

Estimación de parámetros
y
Clasificación de datos

Aplicaciones biomédicas

Juan Pablo Agnelli

Presentado ante la Facultad de Matemática, Astronomía y Física como parte de los
requerimientos para la obtención del grado de Doctor en Matemática de la

Universidad Nacional de Córdoba

Marzo 2011

Directora: Dra. Cristina V. Turner

A mis viejos y hermana.

Resumen

En esta tesis se proponen dos tipos de aplicaciones biomédicas para las cuales hemos empleado diferentes herramientas matemáticas y por lo cual el trabajo está dividido en dos partes. En la primera parte nos hemos abocado a la detección de tumores y estimación de parámetros asociados a regiones tumorales y en la segunda parte a la clasificación de tumores a partir de datos de niveles de expresión genética.

En la primer situación el objetivo es estimar la localización, tamaño y parámetros térmicos asociados a un tumor utilizando como información perfiles de temperaturas medidos sobre la superficie corporal. Desde el punto de vista matemático el estudio de estos problemas implica el planteo de problemas inversos, su análisis y el desarrollo de métodos numéricos para su resolución. En primera instancia, utilizamos ecuaciones en derivadas parciales para modelar la transferencia de calor en tejidos vivos, más precisamente consideramos la ecuación de Pennes estacionaria con condiciones de borde mixtas. Para esta ecuación elíptica probamos la existencia y unicidad de la solución y para resolver este problema directo utilizamos un esquema de diferencias finitas de segundo orden. Luego, para resolver los problemas inversos estos fueron reformulados como problemas de optimización y para resolver estos nuevos problemas presentamos dos metodologías diferentes. La primera se basa en el uso del algoritmo Pattern Search para minimizar la función costo. Este es un algoritmo de búsqueda directa que no hace uso de las derivadas y por ende es de fácil implementación. La segunda metodología que presentamos hace uso de la información que provee la derivada de la función a minimizar con respecto a las distintas variables a estimar. Para calcular tal derivada, deberemos acudir al análisis de sensibilidad.

En la segunda parte del trabajo, el objetivo es desarrollar un algoritmo capaz de extraer, de una gran base de datos, información que reside de manera implícita en estos. Dicha información es previamente desconocida y puede resultar útil para describir el proceso o fenómeno que está bajo análisis o estudio. En particular, aquí se aplica para la clasificación de distintos tipos de tumores usando como base de datos niveles de expresión genéticas. La metodología que se propone está basada en tres pilares fundamentales: 1) la eliminación de una distinción estricta entre datos de entrenamiento y datos de prueba, mediante una asignación de estos últimos a las distintas clases, en un espíritu del algoritmo Expectation-Maximization; 2) un procedimiento para la estimación de densidades que transforma la distribución de probabilidad original en una distribución Gaussiana isotrópica y 3) una medida de la capacidad de agrupamiento asociada a un conjunto de variables, medida que produce de manera natural en un procedimiento para la elección de variables. Esta metodología resulta particularmente útil en situaciones donde hay relativamente muy pocas observaciones correspondientes a un fenómeno que es descrito por medio de una gran cantidad de variables y del cual no se tiene un conocimiento a priori que permita asociar un subconjunto de estas variables para realizar la clasificación requerida.

De acuerdo a los resultados obtenidos podemos decir que las metodologías propuestas en la primera parte pueden ser consideradas de potencial ayuda para localizar regiones tumorales, principalmente melanomas nodulares, como así también para estimar parámetros relacionados con estos. La misma conclusión vale para la metodología empleada en la segunda parte a la hora de realizar diagnóstico, prevención y tratamiento de las enfermedades a partir del conocimiento de los niveles de expresión genéticas del paciente.

Palabras Claves: Transferencia de calor, Problema Inverso, Optimización de formas, Estimación de densidad de probabilidad, Máxima verosimilitud, Teorema de Bayes.

Clasificación matemática (MSC 2010): 35Q92, 45Q05, 62G07, 62H30.

Abstract

In this thesis we propose two main areas of study, so the work is divided into two parts. The first one is related with tumor location and estimation of parameters related with tumor regions and the second part is concerned with the development of an algorithm for tumor classification from gene expression levels.

In the first situation the goal is to estimate position, size and thermal parameters of a tumor using temperature profiles that have been measured on the top boundary of the domain using a thermography camera. From the mathematical point of view the study of these problems imply to pose and analyze inverse problems and also to develop numerical methods to solve it. In a first stage, we use partial differential equations to model heat transfer in living tissue, more precisely we consider the stationary Pennes equation with mixed boundary conditions. For this elliptical equation we have proved existence and uniqueness of the solution and to solve this direct problem a finite difference scheme of second order is considered. Then, to solve the inverse problems these problems were reformulated as optimization problems and to solve these new problems two different methodologies will be presented. The first one, is based on the use of the Pattern Search algorithm. This is a direct search algorithm, so it does not make use of derivatives and therefore is very easy to implement. The second methodology that we present makes use of the information provided by the derivative of the function to minimize with respect to the different variables to be estimated. To calculate this derivative we consider some sensitivity analysis tools.

In the second part of the work, the goal is to build an algorithm capable to extract, from a large database, useful information that resides implicitly. This information is previously unknown and may be useful to describe the process or phenomenon that is under analysis or study. In particular, here we are interested in classify different types of tumors using gene expression levels. The proposed methodology is based on three main ingredients: 1) the blurring of distinctions between training and testing populations, through the soft assignment of the latter to classes, in an expectation-maximization framework, 2) a procedure for density estimation through a descent flow, that transforms the original distribution into an isotropic Gaussian distribution and 3) a measure of the clustering capability of a set of variables, which leads to an effective procedure for variable selection. The methodology is particularly useful in situations where there are relatively few observations for a phenomenon that is described by a large amount of variables, and no a priori knowledge that strongly links a small subset of these variables to the classification sought.

According to the results obtained the methodologies proposed in the first part of this work can be considered as a potential tool to locate tumor regions, like nodular melanomas, as well as to estimate parameters associated with them that could be useful and important to study the tumor evolution after a treatment procedure. The same conclusion applies to the methodology developed in the second part in order to diagnose, prevent and treat different diseases based on gene expression levels.

Key words: Heat transfer equations, Inverse problem, Shape optimization, Density estimation, Maximum likelihood, Bayes' Theorem.

Mathematics Subject Classification 2010: 35Q92, 45Q05, 62G07, 62H30.

Agradecimientos

Agradezco principalmente a Cristina por sus enseñanzas y por toda su confianza que depositó en mí para realizar el presente trabajo. Le agradezco además por el entusiasmo y la alegría con que hemos trabajado durante todos estos años, por su enorme generosidad y por sus valiosos consejos que no sólo me han ayudado en lo académico sino también en lo humano. Para ella, mi mayor admiración.

Deseo agradecer también a todos aquellos profesores de la FaMAF que a lo largo de estos años han contribuido e influenciado mi formación académica. En especial a Andrés Barrea, por su generosidad, su humildad, su buen humor y por sus consejos y charlas tenidas a lo largo de estos años, motivos entre otros tantos que me hacen considerarlo un amigo. A Tomás Godoy y Uriel Kauffman por la constante ayuda brindada y total predisposición mostrada para resolver mis dudas. A cada uno de los integrantes del grupo de Análisis Numérico y Computación por su generosidad, apoyo y buenos momentos compartidos.

A Esteban Tabak, quien con gran generosidad me permitió realizar un estadía en el Instituto Courant donde trabajé bajo su dirección. Le agradezco enormemente todas sus enseñanzas y por sobre todo destaco su humildad, su paciencia y su excelente humor que hicieron de esta estadía una experiencia maravillosa.

A los miembros del jurado Carlos Condat, Germán Torres y Claudio Padra por su predisposición a leer minuciosamente este trabajo y aconsejarme para mejorarlo.

A los miembros de la 334 y los vecinos de la 332 y la 333 por los cafes, mates y principalmente los excelentes momentos compartidos.

A los incondicionales de siempre, mis amig@s: Aldana, Any, Belén, Chan, Leti, Yami G., Yami Y., Fede V., Fede P., Franco, Emilio, Agustín, Pablo R., Chun, Renato y a vos que leíste y no te encontraste, pero que no te vas a ofender por mi olvido ... por algo me consideras tu amigo y te buscaste aquí.

Agradezco también a los organismos oficiales que con su financiación me permitieron llevar a cabo este doctorado: FaMAF, CIEM, CONICET y Secyt-UNC.

Finalmente, mi mayor gratitud para quienes dedico esta tesis, mis padres Maria Inés y Hector y a mi hermana Silvina por ser un pilar de apoyo y una fuente de cariño constante.

Índice general

1. Introducción	1
I Estimación de parámetros asociados a tumores	7
2. Localización de Tumores	9
2.1. Motivación del problema	9
2.2. Modelo matemático.	10
2.3. Existencia y unicidad de la solución.	12
2.3.1. Conceptos básicos.	12
2.3.2. Existencia y unicidad de la solución.	16
2.4. Resolución Numérica	19
2.5. Ejemplos numéricos	22
3. Localización: metodología 1	25
3.1. Problemas Inversos	25
3.2. Definición de los problemas inversos.	28
3.2.1. Problema Inverso 1: Estimación de la posición y tamaño del tumor.	28
3.2.2. Problema Inverso 2: Estimación de la fuente de calor metabólico.	30
3.3. Resultados Numéricos.	32
3.3.1. Resultados: PI 1 Estimación de la posición y tamaño del tumor.	33
3.3.2. Resultados: PI 2 Estimación de la intensidad de la fuente de calor metabólico.	35
3.4. Conclusión del capítulo	37
4. Localización: metodología 2	39
4.1. Introducción al Análisis de Sensibilidad.	39
4.2. Análisis de sensibilidad al cambio en los parámetros.	41
4.2.1. Una metodología para calcular la derivada con respecto a las variables de diseño.	42
4.3. Análisis de sensibilidad al cambio de forma	44
4.3.1. Una metodología para calcular la derivada de forma.	45
4.3.1.A. Operaciones básicas de derivación.	46
4.3.1.B. Cálculo de la derivada de forma.	47
4.4. Resolución del problema: PI 1	49
4.4.1. Cálculo de la derivada de forma	49
4.4.2. Resolución numérica.	52
4.4.3. Resultados.	54
4.5. Resolución del problema: PI 2	56

4.5.1.	Cálculo de la derivada con respecto a la intensidad de calor metabólico q_0	56
4.5.2.	Resolución Numérica	58
4.5.3.	Resultados	59
4.6.	Conclusión del capítulo	60
4.7.	Apéndice: Derivadas de Fréchet y Gâteaux.	61
II	Clasificación y agrupamiento de datos	63
5.	Clasificación y agrupamiento de datos	65
5.1.	Motivación del problema	65
5.2.	Clasificación y Agrupamiento: una visión unificada	67
5.2.1.	Clasificación y Agrupamiento mediante estimación de densidades	68
5.3.	Estimación de densidades	71
5.3.1.	Un flujo normalizador	72
5.3.2.	Algoritmo	75
5.4.	Elección de variables y evaluación del agrupamiento	77
5.5.	Ejemplos clínicos: clasificación de tumores	79
5.5.1.	Diagnóstico de una muestra por vez	79
5.5.2.	Diagnóstico de múltiples muestras	81
5.5.3.	Agrupamiento	84
5.6.	Conclusiones del capítulo	88
6.	Discusión y Conclusiones generales	89

