcup \{e \in E^{(1)}(H) : h \notin e\},$$

$$E^{(1)}(H') = \{e \setminus \{h\} : e \in E(H), \#(e) = 2, h \in e\} \sqcup \{e \in E^{(1)}(H) : h \notin e\}.$$

Por lo tanto,

$$\#(E^{(1)}(H)) - \#(E^{(1)}(H')) = (n - \text{deg}(h)) - e_2(h, H),$$

ya que:

$$e_2(h, H) = \#\{e \setminus \{h\} : e \in E(H), \#(e) = 2, h \in e\},$$

y

$$n - \text{deg}(h) = \#\{e \in E^{(1)}(H) : h \in e\}.$$

Por consiguiente, obtenemos:

$$(\#(H)) \frac{\#(E^{(1)}(H))}{\#(H)} - (\#(H) - 1) \frac{\#(E^{(1)}(H'))}{\#(H) - 1} = (n - \text{deg}(h)) - e_2(h, H).$$

Y, por lo tanto, concluimos:

$$(\sharp(H) - 1) \left(\frac{\sharp(E^{(1)}(H))}{\sharp(H)} - \frac{\sharp(E^{(1)}(H'))}{\sharp(H) - 1} \right) = (n - \deg(h)) - e_2(h, H) - \frac{\sharp(E^{(1)}(H))}{\sharp(H)}.$$

Gracias al Ítem *i*) de la Proposición 6.2.5, tenemos que:

$$(\sharp(H) - 1) \left(\frac{\sharp(E^{(1)}(H))}{\sharp(H)} - \frac{\sharp(E^{(1)}(H'))}{\sharp(H) - 1} \right) = \text{avd}(H) - (\deg(h)) + e_2(h, H).$$

Como $\sharp(H') = \sharp(H) - 1$, de nuestras hipótesis obtenemos que:

$$\frac{\sharp(E^{(1)}(H))}{\sharp(H)} - \frac{\sharp(E^{(1)}(H'))}{\sharp(H')} \geq 0.$$

De nuevo, por el Ítem *i*) de la Proposición 6.2.5, concluimos:

$$\text{avd}(H) = n - \frac{\sharp(E^{(1)}(H))}{\sharp(H)} \leq n - \frac{\sharp(E^{(1)}(H'))}{\sharp(H')} = \text{avd}(H').$$

- *Ítem iii*): Siguiendo un razonamiento similar al del Lema 6.2.15, demostramos que:

$$(6.2.17) \quad \sharp(E(H)) = \sharp(E(H')) + (n - \deg(h)).$$

La razón es que podemos descomponer $E(H')$ como la siguiente unión disjunta de subconjuntos:

$$\begin{aligned} E(H') &= \{e \setminus \{h\} : e \in E(H), h \in e, \sharp(e) \geq 2\} \\ \sqcup &\quad \{e : e \in E(H), h \notin e\}. \end{aligned}$$

Por otra parte, tenemos la siguiente partición de $E(H)$:

$$\begin{aligned} E(H) &= \{e \in E(H) : h \in e, \sharp(e) \geq 2\} \\ \sqcup &\quad \{e \in E(H) : h \notin e\} \\ \sqcup &\quad \{e \in E(H) : h \in e, \sharp(e) = 1\}. \end{aligned}$$

Por lo tanto,

$$\sharp(E(H)) - \sharp(E(H')) = \sharp(\{e \in E(H) : h \in e, \sharp(e) = 1\}),$$

y se deduce la Identidad (6.2.17). A continuación, nótese que la Identidad (6.2.17) nos proporciona la siguiente igualdad:

$$\sharp(H) \frac{\sharp(E(H))}{\sharp(H)} = (\sharp(H) - 1) \frac{\sharp(E(H'))}{\sharp(H) - 1} + (n - \deg(h)).$$

Por lo tanto, obtenemos:

$$(\sharp(H) - 1) \left(\frac{\sharp(E(H))}{\sharp(H)} - \frac{\sharp(E(H'))}{\sharp(H) - 1} \right) = n - \deg(h) - \frac{\sharp(E(H))}{\sharp(H)}.$$

Aplicando el Ítem *ii*) de la Proposición 6.2.5, tenemos que:

$$(\sharp(H) - 1) \left(\frac{\sharp(E(H))}{\sharp(H)} - \frac{\sharp(E(H'))}{\sharp(H')} \right) = E_H[\text{outdeg}] - \deg(h).$$

Como, por hipótesis, $E_H[\text{outdeg}] \leq \deg(h)$, el lado derecho de esta igualdad es negativo y, por lo tanto, concluimos:

$$\frac{\sharp(E(H))}{\sharp(H)} \leq \frac{\sharp(E(H'))}{\sharp(H')}.$$

De nuevo, por el Ítem *ii*) de la Proposición 6.2.5, concluimos:

$$E_{H'}[\text{outdeg}] = n - \frac{\sharp(E(H'))}{\sharp(H')} \leq n - \frac{\sharp(E(H))}{\sharp(H)} = E_H[\text{outdeg}].$$

- *Ítem iv*): Empleando un argumento similar al usado en el Ítem *iii*), concluimos:

$$(\sharp(H) - 1) \left(\frac{\sharp(E(H))}{\sharp(H)} - \frac{\sharp(E(H'))}{\sharp(H) - 1} \right) = n - \deg(h) - \frac{\sharp(E(H))}{\sharp(H)}.$$

Por lo tanto, si $\deg(h) \leq E_H[\text{outdeg}]$, entonces el lado derecho de la expresión anterior es positivo. Esto implica que:

$$\frac{\sharp(E(H))}{\sharp(H)} \geq \frac{\sharp(E(H'))}{\sharp(H')},$$

y concluimos la demostración del segundo resultado del Ítem *iv*):

$$E_{H'}[\text{outdeg}] = n - \frac{\sharp(E(H'))}{\sharp(H')} \geq n - \frac{\sharp(E(H))}{\sharp(H)} = E_H[\text{outdeg}].$$

- *Ítem v*): Para el último ítem, el mismo ejemplo usado en el Ítem *ii*) para probar que $\text{avd}(H)$ no es una función monótona con respecto a la inclusión, muestra que el grado de salida promedio tampoco es una función monótona con respecto a la inclusión de conjuntos. Esto es, tomando $\mathcal{Y} := \{0, 1, 2\}$, la clase de conceptos $H \subseteq \mathcal{Y}^{[2]}$ dada por:

$$H := \{(0, 0), (0, 1), (0, 2), (1, 0), (1, 1), (1, 2), (2, 2)\},$$

$h = (2, 2)$, con $\text{deg}(h) = 1$, y $H' := H \setminus \{h\}$ obtenemos, aplicando el Ítem *ii*) de la Proposición 6.2.5:

$$E_H[\text{outdeg}] = 2 - 6/7 < 2 - 5/6 = E_{H'}[\text{outdeg}].$$

De la expresión anterior, se deduce que el grado de salida promedio no es una función monótona respecto a la inclusión de conjuntos. □

La proposición anterior indica que no podemos emplear argumentos inductivos basados únicamente en la inclusión de conjuntos para estudiar propiedades del grado promedio o del grado de salida promedio, lo que dificulta el análisis de estas nociones.

6.2.2. Grado promedio, grado de salida promedio y conjuntos cerrados hacia abajo. Para cada $h := (h_1, \dots, h_n) \in \mathbb{N}^n$, definimos el conjunto:

$$(6.2.18) \quad DS(h) := \{i \in [n] : h_i \geq 1\}.$$

LEMA 6.2.17. *Con las notaciones anteriores, tenemos que:*

- Sean $f, g \in \mathbb{N}^n$. Si $f \preceq g$ con respecto al pre-orden de Dickson, entonces $DS(f) \subseteq DS(g)$ y, por lo tanto, $\sharp(DS(f)) \leq \sharp(DS(g))$.
- Si $H \subseteq \mathcal{Y}^n$ es una clase cerrada hacia abajo y $h \in H$, entonces $\text{deg}(h) \geq \sharp(DS(h))$. La desigualdad anterior se convierte en una igualdad cuando h es un elemento maximal de H con respecto al pre-orden de Dickson.

DEMOSTRACIÓN. El primer ítem se deduce inmediatamente de la definición del pre-orden de Dickson. Para la primera parte del Ítem *ii*), como H es un conjunto cerrado hacia abajo, se cumple que para todo $h \in H$ se verifica la siguiente inclusión:

$$\{e_{\pi_i(h), i} : h(i) \geq 1\} \subseteq \{e \in E(H) : h \in e, \sharp(e) \geq 2\},$$

y, por lo tanto, se tiene que $\text{deg}(h) \geq \sharp(DS(h))$. Si h es un elemento maximal de H con respecto al pre-orden de Dickson, los conjuntos anteriores coinciden y se verifica la última parte del Ítem *ii*). □

6.2.2.1. *Ejemplo: escalones \mathcal{E}_h .* Retomamos el ejemplo del escalón (cerrado hacia abajo) \mathcal{E}_h , determinado por algún $h \in \mathbb{N}^n$, definido en la Identidad (6.1.2). Consideramos también el conjunto $DS(h)$ definido en la Identidad (6.2.18) precedente.

PROPOSICIÓN 6.2.18. *Sea $n \in \mathbb{N}$ un entero positivo, $n \geq 2$, $h \in \mathbb{N}^n$ un exponente monomial y \mathcal{E}_h la clase de conceptos definida como el escalón determinado por h . Entonces, se verifican las siguientes propiedades:*

- El grado promedio de \mathcal{E}_h , el grado de h en el hiper-grafo de una inclusión $G(\mathcal{E}_h)$ y el cardinal del conjunto $DS(h)$ coinciden. Es decir, tenemos que:

$$\text{deg}(h) = \sharp(DS(h)) = \text{avd}(\mathcal{E}_h).$$

- El cardinal de \mathcal{E}_h viene dado por la siguiente igualdad:

$$\sharp(\mathcal{E}_h) = \prod_{i \in DS(h)} (1 + h_i).$$

- El número total de aristas en $E(\mathcal{E}_h)$ viene dado por:

$$\sharp(E(\mathcal{E}_h)) = (n - \sharp(DS(h)))\sharp(\mathcal{E}_h) + \sum_{i \in DS(h)} \left(\prod_{j \in DS(h), j \neq i} (1 + h_j) \right).$$

- El grado de salida promedio de \mathcal{E}_h verifica:

$$E_{\mathcal{E}_h}[\text{outdeg}] = \sharp(DS(h)) - \sum_{i \in DS(h)} \frac{1}{1 + h_i}.$$

DEMOSTRACIÓN. Demostramos cada ítem por separado:

- *Ítem i)*: La primera igualdad ya fue discutida en el Lema 6.2.17 (basta tener en cuenta que h es un elemento maximal de \mathcal{E}_h con respecto al pre-orden de Dickson). Para la segunda igualdad, sin pérdida de generalidad, supongamos que $\text{DS}(h) = [m] = \{1, \dots, m\}$, para algún $m \leq n$. Entonces, se verifica la siguiente igualdad:

$$(6.2.19) \quad \mathcal{E}_h := \mathcal{E}_h|_{[m]} \times \{0\}^{n-m}.$$

Estudiamos, a continuación, las aristas en $E^{(1)}(\mathcal{E}_h)$, esto es, las aristas unarias del hiper-grafo de una inclusión $G(\mathcal{E}_h)$. Distinguiamos dos casos:

- *Caso i* $\in [m]$: Para cada $g \in \mathcal{E}_h$ y para cada $i \in [m]$, la arista $e_{\pi_i(g),i}$ tiene exactamente $1 + h_i \geq 2$ elementos.
- *Caso j* $\notin [m]$: Para cada $g \in \mathcal{E}_h$, la arista $e_{\pi_j(g),j}$ contiene un único punto ($\{g\}$) en \mathcal{E}_h , que es cerrado hacia abajo.

De lo anterior, se deduce que el conjunto de aristas unarias en \mathcal{E}_h viene dado por:

$$E^{(1)}(\mathcal{E}_h) = \{e_{\pi_j(g),j} : g \in \mathcal{E}_h, j \notin [m]\}.$$

Por lo tanto, tenemos la siguiente expresión para el cardinal del conjunto de aristas unarias:

$$\#(E^{(1)}(\mathcal{E}_h)) = (n - m)\#(\mathcal{E}_h).$$

Finalmente, aplicando el Ítem *i*) de la Proposición 6.2.5, obtenemos:

$$\text{avd}(\mathcal{E}_h) = n - \frac{\#(E^{(1)}(\mathcal{E}_h))}{\#(\mathcal{E}_h)} = n - \frac{(n - m)\#(\mathcal{E}_h)}{\#(\mathcal{E}_h)} = m = \#(\text{DS}(h)).$$

- *Ítem ii)*: De la definición de \mathcal{E}_h se deduce que:

$$\mathcal{E}_h = \{(g_1, \dots, g_n) : g_i \leq h_i, \forall i \in [n]\}.$$

Obviamente, tenemos que:

$$\#(\mathcal{E}_h) = \prod_{i \in \text{DS}(h)} (1 + h_i),$$

donde se han omitido aquellas coordenadas nulas de h , ya que no afectan al producto.

- *Ítem iii)*: Al igual que antes, supongamos, sin pérdida de generalidad, que $\text{DS}(h) = [m]$ para algún $m \leq n$. Consideramos la clase de aristas $E_i(\mathcal{E}_h)$ determinada por la proyección π_i . A continuación, distinguimos dos casos:

- *Caso 1*: Si $i \in [m]$, entonces:

$$(6.2.20) \quad \#(E_i(\mathcal{E}_h)) = \prod_{\substack{1 \leq j \leq m \\ j \neq i}} (1 + h_j).$$

Para demostrar la identidad anterior, basta observar que $\#(E_i(\mathcal{E}_h)) = \#(\pi_i(\mathcal{E}_h))$ y que

$$\pi_i(\mathcal{E}_h) = \prod_{j=1, j \neq i}^n \{0, 1, \dots, h_j\} \times \{0\}^{n-m}.$$

- *Caso 2*: Si $j \notin [m]$, tenemos que:

$$(6.2.21) \quad \#(E_j(\mathcal{E}_h)) = \prod_{i=1}^m (1 + h_i) = \#(\mathcal{E}_h).$$

Esta igualdad se prueba mediante argumentos similares al Caso 1. Aplicando la Igualdad (6.2.19), concluimos que si $j \notin [m]$, tenemos que la proyección de \mathcal{E}_h en la dirección determinada por π_j verifica:

$$\pi_j(\mathcal{E}_h) = \mathcal{E}_h|_{[m]} \times \{0\}^{n-m-1}.$$

Por lo tanto, obtenemos:

$$\#(E_j(\mathcal{E}_h)) = \#(\pi_j(\mathcal{E}_h)) = \prod_{i=1}^m (1 + h_i) = \#(\mathcal{E}_h).$$

Combinando las Igualdades (6.2.20) y(6.2.21), concluimos:

$$\sharp(E(\mathcal{E}_h)) = \sum_{i=1}^m \sharp(E_i(\mathcal{E}_h)) + \sum_{j=m+1}^n \sharp(E_j(\mathcal{E}_h)) = \sum_{i=1}^m \left(\prod_{\substack{1 \leq j \leq m \\ j \neq i}} (1 + h_j) \right) + (n - m)\sharp(\mathcal{E}_h).$$

- *Ítem iv)*: En primer lugar, por el Ítem *ii)* de la Proposición 6.2.5, obtenemos:

$$E_{\mathcal{E}_h}[\text{outdeg}] = n - \frac{\sharp(E(\mathcal{E}_h))}{\sharp(\mathcal{E}_h)}.$$

Aplicando ahora el Ítem *iii)* anterior y tras cancelar términos iguales en numeradores y denominadores, concluimos:

$$E_{\mathcal{E}_h}[\text{outdeg}] = n - \sum_{i=1}^m \frac{1}{1 + h_i} - (n - m) = m - \sum_{i=1}^m \frac{1}{1 + h_i},$$

donde, al igual que antes, $m = \sharp(\text{DS}(h))$.

□

6.3. Pseudo-cubos, dimensión DS y dimensión exponencial

6.3.1. Pseudo-cubos: definición, caracterización y propiedades básicas. El término *pseudo-cubo* fue empleado por primera vez en [BCDMY, 22], donde se utiliza para designar una noción que aparece implícita en la noción de dimensión de [DS, 14]. Esta noción es clave para poder definir uno de los conceptos de dimensión que surgen en el Aprendizaje multiclase. En [BCDMY, 22], esta dimensión recibe el nombre “*dimensión de Daniely-Shalev-Schwartz*” (o simplemente dimensión DS). El objetivo de esta subsección es caracterizar los pseudo-cubos en términos de los invariantes del hiper-grafo de una inclusión introducidos y discutidos en la Sección 6.2.

Comenzamos la subsección introduciendo las nociones de “cubo” y “poli-cubo”:

DEFINICIÓN 41 (Cubo). *Un cubo $H \subseteq \mathcal{Y}^n$ de dimensión n es cualquier subconjunto de modo que existe $\mathcal{Z} \subseteq \mathcal{Y}$ tal que $H = \mathcal{Z}^n$.*

OBSERVACIÓN 6.3.1. Puesto que estamos tratado con conjuntos finitos, resulta más complicado usar términos como politopo, poldisco o rectángulo, entre otros. Por ello, emplearemos el término *poli-cubo* para referirnos al producto cartesiano de cubos, esto es, $H \subseteq \mathcal{Y}^n$ es un poli-cubo si existen $\mathcal{Z}_1, \dots, \mathcal{Z}_n \subseteq \mathcal{Y}$ tales que:

$$(6.3.1) \quad H := \prod_{i=1}^n \mathcal{Z}_i.$$

En [BCDMY, 22], inspirados en [DS, 14], están interesados en la siguiente “generalización” de la noción de “cubo”:

DEFINICIÓN 42 (Pseudo-cubo). *Con las notaciones de las secciones precedentes, dados $n \in \mathbb{N}$, $n \geq 2$, y una clase de conceptos finita $H \subseteq \mathcal{Y}^{[n]}$, decimos que el hiper-grafo de una inclusión $G(H)$ asociado a H es un pseudo-cubo (de dimensión n) si y solo si se verifica la siguiente propiedad:*

Para todo $i \in [n]$ y para todo $h \in H$, existe un i -vecino de h en H (i.e. existe $g \in H$ tal que $h(i) \neq g(i)$, pero $h(j) = g(j)$ para todo $j \in [n] \setminus \{i\}$).

En el caso $n = 1$, llamamos pseudo-cubo de dimensión 1 a cualquier hiper-grafo de una inclusión $G(H)$ asociado a un subconjunto $H \subseteq \mathcal{Y}$ de cardinal mayor o igual que 2. Finalmente, llamamos pseudo-cubo de dimensión 0 a $G(\emptyset)$.

OBSERVACIÓN 6.3.2 (Vecinos, distancia de Hamming y aristas del hiper-grafo de una inclusión). En la definición anterior, hemos introducido la noción de i -vecino y la hemos empleado. A continuación, proporcionamos definiciones equivalentes de esta noción utilizando elementos que hemos descrito con anterioridad. Esto es, dos elementos $g, h \in H$ son i -vecinos si y solo si se verifica cualquiera de las dos siguientes propiedades:

- Ambos elementos son *vecinos*, i.e. la distancia de Hamming entre f y g es 1 (i.e. $d_{\text{Ham}}(f, g) = 1$) o, en la terminología de [Ha, 95] (cf. Observación 6.2.2), la distancia ρ de D. Haussler entre f y g verifica:

$$\rho(f, g) = \frac{1}{n}.$$

- ii) Tenemos que $f \neq g$ y existe $i \in [n]$ tal que, considerando la proyección que “olvida” la coordenada i -ésima, $\pi_i : \mathcal{Y}^n \rightarrow \mathcal{Y}^{n-1}$, se cumple que:

$$\pi_i(f) = \pi_i(g).$$

En términos de las aristas del hiper-grafo de una inclusión $G(H)$, dos elementos distintos $g, h \in H$ son i -vecinos si y solo si están en la misma arista i -ésima de $G(H)$ para algún $i \in [n]$ (i.e. si y sólo si $g, h \in e_{f,i}$, para algún $f = \pi_i(h) = \pi_i(g)$).

OBSERVACIÓN 6.3.3 (Pseudo-cubo: casos defectivos). Nótese que los casos defectivos $n = 1$ y $H = \emptyset$ han sido definidos por los siguientes motivos:

- En el caso $n = 1$ y $\sharp(H) = 1$, es habitual decir que $G(H)$ es un cubo de dimensión 1. Sin embargo, $G(H)$ no es propiamente un pseudo-cubo ya que no posee ningún vecino distinto en la única dirección de proyección posible.
- En el caso $n = 1$ y $H \subseteq \mathcal{Y}$, con $\sharp(H) \geq 2$, podemos suponer que hay una única arista (ya que no hay proyecciones) y, entonces, tenemos que $\deg(h) = 1$ para todo $h \in H$. Evidentemente, obtenemos:

$$\text{avd}(H) = 1, \quad \text{gd}(H) = \frac{1}{\sharp(H)},$$

y el grado de salida promedio de cada orientación $\sigma : E(H) = \{e\} \rightarrow H$ verifica:

$$\text{outdeg}(\sigma, h) := \begin{cases} 1, & \text{si } h \neq \sigma(e) \\ 0, & \text{en otro caso} \end{cases}.$$

Por consiguiente,

$$E_H[\text{outdeg}(\sigma)] = \frac{1}{\sharp(H)} \sum_{h \in H} \text{outdeg}(\sigma, h) = \frac{\sharp(H) - 1}{\sharp(H)},$$

y, entonces, en este caso, se cumple que:

$$\text{avd}(H) - E_H[\text{outdeg}] - \text{gd}(H) = 0.$$

- El caso $H = \emptyset$ es usado como caso defectivo del caso defectivo. Si no hay ningún punto, todos los puntos poseen vecinos distintos.

OBSERVACIÓN 6.3.4. Nótese que si $h \in H$ e $i \in [n]$, la clase de los i -vecinos de h en H son exactamente los elementos de la arista $e_{\pi_i(h),i}$. Por lo tanto, ser un pseudo-cubo implica que todas las aristas $e \in E(H)$ verifican $\sharp(e) \geq 2$.

PROPOSICIÓN 6.3.5 (Caracterización de los pseudo-cubos de dimensión n). Con las notaciones precedentes, sea $n \in \mathbb{N}$, $n \geq 1$, $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita, con $\sharp(H) \geq 2$, y $G(H) = (H, E(H))$ el hiper-grafo de una inclusión asociado a H . Las siguientes afirmaciones son equivalentes:

- El hiper-grafo de una inclusión $G(H)$ es un pseudo-cubo de dimensión n .
- Para cada arista $e \in E(H)$, $\sharp(e) \geq 2$.
- El grado promedio verifica $\text{avd}(H) = n$.
- Para todo $h \in H$, $\deg(h) = n$.
- El grado promedio, el grado de salida promedio y la densidad del hiper-grafo cumplen la siguiente igualdad:

$$\text{avd}(H) = E_H[\text{outdeg}] + \text{gd}(H).$$

En particular, si H es un pseudo-cubo de dimensión n y $n \geq 2$, tenemos que:

$$(6.3.2) \quad \frac{n}{2} \leq E_H[\text{outdeg}] < n.$$

DEMOSTRACIÓN. El caso $n = 1$ ya fue discutido en la Observación 6.3.3. Por lo tanto, supondremos que $n \geq 2$.

En primer lugar, las Observaciones 6.3.2 y 6.3.4 muestran que el Ítem *i*) implica al Ítem *ii*). A continuación, demostramos que el Ítem *ii*) implica al Ítem *i*). Dados $i \in [n]$ y $h \in H$, la arista $e := e_{\pi_i(h),i}$ contiene todos los i -vecinos de h . Por lo tanto, como $\sharp(e) \geq 2$ y $h \in e$, existe al menos un i -vecino de h en H .

Por el Ítem *i*) de la Proposición 6.2.5, tenemos que los Ítems *ii*) y *iii*) del enunciado son equivalentes.

Obviamente, el Ítem *iv*) implica al Ítem *iii*), mientras que el Ítem *ii*) implica de forma inmediata al Ítem *iv*).

La equivalencia de los Ítems *iii*) y *v*) del enunciado se sigue de los Ítems *i*) y *ii*) de la Proposición 6.2.5.

