Título:	Condiciones de ε-eficiencia en optimización vectorial.
Doctorando:	César Gutiérrez Vaquero.
Director/es:	Vicente Novo Sanjurjo.
Defensa:	18 de Noviembre de 2004, U.N.E.D.
Calificación:	Sobresaliente Cum Laude por unanimidad.

Resumen: En esta memoria se estudian las soluciones aproximadas en problemas de optimización vectorial. Su contenido se enmarca en la resolución de este tipo de problemas a través de conceptos de solución que generalizan la noción clásica de eficiencia.

En los tres primeros capítulos se analiza la noción de "elemento aproximado" o ε-eficiente del conjunto final de un problema de optimización vectorial (POV). Se interpreta cada conjunto de puntos ε-eficientes como imagen de una multifunción y se introducen varias propiedades estos elementos que unifican otras analizadas en la literatura y que permiten estudiar distintas conceptos de ε-eficiencia. El estudio de dichas propiedades ha motivado la introducción de una nueva noción de ε-eficiencia y de una propiedad que relaciona las soluciones exactas y ε-eficientes.

Se establecen condiciones necesarias y suficientes que relacionan las soluciones aproximadas obtenidas por escalarización con los elementos ε-eficientes del conjunto final de un POV. Estas condiciones extienden y complementan algunas de las descritas en la literatura, bien porque se consiguen en un contexto más general (no paretiano y no convexo), o bien porque se refieren a conceptos de ε-eficiencia no analizados hasta el momento en este aspecto. En la demostración de las condiciones necesarias se emplean distintas versiones del funcional de Minkowski (funciones calibre o "gauge") y para las suficientes se utilizan funciones con alguna propiedad de monotonía generalizada.

Las condiciones necesarias extienden los procedimientos de separación convexos y no convexos clásicos de forma que, implícitamente, constituyen lo que podrían denominarse teoremas de ε-separación. Los procedimientos de escalarización analizados nos permiten caracterizar las soluciones eficientes aproximadas de un POV mediante soluciones aproximadas de problemas de optimización escalar y motivan la definición de un concepto de parametrización para la ε-eficiencia. Se obtienen diversas parametrizaciones para distintas nociones de ε-eficiencia.

En el capítulo 5, el interés se centra en la caracterización de las soluciones eficientes aproximadas del POV mediante condiciones de tipo Fritz John y de tipo Kuhn-Tucker. Las reglas de multiplicadores para las soluciones de problemas de optimización constituyen un tema central en las teorías de optimización. Aquí se analiza esta cuestión en problemas multiobjetivo paretianos convexos con restricciones abstractas, de desigualdad e igualdad. Previamente, en el capítulo 4, se obtiene una regla de la cadena para la ε-subdiferencial que está motivada por las técnicas de escalarización empleadas.

Women in Mathematics

International Council for Industrial and Applied Mathematics

http://www.iciam.org

Ante los comentarios realizados por el Presidente de la Universidad de Harvard insinuando que las mujeres tienen capacidades inferiores a los hombres para las Ciencias y las Matemáticas, ICIAM ha hecho público el siguiente comunicado mostrando su desacuerdo con tales afirmaciones.

ICIAM Statement on Women in Mathematics

Recent remarks of the President of Harvard University have led to media speculation that innate differences in the mathematical abilities of men and women make it less likely that women will succeed in science and mathematics. ICIAM does not accept this notion.

ICIAM members are well aware that there are many barriers (whether financial, cultural, or practical) that face women who want to pursue mathematical or scientific careers at the highest levels. The unbroken career paths that are typical of successful male careers in mathematics take no account of the specific responsibities of women related to child bearing and family.

As an international organisation representing the world's applied mathematicians, ICIAM is committed to removing the educational inequalities in mathematics that exist in many parts of the world, and to improving the access to careers in the mathematical sciences for all men and women. ICIAM highlights the accomplishments of all applied mathematicians, women and men, at our quadrennial International Congress, the premier event world-wide in applied and industrial mathematics.

ICIAM, the International Council for Industrial and Applied Mathematics, is the world organisation for applied mathematics and computational science. Its members are mathematical sciences societies based in more than 20 countries. For more information, see the Council's web page at http://www.iciam.org/

SEMA, como miembro de ICIAM, valora positivamente la iniciativa y hace suyo el comunicado en los términos generales del mismo. No obstante, en

el comunicado se indica que las mujeres tienen responsabilidades específicas, relacionadas con el cuidado de los niños y la familia. Entendemos que en la sociedad actual tal afirmación no tiene cabida. El hecho de que a lo largo de la historia se haya asignado a las mujeres esa tarea ha supuesto para ellas una dificultad adicional para su participación en el desarrollo de la ciencia y, en particular, de las matemáticas. Por ello, es responsabilidad de la comunidad científica, que no puede evadirse de los problemas sociales, situarse en la vanguardia de la defensa de la igualdad de oportunidades.

Breve presentación del Departamento de Matemática Aplicada de la Universidad de Salamanca

Resumen

Iniciamos en este número una nueva sección en la que invitaremos a los Departamentos de Matemática Aplicada y otros Departamentos cuya actividad esté relacionada con la misma a realizar una presentación de su historia y actividades con el fin de darlas a conocer a la comunidad científica.

El Departamento de Matemática Aplicada de la Universidad de Salamanca (http://matapli.usal.es) fue creado por resolución de la Junta de Gobierno del 30 de octubre de 1997 agrupando en su seno a los profesores del área que en aquel entonces estaban repartidos, por una parte, en el Departamento de Matemática Pura y Aplicada y, por otra, en el Departamento de Estadística y Matemática Aplicada. Consecuencia de esta reorganización, el profesorado del área se agrupó en un solo departamento, lo que mejoraba substancialmente las posibilidades de desarrollo del área al tener mayor autonomía en la toma de decisiones y disponer de más recursos económicos.

En el momento de su creación el Departamento lo formaban: 1 C.U., 2 P.T.U., 7 P.T.E.U. (de los cuales uno era doctor), 3 Ayudantes de E.U., 3 Asociados, 1 Administrativo. En la actualidad, forman parte del mismo: 1 C.U., 1 C.E.U., 5 P.T.U., 10 P.T.E.U. de los cuales 5 son doctores, 1 Ayudante doctor, 2 Profesores Colaboradores, 4 Asociados, 1 Becario de F.P.I. y 1 Administrativo.

El Profesorado del departamento realiza su actividad docente en distintos Centros y muy diversas Titulaciones: Así, se imparten las materias adscritas al área en la Licenciatura de Matemáticas, Diplomatura de Estadística, Ingeniero Geólogo de la Facultad de Ciencias, Ingeniero Químico y Licenciado en Químicas de la Facultad de Ciencias Químicas, Licenciado en Ciencias Ambientales e Ingeniero Técnico Agrícola de la Facultad de Ciencias Agrarias y Ambientales, todas ellas en Salamanca. En la Escuela Técnica Superior de Ingenieros Industriales de Béjar, en Ingeniería Técnica Industrial (especialidades de Electricidad, Electrónica, Mecánica y Textil) y el segundo ciclo de Ingeniería Industrial Superior. En la Escuela Politécnica Superior de Zamora en Ingeniería Técnica Industrial (especialidad Mecánica), Ingeniero Técnico en Obras Públicas (especialidad Construcciones Civiles), Ingeniero Técnico Agrícola (Especialidad en Industrias Agrarias y Alimentarias), Arquitecto Técnico, Ingeniero Técnico en Informática de Gestión y el segundo ciclo de Ingeniería Superior de Materiales. En la Escuela Técnica Superior de

Ávila, los estudios correspondientes a los títulos de Ingeniero Técnico de Obras Públicas (especialidad de Hidrología), Ingeniero Técnico en Topografía, Ingeniero Técnico de Minas (especialidad en Sondeos y Prospecciones Mineras) y el 2º ciclo de Ingeniero Superior en Geodesia y Cartografía.

Durante su breve andadura se han puesto en marcha en el Departamento diversas líneas de investigación alrededor de varios grupos de investigación, habiendo concurrido con éxito a las convocatorias competitivas de proyectos tanto regionales como nacionales, y dando lugar a colaboraciones externas con grupos e investigadores extranjeros. Los grupos de investigación son:

- **Grupo de Simulación Numérica y Cálculo Científico** (SINUCC).

  (http://matapli.usal.es/sinucc.html)

  Sus principales líneas de investigación son los Métodos numéricos para problemas de convección-difusión-reacción, métodos adaptativos para E.D.P. y la modelización numérica de problemas medioambientales. El grupo mantiene relaciones externas en el ámbito nacional: LACAN (Laboratori de Càlcul Numèric, UPC), IUSIANI (Instituto Universitario de Sistemas Inteligentes y Aplicaciones Numéricas en Ingeniería, ULPGC), IRNASA (Instituto de Recursos Naturales y Agrobiología, CSIC), y en el ámbito internacional con el IMATI-CNR (Istituto di Matematica Applicata e Tecnologie Informatiche, Consiglio Nazionale delle Ricerche, Pavia), Department of Mathematics and Institute for Physical Science and Technology, U. of Maryland, Laboratoire de Mathématiques Apliquées, Université Blaise Pascal, Clermont Ferrand.

- **Grupo de Investigación en Autómatas celulares y Aplicaciones**(GIACA).

  (http://matapli.usal.es/investig.html)

  Sus principales líneas de investigación son los Autómatas Celulares y sus aplicaciones en Criptografía, problemas de medioambiente, etc. El grupo mantiene una estrecha colaboración con el Departamento de Tratamiento de la Información y Codificación del Instituto de Física Aplicada del CSIC, Universidad Politécnica de Madrid, Universidad Técnica de Bratislava y Buskerud University College de Noruega.

- **Grupo de Investigación en Métodos Numéricos en E.D.O. y Métodos Adaptados**

  Con diversas colaboraciones nacionales e internacionales.

Otros líneas de investigación provienen de la actividad anterior de los profesores antes de su incorporación al departamento y cuya actividad se mantiene en estrecha colaboración con investigadores de otras áreas de conocimiento, como es el caso de las líneas de investigación sobre Solitones en Teoría de Campos y sus aplicaciones en colaboración con el área de Física Teórica. Relatividad y teorías gauge con profesores de Geometría Diferencial y Física Teórica. Metereología con Física del aire y de la atmósfera.

En el ámbito organizativo el departamento organizó el XVII CEDYA / VII CMA, que tuvo lugar en septiembre de 2001, lo que supuso un reto para todos los miembros del departamento a los pocos años de su creación.

Dentro de este espíritu de servicio a la comunidad científica destacar la labor desinteresada de un grupo de profesores del departamento que ha asumido la edición del Boletín de SëMA, del que el presente número es la primera muestra.

En su todavía corta vida han pasado por el departamento en calidad de profesores visitantes, los profesores Carlos Conca de la Universidad de Chile, y el profesor Jacques Simon de la Universidad de Clermont-Ferrand, y nos han visitado para la impartición de cursos cortos o seminarios, entre otros, el Pr. Enrique Zuazua (UCM), Pr. Mario Ahues (U. Saint Etienne), Pr. Giancarlo Sangalli (IMATI-CNR Pavia), Pra. Cecilia Rivara (U. de Chile), los profesores Rodolfo Rodriguez, Mauricio Barrientos y la Pra. Galina García de la U. de Concepción de Chile, Pr. Askold Peremelov (Institut for Theoretical and Experimental Physics, Moscou), Pr. Erwin Carlos Hernández (U. de Chile), Pr. Raúl Durán Díaz (U. Alcalá de Henares), Dr. Luis Hernández Encinas (CSIC).

También y dentro de la dinámica propia de los grupos de investigación en el departamento se promueven estancias de investigación de los profesores del departamento en centros internacionales del ámbito de la matemática aplicada, así profesores del departamento han realizado estancias de larga duración en el IMATI-CNR Pavia, U. de Maryland (Department of Mathematics), CALTECH (California Institute of Technology, Control and Dynamical Systems Department), Centre of Mathematical Science (U. of Cambridge), así como en diversos centros de universidades españolas y el CSIC.

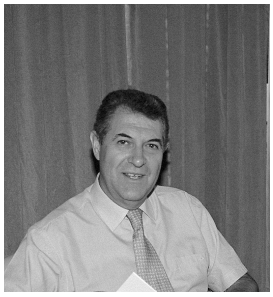
El departamento no tiene programa propio de tercer ciclo, sin embargo participa en dos programas de tercer ciclo interdepartamentales: El programa de Doctorado en Física y Matemáticas con los departamentos de Matemáticas, Física Aplicada, Física, Ingeniería y Radiología Médica y el programa de Doctorado Ciencia y Tecnología de la Ingeniería Geodésica y Cartográfica en el que participan también el departamento de Ingeniería Cartográfica de la U. de Salamanca y departamentos de las universidades Complutense y Alcalá de Henares de Madrid, U. Politécnica de Valencia y de la U. de Cantabria.

La actividad de nuestro profesorado se completa también con una significativa oferta de cursos extraordinarios.

En la actualidad, el departamento afronta nuevos retos: Dada la reforma inminente de los planes de estudios y la amplia participación de nuestro departamento en muchos de ellos, se requerirá una reorganización profunda de la enseñanzas impartidas. También un cambio de mentalidad por parte del profesorado ante la reinterpretación del crédito y la nueva estructura de los

ciclos formativos. Por otra parte, los mecanismos de evaluación de la calidad y criterios objetivos en la evaluación de rendimientos que se nos aplicarán en breve desde instancias superiores nos obliga a optimizar recursos, a reorganizar nuestra actividad, estableciendo protocolos de actuación y mecanismos de control de la misma. En el presente curso estamos poniendo en marcha los protocolos de evaluación de la actividad docente e investigadora de los miembros del departamento. Esto será, sin duda, la base sobre la que mejorar nuestra actividad tanto docente como investigadora y la condición necesaria para que nuestro servicio y aportación a la comunidad universitaria y científica sea significativa y de calidad.

Una mirada al Centre de Recerca Matemàtica



Prof. M. Castellet

Resumen

Se presenta en este breve artículo una descripción de la génesis, desarrollo y estado del *Centre de Recerca Matemàtica*. Los datos han sido proporcionados por el Prof. Manuel Castellet, a quien el Grupo Editor agradece las facilidades prestadas.

En 1964, Beno Eckmann creaba en la ETH de Zürich el *Forschungsinstitut für Mathematik* (FIM). Veinte años más tarde, Manuel Castellet introducía esta estructura de investigación en Barcelona, creando el *Centre de Recerca Matemàtica* (CRM).

Ambos Institutos poseen unas características básicas análogas, al no disponer de personal investigador propio permanente, fundamentándose en los investigadores visitantes y becarios postdoctorales. La idea esencial es establecer un verdadero "laboratorio matemático", en el que el contacto personal y la transmisión de conocimiento entre investigadores en campos afines aportan, sin duda, grandes beneficios al progreso de la investigación matemática, al tratarse ésta de una Ciencia que para muchas tareas apenas requiere de instrumental y en cambio se basa mucho más en las capacidades de los individuos.

Otra característica que comparten los dos Institutos es el bajo coste de su infraestructura, lo que permite dedicar un elevado porcentaje de su presupuesto

directamente a los investigadores y a las actividades que facilitan y estimulan su trabajo.

Sin embargo, existe una diferencia esencial entre ambos centros: mientras que el FIM depende de la ETH, el CRM fue concebido como un Instituto independiente de las Universidades catalanas (si bien trabaja en contacto con todas ellas y contribuye a fortalecer todos los grupos y todas las líneas de investigación), aunque físicamente esté ubicado en el campus de la Universitat Autònoma de Barcelona. En sus inicios, el CRM actuó jurídicamente bajo la tutela del *Institut d'Estudis Catalans*, la primera institución académica de Cataluña y, desde el año 2002, ya con personalidad jurídica propia, bajo la forma de un consorcio entre esta institución y el Gobierno de Cataluña.

Un *Director*, posteriormente con la colaboración de dos *Adjuntos de Dirección* y un *Consejo Científico Asesor*, y un reducido equipo de administración y secretaría, que ha pasado de una a cinco personas, ,llevan a cabo todo el trabajo que no corresponde ni debería corresponder a los investigadores con un único objetivo: establecer o facilitar las condiciones ideales para el trabajo de éstos.

A lo largo de los años, la tipología de las actividades ha ido adaptándose no sólo a las disponibilidades presupuestarias sino, sobre todo, a los intereses de los matemáticos catalanes. La actividad del CRM, iniciada en septiembre de 1984, se ha basado siempre en acoger a investigadores foráneos que trabajen conjuntamente con los locales, potenciando así grupos ya más o menos consolidados o promocionando grupos emergentes en áreas escasamente cultivadas en España, como es el caso, por ejemplo, de las recientes actuaciones en el campo de las Neurociencias.

Desde 1987, el CRM acoge becarios postdoctorales financiados por el propio CRM, por los Gobiernos de Cataluña o de España, por la Comisión Europea (becas Leibniz a principios de la década de los 90, becas Marie Curie a partir de 1996), o a través del *European Post-Doctoral Institute for the Mathematical Sciences* (EPDI), del que el CRM forma parte desde el año 2000 (siendo el único miembro en una latitud geográfica inferior a los $48^{\underline{o}}$ N).

La organización de congresos y seminarios fue escasa en los primeros años, pero se incrementó fuertemente a partir de 1994: 21 cursos avanzados, 24 congresos y 22 workshops en los que han participado más de 3.500 investigadores y estudiantes de doctorado de 72 países distintos acreditan un

volumen de actividad superior a la de la mayoría de Institutos de investigación europeos. Una actividad que, sumada a la de los más de 1.000 investigadores de 58 países de los cinco continentes que han trabajado en el CRM, ha permitido mejorar cuantitativa y cualitativamente la investigación matemática en Cataluña y, por extensión, en España.

Dada la calidad de los cursos avanzados organizados por el CRM, a nivel predoctoral avanzado y postdoctoral reciente, la editorial suiza Birkhäuser-Verlag decidió hace casi cuatro años lanzar una serie de textos de difusión internacional que, bajo el nombre *Advanced Courses in Mathematics CRM Barcelona*, recoge el material que corresponde a los más significativos. Esta serie lleva publicados siete volúmenes y dos más están actualmente en proceso de impresión.

El crecimiento del CRM ha sido progresivo, tanto en número de investigadores, como de becarios postdoctorales o actividades diversas, pasando de 30 meses de investigador en 1985 a 280 meses en el año 2004. Todo ello con un presupuesto que en los últimos años se ha situado alrededor de los 750.000 Euros, procedentes de las tres fuentes principales de financiación: Gobierno de Cataluña, Gobierno de España y Comisión Europea.

En estos momentos en que el Ministerio de Educación y Ciencia ha manifestado explícitamente su disposición a potenciar la investigación matemática en todo el territorio español, creando posiblemente nuevos centros o estructuras en forma de red, es importante destacar que el CRM es el único Instituto de investigación matemática autónomo en todo el Estado Español. Aunque ha centrado su actividad en el entorno de los matemáticos catalanes, su hábitat natural, ha colaborado con las Universidades exteriores a Cataluña y, más generalmente, con los otros matemáticos españoles cuando ha sido requerido, ya sea facilitando la estancia de investigadores en otros centros españoles, o bien organizando actividades, como ha ocurrido por ejemplo con la Universitat Jaume I de Castelló y con la Universidad de Almería. Con ocasión de la celebración del *International Congress of Mathematicians* en Madrid en el año 2006, el CRM organizará un curso avanzado sobre *Geometría Computacional* en Alcalá de Henares, un congreso sobre *Neurociencia Matemática* en Andorra y un trimestre de investigación sobre *Análisis Armónico* en las Universidades Autónoma de Madrid y Autònoma de Barcelona. Una muestra de la voluntad de servicio a toda la comunidad matemática.

Veinte años para un Instituto de investigación cuyo nacimiento no fue consecuencia de una decisión política sino de la iniciativa de un sector de

la comunidad matemática, representan ya una larga vida y una intensa experiencia. Una vida que ha permitido que ocho matemáticos galardonados con la *Medalla Fields* trabajaran temporalmente en España y que el primer *Premio Abel* de la historia, Jean-Pierre Serre, disertara magistralmente el pasado día 9 de noviembre en el acto de conmemoración del vigésimo aniversario. Actualmente, por el volumen de investigadores visitantes y becarios postdoctorales y por las actividades que anualmente organiza, el CRM se sitúa entre los Institutos europeos más destacados de la red ERCOM (European Research Centres on Mathematics), cuyo *Chairman* desde el año 2002 es el Director del CRM.

Matemáticas y CSIC: El reencuentro

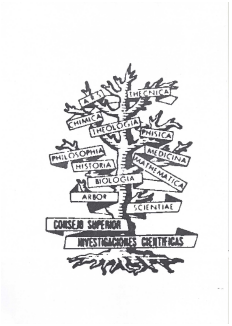
Manuel de León

Dpto. de Matemáticas
Instituto de Matemáticas y Física Fundamental
Consejo Superior de Investigaciones Científicas
Serrano 123, 28006 Madrid

Resumen

Se hace una breve descripción histórica de las matemáticas en el CSIC, señalando su realidad actual y su futuro, vinculado a la creación de un Instituto de Matemáticas y a un área de investigación diferenciada.

1 Breve historia de una decisión



Como de todos es bien sabido, el emblema del Consejo Superior de Investigaciones Científicas es un árbol, el árbol de la ciencia. Una de sus ramas esenciales, la de las Matemáticas, fue podada hace ahora 20 años. El CSIC tiene una larga historia que comienza en 1907 con la Junta de Ampliación de Estudios y tiene continuidad en 1939 con la creación del actual CSIC, así que en 2007 celebrará su centenario. En esta dilatada travesía, el CSIC ha creado nuevos Institutos, aparte de los iniciales, y ha cerrado o reformado algunos otros, pero nunca ha eliminado un área entera de un plumazo. Ese fue el resultado de una desafortunada decisión tomada por los responsables de la institución en 1984 con la supresión del Instituto Jorge Juan de Matemáticas. ¿Cómo pudo llegarse a esta situación? Y sobre todo, ¿por qué la comunidad matemática española aceptó esta medida?, son preguntas que intentaré contestar a continuación.

Debemos recordar que el panorama matemático español de los años ochenta era muy diferente del actual. El Instituto Jorge Juan de Matemáticas desempeñó un papel de referencia desde su fundación hasta principios de los setenta, con una biblioteca importante a la que los matemáticos españoles de entonces acudían en búsqueda de artículos y libros para llevar adelante su investigación. La Real Sociedad Matemática Española tenía allí su sede; en realidad, podríamos hablar de historias paralelas para las dos instituciones.

Desgraciadamente, el Instituto no llevó adelante su renovación cuando, a finales de los setenta y principios de los ochenta, comenzaron a aparecer jóvenes matemáticos con una formación moderna, desarrollando ya una investigación homologable internacionalmente. El CSIC cometió entonces el error de no proceder a la imprescindible transición, y los cambios políticos (que como todos sabemos aparejan en nuestro país cambios en las políticas científicas) impidieron más tarde la refundación, cuando ya se habían dado algunos pasos en la buena dirección.

El proceso degenerativo que simultáneamente sufrió la Real Sociedad Matemática Española impidió que se celebrase un debate nacional sobre las consecuencias de la desaparición del Jorge Juan. Recordemos que en esa época la presencia de otras sociedades Matemáticas era mucho más reducida, y alguna, como SëMA, ni existía todavía. Estos hechos, que han tenido unas consecuencias muy negativas para el desarrollo de las Matemáticas españolas, deberían hacernos reflexionar sobre la importancia de una vertebración social de las mismas, basada en la colaboración y la mutua confianza.

2 Veinte años de ausencia (y lucha)

La historia personal en el CSIC del firmante de este artículo comienza en enero de 1986, con la incorporación a la Confederación Española de Centros de Investigación Matemática y Estadística (CECIME). La CECIME era una estructura condenada a perecer. Intentaba en definitiva resucitar la vieja estructura de los Seminarios Matemáticos en las cuatro Facultades "históricas"de Matemáticas: Complutense, Zaragoza, Central de Barcelona y Santiago de Compostela. Coordinar ahora todas las Facultades de Matemáticas del país (pues a eso se reducían los "centros de investigación" de los que se hablaba) era una empresa sin porvenir. Paradójicamente, los escasos efectivos de plantilla del CSIC en Madrid quedaban fuera de los estatutos de la CECIME, circunstancia que se hizo notar sin resultado a los responsables de la época en el CSIC.

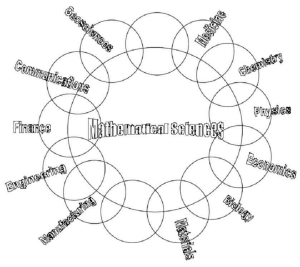
En 1989 desaparece la CECIME y se integra a los matemáticos del CSIC en una Unidad dependiente directamente de la Presidencia, Unidad a la que se le niega la denominación de Matemáticas y que tras unas conversaciones con el Vicepresidente de Investigación pasa a llamarse Unidad de Topología, Álgebra, Geometría y Sistemas. Las causas de esta negativa se escapan al contenido de este artículo y están ligadas al fracaso de un segundo intento de constitución de un Instituto de Matemáticas al margen de los matemáticos del CSIC. Finalmente, es en 1992 cuando se crea el Instituto de Matemáticas y Física Fundamental (IMAFF), que surge de la fusión de los físicos integrados en un proyecto fallido de Instituto de Física Fundamental entre el CSIC y la Universidad Complutense de Madrid, y los matemáticos del CSIC. En estos doce últimos años, los matemáticos encuadrados en el Departamento de Matemáticas del IMAFF hemos sufrido un auténtico calvario, padeciendo no sólo la marginación en el ámbito general del CSIC sino también la desafortunada

gestión del citado Instituto.

Justo es reconocer el apoyo que la Presidencia de César Nombela proporcionó al Departamento de Matemáticas del IMAFF, con la provisión de algunas plazas de turno libre. Estas plazas son las que han permitido en estos momentos que este Departamento esté integrado por un plantel excelente de jóvenes investigadores. Sin embargo, las medidas fueron tímidas y no resuelven de una manera definitiva el problema de las Matemáticas en el CSIC; esta solución pasaba por la independencia de las mismas en un Instituto propio en formación (figura que el CSIC usa para llevar adelante proyectos de Instituto que sin poseer todavía una entidad suficiente son tutelados por la Presidencia hasta alcanzar la madurez) y una política más generosa de plazas específicas. Las Presidencias sucesivas de Rolf Tarrach y Emilio Lora fueron en cierta manera continuistas en esta política de apoyo de perfil bajo.

De nuevo, la comunidad Matemática española asiste con cierta pasividad a estos procesos, no advirtiendo la gravedad de la situación y la pérdida de unos recursos que hoy paliarían las escasas perspectivas de nuestros jóvenes matemáticos, recursos que podría haber proporcionado la Oferta de Empleo Público del CSIC. Otra de las consecuencias negativas ha sido además la menor presencia de una manera colectiva en la elaboración de los Planes Nacionales y su coordinación con los Programas Marco europeos. La comunidad matemática española ha estado además hasta ahora casi ausente (excepto el caso del CRM catalán) de importantes iniciativas europeas, como el European Research Centres of Mathematics (ERCOM), una comisión de la European Mathematical Society.

Pero esto es un camino de ida y vuelta, y el CSIC también ha perdido mucho al eliminar las Matemáticas de su entorno. Las Matemáticas son interdisciplinares, están en el corazón de la investigación científica y el desarrollo tecnológico. Permítasenos reproducir este esquema que encontramos en un reciente estudio de la National Science Foundation (NSF) y que refleja ese carácter de las Matemáticas.



Los investigadores del CSIC han perdido así un valor añadido a su investigación y la institución un lugar central de referencia al que acudir en búsqueda de conocimientos matemáticos. Todos hemos perdido en estos últimos años, todos nos hemos empobrecido.

3 La situación actual

La situación cambia drásticamente con el nuevo equipo directivo del CSIC en 2004 y se comienza a ver la luz. Se anuncia la creación de un Instituto de Matemáticas, mixto con algunas de las Universidades de Madrid y con grandes probabilidades de convertirse en una pieza importante en el esquema de un

futuro Centro Nacional de Matemáticas.

¿Cuál es la composición del Departamento de Matemáticas a día de hoy y cómo pretende afrontar la nueva etapa? Hace algunos años, en 1998, el autor de este artículo se hacía públicamente la siguiente pregunta: ¿Hay vida matemática en el CSIC? Hoy podemos ser más optimistas que entonces. Debemos decir que en los últimos cinco años, las incorporaciones al CSIC en el área de las Matemáticas han fructificado en cuatro jóvenes investigadores de plantilla, cinco contratados del Programa Ramón y Cajal y uno del Programa Juan de la Cierva. Debe decirse que, en total, el CSIC ha incorporado en estos cuatro años del Programa Ramón y Cajal a 7 jóvenes matemáticos, de los cuáles uno es ya personal de plantilla y otro ha optado finalmente por un puesto en el extranjero. Debemos recordar aquí que las escalas del CSIC (Profesor de Investigación, PI; Investigador Científico, IC; y Científico Titular, CT) se corresponden a los antiguos Cuerpos universitarios de Catedrático, Agregado y Adjunto. Con las abreviaturas obvias, éste es el equipo que entraría a formar parte por parte del CSIC en el nuevo Instituto de Matemáticas: 1 PI, 3 IC, 1 CT, 5 RyC y 1 JdlC, a los que hay que añadir 6 Becarios de diferentes programas.

Digamos también que las temáticas a las que se dedican estos investigadores configuran tres grandes líneas:

- Geometría y Física Matemática.
- Ecuaciones en Derivadas Parciales y Mecánica de Fluidos.
- Mecánica Geométrica y Control.

Estas incorporaciones han sido programadas en función de la calidad y la oportunidad, sin tener en cuenta hasta el momento un plan de líneas de investigación prioritarias. El lema ha sido "atraer buenos y jóvenes investigadores, no importa su área de trabajo". Entre los objetivos estratégicos del grupo está el reforzar las líneas existentes junto con la inserción de otras nuevas. La política seguida de incorporar investigadores jóvenes ha llevado a la creación de un grupo de excelencia, tal y como puede comprobarse en la tabla, que recoge la producción y el impacto de 167 instituciones mundiales en Matemáticas; los datos del CSIC han sido obtenidos por el mismo procedimiento al usado por la *Web of Knowledge* en la sección Essential Science Indicators y el puesto sería el hipotéticamente atribuido en ese caso.

El resultado es doblemente meritorio si se tiene en cuenta que la juventud del grupo limita las citas de sus artículos; en efecto, varios de ellos han comenzado a publicar hace 6 u 8 años.

Los investigadores de este Departamento dirigen cuatro proyectos del Programa Nacional de Matemáticas, participan en numerosos proyectos

Puesto	Institución	Artículos	Citas	Citas por artículo
# 103	Matemáticas, CSIC	173	592	3.42
# 109	UAM	468	1475	3.15
# 124	UCM	783	2356	3.01
# 138	UAB	447	1206	2.70
# 147	UB	459	1195	2.60
# 150	UGR	840	2090	2.49

Datos de las cinco instituciones españolas incluídas en el Essential Science Indicators, Enero 1, 1994 - Octubre 31, 2004 (Total 167)

internacionales, en comités editoriales de revistas y en comités científicos y organizadores de eventos matemáticos. Debe destacarse además la gran implicación en los temas de política científica nacional relacionados con las Matemáticas y su organización social.

El Departamento de Matemáticas del CSIC ha elaborado su propia página web `www.mat.csic.es`, en la que se pueden encontrar detalles de sus componentes, así como de las actividades que desarrolla: seminarios, coloquio y proyectos de investigación. También puede allí encontrarse la memoria de actividades.

Hay, sin embargo, circunstancias que pueden poner en peligro esta realidad actual. No existe un área de Matemáticas. De un área de Matemáticas, Física y Química, se pasó a una de Ciencias Físicas y Tecnologías Físicas, en otro desatino más que pareciera no haber preocupado a nadie, ni intramuros ni extramuros. ¿Nos podemos imaginar lo que hubiera dicho la comunidad de físicos españoles si hubiese sido al revés? Esto motiva que conseguir plazas de promoción interna de CT e IC a IC y PI, respectivamente, sea una empresa casi imposible. Imaginemos un tribunal constituido por cinco físicos juzgando a diez físicos y a un matemático; el resultado sería obvio. Esta situación la hemos padecido por 20 largos años.

Por otra parte, el número de contratados Ramón y Cajal es alto y debemos ofrecerles la oportunidad de optar a una plaza permanente. Es decir, necesitamos que haya nuevas plazas de CT, que son de turno libre, así que podrían acceder a ellas investigadores procedentes de las Universidades, en natural competencia que nuestros contratados. Afortunadamente, nos consta que el nuevo equipo directivo tiene a este colectivo entre sus objetivos prioritarios.

4 Y ahora, el futuro

El CSIC es una pieza clave en el sistema español de I+D+i, por su implantación estatal, su interdisciplinariedad y su tamaño. La creación de un Instituto de Matemáticas debe ser el primer paso para la reimplantación de un área de Matemáticas. Conviene recordar que el CNRS francés posee unos 375 matemáticos de plantilla, unos 75 técnicos de apoyo a estos matemáticos y un centenar de unidades que van desde un laboratorio con unos pocos

investigadores de plantilla hasta grandes Institutos de investigación, pasando por su apoyo importante a centros de encuentros como el CIRM de Marsella o suministrando la documentación científica matemática. Y es a este modelo del CNRS al que se podría tender, con una buena gestión de un Centro Nacional de Matemáticas y un adecuado compromiso de nuestras Comunidades Autónomas y Universidades.

Debemos recordar la situación de las Matemáticas españolas, que han experimentado un crecimiento espectacular en los últimos años, pasando de una producción del 0,3 % en artículos en revistas incluídas en el Journal Citation Reports (JCR) en 1980 hasta el 4,65 % del quinquenio 1999–2003.

Este crecimiento está amenazado por dos factores:

- La saturación de plazas en nuestras Universidades, acompañada de la falta de una carrera científica que impide la consolidación de nuestros jóvenes matemáticos.
- La falta de orientación de la investigación matemática española, que la hace poco efectiva para orientarse a campos emergentes y transversales.

La recuperación del espacio matemático en el CSIC puede ser fundamental para combatir estos dos puntos débiles, al poner a disposición de la comunidad matemática española una parte de su Oferta de Empleo Público, a la vez que facilitaría la interacción con otras áreas científicas y tecnológicas. El CSIC está además emprendiendo reformas de enorme alcance, en su estructura jurídica y en sus objetivos, que pueden contribuir en gran manera a dinamizar la investigación matemática.

Desde el punto de vista institucional, esta recuperación supondría la normalización de las Matemáticas en nuestro país. No debe olvidarse que el CSIC es una pieza importante en el organigrama del Ministerio de Educación y Ciencia y en la coordinación con los Programas Marco europeos, realizando una importante tarea de vertebración de la investigación española. No cabe duda del beneficio para las Matemáticas.

Si volvemos a la tabla, vemos que la investigación matemática española, tal y como ocurre en el resto de las áreas científicas, es institucionalmente débil. Cuenta, por supuesto, con excelentes individualidades y grupos establecidos de calidad, pero no tenemos instituciones influyentes. La existencia de un Instituto de Matemáticas del CSIC, en coordinación con las Universidades, puede significar un cambio importante si consigue convertirse en un referente nacional e internacional. En el próximo quinquenio tendremos ocasión de ver si estas expectativas se cumplen.

Entrevista a Manuel de León, Presidente del Comité Ejecutivo del ICM2006

Ofrecemos a continuación una entrevista a Manuel de León, Profesor de Investigación del CSIC y Presidente del Comité Ejecutivo del próximo *International Congress of Mathematicians*, cuya celebración está prevista en Madrid en Agosto de 2006.

Manuel de León es también Coordinador de Matemáticas de la ANEP, Vicepresidente de la RSME y Presidente del Comité Español de Matemáticas (CEMat). Su entusiasmo constante y la dedicación que pone a la organización del ICM2006 parecen capaces de vencer todas las dificultades y auguran un exitoso desarrollo de éste. Desde el grupo editor de SēMA es nuestra obligación aprovechar esta entrevista para pedir a todos los socios su ayuda y colaboración en esta empresa.

Pregunta: ¿A quién representa el Comité Ejecutivo del ICM2006?

El Comité Ejecutivo del ICM2006 fue nombrado por el Comité Español de Matemáticas (CEMat), antiguo Comité IMU-España, el cual representa a España en la Unión Matemática Internacional (IMU en sus siglas inglesas). IMU es una de las uniones científicas de la ICSU (International Council of Scientific Unions). El CEMat engloba a todas las Sociedades Matemáticas españolas, y está nombrado oficialmente por el MEC. Puede decirse con toda propiedad que el CEMat representa a toda la comunidad matemática española, así que el Comité Ejecutivo del ICM2006 hereda de alguna manera esa legitimidad. Para darle más agilidad de gestión al Comité, se constituyó una Asociación, inscrita en el Registro Nacional de Asociaciones del Ministerio de Interior.

P: ¿Cómo podría enjuiciarse la situación actual de la investigación matemática en nuestro país? ¿Existe actualmente una planificación clara? ¿Existen metas a corto o medio plazo?

La investigación matemática española ha crecido vertiginosamente en los últimos 25 años, como no nos hemos cansado de repetir, de un 0,3 % del total mundial en artículos en ISI en 1980, hasta el 4,65 % en el quinquenio 1999–2003. El impacto de nuestra investigación crece, pero lentamente. Tenemos grupos excelentes así como individualidades de gran valor. Fallamos en el conjunto,

mal generalizado en la Ciencia española, fruto de una falta de orientación en la investigación y de estructuras de excelencia. El Programa Nacional de Matemáticas (y quizás un Centro Nacional de Matemáticas) puede ser el inicio de un cambio cualitativo. Tenemos una cantera extraordinaria de jóvenes matemáticos que corremos el riesgo de perder. En este país, nos falta una cultura científica que sea capaz de generar sinergias y no actitudes cainitas, atavismos de un pasado que deberíamos ya olvidar. Si nos ponemos las pilas, las Matemáticas españolas tienen un gran futuro.

P: ¿Qué significa para la comunidad matemática española que el próximo ICM2006 se celebre en España?

Es un reconocimiento de la mejora de nuestra investigación, nuestra puesta de largo internacional. La refundación de la RSME ha sido clave para coordinar a las Sociedades en esta dirección. En 1998 reconstruimos el Comité IMU-España y se dieron los primeros pasos para la *rentrée* de España en el colectivo internacional. En el Año Mundial de las Matemáticas se invitó el Comité Ejecutivo de IMU a celebrar su reunión anual en Madrid y en 2002 presentamos nuestra candidatura al ICM. Este año acabamos de pasar del Grupo III de IMU al Grupo IV. Nadie lo hubiera soñado hace 7 u 8 años, pero se ha trabajado duro y los resultados están ahí. El ICM2006 es además una gran oportunidad para hacer una reflexión sobre nuestras Matemáticas, en investigación y educación, y de llamar la atención sobre esta Ciencia por parte de las administraciones y la propia sociedad. Tenemos que aprovecharla, y una manera inmediata de colaborar es preinscribiéndose ahora y participando después. Ningún matemático español debería mantenerse al margen de esta gran fiesta.

P: Aparte de la progresión de la Matemática española y del deseo de reconocimiento internacional de este hecho, ¿existen o existieron razones adicionales para elegir Madrid como sede?

Mi impresión es que hubo un acuerdo implícito de que debía ser Madrid. Este tipo de acontecimientos suele organizarse en grandes ciudades, porque requieren una gran infraestructura y a la vez unas excelentes comunicaciones

internacionales. Barcelona acababa de organizar el tercer *European Congress of Mathematics* y Madrid era la candidata natural. Debo recordar también que éste es el Congreso de IMU, e IMU concede la organización al país que gana la candidatura y esa candidatura va ya asignada a una ciudad. La propuesta de Madrid hecha por el Comité IMU de España fue unánime. Madrid aglutina la cuarta parte de la investigación matemática española y reúne además todas esas condiciones. Como anécdota, es interesante decir que hace unos cuantos años el Ayuntamiento de Madrid se dirigió a mí con la petición de que los matemáticos madrileños presentaran la candidatura de Madrid para un ICM, porque éste era uno de los Congresos importantes que tenían en la lista. Respondí entonces que debían consultarlo al Ministerio de Educación y de allí los encaminaron de nuevo a mí. A la vista de esto, sugerí que contactaran con matemáticos prestigiosos de las Universidades de Madrid, a fin de poder organizar un Comité de Candidatura. Yo estaba recién llegado a Madrid al CSIC y en aquellos tiempos no se había comenzado la vertebración social de las Matemáticas que se propició con la refundación de la RSME y el Comité IMU-España. La iniciativa desgraciadamente no llegó a nacer.

P: ¿Qué repercusiones cabe esperar tras la celebración del ICM2006? En particular, ¿puede sensibilizar este evento al resto de la comunidad científica?

Como decía antes, este evento es una ocasión para que las Matemáticas estén en el candelero y debemos aprovecharlo. La interacción con otras Ciencias que los ICM permiten debería ser utilizada para señalar ese aspecto transversal de nuestra Ciencia, ese doble papel de las Matemáticas como valor añadido a la investigación y como papel central de soporte intelectual de la misma.

P: ¿Cómo se ve representada la Matemática Aplicada que se hace en España en este evento?

La Matemática Aplicada participa en el Comité Ejecutivo a través de SeMA y además hay varias sesiones que entran de lleno en los temas habituales de la Matemática Aplicada. Se entregará el *Premio Nevalinna*, en el área de las aplicaciones de las Matemáticas a las Ciencias de la Información y habrá una sesión de *software matemático*. Es verdad que existen los ICIAM, y a veces se piensa que los ICM responden mejor a los patrones de una Matemática más fundamental. Yo discrepo radicalmente de esta visión. El ICM engloba a TODAS las Matemáticas, y uno de sus mayores objetivos debe ser mostrar esa gran unidad interna y conceptual de las Matemáticas, unidad que es su gran fuerza y garantía de futuro.

P: Por otra parte, ¿hay indicios de que la celebración del ICM2006 contribuya a dar más sentido e importancia al papel del matemático en la sociedad? ¿Cabe esperar que las relaciones universidad-empresa se vean fortalecidas?

Es evidente que habrá una mayor presencia mediática de las Matemáticas, y que muchos se preguntarán sobre las razones de las mismas. Eso será bueno para nosotros, salir de esa invisibilidad. Está costando mucho convencer a las empresas de la importancia de las Matemáticas para éstas, es una tarea de largo alcance, pero es un buen momento para dar pasos, para construir futuro.

P: ¿Hay alguna parcela de las Matemáticas para la que la celebración del ICM2006 posea un especial interés?

Decía antes que los ICM no priorizan ningún tipo de Matemáticas, los ICM "son" las Matemáticas en su conjunto. Cada vez se ve más este fenómeno de unificación en las Matemáticas. ¿No estamos aplicando las Matemáticas que se consideraban más puras hace 50 años a problemas tecnológicos de primera magnitud? Pienso que, aunque hay una gran especialización, sin embargo, hay un cuerpo de doctrina único que es nuestra gran fuerza. Es verdad que la concesión de las *medallas Fields* focaliza la atención a los campos de trabajo de los premiados, pero eso es natural.

P: ¿En qué situación nos encontramos en este momento? ¿Se ha captado ya suficiente interés por parte de las autoridades? ¿Está ya la comunidad matemática preparada mentalmente para la celebración?

Creo que las autoridades ya han comenzado a darse cuenta de que este Congreso lo ha ganado España y lo tiene que organizar España, que se juega su prestigio internacional. Lo mismo está pasando con el Ayuntamiento y la Comunidad de Madrid. Yo soy ahora más optimista. Estoy más preocupado con la comunidad matemática española, que debería estar preinscribiéndose en masa. Desde aquí animo a todos nuestros profesores, investigadores y estudiantes a hacerlo, y acudir en 2006 a Madrid. La presencia española debería ser masiva, demostrar que nuestras Matemáticas son de verdad potentes. Y me quisiera dirigir a los profesores de Secundaria: éste es también su Congreso, no es un congreso elitista que organicemos "los de la universidad". Hay una sesión sobre Educación Matemática, es una gran ocasión de escuchar a las grandes figuras internacionales y todas las novedades bibliográficas estarán expuestas en Madrid esos días. Todos saldremos matemáticamente enriquecidos del ICM2006.

P: En el aspecto organizativo, ¿qué se puede decir sobre lo que ya se ha hecho y qué es lo más importante todavía por hacer?

Se ha hecho mucho, y queda mucho por hacer. Nos hemos centrado hasta ahora en los aspectos de búsqueda de financiación y difusión. Lo peor está por venir, aunque el equipo está preparado.

P: ¿Se puede mencionar algún problema especial con el que se haya encontrado el CE?

La estructura social de las Matemáticas españolas es muy compleja, hay muchas Sociedades y una estructura de Comunidades Autónomas que siempre está presente. La ausencia del panorama de la RSME ha causado un gran daño, al no ejercer un papel de cohesión, que le corresponde de manera natural. A partir de 1997 y 1998, la RSME comienza a desempeñarlo y en 2000 el Año Mundial de las Matemáticas nos permitió avanzar mucho en esa necesaria coordinación. Nos hemos encontrado con un AMM2000 y, sin tiempo para el descanso, con el ICM2006. Diría que se ha hecho lo imposible.

Un CE debe ser eso, ejecutivo, y no una comisión de cuotas. Hemos tenido que conjugar ambas cosas. Estoy seguro de que el ICM2006 va a ser un éxito para las Matemáticas españolas, pero tenemos que seguir avanzando en las estructuras matemáticas que nos permitan afrontar estos desafíos sin agobios ni tensiones. Mi idea fue siempre que el Comité IMU-España (ahora el CEMat) fuese independiente del CE del ICM2006 y, una vez que se nombrara este CE, permaneciese como observador interviniendo sólo si había que resolver algún conflicto. Desgraciadamente, ambos Comités se solapaban. Cuando hay algún problema difícil (y la organización de un evento de esta magnitud los genera sin duda) debe haber un organismo independiente que pueda intervenir y resolverlos tomando las decisiones oportunas. Hay que separar los "poderes". Pero también debe decirse que las circunstancias eran excepcionales y que había que organizar el ICM y eso es lo que intentamos entre todos. Eso sí, deberíamos todos sacar lecciones para el futuro.

