



Universidad de Valladolid

Facultad de Ciencias

TRABAJO FIN DE MÁSTER

Máster en Investigación en Matemáticas

**Estimación de variedades algebraicas
a partir de conjuntos de datos
y el estudio de su dimensión**

Autor: Álvaro Samperio Valdivieso

Tutor/es: Antonio Campillo López

Índice

1. Dimensión de una variedad algebraica.	4
1.1. Dimensión de Krull.	5
1.2. Dimensión en variedades definidas por ideales monomiales.	10
1.3. Dimensión de Hilbert	21
2. Estimación de la dimensión de un conjunto de puntos	27
2.1. Caso lineal: Análisis de Componentes Principales	27
2.2. Caso general	31
2.2.1. Diagramas de dimensión	31
2.2.2. Estimadores de dimensión	32
3. Estimación de las ecuaciones que definen una variedad.	36
4. Resultados de los experimentos	47
4.1. Subespacio afín	47
4.1.1. Datos pertenecientes al subespacio	47
4.1.2. Datos con ruido	49
4.1.3. Datos con información faltante	51
4.2. Circunferencia	52
4.3. Grupo de rotaciones $SO(3)$	57
4.4. Cúbica alabeada	61

Introducción

El TFM estudia el ajuste de variedades algebraicas a conjuntos de datos desde la perspectiva de la geometría algebraica computacional y la ciencia de datos, haciendo especial enfoque en la dimensión de estos modelos de ajuste.

Uno de los principales problemas de la ciencia de datos es, partiendo de un conjunto C de m puntos, típicamente en \mathbb{R}^n , la búsqueda de un modelo de ajuste a los datos consistente en el conjunto de ceros de ecuaciones que relacionen las n variables, de tal manera que estas ecuaciones se cumplan en los datos excepto por un error razonable. Estas relaciones pueden ser muy útiles en aplicaciones prácticas para obtener conclusiones sobre el conjunto de partida que no se ven a simple vista, y/o para realizar una transformación de los datos que facilite un análisis posterior, especialmente si n es grande. Consideramos la posibilidad de que las ecuaciones se cumplan con un error, ya que en la mayoría de aplicaciones prácticas, los datos están representados con una cantidad finita de decimales, y pueden tener errores (por ejemplo debidos a un proceso de medida), aunque suponemos a lo largo del texto que en los datos del conjunto C hemos detectado y tratado los outliers antes de buscar cualquier modelo, ya que podrían tener un serio impacto en su búsqueda.

Los paradigmas de modelos son los subespacios lineales de dimensión $q < n$, $V \subset \mathbb{R}^n$, que se corresponden con $n-q$ ecuaciones lineales linealmente independientes entre las variables. La búsqueda de este tipo de modelos se realiza con Análisis de Componentes Principales (ACP), que es una herramienta de análisis multivariante que fue introducida por Pearson (1901), y desarrollada de manera independiente por Hotelling (1933). Los modelos lineales tienen un gran atractivo, ya que las relaciones lineales frecuentemente son fáciles de interpretar, y la proyección de los datos en el subespacio obtenido es especialmente sencilla y nos proporciona un conjunto de datos en dimensión menor que contiene gran parte de la información del conjunto de partida. Además, el ACP es una técnica ampliamente utilizada y estudiada que se puede realizar con métodos muy eficientes que están disponibles en muchos lenguajes de programación. Estos métodos son estables frente a la presencia de ruido en los datos o frente a la presencia de datos incompletos.

Sin embargo, en muchas aplicaciones prácticas existen variables con relaciones no lineales, por lo que en esos casos la aproximación anterior no es suficiente para encontrar un buen modelo. Una primera generalización de la búsqueda de modelos consiste en realizar una transformación de las variables previa a la aplicación del

ACP, que haga que los datos transformados pertenezcan a un subespacio lineal del espacio de llegada de la transformación. Tal es el caso, por ejemplo, de las variedades tóricas, que se transforman en un subespacio vectorial tras tomar logaritmos, por lo que se pueden estimar mediante ACP. Sin embargo, en general, encontrar este tipo de transformaciones es difícil a partir de los datos, y además la inclusión de potencias, logaritmos, etc, de las variables originales en un ACP puede hacer difícil la interpretación de sus resultados [9].

En un enfoque diferente, Paul Breiding, Sara Kalisnik, Bernd Sturmfels y Madeleine Weinstein, en 2018 en su artículo [2] proponen la búsqueda de modelos formados por ecuaciones polinomiales de cualquier grado utilizando técnicas de geometría algebraica computacional y topología. Al igual que en el caso lineal, nuestro modelo V entonces será el conjunto de ceros del ideal generado por estas ecuaciones, es decir, una variedad algebraica afín o proyectiva, que es la generalización natural de subespacio a grado mayor que 1.

El objetivo de este trabajo es realizar una aproximación a la estimación de variedades algebraicas reales a partir de conjuntos de datos sin hacer ningún supuesto sobre la distribución de los puntos en V . Para ello presentaremos brevemente algunos resultados de ACP para tratar el caso lineal, dado que las variedades lineales son el primer tipo de modelos que recomendamos buscar, y presentando algunas de las técnicas de [2] que se pueden aplicar posteriormente, si es necesario. En este extenso artículo se proponen técnicas de interés para la estimación de variedades basadas en diversas características de las mismas, como la dimensión, las ecuaciones, los grupos de homología, el grado, o las componentes irreducibles, de las cuales discutiremos acerca de las basadas en las dos primeras características, por considerar que se complementan especialmente bien para proporcionar información sobre los modelos a estimar.

La dimensión de una variedad algebraica es uno de sus invariantes más importantes, y admite múltiples caracterizaciones equivalentes. El primer capítulo del trabajo está dedicado al estudio de la dimensión, empezando por presentar la definición de dimensión de Krull, en términos del ideal de polinomios que se anulan en V . Esta caracterización, junto con algunos resultados de geometría algebraica sobre la relación entre ideales y polinomios nos permitirá entender el concepto de dimensión como una generalización natural de la dimensión lineal. Después presentaremos una caracterización equivalente a partir del polinomio de Hilbert, que nos permitirá establecer un algoritmo para calcularla en la práctica, en el que los ideales monomiales tendrán un papel fundamental.

Los dos capítulos siguientes están dedicados al estudio de varias técnicas que nos permitirán, a partir de un conjunto de datos C perteneciente a una variedad algebraica V salvo quizá un pequeño ruido, obtener información sobre el modelo V y eventualmente estimarlo. En el Capítulo 2 obtendremos estimadores de la dimensión de V a partir de C , empezando por el ACP para el caso lineal, y continuando con diagramas de dimensión construidos a partir de diversos estimadores para el caso no lineal.

El objetivo del Capítulo 3 es encontrar las ecuaciones del modelo V a partir de C . En el caso lineal el ACP nos proporciona también las ecuaciones del modelo, pero su obtención está estrechamente relacionada con la estimación de la dimensión, por lo que detallamos todo el desarrollo del ACP en el capítulo 2 para no entorpecer el texto. Para el caso general, se emplearán matrices de Vandermonde generalizadas, técnicas de álgebra lineal numérica y la información obtenida en la estimación de la dimensión. Además, aunque el objetivo de este texto no es realizar un estudio estadístico sobre la estimación de variedades algebraicas, en este capítulo 3 discutiremos sobre la evaluación de los modelos, su capacidad de generalización (estamos interesados en encontrar modelos con ecuaciones que se cumplan con un error razonable en observaciones de las mismas variables que no hayan sido usadas para obtener dicho modelo), y propondremos unas líneas generales sobre cómo podría plantearse el proceso de búsqueda y evaluación de modelos en un experimento en el que la variedad V sea desconocida.

Por último, en el Capítulo 4 presentaremos diversos experimentos de estimación de variedades conocidas a partir de muestras de puntos contenidas en ellas (salvo posiblemente por un pequeño ruido), y enfocaremos cada experimento en varias diferentes técnicas de entre todas las que hemos presentado anteriormente, utilizando programación en lenguaje R, Julia y Singular, para ver cómo contribuyen en la estimación de la variedad a partir de los puntos.

1. Dimensión de una variedad algebraica.

En este capítulo abordaremos el concepto de dimensión de una variedad algebraica, la cual es una de sus propiedades más importantes, y admite múltiples caracterizaciones equivalentes. Comenzaremos presentando la definición de dimensión de Krull, debido a que quizá resulta la caracterización más intuitiva. A continuación, estudiaremos el caso particular de la dimensión en un ideal monomial, que servirá de ejemplo sencillo de aplicación de la caracterización de Krull y de introducción de la caracterización de dimensión a partir del polinomio de Hilbert. Después desarrollaremos esta última en el caso general, y por último estableceremos la equivalencia entre ambas caracterizaciones. El estudio del polinomio de Hilbert de un ideal nos permitirá establecer un algoritmo para calcular la dimensión de una variedad en la práctica y además nos proporcionará información sobre otras de sus propiedades, como el grado o el género aritmético.

Previamente estableceremos varios conceptos elementales que serán recurrentes a lo largo del texto.

Un ideal I del anillo conmutativo con unidad 1 $\mathbb{K}[x_1, \dots, x_n]$, con \mathbb{K} un cuerpo, es un conjunto que cumple las siguientes propiedades:

1. $0 \in I$.
2. Si $f, g \in I$, entonces $f+g \in I$.
3. Si $f \in I, f \in \mathbb{K}[x_1, \dots, x_n] I$, entonces $fg \in I$.

Dado que -1 pertenece al anillo de polinomios, la propiedad 2 puede sustituirse por la condición: Si $f, g \in I$, entonces $f - g \in I$. Decimos que un ideal es primo si para cualquier $f, g \in \mathbb{K}[x_1, \dots, x_n]$ con $fg \in I$, se tiene que $f \in I$ o $g \in I$.

El anillo $\mathbb{K}[x_1, \dots, x_n]$ es noetheriano, luego cualquier ideal I en él es finitamente generado, es decir, existen polinomios f_1, \dots, f_m tal que:

$$I = \langle f_1, \dots, f_m \rangle := \left\{ \sum_{i=1}^m g_i f_i, \quad g_1, \dots, g_m \in \mathbb{K}[x_1, \dots, x_n] \right\}.$$

Dado un ideal $I \subset \mathbb{K}[x_1, \dots, x_n]$, una variedad algebraica afín, o variedad afín, $V(I)$ es el subconjunto de \mathbb{K}^n de los ceros comunes de todos los polinomios del ideal, es decir el conjunto de los elementos (a_1, \dots, a_n) de \mathbb{K}^n tales que $f(a_1, \dots, a_n) = 0$ para todo f de I . Si $I = \langle f_1, \dots, f_m \rangle$, $V(I)$ es igual al conjunto de ceros de f_1, \dots, f_m , y escribimos $V(f_1, \dots, f_m) := V(I)$. Si el ideal I es homogéneo, es decir, generado por polinomios homogéneos, llamamos variedad proyectiva al conjunto de ceros $V(I)$ en el espacio proyectivo \mathbb{P}^{n-1} .

Por último, denotamos por $I(V)$ al ideal generado por una variedad $V \subset \mathbb{K}^n$ o \mathbb{P}^{n-1} , que es el ideal de polinomios de $\mathbb{K}[x_1, \dots, x_n]$ que se anulan en V , es decir, en todos los puntos de V . Para cualquier ideal J , se tiene $J \subset I(V(J))$, aunque no tiene por qué darse la igualdad entre estos conjuntos. En el caso particular de que \mathbb{K} sea algebraicamente cerrado, el Teorema de los ceros de Hilbert [8] dice que, para todo J :

$$I(V(J)) = \sqrt{J} := \{f \in \mathbb{K}[x_1, \dots, x_n], \text{ tal que existe un } m \text{ con } f^m \in J\}.$$

Al conjunto \sqrt{J} lo llamamos ideal radical de J , y decimos que J es un ideal radical si $J = \sqrt{J}$.

Para no entorpecer el desarrollo del capítulo, expondremos los resultados y definiciones para variedades afines, (que llamaremos simplemente variedades), ya que la mayoría de estos tienen una contraparte proyectiva con pocos cambios con respecto a la afín. Al final de la sección 1.3 comentaremos los cambios más relevantes para el caso proyectivo, y relacionaremos la dimensión afín y proyectiva.

1.1. Dimensión de Krull.

Definición 1.1 *Sea una variedad $V \subset \mathbb{K}^n$. Llamamos dimensión de Krull de V [1] a la dimensión de Krull del anillo cociente $\mathbb{K}[x_1, \dots, x_n]/I(V)$, es decir, a la longitud d de la cadena más larga de ideales primos p_i tal que $p_0 \subsetneq p_1 \subsetneq \dots \subsetneq p_d \subset \mathbb{K}[x_1, \dots, x_n]/I(V)$.*

En los casos en que el cuerpo \mathbb{K} sea real o complejo, se puede probar la equivalencia de esta definición con una formulación geométrica muy intuitiva, a partir de la relación existente entre ideales primos y variedades irreducibles, que exponemos en los siguientes resultados, algunos de los cuales se pueden encontrar en [4] o [8].

Definición 1.2 *Decimos que una variedad $V \subset \mathbb{K}^n$ es irreducible si para cualquier expresión de V como unión de dos variedades, $V = V_1 \cup V_2$, se tiene que $V = V_1$ o $V = V_2$.*

En la introducción del capítulo hemos definido el concepto de variedad $V=V(I)$ a partir de un ideal I , y también hemos asignado a cada variedad el ideal $I(V)$ generado por esta. Entendiendo V como una aplicación del conjunto de ideales de $\mathbb{K}[x_1, \dots, x_n]$ en el de variedades de \mathbb{K}^n , e I como una aplicación de este último conjunto en el primero, obtenemos la siguiente relación entre ambos conjuntos.

Teorema 1.3 *Sea \mathbb{K} un cuerpo arbitrario. Definimos las siguientes aplicaciones,*

$$I : \{\text{Variedades en } \mathbb{K}^n\} \rightarrow \{\text{Ideales en } \mathbb{K}[x_1, \dots, x_n]\}$$

$$V : \{\text{Ideales en } \mathbb{K}[x_1, \dots, x_n]\} \rightarrow \{\text{Variedades en } \mathbb{K}^n\},$$

que asignan a cada variedad V el ideal $I(V)$ generado por ella, y a cada ideal I la variedad $V(I)$ que la define, respectivamente.

Estas aplicaciones invierten las relaciones de orden de inclusión de variedades y de ideales. Es decir, si $V_1 \subset V_2$, entonces $I(V_2) \subset I(V_1)$; y si $I_1 \subset I_2$, entonces $V(I_2) \subset V(I_1)$. Además, se tiene que para cualquier variedad W , $V(I(W)) = W$, por lo que I es inyectiva.

Demostración: Si $V_1 \subset V_2$, y f es un polinomio en $I(V_2)$, entonces se anula en todos los puntos de V_2 , y en particular en los de V_1 , luego $f \in I(V_1)$. Análogamente, si $I_1 \subset I_2$, $V(I_2)$ es el conjunto de ceros comunes a todos los polinomios de I_2 . En particular, los puntos de este conjunto son ceros comunes para todos los polinomios de I_1 , luego $V(I_2) \subset V(I_1)$.

Por último, dado un ideal $J \in \mathbb{K}[x_1, \dots, x_n]$, tomamos la variedad $W = V(J) \subset \mathbb{K}^n$, y veamos que $V(I(W)) = W$. Por definición, todos los polinomios en $I(W)$ se anulan en W , luego $W \subset V(I(W))$. Por otra parte, $J \subset I(W)$, ya que los polinomios de J se anulan en W . La inversión de orden demostrada en el párrafo anterior nos da la otra inclusión, $V(I(W)) \subset V(J) = W$. Hemos probado que $V \circ I = 1$, es decir, V es una inversa de I por la izquierda, por lo que I es inyectiva. \square

El siguiente lema nos permitirá probar a continuación un corolario del teorema anterior, que es una versión más fuerte del mismo para cuerpos algebraicamente cerrados.

Lema 1.4 *El ideal $I(V)$ generado por cualquier variedad V es un ideal radical*

Demostración: Tenemos que probar que $\sqrt{I(V)} = I(V)$. La inclusión de $I(V)$ en su radical es obvia a partir de la definición, y es una propiedad que se da para cualquier ideal. Para probar la otra inclusión, tomamos un elemento f de $\sqrt{I(V)}$. Entonces existe un m natural tal que $f^m \in I(V)$, por lo que, para cualquier $x \in V$, $f^m(x) = 0$. Esto implica que $f(x) = 0$, luego f es idénticamente nula en V , y $f \in I(V)$. \square

Corolario 1.5 *Sea \mathbb{K} un cuerpo algebraicamente cerrado. Las aplicaciones I y V del teorema anterior restringidas a los conjuntos siguientes,*

$$I : \{\text{Variedades en } \mathbb{K}^n\} \rightarrow \{\text{Ideales radicales en } \mathbb{K}[x_1, \dots, x_n]\}$$

$$V : \{\text{Ideales radicales en } \mathbb{K}[x_1, \dots, x_n]\} \rightarrow \{\text{Variedades en } \mathbb{K}^n\},$$

son biyecciones y son una la inversa de la otra.

Demostración: Por el lema anterior, la imagen por I de cualquier variedad es radical, luego podemos ver I como una aplicación entre estos conjuntos. Sabemos

que $V \circ I = \text{Id}$, por lo que únicamente resta probar que V es inversa de I por la derecha también. Para cualquier ideal J , el Teorema de los ceros de Hilbert nos dice que $I(V(J)) = \sqrt{J}$, y como el conjunto de partida de V es el de ideales radicales, $\sqrt{J} = J$, con lo que tenemos el resultado. \square

El Teorema 1.3 y su corolario en el caso algebraicamente cerrado (1.5), nos permiten obtener el siguiente resultado que relaciona las variedades irreducibles en \mathbb{K}^n y los ideales primos en $\mathbb{K}[x_1, \dots, x_n]$, y su respectivo corolario en el caso algebraicamente cerrado.

Proposición 1.6 *Sea \mathbb{K} un cuerpo y $V \subset \mathbb{K}^n$ una variedad. Entonces V es irreducible si y solo si el ideal $I(V)$ que genera es primo.*

Demostración: Para probar la primera implicación, si V es irreducible y $fg \in I(V)$, entonces para cualquier punto de la variedad $x \in V$, se tiene $f(x)=0$ o $g(x)=0$, por lo que $V \subset V(f) \cup V(g)$. Haciendo la intersección de este último conjunto con V , podemos escribir $V = (V(f) \cup V(g)) \cap V = (V(f) \cap V) \cup (V(g) \cap V)$. Esta última expresión de V es la unión de dos variedades, ya que la intersección de dos variedades es una variedad (trivialmente a partir de la definición de variedad). Por hipótesis, V debe ser igual a una de estas variedades. Pongamos por ejemplo $V = (V(f) \cap V)$, lo que implica que $f=0$ en V , y por tanto $f \in I(V)$, con lo que $I(V)$ es primo.

Para probar la otra implicación, suponemos que $V = V_1 \cup V_2$, con $V \neq V_1$, y $I(V)$ es primo. Si probamos que $I(V) = I(V_2)$, tenemos el resultado, ya que $V = V_2$ por ser I inyectiva. Como $V_2 \subset V$, por la inversión de orden de la aplicación I , $I(V) \subset I(V_2)$. Para demostrar la inclusión opuesta, tomamos $g \in I(V_2)$. Como $V_1 \subsetneq V$, debe ser $I(V) \subsetneq I(V_1)$, por la inyectividad de I ; luego podemos tomar $f \in I(V_1) - I(V)$. Cualquier elemento de $V = V_1 \cup V_2$ es un cero de f o de g , luego $fg=0$ en V , y por tanto $fg \in I(V)$. Por hipótesis $I(V)$ es primo, luego $g \in I(V)$, ya que hemos tomado $f \notin I(V)$. \square

Corolario 1.7 *Sea \mathbb{K} un cuerpo algebraicamente cerrado. Las aplicaciones I y V inducen una correspondencia uno a uno entre variedades en \mathbb{K}^n e ideales primos en $\mathbb{K}[x_1, \dots, x_n]$.*

Demostración: Cualquier ideal I primo es radical, ya que si $f^m \in I$, escribiendo $f^m = f^{m-1}f$, o bien $f \in I$, en cuyo caso I es radical, o bien $f \notin I$, lo que implicaría $f^{m-1} \in I$, pero si se diera este último caso, podríamos escribir $f^{m-1} = f^{m-2}f$, y tendríamos $f^{m-2} \in I$, e iterando este proceso, $f^j \in I$ para cualquier natural $j < m$, lo cual es absurdo.

Por lo tanto, podemos restringir las aplicaciones I y V del Corolario 1.5 al conjunto de variedades irreducibles y al conjunto de ideales primos, lo que nos da la biyección entre estos. \square

Por otro lado, una de las propiedades de la definición de dimensión de Krull es que es finita para cualquier variedad algebraica.

Proposición 1.8 *Sean V_1 y V_2 variedades en \mathbb{K}^n . Si $V_1 \subset V_2$, entonces $\dim(V_1) \leq \dim(V_2)$. En particular, la dimensión de cualquier variedad en \mathbb{K}^n es menor o igual que n .*

Demostración: Por la inversión del orden de la aplicación I , los ideales generados por V_1 y V_2 cumplen $I(V_2) \subset I(V_1)$. Las cadenas de ideales primos en $\mathbb{K}[x_1, \dots, x_n]$ que contienen a un ideal I se corresponden con las cadenas de ideales primos en el cociente $\mathbb{K}[x_1, \dots, x_n]/I$ [8]. Si una cadena de ideales primos contiene a $I(V_1)$, también contiene a $I(V_2)$, por lo que el máximo de longitud de las cadenas que contienen a $I(V_1)$ no puede ser estrictamente mayor que el máximo de las que contienen a $I(V_2)$, lo que nos da el primer resultado, $\dim(V_1) \leq \dim(V_2)$.