Las últimas desigualdades se deducen de la Proposición 6.2.5:

$$E_H[\text{outdeg}] \leq \text{avd}(H) = n \leq 2E_H[\text{outdeg}],$$

lo que implica:

$$\frac{n}{2} = \frac{\text{avd}(H)}{2} \leq E_H[\text{outdeg}] \leq n = \text{avd}(H).$$

Por la última dicotomía de la Proposición 6.2.5, sabemos que si $n \geq 1$, entonces:

$$E_H[\text{outdeg}] < \text{avd}(H),$$

lo que prueba la última desigualdad. \square

EJEMPLO 6.3.6 (Cubos, poli-cubos y conjuntos cerrados hacia abajo: su relación con los pseudo-cubos). Estudiamos, a continuación, la relación existente entre los cubos, poli-cubos y conjuntos cerrados hacia abajo con los pseudo-cubos. Abordamos cada situación por separado y nos restringimos al caso $n \geq 2$:

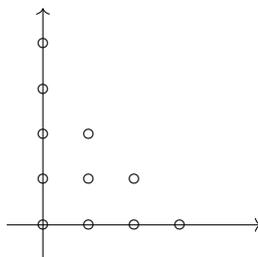
- i) Todo cubo no vacío $H := \mathcal{Z}^n$, con $n \geq 2$, es un pseudo-cubo si y solo si $\sharp(\mathcal{Z}) \geq 2$.
- ii) Todo poli-cubo no vacío $H := \prod_{i=1}^n \mathcal{Z}_i$ es un pseudo-cubo si y solo si $\sharp(\mathcal{Z}_i) \geq 2$. Nótese que si todos los \mathcal{Z}_i 's tienen cardinal al menos 2, entonces en la dirección de la proyección π_i , todo punto $(x_1, \dots, x_n) \in H$ tiene un i -vecino reemplazando $x_i \in \mathcal{Z}_i$ por otro elemento de \mathcal{Z}_i distinto de x_i , que debe existir en \mathcal{Z}_i ya que $\sharp(\mathcal{Z}_i) \geq 2$.
- iii) Un subconjunto cerrado hacia abajo $H \subseteq \mathcal{Y}^n$ es un pseudo-cubo si y solo si ninguno de sus elementos maximales (con respecto del pre-orden de Dickson \preceq) se encuentra sobre alguno de los ejes. Esto es, si no hay elementos maximales de H en:

$$\bigcup_{i=1}^n \{x = (x_1, \dots, x_n) \in \mathbb{N}^n : x_i = 0\}.$$

Obviamente, si un elemento maximal tiene una coordenada nula, por ejemplo, si $h = (0, x_2, \dots, x_n) \in H$ es un elemento maximal, y H es cerrado hacia abajo, entonces no puede haber dos puntos en la arista $e_{\pi_1(h),1} = \{h\}$. Si hubiese un segundo punto en dicha arista de H , tendría que ser de la forma (a, x_2, \dots, x_n) con $a > 0$, lo que impediría que h fuese un elemento maximal. Mostramos ahora un ejemplo de un conjunto cerrado hacia abajo que no es un pseudo-cubo. Sea $\mathcal{Y} := \{0, 1, 2, 3, 4\}$ y consideramos la clase de conceptos $H \subseteq \mathcal{Y}^{[2]}$ dada por:

$$H := \{(0, 0), (0, 1), (0, 2), (0, 3), (0, 4), (1, 0), (1, 1), (1, 2), (2, 0), (2, 1), (3, 0)\}.$$

Una imagen de una clase de conceptos anterior viene dada por:



Los vértices $(0, 4)$ y $(3, 0)$ son elementos maximales de H , con respecto al pre-orden de Dickson \preceq , que no tienen vecinos en alguna de las direcciones: $(0, 4)$ no tiene 1-vecinos (en la dirección de π_1 , la proyección que “olvida” la primera coordenada), mientras que $(3, 0)$ no posee 2-vecinos.

OBSERVACIÓN 6.3.7. Si H es un pseudo-cubo de dimensión n , entonces para cualquier orientación $\sigma : G(H) \rightarrow H$, tenemos que:

$$\frac{n}{2} \leq E_H[\text{outdeg}] \leq \text{outdeg}_{\max}(\sigma).$$

La desigualdad simplificada (i.e. $n/2 \leq \text{outdeg}_{\max}(\sigma)$ para orientaciones de pseudo-cubos de dimensión n) corresponde al Lema 12 de [BCDMY, 22], aunque sin reconocer el papel esencial que juega $E_H[\text{outdeg}]$ en esta desigualdad. Esto se debe a que estos autores no reconocen la presencia y singularidad de $E_H[\text{outdeg}]$, ni son capaces de observar que esa cantidad es independiente de la orientación σ . En suma, esta cantidad está implícita en su Lema 12 pero no la perciben.

6.3.2. Dimensiones. Estudiamos ahora algunas de las dimensiones relacionadas con el Aprendizaje multiclase. Esencialmente, estas dimensiones indican el número mínimo de puntos de muestreo necesarios para lograr un aprendizaje aproximadamente correcto (cf. [DS, 14] y [BCDMY, 22] para más detalles). El lector podrá observar cómo algunas de estas nociones se inspiran “vagamente” en la dimensión VC descrita en la Subsección 1.3.

Comenzamos introduciendo la noción de dimensión de [DS, 14], la cual, al igual que en [BCDMY, 22], definimos basándonos en la noción de pseudo-cubo discutida en la subsección previa.

DEFINICIÓN 43 (Dimensión DS, [DS, 14]). Dada una clase de conceptos finita $H \subseteq \mathcal{Y}^{[n]}$ y dado $S \subseteq [n]$, $\sharp(S) = d$, decimos que S es DS -desmenuzado por H si $H|_S$ contiene un pseudo-cubo de dimensión d . Definimos la dimensión DS de H como el máximo d tal que existe S , $\sharp(S) = d$, de modo que S es DS -desmenuzado por H . Denotamos a la dimensión DS de H mediante $\dim_{DS}(H)$.

OBSERVACIÓN 6.3.8 (La dimensión DS es monótona con respecto a la inclusión de conjuntos). A partir de la definición anterior, se deduce de forma sencilla que la dimensión DS es una función monótona con respecto a la inclusión de conjuntos. Esto es, dados $G \subseteq H \subseteq \mathcal{Y}^{[n]}$, tenemos que:

$$\dim_{DS}(G) \leq \dim_{DS}(H).$$

Esto se debe a que si $S \subseteq [n]$ es un subconjunto de cardinalidad máxima con la propiedad $G|_S$ contiene un pseudo-cubo de dimensión $\sharp(S)$, entonces $H|_S \supseteq G|_S$ también contiene un pseudo-cubo de dimensión $\sharp(S)$.

En [BCDMY, 22], los autores introducen otra noción de dimensión para una clase de conceptos finita:

DEFINICIÓN 44 (Dimensión exponencial, [BCDMY, 22]). Dada una clase de conceptos finita $H \subseteq \mathcal{Y}^{[n]}$, definimos la dimensión exponencial de H como la siguiente cantidad:

$$\dim_E(H) := \max\{\sharp(S) : \sharp(H|_S) \geq 2^{\sharp(S)}\}.$$

OBSERVACIÓN 6.3.9 (La dimensión exponencial es hereditaria). Nótese que la dimensión exponencial es hereditaria para subconjuntos y restricciones. Esto es:

- Dado $G \subseteq H \subseteq \mathcal{Y}^{[n]}$, tenemos que:

$$\dim_E(G) \leq \dim_E(H) \leq n.$$

- Dado $S \subseteq [n]$, también se cumple:

$$\dim_E(H|_S) \leq \dim_E(H) \leq n.$$

El Lema 12 de [BCDMY, 22] es el siguiente corolario de nuestros resultados previos:

COROLARIO 6.3.10 (Lema 12 de [BCDMY, 22]). Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita de dimensión DS igual a n . Entonces, para cada orientación $\sigma : E(H) \rightarrow H$, tenemos que:

$$\frac{n}{2} = \frac{\dim_{DS}(H)}{2} \leq \text{outdeg}_{\max}(\sigma).$$

DEMOSTRACIÓN. Como $\dim_{DS}(H) = n$, existe un pseudo-cubo maximal $\mathcal{C} \subseteq H$ de dimensión n . Por el Ítem v) de la Proposición 6.2.5, para cada orientación $\sigma : E(H) \rightarrow H$, existe otra orientación $\sigma' : E(\mathcal{C}) \rightarrow \mathcal{C}$ tal que:

$$\text{outdeg}_{\max}(\sigma') \leq \text{outdeg}_{\max}(\sigma).$$

Entonces, por la primera desigualdad la Identidad (6.3.2) de la Proposición 6.3.5, concluimos:

$$\frac{n}{2} \leq E_{\mathcal{C}}[\text{outdeg}] \leq \text{outdeg}_{\max}(\sigma') \leq \text{outdeg}_{\max}(\sigma),$$

de lo que se deduce la expresión del enunciado. \square

Una pregunta que surge de forma natural es si la Desigualdad (6.3.2) sigue siendo cierta si H no es un pseudo-cubo. A continuación, incluimos un ejemplo que muestra que esta última desigualdad no es cierta en general (en particular, no es cierta si H no es un pseudo-cubo de dimensión n).

EJEMPLO 6.3.11. En este ejemplo, mostramos que la primera parte de la Desigualdad (6.3.2) no sigue siendo cierta si H no es un pseudo-cubo. Sea $p \geq 3$. Dado $\mathcal{Y}_p := \{0, \dots, p-1\}$, vamos a demostrar que existe una sucesión de clases asociadas $H_p \subseteq \mathcal{Y}_p^{[2]}$ que verifica lo siguiente:

$$\lim_{p \rightarrow \infty} E_{H_p}[\text{outdeg}] = \lim_{p \rightarrow \infty} \text{avd}(H_p) = 0 < 2 = \dim_{DS}(H_p) = \dim_E(H_p).$$

Consideramos ahora el conjunto:

$$H_p := \{(0, 0), (0, 1), (1, 0), (1, 1)\} \cup \{(k, k) : 2 \leq k \leq p-1\}.$$

Tenemos que $\sharp(H_p) = p+2$. Obviamente, H_p contiene un cubo de dimensión 2 y su dimensión exponencial no puede ser mayor que 2. Por lo tanto, $\dim_{DS}(H_p) = \dim_E(H_p) = 2$. Además, H_p no es un pseudo-cubo. El número total de aristas verifica:

$$\sharp(E(H_p)) = 4 + 2(p-2) = 2p,$$

mientras que el número total de aristas unarias viene dado por:

$$\# \left(E^{(1)}(H_n) \right) = 2(p-2).$$

Entonces,

$$E_{H_p}[\text{outdeg}] = 2 - \frac{2p}{p+2} = \frac{4}{p+2},$$

mientras que el grado promedio satisface:

$$\text{avd}(H_p) = 2 - \frac{2p-4}{p+2} = \frac{8}{p+2}.$$

Por lo tanto, tomando límites, concluimos:

$$\lim_{p \rightarrow \infty} E_{H_p}[\text{outdeg}] = \lim_{p \rightarrow \infty} \text{avd}(H_p) = 0 < 2 = \dim_{\text{DS}}(H_p) = \dim_{\text{E}}(H_p),$$

y no se verifica la primera parte de la Desigualdad (6.3.2).

La noción empleada tradicionalmente en el Aprendizaje multiclase es la noción de dimensión de Natarajan. Sin embargo, como se observó en [BCDMY, 22], la dimensión de Natarajan resulta insuficiente para caracterizar el aprendizaje en el caso de que $\#(\mathcal{Y})$ no esté acotado. A continuación, mostramos la relación de esta dimensión con la dimensión exponencial.

OBSERVACIÓN 6.3.12 (Dimensión de Natarajan). En [Na, 88] y [Na, 89], B. K. Natarajan introdujo una primera noción de dimensión para el contexto del Aprendizaje multiclase. Esta noción no será central en nuestras discusiones, sin embargo, merece cierto reconocimiento histórico. Expondremos esta noción siguiendo el enfoque de [BCDMY, 22].

Al igual que antes, sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Dado $q \in \mathbb{N}$, $1 \leq q \leq n$, y dado $S = (s_1, \dots, s_q) \in [n]^q$, decimos que S es N -desmenuzado por H si existen $f, g \in H$ tales que $f(s_i) \neq g(s_i)$ para todo $i \in [q]$ y tales que:

$$H(S) \supseteq \{f(s_1), g(s_1)\} \times \dots \times \{f(s_q), g(s_q)\},$$

donde:

$$H(S) := \{(h(s_1), \dots, h(s_q)) \in \mathcal{Y}^q : h \in H\}.$$

Nótese que la condición anterior es equivalente a que $H(S)$ contenga un poli-cubo (cf. Observación 6.3.1), dado como el producto cartesiano de subconjuntos de cardinalidad 2, aunque sus elementos vengan de funciones distintas de H restringidas a S . La dimensión de Natarajan de H es el máximo de los q 's para los cuales existe un subconjunto $S \in [n]^q$ que es N -desmenuzado por H . Denotamos a la dimensión de Natarajan de H por $\dim_N(H)$.

Con las notaciones precedentes, sea $\tilde{S} := \{s_i : i \in [q]\}$ el conjunto formado las diferentes coordenadas de una lista $S = (s_1, \dots, s_q) \in [n]^q$. Evidentemente,

$$\#(\tilde{S}) = q.$$

Si S es N -desmenuzado por H , entonces $s_i \neq s_j$, para todo $i \neq j$. Nótese que si $s_i = s_j$, con $i \neq j$, entonces, existe $h \in H$ de modo que se verifican las siguientes propiedades:

$$h(s_i) = f(s_i), h(s_j) = g(s_j), h(s_k) \in \{f(s_k), g(s_k)\}, \text{ para } k \notin \{i, j\}.$$

Pero, si $s_i = s_j$, entonces $f(s_i) = f(s_j) = h(s_i)$ mientras que $g(s_i) = g(s_j) = h(s_j) = h(s_i)$ y, por lo tanto, $f(s_i) = g(s_i)$, lo cual no puede ser posible si S es N -desmenuzado por H .

Por consiguiente, si $S \in [n]^q$ es N -desmenuzado por H , entonces:

$$(6.3.3) \quad \#(H|_{\tilde{S}}) \geq 2^{\#(\tilde{S})} = 2^q.$$

En particular, tenemos que:

$$(6.3.4) \quad \dim_N(H) \leq \dim_{\text{E}}(H).$$

6.3.3. Pseudo-dimensión DS . Mientras que la dimensión exponencial depende únicamente de sus proyecciones, la dimensión DS ha sido definida para ser monótona con respecto a la inclusión de conjuntos al incluir el término “contiene” en su definición. Esto nos lleva a definir la noción de pseudo-dimensión DS .

DEFINICIÓN 45 (Pseudo-dimensión DS). Con las notaciones precedentes, dada una clase de conceptos finita $H \subseteq \mathcal{Y}^{[n]}$, definimos su pseudo-dimensión DS como el máximo de los cardinales de cualquier subconjunto $S \subseteq [n]$ tal que $H|_S$ sea un pseudo-cubo de dimensión $\sharp(S)$. Es decir, definimos:

$$p \dim_{DS}(H) := \max\{\sharp(S) : S \subseteq [n], H|_S \text{ es un pseudo-cubo de dimensión } \sharp(S)\}.$$

Si no hay ningún subconjunto $S \subseteq [n]$ tal que $H|_S$ es un pseudo-cubo, decimos que su pseudo-dimensión DS es 0.

La pseudo-dimensión DS está caracterizada por la cardinalidad de un conjunto que maximiza una cierta función definida en $2^{[n]}$.

PROPOSICIÓN 6.3.13 (La pseudo-dimensión DS es el valor donde se maximiza cierta función). Con las mismas notaciones que antes, sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Definimos la siguiente función:

$$\psi_H : 2^{[n]} \longrightarrow \mathbb{R}$$

$$S \longmapsto \begin{cases} \text{avd}(H|_S) - E_{H|_S}[\text{outdeg}] - \text{gd}(H|_S), & \text{si } n \geq 2 \text{ o } \sharp(H) \geq 2 \\ 0, & \text{si } n = 1 \text{ u } \sharp(H) = 1 \\ 0, & \text{si } H = \emptyset. \end{cases}$$

Entonces, tenemos que:

$$p \dim_{DS}(H) = \arg \max_{S \subseteq [n]} \psi_H(S).$$

DEMOSTRACIÓN. Evitamos los casos defectivos $n = 1$ y $\sharp(H) = 1$ o $\sharp(H) = 0$.

De acuerdo con nuestros resultados previos, tenemos que para cada subconjunto $S \subseteq [n]$, se verifica la siguiente propiedad:

$$\psi_H(S) = -\frac{\sharp(E^{(1)}(H|_S))}{\sharp(H|_S)} \leq 0.$$

Además, $\psi_H(S) = 0$ si y solo si $H|_S$ es un pseudo-cubo de dimensión $\sharp(S)$. Por lo tanto, S maximiza ψ_H si y solo si $H|_S$ es un pseudo-cubo. \square

EJEMPLO 6.3.14 (La pseudo-dimensión DS no es una función monótona con respecto a la inclusión de conjuntos). El problema con la pseudo-dimensión DS es que no es una función monótona con respecto a la inclusión de conjuntos como muestra el siguiente ejemplo. Consideramos la clase de conceptos finita:

$$H := \{(0, 0), (0, 1), (1, 0), (1, 1), (0, 2)\} \subseteq \{0, 1, 2\}^{[2]},$$

y el subconjunto $G := H \setminus \{(0, 2)\}$. Nótese que $G = \mathcal{Z}^2$, donde $\mathcal{Z} = \{0, 1\}$, es un cubo y, por lo tanto, es un pseudo-cubo de dimensión 2 (cf. Ejemplo 6.3.6). Por otro lado, H no es un pseudo-cubo ya que tiene una arista unaria $\{(0, 2)\}$ (cf. Proposición 6.3.5) y, en consecuencia,

$$p \dim_{DS}(H) < 2.$$

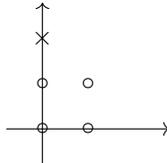
Si consideramos la restricción de H a $S = \{1\}$, obtenemos $H|_S = \{0, 1\}$, que es un pseudo-cubo debido a que tiene más de un elemento (ver Definición 42). Por lo tanto,

$$p \dim_{DS}(H) = 1.$$

Esto es,

$$1 = p \dim_{DS}(H) < p \dim_{DS}(G) = 2,$$

pero $G \subseteq H$. La siguiente imagen muestra todos los puntos de H y representa mediante \times el punto $\{(0, 2)\}$ que se elimina para obtener G :



Por otro lado, la pseudo-dimensión DS da lugar a la dimensión DS como un máximo. Es decir, tenemos que:

$$(6.3.5) \quad \dim_{DS}(H) = \max \left\{ \arg \max_{S \subseteq [n]} \psi_G(S) : G \subseteq H \right\}.$$

6.3.4. El escalón \mathcal{E}_h y sus dimensiones. En esta subsección retomamos las notaciones introducidas en la Subsección 6.2.2.1. La siguiente proposición complementa, en términos de dimensiones, el análisis de los invariantes del escalón \mathcal{E}_h iniciado en la Proposición 6.2.18.

PROPOSICIÓN 6.3.15. *Sea $n \in \mathbb{N}$ un entero positivo, $n \geq 2$, $h \in \mathbb{N}^n$ un exponente monomial y \mathcal{E}_h la clase de conceptos definida como el escalón determinado por h . Entonces, se verifican las siguientes propiedades:*

- i) *El grado promedio de \mathcal{E}_h , el grado de h en $G(\mathcal{E}_h)$ y la dimensión DS de \mathcal{E}_h coinciden. Es decir, tenemos que:*

$$\deg(h) = \sharp(DS(h)) = \text{avd}(\mathcal{E}_h) = \dim_{DS}(\mathcal{E}_h).$$

- ii) *La dimensión exponencial de \mathcal{E}_h verifica:*

$$\dim_E(\mathcal{E}_h) := \min \left\{ n, \left\lfloor \left(\sum_{i \in DS(h)} \log_2(1 + h_i) \right) \right\rfloor \right\}.$$

En particular, se cumplen las siguientes desigualdades:

$$E_{\mathcal{E}_h}[\text{outdeg}] < \text{avd}(\mathcal{E}_h) = \dim_{DS}(\mathcal{E}_h) \leq \dim_E(\mathcal{E}_h).$$

La última desigualdad se convierte en una igualdad si y solo si \mathcal{E}_h es un pseudo-cubo de dimensión n (o, equivalentemente, si $h \in (\mathbb{Z}^\times)^n$, donde $\mathbb{Z}^\times := \mathbb{Z} \setminus \{0\}$, o, también, si $DS(h) = [n]$).

DEMOSTRACIÓN. Demostramos cada ítem por separado:

- *Ítem i):* Ya hemos demostrado las dos primeras igualdades en la Proposición 6.2.18. Al igual que en la demostración de dicha proposición, supongamos, sin pérdida de generalidad, que $DS(h) = [m] = \{1, \dots, m\}$, para algún $m \leq n$. Por lo tanto:

$$(6.3.6) \quad \mathcal{E}_h := \mathcal{E}_h|_{[m]} \times \{0\}^{n-m}.$$

Bajo estas hipótesis, en la Proposición 6.2.18 ya estudiamos la clase de aristas de $G(\mathcal{E}_h)$ que son unarias:

$$E^{(1)}(\mathcal{E}_h) = \{e_{\pi_j(g),j} : g \in \mathcal{E}_h, j \notin [m]\}.$$

Por lo tanto, tomando $S = [m]$, tenemos que el conjunto:

$$\mathcal{E}_h|_{[m]} = \mathcal{E}_h|_{[m]}$$

es un pseudo-cubo cerrado hacia abajo de dimensión DS igual a m . Por otro lado, para cada $S \subseteq [n]$, si existe $j \in S \setminus [m]$, entonces, para todo $g \in \mathcal{E}_h$, la arista $e_{\pi_j(g),j}$ es unaria en \mathcal{E}_h (y, por lo tanto, en $\mathcal{E}_h|_S$). En particular, si $S \cap [n] \setminus [m] \neq \emptyset$, $\mathcal{E}_h|_S$ no puede ser un pseudo-cubo y $[m]$ es máximo con la propiedad $\mathcal{E}_h|_S$ contiene un pseudo-cubo de dimensión $\sharp(S)$. Es decir, tenemos que:

$$\deg(h) = m = \dim_{DS}(\mathcal{E}_h).$$

- *Ítem ii):* Supongamos, al igual que en el Ítem i), que $DS(h) = [m]$ para algún $m \leq n$. Sea $S \subseteq [n]$ cualquier subconjunto tal que $[m] \subseteq S$. Gracias a la Identidad (6.2.19), concluimos de forma sencilla que:

$$\mathcal{E}_h|_S = \mathcal{E}_h|_{[m]} \times \{0\}^{S \setminus [m]}.$$

Por lo tanto, tenemos que:

$$\prod_{i=1}^m (1 + h_i) = \sharp(\mathcal{E}_h|_S) \geq 2^{\sharp(S)}$$

se verifica si y sólo si:

$$\sharp(S) \leq \left\lfloor \sum_{i=1}^m \log_2(1 + h_i) \right\rfloor.$$

Como la dimensión exponencial es siempre menor o igual que n , concluimos la igualdad del enunciado. Las últimas desigualdades se deducen de forma inmediata de la Proposición 6.2.18 y los ítems anteriores. \square

6.3.5. Pseudo-cubos cerrados hacia abajo. Comenzamos la subsección mostrando la siguiente caracterización de los pseudo-cubos cerrados hacia abajo:

PROPOSICIÓN 6.3.16 (Pseudo-cubos cerrados hacia abajo). *Sea $H \subseteq \mathcal{Y}^{[n]}$, $n \geq 1$, una clase de conceptos finita. Las siguientes dos propiedades son equivalentes:*

- i) H es un pseudo-cubo cerrado hacia abajo de dimensión n .*
- ii) H es la unión finita de escalones de dimensión DS igual a n . Es decir, existe una familia finita $h_1, \dots, h_s \in \mathbb{N}^n$ de exponentes monomiales que verifican:*
 - (a) El grado promedio, la dimensión DS y la dimensión exponencial de cada uno de los escalones determinados por cada h_k es igual a n , i.e.*

$$\forall k \in \{1, \dots, s\}, \text{avd}(\mathcal{E}_{h_k}) = \dim_{DS}(\mathcal{E}_{h_k}) = \dim_E(\mathcal{E}_{h_k}) = n.$$

(b) H satisfice:

$$(6.3.7) \quad H := \bigcup_{i=1}^s \mathcal{E}_{h_i}.$$

DEMOSTRACIÓN. Demostramos las dos implicaciones:

- $i) \implies ii)$:* Si H es cerrado hacia abajo, la Proposición 6.1.2 implica que H admite una descomposición como en la Identidad (6.3.7), donde h_1, \dots, h_s están determinados de forma única y son los elementos maximales (distintos dos a dos) de H con respecto al pre-orden de Dickson \preceq . Supongamos que H es un pseudo-cubo cerrado hacia abajo y que contamos con dicha descomposición. A continuación, demostramos, por reducción al absurdo, que $\sharp(DS(h_i)) = n$ para cada $i \in \{1, \dots, s\}$. Sin pérdida de generalidad, supongamos que $\dim_{DS}(\mathcal{E}_{h_1}) < n$. Entonces, de acuerdo a la Proposición 6.2.18, tenemos que $\sharp(DS)(h_1) < n$. Equivalentemente, suponiendo que $h_1 = (h_{1,1}, \dots, h_{1,n})$, debería existir algún $i \in [n]$ tal que $h_{1,i} = 0$. Consideramos ahora la proyección $\pi_i : H \rightarrow \mathcal{Y}^{n-1}$ y supongamos que $f = \pi_i(h_1)$ y sea $e_{f,i}$ la arista asociada en H . Obviamente, $h_1 \in e_{f,i}$. Vamos a demostrar que $e_{f,i} = \{h_1\}$. Supongamos que $g = (g_1, \dots, g_n) \in e_{f,i}$. Entonces, $\pi_i(g) = \pi_i(h_1)$ y, por lo tanto, tenemos que $g_j = h_{1,j}$, $\forall j \neq i$. Como $h_1 = (g_1, \dots, g_{i-1}, 0, g_{i+1}, \dots, g_n)$ y $g_i \geq 0$, inmediatamente concluimos que $h_1 \preceq g$. Al ser h_1 un elemento maximal en H , entonces $g = h_1$ y $e_{f,i} = \{h_1\}$. Esto implica que H no puede ser un pseudo-cubo de dimensión n . Es decir, si H es un pseudo-cubo cerrado hacia abajo, entonces por la Proposición 6.3.15 concluimos:

$$\text{avd}(\mathcal{E}_{h_k}) = \dim_{DS}(\mathcal{E}_{h_k}) = n, \forall k \in \{1, \dots, s\}.$$

Como la dimensión exponencial no puede ser estrictamente mayor que n , también tenemos que $\dim_E(\mathcal{E}_{h_k}) = n$ para todo k y se cumplen las Propiedades (a) y (b) del Ítem *ii)*.