Índice de figuras

2.1. Dominio 2 dimensional: $\Omega - \bar{\omega}$ tejido sano y ω región tumoral.	11
2.2. Ejemplo 2D de la distribución de temperatura y del perfil de temperatura sobre la superficie corporal	23
2.3. Ejemplo 2D de la distribución de temperatura y del perfil de temperatura sobre la superficie corporal	23
2.4. Ejemplo 3D del perfil de temperatura sobre la superficie corporal	24
2.5. Ejemplo 3D del perfil de temperatura sobre la superficie corporal	24
3.1. Ejemplo de estimación de la localización y tamaño de un tumor usando la metodología 1.	34
4.1. Campo vectorial 2D.	53
4.2. Ejemplo de estimación de la localización y tamaño de un tumor usando la metodología 2.	55
4.3. Ejemplo de estimación de la localización y tamaño de un tumor, comparación de ambas metodologías.	56
5.1. Población de entrenamiento y observación de prueba graficados considerando los 2 genes seleccionados por el algoritmo de elección de variables.	81
5.2. Evolución de las observaciones de acuerdo a los flujos z_k	82
5.3. Evolución de las probabilidades p_k	83
5.4. Población de entrenamiento y observación de prueba graficados considerando los 2 genes seleccionados por el algoritmo de elección de variables.	83
5.5. Población de entrenamiento y observación de prueba graficados considerando los 2 genes seleccionados por el algoritmo de elección de variables.	84
5.6. Evolución de las probabilidades p_k	84
5.7. Agrupamiento de las muestras correspondientes a tumores de células redondas y pequeñas.	86
5.8. Agrupamiento de las muestras correspondientes a los distintos tipos de leucemia.	86
5.9. Agrupamiento de las muestras correspondientes a los distintos tipos de tumores de células redondas.	87
5.10. Agrupamiento de las muestras correspondientes a los distintos tipos de leucemia.	88

Índice de cuadros

3.1. Estimación del centro y radio del tumor en 2D. Las diferentes columnas muestran los resultados obtenidos considerando los datos de entrada sin ruido y con 10 % y 15 % de ruido uniformemente distribuido respectivamente.	35
3.2. Estimación del centro y radio del tumor en 3D. Las diferentes columnas muestran los resultados obtenidos considerando los datos de entrada sin y con un ruido aleatorio del 10 % y 15 % respectivamente.	35
3.3. Estimación de la intensidad del calor metabólico dentro del tumor, caso 2D. En todos los ejemplos el tumor fue centrado en el punto (0.045,0.02) y su radio igual a 0.005.	36
3.4. Estimación de la intensidad del calor metabólico dentro del tumor, caso 3D. En todos los ejemplos el tumor fue centrado en el punto (0.045,0.045,0.015) y su radio igual a 0.005.	36
4.1. Estimación del centro y radio del tumor en 2D. Las diferentes columnas muestran los resultados obtenidos considerando los datos sin y con 10 % y 15 % de ruido aleatorio respectivamente.	56
4.2. Estimación de la intensidad del calor metabólico dentro del tumor, caso 2D. En todos los ejemplos el tumor fue centrado en el punto (0.045,0.02) y su radio igual a 0.005.	59

Capítulo 1

Introducción

La interacción entre la matemática y ciencias como la biología y medicina no es reciente, sino que ha existido desde hace décadas. Esta interacción resulta beneficiosa para ambas partes; para el matemático tratar de resolver los problemas que surgen en biología puede implicar el desarrollo de nuevas ideas, aplicaciones y teorías, mientras que para los biólogos o médicos el modelado matemático puede convertirse en una nueva herramienta y una poderosa técnica de experimentación. Entre los acontecimientos que muestran los beneficios de esta interacción podemos destacar, sin ánimo de ser exhaustivos, los siguientes:

- En la primera mitad del Siglo XIX, el botánico escocés Robert Brown estudiaba el proceso de fecundación de una planta cuando percibió un movimiento oscilatorio extremadamente rápido y cambiante en los granos de polen cuando estos se encontraban suspendidos en agua. Brown, en un principio pensó que se trataba de una manifestación de vitalidad del polen, pero luego corroboró que el mismo movimiento se daba en partículas de polvo, por lo tanto esto anulaba su anterior hipótesis que el movimiento se debía a que el polen tenía vida. Él mismo no pudo dar una teoría que explicara el porque de este movimiento. Recién en 1905, Einstein publicó la formalización y explicación teórica del mismo fenómeno. Dicha teoría hoy es conocida como movimiento browniano y la formulación matemática de Einstein es la base de las teorías matemáticas contemporáneas de difusión y caminatas aleatorias y además es de suma importancia en la teoría de la probabilidad.
- Uno de los pioneros en utilizar modelos matemáticos para describir el comportamiento de las epidemias fue el matemático, médico, naturalista y zoólogo escocés Ronald Ross, quien además en 1902 fue galardonado con el Premio Nobel de Fisiología y Medicina por demostrar que la malaria era contagiada por los mosquitos. Ross, en el año 1911, formuló un modelo matemático sencillo como apoyo de su argumentación que para erradicar el paludismo era suficiente con disminuir la población de mosquitos a un nivel bajo, sin necesariamente extinguirlos. Este modelo se basó en la hoy denominada ley de acción de masas, la cual establece que el número de contactos infecciosos por unidad de tiempo, es decir que producen enfermedad, es proporcional al número total de contactos entre individuos infecciosos y sanos. Con el paso de los años se han desarrollado modelos más generales y complejos para describir ciertas enfermedades infecciosas y a partir de esto modelos es posible predecir el surgimiento de epidemias y sugerir entonces por ejemplo políticas de vacunación.
- En 1952 los fisiólogos y biofísicos ingleses Alan Lloyd Hodgkin y Andrew Fielding Huxley propusieron un modelo matemático que describe los mecanismos iónicos que subyacen en la iniciación y propagación de los potenciales de acción en el axón gigante del calamar.

Consiste en un conjunto de ecuaciones diferenciales ordinarias no lineales que aproxima las características eléctricas de células excitables como las neuronas o los miocitos cardíacos. Este trabajo es reconocido como un avance crucial para el entendimiento de la excitación nerviosa y fue realmente significativo en el desarrollo de las neurociencias de la segunda mitad del Siglo XX. Por este trabajo Hodgkin y Huxley, recibieron el año 1963, el Premio Nobel en Fisiología y Medicina.

Tal como dijimos, la interacción matemática y biología o medicina no es reciente. No obstante esta interacción se ha incrementado notoriamente en los últimos años. Son varios los factores que han contribuido a que esto se lleve a cabo, entre los cuales se puede nombrar, por ejemplo, el alto grado de desarrollo tecnológico que hace que hoy sea posible realizar experimentos y mediciones antes inaccesibles y por lo tanto poder contar con grandes bases de datos e información, información que resulta difícil de entender sin el uso de herramientas matemáticas adecuadas. Otro factor, también ligado al anterior, es el desarrollo computacional actual que permite hacer cálculos y simulaciones que años atrás no eran posibles. Por último, también hay que destacar el desarrollo de herramientas y teorías matemáticas que resultan de gran utilidad a la hora de analizar y estudiar este tipo de fenómenos.

Los modelos matemáticos son imprescindibles para el estudio, análisis y comprensión de diferentes procesos y fenómenos que surgen en las ciencias médicas y biológicas. A partir de los resultados y simulaciones obtenidas basadas en modelos matemáticos es posible tener un mejor entendimiento del fenómeno analizado y también estos resultados pueden ser utilizados para estimar parámetros correspondientes al proceso en estudio o descubrir propiedades que no eran evidentes desde la experimentación. Las ecuaciones diferenciales en derivadas parciales, el modelado a partir de éstas, su resolución numérica y la creación de algoritmos son herramientas muy útiles, entre las varias existentes, para realizar lo mencionado anteriormente.

En esta trabajo se proponen principalmente dos tipos de aplicaciones biomédicas para las cuales hemos empleado diferentes herramientas matemáticas y por lo cual el trabajo está dividido en dos partes.

La primera parte está relacionada con la detección de tumores y la estimación de parámetros asociados a regiones tumorales. El objetivo es el desarrollo de una metodología para estimar la localización, tamaño y parámetros térmicos asociados a un tumor, inmerso en el tejido corporal, usando como información perfiles de temperaturas medidas sobre la superficie corporal en la zona próxima al tumor.

La relación entre la temperatura del cuerpo humano y la aparición de alguna enfermedad ha sido estudiada desde hace varios años. A grandes rasgos se puede decir que la temperatura corporal y superficial de una persona está principalmente controlada por la circulación sanguínea, el metabolismo local y el intercambio que se da entre la superficie corporal y el medio ambiente en el que se encuentra la persona [15, 19, 21]. Cualquier anormalidad que exista dentro del organismo humano que implique algún tipo de cambio en alguno de estos parámetros puede producir variaciones en la temperatura y en el flujo de calor existente en la superficie corporal. La particular estructura interna que poseen los tumores y el proceso de angiogénesis originan un incremento de la temperatura. Inflamación, aumento de la actividad metabólica, anormal morfología de nuevos vasos sanguíneos y falta de respuesta a señales de homeostasis son algunas de las características que hacen que un tumor se comporte de manera anormal con respecto al tejido sano en términos de producción y disipación de calor. Se ha determinado que la presencia de un tumor de mamas o melanoma de piel (tumor maligno que deriva de la transformación o proliferación de los melanocitos y que se encuentra predominantemente en la piel), produce un incremento de la temperatura en las zonas que lo rodean y en la superficie corporal próxima a este [51, 52, 57, 67]. Por lo tanto,

la detección de un perfil de temperatura anormal sobre la piel puede indicar la presencia interna de un tumor y esta información puede ser utilizada para predecir la localización y tamaño del tumor, como así también para determinar valores de parámetros térmicos asociados a la región tumoral.

Para modelar la transferencia de calor en tejidos humanos vamos a considerar la ecuación propuesta por Pennes en el año 1948 [62] y que hoy es considerada como un pilar fundamental en la mayoría de los estudios relacionados con aplicaciones a terapias térmicas. En el Capítulo 2 presentaremos en detalle esta ecuación. Además, considerando condiciones de bordes mixtas, probaremos la existencia y unicidad de su solución y luego propondremos un esquema de diferencias finitas de segundo orden para resolver este problema directo.

Existen varios estudios relacionados que utilizan como información mediciones realizadas sobre la superficie de un objeto para determinar características internas de éste. Por ejemplo en Ingeniería, ver [65], utilizan mediciones de temperaturas para detectar piezas defectuosas. La pieza, por ejemplo metálica, es calentada y luego su temperatura superficial es medida a través de una cámara infrarroja. Imperfecciones tales como grietas internas o superficiales causan una distribución no uniforme de la temperatura y en consecuencia, pueden estas imperfecciones ser determinadas haciendo uso de imágenes infrarrojas. En el ámbito de la Medicina, en [60] estudian la clasificación de quemaduras en la piel. Debido a que el tipo de tratamiento a seguir varía de acuerdo al grado de la lesión, es necesario saber con certeza la profundidad de la lesión para determinar el grado de la quemadura. En este trabajo exploran el uso de la temperatura de superficie de la piel para la determinación de la profundidad de las quemaduras. En cuanto a la cantidad de trabajos relacionados con la detección de tumores y estimación de parámetros asociados a éstos, podemos decir que no es tan significativa y entre los principales trabajos podemos citar a [56, 61]. Desde el punto de vista matemático, estimar la localización y tamaño o parámetros térmicos asociados a una región tumoral, inmersa en el tejido corporal, usando como información perfiles de temperaturas medidas sobre la superficie corporal, puede ser definido como un problema inverso. Dada información sobre el borde, la idea es determinar dónde está localizado el tumor y cual es su tamaño o en caso de conocer la localización del tumor la idea es poder determinar qué valor deben tener algunos de los parámetros fisiológicos asociados a esta región.

