

MATERIAL DE TRABAJO



Matemáticas y Tecnología

17-28 de marzo de 2003

**Aula Magna de las Facultades de
Matemáticas y Física**

<http://www.anamat.ull.es/sctm03>



**Cursos Universitarios Interdisciplinares 2003
Vicerrectorado de Extensión Universitaria
Universidad de La Laguna**

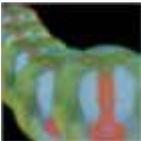
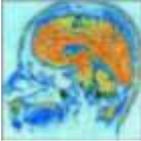
Programa

Módulo 3: *Matemáticas y Tecnología*

17-28 de marzo, 16:00-18:00 horas

Coordinadores: *María Isabel Marrero Rodríguez , Rodrigo Trujillo González*

- lunes 17  ***Aplicaciones matemáticas en los estudios de poblaciones de seres marinos sometidos a explotación pesquera***
Javier Ariz Tellería
Licenciado en Ciencias Biológicas e Investigador del Área de Pesca del Centro Oceanográfico de Canarias
- martes 18  ***Diseño de un controlador para el espejo primario del Gran Telescopio Canarias***
Marta Sigut Saavedra
Profesora Asociada de Ingeniería de Sistemas y Automática de la Universidad de La Laguna
- miércoles 19  ***Seguridad matemática en la Sociedad de la Información***
Pino T. Caballero Gil
Profesora Titular de Ciencias de la Computación e Inteligencia Artificial de la Universidad de La Laguna
- jueves 20  ***Diseño Geométrico Asistido por Ordenador***
Laureano González Vega
Catedrático de Álgebra y Decano de la Facultad de Ciencias de la Universidad de Cantabria
- viernes 21  ***Modelos matemáticos en la industria: Ejemplos y perspectivas***
Alfredo Bermúdez de Castro
Catedrático y Director del Departamento de Matemática Aplicada de la Universidad de Santiago de Compostela

-
- lunes 24  **Generación automática de mallas tridimensionales para la simulación numérica de procesos medioambientales**
Rafael A. Montenegro Armas
Catedrático de Matemática Aplicada y Director de la División de Discretización y Aplicaciones del Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería de la Universidad de Las Palmas de Gran Canaria
- martes 25  **Matemáticas y Visión por Ordenador**
Luis Álvarez León
Catedrático de Ciencias de la Computación e Inteligencia Artificial de la Universidad de Las Palmas de Gran Canaria
- miércoles 26  **Aspectos matemáticos del proyecto multidisciplinar de I+D “Espacio Acústico Virtual”**
Hareesh M. Chulani
Ingeniero Electrónico del Área de Instrumentación del Instituto de Astrofísica de Canarias
Antonio F. Rodríguez Hernández
Licenciado en Medicina y Cirugía e Investigador de Proyecto del Departamento de Fisiología de la Universidad de La Laguna
- jueves 27  **Radiaciones y Vida**
José Hernández Armas
Catedrático de Radiología y Medicina Física y Director del Laboratorio de Física Médica y Radiactividad Ambiental del Departamento de Medicina Física y Farmacología de la Universidad de La Laguna. Jefe del Servicio de Física Médica y Protección Radiológica del Hospital Universitario de Canarias.
- viernes 28  **Resolución de Problemas en Paralelo**
Coromoto León Hernández
Profesora Titular de Lenguajes y Sistemas Informáticos de la Universidad de La Laguna

Diseño de un controlador para el espejo primario del Gran Telescopio Canarias



Marta Sigut Saavedra

Profesora Asociada de Ingeniería de Sistemas y Automática
Departamento de Física Fundamental y Experimental, Electrónica y Sistemas
Universidad de La Laguna

Introducción

El Gran Telescopio de Canarias (GTC), que se encuentra actualmente en fase de construcción, se convertirá, una vez que ésta finalice, en el telescopio que, junto con el Keck en Hawai, posee el espejo primario de mayor diámetro del mundo (10m) [Álvarez, 1997]. Precisamente debido a su gran tamaño es preciso segmentarlo, por lo que se hace imprescindible la implementación de algún tipo de política de control que permita garantizar, en todo momento, el perfecto alineamiento de las piezas del espejo, llamadas segmentos. La disposición de estos segmentos hexagonales así como el tamaño del primario del GTC en relación al de una persona pueden observarse en la Figura 1.

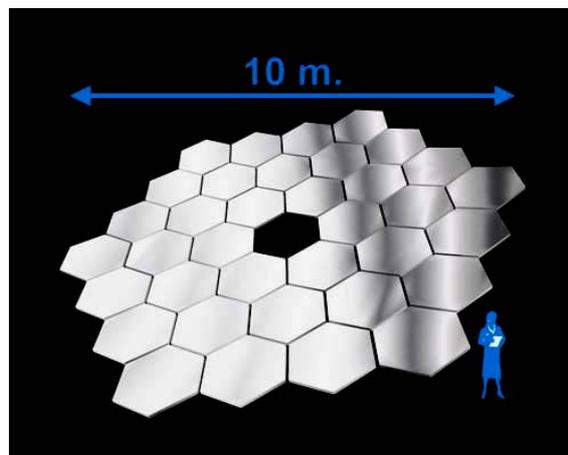


Figura 1: Dimensiones del espejo primario del GTC y disposición de los 36 segmentos.
(Copyright de GRANTECAN, S.A.)

Nos encontramos ante un sistema de gran escala, provisto de un alto número de actuadores y sensores, y complejo debido, sobre todo, al fuerte acoplamiento que presenta. En estas condiciones, diseñar un controlador multivariable, sujeto además a restricciones en cuanto a la frecuencia a la que se puede aplicar el comando, dado el volumen de información que se procesa, y teniendo en cuenta la tecnología de buses empleada, no resulta una tarea sencilla.

I. Obtención de un modelo para el espejo primario del GTC

Un aspecto fundamental en el diseño de un sistema de control para una planta es el desarrollo de un buen modelo matemático para ésta. El estudio de un sistema complejo, como es el espejo primario de diez metros de diámetro del Gran Telescopio de Canarias (GTC), requiere el uso de técnicas de modelado con el objeto de construir una representación matemática adecuada de la estructura, el sistema de actuación y los segmentos que componen el espejo. La técnica de análisis por elementos finitos es, sin lugar a dudas, una herramienta básica de cara a generar modelos dinámicos de estructuras complejas.

I.1. El problema de las dos masas. Sin necesidad de aplicar el método de los elementos finitos, podemos tener una idea de cómo se pueden obtener las matrices de estado y salida de un sistema como el espejo primario del GTC acudiendo a dos problemas sencillos. El primero es el llamado ‘problema de las dos masas’ que se ilustra en la Figura 2:

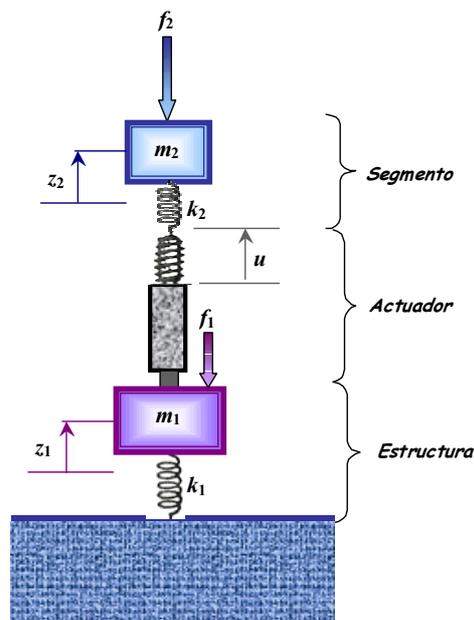


Figura 2: Representación del sistema que consideramos en el problema de las dos masas.

Imponiendo el equilibrio de fuerzas a las dos masas que componen este sistema y tomando como variables de estado del mismo los desplazamientos y las velocidades de dichas masas, se encuentran las ecuaciones de estado y salida que siguen:

$$\begin{bmatrix} \ddot{z}_1 \\ \ddot{z}_2 \\ \dot{z}_1 \\ \dot{z}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & \frac{-(k_1+k_2)}{m_1} & k_2 \\ 0 & 0 & k_2 & -k_2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} \frac{1}{m_1} & 0 \\ 0 & \frac{1}{m_2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

$$z_2 = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ z_1 \\ z_2 \end{bmatrix}$$

I. 2. Influencia del acoplamiento entre segmentos. Para ver cómo afectan a las matrices que representan la dinámica del sistema el hecho de considerar un conjunto de n segmentos que se acoplan entre sí a través de la estructura sobre la que se apoyan, acudimos al problema que se ilustra en la Figura 3:

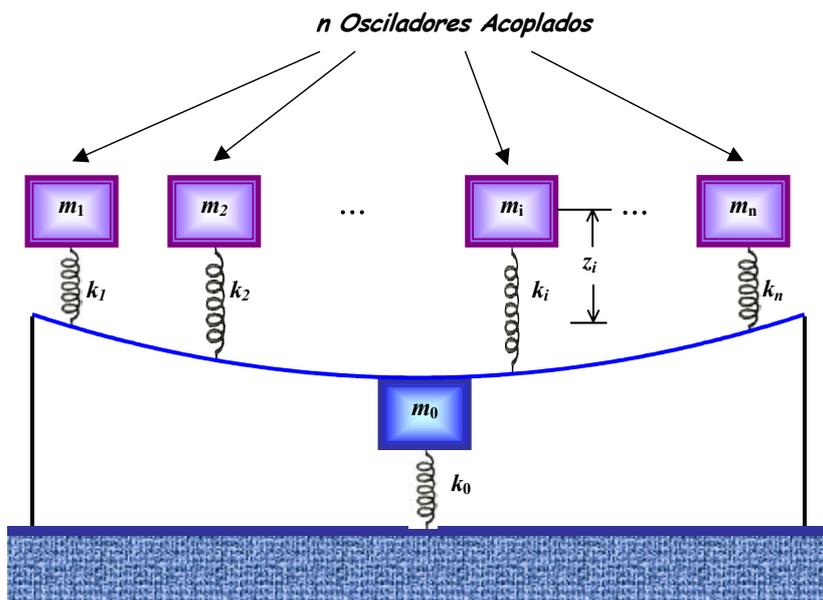


Figura 3: Modelo que representa el acoplamiento entre segmentos del espejo primario del Telescopio.

Procediendo de forma análoga a como lo hicimos en el problema de las dos masas, encontramos que las ecuaciones de estado y salida de este sistema son, ahora, las que mostramos a continuación:

$$\begin{bmatrix} \ddot{z}_0 \\ \ddot{z}_1 \\ \vdots \\ \ddot{z}_n \\ \dot{z}_0 \\ \dot{z}_1 \\ \vdots \\ \dot{z}_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & \dots & 0 & \frac{-(k_0 + \sum_{j=1}^n k_j \alpha_j)}{m_0} & k_1 & \dots & k_n \\ 0 & 0 & \dots & 0 & k_1 \alpha_1 & -\frac{k_1}{m_1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & k_n \alpha_n & 0 & \dots & -\frac{k_n}{m_n} \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \dot{z}_0 \\ \dot{z}_1 \\ \vdots \\ \dot{z}_n \\ z_0 \\ z_1 \\ \vdots \\ z_n \end{bmatrix} + \begin{bmatrix} \frac{1}{m_0} & 0 & \dots & 0 \\ 0 & \frac{1}{m_1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{m_n} \\ 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} 0 \\ u_1 \\ \vdots \\ u_n \end{bmatrix}$$

$$[z_1 \ \dots \ z_n] = [0 \ 0 \ \dots \ 0 \ 0 \ 1 \ \dots \ 1] \begin{bmatrix} \dot{z}_0 \\ \dot{z}_1 \\ \vdots \\ \dot{z}_n \\ z_0 \\ z_1 \\ \vdots \\ z_n \end{bmatrix}$$

II. Características dinámicas del primario del GTC

II. 1. El modelo con el que hemos trabajado. En el caso del espejo primario del GTC, la dinámica de los tres elementos que componen el sistema: estructura, segmentos y actuadores (ver Figura 4) viene descrita, en el espacio de estados, a través de 708 estados. Las ecuaciones de estado y de salida del sistema son las siguientes:

$$\frac{d}{dt} \begin{bmatrix} a \\ \dot{a} \\ xe \\ m_s \dot{x}e \\ r_1 \\ r_2 \\ q \\ \dot{q} \end{bmatrix} = A \begin{bmatrix} a \\ \dot{a} \\ xe \\ m_s \dot{x}e \\ r_1 \\ r_2 \\ q \\ \dot{q} \end{bmatrix} + B * U + \begin{bmatrix} 0_{30 \times 1} \\ 0_{30 \times 1} \\ 0_{108 \times 1} \\ P \\ 0_{108 \times 1} \\ 0_{108 \times 1} \\ 0_{108 \times 1} \\ 0_{108 \times 1} \end{bmatrix} \quad ; \quad S = C * \begin{bmatrix} a \\ \dot{a} \\ xe \\ m_s \dot{x}e \\ r_1 \\ r_2 \\ q \\ \dot{q} \end{bmatrix} + D * U$$

siendo:

$$\mathbf{A} = \begin{bmatrix}
 0_{30 \times 30} & I_{30 \times 30} & 0_{30 \times 108} \\
 -W^2 - k_s V^T V & -C - c_s V^T V & k_s V^T & \frac{c_s}{m_s} V^T & 0_{30 \times 108} & 0_{30 \times 108} & -k_s V^T & -c_s V^T \\
 0_{108 \times 30} & 0_{108 \times 30} & 0_{108 \times 108} & \frac{1}{m_s} I_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} \\
 k_s V & c_s V & -k_s I_{108 \times 108} & -\frac{c_s}{m_s} I_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} & k_s I_{108 \times 108} & c_s I_{108 \times 108} \\
 0_{108 \times 30} & 0_{108 \times 30} & 0_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} & I_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} \\
 0_{108 \times 30} & 0_{108 \times 30} & 0_{108 \times 108} & 0_{108 \times 108} & -\frac{8}{T^2} I_{108 \times 108} & -\frac{4}{T} I_{108 \times 108} & 0_{108 \times 108} & 0_{108 \times 108} \\
 0_{108 \times 30} & 0_{108 \times 30} & 0_{108 \times 108} & I_{108 \times 108} \\
 0_{108 \times 30} & 0_{108 \times 30} & 0_{108 \times 108} & 0_{108 \times 108} & w_q^2 I_{108 \times 108} & 0_{108 \times 108} & -w_q^2 I_{108 \times 108} & -2\xi_q w_q I_{108 \times 108}
 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix}
 0_{30 \times 108} \\
 0_{30 \times 108} \\
 0_{108 \times 108} \\
 0_{108 \times 108} \\
 -\frac{8}{T} I_{108 \times 108} \\
 \frac{32}{T^2} I_{108 \times 108} \\
 0_{108 \times 108} \\
 w_q^2 I_{108 \times 108}
 \end{bmatrix}$$

$$\mathbf{C} = [0_{168 \times 30} \quad 0_{168 \times 30} \quad M \quad 0_{168 \times 108} \quad 0_{168 \times 108} \quad 0_{168 \times 108} \quad 0_{168 \times 108} \quad 0_{168 \times 108}]$$

$$\mathbf{D} = 0_{168 \times 108}$$

donde el significado de las variables que aparecen en estas ecuaciones es el siguiente:

$V(108 \times 30)$: matriz de desplazamientos modales de la celda.

$W(30 \times 30)$: matriz diagonal de frecuencias modales angulares.

$C(30 \times 30)$: matriz diagonal de amortiguamientos ($c_i = 2\xi_i w_i$, ξ_i = amortiguamiento relativo del modo i de la estructura).

$k_s = 6N/\mu m$: rigidez de los soportes de los segmentos.

$c_s = 2000kg/sg$: amortiguamiento de los soportes de los segmentos.

$m_s = 180kg$: masa del segmento asociado a un soporte (1/3 de la masa de un segmento).

$w_q = 2\pi \times 60sg^{-1}$: frecuencia natural de los actuadores.

$\xi_q = 1/\sqrt{2}$: amortiguamiento de los actuadores.

$T = 5ms$: retardo de los actuadores.

$P(108 \times 1)$: perturbaciones debidas al viento.

$M(168 \times 108)$: matriz de relación geométrica entre las medidas realizadas por los sensores del sistema y la posición de los segmentos.

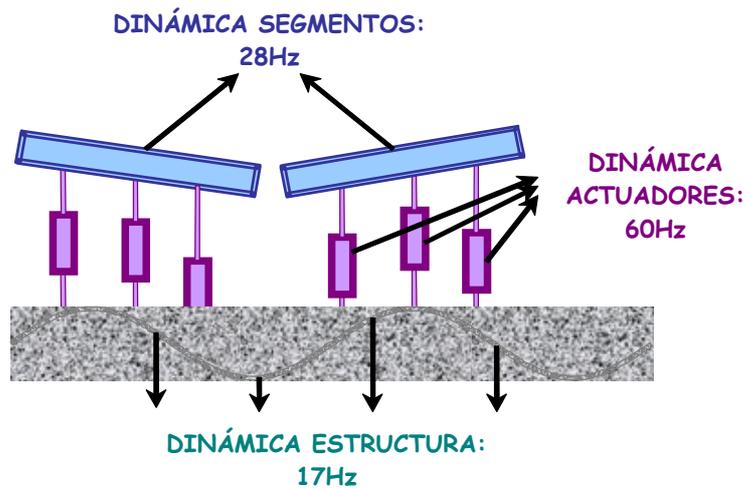


Figura 4: Elementos que componen el espejo primario del GTC.

El diagrama de bloques del sistema en lazo abierto es el que mostramos a continuación:

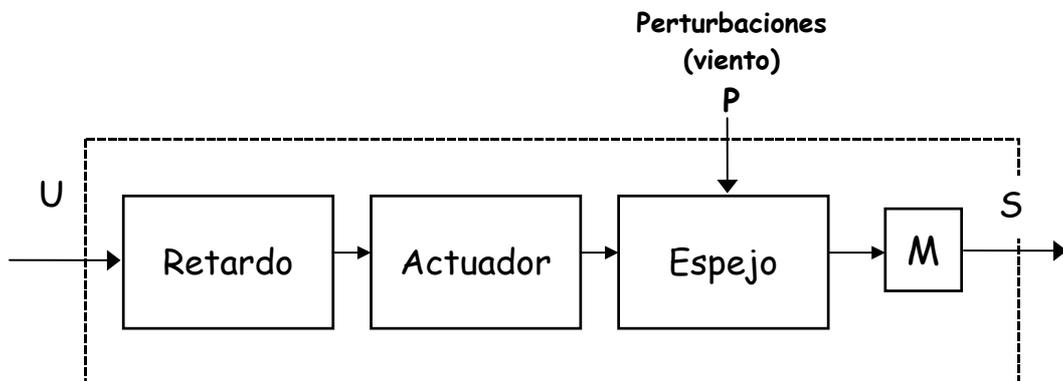


Figura 5: Diagrama de bloques del sistema en lazo abierto.

II.2. La dinámica del sistema. Desde el punto de vista dinámico, las características más destacables de este sistema son oscilaciones amortiguadas en la frecuencia dominante de la estructura (17hz) y del modo de los segmentos (28hz) y un tiempo de establecimiento algo superior a los dos segundos. En las siguientes gráficas se observa la respuesta del sistema en lazo abierto partiendo de condiciones iniciales distintas:

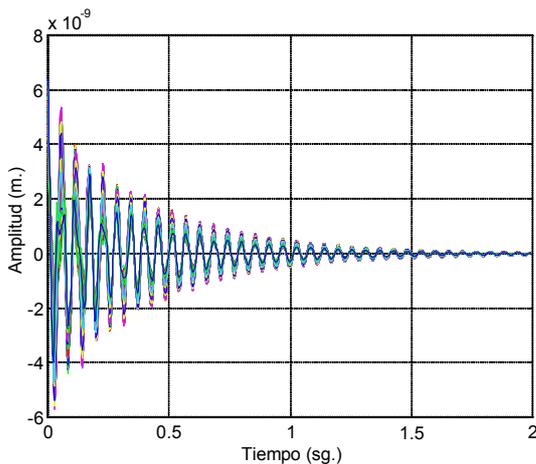


Figura 6: Evolución de los segmentos del espejo primario en lazo abierto partiendo del plano (4nm,5nm,6nm).

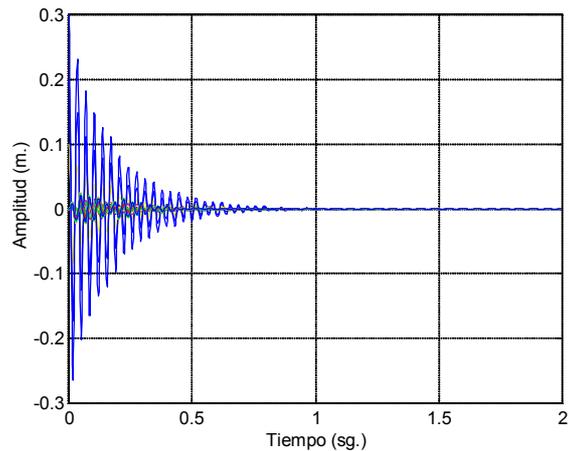


Figura 7: Evolución de los segmentos en lazo abierto cuando sólo 3 de ellos parten de fuera del equilibrio.

En la Figura 6 los segmentos parten de una posición inicial en la que están alineados y evolucionan como un espejo monolítico hacia el estado de equilibrio definido por el plano (0,0,0), por lo que no se excita el modo correspondiente a los segmentos (28hz) y el dominante es el de la estructura en 17hz, tal y como se aprecia en la Figura. Por el contrario, en la Figura 7 todos los segmentos parten de la posición de equilibrio excepto tres de ellos, es decir, la condición inicial no corresponde a un estado en el que todos los segmentos están alineados. Es por esto por lo que sí se excita el modo de 28hz, que es el dominante en este segundo caso.

III. Estrategia de control local-global

Diseñar un controlador multivariable para un sistema con un número tan elevado de entradas y salidas (108) es una labor compleja. Además, hay que tener en cuenta las restricciones con respecto a la frecuencia más alta a la que podemos aplicar el comando impuestas por la tecnología de buses empleada en el diseño del Telescopio. Por ello, se propone una estrategia de control local-global que consiste en aplicar al sistema conjuntamente dos acciones de control a distinto nivel: local y global, donde el comando local se puede aplicar a una frecuencia muy superior al global debido a que su cálculo involucra la consideración de la información suministrada por sólo tres de los sensores de posición de cada uno de los

segmentos. El esquema de aplicación conjunta de los comandos local y global se ilustra en la Figura 8:

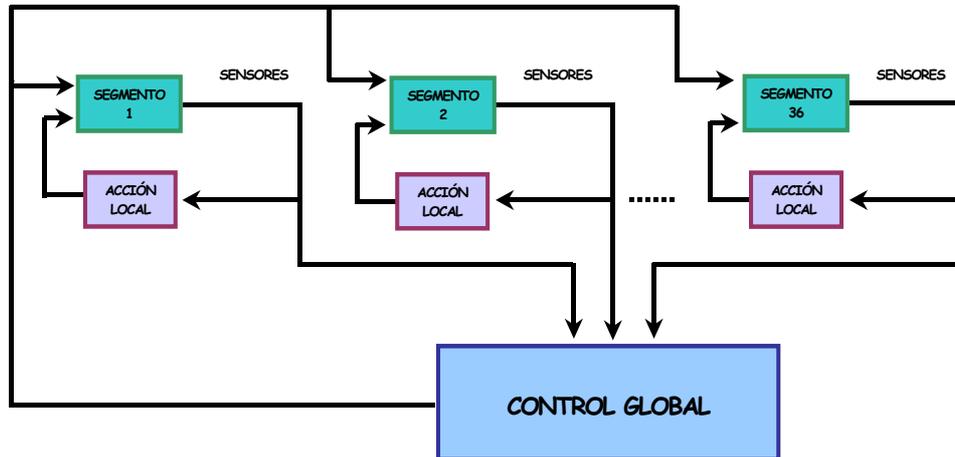


Figura 8: Esquema de los dos niveles de control: local y global.

Otro factor que dificulta en gran medida el diseño de un controlador para el sistema es el alto nivel de acoplamiento existente entre los distintos segmentos que componen el espejo.

Los objetivos que se plantean con un controlador multivariable local-global son dos: cancelar las oscilaciones del sistema consiguiendo, al mismo tiempo, que éste alcance el estado de consigna en el menor tiempo posible.

III.1. Un controlador local-global por asignación de polos. De entre los controladores multivariables implementados destaca uno por asignación de polos [Acosta, 1999] cuyo diseño permite, además, estudiar las simetrías presentes en el problema. El comando aplicado al sistema viene dado por la siguiente expresión:

$$U = -K_l * S - F_l * \dot{S} - K_g * x - F_g * \dot{x}$$

siendo:

$S(168 \times 1)$: señales provenientes de los 168 sensores,

$\dot{S}(168 \times 1)$: derivada 1ª en el tiempo de las señales provenientes de los 168 sensores,

$x(708 \times 1)$: vector de estados del sistema,

$\dot{x}(708 \times 1)$: derivada 1ª en el tiempo del vector de estados del sistema,

$K_l(108 \times 168), F_l(108 \times 168)$: parámetros del controlador local,

$K_g(108 \times 708), F_g(108 \times 708)$: parámetros del controlador global,

$U(108 \times 1)$: comandos que se aplican a los 108 actuadores.

En las Figuras 9 y 10 se muestra la respuesta del sistema en lazo cerrado con el controlador por asignación de polos para distintas frecuencias de los controladores local y global.

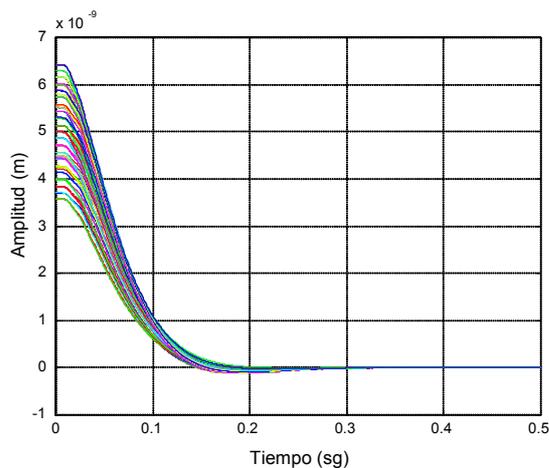


Figura 9: Evolución de los segmentos para una frecuencia local y global de 2000hz

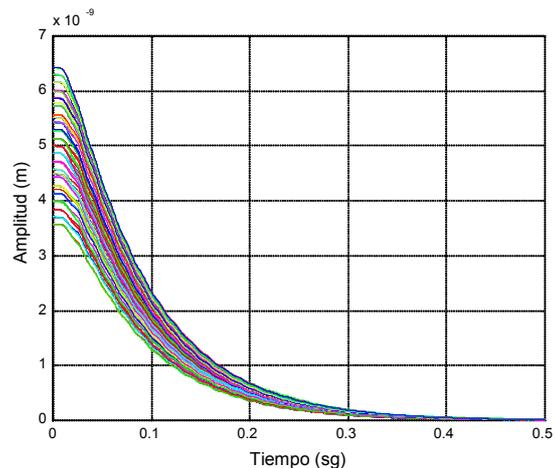


Figura 10: Evolución de los segmentos para una frecuencia local de 2000hz y una global de 200hz.

En estas dos Figuras se aprecia cómo la componente local del controlador por asignación de polos elimina por completo las oscilaciones en la respuesta de los segmentos y sólo al bajar la frecuencia de la acción global a 200hz se observa un rizado mínimo. Asimismo, vemos que el tiempo de establecimiento del espejo se reduce sensiblemente respecto al del lazo abierto, aunque éste aumenta a medida que disminuye la frecuencia de la acción global.

III.2. Estudio de la estabilidad del sistema con el controlador local-global. Operadores de lifting y lifting inverso. Para llevar a cabo el estudio de la estabilidad del sistema en lazo cerrado con el control local-global empleamos dos operadores denominados ‘de lifting’ y ‘de lifting inverso’ [Chen, 1995], que nos permiten expresar el sistema multifrecuencias en lazo cerrado en función de una única frecuencia de muestreo y, por tanto, aplicar los métodos clásicos para el análisis de la estabilidad de los sistemas lineales. Introduciendo estos operadores, el diagrama de bloques de nuestro sistema con el controlador local-global por asignación de polos queda como se muestra en la siguiente Figura:

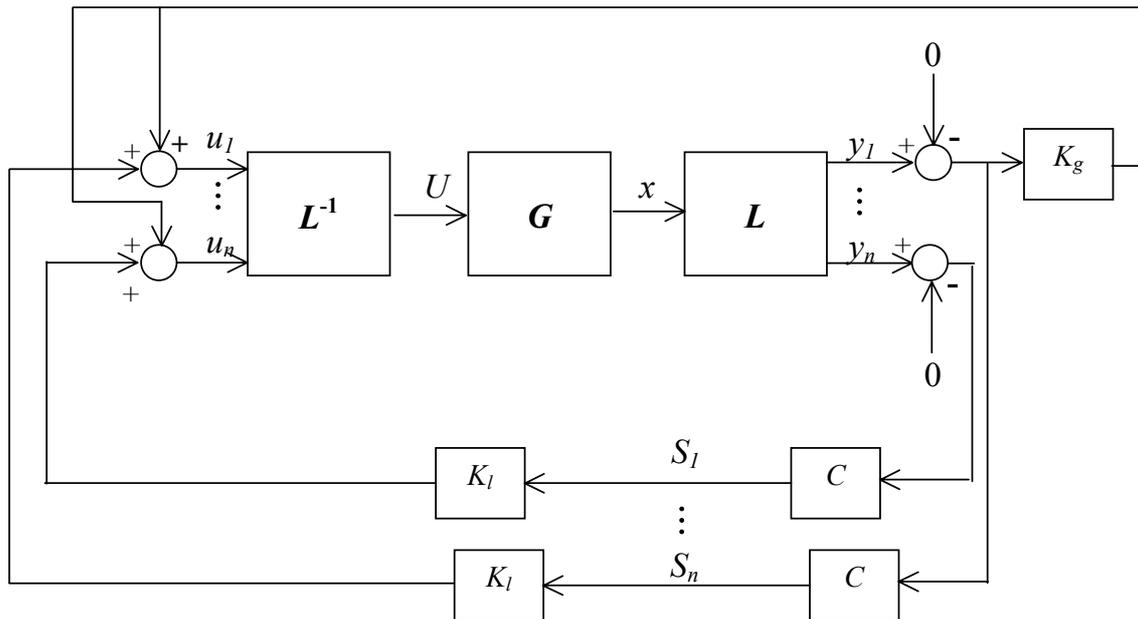


Figura 11: Diagrama de bloques del sistema en lazo cerrado con el controlador por asignación de polos introduciendo los operadores de lifting y lifting inverso.

Dicho estudio nos permite concluir que las frecuencias más bajas a las que se pueden aplicar los comandos local y global para que el sistema en lazo cerrado sea estable son 500 y 200hz, respectivamente.

IV. Desacoplo de la dinámica del espejo primario del GTC

Frente al diseño de un controlador multivariable para un sistema de larga escala [Jamshidi, 1983], [Patel, 1981], [Safonov, 1980], [Skogestad, 1998] y con la intención de alcanzar objetivos más ambiciosos desde el punto de vista del control, tales como robustez frente a incertidumbres en la dinámica del espejo, rechazo al ruido o ancho de banda, se impone reducir la dimensionalidad del problema de control al que nos enfrentamos.

IV.1. Procedimiento de desacoplo. Para llevar a cabo dicha reducción abordamos el problema del desacoplo de la dinámica del espejo primario del GTC [Acosta, 2000], [Acosta, 2001]. El procedimiento de desacoplo desarrollado consiste, básicamente, en un cambio apropiado de la base de vectores propios del sistema [Sigut, 2002] de forma tal que, una vez que dicho procedimiento se lleva a cabo, el sistema multivariable original con 108 entradas y salidas interconectadas entre sí se descompone en un sistema MIMO (con múltiples entradas y salidas) con $4N$ estados, N entradas y N salidas acopladas entre sí (donde N representa el número de modos que se considera para la descripción de la dinámica de la estructura del espejo) y un conjunto del 108- N subsistemas SISO (con una única entrada y una única salida) independientes entre sí. Nótese que en el caso particular de tomar $N=1$ el desacoplo del sistema es total aunque, en cualquier caso, la reducción que se consigue en el problema del diseño de un controlador para el espejo primario del GTC es notable.

IV.2. Diseño de controladores para el sistema desacoplado. Una vez que el sistema está desacoplado, el problema del control del espejo primario del GTC [Acosta, 2002] se reduce a diseñar una serie de controladores de baja dimensionalidad cuyos parámetros se ajustan de forma independiente como consecuencia de la ausencia de interacción entre las entradas y salidas de los distintos subsistemas resultantes del desacoplo [Siljak, 1991], [Skogestad, 1989]. En las gráficas que mostramos a continuación, se muestra la evolución de los segmentos del espejo en lazo cerrado cuando aplicamos al sistema un conjunto de 108 controladores SISO proporcionales, integrales y proporcionales-integrales [Åström, 1995]. Para ello hemos caracterizado la dinámica de la estructura del espejo con un único modo.

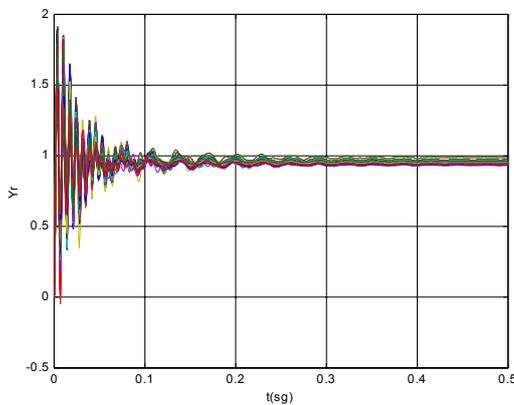


Figura 12: Respuesta de los segmentos con 108 controladores proporcionales.

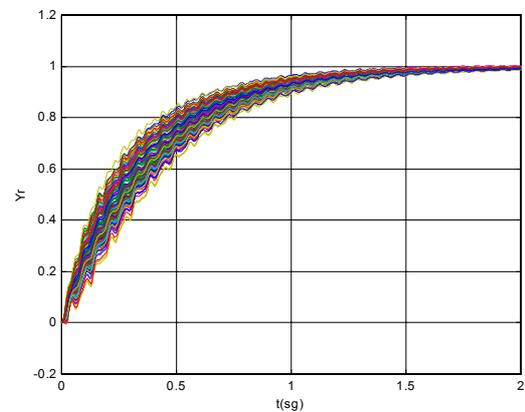


Figura 13: Respuesta de los segmentos con 108 controladores integrales.

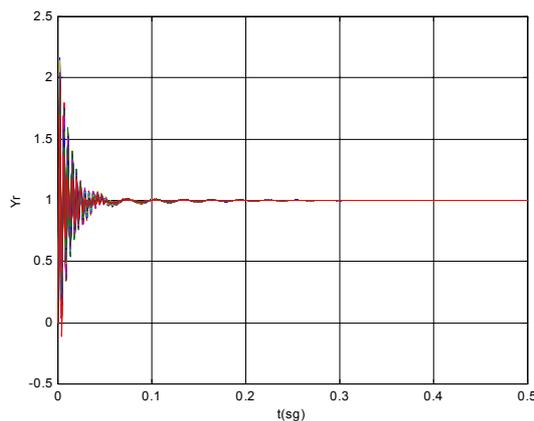


Figura 14: Respuesta de los segmentos con 108 controladores proporcionales-integrales.

El desacoplo del sistema no sólo hace más sencillo el diseño de un controlador o, mejor dicho, de un conjunto de controladores de baja dimensionalidad para el mismo, sino que posibilita la aplicación de algunas políticas de control, tales como el control robusto [Chiang, 1992], [Safonov, 1988], que resultan prohibitivas para el sistema multivariable original dado el tamaño de las matrices que describen la dinámica de la planta.

Bibliografía

- [Acosta, 1999] Acosta L., Sigut M., Hamilton A., Méndez J.A., Marichal G.N. and Moreno L. *Design of a Multivariable Pole-Placement Controller for the Primary Mirror of the 10m Grantecan Telescope*. Proceedings of the 7th Mediterranean Conference on Control and Automation. Haifa, Israel, Junio 1999.
- [Acosta, 2000] Acosta L., Sigut M., Hamilton A., Méndez J.A., Marichal G.N. and Moreno L. *Decoupling of the 10m GRANTECAN telescope's primary mirror dynamics and design of a controller for noise rejection*. Proceedings of the 2000 American Control Conference. Chicago, USA, Junio 2000.
- [Acosta, 2001] Acosta L., Sigut M., Méndez J.A., Hamilton A., Marichal G.N., Torres S. and Moreno L. *Rejecting the low-frequency noise in the 10m diameter primary mirror in the GRANTECAN telescope*. Proceedings of the 2001 IEEE International Conference on Control Applications. Ciudad de Méjico, Méjico, Septiembre 2001.
- [Acosta, 2002] Acosta L., Sigut M., Martín F., Méndez J.A. and Marichal G.N. *Noise Rejection Strategy for a Large Space Structure: The GRANTECAN Telescope*. Proceedings of the 15th IFAC World Congress on Automatic Control. Barcelona, España, Julio 2002.
- [Álvarez, 1997] Álvarez P. et al. *Gran Telescopio Canarias, Conceptual Design*. 1997.
- [Åström, 1995] Åström K.J. and Hagglund T. *PID Controllers: Theory, Design, and Tuning*. ISA - The Instrumentations, Systems, and Automation Society, 1995.
- [Chen, 1995] Chen T. and Francis B. *Optimal Sampled-Data Control Systems*. Springer-Verlag, London, 1995.
- [Chiang, 1992] Chiang R.Y. and Safonov M.G. *Robust Control Toolbox User's Guide*. Mathworks, South Natick, 1992.
- [Jamshidi, 1983] Jamshidi M. *Large Scale Systems Modelling and Control*. Nort-Holland, 1983.
- [Patel, 1981] Patel R. and Munro N. *Multivariable Control: Theory and Design*. Pergamon Press, 1981.
- [Safonov, 1980] Safonov M.G. *Stability and Robustness of Multivariable Feedback Systems*. The MIT Press, Cambridge, Massachusetts and London, 1980.
- [Safonov, 1988] Safonov M.G., Chiang R. and Flashner H. *H[∞] Robust Control Synthesis for a Large Space Structure*. Proceedings of the American Control Conference. Atlanta, GA, June 15-17, 1988.

- [Sigut, 2002] Sigut M., Acosta L., Martín F., Moreno L. and Hamilton A. *An Algebraic Approach Based on Symmetry Relations of State Space Matrices for Decoupling the Dynamics of a Large Space Structure*. Proceedings of the 2002 American Control Conference. Anchorage, USA, Mayo 2002.
- [Siljak, 1991] Siljak D. D. *Decentralized control of complex systems*. Academic Press, 1991.
- [Skogestad, 1989] Skogestad S. and Morari M. *Robust performance of decentralized control systems by independent designs*. *Automatica* **25**(1):119-125, 1989.
- [Skogestad, 1998] Skogestad S. and Postlethwaite I. *Multivariable feedback control*. John Wiley & Sons, Chichester, 1998.

En Internet

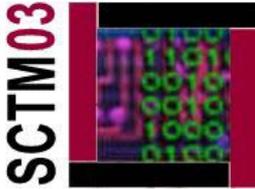
<http://www.cyc.dfis.ull.es>

Grupo de Computadoras y Control de la Universidad de La Laguna

<http://www.gtc.iac.es>

Proyecto Gran Telescopio Canarias

Seguridad matemática en la Sociedad de la Información



Pino T. Caballero Gil

Profesora Titular de Ciencias de la Computación e Inteligencia Artificial
Departamento de Estadística, Investigación Operativa y Computación
Universidad de La Laguna

Introducción

La ciencia de la Criptografía estudia los sistemas de cifrado y descifrado, la gestión de claves y toda una variedad de aplicaciones útiles en el medio hostil y vulnerable de las comunicaciones digitales electrónicas. Uno de los hitos más importantes de la larga y fascinante historia de la Criptografía ocurrió en 1976 cuando se introdujo el revolucionario concepto de la Clave Pública, que marcó el comienzo de la moderna época de la Criptografía de Clave Pública, frente a la época de la Criptografía de Clave Secreta. Mientras en este clásico tipo de sistemas, las operaciones de cifrado y descifrado son imposibles sin la clave secreta compartida por emisor y receptor, en la criptografía de clave pública existen dos piezas de información (claves privada y pública) tales que una de ellas (clave privada de descifrado) es imposible de obtener a partir de la otra (clave pública de cifrado). De esta sencilla manera se elimina la compleja tarea de la distribución de claves secretas.

La revolucionaria idea de la clave pública se puede llevar a cabo de forma sencilla gracias al análogo matemático de las calles de un sólo sentido, que son las funciones unidireccionales las cuales resultan fáciles de aplicar pero imposibles o difíciles de invertir. Así, la seguridad de los cifrados de clave pública reside en la dificultad o intratabilidad computacional de varios problemas matemáticos, por lo que se puede afirmar que toda la criptografía moderna se sustenta en el pilar de la Teoría de la Complejidad Computacional. La mayoría de cifrados de clave pública en la actualidad, y concretamente el que es el más difundido, el RSA, se apoya en la estructura matemática más socorrida en Criptografía que es el grupo multiplicativo de los enteros módulo un gran número primo, y se basa en la hipótesis de que factorizar números grandes es computacionalmente intratable. A este respecto merece la pena mencionar la computación cuántica cuyo estudio está en auge actualmente, ya que se ha probado que con ordenadores cuánticos es posible factorizar números grandes mucho más rápido que con ordenadores digitales convencionales. También merecen ser resaltados los sistemas basados en la dificultad del problema del logaritmo discreto en el grupo abeliano de una curva elíptica en un campo finito, ya que representan uno de los cifrados de clave pública más prometedores debido a su eficiencia y seguridad.

Firma Digital y Control de Accesos

Para el uso seguro de criptografía de clave pública es importante garantizar la autenticidad de las claves públicas. Una elegante solución a este problema viene de la mano de la firma digital, que puede verse como la analogía electrónica de la firma manual. Este tipo de esquemas puede basarse de forma sencilla en la criptografía de clave pública sin más que considerar que una usuaria A (*Alice*) siempre puede firmar un documento mediante su propia clave secreta, de forma que otro usuario B (*Bob*) siempre puede validar dicha firma usando la

clave pública de A. De esta forma, la aplicación de la firma digital en la certificación de claves públicas se basa en unas Terceras Partes de Confianza conocidas como Autoridades de Certificación cuya misión es firmar digitalmente certificados de autenticidad de las claves públicas de los usuarios.

En la práctica, cuando se pretende firmar un mensaje digital hay que considerar que normalmente éste tiene una longitud variable mientras que la entrada al algoritmo de firma suele ser de longitud fija, y por lo general más corta que el mensaje. Para resolver este problema se recurre al uso de funciones *hash*, que se han utilizado tradicionalmente en Ciencias de la Computación sobre todo para procedimientos de búsqueda. Esta solución consiste en aplicar una función hash al mensaje antes de aplicar el algoritmo de cifrado. Dicha función hash debe ser una función unidireccional, simple y eficiente que convierta cualquier mensaje de longitud arbitraria en uno de longitud fija, sea computacionalmente imposible encontrar dos mensajes con igual resumen, y tal que pequeños cambios en el mensaje producen cambios significativos en el resumen. El ataque por búsqueda exhaustiva contra las funciones hash es inabordable para las que producen resúmenes de tamaño 128, ya que por término medio habría que probar 2^{64} mensajes antes de conseguir uno que genere un resumen predeterminado. Sin embargo estas funciones sí son sensibles al ataque del cumpleaños (llamado así por la conocida paradoja probabilística del cumpleaños) ya que sólo se necesita generar $2^{n/2}$ mensajes para obtener dos que generen el mismo resumen.

En cuanto a la identificación para el control de accesos a sistemas, los métodos existentes tienen grados de seguridad muy diferentes. El método más simple es el habitual sistema de contraseñas fijas, que constituye el método de identificación más débil. Dado que el sistema realmente no necesita almacenar las contraseñas, sino sólo diferenciar las válidas de las inválidas, puede mejorarse un poco el esquema sin más que aplicar directamente una función unidireccional sobre las contraseñas introducidas de manera que el sistema sólo necesite almacenar las imágenes para compararlas. Por otra parte están los métodos de identificación fuerte basados en un protocolo criptográfico conocido como Demostración de Conocimiento Nulo. Este tipo de protocolos impiden posibles ataques de suplantación ya que permiten a los usuarios demostrar su conocimiento de la contraseña secreta sin revelar ninguna información sobre ella. Se trata de protocolos interactivos en los que se realiza un intercambio de mensajes (retos aleatorios y respuestas), tras el cual el sistema acepta o rechaza la demostración de conocimiento con una cierta probabilidad de acierto. Las dos demostraciones de conocimiento nulo más estudiadas se basan en dos de los problemas más utilizados en criptografía, la residuosidad cuadrática y los logaritmos discretos.

Protocolos Criptográficos

En general los protocolos criptográficos se pueden definir como algoritmos utilizados por dos o más participantes con una meta común, e implementados en entornos distribuidos inseguros. Formalmente un protocolo k-partito es una función pública $f: (\{0,1\}^*)^k \rightarrow (\{0,1\}^*)^k$ computable en tiempo polinomial, tal que la entrada es información privada de cada uno de los k participantes, y la salida es un vector de los valores que requieren los participantes para resolver el problema. A continuación describimos algunos de los algoritmos más representativos de cada uno de los protocolos estudiados, observando la frecuencia con que se utilizan como base problemas matemáticos como el de la factorización, la residuosidad cuadrática o el logaritmo discreto, todos ellos provenientes de la Teoría de Números.

En los protocolos básicos conocidos como de Transferencia Inconsciente la situación a resolver es la siguiente: A quiere transferir un secreto a B de forma que dicha información se

transfiera con probabilidad $\frac{1}{2}$, y que al final B sepa con certeza si la ha recibido, pero A no. Uno de los algoritmos de Transferencia Inconsciente más sencillos y conocidos es el Protocolo de Rabin, en el que el secreto que se pretende transferir inconscientemente es la factorización del producto de dos grandes números primos, y que se basa en el problema de la residuosidad cuadrática. Dicho protocolo se puede describir como sigue:

1. A escoge al azar dos grandes números primos p y q , y calcula y y envía a B, $N=p \cdot q$.
2. B escoge un entero x al azar entre 1 y $N-1$ y primo con N , y envía $x^2 \pmod{N}$ a A.
3. A calcula gracias a su conocimiento de p y q , las cuatro raíces cuadradas diferentes de $x^2 \pmod{N}$, $\{x, N-x, y, N-y\}$, escoge una al azar y se la envía a B.
4. Si B recibe y o $N-y$, entonces puede calcular p y q gracias a que $\text{m.c.d.}((x+y), N)$ es p o q . Si B recibe x o $N-x$, entonces no puede.

Otro algoritmo de Transferencia Inconsciente, esta vez basado en un criptosistema de clave pública es el Protocolo de Pflieger, que se define de la forma siguiente:

1. A genera dos parejas de claves pública y privada.
2. B escoge una clave secreta propia k_B y una de las dos claves públicas de A al azar, cifrando con ésta su clave k_B y enviando el resultado a A.
3. A escoge al azar una de sus dos claves privadas, descifra con ella el mensaje recibido, cifra con el resultado de este descifrado un mensaje secreto y envía lo que obtiene a B.
4. Si la clave elegida por A en el paso 3 se corresponde con la clave pública escogida por B en el paso 2, B podrá descifrar el mensaje secreto de A. Si no, no podrá.

Los protocolos de Compromiso de Bits configuran la segunda piedra angular del área de los protocolos criptográficos. El objetivo de los esquemas de Compromiso de Bits es el siguiente: A se quiere comprometer frente a B con un valor, de forma que A no pueda cambiarlo, y B no pueda descubrir el valor hasta que A abra el compromiso. En la definición de estos esquemas es habitual plantear como analogía un sobre cerrado, que resulta a la vez inalterable e ilegible. Además, la apertura del compromiso puede verse como una correspondencia definida desde un gran dominio sobre el conjunto binario $\{0,1\}$, de forma que un bit se considera comprometido cuando el origen de la correspondencia para ese valor de salida puede ser cualquier elemento aleatorio. El esquema definido a partir de dicha correspondencia es inalterable cuando la correspondencia es función, es decir cuando cada elemento posee una única imagen, y es ilegible cuando las distribuciones del conjunto origen del 0 y las del conjunto origen del 1 son indistinguibles. Existen muchos algoritmos interesantes de Compromiso de Bits basados en diversas herramientas criptográficas típicas tales como cifrados de clave secreta, logaritmos discretos, o residuos cuadráticos. El primer esquema analizado a continuación es una propuesta de Compromiso de Bits basada en la otra primitiva criptográfica de Transferencia Inconsciente. En todos los algoritmos se repiten los pasos de compromiso, apertura y verificación, y están garantizadas las propiedades de ilegibilidad e inalterabilidad. A continuación vemos brevemente la definición de cada uno de los esquemas con el objetivo de comprometer en cada caso un bit b .

En el Compromiso de Bits basado en Transferencia Inconsciente que se describe a continuación la probabilidad de fraude de A es $\leq \frac{1}{2}$, pero esta cota puede bajarse a 2^{-m} si se ejecuta m veces el algoritmo.

1. Compromiso: A escoge n bits aleatorios b_i , tales que $b_1+b_2+\dots+b_n=b$, y envía cada b_i por orden mediante Transferencia Inconsciente.
2. Apertura: A envía a B los bits b_i
3. Verificación: B compara los bits b_i recibidos en el paso de compromiso con los correspondientes de la apertura.

El siguiente esquema de Compromiso de Bits se basa en un Cifrado de Clave Secreta en el que en primer lugar A y B acuerdan una secuencia pseudoaleatoria R:

1. Compromiso: A cifra con su clave secreta k , la secuencia R y el bit b , y envía a B el resultado $E_k(R,b)$.
2. Apertura: A envía a B la clave k .
3. Verificación: B descifra el mensaje, comprueba la secuencia R, y descubre el bit b .

El siguiente esquema de Compromiso de Bits está basado en el problema de los Logaritmos Discretos. En primer lugar A y B acuerdan un número primo p , un elemento a generador de Z_p y un elemento aleatorio s de Z_p :

1. Compromiso: A escoge al azar un entero y entre 0 y $p-2$, y calcula y envía a B el número $x=s^b a^y \pmod{p}$.
2. Apertura: A envía a B el entero y .
3. Verificación: B obtiene b y comprueba x .

El último esquema descrito se basa en el problema de la Residuosidad Cuadrática. Primero A y B acuerdan $n=p*q$, siendo $p=q=3 \pmod{4}$ información secreta de A:

1. Compromiso: A genera y envía a B:
 - a. un residuo cuadrático mod n , si $b=0$.
 - b. un no residuo cuadrático mod n , si $b=1$.
2. Apertura: A envía a B la factorización (p,q) de n .
3. Verificación: B obtiene b y comprueba si el número enviado es residuo o no.

Tal y como se puede observar el primer esquema descrito tiene la ventaja de requerir únicamente una comunicación unidireccional, mientras que en el resto se hace necesario un contacto previo entre las partes para acordar determinadas piezas de información que se usan durante el desarrollo del correspondiente protocolo.

Los protocolos bipartitos conocidos como Firmas de Contratos se caracterizan por que A y B quieren firmar simultáneamente un contrato a través de una red de comunicaciones de forma que ninguno pueda obtener la firma del otro sin haber firmado el contrato, garantizando además que ninguno pueda repudiar su propia firma. En un intento simple de solución al problema, A y B se podrían intercambiar alternativamente bits de sus firmas digitales del contrato, de forma que si uno interrumpe el proceso, ambos tienen prácticamente la misma cantidad de firma del otro. El inconveniente de esta solución es que si uno envía basura en lugar de su firma, el otro no lo detecta hasta el final. De hecho, en general se ha demostrado que es imposible diseñar un protocolo determinístico que no requiera la participación de una tercera parte de confianza. Por tanto, el diseño de una Firmas de Contratos independiente de terceras partes de confianza ha de basarse en un proceso de aleatorización. En general las Firmas de Contratos aleatorizadas usan esquemas de Compromiso de Bits y se basan en Transferencias

Inconscientes. De hecho, concretamente cualquier protocolo de Transferencia Inconsciente puede ser adaptado a una Firma de Contratos sin más que aplicar la Transferencia Inconsciente sucesivamente, y considerar el contrato firmado al final de la ejecución si ambos usuarios logran conocer la información secreta del otro.

Los Lanzamientos de Monedas son protocolos bipartitos en los que se trata de generar una secuencia aleatoria común a dos usuarios de forma que el que actúa como generador de la secuencia no pueda elegir una secuencia no aleatoria particular, y a la vez el otro usuario no pueda anticipar el resultado del lanzamiento. La aplicación práctica más inmediata de este tipo de protocolos se encuentra en la generación de claves secretas pseudoaleatorias compartidas entre dos usuarios, imprescindibles en los sistemas de clave secreta; aunque también los Lanzamientos de Monedas resultan útiles a la hora de diseñar protocolos multipartitos tales como el Póquer Mental. Se puede diseñar un Lanzamiento de Monedas a partir de cualquier Transferencia Inconsciente de manera que el usuario B gana si logra recibir el secreto transferido. También es posible definir un Lanzamiento de Monedas basándose en un esquema cualquiera de Compromiso de Bits. En este caso A y B escogen al azar sendos bits aleatorios a y b , y se los intercambian mediante Compromiso de Bits. Tras las fases de apertura y verificación de los compromisos, ambos participantes toman como resultado del lanzamiento $a+b$. Aparte de los esquemas generales mencionados, basados en Transferencia Inconsciente y Compromiso de Bits, existen otras muchas propuestas basadas en herramientas criptográficas típicas como funciones hash, criptografía de clave secreta, o el problema de la residuosidad cuadrática, por ejemplo.

En los esquemas multipartitos de Compartición de Secretos un secreto se divide en partes (sombras) que se distribuyen entre varios participantes de forma que sólo cuando un número de ellas (no necesariamente todas) se reúnen es posible reconstruir el secreto. Si P es el conjunto de sombras, se llaman estructuras de acceso $\mathbb{G}2^P$ a los distintos subconjuntos de sombras que permiten calcular el secreto, y agrupación autorizada a cada uno de esos subconjuntos. Un esquema perfecto es el que permite a las agrupaciones autorizadas obtener el secreto, y no permite a ninguna agrupación no autorizada conseguir ninguna información sobre el secreto. Se ha demostrado que siempre se puede diseñar un esquema perfecto a partir de cualquier estructura de acceso. La base del esquema es el subconjunto de agrupaciones autorizadas minimales, y la tasa de información mide la cantidad de información que tienen los participantes. El esquema de Compartición de Secretos más conocido es el llamado esquema umbral (t, w) (con $t \leq w$) en el que un secreto k puede reconstruirse a partir de cualesquiera t de las w sombras, y k no puede reconstruirse mediante ningún subconjunto de $t-1$ o menos sombras. El esquema umbral más conocido es el Esquema de Shamir, que se basa en la interpolación polinomial para reconstruir una curva de grado $t-1$ a partir de t puntos. El siguiente esquema más conocido se conoce como Esquema Umbral de las Sombras Congruentes, y se basa en el Teorema de los Restos Chinos Congruentes. Una de las principales aplicaciones de la Compartición de Secretos es la Criptografía Visual, usada con el propósito de cifrar una imagen que podrá ser reconstruida mediante la superposición de un número mínimo de sombras.

Conclusiones

Hemos analizado aquí diversas aplicaciones criptográficas que permiten resolver en el mundo digital varios problemas habituales en el mundo real, llegando a mejorar en muchos casos las condiciones de seguridad. Hemos estudiado varios protocolos diseñados a partir de herramientas matemáticas para tareas específicas diferentes, aunque íntimamente relacionados

en su diseño. De hecho, se han investigado tanto las relaciones generales como las provenientes de diseños particulares entre los diferentes protocolos analizados.

Bibliografía

- P. Caballero Gil, C. Hernández Goya, C. Bruno Castañeda: *Cryptographic Applications*. Computational Mathematics, Narosha Publishing House, New Delhi, India, 2002.
B. Schneier: *Applied Cryptography*. J. Wiley & Sons, 1993.

En Internet

<http://www.kriptopolis.com>

Kriptópolis

Revista independiente sobre privacidad y seguridad en Internet.

<http://www.iec.csic.es/criptonomicon>

Criptonomicón

Página del Instituto de Física Aplicada del CSIC dedicada a la criptografía y la seguridad en Internet.

<http://www.criptored.upm.es>

CriptoRed

Red Temática Iberoamericana de Criptografía y Seguridad de la Información.

Diseño Geométrico Asistido por Ordenador

SCTM03



Laureano González Vega

Catedrático de Álgebra del Departamento de Matemáticas, Estadística y Computación y Decano de la Facultad de Ciencias de la Universidad de Cantabria

<http://frisco.matesco.unican.es/~gvega>

Resumen

La utilidad de los sistemas de CAD/CAM como un medio para incrementar la eficiencia en los procesos de simulación y diseño dentro del sector productivo es en la actualidad irrefutable. Ventajas como la reducción en tiempo de producción, mejora en la calidad del producto final y reducción de costes al disminuir el tiempo de implementación de cambios en el proceso de diseño son frecuentemente citadas como los mayores beneficios que produce la introducción de los sistemas de CAD/CAM en un entorno industrial.

Utilizando el Diseño Geométrico Asistido por Ordenador como línea argumental, este minicurso intenta proporcionar una introducción básica en esta disciplina que se encuentra en la intersección de las Matemáticas, la Informática y la Ingeniería junto con la presentación de cómo algunas técnicas algebraicas pueden ser de gran utilidad a la hora de resolver de forma satisfactoria problemas de manipulación de curvas y superficies en un entorno CAD/CAM.

1. Introducción

En este minicurso se persiguen tres objetivos bien definidos:

- La introducción de los conceptos y técnicas básicas en Diseño Geométrico Asistido por Ordenador.
- La presentación e integración de cómo las técnicas algebraicas de manipulación de los conjuntos solución de sistemas de ecuaciones algebraicas que proporciona el Álgebra Computacional pueden ser aplicadas a la resolución eficiente de problemas en la manipulación de curvas y superficies paramétricas.
- El estudio y potenciación de las herramientas gráficas que proporcionan los Sistemas de Cálculo Simbólico de propósito general (Maple en nuestro caso particular) junto con el desarrollo, en este sistema, de módulos de simulación específicos para modelado geométrico y visualización.

Este minicurso se centra en mostrar, primero, cómo técnicas procedentes del Álgebra Computacional pueden ser de gran utilidad en la resolución de problemas en Diseño Geométrico Asistido por Ordenador y, segundo, cómo cualquier Sistema de Cálculo Simbólico (Maple en nuestro caso) puede ser utilizado como un sistema de simulación donde probar la eficiencia y validez de métodos y algoritmos en Diseño Geométrico Asistido por Ordenador. Asimismo se mostrará cómo integrar de forma sencilla estas técnicas algebraicas en un entorno CAD/CAM como puede ser ACIS (ver <http://www.spatial.com>), producto de libre distribución en instituciones educativas.

2. Introducción al Diseño Geométrico

Las entidades B-spline (y Bézier) representan actualmente un standard industrial para la representación, el diseño y el intercambio de datos de informaciones geométricas mediante medios informáticos. La sección comienza con las definiciones y propiedades generales de las curvas Bézier y B-spline (racionales y polinomiales).

Si N es un número entero positivo entonces los **polinomios de Bernstein** $B_{i,(N)}$ de orden N (grado $N - 1$) se definen de la siguiente manera:

$$B_{i,(N)}(u) = \binom{N-1}{i} u^i (1-u)^{N-1-i}, \quad i \in \{0, 1, \dots, N-1\}.$$

Una **curva Bézier racional** de orden N (grado $N - 1$), $C_N(u)$, se define como:

$$C_N(u) = \frac{\sum_{i=0}^{N-1} B_{i,(N)}(u) w_i P_i}{\sum_{i=0}^{N-1} B_{i,(N)}(u) w_i}, \quad 0 \leq u \leq 1,$$

donde $\{P_i : 0 \leq i \leq N - 1\}$ representan los puntos de control (que forman el polígono de control), $\{w_i : 0 \leq i \leq N - 1\}$ representan los pesos $\{w_i \geq 0 : 0 \leq i \leq N - 1\}$ y $\{B_{i,(N)} : 0 \leq i \leq N - 1\}$ son los polinomios de Bernstein de orden N . Si todos los w_i son iguales y distintos de cero entonces la curva Bézier es una curva **polinomial**.

Se introducen a continuación las curvas B-spline, lo que requiere la definición de las funciones básicas. Sea $U = \{u_0, \dots, u_m\}$ una sucesión no decreciente de números reales ($u_i \leq u_{i+1}$, $i \in \{0, \dots, m - 1\}$), que se denominan nodos. Las **funciones básicas** de orden N se definen recursivamente sobre el vector de nodos U de la siguiente forma:

$$B_{i,1}(u) = \begin{cases} 1 & \text{si } u_i \leq u < u_{i+1} \\ 0 & \text{en otro caso} \end{cases}$$

$$B_{i,N}(u) = \frac{u - u_i}{u_{i+N-1} - u_i} B_{i,N-1}(u) + \frac{u_{i+N} - u}{u_{i+N} - u_{i+1}} B_{i+1,N-1}(u).$$

Se observa que en la definición recursiva anterior pueden aparecer cocientes $0/0$ que, por convenio, se definen como 0. Las funciones $B_{i,N}(u)$ son polinomios a trozos, definidos sobre \mathbb{R} , pero solamente el intervalo $[u_0, u_m]$ presenta interés. Se estudian a continuación las propiedades más importantes de las funciones básicas que determinaran las características geométricas de las curvas y superficies B-spline.

Una **curva B-spline racional** de orden N , $C_N(u)$, se define como:

$$\mathbf{C}_N(u) = \frac{\sum_{i=0}^n B_{i,N}(u)w_i\mathbf{P}_i}{\sum_{i=0}^n B_{i,N}(u)w_i}, \quad a \leq u \leq b,$$

donde $\{\mathbf{P}_i : 0 \leq i \leq n\}$ representan los puntos de control (que forman el polígono de control), $\{w_i : 0 \leq i \leq n\}$ representan los pesos $\{w_i \geq 0 : 0 \leq i \leq n\}$ y $\{B_{i,N}(u) : 0 \leq i \leq n\}$ son las funciones básicas de orden N definidas sobre el vector no periódico de nodos

$$U = \{u_0, \dots, u_{n+N}\}$$

con $u_0 = u_1 = \dots = u_{N_1-1} = a$ y $u_{n_1} = u_{n_1+1} = \dots = u_{n_1+N_1-1} = b$. Si todos los w_i son iguales y distintos de cero entonces la curva B-spline es una curva **polinomial**. Si el vector de nodos U no tiene nodos internos (i.e. si $n = N - 1$) entonces la curva B-spline (racional) se convierte en una curva Bézier (racional).

Definiendo las funciones básicas racionales

$$R_{i,N}(u) = \frac{B_{i,N}(u)w_i}{\sum_{j=0}^n B_{j,N}(u)w_j}$$

se puede reescribir la definición de curva B-spline de la siguiente forma:

$$\mathbf{C}_N(u) = \sum_{i=0}^n R_{i,N}(u)\mathbf{P}_i, \quad a \leq u \leq b.$$

A continuación se presentan una serie de propiedades de las curvas B-spline, que se derivan de las propiedades de las funciones básicas:

- $\mathbf{C}_N(0) = \mathbf{P}_0$, $\mathbf{C}_N(1) = \mathbf{P}_n$.
- $R_{i,N}(u) = 0$ si $u \notin [u_i, u_{i+N})$.
- En cualquier intervalo $[u_j, u_{j+1})$ a lo sumo N de las funciones básicas racionales $R_{i,N}(u)$ son no nulas: $R_{j-N+1,N}(u), \dots, R_{j+1,N}(u)$ y $R_{j,N}(u)$.
- Las derivadas de una función básica racional $R_{i,N}(u)$ están definidas en el interior de cada intervalo de nodos distintos. En un nodo u_j , $R_{i,N}(u)$ (y por lo tanto $\mathbf{C}_N(u)$) es derivable $N - 1 - k$ veces, donde k es la multiplicidad del nodo u_j .
- Aplicar una transformación afín a una curva B-spline racional se reduce a la aplicación de dicha transformación a todos sus puntos de control.

- Si $u \in [u_i, u_{i+1})$ entonces $C_N(u)$ está dentro de la envolvente convexa generada por los puntos de control $\mathbf{P}_{i-N+1}, \dots, \mathbf{P}_i$.
- El número de intersecciones de cualquier plano (en el caso de las curvas tridimensionales) o cualquier recta (en el caso de las curvas bidimensionales) con una curva B-spline racional está acotado superiormente por el número de intersecciones del plano (o de la recta) con el polígono de control.
- Si se modifican los valores del punto de control \mathbf{P}_i o del peso w_i , la modificación se refleja sólo en aquella parte de la curva correspondiente al intervalo $[u_i, u_{i+N})$.

Las curvas B-spline racionales (al ser curvas racionales con un denominador común) tienen una interpretación geométrica elegante que determina una manera eficiente de almacenarlas y procesarlas. Así, se utilizarán las coordenadas homogéneas para representar una curva B-spline racional en el espacio n -dimensional como una curva B-spline polinomial en el espacio $(n + 1)$ -dimensional. Esta presentación permite derivar algoritmos muy eficientes de cálculo de las derivadas de las curvas B-spline racionales.

Se finaliza la sección presentando las definiciones y propiedades generales de las superficies Bézier y B-spline (racionales y polinomiales). **Una superficie Bézier racional** de orden (N_1, N_2) se define como

$$S_{N_1, N_2}(u, v) = \frac{\sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} B_{i, (N_1)}(u) B_{j, (N_2)}(v) w_{i, j} \mathbf{P}_{i, j}}{\sum_{i=0}^{N_1-1} \sum_{j=0}^{N_2-1} B_{i, (N_1)}(u) B_{j, (N_2)}(v) w_{i, j}}, \quad a \leq u \leq b, \quad c \leq v \leq d,$$

donde $\{\mathbf{P}_{i, j} : 0 \leq i \leq N_1 - 1, 0 \leq j \leq N_2 - 1\}$ representan los puntos de control,

$$\{w_{i, j} : 0 \leq i \leq N_1 - 1, 0 \leq j \leq N_2 - 1\}$$

representan los pesos $\{w_{i, j} \geq 0 : 0 \leq i \leq N_1 - 1, 0 \leq j \leq N_2 - 1\}$, y

$$\{B_{i, (N_1)} : 0 \leq i \leq N_1 - 1\},$$

$$\{B_{j, (N_2)} : 0 \leq j \leq N_2 - 1\}$$

son los polinomios de Bernstein de orden N_1 y N_2 , respectivamente. Si todos los $w_{i, j}$ son iguales y distintos de cero entonces la superficie Bézier es una superficie **polinomial**.

Una superficie B-spline racional de orden (N_1, N_2) se define como

$$S_{N_1, N_2}(u, v) = \frac{\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} B_{i, N_1}(u) B_{j, N_2}(v) w_{i, j} \mathbf{P}_{i, j}}{\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} B_{i, N_1}(u) B_{j, N_2}(v) w_{i, j}}, \quad a \leq u \leq b, \quad c \leq v \leq d,$$

donde $\{\mathbf{P}_{i,j} : 0 \leq i \leq n_1 - 1, 0 \leq j \leq n_2 - 1\}$ representan los puntos de control,

$$\{w_{i,j} : 0 \leq i \leq n_1 - 1, 0 \leq j \leq n_2 - 1\}$$

representan los pesos $\{w_{i,j} \geq 0; 0 \leq i \leq n_1 - 1, 0 \leq j \leq n_2 - 1\}$, y $\{B_{i,N_1} : 0 \leq i \leq n_1\}$, $\{B_{j,N_2} : 0 \leq j \leq n_2\}$ son las funciones básicas de orden N_1 y N_2 respectivamente, definidas sobre los vectores no periódicos de nodos

$$U = \{u_0, \dots, u_{n_1+N_1}\} \text{ y } V = \{v_0, \dots, v_{n_2+N_2}\}$$

respectivamente, con

$$\begin{aligned} u_0 &= \dots = u_{N_1-1} = a, \quad u_{n_1} = \dots = u_{n_1+N_1-1} = b, \\ v_0 &= \dots = v_{N_2-1} = c, \quad v_{n_2} = \dots = v_{n_2+N_2-1} = d. \end{aligned}$$

Si todos los $w_{i,j}$ son iguales y distintos de cero, entonces la superficie B-spline es una superficie **polinomial**.

A continuación se presentan una serie de propiedades de las superficies B-spline racionales, que se derivan de las propiedades de las funciones básicas y que son similares a las propiedades ya estudiadas de las curvas B-spline racionales:

- $\mathbf{S}_{N_1, N_2}(0, 0) = \mathbf{P}_{0,0}$ y $\mathbf{S}_{N_1, N_2}(1, 0) = \mathbf{P}_{n_1,0}$
- $\mathbf{S}_{N_1, N_2}(0, 1) = \mathbf{P}_{0,n_2}$ y $\mathbf{S}_{N_1, N_2}(1, 1) = \mathbf{P}_{n_1, n_2}$
- Si $n_1 = N_1 - 1$ y $n_2 = N_2 - 1$ y U y V son vectores normalizados entonces $\mathbf{S}_{N_1, N_2}(u, v)$ es una superficie Bézier.
- Aplicar una transformación afín a una superficie B-spline racional se reduce a la aplicación de dicha transformación a todos los puntos de control.
- Si $(u, v) \in [u_i, u_{i+1}] \times [v_j, v_{j+1}]$ entonces $\mathbf{S}_{N_1, N_2}(u, v)$ está dentro de la envolvente convexa generada por los puntos $\mathbf{P}_{k,l}$, $i - N_1 + 1 \leq k \leq i$, $j - N_2 + 1 \leq l \leq j$.
- Si se modifican los valores del punto de control $\mathbf{P}_{i,j}$ o del peso $w_{i,j}$, la modificación se refleja solamente en aquella parte de la superficie correspondiente al rectángulo $[u_i, u_{i+N_1}] \times [v_j, v_{j+N_2}]$.

Como ocurría con las curvas B-spline, la utilización de coordenadas homogéneas para representar superficies B-spline es de gran utilidad computacional cuando se desean calcular derivadas o realizar otras operaciones de carácter computacional.

3. Implicitación de curvas y superficies racionales

Uno de los problemas fundamentales en el manejo de curvas y superficies paramétricas en Diseño Geométrico Asistido por Ordenador es la búsqueda de algoritmos eficientes de implicitación de curvas y superficies parametrizadas mediante funciones racionales. Aunque la representación paramétrica es la más adecuada para el trazado, la representación implícita debe ser considerada, por ejemplo, para decidir de forma eficiente la posición de un punto dado con respecto a una curva o superficie considerada.

Se notará aquí que el proceso de implicitación, para una curva o superficie B-spline, será aplicable sólo a cada uno de los trozos de ésta donde la función racional que proporciona la parametrización está determinada de forma unívoca. Sin embargo, la disposición del polígono de control y de la parametrización, junto con la ecuación implícita, motiva la reciente introducción en Diseño Geométrico Asistido por Ordenador de lo que se conocen como A-splines o, en otras palabras, los splines implícitos que están resultando ser una herramienta muy útil en lo que se denomina modelado de superficies con forma libre (ver [1]).

