Espacios de búsquedas geométricamente separables

Edilma Olinda Gagliardi

Departamento de Informática

Facultad de Ciencias Físico, Matemáticas y Naturales

Universidad Nacional de San Luis, Argentina

oli@unsl.edu.ar

Fax: 54-2652-430224

Gregorio Hernández Peñalver

Departamento de Matemática Aplicada

Facultad de Informática

Universidad Politécnica de Madrid, España

gregorio@fi.upm.es

Fax: 34-91-3367426

Resumen

Una temática abordada a menudo en Bases de datos es el estudio de los rangos y las consultas por rangos, denominado *Búsquedas por Rangos*. Este problema tratado desde una perspectiva geométrica nos permite diseñar y analizar algoritmos y estructuras de datos con herramientas propias de la Geometría Computacional.

En el ámbito de la geometría, el estudio de *Separabilidad Geométrica* es de utilidad en campos de aplicación donde se requiere discriminar y/o separar objetos. En este sentido, las regiones se obtienen basándose en características propias de los objetos y de su ubicación en el espacio considerado.

Podemos unificar las nociones de búsquedas por rangos con las de separabilidad geométrica. Tenemos conjuntos disjuntos de objetos en el espacio y nos interesan particularmente las descripciones de las curvas que determinan las regiones que contienen tales conjuntos, puesto que ellas constituyen los separadores geométricos. En este sentido, la búsqueda por rangos puede aprovechar estas particiones del espacio para la recuperación de objetos.

Dado que la selección de los separadores geométricos a ser aplicados para obtener la partición del espacio es un problema difícil, por ser de tipo combinatorio, proponemos el uso de herramientas no tradicionales como las *Metaheurísticas*, donde la partición pueda ser guiada.

En este trabajo de investigación presentamos los aspectos teóricos y prácticos relevantes para las búsquedas por rangos en espacios de búsquedas geométricamente separables proponiendo la aplicación de metaheurísticas.

Palabras Claves: Bases de datos, Búsquedas por rangos, Separabilidad geométrica, Metaheurísticas.

1. Introducción

El problema de las Búsquedas por Rangos en Bases de Datos puede ser tratado en el marco de la Geometría Computacional, dado que nos brinda nuevas posibilidades de diseño de algoritmos y estructuras de datos utilizando herramientas geométricas [1] [2] [3] [6] [7] [18] [19].

Una de tales posibilidades es la aplicación de Separabilidad Geométrica, que nos permite proponer nuevos algoritmos de partición del espacio de búsqueda por medio de la aplicación de diversos criterios de separabilidad (rectas, cuñas, bandas, entre otras) [4] [5] [9] [14] [15] [16] [17] [20]. La aplicación de tales separadores geométricos debe realizarse con relación a la instancia particular del espacio de búsqueda, lo cual hace presuponer, o bien conocimientos de su configuración, o bien el uso de herramientas no tradicionales que ayuden a la selección apropiada de tales criterios de separabilidad.

Proyecto Tecnologías Avanzadas de Bases de Datos 22/F314, UNSL.

Proyecto Geometría Computacional AL2002-1010-2.43/ AL2003-1010-2.55/ AL2004-1010-2.53, UPM.

Por tanto, en este trabajo presentamos una vinculación entre las búsquedas por rangos y la separabilidad geométrica, presentando aspectos teóricos relevantes y proponiendo nuevas ideas para el diseño de estructuras de datos para las búsquedas por rangos. También planteamos trabajar con metaheurísticas para lograr selecciones adecuadas de separadores geométricos.

Este artículo está organizado del siguiente modo: primeramente, introducimos nociones relacionadas a la separabilidad geométrica y a las búsquedas por rangos, proponiendo una vinculación entre ambas temáticas. Posteriormente, brindamos una propuesta de investigación donde, explícitamente, para esta vinculación proponemos la utilización de metaheurísticas para la obtención de resultados satisfactorios en la elección de los separadores geométricos.

2. Separabilidad Geométrica

Los trabajos de separabilidad están orientados a dos o más conjuntos disjuntos de objetos geométricos, básicamente para puntos, bajo diversos criterios de separabilidad como bandas, cuñas, sectores, entre otros. Estos criterios de separabilidad tienen aplicaciones interesantes, como por ejemplo el análisis de imágenes, clasificación de datos, etc. Siempre que sea necesario discriminar objetos en un espacio de trabajo, por algún atributo del mismo, los criterios de separabilidad pueden tener un papel importante.

Para modelar este problema, se dan dos conjuntos disjuntos P y Q de objetos en el plano, los que se referirán como objetos *rojos* y *azules*, respectivamente. Eventualmente, los objetos pueden ser: puntos, segmentos, polígonos o círculos. En cualquiera de los casos, n es el tamaño de la entrada. Entonces, dada una familia C de curvas en el plano, decimos que los conjuntos P y Q son *C-separables* si existe una curva $C_i \in C$, llamada *separador*, tal que cada componente conexa de $\mathbf{R}^d - C_i$ contiene solamente objetos, o bien de P, o bien de Q. Se dice que un conjunto finito S de superficies es un separador de conjuntos de objetos en $\mathbf{R}^d$ si las componentes conexas de $\mathbf{R}^d - S$ contienen objetos de un solo conjunto.