Por otra parte, $I(\mathbb{K}^n) = \langle 0 \rangle$, ya que 0 es el único polinomio que se anula en todo \mathbb{K}^n . El anillo $\mathbb{K}[x_1, \dots, x_n]/\langle 0 \rangle$ es isomorfo a $\mathbb{K}[x_1, \dots, x_n]$, el cual, por la sección 14 del capítulo 5 de [11], es un anillo catenario, es decir, las cadenas de ideales primos irrefinables contenidas en él tienen todas la misma longitud. Claramente la cadena de ideales primos $\langle 0 \rangle \subset \langle x_1 \rangle \subset \langle x_1, x_2 \rangle \subset \dots \subset \langle x_1, \dots, x_n \rangle$ es irrefinable, por lo que $\mathbb{K}[x_1, \dots, x_n]$ tiene dimensión de Krull n , luego la dimensión de \mathbb{K}^n es n . Aplicando la primera parte de esta proposición, tenemos que la dimensión de cualquier variedad contenida en \mathbb{K}^n es menor o igual que n . \square

Los resultados de esta sección nos sirven para exponer la caracterización geométrica que estábamos buscando de la definición de dimensión de Krull, en los casos complejo y real, en términos de variedades irreducibles.

Teorema 1.9 *Sea una variedad $V \subset \mathbb{K}^n$, con \mathbb{K} igual a \mathbb{R} o \mathbb{C} . La dimensión de V es igual a la longitud d de la cadena más larga de variedades irreducibles V_i tal que $V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_d \subset V$.*

Demostración: Tomamos una variedad $V \subset \mathbb{K}^n$ de dimensión d . Trataremos primero el caso en que \mathbb{K} es algebraicamente cerrado. Como podemos ver en [8], las cadenas de ideales primos en $\mathbb{K}[x_1, \dots, x_n]/I(V)$ se corresponden biyectivamente con las cadenas de ideales primos en $\mathbb{K}[x_1, \dots, x_n]$ que contienen a $I(V)$; y estas a su vez, por el Corolario 1.7, se corresponden con las cadenas de variedades irreducibles contenidas

en V , Por lo tanto, la longitud máxima de estas últimas es igual a la dimensión de Krull.

Por otra parte, veamos el caso real. Tomamos $V \subset \mathbb{R}^n$ de dimensión d . Para empezar, no puede existir una cadena de variedades irreducibles de longitud $m > d$ tal que $V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_m \subset V$, ya que la cadena de ideales generados por las variedades de esta sería $I(V_0) \supsetneq I(V_1) \supsetneq \dots \supsetneq I(V_m) \supset I(V)$, de longitud m por la inyectividad de I , y tomando el cociente, tendríamos una cadena de ideales primos en $\mathbb{K}[x_1, \dots, x_n]/I(V)$, $\mathbb{K}[x_1, \dots, x_n]/I(V) \supsetneq I(V_0)/I(V) \supsetneq I(V_1)/I(V) \supsetneq \dots \supsetneq I(V_m)/I(V) \supset \langle 0 \rangle$, de longitud $m > d$, lo que contradice que la dimensión de Krull de V sea d .

Hemos probado que la longitud de la cadena más larga de variedades irreducibles es menor o igual que d . Para probar que esta longitud es exactamente igual a d , no podemos utilizar el mismo argumento que en el caso algebraicamente cerrado, ya que la aplicación V no tiene por qué ser inyectiva. En su lugar, utilizamos un resultado de la sección 3.1 de [12], que dice que, dada una variedad $V \subset \mathbb{R}^n$ de dimensión d , existe una cadena de variedades irreducibles de longitud d , $V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_d \subset V$, utilizando un resultado de Dubois y Efrogmson, con lo que tenemos el resultado. \square

La siguiente proposición nos ayudará a complementar esta interpretación geométrica de la dimensión.

Proposición 1.10 *Cualquier variedad $V \subset \mathbb{K}^n$ se puede expresar de forma única como unión finita de variedades irreducibles, $V = V_1 \cup V_2 \cup \dots \cup V_m$ de tal manera que ninguna de estas variedades esté contenida en otra, $V_i \not\subset V_j$ para todo i distinto de j . Además, la dimensión de V es igual al máximo de las dimensiones de las variedades de esta descomposición, (que denominaremos descomposición minimal).*

Demostración: Para probar la existencia una descomposición de V procedemos por reducción al absurdo. Si V no se puede escribir como unión finita de variedades irreducibles, en particular no es irreducible, luego existe una descomposición $V = V_1 \cup V'_1$, donde $V \neq V_1$ y $V \neq V'_1$. Al menos una de las dos variedades, (pongamos V_1), no se puede escribir como unión finita de variedades irreducibles, con lo que existe una descomposición $V_1 = V_2 \cup V'_2$, donde $V_1 \neq V_2$ y $V_1 \neq V'_2$. Iterando este razonamiento, obtenemos una cadena infinita de variedades $V \supsetneq V_1 \supsetneq \dots \supsetneq V_m \dots$

Por la inyectividad de I , la cadena de ideales generados por estas variedades es infinita $I(V) \subsetneq I(V_1) \subsetneq \dots \subsetneq I(V_m) \dots$, sin embargo, no pueden existir cadenas infinitas de ideales en $\mathbb{K}[x_1, \dots, x_n]$ por noetherianidad. En efecto, el ideal consistente en la unión de todos los ideales de la cadena debe ser finitamente generado, es decir, existen polinomios f_1, \dots, f_t tal que $I(V) \cup I(V_1) \cup \dots \cup I(V_m) \dots = \langle f_1, \dots, f_t \rangle$. Como cada uno de los polinomios generadores f_i pertenece a esta unión, debe existir un índice $j(i)$ tal que $f_i \in V_{j(i)}$ (llamando $V_j := V$). Tomando un índice N mayor que los índices $j(i)$ de

todos los polinomios, estos pertenecen a $I(V_N)$, luego $\langle f_1, \dots, f_t \rangle = I(V_N) = I(V_{N+1}) = \dots$, con lo que llegamos a contradicción.

Hemos probado que, para cualquier variedad V existe una expresión como unión finita de variedades irreducibles. Eliminando las variedades irreducibles que estén contenidas en otra, conseguimos $V = V_1 \cup V_2 \cup \dots \cup V_m$ con $V_i \not\subset V_j$ para todo i distinto de j . Para probar la unicidad de la descomposición con esta propiedad, supongamos que $V = V'_1 \cup V'_2 \cup \dots \cup V'_m$ es otra descomposición con esta propiedad. Entonces $V'_1 \subset V = V_1 \cup V_2 \cup \dots \cup V_m$, luego $V'_1 = \bigcup (V'_1 \cap V_i)$. V'_1 es irreducible, por lo que existe un i con $V'_1 \subset V_i$ (cambiando de nombre a las variedades, podemos poner $i=1$), y por un razonamiento análogo, existe un j tal que $V_1 \subset V'_j$. Combinando las dos contenciones, tenemos $V'_1 \subset V'_j$, por lo que debe ser $j=1$ y $V'_1 = V_1$. Eliminando esta variedad, tenemos $V_2 \cup \dots \cup V_m = V'_2 \cup \dots \cup V'_m$, con lo que podemos aplicar el mismo proceso a esta variedad, y por recurrencia, probamos la unicidad.

Por último, la propiedad $\dim(V) = \max(\dim(V_1), \dim(V_2), \dots, \dim(V_m))$ se obtiene a partir de la igualdad $I(V) = I(V_1) \cap I(V_2) \cap \dots \cap I(V_m)$ [1]. \square

Estos dos últimos resultados nos permiten ver una variedad real o compleja como una unión finita de variedades irreducibles, que podemos entender como los «bloques básicos» en el conjunto de variedades, cada una de las cuales puede ser de diferente dimensión. Si la dimensión de una de estas variedades irreducibles V es d , significa que existe una cadena de variedades irreducibles $V_0 \subsetneq V_1 \subsetneq \dots \subsetneq V_d = V$. En particular, V contiene «bloques» de dimensión $d-1$, pero no es la unión de una cantidad finita de ellos, de igual manera que un plano, que tiene dimensión 2 como espacio vectorial contiene rectas, que son de dimensión 1, pero no está formado por una cantidad finita de ellas. Estos bloques de dimensión $d-1$ contienen a su vez bloques de dimensión $d-2$, y este proceso de pasar a los bloques de una dimensión menor se puede realizar d veces. Vemos que de esta manera, la dimensión de Krull es igual a la dimensión como espacio vectorial en espacios vectoriales y es una extensión natural de esta a variedades algebraicas arbitrarias.

1.2. Dimensión en variedades definidas por ideales monomiales.

La definición de Krull nos ha permitido entender la dimensión de una variedad algebraica de una forma relativamente sencilla y extraer conclusiones geométricas interesantes, pero su cálculo de manera directa, buscando la máxima longitud de ideales primos en un anillo cociente, puede ser realmente complejo o imposible. La definición de dimensión a partir del polinomio de Hilbert, aunque aparentemente más opaca que la anterior, es mucho más fácil de manejar desde un punto de vista

algorítmico.

En esta sección estudiaremos el concepto de dimensión en el caso de variedades definidas por ideales monomiales, es decir, variedades $V(I)$ donde $I \subsetneq \mathbb{K}[x_1, \dots, x_n]$ es un ideal que admite un sistema de monomios generadores, $I = \langle m_1, \dots, m_t \rangle$, que son un ejemplo en el que es sencillo probar la equivalencia entre ambas definiciones, y además juegan un papel clave en el algoritmo de cálculo de dimensión para variedades generales que presentaremos en la siguiente sección, lo que justifica su estudio detallado. No consideramos en la discusión la variedad definida por el ideal total $I = \mathbb{K}[x_1, \dots, x_n]$, que es el conjunto vacío.

Para empezar, estudiaremos cómo es la variedad definida por un ideal monomial. Algunos de los resultados se pueden encontrar en el capítulo 9 de [4]. Comenzaremos dando la siguiente definición.

Definición 1.11 *Sea un espacio vectorial \mathbb{K}^n . Llamamos subespacio coordinado a cualquier subespacio vectorial de \mathbb{K}^n definido fijando un subconjunto x_{i_1}, \dots, x_{i_k} de las coordenadas x_1, \dots, x_n igual a 0. Es decir, a cualquier variedad algebraica de la forma $V(x_{i_1}, \dots, x_{i_k})$.*

A partir de la definición, es sencillo ver que cualquier subespacio coordinado tiene dimensión $n-k$ como espacio vectorial, con k el número de coordenadas fijadas a 0, ya que es la intersección de los k hiperplanos de ecuación $x_{i_j}=0$. Su dimensión de Krull también es $n-k$. Para verlo, supongamos sin pérdida de generalidad que las coordenadas que se anulan son las k primeras, es decir, el subespacio coordinado es $V(x_1, \dots, x_k)$, entonces $I(V) = \langle x_1, \dots, x_k \rangle$, por lo que el anillo cociente asignado a este subespacio es $\mathbb{K}[x_1, \dots, x_n]/I(V) \cong \mathbb{K}[x_{k+1}, \dots, x_n]$, que es de dimensión de Krull $n-k$, como comentamos en la prueba de la Proposición 1.8.

Las variedades definidas por ideales monomiales tienen una estructura especialmente sencilla: son uniones finitas de subespacios coordinados.

Proposición 1.12 *Sea I un ideal monomial en $\mathbb{K}[x_1, \dots, x_n]$. Entonces la descomposición minimal de $V(I)$ es una unión finita de subespacios coordinados de \mathbb{K}^n .*

Demostración: La variedad que define cualquier monomio $x_{i_1}^{a_1} \dots x_{i_r}^{a_r} \in \mathbb{K}[x_1, \dots, x_n]$ con exponentes $a_j > 0$, es la unión de subespacios coordinados $V(x_{i_1}^{a_1} \dots x_{i_r}^{a_r}) = V(x_{i_1}) \cup \dots \cup V(x_{i_r})$, ya que el producto de potencias de coordenadas de un punto se anula si se anula alguna de estas coordenadas y viceversa. Un ideal monomial I está generado por una cantidad finita de monomios, por lo que la variedad $V(I)$ que define es la intersección finita de uniones finitas de subespacios coordinados. Por la propiedad distributiva de unión e intersección, podemos escribir la variedad como unión finita de intersecciones de subespacios coordinados, que son subespacios coordinados, y

por tanto, variedades irreducibles. Eliminando de la unión de subespacios resultante aquellos que estén contenidos en otro, tenemos la descomposición minimal de $V(I)$.
 \square

Aplicando la Proposición 1.10, la dimensión de Krull de una variedad en \mathbb{K}^n definida por un ideal monomial es la dimensión del subespacio coordinado de mayor dimensión. Cada subespacio coordinado tiene dimensión $n-k$, donde k es el número de coordenadas que se anulan en él, luego la dimensión de la variedad es n menos el mínimo de los números de coordenadas que se anulan en cada subespacio de la descomposición minimal. A continuación presentamos un algoritmo para calcular esta dimensión, que será de interés en la siguiente sección para el cálculo de dimensión de variedades arbitrarias.

Algoritmo 1.13 *Sea I un ideal monomial en $\mathbb{K}[x_1, \dots, x_n]$ generado por los monomios m_1, \dots, m_t . Definimos, para $1 \leq j \leq t$ el conjunto*

$$M_j = \{q \in \{1, \dots, n\} \text{ tal que } x_q \text{ divide a } m_j\},$$

que consta de los índices de las variables que son factores de m_j . A continuación, tomamos el conjunto formado por todos los subconjuntos de $\{1, \dots, n\}$ que tienen intersección no vacía con cada uno de los conjuntos M_j anteriores,

$$L = \{C \subset \{1, \dots, n\} \text{ tal que } C \cap M_j \neq \emptyset \text{ para } 1 \leq j \leq t\}.$$

Entonces la dimensión de la variedad definida por I es:

$$\dim(V(I)) = n - \min(|C| \text{ tal que } C \in L),$$

donde $|C|$ denota el número de elementos de C .

Demostración: Comencemos viendo que esta expresión de la dimensión está bien definida. En efecto, al no considerar el caso en el que la variedad está generada por el anillo total $\mathbb{K}[x_1, \dots, x_n]$, todos los monomios m_j son distintos de 1, por lo que todos los M_j son no vacíos. Como consecuencia, L es no vacío, ya que $\{1, \dots, n\} \in L$, lo que implica que $n - \min(|C| \text{ tal que } C \in L)$ es un número entre 0 y n .

Para empezar, veamos que el subconjunto de índices $C \in L$ si y solo si el subespacio coordinado en el que se anulan exactamente las variables con índices en C está contenido en $V(I)$. Primeramente, podemos escribir $V(I = \langle m_1, \dots, m_t \rangle) = V(m_1) \cap \dots \cap V(m_t)$. Si tomamos $C = \{i_1, \dots, i_k\} \in L$, cualquier elemento $p \in V(x_{i_1}, \dots, x_{i_k})$ tiene las coordenadas $x_{i_1}, \dots, x_{i_k} = 0$. Como $C \in L$, para cualquier j hay alguna de estas coordenadas que es un factor de m_j , luego $m_j(p) = 0$ para todo j , y por tanto $V(x_{i_1}, \dots, x_{i_k}) \subset V(I)$. Para probar la implicación contraria, si tomamos $C = \{i_1, \dots, i_k\}$

$\notin L$, entonces existe un j_0 tal que ninguno de los x_{i_1}, \dots, x_{i_k} divide a m_{j_0} . Tomando $p \in V(x_{i_1}, \dots, x_{i_k})$ con coordenadas igual a 0 en las variables x_{i_1}, \dots, x_{i_k} y coordenadas igual a 1 en las demás variables, tenemos $m_{j_0}(p) = 1$, luego $p \notin V(m_{j_0})$ y por consiguiente, $V(x_{i_1}, \dots, x_{i_k}) \not\subset V(I)$.

Por último, por los comentarios previos a la preposición, la dimensión del subespacio coordinado definido igualando a 0 las variables con índices en C es $n - |C|$, y la dimensión de $V(I)$ es el máximo de las dimensiones de los subespacios coordinados de la descomposición minimal de $V(I)$. Este máximo se puede tomar sobre todos los subespacios coordinados contenidos en $V(I)$, notando que todos ellos están contenidos en alguno de los subespacios de la descomposición minimal y aplicando la Proposición 1.10, luego

$$\dim(V) = \max(n - |C| \text{ tal que } C \in L) = n - \min(|C| \text{ tal que } C \in L).$$

□

Con este algoritmo queda resuelto el problema de calcular la dimensión de la variedad generada por un ideal monomial desde el punto de vista de la definición de Krull, esto es, la longitud d de la cadena más larga de ideales primos en $\mathbb{K}[x_1, \dots, x_n]/I(V)$. Alternativamente, Hilbert propone una definición equivalente de dimensión basada en la estructura de este anillo cociente como \mathbb{K} -espacio vectorial.

El cociente $\mathbb{K}[x_1, \dots, x_n]/I(V)$ tiene dimensión infinita en general como \mathbb{K} -espacio vectorial, pero si restringimos el anillo total $\mathbb{K}[x_1, \dots, x_n]$ al espacio vectorial de polinomios de grado menor o igual que el número s , para cualquier $s \geq 0$, que llamaremos $\mathbb{K}[x_1, \dots, x_n]_{\leq s}$, y realizamos el espacio cociente de este espacio vectorial módulo el subespacio vectorial de polinomios de $I(V)$ de grado menor o igual que s , $I(V)_{\leq s}$, obtenemos una restricción al grado s que, cuando s es suficientemente grande, determina una aproximación del anillo cociente por un espacio vectorial cociente de dimensión finita. La idea fundamental de la caracterización de Hilbert es que la velocidad a la que crece la dimensión de estas «restricciones» $\mathbb{K}[x_1, \dots, x_n]_{\leq s}/I(V)_{\leq s}$ al aumentar el grado s está estrechamente ligada a la dimensión de la variedad.

Antes de formalizar esta idea, recordemos un resultado estándar de álgebra lineal [4] sobre la dimensión de espacios vectorial cociente en general.

Lema 1.14 *Sea W un subespacio vectorial de un espacio de dimensión finita V . Entonces V/W es un espacio vectorial de dimensión finita y*

$$\dim(V) = \dim(W) + \dim(V/W).$$

Demostración: Llamamos m a la dimensión de W y $m+n$ a la dimensión de V . Para la prueba es suficiente con ver que si tomamos una base de W $\{v_1, v_2, \dots, v_m\}$,

y la ampliamos a una base cualquiera de V , $\{v_1, \dots, v_m, v_{m+1}, \dots, v_{m+n}\}$ entonces las clases en el cociente V/W de los vectores que hemos añadido, $\{[v_{m+1}], \dots, [v_{m+n}]\}$, son una base del mismo.

Si tomamos una combinación lineal nula $a_1[v_{m+1}] + \dots + a_n[v_{m+n}] = 0$, entonces $[a_1v_{m+1} + \dots + a_nv_{m+n}] = 0$, con lo que $a_1v_{m+1} + \dots + a_nv_{m+n} \in W$, ya que su clase en el cociente es 0. Como $\{v_1, \dots, v_m\}$ es base de W , existirán b_1, \dots, b_m tal que $a_1v_{m+1} + \dots + a_nv_{m+n} = b_1v_1 + \dots + b_mv_m$. Por ser $\{v_1, \dots, v_m, v_{m+1}, \dots, v_{m+n}\}$ base de V , todos los coeficientes en la expresión anterior se deben anular, en particular $a_1, \dots, a_n = 0$, luego $a_1[v_{m+1}] + \dots + a_n[v_{m+n}]$ son linealmente independientes.

Por otro lado, para ver que $\{[v_{m+1}], \dots, [v_{m+n}]\}$ genera V/W , tomamos $[v] \in V/W$. Escribiendo v en la base de V que hemos tomando anteriormente, tenemos que existen a_i , $1 \leq i \leq m+n$, tal que

$$\begin{aligned} [v] &= [a_1v_1 + \dots + a_mv_m + a_{m+1}v_{m+1} + \dots + a_{m+n}v_{m+n}] \\ &= a_{m+1}[v_{m+1}] + \dots + a_{m+n}[v_{m+n}], \end{aligned}$$

ya que la clase de $a_1v_1 + \dots + a_mv_m$ es 0 porque pertenece a W , con lo que llegamos al resultado. \square

Empezamos la caracterización de la dimensión de Hilbert con la siguiente definición [4], que es válida para cualquier ideal en $\mathbb{K}[x_1, \dots, x_n]$.

Definición 1.15 *Sea I un ideal en $\mathbb{K}[x_1, \dots, x_n]$. La función de Hilbert de I es la función $HF_I : \mathbb{N} \rightarrow \mathbb{N}$ definida como:*

$$\begin{aligned} HF_I(s) &= \dim \mathbb{K}[x_1, \dots, x_n]_{\leq s} / I_{\leq s} \\ &= \dim \mathbb{K}[x_1, \dots, x_n]_{\leq s} - \dim I_{\leq s}, \end{aligned}$$

donde \dim denota la dimensión como \mathbb{K} -espacio vectorial (y la última igualdad es consecuencia del Lema 1.14).

En la próxima sección estudiaremos cómo es esta función para cualquier ideal, mientras que en el resto de la sección volveremos al caso de los ideales monomiales y las variedades definidas por estos. En este caso, la función de Hilbert se caracteriza de la siguiente manera.

Proposición 1.16 *Sea I un ideal monomial en $\mathbb{K}[x_1, \dots, x_n]$. Entonces, para todo $s \geq 0$, $HF_I(s)$ es igual al número de monomios de grado menor o igual que s que no pertenecen a I .*

Demostración: El conjunto de monomios de grado menor o igual que s es una base de $\mathbb{K}[x_1, \dots, x_n]_{\leq s}$, y el conjunto de monomios de grado menor o igual que s que

pertenecen a I es una base de $I_{\leq s}$, ya que un monomio pertenece al ideal I si y solo si es divisible por alguno de sus generadores y un polinomio pertenece a I si y solo si todos sus términos pertenecen a I .

Por lo tanto, la base que hemos tomado en $\mathbb{K}[x_1, \dots, x_n]_{\leq s}$ es una extensión de la que hemos tomado en $I_{\leq s}$, y por la demostración del Lema 1.14, las clases en el cociente $\mathbb{K}[x_1, \dots, x_n]_{\leq s}/I(V)_{\leq s}$ de los monomios de grado menor o igual que s que no pertenecen a $I_{\leq s}$ son una base del mismo, lo que nos da el resultado. \square

En el caso de variedades definidas por un ideal monomial que sean irreducibles (es decir, subespacios coordenados), podemos calcular la función de Hilbert de $I(V)$ con un argumento combinatorio elemental.