P: ¿Es previsible que este evento influya en el futuro inmediato sobre el impulso de las Matemáticas en los países en vías de desarrollo y en el Tercer Mundo?

Precisamente, uno de los objetivos que persigue IMU es fomentar el desarrollo de las Matemáticas en los países en vías de desarrollo, en el convencimiento de que las Matemáticas son la clave del desarrollo tecnológico; de hecho, éste era uno de los puntos señalados en la Declaración de Rio de Janeiro del Año Mundial de las Matemáticas. IMU persigue además que cada vez más países alcancen un nivel que los capacite para entrar en la Unión. Un ICM es un foco de atención mundial y a la vez una gran herramienta para fomentar la participación de matemáticos de estos países y ponerlos en contacto con las grandes líneas de investigación. El ICM2006 debería seguir esta tendencia, en particular por la especial situación geopolítica española.

P: ¿Puede ser la celebración del ICM2006 una razón para intensificar la influencia española en los países iberoamericanos?

Uno de los ejes de actuación del ICM2006 es el iberoamericano. El Comité Ejecutivo está trabajando con la Agencia Española de Cooperación Internacional (AECI), intentado que la presencia de los matemáticos de estos países sea importante. El ICM2006 está sirviendo para señalar que España debe

intensificar los contactos y yo hablaría, no de influencia sino de colaboración. A fin de cuentas, son compañeros naturales de viaje.

P: ¿En qué modo puede el ICM2006 influir en las publicaciones matemáticas españolas y en las internacionales lideradas por investigadores españoles?

El ICM2006 va a decirle al mundo matemático que España ha hecho grandes progresos, que no sólo hay individualidades sino que contamos con muchos grupos competentes, y que, además, tenemos una cantera excepcional de jóvenes matemáticos. Los matemáticos españoles ya no viajan fuera para aprender, ahora también viajan para enseñar lo que han hecho en casa.

Este evento es un gran escaparate para nuestros matemáticos que, a medio plazo, producirá sin duda excelentes retornos. Por otra parte, es importante que nuestras propias revistas sean más conocidas internacionalmente, de modo que los matemáticos internacionales encuentren atractivo publicar en ellas. A este fin, se ha puesto en marcha una *Plataforma de Revistas Científicas Españolas*, REVICIEN, al estilo de las grandes editoriales, persiguiendo este objetivo: dar visibilidad a las publicaciones nacionales y conseguir una mejora de las mismas en lo que se refiere a sus soportes digitales. Además, existe una zona de prensa en la que los directores señalan artículos de interés mediático, al estilo de *Science* o *Nature*. Ahora se está procediendo a la creación de una *Asociación de Revistas Científicas Españolas*, una de cuyas misiones sería gestionar REVICIEN.

P: ¿Existen noticias de que las autoridades acompañen o aprovechen la celebración del ICM2006 con algún programa de carácter extraordinario que tenga que ver con la investigación matemática?

El ICM2006 ha supuesto una mayor sensibilización hacia las Matemáticas. En 2006 tendremos seguramente un *Instituto de Matemáticas* en el CSIC, tras más de 20 años de sequía, y el MEC está considerando muy seriamente la creación de un *Centro Nacional de Matemáticas* que podría estar listo para la inauguración del ICM2006. Crucemos los dedos y rememos todos en la misma dirección.

P: ¿Se ha dejado sentir el cambio de Gobierno en las iniciativas en el ámbito de las Matemáticas y en particular en la actitud ante la celebración del ICM2006? En particular, ¿qué ha supuesto la reestructuración del Ministerio de Ciencia y Tecnología?

Para un evento que requiere tanto tiempo de preparación, los cambios son siempre un trabajo añadido. El anterior MCYT había ya comprometido una ayuda importante para el ICM2006 y también el antiguo MECD estaba ya colaborando. Hemos tenido que retomar el contacto con el nuevo equipo del

MEC y volver a explicar de qué se trataba. Pero además en Madrid hemos tenido elecciones municipales y elecciones autonómicas (recuerdo que éstas últimas se tuvieron que repetir). Estos cambios han supuesto un trabajo doble, que lo que demuestra es la escasa tradición científica española y las carencias de nuestro sistema.

Necesitamos que los grandes temas como es el caso de la Educación y la Investigación estén al margen de los cambios políticos si queremos ser un país de la primera división. También es justo decir que somos nosotros los que tenemos que ponernos a trabajar para cambiar la situación. No nos podemos limitar a quedarnos en nuestros despachos lamentándonos con los colegas de la situación, debemos ser activos. Y el CE del ICM2006 así lo ha entendido. Nadie va a venir a nuestros despachos para conceder nuestros deseos, somos un colectivo importante, bien organizado, bien visto por las administraciones, hagamos propuestas interesantes y vayamos a defenderlas. Olvidemos los protagonismos individuales y trabajemos colectivamente. Tenemos mucho que ganar y mucho que decir. En cierta manera, somos un colectivo que muchas veces la propia administración toma como ejemplo para otras áreas. Ahora tenemos una gran oportunidad de mostrar nuestra madurez como colectivo, con un Programa Nacional de Matemáticas, un ICM en 2006, un Espacio Europeo de Educación Superior que nos abre un enorme abanico de posibilidades, un posible Centro Nacional de Matemáticas. Aprovechemos estas oportunidades.

P: ¿Habrá innovaciones de algún tipo en esta edición del ICM? En particular, ¿tendrán impacto de algún modo las nuevas tecnologías?

Volverá a haber como en Berlín 1998 una sesión de *software matemático*. Además, se transmitirán todas las sesiones *online* via internet, una gran novedad.

P: Aparte del ICM2006 y en torno a las mismas fechas, tendrá lugar la celebración de congresos-satélite. ¿Cuál es la estructura y significado de éstos? ¿En qué condiciones tendrán lugar? A día de hoy, ¿cuántos congresos-satélite están previstos ya?

Los congresos satélites son una actividad complementaria de los ICM, con un contenido temático especializado. Son *workshops* tradicionales que aprovechan la movilización que supone un ICM. Tendrán lugar desde mediados de junio a mediados de septiembre, y no tienen que celebrarse necesariamente en España. Hasta ahora hay unos 15 satélites aprobados. Aprovecho la ocasión para hacer un llamamiento a nuestros matemáticos para que presenten nuevas propuestas.

P: ¿Se verán acompañadas las distintas sesiones del ICM2006 con actividades de carácter cultural y social?

Como comentaba antes, queremos aprovechar la oportunidad de mejorar la imagen de las Matemáticas. Están programas actividades culturales, alguna puede ser de gran alcance. Se editará un sello conmemorativo (una tradición en los ICM), un billete de lotería, habrá exposiciones, ciclos de conferencias, queremos que desde junio a septiembre de 2006 España sea una gran fiesta matemática. Seguro que será así.

Bol. Soc. Esp. Mat. Apl. nº 31(2005), 31–52

Sobre la paralelización de problemas elípticos

M. C. Calzada[1], I. I. Albarreal[2],
J. L. Cruz[1], E. Fernández-Cara[2],
J. R. Galo[1], M. Marín[1]

[1] Dpto. de Informática y Análisis Numérico, Universidad de Córdoba, Campus de Rabanales, Ed. C2-3, E-14071 Córdoba
[2] Dpto. de Ecuaciones Diferenciales y Análisis Numérico, Universidad de Sevilla, Aptdo. 1106, 41080 Sevilla

ma1canam@uco.es, iignacio@us.es, jlcruz@uco.es, cara@us.es, ma1gasaj@uco.es, merche@uco.es

Resumen

En este trabajo se recogen dos de los problemas que aparecen a la hora de resolver numéricamente las ecuaciones deNavier-Stokes. La discretización en la variabletemporal en dichas ecuaciones nos conduce, en cada etapa de tiempo, a un problema de Burgers y un problema de Stokes generalizado, en las variables espaciales. La aplicación de un método de punto fijo y un método de tipo gradiente conjugado, respectivamente, nos lleva a la resolución de un elevado número de problemas de tipo Helmholtz. Así, en una primera parte, recordamos un algoritmo paralelo que resuelve numéricamente la ecuación de Helmholtz. Por otro lado, la utilización de diferencias finitas centradas sobre un mismo mallado regular para la velocidad y la presión en la resolución del problema de Stokes puede generar soluciones oscilantes, en particular presiones espúreas. En una segunda parte, por tanto, se presenta un esquema de discretización para el problema de Stokes, demostrándose mediante un análisis asintótico del error que éste regulariza la solución[1].

Fecha de recepción: 31/01/05

[1] Una versión previa de este trabajo aparecerá en las Actas del I Workshop sobre "Recientes Avanaces en el Análisis y Control de Ecuaciones Diferenciales No Lineales", celebrado en Córdoba en febrero de 2004.

1 Introducción

En la resolución de muchos problemas de gran complejidad, aparecen de forma frecuente ecuaciones de tipo Helmholtz:

$$\alpha u - \beta \Delta u = f \quad \text{en } \Omega \subset \mathbb{R}^N, \tag{1}$$

siendo $\alpha \geq 0$, $\beta > 0$, f una función conocida y Ω un dominio acotado bi o tridimensional. La ecuación se completa con condiciones de contorno, que pueden ser de tipo Dirichlet,

$$u = g \text{ sobre } \Gamma = \partial\Omega, \tag{2}$$

de tipo Neumann, Fourier o una mezcla de ambas. Aunque éste es uno de los problemas más conocidos y estudiados, es claro que una pequeña ganancia en el tiempo de resolución de (1) puede en la práctica suponer una gran mejora en el tratamiento de problemas más complejos. Pensemos en las ecuaciones de Navier-Stokes, para las cuales hemos desarrollado varios algoritmos donde este hecho se pone de manifiesto (véase por ejemplo [3, 4]). El método que se presenta sigue la idea de los denominados métodos ADI (***A**lternating **D**irections **I**mplicit methods*). Estos métodos fueron formulados inicialmente por Peaceman y Rachford [15] en el caso bidimensional y por Douglas y Rachford [7] en el caso tridimensional y, posteriormente, fueron generalizados al caso N-dimensional por Douglas y Gunn [6]. Para describir la idea de estos métodos, consideremos por ejemplo el problema de Helmholtz (1)–(2). Consideremos fijados una descomposición del operador $\mathrm{L} = \alpha\mathrm{I} - \beta\Delta$ en la forma

$$\mathrm{L} = \mathrm{L}_1 + \mathrm{L}_2 \quad \text{con} \quad \mathrm{L}_n = \frac{\alpha}{2}\mathrm{I} - \beta\frac{\partial^2}{\partial x_n^2} \quad (n = 1, 2) \tag{3}$$

y un esquema en diferencias finitas centradas con paso de discretización h en las dos direcciones espaciales. Un método de paso fraccionado (de tipo de Peaceman-Rachford) para esta descomposición es un método iterativo que comienza con U^0 arbitrario y, para $m \geq 0$, calcula U^{m+1} en dos pasos. Así, para U^m dado, primero se obtiene $U^{m+\frac{1}{2}}$ como solución de

$$(I + \tau A_1)\, U^{m+\frac{1}{2}} = \tau f + (I - \tau A_2)\, U^m \tag{4}$$

y en segundo lugar se resuelve

$$(I + \tau A_2)\, U^{m+1} = \tau f + (I - \tau A_1)\, U^{m+\frac{1}{2}}. \tag{5}$$

Aquí, $\tau > 0$, y A_n es la matriz correspondiente a la discretización del operador ${\mathrm{L}_n}^2$. Así, en el primer paso se resuelve un problema discreto correspondiente al operador

$$\frac{\alpha}{2}\mathrm{I} - \beta\frac{\partial^2}{\partial x_1^2}\,,$$

[2]Por simplicidad, hemos denotado de nuevo f la función discretizada.

es decir, un problema en la dirección x_1. Análogamente, en el segundo paso se resuelve un problema en la dirección x_2. Esto es lo que motiva el nombre de **método de direcciones alternadas**. Con una adecuada ordenación de los nodos del mallado, se consigue que las matrices $I + \tau A_n$ $(n = 1, 2)$ sean diagonales por bloques, siendo cada bloque una matriz tridiagonal. Esta idea se puede aplicar, en general, a la hora de resolver un sistema algebraico

$$Au = f \tag{6}$$

resultante (por ejemplo) de la discretización de un problema elíptico, siempre que la matriz A se pueda descomponer en la forma[3]

$$A = \sum_{n=1}^{p} A_n \,, \tag{7}$$

siendo las matrices A_n $(1 \leq n \leq p)$ más fáciles de invertir que la propia A. Esta generalización da lugar a los llamados **métodos de descomposición** o **métodos secuenciales de paso fraccionado** (véase [13]). En este caso, para $m \geq 0$, conocido U^m, el cálculo de U^{m+1} se realiza resolviendo los siguientes sistemas:

$$\begin{aligned} \frac{U^{m+\frac{1}{p}} - U^m}{\tau_m} + A_1 \left(U^{m+\frac{1}{p}} - U^m \right) &= -\alpha \left(\sum_{n=1}^{p} A_n U^m - f \right), \\ \frac{U^{m+\frac{n}{p}} - U^{m+\frac{n-1}{p}}}{\tau_m} + A_n \left(U^{m+\frac{n}{p}} - U^m \right) &= 0, \qquad n = 2, \ldots, p \end{aligned} \tag{8}$$

siendo $\alpha > 0$ y $\{\tau_m\}$ una sucesión de parámetros reales. Como acabamos de ver, los métodos de paso fraccionado (8) son algoritmos iterativos *secuenciales*, en los que cada iteración se subdivide en p pasos que se realizan sucesivamente según el esquema:

$$U^m \to U^{m+\frac{1}{p}} \to U^{m+\frac{2}{p}} \to \cdots \to U^{m+\frac{p-1}{p}} \to U^{m+1}. \tag{9}$$

Es decir, hasta que no se calcula $U^{m+\frac{i}{p}}$ no se puede calcular $U^{m+\frac{i+1}{p}}$. Por ello, la paralelización de estos algoritmos sólo puede llevarse a cabo al más bajo nivel, a la hora de resolver los sistemas (ver por ejemplo [18] y [11]). Nosotros abordamos la resolución de los problemas de Helmholtz generalizando y ampliando a dominios arbitrarios un método introducido por Lu y otros [12] formulado para problemas con condiciones de tipo Dirichlet y para dominios que resultan de la unión de rectángulos. Nuestro objetivo global es desarrollar un esquema numérico que permita la paralelización al nivel de las variables espaciales. Este método, que denominamos **método implícito de direcciones simultáneas** o **SDI** (***S**imultaneous **D**irections **I**mplicit metho*d), descompone

[3] En general, p no tiene porqué coincidir con la dimensión N en la que se trabaje.

el problema original, independientemente de la dimensión espacial del dominio en el que esté formulado, en un conjunto de problemas unidimensionales. Se sustituye así un problema en derivadas parciales por una familia de problemas diferenciales ordinarios todos ellos con la misma estructura y por tanto con las mismas dificultades numéricas, independientemente de la dimensión espacial en la que se trabaje. Además, los problemas resultantes son independientes entre sí y pueden resolverse en paralelo. El método SDI que hemos desarrollado es aplicable a dominios Ω arbitrarios. En [17] se aplica en el caso bidimensional, efectuándose previamente una "rectificación"del abierto Ω, de tal manera que se aproxima el dominio original por una unión de rectángulos. En nuestro caso no se efectúa esta rectificación y el mallado se construye a partir de una retícula ortogonal de partición arbitraria, admitiendo todo tipo de condiciones de contorno. Para ello se ha desarrollado un mallador (ver [9]), tanto en el caso 2D como en el 3D, que permite disponer de una estructura de datos que contribuye a una mayor eficiencia del método. El método es económico en sus necesidades de almacenamiento, por lo que es compatible con problemas de gran talla. Además, en [9] se demuestra su carácter "smoother", que explica que se adapte muy bien al uso de técnicas multigrid para acelerar la convergencia sea cual sea la geometría del dominio considerado. Aquí solo expondremos el algoritmo paralelo propuesto cuando la ecuación de Helmholtz se completa con condiciones de contorno de tipo Dirichlet. La consideración de otro tipo de condiciones de contorno rompe en general, con fronteras no necesariamente paralelas a los ejes coordenados, la independencia por direcciones al ser condiciones netamente interespaciales. En [9] se encuentra desarrollada una técnica válida para condiciones de otro tipo. Ésta consiste en un método iterativo que inicialmente necesita la resolución de una sucesión de problemas con condiciones de Dirichlet y posteriormente se reduce a la aplicación del método SDI a un único problema con condiciones de Dirichlet dependientes de la iteración. Recientemente se ha demostrado la convergencia de este método. Hay que destacar que la resolución de los problemas de Helmholtz con condiciones de Neumann tiene interés no sólo por sí misma, sino también cuando se buscan precondicionadores que permiten acelerar la convergencia del gradiente conjugado al aplicar éste, por ejemplo, a la resolución de un problema de Stokes. Como es de preveer, disponer de un método paralelo eficiente para la resolución de los problemas de Helmholtz conduce a una notable mejora en la resolución de problemas más complejos, como puede ser el caso de los problemas de Stokes o Navier-Stokes. En nuestro caso, para resolver el problema de Stokes, reformulamos éste como un problema de mínimos al que aplicamos un métoddo iterativo de tipo gradiente conjugado propuesto en [10] y [5]. Este método conduce a la resolución de un gran número de problemas de Helmholtz y/o Poisson. Por lo tanto, disponemos de una manera de **resolver un problema de Stokes N-dimensional exclusivamente a partir de problemas unidimensionales independientes** (problemas diferenciales ordinarios), con igual esquema de resolución y dificultad numérica independientemente de la dimensión del problema original. Se observa entonces que la paralelización del método es posible a tres niveles, respectivamente correspondientes a la

descomposición de los problemas vectoriales por componentes y a los dos niveles de paralelización implícitos al método SDI. En [5], se resuelven los problemas de Helmholtz y Poisson discretizando por elementos finitos, usándose un mallado el doble de fino para las velocidades que para la presión, al objeto de verificar la condición *inf-sup*. En nuestro caso, discretizamos los problemas unidimensionales resultantes utilizando diferencias finitas y el mismo mallado para todas las incógnitas. Esto hace que no se verifique la condición *inf-sup,* lo que provoca la aparición de presiones *espúreas*, que disminuyen la precisión de la aproximación de la presión calculada sin que la aproximación de la velocidad tenga que verse afectada. Las presiones espúreas se manifiestan en la representación gráfica de las isobaras como líneas de contorno con oscilaciones. En la práctica, según la discretización que se efectúe, el tipo de "test" considerado y el carácter de su solución, puede que explícitamente no se observe la presencia de las mismas. En la sección 3.2 efectuaremos un análisis de esta dificultad introduciendo el desarrollo asintótico del error cometido y siguiendo la metodología propuesta por Wetton [19] en su estudio del método de proyección. Wetton explica la causa de la aparición de presiones espúreas en base a la existencia de "errores alternantes" unidireccionales. Aquí realizaremos un análisis similar, contemplando la existencia de estos errores pero con un carácter multidireccional. Identificaremos los modos espúreos y, de esta manera, podremos buscar el filtrado adecuado para las presiones que haga que las oscilaciones se reduzcan. Este filtrado (ver [9]) se realiza siguiendo cada una de las direcciones espaciales, manteniendo el esquema conceptual establecido de reducir todo el tratamiento a problemas unidimensionales.

2 Un esquema paralelo para la ecuación de Helmholtz

Como acabamos de ver, los métodos de paso fraccionado (8) son algoritmos iterativos *secuenciales*, en los que cada iteración se subdivide en p subetapas que se realizan sucesivamente. En [12], Lu et al. proponen un esquema de descomposición para la resolución de (6) en el que los pasos fraccionados son independientes entre sí. A partir de una descomposición de la matriz A dada por (7) y de un parámetro $\tau > 0$, el paso de U^m a U^{m+1} se realiza calculando $U^{m+1,n}$ $(1 \leq n \leq p)$ a partir de

$$(I + \tau A_n)\, U^{m+1,n} = \left(I - \tau \sum_{k=1, k\neq n}^{p} A_k \right) U^m + \tau f, \quad n = 1, \ldots, p, \tag{10}$$

y haciendo, posteriormente,

$$U^{m+1} = \frac{\omega}{p} \sum_{n=1}^{p} U^{m+1,n} + (1-\omega) U^m, \tag{11}$$

donde ω es un parámetro dado. Se trata de un método de descomposición en el cual los p pasos fraccionados pueden resolverse simultáneamente y la

paralelización se realiza a un nivel más alto, siendo esto mucho más interesante y efectivo. A este método lo denominaremos **método paralelo de paso fraccionado** o **PFS (*P*arallel *F*ractional *S*tep method)**. Esquemáticamente puede representarse como sigue:

$$U^m \begin{array}{c} \nearrow \\ \searrow \end{array} \begin{array}{c} U^{m+1,1} \\ \cdots \\ \cdots \\ U^{m+1,p} \end{array} \begin{array}{c} \searrow \\ \nearrow \end{array} U^{m+1} \tag{12}$$

Cuando se aplica este método al problema de Helmholtz-Dirichlet N dimensional

$$\begin{cases} \mathrm{L}u = \alpha u - \beta \Delta u = f & \text{en } \Omega \subset \mathbb{R}^N, \\ u = g & \text{sobre } \partial\Omega \end{cases} \tag{13}$$

considerando la siguiente descomposición del operador

$$\mathrm{L} = \sum_{n=1}^{N} \mathrm{L}_n = \sum_{n=1}^{N} \left(\frac{\alpha}{N} \mathrm{I} - \frac{\partial^2}{\partial x_n^2} \right),$$

se tiene que, en cada paso (10) se resuelven N problemas independientes correspondiente cada uno de ellos al operador

$$\frac{\alpha}{N} \mathrm{I} - \beta \frac{\partial^2}{\partial x_n^2},$$

al igual que sucedía con el método ADI pero, en este caso, los cálculos se hacen en todas las direcciones simultáneamente. De aquí que a este método lo denominemos **método implícito de direcciones simultáneas** o **SDI** (***S**imultaneous **D**irections **I**mplicit method*)[4]. Como ya indicamos en el método ADI, la descomposición elegida es conveniente, ya que las matrices resultantes $I + \tau A_n$ son más fáciles de invertir que la matriz A del problema original. En este caso, la obtención de cada $U^{m+1,n}$ puede efectuarse resolviendo en paralelo s sistemas independientes con s del orden de h^{-1}, lo que se corresponde con la resolución simultánea en cada una de las líneas del mallado en la dirección x_n. Se tienen por tanto dos niveles de paralelismo que conducen a $N \cdot s$ posibles procesos simultáneos y, dado que s es un número elevado al ser en la práctica h pequeño, obtenemos un algoritmo con un alto índice de paralelización. Además, los problemas a resolver para la obtención de las $U^{m+1,n}$ tienen todos la misma estructura y presentan idénticas dificultades numéricas. Los métodos paralelos de paso fraccionado en general y los SDI en particular, según la descripción (10)–(11) efectuada, dependen de dos parámetros (que pueden variar en cada iteración en el caso no estacionario); el parámetro τ denominado **parámetro de evolución** y el parámetro ω, conocido como **parámetro de coordinación**. En [9] se encuentra un estudio exhaustivo de los valores de estos parámetros,

[4]En general, llamaremos métodos SDI a los métodos paralelos de paso fraccionado aplicados a problemas en los que en cada L_n intervengan derivadas en una única dirección.

que hacen que el método sea convergente y además su velocidad de convergencia sea lo mayor posible. En [1] aparece un estudio del error para un caso particular del algoritmo totalmente discretizado.

2.1 Resolución efectiva

En esta sección efectuamos un análisis de los elementos necesarios para la resolución efectiva de los problemas de tipo Helmhotz-Dirichlet.

Mallado

Como es natural, para la resolución del problema (13) efectuaremos una discretización espacial. A partir de una retícula cartesiana $R(\Delta)$, obtenemos un mallado de Dirichlet $M(\Delta) = M_1(\Delta) \cup ... \cup M_N(\Delta)$, de nodos $N(\Delta) = N_I(\Delta) \cup N_F(\Delta)$, donde $N_I(\Delta)$ contiene los nodos interiores, $N_F(\Delta)$ contiene los nodos frontera y

$$M_n(\Delta) = \bigcup_{j=1}^{M_n} \left(\bigcup_{l=1}^{\widetilde{n}(n,j)} I_n^{j,l} \right), \qquad 1 \leq n \leq N,$$

siendo $I_n^{j,l} = ((a_l), (b_l))$ un segmento de $\mathbb{R}^N$ en la dirección espacial e_n isomorfo al intervalo de $\mathbb{R}$ (a_l, b_l) . En las Figuras 1 y 2, aparece un ejemplo de $M(\Delta)$ para un dominio bidimensional y en la Figura 3 un dominio tridimensional análogo.

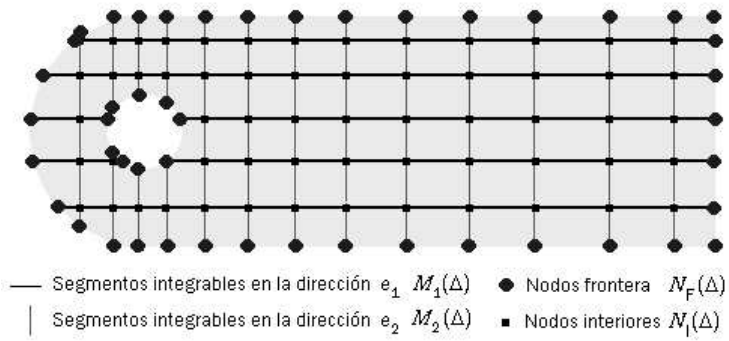


Figura 1: Un mallado $M(\Delta) = M_1(\Delta) \cup M_2(\Delta)$, para un dominio bidimensional.

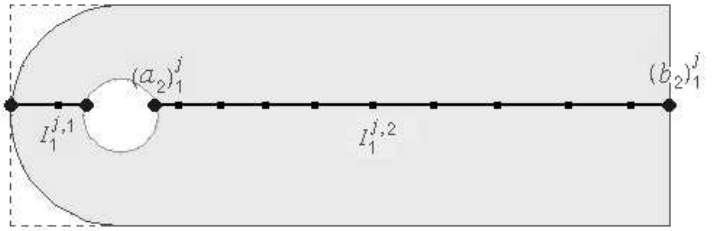


Figura 2: Dos segmentos integrables en la dirección e_1.

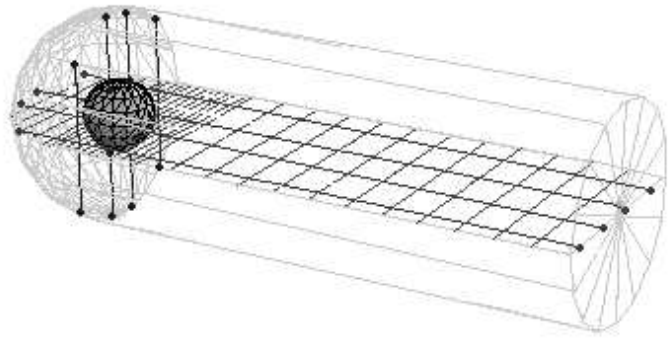


Figura 3: Algunos segmentos integrables en un dominio 3D.

Detalle del algoritmo

$*$ *Etapa de resolución*

La etapa de resolución del algoritmo (10), aplicada al problema (13), se corresponde con N subproblemas del tipo

$$\begin{cases} (\mathrm{I}+\tau \mathrm{L}_n)\, u^{m+1,n} &= \left(\mathrm{I}-\tau \sum_{k=1,k\neq n}^{N} \mathrm{L}_k\right) u^m + \tau f \quad \text{en } \Omega \subset \mathbb{R}^N, \\ u^{m+1,n} &= g \qquad\qquad\qquad\qquad\qquad\quad \text{sobre } \Gamma = \partial\Omega, \end{cases} \tag{14}$$

donde $\mathrm{L}_n = \dfrac{\alpha}{N}\mathrm{I} - \dfrac{\partial^2}{\partial x_n^2}$, $1 \leq n \leq N$ y u^m es conocida. Para el mallado $M(\Delta)$, como L_n es unidimensional, para cada uno de los problemas (14) se tiene el correspondiente problema discreto

$$\begin{cases} (\mathrm{I}+\tau \delta_n)\, U^{m+1,n} &= \left(\mathrm{I}-\tau \sum_{k=1,k\neq n}^{N} \delta_k\right) U^m + \tau f \quad \text{en } M_n(\Delta), \\ U^{m+1,n} &= g \qquad\qquad\qquad\qquad\quad \text{en } N_F(\Delta), \end{cases} \tag{15}$$

donde δ_n es el operador discreto correspondiente a L_n. Dado que las condiciones de contorno son de tipo Dirichlet, (15) es un conjunto de problemas

independientes[5]

$$\begin{cases} (I+\tau\delta_n)\,U^{m+1,n} = \left(I-\tau\sum_{k=1,k\neq n}^{N}\delta_k\right)U^m + \tau f \quad \text{en } I_n^{j,l} = ((a_l),(b_l))\,, \\ U^{m+1,n}\left((a_l)\right) = g\left((a_l)\right)\,, \quad U^{m+1,n}((b_l)) = g((b_l)), \end{cases} \tag{16}$$

definidos en cada uno de los segmentos integrables $I_n^{j,l}$ que constituyen $M_n(\Delta)$. A su vez, cada problema (16) es un problema discreto correspondiente a un problema de contorno ordinario (unidimensional) en la variable x_n del tipo

$$\begin{cases} \left(1+\tau\dfrac{\alpha}{N}\right)v - \tau\beta v'' = \tilde{f} \quad \text{en} \quad (a_l,b_l), \\ v(a_l) = v_a\,, \qquad v(b_l) = v_b, \end{cases} \tag{17}$$

donde v'' es la segunda derivada en la variable x_n y $\tilde{f}$ es un segundo miembro genérico. Así pues, $U^{m+1,n}$ en (16) queda determinada mediante la resolución de, al menos, M_n problemas de contorno ordinarios del tipo (17), cada uno de ellos en el correspondiente intervalo de $\mathbb{R}$, isomorfo al segmento integrable $I_n^{j,l} \in M_n(\Delta)$. Haciendo $n = 1, ..., N$, vemos que (16) es una familia de, al menos, $M = \sum_{n=1}^{N} M_n$ problemas del tipo (17). Todos ellos poseen la misma estructura y, por tanto, presentan igual dificultad numérica. Además, son independientes entre sí por lo que pueden teóricamente resolverse de forma simultánea. Hagamos notar que M es un número relativamente grande, ya que depende del número de puntos de las particiones consideradas en cada dirección espacial. Así pues, se tiene un alto nivel de paralelismo posible.

∗ *Etapa de coordinación*

Determinadas las $U^{m+1,n}$, con $n = 1, ..., N$, en los nodos $N(\Delta)$ del mallado, la etapa de coordinación del algoritmo (11) es también un proceso paralelizable ya que el cálculo

$$U^{m+1}(P) = \frac{\omega}{N}\sum_{n=1}^{N} U^{m+1,n}(P) + (1-\omega)U^m(P) \tag{18}$$

se puede realizar de forma simultánea para cada $P \in N(\Delta)$. Esta etapa de coordinación es realmente la que nos obliga a utilizar una retícula cartesiana como base para la construcción del mallado.

∗ *"Test" de parada*

El "test" de parada considerado en el proceso iterativo es:

$$\|U^m - U^{m+1}\|_\infty = \max_{x\in N(\Delta)} |U^m(x) - U^{m+1}(x)| < \varepsilon,$$

donde ε es una tolerancia dada.

[5]La matriz del problema discreto es una matriz diagonal por bloques.

2.2 Resultados numéricos

En este apartado se presentan los resultados obtenidos al aplicar el método SDI a los siguientes problemas "test":

- Un problema bidimensional ($N = 2$). Se ha resuelto el problema (13) en $\Omega = (0,1) \times (0,1)$ con $\alpha = 0$, $\beta = 1$ y $f \equiv 2$. La solución exacta viene dada por
$$u(x_1, x_2) = \text{senh}(\pi x_1)\ \text{sen}(\pi x_2) + x_1(1 - x_1).$$
- Un problema tridimensional ($N = 3$). En este caso $\Omega = (0,1) \times (0,1) \times (0,1)$, $\alpha = 0$, $\beta = 1$ y $f \equiv 4$. La solución exacta es la función
$$\begin{aligned} u(x_1, x_2, x_3) &= \text{senh}(\pi x_1)\ \text{sen}(\pi x_2) + \text{senh}(\pi x_1)\ \text{sen}(\pi x_3) \\ &\quad + x_1(1 - x_1) + x_3(1 - x_3)\,. \end{aligned}$$

Las condiciones de contorno impuestas en ambos casos se deducen a partir de la restricción de la solución a $\partial\Omega$. Aunque, como ya indicamos, el algoritmo es aplicable a dominios arbitrarios, hemos considerado aquí los casos más sencillos, un cuadrado y un cubo, con objeto de poder obtener un análisis comparativo para diferentes particiones regulares, ya que el tamaño de los problemas unidimensionales (número de puntos de la partición en cada intervalo) y el número de éstos, tiene una incidencia directa en los resultados obtenidos. El algoritmo ha sido implementado en un ordenador SGI Origin 2000 con 8 procesadores, utilizando el modelo de computación paralela del Consorcio OpenMP. La versión de las herramientas de paralelización utilizadas no admite regiones paralelas anidadas, por lo que no se ha podido aprovechar todo el paralelismo inherente al método descrito. Es de esperar, por tanto, que los resultados que se presentan puedan mejorarse. Para medir el comportamiento del algoritmo paralelo se introducen dos parámetros, la ganancia en velocidad S_{NP}, defifinida por

$$S_{NP} = \frac{\text{Tiempo de resolución con 1 procesador}}{\text{Tiempo de resolución con NP procesadores}}$$

y la eficiencia

$$\eta = \frac{S_{NP}}{NP}\,,$$

que representa la ganancia en velocidad por procesador. Idealmente, si todo el código se ejecutara en paralelo, la ganancia en velocidad debería coincidir con el número de procesadores empleados. Pero siempre hay un fracción que se debe ejecutar en forma secuencial, lo que hace que este ideal teórico nunca se alcance en la práctica. En las Figuras 4 y 5 se muestran la ganancia en velocidad y la eficiencia correspondiente a los dos "tests" anteriormente señalados. El comportamiento observado es muy similar en dimensión dos y tres. Para mallados gruesos, la paralelización no mejora sustancialmente los resultados numéricos. De hecho, en el caso 2D de un mallado 65×65, los resultados

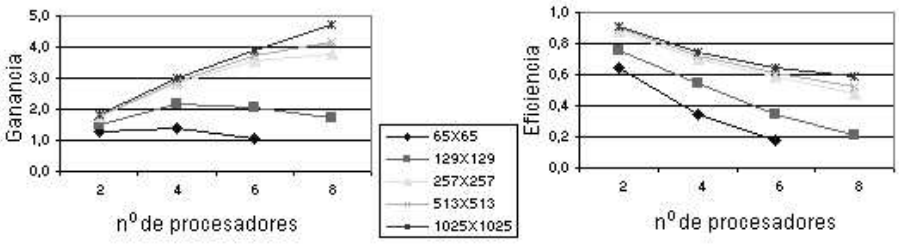


Figura 4: Ganancia en velocidad y eficiencia. "Test" 2D.

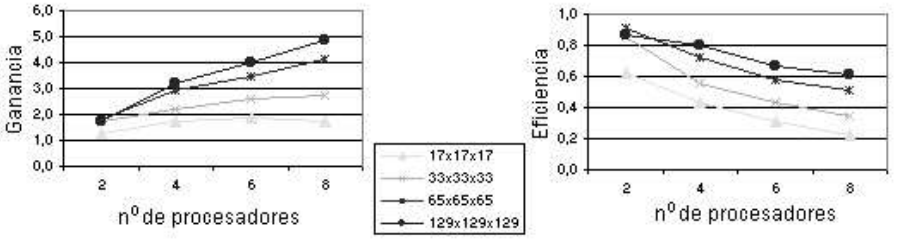


Figura 5: Ganancia en velocidad y eficiencia. "Test" 3D.

correspondientes a 8 procesadores son peores que los que se obtienen trabajando en forma secuencial. Esto se justifica por el hecho de que, para este mallado, el coste de lanzar los procesos es posiblemente mayor que el beneficio obtenido, ya que el trabajo computacional de cada procesador es demasiado pequeño. Por el contrario, cuando el número de nodos es grande, tanto la ganancia en velocidad como la eficiencia aumentan de forma considerable.

Hemos comprobado que el método SDI es eficiente, pero también necesitamos acelerar su convergencia en las situaciones de mayor interés. Para ello nos planteamos la inclusión adicional de técnicas multigrid. En [9] se realiza un estudio detallado del algoritmo, buscando su buena adaptabilidad a este tipo de técnicas, demostrándose mediante un análisis local de Fourier que, efectivamente, el método SDI aplicado al problema de Helmholtz se comporta como un "smoother". En las aplicaciones prácticas se incorpora la técnica de *Iteración Anidada* (IA), que consiste en comenzar con un grid "grueso" e ir refinando hasta llegar al grid deseado, de tal manera que la solución parcial en un nivel se obtiene partiendo de la solución encontrada en el nivel anterior. Con esta estrategia conseguimos obtener unas buenas aproximaciones iniciales con poco coste computacional. Aunque veremos que se acelera considerablemente la convergencia, no se aprovecha totalmente el carácter "smoother" que posee el método. Para poder comparar los resultados obtenidos considerando iteración

anidada, definimos el factor de mejora

$$\text{Mejora} = \frac{\text{Tiempo sin IA y P procesadores}}{\text{Tiempo con IA y P procesadores}},$$

que se calcula para cada número de procesadores empleando la paralelización y refleja la mejora obtenida por el uso de esta técnica. Análogamente, el factor de mejora conjunta

$$\text{Mejora conjunta} = \frac{\text{Tiempo secuencial sin IA}}{\text{Tiempo paralelo con IA y P procesadores}}$$

refleja la mejora obtenida cuando usamos conjuntamente paralelización e iteración anidada. En las gráficas 6 y 7 se comparan los resultados obtenidos con iteración anidada y sin iteración anidada, de los ejemplos "tests" anteriores con un mallado de 1025×1025 nodos en el caso 2D y de $129 \times 129 \times 129$ en 3D.

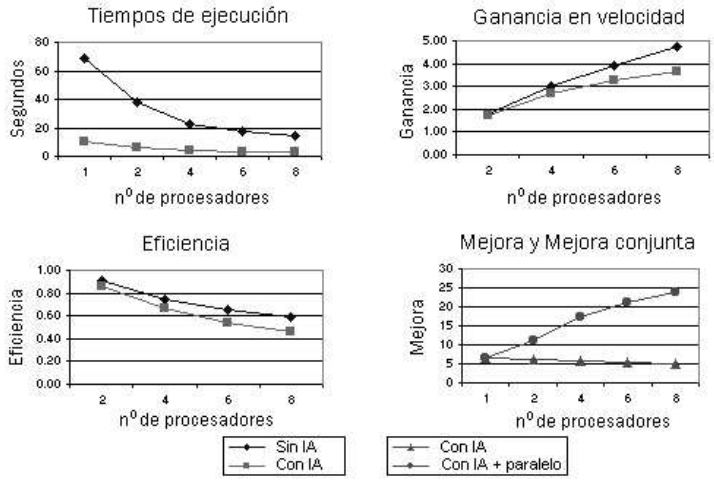


Figura 6: Comparación algoritmo sin IA y con IA. Caso $N = 2$.

Del análisis de estos datos, podemos observar que, la aplicación de iteración anidada (IA) conduce a una sustancial reducción de los tiempos de ejecución. El tiempo en secuencial en este caso es incluso inferior que sin IA utilizando ocho procesadores. La ganancia en velocidad, donde reflejamos la ganancia correspondiente a la paralelización de cada uno de los procesos de manera independiente, es algo inferior en la gráfica 6 al aplicar IA, debido a que el mayor número de iteraciones se efectúa en los niveles con menor número de nodos, que ya indicamos eran menos eficientes. El factor de mejora disminuye al aumentar el número de procesadores y la mejora conjunta aumenta al aumentar el número de procesadores, siendo en el caso bidimensional 20 veces más rápido (30 en 3D) con IA y 8 procesadores que sin IA y en secuencial. Presentamos

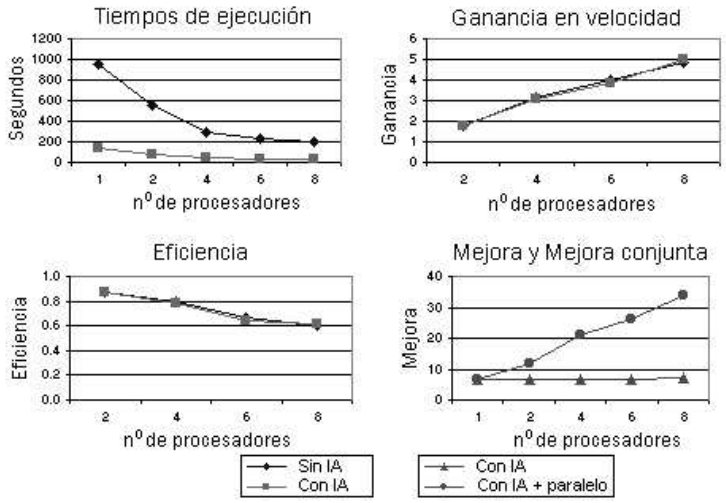


Figura 7: Comparación algoritmo sin IA y con IA. Caso $N = 3$.

por último en la gráfica 8 una comparación de los resultados que se obtienen cuando aplicamos el método SDI con y sin iteración anidada a otros problemas bidimensionales cuyas soluciones son, respectivamente, una función exponencial, una función senusoidal, una función parabólica y la solución del problema "test" 2D reflejado en los estudios anteriores (un seno hiperbólico). En todos los casos se obtiene una mejora conjunta superior a 20.

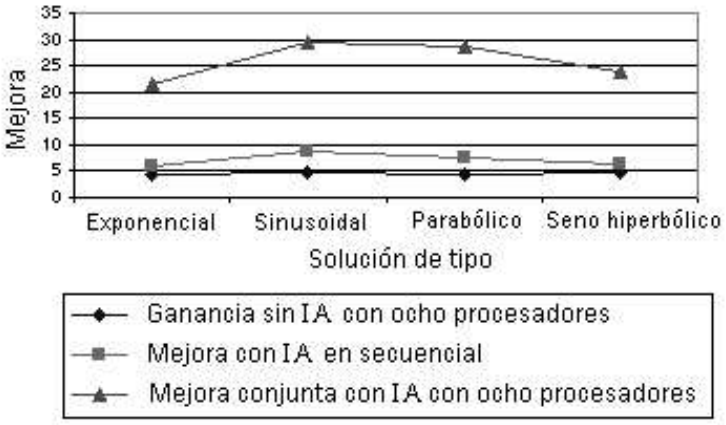


Figura 8: Mejora obtenida para otros "tests" 2D.

3 El problema de Stokes

Consideremos el problema de Stokes estacionario

$$\begin{cases} \alpha\mathbf{u} - \nu\Delta\mathbf{u} + \nabla p = \mathbf{f} & \text{en } \Omega, \\ \nabla \cdot \mathbf{u} = 0 & \text{en } \Omega, \\ \mathbf{u} = \mathbf{0} & \text{sobre } \Gamma = \partial\Omega, \end{cases} \tag{19}$$

donde $\Omega \subset \mathbb{R}^N$ es un dominio acotado de frontera Γ suficientemente regular, $\alpha \geq 0$, $\nu > 0$, $\mathbf{f} : \Omega \subset \mathbb{R}^N \to \mathbb{R}^N$ es una función conocida y las incógnitas son

$$\mathbf{u} : \Omega \subset \mathbb{R}^N \to \mathbb{R}^N \quad \text{y} \quad p : \Omega \subset \mathbb{R}^N \to \mathbb{R}.$$

El problema (19) es un sistema lineal, elíptico de segundo orden, caracterizado porque una de sus incógnitas, la presión p, sólo aparece a través de sus derivadas de primer orden y no está sometida a condiciones de contorno y también por estar acopladas las diferentes componentes de la velocidad $\mathbf{u}$ por la condición de incompresibilidad $\nabla \cdot \mathbf{u} = 0$. Se trata de un modelo estacionario para el estudio del fluidos incompresibles con viscosidades grandes (y por tanto velocidades lentas). También se corresponde con un modelo físico para problemas de elasticidad incompresible. Independientemente de ello, la resolución de este problema es un paso previo a la resolución de las ecuaciones evolutivas de Navier-Stokes, ya que sirve como modelo lineal que incluye las dificultades inherentes al tratamiento de la condición de incompresibilidad. Si Ω_h es una discretización regular del dominio Ω de paso h, el problema (19) puede ser aproximado por un problema discreto (20), que depende de la elección que efectuemos de los *operadores discretos* Δ_h , $\mathbf{G}_h$ y $\mathbf{D}_h$, siendo éstos aproximaciones consistentes de los operadores Laplaciano, gradiente y divergencia habituales, respectivamente:

$$\begin{cases} \alpha\mathbf{U} - \nu\Delta_h\mathbf{U} + \mathbf{G}_h P = \mathbf{f} & \text{en } \Omega_h, \\ \mathbf{D}_h \cdot \mathbf{U} = 0 & \text{en } \Omega_h, \\ \mathbf{U} = \mathbf{0} & \text{sobre } \Gamma_h = \partial\Omega_h. \end{cases} \tag{20}$$

La aplicación en el mallado Ω_h de estos operadores discretos requiere, en general, una aproximación adecuada sobre la frontera Γ_h y la elección de ésta determina problemas discretos diferentes. En el contexto de las técnicas de diferencias finitas, la elección más simple es utilizar diferencias centradas sobre un mismo mallado regular tanto para la velocidad como la presión ("non-staggered grid"); pero esta elección puede generar soluciones oscilantes, ver [8], que como ya hemos comentado suelen manifestarse en la presión. Para no generar dichas soluciones oscilantes, una técnica ampliamente utilizada consiste en acoplar las componentes de la velocidad y la presión con mallados diferentes ("staggered grid" o "semi-staggered grid"). Como ejemplo, podemos citar el método "Marker and Cell" (MAC), formulado por Harlow and Welch en 1965, véase [8]. Sin embargo, el uso de mallados de este tipo lleva consigo algunas desventajas, principalmente a nivel de codificación e implementación, alcanzando su máxima dificultad cuando se consideran dominios no rectangulares. A continuación,

efectuaremos el análisis del error cometido al resolver el problema de Stokes utilizando la técnica del desarrollo asintótico del error de Strang [16], distinguiendo en este desarrollo los denominados términos "regulares" señalados por este autor y términos "alternantes" señalados por Wetton [19]. Hay que destacar el carácter descriptivo de la técnica empleada, que permite identificar qué ocurre en la discretización de un problema continuo, cómo y cuándo se generan presiones espúreas en el problema de Stokes y por qué éstas no se manifiestan explícitamente en determinados casos en los que deberían presentarse. Este estudio es más difícil de abordar si, en lugar de utilizar diferencias finitas se emplearan, por ejemplo, elementos finitos para discretizar las variables espaciales.