El problema de la implicitación, en su formulación algebraica más general, se puede plantear en los siguientes términos (salvo que se diga lo contrario \mathbb{K} denota un cuerpo cualquiera, pero nuestro caso principal es $\mathbb{K} = \mathbb{R}$):

Dados $p_1, \dots, p_n, q_1, \dots, q_n \in \mathbb{K}[y_1, \dots, y_m]$, con $q_i \neq 0$ y $\text{mcd}(p_i, q_i) = 1$, el problema de la implicitación consiste en encontrar un subconjunto \mathcal{F} de $\mathbb{K}[x_1, \dots, x_n]$ tal que el conjunto

$$S = \left\{ \left(\frac{p_1(b)}{q_1(b)}, \dots, \frac{p_n(b)}{q_n(b)} \right) \mid b \in \mathbb{K}^m, q_j(b) \neq 0 \right\}$$

se expresa implícitamente como el conjunto de los ceros comunes a los polinomios de \mathcal{F} .

Como no es posible determinar \mathcal{F} en $\mathbb{K}[x_1, \dots, x_n]$ tal que $V(\mathcal{F}) = S$, puesto que algunos puntos de $V(\mathcal{F})$ se pierden al considerar la representación paramétrica, se trata de obtener el subconjunto finito \mathcal{F} de $\mathbb{K}[x_1, \dots, x_n]$ tal que $V(\mathcal{F})$ es el conjunto algebraico más pequeño que contiene a S .

La explicación de la resolución del problema general comienza con el caso más sencillo en el que se asume que la parametrización considerada es polinomial. En esta situación, si la parametrización es

$$\{x_1 = p_1(y_1, \dots, y_m), \dots, x_n = p_n(y_1, \dots, y_m)\}$$

la representación implícita buscada puede determinarse mediante el cálculo de una Base de Gröbner. De forma más precisa, se considera el ideal \mathfrak{a} generado por

$$\{x_1 - p_1(y_1, \dots, y_m), \dots, x_n - p_n(y_1, \dots, y_m)\}$$

en $\mathbb{K}[x_1, \dots, x_n, y_1, \dots, y_m]$ y se calcula una Base de Gröbner \mathcal{G} de \mathfrak{a} respecto al orden lexicográfico, con $x_1 < \dots < x_n < y_1 < \dots < y_m$. Entonces se tiene que el conjunto \mathcal{F} de ecuaciones implícitas es $\mathcal{G} \cap \mathbb{K}[x_1, \dots, x_n]$.

A continuación se considera el caso más general que corresponde a las parametrizaciones racionales

$$\left\{ x_1 = \frac{p_1(y_1, \dots, y_m)}{q_1(y_1, \dots, y_m)}, \dots, x_n = \frac{p_n(y_1, \dots, y_m)}{q_n(y_1, \dots, y_m)} \right\}.$$

En esta situación el proceso anterior no es válido y requiere un estudio más detallado. Se considera la expresión libre de cuadrados q del polinomio $\prod_{i=1}^n q_i$ y se calcula una Base de Gröbner \mathcal{B} del ideal generado por

$$\{q(y_1, \dots, y_m)z - 1, q_1(y_1, \dots, y_m)x_1 - p_1(y_1, \dots, y_m), \dots, q_n(y_1, \dots, y_m)x_n - p_n(y_1, \dots, y_m)\}$$

en $\mathbb{K}[x_1, \dots, x_n, y_1, \dots, y_m, z]$, respecto al orden lexicográfico con

$$x_1 < \dots < x_n < y_1 < \dots < y_m < z.$$

Entonces se tiene que el conjunto \mathcal{F} de ecuaciones implícitas es $\mathcal{B} \cap \mathbb{K}[x_1, \dots, x_n]$.

Claramente, el proceso descrito anteriormente es válido para implicitar curvas y superficies paramétricas. De hecho, para curvas en el espacio se actúa de este modo. Sin embargo, para el caso de curvas planas y de superficies existen varias técnicas, basadas en distintas versiones de la clásica resultante, que evitan la utilización de Bases de Gröbner y, por tanto, conducen en algunos casos a algoritmos de implicitación más eficientes. En el caso de curvas planas, es decir, si la parametrización es

$$\mathcal{P}(t) = \left(\frac{p_1(t)}{q_1(t)}, \frac{p_2(t)}{q_2(t)} \right)$$

se demuestra que la ecuación implícita $f(x, y)$ de la curva se puede obtener mediante el cálculo de una resultante. Se prueba así que

$$\text{resultante}_t(q_1(t)x - p_1(t), q_2(t)y - p_2(t)) = (f(x, y))^{\text{grado}(\varphi_{\mathcal{P}})},$$

donde φ_p es la aplicación racional inducida por la parametrización. Es decir, la resultante es la ecuación implícita elevada al índice de trazado de la parametrización. Por tanto, teniendo en cuenta este resultado, para calcular la ecuación implícita $f(x, y)$ de la curva considerada se calcula la resultante antes indicada y se determina su factorización en factores libres de cuadrados. Así, como $f(x, y)$ es irreducible, la factorización genera un único factor que es la ecuación implícita de la curva.

Otra opción de cálculo de la ecuación implícita sería reparametrizar $\mathcal{P}(t)$ con el Algoritmo de Sederberg (ver [20]) para transformar la parametrización inicial en una parametrización propia y, por tanto, con índice de trazado 1. En esta situación, la resultante de Sylvester de la nueva parametrización es exactamente la ecuación implícita de la curva.

Para completar este estudio cabe también mencionar que existen otras técnicas más recientes como las basadas en el “*moving curves and surfaces*” (véase [22]) donde se determinan expresiones de la resultante mediante matrices de tamaño menor que el de la Matriz de Sylvester pero con coeficientes polinomiales algo más complicados. No está aún claro qué alternativa es la más eficiente desde el punto de vista computacional, por lo que en esta sección nos limitaremos a mencionar y justificar su existencia junto con un estudio experimental de éstas.

La implicitación de superficies se resuelve también mediante la utilización de resultantes multipolynomiales o multivariadas si la parametrización no tiene puntos base (véase [3]). De forma más precisa, dada una parametrización racional

$$\mathcal{P}(t, h) = \left(\frac{p_1(t, h)}{q(t, h)}, \frac{p_2(t, h)}{q(t, h)}, \frac{p_3(t, h)}{q(t, h)} \right),$$

con $\text{mcd}(p_1, p_2, p_3, q) = 1$, se definen los polinomios auxiliares

$$G_1(x, t, h) = q(t, h)x - p_1(t, h)$$

$$G_2(y, t, h) = q(t, h)y - p_2(t, h)$$

$$G_3(x, t, h) = q(t, h)z - p_3(t, h)$$

como polinomios en $\mathbb{K}(x, y, z)[t, h]$, y se calcula la resultante multivariada

$$R(x, y, z) = \text{resultante}(G_1, G_2, G_3).$$

A continuación se introduce la noción de puntos base de una parametrización $\mathcal{P}(t, h)$ como los puntos de intersección de las curvas definidas por los polinomios $p_i(t, h) = 0$ y $q(t, h) = 0$, y se indica que la existencia de puntos base imposibilita la aplicación directa de la resultante multivariable para resolver el problema de la implicitación. Para ello se demuestra que $R(x, y, z)$ es idénticamente nulo si y sólo si la parametrización $\mathcal{P}(t, h)$ tiene puntos base.

Seguidamente, mediante un razonamiento similar al realizado en el caso de las curvas planas, se deduce el siguiente resultado que se puede considerar como la generalización natural del resultado correspondiente para curvas planas: si la parametrización $\mathcal{P}(t, h)$ no tiene puntos base entonces

$$R(x, y, z) = (f(x, y, z))^\ell$$

donde $f(x, y, z)$ es la ecuación implícita de la superficie y $\ell \in \mathbb{N}$.

Por tanto, a partir de este resultado, si no hay puntos base, el cálculo de la resultante multivariable y de su factorización en factores libres de cuadrados proporciona la ecuación implícita deseada.

Asimismo, existen otras técnicas como las resultantes perturbadas (véase [3]) o las resultantes residuales (véase [2]) que usan presentaciones alternativas de la resultante multivariable y que evitan los problemas que proporcionan los puntos base pero originan expresiones algebraicas más complicadas de generar y manipular.

En la práctica la utilización de los algoritmos de implicitación basados en técnicas de eliminación causa ciertos problemas que se pueden poner de manifiesto con el siguiente ejemplo. Se considera la superficie bicúbica \mathcal{S} (véase la Figura 1) definida por las siguientes ecuaciones paramétricas:

$$x(u, v) = 3v(v - 1)^2 + (u - 1)^3 + 3u$$

$$y(u, v) = 3u(u - 1)^2 + v^3 + 3v$$

$$z(u, v) = 3u(5u - 5 - u^2)v^3 - 3(u^3 + 6u^2 - 9u + 1)v^2 + v(6u^3 + 9u^2 - 18u + 3) - 3u(u - 1)$$

donde $u, v \in [0, 1]$.

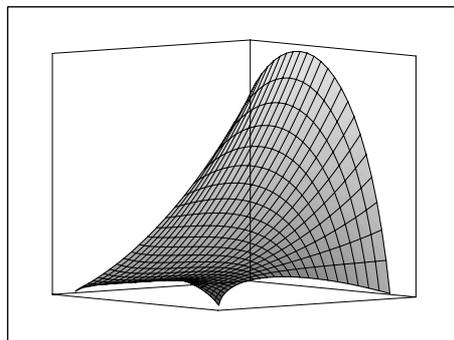


Figura 1. Superficie bicúbica.

Uno de los primeros problemas que se pueden plantear en esta situación es determinar (si existen) los puntos de intersección de la superficie \mathcal{S} con la recta $x = u, y = u, z = u$. Esto implica resolver el siguiente sistema de ecuaciones:

$$\begin{aligned} u &= 3t(t - 1)^2 + (s - 1)^3 + 3s \\ u &= 3s(s - 1)^2 + t^3 + 3t \\ u &= -3s(s^2 - 5s + 5)t^3 - 3(s^3 + 6s^2 - 9s + 1)t^2 + t(6s^3 + 9s^2 - 18s + 3) - 3s(s - 1) \end{aligned}$$

En este caso particular se obtiene sólomente un punto de intersección, con coordenadas

$$(0.5561, 0.5561, 0.5561)$$

que corresponde a $s = 0.2748$ y $t = 0.0408$.

Otra manera de resolver este problema consiste en determinar primero la ecuación implícita $H(x, y, z) = 0$ de \mathcal{S} . Una vez calculada ésta el problema anterior se reduce a resolver la ecuación (en una variable) $H(u, u, u) = 0$. Análogamente, cualquier intersección de \mathcal{S} con una curva se puede abordar resolviendo una ecuación en una variable. En este caso particular, la ecuación implícita de \mathcal{S} es:

$$H(x, y, z) = z^9 + \sum_{i=1}^9 \mathbf{r}_i(x, y)z^{9-i},$$

donde

$$\begin{aligned} \mathbf{r}_1(x, y) &= -\frac{233469x}{2048} + \frac{188595y}{2048} - \frac{112832595}{262144} - \frac{81x^2}{64} + \frac{135xy}{32} - \frac{81y^2}{64} \\ \mathbf{r}_2(x, y) &= -\frac{20972672709381x}{536870912} + \frac{17975329363179y}{536870912} - \frac{729y^4}{8192} - \frac{729x^4}{8192} + \frac{1215x^3y}{2048} - \frac{4779x^2y^2}{4096} + \frac{1215xy^3}{2048} - \\ &\quad \frac{4105971x^3}{65536} + \frac{3129597y^3}{65536} + \frac{14456151x^2y}{65536} - \frac{13181049xy^2}{65536} - \frac{54187594407x^2}{16777216} + \frac{48101467761xy}{8388608} - \\ &\quad \frac{38812918311y^2}{16777216} - \frac{22656991982391171}{137438953472} - \frac{1}{2} \left(\frac{233469x}{2048} - \frac{188595y}{2048} + \frac{112832595}{262144} + \frac{81x^2}{64} - \right. \\ &\quad \left. \frac{135xy}{32} + \frac{81y^2}{64} \right) \left(-\frac{233469x}{2048} + \frac{188595y}{2048} - \frac{112832595}{262144} - \frac{81x^2}{64} + \frac{135xy}{32} - \frac{81y^2}{64} \right) \\ \mathbf{r}_3(x, y) &= \dots\dots\dots \end{aligned}$$

Substituyendo $x = u, y = u, z = u$ se obtiene una ecuación de grado 18 en u , muy fácil de resolver:

$$5159780352u^{18} - 609499054080u^{17} + \dots + 3707912273492242256259566313 = 0.$$

El problema que surge con esta filosofía de trabajo consiste en que, aunque hay métodos para calcular la ecuación implícita $H(x, y, z) = 0$, estos métodos son muy ineficientes y difícilmente se pueden integrar en los softwares CAD/CAM, donde los usuarios deben recibir las respuestas a sus demandas computacionales en tiempo real. Se finaliza la sección mostrando algunas alternativas que permiten la utilización de ecuaciones implícitas como puede ser la implicitación genérica (ver [17]) o la implicitación aproximada (ver [7] y [8]).

4. Manipulación de curvas planas definidas implícitamente

Muchos problemas en Diseño Geométrico Asistido por Ordenador se reducen al cálculo del grafo de una curva plana algebraica definida implícitamente. Por ejemplo, si se quiere seccionar la superficie

$$x = \frac{X(s, t)}{W(s, t)}, \quad y = \frac{Y(s, t)}{W(s, t)}, \quad z = \frac{Z(s, t)}{W(s, t)}; \quad s, t \in [0, 1]$$

(definida por los polinomios reales $X(s, t)$, $Y(s, t)$, $Z(s, t)$ y $W(s, t)$) con respecto al plano $x = x_0$, hay dos posibilidades: o se dibuja en el cuadrado unidad $[0, 1] \times [0, 1]$ la curva algebraica plana definida por

$$X(s, t) - x_0 W(s, t) = 0$$

y luego se eleva este dibujo a la superficie considerada, o, si la ecuación implícita $H(x, y, z)$ de la superficie considerada ha sido calculada, el proceso de “elevar el dibujo” se puede evitar calculando el grafo de la curva algebraica plana definida por $H(x_0, y, z) = 0$. De todas formas, las dos posibilidades requieren dibujar una curva algebraica plana definida implícitamente.

En esta sección se presentan varios algoritmos que, dado un polinomio $f(x, y)$ en $\mathbb{R}[x, y]$, calcula un grafo lineal que representa la estructura topológica de la curva definida por el polinomio considerado (incluso en el caso que la curva considerada tenga singularidades “complicadas”). El principal objetivo de este algoritmo es generar (de la forma más rápida posible) información correcta sobre el conjunto $f(x, y) = 0$ que puede ser utilizada posteriormente para abordar el trazado de la curva considerada. Por ejemplo, si se considera el polinomio

$$\begin{aligned} g(x, y) = & 279756x + 279936xy^4 - 559692y^2x - 15583x^3 + 217x^5 + \frac{130218}{5}x^2 - \frac{23039}{10}x^4 + \\ & + \frac{359}{10}x^6 + 370656y^4 - \frac{3726432}{5}y^2 - 72774y^2x^2 + \frac{12947}{5}y^2x^4 + 1296y^6 + \\ & + 46728y^4x^2 + 15588y^2x^3 + \frac{37333439}{100} \end{aligned}$$

se obtiene un caso de decisión topológica complicado: de su dibujo (Figura 2, izquierda), difícilmente se puede deducir cuál es la estructura topológica de la curva (véase la Figura 2, derecha).



Figura 2. $g(x, y) = 0$.

Determinar (incluso topológicamente) el grafo correspondiente a una curva algebraica plana definida implícitamente representa un problema de especial interés, en primer lugar, en Diseño Geométrico Asistido por Ordenador por ser un subproblema básico que aparece con frecuencia en la práctica y, en segundo lugar, en Álgebra Computacional, siendo este problema el “responsable” de muchos de los avances sobre subresultantes, cálculo simbólico del número de raíces reales, cálculo con infinitésimos, etc. No es tampoco desdeñable su conexión, como herramienta computacional, con la primera parte del Problema XVI de Hilbert sobre las distribuciones posibles de los óvalos de una curva algebraica plana (sin puntos singulares) definida por un polinomio de grado n .

La estrategia usual para determinar (desde un punto de vista topológico) el grafo de una curva algebraica plana definida implícitamente por un polinomio $f(x, y) \in [x, y]$ es la siguiente:

- ₁ Calcular el discriminante de f con respecto a y , $R(x)$, y determinar sus raíces reales, $\alpha_1 < \dots < \alpha_r$.
- ₂ Para cada α_i , determinar las raíces reales de $f(\alpha_i, y)$, $\beta_{i,1} < \dots < \beta_{i,s_i}$.
- ₃ Para cada α_i y $\beta_{i,j}$, calcular el número de ramas a la derecha y a la izquierda del punto $(\alpha_i, \beta_{i,j})$.

Toda esta información es suficiente para determinar la estructura topológica de la curva definida por $f(x, y)$.

Para evitar los posibles problemas numéricos en el cálculo de las raíces de $R(x)$ y sobre todo de cada $f(\alpha_i, y)$ que tiene siempre raíces múltiples, antes de empezar los cálculos se realiza un cambio lineal y genérico de coordenadas con el fin de que se cumpla la siguiente condición para cada $\alpha \in \mathbb{R}$:

$$\#\{\beta \in \mathbb{C} : f(\alpha, \beta) = 0, \frac{\partial f}{\partial y}(\alpha, \beta) = 0\} \leq 1.$$

Esta condición nos asegura que para cada α_i raíz real de $R(x)$, hay, a lo sumo, un punto crítico (un punto singular ó un punto de tangente vertical) de la curva en la línea vertical $x = \alpha_i$ y que la coordenada en y de tal punto se puede describir de forma racional en términos de α_i mediante la utilización de subresultantes o de técnicas matriciales que involucran la determinación de los valores propios generalizados de una haz de matrices que se construye a partir de la matriz de Sylvester. Además, esta condición permite construir de forma simbólica, para cada $f(\alpha_i, y)$, un polinomio libre de cuadrados $g(\alpha_i, y)$ cuyas raíces reales se calculan para finalizar el paso \bullet_2 . De esta forma, el paso \bullet_3 se reduce al cálculo del número de raíces reales de los polinomios libres de cuadrados $f(\gamma_i, y)$ ($i \in \{0, 1, \dots, r\}$) con $\gamma_0 = -\infty$, $\gamma_r = \infty$ y γ_i cualquier número real en el intervalo abierto (α_i, α_{i+1}) . Se obtiene así un grafo de la curva considerada, que será muy útil a la hora de trazar de forma numérica dicha curva, ya que se sabe exactamente cómo proceder al acercarse a un punto “complicado”.

4. Offsetting

Esta sección se dedica al desarrollo de algoritmos para la construcción y manipulación de offsets. En [14], C. Hoffmann inició una línea de trabajo que trata de reducir problemas de cálculo de esta construcción geométrica a problemas algebraicos de eliminación y que constituye nuestro punto de partida.

Las offsets ocupan un lugar destacado en el diseño geométrico asistido por ordenador, principalmente por sus múltiples aplicaciones, entre las que destacan la planificación de trayectorias de robots, el análisis de tolerancia, el modelado geométrico, o la generalización de diagramas de Voronoi.

La sección comienza con la definición de offset. Para ello, se distinguen dos niveles. En el primero se da una idea intuitiva de la noción geométrica de offset, y en el segundo se muestra cómo el resultado de esta construcción provoca un aumento considerable en la complejidad del objeto resultante.

De forma intuitiva la noción de offset de una curva o una superficie se introduce como el lugar geométrico de los puntos a distancia d de la curva o superficie considerada. Es decir, la offset de una curva (resp. una superficie) se puede considerar como la envolvente del sistema de circunferencias (resp. esferas) de radio d centradas en los puntos de la curva (resp. superficie) original. Por tanto, la offset está formada por los puntos intersección de las circunferencias (resp. esferas) y la recta normal a la conjunto inicial en el centro. Así, si $\mathcal{P}(\bar{t})$ es una parametrización de la curva o superficie inicial, los puntos en la offset vienen caracterizados por la fórmula

$$\mathcal{P}(\bar{t}) \pm \frac{d}{\|\mathcal{N}(\bar{t})\|} \mathcal{N}(\bar{t})$$

donde $\mathcal{N}(\bar{t})$ es el vector normal a la curva o superficie inicial en el punto $\mathcal{P}(\bar{t})$. Se muestra entonces el ejemplo clásico de la offset de una parábola a distancia 1 (véase Figura 3). En el ejemplo se observa que, aunque se ha partido de una parábola, la curva resultante es bastante más complicada en grado (se pasa de grado total 2 a grado total 6) y en singularidades (se pasa de una cónica irreducible y por tanto sin singularidades a una curva con varias singularidades). Se indica que esta situación no es casual y que, en general, el resultado de la construcción de offsets origina este tipo de problemas.

Asimismo, este ejemplo permite mostrar cómo las técnicas de implicitación desarrolladas en la sección 2 proporcionan la ecuación implícita de la offset cuando la curva o superficie original viene dada de forma implícita ($y = x^2$) o en forma paramétrica ((t, t^2)).

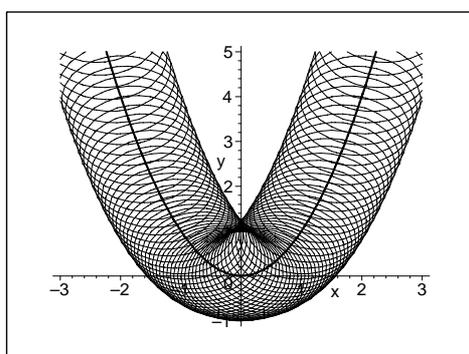


Figura 3. Offset de la parábola $y = x^2$ a distancia 1.

Como se muestra en la Figura 4 el análisis de la variación de las características topológicas de la offset es un problema que está recientemente tomando un creciente interés. La aplicación de las técnicas mostradas en la sección 3 permite, por ejemplo, resolver de forma completamente satisfactoria este problema para el caso de las cónicas y las cúbicas en el plano (ver [12]).

Se termina esta introducción indicando que, en la actualidad, la mayoría de las aplicaciones requiere la construcción de offsets racionales y del cálculo de sus parametrizaciones y se indica que, en general, la racionalidad no se hereda mediante el cálculo de offsets.

Seguidamente, se procede a la definición rigurosa de curva o superficie offset. Para ello se observa que la construcción intuitiva anterior, aunque válida en la mayoría de los casos, conduce a situaciones no esperadas. Así, por ejemplo, si se toma como centro de la circunferencia o esfera una singularidad de la curva o superficie inicial aparecen dificultades al considerar el correspondiente vector normal (el ejemplo de la superficie $x^2 + y^2 + z^2 = 0$ es especialmente significativo en esta situación).

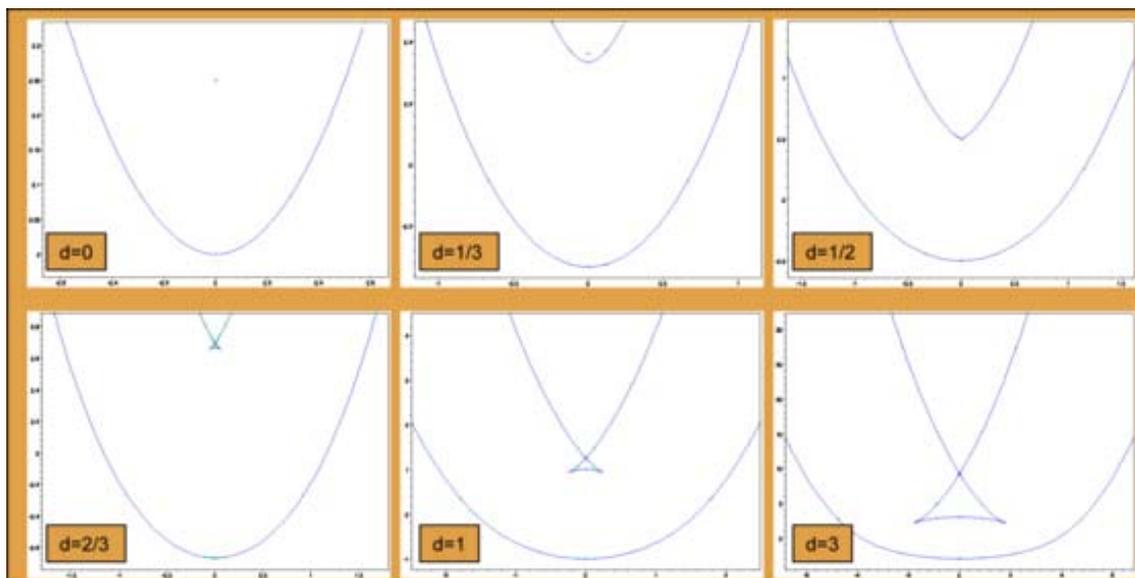


Figura 4. Variación de la topología de la offset de la parábola.

Para resolver estas dificultades se consideran, en primer lugar, sólo curvas o superficies reales para las cuales existe un abierto de puntos regulares. El problema de las singularidades se resuelve tomando el sistema de circunferencias o esferas con centros dentro del abierto de puntos regulares y considerando la clausura de Zariski de dicho conjunto. Asimismo, se demuestra que la offset de una curva o de una superficie es irreducible o tiene dos componentes que se corresponden con la idea intuitiva de offset interior y exterior.

5. Ejemplos

Se muestra en esta sección un conjunto de programas (junto con su aplicación a la resolución de problemas de manipulación de curvas y superficies) que implementan algunos de los algoritmos que se han desarrollado a lo largo de este minicurso.

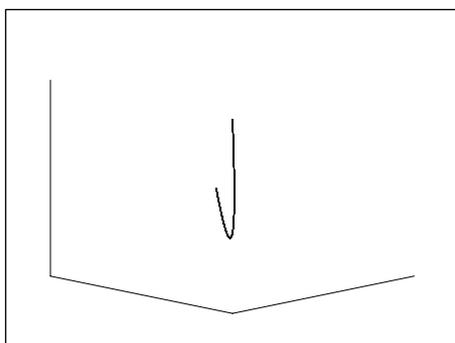
Implicitación genérica de superficies de revolución

Se determina la ecuación implícita genérica de la superficie de revolución que se obtiene al hacer girar la curva que se introduce a continuación alrededor del eje X.

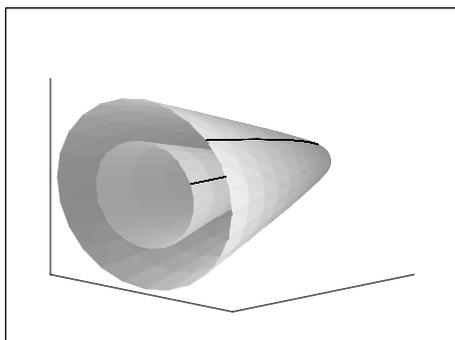
```
> c[1] := (c[1,2]*t**2+c[1,0]);
> c[2] := (c[2,2]*t**2+c[2,1]*t);
> c[3] := (c[3,2]*t**2+c[3,1]*t+c[3,0]);
> Eq := sort(collect(ImplicitSuperfRevolX(c[1],c[2],c[3],x,y,z,t),
> [x,y,z],distributed), [x,y,z]);
```

Se especializan los parámetros y se estudia cómo es la superficie de revolución que se ha obtenido.

```
> c[1,2]:=rand(1..50)/50:c[1,0]:=rand(1..50)/50:
> c[2,2]:=rand(1..50)/50:c[2,1]:=rand(1..50)/50:
> c[3,2]:=rand(1..50)/50:c[3,1]:=rand(1..50)/50:
> c[3,0]:=rand(1..50)/50:
> f[1]:=t -> c[1]:f[2]:=t -> c[2]:f[3]:=t -> c[3]:
> plot3d([f[1](t),f[2](t),f[3](t)],t=-6..6,u=0..1,axes=FRAME,
> orientation=[45,75],tickmarks=[0,0,0]);
```



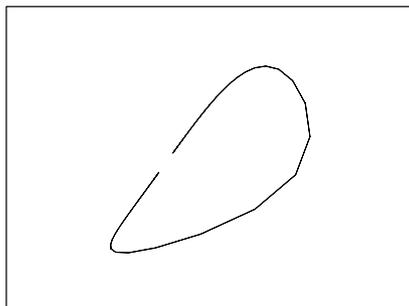
```
> plots[display]([
> plot3d([[f[1](t),[f[2](t),f3(t)],t=-2..2, u=0..1, axes=FRAME,
> orientation=[45,75]),
> plot3d([[f[1](t),[f[2](t)*cos(u)+[f[3](t)*sin(u),[f[3](t)*cos(u)-
> [f[2](t)*sin(u)],t=-2..2,u=0..2*Pi,axes=FRAME,orientation=[45,75],
> style=PATCHNOGRID,tickmarks=[0,0,0]])]);
```



Seccionado de superficies de revolución

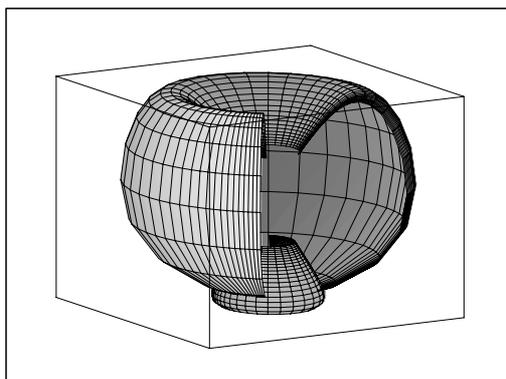
La curva C en $Y=0$ definida por $x=C_1(t)$, $z=C_2(t)$ se va a rotar alrededor del eje OZ .

```
> C[1] := (2*t-1) / (1+t**2) ; C[2] := 0 ; C[3] := (1-t+t**2) / (2+t+t**2) ;
> plot3d([C[1],C[2],C[3]],t=-25..25,s=-1..1,
> orientation=[60,75],grid=[175,2],tickmarks=[0,0,0]);
```



Se muestra a continuación la superficie de revolución que genera la curva C .

```
> T[1] := C[1] * (2*s) / (1+s**2) - C[2] * (1-s**2) / (1+s**2) :
> T[2] := C[2] * (2*s) / (1+s**2) + C[1] * (1-s**2) / (1+s**2) : T[3] := C[3] :
> P1 := (a,b) -> plot3d([T[1],T[2],T[3]],s=a..b,t=-6..6,grid=[50,50],
> axes=BOXED,scaling=UNCONSTRAINED,projection=1,tickmarks=[0,0,0],
> orientation=[59,75]) : P1(-3,3);
```



La ecuación implícita de esta superficie de revolución se calcula de forma genérica.

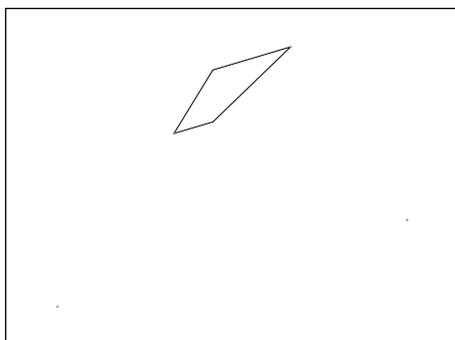
```
> toro := ImplicitSuperfRevolZ(C[1],C[2],C[3],x,y,z,t);
```

Se determina la que se obtiene al intersecar esta superficie de revolución con el plano $y=1$.

```
> Seccion[1] := subs(y=1, toro);
```

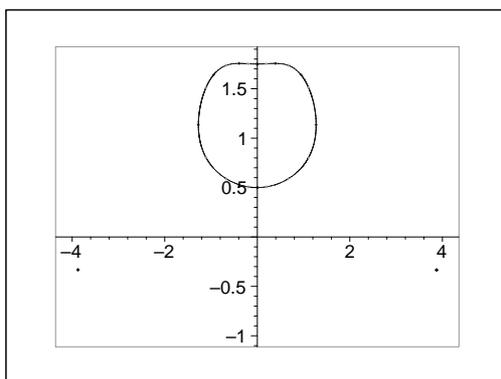
Se calcula la topología de esta sección que muestra la presencia de dos puntos aislados.

```
> principal(Seccion1,x,z,10,'black');
```

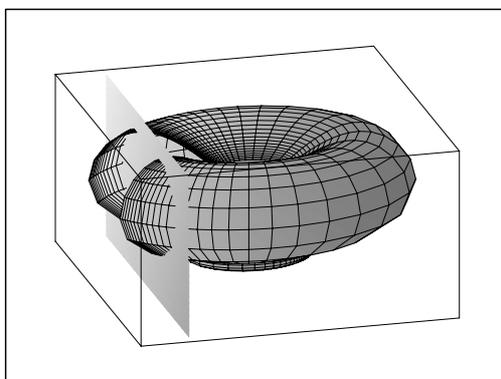


Se muestra finalmente la sección plana junto con su ubicación en la superficie de revolución inicial.

```
> plot_real_curve(Seccion[1],x,z,scaling=UNCONSTRAINED);
```



```
> P[2]:=y->plot3d([s,y,t],s=-2..2,t=0..2,style=PATCHNOGRID);
> plots[display](P[1](-4,4),P[2](1),grid=[50,50],axes=BOXED,
> scaling=UNCONSTRAINED,projection=1,tickmarks=[0,0,0],
> orientation=[165,57]);
```

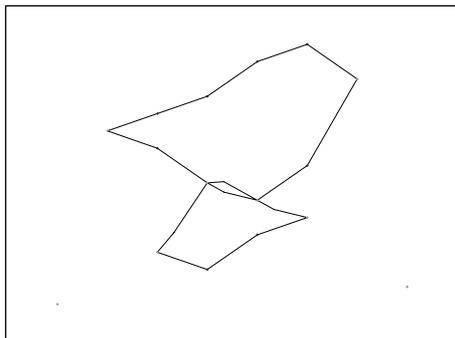


Se calcula a continuación la sección por el plano $y=1/2$:

```
> Seccion[2] := subs (y=1/2, toro) ;
```

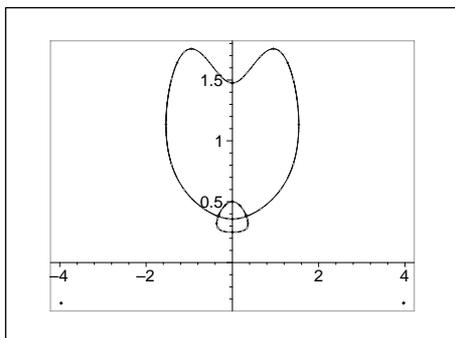
Topología de la sección:

```
> principal (Seccion[2], x, z, 20, 'black') ;
```

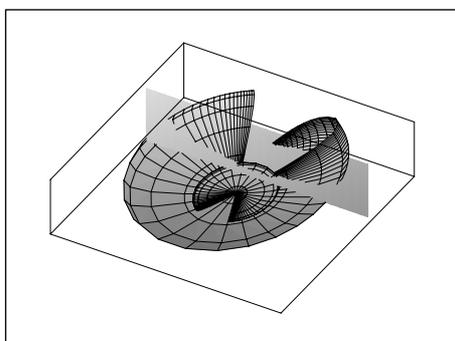


Grafo de la sección:

```
> plot_real_curve (Seccion[2], x, z, scaling=UNCONSTRAINED) ;
```



```
> plots[display] (P[1] (-4,4), P[2] (1/2), grid=[50,50], axes=BOXED,
> scaling=UNCONSTRAINED, projection=1, tickmarks=[0,0,0],
> orientation=[60,161]) ;
```



6. Reconocimientos

Parcialmente subvencionado por DGESIC PB 98-0713-C02-02 (Ministerio de Educación y Cultura) y Proyecto GAIA II (Unión Europea).

7. Bibliografía

Aparte de las citas explícitas que se han hecho en algunas de las secciones de este minicurso, se debe indicar que recientemente han aparecido una serie de manuales que pueden ser de gran utilidad a la hora de elaborar un curso de estas características. Para la parte más ligada al Diseño Geométrico Asistido por Ordenador los manuales [19], [9] y [15] son especialmente recomendables. La exposición de C. Hoffmann en el manual [13] es la primera en la que aparece la utilización de técnicas algebraicas para la resolución de problemas en Diseño Geométrico Asistido por Ordenador.

Hay que destacar que la línea argumental de este minicurso está claramente inspirada por el excelente artículo de T. Sederberg (ver [21]) y por el “survey” que aparece en [10].

Referencias

- [1] C. Bajaj, G. Xu, R. Holt, A. Netravali: NURBS approximation of A-splines and A-patches. *International Journal of Computational Geometry and Applications*, 2003.
- [2] L. Busé: *Residual resultant over the projective plane and the implicitization problem*. Proceedings of ISSAC 2001, pp. 48-55, 2001.
- [3] J. Canny, I. Emiris: Effective Incremental Algorithms for the Sparse Resultant and the Mixed Volume. *Journal of Symbolic Computation* **20** (1995), 117-149.
- [4] E.W. Chionh, R.N. Goldman: Using Multivariate Resultants to Find the Implicit Equation of a Rational Surface. *The Visual Computer* **8** (1992), 171-180.
- [5] Cohen, H. Cuyper, H. Sterk: *Some tapas of computer algebra*. Algorithms and Computations in Mathematics **4**. Springer-Verlag, 1999.
- [6] D. Cox, J. Little, D. O’Shea: *Using Algebraic Geometry*. Graduate Texts in Mathematics **185**. Springer-Verlag, 1998.
- [7] T. Dokken: *Aspects of intersection algorithms and approximation*. Tesis Doctoral, University of Oslo (Noruega), 1997.
- [8] T. Dokken: *Approximate Implicitization*. *Mathematical Methods in CAGD*. Vanderbilt University Press, pp. 81-102, 2001.
- [9] G. Farin, J. Hoschek, M.-S. Koo: *Handbook of Computer Aided Geometric Design*. Elsevier Science, 2002. [Disponible en <http://cagd.snu.ac.kr>].
- [10] L. González-Vega: Computer Aided Design and Modelling. En J. Grabmeier, E. Kaltofen, V. Weispfenning (editores): *Computer Algebra Handbook: Foundations, Applications, Systems*. Springer-Verlag, 2003.
- [11] L. González-Vega, I. Necula: Efficient topology determination of implicitly defined algebraic plane curves. *Computer Aided Geometric Design* **19** (2003), 719-743.
- [12] L. González-Vega, R. Sendra: Topological characterization of offset degenerations for conics and cubic plane curves. *Preprint*, 2002.
- [13] C.M. Hoffmann: *Geometric and Solid Modelling: An Introduction*. Morgan Kaufmann Publishers, 1989.

-
- [14] C.M. Hoffmann: Algebraic and numerical techniques for offsets and blends. En *Computation of Curves and Surfaces*, NATO ASI Series **37**, pp. 499-528. Kluwer Academic Publishers, 1989.
- [15] T. Maekawa, N. Patrikalakis: Shape Interrogation for Computer Aided Design and Manufacturing. Springer-Verlag, 2002.
- [16] R. McLeod, M.L. Baart: *Geometry and Interpolation of Curves and Surfaces*. Cambridge University Press, 1998.
- [17] I. Necula: *Manipulación de curvas y superficies en CAGD mediante técnicas simbólico-numéricas*. Tesis Doctoral, Universidad de Cantabria, 2003.
- [18] V. Rovenski: *Geometry of Curves and Surfaces with Maple*. Birkhäuser, 2000.
- [19] D. Salomon: *Computer Graphics & Geometric Modeling*. Springer-Verlag, 1999.
- [20] T.W. Sederberg: Improperly Parametrized Rational Curves. *Computer Aided Geometric Design* **3** (1986), 67-75.
- [21] T.W. Sederberg: Applications to computer aided geometric design. En D.A. Cox, B. Sturmfels (editores): *Applications of Computational Algebraic Geometry*. Proceedings of Symposia in Applied Mathematics **53**, pp. 67-89. AMS, 1998.
- [22] T.W. Sederberg, R.N. Goldman, H. Du: Implicitizing Rational Curves by the Method of Moving Algebraic Curves. *Journal of Symbolic Computation* **23** (1997), 153-175.

En Internet

<http://www.can.nl>

CAIN Europe

Computer Algebra Information Network.

<http://www.math.sintef.no/gaiatwo>

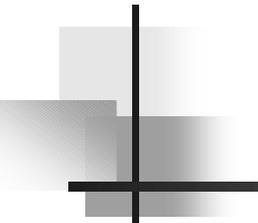
Intersection algorithms for geometry based IT-applications using approximate algebraic methods

Página web del proyecto GAIA II.

<http://www.spatial.com>

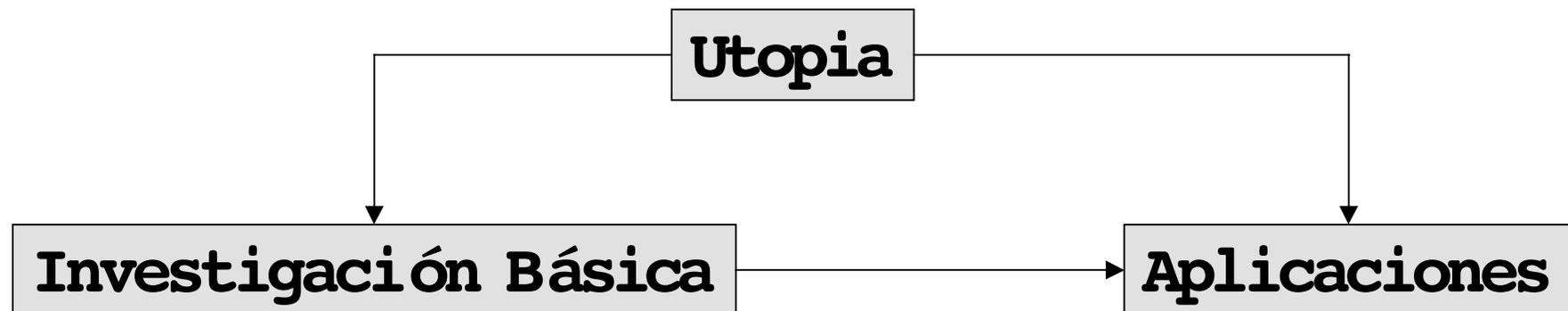
3D Software Development Technologies

Tecnologías de modelado en 3D.



Las Matemáticas del Diseño Geométrico Asistido por Ordenador y sus Aplicaciones en la Industria

Matemáticas e Industria



Laureano González Vega
Departamento de Matemáticas, Estadística y Computación

FEDER Project (1FD97-0409), 1999-2001

Integrating new algebraic-numerical techniques in Computer Aided Geometric Design (CAGD): Developing Problem Solving Environments and Implementation into an industrial CAD/CAM framework



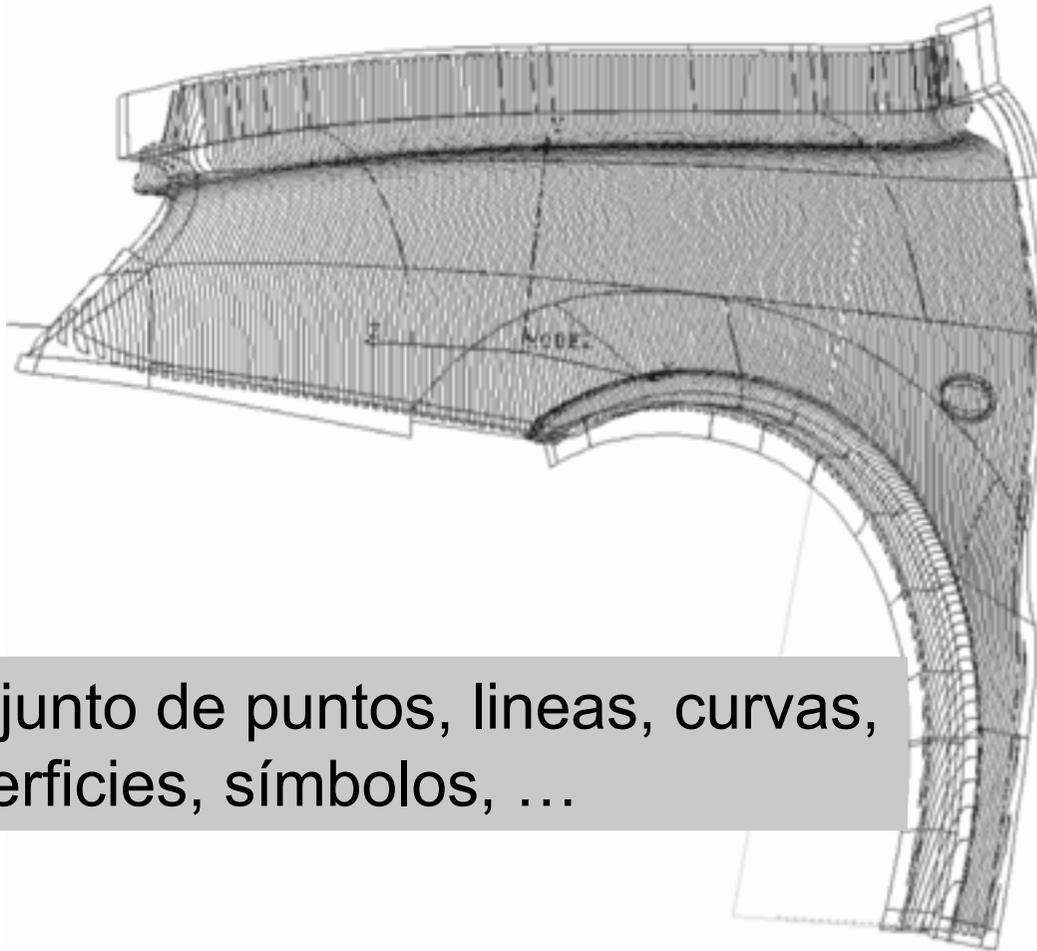
Esquema

- **Motivación**
- **Implicitación en la práctica:**
 - ❖ **Implicitación genérica**
 - ❖ **Optimización de su evaluación**
- **Offsets**
- **Cálculo de topología de curvas definidas de forma implícita**
- **Aplicaciones:**
 - ❖ **Control del error de forma**
 - ❖ **Iluminación**
- **Un entorno de trabajo numérico/simbólico para CAGD**





Definición matemática del troquel

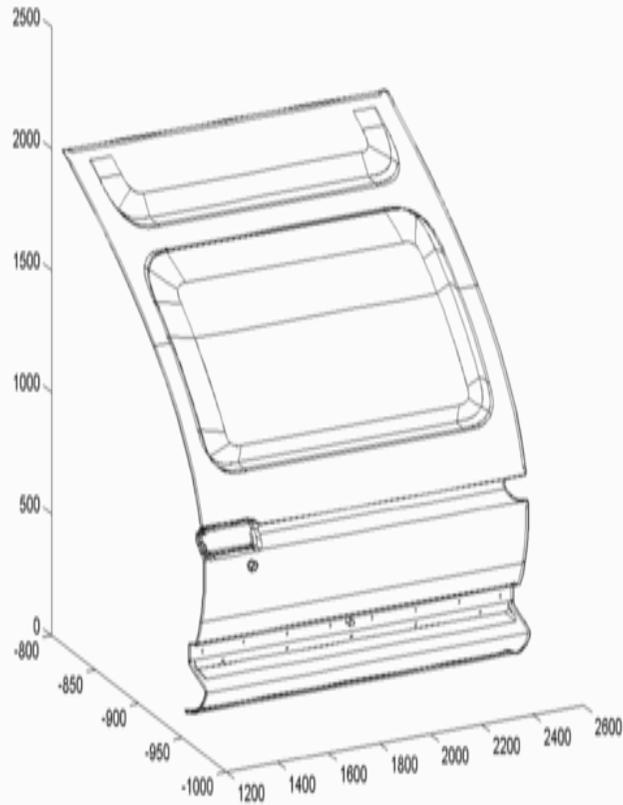


Conjunto de puntos, líneas, curvas, superficies, símbolos, ...

Ficheros ASCII

CAD formatos:

- IGES
- VDA
- STEP
-



This line is the first line of the start section. You can replace these lines and add as many as you want.

```

, ,10Hpasso.felix,38H/users/cadds5/parts/iges/_bcd/top1.igs,44HCOMPUTERVISG
ION CADD5 REV 5.0 GRAPHIC SYSTEM,39HCV Translator Products, Revision 1.G
2.1 ,32,38,6,308,15,10Hpasso.felix,1.0D0,2,2HMM,32767,32.767D0,13H970226.G
163500, .00254D0,0.0D0,6HAuthor,12HOrganization,10,0,13H700101.010000;
314 1 0 0 0 0 0 0 200D 1
314 0 7 1 0 0 0 0 OD 2
314 2 0 0 0 0 0 0 200D 3
314 0 8 1 0 0 0 0 OD 4
314 3 0 0 0 0 0 0 200D 5
314 0 0 1 0 0 0 0 OD 6

```

```

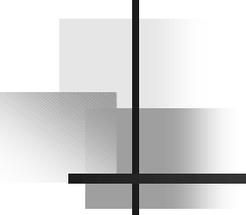
-39.8353900668919E0,47.5E0,-339.792802669565E0, 39P 51
-40.5903342877884E0,47.5E0,-339.904784107237E0, 39P 52
-40.9945510170104E0,47.5E0,-339.985850987103E0, 39P 53
-41.6323696207208E0,47.5E0,-339.994611986241E0, 39P 54
-41.8603348249209E0,47.5E0,-339.985787668859E0, 39P 55
-42.0878723928957E0,47.5000000000001E0,12.4973267384271E0, 39P 56
19.2446820045842E0,0.0E0,0.0E0,1.0E0; 39P 57
126,13,3,1,0,1,0,12.4973267384271E0,12.4973267384271E0, 41P 58
12.4973267384271E0,12.4973267384271E0,13.6287982402182E0, 41P 59
13.6287982402182E0,14.8518171484082E0,14.8518171484082E0, 41P 60
16.0816380356578E0,16.0816380356578E0,17.3158326864021E0, 41P 61
17.3158326864021E0,18.5608627450732E0,18.5608627450732E0, 41P 62

```

```

47.4999990995075E0,-327.225327496092E0,-52.587704489423E0, 523P 4982
47.4999990995075E0,-1.11551352287571E-2,9.30515325886385E0, 523P 4983
-17.9176207830942E0,9.05452635922904E-7; 523P 4984
S 2G 4D 524P 4984 T 1

```



Realización del troquel

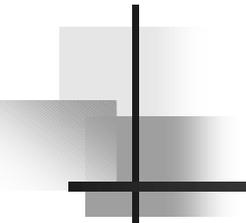
Un primer modelo real es construido usando su definición matemática.

Este modelo es mejorado usando operaciones tales como:

- seccionado,
- offsetting,
- blending,

Estas operaciones se definen inicialmente, en forma matemática, sobre el modelo teórico.

El control de calidad se realiza para verificar que el troquel construido se ajusta a las especificaciones iniciales de éste.



Implicitación y sus Aplicaciones

El problema algebraico

Cálculo de la ecuación implícita de una curva racional en \mathbb{R}^2 ,
de una superficie racional en \mathbb{R}^3 ,

EJEMPLO

Let \mathcal{S} be an elementary bicubic B -spline surface defined by:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = (n_0(u) \ n_1(u) \ n_2(u) \ n_3(u)) \cdot \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{pmatrix} \cdot \begin{pmatrix} n_0(v) \\ n_1(v) \\ n_2(v) \\ n_3(v) \end{pmatrix}$$

with $0 \leq u, v \leq 1$, $P_{ij} \in \mathbb{R}^3$, and

$$\begin{aligned} n_0(s) &= \frac{(1-s)^3}{6} & n_1(s) &= \frac{3s^3-6s^2+4}{6} \\ n_2(s) &= \frac{-3s^3+3s^2+3s+1}{6} & n_3(s) &= \frac{s^3}{6} \end{aligned}$$

Determinar un polinomio $H(x,y,z)$ con el menor grado posible y tal que \mathcal{S} está contenido en:

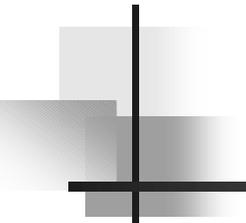
$$\{(\alpha, \beta, \gamma) \in \mathbb{R}^3 : H(\alpha, \beta, \gamma) = 0\}$$

La Implicitación es un problema clásico en Geometría Algebraica (Teoría de la Eliminación):

- mediante resultantes multivariadas
Manocha&Canny, Sederberg, Chionh, Bajaj,...
- mediante Bases de Gröbner
Buchberger, Cox et al, Hoffmann, Kalkbrenner,
Gao&Chou, Fing et al,...
- mediante “moving curves and surfaces”
Cox, Sederberg, Goldman, Zhang, Chen,...
- mediante deformaciones y técnicas “ad-hoc”
Canny, Gonzalez-Vega, ...

Cuando se dispone de ella, la ecuación implícita es muy útil para:

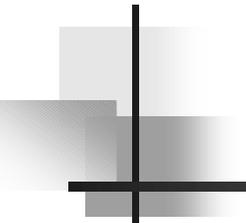
- Posicionamiento Punto-Superficie
- Intersecciones Superficie-Superficie (incluyendo seccionado)
- Intersecciones Superficie-Curva
- Manipulación de Superficies “Trimmed” “Sculptured Solids”
-



La Implicitación no se utiliza o no aparece en los sistemas actuales CAD/CAM debido a dos razones:

Es una operación algebraica muy costosa.

Los coeficientes de los polinomios que definen la parametrización son números reales en coma flotante.



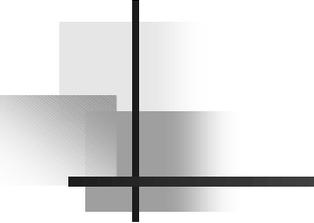
Nuevas Aproximaciones Numéricas:

- mediante Calculo de Variaciones y como un problema de valores propios:
 - ❖ Corless, Giesbrecht, Kotsireas & Watt (2000)

- mediante un problema SVD sobre la matriz de composición y monoides:
 - ❖ Dokken (1997)
 - ❖ Sederberg, Zheng, Klimaszewski & Dokken (1999)

- mediante el método de Newton:
 - ❖ Hartmann (1998)

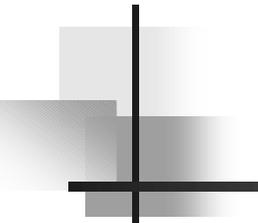
Como evitar estas dificultades (I):



Números reales en coma flotante
como coeficientes

Un objeto concreto a ser modelizado (para luego ser construido) está formado por varios cientos (o miles) de pequeños parches, **todos ellos compartiendo la misma estructura algebraica.**

Para cada objeto, se construye una base de datos que contiene la ecuación implícita de cada tipo de parche que aparece en su definición.



La Base de Datos contiene las formulas de inversión:

$$u=U(x,y,z) \quad v=V(x,y,z)$$

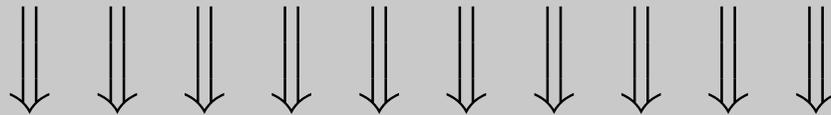
La Base de Datos se filtra “mediante evaluación” para evitar problemas de especialización.

El proceso de implicitación se realiza mediante el uso de Resultantes de Sylvester, cálculo de Bases de Gröbner y técnicas ad-hoc para casos específicos.

$$x = x_{00} \frac{t_2 - t}{t_2 - t_1} + x_{01} \frac{t - t_1}{t_2 - t_1}$$

$$y = y_{00} \frac{s_2 - s}{s_2 - s_1} + y_{11} \frac{s - s_1}{s_2 - s_1}$$

$$z = \left[z_{00} \frac{s_2 - s}{s_2 - s_1} + z_{10} \frac{s - s_1}{s_2 - s_1} \right] \frac{t_2 - t}{t_2 - t_1} + \left[z_{10} \frac{s_2 - s}{s_2 - s_1} + z_{11} \frac{s - s_1}{s_2 - s_1} \right] \frac{t - t_1}{t_2 - t_1}$$



$$\begin{aligned}
 & (z_{00} - 2z_{10} + z_{11})xy + (y_{00}z_{10} + y_{11}z_{10} - y_{00}z_{11} - y_{11}z_{00})x + \\
 & + (x_{00}z_{10} + x_{01}z_{10} - x_{00}z_{11} - x_{01}z_{00})y + \\
 & + (x_{00}y_{11} + x_{01}y_{00} - x_{00}y_{00} - x_{01}y_{11})z + \\
 & + x_{01}y_{11}z_{00} + x_{00}y_{00}z_{11} - x_{00}y_{11}z_{10} - x_{01}y_{00}z_{10}
 \end{aligned}$$

Como evitar estas dificultades (II):

Tiempo de cálculo

Como la implicitación "genérica" es un proceso que se realiza a priori, en la práctica, el tiempo de cálculo se dedica en exclusiva:

- ❖ a la búsqueda en la base de datos,
 - ❖ a la evaluación de la ecuación implícita, y
 - ❖ a la verificación de esta
- mientras el usuario CAD/CAM está trabajando.

Un ejemplo real en CANDEMAT

La superficie considerada viene definida por un conjunto de 9 parches, cada uno de ellos dado por parametrizaciones del tipo

$$x = p_1(s, t), \quad y = p_2(s, t), \quad z = p_3(s, t)$$

definidas en el dominio:

$$s_0 = s_1 < s_2 < s_3 < s_4 = s_5$$

$$t_0 = t_1 < t_2 < t_3 < t_4 = t_5$$

y donde:

$$\begin{aligned} p_1 &= \left(x_{00}f_0 + x_{10}f_1 + x_{20}f_2 + x_{30}f_3 \right)g_0 + \left(x_{01}f_0 + x_{11}f_1 + x_{21}f_2 + x_{31}f_3 \right)g_1 + \\ &\quad \left(x_{02}f_0 + x_{12}f_1 + x_{22}f_2 + x_{32}f_3 \right)g_2 + \left(x_{03}f_0 + x_{13}f_1 + x_{23}f_2 + x_{33}f_3 \right)g_3 \\ p_2 &= \left(y_{00}f_0 + y_{10}f_1 + y_{20}f_2 + y_{30}f_3 \right)g_0 + \left(y_{01}f_0 + y_{11}f_1 + y_{21}f_2 + y_{31}f_3 \right)g_1 + \\ &\quad \left(y_{02}f_0 + y_{12}f_1 + y_{22}f_2 + y_{32}f_3 \right)g_2 + \left(y_{03}f_0 + y_{13}f_1 + y_{23}f_2 + y_{33}f_3 \right)g_3 \\ p_3 &= \left(z_{00}f_0 + z_{10}f_1 + z_{20}f_2 + z_{30}f_3 \right)g_0 + \left(z_{01}f_0 + z_{11}f_1 + z_{21}f_2 + z_{31}f_3 \right)g_1 + \\ &\quad \left(z_{02}f_0 + z_{12}f_1 + z_{22}f_2 + z_{32}f_3 \right)g_2 + \left(z_{03}f_0 + z_{13}f_1 + z_{23}f_2 + z_{33}f_3 \right)g_3 \end{aligned}$$

$$\left. \begin{array}{l} f_0 = \frac{s_2-s}{s_2-s_0} \\ f_1 = \frac{s-s_0}{s_2-s_0} \\ f_2 = 0.0 \\ f_3 = 0.0 \end{array} \right\} \text{ si } s_0 \leq s < s_2$$

$$\left. \begin{array}{l} f_0 = 0.0 \\ f_1 = \frac{s_3-s}{s_3-s_2} \\ f_2 = \frac{s-s_2}{s_3-s_2} \\ f_3 = 0.0 \end{array} \right\} \text{ si } s_2 \leq s < s_3$$

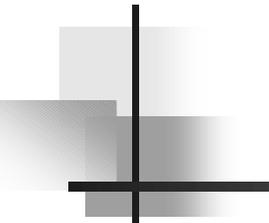
$$\left. \begin{array}{l} f_0 = 0.0 \\ f_1 = 0.0 \\ f_2 = \frac{s_4-s}{s_4-s_3} \\ f_3 = \frac{s-s_3}{s_4-s_3} \end{array} \right\} \text{ si } s_3 \leq s < s_4$$

$$\left. \begin{array}{l} g_0 = \frac{t_2-t}{t_2-t_0} \\ g_1 = \frac{t-t_0}{t_2-t_0} \\ g_2 = 0.0 \\ g_3 = 0.0 \end{array} \right\} \text{ si } t_0 \leq t < t_2$$

$$\left. \begin{array}{l} g_0 = 0.0 \\ g_1 = \frac{t_3-t}{t_3-t_2} \\ g_2 = \frac{t-t_2}{t_3-t_2} \\ g_3 = 0.0 \end{array} \right\} \text{ si } t_2 \leq t < t_3$$

$$\left. \begin{array}{l} g_0 = 0.0 \\ g_1 = 0.0 \\ g_2 = \frac{t_4-t}{t_4-t_3} \\ g_3 = \frac{t-t_3}{t_4-t_3} \end{array} \right\} \text{ si } t_3 \leq t < t_4$$

Los valores concretos



$x_{00} = -402.396210422345$	$y_{00} = 125.604939074148$	$z_{00} = 76.0994513388919$
$x_{10} = -261.17588628063$	$y_{10} = 120.007923434398$	$z_{10} = 80.2658802206019$
$x_{20} = -119.951398608196$	$y_{20} = 114.410862351266$	$z_{20} = 84.4324639780064$
$x_{30} = 21.2789144514636$	$y_{30} = 108.813733891507$	$z_{30} = 88.5992634123272$
$x_{01} = -402.301520929449$	$y_{01} = 57.6898233810474$	$z_{01} = 57.9054501358602$
$x_{11} = -261.108042469145$	$y_{11} = 53.9585848013379$	$z_{11} = 62.570735352082$
$x_{21} = -119.914564008841$	$y_{21} = 50.2273462216285$	$z_{21} = 67.2360205683038$
$x_{31} = 21.2789144514635$	$y_{31} = 46.496107641919$	$z_{31} = 71.9013057845257$
$x_{02} = -402.274168598392$	$y_{02} = 38.0716305763488$	$z_{02} = 52.6498686670318$
$x_{12} = -261.089807581774$	$y_{12} = 36.206011286494$	$z_{12} = 57.8146792486379$
$x_{22} = -119.905446565155$	$y_{22} = 34.3403919966393$	$z_{22} = 62.979489830244$
$x_{32} = 21.2789144514634$	$y_{32} = 32.4747727067845$	$z_{32} = 68.1443004118501$
$x_{03} = -402.179479105497$	$y_{03} = -29.8434851167516$	$z_{03} = 34.4558674640001$
$x_{13} = -261.021963770289$	$y_{13} = -29.8433273465661$	$z_{13} = 40.119534380118$
$x_{23} = -119.8686119658$	$y_{23} = -29.8431241329984$	$z_{23} = 45.7830464205415$
$x_{33} = 21.2789144514633$	$y_{33} = -29.8428535428032$	$z_{33} = 51.4463427840486$

$$s_0 = 0.0$$

$$s_2 = 0.3333333333333333$$

$$s_3 = 0.6666666666666666$$

$$s_4 = 1.0$$

$$t_0 = -1.14816198642716$$

$$t_2 = 0.3333333333333333$$

$$t_3 = 0.6666666666666666$$

$$t_4 = 2.14816198642715$$

$$s_0 = s_1, \quad s_4 = s_5, \quad t_0 = t_1, \quad t_4 = t_5$$

La ecuación implícita genérica

$$c_{10}^2 c_{22}^2 + (-c_{10} c_{11} c_{21} + (-2c_{10} c_{12} + c_{11}^2) c_{20}) c_{22} + c_{10} c_{12} c_{21}^2 - c_{11} c_{12} c_{20} c_{21} + c_{12}^2 c_{20}^2$$

donde

$$c_{10} = ((s^2 - s_0) * t_0 * t_0 * x_{01} + (-s^2 + s_0) * t_0 * t_2 * x_{00} + ((s^2 - s_0) * t_0 * t_2 + (-s^2 + s_0) * t_0 * t_0) * x) * z_{11} + ((-s^2 + s_0) * t_0 * t_2 * x_{01} + (s^2 - s_0) * t_2 * t_2 * x_{00} + ((-s^2 + s_0) * t_2 * t_2 + (s^2 - s_0) * t_0 * t_2) * x) * z_{10} + ((-s^2 + s_0) * t_0 * t_0 * x_{11} + (s^2 - s_0) * t_0 * t_2 * x_{10} + ((-s^2 + s_0) * t_0 * t_2 + (s^2 - s_0) * t_0 * t_0) * x) * z_{01} + ((s^2 - s_0) * t_0 * t_2 * x_{11} + (-s^2 + s_0) * t_2 * t_2 * x_{10} + ((s^2 - s_0) * t_2 * t_2 + (-s^2 + s_0) * t_0 * t_2) * x) * z_{00} + (((-s^2 + s_0) * t_0 * t_2 + (s^2 - s_0) * t_0 * t_0) * x_{11} + ((s^2 - s_0) * t_2 * t_2 + (-s^2 + s_0) * t_0 * t_2) * x_{10} + ((s^2 - s_0) * t_0 * t_2 + (-s^2 + s_0) * t_0 * t_0) * x_{01} + ((-s^2 + s_0) * t_2 * t_2 + (s^2 - s_0) * t_0 * t_2) * x_{00}) * z$$

$$c_{11} = (((-2 * s^2) + 2 * s_0) * t_0 * x_{01} + ((s^2 - s_0) * t_2 + (s^2 - s_0) * t_0) * x_{00} + ((-s^2 + s_0) * t_2 + (s^2 - s_0) * t_0) * x) * z_{11} + (((s^2 - s_0) * t_2 + (s^2 - s_0) * t_0) * x_{01} + ((-2 * s^2) + 2 * s_0) * t_2 * x_{00} + ((s^2 - s_0) * t_2 + (-s^2 + s_0) * t_0) * x) * z_{10} + ((2 * s^2 + (-2 * s_0)) * t_0 * x_{11} + ((-s^2 + s_0) * t_2 + (-s^2 + s_0) * t_0) * x_{10} + ((s^2 - s_0) * t_2 + (-s^2 + s_0) * t_0) * x) * z_{01} + (((-s^2 + s_0) * t_2 + (-s^2 + s_0) * t_0) * x_{11} + (2 * s^2 + (-2 * s_0)) * t_2 * x_{10} + ((-s^2 + s_0) * t_2 + (s^2 - s_0) * t_0) * x) * z_{00} + (((s^2 - s_0) * t_2 + (-s^2 + s_0) * t_0) * x_{11} + ((-s^2 + s_0) * t_2 + (s^2 - s_0) * t_0) * x_{10} + ((-s^2 + s_0) * t_2 + (s^2 - s_0) * t_0) * x_{01} + ((s^2 - s_0) * t_2 + (-s^2 + s_0) * t_0) * x_{00}) * z$$

$$c12 = ((s2-s0)*x01 + (-s2+s0)*x00)*z11 + ((-s2+s0)*x01 + (s2-s0)*x00)*z10 + ((-s2+s0)*x11 + (s2-s0)*x10)*z01 + ((s2-s0)*x11 + (-s2+s0)*x10)*z00$$

$$c20 = ((s2-s0)*t0*t0*y01 + (-s2+s0)*t0*t2*y00 + ((s2-s0)*t0*t2 + (-s2+s0)*t0*t0)*y)*z11 + ((-s2+s0)*t0*t2*y01 + (s2-s0)*t2*t2*y00 + ((-s2+s0)*t2*t2 + (s2-s0)*t0*t2)*y)*z10 + ((-s2+s0)*t0*t0*y11 + (s2-s0)*t0*t2*y10 + ((-s2+s0)*t0*t2 + (s2-s0)*t0*t0)*y)*z01 + ((s2-s0)*t0*t2*y11 + (-s2+s0)*t2*t2*y10 + ((s2-s0)*t2*t2 + (-s2+s0)*t0*t2)*y)*z00 + (((-s2+s0)*t0*t2 + (s2-s0)*t0*t0)*y11 + ((s2-s0)*t2*t2 + (-s2+s0)*t0*t2)*y10 + ((s2-s0)*t0*t2 + (-s2+s0)*t0*t0)*y01 + ((-s2+s0)*t2*t2 + (s2-s0)*t0*t2)*y00)*z$$

$$c21 = (((-2*s2) + 2*s0)*t0*y01 + ((s2-s0)*t2 + (s2-s0)*t0)*y00 + ((-s2+s0)*t2 + (s2-s0)*t0)*y)*z11 + (((s2-s0)*t2 + (s2-s0)*t0)*y01 + ((-2*s2) + 2*s0)*t2*y00 + ((s2-s0)*t2 + (-s2+s0)*t0)*y)*z10 + ((2*s2 + (-2*s0))*t0*y11 + ((-s2+s0)*t2 + (-s2+s0)*t0)*y10 + ((s2-s0)*t2 + (-s2+s0)*t0)*y)*z01 + (((-s2+s0)*t2 + (-s2+s0)*t0)*y11 + (2*s2 + (-2*s0))*t2*y10 + ((-s2+s0)*t2 + (s2-s0)*t0)*y)*z00 + (((s2-s0)*t2 + (-s2+s0)*t0)*y11 + ((-s2+s0)*t2 + (s2-s0)*t0)*y10 + ((-s2+s0)*t2 + (s2-s0)*t0)*y01 + ((s2-s0)*t2 + (-s2+s0)*t0)*y00)*z$$

$$c22 = ((s2-s0)*y01 + (-s2+s0)*y00)*z11 + ((-s2+s0)*y01 + (s2-s0)*y00)*z10 + ((-s2+s0)*y11 + (s2-s0)*y10)*z01 + ((s2-s0)*y11 + (-s2+s0)*y10)*z00$$

La Ecuación Implícita Concreta

$$\begin{aligned} & \left[-12.86405481303769 \right] z^3 + \\ & \left[6.8938261970858576y + 1.0322819534013938x + 4441.0953941817452 \right] z^2 + \\ & \left[-0.92359758113477841y^2 + (-0.27659911577510787x - 1976.073857399886)y - \right. \\ & \quad \left. 0.020708984196767578x^2 - 295.89741941049658x - 387924.22771793336 \right] z + \\ & \left[(3.4188084068007236 \cdot 10^{-18}x + 210.63109093399621)y^2 + \right. \\ & \quad (-6.1121850507742521 \cdot 10^{-19}x^2 + 63.079824695413464x + \\ & \quad 92111.888130195322)y + 4.7227883903081773x^2 + 13792.839722405979x + \\ & \quad \left. 10070450.541378483 \right] \end{aligned}$$

El control del error

$$H(p_1(s, t), p_2(s, t), p_3(s, t)) =$$
$$\begin{aligned} & (-1.7970904234100502 \cdot 10^{-15} s^3 - 3.5539035642606386 \cdot 10^{-14} s^2 + \\ & 1.3866796036202881 \cdot 10^{-12} s - 8.3765572022996195 \cdot 10^{-12}) t^3 + \\ & (5.3954545200343888 \cdot 10^{-14} s^3 - 2.7284841053187847 \cdot 10^{-12} s^2 + \\ & 1.1641532182693481 \cdot 10^{-10}) t^2 + (-1.1527137854777534 \cdot 10^{-12} s^3 + \\ & 2.9103830456733704 \cdot 10^{-11} s^2 - 1.1641532182693481 \cdot 10^{-10} s - \\ & 2.7939677238464355 \cdot 10^{-9}) t - 1.7043986506752499 \cdot 10^{-11} s^3 + \\ & 4.6566128730773926 \cdot 10^{-9} s + 3.7252902984619141 \cdot 10^{-9} \end{aligned}$$

$$\frac{\sum_{i=1}^N H(x_i, y_i, z_i)}{N} = 5.5134296417236327 \cdot 10^{-9}$$

Generación de ecuaciones implícitas optimizadas para su evaluación

Mediante resultantes, Bases de Gröbner, ... la ecuación implícita se obtiene en forma expandida:

- Una expresión algebraica enorme.
- Una expresión difícil a evaluar.
- Ausencia de relación clara entre las ecuaciones paramétricas iniciales y la ecuación implícita obtenida

$$f(T) = a_0 T^m + a_1 T^{m-1} + \dots + a_{m-1} T + a_m = a_0 \prod_{j=1}^s (T - \alpha_j)^{e_j}$$