Lema 1.17 *Sea V un subespacio coordenado de \mathbb{K}^n de dimensión d . Entonces, para todo $s \geq 0$, la función de Hilbert del ideal $I(V)$ asociado es*

$$HF_{I(V)}(s) = \binom{d+s}{s},$$

y por tanto, es un polinomio en la variable s de grado d .

Demostración: Cualquier subespacio coordenado V de dimensión d en \mathbb{K}^n es el conjunto de ceros de $n-d$ variables. Sin pérdida de generalidad, tomamos las coordenadas de tal manera que $V = V(x_{d+1}, \dots, x_n)$. Por lo tanto, $I(V) = \langle x_{d+1}, \dots, x_n \rangle$, y para cada $s \geq 0$, tenemos $\mathbb{K}[x_1, \dots, x_n]_{\leq s}/I(V)_{\leq s} \cong \mathbb{K}[x_1, \dots, x_d]_{\leq s}$, identificando cada monomio en únicamente las primeras d variables con su clase en el cociente, y es un resultado conocido de combinatoria que el número de monomios en d variables de grado menor o igual que s es $\binom{d+s}{s}$. \square

Si bien, para cualquier variedad V definida por un ideal monomial podríamos utilizar la función de Hilbert para definir la dimensión de V como el grado de la función de Hilbert del ideal asociado una subvariedad de dimensión máxima de su descomposición minimal, (y de esta manera es obvio que la definición coincidiría con la de Krull), resulta más interesante a la hora de desarrollar resultados posteriores estudiar la función de Hilbert del ideal $I(V)$ y, más aún, de la del propio ideal monomial I .

Sin embargo, en el caso de que V sea reducible, la generalización del resultado interior no es inmediata, ya que si su descomposición minimal es $V = V_1 \cup \dots \cup V_m$, el ideal $I(V)$ es $I(V) = I(V_1) \cap \dots \cap I(V_m)$, y por la Proposición 1.16, tenemos $HF_{I(V)}(s) = \text{número de monomios que no están en } I(V)_{\leq s}$ (que denominaremos $|I(V)_{\leq s}^c|$). Escribiendo el complementario de una intersección como unión de complementarios, tenemos

$$HF_{I(V)}(s) = |I(V_1)_{\leq s}^c \cup \dots \cup I(V_m)_{\leq s}^c|,$$

es decir, la función de Hilbert evaluada en s es el número de monomios en una unión de conjuntos. Como esta unión puede tener intersección no vacía, no es tan sencillo contar el número de monomios como en el caso irreducible. En este caso general, veremos que la función de Hilbert es polinómica a partir de un cierto s_0 , para lo cual necesitaremos una versión graduada del Teorema de las sizigias de Hilbert.

Antes de presentar este teorema, definimos algunos conceptos sobre módulos y graduación, que podemos encontrar en detalle en [5].

Definición 1.18 *Sea A un anillo conmutativo y con elemento unidad 1_A . Un A -módulo $(M, +)$ es un grupo abeliano junto con una operación $\circ: A \times M \rightarrow M$, (escribimos $a \circ m := \circ(a, m)$), tal que para todo $a, b \in A$ y todo $x, y \in M$:*

1. $a \circ (x + y) = a \circ x + a \circ y$.
2. $(a + b) \circ x = a \circ x + b \circ x$.
3. $(ab) \circ x = a \circ (b \circ x)$.
4. $1_A \circ x = x$.

Decimos que un módulo es libre si tiene una base, es decir, si existe un conjunto $F \subset M$ tal que cualquier elemento $f \in M$ puede escribirse de manera única como una combinación lineal

$$f = a_1 f_1 + \dots + a_n f_n,$$

con $a_i \in A$ y $f_i \in F$.

Como espacio vectorial, un anillo de polinomios $\mathbb{K}[x_1, \dots, x_n]$ tiene una descomposición como suma directa de los subespacios vectoriales $\mathbb{K}[x_1, \dots, x_n]_s$ generados por los monomios de grado s , $\mathbb{K}[x_1, \dots, x_n] = \bigoplus_{s \in \mathbb{N}} \mathbb{K}[x_1, \dots, x_n]_s$. A continuación presentamos el concepto de graduación en módulos, que es una generalización de esta idea.

Definición 1.19 *Un A -módulo graduado M sobre R es un módulo con una familia de subgrupos $\{M_s, s \in \mathbb{Z}\}$ del grupo aditivo de M tal que:*

1. $M = \bigoplus_{s \in \mathbb{Z}} M_s$.
2. La descomposición anterior de M es compatible con el producto por elementos de A : $A_t M_s \subset M_{t+s}$ para todo $t \geq 0$ y todo $s \in \mathbb{Z}$.

Si tenemos un módulo graduado M , podemos cambiar los índices de la graduación, restando un número fijo t a todos ellos. La estructura resultante es un módulo graduado que llamaremos $M(t)$, y definimos como:

$$M(t) = \bigoplus_{s \in \mathbb{Z}} M(t)_s,$$

donde $M(t)_s = M_{t+s}$.

Además, si tenemos varios módulos graduados M_1, \dots, M_l , la suma directa de todos ellos es un módulo graduado [5], si tomamos, para todo s , los siguientes subgrupos:

$$(M_1 \oplus \dots \oplus M_l)_s = (M_1)_s \oplus \dots \oplus (M_l)_s.$$

Por supuesto, un anillo $A = \mathbb{K}[x_1, \dots, x_n]_s$, visto como A -módulo, es el ejemplo más sencillo de módulo graduado, con la graduación usual. La aplicación de los dos resultados anteriores nos permite ver que los módulos libres de la forma:

$$A(t_1) \oplus \dots \oplus A(t_l),$$

donde $d_1, \dots, d_l \in \mathbb{Z}$, son también módulos graduados. Este tipo de módulos aparecen en la siguiente definición.

Definición 1.20 *Sea M un A -módulo graduado. Una resolución libre graduada de M es una cadena de la forma*

$$\dots \rightarrow F_2 \rightarrow F_1 \rightarrow F_0 \rightarrow M \rightarrow 0$$

donde los conjuntos F_i son módulos libres graduados de la forma $A(t_1) \oplus \dots \oplus A(t_{l(i)})$; y los homomorfismos ϕ_j son graduados de grado 0 (es decir, para todo t , $\phi_j((F_j)_t) \subset (F_{j-1})_t$); y la sucesión es exacta (es decir, para todo j , $\text{Im}(\phi_{j+1}) = \text{ker}(\phi_j)$).

Si existe un m tal que $F_{m+1} = F_{m+2} = \dots = 0$, pero $F_m \neq 0$, decimos que la resolución es de longitud m .

Estas resoluciones libres están relacionadas con los módulos de sizigias de M , y su estudio detallado es un trabajo extenso que podemos encontrar en [5]. En este texto presentamos sin demostración uno de los resultados fundamentales de este estudio.

Teorema 1.21 *de las sizigias de Hilbert (versión graduada)*

Sea $A = \mathbb{K}[x_1, \dots, x_n]$. Entonces cualquier A -módulo graduado finitamente generado tiene una resolución libre graduada de longitud menor o igual que n .

La existencia de una resolución libre de longitud finita es el resultado clave que nos permitirá obtener información sobre la dimensión de los submódulos M_s de un módulo graduado M .

Proposición 1.22 *Sea $M = \bigoplus_{s \in \mathbb{Z}} M_s$ un A -módulo graduado finitamente generado, $A = \mathbb{K}[x_1, \dots, x_n]$. Entonces, existe un único polinomio $p(s)$, cuyos coeficientes son números racionales, y un número natural s_0 tal que, para todo $s \geq s_0$,*

$$\dim(M_s) = p(s).$$

Demostración: El teorema de las sizigias de Hilbert nos asegura que existe una resolución libre graduada de M de longitud $m \leq n$, que denotamos (omitiendo los términos de F_{m+2} en adelante),

$$0 \rightarrow F_m \rightarrow \dots \rightarrow F_1 \rightarrow F_0 \rightarrow M \rightarrow 0.$$

Como la resolución es graduada, podemos restringir cada término en la sucesión exacta anterior a los subgrupos de grado t en cada módulo, quedándonos, para grado s , con la sucesión:

$$0 \rightarrow (F_m)_s \rightarrow \dots \rightarrow (F_1)_s \rightarrow (F_0)_s \rightarrow M_s \rightarrow 0.$$

Como esta sucesión es exacta, la suma alternada de las dimensiones es 0 [5], es decir,

$$\dim(M_s) = \sum_{j=0}^m (-1)^j \dim((F_j)_s).$$

Cada uno de estos $(F_j)_s$ es un módulo graduado libre de la forma $A(t_1)_s \oplus \dots \oplus A(t_l)_s$. Por ser esta expresión una suma directa, la dimensión de $(F_j)_s$ es la suma de las dimensiones de estos subconjuntos $A(t_i)_s$. Definimos anteriormente el cambio de graduación del anillo A a $A(t_i)$ de tal manera que $A(t_i)_s = A_{t_i+s}$. La dimensión $\dim(A_{t_i+s})$ es el número de monomios de grado $t_i + s$ en n variables, que es igual a $\binom{n+(t_i+s)-1}{t_i+s}$, por un argumento combinatorio similar al que utilizamos en el Lema 1.17.

Existe un número s_i (diferente en general) tal que para s mayor o igual que él, cada número combinatorio $\binom{n+(t_i+s)-1}{t_i+s}$ es un polinomio en s . Tomando s mayor que cada uno de estos s_i asociados a cada $A(t_i)_s$ (para todo i), tenemos que $\dim((F_j)_s)$ es suma de polinomios, y por tanto, un polinomio. Análogamente, tomando un s lo bastante grande como para que $\dim((F_j)_s)$ para todo j sea un polinomio, tenemos que $\dim(M_s)$ es suma alternada de polinomios, con lo que probamos el resultado. \square

La aplicación de este resultado a cocientes $\mathbb{K}[x_1, \dots, x_n]/I$ de un anillo de polinomios sobre un ideal monomial I (o sobre un ideal homogéneo, que es un caso que incluye al monomial), nos permite obtener la información que buscábamos sobre la función de Hilbert de I . Si I es un ideal homogéneo, el cociente $\mathbb{K}[x_1, \dots, x_n]/I$ es un $\mathbb{K}[x_1, \dots, x_n]$ -módulo graduado finitamente generado, como podemos ver en [5]. Tomamos la graduación de manera que $\mathbb{K}[x_1, \dots, x_n]/I \cong \bigoplus_{t \in \mathbb{Z}} (\mathbb{K}[x_1, \dots, x_n]_t / I_t)$. Denominamos función de Hilbert proyectiva de I a la función $HFP_I : \mathbb{N} \rightarrow \mathbb{N}$ definida como:

$$\begin{aligned} HFP_I(s) &= \dim \mathbb{K}[x_1, \dots, x_n]_s / I_s \\ &= \dim \mathbb{K}[x_1, \dots, x_n]_s - \dim I_s. \end{aligned}$$

Aplicando el resultado anterior, existe un único polinomio $HPP(s)$, (que llamaremos polinomio de Hilbert proyectivo), y un natural s_0 tal que, para todo $s \geq s_0$ es igual a la función de Hilbert proyectiva de I . Utilizaremos este polinomio para definir la dimensión de una variedad proyectiva al final de la sección 1.3. Esto nos permite obtener un resultado análogo para la función de Hilbert afín.

Corolario 1.23 *Sea I un ideal homogéneo en $\mathbb{K}[x_1, \dots, x_n]$. Entonces existe un único polinomio $HP_I(s)$, (que llamaremos polinomio de Hilbert, o polinomio de Hilbert afín), y un natural s_0 tal que, para todo $s \geq s_0$ es igual a la función de Hilbert de I , $HF_I(s)$.*

Demostración: A partir de las definiciones de función de Hilbert (afín) y función de Hilbert proyectiva, se tiene ([4], Teorema 12(i) de la sección 9.3):

$$HFP_I(s) = HF_I(s) - HF_I(s - 1).$$

Como existe un entero a partir del cual la función de Hilbert proyectiva es polinómica, también existirá un entero con esta propiedad para la función de Hilbert (afín). \square

En particular, en lo que concierne al tema de esta sección, hemos probado que si I es un ideal monomial y V la variedad definida por él, la función de Hilbert de $I(V)$, $HF_{I(V)}(s)$ coincide con un polinomio, el polinomio de Hilbert $HP_{I(V)}(s)$, para s suficientemente grande. Por tanto, para este tipo de variedades podemos establecer la definición de dimensión de Hilbert.

Definición 1.24 *Sea una variedad $V \subset \mathbb{K}^n$ definida por un ideal monomial I . Llamamos dimensión de Hilbert de V al grado del polinomio de Hilbert del ideal generado por V , $I(V)$.*

A continuación probaremos la equivalencia entre esta caracterización de dimensión y la de Krull en variedades definidas por un ideal monomial.

Teorema 1.25 *Sea $V \subset \mathbb{K}^n$ una variedad definida por un ideal monomial. Entonces la dimensión de Krull y la dimensión de Hilbert de V son iguales.*

Demostración: En el caso de que V sea irreducible, el Lema 1.7 nos da la igualdad entre ambas caracterizaciones. En caso contrario, tomamos una variedad reducible V de dimensión de Krull d y descomposición minimal $V = V_1 \cup \dots \cup V_m$, donde V_1 es una variedad irreducible de dimensión máxima en esta descomposición.

Retomando la argumentación posterior al Lema 1.17, tenemos $I(V) = I(V_1) \cap \dots \cap I(V_m)$, y $HF_{I(V)}(s) = \text{número de monomios que no están en } I(V)_{\leq s} = |I(V)_{\leq s}^c|$, y

escribiendo el complementario de una intersección como unión de complementarios, tenemos

$$HF_{I(V)}(s) = |I(V_1)_{\leq s}^c \cup \dots \cup I(V_m)_{\leq s}^c|,$$

Además, el número de monomios que no están en $I(V)_{\leq s}$ es mayor o igual que el número de monomios que no están en $I(V_1)_{\leq s}$, y es menor o igual que la suma para $1 \leq j \leq m$ del número de monomios que no están en cada $I(V_j)_{\leq s}^c$, ya que los conjuntos de la unión que aparece en la expresión anterior pueden tener monomios en común. Es decir, para todo $s \geq 0$,

$$|I(V_1)_{\leq s}^c| \leq |I(V_1)_{\leq s}^c \cup \dots \cup I(V_m)_{\leq s}^c| \leq |I(V_1)_{\leq s}^c| + \dots + |I(V_m)_{\leq s}^c|,$$

o, equivalentemente,

$$HF_{I(V_1)}(s) \leq HF_{I(V)}(s) \leq HF_{I(V_1)}(s) + \dots + HF_{I(V_m)}(s).$$

El corolario 1.23 nos dice que existe un s_0 a partir del cual estas funciones de Hilbert coinciden con su respectivo polinomio de Hilbert, luego, para $s \geq s_0$,

$$HP_{I(V_1)}(s) \leq HP_{I(V)}(s) \leq HP_{I(V_1)}(s) + \dots + HP_{I(V_m)}(s).$$

Por el Lema 1.17 y la proposición 1.10, el primer término de la desigualdad es un polinomio en s con coeficiente dominante positivo y de grado igual a la dimensión de Krull d , ya que la dimensión de la variedad irreducible de dimensión máxima coincide con la de V .

Por el Lema 1.17, para cada j , $HP_{I(V_j)}(s)$ también debe ser un polinomio en s , y en este caso, la proposición 1.10 nos dice que su grado debe ser menor o igual que d , ya que en caso contrario, V_j sería de dimensión estrictamente mayor que V_1 . El tercer término en la desigualdad es la suma de estos polinomios, luego también es un polinomio de grado menor o igual que d con coeficiente dominante positivo. Además tenemos $HP_{I(V_1)}(s) \leq HP_{I(V_1)}(s) + \dots + HP_{I(V_m)}(s)$ para $s \geq s_0$, y al tomar s tendiendo a infinito, vemos que el término de la derecha debe tener grado exactamente d .

Por último, el término central $HP_{I(V)}(s)$ es un polinomio que debe ser menor o igual que un polinomio de grado d y coeficiente positivo y mayor o igual que otro polinomio con las mismas características, para $s \geq s_0$, y al tomar s tendiendo a infinito, también vemos que debe ser de grado d , con lo que completamos el resultado. \square

Concluimos esta sección presentando un resultado que relaciona las funciones de Hilbert de dos ideales que definen una variedad, que será de utilidad en el cálculo de la dimensión de variedades arbitrarias.

Proposición 1.26 *Sea $I \in \mathbb{K}[x_1, \dots, x_n]$ un ideal monomial propio. Entonces el grado del polinomio de Hilbert de I es igual a la dimensión de Hilbert de la variedad $V(I)$.*

La demostración de esta proposición se puede encontrar en el Teorema 9.2.6 y la Proposición 9.3.3 y de [4] y se basa, de manera análoga a la demostración del Teorema 1.25, en la expresión de los monomios que no están contenidos en $I_{\leq s}$ como una unión de ciertos conjuntos, y en un razonamiento sobre el grado de los polinomios que aparecen al contar parte de ellos. Este resultado nos indica que podemos utilizar en el cálculo de la dimensión de una variedad V definida por un ideal monomial cualquier ideal J que la defina, ($V = V(J)$), en vez del ideal $I(V)$.

Como observación, comentamos que, para una variedad V definida por un ideal monomial I , también coinciden las respectivas dimensiones de Krull de los anillos cocientes $\mathbb{K}[x_1, \dots, x_n]/I$ y $\mathbb{K}[x_1, \dots, x_n]/I(V)$, ya que el ideal $I(V)$ es el radical del ideal I y, puesto que los ideales primos son radicales, se tiene que coinciden las respectivas familias de ideales primos de $\mathbb{K}[x_1, \dots, x_n]$ que contienen a I y a $I(V)$. Por lo tanto, las dimensiones de Krull y de Hilbert para I y para $I(V)$ son iguales.

Veremos en la sección siguiente que este último resultado no es cierto en general para cualquier variedad, lo que nos da un ejemplo de que la generalización de los resultados de esta sección a variedades arbitrarias no es trivial.

1.3. Dimensión de Hilbert

En esta sección veremos cómo la definición de dimensión de Hilbert para ideales monomiales se puede extender de manera natural a variedades arbitrarias, de manera que ésta sea equivalente a la dimensión de Krull, y proporcionaremos un algoritmo para calcularla. Por último señalaremos información adicional que podemos extraer de la función de Hilbert.

En esta sección necesitaremos algunos conceptos básicos sobre orden monomial y bases de Groebner que presentamos brevemente y que se pueden encontrar en detalle en los capítulos 2 y 8 de [4]. Un orden monomial $>$ es un orden total en el conjunto de monomios de $\mathbb{K}[x_1, \dots, x_n]$ que es compatible con la multiplicación de monomios (en el sentido de que si tenemos una desigualdad de monomios $m_1 > m_2$, multiplicar a ambos lados de la desigualdad por un mismo monomio no altera la desigualdad), y que es un buen orden, (es decir, en cualquier subconjunto de monomios existe un elemento que es el más pequeño de ellos). Decimos que un orden es graduado si siempre que tengamos dos monomios m_1 y m_2 con $\text{grado}(m_1) > \text{grado}(m_2)$, entonces $m_1 > m_2$.

Sea $>$ un orden monomial. Dado un polinomio no nulo f , llamamos $LT(f)$ al término dominante, es decir, al monomio en f que sea mayor que todos los demás. Además, llamamos $LM(f)$ al monomio dominante de f , es decir, a $LT(f)$ dividido por su coeficiente en \mathbb{K} . Si I es un ideal, llamamos $LT(I)$ al conjunto de los términos dominantes de los polinomios no nulos de I .

Por otra parte, dado un ideal I y un orden monomial $>$, decimos que conjunto de polinomios $\{f_1, \dots, f_m\}$ es una base de Groebner del ideal I con respecto a $>$ si $\langle LT(I) \rangle = \langle LT(f_1), \dots, LT(f_m) \rangle$. Cualquier base de Groebner es un sistema de generadores de I , es decir, $I = \langle f_1, \dots, f_m \rangle$. Para cualquier ideal no nulo existen bases de Groebner, y existen algoritmos para calcular una de ellas, como el Algoritmo de Buchberger.

El principal resultado de esta sección [4] es el siguiente, que nos permitirá llevar el problema de cálculo de dimensión de variedades arbitrarias a variedades definidas por ideales monomiales.

Teorema 1.27 *Sea I un ideal en $\mathbb{K}[x_1, \dots, x_n]$, y sea $>$ un orden monomial graduado en $\mathbb{K}[x_1, \dots, x_n]$. Entonces el ideal monomial $\langle LT(I) \rangle$ tiene la misma función de Hilbert que I .*

Demostración: Sea s fijo. Por definición de función de Hilbert y el Lema 1.14, se tiene,

$$\begin{aligned} HF_I(s) &= \dim \mathbb{K}[x_1, \dots, x_n]_{\leq s} - \dim I_{\leq s} \\ HF_{\langle LT(I) \rangle}(s) &= \dim \mathbb{K}[x_1, \dots, x_n]_{\leq s} - \dim \langle LT(I) \rangle_{\leq s}, \end{aligned}$$

por lo que será suficiente ver que $I_{\leq s}$ y $\langle LT(I) \rangle_{\leq s}$ tienen la misma dimensión (como espacio vectorial).

Comenzamos tomando el conjunto de monomios dominantes de todos los polinomios de $I_{\leq s}$. Hay una cantidad finita de ellos, luego eliminando los polinomios $f \in I_{\leq s}$ que tengan el monomio dominante duplicado, existen $f_1, \dots, f_m \in I_{\leq s}$ tal que

$$\{LM(f) : f \in I_{\leq s}\} = \{LM(f_1), \dots, LM(f_m)\}.$$

Además, renombrando los polinomios, podemos suponer $LM(f_1) > \dots > LM(f_m)$.