3.1 Desarrollo asintótico del error en un problema discreto

Para introducirnos en el tema, consideremos el problema de contorno unidimensional

$$\begin{cases} \mathcal{L}\,u = f & \text{en } \Omega, \\ \ell u = g & \text{sobre } \Gamma = \partial\Omega, \end{cases} \tag{21}$$

donde $\Omega = (0,1)$. Tomemos una discretización regular Ω_h, de paso h, constituida por $n+1$ nodos $\{(ih)\}_{i=0}^{n}$, por simplicidad tomaremos n par. Aproximamos (21) por el problema discreto

$$\begin{cases} \mathcal{L}_h\,U^h = f_h & \text{en } \Omega_h, \\ \ell_h\,U^h = g_h & \text{sobre } \Gamma_h, \end{cases} \tag{22}$$

siendo $\mathcal{L}_h$ y ℓ_h aproximaciones consistentes de orden h^s, de los operadores continuos $\mathcal{L}$ y ℓ, respectivamente, obtenidas utilizando diferencias finitas. Denotando U_i^h una aproximación de $u(ih)$, diremos que la solución discreta U_i^h converge con orden h^s y con un desarrollo asintótico regular (véase [16]), si

$$U_i^h = u(ih) + h^s u^{(s)}(ih) + O(h^{s+1}), \quad \text{para} \quad 0 \le i \le n \tag{23}$$

donde $u^{(s)}$ es una función regular, independiente de h. Si el operador $\mathcal{L}_h$ desacopla el mallado Ω_h en dos submallados independientes constituidos, respectivamente, por los nodos de numeración par e impar y, además, la aproximación en la frontera respeta este desacoplamiento, introduciendo aproximaciones diferentes en cada submallado, obtenemos un problema diferente en cada uno de ellos. De esta manera, la solución aproximada U_i^h en el *submallado par* tiene un desarrollo regular

$$u + h^s u^{(sp)} + \cdots$$

y en el *impar*

$$u + h^s u^{(si)} + \cdots .$$

Las funciones $u^{(sp)}$ y $u^{(si)}$ son regulares e independientes de h, pero diferentes entre sí, por lo que la solución U_i^h de (22) tiene un desarrollo asintótico que puede expresarse como:

$$U_i^h = u(ih) + h^s u^{(s)}(ih) + h^s \widehat{u}^{(s)}(ih)(-1)^i + \cdots , \tag{24}$$

donde las

$$\widehat{u}^{(s)} = \frac{u^{(sp)} - u^{(si)}}{2}, \quad u^{(s)} = \frac{u^{(sp)} + u^{(si)}}{2}$$

son regulares e independientes de h. Decimos que este tipo de desarrollo es *alternante*. Esto da una explicación precisa de qué es lo que ocurre cuando se pierde la regularidad discreta. Así pues, aparece una solución global que oscila al cambiar de submallado ya que el desarrollo asintótico del error contiene términos *alternantes* que, en caso de que predominen frente a los términos *regulares*, causan la pérdida de precisión, véase [19]. Notemos que el desacoplamiento es debido a la discretización elegida del operador y el orden en el que aparece el término alternante es el orden en el que difieren las aproximaciones frontera consideradas en cada submallado. Los tipos de desarrollo asintótico del error indicados pueden extenderse a cualquier problema de contorno bi o tridimensional. Como conclusión, se tiene que mediante este análisis podemos averiguar si un esquema de discretización es regularizante y, en el caso de tener un desarrollo alternante, se puede determinar, mediante una adecuada ponderación de los valores de nuestra incógnita en un nodo y sus adyacentes, un patrón de filtrado que permita aminorar la pérdida de precisión indicada. En [19] se efectúa un análisis del esquema de discretización de Chorin [2] aplicado, en este caso, al problema de Stokes. Se demuestra que el error es de segundo orden para la velocidad (con errores de tipo alternante de tercer orden) y sólo de primer orden en la presión debido a la presencia de errores alternantes, que predominan sobre el término regular de segundo orden. Sin embargo, este esquema es aplicable sólo a dominios rectangulares o compuestos por unión de rectángulos, dado que es preciso calcular una *divergencia discreta* en los nodos frontera.

3.2 El esquema regularizante propuesto

Para no extender la exposición, haremos el desarrollo en el caso bidimensional. De forma análoga puede hacerse para el caso N-dimensional. Para la resolución de (19), definido en un dominio arbitrario, con objeto de obtener un esquema regularizante, consideraremos el problema discreto (20), donde:

- Ω_h es el dominio discreto determinado a partir de una retícula cartesiana regular, que será el mismo para la velocidad y la presión.

- Δ_h es el operador de Laplace discreto bidimensional, definido por la aproximación de cinco puntos de orden h^2, es decir:

$$\Delta_h z_{i,j} = \frac{z_{i-1,j} + z_{i+1,j} - 4z_{i,j} + z_{i,j-1} + z_{i,j+1}}{h^2}. \tag{25}$$

- $\mathbf{G}_h = (G_h^x, G_h^y)$ y $\mathbf{D}_h$ son los operadores gradiente y divergencia discreta, obtenidos mediante aproximaciones de segundo orden basadas

en diferencias centradas:

$$G_h^x P_{i,j} = \frac{P_{i+1,j} - P_{i-1,j}}{2h}, \tag{26}$$

$$G_h^y P_{i,j} = \frac{P_{i,j+1} - P_{i,j-1}}{2h}, \tag{27}$$

$$\begin{aligned} \mathbf{D}_h \cdot \mathbf{U} &= \mathbf{D}_h \cdot (U,V)_{i,j} \\ &= \frac{U_{i+1,j} - U_{i-1,j}}{2h} + \frac{V_{i,j+1} - V_{i,j-1}}{2h}. \end{aligned} \tag{28}$$

Calcularemos *la presión exclusivamente en los nodos interiores*, por lo que sólo es necesario evaluar la divergencia y el gradiente discreto en dichos nodos. De aquí que este esquema puede aplicarse en problemas con dominios no necesariamente rectangulares. Además, se tiene lo siguiente:

- El operador divergencia discreta $\mathbf{D}_h$ desacopla el mallado 2D en cuatro submallados (ocho en 3D), pero éste sólo se aplica en los nodos interiores. Dado que los valores de la velocidad son conocidos sobre Γ_h, todos los submallados "ven" igual aproximación en la frontera.

- El operador gradiente discreto $\mathbf{G}_h$ (26)–(27) respeta los submallados anteriores, pero su aplicación en los nodos adyacentes a la frontera requiere la elección de una adecuada aproximación. Por ejemplo, en los nodos de referencia $(1,j)$, según la notación habitual, la aplicación de (26) conduciría a

$$G_h^x P_{1,j} = \frac{P_{2,j} - P_{0,j}}{2h}, \tag{29}$$

es decir, sería necesario conocer la presión $P_{0,j}$ en los nodos frontera de referencia $(0,j)$. Pero en estos nodos hemos excluído el cálculo de la presión y, por consiguiente, debemos realizar una aproximación que efectuaremos mediante una extrapolación unidireccional, dada por

$$P_{0,j} = 3P_{1,j} - 3P_{2,j} + P_{3,j}. \tag{30}$$

El error de truncamiento asociado es

$$e_{0,j} = p(0,jh) - P_{0,j} = h^3 p_{xxx} + O(h^4). \tag{31}$$

Así pues, de (29) y (30) se tiene que

$$G_h^x P_{1,j} = \frac{-3P_{1,j} + 4P_{2,j} - P_{3,j}}{2h}, \tag{32}$$

que no es más que la fórmula de *diferencias progresivas* de segundo orden. Fórmulas análogas a ésta (diferencias progresivas o regresivas) se obtienen en los nodos adyacentes al resto de la frontera.

Según (31), (32) y expresiones análogas, cada uno de los submallados antes citados "ve" diferentes condiciones de contorno para la presión, que se

diferencian en términos de orden h^3. Por consiguiente, el desarrollo asintótico de la presión, aproximada por este esquema, tiene un término alternante de orden h^3, es decir, un término de la forma

$$\widehat{p}_{i,j}^{(3)} = p^{(3)}(ih, jh) + (-1)^i p^{(3i)}(ih, jh) + (-1)^j p^{(3j)}(ih, jh) + (-1)^{i+j} p^{(3ij)}(ih, jh)$$

donde las $p^{(3)}$, $p^{(3i)}$, $p^{(3j)}$ y $p^{(3ij)}$ son funciones regulares e independientes de h. Se tiene entonces el siguiente resultado, cuya demostración puede encontrarse en [9]:

Teorema 1 *Si la solución* $(\mathbf{u}, p)$ *del problema de Stokes* (19) *es suficientemente regular, entonces la solución* $(\mathbf{U}, P)$ *del problema discreto* (20)*, correspondiente a los operadores* (25)–(28) *actuando en los nodos interiores y tomando sobre la frontera los valores extrapolados para la presión, tiene un desarrollo asintótico del error dado por*

$$\mathbf{U} = \mathbf{u} + h^2 \mathbf{u}^{(2)} + h^5 \widehat{\mathbf{u}}^{(5)} + \cdots, \tag{33}$$

$$P = p + h^2 p^{(2)} + h^3 \widehat{p}^{(3)} + \cdots, \tag{34}$$

donde $\mathbf{u}^{(2)}$ *y* $p^{(2)}$ *son términos regulares y* $\widehat{\mathbf{u}}^{(5)}$ *y* $\widehat{p}^{(3)}$ *son términos alternantes. Por tanto, las aproximaciones de* $\mathbf{U}$ *y de* P *son ambas de segundo orden en* h*.*

3.3 Experiencias numéricas

"Test" en dominio rectangular

Con objeto de comprobar el comportamiento numérico del esquema propuesto y compararlo con el esquema de discretización de Chorin (aplicable sólo en dominios rectangulares), consideraremos inicialmente el "test" con solución analítica propuesto por J.T. Oden en [14]. Consiste en la resolución de (19) en el cuadrado $\Omega = (0,1) \times (0,1)$, con $\alpha = 0$, $\nu = 1$ y $\mathbf{f}$ elegida para que la solución exacta venga dada por:

$$\begin{aligned} u_1(x,y) &= 8x^2(1-x)^2 y(y-1)(2y-1), \\ u_2(x,y) &= 8y^2(1-y)^2 x(x-1)(1-2x), \\ p(x,y) &= 4(x-x^2). \end{aligned} \tag{35}$$

La presión toma valores en el intervalo $[0,1]$, siendo las isobaras segmentos verticales. La implementación práctica se efectúa aplicando un algoritmo tipo gradiente conjugado en el que se integra el método SDI (ver [9]), de aquí que el método sea iterativo. En la Figura 9 resumimos los resultados obtenidos al aplicar los esquemas anteriormente citados: el método de Chorin sin filtrar las presiones, el método de Chorin filtrando las presiones según el patrón obtenido en [9] y el esquema regularizante propuesto. En concreto, en cada columna de la gráfica, representamos:

- El comportamiento, respecto al número de iteraciones, de la norma infinito del error exacto tanto del campo de velocidades como de la presión, para diferentes valores del paso de discretización h.

- Las isobaras para el caso $h = 0{,}05$.
- El error exacto de la presión mediante un gráfico de superficie cuando $h = 0{,}05$.

Para el primer esquema (primera fila de gráficos en la Figura 9), podemos observar que la existencia de un término alternante de primer orden en la presión conduce a la pérdida de precisión, que se refleja en la situación final alcanzada en el error exacto y que es del orden teórico demostrado en [19]. La velocidad no se ve afectada. Son muy significativas las oscilaciones de las isobaras, así como el predominio del error de tipo alternante frente al regular, que se aprecia en el gráfico de superficie. También se puede observar que el error está asociado a los diferentes submallados. En el segundo esquema, el filtro aplicado a la presión logra que el término alternante de primer orden se traslade al tercer orden, por lo que pasa a predominar el término regular de segundo orden y, por tanto, se regulariza la presión (este hecho está demostrado en [9]).

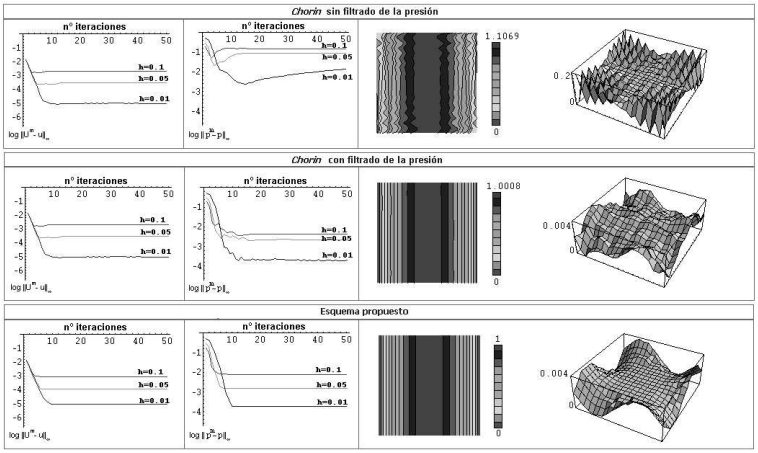


Figura 9: Resultados del "test" de Oden para diferentes esquemas.

Podemos observar en la segunda fila de esa misma figura que el comportamiento de la velocidad es el mismo, pero la presión mejora su precisión, alcanzándose el orden indicado y manifestándose pequeños cambios en la convergencia debido al término alternante, ya de tercer orden. Las isobaras reflejan leves oscilaciones, debido más al error propio de un método aproximado que a la presencia de presiones espúreas. En el gráfico de superficie, se observa el predominio del término regular de segundo orden y la presencia del error alternante en un orden superior. No se aprecia ya distinción entre los diferentes submallados. En el tercer caso, de acuerdo con lo demostrado anteriormente, se logra alcanzar el error regular de segundo orden, tanto para la velocidad como para la presión, no apareciendo oscilaciones en las isobaras. El gráfico de superficie también refleja el comportamiento regular del error y no se observa

dependencia de los submallados. Hemos realizado un número de iteraciones muy superior al necesario, para poder intuir el comportamiento asintótico del algoritmo. Omitimos la representación gráfica del campo de velocidades dado que, según lo descrito anteriormente, éste no se ve afectado en la práctica.

"Tests" 2D y 3D en dominios no rectangulares

Como ejemplo de "tests" bidimensional y tridimensional definidos en dominios no rectangulares, consideraremos la determinación de un fluido estático ("No Flow") sobre el que actúa una fuerza gravitacional. En la Figura 10, reflejamos las presiones obtenidas mediante líneas y superficies isobáricas, respectivamente. Se puede comprobar, de nuevo, el carácter regularizante del esquema propuesto, no apareciendo presiones espúreas. En [9], pueden encontrarse resultados con otros "tests" efectuados.

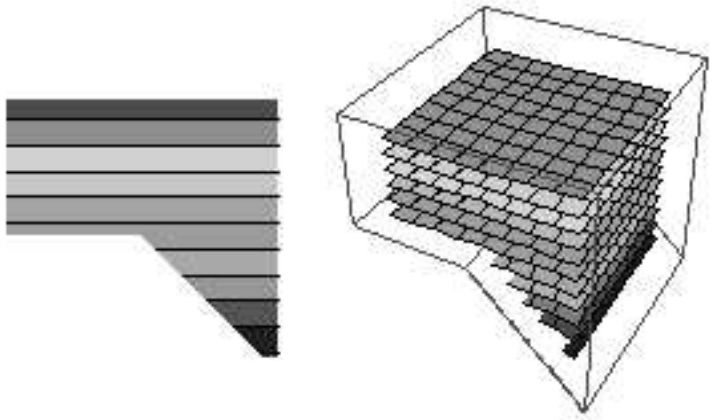


Figura 10: Isobaras y superficies isobáricas en los "tests" "No Flow" 2D y 3D.

Agradecimientos

Trabajo parcialmente financiado por la D.G.E.S., Proyecto BFM2003–06446–C02–02 y por la Junta de Andalucía, ACC–1204–FQM124.

Referencias

[1] I.I. Albarreal - *Paralelización en tiempo y espacio de la resolución numérica de algunas ecuaciones en derivadas parciales*, Tesis, Universidad de Sevilla (2004).

[2] A.J. Chorin - *On the convergence of discrete approximations to the Navier-Stokes equations*, Math. Comp, **22** (1968), 745–762.

[3] J.L. Cruz, M.C. Calzada, E. Fernández-Cara, M. Marín - *A parallel algorithm for solving the incompressible Navier-Stokes problem*, Comp. Math. with Appl., **25 (9)** (1993), 51–28.

[4] J.L. Cruz, M.C. Calzada, E. Fernández-Cara, M. Marín - *A convergence result for a parallel algorithm for solving the Navier-Stokes equations*, Comp. Math. with Appl., **35 (4)** (1998), 71–88. 1995.

[5] E.J. Dean, R. Glowinski - *On Some Finite Element Methods for the Numerical Simulation of Incompressible Viscous Flow*, Incompressible computational fluid dynamics (M.D. Gunzburger and R.A. Nicolaides, Eds.), Cambridge U.P., USA 1993, 17–65.

[6] J. Douglas Jr., J.E. Gunn - *Alternating direction methods for parabolic systems in m space variables*, J. Assoc. Comput. Mach., **9** (1962), 450–456.

[7] J. Douglas Jr., H.H. Rachford Jr. - *On the numerical solution of heat conduction problems in two and three space variables*, Trans. Amer. Math. Soc., **82** (1956), 421–439.

[8] C. Fletcher - *Computational techniques for Fluids Dinamics. Vol 2. Specific techniques for different flow categories*, Springer-Verlag, Berlin (1991).

[9] J.R. Galo - *Resolución en paralelo de las ecuaciones de Navier-Stokes 2D y 3D mediante el método de direcciones simultáneas*, Tesis, Universidad de Sevilla (2002).

[10] R. Glowinski - *Supercomputing and the finite element approximation of the Navier-Stokes equations for incompressible viscous fluids*, Recent advances in computational fluid dynamics (Princeton, NJ, 1988), Springer, Berlin 1989, 277–315.

[11] S.L. Johnsson, Y. Saad, M.H. Schultz - *Alternating direction methods on multiprocessors*, SIAM J. Sci. Statist. Comput., **8 (5)** (1987), 686–700.

[12] T. Lu, P. Neittaanmaki, X.C. Tai - *A parallel splitting-up method for partial differential equations and its applications to Navier-Stokes equations*, RAIRO Modél. Math. Anal. Numér., **Vol. 26 (6)** (1992), 673–708.

[13] G.I. Marchuk - *Splitting and alternating direction methods*, Handbook of numerical analysis, Vol. I, North-Holland, Amsterdam 1990, 197–462.

[14] J.T. Oden, O. Jacquote - *Stability of some mixed finite element methods for Stokesian flows*, Comput. Methods Appl. Mech. Eng., **43(2)** (1984), 231–247.

[15] D.W. Peaceman, H.H. Rachford Jr. - *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Indust. Appl. Math., **3** (1955), 28–41.

[16] G. Strang - *Accurate partial difference methods. I. Non-linear problems*, Numer. Math., **6** (1964), 37–46.

[17] X.C. Tai, P.Ñeittaanmaki - *Parallel finite element splitting-up method for parabolic problems*, **7 (3)** (1991), 209–225.

[18] R.S. Varga - *Matrix iterative analysis*, Springer-Verlag, Berlin 2000.

[19] B.R. Wetton - *Analysis of the spatial error for a class of finite difference methods for viscous incompressible flow*, SIAM J. Numer. Anal., **2** (1997), 723–755.

Bol. Soc. Esp. Mat. Apl. nº 31(2005), 53–72

Análisis de algunos modelos de dinámica de poblaciones no lineales estructurados en edad y espacio*

M. Delgado, M. Molina-Becerra y A. Suárez

Departamento de Ecuaciones Diferenciales y Análisis Numérico, Universidad de Sevilla

madelgado@us.es, monica@us.es y suarez@us.es

Resumen

En este trabajo estudiamos modelos de la dinámica de poblaciones estructurados en edad con difusión. Construiremos un método de sub-supersolución y veremos que, efectivamente, bajo la hipótesis de la existencia de un par de sub-supersoluciones, encontramos existencia de solución entre dicho par. Después, aplicaremos este método a un problema logístico generalizado y a un modelo de tipo Holling-Tanner.

Palabras clave: *modelos estructurados en edad, método de sub-supersolución, condición inicial no local*

Clasificación por materias AMS: *95B25, 35K57, 35B05, 35Q80*

1 Introducción

1.1 Una breve introducción a los modelos estructurados en edad sin difusión

En los últimos doscientos años se han usado diversas teorías matemáticas para modelar y comprender el comportamiento de poblaciones humanas, animales, células, epidemias...

La primera contribución significativa a la teoría de la dinámica de poblaciones fue la de T. Malthus, quien en 1798 publicó su célebre *Ensayo sobre el principio de la población*. Su modelo lineal era satisfactorio siempre que la población no fuese demasiado grande. Es importante señalar que cuando la

*Este trabajo ha sido parcialmente financiado por el Ministerio de Ciencia y Tecnología bajo el proyecto BFM2003-06446-C02-01

Fecha de recepción: 25/01/05

población es "grande", los modelos lineales no pueden ser exactos ya que no contemplan el hecho de que los individuos compiten entre sí por los recursos (como puede ser comida, espacio...). Posteriormente, en 1837, P. F. Verhulst añadió un término no lineal a la ley de Malthus que tenía en cuenta esta limitación de crecimiento, formulando un modelo conocido como *la ley logística*.

Más adelante se introduce la consideración de la edad de la población, que es un dato decisivo por ejemplo para las tasas de natalidad y mortalidad de las especies. F. R. Sharpe, A. Lotka (1911) y A. G. McKendrick (1926) (véanse [33, 29]) fueron los primeros en introducir la edad en los modelos de dinámica de poblaciones.

El *modelo de Sharpe-Lotka-McKendrick*, también llamado *modelo de von Foerster-McKendrick* o simplemente *ecuación de McKendrick*, supone que una población asexual puede ser descrita por una función de dos variables: edad y tiempo. En él se denota $\rho(a,t)$ la densidad de población de edad a en el instante de tiempo t. Entonces se tiene que

$$D\rho := \lim_{h\to 0} \frac{\rho(a+h,t+h)-\rho(a,t)}{h}$$

es la razón con la que la población de edad a cambia en tiempo. Cuando ρ es diferenciable, claramente, se tiene que

$$D\rho = \rho_t + \rho_a.$$

Se denota $d(a,t)$ el número de individuos de edad a que mueren en el tiempo t. Se supone que

$$d(a,t) = \mu(a)\,\rho(a,t),$$

donde $\mu(a)$ es la tasa de mortalidad para los individuos de edad a. A su vez, se supone que el proceso de nacimiento es descrito por la ecuación de renovación

$$\rho(0,t) = \int_0^\infty \beta(a)\rho(a,t)\,\mathrm{d}a,$$

donde $\beta(a)$ es la tasa de descendencia que tiene un individuo de edad a. Por último, se añade una condición inicial

$$\rho(a,0) = \rho_0(a).$$

El modelo resultante es

$$\begin{cases} D\rho(a,t) + \mu(a)\rho(a,t) = 0, \\ \rho(a,0) = \rho_0(a), \\ \rho(0,t) = \displaystyle\int_0^\infty \beta(a)\,\rho(a,t)\,\mathrm{d}a. \end{cases}$$

Este modelo es lineal y, como hemos comentado, no es aplicable a muchas situaciones reales. Así, como sucede habitualmente, el modelo ha ido mejorando desde su introducción, quedando incorporados nuevos aspectos y quedando

planteadas nuevas dificultades matemáticas. Así, un modelo más de acuerdo con la realidad hace pensar en variar las tasas de mortalidad y fertilidad por ejemplo con la población total, P (parece lógico suponer que a mayor población mayor será su tasa de mortalidad).

En 1974, Gurtin y MacCamy, [18], introducen una formulación más realista de un modelo de dinámica de poblaciones no lineal, determinista y estructurado en edad. Consideran que tanto la tasa de natalidad como la tasa de mortalidad son funciones no lineales de P. El resultado es entonces el siguiente:

$$\begin{cases} \dfrac{\partial \rho}{\partial t}(a,t) + \dfrac{\partial \rho}{\partial a}(a,t) + \mu(a,P(t))\rho(a,t) = 0, \\ \rho(a,0) = \rho_0(a), \\ \rho(0,t) = \displaystyle\int_0^\infty \beta(a,P(t))\,\rho(a,t)\,\mathrm{d}a \end{cases} \tag{1}$$

donde la población total P está dada por

$$P(t) := \int_0^\infty \rho(a,t)\,\mathrm{d}a.$$

En los últimos años se ha avanzado en el estudio de estos modelos. Se tienen resultados de existencia y unicidad de solución para distintas expresiones de las tasas de nacimiento y mortalidad, en particular modelos logísticos [7], modelos con tasas de natalidad y mortalidad no acotada en tiempo [9], etc.

También se han estudiado otras cuestiones, como la existencia de estados de equilibrios, su estabilidad, el comportamiento asintótico global de las soluciones, etc.

Las técnicas básicas para el estudio de este tipo de modelos son las siguientes:

- Para el estudio de la existencia y unicidad de solución, principalmente, se utiliza un método de integración a lo largo de las líneas características $a+t = cte.$ que nos permite reducir el problema a un sistema de ecuaciones integrales para otras variables, cuya existencia y unicidad de solución se sigue de un teorema de punto fijo.

- Para el comportamiento asintótico, en el caso de procesos lineales, se suele buscar la existencia de soluciones *persistentes*, que tienen la forma $\rho(a,t) = T(t)\,A(a)$ (cf. [5]).

  Para modelos no lineales que no son de ecuaciones separables, los mejores resultados se siguen de la teoría de semigrupos de operadores no lineales (véase por ejemplo [32, 35]).

Ya que muchos problemas de población implican interacciones entre subclases de poblaciones, también se han estudiados modelos de sistemas con estructura de edad [35]. Esta formulación de sistemas se ha utilizado, entre otras muchas, para modelar epidemias. En estos modelos, se divide habitualmente la población p en tres subpoblaciones s, i y r, individuos susceptibles de contraer la enfermedad, individuos infectados por la misma

e individuos inmunes (tras haber contraído la enfermedad y haber sanado), respectivamente. Evidentemente, se tiene que $p = s + i + r$. Se han considerado modelos denotados en la literatura por SIS (susceptibles-infectados-susceptibles) donde pasar la enfermedad no implica inmunidad y por SIR (susceptibles-infectados-inmunes) donde ocurre lo contrario. Modelos como éstos han sido estudiados por ejemplo en [6] y [20].

También, se han incorporado modelos de dinámicas de poblaciones en el que interactúan dos especies diferentes con estructura de edad, como por ejemplo modelos de presa-depredador (cf. [17, 34]). En otros modelos introducidos se mezclan sistemas de presa-depredador con sistemas de epidemias (véase por ejemplo [8] sin estructura en edad y [4, 11] con estructura en edad en la presa).

Pero, en general, la abundante literatura resulta poco sistemática (véase [31]), puesto que es bastante complicado desarrollar argumentos que sirvan para cualquier tipo de problemas con estructura en edad.

1.2 Introducción a los modelos estructurados en edad con difusión

En [16], Gurtin introduce la edad en los modelos de dinámica de poblaciones con difusión. Así, supuso que la evolución de la población estaba gobernada por la ley

$$\rho_t + \rho_a = -div\, q + s,$$

donde $\rho(x, a, t)$ denota la función distribución de la población, $x \in \Omega$ (Ω un abierto acotado y regular de $\mathbb{R}^N$), q el flujo de la población debido a la dispersión y s representa la variación neta de los individuos, debida generalmente a la mortalidad de la especie.

En 1977, Gurtin y MacCamy [19] diferenciaron entre dos clases de difusión: la difusión debida a una dispersión aleatoria (lineal) y la difusión para evitar el hacinamiento, es decir aquélla en la que las especies migran de zonas de alta a baja densidad de población (difusión no lineal). Los autores aplicaron los resultados que se conocían para el caso sin estructura en edad, a una extensión no lineal de un modelo con estructura en edad aparecido en [16]. Supusieron que la tasa de variación de individuos era debida exclusivamente a la muerte en la población y tomaba la forma:

$$s(x, a, t) := -\mu(a, P)\rho,$$

donde μ representa la tasa de mortalidad de la especie y P denota la población total, es decir

$$P(x, t) = \int_0^\infty \rho(x, a, t)\, \mathrm{d}a.$$

El proceso de nacimiento venía modelado por la siguiente ecuación:

$$\rho(x, 0, t) = \int_0^\infty \beta(a, P)\rho(x, a, t)\, \mathrm{d}a.$$

Además, el flujo de la población debido a la dispersión seguía la ley usual

$$q = \rho \mathbf{v},$$

siendo $\mathbf{v}$ la velocidad de difusión. Puesto que estaban interesados en situaciones en las que la difusión evitaba la concentración, asumieron que:

$$\mathbf{v} = -k(a, \rho, P)\nabla P.$$

Así, suponiendo que tanto μ como β y k son independientes de a y además k es independiente de ρ, obtuvieron el sistema de ecuaciones:

$$\begin{cases} \rho_t + \rho_a + \mu(P)\rho = div\left[\rho k(P)\nabla P\right], \\ \rho(x, 0, t) = \beta(P)P(x, t). \end{cases}$$

Integrando en la variable edad, se tiene una ecuación sin estructura en edad.

A partir de la década de los años 80, estos modelos empiezan a desarrollarse y así muchos trabajos aparecen en la literatura [14, 15, 23, 26, 28]. Estudian modelos lineales y no lineales, a causa de la dependencia de la población total en las tasas de mortalidad y fertilidad y también consideran distintos tipos de difusión.

Más recientemente, Kubo y Langlais [21, 22] introducen en el término de variación neta una función lineal positiva, es decir, suponen que la densidad de población puede aumentar por agentes externos. Además, bajo ciertas hipótesis de periocidad en los datos, dan resultados de existencia y no existencia de solución periódica, aplicando los resultados obtenidos a un sistema de epidemias con difusión.

Pero, que nosotros sepamos, en ninguno de estos modelos se estudia el caso en el que la tasa de variación neta de la especie puede deberse tanto a entrada como a salida de individuos (no sólo natalidad y mortalidad, sino también inmigración y emigración) y además es un término no lineal dependiente de la densidad de población.

Últimamente, se ha comenzado con el estudio de los problemas de control y controlabilidad aplicados a problemas lineales y no lineales con difusión (véanse por ejemplo los trabajos de Ainseba, Langlais y Aniţa [1, 2, 3] entre otros).

En este trabajo analizaremos modelos evolutivos no lineales con dependencia en edad y espacio [12]. En la sección siguiente veremos que un método de sub-supersolución para este tipo de problemas funciona y después se lo aplicaremos a un par de problemas ecológicos: a un modelo logístico generalizado y a un modelo de tipo Holling-Tanner. En la última sección, analizaremos los problemas estacionarios en tiempo asociados [10].

2 Modelos evolutivos dependientes en edad con difusión

En esta sección vamos a considerar un modelo semilineal que describe la dinámica de una especie con dependencia en edad y estructura espacial.

El modelo que estudiamos es una generalización de modelos estudiados anteriormente (véanse por ejemplo los modelos aparecidos en los trabajos de Langlais [25, 27, 28]). En éstos se supone que la variación neta de individuos viene únicamente dada por la mortalidad de la especie, es decir es un término de

salida de individuos, mientras que en el modelo que detallaremos a continuación aparece además un término de reacción con el medio donde habita la especie no lineal.

Sea $u(x,a,t)$ la densidad de la población de edad $a>0$, en el instante de tiempo $t>0$ y en la posición $x\in\Omega$, donde Ω es un dominio acotado de $\mathbb{R}^N$, con frontera, $\partial\Omega$, regular. Supondremos que la difusión es lineal, i.e. el flujo de población viene dado por ∇u, donde ∇ es el gradiente con respecto a la variable espacial. Además, asumiremos que la entrada-salida de individuos viene dada por un término de mortalidad y un término de reacción, i.e.

$$s := -\mu(x,a,t)u + f(x,a,t,u),$$

donde $\mu(x,a,t)$ es la tasa de la mortalidad natural de la especie y f describe el efecto del entorno en la población, es decir migración e inmigración de la especie. Tendremos que f es positiva cuando el entorno es favorable y negativa cuando es hostil.

Además, supondremos que los individuos de la población no alcanzan la edad máxima, $A_\dagger$; es decir, mueren antes de alcanzar dicha edad.

Asumiremos que el proceso de nacimiento viene dado por la ecuación

$$u(x,0,t) = \int_0^{A_\dagger} \beta(x,a,t)u(x,a,t)\,\mathrm{d}a,$$

donde $\beta(x,a,t)$ representa la tasa de fertilidad.

Finalmente, supondremos que la frontera $\partial\Omega$ del dominio Ω es inhabitable, lo que implica una condición frontera de tipo Dirichlet homogénea.

Sea T un número real positivo y denotemos $\mathcal{O}$ el abierto $(0,A_\dagger)\times(0,T)$. Entonces, el modelo que consideramos es el siguiente:

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial u}{\partial a} - \Delta u + \mu(x,a,t)u = f(x,a,t,u) & \text{en } \Omega\times\mathcal{O}, \\ u(x,a,t) = 0 & \text{sobre } \partial\Omega\times\mathcal{O}, \\ u(x,a,0) = u_0(x,a) & \text{en } \Omega\times(0,A_\dagger), \\ u(x,0,t) = \displaystyle\int_0^{A_\dagger} \beta(x,a,t)u(x,a,t)\mathrm{d}a & \text{en } \Omega\times(0,T). \end{cases} \tag{2}$$

En primer lugar vamos a dar el concepto de solución que emplearemos a lo largo de la sección:

Definición 1 *Una función $u:\Omega\times\mathcal{O}\to\mathbb{R}$ es una solución del problema* (2) *si $u\in L^2(\mathcal{O};H_0^1(\Omega))$ y además verifica*

$$(\partial_t+\partial_a)u + \mu u \in L^2(\mathcal{O};H^{-1}(\Omega)),$$
$$f(\cdot,\cdot,\cdot,u)\in L^2(\mathcal{O};H^{-1}(\Omega))$$

y, para cualquier $w \in L^2(\mathcal{O}; H_0^1(\Omega))$*, se tiene que*

$$\begin{aligned}\iint_{\mathcal{O}} \langle(\partial_t + \partial_a)u + \mu u, w\rangle \mathrm{d}a\,\mathrm{d}t + \iiint_{\Omega\times\mathcal{O}} \nabla u \cdot \nabla w\,\mathrm{d}x\,\mathrm{d}a\,\mathrm{d}t \\ = \iint_{\mathcal{O}} \langle f(\cdot, a, t, u), w\rangle \mathrm{d}a\,\mathrm{d}t,\end{aligned} \tag{3}$$

donde $\langle\cdot\,,\cdot\rangle$ *denota el producto de dualidad entre* $H^{-1}(\Omega)$ *y* $H_0^1(\Omega)$*.*

Gracias a la regularidad pedida a una solución u, las condiciones iniciales tienen sentido en $L^2(\Omega\times(0,A_\dagger))$ y $L^2(\Omega\times(0,T))$, respectivamente y, por tanto, u tiene que verificar:

$$u(x,a,0) = u_0(x,a), \text{ en } L^2(\Omega\times(0,A_\dagger))$$

$$u(x,0,t) = \int_0^{A_\dagger} \beta(x,a,t)u(x,a,t)\mathrm{d}a, \text{ en } L^2(\Omega\times(0,T)).$$

Vamos a suponer que

($\mathcal{H}_\mu$) La tasa de mortalidad verifica

$$\mu \in \mathcal{C}^0(\overline{\Omega}\times[0,A_\dagger)\times[0,T]), \qquad \mu(x,a,t) \geq 0 \text{ en } \Omega\times\mathcal{O} \tag{4}$$

y su comportamiento en $a = A_\dagger$ viene dado por la "condición de divergencia"(véase por ejemplo [24]):

$$\begin{cases} 0 < t < A_\dagger, \quad x \in \Omega, \quad \lim\limits_{a\to A_\dagger} \displaystyle\int_0^t \mu(x, a-t+\tau, \tau)\mathrm{d}\tau = +\infty, \\ A_\dagger < t < T, \quad x \in \Omega, \quad \lim\limits_{a\to A_\dagger} \displaystyle\int_0^a \mu(x, \alpha, t-a+\alpha)\mathrm{d}\alpha = +\infty. \end{cases} \tag{5}$$

($\mathcal{H}_\beta$) La tasa de natalidad β, definida en $\Omega\times\mathcal{O}$, verifica

$$\beta \in L^\infty(\Omega\times\mathcal{O}), \qquad \beta(x,a,t) \geq 0 \ \text{ e.c.t } \ \Omega\times\mathcal{O}. \tag{6}$$

Pondremos

$$\bar{\beta} := \sup\{\beta(x,a,t) : (x,a,t) \in \Omega\times\mathcal{O}\} \tag{7}$$

($\mathcal{H}_0$) Asumiremos que la condición inicial, u_0, verifica

$$u_0 \in L^2(\Omega\times(0,A_\dagger)). \tag{8}$$

Nota 1 La condición (5) nos garantiza que, bajo ciertas hipótesis sobre f, la solución del problema (2) se anula en $a = A_\dagger$. Es decir, la población muere al alcanzar la edad $a = A_\dagger$ (véase por ejemplo [15, teorema 3]).

El siguiente resultado nos da la existencia y unicidad de solución de (2) bajo la hipótesis de lipschitzianidad global de f. La demostración está basada en la definición de una aplicación y la búsqueda de su único punto fijo (basada en el teorema del punto fijo de Banach), que será justamente la solución del problema.

Teorema 1 *Supongamos* $(\mathcal{H}_\mu)$, $(\mathcal{H}_\beta)$, $(\mathcal{H}_0)$ *y además*

$(\mathcal{H}_f)$ f *es lipschitziana con respecto a la cuarta variable, i.e. existe una constante* L *positiva tal que e.c.t.* $(x,a,t) \in \Omega \times \mathcal{O}$

$$|f(x,a,t,s_1) - f(x,a,t,s_2)| \leq L|s_1 - s_2| \ \textit{e.c.t}\ s_1, s_2 \in \mathbb{R}.$$

Además suponemos que

$$f(\cdot,\cdot,\cdot,0) \in L^2(\Omega \times \mathcal{O}). \tag{9}$$

Entonces existe una única solución, u*, del problema* (2)*.*

2.1 El método de sub-supersolución

En una segunda etapa hemos establecido un método de sub-supersolución para problemas del tipo (2). Dicho método nos permitirá debilitar la hipótesis de lipschitzianidad global de f a carácter Lipschitz en la zona comprendida entre la sub y la supersolución. También nos ayudará en el estudio del comportamiento asintótico de la solución.

Definición 2 *Se dice que una función* $\underline{u} \in L^2(\mathcal{O}; H^1(\Omega))$ *es una subsolución del problema* (2) *si*

$$\begin{aligned} &(\partial_t + \partial_a)\underline{u} + \mu\underline{u} \in L^2(\mathcal{O}; (H^1(\Omega))'), \\ &f(\cdot,\cdot,\cdot,\underline{u}) \in L^2(\Omega \times \mathcal{O}) \end{aligned}$$

y además verifica

a) Para toda $v \in L^2(\mathcal{O}; H_0^1(\Omega))$ *positiva*

$$\begin{aligned} \iint_{\mathcal{O}} \langle (\partial_t + \partial_a)\,\underline{u} + \mu\underline{u}, v\rangle \,\mathrm{d}a\,\mathrm{d}t + \iiint_{\Omega\times\mathcal{O}} \nabla\underline{u} \cdot \nabla v \,\mathrm{d}x\,\mathrm{d}a\,\mathrm{d}t \\ \leq \iint_{\mathcal{O}} f(x,a,t,\underline{u})\, v \,\mathrm{d}a\,\mathrm{d}t, \end{aligned} \tag{10}$$

b) $\underline{u}(x,a,t) \leq 0$ *sobre* $\partial\Omega \times \mathcal{O}$*, en el sentido débil,*

c) $\underline{u}(x,0,t) \leq \displaystyle\int_0^{A_\dagger} \beta(x,a,t)\underline{u}(x,a,t)\,\mathrm{d}a$ *en* $\Omega \times (0,T)$*,*

d) $\underline{u}(x,a,0) \leq u_0(x,a)$ *en* $\Omega \times (0, A_\dagger)$*.*

Análogamente, se define una supersolución, $\overline{u}$*, invirtiendo las desigualdades anteriores.*

El siguiente resultado proporciona la existencia y unicidad de solución de (2) entre la sub y la supersolución:

Teorema 2 *Supongamos que existe un par de sub-supersoluciones de* (2), $\underline{u}$, $\overline{u}$, *y que* f *verifica*

$$|f(x,a,t,s_1) - f(x,a,t,s_2)| \leq L|s_1 - s_2| \ \textit{p.c.t.}\ s_1, s_2 \in [u_*, u^*]\,, \tag{11}$$

con

$$\begin{aligned} u_* &= \inf_{(x,a,t)\in\Omega\times\mathcal{O}} \{\underline{u}(x,a,t), \overline{u}(x,a,t)\}, \\ u^* &= \sup_{(x,a,t)\in\Omega\times\mathcal{O}} \{\underline{u}(x,a,t), \overline{u}(x,a,t)\}. \end{aligned} \tag{12}$$

Entonces

$$\underline{u} \leq \overline{u}.$$

Además, (2) *posee una única solución,* u, *tal que*

$$\underline{u} \leq u \leq \overline{u}. \tag{13}$$

La prueba de este resultado está basada en un principio del máximo y en la aplicación de un método de punto fijo; para una demostración detallada, ver [12, teoremas 3.2 y 3.4].

2.2 Aplicación a algunos modelos ecológicos

Vamos a aplicar el método de sub-supersolución descrito anteriormente a algunos modelos ecológicos. En concreto, consideraremos el modelo

$$\begin{cases} \partial_t u + \partial_a u - \Delta u + \mu(a)u = \lambda u + F(u) & \text{en } \Omega\times\mathcal{O}, \\ u(x,a,t) = 0 & \text{sobre } \partial\Omega\times\mathcal{O}, \\ u(x,a,0) = u_0(x,a) & \text{en } \Omega\times(0,A_\dagger), \\ u(x,0,t) = \displaystyle\int_0^{A_\dagger} \beta(a)u(x,a,t)\,\mathrm{d}a & \text{en } \Omega\times(0,T), \end{cases} \tag{14}$$

con las particularizaciones siguientes:

- Un modelo **logístico generalizado**, es decir

$$F(s) \equiv -g(s),$$

donde g satisface la siguiente condición

($\mathcal{H}_g$) g es localmente lipschitziana, $g(0) = 0$, $g(s) \geq 0$ para todo $s \in \mathbb{R}_+$,

$$\lim_{s\to 0} \frac{g(s)}{s} = 0, \tag{15}$$

$$\lim_{s\to+\infty} \frac{g(s)}{s} = +\infty. \tag{16}$$

El caso típico es $g(s) \equiv s^p$, con $p > 1$.

- Un modelo de **tipo Holling-Tanner**, es decir,

$$F(s) \equiv \frac{s}{1+s}.$$

Obsérvese que en este caso podemos escribir que

$$\lambda s + F(s) \equiv (\lambda + 1)s - \frac{s^2}{1+s}$$

y es trivial comprobar que este caso no está incluido en el precedente, pues no se verifica la condición (16).

En primer lugar, analicemos el problema lineal asociado a (14), es decir el caso en que $F \equiv 0$. El problema es el que sigue:

$$\begin{cases} \partial_t \omega + \partial_a \omega - \Delta\omega + \mu(a)\omega = \lambda\omega & \text{en } \Omega \times \mathcal{O}, \\ \omega(x,a,t) = 0 & \text{sobre } \partial\Omega \times \mathcal{O}, \\ \omega(x,a,0) = u_0(x,a) & \text{en } \Omega \times (0, A_\dagger), \\ \omega(x,0,t) = \displaystyle\int_0^{A_\dagger} \beta(a)\omega(x,a,t)\,\mathrm{d}a & \text{en } \Omega \times (0,T), \end{cases} \tag{17}$$

donde β satisface la condición $(\mathcal{H}_\beta)$, $\lambda \in \mathbb{R}$ y u_0 y μ verifican

$(\mathcal{H}_0^*)$ $u_0 \in L^\infty(\Omega \times (0, A_\dagger))$ y $u_0(x,a) \geq 0$ e.c.t. $(x,a) \in \Omega \times (0, A_\dagger)$,

$(\mathcal{H}_\mu^*)$ μ es una función tal que $\mu \in L^\infty(0,r)$ para $r < A_\dagger$ y

$$\int_0^{A_\dagger} \mu(a)\,\mathrm{d}a = +\infty. \tag{18}$$

Observemos que (18) es equivalente a (5) cuando $\mu \equiv \mu(a)$. Este problema fue estudiado en 1988 por Langlais [28]. Siguiendo la notación de dicho trabajo, pondremos

$$\pi(a) = \exp\left(-\int_0^a \mu(\sigma)\,\mathrm{d}\sigma\right),$$

y denotaremos r_μ la única solución real de la ecuación

$$\int_0^{A_\dagger} \beta(a)\pi(a)e^{-ra}\,\mathrm{d}a = 1. \tag{19}$$

Finalmente, λ_1 denotará el primer autovalor del problema de Dirichlet homogéneo para el operador $-\Delta$ en Ω.