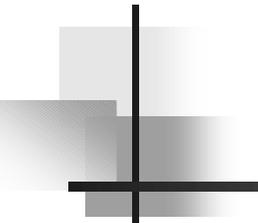
Sumas de Newton ($k \geq 0$):

$$\mathbf{S}_k = e_1 \alpha_1^k + \dots + e_s \alpha_s^k$$

Identidades de Newton ($j \geq 1$):

$$j a_j = -\mathbf{S}_j a_0 - \mathbf{S}_{j-1} a_1 - \dots - \mathbf{S}_1 a_{j-1}$$

(si $i > m$ entonces $a_i = 0$)



$$\frac{f'(T)}{f(T)} = \frac{\mathbf{S}_0(\{\alpha_j\}_j)}{T} + \frac{\mathbf{S}_1(\{\alpha_j\}_j)}{T^2} + \frac{\mathbf{S}_2(\{\alpha_j\}_j)}{T^3} + \dots$$

$$\frac{f'(T)g(T)}{f(T)} = \dots + \frac{\mathbf{S}_0(\{g(\alpha_j)\}_j)}{T} + \frac{\mathbf{S}_1(\{g(\alpha_j)\}_j)}{T^2} + \dots$$

Implicitación no expandida de curvas

The implicit equation of \mathcal{C} defined by

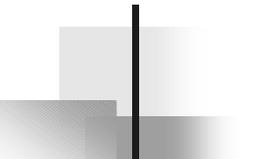
$$x = f(t) \qquad y = g(t)$$

is a divisor of

$$\mathcal{H}_{\mathcal{C}}(x, y) = \text{Resultant}_t(f(t) - x, g(t) - y) = \prod_{\{\tilde{t} : g(\tilde{t}) - y = 0\}} (x - f(\tilde{t}))$$

The coefficients of $\mathcal{H}_{\mathcal{C}}(x, y)$ (as polynomial in x) are determined by computing the Newton Sums on the t -solutions of

$$g(t) - y = 0$$


$$\mathbf{x=f(t), y=g(t)}$$

$$n = \deg(g).$$

$$\mathcal{H}_C(x, y) = x^n + r_1(y)x^{n-1} + \dots + r_n(y)$$

$$k \in \{1, \dots, n\}:$$

$$k \cdot r_k(y) = \mathbf{S}_k(y) - \mathbf{S}_{k-1}(y) \cdot r_1(y) - \dots - \mathbf{S}_1(y) \cdot r_{k-1}(y)$$

Every $\mathbf{S}_j(y)$ is the coefficient of t^{-1} in the expansion in negative powers of t of:

$$\frac{(f(t))^j g'(t)}{g(t) - y}$$

Otro ejemplo

Un ejemplo

$$x = f(t) = t^4 - t + 1 \quad y = g(t) = t^3 + t + 1$$

$$\mathcal{H}_C(x, y) = x^3 + r_1(y)x^2 + r_2(y)x + r_3(y)$$

$$\begin{aligned} \frac{3t^2+1}{t^3+t+1-y} &= \frac{3}{t} - \frac{2}{t^3} - \frac{-3+3y}{t^4} + \frac{2}{t^5} + \frac{5-5y}{t^6} + \frac{1-6y+3y^2}{t^7} - \frac{-7+7y}{t^8} - \frac{-6+16y-8y^2}{t^9} \\ &+ \frac{6-9y^2+3y^3}{t^{10}} + \frac{13-30y+15y^2}{t^{11}} - \frac{-22y+33y^2-11y^3}{t^{12}} - \frac{-19+36y-6y^2-12y^3+3y^4}{t^{13}} \\ &- \frac{-13+65y-78y^2+26y^3}{t^{14}} + \dots \end{aligned}$$

$$\mathbf{S}_1(y) = 5$$

$$\mathbf{S}_2(y) = -11 + 26y - 8y^2$$

$$\mathbf{S}_3(y) = -76 + 93y + 6y^2 - 21y^3 + 3y^4$$

$$\mathcal{H}_C(x, y) = x^3 - 5x^2 + (18 - 13y + 4y^2)x - 23 + 34y - 22y^2 + 7y^3 - y^4$$

Implicitación no expandida de superficies

$$\mathcal{S} : \begin{cases} x = f(t_1, t_2) \\ y = g(t_1, t_2) \\ z = h(t_1, t_2) \end{cases}$$

$$\mathcal{H}_{\mathcal{S}}(x, y, z) = \prod_{(\tilde{t}_1, \tilde{t}_2) \in \mathcal{V}} (x - f(\tilde{t}_1, \tilde{t}_2))$$

$$\mathcal{V} = \left\{ (\tilde{t}_1, \tilde{t}_2) \in \mathbb{C}(\widetilde{y, z}) : \begin{array}{l} g(\tilde{t}_1, \tilde{t}_2) - y = 0 \\ h(\tilde{t}_1, \tilde{t}_2) - z = 0 \end{array} \right\}$$

$$x = f \quad y = t_1^{n_1} + Q_1(t_1, t_2) \quad z = t_2^{n_2} + Q_2(t_1, t_2)$$

$$\deg(Q_i) < n_i, n = \deg(f), m = n_1 n_2$$

$$\mathcal{H}_S(x, y, z) = x^m + r_1(y, z)x^{m-1} + \dots + r_{m-1}(y, z)x + r_m(y, z)$$

$k \in \{1, \dots, n\}$:

$$k \cdot r_k(y, z) = -\mathbf{S}_k(y, z) - \mathbf{S}_{k-1}(y, z) \cdot r_1(y, z) - \dots - \mathbf{S}_1(y, z) \cdot r_{k-1}(y, z)$$

Every \mathbf{S}_j is the coefficient of $t_1^{-1}t_2^{-1}$ in:

$$\frac{(f(t_1, t_2))^j \mathbf{Jac}}{t_1^{n_1} t_2^{n_2}} \sum_{k=0}^{jn} (-1)^k \sum_{\alpha_1 + \alpha_2 = k} \left(\frac{Q_1(t_1, t_2) - y}{t_1^{n_1}} \right)^{\alpha_1} \left(\frac{Q_2(t_1, t_2) - z}{t_2^{n_2}} \right)^{\alpha_2}$$

Jac: Jacobian Determinant of $g(t_1, t_2)$ and $h(t_1, t_2)$

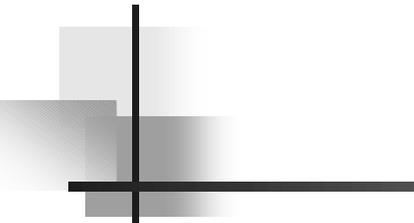
Un ejemplo

$$x = t_1^2 + t_2^2 \quad y = t_1^3 + t_2^2 \quad z = t_2^3 + t_1^2$$

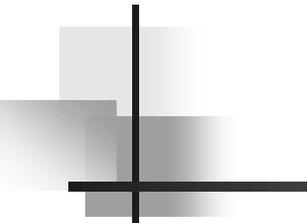
$$\mathcal{H}_S(x, y, z) = x^9 + \sum_{i=1}^9 r_i(y, z)x^{9-i}$$

$$\mathbf{Jac}(t_1, t_2) = t_1 t_2 (9t_1 t_2 - 4)$$

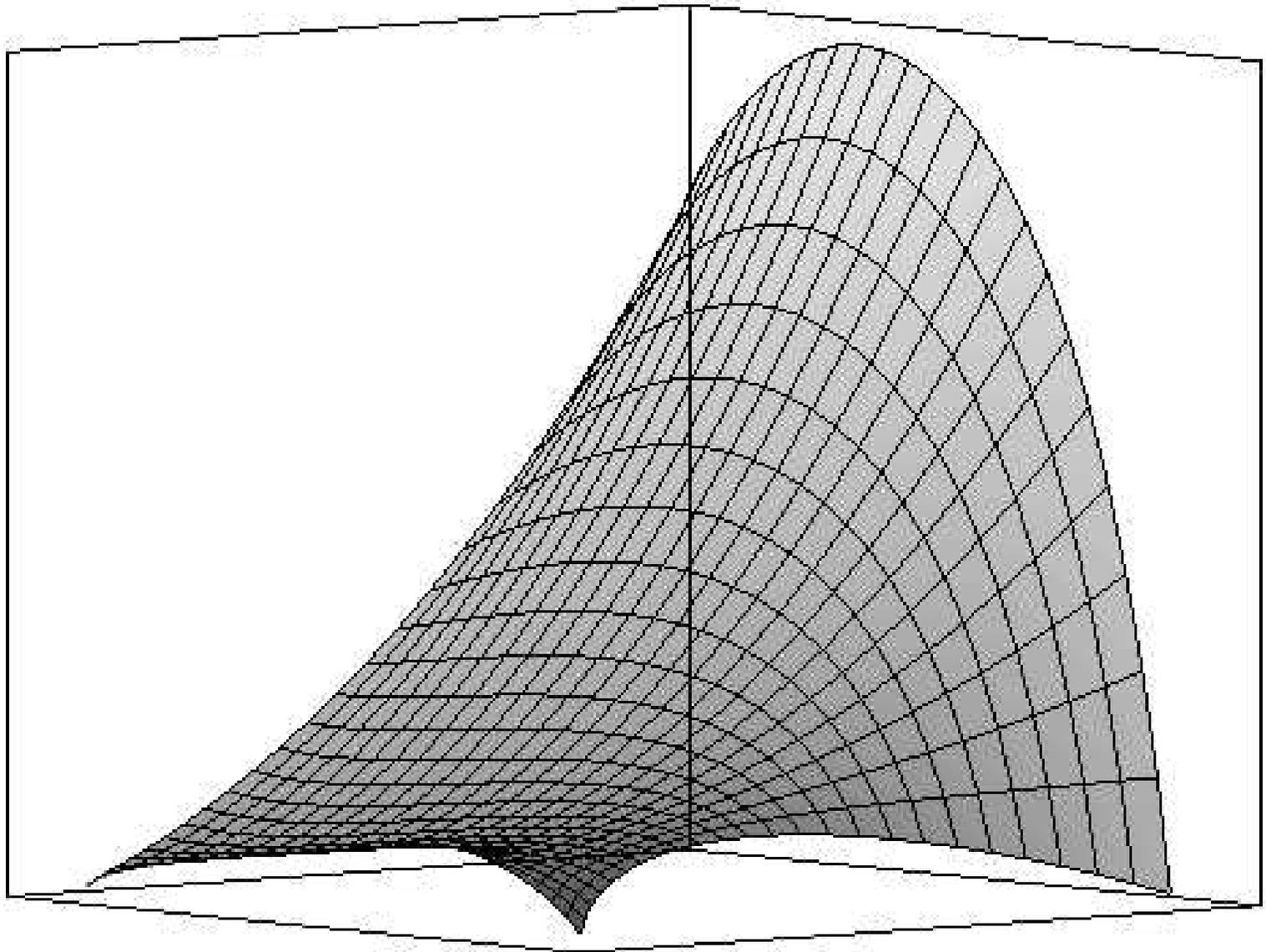
$$\frac{(9t_1 t_2 - 4)(t_1^2 + t_2^2)^j}{t_1^2 t_2^2} \cdot \sum_{k=0}^{2j} (-1)^k \sum_{i=0}^k \left(\frac{t_2^2 - y}{t_1^3} \right)^i \left(\frac{t_1^2 - z}{t_2^3} \right)^{k-i}$$



$$\begin{aligned}
\mathbf{S}_1(y) &= 0 \\
\mathbf{S}_2(y) &= 10 \\
\mathbf{S}_3(y) &= 9x^2 - 42z - 42y + 9y^2 \\
\mathbf{S}_4(y) &= 36x^2 - 24y + 192yz - 24z + 30 + 36y^2 \\
\mathbf{S}_5(y) &= 210x^2 + 10 - 180yx^2 - 240z - 240y \\
&\quad + 120yz + 210y^2 - 180y^2z \\
\mathbf{S}_6(y) &= -612y^2z + 9x^4 + 630y^2 + 100 + 9y^4 \\
&\quad - 612yx^2 + 2160yz - 480x^3 - 480y^3 \\
&\quad - 228z + 180y^2x^2 + 630x^2 - 228y \\
\mathbf{S}_7(y) &= 1428y^3z + 70 + 2037y^2 + 1428yx^3 \\
&\quad - 6300yx^2 + 2037x^2 - 6300y^2z \\
&\quad + 2184yz + 756y^2x^2 - 812y^3 - 1190y \\
&\quad - 812x^3 + 315x^4 - 1190z + 315y^4 \\
\mathbf{S}_8(y) &= 350 - 1008y^3x^2 - 1504y + 5488y^2 \\
&\quad + 1172x^4 + 5488x^2 - 1504z - 7872y^3 \\
&\quad - 1008y^4z - 1008y^2x^3 + 1172y^4 \\
&\quad + 7712yx^3 + 7712y^3z + 20160y^2x^2 \\
&\quad - 7872x^3 - 1008yx^4 - 13872y^2z \\
&\quad - 13872yx^2 + 15680yz \\
\mathbf{S}_9(y) &= -5598z - 5598y + 14499y^4 - 15036x^3 \\
&\quad + 14499x^4 + 360 - 15036y^3 + 14967y^2 \\
&\quad + 14967x^2 + 21024yz + 9y^6 \\
&\quad - 1800y^5 + 52164yx^3 - 77616y^2z \\
&\quad - 25884y^2x^3 + 52164y^3z - 1800x^5 \\
&\quad - 5274yx^4 - 25884y^3x^2 + 49680y^2x^2 \\
&\quad + 756x^4y^2 + 756x^2y^4 - 5274y^4z + 9x^6 \\
&\quad - 77616yx^2
\end{aligned}$$



$$\begin{aligned}
& -2y^4 - 2z^4 - 96x^2yz^2 - 24x^2yz - 21x^3y^2z^2 \\
& + 44xy^3z + 18xy^2z^3 + 60x^3y^2z + 60x^3yz^2 \\
& + 76x^3yz + 18xy^4z + 24xy^2z - 54x^2y^2z^2 \\
& - 60x^2y^3z - 96x^2y^2z - 60x^2yz^3 + 78xy^2z^2 \\
& + 18xy^3z^2 + 18xyz^4 + 3x^3z^4 + 38x^3y^2 \\
& - 22x^4y - 24x^4yz - 22x^4z + 36x^4y^2z \\
& - 9x^5y^2 + 36x^4yz^2 + 6x^5y + 14x^6y - 9x^5z^2 \\
& - 48x^5yz - 3x^6y^2 - 3x^6z^2 + 14x^6z - 12x^2y^2 \\
& + 44xyz^3 - 12x^2z^2 + 38x^3y^3 - 18x^2z^4 \\
& - 18x^2y^4 - 28x^2z^3 + 8x^3y + 8x^3z + 38x^3z^2 \\
& + 38x^3z^3 + 3x^3y^4 + 24xyz^2 + 5xz^4 + 8xy^3 \\
& + 8xz^3 + x^9 - 5x^7 + 5x^5 - 2x^4 - y^6 + 2y^5 \\
& - 2y^4z - 16y^2z^3 - 8y^3z - z^6 + 2z^5 - 2yz^4 \\
& + 5xy^4 - 8yz^3 - 16y^3z^2 - 12y^2z^2 - 3z^4y^2 \\
& - 27x^4y^2 - 27x^4z^2 - 28x^2y^3 + 6x^5z - 3z^2y^4
\end{aligned}$$



$$F_1 = \frac{-H_1 + 3H_2}{8} = s^3 + \frac{3t^2}{4} - \frac{15s^2}{8} + \frac{3t}{4} - \frac{3y}{8} + \frac{x}{8} + \frac{3s}{8} + \frac{1}{8}$$

$$F_2 = \frac{3H_1 - H_2}{8} = t^3 - \frac{9t^2}{4} - \frac{3s^2}{8} + \frac{3t}{4} + \frac{y}{8} - \frac{3x}{8} + \frac{15s}{8} - \frac{3}{8}$$

The computing time was 15 seconds by using Maple 6 on a Sparc-15 Workstation (versus the 10^6 seconds in Hoffmann's solution).

$$r_1(x, y) = -\frac{54187594407x^2}{2048} + \frac{48101467761xy}{2048} - \frac{38812918311y^2}{262144} - \frac{8388608}{64} + \frac{16777216}{32} - \frac{81x^2}{64}$$

The size of the file containing the implicit equation (in non expanded form) is around 600 Kbytes.

$$\begin{aligned} & -\frac{54187594407x^2}{16777216} + \frac{48101467761xy}{8388608} - \frac{38812918311y^2}{16777216} \\ & - \frac{22656991982391171}{137438953472} - \frac{1}{2} \left(\frac{233469x}{2048} - \frac{188595y}{2048} \right) \left(-\frac{233469x}{2048} \right. \\ & \left. + \frac{112832595}{262144} + \frac{81x^2}{64} - \frac{135xy}{32} + \frac{81y^2}{64} \right) \left(-\frac{233469x}{2048} \right. \\ & \left. + \frac{188595y}{2048} - \frac{112832595}{262144} - \frac{81x^2}{64} + \frac{135xy}{32} - \frac{81y^2}{64} \right) \end{aligned}$$

La utilización de esta estrategia ha producido una mejora significativa en el software CAD/CAM/CAE CSIS que usa y desarrolla la compañía CANDEMAT:

Para los objetos considerados el tiempo de cálculo de seccionado se ha dividido por 3.

$$S : \begin{cases} x = \frac{f_1(s, t)}{q(s, t)} \\ y = \frac{f_2(s, t)}{q(s, t)} \\ z = \frac{f_3(s, t)}{q(s, t)} \end{cases} \Rightarrow H(x, y, z) = 0$$

$$x = A \Rightarrow H(A, y, z) = 0$$

Algunas estructuras algebraicas no son sencillas de implicitar genericamente:

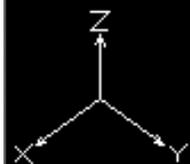
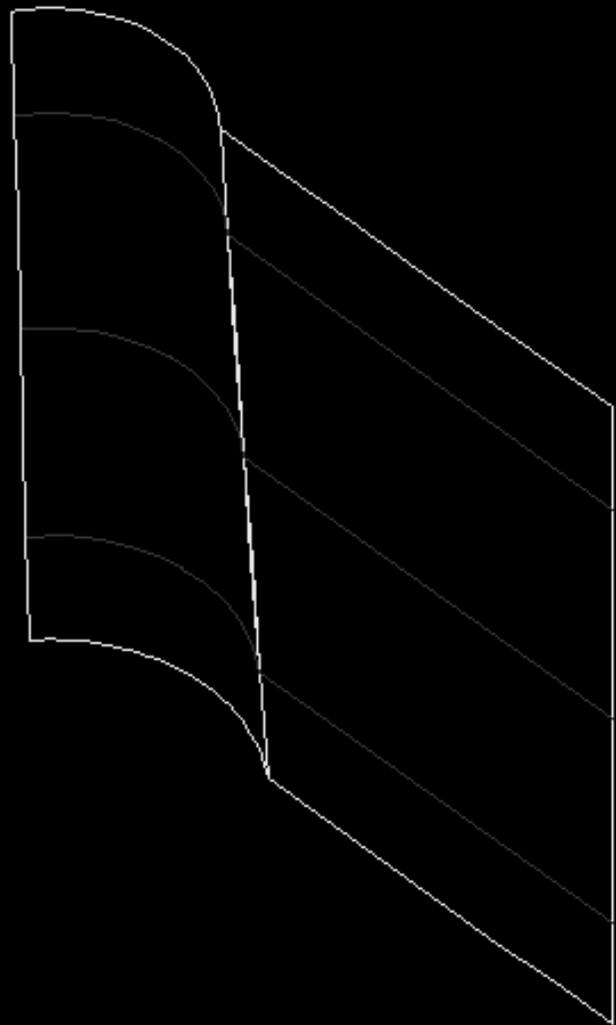
$$x = \frac{f_1(s, t)}{q(s, t)}, \quad y = \frac{f_2(s, t)}{q(s, t)}, \quad z = \frac{f_3(s, t)}{q(s, t)}$$

donde:

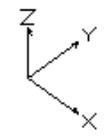
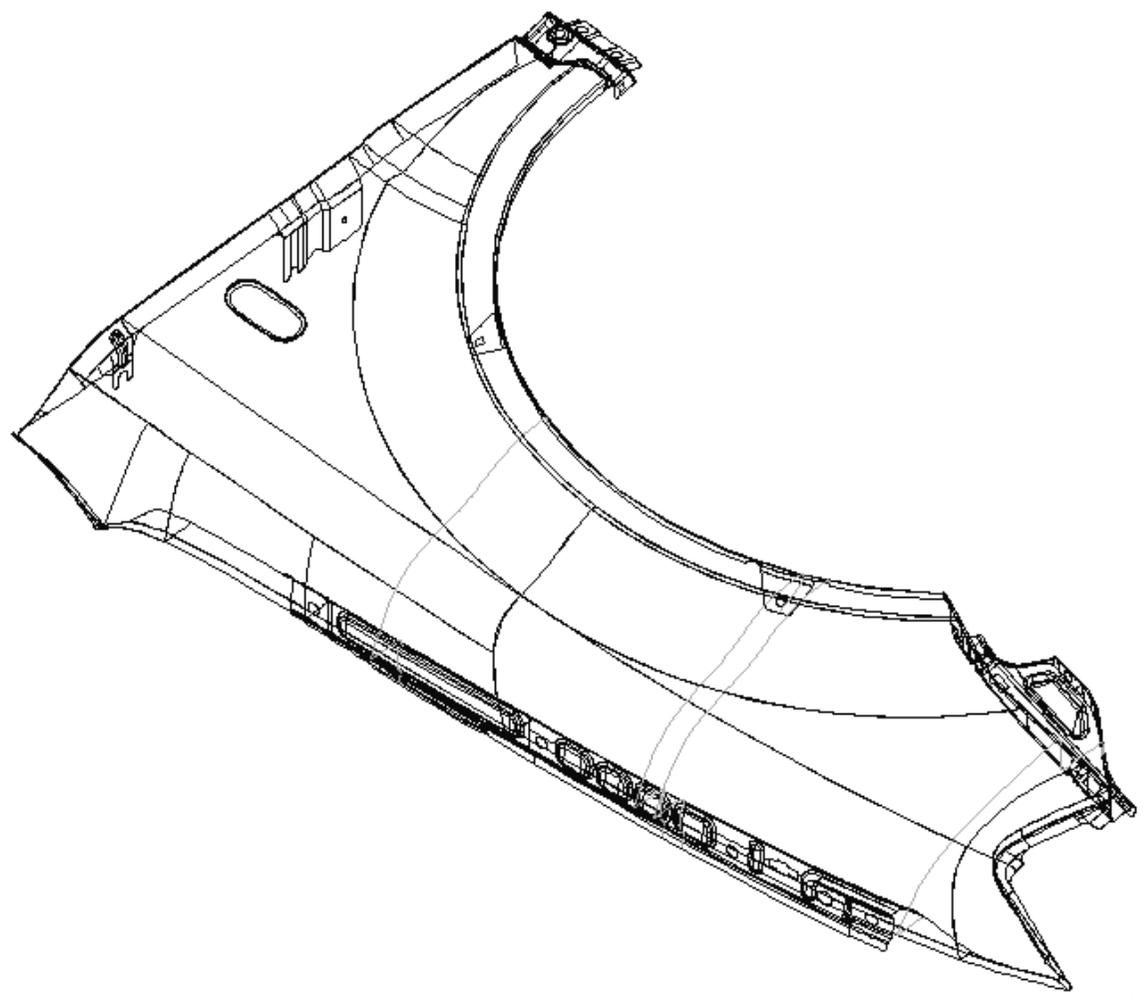
$$q(s, t) = s^3(A_3t^2 + B_3t + C_3) + s^2(A_2t^2 + B_2t + C_2) \\ + s(A_1t^2 + B_1t + C_1) + (A_0t^2 + B_0t + C_0)$$

$$f_i(s, t) = s^3(\alpha_{3i}t^2 + \beta_{3i}t + \gamma_{3i}) + s^2(\alpha_{2i}t^2 + \beta_{2i}t + \gamma_{2i}) \\ + s(\alpha_{1i}t^2 + \beta_{1i}t + \gamma_{1i}) + (\alpha_{0i}t^2 + \beta_{0i}t + \gamma_{0i})$$

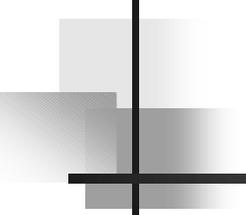
Reducción de grado: Patrikalakis, Hoschek, etc.



X 357.193 Y 83.566



X 481.075 Y -288.2



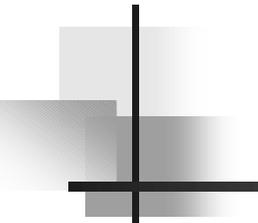
Conclusiones

- ❖ Se obtiene un muy buen comportamiento sobre superficies regladas, superficies de revolución, etc. donde la implicitación generica se comporta muy bien.
- ❖ Los tiempos de seccionado se han reducido.
- ❖ La representacion implícita está disponible para otras tareas CAD (offsetting, blending, etc).

The approximate implicitization factorization

- ❖ [Dokken 1997,2001] Approximate Implicitization
 - ❖ Assume that the surface $\mathbf{p}(s,t)$ has degree (n_1, n_2)
 - ❖ Assume that \mathbf{q} has total degree m and that \mathbf{b} is a vector containing the unknown coefficients of \mathbf{q}
 - ❖ The combination $\mathbf{q}(\mathbf{p}(s,t))$ is a polynomial of degrees (mn_1, mn_2)
 - ❖ Collect basis functions of degree (mn_1, mn_2) in $\alpha(s,t)$
- ⊗ Then $\mathbf{q}(\mathbf{p}(s,t))$ can be factorized

$$\mathbf{q}(\mathbf{p}(s,t)) = (\mathbf{D}\mathbf{b})^T \alpha(s,t)$$



The Factorization

$$q(\mathbf{p}(s,t)) = (\mathbf{D}\mathbf{b})^T \alpha(s,t)$$

- ❖ The matrix \mathbf{D} is built from products of the coefficients of $\mathbf{p}(s,t)$.
- ❖ An element in \mathbf{D} is the product of a maximum of m such coefficients, where m the total degree of q .
- ❖ If $\mathbf{p}(s,t)$ was described in a Bernstein basis of degree (n_1, n_2) then $\alpha(s,t)$ contains the Bernstein basis of degree (mn_1, mn_2) .
- ❖ The first step of moving curves and surfaces use the same factorization.

Properties of the factorization

$$q(\mathbf{p}(s,t)) = (\mathbf{D}\mathbf{b})^T \alpha(s,t)$$

- ❖ If $\mathbf{D}\mathbf{b}=0$ and $\mathbf{b}\neq\mathbf{0}$ then $q(\mathbf{p}(s,t))=0$ and \mathbf{b} describes an implicitization q of $\mathbf{p}(s,t)$.
- ❖ If $\alpha(s,t)$ describes a Bernstein basis then

$$\|\alpha(s,t)\|_2 \leq 1$$

$$|q(\mathbf{p}(s,t))| = |(\mathbf{D}\mathbf{b})^T \alpha(s,t)| \leq \|\mathbf{D}\mathbf{b}\|_2$$

Properties of the inequality

$$|q(\mathbf{p}(s,t))| = |(\mathbf{D}\mathbf{b})^T \alpha(s,t)| \leq \|\mathbf{D}\mathbf{b}\|_2.$$

Let σ_1 be the smallest singular value of \mathbf{D} then

$$\min_{\|\mathbf{b}\|_2=1} \max_{(s,t) \in \Omega} |q(\mathbf{p}(s,t))| \leq \sigma_1.$$

Singular value decomposition (SVD) can be used to find approximate solutions of the implicitization problem.

Piecewise polynomials can be approximated

- ❖ Approximation of multiple manifolds

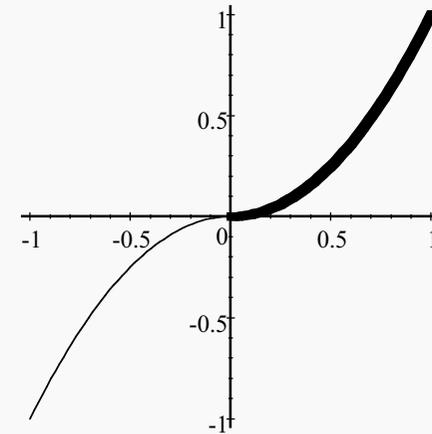
$$\sum_{i=1}^r (q(\mathbf{p}_i(s_i, t_i)))^2 \leq \left\| \begin{pmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_r \end{pmatrix} \mathbf{b} \right\|_2^2.$$

- ❖ D_1 can be based on a curve, D_2 on a surface,...
- ❖ Separation of two manifolds by approximate implicitization of one of the manifolds.

Example: Two Parabolas

$$\mathbf{p}_1(s) = (-1, -1)(1 - s)^2 + \left(-\frac{1}{2}, 0\right) 2(1 - s)s + (0, 0)s^2$$

$$\mathbf{p}_2(s) = (0, 0)(1 - s)^2 + \left(\frac{1}{2}, 0\right) 2(1 - s)s + (1, 1)s^2$$



❖ We want to approximate both curve segments at the same time with one algebraic curve of degree 3.

❖ Thus we will make

- $q(\mathbf{p}_1(s)) = (\mathbf{D}_1 \mathbf{b}) \alpha_1(s)$

- $q(\mathbf{p}_2(s)) = (\mathbf{D}_2 \mathbf{b}) \alpha_2(s)$

⊗ and combine the matrices:

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{pmatrix}$$

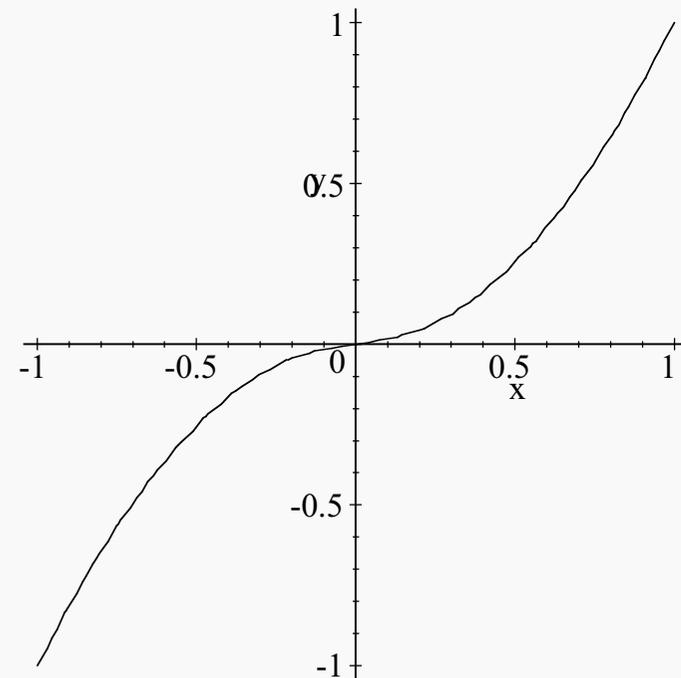
Implicit from combined D

Eigenvalues:

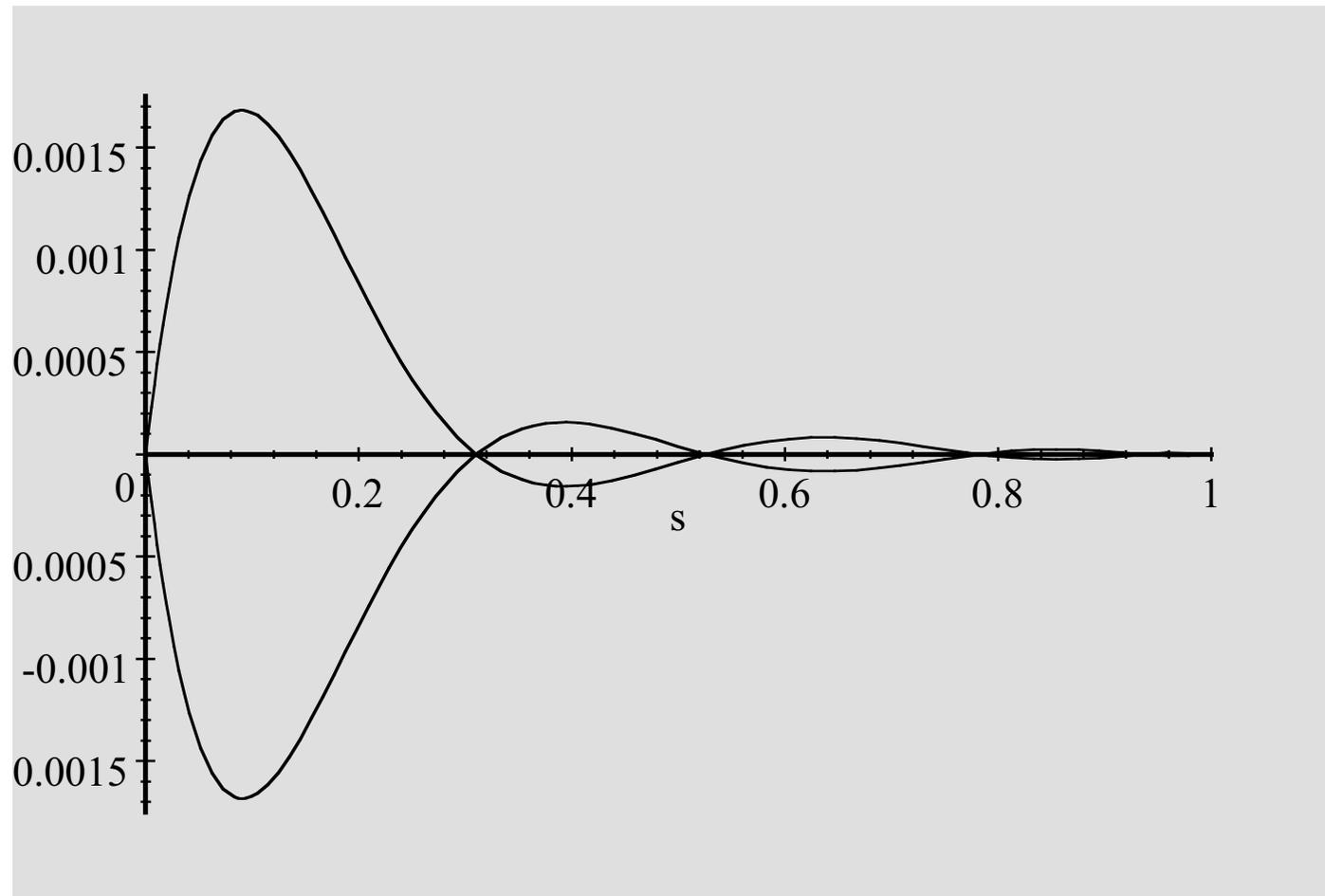
4.25, 3.91, 1.98, 1.31, 0.38, 0.37, 0.11, 0.05, 0.03, 0.007937

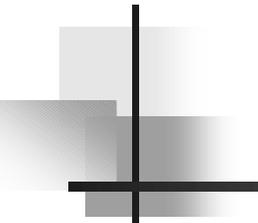
Combining eigenvector with basis functions and plot implicit:

$$\begin{pmatrix} x^3 \\ x^2y \\ xy^2 \\ y^3 \\ x^2 \\ xy \\ y^2 \\ x \\ y \\ 1 \end{pmatrix} \cdot \begin{pmatrix} -.4602815334 \\ .6983314313 \\ -.5087222523 \\ .1433800204 \\ 0 \\ 0 \\ 0 \\ -.0170228764 \\ .1443197267 \\ 0 \end{pmatrix} = 0$$



Error of $p_1(s)$ and $p_2(s)$





OFFSETS

Motivación y definición

Generación ecuaciones implícitas

Control de la topología de la offset

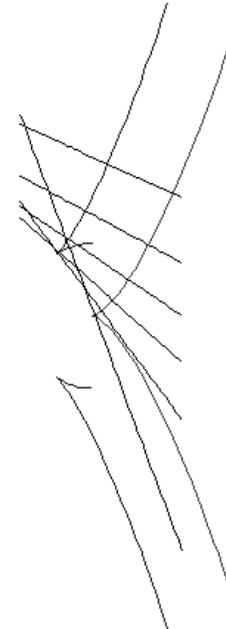
Offsets locales con topología controlada

Definición geométrica de offset de una curva a distancia d

Conjunto de puntos cuya distancia euclídea al correspondiente punto de la curva es d



$$\begin{aligned} y(s) &= s^3 \\ x(s) &= s^2 \end{aligned}$$



Offsets de curvas planas: definiciones algebraicas

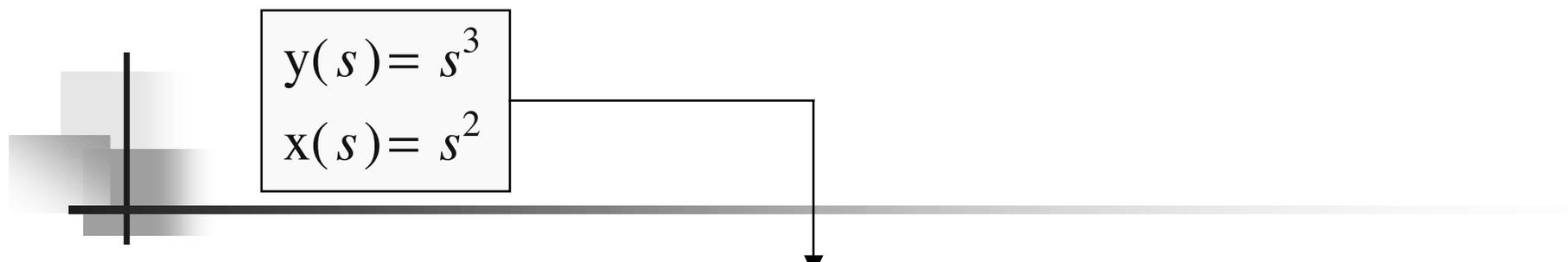
$$f(u, v) := 0$$

$$\begin{aligned} f(u, v) &= 0 \\ (x - u)^2 + (y - v)^2 - d^2 &= 0 \\ (x - u) \left(\frac{\partial}{\partial v} f \right) (u, v) - (y - v) \left(\frac{\partial}{\partial u} f \right) (u, v) &= 0 \end{aligned}$$

$$u = u(t)$$

$$v = v(t)$$

$$\begin{aligned} (u(t) - x)^2 + (v(t) - y)^2 - d^2 &= 0 \\ (x - u(t)) \left(\frac{\partial}{\partial t} u \right) (t) + (y - v(t)) \left(\frac{\partial}{\partial t} v \right) (t) &= 0 \end{aligned}$$


$$y(s) = s^3$$

$$x(s) = s^2$$

$$\begin{aligned} & (x^2 + y^2 - R^2) (729 R^8 + 16 R^4 - 216 R^6 + (-1944 R^6 + 288 R^4) x \\ & + (-32 R^2 + 1188 R^4 - 729 R^6) y^2 + (-2916 R^6 + 1728 R^4) x^2 \\ & + (-504 R^2 + 6318 R^4) y^2 x + (4104 R^4 - 32 R^2) x^3 + (-1701 R^2 + 16) y^4 \\ & + (-2484 R^2 + 2187 R^4) y^2 x^2 + 729 y^6 + (4374 R^4 - 504 R^2) x^4 - 2376 x^5 R^2 + 729 x^8 \\ & + 216 x^7 + (-2916 R^2 + 16) x^6 + (216 - 4374 R^2) y^4 x + (-4860 R^2 - 32) y^2 x^3 \\ & + 729 y^2 x^6 - 1458 y^2 x^5 + (-432 - 2187 R^2) y^2 x^4 - 1458 y^4 x^3 + 729 y^4 x^2) \end{aligned}$$

Factores Extraños

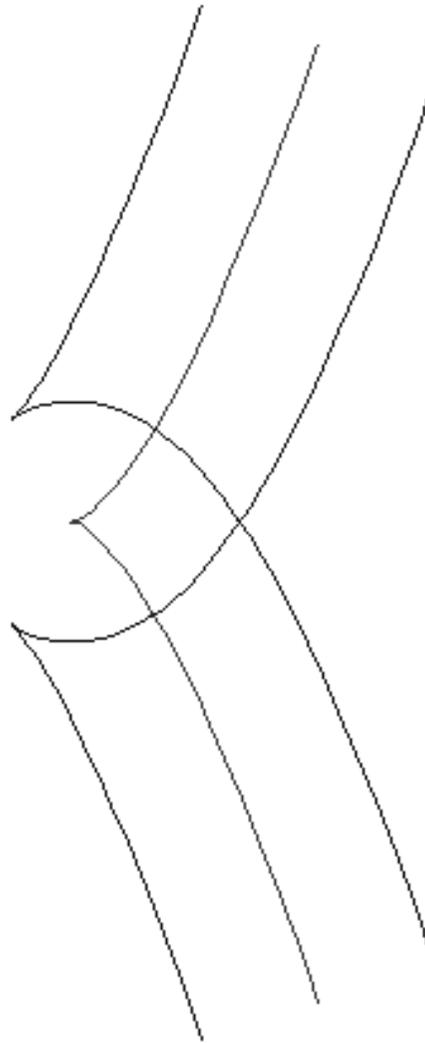
En el caso polinomial, se evitan dividiendo por

$$\gcd\left(\frac{\partial}{\partial s} X(s), \frac{\partial}{\partial s} Y(s)\right)$$

$$\begin{aligned}
& 729x^8 + 729x^6y^2 + 216x^7 - 1458x^5y^2 - 1458x^3y^4 - 11648x^6 - 9180x^4y^2 + 729x^2y^4 + 729y^6 - 9504x^5 \\
& - 19472x^3y^2 - 17280xy^4 + 67968x^4 + 25056x^2y^2 - 6788y^4 + 65536x^3 + 99072xy^2 - 158976x^2 \\
& - 27776y^2 - 119808x + 173056
\end{aligned}$$

$$y(s) = s^3$$

$$x(s) = s^2$$



$$R = 2$$

Offsets de superficies: definiciones algebraicas

$$f(u, v, w) = 0$$

$$f(u, v, w) = 0$$

$$(x - u)^2 + (y - v)^2 + (z - w)^2 - R^2 = 0$$

$$\left[\left(\frac{\partial}{\partial v} f \right) (u, v, w) (x - u) - \left(\frac{\partial}{\partial u} f \right) (u, v, w) (y - v) = 0 \right.$$

$$\left. \left(\frac{\partial}{\partial w} f \right) (u, v, w) (y - v) - \left(\frac{\partial}{\partial v} f \right) (u, v, w) (z - w) = 0 \right.$$

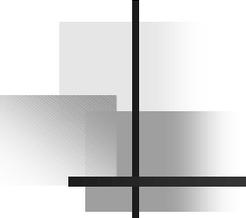
$$\left. \left(\frac{\partial}{\partial u} f \right) (u, v, w) (z - w) - \left(\frac{\partial}{\partial w} f \right) (u, v, w) (x - u) = 0 \right]$$

Offsets de superficies: definiciones algebraicas

$$u = u(s, t), v = v(s, t), w = w(s, t)$$

$$(x - u(s, t))^2 + (y - v(s, t))^2 + (z - w(s, t))^2 - R^2 = 0$$

$$\left[\begin{aligned} \left(\frac{\partial}{\partial s} u \right) (s, t) (x - u(s, t)) + \left(\frac{\partial}{\partial s} v \right) (s, t) (y - v(s, t)) + \left(\frac{\partial}{\partial s} w \right) (s, t) (z - w(s, t)) &= 0 \\ \left(\frac{\partial}{\partial t} u \right) (s, t) (x - u(s, t)) + \left(\frac{\partial}{\partial t} v \right) (s, t) (y - v(s, t)) + \left(\frac{\partial}{\partial t} w \right) (s, t) (z - w(s, t)) &= 0 \end{aligned} \right]$$



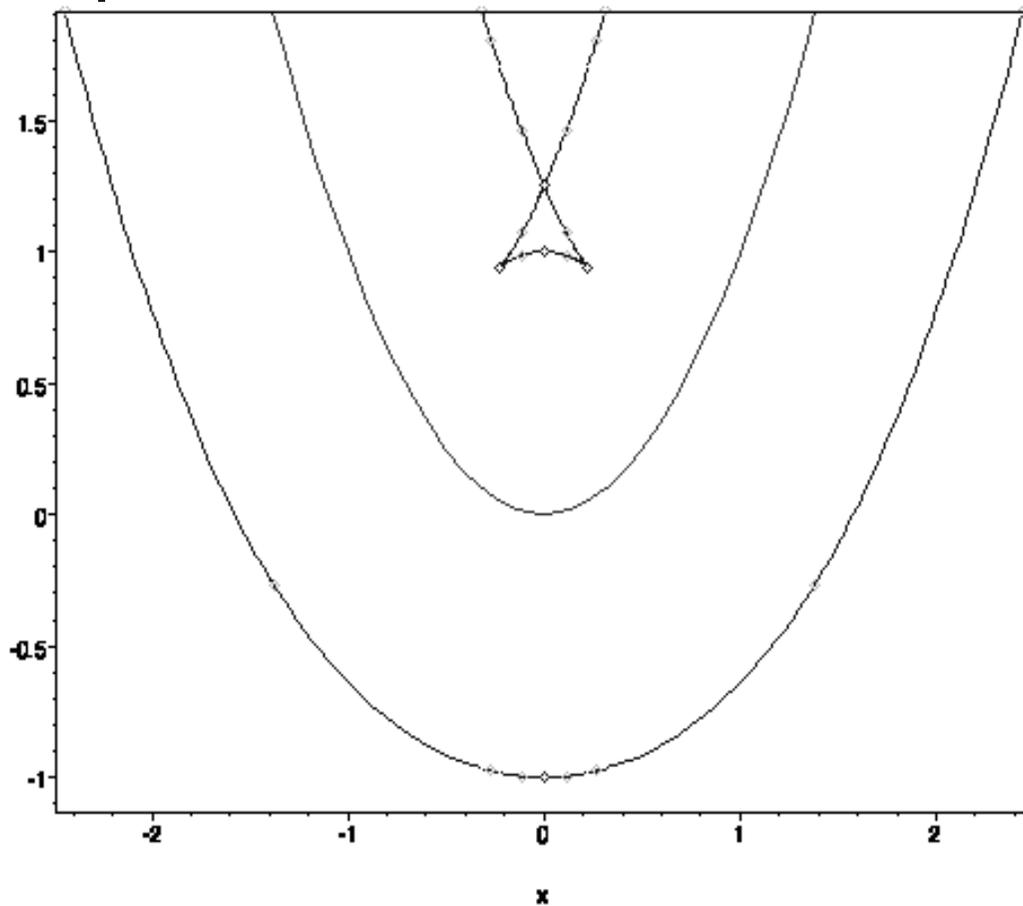
Inconvenientes

- Offsets de curvas/superficies racionales no son en general curvas/superficies racionales.
- El cálculo de Offsets se basa en la eliminación algebraica que requiere coeficientes y aritmética exacta.
- Las ecuaciones implícitas de las Offsets son en general muy grandes y difíciles de manipular.
- El cálculo de Offsets produce en muchos casos factores extraños.

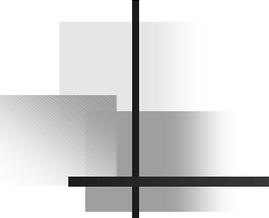
Y además

$$y = x^2 \quad R = 1$$

$$16x^6 + 16x^4y^2 - 40x^4y - 32x^2y^3 - 47x^4 + 16y^4 + 6x^2y - 40y^3 + 28x^2 + 9y^2 + 40y - 25$$



La offset de esta parábola contiene puntos singulares y la forma de una de sus componentes dista mucho de parecerse a una parábola.



Generación de ecuaciones implícitas de Offsets optimizadas para su evaluación

Mediante resultantes, Bases de Gröbner, ...
la ecuación implícita de la Offset se obtiene
en forma expandida:

- Una expresión algebraica enorme.
- Una expresión difícil a evaluar.
- Imposible de calcular cuando los coeficientes de la curva de partida no se conocen de forma exacta.

Implicitación no expandida de Offsets

The implicit equation of the d -offset for the curve \mathcal{C} defined by

$$x = u(t) \qquad y = v(t)$$

is a divisor of

$$\mathcal{H}_{\mathcal{C}}(x, y) = \text{Resultant}_t \left[\begin{array}{l} (u(t) - x)^2 + (v(t) - y)^2 - d^2 \\ (u(t) - x) \frac{\partial u}{\partial t}(t) + (v(t) - y) \frac{\partial v}{\partial t}(t) \end{array} \right] =$$

$$= \prod_{\{\tilde{t} : G(x, y; \tilde{t}) = 0\}} \left((u(\tilde{t}) - x) \frac{\partial u}{\partial t}(\tilde{t}) + (v(\tilde{t}) - y) \frac{\partial v}{\partial t}(\tilde{t}) \right)$$

$$G(x, y; t) = (u(t) - x)^2 + (v(t) - y)^2 - d^2 = 0$$

Control de Topología de Offsets

¿Cuándo la d-offset

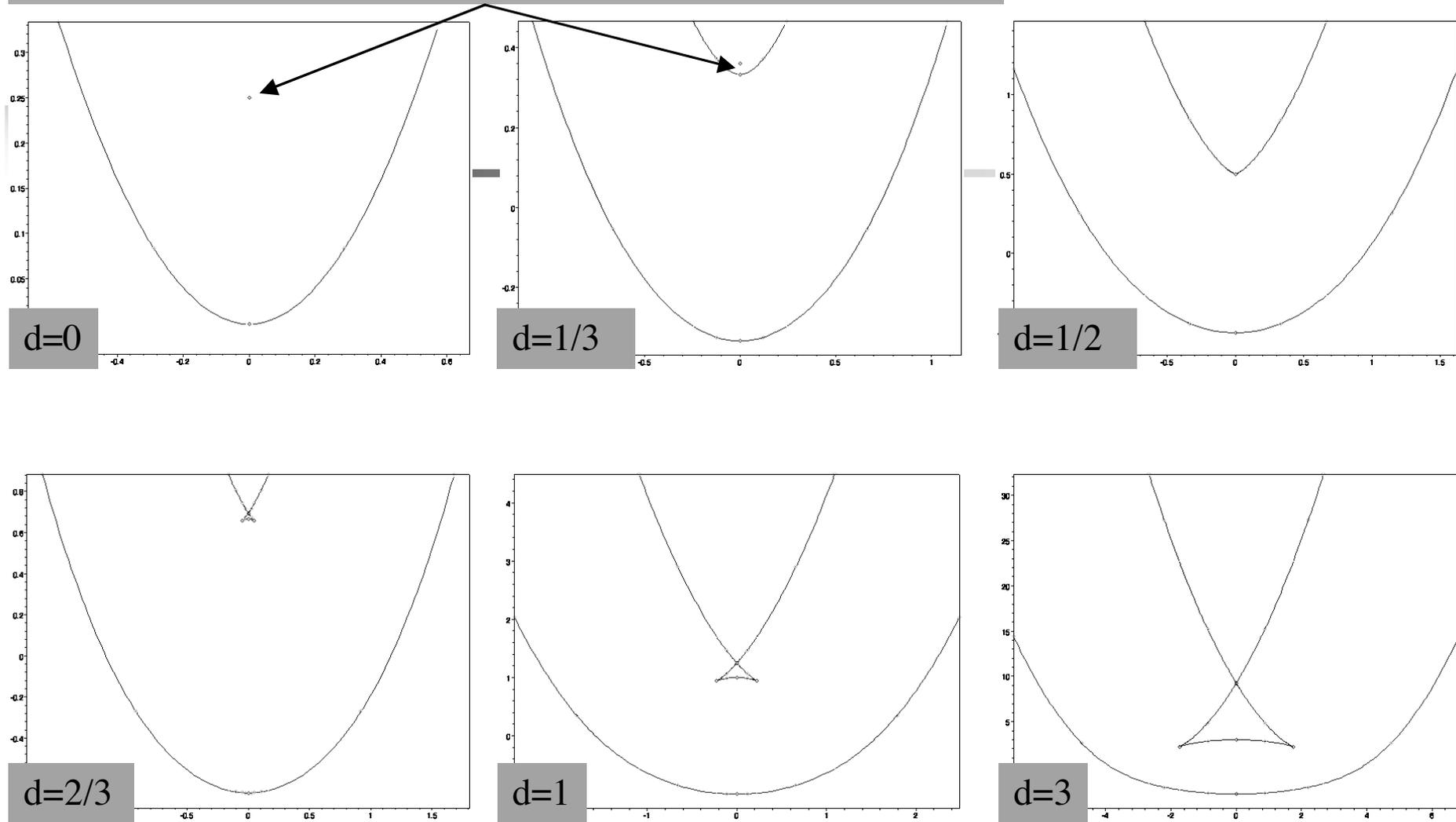
$$\begin{aligned} &16x^6 + 16x^4y^2 - 40x^4y - 32x^2y^3 + (-48d^2 + 1)x^4 + (-32d^2 + 32)x^2y^2 \\ &+ 16y^4 + (-2 + 8d^2)x^2y + (-8 - 32d^2)y^3 + (-20d^2 + 48d^4)x^2 \\ &+ (16d^4 - 8d^2 + 1)y^2 + (32d^4 + 8d^2)y - 16d^6 - 8d^4 - d^2 \end{aligned}$$

de la parábola

$$u(t) = t, v(t) = t^2$$

no es topológicamente una parábola?

El punto aislado viene de la parte compleja



Los cálculos:

Una Eliminación de Cuantificadores sencilla

La offset no es topológicamente una parábola cuando su discriminante con respecto y :

$$x^2 d^2 (64 x^6 + (48 - 192 d^2) x^4 + (192 d^4 + 336 d^2 + 12) x^2 - 64 d^6 - 12 d^2 + 48 d^4 + 1)^3$$

tiene más de 1 raíz real. O, equivalentemente, si el polinomio

$$64 x^3 + (48 - 192 d^2) x^2 + (192 d^4 + 336 d^2 + 12) x - 64 d^6 - 12 d^2 + 48 d^4 + 1$$

tiene 1 raíz real positiva.

Los cálculos:

Una Eliminación de Cuantificadores sencilla

Para cualquier d el número de raíces reales de

$$64 x^3 + (48 - 192 d^2) x^2 + (192 d^4 + 336 d^2 + 12) x - 64 d^6 - 12 d^2 + 48 d^4 + 1$$

es siempre igual a 1: $[1, 1, -d^2, -d^4 (4 d^2 + 1)^2]$

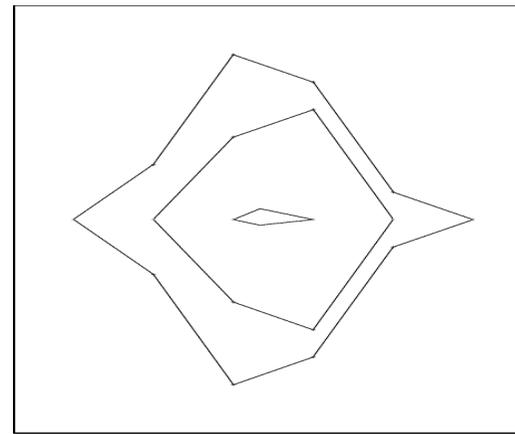
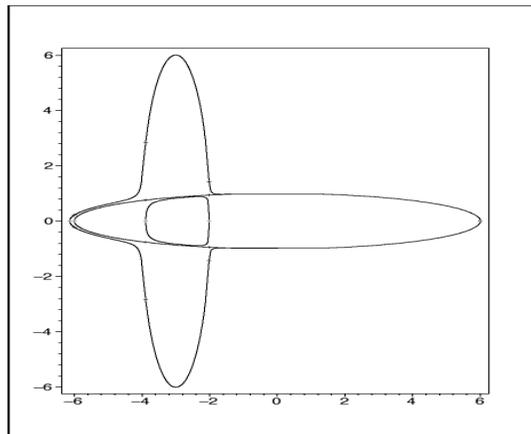
Para cualquier $d > 1/2$ el número de raíces reales de

$$64 x^3 + (48 - 192 d^2) x^2 + (192 d^4 + 336 d^2 + 12) x - 64 d^6 - 12 d^2 + 48 d^4 + 1$$

con $x > 0$ es siempre igual a 1:

$$[1, (2 d - 1) (2 d + 1), -d^2 (20 d^2 + 1) (4 d^2 + 5), -d^4 (4 d^2 + 1)^2 (2 d - 1)^3 (2 d + 1)^3]$$

Cálculo de la topología de una curva definida implícitamente

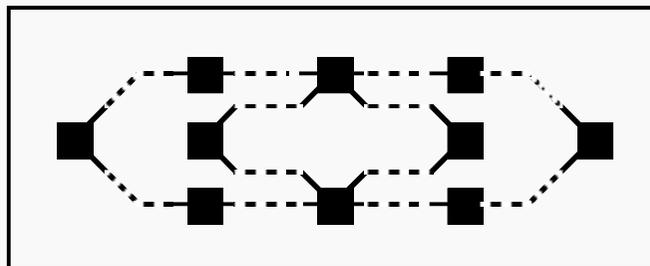
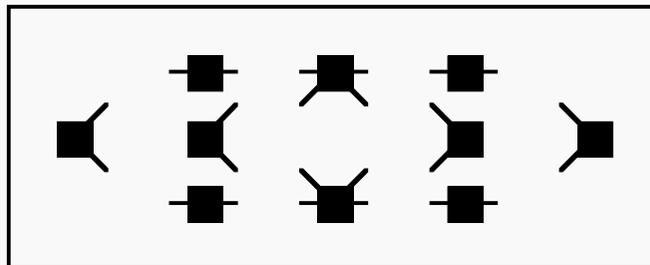
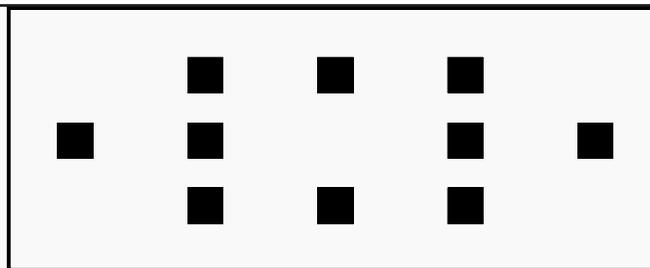


$$g(x,y) = 279756x + 279936xy^4 - 559692y^2x - 15583x^3 + 217x^5 + \frac{130218}{5}x^2 - \frac{23039}{10}x^4 + \frac{359}{10}x^6 + 370656y^4 - \frac{3726432}{5}y^2 - 72774y^2x^2 + \frac{12947}{5}y^2x^4 + 1296y^6 + 46728y^4x^2 + 15588y^2x^3 + \frac{37333439}{100}$$

La Estrategia

$$P(x, y) := 2x^4 - 3x^2y + y^4 - 2y^3 + y^2$$

$$R(x) = x(2048x^6 - 4608x^4 + 37x^2 + 12)$$



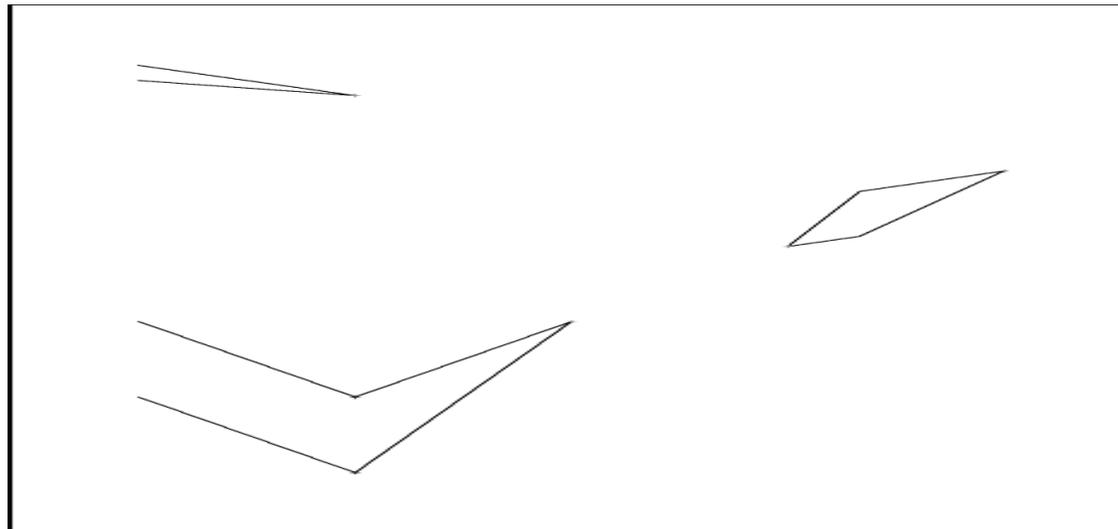
	Time	d	r	GP	Precision
Pol ₁	2.481	8	28	y	20
Pol ₂	0.130	5	4	y	15
Pol ₃	0.269	5	5	y	15
Pol ₄	0.171	4	4	n	15
Pol ₅	0.061	4	3	n	10
Pol ₆	1.170	8	8	n	20
Pol ₇	0.120	4	2	n	10
Pol ₈	0.360	6	6	y	15
Pol ₉	0.869	8	13	n	15
Pol ₁₀	0.351	6	5	n	10
Pol ₁₁	2.409	6	4	y	15
Pol ₁₂	116.959	8	4	y	40
Pol ₁₃	0.110	8	3	y	10
Pol ₁₄	3.569	6	8	n	30
Pol ₁₅	0.180	4	6	n	10
Pol ₁₆	0.230	6	7	n	10

Table 1: Experimental results

$$\text{Pol}_{12} = y^8 + 25y^4x^4 + 61y^6 + 62x^5y^4 + x^6 - 5x^2y^5 - 11xy + \frac{1}{10000000}y$$

Digits	α_2 and α_3
10	$0.9090909091 \cdot 10^{-8}$ $0.9090909091 \cdot 10^{-8}$
20	$0.909090909090909091 \cdot 10^{-8}$ $0.909090909090909091 \cdot 10^{-8}$
30	$0.90909090909090909090909091 \cdot 10^{-8}$ $0.909090909090909090909091 \cdot 10^{-8}$
40	$0.9090909090909090909090909073334909 \cdot 10^{-8}$ $0.9090909090909090909090909108483272 \cdot 10^{-8}$

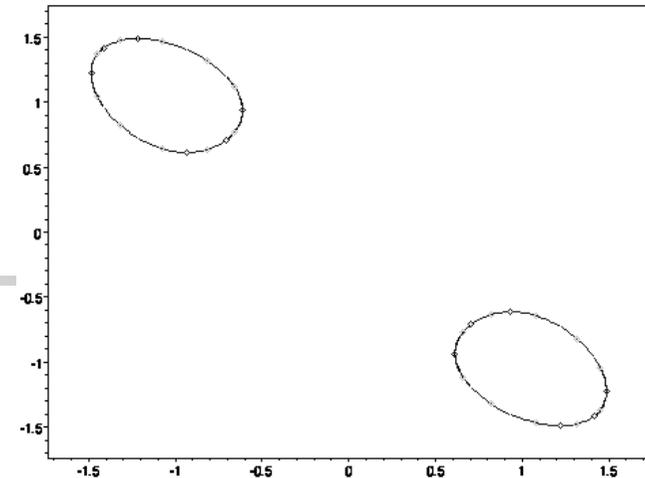
Table 2: Two real roots of the discriminant very close



Una aplicación:
Secciones planas de
superficies implícitas

$$(x^2 + y^2 + z^2 + 2)^2 - 9x^2 - 9y^2 = 0$$

$$x + y + z = 0$$



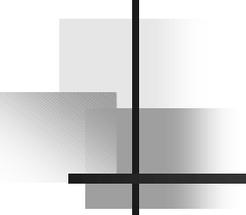
El análisis algebraico/topológico concluye que:

- ❖ La curva sección no tiene puntos singulares
- ❖ La curva sección tiene dos componentes cerradas empezando en:

$$[.6126044921, -.9332990412, .3206945491]$$

$$[-1.486514595, 1.218778958, .267735637]$$

$$\frac{\partial}{\partial s} x(s) = \frac{\frac{\partial}{\partial y} f}{\sqrt{\left(\frac{\partial}{\partial x} f\right)^2 + \left(\frac{\partial}{\partial y} f\right)^2}}, \quad \frac{\partial}{\partial s} y(s) = -\frac{\frac{\partial}{\partial x} f}{\sqrt{\left(\frac{\partial}{\partial x} f\right)^2 + \left(\frac{\partial}{\partial y} f\right)^2}}$$



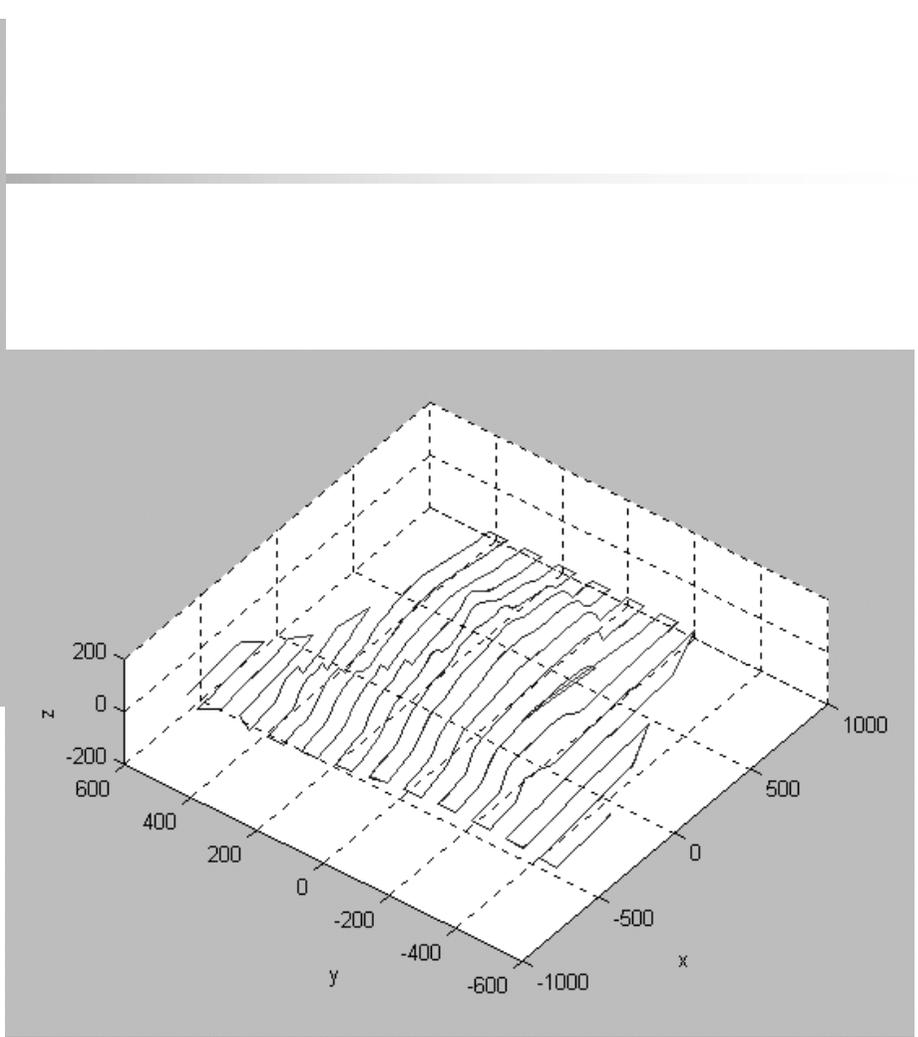
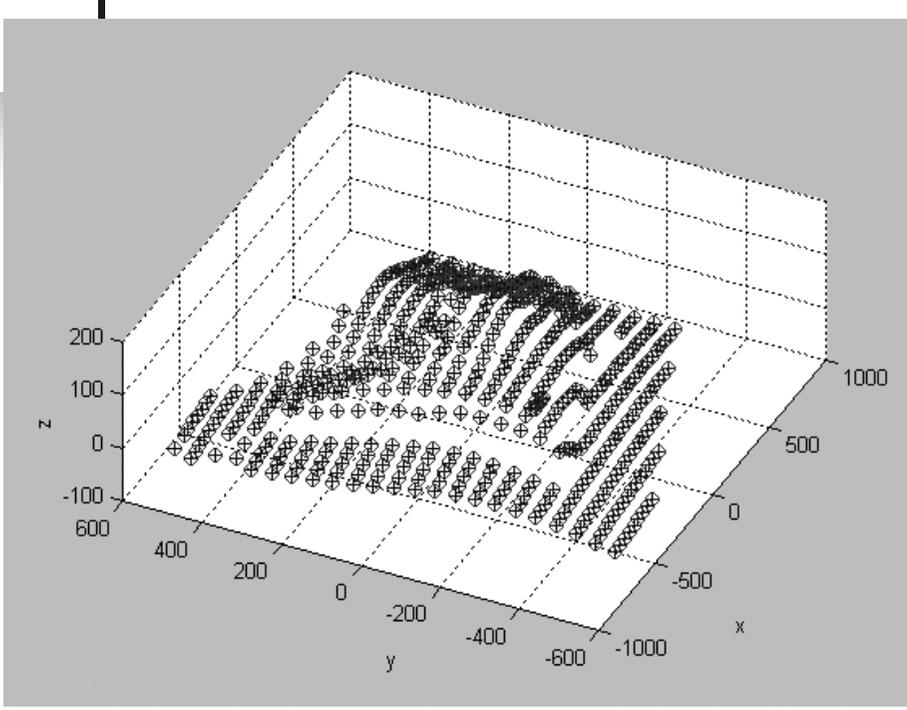
Cálculo del Error de Forma

El problema algebraico

Dados dos conjuntos de n puntos X e Y en \mathbb{R}^3 , determinar el movimiento F tal que $F(X)$ e Y están lo más cerca posible.