Por un lado, veamos que $\{LM(f_1), \dots, LM(f_m)\}$ es una base de $\langle LT(I) \rangle_{\leq s}$. Es trivial que estos monomios son linealmente independientes. Para ver que son una base, por la Proposición 1.16, basta con ver que éstos son los monomios de grado menor o igual que s que pertenecen a $\langle LT(I) \rangle$, es decir, probar la igualdad,

$$\{LM(f) : f \in I \text{ y } \text{grado}(LM(f)) \leq s\} = \{LM(f_1), \dots, LM(f_m)\}.$$

Como el orden monomial es graduado, el grado de cualquier polinomio f y de su monomio dominante es el mismo, luego esta igualdad entre conjuntos se sigue a partir de la igualdad anterior.

Hemos probado que la dimensión de $\langle LT(I) \rangle_{\leq s}$ es m . Veamos que $\{f_1, \dots, f_m\}$ son una base de $I_{\leq s}$ para completar la demostración. Si tomamos cualquier combinación lineal nula $a_1 f_1 + \dots + a_m f_m = 0$. Si existiera algún coeficiente a_i no nulo,

tomando el i más pequeño para el que $a_i \neq 0$, no hay ningún término de ningún polinomio en la combinación que pueda cancelar el monomio $a_i f_i$, porque, al ser el orden graduado, todos los términos son de menor grado que $a_i f_i$, lo cual es absurdo; y por tanto, el conjunto es linealmente independiente. Para probar que $\{f_1, \dots, f_m\}$ genera $I_{\leq s}$, razonamos también por reducción al absurdo, tomamos un polinomio f de grado mínimo de entre los que pertenecen a $I_{\leq s}$ y no son combinación lineal de $\{f_1, \dots, f_m\}$. Por cómo hemos definido f_1, \dots, f_m , $LT(f)$ debe ser múltiplo de uno de ellos, $LT(f) = cf_i$. Entonces el polinomio $f - cf_i$ tiene grado menor que f , porque hemos cancelado su término dominante. Por tanto, $f - cf_i$ debe ser combinación lineal de f_1, \dots, f_m , lo que implica que f también, y llegamos a una contradicción. \square

En particular, este teorema implica que existe un s_0 tal que la función de Hilbert $HF_I(s)$ de cualquier ideal I es igual a un polinomio en s para $s \geq s_0$, que llamamos polinomio de Hilbert, como en el caso de ideales monomiales. Este hecho nos permite extender la definición de dimensión de Hilbert a variedades arbitrarias.

Definición 1.28 *Sea $V \subset \mathbb{K}^n$ una variedad algebraica. Llamamos dimensión de Hilbert de V al grado del polinomio de Hilbert del ideal generado por V , $I(V)$.*

Con esta definición, la caracterización de dimensión de Hilbert y de Krull también son equivalentes en variedades arbitrarias, como indicamos en la siguiente generalización del Teorema 1.25, cuya demostración se puede encontrar en el Teorema 5.6.36 de [10].

Teorema 1.29 *Sea $V \subset \mathbb{K}^n$ una variedad algebraica. Entonces la dimensión de Krull y la dimensión de Hilbert de V son iguales.*

Volviendo al problema de cálculo de la dimensión, una de las ventajas de la caracterización de Hilbert es que nos permite ver, aplicando el Teorema 1.27, que la dimensión de cualquier variedad es igual a la dimensión de una variedad definida por un ideal monomial, que podemos calcular con el Algoritmo 1.13. A continuación presentamos un algoritmo para calcular la dimensión de una variedad arbitraria V a partir del ideal que genera, $I(V)$.

Algoritmo 1.30 *Sea $V \subset \mathbb{K}^n$ una variedad algebraica, $I(V)$ el ideal generado por esta, y $>$ un orden monomial graduado. Obtenemos una base de Groebner respecto a $>$ para $I(V)$, $I(V) = \langle f_1, \dots, f_m \rangle$. A continuación aplicamos el Algoritmo 1.13 para obtener la dimensión de la variedad, (definida por un ideal monomial), $V(\langle LT(f_1), \dots, LT(f_m) \rangle)$, que es igual a la dimensión de V .*

Demostración: La dimensión de V es el grado del polinomio de Hilbert $HP_{I(V)}(s)$. Por el Teorema 1.27, este polinomio es igual a $HP_{\langle LT(I(V)) \rangle}(s)$. Por definición de

base de Groebner, $\langle LT(I(V)) \rangle = \langle LT(f_1), \dots, LT(f_m) \rangle$, por lo que la dimensión de V es el grado del polinomio de Hilbert del ideal $\langle LT(f_1), \dots, LT(f_m) \rangle$. Aplicando la Proposición 1.26, este grado es igual al de la variedad $V(\langle LT(f_1), \dots, LT(f_m) \rangle)$. Al ser esta variedad definida por un ideal monomial, el Algoritmo 1.13 permite obtener su dimensión, con lo que queda probado el resultado. \square

Si bien este algoritmo es sencillo, exige conocer el ideal generado por V , $I(V)$, lo cual no siempre es fácil en la práctica. Sin embargo, en el caso algebraicamente cerrado, esta dificultad desaparece por completo, pudiendo utilizar como entrada del algoritmo cualquier ideal que defina la variedad, en sustitución de $I(V)$. Este hecho es consecuencia del siguiente resultado, cuya demostración encontramos en [4].

Proposición 1.31 *Sea I un ideal en $\mathbb{K}[x_1, \dots, x_n]$. Entonces los polinomios de Hilbert de I y \sqrt{I} tienen el mismo grado.*

En el caso algebraicamente cerrado, por el Teorema de los ceros de Hilbert se tiene, para todo ideal J , $I(V(J)) = \sqrt{J}$, y aplicando la proposición anterior, tenemos;

$$\dim(V(J)) = \text{grado}(HP_{I(V(J))}) = \text{grado}(HP_{\sqrt{J}}) = \text{grado}(HP_J),$$

lo que prueba el siguiente corolario.

Corolario 1.32 *Sea I un ideal en $\mathbb{K}[x_1, \dots, x_n]$, con \mathbb{K} algebraicamente cerrado. Entonces la dimensión de la variedad $V(I)$ es el grado de HP_I .*

Una vez tratado el concepto de dimensión en variedades afines, comentamos brevemente el caso de variedades proyectivas, que podemos encontrar desarrollado en el capítulo 9 de [4]. La dimensión de una variedad proyectiva V , que genera un ideal $I(V)$, es igual al grado del polinomio proyectivo de Hilbert $PHP_{I(V)}$. Además, se tiene un resultado análogo al Teorema 1.27, que nos dice que este polinomio es igual a $PHP_{\langle LT(I(V)) \rangle}$.

El ideal de este polinomio es monomial, y para variedades proyectivas definidas por un ideal monomial, se tiene una descomposición como unión finita de subespacios coordenados proyectivos, en analogía a la Proposición 1.12, con la diferencia de que estos subespacios tienen dimensión igual a la de sus contrapartes afines menos 1. Además, es cierto un resultado análogo a la Proposición 1.9, con lo que la dimensión de V es igual a la del subespacio coordenado proyectivo de máxima dimensión contenido en $V(\langle LT(I(V)) \rangle)$. Esto nos permite establecer un algoritmo prácticamente igual al Algoritmo 1.30 para calcular la dimensión de una variedad proyectiva. También es posible definir la dimensión de una variedad proyectiva desde el punto de vista de Krull, como la dimensión de Krull del anillo cociente $\mathbb{K}[x_1, \dots, x_n]/I(V)$ menos 1.

Para dar un ejemplo de cálculo de dimensión de una variedad afín utilizaremos Singular con una función que calcula el polinomio proyectivo de Hilbert de un ideal homogéneo. Como estamos interesados en un polinomio afín de Hilbert, previamente veamos un resultado que relaciona ambos polinomios.

Dado un ideal I en $\mathbb{K}[x_1, \dots, x_n]$, definimos la homogeinización de I , $I^h \subset \mathbb{K}[x_1, \dots, x_n]$, como el conjunto de todos los polinomios homogéneos que se obtienen tomando un polinomio en I y multiplicando cada uno de sus términos por una potencia de x_0 para que sea homogéneo. El teorema 4 de la sección 8.4 de [4] nos dice que la homogeinización de los polinomios de una base de Groebner de I con respecto a un orden monomial graduado nos da una base de I^h . El siguiente resultado, (Teorema 12-(i) de la sección 9.3 de [4]), nos da la relación de la que hablábamos en el párrafo anterior:

Teorema 1.33 *Sea I un ideal en $\mathbb{K}[x_1, \dots, x_n]$, entonces, para $s \geq 0$,*

$$HF_I(s) = PHF_{I^h}(s).$$

Ejemplo 1.34 *Como ejemplo del estudio realizado en este texto, vamos a calcular la dimensión de la «Cúbica alabeada» («twisted cubic», en inglés), que es la variedad afín $V = V(x_2 - x_1^2, x_3 - x_1^3) \subset \mathbb{R}^3$. En el capítulo 1 de [4] se prueba que $I(V) = \langle x_2 - x_1^2, x_3 - x_1^3 \rangle$.*

Por un lado, calculemos la dimensión siguiendo el algoritmo 1.30. Tomamos el orden grevlex, y el cálculo de una base de Groebner de I nos da

$I = \langle x_2^2 - x_1x_3, x_1x_2 - x_3, x_1^2 - x_2 \rangle$. A continuación, nos quedamos con los términos dominantes de estos generadores para calcular $\langle LT(I) \rangle = \langle x_2^2, x_1x_2, x_1^2 \rangle$. A continuación calculamos la dimensión de $V(\langle LT(I) \rangle)$ con el Algoritmo 1.13. Con la terminología del algoritmo, $M_1 = \{2\}$, $M_2 = \{1, 2\}$, y $M_3 = \{1\}$. El conjunto L es $L = \{\{1, 2, 3\}, \{1, 2\}\}$, por lo que $\dim(V) = n - |\{1, 2\}| = 3 - 2 = 1$. Algoritmos similares a este son utilizados en diversos lenguajes de programación para calcular la dimensión [4]. En el Ejemplo 4.4 calculamos la dimensión de esta variedad con la función \dim de Singular.

Por otro lado, partiendo de la base de Groebner $I = \langle x_2^2 - x_1x_3, x_1x_2 - x_3, x_1^2 - x_2 \rangle$, la homogeinización de I es $I^h = \langle x_2^2 - x_1x_3, x_1x_2 - x_3x_0, x_1^2 - x_2x_0 \rangle$ y llamamos «Cúbica alabeada proyectiva» a la variedad $V(I^h) \subset \mathbb{R}^3$. Por el Teorema 1.33, el polinomio proyectivo de Hilbert de I^h es igual al polinomio de Hilbert de I . En el mismo Ejemplo 4.4 calcularemos que este polinomio es $3s + 1$ en Singular (también es sencillo calcularlo directamente a partir de una resolución graduada exacta), por lo que la variedad tiene dimensión 1, lo que nos sirve para comprobar el resultado del párrafo anterior. Además a partir de este polinomio vemos que la «Cúbica alabeada proyectiva» tiene dimensión 1 y grado 3 (esto último lo vemos en el siguiente párrafo).

Para terminar este capítulo, comentamos que a partir de la función de Hilbert de un ideal se puede extraer mucha información útil, además de la relacionada con la dimensión de una variedad. Resaltamos por ejemplo [6], que si tenemos un ideal homogéneo $I \subset \mathbb{K}[x_0, \dots, x_n]$, si su polinomio de Hilbert proyectivo PHP_I tiene grado d , entonces el coeficiente de su término dominante es $D/d!$, donde D es un entero positivo al que llamamos grado de la variedad proyectiva $V(I) \subset \mathbb{P}^n$, que es igual al número de puntos donde V interseca un subespacio lineal genérico de dimensión $n-d$ en \mathbb{P}^n ; y su género aritmético es $g = (-1)^d (PHP_I(0) - 1)$.

2. Estimación de la dimensión de un conjunto de puntos

Dado un conjunto de m datos C contenido en una variedad real afín o proyectiva V , en este capítulo trataremos de obtener una estimación para la dimensión de V basándonos únicamente en los datos, sin conocer las ecuaciones que definen V . Estas estimaciones serán de utilidad en el siguiente capítulo, dedicado a la determinación de los polinomios del ideal asociado a la variedad, $I(V)$, ya que con técnicas como las del capítulo 1 podremos calcular la dimensión de V a partir de $I(V)$, y descartar los candidatos a $I(V)$ cuya dimensión no coincida con la estimada. Comenzaremos estudiando el caso en que V es una variedad lineal, para el que utilizaremos el Análisis de Componentes Principales, y después presentaremos las herramientas propuestas en [2] para el caso general.

2.1. Caso lineal: Análisis de Componentes Principales

El Análisis de Componentes Principales (ACP, o PCA en inglés) [9] es un procedimiento ampliamente usado en el análisis de datos para la estimación y para la reducción de su dimensión. Partiendo de un conjunto de datos $C \subset \mathbb{R}^n$, el objetivo de esta técnica es encontrar el subespacio de dimensión $q < n$ que mejor ajuste a los datos y dar una medida de la «bondad de ese ajuste» en el sentido de la regresión lineal de mínimos cuadrados, que precisaremos en el Teorema 2.1. Antes de presentar este resultado, introduciremos la siguiente notación.

Llamamos \bar{X} a la matriz de tamaño $m \times n$ cuyas filas son los elementos de C , y llamamos X a la matriz que resulta de restar a todos los elementos de cada columna de \bar{X} la media de dicha columna. Interpretando cada dato de C como una muestra de un vector aleatorio de n variables, el paso de \bar{X} a X consiste en centrar los datos de cada variable para que tenga media 0. Siguiendo con esta interpretación, llamamos S a la matriz de covarianzas de los datos, es decir, la matriz de tamaño $n \times n$ cuyo elemento (i,j) -ésimo es la covarianza (muestral) de las variables i -ésima y j -ésima. Como hemos centrado las variables, $S = \frac{1}{m-1} X^t X$, y a partir de esta expresión es trivial ver que S es simétrica y semidefinida positiva, luego por el Teorema Espectral S diagonaliza en una base ortonormal $\{a_1, \dots, a_n\}$ y sus autovalores $\lambda_1 \geq \dots \geq \lambda_n$ son reales no negativos.

Para cada i , llamamos i -ésima componente principal de X (o de C , abusando de la notación) a un autovector cualquiera a_i de S , que sea ortonormal y asociado al i -ésimo autovalor. La definición de las componentes principales no las determina de forma única: por un lado si a_i es la i -ésima componente principal, el vector $-a_i$

también cumple la definición, pero cualquiera de las dos elecciones tendrá las mismas propiedades que exponremos a continuación; y por otro lado si hay dos autovalores que tengan el mismo valor (o con una diferencia entre ellos muy pequeña que pueda ser enmascarada por los errores de los procedimientos numéricos para calcular los autovectores), habrá más elecciones posibles para elegir las componentes principales (también con las mismas propiedades), lo que es un aspecto a tener en cuenta desde el punto de vista numérico y de interpretación de los resultados ([9], capítulo 1.30), que no entraremos a discutir en este texto.

Por otra parte, para cada i , la proyección de los datos en el subespacio de dimensión 1 generado por la componente principal i -ésima, Xa_i , tiene varianza (muestral) igual al i -ésimo autovalor de S , λ_i . A continuación presentamos el resultado geométrico sobre el ACP ([9], Propiedad G3), que será fundamental para los objetivos de este texto, cuya demostración es sencilla y se puede encontrar en la misma referencia.

Teorema 2.1 *Supongamos que tenemos un conjunto de puntos $C = \{x_1, x_2, \dots, x_m\}$ (que tomamos como vectores fila, siguiendo con el criterio de los párrafos anteriores) y una transformación de la forma $y_i = x_i B$, donde B es una matriz de tamaño $n \times q$ con columnas ortonormales; de tal manera que y_1, y_2, \dots, y_m son proyecciones de los puntos de C en un subespacio de dimensión q . Si medimos la «bondad del ajuste» de este subespacio de dimensión q a los puntos de C como la suma de los cuadrados de las distancias perpendiculares de los puntos al subespacio, esta medida se minimiza cuando B es la matriz cuya fila j -ésima es la j -ésima componente principal de C , es decir, cuando el subespacio está generado por las q primeras componentes principales de C .*

Desde un punto de vista equivalente al del resultado anterior [9], el resultado del ACP nos proporciona una serie de subespacios de dimensión 1, ..., q , ..., $n-1$, de tal manera que la proyección de los datos en ellos retiene la máxima variación posible de entre los subespacios de cada dimensión. El porcentaje de variación en los datos recogido por la proyección del subespacio definido por las q primeras componentes principales es $100 \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^n \lambda_i} \%$.

La estimación de V mediante PCA consiste en tomar V igual al subespacio generado por las q primeras componentes principales y aproximar que los autovalores asociados a las $n-q$ componentes restantes son 0 (es decir, son direcciones sin variación en los datos, y ortogonales a V). La dimensión estimada de V será por tanto igual a q .

A la hora de decidir cuántas componentes principales tomar como significativas existen múltiples criterios que se utilizan comúnmente, como despreciar las componentes principales cuya varianza esté por debajo de cierta tolerancia, o retener las

primeras q componentes principales de tal manera que la proyección en el subespacio que generan recoja un cierto porcentaje de la variación de los datos. La elección que tomamos es quedarnos con las q primeras componentes, tal que

$$q = \operatorname{argmax}_{1 \leq i \leq \min\{n,m\}-1} |\log_{10}(\lambda_{i+1}) - \log_{10}(\lambda_i)|,$$

si existe un índice i para el que esta diferencia en los logaritmos sea significativa, lo que según [2] mejora la estimación de la dimensión q de V con respecto a obtenerla estableciendo una tolerancia fija. La definición de q anterior no varía si se sustituyen las varianzas recogidas por cada componente principal λ_i por sus desviaciones estándar $\sqrt{\lambda_i}$ (lo que haremos en la práctica en el capítulo 4 porque algunos algoritmos proporcionan estas últimas como salida, en vez de las varianzas). Consideraremos que ninguna diferencia $|\log_{10}(\lambda_{i+1}) - \log_{10}(\lambda_i)|$ es significativa si la diferencia entre los logaritmos del primer y último autovalor, $|\log_{10}(\lambda_n) - \log_{10}(\lambda_1)|$, es menor que 1, y en ese caso concluiremos que no existe ninguna variedad lineal V que sea un buen modelo para los datos.

Por último, mencionamos que en aplicaciones prácticas en las que el objetivo fundamental del ACP es la reducción de la dimensión de un conjunto de datos, más que la estimación de la dimensión de un subespacio V que sea un modelo lineal para los datos sin apenas pérdida de información, es frecuente descartar componentes principales en las que la proyección de los datos tienen una variación significativa. De especial interés es el caso en el que se eligen únicamente las dos primeras componentes, lo que permite la representación gráfica en dos dimensiones de los datos que da cuenta del mayor porcentaje de variación posible y puede servir para realizar una exploración visual de los datos si este porcentaje es grande; y también el caso en el que se escoge únicamente la primera componente principal, muy frecuente en contextos de evaluación de indicadores, donde cada dato es una observación de n variables (o indicadores) en un individuo, y se define su indicador compuesto como la proyección de este dato en la primera componente principal, de tal manera que los indicadores compuestos son el mejor resumen unidimensional de los datos, y se pueden interpretar como una «evaluación global» de los individuos que permiten establecer un orden total (y un ranking) entre ellos.

Además, existe una derivación alternativa del ACP que consiste en definir X como la matriz de datos con las columnas estandarizadas (dividiendo cada columna por su desviación estándar) además de centradas, y definiendo las componentes principales como los autovectores de $\frac{1}{m-1} X^t X$, que en este caso es la matriz de correlación de los datos, es decir, la matriz de tamaño $n \times n$ cuyo elemento (i,j) -ésimo es la correlación (muestral) de las variables i -ésima y j -ésima; de tal manera que el ACP pasa a ser independiente de las unidades de medida de las variables, lo que es conveniente en

ciertas aplicaciones como se discute en [9].

Para terminar el caso lineal, presentaremos los métodos que utilizaremos en este texto para realizar el ACP. En el caso de disponer de datos completos (no tener ningún dato con alguna coordenada desconocida), calcularemos el ACP realizando la descomposición en valores singulares de la matriz de datos centrada X . Si M es una matriz de dimensión $m \times n$ de rango q , su descomposición en valores singulares es $X = ULA^t$, donde U y A son matrices con columnas ortonormales de tamaño $m \times m$ y $n \times n$, respectivamente; y L es una matriz diagonal (rectangular) de tamaño $m \times n$ cuyos elementos se denominan valores singulares, y los valores singulares distintos de cero son los q primeros. Si realizamos la descomposición en valores singulares de una matriz de datos centrados, obtenemos los resultados del ACP: su i -ésimo valor singular es igual a $\sqrt{m-1}\sqrt{\lambda_i}$, (es decir, proporcional a la desviación estándar de la proyección de los datos en la i -ésima componente principal de X), y la i -ésima columna de A es la i -ésima componente principal de X . Esta relación entre el ACP y la descomposición en valores singulares nos permitirá calcularlo con algoritmos eficientes de descomposición en valores singulares como el algoritmo estándar de R, que usaremos en el ejemplo 4.1.

Por último, además de encontrar la mejor aproximación a una variedad lineal en presencia de «ruido», el ACP es capaz de manejar la presencia de datos con alguna variable desconocida (es decir, puntos de los que no conocemos alguna coordenada) en una cantidad moderada, utilizando el algoritmo NIPALS [7] (Nonlinear Iterative Partial Least Squares), lo cual es muy útil en aplicaciones prácticas, donde es frecuente el estudio de conjuntos de datos incompletos. El método NIPALS consiste en realizar la diagonalización de la matriz de covarianzas de X realizando un Método de la Potencia [14] sin construir la matriz de covarianzas explícitamente para no multiplicar la cantidad de datos faltantes.