Si C es la familia de rectas en el plano, tenemos la *separabilidad lineal*. Dos conjuntos P y Q son *linealmente separables* si y sólo si sus cierres convexos no se intersecan.

Dos conjuntos disjuntos P y Q de objetos en el plano son *separables por una cuña* si existe una cuña que contiene solamente todos los objetos de uno de los conjuntos. Aquí se estudia el problema de decidir la separabilidad por cuñas calculando la ubicación de los posibles ápices de cuñas; como una extensión al problema, se estudia hallar la cuña de ángulo mínimo (máximo).

Dos conjuntos disjuntos P y Q de objetos en el plano son *separables por banda* si existe una banda, determinada por dos rectas paralelas, que contiene solamente todos los objetos de uno de los conjuntos. Aquí se estudia el problema de decidir la separabilidad por bandas calculando el conjunto de intervalos de pendientes de bandas separadoras; como una extensión al problema, se estudia hallar la banda mínima (máxima).

Los algoritmos para resolver los problemas de decisión y optimización propuestos anteriormente se ejecutan en tiempo $O(n \log n)$; salvo en el caso de la separación lineal, que toma tiempo $O(n)$ [20]. La figura 1 nos ilustra estos tres criterios de separabilidad.

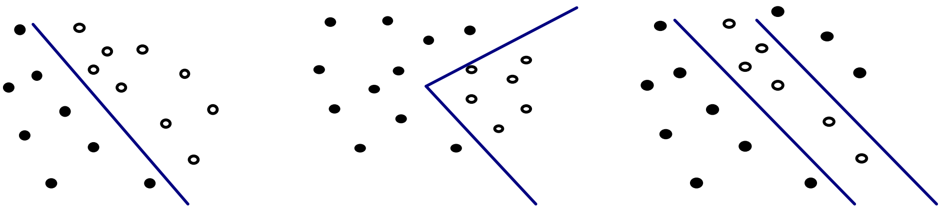


Figura 1: diversos separadores geométricos.

Los principales resultados sobre separabilidad con bandas y cuñas, con diferentes versiones sobre combinaciones y propiedades de ellas, son algoritmos eficientes para decidir la existencia de los separadores y calcular soluciones factibles. Algunos de estos resultados pueden ser extendidos en el plano al uso de otros objetos geométricos que no sean puntos y al espacio tridimensional en donde sólo existen algunos algoritmos para resolver determinados problemas de separación[17].

Los estudios sobre separabilidad en el plano han avanzado sobre separaciones con varias bandas, cuñas y sectores. Así tenemos *separabilidad por dos bandas*, asociado al problema de optimización de hallar las dos bandas de ancho mínimo (máximo). Este problema puede ser resuelto en $O(n^3 \log n)$ tiempo. En el mismo tiempo se puede decidir si las bandas son de mínima (máxima) anchura. En cambio, si se proporciona la pendiente de una de las bandas, el tiempo se reduce a $O(n^2 \log n)$, pudiendo hallarse las dos bandas de mínima (máxima) anchura en el mismo tiempo. Aun mejor si las dos pendientes son conocidas, puesto que se reduce a $O(n)$ tiempo. Como resultado, tenemos que dados dos conjuntos disjuntos P y Q de puntos en el plano, el mínimo número de bandas paralelas (rectas) que los separan y los correspondientes intervalos de pendientes, puede ser determinado en $O(n^2 \log n)$ tiempo.

También, tenemos *separabilidad por dos cuñas*, que en caso de existir el par de cuñas, podemos tener cuñas disjuntos o no; podemos determinarla en $O(n^5 \log n)$ tiempo; pero, en caso de que las cuñas sean disjuntas, podemos decidir la separabilidad en $O(n^3 \log n)$tiempo. Hay más variantes asociadas, como la *separabilidad por sectores*, si todas las cuñas separadoras tienen el mismo ápice.

En lugar de tener una nube bicolor, podemos considerar diversos colores. Dados $C_1, C_2, ..., C_k$ conjuntos disjuntos en el plano, decimos que el conjunto C_i tiene el color c_i, denotamos con n_i la cardinalidad de C_i y $n_1 + n_2 + ... + n_k = n$. Entonces, un conjunto finito S de curvas en el plano es un separador para los conjuntos $C_1, C_2, ..., C_k$ si cada componente conexa en $R^2 - S$ contienen solamente objetos de C_i. También decimos que cada componente es *monocromática*. Obviamente, cuando k=2, son los casos estudiados previamente.

Dados los conjuntos disjuntos $C_1, C_2, ..., C_k$ coloreados, con n puntos en total, en el plano, la mínima cantidad de rectas paralelas separadoras de los puntos en bandas monocromáticas y los intervalos de pendientes de todas las posibles soluciones, pueden calcularse en $O(n^2 \log n)$ tiempo. En el caso de permitir k-1 rectas paralelas únicamente, conseguir los intervalos de pendientes de todas las posibles soluciones, pueden calcularse en $O(kn) + O(k^2 \log k)$ tiempo.