- $ii) \implies i)$:* Como H es la unión finita de escalones, tenemos, de forma inmediata, que es cerrado hacia abajo. De acuerdo a la Observación 6.1.3, toda descomposición de un conjunto cerrado hacia abajo como en el Identidad (6.3.7) puede refinarse hasta obtener una descomposición en la que h_1, \dots, h_s son elementos maximales distintos de H con respecto al pre-orden de Dickson \preceq . Supongamos que la descomposición mostrada en la Identidad (6.3.7) es dicha descomposición única de H . A continuación, vamos a demostrar que H es un pseudo-cubo de dimensión DS igual a n . Para ello, consideramos $i \in [n]$ y $h \in H$. Entonces, debe existir un elemento maximal h_k de H tal que $h \in \mathcal{E}_{h_k}$. Consideramos también la proyección $f := \pi_i(h)$ de h y la arista $e_{f,i} \in E(H)$ determinada por esta proyección. Tenemos que

$$h \in \mathcal{E}_{h_k} \cap e_{f,i} \subseteq e_{f,i},$$

y, por la hipótesis (a), sabemos que \mathcal{E}_{h_k} es un pseudo-cubo de dimensión n . Entonces, tenemos que:

$$\sharp(e_{f,i}) \geq \sharp(\mathcal{E}_{h_k} \cap e_{f,i}) \geq 2.$$

Por lo tanto, H es un pseudo-cubo de dimensión n . □

COROLARIO 6.3.17. *Si $H \subseteq \mathcal{Y}^{[n]}$, $n \geq 1$, es un conjunto cerrado hacia abajo, las siguientes propiedades son equivalentes:*

- i) H es un pseudo-cubo de dimensión n .*
- ii) Se verifica lo siguiente:*

$$E_H[\text{outdeg}] < \text{avd}(H) = \dim_{DS}(H) = \dim_E(H) = n \leq 2E_H[\text{outdeg}].$$

iii) Dada una descomposición H en sus componentes (escalones) maximales:

$$H =: \bigcup_{i=1}^s \mathcal{E}_{h_i},$$

entonces, se verifica lo siguiente:

$$\text{avd}(H) = \frac{1}{s} \sum_{i=1}^s \deg(h_i) = \max\{\deg(h_1), \dots, \deg(h_s)\} = \dim_{\text{DS}}(H) = \dim_{\text{E}}(H) = n.$$

iv) Dada una descomposición de H en sus componentes (escalones) maximales:

$$H =: \bigcup_{i=1}^s \mathcal{E}_{h_i},$$

entonces, para cada i , $1 \leq i \leq s$, se cumple que:

$$\text{avd}(H) = \deg(h_i) = \dim_{\text{DS}}(\mathcal{E}_{h_i}) = \dim_{\text{E}}(\mathcal{E}_{h_i}) = \dim_{\text{DS}}(H) = \dim_{\text{E}}(H) = n.$$

DEMOSTRACIÓN. Demostramos las equivalencias de los ítems anteriores como sigue:

- Ítem i) \iff Ítem ii): Por la Observación 6.3.9, la dimensión exponencial de H es menor o igual que n y es mayor o igual que la dimensión exponencial de cualquiera de sus subconjuntos (por ser hereditaria). Por lo tanto, aplicando la Proposición 6.3.5 y el Ítem ii) de la Proposición 6.3.16, concluimos de forma inmediata que si H es un pseudo-cubo de dimensión n , entonces se verifica la siguiente igualdad:

$$\text{avd}(H) = \dim_{\text{DS}}(H) = \dim_{\text{E}}(H).$$

Como $n \geq 1$, la última dicotomía de la Proposición 6.2.5 implica que:

$$E_H[\text{outdeg}] < \text{avd}(H).$$

La otra desigualdad es consecuencia de la Proposición 6.2.5. Para el recíproco, si $\text{avd}(H) = \dim_{\text{DS}}(H) = n$, entonces por la Proposición 6.3.5, tenemos que H es un pseudo-cubo.

- Ítem iii) \iff Ítem iv): Es inmediata.
- Ítem ii) \implies Ítem iv): Como la dimensión DS es monótona con respecto a la inclusión de conjuntos (cf. Observación 6.3.8), tenemos que:

$$(6.3.8) \quad \deg(h_i) = \dim_{\text{DS}}(\mathcal{E}_{h_i}) \leq \max\{\dim_{\text{DS}}(\mathcal{E}_{h_i}) : 1 \leq i \leq s\} \leq \dim_{\text{DS}}(H) = \text{avd}(H) = n,$$

donde la primera igualdad es consecuencia del Ítem i) de la Proposición 6.3.15 (nótese que h_i es un elemento maximal de H con respecto al pre-orden de Dickson \preceq). Como $\text{avd}(H) = n$, tenemos que $\deg(h_i) = n$ y concluimos que las desigualdades de la Identidad 6.3.8 son igualdades. La igualdad $\dim_{\text{DS}}(\mathcal{E}_{h_i}) = \dim_{\text{E}}(\mathcal{E}_{h_i})$ es consecuencia inmediata de la última parte de la Proposición 6.3.15.

- Ítem iv) \implies Ítem ii): Es una consecuencia inmediata de la Proposición 6.3.16.

□

EJEMPLO 6.3.18. Los resultados anteriores muestran que en el caso de pseudo-cubos cerrados hacia abajo, $\text{avd}(H)$ coincide con las dimensiones exponencial y DS . Sin embargo, esto no es cierto, en general, cuando H no es un pseudo-cubo. A continuación, mostramos un contra-ejemplo.

Dado $\mathcal{Y}_p := \{0, 1, \dots, p-1\}$, entonces existe una sucesión $F_p \subseteq \mathcal{Y}_p^{[2]}$ de clases de funciones cerradas hacia abajo que verifica:

$$1 = E_{F_p}[\text{outdeg}] = \lim_{p \rightarrow \infty} \text{avd}(F_p) < \dim_{\text{DS}}(F_p) = \dim_{\text{E}}(F_p) = 2 \leq 2E_{F_p}[\text{outdeg}].$$

Consideramos ahora las siguientes clases:

$$F_p := \{(0, 0), (0, 1), (1, 0), (1, 1)\} \bigcup \{(0, k) : 2 \leq k \leq p-2\}.$$

Estos conjuntos son cerrados hacia abajo, pero no son pseudo-cubos para $p \geq 2$. Al igual que en el Ejemplo 6.3.11, contienen un pseudo-cubo de dimensión 2 y, por lo tanto,

$$\dim_{\text{DS}}(F_p) = \dim_{\text{E}}(F_p) = 2.$$

El número de vértices verifica $\#(F_p) = p + 2$. Por otro lado, el número de aristas viene dado por:

$$\#(E(F_p)) = p + 2, \quad \#(E^{(1)}(\mathcal{F}_p)) = p - 2.$$

Entonces, tenemos que:

$$E_{F_p}[\text{outdeg}] = 2 - \frac{p+2}{p+2} = 1, \quad \text{avd}(F_p) = 2 - \frac{p-2}{p+2} = 1 + \frac{4}{p+2}.$$

Por lo tanto, concluimos:

$$1 = E_{F_p}[\text{outdeg}] = \lim_{p \rightarrow \infty} \text{avd}(F_p) < \dim_{\text{DS}}(F_p) = \dim_{\text{E}}(F_p) = 2 \leq 2E_{F_p}[\text{outdeg}].$$

LEMA 6.3.19. *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita y $S \subseteq [n]$ un subconjunto no vacío. Sea $f \in H$ una de las funciones en H . Denotamos por $f' := f|_S$ a la restricción de f al subconjunto S . Entonces, tenemos que:*

$$\mathcal{E}_f|_S = \{g|_S : g \in \mathcal{E}_f\} = \{g : S \rightarrow \mathcal{Y} : g \preceq f'\} = \mathcal{E}_{f'} \subseteq H|_S.$$

DEMOSTRACIÓN. La inclusión \subseteq es inmediata a partir de la definición del pre-orden de Dickson. Si $h \preceq f$, entonces $h|_S \preceq f|_S$.

Para la otra inclusión, supongamos, sin pérdida de generalidad, que $S = [m]$ para algún $m \in [n]$. Supongamos también que $h' \in \mathcal{E}_{f'}$. Entonces, tenemos que:

$$h' = (h'(1), \dots, h'(m)) \preceq (f(1), \dots, f(m)) = f'.$$

Consideramos $h \in \mathcal{Y}^{[n]}$ dado por:

$$h \leftrightarrow (h'(1), \dots, h'(m), 0, \dots, 0).$$

Nótese que:

$$h \preceq (f(1), \dots, f(m), f(m+1), \dots, f(n)) = f,$$

ya que $f(j) \geq 0$ para todo $j \in [n]$ y $f(i) = f'(i)$ para $1 \leq i \leq m$. Como \mathcal{E}_f es cerrado hacia abajo, entonces $h \in \mathcal{E}_f$ y $h|_S = h'$. Esto prueba la inclusión \supseteq y concluye la demostración del lema. \square

Antes de continuar, introducimos el siguiente resultado técnico conocido, que nos será útil para demostrar la Proposición 6.3.21.

LEMA 6.3.20. *Sean $a_1, \dots, a_r, b_1, \dots, b_m \in \mathbb{R}$ números reales no negativos que verifican:*

$$b_1 \leq b_2 \leq \dots \leq b_m \leq a_1 \leq \dots \leq a_r.$$

Entonces, tenemos que:

$$\frac{a_1 + a_2 + \dots + a_r + b_1 + \dots + b_m}{m+r} \leq \frac{a_1 + \dots + a_r}{r}.$$

DEMOSTRACIÓN. Bajo nuestras hipótesis, se cumple que:

$$\frac{a_1 + \dots + a_r}{r} \geq \frac{ra_1}{r} = \frac{ma_1}{m} \geq \frac{b_1 + \dots + b_m}{m}.$$

Entonces, obtenemos:

$$b_1 + \dots + b_m \leq m \left(\frac{a_1 + \dots + a_r}{r} \right).$$

Finalmente, concluimos:

$$b_1 + \dots + b_m + a_1 + \dots + a_r \leq m \left(\frac{a_1 + \dots + a_r}{r} \right) + r \left(\frac{a_1 + \dots + a_r}{r} \right) = (m+r) \left(\frac{a_1 + \dots + a_r}{r} \right),$$

de lo que se deduce el enunciado. \square

La siguiente proposición mejora y amplía, en el caso de clases cerradas hacia abajo, la Proposición 27 de [BCDMY, 22]:

PROPOSICIÓN 6.3.21. *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita cerrada hacia abajo. Entonces, tenemos que:*

i) *Si $\dim_{\text{DS}}(H) \geq 1$, entonces:*

$$E_H[\text{outdeg}] < \text{avd}(H) \leq \dim_{\text{DS}}(H).$$

ii) *Si $\dim_{\text{E}}(H) \geq 1$, entonces:*

$$E_H[\text{outdeg}] < \text{avd}(H) \leq \dim_{\text{E}}(H).$$

DEMOSTRACIÓN. La primera desigualdad de ambos ítems es consecuencia inmediata de la Proposición 6.2.5.

Consideramos ahora una descomposición de H en sus componentes maximales (como en la Proposición 6.1.2):

$$H := \bigcup_{i=1}^s \mathcal{E}_{h_i},$$

donde $\{h_1, \dots, h_s\}$ son los elementos maximales (distintos dos a dos) de H con respecto al pre-orden de Dickson \preceq . Salvo reordenación de los índices, supongamos que $r, 1 \leq r \leq s$, es tal que:

$$T := \#(DS(h_i)) = \max\{\#(DS(h_i)) : 1 \leq i \leq s\} \Leftrightarrow 1 \leq i \leq r.$$

Y $\#(DS(h_j)) < T$ para $r+1 \leq j \leq s$. Consideramos $S_i := DS(h_i)$, donde $1 \leq i \leq r$ y $\#(S_i) = T$. La restricción de H a S_i verifica:

$$H|_{S_i} := \bigcup_{k=1}^s \mathcal{E}_{h_k}|_{S_i} \supseteq \mathcal{E}_{h_i}|_{DS(h_i)}.$$

Consideramos ahora $h'_i := h_i|_{DS(h_i)}$ y tenemos que:

$$H|_{S_i} \supseteq \mathcal{E}_{h'_i},$$

y $h'_i(j) \geq 1$ para todo $j \in DS(h_i)$. Además, $\deg(h'_i)$ en $G(\mathcal{E}_{h'_i})$ cumple que $T = \deg(h'_i) = \deg(h_i)$. Entonces, $\mathcal{E}_{h'_i}$ es un pseudo-cubo cerrado hacia abajo que verifica:

$$\deg(h'_i) = \#(DS(h_i)) = T = \dim_{DS}(\mathcal{E}_{h'_i}).$$

Por lo tanto, $T \leq \dim_{DS}(H)$. Adicionalmente, para cada $h \in H$, tenemos que:

$$\deg(h) \leq \max\{\deg(h_1), \dots, \deg(h_s)\}.$$

Por consiguiente:

$$\deg(h) \leq \min\{\deg(h_1), \dots, \deg(h_r)\} = T = \max\{\deg(h_1), \dots, \deg(h_r)\} = \frac{\deg(h_1) + \dots + \deg(h_r)}{r}.$$

Aplicando el Lema 6.3.20, salvo reordenación de los índices, concluimos:

$$\text{avd}(H) = \frac{1}{\#(H)} \left(\sum_{h \in H} \deg(h) \right) \leq \frac{\deg(h_1) + \dots + \deg(h_r)}{r} = T \leq \dim_{DS}(H).$$

Esto demuestra la segunda desigualdad del Ítem *i*).

Para probar la segunda desigualdad del Ítem *ii*), procedemos de forma análoga y concluimos que para todo i , $1 \leq i \leq r$, tenemos que:

$$2^T \leq \#(\mathcal{E}_{h'_i}) = \prod_{k \in DS(h_i)} (h_i(k) + 1),$$

y

$$\mathcal{E}_{h'_i} = \mathcal{E}_{h_i}|_{DS(h_i)} \subseteq H|_{DS(h_i)}.$$

Como $T = \#(DS(h_i))$, obtenemos:

$$2^T = 2^{\#(DS(h_i))} \leq \#(H|_{DS(h_i)}).$$

De lo anterior, se deduce que $T \leq \dim_E(H)$ como queríamos demostrar. \square

COROLARIO 6.3.22. *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita cerrada hacia abajo y $m := \dim_{DS}(H)$ su dimensión DS. Si $S \subseteq [n]$ es un subconjunto tal que $\#(S) = m$ y $H|_S$ contiene un pseudo-cubo de dimensión m , entonces existe un subconjunto $\mathcal{C} \subseteq H|_S$ que verifica las siguientes propiedades:*

- \mathcal{C} es un pseudo-cubo cerrado hacia abajo.
- Se cumple que:

$$\text{avd}(\mathcal{C}) = \dim_{DS}(H) = \dim_E(\mathcal{C}) \leq \dim_E(H).$$

- \mathcal{C} es maximal con respecto a la inclusión de conjuntos (\subseteq) verificando las dos propiedades anteriores.

Además, el propio $H|_S$ es un pseudo-cubo si y solo si:

$$\text{avd}(H|_S) = \dim_{DS}(H).$$

DEMOSTRACIÓN. Por el Lema 6.1.5, si H es cerrado hacia abajo y $S \subseteq [n]$ es un subconjunto, entonces $H|_S$ es también cerrado hacia abajo. Gracias a la Proposición 6.1.2, tenemos que $H|_S$ admite una descomposición en componentes maximales:

$$H|_S := \bigcup_{k=1}^s \mathcal{E}_{h_k},$$

de modo que para cada $k \in \{1, \dots, s\}$:

$$h_k \notin \bigcup_{j \neq k} \mathcal{E}_{h_j}.$$

Salvo reordenación de los índices, podemos suponer que existe $t \in [s]$ tal que:

$$\dim_{DS}(\mathcal{E}_k) = \#(S) = \dim_{DS}(H) = m \iff k \leq t.$$

Definimos $\mathcal{C} \subseteq H|_S$ de la siguiente forma:

$$\mathcal{C} := \bigcup_{k=1}^t \mathcal{C}_{h_k}.$$

La Proposición 6.3.16 implica que \mathcal{C} es un pseudo-cubo cerrado hacia abajo que verifica las propiedades del enunciado. \square

OBSERVACIÓN 6.3.23. Del corolario anterior se deduce la siguiente caracterización de la dimensión DS para subconjuntos cerrados hacia abajo $H \subseteq \mathcal{Y}^{[n]}$:

$$\dim_{DS}(H) := \max\{\#(S) : S \subseteq [n], \exists \mathcal{C} \subseteq H|_S, \text{avd}(\mathcal{C}) = \#(S)\}.$$

6.4. La técnica del “shifting” en el Aprendizaje multiclase

6.4.1. La técnica del shifting. La técnica del “shifting” fue empleada por D. Haussler en [Ha, 95] para demostrar el Lema de Sauer-Shelah-Perles (cf. Lema 1.3.1) y otros resultados en el contexto del Aprendizaje binario. Esta técnica fue definida en [Fr, 83] e, independientemente, en [Al, 83] (véase también el papel del shifting en relación con la traza y la dualidad en [Pa, 24]). En [RBR, 06], los autores extienden esta técnica al caso multiclase. Posteriormente, en [BCDMY, 22] se vuelve a analizar esta técnica. Utilizaremos las notaciones, con algunas variaciones, de esta última referencia.

DEFINICIÓN 46 (Shifting de una arista). Sea $H \subseteq \mathcal{Y}^n$ una clase de conceptos finita y $G(H) = (H, E(H))$ el hiper-grafo de una inclusión asociado a H . Para cada $i \in [n]$, sea $\pi_i : \mathcal{Y}^n \rightarrow \mathcal{Y}^{n-1}$ la proyección que “olvida” la coordenada i -ésima. Para cada $f \in \mathcal{Y}^{n-1}$, sea $e_{f,i} := \pi_i^{-1}(\{f\}) \cap H$ la fibra sobre f en la dirección determinada por π_i , que es el subtrato como subconjunto de H de la arista correspondiente en $E(H)$. Para cada $i \in [n]$, definimos el shift de la arista $e_{f,i}$ en la dirección de π_i de la siguiente forma:

$$\mathbb{S}_i(e_{f,i}) := \{g \in \mathcal{Y}^n : \pi_i(g) = f, g(i) \in \{0, \dots, \#(e_{f,i}) - 1\}\} \subseteq \mathcal{Y}^n.$$

Dicho de otra forma, el operador \mathbb{S}_i modifica la arista $e_{f,i}$ para obtener otra arista del mismo cardinal, en la misma dirección y cerrada hacia abajo en la dirección de π_i , preservando el pre-orden de Dickson \preceq sobre sus elementos. Es decir, “empuja hacia abajo” la coordenada i -ésima de cada elemento en $e_{f,i}$ mientras preserve tanto la cardinalidad $\#(e_{f,i})$ como la proyección (i.e. f) en la dirección de π_i . En particular, $e_{f,i}$ y $\mathbb{S}_i(e_{f,i})$ son biyectables mediante una biyección que preserva el pre-orden de Dickson \preceq y que denotamos por:

$$\widetilde{\mathbb{S}}_{f,i} : e_{f,i} \rightarrow \mathbb{S}_i(e_{f,i}).$$

DEFINICIÓN 47 (Shifting de una clase de conceptos). Con las notaciones anteriores, definimos el shift de H en la dirección de π_i como el siguiente conjunto:

$$\mathbb{S}_i(H) := \bigcup_{f \in \pi_i(H)} \mathbb{S}_i(e_{f,i}).$$

Esto es, tenemos una aplicación:

$$\mathbb{S}_i : 2^{\mathcal{Y}^n} \rightarrow 2^{\mathcal{Y}^n},$$

donde $2^{\mathcal{Y}^n}$ es la clase de subconjuntos de \mathcal{Y}^n .

En el lado positivo, el shifting no incrementa la dimensión exponencial (cf. Corolario 23 de [BCDMY, 22]).

PROPOSICIÓN 6.4.1 (Afirmación 22 y Corolario 23 de [BCDMY, 22]). Con las notaciones precedentes, sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Para cada $i \in [n]$ y para cada $S \subseteq [n]$, tenemos que:

$$\#(\mathbb{S}_i(H)|_S) \leq \#(H|_S).$$

Además, también se cumple que:

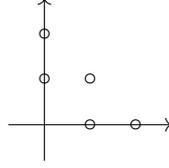
$$\dim_E(\mathbb{S}_i(H)) \leq \dim_E(H).$$

A continuación, exponemos (con todos los detalles) el Ejemplo 19 de [BCDMY, 22].