El método de la potencia consiste en partir de un vector aleatorio y multiplicarlo sucesivamente por una matriz hasta que el producto converge a un autovector correspondiente al autovalor de mayor módulo de ésta. En nuestro caso, se parte de un vector aleatorio y se multiplica sucesivamente por X y X^t , saltando los valores desconocidos en el cálculo de cada componente principal (es decir, tomándolos en este paso iguales a 0, que es la media de todas las variables). Una vez que el producto converge a la primera componente principal (autovector correspondiente al autovalor más grande de la matriz de covarianza), se elimina de la matriz X su proyección en la primera componente principal, y se aplica el método sucesivamente para obtener (por orden) las componentes principales y la descomposición de la matriz de datos como suma de sus proyecciones en cada componente principal [7], a partir de la cual se pueden estimar los datos faltantes. En el Ejemplo 4.1.3 estimamos una variedad

lineal aplicando este método.

2.2. Caso general

2.2.1. Diagramas de dimensión

En el caso general de variedades algebraicas, estimamos la dimensión de V utilizando algunas de las técnicas que se proponen en el Capítulo 3 de [2] y presentamos brevemente a continuación. Los estimadores de dimensión que se utilizan en este capítulo computan la dimensión intrínseca del conjunto de datos C , la cual es un número real positivo que aproxima la dimensión de Hausdorff de la variedad V . La dimensión de Hausdorff de V es una dimensión fractal real que es una generalización métrica de la dimensión de un espacio topológico y mide la dimensión local de un espacio a partir de las distancias entre puntos cercanos.

En esta sección asumimos que V es una variedad real afín o proyectiva tal que el conjunto de puntos reales sea denso en la topología de Zariski (que es la topología cuyos conjuntos cerrados son exactamente las variedades) en cada componente irreducible de V . Si V es irreducible, su conjunto de puntos singulares $\text{Sing}(V)$ es una subvariedad propia, y por tanto tiene medida cero. Además, $V - \text{Sing}(V)$ es una variedad regular, y cada una de sus componentes conexas es una variedad regular de dimensión igual a la dimensión de V . Bajo estas hipótesis naturales, la dimensión de Hausdorff de V coincide con la dimensión (algebraica) de V [2], y por tanto la dimensión intrínseca de $C \subset V$ será una aproximación a esta última.

Como la dimensión intrínseca es un método que aproxima la dimensión local de un espacio, necesitamos definir una distancia en el espacio métrico finito C y un criterio para determinar cuál es el entorno local de un punto. La distancia que utilizaremos en el caso afín es la euclídea escalada:

$$\text{dist}_{\mathbb{R}^n}(u, v) := \frac{\|u - v\|}{\max_{x, y \in C} \|x - y\|}$$

Y la distancia que utilizaremos en el caso proyectivo es la distancia escalada de Fubini-Study (FS):

$$\text{dist}_{\mathbb{P}^{n-1}}(u, v) := \frac{\text{dist}_{FS}(u, v)}{\max_{x, y \in C} \text{dist}_{FS}(u, v)},$$

donde la distancia de Fubini-Study es el ángulo entre las rectas generadas por vectores representantes de u y v en \mathbb{R}^n :

$$\text{dist}_{FS}(u, v) := \arccos \frac{|\langle u - v \rangle|}{\|x\| \|y\|}.$$

Fijando un valor de ϵ , realizamos una partición de C en l clústers $C_\epsilon^1, \dots, C_\epsilon^l$; y la localidad se define en términos de pertenencia a un clúster, de manera que los puntos en el entorno local de cada dato son los que están en su mismo clúster. Los métodos de estimación se aplican en cada clúster por separado, de tal manera que el parámetro ϵ determina el rango de información utilizada para computar la dimensión intrínseca, variando desde $\epsilon = 0$, donde cada punto se considera de manera independiente, hasta $\epsilon = 1$, donde se considera la muestra entera.

Muchos de los métodos de dimensión intrínseca que utilizaremos son consistentes, es decir, es posible computar un valor del parámetro ϵ a partir del conjunto de datos C , de tal manera que la dimensión intrínseca converja a la dimensión (algebraica) de V cuando el número de puntos m tiende a infinito si el muestreo de los datos es suficientemente denso. Sin embargo, en el planteamiento de nuestro problema el número de puntos m es fijo, y no hay hipótesis sobre su distribución en V : nuestro conjunto de datos C casi nunca satisfará las hipótesis necesarias como para obtener una buena aproximación a la dimensión de V aplicando los métodos en un valor apropiado de ϵ , y obtener más puntos y/o elegir su distribución no es una opción.

El método alternativo planteado por [2] para computar un valor de dimensión a partir de un conjunto de datos con las hipótesis de nuestro problema es crear diagramas de dimensión.

Definición 2.2 *Sea $\dim(C, \epsilon)$ la dimensión del conjunto de datos C obtenida por un estimador con parámetro ϵ . El diagrama de dimensión de C es el grafo de la función $(0, 1] \rightarrow \mathbb{R}_{\geq 0}$, $\epsilon \rightarrow \dim(C, \epsilon)$.*

La estrategia que utilizaremos para estimar la dimensión de V es graficar los diagramas de dimensión de varios estimadores, y elegir el valor de dimensión que sea más cercano al dado por la mayoría de indicadores para un rango significativo de valores del parámetro ϵ . En el Capítulo 4 podemos ver la aplicación de esta estrategia para estimar la dimensión de diferentes variedades.

Para terminar este capítulo, presentaremos brevemente los estimadores que utilizaremos en los ejemplos del capítulo 4 y que podemos encontrar en el Capítulo 3.2 de [2]. Mostraremos únicamente la versión afín de los indicadores, que es la que utilizaremos en los ejemplos, aunque todos los indicadores tienen una versión para variedades proyectivas con pocas diferencias.

2.2.2. Estimadores de dimensión

Definición 2.3 (ACP No Lineal) *El estimador de ACP No Lineal (Nonlinear PCA, o NPCA; en inglés) se basa en la aproximación de que $V\text{-Sing}(V)$ es aproximadamente igual localmente a su espacio tangente de Zariski en un punto, que es*

una variedad lineal de dimensión $\dim(V)$. Consiste en realizar una partición de C en clústers, que en nuestro caso los definimos con el método de «enlace único», que consiste en que los clústers son las componentes conexas en el grafo con conjunto de vértices C cuyas aristas son los pares de puntos cuya distancia es menor o igual que ϵ ; realizar en cada clúster la estimación de la dimensión lineal con ACP y hacer la media ponderada de las estimaciones en cada clúster, con pesos proporcionales a la cantidad de datos en cada uno de ellos:

$$\dim_{NPCA}(C, \epsilon) := \frac{1}{\sum_{i=1}^m |C_\epsilon^i|} \sum_{i=1}^l |C_\epsilon^i| \dim_{PCA}(C_\epsilon^i).$$

Definición 2.4 (Dimensión de Conteo de Cajas) Si tomamos un paralelepípedo $[a^-, a^+]$ en \mathbb{R}^n , con vértice inferior $a^- = \min_{(a_1, \dots, a_n) \in C} (a_1, \dots, a_n)$ y vértice superior $a^+ = \max_{(a_1, \dots, a_n) \in C} (a_1, \dots, a_n)$, donde «min» y «max» significan el mínimo y máximo coordenada a coordenada, respectivamente; dividimos cada intervalo $\mathbb{R} [a_i^-, a_i^+]$ en $R(\epsilon)^d$ intervalos de igual tamaño (que depende de ϵ) y definimos como «cajas» los conjuntos producto de estos intervalos esperamos que localmente un objeto de dimensión d tenga puntos en $R(\epsilon)^d$ cajas.

Basándonos en esta idea, definimos el estimador de Conteo de Cajas (o BoxCounting, del inglés «Box Counting») como:

$$\dim_{BoxCounting}(C, \epsilon) := \frac{\log(\mu)}{\log(R(\epsilon))},$$

donde μ es el número de cajas que contienen algún punto de C .

Definición 2.5 (Dimensión de Correlación) La idea que subyace detrás de este estimador es la siguiente. Si C está uniformemente distribuida en la bola unidad, para puntos $a, b \in C$, $\text{Prob}(\text{dist}(a, b) < \epsilon) = \epsilon^{\dim(V)}$. Denotando

$$Z(\epsilon) := \frac{1}{\binom{m}{2}} \text{sum}_{1 \leq i < j \leq m} \mathbf{1}(\text{dist}_{\mathbb{R}^n}(a_i, b_j) < \epsilon),$$

donde $\mathbf{1}$ es la función indicador, esperamos que la distribución empírica de $Z(\epsilon)$ sea aproximadamente $\epsilon^{\dim(V)}$. Por lo tanto se podría estimar la Dimensión de Correlación, (que llamamos «CorrSum») como $\frac{\log(Z(\epsilon))}{\log(\epsilon)}$, aunque un estimador más práctico [2] es:

$$\dim_{CorrSum}(C, \epsilon) := \frac{\log(Z(\epsilon)) - \log(Z(\epsilon + h))}{\log(\epsilon) - \log(\epsilon + h)},$$

donde h es la menor distancia entre los valores del parámetro ϵ para los que computaremos la dimensión al hacer el diagrama.

Definición 2.6 (Dimensión del Estimador de Máxima Verosimilitud) *Este estimador está basado en el siguiente estimador introducido por Levina y Bickel para estimar la dimensión de una variedad en un punto $a^* \in C$:*

$$\dim_{LB}(C, \epsilon, a^*) := \left(\frac{1}{k} \sum_{i=1}^k \log \frac{\epsilon}{T_i(a^*)} \right)^{-1},$$

donde $k = k(\epsilon)$ es el número de puntos de C que están a distancia menor o igual que ϵ de a^* , que por hipótesis sigue un proceso de Poisson en el entorno de puntos $\{a \in C : \text{dist}(a, a^*) \leq \epsilon\}$, en el que a tiene una distribución uniforme; y $T_i(a^*)$, que es la distancia $\text{dist}(a^i, a^*)$, donde a^i es el i -ésimo punto de C más cercano a a^* . Esta fórmula se obtiene resolviendo las ecuaciones de verosimilitud para este proceso de Poisson.

Para obtener el estimador de máxima verosimilitud de la dimensión de V (Maximum Likelihood Estimator, o MLE; en inglés), realizamos una media ponderada de la dimensión de Levina y Bickel en cada punto, asignando a cada punto un peso proporcional a la cantidad de puntos en

$$C_\epsilon^i := \{a \in C : \text{dist}(a, a^i) \leq \epsilon\}:$$

$$\dim_{MLE}(C, \epsilon) := \frac{1}{\sum_{i=1}^m |C_\epsilon^i|} \sum_{i=1}^m |C_\epsilon^i| \dim_{LB}(C_\epsilon^i, \epsilon, a^i).$$

Definición 2.7 (Dimensión de Análisis de Varianza) *Este estimador se debe a Diaz, Quiroz y Velasco, y está basado, al igual que el anterior, en estimaciones de la dimensión en cada punto $a^* \in C$. Denotamos por a^1, \dots, a^k a los puntos de C cuya distancia respecto a a^* es menor o igual que ϵ . Llamamos θ_{ij} al ángulo entre $a^i - a^*$ y $a^j - a^*$, cuya covarianza es:*

$$S = \frac{1}{\binom{k}{2}} \sum_{1 \leq i < j \leq k} \left(\theta_{ij} - \frac{\pi}{2} \right)^2.$$

Como se establece en [2], para ϵ pequeño y C una muestra de una variedad regular de dimensión d , los ángulos θ_{ij} están distribuidos aproximadamente como el ángulo entre dos puntos elegidos aleatoriamente en una esfera de dimensión $d-1$, cuya varianza es β_d , donde, para s un número natural:

$$\beta_{2s-1} = \frac{\pi^2}{4} - 2 \sum_{j=0}^s \frac{1}{(2j+1)^2}, \quad y$$

$$\beta_{2s} = \frac{\pi^2}{12} - 2 \sum_{j=0}^s \frac{1}{(2j)^2}.$$

El estimador de análisis de varianza, (Analysis of Variance, o ANOVA; en inglés), en un punto es el valor de d que minimiza la diferencia entre β_d y S :

$$dim_{ANOVA(puntual)}(C_\epsilon^i, \epsilon, a^*) := \operatorname{argmin}_d |\beta_d - S|.$$

y el estimador dim_{ANOVA} de la dimensión de V se obtiene promediando las dimensiones obtenidas en cada punto.

3. Estimación de las ecuaciones que definen una variedad.

En este capítulo abordamos el problema de, dado un conjunto de m datos C contenido en una variedad real afín o proyectiva V , encontrar las ecuaciones que definen dicha variedad, es decir, generadores de $I(V) \subset \mathbb{R}[x_1, \dots, x_n]$. Para hallar dichas ecuaciones, presentaremos métodos numéricos que utilizaremos junto con los estimadores de dimensión del capítulo anterior. Estos métodos pueden lidiar con un pequeño «ruido» en los datos, es decir, si los datos no pertenecen exactamente a V , obtendremos las ecuaciones de una variedad que ajuste los datos, en el sentido de que casi se anulen en C .

Para empezar, en el caso de una variedad con una transformación sencilla a un subespacio vectorial, el Análisis de Componentes Principales, que utilizamos en el capítulo anterior para estimar la dimensión d de la variedad, proporciona sus ecuaciones paramétricas: la variedad es el subespacio generado por las d primeras componentes principales. El paso de ecuaciones paramétricas a implícitas nos da la solución del problema en este caso.

En el caso de variedades generales, seguiremos la estrategia de [2] consistente en estudiar ciertos subespacios del ideal de todos los polinomios que se anulan en C , $I(C)$, que contiene al ideal $I(V)$ que estamos buscando. Si tomamos un subconjunto finito y linealmente independiente M de polinomios de $\mathbb{R}[x_1, \dots, x_n]$, y llamamos R_M al espacio vectorial con base M , podemos determinar el subespacio $I(C) \cap R_M$ utilizando matrices multivariantes de Vandermonde.

Definición 3.1 *Sea M un subconjunto finito ordenado de $\mathbb{R}[x_1, \dots, x_n]$, y C un conjunto de m puntos en \mathbb{R}^n (el caso proyectivo se define de manera análoga). Llamamos matriz multivariante de Vandermonde de M evaluada en C a la matriz $U_M(C)$ de tamaño $m \times |M|$ cuyo elemento en la fila i -ésima y columna j -ésima consiste en la evaluación del polinomio j -ésimo de M en el punto i -ésimo de C . En particular, denotaremos la matriz de Vandermonde como $U_{\leq d}(C)$ si M es el conjunto de monomios de grado menor o igual que d , y $U_d(C)$ si M es el conjunto de monomios de grado d .*

Veamos un ejemplo muy sencillo de esta definición.

Ejemplo 3.2 *Sea $C = \{(-1, 0), (1, 0), (0, -1)\}$, y $M = \{1, x, y, x^2, xy, y^2\}$. Entonces,*

$$U_{\leq 2}(C) = U_M(C) = \begin{pmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 1 \end{pmatrix}$$

Trivialmente, el núcleo de la matriz de Vandermonde $U_M(C)$ es igual al subespacio $I(C) \cap R_M$. El cálculo de el núcleo de este tipo de matrices es un problema de álgebra lineal numérica en absoluto trivial para el que existen varios algoritmos, y de la elección de un algoritmo adecuado depende parte del éxito que podemos tener en la estimación de variedades algebraicas. Para empezar, las matrices de Vandermonde pueden estar mal acondicionadas, por lo que la solución de la computación de su núcleo puede verse muy distorsionada por la presencia de pequeños errores en los datos de partida. Por ejemplo en el caso en el que los puntos estén en la recta real (en dimensión 1), las matrices de Vandermonde $U_{\leq d}(C)$ son matrices de Vandermonde univariantes, y es bien conocido [2] que este tipo de matrices están muy mal acondicionadas. En general, para aliviar este problema, priorizaremos en la elección de algoritmo el hecho de que sea estable y preciso.

Otro aspecto a tener en cuenta para la elección del algoritmo es si es disperso, es decir, si en general éste proporciona polinomios dispersos (es decir, con varios coeficientes nulos). Esta propiedad resulta de mucha utilidad en algunos casos, pues los polinomios dispersos en general son más fáciles de interpretar, como veremos en el ejemplo 4.2.

En general, los métodos más dispersos resultan ser menos precisos, y viceversa. Por lo tanto, en [2] se proponen tres métodos clásicos de álgebra lineal numérica, de manera que tenemos diferentes elecciones posibles, en función de si queremos priorizar la precisión o la estabilidad en el cálculo. El método más disperso de los tres consiste en computar la forma de matriz escalonada reducida (RREF) de $U_M(C)$, pero no utilizaremos este método en este texto, ya que no tiene garantías de precisión, y algunas pruebas realizadas proporcionan muy malos resultados en presencia de ruido. En el extremo opuesto, tenemos la descomposición en valores singulares, utilizada en el capítulo anterior, que es mucho más precisa y rápida que la anterior, por lo que será el algoritmo que utilizaremos en la mayoría de casos en el capítulo 4, aunque el resultado de la computación es una base de polinomios ortonormal y por tanto es un método muy poco disperso. Si la descomposición en valores singulares de una matriz de Vandermonde es $U_M(C) = ULA^t$, donde U y A son matrices con columnas ortonormales, L es la matriz diagonal de valores singulares y aproximamos a 0 los k últimos valores singulares, las k últimas filas de A son la base de $\ker(U_M(C))$ que obtenemos [2].

El tercer método, la descomposición QR, tiene propiedades intermedias a los anteriores, es algo más lento y menos preciso que la descomposición en valores singulares, pero es algo más disperso que ésta. Dada una matriz U , su descomposición QR es la expresión de $U = QR$ como producto de una matriz ortogonal Q y una matriz triangular R , y en el capítulo 5 de [2] se proporciona un algoritmo para calcular esta

descomposición y una base de $\ker(U_M(C))$ a partir de esta. Todos los métodos requieren la especificación de una tolerancia para determinar el rango (numérico) de la matriz de Vandermonde. Como veremos en el capítulo 4, los programas que ejecutan estos algoritmos en Julia tienen un valor de tolerancia fijo por defecto, pero en situaciones en las que hay ruido en los datos elegiremos el valor de la tolerancia de tal manera que se aproximen a 0 los valores singulares a partir del mayor salto en la diferencia entre el logaritmo de un valor singular y el logaritmo del siguiente (si esta diferencia es significativa), con el mismo criterio utilizado en el Capítulo 1.27 para estimar la dimensión de un subespacio lineal por ACP.

Por último, comentamos que según la discusión del final del capítulo 5 de [2], los autores del artículo son optimistas respecto a la posibilidad de desarrollar en el futuro un método numéricamente estable y preciso para computar el núcleo de matrices multivariantes de Vandermonde, lo que supondría un gran avance en el objetivo de este texto.

Una vez que hemos visto cómo podemos calcular los subespacios de la forma $I(C) \cap R_M$, reflexionemos sobre la información que se puede obtener a partir ellos sobre el ideal $I(V)$.

Por un lado, se tiene $I(V) \cap R_M \subset I(C) \cap R_M$. En consecuencia, si elegimos M de tal manera que $I(V)$ esté generado por su subespacio $I(V) \cap R_M$, el subespacio $I(C) \cap R_M$ que obtendremos mediante computación contendrá un sistema de generadores del ideal que estamos buscando. El Teorema de la base de Hilbert nos asegura que siempre existen conjuntos M con esta propiedad.

Por otra parte, dado que $C \subsetneq V$, la contención $I(V) \cap R_M \subset I(C) \cap R_M$ no tiene por qué ser una igualdad, aunque de acuerdo con [2], en muchos casos, esta igualdad se cumple si M está formado por polinomios de grado bajo. Además, la inclusión de polinomios de grado alto en M puede llevar a la obtención de ecuaciones que supongan un modelo que sobreajuste los datos y no generalice correctamente.

Desde el punto de vista estadístico, el conjunto de datos de partida es una muestra de un vector aleatorio, y la variedad algebraica que queremos obtener es un modelo para el vector aleatorio, es decir, las ecuaciones del modelo deben generalizar: deben cumplirse para muestras de ese vector aleatorio que no hayan sido utilizadas para ajustar el modelo (con un error similar al que tienen en el conjunto de datos que sí ha sido utilizado). Para tratar el problema de la generalización, una aproximación frecuente en la literatura de selección y evaluación de modelos [3] es dividir el conjunto de datos de partida en tres subconjuntos, llamados de entrenamiento, validación y test, y que suelen constar aproximadamente del 60, 20 y 20 % de datos, respectivamente. El proceso de selección del modelo comienza fijando sus hiperparámetros (por ejemplo, en nuestro caso, en el caso afín, el conjunto de polinomios M

asociado a la matriz de Vandermonde cuyo núcleo vamos a calcular), obteniendo el respectivo modelo con el conjunto de entrenamiento, y evaluando qué elección de los hiperparámetros generaliza mejor con el conjunto de validación. Una vez elegidos los hiperparámetros, opcionalmente se puede obtener el modelo definitivo a partir de los datos de entrenamiento y validación. Por último, el modelo se evalúa con el conjunto de test.