Los principales resultados sobre separabilidad con bandas y cuñas, con diferentes versiones sobre combinaciones y propiedades de ellas, son algoritmos eficientes para decidir la existencia de los separadores y calcular soluciones factibles. Algunos de estos resultados pueden ser extendidos en el plano al uso de otros objetos geométricos que no sean puntos y al espacio tridimensional en donde sólo existen algunos algoritmos para resolver algunos de los problemas vistos. En este marco, también queda abierto el estudio en espacios de mayor dimensión sobre nuevos criterios de separabilidad. Es decir, no sólo usar planos o semiplanos, sino otros objetos como pirámides dobles (bipirámides), cilindros, cónicas, cónicas dobles, etc. En todos los casos resueltos, la complejidad crece polinomialmente respecto del tamaño de entrada, por lo cual los algoritmos propuestos caen en la clase de los algoritmos polinomiales.

3. Búsquedas por Rangos

Este es un problema central en Geometría dado que diversas aplicaciones hacen uso de las búsquedas por rangos, tal como las *Búsquedas de Intersección.* Dentro de esta categoría se tiene la búsqueda de intersección de puntos, o de segmentos, o de rectángulos, entre otras.

Para introducirnos en el tema, comencemos con las nociones básicas.

Un *Espacio de Rangos* se define como un sistema $\Omega = (U, F)$, donde U es un conjunto de objetos geométricos y F es una familia de subconjuntos de U.

Los elementos de F son llamados *Rangos* de Ω.

Por ejemplo: $\Omega_1 = (\mathbf{R}^d, \{ h/ h$ es un semiespacio en $\mathbf{R}^d \})$; $\Omega_2 = (\mathbf{R}^d, \{ h/ h$ es una bola en $\mathbf{R}^d \})$, etc.

Dados un espacio de rangos $\Omega = (U, F)$, S un subconjunto de objetos de U y R un rango de F, un problema de búsquedas por rangos puede expresarse como *consultar los objetos geométricos que están en $S \cap R$.* En este caso, a R suele llamárselo *rango de consulta* (*query range*).

La búsqueda por rangos, esencialmente, consiste en buscar los objetos geométricos que contiene una determinada región (rango) del espacio de objetos geométricos. Entonces, observando la forma de las regiones correspondientes a los rangos de consultas, podemos hacer una clasificación en cuatro tipos clásicos de búsqueda por rangos: por Rangos Ortogonales (BRO), por Rangos SemiEspaciales (BRSp), por Rangos Simpliciales (BRSx) y por Rangos SemiAlgebraicos (BRSa).

Nuestro interés radica esencialmente en realizar consultas varias veces sobre el mismo conjunto, por lo cual es conveniente pre-procesar el mismo en una estructura de datos. La mayoría de las estructuras de datos para consultas por rango, se construyen en forma recursiva, dividiendo el espacio de objetos geométricos en varias regiones, con propiedades geométricas deseables sobre ellas. Estas estructuras de datos son referidas como esquemas de descomposición jerárquicos.

En las búsquedas por rangos ortogonales, los rangos son d-rectángulos, cada uno de la forma $\Pi_{i=1..d}[a_i, b_i]$ con a_i, $b_i \in \mathbf{R}^d$, con lados paralelos a los ejes de coordenadas del espacio subyacente. Ésta es una abstracción del problema de búsqueda por múltiples claves, que a causa de sus numerosas aplicaciones, sobre todo en Bases de Datos, ha sido estudiado por muchos investigadores, a lo largo de las tres últimas décadas, buscando bajar las cotas en el tiempo de respuesta y en el espacio requerido.

Con respecto a las búsquedas por rangos semialgebraicos, los rangos están definidos por conjuntos semialgebraicos [1] [2] [23] [24][25]. Por ejemplo, el conjunto de todas las bolas cerradas de $\mathbf{R}^2$, o el conjunto de todos los paraboloides en $\mathbf{R}^3$. Este problema está relacionado a diversos tópicos de investigación como el *problema del vecino más cercano* o el *problema de los k vecinos más cercano.* Chazelle y Welz [8] dieron una solución con espacio lineal para un problema de búsqueda por rango circular en el plano.

Para las búsquedas por rangos semiespaciales, los rangos conforman el conjunto de todos los semiespacios cerrados de $\mathbf{R}^d$. Así, en d=2, tenemos semiplanos. Las principales ideas se deben a

Willard [21] [22] sobre búsquedas por rangos con semiplanos, las que se han generalizado para búsquedas por rangos semiespaciales. Mejores construcciones se han hallado, incluso para dimensiones superiores. Todos los algoritmos de la literatura se basan en el teorema del Ham-Sandwich, consecuencia del teorema de Borsuk-Ulam, quienes utilizan siempre algún esquema de partición para lograr partir el espacio respecto de un conjunto de puntos en varias regiones acotadas por hiperplanos. Una propiedad que se destaca es que para cualquier hiperplano de consulta, las regiones intersecadas contienen una cantidad reducida de los puntos del conjunto total. Teniendo un esquema de partición, es posible construir un árbol de partición en forma estándar [18][25].