X : Puntos teóricos (desde la definición teórica)
 Y : Puntos reales (desde medidas reales)





$$X = \{P_1 = (x_1, y_1, z_1), \dots, P_n = (x_n, y_n, z_n)\}, \quad Y = \{Q_1 = (a_1, b_1, c_1), \dots, Q_n = (a_n, b_n, c_n)\}$$

$$\mathbf{X} = \sum_{i=1}^n x_i \quad \mathbf{Y} = \sum_{i=1}^n y_i \quad \mathbf{Z} = \sum_{i=1}^n z_i$$

$$\mathbf{A} = \sum_{i=1}^n a_i \quad \mathbf{B} = \sum_{i=1}^n b_i \quad \mathbf{C} = \sum_{i=1}^n c_i$$

- $\alpha =$

$$\frac{1}{n}(\mathbf{A} - \mathbf{Z} \cos(\theta) \sin(\eta) \cos(\xi) - \mathbf{Y} \cos(\theta) \sin(\eta) \sin(\xi) - \mathbf{Z} \sin(\theta) \sin(\xi) - \mathbf{X} \cos(\theta) \cos(\eta) + \mathbf{Y} \sin(\theta) \cos(\xi))$$

- $\beta =$

$$\frac{1}{n}(\mathbf{B} - \mathbf{Z} \sin(\theta) \sin(\eta) \cos(\xi) - \mathbf{Y} \sin(\theta) \sin(\eta) \sin(\xi) + \mathbf{Z} \cos(\theta) \sin(\xi) - \mathbf{Y} \cos(\theta) \cos(\xi) - \mathbf{X} \sin(\theta) \cos(\eta))$$

- $\gamma =$

$$\frac{1}{n}(\mathbf{C} - \mathbf{Z} \cos(\eta) \cos(\xi) + \mathbf{X} \sin(\eta) - \mathbf{Y} \cos(\eta) \sin(\xi))$$

$$\begin{aligned}
& (n\mathbf{YA} - \mathbf{Y} \cdot \mathbf{A}) \sin(\theta) \sin(\eta) \sin(\xi) + (n\mathbf{ZA} - \mathbf{Z} \cdot \mathbf{A}) \sin(\theta) \sin(\eta) \cos(\xi) + \\
& + (\mathbf{Y} \cdot \mathbf{B} - n\mathbf{YB}) \cos(\theta) \sin(\eta) \sin(\xi) + (\mathbf{Z} \cdot \mathbf{B} - n\mathbf{ZB}) \cos(\theta) \sin(\eta) \cos(\xi) + \\
& + (n\mathbf{XA} - \mathbf{X} \cdot \mathbf{A}) \sin(\theta) \cos(\eta) + (\mathbf{Z} \cdot \mathbf{B} - n\mathbf{ZB}) \sin(\theta) \sin(\xi) + (n\mathbf{YB} - \mathbf{Y} \cdot \mathbf{B}) \sin(\theta) \cos(\xi) + \\
& + (\mathbf{X} \cdot \mathbf{B} - n\mathbf{XB}) \cos(\theta) \cos(\eta) + (\mathbf{Z} \cdot \mathbf{A} - n\mathbf{ZA}) \cos(\theta) \sin(\xi) + (n\mathbf{YA} - \mathbf{Y} \cdot \mathbf{A}) \cos(\theta) \cos(\xi)
\end{aligned}$$

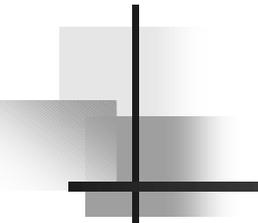
$$\begin{aligned}
& (\mathbf{Y} \cdot \mathbf{B} - n\mathbf{YB}) \sin(\theta) \cos(\eta) \sin(\xi) + (\mathbf{Z} \cdot \mathbf{B} - n\mathbf{ZB}) \sin(\theta) \cos(\eta) \cos(\xi) + \\
& + (\mathbf{Y} \cdot \mathbf{A} - n\mathbf{YA}) \cos(\theta) \cos(\eta) \sin(\xi) + (\mathbf{Z} \cdot \mathbf{A} - n\mathbf{ZA}) \cos(\theta) \cos(\eta) \cos(\xi) + \\
& + (n\mathbf{XB} - \mathbf{X} \cdot \mathbf{B}) \sin(\theta) \sin(\eta) + (n\mathbf{XA} - \mathbf{X} \cdot \mathbf{A}) \cos(\theta) \sin(\eta) + (n\mathbf{YC} - \mathbf{Y} \cdot \mathbf{C}) \sin(\eta) \sin(\xi) + \\
& + (n\mathbf{ZC} - \mathbf{Z} \cdot \mathbf{C}) \sin(\eta) \cos(\xi) + (n\mathbf{XC} - \mathbf{X} \cdot \mathbf{C}) \cos(\eta)
\end{aligned}$$

$$\begin{aligned}
& (n\mathbf{ZB} - \mathbf{Z} \cdot \mathbf{B}) \sin(\theta) \sin(\eta) \sin(\xi) + (\mathbf{Y} \cdot \mathbf{B} - n\mathbf{YB}) \sin(\theta) \sin(\eta) \cos(\xi) + \\
& + (n\mathbf{ZA} - \mathbf{Z} \cdot \mathbf{A}) \cos(\theta) \sin(\eta) \sin(\xi) + (\mathbf{Y} \cdot \mathbf{A} - n\mathbf{YA}) \cos(\theta) \sin(\eta) \cos(\xi) + \\
& + (\mathbf{Y} \cdot \mathbf{A} - n\mathbf{YA}) \sin(\theta) \sin(\xi) + (\mathbf{Z} \cdot \mathbf{A} - n\mathbf{ZA}) \sin(\theta) \cos(\xi) + (n\mathbf{YB} - \mathbf{Y} \cdot \mathbf{B}) \cos(\theta) \sin(\xi) + \\
& + (n\mathbf{ZB} - \mathbf{Z} \cdot \mathbf{B}) \cos(\theta) \cos(\xi) + (n\mathbf{ZC} - \mathbf{Z} \cdot \mathbf{C}) \cos(\eta) \sin(\xi) + (n\mathbf{Y} \cdot \mathbf{C} - \mathbf{YC}) \cos(\eta) \cos(\xi)
\end{aligned}$$

where

$$\mathbf{XA} = \sum_{i=1}^n x_i a_i$$

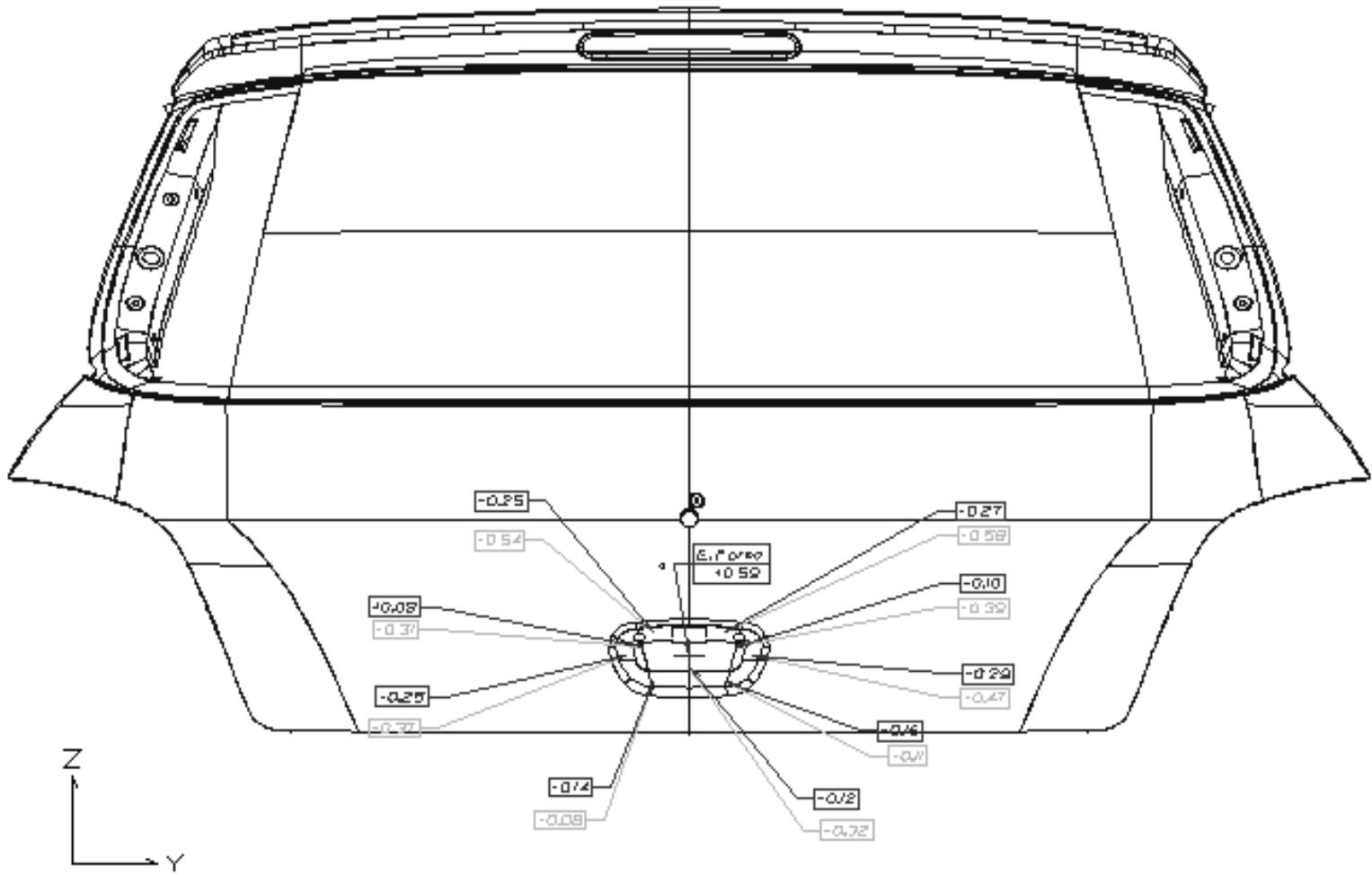
etc.

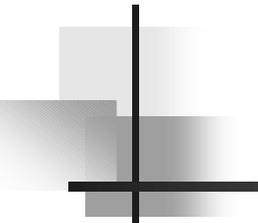


$$\tan(\xi) =$$

$$\begin{aligned}
 & - ((-Y C + Y C n) \cos(\eta) + (Y A n - Y A) \sin(\eta) \cos(\theta) + (-Y B + Y B n) \sin(\eta) \sin(\theta) \\
 & + (-Z A + Z A n) \sin(\theta) + (Z B - Z B n) \cos(\theta)) / ((Z C - Z C n) \cos(\eta) + (-Z A n + Z A) \sin(\eta) \cos(\theta) \\
 & + (Z B - Z B n) \sin(\eta) \sin(\theta) + (Y A n - Y A) \sin(\theta) + (-Y B n + Y B) \cos(\theta))
 \end{aligned}$$

Las otras dos ecuaciones (en dos incógnitas) se han resuelto simbólicamente: proporciona una ecuación en una variable ...





Triangulación de superficies 'trimmed'

e

Iluminación/sombreado

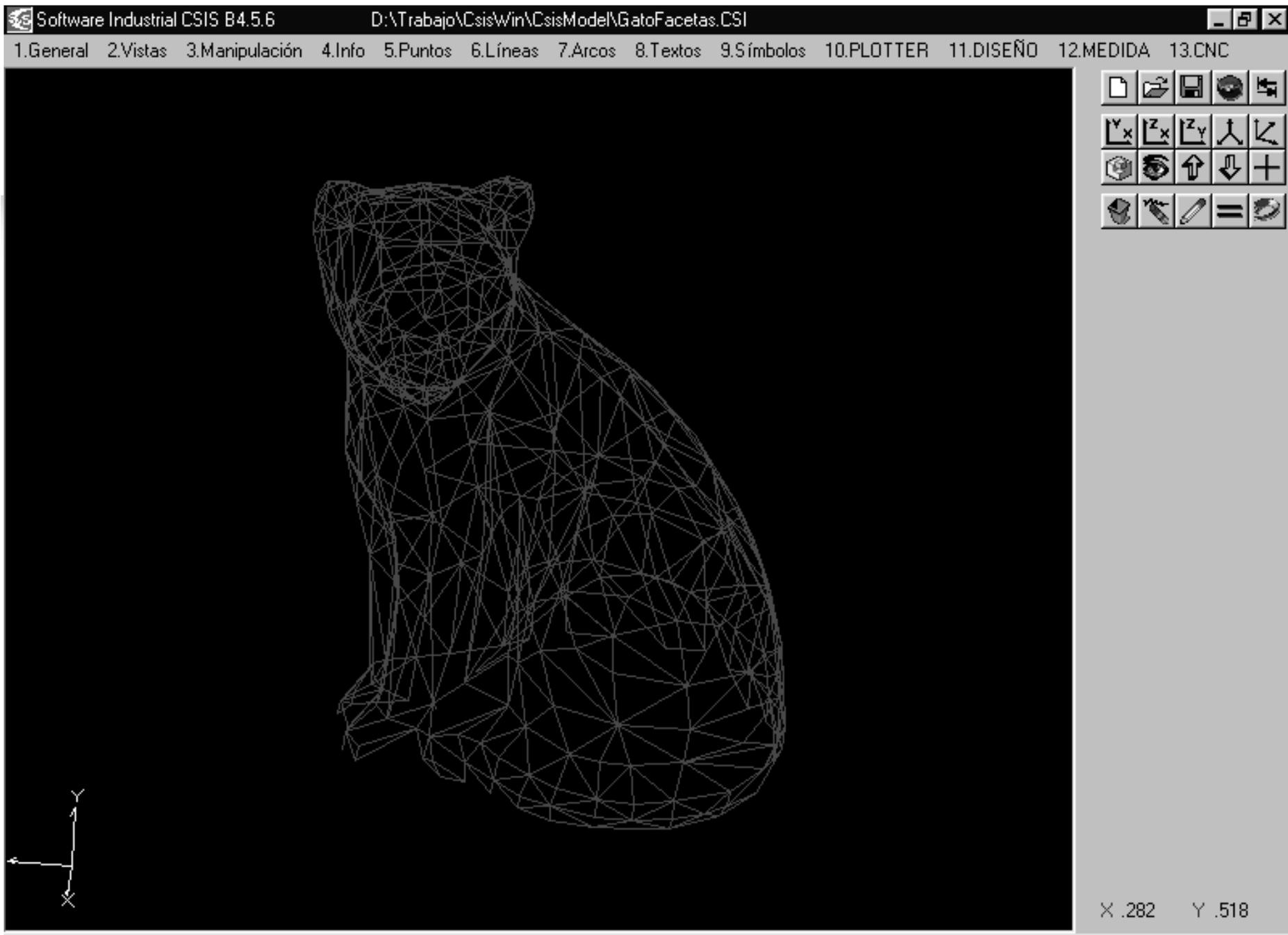
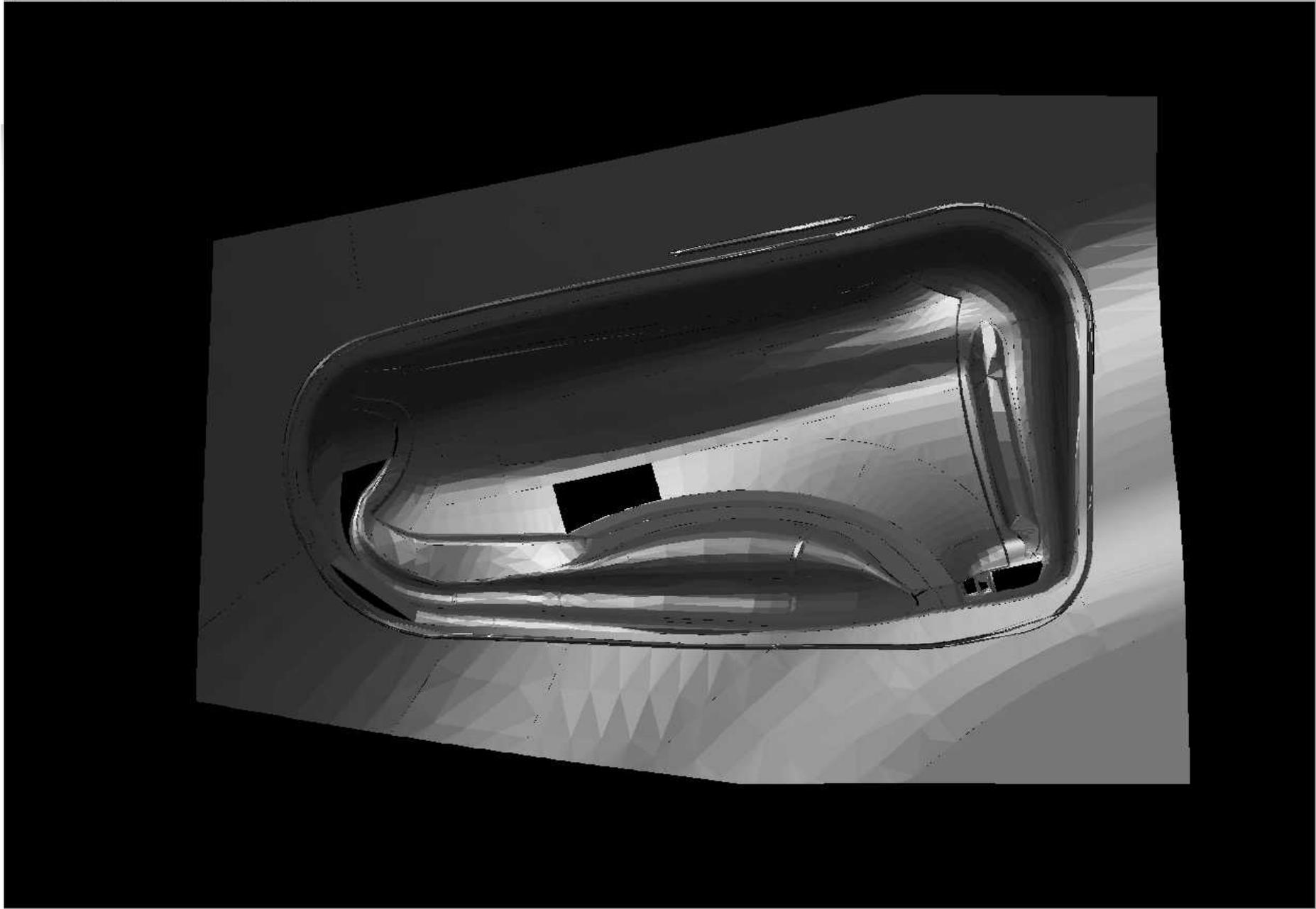


Figure No. 1

File Edit View Insert Tools Window Help



A
Numeric/Symbolic
Problem Solving Environment
for CAGD

Maple/Mathematica/Mupad
Matlab
CSIS/Think3

Some capabilities

GENERATION OF POLYNOMIAL AND RATIONAL B-SPLINE CURVES AND SURFACES:

- Reading IGES/VDA format
- Plotting IGES/VDA format

POLYNOMIAL AND RATIONAL B-SPLINE CURVE/SURFACE MANIPULATION

- Numerical generation of polynomial and rational B-spline curves and surfaces.
- Computation of the B-spline curve/surfaces derivatives, using:
 - basis function method
 - control point method.
- Knot vector refinement.
- Bezier decomposition.
- Approximate implicitization
- Approximation of rational B-spline curves/surfaces with polynomial ones, degree raising, ...

A CAGD SYSTEM WITH EXACT ARITHMETIC

- Generic implicitization
- Symbolic/Numerical Polynomial System Solving

Example 6 We consider the B-spline curves (polynomial and rational) defined by the following parameters:

```
> read "simbC.txt";
```

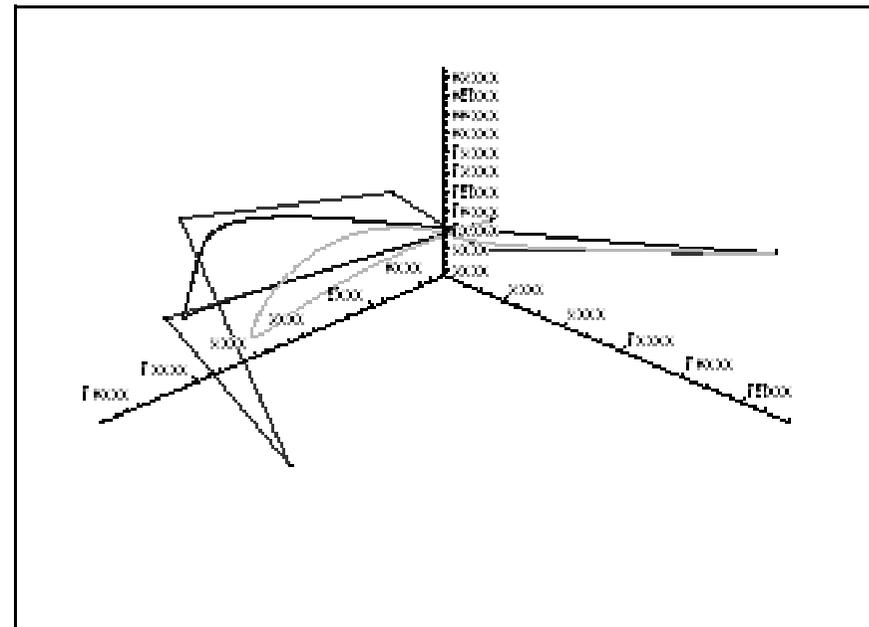
```
Curve order: 6
```

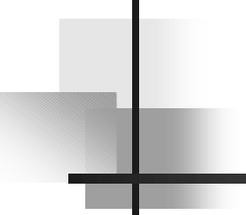
```
Knot vector: [0,0,0,0,0,0,.1210225739,1.,1.,1.,1.,1.,1.]
```

```
Control points: [[1813.213542, 3155.072848, 7292.017654,  
10852.09613, 11970.06777, 13138.01982, 5307.115079],  
[15411.10969, 6888.316017, 7878.233333, 4268.225166,  
9045.040000, 6034.080537, 9315.020747], [22834.01291,  
13631.11796, 26325.11671, 23448.19923, 5654.282051,  
18408.02733, 22853.31278]]
```

```
Weights: [.1883026555, .1081886648e-1, .5431078818e-1,  
.5421364074, .1062056055, .6872239468, .2581224202e-1]
```

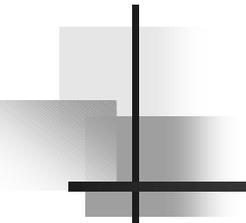
DEMO





PROBLEMAS

- ❖ Detección de autointersecciones para curvas y superficies
- ❖ A-splines
- ❖ Cálculo de la topología de superficies definidas implícitamente
- ❖ Cálculo de familias (definidas de forma paramétrica) de curvas y superficies con topología prefijada de antemano.



EL PROYECTO GAIA II



IST-2001-35512 - GAIA II

Intersection algorithms for geometry based IT-applications using approximate algebraic methods

- **Establish communication between the algebraic geometry and computer aided geometric design communities**
- **Improve the speed, robustness and accuracy of intersection algorithms in CAD by integrating approximate algebraic methods and traditional approaches**
- **Understand better possibilities and limitations of approximate implicitization**
- **Improve the usefulness of algebraic geometry in CAD and other application domains by using knowledge from**
 - **Real and classical algebraic geometry**
 - **Computer Aided Design (CAD)**
 - **Approximation theory**





IST-2001-35512 - GAIA II

Intersection algorithms for geometry based IT-applications using approximate algebraic methods

■ **Participants**

- SINTEF Applied Mathematics, Norway (Coordinator)S
- Johannes Kepler University, Linz, Austria INTEF
- University of Nice Sophia Antipolis and INRIA, France
- University of Cantabria, Santander, Spain
- University of Oslo, Norway
- think3, France & Italy

■ **Start date July 1st 2002, Duration 36 months**

- **Contact details:** SINTEF Applied Mathematics e-mail: tor.dokken@sintef.no
Attention: tor Dokken Phone: Telephone: +47 22 06 76 61
P.O. Box 124 Fax: +47 22 06 73 50
0314 Oslo, Norway <http://www.math.sintef.no/gaiatwo/>



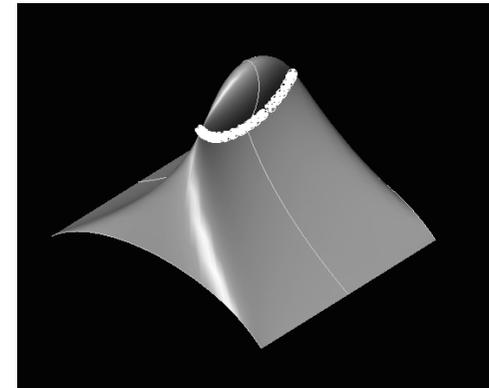
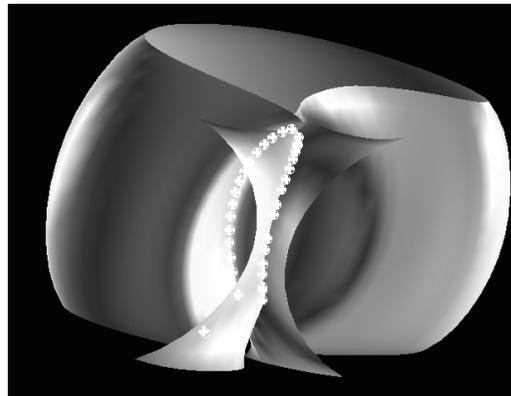
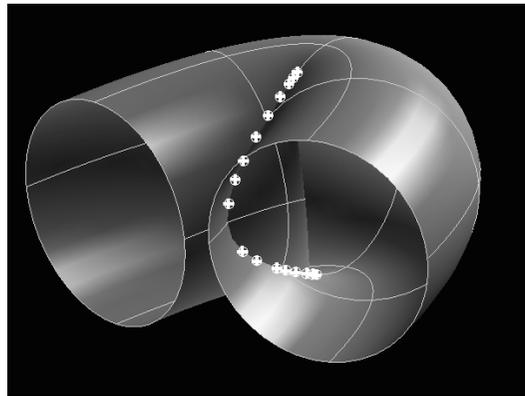
Project funded by the Future and Emerging Technologies arm of the IST Programme
FET-Open scheme

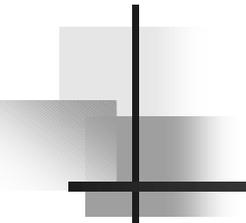


IST-2001-35512 - GAIA II

Intersection algorithms for geometry based IT-applications using approximate algebraic methods

- **Examples of self-intersections from the assessment project preceding GAIA II**

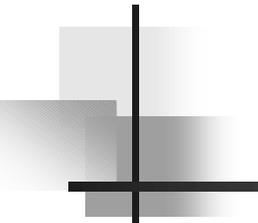




Diseño Geométrico Asistido por Ordenador

*Álgebra Computacional
Cálculo Numérico
Desarrollo de Software
Ecuaciones Diferenciales
Geometría Algebraica
Geometría Algebraica Real
Geometría Diferencial*

.....



El Equipo

D. Bochis	F. Etayo
I. Necula	C. Gómez de Dios
M. Romano	L. González Vega
J. Espínola	E. Mainar
C. Tanasescu	J. Puig-Pey
F. Carreras	T. Recio
S. Corvez	R. Sendra

E_mail: gvega@matesco.unican.es

<http://frisco.matesco.unican.es/~gvega>

Modelos matemáticos en la industria: Ejemplos y perspectivas



Alfredo Bermúdez de Castro

Catedrático y Director del Departamento de Matemática Aplicada de la Universidad de Santiago de Compostela

Resumen

La charla va a tener dos partes diferentes. En la primera se hará una presentación general del modelado de problemas industriales, señalando los aspectos fundamentales de su metodología y dando una panorámica de sus aplicaciones. También se explicitarán las áreas de la matemática más relacionadas con el tema y los conocimientos que un “modelador” debería tener. Estos aspectos llevarán a discutir algunas cuestiones relacionadas con los planes de estudio actuales y las perspectivas en el marco de la adaptación al espacio europeo.

En la segunda parte se estudiarán ejemplos tomados de las industrias metalúrgica y de generación de energía. También se mostrarán casos de aplicación del modelado matemático a problemas medioambientales.

Más concretamente, se presentarán, en primer lugar, trabajos realizados o en curso para la empresa Ferroatlántica I+D, que desarrolla y comercializa tecnología para la producción del silicio. Se introducirán modelos para la simulación del comportamiento de hornos de arco en sus diferentes aspectos: electromagnético, térmico y mecánico. También se tratarán otras aplicaciones relacionadas con procesos de purificación del silicio con vistas a su utilización en células solares. Este ejemplo de colaboración permitirá abordar algunos aspectos generales de las relaciones entre centros de investigación y empresas y otros más específicos en el ámbito de la matemática aplicada.

Otro ejemplo de aplicación se refiere a la simulación y control del ruido en recintos. En particular, las técnicas de “antirruído” encuentran en la teoría matemática del control óptimo un marco adecuado para su análisis. Por otra parte, en muchas ocasiones, la propagación del sonido involucra la interacción mecánica entre un fluido (generalmente el aire) y uno o varios sólidos que transmiten o generan las vibraciones acústicas. Se mostrarán modelos y métodos numéricos para simular estos fenómenos.

Finalmente, se tratarán problemas relacionados con la combustión en calderas de centrales eléctricas. El nuevo marco de competitividad en el que se mueve el mercado de la energía, por una parte, y las normas cada vez más exigentes que limitan las emisiones contaminantes, por otra, obligan a las compañías a optimizar el funcionamiento de sus centrales y, de modo especial, de las calderas en las que tiene lugar la combustión del carbón o el fuel. En este contexto, la simulación numérica es una herramienta fundamental para ayudar a los ingenieros en el diseño o en el establecimiento de condiciones de operación adecuadas.

Bibliografía

- A. Bermúdez de Castro: Simulación numérica de diferentes procesos industriales relacionados con la producción de silicio. Proyectos de colaboración con Ferroatlántica I+D. *Boletín SEMA* (por aparecer).

Simulación numérica de diferentes procesos industriales relacionados con la producción de silicio. Proyectos de colaboración con Ferroatlántica I+D.

Alfredo Bermúdez
Departamento de Matemática Aplicada
Universidad de de Santiago de Compostela

Resumen

El objetivo de este trabajo es la descripción de las líneas de investigación desarrolladas en el marco de colaboración que mantiene la empresa Ferroatlántica I+D con el Departamento de Matemática Aplicada de la Universidad de Santiago de Compostela. En el artículo, se aborda la simulación numérica de diferentes problemas físicos relacionados con la producción del silicio; para ello se describen los modelos matemáticos desarrollados, los métodos numéricos utilizados y los principales resultados obtenidos.

Palabras clave: Simulación numérica, hornos eléctricos, colada de silicio, purificación de silicio, electromagnetismo, transferencia de calor, elementos finitos.

1 Introducción

El Grupo Ferroatlántica está integrado por diferentes empresas dedicadas a la producción de ferroaleaciones. Su actividad le convierte en el primer grupo español del sector y el segundo de la Unión Europea. Además, es el primer productor español independiente de energía eléctrica.

Las actividades de investigación y desarrollo del Grupo Ferroatlántica son gestionadas por la empresa Ferroatlántica I+D, que juega un papel fundamental en el desarrollo y comercialización de nuevas tecnologías dentro del grupo. En los últimos años, Ferroatlántica I+D ha mantenido una estrecha colaboración con la Universidad y otros Centros Públicos de Investigación principalmente en Galicia. En concreto, Ferroatlántica I+D ha iniciado proyectos de investigación con el Departamento de Matemática Aplicada de la Universidad de Santiago de Compostela en el año 1996. En este artículo expondremos los principales temas

de investigación abordados y los resultados obtenidos. La actividad investigadora ha estado orientada al desarrollo de modelos matemáticos y programas de ordenador que permiten simular numéricamente diferentes procesos industriales de interés para la empresa. Así, destacan los siguientes campos:

1. Modelado de electrodos metalúrgicos.
2. Modelado de nuevos sistemas de colada para ferroaleaciones.
3. Modelado de sistemas innovadores para la purificación del silicio.

El equipo de investigación que ha participado en el desarrollo y resolución numérica de los diferentes modelos está integrado por: A. Bermúdez, M. C. Muñiz, R. Leira, F. Pena y P. Salgado por parte del Departamento de Matemática Aplicada y por J. Bullón, M. Lage, A. Lorenzo por parte de Ferroatlántica I+D.

El desarrollo del artículo es el siguiente: la Sección 2 se ocupa de la descripción del proceso de producción de silicio y sus aplicaciones. En la Sección 3 se describe la problemática de los electrodos, piezas clave en un horno de producción de silicio, y se describen los modelos matemáticos desarrollados en este tema. En la Sección 4 se aborda el modelado matemático de un innovador sistema de colada sobre placa vibrante de cobre refrigerada por agua, para la solidificación del silicio. La Sección 5 se ocupa del modelado de sistemas relacionados con la purificación del silicio. Finalmente, en la Sección 6 se presenta un avance de la investigación en curso sobre la simulación numérica de un horno de inducción, también destinado a la purificación de silicio.

2 El Silicio y sus aplicaciones

El silicio (Si) es el segundo elemento más abundante en la corteza terrestre después del oxígeno. Se obtiene industrialmente por reducción del dióxido de silicio, en forma de cuarzo o cuarcita, con carbón, mediante una reacción química que puede escribirse en forma simplificada (ver [16]):



Este proceso tiene lugar en un horno eléctrico de reducción, más concretamente, en un “horno de arco” que se describe en la siguiente sección. El silicio tiene una gran variedad de aplicaciones dependiendo de la cantidad de impurezas presentes en el producto. Así, se distingue:

- **Silicio electrónico:** es el más puro de los empleados industrialmente, denominado 9N (“9 nueves” \equiv 99,999999% de pureza). Se utiliza para la fabricación de los semiconductores, en la que se requiere una precisión en la conducción eléctrica muy alta.
- **Silicio solar:** precisa de menor pureza que el anterior, aunque en la actualidad comparte con el silicio electrónico varias fases del proceso de

producción. Sirve para fabricar las células fotovoltaicas de los paneles solares.

- **Silicio metal:** es el silicio metalúrgico que contiene un 1-2% de otros elementos. Se usa principalmente en aleaciones con otros metales no féreos, como el aluminio.
- **Silicio químico:** es el que se utiliza para la fabricación de siliconas, productos de gran consumo por su variedad de aplicaciones.
- **Ferrosilicio:** es el que contiene más de un 2% de otros elementos, especialmente hierro (de ahí su nombre). Tiene menor precio que el silicio metal y se usa para producir aceros al silicio.

3 Modelado de electrodos metalúrgicos

Un horno eléctrico para la producción de silicio está compuesto, en general, por una cuba cilíndrica que contiene materiales carbonosos y tres electrodos cuyos ejes forman un triángulo equilátero centrado en la cuba (ver Figura 1). Los electrodos son las piezas clave del horno y su propósito es la conducción de la corriente eléctrica, normalmente corriente alterna y trifásica. En la punta de cada electrodo se genera un arco eléctrico, que crea las altas temperaturas necesarias para que tengan lugar las diferentes reacciones químicas del proceso de reducción. El buen funcionamiento del horno depende, en gran medida, de conseguir condiciones de operación adecuadas en los electrodos.

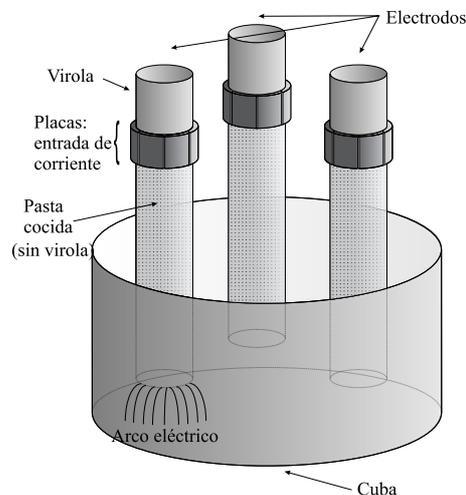


Figura 1: Cuba de un horno de reducción.

Los electrodos clásicos utilizados mayoritariamente en la industria son los electrodos de grafito puro, los precocidos y los Söderberg. Estos últimos, están

compuestos de pasta carbonosa que se va cociendo durante el proceso y son los más empleados en la industria del ferrosilicio. Sus ventajas son que se construyen de tamaños mayores y que son más baratos que los de grafito puro o los precocidos. Sin embargo, los electrodos Söderberg no se pueden utilizar para la producción de silicio metal, ya que durante el proceso contaminan el producto. Así, hasta la pasada década, los electrodos precocidos eran la única alternativa para la producción de silicio metal.

A principios de los años 90, Ferroatlántica I+D desarrolló un nuevo tipo de electrodo, el llamado electrodo ELSA, que es apto para la producción de silicio metal. Si bien los electrodos clásicos están formados por un único material, el electrodo ELSA es un electrodo compuesto que está formado por un núcleo de grafito rodeado de pasta (ver Figura 2). El electrodo ELSA es mucho más barato que el electrodo precocido, por lo que se ha convertido en una alternativa muy interesante para la producción de silicio metal al reducir en más del 10% los costes de producción.

En general, las condiciones de funcionamiento de cualquier tipo de electrodo son complejas, por lo que la simulación numérica constituye una herramienta muy importante para estudiar su comportamiento. Modelar el problema en un ordenador permite estudiar la influencia de los diferentes parámetros que intervienen en el funcionamiento del electrodo, y por lo tanto del horno, sin necesidad de experimentos delicados y costosos. Así, el comportamiento de los electrodos clásicos ha sido estudiado durante décadas (ver por ejemplo [12] y [10]). La mayoría de estos trabajos asumen simetría cilíndrica y utilizan métodos de diferencias y elementos finitos para la resolución de los modelos. Sin embargo, la estructura compuesta del electrodo ELSA hace que su comportamiento termoelectrico sea diferente del de los electrodos clásicos. En particular, el electrodo ELSA combina el grafito, que es muy buen conductor de la electricidad, con la pasta, que solo es buena conductora a temperaturas altas. Por lo tanto, no sólo el grafito es importante en el movimiento de la columna, sino también en la distribución de la corriente. Así, el llamado *efecto piel* es comparable al de electrodos clásicos sólo en la punta. Por otra parte, la pasta cambia de estado durante el proceso, siendo su cocción un fenómeno fundamental; sin embargo, este proceso de cocción también presenta notables diferencias con respecto al que tiene lugar en los electrodos Söderberg (ver [15] para más detalles). Teniendo en cuenta estas diferencias, la simulación numérica del electrodo ELSA ha constituido el objetivo fundamental de muchos de los proyectos y contratos mantenidos con Ferroatlántica I+D. Cabe señalar sin embargo, que los modelos desarrollados son lo suficientemente generales para simular el comportamiento de cualquier tipo de electrodo.

La naturaleza de los modelos desarrollados para estudiar el comportamiento del electrodo ELSA es muy diversa y ha dado lugar al estudio numérico y matemático de un amplio número de problemas. Los resultados más relevantes se recogen en diferentes publicaciones ([1, 2, 7, 5, 6]). Por ello, en esta sección daremos una descripción bastante general de los principales problemas abordados y de los resultados obtenidos.

La simulación del electrodo ELSA ha tenido dos objetivos fundamentales:

conocer y controlar los diferentes parámetros que intervienen en la cocción de la pasta y conocer los factores que pueden provocar roturas en el electrodo. Así, los modelos matemáticos desarrollados tienen como objeto el cálculo de la distribución de la temperatura, de la densidad de corriente y de las tensiones en el electrodo, bajo diferentes condiciones de operación.

El primer paso ha sido el desarrollo de modelos matemáticos basados en la simetría cilíndrica; es decir, despreciando el *efecto proximidad* de los otros dos electrodos y considerando condiciones de contorno axisimétricas. En estas condiciones se puede suponer que los diferentes campos no dependen de la variable angular. Estas hipótesis permiten resolver el problema en un dominio bidimensional, en concreto, en una sección radial del electrodo (ver Figura 2). Así, se han desarrollado y resuelto los siguientes modelos:

- Un modelo **termoeléctrico estacionario**, que permite conocer la distribución de la temperatura y la densidad de corriente en una sección radial del electrodo en estado estacionario. El modelo electromagnético se obtiene a partir de las ecuaciones de Maxwell en régimen armónico de baja frecuencia y el modelo térmico a partir de la ecuación de transferencia de calor en estado estacionario. Se trata de un problema acoplado, ya que la fuente de calor en el problema térmico es el efecto Joule y, por otra parte, los parámetros termoeléctricos dependen de la temperatura.
- Un modelo **termoeléctrico transitorio**, que resuelve las ecuaciones de Maxwell acopladas con la ecuación de transferencia de calor en estado transitorio. Así, puede obtenerse la evolución de la temperatura con el tiempo y tener en cuenta aspectos de la operación que dependen del tiempo como son los descensos del electrodo, las desconexiones de la red eléctrica, etc.
- Un modelo **termomecánico**, que permite conocer la distribución de esfuerzos en el electrodo debido a su peso y a los gradientes térmicos. Los resultados obtenidos han permitido, por ejemplo, mejorar el diseño de los *nipples* que son las piezas que unen las columnas de grafito.

Para resolver numéricamente los diferentes modelos se han escrito programas en Fortran 77, que han sido utilizados para simular el comportamiento de electrodos reales. En las referencias [3], [7] y [15] se presentan algunos de los resultados obtenidos por los diferentes modelos. En la Figura 3 se presenta, por ejemplo, la evolución de la temperatura con respecto al tiempo en un punto situado en el eje del electrodo a la altura de placas. Nótese que en la simulación se tienen en cuenta los deslizamientos y desconexiones del electrodo de la red eléctrica varias horas al día. Además, se ha desarrollado un paquete informático llamado ELSATE, que incluye un menú desplegable que permite al usuario la introducción de datos a través de cuadros de diálogo muy simples. El paquete permite la visualización de una amplia gama de resultados. En la Figura 4 se presenta un ejemplo de los cuadros de diálogo.

Los modelos bidimensionales descritos previamente, han proporcionado información importante sobre algunos de los parámetros que intervienen en el

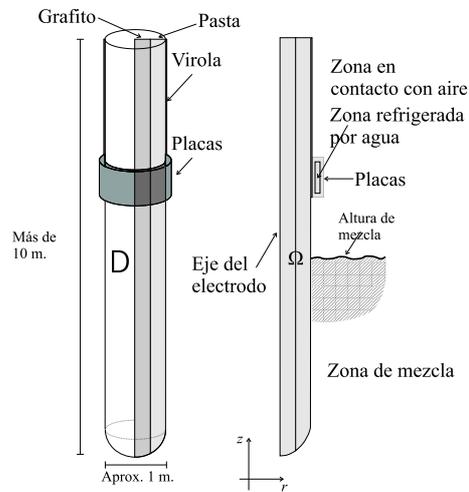


Figura 2: Esquema del electrodo ELSA y dominio bidimensional de los modelos axisimétricos.

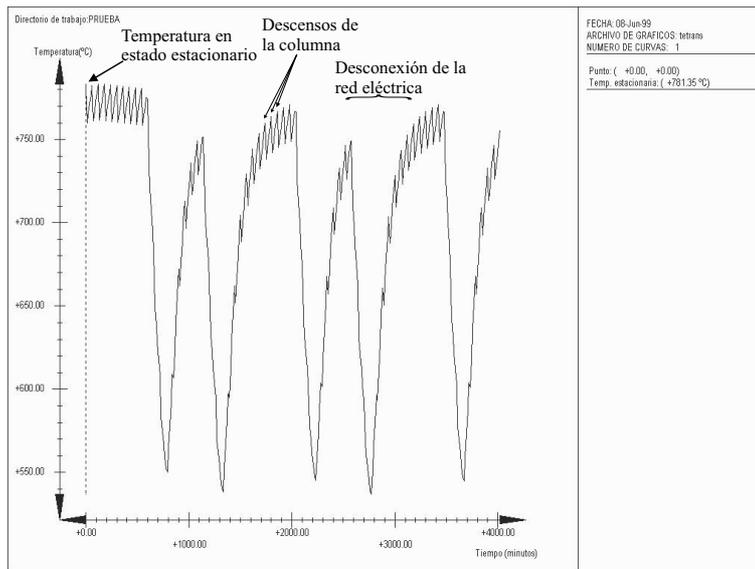


Figura 3: Evolución de la temperatura.

funcionamiento del electrodo ELSA y tienen la ventaja de un importante ahorro computacional con respecto a un modelo puramente tridimensional. Sin embargo, como se ha indicado anteriormente, la hipótesis de simetría cilíndrica obliga a despreciar ciertos hechos como el efecto de los otros dos electrodos (el llamado *efecto de proximidad*) o la no simetría de las condiciones de contorno. Estos

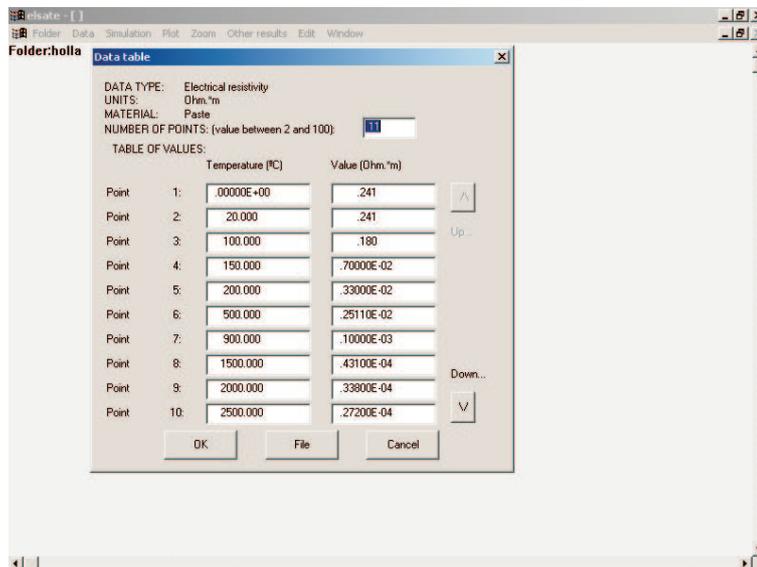


Figura 4: Un ejemplo de menú del paquete ELSATE.

hechos solo pueden ser considerados por un modelo puramente tridimensional. Además, los modelos tridimensionales son necesarios para simular los electrodos Soderberg que por su propia geometría no poseen simetría cilíndrica. Por ello, con el objetivo de tener en cuenta los efectos tridimensionales se ha desarrollado un **modelo termoelectrónico tridimensional**, que es lo suficientemente general para simular cualquier tipo de electrodo e incluso el horno completo. La necesidad de resolver el problema electromagnético a partir de datos realistas, es decir, medibles en la práctica, nos ha llevado a investigar en el campo de las ecuaciones de Maxwell en baja frecuencia (ver [15], para detalles).

4 Modelado de nuevos sistemas de coladas para ferroaleaciones

4.1 Descripción del proceso físico

El silicio producido en los hornos de arco eléctrico se extrae en estado líquido a través de un orificio situado en la parte inferior de la cuba. El sólido que se obtiene tras su enfriamiento debe ser desmenuzado para obtener trozos de silicio del tamaño deseado. El silicio solidifica en forma de cristales. En el frente de cristalización, las impurezas que contiene el metal se desplazan a la parte aún líquida, por lo que la mayor concentración de éstas se dará en las últimas zonas que hayan alcanzado la temperatura de solidificación.

El procedimiento más sencillo consiste en almacenar el silicio líquido en

unos recipientes cilíndricos llamados *cucharas*, para dejarlo enfriar al aire (ver Figura 5). Pero con este sistema el eje central del bloque de silicio contendrá un mayor porcentaje de impurezas que los bordes y al triturarlo no es posible conseguir que la pureza de cada trozo sea homogénea, como exigen algunos de los usos de este producto. En este contexto, en la Sección 5 se presenta un nuevo mecanismo de enfriamiento del silicio a partir de la cuchara tradicional con el objetivo de obtener un silicio con un reparto de impurezas más homogéneo.



Figura 5: Esquema y foto de la colada clásica.

Ferroatlántica I+D ha desarrollado una alternativa a los sistemas tradicionales de colada que consiste en verter el silicio líquido en una placa refrigerada para producir una delgada lámina de metal en la que la concentración de impurezas sea constante en el sentido longitudinal (ver Figura 7). El sistema consiste en un canal que riega la placa en zig-zag y en una placa que está inclinada y posee un mecanismo que la hace vibrar para permitir que la lámina de silicio se deslice sin pegarse a la placa. Una serie de tubos por los que circula agua la enfrían por su parte inferior.

La placa de enfriamiento se compone de varias secciones. Cada una de ellas mide casi 3 metros de largo, cerca de 2 metros de ancho y 50 mm de alto. Las dos primeras son de cobre y las restantes de hierro, hasta alcanzar una longitud de unos 15 metros. El sistema de refrigeración está compuesto por varios grupos de tubos horadados en la placa (ver Figura 6). Una serie de seis tubos se repite a lo largo de toda la placa. Los tubos de mayor diámetro contienen en su interior otro tubo de diámetro menor, soldado a su parte inferior, para que el agua circule por la parte superior del primero.

La capa de silicio cubre el 80% de la anchura de la placa. Se estima que la producción ronda las 8 toneladas de silicio por hora.

La simulación numérica de la transferencia de calor que se produce en la lámina de silicio y en la placa se convierte en una herramienta muy interesante

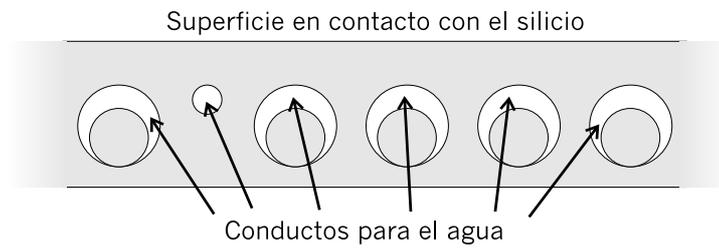


Figura 6: Detalle de un corte longitudinal de la placa.

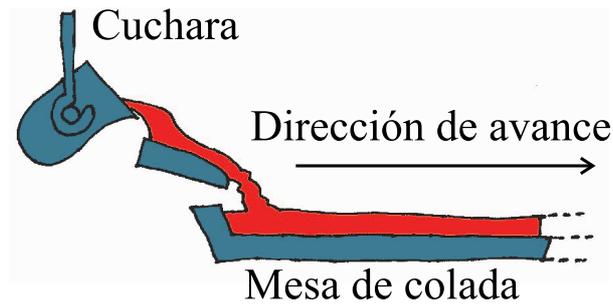


Figura 7: Esquema y foto de una placa de colada.

para conocer mejor el proceso de enfriamiento. Son varios los objetivos que se persiguen: por un lado se desea aumentar el caudal de silicio colado, pero sin producir daños por fusión o deformación en la placa de cobre (nótese que el silicio funde a una temperatura muy superior a la del cobre); por otro se pretende conocer (e intentar controlar) la región de la lámina de silicio que solidifica en último lugar, es decir, donde se concentran la mayor parte de las

impurezas. Para ello, se ha desarrollado un modelo matemático que permite obtener la distribución de temperaturas en la placa de refrigeración cubierta por la lámina de silicio en movimiento continuo.

4.2 El modelo matemático

El problema que se plantea es hallar la temperatura de un cuerpo compuesto por la plancha de cobre refrigerada y la lámina de silicio existente sobre ella.

Dado que la anchura de la placa es mucho mayor que su altura y los gradientes de temperatura son mucho menores en la primera dirección que en la segunda, es factible restringir el problema al estudio de lo que ocurre en una sección longitudinal de la placa (ver Figura 8). Tomaremos como dominio $\Omega \subset \mathbb{R}^2$ la unión de la sección media de la placa de cobre, Ω_c , y de la lámina de silicio, Ω_s .

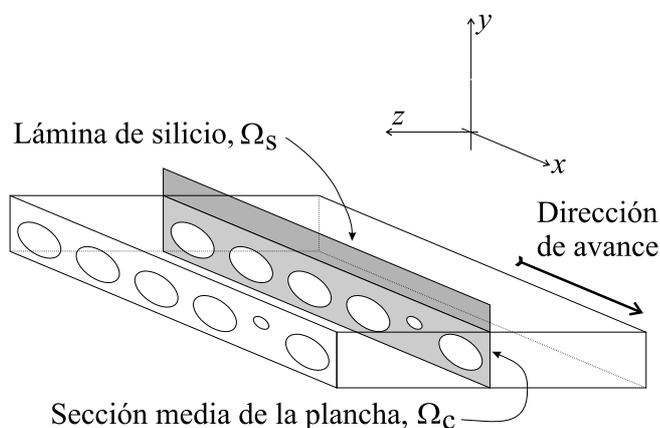


Figura 8: Sección media de la placa.

El modelo se obtiene a partir de la ecuación de transferencia de calor en estado transitorio. Además, como el silicio cambia de estado líquido a sólido debemos tener en cuenta el calor latente liberado al alcanzar la temperatura de solidificación T_S . Por ello, conviene escribir la ecuación de transmisión del calor en términos de la variable *entalpía* e :

$$\frac{\partial e(\mathbf{x})}{\partial t} + \mathbf{v}(\mathbf{x}) \cdot \mathbf{grad} e(\mathbf{x}) - \operatorname{div}((k(\mathbf{x}) \mathbf{grad} T(\mathbf{x}))) = 0, \quad (2)$$

donde \mathbf{x} es la coordenada espacial, T es la temperatura, \mathbf{v} es la velocidad y k la conductividad térmica.

La entalpía en el silicio se expresa como función de la temperatura a través de un operador monótono multivaluado

$$e(\mathbf{x}) \in H(\mathbf{x}, T), \quad (3)$$

siendo

$$\mathcal{H}(\mathbf{x}, T) = \begin{cases} \int_0^T \rho c ds & T < T_S \\ \left[\int_0^T \rho c ds, \int_0^T \rho c ds + \rho(\mathbf{x}, T_S) L \right] & T = T_S \\ \int_0^T \rho c ds + \rho(\mathbf{x}, T_S) L & T > T_S \end{cases} \quad (4)$$

donde L el “calor latente de fusión” o calor por unidad de masa necesario para realizar el cambio de estado, ρ es la densidad del silicio y c su calor específico.

En el caso de la placa, al no haber cambio de estado, la expresión para la entalpía es, simplemente

$$H(\mathbf{x}, T) = \int_0^T \rho(\mathbf{x}, s) c(\mathbf{x}, s) ds. \quad (5)$$

Cabe señalar que todos los parámetros dependen de la posición (por haber dos materiales, silicio y cobre) y de la temperatura.

La velocidad a la que se mueve el silicio se supone constante y con sólo componente horizontal. La placa de cobre, por su parte, está quieta. La velocidad es, por tanto,

$$\mathbf{v} = \begin{cases} v^* \mathbf{e}_x, & \text{en } \Omega_s \\ \mathbf{0}, & \text{en } \Omega_c. \end{cases}$$

El modelo se completa definiendo condiciones de contorno adecuadas. En concreto, en la frontera vertical izquierda del dominio correspondiente a la entrada del silicio, se supone conocida e igual a la temperatura a la que éste sale de la cuchara. En el resto de las fronteras, se consideran condiciones de radiación-convección siendo nula la radiación en la zona en contacto con el agua. Los coeficientes de convección han sido obtenidos a partir de fórmulas semiempíricas de la bibliografía (ver [14] para más detalles).

Por otra parte, la placa posee un mecanismo vibratorio que permite el deslizamiento del silicio. Como la amplitud de las vibraciones es pequeña comparada con el movimiento de la lámina de silicio, hemos supuesto que la velocidad sólo tiene componente horizontal. Sin embargo, un aspecto en el que inciden de forma determinante las vibraciones es la transmisión del calor entre el silicio y la placa. Llamemos Γ_{I+} y Γ_{I-} a las superficies de silicio y cobre en las que se produce el contacto, T_+ a la temperatura del silicio y T_- a la del cobre. Debido a la vibración, no podremos asumir que exista un contacto térmico perfecto entre ambos materiales, sino que supondremos que existe una transmisión imperfecta, es decir, que existe una “resistencia de contacto”, regulada por una ley del tipo siguiente:

$$-k(T_+) \frac{\partial T_+}{\partial \mathbf{n}} = h_r(T_+ - T_-) \text{ en } \Gamma_{I+}, \quad (6)$$

$$-k(T_-) \frac{\partial T_-}{\partial \mathbf{n}} = h_r(T_- - T_+) \text{ en } \Gamma_{I-}. \quad (7)$$

El coeficiente h_r es a priori desconocido y muy difícil de medir de forma directa. Por ello, este parámetro ha sido ajustado numéricamente a partir de medidas experimentales de la temperatura de salida del agua de refrigeración.

La ecuación (2) se discretiza en tiempo utilizando un esquema implícito en el que el término convectivo se discretiza mediante el método de características. La ecuación semidiscretizada se resuelve usando elementos finitos continuos y lineales a trozos sobre su formulación variacional. Para tener en cuenta el calor latente de solidificación del silicio y la no linealidad debida a los parámetros, se ha usado un algoritmo iterativo (ver [4] para más detalle). El hecho de no conocer la resistencia térmica entre silicio y placa nos ha obligado utilizar este parámetro como elemento de ajuste para que el modelo reproduzca el incremento observado en la temperatura de salida del agua de refrigeración.

4.3 Resultados numéricos

El algoritmo para resolver el problema de la colada ha sido implementado en un ordenador mediante un programa escrito en lenguaje Fortran. Presentamos aquí los resultados de una simulación sobre un dominio semejante al de la placa de enfriamiento utilizada por Ferroatlántica S.L.

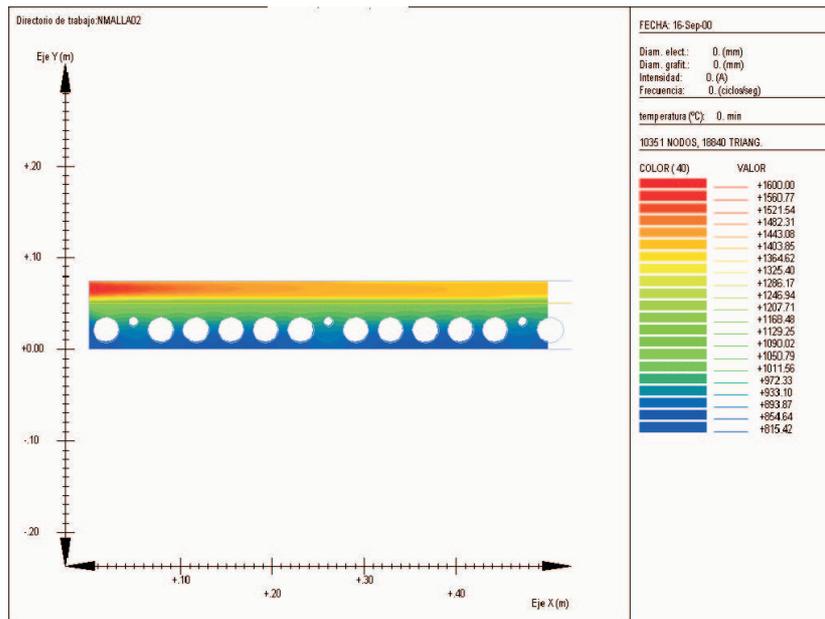


Figura 9: Temperatura a la entrada de la placa

En todas las figuras, el dominio es una sección longitudinal de la placa junto con una sección de la lámina de silicio situado sobre ella. El silicio líquido entra en el dominio por la parte izquierda y el movimiento vibratorio lo desplaza hacia

la derecha. Como se ve en la Figura 9, la temperatura en la lámina de silicio es superior en su interior pues, tanto el aire como el agua a través del cobre, extraen el calor del silicio líquido.

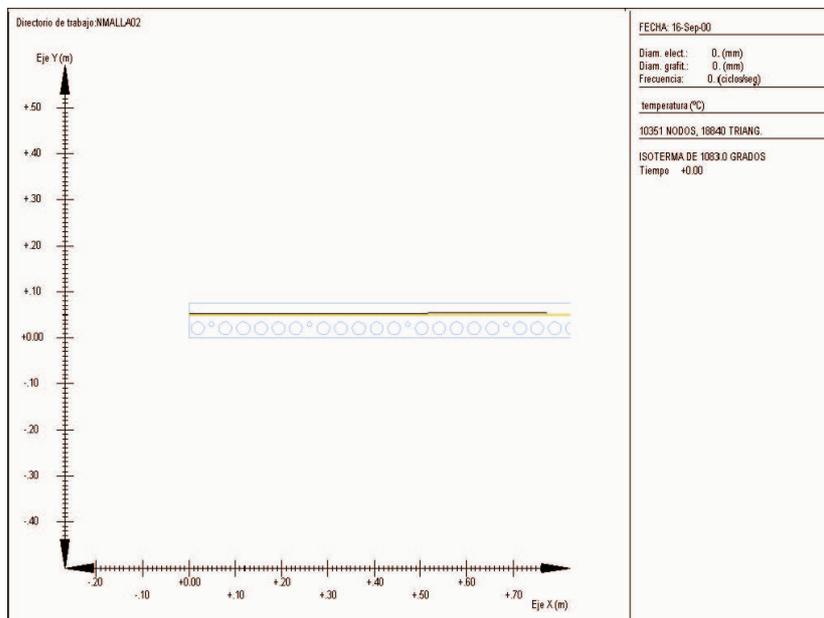


Figura 10: Isotherma de fusión del cobre.

La temperatura de entrada se toma a $1600\text{ }^{\circ}\text{C}$, casi $200\text{ }^{\circ}\text{C}$ por encima de su temperatura de fusión.

Lógicamente, el comienzo de la placa es el lugar donde existe un mayor riesgo de que el cobre se funda o deforme por alcanzar altas temperaturas. Se han hecho varias pruebas para ajustar cuál debería ser el caudal de agua a circular por los tubos de refrigeración para evitar que el cobre alcance su temperatura de fusión, $1083\text{ }^{\circ}\text{C}$. La resistencia de contacto provoca que la temperatura deje de ser continua al pasar a través de la interfase entre grafito y cobre. Como se ve en la Figura 10, la isoterma de $1083\text{ }^{\circ}\text{C}$ no interseca la placa de cobre.

5 Modelos de purificación del silicio

5.1 Descripción del proceso físico

En los últimos años se ha producido un importante crecimiento de la demanda de silicio de alta pureza para la fabricación de placas solares. Actualmente el silicio solar se obtiene como un subproducto de la fabricación de silicio electrónico. Sin embargo las estimaciones de la demanda de silicio solar superan a las del electrónico y es por eso que, en las grandes compañías están llevando a

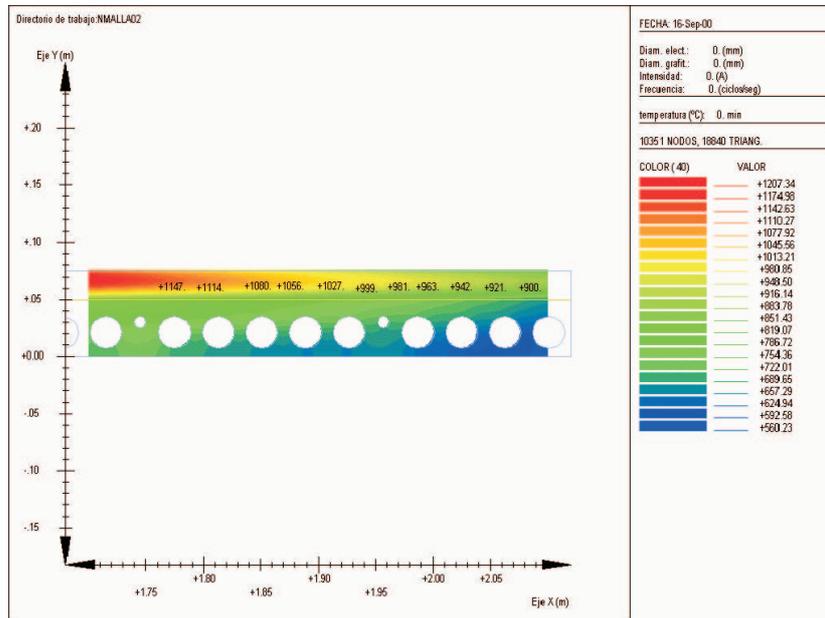


Figura 11: Temperatura en la parte final del dominio.

cabo grandes proyectos de investigación para encontrar métodos alternativos y directos para la purificación del silicio metal de manera que se obtenga silicio solar. En este contexto, el diseño de nuevos sistemas de colada como el visto en la sección anterior u otros procedimientos como el que se describe en esta sección persiguen obtener siempre un silicio más puro.

Un sistema utilizado por Ferroatlántica para enfriar el silicio que sale del horno consiste en un recipiente cilíndrico que contiene en su interior un crisol (ver Figura 12). En el crisol se introduce silicio fundido que irá solidificando por la pérdida de calor, creando un frente de solidificación que progresa de la parte inferior a la superior. Es importante que este frente de solidificación sea lo más horizontal posible de modo que las impurezas emigren hacia las capas altas para ser posteriormente eliminadas. Por ello, conocer y controlar la distribución de temperaturas en el silicio bajo diferentes condiciones de operación y con diferentes geometrías, son los objetivos que se persiguen con la simulación numérica de este proceso.

5.2 El modelo matemático

El modelo matemático que permite obtener la distribución de temperaturas en el recipiente descrito anteriormente y en el silicio contenido en su interior es la ecuación de transmisión del calor en estado transitorio sin convección, es decir con velocidad nula. Supondremos que los campos no dependen de la coordenada angular, lo cual permite resolver el problema en un dominio

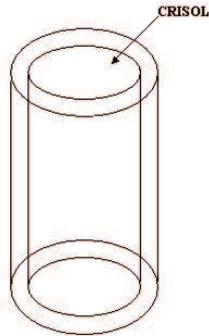


Figura 12: Esquema del dominio.

bidimensional. En concreto, el dominio de cálculo es la sección del dominio cilíndrico que se presenta en la Figura 13. Nótese que el interior del cilindro está formado por diferentes materiales: la capa externa es una manta de alúmina aislante, mientras que la interior está formada por cemento.

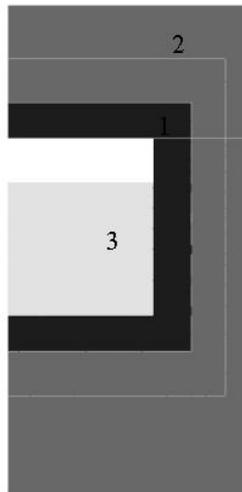


Figura 13: Sección radial de la cuchara: manta de alúmina (2), cemento (1), silicio (3).

Dado que el silicio cambia de estado durante el proceso es necesario utilizar la ecuación de transmisión del calor en términos de la entalpía, que escrita en

coordenadas cilíndricas queda de la forma:

$$\dot{e} - \frac{1}{r} \frac{\partial}{\partial r} \left(k \frac{\partial T}{\partial r} \right) - \frac{\partial}{\partial z} \left(k \frac{\partial T}{\partial z} \right) = 0. \quad (8)$$

La expresión de la entalpía en términos de la temperatura es la descrita en la sección anterior dependiendo de si el material experimenta o no un cambio de estado.

Con respecto a las condiciones de contorno, en las fronteras exteriores del recipiente, se considera una condición de convección mientras que en el hueco que queda entre la tapa y el crisol se considera una condición de radiación. En el resto de las fronteras se proporciona el flujo de calor, que es conocido a partir de resistencias internas que calientan el recipiente.

El esquema numérico utilizado es el mismo que el descrito en la sección anterior.

5.3 Resultados numéricos

Se ha desarrollado un programa Fortran que permite resolver la ecuación de transferencia de calor en la cuchara descrita anteriormente. En esta sección se presentan los resultados obtenidos al estudiar el comportamiento de la cuchara bajo determinadas condiciones de operación. En concreto, se realiza un precalentamiento de la cuchara mediante resistencias internas durante 10 horas. A continuación, se introduce el silicio en el crisol a una temperatura de 1480 °C y se conectan resistencias en la parte inferior de la tapa. La potencia máxima de las resistencias internas es de 10 Kw, manteniendo esta potencia mientras las temperaturas en ellas sean inferiores a 1450 °C. A partir de 1450 °C, la potencia de las resistencias disminuye linealmente hasta ser apagadas definitivamente a 1550 °C.

Con la composición de materiales descrita en la sección anterior, la Figura 14 muestra el frente de solidificación obtenido (temperaturas entre 1412 y 1414 °C) a las 10 horas.

Dado que el principal objetivo consistía en obtener un frente de solidificación más horizontal, se abordó la posibilidad de introducir un nuevo material muy conductor tal y como se indica en la Figura 15. La conclusión es que cuanto mayor sea la conductividad de dicho material, el frente es más horizontal. La Figura 16 presenta el frente de solidificación obtenido con esta nueva geometría al cabo de 10 horas, siendo la conductividad eléctrica del material 1000 W/mK.

6 Líneas de investigación en curso: modelado de un horno de inducción para la purificación de silicio.

En el contexto de la sección anterior, actualmente la empresa Ferroatlántica I+D está investigando nuevos sistemas de purificación de silicio. En concreto, el

Rango de isothermas entre 1412°C y 1414°C, a las 10 horas

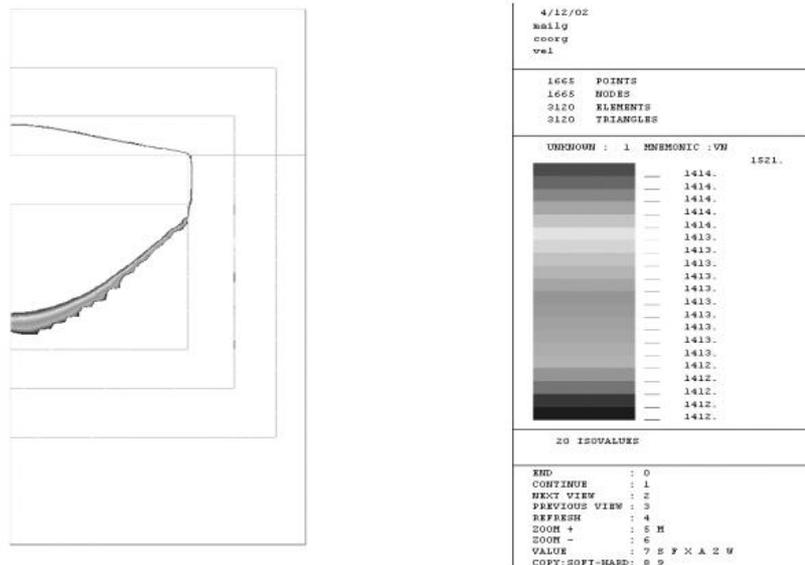


Figura 14: Frente de solidificación utilizando 3 materiales.

diseño de un horno destinado al calentamiento por inducción electromagnética, del silicio contenido en un crisol, es uno de los actuales temas de investigación. En esta sección, se describen brevemente los modelos matemáticos que se están desarrollando por parte del Departamento de Matemática Aplicada para contribuir al diseño de dicho horno.

Cabe señalar que el calentamiento por inducción electromagnética es ampliamente utilizado en la industria actual en procesos como fundición de metales, precalentamiento para operaciones de soldadura, y en general, en aquellos procesos que requieren una velocidad alta de calentamiento en zonas localizadas de la pieza de un material conductor. Se trata de un proceso complejo que involucra fenómenos electromagnéticos, fenómenos térmicos con cambio de estado y fenómenos hidrodinámicos en las fases líquidas del metal. Por ello, la simulación numérica es un herramienta muy eficaz en este campo. En el caso concreto del horno de inducción, la simulación numérica permite conocer y controlar los parámetros que intervienen en su funcionamiento.

El objetivo que se persigue es estudiar los diferentes modelos matemáticos que permiten simular los procesos físico-químicos mencionados anteriormente y desarrollar métodos numéricos adecuados para su resolución numérica. Se pretende, por tanto, la realización de un programa de ordenador que resuelva numéricamente las ecuaciones de la termo-electromagneto-hidrodinámica de

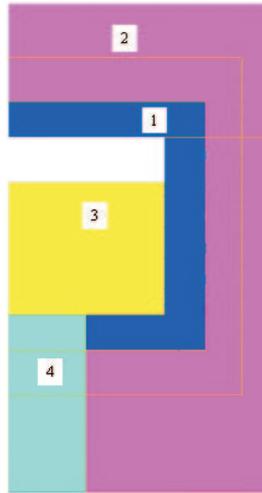


Figura 15: Sección radial de la cuchara: manta de alúmina (2), cemento (1), silicio (3), material muy conductor (4).

modo que se puedan obtener, en un tiempo de cálculo razonable, las corrientes inducidas, la distribución de la temperatura junto con la posición de la interfase sólido-líquido y el campo de velocidades en el material líquido.

Desde un punto de vista matemático, el modelado completo de un horno de inducción implica el acoplamiento de las ecuaciones de Maxwell, con la ecuación del calor con términos de convección-difusión y cambio de estado, así como las ecuaciones de Navier-Stokes para un fluido incompresible. Las relaciones entre los modelos anteriores se esquematizan en el diagrama que se presenta en la Figura 17.

Este proceso presenta características muy complejas entre las que cabe destacar:

• **Acoplamiento de los fenómenos:**

- Modelo Electromagnético - Modelo Térmico: el campo de temperaturas depende de la densidad de corriente a través del efecto Joule, mientras que las conductividades eléctricas de los diferentes materiales dependen de la temperatura.
- Modelo Electromagnético - Modelo hidrodinámico: la fuerza de Lorentz combinada con la gravitatoria y las diferencias de densidad con la temperatura, provocan el movimiento del metal fundido y dicho movimiento produce, a su vez, variaciones de los campos electromagnéticos a través de la Ley de Ohm.
- Modelo térmico - Modelo hidrodinámico: La resolución del primero dará como resultado la posición del frente de solidificación del metal y,

- La fuente de calor en el modelo térmico es la densidad de potencia del efecto Joule que es proporcional al cuadrado de la densidad de corriente.
 - Las propiedades físicas de los materiales que forman el horno y del propio metal que se funde dependen, en general de forma no lineal, de la temperatura.
 - La fusión del metal dentro del horno requiere que se resuelva la ecuación del calor evolutiva con cambio de estado, lo que introduce una fuerte no linealidad en el problema.
 - Las condiciones de contorno del modelo térmico serán de convección-radiación con lo cual el flujo de calor en la frontera depende de la potencia cuarta de la temperatura.
- **Existencia de fronteras libres:** la interfase sólido-líquido del material que se introduce en el horno de inducción. Téngase en cuenta que el conocimiento de dicha interfase es crucial para los modelos electromagnético e hidrodinámico, puesto que las propiedades físicas varían considerablemente al cambiar de estado y, además, el modelo hidrodinámico solo se desarrolla en la fase líquida.