En el Capítulo 3 se presentará con más detalle la definición de problema inverso y también serán explicados de manera detallada los problemas inversos que son de interés en este trabajo. Luego de analizar la existencia, unicidad y estabilidad de las soluciones para cada uno de los problemas inversos, nos enfocaremos en nuestro objetivo principal de desarrollar una metodología para resolver tales problemas. En el mismo Capítulo y luego de reformular los problemas inversos como problemas de minimización, proponemos como primera metodología la utilización del algoritmo Pattern Search [25, 76]. Este algoritmo pertenece a la clase de algoritmos de optimización conocidos como algoritmos de búsqueda directa o también llamados de orden cero, ya que no hace uso de las derivadas para encontrar el mínimo deseado, lo cual lo convierte en un algoritmo de muy fácil implementación. La utilización de algoritmos libre del cálculo de derivadas ha sido también propuesta en [61], quienes utilizan algoritmos del tipo genético y en [56] proponen utilizar redes neuronales. La metodología que aquí proponemos y la cual está basada en el trabajo [1], es de más fácil implementación que las anteriores y sin embargo se obtienen excelentes resultados.

En el Capítulo 4 desarrollamos la segunda metodología propuesta. Esta hace uso de la información que provee la derivada del funcional a minimizar con respecto a las distintas variables a estimar. Para calcular tal derivada, debemos acudir al análisis de sensibilidad. El objetivo principal del análisis de sensibilidad es determinar de manera cuantitativa el cambio en el comportamiento de un sistema cuando, de alguna manera, un conjunto de sus variables de control son modificadas. Primero introduciremos los conceptos básicos referidos a la sensibilidad al cambio de los parámetros y con esto podremos resolver el problema de la estimación del valor de la fuente de

calor metabólico. Luego, para el problema particular de localización y estimación del tamaño del tumor, cuya variable de control es el dominio donde el problema está definido, introduciremos algunos de los conceptos relacionados con el análisis al cambio de forma. El análisis de sensibilidad al cambio de forma tradicionalmente ha estado muy ligado a la mecánica de las estructuras existiendo varias aplicaciones [33, 66]. Sin embargo, como veremos en este trabajo, estas herramientas pueden ser aplicadas en un problema relacionado a la medicina obteniendo excelentes resultados. El cálculo de la derivada de la función costo y su posterior uso permite obtener una metodología más eficiente y la cual arroja, como mostraremos en varios ejemplos, mejores resultados que la primer metodología propuesta.

El cálculo de la derivada del funcional costo y su posterior uso permitió obtener una metodología más eficiente, la cual arrojó, de acuerdo a los ejemplos mostrados, mejores resultados que la propuesta en el capítulo anterior.

La segunda parte del trabajo está relacionada con el desarrollo de una metodología destinada a la clasificación y agrupación de datos. El objetivo principal es desarrollar un algoritmo capaz de extraer, de una gran base de datos, información que reside de manera implícita en estos datos. Dicha información es previamente desconocida y puede resultar útil para describir o tomar decisiones sobre el proceso o fenómeno que está bajo análisis. En particular, aquí se aplicará para la clasificación de distintos tipos de tumores usando como información niveles de expresión genética correspondientes a distintos pacientes.

El problema de hallar patrones dentro de un conjunto de datos ha sido de gran importancia y de mucho éxito a lo largo de la historia. Por ejemplo, fue a partir de los datos de observaciones planetarias que disponía el astrónomo Tycho Brahe, que a principios del Siglo XVII el astrónomo Johannes Kepler elaboró sus tres leyes para describir el movimiento de los planetas en órbitas alrededor del Sol. Leyes que han resultado ser un pilar fundamental para el desarrollo de la mecánica celeste. En la actualidad, por ejemplo, se cuenta con una gran cantidad de estaciones meteorológicas donde se realizan mediciones de diversas variables como velocidad del viento, radiación solar, concentración de CO_2 en la atmósfera y temperatura sobre la superficie del mar, entre otras. Mediante el estudio de los fenómenos que ocurren en la atmósfera, la meteorología trata de predecir el tiempo y todo tipo de variación climática. El conocimiento de las variaciones climáticas ha sido siempre de suma importancia para el desarrollo de la agricultura, la navegación y la vida en general. En el ámbito de la medicina, más precisamente en el área de genética, un objetivo importante es entender la relación que existe entre la variación que posee un ser humano en sus secuencia de ADN y la susceptibilidad que puede tener a cierto tipo de enfermedad. Es decir, cómo los cambios en la secuencia de ADN de un individuo influyen en el riesgo de desarrollar enfermedades, como por ejemplo el cáncer. Esto es de suma importancia para ayudar a mejorar el diagnóstico, prevención y tratamiento de las enfermedades.

Podríamos decir que el campo de reconocimiento de patrones se refiere a la detección automática de regularidades existente en los datos mediante el uso de algoritmos numéricos y en base a estas regularidades realizar acciones tales como la clasificación o agrupamiento de datos.

Usualmente, en estos problemas se dispone de un conjunto de de datos $\{x_1, \dots, x_n\}$ que es denominado conjunto de entrenamiento o aprendizaje. Supongamos por ejemplo que estamos interesados en diagnosticar si una persona padece o no una cierta enfermedad denominada "E". En este caso, podemos entonces suponer que el conjunto de datos lo componen valores de ciertos análisis clínicos que se les realizaron a distintos pacientes. Usualmente, la clase o etiqueta correspondiente a cada uno de los datos de entrenamiento son conocidas de antemano, es decir para cada paciente es sabido si padece o no la enfermedad E. La idea es usar estos datos para ajustar los parámetros correspondientes a un modelo adaptativo de manera tal que cuando venga un nuevo paciente, usando los valores que arrojaron sus análisis, se pueda diagnosticar si este padece o no

la enfermedad.

El resultado de correr el algoritmo puede ser expresado como una función $y(x)$, o sea para cada entrada x tenemos como resultado un valor de salida y . La forma precisa de la función $y(x)$ es determinada durante una etapa de aprendizaje en la que los datos de aprendizaje son utilizados. Una vez que el modelo es entrenado puede ser luego empleado para diagnosticar a un paciente nuevo. A estos nuevos pacientes se los llama grupo de prueba. La capacidad del algoritmo para clasificar correctamente las nuevas observaciones, las cuales difieren de las usadas durante el entrenamiento, se denomina generalización. En las aplicaciones resulta a veces muy difícil disponer de una gran cantidad de datos de aprendizaje o a veces el problema puede ser que solo se dispone de una muestra muy pequeña de una cierta clase, por lo cual la generalización es un objetivo esencial en el reconocimiento de patrones.

Los problemas en los que las clases o etiquetas correspondientes a los datos del conjunto de aprendizaje son conocidos de antemano se conocen como problemas de aprendizaje supervisado. En el caso en que las nuevas observaciones pueden ser asignadas a un número finito de clases o categorías, estos problemas son denominados problemas de clasificación.

Otro tipo de problema se da cuando para el conjunto de datos de entrenamiento no se conocen sus etiquetas. El objetivo en este tipo de problemas de aprendizaje no supervisado, consiste en asignar las observaciones a distintos subconjuntos de manera tal que las observaciones en de cada subgrupo posean propiedades similares o sean “cercanas” en algún sentido.

En el Capítulo 5 presentamos una metodología destinada a resolver problemas de clasificación y agrupamiento de datos. La metodología propuesta implica considerar las poblaciones de entrenamiento y de prueba de manera casi indistinta. Este hecho resulta no sólo de utilidad desde el punto de vista de poder unificar los procedimientos para clasificación y agrupamiento de datos, sino que además resulta de mucha utilidad para atenuar el problema de la dimensión. Cuando se está trabajando con una muestra pequeña de datos de dimensión muy grande, de acuerdo a las ideas que aquí proponemos, es posible considerar de manera indistinta tanto a los datos de entrenamiento como a los de prueba y de esta forma contar con una única población de mayor volumen. La metodología está basada en el espíritu del algoritmo Expectation-Maximization y uno de sus pilares es la estimación de densidades de probabilidad. Este procedimiento consiste en la transformación de las muestras a través de “flujos”, a muestras correspondientes a una distribución Gaussiana isotrópica. En este contexto, las observaciones actúan como marcadores que guían el flujo. A medida que una observación es claramente asignada a una clase, esta observación comienza a cumplir un papel más activo en el flujo correspondiente a esa clase y por lo tanto actuar de manera casi pasiva en el resto.

Cabe destacar, que de la metodología propuesta se desprende un procedimiento destinado a la selección de un conjunto de variables basado en la capacidad que estas poseen para particionar la muestras en clases distintas.

Con el objetivo de demostrar el desempeño de la metodología propuesta, cerramos el Capítulo 5 utilizándola en dos ejemplos de aplicaciones médicas relacionados con la clasificación de tumores a partir de datos de niveles de expresión genética.

Finalmente, en el Capítulo 6, se presentan las conclusiones y comentarios generales sobre el presente trabajo. Además describimos brevemente la relación entre las dos partes de este trabajo y comentamos cómo la ideas y herramientas desarrolladas en la segunda parte pueden ser aplicadas para resolver los problemas inversos presentados en la primera parte utilizando un enfoque Bayesiano.

Parte I

Estimación de parámetros asociados a tumores

Capítulo 2

Localización de Tumores

2.1. Motivación del problema

La temperatura corporal es un indicador de la salud de una persona. A grandes rasgos se puede decir que la temperatura corporal y superficial de una persona está principalmente controlada por la circulación sanguínea, el metabolismo local y el intercambio que se da entre la superficie corporal y el medio ambiente en el que se encuentra la persona [15, 19, 21]. Cualquier anomalía que exista dentro del organismo humano que implique algún tipo de cambio en alguno de estos parámetros puede producir variaciones en la temperatura y en el flujo de calor existente en la superficie corporal. La particular estructura interna que poseen los tumores y el proceso de angiogénesis producen un incremento de la temperatura. Inflamación, aumento de la actividad metabólica, anormal morfología de nuevos vasos sanguíneos y falta de respuesta a señales de homeostasis son algunas de las características que hacen que un tumor se comporte de manera anormal con respecto al tejido sano en términos de producción y disipación de calor. Se ha determinado que la presencia de un tumor de mamas o melanoma de piel (tumor maligno que deriva de la transformación o proliferación de los melanocitos y se encuentra predominantemente en la piel), produce un incremento de la temperatura en las zonas que lo rodean y en la superficie corporal próxima a éste [51, 52, 57, 67]. Por lo tanto, la detección de un perfil de temperatura anormal sobre la piel podría indicar la presencia interna de un tumor y esta información puede ser utilizada para predecir la localización y su tamaño, como así también para determinar valores de parámetros térmicos de la región tumoral. El conocimiento de estos datos resulta de gran utilidad para hacer un seguimiento de la evolución de un paciente luego de haber recibido algún tipo de tratamiento o terapia con el fin de eliminar el tumor o intentar frenar su crecimiento.

En el diagnóstico por medio de imágenes térmicas, se utilizan tecnologías capaces de medir la radiación infrarroja emitida por un cuerpo y así obtener información sobre cualquier cambio de temperatura que exista en su superficie. Una de estas tecnologías es la termografía. Este es un método no invasivo, el cual consiste en medir, mediante el uso de una cámara, la radiación infrarroja emitida por la superficie corporal del paciente. Los diferentes patrones de temperatura no sólo dependen de diferentes parámetros como lo es el coeficiente de sensibilidad de la piel, sino que también están relacionados con cuestiones asociadas al proceso homeostático y metabólico, como también la estructura y dinámica de los sistemas vascular y nervioso.

El uso de imágenes termográficas en la medicina no es un fenómeno reciente sino que comenzó alrededor del año 1950. Sin embargo, su impacto no fue mayor debido principalmente a la complejidad, alto costo y escasa definición y sensibilidad que tenían las cámaras termográficas de esa época. En la actualidad, los avances relacionados con la tecnología infrarroja alientan el uso de las cámaras termográficas en el campo de la medicina como una herramienta no invasiva,

de bajo costo y de alta precisión. Un claro ejemplo de uso masivo de estas cámaras se dio en el año 2009 cuando se desató la pandemia de gripe A H1N1 más conocida como gripe porcina. Estas cámaras fueron utilizadas en la mayoría de los aeropuertos internacionales con el objetivo de detectar posibles personas que padecían la infección.