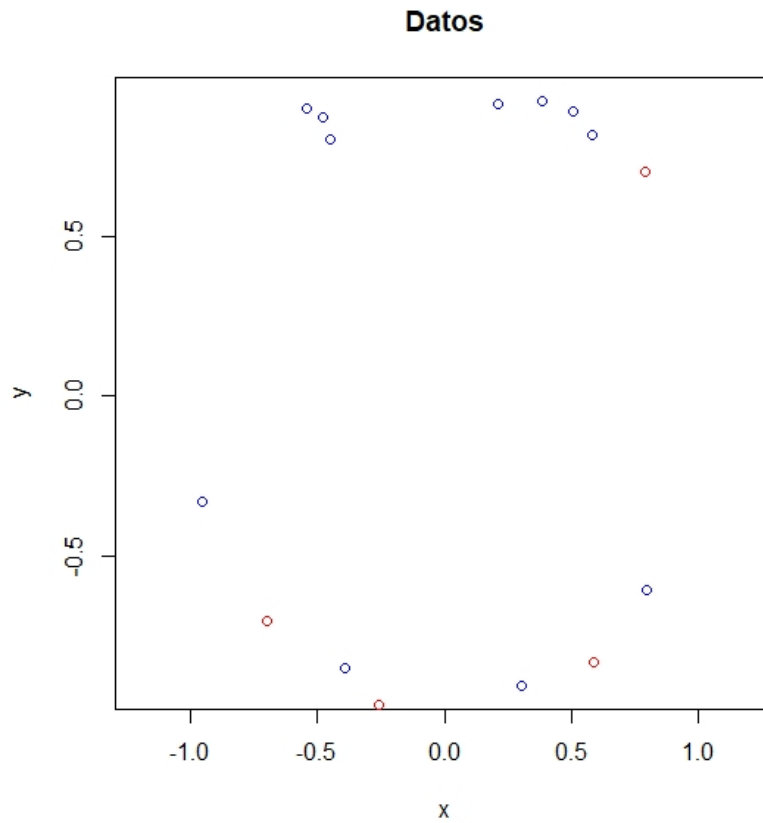
Según [3], intuitivamente, al realizar diferentes elecciones de los hiperparámetros aumentando sucesivamente la complejidad del modelo, el error de los modelos disminuye en los conjuntos de entrenamiento y validación (siendo generalmente un poco superior en este último), hasta un cierto nivel de complejidad a partir del cual, aumentar la complejidad de los modelos disminuye el error en los datos de entrenamiento, pero lo aumenta en los de validación: esta es la situación de sobreajuste. Si bien el objetivo de este texto no es el estudio de las propiedades de generalización de las variedades estimadas, presentamos un ejemplo sencillo a continuación de cómo incluir polinomios de grado alto en el conjunto M puede llevar a situaciones de sobreajuste.

Ejemplo 3.3 *Tomamos un conjunto C de 15 puntos en \mathbb{R}^2 , cada uno obtenido como suma de un punto aleatorio de la circunferencia $x^2 + y^2 = 1$ y un pequeño error. Concretamente, tomamos puntos de la forma:*

$$(x, y) = (\cos(t) + e1, \sin(t) + e2)$$

donde t sigue una distribución uniforme en $[0, 2\pi]$, y $e1$ y $e2$ son los términos de error, de distribución uniforme en $[-0.05, 0.05]$.

En la siguiente gráfica podemos ver el conjunto de puntos, que hemos dividido aleatoriamente en un subconjunto E de 11 puntos que usaremos para obtener ecuaciones (y representamos en color azul), y otro de 4 puntos, que nos servirá como conjunto de test.

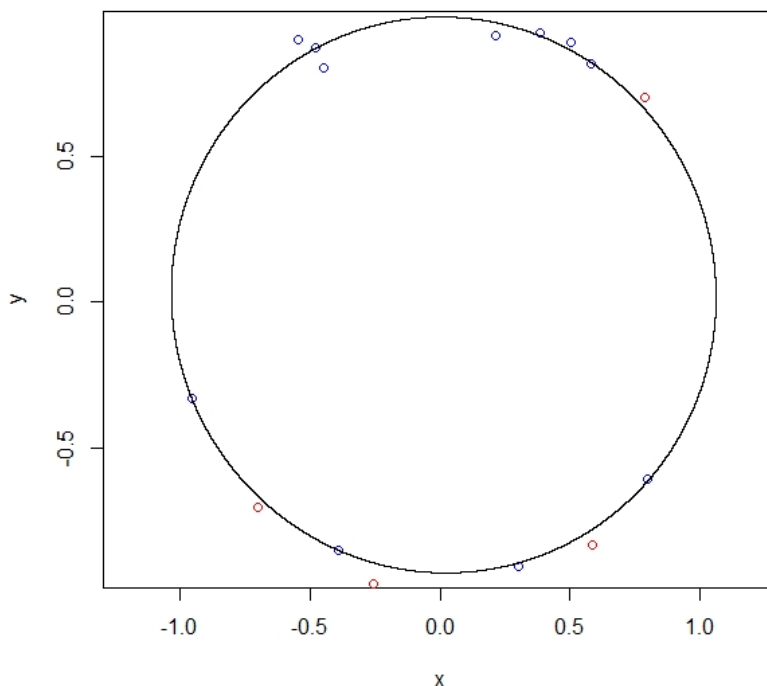


En el ejemplo 4.2 tratamos el problema de estimación de una variedad algebraica a partir de un conjunto de datos generado de manera similar, detallando la computación de las ecuaciones en Julia. En este caso, utilizando una computación análoga con descomposición en valores singulares, calculamos numéricamente una aproximación al ker de la matriz de Vandermonde $U_{\leq 2}(E)$, y obtenemos como única ecuación la elipse:

$$f = 0,5198x^2 + 0,0106xy + 0,6340y^2 - 0,0169x - 0,0286y - 0,5715 = 0,$$

que podemos ver en la siguiente gráfica.

Modelo con buena generalización



A consecuencia de la aproximación numérica, la variedad algebraica $V(f)$ obtenida no contiene exactamente a los puntos, sin embargo, la media del valor absoluto de la evaluación de la ecuación f en los puntos del conjunto de entrenamiento es 0.02521, y en los puntos del conjunto de test es 0.0516, del mismo orden de magnitud; por lo que se puede considerar que el modelo generaliza bien y es adecuado, lo cual era previsible sabiendo que los puntos pertenecen a una circunferencia de ecuación muy similar a la de la elipse obtenida salvo por un pequeño error.

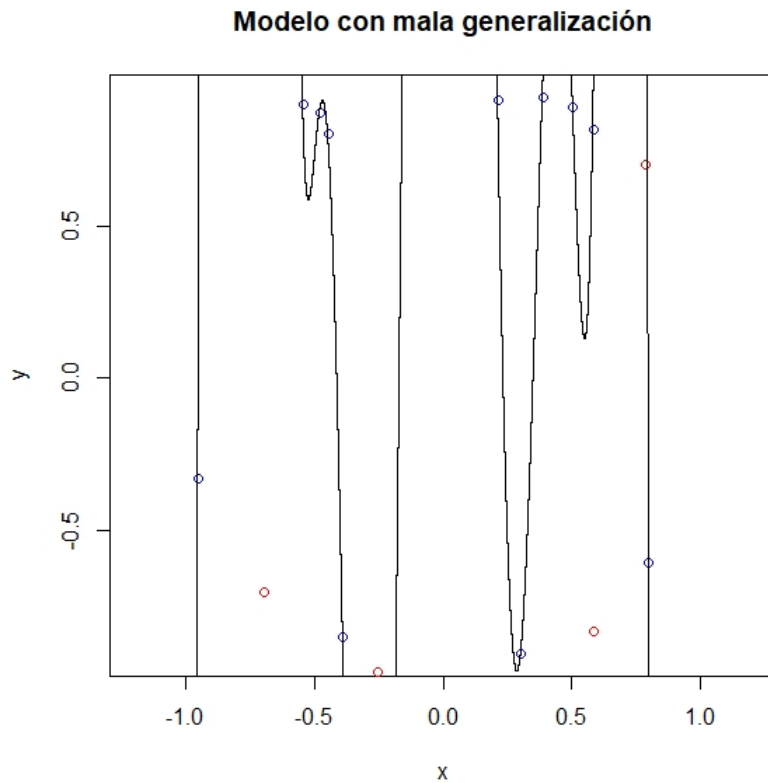
Sin embargo, en un caso de aplicación real en el que no disponemos de esta información sobre el mecanismo de generación de los puntos, podríamos realizar otra elección diferente del conjunto M de polinomios para la matriz de Vandermonde, como por ejemplo $U_{\leq 10}(E)$, lo que nos llevaría a la obtención de, al menos, una ecuación que ajusta a los datos de entrenamiento de manera perfecta: $g = y - p(x) = 0$, donde

$$p(x) = 11,22573 + 12,68781x - 458,403x^2 - 83,51741x^3 + 5149,59x^4 - 67,42738x^5 - 23432,37x^6 + 1600,576x^7 + 45369,99x^8 - 3152,173x^9 - 29398,87x^{10},$$

es el polinomio de interpolación de Lagrange, que tiene grado 10, en los 11 puntos del conjunto de entrenamiento. Dados n puntos $\{(x_i, y_i)\}_{i=1}^n \subset \mathbb{R}^2$ con la primera coordenada distinta en todos ellos, llamamos polinomio de interpolación de Lagrange

al único polinomio existente [13] de grado menor o igual que $n-1$ tal que $p(x_i) = y_i$ para todo i .

En este ejemplo, la exploración visual de los datos nos lleva a pensar que la relación entre las variables x e y no es explícita, pero incluso en los casos en los que sí existe una relación explícita de la variable y como función de x , La interpolación polinómica por polinomios de grado alto presenta grandes oscilaciones en los extremos del intervalo $[\min(x_i), \max(x_i)]$ en muchos casos, lo que hace que el problema esté mal acondicionado, es decir, pequeños cambios en los datos producen grandes cambios en el interpolante. Este problema, llamado Fenómeno de Runge puede ser evitado tomando puntos cuya distribución en el intervalo $[\min(x_i), \max(x_i)]$ tenga ciertas propiedades [13], pero en el planteamiento de nuestro problema el conjunto de datos es fijo, por lo que la interpolación polinómica por polinomios de grado alto será un modelo inadecuado en la gran mayoría de casos. En la siguiente figura podemos ver el modelo $V(g)$ para los datos de este ejemplo.



A pesar de que el ajuste a los datos de entrenamiento es perfecto, ya que $g = 0$ en ellos, el modelo no generaliza: la media del valor absoluto de la evaluación de g en los puntos del conjunto de test es $24.2722 \gg 0.0516$. Este sencillo ejemplo ilustra por un lado la necesidad de validar la capacidad de generalización de un modelo y por otra

parte desaconseja la inclusión en M de polinomios de grado alto, especialmente si el tamaño m del conjunto de datos es pequeño.

Además del grado de los polinomios, otro aspecto a considerar para la elección del conjunto M es su tamaño. Claramente, la elección de un gran número de polinomios en M puede favorecer que obtengamos ecuaciones en $I(C) \cap R_M$ que no están en $I(V) \cap R_M$. De hecho, se tiene el siguiente resultado [2].

Proposición 3.4 *Si $|M| > m + \dim(I(V) \cap R_M)$, entonces $I(V) \cap R_M \subsetneq I(C) \cap R_M$.*

Demostración: El hecho de fijar un orden en los polinomios de M , nos permite identificar R_M con el espacio vectorial $\mathbb{R}^{|M|}$. Por lo tanto, podemos ver la matriz de Vandermonde $U_M(C)$ como una matriz asociada a una aplicación lineal de $\mathbb{R}^{|M|}$ a \mathbb{R}^m . Por lo tanto

$$M = \dim(\ker(U_M(C))) + \dim(\text{im}(U_M(C)))$$

La dimensión de la imagen debe ser menor o igual a m , que es la dimensión del espacio de llegada. Si tuviéramos $I(V) \cap R_M = I(C) \cap R_M$, entonces

$$M - \dim(\ker(U_M(C))) = M - \dim(I(V) \cap R_M) = \dim(\text{im}(U_M(C))) \leq m,$$

lo que nos da un resultado equivalente al del enunciado del lema. \square

Dado que en el planteamiento de nuestro problema la cantidad de puntos m es fija, este resultado nos da una cota (de valor desconocido, ya que desconocemos V) para la elección del número de elementos de M si queremos que $I(V) \cap R_M = I(C) \cap R_M$, que es menor cuanto menor sea el número de puntos m para una misma variedad. Este resultado nos indica que deberíamos evitar en la medida de lo posible elegir M con un número muy grande de polinomios para evitar la obtención de ecuaciones no deseadas, especialmente si el número de puntos es pequeño.

Teniendo en cuenta las consideraciones anteriores, en el caso de no disponer de información a priori sobre la variedad V que nos dé una pista sobre qué elecciones de M pueden ser adecuadas, proponemos el siguiente procedimiento para la búsqueda de ecuaciones de una variedad algebraica V , que debe tomarse como una sugerencia para enfocar el problema y no como un algoritmo cerrado.

Comenzamos eligiendo aleatoriamente un subconjunto $T \subset C$ de test que contenga aproximadamente el 20 % de los datos del conjunto de datos total C . A continuación, elegimos un método de validación, por ejemplo la validación cruzada de k iteraciones, que consiste en dividir el 80 % de datos restante aleatoriamente en k grupos, y realizar k pasos de entrenamiento de modelo y validación con los mismos hiperparámetros, eligiendo en cada paso uno de los k grupos como conjunto de validación y los $k-1$

grupos restantes como conjunto de entrenamiento; y tomar como resultado de validación de la elección de esos hiperparámetros la media de los resultados de validación en los k pasos. Por ejemplo podríamos utilizar este método con $k = 4$, de manera que tendríamos el 60 % de los datos de C de entrenamiento y el 20 % de validación en cada paso.

Luego, en el caso de que V sea una variedad afín, empezamos la búsqueda de un hiperparámetro adecuado con un conjunto de hiperparámetros A vacío, y fijando el grado $d=1$ y la elección del hiperparámetro M igual al conjunto de monomios de grado menor o igual que d , que se corresponde con la matriz de Vandermonde generalizada $U_{\leq d}(E)$, con E un conjunto de entrenamiento. Realizamos el proceso de validación cruzada de $k=4$ iteraciones, donde variamos el conjunto de entrenamiento E en cada una de las 4 iteraciones, calculamos una base del subespacio $\ker(U_{\leq d}(E))$, y calculamos la dimensión del ideal generado por esta base. Si en alguna de las 4 iteraciones esta dimensión no coincide con la dimensión de V indicada por los estimadores de dimensión del capítulo anterior, descartamos el hiperparámetro M ; y en caso contrario, añadimos M al conjunto de hiperparámetros A , y calculamos su valor de validación cruzada. Además, si a continuación encontramos algún conjunto M' formado por polinomios de grado menor o igual que d tal que el ideal generado por el subespacio $\ker(U_{M'}(E)) \subset \ker(U_{\leq d}(E))$ tenga dimensión igual a la de V para los cuatro conjuntos de entrenamiento E de la validación cruzada anterior, añadimos también M' al conjunto A y calculamos su valor de validación cruzada.

Después, se itera este proceso, aumentando d en una unidad en cada paso. Terminamos el proceso de búsqueda si tras varios pasos aumentando el valor de d los nuevos hiperparámetros nos dan peores resultados en el conjunto de validación que algún hiperparámetro obtenido en un paso anterior; o si llegamos a un hiperparámetro cuya capacidad de generalización consideramos adecuada; o bien, si llegamos a un valor alto de d sin encontrar en ningún paso un modelo de dimensión correcta que generalice, en cuyo caso terminamos la búsqueda con el resultado de que el método no encuentra ninguna variedad generada por polinomios de grado menor o igual que ese valor máximo. Por último, tomamos como hiperparámetro óptimo M el que tenga mejor resultado de validación cruzada de entre los elementos del conjunto A , calculamos el modelo definitivo con ese conjunto M y con todos los datos excepto los del conjunto de test y lo evaluamos en el conjunto de test. En el primer paso (el caso lineal, $d = 1$) es posible reemplazar los estimadores de dimensión del caso general y las matrices de Vandermonde por el ACP.

En el caso proyectivo, el proceso es similar, con la diferencia de que buscaremos generadores homogéneos de grado d para $I(V)$, y por tanto, computaremos el núcleo de $U_d(E)$ en cada paso. Además, $I(V)$ puede estar generado por polinomios homogéneos

de diferente grado, por lo que en este caso «guardaremos» los polinomios que vayamos obteniendo en cada paso, incluso si no generan un ideal de la dimensión adecuada por sí solos. En este caso, los hiperparámetros son sucesiones $(M_1, M_2, \dots, M_d, \dots)$, donde M_d es un conjunto de polinomios homogéneos de grado d , y solo una cantidad finita de estos conjuntos son no vacíos; y el modelo asociado a un hiperparámetro será el ideal generado por la unión de las bases de $\ker(U_{M_d}(E))$ para todo d .

Concretamente, empezamos la búsqueda de un hiperparámetro adecuado con un conjunto de hiperparámetros A vacío, y fijando el grado $d=1$ y la elección de M_d igual al conjunto de monomios de grado igual a d , que se corresponde con la matriz $U_d(E)$, con E un conjunto de entrenamiento. Si las ecuaciones obtenidas no generan un ideal de dimensión acorde a la obtenida para V por los estimadores, aumentamos d en una unidad y añadimos las ecuaciones del grado siguiente. Al igual que en el caso afín, si llegamos a un hiperparámetro $(M_1, M_2, \dots, M_d, \dots)$, donde cada M_d es igual al conjunto de monomios de grado igual a d , de tal manera que el modelo resultante tenga la dimensión adecuada, realizamos la validación cruzada con este hiperparámetro y lo guardamos en el conjunto A y calculamos su resultado de validación si la dimensión es la correcta en todos los conjuntos de entrenamiento. Por supuesto, análogamente al caso anterior, también se pueden elegir hiperparámetros de manera que algún M_d esté formado por polinomios homogéneos de grado d que generen un subespacio contenido en el espacio generado por el conjunto de monomios de grado igual a d . Después, se itera este proceso, aumentando d en una unidad en cada paso hasta llegar a otro modelo de dimensión adecuada. Por último, el criterio para parar la búsqueda, la elección del hiperparámetro óptimo, la computación del modelo definitivo, y su evaluación son los mismos que en el caso afín.

La metodología anterior tiene la ventaja de que comienza la búsqueda incluyendo únicamente en M polinomios de grado bajo, lo que además hace que el tamaño de M sea pequeño en los primeros pasos, por lo que, por los comentarios anteriores acerca del grado y cantidad de polinomios de M , si existe una variedad de ajuste V generada por polinomios de grado d bajo, esperamos encontrar una solución satisfactoria en pocos pasos, que generalice y que cumpla $I(V) \cap R_M = I(C) \cap R_M$, con M el conjunto de monomios de grado menor o igual que d . En otros casos en los que las variedades algebraicas adecuadas no admitan generadores de grado bajo y/o el número de puntos m sea pequeño, será más probable que tengamos que buscar como hiperparámetro un conjunto M' que no sea el de monomios de grado menor o igual que algún número. Creemos que la búsqueda de ecuaciones con algoritmos dispersos en este caso puede ayudar a buscar este tipo de hiperparámetros, por ejemplo llevándonos a buscar un conjunto M' que elimine ciertas ecuaciones que aparecen al computar un núcleo de la forma $\ker(U_{\leq d}(C))$ porque el resto de ecuaciones correspondan a una variedad

algebraica conocida, como las que podemos ver en el capítulo 2 de [2]. Por último comentamos que el estudio de otras propiedades de las variedades algebraicas que se contempla en este mismo artículo, tales como su grado o componentes irreducibles puede ser útil para enriquecer esta metodología. Dejamos para un trabajo futuro la aplicación de una metodología de este tipo, con una evaluación de los modelos obtenidos y un análisis de su capacidad de generalización.

En el capítulo siguiente mostraremos varios ejemplos de estimación de variedades algebraicas a partir de conjuntos de datos, pero con un diseño experimental más sencillo: tomaremos un conjunto C de m puntos en una variedad V que admita un sistema de generadores de grado bajo, a los que en algunos casos añadiremos un pequeño error, y supondremos que la cantidad de puntos es suficientemente grande (de un orden de magnitud mayor que la cantidad de puntos del ejemplo de este capítulo), como para que $I(V) \cap R_M = I(C) \cap R_M$ si M está formado por polinomios de grado bajo. Realizaremos una búsqueda de ecuaciones en $\ker(U_{\leq d}(C))$ en el caso afín y $\ker(U_d(C))$ en el caso proyectivo empezando en $d=1$ y aumentando el grado d en una unidad en cada paso. Consideraremos que el único modelo de grado bajo que obtendremos es correcto y generaliza bien si, tras un redondeo en los coeficientes, llegamos a ecuaciones que definan la variedad algebraica V , por lo que prescindiremos de conjuntos de validación y test, y centraremos la exposición de los ejemplos en la computación de las ecuaciones con el software utilizado y la concordancia entre la dimensión de V proporcionada por los estimadores del capítulo 2 y la dimensión calculada en Singular a través del ideal $I(V)$ según las definiciones del capítulo 1.

4. Resultados de los experimentos

En este capítulo aplicaremos la metodología propuesta al final de la sección 3.2 del capítulo anterior a ejemplos prácticos de estimación de variedades V a partir de conjuntos de datos que pertenecen a V salvo un pequeño error y discutiremos los resultados obtenidos. Presentaremos cuatro ejemplos sencillos que centraremos en diferentes técnicas desarrolladas durante el trabajo.

En el primer ejemplo, V será un subespacio afín, lo que nos servirá como ejemplo de aplicación del Análisis de Componentes Principales utilizando R y para mostrar la gran estabilidad numérica de su cálculo, en presencia de ruido y de datos faltantes. En el segundo ejemplo, V será una circunferencia, lo que nos permitirá analizar un ejemplo con un conjunto de datos muy parecido al del Ejemplo 3.3, pero centrándonos en la computación de las ecuaciones y la sensibilidad de los coeficientes al ruido en los datos en Julia en vez de en la generalización y sobreajuste de los modelos.

En el tercer ejemplo, la variedad V será el grupo de matrices de rotación $SO(3)$. Este ejemplo está tomado de [2] y lo desarrollaremos poniendo el enfoque en los diagramas de dimensión que obtendremos en Julia, y en la diferencia que puede existir entre las ecuaciones estimadas en función del algoritmo utilizado para la computación del núcleo de las matrices de Vandermonde. Y, por último, en el cuarto ejemplo, V será la «cúbica retorcida» («twisted cubic», en inglés), que es una variedad proyectiva, y nos servirá para probar la computación de ecuaciones de V en el caso proyectivo, y la computación de la dimensión en Singular y la relación entre el polinomio de Hilbert proyectivo y afín que discutimos en el capítulo 1.

4.1. Subespacio afín

En esta sección la variedad V a estimar es el subespacio afín de \mathbb{R}^6 de dimensión 2, dado en ecuaciones paramétricas, con parámetros t y u , por:

$$(x_1, x_2, x_3, x_4, x_5, x_6) = (7, -2, 0, 11, 2, -1) + t(2, 3, -4, 0, 0, -1) + u(0, -4, 1, 1, -5, 0).$$

Realizaremos tres experimentos diferentes con esta variedad, con computación en lenguaje R.