Los resultados sobre el comportamiento asintótico de (17) son los siguientes (para una demostración ver [28, secciones 3 y 4]):

Teorema 3 *Supongamos que se verifican las hipótesis* $(\mathcal{H}_\beta)$, $(\mathcal{H}_0^*)$ *y* $(\mathcal{H}_\mu^*)$ *y que existe* $0 \leq A_0 \leq A_\dagger$ *tal que* $\operatorname{sop}(\beta) \subset [0, A_0]$. *Entonces existe una única solución,* ω_λ, *de* (17). *Además, fijado* $T > 0$, ω_λ *está acotada en* $\Omega \times \mathcal{O}$. *Además,*

1. *Si* $\operatorname{sop}(u_0) \subset \bar{\Omega} \times (A_0, A_\dagger)$, *la solución de* (17) *satisface* $\omega_\lambda(x, a, t) = 0$ *para* $t > a$ *y* $x \in \Omega$. *Además, para cada* $A > 0$

$$\omega_\lambda(\cdot, \cdot, t) \underset{t \to +\infty}{\longrightarrow} 0 \text{ uniformemente en } \bar{\Omega} \times [0, A].$$

2. *Supongamos que la condición inicial* u_0 *verifica*

$$\operatorname{sop}(u_0) \cap (\Omega \times (0, A_0)) \neq \emptyset.$$

Entonces

i) Si $\lambda < \lambda_1 - r_\mu$, *la solución de* (17) *verifica*

$$\omega_\lambda(\cdot, \cdot, t) \underset{t \to +\infty}{\longrightarrow} 0 \text{ uniformemente en } \bar{\Omega} \times [0, A_\dagger].$$

ii) Si $\lambda = \lambda_1 - r_\mu$, *la solución de* (17) *verifica*

$$\omega_\lambda(x, a, t) \underset{t \to +\infty}{\longrightarrow} g(x, a) \text{ en } L^2(\Omega \times (0, A_\dagger))$$

para una cierta función $g \in L^2_+(\Omega \times (0, A_\dagger))$.

iii) Si $\lambda > \lambda_1 - r_\mu$, *la solución de* (17) *verifica*

$$\omega_\lambda(x, a, t) \underset{t \to +\infty}{\longrightarrow} +\infty \text{ en } L^2(\Omega \times (0, A_\dagger)).$$

Para una descripción de la función g que aparece en este resultado, véase [28, Teorema 4.9].

2.2.1 Un problema logístico generalizado

Primero aplicaremos los resultados del teorema 2 al problema logístico generalizado

$$\begin{cases} \partial_t u + \partial_a u - \Delta u + \mu(a)u = \lambda u - g(u) & \text{en } \Omega \times \mathcal{O}, \\ u(x, a, t) = 0 & \text{sobre } \partial\Omega \times \mathcal{O}, \\ u(x, a, 0) = u_0(x, a) & \text{en } \Omega \times (0, A_\dagger), \\ u(x, 0, t) = \displaystyle\int_0^{A_\dagger} \beta(a) u(x, a, t)\, \mathrm{d}a & \text{en } \Omega \times (0, T), \end{cases} \tag{20}$$

donde $\lambda \in \mathbb{R}$.

Teorema 4 *Supongamos que se verifican las hipótesis* $(\mathcal{H}_\beta)$, $(\mathcal{H}_0^*)$, $(\mathcal{H}_g)$ *y* $(\mathcal{H}_\mu^*)$. *Entonces existe una única solución positiva,* u, *del problema* (20). *Supongamos que existe* $0 \leq A_0 \leq A_\dagger$ *tal que* $\operatorname{sop}(\beta) \subset [0, A_0]$. *Entonces*

1. *Si* $\operatorname{sop}(u_0) \subset \bar{\Omega} \times (A_0, A_\dagger)$*, la solución de* (20) *satisface* $u(x,a,t) = 0$ *para* $t > a$ *y* $x \in \Omega$*. Además, para cada* $A > 0$

$$u(\cdot,\cdot,t) \underset{t\to+\infty}{\longrightarrow} 0 \text{ uniformemente en } \bar{\Omega} \times [0, A].$$

2. *Supongamos que la condición inicial,* u_0*, verifica*

$$\operatorname{sop}(u_0) \cap (\Omega \times (0, A_0)) \neq \emptyset.$$

Entonces

i) Si $\lambda < \lambda_1 - r_\mu$*, la solución de* (20) *satisface*

$$u(\cdot,\cdot,t) \underset{t\to+\infty}{\longrightarrow} 0 \text{ uniformemente en } \bar{\Omega} \times [0, A_\dagger].$$

ii) Si $\lambda > \lambda_1 - r_\mu$ *y* $u_0(x,a) > 0$ *en* $\Omega \times (0, A_\dagger)$*, entonces el modelo* (20) *es permanente en el sentido siguiente: existe una subsolución* $\underline{u}$ *y una supersolución* $\overline{u}$ *de* (20) *tales que p.c.t.* $(x,a,t) \in \Omega \times (0, A_\dagger) \times (0, +\infty)$*:*

$$\underline{u}(x,a) \leq u(x,a,t) \leq \overline{u}(x,a). \tag{21}$$

Demostración. Vamos a dar a continuación una idea de la demostración.

Claramente, se tiene que 0 y la solución ω_λ de (17) son un par de sub-supersoluciones de (20). Luego una aplicación directa del teorema 2 nos da la existencia y unicidad de solución positiva. Además, aplicando el teorema 3, obtenemos *1* y *2 i)*.

Para la demostración de (21), consideramos como subsolución la función:

$$\underline{u}(x,a) := \varepsilon \exp\left(-r_\mu a - \int_0^a \mu(s)\,\mathrm{d}s\right) \varphi_1(x)$$

con $\varepsilon > 0$ suficientemente pequeño y φ_1 una autofunción positiva asociada al autovalor λ_1. Y como supersolución consideramos la función:

$$\overline{u}(x,a) := K \exp\left(-r_m a - \int_0^a m(s)\,\mathrm{d}s\right) \widetilde{\varphi_1}(x),$$

donde $\widetilde{\varphi_1}$ es la autofunción positiva asociada a $\widetilde{\lambda_1}$, el primer autovalor del problema de Dirichlet homogéneo para $-\Delta$ en el dominio $\widetilde{\Omega}$, siendo $\Omega \subset\subset \widetilde{\Omega}$ y $m \in \mathcal{C}^0([0, A_\dagger])$ tal que $m(a) \leq \mu(a)$ para todo $a \in [0, A_\dagger]$. Entonces, aplicando (13) se obtiene (21). □

2.2.2 Un modelo de tipo Holling-Tanner

Ahora, vamos a aplicar el teorema 2 a un modelo de Holling-Tanner, es decir al problema

$$\begin{cases} \partial_t u + \partial_a u - \Delta u + \mu(a)u = \lambda u + \dfrac{u}{1+u} & \text{en } \Omega \times \mathcal{O}, \\ u(x,a,t) = 0 & \text{sobre } \partial\Omega \times \mathcal{O}, \\ u(x,a,0) = u_0(x,a) & \text{en } \Omega \times (0, A_\dagger), \\ u(x,0,t) = \displaystyle\int_0^{A_\dagger} \beta(a)u(x,a,t)\,\mathrm{d}a & \text{en } \Omega \times (0,T), \end{cases} \tag{22}$$

donde $\lambda \in \mathbb{R}$.

Teorema 5 *Supongamos que se verifican las hipótesis* $(\mathcal{H}_\beta)$, $(\mathcal{H}_0^*)$, $(\mathcal{H}_\mu^*)$ *y que* $\mu \in \mathcal{C}^0([0, A_\dagger))$ *es una función creciente. Entonces existe una única solución positiva* u *de* (22).

Supongamos además que existe $0 \leq A_0 \leq A_\dagger$ *tal que* $\mathrm{sop}(\beta) \subset [0, A_0]$. *Entonces:*

1. *Si* $\mathrm{sop}(u_0) \subset \bar{\Omega} \times (A_0, A_\dagger)$, *la solución de* (22) *verifica* $u(x,a,t) = 0$ *para* $t > a$ *y* $x \in \Omega$. *Además, para cada* $A > 0$,

$$u(\cdot,\cdot,t) \to 0 \text{ uniformemente } \bar{\Omega} \times [0, A] \text{ cuando } t \to +\infty.$$

2. *Supongamos que la condición inicial* u_0 *verifica*

$$\mathrm{sop}(u_0) \cap (\Omega \times (0, A_0)) \neq \emptyset.$$

Entonces

i) Si $\lambda < \lambda_1 - r_\mu - 1$, *la solución de* (22) *verifica*

$$u(\cdot,\cdot,t) \to 0 \text{ uniformemente en } \bar{\Omega} \times [0, A_\dagger] \text{ cuando } t \to +\infty.$$

ii) Si $\lambda > \lambda_1 - r_\mu$, *la solución de* (22) *verifica*

$$u(x,a,t) \to +\infty \text{ en } L^2(\Omega \times (0, A_\dagger)) \text{ cuando } t \to +\infty.$$

iii) Si $\lambda \in (\lambda_1 - r_\mu - 1, \lambda_1 - r_\mu)$ *y* $u_0(x,a) > 0$ *en* $\Omega \times (0, A_\dagger)$, *entonces* (22) *es un sistema permanente, en el sentido que se explica en el teorema* 4.

Demostración. Daremos a continuación una idea de la prueba.

Primero, para ver la existencia y unicidad de solución positiva del problema, tomamos como subsolución la función idénticamente nula y como supersolución $\omega_{\lambda+1}$, solución del problema (17) donde en el término de la derecha en vez de λ tenemos $\lambda + 1$. Con este par de sub-supersoluciones y una aplicación directa de los teoremas 2 y 3, se tienen *1* y *2 i)*.

Por otra parte, tenemos que la solución u de (22) es una supersolución de (17) y por consiguiente $\omega_\lambda \leq u$; y de aquí se sigue *2 ii)*.

Para la demostración de *2 iii)*, tomaremos como subsolución la misma que en el teorema 4, es decir, la función

$$\underline{u}(x,a) := \varepsilon \exp\left(-r_\mu a - \int_0^a \mu(s)\,\mathrm{d}s\right)\varphi_1(x).$$

Y tomaremos como supersolución

$$\overline{u}(x,a) := K \exp\left(-r_{\mu_n} a - \int_0^a \mu_n(s)\,\mathrm{d}s\right)\overline{\varphi}_1(x)$$

con $K > 0$ suficientemente grande, donde

$$\mu_n(s) := \begin{cases} \mu(0) & 0 \leq s \leq 1/n, \\ \mu(s-1/n) & 1/n < s < A_\dagger + 1/n, \end{cases}$$

$\overline{\varphi}_1$ es la autofunción positiva asociada a $\overline{\lambda}_1$, $\overline{\lambda}_1$ es el primer autovalor del problema de Dirichlet homogéneo para $-\Delta$ en el dominio Ω_1 y Ω_1 está elegido de forma que $\Omega \subset\subset \Omega_1$. Y aplicando (13) se obtiene directamente *2 iii)*.

□

3 Estudio del problema estacionario asociado al modelo con difusión

Vamos a estudiar el problema estacionario asociado al problema (2). Por tanto, supondremos que la tasa de mortalidad μ es independiente del tiempo. Así, conservando la notación de la sección anterior, estudiamos el problema no lineal siguiente:

$$\begin{cases} \partial_a u - \Delta u + \mu(x,a)u = f(x,a,u) & \text{en } \Omega \times (0, A_\dagger), \\ u(x,a) = 0 & \text{sobre } \partial\Omega \times (0, A_\dagger), \\ u(x,0) = \displaystyle\int_0^{A_\dagger} \beta(x,a)u(x,a)\mathrm{d}a & \text{en } \Omega. \end{cases} \tag{23}$$

Una de las principales dificultades que nos encontramos en el estudio del problema (23) es que vamos a suponer que μ explota en edad finita, para poder garantizar que la población se extingue al llegar a la edad máxima $A_\dagger$ y además la condición inicial es no local. Es por ello que no podemos aplicar el método clásico de sub-supersoluciones para problemas parabólicos (véase por ejemplo [13] y [30]). Por esta razón, en esta sección, análogamente a como se ha hecho en la sección precedente, vamos a justificar un método de sub-supersolución.

El concepto de solución es ahora el siguiente:

Definición 3 *Diremos que una función u es una solución del problema* (23) *si* $u \in L^2(0, A_\dagger; H^1(\Omega))$, *se tiene que*

$$\partial_a u + \mu u \in L^2(0, A_\dagger; H^{-1}(\Omega)),$$
$$f(\cdot, \cdot, u) \in L^2(0, A_\dagger; H^{-1}(\Omega)),$$

para cualquier $v \in L^2(0, A_\dagger; H_0^1(\Omega))$ *se tiene que*

$$\int_0^{A_\dagger} \langle \partial_a u + \mu u, v \rangle \, \mathrm{d}a + \iint_{\Omega \times (0, A_\dagger)} \nabla u \cdot \nabla v \, \mathrm{d}x \, \mathrm{d}a$$
$$= \iint_{\Omega \times (0, A_\dagger)} f(x, a, u) v \, \mathrm{d}x \, \mathrm{d}a,$$

$$u(x, a) = 0 \text{ sobre } \partial\Omega \times (0, A_\dagger)$$

y

$$u(x, 0) = \int_0^{A_\dagger} \beta(x, a) u(x, a) \, \mathrm{d}a, \text{ en } \Omega.$$

Supondremos que se cumplen las siguientes hipótesis:

$(\widetilde{\mathcal{H}_\mu})$ μ es una función tal que $\mu \in L^\infty(\overline{\Omega} \times (0, r))$ para $r < A_\dagger$,

$$\int_0^r \mu_M(a) \, \mathrm{d}a < \infty \quad \text{y} \quad \int_0^{A_\dagger} \mu_L(a) \, \mathrm{d}a = +\infty, \tag{24}$$

donde

$$\mu_L(a) := \inf_{x \in \overline{\Omega}} \mu(x, a) \quad \text{y} \quad \mu_M(a) := \sup_{x \in \overline{\Omega}} \mu(x, a).$$

$(\widetilde{\mathcal{H}_\beta})$ $\beta \in L^\infty(\Omega \times (0, A_\dagger))$ es positiva y no trivial.

3.1 El método de sub-supersolución

Vamos a presentar a continuación un método de sub-supersolución que conducirá a la existencia de una solución minimal y una solución maximal entre la subsolución y la supersolución. Primero, digamos qué entendemos por un par de sub-supersoluciones:

Definición 4 *Diremos que una función* $\overline{u} \in L^2(0, A_\dagger; H^1(\Omega))$ *es una supersolución de* (23) *si*

$$\partial_a \overline{u} + \mu \overline{u} \in L^2(0, A_\dagger; (H^1(\Omega))'),$$

$$f(\cdot, \cdot, \overline{u}) \in L^2(\Omega \times (0, A_\dagger)),$$

para toda $v \in L^2(0, A_\dagger; H_0^1(\Omega))$ *positiva se tiene que*

$$\int_0^{A_\dagger} \langle \partial_a \overline{u} + \mu \overline{u}, v \rangle \, \mathrm{d}a + \iint_{\Omega \times (0, A_\dagger)} \nabla \overline{u} \cdot \nabla v \, \mathrm{d}x \, \mathrm{d}a \geq \iint_{\Omega \times (0, A_\dagger)} f(x, a, \overline{u}) v \, \mathrm{d}x \, \mathrm{d}a,$$

$$\overline{u}(x, a) \geq 0 \text{ sobre } \partial\Omega \times (0, A_\dagger)$$

y

$$\overline{u}(x, 0) \geq \int_0^{A_\dagger} \beta(x, a) \overline{u}(x, a) \, \mathrm{d}a \text{ en } \Omega.$$

Análogamente se define el concepto de subsolución (cambiando las desigualdades anteriores por sus contrarias).

Recordaremos ahora un resultado de existencia de solución bajo la hipótesis de existencia de un par de sub-supersoluciones. Dicho resultado está basado en la construcción de un par de sucesiones que arrancan respectivamente en la subsolución y supersolución.

Teorema 6 *Supongamos que se verifican las condiciones* $(\widetilde{\mathcal{H}_\mu})$ *y* $(\widetilde{\mathcal{H}_\beta})$ *y que además* f *verifica*

$$|f(x, a, s_1) - f(x, a, s_2)| \leq L|s_1 - s_2|, \text{ e.c.t. } (x, a) \in \Omega \times (0, A_\dagger), s_1, s_2 \in \mathbb{R}. \quad (25)$$

Si existe un par de sub-supersoluciones de (23) *tales que* $\underline{u} \leq \overline{u}$*, entonces existe una solución minimal* u_* *y una solución maximal* u^* *de* (23) *en el sentido siguiente: para cualquier otra solución*

$$u \in [\underline{u}, \overline{u}] := \{u \in L^2(\Omega \times (0, A_\dagger)) : \underline{u} \leq u \leq \overline{u}\},$$

se verifica que

$$\underline{u} \leq u_* \leq u \leq u^* \leq \overline{u}.$$

3.2 El problema de autovalores

Análogamente a como hemos hecho en la sección anterior, queremos aplicar el teorema 6 a un problema logístico generalizado. Como hicimos en el caso evolutivo, la primera etapa consistirá en estudiar el problema lineal asociado, es decir,

$$\begin{cases} \partial_a u - \Delta u + \mu(x, a) u = \lambda u & \text{en } \Omega \times (0, A_\dagger), \\ u(x, a) = 0 & \text{sobre } \partial\Omega \times (0, A_\dagger), \\ u(x, 0) = \displaystyle\int_0^{A_\dagger} \beta(x, a) u(x, a) \, \mathrm{d}a & \text{en } \Omega. \end{cases} \quad (26)$$

La situación no es la misma que la que corresponde al caso parabólico clásico, donde la condición inicial tiene carácter local ($u(x, 0) = u_0(x) > 0$).

Para el parabólico clásico sabemos que el problema (26) posee solución positiva para todo $\lambda \in \mathbb{R}$. Por el contrario, nos encontramos aquí ante un problema de autovalores.

Definición 5 *Diremos que λ es un autovalor de* (26) *si existe una solución u de* (26)*. Diremos que λ es un autovalor principal si existe una solución u que verifica $u > 0$ en $\Omega \times (0, A_\dagger)$.*

En el siguiente resultado, que reposa sobre el teorema de Krein-Rutman, queda asegurada la existencia de un único autovalor principal:

Teorema 7 *Supongamos que se verifican las hipótesis $(\widetilde{\mathcal{H}_\mu})$ y $(\widetilde{\mathcal{H}_\beta})$. Entonces existe un único autovalor principal de* (26)*, que será denotado por $\lambda_0(\mu)$, que es simple y es el único que tiene asociada una autofunción positiva. Además, las autofunciones positivas pueden elegirse acotadas. Finalmente, la aplicación $\mu \mapsto \lambda_0(\mu)$ es creciente.*

3.3 Aplicación a un problema logístico generalizado

Ahora vamos a aplicar los teoremas 6 y 7 al problema logístico generalizado siguiente:

$$\begin{cases} \partial_a u - \Delta u + \mu(x,a)u = \lambda u - g(u) & \text{en } \Omega \times (0, A_\dagger), \\ u(x,a) = 0 & \text{sobre } \partial\Omega \times (0, A_\dagger), \\ u(x,0) = \displaystyle\int_0^{A_\dagger} \beta(x,a)u(x,a)\,\mathrm{d}a & \text{en } \Omega, \end{cases} \tag{27}$$

con $\lambda \in \mathbb{R}$ y g verificando las hipótesis $(\mathcal{H}_g)$, (15) y (16). Supondremos además, para obtener unicidad, que $g(s)/s$ es una función estrictamente creciente.

El siguiente teorema es un resultado de existencia de solución para ciertos valores de λ donde, de nuevo, observamos un destacable cambio con respecto al problema parabólico clásico.

Teorema 8 *El problema* (27) *tiene una solución positiva si y sólo si $\lambda > \lambda_0(\mu)$. Además, caso de que exista la solución, ésta es única.*

Demostración. De nuevo nos limitaremos a dar una idea de la demostración.

- Supongamos en primer lugar que existe $u > 0$ solución de (27). Entonces es fácil comprobar que

$$\lambda = \lambda_0(\mu + g(u)/u) > \lambda_0(\mu). \tag{28}$$

- Supongamos ahora que $\lambda > \lambda_0(\mu)$. Veamos entonces la existencia de solución de (27). Para ello, tomamos como subsolución la función

$$\underline{u} := \varepsilon\varphi(x,a)$$

con $\varepsilon > 0$ suficientemente pequeño y φ una autofunción positiva asociada a $\lambda_0(\mu)$. Y como supersolución la función

$$\overline{u} := K\widetilde{\varphi}(x, a),$$

donde $K > 0$ suficientemente grande y $\widetilde{\varphi}(x, a)$ es la autofunción positiva asociada a $\widetilde{\lambda_0}(\mu)$, el autovalor principal del problema (26) correspondiente a un abierto $\widetilde{\Omega}$ tal que $\Omega \subset\subset \widetilde{\Omega}$.

Para demostrar la unicidad, supongamos que existen dos soluciones distintas u_1, u_2 de (26). Entonces se comprueba que en este caso $\lambda > \lambda_0(\mu + g(u_1)/u_1)$, lo cual está en contradicción con que u_1 sea solución de (26) por (28).

□

Referencias

[1] B. AINSEBA, Exact and approximate controllability of the age and space population dynamics structured model. *J. Math. Anal. Appl.*, 275:562–574, 2002.

[2] B. AINSEBA, M. LANGLAIS, On a population dynamics control problem with age dependence and spatial structure. *J. Math. Anal. Appl.*, 248:455–474, 2000.

[3] S. ANIŢA, *Analysis and control of age-dependent population dynamics*, volume 11 of *Mathematical Modelling: Theory and Applications*. Kluwer Academic Publishers, Dordrecht, 2000.

[4] O. ARINO, M. DELGADO, M. MOLINA-BECERRA, Asymptotic behavior of disease-free equilibriums of an age-structured predator-prey model with disease in the prey. *Discrete Contin. Dyn. Syst. Ser. B*, 4:501–515, 2004.

[5] S. BUSENBERG, M. IANNELLI, Separable models in age-dependent population dynamics. *J. Math. Biol.*, 22:145–173, 1985.

[6] S. BUSENBERG, M. IANNELLI, H. R. THIEME, Global behavior of an age-structured epidemic model. *SIAM J. Math. Anal.*, 22:1065–1080, 1991.

[7] W. L. CHAN, B. Z. GUO, Global behaviour of age-dependent logistic population models. *J. Math. Biol.*, 28:225–235, 1990.

[8] J. CHATTOPADHYAY, O. ARINO, A predator-prey model with disease in the prey. *Nonlinear Anal.*, 36:747–766, 1999.

[9] M. CHIPOT, On the equations of age-dependent population dynamics. *Arch. Rat. Mech. Anal.*, 82:13–25, 1983.

[10] M. Delgado, M. Molina-Becerra, A. Suárez, A nonlinear age-dependent model with spatial diffusion. Sometido a publicación.

[11] M. Delgado, M. Molina-Becerra, A. Suárez, Relating disease and predation: equilibria of an epidemic model. *Math. Methods Appl. Sci.*, 28:349–362, 2005.

[12] M. Delgado, M. Molina-Becerra, A. Suárez, The sub-supersolution method for an evolutionary reaction-diffusion age-dependent problem. *Differential Integral Equations*, 18:155–168, 2005.

[13] J. Deuel, P. Hess, Nonlinear parabolic boundary value problems with upper and lower solutions. *Israel J. Math.*, 29:92–104, 1978.

[14] G. Di Blasio, L. Lamberti, An initial-boundary value problem for age-dependent population diffusion. *SIAM J. Appl. Math.*, 35:593–615, 1978.

[15] M. G. Garroni, M. Langlais, Age-dependent population diffusion with external constraint. *J. Math. Biol.*, 14:77–94, 1982.

[16] M. E. Gurtin, A system of equations for age dependent population diffusion. *J. Theor. Biol.*, 40:389–392, 1973.

[17] M. E. Gurtin, D. S. Levine, On predator-prey interactions with predation dependent on age of prey. *Math. Biosc.*, 47:207–219, 1979.

[18] M. E. Gurtin, R. C. MacCamy, Non-linear age-dependent population dynamics. *Arch. Ration. Mech. Anal.*, 54:281–300, 1974.

[19] M. E. Gurtin, R. C. MacCamy, On the diffusion of biological populations. *Math. Biosci.*, 33:35–49, 1977.

[20] M. Iannelli, M. Y. Kim, E. J. Park, Asymptotic behavior for an SIS epidemic model and its approximation. *Nonlinear Anal.*, 35:797–814, 1999.

[21] M. Kubo, M. Langlais, Periodic solutions for a population dynamics problem with age-dependence and spatial structure. *J. Math. Biol.*, 29:363–378, 1991.

[22] M. Kubo, M. Langlais, Periodic solutions for nonlinear population dynamics models with age-dependence and spatial structure. *J. Differential Equations*, 109:274–294, 1994.

[23] K. Kunisch, W. Schappacher, G. F. Webb, Nonlinear age-dependent population dynamics with random diffusion. *Comput. Math. Appl.*, 11:155–173, 1985.

[24] H. L. Langhaar, General population theory in the age-time continuum. *J. Franklin Inst.*, 293:199–214, 1972.

[25] M. LANGLAIS, Solutions fortes pour une classe de problèmes aux limites du second ordre dégénérés. *Comm. Partial Differential Equations*, 4:869–897, 1979.

[26] M. LANGLAIS, On a linear age-dependent population diffusion model. *Quart. Appl. Math.*, 40:447–460, 1982/83.

[27] M. LANGLAIS, A nonlinear problem in age-dependent population diffusion. *SIAM J. Math. Anal.*, 16:510–529, 1985.

[28] M. LANGLAIS, Large time behavior in a nonlinear age-dependent population dynamics problem with spatial diffusion. *J. Math. Biol.*, 26:319–346, 1988.

[29] A. G. MCKENDRICK, Applications of mathematics to medical problems. *Proc. Edin. Math. Soc.*, 98–130, 1926.

[30] C. V. PAO, *Nonlinear parabolic and elliptic equations*. Plenum Press, New York, 1992.

[31] B. PERTHAME, Quelques équations de transport apparaissant en biologie. *Boletín de la Sociedad Española de Matemática Aplicada*, 28:71–98, 2004.

[32] J. PRÜSS, Stability analysis for equilibria in age-specific population dynamics. *Nonlinear Anal.*, 7:1291–1313, 1983.

[33] F. R. SHARPE, A. J. LOTKA, A problem in age-distribution. *Philosophical Magazine*, 435–438, 1911.

[34] E. VENTURINO, Age-structured predator-prey models. *Math. Modelling*, 5:117–128, 1984.

[35] G. F. WEBB, *Theory of nonlinear Age-dependent Population Dynamics*. Pure Appl. Math. Monographs, Marcel Dekker, New York, 1985.

Bol. Soc. Esp. Mat. Apl. nº31(2005), 73–108

Harten's framework for multiresolution with applications: from conservation laws to image compression

F. Aràndiga[1], G. Chiavassa[2],
R. Donat[1]

[1] Dept. Matemática Aplicada, Universitat de Valencia, Spain
[2] EGIM and LATP, Marseille. France

arandiga@uv.es, guillaume.chiavassa@esm2.imt-mrs.fr,
donat@uv.es

Abstract

We briefly review Harten's framework for multiresolution decompositions and describe two situations in which two different instances of the general framework have been used with success. In the numerical simulation of Hyperbolic Conservation Laws, the simple point-value setting with a linear centered interpolatory reconstruction is used to design a multilevel algorithm that effectively helps to reduce the computational expense associated to state of the art high resolution shock capturing schemes. The possibility of using nonlinear, data dependent, reconstruction techniques is explored in image compression. For piecewise smooth images, ENO reconstruction techniques outperform classical algorithms based on biorthogonal wavelets.

Palabras clave: *Conservation laws, multiresolution, shock capturing schemes, image compression.*

Clasificación por materias AMS: *65M06, 65D05, 35L65, 94A08.*

1 Introduction

Fourier analysis provides a way to represent square integrable functions in terms of their sinusoidal components. Fourier decomposition techniques have become basic tools for a great variety of applications in many fields of science, however, there is a serious drawback that renders the Fourier decomposition of

Fecha de recepción: 07/07/03

an *irregular* function useless in many situations: an isolated singularity affects all coefficients in the decomposition and, as a result, an $O(1)$ oscillatory behavior can be observed in all functional approximations obtained by considering truncated Fourier series. This typical oscillatory behavior, known as Gibb's phenomenon, is a direct consequence of the fact that the basis functions (sines and cosines in Fourier analysis) have non-compact support.

Wavelet basis fare much better in this respect. Orthonormal and biorthogonal wavelet basis can be used to analyze a square integrable function much in the same way as in Fourier analysis [17]. In this framework, however, each basis function has compact support; in fact, the support of each wavelet basis function is *localized* around a particular point in space, and it identifies oscillatory behavior in the input function at a particular scale. A wavelet representation of a given function f is also known, due to this particular feature, as a multiresolution (or multiscale) decomposition of the function.

Multiscale decompositions have an important role in numerical analysis. A wavelet type decomposition of a function can be used to reduce the cost of many numerical algorithms either by applying it to the numerical solution operator to obtain an approximate sparse form [10, 24, 4, 33] or by applying it to the numerical solution itself to obtain an approximate reduced representation in order to solve for fewer quantities [29, 9].

The building block of the wavelet theory is a square integrable function whose dilates and translates form an orthonormal basis of the space of square-integrable functions. Wavelet orthonormal basis are composed of dilates and translates of a single function, the *wavelet*. The wavelet is intimately linked to the *scaling function*, sometimes called *mother wavelet*. This function satisfies a dilation relation which is in fact responsible for the properties of the multiscale representation. In a way, the construction of particular orthonormal wavelet basis becomes equivalent to a search for solutions in $L^2(\mathbb{R})$ or very particular dilation relations [38, 18, 19].

The uniformity underlying the classical wavelet theory leads to conceptual difficulties in extending wavelets to bounded domains and general geometries [15]. Moreover, it is even harder to obtain adaptive (data-dependent) multiresolution representations that fit the approximation to the local nature of the data.

A combination of ideas from multigrid methods, numerical techniques for conservation laws, hierarchical basis in finite element spaces, subdivision schemes in Computer-Aided Design (CAD) and, of course, the theory of wavelets, led Ami Harten to the development of a "General Framework" for multiresolution representation of (discrete) data.

In Harten's framework, the dilation relation is no longer the main design tool. Instead, the connection between adjacent resolution levels isspecified by means of two operators: Decimation and Prediction. In turn, these two interscale operators rely on the two very basic building blocks of Harten's framework: The *Discretization* and *Reconstruction* operators. The first one obtains discrete information from a given signal (belonging to a particular function space). The last one produces a functional approximation (in the same function space) from

the discrete information contents of the original signal.

The relation between prediction and reconstruction gives Harten's framework a degree of flexibility lacking in standard wavelet theory. The reconstruction operator is allowed to be data-dependent (hence nonlinear). A nonlinear reconstruction operator leads to a nonlinear prediction scheme and to a nonlinear multiresolution transform. Even in the linear case (i.e. data-independent reconstruction operators) it is a conceptually simple matter to deal with bounded domains: It is enough to design the reconstruction operator in such a way that it solves an approximation problem that fits the boundary data.

Because of the essential role played by the discretization and reconstruction operators in Harten's framework, building multiresolution schemes that are appropriate for a given application becomes a task which is very familiar to a numerical analyst: The first step is to identify a sense of discretization which is appropriate for the given application; the second step consists in solving a problem in approximation theory.

2 The general framework

Multiscale decompositions aim at a 'rearrangement' of the information contents in a set of discrete data. To achieve such a 'rearrangement', Harten's general framework for multiresolution [30] relies on two operators that define the basic interscale relations: Decimation and Prediction. These operators act on finite dimensional linear vector spaces, V^j, that represent the different resolution levels (j increasing implies more resolution)

$$\text{(a) } D_j^{j-1} : V^j \to V^{j-1}, \qquad \text{(b) } P_{j-1}^j : V^{j-1} \to V^j, \tag{1}$$

and must satisfy two requirements of algebraic nature: D_j^{j-1} needs to be a *linear* operator and $D_j^{j-1} P_{j-1}^j = I_{V^{j-1}}$, i.e. the identity operator on the lower resolution level represented by V^{j-1}. The second condition is a *consistency* requirement: the prediction operator P_{j-1}^j does not change the discrete info at the lower resolution level $j-1$.

Using these two operators, a vector (i.e. a discrete sequence) $v^j \in V^j$ can be decomposed and reassembled as follows

$$\begin{array}{ccccc} v^j & \to & v^{j-1} & = & D_j^{j-1} v^j \\ & \searrow & e^j & = & v^j - P_{j-1}^j v^{j-1} \end{array} , \qquad v^j = P_{j-1}^j v^{j-1} + e^j$$

The vector $D_j^{j-1} v^j = v^{j-1}$ represents the discrete information contents of v^j at the lower resolution level $j-1$, while $e^j = (I_{V^j} - P_{j-1}^j D_j^{j-1}) v^j := Q_j v^j$ represents de *prediction error*, that is the error committed in trying to predict v^j from the low-resolution vector v^{j-1} via P_{j-1}^j .

Notice that the consistency requirement $D_j^{j-1} P_{j-1}^j = I_{V^{j-1}}$ implies that $D_j^{j-1} e^j = 0$, therefore its representation in terms of a basis of V^j is redundant.

Since D_j^{j-1} has full rank, $\dim\mathcal{N}(D_j^{j-1}) = \dim V^j - \dim V^{j-1}$, and redundancy can be eliminated by expressing e^j in terms of a basis of the null space $\mathcal{N}(D_j^{j-1})$. Let us introduce the operator G_j that computes the coordinates of e^j in a basis of $\mathcal{N}(D_j^{j-1})$ and call $d^j = G_j e^j$, then the sets v^j and $\{v^{j-1}, d^j\}$ have the same cardinality and are algebraically equivalent. The one to one correspondence between these two sets can be described using one more operator E_j, the canonical injection $\mathcal{N}(D_j^{j-1}) \hookrightarrow V^j$,

$$P_{j-1}^j v^{j-1} + E_j d^j = v^j \quad \longleftrightarrow \quad \{d^j, v^{j-1}\} \begin{cases} v^{j-1} &= D_j^{j-1} v^j \\ d^j &= G_j Q_j v^j. \end{cases} \tag{2}$$

This purely algebraic description can be recursively applied to 'rearrange' the information contents of a discrete sequence v^L, containing information on a very fine scale, as v^0, the information on a much coarser scale obtained by successive decimation of v^L, plus a sequence of (non-redundant) prediction errors at each resolution level

$$v^L \leftrightarrow Mv^L = (v^0, d^1, \ldots, d^L), \qquad \begin{array}{ccccccc} v^L & \to & v^{L-1} & \to & v^{L-2} & \to & \ldots \to v^0 \\ & \searrow & d^L & \searrow & d^{L-1} & \searrow & \ldots \searrow d^1 \end{array} \tag{3}$$

In electrical engineering terms, the basic encoding and decoding steps (2) are the analysis and synthesis steps of a sub-band filtering scheme with exact reconstruction [16]. The operator D_j^{j-1} plays the role of a low-pass filter and the operator $G_j(I_j - P_{j-1}^j D_j^{j-1})$ that of a band-pass filter. Usually, these filters are linear and of convolution type (this is exactly the situation within the wavelet framework).

In a multiresolution scheme *à la Harten*, D_j^{j-1} and P_{j-1}^j are constructed using two operators that relate discrete data to signals in a particular function space. These basic 'building blocks' are the *Discretization* and *Reconstruction* operators

$$\text{(a) } \mathcal{D}_j : \mathcal{F} \to V^j, \qquad \text{(b) } \mathcal{R}_j : V^j \to \mathcal{F}. \tag{4}$$

The discretization operators $\mathcal{D}_j$ yield discrete information at the resolution level specified by V^j, and the reconstruction operators $\mathcal{R}_j$ relate a set of discrete data in V^j to functions in $\mathcal{F}$.

These operators also have to satisfy some basic requirements of algebraic nature: $\mathcal{D}_j$ must be a linear operator and $\mathcal{D}_j(\mathcal{F}) = V^j$. In addition, $\mathcal{D}_j\mathcal{R}_j = I_{V^j}$, which is again a consistency relation.

It is also necessary that the sequence of discretization operators $\{\mathcal{D}_j\}$ be *nested*, that is

$$\mathcal{D}_j f = 0 \Rightarrow \mathcal{D}_{j-1} f = 0, \quad \forall f \in \mathcal{F}. \tag{5}$$

Given a nested sequence of linear discretization operators $\{\mathcal{D}_j\}$ (4)-(a), and a sequence of consistent reconstruction operators, $\{\mathcal{R}_j\}$ (4)-(b), the interscale operators are defined as follows:

$$D_j^{j-1} = \mathcal{D}_{j-1}\mathcal{R}_j, \qquad P_{j-1}^j = \mathcal{D}_j\mathcal{R}_{j-1}. \tag{6}$$

The dependence of D_j^{j-1} on $\mathcal{R}_j$ is 'fictitious'. For a nested sequence of discretization, D_j^{j-1} is completely specified by the sequence of discretization operators.

Proposición 1 *If $\{\mathcal{D}_j\}$ is a nested sequence of discretization and $\{\mathcal{R}_j\}$, $\{\tilde{\mathcal{R}}_j\}$ two sequences of consistent reconstruction operators, i.e. such that $\mathcal{D}_j\mathcal{R}_j = I_{V^j} = \mathcal{D}_j\tilde{\mathcal{R}}_j$, then*

$$\mathcal{D}_{j-1}\mathcal{R}_j = \mathcal{D}_{j-1}\tilde{\mathcal{R}}_j \tag{7}$$

Demostración. For any $v^j \in V^j$ we have $\mathcal{D}_j\mathcal{R}_j v^j = v^j = \mathcal{D}_j\tilde{\mathcal{R}}_j v^j$, thus

$$\mathcal{D}_j(\mathcal{R}_j v^j - \tilde{\mathcal{R}}_j v^j) = 0 \quad \Longrightarrow \quad \mathcal{D}_{j-1}(\mathcal{R}_j v^j - \tilde{\mathcal{R}}_j v^j) = 0.$$

□

Nestedness of the sequence $\{\mathcal{D}_j\}$ is an essential property of the discretization sequence, and has other important consequences:

Proposición 2 *If $\{\mathcal{D}_j\}$ is a nested sequence of discretization, and $\{\mathcal{R}_j\}$ is a consistent sequence of reconstruction operators then*

$$\mathcal{D}_l(\mathcal{R}_m\mathcal{D}_m) = \mathcal{D}_l \quad l \le m \tag{8}$$

Demostración. For any $f \in \mathcal{F}$, let $g = \mathcal{R}_m\mathcal{D}_m f$. Then

$$\mathcal{D}_m g = \mathcal{D}_m(\mathcal{R}_m\mathcal{D}_m)f = (\mathcal{D}_m\mathcal{R}_m)\mathcal{D}_m f = \mathcal{D}_m f \quad \rightarrow \quad \mathcal{D}_m(g - f) = 0$$

thus, the nested property implies $\mathcal{D}_l(g - f) = 0 \quad l \le m$, i.e.

$$\mathcal{D}_l f = \mathcal{D}_l(\mathcal{R}_m\mathcal{D}_m)f$$

□

3 Discretization by local averages

In Numerical Analysis, discretization processes are used as tools to go from a function in some functional space to a set of discrete values which give sensible information about the function. As an example, the simplest discretization process when dealing with continuous functions is that of point-wise evaluation, since the values of a continuous function at a given set of points give relevant information about the function. This information can be effectively used to represent the function or to reconstruct it approximately if necessary (via interpolation, for example).

On the other hand, point-wise evaluation does not provide an appropriate discretization setting for functions that are only piecewise continuous, since the values of the function at particular points might not be well defined; in addition, information about the exact location of discontinuities that fall between grid-points is completely lost. A discretization process often used when this space of

functions is relevant is the so-called cell-average discretization which considers the mean values of the function in each interval of the underlying grid.

Many discretization procedures in Numerical Analysis, and in particular the two mentioned above, can be described as the process of obtaining local averages with respect to a weight function, which is imposed by the underlying context. The weight function, $\omega(x)$, is a function with compact support and it is usually required that

$$\int \omega(x)dx = 1. \tag{9}$$

To be more specific, while keeping it simple at the same time, let us consider a grid X composed of a finite sequence of equally spaced points in $[0,1]$:

$$X = \{x_i\}, \qquad x_i \in [0,1] \qquad h = x_i - x_{i-1}.$$

For each function f in $\mathcal{F}$, a discretization operator, $\mathcal{D}$, associated to the resolution level defined by the grid X is defined as follows:

$$(\mathcal{D}f)_i \equiv \bar{f}_i := \frac{1}{h}\int f(x)\omega(\frac{x-x_i}{h})dx =: < f, \frac{1}{h}\omega(\frac{x-x_i}{h}) >, \quad x_i \in X. \tag{10}$$

The (linear) operator $\mathcal{D}$ acts naturally on a space of functions $\mathcal{F}$ for which the integral in (10) is well defined, and produces discrete values which give local information on the behavior of the function f at the resolution level specified by the grid X.

Weighted averages against the 'generalized' function $\omega(x) = \delta(x)$, *Dirac's* delta-function, correspond precisely to the process of discretizing a continuous function by considering its values at the points of the grid X. A rather 'natural' function space $\mathcal{F}$ for this discretization operator is $\mathcal{C}[0,1]$. On the other hand, the cell-average discretization procedure corresponds to the local-average discretization (10) with $\omega(x)$ being the box function:

$$\omega(x) = \begin{cases} 1 & -1 \le x < 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The natural function space in this context is $L^1[0,1]$.

These weight functions satisfy a two scale relation of the following type,

$$\omega(y) = 2\sum_l \alpha_l \omega(2y - l). \tag{12}$$

and this is a key feature to obtain a nested sequence of discretization-by- local-averages operators. The other key feature is a nested structure for the sequence of underlying grids that serves to define the sequence $\{\mathcal{D}_j\}$.

Let us consider the set of nested dyadic grids on $[0,1]$ (to which we associate the increasing levels of resolution) $\{X^j\}$, $j \ge 0$ of size $h_j = 2^{-j}h_0$:

$$X^j = \{x_i^j\} \qquad x_i^j = i \cdot h_j, \quad i = 0, \ldots, J_j \qquad J_j \cdot h_j = 1 \tag{13}$$

(notice that $x_{2i}^j = x_i^{j-1}$), and the sequence of discretization operators

$$\mathcal{D}_j : \mathcal{F} \to V^j, \quad (\mathcal{D}_j f)_i = \bar{f}_i^j =< f, \omega_i^j >, \qquad \omega_i^j = \frac{1}{h_j}\omega(\frac{x - x_i^j}{h_j}) \tag{14}$$

where V^j is a space of sequences of dimension N_j related to J_j (e.g. for $\omega(x) = \delta(x)$, $N_j = J_j + 1$, while for $\omega(x)$ in (11), $N_j = J_j$, see sections 3.1 and 3.2).

Taking $y = (x - x_i^{j-1})/h_{j-1}$, we can rewrite (12) as

$$\omega(\frac{x - x_i^{j-1}}{h_{j-1}}) = 2\sum_l \alpha_l \omega(2\frac{x - x_i^{j-1}}{h_{j-1}} - l) = 2\sum_l \alpha_l \omega(\frac{x - x_{2i}^j}{h_j} - l)$$

or, equivalently

$$\omega_i^{j-1} = \sum_l \alpha_l \omega_{2i+l}^j = \sum_l \alpha_{l-2i} \omega_l^j. \tag{15}$$

This relation can be expressed in terms of the discretization operators as

$$(\mathcal{D}_{j-1} f)_i = \sum_j \alpha_{j-2i} (\mathcal{D}_j f)_j. \tag{16}$$

which implies that the sequence $\{\mathcal{D}_j\}$ in (14) is nested. Therefore, discretizing by local averages with respect to a function that satisfies a dilation relation becomes a particular way of obtaining a nested sequence of discretization operators.

Nota 1 *Formulas (15) or (16) also imply that the decimation operator,* D_j^{j-1} *, can be described by a matrix whose elements are* $(D_j^{j-1})_{il} = \alpha_{l-2i}$*. Observe that* D_j^{j-1} *is independent of the level of resolution (except for its dimension).*

Once the weight function is fixed, the primary choice, that of the decimation operator, is already made. To construct an adequate multiresolution scheme, we still have two more *independent* choices to make:

1. A basis for the null space or, equivalently, an operative definition of the transfer operators G_j and E_j.

2. A prediction operator P_{j-1}^j , which is a right inverse of D_j^{j-1} . This amounts to choosing appropriate reconstruction operators at each resolution level.

