Orografía homológica cúbica para la minería de datos

por

Luis Javier Hernández Paricio

Departamento de Matemáticas y Computación, Universidad de La Rioja, C/Luis de Ulloa s/n. Edificio Vives. 26 004 Logroño, La Rioja, España.
email: luis-javier.hernandez@dmc.unirioja.es
página web: http://www.unirioja.es/dptos/dmc/luhernan/index.html

Resumen

Se introduce una nueva técnica para el estudio y clasificación de muestras de datos. Las ideas básicas están basadas en la llamada Teoría de la Forma introducida por Borsuk y que aproxima espacios localmente patológicos por una sucesión de poliedros. Aquí modificamos el modelo anterior tomando una aproximación poliedral (o un número finito) que además dispone de una aplicación densidad. Utilizamos poliedros construidos con cubos que llamaremos complejos cúbicos. Analizando el complejo cúbico y la orografía de la aplicación densidad con métodos homológicos se descubren interesantes propiedades del conjunto de datos. Además de disponer de herramientes para dividir el conjunto de datos en diferentes clases de cúmulos, de análizar posibles datos espurios y de calcular los invariantes de Betti y de torsión, las densidades cúbicas admiten derivadas parciales utilizables para ser interpolarladas mediante polinomios de varias variables. Todo ello hace que la herramienta sea muy completa y que sea un lugar de encuentro de los métodos algebraicos, geométricos, topológicos, analíticos, lógicos y algorítmicos. Un primer paquete en *Mathematica* para algunos aspectos de la Orografía Homológica Cúbica puede verse en la página web del autor.

Palabras clave

Mimería de datos, homología, números de Betti, coeficientes de torsión, orografía homológica cúbica, cúmulos, datos espurios, densidad cúbica.

Data Mining, homology, Betti numbers, torsion coefficients, cubical homological orography, clustering, outliers, cubical density,

Agradecimientos

Quiero agradecer la ayuda y orientaciones de Juan Felix Sanjuan sobre el uso de la aplicación Mathematica y algunos trucos para la optimización de algoritmos. También a Manuel Castejon por sus sugerentes comentarios que fueron los que nos motivaron a abordar este proyecto, a Francisco Javier Díaz por su trabajo de programación en las primeras versiones de algunos algoritmos para el estudio de complejos cúbicos y a Eduardo Sáenz de Cabezón por su entusiamo en la programación en los algoritmos de interpolación. Finalmente, a M.T. Rivas, J.I. Extremiana y otros compañeros que con sus críticas y comentarios han hecho posible los primeros pasos de la Orografía Homológica Cúbica.

El proyecto está financiado por la Dirección General de Investigación, por la Consejería de Educación, Cultura, Juventud y Deportes del Gobierno de la Rioja y también de la Universidad de La Rioja.

0 Introducción

De entre los posibles enfoques que un observador puede adoptar al analizar un conjunto de datos numéricos, merece especial importancia, tanto por su originalidad como por la importancia de los resultados obtenidos, aquél que estudia las características geométricas y topológicas de los poliedros asociados a los puntos de la muestra (que también llamaremos conjunto de datos). Esta aproximación permite analizar múltiples aspectos de los procesos objeto de estudio de una forma computacionalmente más sencilla que las actuales alternativas, proponiendo un enfoque integral, por permitir abordar múltiples dificultades a través de una única técnica, y complementario, pues, si bien los resultados obtenidos a partir de las técnicas de Orografía Homológica Cúbica gozan de interés ya por si mismas, nada obsta para que la información revelada nutra al resto de técnicas interesadas, proporcionándoles información de excepcional calidad concernientes a su geometría y topología.

Las sólidas bases matemáticas en las que se fundamenta la familia de algoritmos desarrollados, fiables y capaces de manejar de forma ágil grandes volúmenes de datos, nos prometen una larga serie de interesantes resultados en un campo en clara expansión y de innumerables dominios de competencia y aplicación como es la Minería de Datos.

Ya de forma inmediata y por su mera construcción, los poliedros facilitan la etapa de análisis exploratorio de los datos, acotando la ubicación de unas aisladas regiones de interés (aquellas donde se concentran los equivalentes numéricos de las ocurrencias observadas en la naturaleza) inmersas en un océano multidimensional cuya extensión parece crecer a pasos de gigante (efecto conocido como ``la maldición de la dimensionalidad'') con el número de variables observadas.

Además el estudio de sus propiedades geométricas y topológicas permite profundizar en el conocimiento de las estructuras ocultas; ocultas, precisamente, por la superabundancia de datos. Se revela entonces a nuestros ojos la información útil que permitirá posteriormente elaborar conclusiones basadas en criterios objetivos, que confirmarán o recusarán el conocimiento previo basado en los usos y costumbres.

En este trabajo, basándonos en los métodos poliedrales, pretendemos esbozar una nueva técnica para estudiar y clasificar conjuntos de datos que podríamos llamar Orografía Homológica Cúbica. Se trata de asociar a cada conjunto de datos (muestra) una estructura matemática que hemos llamado aplicación densidad cúbica que consiste en una aplicación $f:\mathbf{K}\rightarrow\mathbf{Z}$, donde $\mathbf{K}$ es un complejo cúbico y $\mathbf{Z}$ es el anillo de los enteros.

Nuestros objetivos son, por un lado, dar una descripción asequible de las herramientas matemáticas que se utilizan en este modelo, y por otro, enfatizar sobre la importancia que tienen los invariantes (co)homógicos. Un invariante de un conjunto de datos (muestra) es una estructura generalmente algebraica que no cambia de valor aunque se realicen pequeños cambios en el conjunto de datos inicial, salvo en el caso que cambios se realicen en la proximidad de puntos críticos de las muestra. Aquí proponemos estudiar invariantes numéricos que consisten en un conjunto ordenado finito de enteros no negativos que están asociados al conjunto de datos y que se denominan números de Betti y coeficientes de torsión.

Pretendemos clarificar dos cuestiones: La primera es la utilidad de los invariantes homológicos anteriores como una herramienta de clasificación y la segunda es la de mostrar que los cambios de valor de los números de Betti y de torsión son causados por cambios en los factores principales que regulan el sistema que ha producido estos datos. Es decir que los cambios de valor en los números de Betti y los coeficientes de torsión se deben a modificaciones importantes en la distribución de los datos o a cambios pequeños pero situados en las proximidades de los valores críticos de los parámetros.

Por otra parte es importante observar como aspectos tan importantes en la Minería de Datos como particiones en clases (Clusterización) y localización de datos espurios están relacionados con los valores críticos de la aplicación den-

sidad. En este trabajo se proponen varios métodos de asociar invariantes homológicos a las aplicaciones densidad, de manera que dispongamos de una técnica más fina y adecuada que el mero análisis de los complejos cúbicos que aproximan conjuntos de datos. Para ello utilizamos varios técnicas: proponemos estudiar las fibras de las aplicación densidad, sus mesetas, planicies y hoyas. Definimos el concepto de flexión direccional que divide el complejo cúbico dominio de la aplicación densidad en subcomplejos cuyos n-cubos tienen un comportamiento similar respecto los cambios de valor de la densidad en determinadas direcciones.

Para todas los métodos anteriores existen algoritmos "razonables" que pueden ser programados con facilidad. Como un botón de muestra, incluimos como referencia la siguiente página web que contiene un paquete en *Mathematica* con alguno de estos algoritmos; en particular , aproxima un conjunto de datos por una aplicación densidad cúbica, divide un complejo en pseudocomponentes, calcula complejos duales y es capaz de calcular la homología. El autor y otros colaboradores estamos elaborando una segunda versión que corrige muchos de los defectos de la primera y que podrá calcular otros muchos elementos importantes. Hasta que no este lista la nueva versión el lector interesado puede analizar la primera en:

http://www.unirioja.es/dptos/dmc/luhernan/index.html

Las páginas siguientes nos harán recorrer una senda en la que apreciaremos, primero que se entiende por complejo cúbico y aplicación densidad y como se construyen a partir de la materia prima con la que trabajamos: conjuntos de datos procedentes del muestreo de un proceso real. Una vez asimilado el concepto de complejo cúbico y aplicación densidad expondremos otras nociones necesarias para determinar las propiedades fundamentales que nos permitirán extraer, de la plétora numérica, información condensada útil, acerca de la geometría y topología del problema. También mostraremos como esta información puede ser aprovechada de forma práctica desde el punto de vista de la Minería de Datos. Concluimos el trabajo con un modelo analítico en el que toda la información condensada en una aplicación densidad es transformada en un equivalente analítico formado por un polinomio de varias variables junto con su cuboide dominio de definición. Este polinomio da valores a los puntos de un cubiode, de manera que para valores positivos mayores se tiene una mayor densidad de datos en el área próxima al punto, si los valores son negativos significa que no hay en la zona local puntos de la muestra. La hipersuperficie obtenida igualando a cero es polinomio junto con las condiciones de estar dentro del cubiode constituyen lo que podríamos denominar como ecuación del borde de la muestra.

1 Aproximación de conjuntos de datos por aplicaciones densidad cúbicas

Descripción de la materia prima

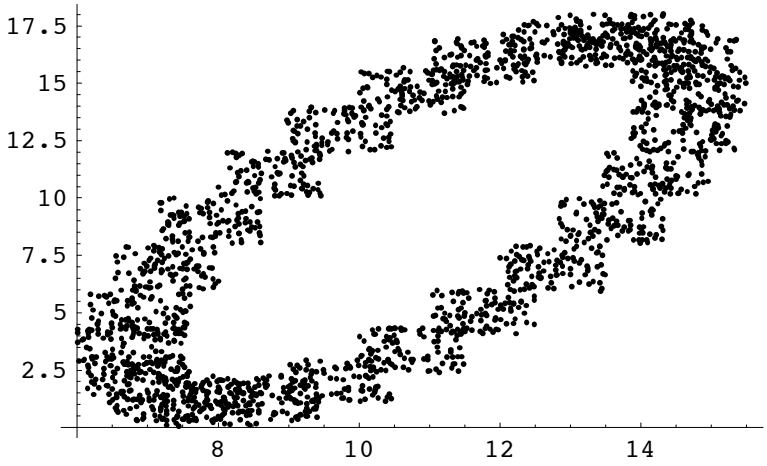
Consideremos un conjunto de datos d que viene dado por una lista de vectores de $\mathbf{R}^n$. Por ejemplo, para n=2, el conjunto de datos ad puede ser un fichero con 2232 vectores, de los cuales los cuatros cuatro primeros son los siguientes:

ad={{11.2858,5.49001},{10.4903,3.36277},{9.91928,2.41352},{8.55568,2.12325}, ...}

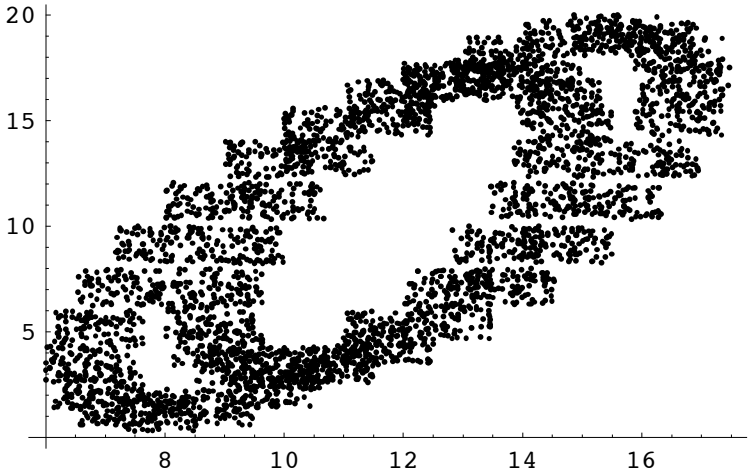
En primer lugar, notemos que cada vector (dato) es una 2-tupla de "números reales" en este caso con varias cifras decimales. El número de cifras decimales puede estar relacionado con la sensibilidad de los sensores que hayan generado esta lista de datos. La lista que nos han dado tiene además un orden natural, el primer vector dado, el segundo, etc. Quizás el primer dato se generó en el instante t=1, el segundo para t=2, etc. En el estudio que vamos a realizar no vamos a tener en cuenta este orden, de modo que si permutamos los dos primeros vectores, seguimos teniendo en mismo conjunto de datos.

Vamos a considerar cuatro conjuntos de datos, de los que de momento no conocemos su procedencia y a pesar de ello vamos a intentar descubrir los secretos que puedan contener. La lista ad tiene 2232 datos, la bd tiene 4392, la cd 6552 y la lista dd tiene 2232 datos. Como estamos en dimensión dos, una primera impresión se puede extraer representando gráficamente los cuatro conjuntos de datos.