EJEMPLO 6.4.2 (Ejemplo 19 de [BCDMY, 22]). El grado promedio y la dimensión DS pueden aumentar después de un shift como se muestra en el Ejemplo 19 de [BCDMY, 22]. Sea $\mathcal{Y} := \{0, 1, 2\}$. Consideramos la clase de conceptos $H \subseteq \mathcal{Y}^2$ dada por:

$$H := \{(0, 1), (0, 2), (1, 0), (1, 1), (2, 0)\},$$

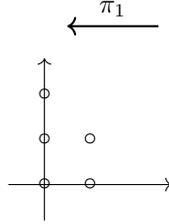
cuya imagen es:



Aplicamos un shift en la dirección de la proyección π_1 (que “olvida” la primera coordenada) y obtenemos:

$$H' := \mathbb{S}_1(H) := \{(0,0), (0,1), (0,2), (1,0), (1,1)\},$$

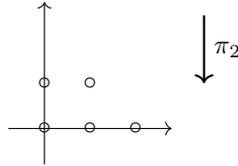
cuya imagen resulta ser:



Ocurre algo similar si aplicamos un shift en la dirección de π_2 :

$$H'' := \mathbb{S}_2(H) := \{(0,0), (0,1), (1,0), (1,1), (2,0)\},$$

cuya imagen viene dada por:



En ambos casos, obtenemos conjuntos cerrados hacia abajo que no son pseudo-cubos, aunque ambas transformaciones contienen un pseudo-cubo de dimensión 2. Por lo tanto, se cumple la siguiente igualdad:

$$\dim_{\text{DS}}(H') = \dim_{\text{E}}(H') = \dim_{\text{DS}}(H'') = \dim_{\text{E}}(H'') = 2 > \dim_{\text{DS}}(H) = 1.$$

Además, en ambos casos, el número de aristas disminuye (esto es estándar en el shifting, como veremos más adelante), y también lo hace el número de aristas unarias. Por consiguiente, tenemos que:

$$1 = 2 - 5/5 = E_{H''}[\text{outdeg}] = E_{H'}[\text{outdeg}] > E_H[\text{outdeg}] = 2 - 6/5 = 4/5$$

y

$$9/5 = 2 - 1/5 = \text{avd}(H'') = \text{avd}(H') > \text{avd}(H) = 2 - 2/5 = 8/5.$$

OBSERVACIÓN 6.4.3. En primer lugar, es evidente que el shifting no modifica las proyecciones de los conjuntos en la dirección determinada por la proyección i -ésima. Es decir, se cumple que:

$$\pi_i(H) = \pi_i(\mathbb{S}_i(H)).$$

Además, dada una proyección π_i , tenemos una partición de la clase de conceptos basada en las fibras de esta proyección:

$$H := \bigsqcup_{f \in \pi_i(H)} e_{f,i},$$

que nos permite construir una biyección natural a partir de las biyecciones sobre cada una de las fibras de la partición. Denotamos a esta biyección por:

$$(6.4.1) \quad \widetilde{\mathbb{S}}_{H,i} : H \longrightarrow \mathbb{S}_i(H),$$

dada la propiedad:

$$\widetilde{\mathbb{S}}_{H,i} \Big|_{e_{f,i}} = \widetilde{\mathbb{S}}_{f,i},$$

para cada $f \in \pi_i(H) = \pi_i(\mathbb{S}_i(H))$. Nótese que $\widetilde{\mathbb{S}}_{H,i}$ es una biyección que preserva el pre-orden de Dickson \preceq y, por lo tanto, preserva los cardinales, i.e.

$$\sharp(H) = \sharp(\mathbb{S}_i(H)).$$

Además, tenemos una biyección entre las aristas asociadas con la dirección de π_i de cada uno de los hiper-grafos de una inclusión asociados con cada uno de los conjuntos, i.e. $\widetilde{\mathbb{S}}_{H,i}$ define una biyección entre $E_i(H)$ y $E_i(\mathbb{S}_i(H))$. En particular, se cumple la siguiente igualdad:

$$\sharp(E_i(H)) = \sharp(E_i(\mathbb{S}_i(H))).$$

Finalmente, también tenemos que $\mathbb{S}_i(H)$ es cerrado hacia abajo en la dirección de π_i . Sin embargo, no debemos olvidar que la biyección entre H y $\mathbb{S}_i(H)$ depende del conjunto H involucrado.

En [Ha, 95], el autor no aplica únicamente un shifting, sino que utiliza secuencias de shiftings como definimos a continuación:

DEFINICIÓN 48 (Secuencias de shiftings). Sea $H \subseteq \mathcal{Y}^n$ una clase de conceptos finita y sea

$$\omega := \omega_1, \dots, \omega_k \in [n]^*$$

una palabra sobre el alfabeto $[n]$. Definimos el shift determinado por la palabra ω como la siguiente composición:

$$(6.4.2) \quad \mathbb{S}_\omega := \mathbb{S}_{\omega_1} \circ \dots \circ \mathbb{S}_{\omega_r} : 2^{\mathcal{Y}^n} \longrightarrow 2^{\mathcal{Y}^n}$$

Por lo tanto, tenemos que:

$$(6.4.3) \quad \mathbb{S}_\omega(H) := \mathbb{S}_{\omega_1}(\mathbb{S}_{\omega_2}(\dots(\mathbb{S}_{\omega_k}(H)))).$$

OBSERVACIÓN 6.4.4 (Las secuencias de shiftings determinan biyecciones que preservan \preceq). Un shift \mathbb{S}_ω determinado por una palabra $\omega = \omega_1 \dots \omega_k \in [n]^*$ determina una biyección entre los conjuntos $H \subseteq \mathcal{Y}^n$ y $\mathbb{S}_\omega(H)$, pero no es literalmente una composición de funciones, sino que toma la siguiente forma:

$$(6.4.4) \quad \widetilde{\mathbb{S}}_{H,\omega} := \widetilde{\mathbb{S}}_{H_1,\omega_1} \circ \widetilde{\mathbb{S}}_{H_2,\omega_2} \circ \dots \circ \widetilde{\mathbb{S}}_{H_k,\omega_k},$$

donde $H_1 = H$ y para cada j , $2 \leq j \leq k$, tenemos que: $H_j := \mathbb{S}_i(H_{j-1})$. Por lo tanto, podemos concluir:

- i) Las aplicaciones $\widetilde{\mathbb{S}}_{H,\omega}$ son biyecciones entre H y $\mathbb{S}_\omega(H)$ que preservan \preceq (porque lo preserva cada una de las biyecciones que interviene en su definición).
- ii) $\sharp(H) = \sharp(\mathbb{S}_\omega(H))$.

OBSERVACIÓN 6.4.5. En [Ha, 95] se demuestra (usando inducción y las observaciones precedentes) que, para cada clase de conceptos finita $H \subseteq \mathcal{Y}^n$, existe una palabra $\omega \in [n]^*$ que define una secuencia de shifts \mathbb{S}_ω , de modo que la clase $H_\omega^* := \mathbb{S}_\omega(H)$ es cerrada hacia abajo con respecto al pre-orden de Dickson \preceq . Nótese que cada H tendrá asociada una o varias palabras $\omega := \omega(H)$ y que no es el mismo ω para todo $H \subseteq \mathcal{Y}^n$. Además, el conjunto cerrado hacia abajo biyectable con H no es único y los invariantes pueden cambiar según la palabra escogida.

Como ejemplo, podemos considerar la clase de conceptos H descrita en el Ejemplo 6.4.2. Tomando la palabra $\omega := 1 \in [n]^*$, obtenemos una clase $H_\omega^* := \mathbb{S}_\omega(H)$ que es cerrada hacia abajo con respecto a \preceq . Análogamente, tomando la palabra $\omega' := 2 \in [n]^*$, tenemos la clase $H_{\omega'}^* := \mathbb{S}_{\omega'}(H) = \mathbb{S}_2(H)$, que es también cerrada hacia abajo. Finalmente, nótese que las clases H_ω^* y $H_{\omega'}^*$, obtenidas son distintas.

6.4.2. Shifting, clases cerradas hacia abajo y bases monomiales. Al final de la Subsección 6.1.2, concluíamos que las escaleras finitas pueden obtenerse como complementarios de conjuntos de exponentes asociados a ciertos ideales, independientemente del orden monomial escogido (ver Corolario 6.1.11). En esta subsección, seguiremos de alguna forma el “camino inverso”.

Fijamos un orden monomial (usaremos órdenes lexicográficos) y tratamos de asociar escaleras finitas con ideales “especiales”. Estos ideales “especiales” resultan ser los ideales de polinomios que se anulan en los puntos de la propia escalera. Estas son las ideas de [Me, 19] y su predecesor [FRR, 06]. La demostración de [Me, 19] se apoya fuertemente en un resultado de [FRR, 06], quien, a su vez, está altamente inspirado en elementos de la Teoría de Juegos (en particular, del *lex game*). Hemos construido nuestra propia demostración del resultado de [Me, 19] usando únicamente elementos propios de la División de Weierstarss-Hironaka y las Bases de Gröbner. Comenzamos estudiando el ideal asociado a un subconjunto finito:

LEMA 6.4.6. Sea K un cuerpo (no necesariamente algebraicamente cerrado), $V \subseteq K^n$ un conjunto finito e $I(V) \subseteq K[X_1, \dots, X_n]$ el ideal dado por:

$$I(V) := \{f \in K[X_1, \dots, X_n] : f|_V \equiv 0\}.$$

Entonces, para cualquier orden monomial \leq sobre el monoide $(\mathbb{N}^n, +)$, tenemos que $\text{Nexp}_{\leq}(I(V))$ es un conjunto finito y:

$$\sharp(V) = \dim_K(K[V]) = \sharp(\text{Nexp}_{\leq}(I(V))).$$

Además, como en el Teorema 6.1.9, el conjunto:

$$\mathcal{B}/I(V) := \{X_1^{\mu_1} \cdots X_n^{\mu_n} + I(V) : \mu := (\mu_1, \dots, \mu_n) \in \text{Nexp}_{\leq}(I(V))\},$$

es una base de $K[V]$ como K -espacio vectorial, denominada “base monomial” asociada al orden monomial \leq y está en biyección con $\text{Nexp}_{\leq}(I(V))$.

DEMOSTRACIÓN. Sea $D = \sharp(V)$. Supongamos que $V = \{\zeta_i : 1 \leq i \leq D\}$. Para cada i , $1 \leq i \leq D$, consideramos el ideal maximal \mathfrak{m}_i asociado al punto afín $\zeta_i \in \mathbb{A}^n(K)$, i.e.

$$\mathfrak{m}_i := I(\{\zeta_i\}) = \{f \in K[X_1, \dots, X_n] : f(\zeta_i) = 0\}.$$

Nótese que \mathfrak{m}_i es el núcleo del siguiente epimorfismo de anillos:

$$\begin{array}{ccc} \text{ev}_{\zeta_i} : K[X_1, \dots, X_n] & \longrightarrow & K \\ f & \longmapsto & f(\zeta_i). \end{array}$$

Por el Teorema Chino de los Restos, concluimos (ya que los ideales \mathfrak{m}_i son co-maximales dos a dos, por ser cada uno de ellos maximal):

- i) $\bigcap_{i=1}^D \mathfrak{m}_i = \prod_{i=1}^D \mathfrak{m}_i$.
- ii) La siguiente aplicación es un isomorfismo de K -álgebras:

$$\begin{array}{ccc} \text{ev}_V : K[X_1, \dots, X_n] / \prod_{i=1}^D \mathfrak{m}_i & \longrightarrow & K^D \\ f + \prod_{i=1}^D \mathfrak{m}_i & \longmapsto & (f(\zeta_1), \dots, f(\zeta_D)). \end{array}$$

En otras palabras, tenemos que:

$$I(V) = \prod_{i=1}^D I(\{\zeta_i\}) = \bigcap_{i=1}^D \mathfrak{m}_i.$$

En particular, $D = \sharp(V) = \dim_K(K[V])$. Como $\text{Nexp}_{\leq}(I(V))$ determina (de forma biyectiva) una base del anillo cociente $K[X_1, \dots, X_n]/I(V)$ como K -espacio vectorial, concluimos:

$$\sharp(V) = \dim_K(K[V]) = \sharp(\text{Nexp}_{\leq}(I(V))).$$

□

Como ya hemos visto en la Proposición 6.1.7, al ser $I(V)$ un ideal del anillo de polinomios $K[X_1, \dots, X_n]$, tenemos que el siguiente conjunto es cerrado hacia abajo:

$$\text{Nexp}_{\leq}(I(V)) = \mathbb{N}^n \setminus \text{exp}_{\leq}(I(V)),$$

para cualquier orden monomial \leq . Dedicaremos el resto de la subsección al estudio de una forma de “recíproco”, para lo cual nos restringiremos al orden lexicográfico canónico \leq_{lex} , introducido en la Definición 36, en el que las variables están ordenadas de la siguiente forma:

$$X_1 >_{\text{lex}} X_2 >_{\text{lex}} \cdots >_{\text{lex}} X_n.$$

Con el objetivo de simplificar la terminología, emplearemos las siguientes notaciones para cualquier ideal $\mathfrak{a} \subseteq K[X_1, \dots, X_n]$:

$$\text{exp}_{\text{lex}}(\mathfrak{a}) := \text{exp}_{\leq_{\text{lex}}}(\mathfrak{a}), \quad \text{Nexp}_{\text{lex}}(\mathfrak{a}) := \mathbb{N}^n \setminus \text{exp}_{\leq_{\text{lex}}}(\mathfrak{a}).$$

Sea $\mathcal{Y} := \{0, \dots, p-1\}$. Supongamos que el cuerpo K verifica $\mathcal{Y} \subseteq K$ (con cualquier inmersión razonable) y sea $H \subseteq \mathcal{Y}^n$ una clase de conceptos finita. Para cada $i \in [n]$, a diferencia de las proyecciones usadas en las secciones anteriores, consideramos la proyección $\tilde{\pi}_i : \mathcal{Y}^n \longrightarrow \mathcal{Y}^{n-i}$ que “olvida” las primeras i coordenadas de los elementos de \mathcal{Y}^n , i.e.

$$\tilde{\pi}_i(x_1, \dots, x_n) := (x_{i+1}, \dots, x_n), \quad \forall (x_1, \dots, x_n) \in \mathcal{Y}^n.$$

Dado $q := (q_1, \dots, q_{n-i}) \in \mathcal{Y}^{n-i}$, denotamos por $H(q)$ a la parte i -inicial de la fibra en H sobre q a partir de $\tilde{\pi}_i$, i.e.

$$(6.4.5) \quad H(q) := \{(\theta_1, \dots, \theta_i) \in \mathcal{Y}^i : (\theta_1, \dots, \theta_i, q_1, \dots, q_{n-i}) \in \tilde{\pi}_i^{-1}(\{q\}) \cap H\}.$$

LEMA 6.4.7 (de eliminación cuantificada). Con las notaciones anteriores, sea $\nu := (\nu_1, \dots, \nu_n) \in \mathbb{N}^n$ un exponente monomial. Supongamos que se verifica:

$$\sharp(\{q \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{\text{lex}}(I(H(q)))\}) \geq \nu_n + 1.$$

Entonces, $\nu \in \text{Nexp}_{\text{lex}}(I(H))$.

DEMOSTRACIÓN. El caso $n = 1$ es inmediato y no requiere discusión adicional.

Consideramos el caso $n \geq 2$ y definimos:

$$B_1 := \{q \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{1\text{ex}}(I(H(q)))\}.$$

Supongamos que $\#(B_1) \geq \nu_n + 1$ y procedemos por reducción al absurdo. Supongamos que $\nu \in \text{exp}_{1\text{ex}}(I(H))$. Entonces, existe un polinomio $f \in I(H)$ de la siguiente forma:

$$(6.4.6) \quad f(X_1, \dots, X_n) := X_1^{\nu_1} \cdots X_n^{\nu_n} + g(X_1, \dots, X_n),$$

donde los términos no nulos de g tienen exponentes estrictamente menores que ν con respecto a $\leq_{1\text{ex}}$, i.e.

$$(6.4.7) \quad g(X_1, \dots, X_n) := \sum_{\mu <_{1\text{ex}} \nu} a_\mu X_1^{\mu_1} \cdots X_n^{\mu_n}.$$

A continuación, distinguimos los siguientes casos para los términos de g :

- *Caso 1:* Términos cuyo exponente monomial es de la forma $(\nu_1, \dots, \nu_{n-1}, k)$, donde $k \leq \nu_n - 1$.
- *Caso 2:* $\mu := (\mu_1, \dots, \mu_{n-1}, r) \in \mathbb{N}^n$, para cualquier r , pero verificando:

$$(\mu_1, \dots, \mu_{n-1}) <_{1\text{ex}} (\nu_1, \dots, \nu_{n-1}).$$

Podemos agrupar los términos correspondientes al Caso 1 y considerar el siguiente polinomio univariado:

$$h_n(X_n) := X_n^{\nu_n} + \sum_{k=0}^{\nu_n-1} a_{(\nu_1, \dots, \nu_{n-1}, k)} X_n^k.$$

Entonces, reescribimos el polinomio f descrito en la Identidad (6.4.6) de la siguiente manera:

$$f(X_1, \dots, X_n) := X_1^{\nu_1} \cdots X_{n-1}^{\nu_{n-1}} \cdot h_n(X_n) + h(X_1, \dots, X_n),$$

donde todos los términos no nulos de h pertenecen al Caso 2. En otras palabras, h tiene la siguiente forma:

$$h(X_1, \dots, X_n) := \sum_{\theta \in J} a_\theta X_1^{\theta_1} \cdots X_n^{\theta_n},$$

donde $J \subseteq \mathbb{N}^n$ es un conjunto finito y $\theta := (\theta_1, \dots, \theta_n) \in J$ si y solo si:

$$(\theta_1, \dots, \theta_{n-1}) <_{1\text{ex}} (\nu_1, \dots, \nu_{n-1}).$$

Consideramos ahora $b \in \mathcal{Y}$ y el polinomio:

$$f(X_1, \dots, X_{n-1}, b) := X_1^{\nu_1} \cdots X_{n-1}^{\nu_{n-1}} \cdot h_n(b) + h(X_1, \dots, X_{n-1}, b).$$

Debido a la forma en que hemos construido f , tenemos que:

- $f(X_1, \dots, X_{n-1}, b)$ se anula en $H(b)$.
- Si $h_n(b) \neq 0$, el monomio director de $f(X_1, \dots, X_{n-1}, b)$ con respecto a $\leq_{1\text{ex}}$ es $(\nu_1, \dots, \nu_{n-1})$.

Por lo tanto, si $h_n(b) \neq 0$, entonces:

$$(\nu_1, \dots, \nu_{n-1}) \in \text{exp}_{1\text{ex}}(I(H(b)))$$

y, por lo tanto, $b \notin B_1$. En conclusión,

$$B_1 \subseteq \{b \in \mathcal{Y} : h_n(b) = 0\},$$

pero como h_n es un polinomio univariado de grado ν_n , el número de posibles raíces está acotado por ν_n , esto es,

$$\#(B_1) \leq \nu_n,$$

lo que contradice nuestra hipótesis. Esto completa la demostración del lema. \square

LEMA 6.4.8 (existencia de factores triangulares). *Con las notaciones precedentes, sea $\nu = (\nu_1, \dots, \nu_n) \in \mathbb{N}^n$ un exponente monomial. Supongamos que $n \geq 2$ y que se verifica lo siguiente:*

$$\#\{q \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{1\text{ex}}(I(H(q)))\} \leq \nu_n.$$

Entonces, existen $g_1, \dots, g_n \in K[X_1, \dots, X_n]$ tales que:

- $g_i \in K[X_i, X_{i+1}, \dots, X_n]$ es un polinomio mónico con respecto a la variable X_i que satisface $\deg_{X_i}(g_i) \leq \nu_i$, para cada i , $1 \leq i \leq n$.
- El polinomio:

$$h := \prod_{i=1}^n g_i$$

se anula en H , esto es, $h \in I(H)$.

En particular,

$$(\nu_1, \dots, \nu_n) = \exp_{\leq \text{lex}}(h) \in \exp_{\text{lex}}(I(H)).$$

DEMOSTRACIÓN. Procedemos por inducción en n . Comenzamos con el caso $n = 2$. Definimos el conjunto $B_1 \subseteq \mathcal{Y}$ como:

$$B_1 := \{b \in \mathcal{Y} : \nu_1 \in \text{Nexp}_{\text{lex}}(I(H(b)))\},$$

y supongamos que $\sharp(B_1) \leq \nu_2$. Entonces, existe un polinomio mónico $g_2 \in K[X_2]$ de grado acotado por ν_2 de modo que g_2 se anula en B_1 . Nótese que, en particular, g_2 se anula en:

$$\bigcup_{b \in B_1} H(b) \times \{b\}.$$

Consideramos ahora el conjunto:

$$B_2 := \mathcal{Y} \setminus B_1 = \{a \in \mathcal{Y} : \nu_1 \in \exp_{\text{lex}}(I(H(a)))\}.$$

Como $\nu_1 \in \exp_{\text{lex}}(I(H(a)))$, para cada $a \in B_2$, existe un polinomio mónico univariado $g_1^{(a)} \in K[X_1]$ de grado $\deg(g_1^{(a)}) \leq \nu_1$ tal que $g_1^{(a)}$ se anula en $H(a)$. Esto es, para cada $a \in B_2$, podemos escribir, sin pérdida de generalidad:

$$g_1^{(a)}(X_1) = X_1^{\nu_1} + \sum_{k=0}^{\nu_1-1} a_k^{(a)} X_1^k.$$

Si el grado $d_1 := \deg(g_1^{(a)})$ fuese estrictamente menor que ν_1 , podríamos reemplazarlo por:

$$X_1^{\nu_1 - \deg(g_1^{(a)})} \cdot g_1^{(a)}(X_1).$$

A continuación, aplicamos interpolación de Lagrange. Por lo tanto, consideramos los siguientes polinomios:

$$g_{1,k}(X_2) := \sum_{a \in B_2} a_k^{(a)} \frac{\prod_{a' \in B_2 \setminus \{a\}} (X_2 - a')}{\prod_{a' \in B_2 \setminus \{a\}} (a - a')}.$$

Nótese que $g_{1,k}(a) = a_k^{(a)}$ para cada $a \in B_2$. Finalmente, consideramos el siguiente polinomio:

$$g_1(X_1, X_2) := X_1^{\nu_1} + \sum_{k=0}^{\nu_1-1} g_{1,k}(X_2) \cdot X_1^k.$$

Además, tenemos que $g_1(X_1, a) = g_1^{(a)}(X_1)$, $\forall a \in B_2$ y, por lo tanto, g_1 se anula en la siguiente unión:

$$\bigcup_{a \in B_2} H(a) \times \{a\}.$$

En particular, el polinomio $h = g_1(X_1, X_2) \cdot g_2(X_2)$ se anula en todo H ya que:

$$H = \left(\bigcup_{b \in B_1} H(b) \times \{b\} \right) \cup \left(\bigcup_{a \in B_2} H(a) \times \{a\} \right).$$

Por otro lado, es evidente que usando el orden lexicográfico canónico obtenemos:

$$\text{LM}_{\leq \text{lex}}(h) = (\nu_1, \nu_2),$$

ya que $h(X_1, X_2) = X_1^{\nu_1} X_2^{\nu_2} + \ell$, donde los términos no nulos de ℓ tienen grado estrictamente menor que (ν_1, ν_2) con respecto al orden lexicográfico canónico.

Supongamos ahora que $n > 2$. Al igual que antes, definimos el conjunto:

$$B_1 := \{b \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{\text{lex}}(I(H(b)))\}.$$

y supongamos que $\sharp(B_1) \leq \nu_n$. Como en el caso $n = 2$, existe un polinomio univariado $g_n \in K[X_n]$ de grado ν_n , mónico con respecto a la variable X_n , tal que g_n se anula en B_1 . En particular, tenemos que:

$$\bigcup_{b \in B_1} H(b) \times \{b\} \subseteq V_{\mathbb{A}^n(K)}(g_n) := \{\zeta = (\zeta_1, \dots, \zeta_n) \in \mathcal{Y}^n : g_n(\zeta) = g_n(\zeta_n) = 0\}.$$

Consideramos ahora el conjunto $B_2 := \mathcal{Y} \setminus B_1$. Por definición, tenemos que:

$$B_2 := \{a \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \exp_{\text{lex}}(I(H(a)))\}.$$

Dado $a \in B_1$, si aplicamos el Lema 6.4.7 anterior a $H(a)$ (ya que $(\nu_1, \dots, \nu_{n-1}) \in \exp_{\text{lex}}(I(H(a)))$), entonces el siguiente conjunto no puede tener cardinal mayor o igual que $\nu_{n-1} + 1$:

$$B_1(a) := \{c \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-2}) \in \text{Nexp}_{\text{lex}}(I(H(a, c)))\}.$$

Como $\sharp(B_1(a)) \leq \nu_{n-1}$, aplicando la hipótesis de inducción, existirán polinomios

$$g_1^{(a)}, \dots, g_{n-1}^{(a)} \in K[X_1, \dots, X_{n-1}]$$

tales que:

- i) $g_i^{(a)} \in K[X_i, X_{i+1}, \dots, X_{n-1}]$ es un polinomio mónico con respecto a la variable X_i que verifica $\deg_{X_i}(g_i^{(a)}) \leq \nu_i$, para cada i , $1 \leq i \leq n$.
- ii) El polinomio:

$$h^{(a)} := \prod_{i=1}^{k-1} g_i^{(a)}$$

se anula en $H(a)$, i.e. $h^{(a)} \in I(H(a))$.