Finalmente, en las búsquedas por rangos simpliciales, las consultas en este contexto son consultas en polígonos; es decir, los rangos son polígonos. Todo polígono puede descomponerse en piezas más pequeñas, realizando una triangulación del mismo. De esta forma, la consulta en el polígono se transforma en consultas sobre triángulos de la triangulación. El resultado de la consulta es la unión de las respuestas de cada triángulo. Este concepto puede extenderse al espacio d-dimensional, donde tenemos símplices. Un símplice es el cierre convexo de d+1 puntos independientes [18][19][22].

Estos últimos tres tipos de búsqueda son los que han recibido más atención recientemente, puesto que además de sus aplicaciones directas, han brindado soluciones a problemas subyacentes en problemas geométricos de mayor jerarquía. Aunque las complejidades son mayores que las de *BRO*, ocurre que ellas han brindado soluciones a problemas, no resueltos por las *BRO*.

En cuanto a las consultas, dado un rango de consulta puede interesarnos efectuar alguna de las siguientes clases sobre él: Consultas de Recuento (*range counting*), Consultas de Reporte (*range reporting*), Consultas Booleanas (*range emptiness*) y Consultas Extremales.

El problema real para el cual tiene sentido este estudio, es que el conjunto de objetos geométricos, *S*, es dado con anterioridad y luego, repetidas veces, variando el rango *R*, se consulta por la intersección. Entonces, no es suficiente con diseñar un buen algoritmo sino que también se debe diseñar una estructura de datos apropiada para almacenar el conjunto *S*, de forma tal que cada consulta por un rango *R*, pueda ser respondida eficientemente.

Los análisis de las complejidades tanto en tiempo de construcción de la estructura de datos, en espacio que ocupa y en tiempos de respuestas de las consultas, depende de la dimensión del espacio, de la cardinalidad de *S*, de la estrategia de partición de *S* y del tamaño de la respuesta de la consulta.

4. Combinando Separabilidad Geométrica con las Búsquedas por Rangos

Innovando en nuevos métodos de partición, incorporamos las nociones separabilidad geométrica como posibles criterios de separación de objetos en el espacio o partición del espacio.

Como hemos visto acerca de separabilidad geométrica, tenemos conjuntos disjuntos de objetos en el plano, los cuales queremos separar por algún criterio de separabilidad. Los conjuntos son disjuntos debido a alguna propiedad específica de los objetos y de su posición en el espacio. Nos interesan particularmente las descripciones de las curvas que determinan las regiones que contienen tales conjuntos disjuntos, dado que ellas constituyen los separadores geométricos. Con esta información podemos obtener un esquema de partición, en cuyo caso estamos en condiciones de crear una estructura de datos apropiada para recuperar los puntos por medio de diversos rangos. En la figura 2 podemos visualizar estas ideas:

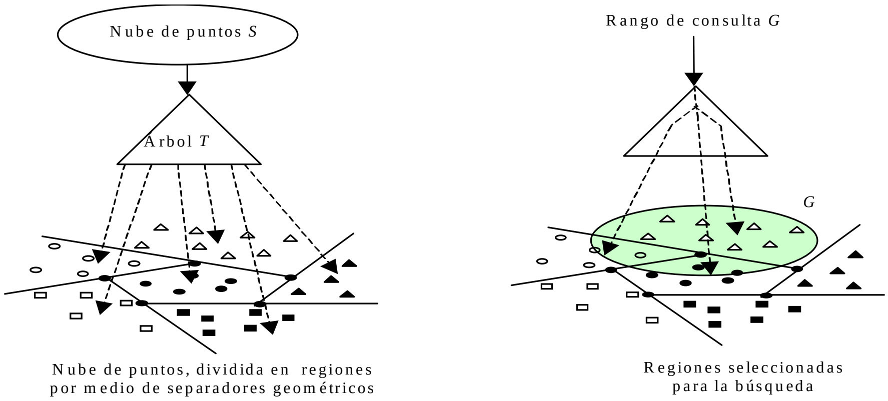


Figura 2: esquema de partición, estructura de acceso y consulta por rango.

Las consultas se pueden realizar en función de las características dadas de los puntos, o sino, por rangos. En el primer caso, por ejemplo, podemos solicitar las regiones correspondientes a los puntos rojos, pudiendo obtener como respuesta una o varias regiones; o también, consultar cuáles son los diferentes colores existentes. En el caso de los rangos, podemos utilizar un rango de consulta que abarque parte de una región o varias regiones. Por ejemplo, un rango de consulta dado por un semiplano H, definido por una recta h, puede corresponderse con partes de regiones, con regiones completas, y posiblemente (es deseable), dejar excluidas regiones íntegras.

A continuación veremos la construcción del primer nivel de una estructura jerárquica, considerando una partición obtenida por criterios de separabilidad geométrica, y luego analizaremos la resolución de consultas.

Cualquier característica que interese acerca de los puntos, puede verse como una aplicación w:$S \to C$, donde $C=\{c_1, c_2,..., c_k\}$, $1 \leq i \leq k$, para todo p perteneciente a S. Un ejemplo es considerar el color de los puntos.

