Material de divulgació del CRM

CRM
CENTRE DE RECERCA MATEMÀTICA

Les bases matemàtiques de la civilització tecnològica

Cicle Ferran Sunyer i Balaguer

Aula de Ciència i Cultura, 10

AULA DE CIÈNCIA
I CULTURA, 10

Les bases matemàtiques de la civilització tecnològica

Cicle Ferran Sunyer i Balaguer

Jaume Aguadé (ed.)

José Luis Fernández — *Las matemáticas y los riesgos financieros*

Paz Morillo i Carles Padró — *Protocols criptogràfics distribuïts, esquemes per compartir secrets*

Helmut Neunzert — *Mathematics as a key to key technologies*

Jesús Mosterín — *Modelos simples de un mundo complejo*

Joan Girbau — *Si no es pot representar cap terreny a escala, què fan els cartògrafs?*

Ismael Colomina — *Les eines matemàtiques de la geomàtica*

AULA DE CIÈNCIA
I CULTURA, 10

Les bases matemàtiques de la civilització tecnològica

Cicle en homenatge a Ferran Sunyer i Balaguer

(Del 17 de febrer al 24 de març de 1999)

Jaume Aguadé (ed.)

Catalogació en publicació de la
FUNDACIÓ CAIXA DE SABADELL

Les bases matemàtiques de la civilització tecnològica
(17 de febrer - 24 de març de 1999). - (Aula de ciència i cultura ; 10)

Conté: Presentació del cicle / Jaume Aguadé ; Las matemáticas y los riesgos financieros / José Luis Fernández ; Protocols criptogràfics distribuïts, esquemes per compartir secrets / Paz Morillo i Carles Padró ; Mathematics as a key to key technologies / Helmut Neunzert ; Modelos simples de un mundo complejo / Jesús Mosterín ; Si no es pot representar cap terreny a escala, què fan els cartògrafs? / Joan Girbau ; Les eines matemàtiques de la geomàtica / Ismael Colomina.

II. Fundació Caixa de Sabadell
1. Matemàtiques - Conferències
2. Fundació Caixa de Sabadell - Conferències
51(042)

Primera edició: novembre de 1999

© J.L. Fernández, P. Morillo, C. Padró, H. Neunzert, J. Mosterín, J. Girbau, I. Colomina, J. Aguadé.

Per a aquesta edició:
© Fundació Caixa de Sabadell, 1999
carrer d'En Font, 25, 08201 - Sabadell

Fotocomposició i producció: Addenda, s.c.c.l., Pau Claris, 92, Barcelona

Impressió: Winihard Gràfics, S.L., Polígon industrial, Av. del Prat s/n, Moià

És rigorosament prohibida, sense l'autorització escrita dels titulars del copyright, sota les sancions establertes a la llei, la reproducció total o parcial d'aquesta obra per qualsevol procediment, incloent-hi la reprografia i el tractament informàtic i la distribució d'exemplars mitjançant lloguer o préstecs públics.

Dipòsit legal: B-46353-1999

ISBN: 84-95166-10-0

Índex

7 Paraules d'obertura

15 Jaume Aguadé: *Presentació del cicle*

17 José Luis Fernández: *Las matemáticas y los riesgos financieros*

31 Paz Morillo i Carles Padró: *Protocols criptogràfics distribuïts, esquemes per compartir secrets*

41 Helmut Neunzert: *Mathematics as a key to key technologies*

59 Jesús Mosterín: *Modelos simples de un mundo complejo*

71 Joan Girbau: *Si no es pot representar cap terreny a escala, què fan els cartògrafs?*

83 Ismael Colomina: *Les eines matemàtiques de la geomàtica*

Paraules d'obertura

Hi intervenen:

Manuel Grau i Gasulla
President de la Fundació Caixa de Sabadell

Manuel Castellet
President de l'Institut d'Estudis Catalans

MANUEL GRAU I GASULLA

President de la Fundació Caixa de Sabadell

Com a president de la Fundació Caixa de Sabadell, em complau presentar avui el cicle: «Les bases matemàtiques de la civilització tecnològica.» Aquest cicle té una doble finalitat, d'una banda, retre homenatge a Ferran Sunyer i Balaguer, investigador i matemàtic nascut a l'Empordà i mort l'any 1967, el qual mereix el nostre reconeixement per la tasca que va dur a terme al llarg de la seva vida i, especialment, per les dificultats amb què es va trobar a causa de la seva malaltia, ja que era tetraplègic de naixement.

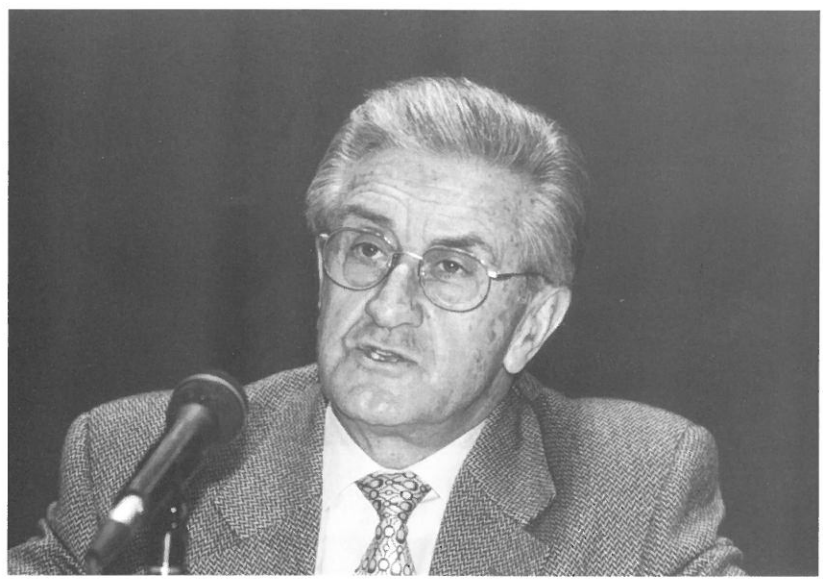
La segona finalitat que ens proposem és fer palesa la importància que tenen les matemàtiques en la vida moderna, és a dir, descobrir fins a quin punt una ciència exacta com aquesta fa possible que puguem utilitzar totes i cadascuna de les eines que ens posa a les mans la tecnologia d'avui dia.

Segurament, ningú no s'imagina que les matemàtiques, tan desconegudes per a molts –gairebé diria que per a tots–, estiguin tan a prop de nosaltres, però de fet ho estan, i això és precisament el que vol demostrar aquest curs que la Fundació Caixa de Sabadell ha organitzat conjuntament amb el Centre de Recerca Matemàtica i l'Institut d'Estudis Catalans, i amb el qual es proposa crear un espai de reflexió i diàleg obert per poder, alhora, enriquir el debat social, amb aportacions rigoroses, i abordar també temàtiques properes a la vida social i cultural que compartim.

Estem molt satisfets de comprovar que aquest cicle ha tingut una molt bona rebuda, no solament entre els estudiants, atès el valor acadèmic que té, sinó també entre altres persones, afins o no a les matemàtiques, que volen estar presents en aquestes conferències perquè consideren que poden ser interessants.

Per dur a terme aquest cicle comptarem durant algunes setmanes amb la participació de diverses personalitats, grans coneixedores del tema, que oferiran les aportacions des de diferents punts de vista. En total seran set ponents, alguns d'ells procedents del món de la docència, com és el cas de José Luis Fernández, de la Universitat Autònoma de Madrid, al qual vull donar una especial benvinguda. Ell pronunciarà, tot seguit, la primera conferència del cicle, que porta com a títol «Les matemàtiques i els riscos financers».

També voldria agrair a l'Institut d'Estudis Catalans, i especialment al seu president, el senyor Manuel Castellet, i al senyor Jaume Aguadé, coordinador del cicle, la seva tasca en l'organització del cicle i la seva presència avui a l'Auditori de la Caixa de Sabadell. A tots vostès els desitjo que gaudeixin del curs, que vegin complertes les expectatives que s'han creat en matricular-s'hi i que en treguin tant benefici personal com intel·lectual.



La Fundació Caixa de Sabadell no ha oblidat les matemàtiques entre els temes dels seus cicles.

MANUEL CASTELLET

President de l'Institut d'Estudis Catalans

Senyor Manuel Grau, president de la Fundació Caixa de Sabadell; senyor Fermí Pons, gerent de la Fundació; senyor José Luis Fernández; senyor Jaume Aguadé.

He de ser forçosament molt breu perquè qui ha de tenir la paraula avui, sobretot, ha de ser el conferenciant. Però voldria dir quatre mots d'agraïment a la Fundació Caixa de Sabadell per la bona disposició que mostra i ha mostrat en realitzar activitats conjuntes amb l'Institut d'Estudis Catalans.

No és aquesta la primera vegada que es produeix una bona col·laboració. Per exemple, la Fundació ja patrocina un dels premis que atorga l'Institut cada any i també va participar en unes jornades sobre Pompeu Fabra que van tenir lloc el novembre a Barcelona. Avui mateix hem estat a Palma presentant un cicle de conferències sobre Ramon Llull, organitzat conjuntament per la Fundació Caixa de Sabadell, l'Obra social i cultural Sa Nostra i l'Institut d'Estudis Catalans, un curs que es farà simultàniament a Palma i a Barcelona. Hi ha moltes més activitats que ja estan en camí.

Quan el gerent de la Fundació, el senyor Fermí Pons, em va suggerir –perquè va ser una idea d'ell– de fer un cicle de conferències sobre matemàtiques amb l'objectiu de treure una mica la por escènica que hi ha vers les matemàtiques al món– ens va semblar una excel·lent idea i ens vam posar immediatament a treballar-hi amb en Jaume Aguadé. El resultat ha estat una programació de conferències que seran extraordinàriament atractives tant pels temes com per les persones que les imparteixen.

Però, a més a més, jo vull destacar la sensibilitat de la Fundació de voler dedicar aquest cicle de conferències a Ferran Sunyer i Balaguer.

Ferran Sunyer i Balaguer fou un matemàtic –nascut a l'Empordà prop de Figueres, a Vall Déu, si mal no recordo– que no va anar mai a escola perquè era tetraplègic; no va fer l'ensenyament primari, ni el batxillerat i tampoc no va estudiar cap llicenciatura, però va ser educat amb molta cura pels seus pares, especialment per la seva mare. Tanta cura va rebre que això li va permetre assolir una educació gairebé normal per a una persona que no articulava bé les paraules, que no podia escriure i que amb prou feines podia passar la pàgina d'un llibre.

Ferran Sunyer tenia una intel·ligència fora del normal. Solia passar les tardes amb un cosí que estudiava enginyeria industrial, i així li van caure a les mans treballs de matemàtiques; es va adonar que en aquells llibres hi havia una cosa que no funcionava. Va escriure a Jacques Hadamard, i d'aquesta manera es va iniciar una



El president de l'Institut d'Estudis Catalans va glossar la figura de Ferran Sunyer i Balaguer.

correspondència molt rica entre matemàtics, francesos sobretot, i Ferran Sunyer, que tenia una visió extraordinària del que podien ser les matemàtiques, encara que no les escrivia. Primer l'ajudava a escriure la seva mare, i després unes cosines que es van ocupar d'ell.

Els anys quaranta, la comunitat local va descobrir el fet i va veure que s'havia de fer alguna cosa a fi que Sunyer pogués treballar. No podia fer de professor, i per tant es va provar que aconseguís una plaça d'investigador al Consell Superior d'Investigacions Científiques. Els anys quaranta i cinquanta la situació del Consell no era pas bona sinó molt polititzada en una direcció; mentre que Ferran Sunyer tenia idees molt clares en un altre sentit, la qual cosa li va comportar dificultats, però va mantenir una relació molt intensa amb la comunitat científica, i això va fer que pogués anar publicant articles, alguns dels quals premiats per l'Institut d'Estudis Catalans de l'època. Tot plegat va fer que aconseguís un contracte d'investigació amb la marina americana del qual va viure una bona temporada. Ferran Sunyer era una persona molt agradable –no físicament, sinó pel que fa a la mentalitat– i científicament activa, que assistia amb regularitat a congressos internacionals. Potser ha estat un dels investigadors més importants d'aquest període. Em plau molt que la Fundació Caixa de Sabadell li hagi volgut dedicar aquest cicle.

Fa uns anys, la família del matemàtic va crear una fundació, la Fundació Ferran Sunyer Balaguer, que dota cada any amb un premi internacional de matemàtiques una monografia que posi al dia els avenços en alguna àrea en què s'hagi treballat recentment. Ja anem per la setena edició del premi, un premi amb gran ressò internacional; a part del premi en metàl·lic, l'obra es publica en la sèrie «Progress in Mathematics» de l'editorial Birkhäuser.

M'he permès allargar-me parlant de la figura de Ferran Sunyer i Balaguer. En Jaume Aguadé ens orientarà ara sobre el cicle en general i sobre el ponent d'avui.



L'Auditori de la Caixa de Sabadell en un moment del cicle.

Les bases matemàtiques de la civilització tecnològica

– Conferències –

Presentació del cicle	Jaume Aguadé
Las matemáticas y los riesgos financieros	José Luis Fernández
Protocols criptogràfics distribuïts, esquemes per compartir secrets	Paz Morillo i Carles Padró
Mathematics as a key to key technologies	Helmut Neunzert
Modelos simples de un mundo complejo	Jesús Mosterín
Si no es pot representar cap terreny a escala, què fan els cartògrafs?	Joan Girbau
Les eines matemàtiques de la geomàtica	Ismael Colomina

JAUME AGUADÉ
Presentació del cicle

El 1917, el matemàtic Johann Radon publica un teorema que demostra que una funció de dues variables es pot recuperar a partir de les seves integrals al llarg de les rectes del pla. L'any 1972 es patenta el primer escàner que permet, per exemple, obtenir imatges de l'interior del cervell, i que té per fonament teòric el teorema de Radon. Situacions d'aquesta mena s'estan produint constantment.

Fa un parell d'anys, Peter Shor (que va rebre el Premi Nevanlinna al Congrés Internacional de Matemàtiques del 1998), publica un teorema que demostra que existeix un algorisme d'eficiència polinòmica per a la factorització d'un nombre en factors primers. Encara que no ho sembli, és un resultat crucial! Ara mateix, tota la seguretat del nostre sistema (de les transaccions financeres a l'aparell bèl·lic) depèn de la conjectura –no demostrada matemàticament– que un algorisme d'aquesta mena no pot existir. Shor n'ha trobat un, però aquest algorisme no funciona en cap dels ordinadors actuals, si bé es podria implementar en un ordinador quàntic, una màquina –conjectural, ara per ara– que podria existir abans de vint anys. Els països més desenvolupats estan invertint pressupostos de deu xifres en projectes relacionats amb el disseny d'aquestes màquines hipotètiques.

Hi ha massa poca gent que s'adoni que l'alta tecnologia de què tant ens enorgullim és, essencialment, una tecnologia matemàtica. Aquestes no són pas paraules d'un matemàtic que assaja de vendre el seu producte de la millor manera possible, sinó que pertanyen al qui fou fins fa poc director de recerca i desenvolupament de la petroliera Exxon, una companyia que té un pressupost anual de recerca de 500 milions de dòlars. No ens hauria de sorprendre gens que fos precisament la gent que està més en contacte amb el *món real* la que s'adonés amb més facilitat de fins a quin punt la nostra civilització tecnològica –com també el nostre model de pensament i de cultura– té unes profundes bases matemàtiques.

Quan escoltem la música enregistrada en un CD, quan ens sotmetem a una ressonància magnètica o utilitzem un telèfon mòbil, la nostra atenció se centra en els sofisticats processos d'enginyeria que sospitem que es troben al darrere de tota aquesta tecnologia, però ens resulta difícil relacionar aquests avenços tecnològics amb avenços matemàtics crucials, que s'han produït recentment o que s'estan produint virtualment ara mateix. I si això succeeix fins i tot als matemàtics professionals, com podem aconseguir que l'extraordinària força emergent de la matemàtica es faci evident als ulls de tota la societat?

De tot això és del que volem parlar en aquest cicle. Volem posar de relleu com la matemàtica –pura, abstracta, mil·lenària– està en la base de la nostra civilització tecnològica. En la base dels grans salts tecnològics hi ha, cada cop més, avanços matemàtics significatius. La matemàtica és la gran arma secreta del nostre temps.

Mentre, paradoxalment i lamentablement, va perdent de mica en mica el paper preeminent que ha tingut, al llarg del temps, en la formació cultural de les persones, la matemàtica controla el funcionament del nostre món –de les finances a la diagnosi mèdica. Són els teoremes de la teoria de nombres –considerada durant segles com la més «inútil» de les ciències matemàtiques– els que, en darrera instància, garanteixen el funcionament d'una targeta de crèdit o la inviolabilitat de les transaccions comercials electròniques. El desenvolupament del càlcul infinitesimal al segle XVII va anar lligat a la matematització de la naturalesa i va conduir a una primera revolució tecnològica. Ara, novament, és la pràctica totalitat de la matemàtica la que està matematitzant, globalment, la nostra societat.

Em complau saber que en aquest auditori hi ha força estudiants de matemàtiques. Els volem dir que necessitem matemàtics multidisciplinaris que vulguin considerar problemes del món real –i que tinguin el coratge de fer-ho–, i també els volem dir que podran fer-ho sense renunciar a utilitzar matemàtica del nivell més alt possible. A aquests estudiants que, inevitablement, reben un ensenyament esmicolat en petits àtoms escassament relacionats entre ells, potser aquest seminari els ajudarà a copsar la unitat intrínseca del coneixement matemàtic i la il·limitada potència de l'eina que estem posant en les seves mans.

Las matemáticas y los riesgos financieros

José Luis Fernández Pérez

Se afirma con frecuencia que no hay pensamiento sin lenguaje, que no podemos pensar conceptos si no es a través de su expresión articulada mediante el lenguaje. En exacto paralelismo, se afirma también que, dentro del ámbito general del conocimiento, la ciencia se caracteriza por ser expresable en matemáticas: las matemáticas *son* el lenguaje de la ciencia. O mejor, en la frase clásica de Galileo:

> Sólo pueden leer el gran libro de la naturaleza aquellos que conocen el lenguaje en que está escrito. Y este lenguaje es la matemática.

Formidable poder de las matemáticas que le lleva a uno a admirarse, con Wigner [18], de

> The unreasonable effectiveness of mathematics in the natural sciences

al constatar la universalidad de las matemáticas como lenguaje de todas las ciencias y su asombroso rango de aplicabilidad.

En esta conferencia nos centraremos en el uso cada vez más demandado de las matemáticas en el *estudio y análisis de los riesgos financieros*.

Matemática e incertidumbre

La humanidad siempre se ha sentido a la vez impotente y fascinada por la incertidumbre inherente al futuro. La regularidad de las estaciones o la sucesión permanente del día y de la noche confortaban y daban seguridad a las civilizaciones primitivas, mientras que la (entonces) imprevisible aparición de un cometa era universalmente considerada como presagio de mal agüero.

Supuso toda una revolución para la humanidad, un reto a la divinidad [5], atreverse a buscar reglas en el comportamiento de fenómenos aleatorios, casi una contradicción. Para los griegos, *aleatorio* significaba «ausencia absoluta de orden». No fue hasta el siglo XVII en que Huygens, Pascal, Fermat y De Moivre crean el cálculo de probabilidades, que proporciona un marco conceptual y las herramientas pertinentes, con las que analizar los fenómenos aleatorios [7].

Matemáticas y comercio

Es bien posible que los primeros cálculos matemáticos nacieran de necesidades de comercio, de trueque o intercambio. Esta relación entre matemáticas y comercio se ha mantenido siempre a lo largo de la historia, y como pequeña muestra basten estos ejemplos:

- la agrimensura en Mesopotamia y en el valle del Nilo, o el reparto conforme a la tradición coránica de las herencias en la España musulmana, favorecieron desarrollos respectivos de la trigonometría y del álgebra,
- el insigne número *e* vio la luz, entre otras razones, cuando se hizo preciso calcular intereses continuos de capital.

Matemática actuarial

La matemática, una vez que estuvo dotada del cálculo de probabilidades, permitió el desarrollo de la ciencias del seguro, de la llamada *matemática actuarial* [10].

Este uso de las matemáticas ya trata de hacer frente a la incertidumbre que suponen los riesgos asegurados. En su forma más simple, una compañía aseguradora opera de la manera siguiente: calcula el riesgo total que está asegurando en un cierto período de tiempo en función de ciertos datos estadísticos y distribuye ese montante total, confiando en las leyes de los grandes números, entre todos los asegurados.

Matemáticas y finanzas

Daniel Bernoulli, en su análisis de la paradoja de San Petersburgo [4], introduce las funciones de utilidad y se plantea preguntas tales como: ¿cómo medir la incertidumbre?, ¿cómo enfrentarse racionalmente a decisiones sobre eventos inciertos? Desde ahí parte un hilo conductor que, pasando por la teoría de juegos

creada por Von Neumann y Morgenstern, llega a la teoría de carteras de Markowitz y Sharpe, Premios Nobel de Economía.

En su forma más simple, la cuestión es la siguiente. Tenemos dos activos, por ejemplo, acciones del BSCH y bonos del Tesoro, cuyos valores dentro de un cierto período de tiempo son inciertos. Las rentabilidades que pudieran tener esos activos en ese período de tiempo son, pues, variables aleatorias. Análisis previos nos dicen qué rentabilidad media podemos esperar de cada uno, y cuál es su nivel de riesgo. Supongamos que ya hemos determinado el montante total a invertir y que hemos fijado la utilidad que la riqueza tiene para nosotros: ¿cómo debemos repartir de manera óptima nuestra inversión entre esos activos?

Riesgo financiero

En finanzas, el término *riesgo* se emplea de manera precisa para referirse a todo tipo de incertidumbre sobre el precio o valor de un determinado activo o posición.

Si supiéramos con certeza que un determinado activo va a perder un 10 % de su valor, no habría riesgo.

Sin embargo, si de un activo supiéramos que va a subir entre un 10 % y un 20 %, pero no supiéramos cuánto con precisión, ahí tendríamos riesgo.

Los riesgos son múltiples:

- Riesgo de divisas. Tenemos dólares que queremos usar para invertir en euros dentro de un mes: el tipo de cambio de dólares a euros es incierto.
- Riesgo de reinversión. ¿A qué tipo de interés podremos reinvertir los cupones que vamos a percibir por nuestra cartera de bonos del Tesoro?
- Riesgo de crédito. No hace mucho, Rusia decidió no cumplir con el pago de los cupones de su deuda.
- [...]

¿Cómo medimos la incertidumbre? En una situación estocástica, es decir, si entendemos los valores, flujos o niveles como variables aleatorias, el *riesgo* se mide con la *varianza*. Cuanto mayor es la *varianza*, mayor es la dispersión de los posibles valores, y, por tanto, hay mayor incertidumbre.

Instrumentos financieros diseñados para controlar/transferir riesgos

El mundo financiero ha creado instrumentos que tienen como misión transferir riesgos financieros entre dos partes. Una parte elimina su incertidumbre transfiriéndosela a otra parte. Vaya por delante que eliminar la incertidumbre puede suponer tanto perder como ganar, en cuanto a valor.

Veamos dos ejemplos básicos: los futuros y las opciones.

- *Futuros sobre cítricos*

  Como productor de naranjas, contrato *hoy* la venta dentro de tres meses (al término de la campaña) de una parte de mi producción a un determinado precio K por tonelada a un comercializador. Esto es un contrato de futuro (a plazo). Ambas partes del contrato están obligadas a ejecutar la transacción. Esto me permite programar mis finanzas: compras, deudas...

  Si dentro de tres meses el precio por tonelada resulta ser P, y $P > K$, he perdido dinero: me hubiera compensado no entrar en ese contrato de futuro. Por el contrario, si $P < K$ habré ganado.

  En suma, cuando vence el contrato habré incurrido en una ganancia neta de:

$$K - P$$

  pero he eliminado la incertidumbre del precio.

  Quedan riesgos tales como posibles incumplimientos de contrato, por ejemplo. Esa es la razón de que exista un mercado organizado.

- *Opciones sobre acciones*

  Una opción *put* (opción de venta) me da el derecho (pero no la obligación) a vender a un determinado precio de ejercicio K, que en la jerga financiera se conoce como *strike*, una acción, por ejemplo, de BSCH, y en un determinado tiempo futuro, digamos dentro de 3 meses.

  ¿Cuál es el uso de esa opción *put*? He aquí un ejemplo:
  — Dentro de 3 meses necesitaré liquidez,
  — para ello venderé mi cartera de BSCH esperando que el precio esté por encima de K,
  — podría vender ahora,
  — pero creo que su cotización va a subir y no quiero vender ahora,
  — pero BSCH está muy volátil ¿y si su cotización baja dentro de 3 meses?

  La compra de una opción *put* me garantiza un precio mínimo al que podré vender. He comprado un seguro. ¿A quién? A un banco o a un particular o a una contrapartida en un mercado organizado, como MEFF.

En un contrato de futuro una parte ha de vender al precio K y la otra ha de comprar a ese precio. En una opción *put* una parte tiene el derecho a vender al precio K (pero no la obligación: no venderá a ese precio si no le interesa) y la otra tiene la obligación de comprar (si la otra parte lo exige) a ese precio K (pero no tiene el derecho).

Las opciones

El derecho que he adquirido al comprar una opción *put* no es gratuito, he de pagar una prima.

¿Cuánto debe ser esa prima? Este es el problema de *valoración*.

El banco que me ha vendido esa *put* asume mis riesgos, y debe ser capaz de cubrirlos con la prima percibida: ¿qué ha de hacer para «cubrir» el compromiso que ha adquirido? Esta es la cuestión de *cobertura*.

Como veremos, valorar y cubrir son esencialmente equivalentes.

Podemos codificar el pago que supone esa opción *put* a vencimiento como:

máximo entre $(K - P)$ y 0

siendo P el precio de dentro de 3 meses, que por supuesto es desconocido hoy.

- A los futuros y las opciones se los conoce como instrumentos *derivados*. El pago

$$\max\{(K - P), 0\}$$

de una *put* y por tanto la prima correspondiente como veremos se *deriva* del nivel de precios de otro activo (las acciones).

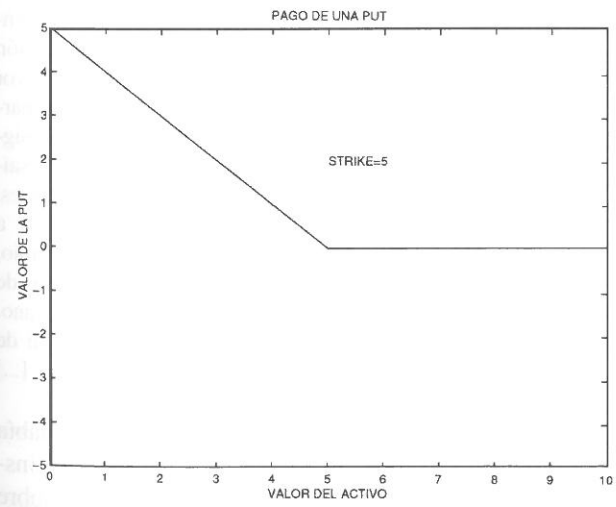


Figura 1: Fujos de una put.

- Una opción *call* es una opción de compra. Obsérvese que estas opciones de venta y de compra pueden ser compradas y vendidas.
- Las opciones son instrumentos más sutiles y delicados de lo que pudieran aparecer a primera vista:

— Apalancamiento. La prima de una *put* es una fracción del precio de la acción sobre la que se basa. A vencimiento, el comprador puede perder el valor de esa prima pero puede llegar a ganar hasta el precio de ejercicio. Una inversión pequeña supone una ganancia potencial grande.

— Los pagos tales como

máximo entre $(K - P)$ y 0

no son lineales en P, es decir, las variaciones en esos pagos no son proporcionales a las variaciones en el valor de P. Esto dificulta su valoración y cobertura y hace que sean instrumentos delicados de manejar.

— De vez en cuando aparecen en la prensa ejemplos de desastres financieros que han sido consecuencia de gestión especulativa (combinada, quizá, con avaricia y arrogancia) de instrumentos derivados: la banca Barings y el fondo de riesgo Long Term Capital son ejemplos recientes que han recibido mucha publicidad.

- *Opciones por todas partes*

Pero no creamos que los derivados son instrumentos esotéricos que un reducido número de inversores manejan. Nada más lejos de la realidad. Los derivados están por todas partes.

— Fondos de inversión en los que nos garantizan, por una inversión inicial de 100 hoy, que dentro de un año la inversión se ha revalorizado en al menos un 80 % de lo que se ha revalorizado un índice de bolsa como el Ibex.

OPCIONES IBEX 35 PLUS

	Precio Ejerc.	Primas Dem.	Cierre Ofer.	Ultim. Cruz.	Max Sesión	Min. Sesión	Volat. Cierre	Delta Cierre	Volum. Contrat.	Posic. Abierta
CALL ▫ ♦ 5.3 000,1										
Nov 1999	9.300	195	217	215	330	215	22	51	6	105
Dic 1999	9.300	315	331	0	0	0	23	52	0	0
Ene 2000	9.300	411	429	427	530	427	23	52	9	3
Mar 2000	9.150	548	567	0	0	0	24	52	0	0
Jun 2000	9.400	725	743	0	0	0	25	53	0	4
Sep 2000	9.400	882	900	0	0	0	25	54	0	0
PUT ▫ ♦ 5.3 000,2										
Nov 1999	9.300	204	222	219	220	140	23	49	25	286
Dic 1999	9.300	302	320	0	0	0	23	47	0	360
Ene 2000	9.300	402	420	405	405	347	23	47	9	3
Mar 2000	9.350	553	571	0	0	0	24	46	0	0
Jun 2000	9.400	724	742	0	0	0	25	45	0	0
Sep 2000	9.400	648	966	0	0	0	25	43	0	0
Total Contratos		Opciones	2.474,00		Futuros	10.635,00		Total		13.109,00
Posiciones Abiertas		Opciones	32.186,00		Futuros	46.374,00		Total		138.560,00
Ratio CALL / PUT		2,71								

Figura 2: Cotizaciones de derivados.

— Amortización anticipada de un crédito. Por el derecho a anticipar la amortización de un crédito nos imputan una prima. Se trata de una opción sobre el montante de ese crédito.

— Préstamos con techos. Son préstamos en los que pagamos un tipo de interés variable, como el tipo Euribor a 1 año; pero si este tipo Euribor pasa de, por ejemplo, 6 % entonces pagamos só-

lo un 6 %. Esa garantía adicional es una opción, que cuesta una prima.

— En un apéndice describimos un ejemplo de un contrato de opción que se diseñó para cubrir los riesgos asociados al pago de las primas que el Comité Olímpico Español había acordado con nuestros representantes en la Olimpiada de Atlanta.

La variedad de opciones es un mundo enorme que no para de crecer. Los derivados se diseñan a medida del riesgo específico que se quiere transferir.

Un poco de historia

Asistimos en estos tiempos, como hemos indicado más arriba, a un uso muy extendido de este tipo de instrumentos financieros. Pero los futuros y las opciones tienen una larga, rica e interesante historia cuyos hitos más señalados pasamos a comentar a continuación.

ARISTÓTELES. En el libro I de su *Política* aparece una definición de opción como instrumento financiero.

TALES. Es tradición que el insigne matemático y filósofo Tales de Mileto se hizo rico negociando con opciones. Tales previó, no sabemos exactamente cómo, que la cosecha de aceituna iba a ser particularmente buena en cantidad. Decidió especular sobre esta idea y compró opciones de uso de molinos de aceite por pequeñas cantidades de dinero (apalancamiento). Tras la cosecha, esos derechos de uso adquirieron, como había especulado Tales, un valor extraordinario.

JOSÉ DE LA VEGA. En el siglo XVII había en Holanda un mercado de opciones muy activo y sofisticado (en los dos sentidos actuales de este calificativo) que llegaba hasta a incluir opciones donde el subyacente era a su vez otra opción.

Hubo incluso un mercado de opciones sobre el jaspeado de los tulipanes que se importaban de Turquía. Este mercado llegó a un desfase tal entre valor y precio que cuando se desinfló como una burbuja supuso una profunda crisis financiera.

El español José de la Vega publicó en Amsterdam en 1688 un jugoso libro describiendo el negocio de acciones y de opciones en esa plaza con el conspicuo título de *Confusión de confusiones* [17].

> Dad OPSIES, y sabréis el límite de la pérdida, pudiendo exceder la ganancia a la fantasía, y ser mayor aún el aumento que la esperanza. [...]
>
> Tocante a los OPSIES, son unos premios o cantidades que se dan para asegurar las partidas, o conquistar los avances; los cuales sirven de velas para navegar felices en las bonanzas, y de áncoras para navegar seguros en las tormentas. [...]

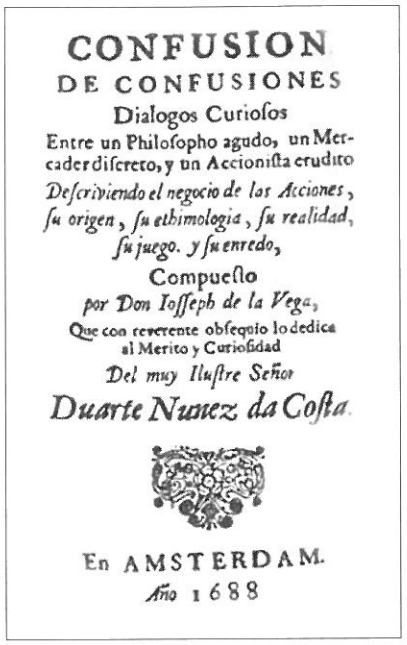

CONFUSION
DE CONFUSIONES
Dialogos Curioſos
Entre un Philoſopho agudo, un Mercader diſcreto, y un Accioniſta erudito
Deſcribiendo el negocio de las Acciones, ſu origen, ſu ethimologia, ſu realidad, ſu juego. y ſu enredo,
Compueſto
por Don Ioſſeph de la Vega,
Que con reverente obſequio lo dedica al Merito y Curioſidad
Del muy Iluſtre Señor
Duarte Nunez da Coſta.
En AMSTERDAM.
Año 1688

Figura 3: Confusión de confusiones.

> Llamáronle los flamencos OPSIE, derivado del verbo latino *Optio Optionis*, que significa elección por quedar a elección del que lo da, el poder pedir, o entregar la partida, al que lo recibe: [...] pues desea el que se desembolsa el premio, elegir lo que más convenga, y en falta siempre puede dejar de elegir lo que desea. [...]
>
> Están las acciones al presente en precio de 580, paréceme que por el gran retorno que se espera de la India, aumento de la Compañía, reputación de los géneros, repartición que se promete, y paz de la Europa, subirán a mucho mayor del que logran. No me delibero sin embargo, a comprar partidas efectivas, porque temo que si me faltaren estos designios, podrá alcanzarme un despeño, o sucederme un desaire: llégome pues a los que me dicen que toman esos Opsies, propóngoles cuánto quieren por quedarme obligados a entregar cada partida a 600 hasta tal plazo, ajusto el premio, escríbolo luego en Banco, y sé que no puedo perder más de lo que desembolso, con que todo lo que suben de 600 gano, y lo que bajan, no me sirve de ansia para el juicio, ni de inquietud para la honra, ni de sobresalto para el sosiego. [...]

NUEVA YORK. A finales del siglo XVIII ya se había desarrollado un activo mercado de opciones sobre instrumentos financieros tales como acciones, y no sobre mercanderías o materias primas. Se negociaba privadamente, y no a través de una bolsa o cámara de compensación que garantizara los contratos. El carácter eminentemente especulativo y la dimensión de las inversiones hacía que con frecuencia no se respetaran los contratos, y esto hizo que en varias ocasiones se prohibiera la negociación de opciones.

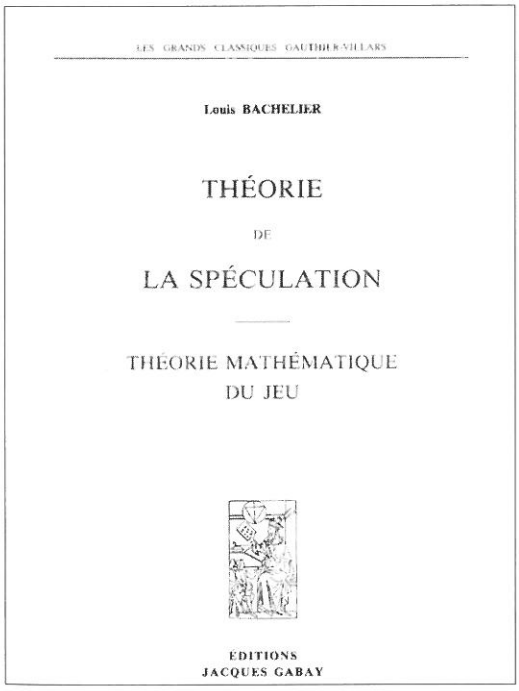
CHICAGO. En 1973 comienza a funcionar el primer mercado organizado de derivados financieros. Desde sus comienzos tuvo un éxito espectacular y su volumen de negocio no ha parado de crecer.

Por su parte, el mercado español de futuros y de opciones financieras, MEFF Renta fija, que tiene su sede

en Barcelona, y MEFF Renta variable, que la tiene en Madrid, comenzó a operar a principios de los años noventa.