En la literatura pueden encontrarse algunos trabajos donde se acoplan el modelo térmico y el modelo electromagnético (ver por ejemplo [8], [9]) o bien el modelo electromagnético y el modelo hidrodinámico ([13]). Sin embargo, el acoplamiento de los tres modelos constituye una línea de investigación novedosa. Recientemente, en [11] se presenta la simulación numérica de un horno de inducción, también destinado a la purificación del silicio, que se lleva a cabo mediante códigos comerciales.

Finalmente, las técnicas numéricas que se emplearán para la resolución de las ecuaciones que modelan los tres procesos anteriormente descritos son:

- Modelo electromagnético: se utilizará un método híbrido de elementos finitos (FEM) y elementos de contorno (BEM). La combinación de ambas técnicas permite utilizar sus respectivas ventajas numéricas: el método de elementos finitos nos permite aproximar bien la geometría y tratar eficazmente los términos no lineales, y las matrices resultantes son "huecas". Por otro lado, los elementos de contorno proporcionan un método eficaz para tratar dominios homogéneos no acotados. Los BEM reducen el número de incógnitas pero tienen el inconveniente de que dan lugar a matrices llenas.
- Modelo térmico: La ecuación del calor se integrará en tiempo mediante un esquema implícito mientras que, en espacio, se utilizará el método de los elementos finitos. Además, se emplearán técnicas de operadores maximales monótonos para el tratamiento del cambio de estado.
- Modelo hidrodinámico: se integrarán en tiempo las ecuaciones de Navier-Stokes utilizando un esquema implícito donde el término inercial será

aproximado mediante un método de características. En cada paso de tiempo, se utilizará un método de elementos finitos; en particular, se considerarán elementos continuos lineales a trozos con “burbujas” para las componentes de la velocidad y lineales a trozos para la presión.

Agradecimientos

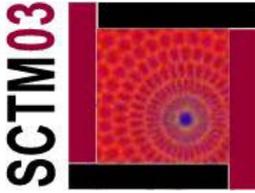
Este trabajo ha sido financiado por FERROATLANTICA I+D y los proyectos FEDER-CICYT 1FD97.0280 y PGIDT00PXI20701PR.

Referencias

- [1] A. Bermúdez, J. Bullón and F. Pena, “A finite element method for the thermoelectrical modelling of electrodes”, *Commun. Numer. Meth. Engng.* **14**, (1998), 581-593.
- [2] A. Bermúdez and R. Muñoz, “Existence of solution of a coupled problem arising in the thermoelectrical simulation of an electrode”, *Quart. of Appl. Math.* **57**, (1999), no. 4, 621-636.
- [3] A. Bermúdez, J. Bullón, F. Pena, P. Salgado, “Modelado y simulación numérica de electrodos para hornos metalúrgicos”, CD-ROM Actas del XVII Congreso de Ecuaciones Diferenciales y Aplicaciones/VII Congreso de Matemática Aplicada (XVII CEDYA/VII CMA) Salamanca, España, 2001.
- [4] A. Bermúdez, J. Bullón y F. Pena, “Simulación estacionaria de una colada de silicio” CD-ROM Actas del XVII Congreso de Ecuaciones Diferenciales y Aplicaciones/VII Congreso de Matemática Aplicada (XVII CEDYA/VII CMA) Salamanca, España, 2001.
- [5] A. Bermúdez, R. Rodríguez and P. Salgado, “A finite element method with Lagrange multipliers for low-frequency harmonic Maxwell equations”, *SIAM J. Numer. Anal.*, **40**, (2002), 1823-1849.
- [6] A. Bermúdez, R. Rodríguez and P. Salgado, “Numerical treatment of realistic boundary conditions for the eddy current problem in an electrode via Lagrange multipliers”, Preprint 2002-11, Universidad de Concepción, Chile, 2002.
- [7] A. Bermúdez, J. Bullón, F. Pena and P. Salgado, “A numerical method for transient simulation of metallurgical compound electrodes”, *Finite Elem. Anal. Des.*, **39**, (2003), 283-299.
- [8] C. Chaboudez, S. Clain, R. Glardon, D. Mari, J. Rappaz, J. and M. Swierkosz, “Numerical Modeling in Induction Heating for axisymmetric geometries”, *IEEE Transactions on Magnetics*, **33**(1), (1997), 739-745.

- [9] S. Clain, J. Rappaz, M. Swierkosz and R. Touzani “Numerical modelling of induction heating for two-dimensional geometries”, *Mathematical Models and Methods in Applied Sciences*, **3**(6), (1993), 805-822.
- [10] P. D’Ambrosio and I. Letizia, “Temperature and stress distribution on carbon electrodes for silicon metal production under transient temperature conditions”, *16th Biennial Conference on Carbon*, Baden-Baden, (1983).
- [11] Y. Delannoy, C. Alemany, K.-I. Li, P. Proulx and C. Trassy, “ Plasma-refining process to provide solar-grade silicon”, *Solar Energy Materials & Solar Cells*, **72**, (2002), 69-75.
- [12] R. Innvær and L. Olsen, “Practical use of mathematical models for Søderberg electrodes”, *Elkem Carbon Technical Paper presented at the A.I.M.E. Conference* (1980).
- [13] T.T. Natarajan and N. El-Kaddah, “A methodology for two-dimensional finite element analysis of electromagnetically driven flow in induction stirring systems”, *IEEE Transactions on Magnetism*, **35**(3), (1999), 1773-1776.
- [14] F. Pena, *Contribución al modelado matemático de algunos problemas de la metalurgia del silicio*, Tesis Doctoral, Universidade de Santiago de Compostela, 2003
- [15] P. Salgado, *Mathematical and numerical analysis of some electromagnetic problems. Application to the simulation of metallurgical electrodes*, Tesis Doctoral, Universidade de Santiago de Compostela, 2002
- [16] A. Schei, J.K. Tuset and H. Tveit, *High Silicon alloys*, Tapir Forlag, Trondheim, Noruega, 1998.

Generación automática de mallas tridimensionales para la simulación numérica de procesos medioambientales



Rafael A. Montenegro Armas

Catedrático de Matemática Aplicada y Director de la División de Discretización y Aplicaciones del Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería de la Universidad de Las Palmas de Gran Canaria

Resumen

El objetivo de este trabajo es presentar los principales aspectos de un código capaz de generar automáticamente mallas de tetraedros a partir de una distribución adecuada de nodos en el dominio de estudio, así como plantear una versión de un algoritmo de refinamiento local basado en la subdivisión en 8-subtetraedros. En concreto, el dominio está limitado en su parte inferior por la superficie del terreno considerada como $z = f(x, y)$, tal que $\forall (x, y) \in [a, b] \times [c, d] \subset R^2 \Rightarrow \exists ! z \in R$, y en su parte superior por un plano horizontal. Las paredes laterales están formadas por cuatro planos verticales, paralelos dos a dos. Las ideas básicas para la construcción de la malla combinan, por un lado, la utilización de un algoritmo de refinamiento y desrefinamiento para dominios bidimensionales [6], por otro lado, un algoritmo de generación de mallas de tetraedros basado en la triangulación de Delaunay [5], y además, una función para definir el espaciado vertical de los nodos distribuidos en el dominio. Este código ha sido utilizado anteriormente para generar mallas donde la superficie $z = f(x, y)$ correspondía a una digitalización topográfica del sur de la isla de la Palma, [19] y [20]. Asimismo, en el generador de mallas entran en juego un procedimiento para optimizar la malla resultante y diferentes estrategias para construir la nube de puntos. Una vez que se ha generado la malla atendiendo a la geometría del dominio, es necesario adaptarla a la solución numérica del problema. Por ello, dedicamos una segunda parte de este trabajo al desarrollo e implementación de un algoritmo de refinamiento local de mallas encajadas de tetraedros, cuyo principal interés es su aplicabilidad, rapidez y sencillez. El algoritmo propuesto se basa en la división expuesta en [16] cuyo análisis de calidad se estudia en [15].

1. Introducción

El trabajo se ha dividido en dos partes. La primera se desarrolla en la sección 2 y se dedica al análisis del generador automático de mallas de tetraedros de un dominio tridimensional definido sobre un terreno irregular. En la sección 3 presentamos el algoritmo de refinamiento local de mallas de tetraedros. Estas dos grandes líneas se engloban entre los aspectos básicos necesarios para la simulación numérica del problema de dispersión de contaminantes en la atmósfera, que pretendemos afrontar.

2. Generación de mallas sobre orografía irregular

Inicialmente, nos planteamos crear una malla de tetraedros que respete la topografía de la región de estudio con una precisión determinada. Para ello, se dispone únicamente de la

información digitalizada del terreno. Por otra parte, deseamos que la malla esté adaptada con una mayor densidad de nodos donde sea necesario para definir las características geométricas de nuestro dominio a partir de una interpolación lineal a trozos. La malla generada podrá utilizarse como malla base para la simulación numérica de procesos naturales en el dominio; por ejemplo, ajuste de campos de viento [27] y [21], propagación de fuego [18], contaminación atmosférica, etc. Estos fenómenos tienen su mayor efecto en las zonas próximas al terreno, de ahí que también sea deseable que la densidad de nodos aumente al acercarnos a éste. Sobre esta malla base, que únicamente se adapta a las características geométricas del dominio, se podrán aplicar posteriormente algoritmos de refinamiento y desrefinamiento de tetraedros para mejorar la solución numérica del problema [16] y [15]. Estos algoritmos tendrán un especial interés en los problemas evolutivos.

Es bien conocido que para construir una triangulación de Delaunay es necesario definir una nube de puntos en el dominio y su frontera. Estos nodos serán precisamente los vértices de los tetraedros que conforman la malla. La generación de puntos en nuestro dominio se realizará sobre diferentes capas, reales o ficticias, definidas desde el terreno hasta la parte superior del dominio. En concreto, se propone plantear una malla rectangular con una distribución uniforme de puntos en el plano superior del dominio. Esta malla bidimensional puede ser obtenida a partir de la realización de un cierto número de refinamientos globales sobre una malla simple definida en la entrada de datos, o por ejemplo, puede también construirse realizando una triangulación de Delaunay sobre la distribución uniforme de puntos establecida. Consideraremos la malla obtenida como el nivel más bajo de la secuencia que define la distribución de los puntos en el resto de las capas. Sobre esta malla regular aplicamos a continuación el algoritmo de refinamiento y desrefinamiento, [6] y [23], para definir la distribución de los nodos de la capa correspondiente a la superficie del terreno. Para ello, en primer lugar se construye una función que interpola las cotas obtenidas a partir de una digitalización de la topografía de la zona rectangular estudiada. En segundo lugar, realizamos una serie de refinamientos globales sobre la malla uniforme hasta conseguir una malla regular capaz de captar la variación topográfica del terreno. El máximo grado de discretización viene definido por el nivel de detalle de la digitalización. Posteriormente, se realizará un desrefinamiento sobre estos últimos niveles de malla utilizando como parámetro de desrefinamiento el máximo error de cotas permitido entre la superficie real del terreno y la superficie definida mediante la interpolación a trozos obtenida con la malla resultante.

Una vez que se ha definido la distribución de nodos sobre el terreno y sobre el plano superior del dominio, comenzamos a distribuir los nodos situados entre ambas capas. Esta distribución se puede realizar mediante diferentes estrategias, en las que interviene una función de espaciado vertical. La característica fundamental de esta función es que el grado de discretización obtenido sobre la vertical debe disminuir con la altura, o a lo sumo mantenerse constante.

La distribución de puntos en el dominio entrará como dato en el mallador tridimensional basado en la triangulación de Delaunay. Para evitar posibles problemas de conformidad con la superficie del terreno, se propone construir la malla de tetraedros con la ayuda de un paralelepípedo auxiliar. Sobre su cara inferior se sitúan todos los nodos distribuidos sobre el terreno, proyectados sobre un plano horizontal situado a la altura definida por la cota inferior de la región de estudio, y sobre su cara superior se sitúan los puntos distribuidos en el plano superior del dominio a su altura real. Esto conlleva a realizar una transformación, atendiendo a la función de espaciado sobre cada vertical, para situar el resto de puntos en el paralelepípedo auxiliar. Estos detalles nos asegurarán que la distancia máxima entre dos puntos consecutivos

sobre la misma vertical del dominio real será siempre igual o inferior que la correspondiente distancia establecida en el paralelepípedo auxiliar.

La calidad de la malla final, obtenida mediante la transformación inversa al dominio real, depende de la distribución de los puntos definida en ambos dominios, ya que respetaremos la topología de la malla obtenida en la triangulación del paralelepípedo auxiliar. Proponemos una estrategia para determinar el número de puntos generados sobre la vertical de cada nodo de la malla bidimensional adaptada a la superficie del terreno, y analizamos sus características fundamentales. Con esta estrategia se generan capas virtuales, es decir, no se define un número concreto de superficies interiores al dominio sobre las que se sitúan los puntos. Por ello, diremos que el número de capas es variable, y será calculado automáticamente en función de los tamaños de los elementos existentes en la malla bidimensional que define el terreno, y en la correspondiente a la frontera superior del dominio. En concreto, se determina automáticamente, para cada nodo del terreno, una función de espaciado vertical con el objeto de respetar las distancias desde el primer punto generado hasta el terreno, y desde el último punto generado hasta la parte superior del dominio, en función de los tamaños de los elementos existentes sobre ambas superficies.

Una vez que se ha construido la triangulación de Delaunay de la nube de puntos en el paralelepípedo, procedemos a situar los puntos en sus posiciones reales manteniendo la topología de la malla. Hay que tener en cuenta que este proceso de compresión de la malla puede dar lugar a cruces de tetraedros que habrá que deshacer posteriormente. Asimismo, será aconsejable aplicar una etapa de suavizado para mejorar la calidad de los elementos de la malla resultante.

2.1. Experimentos numéricos. Como aplicación real del generador de mallas se ha considerado una región rectangular del sur de la Isla de La Palma de 22.8×15.6 km, en la que las cotas extremas varían de 0 a 2279 m de altitud. La parte superior del dominio se ha establecido a una altitud $h = 6$ km. Para definir la topografía se dispuso de una digitalización de la zona en la que las alturas estaban definidas sobre una cuadrícula con un paso espacial de 200 m según los ejes x e y .

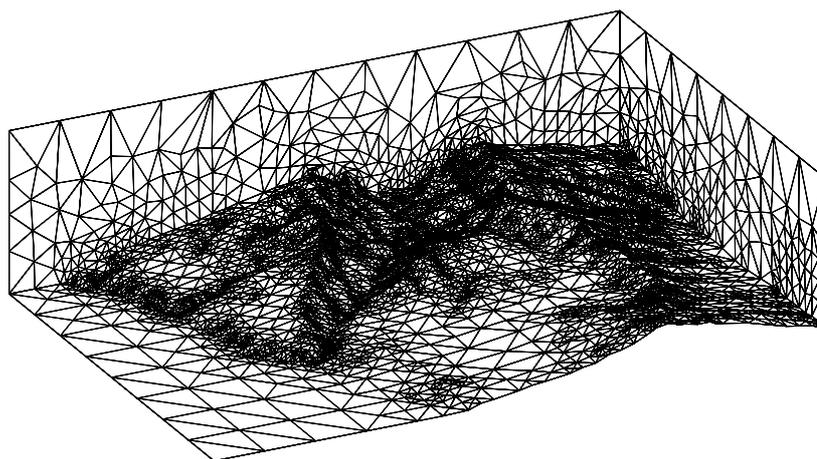


Figura 1. Malla resultante después de diez pasos del proceso de optimización para una región de la Isla de La Palma.

A partir de una malla uniforme de la región rectangular con un tamaño de elementos aproximadamente de $2 \times 2 \text{ km}$, se realizaron seis refinamientos globales utilizando el algoritmo 4-T de Rivara [25]. Una vez que se interpolaron los datos digitalizados sobre esta malla refinada, se empleó el algoritmo de desrefinamiento desarrollado en [6] y [23] con un parámetro de desrefinamiento $\varepsilon = 25 \text{ m}$. Esto asegura que la malla adaptada aproxima la superficie del terreno con un error menor que este valor. La distribución de nodos de la malla base regular utilizada antes de los refinamientos globales es la que se consideró sobre la frontera superior del dominio. En la figura 1 se representa la malla obtenida después de diez pasos del proceso de optimización.

También mostramos otras aplicaciones para una serie de superficies *test*. Las mallas que se exponen a continuación han sido generadas por nuestro código antes de la aplicación del proceso de optimización. Todas las superficies han sido definidas en una región rectangular de $10 \times 5 \text{ km}$. La parte superior del dominio se ha establecido a una altitud $h = 5 \text{ km}$. La primera aplicación corresponde a una superficie gaussiana relativamente suave y se representa en la figura 2. Hemos fijado un parámetro de desrefinamiento $\varepsilon = 20 \text{ m}$. Esta estrategia introduce 6 capas, 1155 nodos y no produce ningún tetraedro invertido.

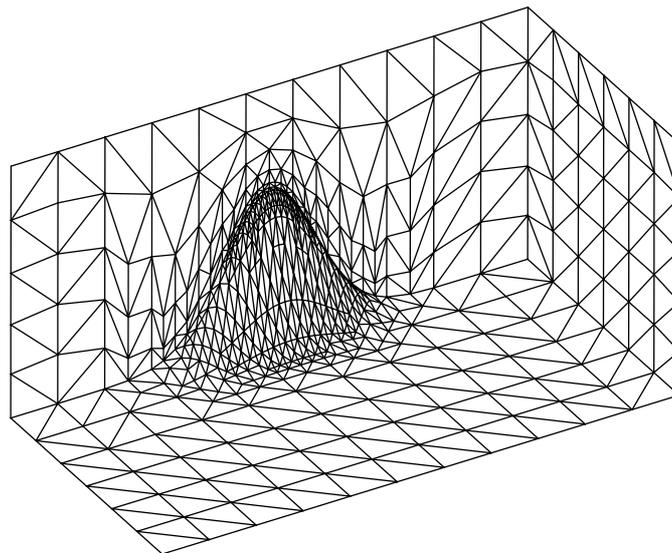


Figura 2. Malla resultante, antes de aplicar el proceso de optimización, para una superficie gaussiana.

Planteamos ahora una superficie con forma de *volcán* que fue definida mediante la composición de dos superficies gaussianas; la cota máxima en el dominio es $h = 1 \text{ km}$, mientras que la cota mínima del terreno es $z_{\min} = -1 \text{ km}$. En la figura 3 se representa un detalle de la malla obtenida con un parámetro de desrefinamiento $\varepsilon = 5 \text{ m}$. La malla resultante posee 3973 nodos sin existencia de cruces de tetraedros.

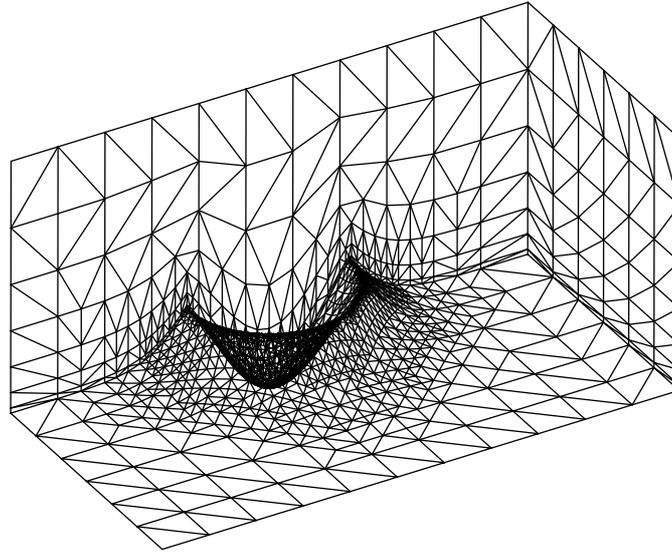


Figura 3. Malla resultante, antes de aplicar el proceso de optimización, para una superficie con forma de volcán.

2.2. Conclusiones. Hemos establecido y analizado los aspectos principales para generar una malla tridimensional de tetraedros que se adapta a una superficie definida sobre una región rectangular con una mínima intervención del usuario. En concreto se ha planteado una generación de puntos, bien distribuidos en el dominio de estudio, capaz de captar la información de la superficie y que posee una densidad menor a medida que aumenta la altura con respecto al terreno. Los puntos se generan aplicando técnicas de refinamiento/desrefinamiento en 2-D, una función de espaciado vertical y diferentes estrategias. Seguidamente, con la ayuda de un paralelepípedo auxiliar, se ha planteado un procedimiento basado en la triangulación de Delaunay para construir automáticamente la malla, asegurando la conformidad con la superficie del terreno. No obstante, la distribución de puntos obtenida también podría tener interés para generar la malla tridimensional con otras técnicas clásicas, tales como avance frontal [11] y *normal offsetting* [12]. Finalmente, el procedimiento propuesto para optimizar la malla generada ha permitido resolver al mismo tiempo los problemas de cruces de tetraedros y de calidad de la malla.

3. Refinamiento local de mallas de tetraedros

En la actualidad, la mayor parte de los programas que utilizan el método de elementos finitos se apoya en técnicas adaptables basadas en una estimación del error cometido con nuestra solución numérica, o al menos en indicadores de error fiables que nos señalen los elementos que deben ser refinados o desrefinados en la malla.

En generación de mallas adaptables podemos considerar dos aspectos diferentes: la discretización del dominio atendiendo a su geometría o a la solución numérica. Existen muchas formas de abordar estos aspectos. La primera cuestión es: ¿mallas estructuradas o no estructuradas?. En este sentido, está claro que el uso de mallas no estructuradas nos proporciona más flexibilidad a la hora de mallar geometrías complejas utilizando un número óptimo de nodos. En este caso, los métodos más clásicos para la obtención de triangulaciones tridimensionales se basan fundamentalmente en algoritmos de avance frontal [17] o en

algoritmos basados en la triangulación de Delaunay [10] y [5]. Una vez que se ha discretizado la geometría del dominio, la malla debe adaptarse atendiendo a las singularidades de la solución numérica. Este proceso implica la introducción (refinamiento) o eliminación (desrefinamiento) de nodos de la malla actual. Los cambios pueden afectar a la malla actual de forma local o global, dependiendo del método de triangulación elegido. Diferentes estrategias de refinamiento han sido desarrolladas para triangulaciones en 2-D, y han sido generalizadas a 3-D. Si se ha optado por un refinamiento que afecte localmente a la malla actual, cabe plantearse otra cuestión: ¿mallas encajadas o no encajadas?. La respuesta en este caso no es tan clara. El uso de mallas encajadas tiene varias ventajas importantes. Podemos conseguir familias de secuencias de mallas encajadas en un mínimo tiempo de CPU. Además, se puede aplicar más fácilmente el método multimalla para resolver el sistema de ecuaciones asociado al problema. Por otra parte, se puede controlar automáticamente la suavidad y la degeneración de la malla, y el mantenimiento de las superficies definidas en el dominio, en función de las características de la malla inicial. Si el dominio posee una geometría compleja, un buen modo de proceder es obtener la malla inicial empleando un generador de mallas no estructuradas y, posteriormente, aplicar una técnica de refinamiento y desrefinamiento local de mallas encajadas atendiendo a un indicador de error apropiado al problema. Además, si tratamos de resolver un problema evolutivo, podemos aproximar automáticamente cualquier solución inicial definida en el dominio. Con la técnica de refinamiento y desrefinamiento conseguimos un óptimo soporte de interpolación a trozos capaz de aproximar esta solución con la precisión deseada. En general, podría aplicarse esta técnica para cualquier función definida en el dominio de forma discreta o analítica.

La elección particular del algoritmo de refinamiento es muy importante, puesto que el algoritmo de desrefinamiento puede entenderse como el inverso del algoritmo de refinamiento. El algoritmo de refinamiento 4-T de Rivara posee buenas propiedades en cuanto a la suavidad y degeneración de la malla. Además de esto, el número de posibilidades que aparecen en la relación entre un elemento padre y sus hijos es menor que con otros algoritmos de refinamiento en 2-D, tras asegurar la conformidad de la malla. Por ejemplo, sería más complicado desarrollar un algoritmo de desrefinamiento, acoplado con el algoritmo de refinamiento local propuesto en [2]; todos los triángulos que deben ser refinados, atendiendo al indicador de error, se dividen en cuatro subtriángulos mediante la introducción de un nuevo nodo en los centros de sus lados y uniéndolos entre sí.

En 3-D, el problema es diferente. Aunque parezca paradójico, la extensión de un algoritmo adaptable que sea más simple que otro en 2-D, no tiene por qué ser también más simple en 3-D. Así, entre los algoritmos de refinamiento desarrollados en 3-D podemos mencionar los que se basan en la bisección del tetraedro [1], [26], [24], y los que utilizan la subdivisión en 8-subtetraedros [3], [15], [16]. En concreto, el algoritmo desarrollado en [24] se puede entender como la generalización a 3-D del algoritmo 4-T de Rivara, que a su vez está basado en la bisección del triángulo por su lado mayor. El problema que se produce en esta extensión a 3-D es el gran número de casos posibles en los que puede quedar dividido un tetraedro, respetando las diferentes posibilidades de la división 4-T en sus cuatro caras, durante el proceso de conformidad de la malla. Sin embargo, los algoritmos analizados en [3], [15], [16], que a su vez generalizan a 3-D la partición en cuatro subtriángulos propuesta en [2], son más sencillos debido a que el número de particiones posibles de un tetraedro es mucho menor que en el caso de la generalización del algoritmo 4-T. Por otra parte, puesto que la calidad de la malla está asegurada en todos estos casos, hemos optado por implementar en C++ una versión del algoritmo que utiliza la subdivisión en 8-subtetraedros.

3.1. Aplicaciones. Como ejemplo de aplicación del algoritmo de refinamiento, se presenta en la figura 4(a) una triangulación inicial τ_1 formada por 5072 tetraedros y 1140 nodos, que fue generada a partir del código introducido en [5]. Para proceder a su refinamiento se ha utilizado un indicador de error atendiendo a las distancias desde los centros de gravedad de los tetraedros a un vértice del dominio. En las figuras 4(b) y 4(c) se representan las mallas refinadas después de 1 y 2 etapas de refinamiento, compuestas por 5386 tetraedros y 1201 nodos, y 6270 tetraedros y 1433 nodos, respectivamente. Destacamos la rapidez del proceso de refinamiento, así como la calidad de la malla resultante.

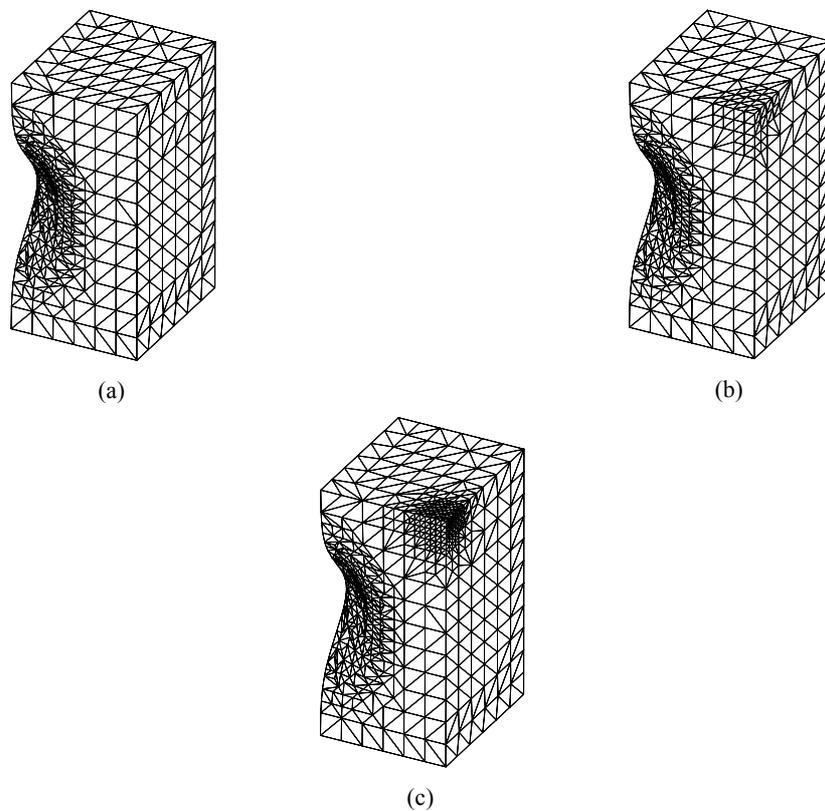


Figura 4. Aplicación del algoritmo de refinamiento: (a) malla inicial; (b) y (c) mallas resultantes tras 1 y 2 etapas de refinamiento, respectivamente.

3.2. Conclusiones. En este apartado se estudia un algoritmo de refinamiento local de mallas de tetraedros basado en la subdivisión en 8-subtetraedros propuesta en [3], [15] y [16]. Se ha aplicado de forma eficiente en mallas tridimensionales generadas mediante una versión del método de triangulación de Delaunay [5]. En trabajos futuros se propone desarrollar el correspondiente algoritmo de desrefinamiento y aplicarlo especialmente en problemas evolutivos de dispersión de contaminantes en la atmósfera.

Reconocimientos

El trabajo que aquí se expone es fruto de una investigación conjunta con G. Montero, J.M. Escobar, E. Rodríguez y J.M. González-Yuste, parcialmente subvencionada por el Ministerio de Ciencia y Tecnología y fondos FEDER a través del proyecto REN2001-0925-C03-02/CLI.

Bibliografía

- [1] D.N. Arnold, A. Mukherjee y L. Pouly, Locally adapted tetrahedral meshes using bisection, *SIAM J. Sci. Comput.*, **22**, 2, 431-448 (2000).
- [2] R.E. Bank, A.H. Sherman y A. Weiser, Refinement algorithms and data structures for regular local mesh refinement, in *Scientific Computing IMACS*, Amsterdam, North-Holland, (1983), 3-17.
- [3] F. Bornemann, B. Erdmann y R. Kornhuber, Adaptive multilevel methods in three space dimensions", *Int. J. Numer. Meth. Eng.*, **36**, 3187-3203 (1993).
- [4] H.N. Djidjev, *Force-directed methods for smoothing unstructured triangular and tetrahedral meshes*, Tech. Report, Dep. of Computer Science, Univ. of Warwick, Coventry, UK, (2000).
[Disponible en <http://www.andrew.cmu.edu/user/sowen/topics/new.html>].
- [5] J.M. Escobar y R. Montenegro, Several aspects of three-dimensional Delaunay triangulation, *Advances in Engineering Software*, **27**, 1/2, 27-39 (1996).
- [6] L. Ferragut, R. Montenegro y A. Plaza, Efficient refinement/derefinement algorithm of nested meshes to solve evolution problems, *Comm. Num. Meth. Eng.*, **10**, 403-412 (1994).
- [7] L.A. Freitag y P.M. Knupp, *Tetrahedral element shape optimization via the jacobian determinant and condition number*, en *Proceedings of the Eighth International Meshing Roundtable*, Sandia National Laboratories (1996), pp. 247-258.
- [8] L.A. Freitag y P. Plassmann, Local optimization-based simplicial mesh untangling and improvement, *Int. J. Numer. Methods Engng.*, **49**, 109-125 (2000).
- [9] L.A. Freitag y P.M. Knupp, Tetrahedral mesh improvement via optimization of the element condition number, *Int. J. Numer. Methods Engng.*, **53**, 1377-1391 (2002).
- [10] P.L. George, F. Hecht y E. Saltel, Automatic mesh generation with specified boundary, *Comp. Meth. Appl. Mech. Eng.*, **92**, 269-288 (1991).
- [11] H. Jin y R.I. Tanner, Generation of unstructured tetrahedral meshes by advancing front technique, *Int. J. Num. Meth. Eng.*, **36**, 1805-1823 (1993).
- [12] B.P. Johnston y J.M. Sullivan, Jr., A normal offsetting technique for automatic mesh generation in three dimensions, *Int. J. Num. Meth. Eng.*, **36**, 1717-1734 (1993).
- [13] P.M. Knupp, Algebraic mesh quality metrics, *SIAM J. Sci. Comput.*, **23**, 193-218 (2001).
- [14] P.M. Knupp, Achieving finite element mesh quality via optimization of the Jacobian matrix norm and associated quantities. Part II-A frame work for volume mesh optimization and the condition number of the Jacobian matrix, *Int. J. Numer. Methods Engng.*, **48**, 1165-1185 (2000).
- [15] A. Liu y B. Joe, Quality local refinement of tetrahedral meshes based on 8-subtetrahedron subdivision, *Mathematics of Computations*, **65**, 215, 1183-1200 (1996).
- [16] R. Löhner y J.D. Baum, Adaptive *h*-refinement on 3D unstructured grids for transient problems, *Int. J. Num. Meth. Fluids*, **14**, 1407-1419 (1992).

-
- [17] R. Löhner y P. Parikh, Three-dimensional grid generation by advancing front method, *Int. J. Num. Meth. Fluids*, **8**, 1135-1149 (1988).
- [18] R. Montenegro, A. Plaza, L. Ferragut y I. Asensio, Application of a nonlinear evolution model to fire propagation, *Nonlinear Analysis, Th., Meth. & App.*, **30**, 5, 2873-2882 (1997).
- [19] R. Montenegro, G. Montero, J.M. Escobar, E. Rodríguez y J.M. González-Yuste, Tetrahedral mesh generation for environmental problems over complex terrains, *Lecture Notes in Computer Science*, **2329**, 335-344 (2002).
- [20] R. Montenegro, G. Montero, J.M. Escobar y E. Rodríguez, Efficient strategies for adaptive 3-D mesh generation over complex orography, *Neural, Parallel & Scientific Computation*, **10**, 57-76 (2002).
- [21] G. Montero, R. Montenegro y J.M. Escobar, A 3-D diagnostic model for wind field adjustment, *J. of Wind Eng. and Ind. Aerodynamics*, **74-76**, 249-261 (1998).
- [22] M. Murphy, D.M. Mount y C.W. Gable, *A point-placement strategy for conforming Delaunay tetrahedralization*, en *Symposium on Discrete Algorithms*, (2000), pp. 67-74.
- [23] A. Plaza, R. Montenegro and L. Ferragut, An improved derefinement algorithm of nested meshes, *Advances in Engineering Software*, **27**, 1/2, 51-57 (1996).
- [24] A. Plaza y G.F. Carey, Local refinement of simplicial grids based on the skeleton, *Appl. Numer. Math.*, **32**, 195-218 (2000).
- [25] M.C. Rivara, A grid generator based on 4-triangles conforming. Mesh-refinement algorithms, *Int. J. Num. Meth. Eng.*, **24**, 1343-1354 (1987).
- [26] M.C. Rivara y C. Levin, A 3-d refinement algorithm suitable for adaptive multigrid techniques, *J. Comm. Appl. Numer. Meth.*, **8**, 281-290 (1992).
- [27] G. Winter, G. Montero, L. Ferragut y R. Montenegro, Adaptive strategies using standard and mixed finite elements for wind field adjustment, *Solar Energy*, **54**, 1, 49-56 (1995).

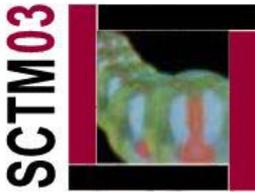
En Internet

<http://ceani.ulpgc.es/index.html>

CEANI-IUSIANI

Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería de la Universidad de Las Palmas de Gran Canaria.

Matemáticas y Visión por Ordenador



Luis Álvarez León

Catedrático de Ciencias de la Computación e Inteligencia Artificial
Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria

<http://serdis.dis.ulpgc.es/~lalvarez>

Introducción

En los últimos años, los enormes avances tecnológicos en el campo de los ordenadores, fotografía y video digital, etc., han permitido el desarrollo de nuevas disciplinas científicas como es el caso de la visión por ordenador. El sistema de visión humano es muy eficiente y nos suministra información útil y muy variada sobre nuestro entorno. Somos capaces, por ejemplo, de identificar y reconocer fácilmente objetos y formas, también podemos detectar y seguir con facilidad objetos en movimiento en nuestro campo de visión, gracias a la visión binocular podemos estimar la distancia que nos separa de los objetos presentes en nuestro entorno, etc.. Todas estas capacidades de la visión humana y muchas otras que no mencionamos aquí son extremadamente complejas de modelizar y formalizar desde el punto de vista matemático. La visión por ordenador es una disciplina científica de reciente desarrollo que estudia la modelización e implementación en ordenador de procesos propios de la visión. La visión por ordenador empezó siendo una disciplina esencialmente tecnológica donde se creía que todos los problemas se irían resolviendo fácilmente en base al incremento de potencia de cálculo de los ordenadores y las mejoras en la calidad de los dispositivos de adquisición como son las cámaras digitales. Pero pronto se descubrió que el tipo de problemas que aparecen son muy complejos y difíciles de analizar. De tal forma que para muchos de ellos no existe actualmente una solución plenamente satisfactoria. Esta complejidad de los problemas ha llevado a un esfuerzo científico muy importante en la disciplina donde las Matemáticas juegan un papel fundamental. De hecho, actualmente, el perfil de un investigador en la disciplina de visión por ordenador es un perfil mucho más científico que técnico y con una sólida formación matemática.

El objetivo de este seminario es presentar algunas reflexiones sobre la importancia del papel de las Matemáticas en el campo de la visión por ordenador. La exposición se organizará en base a diferentes disciplinas matemáticas y sus relaciones con diferentes problemas de visión por ordenador. Concretamente trataremos las siguientes materias:

- La Geometría Proyectiva.
- Las Transformadas Integrales.
- El Cálculo Variacional.
- Los Modelos Probabilísticos.
- Las Ecuaciones en Derivadas Parciales Geométricas.
- La Optimización y el Análisis Numérico.

Esta lista de disciplinas matemáticas no pretende ser exhaustiva y sólo representa algunos temas con los que quien suscribe está familiarizado. La organización del seminario será la siguiente: iremos visitando cada una de las disciplinas matemáticas mencionadas arriba y veremos algunas de sus aplicaciones en el campo de la visión por ordenador.

Respecto a la bibliografía descrita al final del texto, he optado por citar unas pocas referencias [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11] que, aunque no cubren por completo el espectro de conocimientos presentado en esta memoria (lo que obligaría a llenar varias páginas de referencias), ilustran bastante bien algunas de las materias presentadas.

1. La geometría proyectiva

El modelo de cámara proyectivo es el más sencillo y el que más se utiliza habitualmente. Viene definido por un plano de proyección que representa el plano imagen y un foco. La proyección de un punto 3D en la imagen viene dada por la intersección entre la recta que une dicho punto y el foco, y el plano de proyección. En la Figura 1, podemos observar una ilustración de cómo se aplica este modelo en pintura. El pintor mira a través de un foco una escena 3D y ha intercalado entre él y la escena una cuadrícula que representa el plano de proyección. Apoyándose en la proyección de la escena 3D sobre esta cuadrícula va pintando la escena.



Figura 1. Ilustración de Albrecht Dürer, pintor alemán del Renacimiento (año 1527).

En principio, las coordenadas en las que se representan los puntos son coordenadas euclídeas, esto es, 3 coordenadas para un punto 3D y 2 coordenadas para un punto en el plano. Sin embargo, cuando buscamos la expresión matemática que determina cómo se proyecta un punto 3D en el plano, resulta mucho más conveniente trabajar en los correspondientes espacios proyectivos, de tal manera que un punto 3D se identifica con 4 coordenadas $\mathbf{X} = (x, y, z, v)$; cuando v es distinto de 0, recuperamos las coordenadas euclídeas del punto haciendo $(x/v, y/v, z/v)$. Trabajar en el espacio proyectivo nos permite manejar correctamente puntos en el infinito ($v = 0$). Además, la aplicación que determina cómo se proyecta un punto 3D en el plano, que en coordenadas euclídeas es no-lineal, en coordenadas proyectivas es lineal y viene dada por una matriz $P \in M_{3 \times 4}$. De tal manera que dado un punto 3D \mathbf{X} en coordenadas proyectivas, su correspondiente proyección \mathbf{x} en el plano viene dada por la expresión

$$\mathbf{x} = P\mathbf{X}$$

La matriz P depende de la posición y orientación del plano de proyección en el espacio 3D y del sistema de referencia elegido en el plano de proyección. Un problema fundamental que surge con mucha frecuencia en visión por ordenador es el de calibrar una cámara, que consiste

básicamente en encontrar la matriz P . La primera pregunta que surge es si cualquier matriz P no nula determina una proyección admisible en el sentido de que corresponda físicamente a una configuración real del modelo de proyección. La respuesta a esta pregunta es negativa. Para que P determine una proyección admisible es necesario que sus coeficientes verifiquen una cierta relación algebraica. El estudio de las relaciones algebraicas que aparecen de forma natural al plantear el problema de calibración de cámaras involucra técnicas de álgebra computacional para su resolución.

Existen muchas formas distintas de plantear el problema de calibración en función del tipo de información y número de cámaras que queramos calibrar. La manera más simple es el uso de un calibrador. Un calibrador es un objeto como el representado en la imagen de la Figura 2, del cual conocemos con exactitud las coordenadas 3D de algunos puntos. Por ejemplo, en este caso, conocemos (porque lo hemos medido físicamente) las coordenadas 3D de los puntos que corresponden a las esquinas de los rectángulos negros. Por otro lado, para cada esquina 3D X_i calculamos en la imagen las coordenadas 2D (en pixels) de su proyección x_i en el plano imagen. A partir de las correspondencias $X_i \longleftrightarrow x_i$ es posible calcular P resolviendo un cierto sistema de ecuaciones lineales.

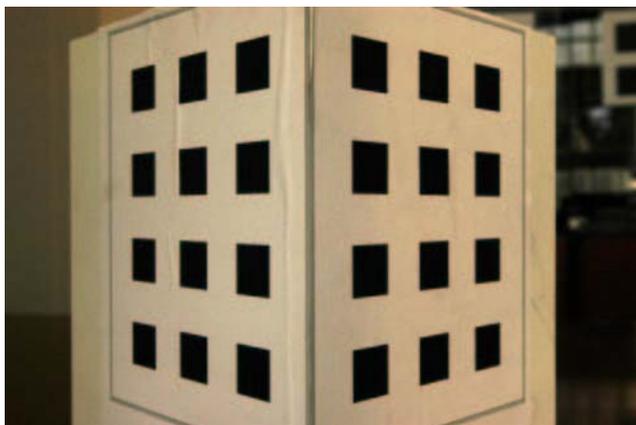


Figura 2. Imagen de un calibrador.

En muchas ocasiones no es posible la utilización de calibradores para calibrar las cámaras. En general, el estudio del problema de la calibración de un sistema de varias cámaras que están observando una misma escena 3D requiere de un fuerte aparato matemático basado principalmente en la geometría y el álgebra. Del hecho de que las cámaras están observando una misma escena se pueden derivar muchos tipos de relaciones, en su mayoría algebraicas entre las matrices de proyección de las diferentes cámaras.

2. Las transformadas integrales

Las transformadas integrales son una herramienta muy poderosa que posee múltiples utilidades en visión por ordenador. En esta sección veremos algunas aplicaciones concretas relacionadas con el análisis multiescala lineal, el reconocimiento de formas planas y la compresión de imágenes.

2.1. El análisis multiescala lineal. La transformada integral por excelencia viene dada por la convolución de una función de entrada con un cierto núcleo de convolución. En nuestro caso, la función de entrada será una imagen bidimensional que formalmente definimos como una aplicación $I : \mathbb{R}^2 \rightarrow \mathbb{R}$, donde $I(x,y)$ representa el nivel de gris de la imagen en dicho punto. Por simplicidad en la exposición supondremos que el dominio donde está definida la imagen es todo \mathbb{R}^2 y que la imagen es de niveles de gris (y no de color, en cuyo caso tendría 3 canales). En una misma imagen coexisten informaciones a diferentes escalas sobre la escena que estamos observando. Por ejemplo, en una panorámica de un bosque podemos observar el bosque en su conjunto, los árboles individuales, e incluso, si la resolución de la imagen es suficientemente buena, los detalles de las hojas de los árboles. La existencia de una información tan variada y compleja en una imagen hace muy complicado su análisis. Para intentar simplificar la información presente en una imagen y poder analizarla más fácilmente, una herramienta muy utilizada consiste en extraer de la imagen original una secuencia de imágenes, que representan la imagen a escalas diferentes. La manera más sencilla de hacer esto es convolucionar la imagen original con un núcleo de convolución gaussiano, es decir, construimos la secuencia

$$I_\sigma(x, y) = \int_{\mathbb{R}^2} G_\sigma(x - u, y - v) I(u, v) du dv$$

donde

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Aquí, σ representa la escala del análisis y cuanto mayor sea, mayor simplificación se produce en la imagen original.

2.2. El reconocimiento de formas planas. La transformada de Fourier es una herramienta muy útil en Matemáticas, tanto desde el punto de vista teórico como práctico. Desgraciadamente, en general, en las Facultades de Matemáticas no se dedica en los planes de estudio el tiempo suficiente para enseñar correctamente el interés y alcance de esta herramienta, o, al menos, esto era así cuando yo estudié. La transformada de Fourier de una función $f(x)$ en el intervalo $[a, b]$ viene dada por la expresión

$$\hat{f}(w) = \int_a^b f(x) e^{-iwx} dx.$$

A continuación veremos una sencilla aplicación de la transformada de Fourier a la descripción y reconocimiento de formas planas. Consideraremos formas sencillas definidas por un conjunto en el plano cuya frontera es una única curva cerrada como se muestra en la Figura 3. Las formas de este tipo vienen caracterizadas por la geometría de su contorno. Con objeto de

utilizar una representación del contorno adecuada para comparar contornos entre sí, vamos a asociar a la curva cerrada que delimita el contorno la función unidimensional $\theta(s)$ que para cada punto de la curva determina su orientación, es decir, el ángulo que forma la normal a la curva en dicho punto con el eje x ; en la Figura 3 puede verse tal representación. Por tanto, para comparar dos formas lo que haremos será comparar las funciones de orientación asociadas.

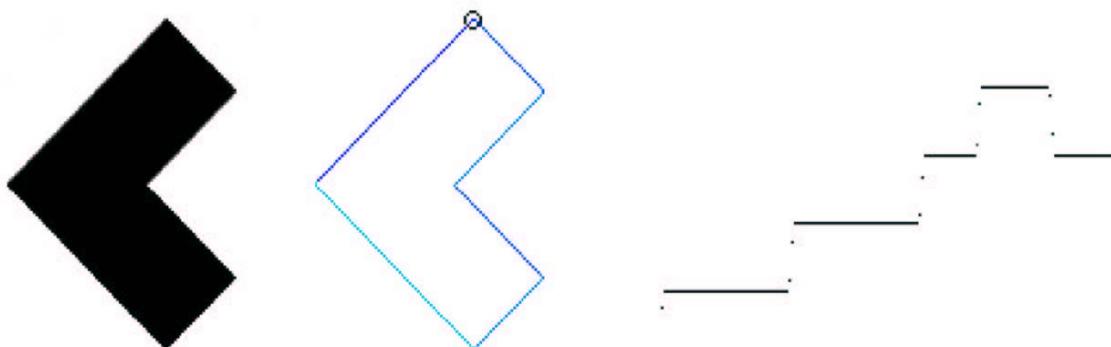


Figura 3. Ilustración de una forma y la función de orientación de su contorno tomando como punto inicial la esquina marcada.

Ahora bien, queremos que dicha comparación sea independiente del tamaño, posición y orientación de los objetos y por supuesto, independiente del punto inicial que se ha tomado en la curva para obtener su función de orientación. Dado que estamos comparando funciones periódicas, la transformada de Fourier resulta una herramienta idónea y realizamos la comparación de las funciones de orientación de 2 formas utilizando su representación en el espacio de Fourier. Comparando los coeficientes de Fourier podemos asociar a 2 formas una medida de similitud entre ellas, de tal forma que cuanto menor sea esta medida mayor parecido habrá entre las formas. En la Figura 4 se ilustra esta técnica.

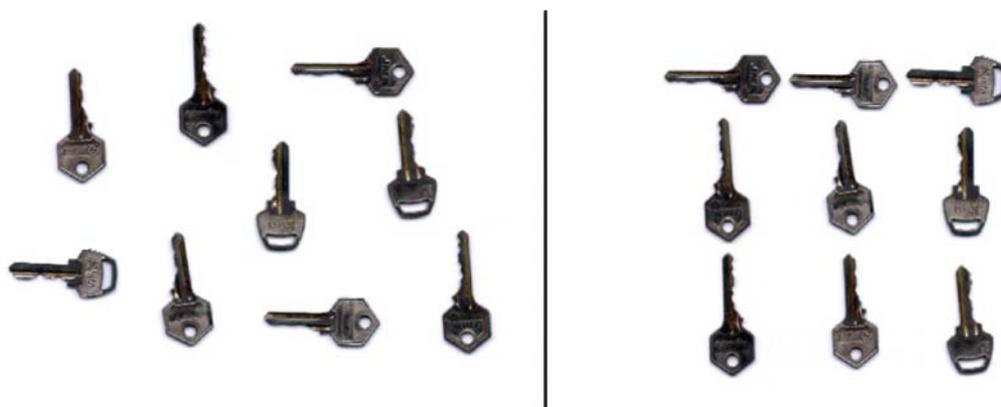


Figura 4. Reconocimiento de formas: a la izquierda observamos un grupo de llaves, a la derecha su clasificación por criterios de similitud. Las llaves idénticas se encuentran ordenadas por columnas.

2.3. La compresión de imágenes. La transformada de Fourier es una herramienta muy útil cuando analizamos fenómenos ondulatorios y periódicos, pero no es adecuada para analizar todo tipo de funciones debido principalmente a que en el espacio de Fourier se pierde totalmente la información espacial de la señal, en el sentido de que la información suministrada por la transformada de Fourier es global, y es imposible extraer información del carácter local sobre el comportamiento de la función. En los últimos años se ha desarrollado la teoría del análisis multiresolución y las *ondelettes* que permiten una representación muy interesante de una función, incluyendo la localización tanto en espacio como en frecuencia. Un análisis multiresolución viene dado por 2 funciones $\psi(s)$ (la ondelette madre) y $\phi(s)$ (la ondelette padre), tales que el conjunto

$$\psi_{m,n}(s) = 2^{m/2} \psi(2^m(s - 2^{-m}n)) \quad m, n \in \mathbb{Z}$$

representa una base ortonormal de $L^2(\mathbb{R})$; y para todo $m \in \mathbb{Z}$ el conjunto de funciones $\{\psi_{m,n}(s)\}_{n \in \mathbb{Z}} \cup \{\phi_{m,n}(s)\}_{n \in \mathbb{Z}}$ es una base ortonormal del espacio $\{\phi_{m+1,n}(s)\}_{n \in \mathbb{Z}}$.

Los denominados coeficientes de ondelettes de una función $f(\cdot)$ vienen dados por:

$$f_\psi(m, n) = \int_{-\infty}^{\infty} \psi_{m,n}(s) f(s) ds, \quad f_\phi(m, n) = \int_{-\infty}^{\infty} \phi_{m,n}(s) f(s) ds.$$

El nombre de análisis multiresolución viene del hecho de que podemos considerar los coeficientes $\{f_\phi(m, n)\}_{n \in \mathbb{Z}}$ como el análisis de la función $f(\cdot)$ a la resolución m , y $\{f_\psi(m, n)\}_{n \in \mathbb{Z}}$ representa la información de detalles necesaria para pasar de la resolución m a la resolución $m+1$. En dimensión 2, las ondelettes de base se definen como:

$$\begin{aligned} \Psi_1^{m,n,k,l}(x, y) &= 2^{m/2} \phi(2^m x - n) 2^{k/2} \phi(2^k y - l) \\ \Psi_2^{m,n,k,l}(x, y) &= 2^{m/2} \phi(2^m x - n) 2^{k/2} \psi(2^k y - l) \\ \Psi_3^{m,n,k,l}(x, y) &= 2^{m/2} \psi(2^m x - n) 2^{k/2} \phi(2^k y - l) \\ \Psi_4^{m,n,k,l}(x, y) &= 2^{m/2} \psi(2^m x - n) 2^{k/2} \psi(2^k y - l) \end{aligned}$$

Si consideramos que una imagen discreta $I(n, l)$ representa los coeficientes de ondelettes de una función $I(x, y)$ para la ondelette $\Psi_1^{1,n,1,l}(x, y)$ a partir de $I(n, l)$ podemos descomponer la imagen calculando los coeficientes para las 4 ondelettes $\Psi_i^{0,n,0,l}(x, y)$. En la Figura 5 se ilustra el resultado de tal descomposición. Como puede observarse, muchos de los coeficientes de ondelettes son nulos o muy pequeños (el tono de gris blanco representa un valor nulo), por tanto si sólo almacenamos los coeficientes de ondelettes no nulos podemos ahorrar memoria al almacenar la imagen.

3. El cálculo variacional

La modelización de la solución de un problema como el mínimo de un cierto funcional de energía es una herramienta de gran utilidad en el campo de la visión por ordenador. En esta sección vamos a ver algunas aplicaciones de esta técnica a problemas como la eliminación de ruido en una imagen, el seguimiento de objetos en una secuencia video o la reconstrucción 3D de objetos a partir de un par estéreo de imágenes.

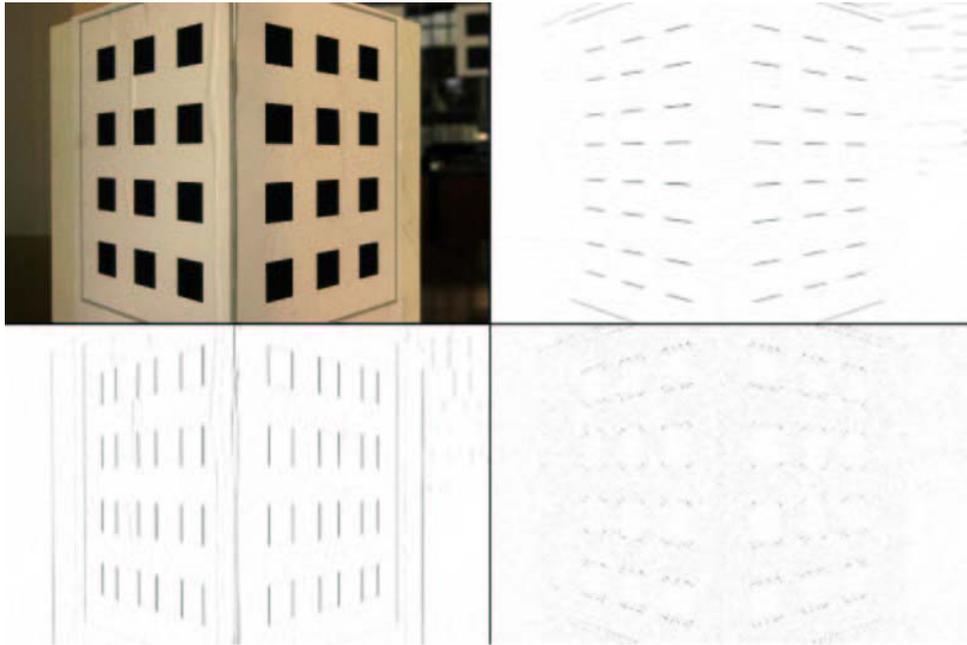


Figura 5. Descomposición por ondelettes de la imagen del calibrador.

3.1. La eliminación de ruido en una imagen. Normalmente, existe siempre un cierto ruido en las imágenes producido por los dispositivos de adquisición, manipulaciones posteriores, etc. El ruido en una imagen se presenta habitualmente en forma de irregularidades de la función imagen. Formularemos el problema de la eliminación de ruido de la siguiente forma: Dada una imagen $I(x,y)$, pretendemos encontrar otra imagen $I'(x,y)$ tal que por un lado sea una función más regular que $I(x,y)$, y por otro lado se parezca a $I(x,y)$ lo más posible. Existen muchas formas de expresar matemáticamente estas condiciones. Una forma sencilla consiste en buscar $I'(x,y)$ minimizando el funcional

$$E(I') = \int_{\mathbb{R}^2} (I' - I)^2 + \alpha \int_{\mathbb{R}^2} \Phi(\|\nabla I'\|)^2$$

donde $\Phi(\cdot)$ es una función positiva y α representa el balance entre el término que ajusta I a I' y el término que suaviza la función I' minimizando globalmente su gradiente, módulo la función $\Phi(\cdot)$.

Para calcular los posibles mínimos de esta energía se iguala a cero la derivada del funcional de energía, lo que da lugar a la denominada ecuación en derivadas parciales de Euler-Lagrange del funcional. A título de ejemplo, en el caso del funcional anterior, la ecuación de Euler-Lagrange resultante es

$$(I' - I) - \alpha \operatorname{div} \left(\frac{\Phi'(\|\nabla I'\|)}{\|\nabla I'\|} \nabla I' \right) = 0.$$

Por tanto, observamos cómo la minimización del funcional se reduce a resolver una ecuación en derivadas parciales. En la Figura 6 se ilustra el resultado de aplicar esta técnica.

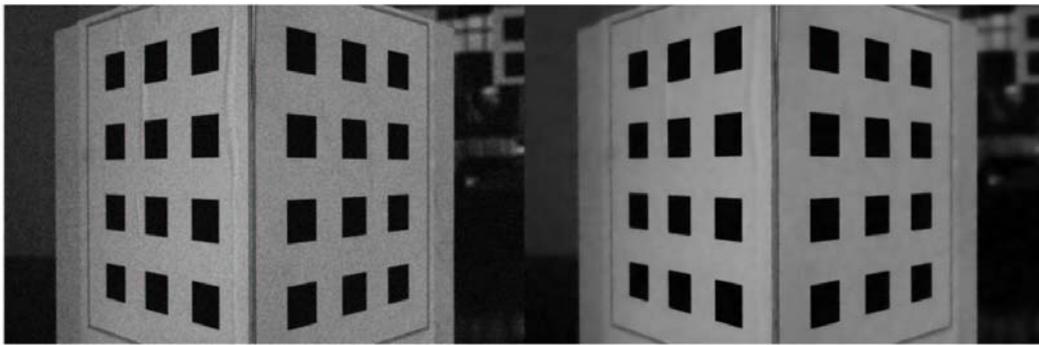


Figura 6. A la izquierda, imagen del calibrador a la que hemos añadido ruido artificialmente; a la derecha, la imagen restaurada.

3.2. Seguimiento de objetos en una secuencia video. Cuando filmamos una secuencia de imágenes, se producen movimientos debidos al propio movimiento de la cámara o al desplazamiento de objetos en la escena. Dadas 2 imágenes consecutivas de la secuencia que denotaremos por $I_1(x,y)$ e $I_2(x,y)$ consideramos el problema de encontrar una función $h(x,y) = (u(x,y), v(x,y))$ denominada flujo, que determina el movimiento de cada punto entre la imagen I_1 e I_2 . Utilizaremos como hipótesis de base que el nivel de gris de un punto no varía cuando el punto cambia de posición, es decir: $I_1(x,y) = I_2((x,y) + h(x,y))$. Ahora bien, esta hipótesis no es suficiente para definir $h(x,y)$, pues, en general, para cada punto en I_1 , hay múltiples combinaciones de puntos en I_2 que verifican la hipótesis. Para poder resolver el problema, añadiremos como condición que el flujo $h(x,y)$ sea regular. Ello nos lleva a minimizar el funcional

$$E(h) = \int_{\mathbb{R}^2} (I_1(x,y) - I_2((x,y) + h(x,y)))^2 + \alpha \int_{\mathbb{R}^2} \operatorname{traza}(\nabla h^T D(\nabla I_1) \nabla h).$$

El término de regularización es un poco especial, porque regulariza el flujo h pero permitiendo discontinuidades en los bordes de los objetos de I_1 . En la Figura 7 se ilustran algunos resultados obtenidos con este tipo de técnicas.



Figura 7. En la parte inferior observamos 2 frames de una secuencia video. En la parte superior representamos las dos componentes del desplazamiento encontrado. El nivel de gris medio indica desplazamiento 0. Un tono claro indica desplazamiento en dirección positiva, y un tono oscuro desplazamiento en dirección negativa.

3.3. Reconstrucción 3D de objetos a partir de un par estéreo de imágenes. Una vez calibrado un sistema de cámaras, el principal problema a resolver es la puesta en correspondencia de puntos en ambas imágenes. El problema se puede formular en los mismos términos que el cálculo del flujo en una secuencia video; la única diferencia es que podemos reducir la complejidad del problema utilizando la información suministrada por la denominada Geometría Epipolar. En un par estéreo, dado un punto en una imagen, su correspondiente en la otra imagen debe estar sobre una recta denominada recta epipolar. Si las cámaras están calibradas, la recta epipolar que corresponde a cada punto es conocida y por tanto ello simplifica considerablemente la complejidad del problema. En la Figura 8 se muestra un resultado de aplicar este tipo de técnicas a la reconstrucción 3D.

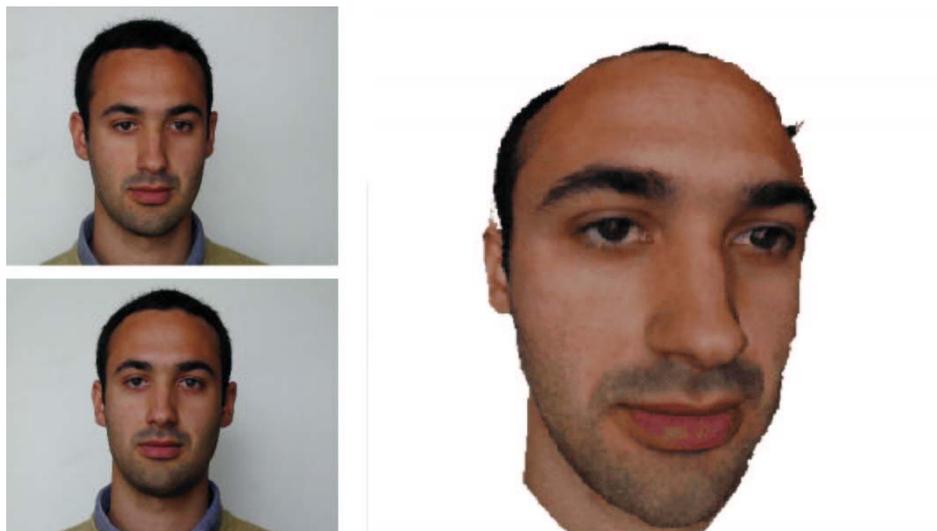


Figura 8. A la izquierda, un par estéreo de una cara humana; a la derecha, la reconstrucción 3D obtenida.

4. Los modelos probabilísticos

Sin duda, los modelos probabilísticos son de una gran utilidad en visión por ordenador y existe toda una comunidad de investigadores en visión que utilizan como base dichos modelos. Una imagen puede modelizarse como la realización de una cierta variable aleatoria, y muchos filtros y transformaciones en imágenes pueden modelizarse como procesos estocásticos. En esta sección vamos a ilustrar cómo se utilizan los modelos probabilísticos para modelizar la formación de escenas naturales.

En concreto veremos el denominado “modelo de hojas muertas” que modeliza la estructura resultante de la superposición de objetos que se acumulan de forma aleatoria; por ejemplo, y de ahí su nombre, la estructura de hojas caídas en un bosque en otoño. La estructura del objeto que va cayendo (la hoja) se modeliza a través de los denominados conjuntos aleatorios. El objetivo que se plantea es estudiar las características geométricas de las partes visibles (denominadas celdas) de la estructura resultante después de superponer aleatoriamente las hojas. Existen resultados realmente interesantes como que la longitud media de los segmentos resultantes de intersectar una recta con las celdas es exactamente la mitad de la longitud media de los segmentos resultantes de interceptar las hojas originales. En la Figura 9 se ilustra un modelo de hojas muertas donde la hoja esta compuesta por rectángulos de tamaño y orientación uniformemente distribuidos.

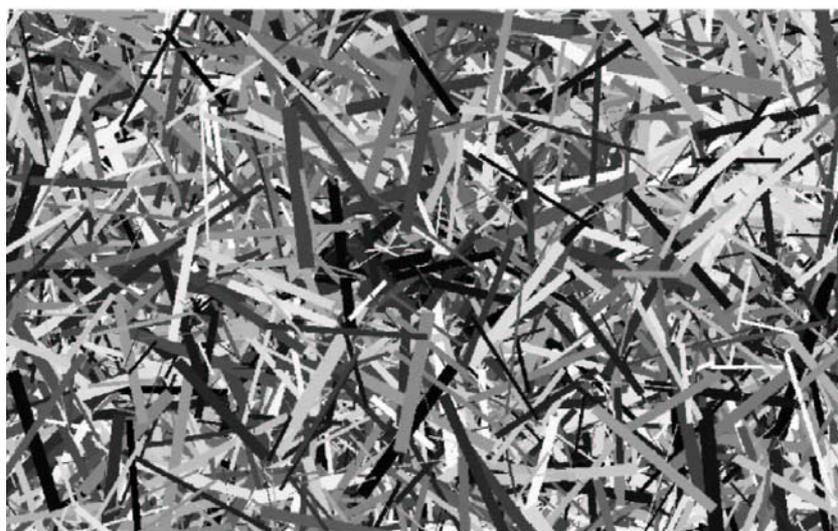


Figura 9. Ilustración de un modelo de hojas muertas.

5. Las ecuaciones en derivadas parciales geométricas

Una de las características más importantes de nuestro sistema de visión es que la percepción que tenemos de nuestro entorno es invariante respecto a múltiples transformaciones. Somos capaces de reconocer el mismo objeto cuando cambiamos la luminosidad, su orientación posición o tamaño. Intentar que los modelos de visión artificiales posean también de alguna forma este tipo de invariantes es algo realmente complicado. Por ejemplo, si buscamos filtros

que sean invariantes por el tipo de transformaciones mencionadas nos encontramos con que dichos filtros están asociados a un tipo muy especial de ecuaciones en derivadas parciales no lineales que tienen la expresión general

$$\frac{\partial G}{\partial t} = G \left(\operatorname{div} \left(\frac{\nabla u}{\|\nabla u\|} \right), t \right) \|\nabla u\|,$$

donde $G(s,t)$ es una función creciente respecto a su primer argumento. Este tipo de ecuaciones se denominan geométricas porque su evolución sólo depende de la geometría de la frontera de los conjuntos de nivel de u .

6. La optimización y el análisis numérico

La mayor parte de los modelos formulados para analizar los diferentes problemas que aparecen en la visión por ordenador requieren una fase de validación a través de su implementación en ordenador. Para realizar esta implementación es necesario hacer un uso intensivo de técnicas de optimización y análisis numérico. En concreto, resulta necesario con frecuencia abordar los siguientes problemas:

- Interpolación de funciones.
- Resolución de grandes sistemas de ecuaciones lineales.
- Cálculo de autovalores y autovectores.
- Descomposición SVD de matrices.
- Análisis numérico de ecuaciones en derivadas parciales.
- Estimación robusta de parámetros.
- Ajuste de parámetros utilizando Levenberg-Marquardt.

A título de conclusión de este seminario, podemos constatar que la investigación en visión por ordenador requiere de una sólida formación matemática y que el número de problemas abiertos es muy amplio. Además, dado el interés tecnológico de este campo es de prever en el futuro un esfuerzo considerable de las administraciones públicas y empresas privadas en términos de financiación de la investigación. Un joven licenciado en matemáticas con cierta sensibilidad hacia los problemas aplicados y la experimentación es un perfil de investigador idóneo en visión por ordenador. Para iniciarse como investigador en este campo lo más adecuado es seguir un programa de cursos de doctorado específico. En este sentido, los mejores programas se imparten en el extranjero. Por ejemplo, en Francia, que es el país que mejor conozco, los cursos de doctorado se hacen todos en un único año, están bien organizados y un alumno puede dedicarse en exclusiva a seguir los cursos sin aburrirse; una vez terminados los cursos de doctorado el alumno decide si quiere continuar con la tesis o no. En España, los cursos de doctorado se imparten por bienios, son mucho más dispersos en el tiempo, no están diseñados para que el alumno se dedique en exclusiva a ellos, y la realización de la tesis se empieza a hacer muchas veces en paralelo con los cursos.

Por último, me gustaría terminar esta memoria con una invitación a todos aquellos que tuvieran interés en iniciarse en este campo, en especial a jóvenes licenciados o alumnos de último curso de Matemáticas, a contactar conmigo, y gustosamente, dentro de mis limitados conocimientos, les intentaría orientar sobre los pasos a seguir.

Bibliografía

- L. Alvarez, R. Deriche, J. Sánchez, J. Weickert: Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation* **13** (2002), 3-21.
- L. Alvarez, F. Guichard, P.L. Lions, J.M. Morel: Axioms and fundamental equations of image processing. *Arch. Rat. Mech. Anal.* **123** (1993), 199-257.
- G. Aubert, P. Kornprobst: *Mathematical Problem in Image Processing*. Springer, 2002.
- O. Faugeras, Q.T. Long, T. Papadopoulos: *The Geometry of Multiple Images*. MIT Press, 2001.
- O. Faugeras, Q.T. Long, T. Papadopoulos: *Three-Dimensional Computer Vision, a Geometric Viewpoint*. MIT Press, 2001.
- R. Hartley, A. Zisserman: *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.
- T. Lindeberg: *Scale Space Theory in Computer Vision*. Kluwer, 1994.
- Y. Meyer: *Ondelettes et Operateurs I*. Hermann, Paris, 1990.
- J.M. Morel, S. Solimini: *Variational Methods in Image Segmentation*. Birkhauser, 1995.
- Bart M.ter Haar Romeny (editor): *Geometry-Driven Diffusion in Computer Vision*. Kluwer, 1994.
- J. Weickert: *Anisotropic Diffusion in Image Processing*. Teubner, Stuttgart, 1998.

En Internet

- <http://serdis.dis.ulpgc.es/~lalvarez/ami/index.html>
Grupo de Análisis Matemático de Imágenes
Universidad de Las Palmas de Gran Canaria.
- <http://www.cmla.ens-cachan.fr/Cmla/index.html>
CMLA
Centre de Mathématiques et de Leurs Applications.



Las Matemáticas y La Visión por Ordenador

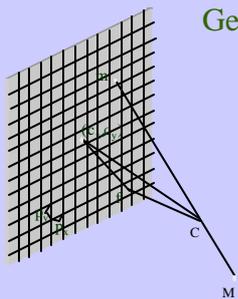
Luis Álvarez León

Departamento de Informática y Sistemas
Universidad de Las Palmas de Gran Canaria

Contenido

- Geometría proyectiva
- Las transformadas integrales
- El cálculo variacional
- Los modelos probabilísticos
- Ecuaciones en derivadas parciales geométricas
- La optimización y el análisis numérico

Geometría Proyectiva



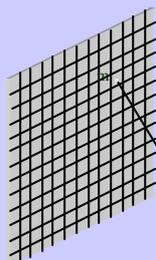
Parámetros Intrínsecos

(c_x, c_y) Proyección ortogonal del foco C en la imagen

$$a_x = \frac{p_x}{f} = \frac{\text{tamaño horizontal pixel}}{\text{distancia focal}}$$

$$a_y = \frac{p_y}{f} = \frac{\text{tamaño vertical pixel}}{\text{distancia focal}}$$

Parámetros Extrínsecos



$t = (t_x, t_y, t_z)$ vector \overline{CO} de traslación entre los sistemas de referencia

$R = (r_{ij})$ matriz de rotación entre los sistemas de referencia

Sistema de Referencia Universal

Ilustración Pintor Alemán del Renacimiento (Albrecht Dürer, 1527)



Los espacios proyectivos P^4 y P^3

$$M = (M_1, M_2, M_3) \in R^3 \rightarrow \tilde{M} = (M_1, M_2, M_3, 1) \in P^4$$

$$m = (m_1, m_2) \in R^2 \rightarrow \tilde{m} = (m_1, m_2, 1) \in P^3$$

Puntos en el infinito

$$\tilde{M} = (M_1, M_2, M_3, 0) \in P^4$$

Intersección rectas paralelas
(vector director recta)

Matriz de proyección P asociada a una cámara

Teorema (Faugeras). Sea $\tilde{M}=(M_1, M_2, M_3, M_4) \in P^4$ y $\tilde{m}=(m_1, m_2, m_3) \in P^3$ su proyección en la imagen, entonces:

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \underbrace{\begin{pmatrix} \mathbf{a}_x r_1 + c_x r_3 & \mathbf{a}_x t_x + c_x t_z \\ \mathbf{a}_y r_2 + c_y r_3 & \mathbf{a}_y t_y + c_y t_z \\ r_3 & t_z \end{pmatrix}}_P \begin{pmatrix} M_1 \\ M_2 \\ M_3 \\ M_4 \end{pmatrix}$$

Donde r_1, r_2 y r_3 son los vectores columna de la matriz R.