Dado un conjunto S de n puntos en el plano y una aplicación total w como definimos anteriormente, generamos una *Partición Poligonal* de S, $\Pi=\{(S_1,P_1),(S_2,P_2),...,(S_m,P_m)\}$, con las siguientes propiedades:

- $S_i \subseteq S$, $S_i \subseteq P_i$, $S_i \cap S_j = \varnothing$, $P_i \cap P_j = \varnothing$ para $i \neq j$, $1 \leq i,j \leq$m.
- Para cada i existe un h tal que si p está en P_i, $w(p)=c_h$.
- n_i es la cardinalidad de S_i y $n_1+ n_2+... + n_m= n$.
- Las regiones poligonales son inducidas por separadores geométricos.

Es interesante y deseable pensar que las nubes de puntos acepten criterios de separabilidad acotados a complejidades bajas, aunque no es relevante, porque la principal presunción que podemos hacer es que la estructura se genera una única vez y no es dinámica. Puede ser prioritario crear estructuras con bajo costo de almacenamiento o bien crear estructuras que privilegien los tiempos de respuestas. Por asumir, justamente, que la estructura se crea una vez y que es estática, es que los costos de construcción son amortizados basándose en el criterio de ahorro elegido. Además, el objetivo principal es realizar muchas consultas por diferentes regiones sobre la misma nube de puntos.

Teniendo una partición poligonal $\Pi=\{(S_1,P_1),(S_2,P_2),...,(S_m,P_m)\}$ para una nube de puntos S, se construye un *Árbol de Partición Poligonal*. La raíz tiene m hijos, siendo cada uno la raíz de un árbol de partición poligonal, definido recursivamente para cada una de las clases de la partición poligonal del siguiente modo:

- Si S contiene un solo punto p, el árbol de partición consiste de una hoja donde p se almacena explícitamente. En este caso, la partición poligonal es $\Pi=\{(S_1,P_1)\}$.
- En otro caso, la estructura es un árbol T con m hijos. Cada hijo de la raíz corresponde a una de las regiones poligonales de la partición de S. La región poligonal correspondiente a un nodo v, hijo de la raíz, se denota P_v, y mantiene la curva separadora. La clase respectiva, S_v, se llama clase canónica de v.
- En cada hijo de la raíz almacenamos información correspondiente a la clase S_v y la descripción de la región poligonal P_v.

Con respecto a las consultas, tenemos básicamente del tipo: *¿cuáles son las regiones de un determinado color?*, o, *¿cuáles son los diferentes colores que se hallan en el plano?* O bien, las *consultas por rango*.

Para la primera, basta con consultar por el color correspondiente a cada región poligonal P_v, se describe en S_v. Es una consulta que toma tiempo $O(m)$, siendo m la cantidad de regiones poligonales que obtuvimos. Para la segunda consulta, es la misma idea, salvo que por cada región poligonal, registramos qué color almacena y finalmente, se concluye en la unión de todos los colores existentes. También su complejidad es lineal.

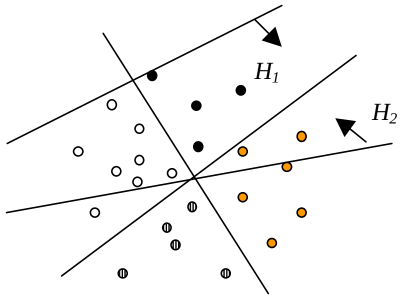
Acerca de las consultas por rangos, podemos tener diferentes tipos. Entonces, dado un conjunto de puntos en el plano, se desea recuperar y/o contar los puntos que caen en una determinada región definida por un rango de consulta.

Asumimos que en el caso de rangos polígonos, estos son simples; o en su defecto, siempre se pueden aproximar. En estos casos, la región del rango de consulta se triangula y así la consulta es derivada en consultas sobre cada triángulo resultante. El resultado final es la unión de los resultados parciales obtenidos por cada triángulo. Queda precisar cómo se resuelve la consulta en un triángulo. Dado que un triángulo es la región determinada por la intersección de tres semiplanos, la consulta se descompone en tres consultas por semiplanos y el resultado final de la consulta en el triángulo es la intersección de los resultados parciales obtenidos por cada semiplano.

Además, proponemos otros rangos de consulta basándonos en los separadores geométricos. Así se pueden tener rangos cuñas, rangos bandas, etc. Cualquiera de estos rangos, puede considerarse como la unión/intersección de semiplanos, por lo que básicamente las consultas se reducen a consultas por semiplanos.

Supongamos tener un semiplano H, definido por una recta h. Si una región poligonal P_i no es cortada por la recta h, una de las siguientes situaciones ocurre: $P_i \subseteq H$ o bien $P_i \cap H = \emptyset$. En el primer caso, todo P_i es parte de la respuesta, mientras que en el segundo, P_i se descarta totalmente. Queda por ver el caso en que P_i es cortado por h, en cuyo caso basta revisar la clase S_i y determinar qué puntos quedan contenidos en el semiplano H. En este último caso, la búsqueda es lineal sobre S_i. Como sólo tenemos un nivel del árbol, hasta aquí en el peor caso, el orden de complejidad de la consulta es lineal.