BACHELIER. Louis Bachelier defendió su tesis doctoral *Théorie de la spéculation* [1] en el año 1900. El supervisor de su trabajo doctoral fue el insigne Henri Poincaré, quien no gustaba del tema y no supo percibir la innovación conceptual que las ideas expuestas por Bachelier suponían. La importancia de esta tesis doctoral no fue apreciada –quizá por culpa de la poca estima que le tuvo Poincaré– hasta los años sesenta, en que fue rescatada del olvido por Benoit Mandelbrojt, el inventor de los fractales.

La tesis de Bachelier postula que: *los precios de los activos siguen un movimiento browniano con deriva.* Más adelante explicamos qué es el movimiento browniano. Pero baste por ahora decir que Bachelier *¡Inventa/descubre el movimiento browniano!* Bachelier, en su tesis, da ya ciertas fórmulas para valorar opciones de compra y de venta sobre acciones.


LES GRANDS CLASSIQUES GAUTHIER-VILLARS

Louis BACHELIER

THÉORIE

DE

LA SPÉCULATION

THÉORIE MATHÉMATIQUE
DU JEU

ÉDITIONS
JACQUES GABAY

Figura 4: Reedición de la tesis doctoral de Louis Bachelier.

SAMUELSON. Premio Nobel de Economía. Aporta un matiz fundamental a las tesis de Bachelier y postula que: *Los rendimientos, y no los precios, de los activos siguen un movimiento browniano con deriva.* Su modelo no identifica qué deriva.

En colaboración con H. McKean valora opciones americanas perpetuas, estableciendo la conexión con la ecuación del calor.

BLACK († 1995), SCHOLES, MERTON (Premio Nobel 1997). Black y Scholes, en colaboración [3], y Merton [14], a comienzos de los años setenta completan el marco conceptual vigente que permite analizar en profundidad los complejos mercados financieros modernos al situar en un papel central el concepto de ausencia de oportunidad de arbitraje. Muy crípticamente, introducen la replicación dinámica, identifican la deriva en el modelo de Samuelson y sitúan a la volatilidad en el papel central que ahora tiene. El artículo [3] figura entre los más citados de toda la literatura científica de todos los tiempos.

La fórmula de Black-Scholes, que comentaremos más adelante, traduce volatilidad en precios. Por estos trabajos les fue concedido a Scholes y a Merton el Premio Nobel de Economía 1997; Black había fallecido dos años antes.

Modelización matemática de mercados financieros

La modelización matemática de los mercados financieros combina fundamentalmente dos herramientas: el *movimiento browniano* y la hipótesis de *ausencia de oportunidad de arbitraje.*

Esta modelización se basa en la premisa de que los mercados financieros se comportan como un sistema físico. Los operadores e inversores se conceptúan como pequeñísimas unidades que usan la información común disponible por igual a todos ellos para tomar decisiones de inversión racionales que tiran de los precios, o mejor, de las rentabilidades de los precios. Visto de esta manera, el análisis del mercado financiero no es sino mecánica estadística. Todo el proceso de formación de precios está supervisado por un principio de equilibrio: la hipótesis de ausencia de oportunidad de arbitraje, que actúa de manera eficiente.

Enseguida veremos cómo se relacionan entre sí estas ideas.

Este marco conceptual ha tenido un éxito innegable al ayudar a clarificar el funcionamento de ciertos aspectos de los mercados financieros. Y es, hoy por hoy, una herramienta estándar de esos mercados.

Pero conviene advertir ahora que se trata de un modelo físico-matemático que intenta reflejar un comportamiento humano. Todas las hipótesis implícitas en la descripción de más arriba, a saber, eficiencia, transparencia de la información, ausencia de oportunidad de arbitraje, racionalidad, similitud y tamaño de los operadores e inversores, son y deben ser siempre tenidas en cuenta y revisadas. Digamos que funcionan razonablemente bien en situaciones *normales*.

Dos referencias interesantes que ofrecen perspectivas sobre las carencias de esta modelización son: [16] y [6].

Por supuesto que los investigadores de la modelización físico-matemática de las finanzas son conscientes de estas carencias y los modernos estudios van encaminados a mejorar estos aspectos.

El movimiento browniano

Los precios y rentabilidades, y, en general, las variables financieras son cantidades aleatorias que varían con el tiempo. El marco matemático para analizar esta combinación de azar y tiempo es la teoría de procesos estocásticos. Los procesos de difusión, en particular, y fundamentalmente, el movimiento browniano, y el aparato para calcular con ellos, el cálculo estocástico, constituyen sus herramientas fundamentales. De hecho, el creador del cálculo estocástico, el matemático japonés Itô, inició sus investigaciones para construir un marco conceptual en el que desarrollar las propuestas de Bachelier sobre el comportamiento de los mercados financieros.

¿Qué es el movimiento browniano? Para comenzar, una cita:

> El movimiento browniano, una joya de la física moderna, es un descubrimiento botánico cuya teoría matemática tiene su origen en las ciencias sociales.

En la literatura científica, el movimiento browniano aparece por primera vez, y de ahí recibe su nombre, en trabajos del botanista Brown. El trabajo de Einstein sobre el movimiento browniano se encuentra citado entre los méritos por los que le concedieron el Premio Nobel en 1905. Su primera descripción matemática, sin embargo, es debida, como hemos comentado más arriba, a Bachelier. Aunque la formulación moderna hubo de esperar a Wiener, en los años veinte.

Pero ya Tito Lucrecio en su *De rerum natura* del siglo primero antes de Cristo nos da una descripción literaria de fenómenos físicos que pueden ser descritos como de movimiento browniano:

> Observe el fenómeno que se produce cuando un rayo de luz penetra en una casa y atraviesa una habitación a oscuras. Verá entonces una multitud de partículas que se desplazan en todas direcciones por el espacio iluminado y que, como si de una batalla incesante se tratara, se abalanzan las unas sobre las otras, oleada tras oleada, en una rápida sucesión de combates singulares. Uno puede imaginar cuál es el destino de átomos perpetuamente empujados unos sobre otros en el vacío ilimitado... Choques repetidos e invisibles de los átomos impulsan a las partículas, las cuales asestan golpes poderosos en su derredor. De manera que aunque la agitación proviene de los átomos, se manifiesta ante nuestros sentidos por intermedio de cuerpos movidos como por colisiones invisibles.

La idea básica del movimiento browniano es la de *paseo aleatorio*:

En cada momento lanzamos una moneda para decidir si aumentamos o decrecemos en una unidad. Las posiciones sucesivas que así vamos obteniendo describen una de las posibles trayectorias del paseo aleatorio. Imaginemos que nos situamos en una recta y que damos pasos hacia delante o hacia atrás según salga cara o cruz. El recorrido que realizamos es aleatorio: es un paseo aleatorio.

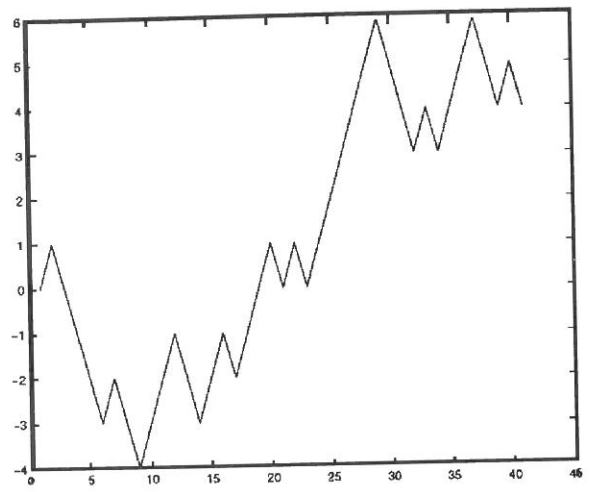


Figura 5: Camino aleatorio.

Matemáticamente: nuestra posición en tiempo n viene dada por

$$S_n = S_{n-1} + X_n, \text{ partiendo de } S_0 = \text{posición inicial}$$

y donde las X_n son variables aleatorias independientes y con distribución común de probabilidad dada por:

$$X_n = \begin{cases} +1, \text{ con probabilidad } \dfrac{1}{2} \\ -1, \text{ con probabilidad } \dfrac{1}{2} \end{cases}$$

Obsérvese que, a priori, no sabemos qué valor concreto va a tomar S_n, es decir, en qué posición nos vamos a encontrar en tiempo n. Pero sí sabemos qué valores son posibles y podemos calcular la probabilidad que tiene cada uno de ellos de ser alcanzados.

Pero necesitamos un proceso matemático de la misma naturaleza pero donde las transiciones tengan lugar en pequeños intervalos de tiempo, e idealmente de manera continua en el tiempo. Denotemos ese pequeño intervalo de tiempo por Δt. El nivel o posición valor en tiempo $n\Delta t$ se denotará por $S_{n\Delta t}$. La escala espacial también tiene que ser reajustada pero será necesario tomar múltiplos de $\sqrt{\Delta t}$ y no de Δt. es decir, proponemos considerar las transiciones

$$S_{n\Delta t} = S_{(n-1)\Delta t} + \sqrt{\Delta t} X_n$$

Pero, ¿por qué $\sqrt{\Delta t}$? Pongamos que en una unidad de tiempo, digamos en 1 año, hay N periodos de tiem-

po de tamaño Δt, $N\Delta t = 1$. Así que $S_{N\Delta t}$ representa los distintos niveles de dentro de 1 año, y éstos se alcanzan dando N saltos:

$$S_{n\Delta t} = \sum_{k=1}^{N} \left[S_{k\Delta t} - S_{(k-1)\Delta t} \right] + S_0 \,,$$

pero como esos saltos son independientes (e idénticamente distribuidos), la varianza de $S_{N\Delta t}$ es

$$N \cdot \text{varianza de un salto} = N \cdot \Delta t = 1$$

Es decir, sea cual sea el pequeño intervalo de tiempo Δt utilizado, la varianza de los valores de dentro de un año es fija, e igual a 1.

Esto es importante puesto que querremos hacer que Δt tienda a cero. Si hubiéramos tomado una escala espacial con otra potencia de Δt distinta de 1/2, la varianza se aproximaría a infinito o a cero (según la potencia). Podríamos, eso sí, tomar un múltiplo de Δt, y eso será útil enseguida.

Si tomamos Δt muy pequeño, las evoluciones se corresponderán aproximadamente con dar saltos de manera continua en el tiempo. Esa evolución límite es el *movimiento browniano*. Es decir,

$$\Delta t \to 0$$

$$S_{n\Delta t} = S_{(n-1)\Delta t} + \sqrt{\Delta t} X_n$$

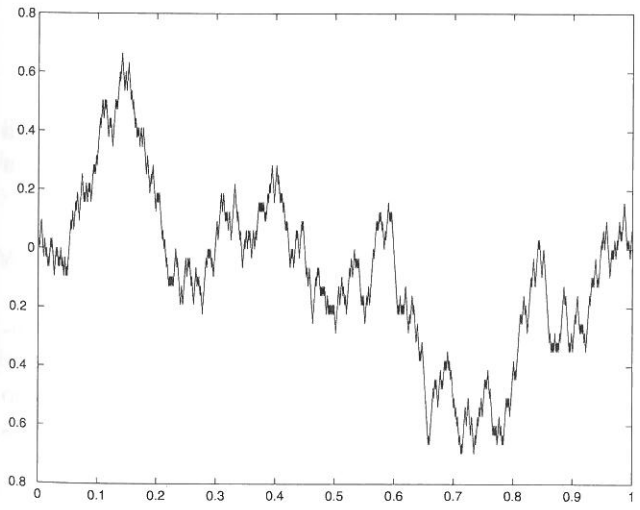


Figura 6: Movimiento browniano.

Podemos además imponer una *deriva* μ y un *volatilidad* σ a la evolución, es decir, considerar

$$\Delta t \to 0$$

$$S_{n\Delta t} = S_{(n-1)\Delta t} + \mu\Delta t + \sigma\sqrt{\Delta t} X_n$$

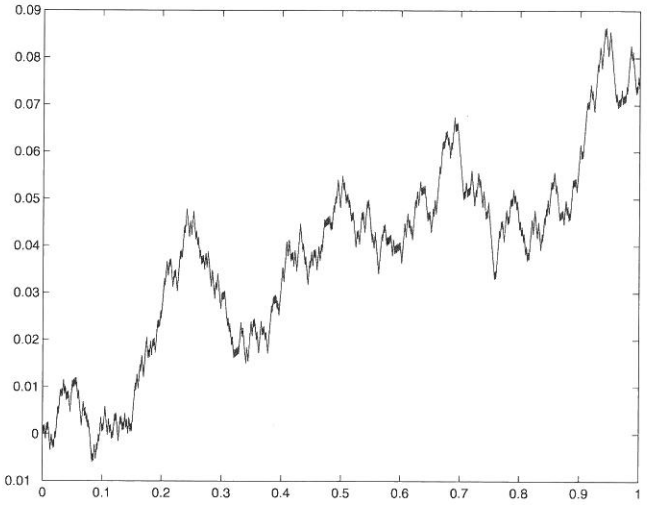


Figura 7: Movimiento browniano con $\mu = 0{,}1$ y σ

Finalmente, introducimos la versión multiplicativa, que no aditiva. El *movimiento browniano geométrico* con deriva μ y volatilidad σ viene dado por:

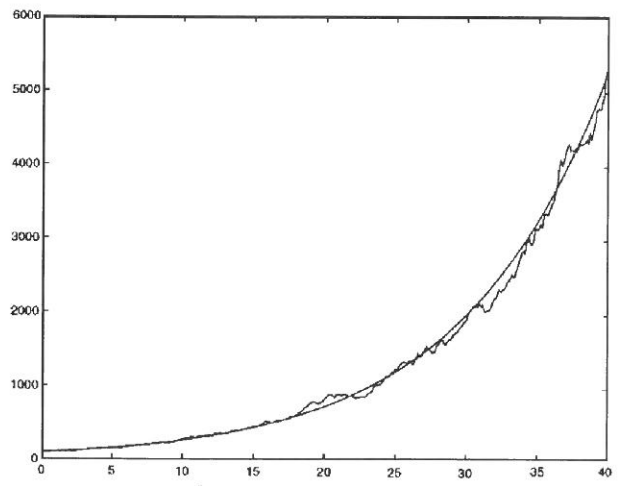


Figura 8: Movimiento browniano geométrico.

$\Delta t \to 0$

$$G_{n\Delta t} = G_{(n-1)\Delta t}\left(e^{\mu\Delta t + \sigma\sqrt{\Delta t}X_n}\right)$$

El movimiento browniano geométrico es el proceso que usaremos para describir las sucesivas posiciones de la cotización de una acción.

Observemos que en el movimiento browniano geométrico, la rentabilidad en un intervalo de tiempo Δt,

$$\text{rentabilidad en el intervalo } [(n-1)\,\Delta t,\, n\Delta t] = \log \frac{G_{n\Delta t}}{G_{(n-1)\Delta t}},$$

es una variable normal con media $\mu\Delta t$ y desviación típica $\sigma\sqrt{\Delta t}$, y que las rentabilidades en sucesivos períodos de tiempo son independientes unas de otras.

Cuanto más grande es σ mayor es la amplitud de los saltos que da la rentabilidad, es decir, mayor incertidumbre tenemos sobre los movimientos del activo. Mercados volátiles son aquellos en los que se observan grandes oscilaciones, con gran incertidumbre; revueltos, en suma, y ya se sabe que a «río revuelto, ganancia de pescadores».

Es natural que en un mercado con gran volatilidad, opciones *put* , que podemos entender como seguros, tengan primas altas.

Descripciones detalladas del cálculo estocástico y de su aplicación a los mercados financieros pueden verse en las referencias [15], [9] y [8].

Ausencia de oportunidad de arbitraje

El principio de fuerzas y equilibrio que determinan la formación de precios es la *ausencia de oportunidad de arbitraje*. La siguiente cita de Krugman [12], p. 72, explica en lenguaje llano el significado de esta hipótesis.

> En su base, la teoría económica descansa en dos observaciones:
>
> - las oportunidades de beneficio obvias raramente se dejan sin explotar,
> - y las cosas cuadran.
>
> O, tal como lo digo a veces,
>
> - un billete de veinte dólares no permanece en el suelo de una calle transitada durante mucho tiempo,
> - y cada venta es también una compra.
>
> Cuando uno se propone construir un modelo matemático formal, estos principios tan toscos se convierten en las ideas más exactas de maximización (de algo) y de equilibrio (de alguna forma).
>
> Sin embargo, es aconsejable no olvidar su versión más imprecisa, por dos razones opuestas: para acordarnos de no tomar ninguna formalización matemática demasiado en serio, y para acordarnos de que los principios básicos de la economía no son del todo estúpidos o poco razonables.

En términos precisos, una oportunidad de arbitraje es una estrategia de inversión que con una inversión inicial nula garantiza una ganancia segura.

EJEMPLO. Datos de mercado:
- Un contrato de futuro con *strike* K, a tres meses,
- Sobre una acción que hoy vale S
- El tipo de interés simple a 3 meses es R
- Supongamos que $K > S\,(1 + R)$

Hay una oportunidad de arbitraje:
- Entrar en un contrato de futuro (como vendedor) (no cuesta dinero)
- Compramos una acción S
- con dinero prestado

La inversión inicial es 0. Pero dentro de 3 meses, si la acción cotiza en un valor $\overline{S}$, la situación de nuestra cartera es la siguiente:

- el futuro nos da $K - \overline{S}$
- la acción vale $\overline{S}$
- y debemos $S\,(1 + R)$,

lo que da un total (seguro, independiente de cuanto sea $\overline{S}$) de

$$K - S\,(1 + R)$$

La hipótesis contraria, $K < S\,(1 + R)$, también da lugar a oportunidad de arbitraje. Este análisis fuerza a que, si no hay oportunidades de arbitraje:

$$K = S\,(1 + R)$$

es decir, nos permite determinar unos precios en términos de otros.

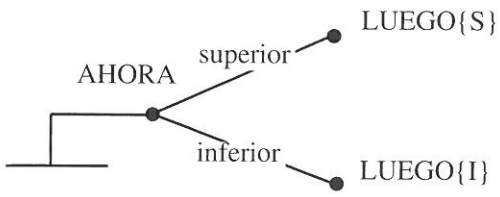
Las oportunidades de arbitraje existen, pero como afirma Krugman, duran sólo unos instantes. Podemos suponer que no las hay.

1 período, 2 estados, 3 activos

La interacción entre la descripción de la rentabilidad en términos del paseo aleatorio o del movimiento browniano y la hipótesis de ausencia de oportunidad de arbitraje puede ser descrita en términos sencillos de la siguiente manera.

Disponemos de *tres* activos *distintos*. Cada activo se identifica con su flujo, es decir, con lo que paga o vale.

Sólo analizamos *un* intervalo de tiempo (Δt). Y suponemos que el mundo sólo puede evolucionar del estado actual a *dos* estados futuros, que denotamos convencionalmente por s = *superior* e i = *inferior*.



Representemos los valores de los activos A, B y C en el futuro en cada uno de los estados s e i por A_s, A_i; B_s, B_i; C_s, C_i y resolvamos el sistema (con α y β como incógnitas).

$$\begin{cases} C_s = \alpha A_s + \beta B_s \\ C_i = \alpha A_i + \beta B_i \end{cases}$$

Las dos carteras siguientes:

1) 1 unidad del activo C
2) α unidades de A más β unidades de B

tienen los mismos flujos. Por tanto, si no hay oportunidad de arbitraje, debe darse además que sus precios en el tiempo presente deben coincidir:

$$C_{\text{ahora}} = \alpha A_{\text{ahora}} + \beta B_{\text{ahora}}$$

La conclusión es que si «conocemos» los activos A y B y además conocemos los flujos del activo C podemos hallar su valor actual. En otros términos, si conocemos A y B podremos valorar C en términos de sus flujos.

Valoración de un derivado

Concretemos el análisis anterior en un caso particular, cuyos ingredientes son:

A: *El* activo sin riesgo: $1 \rightarrow r$
B: Un activo subyacente
C: Un activo derivado

El activo A es el dinero que tiene una rentabilidad segura r, $r > 1$. Es decir, por un euro de hoy nos devuelven r euros al final de ese período de tiempo (por ejemplo, en letras del Tesoro). B es el activo (por ejemplo, una acción de BSCH) sobre el que está escrito el contrato derivado (por ejemplo, una *call*).

Los estados futuros del mundo vendrán determinados por lo único relevante: la *cotización de B*, de la que supondremos que si S es su cotización actual entonces los valores futuros podrán ser sólo

$$\text{Superior: } S \rightarrow S \cdot R_s$$
$$\text{Inferior: } S \rightarrow S \cdot R_i$$

R_s y R_i son las rentabilidades posibles de B.

Como «conocemos» A y B, podemos valorar C en términos de sus contingentes valores futuros. Tras resolver el sistema con incógnitas α y β

$$\begin{cases} C_s = \alpha r + \beta S R_s \\ C_i = \alpha r + \beta S R_i \end{cases}$$

y haciendo el álgebra correspondiente se obtiene:

$$C_{\text{ahora}} = \frac{1}{r}\left[(1-p)\, C_{\text{inf}} + p\, C_{\text{sup}}\right]$$

donde

$$p = \frac{1}{2} - \frac{1}{2}\left(\frac{R_s + R_i - 2r}{R_s - R_i}\right)$$

Importa observar que

- $0 < p < 1$, y *podemos interpretar* p como una probabilidad: p es la probabilidad de subir y $1 - p$ es la probabilidad de bajar.
- C_{ahora} resulta ser *el valor promediado* (*con* p *y* $1 - p$) *actualizado* del *cash-flow*.
- Además

$$r = pR_s + (1 - p)\, R_i$$

Es decir, la probabilidad p es aquella para la que la rentabilidad promediada del activo subyacente es justamente la del activo sin riesgo. Por eso a p se le conoce como la probabilidad *riesgo-neutro*: es la probabilidad subjetiva que un inversor indiferente al riesgo le adjudicaría a la evolución de B.

En palabras:

prima = descuento *valor medio de pagos

promediando de manera que se cumpla

rendimiento *medio* del activo con riesgo = rendimiento *seguro* del activo sin riesgo

Es importante señalar que en el argumento anterior

no sólo se dice qué prima debe tener el derivado (la valoración), sino que además nos dice cómo invertir esa prima (la replicación), basta resolver α y β.

La valoración y la cobertura o replicación son aspectos inseparables de la misma cuestión.

La cobertura es *dinámica*, es decir, quizá sea necesario cambiar la composición de la cartera replicante, y *autofinanciada*, es decir, en estos estadios intermedios no hace falta aportar ni se permite retirar dinero o acciones de la cartera.

El premio Nobel de Black-Scholes-Merton

En los años sesenta, Samuelson postuló una evolución estocástica de los precios de las acciones dada por el movimiento browniano geométrico

$$dS_t = S_t\,(\mu dt + \sigma dW_t)$$

donde

- la unidad de tiempo es el año,
- S_t denota el precio de la acción en tiempo t
- μ es el rendimiento medio (instantáneo)
- σ es la *volatilidad*, la desviación típica (instantánea) del rendimiento

El análisis de la sección anterior llevó a Black y Scholes, y a Merton a observar que

- *para valorar* podemos suponer que
  $\mu = R$ (el tipo de interés continuo anual sin riesgo).

Esto determina la ley de probabilidad del precio de la acción en cualquier momento futuro, y como la prima de la opción depende de esa distribución de valores, se obtiene una fórmula para la prima de la *put* de BSCH:

$$\text{prima} = Ke^{-RT}\,\Phi\,(-d_-) - S\,\Phi\,(-d_+)$$

donde Φ es la función de distribución de la normal y donde:

$$d_\pm = \frac{1}{\sigma\sqrt{T}}\left(\log\,(S/K) + (R \pm \frac{\sigma^2}{2})\sqrt{T}\right)$$

La fórmula desempeña en el mundo financiero un papel fundamental y es comparable a algo tan simple como la noción de interés continuo.

Lo espectacular, contraintuitivo (¡la intuición se educa, claro!) de esta fórmula es que, aparte de los datos del contrato, todo queda determinado por la volatilidad, por el nivel de riesgo, y no por la rentabilidad, que no aparece. En un apéndice vemos un ejemplo, las quinielas, que quizás ayuden a explicar por qué no aparece, en la fórmula de Black-Scholes, la rentabilidad del activo.

Y, ¿ahora?

La fórmula de Black-Scholes es tan sólo el principio. Las aplicaciones de las matemáticas a las finanzas y a los riesgos financieros es un campo de investigación muy activo. Aquí, claro, sólo hemos señalado los fundamentos básicos y de una faceta concreta: los derivados.

Siguen algunas pinceladas sobre qué aplicaciones tienen estas ideas y qué líneas de investigación están en marcha.

- Las hipótesis de lo que podemos llamar entorno de Black-Scholes son muy específicas. Uno puede incluir pagos de dividendos, costes de transacción, tipos de interés y volatilidades que son inciertos o estocásticos, topes en ventas al descubierto y restricciones de volumen de contratación y márgenes financieros...
- Uno supone que los contratos se cumplen, pero el mercado cobra una prima adicional si percibe que éste pueda no ser el caso. Hay derivados que compensan de pérdidas por insolvencia. El riesgo de crédito es ya parte de esta metodología.
- Los futuros y las opciones nacieron como contratos sobre mercaderías, y no sobre instrumentos financieros. En ciertos aspectos, los derivados sobre mercaderías son más complejos y esto sobre todo porque entra en acción la llamada *tasa de conveniencia*: el interés que el poseedor del bien en cuestión pueda tener de no venderlo por su posible consumo ulterior. Su importancia va en aumento. En España hay mercados sobre cítricos y sobre aceite, y hay planes para poner en marcha un mercado sobre electricidad y sobre gas. La modelización que se ha desarrollado para analizar estos mercados es todavía poco satisfactoria.
- Ya comentamos las carencias de esta modelización en situaciones de mercados convulsos, que los tiempos recientes nos llevan a considerar ya como normales. Se empieza a progresar en la aplicación de ideas de física, y ya se empieza hablar de econofísica: transición de fase, ruptura de escala... Para esta próxima primavera se ha convocado un congreso sobre *Turbulencia y Finanzas*.
- ...

Apéndice 1: Argumento 1X2. Las matemáticas de la porra, de las quinielas

El siguiente ejemplo quizá ayude a entender por qué la ponderación de los posibles valores se debe hacer con la probabilidad riesgo-neutro obtenida de datos de mercado, y no, por ejemplo, de datos históricos o de cualquier otro análisis.

Las quinielas se valoran precisamente así: usando la probabilidad riesgo-neutro obtenida de los datos del mercado.

En las quinielas (si hacemos caso omiso de la sustanciosa comisión que cobra el Estado) todo lo recaudado se distribuye entre los acertantes. Es decir, se trata de una (gigantesca) porra.

Supongamos que sólo hay dos resultados en, por ejemplo, una final de Copa: victoria de X o victoria de Y. Y que los datos de apuestas son como sigue:

- Cada apuesta cuesta 1 euro
- Hay un total de 1 millón de apuestas
- 30 % apuestan por X
- 70 % apuestan por Y

Las distintas decisiones individuales de apuesta ya recogen toda la información y los diversos análisis personales técnicos o fundamentales.

Las estadísticas de partidos similares arrojan el siguiente balance:

- Victoria de X en un 40 %
- Victoria de Y en un 60 %

Pero, como veremos, en asuntos de valoración y cobertura de derivados estos datos resultan irrelevantes. Aunque, claro, no son irrelevantes como información para decidir por quién apostar.

Una apuesta por X. El valor hoy, 1 euro, debe ser la media de los flujos descontados, promediados con una cierta probabilidad. Los datos son:

- precio: 1 euro
- no hay descuento

$$\text{pagos}: \begin{cases} \text{Si } X \text{ gana, pago de } \dfrac{1{,}000.000}{300.000} = \dfrac{10}{3} \\ \text{Si } Y \text{ gana, pago de } 0 \end{cases}$$

El cálculo es el siguiente:

$$1 = \left[\underbrace{\frac{10}{3}}_{\text{Pago}} \cdot \underbrace{\frac{1}{}}_{\text{descuento}} \cdot \textbf{Prob}\,(X \text{ gana}) \right] + [0 \cdot 1 \cdot \textbf{Prob}\,(Y \text{ gana})]$$

Lo que nos permite inferir que:

$$\textbf{Prob}\,(X) = 0,3$$

y, claro, también

$$\textbf{Prob}\,(Y) = 0,7$$

Derivados. Veamos ahora cómo utilizaríamos esa información para valorar un derivado.

Supongamos que un revendedor me ofrece el siguiente contrato:

- si gana X, me pagan 5
- si gana Y, me pagan 3

¿Cómo valorar esta oferta? ¿Cómo obtener su precio (justo de mercado)? (Hacemos caso omiso de comisiones.)

Descontando y promediando expectativas obtenemos que

$$\text{precio} = 5 * \textbf{Prob}\,(X) + 3 * \textbf{Prob}\,(Y) = 3,6$$

Justificación: la replicación. Una *cartera* que se compone de

$$1,5 \text{ de boletos de } X \text{ y } 2,1 \text{ de boletos de } Y$$

cuesta precisamente 3,6 euros. Y, esto es lo importante, equivale al contrato que ofrece el revendedor, pues

- si sale X, paga $(1,5)\cdot 10/3 = 5$ euros,
- si sale Y, paga $(2,1)\cdot 10/7 = 3$ euros.

En suma, el derivado se valora con la probabilidad que da el mercado, es decir, con la probabilidad que inferimos a partir de los precios. Esa valoración equivale y es consistente con la replicación.

Apéndice 2: Opciones y Olimpiadas

Con motivo de la última Olimpiada en Atlanta, el COE y su programa ADO habían prometido a nuestros deportistas olímpicos primas sustanciosas. Estas primas dependían del metal de la medalla y de si se trataba de un deporte individual o de una competición por equipos.

En cualquier caso, para cubrir el riesgo económico y así eliminar la incertidumbre que éste conlleva, el ADO contrató con un consorcio de aseguradoras un seguro en los siguientes términos:

si pago total < 250 M	ADO lo asumía todo
si pago total entre 250 M y 400 M	ADO pagaba 250 M y el resto el consorcio
si pago > 400 M	ADO lo asumía todo

Al ADO le preocupaba tener que pagar una cifra superior a 250 M, y quería cubrir ese riesgo (económico, que no deportivo). Pero parecía estar seguro de que no tendría que pagar más allá de 400 M, y por eso no quería asegurarse contra un riesgo de algo que no creía que fuera a ocurrir.

En la terminología al uso: el contrato del que hablamos es una *call* con precio de ejercicio 250 M con una barrera *knock-out* (de desactivación) situada en 400 M. Si bien se trata de una opción particularmente sencilla, pues sólo involucra un intervalo de tiempo.

Referencias

[1] BACHELIER, L. (1900). *Theorie de la spéculation*. Annales Scientifiques de l'Ecole Normale Superieur, tome 17, pp. 21-86. Existe una edición facsímil de editions Gabay, París.

[2] BAXTER, M.; RENNIE, A. (1996). *Financial Calculus*. Cambridge.

[3] BLACK, F.; SCHOLES, M. (1973). *The pricing of Options and Corporate Liabilities*. Journal of Political Economy.

[4] BERNOULLI, D. (1738). *Specimen Theoriae Novae de Mensura Sortis*. Commentarii Academiae scientarum Imperailis Petropolitanae 5, pp. 175-192. Hay traducción al inglés en *Econometrica*.

[5] BERNSTEIN, P. (1996). *Against the Gods*, Wiley, J.

[6] DE LA DEHESA, G. *Los paradigmas financieros en tiempos de crisis*. Artículo aparecido en el diario *El País*, 30 de enero de 1999.

[7] DEHEUVELS, P. *La probabilité, le hasard et la certitude*. Colección Que sais je? Presses Universitaires de France.

[8] KARATZAS, I. (1996). *Lectures on the Mathematics of Finance*. American Mathematical Society.

[9] KARATZAS, I.; SHREVE, S. (1991). *Brownian motion and stochastic calculus*. I. Springer Verlag.

[10] GRANDELL, J. (1991). *Aspects of Risk Theory*. Springer Verlag.

[11] HULL, J. (1993). *Options and futures and other derivatives*. Prentice Hall.

[12] KRUGMAN, P. *Desarrollo, geografía y teoría económica*. Antoni Bosch, Editor.

[13] MALKIEL, B. (1997). *Un paseo aleatorio por Wall Street*. Alianza Editorial.

[14] MERTON, R. (1990). *Continuous time finance*. Blackwell.

[15] REVUZ, D.; YOR, M. (1996). *Continuous Martingales and Brownian Motion*. Springer.

[16] SOROS, G. (1999). *La crisis del capitalismo global*. Editorial Debate.

[17] VEGA, J. DE LA. (1688). *Confusión de confusiones*. Amsterdam.

[18] WIGNER, E. (1960). *The unreasonable effectiveness of mathematics in the natural sciences*. Communications on pure and applied mathematics, 23, pp. 1-14.

Protocols criptogràfics distribuïts i esquemes per compartir secrets

Paz Morillo i Carles Padró

1. Introducció

Encara que es podria afirmar que la criptografia és tan antiga com l'escriptura, el naixement de la *criptologia* científica l'hem de situar en la segona guerra mundial. La criptologia té dues vessants: la *criptografia*, que s'encarrega del disseny de sistemes d'escriptura secreta, i la *criptoanàlisi*, que estudia les debilitats d'aquests sistemes per tal de trencar-ne el secret.

La matematització de la criptologia és el fet que marca l'inici d'aquesta disciplina com a ciència. Un dels moments claus va ser l'aparició de la teoria matemàtica de la informació de Shannon, de la qual l'any passat celebràvem el cinquantè aniversari del seu article fundacional [26], i del seu treball sobre el secret perfecte en els sistemes criptogràfics [27].

Fins a èpoques relativament recents, l'ús de la criptografia es reduïa als àmbits militar i diplomàtic. El volum d'informació que s'havia de transmetre de manera secreta, així com el nombre d'usuaris involucrats eren bastant reduïts. A més, el recursos computacionals eren molt limitats, tant per als qui havien de xifrar i desxifrar els missatges com per als espies que volien descobrir-ne els continguts.

El gran avenç que s'ha produït en la criptologia en les darreres dècades té dues causes: en primer lloc, l'existència d'ordinadors amb cada vegada més potència de càlcul, i en segon lloc, però potser més determinant, la generalització de les comunicacions i altres transaccions telemàtiques que han fet necessari l'ús de la criptografia per part d'un nombre cada vegada més gran d'usuaris. En efecte, els grans canvis en la informàtica i les comunicacions que hem viscut darrerament han fet augmentar molt considerablement el volum d'informació que ha de ser transmesa o emmagatzemada de forma secreta: transaccions bancàries, comerç electrònic, informació sanitària...

Un tret comú de la criptologia amb altres àrees de la informàtica i les comunicacions, com, per exemple, la teoria de codis o la teoria de la complexitat, és que les seves eines matemàtiques provenen principalment de la matemàtica discreta. Veiem, doncs, com àrees de les matemàtiques que es consideraven pures o poc aplicades, com, per exemple, la teoria de nombres, la teoria de grafs, la combinatòria, grups finits i geometries finites, han esdevingut una peça clau en avenços tecnològics molt importants. I és precisament per això que aquestes àrees de les matemàtiques han experimentat un creixement tan destacat en les darreres dècades.

Aquest treball està dividit en tres parts. En la primera part (Secció 2) presentem les idees principals sobre la criptografia de clau pública, la introducció de la qual l'any 1976 va suposar una revolució transcendent en el disseny de sistemes criptogràfics.

La segona part (Secció 3) tracta sobre els esquemes per compartir secrets, en què un secret és repartit entre un grup d'usuaris de manera que només els subconjunts autoritzats poden recuperar-ne el valor.

En aquestes dues primeres parts es presenten, doncs, dues eines criptogràfiques de naturalesa molt diferent. En efecte, així com la seguretat dels criptosistemes de clau pública està basada en la dificultat computacional de certs problemes matemàtics com, per exemple, la descomposició en factors primers de nombres enters grans, la seguretat dels esquemes per compartir secrets és incondicional, és a dir, no depèn dels recursos computacionals dels usuaris. Una altra diferència important són les eines matemàtiques: en la criptografia de clau pública intervenen sobretot la teoria de nombres i la teoria de Galois, mentre que en la compartició de secrets la combinatòria i l'àlgebra lineal hi juguen el paper principal.

En la tercera part (Secció 4) presentem breument alguns protocols criptogràfics distribuïts, és a dir, protocols destinats a resoldre problemes relacionats amb les comunicacions col·lectives en un grup d'usuaris. Les dues eines criptogràfiques que hem presentat en les seccions anteriors es combinen en aquests protocols.

2. Criptografia de clau pública

L'aparició de la criptografia de clau pública, introduïda per Diffie i Hellman [13, 14] l'any 1976, ha estat el punt d'inflexió més important en la criptologia.

Anteriorment, s'utilitzava només la criptografia de clau privada. Els sistemes criptogràfics de clau privada

han de permetre la comunicació secreta entre dos usuaris que comparteixen una clau de treball. La mateixa clau secreta serveix per xifrar i desxifrar els missatges i, naturalment, l'emisor i el receptor han d'acordar la clau secreta a través d'un canal totalment segur.

La dificultat d'aquest pas inicial és el principal inconvenient de la criptografia de clau privada. Pensem, per exemple, en la distribució de les claus secretes de treball en una xarxa de n usuaris. Es necessiten $(n^2 - n)/2$ claus privades, és a dir, en un sistema amb un miler d'usuaris, calen 450.000 connexions totalment segures per poder distribuir les claus.