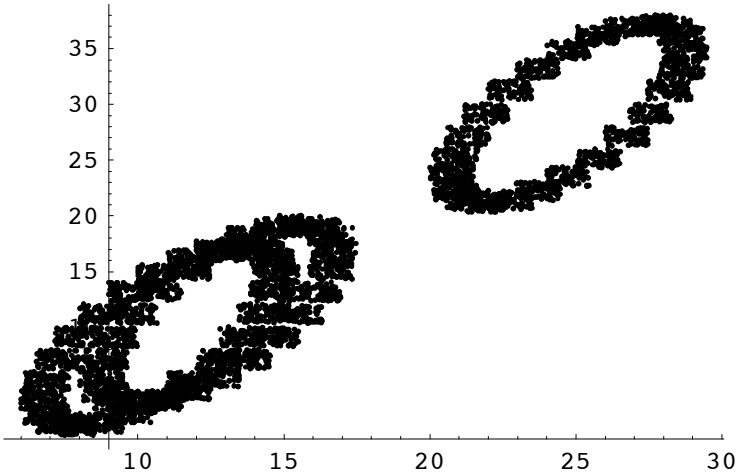
Gráfica del conjunto de datos ad



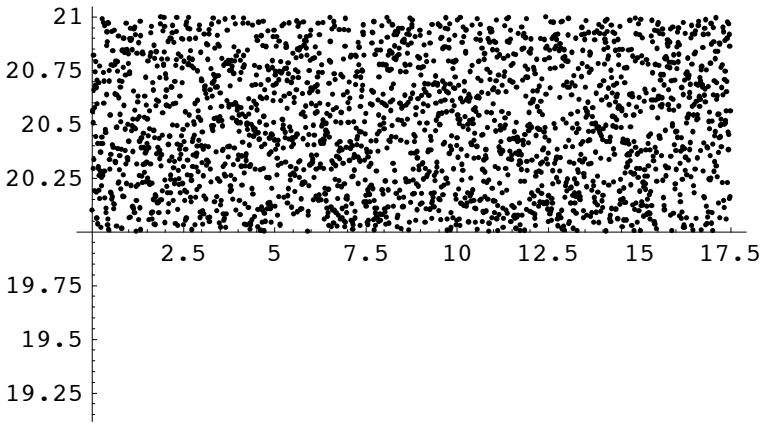
Gráfica del conjunto de datos bd



Gráfica del conjunto de datos cd



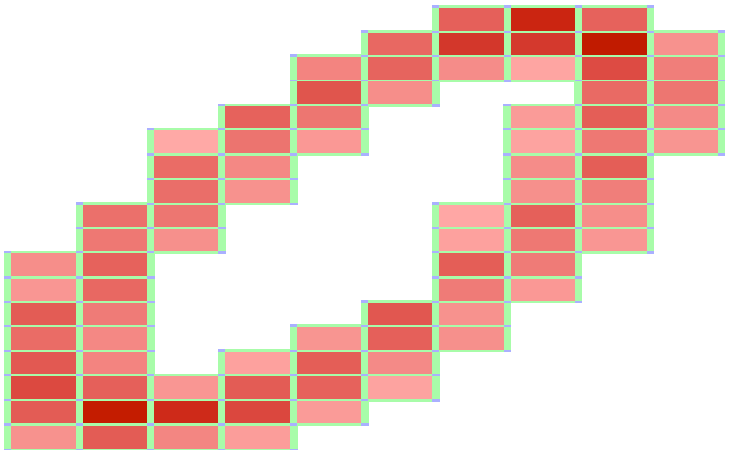
Gráfica del conjunto de datos dd



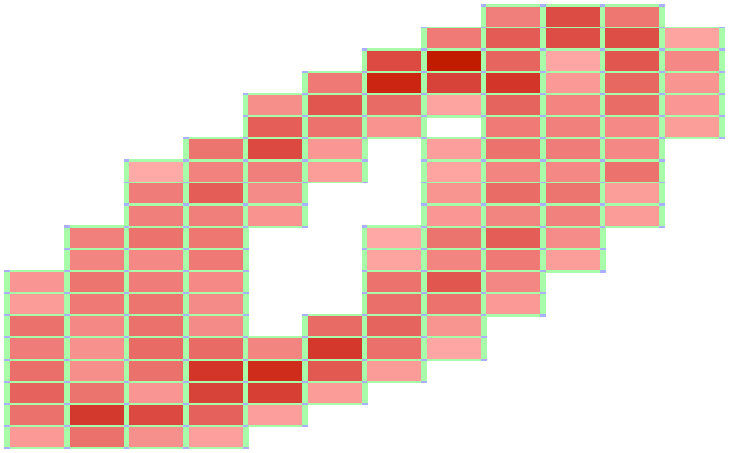
Aplicaciones densidad cúbicas

En esta subsección describimos un método para realizar el estudio de estos conjuntos de datos. Se trata de aproximar el conjunto de datos a través de un objeto matemático estructurado de modo que esté determinado con una codificación que con menos contenido preserve la información más relevante que incluyeran estos datos, e incluso que "prácticamente" se puedan recuperar los datos iniciales. El modelo matemático que proponemos verifica alguna de estas propiedades y por supuesto olvida otras que para algunos interesados en estos conjuntos de datos considerarán imprescindibles. El modelo que proponemos lo vamos a llamar aplicación densidad cúbica, consiste en una aplicación de un complejo cúbico en un anillo unitario conmutativo. Necesitaremos definir la noción de complejo cúbico y en este caso vamos a tomar en anillo de los enteros. Veamos primero complejos cúbicos asociados a los conjuntos de datos anteriores y después daremos su definición:

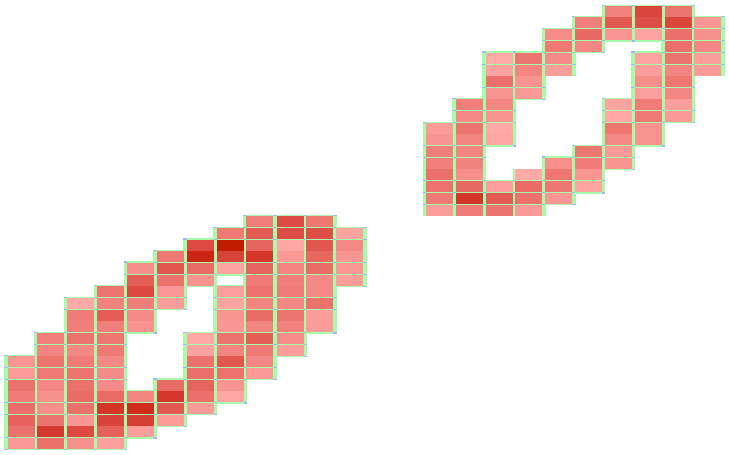
El complejo cúbico asociado al conjunto ad



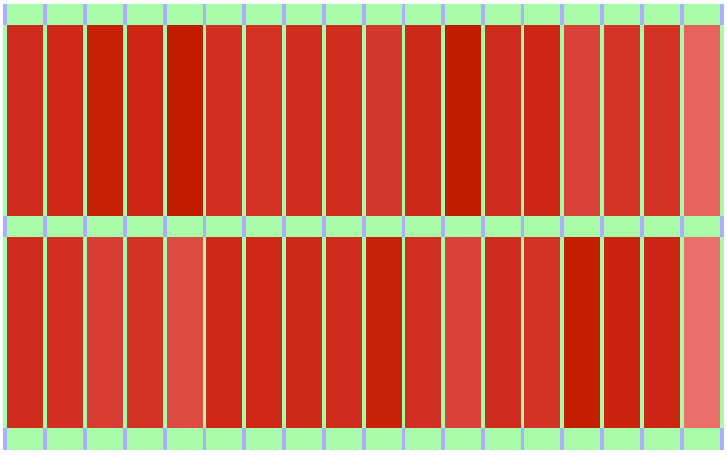
El complejo cúbico asociado al conjunto bd



El complejo cúbico asociado al conjunto cd



El complejo cúbico asociado al conjunto dd



En la figuras anteriores observamos una serie de rectángulos con diversos colores y diferente intensidad. Representan cuatro complejos cúbicos. Un complejo cúbico en $\mathbf{R}^2$ es un conjunto de p-cubos donde p denotará la dimensión del cubos, en este caso p puede ser 0, 1 ó 2.

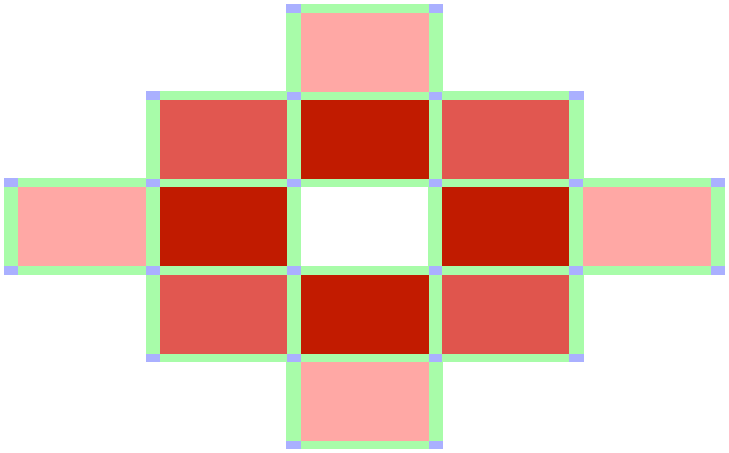
En general podemos codificar un *p-cubo* en $\mathbf{R}^n$ como una pareja $\{l,v\}$, donde $l=\{l_1, ..., l_n\}$ es un *cuboide direccionado* y cada l_i toma como valores 0 ó 1; v, *el vértice principal de p-cubo*, es un "vector" de $\mathbf{Z}^n$; es decir, que cada una de sus coordenadas es un entero, además p es igual a la suma de los l_i desde $i=1$ hasta $i=n$. Dado un p-cubo $\{l,v\}$ con $l_i=1$, diremos que sus caras en la dirección i son los (p-1)-cubos $\{\{l_1, ..., l'_i, ..., l_n\},\{v_1,..., v_i, ..., v_n\}\}$ y $\{\{l_1, ..., l'_i, ..., l_n\},\{v_1,..., v_i+1, ..., v_n\}\}$, donce $l'_i=0$. Puesto que $\mathbf{Z}/2$ tiene dos elementos (el 0 y el 1) entonces la estructura cúbica que estamos considerando en $\mathbf{R}^n$ es precisamente $(\mathbf{Z}/2)^n \times \mathbf{Z}^n$ que denotaremos por $\mathbf{E}^n$ (la **E** viene de Euclides).

Definición. Un *complejo cúbico* **K** de $\mathbf{E}^n$ es un conjunto de cubos de $\mathbf{E}^n$ tal que cada cubo del complejo se verifica que también contiene a sus caras. Para cada entero p, denotaremos por $\mathbf{K}^p$ el subcomplejo formado por los i-cubos con dimensión i menor o igual que p y por $\mathbf{K}_p$ el conjunto de los p-cubos de **K** .

Ejemplo: El siguiente conjunto de cubos es un complejo cúbico en $\mathbf{E}^2$

```
K = {{{0, 0}, {-2, 0}}, {{0, 0}, {-2, 1}}, {{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}},
    {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}}, {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}},
    {{0, 0}, {0, 0}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}}, {{0, 0}, {0, 3}}, {{0, 0}, {1, -2}},
    {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}}, {{0, 0}, {1, 2}}, {{0, 0}, {1, 3}},
    {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}}, {{0, 0}, {2, 2}}, {{0, 0}, {3, 0}},
    {{0, 0}, {3, 1}}, {{0, 1}, {-2, 0}}, {{0, 1}, {-1, -1}}, {{0, 1}, {-1, 0}},
    {{0, 1}, {-1, 1}}, {{0, 1}, {0, -2}}, {{0, 1}, {0, -1}}, {{0, 1}, {0, 0}},
    {{0, 1}, {0, 1}}, {{0, 1}, {0, 2}}, {{0, 1}, {1, -2}}, {{0, 1}, {1, -1}}, {{0, 1}, {1, 0}},
    {{0, 1}, {1, 1}}, {{0, 1}, {1, 2}}, {{0, 1}, {2, -1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 1}},
    {{0, 1}, {3, 0}}, {{1, 0}, {-2, 0}}, {{1, 0}, {-2, 1}}, {{1, 0}, {-1, -1}},
    {{1, 0}, {-1, 0}}, {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 2}}, {{1, 0}, {0, -2}},
    {{1, 0}, {0, -1}}, {{1, 0}, {0, 0}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}}, {{1, 0}, {0, 3}},
    {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}}, {{1, 0}, {1, 2}}, {{1, 0}, {2, 0}},
    {{1, 0}, {2, 1}}, {{1, 1}, {-2, 0}}, {{1, 1}, {-1, -1}}, {{1, 1}, {-1, 0}},
    {{1, 1}, {-1, 1}}, {{1, 1}, {0, -2}}, {{1, 1}, {0, -1}}, {{1, 1}, {0, 1}},
    {{1, 1}, {0, 2}}, {{1, 1}, {1, -1}}, {{1, 1}, {1, 0}}, {{1, 1}, {1, 1}}, {{1, 1}, {2, 0}}}}
```

La representación gráfica del complejo anterior es la siguiente:

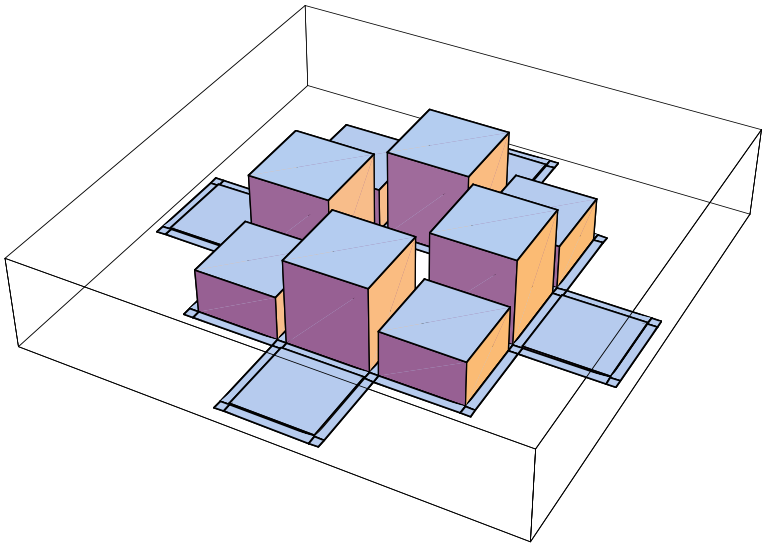


Definición. Un *aplicacion densidad cúbica* con valores en un anillo R es una aplicación f:**K**->R , donde **K** es un conjunto cúbico de $\mathbf{E}^n$ y R es un anillo conmutativo con unidad.