For sequences $\{\mathcal{D}_j\}$ as in (14), the null space of D_j^{j-1} is easily characterized,

$$\mathcal{N}(D_j^{j-1}) = \{v^j \in V^j \mid D_j^{j-1} v^j = 0\} = \{v^j \in V^j \mid \sum_l \alpha_{l-2i} v_l^j = 0 \; \forall i\}. \tag{17}$$

Thus, the prediction errors always satisfy the following system of equations:

$$\sum_l \alpha_l e^j_{2i+l} = 0 \qquad i = 1, \ldots, N_{j-1}. \tag{18}$$

For the weight functions mentioned so far in this section it is possible to store the values e^j_i with odd indices, and use (18) in order to formulate a system of equations for the unknowns (e^j_{2i}), i.e.

$$d^j_i = e^j_{2i-1}, \quad \sum_l \alpha_{2l} e^j_{2i+2l} = -\sum_l \alpha_{2l-1} e^j_{2i+2l-1} = \sum_l \alpha_{2l-1} d_{i+l} \tag{19}$$

The definition of scale coefficients given in (19) leads to a simple definition for the operator G_j:

$$(G_j)_{lm} = \delta_{2l-1,m}. \tag{20}$$

The operator E_j is then obtained from (19) (notice that its columns provide a set of basis vectors for $\mathcal{N}(D_j^{j-1})$).

Nota 2 *The δ-function and the box function are the first two elements of a hierarchy of functions $\omega^m(x)$ obtained by repeated convolution with a characteristic function (see also [27]),*

$$\omega^{m+1}(x) = \omega^m(x) * \chi_{[-1+s_m, s_m]}, \quad s_m = \frac{1}{2}[1 - (-1)^m], \quad \omega^0 = \delta(x)$$

All of the functions in this chain satisfy a dilation relation of the type (12) and serve to specify an appropriate multiresolution setting. The multiresolution framework that correspond to the next member in the hierarchy $\omega^2(x)$ (the hat function) has been analyzed in [6, 7]

Nota 3 *The discretization operator specifies the process of generation of discrete data and, hence, it determines the* nature *of the discrete data to be analyzed. When reinterpreted within Harten's framework, the discretization operator in a wavelet decomposition is obtained by (14) with $\omega(x)$ being the scaling function. In this context the space $\mathcal{F}$ is $L^2(R)$ thus, in reward, we have all the geometric properties of a Hilbert space but in turn, many weight functions are automatically ruled out, in particular the δ-function which does not belong to $L^2(R)$. We shall see later that, nevertheless, the δ-function provides an appropriate setting for multiresolution .*

3.1 The point-value setting: interpolatory multiresolution

When we consider $w(x) = \delta(x)$, the discretization sequence in (14) is nothing but point-wise evaluation, on the points of the underlying grid that specifies the resolution level, i.e. $\mathcal{D}_j(f) = \{v^j_i\}$ where $v^j_i = f(x^j_i)$. Clearly, an appropriate functional space is $\mathcal{F} = \mathcal{C}[0,1]$ (keeping the simple setting of the previous section).

Any *consistent* sequence of reconstruction operators must satisfy

$$\mathcal{D}_j \mathcal{R}_j v^j = v^j \quad \longrightarrow \quad (\mathcal{R}_j v^j)(x_j^j) = v_j^j = f(x_j^j),$$

that is, the reconstruction operators must interpolate the data on which they act.

Since Dirac's δ function satisfies $\delta(x) = 2\delta(2x)$, $\alpha_0 = 2$ and the decimation operator is simply $(D_j^{j-1} v^j)_i = v_i^{j-1}$. Thus, following the guidelines in last section, the one to one correspondence in (2) can be written as

$$\left\{ \begin{array}{rcl} v_i^{j-1} & = & v_{2i}^j \\ d_i^j & = & v_{2i-1}^j - \mathcal{I}(x_{2i-1}^j, v^{j-1}) \end{array} \right\} \quad \longleftrightarrow \quad \left\{ \begin{array}{rcl} v_{2i}^j & = & v_i^{j-1} \\ v_{2i-1}^j & = & \mathcal{I}(x_{2i-1}^j, v^{j-1}) + d_i^j \end{array} \right\}$$

where the notation $\mathcal{I}(x, v^j)$ has been used instead of $\mathcal{R}_j v^j(x)$. Notice that the prediction errors are simply interpolation errors and that they only need to be computed at odd points on the fine grid X^j.

3.2 The cell-average setting

When we consider $w(x)$ to be the box function (11), the discretization sequence in (14) becomes

$$\mathcal{D}_j : L^1[0,1] \longrightarrow V^j, \qquad (\mathcal{D}_j f)_i = \frac{1}{h_j} \int_{x_{i-1}^j}^{x_i^j} f(x) dx,$$

where V^j is a space of sequences whose dimension is equal to the number of intervals on the jth grid and h_j is the uniform mesh spacing on $X^j = \{x_i^j\}_i$. Clearly, an appropriate functional space is $\mathcal{F} = L^1[0,1]$ (keeping the simple setting of the previous sections).

Since the box function satisfies the dilation relation $w(x) = w(2x) + w(2x+1)$, which implies $\alpha_0 = \alpha_{-1} = 1/2$ in the notation of section 3, we have $v_i^{j-1} = (D_j^{j-1} v^j)_i = (v_{2i}^j + v_{2i-1}^j)/2$.

The consistency requirement for the reconstruction operators, i.e. $\mathcal{D}_j \mathcal{R}_j = I_{V^j}$, can be expressed as

$$v_i^j = \frac{1}{h_j} \int_{x_{i-1}^j}^{x_i^j} \mathcal{R}_j(x; v^j) dx. \tag{21}$$

In one dimension, there is a simple way to obtain reconstruction operators satisfying (21). The basic observation is that if $\{v_i^j\}$ represent the cell averages of a function $f(x) \in L^1[0,1]$, we can define

$$F_i^j := h_j \sum_{s=1}^{i} v_s^j = \int_0^{x_i^j} f(x) dx = F(x_i^j)$$

where $F(x) := \int_0^x f(x) dx$ is a primitive of $f(x)$. Using the sequence $\{F_i^j\}_i$, we approximate $F(x)$ by interpolation, i.e. construct $\mathcal{I}(x, F^j)$, and then obtain $\mathcal{R}_j$

satisfying (21) as follows [30, 5]:

$$\mathcal{R}_j(x, v^j) = \frac{d}{dx}\mathcal{I}(x, F^j) \tag{22}$$

The reconstruction operator defined above is *consistent* with the discretization operator (21) since

$$(\mathcal{D}_j\mathcal{R}_j v^j)_i = \frac{1}{h_j}\int_{x_{i-1}^j}^{x_i^j} \frac{d}{dx}\mathcal{I}(x; F^j) = \frac{F_i^j - F_{i-1}^j}{h_j} = v^j.$$

Since $\mathcal{I}(x; F^j)$ is a continuous piecewise smooth function, $\mathcal{R}_j(x; v^j) \in L^1[0, 1]$.

With these definitions and the considerations of section 3, the one to one correspondence in (2) can be written as

$$\left\{ \begin{array}{rcl} v_i^{j-1} & = & (v_{2i}^j + v_{2i-1}^j)/2 \\ d_i^j & = & v_{2i-1}^j - (P_{j-1}^j v^{j-1})_{2i-1} \end{array} \right\} \longleftrightarrow \left\{ \begin{array}{rcl} v_{2i-1}^j & = & (P_{j-1}^j v^{j-1})_{2i-1} + d_i^j \\ v_{2i}^j & = & 2v_i^{j-1} - v_{2i-1}^j \end{array} \right\}$$

It must be noted that the reconstruction *via* primitive function is used as a design tool, but the practical implementation of the prediction operator in the cell-average settings never requires the explicit computation of $\mathcal{I}(x; F^j)$. In all the numerical experiments reported in this paper (and in the references) $\mathcal{R}_j(x; v^j)$ can be explicitely written directly in terms of the cell-averages v^j (see e.g. [30, 5]).

4 The role of $\mathcal{R}_j$: linear versus nonlinear

The description of the prediction operator as $P_{j-1}^j = \mathcal{D}_j\mathcal{R}_{j-1}$ opens up a tremendous number of possibilities in designing multiresolution schemes specifically adapted to particular applications. In principle, the general framework allows for nonlinear reconstruction operators, which in turn lead to nonlinear prediction operators and, thus, to nonlinear multiresolution transforms. The reconstruction process becomes, then, a key step, while the discretization operator selects a particular setting for the multiresolution decomposition.

The problem of reconstructing a function from a discrete set of data which gives relevant information on the function is purely a problem in approximation theory, which depends on the *interpretation* we assign to the discrete data. Probably the simplest case is that of point-value interpolation: When the given set of data are the values of a function at a finite set of points, the function can be approximated by an *interpolant*, i.e. another function whose values at the given set of points are exactly those of the original one.

Interpolatory techniques are a well understood subject, studied nowadays in elementary numerical analysis courses. The interpolant is usually a function easy to work with, thus the most usual interpolatory techniques involve polynomials or trigonometric functions. The case of one dimensional Lagrange

polynomial interpolation is particularly simple and serves to illustrate the effect of using linear (i.e. data-independent) techniques versus nonlinear (i.e. data-dependent) techniques.

4.1 Data-independent piecewise polynomial (DIPP) interpolation

Let $\mathcal{S}$ denote the stencil

$$\mathcal{S} = \mathcal{S}(r,s) = \{-s, -s+1 \cdots, -s+r\}, \qquad r \geq s > 0, \quad r \geq 1$$

and let $\{L_m(y)\}_{m\in\mathcal{S}}$ denote the Lagrange interpolation polynomials for this stencil

$$L_m(y) = \prod_{l=-s, l\neq m}^{-s+r} \left(\frac{y-l}{m-l}\right), \qquad L_m(i) = \delta_{i,m}, \quad i \in \mathcal{S}.$$

If we consider, as before, a grid $X = \{x_i\}$ composed of a finite sequence of equally spaced points on $[0,1]$, it is clear that

$$q_i(x; f, r, s) = \sum_{m=-s}^{-s+r} \bar{f}_{i+m} L_m\left(\frac{x-x_i}{h}\right),$$

satisfies $q_i(x_l) = f_l$, for $i - s \leq l \leq i - s + r$. When $f_l = f(x_l)$, we say that $q_i(x)$ interpolates $f(x)$ at the *stencil* of points $\mathcal{S}_i = \{x_{i-s}, \cdots, x_{i-s+r}\}$.

Hence,

$$\mathcal{I}(x,f) := q_i(x; f, r, s) \qquad x \in [x_{i-1}, x_i], \quad 1 \leq i \leq J \tag{23}$$

is a piecewise polynomial function that interpolates $f(x)$ on the grid X.

It is well known that

$$f(x) = q_i(x) + f[\mathcal{S}_i, x]\Pi_{x_m\in\mathcal{S}_i}(x - x_m) \tag{24}$$

where $f[\mathcal{S}, x] = f[x_{i-s}, \ldots, x_{i-s+r}, x]$ is the $r+1$st divided difference of $f(x)$ at the points of the stencil and x. If $f(x)$ is sufficiently smooth (at least differentiable $r+1$ times) then

$$f[\mathcal{S}_i, x] = \frac{f^{(r+1)}(\xi_x)}{(r+1)!},$$

where ξ_x is a point in the convex hull of the set $\{\mathcal{S}_i, x\}$. Therefore for sufficiently smooth functions, the piecewise polynomial interpolatory technique just described satisfies

$$\mathcal{I}(x; f) = f(x) + O(h^{r+1}), \qquad \forall x \in [0,1].$$

The O-term depends on $f^{(r+1)}$, which ensures that the piecewise polynomial interpolant is exact when $f(x)$ is a polynomial function of degree less than or equal to r.

Lack of smoothness in $f(x)$ can, however, have a dramatic effect on the *quality* of the approximation. If $f^{(p)}(x)$ has a jump discontinuity in the convex hull of a set of points $\{x_l, \ldots, x_{l+m}\}$, ($p \geq 0$ is the first index in which the singularity appears for the set of points considered) then one can prove that

$$f[x_l, \ldots, x_{l+m}] = \begin{cases} O([f^{(p)}])/h^{m-p} & \text{if } m > p \\ O(||f^{(m)}||) & \text{else} \end{cases} \tag{25}$$

where $||f^{(m)}||$ stands for the max-norm of $f^{(m)}$ in the convex hull of the considered set of points, and $[f^{(p)}]$ is the value of the jump at the discontinuity.

An isolated discontinuity in $f^{(p)}$ affects the accuracy of each one of the polynomial pieces of $\mathcal{I}(x, f)$ whose stencil crosses the discontinuity, and the degradation of the accuracy depends on the strength of the singularity. The worst case is of course when $p = 0$, in this case the global error becomes $O(1)$. It also happens that increasing r results in an interpolant with a larger "polluted region " (see section 4.5 and also [5]).

It is obvious that a smooth polynomial function cannot provide an accurate approximation at any interval containing a singularity; however, at a singularity-free interval the function is smooth, and one would like to use a polynomial piece that is as accurate as possible. As long as the convex hull of the polynomial stencil is contained in a singularity-free region of $f(x)$, divided differences can be considered as derivatives, and the interpolation error formula guarantees that we get full accuracy. Hence, to maintain accuracy we need to be able to choose singularity-free stencils, whenever this is possible.

Nota 4 *Because of the form of the interpolation error (24), centered stencils are always the preferred choice. In our notation, this corresponds to the choice* $r = 2s - 1$, *thus* $\mathcal{S}_i = \{x_{i-s}, \cdots, x_{i+s-1}\}$.

When the given function is periodic, i.e. $f_{-i} = f_{J_j - i}$, $f_{J+i+1} = f_{i+1}$, $0 \leq i \leq J$, *the data to construct the polynomial function* $\mathcal{I}(x, f)$ *using centered stencils is always available. If the function is not periodic, one can simply choose one sided stencils, of* $r + 1 = 2s$ *points, at intervals where the centered-stencil choice would require function values which are not available.*

4.2 Data-dependent piecewise polynomial interpolation: essentially non oscillatory (ENO) interpolation

The essential feature of the ENO interpolatory technique [31] is the stencil selection procedure. For each subinterval $I_i = [x_{i-1}, x_i]$ where f is smooth, the goal is to design a strategy that leads to a stencil $\mathcal{S}_i$ which does not 'cross singularities'. Such a process needs smoothness indicators, and the observations of last section (in particular (25)) point out that the divided differences could be used as such.

We will denote the ENO stencil for the ith interval, in a $r + 1$st order technique, as follows

$$\mathcal{S}_i^{ENO} = \{x_{s_i - 1}, x_{x_{s_i}}, \ldots, x_{s_i + r - 1}\}$$

There are several ENO strategies for the selection of the stencil. Here we shall only describe the one most commonly used:

Algorithm I. *Hierarchical* choice of stencil:

For each $i = 1, \ldots, J$

Set $s_0 = i$
for $l = 0, ..., r-2$
if $|f[x_{s_l-2}, \ldots, x_{s_l+l}]| < |f[x_{s_l-1}, \ldots, x_{s_l+l+1}]|$ then $s_{l+1} = s_l - 1$
end
$s_i = s_{r-1}$

Observe that x_{i-1}, x_i always belong to $\mathcal{S}_i$. When $r = 1$, $\mathcal{S}_i = \{x_{i-1}, x_i\}$ and no stencil selection is needed.

The effect of the ENO choice of stencil in the presence of singularities can be easily appreciated: Let us consider again the case of a smooth function except for an isolated jump discontinuity, $x_d \in I_j$. Let $\mathcal{S}$ be a stencil of $s+1$ points which does not cross the singularity (does not contain *both* x_j *and* x_{j-1}), and $\mathcal{S}^*$ a singularity-crossing stencil with the same number of points. Because of (25),

$$f[\mathcal{S}] = O(1), \qquad f[\mathcal{S}^*] = O(\frac{1}{h^s}).$$

Hence, divided differences based on singularity-crossing stencils are always larger than divided differences of the same order whose stencil is included entirely in a region of smoothness of $f(x)$ (at least for sufficiently small h). Because of this fact, Algorithm I would lead to stencils which *move away* from a jump discontinuity for any I_i, $i \neq j$.

In the case of a corner (i.e. a jump discontinuity in $f'(x)$) the situation is quite similar. With the same notation as in the paragraph above, because of (25), we have for $s \geq 2$

$$f[\mathcal{S}] = O(1), \qquad f[\mathcal{S}^*] = O(\frac{1}{h^{s-1}})$$

Since the first step in Algorithm I involves divided differences of second order, singularity-crossing stencils lead also to larger divided differences. Once again the stencils obtained with both algorithms *move away* from the singularity.

Thus, for jump discontinuities and corners, the ENO stencil selection procedure of Algorithm I leads to interpolatory polynomials that satisfy

$$f(x) = q_l(x) + O(h^{r+1}) \qquad x \in [x_{l-1}, x_l], \quad l \leq j-1, \quad l \geq j+1 \qquad (26)$$

and all polynomial pieces are fully accurate, with the exception of q_j.

Independently of the way in which it is constructed, a smooth polynomial piece $q_l(x)$ can only be a poor approximation to a (non-smooth) function $f(x)$ at a cell containing a singularity. When the singularity is a jump discontinuity, there is *always* one grid-cell where the accuracy is completely lost. For corners, there is one special situation where the accuracy loss will not occur at all: If

the singularity falls on a grid point, i.e. $x_d = x_j$. A stencil selection process such that

$$\{x_{j+1}\} \cap \mathcal{S}_l = \emptyset, \quad l \leq j, \qquad \{x_{j-1}\} \cap \mathcal{S}_l = \emptyset, \quad l \geq j+1 \tag{27}$$

guarantees $q_l(x) = f(x) + O(h^{r+1}) \quad x \in [x_{l-1}, x_l], \quad \forall l$, that is, we obtain an accurate approximation of the original function $f(x)$ in the *entire* interval.

It is unlikely that a singularity falls exactly on a grid point. However, if we happen to know the location of the singularity within the cell (or a sufficiently good approximation to it), the definition of the piecewise interpolants $\mathcal{I}(x)$ can be modified to keep the relation $\mathcal{I}(x) = f(x) + O(h^{r+1})$ valid over a region that contains almost all the interval where the singularity lies. This is the basic idea behind Harten's Subcell Resolution (SR) technique, which we describe next.

4.3 The subcell resolution technique

It is clear that when considering the point-values of a piecewise smooth function, all information on the *location* of the discontinuities is completely lost. On the other hand, it is quite easy to see that there is a direct connection between the location of the discontinuities and the cell-averages of a piecewise smooth function in one dimension. The connection was used by Ami Harten within the context of ENO schemes for Hyperbolic Conservation Laws to sharpen the profiles of contact discontinuities [26]. Harten's original technique uses discrete information on the cell-averages of a function $f(x)$ on a given grid, to recover (approximately) the location of an isolated jump discontinuity in $f(x)$, and because of this he named it *Subcell Resolution*.

Because of the relation between the cell-averages of $f(x)$ and the point-values of its primitive $F(x) = \int_0^x f(s)ds$ (see also section 3.2), Harten's SR is best described as an approximation technique that uses discrete information on the *point-values* of a continuous functions $f(x)$ to recover the location of an isolated discontinuity in $f'(x)$.

Let us assume that $f(x)$ is a continuous, piecewise smooth, function with a *corner* at $x_d \in I_j = (x_{j-1}, x_j)$, i.e.

$$f(x) = \begin{cases} P_L(x) & x \leq x_d \\ P_R(x) & x \geq x_d \end{cases} \quad \text{where} \quad \begin{cases} P_L(x_d) & = & P_R(x_d), \\ P'_L(x_d) & \neq & P'_R(x_d). \end{cases} \tag{28}$$

with $P_L(x)$ and $P_R(x)$ sufficiently smooth functions. Assume also that the polynomial pieces $q_{j\pm1}(x)$ are fully accurate, i.e. they satisfy

$$q_{j-1}(x) = P_L(x) + O(h^{r+1}), \quad q_{j+1}(x) = P_R(x) + O(h^{r+1}). \tag{29}$$

The function $G_j(x) := q_{j+1}(x) - q_{j-1}(x)$, satisfies $G_j(x) = P_R(x) - P_L(x) + O(h^{r+1})$ for $x \in I_j$. A Taylor expansion exercise easily leads to

$$G_j(x) = (x - x_d)[f']_{x_d} + O((x - x_d)^2) + O(h^{r+1})$$

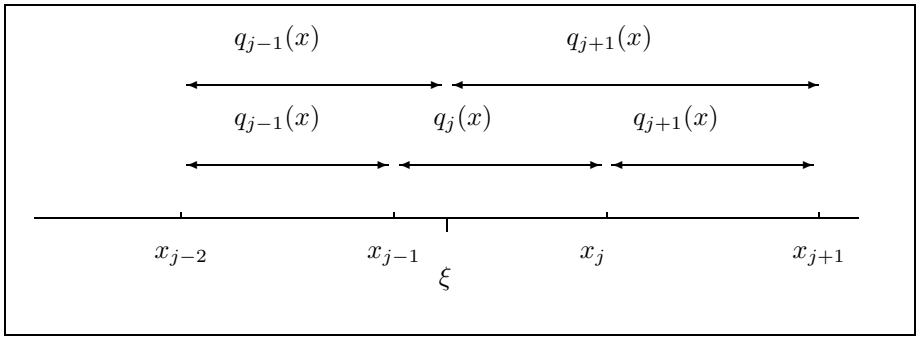


Figure 1: The polynomial pieces and their modification in SR

therefore, if h is sufficiently small, and $x \in I_j$, then $G_j(x) \approx (x - x_d)[f']_{x_d}$ and there must be only one root of $G_j(x)$ in I_j. Let $\xi \in (x_{j-1}, x_j)$ be such that $G_j(\xi) = 0$.

Observe that if P_L and P_R are polynomial functions of degree at most r, and the polynomial pieces $q_{j\pm1}(x)$ are fully accurate, we must have $q_{j-1}(x) = P_L(x)$ and $q_{j+1}(x) = P_R(x)$, which implies that $\xi = x_d$. In the general case (29), it is not hard to prove that $\xi = x_d + O(h^{r+1})$, at least for h sufficiently small [5, 20]. Thus, using *fully accurate* polynomial pieces at each side of a corner, we can recover its location up to the order of the truncation error.

The approximate location of the corner can be used to modify $\mathcal{I}(x)$ as follows: Instead of taking the polynomial $q_j(x)$ as the approximation of $f(x)$ in I_j, we extend the polynomial pieces at the left and right neighboring intervals up to the point ξ, where they intersect. The new piecewise polynomial interpolant has the following form (see Figure 1):

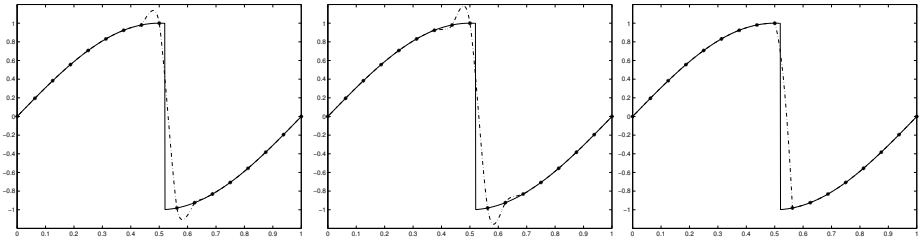
$$\mathcal{I}^{SR}(x) = \begin{cases} q_l(x) & x \in [x_{l-1}, x_l], \quad l \neq i \\ q_{j-1}(x) & x \in [x_{j-1}, \xi] \\ q_{j+1}(x) & x \in [\xi, x_j]. \end{cases} \tag{30}$$

It is clear that $\mathcal{I}^{SR}(x_l) = f(x_l)$, $\forall x_l \in X$ and that $\mathcal{I}^{SR}(x) = f(x) + O(h^{r+1})$, at all points except for an $O(h^{r+1})$ band around x_d which is now the only region in which the accuracy is degraded (instead of the whole interval $[x_{j-1}, x_j]$). Since ENO techniques lead to fully accurate polynomial pieces except for the cell containing the singularity, ENO-SR techniques lead to piecewise polynomial interpolatory functions with the *largest possible* region of high accuracy.

Nota 5 *It follows that if $f(x)$ is as in (28), with $P_{L,R}$ polynomials of degree up to r, then $\mathcal{I}^{SR}(x) = f(x)$, i.e., the modified reconstruction is exact.*

4.4 Linear versus nonlinear reconstructions: numerical performance

The performance of the reconstruction procedures described above can be quite different in the presence of singularities. The following simple example serves to illustrate their behavior.



(a) (b) (c)

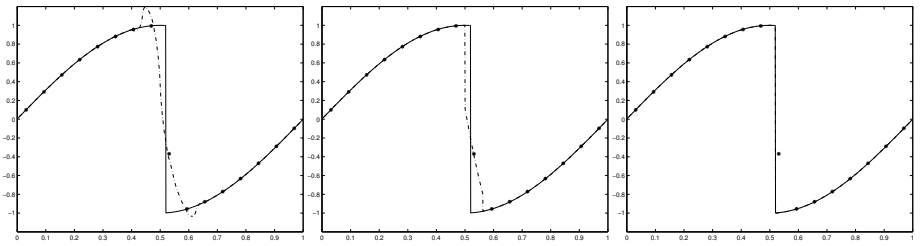
Figure 2: Solid line: A piecewise smooth signal (solid line) and its point-values on a uniform grid (dots on solid line). Dotted lines: (a)-(b) Linear Interpolatory reconstructions (a)-$r = 3$, (b)-$r = 5$; (b) Nonlinear ENO interpolatory reconstruction $r = 4$.

Let us consider the 1D signal displayed as a solid line in Figures 2 and 3: a periodic sinusoidal discontinuous function, and assume that we discretize it by considering its point-values on a discrete grid with 16 equally spaced intervals $v_i^l = f(x_i^l)$ and $X^l = \{x_i^l\}_{i=0}^{16}$, $h_l = 1/16$ (the discrete values are shown by dots in the plots in Figure 2.

Figures 2(a)-(b) show $\mathcal{I}(x, v^l)$ obtained using DIPP interpolation as specified in section 4.1 and $r = 3, s = 2$ in (a) and $r = 5, s = 3$ in (b) (in both cases it corresponds to a centered choice of the stencil). The *quality* of the approximation should be compared with the one obtained in Figure 2(c), where we used the nonlinear ENO interpolatory technique of Algorithm I for the selection of the stencil ($r = 3$ here). Observe that the $\mathcal{I}^{ENO}(x, v^l)$ keeps a fully accurate approximation right up to the interval, on X^l, where the discontinuity is located.

It should be noted that the linear interpolatory reconstruction obtained with $r = 5$ is *worse* than the one obtained with $r = 3$ not only in the two intervals next to the discontinuity; the effect of the discontinuity is felt in 5 intervals instead of only 3 for $r = 3$; again, this should be compared with the ENO interpolant, for which the plot corresponding to $r = 5$ (not shown) is indistinguishable from that of $r = 3$.

Assume now that we choose to discretize the signal by considering its cell averages on the same grid as before. In Figure 3, the dots, which have been placed at the center of each subinterval, represent the cell-averages of $f(x)$ on the same grid as in Figure 2. The dotted line on Figure 3-(a) displays $\mathcal{R}(x, v^j)$ in (22) when $\mathcal{I}(x; F^j)$ is constructed using the linear DIPP technique used before with $r = 3$ (hence, the polynomial pieces of $\mathcal{R}(x, v^j)$ in (22) are parabolic). Again, the *quality* of the approximation should be compared with those in Figures 3-(b) and (c) in which nonlinear interpolation techniques have been used in the reconstruction process. In Figure 3(b), we used $\mathcal{I}^{ENO}(x; F^j)$ and in Figure 3(c) we used $\mathcal{I}^{ENO-SR}(x; F^j)$ (same degree). Observe that the discontinuity in $f(x)$ becomes a corner in $F(x)$, the primitive function, hence



(a) (b) (c)

Figure 3: Solid line: A piecewise smooth function and its cell-averages on a uniform grid (dots on solid line). Dotted lines: (a) Linear reconstruction technique; (b) Nonlinear ENO technique; (c) Nonlinear ENO-SR technique

the SR technique described in section 4.3 leads to a reconstruction which is fully accurate right up to the discontinuity.

Figures 2 and 3 show that nonlinear techniques can be a powerful tool in approximating discontinuous functions. In particular, when the data are interpreted as cell-averages of a given piecewise smooth function, the combination of Harten's Subcell Resolution technique with the ENO interpolation (ENO-SR) allows for an almost perfect description of piecewise smooth functions with (sufficiently separated) jump discontinuities.

4.5 Linear versus nonlinear transforms: numerical performance

In Harten's framework, a sequence of nonlinear reconstruction operators $\{\mathcal{R}_j\}$ which is consistent with a given sequence of discretization operators $\{\mathcal{D}_j\}$ produces a nonlinear multiresolution transform. The decimation operators depend directly on the discretization sequence, and are always linear, but the prediction operator is nonlinear when the reconstruction is so.

The behavior of linear multiresolution transforms can be quite different from that of nonlinear multiresolution transforms. Recall that a scale coefficient d_i^j represents the (nonredundant) error committed by the prediction scheme at a particular location on the jth scale. In both the interpolatory and cell-average frameworks, these errors are directly related to interpolation errors, which are small in regions of smoothness. We have observed how linear reconstruction operators lead to regions of poor accuracy around singularities, which in turn produce large scale coefficients that *pile up* in a neighborhood of the singularity. Singularities have, thus, an associated *signature* which essentially measures the extent of the low accuracy region for the reconstruction process that defines the prediction.

We consider again the simple example of last section and discretize it in the cell-average setting to illustrate our point. The input of the multiresolution transform are the cell-averages of the function on a very fine grid X^L, where $J_L = 1024$. We consider the coarsest level to be specified by a grid X^0 with

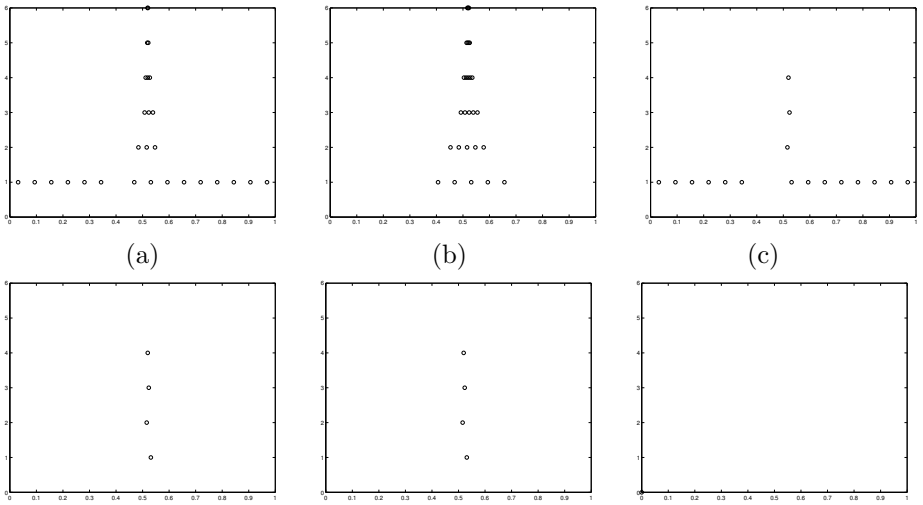


Figure 4: Detail coefficients above ϵ_j, $j = 1, \ldots, 6$. Top: (a)DIPP-(r=3,s=2), (b)DIPP-(r=5,s=3), (c)ENO-(r=3) Bottom: (d)ENO-(r=5), (e)ENO-SR-(r=3), (f)ENO-SR-(r=5)

$J_0 = 16$, thus the multiresolution transform involves $L = 6$ levels.

In Figure 4 we display, for each resolution level, the location of the detail coefficients that are above a level-dependent threshold specified a priori. We do this by plotting a little circle at the position i where $|d_i^j| > \epsilon_j$ for each j. In this particular case we chose $\epsilon_L = 10^{-2}$ and $\epsilon_k = \epsilon_{k+1}/2$, which is an appropriate strategy for the cell-average setting [27, 30, 5]. We readily observe that the presence of the discontinuity is felt at all resolution levels in the linear transforms. The fixed (centered in this case) choice of stencil leads to large interpolation errors, and thus to large scale coefficients, at those intervals with a *singularity-crossing* stencil. It can be clearly appreciated in Figures 4(a)-(b) that increasing the degree of the interpolation leads to a larger "polluted" region. On the other hand, the ability of the ENO strategy to construct *singularity-free* stencils can be clearly appreciated in Figures 4(c)-(d). The SR-technique for $r = 5$ is so accurate that no detail coefficients are above the prescribed tolerance.

Thus, linear multiresolution transforms are good tools to locate singularities, while nonlinear multiresolution transforms will be superior with respect to compression capabilities. We shall use these features in the applications of sections 6 and 7.

5 The question of stability

Often, the purpose behind a multiresolution decomposition is not just to decompose and reconstruct, the goal is to do some processing (truncation or

quantization) between the decomposition and reconstruction stages. Starting with a sequence v^L, we compute its multiresolution decomposition $Mv^L = (v^0, d^1, \ldots, d^L)$ and *process* it to obtain a *compressed* representation that is to be stored, transmitted or used in some way. The processing step involves a sequence of pre-determined tolerance levels $\bar{\epsilon} = (\epsilon_0, \epsilon_1, \ldots, \epsilon_L)$ and it is designed in such a way that the compressed representation $M_{\bar{\epsilon}} v^L = (\hat{v}^0, \hat{d}^1, \ldots, \hat{d}^L)$ satisfies $||v^0 - \hat{v}^0|| \leq \epsilon_0$ and $||d^j - \hat{d}^j|| \leq \epsilon_j$.

The goal is to be able to control *also* the difference between the 'uncompressed' sequence $\hat{v}^L = M^{-1}(\hat{v}^0, \hat{d}^1, \ldots, \hat{d}^L)$ and the original one. Stability is a crucial issue in examining the effect of using 'perturbed values' $\hat{v}^0, \hat{d}^j$ instead of v^0, d^j in the input of M^{-1}. To ensure stability, we need to have

$$||v^L - M^{-1}(\hat{v}^0, \hat{d}^1, \ldots, \hat{d}^L)|| = \sigma(\bar{\epsilon}), \qquad \text{with } \lim_{\bar{\epsilon} \to 0} \sigma(\bar{\epsilon}) = 0$$

The stability of the repeated application of (2) is seen to be tightly connected to the properties of the reconstruction sequence [30]. When $\mathcal{R}_j$ are linear operators (i.e. data independent), the multiresolution transform can be seen as a change of basis functions in V^L, and the stability of M^{-1} can be studied with linear techniques as in wavelets or subdivision schemes [30, 6]. However, when $\mathcal{R}_j$ are data-dependent, nonlinear operators, these techniques no longer apply and one needs a different approach to guarantee stability and control of the error.

5.1 Linear multiresolution algorithms

The decimation operators D_j^{j-1} are always linear. When the reconstruction operators, $\mathcal{R}_j$, are linear, the prediction operators P_{j-1}^j are linear too. In this case, the multiresolution transform becomes a linear operator describing a change of basis vectors in $\mathcal{D}_L(\mathcal{F})$, and the question of stability admits a relatively simple approach.

To see this, let us introduce the linear operator B_L^j of successive decimation

$$B_L^j = D_{j+1}^j \cdots D_L^{L-1} : V^L \to V^j \tag{31}$$

and observe that v^j in (3) can be written as $v^j = B_L^j v^L$. The direct multiresolution transform $v^L \mapsto M(v^L)$ can, thus, be expressed as

$$v^0 = B_L^0 v^L, \qquad d^j = G_j Q_j B_L^j v^L, \qquad 1 \leq j \leq L. \tag{32}$$

Likewise, let us introduce the operator A_j^L of successive prediction as

$$A_j^L = P_{L-1}^L \cdots P_j^{k+1} : V^j \to V^L. \tag{33}$$

When $\mathcal{R}_j$ is linear $\forall j$ these operators are also linear, and this fact allows us to express the inverse multiresolution transform directly in terms of $M(v^L)$ as follows:

$$v^L = A_0^L v^0 + \sum_{j=1}^{L} A_j^L E_j d^j. \tag{34}$$

Expressions (32) and (34) allow us examine the influence of perturbations in the input data of the direct and inverse multiresolution transforms rather easily.

For purposes of analysis, if v^L is replaced by a perturbed v_ϵ^L, stability of the direct multiresolution transform should imply that the perturbation in the resulting scale coefficients and low-level approximation has to be bounded by the perturbation in the input. Under our linearity assumptions, we can write

$$\delta(v^0) = v_\epsilon^0 - v^0 = B_L^0(v_\epsilon^L - v^L), \quad \delta(d^j) = d_\epsilon^j - d^j = G_j Q_j B_L^j(v_\epsilon^L - v^L). \quad (35)$$

These relations show that the perturbation in the input is subject to successive decimation D_{m-1}^m for $m = L, \ldots, j+1$, and in the case of the scale coefficients, projected into $\mathcal{N}(D_j^{j-1})$ and represented in some basis there. Clearly the 'dangerous' process that needs to be controlled, from the point of view of error-amplification, is that of successive decimation.

Similarly, for purposes of data compression if the scale coefficients $\{d^j\}$ are replaced by $\{d_\epsilon^j\}$ which are obtained either by quantization or truncation, we want the perturbation in the output of the algorithm, the decompressed v_ϵ^L, to be bounded by the perturbation in the scale coefficients. Linearity of all operators involved leads now to

$$\delta(v^L) = v_\epsilon^L - v^L = A_0^L(v_\epsilon^0 - v^0) + \sum_{j=1}^{L} A_j^L E_j(d_\epsilon^j - d^j), \quad (36)$$

which shows that the perturbation in the scale coefficients is 'translated' into a perturbation in the prediction error and then transmitted into higher levels of resolution by successive prediction P_{m-1}^m for $m = j+1, \ldots L$. The danger here is that the perturbation could be amplified by the process of successive prediction.

Thus, for linear multiresolution transforms, the successive decimation operator B_L^j controls the stability of the direct multiresolution transform, while the successive prediction operator A_j^L controls the stability of the inverse multiresolution transform.

The nested character of the sequence of discretization suffices to eliminate the possibility of amplification due to successive decimation. This is a consequence of proposition 2 since

$$B_L^j \mathcal{D}_L = D_{j+1}^j \cdots D_L^{L-1} \mathcal{D}_L = \mathcal{D}_j \mathcal{R}_{j+1} \mathcal{D}_{j+1} \cdots \mathcal{D}_{L-1} \mathcal{R}_L \mathcal{D}_L = \mathcal{D}_j. \quad (37)$$

The stability of the successive decimation step hinges on this purely algebraic relation. It essentially means that if we start at a given resolution level, L, and apply a number of decimation sweeps, say m, the discrete information we obtain is precisely what corresponds to the $L - m$ resolution level, in other words, the decimation operator does not introduce additional information or amplify noise.

Stability of the inverse multiresolution transform is usually a more involved matter. There is one situation, however, where the analysis is particularly simple:

5.1.1 Hierarchical sequences.

We say that the sequence $\{\mathcal{R}_j\mathcal{D}_j\}$ is hierarchical, if for all j

$$(\mathcal{R}_j\mathcal{D}_j)\mathcal{R}_{j-1} = \mathcal{R}_{j-1} \qquad \equiv \qquad \mathcal{R}_j P_{j-1}^j = \mathcal{R}_{j-1} \tag{38}$$

Note that for a hierarchical sequence

$$\mathcal{R}_L A_j^L = \mathcal{R}_L\mathcal{D}_L\mathcal{R}_{L-1}\cdots\mathcal{D}_{j+1}\mathcal{R}_j = \mathcal{R}_j. \tag{39}$$

Relation (39) shows that a hierarchical structure in the sequence $\{\mathcal{R}_j\mathcal{D}_j\}$ prevents the amplification of perturbations due to successive prediction in the same way nestedness, i.e. $\mathcal{D}_{j-1}(\mathcal{R}_j\mathcal{D}_j) = \mathcal{D}_{j-1}$, prevents excessive perturbation growth in the successive decimation step.

The algebraic relation (39) is the equivalent to (37) for the successive prediction operator. It means that after a finite number of applications of the prediction operator the reconstruction from the discrete information obtained is the same as the reconstruction obtained with the discrete data we started with. This is enough to ensure that the successive prediction step does not introduce spurious information or amplify existing noise (see [30] for specific error bounds).

5.1.2 Non-hierarchical sequences: The hierarchical form.

Hierarchical reconstructions are guaranteed to be stable. However, many reconstruction techniques used in numerical analysis are not hierarchical. For example the DIPP interpolation of section 4.1, one of the most common procedures in numerical analysis, does not lead to hierarchical reconstruction procedures when the polynomial pieces are of degree strictly larger than one (see [28, 30]). However, a sequence of approximation that is not hierarchical to begin with, has, in many cases, a *hierarchical form* which is obtained by considering a limiting process akin to refinement in subdivision schemes [12, 23]. This hierarchical reconstruction leads to the *same* prediction scheme as the original one (hence to the same multiresolution transform); thus the stability properties derived from the structure of the hierarchical reconstruction sequence are also inherited by the original (usually more transparent) one.

Teorema 3 *Let $\{\mathcal{R}_j\mathcal{D}_j\}$, $\{\mathcal{D}_j\}$ be sequences of bounded linear operators and assume that*

$$\mathcal{R}_j^H : V^j \to \mathcal{F} \qquad\qquad \mathcal{R}_j^H v^j = \lim_{L\to\infty} \mathcal{R}_L A_j^L v^j. \tag{40}$$

is a well defined operator for each j, i.e. the limit exists for all $v^j \in V^j$ and for all j. Then

1. *$\mathcal{R}_j^H$ is a reconstruction operator consistent with $\mathcal{D}_j$.*
2. *$(P^H)_{j-1}^j := \mathcal{D}_j\mathcal{R}_{j-1}^H = \mathcal{D}_j\mathcal{R}_{j-1} = P_{j-1}^j$;*

3. $\{\mathcal{R}_j^H \mathcal{D}_j\}$ *is a hierarchical sequence, i.e.* $(\mathcal{R}_{j+1}^H \mathcal{D}_{j+1})\mathcal{R}_j^H = \mathcal{R}_j^H$.

Demostración. To prove the theorem, we simple use proposition 2, the linearity (and boundedness) of the operators and the definition of the prediction operator in terms of the discretization and reconstruction operators.

$$\begin{aligned}
\mathcal{D}_j \mathcal{R}_j^H v^j &= \mathcal{D}_j \lim_{L\to\infty} \mathcal{R}_L A_j^L v^j = \lim_{L\to\infty} \mathcal{D}_j \mathcal{R}_L A_j^L = \mathcal{D}_j \mathcal{R}_j v^j = v^j \\
\mathcal{D}_j \mathcal{R}_{j-1}^H v^{j-1} &= \mathcal{D}_j \lim_{L\to\infty} \mathcal{R}_L A_{j-1}^L v^{j-1} = \mathcal{D}_j \mathcal{R}_j \mathcal{D}_j \mathcal{R}_{j-1} v^{j-1} = \mathcal{D}_j \mathcal{R}_{j-1} v^{j-1} \\
(\mathcal{R}_{j+1}^H \mathcal{D}_{j+1})\mathcal{R}_j^H v^j &= \mathcal{R}_{j+1}^H (\mathcal{D}_{j+1}\mathcal{R}_j^H) v^j = \mathcal{R}_{j+1}^H \mathcal{D}_{j+1} \mathcal{R}_j v^j \\
&= \lim_{L\to\infty} \mathcal{R}_L A_{j+1}^L P_j^{j+1} v^j = \lim_{L\to\infty} \mathcal{R}_L A_j^L v^j = \mathcal{R}_j^H v^j
\end{aligned}$$

□

Theorem 3 states that the existence of the limiting process in (40) implies, in turn, the existence of a hierarchical reconstruction procedure that produces exactly the same prediction operator as the original one. Hierarchical reconstructions lead naturally to stable multiresolution transforms, hence stability of the original scheme is a direct consequence of the *existence* of the limit in (40).

Since V^j is a finite dimensional vector space, $\mathcal{R}_j^H$, when it exists, is completely specified by the hierarchical reconstructions of a set of basis vectors. Indeed, let $\{\eta_i^j\}$ be a basis of V^j and $\varphi_i^j = \mathcal{R}_j^H \eta_i^j$, then if $v^j = \sum_i \hat{v}_i^j \eta_i^j$, because of linearity we have $\mathcal{R}_j^H v^j = \sum_i \hat{v}_i^j \mathcal{R}_j^H \eta_i^j = \sum_i \hat{v}_i^j \varphi_i^j$.

Thus, the existence of the limit functions

$$\varphi_i^j = \lim_{L\to\infty} \mathcal{R}_L A_j^L \eta_i^j \tag{41}$$

becomes a test for the stability of the multiresolution scheme derived from a particular sequence of discretization and reconstruction operators. For all practical purposes it is not important to know the explicit expression of the hierarchical form, however knowledge of its *existence* is essential because it implies stability of the original multiresolution scheme.

In the discretization by local averages framework of section 3, V^j are spaces of sequences, thus we can consider $\eta_i^j = \delta_i^j$, where δ_i^j is a sequence of $\dim V^j$ elements, all of which are zero except for the i-th position, whose value is 1. Then it is a simple matter to check, numerically, the convergence properties of the sequence $A_j^L \delta_i^j$. For this, we simply apply the inverse multiresolution transform with L levels to the sequence $(u^0, 0, \ldots, 0)$ with $u^0 = \delta_i^j$, i.e. taking the starting grid as the j-th grid, and all the scale coefficients as zero and applying L times the prediction scheme.