Teorema (Caracterización de P): P se puede expresar en función de un conjunto de parámetros intrínsecos y extrínsecos si sólo si

$$\begin{cases} p_{31}^2 + p_{32}^2 + p_{33}^2 = 1 \\ \left(\begin{pmatrix} p_{11} \\ p_{12} \\ p_{13} \end{pmatrix} \wedge \begin{pmatrix} p_{31} \\ p_{32} \\ p_{33} \end{pmatrix} \right) \cdot \left(\begin{pmatrix} p_{21} \\ p_{22} \\ p_{23} \end{pmatrix} \wedge \begin{pmatrix} p_{31} \\ p_{32} \\ p_{33} \end{pmatrix} \right) = 0 \end{cases}$$

Información Euclídea a partir de P

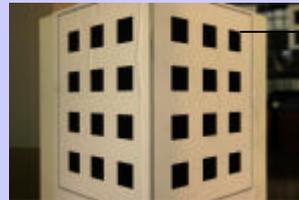
Foco $\tilde{C}=(C_1, C_2, C_3, 1) \rightarrow P\tilde{C}=0$

vector director recta $V=mC$

$$\hat{V}=(V_1, V_2, V_3, 0) \rightarrow P\hat{V}=\tilde{m}$$

Utilización de un calibrador para encontrar los pares

$$(\tilde{M}, \tilde{m})$$



$$\begin{cases} \tilde{M}=(-12.4, 0.8, 4) \text{ en cm} \\ \tilde{m}=(180, 20) \text{ en pixels} \end{cases}$$

Cada par (\tilde{M}, \tilde{m}) determina 2 ecuaciones lineales homogéneas en los coeficientes de P.

Problema de Calibración Fuerte: Encontrar P a partir de un conjunto N de correspondencias $(\tilde{M}_i, \tilde{m}_i)$

Si $N > 5$ obtenemos $2N > 11$ ecuaciones lineales en los coeficientes de P. Es decir $AP=0$ donde A es una matriz de $(2N \times 12)$.

$$P = \min_{\|P\|=1} \|AQ\|^2 = \text{auvector de } A^T A \text{ del autovalor más pequeño.}$$

Si $N=5$ obtenemos 10 ecuaciones lineales en los coeficientes de P, el núcleo de A tiene dimensión > 1 y el método anterior no funciona.

Transformadas Integrales. Convolución con una Gaussiana.

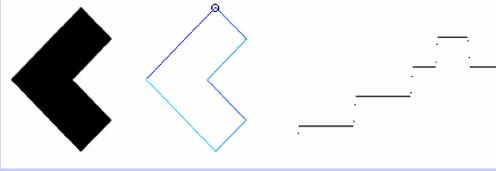
$$I_s(x, y) = \frac{1}{2\pi s^2} \int e^{-\frac{(x-u)^2 + (y-v)^2}{2s^2}} I(u, v) dudv$$



Transformadas Integrales . Transformada de Fourier.

$$\hat{f}(w) = \int_a^b f(x)e^{-iwx} dx$$

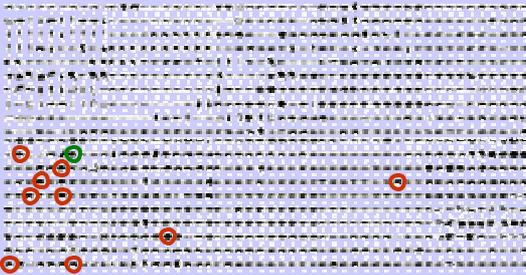
Reconocimiento de formas planas



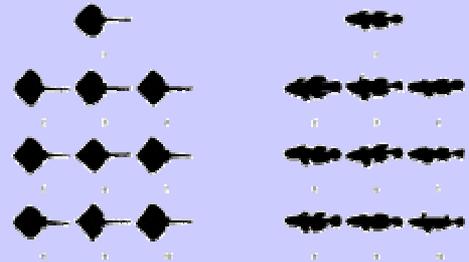
Clasificación de objetos



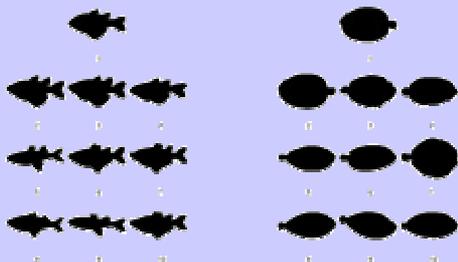
Clasificación de objetos



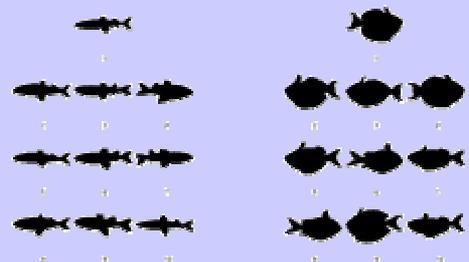
Clasificación de objetos



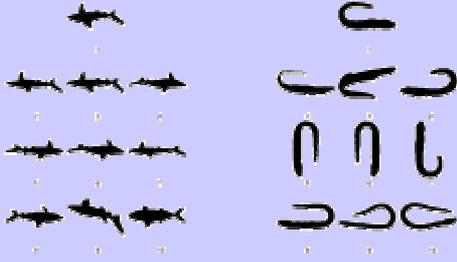
Clasificación de objetos



Clasificación de objetos



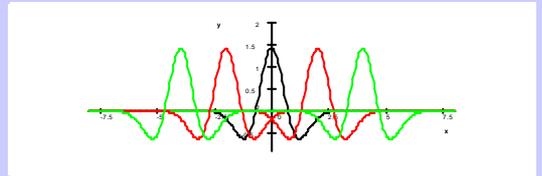
Clasificación de objetos



Transformadas Integrales . Transformada en Ondelettes

$$y_{m,n}(s) = 2^{\frac{m}{2}} y(2^m(s - 2^{-m}n)) \quad m, n \in \mathbb{Z} \quad \text{base } L^2(\mathbb{R})$$

m=01



Análisis MultiResolución .

$$f_y(m,n) = \int f(s) y_{m,n}(s) ds \quad f_f(m,n) = \int f(s) f_{m,n}(s) ds$$

$$\{y_{m,n}(s)\}_{n \in \mathbb{Z}} \cup \{f_{m,n}(s)\}_{n \in \mathbb{Z}} \quad \text{base de } \{f_{m+1,n}(s)\}_{n \in \mathbb{Z}}$$

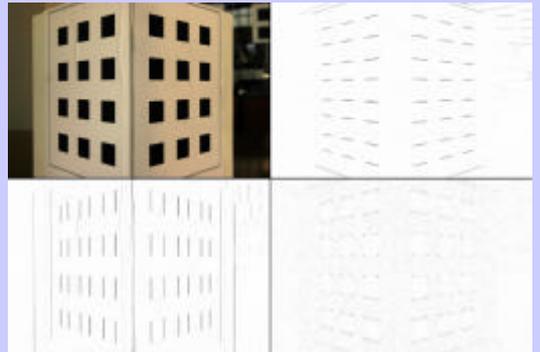
Dimensión 2

$$\left\{ \begin{array}{l} \{y_{m,n}(x) f_{m,l}(y)\} \quad \{f_{m,n}(x) f_{m,l}(y)\} \\ \{y_{m,n}(x) y_{m,l}(y)\} \quad \{f_{m,n}(x) y_{m,l}(y)\} \end{array} \right\}_{n,l \in \mathbb{Z}}$$

base de

$$\{f_{m+1,n}(x) f_{m+1,l}(y)\}_{n,l \in \mathbb{Z}}$$

Codificación de Imágenes



Cálculo Variacional. Eliminación de ruido

$$E(I') = \int (I(x) - I'(x))^2 + \alpha \int \Phi(\|\nabla I'\|)$$

Derivada del funcional $E(I')$

$$\lim_{t \rightarrow 0} \frac{E(I'+tI'') - E(I')}{t} = \langle \nabla E(I'), I'' \rangle = \int \nabla E(I')(\bar{x}) I''(\bar{x})$$

Ecuaciones de Euler-Lagrange

$$\nabla E(I') = -\text{div} \left(\frac{\Phi'(\|\nabla I'\|)}{\|\nabla I'\|} \nabla I' \right) + (I - I') = 0$$

Eliminación de Ruido

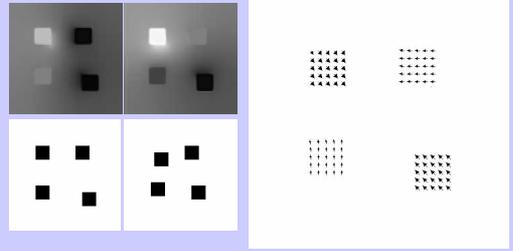


Cálculo Flujo Óptico

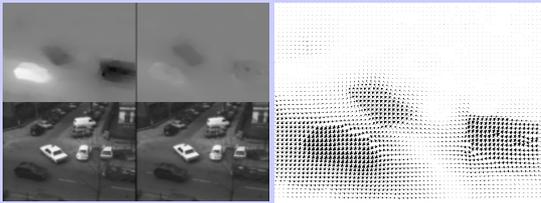


$$E(\vec{h}) = \int_{\Omega} (I_1(x) - I_2(x + \vec{h}))^2 + \alpha \int_{\Omega} \Phi(\nabla \vec{h}) \cdot d\omega$$

Flujo óptico



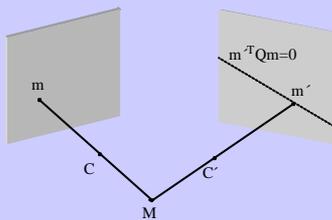
Cálculo flujo óptico



Visión Stereo



Geometría Epipolar



$$Q = \begin{pmatrix} r_{13}f_x - r_{12}f_z & r_{11}f_x - r_{13}f_z & r_{12}f_x - r_{11}f_z \\ r_{23}f_x - r_{22}f_z & r_{21}f_x - r_{23}f_z & r_{22}f_x - r_{21}f_z \\ r_{33}f_x - r_{32}f_z & r_{31}f_x - r_{33}f_z & r_{32}f_x - r_{31}f_z \end{pmatrix}$$

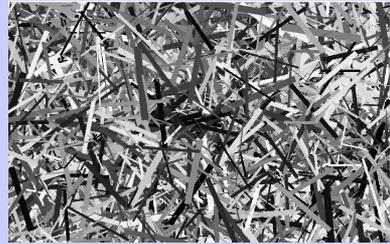
Reconstrucción de la Geometría 3D



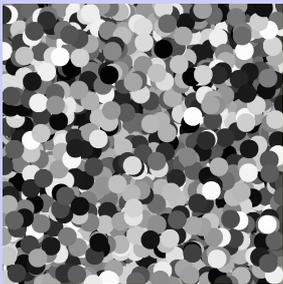
Reconstrucción de la Geometría 3D



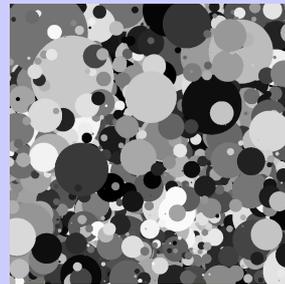
Modelos Probabilísticos. El modelo de hojas muertas



Modelos Probabilísticos. El modelo de hojas muertas



Modelos Probabilísticos. El modelo de hojas muertas



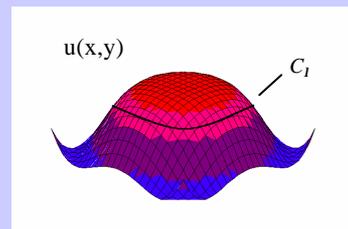
Modelos Probabilísticos. El modelo de hojas muertas

Longitud media segmentos partes visible de las celdas

||

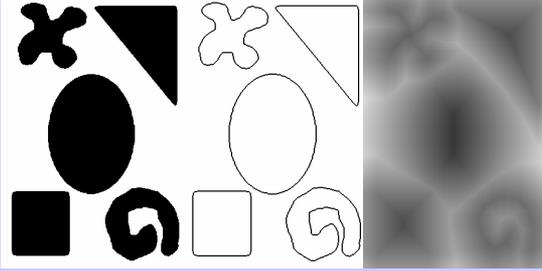
$0.5(\text{Longitud media segmentos grano})$

Ecuaciones diferenciales geométricas. Lineas de nivel



línea de nivel $C_1 \stackrel{\text{def}}{=} \partial\{(x, y) : u(x, y) > I\}$

Lineas de Nivel



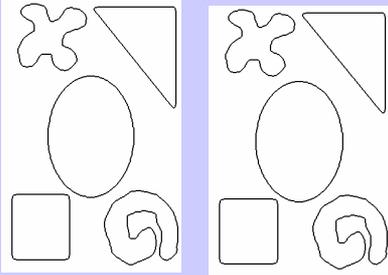
Ecuaciones en derivadas parciales geométricas

$$\frac{\partial u}{\partial t} = G\left(t, \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)\right) \|\nabla u\|$$

$$\operatorname{curv}(u) = \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right)$$

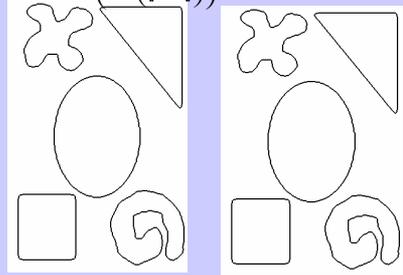
Ecuación de la Curvatura Media

$$\frac{\partial u}{\partial t} = \|\nabla u\| \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \leftrightarrow \frac{\partial C_L}{\partial t} = -k\bar{N}$$



Ecuación Invariante Afin

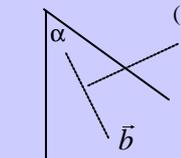
$$\frac{\partial u}{\partial t} = \|\nabla u\| \left(\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \right)^{\frac{1}{3}} \leftrightarrow \frac{\partial C_L}{\partial t} = -k^{\frac{1}{3}}\bar{N}$$



Detector morfológico de esquinas

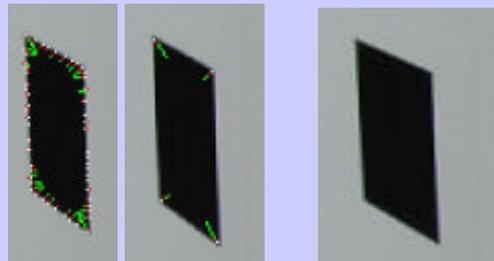
$$\frac{\partial u}{\partial t} = \|\nabla u\| \left(\operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \right)^{\frac{1}{3}} \leftrightarrow \frac{\partial C_L}{\partial t} = -k^{\frac{1}{3}}\bar{N}$$

(x_0, y_0)



$$(x(t), y(t)) = (x_0, y_0) + \frac{t}{\sqrt{\tan \frac{\alpha}{2}}} (b_x, b_y)$$

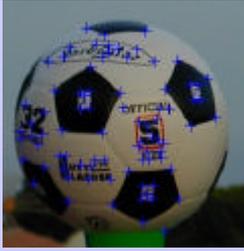
Aplicación del detector morfológico de esquinas (1)



$t_0=0$

$t_0=6$

Aplicación del detector morfológico de esquinas (2)

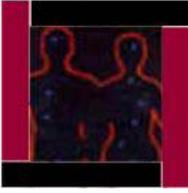


La optimización y el análisis numérico

- Interpolación de funciones
- Resolución de grandes sistemas de ecuaciones
- Cálculo de autovalores y autovectores
- Descomposición de SVD de matrices
- Análisis Numérico de Ecuaciones en Derivadas Parciales
- Estimación robusta de parámetros
- Ajuste de parámetros (Levenberg-Marquardt)
- Estimadores robustos

Aspectos matemáticos del proyecto multidisciplinar de I+D “Espacio Acústico Virtual”

SCTM03



Haresh M. Chulani

Ingeniero Electrónico del Area de Instrumentación del Instituto de Astrofísica de Canarias

Antonio F. Rodríguez Hernández

Licenciado en Medicina y Cirugía e Investigador de Proyecto del Departamento de Fisiología de la Universidad de La Laguna

1. Introducción

Es proverbial la capacidad del invidente para moverse, orientarse y reconocer e interactuar con el entorno apoyándose en los otros sentidos. Braille, con el desarrollo de su sistema de lectura, nos muestra que es posible realizar de una manera efectiva y útil tareas que habitualmente consideramos propias, si no exclusivas, de un sentido a través de otra modalidad sensorial. El presente proyecto incide en esta concepción y aborda una línea de investigación y desarrollo en la percepción del entorno inmediato del invidente a partir de sonidos.

1.1. “Sonidos que se extienden”. La hipótesis se fundamenta en el siguiente razonamiento. La imagen visual que experimentamos se genera a partir de las imágenes del entorno impresionadas en nuestras retinas. Cada punto de una imagen retiniana dada se corresponde con un punto concreto del espacio abarcado por la mirada. La información (en forma de onda luminosa) procedente de cada uno de los puntos del campo perceptivo visual es la que finalmente utiliza el cerebro para generar la imagen visual que experimentamos. Si esa misma información espacial procedente del objeto viaja en forma de onda sonora, se plantea la hipótesis de si es posible que se genere también en el receptor una imagen de la ocupación espacial por parte del objeto u objetos presentes en el campo perceptivo, útil para el reconocimiento y la interacción con los mismos.

Este razonamiento, junto a toda una serie de indicios científicos que parecen apoyarlo, procedentes tanto del campo de la neurofisiología como de la psicología y la psicoacústica (por ejemplo: Andersen, 1997; Konishi, 1986; Takahashi, 1994; Lakatos, 1993; Bach-Y-Rita, 1969; Rice, 1965) hace gestar pues la idea de que:

“A través del estímulo sonoro adecuado, un sujeto puede percibir de una forma global, de forma similar a como ocurre con la visión, una imagen de los objetos del entorno, con la opción de percatarse de sus características espaciales como su posición, las dimensiones en los tres ejes, sus límites, etc.”

Se aborda así una línea de investigación dirigida a validar tal hipótesis y a establecer las propiedades que debe cumplir dicho estímulo. Por la aplicación última de la misma y por la indicación de los propios argumentos neurofisiológicos y psicológicos referidos, se plantea desde el inicio explorar la viabilidad de la hipótesis en el grupo de la población invidente, sin cuya valiosa colaboración ninguno de los avances habría sido posible.

Lo que se pretende es, pues, generar una ***percepción global y mantenida en el tiempo del conjunto del entorno captado***, pero para ser experimentada ***como una imagen*** de la

ocupación espacial por parte de los objetos de dicho entorno, lo cual se acerca más a la experiencia de la visión que a la propia de la audición.

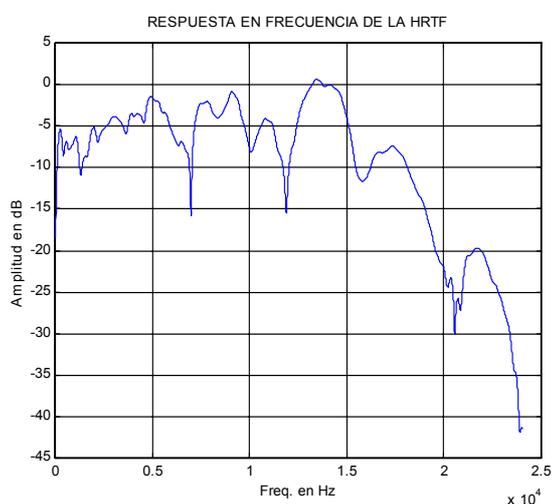
Existen múltiples líneas de investigación sobre percepción del entorno mediante sonidos. En algunas se ha utilizado el sonido únicamente como indicador de presencia o no de un objeto y su posible distancia, en el caso de los denominados detectores de obstáculos [por ejemplo: Doods, 1981; Benjamin, 1974; Loomis JM, Golledge RG, Klatzky RL, 1998]. Las líneas más cercanas a nuestro planteamiento pretenden generar un conocimiento de la escena en sí de una manera global, pero tras aprender un código y basando el reconocimiento de los objetos en el aprendizaje de un patrón sonoro concreto o específico [Kay, 1984; Capelle et al, 1998].

1.2. Las bases de la localización humana de fuentes sonoras. La información espacial es transmitida en nuestro caso utilizando el código natural, es decir: **codificamos cada punto del espacio ocupado por un objeto mediante un sonido que es percibido por el sujeto como si procediera de dicha posición.** El estímulo sonoro final correspondiente a uno o varios objetos suficientemente grandes de una escena, equivale pues, en líneas generales, al efecto de cubrir con pequeñas fuentes sonoras la superficie de los objetos a percibir. Para entender cómo es posible generar en un sujeto por medio del sonido que se le envía a través de unos auriculares, la ilusión perceptiva de que tal sonido procede de múltiples posiciones que componen la superficie de un objeto, es conveniente repasar las bases de la capacidad humana de localizar fuentes sonoras aisladas.

Para localizar una fuente sonora externa, el sistema auditivo humano se apoya en una serie de pistas contenidas en el sonido que está recibiendo. Tales son:

- La diferencia interaural de tiempo de llegada de un mismo sonido a un oído y al otro (ITD).
- La diferencia interaural de intensidad con que llega un mismo sonido a un oído frente al otro (IID).
- Las modificaciones que sufre el sonido en cada oreja por efecto de reflexiones, difracciones, etc. en sus repliegues (HRTF).

Dada una fuente sonora a nuestra derecha, el sonido emitido llega antes al oído derecho que al izquierdo. La diferencia del tiempo de llegada depende de cuánto a la derecha esté la fuente respecto a nosotros. Igualmente ocurre con la intensidad: la presencia de los relieves faciales y craneales ejercen un efecto de sombra sobre determinadas frecuencias procedentes de una fuente sonora. Se genera así una diferencia de intensidad percibida en cada oído que el cerebro reconoce como correspondiente a una determinada posición en el espacio (concretamente en el eje horizontal). Además, cada oreja, junto al resto de la cabeza y del cuerpo en menor medida, da lugar a una serie de modificaciones en la señal acústica que llega hasta ella y la sobrepasa para llegar al tímpano. Estas modificaciones son dependientes y específicas de la dirección relativa de



aproximación de la onda acústica a la oreja, y el cerebro es capaz de extraerlas de la señal que finalmente le llega al tímpano. Estas modificaciones constituyen una pista más de localización, en particular para determinar la altura de la que procede el sonido, si se encuentra delante o detrás de nosotros, y el hecho de que proceda de fuera de nosotros y no sea un sonido de producción corporal. Con la técnica apropiada (que se abordará más adelante), dichas modificaciones pueden ser medidas y representadas en forma de una función matemática. Esta función corresponde a las modificaciones espectrales que ejerce la oreja y el resto del cuerpo sobre la onda sonora incidente, y consisten en una serie de picos y de depresiones de la función de la intensidad sonora respecto de la frecuencia. Es conocida por las siglas HRTF (de Head Related Transfer Function).

1.3. “Sonidos que se ven”. Varios de los colaboradores invidentes han referido que perciben en el punto donde acústicamente localizaban la fuente sonora, unos destellos luminosos que describen como “pequeñas lucecitas o estrellitas”, simultáneos a la percepción acústica e inducidos por esta. Este fenómeno se ha repetido a lo largo del tiempo, siempre que no se han variado las condiciones del estímulo. Con la utilización del prototipo, se ha reproducido en sólo una ocasión en el caso de la única colaboradora de aquellas dos que lo ha utilizado, y refiere diferencias cualitativas con respecto a las demás ocasiones en que no lo ha percibido, en el sentido de una mayor facilidad para la percepción de dimensiones y formas, aparte de producirle una impresión muy agradable y que le recordaba la experiencia de la visión anterior. Se ha iniciado un estudio de la naturaleza de este fenómeno así como sobre las condiciones en que se puede dar.

2. Aspectos metodológicos

Dada la hipótesis, se plantea una experimentación en la que se explora la percepción y la capacidad de reconocimiento de patrones espaciales de fuentes sonoras reales, i.e., figuras construidas mediante altavoces y que se hacen sonar de una determinada manera. Posteriormente se desarrolla un prototipo de laboratorio (descrito más adelante) que hace percibir a un sujeto la situación virtual de que los objetos presentes en el campo perceptivo frontal del mismo, sin estarlo, parezcan estar recubiertos de fuentes sonoras emitiendo el estímulo. Las recientes técnicas de **auralización o tridimensionalización** de sonidos [Wightman & Kistler, 1989; para una introducción general al tema: Begault, 1994] y el desarrollo de los sistemas de **captación y reconocimiento del entorno** de la visión artificial [Faugheras, 1993], ofrecen la posibilidad de una generación de dicho efecto virtual. De esta manera, se explora si esta codificación particular del entorno puede llegar a ser de utilidad práctica para el invidente.

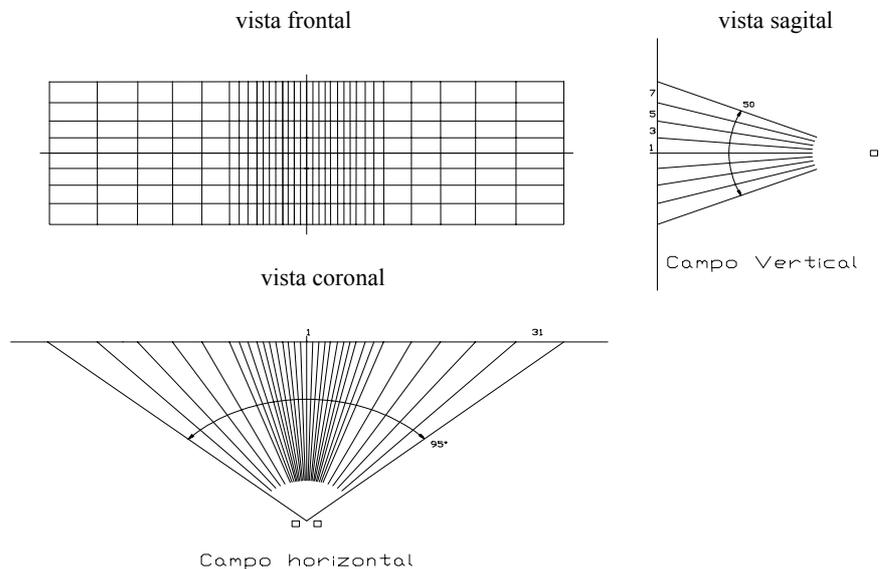
La técnica de auralización empleada, como se explicará más adelante, requiere la realización de sesiones de grabación para cada sujeto de los sonidos correspondientes a las posiciones que se quieren presentar de forma virtual. Estas sesiones pueden prolongarse excesivamente en el tiempo. Entre distintos abordajes posibles para solucionar esta cuestión (uno de ellos, la interpolación de unas funciones a partir de un número menor de funciones medidas), se han realizado estudios para evaluar la posibilidad de sustituir a cada sujeto por una reproducción anatómica precisa del mismo. En colaboración con el departamento de Escultura y Pintura de la facultad de Bellas Artes de la Universidad de la Laguna, se ha elaborado en silicona un maniquí individualizado, y se han realizado las primeras pruebas de comparación del comportamiento acústico del mismo en relación al sujeto original. Los resultados iniciales se

muestran favorables a la hipótesis. Actualmente trabajamos en la cuantificación matemática de la similitud gráfica observada.

3. Descripción del prototipo realizado

Se ha construido un prototipo que obtiene información de las distancias de los objetos situados delante del usuario y reproduce un conjunto de sonidos virtualmente espacializados de manera que se perciban como si estuviesen situados donde dichos objetos. Dos microcámaras situadas en un soporte adecuado que el usuario porta en su cabeza recogen la información del entorno y la envían a los ordenadores, los cuales tras las oportunas operaciones enviarán a los auriculares que el usuario tiene puestos el conjunto de sonidos correspondiente. El usuario tiene así la sensación de que los objetos emiten sonidos, y generan todos ellos un mapa de sonidos virtuales que le permiten extraer información de la forma y volumen de su entorno.

El espacio que rodea al sujeto debe ser discretizado, es decir, transformado de un todo continuo con infinitas distancias a considerar, a un entorno discontinuo con un número finito de distancias a reproducir, con el objeto de que sea manejable electrónica/informáticamente. A tal efecto se trabaja con un sistema de coordenadas esféricas, cuyo centro es la cabeza del sujeto, y del que únicamente reviste interés el sector esférico que se encuentra inmediatamente delante del individuo. Dicho sector se descompone en "estereopíxeles", en cada uno de los cuales estará situada una porción del entorno, con un valor de profundidad (distancia respecto al usuario) determinado.

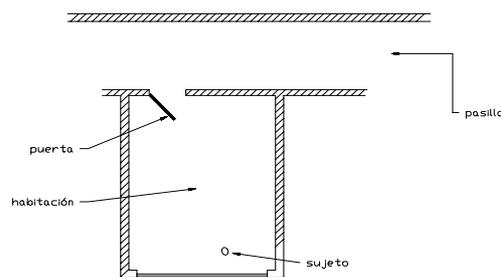


Descomposición en estereopíxeles de un campo rectangular.

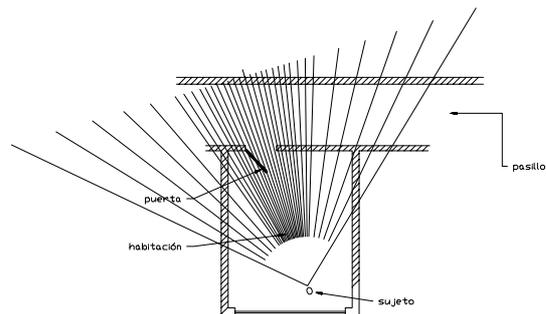
La figura anterior representa esquemáticamente un ejemplo de distribución de estereopíxeles en un campo rectangular, de 32×8 elementos, los cuales tienen ángulos diferentes desde el punto de vista del sujeto, de tal forma que la resolución en el centro del campo aparece reforzada frente a los laterales.

A modo de ejemplo, y al objeto de visualizar el efecto de la división del entorno en estereopíxeles sobre la percepción del mismo, seguidamente se representa un caso en el que el sujeto se encuentra en una habitación normal, dirigiendo su campo de visión a una puerta entreabierta.

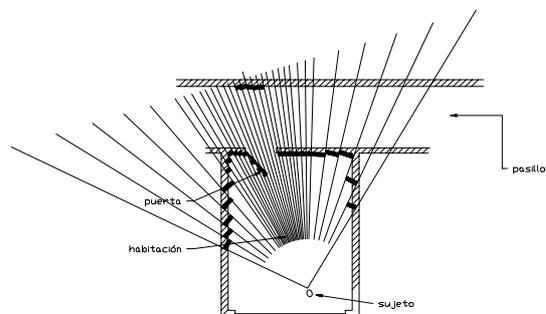
a) Entorno ejemplo:



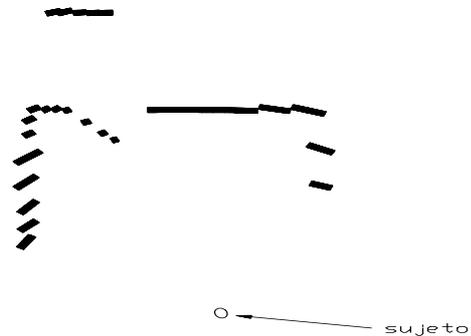
b) Trazado de los estereopíxeles:



c) Determinación de valores de los estereopíxeles:



d) Percepción resultante:



Puede observarse la manera en la que sobre el entorno inicial (a), tras el trazado de los estereopíxeles (b), se pueden obtener los valores de las distancias de las porciones del entorno encerradas por los mismos (c), y con ello la percepción a la que el conjunto daría lugar (d).

El número de estereopíxeles finalmente escogido ha sido de 17 en azimut (horizontal) \times 9 en elevación, con 8 profundidades posibles por estereopíxel.

La solución técnica conceptual para el prototipo consiste en dos subsistemas independientes, Subsistema de Visión y Subsistema de Sonidos, los cuales se comunican mediante un enlace de comunicaciones para la transmisión de las informaciones pertinentes entre ellos. Esto se muestra en la figura siguiente. El subsistema de captación o sistema de visión artificial registra las formas del entorno del sujeto mediante dos microcámaras de video situadas en la cabeza a la altura de los ojos, colocadas adecuadamente sobre un soporte diseñado al efecto. Una vez el subsistema de visión ha calculado el mapa de profundidades del entorno, envía esta información al subsistema de sonidos. Éste selecciona y suma los sonidos adecuados a reproducir, y envía esta señal acústica a unos auriculares que también forman parte del soporte que el usuario porta. Toda esta operación se realiza unas 10 veces por segundo.

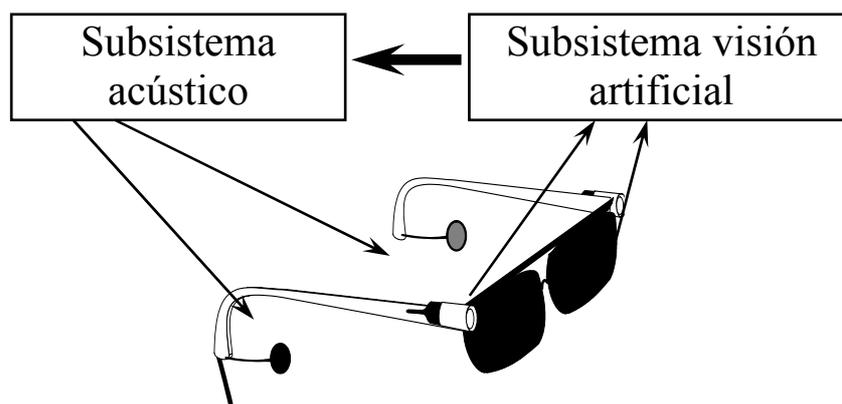


Diagrama conceptual del prototipo desarrollado.

Para que los sonidos se escuchen por los auriculares pero se perciban provenientes del entorno exterior, es necesario haberlos procesado previamente con las denominadas HRTFs (Head Related Transfer Functions), que deben haber sido medidas para cada usuario del prototipo, para cada una de las posiciones de interés.

3.1. Subsistema de visión. Para el sistema de captación del entorno en los comienzos del proyecto se estudiaron varias posibilidades, considerándose en un principio como requerimientos la utilización de sistemas pasivos de percepción del entorno que ofreciesen un rango de medidas amplio, posibilidades de mejora considerable en requerimientos de resolución en cualquiera de las tres dimensiones espaciales, y captación de características del entorno que aportasen información adicional a la espacial.

La solución escogida fue la utilización de sistemas pasivos de visión artificial basados en la adquisición de imágenes con una o varias cámaras CCD y su posterior procesamiento mediante computadores digitales, para la determinación del mapa de profundidades y otras propiedades de los objetos como color, textura, etc. El trabajo inicial se ha centrado únicamente en la obtención del mapa de profundidades, para el envío mediante sonidos de información simplemente espacial, con idea de validar la idea de percepción de superficies y formas mediante sonidos.

Se han estudiado dos técnicas para la obtención del mapa de profundidades de la escena:

- **Estereovisión o visión estereoscópica.** Se basa en la diferencia existente entre las imágenes de un objeto visto simultáneamente desde dos posiciones diferentes.
- **Profundidad a partir del desenfoque (DFD depth from defocus).** Estos métodos utilizan dos imágenes simultáneas de un objeto tomadas desde la misma posición y con distintos parámetros ópticos (focal de las lentes, diámetro del diafragma de apertura, distancia entre lente y sensor), lo cual da lugar a que el desenfoque en cada punto sea distinto en ambas imágenes y dependiente de la distancia entre la cámara y el objeto.

Actualmente el prototipo implementa únicamente el método de estereovisión. En un futuro se pretende fusionar ambos métodos para aprovechar las ventajas de cada uno de ellos.

Estereovisión. Suponemos dos cámaras con objetivos de distancia focal f separadas una cierta distancia b . Un punto (x, y, z) del espacio se proyectará en los puntos (x_1, y_1) y (x_2, y_2) de los planos imagen de cada cámara. Suponiendo que las cámaras tienen los ejes ópticos paralelos entre sí y perpendiculares a los planos imagen, la distancia z se puede obtener a partir de la diferencia en las coordenadas x : $z = bf/d$, donde $d = x_1 - x_2$ se conoce como disparidad. Para conocer la distancia, por tanto, tenemos que encontrar en una imagen el punto que corresponde a un punto dado de la otra imagen: es el llamado problema de la correspondencia.

En la práctica es necesario un paso previo al proceso de correspondencia, que es el proceso de rectificación y consiste en una transformación de las imágenes obtenidas de manera que se consigan las imágenes que corresponderían al caso ideal (cámaras paralelas, lentes sin distorsión). En las imágenes rectificadas, los pares de puntos correspondientes tienen la misma coordenada y , lo cual permite restringir la búsqueda a los píxeles que están en la misma fila que el pixel original.

Para poder aplicar la rectificación y para poder calcular la profundidad z a partir de la disparidad d necesitamos conocer los valores de los parámetros ópticos del sistema (coeficientes de distorsión y distancia focal de las lentes, ángulos de rotación de una cámara respecto a la otra, etc.). Estos valores se calculan en un procedimiento previo de calibración del sistema, basado en tomar 5 pares de imágenes de un patrón de calibración (consistente en cuadros blancos y negros de un determinado tamaño previamente conocido) y buscar las coordenadas imagen de sus esquinas.

El problema fundamental de la obtención de profundidad mediante estereovisión es, pues, el problema de la correspondencia; es decir, dado un pixel de una imagen, buscar (en la

misma fila) el pixel correspondiente en la otra imagen y obtener de esta manera un mapa de disparidades a partir del cual se puede obtener el mapa de profundidades. El algoritmo de correspondencia usado por nosotros consta de los siguientes pasos:

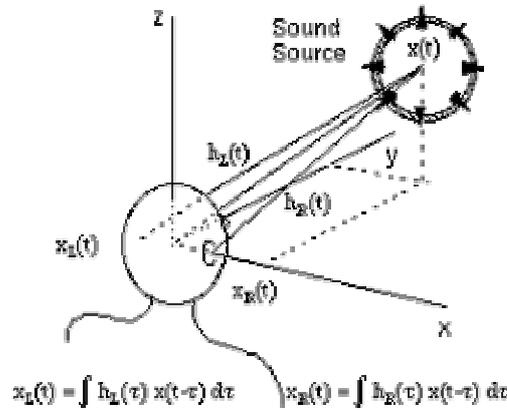
- Rectificación de las imágenes tomadas por las cámaras.
- Extracción de características: convolución con el laplaciano de la gaussiana. El par de imágenes así obtenido contiene información sobre los contornos de las imágenes originales.
- Correlación mediante la suma del valor absoluto de las diferencias de ambas imágenes a lo largo de una pequeña ventana. Para cada pixel de la imagen izquierda se busca el máximo de la correlación en la imagen derecha, hasta un valor máximo de disparidad. Se interpola la posición del máximo de correlación hasta una precisión de $\frac{1}{4}$ pixel.
- Filtrado: se eliminan correspondencias poco fiables. En este paso se emplea una medida de confianza basada en la energía del contorno y también se hace una prueba de consistencia derecha/izquierda (es decir, se busca el pixel en la imagen izquierda que corresponde a un determinado pixel de la imagen derecha y se comprueba que la correspondencia sea la misma en ambos sentidos).

3.2. Subsistema de sonidos. Como repuesta a la información proporcionada por el sistema de captación del entorno, el subsistema acústico se encarga de reproducir sonidos espacializados, es decir, escuchados por el usuario a través de los auriculares pero percibidos como si estuviesen situados *en* los objetos del entorno. Los sonidos han sido procesados previamente fuera de tiempo real de manera que se perciban situados en el entorno exterior al sujeto. Posteriormente se han almacenado en la memoria del subsistema de sonidos. El subsistema selecciona en cada momento los sonidos que correspondan, los suma y reproduce el resultado por los auriculares.

Las tareas que realiza el subsistema de sonidos son:

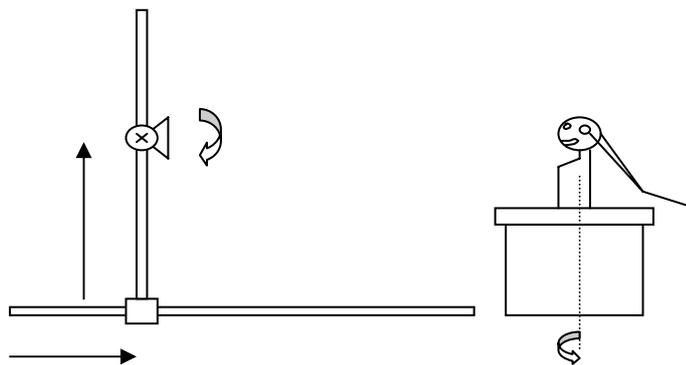
- Recepción de los mapas de profundidades enviados por el subsistema de visión
- Cálculo del orden de presentación de sonidos correspondientes al conjunto de estereopíxeles recibidos, aleatorio y diferente para cada imagen. Esta aproximación evita patrones sonoros repetitivos que distraen la atención del usuario respecto de la información espacial contenida en los sonidos.
- Reproducción de los sonidos correspondientes a cada estereopíxel, uno a uno, manteniendo entre ellos una separación temporal programable. En este momento sólo se utiliza un tipo de sonido, corto e impulsivo, sin cualidades tonales.

Para la correcta percepción del sonido en tres dimensiones, es necesario procesarlo con las denominadas HRTF (Head Related Transfer Functions). La transformación que sufre una señal acústica desde que se emite en una determinada posición espacial relativa al oyente, rebota en todas las superficies del entorno, incluidas las del cuerpo del oyente, hasta que excita el tímpano de este último, se considera una transformación lineal, y por lo tanto, caracterizable por medio de una función de transferencia. Para cada posición espacial relativa al oyente tendremos un par de funciones de transferencia, una para cada oído, y las llamamos HRTFs.



Head Related Transfer Functions.

Se ha desarrollado un sistema de medidas de HRTFs que permite medir HRTFs individuales para cada usuario del prototipo. Un robot manipulador sitúa un altavoz respecto del usuario en la posición espacial relativa de interés. Se reproduce un sonido especial de medida, denominado MLBS (Maximum Length Binary Sequence) a través del altavoz y se graban las respuestas de dos micrófonos miniatura situados en los canales auditivos del usuario (o bien de un maniquí réplica del busto del mismo). El proceso se repite para todas las posiciones espaciales de interés, y además para el caso en que el sonido se reproduzca desde los auriculares seleccionados para el prototipo. Esto es así porque los sonidos que finalmente se reproducen han de estar ecualizados para compensar la reproducción a través de esos auriculares en concreto. El sistema descrito se sitúa dentro de una sala acondicionada para tener el suficiente aislamiento acústico y un pequeño grado de reverberación.



Grados de libertad de movimiento del robot manipulador de altavoz y voluntario.

La HRTF pura no la podemos medir, puesto que las funciones que medimos siempre estarán contaminadas de la respuesta de altavoz, micrófono y demás elementos de la cadena de audio. Sin embargo, el método es válido, y para entenderlo, quizás convenga formular matemáticamente el problema.

Sea $x_1(n)$ el registro digital del sonido a reproducir por un altavoz, y sea $y_1(n)$ el registro resultante de grabar con el micrófono situado en el canal auditivo del voluntario la respuesta a $x_1(n)$. Similarmente, sea $x_2(n)$ el sonido a reproducir por un auricular e $y_2(n)$ la respuesta grabada del mismo. Nuestro objetivo es entonces, para un determinado $x_1(n)$, generar un $x_2(n)$ tal que la resultante $y_2(n)$ sea igual a $y_1(n)$, ya que así habremos conseguido que un estímulo sonoro generado desde un altavoz y otro generado desde un auricular produzcan la misma respuesta en el canal auditivo del individuo, y por tanto, la misma impresión acústica y espacial. Esto lo hacemos diseñando un filtro digital que transforme $x_1(n)$ en $x_2(n)$.

La descripción del diseño del filtro es más sencilla en el dominio de la frecuencia. Llamemos X_1 a la representación de $x_1(n)$ en el dominio transformado de la frecuencia, Y_1 a la de $y_1(n)$, etc. La respuesta registrada como resultado de reproducir $x_1(n)$ se puede escribir como:

$$Y_1 = X_1 LFM \quad (1)$$

donde L , por simplicidad, es la función de transferencia conjunta del altavoz más toda la cadena de audio de reproducción, F es la función de transferencia del espacio libre entre altavoz y canal auditivo, que hemos llamado anteriormente *HRTF*, y M es la función de transferencia conjunta de micrófono más toda la cadena de audio de grabación. La respuesta registrada por el micrófono al reproducir $x_2(n)$ a través del auricular se puede expresar como:

$$Y_2 = X_2 HM \quad (2)$$

donde H es la función de transferencia del auricular (y toda la cadena de reproducción) al canal auditivo. Haciendo $Y_1 = Y_2$ y despejando X_2 obtenemos:

$$X_2 = X_1 LF / H \quad (3)$$

lo cual muestra que el filtro digital que buscamos viene dado por:

$$T = LF / H \quad (4)$$

Por tanto, si la señal $x_1(n)$ la filtramos con este filtro y la señal resultante $x_2(n)$ la reproducimos por el auricular, la señal registrada por el micrófono situado en el canal auditivo cabe esperarse sea $y_1(n)$, la misma que se registraría al reproducir $x_1(n)$ por el altavoz.

El filtro descrito por (4) se refiere a un único altavoz en una determinada posición espacial y un sólo oído. Necesitaremos dos filtros, uno para cada oído, para sintetizar el estímulo que simule cada fuente sonora en una determinada posición espacial.

Medimos las funciones de transferencia del tipo Y_1 / X_1 , a las que denominamos ‘funciones de transferencia altavoz-micrófono’ G_{LM} , para las diferentes posiciones espaciales de interés y para los dos oídos simultáneamente:

$$G_{LM} = \frac{Y_1}{X_1} = L \cdot F \cdot M \quad (5)$$

Posteriormente medimos para los dos oídos, simultáneamente, las dos funciones de transferencia Y_2 / X_2 , a las que denominamos ‘funciones de transferencia auricular-micrófono’ G_{HM} :

$$G_{HM} = \frac{Y_2}{X_2} = H \cdot M \quad (6)$$

Los filtros digitales necesarios para la simulación de un sonido proveniente de las posiciones espaciales de interés (a los que denominamos ‘filtros totales’) se obtienen multiplicando las funciones altavoz-micrófono G_{LM} de cada oído por la inversa de la función auricular-micrófono G_{HM} del mismo oído (ver ecuación (4)), así para ambos oídos:

$$T = \frac{G_{LM}}{G_{HM}} = \frac{L \cdot F \cdot M}{H \cdot M} = \frac{L \cdot F}{H} \quad (7)$$

Tanto para la medida de la función de transferencia altavoz-micrófono como para la del auricular-micrófono se ha utilizado la técnica de medida de respuestas impulsionales por medio de secuencias de máxima longitud (MLBS, de ‘Maximum Length Binary Sequences’), con la posterior correlación cruzada entre la respuesta del sistema y la MLBS de entrada.

La respuesta impulsional del sistema a identificar se obtiene por medio de una correlación cruzada circular entre la MLBS de entrada al sistema y la respuesta a la salida. Es decir, si excitamos al sistema con una MLBS que denominamos $s(n)$, y medimos la a la salida la señal $y(n)$ durante el intervalo de tiempo que dura la MLBS, la respuesta impulsional $h(n)$ viene dada por:

$$h(n) = \Omega_{sy}(n) = s(n) \Phi y(n) = \frac{1}{L+1} \sum_{k=0}^{L-1} s(k) \cdot y(n+k) \quad (8)$$

donde Φ representa la operación correlación cruzada circular o periódica; por tanto, $h(n)$ es una respuesta impulsional periódica, contaminada de aliasing temporal, y no una respuesta impulsional pura. Siempre y cuando se escoja una secuencia lo suficientemente larga, el aliasing resultante se podrá despreciar.

Puesto que la implementación directa de la ecuación para secuencias largas es inviable por el tiempo de cálculo que ello conllevaría, se ha utilizado la equivalencia entre correlación y convolución cruzadas periódicas y se ha pasado al dominio frecuencial, donde la operación de convolución se traduce en una multiplicación de vectores. La mencionada equivalencia es:

$$a(n) \Phi b(n) = \frac{1}{L+1} a(-n) * b(n) \quad (9)$$

donde la inversión de la primera secuencia es circular, al igual que la convolución. Aún así el tiempo de cómputo es excesivo, puesto que las FFTs utilizadas son de longitud 2^k-1 , y no de longitud potencia de dos, que es cuando realmente son eficientes.

Finalmente, se ha conseguido reducir el tiempo de cómputo en dos órdenes de magnitud, empleando la FHT (Fast Hadamard Transform). La ecuación que rige el cálculo de la respuesta impulsional $h(n)$ es:

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \langle S_2 \{ H_{L+1} [S_1 (P_1 y(n))] \} \rangle \quad (10)$$

donde se ha excitado el sistema con una MLBS $s(n)$ de longitud L , y se ha registrado la respuesta $y(n)$. Las matrices P son matrices de permutación, las matrices S son matrices de redimensionamiento, y H_{L+1} es la matriz de Hadamard de orden $L+1$.



Subsistema de generación/procesado de sonidos. A la izquierda, rack donde está ubicado el sistema Huron, con 40 salidas analógicas, 8 entradas analógicas y 8 DSPs 56002. Se ha utilizado para experimentos como el “multifuente”, para el sistema de medida de HRTFs y para el subsistema de sonidos del prototipo. A la derecha, ordenador en el que se calculan las HRTFs y los sonidos espacializados off-line.

4. Resultados con el prototipo actual

Los primeros resultados se basan en las respuestas de una muestra de diez colaboradores invidentes, heterogénea en cuanto a sexo, edad, grado, duración y causa de la ceguera, etc., si bien en su mayoría son personas caracterizadas por tener desarrollado un fino sentido de la audición y de la orientación en entornos conocidos. Tal respuesta se ha medido a partir de la descripción verbal y del registro informatizado de gestos o señalamientos realizados por el voluntario para describir las distintas figuras y patrones espaciales presentados, así como en la realización de tareas de orientación y movilidad concretas, tanto ante fuentes reales como virtuales.

Presentados de una manera general, y considerando que los objetos a percibir varían desde formas puntuales hasta, mono, bi y tridimensionales (lo cual incluye, líneas horizontales o verticales, superficies planas y curvas cóncavas o convexas, aisladas o agrupadas componiendo figuras, como por ejemplo, cuadrados, o columnas o filas paralelas, etc.), podríamos destacar los siguientes aspectos observados:

- Un objeto presente en el campo de percepción del usuario, a partir de la información sonora recibida, puede ser descrito correctamente, en aspectos espaciales significativos como son: su posición, su distancia y las dimensiones en los ejes horizontal y vertical e incluso en el eje z de profundidad.
- Dos objetos separados por un determinado espacio, todo dentro del campo perceptivo captado por el sistema, pueden ser percibidos igualmente en sus posiciones, distancias - que pueden ser distintas, más próximas o más lejanas -, y con sus límites externos e internos.
- Los objetos, transcurrido un breve periodo de exploración, habitualmente inmediato, son percibidos en su disposición espacial de una manera global, i.e., presentes como un todo en el seno del campo perceptivo (por ejemplo, los lados de un cuadrado presentados en las distintas combinaciones posibles: una o dos líneas horizontales; una o dos verticales, etc., hasta el cuadrado total).

El invidente puede reconocer la presencia de un panel o superficie rectangular en su posición, a su distancia correcta, y con sus dimensiones en anchura y en altura. Se percibe además la condición de *superficie*, de continuidad espacial, en un grado suficiente –si bien esperamos mejorable- como para, como suelen referir los sujetos, poder decir que “por ahí no pasaría”. Dos paneles a modo de paredes formando un pasillo, son percibidos como objetos a ambos lados, con sus dimensiones verticales y de profundidad, y que dejan en medio un espacio de silencio por el que se puede pasar y de hecho pasa el invidente.

Los resultados son pues muy alentadores en las pruebas realizadas con personas ciegas. Se ha comprobado que es posible a partir del estímulo acústico recibido notar la presencia, en su sitio y con sus dimensiones, de los objetos captados y que no eran en un principio sonoros en sí mismos.

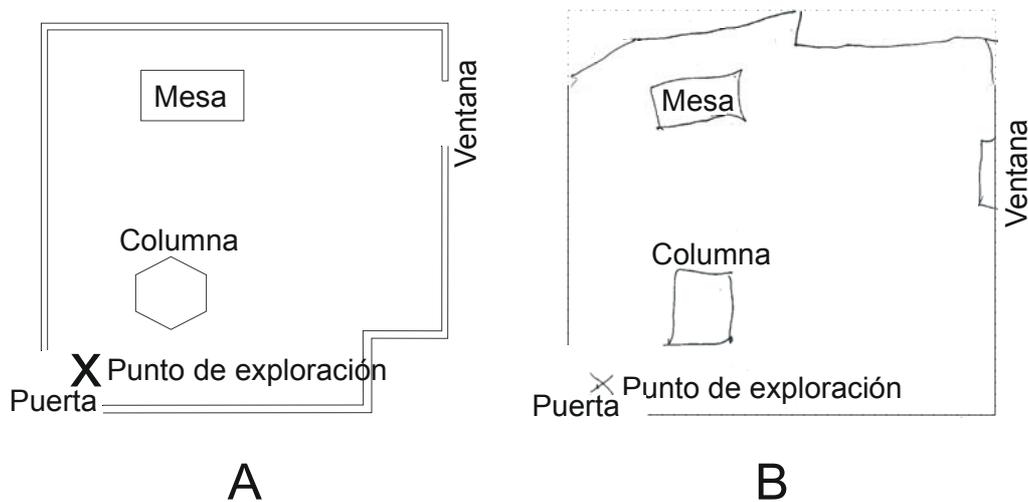
No se trata de que se localizan muchos sonidos en sus posiciones más o menos alejadas cada uno, como los sonidos que se pueden apreciar en una reunión multitudinaria, sino que **se percibe un solo sonido que se extiende, forma una superficie continua y termina en unos bordes netos que puedes delimitar, y que además tiene volumen.**



Voluntario señalando los límites de una apertura a modo de ventana en un panel del laboratorio.

De esta forma, la persona es capaz por medio del sonido de reconocer una **columna** o la presencia y dimensiones del hueco de una **ventana** en el seno de una **pared**, pasar a través del hueco de una **puerta** o caminar dejando al lado la pared de un **pasillo**.

Simulando las condiciones de la vida cotidiana en una habitación experimental relativamente simple, ha sido posible para la persona ciega moverse y extraer información suficiente para hacer una descripción verbal y gráfica de la escena global .



A. Croquis de una habitación experimental con una disposición de objetos desconocida para una usuaria invidente. B. Dibujo de la habitación con la distribución de los objetos reconocidos por la sujeto usuaria del dispositivo tras una exploración del entorno sin apoyarse en el tacto.

Por otro lado, el estímulo utilizado permite la captación y comprensión de otros sonidos procedentes del ambiente como ruidos o una conversación.

Otro resultado obtenido en las pruebas ha sido la aparición de los denominados **fosfenos** evocados por sonidos. En algunas de las personas ciegas, determinados sonidos con los que hemos estado trabajando provocan una sensación de luz (denominada fosfeno). Son como luciérnagas o estrellitas que la persona percibe simultáneamente en el mismo punto donde oye la fuente sonora, o el escenario acústico.

Existen referencias a este fenómeno pero con muy pocos casos y además la mayoría transitorios y sin esa relación espacial tan marcada entre posición oída y posición de las luces [Lessel & Cohen, 1979; Page et al, 1982].

Este hecho ha dado lugar a la apertura de una línea de investigación que pensamos podría tener importantes implicaciones en nuestro objetivo global de ayuda a la orientación y la movilidad, así como en otros aspectos clínicos y de rehabilitación en la ceguera.

Los primeros estudios con la técnica de **Resonancia Magnética funcional**, que muestra cortes del cerebro donde se aprecian las áreas cerebrales que se activan durante la realización de una determinada tarea mental (perceptiva, imaginativa, memorística, de control de movimientos, etc.) muestran una activación de la corteza visual cuando el sujeto, mediante el estímulo sonoro, percibe fosfenos.

Como en toda investigación, especialmente en las ciencias de la vida, la estadística juega un papel muy importante tanto en el diseño experimental de las pruebas planteadas como en el análisis de los resultados obtenidos.

5. Líneas de investigación y desarrollo abiertas

En la actualidad se está culminando el desarrollo de un **dispositivo portátil** para su utilización cotidiana en **ambientes habituales**, es decir, ya fuera de los límites del laboratorio, por personas ciegas. Esto permitirá la consiguiente evaluación de validez y utilidad del prototipo.

Además, se continúa la necesaria línea de investigación para la mejora del estímulo sonoro de cara a conseguir las percepciones de una manera más fina, así como de los distintos elementos implicados en el fenómeno perceptivo bajo estudio (papel de los factores individuales, papel del aprendizaje, etc.).

Paralelamente se mantiene la línea de investigación neurofisiológica destinada a valorar el potencial de adaptación del cerebro humano ante este tipo de estímulo particular y ante la posibilidad de generar sensaciones luminosas de forma controlada mediante estímulos auditivos.

Las matemáticas pueden ejercer un papel determinante en la consecución de toda una serie de mejoras del dispositivo actual. La obtención de las HRTFs individuales, puede beneficiarse del desarrollo de los oportunos métodos de obtención de funciones a partir de un número limitado de las mismas, normalmente referidos en la literatura científica como métodos de interpolación. La obtención de una información de profundidad (mapa de profundidad) estable y precisa, al menos con respecto a las técnicas actuales de estereovisión, se puede beneficiar del desarrollo de métodos adecuados para la eliminación de ruidos por inestabilidades periódicas de las imágenes de partida, etc. etc.

Ha surgido también la idea de utilizar este estímulo acústico para generar una interfaz entre la persona invidente y el ordenador, concretamente en forma de un juego para los niños. Éste podría tener valor, además del meramente lúdico, en la por otro lado dificultosa introducción de los niños al ferviente mundo actual de la informática..

Creemos que nuestro prototipo será una herramienta de ayuda más a la movilidad de personas ciegas o con deficiencia visual severa, como lo es por ejemplo también el bastón, con el que se podría utilizar de forma conjunta y complementaria. Además de **seguridad** en los

desplazamientos aporta **nuevas sensaciones**, como son la de **volumen, globalidad**, etcétera. **Es un dispositivo con el que puedes seleccionar el objeto o la escena que quieres reconocer con sólo girar la cara hacia donde están situados.**

También consideramos que puede ser de gran interés la aplicación de un sistema de este tipo como **herramienta de rehabilitación** en el **desarrollo del niño** ciego congénito o con déficit visual severo temprano. La ausencia de información del mundo distante, con todas las repercusiones que ello conlleva para el desarrollo tanto psicomotor como cognitivo e incluso emocional, podría compensarse con un estímulo acústico que se comporta como lo haría uno visual.

Los resultados, muy alentadores, no nos ocultan sin embargo que estamos a las puertas de un área de investigación muy reciente y en la que queda mucho por desvelar.

Reconocimientos

Este es un proyecto desarrollado por la Universidad de la Laguna y el Instituto Astrofísico de Canarias en colaboración con el Instituto Tecnológico de Canarias, la O.N.C.E., el Hospital Universitario de Canarias, el Instituto Canario de Investigación y Desarrollo, Imásde y pixel 486, con financiación a lo largo de sus diferentes etapas por fondos de la Consejería de Economía y Hacienda del Gobierno de las Islas Canarias, IMSERSO (Ministerio Español de Trabajo y Asuntos Sociales), fondos FEDER y fondos del Ministerio Español de Ciencia y Tecnología (TIC2001-3916 y FIT-070000-2002-549). **Nuestro mayor agradecimiento a las personas colaboradoras en la investigación, sin las cuales no hubiera sido posible su realización.**

Bibliografía

- Andersen, R. A. MULTIMODAL REPRESENTATION OF SPACE IN THE POSTERIOR PARIETAL CORTEX AND ITS USE IN PLANNING MOVEMENTS. *Annual Review Neuroscience*, **20**, 303-330 (1997).
- Bach-Y-Rita, P. et al. VISUAL SUBSTITUTION BY TACTILE IMAGE PROJECTION. *Nature*, **1221**, 963-964 (1969).
- Begault, D. R. 3-D SOUND FOR VIRTUAL REALITY AND MULTIMEDIA, 1994.
- Benjamin, J. M. THE LASER CANE. *Bulletin of Prosthetics Research, Veterans Administration, Washington*, 1974 .
- Bregman, A. S. AUDITORY SCENE ANALYSIS: THE PERCEPTUAL ORGANIZATION OF SOUND. The MIT Press, 1990.
- Doods, A. G. et al. THE NOTTINGHAM OBSTACLE DETECTOR: DEVELOPMENT AND EVALUATION. *Ibid.*, **75**, 203-209 (1981).
- Faugeras, O. THREE-DIMENSIONAL COMPUTER VISION: A GEOMETRIC VIEWPOINT. The MIT Press, 1993.
- González Mora, J.L., Rodríguez-Hernández, A., Rodríguez-Ramos, L.F., Díaz-Saco, L., Sosa, N. DEVELOPMENT OF A NEW SPACE PERCEPTION SYSTEM FOR BLIND PEOPLE, BASED ON THE CREATION OF A VIRTUAL ACOUSTIC SPACE. *Proceedings of the International Work Conference on Artificial and Natural Networks*, Vol. 2, pp 321-330. Springer, 1999.

- Heikkila, Silven. A FOUR-STEP CAMERA CALIBRATION PROCEDURE WITH IMPLICIT IMAGE CORRECTION.
- Horn. ROBOT VISION. The MIT Press, 1986.
- Kanade, Kano, Kimura, Yoshida, Oda. DEVELOPMENT OF A VIDEO-RATE STEREO MACHINE. Proceedings of International Robotics and Systems Conference (IROS 95), August 1995.
- Kanade, Okutomi. A STEREO MATCHING ALGORITHM WITH AN ADAPTIVE WINDOW: THEORY AND EXPERIMENT. PAMI, **16**, no. 9, September 1994.
- Kay, L. ELECTRONICS AIDS FOR BLIND PERSONS: AN INTERDISCIPLINARY SUBJECT. IEEE Review. IEEE Proceedings, **131**, pt. A, no. 7, September 1984.
- Kistler, Doris J. and Wightman, Frederic L. A MODEL OF HEAD-RELATED TRANSFER FUNCTIONS BASED ON PRINCIPAL COMPONENT ANALYSIS AND MINIMUM-PHASE RECONSTRUCTION. J. Acoust. Soc. Am., **91**, 1637-1647(1992).
- Konishi, M. CENTRALLY SYNTHESIZED MAPS OF SENSORY SPACE. TINS, 163-168, April 1986.
- Krotov. ACTIVE COMPUTER VISION BY COOPERATIVE FOCUS AND STEREO. Springer-Verlag, 1989.
- Lai, Chang. A GENERALIZED DEPTH ESTIMATION ALGORITHM WITH A SINGLE IMAGE. PAMI, **14**, no. 4, April 1992.
- Lakatos, S. RECOGNITION OF COMPLEX AUDITORY SPATIAL PATTERNS. Perception. **22**, 363-374 (1993).
- Lessell, S. and Cohen, M.M. PHOSPHENES INDUCED BY SOUND. Neurology, **29**, 1524-1527 (1979).
- Loomis, J.M., Golledge, R.G., Klatzky, R.L. NAVIGATING SYSTEM FOR THE BLIND: AUDITORY DISPLAY MODES AND GUIDANCE. Presence, **7**(2), 193-203 (1998).
- Marr, D. VISION: A COMPUTATIONAL INVESTIGATION INTO THE HUMAN REPRESENTATION AND PROCESSING OF VISUAL INFORMATION. W. H. Freeman and Company, New York, 1982.
- Martens, W. L. PRINCIPAL COMPONENTS ANALYSIS AND RESYNTHESIS OF SPECTRAL CUES TO PERCEIVED DIRECTION. En J. Beauchamp (editor): Proceedings of the International Computer Music Conference, pp. 274-281. International Music Association, San Francisco, CA (1987).
- Nishihara, Poggio. STEREOVISION FOR ROBOTICS. En R. Paul and M. Bradu (editores): Robotics Research, The First International Symposium, pp. 489-505. MIT Press, 1984.
- Page N. G. R. et al. AUDITORY EVOKED PHOSPHENES IN OPTIC NERVE DISEASE. Journal of Neurology, Neurosurgery and Psychiatry, **45**, 7-12 (1982).
- Pentland, Scherrock, Darrell, Girod. SIMPLE RANGE CAMERAS BASED ON FOCAL ERROR. Journal Optical Society of America A, **11**, no. 11.
- Rice, Ch. E. et al. ECHO-DETECTION ABILITY OF THE BLIND: SIZE AND DISTANCE FACTORS. Journal of Experimental Psychology, **70**, 3, 246-251 (1965).
- Rodríguez Hernández, A; Sosa, N , Rodríguez Ramos, L.F. , Chulani, H. Díaz Saco, L, y González-Mora, J.L. PERCEPCIÓN DEL ENTORNO EN PERSONAS CIEGAS A TRAVÉS DE UN ESTÍMULO SONORO ESPACIAL VIRTUAL GENERADO POR COMPUTADOR. Actas del Congreso Iberoamericano 3º de CAA, 1º de Tecnologías de Apoyo para la Discapacidad, pp. 81-84, Octubre 2000.
- Rodríguez Hernández, A.F, González Mora, J.L., Pujol, J., López García, J.A., Muñoz Montes, J.R., Rodríguez Ramos, L.F., Chulani, H.M. and Burunat, E. FUNCTIONAL MAGNETIC RESONANCE IMAGING OF SPATIAL AUDITORY PERCEPTION IN

- BLIND AND SIGHTED HUMAN SUBJECTS. Aceptado para presentación en el XXXII Congreso de la Sociedad Española de Ciencias Fisiológicas, 2003.
- Rodríguez-Ramos, L.F., Chulani, H.M., Díaz-Saco, L., Sosa, N., Rodríguez-Hernández, A., González Mora, J.L. IMAGE AND SOUND PROCESSING FOR THE CREATION OF A VIRTUAL ACOUSTIC SPACE FOR THE BLIND PEOPLE. *Signal Processing and Communications*, 472-475 (1997).
- Sadato, N., Pascual-Leone, A., Grafman, J., Ibáñez, V., Daiber, M.P., Dold, G., Hallett, M. (1996). ACTIVATION OF PRIMARY VISUAL CORTEX BY BRAILLE READING IN BLIND PEOPLE. *Nature*, **380**, 526-527 (1996).
- Schlag, Sanderson, Neumann, Wimberly. IMPLEMENTATION OF AUTOMATIC FOCUSING ALGORITHMS FOR A COMPUTER VISION SYSTEM WITH CAMERA CONTROL. Technical Report CMU-RI-TR **314**, Carnegie-Mellon University, August 1983.
- Shih, Hung, Lin. ACCURATE LINEAR TECHNIQUE FOR CAMERA CALIBRATION CONSIDERING LENS DISTORTION BY SOLVING AN EIGENVALUE PROBLEM. *Optical Engineering* **32**(1), 138-149 (1993).
- Shirai. THREE-DIMENSIONAL COMPUTER VISION. Springer-Verlag, 1987.
- Takahashi, T. T. et al. REPRESENTATION OF MULTIPLE SOUND SOURCES IN THE OWL'S AUDITORY SPACE MAP. *Journal of Neuroscience*, **14**, no. 8, 4780-4793 (1994).
- Thurlow, W. R. et al. HEAD MOVEMENTS DURING SOUND LOCALIZATION. *Journal of the Acoustic Society of America*, **42**, 489-493 (1967).
- Tsai. A VERSATILE CAMERA CALIBRATION TECHNIQUE FOR HIGH-ACCURACY 3D MACHINE VISION METROLOGY USING OFF-THE-SHELF TV CAMERAS AND LENSES. *IEEE Journal of Robotics and Automation* RA -3 (4): 323-344.
- Tsai. AN EFFICIENT AND ACCURATE CAMERA CALIBRATION TECHNIQUE FOR 3-D MACHINE VISION. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 364-374 (1986).
- Wightman, F., Kistler, D.J. HEADPHONE SIMULATION OF FREE FIELD LISTENING I: STIMULUS SYNTHESIS. *Journal of the Acoustical Society of America*, **85**, 858-867 (1989).
- Wei, De, Ma. IMPLICIT AND EXPLICIT CAMERA CALIBRATION: THEORY AND EXPERIMENTS. *IEEE Transactions on PAMI*, **16**, no. 5 (1994).
- Wilson. FOURIER SERIES AND OPTICAL TRANSFORM TECHNIQUES IN CONTEMPORARY OPTICS. Wiley-Interscience. 1995
- Xu, Zang. EPIPOLAR GEOMETRY IN STEREO, MOTION AND OBJECT RECOGNITION: A UNIFIED APPROACH. Kluwer Academic Publishers, 1996.
- Zölzer, Udo. DIGITAL AUDIO SIGNAL PROCESSING. John Wiley & Sons, 1997.

En Internet

<http://www.iac.es/eav>

Espacio Acústico Virtual

Sitio web del Proyecto.

La idea

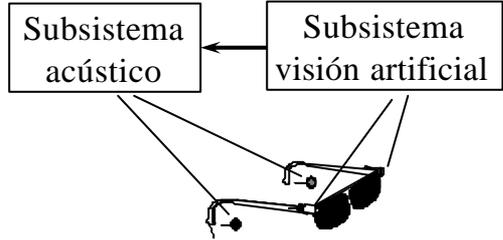
- Si el cerebro recibiese la información espacial que normalmente le proporciona la vista, pero a través del oído, ¿podría componer una imagen 3D del entorno?

El desarrollo tecnológico

- Un aparato de ayuda a la movilidad y percepción 3D para los ciegos

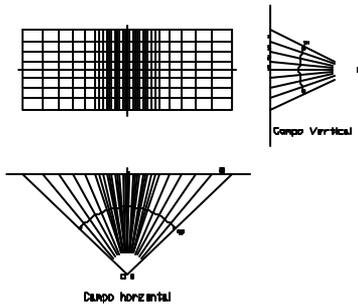
1

El desarrollo tecnológico. El concepto.



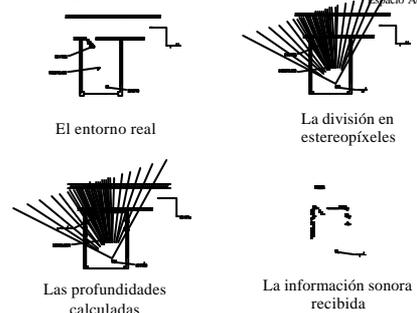
2

El estereopíxel



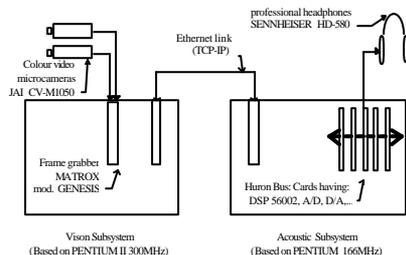
3

Ejemplo 2D



4

El prototipo de validación en laboratorio



5

Sala de control



Subsistema de generación/procesado de sonidos. A la izquierda, rack donde está ubicado el sistema Harman, con 40 salidas analógicas, 8 entradas analógicas y 8 DSPs 56002. Se ha utilizado para experimentos como el "maquillaje", para el sistema de medida de HRTFs, y para el subsistema de sonidos del prototipo. A la derecha, ordenador en el que se calculan las HRTFs y los sonidos especializados off-line.