En la criptografia de clau pública, cada usuari A escull una clau pública i una clau privada. Qualsevol altre usuari pot xifrar un missatge adreçat a A utilitzant la clau pública de A i només el participant A pot desxifrar aquest missatge utilitzant la seva clau privada. Ens podem imaginar, doncs, que cada participant té una bústia on tothom hi pot dipositar missatges però només ell té la clau per obrir-la.

Els criptosistemes de clau pública estan basats en les anomenades *funcions unidireccionals amb trampa*, que són funcions fàcils de calcular però tals que la seva inversa és difícil de calcular si no es coneix alguna informació addicional. Així, cada participant A té la seva funció de xifrat e_A (pública) tal que la seva inversa d_A, la funció de desxifrat, és difícil de calcular si no es coneix la clau privada de A.

Veiem, doncs, que la seguretat dels criptosistemes de clau pública és condicional, és a dir, depèn de la capacitat computacional dels participants. Per tant, s'han de considerar funcions unidireccionals tals que el temps necessari per calcular-ne la inversa, amb els recursos computacionals dels participants i el millor algorisme conegut, sigui prou gran. De fet, la seguretat incondicional és impossible en aquest cas, ja que un participant amb recursos computacionals infinits que conegui el missatge xifrat $y = e_A(x)$ pot anar provant tots el valors possibles del missatge fins que trobi el valor de x tal que $y = e_A(x)$.

La seguretat dels criptosistemes de clau pública està basada, doncs, en la dificultat d'algun problema. Així, per exemple, la seguretat del criptosistema RSA està basada en la dificultat de factoritzar un nombre enter producte de dos primers grans. Altres criptosistemes es basen en la dificultat de calcular el logaritme discret en certs grups finits, com el grup multiplicatiu del cos $\mathbb{Z}/(p)$ o el grup de punts d'una corba el·líptica. Estudiarem més endavant el funcionament d'aquests criptosistemes.

En la pràctica, la velocitat de xifrat dels sistemes de clau pública coneguts és molt menor que la dels sistemes de clau privada. Per tant, si s'han de xifrar grans volums d'informació, es fan servir sistemes de clau privada. En aquest cas, s'utilitza un criptosistema de clau pública per distribuir les claus del criptosistema de clau privada a través de canals insegurs.

2.1. El criptosistema RSA

La primera realització de la idea de la criptografia de clau pública de Diffie i Hellman [13, 14] va ser el criptosistema proposat l'any 1978 per Rivest, Shamir i Adleman [24], conegut per RSA. La seguretat d'aquest sistema es basa en la dificultat de factoritzar un nombre enter producte de dos primers grans.

En el criptosistema RSA, cada participant ha d'escollir un nombre enter producte de dos nombres primers grans, $n = pq$, i dos enters a, b tals que $ab \equiv 1 \pmod{\Phi(n)}$, on Φ és la funció d'Euler i, per tant, $\Phi(n) = (p - 1)(q - 1)$. La clau pública del participant és (n, b) i la clau privada és (p, q, a).

El mètode de xifrat i desxifrat utilitza el Teorema d'Euler, que afirma que, si x es un enter primer amb n, aleshores $x^{\Phi(n)} \equiv 1 \pmod{n}$.

Quan A vol enviar un missatge xifrat a B, ha de prendre la seva clau pública (n_B, b_B) i codificar el missatge de manera que sigui un element $x \in \mathbb{Z}/(n_B)$. El missatge xifrat que A envia a B és $y = e_B(x) \equiv x^{b_B} \pmod{n_B}$. Per desxifrar el missatge, B calcula $d_B(x) = y^{a_B}$. Veiem que, pel Teorema d'Euler, $d_B(x) = y^{a_B} = x^{a_B b_B} \equiv x \pmod{n_B}$.

Observem que per calcular la funció de desxifrat d_B s'ha de conèixer la clau privada a_B. Ara bé, si un oponent és capaç de factoritzar $n_B = p_B q_B$, aleshores pot calcular $\Phi(n_B)$ i, per tant, pot trobar a_B i desxifrar qualsevol missatge adreçat a B.

Així doncs, la mida dels nombres primers a partir dels quals un usuari construeix les seves claus ha de ser suficient per tal que el temps necessari per factoritzar $n = pq$ sigui prou gran. Actualment, se solen prendre nombres primers d'uns 80 dígits decimals, però pel que sembla comença a considerar-se que aquesta mida no és prou segura. Hem dit abans que la seguretat del criptosistema RSA està basada en la dificultat de factoritzar $n = pq$. Per tant, hauríem de veure que trencar el RSA és tan difícil com factoritzar.

S'ha demostrat que trobar la clau privada a, que permet desxifrar qualsevol missatge, és tan difícil com factoritzar n. Més concretament, qualsevol algorisme que calculi a a partir de n i b pot ser utilitzat com a subrutina en un algorisme probabilístic eficient per factoritzar n.

Ara bé, podria ser que es pogués desxifrar un missatge sense haver de calcular valor de a. No s'ha demostrat, doncs, que trencar el RSA sigui tan difícil com factoritzar i, de fet, alguns treballs recents semblen indicar el contrari [9, 4].

2.2. Criptosistemes basats en el logaritme discret

Un altre problema difícil que s'ha utilitzat per dissenyar criptosistemes de clau pública és el càlcul del logaritme discret en certs grups finits. És a dir, donats α, $\beta \in G$ trobar un enter a tal que $\alpha^a = \beta$. En aquest cas direm que $a = \lg_\alpha \beta$.

El criptosistema proposat per ElGamal [15] l'any 1985, conegut precisament com a *criptosistema ElGamal*, es basa en la dificultat de calcular el logaritme discret en el grup multiplicatiu $(\mathbb{Z}/(p))^*$, on p és un nombre primer adequat. Concretament, tenint en compte els mètodes i la capacitat de càlcul actuals, p ha de ser un nombre primer de més de 150 dígits decimals tal que $p - 1$ tingui com a mínim un factor primer gran.

En el criptosistema ElGamal, cada usuari escull un nombre primer p amb les característiques esmentades abans, un generador α del grup multiplicatiu $(\mathbb{Z}/(p))^*$, que és cíclic, i un enter a tal $0 \leq a \leq p - 2$, i calcula $\beta = \alpha^a$. La clau pública de l'usuari és (p, α, β) i la seva clau privada és a. Si es vol xifrar un missatge $x \in (\mathbb{Z}/(p))^*$ adreçat a un usuari A amb clau pública (p, α, β), es pren a l'atzar un element $k \in \mathbb{Z}/(p - 1)$, que es mantindrà secret, i es calcula el missatge xifrat $e_A(x, k) = (y_1, y_2)$ on $y_1 = \alpha^k$ i $y_2 = x\beta^k$. L'usuari A pot desxifrar el missatge utilitzant la clau privada a. En efecte, la funció de desxifrat és $d_A(y_1, y_2) = y_2(y_1^a)^{-1} = x$. Tots els càlculs es realitzen en el cos $\mathbb{Z}/(p)$.

A diferència del criptosistema RSA, el criptosistema ElGamal és no determinista. En efecte, ja que el missatge xifrat depèn de l'elecció aleatòria de k, hi ha molts missatges xifrats que corresponen a un sol missatge en clar.

Observem que, de fet, es pot construir un criptosistema com l'ElGamal en qualsevol grup finit on el problema del logaritme discret sigui intractable. Per exemple, Koblitz [17] i Miller [19] van proposar l'any 1986 utilitzar el grup de punts d'una corba el·líptica.

2.3. Signatures electròniques

Un dels avantatges principals de la criptografia de clau pública és que permet implementar altres aplicacions a part del xifrat de missatges com, per exemple, les signatures electròniques o els esquemes d'identificació. Vegem, per exemple, com es pot implementar un esquema de signatura electrònica a partir d'un criptosistema de clau pública determinista, és a dir, tal que hi ha un únic missatge xifrat per a cada missatge en clar. El criptosistema RSA, per exemple, compleix aquesta propietat. Suposem que l'usuari A, amb funció de xifrat e_A (pública) i funció de desxifrat d_A (privada) vol enviar un missatge signat a l'usuari B, que té funcions de xifrat i desxifrat e_B i d_B respectivament. Per fer-ho, l'usuari A pot enviar a B $(x, y) = (x, d_A(x))$. Recordem que només A pot calcular d_A. L'usuari B pot verificar que efectivament A ha signat el missatge comprovant que $x = e_A(y)$.

Una signatura ha d'estar lligada al document que se signa. En un document sobre paper, la signatura ha d'estar en el mateix full. En les signatures electròniques, la signatura ha de dependre del contingut del missatge. A més, qualsevol ha de poder verificar una signatura, no només aquell a qui va adreçat el missatge. Observem que en l'esquema que hem descrit abans, la verificació es fa utilitzant la funció e_A, que és pública.

Atès que el criptosistema ElGamal no és determinista, no es pot utilitzar de la manera que hem descrit abans per implementar un esquema de signatura electrònica. De tota manera, en el mateix treball de 1985 en què proposa el criptosistema [15], ElGamal presenta un esquema de signatura electrònica basat també en la dificultat de calcular el logaritme discret. Una modificació de l'esquema de signatura ElGamal, el Digital Signature Standard va ser adoptat com a estàndard l'any 1994 pel National Institute of Standards and Technology dels Estats Units [20].

3. Esquemes per compartir secrets

Suposem que en un col·lectiu de persones es necessita l'acord d'algun grup qualificat per iniciar certes accions o prendre algunes decisions. Aquesta situació es pot resoldre amb un esquema per compartir secrets.

En un *esquema per compartir secrets* es reparteix un valor $k \in K$ secret en fragments entre els participants d'un conjunt P, de manera que només els subconjunts autoritzats de participants poden reconstruir el secret a partir dels seus fragments. Els esquemes per compartir secrets van ser introduïts independentment per Blakley [2] i Shamir [25] l'any 1979. Podeu trobar una introducció a la compartició de secrets en [29] o bé en [31].

La família Γ de subconjunts qualificats de P s'anomena *estructura d'accés* i ha de ser *monòtona*. És a dir, tots els subconjunts que contenen un subconjunt autoritzat són autoritzats.

En un *esquema per compartir secrets* amb *estructura d'accés* Γ i conjunt de possibles valors del secret K, a partir d'un valor secret $k \in K$ i d'una certa elecció aleatòria, cada participant $p \in P$ rep un fragment $s_p \in S_p$ on S_p denota el conjunt de possibles valors del fragment s_p, de manera que:

1. Els subconjunts autoritzats de P poden reconstruir el valor del secret k a partir dels seus fragments. És a dir, si $A \in \Gamma$, existeix un únic valor $k \in K$ per al qual els fragments $\{s_p : p \in A\}$ són possibles.
2. Els participants d'un subconjunt no autoritzat no poden obtenir *cap informació* sobre el valor del secret a partir dels seus fragments. És a dir, si, $A \notin \Gamma$, tots els valors possibles del secret són igualment probables tot i coneixent els fragments $\{s_p : p \in A\}$.

3.1. L'esquema de Shamir

L'esquema de Shamir [25], és un dels primers esquemes per compartir secrets que es van proposar. L'estructura d'accés d'aquest esquema està formada pels subconjunts amb, com a mínim, t participants d'un conjunt P amb N participants. Els esquemes que tenen

aquesta estructura d'accés s'anomenen *esquemes de llindar* (t, N).

Considerem el conjunt de participants $P = \{1, 2, \ldots N\}$ i un enter t tal que $1 \leq t \leq N$. El valor secret que es reparteix en aquest esquema és un element del cos finit $\mathbb{F}_q$, on $q = p^r > N$ i p és primer.

En la fase inicial s'assigna a cada participant $i \in P$ un element del cos $x_i \in \mathbb{F}_q$, on $x_1, x_2, \ldots, x_N$ són no nuls i diferents dos a dos. Els valors dels x_i són públics. Per distribuir un secret $k \in \mathbb{F}_q$, es pren a l'atzar un polinomi de grau menor o igual que $t - 1$,

$$f = k + a_1 x + \ldots a_{t-1} x^{t-1} \in \mathbb{F}_q [x]$$

tal que $f(0) = k$ i cada participant rep el seu fragment $s_i = f(x_i) \in \mathbb{F}_q$. Atès que els valors de k i dels coeficients a_j s'han de mantenir secrets, aquesta part del protocol l'ha de dur a terme un participant especial, $D \notin P$, anomenat *distribuïdor*.

Vegem que, efectivament, l'esquema de Shamir és un esquema per compartir secrets amb estructura d'accés $\Gamma = \{A \subset P : |A| \geq t\}$. Qualsevol subconjunt de t participants pot trobar el polinomi f i, per tant, el secret k, a partir dels seus fragments $s_{ij} = f(x_{ij})$, on $j = 1, \ldots, t$. Poden fer-ho, per exemple, utilitzant la interpolació de Lagrange. D'altra banda, si només es coneixen $t - 1$ fragments $s_{ij} = f(x_{ij})$, on $j = 1, \ldots, t - 1$, per a qualsevol possible valor del secret $k' \in \mathbb{F}_q$, existeix un únic polinomi $g(x) \in \mathbb{F}_q [x]$ amb grau menor o igual que $t - 1$ tal que $g(0) = k'$ i $g(x_{ij}) = s_{ij}$ per a tot $j = 1, \ldots, t - 1$. Per tant, els fragments dels participants d'un subconjunt no autoritzat no donen cap informació sobre el valor del secret.

Observem que la seguretat de l'esquema de Shamir és incondicional, és a dir, no està basada en cap supòsit computacional. Dit d'una altra manera, els participants d'un subconjunt no autoritzat no obtindrien cap informació sobre el secret encara que tinguessin recursos computacionals infinits. De fet, per la definició, aquesta és una propietat que ha de tenir qualsevol esquema per compartir secrets.

3.2. Esquemes lineals

El principal inconvenient de l'esquema de Shamir és que només és vàlid per a estructures d'accés de llindar. Els esquemes lineals, proposats per Brickell [6] l'any 1989, poden ser implementats en una família més àmplia d'estructures d'accés.

Considerem $\mathbb{F}_q$ un cos finit i E un espai vectorial de dimensió finita sobre $\mathbb{F}_q$. Considerem un conjunt de participants P i un participant especial, el distribuïdor, $D \notin P$. Per a cada $p \in P \cup \{D\}$, considerem $\pi_p \in E^*$ un element de l'espai dual de E. Donat un valor secret $k \in \mathbb{F}_q$, el distribuïdor pren a l'atzar un vector $v \in E$ tal que $k = \pi_D(v)$ i calcula, per a cada participant $p \in P$, el fragment $s_p = \pi_p(v) \in \mathbb{F}_q$.

D'aquesta manera es defineix un esquema per compartir secrets l'estructura d'accés Γ del qual està formada pels subconjunts $A \subset P$ tals que π_D és combinació lineal dels vectors π_p amb $p \in A$. En efecte, si $\pi_D = \Sigma_{p \in A} \lambda_p \pi_p$, aleshores

$$k = \pi_D(v) = \sum_{p \in A} \lambda_p \pi_p(v) = \sum_{p \in A} \lambda_p s_p .$$

Observem que els escalars λ_p poden ser calculats per qualsevol participant ja que els vectors π_p són coneguts públicament. No és difícil provar que els participants d'un subconjunt no autoritzat no poden obtenir cap informació sobre el valor del secret.

Els esquemes per compartir secrets construïts d'aquesta manera s'anomenen *esquemes lineals ideals*. Observem que, per a un subconjunt autoritzat minimal donat, el secret és una funció lineal dels fragments. El terme ideal es refereix al fet que, en aquests esquemes, el conjunt de possibles valors del secret coincideix amb el conjunt de possibles valors del fragment de cada participant. En el nostre cas, $K = S_p = \mathbb{F}_q$ per a tot $p \in P$. L'esquema de llindar de Shamir també és ideal, ja que també compleix aquesta propietat.

De fet, l'esquema de Shamir és un cas particular d'esquema lineal. S'obté prenent $E = (\mathbb{F}_q)^t$, $\pi_D = (1, 0, \ldots, 0) \in E^*$ i $\pi_i = (1, x_i, x_i^2, \ldots, x_i^{t-1}) \in E^*$ per a tot $i = 1, \ldots, N$.

La família d'estructures d'accés que es poden implementar amb un esquema lineal ideal, tot i que és més àmplia que la de les estructures de llindar, és força restringida. Per poder trobar un esquema per compartir secrets per a qualsevol estructura d'accés, hem de considerar una construcció més general: els esquemes lineals (no necessàriament ideals).

Així com en els esquemes lineals ideals a cada participant li correspon un vector de E^*, en un esquema lineal general s'assigna a cada participant $p \in P \cup \{D\}$ un conjunt de vectors, o, equivalentment, el subespai $V_p \subset E^*$ que generen. Aquests subespais han de ser tals que, per a tot $A \subset P$,

$$V_D \subset \sum_{p \in A} V_p \quad \text{o bé} \quad V_D \cap \sum_{p \in A} V_p = \{0\} \qquad (1)$$

I, en aquest cas, l'estructura d'accés Γ és

$$\Gamma = \left\{ A \subset P : V_D \subset \sum_{p \in A} V_p \right\}$$

El conjunt de secrets és l'espai vectorial $(\mathbb{F}_q)^{r_D}$ on $r_D = \dim V_D$ i el fragment del participant p és un element de l'espai vectorial $(\mathbb{F}_q)^{r_p}$ on $r_p = \dim V_p$.

Simmons, Jackson i Martin [28] van provar que qualsevol estructura d'accés Γ es pot implementar amb un esquema lineal. La demostració que donen d'aquest fet és constructiva.

3.3. Taxa d'informació

En un esquema lineal, podem considerar que dim V_D i dim V_p mesuren, respectivament, la mida del secret i la mida del fragment que rep el participant p. A partir de la propietat (1), es pot veure que dim $V_D \leq$ dim V_p per a tot $p \in P$. És a dir, el fragment de qualsevol participant és com a mínim tan gran com el secret. Això no només es compleix per als esquemes lineals, sinó que és cert en qualsevol esquema per compartir secrets.

El principal problema del mètode proposat en [28] per construir un esquema lineal per a una estructura d'accés arbitrària és que, en els esquemes obtinguts, la mida dels fragments que reben els participants pot ser molt més gran que la mida del secret. És a dir, la dimensió de V_p pot ser molt més gran que la dimensió de V_D. Això no és desitjable ja que, en general, la seguretat d'un sistema disminueix amb la quantitat d'informació que s'ha de mantenir en secret. A més, si la mida dels fragments és molt gran, augmenta el nombre d'operacions necessàries per calcular-los i també el nombre d'operacions per calcular el secret a partir dels fragments d'un subconjunt autoritzat.

Així doncs, un dels paràmetres bàsics en el disseny d'esquemes per compartir secrets és la taxa d'informació, que es defineix com la raó entre la llargada (en bits) del secret i la llargada màxima dels fragments que reben els participants. És a dir, si Σ és un esquema per compartir secrets amb estructura d'accés Γ i conjunt de possibles valors del secret K, la *taxa d'informació* $\rho(\Sigma, \Gamma, K)$ de l'esquema Σ és

$$\rho(\Sigma, \Gamma, K) = \frac{\log |K|}{\max_{p \in P} \log |S_p|}$$

on S_p és el conjunt de possibles valors del fragment del participant p. Observem que la taxa d'informació d'un esquema lineal és igual a

$$\frac{\dim V_D}{\max_{p \in P} \dim V_p}$$

Direm que un esquema per compartir secrets és *ideal* si té taxa d'informació igual a 1, que és el valor màxim que pot prendre aquest paràmetre. En aquest cas, la llargada dels fragments és igual a la llargada del secret. Diem que una estructura d'accés és ideal si admet un esquema ideal.

La caracterització de les estructures d'accés ideals és encara un problema obert. Els resultats obtinguts en la recerca sobre aquest problema [7, 22] fan pensar que una condició necessària i suficient perquè una estructura d'accés sigui ideal és l'existència d'un esquema lineal ideal que la realitzi. A més, en el cas que l'estructura no admeti un esquema ideal, es pot comprovar que, amb les millors tècniques conegudes per trobar esquemes amb taxa d'informació elevada [8, 3, 30] s'obtenen esquemes lineals.

3.4. Detecció de mentiders

Els esquemes lineals presenten, però, un greu inconvenient: la seva vulnerabilitat a l'acció de participants mentiders. En efecte, per a qualsevol subconjunt autoritzat minimal $A = \{p_1, ..., p_r\} \in \Gamma$ existeix una aplicació lineal $\chi_A : E_{p_1} \times ... \times E_{p_r} \rightarrow E_D$ que ens permet obtenir el secret a partir dels fragments dels participants de A. Observem que l'aplicació χ_A és pública, és a dir, pot ser calculada per qualsevol participant. Per tant, en el moment de recuperar el secret, el participant p_1 pot donar un fragment fals $s^*_{p_1}$ de manera que $\chi_A(s^*_{p_1}, s_{p_2}, ..., s_{p_r}) = k^* \neq k$. D'aquesta manera, el participant p_1 enganya els altres participants i, a més, pot trobar el valor correcte del secret, $k = k^* + \chi_A(s_{p_1} - s^*_{p_1}, 0, ..., 0)$.

Ara bé, els esquemes lineals es poden modificar per tal d'obtenir esquemes *incondicionalment segurs contra mentiders*. En un esquema incondicionalment segur, la probabilitat que una coalició de participants que no coneixen el secret puguin mentir sense ser detectats no depèn dels recursos computacionals d'aquests participants.

Per exemple, a partir d'un esquema lineal tal que $E_D \cong \mathbb{F}_q$, es pot construir un esquema segur de la forma següent: donat un secret $k \in \mathbb{F}_q$ es reparteixen fragments corresponents a k i al seu quadrat k^2. Es pot veure que d'aquesta manera s'obté un esquema incondicionalment segur amb probabilitat de mentir sense detecció igual a $1/q$. És a dir, és tan difícil mentir amb èxit com endevinar el valor del secret a l'atzar. Observem que per tal d'obtenir seguretat contra tramposos hem hagut de reduir la taxa d'informació a la meitat.

Aquesta solució està basada en l'esquema segur que es proposa en [23] per al cas d'esquemes lineals ideals. En aquest cas, si comparem la taxa d'informació amb la fita donada [21], veiem que és gairebé òptima en relació amb la probabilitat de mentir. Ogata i Kurosawa [21] proposen un esquema segur amb la mateixa probabilitat de mentir i taxa d'informació òptima. Aquest esquema, que està basat en conjunts de diferències planars, és computacionalment menys eficient que el proposat en [23].

4. Protocols criptogràfics distribuïts

En sistemes de comunicació que involucren un nombre important d'usuaris, apareixen diverses situacions en què els criptosistemes habituals, que permeten la comunicació secreta entre dos usuaris, no són suficients. Les votacions electròniques o la televisió de pagament són exemples d'aquestes situacions.

En aquesta secció, presentem algunes d'aquestes situacions i algunes de les solucions criptogràfiques que s'han proposat. Aquestes solucions s'engloben en el que s'anomenen *protocols criptogràfics distribuïts*, és a dir, protocols en què participen col·lectivament un grup d'usuaris.

Els esquemes per compartir secrets són un exemple d'aquesta mena de protocols. Com hem dit abans, poden ser utilitzats en un col·lectiu quan certes decisions s'han de prendre amb l'acord d'algun grup autoritzat. Ara bé, per la seva naturalesa, els esquemes per compartir secrets són una eina important en el disseny de protocols criptogràfics distribuïts que s'apliquen en moltes altres situacions. La criptografia de clau pública és l'altre component important d'aquests protocols.

4.1. Custòdia de claus secretes i criptografia de llindar

Un tema que s'ha estat debatent intensament des de fa uns quants anys, sobretot als Estats Units, és la possibilitat que, per ordre judicial, es puguin desxifrar certs missatges si se sospita que estan relacionats amb alguna activitat delictiva.

Així, qualsevol usuari d'un sistema criptogràfic estaria obligat a lliurar les seves claus secretes a alguna autoritat. Aquest dipositari de les claus s'hauria de comprometre a utilitzar-les només per ordre judicial.

Ara bé, és clar que els beneficis que el dipositari podria obtenir d'una utilització il·lícita de les claus secretes serien immensos i, per tant, no es pot demanar a ningú que sigui tan honrat com per evitar caure en aquesta temptació.

Una possible solució es utilitzar un esquema per compartir secrets i repartir fragments de les claus entre diverses autoritats, si pot ser amb interessos enfrontats. Així, si un jutge creu necessari desxifrar les comunicacions adreçades a algun usuari, s'han d'ajuntar els fragments d'un grup qualificat de dipositaris per reconstruir la clau secreta del sospitós.

De tota manera, aquesta solució presenta un problema important: un cop recuperada la clau secreta es podrien desxifrar tots els missatges xifrats amb aquesta clau encara que l'ordre judicial autoritzés desxifrar només alguns d'aquests missatges.

Una solució millor l'ofereix la *criptografia de llindar*, que va ser introduïda per Boyd [5] i Desmedt [11]. En un criptosistema de llindar hi ha una clau pública i una clau secreta que es reparteix en fragments en un grup de participants. A partir d'un missatge xifrat amb la clau pública, cada participant obté, amb el seu fragment de clau secreta, un fragment del missatge de manera que el missatge en clar pot ser reconstruït a partir dels fragments dels participants en un subconjunt qualificat. Observem, doncs, que no es recupera la clau secreta, sinó només el contingut del missatge xifrat. S'han proposat criptosistemes de llindar basats en el problema del logaritme discret [12] i basats en el criptosistema RSA [10].

4.2. Emissions xifrades

En un sistema de televisió de pagament, tots els usuaris reben el mateix senyal, però només els usuaris autoritzats l'han de poder desxifrar. Observem que, en aquest cas, el conjunt d'usuaris autoritzats és variable.

Els *esquemes de xifrat d'emissions* ofereixen solucions per a situacions com aquesta. En un d'aquests esquemes, l'emissor escull un conjunt de claus i assigna a cada usuari d'un conjunt U un subconjunt d'aquestes claus. Per a cada possible subconjunt $P \subset U$ d'usuaris autoritzats, existeix una funció de xifrat e_P de manera que un missatge xifrat $e_P(x)$ només pot ser desxifrat pels participants de P. Periòdicament, l'emissor escull una clau secreta S per xifrar l'emissió amb un criptosistema de clau privada. L'emissor calcula la clau xifrada $e_P(S)$, on P és el conjunt de participants autoritzats en aquell moment, i l'emet pel mateix canal que l'emissió xifrada. Només els participants de P podran obtenir S i, per tant, desxifrar l'emissió.

El primer esquema de xifrat d'emissions va ser proposat per Berkovits [1] l'any 1991. Darrerament, s'han proposat diversos esquemes amb diferents prestacions [32, 33, 18].

Referències

[1] BERKOVITS, S. (1992). «How to broadcast a secret». *Advances in Cryptology CRYPTO'91. Lecture Notes in Computer Science,* núm. 547, p. 536-541.

[2] BLAKLEY, G. R. (1979). «Safeguarding cryptographic keys». *AFIPS Conference Proceedings,* núm. 48, p. 313-317.

[3] BLUNDO, C.; DE SANTIS, A.; STINSON, D. R.; VACCARO, U. (1995). «Graph Decompositions and Secret Sharing Schemes». *J. Cryptology,* núm. 8; p. 39-64.

[4] BONEH, D.; VENKATESAN, R. (1998). «Breaking RSA may not be equivalent to factoring». *Advances in Cryptology EUROCRYPT'98. Lecture Notes in Computer Science,* núm. 1403, p. 59-71.

[5] BOYD, C. (1986). «Digital multisignatures». A: *Cryptography and Coding*, H. Baker and F. Piper, editors. Clarendon Press, p. 241-246.

[6] BRICKELL, E. F. (1989). «Some ideal secret sharing schemes». *J. Combin. Math. and Combin. Comput,* núm. 9, p. 105-113.

[7] —; DAVENPORT, D. M. (1991). «On the Classification of Ideal Secret Sharing Schemes». *J. Cryptology,* núm. 4, p. 123-134.

[8] —; STINSON, D. R. (1992). «Some Improved Bounds on the Information Rate of Perfect Secret Sharing Schemes». *J. Cryptology,* núm. 5, p. 153-166.

[9] COPPERSMITH, D. (1998). «Finding a small root of a univariate modular equation». *Advances in Cryptology EUROCRYPT'96. Lecture Notes in Computer Science,* p. 155-165.

[10] DE SANTIS, A.; DESMEDT, Y.; FRANKEL, Y.; YUNG, M. (1994). «How to share a function securely». *26th Annual ACM Symposium on Theory of Computing*, p. 522-533.

[11] DESMEDT, Y. (1987). «Society and group oriented crytpography: a new concept». *Advances in Cryptology CRYPTO'87. Lecture Notes in Computer Science*, p. 120-127.

[12] —; FRANKEL, Y. (1989). «Threshold cryptosystems». *Advances in Cryptology CRYPTO'89. Lecture Notes in Computer Science,* p. 307-315.

[13] DIFFIE, W.; HELLMAN, M. E. (1976). «Multiuser cryptographic techniques». AFIPS *Conference Proceedings,* núm. 45, p. 109-112.

[14] DIFFIE, W.; HELLMAN, M. E. (1976). «New directions in cryptography». *IEEE Transactions on Information Theory,* núm. 22, p. 644-654.

[15] ELGAMAL, T. (1985). «A public key cryptosystem and a signature scheme based on discrete logarithms». *IEEE Transactions on Information Theory,* núm. 31, p. 469-472.

[16] KOBLITZ, N. (1987). *A Course in Number Theory and Cryptography*. Springer-Verlag.

[17] —; (1987). «Elliptic curve cryptosystems». *Mathematics of Computation,* núm. 48, p. 203-209.

[18] KUROSAWA, K.; DESMEDT, Y. (1998). «Optimum traitor tracing and asymmetric schemes». *Advances in Cryptology EUROCRYPT'98. Lecture Notes in Computer Science,* núm. 1403, p. 145-157.

[19] MILLER, V. (1985). «Uses of elliptic curves in cryptography». *Advances in Cryptology CRYPTO'85. Lecture Notes in Computer Science,* núm. 218, p. 417-426.

[20] (1994). *Digital signature standard*. National Bureau of Standards. FIPS Publication 186.

[21] OGATA, W.; KUROSAWA, K. (1996). «Optimum Secret Sharing Scheme Secure against Cheating». *Advances in Cryptology EUROCRYPT 96. Lecture Notes in Computer Science,* núm. 1070, p. 200-211.

[22] PADRÓ, C.; SÁEZ, G. (1998). «Secret sharing schemes with bipartite access structure». *Advances in Cryptology EUROCRYPT'98. Lecture Notes in Computer Science,* núm. 1403, p. 500-511.

[23] PADRÓ, C.; SÁEZ, G.; VILLAR, J. L. (En premsa). «Detection of cheaters in vector space secret sharing schemes». *Designs, Codes and Cryptography.*

[24] RIVEST, R. L.; SHAMIR, A.; ADLEMAN, L. (1978). «A method for obtaining digital signatures and public key cryptosystems». *Communications of the ACM,* núm. 21, p. 120-126.

[25] A. SHAMIR. (1979). «How to share a secret». *Comm. of the ACM*, núm. 22, p. 612-613.

[26] SHANNON, C. E. (1948). «A mathematical theory of communication». *Bell Systems Technical Journal,* núm. 27, p. 379-423, 623-656.

[27] —; (1949). «Communication theory of secrecy systems». *Bell Systems Technical Journal,* núm. 28, p. 656-715.

[28] SIMMONS, G. J.; JACKSON, W.; MARTIN, K. (1991). «The geometry of secret sharing schemes». *Bulletin of the ICA,* núm. 1, p. 71-88.

[29] STINSON, D. R. (1992). «An explication of secret sharing schemes». *Designs, Codes and Cryptography,* núm. 2, p. 357-390.

[30] —; (1994). «Decomposition Constructions for Secret-Sharing Schemes». *IEEE Trans. on Information Theory,* f 40, p. 118-125.

[31] —; (1995) *Cryptography: Theory and Practice*. CRC Press Inc., Boca Raton.

[32] —; (1997) «On some methods for unconditionally secure key distribution and broadcast encryption». *Designs, Codes and Cryptography,* núm. 12, p. 215-243.

[33] —; WEI, R. (1998). «Combinatorial properties and constructions of traceability schemes and frameproof codes». *SIAM Journal on Discrete Mathematics,* núm. 11, p. 41-53.

Mathematics as a key to key technologies

Helmut Neunzert

In memory of a very precious friend,
Henrik Martens, who loved mathematics
and its applications, Spanish guitar music,
Norwegian mountains and Oberwolfach
–even more than I do.

Asking people on the road, in parliaments, in ministries about useful sciences, one cannot expect that mathematics is very often named in a top list. Of course, it is useful to practise the brain of pupils; to be good in mathematics may be regarded as a sign for some intelligence, but for example in my country even many well educated people seem to be proud that mathematics was their horror subject in school («but see, what I have nevertheless achieved»). There are, however, countries like France, where the prestige of mathematics is really very much higher; I quote a sentence from a letter of the former French president Mitterand to the assembly of mathematicians in 1992: «Your discussions refer to a very essential domain of research, which commands the scientific and technological progress of a country and they refer to a central discipline of our educational system.» Mathematics commands the development of science and technology of the country –maybe a reality or at least a dream in France, but what about other countries?

In Germany some of the research directors of big companies see it like Mitterand. In 1994 I asked 5 of them in a letter, what they think about the economic benefits of mathematics; not all replied, but for example Prof. Weule, by then research director of Daimler, wrote: «The demand for a maximum output of industrial research and development today can only be fulfilled by an increasing use of mathematical methods.» And he explained why: «Examples are simulation methods, which allow to reduce the experimental and constructive effort for the development of complex products drastically.» Simulation here is the key word –computer simulation substitutes more and more real experiments; there is still a huge need of new simulation tools, which of course must always be evaluated by experiments. The main message is that:

Mathematics always is the core of any computer simulation.

Those in industry, who almost daily use a certain software package to simulate a process or the behaviour of a product, sometimes forget this simple fact –a disadvantage for them, since they may forget the limitations of their tool too and a disadvantage for mathematicians, since their key role comes out of view.

We will discuss the role of mathematics in simulation –and therefore in control and optimisation– of technical processes and products more in details. The general task was clearly formulated in a lecture by J.L. Lions in 1994:

Mathematics helps to make things better, faster, safer, cheaper by the simulation of complex phenomena, the reduction of the flood of data, visualisation.

This means hope and challenge for us! For 20 years I have tried to work in this direction; I had my first projects with industry in 1978 and have, since then, found and worked on over 200 projects from companies from all over the world. In 1980 we developed at Kaiserslautern a new programme called «Technomathematics», which combines proper mathematics with computer science and engineering and is now offered at more than 20 German and some foreign universities. In 1986 I was a co-founder of the European Consortium for Industrial Mathematics, which still bundles the efforts of European institutes and universities in this field. During that time we realised that Industrial Mathematics, with proper modifications, may also be of great interest for scientists in 3rd world countries and started in 1987 a 2 year master programme with today 30 beginners per year. Finally, in 1995 we founded an «Institute for Industrial Mathematics» (ITWM) outside of the university, which has to earn 75 % of its total budget through industrial and public projects. For 1999 we expect a total budget of 10 Million DM, of which 7.5 have to be earned; we shall do that with almost 70 fully em-

ployed postdoc scientists and about 40 part time scientists and PhD students. We expect to become the first «Fraunhofer-Institut» for Mathematics in 2000.

These experiences are the basis for my analysis of the following questions:

A) Why has mathematics become so important for industry and commerce during the last 20 years and which kind of mathematics is needed?

B) What are the consequences for research and education in the field of mathematics?

The main part of this paper will consist of examples, how mathematics really helps to solve industrial problems; these examples are taken from our Institute for Industrial Mathematics, from research in the Technomathematics group at my university, but also from ECMI groups and a company called Tecmath, which originated 10 years ago from my university group and has already a very successful history.

A) Simulation, data reduction, control, optimisation, visualisation

Key words for the role of mathematics in industry. And mathematics as a key to these keywords: all these statements are already made in the introduction. But why just now during the last decades of the 20th century? The only reason is that modern computers allow the evaluation of realistic mathematical models. Therefore, the behaviour of complex, 3-dimensional technical systems can be predicted, designed and optimised in a virtual reality, which is astonishingly near to «real» reality. «Virtual prototyping» is very fashionable in industry –and again: Its core is mathematics. More precise: The main components in virtual prototyping and in industrial mathematics are:

Mathematical Modelling and ***Scientific Computing***

By *modelling* we mean the transformation of real objects into mathematics by neglecting details, which are unimportant with respect to the questions we pose. (I know that there are constructivists, who see the world different: Our models are the only reality we have. They may be right –but it is hard to explain this philosophy to an industrialist, who wants to sell his «real» product.) Heinrich Hertz, the famous physicist, explained modelling in his «Introduction to Mechanics» from 1897: «We make images or symbols of objects in such a way that the logical consequences of the images are again images of the natural consequences of the objects.» His models should be «correct» (in the sense just defined above) and «logically admissible» (not leading to a contradiction). Models are not unique and it is the art of the scientist to choose the «cheapest» model. And it was clear to Hertz: The raw material of these images is, these models «are made of», mathematics. Of course: The better the raw material, the better the model. One needs good «pure» mathematics to do good modelling: «Engineering mathematics» is not enough. There are hierarchies of models: Very complicated ones, which could be very precise –if you have a very fast computer to evaluate them and if you have information about all the parameters involved. And simpler ones, deduced from the top model by simplifications of geometry or by taking asymptotic limits. This deduction is sometimes called modelling too and is a real mathematical art. Simple models are in general easier to evaluate (often even analytically, the only way of evaluation before the computer age), but provide you only with a qualitative prediction, seldom with reliable data. They are useful to explain certain phenomena, but industrialists often want the optimal design and the optimal parameters of a 3-dimensional technical system. To get these things, you need to evaluate the more complex model and we do that with aid of *scientific computing*. By this we mean to design an algorithm to solve the equation representing the model, to implement it in a proper computer system and to organise the data produced by the computer so that we are able to get the information we want. The Russian mathematician A.A. Samarskii called the whole procedure a «numerical experiment» and claimed it to be «a new scientific method, which determines the style of thinking of a modern scientist as well as the kind of problems he is able to attack.»