En [67] Santa Cruz *et al.* han estudiado mediante el uso de termografía la correlación que existe, en pacientes tratados con la terapia por captura neutrónica en Boro (BNCT de sus siglas en inglés), entre la reacción cutánea aguda y la distribución de dosis superficial que les fue suministrada, con el objetivo de determinar un grado de tolerancia a las dosis suministradas y en consecuencia optimizar el tratamiento por BNCT. En este trabajo también destacan la capacidad que tienen las imágenes termográficas para observar distintos aspectos funcionales del tejido humano y concluyen que el uso de la termografía puede ser de gran ayuda para localizar zonas de temperaturas anormales y con lo cual detectar nódulos de melanomas que son invisibles o difíciles de ver mediante imágenes de tomografía computarizada. Mayores detalles sobre técnicas de detección de cáncer de piel mediante el uso de imágenes infrarrojas pueden ser encontradas en el libro de Diakides and Bronzino [27].

El objetivo de esta parte del trabajo es el desarrollo de una metodología para estimar la localización, tamaño y parámetros térmicos asociados a una región tumoral, principalmente correspondiente a un melanoma maligno, inmersa en el tejido corporal, usando como información perfiles de temperaturas medidas sobre la superficie corporal en la zona próxima al tumor. Desde el punto de vista matemático, el problema a resolver puede ser definido como un problema inverso. Dada información sobre parte del borde del dominio, en este caso mediciones sobre la superficie corporal, la idea es determinar dónde debe estar localizado el tumor y su tamaño o en caso de conocer la localización del tumor entonces determinar qué valor deben tener alguno de los parámetros fisiológicos asociados a esta región, o sea en ambos casos determinar características internas del dominio. En el Capítulo 3 se presentará con más detalle la definición de problema inverso y serán explicados también detalladamente los problemas inversos que son de interés en este trabajo.

Luego de haber introducido algunos de los aspectos médicos sobre el cáncer de piel y su relación con la temperatura del cuerpo, el plan para el resto de este Capítulo 2 es el siguiente. En la próxima sección describiremos la ecuación diferencial propuesta para modelar la transferencia de calor en el cuerpo humano y a continuación, en la sección 3, probamos la existencia y unicidad de la solución para este modelo propuesto. Luego, en la sección 4 describimos el método numérico utilizado para resolver el modelo de transferencia de calor previamente planteado y por último en la sección 5 mostramos algunas simulaciones numéricas. Todos estos contenidos serán de suma importancia en el Capítulo siguiente.

2.2. Modelo matemático.

En Agosto del año 1948 Harry Pennes publica en la revista *Journal of Applied Physiology* su trabajo “Analysis of tissue and arterial blood temperatures in the resting human forearm” [62]. En este trabajo analiza la distribución de temperatura en el cuerpo humano, más precisamente considera el antebrazo humano como región de estudio para realizar sus experimentos sobre transferencia térmica. Los resultados obtenidos en este trabajo y el modelo matemático allí propuesto se han convertido en un pilar fundamental en lo que respecta a transferencia de calor en tejidos humanos. Este trabajo ha influenciado a un gran número de autores que han realizado estudios relacionados con aplicaciones a terapias térmicas utilizando la ecuación de Pennes como modelo. Algunos ejemplos que podemos citar son [24, 56, 61].

La contribución principal que realizó Pennes fue considerar la tasa de transferencia de calor

existente entre la sangre y el tejido corporal como proporcional al producto de la tasa de perfusión¹ por unidad de volumen de tejido y la diferencia entre la temperatura arterial y la temperatura del tejido local, o sea:

$$h_b = G_b \rho_b c_b (1 - \kappa)(T_b - u)$$

donde h_b representa la tasa de transferencia de calor por unidad de volumen de tejido, G_b es la tasa de perfusión por unidad de volumen de tejido, ρ_b es la densidad de la sangre, c_b es el calor específico de la sangre, κ es un factor que representa la falta de equilibrio térmico entre la sangre y el tejido, T_b es la temperatura de sangre en arteria y u es la temperatura local del tejido.

Más allá de que en un principio κ sea tal que $0 \leq \kappa \leq 1$, en [62] y en la mayoría de los trabajos se asume $\kappa = 0$, como así también considera que la temperatura arterial T_a es constante.

La ecuación propuesta incluye un término que representa la transferencia de calor por conducción a través del tejido y un término que representa la generación de calor metabólico. Estos dos términos sumados al término que representa la transferencia de calor debida a la circulación de sangre dan lugar a la siguiente ecuación de Pennes estacionaria:

$$- \operatorname{div}(\sigma(x)\nabla u(x)) + k(x)(u(x) - T_b) = q(x), \quad x \in \Omega \subset \mathbb{R}^n, \quad n = 2, 3. \quad (2.1)$$

donde σ es el coeficiente de conductividad térmica, $k = G_b \rho_b c_b$ es el coeficiente de perfusión, q es la fuente de calor metabólico y T_b es la temperatura de la sangre que suponemos de valor constante.

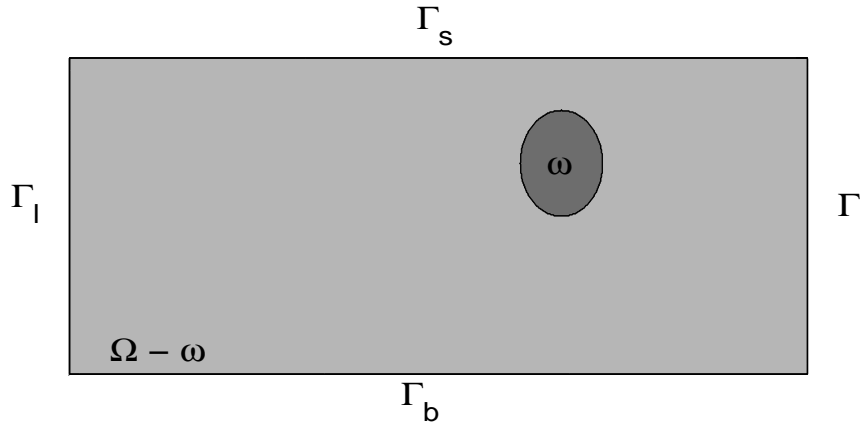


Figura 2.1: Dominio 2 dimensional: $\Omega - \bar{\omega}$ tejido sano y ω región tumoral.

Teniendo en cuenta el hecho de que los valores correspondientes a la conductividad, la perfusión y la actividad metabólica en la región tumoral son significativamente mayores a los valores encontrados en tejido sano [24, 61], vamos a modelar estos coeficientes como funciones continuas a trozos. Así por ejemplo, el coeficiente de conductividad térmica está definido como:

$$\sigma(x) = \begin{cases} \sigma_1, & \text{si } x \in \Omega - \bar{\omega}, \\ \sigma_0, & \text{si } x \in \omega, \end{cases} \quad (2.2)$$

donde ω representa la región tumoral y $\Omega - \bar{\omega}$ el tejido sano (ver Fig.1).

¹Perfusión sanguínea: llevar oxígeno y nutrientes a un tejido por medio de la sangre.

Las condiciones de contorno que adoptaremos varían en los diferentes bordes del dominio de manera de describir en forma aproximada la realidad del fenómeno que estamos considerando. En el borde inferior, el cual denotamos como Γ_b , suponemos una condición Dirichlet no homogénea:

$$u(x) = T_b, \quad x \in \Gamma_b.$$

Esto se debe a que este borde se encuentra inmerso en el cuerpo y por lo tanto suponemos que la temperatura corporal interior es constante, más precisamente de 37° .

En los laterales del dominio, la condición de borde considerada es de tipo Neumann homogénea o de flujo cero, es decir:

$$-\sigma(x) \frac{\partial u(x)}{\partial n} = 0, \quad x \in \Gamma_l,$$

donde Γ_l denota los lados laterales del dominio para $n = 2$ y las caras laterales en el caso $n = 3$. El motivo para considerar condiciones adiabáticas en los laterales se debe a que suponemos que en lugares lejos del centro del dominio, o más precisamente alejados del centro de la región tumoral, la temperatura no es afectada por ninguna fuente de calor existente dentro o fuera del dominio, por lo cual no existe un flujo de intercambio de calor.

Finalmente, en el borde superior del dominio, que representa la superficie corporal y denotamos por Γ_s , suponemos la siguiente condición convectiva o también conocida como Ley de enfriamiento de Newton:

$$-\sigma(x) \frac{\partial u(x)}{\partial n} = \alpha(u(x) - T_a) \quad x \in \Gamma_s,$$

donde α es el coeficiente de transferencia de calor y T_a es la temperatura ambiente. Esta ecuación simula el intercambio de calor que se da entre el cuerpo y el medio ambiente.

Sería posible también considerar la evaporación sobre la superficie de la piel, pero en este caso se debería considerar la evolución temporal del problema debido a que el coeficiente de humedad depende del tiempo.

Teniendo en cuenta todas las consideraciones anteriores el problema de borde que consideraremos para modelar la transferencia de calor en tejidos vivos es el siguiente:

$$\left\{ \begin{array}{ll} -\operatorname{div}(\sigma(x)\nabla u(x)) + k(x)u(x) = q(x) + k(x)T_b, & \text{en } \Omega, \\ -\sigma(x) \frac{\partial u(x)}{\partial \eta} = \alpha(u(x) - T_a), & \text{sobre } \Gamma_s, \\ -\sigma(x) \frac{\partial u(x)}{\partial \eta} = 0, & \text{sobre } \Gamma_l, \\ u(x) = T_b, & \text{sobre } \Gamma_b. \end{array} \right. \quad (2.3)$$

El modelo (2.3) recién presentado se enmarca dentro del conjunto de problemas elípticos con condiciones de borde mixtas y coeficientes discontinuos. La existencia y unicidad de la solución a este problema será tratada en la siguiente sección.

2.3. Existencia y unicidad de la solución.

2.3.1. Conceptos básicos.

Para demostrar la existencia y unicidad de la solución al problema (2.3) primero introduciremos los conceptos de derivada en sentido débil de una función, formulación débil de una ecuación

diferencial y de solución débil o generalizada. Estos conceptos son de suma importancia en la teoría de ecuaciones diferenciales debido a varias razones. Por ejemplo, la idea de solución generalizada permite estudiar ecuaciones diferenciales de la forma (2.1) donde los coeficientes no son diferenciables y más aún ni siquiera continuos, de manera tal que el lado izquierdo de la ecuación en un principio carecería de sentido. Este tipo de ecuaciones son halladas en un gran número de problemas, tal es el caso en los problemas de transmisión. Por otra parte, al plantear la formulación débil de un problema es posible aplicar herramientas y resultados bien establecidos correspondientes a la teoría del Análisis Funcional, lo cual en varios casos facilita la tarea de decidir la existencia de soluciones.

Para definir el concepto derivada en sentido débil recordemos previamente la definición de los siguientes espacios de funciones. Dado $p \geq 1$ un número real y Ω un subconjunto abierto de \mathbb{R}^n podemos introducir los siguientes espacios de funciones

$$L^p(\Omega) = \left\{ \text{“Clases de funciones” } u : \Omega \rightarrow \mathbb{R} / \int_{\Omega} |u(x)|^p dx < +\infty \right\},$$

$$L^\infty(\Omega) = \left\{ \text{“Clases de funciones” } u : \Omega \rightarrow \mathbb{R} / \exists C \text{ con } |u(x)| \leq C \text{ a.e. } x \in \Omega \right\}.$$

Ambos espacios resultan Banach con las respectivas normas

$$\|u\|_{L^p(\Omega)} = \left(\int_{\Omega} |u(x)|^p dx \right)^{\frac{1}{p}},$$

$$\|u\|_{L^\infty(\Omega)} = \inf \{ C / |u(x)| \leq C \text{ a.e. } x \in \Omega \}.$$

Denotaremos por $L^p_{loc}(\Omega)$ al conjunto de funciones u definidas en Ω y tales que para cada subconjunto abierto y acotado $\Omega' \subset \Omega$ se tiene que $u \in L^p(\Omega')$.