6



Sala de experimentación, detalle



El prototipo portátil



Maqueta de las gafas



Maqueta de las gafas



Subsistema de visión

Captación del entorno y generación de un mapa de profundidades



Métodos abordados

- **Visión estereoscópica o estereovisión**
Dos imágenes de la misma escena tomadas desde dos posiciones diferentes
- **Profundidad a partir del desenfoque (DFD Depth From Defocus)**
Dos imágenes de la misma escena con diferente desenfoque

Visión estereoscópica. Operaciones.



Espacio Acústico Virtual

- **Calibración:** Obtención de los parámetros ópticos
- **Rectificación:** Transformación al caso ideal (cámaras paralelas, lentes sin distorsión)
- **Correspondencia:** Obtención de mapa de disparidades
 - Extracción de características
 - Correlación e interpolación
 - Filtrado
- **Cálculo de la profundidad:**

$$z = bf/d, \text{ donde } d = x_1 - x_2$$

13

Depth from Defocus. Resultados.



Espacio Acústico Virtual



Entorno real. Dos imágenes con distinto desenfoque.

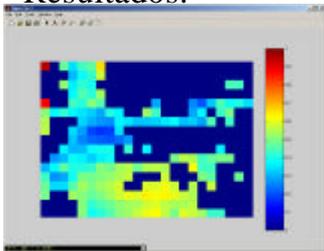
(Centre de Visió per Computador)

14

Depth from Defocus. Resultados.



Espacio Acústico Virtual



Distancias lejanas en azul. Medias en verde. Cercanas en rojo.

(Centre de Visió per Computador)

15

Subsistema acústico

Generación de un estímulo sonoro
formado por un mapa de sonidos 3D



Espacio Acústico Virtual

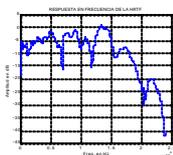
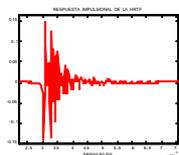
16

Respuesta impulsional y función de transferencia



Espacio Acústico Virtual

Respuesta impulsional: dominio del tiempo Función de transferencia: dominio de la frecuencia



17

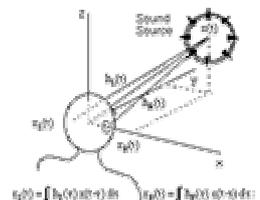
H.R.T.F. Definición.



Espacio Acústico Virtual

HRTF: Head Related
Transfer Function
(Función de
Transferencia Relativa
a la Cabeza)

Depende de la posición
relativa entre fuente
sonora y oyente

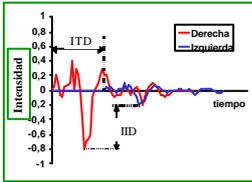


18



Información de azimut

RESPUESTAS IMPULSIONALES



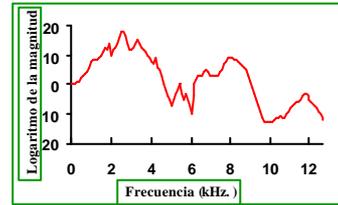
IID: Diferencia interaural de intensidad

ITD: Diferencia interaural de tiempo



Información de elevación

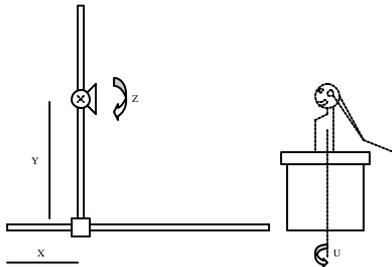
MAGNITUD DE LA HRTF



Crestas y valles en el dominio frecuencial



Stma automatizado de medidas de HRTFs



M.L.B.S. (Maximum Length Binary Sequences)

Entrada (M.L.B.S. de longitud L): $s(n)$ Salida: $y(n)$

Respuesta impulsional:

$$h(n) = \Omega_{sy}(n) = s(n) \Phi \quad y(n) = \frac{1}{L+1} \sum_{k=0}^{L-1} s(k) \cdot y(n+k)$$

Propiedad:

$$\Omega_{ss}(n) = s(n) \Phi s(n) \equiv \mathbf{d}(n)$$



Procesado de las medidas

L : Altavoz (más la cadena de reproducción)

F : HRTF

M : Micrófono (más la cadena de grabación)

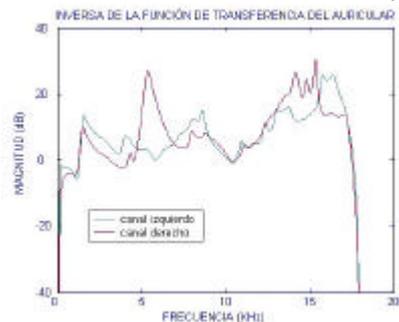
H : Auriculares (más la cadena de reproducción)

Filtro buscado:

$$T = \frac{G_{LM}}{G_{HM}} = \frac{L \cdot F \cdot M}{H \cdot M} = \frac{L \cdot F}{H}$$



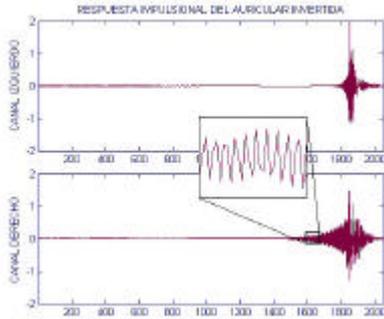
Ejemplo problemático



Ejemplo problemático



Espacio Acústico Virtual

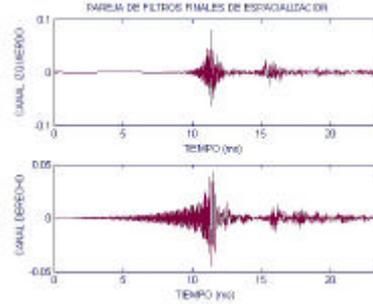


25

Ejemplo problemático



Espacio Acústico Virtual



26

Fast Hadamard Transform



Espacio Acústico Virtual

Problema con la MLBS: Su longitud es $L = 2^k - 1$
 \rightarrow FFT's de longitud $2^k - 1$

Reformulación del cálculo de $h(n)$ con la F.H.T.

• H_{L+1} : Matriz de Hadamard de orden $L+1$, relacionada con la M.L.B.S. $s(n)$

• P_i : Matrices de permutación

• S_i : Matrices de redimensionamiento

$$h(n) = \frac{1}{(L+1)s[0]} P_2 \{ S_2 \{ H_{L+1} [S_1 \{ P_1 y(n) \}] \} \}$$

27

HRTFs en plano horizontal



Espacio Acústico Virtual

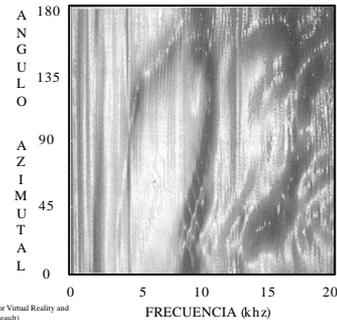


Foto tomada de "3D Sound for Virtual Reality and Multimedia" (Durand R. Begault)

28

HRTFs en plano medio



Espacio Acústico Virtual

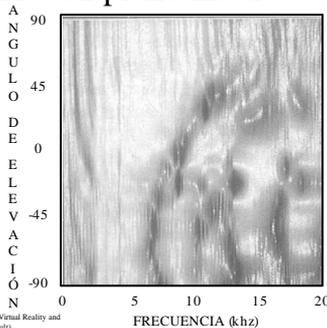


Foto tomada de "3D Sound for Virtual Reality and Multimedia" (Durand R. Begault)

29

Modelado de HRTF's



Espacio Acústico Virtual

- Objetivo: Obtener una representación funcional de las HRTFs dependiente de la frecuencia, posición espacial e individuo.
- Métodos recogidos en la bibliografía:
 - Representación de las HRTFs en función de formas espectrales básicas (funciones básicas)
 - Análisis de componentes principales sobre la magnitud de HRTFs de mínima fase (Wightman & Kistler, University of Wisconsin, Madison, 1991)
 - SFER (Spatial Features Extraction and Regularization) sobre magnitud y fase de las HRTFs (Chen, J. et al., University of Wisconsin, Madison, 1995).
 - Ajuste polinomial sobre los coeficientes (spline, lineal ...)

30



Modelado de HRTF's

- Resultados:
 - Utilización eficiente de la cantidad de memoria y potencia de cálculo disponibles.
 - HRTFs generales (no individuales)
 - Posibilidad de medir la similitud entre HRTFs de distintos individuos
 - distancia euclídeana entre coeficientes de las funciones base
 - análisis de escalado multidimensional
 - Interpolación espacial de HRTFs

PERCEPCIÓN GLOBAL DEL ENTORNO A TRAVÉS DE SONIDOS

PROYECTO ESPACIO ACÚSTICO VIRTUAL

LÍNEAS ACTUALES DE ACTUACIÓN ANTE LA CEGUERA

- PREVENCIÓN.
- RESTABLECIMIENTO DE LA VÍA visual NATURAL.
- RESTABLECIMIENTO DE LA VÍA visual POR PROTESIS SENSORIALES.
- APROVECHAMIENTO Y POTENCIACIÓN DE LOS **OTROS SENTIDOS**.

APROVECHAMIENTO DE LAS CAPACIDADES SENSORIALES INDEMNES

Es posible la realización de tareas propias de un sentido a través de otro u otros:
SUSTITUCIÓN SENSORIAL.

- Orientación y movilidad.
- Desarrollo del niño.
- Lectura.
- Reconocimiento de objetos.
- Interacción hombre - máquina.

PERCEPCIÓN DEL ENTORNO A TRAVÉS DE SONIDOS

OBJETOS SONOROS

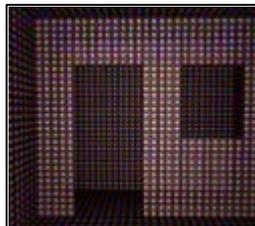
- Localización.
- Identificación.
- Percepción de objetos en su posición y forma.

OBJETOS NO SONOROS

- Dimensiones del entorno.
- Ecolocación.
- Bastón, cascabeles, etc.
- **Dispositivos electrónicos:**
 - Detectores de obstáculos.
 - Codificación de escenas.

PERCEPCIÓN DE SUPERFICIES SONORAS

- La hipótesis.
- Fundamentos:
 - **Procesamiento amodal**
 - **Reproducción de estrategia.**
 - **Condiciones adecuadas de la audición.**
 - **Precedentes.**



ABORDAJE EXPERIMENTAL DE LA HIPÓTESIS

- Estudio con fuentes sonoras reales
- Estudio con fuentes sonoras virtuales



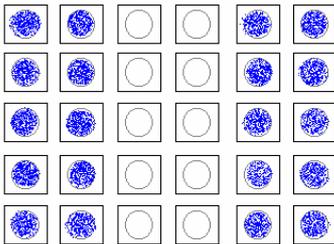


MATRIZ DE ALTAVOCES

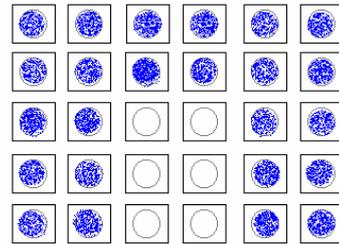
FUENTES SONORAS REALES: EXPERIMENTO MULTIFUENTE

Valoración de la percepción de figuras en el espacio a partir de superficies sonoras.

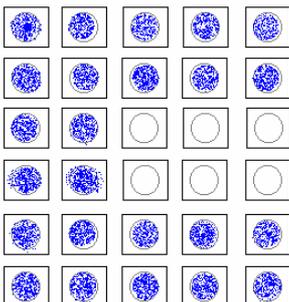
- Cada imagen definida por el conjunto de puntos de la superficie o escena a percibir.
- Imágenes cada 30, 90 y 300 ms.
- Sonido muy localizable.
- Secuencia espacial de presentación al azar.



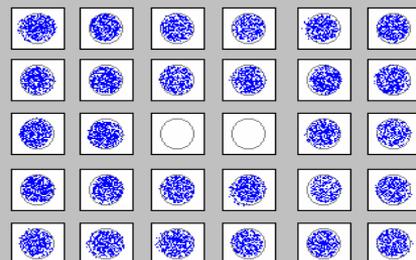
MATRIZ SONORA: BARRAS VERTICALES



MATRIZ SONORA: U INVERTIDA

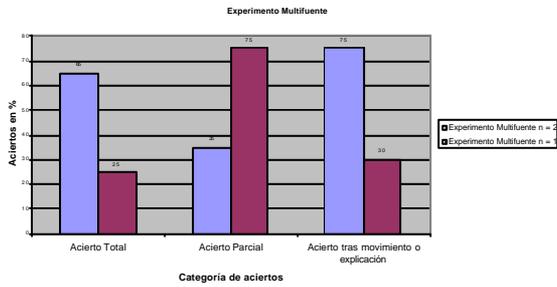


MATRIZ SONORA: C A LA DERECHA



MATRIZ SONORA: VENTANA ABIERTA

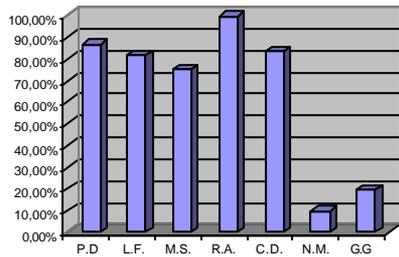
RESULTADOS PRELIMINARES



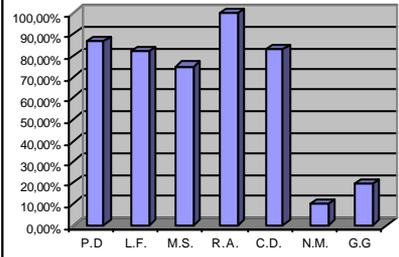
Resultados



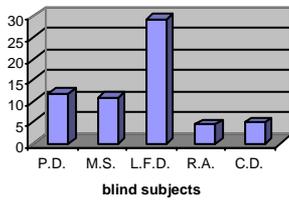
Percentage of correct responses about the shape of the figures



Percentage of correct responses about the shape and orientation of the figures



Mean error in the perception of the dimensions of the figures (cm)



Blind person pointing to two points horizontally located. Both them are perceived simultaneously present

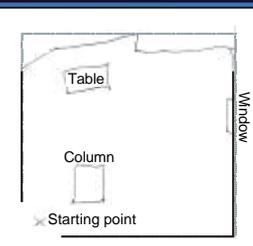
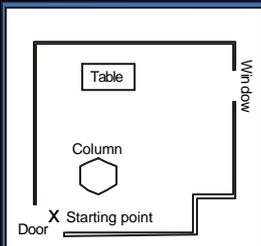
Resultados



- Detección del entorno
- Reconocimiento de formas con sus tamaños
- Reconocimiento de la habitación con su contenido

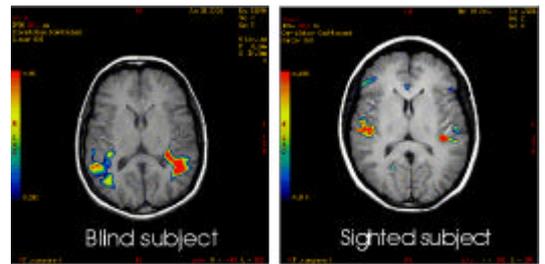


Resultados



Schematic representation of the experimental room, with a particular objects distributions.

Drawing made by a blind person after a very short exploration, using the prototype, without relying in touch



Imágenes de Resonancia Magnética Funcional mostrando las áreas cerebrales implicadas en la localización de una fuente sonora virtual aislada

El futuro



- Continuar el desarrollo del programa de investigación optimizando sonidos, vision artificial, y capacidad del sistema
- Poner a punto el prototipo portátil desarrollado, con voluntarios ciegos.
- Realizar un programa de transferencia tecnológica para el desarrollo de un prototipo comercial.

Reconocimientos

- Gobierno de las Islas Canarias
- Ministerio español de Asuntos Sociales
- Comunidad Europea
- Organización Nacional de Ciegos de España (ONCE)
- Ministerio Español de Ciencia y Tecnología
- Instituto Tecnológico de las Islas Canarias
- Hospital universitario de Canarias
- ImasDe Canarias SA
- Pixel SL

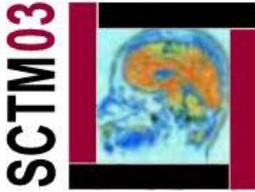


El equipo

Gracias por su atención.

www.iac.es/eav

Radiaciones y Vida



José Hernández Armas

Catedrático de Radiología y Medicina Física y Director del Laboratorio de Física Médica y Radiactividad Ambiental del Departamento de Medicina Física y Farmacología de la Universidad de La Laguna.
Jefe del Servicio de Física Médica y Protección Radiológica del Hospital Universitario de Canarias.

1. Introducción

Los seres vivos han recibido siempre radiación debida a fuentes naturales, tales como la radiactividad en suelo y en alimento o bebida, así como rayos cósmicos procedentes del espacio exterior. Toda la historia de la humanidad ha estado presidida por desarrollarse permanentemente en presencia de radiaciones.

Las radiaciones artificiales ocasionan una exposición adicional sobre la anterior, dando lugar a un incremento de la que recibimos desde la naturaleza. Como resultado de la vida moderna, millones de personas cada año vuelan en aviones comerciales recibiendo por ello una exposición extra a los rayos cósmicos.

El **uso** de la radiación forma actualmente parte integral de nuestra vida. En particular, las aplicaciones médicas de la radiación se consideran como algo común, abarcando tanto el diagnóstico radiológico como el tratamiento del cáncer.

El uso de la radiación en medicina y en la industria ocasiona, sin duda, resultados tremendamente beneficiosos para la sociedad. Al mismo tiempo, se puede asegurar que una dosis excesiva de radiación puede tener consecuencias desastrosas sobre la salud y el bienestar. La radiación puede causar enfermedades, y curarlas también. ¿Qué nivel de riesgo estamos dispuestos a aceptar por determinados beneficios?

El **miedo** a la radiación ocasiona las respuestas más irracionales de personas que, en los demás temas, responden de manera tranquila e inteligente. De alguna forma es comprensible, ya que la radiación es invisible y produce cambios biológicos sutiles que normalmente no son observables hasta pasados muchos años. Por otra parte, la radiación suele asociarse inmediatamente con bombas, lluvia radiactiva, muerte y destrucción. **El conocimiento es el mejor antídoto contra el miedo.**

En muchos ámbitos de la Medicina se están aplicando directa o indirectamente los resultados de usar radiaciones ionizantes sobre los pacientes. Este uso tiene su exponente principal en el **Radiodiagnóstico**, pero también las radiaciones constituyen la base de la especialidad médica **Medicina Nuclear**. Además, las radiaciones constituyen un arma esencial, a veces única, en el tratamiento de diversos tipos de cáncer (**Radioterapia**). Todos estos usos se fundan en los incontables beneficios para la realización de diagnósticos o para la ejecución de tratamientos de enfermedades. Como quiera que, de manera indefectible, el uso siempre va acompañado de efectos lesivos, es imprescindible que la utilización se realice respetando principios de **Protección Radiológica** que eviten, dentro de lo razonablemente posible, los efectos indeseados.

Cualquiera que sea el tipo de utilización, el efecto buscado es siempre una consecuencia de la interacción entre la radiación ionizante y la materia, particularizando habitualmente para la materia viva. Todos los procesos de interacción radiación-materia se expresan en el lenguaje de

las matemáticas. El uso de las matemáticas en los ámbitos clínicos de utilización de las radiaciones es una característica diferenciadora entre los mismos y los demás ámbitos de la práctica médica.

El principal propósito de esta ponencia es exponer un conjunto de ejemplos que pongan de manifiesto algunos aspectos de las matemáticas (especialmente el cálculo) que han sido y son útiles para las personas que trabajan con radiaciones electromagnéticas en Medicina.

2. Matemáticas en el uso terapéutico de radiaciones (Radioterapia)

El proceso de planificar un tratamiento en Radioterapia implica la elección y disposición de haces de radiación alrededor del paciente, a fin de proporcionarle una alta dosis en el lugar donde se le haya desarrollado un tumor y una dosis tan baja como sea posible en los demás tejidos.

Veamos algún ejemplo concreto de aplicación sencilla: *haces únicos*.

En muchos casos de leucemia, el paciente se trata con un “baño” de radiación a todo el cuerpo. En estas circunstancias no hay un lugar específico para el tumor, de forma que no puede especificarse una dosis de radiación determinada para el tumor.

Se trata de proporcionar una dosis uniforme a todo el cuerpo. Esto puede hacerse mediante uno o dos grandes campos de radiación o con diferentes combinaciones de campos más pequeños, administrados bien secuencialmente o usando un número de fuentes simultáneamente.

Veamos un caso concreto.

Se usa un gran haz como parte de un tratamiento de irradiación corporal total. La superficie anterior del paciente se encuentra a una distancia de $2.25m$ de la fuente de radiación. El tamaño del haz a esta distancia es de $70 \times 50cm$. La tasa de dosis, medida en el espacio libre en una pequeña masa de tejido a una distancia de $80.5cm$, es de $1Gy$ por minuto. Hay que establecer la dosis máxima debida a este haz, la dosis en la línea media del paciente y la dosis a la salida por la espalda del paciente, cuyo grosor es de $24cm$, teniendo en cuenta que se usa el mismo haz para un campo posterior y que cada haz se aplica durante 1 minuto.

Para el cálculo de la tasa de dosis en el espacio libre inmediato a la superficie del paciente, la aplicación de la ley de la inversa del cuadrado de la distancia conduce a:

$$1Gy/min (80.5/225)^2 = 0.128Gy/min$$

La relación tejido-aire (relación de la dosis en un punto del tejido y la dosis en un punto del aire situado a la misma distancia de la fuente) para este campo es de 1.09 (tablas). Por tanto, la tasa máxima de dosis es igual a $0.128 \times 1.09 = 0.140Gy/min$.

Como el paciente tiene un grosor en el tronco de $24cm$, la dosis en el punto medio en el espacio libre es: $1Gy/min (80.5/237)^2 = 0.115Gy/min$. Ya que la relación tejido-aire para una profundidad de $12cm$ es 0.81 (tablas), la tasa de dosis en línea media es: $0.115 \times 0.81 = 0.093Gy/min$.

La tasa de dosis en el espacio libre en la superficie de salida es de $1Gy/min (80.5/249)^2 = 0.104Gy/min$, y como la relación tejido-aire es 0.53, la tasa de dosis en la superficie de salida es de $0.104 \times 0.53 = 0.055Gy/min$. Así, la dosis en la superficie (a $0.5cm$) es: $0.104 + 0.055 = 0.159Gy$. La dosis en línea media es: $0.093 + 0.093 = 0.186Gy$.

Los cálculos de las planificaciones de los tratamientos radiológicos se han ido haciendo progresivamente más complicados, hasta el punto de que hoy ya no puede concebirse el efectuarlos de la manera anterior, más que en casos muy concretos. Estos cálculos se realizan

actualmente mediante complejos programas informáticos que se alimentan con los datos proporcionados por simulaciones Monte Carlo de los procesos de interacción de las radiaciones con los átomos y moléculas constituyentes del cuerpo humano.

En el Hospital Universitario de Canarias se utiliza un sistema de planificación de tratamientos radiológicos en el que hay que suministrar, lógicamente, los datos correspondientes a la configuración del paciente, los cuales se obtienen bien a través de imágenes obtenidas en un tomógrafo computerizado o bien mediante medidas y tomas de contornos realizadas sobre el propio paciente.

Uno de los métodos de cálculo de la dosis absorbida está basado en el conocimiento de los valores porcentuales de dosis en profundidad. En esencia:

(Dosis absorbida en un paciente) =
 (Rendimiento del equipo irradiador) × (%Dosis en profundidad) × (Tasa de dosis normalizada al eje del campo) × (Factor de corrección por filtro del haz de radiación) × (Factor de conversión Gy/Unidades monitor) × (Nº Unidades monitor):

$$D(x, y, z) = S_{c,p}(v_c, v_p | filter) * PDD(v_p, z, SSD) \\ * R_{oa}(x, y, z, v_z) * C_{filter} * R_{cGy/MU} * N_{MU}$$

donde z indica la profundidad del punto, S es el factor relativo para dar la radiación producida a la profundidad z , v_c es el tamaño del campo en la superficie de agua o a la distancia fuente-superficie, y v_p es el tamaño del campo en el isocentro o distancia fuente-eje.

Los parámetros a considerar son:

- Distancia fuente-superficie del paciente (SSD)
- Distancia fuente-isocentro del equipo (SAD)
- % Dosis en profundidad (PDD)
- Factor pico dispersión del haz (PSF)
- Factor de conversión Gy/UM : R

Cuando SSD no es igual a SAD es necesario corregir PDD y R. Para ello hay que aplicar la ley de la inversa del cuadrado de la distancia y los factores pico de dispersión:

$$PDD(v_p, z, SSD) = F(z) * \frac{PSF(v_F)}{PSF(v_p)} * PDD(v_F, z, SSD_m)$$

$$F(z) = \left(\left[\frac{SSD_m + z}{SSD + z} \right] * \left[\frac{SSD + d_{max}}{SSD_m + d_{max}} \right] \right)^2$$

$$v_F = \frac{v_p}{\sqrt{F(z)}}$$

$$R_{cGy/M}(SSD) = \left(\frac{SSD_m + d_{max}}{SSD + d_{max}} \right)^2 * R_{cGy/M}(SSD_m)$$

Las sencillas expresiones anteriores no tienen en cuenta situaciones muy frecuentes en la práctica clínica: inhomogeneidades y campos irregulares, ya que esto ocasiona cambios en el comportamiento de los componentes del haz de radiación. Asimismo, la incidencia oblicua del haz es muy difícil de introducir con métodos de corrección adecuados.

El cálculo habitual se realiza utilizando un algoritmo que hace uso de la teoría del haz pincel. En ella, el haz de tratamiento se divide en pequeños haces, llamados haces pincel. Cada pequeña sección del haz se considera constituida por tres finos haces pincel:

- el correspondiente a la radiación primaria
- el correspondiente a la radiación dispersa simple
- el correspondiente a la radiación dispersa múltiple

En este modelo, la distribución transversal de la dosis para cada uno de los tres haces pincel se modela mediante una función gaussiana $K=K_\sigma$, llamada kernel del haz pincel. El subíndice sigma establece el ancho de la gaussiana: es diferente para cada haz pincel y así, el haz pincel primario tendrá una menor sigma que los haces pincel dispersos.

La dosis para un campo irregular puede escribirse en términos del kernel del haz pincel mediante la expresión

$$D_{PBI}(x, y, z, A) = N(x, y, z) \iint_{A'} a(x', y') * PDD(x', y', z) * K(x - x', y - y', z) dx' dy'$$

La ecuación expresa la dosis debida a fotones para cualquier situación clínica. N es un factor de conversión para expresar la dosis real. Es igual para todos los puntos situados a una profundidad dada en caso de incidencia normal del haz. Todos los haces se ponderan con el valor del porcentaje de dosis en profundidad para cada profundidad. El factor $a(x', y')$ permite corregir la situación en que los haces de radiación pasen a través de bloques o inhomogeneidades o en caso de incidencia oblicua.

La función kernel puede escribirse en función de la fluencia de fotones en la superficie, W :

$$K(x - x', y - y', z) = W(r') K_p(x - x', y - y', z) + R_1(z) K_{s_1}(x - x', y - y', z) + R_2(z) K_{s_2}(x - x', y - y', z) + \dots$$

Tras un cambio de variable y aplicando los límites de integración correctos, la dosis se calcula mediante la expresión:

$$D_{PBI}(x, y, z, ref) = N(z) * PDD(z) \frac{1}{2\pi\sigma^2} \int_{(-\frac{1}{2}v-x)/\sigma_p(z)}^{(\frac{1}{2}v-x)/\sigma_p(z)} e^{-t^2/2} dt \int_{(-\frac{1}{2}v-y)/\sigma_p(z)}^{(\frac{1}{2}v-y)/\sigma_p(z)} e^{-t^2/2} dt$$

Las integrales pueden reescribirse en términos de funciones de error:

$$\begin{aligned} D_{PBI}(x, y, z(x', y')) &= N(x, y, z) * PDD(x', y') * a(x', y') * \frac{1}{2\pi\sigma^2} \int_{(x_1-x)/\sigma_p(z)}^{(x_2-x)/\sigma_p(z)} e^{-t^2/2} dt \int_{(y_1-y)/\sigma_p(z)}^{(y_2-y)/\sigma_p(z)} e^{-t^2/2} dt \\ &= N(x, y, z) * PDD(x', y') * a(x', y') * \left[\text{erf}\left(\frac{x_2 - x}{\sigma_p(z)}\right) - \text{erf}\left(\frac{x_1 - x}{\sigma_p(z)}\right) \right] * \\ &\quad \left[\text{erf}\left(\frac{y_2 - y}{\sigma_p(z)}\right) - \text{erf}\left(\frac{y_1 - y}{\sigma_p(z)}\right) \right] \end{aligned}$$

Esto significa que dentro de la pequeña región se hace la hipótesis de que hay una cantidad infinita de haces pincel, todos con las mismas características. La ecuación establece solamente la contribución de la región $a \times b$ con centro en (x', y') al punto $P(x, y, z)$. Para calcular la dosis total en el punto P , debe sumarse la contribución de todas las regiones.

3. Matemáticas básicas en la práctica diaria de la Medicina Nuclear

Puede definirse la especialidad médica de la Medicina Nuclear como aquella que mediante el uso de sustancias radiactivas no encapsuladas obtiene información diagnóstica de pacientes mediante el análisis de la distribución de aquellas sustancias en el cuerpo del paciente.

El personal médico y sanitario ha de estar familiarizado con los conceptos básicos del fenómeno físico que utiliza: la desintegración radiactiva, que está regido por la ley general:

$$A = A_0 e^{-\lambda t}$$

de la que se deriva el periodo de semidesintegración $T_{1/2} = 0.693/\lambda$ y el de vida media, igual a $1.44T_{1/2}$.

La cantidad de núcleos radiactivos que experimentan transiciones en un periodo de tiempo se llama actividad acumulada, que se expresa en la unidad de actividad, Becquerelios, por la unidad de tiempo hora. Es inmediato tener en cuenta que:

$$\text{Actividad acumulada (Bq-horas)} = A \text{ (Bq)} \times T_{1/2} \text{ (horas)} \times 1.44$$

La calidad de las imágenes que se obtienen cuando se registra la actividad radiactiva administrada a un paciente está en general relacionada con la actividad, mientras que la energía depositada en el paciente depende de la actividad acumulada. Por ello, la relación anterior es manejada continuamente por el personal médico y de Radiofísica Hospitalaria, pues establece que un radionúclido con una vida media corta tiene más actividad con menor material radiactivo, lo que suele ser deseable para estudios clínicos puesto que la relación (calidad de imagen / dosis al paciente) es inversamente proporcional a la vida media.

La utilización adecuada de materiales radiactivos administrados al paciente y la interpretación de los resultados consiguientes implica el apoyarse en modelos matemáticos capaces de dar una idea de lo que ha ocurrido con el material radiactivo dentro del cuerpo del paciente. De manera muy simplificada se puede hablar de un modelo de compartimentos, tanto cuando se considera un compartimento cerrado como cuando se trata de uno abierto.

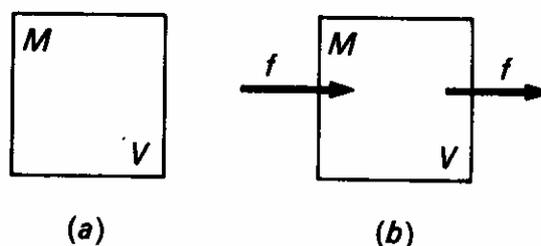


Figura 1. Modelos de compartimentos: (a) cerrado; (b) abierto.

Modelo de compartimento cerrado. Considérese un recipiente cerrado con un volumen V de agua. En él se disuelve una masa M de una sustancia. Se introduce un trazador radiactivo de actividad radiactiva R conocida y masa despreciable. Supóngase reparto homogéneo e inmediato del trazador.

Se toma una muestra de volumen v , que contiene una masa m . La radiactividad en la muestra es r .

Como el reparto del trazador es homogéneo, la concentración es igual en todo el recipiente y se tiene: $r/v = R/V = \omega$; o bien por unidad de masa: $r/m = R/M = x$.

Este sencillo método se utiliza para determinar el volumen sanguíneo inyectando hemafís marcados con Cr-51. También el agua total del organismo, inyectando agua tritiada.

Desgraciadamente, un compartimento biológico nunca es totalmente cerrado, y se puede llegar a resultados absurdos si se desprecian los intercambios metabólicos entre compartimentos. Así, por ejemplo, se administran a un sujeto normal 100 microcurios de yoduro marcado (con I-131). Se mide la radiactividad plasmática al cabo de 24 horas y se obtienen 0.25 microcurios por litro. Entonces, según la ecuación de dilución, el volumen de difusión sería de 400 litros, lo que es absurdo. El modelo es incorrecto, pues no ha tenido en cuenta la fijación del yodo en el tiroides.

$$\boxed{\frac{r}{v} = \frac{R}{V} = \omega} \quad (\text{concentración radiactiva})$$

$$\boxed{\frac{r}{m} = \frac{R}{M} = x} \quad (\text{radiactividad másica})$$

Figura 2. Modelo de compartimentos: ecuación de dilución.

Modelo de compartimento abierto. Consideremos el mismo recipiente que antes, pero ahora en comunicación con el exterior. Entra y sale de él por unidad de tiempo la cantidad f de materia; f es una constante, al igual que la masa M : hay equilibrio entre las entradas y las salidas (estado de equilibrio dinámico).

El valor de f depende al mismo tiempo de la masa M del recipiente y de la cinética del intercambio: una misma tasa f puede corresponder a un intercambio rápido si M es pequeña, o a un intercambio lento si M es grande. Por esta razón se define otra magnitud más directamente cinética: la tasa de renovación $k = f/M$.

Las cantidades M , V , k y f caracterizan al modelo con independencia del trazador. Durante un tiempo t entra y sale la cantidad ft , es decir, una fracción kt de la masa del compartimento.

La introducción en el compartimento de un trazador de actividad R_0 en el instante $t=0$ va seguida de una eliminación progresiva, y la actividad R del compartimento disminuirá a lo largo del tiempo. Supóngase que la mezcla es inmediata y homogénea.

La actividad que sale del compartimento durante el tiempo dt es $-dR$. Esta actividad será proporcional a f y a la actividad específica $x = R/M$.

Cuando sale del compartimento la actividad dR , la actividad y por tanto la actividad másica disminuyen en el compartimento y se tiene $-dx = -dR/M$, de donde:

$$-(dx/dt) = kx.$$

La solución de esta ecuación diferencial es $x = x_0 e^{-kt}$, y análogamente se obtendrá para R y para ω .

Después de haber introducido la actividad R_0 en un compartimento se procede a hacer medidas sucesivas de la actividad másica x en muestras tomadas en el transcurso del tiempo y se obtienen resultados experimentales que se sitúan sobre una recta descendente en coordenadas semilogarítmicas, lo que permite deducir que es correcta la suposición de modelo de compartimento abierto.

En la recta, el tiempo que separa dos puntos x_1 y $x_1/2$ se llama periodo biológico, que está relacionado con la tasa de renovación por la expresión $K = 0.693/T_b$.

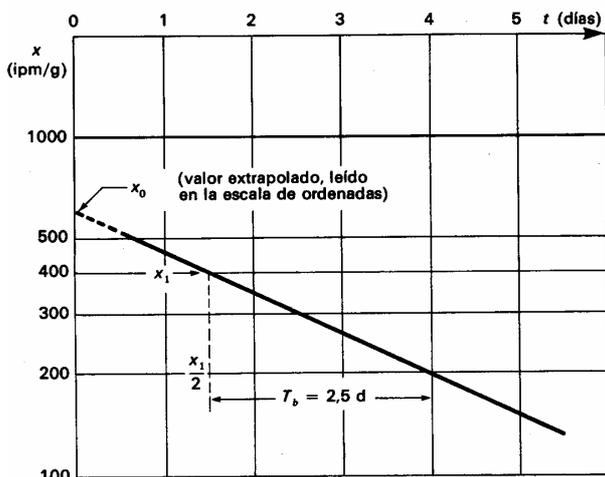


Figura 3. Curva experimental de un compartimento abierto.

Al cabo del tiempo T_b la cantidad de materia que ha salido es fT_b , es decir, kMT_b o $0.693M$. Esta materia ha sido reemplazada por la misma cantidad que ha entrado en el compartimento. Al cabo de un tiempo igual a $1/k = 1.44T_b$ ha entrado y salido una cantidad de materia igual a la masa M del compartimento. A este tiempo se le llama tiempo de renovación. Entonces, conociendo x_0 o ω_0 se puede determinar, por la ecuación de dilución, o bien M , o bien V .

En Biología es necesario con frecuencia utilizar modelos más complejos. Son frecuentes los modelos catenario y mamilar.

Si en el instante $t=0$ se inyecta la radiactividad R_0 en el primer compartimento, se puede determinar la ley de crecimiento de $R_1=f(t)$ que, en general, es una suma de exponenciales decrecientes, del tipo

$$R_1 = K_1 e^{-\lambda_1 t} + K_2 e^{-\lambda_2 t}.$$

Los parámetros K_1 , K_2 , λ_1 y λ_2 dependen de M_1 , M_2 , f_1 , f_2 , f_e y f_s .

Los sistemas fisiológicos corresponden con frecuencia a modelos más complejos, y los resultados experimentales son difíciles de analizar e interpretar.

Con demasiada frecuencia el compartimento se asimila a una entidad anatómica o a un conjunto funcional dado: en algunos casos pueden coincidir, pero no tiene que ser así obligatoriamente. En realidad, un compartimento es un conjunto de moléculas que tienen la misma probabilidad de experimentar un suceso dado.

En la práctica médica habitual sólo sirven los modelos simples en los que la ecuación de dilución es inmediatamente utilizable; los modelos más complejos son muy útiles para conocer el metabolismo de numerosas sustancias, pero no se emplean en su forma completa.

Medida de las tasas. Si se inyecta bruscamente la actividad R en una vena, el producto radiactivo se difunde y “se extiende”, mientras que el conjunto del flujo sanguíneo se desplaza hacia el contador colocado, por ejemplo, frente a las cavidades cardiacas.

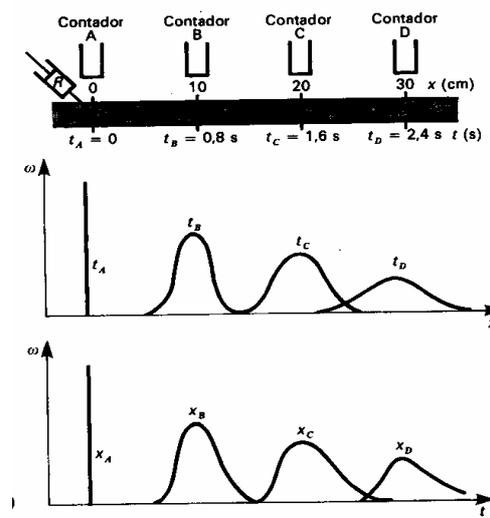


Figura 4. Medida del flujo cardíaco: posición de los contadores a lo largo de la vena; difusión de la radiactividad; registros temporales.

Durante el tiempo dt , el contador registra la actividad dR del volumen dV que pasa ante él. Se tendrá entonces:

$$dR/dt = (dR/dV)(dV/dt).$$

Como $dR/dV = \omega$ y $dV/dt = F$ (tasa sanguínea, en este caso tasa cardíaca), entonces:

$$dR = F \omega dt.$$

La actividad total registrada por el contador durante el paso de la zona radiactiva está dada por:

$$R = \int_0^\infty dR = F \int_0^\infty \omega dt,$$

así que

$$F = R / \int_0^\infty \omega dt \quad (\text{ecuación de Stewart-Hamilton}).$$

Si el contador puede medir ω en cada instante, la integral permite calcular entonces F .

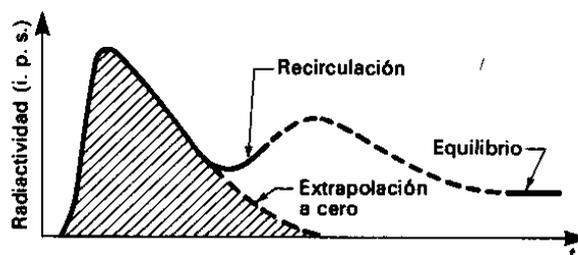


Figura 5. Medida del flujo cardiaco. Curva registrada al nivel de un contador.

Dosis a pacientes debidas a radionúclidos administrados. Se trata de calcular las dosis a tejidos en un órgano determinado debidas a material radiactivo que se encuentra en ese órgano o en otro órgano del cuerpo. Estos cálculos son importantes, ya que es esencial para la salud presente y futura del paciente conocer las dosis a órganos críticos debidas a la actividad situada en algún órgano.

El llamado comité MIRD (*Medical Internal Radiation Dose*) ha asumido, desde 1968, la responsabilidad de crear tablas de manera que los cálculos de dosis puedan hacerse para los radionúclidos usados.

El punto de partida para cualquier cálculo de dosis es el disponer de información detallada sobre los factores físicos implicados en la desintegración de los radionúclidos incluyendo los tipos de “partículas” emitidas, sus números relativos y la energía. Pueden observarse los datos correspondientes dados por MIRD 10 sobre el Tc-99m. A partir de estos datos se puede calcular la energía emitida por desintegración.

4. Ejemplos de uso de las matemáticas en Radiodiagnóstico

El Radiodiagnóstico se ha definido como la obtención de imágenes del interior del organismo humano mediante la utilización de radiaciones. Actualmente, esta definición es incompleta ya que dentro del Radiodiagnóstico se encuentra también la obtención de imágenes con otros agentes físicos, tales como los ultrasonidos o los campos magnéticos. Actualmente, el pujante uso de las radiaciones ionizantes en Medicina ha dado lugar a una nueva especialidad que, además de permitir la realización de diagnósticos mediante la utilización de catéteres guiados por rayos X, realiza también innumerables intervenciones con finalidad terapéutica que han suplantado muchas intervenciones quirúrgicas agresivas que tenían que llevarse a cabo en el pasado para alcanzar el mismo objetivo: estamos hablando de la Radiología Vasculare Intervencionista, dentro de la que existe una subespecialidad que es la Cardiología Hemodinámica Intervencionista.

Es conocido que el Radiodiagnóstico interviene en una u otra forma en más del 70% de los diagnósticos médicos. Es aún más conocido que esta especialidad médica se desarrolló casi inmediatamente después de que el Profesor Rontgen comunicó, en diciembre de 1895, que había descubierto “una nueva clase de rayos”. El desarrollo meteórico de la utilización de los rayos X en Medicina no tiene comparación con ningún otro agente físico o químico que se utilice en la práctica médica. Desde 1896 hasta 1973 el Radiodiagnóstico podía definirse como la obtención

de una imagen radiante en una película radiográfica constituida por un soporte plástico impregnado con pequeños cristales de haluro de plata. Este tipo de Radiodiagnóstico aún está en vigor, y constituye todavía una buena parte del conjunto de los usos diagnósticos de los rayos X. Pero en la última fecha indicada se comenzó la utilización de un equipo técnico inventado por el Dr. Hounsfield, científico e ingeniero inglés. Ese equipo se denominó inicialmente Tomografía Axial Computerizada, y actualmente TC. Fue la irrupción de la imagen digital en el Radiodiagnóstico, que ha tenido desde entonces un crecimiento continuo. Es notorio el hecho de que la Radiología Vascular e Intervencionista ya no se concibe más que en un entorno de imágenes digitales. Y esta es también la situación de las imágenes de Ultrasonografía o las de Resonancia Magnética.

Se comprenderá que todo este desarrollo no puede entenderse sin la participación activa de las matemáticas: tanto para la descripción precisa del proceso de interacción agente físico-materia como en el proceso de generación de imágenes digitales. Es imposible en el espacio de estas líneas de hacer un recorrido por todo este panorama de tecnología aplicada a la medicina. Sin embargo, espero someter a la consideración del lector algunos ejemplos prácticos que ilustren cómo el cálculo matemático está presente en el día a día del uso clínico de los rayos X.

Radiodiagnóstico general clásico. Cuando se usa una película en Radiodiagnóstico, casi siempre se coloca entre dos pantallas fluorescentes que responden principalmente a la luz emitida por la sustancia fluorescente. El ennegrecimiento de la película debida a la acción directa de los rayos X es normalmente menos del 2% del efecto total.

Después de haberla procesado, la película se observa colocándola enfrente de un negatoscopio uniformemente iluminado, donde la sección de la película expuesta a la radiación se observa negra, y la no expuesta, blanca. Si el brillo de la luz del negatoscopio es B_0 y la cantidad transmitida es B , entonces la densidad de la película viene definida por $D = \log(B_0/B)$, o bien $B = B_0 e^{-2.303D}$. Si la película transmite 1/10 de la luz incidente, la densidad es 1 ($B_0/B = 10$).

La razón para esta elección de definición de densidad se explica considerando que si una película con densidad 1.0 transmite 1/10 de la luz incidente, dos de estas películas transmitirán $0.1 \times 0.1 = 0.01$ de la luz, y entonces tendrá una densidad 2. Dos de estas películas tienen el doble de cantidad de granos de plata reducida que una película simple. Por tanto, es razonable expresar la densidad combinada de las películas como 2.0. De esta forma la densidad es proporcional a la cantidad de plata procesada en la película.

La cantidad de granos de haluro de plata reducidos a plata metálica es directamente proporcional a la cantidad de radiación que alcanza a la película. Dicha cantidad viene dada por el miliamperaje de la corriente electrónica que generó el haz de rayos X y por el tiempo que estuvo actuando: es lo que suele expresarse como el *mAs* del haz de rayos. Asimismo, la energía depositada por la radiación en la unidad de masa, es decir, la dosis, es directamente proporcional a la cantidad de radiación.

Cada día se producen unas 200000 imágenes médicas, lo que implica que un número similar de decisiones se toman basadas en la calidad de estas imágenes.

La calidad de la imagen radiográfica significa cuantificar la perfección de la misma, para lo que hay que establecer los valores que en ella toman los siguientes parámetros:

- Contraste
- Borrosidad
- Ruido

El término contraste está directamente ligado a la diferencia de los coeficientes lineales de atenuación de los medios atravesados por la radiación. El contraste radiográfico se utiliza para describir la diferencia de densidad óptica entre dos áreas de una radiografía. En muchos casos, un contraste radiográfico alto mejora la capacidad de diferenciar estructuras del objeto en la radiografía. El contraste radiográfico es el resultado de dos factores: el contraste de la imagen y el contraste de la película o sistema receptor de la imagen.

Se denomina curva característica de una película a la curva que representa los valores de densidad óptica medida para cada zona de la película frente al logaritmo de la exposición relativa de dicha zona (en mR). Puede considerarse como una descripción de la respuesta del sistema de registro de la imagen (en este caso película) a unas ciertas condiciones de entrada. El contraste de la película está asociado con la pendiente de la curva característica; mayor pendiente significa película más sensible o más rápida, es decir, que requiere menos exposición.

Las curvas características son las curvas estándar usadas para presentar las propiedades de una película. Estas curvas tienen cuatro regiones importantes:

La porción plana en el extremo izquierdo, donde la densidad es independiente de la exposición, es debida a la densidad de la película base más el velo de fondo: hay densidad tanto si la película ha sido expuesta como si no. El velo tiende a aumentar si la película es vieja o si se almacena a temperaturas altas.

En la siguiente región de la curva la densidad aumenta rápidamente con la exposición. Esta región va seguida por una porción recta donde la densidad aumenta linealmente con el logaritmo de la exposición. Finalmente la curva comienza a aplanarse a alta exposición, se produce saturación y región de exposición correcta.

La velocidad de una película puede determinarse hallando el valor de la exposición necesaria para producir una densidad de 1 mayor que la densidad de la base más velo.

La sensibilidad de estas películas es la recíproca de estas exposiciones. Todo lo dicho para películas es válido para combinaciones películas-pantallas intensificadoras.

Lo frecuente es usar películas en la parte lineal del gráfico, con densidades comprendidas entre 0.4 y 2.0.

La mínima diferencia en exposición detectable dependerá de la pendiente de la curva característica. La pendiente de la porción recta de la curva se suele denominar la gamma de la película: $\text{Gamma} = (D_2 - D_1) / \log(X_2/X_1)$, correspondiendo los subíndices a los valores de dos puntos situados en dicha zona recta.

La exposición a la radiación debe ser tal que todas las partes de la radiografía se encuentren en la porción recta de la curva característica, lo cual asegurará el mismo contraste para todas las densidades.

El término borrosidad va ligado al grado de detalle observable en una imagen registrada. La borrosidad tiene tres causas principales: tamaño del foco productor de radiación; sistema de detección de radiación; movimiento de la estructura radiografiada. Estas tres causas de imprecisión suelen cuantificarse mediante el uso de un dispositivo de prueba, llamado habitualmente modelo de barras. Pero existe un método para describir la borrosidad de la imagen de una forma precisa y práctica e independiente del observador. Consiste en analizar las características de la imagen de la línea producida por un objeto de prueba con una hendidura. El gráfico que representa los resultados de la exploración de la imagen con un microdensitómetro, es decir, la distribución espacial de la densidad óptica en la imagen de la ranura, se conoce como función de dispersión lineal.

Una forma de obtener más información sería utilizar la dada por la función de dispersión lineal y aplicarle una transformada de Fourier. El resultado es la función de la

transferencia de modulación, que representa la capacidad de resolución de un sistema de imágenes.

La frecuencia espacial es una relación del cambio de la densidad óptica con relación a la distancia (es similar al concepto de pares de líneas por *mm*). Las porciones de una imagen en las que las densidades ópticas varían de valores altos a bajos en una distancia corta se dice que contienen información de frecuencia espacial alta.

La función de transferencia de modulación describe la capacidad de un sistema de formación de imágenes para reproducir información de frecuencia espacial baja, media y alta.

En las imágenes médicas, ambas frecuencias espaciales, baja y alta, son clínicamente importantes. Las componentes de la imagen con frecuencia espacial baja proporcionan información útil para el diagnóstico cuando se examina una imagen del abdomen, donde los cambios graduales en las densidades ópticas en algunas regiones anatómicas pueden indicar la presencia de patología. La información de frecuencia espacial alta en una imagen de extremidades o de una mama indica los detalles finos y las estructuras dentro de estos órganos.

Es importante distinguir entre la curva de la función de dispersión lineal y la curva de la función de transferencia de modulación. El eje *X* para la función de dispersión lineal está en unidades de longitud, mientras que para la función de transferencia de modulación está en unidades de frecuencia espacial (normalmente en ciclos por *mm*). Al comparar sistemas de imagen, la función de dispersión lineal de un sistema con borrosidad mayor será más ancha que la de un sistema nítido. La función de transferencia de modulación de un sistema mejor que otro (más nítido) estará por encima de la función de transferencia de un sistema de mayor borrosidad.

Las funciones de transferencia de modulación pueden determinarse por separado para cada uno de los factores que influyen en el sistema de imagen. Como cada valor en una curva de función de transferencia de modulación (MTF) está entre 0 y 100%, el MTF del sistema puede determinarse multiplicando las MTF de cada uno de los componentes. La MTF de todo sistema siempre estará por debajo de las curvas de cada componente aislado.

Fijando nuevos parámetros de operación. Un médico o un técnico que obtenga radiografías de un paciente ha de tener presente los anteriores conceptos para ser capaz de establecer de manera inmediata el cambio que ha de experimentar el valor del *mAs* a fijar en la consola de un equipo de rayos X para obtener una radiografía si la distancia foco-película cambia. Los conceptos anteriores han de ligarse con el hecho de que la energía emitida desde el foco emisor de radiación se extiende por la superficie delimitada por el ángulo sólido formado por dicho punto y el lugar de interacción de la radiación. Dicha superficie es, evidentemente, proporcional al cuadrado de la distancia a la fuente de radiación. Como la misma cantidad de energía ha de repartirse en superficies progresivamente crecientes con la distancia, se comprenderá que es de aplicación la denominada ley de la inversa del cuadrado de la distancia, que, en términos matemáticos, se expresa:

$$D_1 / D_2 = (d_2 / d_1)^2,$$

siendo D_i los valores de dosis a las distancias d_i . Entonces, para mantener un valor de dosis necesario para obtener una imagen similar al variar la distancia, se tendrá:

$$(\text{Nueva distancia})^2 / (\text{Antigua distancia})^2 = \text{Nuevo valor de } mAs / \text{Antiguo valor de } mAs.$$

De acuerdo con lo anteriormente expresado, para mantener constante una determinada cantidad de radiación ha de mantenerse constante el producto de los valores de miliamperios y tiempo, es decir, la carga eléctrica configurada por el haz de electrones generados en el tubo de rayos X.

Sensitometría. Los conceptos anteriores también han de manejarse con soltura por el médico o el técnico a la hora de buscar una imagen radiográfica con mayor contraste, que permita diferenciar mejor las zonas anatómicas observadas en función de las diferencias de densidad óptica. Para ello suele usarse el concepto gradiente, que es similar al concepto gamma anterior, si bien suelen fijarse los valores 2 y 0.25 como los correspondientes a los valores indicados con los subíndices 2 y 1 anteriores.

$$\text{Gradiente medio} = (\text{Densidad } 2.0 - \text{Densidad } 0.25) / (\log X_{D_2} - \log X_{D_1})$$

Cuanto mayor sea el valor del gradiente para una determinada película o combinación película-pantalla, mayor será el contraste.

Rejillas o parrillas. Cuando el haz de rayos X pasa a través de un paciente, el haz es absorbido y dispersado. Mientras que el haz atenuado produce una “sombra” útil, el haz dispersado por el paciente tiende a enmascarar los efectos beneficiosos. La radiación dispersada se elimina colocando una rejilla entre la película y el paciente.

La rejilla consiste en una serie de tiras de plomo de grosor c y de altura h separadas por espaciadores de material de baja densidad de anchura b . En una rejilla típica, las tiras pueden ser de 0.05mm de grosor, de altura 2.5mm y separadas por espaciadores de 0.35mm de ancho.

Se define la razón de rejilla como h/b .

Las rejillas pueden ser paralelas o focalizadas. Pueden ser estacionarias o en movimiento.

En una rejilla paralela el haz primario se cortará por las tiras de plomo a una distancia w , que está relacionada con la distancia foco-película, d : $w/d = b/h$, es decir, $w = d/\text{razón de rejilla}$. Usando esta sencilla ecuación, puede calcularse el ancho máximo de campo que puede radiografiarse con una determinada rejilla.

El uso de una rejilla siempre incrementa la exposición debido a que absorberá algo de la radiación primaria. Por ejemplo, al obtener una imagen de un maniquí de 20cm de agua a 100 kV , una rejilla 8:1 reduce el haz primario por un factor de 1.8, y el primario más el disperso por 3.7. Para obtener la misma densidad en la película con rejilla que sin ella la exposición tendrá que incrementarse por un factor de 3.7. A esto se le llama el factor de rejilla.

Como la exposición es directamente proporcional al producto mAs , cada vez que se use una rejilla se tendrá que realizar el siguiente cálculo:

- 1) Nuevo mAs con rejilla = Viejo mAs sin rejilla \times factor de rejilla
- 2) mAs con nueva rejilla = (Nuevo factor de rejilla \times mAs con vieja rejilla / factor de vieja rejilla)

A fin de reducir exposiciones, las rejillas con menores razones están siendo usadas más frecuentemente.

Digitalización. Actualmente, el Radiodiagnóstico está siendo cada vez más el reino de las imágenes digitales. Como es bien sabido, la palabra “digital” se refiere a “cálculo por

métodos numéricos o por unidades discretas”. Una imagen digital es una representación numérica de un objeto (que también puede ser una imagen). Consiste en una disposición bidimensional de elementos de la imagen (pixel) representando una subdivisión en las direcciones X e Y del área de la imagen.

El proceso de digitalización de una imagen siempre consta de tres pasos:

- Muestreo
- Cuantificación
- Codificación

El muestreo consiste en obtener un número finito de muestras del brillo según posiciones predefinidas en la imagen.

En el proceso de cuantificación y codificación, se convierte una variable continua en una discreta, de forma que toma valores en un conjunto finito de números según una función de asignación. Las reglas de cuantificación definen los niveles de transición entre intervalos. El número de intervalos está directamente relacionado con el método que se utilice para la codificación binaria de los niveles de brillo. Si se dispone de n bits para la codificación del brillo tendremos 2^n niveles distintos para cuantificar la señal, o sea, el brillo.

La resolución del brillo es la capacidad para distinguir entre dos niveles de grises consecutivos. En radiografía digital se utilizan habitualmente 256 niveles de grises, o sea, para cada pixel se necesitan 8 bits (1 byte) para codificar el brillo. El proceso de cuantificación/codificación influye en el contraste.

Así, por ejemplo, si decimos que una imagen se ha digitalizado con una rejilla cuadrada de 4×4 elementos (2^m , $m=2$) y se ha cuantificado con un código Gray de (3,3) (que define el número de niveles de grises), estamos indicando que:

- a) Estamos manejando una escala de grises de $2^3 = 8$ niveles (siendo el nivel 0 para el blanco y 7 para el negro)
- b) Cada nivel de brillo se puede codificar mediante $N = 3$ bits.

En general, si suponemos que el número de pixels es 2^m , y que el código Gray es de (n,n) , el número de bits necesarios para almacenar cada imagen viene dado por $2^{(2m+n)}$.

Una vez las imágenes en forma digital se aplica normalmente una transformada lineal o logarítmica que comprime el rango de la imagen, haciéndola más adecuada para su proyección en una pantalla de un monitor.

Procesado de imágenes. El procesado digital de imágenes consiste en la utilización de algoritmos para la modificación de los valores de los pixeles de una imagen digital. Se persigue con ello: a) mejorar la imagen desde el punto de vista subjetivo, o b) interpretar automáticamente las imágenes con un ordenador. El procesamiento puede ser una operación puntual, local o global.

Entre las técnicas más comunes en el procesado digital de imágenes que se utilizan en la práctica médica tenemos:

1) *Modificación de contraste.* Es el nivel más básico del procesamiento de imágenes. El contraste de una imagen puede apreciarse en un histograma en que se representen los niveles de gris frente a su frecuencia de aparición. La expansión del histograma puede realizarse de forma

lineal o no lineal. Entre las funciones no lineales se encuentran la exponencial y la ecualización de histograma.

Si se denomina $I(x,y)$ el valor digital en la localización (x,y) , la modificación de contraste puede representarse por $I_{out}(x,y) = c I_{in}(x,y) + d$.

2) *Operaciones aritméticas y geométricas.* Las operaciones aritméticas con imágenes se utilizan con mucha frecuencia. En Radiología es particularmente frecuente la sustracción de imágenes, sobre todo en angiografía, empleando imágenes obtenidas antes y después de la inyección de contraste intravascular, de forma que se visualizan los vasos sanguíneos sin la superposición de estructuras óseas. La forma más simple de angiografía digital de sustracción consiste en sustraer la imagen previa a la llegada de contraste a la imagen correspondiente al pico máximo de contraste. Para ello se restan a la segunda imagen, pixel a pixel, los datos correspondientes a la imagen antes del contraste (máscara):

$$I_{out}(x,y) = I_1(x,y) - I_2(x,y) + c,$$

siendo I_1 e I_2 las imágenes amplificadas logarítmicamente obtenidas antes y después de la inyección de contraste, y c un valor que permite que tanto valores positivos como negativos se presenten siempre como enteros positivos.

La amplificación logarítmica antes de la sustracción permite linealizar las señales de atenuación debidas al contraste.

3) *Promedio temporal.* El promedio de una secuencia de imágenes implica la suma de imágenes pixel a pixel. Mediante esta operación, el nivel de ruido en la imagen obtenida se reduce en un factor igual a la raíz cuadrada de N (N imágenes sumadas en cada posición).

4) *Filtrado espacial.* En esta operación se utilizan los valores de varios píxeles de la imagen de entrada para determinar un único valor de pixel en la imagen de salida. La función lineal usada más habitualmente es la convolución o combinación lineal de los niveles de gris de varios píxeles. La operación de convolución consiste en una serie de multiplicaciones punto a punto y sumas de los productos para derivar una imagen a partir de una imagen de entrada.

La transformada de Fourier convierte una distribución en el dominio espacial en una descripción equivalente en el dominio de frecuencias. Es frecuente realizar este tipo de transformación por las ventajas que significa trabajar con funciones en el dominio de frecuencias. La transformada de la convolución de dos funciones es el producto de las transformadas de Fourier de las dos funciones.

5) *Wavelets.* Se están utilizando estas transformadas en el ámbito de las imágenes médicas. Sus aplicaciones pueden ser muy diversas, incluyendo el realce del contraste, la detección de microcalcificaciones en mamografías, la compresión de imágenes para su transmisión eficaz en redes, etc.

Se trata de funciones matemáticas “con forma de onda” que permiten la descomposición de una imagen en distintas aproximaciones con diferente escala o resolución. Así son capaces de descomponer una imagen en componentes de diferente frecuencia, estudiando cada componente con una resolución adaptada a su escala. En cada resolución puede aplicarse una técnica de procesamiento de imagen diferente, adaptándola a la resolución. Por último, las imágenes procesadas de distinta escala pueden reconstruir de nuevo la imagen médica buscada,

empleando la transformada “wavelet” inversa. De este modo se pueden aplicar con facilidad diferentes algoritmos adaptados al tamaño de los objetos que se encuentran en la imagen.

Mamografía digital. Es un método efectivo para la detección y diagnóstico de cáncer de mama. Las diferencias entre los coeficientes de atenuación del tejido mamario fibroglandular, de la grasa y del carcinoma mamario son muy pequeñas, por lo que es difícil llegar a un diagnóstico acertado de tumores rodeados de tejido fibroglandular. El contraste del objeto disminuye rápidamente con el incremento de la energía, por lo que se usan rayos X de muy baja energía (26 a 30 keV).

Las propiedades del contraste se describen mediante la pendiente o gradiente de la curva característica. Debido a la forma típica, el rango de exposiciones para los que el gradiente de la proyección es significativo, o sea, la latitud de la imagen, está limitada a un factor de, aproximadamente, 25. Como el rango de exposiciones transmitidas puede ser de 100 a 1, la pérdida de información puede ser particularmente importante para pacientes con mamas que contengan gran proporción de tejido fibroglandular.

La mamografía digital da lugar a una imagen “muestreada”. El muestreo se hace tanto en el dominio espacial como en el de intensidades. La intensidad de la imagen se digitaliza en términos de un número finito de niveles de señal que abarcan desde 0 a $2^n - 1$, siendo n el número de bits usados para digitalizar las variaciones de la señal detectada. En mamografía digital el número de bits está entre 12 y 16, lo que da lugar a un número comprendido entre 4096 y 64000 valores de intensidad. La calidad de la imagen digital depende críticamente de estos intervalos de muestreo.

Actualmente, los detectores usados en mamografía digital tienen una resolución espacial inferior a la clásica de película pantalla. En efecto, se necesitaría para ello unos 70×10^6 pixels. Sin embargo, con menos pixels el sistema digital permite percibir mejor estructuras de moderado y de bajo contraste.

Hoy ya existe respuesta técnica para la resolución espacial realmente necesaria en mamografía digital. Con los actuales monitores ya se alcanza una adecuada resolución espacial y se consigue, mediante el sistema digital, el empleo de sistemas de archivo de imágenes y comunicaciones (PACS).

5. Protección Radiológica: Consideraciones matemáticas. Cálculo de blindajes.

Toda ella descansa en el denominado Sistema de Limitación de Dosis o Sistema de Protección Radiológica, que establece que todo uso de la radiación sobre personas debe hacerse de forma que se cumplan premisas tales como Justificación y Optimización.

En el caso de que la radiación pudiera afectar a personal trabajador profesionalmente expuesto o al público en general ha de cumplirse, además, el principio de no superación de unos valores límite de dosis individuales.

Tanto el principio de Justificación como el de Optimización alcanzan su exacta formulación cuando se expresan en el lenguaje matemático que resulta de la aplicación de un análisis costes-beneficios al uso de la radiación sobre los humanos. Aunque este análisis es frecuente en las ciencias económicas, parece a primera vista extraño que sea la verdadera base sobre la que se apoya el actual reconocimiento de la pertinencia del uso de las radiaciones en la práctica médica.

Hagamos unos comentarios sobre ambos principios.

Justificación. En el análisis costo-beneficios, los beneficios a tener en cuenta han de comprender todos los beneficios recibidos por la sociedad, tanto si son cuantificables como si no. Asimismo, se considera que los costes comprenden la suma total de todos los aspectos negativos de una operación expresados en costes monetarios y todos los daños para la salud del hombre o daños al medio ambiente.

La justificación exige que el beneficio neto que se obtiene a partir de la fórmula que se indica a continuación, sea positivo:

$$B = V - (P + X + Y),$$

siendo B el beneficio neto, V el beneficio bruto o valor de la práctica con radiaciones, P el costo base de producción o coste de introducción de la práctica, X el costo de un nivel determinado de protección e Y el costo del detrimento ligado a esa operación como consecuencia de la exposición que, como todo costo, podrá expresarse por el producto de valor de la unidad de esa magnitud de costo multiplicado por el valor absoluto de dicha magnitud.

La unidad de costo del detrimento suele expresarse como α_j (valoración de la unidad persona-Sievert para justificación).

En general, la justificación suele determinarse examinando la diferencia entre una nueva práctica (n) y una condición de referencia (r) en forma de la ecuación:

$$(B_n - B_r) = (V_n - V_r) - (P_n - P_r) - (X_n - X_r) - \alpha_j (Y_n - Y_r)$$

En la mayoría de los casos los valores de ΔV , ΔP y α_j son, precisamente, los parámetros definidos con menor precisión. Esto tiene como consecuencia que las decisiones tales como la justificación de una práctica no pueden ser más precisas que los parámetros de los que depende.

Sin embargo, es posible usar la ecuación de justificación para realizar un análisis de condiciones límite para evaluar estos parámetros. Muchas situaciones ocasionan grandes variaciones en estas condiciones, lo cual hace obvia la elección, o bien el análisis ayuda a identificar los parámetros críticos. El análisis de límites tiene lugar cuando $\Delta B = 0$, lo que ocasiona que:

$$(\Delta V - \Delta P) = \Delta X + \alpha_j \Delta Y$$

En el límite, la justificación puede usarse para investigar los posibles rangos de valores de α , lo cual puede ser apropiado en aquellos casos en que no se conocida la propia α . En esta situación, se considera la siguiente ecuación en el límite para los rangos variables de incertidumbre de los parámetros:

$$((\Delta V - \Delta P) - \Delta X) / \Delta Y = \alpha_j$$

La justificación necesita asignar un valor monetario para α como unidad de dosis colectiva. Se han publicado los valores usados en distintos países. Está claro que α puede tener distintos valores como resultado de diversos factores. Esto es particularmente evidente cuando se considera la modificación de exposición debida a fuentes de radiación natural.

Uno de los factores clave en el uso correcto de la radiación es el factor detrimento. El detrimento en un grupo de población irradiada se define como la esperanza matemática del daño que la radiación puede ocasionar, teniendo en cuenta no sólo las probabilidades de cada tipo de efecto dañino, sino también la gravedad de los efectos. Así, si p_i es el riesgo de experimentar el

efecto i , cuya gravedad se mide con un factor g_i , entonces el detrimento Y para un grupo de N personas es $Y_i = N \sum p_i g_i$.

Optimización. Es un elemento esencial de la radioprotección aplicada. Para la optimización, el análisis costo-beneficios es diferencial y corresponde a un análisis costo-eficacia. Para aplicarla ha de tenerse en cuenta que la función beneficio depende de la variable dosis colectiva, S , resultante de la operación. La condición de máximo se expresa como $dB / dS = 0$, es decir:

$$(dV/dS) - ((dP/dS) + (dX/dS) + (dY/dS)) = 0.$$

Como se puede considerar que tanto V como P son independientes de S para una práctica radiológica dada, la condición de optimización queda reducida a:

$$(dX/dS) = - (dY/dS).$$

Es conveniente señalar que ya se ha indicado que el detrimento da lugar a un término económico cuando se multiplica su valor por el de la unidad de dosis efectiva colectiva (persona-Sievert). En rigor, por tanto, y desde el punto de vista del costo del detrimento, el mismo puede calcularse multiplicando el costo monetario asignado a la unidad de dosis efectiva colectiva por la cuantía del compromiso de dosis efectiva colectiva asociado con el nivel de protección considerado: $Y = \alpha S$. Pues bien, el compromiso de dosis efectiva colectiva debido a un suceso determinado, decisión o práctica, se define como la integral extendida a un intervalo de tiempo entre 0 e infinito de la dosis efectiva colectiva debida a ese evento, decisión o práctica.

A su vez, la dosis efectiva colectiva en una población se define por la expresión

$$S = N(E) dE,$$

siendo $N(E)$ el espectro de la población en relación con la dosis efectiva y $N(E) dE$ el número de individuos que reciben una dosis efectiva en el rango E a $E + dE$. Por consiguiente, la dosis efectiva colectiva es el producto ponderado de la dosis efectiva y el número de individuos en la población expuesta.

Teniendo en cuenta las expresiones anteriores, se concluye que $(dX / dS)_{S_0} = -\alpha$, siendo S_0 el valor correspondiente a un parámetro que da protección óptima y un costo de protección determinado.

Estos aspectos matemáticos no tienen una utilización inmediata cuando se están usando radiaciones ionizantes. ¡No es comprensible solicitar a un facultativo médico que haga un análisis costo-beneficios cada vez que vaya a prescribir una exploración radiológica! Sin embargo, los especialistas en Protección Radiológica están obligados a conocerlos para poder convencer al personal médico y sanitario de la necesidad de la observación de los principios de Protección Radiológica, ya que, en caso contrario, el recurso sería la pura imposición. Además, ante una situación como la presente en que se han promulgado recientemente unos cambios en los límites de dosis tanto del personal expuesto como del público en general, la necesidad de conocer las bases matemáticas que subyacen en las premisas del sistema de Protección Radiológica se convierte en algo esencial.

Cálculo de blindajes. El uso de radiaciones ionizantes en la práctica médica está determinado por el cumplimiento del sistema de limitación de dosis que, después de las premisas de justificación y optimización de cualquier uso de radiación que se haga, establece la obligatoriedad, para todo centro sanitario en el que se use radiación, de poner en marcha medidas que eviten que, en condiciones normales de trabajo, se superen los límites de dosis establecidos para el personal trabajador profesionalmente expuesto a radiaciones.

Entre esas medidas se encuentra la colocación de blindajes o barreras de material interpuestas entre el haz de radiación y el trabajador, a fin de que que atenúen la intensidad de la radiación que alcance al trabajador de forma que no ocasionen la superación de aquellos valores límite, en condiciones normales de actividad laboral. Estas medidas de protección también tienen que ser capaces de evitar que el público en general que visite un centro hospitalario supere los límites de dosis establecidos para ellos por la legislación vigente ($1mSv/año$). Conseguir estos dos objetivos es un imperativo legal que habitualmente pasa desapercibido tanto para el personal sanitario no expuesto profesionalmente a radiaciones como al público en general y que, sin embargo, requiere un cuidadoso análisis que implica el uso de capacidades matemáticas evidentes, como puede comprobarse con la presentación siguiente.

En condiciones de buena geometría (haz de radiación monoenergético, estrecho y bien colimado), la atenuación de un haz de radiación gamma viene dada por $I = I_0 e^{-\mu x}$, siendo I la intensidad de la radiación después de atravesar un espesor x de material, I_0 el valor de la intensidad de la radiación incidente en el material y μ el denominado coeficiente de atenuación lineal de la radiación, que es un valor numérico (con dimensiones de inversa de longitud) que depende tanto del tipo de material interpuesto en el camino del haz de radiación como de las características de la radiación.