Como $B_2 \subseteq \mathcal{Y} \subseteq K$ y K es un cuerpo, podemos considerar los anillos $R_i := K[X_i, \dots, X_{n-1}]$ y aplicar interpolación de Lagrange sobre $R_i[X_n]$ siempre que el denominador sea una constante no nula en K . En otras palabras, supongamos que para cada $a \in B_2$, tenemos que:

$$g_i^{(a)} = X_i^{\nu_i} + \sum_{k=0}^{\nu_i-1} g_{i,k}^{(a)} \cdot X_i^k,$$

donde $g_{i,k}^{(a)} \in R_{i+1} = K[X_{i+1}, \dots, X_n]$ son los coeficientes con respecto a la variable X_i . Aplicando interpolación de Lagrange, consideramos los siguientes polinomios:

$$g_{i,k}(X_{i+1}, \dots, X_n) := \sum_{a \in B_2} g_{i,k}^{(a)}(X_{i+1}, \dots, X_{n-1}) \cdot \frac{\prod_{a' \in B_2 \setminus \{a\}} (X_n - a')}{\prod_{a' \in B_2 \setminus \{a\}} (a - a')}.$$

Nótese que podemos hacer esta interpolación porque el discriminante es un elemento no nulo de K . Además, tenemos que:

- i) $g_{i,k} \in K[X_{i+1}, \dots, X_n]$.
- ii) Para cada $a \in B_2$:

$$g_{i,k}(X_{i+1}, \dots, X_{n-1}, a) = g_{i,k}^{(a)}(X_{i+1}, \dots, X_{n-1}).$$

Finalmente, definimos el siguiente polinomio:

$$g_i(X_i, \dots, X_n) := X_i^{\nu_i} + \sum_{k=0}^{\nu_i-1} g_{i,k}(X_{i+1}, \dots, X_n) \cdot X_i^k.$$

Nótese que:

$$g_i(X_i, X_{i+1}, \dots, X_{n-1}, a) = g_i^{(a)}(X_i, \dots, X_{n-1}).$$

Definiendo:

$$h_0 := \prod_{i=1}^{n-1} g_i \in K[X_1, \dots, X_n],$$

concluimos que $\forall a \in B_2$ y $\forall (\zeta_1, \dots, \zeta_{n-1}) \in H(a)$, tenemos que:

$$h_0(\zeta_1, \dots, \zeta_{n-1}, a) = \prod_{i=1}^{n-1} g_i(\zeta_1, \dots, \zeta_{n-1}, a) = \prod_{i=1}^{n-1} g_i^{(a)}(\zeta_1, \dots, \zeta_{n-1}) = h^{(a)}(\zeta_1, \dots, \zeta_{n-1}) = 0.$$

Por lo tanto, $h_0 \in I(H(a) \times \{a\})$ para todo $a \in B_2$. Consideramos ahora la familia de polinomios $g_1, \dots, g_n \in K[X_1, \dots, X_n]$. Tenemos que:

$$g_i \in K[X_i, X_{i+1}, \dots, X_n],$$

y g_i es un polinomio mónico con respecto a X_i de grado $\deg_{X_i}(g_i) \leq \nu_i$. Si definimos:

$$h := h_0 \cdot g_n(X_n) = \prod_{i=1}^n g_i \in K[X_1, \dots, X_n],$$

concluimos que para cada $(\zeta_1, \dots, \zeta_n) \in H$, tenemos que:

- i) Si $\zeta_n \in B_1$, entonces $g_n(\zeta_n) = 0$, i.e. g_n se anula en $H(\zeta_n)$. Por lo tanto, $h(\zeta_1, \dots, \zeta_n) = 0$.
- ii) Si, por otro lado, $\zeta_n \in B_2 = \mathcal{Y} \setminus B_1$, entonces $(\zeta_1, \dots, \zeta_{n-1}) \in H(\zeta_n)$. Por lo tanto,

$$h^{(\zeta_n)}(\zeta_1, \dots, \zeta_{n-1}) = 0$$

y, en particular:

$$h_0(\zeta_1, \dots, \zeta_n) = h^{(\zeta_n)}(\zeta_1, \dots, \zeta_{n-1}) = 0.$$

Nótese que se verifican las condiciones para el caso n . Por lo tanto, $h \in I(H)$ y se deduce de forma inmediata que:

$$\text{LM}_{\leq 1\text{ex}}(h) = (\nu_1, \dots, \nu_n).$$

□

Los dos lemas anteriores nos permiten concluir el siguiente teorema debido a [FRR, 06]:

TEOREMA 6.4.9 (Corolario 8 de [FRR, 06]). *Sea \mathcal{Y} un conjunto finito y K un cuerpo. Supongamos que $\mathcal{Y} \subseteq K$ (con cualquier inmersión razonable) y sea $H \subseteq \mathcal{Y}^n$. Los siguientes ítems son equivalentes para un exponente monomial $\nu := (\nu_1, \dots, \nu_n) \in \mathbb{N}^n$:*

- i) $\nu \in \text{Nexp}_{1\text{ex}}(I(H))$,
- ii) se verifica la siguiente desigualdad:

$$\#\{q \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{1\text{ex}}(I(H(q)))\} \geq \nu_n + 1.$$

DEMOSTRACIÓN. Cada implicación es uno de los lemas anteriores:

- $ii) \Rightarrow i)$ es el Lema 6.4.7.
- $i) \Rightarrow ii)$ es el contrarecíproco del Lema 6.4.8 tras recordar que:

$$\text{Nexp}_{1\text{ex}}(I(H)) = \mathbb{N}^n \setminus \text{exp}_{1\text{ex}}(I(H)).$$

□

6.4.2.1. *Las ideas de [Me, 19].* Al igual que en las secciones previas, sea $\mathcal{Y} := \{0, \dots, p-1\}$ un conjunto finito, K un cuerpo que contiene a \mathcal{Y} (esto es, suponemos que la característica de K es mayor o igual que p o cero) y $H \subseteq \mathcal{Y}^n$ una clase conceptos finita. Consideramos la estructura de hiper-grafo de una inclusión discutida en la Sección 6.2 y los operadores de shifting:

$$\mathbb{S}_i : 2^{\mathcal{Y}^n} \longrightarrow 2^{\mathcal{Y}^n},$$

para cada $i \in [n]$ definidos previamente. Finalmente, para cada $b \in \mathcal{Y}$, sea $H(b)$ el conjunto definido en la Identidad (6.4.5):

$$H(b) := \{(q_1, \dots, q_{n-1}) \in \mathcal{Y}^{n-1} : (q_1, \dots, q_{n-1}, b) \in H\}.$$

Dados $H \subseteq \mathcal{Y}^n$ y $b \in \mathcal{Y}$, tenemos las siguientes estructuras de hiper-grafo de una inclusión:

- $G(H) := (H, E(H))$,
- $G(H(b)) := (H(b), E(H(b)))$.

Sobre ambos hiper-grafos, podemos definir las correspondientes operaciones de shifting y las palabras asociadas en $[n]^*$. En particular, estamos interesados en los siguientes dos conjuntos:

$$\mathbb{S}_{x'}(H), \quad \mathbb{S}_{x'}(H(b)),$$

donde: $x' := n-1 \ n-2 \ \dots \ 2 \ 1 \in [n]^*$. La siguiente proposición nos permite relacionar el resultado de ambos operadores de shifting:

PROPOSICIÓN 6.4.10. *Con las mismas notaciones que antes, tenemos que:*

$$\mathbb{S}_{x'}(H(b)) \times \{b\} \subseteq \mathbb{S}_{x'}(H).$$

Además, $\mathbb{S}_{x'} \times \{b\}$ está contenido en la siguiente fibra:

$$\{(\mu_1, \dots, \mu_n) \in \mathbb{S}_{x'}(H) : \mu_n = b\} \subseteq \mathbb{S}_{x'}(H).$$

DEMOSTRACIÓN. Basta observar que los shifts involucrados en $\mathbb{S}_{x'}$ no afectan a la última coordenada de los elementos de \mathcal{Y}^n , que permanece inalterada. □

Finalmente, enunciamos y demostramos el teorema que justifica esta subsección.

TEOREMA 6.4.11 (Teorema 2 de [Me, 19]). *Con las notaciones precedentes, tenemos que:*

$$(6.4.8) \quad \text{Nexp}_{1\text{ex}}(I(H)) = \mathbb{S}_{n \dots 1}(H).$$

DEMOSTRACIÓN. Procedemos por inducción en n . El caso $n = 1$ es evidente, por lo que supondremos que $n \geq 2$. En primer lugar, combinando la Observación 6.4.4 con el Lema 6.4.6, obtenemos la siguiente cadena de igualdades:

$$\#\{\mathbb{S}_{n \dots 1}(H)\} = \#\{H\} = \dim_K(K[H]) = \#\{\text{Nexp}_{1\text{ex}}(I(H))\}.$$

Por lo tanto, será suficiente con probar una de las inclusiones de la Identidad (6.4.8) para obtener la igualdad ya que son conjuntos que tienen el mismo cardinal. Por consiguiente, nos limitamos a demostrar la siguiente inclusión:

$$\text{Nexp}_{1\text{ex}}(I(H)) \subseteq \mathbb{S}_{n \ n-1 \ \dots \ 1}(H).$$

Consideramos $(\nu_1, \dots, \nu_n) \in \text{Nexp}_{1\text{ex}}(I(H))$ y la palabra:

$$x' := n - 1 \ n - 2 \ \cdots \ 2 \ 1 \in [n]^*.$$

Definimos el conjunto:

$$B_1 := \{b \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{1\text{ex}}(I(H(b)))\}.$$

Por el Teorema 6.4.9 anterior, tenemos que:

$$\#(B_1) = \#(\{b \in \mathcal{Y} : (\nu_1, \dots, \nu_{n-1}) \in \text{Nexp}_{1\text{ex}}(I(H(b)))\}) \geq \nu_n + 1.$$

Por otro lado, aplicando la hipótesis de inducción, para cada $b \in \mathcal{Y}$ tal que $H(b) \neq \emptyset$, tenemos que:

$$\text{Nexp}_{1\text{ex}}(I(H(b))) = \mathbb{S}_{x'}(H(b)).$$

Como $x' = n - 1 \ n - 2 \ \cdots \ 2 \ 1 \in [n]^*$, la Proposición 6.4.10 implica que para cada $b \in B_1$, obtenemos:

$$(6.4.9) \quad (\nu_1, \dots, \nu_{n-1}, b) \in \mathbb{S}_{x'}(H(b)) \times \{b\} \subseteq \mathbb{S}_{x'}(H).$$

Tomando $f := (\nu_1, \dots, \nu_{n-1}) \in \mathcal{Y}^{n-1}$, podemos considerar la arista $e_{f,n}$ del hiper-grafo de una inclusión $G(\mathbb{S}_{x'}(H))$ dada por:

$$e_{f,n} = \mathbb{S}_{x'}(H) \cap \pi_n^{-1}(\{f\}),$$

donde $\pi_n : \mathcal{Y}^n \rightarrow \mathcal{Y}^{n-1}$ es la proyección que “olvida” la última coordenada. Por la inclusión descrita en la Identidad (6.4.9), esto significa que:

$$\{f\} \times \{b\} = (\nu_1, \dots, \nu_{n-1}, b) \in \mathbb{S}_{x'}(H)$$

y, obviamente, π_n proyecta $(\nu_1, \dots, \nu_{n-1}, b)$ sobre f . Esto es, hemos probado la siguiente inclusión:

$$e_{f,n} \supseteq \{(\nu_1, \dots, \nu_{n-1})\} \times B_1.$$

Por lo tanto, concluimos la siguiente desigualdad de cardinales:

$$\#(e_{f,n}) \geq \#(B_1) \geq \nu_n + 1.$$

Entonces,

$$\mathbb{S}_n(e_{f,n}) \supseteq \{(\nu_1, \dots, \nu_{n-1}, t) : 0 \leq t \leq \nu_n\}.$$

Por lo tanto,

$$(\nu_1, \dots, \nu_n) \in \mathbb{S}_n(\mathbb{S}_{x'}(H)) = \mathbb{S}_{n \ n-1 \ \dots \ 1}(H).$$

De lo anterior, se deduce que:

$$\text{Nexp}_{1\text{ex}}(I(H)) \subseteq \mathbb{S}_{n \ n-1 \ \dots \ 1}(H),$$

y el resultado se sigue debido a que ambos conjuntos tienen el mismo cardinal. \square

En el Corolario 6.1.11 obtuvimos una caracterización de los conjuntos finitos cerrados hacia abajo como complementarios de conjuntos de exponentes monomiales de ideales cuyos anillos residuales son K -álgebras de Artin. El siguiente resultado nos permite ser más precisos:

COROLARIO 6.4.12 (H es un conjunto cerrado hacia abajo si y solo si $H = \text{Nexp}_{1\text{ex}}(I(H))$). Con las notaciones precedentes, un subconjunto $H \subseteq \mathcal{Y}^n$ es cerrado hacia abajo si y solo si:

$$H = \text{Nexp}_{1\text{ex}}(I(H)),$$

donde $I(H)$ es el ideal de todos los polinomios $p \in K[X_1, \dots, X_n]$ que se anulan en H .

DEMOSTRACIÓN. En primer lugar, en la Proposición 6.1.7 ya demostramos que los conjuntos de la forma $\text{Nexp}_{1\text{ex}}(I(H))$ son cerrados hacia abajo. De hecho, lo probamos para cualquier ideal. Por otro lado, en el Corolario 6.1.11 probamos que todo conjunto finito cerrado hacia abajo es el complementario de $\text{exp}_{\leq}(\mathfrak{a})$ para cualquier orden monomial y un ideal \mathfrak{a} construido ad hoc.

Nótese que si H es cerrado hacia abajo, entonces es invariante por la acción de los shifts. Es decir, para cualquier $i \in [n]$, tenemos que $\mathbb{S}_i(H) = H$. Entonces, usando inducción, podemos demostrar fácilmente que si H es cerrado hacia abajo, tenemos que:

$$\mathbb{S}_{n \ \dots \ 1}(H) = H.$$

El Teorema 6.4.11 especifica un orden y un ideal monomial cuya base monomial coincide con H , esto es,

$$H = \text{Nexp}_{1\text{ex}}(I(H)),$$

lo que prueba la equivalencia buscada. \square

6.4.3. Shifting e invariantes de clases de conceptos finitas. En esta subsección trataremos de completar el estudio, iniciado en la Subsección 6.4.1, sobre el comportamiento de los invariantes del hiper-grafo de una inclusión después aplicar una operación de shifting.

Como ya vimos en el Corolario 6.4.1, la dimensión exponencial no se incrementa al realizar una operación de shifting. Por otro lado, en el Ejemplo 6.4.2 hemos observado que tanto el grado promedio como la dimensión DS pueden incrementarse.

En particular, demostraremos que el grado de salida promedio no se incrementa tras un shift (Afirmación 26 de [BCDMY, 22]) y obtendremos algunos resultados que nos permitirán entender mejor el comportamiento “impredecible” de grado promedio tras un shift. Estos resultados nos permitirán mejorar la Proposición 27 y el Corolario 28 de [BCDMY, 22]. Además, ampliamos el Lema 3 de [DS, 14] proporcionando una cota superior que es estable por shifting y que permite controlar el $\text{outdeg}_{\max}(\sigma)$.

6.4.3.1. *Shifting: grado de salida promedio.* Nuestro primer objetivo será analizar el comportamiento del grado de salida promedio después de aplicar un shift.

Dados $i, j \in [n]$, con $i \neq j$ y, $H \subseteq \mathcal{Y}^{[n]}$, denotamos por $\pi_{i,j}$ a la proyección que “olvida” las coordenadas i y j , i.e.

$$(6.4.10) \quad \begin{aligned} \pi_{i,j} : \mathcal{Y}^{[n]} &\longrightarrow \mathcal{Y}^{[n] \setminus \{i,j\}} \\ h &\longmapsto h|_{[n] \setminus \{i,j\}}, \end{aligned}$$

donde $h|_{[n] \setminus \{i,j\}}$ quiere decir que hemos eliminado las coordenadas i y j del h original.

Comenzamos introduciendo el siguiente lema cuya demostración se sigue de las definiciones:

LEMA 6.4.13. *Con las notaciones anteriores, dados $i \neq j$, tenemos que:*

$$\pi_{i,j}(H) = \pi_{i,j}(\mathbb{S}_i(H)).$$

Introducimos a continuación dos subconjuntos para $\ell \in \pi_{i,j}(H)$:

$$(6.4.11) \quad A_{\ell,i} := \bigcup_{\substack{f \in \pi_i(H) \\ f|_{[n] \setminus \{i,j\}} = \ell}} e_{f,i}.$$

$$(6.4.12) \quad B_{\ell,i} := \bigcup_{\substack{f \in \pi_i(H) \\ f|_{[n] \setminus \{i,j\}} = \ell}} \mathbb{S}_i(e_{f,i}).$$

Nótese que las uniones (de aristas en la dirección determinada por π_i) que aparecen en el lado derecho de las Igualdades (6.4.11) y (6.4.12) son disjuntas. Además, tenemos que:

LEMA 6.4.14. *Con las notaciones precedentes, las dos siguientes igualdades muestran respectivamente particiones de H y $\mathbb{S}_i(H)$:*

$$(6.4.13) \quad H := \bigsqcup_{\ell \in \pi_{i,j}(H)} A_{\ell,i}.$$

$$(6.4.14) \quad \mathbb{S}_i(H) := \bigsqcup_{\ell \in \pi_{i,j}(H)} B_{\ell,i}.$$

DEMOSTRACIÓN. En primer lugar, vamos a demostrar que ambas uniones son disjuntas:

- Dados $\ell_1, \ell_2 \in \pi_{i,j}(H)$, si $A_{\ell_1,i} \cap A_{\ell_2,i} \neq \emptyset$, entonces existe $h \in A_{\ell_1,i} \cap A_{\ell_2,i}$. Por lo tanto, de acuerdo con la Identidad (6.4.11), tenemos que:

$$\ell_1 = \pi_i(h)|_{[n] \setminus \{i,j\}} = \pi_{i,j}(h) = \ell_2.$$

Por consiguiente, la parte derecha de la Identidad (6.4.13) describe una unión disjunta.

- Análogamente, dados $\ell_1, \ell_2 \in \pi_{i,j}(H) = \pi_{i,j}(\mathbb{S}_i(H))$, si existe $h' \in B_{\ell_1,i} \cap B_{\ell_2,i} \neq \emptyset$, entonces tenemos que:

$$\ell_1 = \pi_{i,j}(h') = \ell_2,$$

y la parte derecha de la Identidad (6.4.14) también describe una unión disjunta.

Obviamente, $A_{\ell,i} \subseteq H$ y $B_{\ell,i} \subseteq \mathbb{S}_i(H)$. De lo anterior, se deducen las inclusiones en la dirección \supseteq en ambas identidades.

Respecto a las otras inclusiones, probaremos únicamente la segunda ya que la primera se obtiene utilizando argumentos similares. Sea $h' \in \mathbb{S}_i(H)$. Consideramos la proyección $f := \pi_i(h')$ y tenemos que $h' \in \mathbb{S}_i(e_{f,i})$. Por lo tanto, tomando $\ell := f|_{[n] \setminus \{i,j\}} = \pi_{i,j}(h)$, tenemos que:

$$h' \in \mathbb{S}_i(e_{f,i}) \subseteq B_{\ell,i},$$

y se deduce la inclusión en la dirección \subseteq . □

LEMA 6.4.15. *Con las mismas notaciones que antes, dados $i, j \in [n]$, con $i \neq j$, tenemos que:*

$$(6.4.15) \quad \#(\pi_j(H)) = \sum_{\ell \in \pi_{i,j}(H)} \#(\pi_j(A_{\ell,i})),$$

$$(6.4.16) \quad \#(\pi_j(\mathbb{S}_i(H))) = \sum_{\ell \in \pi_{i,j}(H)} \#(\pi_j(B_{\ell,i})).$$

DEMOSTRACIÓN. Por las definiciones y los resultados anteriores, lo único que tenemos que demostrar es que π_j respeta las uniones disjuntas descritas en las Identidades (6.4.13) y (6.4.14). Es decir, tenemos que probar que las siguientes uniones son disjuntas:

$$(6.4.17) \quad \pi_j(H) = \bigsqcup_{\ell \in \pi_{i,j}(H)} \pi_j(A_{\ell,i}),$$

$$(6.4.18) \quad \pi_j(\mathbb{S}_i(H)) = \bigsqcup_{\ell \in \pi_{i,j}(H)} \pi_j(B_{\ell,i}).$$

Estas uniones disjuntas se deducen a partir de las siguientes afirmaciones:

- Si $\pi_j(A_{\ell_1,i}) \cap \pi_j(A_{\ell_2,i}) \neq \emptyset$, entonces $\ell_1 = \ell_2$.
- Si $\pi_j(B_{\ell_1,i}) \cap \pi_j(B_{\ell_2,i}) \neq \emptyset$, entonces $\ell_1 = \ell_2$.

Ambas afirmaciones se demuestran empleando argumentos similares. Si existe $g \in \pi_j(A_{\ell_1,i}) \cap \pi_j(A_{\ell_2,i}) \neq \emptyset$ (respectivamente, $g' \in \pi_j(B_{\ell_1,i}) \cap \pi_j(B_{\ell_2,i}) \neq \emptyset$) es debido a que existe $f \in \pi_i(H)$ (resp. $f' \in \pi_i(\mathbb{S}_i(H)) = \pi_i(H)$) tal que $h \in e_{f,i}$ (resp. $h' \in \mathbb{S}_i(e_{f,i})$) y $\ell_1 = f|_{[n] \setminus \{i,j\}} = \pi_{i,j}(h) = \ell_2$ (resp. $\ell_1 = f'|_{[n] \setminus \{i,j\}} = \pi_{i,j}(h') = \ell_2$). Las igualdades entre los cardinales se siguen de manera obvia por tratarse de uniones disjuntas. □

Introducimos ahora dos cantidades relacionadas con la proyección $\pi_{i,j}$. Dados $i, j \in [n]$, $i \neq j$, y $\ell \in \pi_{i,j}(H)$, denotamos por $H(\ell)$ a la fibra de ℓ en H :

$$H(\ell) := \pi_{i,j}^{-1}(\ell) \cap H := \{h \in H : \pi_{i,j}(h) = \ell\}.$$

Nótese que:

$$\pi_i(H(\ell)) = \{\pi_i(h) : h \in H, h|_{[n] \setminus \{i,j\}} = \ell\}.$$

DEFINICIÓN 49. *Con las notaciones anteriores, dados $i, j \in [n]$, $i \neq j$, y $\ell \in \pi_{i,j}(H)$, definimos el máximo de los cardinales de las aristas que se proyectan sobre ℓ como:*

$$(6.4.19) \quad \mu_i(\ell) := \max\{\#(e_{f,i}) : f \in \pi_i(H(\ell))\}.$$

DEFINICIÓN 50. *Con las mismas notaciones, dados $i, j \in [n]$, $i \neq j$, y $\ell \in \pi_{i,j}(H)$, elijamos $f_0 \in \pi_i(H) \cap H(\ell)$ tal que $\#(e_{f_0,i}) = \mu_i(\ell)$. Definimos el “segundo” cardinal máximo de las aristas que se proyectan sobre ℓ de la siguiente forma:*

- Si $\#(\pi_i(H(\ell))) \geq 2$, entonces definimos:

$$(6.4.20) \quad \mu_i^{(2)}(\ell) := \max\{\#(e_{f,i}) : f \in \pi_i(H(\ell)), f \neq f_0\},$$

- en otro caso, esto es, si $\#(\pi_i(H(\ell))) = 1$, definimos $\mu_i^{(2)}(\ell) := 0$.

Los siguientes lemas explican, parcialmente, las variaciones en el número de aristas del hiper-grafo de una inclusión después de aplicar un shift.