Model + **A**lgorithm + **P**rogramme = **MAP**

a map from the real into the virtual, mathematical world.

One word more about the relation between asymptotical and numerical methods (which sometimes seems to create a separation between mathematical cultures): I think that both methods supplement each other and that it is an essential feature of an industrial mathematician to be knowledgeable in both disciplines. Of course we have to choose the simplest model, which is sufficiently exact; and to get it, we need analytical (= asymptotic) methods. «What is complex is to make things simple» (Paul Valéry). But very often it is not possible to get sufficiently exact simple models: Then we need scientific computing. It happens quite often that in one region of the (spatial or time) domain of our problem a simpler model may be used, but in another we need to apply a more complex description: This leads to domain decomposition and to the task of matching different algorithms for the different models. The American applied mathematician J. Keller finished his invited lecture at ICIAM 95 in Hamburg by claiming that, while the first half of this century was dominated by asymptotical methods, the second by numerical methods, we are now entering a century dominated by the combination and interaction of these methods (I cite it from my memory, since he has not published the lecture). I will show some applications of these ideas in our context. I agree completely with Keller: Our chances are not «either-or», but «and».

If we want to see how this MAP works, we have to consider examples from the past; even more interesting, however, is the question about the future of this research: What are the most challenging problems to be solved during the first decades of the next century? Of course, I cannot do what David Hilbert did at the International Conference in Paris in 1900: He posed 23 yet unsolved problems mainly in pure mathematics, which influenced the mathematical research during the last century enormously. I even believe that nobody can do something similar today. What is possible are «Reflections on the Future of Mathematics», like Avner Friedman made at the last International Conference in Berlin, 1998 (see Berlin Intelligencer, Springer 1998). He described, in a spirit very near to my own, what mathematics can do in fields like material sciences, biology and multimedia. I shall try to give examples and a paradigmatic problem for the future in the fields of

I. Design of materials
II. Design of products
III. Simulation and Optimisation of production processes
IV. Discovery of Patterns in data for quality control, economic data prediction and medical diagnosis.

I hope to be able to show which kind of questions in these fields mathematics can help to answer. I could also name mathematical research domains, which occur most often in these projects –domains like multiscale analysis, free boundary value problems, computational fluid dynamics, stochastic systems, optimisation or signal– and image processing. And we could make a matrix about what is used where. But this bookkeeping does not tell us much –I believe examples to be more illustrative.

I. *Design of materials*

This is the main goal of material sciences. «Material sciences is concerned with properties and the use of materials. The objectives are the synthesis and manufacture of new materials, the modification of materials, the understanding and prediction of material properties, and the evolution and control of these properties over a time period. Until recently, material sciences was primarily an empirical study in metallurgy, ceramics and plastics. Today it is a vast growing body of knowledge based on physical sciences, engineering and mathematics.» (A. Friedman.)

One problem we encountered arises in dental technology: We need to compose a material for the filling of tooth cavities out of a certain kind of «cement» and small pieces of glass. The composite should have as little shrinkage and as much stability as possible: How many glass particles of which size should the mixture contain. The process is quite complicated, since the acidity in the cement triggers a kind of sintering in the glass.

How can we predict the behaviour of such a composite, not to talk about finding an optimal design? We can model the material to be linear viscoelastic with aging and isotropic shrinkage characterized by different parameters in glass and cement. We get integrodifferential equations with spatially very inhomogeneous coefficients: Could we really think of FEM approximations, resolving the microstructure of the composite? Of course we can't. This is a point where multiscale analysis comes in: If the microstructure is very fine compared to the total dimensions (the size of the hole), the details will not play a big role and averaged quantities will be sufficient. Mathematically we get them in a limiting process called homogenisation, when the relation of the microscopic to the macroscopic scales tends to zero (see for example: E. Sánchez-Palencia: *Non-homogeneous Media and Vibration Theory*, Springer 1980 or V.V. Jikov, S. Kozlov, O. Oleinik: *Homogenization of Differential Operators and Integral Functionals*, Springer 1994). Homogenisation needs very tricky mathematics, new concepts of convergence; but every new problem poses new tasks –here, to include aging and shrinkage. If this is done, we may talk about the optimal volume fraction of glass, etc. Much depends on a good model –dental laboratories often have not enough facilities to measure the material properties.

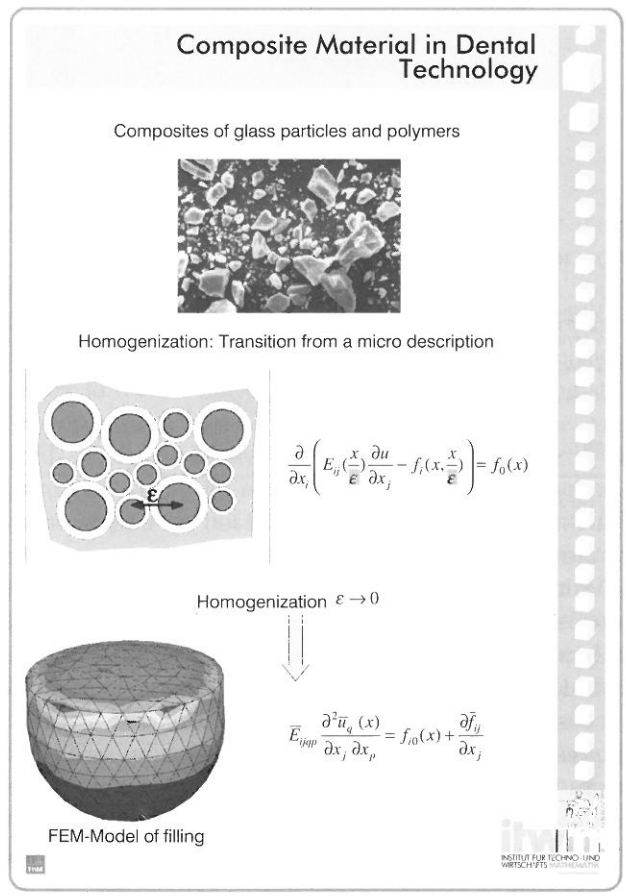


Figure 1: Composite material in dental technology.

So much about a problem we are presently working on. As a big challenge for the future research in this area I consider «fatigue life analysis». During a periodic or irregular loading material dissipates energy, which leads to internal fractures and finally to a local break down. The main interest for fatigue life analysis is shown by car industry: How long will critical safety components of a car survive given a «normal» driver on a «normal» road? Damage events are characterised by hysteresis loops in the stress-strain relation; this is fine for one-dimensional strains and stresses and leads to a method called rainflow counting (see e.g. A. Beste, M. Brokate, K. Dressler: *Kann man berechnen, wie lange ein Auto hält*, in «Mathematik in der Praxis», pp. 3-24, Springer 1995).

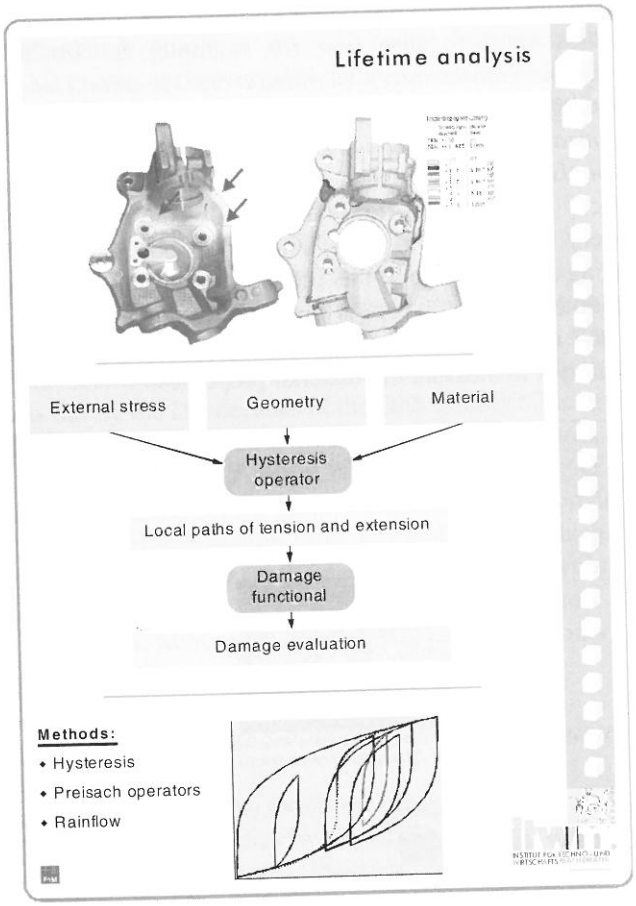


Figure 2: Service life time.

For more-dimensional situations one needs new concepts –the research started at TecMath and led to practically very successful tools provided now by an off spring, LMS Durability at Kaiserslautern and used in almost any car company in the world. It is good, but not perfect –and it has to be combined with other research branches like fracture analysis. Cracks show a very complex, a «fractal» geometry and the questions, where they initiate, how they evolve and when they branch into several cracks are still being researched. I don't know about many interactions between fatigue life and fracture analysis. But the big challenge in this area is again to design materials with a maximal duration under given loading. These may again be composite materials and we get a high dimensional parameter manifold, on which we have to find the optimum. Reliable tools to predict the life time of materials and to find an optimal design –a big task not only for mathematicians, but for teams of scientists including mathematicians.

A wide research field inside the material sciences is related to polymers. Polymers in dilute solutions give rise to very high-dimensional parabolic equations (Fokker-Planck equations), where the dimension is related to the «length» of a molecule and to its linear, branched or network structure. As financial mathematics (see IV). has posed new problems how to compute 2000-dimensional integrals, polymers pose problems how to compute solutions of say 100-dimensional parabolic equations. And as in finance mathematics we have to develop numerical methods based on pseudo random numbers (and therefore on number theory) to solve these problems (see: J.F. Traub, A.G. Werschulz: *Complexity and Information*, Cambridge Univ. Press 1998).

If we have a «pure» polymeric substance, we have to study its transition from the fluid to the solid form, which is a kind of crystallisation. This process poses exciting mathematical (mainly stochastical) problems –there is an ECMI «special interest group» based in Milano dealing with this topic (see e.g. V. Capasso *et al*: *Stochastic modelling and statistics of polymer crystallization processes*, Surveys on Math. for Industry, vol. 6, n. 2, 1996).

Let me, last not least in this chapter, mention that even classical problems from metallurgy pose new and exciting tasks for mathematicians. Consider casting processes in steel production plants or in «normal foundries». In continuous casting for the production of rolled steel, one has to control the solidification front by a proper cooling; this leads to an inverse, i.e. ill-posed problem, where the «free» boundary, the solidification front is prescribed and the cooling conditions at a fixed boundary have to be determined. «Inverse problems» is, like optimisation and many others, a very important subject, «transversal» to our classification of problems; in ECMI it is mainly represented in Linz (see H.W. Engl, M. Hanke, A. Neubauer: *Regularization of Inverse Problems*, Kluwer 1996).

In my institute we cooperate with the software company MAGMA, which offers a software tool to simulate the casting of iron in complex moulds. The package, while providing excellent simulations and therefore being internationally top, has a draw-back: on normal work stations it may take weeks to get one simulation finished. This defines another task for mathematicians: Find parallel algorithms for work-stations, that do the same job in one night.

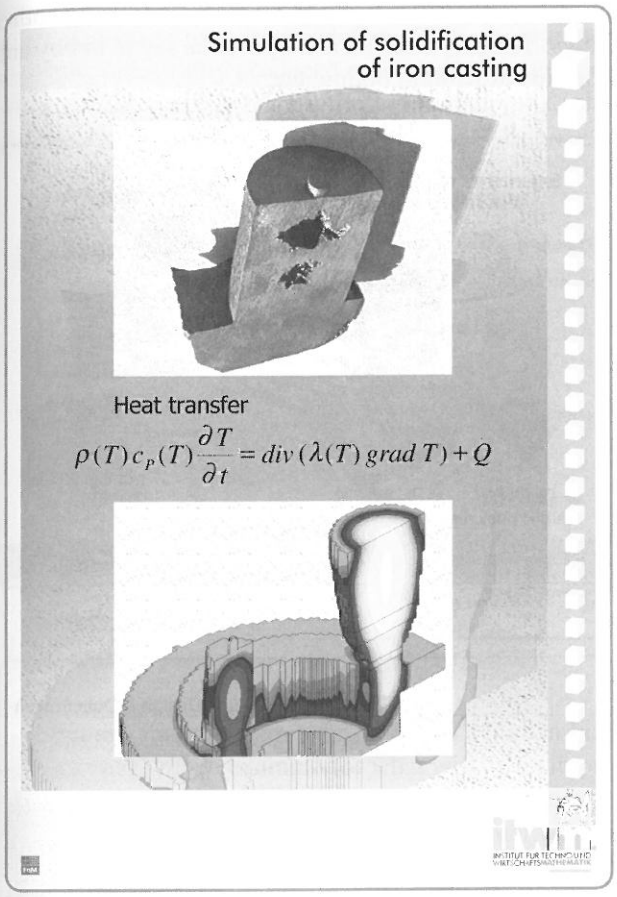


Figure 3: Solidification in iron casting.

You see: Not only the so-called high technology needs mathematics –the basic technologies do it as well. Especially if they try to approach an optimum: Maximal quality at minimal costs (and minimal time). And since even small and mean size enterprises today are forced to do that, the role of mathematics (as I see it) becomes more and more crucial.

II. *Design of products*

Of course, the border line between the design of materials and the design of products is as «fuzzy» as the border line between products and production processes: The material is part of the product –and the product depends on the production process. But, in the virtual world of simulation, you may first think of an optimal product, where materials and production processes are virtual; then, in a second step, you may try to realize appropriate materials and processes. In this sense, the design of optimal products is the key task. Of course, there are many examples, where mathematics has helped in the design of products and I will only choose some of our experience, which I consider to be quite surprising.

With the first problem –the optimal design of baby diapers– we have struggled for quite a while. Diapers consist of cellulose and absorbing grains; the transport of liquid in diapers is modelled by Darcy's law supplemented by ordinary differential equations describing the absorption process. The permeability, however, depends strongly on the amount of transported or absorbed liquid –the problem is ugly nonlinear. The model has to be checked to be «logically correct» (in the sense of Hertz): We have to show that it defines a solution uniquely. Then we have to develop numerical algorithms able to deal with these nonlinearities; and finally we have to define the proper material (distribution of the granulate, kind of cellulose) in order to get optimal properties of the diaper (minimal surface humidity f.e.). Why is this a never ending story? Told in details it would show many quite typical difficulties for the interaction of industry with mathematics. Often, mathematics is behind the practical needs –sometimes it is ahead. It took us almost ten years until industry believed that our striving for optimality was justified. Even mathematics benefits sometimes from industrial competition.



$$\frac{\partial u}{\partial t} = \nabla\cdot(\lambda(v)D(\tilde{u})\nabla\tilde{u}) - A(u,v)$$

$$\frac{\partial v}{\partial t} = A(u,v)$$

Figure 4: The diaper problem.

Flow in porous media, especially in woven or non-woven fabrics, is very common in industry: Think about the flow of oil in filters, of water in the felts of a paper machine, of sweat in shirts, of resin through a periodic matrix to form a composite, of blood in measuring stripes, etc. Darcy's law is not always correct –one may need Darcy with memory or Darcy with inertia or Brinkman's law; the mathematical theory of homogenisation helps a lot, but here again (as in I.) a lot of modelling and computing has to be done (see e.g. U. Hornung, ed: *Homogenization and Porous Media*, Springer 1997). Think only about the problem that a baby in fact deforms the diaper. A simple experiment still to be simulated is: To press out a sponge! Do you see that mathematics literally is everywhere?

I want to show a second example, from Tecmath. It originated about 12 years ago, when, after my talk at AUDI about the «possibilities of mathematics in car industry», a responsible engineer formulated the need of a computer model of a human being, including a detailed skin model and including postural comfort. If you

scan a surface of a human body you get a cloud of millions of points; if you want to simulate a change of the posture in a computer, you cannot do it operating on this cloud. At least not directly: You need a data reduction –and this can only be done by finding a «lower dimensional basis» i.e. typical modes of posture; these modes make a data reduction to less than 100 parameters possible.

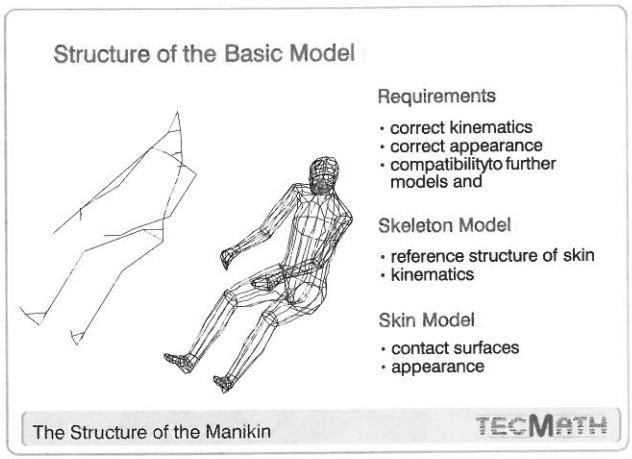


Figure 5: The basic modes: RAMSIS.

RAMSIS was developed using careful anthropometric measurements. If this reduction is done, one may define comfort functions (depending on these parameters) and find postures maximising the comfort.

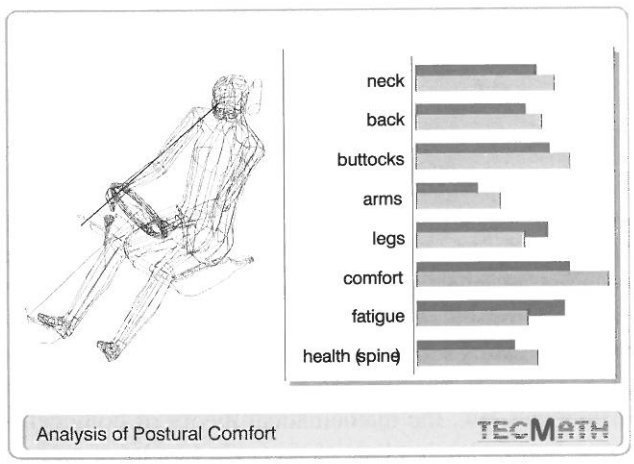


Figure 6: Postural comfort.

Of course, it is a multicriterial optimisation and you need tricky algorithms (and representation of surfaces) to do it properly. Then, you may use it to validate the ergonomy of car seat: What does the driver see, what can he reach –even, how is the secular change of the data of a typical driver? How are belts operating– and maybe even airbags (if our institute will be able to simulate the expansion of an airbag –see III).

The tool is now operating in almost every car company in the world. But not only that: To have a good model for the human body is useful for many other things: To study characteristic postures say in hobby and working, at fishing, biking, etc.

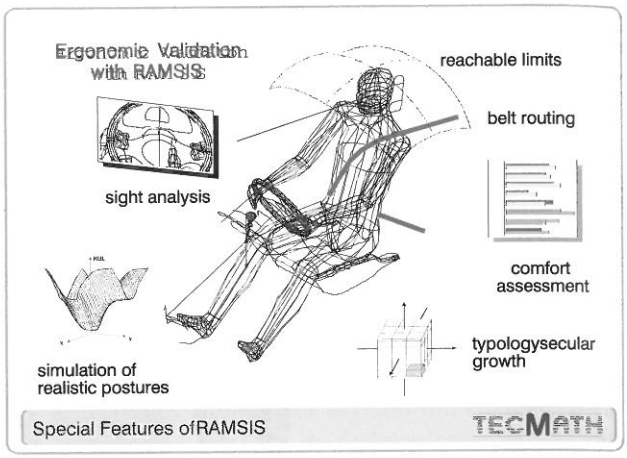


Figure 7: Features of RAMSIS.

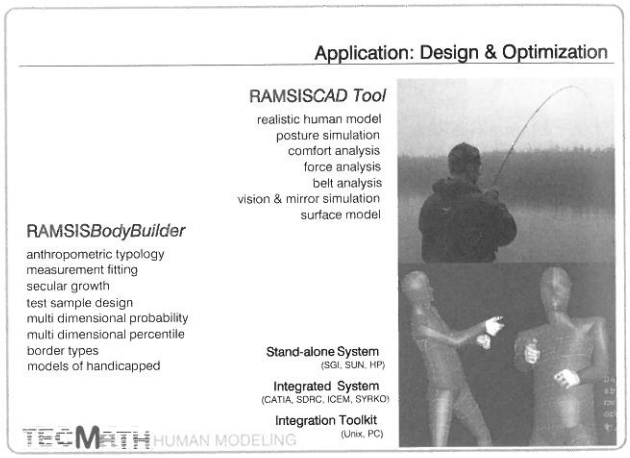


Figure 8: Applications.

A new and very profitable application is in tailoring: The body is scanned (the point cloud is produced), the scan is projected to the manifold of RAMSIS modes, clothes are virtually put on the body, they are designed and inspected in different postures, the patterns for tai-

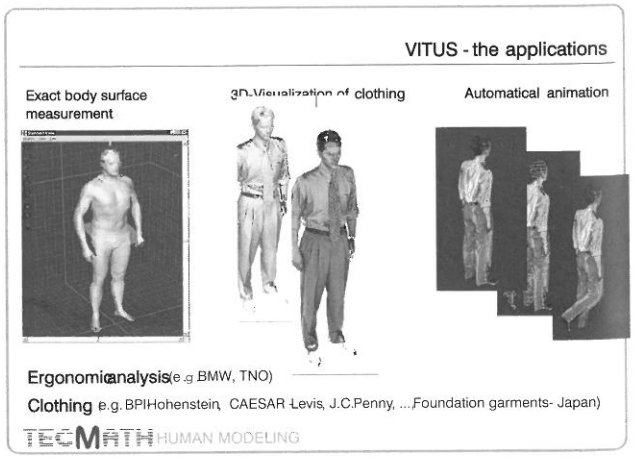


Figure 9: Made-to-mesurclothing.

loring are designed by the computer and the clothes are produced: made-to-measure clothes, 20 % more expensive than industrially produced ones. And of course possibilities for extensions are infinite, for example to shoe making.

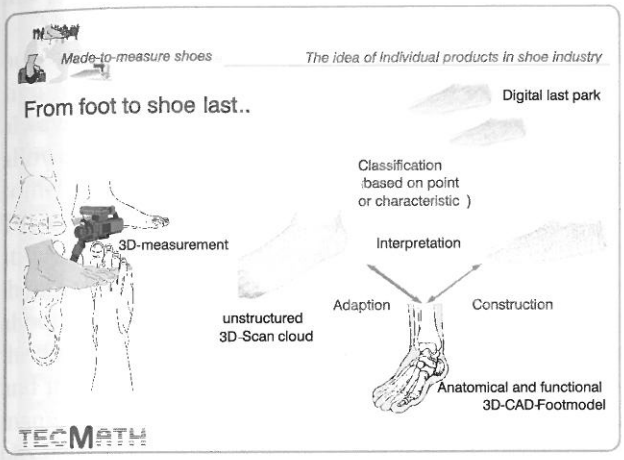


Figure 10: Made-to measure shoes.

I believe that this computer model of a human body is just a very typical example for what will be an important task during the next century: Individual products, designed according to the needs of individual persons (think of children or old people) and producible at costs not much higher than mass ready-made products. Of course the geometric shape of the human body is important, but by far it is not everything: Operating the products and the functioning of the product may be adapted to the individual capacities of the user.

A field, where mathematics has become very popular during the last 10 years, finance mathematics, seems to have similar goals: To design financial products like options, block of shares according to the wishes of the customer. Mathematics here is different: Time series analysis, theory of martingales, free boundary value problems for Black Sholes equations. Mathematical chairs for and books on this subject shoot up at present, new study programmes are created and high salaries in banks offered to our graduates. It is wonderful that quite advanced mathematics found a new, prosperous field of application; but I believe that quite soon we will reach a more modest saturation level. Altogether, the design of individual products (who have to be produced nearby, not at far east) may essentially contribute to the bright future of mathematics (see e.g. P. Wilmott, J. Dewynne, S. Howison: *Option Pricing*, Oxford 1993).

III. *Simulation and optimisation of production processes*

When the product is designed, it has to be produced. I said already that design and production are strongly correlated. And, of course, the machines for production are again products of another company, products, which should be designed optimally with respect to their task, to produce these products. To improve the quality of paper is a task for the paper manufacturer as well as for the paper machine manufacturer.

I want to take again examples from our own experience: from paper and from glass industry.

If we want to model complex production processes, we have to realise that again we have a hierarchy of models. Look for example at a paper machine:

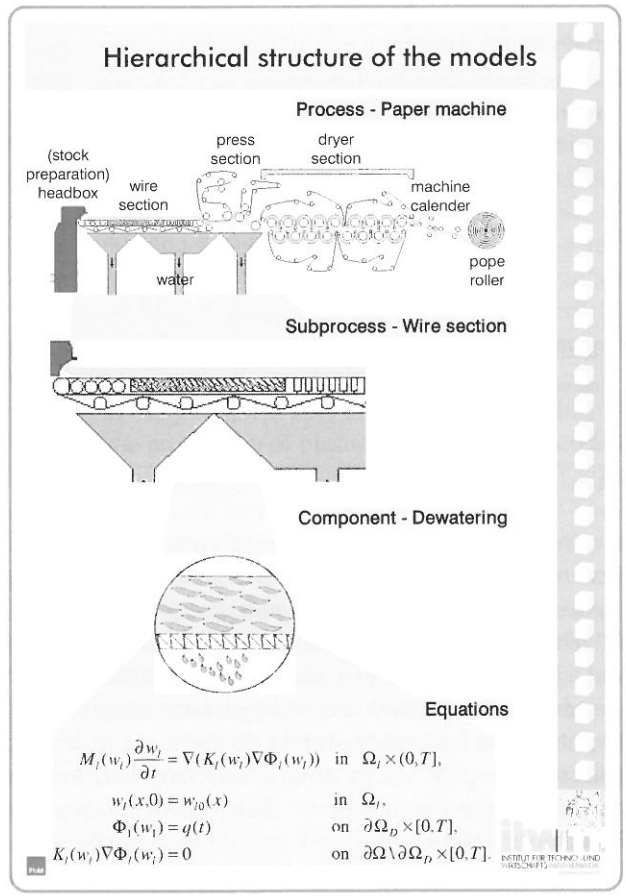


Figure 11: Hierarchical structure of models for a paper machine.

You may have the whole process in your focus –we call that again the macroscopic view. Then you may consider the sub-processes just as input-output systems with capacities and buffers in between, forming a network with stochastic perturbations, etc. Stochastic manufacturing systems may be considered as a network of queues and simulated as such; under the so-called heavy traffic assumption (the buffers are never empty), a diffusion approximation for the network may be derived, which has as many variables as the system buffers. Again we end up with 50-dimensional diffusion equations –an exciting task and similar to the polymer problem of section I.

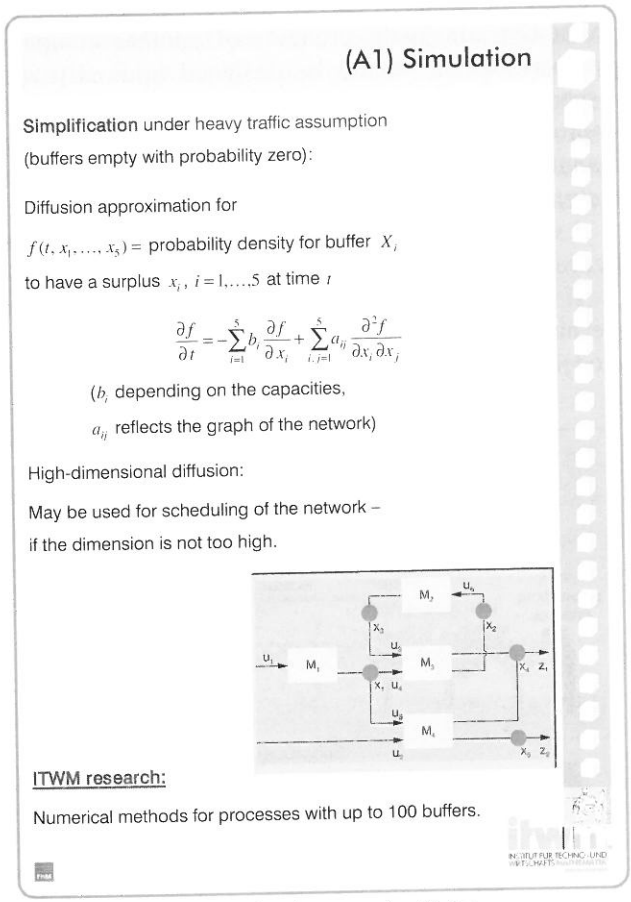


Figure 12: Simulation of manufacturing systems by diffusion.

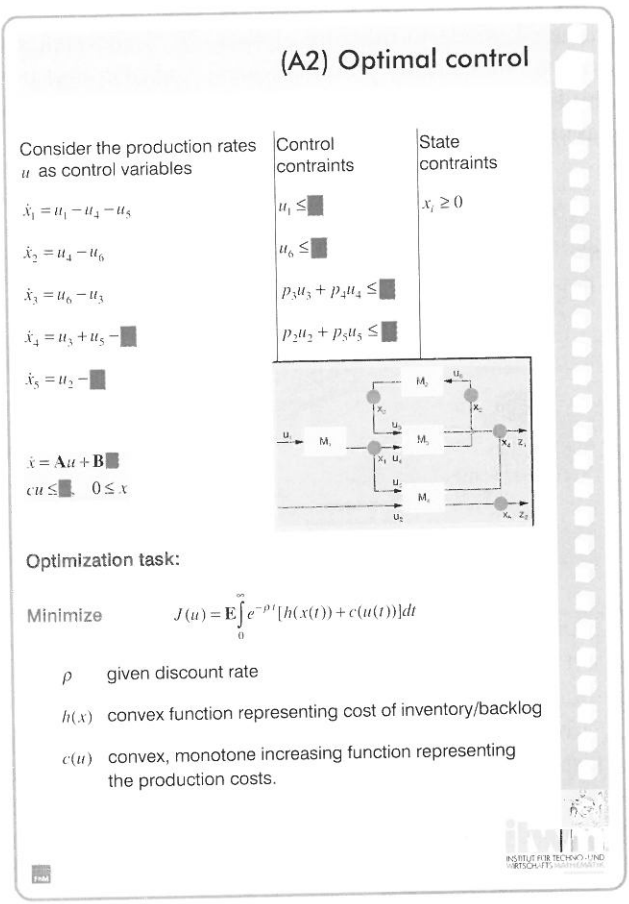


Figure 13: Optimal control of a manufacturing system.

We may also think of an optimal control of the whole system, where we have only one control variable per input-output system (machine unit).

This leads to Hamilton-Jacobi-Bellman equations and a lot of tricky analysis and numerics. Even questions concerning the optimal time to buy a new machine could be integrated and management decision could be supported (see e.g. S.P. Sethi & Q. Zhang: *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser 1994).

On an intermediate level we look at the subsystems more closely; instead of modelling them as a black box, we try to model the flow of energy and material in order to derive a dynamical system with controls like input of material, temperature, etc. A lot is done on this level, which we call the mesoscopic view. New research turns into the direction of adaptive or learning systems, where unknown or changing parameters of the system are updated automatically. System theory, neuronal networks and fuzzy logic to integrate expert rules are modern key words in this field.

Finally we may look at the finest level, the physical processes occurring in a paper machine; we call it the microscopic view.

None of these processes –grinding, spreading, dewatering, pressing, drying, smoothing, coating and poping– is simple to model. In spite of the fact that quite a lot of work is done, for example at the ECMI centre in Finland at Jyväskylä, there is even still more to do: Multiphase flows with free boundaries, sedimentation, flow in deformed porous media, again free boundary value problems, thin layer approximation... If you really want to understand the task, you have to look at a paper-mill, a huge monster, which you want to simulate.

What is really needed is a proper combination of all these hierarchical levels in order to get a *tool for diag-*

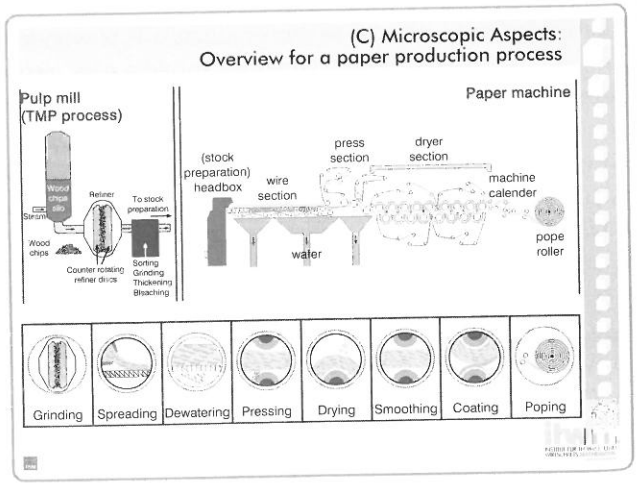


Figure 14: Microscopic aspects for a paper machine.

nosis: How can you influence the quality of the final product, how can you avoid failures like tearing. Risk, or better diagnosis tools for the total system: That is what industry is looking for, not only in paper-mills or printing machines or polymer fibre plants –even in systems like planes or fast trains, etc.

Much of the work industry or research institutes like ours are doing in this field is classified. It would be a bad sign for the commercial value of our work, if it would always be pre-competitive. Sometimes we are not even allowed to name the company we are cooperating with, sometimes only the tricky details are kept secret. Therefore I select an example, a cooperation of our institute with a company, which runs already over more than 3 years and has been so successful that we may tell about the cooperation –maybe not the latest, but at least the fundamental ideas. The companies name is Schott and it is one of the best known glass producers in Germany; and the problem is old –so old that the German physicist Fraunhofer, who gave the name to the research association we are related to, already struggled with it 200 years ago: During the cooling of glass, heat is to an essential part transported by radiation; if one wants to control the cooling process (in order to avoid cracks), one has to predict the temperature evolution and therefore the radiation. Glass is a semitransparent medium:

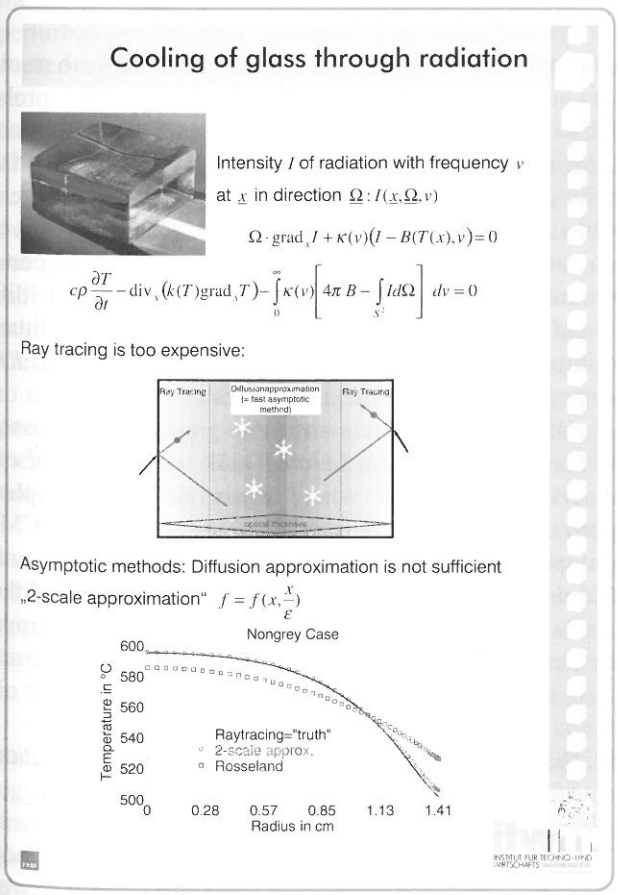


Figure 15: Cooling of glass through radiation.

Each point in the glass absorbs and emits rays at the same time. Ray tracing, relatively easy for surface-to-surface radiation, where the points in between do not influence anything, becomes very expensive, in realistic 3-dimensional configurations it is not feasible. Mathematically we have to solve a 6-dimensional integrodifferential equation, coupled with the heat equation.