Un valor de grosor de material muy usado es el de capa hemirreductora, es decir, aquel grosor de material que reduce la intensidad de la radiación incidente a la mitad. Es evidente que

$$CHR = 0.693 / \mu$$

Análogamente se define el concepto de capa décimo-reductora, CDR, la cual es

$$CDR = \ln 10 / \mu$$

Blindaje para rayos X. El blindaje estructural contra rayos X se diseña para proteger frente a los rayos X útiles, la radiación de fuga y la radiación dispersa. Siempre se diseña para proteger a la gente en un área ocupada que está fuera del área o dependencia en que se usa el equipo emisor de radiación. Los requisitos para una instalación dada están determinados por:

- 1) El kilovoltaje máximo al que pueda operar el tubo de rayos X.
- 2) El miliamperaje máximo de la corriente del haz.
- 3) La carga de trabajo (W), que es una medida de uso que se da a la máquina de rayos X. Es frecuente expresar la carga de trabajo en unidades de miliamperios-minuto por semana.
- 4) El factor de uso de una dirección, U , que es la fracción de la carga de trabajo durante la que el haz útil de radiación tiene la dirección que se esté considerando.
- 5) El factor de ocupación de un área tras el blindaje, T , que es el factor por el que se debe multiplicar la carga de trabajo para corregir la misma teniendo en cuenta el grado o tipo de ocupación del área en cuestión.

Cálculo de blindajes para equipos de megavoltaje. Con objeto de reducir la tasa de dosis a niveles que no ocasionen la superación de los límites de dosis establecidos para el personal que usa estos equipos, es usualmente necesario colocar barreras de plomo, de hormigón o de otros materiales entre las personas y las fuentes de radiación.

Analicemos, como ejemplo de una posible configuración de un acelerador lineal de electrones de uso médico, el grosor de paredes necesario para conseguir los niveles de protección requeridos. Se necesitará para ello conocer la producción de la máquina, el tipo de radiación emitido, la distancia desde la fuente hasta el punto a proteger, el tiempo que el haz de radiación apunta en esa dirección y las características de las personas a proteger.

Los tipos de radiación a considerar son:

- Radiación primaria
- Radiación de fuga
- Radiación dispersa

Para diseñar las barreras hay que disponer de los siguientes datos:

- 1) Carga de trabajo: Aunque un acelerador puede proporcionar tasas de dosis de hasta $5\text{Gy}/\text{min}$ a 1m , el tiempo necesario para la colocación de los pacientes hace que el número máximo de personas tratadas sea de unas 30 en sesiones de 7 horas de trabajo. Como la dosis media por pacientes es de unos 2.5Gy , la carga de trabajo a 1m es:

$$W = (2.5 \text{ Gy}/\text{pac}) \times (30 \text{ pac}/\text{día}) \times (5 \text{ días}/\text{semana}) \times (50 \text{ semanas}/\text{año})$$

- 2) Ley de la inversa del cuadrado de la distancia.
- 3) Factor de uso.
- 4) Factor de ocupación.

En los estudios de protección radiológica conviene especificar la barrera en términos de número de capas decimorreductoras.

Dosis a pacientes en Radiodiagnóstico. El incremento progresivo en la utilización de los rayos X en Medicina significa que sean responsables de cerca del 15% de la dosis total que recibe de promedio una persona en un país de desarrollo sanitario similar al nuestro. En otras palabras, el Radiodiagnóstico es la principal causa de irradiación debida a agentes no naturales a que está expuesta la especie humana. Por ello y a fin de evitar tanto como sea posible los efectos perjudiciales de la radiación, se ha desarrollado una normativa aplicable a todo uso de radiación que esencialmente trata de conseguir que las dosis de radiación recibidas sean tan bajas como sea razonablemente posible. Es imperativo, pues, medir las dosis de radiación que se imparten en los centros sanitarios a los pacientes con objeto de determinar si exceden unos valores medios de referencia que se han establecido para exploraciones simples o bien si exceden los valores de referencia de cada centro para las exploraciones complejas.

El asunto es de tal importancia que se ha tratado de establecer un modelo matemático basado en el método Monte Carlo para poder conocer las dosis de radiación recibidas por los pacientes. Este modelo ha dado lugar a un programa informático que, al menos para exploraciones simples, permite fijar el valor de dosis recibida por el paciente tanto en órganos individuales como la dosis efectiva correspondiente a la globalidad del individuo. Uno de estos programas es el que se denomina *Effdose*, el cual es utilizado habitualmente por personal

técnico y por personal Especialista en Radiofísica Hospitalaria que, obviamente, han de estar familiarizados con los fundamentos matemáticos que han permitido su construcción.

6. A modo de conclusión

Este recorrido, casi de puntillas, por el uso médico de las radiaciones ionizantes ha permitido observar que todo el personal implicado en su utilización: médicos especialistas, radiofísicos hospitalarios, técnicos, etc. están haciendo uso, en uno u otro grado, durante todos los momentos de su actividad profesional, de conceptos, lenguaje y herramientas matemáticas. Parodiando a Platón puede afirmarse: “Nadie puede entrar en el ámbito del uso médico de la radiación si no sabe matemáticas”. Sin dicho conocimiento los perjuicios para la salud de las personas pueden ser tremendos, habida cuenta del carácter lesivo de las radiaciones ionizantes.

En Internet

<http://webpages.ull.es/users/jjgranad/welcome.html>

Laboratorio de Física Médica y Radiactividad Ambiental
Universidad de La Laguna

Resolución de Problemas en Paralelo



Coromoto León Hernández

Profesora Titular de Lenguajes y Sistemas Informáticos
Departamento de Estadística, Investigación Operativa y Computación
Universidad de La Laguna

<http://nereida.deioc.ull.es/~cleon>

Resumen

En este trabajo se abordará la resolución de algunos problemas de optimización utilizando *esqueletos* paralelos basados en las técnicas Divide y Vencerás y Ramificación y Acotación. Las implementaciones están hechas utilizando el lenguaje de programación C++, y el usuario sólo tiene que especificar el tipo de problema, el tipo de solución y las características específicas de la técnica algorítmica que quiere utilizar. Esta información se combina con los esqueletos de resolución que se proporcionan para obtener programas secuenciales y paralelos tanto para arquitecturas de Memoria Compartida, como para las de Memoria Distribuida.

Palabras Clave

Técnicas Algorítmicas, Ramificación y Acotación, Divide y Vencerás, Esqueletos, Paso de Mensajes, Memoria Compartida.

1. Introducción

El ser humano siente una gran curiosidad por comprender el funcionamiento de las cosas que le rodean, lo que ha traído como consecuencia el desarrollo del método científico. El proceso empieza con una observación cuidadosa del fenómeno a estudiar y la formulación del problema. A continuación se aborda la construcción del modelo científico, en general matemático, que intenta abstraer la esencia del problema real. Llegado este punto, se propone la hipótesis de que el modelo es una representación lo suficientemente precisa de las características esenciales de la situación como para que las soluciones obtenidas sean válidas también para el problema real. Esta hipótesis se verifica y modifica mediante las pruebas adecuadas. Sin embargo, el objetivo no es sólo encontrar una solución al problema en cuestión, la meta es encontrar la mejor solución, la *solución óptima*. El estudio de este tipo de problemas es parte del campo de acción de la rama de las Matemáticas Aplicadas denominada *Investigación Operativa*. Concretamente, la *Programación Matemática* trata de resolver problemas de decisión en los que se deben determinar acciones que optimicen un determinado objetivo, pero satisfaciendo ciertas limitaciones en los recursos disponibles. La *Optimización Combinatoria* es una rama de la Programación Matemática que intenta la resolución de problemas de optimización caracterizados por tener un número finito de posibles soluciones ... pero dicho número es muy grande. Y a este tipo de problemas es al que hace referencia el título de este trabajo.

Fijado el tipo de problema, se trata de proporcionar el conjunto de operaciones y procedimientos que deben seguirse para resolverlo. Al conjunto de todas las operaciones a realizar y el orden en el que deben efectuarse se le denomina *Algoritmo*. El diseño de un algoritmo que resuelva un problema es, en general, una tarea difícil. Sin embargo, no es necesario partir de cero; se puede facilitar esta labor recurriendo a técnicas conocidas de diseño de algoritmos, es decir, a esquemas muy generales que pueden adaptarse a un problema particular (*Técnicas Algorítmicas*). Muchos problemas pueden resolverse buscando una solución fácil y directa. Este método, denominado de *fuerza bruta*, puede ser muy directo, pero con un análisis superficial se concluye que los algoritmos no son eficientes. El esquema algorítmico más sencillo es el llamado *Divide y Vencerás*, basado en la descomposición de un problema en subproblemas. Hay problemas cuya solución sólo puede hallarse mediante un proceso de búsqueda. A pesar de lo complejas que son las operaciones de búsqueda, su uso adecuado mediante el esquema de *Búsqueda con Retroceso* (o backtracking) permite ganar gran eficiencia respecto a soluciones de fuerza bruta. La técnica de *Ramificación y Acotación* constituye una variante del método de retroceso para problemas donde se trata de encontrar el valor máximo o mínimo de cierta función objetivo.

El papel que juegan los *ordenadores* en el proceso de diseño e implementación de los algoritmos para resolver problemas de optimización es fundamental. Sin embargo, aún existen problemas tan grandes que no son resolubles en un tiempo razonable. Una forma de proceder consiste en utilizar *paralelismo*. Existen diferentes tipos de máquinas paralelas comerciales. Las que pueden ejecutar múltiples instrucciones sobre múltiples datos (MIMD) se dividen en dos subclases, dependiendo de la relación entre la memoria y los procesadores. Las máquinas paralelas de *Memoria Compartida* permiten a cualquier procesador acceder a cualquier módulo de memoria global, mientras que las de *Memoria Distribuida* conectan procesadores individuales con sus propios módulos de memoria e implementan los accesos a memoria remota mediante el *paso de mensajes* entre procesadores. Para cada una de estas arquitecturas se ha desarrollado software paralelo, pero de forma dependiente de la misma.

La solución exacta de la mayoría de los problemas combinatorios implica la enumeración de los elementos de un espacio de soluciones exponencial. Fijado un problema, la búsqueda exhaustiva de una solución se puede acelerar a través de técnicas exactas como *Divide y Vencerás*, *Ramificación y Acotación* o *Programación Dinámica*. El esquema de trabajo de estas técnicas se puede generalizar para desarrollar herramientas software que permitan el análisis, diseño e implementación de resolutores para problemas concretos. Desafortunadamente, en muchas aplicaciones aparecen casos de tamaño irresoluble, esto es, no es posible encontrar la solución óptima en tiempo razonable. Sin embargo, es posible incrementar el tamaño de los problemas que se pueden abordar por medio de técnicas de paralelización.

En este trabajo se presentan los esqueletos orientados a objeto DnC y BnB para la resolución de problemas de optimización combinatoria con las técnicas *Divide y Vencerás* y *Ramificación y Acotación* respectivamente. La implementación de los esqueletos se ha realizado en C++. Se proporciona el código secuencial y el código paralelo de la parte invariante de los resolutores para ambos paradigmas. Además, se pone a disposición del usuario la plantilla de la parte que ha de rellenar. Las clases que componen dicha plantilla sirven para establecer la relación entre el resolutor principal y el problema instanciado. Una vez que el usuario ha instanciado su problema, obtiene gratis un resolutor paralelo de su algoritmo, sin realizar ningún esfuerzo adicional.

Los esqueletos proporcionan modularidad al diseño de algoritmos exactos, lo que supone una gran ventaja respecto a la implementación directa del algoritmo desde el principio, no sólo en términos de reutilización de código sino también en metodología y claridad de conceptos.

Se han encontrado en la revisión bibliográfica varias herramientas para la implementación paralela de algoritmos generales de Ramificación y Acotación. PPBB (Portable Parallel Branch-and-Bound Library) [14] y PUBB (Parallelization Utility for Branch and Bound algorithms) [13] proponen implementaciones en el lenguaje de programación C. BOB [10] y PICO (An Object-Oriented Framework for Parallel Branch-and-Bound) [4] están desarrollados en C++ y en ambos casos se proporciona una jerarquía de clases que el usuario ha de extender para resolver su problema particular. En cuanto a los sistemas de programación paralelos orientados a la metodología Divide y Vencerás podemos citar los siguientes: Cilk [7] basado en el lenguaje C y Satin [8] basado en Java. Sin embargo, no hemos encontrado ninguna herramienta que permita trabajar de forma integrada con más de una técnica algorítmica, y esta es una de las principales aportaciones de nuestra propuesta. Los esqueletos DnC y BnB forman parte del proyecto MaLLBa [1,2,3,9].

El resto del artículo se organiza como sigue. En la segunda sección se presenta la arquitectura MaLLBa y las interfaces particulares de DnC y BnB. A continuación, en el tercer epígrafe, se describen los esqueletos secuenciales y paralelos. En la cuarta sección se utilizan DnC y BnB de forma integrada en la resolución del Problema de la Mochila 0/1. Finalmente, se presentan las conclusiones y los trabajos futuros.

2. Descripción de la arquitectura MaLLBa

Se denomina *esqueleto algorítmico* a un conjunto de procedimientos que constituyen el armazón con el que desarrollar programas para resolver un problema dado utilizando una técnica particular. Este software presenta declaraciones de clases vacías que el usuario ha de rellenar para adaptarlo a la resolución de un problema concreto.

La librería MaLLBa se ha diseñado de forma que la tarea de la persona que adapta el problema real al esqueleto sea lo más simple posible. En un esqueleto MaLLBa se distinguen dos partes principales: una parte que implementa el *patrón de resolución* y que es completamente proporcionada por la librería, y una parte que el usuario ha de acabar de completar con las características particulares del *problema a resolver* y que será utilizada por el patrón de resolución.

En el diseño de los esqueletos se ha utilizado la Orientación a Objetos (OO). La facilidad de interpretación del esqueleto es una de las ventajas que aporta esta filosofía, además de las de modularidad, reusabilidad y modificabilidad. Existe una relación bastante intuitiva entre las entidades participantes en el patrón de resolución y las clases que ha de implementar el usuario.

La parte proporcionada por el esqueleto, esto es, los patrones de resolución, se implementan mediante clases en las que explícitamente se indica que son definidas por el esqueleto. A estas clases se les da el nombre de clases *proporcionadas* y aparecen en el código con el calificativo *provides*.

La parte que ha de completar el usuario con la instanciación de su problema particular se implementa mediante clases marcadas con el calificativo *requires*, a las que se denominarán clases *requeridas*. La adaptación que ha de realizar el usuario consiste en representar mediante las estructuras de datos necesarias su problema e implementar (en función de dicha

representación) las funcionalidades requeridas por los métodos de las clases. Estas clases serán invocadas desde el patrón de resolución particular (porque conoce la interfaz para dichas clases) de forma que, cuando la instanciación se ha completado, se obtienen las funcionalidades esperadas aplicadas al problema concreto del usuario.

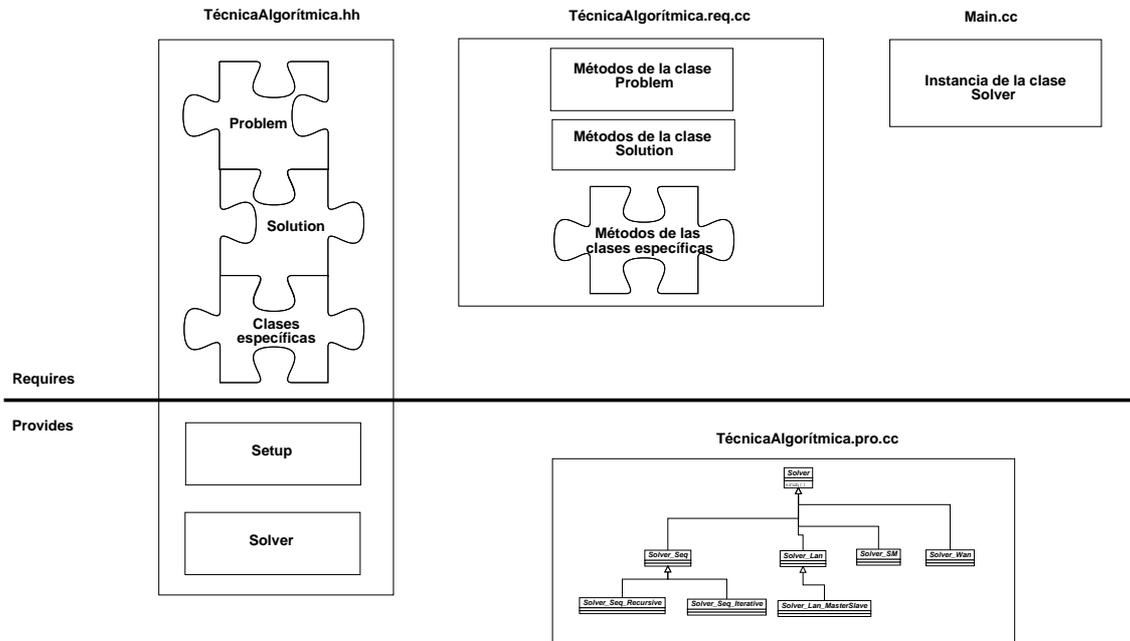


Figura 1. Arquitectura de los Esqueletos MaLLBa.

La Figura 1 presenta la arquitectura del esqueleto. En el fichero con extensión `.hh` se definen las cabeceras de las clases requeridas y proporcionadas. El fichero `.pro.cc` contiene las implementaciones C++ de los patrones de resolución, mientras que el fichero `.req.cc` es el que ha de implementar el usuario.

2.1. Clases requeridas. La solución de un problema de optimización combinatoria en general consiste en un vector de enteros que cumple un número de restricciones y optimiza una función objetivo. La función objetivo debe ser maximizada o minimizada. Partiendo de esta descripción, se abstrae que en un esqueleto deben representarse mediante clases tanto el *problema* como la *solución*.

Las clases requeridas se utilizan para almacenar los datos básicos del algoritmo. Con ellas se representan el problema, la solución, los estados del espacio de búsqueda y la entrada/salida. Todos los esqueletos MaLLBa han de definir las siguientes clases:

- **Problem:** define la interfaz mínima estándar necesaria para representar un problema.
- **Solution:** define la interfaz mínima estándar necesaria para representar una solución.

Cada patrón de resolución requiere además un conjunto de clases propias de la técnica algorítmica. En los epígrafes siguientes se describen las clases que el usuario ha de implementar cuando utilice DnC y BnB.

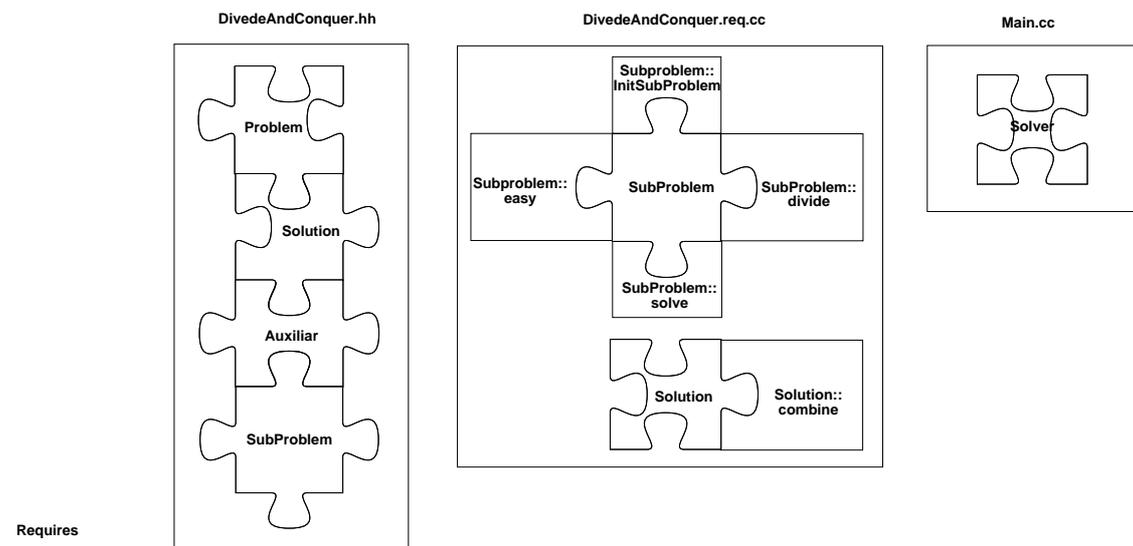


Figura 2. Clases requeridas por el esqueleto de Divide y Vencerás.

2.1.1. Interfaz del esqueleto de Divide y Vencerás. Para instanciar un esqueleto DnC además de las clases `Problem` y `Solution`, es necesario implementar las clases `SubProblem` y `Auxiliar` (véase la Figura 2).

- La clase `SubProblem`: representa una partición de un problema en partes disjuntas. Esta clase se ha introducido en aras de una mayor eficiencia en la implementación. La alternativa hubiera sido representar a los subproblemas para que fueran del mismo tipo que los problemas. Esta clase debe proporcionar las siguientes funcionalidades:
 - `initSubProblem(pbm)`: inicializa el subproblema de partida a partir del problema original.
 - `easy(pbm)`: determina si un subproblema es lo suficientemente pequeño para ser considerado simple.
 - `solve(pbm, sol)`: proporciona el algoritmo básico de resolución para subproblemas fáciles.
 - `divide(pbm, sps, aux)`: genera una lista de subproblemas.
- La clase `Auxiliar`: en algunos casos es necesario el uso de una clase auxiliar. Su necesidad se hace patente cuando como resultado de dividir el problema se obtiene una partición que no constituye un subproblema.

- Además la clase `Solution` debe proporcionar la siguiente funcionalidad:
 - `combine(pbm, sols, aux)` : dado un conjunto de soluciones parciales de subproblemas, este método ha de combinarlas para crear una nueva solución parcial más general.

2.1.2. *Interfaz del esqueleto de Ramificación y Acotación.* Los algoritmos de Ramificación y Acotación dividen el área de soluciones paso a paso y calculan una cota del posible valor de aquellas soluciones que pudieran encontrarse más adelante. Si la cota muestra que cualquiera de estas soluciones tiene que ser necesariamente peor que la mejor solución hallada hasta el momento, entonces no es necesario seguir explorando esta parte del árbol. Por lo general, el cálculo de cotas se combina con un recorrido en anchura o en profundidad. Para representar este proceso en un esqueleto, se utilizará la clase `Subproblem`. Así pues, el esqueleto BnB a diferencia de DnC sólo requiere que el usuario le suministre dicha clase (véase la Figura 3). Además el usuario debe especificar en la definición de la clase `Problem` si el problema a resolver es de máximo o de mínimo.

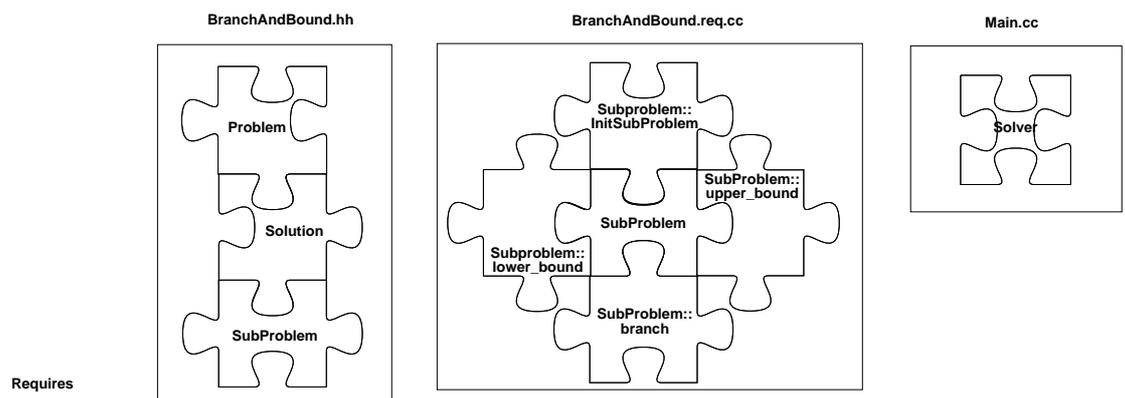


Figura 3. Clases requeridas por el esqueleto de Ramificación y Acotación.

- La clase `Subproblem`: representa el área de soluciones no exploradas. Esta clase debe proporcionar las siguientes funcionalidades:
 - `initSubProblem(pbm)` : inicializa el subproblema de partida a partir del problema original.
 - `solve(pbm)` : este método booleano devuelve cierto cuando el área de soluciones factibles a ser explorada es vacía.
 - `branch(pbm, sps, sol)` : ha de proporcionar un algoritmo para dividir el área de soluciones factibles; esto implica generar a partir del subproblema actual un subconjunto de problemas a ser explorados.
 - `lower_bound(pbm, sol)` : este método calcula una cota inferior del mejor valor de la función objetivo que podría ser obtenido para un subproblema dado.
 - `upper_bound(pbm, sol)` : esta función calcula una cota superior del mejor valor de la función objetivo que podría ser obtenido para un subproblema dado.

2.2. Clases proporcionadas. Las clases proporcionadas son las que el usuario ha de instanciar cuando utilice un esqueleto, esto es, sólo ha de utilizarlas. Dichas clases son:

- **Setup:** agrupa los parámetros de configuración del esqueleto. Por ejemplo, en esta clase en el esqueleto de Ramificación y Acotación se especifica si la búsqueda en el área de soluciones será en profundidad, primero el mejor, etc.
- **Solver:** esta clase está comprendida por ella misma y sus hijas `Solver_Seq`, `Solver_Lan` y `Solver_Wan`. Implementa la estrategia a seguir: Divide y Vencerás, Ramificación y Acotación, etc., y mantiene información actualizada del estado de la exploración durante la ejecución. La ejecución se lleva a cabo mediante una llamada al método `run()`.

Para elegir un patrón de resolución determinado, el usuario debe instanciar en el método `main()` la clase correspondiente. El siguiente código muestra un ejemplo de instanciación de un resolutor secuencial, para el esqueleto de Ramificación y Acotación.

```
int main (int argc, char** argv) \{
    using skeleton BranchAndBound;
    Problem pbm;
    Solution sol;
    Bound bs;

    //instancia del esqueleto secuencial
    Solver_Seq sv(pbm);
    //se ejecuta el esqueleto secuencial
    sv.run();
    //se obtiene la mejor solución
    bs = sv.bestSolution();
    //se obtiene la solución
    sol = sv.solution();
}
```

En primer lugar, se indica que se va a utilizar un esqueleto de Ramificación y Acotación. A continuación se crea una instancia del esqueleto secuencial (`Solver_Seq`). Para que se ejecute, es necesario realizar una llamada al método `run()`. Se invoca al método `bestSolution()` para obtener la mejor solución. La solución se obtiene utilizando el método `solution()` que proporciona el esqueleto.

3. Implementación de los patrones de resolución

En esta sección se describe la forma en la que se han diseñado e implementado los patrones de resolución generales tanto en secuencial como en paralelo.

Las implementaciones de DnC y BnB se han realizado mediante un esquema *Maestro-Eslavo*. En el patrón maestro-esclavo un procesador *Maestro (Master)* controla la actividad de un grupo de procesadores *Esclavos (Slaves)*, asignando el trabajo que se ha de realizar en paralelo y ocupándose también de las operaciones de entrada/salida.

Dado un problema de grandes dimensiones, se divide y se distribuye sobre los esclavos disponibles. Cada Esclavo calcula los resultados parciales de manera independiente y se los envía al Maestro (véase la Figura 4).

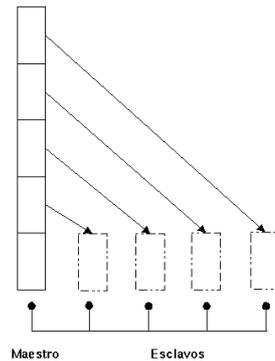


Figura 4. Esquema Maestro-Eslavo.

Las tareas del Maestro consisten en distribuir trabajo a los componentes Esclavos, arrancar la ejecución de los esclavos y generar el resultado final usando los resultados de los diferentes esclavos. Un esclavo ha de implementar la tarea que le indique el maestro.

3.1. Esqueletos Divide y Vencerás. La estrategia Divide y Vencerás consiste en descomponer un problema en subproblemas más simples del mismo tipo, resolverlos de forma independiente y una vez obtenidas las soluciones parciales combinarlas para obtener la solución del problema original. El siguiente código refleja la estructura general de un algoritmo basado en la estrategia Divide y Vencerás:

```

procedure DandC(pbm, sol)
  local var aux;
  begin
    if easy(pbm) then solve(pbm)
    else begin
      divide(pbm, subpbm, aux);
      for i := 1 to numProblem() do
        DandC(subpbm[i], subsol[i]);
      combine(subsol, aux, sol);
    end
  end;
end;

```

En el caso de DnC se ha utilizado el patrón Maestro/Esclavo dos veces, uno para la *Fase de División* del problema original en subproblemas y una segunda vez en la *Fase de Combinación* de soluciones parciales para obtener la solución del problema original. La división del trabajo entre las subtareas y la evaluación conjunta de los resultados de las subtareas están completamente separados del procesamiento individual de cada subtarea.

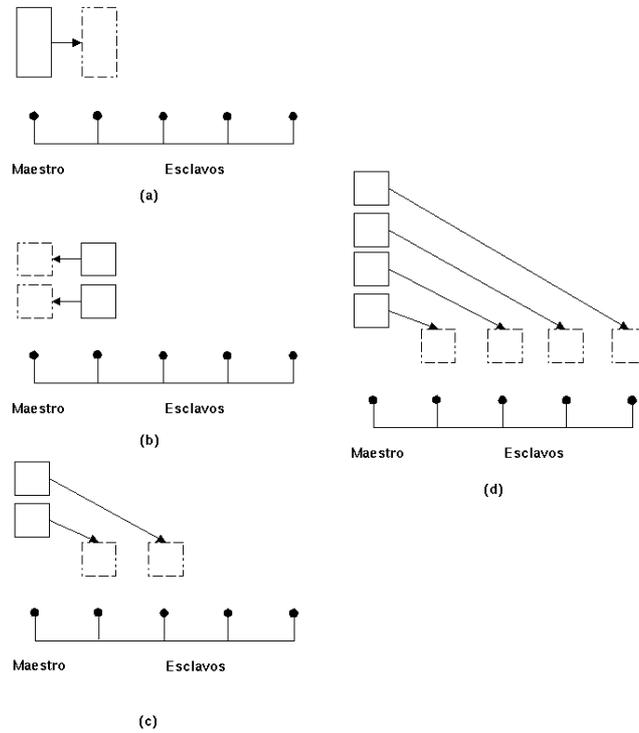


Figura 5. Esquema de la Fase de División.

En la Figura 5 (a) el maestro envía el problema al primer esclavo libre. Éste divide el problema, en este caso en dos, y le devuelve los dos nuevos subproblemas al maestro (b). El maestro vuelve a distribuir los problemas que le quedan sin resolver entre los procesadores que están ociosos (c). Finalmente todos los esclavos estarán trabajando en la división de algún subproblema (d).

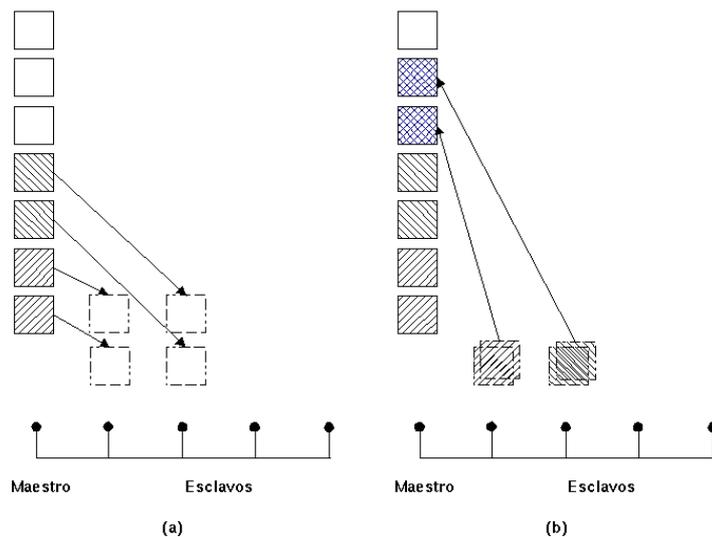


Figura 6. Esquema de la Fase de Combinación.

En la Figura 6 (a) el maestro envía las soluciones hermanas a los esclavos libres. Los esclavos combinan las soluciones, en este caso dos, y le devuelve la nueva solución al maestro (b) quien la coloca en el lugar correspondiente de la estructura de datos para repetir el esquema y enviar a un esclavo un nuevo conjunto de soluciones hermanas.

3.2. Esqueletos Ramificación y Acotación. El siguiente código muestra un esquema de Ramificación y Acotación para un problema de minimización:

```

procedure BandB (pbm, sp, sol)
begin
  SubProblem Queue[];
  Solution sl;
  Bound high, low, bestSol;

  bestSol := INFINITE;
  Queue.insert(sp);
  while not Queue.empty() do begin
    sp := Queue.extract();
    high := sp.upper_bound (pbm, sl);
    if high > bestSol then begin
      low := sp.lower_bound(pbm, sl);
      if low > bestSol then begin
        bestsol := low;
        sol := sl;
      end;
      if not sp.solve(pbm) then
        sp.branch(pbm, HP);
      end;
    end;
  return bestSol;
end;

```

La implementación paralela de BnB se realiza siguiendo el mismo esquema que el de la *Fase de División* de la técnica Divide y Vencerás.

4. Un caso de estudio: Problema de la Mochila 0/1

Se ha implementado utilizando los esqueletos Divide y Vencerás y Ramificación y Acotación el algoritmo que describen Martelo y Toth [11] para la resolución del problema de la Mochila 0/1 clásico. En dicho algoritmo, es necesario que los elementos estén ordenados de forma ascendente mediante la relación peso/beneficio. Haciendo uso de este requisito se ha implementado el método de ordenación Quicksort [6] con DnC para ordenar los elementos a insertar en la Mochila. A continuación mediante el uso de BnB se ha implementado el algoritmo de Ramificación y Acotación, lo que nos ha permitido utilizar de forma integrada ambos esqueletos.

4.1. Definición del problema. Se puede plantear el Problema de la Mochila 0/1 de la siguiente forma:

“Se dispone de una mochila de capacidad C y de un conjunto de N objetos. Se supone que los objetos no se pueden fragmentar en trozos más pequeños; así pues, se puede decidir si se toma un objeto o si se deja, pero no se puede tomar una fracción de un objeto. Supóngase además que el objeto k tiene beneficio b_k y peso p_k , para $k = 1, 2, \dots, N$. El problema consiste en averiguar qué objetos se han de insertar en la mochila sin exceder su capacidad, de manera que el beneficio que se obtenga sea máximo.”

Su formulación como un problema de optimización es:

$$\max \sum_{k=1}^N p_k x_k \leq C$$

sujeto a:

$$x_k \in \{0, 1\}, k \in \{1, \dots, N\}$$

4.2. Usando DnC. Para que el algoritmo propuesto funcione correctamente los elementos a insertar en la mochila se han de ordenar en función de la relación peso/beneficio: $b_1/p_1 \geq b_2/p_2 \geq \dots \geq b_N/p_N$.

El *Quicksort* trabaja particionando el conjunto a ordenar en dos partes, para después ordenar dichas partes independientemente. El punto clave del algoritmo está en el procedimiento que divide el conjunto. El proceso de división del conjunto debe cumplir las siguientes tres condiciones:

- El elemento pivote= $a[i]$ está en su posición final en el array para algún índice i .
- Todos los elementos en $a[\text{first}]$, ..., $a[i-1]$ son menores o iguales a $a[i]$.
- Todos los elementos en $a[i+1]$, ..., $a[\text{last}]$ son mayores que $a[i]$.

En este punto se aplica el mismo método recursivamente a los dos subproblemas generados: $a[\text{first}]$, ... , $a[i-1]$ y $a[i+1]$, ... , $a[\text{last}]$. El resultado final será una matriz completamente ordenada, y por tanto no hace falta un paso subsiguiente de combinación.

La clase `Problem` estará formada por una variable miembro que representa el vector de enteros a ordenar. Dicha variable se ha declarado con el tipo `vector` de la librería STL de C++. Tanto la clase `SubProblem` como la clase `Solution` también tendrán una variable miembro del mismo tipo, que contiene en el primer caso un vector, de tamaño menor que el original, a ordenar, y en el segundo caso un vector de enteros ordenados que representan la solución parcial a un determinado subproblema.

`initSubProblem()` : El subproblema de partida tendrá el mismo contenido que el problema original. Por tanto, este método de la clase `SubProblem` simplemente copia los elementos del problema original en el subproblema.

```

SubProblem::initSubProblem (const Problem& pbm) {
    for (unsigned i = 0; i < pbm.l.size(); i++)
        l.push_back(pbm.l[i]);
}

```

`easy()` : Se considera que un subproblema es fácil para resolverlo por un algoritmo trivial cuando su tamaño es menor o igual que uno. Por tanto, la implementación del método queda como sigue:

```

bool SubProblem::easy () {
    return (l.size() <= 1);
}

```

`divide()` : La división del subproblema da lugar a dos nuevos subproblemas, donde el primero de ellos contiene los elementos menores que el pivote, mientras que el segundo contendrá los elementos mayores al pivote. A su vez, los elementos iguales al pivote se almacenarán en el objeto Auxiliar asociado a dicho subproblema. Se ha escogido como elemento pivote para la división el primer elemento del array de enteros a ordenar.

```

SubProblem::divide (const Problem& pbm,
    vector<SubProblem>& subpbms,
    Auxiliar& aux){
    SubProblem sp1; // Elms. menores al pivote.
    SubProblem sp2; // Elms. mayores al pivote.
    // pivote = primer elemento de la lista.
    const int& pivot = l[0];
    // Division del subproblem.
    for (unsigned i = 0; i < l.size(); i++) {
        if (l[i] < pivot) sp1.l.push_back(l[i]);
        if (l[i] == pivot) aux.l.push_back(l[i]);
        if (l[i] > pivot) sp2.l.push_back(l[i]);
    }
    subpbms.push_back(sp1);
    subpbms.push_back(sp2);
}

```

`combine()` : Para llevar a cabo la combinación, como las dos soluciones a combinar ya contienen los elementos ordenados correspondientes a dos subproblemas para los que los elementos del primero son menores que los elementos del segundo, simplemente hay que concatenar el contenido de la primera solución con el del auxiliar (elementos iguales), y finalmente con los elementos de la segunda solución.

```

Solution::combine (const Problem& pbm,
    const Auxiliar& aux),
    const vector<Solution>& subsols) {
    unsigned i;
    for (i = 0; i < subsols[0].l.size(); i++)
        l.push_back(subsols[0].l[i]);
}

```

```

        for (i=0; i<aux.l.size(); i++)
            l.push_back(aux.l[i]);
        for (i=0; i<subsols[1].l.size(); i++)
            l.push_back(subsols[1].l[i]);
    }

```

4.3. Implementación con BnB. La instanciación con MaLLBa del problema de la mochila queda de la siguiente forma:

Problem: se definen los datos del problema, el número de objetos, la capacidad de la mochila y el peso y beneficio de cada objeto.

```

requires class Problem {
    public:
        Number      C, // Capacidad
                  N; // Número de elementos
        vector<Number> p, //beneficios
                  w; //pesos
        Problem (); //Constructor
        inline Direction direction() const
        { return Maximize; }
        ...
}

```

Solution: se representa mediante un vector que contiene un verdadero o falso dependiendo de si el objeto pertenece o no a la solución.

```

requires class Solution {
    public:
        vector<bool> s;
        ...
}

```

SubProblem: en cada subproblema se usarán los siguientes datos: la capacidad actual, el siguiente objeto a considerar, el beneficio actual y la solución actual.

```

requires class SubProblem {
    public:
        Number      CRest, // capacidad actual
                  obj, // siguiente objeto
                  profit; // beneficio actual
        Solution sol; // solucion actual
        SubProblem (); // constructor
        ...
}

```

initSubProblem(): el primer subproblema considera toda la capacidad de la mochila; no se ha insertado ningún objeto, por lo tanto el beneficio obtenido es nulo.

```

void SubProblem::initSubProblem (
    const Problem& pbm){
    CRest = pbm.C;
    obj = 0;
    profit = 0;
}

```

branch(): partiendo de un subproblema, genera dos nuevos, considerando la inserción o no de otro elemento.

```

void SubProblem::branch (
    const Problem& pbm,
    container<SubProblem>& subpbms){
    SubProblem spNO; SubProblem spYES;

    spNO = SubProblem(CRest, obj+1, profit);
    spNO.sol.update(false, sol);
    subpbms.insert(spNO);

    Number newC = CRest - pbm.w[obj];
    if (newC >= 0) {
        Number newPf = profit + pbm.p[obj];
        spYES = SubProblem(newC, obj+1, newPf);
        spYES.sol.update(true, sol);
        subpbms.insert(spYES);
    }
}

```

upper_bound(): calcula la cota superior incluyendo objetos hasta que la capacidad actual sea alcanzada, y para este último objeto sólo se incluye la proporción de capacidad que quepa.

```

Bound SubProblem::upper_bound (
    Const Problem& pbm){
    Bound upper, weigth, pft; Number i;

    for(i=obj, weigth=0, pft=profit;
        weigth<=CRest; i++) {
        weigth += pbm.w[i];
        pft += pbm.p[i];
    }
    i--;
    weigth -= pbm.w[i]; pft -= pbm.p[i];
    upper = pft + (Number)((pbm.p[i]*
        (CRest- eigth))/pbm.w[i]);
    return(upper);
}

```

`lower_bound()`: calcula la cota inferior incluyendo los objetos hasta que la capacidad actual sea alcanzada. El último objeto que haga que se supere la capacidad no será incluido.

```
Bound SubProblem::lower_bound (
    const Problem& pbm, Solution& us){
    Bound lower, weight, pft; Number i, tmp;
    us = sol;

    for(i = obj, weight = 0, pft = profit;
        weight <= CRest; i++){
        weight += pbm.w[i];
        pft += pbm.p[i];
        us.s.push_back(true);
        i--;
        weight -= pbm.w[i]; pft -=pbm.p[i];
        us.s.pop_back();
        tmp = pbm.N - us.s.size();
        for (Number j = 0; j < tmp; j++)
            us.s.push_back(false);
        lower = pft;
        return(lower);
    }
}
```

`solve()`: el problema está resuelto cuando no quedan objetos o cuando la mochila está llena.

```
bool SubProblem::solve(const Problem& pbm){
    return ((obj >= pbm.N) || (CRest <= 0));
}
```

5. Conclusiones

Se han presentado los esqueletos de la librería MaLLBa para la estrategia Divide y Vencerás (DnC) y para el paradigma de Ramificación y Acotación (BnB). La herramienta MaLLBa ofrece al usuario libertad para implementar las estructuras de datos que representan su problema, pero proporciona unos patrones de resolución que controlan el flujo de la ejecución.

El principal objetivo de MaLLBa es simplificar la tarea de los investigadores o usuarios que han de implementar un algoritmo usando una técnica algorítmica particular. MaLLBa proporciona al usuario un valor añadido no sólo en términos de la cantidad de código que ha de escribir, que es mucho menor, sino también en modularidad y claridad conceptual. Si se utiliza MaLLBa, o cualquier otro sistema orientado a objetos, el usuario ha de colocar cada pieza de código en la posición correcta.

Otra de las características de MaLLBa es la modularidad: una vez que las estructuras de datos básicas y las funcionalidades se han definido e insertado en el código, el esqueleto proporciona una implementación gratuita en paralelo.

MaLLBa hace un uso equilibrado de la programación orientada a objetos, en el sentido de que proporciona las mínimas plantillas necesarias para conseguir una eficiencia computacional buena y al mismo tiempo hacer el proceso de compilación seguro.

Además el sistema pone a disposición del usuario la posibilidad de generar y experimentar con nuevos algoritmos que permitan la integración de técnicas. En este trabajo se ha mostrado la posibilidad de esta integración mediante el ejemplo del Problema de la Mochila.

Reconocimientos

Han colaborado en la realización del proyecto MaLLBa::DnC y MaLLBa::BnB Isabel Dorta González, Casiano Rodríguez León y Angélica Rojas Rodríguez.

Bibliografía

- [1] Alba, E., Almeida, F., Blesa, M., Cabeza, J., Cotta, C., Díaz, M., Dorta, I., Gabarró, J., León, C., Luna, J., Moreno, L., Petit, J., Rojas, A., Xhafa, F.: MaLLBa: A Library of skeletons for combinatorial optimization. En *Proceedings of the International Euro-Par Conference (Paderborn, Germany)*. LNCS **2400** (2002), 927-932.
- [2] Alba, E., Almeida, F., Blesa, M., Cotta, C., Díaz, M., Dorta, I., Gabarró, J., González, J., León, C., Moreno, L., Petit, J., Roda, J., Rojas, A., Xhafa, F.: MaLLBa: Towards a Combinatorial Optimization Library for Geographically Distributed Systems. Actas de las XII Jornadas de Paralelismo (2001), 105-110.
- [3] Dorta, I., Rojas, A., León, C., Dorta P.: *Utilización de software en la docencia de técnicas algorítmicas*. Actas de las VII Jornadas de Enseñanza Universitaria de la Informática, 2001 .
- [4] Eckstein, J., Phillips, C.A., Hart, W.E.: *PICO: An Object-Oriented Framework for Parallel Branch and Bound*. Rutcor Research Report (2000).
- [5] High Performance Fortran Forum: *High Performance Fortran Specification*, 1993.
- [6] Hoare, C.A.R., Algorithm 64: Quicksort. *Communications of the ACM*, **4**:321 (1961).
- [7] Kielmann, T., Nieuwpoort, R., Bal, H.: *Cilk-5.3 Reference Manual*. Supercomputing Technologies Group, 2000.
- [8] Kielmann, T., Nieuwpoort, R., Bal, H.: Satin: *Efficient Parallel Divide-and-Conquer in Java*. Euro-Par 2000 (2000), 690-699.
- [9] León, C., Rojas A.: *Un Esqueleto Paralelo para el Paradigma Divide y Vencerás*. Documento de Trabajo Interno. DEIOC-01-01, 2001.
- [10] B. Le Cun, C. Roucairol, The PNN Team: *BOB: A unified platform for implementing branch-and-bound like algorithms*. Rapport de Recherche 95/16 (1999).
- [11] Martello, S., Toth, P.: *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons Ltd., 1990.
- [12] OpenMP Architecture Review Board: *OpenMP C and C++ Application Program Interface, Version 1.0* (1998). [Disponible en <http://www.openmp.org>].
- [13] Shinano, Y., Higaki, M., Hirabayashi, R.: *A Generalized Utility for Parallel Branch and Bound Algorithms*. IEEE Computer Society Press (1995), 392-401.
- [14] Tschöke, S., Polzer, T.: *Portable Parallel Branch-and-Bound Library*. User Manual Library Version 2.0, Paderborn, 1995.

En Internet

<http://cvs.deioc.ull.es>

PCGULL

Parallel Computing Group at the University of La Laguna.

<http://www.cs.sandia.gov/~caphill/proj/pico.html>

PICO

Parallel Integer and Combinatorial Optimization.

<http://www.prism.uvsq.fr/~blec/Research/BOBO>

Bob++

Parallel and sequential parallel branch-and-bound.

Resolución de Problemas en Paralelo

Coromoto León Hernández
Sociedad, Ciencia, Tecnología y Matemáticas.
28 de Marzo de 2003
URL: <http://nereida.deioc.uill.es/~cleon>
e-mail: cleon@ull.es



A mis padres

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 2

¿A qué tipo de problemas hace referencia el título?



- Problemas de **Optimización**
 - Existe una entrada de tamaño n en la que están los candidatos a formar parte de la solución
 - Existe un subconjunto de esos n candidatos que satisface ciertas restricciones (**soluciones factibles**)
 - Hay que obtener la solución factible que minimiza o maximiza una cierta función objetivo (**solución óptima**)

$$\min \quad | \quad \max \quad f(x)$$

sujeto a: $x \in S$

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 3

¿Cómo proponemos resolverlos?

- Un **Algoritmo** es un conjunto ordenado de operaciones que deben efectuarse para resolver un problema dado.



28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 4

¿Cómo podemos abordar la resolución de un problema dado?

- Con lápiz y papel
- Con un ordenador secuencial
- Con una máquina paralela





28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 5

Proceso de resolución



28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 6

Contenido de la charla

- Un caso de estudio: El Problema de Mochila 0/1
- Algoritmos...
- Ordenadores un poco de historia
- Técnicas Algorítmicas
 - Ramificación y Acotación
 - Divide y Vencerás
- Esqueletos Algorítmicos:
 - Secuenciales
 - Paralelos
- Conclusiones



28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

7

El Problema de la Mochila

$b = 100$, $p = 20$ (key)
 $b = 60$, $p = 30$ (screwdriver)
 $b = 15$, $p = 2$ (umbrella)
 $b = 90$, $p = 20$ (axe)
 $b = 1$, $p = 10$ (scarf)
 $b = 40$, $p = 40$ (flower)
 $b = 15$, $p = 30$ (book)
 $b = 10$, $p = 60$ (box)
 La capacidad C es 102
 Los elementos N son 8
 Beneficios $b_k = \{15, 100, 90, 60, 40, 15, 10, 1\}$
 Pesos $p_k = \{2, 20, 20, 30, 40, 30, 60, 10\}$

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

8

El Problema de la Mochila

Este problema lo podemos formular de la siguiente forma: "Se dispone de una mochila de capacidad C y de un conjunto de N objetos, siendo $b[k]$ y $p[k]$ el beneficio y el peso asociado al objeto k . Sin exceder la capacidad de la mochila, los objetos deben ser insertados en la mochila proporcionando el máximo beneficio"

$$\begin{aligned}
 \text{max} \quad & \sum_{k=1}^N b_k x_k \\
 \text{s.t.} \quad & a: \sum_{k=1}^N p_k x_k \leq C \\
 & x_k \in \{0, 1\} \\
 & \forall k \in \{1, \dots, N\}
 \end{aligned}$$

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

9

Algoritmos (Al Khuwârizmî)

- Intuitivamente un **algoritmo** es una sucesión finita de reglas elementales, regidas por una prescripción precisa y uniforme, que permite efectuar paso a paso, en un encadenamiento estricto y riguroso, ciertas operaciones de tipo ejecutable, con vistas a la resolución de problemas de una misma clase.

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

10

Alan Mathison Turing (1912-1954)

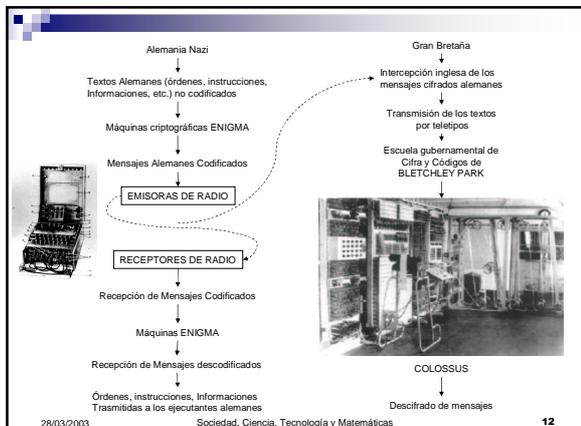
- "Sobre los números calculables y su aplicación al problema de la decidibilidad" (1936)
 - Define rigurosamente un Algoritmo
 - Introduce los conceptos de
 - Máquina de Turing y
 - Autómata Algorítmico Universal



28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

11



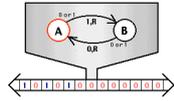
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

12

Máquinas de Turing

- Existe una infinidad de Máquinas de Turing, cada una de las cuales define la estructura de una familia de dispositivos artificiales:
 - La máquina de Pascal
 - Las máquinas de referencias
 - Las calculadoras de tarjetas perforadas
 - La máquina Analítica de Babbage
 - La máquina de criptoanálisis Colossus
 - Los ordenadores



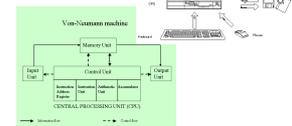
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

13

John von Neumann (1903-1957)

- “First draft of a report on EDVAC” . Publicado el 30 de junio de 1945. (EDVAC = **E**lectronic **D**iscrete **V**ariable **A**utomatic **C**omputer)



Programa Almacenado

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

14



28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

15



Técnicas Algorítmicas

- Son técnicas de diseño de algoritmos, que se adaptan al problema particular que se desea resolver.
- Fuerza bruta
- Divide y Vencerás (*Divide and Conquer*)
- Búsqueda con retroceso (*Backtracking*)
- Ramificación y Acotación (*Branch and Bound*)
- ...

28/03/2003

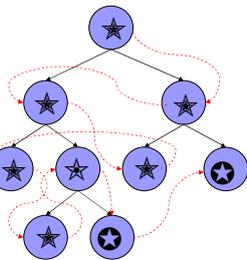
Sociedad, Ciencia, Tecnología y Matemáticas

16



Ramificación y Acotación

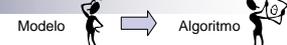
- Árbol de búsqueda
- Nodos:
 - ★ Vivo : no se han generac aún todos sus hijos
 - ☆ Muerto: no se van a generar más hijos
 - ☆ Actual: se están generan sus hijos
- Lista de Nodos vivos



28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

17



Ramificación y Acotación

- Paso 1: Si la lista de nodos vivos está vacía, terminar.
- Paso 2: Extraer un nodo vivo
- Paso 3: Calcular su **coste esperado**.
- Paso 4: Si el coste esperado es peor que el de la mejor solución hasta el momento, matarlo y volver al paso 1.
- Paso 5: Si el coste esperado es mejor que el de la mejor solución actual pero no **es solución**, **generar** todos sus hijos y matarlo. Ir al paso 1.
- Paso 6: Si el coste esperado es mejor que el de la mejor solución actual y es solución, ésta pasa a ser la mejor solución. Ir al paso 1.

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

18

El Problema de la Mochila

$$\max \sum_{i=1}^n p_i x_i$$

$$\text{s.t. } \sum_{i=1}^n p_i x_i \leq C$$

$$x_i \in \{0,1\}$$

$$\forall i \in \{1, \dots, N\}$$

Árbol de búsqueda

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 19

El Problema de la Ordenación

$$\frac{b_1}{p_1} > \frac{b_2}{p_2} > \frac{b_3}{p_3} > \dots > \frac{b_N}{p_N}$$

Beneficios $b_i = \{15, 100, 90, 60, 40, 15, 10, 1\}$
 Pesos $p_i = \{2, 20, 20, 30, 40, 30, 60, 10\}$

1 2 3 4 5 6 7 8

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 20

Árbol de Búsqueda

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 21

Divide y Vencerás

Modelo \rightarrow Algoritmo

- **Descomponer** el problema a resolver en un cierto número de subproblemas más pequeños que el original, pero con la misma estructura
- **Resolver** independientemente cada subproblema
- **Combinar** los resultados obtenidos para obtener la solución al problema original

Aplicar esta técnica recursivamente

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 22

Divide y Vencerás

Modelo \rightarrow Algoritmo

- **Paso 1:** Si x es suficientemente simple resolverlo, devolver la solución e ir al paso 6. En caso contrario ir al paso 2.
- **Paso 2:** descomponer x en casos más pequeños x_1, x_2, \dots, x_k
- **Paso 3:** para $i = 1, \dots, k$ aplicar el mismo procedimiento hasta obtener $y_i = \text{Divide y Vencerás}(x_i)$
- **Paso 4:** recombinar los y_i , para obtener una solución y de x
- **Paso 5:** devolver y
- **Paso 6:** finalizar

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 23

El Problema de la Ordenación

$$\frac{b_1}{p_1} > \frac{b_2}{p_2} > \frac{b_3}{p_3} > \dots > \frac{b_N}{p_N}$$

Fase de división

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 24

El Problema de la Ordenación

$\frac{b_1}{p_1} > \frac{b_2}{p_2} > \frac{b_3}{p_3} > \dots > \frac{b_n}{p_n}$

Fase de combinación

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 25

Algoritmo → Implementación Secuencial

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 26

Esqueleto

La definición que se da de esqueleto en el diccionario de la Lengua Española de la Real Academia es la que sigue:

"Anat. Conjunto de piezas duras y resistentes, por lo regular trabadas y articuladas entre sí, que da consistencia al cuerpo de los animales, sosteniendo o protegiendo sus partes blandas. Armazón que sostiene algo. Col., Cos. Rica, Guat., Méx., y Nic., Modelo o patrón impreso en el que se dejan blancos que se rellenan a mano.

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 27

Esqueletos Algorítmicos

Se denomina esqueleto algorítmico a un conjunto de procedimientos que constituyen el armazón con el que desarrollar programas para resolver un problema dado utilizando una técnica particular.

Este software tiene algunos blancos que se han de rellenar para adaptarlo a la resolución de un problema concreto.

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 28

Librería de Esqueletos MaLLBa

Casiano Rodríguez León

<http://www.lsi.upc.es/~mallba>

LCC - UMA. Málaga

EIOC - ULL. La Laguna

LSI - UPC. Barcelona

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 29

Esqueletos MaLLBa

Técnica.req.cc

Técnica.pro.cc

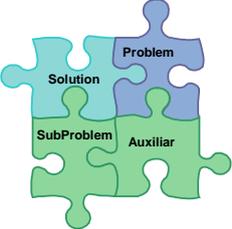
Técnica.hh

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 30

Angélica Rojas Rodríguez

Parte requerida del Esqueleto Divide y Vencerás

- Problem
- SubProblem
 - initSubProblem
 - easy
 - solve
 - divide
- Solution
 - combine
- Auxiliar



28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 31

$\frac{b_1}{p_1} > \frac{b_2}{p_2} > \frac{b_3}{p_3} > \dots > \frac{b_n}{p_n}$

El problema de Ordenación

```

void SubProblem::initSubProblem (const Problem& pbm) {
    for (unsigned i = 0; i < pbm.l.size(); i++)
        l.push_back(pbm.l[i]);
}

bool SubProblem::easy () {
    return (l.size() <= 1);
}

void SubProblem::solve (Solution& sol) {
    sol.l = l;
}
  
```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 32

$\frac{b_1}{p_1} > \frac{b_2}{p_2} > \frac{b_3}{p_3} > \dots > \frac{b_n}{p_n}$

El problema de Ordenación

```

void SubProblem::divide (const Problem& pbm,
                        vector<SubProblem>& subpbms,
                        Auxiliar& aux){

    SubProblem sp1, sp2;
    unsigned i, middle = l.size() / 2;

    for (i = 0; i < middle; i++)
        sp1.l.push_back(l[i]);
    for (i = middle; i < l.size(); i++)
        sp2.l.push_back(l[i]);
    subpbms.push_back(sp1);
    subpbms.push_back(sp2);
}
  
```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 33

$\frac{b_1}{p_1} > \frac{b_2}{p_2} > \frac{b_3}{p_3} > \dots > \frac{b_n}{p_n}$

El problema de Ordenación

```

void Solution::combine (const Problem& pbm,
                       const Auxiliar& aux,
                       const vector<Solution>& subsols) {

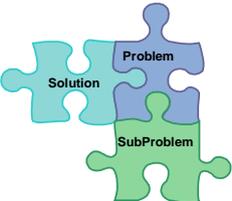
    vector<int>::const_iterator i = subsols[0].l.begin();
    vector<int>::const_iterator j = subsols[1].l.begin();
    while ((i != subsols[0].l.end()) && (j != subsols[1].l.end())) {
        if (*i < *j) { l.push_back(*i); i++; }
        else { l.push_back(*j); j++; }
    }
    while (i != subsols[0].l.end()) { l.push_back(*i); i++; }
    while (j != subsols[1].l.end()) { l.push_back(*j); j++; }
}
  
```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 34

Isabel Dorta González

Parte requerida del Esqueleto de Ramificación y Acotación

- Problem
- SubProblem
 - initSubProblem
 - solve
 - branch
 - lower_bound
 - upper_bound
- Solution



28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 35

$\sum_{i \in [N]} x_i$
 $\sum_{i \in [N]} c_i x_i$
 $x_i \in \{0,1\}$
 $\forall i \in [N]$

El problema de la Mochila

```

void SubProblem::initSubProblem(const Problem& pbm){
    CRest = pbm.C; obj = 0; profit = 0;
}

Bound SubProblem::upper_bound(const Problem& pbm, Solution& sl){
    Number weigth, pft, i;
    for(i = obj, weigth = 0, pft = profit;
        weigth <= CRest; i++){
        weigth += pbm.w[i];
        pft += pbm.p[i];
    }
    if (i != obj){
        i--; weigth -= pbm.w[i]; pft -= pbm.p[i];
    }
    if ((i == obj) && (obj == (pbm.N - 1))) return (Bound)pft;
    return (Bound)(pft);
}
  
```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 36

El problema de la Mochila

$$z = \sum_{i=1}^n \lambda_i x_i$$

$$s.t. \sum_{i=1}^n \lambda_i c_i \leq C$$

$$x_i \in \{0,1\}$$

$$\forall i \in \{1, \dots, n\}$$

```

Bound SubProblem::lower_bound(const Problem& pbm, Solution& us){
    Number weight, pft, i, tmp;
    us = sol;
    i = obj;
    weight = 0;
    pft = profit;
    while ( (weight <= CRest) && (i < pbm.N) ) {
        if (pbm.w[i] <= (CRest-weight)){
            weight += pbm.w[i]; pft += pbm.p[i]; us.s.push_back(true);
        }
        else us.s.push_back(false);
        i++;
    }
    tmp = pbm.N - us.s.size();
    for (Number j = 0; i < tmp; j++)
        us.s.push_back(false);
    return (Bound)pft;
}

```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 37

El problema de la Mochila

$$z = \sum_{i=1}^n \lambda_i x_i$$

$$s.t. \sum_{i=1}^n \lambda_i c_i \leq C$$

$$x_i \in \{0,1\}$$

$$\forall i \in \{1, \dots, n\}$$

```

void SubProblem::branch (const Problem& pbm,
    branchQueue<SubProblem>& subpbms){
    SubProblem spNO, spYES;

    Number nextObject = obj + 1;
    if (nextObject <= pbm.N) {
        spNO = SubProblem(CRest, nextObject, profit);
        spNO.sol.update(false, sol);
        subpbms.insert(spNO);
        Number newC = CRest - pbm.w[obj];
        if (newC >= 0) {
            Number newPf = profit + pbm.p[obj];
            spYES = SubProblem(newC, nextObject, newPf);
            spYES.sol.update(true, sol);
            subpbms.insert(spYES);
        }
    }
}

```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 38

Parte Proporcionalada

- **SetUp**
 - Por ejemplo, el tipo de recorrido del espacio de búsqueda.
- **Solver**
 - Implementaciones secuenciales
 - Implementaciones paralelas

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 39

Esqueleto Divide y Vencerás Secuencial

```

procedure DandC(sp, sol) {
    if (sp.easy()) {
        sol = sp.solve();
    }
    else {
        sp.divide(<array>subpbm);
        <array>subsol;
        for i := 1 to subpbm.numProblem() do
            DandC(subpbm[i], subsol[i]);
        }
        sp.combine(subsol[i], sol);
    }
}

```

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 40

Esqueleto de Ramificación y Acotación Secuencial

```

Bound BB (const Problem& pbm, const SubProblem& sp, Solution& sol){
    branchQueue<SubProblem> bqueue;
    Solution sl;
    Bound high, low;
    SubProblem sp;
    bqueue.insert(sp);
    while (!bqueue.empty()) {
        sp = bqueue.remove();
        high = sp.upper_bound (pbm, sl);
        if (high > bestSol){
            low = sp.lower_bound(pbm, sl);
            if (low > bestSol) {
                bestSol = low;
                sol = sl;
            }
        }
        if (!sp.solve(pbm)) sp.branch(pbm, bqueue);
    }
    return(bestSol);
}

```

Maximización

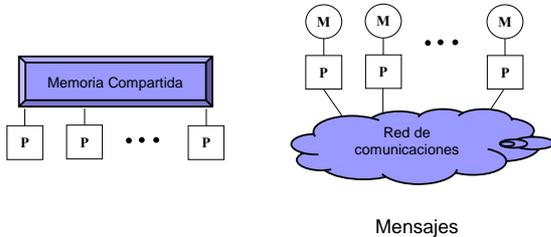
28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 41

Computación de Altas Prestaciones

Scientific Computing Division
History of Supercomputing at the National Center for Atmospheric Research

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 42

Aproximaciones a las máquinas paralela



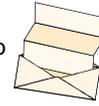
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

43

Paradigma de Paso de Mensajes. MPI (Message Passing Interface)

■ Contenido



Mensaje

■ Sobre



- Qué procesador envía el mensaje
- Dónde están los datos en el procesador emisor
- Qué clase de datos se están enviando y cuántos.
- Qué procesador(es) tienen que recibir el mensaje.
- Dónde se deben dejar los datos en el procesador receptor.
- Cuántos datos debe estar el procesador receptor preparado para recibir

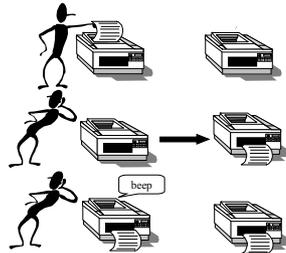
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

44

Comunicaciones Síncronas

Una comunicación síncrona no se completa hasta que el mensaje ha sido recibido



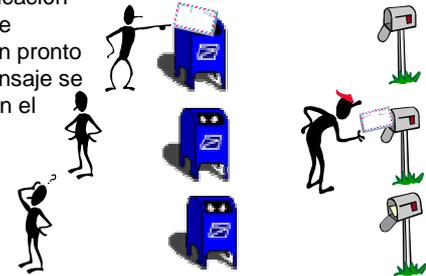
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

45

Comunicaciones Asíncronas

Una comunicación asíncrona se completa tan pronto como el mensaje se ha puesto en el medio de transporte.



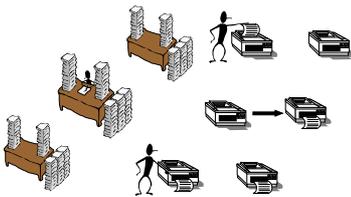
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

46

Comunicaciones bloqueantes y no bloqueantes

- Las comunicaciones **bloqueantes** sólo terminan cuando la correspondiente comunicación se ha completado.
- La comunicaciones **no bloqueantes** permiten realizar trabajo útil mientras se espera que se termine la comunicación.



28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

47

Comunicaciones colectivas (1)

Una **barrera** es una operación que permite sincronizar procesos. No hay intercambio de datos pero la barrera impide el paso hasta que todos los procesadores participantes lleguen ahí.



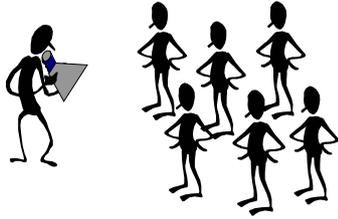
28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

48

Comunicaciones colectivas (2)

La **emisión** (broadcast) es una comunicación uno-a-muchos. Con una sola operación un procesador envía el mismo mensaje a varios.

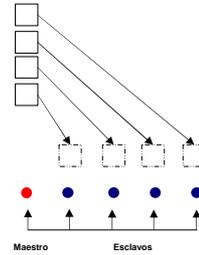


28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

49

Paradigma Maestro/Esclavo

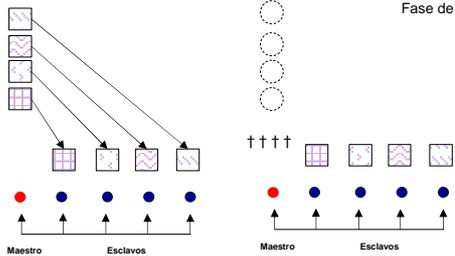


28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

50

Esqueleto Divide y Vencerás Paralelo



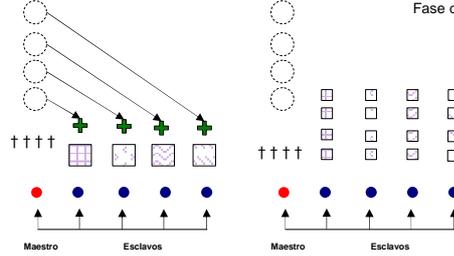
Fase de división

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

51

Esqueleto Divide y Vencerás Paralelo



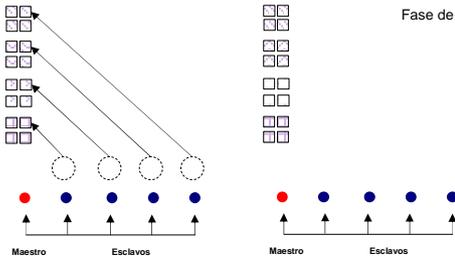
Fase de división

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

52

Esqueleto Divide y Vencerás Paralelo



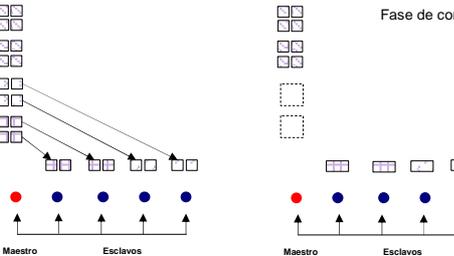
Fase de división

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

53

Esqueleto Divide y Vencerás Paralelo

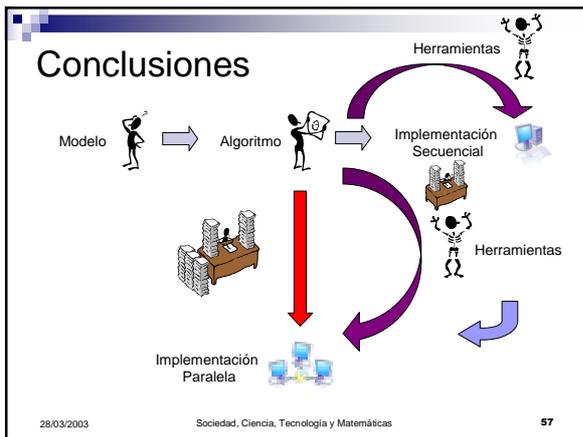
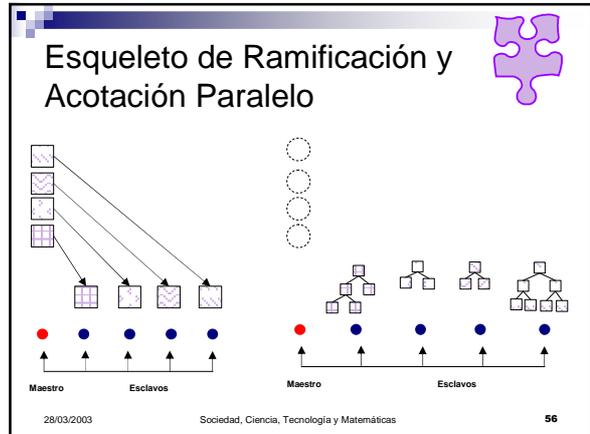
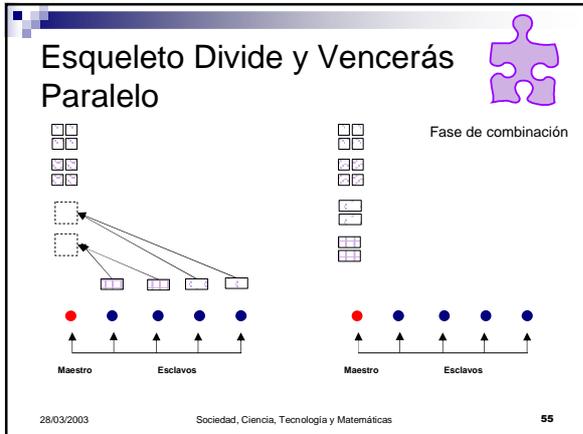


Fase de combinación

28/03/2003

Sociedad, Ciencia, Tecnología y Matemáticas

54



Otras Referencias

Georges Ifrah. Historia Universal de las Cifras. 2ª edición. Editorial Espasa Calpe, S.A.

<http://nereida.deioc.ull.es/~cleon>

28/03/2003 Sociedad, Ciencia, Tecnología y Matemáticas 58