4.1.1. Datos pertenecientes al subespacio

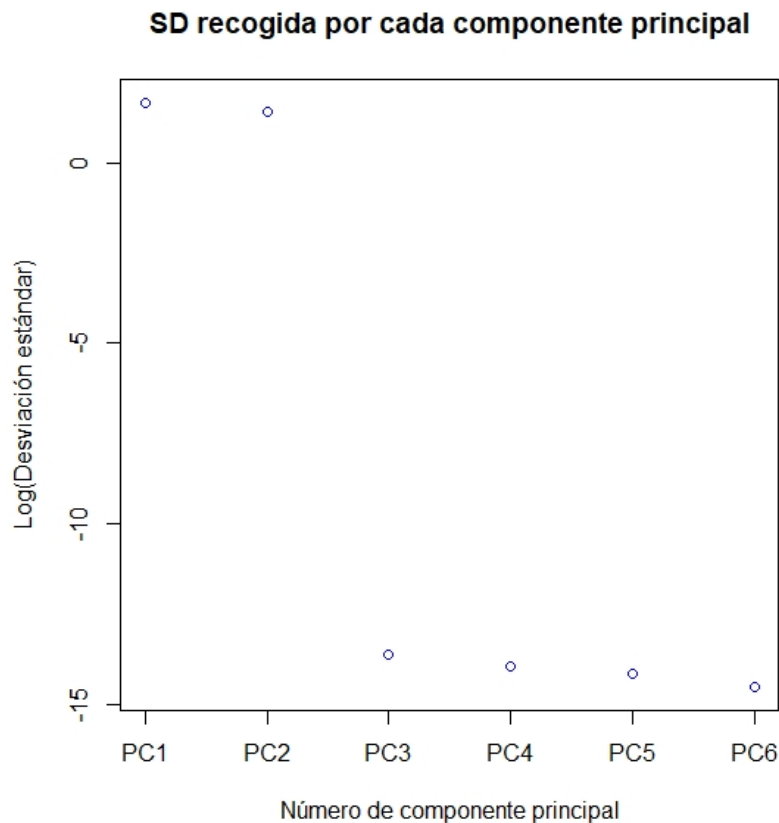
Tomamos un conjunto C de puntos en V recorriendo con un doble bucle todas las combinaciones de t y u con $t \in [-8, 10] \cap \mathbb{Z}$, y con $u \in [-15, 6] \cap \mathbb{Z}$. Esto nos da 19 posibilidades para t y 22 para u , y por tanto un conjunto C de $m = 19 \times 22 = 418$ puntos. Cada uno de estos puntos será una fila de una matriz

de tamaño 418×6 . Restando a cada columna de esta matriz su media (es decir, restando a cada punto el vector de medias, $\text{medias}=(9.0, 19.0, -8.5, 6.5, 24.5, -2.0)$), obtenemos la matriz de datos centrada X , lista para realizar el Análisis de Componentes Principales. El objetivo es estimar el subespacio vectorial que contiene a los datos centrados y trasladarlo después sumando el vector medias.

Para realizar el ACP utilizaremos la función `pca` del paquete `pcaMethods` [15], la cual por defecto computa la descomposición en Valores Singulares si no hay datos faltantes y computa el algoritmo NIPALS en caso contrario. En este caso, ejecutamos

```
A <- pca(X, nPcs = 6, seed = NA, threshold = 1e - 8, maxIterations = 100000).
```

Las desviaciones estándar (SD) asociadas a cada componente principal se obtienen ejecutando `sDev(A)`. En la siguiente figura podemos ver su logaritmo en base 10.



Vemos que las cuatro últimas desviaciones estándar son casi nulas, salvo por un pequeñísimo error de redondeo, ya que los puntos pertenecen realmente a V . Por tanto la dimensión estimada es 2, y las dos primeras componentes principales las obtenemos como:

```
pc1 <- loadings(A)[, 1]  
pc2 <- loadings(A)[, 2],
```

de tal manera que

$$\begin{aligned}pc1 &= (0,109491 \quad 0,670357 \quad -0,345513 \quad -0,126530 \quad 0,632650 \quad -0,054746) \\pc2 &= (-0,392881 \quad -0,133187 \quad 0,671729 \quad -0,114033 \quad 0,570168 \quad 0,196440).\end{aligned}$$

Computando la descomposición QR de la matriz cuyas filas son las dos primeras componentes principales y los dos vectores que generan el subespacio vectorial asociado a V , obtenemos que la matriz tiene rango 2, por lo que este último es igual al subespacio generado por las dos primeras componentes principales. Además, el vector cuyas componentes son las medias de cada columna de la matriz de datos de partida es

$$medias = (7 \quad -2 \quad 0 \quad 11 \quad 2 \quad -1) + 1(2 \quad 3 \quad -4 \quad 0 \quad 0 \quad -1) - 4,5(0 \quad -4 \quad 1 \quad 1 \quad -5 \quad 0),$$

por lo que el subespacio afín que hemos estimado, igual a $(9,0 \quad 19,0 \quad -8,5 \quad 6,5 \quad 24,5 \quad -2,0)$ más el subespacio vectorial generado por $pc1$ y $pc2$ es exactamente V .

4.1.2. Datos con ruido

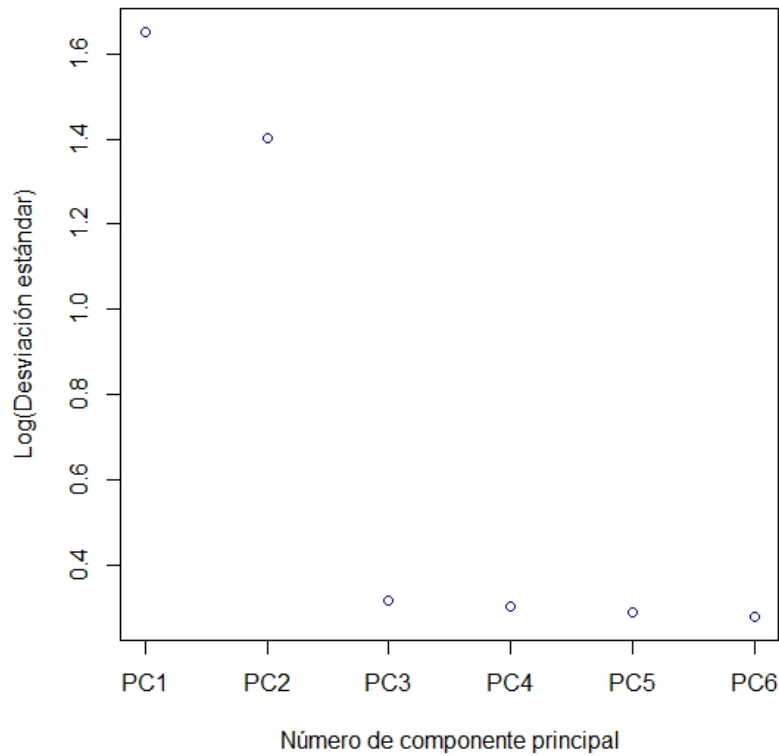
En una segunda prueba, tomamos el conjunto de puntos de la sección anterior y le añadimos a cada punto en cada coordenada un ruido aleatorio generado por una distribución normal de media 0 y desviación estándar 2, es decir, el conjunto de $m=418$ puntos de este experimento está formado por los puntos de la forma:

$$(7, -2, 0, 11, 2, -1) + t(2, 3, -4, 0, 0, -1) + u(0, -4, 1, 1, -5, 0) + ruido,$$

con $ruido = rnorm(n=6, mean=0, sd=2)$, y t y u tomando los mismos valores que en el apartado anterior.

Repitiendo exactamente los mismos comandos de R para realizar el ACP (con descomposición en valores singulares), en este caso el resultado es que las desviaciones estándar asociadas a cada componente principal son las que podemos ver en la siguiente gráfica:

SD recogida por cada componente principal



Si bien en este caso, el ruido hace que aparezca una variación en las últimas cuatro componentes principales que es mucho mayor que en el ejemplo anterior, siguiendo el criterio de estimación de la dimensión en el caso lineal del Capítulo 3.1, estimamos que la dimensión de V es 2 porque el mayor salto entre el logaritmo de la desviación estándar de una componente principal y el logaritmo de la siguiente se da entre la segunda y tercera componentes principales. Las dos primeras componentes principales explican el 90.25% de la desviación estándar.

Las dos primeras componentes principales que generan el subespacio vectorial estimado son:

$$pc1 = (0,108337 \quad 0,671264 \quad -0,344593 \quad -0,125054 \quad 0,632711 \quad -0,054420)$$

$$pc2 = (-0,396646 \quad -0,128607 \quad 0,674107 \quad -0,108328 \quad 0,566969 \quad 0,196271).$$

por lo que vemos un ejemplo de que el ACP con descomposición en valores singulares es muy estable frente a la presencia de errores: los coeficientes de los vectores de las dos primeras componentes principales son casi iguales a los del ejemplo anterior.

4.1.3. Datos con información faltante

En una última prueba, tomamos el conjunto C de puntos de V de la sección 4.1.1 y suprimimos el valor de algunas de las coordenadas de algunos de los datos para dar un ejemplo de cómo se comporta el algoritmo NIPALS con un conjunto de datos con información incompleta. Para ello, programamos una función auxiliar llamada `vaciado(matriz, porcentaje, mincol, minfila)`, que realiza un vaciado aleatorio de una matriz, reemplazando el valor del porcentaje de elementos especificado de la misma por NA (no disponible), y dejando siempre unos porcentajes mínimos `mincol` y `minfila` de elementos disponibles en cada columna y en cada fila, respectivamente. Estos dos últimos porcentajes sirven para evitar que el vaciado aleatorio de la matriz de datos X elimine la mayoría de las coordenadas de algún punto, o los valores de una variable en la mayoría de puntos, ya que en un proceso de análisis de datos real recomendamos eliminar las filas o columnas que estén prácticamente vacías como paso previo a la realización del ACP por considerar que no proporcionan mucha información y porque su eliminación reduce el porcentaje de datos a estimar, haciendo más fiable el análisis del resto de datos.

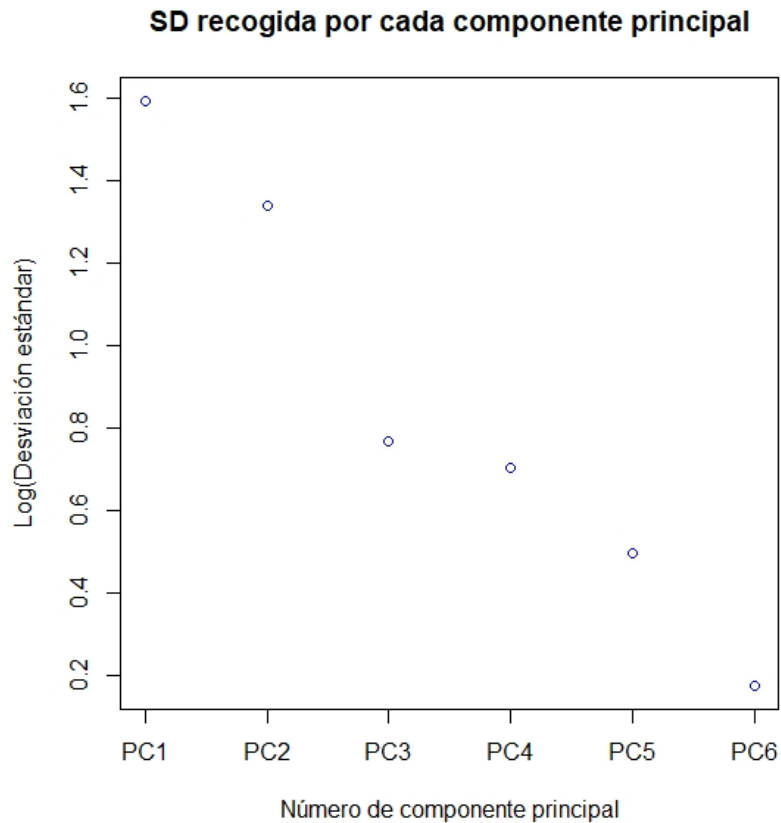
En este caso, realizamos el vaciado de la matriz de datos X del 15% de sus elementos, dejando el 50% de elementos como mínimo en cada fila y en cada columna. Ejecutamos

```
 $X <- \text{vaciado}(\text{matriz} = X, \text{porcentaje} = 15, \text{mincol} = 50, \text{minfila} = 50),$ 
```

y realizamos el ACP de la matriz resultante ejecutando

```
 $A <- \text{pca}(X, \text{nPcs} = 6, \text{seed} = NA, \text{threshold} = 1e - 8, \text{maxIterations} = 100000).$ 
```

Como la matriz de entrada tiene datos faltantes, la función `pca` utiliza el algoritmo NIPALS. Las desviaciones estándar obtenidas asociadas a cada componente principal son:



Vemos que la dimensión estimada para V es 2, como en los casos anteriores y las dos primeras componentes principales, que recogen un 79.65 % de la desviación estándar, son

$$pc1 = (0,102518 \quad 0,681943 \quad -0,342764 \quad -0,117361 \quad 0,624839 \quad -0,052529)$$

$$pc2 = (0,375962 \quad 0,142010 \quad -0,694398 \quad 0,104003 \quad -0,554987 \quad -0,193570).$$

por lo que vemos que el algoritmo NIPALS proporciona una buena estimación al subespacio de partida V con un porcentaje moderado de datos faltantes (15%). Aunque no presentamos más resultados en esta sección, comentamos que incluso en experimentos en los que se parte de un conjunto de datos del subespacio, se añade un ruido del orden de magnitud del añadido en la sección 4.1.2, y posteriormente se suprime un porcentaje de elementos similar al suprimido en esta sección, la aplicación del ACP nos da como resultado dimensión 2 y una estimación razonable de V .

4.2. Circunferencia

En esta sección la variedad V a estimar es la circunferencia en \mathbb{R}^2 de ecuación $f = x^2 + y^2 - 1 = 0$. De forma similar al ejemplo 3.3, tomamos un conjunto C de 100

puntos en la circunferencia $V(f)$ con un error, de la forma:

$$(x, y) = (\cos(t) + e1, \sin(t) + e2)$$

donde t sigue una distribución uniforme en $[0, 2\pi]$, y $e1$ y $e2$ son los términos de error, de distribución uniforme en $[-0.1, 0.1]$.

Este sencillo ejemplo servirá para presentar con detalle la computación en Julia que utilizaremos en el caso de estimación de variedades algebraicas en el caso general, utilizando el paquete `LearningAlgebraicVarieties` desarrollado por los autores del artículo [2]. Comenzaremos añadiendo los siguientes paquetes al iniciar sesión en Julia si no los hemos añadido previamente:

```
import Pkg
Pkg.clone("https://github.com/PBrdng/LearningAlgebraicVarieties.git")
Pkg.add("Distributions")
Pkg.add("JLD")
Pkg.add("JLD2")
Pkg.add("WebIO")
Pkg.build("WebIO")
```

Después de importar la función `Pkg`, hemos descargado los paquetes `LearningAlgebraicVarieties`, que contiene la mayoría de funciones que utilizaremos, como las de estimación de la dimensión o búsqueda de ecuaciones. El paquete `Distributions` nos servirá para poder tomar puntos de distribuciones como la uniforme, normal, etc, que usaremos para obtener conjuntos de datos, y los demás paquetes nos servirán para cargar datos y comunicarnos con páginas web. A continuación, cargamos los paquetes que hemos añadido (cada vez que iniciemos una sesión nueva en Julia), además de otros paquetes auxiliares para manejar datos, o paquetes con funciones de álgebra lineal.

```
using LearningAlgebraicVarieties
using Distributions
import JLD: load
using LinearAlgebra
using WebIO
import HTTP
using HTTP: HTTP
using DelimitedFiles
```

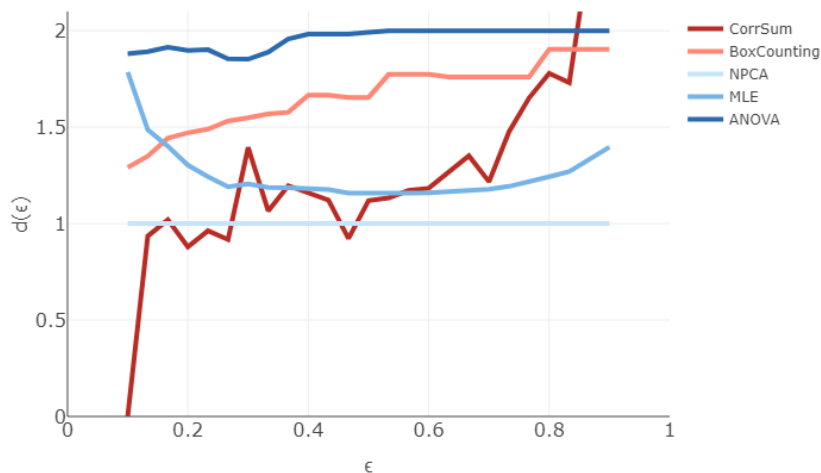
Tras estos pasos previos, comenzaremos el estudio del ejemplo de esta sección generando los datos de la circunferencia en la matriz `circunferencia_datos`, a partir de dos distribuciones uniformes `t_unif` y `ruido_unif` para el parámetro `t` y el ruido, como:

```
t_unif = Uniform(0, 2*pi)
ruido_unif = Uniform(-0.1, 0.1)
circunferencia_datos = zeros(Float64, 2, 100)
parametrizacion_circ=rand(t_unif,100)
circunferencia_datos[1,:]=cos.(parametrizacion_circ).+rand(ruido_unif,100)
circunferencia_datos[2,:]=sin.(parametrizacion_circ).+rand(ruido_unif,100)
```

Para comenzar el análisis, realizamos los diagramas de dimensión de los puntos, que presentamos en el capítulo 2.2, con el comando:

```
DimensionDiagrams(circunferencia_datos, false)
```

El parámetro `false` indica que la variedad es afín, y sería igual a `true` si buscáramos una variedad proyectiva. Obtenemos los siguientes diagramas:



Por un lado, vemos que los métodos de Conteo de Cajas y Dimensión de Correlación sobrestiman para todo ϵ el verdadero valor de la dimensión de V , que es 1 y por otro lado, vemos que los diagramas presentan su zona más estable entre aproximadamente los valores $\epsilon = 0.2$ y 0.6 , donde la estimación de la dimensión de los tres métodos restantes tiene el 1 como valor entero más cercano, luego el criterio de votación de la mayoría que establecimos en el Capítulo 2.2 podemos afirmar que los diagramas estiman correctamente que la dimensión de V es igual a 1.

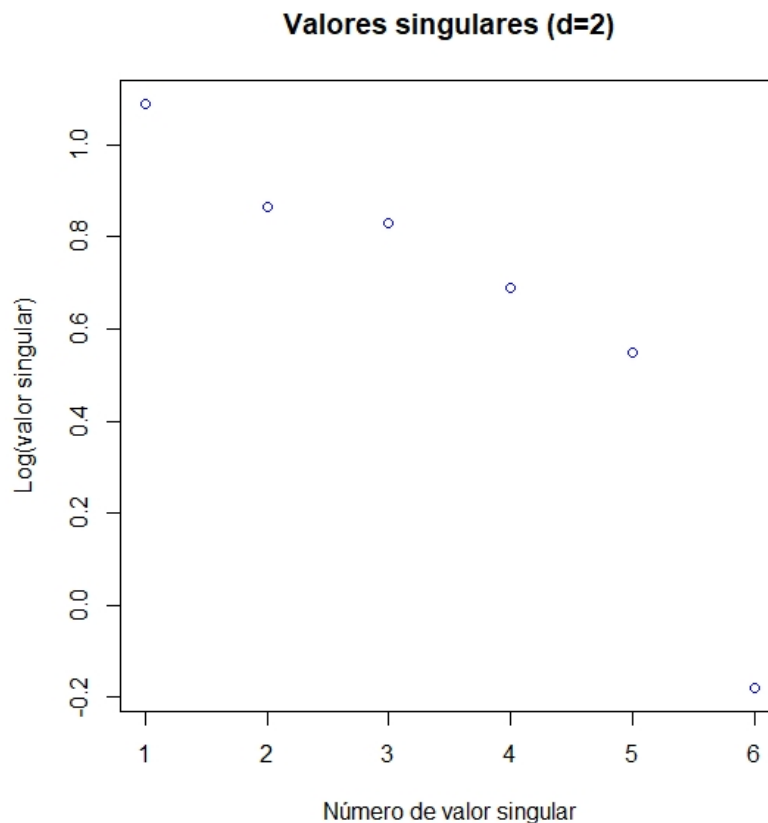
El siguiente paso es la búsqueda de las ecuaciones que se anulan en V . Para ello utilizamos la función `MultivariateVandermondeMatrix(datos, d, proyectivo)`, que

calcula la matriz de Vandermonde multivariante $U_{\leq d}(C)$ si `proyectivo = false`, y la matriz $U_d(C)$ si `proyectivo = true`, donde `datos` es la matriz cuyas columnas son los elementos de C (es la matriz traspuesta de la matriz X de datos de , según la definimos en el Capítulo 2.1). Empezamos por la búsqueda de ecuaciones lineales, de grado $d=1$, con descomposición en valores singulares, escribiendo

```
M = MultivariateVandermondeMatrix(circunferencia_datos, 1, false)
valores = svdvals(M.Vandermonde)
logvalores = log10.(valores),
```

donde la función `svdvals` calcula los valores singulares, que en este caso son: 10.03, 7.29 y 6.74, por lo que la diferencia entre sus logaritmos no es significativa, y llegamos a la conclusión de que no hay ecuaciones lineales que se anulen en los datos. Alternativamente, aplicar ACP a la matriz de datos de C (`circunferencia_datost`) no nos proporciona ninguna ecuación.

Como el análisis lineal no nos ha proporcionado ningún modelo, pasamos a buscar polinomios de grado $d=2$. Ejecutando los comandos anteriores pero cambiando el valor de d de 1 a 2, obtenemos los logaritmos de los valores singulares (variable `logvalores`) que podemos ver en la siguiente figura.