Ejemplo. El primer miembro de la siguiente lista es un complejo cúbico de $\mathbf{E}^2$, el segundo es asocia un entero a cada cubo del complejo. La siguiente densidad cúbica tiene la peculiaridad que sólo asocia enteros no nulos a los 2-cubos.

```
{{{{0, 0}, {-2, 0}}, {{0, 0}, {-2, 1}}, {{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}},
  {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}}, {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}},
  {{0, 0}, {0, 0}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}}, {{0, 0}, {0, 3}}, {{0, 0}, {1, -2}},
  {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}}, {{0, 0}, {1, 2}}, {{0, 0}, {1, 3}},
  {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}}, {{0, 0}, {2, 2}}, {{0, 0}, {3, 0}},
  {{0, 0}, {3, 1}}, {{0, 1}, {-2, 0}}, {{0, 1}, {-1, -1}}, {{0, 1}, {-1, 0}},
  {{0, 1}, {-1, 1}}, {{0, 1}, {0, -2}}, {{0, 1}, {0, -1}}, {{0, 1}, {0, 0}},
  {{0, 1}, {0, 1}}, {{0, 1}, {0, 2}}, {{0, 1}, {1, -2}}, {{0, 1}, {1, -1}}, {{0, 1}, {1, 0}},
  {{0, 1}, {1, 1}}, {{0, 1}, {1, 2}}, {{0, 1}, {2, -1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 1}},
  {{0, 1}, {3, 0}}, {{1, 0}, {-2, 0}}, {{1, 0}, {-2, 1}}, {{1, 0}, {-1, -1}},
  {{1, 0}, {-1, 0}}, {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 2}}, {{1, 0}, {0, -2}},
  {{1, 0}, {0, -1}}, {{1, 0}, {0, 0}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}}, {{1, 0}, {0, 3}},
  {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}}, {{1, 0}, {1, 2}}, {{1, 0}, {2, 0}},
  {{1, 0}, {2, 1}}, {{1, 1}, {-2, 0}}, {{1, 1}, {-1, -1}}, {{1, 1}, {-1, 0}},
  {{1, 1}, {-1, 1}}, {{1, 1}, {0, -2}}, {{1, 1}, {0, -1}}, {{1, 1}, {0, 1}},
  {{1, 1}, {0, 2}}, {{1, 1}, {1, -1}}, {{1, 1}, {1, 0}}, {{1, 1}, {1, 1}}, {{1, 1}, {2, 0}}},
 {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 30, 59, 30, 1, 58, 59, 1, 31, 59, 30, 1}}
```

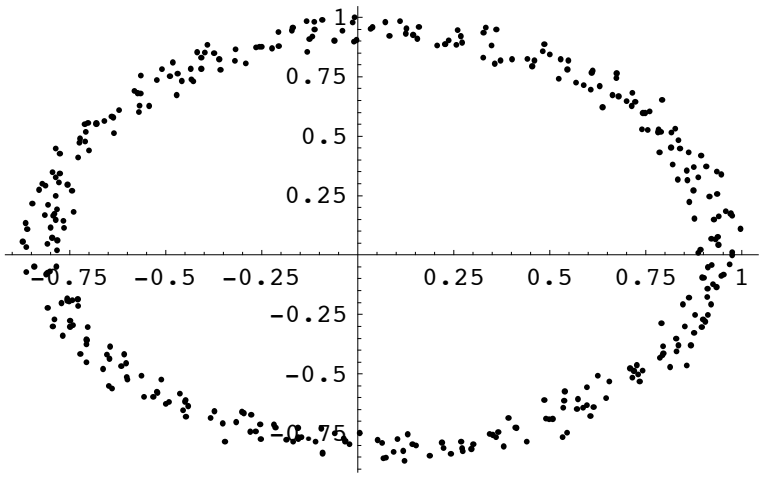
En las representaciones gráficas anteriores, se ha asignado un color a los 0-cubos, otro a los 1-cubos y otro a los 2-cubos, además el color es más intenso si densidad es más elevada. Damos a continuación otra representación gráfica de la densidad cúbica del ejemplo anterior en dimensión tres, en este caso la altura es proporcional a la densidad. Si bien a los cubos con densidad cero se les ha asignado una altura mínima para señalar su presencia.



Conjuntos de datos y aplicaciones densidad cúbicas

Existen varias construcciones que asocian a un conjuntos de datos una aplicación densidad y viceversa. A continuación vamos a describir tres métodos. Aunque lo más probable es que para el estudio de determinadas familias de conjuntos de datos haya que diseñar nuevos métodos que posiblemente sean modificaciones de alguno de los tres que señalamos a continuación. Consideremos el siguiente conjunto de datos:

Conjunto de datos coronad



Le vamos a asociar tres densidades por tres procedimientos distintos:

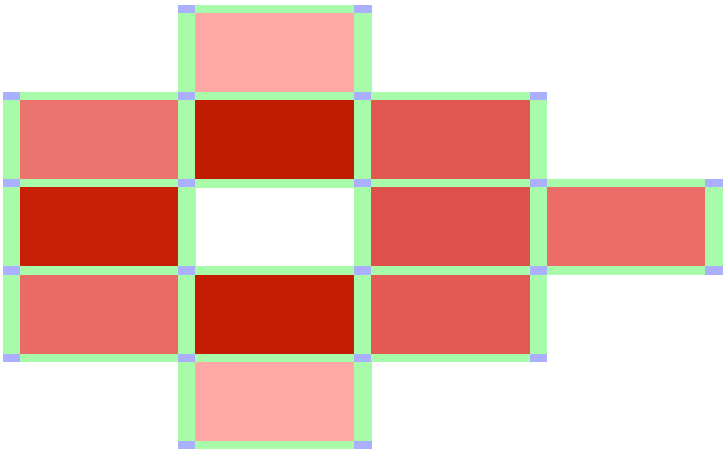
Construcción cero. Llamaremos *escala* $e=\{1/\lambda_1 , \dots , 1/\lambda_n \}$ a una n-tupla de números reales positivos. Vamos a considerar el n-cubo "semiabierto" de la forma $c_0{}^n(e)= [0,\lambda_1) \times [0,\lambda_2) \times \dots \times [0,\lambda_n)$, donde $[a,b)$ denota el intervalo semiabierto

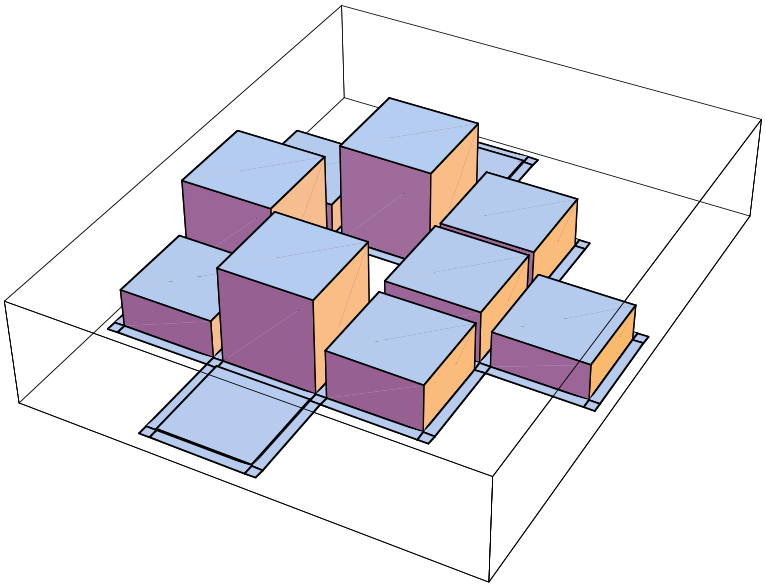
$\{r|$ a menor o igual que r y r menor que $b\}$. Si consideramos todas las traslaciones asociadas a vectores de la forma $v(e)=\{z_1 \lambda_1 , z_2 \lambda_2 , ...z_n \lambda_n \}$, donde $v=\{z_1 , ... ,z_n \}$ con cada z_i entero, se obtiene una teselación de $\mathbf{R}^n$ como una reunión disjunta de los n-cubos trasladados del asociado, $c_0{}^n(e)$, a la escala e dada. La construcción cero, asocia para cada escala e y conjunto de datos d, una densidad cúbica que tiene como dominio el complejo cúbico cuyos n-cubos $\{l,v\}$ verifican que las coordenadas de l son todas unos y $v=\{z_1 , ... ,z_n \}$ es tal que la intersección de d y $v(e)+ c_0{}^n(e)$ es no vacía. Para que efectivamente sea un complejo cúbico se añaden iteradamente las caras necesarias de baja dimensión. La aplicación densidad asocia a cada n-cubo $\{l,v\}$ el cardinal de $(v(e)+ c_0{}^n(e))\bigcap$ d que es mayor o igual que uno y a los p-cubos de dimensión más baja les asocia el cero. Vemos a continuación la densidad asociada al conjunto de datos coronad que corresponde a la figura anterior. En las figuras siguientes puede apreciarse la diferente intensidad de color y diferente altura asociada a los 2-cubos, nótese que la altura los prismas asociados a 1-cubos y 0-cubos debería ser cero, la hacemos un poco mayor para que al menos se puedan ver.

Densidad asociada por la construcción cero al conjuno de datos coronad

```
{{{{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}}, {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}},
  {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}}, {{0, 0}, {0, 0}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}},
  {{0, 0}, {0, 3}}, {{0, 0}, {1, -2}}, {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}},
  {{0, 0}, {1, 2}}, {{0, 0}, {1, 3}}, {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}},
  {{0, 0}, {2, 2}}, {{0, 0}, {3, 0}}, {{0, 0}, {3, 1}}, {{0, 1}, {-1, -1}}, {{0, 1}, {-1, 0}},
  {{0, 1}, {-1, 1}}, {{0, 1}, {0, -2}}, {{0, 1}, {0, -1}}, {{0, 1}, {0, 0}}, {{0, 1}, {0, 1}},
  {{0, 1}, {0, 2}}, {{0, 1}, {1, -2}}, {{0, 1}, {1, -1}}, {{0, 1}, {1, 0}}, {{0, 1}, {1, 1}},
  {{0, 1}, {1, 2}}, {{0, 1}, {2, -1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 1}}, {{0, 1}, {3, 0}},
  {{1, 0}, {-1, -1}}, {{1, 0}, {-1, 0}}, {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 2}},
  {{1, 0}, {0, -2}}, {{1, 0}, {0, -1}}, {{1, 0}, {0, 0}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}},
  {{1, 0}, {0, 3}}, {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}}, {{1, 0}, {1, 2}},
  {{1, 0}, {2, 0}}, {{1, 0}, {2, 1}}, {{1, 1}, {-1, -1}}, {{1, 1}, {-1, 0}},
  {{1, 1}, {-1, 1}}, {{1, 1}, {0, -2}}, {{1, 1}, {0, -1}}, {{1, 1}, {0, 1}},
  {{1, 1}, {0, 2}}, {{1, 1}, {1, -1}}, {{1, 1}, {1, 0}}, {{1, 1}, {1, 1}}, {{1, 1}, {2, 0}}},
 {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 25, 59, 22, 1, 63, 65, 1, 32, 35, 33, 24}}
```

La densidad del ejemplo anterior también puede apreciarse en la siguientes representaciones gráficas:



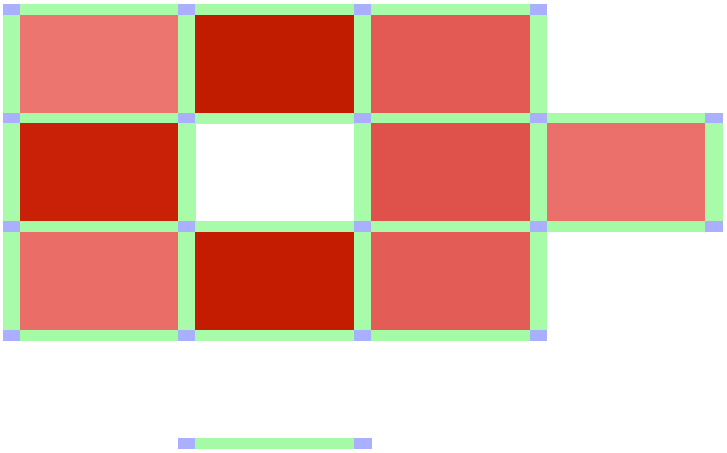


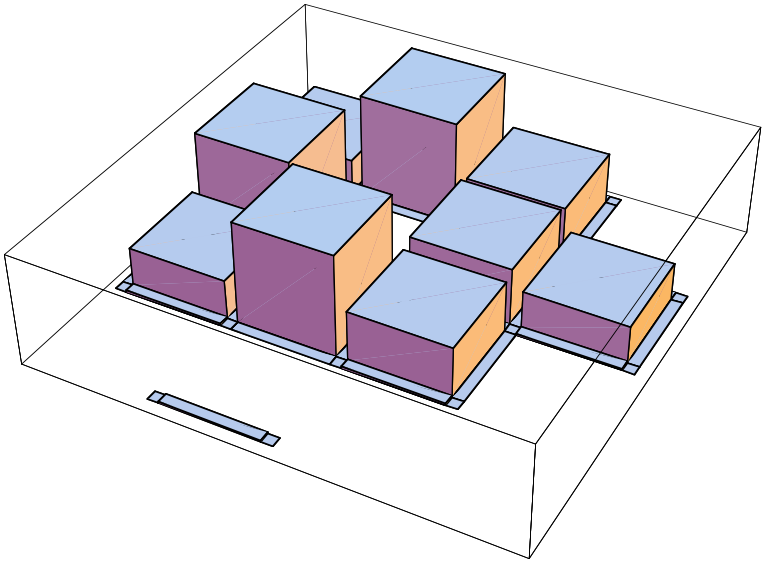
Construcción uno. Para *a, b* reales con *a* menor o igual que *b*, denotemos por int[*a,b*] el intervalo abierto (*a,b*) si *a<b* en el caso que *a=b*, int[*a,b*] es el conjunto atómico {*a*}. Dada una escala $e=\{1/\lambda_1 , \ldots , 1/\lambda_n\}$ y un cuboide direccionado *l* consideramos el subconjunto de la forma $c_1{}^n(e,l)=\text{int}[0,l_1\lambda_1] \times \text{int}[0,l_2\lambda_2] \times \ldots \times \text{int}[0,l_n\lambda_n]$. Si consideramos todas las traslaciones asociadas a vectores de la forma $v(e)=\{z_1\lambda_1 , z_2\lambda_2 , \ldots , z_n\lambda_n\}$, donde $v=\{z_1 , \ldots , z_n\}$ con cada z_i entero, de los conjuntos de la forma $c_1{}^n(e,l)$, se obtiene una teselación de $\mathbf{R}^n$ como una reunión disjunta de los conjuntos trasladados. La construcción uno, asocia para cada escala *e* y conjuntos de datos d, una densidad cúbica que tiene como dominio el complejo cúbico cuyos p-cubos {*l,v*} verifican que la intersección de d y $v(e)+c_1{}^n(e,l)$ es no vacía. Para que efectivamente sea un complejo cúbico se añaden iteradamente las caras necesarias de baja dimensión. La aplicación densidad asocia a cada cubo {*l,v*} en cardinal de $(v(e)+c_1{}^n(e,l))\bigcap$ d que es mayor o igual que uno y al resto de los cubos de dimensión más baja les asocia el cero. Esta construcción uno es la que ha sido utilizada en el ejemplo siguiente:

Densidad asociada por la construcción uno al conjuno de datos coronad

```
{{{{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}}, {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}},
  {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}}, {{0, 0}, {0, 0}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}},
  {{0, 0}, {1, -2}}, {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}}, {{0, 0}, {1, 2}},
  {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}}, {{0, 0}, {2, 2}}, {{0, 0}, {3, 0}},
  {{0, 0}, {3, 1}}, {{0, 1}, {-1, -1}}, {{0, 1}, {-1, 0}}, {{0, 1}, {-1, 1}},
  {{0, 1}, {0, -1}}, {{0, 1}, {0, 0}}, {{0, 1}, {0, 1}}, {{0, 1}, {1, -1}}, {{0, 1}, {1, 0}},
  {{0, 1}, {1, 1}}, {{0, 1}, {2, -1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 1}}, {{0, 1}, {3, 0}},
  {{1, 0}, {-1, -1}}, {{1, 0}, {-1, -1}}, {{1, 0}, {-1, 0}}, {{1, 0}, {-1, 0}},
  {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 2}}, {{1, 0}, {0, -2}},
  {{1, 0}, {0, -1}}, {{1, 0}, {0, 0}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}}, {{1, 0}, {0, 2}},
  {{1, 0}, {1, -1}}, {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}},
  {{1, 0}, {1, 1}}, {{1, 0}, {1, 2}}, {{1, 0}, {2, 0}}, {{1, 0}, {2, 0}}, {{1, 0}, {2, 1}},
  {{1, 1}, {-1, -1}}, {{1, 1}, {-1, 0}}, {{1, 1}, {-1, 1}}, {{1, 1}, {0, -1}},
  {{1, 1}, {0, 1}}, {{1, 1}, {1, -1}}, {{1, 1}, {1, 0}}, {{1, 1}, {1, 1}}, {{1, 1}, {2, 0}}},
 {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 1, 0, 1, 0, 0, 0, 0, 1,
  0, 1, 0, 0, 1, 0, 0, 1, 0, 24, 58, 21, 63, 65, 31, 35, 32, 23}}
```

La densidad del ejemplo anterior ya asocia a algunos 1-cubos un entero no nulo como también puede apreciarse en la siguientes representaciones gráficas:





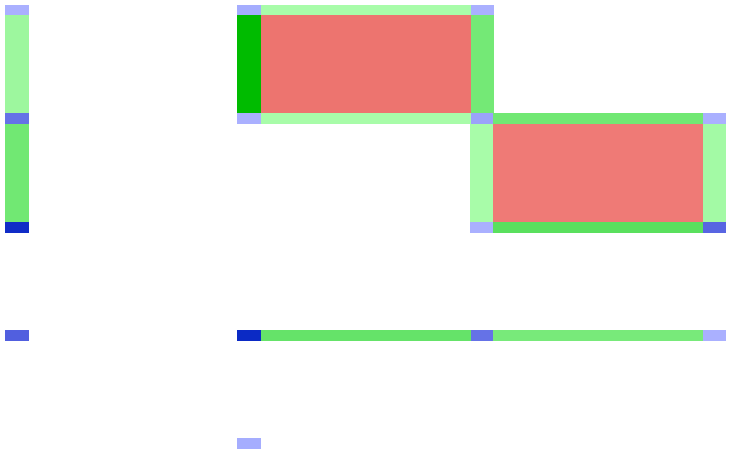
La ausencia completa de cubos de máxima dimensión puede deberse a que los datos originales estén en hiperplanos de dimensión más baja. Otra causa posible es una inadecuada elección de la escala e, de modo que la "sensibilidad" de la teselación asociada sea más fina que la de los sensores que generaron los datos.

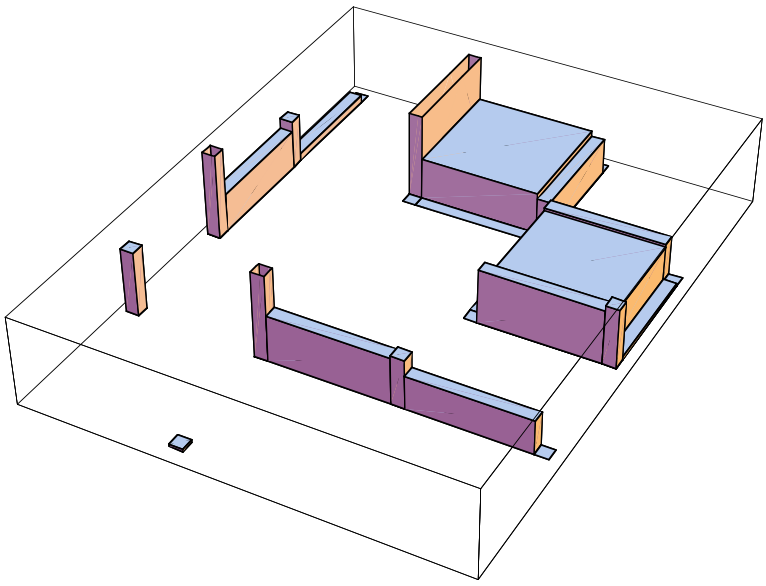
Construcción dos. Dada una escala $e=\{1/\lambda_1 , \ldots , 1/\lambda_n \}$ consideramos el subconjunto de la forma $c_2{}^n(e)= [0,\lambda_1 /2) \times [0,\lambda_2 /2) \times \ldots \times [0,\lambda_n /2)$. Para cada $\{l,v\}$ con $l=\{l_1 , \ldots ,l_n \}$ y $v=\{z_1 , \ldots , z_n \}$ consideramos todas las traslaciones asociadas a vectores de la forma $t(l,v)=\{ z_1 (1+(l_1 /2))\lambda_1 , z_2 (1+(l_2 /2))\lambda_2 , \ldots ,z_n (1+(l_n /2)) \lambda_n \}$, con cada z_i un entero, de los conjuntos de la forma $c_2{}^n(e)$, se obtiene así una teselación de $\mathbf{R}^n$ como una reunión disjunta de los conjuntos trasladados. La construcción dos, asocia para cada escala e y conjuntos de datos d, una densidad cúbica que tiene como dominio el complejo cúbico cuyos p-cubos $\{l,v\}$ verfican que la intersección de d y $t(l,v)+c_2{}^n(e)$ es no vacía. Para que efectivamente sea un complejo cúbico se añaden iteradamente las caras necesarias de baja dimensión. La aplicación densidad asocia a cada n-cubo {l,v} en cardinal de $(t(l,v)+c_2{}^n(e))\bigcap d$ que es mayor o igual que uno y al l resto de los p-cubos de dimensión más baja les asocia el cero. Esta construcción dos es la que ha sido utilizada en el ejemplo siguiente:

Densidad asociada por la construcción dos al conjuno de datos coronad

```
{{{{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}}, {{0, 0}, {-1, 0}}, {{0, 0}, {-1, 1}},
  {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}}, {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}},
  {{0, 0}, {0, -1}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}}, {{0, 0}, {0, 2}}, {{0, 0}, {1, -1}},
  {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}}, {{0, 0}, {1, 1}}, {{0, 0}, {1, 2}},
  {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}},
  {{0, 1}, {-1, 0}}, {{0, 1}, {-1, 1}}, {{0, 1}, {0, 1}}, {{0, 1}, {0, 1}},
  {{0, 1}, {1, 0}}, {{0, 1}, {1, 1}}, {{0, 1}, {1, 1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 0}},
  {{1, 0}, {0, -1}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}}, {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}},
  {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}}, {{1, 0}, {1, 1}}, {{1, 1}, {0, 1}}, {{1, 1}, {1, 0}}},
 {25, 0, 44, 0, 19, 0, 1, 0, 45, 0, 0, 1, 0, 19, 0, 0, 4, 0, 0, 0, 23, 0,
  15, 3, 0, 49, 0, 0, 14, 0, 1, 18, 0, 0, 13, 0, 21, 0, 15, 16, 14}}
```

La densidad del ejemplo anterior también puede apreciarse en las siguientes representaciones gráficas:





Existen otros numerosos métodos de construir densidades. Por ejemplo en vez de complejos cúbicos se pueden utilizar conjuntos de cubos que no tengan la necesidad de contener las caras generadas. En vez de cubos se puede teselar $\mathbf{R}^n$ por otro tipo de poliedro, en dimensión dos por hexágonos y triángulos. Con los criterios que hemos utilizado, cada dato se asigna a un cubo, este criterio se puede modificar asignado cada dato a los cubos "más próximos" de modo que las suma de las asignaciones realizadas sea igual a uno.

2 Información básica de conjuntos de datos asociada a sus aplicaciones densidad cúbicas

En esta sección mencionamos muy brevemente algunos aspectos que son de gran importancia en las técnicas de la minería de datos. Supondremos que hemos aproximado un conjunto de datos mediante la construcción cero, en consecuencia se supone que trabajamos con aplicaciones densidad cuyo soporte está contenido en $\mathbf{K}_n$.

Recuperación de los datos originales

Puede suceder que, por circunstancias tales como ación de un virus informático, nuestro preciado fichero de datos haya desaparecido, pero hayamos tenido la fortuna de conservar su densidad cúbica asociada. Si por ejemplo, habíamos utilizado la construcción cero, la forma más simple de "recuperar" los datos es la de rellenar de un modo aleatorio cada n-cubo con el número de datos que indique la densidad de dicho cubo. No obstante, tenemos el problema de que la escala utilizada en la construcción cero ha dilatado o contraído los datos originales para luego aproximarlos por el complejo cúbico. Entonces, para poder completar el proceso de recuperación debemos aplicar el proceso inverso de dilatación o contracción realizado y para ello debemos conocer la escala utilizada. Es por ello que vamos a utilizar la estructura de datos que llamaremos *densidad cúbica aumentada* que consiste en la pareja $\{f,e\}$ donde f es la densidad y e la escala utilizada.

Ejemplo. En la primera muestra de la subsección anterior se utilizó la escala {2,2} en la construcción cero, por lo que la densidad cúbica aumentada es la siguiente:

Densidad cúbica aumentada asociada por la construcción cero al conjunto de datos coronad

```
{{{{{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}}, {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}},
   {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}}, {{0, 0}, {0, 0}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}},
   {{0, 0}, {0, 3}}, {{0, 0}, {1, -2}}, {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}},
   {{0, 0}, {1, 2}}, {{0, 0}, {1, 3}}, {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}},
   {{0, 0}, {2, 2}}, {{0, 0}, {3, 0}}, {{0, 0}, {3, 1}}, {{0, 1}, {-1, -1}}, {{0, 1}, {-1, 0}},
   {{0, 1}, {-1, 1}}, {{0, 1}, {0, -2}}, {{0, 1}, {0, -1}}, {{0, 1}, {0, 0}}, {{0, 1}, {0, 1}},
   {{0, 1}, {0, 2}}, {{0, 1}, {1, -2}}, {{0, 1}, {1, -1}}, {{0, 1}, {1, 0}}, {{0, 1}, {1, 1}},
   {{0, 1}, {1, 2}}, {{0, 1}, {2, -1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 1}}, {{0, 1}, {3, 0}},
   {{1, 0}, {-1, -1}}, {{1, 0}, {-1, 0}}, {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 2}},
   {{1, 0}, {0, -2}}, {{1, 0}, {0, -1}}, {{1, 0}, {0, 0}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}},
   {{1, 0}, {0, 3}}, {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}}, {{1, 0}, {1, 2}},
   {{1, 0}, {2, 0}}, {{1, 0}, {2, 1}}, {{1, 1}, {-1, -1}}, {{1, 1}, {-1, 0}},
   {{1, 1}, {-1, 1}}, {{1, 1}, {0, -2}}, {{1, 1}, {0, -1}}, {{1, 1}, {0, 1}},
   {{1, 1}, {0, 2}}, {{1, 1}, {1, -1}}, {{1, 1}, {1, 0}}, {{1, 1}, {1, 1}}, {{1, 1}, {2, 0}}},
  {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
   0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
   0, 0, 0, 0, 0, 0, 25, 59, 22, 1, 63, 65, 1, 32, 35, 33, 24}}, {2, 2}}
```

Asociada a una escala e podemos definir la norma $|x|=\max\{ e_1 |x_1|, ... ,e_n |x_n|\}$. Notemos que si utilizamos la construcción cero, se tiene que dos conjuntos de datos d y d' tienen la misma densidad cúbica aumentada si existe una biyección b: d→ d'

tal que |x-b(x)| es menor que 1. Este criterio de clasificación en algunos casos puede ser suficiente, en otros dejará demasiadas clases de datos y será preciso preservar menos información para disminuir el número de clases.

Cúmulos de datos

Es bien conocido el interés que tiene el estudio aquellas regiones del espacio en donde ha habido una mayor acumulación de datos. Puesto que hemos asociado una densidad cúbica, en este caso el problema es equivalente a la elaboración de algoritmos para la búsqueda de aquellos cubos o mesetas maximales de la densidad. Si además de densidades cúbicas conocemos la escala se pueden afinar los algoritmos anteriores para eliminar aquellos datos que no estén contenidos en mesetas maximales.

Las particiones del complejo cúbico en sus componetes conexas se puede interpretar como una partición de datos en diferentes clases. Si hacemos que las coordenadas de la escala aumenten su valor se producen dilataciones de los datos (o equivalentemente, subdivisiónes de la teselación inicial) lo que determina una jerarquía de particiones en componentes conexas que al final degenera a una partición en la que cada dato es una componente conexa. Una zona intermedia en esta jerarquía puede ser la que interese conocer para ciertos conjuntos de datos, en otros casos puede ser la jerarquía completa la que contenga la información deseada.

En los párrafos siguientes veremos que el número de componentes conexas es precisamente un invariante homológico del complejo cúbico que se denomina *0-ésimo número de Betti*. En el caso que el complejo cúbico sea una pseudovariedad n-dimensional existe una partición más fina que la de las componentes conexas y es la que la divide en pseudocomponentes. Si utilizamos la construcción cero, se tiene que dos datos están en la misma pseudocomponente si existe una sucesión de n-cubos tal que el primer n-cubo está asociado al primer dato, cada pareja de n-cubos consecutivos se cortan exactamente en un (n-1)-cubo y el último cubo está determinado por el segundo dato. De nuevo un cambio de escalas produce una jerarquización de las particiones por pseudocomponentes. El número de pseudocomponentes coincide con el *n-ésimo número de Betti* de la cohomología de la pareja formada por la pseudovariedad y su borde. A su vez este coincide con el 0-ésimo número de Betti del interior de la pseudovariedad. En los siguientes párrafos se introducen más pausadamente nociones mencionadas como pseudovariedad, componente y pseudocomponente.

Sin embargo puede que sólo nos interese encontrar aquellas zonas en las que hay una mayor densidad de datos. Éstas corresponden con las mesetas maximales absolutas o relativas cuyo estudio puede ser abordado mediante el estudio de ciertas derivadas direccionales de la aplicación densidad.

El método de las aplicaciones densidad provee herramientas adecuadas para la partición en clases y búsqueda de regiones de alta densidad, que como ya hemos observado están relacionadas con algunos números de Betti. El diseño de algoritmos que realicen las tareas mencionadas no es complicado y su implementación puede realizarse en cualquiera de los lenguajes de programación usuales.