Por otra parte, dado un número $k \geq 0$ entero o $k = \infty$, llamaremos $C^k(\Omega)$ al conjunto de funciones que tienen derivadas de orden $\leq k$ continuas en Ω y por $C^k_0(\Omega)$ nos referiremos al conjunto de funciones en $C^k(\Omega)$ que tienen soporte compacto.

Dadas funciones $u \in C^1(\Omega)$ y $\phi \in C^\infty(\Omega)$ e integrando por partes tenemos que

$$\int_{\Omega} \partial_{x_i} u \phi dx = - \int_{\Omega} u \partial_{x_i} \phi dx + \int_{\partial\Omega} u \phi \cos(\eta, x_i) ds$$

donde ∂_{x_i} denota la derivada con respecto a la variable x_i y η la normal exterior. En el caso particular en que $\phi \in C^\infty_0(\Omega)$, la ecuación anterior se reduce simplemente a

$$\int_{\Omega} \partial_{x_i} u \phi dx = - \int_{\Omega} u \partial_{x_i} \phi dx.$$

La integral sobre el borde desaparece ya que ϕ tiene soporte compacto en Ω con lo cual se anula cerca de $\partial\Omega$. De manera general, dado k un número entero positivo, funciones $u \in C^k(\Omega)$ y $\phi \in C^\infty_0(\Omega)$ y un multi-índice $\alpha = (\alpha_1, \dots, \alpha_n)$ con $\alpha_i \in \mathbb{N}_0$ y $k = |\alpha| = \sum_{j=1}^n \alpha_j$, tenemos que

$$\int_{\Omega} \partial_\alpha u \phi dx = (-1)^{|\alpha|} \int_{\Omega} u \partial_\alpha \phi dx. \quad (2.4)$$

Basándonos en esta última ecuación (2.4), podemos ahora dar la definición de derivada débil o generalizada de una función. Notemos que para que el lado derecho de la ecuación (2.4) tenga sentido sólo basta pedir que u sea localmente integrable. Ahora, si u no está en $C^k(\Omega)$ la expresión $\partial_\alpha u$ en el término izquierdo de la ecuación carece de significado. Sin embargo, podemos resolver

esta dificultad pidiendo que exista una función v que sea localmente integrable y tal que la fórmula (2.4) resulte válida reemplazando v por $\partial_\alpha u$.

Definición: Dadas $u, v \in L^1_{loc}(\Omega)$ y α un multi-índice. Se dice que v es la derivada α -ésima débil de u si

$$\int_{\Omega} v \phi \, dx = (-1)^{|\alpha|} \int_{\Omega} u \partial_\alpha \phi \, dx \quad \forall \phi \in C_0^\infty(\Omega), \quad (2.5)$$

y decimos que $\partial_\alpha u = v$ en sentido débil.

Observación: en el caso en que la función $u \in C^{|\alpha|}(\Omega)$, es decir que existan las derivadas usuales de u de orden $\leq |\alpha|$, la α -ésima derivada en sentido débil de u es justamente la α -ésima derivada de u en el sentido usual.

Definición: Dado Ω subconjunto abierto de \mathbb{R}^n denotamos por $W_p^k(\Omega)$ el espacio de Sobolev definido por

$$W_p^k = \{u \in L^p(\Omega) / \partial_\alpha u \in L^p(\Omega) \quad \forall |\alpha| \leq k\}. \quad (2.6)$$

Cabe aclarar que las derivadas son en sentido débil. En el caso particular en que $p = 2$ denotaremos $W_p^k(\Omega) = H^k(\Omega)$. Para nosotros el espacio de mayor interés será

$$H^1(\Omega) = \{u \in L^2(\Omega) / \partial_{x_i} u \in L^2(\Omega) \quad \forall i = 1, \dots, n\}.$$

Observemos que si u y v están en $H^1(\Omega)$, podemos definir el siguiente producto escalar

$$(u, v)_{H^1(\Omega)} := \int_{\Omega} uv \, dx + \sum_{i=1}^n \int_{\Omega} \partial_{x_i} u \partial_{x_i} v \, dx. \quad (2.7)$$

De esta manera el espacio $H^1(\Omega)$ resulta un espacio de Hilbert con la norma

$$\|u\|_{H^1(\Omega)} := \left(\int_{\Omega} u^2 \, dx + \sum_{i=1}^n \int_{\Omega} (\partial_{x_i} u)^2 \, dx \right)^{\frac{1}{2}}. \quad (2.8)$$

Por último, con $H_0^1(\Omega)$ vamos a representar el subespacio de $H^1(\Omega)$ definido como la clausura de $C_0^\infty(\Omega)$ en $H^1(\Omega)$. Es decir, $u \in H_0^1(\Omega)$ si y sólo si existen funciones $u_m \in C_0^\infty(\Omega)$ tales que $u_m \rightarrow u$ en $H^1(\Omega)$. Otra manera de interpretar $H_0^1(\Omega)$, es pensar en las funciones $u \in H^1(\Omega)$ tales que $u = 0$ sobre $\partial\Omega$ en el sentido de la traza, ver [29].

Para definir ahora los conceptos de formulación débil de una ecuación diferencial y solución generalizada analicemos el siguiente ejemplo.

Consideremos el problema de hallar la posición de equilibrio de una membrana elástica homogénea, la cual está sujeta a lo largo de su borde y que además está expuesta a una fuerza externa f , que por el momento consideramos continua en $\partial\Omega$. La ecuación que modela este problema es la siguiente:

$$\begin{cases} \Delta u = f, & \text{en } \Omega, \\ u = 0, & \text{sobre } \partial\Omega. \end{cases} \quad (2.9)$$

Si $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ satisface (2.9) se dice que u es una solución clásica de este problema. Multiplicado (2.9) por $\phi \in C_0^\infty(\Omega)$ e integrando en Ω obtenemos

$$\int_{\Omega} \Delta u \phi \, dx = \int_{\Omega} f \phi \, dx,$$

ecuación que puede ser reescrita como

$$\int_{\Omega} \sum_{i=1}^n \partial_{x_i}^2 u \phi \, dx = \int_{\Omega} f \phi \, dx.$$

Ahora, si integramos por partes y tenemos en cuenta el hecho que $\phi \in C_0^{\infty}(\Omega)$ obtenemos

$$- \int_{\Omega} \sum_{i=1}^n \partial_{x_i} u \partial_{x_i} \phi \, dx = \int_{\Omega} f \phi \, dx.$$

Por lo tanto, tenemos que si $u \in C^2(\Omega) \cap C^0(\overline{\Omega})$ satisface (2.9) también satisface la ecuación

$$- \int_{\Omega} \nabla u \cdot \nabla \phi \, dx = \int_{\Omega} f \phi \, dx \quad \forall \phi \in C_0^{\infty}(\Omega).$$

Notemos que en esta última ecuación sólo aparecen las derivadas de primer orden de la función u y que además esta ecuación tiene sentido si en vez de considerar $u \in C^2(\Omega)$ y tal que $u = 0$ punto a punto, consideramos $u \in H_0^1(\Omega)$. Además teniendo en cuenta el hecho de que $C_0^{\infty}(\Omega)$ es denso en $H_0^1(\Omega)$, podemos entonces dar la siguiente definición.

Definición: Dada $u \in H_0^1(\Omega)$, se dice que u es una solución débil o generalizada del problema (2.9) si

$$\int_{\Omega} -\nabla u \cdot \nabla v \, dx = \int_{\Omega} f v \, dx \quad \forall v \in H_0^1(\Omega), \quad (2.10)$$

donde las derivadas se entienden en sentido débil.

Es claro entonces que con esta definición toda solución clásica es a su vez solución débil. Recíprocamente si $u \in H_0^1(\Omega)$ satisface la identidad (2.10) para $v \in H_0^1(\Omega)$ arbitraria, entonces u satisface la ecuación (2.9) para casi todo punto. Estos hechos reflejan que el concepto de solución generalizada de la ecuación (2.9) es simplemente una extensión del viejo concepto de solución en sentido clásico.

Observación: En el ejemplo correspondiente a hallar el estado de equilibrio de la membrana, sabemos que la energía potencial de la membrana está dada por la integral

$$I(u) = \int_{\Omega} \left(\frac{1}{2} |\nabla u|^2 + u f \right) dx,$$

y el estado de equilibrio, de acuerdo al principio de Hamilton, está determinado por la condición que en tal estado el funcional $I(u)$ alcanza el menor valor posible comparado con todo estado $v \in H_0^1(\Omega)$. Por lo tanto, en el estado de equilibrio u se tiene que

$$\delta I(u) = \int_{\Omega} (\nabla u \cdot \nabla v + f v) \, dx = 0 \quad \forall v \in H_0^1(\Omega). \quad (2.11)$$

Entonces, de acuerdo al principio de Hamilton, el problema de hallar la posición de equilibrio de la membrana se reduce a hallar la función u en el espacio $H_0^1(\Omega)$ que satisface la identidad integral (2.11). De esta manera podemos descartar el requerimiento extra, no inherente a la naturaleza física del problema, que la solución posea derivadas de orden dos veces mayor a las derivadas que aparecen en la integral de la energía $I(u)$. Este concepto más amplio de solución generalizada permite entonces debilitar las condiciones sobre u y por lo tanto facilita la tarea de hallar una solución al problema en cuestión.

De manera general, dado el siguiente problema de borde:

$$\begin{cases} -\sum_{i,j=1}^n \frac{\partial}{\partial x_i} \left(a_{ij}(x) \frac{\partial u}{\partial x_j} \right) + b(x)u = f(x), & \text{en } \Omega, \\ \sum_{i,j=1}^n a_{ij}(x) \frac{\partial u}{\partial x_j} \eta_i = g(x), & \text{sobre } \Gamma, \\ u = 0, & \text{sobre } \Gamma', \end{cases} \quad (2.12)$$

donde Ω es un abierto acotado de \mathbb{R}^n , $\Gamma, \Gamma' \subset \partial\Omega$ abiertos disjuntos tales que $\partial\Omega = \Gamma \cup \Gamma'$ y los coeficientes a_{ij}, b son acotados y medibles. Dada $u \in H_{\Gamma'}^1(\Omega)$, donde $H_{\Gamma'}^1(\Omega)$ son las funciones v que pertenecen a $H^1(\Omega)$ y $v = 0$ sobre Γ' en el sentido de la traza [29]. Decimos que u es una solución generalizada del problema (2.12) si satisface la identidad:

$$\int_{\Omega} \sum_{i,j=1}^n a_{ij}(x) \partial_{x_j} u \partial_{x_i} v + b(x)uv \, dx = \int_{\Omega} f v \, dx + \int_{\partial\Omega} g(x)v \, ds \quad \forall v \in H_{\Gamma'}^1(\Omega). \quad (2.13)$$

Esta ecuación suele denominarse formulación débil del problema (2.12).