For illustration purposes we display in Figure 5 the $r = 3, s = 2$ DIPP case of section 4.1. We have not assumed periodicity, thus the reconstruction operators use non-centered stencils for the polynomial pieces in the two intervals next to each boundary (see remark 4). The displayed results correspond to $A_0^7 \delta_i^0$,

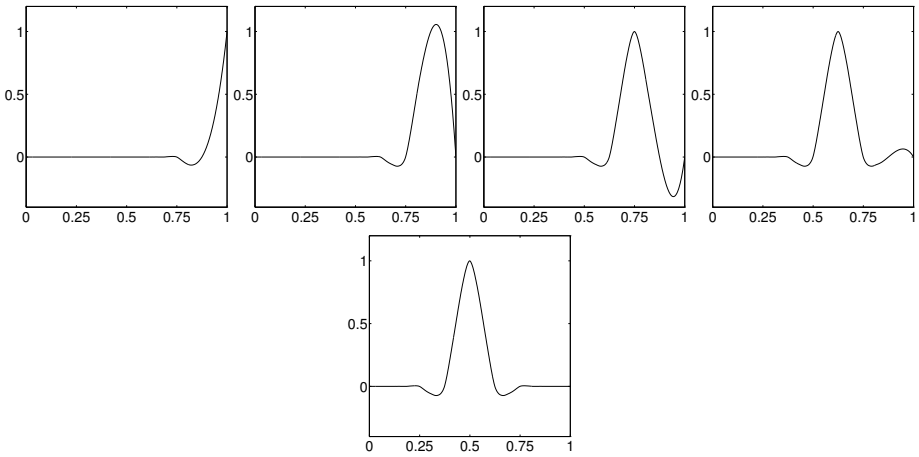


Figure 5: Limiting functions for interpolatory multiresolution. $J_0 = 8$ $r = 3$. Top: 'special' boundary functions. Bottom: periodic case.

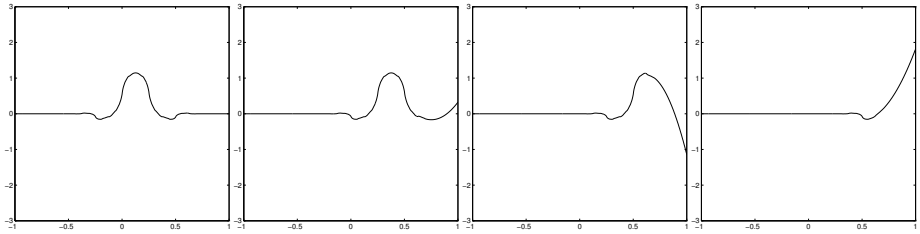


Figure 6: Limiting functions for cell-average multiresolution. $J_0 = 8$, $r = 3$. $\varphi_i^0, i = 5, 6, 7, 8$.

with $J_0 = 8$, $L = 7$, and $i = 4, \ldots, 8$. These are basically indistinguishable from $A_0^L \delta_i^0$ for L larger than 7. The remaining functions (not shown), $\varphi_i^{0,7}, i = 0, \ldots 3$ are specular images (with respect to the left boundary) of $\varphi_i^{0,7}, i = 8, \ldots, 5$.

In Figure 6 we display $A_0^7 \delta_i^0$ for the cell-average framework (with the same parameters and considerations as before). Recall that $\dim V^j = J_j$ now, thus we have 8 basis functions for V^0, the lowest resolution level. The functions φ_i $1, 2, 3, 4$ are specular images of the ones shown here.

Figures 5 and 6 give numerical evidence on the *existence* of these limit functions (hence on the stability of the multiresolution scheme). It is worth mentioning that the limiting process involved in obtaining the hierarchical reconstruction is very much related to the theory of subdivision refinement, $A_j^L v^j$ is computed by applying L times the prediction scheme to the data v^j. The existence of a limit of the successive refinement of a given sequence by a particular prediction process is a well studied subject. We refer the interested reader to the classical references [12, 22]. The connection can be directly exploited to prove the existence (and regularity) of these limit functions [8].

Orthogonal and biorthogonal wavelet algorithms can be seen as particular examples of this general framework [27, 28, 30]. The reconstruction operators used in these algorithms are hierarchical and, as a consequence, the associated compression algorithms are stable. As it turns out, they are the hierarchical form of other reconstruction operators [6, 30], for example the centered DIPP interpolatory techniques of section 4.1 lead to prediction schemes in the cell-average framework that are exactly those corresponding to the Biorthogonal wavelet framework of [14] when the scaling function is the box function. The reconstruction operator in the biorthogonal framework is just $\mathcal{R}_j^H$, where $\mathcal{R}_j$ is the reconstruction operator defined in (21) (see also [25] for more general results).

5.2 Nonlinear multiresolution algorithms: error-control strategies

The linearity assumption for the reconstruction operator is an essential ingredient in all proofs of last section. When $\mathcal{R}_j$ is allowed to be nonlinear, the arguments of last section can no longer be used; for example when the reconstruction operators are based on interpolatory techniques, lack of stability follows from the inability to ensure that the same stencil of points will be used in the decomposition (direct transform) and reconstruction (inverse transform) stages.

For nonlinear reconstruction operators, the question of stability needs a different approach. In his very first paper on multiresolution [27], Harten proposed an interesting strategy to achieve direct control on the difference $||v^L - M^{-1}\hat{v}^L||$, the so-called *error-control algorithms*, which involve a modification of the direct transform mechanism.

In an error-controlled multiresolution transform, the full sequence of decimated values $v^{L-1}, \ldots, v^1$ is computed first. Then, we start at the coarsest level and apply some perturbation process (usually thresholding or quantization) to v^0 to obtain $\hat{v}^0$ in such a way that $||v^0 - \hat{v}^0|| \leq \epsilon_0$, the user-set compression parameter for the coarsest level. From this point on, the idea consists in defining processed details $\hat{d}_j$ weaving them together with *processed values* $\hat{v}^j$, from coarse to fine scales in such a way that the accumulated compression error, $||v^j - \hat{v}^j||$, can be controlled at each step. For the cell-average case, the mechanism is sketched in Figure 7 (see [5, 2] for further details).

For $j = 1, \ldots, L$ the processed details $\hat{d}^j$ represent a perturbation (again it is usually quantization or truncation) of the *real* prediction error $v^j - P_{j-1}^j \hat{v}^{j-1}$ using the processed data $\hat{v}^{j-1}$, the processed data $\hat{v}^j$ is then computed as $\hat{v}^j = P_{j-1}^j \hat{v}^{j-1} + E_j \hat{d}^j$.

It is important to notice that the prediction errors $v^j - P_{j-1}^j \hat{v}^{j-1}$ do not necessarily belong to the null space $\mathcal{N}(D_j^{j-1})$ (in contrast to $e^j = v^j - P_{j-1}^j v^{j-1}$), and therefore one cannot simply define $\hat{d}^j$ as a perturbation of $G_j(v^j - P_{j-1}^j \hat{v}^{j-1})$. In the point-value and cell-average frameworks, there is a rather natural way to define the processed details $\hat{d}^j$ from the errors $v^j - P_{j-1}^j \hat{v}^{j-1}$. In the point-value setting, the compressed details $\hat{d}_i^j$ are

for $i = 1 : N_{k-1}$
$\tilde{d}_i^k = [\bar{f}_{2i-1}^j - (P_{j-1}^j \hat{f}^{k-1})_{2i-1}] - [\bar{f}_{2i}^j - (P_{j-1}^j \hat{f}^{k-1})_{2i}]$ /2;
end

$\hat{d}^k = proc(\tilde{d}^k, \epsilon_k)$

for $i = 1 : N_{k-1}$
$\hat{f}_{2i-1}^k = (P_{j-1}^j \hat{f}^{k-1})_{2i-1} + \hat{d}_i^k$;
$\hat{f}_{2i}^k = (P_{j-1}^j \hat{f}^{k-1})_{2i} - \hat{d}_i^k$ $(\equiv 2\hat{f}_i^{k-1} - \hat{f}_{2i-1}^k)$;
end

Figure 7: Error control algorithm for the Cell-Average setting in 1D

computed by processing (usually quantization or truncation) the differences $\tilde{d}_i^j = (v^j - P_{j-1}^j \hat{v}^{j-1})_{2i+1}$, but in the cell average setting $\hat{d}_i^j$ is obtained by processing instead $\tilde{d}_i^j = [(v^j - P_{j-1}^j \hat{v}^{j-1})_{2i} - (v^j - P_{j-1}^j \hat{v}^{j-1})_{2i+1}]/2$. In this last case, the coefficients $\tilde{d}_i^j$ coincide with the prediction errors e_{2i+1}^j only when $\epsilon_l = 0$, $l = 0, \ldots, L$, i.e. no compression takes place. If this is not the case, they represent a mean value between the prediction errors at odd and even points.

All in all, Harten's error-control strategy involves a modification of the analysis algorithm (the direct transform) that serves to ensure a prescribed accuracy after the application of M^{-1}. We refer the reader to [27], and more specifically to [5] for full details on the 1-D algorithms and to [3] for a full description of the 2-D error-control algorithms as well as the error bounds on the compression error $||v^L - M^{-1}(\hat{v}^0, \hat{d}^1, \ldots, \hat{d}^L)||$ in various norms.

6 The point-value setting: a multilevel strategy for the numerical simulation of Hyperbolic Conservation Laws

The solutions of Hyperbolic systems of Conservation Laws (HCL henceforth) are known to develop discontinuities (shock waves and contact discontinuities) in finite time. The discontinuities may develop spontaneously, and this is a major difficulty in designing numerical schemes that can accurately approximate these solutions.

Numerical schemes that attempt to obtain high order numerical approximations by traditional techniques, i.e. based on Taylor expansions, lead to oscillations around the discontinuities in the solution [34]. This oscillatory behavior is not only highly inaccurate, and very unpleasant, but also extremely dangerous (see [32]).

State of the art numerical schemes for HCL combine high order approximation in smooth regions with sharp transition profiles at discontinuities, and are known as High Resolution Shock Capturing (HRSC henceforth) methods. There are nowadays a number of HRSC schemes that provide reliable numerical approximations in many situations, however, the

perfect scheme has yet to be found (and may very well not even exist [35]), and the search continues for reliable schemes that can cope with (at least some of) the flaws displayed by well known HRSC schemes actually in use.

To keep our description simple, we shall consider a system of conservation laws in 1D

$$u_t + f(u)_x = 0 \tag{42}$$

where u is the vector of conserved variables and $f(u)$ the flux vector. The basic structure of a HRSC scheme for (42) on a uniform grid of mesh size δx is quite simple. In a semi-discrete formulation, it looks like

$$\frac{d}{dt}U_i + (\mathcal{D}U)_i = 0 \tag{43}$$

with the *numerical divergence* $(\mathcal{D}U)_i$ in *conservation form*

$$(\mathcal{D}U)_i = \frac{F_{i+1/2} - F_{i-1/2}}{\delta x}. \tag{44}$$

Here $F_{i+1/2}$ is the *numerical flux function*: the trademark of the scheme and where most of the computational work goes. Some HRSC schemes demand the computation of one spectral decomposition (or even two) of the Jacobian matrix $A(u) = \partial f/\partial u$ at each cell interface. In addition, some sort of reconstruction technique, possibly nonlinear, is also involved in the computation of $F_{i+1/2}$ if one wants to obtain high resolution in the presence of discontinuities.

The high resolution, un-oscillatory behavior that characterizes a HRSC scheme is a direct consequence of the sophisticated, highly complex and often very expensive numerical flux function used. However, every user of HRSC codes is painfully aware that these expensive flux evaluations are only necessary in a neighborhood of an existing discontinuity or in a region where compression, leading to shock formation, is taking place. In smooth regions, sufficiently far from discontinuities (existing or ready to form), a more traditional approach would lead to equally good results.

Any tool that can determine the regions of non-smooth behavior from a careful examination of the discrete data at a given time step could, in principle, be used to reduce the computational expense associated to a HRSC scheme. It is absolutely imperative to use a sophisticated flux formula in critical regions if one wants the full benefits of the scheme, but in smoothness regions one can do something cheaper, as long as it is equally accurate.

The link with multiresolution decompositions comes out naturally, and it was first explored by Harten [29] and pursued by other authors (see [1, 11, 37] and references therein). In [13], we propose an attractive alternative to the algorithms developed in these works.

6.1 The basic strategy: a multilevel evaluation of the numerical divergence

Let us consider again the simplest fully discrete realization of (43),

$$U_i^{n+1} = U_i^n - \delta t(\mathcal{D}U)_i^n \tag{45}$$

on a grid $\mathcal{G}_L$, where the HRSC scheme we have chosen gives us a numerical solution which we consider completely satisfactory. Our goal is to compute this numerical solution (or a 'high resolution' sufficiently close analog) at a much lower cost. In order to do this, we design a strategy that will allow us to obtain the values $\mathcal{D}(U^n)_i$, necessary to compute U^{n+1}, the numerical solution at the next time step, using the numerical flux computation prescribed by the HRSC *only when strictly necessary.*

The main ingredients in our multilevel scheme are the following:

• The multiresolution analysis of U^n, the numerical data available at the beginning of the time step, i.e. MU^n.

The multiresolution framework used to analyze the smoothness of the data is the simplest of all multiresolution frameworks: the interpolatory multiresolution framework. In this case, the *scale coefficients* (prediction errors) are simply interpolation errors, thus, its relation with the smoothness of the underlying function is well understood. We use uniform grids, with h_j being the mesh spacing in grid $\mathcal{G}_j$, and a centered interpolatory technique of degree 3 (see section 3.1), thus we know that

$$\begin{aligned} d_i^j &= O(h_{j-1}^4) \qquad \text{in smooth regions} \\ d_i^j &= O(1) \qquad \text{around jump discontinuities} \end{aligned}$$

• The thresholding algorithm which associates to each scale coefficient (therefore to a particular location in space) a boolean flag b_i^l, whose value (0 or 1) will determine the choice of procedure in the evaluation of the numerical divergence at that location. The thresholding algorithm uses the smoothness information contained in d_i^l to mark out the critical regions (discontinuities and compression regions) of *both* U^n *and* U^{n+1} (unknown at this stage of the computation). The details of the thresholding algorithm can be found in [13], but it is worth mentioning here that the propagation of information is limited by the CFL condition number *on the finest grid*, which bounds the location of moving singularities from one time step to the next.

It should also be mentioned that since the thresholding algorithm measures the size of interpolation errors, the thresholding parameters should be solution-dependent (but independent of the geometry of the computational domain).

• The multilevel evaluation of the numerical divergence To compute the values $(\mathcal{D}U)_i$ on the finest grid $\mathcal{G}_L$, we start by computing them on the lowest resolution grid $\mathcal{G}_0$ using our chosen HRSC scheme . Once the numerical divergence is known on $\mathcal{G}_{l-1}$, i.e. $(\mathcal{D}U)^{l-1}$ has been computed, we only need to determine $(\mathcal{D}U)^l$ at points in $\mathcal{G}_l \setminus \mathcal{G}_{l-1}$. To do this we examine the flag vector:

- if $b_i^l = 1$ compute $(\mathcal{D}U)_i^l$ directly with the HRSC scheme
- if $b_i^l = 0$, $(\mathcal{D}U)_i^l = \mathcal{I}(x_i^l, (\mathcal{D}U)^{l-1})$

letting l go from 1 to L gives the values of $\mathcal{D}(U^n)$ on $\mathcal{G}_L$.

6.2 Numerical tests: quality and efficiency

We remark that the purpose of the multilevel strategy is to be as close as possible to the *reference simulation*, i.e. to the solution that would be obtained with the underlying HRSC scheme on the same fine grid. Hence, the numerical results of the multilevel scheme are to be evaluated in terms of *Quality*, i.e. the difference between the outcome of the multilevel algorithm and the reference simulation, and *Efficiency*, i.e. net savings of the multilevel computation with respect to the reference simulation.

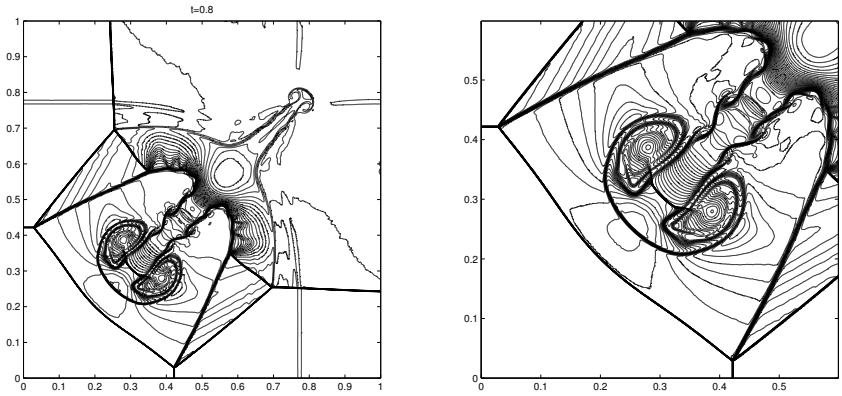
In [13], we use (a 2D tensor product version of) the multilevel scheme to compute numerical approximations to the solution of several benchmark tests for the Euler Equations of Gas Dynamics in 2D. The numerical results reported in [13] indicate that the quality of the multilevel approximation is directly controlled by the tolerance parameter used in the thresholding algorithm. The thresholding tolerance is a user-dependent parameter which is set at the beginning of the computation according to the order of the interpolation technique used in the multiresolution transform, the mesh spacing of the finest grid, and the solution itself.

The efficiency of the multilevel scheme is, of course, problem dependent. In a typical simulation involving a system of HCL, the non-smooth structures of the solution occupy a small percentage of the total computational domain, and in this situation the multilevel strategy is very efficient. We refer again to [13] for a specific evaluation of the efficiency of the multilevel scheme in several situations.

Figures 8 and 9, display high resolution numerical approximations to the solution of 2D-Riemann problems for the Euler equations of gas dynamics in 2D. The initial data is given on the unit square and involves 4 different constant states, one in each quadrant (see [36] for the initial data and a description of the exact the solution).

In Figure 8, the initial configuration gives rise to 4 strong shocks plus a number of interior structures that are very hard to compute accurately unless a high-order, high-resolution scheme is used on a fine mesh. The numerical results shown have been obtained with our multilevel strategy and Marquina's third order scheme [21] as the underlying HRSC scheme. Figure 8-(b) shows a zoom of the region at the left corner of the domain, where Kelvin-Helmholtz type instabilities start to develop, due to the low numerical viscosity in the simulation. It is worth mentioning that the percentage of mesh points on the finest grid where the numerical divergence has been computed with the HRSC scheme grows (in time) from 3% to 22% for this test case. In practice, this leads to an efficiency factor (decrease in cpu-time) of 4.2, with respect to the reference simulation (i.e. without multiresolution).

Decreasing the numerical viscosity in the simulation can have a dramatic effect on the computational results. The initial configuration for Figure 9 leads to the development of 4 contact discontinuities separating the initial states. Refining the grid, and thus reducing the numerical viscosity in the simulation, accentuates the Kelvin-Helmholtz instabilities at contact discontinuities. The



(a) (b)

Figure 8: 4-shock configuration: High Resolution numerical approximation to the density obtained with the Multilevel-Marquina scheme on a 1024×1024 grid

'roll-up' typical of this type of instability is more visible in finer grids. In this case, the percentage of HRSC divergence evaluation goes from 4% to 28 % for the 1024×1024 grid, and an associated decrease in cpu time by a factor of 3.6. In practice, this means that the simulation (on a PC at 350Mhz) can be run in a few days instead of a few weeks.

One of the nicest features of our multilevel algorithm lies in its simplicity: it can be introduced into an existing HRSC code without modifying its essential structure. This feature makes it into a useful tool to explore the possibilities (and the flaws) of state of the art HRSC schemes. It gives the user the possibility to use very fine (uniform) grids to run test problems at low cost: the cost of the user's own numerical technique on a much coarser mesh.

7 The cell-average setting: nonlinear multiresolution transforms for image compression

Images can be understood as discrete data corresponding to piecewise smooth functions. When designing a compression scheme, the approximation properties of the reconstruction sequence play a key role. Highly accurate reconstruction operators lead to small prediction errors, which can be truncated or quantized with very little loss in *real* information contents.

In section 4, we have seen that nonlinear, data dependent reconstruction techniques can be used to obtain fully accurate approximations almost up to discontinuities. In particular, ENO interpolation plus Harten's Subcell Resolution technique in the cell-average setting maximize the region in which the approximation is fully accurate, and will in turn lead to multiresolution schemes with good compression capabilities.

Let us examine first the approximation capabilities of the various reconstruction techniques considered in this paper. For this, let us consider a purely geometric image, the one displayed in Figure 10-(a) (512×512 pixels). By repeated decimation (in the cell-average setting) we obtain a coarse version of the image with only 32×32 pixels (not shown); in the notation of this paper v^L is the original image, v^0 is the coarse representation and $L = 4$ in this case.

To show the advantages of nonlinear reconstruction processes, we compute $M^{-1}(v^0, 0, \ldots, 0)$, where the prediction operator, $P_{j-1}^j = \mathcal{D}_j \mathcal{R}_{j-1}$, is constructed as specified in section 3.2. In particular, in constructing $\mathcal{R}_j$, we shall consider piecewise polynomial techniques of degree 3 for the primitive function and

- A centered piecewise polynomial technique. We recall [30, 6, 25] that the resulting multiresolution transform is that of the *BOW* scheme of [14] ($N = 1, \tilde{N} = 3$).
- A nonlinear ENO piecewise polynomial technique.
- A nonlinear ENO-SR technique.

The Gibbs-like oscillations typical of linear schemes in the presence of discontinuities lead to the blurring of the edges observed in Figure 10-(c). The edges are much sharper in the non-linear case and, in this special situation in which the discontinuities are aligned with the grid, the ENO-SR prediction scheme gives back the exact original image.

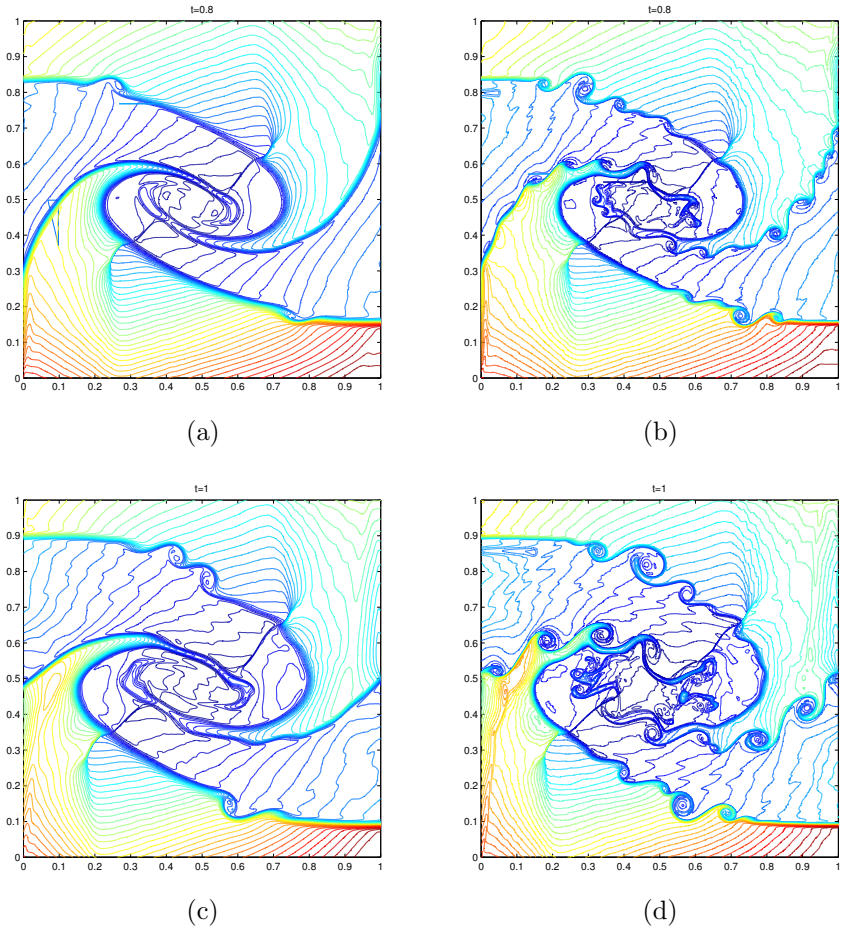


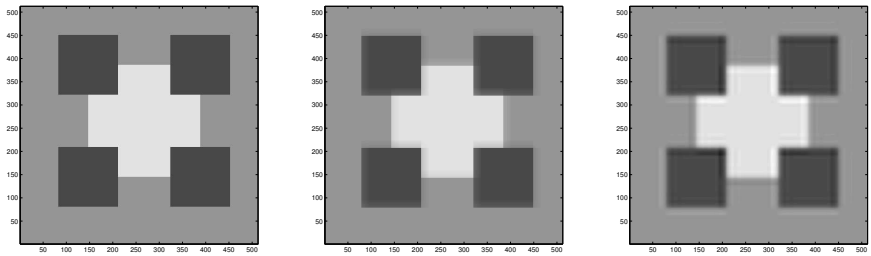
Figure 9: 4-contact configuration: High Resolution numerical approximation to the density obtained with the Multilevel-Marquina scheme. (a) and (b) t=.8, (c) and (d) t=1. (a) and (c) 512×512 mesh-points, (b) and (d) 1024×1024 mesh-points.

The limitations of the nonlinear techniques can be observed by repeating this same experiment but with an image with geometric features not aligned with the tensor-product grids, as in Figure 11. Since $d^j \equiv 0$, $j = 1, \ldots, L$, the data shown in Figures 10 and 11 allow us to see those regions where the reconstruction process leads to large prediction errors. The blurred region is larger in the linear scheme, more localized in the plain ENO alternative and even more so in the ENO-SR case. It is clear that the number of coefficients kept in compressing the original image would be smaller for the nonlinear schemes.

We refer the reader to [2, 3] for a thorough comparison of the capabilities of ENO-type reconstruction operators for image compression with respect to traditional wavelet-based alternatives.

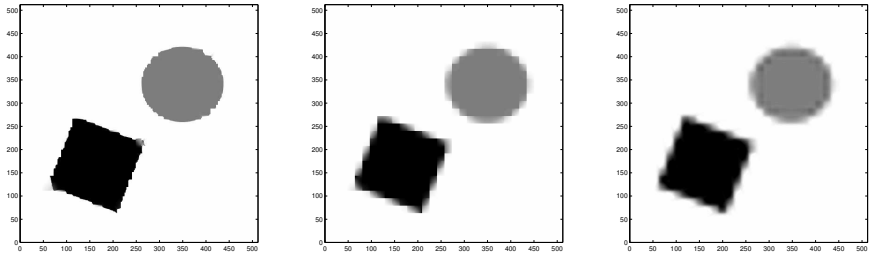
As pointed out in section 5, stability is not granted when using nonlinear multiresolution transforms. To ensure stability for ENO-based compression schemes, we implement the error-control algorithms described in [3] (it is not known at the moment whether or not the *direct* ENO-based nonlinear multiresolution transforms are stable). We end this section with an example that shows the differences in the outcome of the 2D error-control algorithm [3] versus the direct algorithm.

We consider the cell-average setting to compress the familiar image of Lena (512×512 pixels) using the same ENO technique as before for the prediction step. We use the direct algorithm with $L = 4$ to compute Mv^L and compress by truncating the scale coefficients i.e. if $|d_i^j| < \epsilon_j$ then $\hat{d}_i^j = 0$, otherwise $\hat{d}_i^j = d_i^j$. For this test case we have used $\epsilon_L = 12$ and $\epsilon_j = \epsilon_{j-1}/2$. Figure 12-(b) shows $M^{-1}(v^0, \hat{d}^1, \ldots, \hat{d}^L)$. The number of non-zero elements in the compressed image $M_{\bar{\epsilon}} v^L = (v^0, \hat{d}^1, \ldots, \hat{d}^L)$ is 16852, which corresponds to a compression rate of 9.34:1 (.85 *bpp*), and $||v^L - M^{-1} M_{\bar{\epsilon}} v^L||_2 = 8.4$. On the other hand, Figures 12-(a) shows $M^{-1}(v^0, \tilde{d}^1, \ldots, \tilde{d}^L)$, where $M_{\bar{\epsilon}}^m v^L = (v^0, \tilde{d}^1, \ldots, \tilde{d}^L)$, is the outcome of the *modified encoding* that characterizes the error-control algorithms (see [3] for details). Now, the number of non-zero elements in $M_{\bar{\epsilon}}^m v^L$ is 17314, i.e. a compression rate of 9.12:1 (.88 *bpp*) and $||v^L - M^{-1} M_{\bar{\epsilon}}^m v^L||_2 = 5.0$. We remark



(a) (b) (c)

Figure 10: A purely geometric image. (a) original and reconstruction by ENO-SR multiresolution scheme; (b) reconstruction by ENO multiresolution scheme; (c) reconstruction from linear (*bow*) scheme.



(a) (b) (c)

Figure 11: A purely geometric image. (a) reconstruction by ENO-SR multiresolution scheme; (b) reconstruction by ENO multiresolution scheme; (c) reconstruction from linear (*bow*) scheme.

(a) (b)

Figure 12: Images of Lena reconstructed from a compressed nonlinear multiresolution decomposition. (a) with an error-control algorithm; (b) with a direct algorithm

that the error-control algorithm leads to comparable compression rates, while keeping at the same time a noticeable better quality in the reconstructed image.

The stability of the direct nonlinear multiresolution transform is still being investigated, but it should be mentioned that the error-control algorithms lead to compression schemes in which it is possible to guarantee a certain accuracy *a priori*, and thus could be useful even when using linear multiresolution transforms. We refer the interested reader to [3, 2].

Acknowledgments Este trabajo se ha llevado a cabo en el marco del proyecto europeo IHP HPRN-CT-2002-00286 "Nonlinear Approximation and Adaptivity: Breaking complexity in numerical modelling and data representation".

References

[1] R. Abgrall, Multiresolution in unstructured meshes: Aplications to CFD. In *Oxford University Press*, number 5, 1996.

[2] S. Amat, F. Aràndiga, A. Cohen, R. Donat, Tensor product multiresolution analysis with error control for compact image representation. *Signal Processing*, 4:587–608 2002.

[3] S. Amat, F. Aràndiga, A. Cohen, R. Donat, G. Garcia, M. Von Oehsen, Data compression with ENO schemes. *Appl. Comput. Harmon. Anal.*, 11(2):273–288, 2001

[4] F. Aràndiga, V. Candela, R. Donat, Fast multiresolution algorithms for solving linear equations: A comparative study. *SIAM J. Sci. Comput.*, 16:581–600, 1995.

[5] F. Aràndiga, R. Donat, Nonlinear multi-scale decompostitions: The approach of A. Harten. *Numer. Algorith.*, 23:175–216, 2000.

[6] F. Aràndiga, R. Donat, A. Harten, Multiresolution based on weighted averages of the hat function I: Linear reconstruction operators. *SIAM J. Numer. Anal.*, 36(1):160–203, 1999.

[7] F. Aràndiga, R. Donat, A. Harten, Multiresolution based on weighted averages of the hat function II: Nonlinear reconstruction operators. *SIAM J. Sci. Comput.*, 20(3):1053–1093, 1999.

[8] F. Aràndiga, R. Donat, J. Liandrat, Multiscale decompositions by discretization and prediction on the interval. In preparation.

[9] E. Bacry, S. Mallat, G. Papanicolau, A wavelet based space-time adaptive numerical method for partial differential equations. *Math. Modelling and Numer. Anal.*, 26:703–834, 1992.

[10] G. Beylkin, R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms I. *Comm. Pure Appl. Math.*, 44:141–183, 1991.

[11] B. L. Bihari A. Harten, Multiresolution schemes for the numerical solution of 2-D conservation laws. I. *SIAM J. Sci. Comput.*, 18(2):315–354, 1997.

[12] A. Cavaretta, W. Dahmen, C. Micchelli, Stationary subdivision. *AMS Memoirs*, 453, 1991.

[13] G. Chiavassa R. Donat, Numerical experiments with point value miltiresolution for 2d compressible flows. *SIAM J. Sci. Computing.*, 23(3):805–823, 2001.

[14] A. Cohen, I. Daubechies, J. C. Feauveau, Biorthogonal bases of compactly supported wavelets. *Comm. Pure Applied Math.*, 45:485–560, 1992.

[15] A. Cohen, I. Daubechies, P. Vial, Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.*, 1(1):54–81, 1993.

[16] I. Daubechies, Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:909–996, 1988.

[17] I. Daubechies, *Ten Lectures on Wavelets.* Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM, Philadelphia, 1992.

[18] I. Daubechies, J. Lagarias, Two scale difference equations I. *SIAM J. Math. Anal.*, 22:1388–1410, 1991.

[19] I. Daubechies, J. Lagarias, Two scale difference equations II. *SIAM J. Math. Anal.*, 23(1031-107), 1992.

[20] R. Donat, Studies on error propagation for certain nonlinear approximations to hyperbolic equations: Discontinuities in derivative. *SIAM J. Numer. Anal.*, 31:655–679, 1994.

[21] R. Donat, A. Marquina, Capturing shock reflections: An improved flux formula. *J. Comp. Phys.*, 125:42–58, 1996.

[22] N. Dyn, Subdivision schemes in computer-aided geometric design. In *Advances in numerical analysis, Vol. II (Lancaster, 1990)*, p. 36–104. Oxford Univ. Press, New York 1992.

[23] N. Dyn, J. A. Gregory, D. Levin, Analysis of linear binary subdivision schemes for curve design. *Constr. Approx.*, 7:127–147, 1991.

[24] B. Engquist, S. Osher, S. Zhong, Fast wavelet algorithms for linear evolution equations. *SIAM J. Sci. Comput.*, 15:755–775, 1994.

[25] M. Guichaoua, *Analyses Multirésolution Biorthogonales associées à la Résolution d'Equations aux Dérivées Partielles.* PhD thesis, Ecole Superieure de Mecanique de Marseille, Universite de la Mediterranee Aix-Marseille II, 1999.

[26] A. Harten, ENO schemes with subcell resolution. *J. Comput. Phys.*, 83:148–184, 1989.

[27] A. Harten, Discrete multiresolution analysis and generalized wavelets. *J. Applied Num. Math.*, 12:153–193, 1993.

[28] A. Harten, Multiresolution representation of data. Technical report, UCLA CAM Report 93-13, 1993.

[29] A. Harten, Multiresolution algorithms for the numerical solution of hyperbolic conservation laws. *Comm. Pure Appl. Math.*, 48:1305–1342, 1995.

[30] A. Harten, Multiresolution representation of data: A general framework. *SIAM J. Numer. Anal.*, 33:1205–1256, 1996.

[31] A. Harten, B. Engquist, S. Osher, S. R. Chakravarthy, Uniformly high-order accurate essentially nonoscillatory schemes. III. *J. Comput. Phys.*, 71(2):231–303, 1987.

[32] A. Harten, J. M. Hyman, P. D. Lax, On finite-difference approximations and entropy conditions for shocks. *Comm. Pure Appl. Math.*, 29(3):297–322, 1976 (with an appendix by B. Keyfitz).

[33] A. Harten I. Yad-Shalom, Fast multiresolution algorithms for matrix-vector multiplication. *SIAM J. Numer. Anal.*, 31(4):1191–1218, 1994.

[34] R. LeVeque, *Numerical Methods for Conservation Laws.* Birkhäuser Verlag, 1992.

[35] J. Quirk, A contribution to the great riemann solver debate. *Internat. J. Num. Meth. Fluids*, 18:555–574, 1994.

[36] C. W. Schulz-Rinne, J. P. Collins , H. M. Glaz, Numerical solution of the Riemann problem for two-dimensional gas dynamics. *SIAM J. Sci. Comput.*, 14:1394–1414, 1993.

[37] B. Sjögrenn, Numerical experiments with the multiresolution scheme for the compressible euler equations. *J. Comput. Phys.*, 117:251–261, 1995.

[38] G. Strang, Wavelets and dilation relations: A brief introduction. *SIAM Review*, 31:614–627, 1989.

Bol. Soc. Esp. Mat. Apl. nº31(2005), 109–138

Algunos resultados sobre el problema del termistor

M. T. González y F. Ortegón

Dpto. Matemáticas, Universidad de Cádiz

mariateresa.gonzalez@uca.es, francisco.ortegon@uca.es

Resumen

En este trabajo se analizan algunos sistemas de ecuaciones en derivadas parciales no lineales con origen en Electromagnetismo. En concreto, se estudian diversas situaciones del denominado *problema del termistor*, así como algunas variantes del mismo.

Palabras clave: *Termistor, ecuaciones elípticas no lineales y degeneradas, ecuaciones parabólicas degeneradas, soluciones débiles, soluciones renormalizadas, soluciones de capacidad.*

Clasificación por materias AMS: *35M10, 35J60, 35K65.*

1 Introducción

La palabra *termistor* es la adaptación al castellano del vocablo inglés *thermistor*, acrónimo de las palabras *thermally sensitive resistor*, es decir resistencia sensible a cambios de temperatura. Se trata de un dispositivo semiconductor, habitualmente de forma cilíndrica, de unos cinco milímetros de radio y dos milímetros de grosor, conectado dentro de un circuito mediante cables soldados en su parte superior e inferior (figura 1); estas superficies están cubiertas con una fina chapa metálica que actúa como contacto. La característica esencial del termistor es que está fabricado con un material cerámico cuya conductividad eléctrica depende fuertemente de la temperatura.

Dependiendo de que la resistividad (la inversa de la conductividad) de los materiales sea una función creciente o decreciente de la temperatura, los termistores pueden ser de *temperatura característica positiva* (termistores PTC con sus iniciales en inglés) o de *temperatura característica negativa* (termistores NTC). El decrecimiento de la conductividad es rápido, con un cambio típico de

Fecha de recepción: 16/02/05

magnitud de orden cuatro cuando la temperatura crece de 100℃ a 200℃. Por ejemplo, un termistor de 2252 ohmios (Ω) posee una sensibilidad de $-100\ \Omega/℃$ a temperatura ambiente; un termistor con resistencia más alta puede alcanzar una sensibilidad del orden de $-10000\ \Omega/℃$. Estos valores pueden compararse con otro tipo de dispositivos eléctricos, como las resistencias usadas como sensores de temperatura; por ejemplo, una resistencia de platino de 100 Ω posee una sensibilidad de tan solo 0'4 Ω/℃.

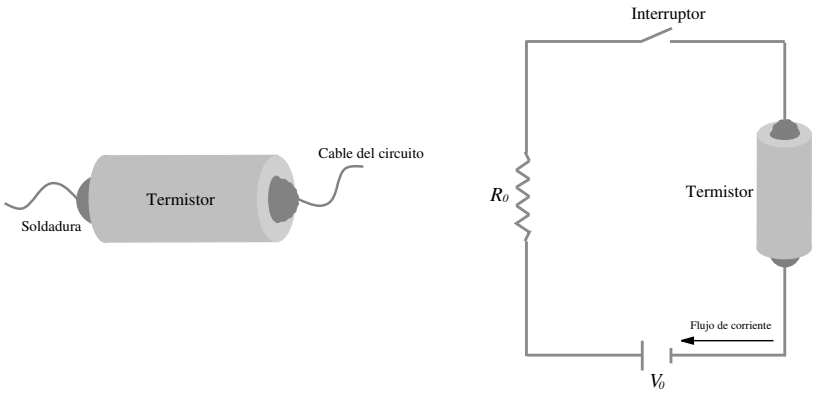


Figura 1: Termistor y circuito con un termistor integrado.

Entre las numerosas aplicaciones del termistor destacan las de regulador de sobretensión eléctrica, fusible, regulación de la corriente, interruptores o conmutadores, análisis de la conductividad térmica, controles y alarmas. También existe un largo historial de significativas aplicaciones en Ingeniería, siendo de particular interés las estructuras del termistor que se originan en aplicaciones de microsensores. El uso y explotación del termistor como un medidor de alta precisión de la temperatura ha tenido un enorme impacto, sobre todo por sus aplicaciones en Medicina y esto ha provocado el interés de muchos fabricantes por desarrollar y diseñar estos dispositivos; véase [35]. Este interés es relativamente reciente (desde hace unos treinta y cinco años), pero desde entonces numerosos investigadores (ingenieros, físicos y matemáticos) han dedicado muchos esfuerzos a estudiar el problema del termistor.

En la figura 1 se ha representado un pequeño circuito cerrado con un interruptor, que causa una corriente producida por un voltaje externo, V_0, para pasar a través de la resistencia R_0 y del termistor, calentándolo. El consecuente decrecimiento de la conductividad eléctrica provoca una caída de corriente hasta que se alcanza el equilibrio y todo el calor generado por el termistor se va propagando por sus alrededores. En un termistor bien diseñado, la corriente final debería ser una pequeña fracción de la inicial.

Existen numerosos problemas de interés práctico. Así, a menudo se requiere

un termistor hecho a medida para un montaje concreto que responda a ciertas necesidades o características, tales como el tiempo de conmutación o interrupción (el tiempo que tarda la corriente en alcanzar $1/e$ veces su valor inicial) y la corriente final. Es fundamental determinar cómo estas características dependen de los parámetros propuestos tales como, por ejemplo, tamaño, aspecto de la proporción, transmisión de calor a la superficie, resistencia externa, etc. Por otro lado, si el voltaje es demasiado grande, el termistor puede estropearse. Se sospecha que esta ruptura está causada por tensiones térmicas y, en consecuencia, es importante localizar dónde pueden producirse elevados gradientes de temperatura.

Cuando actúa como un interruptor automático en un circuito, un termistor funciona como sigue: un incremento de la intensidad de corriente produce más calor (efecto Joule), provocando un aumento de la temperatura del material, lo cual causa un aumento de la resistividad, reduciéndose así la corriente (a lo más hasta cero si la temperatura supera un límite crítico). Cuando el termistor se enfría, su resistividad decrece y se reanuda el funcionamiento normal del circuito.

1.1 Formulación del problema

En este trabajo se analizan diversas cuestiones del problema del termistor desde el punto de vista matemático, esto es, la resolución de las ecuaciones en derivadas parciales que gobiernan éste. Las incógnitas que intervienen en el problema son la temperatura, u, y el potencial eléctrico, φ.

A lo largo del trabajo, $\Omega \subset \mathbb{R}^N$ será un abierto acotado y regular, $N \geq 1$ y $T > 0$. Para un espacio de Banach X y un exponente $1 \leq p \leq +\infty$, denotaremos $L^p(X)$ el espacio de Banach $L^p(0,T;X)$. También emplearemos la función de truncamiento a la altura $K \in \mathbb{R}$, o sea, $T_K(s) = (\operatorname{sign} s)\,\min(K,|s|)$. Finalmente, la letra C designará constantes arbitrarias que sólo dependen de los datos iniciales.

Las ecuaciones que rigen el problema del termistor se deducen a partir de las leyes de conservación de la corriente y de la energía. Sean $\mathcal{J}$ la intensidad de corriente eléctrica, $\mathcal{Q}$ el flujo de calor y $\mathcal{E} = -\nabla\varphi$ el campo eléctrico. Las leyes de Ohm y de Fourier relacionan estos campos y la temperatura mediante las expresiones

$$\mathcal{J} = \bar{\sigma}(u)\mathcal{E}, \quad \mathcal{Q} = -\bar{a}(u)\nabla u,$$

donde $\bar{\sigma}(u)$ y $\bar{a}(u)$ son coeficientes de conductividad eléctrica y térmica, respectivamente. Las leyes de conservación de la corriente y de la energía vienen dadas por

$$\nabla \cdot \mathcal{J} = 0, \quad \rho c \frac{\partial u}{\partial t} + \nabla \cdot \mathcal{Q} = \mathcal{J} \cdot \mathcal{E},$$

donde ρ es la densidad del semiconductor y c su capacidad calorífica (se supondrá que ρ y c son constantes). Finalmente, poniendo $\sigma(u) = \bar{\sigma}(u)/(\rho c)$ (la conductividad eléctrica) y $a(u) = \bar{a}(u)/(\rho c)$ (la conductividad térmica), se

deduce el problema siguiente:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nabla \cdot (a(u)\nabla u) = \sigma(u)|\nabla\varphi|^2 & \text{en } Q = \Omega \times (0,T), \\ \nabla \cdot (\sigma(u)\nabla\varphi) = 0 & \text{en } Q, \\ u = 0 & \text{sobre } \partial\Omega \times (0,T), \\ \varphi = \varphi_0 & \text{sobre } \partial\Omega \times (0,T), \\ u(\cdot,0) = u_0 & \text{en } \Omega, \end{cases} \tag{1}$$

donde Ω (espacio ocupado por el termistor) es un abierto acotado y regular de $\mathbb{R}^N$, $N \geq 1$ y $T > 0$.

Por simplicidad, aquí se han considerado condiciones de contorno de tipo Dirichlet para u y para φ. Obviamente, se pueden tener en cuenta otras condiciones de contorno más generales (Neumann, Fourier, mixtas, etc.). Se observa entonces que el problema del termistor está constituido por dos ecuaciones en derivadas parciales no lineales acopladas, la primera de ellas de tipo parabólico y la segunda de tipo elíptico. Nótese que, aunque la ecuación para el campo eléctrico sea elíptica, la incógnita φ también depende del tiempo y lo hace a través del coeficiente $\sigma(u)$ (por eso indicamos que esta ecuación se cumple en $\Omega \times (0,T)$).

Una versión más general del problema del termistor contempla la situación en la que la ecuación para el campo eléctrico se cumple en un cilindro $\mathcal{O} \times (0,T)$, donde $\mathcal{O} \subset \Omega$ (véase [1]). Éste es el caso en el que el termistor está rodeado por capas de óxidos metálicos. En este trabajo siempre se supondrá que $\mathcal{O} = \Omega$.

También posee cierto interés la versión estacionaria, a saber:

$$\begin{cases} -\nabla \cdot (a(u)\nabla u) = \sigma(u)|\nabla\varphi|^2 & \text{en } \Omega, \\ \nabla \cdot (\sigma(u)\nabla\varphi) = 0 & \text{en } \Omega, \\ u = 0 & \text{sobre } \partial\Omega, \\ \varphi = \varphi_0 & \text{sobre } \partial\Omega. \end{cases} \tag{2}$$

1.2 El problema de evolución

¿Qué hace complicado el estudio del problema del termistor? Al analizar los términos que participan en (1), se observa que el acoplamiento de las incógnitas se realiza a través del segundo miembro de la ecuación del calor $\sigma(u)|\nabla\varphi|^2$ (efecto Joule) y a través de la conductividad eléctrica, que depende de la temperatura $\sigma(u)$. La aparición del término cuadrático $\sigma(u)|\nabla\varphi|^2$ es una de las dificultades en el estudio teórico de este problema. Por ejemplo, con la regularidad $\varphi \in L^2(H^1(\Omega))$ y $\sigma \in L^\infty(\mathbb{R})$ se tiene $\sigma(u)|\nabla\varphi|^2 \in L^1(Q)$ y entraríamos en el marco de las ecuaciones parabólicas con segundo miembro en L^1; este tipo de situaciones se resuelven por aproximación (truncamientos, estimaciones y paso al límite), pero en este contexto haría falta al mismo tiempo la convergencia en casi todo de las temperaturas y la convergencia fuerte en $L^1(Q)$ de los gradientes del campo eléctrico, convergencias que no se tienen en general.