Vemos que el mayor salto en estos logaritmos se produce entre el penúltimo y el último, y es significativo, por lo que elegimos una tolerancia que aproxime a 0 únicamente al último valor singular, por ejemplo:

`tol=(valores[5]+valores[6])/2.`

A partir de esta tolerancia y la matriz de Vandermonde multivariante M , podemos calcular las ecuaciones que son base del ker de esta matriz con la función

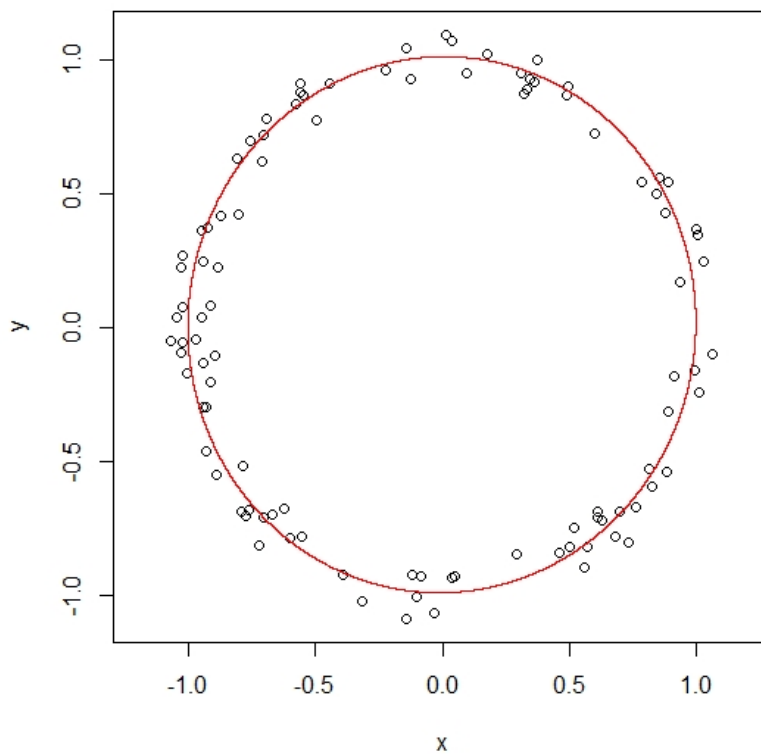
`f = FindEquations(M, :with_svd, tol),`

en la que el segundo argumento es el método utilizado, en este caso la descomposición en valores singulares. Para utilizar los métodos de factorización QR y RREF se puede hacer este argumento igual a `:with_qr` y `:with_rref`, respectivamente.

La ecuación obtenida con este método es

$$f = -0,998330x^2 + 0,019868xy - 0,995072y^2 - 0,001117x + 0,016028y + 1,0 = 0,$$

que es una elipse muy parecida a la circunferencia original V , que dibujamos en color rojo en la siguiente gráfica. Como un redondeo en los coeficientes de f nos da la ecuación de la circunferencia de partida, concluimos que la ecuación encontrada es un buen modelo para los datos.



4.3. Grupo de rotaciones $SO(3)$

En el siguiente ejemplo ([2], Ejemplo 7.1) la variedad V a estudiar es el grupo $SO(3)$ de las matrices de rotación en 3 dimensiones, es decir las matrices R de tamaño 3×3 con determinante igual a 1 y ortogonales. Si llamamos a las entradas de la matriz:

$$R = \begin{pmatrix} x_1 & x_2 & x_3 \\ x_4 & x_5 & x_6 \\ x_7 & x_8 & x_9 \end{pmatrix},$$

$V=SO(3)$ es una variedad en \mathbb{R}^9 definida por las ecuaciones:

$$\begin{aligned} f_1 &= -x_1^2 - x_2^2 - x_3^2 + 1; \\ f_2 &= x_4^2 + x_5^2 + x_6^2 - 1; \\ f_3 &= -x_7^2 - x_8^2 - x_9^2 + 1; \\ f_4 &= x_1 * x_4 + x_2 * x_5 + x_3 * x_6; \\ f_5 &= x_1 * x_7 + x_2 * x_8 + x_3 * x_9; \\ f_6 &= x_4 * x_7 + x_5 * x_8 + x_6 * x_9; \\ f_7 &= x_1 * x_2 + x_4 * x_5 + x_7 * x_8; \\ f_8 &= x_1 * x_3 + x_4 * x_6 + x_7 * x_9; \\ f_9 &= x_2 * x_3 + x_5 * x_6 + x_8 * x_9; \\ f_{10} &= \det(R) - 1. \end{aligned}$$

Las primeras 9 cuádricas indican que R es ortonormal, y la última ecuación es una cúbica. El ideal I generado por estos 10 polinomios es primo [2] y contiene 20 cuádricas linealmente independientes que lo generan.

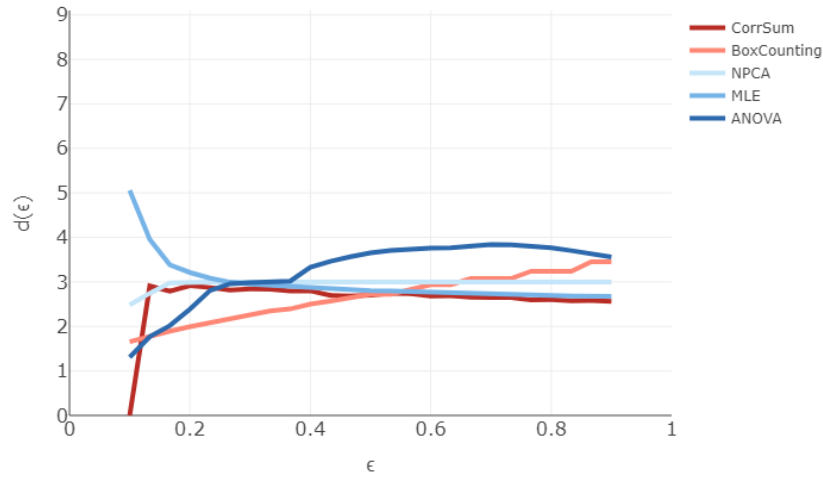
El conjunto C de puntos a estudiar en este ejemplo consta de $m = 887$ puntos en $V = SO(3)$ y está disponible en el paquete `LearningAlgebraicVarieties`. Para empezar, lo descargamos y lo guardamos en la variable `SO3_datos` ejecutando

```
s = string(Pkg.dir("LearningAlgebraicVarieties"),"datasets.jld2")
datasets=load(s)
SO3_datos = datasets["data"]["SO(3)"].
```

Como en el ejemplo anterior, comenzamos ejecutando

```
DimensionDiagrams(SO3_datos, false),
```

lo que nos proporciona los siguientes diagramas de dimensión, que estiman correctamente la dimensión de la variedad que estamos estudiando, que es igual a 3, con un gran acuerdo en todos los estimadores en valores de ϵ entre 0.2 y 0.4, con la excepción del estimador de Conteo de Cajas, que subestima la dimensión ligeramente en ese intervalo.



A continuación buscamos ecuaciones que se anulen en C . Al igual que en el ejemplo anterior, la búsqueda de ecuaciones lineales no nos proporciona ningún resultado, por lo que buscamos ecuaciones de grado $d=2$. En esta ocasión utilizamos la función FindEquations con diferentes argumentos a los empleados en el ejemplo anterior. En este caso no estamos interesados en el ruido (los datos de C pertenecen a V salvo error de redondeo), por lo que dejamos la tolerancia que viene por defecto, y damos como entrada a la función la matriz de datos `SO3_datos` en vez de una matriz de Vandermonde multivariante. Si utilizamos descomposición en valores singulares, con el comando

```
f = FindEquations(SO3_datos, :with_svd, 2, false),
```

el resultado son 20 cuádricas linealmente independientes, pero muy dispersas, y por tanto difíciles de interpretar, como la que presentamos a continuación (con redondeo a 3 decimales).

$$\begin{aligned}
 f_1 = & (0,053) * x_1^2 + (-0,002) * x_1 * x_2 + (-0,014) * x_1 * x_3 + (0,018) * x_1 * x_4 + \\
 & + (-0,085) * x_1 * x_5 + (-0,023) * x_1 * x_6 + (-0,032) * x_1 * x_7 + (-0,035) * x_1 * x_8 + \\
 & + (-0,306) * x_1 * x_9 + (-0,040) * x_2^2 + (-0,016) * x_2 * x_3 + (0,085) * x_2 * x_4 + \\
 & + (0,018) * x_2 * x_5 + (-0,004) * x_2 * x_6 + (0,034) * x_2 * x_7 + (-0,032) * x_2 * x_8 + \\
 & + (-0,014) * x_2 * x_9 + (0,042) * x_3^2 + (0,023) * x_3 * x_4 + (0,004) * x_3 * x_5 +
 \end{aligned}$$

$$\begin{aligned}
&+(0,018) * x_3 * x_6 + (0,306) * x_3 * x_7 + (0,014) * x_3 * x_8 + (-0,032) * x_3 * x_9 + \\
&+(0,285) * x_4^2 + (-0,002) * x_4 * x_5 + (-0,014) * x_4 * x_6 + (0,014) * x_4 * x_7 + \\
&+(0,018) * x_4 * x_8 + (0,020) * x_4 * x_9 + (0,192) * x_5^2 + (-0,017) * x_5 * x_6 + \\
&+(-0,018) * x_5 * x_7 + (0,014) * x_5 * x_8 + (0,371) * x_5 * x_9 + (0,274) * x_6^2 + \\
&+(-0,020) * x_6 * x_7 + (-0,371) * x_6 * x_8 + (0,014) * x_6 * x_9 + (0,016) * x_7^2 + \\
&+(-0,002) * x_7 * x_8 + (-0,014) * x_7 * x_9 + (-0,078) * x_8^2 + (-0,017) * x_8 * x_9 + \\
&+(0,005) * x_9^2 + (-0,371) * x_1 + (0,020) * x_2 + (-0,018) * x_3 + (-0,014) * x_4 + \\
&+(0,306) * x_5 + (-0,035) * x_6 + (0,004) * x_7 + (-0,023) * x_8 + (0,085) * x_9 + (-0,250).
\end{aligned}$$

Sin embargo, si realizamos la búsqueda de ecuaciones con el algoritmo de descomposición QR, y redondeamos los polinomios obtenidos,

```
f = FindEquations(SO3_datos, :with_qr, 2, false)
round.(f),
```

el resultado son las siguientes 20 cuádricas dispersas:

$$\begin{aligned}
f_1 &= x_1 * x_4 + x_2 * x_5 + x_3 * x_6; \\
f_2 &= x_1 * x_7 + x_2 * x_8 + x_3 * x_9; \\
f_3 &= -x_1^2 - x_2^2 - x_3^2 + x_4^2 + x_5^2 + x_6^2; \\
f_4 &= x_4 * x_7 + x_5 * x_8 + x_6 * x_9; \\
f_5 &= -x_2^2 - x_3^2 + x_4^2 + x_7^2; \\
f_6 &= x_1 * x_2 + x_4 * x_5 + x_7 * x_8; \\
f_7 &= x_1 * x_3 + x_4 * x_6 + x_7 * x_9; \\
f_8 &= -x_1^2 - x_3^2 + x_5^2 + x_8^2; \\
f_9 &= x_2 * x_3 + x_5 * x_6 + x_8 * x_9; \\
f_{10} &= x_3^2 - x_4^2 - x_5^2 + x_9^2; \\
f_{11} &= -x_5 * x_9 + x_6 * x_8 + x_1; \\
f_{12} &= x_4 * x_9 - x_6 * x_7 + x_2; \\
f_{13} &= -x_4 * x_8 + x_5 * x_7 + x_3; \\
f_{14} &= x_2 * x_9 - x_3 * x_8 + x_4; \\
f_{15} &= -x_1 * x_9 + x_3 * x_7 + x_5; \\
f_{16} &= x_1 * x_8 - x_2 * x_7 + x_6; \\
f_{17} &= -x_2 * x_6 + x_3 * x_5 + x_7; \\
f_{18} &= x_1 * x_6 - x_3 * x_4 + x_8; \\
f_{19} &= -x_1 * x_5 + x_2 * x_4 + x_9; \\
f_{20} &= -x_1^2 - x_2^2 - x_3^2 + 1.
\end{aligned}$$

En este caso, es sencillo ver que estas cuádricas se corresponden con las 20 cuádricas del ideal I . Por ejemplo, la combinación $f_3 - f_{20}$ indica que la segunda fila de la matriz R tiene norma 1. Con un cambio de base sencillo (con operaciones similares a la anterior), llegamos a una base de 20 cuádricas [2] que contiene a las 9 cuádricas correspondientes a la ortonormalidad de filas y columnas, otras dos cuádricas de la diagonal de $R^t R - \text{Id}_3$, y 9 cuádricas que expresan la regla de la mano derecha para la orientación, como por ejemplo $x_5 * x_9 - x_6 * x_8 - x_1$.

Por último comentamos por un lado que la computación en Singular con la técnica que veremos en el apartado siguiente confirma que el ideal asociado a esta variedad tiene dimensión 3; y por otra parte que, a pesar de que la descomposición QR es menos precisa que la descomposición en valores singulares, hemos realizado un ejemplo añadiendo al conjunto de datos descargado un error que sigue una distribución uniforme en $[-0.01, 0.01]$, y con una tolerancia obtenida a partir de los logaritmos

de los valores singulares de la respectiva matriz de Vandermonde, (al igual que en el apartado anterior), la descomposición QR nos da como resultado las mismas 20 cuádricas, después de un redondeo en los coeficientes.

4.4. Cúbica alabeada

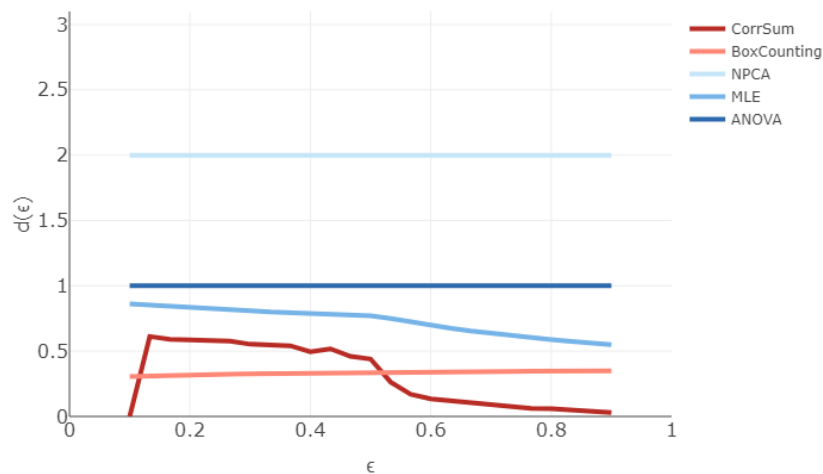
Como último ejemplo, tomamos un conjunto C de $m = 201$ puntos pertenecientes a la «Cúbica alabeada proyectiva» $V = V(\langle x_2^2 - x_1x_3, x_1x_2 - x_3x_0, x_1^2 - x_2x_0 \rangle)$, que presentamos en el último ejemplo del Capítulo 1. Tomamos los puntos en la carta afín $V \cap \{x_0 = 1\}$, que es la «Cúbica alabeada» (afín) $V(x_2 - x_1^2, x_3 - x_1^3) \subset \mathbb{R}^3$, que tiene trivialmente la parametrización $(x_1, x_2, x_3) = (t, t^2, t^3)$. Generamos los puntos tomando valores del parámetro t enteros entre -100 y 100, de la siguiente manera:

```
t = [-100:100][1]
twisted_datos=zeros(Float64, 4, 2*n+1)
twisted_datos[1,:]=twisted_datos[1,:].+1
twisted_datos[2,:]=t
twisted_datos[3,:]=t.^ 2
twisted_datos[4,:]=t.^ 3
```

Empezamos el estudio del conjunto de datos con la estimación de la dimensión. No ha sido posible computar la dimensión con los estimadores de dimensión proyectiva con este conjunto, debido a un error en una de las funciones internas del paquete LearningAlgebraicVarieties, sin embargo, sí es posible realizar los diagramas de dimensión (afín) de la «Cúbica alabeada» afín, utilizando todas las filas en menos la primera, que se corresponde con la variable x_0 , es decir, ejecutando,

```
DimensionDiagrams(twisted_datos[2:4,:], false),
```

lo que nos da los siguientes diagramas, que vemos que muestran bastante disparidad entre los indicadores. El ACP no lineal sobrestima la dimensión y el Conteo de cajas la subestima, para todo ϵ . Aún así, los valores de ϵ para los que más indicadores están próximos son los de un entorno pequeño de $\epsilon = 0.2$, donde el entero más cercano en los tres estimadores restantes es 1, que es el valor de la dimensión real.



Sabemos que la dimensión de la «Cúbica alabeada» afín es realmente 1 porque la calculamos en el último ejemplo del capítulo 1, pero también la podemos calcular en Singular con los comandos:

```
ring r=0,(x_1,x_2,x_3),dp;
ideal i = x_2-x_1^2,x_3-x_1^3;
dim(groebner(i));
```

donde para utilizar la función dim tenemos que especificar como entrada una base de Groebner del ideal i. O equivalentemente, como adelantábamos en el mismo ejemplo del Capítulo 1, podemos calcularla como el grado del polinomio de Hilbert proyectivo (con la función hilbPoly) del ideal i homogeneizado, que resulta ser $3t + 1$, con lo que vemos además que la dimensión de la «Cúbica alabeada proyectiva» V también es 1:

```
LIB "poly.lib";
ring r=0,(x_0,x_1,x_2,x_3),dp;
ideal i=x_2-x_1^2,x_3-x_1^3;
ideal j=groebner(i);
hilbPoly(homog(j,x_0));
```

Una vez que hemos estimado correctamente la dimensión de V, busquemos sus ecuaciones a partir de los datos. Como en el ejemplo 4.2, utilizamos descomposición en valores singulares, con una tolerancia obtenida a partir de los valores singulares, de manera que aproximamos a 0 los últimos valores singulares después de un salto significativo en el valor de sus logaritmos. En este caso buscamos un ideal homogéneo, luego fijamos la entrada proyectivo = true, y ejecutamos:

```

M = MultivariateVandermondeMatrix(twisted_datos, 2, true)
valores = svdvals(M.Vandermonde)
logvalores = log10.(valores)
diferencia=zeros(Float64, (length(logvalores)-1))
for i in [1:(length(logvalores)-1)][1]
diferencia[i]=logvalores[i]-logvalores[i+1]
end
indice=findmax(diferencia)[2]
tol=(valores[indice]+valores[indice+1])/2
f = FindEquations(M, :with_svd, tol)

```

El resultado guardado en la variable f son las tres ecuaciones f_1 , f_2 y f_3 que vemos a continuación, cuyos coeficientes hemos redondeado a tres decimales y las hemos exportado a Singular para estudiar el ideal i que generan. Comenzamos por calcular su polinomio de Hilbert proyectivo:

```

ring r=(real,15),(x_1,x_2,x_3,x_4),dp;
poly f1 = (0.707)*x_2*x_4+(-0.707)*x_3^2;
poly f2 =
(-0.643)*x_1*x_3+(0.293)*x_1*x_4+(0.643)*x_2^2+(-0.293)*x_2*x_3;
poly f3 =
(0.293)*x_1*x_3+(0.643)*x_1*x_4+(-0.293)*x_2^2+(-0.643)*x_2*x_3;
ideal i = f1,f2,f3;
ideal j=groebner(i);
hilbPoly(j);

```

El resultado es $3t + 1$, por lo que el modelo generado por estas tres ecuaciones tiene la misma función de Hilbert proyectiva que V (y por tanto la misma dimensión). Calculando una base de groebner de j con

```
groebner(j);
```

obtenemos la misma base para el ideal que utilizamos para definir la «Cúbica alabeada proyectiva» V , $x_2^2 - x_1x_3$, $x_1x_2 - x_3x_0$, $x_1^2 - x_2x_0$ (salvo un cambio de nombre en las variables), por lo que hemos tenido éxito en la estimación de V .

Referencias

- [1] BOCHNAK, J., COSTE, M., AND ROY, M.-F. *Real Algebraic Geometry*. Springer, 1998.

- [2] BREIDING, P., KALISNIK, S., STURMFELS, B., AND WEINSTEIN, M. Learning algebraic varieties from samples. *Revista Matemática Complutense* (02 2018).
- [3] CARO, G. A. D. *Introduction to Machine Learning. Lecture 24: Nonlinear and Kernel regression 1*, 2019 (accedido el 21 de junio de 2020).
- [4] COX, D. A., LITTLE, J., AND O'SHEA, D. *Ideals, Varieties, and Algorithms: An Introduction to Computational Algebraic Geometry and Commutative Algebra*, second ed. Springer, 1997.
- [5] COX, D. A., LITTLE, J., AND O'SHEA, D. *Using Algebraic Geometry*, second ed., vol. 185 of *Graduate Texts in Mathematics*. Springer, Mar. 2005.
- [6] GATHMANN, A. First applications of scheme theory, 2002 (accedido el 11 de julio de 2020).
- [7] GONZÁLEZ, V. M. *Métodos PLS en Análisis Multibloques: Tratamiento de datos faltantes y multicolinealidad*, 2016 (accedido el 21 de junio de 2020).
- [8] HARTSHORNE, R. *Algebraic Geometry*, vol. 52 of *Graduate Texts in Mathematics*. Springer, 1977.
- [9] JOLLIFFE, I. *Principal component analysis. 2nd ed*, vol. 98. 10 2005.
- [10] KREUZER, M., AND ROBBIANO, L. *Computational Commutative Algebra 2*. 01 2005.
- [11] MATSUMURA, H. *Commutative Algebra*. Math Lecture Notes Series. Benjamin/Cummings Publishing Company, 1980.
- [12] SHEIHAM, D. Invariants of boundary link cobordism. *Memoirs of the American Mathematical Society* 165 (11 2001).
- [13] TREFETHEN, L. Barycentric lagrange interpolation. *SIAM Review* 46 (12 2002).
- [14] UGR. Power method for approximating eigenvalues, Accedido el 23 de junio de 2020.
- [15] WOLFRAM STACKLIES, HENNING REDESTIG, K. W. Package pcamethods, 1 de julio de 2020 (accedido el 02 de julio de 2020).