LEMA 6.4.16. *Con las notaciones precedentes, tenemos que:*

$$\#(\pi_j(\mathbb{S}_i(H))) = \sum_{\ell \in \pi_{i,j}(H)} \mu_i(\ell).$$

Además, se verifica la siguiente igualdad:

$$\#(E(\mathbb{S}_i(H))) = \#(E_i(H)) + \sum_{j \neq i} \left(\sum_{\ell \in \pi_{i,j}(H)} \mu_i(\ell) \right).$$

DEMOSTRACIÓN. Como $\pi_{i,j}(H) = \pi_{i,j}(\mathbb{S}_i(H))$, de acuerdo con el Lemma 6.4.15, tenemos que:

$$\#(\pi_j(\mathbb{S}_i(H))) = \sum_{\ell \in \pi_{i,j}(H)} \#(\pi_j(B_{\ell,i})).$$

Para terminar la demostración de la primera igualdad del lema, únicamente nos falta probar la siguiente igualdad:

$$(6.4.21) \quad \#(\pi_j(B_{\ell,i})) = \mu_i(\ell).$$

Sin pérdida de generalidad, supongamos que $i = 1$ y $j = 2$. Por lo tanto, nótese que dado $f \in \pi_1(H)$ tal que $f|_{[n] \setminus \{1,2\}} = \ell$ y dado $\mathbb{S}_1(e_{f,1})$, tenemos que:

$$\pi_2(\mathbb{S}_1(e_{f,1})) = \{0, \dots, \#(e_{f,1}) - 1\} \times \{(\ell(3), \ell(4), \dots, \ell(n))\}.$$

Por lo tanto, concluimos:

$$\pi_2(B_{\ell,1}) = \bigcup_{\substack{f \in \pi_1(H) \\ f|_{[n] \setminus \{1,2\}} = \ell}} \{0, \dots, \#(e_{f,1}) - 1\} \times \{(\ell(3), \dots, \ell(n))\} = \{0, \dots, \mu_1(\ell) - 1\} \times \{(\ell(3), \dots, \ell(n))\}.$$

Y, entonces, se deduce la Identidad (6.4.21) y la primera igualdad del lema. Para la segunda igualdad, basta observar que:

$$\#(E(\mathbb{S}_i(H))) = \sum_{k=1}^n \#(E_k(\mathbb{S}_i(H))).$$

Como $\#(E_i(\mathbb{S}_i(H))) = \#(E_i(H))$, la última igualdad se sigue de nuestro estudio previo. \square

LEMA 6.4.17. Con las notaciones precedentes, dados $i, j \in [n]$, $i \neq j$, $\ell \in \pi_{i,j}(H)$, se cumple la siguiente igualdad:

$$\#(\pi_j(A_{\ell,i})) \geq \mu_i(\ell).$$

Además, también tenemos que:

$$(6.4.22) \quad \#(\pi_j(H)) \geq \sum_{\ell \in \pi_{i,j}(H)} \mu_i(\ell).$$

DEMOSTRACIÓN. Por la definición de $A_{\ell,i}$ (Identidad (6.4.11) previa), tenemos que:

$$\pi_j(A_{\ell,i}) = \bigcup_{\substack{f \in \pi_i(H) \\ f|_{[n] \setminus \{i,j\}} = \ell}} \pi_j(e_{f,i}).$$

Por lo tanto, concluimos:

$$\#(\pi_j(A_{\ell,i})) \geq \max\{\#(\pi_j(e_{f,i})) : f \in \pi_i(H), f|_{[n] \setminus \{i,j\}} = \ell\}.$$

Nos falta demostrar que la restricción de π_j a $e_{f,i}$ es inyectiva. Esto es inmediato puesto que dados $h \in e_{f,i}$ y $f \in \pi_i(H)$ tales que $f|_{[n] \setminus \{i,j\}} = \ell$, sabemos que $\pi_j(h)(k) = \ell(k)$ para cada $k \neq i, j$. Pero $\pi_j(h)(i) = f(i)$ y, entonces, $\pi_j(h)$ está determinado de forma única por f para cada $h \in e_{f,i}$.

Como π_j restringida a $e_{f,i}$ es inyectiva, concluimos que:

$$\#(\pi_j(e_{f,i})) = \#(e_{f,i})$$

y, por lo tanto,

$$\#(\pi_j(A_{\ell,i})) \geq \max\{\#(e_{f,i}) : f \in \pi_i(H), f|_{[n] \setminus \{i,j\}} = \ell\} = \mu_i(\ell),$$

como queríamos demostrar. \square

Estos resultados técnicos nos permiten demostrar que el grado de salida promedio no decrece tras aplicar un shift (este resultado es la Afirmación 26 de [BCDMY, 22]).

PROPOSICIÓN 6.4.18 (**El grado de salida promedio no decrece tras un shift**). *Con las mismas notaciones que en los resultados anteriores, dados $H \subseteq \mathcal{Y}^{[n]}$ e $i \in [n]$, el número de aristas y la densidad del hiper-grafo de una inclusión $G(H)$ no aumentan tras realizar un shift, i.e.*

$$\begin{aligned} \#(E(\mathbb{S}_i(H))) &\leq \#(E(H)), \\ \text{gd}(\mathbb{S}_i(H)) &\leq \text{gd}(H). \end{aligned}$$

En particular, el grado de salida promedio no decrece después de un shift, i.e.

$$E_{\mathbb{S}_i(H)}[\text{outdeg}] \geq E_H[\text{outdeg}].$$

DEMOSTRACIÓN. De acuerdo con la Identidad (6.2.3), tenemos que:

$$\#(E(\mathbb{S}_i(H))) = \sum_{j=1}^n \#(E_j(\mathbb{S}_i(H))) = \sum_{j=1}^n \#(\pi_j(\mathbb{S}_i(H))).$$

Como $\#(E_i(\mathbb{S}_i(H))) = \#(E_i(H))$, podemos escribir:

$$\#(E(\mathbb{S}_i(H))) = \#(E_i(H)) + \sum_{j \neq i} \#(\pi_j(\mathbb{S}_i(H))).$$

Aplicando el Lema 6.4.16, concluimos:

$$\#(E(\mathbb{S}_i(H))) = \#(E_i(H)) + \sum_{j \neq i} \left(\sum_{\ell \in \pi_{i,j}(H)} \mu_i(\ell) \right).$$

Gracias a la Desigualdad (6.4.22) del Lema 6.4.17, tenemos que:

$$\#(E(\mathbb{S}_i(H))) \leq \#(E_i(H)) + \sum_{j \neq i} \#(\pi_j(H)).$$

De nuevo, aplicando la Identidad (6.2.3), concluimos:

$$\#(E(\mathbb{S}_i(H))) \leq \#(E_i(H)) + \sum_{j \neq i} \#(E_j(H)) = \#(E(H)).$$

Esto prueba la primera desigualdad de la proposición. Como $\#(\mathbb{S}_i(H)) = \#(H)$, la segunda desigualdad, relativa a la densidad del hiper-grafo de una inclusión, es una consecuencia inmediata de la primera. En lo que respecta a la última desigualdad, esta se sigue del Ítem *ii*) de la Proposición 6.2.5. \square

OBSERVACIÓN 6.4.19. De acuerdo con nuestro análisis, la dimensión exponencial y el grado de salida promedio evolucionan en direcciones opuestas cuando aplicamos una secuencia de shifts. Además, el número de aristas del hiper-grafo resultante tras aplicar un shift verifica la siguiente igualdad:

$$\#(E(\mathbb{S}_i(H))) = \#(E_i(H)) + \sum_{j \neq i} \left(\sum_{\ell \in \pi_{i,j}(H)} \mu_j(\ell) \right).$$

Nótese que no podemos “permutar” los sumatorios con índices $j \neq i$ y $\ell \in \pi_{i,j}(H)$ debido a que $\pi_{i,j}(H)$ varía con j .

Finalmente, ya estamos en condiciones de demostrar uno de nuestros resultados principales:

TEOREMA 6.4.20 (**El grado de salida promedio es estrictamente menor que la dimensión exponencial**). *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Entonces, tenemos que:*

$$E_H[\text{outdeg}] < \dim_E(H).$$

DEMOSTRACIÓN. Al igual que en la Observación 6.4.5, seguimos la idea de emplear secuencias de shifts como en [Ha, 95]. Para cada $H \subseteq \mathcal{Y}^{[n]}$, existe una palabra $\omega := \omega_1 \cdots \omega_r \in [n]^*$ que define una composición de una secuencia de shifts:

$$\mathbb{S}_\omega := \mathbb{S}_{\omega_1} \circ \cdots \circ \mathbb{S}_{\omega_r} : 2^{\mathcal{Y}^n} \longrightarrow 2^{\mathcal{Y}^n},$$

de modo que la clase:

$$H_* := \mathbb{S}_\omega(H),$$

es cerrada hacia abajo con respecto al pre-orden de Dickson \preceq . Gracias a la Proposición 6.4.18, obtenemos

$$E_H[\text{outdeg}] \leq E_{H_*}[\text{outdeg}].$$

Por la Proposición 6.4.1, tenemos que:

$$\dim_E(H_*) \leq \dim_E(H).$$

Finalmente, por la Proposición 6.3.21, como H_* es cerrado hacia abajo, obtenemos:

$$E_{H_*}[\text{outdeg}] < \text{avd}(H_*) \leq \dim_E(H_*),$$

lo que concluye la demostración del teorema. \square

A continuación, utilizamos el teorema anterior para obtener una mejora estricta de la Proposición 27 de [BCDMY, 22]. Esta mejora consiste en transformar la desigualdad $\text{avd}(H) \leq 4 \dim_E(H)$ en el Ítem *ii*) del siguiente corolario, i.e.

$$\text{avd}(H) < 2 \dim_E(H).$$

Esta mejora es posible gracias al uso del grado de salida promedio.

COROLARIO 6.4.21 (Mejora de la Proposición 27 de [BCDMY, 22]). *Sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Entonces, tenemos que:*

i) Las dimensiones satisfacen la siguiente desigualdad:

$$(6.4.23) \quad \dim_{\text{DS}}(H) < 2 \dim_E(H).$$

ii) El grado promedio y la dimensión exponencial verifican:

$$(6.4.24) \quad \text{avd}(H) < 2 \dim_E(H).$$

DEMOSTRACIÓN. Sea $S \subseteq [n]$ un subconjunto maximal con la propiedad de que $H|_S$ contiene un pseudo-cubo de dimensión $r = \sharp(S) = \dim_{\text{DS}}(H)$. Sea $\mathcal{C} \subseteq H|_S$ dicho pseudo-cubo. Por la Proposición 6.3.5, tenemos que:

$$\dim_{\text{DS}}(H) = r = \sharp(S) = \text{avd}(\mathcal{C}).$$

Gracias a la Proposición 6.2.5, obtenemos:

$$\text{avd}(\mathcal{C}) \leq 2E_{\mathcal{C}}[\text{outdeg}].$$

Por lo tanto, aplicando el Teorema 6.4.20 anterior, concluimos:

$$\dim_{\text{DS}}(H) = r < 2 \dim_E(\mathcal{C}).$$

A partir de la Observación 6.3.9, obtenemos de forma inmediata que:

$$\dim_{\text{DS}}(H) = r < 2 \dim_E(\mathcal{C}) \leq 2 \dim_E(H|_S) \leq 2 \dim_E(H),$$

y hemos probado la Desigualdad (6.4.23).

En cuanto a la segunda desigualdad, por la Proposición 6.2.5, tenemos que:

$$\text{avd}(H) \leq 2E_H[\text{outdeg}],$$

y el Teorema 6.4.20 anterior implica:

$$\text{avd}(H) < 2 \dim_E(H),$$

lo que demuestra la Desigualdad (6.4.24). \square

EJEMPLO 6.4.22. Si cambiamos la distribución de probabilidad definida en la clase de conceptos H , la segunda desigualdad del corolario anterior no es cierta. Sea $H \subseteq \{0, 1\}^{[3]}$ una clase de conceptos dada por la siguiente igualdad:

$$H := \{(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1)\}.$$

El hiper-grafo de una inclusión asociado tiene 9 aristas:

$$E(H) = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8, e_9\},$$

donde:

$$e_1 = \{(0, 0, 0), (1, 0, 0)\}, e_2 = \{(0, 0, 0), (0, 1, 0)\}, e_3 = \{(0, 0, 0), (0, 0, 1)\}, \\ e_4 = \{(1, 0, 0)\} = e_5, e_6 = \{(0, 1, 0)\} = e_7, e_8 = \{(0, 0, 1)\} = e_9.$$

Los grados de los vértices son:

$$\deg((0, 0, 0)) = 3, \deg((1, 0, 0)) = 1, \deg((0, 1, 0)) = 1, \deg((0, 0, 1)) = 1.$$

Consideramos la siguiente distribución de probabilidad sobre H :

$$\nu_{(0,0,0)} = \frac{4}{7}, \nu_{(1,0,0)} = \frac{1}{7}, \nu_{(0,1,0)} = \frac{1}{7}, \nu_{(0,0,1)} = \frac{1}{7}.$$

Tenemos que:

$$\sum_{h \in H} \deg(h) \cdot \nu_h = \frac{15}{7}.$$

Por otro lado, se cumple que $\dim_{\mathbb{E}}(H) = 1$ ya que cualquier restricción a dos coordenadas tiene cardinalidad 3. Finalmente, concluimos:

$$\frac{15}{7} = \sum_{h \in H} \deg(h) \cdot \nu_h > 2 \dim_{\mathbb{E}}(H) = 2.$$

6.4.3.2. *Shifting: grado promedio.* Analizamos en detalle el comportamiento, aparentemente “impredecible”, del grado promedio tras un shift.

En primer lugar, nótese que, en el caso binario ($p = 2$), $\text{avd}(H)$ no decrece al aplicar el operador de shifting:

COROLARIO 6.4.23. *En el caso $p = 2$, dado $i \in [n]$ y una clase de conceptos finita $H \subseteq \{0, 1\}^{[n]}$, tenemos que:*

$$\text{avd}(\mathbb{S}_i(H)) \geq \text{avd}(H).$$

DEMOSTRACIÓN. De acuerdo con la Identidad (6.2.15) de la Observación 6.2.14, dado un shift \mathbb{S}_i , tenemos que:

$$\text{avd}(\mathbb{S}_i(H)) - E_{\mathbb{S}_i(H)}[\text{outdeg}] = \frac{\sharp(\text{Haussl}(\mathbb{S}_i(H)))}{2\sharp(H)}.$$

Aplicando la Afirmación 2 de la demostración del Lema 2 de [Ha, 95], obtenemos:

$$\sharp(\sharp(\text{Haussl}(\mathbb{S}_i(H)))) \geq \sharp(\text{Haussl}(H)).$$

Por lo tanto, tenemos que:

$$\text{avd}(\mathbb{S}_i(H)) - E_{\mathbb{S}_i(H)}[\text{outdeg}] \geq \frac{\sharp(\text{Haussl}(H))}{2\sharp(H)} = \text{avd}(H) - E_H[\text{outdeg}].$$

Por la Proposición 6.4.18, obtenemos:

$$\text{avd}(\mathbb{S}_i(H)) - \text{avd}(H) \geq E_{\mathbb{S}_i(H)}[\text{outdeg}] - E_H[\text{outdeg}] \geq 0.$$

Finalmente, concluimos:

$$\text{avd}(\mathbb{S}_i(H)) \geq \text{avd}(H). \quad \square$$

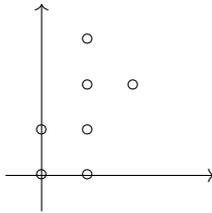
Sin embargo, en el caso $p > 2$, el grado promedio del hiper-grafo de una inclusión $G(H)$ tiene un comportamiento “impredecible” como muestra el siguiente ejemplo.

EJEMPLO 6.4.24 (**El grado promedio puede crecer o decrecer tras un shift en el caso no binario**).

Consideramos $\mathcal{Y} := \{0, 1, 2, 3\}$, $n = 2$ y definimos la clase de conceptos:

$$H := \{(0, 0), (0, 1), (1, 0), (1, 1), (1, 2), (1, 3), (2, 2)\} \subseteq \mathcal{Y}^2,$$

cuya imagen es:



El hiper-grafo de una inclusión $G(H)$ asociado a H tiene 7 aristas y, por lo tanto:

$$E_H[\text{outdeg}] = 2 - \frac{7}{7} = 1.$$

Por otro lado, contiene un cubo de dimensión 2, de lo que se deduce que:

$$\dim_{\text{DS}}(H) = \dim_{\mathbb{E}}(H) = 2.$$

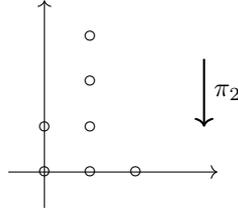
Además, tenemos que $\sharp(E^{(1)}(H)) = 2$ (i.e. hay solo dos aristas en $G(H)$ que son unarias). Por consiguiente,

$$\text{avd}(H) = 2 - \frac{2}{7} = \frac{12}{7}.$$

En primer lugar, realizamos un shift en la dirección de la proyección π_2 (i.e. la proyección que “olvida” la segunda coordenada):

$$H' := \mathbb{S}_2(H) = \{(0, 0), (1, 0), (2, 0), (0, 1), (1, 1), (1, 2), (1, 3)\},$$

cuya imagen es:



En este caso, tanto la dimensión exponencial como la dimensión DS se mantienen invariantes. Lo mismo ocurre con el número total de aristas y , en consecuencia, con la densidad del hiper-grafo y el grado de salida promedio, i.e.

$$E_{H'}[\text{outdeg}] = E_H[\text{outdeg}] = 1.$$

Sin embargo, el número de aristas unarias aumenta:

$$3 = \#(E^{(1)}(H')) > \#(E^{(1)}(H)) = 2.$$

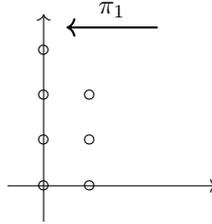
Por lo tanto, el grado promedio disminuye:

$$\frac{11}{7} = \text{avd}(H') < \text{avd}(H) = \frac{12}{7}.$$

Por otra parte, si hacemos un shift en la dirección de la proyección π_1 , obtenemos:

$$H'' := \mathbb{S}_1(H) := \{(0, 0), (1, 0), (0, 1), (1, 1), (0, 2), (1, 2), (0, 3)\}$$

cuya imagen es:



El conjunto H'' es cerrado hacia abajo, pero no es un pseudo-cubo. Tanto la dimensión exponencial como la dimensión DS se preservan. Sin embargo, el número de aristas decrece de forma estricta y , por lo tanto, el grado de salida promedio aumenta. Es decir:

$$\#(E(H'')) = 6, \quad 8/7 = E_{H''}[\text{outdeg}] > E_H[\text{outdeg}] = 1.$$

En esta ocasión, el número de aristas que son unarias también decrece (i.e. $\#(E^{(1)}(H'')) = 1$) y , por lo tanto, el grado promedio aumenta:

$$2 - \frac{1}{7} = \frac{13}{7} = \text{avd}(H'') > \text{avd}(H) = \frac{12}{7}.$$

El principal problema es la evolución (aparentemente) “caótica” e “impredecible” del grado promedio tras aplicar un shift. El siguiente resultado determina, parcialmente, dicho comportamiento:

LEMA 6.4.25. *Con las mismas notaciones e hipótesis que antes, el número de aristas que son unarias tras un shift verifica:*

$$\#(E^{(1)}(\mathbb{S}_i(H))) = \#(E_i(H) \cap E^{(1)}(H)) + \sum_{j \neq i} \left(\sum_{\ell \in \pi_{i,j}(H)} (\mu_i(\ell) - \mu_i^{(2)}(\ell)) \right),$$

donde $\mu_i(\ell)$ y $\mu_i^{(2)}(\ell)$ son las funciones introducidas en las Definiciones 49 y 50 respectivamente.

DEMOSTRACIÓN. El proceso de shifting respeta las cardinalidades de las aristas en la dirección de π_i . Por lo tanto, se deduce de forma inmediata la siguiente igualdad para aristas unarias:

$$\#(E_i(\mathbb{S}_i(H)) \cap E^{(1)}(\mathbb{S}_i(H))) = \#(E_i(H) \cap E^{(1)}(H)).$$

Por consiguiente, consideramos ahora el comportamiento de las aristas en $G(\mathbb{S}_i(H))$ en la dirección π_j , con $j \neq i$. A continuación, vamos a demostrar que dados $i, j \in [n]$, con $i \neq j$, se verifica la siguiente igualdad:

$$(6.4.25) \quad \#(E_j(\mathbb{S}_i(H)) \cap E^{(1)}(\mathbb{S}_1(H))) = \sum_{\ell \in \pi_{i,j}(H)} (\mu_i(\ell) - \mu_i^{(2)}(\ell)),$$

donde $\mu_i(\ell)$ y $\mu_i^{(2)}(\ell)$ son respectivamente el “máximo” (cf. Definición 49) y el “segundo máximo” (cf. Definición 50) de las cardinalidades de las aristas determinadas por $\pi_i(H(\ell))$.

Nótese que tanto $\mathbb{S}_i(H)$ como $\mathbb{S}_i(H)(\ell)$, para todo $\ell \in \pi_{i,j}(H) = \pi_{i,j}(\mathbb{S}_i(H))$, son *conjuntos cerrados hacia abajo con respecto a la coordenada i -ésima*. Es decir, ambos conjuntos verifican la siguiente propiedad: dado $h \in \mathbb{S}_i(H)$ (respectivamente, $h \in \mathbb{S}_i(H)(\ell)$), para todo $g \in \mathcal{Y}^{[n]}$ tal que $\pi_i(g) = \pi_i(h)$ y $g(i) \leq h(i)$, entonces $g \in \mathbb{S}_i(H)$ (resp. $g \in \mathbb{S}_i(H)(\ell)$).

A continuación, demostramos una serie de propiedades relativas a clases de conceptos cerrados hacia abajo con respecto a la i -ésima coordenada:

AFIRMACIÓN 6.4.25.1. *Con las notaciones anteriores, sean $i, j \in [n]$, con $i \neq j$, y sea $\ell \in \pi_{i,j}(H)$ la proyección de alguna función en H bajo $\pi_{i,j}$. Supongamos que $H(\ell)$ es cerrado hacia abajo con respecto a la coordenada i -ésima. Entonces, tenemos que:*

i) Si $\mu_i(\ell) = \mu_i^{(2)}(\ell)$, entonces, para cada $f \in \pi_j(H(\ell))$, la cardinalidad de la arista $e_{f,j} \in E_j(H)$ verifica:

$$\sharp(e_{f,j}) \geq 2.$$

ii) Si $\mu_i(\ell) > \mu_i^{(2)}(\ell)$, entonces, para cada $f \in \pi_j(H(\ell))$ tal que $0 \leq f(i) \leq \mu_i^{(2)}(\ell) - 1$, la cardinalidad de la arista $e_{f,j} \in E_j(H)$ cumple que:

$$\sharp(e_{f,j}) \geq 2.$$

iii) Si $\mu_i(\ell) > \mu_i^{(2)}(\ell)$, entonces, para cada $f \in \pi_j(H(\ell))$ tal que $\mu_i^{(2)}(\ell) \leq f(i) \leq \mu_i(\ell) - 1$, la cardinalidad de la arista $e_{f,j} \in E_j(H)$ satisface:

$$\sharp(e_{f,j}) = 1.$$

En particular, siempre se verifica la siguiente igualdad:

$$(6.4.26) \quad \sharp(\{e_{f,j} \subseteq H : f \in \pi_j(H(\ell)), \sharp(e_{f,j}) = 1\}) = \mu_i(\ell) - \mu_i^{(2)}(\ell).$$

Demostración de la Afirmación. Demostramos cada uno de estos ítems por separado. Sin pérdida de generalidad, podemos suponer que $i = 1$ y $j = 2$.

- *Ítem i):* Consideramos $f \in \pi_2(H(\ell))$. Tenemos que:

$$f = (r, \ell(3), \dots, \ell(n)),$$

donde $0 \leq r \leq \mu_1(\ell) - 1$. Como $\ell \in \pi_{1,2}(H)$, obviamente tenemos que $\mu_1(\ell) \geq 1$ y, por lo tanto, también se cumple que $\mu_1^{(2)}(\ell) \geq 1$. Por consiguiente, debido a la definición de estos dos máximos, existen $g_1, g_2 \in \pi_1(H(\ell))$ tales que $e_{g_1,1} \neq e_{g_2,1}$ y sus cardinalidades satisfacen $\sharp(e_{g_1,1}) = \mu_1(\ell)$ y $\sharp(e_{g_2,1}) = \mu_1^{(2)}(\ell)$. Como $H(\ell)$ es cerrado hacia abajo con respecto a la coordenada i -ésima, y $\mu_1(\ell) = \mu_1^{(2)}(\ell)$, concluimos que:

$$e_{g_1,1} = \{0, \dots, \mu_1(\ell) - 1\} \times \{(g_1(2), \dots, g_1(n))\},$$

$$e_{g_2,1} = \{0, \dots, \mu_1(\ell) - 1\} \times \{(g_2(2), \dots, g_2(n))\},$$

Además, tenemos que:

$$(g_1(2), \dots, g_1(n)) = (g_1(2), \ell(3), \dots, \ell(n)), \quad (g_2(2), \dots, g_2(n)) = (g_2(2), \ell(3), \dots, \ell(n)),$$

donde $g_1(2) \neq g_2(2)$. Definimos ahora las dos siguientes funciones:

$$h_1 = (r, g_1(2), \ell(3), \dots, \ell(n)),$$

$$h_2 = (r, g_2(2), \ell(3), \dots, \ell(n)).$$

Como $g_1(2) \neq g_2(2)$, entonces $h_1 \neq h_2$. Como $0 \leq r \leq \mu_1(\ell) - 1$, tenemos que $h_i \in e_{g_i,1} \subseteq H(\ell)$ para cada $i \in \{1, 2\}$. Por otro lado, $\pi_2(h_1) = \pi_2(h_2) = f$. Entonces, $h_1, h_2 \in e_{f,2}$ y, por lo tanto, $\sharp(e_{f,2}) \geq 2$.