Notemos que el lado izquierdo de (2.13) define una forma bilineal $a : H_{\Gamma'}^1(\Omega) \times H_{\Gamma'}^1(\Omega) \rightarrow \mathbb{R}$ dada por

$$a(u, v) = \int_{\Omega} \sum_{i,j=1}^n a_{ij}(x) \partial_{x_j} u \partial_{x_i} v + b(x)uv \, dx,$$

mientras que en el lado derecho de la identidad tenemos la forma lineal $l : H_{\Gamma'}^1(\Omega) \rightarrow \mathbb{R}$ definida por

$$l(v) = \int_{\Omega} f v \, dx + \int_{\partial\Omega} g(x)v \, ds.$$

Por lo tanto, hallar la solución en sentido débil correspondiente al problema (2.12) equivale a hallar $u \in H_{\Gamma'}^1(\Omega)$ tal que

$$a(u, v) = l(v) \quad \forall v \in H_{\Gamma'}^1(\Omega).$$

Mediante la formulación débil de una ecuación diferencial es posible apelar a la estructura del espacio vectorial donde se hallan las posibles soluciones, por ejemplo los espacios de Sobolev W_p^k , y utilizar distintas herramientas que facilitan la tarea de hallar la solución buscada. Esta manera de escribir el problema será de utilidad en la proxima sección y en capítulos posteriores.

2.3.2. Existencia y unicidad de la solución.

En esta sección probaremos la existencia y unicidad de la solución al problema (2.3). Para esto será necesario utilizar uno de los resultados más conocidos en la teoría del análisis funcional y que enunciamos a continuación.

Teorema(Lax-Milgram): Sean \mathcal{V} un espacio de Hilbert, $a : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ una forma bilineal tal que

1. $|a(u, v)| \leq C \|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}} \quad \forall u, v \in \mathcal{V}$ (continua) y
2. $a(u, u) \geq c \|u\|_{\mathcal{V}}^2 \quad \forall u, v \in \mathcal{V}$ (coercitiva),

donde C_1 y C_2 son dos constantes positivas y $l : \mathcal{V} \rightarrow \mathbb{R}$ una forma lineal perteneciente al espacio dual \mathcal{V}^* . Entonces existe única $u \in \mathcal{V}$ tal que

$$a(u, v) = l(v), \quad \forall v \in \mathcal{V}. \quad (2.14)$$

Para hacer uso del resultado de este Teorema, debemos primero escribir la formulación débil del problema (2.3). De acuerdo a lo presentado en la sección previa decimos que $u \in H^1(\Omega)$ es solución débil de (2.3) si satisface la ecuación

$$\int_{\Omega} (\sigma \nabla u \cdot \nabla v + kuv) dx - \int_{\partial\Omega} \sigma(\nabla u \cdot \eta) d\Gamma = \int_{\Omega} (q + kT_b)v dx \quad \forall v \in H^1(\Omega). \quad (2.15)$$

Ahora como $\partial\Omega = \Gamma_b \cup \Gamma_s \cup \Gamma_l$, y apelando a cada una de las condiciones de borde, la integral sobre la frontera puede ser escrita como

$$\int_{\partial\Omega} \sigma(\nabla u \cdot \eta)v d\Gamma = -\alpha \int_{\Gamma_s} (u - T_a)v d\Gamma + \int_{\Gamma_b} \sigma(\nabla u \cdot \eta)v d\Gamma.$$

El último término en el lado derecho de la ecuación aparece debido a la condición de Dirichlet no homogénea impuesta sobre Γ_b . Sin embargo, podemos escribir nuestro problema como si estuviésemos tratando con uno con condiciones de borde homogéneas sobre Γ_b . En efecto, primero definimos el siguiente espacio de Hilbert²:

$$\mathcal{V}(\Omega) = \{u \in H^1(\Omega) : u = 0 \text{ sobre } \Gamma_b\}, \quad (2.16)$$

y consideramos la siguiente forma bilineal y forma lineal definidas respectivamente por:

$$\begin{aligned} a(u, v) &= \int_{\Omega} \sigma(x) \nabla u \cdot \nabla v dx + \int_{\Omega} k(x)uv dx + \alpha \int_{\Gamma_s} uv d\Gamma, \\ \tilde{l}(v) &= \int_{\Omega} (q(x) + k(x)T_b)v dx + \alpha T_a \int_{\Gamma_s} v d\Gamma. \end{aligned}$$

Luego, si $u_b \in H^1(\Omega)$ es tal que $u_b = T_b$ sobre Γ_b en el sentido de la traza, tenemos que encontrar una solución de (2.3) se reduce a resolver el siguiente problema:

$$\begin{cases} \text{Hallar } u \in \mathcal{V}(\Omega) \text{ tal que} \\ a(u, v) = \tilde{l}(v) - a(u_b, v), \quad \forall v \in \mathcal{V}(\Omega). \end{cases}$$

ya que entonces la función $\tilde{u} = u + u_b$ es solución débil de (2.3). En este trabajo tomaremos $u_b \equiv T_b$ en Ω , con lo cual finalmente la formulación débil que nos interesa es la siguiente:

$$\begin{cases} \text{Hallar } u \in \mathcal{V}(\Omega) \text{ tal que} \\ a(u, v) = l(v), \quad \forall v \in \mathcal{V}(\Omega), \end{cases}$$

donde

$$a(u, v) = \int_{\Omega} \sigma(x) \nabla u \cdot \nabla v dx + \int_{\Omega} k(x)uv dx + \alpha \int_{\Gamma_s} uv d\Gamma, \quad (2.17)$$

$$\begin{aligned} l(v) &= \tilde{l}(v) - a(T_b, v) \\ &= \int_{\Omega} q(x)v dx + \alpha(T_a - T_b) \int_{\Gamma_s} v d\Gamma. \end{aligned} \quad (2.18)$$

²La ecuación sobre Γ_b vale en el sentido de la traza [29]

Ya definidas la forma bilineal y forma lineal el paso siguiente es ver que éstas satisfacen la hipótesis del Teorema de Lax-Milgram.

Primero veamos que la forma bilineal $a : \mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ definida por la ecuación (2.17) es continua. Aplicando la desigualdad de Hölder tenemos que

$$\begin{aligned} |a(u, v)| &\leq \int_{\Omega} (|\sigma| |\nabla u| |\nabla v| + |k| |u| |v|) dx + \int_{\Gamma_s} |u| |v| d\Gamma \\ &\leq c_1 (\|\nabla u\|_{L^2(\Omega)} \|\nabla v\|_{L^2(\Omega)} + \|u\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)}) + c_2 \|u\|_{L^2(\partial\Omega)} \|v\|_{L^2(\partial\Omega)}. \end{aligned}$$

donde c_1 y c_2 son constantes positivas independientes de u y v . Luego, por el Teorema de la Traza [29], sabemos que

$$\|v\|_{L^2(\partial\Omega)} \leq \tilde{C} \|v\|_{H^1(\Omega)} \quad \forall v \in H^1(\Omega),$$

con \tilde{C} constante positiva que no depende de v . Entonces, si aplicamos este resultado a los elementos del subespacio $\mathcal{V}(\Omega)$ de $H^1(\Omega)$ y consideramos la norma que hereda de este espacio, obtenemos que

$$\begin{aligned} |a(u, v)| &\leq \tilde{c}_1 (\|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}} + \|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}}) + \tilde{c}_2 \|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}} \\ &\leq C \|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}}. \end{aligned} \quad (2.19)$$

De este resultado concluimos que a es continua. Para ver que es coercitiva, primero utilizamos la desigualdad de Poincaré y luego usamos que la ecuación es estrictamente elíptica, obteniendo

$$\begin{aligned} c \|u\|_{\mathcal{V}} &\leq \tilde{c} \int_{\Omega} |\nabla u|^2 dx \\ &\leq \int_{\Omega} \sigma(x) \nabla u \cdot \nabla v dx \\ &= a(u, u) - \int_{\Omega} k u^2 dx - \alpha \int_{\Gamma_s} u^2 \\ &\leq a(u, u) \end{aligned} \quad (2.20)$$

En la última desigualdad hemos usado el hecho que los coeficientes k y α son positivos.

Para ver que la forma lineal es continua, aplicamos nuevamente la desigualdad de Hölder, el Teorema de la Traza y la definición de norma en $\mathcal{V}(\Omega)$ obteniendo

$$\begin{aligned} |l(v)| &\leq \int_{\Omega} |q| |v| dx + \alpha |T_a - T_b| \int_{\Gamma_s} |v| d\Gamma \\ &\leq C_1 \|v\|_{L^2(\Omega)} + C_2 \|v\|_{L^2(\partial\Omega)} \\ &\leq C_3 \|v\|_{\mathcal{V}}. \end{aligned} \quad (2.21)$$

Entonces, de (2.19), (2.20) y (2.21) vemos que las hipótesis del teorema de Lax-Milgram son válidas, con lo cual podemos afirmar que el problema (2.3) admite solución $u \in H^1(\Omega)$ y esta solución es única.

Observación: por regularidad [29], la solución positiva $u \in H^1(\Omega)$ también pertenece al espacio $W_{loc}^{2,p}(\Omega)$ para todo $p > 1$ y por lo tanto $u \in C_{loc}^{1,\alpha}(\Omega)$ para $0 < \alpha < 1$. Esto significa que para todo subdominio $\Omega' \subset \overline{\Omega'} \subset \Omega$, la función u tiene primera y segunda derivada en sentido débil perteneciente a $L^p(\Omega)$ y más aún la derivada primera en el sentido usual resulta continua en Ω'

Para cerrar esta sección queremos destacar que si u es la solución del problema (2.3) y si denotamos por $u_1 = u|_{\Omega - \bar{\omega}}$ y por $u_0 = u|_{\omega}$ entonces toda solución de (2.3) es solución del siguiente problema:

$$\left\{ \begin{array}{ll} -\sigma_1 \Delta u_1 + k_1 u_1 = q_1 + k_1 T_b, & \text{en } \Omega - \bar{\omega}, \\ -\sigma_0 \Delta u_0 + k_0 u_0 = q_0 + k_0 T_b, & \text{en } \omega, \\ u_1 = u_0, & \text{sobre } \partial\omega, \\ -\sigma_1 \frac{\partial u_1}{\partial \eta} = -\sigma_0 \frac{\partial u_0}{\partial \eta}, & \text{sobre } \partial\omega, \\ -\sigma_1 \frac{\partial u_1}{\partial \eta} = \alpha(u_1 - T_a), & \text{sobre } \Gamma_s, \\ -\sigma_1 \frac{\partial u_1}{\partial \eta} = 0, & \text{sobre } \Gamma_l, \\ u_1 = T_b, & \text{sobre } \Gamma_b. \end{array} \right. \quad (2.22)$$

Este tipo de problema es conocido en la literatura como problema de transmisión y toma su nombre debido justamente a las condiciones de transmisión impuestas sobre la frontera $\partial\omega$. Principalmente este tipo de formulación corresponde a problemas físicos en los que existen dos o más medios de diferentes características, por ejemplo compuestos por diferentes sustancias.

2.4. Resolución Numérica

En esta sección presentamos y describimos brevemente el método de diferencias finitas, método numérico que emplearemos para resolver el problema de borde planteado en la sección previa.

El método de diferencias finitas (MDF) se basa en aproximar una ecuación diferencial o más precisamente la derivada de una función por una fórmula algebraica. Esta fórmula algebraica representa una discretización del cociente incremental por medio de una ecuación en diferencias.