There are many asymptotic methods, using the optical thickness as small parameter; the most often used Rosseland approximation is quick, but not sufficiently accurate. Combination of ray tracing, where it is necessary (i.e. at the boundary) and diffusion approximation, whenever it is possible, is in the spirit of combining analytical with numerical methods; a proper 2-scale Ansatz is a more elegant similar approach, which permits us now to do accurate 3-dimensional calculations in a reasonable time; this gives our partner a lead and us (besides the money) scientific success. Moreover, radiation in semitransparent medium happens not only in ordinary glass production; even in medicine for example the human body is semitransparent for certain rays –a fact, which is used in radiation therapy. Many more applications are at hand (see e.g. F. Zingsheim: *Numerical Solution Methods for Radiative Transfer in Semitransparent Media,* PhD-Thesis, Kaiserslautern 1999).

A completely different kind of mathematical problem is posed by the production of plastic fibres. Here, air streams flow through an ensemble of very thin fibres, through a filament. How does the air flow interact with the fibres? These fibres are almost one-dimensional objects in a 3-dimensional flow: An Euler model would not even notice them. For wet spinning, where the fibres are drawn by means of rotating rollers, nice work has been done at another ECMI centre, at OCIAM in Oxford.

For quite a while we have worked on air-spinning processes, and tried to find numerical schemes computing the flow and the motions of the fibres at the same time. The main problem of course is the complex geometry of the flow region: The filament forms a moving, complex, interior boundary.

Similar problems appear if you want to predict the behaviour of sails or flags in wind or other similar industrial processes. To compute flows in complex regions with quickly moving (free) boundaries we would prefer to use grid free methods. There are two main ideas to develop those methods.

1. Particle methods like «Smoothed Particle Hydrodynamics». Here «particles», mathematically understood as δ-measures, carry information like velocity, energy, etc.; they move with their velocities, but change these velocities, energies, etc. according to ordinary differential equations (with the number of variables corresponding to the number of particles) in such a way that the particle number density, the particle momentum and energy densities approximate air density, momentum and energy, which solve the Euler or Navier-Stokes equations.

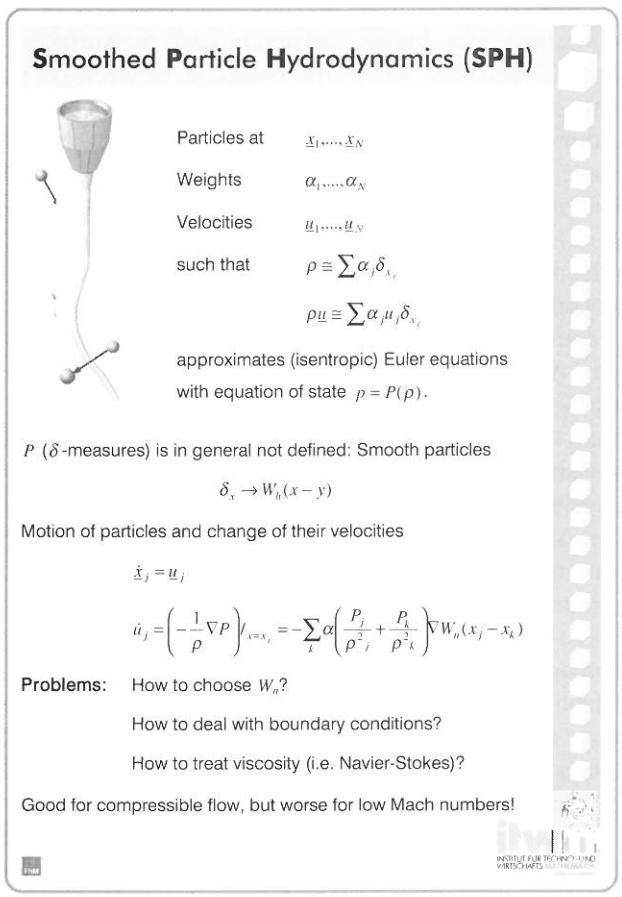


Figure 16: Smoothed particle hydrodynamics.

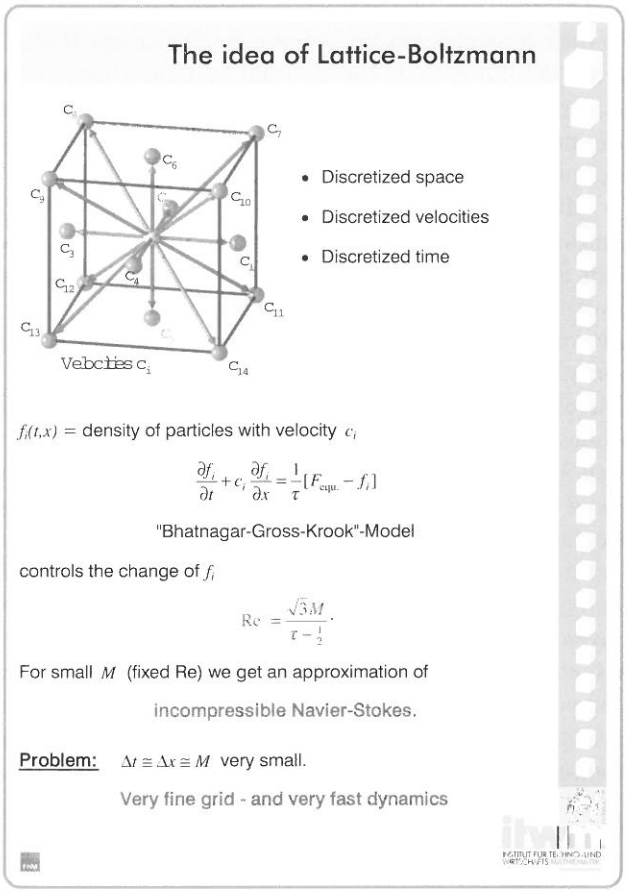


Figure 17: The idea of Lattice Boltzmann.

The method, very popular in astrophysics, is not very accurate, but working even for difficult problems –as long as the flow does not get incompressible, i.e. for sufficiently large Mach numbers. Incompressible Navier-Stokes equations should not be attacked with SPH. There is an alternative method called.

2. Lattice Boltzmann Method (LB). It is by far not grid free. Opposite: As the name says, it works on a lattice in space and we consider particles flowing through the lattice but only with velocities built-in in the lattice.

And we use a kinetic Boltzmann equation like description, which considers the evolution of densities $f(t,x,v)$ in position –and velocity space. Normal density, momentum and energy are moments of f with respect to v. Everything is now discrete: time, position and velocity; the kinetic equation contains an equilibrium F_{equ} and a relaxation parameter –both can be chosen such that the moments approximate incompressible Navier-Stokes equations (Shiyi Chen, G.D. Doolen: *Lattice Boltzmann Methods for Fluid Flows*, Annual Rev. Fluid Mechanics 30, pp. 329-364, 1998).

But again: Δt and even Δx have to be chosen small, when the Mach number of the flow is small –and since we want to describe incompressible gas now, the Mach number is small. So, why is it nevertheless a good idea? The evolution of f is extremely fast; therefore we can use very fine lattices, millions of cells –and a very fine lattice is of course as good as no grid at all. It is able to resolve complicated structures. But still, there are many problems to solve and ITWM has a group of 5 scientists working on the feasibility of these methods for industrial flow problems. The combination of codes for compressible and incompressible flows is necessary, since this is not a real alternative: Industrial flows may well be compressible and incompressible at the same time, but in different regions. There is hope to get progress in this direction using the idea of «kinetic schemes» (M. Junk: *Kinetic Schemes*, PhD-Thesis, Kaiserslautern 1997).

Of course, there are many more production processes in completely different fields: Think for example of oil reservoir modelling, where again flow in complex porous media has to be modelled and simulated. ECMI groups in the north of Europe (Oslo) and in the south (Sardegna) provide European expertise. Or think of the production of a good Italian espresso: The ECMI group in Florence has done marvellous work –you may even taste it.

IV. *Discovery of patterns for quality control, prediction of finance data and diagnosis of medical data*

This is an area where a very efficient tool is already available: our brain. Evolution has taught it to discover patterns in audible or visual signals and to classify them.

But to teach computers to have the same capacity is still out of scope.

Think only of walking through a pedestrial area in your home town, obviously checking hundreds of faces in seconds and discovering that a certain face belongs to a person you knew very well but haven't seen since 10 years. The brain realises the similarity between the face we just see and another, younger face, which we had seen long time ago –one of the maybe 1000 faces it has stored since then. How long will mathematics and/or computer science have to make intense research until a computer can do the same?

More easier and an example from our projects: Look at a veneer and try to classify it according to its quality; that is what men and women do in a furniture manufactory every day at a high speed. But how can we tell a computer what is quality and to determine it?

The main problem is the amount of data: The image given by a veneer consists for example of 256×256 pixels with 256 grey values at each pixel. How should we classify points in 256^3-dimensional spaces? The first problem is always data reduction –as it was with RAMSIS, the model of a human body: We need to find «modes», an underlying structure, which allows us to describe a single image by say 20 (instead of 256^3) parameters. In the veneer case the best way to do it was to ask the biologists: A tree trunk is a perturbed cone and the annual rings as well; a veneer consists therefore of perturbations of conic sections. The perturbations are caused by the axis of the tree being a curve (and not a straight line) and by knots. Using these informations we may reduce the number of parameters in fact to about 25. The next problem is an inverse problem again: Given a real veneer and our 25-dimensional model class –find the 25 parameter values giving the best approximation of our real object. This problem can be solved, but we are still far too slow in comparison to what industry is asking for. The last problem would then be to classify these points on our 25-dimensional manifold according to the quality; this task can be learnt from experienced persons by training a neuronal network according to their judgements. This would be quick and at least as reliable as human beings.

The steps above are typical: A model class, a set of modes, has to be constructed, which allows the reduction of data to a feasible amount; the identification of the best approximation in the model class to a given real object has to be performed; the classification rules for elements in the model class must be learnt.

We will find this job pattern in any pattern recognition job again; there seem to be ideas in neurophysiology, how our brain does it –but besides this very modest transfer of ideas, which is intimated in notions as «neuronal networks», not much is yet done. To increase this transfer is, in my view, one of the big tasks for future research.

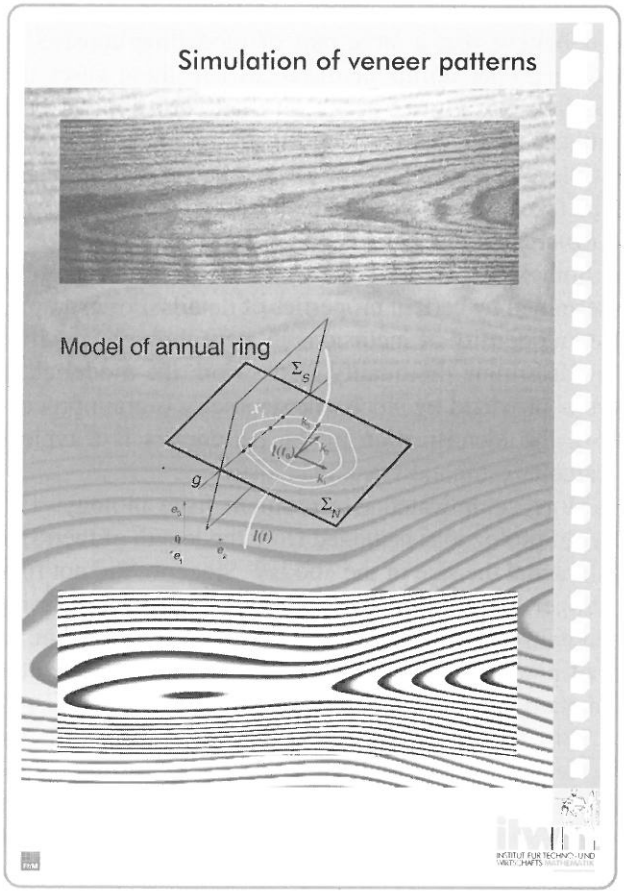
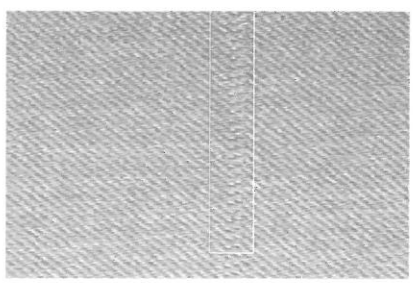
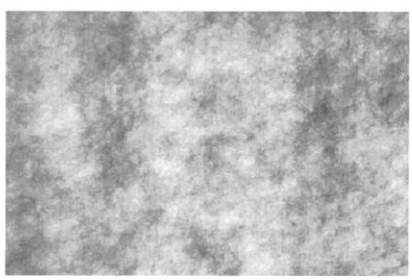


Figure 18: Simulation of veneer patterns.

Other problems of quality control are similar and conceptually even simpler: Find defects in a woven fabric (for example knitwear), judge the quality of a non woven fabric, given as a properly chosen distance to uniformity. («Properly chosen» means to find a metric measuring the distance between two images in such a way



Textile type: Flat fabric



Non woven fabric

Figure 19a: Inspection of fabrics.

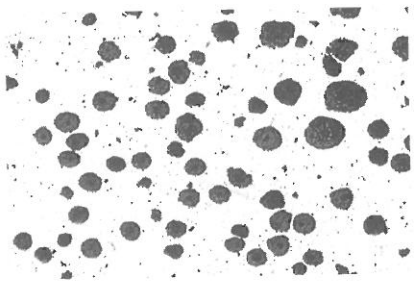
that the judgement of experts is reflected in this distance; I believe that a large part of modelling consists in finding proper norms or distances.) In these cases the model class is given by wavelet filters or a similar local multiscale analysis, the identification is done through a fast wavelet decomposition or another pyramidal algorithm (see e.g. J. Weickert: *Anisotropic Diffusion in Image Processing*, Teubner 1998).

Sometimes, the quality of a «surface» or an image is determined by certain properties of details. For example, size and density of inclusions in a cut through cast iron may determine the quality of the iron; the model class here is provided by stochastic geometry («grain processes»), the identification of the parameters is a typical parameter estimation task in statistics.

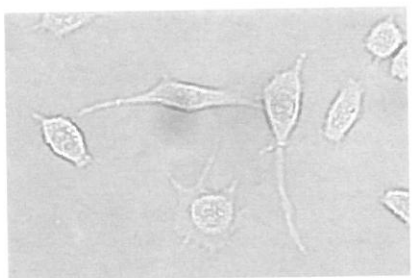
More complicated is a problem from biology: The state of protozoons is judged from the length of their flagellum and the size of the «body». Since we did not find a proper model class we tried to study the «intensity mountain», given by the image of a single bacterium. A lot of nonlinear smoothing has to be done to arrive at a mountain which we can deal with.

Now we have to decide: Is this «side mountain» on the left a flagellum? How do we determine the «ground plan» and by that the size of the mountain area? Even tricky methods as «Active Contour» or «Snake» methods do not work without errors. People in biological laboratories do the job with a lot of stress, but without many errors.

Since we are already near to medicine, we may switch to this field, which I believe provides enormous challenges for the future. The situation in medicine is new with respect to the amount of available data: The use of data bases allows to store the long-term observation data of patients.



Modular cast iron



Size and shape of protozoon's

Figure 19b: Inspection of surfaces.

Consider a human body as an input-output system, for which we have collected many input and output data as medication, food, etc. and temperature, heart beat, blood pressure, etc. respectively. From a careful long term observation we may have gathered hundreds of thousands of data. Now we choose a model class for this input-to-output mapping: For example linear con-

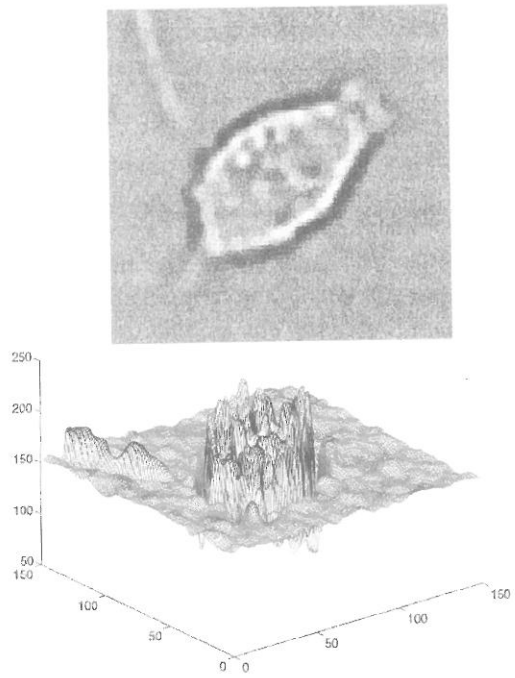


Figure 19c: Size and shape of bacteria.

trol systems or neuronal networks. To identify an individual patient in this model class means the classical task of identification of systems or training of networks. Finally we have to classify the «identified» patients according to their state of health; here we may use experiences from experts, which can be often formulated in a fuzzy setting.

An example from our institute is the task to support the diagnosis from a thermogram. In thermography one measures the surface temperatures of a patient at ca. 100 points; the measurements are done twice, the second ones after cooling down the body a little bit. In this way one gets the system response to a cold stimulus; we see the temperature regulation of the body. A thermogram consists therefore of 2 times 100 data. And we repeat this thermogram and do it with many patients. This data base, together with expert rules, provide an input for a learning system, which at the end is able to make a diagnosis incorporating the knowledge stored in the data and in the expert rules.

The method is now used in oncology and seems to be very promising.

Another medical project we are involved in concerns the evaluation of electrocardiogram data. There again is a lot of expert knowledge, but are there more than fuzzy rules?

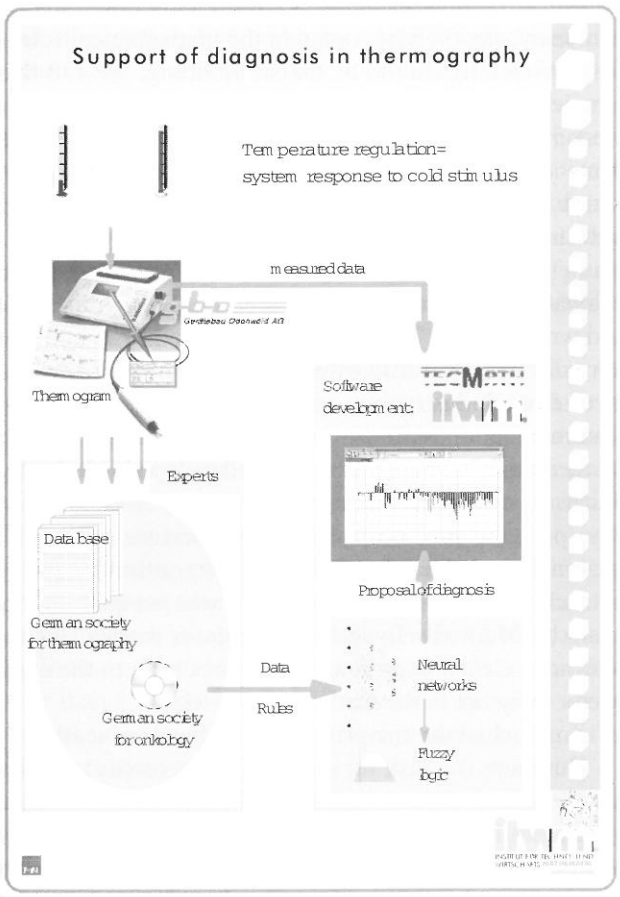


Figure 20: Support of diagnosis in thermography.

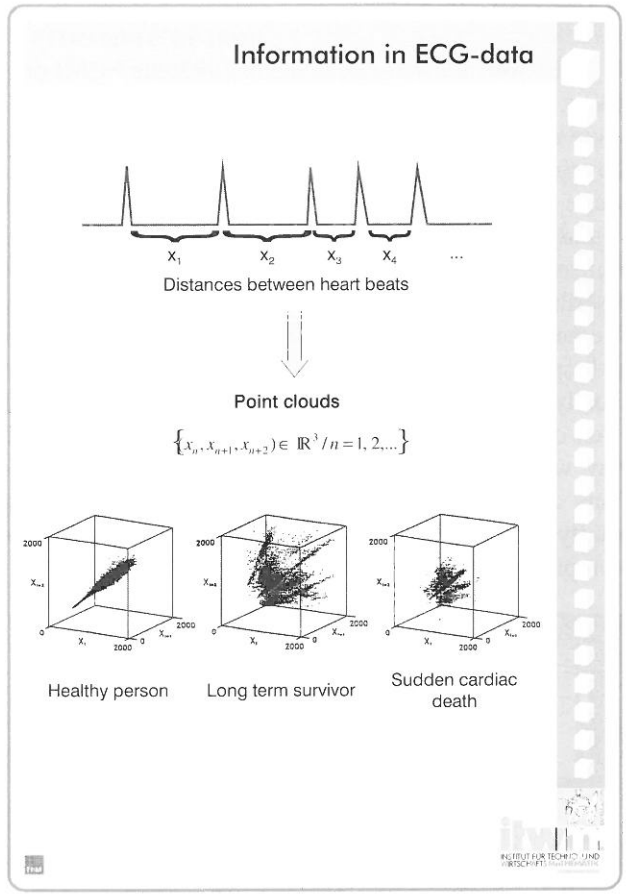


Figure 21: Information in ECG data.

It seems general belief that the series of distances of heart beats contains most of the information; but how do we extract this information from the series x_1, x_2... of beat length? It would be optimal to find a characteristic number assigned to the series, which tells us about the health or illness of the patient. A regular beat would mean $x_1 = x_2 = ...$; but is regularity a good sign?

More information is visible, if we consider clouds of points in $\mathbb{R}^2$ or $\mathbb{R}^3$, which are made of 2 or 3 consecutive beat length: $\{(x_n, x_{n+1}, x_{n+2}) \in \mathbb{R}^3 \mid n = 1, 2...\}$. The cloud degenerates to a point on the diagonal for complete regularity and it will be a «slim club» along this diagonal, if the beat changes slowly. If the club is too slim, experts consider it as a bad sign.

But what are the proper characteristics? Again, we have much long term data, for which we know the fate of the patient. Therefore, if we would have a suggestion, we can check, how good it is. Astrophysicists have proposed a number –they used their experience in dealing with point clouds; and it works rather well, but optimal? One may again think of many model classes –those from stochastic geometry (my favourite) or from fractal geometry (there are many attempts in this direction –they speak of «chaos» in heart beat; but it seems that they are not better than the astrophysicists) or something completely new. A very exciting research field: We have to find order, structure in seeming by unstructured patterns: Who else if not mathematicians, whose profession it is to find ordering structures, should try to solve those problems.

Another field of finding patterns in data belongs to finance and commerce, which I already mentioned in II. The data are provided by the time evolution of the prices of financial assets; the models are taken from stochastics. The evaluation of the models, the answer to questions about arbitrage opportunities or about option pricing leads to new concepts and theories in probability theory, functional analysis, partial differential equations, etc.

Even completely new questions, maybe from other fields, where actions of many human beings are decisive, may come up: Nobody would have predicted the boom of finance mathematics 25 years ago! Maybe we have «sociological mathematics» after 25 years from now? The Austrian novelist Robert Musil in his «Man without Properties» (chapter 103) talks about thermodynamics of moral: In societies, the stochastic behaviour of the individuals, in spite of being very important for the individuals themselves, does not contribute very much to the evolution of the ideas in the society. Individuals as Brownian particles, the society as a kind of gas of particles: Are more mathematical models in sociology possible and even useful? Finance mathematics seems only to be a start in this direction.

B) The task ahead

A few months before he died in 1993, the ECMI president, Henrik Martens from Trondheim in Norway, wrote an article with just this title «The task ahead» (see Surveys on Mathematics for Industry, vol. 4, n. 1, 1994), where he focussed on a problem he considered to be of primary interest: How do we prepare our students for a profession as an «industrial mathematician»; I tried to describe the corresponding «discipline» in this article. The consequence of this new discipline is a profession with special «attitudes, habits and skills, which are not usually emphasised in the mathematical community, and often quite foreign to the research mathematicians. We may well see the emergence of industrial mathematicians as members of a new technological profession, solidly rooted in the mathematical sciences, but with its own professional profile and goals» (H. Martens).



Figure 22: The emergence of an industrial mathematician.

These attitudes, habits and skills of a mathematician *in* industry (not necessarily an «industrial mathematician», who makes mathematics *for* industry) were described on a solid base of success stories in US Industry by the SIAM Report on Mathematics in Industry, 1995. I just cite a few sentences from the SIAM NEWS article about this report; it runs under the headline «Change is important» and evaluates the responses of mathematicians in industry: «More than 30 % of PhD respondents, however, assessed their preparation (at the university) as less than good in some of the very areas identified as crucial for success in a non-academic setting: *computing, the flexibility to work effectively on a range of varied problems, breadth of scientific background, teamwork and communication*. An overwhelming majority of mathematicians surveyed (90 %) indicated that change in the graduate education of mathematical scientists is important.» Even if we are willing to change: How should we educate our mathematics students (or even our engineering students during their mathematical education)? The key words are again «modelling» and «scientific computing». When the students, after the first two years, are «solidly rooted in the mathematical sciences», modelling should be taught by doing. We call that «modelling seminars» or «project seminars»: Students work in (if possible: interdisciplinary) groups of 5 students during one semester on a real world problem, which originates from our multiple contacts with industry. The problems are given in their original non-mathematical settings, they have to be modelled, to be evaluated with a computer and presented in a proper oral and written form. These seminars are the key features of our education and lead often to master theses with subjects from industry but supervision at the university. Communication skills, normally not a strong point of mathematicians, have to be better developed: Mathematicians have to communicate with people from industry, who pose the problem, with other scientists, even with «common people». The obligation for someone engaged in teaching mathematics to engineers was put quite simply again by Martens «To seek out whatever mathematics is relevant to technology, and make it available to the engineer». Why not to the mathematician too?

So much about the primary task ahead: education.

But there is a task for science (or university) politics too: Research in this interdisciplinary field, research, which «falls among all stools», must be academically rewarding. To develop and implement a tricky algorithm should be acknowledged as a scientific achievement completely equivalent to proving a new theorem. Do not misunderstand: Not every mathematician should work that way. But «it seems to me that you have to do very good theoretical mathematics in order to justify a total lack of interests in the problems of society» (Martens). We need pure mathematicians, we need them urgently –but only the very good ones. And we need more «bridge builders», people with knowledge of both sides, of mathematics and of industrial requests; we need them in industry, but also as professors at universities. Academia has to reconsider its criteria for the promotion of young mathematicians; if not, I do not think that modern society will stand in awe of such a situation for very long.

Finally there is a task for the responsible managers in industry and commerce: Not many of them have yet realised the power of mathematics. Whenever I visit a new company I experience an attitude of scepticism with respect to mathematicians. «How can these theoreticians help us in solving our practical problems?» The image of mathematics is often formed by school mathematics, and it appears therefore to be an ivory tower science. It is an advantage that we can show the benefits other companies have from mathematics; how could we earn several millions of Euro every year from industry, mainly from small and mean size enterprises, if these companies would not benefit from our work! Others do not trust in simulations at all –an experiment has to be real and not

virtual. And sometimes these sceptics are right: «Mathematical modellers look to an uncertain future» headlines the last issue of SIAM NEWS, December 1998. What it means is that complex models, like models for global climate, multiphase flow or urban traffic, carry some uncertainty, which may be quite prominent. If problems are inherently stochastic and/or highly nonlinear, we cannot make the models as correct as we like just by adding more terms. The request to find methods to quantify and predict this uncertainty may even reshape our scientific methods.

Even if the uncertainty is small, each model has its assumptions; a model is only correct (in the sense of Hertz) if these assumptions are justified. Scientists have to be honest in stressing these restrictions, especially if their results become relevant in political discussions. I think that most mathematicians are!

But, finally, many simulations are astonishingly correct. In most cases I have seen, the errors of simulation were much smaller and not larger than I had expected: Industrialists rarely ask for results with a precision better than 1 %; they are often content with a 5 % error, sometimes even with a qualitative understanding of a process.

And therefore, the managers should give a chance to simulation: If they really want to optimise a process or product (of course: in order to be better than the competitor), they have to develop a simulation tool timely. If the competitor has already succeeded in improving his product, it is too late to start with simulation.

If we are successful in achieving these goals, mathematics will not only be the key to key technologies, but one of the most important technologies itself.

Therefore we need

- open-minded mathematicians with a spirit of adventure,
- industrial managers with curiosity for and confidence in science and mathematics and will to invest in future,
- students well trained in modelling and scientific computing with ability to formulate and solve problems in a wide variety of areas, skills in communicating and the ability to work in interdisciplinary temas.

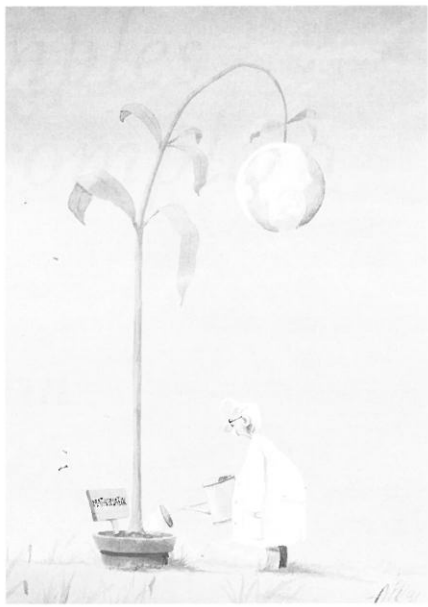


Figure 23.

Modelos simples de un mundo complejo

Jesús Mosterín

La realidad es endiabladamente complicada, llena de recovecos insondables y oscilaciones imprevisibles, inefable en su inagotable riqueza e inasible con nuestras redes conceptuales. Pretender abarcarla y conocerla en su exacto detalle es una meta utópica, que sólo conduce al mareo y la esterilidad intelectual. Lo único que nos resulta transparente y perfectamente comprensible es el mundo ficticio de la matemática, hechura y creación nuestra, que conocemos como si lo hubiéramos parido (pues, en efecto, lo hemos parido, aunque no sea más que con la cabeza, como Zeus a Palas Atenea). Por ello el método de la ciencia moderna, y en especial el de la física, pasa por la construcción de modelos matemáticos de los sistemas empíricos. Esos modelos sólo corresponden a la realidad que modelan con grandes dosis de aproximación, abstracción y simplificación. Pero nos permiten comprender una realidad de otro modo incomprensible, formular preguntas precisas, computar las respuestas utilizando toda la artillería del cálculo matemático, y finalmente obtener respuestas aproximadamente correctas.

Incluso las cuestiones más triviales, como «¿cuánto pesas?», requerirían respuestas de una complejidad infinita, si no las simplificásemos. Nuestro peso varía constantemente, cada fracción de segundo, cada vez que respiramos, comemos, bebemos, sudamos o excretamos, varía con cada paso que damos, con cada palabra que pronunciamos. Construyendo un modelo numérico groseramente simplificado, contestamos: «70 kilos».

Pensemos en la superficie de nuestro planeta. En una primera aproximación decimos que la Tierra es una esfera euclídea, con lo cual podemos calcular su superficie y volumen, conociendo su radio o su meridiano, podemos computar los ángulos de la navegación aérea y podemos representarla cartográficamente mediante proyecciones adecuadas. Pero, obviamente, la Tierra no es una esfera. En una segunda aproximación decimos que es una esfera achatada por los polos, lo cual tampoco es cierto, aunque ya se aproxima un poco más. Así, podemos ir definiendo una sucesión de figuras geométricas cada vez un poco más complejas y un poco más próximas a la realidad, pero siempre grotescamente alejadas de ella, si uno se pone a fijarse en los detalles. La forma real de la superficie terrestre incluye cada hoja de cada árbol y cada pelo de cada mamífero y cada espuma de cada ola. Esta forma varía constantemente con cada brisa que sopla, con cada automóvil que pasa, con cada gota de lluvia que cae. Ninguna geometría conocida ni imaginable podría dar cuenta cabal de forma tan prodigiosamente compleja. Ni falta que hace. Como ya sabía Borges, sólo la Tierra entera es un mapa completamente exacto de la Tierra. Cuando buscamos otro mapa, lo que pretendemos es obtener una representación simplificada y manejable, que podamos entender y que nos permita contestar con suficiente (pero no excesiva) precisión a las preguntas que nos hacemos; en definitiva, lo que buscamos es un modelo simple de este planeta complejo.

Si nos elevamos a la contemplación del universo en su conjunto, y el vértigo no nos deja atónitos, sólo mediante la teoría general de la relatividad podemos esperar describir algo tan inmenso. Sin embargo, esa teoría, aunque ella misma ya es de por sí una simplificación matemática, sigue siendo apabullantemente compleja (menos que la realidad, claro, pero todavía demasiado para nuestras limitadas entendederas). No sabemos cómo computar y cómo resolver ecuaciones en todos los casos que la teoría admite, sino sólo en los casos extraordinariamente simples de la métrica de Friedman-Robertson-Walker, que corresponden a un universo perfectamente isótropo y homogéneo. Y así, simplificando todavía más, suponemos que el universo es isótropo y homogéneo, aunque sabemos que en realidad no lo es, sino que más bien tiene estructura esponjosa formada por enormes vacíos rodeados de increíbles concentraciones de galaxias. Gracias a ello podemos hacer cosmología. Pero nuestros modelos cosmológicos son meros modelos matemáticos, simplificaciones drásticas aunque inevitables, y no reflejan la compleja realidad de la distribución de masas en el espacio-tiempo más que de un modo vagamente aproximado y *cum granu salis*.

Lo mismo ocurre con las otras ciencias. El mayor triunfo de la biología de nuestro siglo ha consistido en desentrañar el código genético, pero nuestro modelo del sistema genético simplifica considerablemente lo que de hecho ocurre en la célula. Las gramáticas generativas y

fonologías estructurales de la lingüística son modelos abstractos, simples e idealizados de la complejísima realidad del habla. Especialmente conspicuo es el carácter simplificador de los modelos de la teoría económica. De todos modos, nada de esto debe considerarse como una crítica. Más vale pájaro en mano que ciento volando. Más vale una aproximación computable que una verdad absoluta incomputable. Más vale el arduo progreso de la ciencia real que la fatua añoranza de un saber imposible.

Los conceptos científicos

Así como no se puede dibujar sin líneas, ni se puede pintar sin colores, tampoco se puede hablar ni pensar sin conceptos. Esto vale tanto para la vida cotidiana como para la actividad científica. De hecho, muchos de los conceptos científicos actuales provienen de conceptos cotidianos, aunque durante el viaje se han transformado, ganando sobre todo en precisión. Así, las nociones químicas de hierro (átomo con 26 protones en su núcleo) o de agua (H_2O) precisan nociones previas del lenguaje ordinario.

Es usual dividir los conceptos científicos en clasificatorios, comparativos y métricos.

Conceptos clasificatorios

Los conceptos clasificatorios atribuyen propiedades a los individuos del dominio que clasifican, de tal modo que a cada individuo se le atribuye una y sólo una de esas propiedades. Una clasificación de un dominio es un conjunto de conceptos clasificatorios que clasifican ese dominio, es decir, que lo dividen en clases disjuntas y conjuntamente exhaustivas. Esas clases (es decir, las extensiones de los conceptos clasificatorios en cuestión) constituyen una partición.

Una familia de conjuntos $\{B_1 \ldots B_n\}$ es una *partición* de un conjunto A si y sólo si (*i*) para cada i, j $(1 \leq i \neq j \leq n)$: $B_i \cap B_j = \varnothing$, y (*ii*) $B_1 \cup \ldots \cup B_n = A$.

Una relación binaria ~ entre objetos de un dominio A es una *relación de equivalencia* si y sólo si es reflexiva, simétrica y transitiva en ese dominio (es decir, si y sólo si para cada $x, y, z \in A$: (*i*) $x \sim x$; (*ii*) $x \sim y \Rightarrow y \sim x$; (*iii*) $x \sim y \wedge y \sim z \Rightarrow x \sim z$). Dada una relación de equivalencia en A, llamamos clase de equivalencia de un elemento $x \in A$, $[x]$, a la clase de todos los elementos de A que están relacionados con x en esa relación de equivalencia, $[x] = \{y \in A | \, y \sim x\}$.

Toda relación de equivalencia ~ sobre un dominio A induce una partición de ese dominio en clases de equivalencia, llamada el espacio cociente de A por la relación ~, y simbolizada como $A/\sim$. Este hecho se usa con frecuencia para clasificar un dominio mediante la previa introducción de una clase de equivalencia. Consideremos la siguiente relación de equivalencia $\sim_p$ sobre el dominio A de los átomos. Para cada átomo $x, y \in A$: $x \sim_p y \Leftrightarrow x$ tiene el mismo número de protones en su núcleo que y. La clase de equivalencia (respecto a esta relación) de un átomo determinado es el conjunto de todos los átomos que tienen su mismo número de protones en el núcleo, es decir, es un elemento químico. Así, el elemento químico carbono es la clase de todos los átomos que tienen 6 protones en su núcleo, el elemento químico nitrógeno es la clase de todos los átomos que tienen 7 protones en su núcleo, el elemento químico oxígeno es la clase de todos los átomos que tienen 8 protones en su núcleo, etc. El espacio cociente $A/\sim_p$ es el conjunto de los elementos químicos.

Sistemas comparativos

Algunas cuestiones exigen una respuesta binaria, de sí o no. Por ejemplo, si un átomo determinado es carbono, si un mamífero determinado es macho o hembra, si un número natural es primo o no. Otras cuestiones más bien se resisten a ese tipo de tratamiento. Si nos interesa la altura de las personas, podríamos calificarlas –con el lenguaje ordinario– en altas y bajas. Pero esa clasificación no nos lleva muy lejos. Ya el mismo lenguaje ordinario nos invita a ir más allá, estableciendo comparaciones de altura mediante el llamado grado comparativo de los adjetivos. Aunque Fulano y Mengano son ambos altos (o ambos bajos), lo que nos interesa es saber si Fulano es más o menos alto que Mengano. El concepto de ser más bajo (o más alto) es un concepto comparativo. Otros conceptos comparativos son el de ser más duro (entre minerales), el de ser más antiguo (entre estratos geológicos), o el de ser más rápido (entre corredores).