En la figura 3, mostramos dos semiplanos de consulta.

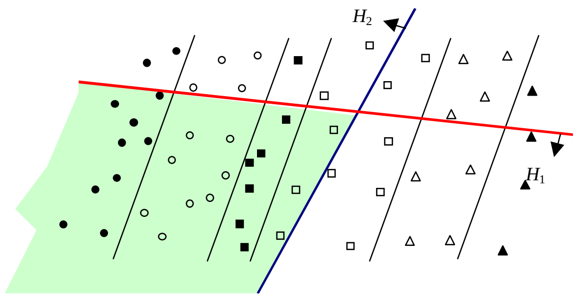


El semiplano H_1 tiene peores condiciones, puesto que por cada sector ocurre: $P_i \cap H \neq \emptyset$ y $P_i \not\subset H$, y por lo tanto es necesario determinar para todos los puntos de S en qué semiplano quedan contenidos.

Mientras que para H_2, solamente hay que mirar en detalle los dos sectores que corta. El resto o queda fuera o queda totalmente contenido en el semiplano de consulta.

Figura 3: consultas por semiplanos en un espacio partido por una cuña doble.

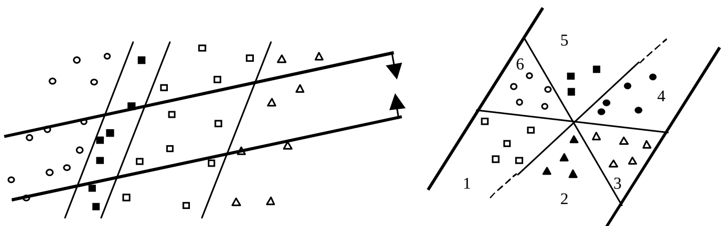
Veamos otro ejemplo, supongamos que proponemos un *rango cuña*. Una cuña es la región correspondiente a la intersección de dos semiplanos opuestos; precisamente es alguno de los cuadrantes descriptos por dos rectas que se cortan en algún punto. En la figura 4, mostramos un rango cuña sombreado.



Una consulta de este tipo se descompone en dos consultas por semiplanos, y luego se realiza la intersección de los cuadrantes correspondientes.

Figura 4: consultas por rango cuña en un espacio partido por múltiples bandas.

Un *rango por bandas*, no difiere mucho en su resolución y complejidad. Supongamos que hemos realizado en el primer nivel una separación por varias bandas. Todas las regiones obtenidas son parcialmente acotadas, dando origen a un árbol de aridad m en el primer nivel. En este caso, m=k. Luego, por cada hijo analizamos qué posibilidades de separación existen y basándose en ello, volvemos a aplicar algunas de las vistas previamente. Por ejemplo, supongamos que tomamos la banda i, y que vale separarla en sectores.



En el primer nivel realizamos una Separación por varias bandas y presentamos un rango banda

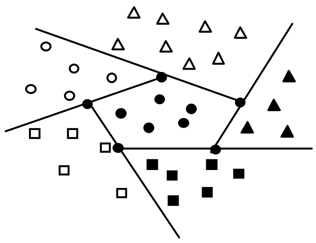
En la banda *i*, realizamos una Separación por sectores

Los sectores 2 y 5 son parcialmente acotados por las rectas que definen las cuñas. Mientras que, los sectores 1, 3, 4 y 6 están totalmente acotados por las rectas que definen la banda y las que definen las cuñas

Figura 5: partición en dos niveles (bandas y cuñas).

Con respecto a la banda i, del i-ésimo hijo del nodo raíz se desprenderán seis nodos hijos, uno para cada sector obtenido por la segunda separación. Repitiendo este tipo de proceso con cada hijo, estamos en condiciones de conformar el segundo nivel del árbol. El rango por bandas se resuelve como la intersección de dos semiplanos.

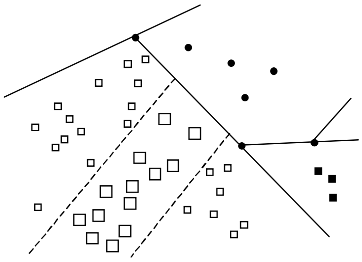
En el siguiente ejemplo, figura 6, supongamos que tomamos la siguiente descomposición en primera instancia.



La nube S está descompuesta en seis regiones. Tenemos k=6 conjuntos disjuntos de puntos, coloreados diferentes. El ejemplo muestra un cierre convexo para el conjunto C_1. Por cada lado del cierre, se disparan rayos que parten la zona exterior en varias regiones que cubren el espacio. En cada región se alberga un C_i, con $2 \leq i \leq k$. También en este caso tenemos m=k.

Figura 6: partición compuesta por cierre convexo y cuñas.

El primer nivel del árbol de partición poligonal queda compuesto por seis nodos pendientes de la raíz, cada uno representando una región poligonal. Elegimos una región cualquiera y trabajamos sobre ella. Si la región corresponde al i-ésimo hijo de la raíz, éste será el padre de tres nodos, correspondientes a cada banda obtenida en la segunda separación, como mostramos en la figura 7:



El separador es una banda, posiblemente de ancho máximo.