- *Ítem ii):* Supongamos ahora que $\ell \in \pi_{1,2}(H)$ es tal que $\mu_1(\ell) > \mu_1^{(2)}(\ell)$. Sea $f \in \pi_2(H(\ell))$, tal que $0 \leq f(1) = r \leq \mu_1^{(2)}(\ell) - 1$. Existen $g_1, g_2 \in H(\ell)$ de modo que $\sharp(e_{g_1,1}) = \mu_1(\ell)$ y $\sharp(e_{g_2,1}) = \mu_1^{(2)}(\ell)$. Como $\mu_1(\ell) \neq \mu_1^{(2)}(\ell)$, entonces $g_1 \neq g_2$. Además, como $H(\ell)$ es cerrado hacia abajo con respecto a la coordenada i -ésima, tenemos que:

$$e_{g_1,1} = \{0, \dots, \mu_1(\ell) - 1\} \times \{(g_1(2), \dots, g_1(n))\},$$

$$e_{g_2,1} = \{0, \dots, \mu_1^{(2)}(\ell) - 1\} \times \{(g_1(2), \dots, g_1(n))\}.$$

Al ser $g_1 \neq g_2$ y $g_1, g_2 \in \pi_i(H(\ell))$, entonces $g_1(k) = g_2(k) = \ell(k)$, donde $3 \leq k \leq n$, y $g_1(2) \neq g_2(2)$. A continuación, definimos las dos funciones siguientes:

$$h_1 = (r, g_1(2), \ell(3), \dots, \ell(n)),$$

$$h_2 = (r, g_2(2), \ell(3), \dots, \ell(n)).$$

Por lo tanto, $h_1 \in e_{g_1,1}$, $h_2 \in e_{g_2,1}$ y $h_1 \neq h_2$. En particular, $h_1, h_2 \in H(\ell)$ y, obviamente, $h_1, h_2 \in e_{f,2}$. Por consiguiente, $\sharp(e_{f,2}) \geq 2$ como queríamos demostrar.

- *Ítem iii):* Finalmente, supongamos que $\ell \in \pi_{1,2}(H)$ cumple que $\mu_1(\ell) > \mu_1^{(2)}(\ell)$. Sea $f \in \pi_2(H)$ tal que $\mu_1^{(2)}(\ell) \leq f(1) = r \leq \mu_1(\ell) - 1$. Supongamos que existen $h_1, h_2 \in e_{f,2}$ tales que $h_1 \neq h_2$. Como $h_1, h_2 \in e_{f,2}$, entonces concluimos:

$$- h_1(k) = h_2(k) = f(k), \text{ para } k \neq 2,$$

$$- h_1(2) \neq h_2(2).$$

Consideramos $g_1 = \pi_1(h_1)$ y $g_2 = \pi_1(h_2)$. Como $h_1(2) \neq h_2(2)$, entonces $g_1 \neq g_2$. Además,

$$g_1 = (h_1(2), \ell(3), \dots, \ell(n)),$$

$$g_2 = (h_2(2), \ell(3), \dots, \ell(n)).$$

Entonces, tenemos que:

$$e_{g_1,1} \subseteq \{0, \dots, \mu_1(\ell) - 1\} \times \{(h_1(2), \ell(3), \dots, \ell(n))\},$$

$$e_{g_2,1} \subseteq \{0, \dots, \mu_1(\ell) - 1\} \times \{(h_2(2), \ell(3), \dots, \ell(n))\}.$$

Adicionalmente, por razones obvias, se cumple que $h_1 \in e_{g_1,1}$ y $h_2 \in e_{g_2,1}$. Pero $h_1(1) = f(1) \geq \mu_i^{(2)}(\ell)$ y $h_2(1) = f(1) \geq \mu_i^{(2)}(\ell)$. Por lo tanto, para todo $i \in \{1, 2\}$, concluimos:

$$\{0, \dots, \mu_1^{(2)}(\ell) - 1\} \times \{(h_1(2), \ell(3), \dots, \ell(n))\} \subsetneq e_{g_1,1},$$

y

$$\{0, \dots, \mu_1^{(2)}(\ell) - 1\} \times \{(h_2(2), \ell(3), \dots, \ell(n))\} \subsetneq e_{g_2,1}.$$

Así que, necesariamente:

$$\sharp(e_{g_1,1}) = \sharp(e_{g_2,1}) = \mu_1(\ell)$$

con $g_1 \neq g_2$. Sin embargo, esto no es posible debido a la definición de ambos máximos. Por lo tanto, no pueden existir $h_1, h_2 \in e_{f,2}$ con $h_1 \neq h_2$ y, entonces, en este caso tenemos que

$$\sharp(e_{f,2}) = 1.$$

La Identidad (6.4.26) se sigue inmediatamente de los Ítems *i)*, *ii)* y *iii)* anteriores. ■

Con el objetivo de simplificar las notaciones, consideramos $i, j \in [n]$, con $i \neq j$, y escribimos $H' := \mathbb{S}_i(H)$. Sea $\ell \in \pi_{i,j}(\mathbb{S}_i(H)) = \pi_{i,j}(H')$. Por definición, tanto H' como $H'(\ell)$ son conjuntos cerrados hacia abajo con respecto a la coordenada i -ésima. Denotamos por ν_i y $\nu_i^{(2)}$ respectivamente al “máximo” y al “segundo máximo” de las cardinalidades de las aristas determinadas por $\pi_i(H'(\ell)) = H'(\ell)$. Con estas notaciones, tenemos que:

$$\mu_i(\ell) = \nu_i(\ell), \quad \mu_i^{(2)}(\ell) = \nu_i^{(2)}(\ell), \quad \forall \ell \in \pi_{i,j}(H) = \pi_{i,j}(H').$$

Consideramos ahora el conjunto:

$$E_j^{(1)}(\mathbb{S}_i(H)) := \{e_{f,j} \in E(H') : e_{f,j} \in E^{(1)}(\mathbb{S}_i(H))\}.$$

La siguiente igualdad muestra una partición de $E_j^{(1)}(\mathbb{S}_i(H))$:

$$E_j^{(1)}(\mathbb{S}_i(H)) = \bigsqcup_{\ell \in \pi_{i,j}(H)} \{e_{f,j} \subseteq H' : f \in \pi_j(H'(\ell)), \sharp(e_{f,j}) = 1\}.$$

Por lo tanto, tenemos que:

$$\sharp(E_j^{(1)}(\mathbb{S}_i(H))) = \sum_{\ell \in \pi_{i,j}(H)} \sharp(\{e_{f,j} \subseteq H' : f \in \pi_j(H'(\ell)), \sharp(e_{f,j}) = 1\}).$$

Dado que tanto H' como $H'(\ell)$ son cerrados hacia abajo con respecto a la coordenada i -ésima, la afirmación anterior implica:

$$\sharp(\{e_{f,j} \subseteq H' : f \in \pi_j(H'(\ell)), \sharp(e_{f,j}) = 1\}) = \nu_i(\ell) - \nu_i^{(2)}(\ell) = \mu_i(\ell) - \mu_i^{(2)}(\ell).$$

Por lo tanto, concluimos:

$$\# \left(E(\mathbb{S}_i(H)) \cap E^{(1)}(\mathbb{S}_i(H)) \right) = \sum_{\ell \in \pi_{i,j}(H)} \left(\mu_i(\ell) - \mu_i^{(2)}(\ell) \right),$$

que es la Identidad (6.4.25). El lema se verifica ya que:

$$\# \left(E^{(1)}(\mathbb{S}_i(H)) \right) = \# \left(E_i(\mathbb{S}_i(H)) \cap E^{(1)}(\mathbb{S}_i(H)) \right) + \sum_{j \neq i} \# \left(E_j(\mathbb{S}_i(H)) \cap E^{(1)}(\mathbb{S}_i(H)) \right).$$

□

Este resultado técnico implica la siguiente expresión para la variación del grado promedio y el grado de salida promedio tras un shift:

PROPOSICIÓN 6.4.26. *Con las notaciones anteriores, tenemos que:*

$$\text{avd}(\mathbb{S}_i(H)) - E_{\mathbb{S}_i(H)}[\text{outdeg}] = \frac{1}{\#(H)} \left(\# \left(E_i(H) \setminus E^{(1)}(H) \right) + \sum_{j \neq i} \left(\sum_{\ell \in \pi_{i,j}(H)} \mu_i^{(2)}(\ell) \right) \right).$$

DEMOSTRACIÓN. De la Proposición 6.2.5 se deduce que:

$$\text{avd}(\mathbb{S}_i(H)) - E_{\mathbb{S}_i(H)}[\text{outdeg}] = \frac{1}{\#(H)} \left(\#(E(\mathbb{S}_i(H))) - \#(E^{(1)}(\mathbb{S}_i(H))) \right).$$

Aplicando los Lemas 6.4.16 y 6.4.25 a la expresión anterior, se obtiene la igualdad del enunciado. □

6.4.3.3. *Orientaciones con grado de salida máximo controlado.* El Corolario 6.4.21 nos permite probar una mejora estricta del Corolario 28 de [BCDMY, 22] sobre la existencia de orientaciones con grado de salida máximo controlado. En el Corolario 28 de [BCDMY, 22], los autores prueban la existencia de una orientación $\sigma : E(H) \rightarrow H$ que verifica la siguiente desigualdad:

$$\text{outdeg}_{\max}(\sigma) \leq 4 \dim_E(H).$$

Establecemos esta mejora mediante nuestro Teorema 6.4.27, especialmente su Ítem *i*) y la Desigualdad (6.4.31). A continuación, introducimos una serie de nociones que necesitaremos para nuestro Teorema 6.4.27. En primer lugar, definimos el grado mínimo de una clase de conceptos $H \subseteq \mathcal{Y}^{[n]}$ como el mínimo de los grados de sus elementos en el hiper-grafo de una inclusión $G(H)$, i.e.

$$\text{deg}_{\min}(H) := \min\{\text{deg}(h) : h \in H\}.$$

En [DS, 14], los autores consideran también el grado promedio “maximal”:

$$\text{avd}_{\max}(H) := \max\{\text{avd}(F) : F \subseteq H\},$$

que es obviamente una función monótona para la inclusión de conjuntos. Análogamente, consideramos el grado de salida promedio “maximal” definido mediante la siguiente expresión:

$$E_H^{(\max)}[\text{outdeg}] := \max\{E_F[\text{outdeg}] : F \subseteq H\}.$$

que es, de nuevo, una función monótona con respecto a la inclusión de conjuntos. Obviamente, como $\text{avd}(F) \leq 2E_F[\text{outdeg}]$ para todo $F \subseteq H$, también concluimos que:

$$(6.4.27) \quad \text{avd}_{\max}(H) \leq 2E_H^{(\max)}[\text{outdeg}].$$

Finalmente, introducimos la siguiente cantidad:

$$\min - \text{outdeg}_{\max}(H) = \min\{\text{outdeg}_{\max}(\sigma) : \sigma : E(H) \rightarrow H \text{ es una orientación}\}.$$

TEOREMA 6.4.27 (**Mejora estricta del Corolario 28 de [BCDMY, 22]**). *Con las mismas notaciones que antes, sea $H \subseteq \mathcal{Y}^{[n]}$ una clase de conceptos finita. Sea $\varphi : 2^H \rightarrow \mathbb{R}$ una función monótona con respecto a la inclusión en 2^H . Supongamos que la función φ también verifica las siguiente hipótesis:*

$$(6.4.28) \quad \forall F \in 2^H, \text{deg}_{\min}(F) \leq \varphi(F).$$

Entonces, existe una orientación $\sigma : E(H) \rightarrow H$ que cumple la siguiente desigualdad:

$$(6.4.29) \quad \text{outdeg}_{\max}(\sigma) \leq \varphi(H).$$

Además, tenemos que:

i) La Hipótesis (6.4.28) se verifica siempre que la siguiente condición sea cierta:

$$(6.4.30) \quad \text{avd}(F) \leq \varphi(F), \quad \forall F \subseteq H.$$

En particular, la función monótona $\varphi(F) := 2 \dim_E(F)$ satisface la Desigualdad (6.4.30) y, por lo tanto, existe una orientación $\sigma : E(H) \rightarrow H$ tal que:

$$(6.4.31) \quad \text{outdeg}_{\max}(\sigma) \leq 2 \dim_E(H).$$

ii) El grado promedio maximal también verifica la Hipótesis (6.4.30) y, por consiguiente, existe una orientación $\sigma : E(H) \rightarrow H$ de modo que:

$$\text{outdeg}_{\max}(\sigma) \leq \text{avd}_{\max}(H) \leq 2E_H^{(\max)}[\text{outdeg}].$$

iii) El $E_H[\text{outdeg}]$ es la cantidad que controla la complejidad del algoritmo OIG (cf. Algoritmo 0.2.2) ya que:

$$E_H[\text{outdeg}] \leq \min - \text{outdeg}_{\max}(H) \leq 2E_H^{(\max)}[\text{outdeg}].$$

DEMOSTRACIÓN. Procedemos por inducción en el cardinal de H (formalizando con cuidado los argumentos de [BCDMY, 22]). El caso $\sharp(H) = 1$ es evidente.

De acuerdo con la Identidad (6.4.28), sea $h \in H$ tal que $\deg(h) \leq \varphi(H)$. Consideramos la clase $H' := H \setminus \{h\}$ y el hiper-grafo de una inclusión $G(H')$ asociado. A continuación, clasificamos las aristas de $E(H)$ en las siguientes tres clases disjuntas:

- La clase de aristas que no contienen a h :

$$E_h^{(0)}(H) := \{e \in E(H) : h \notin e\}.$$

- La clase de aristas que contienen a h pero que no son unarias:

$$E_h^{(2)}(H) := \{e \in E(H) : h \in e, \sharp(e) \geq 2\}.$$

- La clase de aristas unarias cuyo único elemento es h :

$$E_h^{(1)}(H) := \{e \in E(H) : e = \{h\}\}.$$

Tenemos la siguiente partición de $E(H)$:

$$E(H) = E_h^{(0)}(H) \sqcup E_h^{(1)}(H) \sqcup E_h^{(2)}(H).$$

Aplicamos ahora la hipótesis de inducción y suponemos que existe una orientación $\sigma' : E(H') \rightarrow H'$ tal que:

$$\text{outdeg}_{\max}(\sigma') \leq \varphi(H').$$

A continuación, extendemos σ' a una orientación $\sigma : E(H) \rightarrow H$ de la siguiente forma:

- Si

$$e \in E_h^{(0)}(H) \sqcup E_h^{(2)}(H),$$

entonces definimos:

$$\sigma(e) := \sigma'(e \setminus \{h\}) \in H' \subseteq H.$$

- En otro caso, es decir, si $e \in E_h^{(1)}(H)$, definimos:

$$\sigma(e) := h.$$

Nótese que para cada $g \in H'$ (i.e. $g \neq h$), tenemos que:

$$\text{outdeg}(\sigma, g) = \text{outdeg}(\sigma', g).$$

Por otro lado, se verifica la siguiente igualdad:

$$\sharp(\{e \in E(H) : h \in e, \sigma(e) \neq h\}) = \sharp(E_h^{(2)}(H)) = \deg(h).$$

Entonces,

$$\text{outdeg}(\sigma, h) = \deg(h)$$

y, por consiguiente, obtenemos:

$$\text{outdeg}_{\max}(\sigma) := \max\{\text{outdeg}_{\max}(\sigma'), \deg(h)\}.$$

Además, también tenemos que:

- Por la hipótesis de inducción y el hecho de que φ es monótona, se cumple la siguiente desigualdad:

$$\text{outdeg}_{\max}(\sigma') \leq \varphi(H') \leq \varphi(H).$$

- Por nuestra Hipótesis (6.4.28), h ha sido escogido de modo que satisfice:

$$\deg(h) \leq \varphi(H).$$

Por lo tanto, concluimos que:

$$\text{outdeg}_{\max}(\sigma) \leq \varphi(H).$$

Para el Ítem *i*), nótese que si $\text{avd}(F) \leq \varphi(F)$ es cierto para todo $F \subseteq H$, entonces la Hipótesis (6.4.28) también se verifica. Finalmente, como la función $\varphi(H) := 2 \dim_E(H)$ es monótona con respecto a la inclusión de conjuntos, gracias al Corolario 6.4.21, tenemos que esta función satisface la Hipótesis (6.4.28), lo que implica el Ítem *i*).

La primera desigualdad del Ítem *ii*) se verifica porque la función avd_{\max} es monótona con respecto a la inclusión de conjuntos y, evidentemente, verifica la Hipótesis 6.4.30. Por otro lado, la segunda desigualdad de este ítem es la Desigualdad 6.4.27. Por último, el Ítem *iii*) es una consecuencia inmediata de los resultados anteriores. \square

El Ítem *i*) de este último teorema mejora el Corolario 28 de [BCDMY, 22] por un factor estrictamente mayor que 2. Por otra parte, el Ítem *ii*) extiende el Lema 3 de [DS, 14] incluyendo dos veces el grado de salida promedio maximal. Sin embargo, no hemos sido capaces de probar resultados similares para la función (naturalmente monótona) dim_{DS} .

El siguiente ejemplo muestra que la Hipótesis (6.4.30) no es satisfecha por la dimensión DS . Además, este ejemplo prueba también que $\text{avd}_{\max}(H)$ puede ser mayor que $\text{dim}_{\text{DS}}(H)$. Por lo tanto, el Lema 3 de [DS, 14] no puede utilizarse para producir orientaciones cuyo grado de salida máximo esté acotado por $\text{dim}_{\text{DS}}(H)$. En consecuencia, parece que:

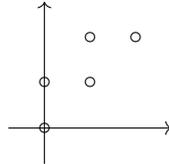
$$2E_H^{(\max)}[\text{outdeg}]$$

es la cantidad “más fina” que podemos encontrar de modo que sea estable por shifting y que permita controlar grado de salida máximo de alguna orientación en el hiper-grafo de una inclusión $G(H)$ asociado a H .

EJEMPLO 6.4.28 (El grado promedio no es siempre menor o igual que la dimensión DS). Consideramos $n = 2$, $\mathcal{Y} := \{0, 1, 2\}$ y la clase de conceptos $H \subseteq \mathcal{Y}^{[2]}$ dada por la siguiente igualdad:

$$H := \{(0, 0), (0, 1), (1, 1), (1, 2), (2, 2)\},$$

cuya imagen es:



Unos cálculos sencillos nos permiten concluir que:

$$\sharp(E(H)) = 6, \quad \sharp(E^{(1)}(H)) = 2.$$

Por lo tanto,

$$\text{avd}(H) = \frac{8}{5} > 1, \quad E_H[\text{outdeg}] = \frac{4}{5} < 1.$$

Como $\sharp(H) = 5 \geq 2^2$, inmediatamente tenemos que $\text{dim}_E(H) = 2$. Sin embargo, H no es un pseudo-cubo. Una discusión de diversos casos, muestra que si eliminamos un solo punto de H , los cuatro puntos restantes no forman un pseudo-cubo. Si eliminamos dos puntos de H , tampoco obtenemos una clase H' que sea un pseudo-cubo. En consecuencia, concluimos que H solo contiene pseudo-cubos de dimensión 1. Por consiguiente, $\text{dim}_{\text{DS}}(H) = 1$ y este ejemplo verifica que:

$$E_H[\text{outdeg}] = \frac{4}{5} < \text{dim}_{\text{DS}}(H) = 1 < \text{avd}(H) = \frac{8}{5} < \text{dim}_E(H) = 2.$$

Por lo tanto, la Hipótesis (6.4.30) no se verifica para $\text{dim}_{\text{DS}}(H)$. Tampoco el Lema 3 de [DS, 14] ayuda con este ejemplo, ya que $\text{dim}_{\text{DS}}(H) < \text{avd}_{\max}(H)$. Sin embargo, la Hipótesis (6.4.28) se cumple en este ejemplo y la tesis del Lema 13 de [BCDMY, 22] sigue siendo cierta en este caso. Una cuestión que queda abierta es si el Lema 13 de [BCDMY, 22] sigue siendo válido en un contexto más general.