Datos espurios

Entendemos por tales aquellos que se han podido incluir en los ficheros de datos originales pero de modo no estándar. Errores de los sensores, condiciones ambientales inusuales, etc. Es esperable que su proporción sea escasa en comparación con el total, o por lo menos trabajaremos bajo esta hipótesis. Sin otra información adicional podemos suponer que se corresponden con aquellos cubos o mesetas en los que la densidad alcanza valores minimales. El estudio de datos espurios puede abordarse con diferentes herramientas, una de ellas puede ser el análisis de mesetas minimales de aplicaciones densidad que puede ser completada con técnicas como la distancia a otras componentes, a zonas de densidad maximal u otros criterios para su detección. Por lo tanto consideramos también interesante el diseño de algoritmos para la búsqueda de zonas (subcomplejos) en los que la densidad sea minimal, y por supuesto la selección de aquellos datos que se encuentren en estos subcomplejos con densidad minimal.

Puertos de la gráfica de la aplicación densidad

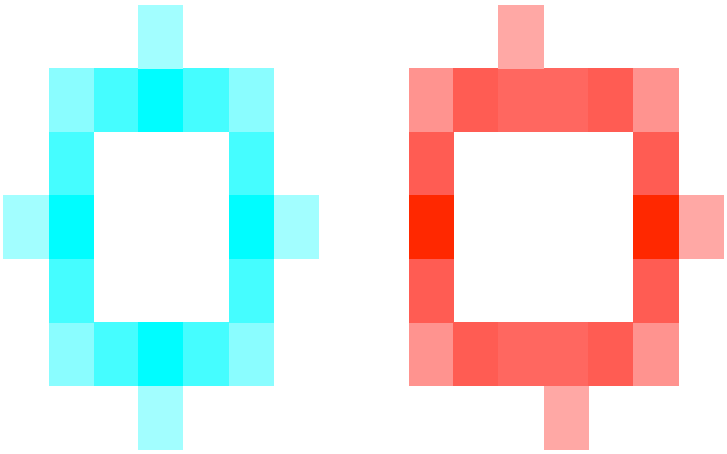
Se trata del "resto" del complejo cúbico es decir el "complementario" de las mesetas maximales y minimales. En este trabajo generamos herramientas para su estudio como es la noción de flexión direccional que permite un estudio más intensivo de estas regiones sobre las que la gráfica de la densidad cambia más intensamente sus valores, pudiendo crecer localmente en unas direcciones, decrecer localmente en otras y ser localmente constante en las direcciones restantes.

3 Invariantes homológicos del complejo inicial de una aplicación densidad de un conjuntos de datos

En primer lugar estudiaremos únicamente los invariantes asociados al dominio de la aplicación densidad que tiene una estructura natural de complejo cúbico.

El 0-ésimo número de Betti y Componentes conexas por caminos

Empezando por un área local pequeña que tenga cierta densidad, a continuación se puede emprender una expansión mientras se mantenga continuamente la existencia mínima de datos hasta que la ausencia de éstos fuera de una región maximal impida expandir más la región expandida. Cada una de estas regiones es una componente del conjunto de datos o si ya está representada por una densidad cúbica es una componente conexa por caminos del conjunto inicial de la aplicación densidad. El número de estas componentes es el *0-ésimo número de Betti*. En la siguiente figura representamos un complejo con dos componentes conexas por caminos



Son de utilidad aquellos algoritmos que permitan

1) Conocer el número de componentes del conjunto inicial de una densidad cúbica.

2) Determinar la importancia relativa de cada componente respecto el total de datos (volumen determinado por la densidad en la componente).

3) Explicitación de los cubos que forman cada componente.

4) Extracción de los datos que están en una componente.

De toda la información anterior la dada en el apartado 1) es un invariante homotópico. Esta información ocupa poco lugar, se trata de un entero no negativo y en general es más estable que la información sugerida en los apartados 2) 3) y 4).

Definición. Sea **C** un complejo cúbico. Diremos que **C** es *conexo por caminos* si dados dos 0-cubos V , V' existen un número finito de 1-cubos α_0 , ... , α_k tales que el el vértice inicial de α_0 es V , el vertice final de α_k es V' y para $k \geq i > 0$ el vérice final de α_{i-1} coincide con el inicial de α_i . Si **C** es un subcomplejo de **K** diremos que es una componente conexa por caminos si **C** es conexa por caminos y si $\mathbf{C} \subseteq \mathbf{C}'$ y $\mathbf{C}'$ es conexa por caminos, implica que $\mathbf{C}=\mathbf{C}'$.

Un complejo cúbico **K** siempre admite una partición canónica en subcomplejos conexos por caminos. Como ya lo hemos mencionado el *0-ésimo número de Betti* coincide con el número de componentes conexas por caminos de **K**. Normalmente para reducir el tamaño de las expresiones llamaremos componente a una componete conexa por caminos.

El 1-ésimo número de Betti y otros invariantes homológicos

El siguiente invariante homológico es el 1-ésimo número de Betti, se puede dar una primera explicación de su significado diciendo que es nulo cuando todo ciclo de complejo cúbico (una familia finta de curvas cerradas) es borde de una superficie con singularidades contenida en el complejo cúbico. En complejos cúbicos del plano como los que aparecen en la sección uno, la interpretación es un poco más sencilla, coincide con el número que se obtiene a restar uno al número de componentes del complemento de complejo cúbico en el plano. Vamos a convenir en llamarlo el número de "agujeros" que tiene tienen este tipo de complejos cúbicos 2-dimensionales. En el ejemplo anterior se observa que el complementario de la figura en el plano tiene tres componentes conexas, ello significa que el complejo cúbico tiene dos "dos agujeros", en consecuencia el 1-ésimo número de Betti del complejo dado en la figura anterior es dos.

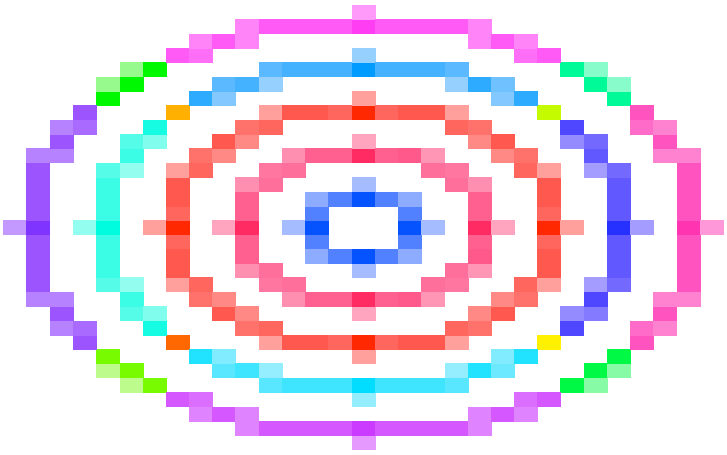
La interpretación geométrica del resto de los números de Betti y de torsión requiere de nociones como pseudovariedades y pseudovariedades con borde o complejos de cadenas, así que para su explicación detallada referimos al lector a un curso elemental de Topología Algebraica, véase [D], [G], [M] ó [Mu] .

Componentes pseudoconexas de complejos cúbicos puros

Un *complejo cúbico n-puro* consiste en un complejo cúbico tal que cualquiera de sus p-cubos tiene la propiedad de incidir con al menos un n-cubo. Un complejo cúbico n-puro decimos que es una *pseudovariedad* si cada (n-1)-cubo incide a lo sumo con dos n-cubos (notemos que por ser n-puro cada (n-1)-cubo al menos es cara de un n-cubo). Se llama *borde* de una pseudovariedad al subcomplejo generado por aquellos (n-1)-cubos que únicamente inciden con un n-cubo (con el sistema de definiciones dado aquí el borde de una pseudovariedad en general no es una pseudovariedad). Una pseudovariedad es *pseudoconexa* si además verifica que dados dos n-cubos C y C', existe una cadena de n-cubos C_0 , C_1 , ... , C_n tal que C_0 =C y C_n =C' y de modo que para $i=1$, ..., n, C_{i-1} corta a C_i en un (n-1)-cubo. En una pseudovariedad no conexa, siempre se puede decir que dos n-cubos están relacionados si existe una cadena del uno al otro del tipo anterior. Esta relación divide a la pseudovariedad (con borde) en una reunión de pseudovariedades pseudoconexas(con borde) que llamaremos *pseudocomponentes* conexas de la pseudovariedad. Notemos que dos pseudocomponentes se cortan en un subcomplejo (posiblemente vacío) de dimensión menor o igual que (n-2).

Puesto que una pseudocomponente es siempre conexa se tiene que siempre esta contenida en una componente conexa.

En el siguiente ejemplo vemos como el complejo cúbico se ha dividido en componentes conexas y a su vez cada componente se ha dividido en pseudocomponentes. Hemos coloreado cada pseudocomponente con un color distinto, si en color de un n-cubo es más intenso denota que su densidad es mayor.



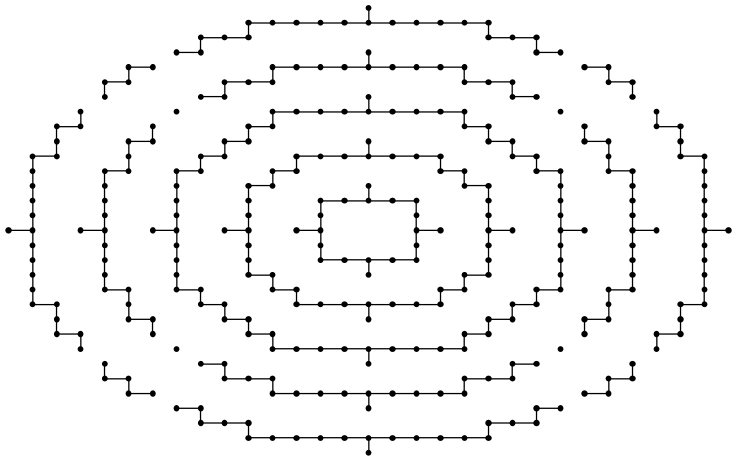
Notemos que las tres componentes más internas son pseudocomponentes. Sin embargo cada una de las dos componentes más externas se divide a su vez en ocho pseudocomponentes.

El 0-ésimo número de Betti del complejo cúbico anterior es cinco y en este caso también es cinco el 1-ésimo número de Betti. Este complejo tiene 19 pseudocomponentes.

El dual de una pseudovariedad con borde

La siguiente construcción puede ser de utilidad, se trata de calcular del dual de una pseudovariedad (con borde). Más generalmente el dual C^d de un p-cubo C en el n-espacio cúbico se define como el q-cubo ortogonal a C a través de su "centro de gravedad". Dado un p-cubo $\{z,l\}$ con $\sum_{i=1}^{n} l_i = p$ si definimos tilde$\{l\}=u-l$ donde $u=\{1,...,1\}$~. Entonces C^d $=\{z+(l/2)$(l-tilde$\{l\}\}$, tilde$\{l\}\}$. Una pseudovariedad **P** con borde **B** , determina el conjunto de cubos **P\ B** , que tiene la siguiente propiedad: Sean C y C' cubos tales que C es cara de C', si C esta en **P\B**, entonces C' está en **P\ B**. También es fácil comprobar que si C es cara de C', entonces el dual de C' es cara del dual de C. De aquí se sigue que el conjunto de los duales de los cubos de una pseudovariedad menos su borde es un complejo cúbico; es decir que si un cubo está en el complejo también lo están sus caras.

Llamaremos interior de una pseudovariedad a la reunión de los cubos que no están contenidos en el borde. Las pseudovariedades tienen la siguiente interesante propiedad:. El interior de una pseudovariedad tiene el mismo tipo de homotopía que su dual. En la siguiente figura hemos calculado el dual del complejo cúbico de la figura anterior



Señalaremos que los números de Betti del dual coinciden con los interior del la pseudovariedad (con borde) ya que tienen el mismo tipo de homotopía. En este caso $\beta_0 =19$ y $\beta_1 =3$.

Clasificación homológica de datos

Para cada elección de escala y del tipo de construcción, digamos P, que asocia un complejo cúbico a un conjunto de datos, se dispone del siguiente criterio de clasificación: Se dice que dos conjuntos de datos son homológicamente P-equivalentes si los complejos cúbicos asociados tienen la "misma homología" o equivalentemente si tienen los mismos números de Betti y de torsión. Por ejemplo para n=2 escala e={1/3, 1/3}podemos clasificar muestras con los siguientes procedimientos:

Criterio 1. Tomemos la que hemos denominado construcción cero P_0 . Puesto que los subcomplejos 2-dimensionales de la estructura cúbica de $\mathbf{E}^2$ no tienen torsión se tiene que dos conjuntos de datos 2-dimensionales serán P_0 -equivalentes si y sólo si los q-ésimos números de Betti de los complejos cúbicos asociados coinciden para q=0 y q=1 o equivalentemente si y sólo si tienen el mismo número de componentes y característica de Euler.

Criterio 2. Asociemos a cada conjunto de datos 2-dimensional la familia de componentes conexas del complejo cúbico del ejemplo anterior y llamemos P_1 a esta construcción . Puesto que cada componente conexa aporta una unidad en el β_0 , ponemos analizar el β_1 de cada componente, se este modo asociaremos a cada conjunto de datos 2-dimensional un un conjunto finito formado por los β_1 de cada componente. Decimos que dos conjuntos de datos son P_1 -equivalentes si y sólo si tienen los mismos conjuntos finitos asociados.

Ventajas: 1) La clasificación anterior genera una cantidad contable de clases de equivalencia posibles.

2) La homología de complejos cúbicos finitos se puede calcular mediante algunos algoritmos.

Inconvenientes: 1) Es necesario realizar una adecuada elección de escala.

2) En dimensiones altas el tiempo necesario para calcular la homología es elevado.

3) Puede haber dos funciones densidad completamente distinstas que tengan el mismo complejo cúbico como conjunto inicial y aún siendo muy distinta la distribución interna de datos generan los mismos invariantes homológicos.

Algoritmo clásico para el cálculo de la homología

Dado un complejo de cadenas que sea libre sobre un conjunto graduado finito. Si la dimensión del complejo es n, se calculan las matrices de incidencia que tendrán coeficientes enteros, en el caso de complejos simpliciales o cúbicos serán 1,-1 ó 0. Después se diagonolizan estas matrices y basta con examinar la matriz de incidencia de las dimensiones (q+1) y q y la de dimensiones q y (q-1) para poder calcular la homología con coeficientes enteros, en dimensión q. Puesto que es isomorfa a la suma finita de cíclicos, para cada dimensión el correspondiente q-ésimo grupo de homología queda determinado por el número de cíclicos infinito, β_q , que se llama q-ésimo número de Betti y por las torsiones de los cíclicos finitos. Las antiguas ediciones del libro de Seifert y Threlfall [ST] ya describían este algoritmo.