Figura 7: bandas dentro de una cuña

Para concluir con estas ideas, hemos creado un árbol de partición poligonal, obtenido por medio de la aplicación de diversos criterios de separabilidad geométrica. Nos planteamos qué cosas podríamos continuar haciendo, y vemos que se abre un abanico de posibilidades. Por cada hijo de la raíz podemos analizar qué posibilidades de separación existen y basándonos en ello, volvemos a aplicar algunas de las variantes vistas previamente. Repitiendo proceso con cada hijo, estamos en condiciones de conformar los diversos niveles del árbol. Observemos que la clasificación en cada nivel debe realizarse en función de características diferentes de la nube de puntos [10] [11] [12] [13].

5. El uso de Metaheurísticas

Considerar cuáles son los criterios de separabilidad más propicios para la nube de puntos resulta una incógnita a resolver y que dependerá, en la mayoría de los casos, de la instancia de la nube de puntos a tratar. Así, encontrar la secuencia óptima (o al menos aceptable) de separadores geométricos puede resultar un problema difícil, lo cual nos puede llevar a la aplicación de distintos tipos de búsquedas en un espacio considerable de soluciones. Para resolver este problema proponemos el uso de metaheurísticas.

Entonces, supongamos que llamamos E a nuestro espacio de soluciones, conformado por todas las secuencias posibles de cualquier longitud de separadores geométricos. Cada nivel del árbol de partición poligonal está representado en la secuencia indicando cuáles son los separadores aplicados para obtener las diferentes regiones. Obviamente, como el árbol tiene aridad variable en todos los niveles, es necesario considerarlo en la secuencia.

Nuestro objetivo consiste en hallar un $x \in E$ que resulte adecuado; es decir, que represente una secuencia de separadores geométricos adecuada respecto de alguna métrica en particular.

Como casos de métrica, se podrían considerar la longitud de la secuencia. En este caso tendríamos una idea de la profundidad del árbol y de la cantidad de regiones obtenidas. También, otra métrica de interés podría ser la complejidad final de la secuencia. Ésta nos daría una idea de cuál es el costo de resolución del problema, y así podríamos definir más métricas.

Ello nos permitiría obtener una partición de la nube de puntos, que permita crear, finalmente, el árbol de partición poligonal como estructura de datos apropiada para acceder a la misma.

De acuerdo a lo anterior, podemos ver este problema como de optimización combinatoria, donde se pueden tratar los siguientes problemas computacionales: i) caracterización del espacio de búsqueda en cuanto a la posible existencia de soluciones no factibles; ii) optimización, es decir encontrar una secuencia óptima; y iii) evaluación del costo de las soluciones consideradas. Todos estos ítems constituyen un desafío a estudiar, puesto que sobre ellos hay un campo abierto de investigación.

Como métodos de búsqueda en este espacio de soluciones, se pueden considerar diferentes clases. Nosotros en particular proponemos las metaheurísticas, dado que permiten resolver problemas de optimización de interés práctico mediante estudios experimentales que ayuden a observar caracterizaciones que den lugar a la optimización y evaluación de costos de las soluciones obtenidas.

Además, estos enfoques incluyen muchas variaciones, basadas en enfoques de la naturaleza misma, otros en evolución biológica, neurofisiología, y comportamientos biológicos, entre otros. Algunos ejemplos de estos son: *simulated annealing* (SA), *tabu search* (TS), *algoritmos evolutivos* (AE), *optimización basada en colonias de hormigas* (OCH), etc.

6. Conclusiones y Visión de futuro

Este campo de investigación tiene una motivación teórica, dado que resulta ser un problema interesante al pensar en vincular tópicos aparentemente aislados y que juntos devienen en un método que sirve para indagar en nuevas estrategias para partir el espacio de búsqueda, innovando en el uso de separadores geométricos y de metaheurísticas.

Este trabajo de investigación constituye un proyecto de tesis de postgrado con propuestas de diseño que desencadenen en aspectos teóricos y prácticos relevantes para las búsquedas por rangos en espacios geométricamente separables. Se enmarca dentro de la línea de investigación Geometría Computacional y Bases de Datos, perteneciente al Proyecto Tecnologías Avanzadas de Bases de Datos 22/F314, Departamento de Informática, UNSL. Además, el grupo de trabajo en Geometría Computacional de la UNSL, mantiene un proyecto de investigación conjunto con la UPM, Proyecto Geometría Computacional AL2002-1010-2.43/ AL2003-1010-2.55/ AL2004-1010-2.53, con el objetivo principal de consolidar la línea de trabajo en la UNSL, aportando nuevos enfoques y técnicas algorítmicas a las líneas de investigación ya establecidas en su Departamento de Informática. Asimismo, dentro del Departamento de Informática se está trabajando con el apoyo de integrantes del Grupo Laboratorio de Investigación y Desarrollo en Inteligencia Computacional (LIDIC) de la UNSL, quienes tienen experiencia y formación en el desarrollo y trabajo de Metaheurísticas.