No obstante, hay situaciones en que esta dificultad puede ser evitada. Por ejemplo, si $\varphi \in L^2(H^1(\Omega)) \cap L^\infty(Q)$, entonces, para $\phi \in \mathcal{D}(\Omega)$ y $t \in [0,T]$ casi

por doquier,

$$\int_\Omega \sigma(u)\nabla\varphi\cdot\nabla(\phi\varphi)=0,$$

es decir,

$$\int_\Omega \sigma(u)|\nabla\varphi|^2\phi=-\int_\Omega \sigma(u)\varphi\nabla\varphi\cdot\nabla\phi\,.$$

Como $\sigma(u)\varphi\nabla\varphi\in L^2(Q)^N$, se sigue que el término cuadrático $\sigma(u)|\nabla\varphi|^2$ no sólo pertenece a $L^1(Q)$ sino que además está en $L^2(H^{-1}(\Omega))$ y se tiene que

$$\sigma(u)|\nabla\varphi|^2=\nabla\cdot(\sigma(u)\varphi\nabla\varphi)\ \text{ en }\ L^2(H^{-1}(\Omega)).$$

Esto nos lleva a introducir el problema del termistor con término fuente en forma de divergencia:

$$\begin{cases} \dfrac{\partial u}{\partial t}-\nabla\cdot(a(u)\nabla u)=\nabla\cdot(\sigma(u)\varphi\nabla\varphi) & \text{en } Q=\Omega\times(0,T),\\ \nabla\cdot(\sigma(u)\nabla\varphi)=0 & \text{en } Q,\\ u=0 & \text{sobre } \partial\Omega\times(0,T),\\ \varphi=\varphi_0 & \text{sobre } \partial\Omega\times(0,T),\\ u(\cdot,0)=u_0 & \text{en } \Omega, \end{cases} \tag{3}$$

que es equivalente a (1) al menos con la regularidad $\varphi\in L^2(H^1(\Omega))\cap L^\infty(Q)$ y σ acotada. En muchas situaciones, es el problema (3) el que se sabe resolver, pero φ no es lo suficientemente regular como para asegurar la equivalencia con (1).

Sin embargo, lo que hace particularmente complicados los problemas (1) y (3) es el comportamiento de a y σ como funciones de u en los casos en que las ecuaciones sean degeneradas (a o σ se anulan) o no uniformemente elípticas ($a(s)$ o $\sigma(s)$ no están uniformemente acotadas inferiormente por constantes positivas o no están uniformemente acotadas superiormente). Ésta es precisamente la situación en los casos de especial interés práctico. Por ejemplo, si $\boldsymbol{u}>0$ representa la temperatura absoluta, medida en grados Kelvin (°K), entonces una aproximación posible es

$$a(\boldsymbol{u})=\frac{1}{A+B\boldsymbol{u}+C\boldsymbol{u}^2},\quad \sigma(\boldsymbol{u})=D\boldsymbol{u}^\gamma e^{-E/(k\boldsymbol{u})},$$

donde A,B,C,D y E son constantes positivas, $A+B+C>0$, $\gamma\in[-1,1)$ y k es la constante de Boltzmann ($k=1'38066\times10^{-23}$ julios/°K); véanse [10, 36, 28].

Entre los numerosos autores que han estudiado el problema de evolución del termistor encontramos a Chipot y Cimatti [11], que obtuvieron un resultado de unicidad en el caso en que $a\equiv1$ y $\sigma\in W^{1,\infty}(\mathbb{R})$. La clave de éste residía en la deducción de una estimación uniforme de φ en $L^\infty(W^{1,\infty}(\Omega))$. Por otro lado, Cimatti [16] demostró haciendo uso del método de Faedo-Galerkin la existencia de solución débil, suponiendo que $u_0=0$ en Ω, $\varphi_0\in C^2(\bar{Q})$, $a\equiv a_0$ y σ continuamente diferenciable, positiva y acotada; la utilización del teorema de Meyers y de las inyecciones de Sobolev limita la validez de sus resultados a $N=2$.

Antontsev y Chipot [3] trataron este problema suponiendo que $a, \sigma \in C^0(\Omega)$ eran tales que $0 < a_1 \leq a(s) \leq a_2$ y $0 < \sigma_1 \leq \sigma(s) \leq \sigma_2$ para cualquier $s \in \mathbb{R}$; dedujeron un resultado de existencia de solución débil haciendo uso del teorema del punto fijo de Schauder y asumiendo que $\varphi_0 \in L^\infty(H^1(\Omega)) \cap L^\infty(Q)$ y $u_0 \in L^2(\Omega)$. Más aún, analizaron la regularidad de las soluciones débiles y la existencia de soluciones clásicas bajo ciertas hipótesis de regularidad sobre las conductividades y los datos iniciales. Por consiguiente, con estas hipótesis para a y σ, se sabe resolver el problema del termistor. Desgraciadamente, como se ha indicado antes, en los casos de interés práctico estos coeficientes no cumplen estas hipótesis.

En [40] se estableció una estimación L^∞ para la temperatura, u, la cual permitió demostrar la existencia y unicidad de solución débil bajo las hipótesis $a \equiv 1$, $\sigma \in W^{1,\infty}(0, +\infty)$, $0 < \sigma_1 \leq \sigma(s) \leq \sigma_2$ y $|\sigma'(s)| \leq L$, $u_0 \in W^{2-2/p,p}(\Omega)$ y $\varphi_0 \in L^\infty(W^{2-2/p,p}(\Omega))$, con $p > N$.

Ping y Jishan [33] estudiaron el problema del termistor cuando $N = 3$ con condiciones de contorno mixtas. La conductividad térmica vuelve a considerarse constante y σ es acotada y diferenciable. La regularidad de los datos iniciales conduce a la obtención de soluciones también muy regulares. Asimismo, se plantea el problema con condiciones de contorno mixtas en [44], siendo $\Omega \subset \mathbb{R}^N$ un dominio acotado de clase $C^{2+\alpha}$ ($N \geq 1$); de nuevo, las hipótesis sobre los datos implican un resultado de existencia y unicidad de solución débil muy regular.

Xu [42] estudia un problema ligeramente distinto a (1), en el que el término $\nabla \cdot (a(u)\nabla u)$ se sustituye por $\nabla \cdot a(\nabla u)$, siendo $a : \mathbb{R}^N \mapsto \mathbb{R}^N$ una función continua tal que $|a(\xi)| \leq C|\xi|$ para $|\xi|$ suficientemente grande y $[a(\xi) - a(\eta)]\,(\xi - \eta) \geq \alpha|\xi - \eta|^2$ para cualesquiera $\xi, \eta \in \mathbb{R}^N$, $\alpha > 0$; se supone además que $u_0 \in L^2(\Omega)$ y $\varphi_0 \in L^2(H^1(\Omega)) \cap L^\infty(Q)$. Pero lo novedoso de este trabajo reside en que sólo se supone que $0 < \sigma(s) \leq \bar{\sigma}$ para cualquier $s \in \mathbb{R}$, dando lugar a problemas matemáticos de gran complejidad, ya que esta hipótesis deja abierta la posibilidad de que se tenga $\sigma(s) \to 0$ cuando $|s| \to \infty$. Así, si u no está acotada en Q, la ecuación elíptica del problema es degenerada en los puntos donde u es infinita, y no es posible obtener estimaciones *a priori* de $\nabla\varphi$. En consecuencia, el autor se ve obligado a buscar una solución en el espacio $L^2(H_0^1(\Omega)) \times L^p(Q)$, $p \geq 1$, lo cual puede provocar que $\nabla\varphi$ sea simplemente una distribución y el sistema de ecuaciones no pueda tratarse en el sentido de las distribuciones. Sin embargo, considerando $\Phi = \sigma(u)\nabla\varphi$ como una función de L^p, la multiplicación por distribuciones es posible. Para resolver esta situación, Xu introduce el concepto de solución de capacidad de dicho problema. Posteriormente, usó este tipo de solución para la resolución de un problema más general en [41], en el que se suponen hipótesis más débiles, por ejemplo, $\sigma(s) \to 0$ cuando $|s| \to \infty$. La volvió a considerar en [43] para la resolución del problema del termistor, tanto en el caso estacionario como en el de evolución, con $a \equiv 1$ y $\sigma \in C^0(\mathbb{R})$ satisfaciendo $\sigma(s) = 0$ para cualquier $s \geq L$, $L > 0$, y $0 < \sigma(s) < M$ para $-\infty < s < L$, $M > 0$.

En [12] se analiza el problema del termistor unidimensional con conductividad térmica degenerada. Esto incluye en particular el caso en el que $a(u)$

viene dada por la *ley de Wiedemann-Franz*, a saber, $a(u) = L\,u\,\sigma(u)$, donde $L > 0$ es una constante. Los autores demuestran existencia y unicidad de la solución débil, pero su técnica sólo se aplica al caso $N = 1$.

En [4] se aborda también el caso unidimensional, pero esta vez se analizan soluciones periódicas en tiempo en presencia de núcleos muertos para la temperatura.

1.3 El problema estacionario

También son muchos los autores que han investigado el caso estacionario, estableciendo numerosos resultados de existencia y unicidad de solución débil de (2). Entre ellos, Cimatti [13] dedujo un resultado de existencia y unicidad con las siguientes hipótesis: Ω es un abierto de $\mathbb{R}^N$ con frontera de clase C^2, $u_0, \varphi_0 \in C^2(\bar{\Omega})$, $u_0 > 0$ en Ω, $a \equiv 1$, $\sigma \in C^2(\mathbb{R}^+)$ y $0 < \sigma_1 \leq \sigma(s) \leq \sigma_2$ para cualquier $s \geq u_m = \min_{\partial\Omega} u$. Ésta última, aunque muy general, es la hipótesis fundamental del trabajo; no obstante, queda excluído el caso de conducción metálica, muy importante desde el punto de vista físico, en el que $\sigma(u) \sim O(u^{-1})$. Los resultados de [13] se basan en la *transformación de Diesselhorst*, a saber:

$$\xi = \frac{\varphi^2}{2} + \int_{u_m}^{u} \frac{a(s)}{\sigma(s)}\,\mathrm{d}s.$$

Se observa que ξ satisface la ecuación diferencial

$$\nabla \cdot (\sigma(u)\nabla\xi) = 0.$$

Así, si se puede probar que $\xi \in L^\infty(\Omega)$ y se supone además que $\int_{u_m}^{\infty} \frac{a(s)}{\sigma(s)}\,\mathrm{d}s = \infty$, se llega a la conclusión de que u y φ están acotadas, lo cual permite, incluso en el caso no uniformemente elíptico, deducir más regularidad para u y φ. En este contexto, es fácil comprender que la función $\gamma(u) = \int_{u_m}^{u} \frac{a(s)}{\sigma(s)}\,\mathrm{d}s$ desempeña un papel fundamental en el análisis teórico del problema del termistor cuando las conductividades térmica y eléctrica cumplen una hipótesis del tipo $a(s) > 0$ o $\sigma(s) > 0$ para $s \geq 0$.

Cimatti y Prodi [17] demuestran la existencia y unicidad de solución débil suponiendo que Ω es un abierto acotado y regular de $\mathbb{R}^N$ con $N \leq 3$, $a \equiv 1$, $u_0 \in H^2(\Omega)$ es una función armónica en Ω tal que $u_0 \geq u_m > 0$ sobre $\partial\Omega$, $\varphi_0 \in H^2(\Omega)$ con $\nabla\varphi_0 \in L^\infty(\Omega)$ y, además, $\sigma \in C^1(\mathbb{R}^+)$ es tal que $0 < \sigma(s) \leq \sigma_1$ y $0 < \sigma_0 \leq \sigma(s)s$ para cualquier $s \geq u_m$.

En [27] se estableció un resultado de existencia y unicidad para $N \geq 2$ y $a \equiv 1$ pero, en esta ocasión, se el problema considerado poseía condiciones de contorno mixtas; para demostrar la existencia se hizo uso del teorema del punto fijo de Schauder, mientras que para la unicidad se impusieron hipótesis de regularidad sobre los datos iniciales como, por ejemplo, que σ es globalmente Lipschitziana y $\|\nabla\varphi\|_{L^q(\Omega)} \leq C$, para q adecuado ($q > 2$ si $N = 2$ y $q = N$ si $N \geq 3$).

Xie [39] presenta dos resultados muy interesantes con $a \equiv 1$. En primer lugar, supone que $\sigma \in C^1[0, +\infty)$, que $0 < \sigma_0 \leq \sigma(s) \leq \sigma_1$ para cualquier

$s \geq 0$ y que existe una constante $L > 0$ tal que $|\sigma'(s)| \leq L$; también supone que $u_0, \varphi_0 \in W^{2-1/p,p}(\Omega)$, $p > N > 1$, con $u_0 \geq 0$ en Ω; con estas hipótesis demuestra un resultado de existencia. Para la unicidad, impone una restricción adicional sobre los datos iniciales. Seguidamente, proporciona otro resultado de existencia y unicidad con $0 < \sigma(s) \leq \sigma_1$ y $\sigma(s) \to 0$ cuando $s \to \infty$; esta hipótesis refleja un caso de especial interés práctico para algunos dispositivos eléctricos en los que $\sigma(s) \sim Ks^{-q}$ cuando $s \to \infty$ $(q > 0)$.

Asimismo, algunos autores han abordado el problema estacionario doblemente degenerado. Entre ellos, Allegretto y Xie [2] plantean dicho problema con unas condiciones de contorno mixtas que exigen regularidad tanto para u como para φ, siendo $\sigma, a \in C(\mathbb{R}^+)$ positivas y tales que

$$\int_{u^*}^{\infty} \frac{a(s)}{\sigma(s)} \mathrm{d}s = K < \infty,$$

donde $u^* = \max_{\Gamma_D} u_0(x)$ y $\Gamma_D \subset \partial\Omega$. Una de las hipótesis sobre la frontera de Ω es esencial para la demostración que se expone en ese trabajo, obteniéndose que $u, \varphi \in C^{\alpha}(\bar{\Omega})$ con $0 < \alpha < 1$. También tratan el caso en que

$$\int_{u^*}^{\infty} \frac{a(s)}{\sigma(s)} \mathrm{d}s = \infty.$$

Por otro lado, Cimatti [14] también analiza este problema en un abierto $\Omega \subset \mathbb{R}^3$ con frontera de clase C^2 tal que $\partial\Omega = \partial\Omega_1 \cup \partial\Omega_2 \cup \partial\Omega_3$, con $\partial\Omega_1 \cap \partial\Omega_2 = \emptyset$, $\mathrm{int}(\partial\Omega_i) \cap \mathrm{int}(\partial\Omega_j) = \emptyset$, $1 \leq i < j \leq 3$. Las condiciones de frontera que se imponen son: $u = u_0$ y $\varphi = \varphi_1$ sobre $\partial\Omega_1$, $u = u_0$ y $\varphi = \varphi_2$ sobre $\partial\Omega_2$, $\frac{\partial u}{\partial n} = 0$ y $\frac{\partial \varphi}{\partial n} = 0$ sobre $\partial\Omega_3$. Ésta última significa que el conductor está aislado sobre $\partial\Omega_3$ tanto eléctrica como térmicamente, mientras que $\partial\Omega_1$ y $\partial\Omega_2$ representan, respectivamente, los electrodos superior e inferior, a los cuales se les aplica la diferencia de potencial $\varphi_2 - \varphi_1$. Se supone que $u_0, \varphi_1, \varphi_2$ son constantes dadas tales que $\varphi_2 > \varphi_1$. Para lograr el resultado principal se supone que $\sigma, a \in C^2(\mathbb{R})$ son tales que $\sigma(s), a(s) > 0$, para cualquier $s \in \mathbb{R}$. Se busca una solución clásica, y para ello se distinguen los dos casos: $\int_{u_0}^{\infty} \frac{a(s)}{\sigma(s)} \mathrm{d}s = \alpha < \infty$ e $\int_{u_0}^{\infty} \frac{a(s)}{\sigma(s)} \mathrm{d}s = \infty$. Incluso en el caso en que el conductor obedezca la ley de Wiedemann–Franz, el problema posee una única solución.

De nuevo Cimatti [15] demuestra existencia y unicidad de solución clásica del mismo problema, pero esta vez debilita un poco las hipótesis sobre las conductividades: $\sigma, a \in C(\mathbb{R}^+)$, $\sigma(s), a(s) > 0$ para cualquier $s \geq 0$. La técnica que sigue en la demostración es análoga a la del trabajo anterior. Ping y Jishan lo abordan de nuevo en [34], con Ω un abierto acotado y regular de $\mathbb{R}^N$, $N \geq 1$, y tanto u_0 como φ_0 Hölderianas y acotadas; para obtener soluciones con regularidad $C^{\alpha}(\bar{\Omega})$, estos autores vuelven a usar la expresión $\int_{u_m}^{\infty} \frac{a(s)}{\sigma(s)} \mathrm{d}s$, donde $u_m = \inf_{\partial\Omega} u_0$.

1.4 Contribuciones de los autores

En los últimos años nos hemos centrado fundamentalmente en aquellos casos en que alguno o los dos coeficientes de difusión, $a(s)$ y $\sigma(s)$, no están acotados inferiormente por un valor estrictamente positivo o no están acotados superiormente. También hemos analizado el caso en el que el término de difusión de la ecuación parabólica es de la forma $-\nabla \cdot a(x,t,u,\nabla u)$, siendo a un operador de tipo Leray-Lions. Todo ello ha desembocado en una serie de problemas parabólicos-elípticos y elípticos degenerados de cierta complejidad desde el punto de vista matemático y que se describen más adelante.

Cuando la conductividad térmica es de tipo Wiedemann-Franz, esto es, $a(u) = L\,u\,\sigma(u)$, con L una constante positiva y $\sigma \in C(\mathbb{R})$, $a(u)$ puede anularse (al menos sobre $\partial\Omega$). Esto excluye por ejemplo el caso en que hay conducción metálica, para el cual se tiene $\sigma(u) = \sigma_0/u$. Deseamos hacer hincapié en el hecho de que, en los trabajos mencionados en las secciones anteriores y en otros muchos, la función a siempre es constante o acotada. Sin embargo, como puede comprobarse en [21] y [22], nuestro objetivo es la resolución, en el sentido débil, del problema del termistor cuando la conductividad térmica se anula en $u = 0$. Es justamente esta hipótesis la causante del principal obstáculo, puesto que la ecuación parabólica del sistema (1) se convierte en degenerada.

A continuación realizamos el estudio de la existencia de solución débil del problema estacionario del termistor bajo unas hipótesis que se corresponden con los casos físicamente importantes de conducción metálica junto a la ley de Wiedemann-Franz, las cuales se alejan sobremanera de las consideradas en los trabajos que hemos comentado, pues son mucho más débiles. Dichas hipótesis complican en gran medida la resolución del sistema (2), pues éste va estar constituido por dos ecuaciones elípticas degeneradas y acopladas, lo cual nos impedirá deducir la regularidad $\varphi \in H^1(\Omega) \cap L^\infty(\Omega)$ (véanse [21, 23, 24]). Más aún, llevamos a cabo un análisis del mismo problema suponiendo además que la conductividad térmica explota para un valor finito de la temperatura, obteniéndose así un sistema acoplado que, además de degenerado, es singular, y en el que sí se consigue una estimación L^∞ para la temperatura y, en consecuencia, se deduce más regularidad para φ y la misma u (véanse [21, 24]).

Asimismo hemos analizado en [21] un problema más general que se aborda partiendo de hipótesis más débiles que las habituales, más precisamente, suponiendo que la conductividad térmica a no está acotada inferiormente por una constante estrictamente positiva y considerando términos fuente en L^1. Por otro lado, presentamos el mismo problema en [25], pero en esta ocasión ninguno de los coeficientes de difusión está acotado superioremente. En ambas situaciones, se hace uso de la noción de solución renormalizada adaptada a ese contexto.

Por último, en [26] investigamos una generalización del problema del termistor y de un resultado de Xu [42], donde el término de divergencia de la ecuación parabólica es de la forma $-\nabla \cdot a(x,t,u,\nabla u)$ (a es un operador de tipo Leray-Lions) y la conductividad eléctrica σ no está acotada inferiormente por una constante estrictamente positiva (con lo que la ecuación elíptica

no es uniformente elíptica). En estas condiciones no es posible abordar el problema directamente en el marco de las soluciones débiles (en el sentido de las distribuciones); en su lugar, se introduce el concepto de solución de capacidad.

2 El problema evolutivo del termistor con conductividad térmica degenerada

En esta sección se demuestra la existencia de solución débil del problema de evolución del termistor con conductividad térmica degenerada (véanse [21, 22]), donde la ausencia de estimaciones espaciales para la temperatura y de estimaciones temporales para el potencial eléctrico son las principales dificultades.

Sea $a \in C(\mathbb{R})$ satisfaciendo la hipótesis (H.4) que se expone a continuación y definamos la función A, con

$$A(r) = \int_0^r a(s)\mathrm{d}s \quad \forall r \in \mathbb{R}.$$

Claramente, $A \in C^1(\mathbb{R})$ con $A(0) = 0$, es globalmente Lipschitziana y estrictamente creciente. Además, $\nabla \cdot A(\phi) = a(\phi)\nabla\phi$ para cualquier $\phi \in L^2(H^1(\Omega))$. En lugar del problema (1), considérese este otro:

$$\begin{cases} \dfrac{\partial u}{\partial t} - \Delta A(u) = \sigma(u)|\nabla\varphi|^2 & \text{en } Q, \\ \nabla \cdot (\sigma(u)\nabla\varphi) = 0 & \text{en } Q, \\ u = 0 & \text{sobre } \partial\Omega \times (0,T), \\ \varphi = \varphi_0 & \text{sobre } \partial\Omega \times (0,T), \\ u(\cdot,0) = u_0 & \text{en } \Omega \end{cases} \tag{4}$$

y supongamos que se verifican las siguientes hipótesis:

(H.1) $u_0 \in L^2(\Omega)$ y $u_0 \geq 0$ c.p.d. en Ω.

(H.2) $\varphi_0 \in L^\infty(H^1(\Omega)) \cap L^\infty(Q)$.

(H.3) $\sigma \in C(\mathbb{R})$ y $0 < \sigma_1 \leq \sigma(s) \leq \sigma_2$ para cualquier $s \in \mathbb{R}$.

(H.4) $a \in C(\mathbb{R})$ y $0 < a(s) \leq \bar{a}$ para cualquier $s \neq 0$, $a(0) = 0$.

(H.5) Para cada $\delta > 0$ existe una constante $\alpha_\delta > 0$ tal que $\inf_{|s|>\delta} a(s) \geq \alpha_\delta$.

Definición 1 *Diremos que un par (u, φ) es solución débil de (4) si*

$$u \in L^1(Q) \quad \frac{\mathrm{d}u}{\mathrm{d}t} \in L^2(H^{-1}(\Omega)), \quad A(u) \in L^2(H_0^1(\Omega)),$$

$$\varphi - \varphi_0 \in L^\infty(H_0^1(\Omega)) \cap L^\infty(Q), \quad u(\cdot,0) = u_0 \quad en \quad \Omega$$

y, para cualquier $\phi \in L^2(H_0^1(\Omega))$,

$$\int_0^t \langle \frac{\mathrm{d}u}{\mathrm{d}t}, \phi \rangle + \int_0^t \int_\Omega \nabla A(u) \nabla \phi = \int_0^t \int_\Omega \sigma(u) |\nabla \varphi|^2 \phi \quad \forall t \in [0,T],$$
$$\int_\Omega \sigma(u) \nabla \varphi \cdot \nabla \phi = 0, \ \ c.p.d.\ t \in (0,T).$$

El resultado principal de esta sección es el

Teorema 1 *Bajo las hipótesis* (H.1)–(H.5)*, el sistema* (4) *posee solución débil.*

Las secciones que siguen desarrollan la demostración del teorema 1.

2.1 Problemas aproximados

Para cada $n \geq 1$, se definen las funciones $a_n(s) = a(s) + \frac{1}{n} \leq \bar{a} + 1 = \tilde{a}$ y $A_n(r) = \int_0^r a_n(s)\mathrm{d}s = A(r) + \frac{r}{n}$ y se plantea el problema aproximado

$$\begin{cases} \dfrac{\partial u_n}{\partial t} - \nabla \cdot (a_n(u_n) \nabla u_n) = \sigma(u_n) |\nabla \varphi_n|^2 & \text{en } Q, \\ \nabla \cdot (\sigma(u_n) \nabla \varphi_n) = 0 & \text{en } Q, \\ u_n = 0 & \text{sobre } \partial\Omega \times (0,T), \\ \varphi_n = \varphi_0 & \text{sobre } \partial\Omega \times (0,T), \\ u_n(\cdot, 0) = u_0 & \text{en } \Omega. \end{cases} \tag{5}$$

Los resultados clásicos de existencia garantizan que el sistema (5) posee solución débil (u_n, φ_n) tal que

$$u_n \in L^2(H_0^1(\Omega)) \cap C^0([0,T]; L^2(\Omega)), \quad \frac{\mathrm{d}u_n}{\mathrm{d}t} \in L^2(H^{-1}(\Omega)),$$

$$\varphi_n \in L^\infty(H^1(\Omega)) \cap L^\infty(Q)$$

y

$$\int_\Omega |\nabla \varphi_n|^2 \leq C(\sigma_1, \sigma_2, \varphi_0) = C \quad t \in (0,T) \text{ c.p.d.}, \tag{6}$$

$$\|\varphi_n\|_{L^\infty(Q)} \leq \|\varphi_0\|_{L^\infty(Q)}. \tag{7}$$

Es más, gracias a (H.1), se puede probar que $u_n \geq 0$ casi por doquier en Q; véase [3].

2.2 Estimaciones

Usando la ecuación elíptica de (5), es fácil ver que

$$\int_Q \sigma(u_n) |\nabla \varphi_n|^2 \phi = -\int_Q \sigma(u_n) \varphi_n \nabla \varphi_n \cdot \nabla \phi \quad \forall\, \phi \in L^2(H_0^1(\Omega)), \tag{8}$$

donde $(\nabla\cdot(\sigma(u_n)\varphi_n\nabla\varphi_n))\subset L^2(H^{-1}(\Omega))$ está acotada gracias a (H.3), (6) y (7). Tomemos ahora $A_n(u_n)\in L^2\left(0,T;H_0^1(\Omega)\right)$ como función de test y definamos

$$\tilde{A}_n(r)=\int_0^r A_n(s)\,\mathrm{d}s\geq 0\quad\forall r\in\mathbb{R};$$

como $\int_\Omega \tilde{A}_n(u_0)\leq\frac{\tilde{a}}{2}\|u_0\|_{L^2(\Omega)}$, en vista de (8) y de la desigualdad de Young, obtenemos:

$$\int_0^t\int_\Omega|\nabla A_n(u_n)|^2\leq C(\tilde{a},u_0,\sigma_1,\sigma_2,\varphi_0,T)=C\ \text{ c.p.d. en }\ (0,T).\tag{9}$$

De (6) y (9) se deduce que $(\tilde{A}_n(u_n))$ está acotada en $L^\infty(L^1(\Omega))$. Más aún, $\left(\frac{\mathrm{d}u_n}{\mathrm{d}t}\right)$ está acotada en $L^2(H^{-1}(\Omega))$ pues $(\Delta A_n(u_n))$ y $(\nabla\cdot(\sigma(u_n)\varphi_n\nabla\varphi_n))$ lo están en el mismo espacio.

Veamos ahora que

$$(u_n)\text{ está acotada en }L^\infty(L^1(\Omega)).\tag{10}$$

Efectivamente, si se considera la función de test $\eta_\varepsilon(u_n)=\frac{1}{\varepsilon}T_\varepsilon(u_n)$ con $\varepsilon>0$ y se define la función $\Phi_\varepsilon(r)=\int_0^r\eta_\varepsilon(s)\,\mathrm{d}s$, gracias a que $|\eta_\varepsilon(r)|\leq 1$ para cualquier $s\in\mathbb{R}$, tendremos que $\int_\Omega\Phi_\varepsilon(u_n(t))\leq\int_\Omega\Phi_\varepsilon(u_0)+\sigma_2TC$ c.p.d. en $(0,T)$ y, haciendo $\varepsilon\to 0$, obtendremos (10).

Para cada $\delta>0$, se define la función g_δ como sigue: $g_\delta(s)=s+\delta$ si $s<-\delta$, $g_\delta(s)=0$ si $|s|\leq\delta$ y $g_\delta(s)=s-\delta$ si $s>\delta$. Tomemos $g_\delta(u_n)$ como función

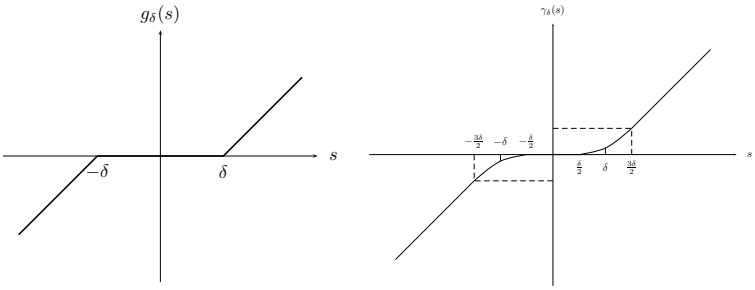


Figura 2: Funciones $g_\delta(s)$ y $\gamma_\delta(s)$.

de test y definamos $G_\delta(r)=\int_0^r g_\delta(s)\,\mathrm{d}s\geq 0$ para cualquier $r\in\mathbb{R}$. Como $\nabla g_\delta(u_n)=\nabla u_n\chi_{\{|u_n|>\delta\}}$, gracias a (H.5) se deduce que

$$\begin{aligned}\alpha_\delta\int_Q|\nabla g_\delta(u_n)|^2&=\alpha_\delta\int_{\{|u_n|>\delta\}}|\nabla u_n|^2\\&\leq\int_\Omega G_\delta(u_0)+\sigma_2\|\varphi_0\|\int_Q\nabla\varphi_n\cdot\nabla g_\delta(u_n).\end{aligned}$$

Para δ suficientemente pequeño $(0 < \delta < 1)$, se tiene que

$$\int_\Omega G_\delta(u_0) \leq \frac{1}{2}\|u_0\|_{L^2(\Omega)};$$

así, aplicando la desigualdad de Young y usando (6), $(g_\delta(u_n))$ está acotada en $L^2(H_0^1(\Omega))$ para cada $\delta > 0$, probándose también que $(G_\delta(u_n))$ está acotada en $L^\infty(L^1(\Omega))$ para cada $\delta > 0$.

Se define ahora una función regularizada de g_δ, a saber γ_δ, tal que $\gamma_\delta \in C^\infty(\mathbb{R})$. Sea $\gamma \in C^\infty(\mathbb{R})$ verificando: $\gamma(s) = 0$ si $s \in [0, \frac{1}{2}]$; $\gamma(s) = s - 1$ si $s > \frac{3}{2}$; $\gamma(s)$ es convexa y $\gamma(s)/s$ es creciente en $s \in [\frac{1}{2}, \frac{3}{2}]$; $\gamma(-s) = -\gamma(s)$ para cualquier $s \in \mathbb{R}$. Entonces hacemos $\gamma_\delta(s) = \delta\gamma\left(\frac{s}{\delta}\right)$. Claramente, la definición de γ_δ implica que $0 \leq \gamma_\delta(s) \leq \gamma_{\delta'}(s)$ para cualesquiera $s \in \mathbb{R}$ y $\delta' \leq \delta$; además $|\gamma_\delta'(s)| \leq 1$, $|\gamma_\delta''(s)| \leq \frac{K}{\delta}$, para cualquier $s \in \mathbb{R}$.

Fijado un $\delta > 0$,

$$\langle \frac{\mathrm{d}\gamma_\delta(u_n)}{\mathrm{d}t}, \phi\rangle = \langle \frac{\mathrm{d}u_n}{\mathrm{d}t}, \gamma_\delta'(u_n)\phi\rangle \quad \forall\, \phi \in \mathcal{D}(Q),$$

de modo que, como $\nabla\left(\gamma_\delta'(u_n)\phi\right) = \gamma_\delta''(u_n)\nabla u_n\phi + \gamma_\delta'(u_n)\nabla\phi$, se tiene que

$$\begin{aligned}\frac{\mathrm{d}\gamma_\delta(u_n)}{\mathrm{d}t} &= -a_n(u_n)\gamma_\delta''(u_n)|\nabla u_n|^2 + \nabla\cdot(a_n(u_n)\gamma_\delta'(u_n)\nabla u_n)\\ &\quad -\sigma(u_n)\varphi_n\nabla\varphi_n\cdot(\gamma_\delta''(u_n)\nabla u_n) + \nabla\cdot(\sigma(u_n)\varphi_n\gamma_\delta'(u_n)\nabla\varphi_n)\\ &= I_{1n} + I_{2n} + I_{3n} + I_{4n} \qquad \text{en } \mathcal{D}'(Q).\end{aligned}$$

Se puede probar que I_{1n} y I_{3n} están acotadas en $L^1(Q)$ mientras que I_{2n} y I_{4n} lo están en $L^2(H^{-1}(\Omega))$, deduciéndose entonces que

$$\left(\frac{\mathrm{d}\gamma_\delta(u_n)}{\mathrm{d}t}\right) \text{ está acotada en } L^1(Q) + L^2(H^{-1}(\Omega)), \text{ para cada } \delta > 0. \tag{11}$$

Más aún, de las propiedades de γ_δ se tiene que

$$(\gamma_\delta(u_n)) \text{ está acotada en } L^2(H_0^1(\Omega)), \text{ para cada } \delta > 0. \tag{12}$$

Por otro lado, gracias a (10) y a la definición de γ_δ, se sigue que

$$\int_\Omega |\gamma_\delta(u_n)| \leq \int_\Omega |u_n| \leq C \ \text{ c.p.d. en } \ (0,T),\ \forall\, n \geq 1,\ \forall\, \delta > 0. \tag{13}$$

2.3 Paso al límite

Haremos uso del siguiente resultado de compacidad (véase [37]):

Lema 2 *Sean X, B e Y tres espacios de Banach tales que $X \hookrightarrow B \hookrightarrow Y$, siendo todas las inyecciones continuas y la primera de ellas compacta. Para $1 \leq p, q < +\infty$, sea $\mathbf{W}$ el espacio de Banach*

$$\mathbf{W} = \{\, v \in L^p(X) : \frac{\mathrm{d}v}{\mathrm{d}t} \in L^q(Y)\,\}.$$

Entonces la inyección $\mathbf{W} \hookrightarrow L^p(B)$ es compacta.

La inyección $L^1(\Omega)+H^{-1}(\Omega) \hookrightarrow W^{-1,q'}(\Omega)$, $q' < \frac{N}{N-1}$ es continua; tomando $X = H_0^1(\Omega) \hookrightarrow B = L^2(\Omega)$ con inyección compacta e $Y = W^{-1,q'}(\Omega)$ en el lema 2, el correspondiente espacio $\mathbf{W} = \{\, v \in L^2(H_0^1(\Omega)) : \frac{\mathrm{d}v}{\mathrm{d}t} \in L^1(W^{-1,q'}(\Omega)) \,\}$ es tal que $\mathbf{W} \hookrightarrow L^2(Q)$ con inyección compacta. Por lo tanto, en virtud de (11) y (12), para cada $\delta > 0$, $(\gamma_\delta(u_n))_n$ es relativamente compacta en $L^2(Q)$, de manera que, para cada $\delta > 0$, existe una función $z_\delta \in L^2(Q)$ tal que, para una subsucesión, $\gamma_\delta(u_n) \to z_\delta$ casi por doquier en Q y

$$\gamma_\delta(u_n) \to z_\delta \text{ fuerte en } L^2(Q). \tag{14}$$

Además, (13) implica que (z_δ) está acotada en $L^\infty(L^1(\Omega))$. Como γ_δ crece si δ decrece, se deduce también que (z_δ) es una sucesión creciente cuando $\delta \downarrow 0$, así que existe una función medible $u : Q \to \mathbb{R}$ tal que $\lim_{\delta\downarrow 0} z_\delta = u$; de este modo, aplicando el teorema de la convergencia monótona, $u \in L^1(Q)$ y $z_\delta \to u$ fuerte en $L^1(Q)$ y c.p.d. para una subsucesión.

Como

$$u_n \to u \ \text{ c.p.d. en } \ Q, \tag{15}$$

$\gamma_\delta(u_n) \to \gamma_\delta(u)$ y, por tanto, debe ser $\gamma_\delta(u) = z_\delta$. Más aún, se prueba que

$$u_n \to u \ \text{ fuertemente en } \ L^1(Q), \tag{16}$$

teniendo en cuenta que, para un $\delta > 0$ fijo, $|\gamma_\delta(s) - s| \le \delta$ para cualquier $s \in \mathbb{R}$ y que de (14) se obtiene la convergencia $\gamma_\delta(u_n) \to z_\delta = \gamma_\delta(u)$ fuerte en $L^1(Q)$.

De (16) se deduce fácilmente que $\frac{\mathrm{d}u_n}{\mathrm{d}t} \rightharpoonup \frac{\mathrm{d}u}{\mathrm{d}t}$ débil en $L^2(H^1(\Omega))$. Por otro lado, $(\sigma(u_n))$ está acotada en $L^\infty(Q)$ y $\sigma \in C(\mathbb{R})$, así que de (15) se deduce, para una subsucesión, que $\sigma(u_n) \to \sigma(u)$ c.p.d. en Q y

$$\sigma(u_n) \rightharpoonup \sigma(u) \ \text{ débilmente-}* \text{ en } \ L^\infty(Q). \tag{17}$$

Gracias a (6) y (7), existen una subsucesión (denotada de igual modo) y una función $\varphi \in L^\infty(H^1(\Omega)) \cap L^\infty(Q)$ tales que $\varphi_n \rightharpoonup \varphi$ débilmente-$*$ en $L^\infty(Q)$ y

$$\varphi_n \rightharpoonup \varphi \ \text{ débilmente-}* \text{ en } \ L^\infty(H^1(\Omega)). \tag{18}$$

Por otro lado,

$$0 \le \sigma_1 \int_Q |\nabla(\varphi_n - \varphi)|^2 \le -\int_Q \sigma(u_n)\nabla\varphi\nabla(\varphi_n - \varphi) = I_n.$$

La convergencia (17) implica que $\sigma(u_n)\nabla\varphi \to \sigma(u)\nabla\varphi$ fuertemente en $L^2(Q)^N$; consecuentemente, a partir de (18), $I_n \to 0$, de donde

$$\nabla\varphi_n \to \nabla\varphi \ \text{ fuertemente en } \ L^2(Q)^N, \tag{19}$$

obteniéndose así que $\varphi_n \to \varphi$ fuertemente en $L^2(H^1(\Omega))$ y c.p.d. en Q. De (17) y (19) concluímos también que $\sigma(u_n)|\nabla\varphi_n|^2 \to \sigma(u)|\nabla\varphi|^2$ fuertemente en $L^1(Q)$.

Por último, (9) y (15) nos conducen a las convergencia $A_n(u_n) \rightharpoonup A(u)$ débil en $L^2(H_0^1(\Omega))$ y c.p.d. en Q, quedando probado el teorema 1.

3 Sistemas elípticos doblemente degenerados

Se considera ahora el problema del termistor con segundo miembro en forma de divergencia

$$\begin{cases} -\nabla\cdot(a(u)\nabla u) = \nabla\cdot(\sigma(u)\varphi\nabla\varphi) & \text{en } \Omega, \\ \nabla\cdot(\sigma(u)\nabla\varphi) = 0 & \text{en } \Omega, \\ u = 0 & \text{sobre } \partial\Omega, \\ \varphi = \varphi_0 & \text{sobre } \partial\Omega, \end{cases} \tag{20}$$

bajo las hipótesis que siguen:

(H.1) $\sigma \in C(\mathbb{R})$ es tal que $0 < \sigma(s) \le \bar{\sigma}$ para cualquier $s \in \mathbb{R}$.

(H.2) $a \in C(\mathbb{R}) \cap L^\infty(\mathbb{R})$ es tal que $\int_0^{+\infty} a(s)\mathrm{d}s = +\infty$ y $A(r) = \int_0^r a(s)\mathrm{d}s$ es estrictamente creciente.

(H.3) $\varphi_0 \in H^1(\Omega)$.

(H.4) Existen un entero $M > 1$ y una función $\alpha : [M, +\infty) \to \mathbb{R}$ tales que $\alpha(s) > 0$ para todo $s \ge M$, α es decreciente y $\sigma(s) \ge \alpha(s) > 0$.

(H.5) Si $2N/N + 2 < p < 2$ para $N \ge 2$, $1 < p < 2$ para $N = 1$ y $p' = 2 - p$, entonces

$$\int_{M-1}^{+\infty} \frac{\mathrm{ds}}{\alpha(s+1)^{p/p'} A(s)^{\bar{q}/2}} < +\infty, \quad \text{con} \quad \begin{cases} \bar{q} = 2^* & \text{si } N \ge 3, \\ \bar{q} \in [2, +\infty) & \text{si } N = 2, \\ \bar{q} \in [1, +\infty) & \text{si } N = 1. \end{cases} \tag{21}$$

Nota 1 Según las hipótesis (H.1) y (H.2), ninguno de los coeficientes de difusión es uniformemente elíptico, con lo que el problema (20) constituye un sistema elíptico doblemente degenerado. Es más, obsérvese que sólo se supone que φ_0 pertenece a $H^1(\Omega)$ y no a $L^\infty(\Omega)$; esto añade una dificultad adicional al análisis de (20). Por último, (H.5) es una condición técnica que permite cerrar el problema: se trata de una restricción que relaciona a los coeficientes de difusión.

En los trabajos mencionados en la Introducción se suponen hipótesis de alta regularidad, bien sobre $\partial\Omega$, bien sobre los datos iniciales, lográndose así la existencia de soluciones clásicas. La situación que planteamos en [21], [23] y [24] es bien distinta: la escasa regularidad impuesta sólo conduce a la existencia de soluciones débiles. Si la conductividad eléctrica, σ, y φ_0 fuesen suficientemente regulares, entonces φ sería muy regular y podríamos deducir que $u \in L^\infty(\Omega)$. En tal caso, $\sigma(u)$ ya no sería degenerada, con lo que vol,veríamos a obtener una función φ muy regular y este ciclo nunca se cerraría. Sin embargo, éste no es el caso que abordamos ahora. Al contrario; tanto a como σ van a ser degeneradas y tendremos que recurrir a otras técnicas para la resolución de (20).

Se tiene el siguiente resultado de existencia:

Teorema 3 *Bajo las hipótesis* (H.1)–(H.5)*, el problema*

$$\begin{cases} -\Delta A(u) = \nabla\cdot(\sigma(u)\varphi\nabla\varphi) & \text{en } \mathcal{D}'(\Omega), \\ \nabla\cdot(\sigma(u)\nabla\varphi) = 0 & \text{en } \Omega, \\ u = 0 & \text{sobre } \partial\Omega, \\ \varphi = \varphi_0 & \text{sobre } \partial\Omega, \end{cases}$$

admite solución débil (u,φ) *en el sentido siguiente:*

$$A(u)\in W_0^{1,q}(\Omega),\quad \varphi-\varphi_0\in W_0^{1,p}(\Omega),\quad \sigma(u)^{1/2}\nabla\varphi\in L^2(\Omega),$$
$$\int_\Omega \nabla A(u)\nabla\xi = -\int_\Omega \sigma(u)\varphi\nabla\varphi\nabla\xi,\ \forall\,\xi\in\mathcal{D}(\Omega),$$
$$\int_\Omega \sigma(u)\nabla\varphi\nabla\phi = 0,\ \forall\,\phi\in H_0^1(\Omega),$$

donde $q<\frac{N}{N-1}$ *si* $N\geq 2$ *y* $q=2$ *si* $N=1$*. Además, el término* $\nabla\cdot(\sigma(u)\varphi\nabla\varphi)$ *es una medida de Radon positiva y* $u\geq 0$ *c.p.d. en* Ω*.*

Demostración. Para cada $n\geq 1$, se definen las funciones $a_n(s)=a(s)+\frac{1}{n}$ y $\sigma_n(s)=\sigma(s)+\frac{1}{n}\leq 1+\bar\sigma=\tilde\sigma$ y se plantea el problema aproximado

$$\begin{cases} -\nabla\cdot(a_n(u_n)\nabla u_n) = \sigma_n(u_n)|\nabla\varphi_n|^2 & \text{en } \Omega, \\ \nabla\cdot(\sigma_n(u_n)\nabla\varphi_n) = 0 & \text{en } \Omega, \\ u_n = 0 & \text{sobre } \partial\Omega, \\ \varphi_n = T_n(\varphi_0) & \text{sobre } \partial\Omega. \end{cases} \tag{22}$$

En vista de [3], sabemos que el sistema (22) admite una solución débil tal que $u_n\in H_0^1(\Omega)$ y $\varphi_n - T_n(\varphi_0)\in H_0^1(\Omega)\cap L^\infty(\Omega)$. Es más, como

$$\int_\Omega \sigma_n(u_n)\nabla\varphi_n\cdot\nabla\phi = 0,\ \forall\,\phi\in H_0^1(\Omega), \tag{23}$$

haciendo $\phi=\varphi_n - T_n(\varphi_0)$ se tiene que

$$\int_\Omega \sigma_n(u_n)|\nabla\varphi_n|^2 \leq \tilde\sigma\int_\Omega |\nabla\varphi_0|^2 \leq \tilde\sigma\|\varphi_0\|^2_{H^1(\Omega)} = C, \tag{24}$$

es decir, $(f_n)=\big(\sigma_n(u_n)|\nabla\varphi_n|^2\big)$ está acotada en $L^1(\Omega)$.