Un concepto clasificatorio de altura nos dice que tanto x como y son altos, por lo que no resulta muy informativo. Un concepto comparativo de altura nos dice que x es más alto que y, lo que ya nos informa más, pero no nos dice cuánto más alto es x que y (si x es sólo un poquitín más alto que y, o si x es el doble de alto que y...). Un concepto métrico, finalmente, nos dice cuál es la altura de x, cuál es la de y, qué diferencia exacta hay entre ambas, etc. Es el concepto más informativo. De todos modos, el pasar por un concepto comparativo es con frecuencia una etapa necesaria para llegar a disponer de un concepto métrico.

Introducir un concepto comparativo en un dominio A requiere especificar una relación de equivalencia ~ y una relación de orden débil $\prec$. Una relación de orden débil es asimétrica, transitiva y ~-conectada. La relación de equivalencia corresponde a la coincidencia o indiferencia respecto a la propiedad de que se trate (altura, dureza...). La relación de orden débil corresponde a la precedencia o inferioridad respecto a esa propiedad. Se supone que las relaciones ~ y $\prec$ son cualitativas y determinables de un modo empírico y operativo (aceptando a veces ciertas idealizaciones). Si el ámbito A está bien definido, y las relaciones ~ y $\prec$ cumplen las condiciones indicadas,

decimos que $<A, \sim, \prec>$ constituye un sistema comparativo.

En general, $<A, \sim, \prec>$ es un *sistema comparativo* si y sólo si para cualesquiera $x, y, z \in A$:

(1) $x \sim x$
(2) $x \sim y \Rightarrow y \sim x$
(3) $x \sim y \wedge y \sim z \Rightarrow x \sim z$
(4) $x \prec y \Rightarrow \neg\, y \prec x$
(5) $x \prec y \wedge y < z \Rightarrow x \prec z$
(6) $x \prec y \vee y \prec x \vee x \sim y$

Escalas ordinales

Un sistema cualitativo empírico es la base sobre la que establecer una escala, que no es sino un homomorfismo de ese sistema empírico en cierto sistema matemático. Ese homomorfismo es una función o aplicación del dominio A del sistema empírico en algún conjunto matemático, por ejemplo, en el conjunto $\mathbb{R}$ de los números reales.

Una escala asigna números (o vectores) a los elementos de un sistema empírico, de tal manera que esos números y sus interrelaciones matemáticas reflejen las interrelaciones empíricas entre los elementos del sistema empírico. El homomorfismo en que consiste la escala es como un traducción al lenguaje y al sistema matemático correspondiente del sistema empírico cualitativo inicial, que así queda cuantificado de alguna manera.

Una función h es una transformación (de cierto tipo) de otra función f, si h se obtiene a partir de f mediante una fórmula del tipo correspondiente. Dada una escala de cierto tipo, son transformaciones permisibles aquellas transformaciones que siempre convierten escalas de ese tipo en otras escalas de ese mismo tipo. Precisamente un tipo de escala puede caracterizarse como cierto grupo de transformaciones.

Una función es una transformación monótona de otra si ambas crecen juntas. Es decir, la función h es una transformación monótona de la función f si y sólo si para cada $x, y \in A$: $h(x) < h(y) \Leftrightarrow f(x) \Leftrightarrow f(y)$.

Las escalas más débiles son las ordinales. Una escala ordinal es una función que se limita a asignar números a los objetos del sistema empírico, de tal manera que, si un objeto precede a otro respecto a la propiedad de que se trate, se le asigne al primero un número menor que al segundo, y si coinciden, se les asigne el mismo número. No hay pretensión alguna de expresar cuantitavamente las diferencias o las proporciones. La escala de Mohs para la dureza de los minerales es un ejemplo de escala ordinal.

Una *escala ordinal* sobre el sistema comparativo $\langle A, \sim, \prec \rangle$ es un homomorfismo de $\langle A, \sim, \prec \rangle$ en $\langle \mathbb{R}, =, < \rangle$, es decir, una función $f{:}A \rightarrow \mathbb{R}$, tal que para cada $x, y \in A$:

$$x \sim y \Rightarrow f(x) = f(y)$$
$$x \prec y \Rightarrow f(x) < f(y)$$

Teorema de representación: Si $\langle A, \sim, \prec \rangle$ es un sistema comparativo, entonces hay al menos una escala ordinal sobre $\langle A, \sim, \prec \rangle$.

Teorema de unicidad: Si $\langle A, \sim, \prec \rangle$ es un sistema comparativo, f es una escala ordinal sobre $\langle A, \sim, \prec \rangle$, y h es una transformación monótona de f, entonces h es también una escala ordinal sobre $\langle A, \sim, \prec \rangle$.

Sistemas extensivos

La estructura de un sistema comparativo es demasiado débil para determinar una función que nos permita no sólo constatar que un objeto es mayor que otro (respecto a cierta propiedad), sino también medir exactamente en qué proporción el primer objeto es mayor que el segundo, en cuánto lo supera. Para ello necesitamos enriquecer la estructura del sistema comparativo, añadiéndole una nueva operación empírica $\perp$ de combinación o concatenación de objetos. Dados dos objetos x, y del dominio, siempre ha de ser posible combinarlos de tal modo que su combinación, $x \perp y$, sea considerada como un nuevo objeto. Además queremos que esa operación de combinación corresponda de alguna manera a la adición de números. La operación de vertir el contenido de dos botellas iguales en un tercer recipiente es «aditiva» respecto a volumen o masa, pero no lo es respecto a temperatura. El volumen y la masa del líquido contenido en el recipiente final es el doble que el volumen o la masa del líquido en una de las botellas, pero la temperatura resultante no es el doble de la temperatura previa, sino la misma temperatura. Sólo las operaciones del primer tipo conducen a sistemas extensivos, que, a su vez, nos permiten luego definir sobre ellos magnitudes aditivas.

Un sistema extensivo es la expansión de un sistema comparativo mediante la introducción de una operación binaria $\perp$ de combinación o concatenación de dos objetos cualesquiera de A para formar otro objeto de A. Esta operación $\perp$ debe ser asociativa, conmutativa respecto a $\sim$, monótona respecto a $\prec$, positiva y arquimediana. Esta última condición exige que, por mucho que y sea inferior a x, siempre habrá un número natural n tal que la concatenación de y consigo mismo n veces sea superior a x. La manera más sencilla de entender esta condición es exigir que haya en A copias exactas de los objetos de A, de tal manera que la concatenación de x consigo mismo sea la concatenación de x con una copia exacta de x. La concatenación de x consigo mismo n veces puede ser definida recursivamente así: (i) $1x = x$; (ii) $(n + 1)\, x = nx \perp x$.

En general, $\langle A, \sim, \prec, \perp \rangle$ es un *sistema extensivo* si y sólo si para cualesquiera $x, y, z \in A$:

(0) $\langle A, \sim, \prec \rangle$ es un sistema comparativo
(1) $x \perp (y \perp z) \sim (x \perp y) \perp z$
(2) $x \perp y \sim y \perp x$
(3) $x \prec y \Leftrightarrow x \perp z \prec y \perp z \Leftrightarrow z \perp x \prec z \perp y$
(4) $x \prec x \perp y$
(5) $\exists\, n \in \mathbb{N}\; x \prec n\, y$

Escalas proporcionales

Las escalas proporcionales son las más informativas. Asignan números a los objetos de un sistema extensivo de tal modo que la función resultante no sólo conserva el orden del sistema empírico, sino también traduce adecuadamente la operación empírica de combinación de objetos como una adición de números. Toda escala proporcional es una escala ordinal, pero no a la inversa.

Una transformación similar de una función es otra función que resulta de multiplicar cada valor de la primera por un número positivo. Es decir, h es una transformación similar de f si y sólo si hay un $k \in \mathbb{R}^+$, tal que para cada $x \in A$: $h(x) = k \cdot f(x)$. Toda transformación similar es monótona, pero no a la inversa.

Una *escala proporcional* sobre un sistema extensivo $\langle A, \sim, \prec, \perp \rangle$ es un homomorfismo de $\langle A, \sim, \prec, \perp \rangle$ en $\langle \mathbb{R}, =, +, < \rangle$, es decir, una función $f: A \to \mathbb{R}$ tal que para cada $x, y \in A$:

$$x \sim y \Rightarrow f(x) = f(y)$$
$$x \prec y \Rightarrow f(x) < f(y)$$
$$f(x \perp y) = f(x) + f(y)$$

Teorema de representación: Si $\langle A, \sim, \prec, \perp \rangle$ es un sistema extensivo, entonces hay al menos una escala proporcional sobre $\langle A, \sim, \prec, \perp \rangle$.

Teorema de unicidad: Si $\langle A, \sim, \prec, \perp \rangle$ es un sistema extensivo, f es una escala proporcional sobre $\langle A, \sim, \prec, \perp \rangle$ y h es una transformación similar de f, entonces h es también una escala proporcional sobre $\langle A, \sim, \prec, \perp \rangle$.

Un sistema extensivo no determina unívocamente una escala proporcional más que hasta transformaciones similares. Si queremos construir una escala concreta, procedemos del siguiente modo. Elegimos un objeto cualquiera (o clase de equivalencia de objetos) del dominio, y le asignamos convencionalmente un número cualquiera (normalmente, el 1). Ese objeto (o clase de objetos equivalentes) es la unidad estándar o patrón de la escala. Una vez efectuada esa elección por nuestra parte, las propiedades del sistema extensivo determinan unívocamente los valores de la función para el resto de los objetos, de tal modo que se preserva el orden y la operación resulta aditiva. Las diversas escalas se basan en la elección de objetos no equivalentes como patrón o en la asignación de números distintos al mismo patrón. En cualquier caso, cada una de esas escalas es una transformación similar de cualquier otra de ellas.

Medida y metrización

Un concepto métrico o magnitud es un conjunto de escalas del mismo tipo entre el mismo sistema empírico y el mismo sistema matemático. Aunque hay otros tipos de sistemas empíricos sobre los que se pueden definir otros tipos de escala, aquí nos limitamos a considerar las escalas proporcionales sobre sistemas extensivos. Y aunque también hay conceptos métricos no escalares (por ejemplo, los vectoriales), aquí nos limitamos a considerar los escalares (que asignan números reales a los objetos del sistema empírico).

Hay que distinguir claramente los problemas de medición de los de metrización. Cuando ya disponemos de un concepto métrico para un ámbito determinado, y de lo que se trata es de averiguar cuál es el valor (el número) que (una escala de) ese concepto asigna a un objeto determinado del dominio, nos encontramos ante una tarea de medida. Cuando, por el contrario, carecemos de un concepto métrico para un ámbito que de momento sólo nos es dado cualitativamente, y de lo que se trata es de introducir por primera vez un concepto métrico que lo cuantifique, nos encontramos ante un problema de metrización.

Metrizar es introducir un concepto métrico donde no lo había. Es una tarea importante, pero que sólo en raras ocasiones es preciso llevar a cabo. Medir es hallar el valor que la función métrica asigna a un objeto. En todos los laboratorios del mundo se realizan constantemente medidas (a veces millones de medidas cada día). Es el trabajo cotidiano de la ciencia experimental.

Dentro de la metrización, se distingue la fundamental de la derivada. En general, cuando introducimos un concepto métrico, lo hacemos sencillamente definiéndolo en función de otros conceptos métricos previamente definidos. Así, por ejemplo, definimos la densidad d como la masa m partida por el volumen V: $d(x) = m(x)/V(x)$. Con ello la densidad queda definida, pero sólo a condición de que previamente ya sepamos qué es la masa y el volumen. Se trata de una metrización derivada.

Naturalmente, no podemos introducir todos los conceptos métricos de un modo derivado. Algunos deberán ser definidos o introducidos de un modo directo, primitivo o fundamental (al menos al principio, y aunque luego experimenten extensiones de su ámbito de aplicación en función de complejas interrelaciones teóricas).

Aquí vamos a considerar someramente la metrización fundamental de los tres conceptos básicos de la mecánica: los de masa, longitud y tiempo.

El sistema extensivo de masa

Cuando sostenemos dos objetos (por ejemplo, dos libros), uno en cada mano, con fecuencia tenemos la impresión subjetiva de que uno de ellos es más pesado que el otro. Puesto que en la superficie terrestre la aceleración gravitatoria es constante, el peso de los objetos es proporcional a su masa. Un libro nos parece más pesado que el otro porque es más pesado que el otro. Y es más pesado porque tiene más masa. Otras veces nos parece que ambos libros coinciden en cuanto a masa.

Algunos japoneses afirman que el resultado de un combate de sumo está casi siempre determinado por la masa de los contendientes. El más masivo es el que ga-

na. Para comprobar esta hipótesis tenemos que disponer de un procedimiento que nos permita comparar sus masas respectivas.

Desde tiempo inmemorial la comparación entre objetos mesoscópicos en cuanto a su masa se ha efectuado con ayuda de la balanza de brazos iguales. Supongamos que queremos introducir un concepto comparativo de masa para un dominio de objetos mesoscópicos manejables, como piedras o cilindros metálicos, y que disponemos de una balanza, en cuyos platillos podemos colocar dichos objetos sin dificultad.

En primer lugar, introducimos una relación $\sim_M$ de coincidencia en cuanto a masa. Por convención, todo objeto coincide en cuanto a masa consigo mismo. Dos objetos distintos coinciden en cuanto a masa si y sólo si, colocados en sendos platillos de la balanza, la equilibran. Esta relación es reflexiva, simétrica y transitiva, y, por tanto, es una relación de equivalencia.

En segundo lugar, introducimos una relación $\prec_M$ de precedencia en cuanto a masa. Por convención, un objeto nunca es menos masivo que él mismo, nunca se precede en cuanto a masa. Dados dos objetos distintos, el primero es menos masivo que el segundo si y sólo si, colocados en sendos platillos de la balanza, ésta se desequilibra a favor del segundo objeto (es decir, el platillo que contiene el segundo objeto se hunde, mientras el otro sube). Esta relación es asimétrica y transitiva, y, por tanto, es una relación de orden parcial estricto.

La relación $\prec_M$ de precedencia en cuanto a masa es $\sim_M$-conectada, es decir, para cada objeto x, y ocurre: $x \prec_M y$ o $y \prec_M x$ o $x \sim_M y$. Por tanto, $\prec_M$ es un orden débil. Dados dos objetos, siempre uno de ellos desequilibra la balanza a su favor, o ambos la equilibran. Así es el mundo. (Podría ser de otra manera, la balanza podría ponerse a oscilar indefinidamente, por ejemplo, pero de hecho eso no ocurre.)

En tercer lugar, introducimos la operación $\perp_M$ de concatenación o combinación empírica de objetos. Dados dos objetos x, y, la combinación $x \perp_M y$ consiste en colocar ambos objetos en el mismo platillo de la balanza (con lo que ambos, juntos, pasan a ser considerados como un nuevo objeto, que es su concatenación). Esta operación $\perp_M$ es asociativa, conmutativa y monótona respecto a $\prec_M$. También vamos a considerar que es arquimediana, aunque esto representa una gran idealización.

El sistema cualitativo formado por el conjunto A de los objetos mesoscópicos manejables, la relación de coincidencia $\sim_M$, la relación de precedencia $\prec_M$, y la operación de concatenación $\perp_M$, $\langle A, \sim_M, \prec_M, \perp_M \rangle$, es un sistema extensivo.

El concepto métrico de masa

Dado el sistema extensivo que acabamos de describir, basta con elegir uno de los objetos (o una clase de equivalencia de ellos) como unidad, estándar o patrón y asignarle un número para determinar unívocamente una escala de masa.

Hasta la Revolución francesa, había una enorme variedad de escalas (mal definidas, pero todas distintas entre sí) tanto para la masa (o, más bien, el peso) como para otras magnitudes, lo cual creaba todo tipo de confusiones, abusos y problemas. Los Estados Generales habían solicitado varias veces acabar con la anarquía de las unidades de medida. En 1791, y a sugerencia de Talleyrand, la Asamblea Constituyente encargó a la Académie des Sciences que diseñara un nuevo y unificado sistema de pesas y medidas. La Académie nombró un ilustre comité, presidido por Borda, del que formaban parte varios de los mejores científicos del momento, como Lagrange, Condorcet, Monge y Laplace, y que mantenía estrecho contacto con Lavoisier. La Asamblea Constituyente aprobó ese mismo año 1991 (antes de que se instaurara el Terror) las propuestas de la Académie. Talleyrand emigró a Inglaterra durante el Terror. Tras su vuelta a París, en 1798 convocó una conferencia internacional de científicos para perfeccionar el sistema métrico decimal, que en 1799 fue declarado sólo sistema legal en Francia. Su actual sucesor se llama desde 1960 el sistema internacional (SI).

En 1799 los padres del sistema métrico decimal eligieron como patrón de masa la de un decímetro cúbico de agua a 4 °C (temperatura de máxima densidad del agua), y, más específicamente, la de un cilindro metálico de esa masa fundido al efecto por encargo de la Académie. En 1889, este viejo cilindro fue reemplazado por otro nuevo. En efecto, la Conferencia General de Pesas y Medidas celebrada ese año proclamó como patrón de masa un cilindro (de 3,98 cm de altura y diámetro) hecho de una aleación de 90 % de platino y 10 % de iridio. Este cilindro, conservado bajo una triple campana de vidrio, y junto a 6 copias, en la Oficina Internacional de Pesas y Medidas de Sèvres, sigue siendo el estándar o patrón de masa en el SI. La masa es la única magnitud básica del SI, cuya unidad (el kilogramo) no se basa en un proceso de la naturaleza, sino en un objeto artificial convencional: el kilogramo patrón.

De todos modos, el concepto métrico de masa, tal y como lo hemos introducido aquí, sólo se aplica a objetos mesoscópicos manejables, no a átomos o estrellas, por ejemplo, que no pueden colocarse en los platillos de una balanza. A partir de este concepto de masa, y mediante una serie de ampliaciones sucesivas (en realidad, una serie de conceptos distintos de dominio o alcance creciente), se extiende su ámbito de aplicación. Estas ampliaciones son extensiones conservativas del concepto anterior, en el sentido de que conservan los mismos valores para los objetos del ámbito previamente metrizado.

La extensión del concepto de masa en la mecánica clásica tiene lugar mediante el establecimiento de relaciones basadas en sus leyes fundamentales. Esto presentaba inicialmente un problema, pues las dos leyes relevantes (la segunda ley de Newton y la ley de la gravita-

ción) parecían dar lugar a dos nociones distintas de masa, las llamadas masa inercial y masa gravitatoria. La masa inercial se determina (en base a la segunda ley de Newton, $F = m(x)\cdot a(x)$) a partir de la aceleración producida por una fuerza conocida:

$$\text{masa inerte de } x = F/a(x)$$

La masa gravitatoria, por el contrario, se determina (en base a la ley de la gravitación universal, $F_{xy} = G\cdot m(x)\cdot m(y)/r^2$, donde r es la distancia entre x e y) a partir de la medición de la fuerza gravitatoria ejercida por la tierra T sobre un cuerpo x:

$$\text{masa gravitatoria de } x = F_{Tx}\cdot r^2/G\cdot m(T)$$

Afortunadamente ambas masas –la inercial y la gravitatoria– son iguales, como R. Eötvös comprobó experimentalmente a principios de nuestro siglo. (No era necesario que lo fueran, pero de hecho lo son.)

Otra cuestión distinta, planteada y respondida afirmativamente por E. Mach dentro de su programa de reducción de la dinámica a la cinemática, es la de si sería posible definir la masa en términos puramente cinemáticos, como la longitud y el tiempo, con lo que su metrización sería derivada, no fundamental. En función de la posición (reducible a la longitud) y el tiempo se define la aceleración (como segunda derivada de la posición por el tiempo). Y en función de la aceleración trató Mach de definir la masa. Dos objetos tienen la misma masa si y sólo si, al interactuar (por ejemplo, mediante una colisión frontal), obtienen ambos la misma aceleración. Un objeto tiene una masa n veces superior a otro si, al interactuar, el segundo adquiere una aceleración n veces mayor que el primero. Esta interesante propuesta de Mach ha tropezado sin embargo con dificultades.

Al pasar a otras teorías no newtonianas, como la relatividad especial, la noción de masa cambia profundamente. La masa de un objeto o de una partícula ya no es invariante respecto a su velocidad, sino que depende esencialmente de ella. Se trata de un concepto muy distinto de masa, que (con buena voluntad) puede considerarse como una ampliación del concepto clásico a objetos que se mueven a velocidades próximas a la de la luz, extensión conservativa (dentro de los márgenes de medida efectiva) respecto a los objetos a baja velocidad.

El sistema extensivo de longitud

En el lenguaje cotidiano decimos que unos humanos son más altos que otros, que una gasolinera está más lejos de aquí que otra, que un barco tiene mayor eslora que otro, que una falda es más corta que otra, etc. Comparamos cosas respecto a su longitud, como más cortas o largas que otras.

Supongamos que queremos introducir un concepto comparativo de longitud para un dominio de barras metálicas rígidas a temperatura constante.

Introducimos una relación $\sim_L$ de coincidencia respecto a longitud del siguiente modo: Cada barra, por convención, coincide consigo misma respecto a longitud. Dos barras distintas son equivalentes o coincidentes respecto a longitud si y sólo si, yuxtapuestas colateralmente la una junto a la otra de tal manera que sus extremos iniciales coincidan, sus extremos finales coinciden también. Esta relación es reflexiva, simétrica y transitiva, es decir, una relación de equivalencia.

Introducimos luego una relación $\prec_L$ de precedencia: Por convención, ninguna barra es más corta que ella misma. Dadas dos barras distintas, y yuxtapuestas colateralmente la una junto a la otra de tal manera que sus extremos iniciales coincidan, la primera barra es más corta que la segunda si y sólo si el extremo final de la segunda sobresale o se extiende más allá que el de la primera. Esta relación es asimétrica y transitiva, y por tanto es un orden parcial estricto.

Además ocurre que, dadas dos barras cualesquiera, siempre ocurre que una de ellas es más corta que la otra, o que la otra es más corta que la una, o que ambas coinciden en cuanto a longitud. Por tanto, la relación $\preceq_L$ de ser más corta es una relación de orden débil.

Finalmente, introducimos también una operación $\perp_L$ de combinación o concatenación de barras, consistente en colocar colinealmente una barra a continuación de la otra, de tal modo que una empiece donde termine la otra, y que ambas estén en la misma recta. Incluso podríamos pensar en un mecanismo para ajustar firmemente una barra a la otra por su extremo, formando una nueva barra rígida. En cualquier caso, consideramos que la concatenación indicada de dos barras es una nueva barra.

El sistema cualitativo formado por el conjunto A de las barras metálicas rígidas, la relación de coincidencia $\sim_L$, la relación de precedencia $\prec_L$, y la operación de concatenación $\perp_L$, $\langle A, \sim_L, \prec_L, \perp_L\rangle$, es un sistema extensivo.

El concepto métrico de longitud

Dado el sistema extensivo que acabamos de describir, basta con elegir una de las barras (o una clase de equivalencia de ellas) como unidad, estándar o patrón y asignarle un número para determinar unívocamente una escala de longitud. Esta escala de longitud puede luego ser extendida hasta abarcar otros objetos rígidos con una arista (yuxtaponible colateralmente a una barra), posiciones en el espacio, distancias, etc., todo lo cual presenta problemas que no vamos a analizar aquí. En cualquier caso, el primer paso consiste en la elección de una unidad estándar.

La unidad de longitud elegida por el Comité de la Académie des Sciences en 1791 fue la diezmillonésima parte del cuadrante de un meridiano terrestre. Como los

meridianos podían ser diferentes, se eligió uno determinado: el que pasa por Dunkerque. Entre 1792 y 1799 se llevó a cabo la medición por triangulación de un arco de diez grados de latitud sobre ese meridiano (el arco comprendido entre Dunkerque y Barcelona). Hoy sabemos que tuvo un error de 2 partes en 10.000. En realidad el cuadrante de ese meridiano terrestre tiene 10.002.288,3 m, no 10.000.000 m. Esta medida tenía una precisión de casi una parte en 10^4. En 1798 se fabricó el prototipo del metro, una barra de platino, depositada en los Archivos Nacionales y aprobada por la Asamblea Legislativa al año siguiente. Con la introducción de este prototipo se aumentó la precisión en un orden de magnitud, alcanzando una parte en 10^5. La unidad de longitud, el metro (como la de masa, el kilogramo), no se basaría en un concepto, sino en un objeto artificial concreto, el metro patrón. El metro era la longitud de esa barra, con independencia del meridiano.

La Conferencia General de Pesas y Medidas de 1889 estableció como unidad estándar de longitud la distancia (a 0 °C de temperatura) entre dos marcas sobre una nueva barra metálica de perfil en forma de «*X*», hecha de una aleación de 90 % de platino e 10 % de iridio, y conservada en la Oficina de Pesas y Medidas de Sèvres. Otras barras-copias eran comparadas con ella mediante un microscopio reversible especial. Este nueva barra estándar y los procedimientos de comparación que la acompañaban hicieron posible incrementar la precisión por otro orden de magnitud, llegándose así a casi una parte en 10^7. Entre 1899 y 1960, esta barra sirvió de patrón fundamental para la medida de longitudes.

En los años cincuenta todo el mundo era consciente de que ninguna barra metálica era completamente estable. Las ondas de luz coherente proporcionarían un estándar mucho más invariable, y diversas lámparas atómicas fueron ensayadas. Finalmente se eligió el kriptón, un gas noble de número atómico 36, que aparece en la naturaleza (en la atmósfera) en forma de diversos isótopos, de los cuales el más frecuente es el kriptón-86. El espectro del kriptón se compone de 36 líneas, la mayoría amarillas o verdes, correspondientes a las transiciones de energía de los 36 electrones del átomo. La línea elegida para la definición del metro estándar es una particular línea (luz) de color anaranjado.

En 1960, la Conferencia General de Pesas y Medidas decidió cambiar el estándar de longitud, redefiniendo el metro como una longitud igual a 1.650.763,73 veces la longitud de onda en el vacío de la radiación correspondiente a la transición entre los niveles $2p_{10}$ y $5d_5$ del isótopo kriptón-86. Con esto la precisión de la medida se multiplicaba por 100 y alcanzaba una parte en 10^9. El comparador de barras mediante microscopio reversible de Sèvres fue sustituido por una compleja instalación que permite la comparación directa con el estándar definido en función de la radiación del kriptón-86. Además el nuevo estándar tenía la ventaja de ser reproducible en cualquier laboratorio adecuadamente equipado del mundo, sin necesidad de ir a Sèvres.

La lámpara de kriptón-86 permitió incrementar la precisión, pero seguía teniendo problemas, relacionados muchos de ellos con la dificultad de conseguir luz suficientemente coherente (que mantuviera su longitud de onda durante suficiente tiempo como para recorrer un metro, por ejemplo). Pronto se vio que el desarrollo de la tecnología del láser permitía conseguir una luz mucho más coherente que la de la lámpara de kriptón, y se pensó en redefinir el estándar de longitud mediante el láser. Pero no llegó a ser así, pues una solución más radical y definitiva acabó imponiéndose. Esta solución se basa en el hecho (comprobado hasta la saciedad y principio fundamental de la teoría especial de la relatividad) de que la velocidad de la luz en el vacío es una constante absoluta. Puesto que la luz en el vacío recorre siempre la misma longitud por unidad de tiempo, y puesto que la medida del tiempo había adquirido una precisión mayor que todas las demás, bastaba con definir el metro como la longitud recorrida por la luz en el vacío en una fracción determinada de segundo.

En octubre de 1983 la Conferencia General de Pesas y Medidas decidió redefinir el metro, incrementando su precisión en una potencia de 10 (un orden de magnitud), y alcanzando una exactitud de una parte en 10^{10}. A partir de entonces, el metro se define oficialmente como la longitud (o distancia) recorrida por la luz en el vacío en la fracción 1/299792458 de segundo.

Con esta definición, el estándar de longitud se define en función del estándar de tiempo (el segundo). Por tanto, y en teoría, podría considerarse que la metrización de la longitud deja de ser primitiva o fundamental, para convertirse en derivada.

El sistema extensivo de tiempo

Así como atribuimos masa y longitud a los objetos, atribuimos duración a los procesos. Unos procesos duran más o menos que otros. Los viajes, las enfermedades y nuestra propia vida son más o menos breves. Todos tenemos una experiencia subjetiva del tiempo y la duración. El lenguaje ordinario tiene recursos (adverbios temporales, tiempos verbales, etc.) para expresar la duración. Sin embargo, la noción objetiva de tiempo y duración está necesariamente relacionada con los relojes. Cualquier proceso periódico o cíclico o repetitivo puede ser considerado como un reloj (más o menos bueno, según que su ciclo sea más o menos regular). Entre los ciclos regulares bien conocidos están las oscilaciones de los péndulos.

Supongamos que queremos introducir un concepto comparativo de tiempo para las oscilaciones de un conjunto de péndulos.

Primero definimos una relación $\sim_T$ de coincidencia en cuanto a duración. Dos péndulos coinciden en la duración de su período si, puestos en marcha a la vez, alcanzan también a la vez el punto inferior de su trayectoria en cada una de sus oscilaciones.

Luego definimos una relación $\prec_T$ de precedencia en cuanto a duración en la que están dos procesos o períodos si el primero es más breve que el segundo. Un péndulo tiene un período más breve que otro si, puestos en marcha a la vez, el primero completa su primer período mientras el segundo todavía no lo ha completado. El procedimiento podría precisarse mediante un detector fotoeléctrico que reaccionase a la interrupción de un rayo luminoso que pase por el punto inferior de la trayectoria del péndulo.

Finalmente introducimos una operación $\perp_T$ de combinación o concatenación de oscilaciones de péndulos distintos. Para concatenar dos péndulos x e y, ponemos en marcha una oscilación o período de y exactamente en el momento en que x completa su oscilación o período. El detector fotoeléctrico puede también ser usado aquí. En cualquier caso, consideramos que esas dos oscilaciones juntas (desde el inicio de la primera hasta el final de la segunda) forman una nueva oscilación o período.

El sistema cualitativo formado por el conjunto A de las oscilaciones de los péndulos, la relación de coincidencia $\sim_T$, la relación de precedencia $\prec_T$, y la operación de concatenación $\perp_T$, $\langle A, \sim_T, \prec_T, \perp_T \rangle$ es un sistema extensivo.

El concepto métrico de tiempo

Dado el sistema extensivo que acabamos de describir, basta con elegir uno de los objetos (o una clase de equivalencia de ellos) como unidad, estándar o patrón y asignarle un número para determinar unívocamente una escala de tiempo. Más tarde esa escala puede ser extendida a otros procesos estrictamente periódicos (es decir, coordinables con las oscilaciones de algún péndulo, aunque sea ideal) y, finalmente, a todo tipo de procesos. El camino es escabroso, pero transitable.

Según Aristóteles, «el tiempo es la medida del movimiento, según lo anterior y lo posterior» [*Physiké A.*, 219 *b*]. Esa medida del movimiento viene dada por el número de ciclos que recorre un reloj mientras dura ese movimiento. El tiempo es lo que miden los relojes (es decir, los sistemas cíclicos estrictamente periódicos). Durante la mayor parte de la historia los únicos relojes fiables eran los astronómicos, los movimientos cíclicos aparentes del sol y de la luna, que correspondían a la rotación de la Tierra en torno a su eje (el día), a la translación orbital de la Luna en torno a la Tierra (el mes) y a la translación orbital de la Tierra en torno al sol (el año). Hoy sabemos que esos relojes celestes no son perfectos, pero hay que reconocer que nos han prestado un buen servicio como aproximaciones satisfactorias.

Para medir procesos más breves que un día los relojes celestes no servían (sobre todo si el día estaba nublado). Por ello el ingenio humano ha producido una serie de relojes o sistemas cíclicos artificiales, que sirviesen para medir tiempos pequeños, como las horas, los minutos o los segundos: relojes de sol, de arena, de velas, de agua o mecánicos. En Europa, hacia 1300, los mejores relojes ganaban o perdían 15 minutos por día, es decir, sólo lograban una precisión de una parte en 100. Sin embargo en China, por la misma época, una larga tradición de perfeccionamiento de los relojes de agua había conducido en algunos casos a relojes con un error de sólo medio minuto por día, es decir, una precisión 30 veces mayor que en Europa. La relojería europea experimentó un gran progreso en el siglo XVI y, tras la incorporación por Huygens de los principios galileanos del péndulo, los nuevos relojes de péndulo redujeron el error a 10 segundos por día, alcanzando así una precisión de una parte en 10^4. De hecho, hasta 1950 los mejores relojes disponibles siguieron siendo los de péndulo.

Los fundadores del sistema métrico no se preocuparon de definir una nueva unidad de tiempo. En vez de ello, propusieron la unidad natural existente, el día, y se limitaron a sugerir múltiplos y submúltiplos decimales del mismo. Pero esa propuesta no prosperó.

Finalmente se adoptó como unidad de tiempo el segundo, definido en función del movimiento rotacional de la Tierra, como la fracción 1/86.400 (= 24·60·60) del día solar medio. Pero el día solar medio no es constante, va creciendo lentamente. Por ello en 1956 el segundo fue redefinido oficialmente en el SI en función del movimiento orbital de la Tierra alrededor del Sol, es decir, del año solar o tropical. El ecuador celeste y la eclíptica sólo se cruzan en dos puntos: los equinoccios. El centro del Sol, en su trayectoria aparente por la eclíptica, cruza el ecuador celeste dos veces al año. El tiempo comprendido entre dos cruces sucesivos por el equinoccio de primavera se llama un año tropical. Lo malo es que este año tropical también varía, se va reduciendo lentamente. Por eso, había que fijar un año determinado, y se eligió el 1900. El segundo sería la fracción 1/31556925974 del año tropical 1900. De todos modos este segundo «efemérico» así definido sólo estuvo oficialmente vigente durante 11 años.

En 1967 se dio una nueva definición del segundo, que aprovechaba los avances de la ciencia y tecnología atómicas. El segundo pasó a ser definido en función de un cierto número de oscilaciones de la radiación generada por un reloj atómico basado en el comportamiento del isótopo cesio-133.

El cesio es un metal alcalino de número atómico 55. Casi todo el cesio presente en la naturaleza tiene la forma de isótopo 133.

Los átomos no pueden encontrarse más que en ciertos niveles de energía bien determinados. Toda transición entre dos de estos niveles se acompaña de la emisión o de la absorción de un fotón u onda electromagnética de frecuencia invariable. En la última capa del áto-

mo de cesio hay un solo electrón. Si el spin de ese electrón tiene dirección opuesta al spin del núcleo, el átomo de cesio está en su nivel de energía más bajo posible. El nivel inmediatamente superior de energía se alcanza si el spin del electrón externo cambia de dirección y se alinea con el del núcleo. El átomo de cesio pasa del primer estado al segundo (es decir, realiza una transición entre dos niveles hiperfinos) absorbiendo la energía de una radiación electromagnética muy determinada (de 9.192.631.770 hertz, o ciclos por segundo). Si pasa del segundo estado al primero, pierde la energía previamente ganada emitiendo un fotón de la misma frecuencia.

El átomo de cesio gira en torno a su eje de rotación o spin. Colocado en un campo magnético de cierta intensidad, su eje de rotación describe un círculo (como una peonza girando en el suelo) o precesión. Esta precesión puede ser detectada y estimulada por una emisión de radio de una frecuencia de 9.192.631.770 vibraciones o ciclos por segundo, que es la que corresponde a la transición entre dos niveles hiperfinos de isótopo cesio-133. En ello se basa la (desde 1967) definición oficial de la unidad de tiempo: «El segundo es la duración de 9.192.631.770 ciclos de la radiación correspondiente a la transición entre dos niveles hiperfinos del átomo de cesio-133».

¿Cómo conseguir una radiación electromagnética de la frecuencia deseada (9.192.631.770 hertz)? Mediante un reloj de cesio, que simplificadamente, consiste en lo siguiente: Un horno eléctrico calienta cesio-33, con lo que se produce un chorro de átomos de cesio. Un selector magnético filtra y deja pasar sólo los átomos en el nivel más bajo de energía. Estos penetran en una cámara sometida a radiación electromagnética procedente de un oscilador de cuarzo, que pretende acercarse lo más posible a la frecuencia deseada de 9.192.631.770 hertz. La radiación de ese tipo transmite su energía a los átomos con los que interacciona, que pasan al nivel siguiente de energía. Un nuevo selector magnético elimina a los átomos de nivel más bajo de energía. Finalmente, un detector cuenta los átomos que llegan hasta el final (los de energía más alta). Un servomecanismo de retroalimentación modula la producción de ondas electromagnéticas en el oscilador. Si la radiación producida es la correcta, muchos átomos realizan la transición y son detectados al final del proceso. Si la radiación se desvía de la correcta, entonces menos átomos son detectados al final, con lo que el detector envía al oscilador una señal que automáticamente altera su frecuencia, hasta que de nuevo se consiga que un número máximo de átomos alcance el detector. Cuando esto ocurre, ello garantiza que la radiación electromagnética producida tiene exactamente la frecuencia deseada. Esta frecuencia puede entonces ser transformada por divisores electrónicos de frecuencia en señales a intervalos exactos elegidos (por ejemplo, cada microsegundo). Estos relojes de cesio permiten una precisión superior a una parte entre 10^{12}, que es la máxima precisión alcanzada hasta ahora en cualquier tipo de medición.