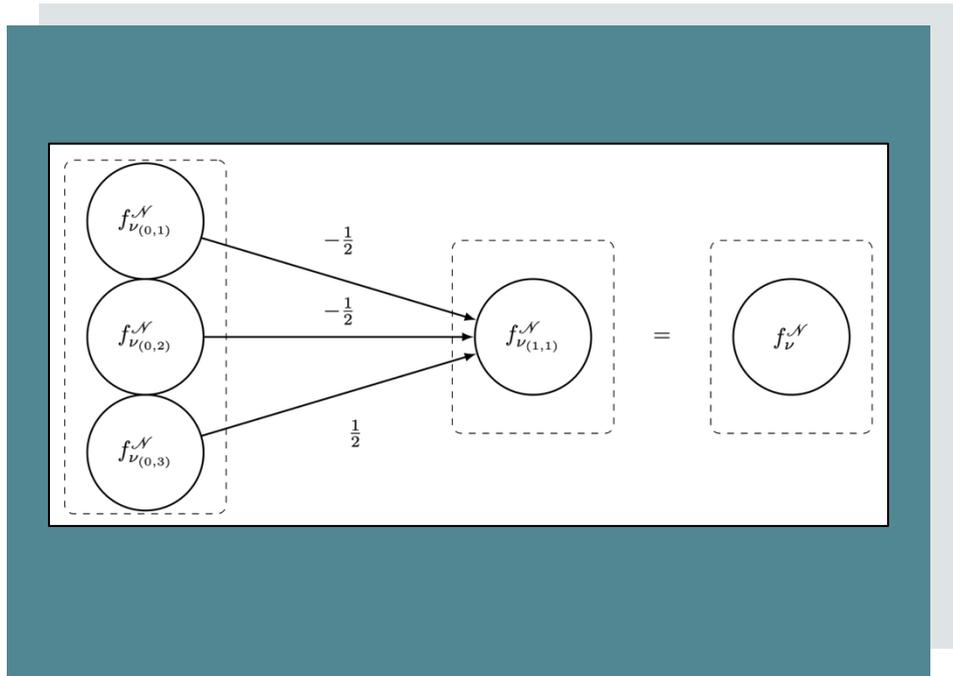
Bibliografía

- [Ad, 78] L. Adleman, *Two theorems on random polynomial time*. En 19th Annual Symposium on Foundations of Computer Science, 1978, 75-83.
- [AK, 06] F. Albiac, N.J. Kalton, “*Topics in Banach Space Theory*”. Graduate Text in Mathematics **233**, Springer, New York, 2006.
- [Al, 83] N. Alon, *On the density of sets of vectors*. Discrete Math. **46** (1983), 199-202.
- [Al, 99] N. Alon, *Combinatorial Nullstellensatz*. Combin. Probab. Comput. **8** (1999), 7-29.
- [Ar, 74] V.I. Arnol’d, “*Équations Différentielles Ordinaires*”. Editions MIR de Moscú, 1974.
- [Ar, 50] N. Aronszajn, *Theory of reproducing kernels*. Trans. Amer. Math. Soc. **68** (1950), 337-404.
- [BP, 07] C. Beltrán, L.M. Pardo, *Estimates on the Distribution of the Condition Number of Singular Matrices*. Found. Comput. Math. **7** (2007), 87-134.
- [BP, 08] C. Beltrán, L. M. Pardo, *On Smale’s 17th problem: a probabilistic positive solution*. Found. Comput. Math. **8** (2008), 1-43.
- [BP, 09] C. Beltrán, L. M. Pardo, *Smale’s 17th problem: average polynomial time to compute affine and projective solutions*. Journal of the American Mathematical Society **22** (2009), 363-385.
- [BP, 11] C. Beltrán, L. M. Pardo, *Fast Linear Homotopy to Find Approximate Zeros of Polynomial Systems*. Foundations of Computational Mathematics **11** (2011), 95-129.
- [BCSS, 98] L. Blum, F. Cucker, M. Shub, S. Smale, “*Complexity and real computation*”. Springer-Verlag, New York, 1998.
- [BSS, 89] L. Blum, M. Shub, S. Smale, *On a theory of computation and complexity over the real numbers: NP-completeness, recursive functions and universal machines*. Bulletin of Amer. Math. Soc **1** (1989), 1-46.
- [BEHW, 89] A. Blumer, A. Ehrenfeucht, D. Haussler, M.K. Warmuth, *Learnability and the Vapnik-Chervonenkis Dimension*. J. of the ACM **36** (1989), 929-965.
- [BCDMY, 22] N. Brukhim, D. Carmon, I. Dinur, S. Moran, A. Yehudayoff, *A Characterization of Multiclass Learnability*. Electronic Colloquium on Computational Complexity, TR22-035, 2022.
- [BJ, 14] L. Busé, J.P. Jouanolou, *On the discriminant scheme of homogeneous polynomials*. Math. in Comput. Sci. **8** (2014), 175-234.
- [CGHMP, 03] D. Castro, M. Giusti, J. Heintz, G. Matera, L. M. Pardo, *The Hardness of Polynomial Equation Solving*. Found. Comput. Math. **3** (2003), 347-420.
- [CD, 05] E. Cattani, A. Dickenstein, *Introduction to residues and resultants*. En A.Dickenstein, I.Z.Emiris (Eds.): “*Solving Polynomial Equations: Foundations, Algorithms, and Applications*”, Algorithms and Comput. Math., **14**, Springer-Verlag, 2005, 1-61.
- [Ch, 64-65] C. Chevalley, “*Introduction à la Théorie des Schémas*”. Cours à l’IHP, Notes de D. Fatiadi, éd. Centre de Physique Théorique de l’École Polytechnique, Pub. N **A75.1268**, 1964-65.
- [CLO, 15] D.A. Cox, J. Little, D. O’Shea, “*Ideals, Varieties, and Algorithms. An Introduction to Computational Algebraic Geometry and Commutative Algebra*”. Springer, 2015.
- [CKKLW,95] F. Cucker, M. Karpinski, P. Koiran, T. Lickteig, K. Werther, *On real Turing machines that toss coins*. In Proc. of the Twenty-Seventh Ann. ACM Symp. on Theor. of Comput. (F. T. Leighton, A. Borodin eds.), ACM 1995, 335-342.
- [CZ, 07] F. Cucker, D. Zhou, “*Learning theory: an approximation theory viewpoint*”. Cambridge Monographs on Applied and Computational Mathematics, 2007.
- [DS, 14] A. Daniely, S. Shalev-Shwartz, *Optimal Learners for Multiclass Problems*. In Proc. COLT’2014, J. Machine Learning Research, **35** (2014), 1-30.
- [DML, 78] R. A. DeMillo, R.J. Lipton, *A probabilistic remark on algebraic program testing*. Information Processing Letters **7** (1978), 193-195.
- [Di, 1913] L. E. Dickson, *Finiteness of the odd perfect and primitive abundant numbers with n distinct prime factors*. American Journal of Mathematics **35** (1913), 413-422.
- [Dv, 09] Z. Dvir, *On the size of Kakeya sets on finite fields*. J. of the A.M.S. **22** (2009), 1093-1097.
- [DKL, 14] Z. Dvir, J. Kollár, S. Lovett, *Variety evasive sets*. Comput. Complexity **23** (2014), 509-529.
- [DKSS, 09] Z. Dvir, S. Kopparty, S. Saraf, M. Sudan, *Extensions to the method of multiplicities, with applications to Kakeya sets and mergers*, 2009 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2009), 181-190, IEEE Computer Soc., Los Alamitos, CA, 2009.
- [FP, 13] M. Fernández, L. M. Pardo, *An Arithmetic Poisson Formula for the Multi-variate Resultant*. Journal of Complexity **29** (2013), 323-350.
- [FRR, 06] B. Felszeghy, B. Ráth, L. Rónyai, *The lex game and some applications*. J. of Symb. Comput. **41** (2006), 663-681.
- [Fr, 83] P. Frankl, *On the Trace of Finite Sets*. J. of Combinatorial Theory, Series A, **34** (1983), 41-45.
- [Fu, 83] W. Fulton, “*Intersection theory*”. New York-Heidelberg-Berlin Tokyo Springer, 1983.
- [GHJ, 54] L. Gillman, M. Henriksen, M. Jerison, *On a theorem of Gelfand and Kolmogoroff concerning maximal ideals in rings of continuous functions*. Proc. of the Amer. Math. Soc. **5** (1954), 447-455.
- [GHMPPP, 25] N. Giménez, J. Heintz, G. Matera, L. M. Pardo, M. Pérez, M. Privitelli, *Kronecker for Locally Closed Sets*. Sometido a Applicable Algebra in Engineering, Communication and Computing, 2025.

- [GH, 91] M. Giusti, J. Heintz, *La détermination des points isolés et de la dimension d'une variété algébrique peut se faire en temps polynomial*. In Computational algebraic geometry and commutative algebra, Cortona (1991), 216-256.
- [GHMP, 95] M. Giusti, J. Heintz, J. E. Morais, L. M. Pardo, *When polynomial equation systems can be "solved" fast?*. In Applied Algebra, Algebraic Algorithms and Error-Correcting Codes: 11th International Symposium, AAECC-11 Paris, France, July 17–22, 1995 Proceedings **11**, 205-231. Springer Berlin Heidelberg (1995).
- [GHHMP, 97] M. Giusti, K. Hägele, J. Heintz, J.E. Morais, J.L. Montaña, L. M. Pardo, *Lower bounds for diophantine approximations*. Journal of Pure and Applied Algebra **117** & **118** (1997), 277-317.
- [GHMP, 97] M. Giusti, J. Heintz, J.E. Morais, Luis M. Pardo, *Le rôle des structures des données dans les problèmes d'élimination*. Comptes Rendues Acad. Sci. Paris, Sér. I **325** (1997), 1223-1228.
- [GHMMP, 98] M. Giusti, J. Heintz, J. E. Morais, J. Morgenstern, L. M. Pardo, *Straight-line programs in geometric elimination theory*. Journal of pure and applied algebra, **124(1-3)**, 101-146.
- [GHS, 93] M. Giusti, J. Heintz, J. Sabia, *On the efficiency of effective Nullstellensätze*. Computational Complexity **3** (1995), 56-95.
- [GHSS, 95] M. Giusti, J. Heintz, J. Sabia, P. Solernó, F. Smietanski, *Sur la complexité du théorème des zéros*. In "Approximation and Optimization (in the Caribbean) **8**, J. Guddat *et al.* eds, Peter Lange, Verlag, 1995, 247-329.
- [GLS, 01] M. Giusti, G. Lecerf, B. Salvy, *A Gröbner free alternative for polynomial system solving*. J. of Complexity **17** (2001), 154-211.
- [GJ, 95] P. Goldberg, M. Jerrum, *Bounding the Vapnik-Chervonenkis dimension of concept classes parameterized by real numbers*. Machine Learning **18** (1995), 131-148.
- [GV, 96] D. Grigor'ev, N. Vorobjov, *Complexity lower bounds for computation trees with elementary transcendental function gates*. Theor. Comput. Sci. **157** (1996), 185-214.
- [Gr, 18] D. Yu. Grigor'ev, *Tropical Newton-Puiseux polynomials*. En Computer Algebra in Scientific Computing: 20th International Workshop, CASC 2018, Lille, France, September 17–21, 2018, Proceedings **20**, 177-186. Springer International Publishing (2018)
- [Gr, 20] D. Yu. Grigor'ev, *Decomposing tropical rational functions*. Journal of Symbolic Computation **101** (2020), 61-72.
- [Guo, 21] z. Guo, *Variety evasive subspace families*. In "36th Computational Complexity Conference", Leibniz International Proceedings in Informatics, 2021, 20:1-20:33.
- [HMPP, 21] A. Hashemi, J. Heintz, L. M. Pardo, P. Solernó, *On Bézout Inequalities for non-homogeneous polynomial ideals*. Journal of Symbolic Computation **106** (2021), 1-22.
- [Ha, 92] J. Harris, *Algebraic geometry: a first course*. Springer Science & Business Media, 1992.
- [Ha, 95] D. Haussler, *Sphere packing numbers for subsets of the Boolean n-cube with bounded Vapnik-Chervonenkis dimension*. J. Comb. Theory Ser. A **69** (1995), 217-232.
- [HLW, 94] D. Haussler, N. Littlestone, M. K. Warmuth, *Predicting $\{0, 1\}$ functions on randomly drawn points*. Information and Computation, **115** (1994), 284-293.
- [HMPS, 00] K. Hägele, J. E. Morais, L. M. Pardo, M. Sombra, *On the intrinsic complexity of the arithmetic Nullstellensatz*. Journal of Pure and Applied Algebra, **146(2)**, 103-183.
- [He, 83] J. Heintz, *Fast Quantifier Elimination for Algebraically Closed Fields*. Theoret. Comput. Sci. **24** (1983), 239-277.
- [He, 85] J. Heintz, *Corrigendum: Definability and Fast Quantifier Elimination in Algebraically Closed Fields*. Theoret. Comput. Sci. **39** (1985), 343.
- [He, 89] J. Heintz, *On the computational complexity of polynomials and bilinear mappings*. Proc. AAECC-5, Springer **356** (1989), 269-300.
- [HMW, 01] J. Heintz, G. Matera, A. Waissbein, *On the time-space complexity of geometric elimination procedures*. Applicable Algebra in Engineering, Communication and Computing **11** (2001), 239-296.
- [HOPRS, 23] J. Heintz, H. Ocar, L. M. Pardo, A. Rojas Paredes, C. Segura, *An unfeasibility view of neural network learning*. J. of Complexity **75** (2023), 101710.
- [HS, 82] J. Heintz, C. P. Schnorr, *Testing polynomials which are easy to compute*. Logic and algorithmic: An international symposium held in honor of Ernst Specker, Monographie No. **30** de l'Enseignement Mathématique, 1982, 237-254.
- [HS, 80] J. Heintz, M. Sieveking, *Lower bounds for polynomials with algebraic coefficients*. Theoret. Comput. Sci. **11** (1980), 321-330.
- [Hew, 48] E. Hewitt, *Rings of real-valued continuous functions. I*. Trans. Amer. Math. Soc. **64** (1948), 45-99.
- [HL, 20] J. van der Hoeven, G. Lecerf, *Fast multivariate multi-point evaluation revisited*. Journal of Complexity **56** (2020), 101405.
- [HL, 21] J. van der Hoeven, G. Lecerf, *Fast amortized multi-point evaluation*. Journal of Complexity **67** (2021), 101574.
- [Je, 05] Z. Jelonek, *On the effective Nullstellensatz*. Invent. math. **162** (2005), 1-17.
- [Jo, 97] J. P. Jouanolou, *Formes d'inertie et résultant: un formulaire*. Advances in Mathematics **126** (1997), 119-250.
- [KM, 97] M. Karpinski, A. Macintyre, *Polynomial Bounds for VC Dimension of Sigmoidal and General Pfaffian Neural Networks*. J. of Comput. & Syst. Sci. **54** (1997), 169-176.
- [KU, 08] K. Kedlaya, C. Umans, *Fast modular composition in any characteristic*. In "49th Annual IEEE Symposium on Foundations of Computer Science", IEEE, 2008, 146-155.
- [KU, 11] K. Kedlaya, C. Umans, *Fast polynomial factorization and modular composition*. SIAM Journal on Computing **40** (2011), 1767-1802.
- [Kh, 91] A. G. Khovanski, *"Fewnomials"*. American Mathematical Society, Providence, RI, 1991. Translated from the Russian by Smilka Zdravkovska.
- [Ko, 99] P. Koiran, *Elimination of Parameters in the Polynomial Hierarchy*. Theor. Comput. Sci. **215** (1999), 289-304.
- [Ko, 88] J. Kollár, *Sharp effective Nullstellensatz*. J. Am. Math. Soc. **1** (1988), 963-975.
- [KP, 94] T. Krick, L. M. Pardo, *Une approche informatique pour l'approximation diophantienne*. Comptes Rendues Acad. Sci. Paris, Sér. I **318** (1994), 407-412.
- [KP, 96] T. Krick, L. M. Pardo, *A computational method for diophantine approximations*. In "Algorithms in Algebraic Geometry", L. González Vega, T. Recio (eds.), Progress in Mathematics **143**, Birkhauser Verlag, 1996, 193-253.
- [LW, 54] S. Lang, A. Weil, *Number of points of varieties in finite fields*. American Journal of Mathematics **76** (1954), 819-827.

- [MM, 82] E. Mayr, E. Meyer, *The complexity of the word problems for commutative semigroups and polynomial ideals*. Advances in Mathematics **46** (1982), 305-329.
- [Me, 19] T. Mészáros, *Standard monomials and extremal point sets*. Discrete Mathematics, **343** (2020), 111785.
- [Mi, 06] G. Mikhalkin, *Tropical geometry and its applications*. En Proc. ICM Madrid, Spain, 2006, Eur. Math. Soc., 2006, 827-852.
- [Mi, 64] J. Milnor, *On the Betti numbers of real varieties*. Proceedings of the American Mathematical Society **15.2** (1964), 275-280.
- [MP, 09] J.L. Montaña, L.M. Pardo, *On the Vapnik-Chervonenkis dimension of computer programs which use transcendental elementary operations*. Annals of Math. and Artif. Intel. **56** (2009), 371-388.
- [MPC, 08] J.L. Montaña, L.M. Pardo, M. Callau, *VC-dimension bounds for analytic algebraic computations*. In Proc. Computing and Combinatorics, 14 Annual International Conference. Lecture Notes in Computer Science, **5092**, 62-71. Springer, Berlin (2008).
- [Mo, 16] T. Mora, “*Solving Polynomial Equation Systems IV: Volume 4, Buchberger Theory and Beyond*”. Encyclopedia of Mathematics and its Applications **158**, Cambridge University Press, 2016.
- [Mo, 55] T. S. Motzkin, *Evaluation of polynomials and evaluation of rational functions*. In Bull. Amer. Math. Soc. **61.163** (1955), 10.
- [Mu, 87] K. Mulmuley, *A fast parallel algorithm to compute the rank of a matrix over an arbitrary field*. Combinatorica **7** (1987), 101-104.
- [Na, 88] B. K. Natarajan, *Some results on learning*. Unpublished manuscript, 1988.
- [Na, 89] B. K. Natarajan, *On learning sets and functions*. Machine Learning **4** (1989), 67-97.
- [NRS, 20] V. Neiger, J. Rosenkilde, G. Solomatov, *Generic bivariate multi-point evaluation, interpolation and modular composition with precomputation*. In “Proceedings of the 45th International Symposium on Symbolic and Algebraic Computation”, A. Mantzaflaris, ed., 2020, 388-395.
- [Ol, 51] O. Oleinik, *Estimates of the Betti numbers of real algebraic hypersurfaces*. Mat. Sb. N.S. **28** (1951), 635-640.
- [OP, 62] O. Oleinik, I. Petrovsky. *On the topology of real algebraic surfaces*. Izv. Akad. Nauk SSSR (in Trans. Am. Math. Soc.) **1** (1962), 399-417.
- [Ore, 22] Ø. Ore, *Über höhere Kongruenzen*. Norsk Mat. Forenings Skrifter, **8** (1922), 15 pages.
- [Os, 54] A. M. Ostrowski, *On two problems in abstract algebra connected with Horner's rule*. In Studies in Mathematics and Mechanics presented to Richard von Mises (1954), 40-48.
- [Pa, 95] L. M. Pardo, *How lower and upper complexity bounds meet in elimination theory*. In “Applied Algebra, Algebraic Algorithms and Error-Correcting Codes”, G. Cohen, M. Giusti & T. Mora, eds., Lecture Notes in Computer Science **948**, Springer Verlag, 1995, 33-69.
- [Pa, 00] L. M. Pardo, *Universal Elimination requires exponential running time*. En “Actas EACA'200”, A. Montes ed., Univ. Politècnica de Cataluña, 2000, 25-51.
- [Pa, 12] L. M. Pardo, *La Conjetura de Cook ($\delta P=NP?$). Parte I: Lo Básico*. La Gaceta de la RSME **5** 2012, 117-147.
- [Pa, 24] L.M. Pardo, *Exploring implications of Trace (Inversion) formula and Artin algebras in extremal combinatorics*. Applicable Algebra in Engineering, Communication and Computing, **35.1** (2024), 71-118.
- [Pa, 25] L.M. Pardo, *Draconis Venationes Enhciridion, Pras Prima*. Manuscrito en permanente preparación, 2025.
- [PSM, 04] L. M. Pardo, J. San Martín, *Deformation techniques to solve Generalised Pham Systems*. Theoretical Computer Science **315** (2004), 593-625.
- [PS, 22] L. M. Pardo, D. Sebastián, *A promenade through correct test sequences I: Degree of constructible sets, Bézout's Inequality and density..* J. of Complexity **68** (2022), 101588.
- [PS, 25] L. M. Pardo, D. Sebastián, *Erzeugungsgrad, VC-Dimension and Neural Networks with rational activation function*. En arXiv:2504.11345 [cs.LG]. <https://doi.org/10.48550/arXiv.2504.11345>. Sometido a Applicable Algebra in Engineering, Communication and Computing, 2025.
- [PSZ, 25] L. M. Pardo, D. Sebastián, J. Zhang, *Some comments on the expected out-degree and the average degree of the one-inclusion hyper-graph in multiclass learning*. Manuscrito en preparación.
- [PMRML, 17] T. Poggio, H. Mhaskar, L. Rosasco, B. Miranda, Q. Lioa, *Why and When Can Deep - bot Not Shallow - Networks Avoid the Curse of Dimensionality: a Review*. CBMM Memo No. **058**, 2017.
- [RBR, 06] B.I. Rubinstein, P.L. Bartlett, J. Hyam Rubinstein. *Shifting, one-inclusion mistake bounds and tight multiclass expected risk bounds*. In “Advances in Neural Information Processing Systems **19**, Proc. of the 2006 Conference”, MIT Press 2006, 1193-1200.
- [Sau, 72] N. Sauer, *On the density of families of sets*. J. of Combinatorial Theory, Series A, **13** (1972), 145-147.
- [Sax, 09] N. Saxena, *Progress on Polynomial Identity Testing*. Bulletin of the EATCS **99** (2009), 49-79.
- [Sax, 14] N. Saxena, *Progress on Polynomial Identity Testing - II*. In “Perspectives in Computational Complexity: The Somenath Biswas Anniversary Volume”, 2014, pages 131-146.
- [Sch, 80] J. P. Schwarz, *Fast Probabilistic Algorithms for Verification of Polynomial Identities*. J. of the A.C.M. **27** (1980), 701-717.
- [Sha, 77] I. R. Shafarevich, “*Basic Algebraic Geometry*”. Grundlehren der mathematischen Wissenschaften **213**, Springer, 1977.
- [She, 72] S. Shelah, *A combinatorial problem; stability and order for models and theories in infinitary languages*. Pacific J. of Mathematics **41** (1972), 247-261.
- [SS, 93] M. Shub, S. Smale, *Complexity of Bezout's theorem I: Geometric aspects*. J. of the Amer. Math. Soc. **6** (1993), 450-501.
- [So, 99] M. Sombra, *A sparse effective Nullstellensatz*. Adv. Appl. Math. **22** (1999), 271-295.
- [St, 69] S. A. Stepanov, *On the number of points of a hyperelliptic curve over a finite prime field*. Mathematics of the USSR-Izvestiya **3** (1969), 1103-1114.
- [St, 73] V. Strassen, *Vermeidung von Divisionen*. Journal für die reine und angewandte Mathematik **264** (1973), 184-202.
- [MS, 15] D. Maclagan, B. Sturmfels, “*Introduction to Tropical Geometry*”. American Mathematical Society, 2015.
- [Tao, 14] T. Tao, *Algebraic Combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory*. EMS Surveys Math. Sci. **1** (2014), 1-46.
- [Th, 65] R. Thom, *Sur l'Homologie des Variétés Algébriques Réelles*. In Differential and Combinatorial Topology (A Symposium in Honor of Marston Morse), pp. 255-265. Princeton University Press, Princeton (1965).
- [Va, 84] L. G. Valiant, *A theory of the Learnable*, Communications of the ACM, **27** (1984), 1134-1142.

- [VC, 71] V. N. Vapnik, A. Ya. Chervonenkis, *On the Uniform Convergence of Relative Frequencies of Events to Their Probabilities*. Theory Probab. Appl. **16** (1971), 264–280.
- [Vo, 84] W. Vogel, *Lectures on Bézout's Theorem*. Tata Inst. for Fundamenta Research, Springer-Verlag, 1984.
- [Wa, 68] H. E. Warren, *Lower bounds for approximation by nonlinear manifold*. Trans. Amer. Math. Soc. **133** (1968), 167-178.
- [We,49] A. Weil, *Numbers of solutions of equations in finite fields*. Bull. Amer. Math. Soc. **55** (1949), 497-508.
- [Wo, 99] T. Wolff, *Recent work connected with Kakeya problem*. In *Prospects in mathematics*, Amer. Math. Soc., 1999, 129-162.
- [ZNL, 18] L. Zhang, G. Naitzat, L. Lim, *Tropical geometry of deep neural networks*. In International Conference on Machine Learning, PMLR (2018), 5824-5832.
- [Zp, 79] R. Zippel, *Probabilistic Algorithms for Sparse Polynomials*. In “Symbolic and Algebraic Computation (EUROSAM’79)”, Lecture Notes in Computer Science **72**, Springer, 1979, 216-266.



RESUMEN EN CASTELLANO Y EN INGLÉS

Esta tesis establece conexiones entre la Geometría Algebraica y la Teoría de Aprendizaje Computacional. En primer lugar, cerramos la Teoría de la Intersección para conjuntos constructibles, aportando dos nociones de grado que satisfacen la Desigualdad de Bézout. Estos resultados nos permiten refinar y extender las técnicas de J. Heintz y C. P. Schnorr para obtener cotas finas sobre la existencia y densidad de conjuntos cuestores. Posteriormente, exploramos las conexiones entre familias de conjuntos constructibles, conjuntos cuestores y la teoría de Vapnik–Chervonenkis. Para ello, generalizamos la noción de Erzeugungsgrad de J. Heintz al caso constructible. Empleando esta noción, demostramos que la dimensión de Vapnik–Chervonenkis de una familia de clasificadores constructibles está linealmente acotada, salvo por unas cantidades logarítmicas basadas en la Teoría de la Intersección, por la dimensión de Krull del espacio de parámetros. Usando la relación anterior, estudiamos la probabilidad de encontrar conjuntos cuestores de longitud apropiada en variedades evasivas de dimensión positiva. Aplicamos los resultados anteriores al análisis de redes neuronales con función de activación racional. Finalmente, abordamos el aprendizaje multiclase. En este contexto, introducimos una nueva noción, el grado de salida promedio, y demostramos que se trata de un invariante central. Además, llevamos a cabo un estudio detallado de la noción de pseudo-cubo y de la técnica del shifting.

This thesis establishes connections between Algebraic Geometry and Computational Learning Theory. First, we complete the Intersection Theory for constructible sets by introducing two notions of degree that satisfy Bézout’s Inequality. These results allow us to refine and extend the techniques of J. Heintz and C. P. Schnorr to obtain sharp bounds on the existence and density of correct test sequences. Next, we explore the connections between families of constructible sets, correct test sequences and the Vapnik–Chervonenkis theory. To this end, we generalize the notion of Erzeugungsgrad, introduced by J. Heintz, to the constructible case. Using this notion, we prove that the Vapnik–Chervonenkis dimension of a family of constructible classifiers is, up to logarithmic factors based on Intersection Theory, linearly bounded by the Krull dimension of the parameter space. Using the previous relation, we analyze the density of short correct test sequences in evasive varieties of positive dimension. We then apply our results to the study of neural networks with rational activation function. Finally, we address multiclass learning. In this context, we introduce a new notion, the expected outdegree, and prove that it is a central invariant. Additionally, we provide a detailed study of the notion of pseudo-cube and the shifting technique.