Para ser más claro, consideremos $f(x)$ una función $n + 1$ veces derivable en el intervalo (a, b) . Por el Teorema de Taylor tenemos que para todo $x \in (a, b)$ y h suficientemente pequeño de manera tal que $x + h \in (a, b)$

$$f(x+h) = f(x) + \frac{f'(x)}{1!}h + \frac{f^{(2)}(x)}{2!}h^2 + \dots + \frac{f^{(n)}(x)}{n!}h^n + R_n(x+h),$$

donde $n!$ denota el factorial de n y $R_n(x)$ es el resto o residuo, el cual representa la diferencia entre el polinomio de Taylor de grado n y el valor original de la función. Ahora, si nos restringimos sólo a la derivada de primer orden de la función tenemos

$$f(x+h) = f(x) + f'(x)h + R_1(x+h),$$

lo cual es equivalente a

$$f'(x) = \frac{f(x+h) - f(x)}{h} - \frac{R_1(x+h)}{h}.$$

Por lo tanto, para h suficientemente pequeño, obtenemos

$$f'(x) \approx \frac{f(x+h) - f(x)}{h}.$$

Esta última ecuación nos servirá como base para definir los distintos tipos de diferencias. Primero, discretizamos el intervalo donde está definida la función obteniendo así una grilla determinada por los puntos o nodos x_i , con $i = 1, \dots, M$ y donde $x_1 = a$ y $x_M = b$. Asumimos además que

$x_i - x_{i-1} = h$, es decir una grilla igualmente espaciada o equiespaciada. Definimos, para $i = 2, \dots, M-1$, la **diferencia hacia adelante** o **forward difference** mediante la ecuación

$$D_+f(x_i) = \frac{f(x_{i+1}) - f(x_i)}{h} = f'(x_i) + \frac{1}{2}f''(\xi)h, \quad (2.23)$$

donde ξ es un punto perteneciente al intervalo (x_i, x_{i+1}) . Del mismo modo definimos la **diferencia hacia atrás** o **backward difference** como:

$$D_-f(x_i) = \frac{f(x_i) - f(x_{i-1})}{h} = f'(x_i) - \frac{1}{2}f''(\xi)h, \quad (2.24)$$

en este caso con $\xi \in (x_{i-1}, x_i)$. Por último, definimos la **diferencia central** o **simétrica** mediante la ecuación:

$$D_0f(x_i) = \frac{f(x_{i+1}) - f(x_{i-1}))}{2h} = f'(x_i) + \frac{h^2}{6}f'''(\xi). \quad (2.25)$$

El error cometido al utilizar un método numérico puede definirse como la diferencia que existe entre el valor exacto de la solución analítica y el valor obtenido utilizando la aproximación numérica. Hay dos fuentes de generación de errores en el método de diferencias finitas y en general en todo método numérico. La primera es la relacionada a los errores de redondeo debido a la pérdida o falta de precisión causada por la representación decimal finita realizada por la computadora. La otra fuente está asociada con el error de truncamiento o discretización, el cual se puede definir como la diferencia entre el valor exacto y el valor obtenido utilizando diferencias finitas suponiendo aritmética perfecta, es decir suponiendo que no existen errores de redondeo. Para ser más precisos si $f'(x_i)$ denota el valor exacto de la derivada en el punto x_i y $Df(x_i)$ representa el valor calculado utilizando diferencias finitas, el error de truncamiento es $f'(x_i) - Df(x_i)$. Por lo tanto, de acuerdo a las ecuaciones (2.23) y (2.24) tenemos que el error de truncamiento cometido usando diferencias hacia adelante o hacia atrás es proporcional al paso h . En este caso se dice que el error es de primer orden y se denota por $O(h)$. Por su parte para el caso de diferencia central el error de truncamiento resulta ser proporcional a h^2 , o sea que para h pequeño se comete un error menor que en los casos anteriores. En este caso se dice que el error es de segundo orden y se denota por $O(h^2)$.

Las fórmulas de diferencias finitas para aproximar derivadas de mayor orden pueden ser obtenidas usando las fórmulas para derivadas de menor orden. Por ejemplo, la fórmula de diferencia central para aproximar una derivada de segundo orden f'' está dada por la siguiente expresión:

$$\begin{aligned} D_0^2f(x_i) &= D_+D_-f(x_i) = D_+\frac{f(x_i) - f(x_{i-1}))}{h} \\ &= \frac{f(x_{i+1}) - 2f(x_i) + f(x_{i-1}))}{h^2} = f''(x_i) + \frac{h^2}{12}f^{iv}(\xi). \end{aligned}$$

Notemos que nuevamente la diferencia central resulta ser una aproximación de segundo orden.

En el caso en que estemos tratando con una función de n variables, las derivadas parciales se definen de manera análoga. Para nuestro caso particular, si consideramos $n = 2$, la aproximación del Laplaciano de la función $u(x, y)$ en el punto (x_i, y_j) está dada por la fórmula

$$\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right)(x_i, y_j) = \frac{u(x_{i+1}, y_j) - 2u(x_i, y_j) + u(x_{i-1}, y_j))}{h^2} + \frac{u(x_i, y_{j+1}) - 2u(x_i, y_j) + u(x_i, y_{j-1}))}{h^2}. \quad (2.26)$$

Por simplicidad en la notación, a continuación desarrollaremos la aproximación mediante el MDF del problema (2.22) para el caso $n = 2$, notando que para dimensión $n = 3$ se procede de manera similar.

Consideremos nuevamente una discretización del dominio en las direcciones x e y tal que $\Delta x = \Delta y = h$ y denotemos los puntos correspondientes a la dirección o eje x por $i = 1, \dots, n_x$ mientras que a los del eje y los denotaremos por $j = 1, \dots, n_y$. Con esta nueva notación la ecuación (2.26) es simplemente:

$$\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} \right) \Big|_{i,j} = \frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2}, \quad (2.27)$$

Usando diferencias centrales, la ecuación diferencial definida en el interior del dominio resulta ser:

$$-\sigma_{i,j} \left(\frac{u_{i+1,j} - 2u_{i,j} + u_{i-1,j}}{h^2} + \frac{u_{i,j+1} - 2u_{i,j} + u_{i,j-1}}{h^2} \right) + k_{i,j} u_{i,j} = q_{i,j} + k_{i,j} T_b,$$

ecuación que puede ser reescrita, operaciones algebraicas mediante, como:

$$u_{i,j} = \frac{\sigma_{i,j}}{4\sigma_{i,j} + h^2 k_{i,j}} (u_{i+1,j} + u_{i-1,j} + u_{i,j+1} + u_{i,j-1}) + \frac{h^2}{4\sigma_{i,j} + h^2 k_{i,j}} (q_{i,j} + k_{i,j} T_b). \quad (2.28)$$

Una vez discretizada la ecuación diferencial en el interior del dominio, debemos ahora aproximar cada una de las condiciones impuestas en los bordes del dominio. Teniendo en cuenta que las derivadas parciales en el Laplaciano fueron aproximadas por un esquema de diferencias finitas de segundo orden, las correspondientes derivadas parciales en los bordes también serán aproximadas por un esquema de segundo orden. Para realizar esto consideramos puntos de grilla ficticios fuera del dominio y luego reemplazamos estos puntos en la ecuación principal. A continuación describimos cada una de las condiciones de borde.

En primer lugar en el lateral izquierdo del dominio, el cual corresponde a los nodos con $i = 1$, tenemos la condición de flujo cero:

$$\frac{\partial u}{\partial x} \Big|_{1,j} = \frac{u_{2,j} - u_{0,j}}{2h} = 0, \quad j = 2, \dots, n_y - 1,$$

por lo tanto $u_{0,j} = u_{2,j}$, luego reemplazando esto en la ecuación (2.28) y por simplicidad denotando $\beta_{i,j} = 4\sigma_{i,j} + h^2 k_{i,j}$, obtenemos:

$$u_{1,j} = \frac{\sigma_{1,j}}{\beta_{1,j}} (u_{1,j+1} + 2u_{2,j} + u_{1,j-1}) + \frac{h^2}{\beta_{1,j}} (q_{1,j} + k_{1,j} T_b) \quad j = 2, \dots, n_y - 1. \quad (2.29)$$

De manera similar, para la condición impuesta en el lateral derecho, donde $i = n_x$, tenemos:

$$u_{n_x,j} = \frac{\sigma_{n_x,j}}{\beta_{n_x,j}} (u_{n_x,j+1} + 2u_{n_x-1,j} + u_{n_x,j-1}) + \frac{h^2}{\beta_{n_x,j}} (q_{n_x,j} + k_{n_x,j} T_b) \quad j = 2, \dots, n_y - 1. \quad (2.30)$$

Por otra parte, la condición convectiva impuesta en el borde superior del dominio se traduce en:

$$u_{i,n_y} = \frac{\sigma_{i,n_y}}{\beta_{i,n_y} + 2h\alpha} (2u_{i,n_y-1} + u_{i+1,n_y} + u_{i-1,n_y} + \frac{2h\alpha T_a}{\lambda_{i,n_y}}) + \frac{h^2}{\beta_{i,n_y} + 2h\alpha} (q_{i,n_y} + k_{i,n_y} T_b) \quad i = 2, \dots, n_x - 1. \quad (2.31)$$

Finalmente, la condición Dirichlet no homogénea correspondiente al borde inferior es simplemente:

$$u_{i,1} = T_b \quad i = 2, \dots, n_x - 1. \quad (2.32)$$

Una vez construida la aproximación numérica del problema de borde, definimos la iteración $u_{i,j}^{n+1} = F(u_{i,j}^n)$, donde F representa el lado derecho correspondiente a las ecuaciones (2.28-2.32). Es decir, dado un valor inicial $u_{i,j}^0$, calculamos los valores de $u_{i,j}^{n+1}$ en todo el dominio hasta que se cumpla

$$\max_{i,j} e(i,j) < \varepsilon,$$

donde $e(i,j)$ es el error relativo en el nodo (i,j) definido por

$$e(i,j) = \frac{|u_{i,j}^{n+1} - u_{i,j}^n|}{|u_{i,j}^n|},$$

y con ε representamos el error de tolerancia deseado.

2.5. Ejemplos numéricos

Para mostrar el funcionamiento del método numérico propuesto en la sección previa presentamos aquí algunas soluciones numéricas de la ecuación (2.3) con sus correspondientes condiciones de borde. En todos los ejemplos, los siguientes valores para los coeficientes y parámetros fisiológicos son considerados [24, 56, 61]:

$$\begin{aligned} \sigma_1 &= 0,5[W/m \text{ } ^\circ C], & k_1 &= 1998,1[W/m^3 \text{ } ^\circ C], & Q_1 &= 4200[W/m^3], \\ \sigma_2 &= 0,75[W/m \text{ } ^\circ C], & k_2 &= 7992,4[W/m^3 \text{ } ^\circ C], & Q_2 &= 42000[W/m^3], \\ T_b &= 37 \text{ } ^\circ C, & T_a &= 25 \text{ } ^\circ C, & \alpha &= 10[W/m^2 \text{ } ^\circ C]. \end{aligned}$$

donde W representa la unidad de potencia considerada en watts, m la unidad de longitud en metros y $^\circ C$ representa la unidad de temperatura en grados Celsius.

En los ejemplos presentados a continuación consideraremos $n = 2$ y un dominio de dimensiones $0.09 \times 0.03[m]$, mientras que el paso de la discretización es igual a $h = 3 * 10^{-4}[m]$.

En el primer ejemplo que mostramos, el centro del tumor se encuentra localizado en el punto $(0.045, 0.020)$ y el radio es igual a $0.005[m]$. En la Figura 2.2a podemos ver la distribución de temperatura sobre todo el dominio. Se observa que comienza a $37^\circ C$ en el borde inferior del dominio y decrece debido a la condición convectiva impuesta en la superficie corporal, pero presenta un importante incremento en la región donde se localiza el tumor. En la Figura 2.2b vemos el perfil de temperatura sobre la superficie de la piel. Se observa una diferencia de temperatura de aproximadamente $1^\circ C$ entre la región que está situada encima del tumor y las regiones que se encuentran distantes de éste. Esto concuerda con el hecho de que la presencia de un tumor altamente vascularizado puede inducir un incremento de la temperatura en la superficie corporal.

En la Figura 2.3 presentamos otro ejemplo 2-dimensional. En este caso el tumor está centrado en el punto $(0.03, 0.022)$ y su radio es igual a $0.004[m]$. En la Figura 2.3a se observa la distribución de temperatura sobre todo el dominio, mientras que en la Figura 2.3b se muestra solamente el perfil de temperatura en el borde superior que representa la superficie corporal. Nuevamente, podemos ver un diferencia de casi $1^\circ C$ entre la zona localizada inmediatamente sobre el tumor y zonas alejadas.