Referencias bibliográficas

[1]. Agarwal, P.; Erickson, J. *Geometric range searching and its relatives*; Advances in Discrete and Comput. Geom. (B. Chazelle,J. Goodman, and R. Pollack, eds.), American Mathematical Society, Providence, 1998.

[2]. Agarwal, P.; Matousek,J.; *On range searching with semialgebraic sets;* Discrete Comput. Geom., 11: 393-418, 1994.

[3]. Agarwal, P.; *Range searching* CRC Handbook of Computational Geometry (J. Goodmand and J. O'Rourke, eds.).

[4]. Arkin, E.M.; Hurtado, F.; Mitchell,S.B.; Seara, C.; Skiena, S.S.; *Some Separability Problems in the Plane*; 16th European Workshop on Computational Geometry, Eilat, Israel, 2000.

[5]. Arkin, E.M.; Hurtado, F.; Mitchell,S.B.; Seara, C.; *Some Lower Bounds on Geometric Separability Problems*; 11th Fall Workshop on Computational Geometry, Polytechnic University, Brooklyn, NY, 2001.

[6]. Bentley, J.L.; Friedman, J.H; *Data Structures for range searching*; ACM Comput. Surv. 11:397-409;1979.

[7]. Boissonnat, J.D.; Yvinec, M. *Algorithmic Geometry*, Cambridge University Press, 1998.

[8]. Chazelle, B.; Welzl, E.; Quasi-optimal range searching in spaces of finite VC-dimension. Discrete - Compuatcional Geometry, 4:387-490, 1989.

[9]. Devillers, O; Hurtado, F.; Mitchell,S.B.; Seara, C.; *Separating Several Point Sets in the Plane*; 13th Canadian Conference on Computational Geometry, 2001.

[10]. Gagliardi, E. O.; Hernández Peñalver, G.; *Separabilidad Geométrica aplicada a las Búsquedas por Rangos*; Workshop de Investigadores en Ciencias de la Computación 2003, WICC 2003. Tándil, Buenos Aires, 22-23 de Mayo de 2003.

[11]. Gagliardi, E. O.; Hernández Peñalver, G.; *Un enfoque propuesto para las Búsquedas por Rangos con Separabilidad Geométrica*. VIII Congreso Argentino de Ciencias de la Computación 2002 (CACIC 2002). Capital Federal, 15-18 de Octubre de 2002.

[12]. Gagliardi, E.O.; Hernández Peñalver, G.; *Geometría Computacional y Bases de datos: Búsquedas por Rangos y Separabilidad Geométrica*. I Workshop de Bases de Datos, Jornadas Chilenas de Ciencias de la Computación, Copiapó, Chile, 4-9 de Noviembre de 2002.

[13]. Gagliardi, E.O.; Hernández Peñalver, G.; *Las búsquedas por rangos en espacios de búsquedas geométricamente separables.* II Workshop de Bases de Datos, Jornadas Chilenas de Ciencias de la Computación, Chillán, Chile, 3-7 de Noviembre de 2003.

[14]. Hurtado, F.; Mora, M.; Ramos, P.A.; Seara, C.; *Separación de Objetos en el Plano por Doble Cuña y por Θ-Poligonal*; VIII Encuentros de Geometría Computacional, Castelló, España, 1999.

[15]. Hurtado, F.; Noy, M.; Ramos, P.A.; Seara, C.; *Separating Objects in the Plane with Wedges and Strips*; Discrete Applied Mathematics, Vol 109, pp 109-138, 2001.

[16]. Hurtado, F.; Noy, M.; Ramos, P.A.; Seara, C.; *Separating Objects in the Plane with Wedges and Strips*; 110th Canadian Conference on computational Geometry, Montreal, Canadá, 1998.

[17]. Hurtado, F.; Seara, C.; Sethia, S.; *Red-Blue separability problems ein 3D;* to appear in the 19th European Workshop on Computational Geometry, Bonn, Germany, March 24-26, 2003.

[18]. Matousek,J.; *Geometric Range searching*; ACM Comput. Survey, 26:421-461;1994.

[19]. Melhorn, K. *Multidimensional Searching and Computational Geometry* Springer Verlag 1984.

[20]. Seara, C; *On geometric separability*; Tesis Doctoral 2002, Univ. Politécnica de Cataluña.

[21]. Willard, D.E.; Lueker, G.S*: kadding range restriction capability to dinamic data structures*. J.ACM, 32:597-617;1985.

[22]. Willard, D.E.; *Polygon retrieval;* SIAM J. Comput., 11;149-165;1982.

[23]. Yao, A.C.;Yao,F.F.; *A general approach to D-dimensional geometric queries*; Proc. 17th Annu. ACM Sympos. Theory Comput. 163-168, 1985.

[24]. Yao,F.F.; *A 3-space partition and its applications.* Proc. 15th Annu. ACM Sympos. Theory Comput. 258-263, 1983.

[25]. Yao,F.F.;Dobkin,D.P.;Edelbrusnner,H; Paterson,M.S.; *Partitioning space for range queries*; SIAM L Compuy. 18:371-384;1989.