El rol de los conceptos métricos en la ciencia

La realidad que nos rodea es enormemente compleja y en gran parte resulta opaca a nuestra comprensión y manipulación intelectual. Sin embargo, el mundo ficticio de la matemática, que nosotros hemos creado, es mucho más transparente y mejor conocido. Además, disponemos de técnicas conceptuales potentísimas para resolver los problemas acerca del mundo matemático formulados en el lenguaje de las matemáticas.

Afortunadamente, y desde el siglo XVII, hemos salido del marasmo en que nos había sumido el intento por comprender directamente la realidad, y hemos aprendido a conquistarla por la ruta indirecta de la modelización cuantitativa. Construimos modelos matemáticos de la realidad empírica, y trasladamos a esos modelos los problemas que la realidad nos plantea. Esos problemas, así traducidos al lenguaje matemático, son susceptibles de ser analizados y resueltos matemáticamente. Y la solución matemática, retraducida al lenguaje empírico, se convierte en una solución satisfactoria de nuestros iniciales problemas reales. Al menos, eso es lo que ocurre, si nuestro modelo matemático es suficientemente bueno. En cualquier caso, son los conceptos métricos los que juegan el papel clave de intermediarios en este taumatúrgico ir y venir entre realidad opaca y ficción transparente.

Resulta sorprendente que ese rodeo por el mundo ficticio de la matemática nos proporcione representaciones fiables del mundo real de los procesos físicos y soluciones eficaces a nuestros problemas empíricos. Parece milagroso que algo tan extravagante funcione. Como dice Eugene Wigner, «el milagro de la adecuación del lenguaje de la matemática para la formulación de las leyes de la física es un don maravilloso que nosotros no entendemos ni merecemos». Quizá no lo merezcamos, pero si, a pesar de todo, tratamos de entenderlo, tendremos que seguir avanzando en nuestra comprensión de la estructura, dinámica y papel de los conceptos métricos en la empresa científica.

Bibliografía sumaria

HEMPEL, Carl (1952). *Fundamentals of Concept Formation in Empirical Science*. Vol II, n 7 of *International Encyclopedia of the Unity of Science*. University of Chicago Press.

KLEIN, Arthur (1974). *The World of Measurements*. Simon and Schuster. New York.

KRANTZ, Luce, SUPPES & TVERSKY (1971). *Foundations of Measurement*, vol I (*Additive and Polynomial Representations*). Vol II, 1989. Vol III, 1990. Academic Press. New York.

KYBURG, Henry, Jr. (1984). *Theory and Measurement*. Cambridge University Press.

MOSTERÍN, Jesús (1987). *Conceptos y teorías en la ciencia* (2.ª ed.). Alianza Editorial. Madrid.

NARENS, Louis (1985). *Abstract Measurement Theory*. The MIT Press. Cambridge (Mass).

ORTH, Bernhard (1974). *Einführung in die Theorie des Messens*. Verlag Kohlhammer. Stuttgart.

PFANZAGL, Johann (1971). *Theory of Measurement*. Physica-Verlag. Würzburg.

ROBERTS, Fred (1979). *Measurement Theory* (with Applications to Decisionmaking, Utility, and the Social Sciences). Vol 7 of the *Encyclopedia of Mathematics and its Applications*. Addison-Wesley Publ. Co. Reading (Mass).

STEGMÜLLER, Wolfgang (1970). *Theorie und Erfahrung*. Erster Halbband. Springer-Verlag. Berlin-Heidelberg. (Traducción española: *Teoría y experiencia*. Ed. Ariel. Barcelona 1979.)

WIGNER, Eugene: «The Unreasonable Effectiveness of Mathematics in the Natural Sciences», in: *Communications in Pure and Applied Mathematics*, vol 13, n.° I (February 1960). New York: John Wiley & Sons, Inc.

Si no es pot representar cap terreny a escala, què fan els cartògrafs?

Joan Girbau

1. Introducció

Moltes vegades he tingut l'oportunitat d'impartir cursos de geometria diferencial de superfícies als alumnes de la llicenciatura de matemàtiques de la Universitat Autònoma de Barcelona. Sempre que dintre d'un tal curs explico als estudiants que com a conseqüència del teorema egregi de Gauss és impossible representar qualsevol trosset d'esfera sobre un pla conservant les distàncies, per petit que sigui el trosset d'esfera, jo mateix em quedo una mica perplex. Encara que fa molt de temps que conec teòricament aquest fet, sempre que agafo un mapa m'oblido completament del que sé i, de manera inconscient, tendeixo a creure –com la resta de mortals que no són matemàtics ni cartògrafs– que per calcular la distància entre dos indrets basta mesurar sobre el mapa la distància que els separa i multiplicar el resultat pel factor d'escala d'aquell mapa. Quan he intentat explicar a alguna persona que és impossible fer mapes a escala d'una determinada regió, el meu interlocutor es pensa primer que el que jo dic afecta només regions del planeta molt i molt grans, però quan jo insisteixo que no es pot fer un mapa de Catalunya a escala, ni tan sols un mapa del Vallès Oriental, per posar un exemple, el meu interlocutor es va tornant cada vegada més incrèdul i no s'acaba de creure les meves afirmacions.

La gent tendeix a pensar que els matemàtics obtenen sempre resultats massa teòrics i poc pràctics i que el millor que es podria fer a l'hora d'elaborar un mapa seria fer servir la intuïció més elemental –allò que a vegades es denomina sentit comú–, passar de resultats matemàtics teòrics, oblidar-se del senyor Gauss i de les seves històries i, tot actuant amb sentit pràctic, tractar de representar a escala –com es pugui– les magnituds del terreny sobre el paper.

Aquesta actitud seria, òbviament, poc intel·ligent i és, per sort, totalment oposada a la que han adoptat els cartògrafs des de temps immemorials. Si una persona quan està davant d'un mapa ben fet pot tenir la impressió que aquell mapa és una reproducció fidel, a escala, de la realitat, és precisament perquè els qui l'han confeccionat han utilitzat en la seva elaboració eines matemàtiques avançades per tal de minorar els efectes certs de la impossibilitat de reproduir la realitat, a escala, sobre un paper.

Fa uns anys, mentre utilitzava un mapa a escala 1:10.000 d'una determinada zona de Catalunya, em vaig adonar que sobre el mapa hi havia una quadrícula. Aquella quadrícula estava formada per unes rectes verticals (que jo hauria dit que eren meridians) i per unes rectes horitzontals (que jo hauria dit que eren paral·lels). Sobre les línies verticals hi havia uns nombres: el 436, el 437, ... , fins al 442. Sobre les línies horitzontals hi havia uns altres nombres: el 4.609, el 4.610, ... , fins al 4.613. En algun lloc hi havia una indicació que aquells nombres representaven quilòmetres. Llavors jo no sabia que aquella quadrícula corresponia a unes coordenades denominades UTM (inicials de les paraules angleses Universal Transversal Mercator). Uns mesos més tard, em vaig assabentar que aquelles coordenades no eren pas una particularitat del mapa que llavors jo estava utilitzant, sinó que la pràctica totalitat de mapes detallats de qualsevol zona del planeta, fora dels casquets polars, les empraven. El que no em vaig imaginar la primera vegada que vaig veure un mapa amb aquella quadrícula és que el mateix concepte de coordenades UTM involucrava, a part d'una gran imaginació, molts coneixements matemàtics que a la nostra universitat s'imparteixen al llarg de diferents cursos de la llicenciatura de matemàtiques.

En aquesta conferència explicaré amb tots els detalls la definició de les coordenades UTM per tal de posar en evidència que l'alta matemàtica sovint està present en els fenòmens més quotidians i més pràctics de la nostra civilització.

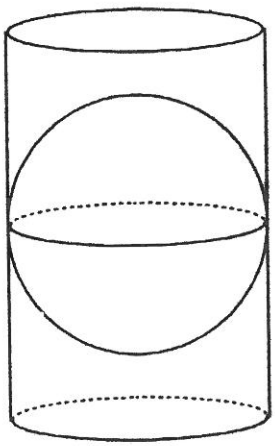
A part de coneixements matemàtics bàsics de càlcul infinitesimal, per seguir aquesta conferència també seran necessàries unes certes nocions de geometria diferencial de superfícies i de funcions holomorfes d'una variable. Les seccions 5 i 6 d'aquesta redacció contenen dos apèndixs on el lector trobarà un resum dels coneixements mínims que necessita d'aquestes dues matèries.

2. Idea intuïtiva de la projecció de Mercator i de la projecció de Mercator transversa

Recordo que quan jo estudiava el batxillerat, algun llibre de geografia dels que havia utilitzat explicava la

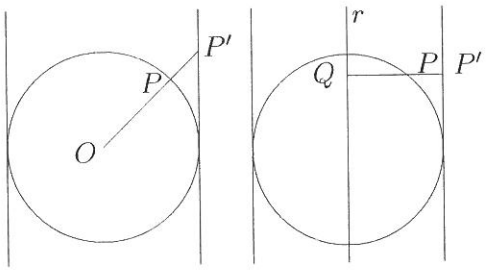
projecció de Mercator (diguem aquí que Gerard Mercator fou un cartògraf holandès de pares alemanys que visqué entre 1512 i 1594). Recordo que aquell meu llibre deia més o menys sobre aquest tema el següent:

Suposeu que la terra és esfèrica (malgrat que en realitat és una mica aplanada pels pols). Imagineu-vos un cilindre circular de radi el de la terra que sigui tangent a la terra per l'equador. Per fer un mapa qualsevol d'un país projecteu els diversos punts d'aquell país des de l'esfera (que és la terra) sobre el cilindre. Així obtindreu una regió del cilindre. Desplegueu després el cilindre sobre un pla tallant-lo prèviament per una generatriu i ja tindreu un mapa pla d'aquell país. Això seria un mapa a escala 1:1. Si voleu un mapa a una altra escala, reduïu el dibuix anterior per un factor d'escala convenient i ja tindreu el mapa desitjat. Aquesta explicació que llavors em semblava d'una claredat meridiana, al cap de molts anys em va semblar una mica confusa. Es projectava cada punt de l'esfera sobre el cilindre. Però, com es projectava? Des de quin punt? De quina manera? Primer se'm va ocórrer que potser es projectava des del centre de l'esfera. Si P era un punt de l'esfera, s'unia P amb el centre O de l'esfera i el punt P' intersecció de la semirecta OP amb el cilindre seria tal vegada la projecció de P sobre el cilindre. De seguida em vaig adonar, però, que per aquest procediment els pols terrestres se n'anaven a l'infinit i que els punts a prop d'un pol se n'anaven, per tant, molt lluny. Segur, doncs, que aquesta no era la manera de projectar que s'utilitzava en la projecció de Mercator (manera de projectar que aquell meu llibre de batxillerat no explicava).



Després se'm va ocórrer que tal vegada es projectava des de l'eix central del cilindre de la manera següent:

Considerem la recta r que uneix el pol nord i el pol sud de la terra (l'eix de la terra). Si P és un punt qualsevol de la terra, des de P tracem una perpendicular a r. Sigui Q el peu d'aquesta perpendicular. Unim Q amb P, i el punt P' on aquesta semirecta talla el cilindre serà la projecció de P sobre el cilindre.



D'aquesta manera els paral·lels de la terra van a parar a rectes horitzontals, i els meridians a rectes verticals. Sobre l'equador es conserven les distàncies, però sobre els altres paral·lels, no. Sobre els altres paral·lels les distàncies de la projecció són més grans que les de la realitat.

Val a dir que aquesta manera de projectar que jo durant anys vaig creure que era la projecció de Mercator, no ho és. Els cartògrafs utilitzen gairebé sempre, amb molt bon criteri, projeccions conformes de la terra sobre un pla. Això vol dir que l'angle entre dues direccions qualssevol de la terra és el mateix que el de les dues direccions corresponents en el pla. Ja que no es poden conservar les distàncies entre la realitat i la projecció, que es conservin almenys els angles. La projecció que acabo de descriure (que jo creia que era la de Mercator) té el gran defecte de no ser conforme i, per tant, no serveix als cartògrafs. La vertadera projecció de Mercator és semblant a la que he descrit abans, però conforme. Té el gran inconvenient (inconvenient relatiu) que el pas d'un punt P de l'esfera al seu homòleg P' del pla no pot ser descrit per una construcció geomètrica del tipus anterior (una construcció que digués «s'uneix P amb tal altre punt, es traça tal o qual recta, es troba la intersecció amb no sé què..., i finalment s'obté P'»). Encara que no es pugui definir per una construcció geomètrica, la vertadera projecció de Mercator es pot definir en paraules, de manera senzilla, tot dient que és l'aplicació de la terra (menys els dos pols) sobre un pla euclidià de coordenades (x, y) que compleix les condicions següents:

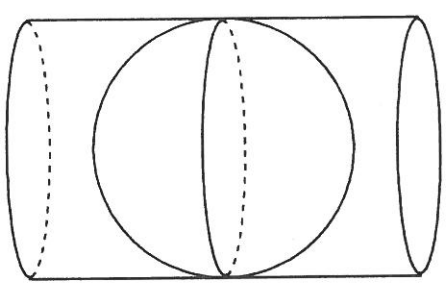
1) Aplica l'equador terrestre en l'eix de les x tot conservant les distàncies sobre l'equador (multiplicades, naturalment, pel factor d'escala del mapa).
2) Aplica el punt de l'equador de longitud geogràfica 0 en l'origen de l'eix de les x. Aplica els punts de la terra pròxims al meridià origen situats a l'est d'aquest, en punts del pla amb $x > 0$, i aplica els punts de l'hemisferi nord en punts del pla amb $y > 0$.
3) L'aplicació és conforme.

Aquesta descripció es pot traduir, naturalment, en fórmules matemàtiques.

La projecció de Mercator és una representació molt

còmoda: és conforme (per definició), els paral·lels es transformen en rectes paral·leles a l'eix de les x i els meridians en rectes paralel·les a l'eix de les y. A més, les fórmules matemàtiques que donen el punt del pla que correspon a un punt de la terra (en funció de la latitud i la longitud) no són gaire complicades. Ara bé, té el gran inconvenient que lluny de l'equador les distàncies queden molt alterades. Per aquesta raó no va bé per fer mapes detallats de zones geogràfiques una mica allunyades de l'equador. Per a aquest tipus de mapes s'ha anat imposant la projecció de Mercator transversa (coneguda com UTM). Si la volguéssim descriure intuïtivament com hem fet abans amb la projecció de Mercator diríem que, en primer lloc, es considera la regió de la terra compresa entre els dos cercles polars (per als casquets polars no s'utilitza la representació UTM). Es divideix aquesta regió en fusos de 6 graus d'amplitud. Així, el primer fus, que conté Catalunya, abraça des del meridià de 0º fins al meridià de 6º de longitud est. El segon fus va des del de 6º fins al de 12º de longitud est, etc. Per a cada un d'aquests fusos es fa una projecció diferent que anem a descriure ara intuïtivament. Es considera un cilindre de radi el de la terra tangent a la terra al llarg del meridià central del fus (en el cas de Catalunya, el meridià de 3º de longitud est). Es projecten els punts del fus sobre el cilindre, de la mateixa manera que es feia en la projecció de Mercator (el paper que allà feia l'equador, aquí el fa el meridià central del fus). Si hom projecta els punts del fus sobre el cilindre des de l'eix del cilindre, s'obtindrà una representació que s'assemblarà molt a la vertadera projecció UTM, però que no ho serà ben bé (perquè no és conforme).

Tal com passava amb la projecció de Mercator, la



projecció UTM no es pot definir per cap construcció geomètrica, però sí que es pot definir en paraules per les següents condicions:

1) Sigui m_0 el meridià central del fus. El meridià m_0 s'aplica sobre l'eix de les y. Sobre aquest meridià es conserven les distàncies entre la realitat i la projecció (multiplicades, naturalment, pel factor d'escala del mapa).
2) La projecció aplica el punt d'intersecció de l'equador amb el meridià m_0, en l'origen de l'eix de les y. Aplica els punts de l'hemisferi nord en punts del pla amb $y > 0$, i els punts del fus a l'est del meridià m_0 en punts amb $x > 0$.
3) L'aplicació es conforme.

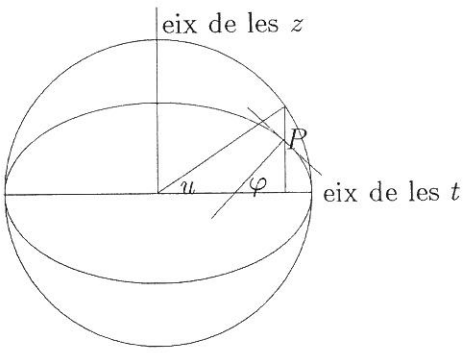
En aquesta projecció, si prescindim del factor d'escala, quan hom s'allunya del meridià central m_0 les distàncies sobre el mapa són més grans que les distàncies vertaderes sobre la terra. En canvi, sobre el meridià central, són iguals. Per tal que sempre les distàncies de la realitat siguin gairebé iguals a les del mapa, l'Army Map Service dels Estats Units (AMS) va proposar multiplicar les distàncies sobre el meridià central pel factor $k = 0{,}9996$. D'aquesta manera a dos punts sobre el meridià central que disten 1 km se'ls assignen punts sobre l'eix de les y del mapa que només disten 999,6 metres. S'escurcen les distàncies sobre el meridià central per tal de compensar l'allargament lluny d'aquest meridià (es pretén que la distància entre punts d'un meridià allunyat del central no s'allargui tant com abans). L'AMS també va proposar sumar 500 km a totes les coordenades x obtingudes pel procediment anterior per tal de no tenir coordenades x negatives dins del fus. Amb aquestes modificacions proposades per l'AMS dels Estats Units, la projecció UTM, que s'utilitza arreu del món llevat de les zones pròximes als pols, pot ser descrita de la següent manera:

1) El meridià central del fus, m_0, s'aplica sobre la recta vertical paral·lela a l'eix de les y d'abscissa $x = 500$ km. Sobre aquest meridià es conserven les distàncies entre la realitat i la projecció, multiplicades pel factor d'escala del mapa i pel factor $k = 0{,}9996$.
2) La projecció aplica el punt d'intersecció de l'equador amb el meridià m_0, en el punt de coordenades (500 km, 0). Aplica els punts de l'hemisferi nord en punts del pla amb $y > 0$, i els punts del fus a l'est del meridià m_0 en punts amb $x > 500$ km.
3) L'aplicació es conforme.

3. Com obtenir les fórmules matemàtiques que donen la projecció UTM: planteig del problema

La forma de la terra s'aproxima bastant a la d'un el·lipsoide de revolució. Recordem que una el·lipse centrada a l'origen, en un pla de coordenades (t, z), ve representada pels punts que compleixen l'equació $(t^2 / a^2) + (z^2 / b^2) = 1$. Aquesta el·lipse es pot representar paramètricament per les equacions $t = a \cos u$, $z = b \sin u$, on, si P és el punt de l'el·lipse de coordenades (t, z), el paràmetre u corresponent a P és l'angle que indica la figura. La latitud φ del punt P és l'angle que forma la perpendicular a la tangent a l'el·lipse en el punt P amb l'eix de les t. La relació que hi ha entre la latitud φ i el paràmetre u és $\tan u = (b/a) \tan \varphi$.

L'el·lipsoide de revolució engendrat per aquesta el·lip-



se a l'espai, en girar entorn de l'eix de les z, vindrà donat paramètricament per

$$\begin{cases} x = t\cos\lambda = a\cos u\cos\lambda \\ y = t\sin\lambda = a\cos u\sin\lambda \\ z = b\sin u \end{cases}$$

Aquesta representació paramètrica ve donada, doncs, per l'aplicació f: $(\lambda, u) \to (a\cos u\cos\lambda,\ a\cos u\sin\lambda,\ b\sin u)$. El paràmetre λ representa la longitud geogràfica i el paràmetre u està relacionat amb la latitud geogràfica (com hem dit abans) per $\tan u = (b/a)\tan\varphi$. Com que volem estudiar la representació UTM d'un determinat fus, prendrem les longituds geogràfiques comptades a partir del meridià central del nostre fus. Tindrem:

$$\partial\vec{f}/\partial\lambda = (-a\cos u\sin\lambda,\ a\cos u\cos\lambda,\ 0)$$
$$\partial\vec{f}/\partial u = (-a\sin u\cos\lambda,\ -a\sin u\sin\lambda,\ b\cos u)$$

La matriu de productes escalars entre aquests dos vectors (primera forma quadràtica fonamental) és

$$\begin{pmatrix} a^2\cos^2 u & 0 \\ 0 & a^2\sin^2 u + b^2\cos^2 u \end{pmatrix}$$

Aquesta matriu es pot escriure així:

$$a^2\cos^2 u \begin{pmatrix} 1 & 0 \\ 0 & \tan^2 u + (b/a)^2 \end{pmatrix}$$

Considerem una funció $\mathcal{L}$ de u que sigui primitiva de $\sqrt{\tan^2 u + (b/a)^2}$. Si canviem la coordenada u que fèiem servir fins ara per $\mathcal{L}$, tindrem $\partial\vec{f}/\partial\mathcal{L} = (\partial\vec{f}/\partial u)\,(du/d\mathcal{L})$. Per tant, el producte escalar de $\partial f/\partial\mathcal{L}$ per ell mateix serà $g(\partial\vec{f}/\partial\mathcal{L},\ \partial\vec{f}/\partial\mathcal{L}) = g(\partial\vec{f}/\partial u,\ \partial\vec{f}/\partial u)\,(du/d\mathcal{L})^2 = a^2\cos^2 u$. Per tant, en les coordenades $(\lambda, \mathcal{L})$ la primera forma quadràtica fonamental s'escriurà

$$a^2\cos^2 u\,(\mathcal{L}) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ens adonem, per tant, que l'aplicació $(\lambda, \mathcal{L}) \to f(\lambda, \mathcal{L})$ és una representació conforme d'un cert domini V del pla de les coordenades $(\lambda, \mathcal{L})$ en el fus d'el·lipsoide que estem considerant:

$$V \xrightarrow{f} \text{Fus d'el·lipsoide}$$
$$(\lambda, \mathcal{L}) \longrightarrow f(\lambda, \mathcal{L})$$

Aquesta representació, però, tot i ser conforme, no compleix totes les propietats exigides a la representació UTM. Ara voldríem trobar una representació conforme que complís totes aquelles propietats. Oblidem-nos, per simplificar, de les condicions suplementàries proposades per l'AMS (factor d'escala $k = 0{,}9996$ i afegir 500 km a la coordenada x). La idea fonamental és buscar una funció holomorfa convenient $F : V \to \mathbb{C}$, de manera que la composició

$$\text{Fus d'el·lipsoide} \xrightarrow{f^{-1}} V \xrightarrow{F} \mathbb{C}$$
$$\text{Punt de l'el·lipsoide} \longrightarrow \lambda + i\mathcal{L} \longrightarrow x + iy = F(\lambda + i\mathcal{L})$$

(que serà conforme per ser composició de dues que ho són) compleixi les propietats exigides a la representació UTM. Com que el meridià central del fus (que correspon a $\lambda = 0$) volem que s'apliqui sobre l'eix de les y (eix imaginari de $\mathbb{C}$), la funció F que busquem haurà de complir $F(i\mathcal{L}) = iy$. Imposem ara que les distàncies sobre el meridià central es conservin en la representació. Per fer això, considerem sobre el meridià central del fus la distància y d'un punt a l'equador expressada en funció de la $\mathcal{L}$ d'aquell punt (recordem que $\mathcal{L}$ és funció de u i u és funció de la latitud φ del punt). Tindrem, doncs, $y = y(\mathcal{L})$, on y és la distància a l'equador sobre el meridià central. Encara que ara no coneguem explícitament la funció $y = y(\mathcal{L})$, no és gaire difícil de veure que serà una funció analítica real, perquè s'obté per composició i integració de funcions analítiques reals. Per tant $y = y(\mathcal{L})$ s'expressarà localment per una sèrie de potències. Considerem la funció F de la variable complexa $\omega = \lambda + i\mathcal{L}$ definida per $F(\omega) = iy(\omega/i)$, on $y(\omega/i)$ significa que en la sèrie real corresponent a la funció $y = y(\mathcal{L})$ se substitueix la variable real $\mathcal{L}$ per la variable complexa $\omega = \lambda + i\mathcal{L}$ dividida per i. Amb aquesta funció holomorfa F, tindrem que la composició

$$\text{Fus d'el·lipsoide} \xrightarrow{f^{-1}} V \xrightarrow{F} \mathbb{C}$$
$$\text{Punt de l'el·lipsoide} \longrightarrow \lambda + i\mathcal{L} \longrightarrow x + iy = F(\lambda + i\mathcal{L})$$

complirà les condicions exigides a la representació UTM. En efecte, un punt qualsevol del meridià central ($\lambda = 0$) s'aplica per fer f^{-1} en $i\mathcal{L}$. I aquest s'aplica per F en $F(i\mathcal{L}) = iy$ ($\mathcal{L}$), que és el punt sobre l'eix imaginari (eix de les y) que té per ordenada la distància a l'equador que hi havia des del punt corresponent del meridià central.

4. Fórmules explícites de la projecció UTM

La funció $\mathcal{L}$ l'hem definida per la integral indefinida

$$1) \qquad \mathcal{L} = \int \sqrt{\tan^2 u + (b/a)^2}\, du\,.$$

Prenguem la primitiva tal que $\mathcal{L}(0) = 0$. De moment, sembla difícil resoldre explícitament la integral anterior. Fem el canvi de variable $\tan u = (b/a) \tan \varphi$, que correspon a expressar $\mathcal{L}$ en funció de la latitud geogràfica φ. D'aquest canvi es dedueixen les fórmules següents que donen el sinus i el cosinus de u en funció del sinus i cosinus de φ

$$2) \qquad \begin{cases} \sin u = \dfrac{b}{a}\, \dfrac{\sin \varphi}{\sqrt{1 - e^2 \sin^2 \varphi}} \\ \cos u = \dfrac{\cos \varphi}{\sqrt{1 - e^2 \sin^2 \varphi}} \end{cases}$$

on e expressa l'excentricitat de l'el·lipse definida per $e^2 = 1 - (b/a)^2$ (se suposa $b < a$). Amb aquest canvi, la integral (1) queda:

$$3) \qquad \mathcal{L} = \int \frac{(1 - e^2)\, d\varphi}{(1 - e^2 \sin^2 \varphi) \cos \varphi}$$

Això és la integral d'una funció racional en sinus i cosinus que se sap resoldre explícitament. La solució que val 0 quan φ és 0 queda:

$$4) \qquad \mathcal{L} = \frac{1}{2} \log \frac{1 + \sin \varphi}{1 - \sin \varphi} - \frac{e}{2} \log \frac{1 + e \sin \varphi}{1 - e \sin \varphi}$$

Per saber explícitament com funciona la projecció UTM,

Fus d'el·lipsoide $\longrightarrow$ $\mathbb{C} \cong \mathbb{R}^2$

Punt del fus $\longrightarrow$ Punt del pla complex

hauríem de donar l'expressió del punt del pla complex que correspon al punt de l'el·lipsoide d'una longitud λ (comptada des del meridià central del fus) i d'una latitud φ. Segons hem dit a la secció anterior, al punt de l'el·lipsoide de coordenades (λ, φ) li associem el punt de $\mathbb{C}$ donat per $\lambda + i\mathcal{L}$ on $\mathcal{L}$ ve donat en funció de φ per la fórmula (4). Llavors el punt de $\mathbb{C}$ associat al punt (λ, φ) de l'el·lipsoide, és $F(\lambda + i\mathcal{L})$, on F és la funció holomorfa definida a la secció anterior. Per a cada punt (λ, φ) del fus d'el·lipsoide, calcularem la seva $\mathcal{L}$ en funció de φ per (4) i usarem un desenvolupament de F per Taylor en el punt $p = i\mathcal{L}$ de $\mathbb{C}$ (punt imaginari pur, que correspon, per tant, a un punt del meridià central de l'el·lipsoide). El desenvolupament de Taylor de F en el punt p serà de la forma

$$5) \qquad \begin{aligned} &F(\lambda' + i\mathcal{L}') = \\ &a_0 + a_1(\lambda' + i(\mathcal{L}' - \mathcal{L})) + a_2(\lambda' + i(\mathcal{L}' - \mathcal{L}))^2 + \ldots \end{aligned}$$

Recordem com havíem definit la funció F a l'apartat anterior. Recordem que si $y(\mathcal{L}')$ indicava la distància a l'equador sobre l'el·lipsoide d'un punt del meridià central corresponent a aquell $\mathcal{L}'$, llavors $F(\omega') = iy(\omega'/i)$. Això vol dir que si desenvolupem la funció real $y(\mathcal{L}')$ per Taylor en un punt $\mathcal{L}$.

$$6) \qquad y(\mathcal{L}') = b_0 + b_1(\mathcal{L}' - \mathcal{L}) + b_2(\mathcal{L}' - \mathcal{L})^2 + \ldots$$

Llavors

$$7) \qquad \begin{aligned} &F(\lambda' + i\mathcal{L}') = \\ &ib_0 + ib_1 \left(\frac{\lambda' + i(\mathcal{L}' - \mathcal{L})}{i}\right) + ib_2 \left(\frac{\lambda' + i(\mathcal{L}' - \mathcal{L})}{i}\right)^2 + \ldots \end{aligned}$$

Comparant (5) amb (7) veiem que $a_0 = ib_0$, $a_1 = b_1$, $a_2 = -ib_2$, i en general $a_k = (-i)^{k-1} b_k$.

A nosaltres ens interessarà l'expressió (5) quan $\mathcal{L}' = \mathcal{L}$. Llavors tindrem

$$8) \qquad x + iy = F(\lambda + i\mathcal{L}) = a_0 + a_1\lambda + a_2\lambda^2 + \ldots$$

Això ens dóna l'expressió que busquem del punt del pla complex que correspon al punt de l'el·lipsoide (λ, φ) o bé $(\lambda, \mathcal{L})$ (tingueu en compte que donar φ equival a donar $\mathcal{L}$).

Per conèixer totalment l'expressió (8) haurem de conèixer el valor dels coeficients a_k, o bé, equivalentment, dels coeficients b_k de l'expressió (6).

El coeficient b_0, és $y(\mathcal{L})$, o sigui, la distància a l'equador mesurada sobre l'el·lipsoide del punt de l'el·lipsoide $(0, \mathcal{L})$, o equivalentment, del punt $(0, \varphi)$. Anem a calcular, doncs, b_0. El meridià central (parametritzat per u) és la corba obinguda fent $\lambda = 0$ en la representació paramètrica de l'el·lipsoide, o sigui la corba $u \to (a \cos u, 0, b \sin u)$. La norma del vector tangent a aquesta corba en un punt

de paràmetre u és $\sqrt{a^2 \cos^2 u + b^2 \sin^2 u}$. O sigui, si y és la longitud d'aquesta corba des de l'equador, $dy/du = \sqrt{a^2 \cos^2 u + b^2 \sin^2 u}$. Si volem expressar y en funció de la latitud φ (en lloc del paràmetre u), utilitzant les fórmules (2) tindrem

$$\frac{dy}{du} = \frac{b}{\sqrt{1 - e^2 \sin^2 \varphi}}$$

Com que $dy/d\varphi = (dy/du)\,(du/d\varphi)$, derivant la relació $\tan u = (b/a) \tan \varphi$ i utilitzant les fórmules (2) s'obté

$$9) \qquad \frac{dy}{d\varphi} = \frac{b^2}{a} (1 - e^2 \sin^2 \varphi)^{-3/2}$$

Per tant,

$$10) \qquad a_0 = ib_0 = i \frac{b^2}{a} \int_0^{\varphi} (1 - e^2 \sin^2 \xi)^{-3/2} d\xi$$

Aquesta és una integral el·líptica que s'ha de calcular o bé numèricament o bé utilitzant algun desenvolupament en sèrie. Per exemple, de la sèrie binòmica s'obté

$$(1 - e^2 \sin^2 \xi)^{-3/2} = 1 + \frac{3}{2} e^2 \sin^2 \xi + \frac{15}{8} e^4 \sin^4 \xi + \frac{35}{16} e^6 \sin^6 \xi + \ldots$$

Si substituïm ara en la fórmula anterior $\sin^2 \xi$, $\sin^4 \xi$ i $\sin^6 \xi$ per

$\sin^2 \xi = (1/2) - (1/2) \cos 2\xi$
$\sin^4 \xi = (3/8) - (1/2) \cos 2\xi + (1/8) \cos 4\xi$
$\sin^6 \xi = (5/16) - (15/32) \cos 2\xi - (3/16) \cos 4\xi - (1/32) \cos 6\xi$

i substituïm després el resultat a (10) i integrem terme a terme, obtenim

$$11) \qquad a_0 = i\,(b^2/a)\,(\alpha\varphi + \beta \sin 2\varphi + \gamma \sin 4\varphi + \delta \sin 6\varphi + \ldots)$$

on els coeficients α, β, γ i δ vénen donats per

$\alpha = 1 + (3/4)e^2 + (45/64)e^4 + (175/256)e^6 + \ldots$
$\beta = -(1/2)\,((3/4)e^2 + (15/16)e^4 + (525/512)e^6 + \ldots)$
$\gamma = (1/4)\,((15/64)e^4 + (105/256)e^6 + \ldots)$
$\delta = -(1/6)\,((35/512)e^6 + \ldots)$

Hem de calcular ara els altres coeficients b_1, b_2, ... de la sèrie (6) per tal de tenir els a_1, a_2 ... de (8). De la fórmula de Taylor es té $b_k = (1/k!)\,(d^k y/d\mathcal{L}^k)$. Les derivades successives de y respecte a $\mathcal{L}$ vénen donades per

$dy/d\mathcal{L} = (dy/d\varphi)\,(d\varphi/d\mathcal{L}) = (dy/d\varphi)\,(d\mathcal{L}/d\varphi)^{-1}$
$d^2y/d\mathcal{L}^2 = (d^2y/d\varphi^2)\,(d\mathcal{L}/d\varphi)^{-2} - (dy/d\varphi)\,(d\mathcal{L}/d\varphi)^{-3}\,(d^2\mathcal{L}/d\varphi^2)$

etc.

Aquí apareixen derivades successives de y respecte a φ que es calculen a partir de l'expressió (9). També apareixen derivades successives de $\mathcal{L}$ respecte a φ que es poden calcular a partir de (3). Donem a continuació les expressions finals dels coeficients $a_k = (-i)^{k-1} b_k$ per a $k = 1, 2, 3$ i 4. Si per abreujar designem per $N = a(1 - e^2 \sin^2 \varphi)^{-1/2}$ i per $\eta^2 = e^2 \cos^2 \varphi / (1 - e^2)$, tindrem

$a_1 = b_1 = N \cos \varphi$
$a_2 = -ib_2 = i(1/2)N \cos^2 \varphi \tan \varphi$
$a_3 = -b_3 = (1/6)N \cos^3 \varphi\,(1 + \tan^2 \varphi - \eta^2)$
$a_4 = ib_4 = (i/24)N \cos^2 \varphi \tan \varphi\,(5 - \tan^2 \varphi + 9\eta^2 + 4\eta^4)$

Hauríem de dir, finalment, que per fer servir les fórmules anteriors amb les dimensions reals de la terra, hauríeu de prendre com a valor del semieix major de l'el·lipse $a = 6.378{,}388$ km, i com a excentricitat $e = 8{,}199191 \times 10^{-2}$.

5. Apèndix 1: repàs de geometria diferencial de superfícies

Una superfície parametritzada de $\mathbb{R}^3$ es pot definir com un parell (V, f) on V és un cert obert de $\mathbb{R}^2$ i f és una aplicació diferenciable f: $V \to \mathbb{R}^3$, injectiva, tal que per a tot $v \in V$ l'aplicació lineal tangent $f'(v) : \mathbb{R}^2 \to \mathbb{R}^3$ és injectiva. El subconjunt $S = f(V) \subset \mathbb{R}^3$ és allò que es dibuixa de la superfície parametritzada i l'aplicació $f : V \to S$ s'anomena parametrització de S. Així, per exemple, si designem per (λ, μ) les coordenades canòniques de $\mathbb{R}^2$ i considerem $V = \{(\lambda, \mu) \in \mathbb{R}^2$ amb $0 < \lambda < 2\pi\}$ i l'aplicació $f : V \to \mathbb{R}^3$ donada per $\vec{f}\,(\lambda, \mu) = (\cos \lambda, \sin \lambda, \mu)$, el conjunt $S = f(V) \subset \mathbb{R}^3$ és cilindre de radi 1 entorn de l'eix de les z (cilindre que no conté la recta vertical $\mu \to (1, 0, \mu)$ perquè no permetem que λ prengui els valors 0 i 2π).