Sea $v_n=A_n(u_n)$ y considérese el problema elíptico

$$\begin{cases} -\Delta v_n = f_n & \text{en } \Omega, \\ v_n = 0 & \text{sobre } \partial\Omega. \end{cases}$$

De las estimaciones de Boccardo-Gallouët, se deduce que

$$(v_n) \text{ está acotada en } W_0^{1,q}(\Omega),\ \forall\, q<\frac{N}{N-1} \text{ si } N\geq 2,\ q=2 \text{ si } N=1 \tag{25}$$

(véanse [9, 20]). De este modo, existen una subsucesión (denotada de igual modo) y una función $v\in W_0^{1,q}(\Omega)$ tales que $v_n \rightharpoonup v$ débil en $W_0^{1,q}(\Omega)$.

Las inyecciones $W_0^{1,q}(\Omega) \hookrightarrow L^r(\Omega)$ para cualquier $r < \frac{N}{N-2}$ si $N \geq 2$ y $W_0^{1,q}(\Omega) = H_0^1(\Omega) \hookrightarrow C(\bar{\Omega})$ si $N = 1$, son compactas, con lo cual podemos suponer también que

$$v_n \to v \text{ fuerte en } L^r(\Omega), \text{ si } N \geq 2, \tag{26}$$

$$v_n \to v \text{ fuerte en } C(\bar{\Omega}), \text{ si } N = 1, \tag{27}$$

$$v_n \to v \text{ c.p.d. en } \Omega. \tag{28}$$

Además, como $f_n \geq 0$ en Ω, asimismo serán $v_n \geq 0$ y $u_n \geq 0$ en Ω.

Usando (25), se prueba que $(A(u_n)) \subset H_0^1(\Omega)$ está acotada en $W_0^{1,q}(\Omega)$. Por tanto, para alguna subsucesión, existe una función $z \in W_0^{1,q}(\Omega)$ tal que $A(u_n) \rightharpoonup z$ débilmente en $W_0^{1,q}(\Omega)$, $A(u_n) \to z$ fuertemente en $L^r(\Omega)$ para cualquier $r < \frac{N}{N-2}$ si $N \geq 2$, $A(u_n) \to z$ fuertemente en $C(\bar{\Omega})$ si $N = 1$ y $A(u_n) \to z$ c.p.d. en Ω. Como A es una función biyectiva,

$$u_n \to A^{-1}(z) = u \text{ c.p.d. en } \Omega, \tag{29}$$

siendo $u \geq 0$ c.p.d.

De la definición de σ_n y (29) se deduce claramente que, para una subsucesión,

$$\sigma_n(u_n) \to \sigma(u) \text{ c.p.d. en } \Omega. \tag{30}$$

Además, en virtud de la hipótesis (H.1), $(\sigma_n(u_n))$ está acotada en $L^\infty(\Omega)$, por lo que teniendo en cuenta (30),

$$\sigma_n(u_n) \rightharpoonup \sigma(u) \text{ débilmente-}* \text{ en } L^\infty(\Omega). \tag{31}$$

Busquemos ahora una estimación de (φ_n) en un espacio $W^{1,p}(\Omega)$, con $1 < p < 2$. Gracias a (H.5), $2/p'$ es el exponente conjugado de $2/p$. Aplicando la desigualdad de Young y teniendo en cuenta (24), obtenemos

$$\int_\Omega |\nabla \varphi_n|^p = \int_\Omega \sigma_n(u_n)^{-p/2} \sigma_n(u_n)^{p/2} |\nabla \varphi_n|^p \leq C^{p/2} \left(\int_\Omega \sigma_n(u_n)^{-p/p'} \right)^{p'/2}.$$

Veamos ahora que

$$\int_\Omega \sigma_n(u_n)^{-p/p'} \leq C. \tag{32}$$

De la definición de σ_n es inmediato que $\tilde{\sigma}^{-p/p'} \leq \sigma_n(s)^{-p/p'} \leq \sigma(s)^{-p/p'}$ para cualquier $s \in \mathbb{R}$ y entonces

$$\int_\Omega \sigma_n(u_n)^{-p/p'} \leq \int_\Omega \sigma(u_n)^{-p/p'} \leq \int_{\{|u_n| \leq M\}} \sigma(u_n)^{-p/p'} + \int_{\{u_n > M\}} \sigma(u_n)^{-p/p'}.$$

Gracias a (H.1), σ está acotada en cada compacto de $\mathbb{R}$; existe pues una constante $C_M > 0$ tal que $\min_{|s| \leq M} \sigma(s) = C_M$, de donde

$$\int_{\{|u_n| \leq M\}} \sigma(u_n)^{-p/p'} \leq C_M^{-p/p'} \operatorname{med} \Omega = C.$$

Por otro lado, de la hipótesis (H.4) se tiene que

$$\int_{\{u_n>M\}} \sigma(u_n)^{-p/p'} \le \int_{\{u_n>M\}} \alpha(u_n)^{-p/p'} \le \sum_{i\ge M} \int_{\{i\le u_n<i+1\}} \alpha(u_n)^{-p/p'}$$
$$\le \sum_{i\ge M} \int_{\{i\le u_n<i+1\}} \alpha(i+1)^{-p/p'} \le \sum_{i\ge M} \alpha(i+1)^{-p/p'} \mathrm{med}\{u_n \ge i\}.$$

Para deducir una estimación de $\mathrm{med}\{u_n \ge i\}$, tomemos como función de test $T_i(v_n)$ en la ecuación para u_n:

$$\int_\Omega \nabla v_n \nabla T_i(v_n) = \int_\Omega \sigma_n(u_n)|\nabla\varphi_n|^2 T_i(v_n) \le C\, i,$$

siendo también $\int_\Omega \nabla v_n \nabla T_i(v_n) = \int_\Omega |\nabla T_i(v_n)|^2 = I_{i,n}$. Teniendo en cuenta la desigualdad de Sobolev,

$$I_{i,n} \ge C\left(\int_\Omega |T_i(v_n)|^{\bar q}\right)^{2/\bar q} \ge C\left(\int_{\{v_n\ge i\}} |T_i(v_n)|^{\bar q}\right)^{2/\bar q} = Ci^2 \mathrm{med}\{v_n \ge i\}^{2/\bar q},$$

donde $\bar q = 2^*$ y $C = C(\Omega, N)$ si $N \ge 3$, $\bar q \in [2, +\infty)$ y $C = C(\Omega, \bar q)$ si $N \le 2$. En consecuencia, $\mathrm{med}\{v_n \ge i\}^{2/\bar q} \le C/i^{\bar q/2}$. Como $u_n \ge 0$ en Ω, $A_n(u_n) \ge A(u_n)$ en Ω y $\mathrm{med}\,\{A(u_n) \ge i\} \le \mathrm{med}\{v_n \ge i\} \le C/i^{\bar q/2}$, entonces $\mathrm{med}\{u_n \ge l\} \le C/A(l)^{\bar q/2}$. Por consiguiente, gracias a (21),

$$\int_{\{u_n>M\}} \sigma(u_n)^{-p/p'} \le C \int_{M-1}^{+\infty} \frac{\mathrm{ds}}{\alpha(s+1)^{p/p'} A(s)^{\bar q/2}} = C.$$

De este modo queda probado (32) y se deduce que $(\varphi_n - T_n(\varphi_0))$ está acotada en $W_0^{1,p}(\Omega)$. Entonces existe una función $\varphi \in W^{1,p}(\Omega)$ tal que

$$\varphi_n \rightharpoonup \varphi \ \text{ débilmente en } \ W^{1,p}(\Omega), \tag{33}$$

$\varphi_n \to \varphi$ fuertemente en $L^{\bar r}(\Omega)$ con $\bar r < p^*$ si $N \ge 2$, $\varphi_n \to \varphi$ fuertemente en $C(\bar\Omega)$ si $N = 1$ y $\varphi_n \to \varphi$ c.p.d. en Ω, todo ello para una subsucesión. En particular,

$$\varphi_n \to \varphi \ \text{ fuertemente en } \ L^2(\Omega). \tag{34}$$

De (24) se colige que $\left(\sigma_n(u_n)^{1/2}\nabla\varphi_n\right)$ está acotada en $L^2(\Omega)^N$; esto implica, en virtud de (31) y (33), que

$$\sigma_n(u_n)^{1/2}\nabla\varphi_n \rightharpoonup \sigma(u)\nabla\varphi_n \text{ débilmente en } L^2(\Omega)^N. \tag{35}$$

Es más, $\sigma_n(u_n)\nabla\varphi_n \rightharpoonup \sigma(u)\nabla\varphi$ débilmente en $L^2(\Omega)^N$ si se tienen en cuenta (H.1), (31) y (35). Consecuentemente, $\nabla\cdot(\sigma(u)\nabla\varphi) \in H^{-1}(\Omega)$ y

$$\langle \nabla\cdot(\sigma(u)\nabla\varphi), \phi\rangle = -\int_\Omega \sigma(u)\nabla\varphi\cdot\nabla\phi = 0 \quad \forall\, \phi \in H_0^1(\Omega),$$

ya que, gracias a la segunda ecuación del problema (22), se concluye (23).

Retomemos la expresión (23) con $\phi = \varphi_n \xi$, siendo $\xi \in \mathcal{D}(\Omega)$. Entonces

$$\begin{aligned} 0 &= \int_\Omega \sigma_n(u_n)\nabla\varphi_n \cdot \nabla(\varphi_n\xi) = \int_\Omega \sigma_n(u_n)|\nabla\varphi_n|^2\xi + \int_\Omega \sigma_n(u_n)\varphi_n\nabla\varphi_n \cdot \nabla\xi \\ &= \int_\Omega \sigma_n(u_n)|\nabla\varphi_n|^2\xi - \int_\Omega \nabla \cdot (\sigma_n(u_n)\varphi_n\nabla\varphi_n)\,\xi, \end{aligned}$$

o sea, $\sigma_n(u_n)|\nabla\varphi_n|^2 = \nabla \cdot (\sigma_n(u_n)\varphi_n\nabla\varphi_n)$ en $\mathcal{D}'(\Omega)$. Ahora bien,

$$\int_\Omega \sigma_n(u_n)\varphi_n\nabla\varphi_n \cdot \nabla\xi = \int_\Omega \sigma_n(u_n)^{1/2}\varphi_n\sigma_n(u_n)^{1/2}\nabla\varphi_n \cdot \nabla\xi$$

para todo $\xi \in \mathcal{D}(\Omega)$. Gracias a las convergencias (30), (34) y (35), haciendo $m \to \infty$, se obtiene que

$$\int_\Omega \sigma(u)^{1/2}\varphi\sigma(u)^{1/2}\nabla\varphi \cdot \nabla\xi = \int_\Omega \sigma(u)\varphi\nabla\varphi \cdot \nabla\xi \quad \forall\, \xi \in \mathcal{D}(\Omega),$$

es decir, $\sigma_n(u_n)|\nabla\varphi_n|^2 = \nabla \cdot (\sigma_n(u_n)\varphi_n\nabla\varphi_n) \to \nabla \cdot (\sigma(u)\varphi\nabla\varphi)$ en $\mathcal{D}'(\Omega)$. Al ser $\sigma_n(u_n)|\nabla\varphi_n|^2 \geq 0$ una sucesión acotada en $L^1(\Omega)$, concluímos también que $\nabla \cdot (\sigma(u)\varphi\nabla\varphi)$ es una medida de Radon positiva. □

Nota 2 ¿ Cuándo se puede asegurar que se cumple la igualdad $\sigma(u)|\nabla\varphi|^2 = \nabla \cdot (\sigma(u)\varphi\nabla\varphi)$? Conocemos algunas situaciones en las que la respuesta es afirmativa; por ejemplo, si $N = 1$. También es cierta en el caso regular, es decir, cuando $\varphi \in H^1(\Omega)$. En el caso no regular, en virtud del teorema 3, tenemos que $\varphi \in W^{1,p}(\Omega)$ y $\sigma(u)^{-1} \in L^{1/(r-1)}(\Omega)$, con $r = 2/p$, y la igualdad se cumple:

- Si $\sigma(u)$ es un peso de tipo Muckenhoupt [38], a saber, si existe una constante C tal que, para cualquier $x \in \mathbb{R}^N$,

$$\left(\frac{1}{|B_R(x)|}\int_{B_R(x)} \sigma(u)\right)\left(\frac{1}{|B_R(x)|}\int_{B_R(x)} \sigma(u)^{-1/(r-1)}\right)^{r-1} \leq C.$$

- O bien si el problema lineal

$$\begin{cases} \psi \in W_0^{1,p}(\Omega),\ \sigma(u)^{1/2}\nabla\psi \in L^2(\Omega)^N, \\ \displaystyle\int_\Omega \sigma(u)\nabla\psi \cdot \nabla\phi = 0, \quad \forall\phi \in H_0^1(\Omega), \end{cases}$$

  sólo posee la solución trivial $\psi = 0$ (nótese que no podemos tomar $\phi = \psi$).

Con la regularidad aquí deducida no se ha conseguido probar si esta igualdad se cumple o no.

4 Un sistema elíptico degenerado y singular

El caso descrito anteriormente no conduce a estimaciones L^∞ para la temperatura u. En esta sección se analiza el caso de una conductividad térmica singular, esto es, $a(s)$ explota para un valor finito $s = \tau > 0$. Bajo ciertas hipótesis sobre los datos iniciales, probamos que la temperatura está acotada en Ω. Concretamente, se prueba que $0 \le u(x) < \tau$ casi por doquier en Ω.

Consideremos el problema del termistor (2) bajo las siguientes hipótesis:

(H.1) $\sigma \in C(\mathbb{R})$ y $\sigma(s) > 0$ para cualquier $s \in \mathbb{R}$.

(H.2) $a \in C(-\infty, \tau)$, $\tau > 0$, $a(0) = 0$, $a(s) > 0$ para cualquier $s \in (0, \tau)$, $a(s) \ge 0$ para todo $s < \tau$ e $\int_0^\tau a(s)\,\mathrm{d}s = +\infty$.

(H.3) Existe $n_0 > 1/\tau$ tal que $a(s)$ es una función creciente en el intervalo $(\tau - 1/n_0, \tau)$.

(H.4) $\varphi_0 \in H^1(\Omega)$.

Nota 3 La hipótesis (H.1) es muy general. En efecto; en la sección anterior σ no era uniformemente elíptica, obteniéndose así un coeficiente de difusión degenerado. Además no asumimos hipótesis alguna sobre el comportamiento asintótico de $\sigma(s)$ para valores grandes de s. Por otro lado, (H.2) y (H.3) implican que $\lim_{s\to\tau^-} a(s) = +\infty$; por lo tanto $a(s)$ es singular para el valor finito $s = \tau$.

Se tiene el siguiente resultado de existencia:

Teorema 4 *Bajo las hipótesis* (H.1)–(H.4)*, el problema* (2) *posee solución débil* (u, φ) *en el sentido siguiente:*

$$
\begin{gathered}
u \in L^\infty(\Omega), \;\; 0 \le u < \tau \;\; c.p.d. \;\; en \;\; \Omega, \\
u \in W^{1,q}_{\mathrm{loc}}(\Omega), \;\; A(u) \in W^{1,q}_0(\Omega), \;\; q < \frac{N}{N-1} \;\; si \;\; N \ge 2, \;\; q = 2 \;\; si \;\; N = 1, \\
\nabla A(u) = a(u)\nabla u, \quad \varphi - \varphi_0 \in H^1_0(\Omega), \\
\int_\Omega a(u)\nabla u \cdot \nabla \xi = \int_\Omega \sigma(u)|\nabla \varphi|^2 \xi \quad \forall \xi \in \mathcal{D}(\Omega), \\
\int_\Omega \sigma(u)\nabla\varphi\nabla\phi = 0 \quad \forall \phi \in H^1_0(\Omega).
\end{gathered}
$$

Demostración. Introduzcamos la función de truncamiento T^n, dada por

$$
T^n(s) = \begin{cases} s & \text{si } s < \tau - 1/n, \\ \tau - 1/n & \text{si } s \ge \tau - 1/n. \end{cases}
$$

Definimos los coeficientes regularizados de difusión $a_n(s) = a(T^n(s)) + \frac{1}{n}$ y $\sigma_\tau(s) = \sigma(T_\tau(s))$ y planteamos los problemas aproximados

$$
\begin{cases}
-\nabla \cdot (a_n(u_n)\nabla u_n) = \sigma_\tau(u_n)|\nabla\varphi_n|^2 & \text{en } \Omega, \\
\nabla \cdot (\sigma_\tau(u_n)\nabla\varphi_n) = 0 & \text{en } \Omega, \\
u_n = 0 & \text{sobre } \partial\Omega, \\
\varphi_n = T_n(\varphi_0) & \text{sobre } \partial\Omega.
\end{cases} \tag{36}
$$

Aplicando resultados clásicos de existencia (véase [3]), se llega a que el problema (36) posee solución débil (u_n, φ_n), con $u_n \in H_0^1(\Omega)$ y $\varphi_n - T_n(\varphi_0) \in H_0^1(\Omega) \cap L^\infty(\Omega)$.

De (H.1) deducimos la existencia de unas constantes, $c_\tau, C_\tau > 0$, tales que $c_\tau \leq \sigma_\tau(s) \leq C_\tau$ para cualquier $s \in \mathbb{R}$. En particular, tomando $\phi = \varphi_n - T_n(\varphi_0)$ en la ecuación para φ, resulta que (φ_n) está acotada en $H^1(\Omega)$.

Por otro lado, haciendo

$$A_n(s) = \int_0^s a_n(t)\,\mathrm{d}t, \quad v_n = A_n(u_n),$$

se obtiene que $v_n \geq 0$ c.p.d. en Ω y (26)–(28) son válidas. Gracias a (H.3) y a que (v_n) está acotada en $W_0^{1,q}(\Omega)$, se prueba que $(A(T^n(u_n))) \subset H_0^1(\Omega)$ está acotada $W_0^{1,q}(\Omega)$ para cualquier $q < \frac{N}{N-1}$. Consecuentemente, existe una función $w \in W_0^{1,q}(\Omega)$ tal que, para una subsucesión, $A(T^n(u_n)) \rightharpoonup w$ débilmente en $W_0^{1,q}(\Omega)$ y c.p.d. en Ω.

Nótese que el conjunto $\{w = +\infty\}$ tiene medida nula; ahora bien, como A y $A^{-1} : [0, +\infty) \to [0, \tau)$ son biyectivas, para cada $x \in \Omega$ con $w(x) < +\infty$, $T^n(u_n) = A^{-1}(A(T^n(u_n))) \to A^{-1}(w(x))$. Pongamos $u(x) = A^{-1}(w(x))$; entonces $w = A(u)$, $0 \leq u(x) < \tau$ y además $u_n \to u$ c.p.d. en Ω.

Como (φ_n) está acotada en $H^1(\Omega)$, existen una subsucesión (que denotaremos de la misma forma) y una función $\varphi \in H^1(\Omega)$ tales que $\varphi = \varphi_0$ sobre $\partial\Omega$ y $\varphi_n \rightharpoonup \varphi$ débilmente en $H^1(\Omega)$. Al ser $\sigma_\tau(u) = \sigma(u)$, se deduce que $\varphi_n \to \varphi$ fuertemente en $H^1(\Omega)$, con lo cual podemos pasar al límite en los problemas aproximados:

$$\begin{cases} \displaystyle\int_\Omega \nabla A(u) \cdot \nabla \xi = \int_\Omega \sigma(u) |\nabla \varphi|^2 \xi \quad \forall\, \xi \in \mathcal{D}(\Omega), \\ \displaystyle\int_\Omega \sigma(u) \nabla \varphi \cdot \nabla \phi = 0 \quad \forall\, \phi \in H_0^1(\Omega). \end{cases}$$

Sólo resta probar que $u \in W^{1,q}_{\mathrm{loc}}(\Omega)$ y $\nabla A(u) = a(u)\nabla u$. Estas propiedades se basan en el siguiente resultado (véanse [21, 24]):

Proposición 5 *Para cada subconjunto compacto $\mathcal{K} \subset \Omega$ existe una constante $\beta_\mathcal{K} > 0$ tal que* $\inf \operatorname{ese}_\mathcal{K} A(u) \geq \beta_\mathcal{K}$.

Como A^{-1} es globalmente Lipschitziana sobre conjuntos de la forma $[\varepsilon, +\infty)$, $\varepsilon > 0$, $w \in W_0^{1,q}(\Omega)$ y $u = A^{-1}(w)$, tenemos que, para cualquier subdominio $\Omega' \subset \Omega$ con clausura compacta en Ω, $u \in W^{1,q}(\Omega')$ y $\nabla u = a(u)^{-1}\nabla w$ en Ω'. Nótese que $a(u)^{-1} \in L^\infty(\Omega')$ gracias a la proposición 5. En conclusión, hemos deducido que $\nabla w = \nabla A(u) = a(u)\nabla u$ en Ω' para cualquier subdominio $\Omega' \subset \Omega$ con clausura compacta en Ω, quedando así probado el teorema 4. $\square$

5 Soluciones renormalizadas de un sistema parabólico-elíptico

Planteamos ahora el problema

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nabla \cdot (a(u)\nabla u) = \sigma(u)|\nabla\varphi|^2 & \text{en } Q, \\ -\nabla \cdot (\sigma(u)\nabla\varphi) = \nabla \cdot F(u) & \text{en } Q, \\ u = 0 & \text{sobre } \partial\Omega \times (0,T), \\ \varphi = 0 & \text{sobre } \partial\Omega \times (0,T), \\ u(\cdot,0) = u_0 & \text{en } \Omega. \end{cases} \tag{37}$$

Este sistema constituye una generalización del problema (1), en cuyo contexto $F(x,t,s) = \sigma(s)\nabla\varphi_0(x,t)$. Supongamos las hipótesis que siguen:

(H.1) σ, $a \in C(\mathbb{R})$ son tales que $0 < \sigma_1 \leq \sigma(s) \leq \sigma_2$, $0 < \alpha \leq a(s)$, para cualquier $s \in \mathbb{R}$.

(H.2) $u_0 \in L^1(\Omega)$.

(H.3) $F : Q \times \mathbb{R} \to \mathbb{R}^N$ es una función de Caratheodory y $F \in L^\infty(Q \times \mathbb{R})$.

Antes de abordar el análisis del problema (37), debemos aclarar qué se entiende por solución del mismo. El hecho de no imponer ninguna hipótesis sobre el comportamiento asintótico del coeficiente de difusión $a(s)$ significa que el marco de soluciones débiles no es el apropiado en este contexto, pues $a(u)$ no tiene por qué pertenecer a ningún espacio L^p. Por otro lado, a esta dificultad se añade la escasa regularidad que se impone sobre los datos y la que se deduce sobre los términos no lineales. La noción de solución renormalizada adaptada al sistema (37) resuelve las cuestiones anteriores. Este concepto fue introducido primeramente por DiPerna y Lions (véanse [18, 19]) en el contexto de las ecuaciones de Fokker-Plank-Boltzman. Más tarde, se aplicó a situaciones más generales: resolución de ecuaciones elípticas no lineales (véanse [8, 31, 32]) o de ecuaciones parabólicas no lineales (véanse [5, 6, 7]).

Definición 2 *Llamamos solución renormalizada de (37) a todo par de funciones (u,φ) para el que se cumplen las siguientes condiciones:*

(R.1) $u \in L^1(Q)$ *y* $\varphi \in L^2(H_0^1(\Omega))$*;*

(R.2) $T_M(u) \in L^2(H_0^1(\Omega))$ *para cualquier* $M > 0$*;*

(R.3) $\lim\limits_{n\to\infty} \displaystyle\int_{\{n<|u|<n+1\}} a(u)|\nabla u|^2 = 0$*;*

(R.4) *Para cualquier* $S \in C^\infty(\mathbb{R})$ *con* $\operatorname{sop} S'$ *compacto se tiene*

$$\frac{\partial S(u)}{\partial t} - \nabla \cdot [a(u)\nabla u S'(u)] + S''(u)a(u)\nabla u \nabla u = \sigma(u)|\nabla\varphi|^2 S'(u) \quad \text{en } \mathcal{D}'(Q),$$
$$S(u(\cdot,0)) = S(u_0) \quad \text{en } \Omega;$$

(R.5) $\displaystyle\int_Q \sigma(u)\nabla\varphi\cdot\nabla\psi = -\int_Q F(u)\cdot\nabla\psi$, *para cualquier* $\psi \in L^\infty(H_0^1(\Omega))$.

Se tiene entonces el siguiente resultado de existencia (véase [21]):

Teorema 6 *Bajo las hipótesis* (H.1)–(H.3)*, el sistema* (37) *posee solución renormalizada. Además, toda solución renormalizada* (u,φ) *es tal que* $u \in L^q(W_0^{1,q}(\Omega))$*, con* $q < \frac{N+2}{N+1}$ *si* $N \geq 2$*,* $q < 2$ *si* $N = 1$ *y* $\varphi \in L^\infty(H_0^1(\Omega))$.

No obstante, se puede deducir un resultado de existencia de solución renormalizada bajo unas hipótesis aún más débiles que las anteriores ([25]):

(H.1) $a, \sigma : Q\times\mathbb{R}\to\mathbb{R}$ y $F : Q\times\mathbb{R}\to\mathbb{R}^N$ son funciones de Caratheodory y $\gamma : \mathbb{R}^+\to\mathbb{R}^+$ es una función creciente satisfaciendo, para cualquier $s\in\mathbb{R}$ y c.p.d. en Q, que $\max(a(x,t,s),\sigma(x,t,s),|F(x,t,s)|)\leq\gamma(|s|)$.

(H.2) Existen dos constantes $a_0,\sigma_0 > 0$ tales que $a(x,t,s)\geq a_0$, $\sigma(x,t,s)\geq\sigma_0$ para cualquier $s\in\mathbb{R}$ y c.p.d. en Q.

(H.3) $\Gamma \in L^1(Q)$ es una función de Caratheodory tal que $|F(x,t,s)|^2 \leq \Gamma(x,t)\sigma(x,t,s)$ para cualquier $s\in\mathbb{R}$ y c.p.d. en Q.

(H.4) $\displaystyle\max_{k\leq|s|\leq 2k}\operatorname*{sup\,ese}_Q \frac{1}{k}\frac{\sigma(x,t,s)}{a(x,t,s)} = \omega(k)$ cuando $k\to\infty$, donde $\displaystyle\lim_{k\to\infty}\omega(k) = 0$.

(H.5) $u_0 \in L^1(\Omega)$.

Nótese que, en este caso, los coeficientes de difusión no están acotados; es más, no estamos asumiendo ningún comportamiento asintótico sobre a, σ y F. Bajo estas hipótesis tan generales, aun cuando u y φ pertenezcan a algún espacio $L^q(W^{1,q}(\Omega))$, los términos $a(u)\nabla u$, $\sigma(u)\nabla\varphi$ o $F(u)$ pueden no pertenecer a ningún $L^r(Q)$ con $r\geq 1$. Es por ello que retomamos la noción de solución renormalizada adaptada a (37). De este modo, tenemos el

Teorema 7 *Bajo las hipótesis* (H.1)–(H.5)*, el sistema (37) admite solución renormalizada* (u,φ) *en el sentido de la definición 2, donde ahora las condiciones* (R.1) *y* (R.5) *serán*

(R.1) $u\in L^1(\Omega)$, $\varphi\in L^2(H_0^1(\Omega))$ *y* $\displaystyle\int_Q\sigma(u)|\nabla\varphi|^2 < +\infty$;

(R.5) $\displaystyle\int_Q\sigma(u)\nabla\varphi\cdot\nabla\psi = -\int_Q F(u)\cdot\nabla\psi$ *para toda* $\psi\in L^2(H_0^1(\Omega))$ *tal que* $\displaystyle\int_Q\sigma(u)|\nabla\psi|^2 < +\infty$.

Nota 4 Las condiciones (R.1)–(R.4) sobre u son las propiedades usuales de las soluciones renormalizadas de ecuaciones parabólicas (véase [6]). Por otro lado, en (R.5) se establece que el conjunto de funciones de test para la ecuación de φ depende de la solución u.

Nota 5 Los coeficientes de difusión a y σ son funciones escalares en el planteamiento descrito por las hipótesis (H.1)–(H.4). Podemos considerar un caso más general en el que a y σ sean matrices de difusión de orden $N \times N$ (véase [25]). Las hipótesis en este caso son análogas a las del caso escalar y el resultado de existencia dado en el teorema 7 continúa siendo cierto.

6 Solución de capacidad de un sistema parabólico-elíptico acoplado no lineal

Se considera el sistema parabólico-elíptico no lineal

$$\begin{cases} \dfrac{\partial u}{\partial t} - \nabla \cdot a(x,t,u,\nabla u) = \sigma(u)|\nabla\varphi|^2 & \text{en } Q, \\ \nabla \cdot (\sigma(u)\nabla\varphi) = 0 & \text{en } Q, \\ u = 0 & \text{sobre } \partial\Omega \times (0,T), \\ \varphi = \varphi_0 & \text{sobre } \partial\Omega \times (0,T), \\ u(\cdot,0) = u_0 & \text{en } \Omega, \end{cases} \tag{38}$$

que puede considerarse como una versión más general del problema del termistor, donde $a(x,t,s,\xi) = \kappa(s)\xi$ (siendo κ la conductividad térmica) y del problema planteado por Xu ([42]). Nuestro objetivo en esta sección es estudiar la existencia de solución de capacidad del problema (38), donde $a : Q \times \mathbb{R} \times \mathbb{R}^N \mapsto \mathbb{R}^N$ es un operador de tipo Leray-Lions ([29, 30]) y $\sigma \in C^0(\mathbb{R})$ es tal que $0 < \sigma(s) \leq \bar{\sigma}$ para cualquier $s \in \mathbb{R}$. En situaciones de interés práctico, σ satisface esta propiedad e incluso $\sigma(s) \to 0$ cuando $|s| \to \infty$.

Bajo estas hipótesis tan generales nos hallamos ante una situación similar a la de [42], comentada en la Introducción, esto es, el marco de las soluciones débiles no puede aplicarse directamente en este contexto y nos vemos obligados a usar la noción de solución de capacidad adaptada a (38).

Éste no es el único obstáculo en la resolución de este sistema. En efecto, el término de divergencia de la ecuación parabólica dificulta en gran medida nuestro análisis, al ser a un operador de Leray-Lions. La existencia de solución se obtiene habitualmente como el límite de soluciones de problemas aproximados con datos iniciales regulares; la convergencia c.p.d. de los gradientes suele ser la etapa más complicada en la demostración de existencia de solución de ecuaciones no lineales con datos iniciales medidas o con regularidad L^1.

6.1 Hipótesis y definición de solución de capacidad

Supongamos las siguientes hipótesis sobre los datos:

(H.1) $a : Q \times \mathbb{R} \times \mathbb{R}^N \mapsto \mathbb{R}^N$ es una función de Caratheodory, esto es, $a(\cdot,\cdot,s,\xi)$ es medible para cualesquiera $s \in \mathbb{R}$ y $\xi \in \mathbb{R}^N$ y $a(x,t,\cdot,\cdot)$ es continua en $\mathbb{R} \times \mathbb{R}^N$ para $(x,t) \in Q$ c.p.d.

(H.2) Existe una constante $\alpha > 0$ tal que, para cualesquiera $s \in \mathbb{R}$ y ξ, $\eta \in \mathbb{R}^N$, $\xi \neq \eta$ y para $(x,t) \in Q$ c.p.d., $[a(x,t,s,\xi) - a(x,t,s,\eta)](\xi - \eta) \geq \alpha|\xi - \eta|^2$.

(H.3) $a(x,t,s,0)=0$.

(H.4) Existen una función positiva $b\in L^2(Q)$ y una constante $\beta>0$ tales que, para cualesquiera $s\in\mathbb{R}$ y $\xi\in\mathbb{R}^N$ y $(x,t)\in Q$ c.p.d., $|a(x,t,s,\xi)|\leq \beta[b(x,t)+|s|+|\xi|]$.

(H.5) $\sigma\in C(\mathbb{R})$ es tal que $0<\sigma(s)\leq\bar{\sigma}$ para cualquier $s\in\mathbb{R}$.

(H.6) $\varphi_0\in L^2(H^1(\Omega))\cap L^\infty(Q)$.

(H.7) $u_0\in L^2(\Omega)$.

Definición 3 *Se dice que una terna (u,φ,Φ) es solución de capacidad del problema (38) si se cumplen las siguientes condiciones:*

(C.1) $u\in L^2(H_0^1(\Omega))$, $\frac{\mathrm{d}v}{\mathrm{d}t}\in L^2(H^{-1}(\Omega))$, $\varphi\in L^\infty(Q)$, $\Phi\in L^2(Q)^N$.

(C.2) (u,φ,Φ) *satisface el sistema de ecuaciones diferenciales*

$$\begin{cases} \dfrac{\mathrm{d}u}{\mathrm{d}t}-\nabla\cdot a(x,t,u,\nabla u)=\nabla\cdot(\varphi\Phi) & \text{en } L^2(H^{-1}(\Omega)),\\ \nabla\cdot\Phi=0 & \text{en } L^2(H^{-1}(\Omega)). \end{cases}$$

(C.3) *Para cualquier $S\in C_0^1(\mathbb{R})$, $S(u)\varphi-S(0)\varphi_0\in L^2(H_0^1(\Omega))$ y $S(u)\Phi=\sigma(u)\,(\nabla(S(u)\varphi)-\varphi\nabla S(u))$.*

(C.4) $u(\cdot,0)=u_0$.

Nota 6 La diferencia fundamental entre una solución de capacidad y una solución débil es que, en la primera, se considera $\nabla\varphi$ c.p.d., mientras que en la segunda, $\nabla\varphi$ es el gradiente en el sentido de las distribuciones. Concretamente, si u está acotada en Q, es obvio que ambas nociones de solución son equivalentes: tomando $S\in C_0^1(\mathbb{R})$ con $S=1$ en el intervalo $[-\|u\|_{L^\infty(Q)},\|u\|_{L^\infty(Q)}]$, (C.3) implica que $\varphi-\varphi_0\in L^2(H_0^1(\Omega))$ y $\Phi=\sigma(u)\nabla\varphi$. Por otro lado, si $u\in L^1(Q)$ no está acotada, tomamos $m>0$ y $S_m\in C_0^1(\mathbb{R})$ tal que $S_m=1$ sobre $[-m,m]$; entonces $S_m(u)\varphi=\varphi$ sobre $\{|u|\leq m\}$, $S_m(u)\varphi\in L^2(H^1(\Omega))$ y podemos definir $\nabla\varphi$ en $\{|u|\leq m\}$ c.p.d. en Q. Además $\nabla\varphi$ es una función medible y $\nabla\varphi\in L^2(\{|u|\leq m\})^N$ para cualquier $m>0$. Por lo tanto, tomando $S=S_m$ en (C.3), y haciendo $m\to\infty$, obtenemos que $\Phi=\sigma(u)\nabla\varphi$ c.p.d. en Q. Por otro lado, podría ocurrir que φ no fuese lo suficientemente regular como para poder definir en sentido clásico su traza sobre $\partial\Omega\times(0,T)$; sin embargo, usando de nuevo (C.3), se puede deducir que $\varphi=\varphi_0$ sobre $\partial\Omega\times(0,T)$. En efecto, eligiendo $S\in C_0^1(\mathbb{R})$ con $S(0)=1$ y en vista de que $S(u)\varphi\in L^2(H^1(\Omega))$ tenemos que $\varphi=\varphi_0$ sobre $\partial\Omega\times(0,T)$.

Se tiene entonces el siguiente resultado (véase [26]):

Teorema 8 *Bajo las hipótesis* (H.1)–(H.7)*, el sistema* (38) *posee solución de capacidad.*

Nota 7 Podemos considerar un operador de tipo Leray-Lions más general actuando sobre $L^p(W_0^{1,p}(\Omega))$, $1 < p < \infty$, donde las hipótesis (H.2) y (H.4) se sustituirían, respectivamente, por

(H.2)′ Existe una constante $\alpha > 0$ tal que, para cualesquiera $s \in \mathbb{R}$ y ξ, $\eta \in \mathbb{R}^N$, $\xi \neq \eta$ y c.p.d. $(x,t) \in Q$, $[a(x,t,s,\xi) - a(x,t,s,\eta)](\xi - \eta) \geq \alpha|\xi - \eta|^p$.

(H.4)′ Existen una función positiva $b \in L^{p'}(Q)$, $1/p + 1/p' = 1$, y una constante $\beta > 0$ tales que, para cualesquiera $s \in \mathbb{R}$ and $\xi \in \mathbb{R}^N$, y c.p.d. $(x,t) \in Q$, $|a(x,t,s,\xi)| \leq \beta[b(x,t) + |s|^{p-1} + |\xi|^{p-1}]$.

Esta situación ha sido considerada recientemente por Badii y Díaz [4] cuando el operador $\nabla \cdot a$ es el p-laplaciano, con $p > 2$.

7 Sobre la unicidad

Nuestras contribuciones al problema del termistor no han abordado la cuestión de la unicidad de soluciones, ya que los resultados de unicidad para el problema del termistor se deducen a partir de la estimación $\varphi \in L^\infty$ (véanse [3, 11, 40, 44]), propiedad que no se tiene en las situaciones contempladas.

Para los sistemas elípticos doblemente degenerado y singular-degenerado, la deducción de la unicidad es complicada, pues las hipótesis consideradas en estos casos son muy débiles. No obstante, en [12] los autores han demostrado un resultado de unicidad en el caso degenerado con $N = 1$.

Tampoco hemos analizado la unicidad de la solución renormalizada ni de capacidad de los problemas (37) y (38), respectivamente, debido a la complejidad que esto supone en ambos casos. De este modo, en el primero precisamos obtener estimaciones L^∞ para u y φ, lo cual se conseguiría asumiendo, por ejemplo, que $a, \sigma, F \in L^\infty$. En el segundo, tendríamos que suponer, por ejemplo, que $\varphi_0 \in L^\infty(W^{1,\infty}(\Omega))$. Pero entonces no habría necesidad de buscar soluciones renormalizadas o de capacidad, ya que la regularidad de las mismas nos conducirían a la existencia de soluciones débiles. En general, no se puede esperar la unicidad de solución sin hipótesis adicionales sobre a y σ; en [15], se exhibe un ejemplo en el que el problema estacionario del termistor no tiene unicidad de solución.

Agradecimientos

Este trabajo ha sido financiado parcialmente por el Grupo de Investigación FQM 315 de la Junta de Andalucía y el Proyecto BFM2003-01187 del Ministerio de Ciencia y Tecnología, con la participación de los Fondos Europeos para el Desarrollo Regional (FEDER).

Referencias

[1] W. Allegretto y H. Xie. *Existence of solutions for the time-dependent thermistor equations*, IMA Journal of Applied Mathematicas, 48: 271–281, 1992.

[2] W. Allegretto y H. Xie. *A non-local thermistor problem*, European J. Appl. Math., 6: 83–94, 1993.

[3] S.N. Antontsev y M. Chipot. *The thermistor problem: existence, smoothness, uniqueness, blowup*, SIAM J. Math. Anal., 25(4): 1128–1156, 1994.

[4] M. Badii y I. J. Díaz *On the thermistor in periodical regime with a free boundary*. Aparecerá.

[5] D. Blanchard. *Truncations and mononicity methods for parabolic equations*, Nonlinear Anal., 21: 725–743, 1993.

[6] D. Blanchard y F. Murat. *Renormalized solutions of nonlinear parabolic problems with L^1-data: existence and uniqueness.* Proccedings of the Royal Society of Edinburgh, Sect. A, 127: 1137–1152, 1997.

[7] D. Blanchard y H. Redwane. *Renormalized solutions for a class of nonlinear evolution problems*, J. Math. Pures Appl., 77: 117–151, 1998.

[8] L. Boccardo, I. Díaz, D. Giachetti y F. Murat. *Existence of a solution for a weaker form of a nonlinear elliptic equation.* Recent Advances in Nonlinear Elliptic and Parabolic Problems. Notes in Math., 208, Logman, Harlow, 1989.

[9] L. Boccardo y T. Gallouët. *Non-linear elliptic and parabolic equations involving measure data.* J. Funct. Anal., 87: 149–169, 1989.

[10] K. Chau, W. Allegretto y L. Ristic. *Thermal modelling of CMOS temperature/flow microsensors*, Canad. J. Physics, 69: 212–216, 1991.

[11] M. Chipot y G. Cimatti. *A uniqueness result for the thermistor problem*, European J. Appl. Math., 2: 97–103, 1991.

[12] M. Chipot, J. I. Díaz y R. Kersner. *On the degenerate thermistor problem.* Aparecerá.

[13] G. Cimatti. *A bound for the temperature in the thermistor problem*, IMA J. Appl. Math., 40: 15–22, 1988.

[14] G. Cimatti. *Remark on existence and uniqueness for the thermistor problem under mixed boundary conditions*, Quart. Appl. Math., XLVII(1): 117–121, 1989.

[15] G. Cimatti. *The stationary thermistor problem with a current limiting device*, Proc. Roy. Soc. Edinburgh Sect. A, 116: 79–84, 1990.

[16] G. Cimatti. *Existence of weak solutions for the nonstationary problem of the Joule heating of a conductor*, Ann. Mat. Pura Appl. (4), CLXII: 33–42, 1992.

[17] G. Cimatti y G. Prodi. *Existence results for a nonlinear elliptic system modelling a temperature dependent electrical resistor*, Ann. Mat. Pura Appl. (4), 152: 227–236, 1988.

[18] R.J. Di Perna y P.L. Lions. *On the Fokker-Plank-Boltzman equations*, Comm. Math. Phys., 120: 1–23, 1988.

[19] R.J. Di Perna y P.L. Lions. *On the Cauchy problem for Boltzman equations: global existence and weak stability*, Ann. of Math., 130: 321–366, 1989.

[20] M. Gómez Mármol. *Estudio matemático de algunos problemas no lineales de la mecánica de fluidos incompresibles*. Tesis, Universidad de Sevilla, 1998.

[21] M. T. González Montesinos. *Estudio matemático de algunos problemas no lineales del electromagnetismo relacionados con el problema del termistor*. Tesis, Universidad de Cádiz, 2002.

[22] M.T. González Montesinos y F. Ortegón Gallego. *The evolution thermistor problem with degenerate thermal conductivity*, Commun. Pure Appl. Anal., 1(3): 313–325, 2002.

[23] M.T. González Montesinos y F. Ortegón Gallego. *A doubly degenerate elliptic problem*, Equadiff 10 CDROM, Masaryk University Publishing House: 169–176, 2002.

[24] M.T. González Montesinos y F. Ortegón Gallego. *On certain non-uniformly and singular non-uniformly elliptic systems*, Nonlinear Anal., 54: 1193–1204, 2003.

[25] M.T. González Montesinos y F. Ortegón Gallego. *Renormalized solutions to a nonlinear parabolic-elliptic system*, Aceptado en SIAM J. Math. Anal., 2004.

[26] M.T. González Montesinos y F. Ortegón Gallego. *Existence of a capacity solution to a coupled nonlinear parabolic-elliptic system*. Aparecerá.

[27] S.D. Howison, J.F. Rodrigues y M. Shilor. *Stationary solutions to the thermistor problem*, J. Math. Anal. Appl., 174: 573–588, 1993.

[28] F. J. Hyde. *Thermistors*, Iliffe Books, London (1971).

[29] J. Leray y J.L. Lions. *Quelques résultats de Višik sur les problèmes elliptiques non linéaires par les méthodes de Minty-Browder*, Bull. Soc. Math. France, 93: 97–107, 1965.

[30] J.L. Lions. *Quelques méthodes de résolution des problèmes aux limites non linéaires*. Dunod et Gauthier-Villars, 1969.

[31] F. Murat. *Soluciones renormalizadas de ecuaciones en derivadas parciales elípticas no lineales*. Publications du laboratoire d'Analyse Numérique Paris VI, R 93023, Cours a l'Université de Sevilla, 1993.

[32] F. Murat. *Équations elliptiques non linéaires avec second membre L^1 ou mesure*. Comptes rendus du 26ème Congrès national d'anayse numérique, Les Karellis, France, 1994.

[33] G. Ping y F. Jishan. *Global existence, uniqueness and asymptotic behavior of the solutions of the thermistor problem*, Nanjing Univ. J. Math. Biquaterly, 13(2): 156–167, 1996.

[34] G. Ping y F. Jishan. *Some remarks on stationary thermistor problem*, J. Southeast Univ. (English Ed.), 12(2): 93–96, 1996.

[35] D. Potter. *Measuring Temperature with thermistors - a Tutorial*, National Instruments, Application Note 065: 1–8, 1996.

[36] S. Selberherr. *Analysis and Simulation of Semiconductor Devices*, New York, Springer, 1984.

[37] J. Simon. *Compact sets in the space $L^p(0,T;B)$*, Ann. Mat. Pura Appl. (4), 146: 65–96, 1987.

[38] B.O. Turesson. *Nonlinear Potential Theory and Weighted Sobolev Spaces*. Lectures Notes in Math. 1736, Springer, Berlin, 2000.

[39] W. Xie. *A nonlinear elliptic system modelling the thermistor problem*, Appl. Anal., 50: 263–276, 1993.

[40] W. Xie. *On the existence and uniqueness for the thermistor problem*, Adv. Math. Sci. Appl., 2(1): 63–73, 1993.

[41] X. Xu. *A degenerate Stefan-like problem with Joule's heating*, SIAM J. Math. Anal., 23: 1417–1438, 1992.

[42] X. Xu. *A strongly degenerate system involving an equation for parabolic type and an equation of elliptic type*, Comm. Partial Differential Equations, 18(1&2): 199–213,1993.

[43] X. Xu. *The thermistor problem with conductivity vanishing for large temperature*, Proc. Roy. Soc. Edinburgh Sect. A, 124: 1–21, 1994.

[44] G. Yuan y Z. Liu. *Existence and uniqueness of the C^{α} solution for the thermistor problem with mixed boundary value*, SIAM J. Appl. Math., 25(4): 1157–1166, 1994.