En la literatura es frecuente el referirse al algoritmo anterior como la diagonalización de Smith; son también numerosos los trabajos existentes sobre algoritmos y su complejidad sobre la diagonalización de Smith.

Hay que tener en cuenta que se pueden utilizar diversos trabajos existentes sobre el cálculo de homología de poliedros. En concreto el algoritmo incremental de Delfinado y Edelsbrunner [DE] y el de Friedman [F] ambos para el cálculo de los números de Betti.

Existen también algunos paquetes que facilitan el cómputo de los grupos de homología. Citemos por ejemplo, el paquete GAP (Grupos, Algoritmos y Programación), véase [DHSW1, DHSW2], y el paquete KENZO desarrollado por J. Rubio y F. Sergeraert [RS] .

4 Interpretación causal de los números de Betti y de torsión

Pretendemos clarificar dos cuestiones: La primera es la utilidad de los invariantes homológicos anteriores como una herramienta de clasificación y la segunda es la de mostrar que los cambios de valor de los números de Betti y de torsión son causados por cambios en los factores principales que regulan el sistema que ha producido estos datos. En general pequeños cambios en el conjunto de datos no altera los valores de los invariantes homológicos, salvo el caso que estos pequeños cambios se realicen cerca de los valores críticos, lo cual refuerza de nuevo la importancia y capacidad sensitiva de estos invariantes. Existen casos más complejos en los que los procesos se caracterizan por tener estos invariantes de Betti y de torsión en determinados rangos. En algunos de estos casos las alteraciones del sistema que genera estos datos se caracterizan más que por el cambio de los valores de los invariantes por el hecho de que se salgan de sus rangos estándar.

En relación con los valores de los números de Betti de los primeros ejemplos de este trabajo señalaremos las siguientes observaciones sobre la causa que produjo tales cambios.

El complejo cubico de primer conjunto de datos ad tiene como números de Betti, $\beta_0 =1$ y $\beta_1 =1$. Topológicamente significa que tiene una componente conexa y que su complementario en el plano tiene dos componentes conexas (es decir, que "tiene un agujero"). En ejemplo bd se obtienen una componente conexa $\beta_0 =1$ y su complemento en el plano tiene tres componentes conexas (dos "agujeros") lo que se refleja en el hecho que $\beta_1 =2$. El ejemplo bc tiene dos componentes conexas y el complementario en el plano tiene tres componentes conexas este hecho queda reflejado en los valores $\beta_0 =2$ y $\beta_1 =2$. En el último ejemplo se tiene que hay una componente conexa y el complementario tiene una componente conexa (sin "agujeros") lo queda expresado con los valores $\beta_0 =1$, $\beta_1 =0$. Los ejemplos cúbicos anteriores no han generado torsión, por lo que el conjunto de invariantes de torsión es vacío.

Intrigados por los cambios que se había en los números de Betti, preguntamos sobre la procedencia de los datos analizados y sus peculiaridades. Se nos dijo lo siguiente:

El conjunto de datos a corresponden a los valores de temperatura en grados centígrados de un almacén , la primera variable es la temperatura en el interior del almacen y la segunda la temperatura en el exterior. Los datos provienen de tres termómetros, el que está en el exterior queda reflejado en la segunda variable y los del interior en la primera variable. En los ejemplos ad, bd, cd corresponden al termómetro exterior y del almacén; en el último caso, dd, corresponde al exterior y a la oficina del vigilante que dispone de calefacción con termostato.

El sistema de sensores realiza una medición cada 20 minutos y los datos corresponden a lo siguiente:

ad) Mediciones (interior, exterior) tomadas durante el mes de Enero.

bd) Se han juntado los datos de las mediciones de los meses de Enero y Abril.

cd) Corresponden a la reunión de los datos de los meses de Enero, Abril y Agosto.

dd) Son los valores de (oficina, exterior) en el mes de Enero.

Una vez conocido el sistema que generó los datos podemos establecer una interpretación de los cambios que se habían producido en los números de Betti, señalaremos las siguientes observaciones:

1) Es destacable que el esquema anterior cuando existe un termostato que regula la calefacción hace que el primer invariante de Betti sea cero, en los demás casos es mayor que cero.

2) La mezcla un poco tanto arbitraria de datos ha sido detectada mediante más cambios en los números de Betti. En el ejemplo bd que nos han mezclado los datos de primavera con los de invierno ha generado que el primer invariante de Betti saltara de 1 a 2. Cuando para despistarnos nos han mezclado valores de temperaturas del mes de Agosto ha hecho que se rompiera la estabilidad del invariante cero de Betti de modo que ha saltado de uno a dos.

Por lo tanto los cambios de los invariantes de Betti se han debido a cambios fundamentales del sistema anterior, unas veces por existencia de sistemas automáticos, otras ha sido el cambio fuerte de los rangos de recorrido habituales de los valores que dan sensores y también por mezclas inadecuadas de conjuntos de datos, sin previo aviso o con criterios caprichosos.

En consecuencia, desde el estudio de este simple ejemplo inferimos que el sistema de invariantes de homología es capaz de detectar cambios fundamentales en un sistema. Si adicionalmente se dispone de un modelo interpretativo causal que determine la procedencia de estos cambios se dispondrá de un sistema más competo de control automático para el sistema estudiado.

5 Orografía de la aplicación densidad de un conjunto de datos

Ya hemos abordado el estudio de los invariantes del conjunto inicial de una aplicación densidad, como se trata de aplicaciones con valores en un anillo R, normalmente el anillo de los enteros, reales o complejos. En relación con el conjunto final puesto que estamos trabajando con complejos cúbicos finitos la imagen contendrá un número finito de valores.

El objetivo de esta sección es la diseñar algunos procedimientos que asignen a cada aplicación densidad invariantes que además de la forma del complejo inicial de la aplicación densidad analicen la orografía de la gráfica de la densidad, de modo los invariantes no cambien por pequeños cambios en la orografía pero sí que debe cambiar si los cambios son cualitativamente importantes.

Para centrar ideas en esta sección supondremos que trabajamos con aplicaciones densidad como las obtenidas por la construcción cero, así que el dominio es una pseudovariedad (con borde) y el soporte de la densidad está contenido en el conjunto de los n-cubos. Señalaremos algunos procedimientos para su estudio:

Discretización de la Teoría de Morse.

Sea $\mathbf{K}$ un complejo cúbico y sea f: $\mathbf{K} \to \mathbf{Z}$ una aplicación densidad. Puesto que la imagen de la aplicación es finita, existe un valor mínimo a y un valor máximo A que por supuesto verifican que $a \leqslant A$. En el caso que $a=A$ se trata de una aplicación densidad constante. La aplicación densidad f determina una partición en $\mathbf{K}_n$ de modo que $\mathbf{K}_n = \bigcup\{ f^{-1}(z) | z \in [a, A]\}$. Una primera dificultad es la de que los conjuntos de cubos $f^{-1}(z)$ no tiene por qué ser subcomplejos.

Supongamos que estamos trabajando con una construcción como aquella que llamamos cero, que tenía la propiedad adicional que determinaba un complejo $\mathbf{K}$ que era n-puro, además la aplicación densidad era nula en dimensiones menores que n y sobre los n-cubos tenia valores estrictamente positivos. En este caso $a=0$, pero claramente se ve que hay otro mínimo m que es el que toma f sobre los n-cubos y que si $\mathbf{K}$ es no vacío se tiene que $a=0<m$. Si llamamos M al máximo de f sobre los n-cubos se sigue que $M=A$. En este caso se tiene que $\mathbf{K}=\bigcup\{ <f^{-1}(z)> \mid z \in [m, M]\}$ donde $<f^{-1}(z)>$ es el menor subcomplejo que contiene a $f^{-1}(z)$. Se observa que la intersección $<f^{-1}(z)> \bigcap <f^{-1}(z')>$ es un subcomplejo de dimensión estrictamente menor que n.

Un primer invariante (conjunto de invariantes numéricos) asociado a la aplicación densidad f puede ser los invariantes homológicos (Betti y torsión) del complejo junto con la familia de invariantes homológicos de los subcomplejos $<f^{-1}(z)>$ para $m \leqslant z \leqslant M$.

Construido el invariante se puede decir que f:$\mathbf{K} \to \mathbf{Z}$ es equivalente a f':$\mathbf{K}' \to \mathbf{Z}$ si los invariantes homológicos de f son isomorfos (números de Betti y torsión iguales) que los de f'.

Este método puede tener la ventaja de reflejar con cierta precisión las propiedades homológicas de la función densidad. En seguida se observa un posible problema y es que la cantidad de invariantes a calcular puede ser muy alto por lo que los tiempos de computo pueden ser muy elevados.

También podemos estudiar el grafo de Reeb asociado a una función densidad. Llamaremos *planicie* de f con valor z a cada componentes conexas de $<f^{-1}(z)>$. El grafo de Reeb de f tiene como vértices todas las planicies de f. Si $\mathbf{P}$ y $\mathbf{Q}$ son planicies de f tales que $\mathbf{P} \bigcap \mathbf{Q} \neq \emptyset$ entonces se incorpora una arista al grafo de Reeb. Llamaremos *pseudoplanie* de f con valor z a cada pseudocomponentes de $<f^{-1}(z)>$.

Para más información sobre grafos de Reeb, véase [BFMPS]. Para otros trabajos sobre teoría de Morse, puntos críticos y discretizaciones, véase [GVV].

Ante el coste temporal de los calculos sugeridos podemos idear algún método alternativo:

Hoyas, mesetas absolutas y relativas y puertos

Recordemos que tenemos una pseudovariedad con borde **K** cuyo borde lo denotamos por B**K**. Para cada $z \in [m,M]$ se tiene que la fibra $<f^{-1}(z)>$ es una pseudovariedad con borde $B<f^{-1}(z)>$. Cada pseudovariedad $<f^{-1}(z)>$ se descompone en pseudocomponentes $<f^{-1}(z)> = <D^1(z)> \cup ... \cup <D^{k(z)}(z)>$ que llamamos pseudoplanicies.

Las pseudocomponentes $<D^l(z)>$ asociadas al mínimo m , tienen una interesante propiedad adicional

"Si σ es un (n-1)-cubo ortogonal al j-ésimo eje que está en $B<D^l(z)>$ y no está en B**K**, y tal que queda a la izquierda de la pseudocomponente entonces $\partial^{-}_{j} f(\sigma) < 0$ y si está a la derecha >0."

Diremos que estás pseudocomponentes son *hoyas* o *depresiones minimales absolutas*. Las pseudocomponentes $<D^i(M)>$ asociadas al máximo, tienen una propiedad análoga cambiando el sentido de las desigualdades. Diremos que son las *mesetas maximales absolutas*.

El subcomplejo hma(f) es la reunión de las hoyas minimales absolutas y el mma(f) la reunión de las mesetas maximales absolutas. Notemos que si f es constante setiene que hma(f) =**K**=mma(f) .

Cualquier otra pseudoplanicie que verifique la propiedad anterior pero no corresponda al mínimo diremos que es una peudocomponente minimal relativa y, las otras, pseudoplanices maximales relativas.

El subcomplejo hoyas(f) es la reunión de todas las pseudoplanicies minimales y subcomplejo es la mesetas(f) es reunión de todas las pseudoplanicies maximales. Notemos que los subcomplejos anteriores verifican la relación

hma(f) $\subset$ hoyas(f) $\subset$ **K** $\cap$ mesetas(f) $\cap$ mma(f)

Estos subcomplejos determinan regiones importantes del conjunto de datos originales. El primer subcomplejo contiene las zonas de densidad más baja. El segundo son zonas de densidad baja relativa que pueden jugar un papel importante. El quinto subcomplejo corresponden a las zonas de densidad más alta , el cuarto a máximos relativos de la densidad que en muchos casos se pueden considerar también como acumulaciones de datos.

El resto de las pseudoplanicies no verifican la propiedad anterior hecho que se puede producir por diferentes motivos.

Si se extiende la aplicación densidad de modo nulo a los n-cubos de $\mathbf{E}^n$ que no están en **K**, entonces se disponen de dos valores mínimos a<m, donde a=0. Notemos que el complejo $\mathbf{E}^n$ no tiene borde y las pseudoplanicies no nulas de la densidad extendida son las pseudoplanicies de la densidad inicial. Notemos que respecto a la densidad extendida las hoyas minimales no nulas que cortan al borde de **K** pueden dejan de ser minimales. Es decir, en general aunque eliminemos la meseta nula de las zonas minimales de la densidad extendida los subcomplejos de la densidad f no coinciden exactamente con los de la densidad extendida.

Sólo hoyas y mesetas.

Es una variante del anterior. Además del complejo inicial de la aplicación densidad, se pueden considerar el subcomplejo minimal que es la reunión de todas la pseudoplanicies minimales (absolutas y relativas) el subcomplejo maximal; es decir, la reunión de las mesetas maximales (absolutas y relativas).

hoyas(f) $\subset$ **K** $\cap$ mesetas(f)

Como antes si en vez de considerar la densidad f no nula consideramos su extensión nula a todo $\mathbf{E}^n$ entonces se tienen que las pseudoplanicies no nulas de la extenmsión de f son las pseudopalnicies de f.

Flexión direccional.

Dada una función densidad f: $\mathbf{K}_n \to \mathbf{Z}$, en primer lugar la extendemos dando valores nulos a los n-cubos que no estén en el complejo, bar{f} : $\mathbf{E}^n{}_n \to \mathbf{Z}$. Ahora definimos la variación de densidad como una aplicación V: $\mathbf{K} \to \{-1,0,1\}^n$, del modo siguiente dado un n-cubo τ de **K** y una dirección j$\in$ {1, ... , n}, V_j(τ) toma valores -1, si la extensión de la densidad es decreciente en τ (basta que sea decreciente a derecha o a izquierda) , el valor 0, si la extensión tiene un mínimo o máximo relativo o es plana a derecha e izquierda, y finalmente, 1 si es creciente (a izquierda o derecha). Entonces

La aplicación V: $\mathbf{K}_n \to \{-1,0,1\}^n$ descompone a $\mathbf{K}_n$ como la reunión disjunta

$\mathbf{K}_n = \bigcup\{ V^{-1}(v) | v \in \{-1,0,1\}^n \}$ y el complejo $\mathbf{K}=\bigcup\{< V^{-1}(v)> | v \in \{-1,0,1\}^n \}\sim$. A su vez a las pseudocomponentes de los subcomplejos $< V^{-1}(v)>$ que podemos llamar v-puertos. Notemos que para v=0, los 0-puertos son pseudoplanicies maximales o minimales que no cortan al borde de **K**.