Sigui (V, f) una superfície parametritzada de $\mathbb{R}^3$. Sigui $p \in V$ i sigui

$$\begin{aligned} (-\varepsilon, \varepsilon) &\to V \\ t &\to v(t) = (v_1(t), v_2(t)) \end{aligned}$$

una petita corba de V tal que $v(0) = p$. Llavors

$$\begin{aligned} (-\varepsilon, \varepsilon) &\to \mathbb{R}^3 \\ t &\to f(v(t)) \end{aligned}$$

serà una petita corba de $S = f(V)$ que passarà per $f(p)$. El vector tangent a aquesta corba de $\mathbb{R}^3$ en el punt $f(p)$ serà

$$\left(\frac{df(v_1(t), v_2(t))}{dt}\right)_{t=0} = \left(\frac{\partial \vec{f}}{\partial v_1}\right)_p \left(\frac{dv_1}{dt}\right)_{t=0} + \left(\frac{\partial \vec{f}}{\partial v_2}\right)_p \left(\frac{dv_2}{dt}\right)_{t=0}$$

Aquest vector tangent és, doncs, combinació lineal dels vectors $(\partial \vec{f}/\partial v_1)_p$ i $(\partial \vec{f}/\partial v_2)_p$. Es defineix l'espai tangent a la superfície S en el punt $q = f(p)$ com el subespai $T_q(S)$ de $\mathbb{R}^3$ format pels vectors tangents en el punt q a totes les corbes contingudes a S que passen per q. El càlcul anterior mostra que $(\partial \vec{f}/\partial v_1)_p$ *i* $(\partial \vec{f}/\partial v_2)_p$ són base de $T_q(S) \subset \mathbb{R}^3$.

El producte escalar usual g de $\mathbb{R}^3$ restringit a $T_q(S)$ dóna un producte escalar a $T_q(S)$ que queda determinat pel coneixement de la matriu simètrica (g_{ij}) de productes escalars.

$$g_{ij} = g\left(\frac{\partial \vec{f}}{\partial v_i}, \frac{\partial \vec{f}}{\partial v_j}\right), \text{ en què } i, j \text{ varien des d'1 fins a 2.}$$

Sovint la matriu (g_{ij}) és denominada *primera forma quadràtica fonamental* de la superfície.

Sigui (V, f) una superfície parametritzada de $\mathbb{R}^3$. Direm que la parametrització f de $S = f(V)$ és una isometria entre l'obert V del pla i la superfície S quan per a tot $v \in V$ l'aplicació lineal tangent $f'(v)$: $\mathbb{R}^2 \to T_{f(v)}(S) \subset \mathbb{R}^3$ és una isometria d'espais vectorials, quan es considera a $\mathbb{R}^2$ el producte escalar habitual, i a $T_{f(v)}(S)$ el que hem considerat anteriorment (l'induït pel producte escalar de $\mathbb{R}^3$). Això vol dir que qualsevol arc de corba de V té la mateixa longitud que l'arc de corba de S que s'obté aplicant f al primer. També l'angle entre els vectors tangents en un punt $p \in V$ a dues corbes de V que passen per p és el mateix que l'angle en $q = f(p)$ dels vectors tangents a les corbes transformades per f. Recordem que l'aplicació lineal tangent $f'(v)$: $\mathbb{R}^2 \to T_{f(v)}(S)$ es defineix de la següent manera: Donat un vector $X \in \mathbb{R}^2$, considerem una corba $v(t) = (v_1(t), v_2(t))$ de $V \subset \mathbb{R}^2$ tal que $v(0) = v$ i $(dv(t)/dt)_{t=0} = X$. Llavors la imatge de X per l'aplicació $f'(v)$ és el vector $(df(v(t)/dt)_{t=0}$. Abans hem vist que els components d'aquest vector de $T_{f(v)}(S) \subset \mathbb{R}^3$ en la base $\{\partial \vec{f}/\partial v_1, \partial \vec{f}/\partial v_2\}$ són $(dv_1(t)/dt)_{t=0}$ i $(dv_2(t)/dt)_{t=0}$. És a dir, els mateixos components que tenia el vector $X = (dv(t)/dt)_{t=0}$ en la base canònica de $\mathbb{R}^2$. Per tant, arribem a la conclusió que quan a $\mathbb{R}^2$ prenem la base canònica i a $T_{f(v)}(S)$ la base $\{\partial \vec{f}/\partial v_1, \partial \vec{f}/\partial v_2\}$, l'aplicació $f'(v)$: $\mathbb{R}^2 \to T_{f(v)}(S)$ ve donada en components per $(X_1, X_2) \to (X_1, X_2)$. Per tant, aquesta aplicació serà una isometria si i només si la matriu (g_{ij}) de productes escalars dels vectors de la base $\{\partial \vec{f}/\partial v_1, \partial \vec{f}/\partial v_2\}$ és la matriu identitat

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Considerem, per exemple, el cilindre parametritzat del que hem parlat abans. Recordem que preníem $V = \{(\lambda, \mu) \in \mathbb{R}^2 \text{ amb } 0 < \lambda < 2\pi\}$ i consideràvem l'aplicació $f: V \to \mathbb{R}^3$ donada per $f(\lambda, \mu) = (\cos \lambda, \sin \lambda, \mu)$. En aquest cas, els vectors $\partial f/\partial \lambda$, i $\partial f/\partial \mu$ són respectivament els vectors de $\mathbb{R}^3$ $(-\sin \lambda, \cos \lambda, 0)$ i $(0, 0, 1)$. La matriu (g_{ij}) de productes escalars d'aquests dos vectors és la matriu identitat. Per tant, aquesta parametrització f del cilindre dóna una isometria entre l'obert V del pla i el cilindre S.

Sigui (V, f) una superfície parametritzada de $\mathbb{R}^3$. Direm que la parametrització f de $S = f(V)$ és una aplicació conforme entre l'obert V del pla i la superfície S quan per a tot $v \in V$ l'aplicació lineal tangent $f'(v)$: $\mathbb{R}^2 \to T_{f(v)}(S) \subset \mathbb{R}^3$ conserva els angles quan es considera a $\mathbb{R}^2$ el producte escalar habitual i a $T_{f(v)}(S)$ l'induït pel producte escalar de $\mathbb{R}^3$. És immediat de veure (seguint el mateix raonament que hem utilitzat abans per a les isometries) que si la matriu (g_{ij}) de productes escalars dels vectors $\partial \vec{f}/\partial v_1$ *i* $\partial \vec{f}/\partial v_2$ és de la forma

$$h(v_1, v_2) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

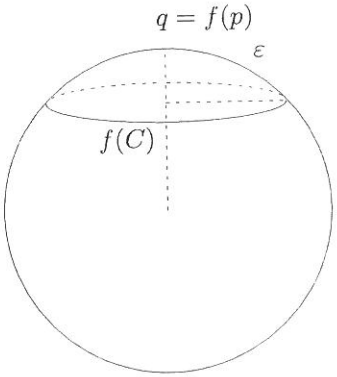
on $h(v_1, v_2)$ és qualsevol funció diferenciable que no s'anul·la en cap punt, llavors la parametrització f és una aplicació conforme entre V i $S = f(V)$. Quan la funció h és la constant 1 com en el cas del cilindre, es té una isometria.

Com hem dit en la introducció, el teorema egregi de Gauss –un teorema absolutament meravellós per merèixer el qualificatiu d'egregi (que vol dir insigne, il·lustre, distingit)– porta com a conseqüència que les úniques superfícies que poden admetre parametritzacions isomètriques per un pla són aquelles que tenen curvatura de Gauss nul·la. En particular, ni una esfera ni un el·lipsoide admeten parametritzacions isomètriques per un pla. Aquest fet és molt important a l'hora de fer mapes de la terra.

Anem a fer ara un raonament simple (sense utilitzar el teorema egregi) que demostri que un tros d'esfera no admet cap parametrització isomètrica per un pla. Imaginem que tenim un obert V de $\mathbb{R}^2$ i una aplicació $f: V \to \mathbb{R}^3$ de manera que la imatge $f(V) = S \subset \mathbb{R}^3$ està continguda en l'esfera de radi R i centre l'origen. Suposem que la parametrització f és una isometria entre V i $S = f(V)$. A S definim una noció de distància entre dos punts q i q' de S com l'inferior de totes les longituds de totes les corbes contingudes a S que els uneixen. La hipòtesi que f és isometria comporta que f conserva les longituds de les corbes i, per tant, per a qualsevol parell de punts p i p' de V es complirà $d(p, p') = d_s(f(p), f(p'))$, on d designa la distància habitual de $\mathbb{R}^2$ i d_s la distància que hem introduït a S (recordeu que f és bijectiva).

Fixem un punt $p \in V$. Considerem a V una circumferència C de centre p i radi ε prou petit per tal que $C \subset V$. Òbviament la imatge per a f de la circumferència C serà el conjunt de punts de l'esfera tals que la distància d'aquests punts a $q = f(p)$ és constant igual a ε. Així com a $\mathbb{R}^2$ la distància entre dos punts es mesura sobre la recta que els uneix, en una esfera la distància entre dos punts es mesura sobre el cercle màxim que els uneix. Així, doncs, si vull veure com és sobre una esfera de radi R el conjunt de punts que disten ε d'un punt donat q, puc ima-

ginar-me que q és un pol d'aquesta esfera des del qual surten meridians (cercles màxims) en totes direccions. Sobre cada un d'aquests meridians prenc una distància ε i, què obtinc? Un paral·lel. Aquest paral·lel és $f(C)$. Però si f és una isometria, la circumferència C de V i el paral·lel $f(C)$ de l'esfera hauran de tenir la mateixa longitud. Però la longitud de C és $2\pi\varepsilon$ i la del paral·lel $f(C)$ de S és $2\pi R\sin(\varepsilon/R)$, que és menor que $2\pi\varepsilon$. D'aquesta manera arribem a contradicció.



6. Apèndix 2: repàs de les nocions bàsiques de funcions holomorfes

Recordem que una funció real f definida en un obert V de $\mathbb{R}$, $f: V \to \mathbb{R}$, es diu analítica quan és desenvolupable en sèrie de Taylor en tot punt de V. Això vol dir que per a tot $x_0 \in V$ es pot trobar un interval $I = (x_0 - \varepsilon, x_0 + \varepsilon)$ contingut a V tal que per a tot $x \in I$ es té

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!} f''(x_0)(x - x_0)^2 + \ldots$$

És un fet remarcable l'existència de funcions infinitament derivables que no són analítiques. Cauchy va donar el següent exemple per tal de posar de manifest aquest fet. Considereu la funció f: $\mathbb{R} \to \mathbb{R}$ definida per

$$f(x) = \begin{cases} e^{-1/x^2} \text{si } x \neq 0 \\ 0 \text{ si } x = 0 \end{cases}$$

Si hom calcula les derivades d'aquesta funció a l'origen hom obté que totes elles són nul·les. O sigui, $0 = f(0) = f'(0) = f''(0) = \ldots$ El polinomi de Taylor $P_n(x)$ de grau n d'aquesta funció a l'origen és idènticament nul per a tot n. La fórmula de Taylor de $f(x)$ fins a l'ordre n a l'origen dirà $f(x) = P_n(x) + T_n(x)$, on $T_n(x)$ és el terme complementari. Com que $P_n(x)$ és sempre nul, el terme complementari $T_n(x)$ és sempre igual a la funció per a tot x. Això mostra que aquesta funció no és desenvolupable en sèrie de Taylor a l'origen (tot i ser infinitament derivable). Per tant, aquesta funció no és analítica.

Dit això, passem ara a veure què s'entén per funció derivable en el cas de funcions definides sobre oberts de $\mathbb{C}$ i que prenen valors a $\mathbb{C}$. Veurem que aquí es produeix una situació molt diferent de la del cas real. El fet de ser una vegada derivable ja implica ser desenvolupable en sèrie de Taylor (o sigui, analítica).

Identificarem el pla complex $\mathbb{C}$ amb $\mathbb{R}^2$. Sigui V un obert de C i sigui $f: V \to \mathbb{C}$ una funció sobre V amb valors complexos. Es diu que f és derivable en un punt $z_0 \in V$ si existeix el límit

$$\lim_{u \to 0,\, u \neq 0} \frac{f(z_0 + u) - f(z_0)}{u}$$

Això equival a dir que f admet una derivada respecte a la variable complexa. Afanyem-nos a dir que aquí s'utilitza una terminologia diferent a la del cas real, i una funció derivable en un punt z_0 s'acostuma a dir que és *holomorfa* en el punt z_0. Una funció es diu holomorfa en un domini V quan és holomorfa (o derivable respecte a la variable complexa) en tots els punts de V. La definició anterior de derivable en z_0 (o holomorfa en aquest punt) es pot reformular de manera equivalent tot dient que existeix $a \in \mathbb{C}$ tal que la funció $\alpha(u) = f(z_0 + u) - f(z_0) - au$ compleix

$$\lim_{u \to 0,\, u \neq 0} \frac{\alpha(u)}{|u|} = 0$$

La a de l'expressió anterior és precisament la derivada complexa $f'(z_0)$. Si posem $z = x + iy$, $u = h + ik$, i considerem f com una funció de les dues variables reals x i y, l'expressió que serveix per definir $\alpha(u)$ s'escriurà així:

$$f(x_0 + h, y_0 + k) - f(x_0, y_0) = a(h + ik) + \alpha(h, k),$$

amb

$$\lim_{(h,k) \to (0,0),(h,k) \neq (0,0)} \frac{\alpha(h,k)}{\sqrt{h^2 + k^2}} = 0$$

Aquesta és justament l'expressió que serveix per definir el concepte de diferenciabilitat en el punt (x_0, y_0) d'una funció de dues variables reals. Es dedueix llavors de la definició que els coeficients de h i de k són precisament les dues derivades parcials de la funció en el punt (x_0, y_0). Com que en el nostre cas aquests dos coeficients són respectivament a i ia, tindrem:

$$\left(\frac{\partial f}{\partial x}\right)_{(x_0,y_0)} = a = f'(z_0) \quad ; \quad \left(\frac{\partial f}{\partial y}\right)_{(x_0,y_0)} = ia = if'(z_0)$$

D'aquí deduïm que f és holomorfa si i només si és diferenciable en tot punt com a funció de dues variables

reals i es compleix la condició (de Cauchy-Riemann) següent:

$$\frac{\partial f}{\partial x} + i\frac{\partial f}{\partial y} = 0$$

El que és un fet sorprenent que ara nosaltres aquí no demostrarem és que la condició anterior ja implica que la funció és infinitament derivable i desenvolupable en sèrie de Taylor complexa en tots els punts del domini V,

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \frac{1}{2!} f''(z_0)(z_0 - z_0)^2 + \dots$$

Recíprocament, tota funció definida per una sèrie de potències convergent és holomorfa.

La raó per la qual les funcions holomorfes s'utilitzen en cartografia és la següent: donen aplicacions conformes d'un domini del pla en un altre domini del pla, si la derivada no s'anul·la en cap punt del domini. Sigui V un obert de $\mathbb{C} \cong \mathbb{R}^2$ i $f : V \to \mathbb{C}$ una funció holomorfa. Suposem que $f'(z) \neq 0$ per $z \in V$. En un punt qualsevol z_0 de V l'aplicació lineal tangent de f en el punt z_0 aplica Z en la part lineal en Z del desenvolupament de Taylor $f(z_0 + Z) - f(z_0) = f'(z_0)\,Z + \dots$ Així doncs, aquesta aplicació lineal tangent serà

$$\begin{array}{ccc} \mathbb{C} \cong \mathbb{R}^2 & \to & \mathbb{C} \cong \mathbb{R}^2 \\ Z & \to & f'(z_0)\,Z \end{array}$$

Aquí $f'(z_0)\,Z$ vol dir la multiplicació del nombre complex $f'(z_0)$ pel nombre complex Z. Aquesta aplicació, com a aplicació de $\mathbb{R}^2$ en $\mathbb{R}^2$, és un gir compost amb una homotècia, ja que l'argument de $f'(z_0)\,Z$ és suma dels arguments de $f'(z_0)$ i de Z i el mòdul és el producte de mòduls. Un gir compost amb una homotècia conserva els angles. És, doncs, una aplicació conforme.

Les eines matemàtiques de la geomàtica

Ismael Colomina

1. Introducció: la geomàtica

La geomàtica és el conjunt disciplinari de ciències i tecnologies que tracten de l'estudi, adquisició, emmagatzematge, organització, anàlisi, disseminació, gestió i explotació de la informació espacial referenciada geogràficament. Les disciplines geomàtiques inclouen la cartografia, la fotogrametria i la teledetecció, la geodèsia, la topografia i els sistemes d'informació geogràfica.

Les disciplines geomàtiques han tingut, de sempre, importants aplicacions en la metrologia industrial i en la navegació. La geomàtica és multidisciplinària. El seu perfil és clàssic i modern alhora. Hereta el rigor de disciplines clàssiques com la geodèsia, la fotogrametria i la cartografia. Incorpora el vigor i la capacitat d'innovació de noves disciplines com la teledetecció, les tecnologies de la informació i les comunicacions. La geomàtica és, en totes les seves facetes, en forta expansió.

Les principals àrees d'aplicació de la geomàtica són la gestió i planificació d'infraestructures del territori, la gestió del medi ambient, la gestió dels recursos naturals i, darrerament, la navegació precisa, particularment la navegació de vehicles terrestres.

L'evolució de la geomàtica està dominada per les tendències de la tecnologia i del mercat.

Les grans tendències tecnològiques són: el ràpid desenvolupament de les tecnologies de la informació i les conseqüències sobre el cost, les dimensions, la capacitat i la velocitat de procés i adquisició de dades; la disponibilitat d'equips GPS (Sistema de Posicionament Global) barats i precisos; el desenvolupament de la tecnologia de sensors i de fusió de dades; la creixent facilitat d'accés a imatgeria aèria i espacial; i el desenvolupament de sistemes de gestió de la informació, en particular els GIS (Sistemes d'Informació Geogràfica).

Les grans tendències de mercat geomàtic són: el creixent ús i necessitat d'informació geogràfica en molt diversos sistemes de gestió; la importància del sector governamental com a principal client de serveis geomàtics; les noves oportunitats degudes a les aplicacions mediambientals i la navegació; l'impacte del gran cost de nodrir –adquirir informació i mantenir-la al dia– les bases cartogràfiques, i el ràpid creixement de les necessitats geomàtiques a l'Europa de l'Est, Àsia i Llatinoamèrica.

2. Geodèsia

Dins la geomàtica, la geodèsia proporciona els marcs geomètric i físic, i les eines per a la georeferenciació. La geodèsia és la ciència que estudia la figura i el camp gravitatori de la Terra, així com la seva variació en el temps (Helmert, 1962; Torge, 1980). La figura de la Terra es descriu mitjançant conjunts de coordenades dels punts de la seva superfície. Els camps gravitatori i de la gravetat es descriuen mitjançant el seu potencial o funcions del potencial (Torge, 1980). Tant la figura com el potencial són funcions que depenen del temps, si més no en els marcs de referència globals, físic i geomètric.

El coneixement de la figura i el camp gravitatori de la Terra són necessaris per a la realització de cartografia i per al seu bon ús. A més, el coneixement de la seva variació ajuda a conèixer els processos dinàmics del nostre planeta, com els moviments de les plaques continentals, la puja del nivell del mar i altres efectes del canvi climàtic.

La geodèsia es relaciona amb altres disciplines com la geofísica, les ciències de l'espai, l'astronomia, l'oceanografia, la meteorologia i la geologia. Però els seus fonaments teòrics descansen en la matemàtica i la física (Vanicek i Krakiwsky, 1986). Com en la resta de disciplines geomàtiques, la informàtica n'és un dels instruments fonamentals.

La geodèsia, a diferència d'altres subdisciplines geomàtiques, pertany al grup de ciències que com la mecànica, l'astronomia, etc. han estat sempre enterament matematitzades. En particular, ha contribuït al desenvolupament de disciplines matemàtiques com la geometria diferencial. En general, la geodèsia ha desenvolupat i/o emprat eines matemàtiques per dues raons ben diferenciades. En uns casos, ha calgut desenvolupar eines per processar les observacions realitzades amb els instruments proveïts per la tecnologia del moment. És el cas de la geodèsia espacial (Kaula, 1966; Seeber, 1993), o de la geodèsia inercial (Schwarz, 1983), o de les xarxes geo-

dèsiques, tant en el seu aspecte numèric (Schwarz i Wade, 1990) com estadístic (Baarda, 1976). En altres casos, les eines han estat més directament relacionades amb els problemes propis de la geodèsia. És el cas dels models de la figura i el camp gravitatori de la Terra (Heiskanen i Moritz, 1967).

No es pot deixar de mencionar el llarg i apassionant camí recorregut per la ciència per a la determinació de la figura de la Terra. Des de la polèmica entre Newton i Cassini sobre la forma de l'el·lipsoide que l'aproxima (definit per dos paràmetres) fins als models actuals del geoide (definits per sèries d'esfèrics harmònics fins a grau i ordre 360, o sigui per uns 130.000 paràmetres) passant pels coneguts treballs de G. G. Stokes per a la solució de l'anomenada equació fonamental de la geodèsia física, de fet, un problema de contorn de la teoria del potencial. (Cal recordar que no es disposà de la primera xarxa geodèsica mundial tridimensional fins fa només vint-i-cinc anys [Schmid, 1974].)

Tant en els casos de tractament de dades observacionals com en el de determinació dels models geodèsics terrestres, la geodèsia ha fet un fort ús de tot tipus d'eines d'estimació, particularment del mètode dels mínims quadrats inventat per K. F. Gauss el 1795 (Gauss, 1967). Cal destacar també l'esforç realitzat en el desenvolupament de les eines estadístiques de la fiabilitat del qual és màxim exponent el geodesista holandès W. Baarda (Baarda, 1967). Les idees dels geodesistes sobre la fiabilitat han estat en bona part adoptades recentment per la comunitat de la navegació amb el nom de *integritat*.

3. Teledetecció i fotogrametria

La teledetecció i la fotogrametria tracten de l'obtenció d'informació geomètrica, espectral i semàntica d'objectes a partir d'imatges bidimensionals. Avui, la fotogrametria és la principal eina de la cartografia per a l'elaboració de mapes a mitjana i gran escala. La teledetecció, inicialment basada en satèl·lits d'observació de la Terra i posteriorment basada també en avions d'observació, és cada vegada més usada com a complement o alternativa a la fotogrametria tradicional. El desenvolupament de sensors multi i hiperespectrals aerotransportables i de sensors fotogramètrics digitals fa que la teledetecció i la fotogrametria convergeixin en una sola disciplina d'observació de la Terra mitjançant imatges.

Des que una imatge es registra fins que es converteix en un mapa o farceix una base d'informació cartogràfica, s'han de seguir algunes etapes fortament matematitzades. La imatge s'ha de normalitzar; és a dir, cal determinar la seva posició i orientació i cal determinar les característiques de la propagació del senyal electromagnètic entre l'objecte emissor d'interès i el sensor. Després la imatge ha d'ésser interpretada, geomètricament i semànticament (Greve, 1996). Cada vegada més, es combinen les característiques complementàries de diferents sensors i, com a conseqüència, s'està investigant el problema de la integració de sensors i la fusió de dades heterogènies (Núñez *et al.*, 1999).

Tots els processos anteriors es basen en l'ús de models matemàtics, físics i cartogràfics complexos. La determinació de la posició i orientació de les imatges és un problema a cavall entre la geodèsia i l'observació de la Terra. Per a aquesta tasca s'utilitzen la geodèsia de satèl·lits, les tècniques de navegació inercials, les tècniques de procés d'imatge i, normalment, tècniques numèriques de solució de grans sistemes d'equacions que, a la vegada, recorren a eines combinatòries (Slama, 1980).

La interpretació geomètrica o restitució de les imatges és la tasca de l'observació de la Terra que determina la figura dels objectes que apareixen en les imatges. Per a això s'acostuma a recórrer a models de la geometria projectiva –les equacions de col·linealitat fotogramètriques–, a tècniques de correspondència d'imatges i, darrerament, a eines de la geometria computacional (Berg *et al.*, 1997). La interpretació geomètrica combinada amb el que es podria denominar «interpretació espectral» produeix les ortoimatges (del territori), una eina extremadament útil i de producció barata.

Un cas especialment interessant i espectacularment precís d'interpretació geomètrica de les imatges és la interferometria radar que, per exemple, és capaç de determinar variacions milimètriques en l'escorça terrestre després d'un terratrèmol (Massonet *et al,* 1994).

La interpretació semàntica és la més jove de les tasques mencionades anteriorment i es basa en multitud de tècniques matemàtiques, de procés d'imatge i d'intel·ligència artificial. A causa d'aquesta joventut, a la complexitat dels models cartogràfics i, en general, a la manca de models de l'espai-objecte en què es realitza la interpretació, aquest és un dels camps amb més activitat investigadora en geomàtica (Förstner, 1993).

A més dels recursos matemàtics clàssics mencionats o omesos en els paràgrafs anteriors, la teledetecció i la fotogrametria moderna estan fent un creixent ús d'eines com la morfologia matemàtica, les *wavelets*, i altres.

4. Cartografia

La cartografia matemàtica és un concepte ben establert de fa temps. És, sobretot, la teoria de les projeccions cartogràfiques, que és una aplicació de la geometria diferencial (Snyder, 1987). És un tema objecte de limitada recerca. Tornem, pero, als principis.

Diuen els diccionaris i les enciclopèdies que la cartografia és l'art, la ciència i la tècnica de fer mapes. Dels mapes, diuen que són representacions planes i reduïdes d'una part o de tota la superfície terrestre. Es poden trobar definicions més precises –de la cartografia i dels mapes– en els llibres de cartografia, però tenen l'incon-

venient de ser fetes pels mateixos cartògrafs. La cartografia, com tantes altres àrees especialitzades de l'activitat humana, ha anat evolucionant des d'obscurs orígens (Brown, 1949) fins al seu estat actual (Arbiol *et al.*, 1998). En el context d'aquest article és convenient mirar-se-la com qualsevol altra ciència i, en particular, com la ciència que estudia la superfície terrestre –i d'altres planetes– i la representa. En termes més precisos, els cartògrafs són especialistes en la creació de models de determinats aspectes de la superfície terrestre i de representar-los. Depenent del tema del mapa, els cartògrafs han de col·laborar amb altres col·lectius, com els geòlegs o els biòlegs, en la recerca dels models i representacions apropiades. Més enllà de l'aspecte essencial de la seva feina –el modelatge i la representació cartogràfica– els cartògrafs han hagut d'adquirir el domini de les tecnologies auxiliars per poder fer els mapes. Des de tècniques de dibuix, fins a tècniques d'impressió, i avui, fins al conjunt d'eines que les tecnologies de la informació han posat al seu abast.

En general, quan una nova tecnologia entra en una determinada àrea de l'activitat científica s'obren nous camins. Però cal no cometre dos errors prou freqüents: confondre el mitjà tecnològic amb el fi, i no adquirir el domini necessari de la nova eina. Les tecnologies de la informació, la informàtica, són l'instrument per tal que la ciència cartogràfica s'alliberi de les limitacions imposades per les tecnologies anteriors. Amb l'ús dels sistemes d'informació (geogràfica) es poden materialitzar complexíssims models cartogràfics i es poden separar, conceptualment i pràctica, de la seva representació. I aquesta representació no s'ha de fer necessàriament sobre paper, ni necessàriament a una escala i amb uns continguts predeterminats.

Les noves eines informàtiques faciliten –en la pràctica, permeten– el tractament o procés de la informació cartogràfica. I el tractament, com en la resta de les ciències, s'ha de basar en models matemàtics de la realitat cartogràfica específica. En la gran majoria dels casos no es disposa d'aquests models. Aquest és el desafiament i, alhora, l'oportunita, tant per a la cartografia com per a la matemàtica.

Un exemple relacionat és el problema de la generalització cartogràfica. Atès un mapa a una determinada escala –per exemple, 1:5.000– és possible generar automàticament un mapa a una escala menor –per exemple, a escala 1:10.000 o 1:25.000?–. En general, malgrat que aparentment un mapa a escala 1:5.000 conté tota la informació necessària per a un mapa a qualsevol escala menor, la resposta és no. Depèn de l'existència de models cartogràfics per a cada escala i de si existeixen lleis per a la transformació d'aquests models. A Baella i Pla (1999) es discuteix aquest problema per a la cartografia topogràfica de Catalunya a escales 1:5.000 i 1:10.000.

Un exemple relacionat amb l'anterior és la representació del relleu a diferents escales. Es pot considerar un subproblema del problema de la generalització cartogràfica restringit a la superfície geomètrica de la Terra. Malgrat tractar-se d'una qüestió «simple» de modelatge i representació d'una superfície, el problema encara és lluny d'estar solucionat. En el tema del modelatge del relleu, la tecnologia de gràfics per ordinador i la geometria computacional estan guanyant un protagonisme creixent (Berg *et al.* 1997).

Navegació

En els darrers anys la navegació, amb l'ús creixent del GPS, ha deixat d'ésser una especialitat dels marins, aviadors i astronautes per convertir-se en una activitat quotidiana de molts ciutadans. La navegació, al llarg de la seva història i al llarg de les seves tecnologies instrumentals, ha estat un camp propici a l'ús de la matemàtica (Maybeck, 1979). Ara, amb el GPS i la cartografia digital, la navegació terrestre és una disciplina difícilment separable de la geomàtica.

La navegació terrestre precisa és una tasca especialment complicada que necessita del posicionament cinemàtic geodèsic, de la cartografia navegable i, segons el nivell d'automatització, de tecnologies de visió per ordinador i intel·ligència artificial. Aquest és el cas de la navegació automàtica amb obstacles que han de realitzar alguns robots. Pel que fa a les eines matemàtiques emprades per la part geomàtica de la navegació cal destacar els sistemes dinàmics estocàstics (SDS) i les equacions diferencials estocàstiques (SDE) juntament amb els mètodes seqüencials d'estimació òptima adaptats als càlculs en temps real com el filtratge de Kalman (Kalman, 1960).

En termes generals, actualment, la navegació precisa descansa en el posicionament per satèl·lits i en els INS (Sistemes de Navegació Inercials). S'acostuma a utilitzar els INS com a font d'observacions d'alta freqüència per generar l'SDS. El GPS proporciona posicions a una freqüència menor que permeten el calibratge dels sensors de l'INS i, així, assolir fins a precisions centimètriques en la determinació de trajectòries. La navegació «de consum» enfoca el problema altrament. Usa el GPS com a font per generar un SDS de baixa freqüència, que en entorns urbans s'ha de corregir amb dades cartogràfiques.

A tall de conclusió

Entre els geodesistes encara es recorda l'afirmació que l'any 1902 féu F. Klein: «la geodèsia és un magnífic exemple del que es pot fer amb la matemàtica en les aplicacions pràctiques i de com s'ha de fer». Després de gairebé un segle, l'afirmació de Klein no només continua essent vàlida sinó que, a més, és a prop de poder-se aplicar a tota la geomàtica. Depèn dels matemàtics.

Referències

[1] ALBIOL, R.; COLOMER, J. L.; COLOMINA, I.; LLEOPART, A.; PLA, M. (1998). *Les noves eines de la cartografia: cap al segle XXI.* L'Avenç, núm. 224, p. 65-70.

[2] BAARDA, W. (1967). *Statistical concepts in geodesy.* Netherlands Geodetic Comission, Delft.

[3] —; (1976). Reliability and precision of networks. Delft Geodetic Institute, Delft.

[4] BAELLA, B.; PLA, M. (1999). *Map generalization to obtain the topographic map of Catalonia 1:10000.* Article presentat al Workshop «Progress in Automated Map Generalization», International Cartographic Association, Ottawa.

[5] BERG, M. DE; KREVELD, M. VAN; OVERMARS, M.; SCHWARZKOPF, O.; FREYSCHMIDT, J. (1997). Computational geometry, algorithms and applications. Berlin, Springer.

[6] BROWN, L. A. (1949). *The story of maps.* New York, Dover.

[7] FÖRSTNER, W. (1993). *Feature extraction in digital photogrammetry.* Photogrammetric Record, Vol. 82, núm. 14, p. 595-611.

[8] GAUSS, K. F., 1967 (1809). *Theory of motion of the heavenly bodies.* New York, Dover.

[9] GREVE, C. [ed.] (1996). *Digital photogrammetry, an addendum to the Manual of Photogrammetry*, ASPRS, MA, Bethesda.

[10] HEISKANEN, W. A.; MORITZ, H. (1967). Physical Geodesy. Freeman.

[11] HELMERT, F. R. 1962 (1880). *Die matematischen und physikalishen Theorien der höheren Geodäsie.* Vol. 1, Minerva GmbH.

[12] KALMAN, R. E. (1960). «A new approach to linear filtering and prediction problems». *Transactions of the ASME. Journal of Basic Engineering*, Vol. 82, p. 34-45.

[13] KAULA, W. (1966). *Theory of satellite geodesy: applications of satellites to geodesy.* Blaisdell.

[14] KLOEDEN, P. E.; PLATEN, E. (1992). *Numerical solution of stochastic differential equations.* Berlin, Springer.

[15] KRAMER, H. J. (1996). *Observation of the Earth and its environment: survey of missions and sensors.* Berlin, Springer.

[16] LEICK, A. (1995). *GPS satellite surveying.* New York, John Wiley and Sons.

[17] MASSONET, D.; FEIGL, K.; ROSSI, M.; ADRAGNA, F. (1994). «Radar interferometric mapping of deformation in the year after the Landers earthquake», *Nature,* Vol. 369, p. 227-230.

[18] MAYBECK, P. S. (1979). *Stochastic models, estimation and control.* Mathematics in Science and Engineering, Vol. 141, New York, Academic Press.

[19] NÚÑEZ, J.; OTAZU, X.; FORS, O.; PRADES, A.; PALÀ, V.; ARBIOL, R. (1999). «Image fusion with additive multiresolution wavelet decomposition. Applications to SPOT+Landsat images». *Journal of the Optical Society of America*, Vol. 16, núm. 3, p. 467-474.

[20] SCHMID, H. H. (1974). «Worlwide geocentric satellite triangulation». *Journal of Geophysical Research*, Vol. 79, núm. 35, p. 5348-5376.

[21] SCHWARZ, C.; WADE, E. B. (1990). «The North American datum of 1983: project methodology and execution». *Bulletin Géodesique*, Vol. 64, p. 28-62.

[22] SCHWARZ, K. P. (1983). «Inertial surveying in geodesy». *Reviews of Geophysics and Space Physics*, Vol. 21, núm. 4, p. 878-890.

[23] SEEBER, G. (1993). *Satellite geodesy.* Berlin, Walter de Gruyter.

[24] SLAMA, C. C. (1980). *Aerotriangulation. Manual of photogrammetry.* ASPRS. VA, Falls Church.

[25] SNYDER, J. P. (1987). *Map projections –a working manual.* US Geological Survey

[26] TORGE, W. (1980). *Geodesy.* Berlin, Walter de Gruyter.

[27] VANICEK, P.; KRAKIWSKY, E. (1986). *Geodesy, the concepts.* Amsterdam, North-Holland.

Notícia dels autors

Jaume Aguadé és catedràtic de Matemàtiques a la Universitat Autònoma de Barcelona i director del grup de recerca en topologia algebraica.

Ismael Colomina és doctor en Matemàtiques per la Universitat de Barcelona. Entre 1982 i 1998 ha treballat a l'Institut Cartogràfic de Catalunya, i ha estat cap de geodèsia d'aquesta institució des de l'any 1990. Actualment és director de l'Institut Català de Geomàtica.

José Luis Fernández és doctor per la Washington University de St. Louis i Catedràtic d'Anàlisi Matemàtica de la Universitat Autònoma de Madrid. Ha estat professor de les universitats de Wisconsin i Maryland. Ha desenvolupat la seva activitat d'investigador en els camps de l'anàlisi complexa i la geometria hiperbòlica. Actualment treballa en les aplicacions de les matemàtiques a les finances. És consultor associat a Analistas Financieros Internacionales (AFI).

Joan Girbau és catedràtic de Matemàtiques a la Universitat Autònoma de Barcelona i membre de l'Institut d'Estudis Catalans. Ha desenvolupat una activitat de recerca molt important dins del camp de la geometria diferencial i la teoria de la relativitat. Posseeix la medalla Narcís Monturiol al mèrit científic i tecnològic.

Paz Morillo és llicenciada en Matemàtiques i doctora en Informàtica per la Universitat Politècnica de Catalunya. Des de 1991 és directora del grup de recerca Matemàtica Aplicada a la Criptografia, a la Universitat Politècnica de Catalunya.

Jesús Mosterín és catedràtic de Lògica i Filosofia de la Ciència a l'Institut de Filosofia del Consell Superior d'Investigacions Científiques. És membre de diverses acadèmies internacionals de filosofia i és autor d'una vintena de llibres sobre lògica i filosofia.

Helmut Neunzert és catedràtic de Matemàtiques a la Universitat de Kaiserslautern i doctor *honoris causa* per la Universitat de Linz. El 1995 va fundar l'Institut de Matemàtica Tecnològica i d'ençà d'aleshores n'és el director. És el president del Consorci Europeu de Matemàtica Industrial (ECMI).

Carles Padró és llicenciat en Matemàtiques per la Universitat de Barcelona i doctor en Matemàtiques per la Universitat Politècnica de Catalunya. Desenvolupa la seva tasca d'investigació dins del grup de recerca Matemàtica Aplicada a la Criptografia, a la Universitat Politècnica de Catalunya.

ISBN 84-95166-10-0

CRM
CENTRE DE RECERCA MATEMÀTICA

INSTITUT D'ESTUDIS CATALANS
BARCELONA

FUNDACIÓ
Caixa de Sabadell



Campus de Bellaterra, Edifici C
08193 Bellaterra, Spain
Tel.: +34 93 581 10 81
Fax: +34 93 581 22 02
crm@crm.cat
www.crm.cat