Si V es la aplicación variedad de densidad podemos definir la *flexión direccional* de un n-cubo como el producto $c(\sigma)=V_1(\sigma)... V_n(\sigma)$.

La aplicación V divide también al complejo **K** en tres subcomplejos, el generado por los n-cubos con flexión direccional positiva, con curvatura cero y con curvatura negativa.

Teoría de puntos críticos para funciones densidad

Dada una función densidad f: $\mathbf{K}_n \to \mathbf{Z}$, en primer lugar la extendemos dando valores nulos a los n-cubos que no estén en el complejo, bar{f} : $\mathbf{E}_n \to \mathbf{Z}$. Puesto que disponemos de los operadores $\partial^-{}_i \partial_j$ podemos considerar la siguiente versión discretizada del Hessiano $\partial_i \partial^-{}_j f$ que nos determina la herramienta necesaria para mimetizar los enunciados de la teoría de puntos críticos de una función.

Clasificación homológica orográfica de una aplicaciones de densidad

Supongamos que estamos trabajando con aplicaciones densidad como las del procedimiento anterior. Entonces dos conjuntos de datos serán equivalentes si los invariantes homológicos (números de Betti y torsión) del complejo inicial de la función densidad y los de cada uno de los subcomplejos definidos anteriormente son isomorfos.

Esta clasificación permite diferenciar en algunos casos funciones densidad distintas definidas sobre el mismo complejo cúbico.

6 Interpolaciones polinómicas de las aplicaciones densidad

En esta sección seguimos considerando aplicaciones densidad sobre pseudovariedades (con borde) n-dimensionales que sólo toman valores no nulos sobre el conjunto de n-cubos.

Haremos referencia al trabajo [EHRS] que determina un procedimiento que asocia a una aplicación densidad f: $\mathbf{Z}^n \to \mathbf{R}$ series polinómicas que al evaluarlas en puntos con coordenadas enteras coinciden con *f*. Para otros aspectos de interpolación en varias variables véase [GS]

Existen diversos tipos de series que dependen de una sucesión básica de polinomios p_α respecto de la cual f se expresa como $f=\sum_\alpha a(\alpha,f)\ p_\alpha \sim$. Además la series de este tipo verifican que en regiones acotadas de $\mathbf{Z}^n$ se anulan todos salvo un número finito de los polinomios anteriores, con lo cual en dicha región la serie se reduce a un polinomio.

Notemos que una serie se puede interpretar como una función de $\mathbf{R}^n$ en $\mathbf{R}$ (también de $\mathbf{C}^n$ en $\mathbf{C}$). Estas series siempre convergen en $\mathbf{Z}^n$ y podemos asociar a cada $z \in \mathbf{Z}^n$ un radio de convergencia $r(z)$. Particularmente interesantes son aquellas series para los que los radios de convergencia respecto de la norma del máximo son mayores que 1/2 ya que entonces el domino de la serie es todo $\mathbf{R}^n$. Dada una $f : \mathbf{Z}^n \to \mathbf{R}$ diremos que es *una aplicación discreta analítica* si los radios de convergencia de la serie que la extiende son mayores que 1/2.

Recordamos que existe un isomorfismo natural de $\mathbf{E}^n{}_n \cong \mathbf{Z}^n$ que hace corresponder a un subcomplejo $\mathbf{K}$ la región que se corresponde con $\mathbf{K}_n \sim$. A continuación sugerimos alguna maneras de extender funciones con dominio regiones $\mathbf{K}_n$ "de $\mathbf{Z}^n$ " a todo $\mathbf{Z}^n$ con lo cual después se puede aplicar este tipo de procedimientos para construir series y polinomios.

Primer procedimiento

Puesto que el complejo cúbico **K** es finito existe un cuboide c(**K**) tal que **K** ⊂ c(**K**) y no está contenido en otro cuboide más pequeño . Podemos extender f a todo c(**K**) mediante un procedimiento estándar pero que de valores estrictamente menores o iguales que cero a los n-cubos de c(**K**) que no estén en **K**.

En el trabajo [EHRS] se construye un polinomio p_K tal que sobre los n-cubos de c(**K**) es igual exactamente a la extensión de la densidad. Entonces si los criterios de extensión de la densidad son estándar, resulta que una aplicación densidad determina (p_K ,c(**K**)) y dada una pareja (p,,c) los n-cubos de complejo se recuperan como aquellos del cuboide c cuyo vértice principal, que tiene coordenadas enteras, verifican que el polimomio les asigna valores estrictamente positivos.

Notemos que si se cambia el criterio de extender con valores menores o iguales que cero, a una aplicación densidad le puede corresponder una pareja (p',c) diferente. Del mismo modo dos parejas (p, c) y (p',c) que tengan el mismo soporte con valores positivos y la restricción de p y p' a este soporte coincida determinan la misma aplicación densidad.

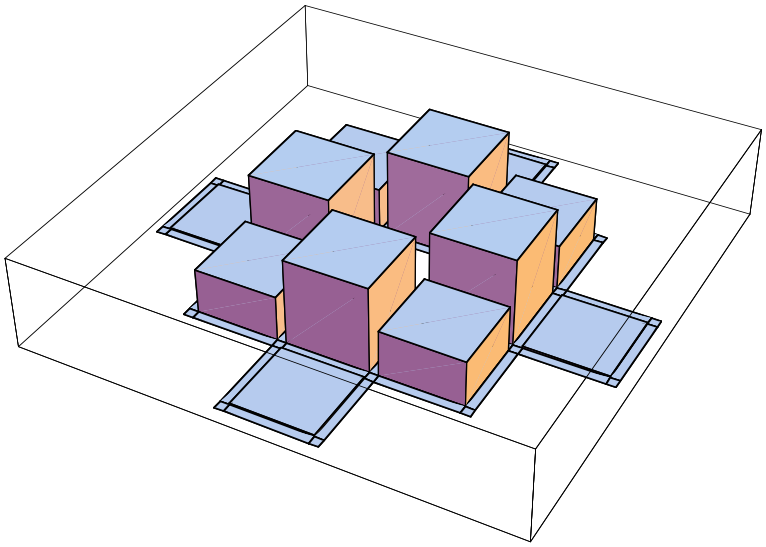
De momento indicamos tres formas canónicas de extender la ampliación densidad, la de asignar a los n-cubos de c(**K**)$_n$ que no estén en **K**$_n$ el valor -1. Otra posibilidad es el asignar el valor 0. Finalmente puede ser interesante la de asignar la distancia del vértice principal al complejo **K** pero con signo negativo.

Notemos que con el primer criterio de extensión, la hipersuperficie singular de ecuación p_K =0 se puede decir que es *una ecuación del borde de la muestra de puntos*. Divide en tres zonas, aquellas con valores positivos regiones de la nube de puntos, el borde de la muestra y las regiones que no contienen datos.

Veamos con un ejemplo del polinomio construido extendiendo con valor -1. Consideremos la siguiente aplicación densidad

```
{{{{0, 0}, {-2, 0}}, {{0, 0}, {-2, 1}}, {{0, 0}, {-1, -1}}, {{0, 0}, {-1, 0}},
  {{0, 0}, {-1, 1}}, {{0, 0}, {-1, 2}}, {{0, 0}, {0, -2}}, {{0, 0}, {0, -1}},
  {{0, 0}, {0, 0}}, {{0, 0}, {0, 1}}, {{0, 0}, {0, 2}}, {{0, 0}, {0, 3}}, {{0, 0}, {1, -2}},
  {{0, 0}, {1, -1}}, {{0, 0}, {1, 0}}, {{0, 0}, {1, 1}}, {{0, 0}, {1, 2}}, {{0, 0}, {1, 3}},
  {{0, 0}, {2, -1}}, {{0, 0}, {2, 0}}, {{0, 0}, {2, 1}}, {{0, 0}, {2, 2}}, {{0, 0}, {3, 0}},
  {{0, 0}, {3, 1}}, {{0, 1}, {-2, 0}}, {{0, 1}, {-1, -1}}, {{0, 1}, {-1, 0}},
  {{0, 1}, {-1, 1}}, {{0, 1}, {0, -2}}, {{0, 1}, {0, -1}}, {{0, 1}, {0, 0}},
  {{0, 1}, {0, 1}}, {{0, 1}, {0, 2}}, {{0, 1}, {1, -2}}, {{0, 1}, {1, -1}}, {{0, 1}, {1, 0}},
  {{0, 1}, {1, 1}}, {{0, 1}, {1, 2}}, {{0, 1}, {2, -1}}, {{0, 1}, {2, 0}}, {{0, 1}, {2, 1}},
  {{0, 1}, {3, 0}}, {{1, 0}, {-2, 0}}, {{1, 0}, {-2, 1}}, {{1, 0}, {-1, -1}},
  {{1, 0}, {-1, 0}}, {{1, 0}, {-1, 1}}, {{1, 0}, {-1, 2}}, {{1, 0}, {0, -2}},
  {{1, 0}, {0, -1}}, {{1, 0}, {0, 0}}, {{1, 0}, {0, 1}}, {{1, 0}, {0, 2}}, {{1, 0}, {0, 3}},
  {{1, 0}, {1, -1}}, {{1, 0}, {1, 0}}, {{1, 0}, {1, 1}}, {{1, 0}, {1, 2}}, {{1, 0}, {2, 0}},
  {{1, 0}, {2, 1}}, {{1, 1}, {-2, 0}}, {{1, 1}, {-1, -1}}, {{1, 1}, {-1, 0}},
  {{1, 1}, {-1, 1}}, {{1, 1}, {0, -2}}, {{1, 1}, {0, -1}}, {{1, 1}, {0, 1}},
  {{1, 1}, {0, 2}}, {{1, 1}, {1, -1}}, {{1, 1}, {1, 0}}, {{1, 1}, {1, 1}}, {{1, 1}, {2, 0}}},
 {0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
  0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 30, 59, 30, 1, 58, 59, 1, 31, 59, 30, 1}}
```

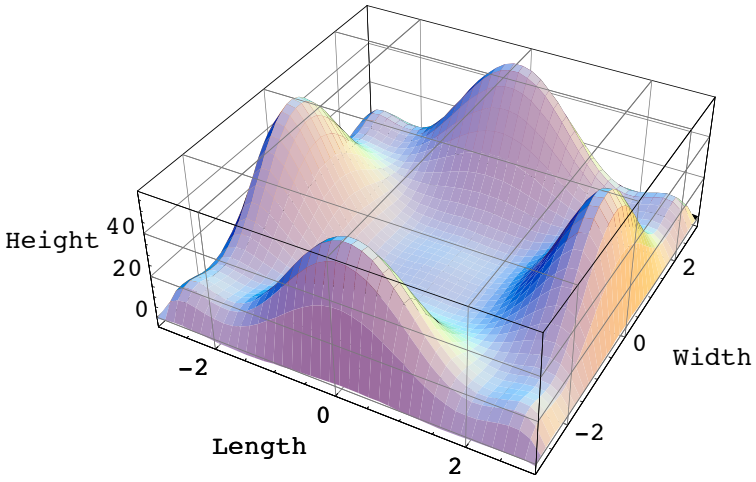
que le corresponde la siguiente representación gráfica:



Extendiendo con valor -1 la función anterior, se obtiene el siguiente polinomio

$$
\begin{aligned}
&-1+\tfrac{1}{3}\,(1+x)\,(2+x)\,(3+x)-\tfrac{7}{24}\,x\,(1+x)\,(2+x)\,(3+x)+\tfrac{1}{8}\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)-\\
&\tfrac{5}{144}\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)+8\,(3+x)\,(3+y)+\tfrac{7}{2}\,(2+x)\,(3+x)\,(3+y)-\\
&\tfrac{4}{3}\,(1+x)\,(2+x)\,(3+x)\,(3+y)-\tfrac{5}{6}\,x\,(1+x)\,(2+x)\,(3+x)\,(3+y)+\\
&\tfrac{13}{20}\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(3+y)-\tfrac{2}{9}\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(3+y)+\\
&\tfrac{7}{2}\,(3+x)\,(2+y)\,(3+y)-\tfrac{59}{4}\,(2+x)\,(3+x)\,(2+y)\,(3+y)+\tfrac{27}{4}\,(1+x)\,(2+x)\,(3+x)\,(2+y)\,(3+y)-\\
&\tfrac{7}{8}\,x\,(1+x)\,(2+x)\,(3+x)\,(2+y)\,(3+y)-\tfrac{19}{120}\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(2+y)\,(3+y)+\\
&\tfrac{1}{18}\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(2+y)\,(3+y)+\tfrac{1}{3}\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{4}{3}\,(3+x)\,(1+y)\,(2+y)\,(3+y)+\tfrac{27}{4}\,(2+x)\,(3+x)\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{109}{36}\,(1+x)\,(2+x)\,(3+x)\,(1+y)\,(2+y)\,(3+y)+\tfrac{5}{16}\,x\,(1+x)\,(2+x)\,(3+x)\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{29}{240}\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{23}{540}\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(1+y)\,(2+y)\,(3+y)-\tfrac{7}{24}\,y\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{5}{6}\,(3+x)\,y\,(1+y)\,(2+y)\,(3+y)-\tfrac{7}{8}\,(2+x)\,(3+x)\,y\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{5}{16}\,(1+x)\,(2+x)\,(3+x)\,y\,(1+y)\,(2+y)\,(3+y)+\tfrac{37}{288}\,x\,(1+x)\,(2+x)\,(3+x)\,y\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{1}{9}\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,y\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{331\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,y\,(1+y)\,(2+y)\,(3+y)}{8640}+\tfrac{1}{8}\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{13}{20}\,(3+x)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)-\tfrac{19}{120}\,(2+x)\,(3+x)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{29}{240}\,(1+x)\,(2+x)\,(3+x)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{1}{9}\,x\,(1+x)\,(2+x)\,(3+x)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)+\tfrac{133\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)}{2400}-\\
&\tfrac{1}{8640}\,(163\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y))-\\
&\tfrac{5}{144}\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)-\tfrac{2}{9}\,(3+x)\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{1}{18}\,(2+x)\,(3+x)\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)-\\
&\tfrac{23}{540}\,(1+x)\,(2+x)\,(3+x)\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)+\\
&\tfrac{331\,x\,(1+x)\,(2+x)\,(3+x)\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y)}{8640}-\\
&\tfrac{1}{8640}\,(163\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y))+\\
&\tfrac{1}{9600}\,(61\,(-2+x)\,(-1+x)\,x\,(1+x)\,(2+x)\,(3+x)\,(-2+y)\,(-1+y)\,y\,(1+y)\,(2+y)\,(3+y))
\end{aligned}
$$

cuya representacion gráfica es la siguiente:



La parte central tiene densidad negativa lo que indica que en dicha zona no hay elementos de nuestro conjunto de datos.

Recordemos que si igualamos a cero el polinomio anterior se obtiene una ecuación del borde de la nube.

Segundo procedimiento

Por simplicidad daremos la explicación en el caso que el complejo **K** sea de dimensión uno. Pero el método es fácilmente generalizable a cualquier dimensión.

Se trata de dar una familia finita de polinomios p_1 , ... , p_m . Estos polinomios tiene asociados dominios de la forma $v_1 + [-a_1 , a_1]$, ... , $v_m + [-a_m , a_m]$. Entonces podemos casi solapar estos polinomios mediante la siguiente expresiones de la forma

$$e^{(x-v_1)^2/(a_1)^2}\, p_1 (x-v_1)+... + e^{(x-v_m)^2/(a_m)^2}\, p_m (x-v_m)$$

Este procedimiento aunque aumenta un poco la descripción de los dominios de los polinomios, reduce el grado de cada uno de ellos, lo que en un sistema en el que la exponencial sea ágil el cálculo del valor de la función en un punto arbitrario será rápido.

En dimensiones superiores se trata de encontrar adecuadas particiones de la unidad asociadas a una reunión de subdominios en cada uno de los cuales disponemos de un polinomio que interpola a la función discreta.

También se pueden considerar descomposiciones "regulares" en cuboides y considerar un polinimio en cada cuboboide. Si los grados son adecuados en dos cuboides contiguos las restricciones de sendos polinomios coinciden de modo que la familia de polinomios determina una exensión continua que es diferenciable en cada cuboide.

7 Algunos antecedentes y técnicas relacionadas

Las técnica sugerida en este trabajo tienen antecedentes de tipo muy diverso. Surge en un intento de encontrar un lenguaje común para realizar procesos de discretización, aproximación, interpolación, estudio de invariantes homológicos y aplicaciones a la minería de datos y visualización.

Una de las nociones básicas que se considera es la de cubo geométrico. Los cubos aparecen ya en la etapa griega; incluso antes, existen interesantes ejemplos del Neolítico, véase [EHR]. Es, precisamente, uno de los cuerpos platónicos el que inspira la técnica que aquí hemos presentado, se trata del único cuerpo platónico autodual.

Algunas de las ideas básicas de estas técnicas ya aparecen en los "Elementos" de Euclides, donde se describe con perfección la geometría de los segmentos, cuadrados y cubos y la existencia de procesos fundamentales como la subdivisión de un segmento.

Obviamente más recientes son las técnicas topológicas introducidas por J.H.C. Whithehead [W1, W2] y las técnicas de la teoría de prehaces para la construcción de conjuntos simpliciales debidas principalmente a D. Kan [K]. Con estos antecedentes, se trata de presentar los complejos cúbicos con un lenguaje más flexible que permita tanto interpretarlos topológicamente como complejos de celdas como expresarlos en términos de prehaces, de tal modo que además tengan una buena conexión con otras técnicas analíticas y algebraicas. Desde el punto de vista topológico, destacaremos también la Teoría de la Forma iniciada por Borsuk [B], que determina diversos procedimientos para aproximar espacios por poliedros.

Algunas de las técnicas básicas que se emplean en este trabajo consisten en la definición de adecuados operadores de derivación discreta, que juegen un papel similar a los empleados en procesos clásicos de discretización. Los operadores discretos clásicos análogos a la diferenciación más usados son los llamados operadores de diferencias (referencias clásicas son Steffensen [S], Jordan[J] o Hildebrand[H]). Este tipo de operadores han sido empleados en el Análisis Numérico para la interpolación y derivación numérica de funciones y, especialmente, para la resolución numérica de Ecuaciones en Derivada Parciales (EDP's). En la actualidad se siguen mejorando los métodos basados en diferencias finitas para la resolución de EDP's (Hyman , Shashkov y Steinberg [HSS], Khan y Ohba [KO]), y se profundiza en técnicas de discretización de operadores del análisis clásico (Hyman y Shashkov[HS1, HS2], Margolin,, Shashkov y Smolarkiewicz [MSS], y Castillo, Hyman, Shashkov y Steinberg [CHSS]). En conexión con estos trabajos, se trata de generalizar algunas de las técnicas, dotándolas de un lenguaje más flexible y a la vez más capaz de extender su uso a dimensiones superiores y a complejos cúbicos más generales.

En relación con las aplicaciones de la minería de datos a los campos científicos y tecnológicos, se cuenta con una amplia experiencia de diversos grupos de investigación, por ejemplo, se dispone de los resultados de los talleres de trabajo de diversos especialistas que han sido recogidos en [GKKK]. Se contará también con las experiencias realizadas por el grupo del área de Proyectos de Ingeniería Mecánica de la Universidad de La Rioja [COCM]. En nuestra propia Universidad se cuentan con otros grupos que tienen una gran experiencia en las aplicaciones de la Minería de Datos a ciertos campos tecnológicos, véase [COCM]~.

Referencias

[B] K, Borsuk, "Concerning homotopy properties of compacta", Fund. Math, 62, 1968.

[BEAACDDEGGHHHJLLPSSWYZ] Marshall Bern, David Eppstein, Pankaj K. Agarwal, Nina Amenta, Paul Chew, Tamal Dey, David P. Dobkin, Herbert Edelsbrunner, Cindy Grimm, Leonidas J. Guibas, John Harer, Joel Hass, Andrew Hicks, Carroll K. Johnson, Gilad Lerman, David Letscher, Paul Plassmann, Eric Sedgwick, Jack Snoeyink, Jeff Weeks, Chee Yap, and Denis Zorin , "Emerging Challenges in Computational Topology", preprint, 2000?.

[BDR] J.N. Bhuyan, J.S.Deogun and V.V. Raghavan, "Algoritms for the Boundary Selection Problem" , Algorítmica, 17, 133-161 (1997).

[BFMPS] S.Biasotti, B. Falcidieno, M. Mortara, G. Patané and M. Spagnuolo, "Computational Topology Techniques for Shape Modelling", ERCIM News, 50, July 2002.

[COCM] M. Castejón, J. Ordieres, J. De Cos y F.J. Martínez, "Metodología para el análisis previo a la modelización de datos en procesos inductriales. Fundamentos teóricos y prácticas con R". Servicio de Publicaciones de la Uniersidad de La Rioja, 2001.

[CHSS] J.E. Castillo, J.M. Hyman, M. Shashkov and S. Steinberg, "Fourth and sixth-order conservative finite difference approximations of the divergence and gradient", Applied Numerical mathematics, 37, 2001.

[DE] C. J, A. Delfinado and H. Edelsbrunner, "An incremental algorithm for Betti numbers of simplicial complexes", Proceedings of the ninth annual symposium on Computational geometry ,p.232-239, May 18-21,1993, San Diego,California,United States

[CDES] H.I. Chen, T.K. Dey, H. Edelsbrunner and J. Sullivan, "Dynamic Skin Triangulations", Discrete Computational Geometry, 23, 523-567 (2001).

[DH] M. Dellnitz and H. Hohmann, "A subdivision algoritm for the computation of unstable manifolds and globals attractors", Num. Math. 75, 293-317 (1997).

[DE] T.K. Dey and H. Edelsbruner , "Counting Triangle Crossings and Halving Planes", Discrete Comput. Geometry 12, 281-289) (1994).

[D] A. Dold, "Lectures on algebraic topology", Springer-Verlag, 1972.

[DHSW1] J.G. Dumas, F. Heckenbach, D. Saunders and V. Welker , "Computing Simplicial Homology Based on Efficient Smith Normal Form Algoritms", preprint.

[DHSW2] J.G. Dumas, F. Heckenbach, D. Saunders and V. Welker, "Simplicial Homology. Aproposed share package for GAP", preprint.

[EHR] J.I. Extremiana, L.J. Hernández y M.T. Rivas, Poliedros, en "Margarita Mathematica en Memoria de José Javier (Chicho) Guadalupe Hernández", 139-166, Servicio de Publicaciones de la Universidad de La Rioja, 2001.

[EHRS] J.I. Extremiana, L.J. Hernández, M.T. Rivas y E. Saenz de Cabezón, "Derivaciones discretas y homología", prepublicación, 2003

[E] H. Edelsbrunner, "The union of balls and its dual shape", Discrete Comput. Geom, 13, (1995), 415-440.

[ER] H. Edelsbrunner and E.A. Ramos, "Edgewise Subdivisión of a Simplex", Discrete Comput. Geom, 24, 707-719 (2000).

[ER] H. Edelsbrunner and E.A. Ramos, "Inclusión-Exclusion Complexes for pseudodisk Collections" , Discrete Comput. Geom, 17, 287-306 (1997).

[ES] H. Edelsbrunner, Shah "Shape Reconstructions with Delaunay Complex", Preprint

[E] H. Edelsbrunner, "Triangulations and meshes in computational geometry", Acta Numérica, 1-81 (2000).

[EFL] H. Edelsbrunner, M. Facello and J. Liang, "On the Definition and Construction of Pockets in Macromoleculaes", preprint, (1998).

[EH] V. Estivill-Castro, M.E. Houle, "Robust Distance-Based Clustering with Applications to Spatial Data Mining", Algoritmica, 30, 216-242 (2001).

[F] J. Friedman, "Computing Betti numbers via combinatorial Laplacians", Proc. 28th Sympos. Theory Com-put., (1996), 386-391.

[GS] M. Gasca and T. Sauer, "Polynomial interpolation in several variables", Advances in Computational Mathematics, 0 (2001).

[GVV] R.W. Ghrist, J.B. Van Den Berg and R.C. Vandervorst, "Morse theory of spaces of braids and Lagrangian Dynamics", prepublicación, 2003.

[GKKK] R. L. Grossman, C. Kamanth , P. Kegelmeyer y V. Kumar (Editores), Data Mining for Scientific and Engineering Applications, Kluwer Academic Publishers, 2001.

[G] M. Greenberg,, "Lectures on algebraic topology", Benjamin, 1967.

[R] R. González-Díaz ,"Algebraic tools for computing Betti numbers"

[GLR1] R. González-Díaz , F. Leal and P. Real ,"Computing the Cohomology on Digital Images", prepublicación, 2003.

[GLR2] R. González-Díaz , F. Leal and P. Real, "Towards Digital Cohomology", prepublicación, 2003.

[H] F.B. Hildebrand "Introduction to numerical análisis" (2nd ed.), Dover, New York, 1987.

[HSS] J.M. Hyman, M. Shashkov, S. Steinberg, "The effect of inner products for discrete vector fields on the accuracy of mimetic finite difference methods", Computer and mathematics with applications, 42, (2001).

[HS1] J.M. Hyman, M. Shashkov, "Natural discretizations of the divergence, gradient and curl on logically rectangular grids",

[HS2] J.M. Hyman, M. Shashkov, "Adjoint operators for the natural discretizations of the divergence, gradient and curl on logically rectangular grids", , Applied Numerical Mathematics, 25 (1997) 413-442.

[HMM] R. Holley, J. Malitz and S. Malitz, "An Approach to Hierarchical Clustering via Level Sufaces and Convexity", Discrete Comput. Geometry 25, 221-233 (2001).

[K] D. M. Kan, "On homotopy theory and c.s.s. groups", Annals of Math. 68(1), 38-53(1958).

[KMW] W.D. Kalies, K. Mischaikov G. Watson, "Cubical aproximations and computation of homology", Banach Center Pubications, 47, 115-131 (1999).

[KL] D. Krznaric and C. Levcopoulos, "Fast Algoritms for Complete Linkage Clustering", Discrete Comput. Geom.. 19, 131-145 (1998).

[J] C. Jordan "Calculus of Finite differences" (3 ed.), Chelsea Publishing Company, New York,1965.

[K] M. Kantardzid "Data Mining", Concepts, Models, Methods and Algoritms, Wiley and Sons, Inc, 2003.

[K-B] R. Kanna, A. Bachem, "Polynomial algorithms for computing the Smith and Hermite normal forms of an integer matriz" SIAM Jour. comp., 8, (1979), 499-507.

[KOH] I.R. Khan, R. Ohba, N. Hozumi, "Mathematical Proof of closed form expressions for finite difference approximations based on Taylor series", Journal of Computational and Applied Mathematics, 150, 2003.

[KO] I.R. Khan., R. Ohba, "Closed form expressions for finite difference approximations of first and higher derivatives based on Taylor series", Journal of Computational and Applied Mathematics, 107, 1999.

[M] W. S. Massey, "Singular Homology Theory", Graduate Texts in Mathematics, 70, Springer-Verlag, New York-Berlin, 1980.

[MSS] L.G. Margolin, M. Shashkov and P.K. Smolarkiewicz., "A discrete operator calculus for finite difference approximations", Computer methods in applied mechanics and engineering, 187, 2000.

[M] J. Matousek, "On Approximate Geometric k-Clustering", Discrete Comput. Geometry, 24 61-84 (2000).

[Mu] J. R. Munkres, "Elements of Algebraic Topology", Addison-Wesley, 1984.

[M] M. Niethammer, A.N. Stein, , W.D. Kalie, P. Pilarczyk, Mischaikow, K. Tannnenbaum, "Analasis of Blood Vessel Topology by Cubical Homology",Springer-Verlag,1980.

[N] S. Novikov, "Discrete Schrondinger Operator on Topology",arXiv:math-ph/990325

[P] R. Pollack,"On the Betti numbers of semi-algebraic sets"

[R] J. Rubio,"Topología Algebraica Efectiva", La Gaceta, 198-208 (1999).

[RS] J. Rubio and Sergeraert, "KENZO a Symbolic Sofware for Effective Homology Computation", ftp fourier.ujf-grenoble.fr en /pub/KENZO

[R] V. Robins, "Foundations for computing topology from finite approximations",Concepts, Models, Methods and Algoritms, Wiley and Sons, Inc, 2003.

[ST] Seifert and Threlfall, "A textbook of topology", Academic Press, 1980.

[S] J.F Steffensen, "Interpolation" (2nd ed.), Chelsea Publishing Company, New York, 1950.

[V] G. Vegter, "Computational topology", In Jacob E. Goodman and Joseph O'Rourke, editors, Handbook of Discrete and Computational Geometry, chapter 28, pages 517-536. CRC Press, Boca Raton,FL, 1997.

[W1] J.H.C. Whitehead, "Combinatorial homotopy I", Bull. Amer. Math. Soc., 55 (1949) 213-245.

[W2] J.H.C. Whitehead, "Combinatorial homotopy. II", Bull Amer. Math. Soc., 55 (1949) 453-496.

[Y] JCSS Yao, "Decision Tree Complexity and Betti Numbers"