Para el caso en dimensión 3, presentamos dos ejemplos que muestran los resultados utilizando el método de resolución propuesto. En ambos ejemplos las dimensiones del dominio asumidas son $0.09 \times 0.09 \times 0.03 [m]$ y nuevamente un paso de discretización $h = 3 * 10^{-4}[m]$. En el primer ejemplo el centro es localizado en el punto $(0.045, 0.045, 0.002)$ mientras que el radio es igual a $0.06 [m]$. En la Figura 2.4 vemos el perfil de temperatura sobre la superficie corporal. Esta Figura

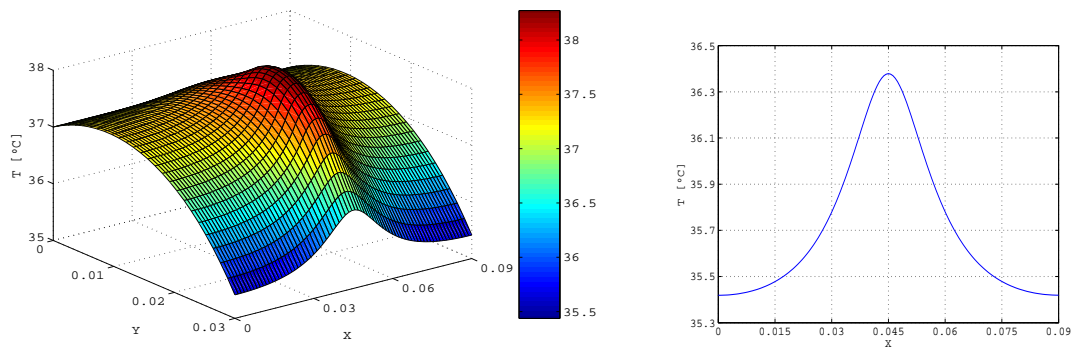


Figura 2.2: (a) Distribución de la temperatura. (b) Perfil de la temperatura superficial.

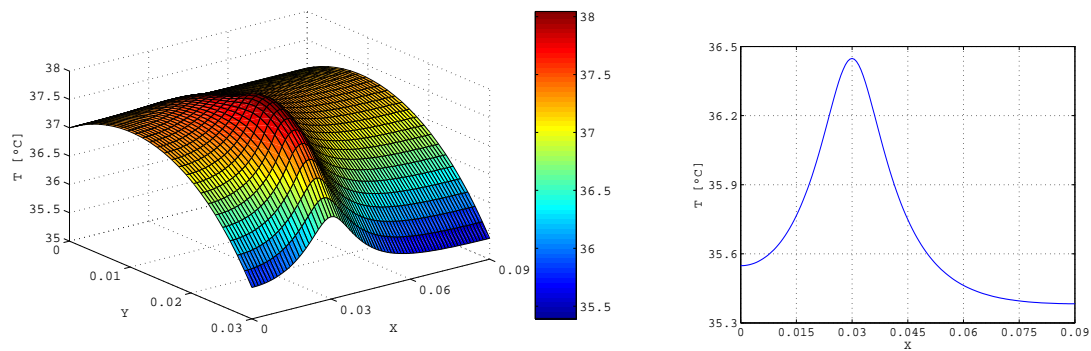


Figura 2.3: (a) Distribución de la temperatura. (b) Perfil de la temperatura superficial.

revela una diferencia de aproximadamente 0.5°C entre áreas sobre la región tumoral y áreas sobre tejido sano.

Por último, la Figura 2.5 muestra otro ejemplo 3-dimensional. En este caso el centro del tumor se halla en el punto $(0.06, 0.06, 0.02)$ y el radio es $0.006 [m]$. Nuevamente se nota el incremento de temperatura en el área sobre la región tumoral.

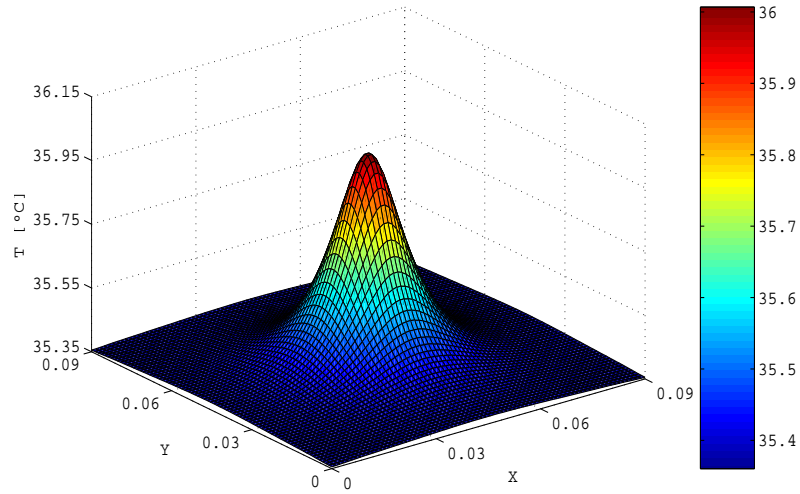


Figura 2.4: Perfil de temperatura sobre la superficie de la piel.

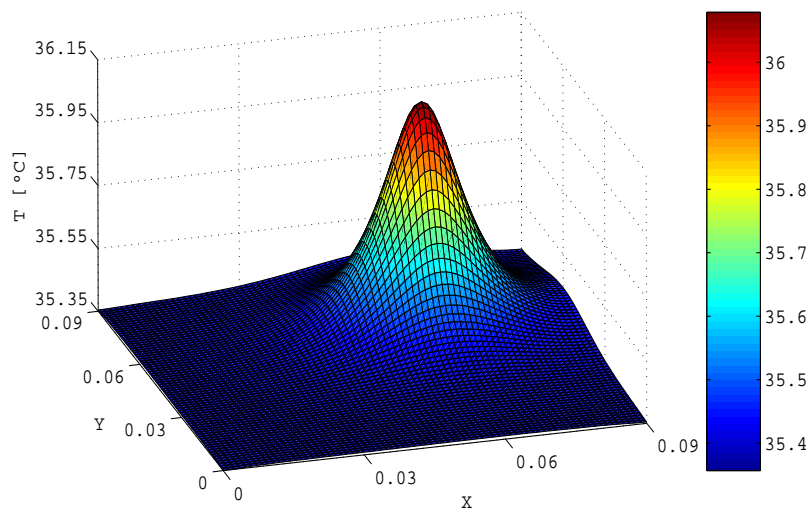


Figura 2.5: Perfil de temperatura sobre la superficie de la piel.

Capítulo 3

Localización de tumores y estimación de parámetros: metodología 1

3.1. Problemas Inversos

De acuerdo a Keller [47] podemos decir que dos problemas son inversos el uno del otro si la formulación de cada uno de estos requiere un conocimiento total o parcial del otro. Generalmente se acostumbra a llamar a uno de los dos problemas, usualmente el que ha sido estudiado con anterioridad o con mayor detalle, el problema directo mientras que al otro se lo denomina el problema inverso. A pesar de parecer esta una distinción un tanto arbitraria, en el caso en que el problema matemático resulta ser la descripción de algún fenómeno del mundo real, la distinción entre directo e inverso resulta en la mayoría de los casos más natural. Por ejemplo, si uno quiere describir el comportamiento futuro de un sistema físico, a partir del conocimiento del estado actual de este y de las leyes físicas que lo gobiernan (incluyendo el conocimiento concreto de los valores de parámetros que resulten de relevancia), es claro llamar a este el problema directo, pues conociendo las causas es posible averiguar el efecto. Ahora si el problema por ejemplo consiste en identificar el valor de ciertos parámetros físicos a partir de observaciones de la evolución del sistema, es más razonable llamar a éste el problema inverso, aquí nos interesa conocer cuál fue la causa que provocó las consecuencias observadas.

Desde el punto de vista de las aplicaciones se podrían destacar dos motivaciones para estudiar problemas inversos: la primera tiene que ver con el interés en conocer estados pasados o parámetros del sistema físico a partir de observaciones actuales. La segunda motivación está relacionada con el hecho de saber como influenciar un sistema de manera tal de conducirlo a un estado deseado en el futuro. Se podría decir entonces que los problemas inversos están relacionados con el hecho de determinar las causas que generan un efecto deseado u observado. A continuación, a modo de ejemplificar, exponemos brevemente algunos de los problemas inversos más conocidos:

- *Tomografía*: es una técnica que permite el registro de imágenes correspondientes a un plano o a una sección determinada de un objeto dado. El registro de las imágenes está basado en los datos medidos al iluminar el objeto desde varias direcciones distintas. Esta técnica tuvo un impacto muy grande cuando se la aplicó a la Medicina, ya que brindó a los médicos la posibilidad de explorar el interior del cuerpo humano y distinguir diferentes órganos con una gran precisión y en cierta medida de una manera segura para el paciente, convirtiéndose así en una metodología de diagnóstico muy utilizada. Varias técnicas de imágenes han sido empleadas también en ciencias como Arqueología, Biología, Geofísica, Oceanografía entre otras [58, 68]. Desde el punto de vista matemático el objetivo en el problema de tomografía

consiste en reconstruir una función f a partir del conocimiento de la integral

$$\int_{\gamma} f d\gamma$$

sobre una familia de variedades γ . El caso en que γ representa rectas en \mathbb{R}^2 es de suma importancia pues modela la tomografía de rayos X. La unicidad en la reconstrucción de la función f y una fórmula explícita de reconstrucción fue dada por Radon en 1917 [64], por ende muy a menudo este problema lleva su nombre. No obstante este problema recién cobró importancia gracias a Cormak y Hounsfield quienes desarrollaron en los años 60', basados en resoluciones numéricas y algoritmos de reconstrucción, la metodología hoy conocida como tomografía de rayos X. En el año 1979 ellos recibieron el premio Nobel por este trabajo.

- *Problemas espectrales*: uno de los problemas espectrales más conocido es el de “escuchar la forma de un tambor”. Esto es, inferir información sobre la forma de un tambor escuchando el sonido que este emite. Varias personas han puesto su atención en este tipo de problemas, podría decirse que uno de los primeros fue Sir Franz Arthur Shuster por el año 1882, pero fue recién por el año 1950 que Bochner lo planteó de manera más rigurosa y ya en el año 1966 fue Marc Kac el que popularizó este problema con su bien conocido trabajo “Can one hear the shape of a drum?” [43]. Desde el punto de vista matemático el tambor es concebido como una membrana elástica cuyo borde está sujeto o atornillado. La pregunta se traduce entonces en la siguiente: es posible determinar el dominio D a partir de los autovalores $\lambda = \lambda_k$ del problema de Dirichlet

$$\begin{cases} -\Delta u + \lambda u = 0, & \text{en } D, \\ u = 0, & \text{sobre } \partial D. \end{cases} \quad (3.1)$$

Los autovalores corresponden a las frecuencias de resonancia, por lo tanto $\{\lambda_k\}$ pueden ser interpretados como información ‘exterior’ natural. Es claro que los autovalores no cambian bajo la acción de isometrías (rotaciones, traslaciones o reflexiones), por ende sólo la forma de D puede ser determinada.

Aún antes de que el trabajo de Kac fuera publicado, John Milnor halló un contraejemplo. El probó la existencia de un par de toros en \mathbb{R}^{16} que tenían los mismos autovalores pero diferentes formas. Luego, Vigneras [80] también halló ejemplos de variedades compactas n -dimensionales, $n = 2, 3$ no isométricas pero con el mismo espectro. Más recientemente, ya en 1992 Grodon et al. [37] hallaron dos polígonos no isométricos pero sí isoespectrales, es decir con iguales autovalores, lo cual da una respuesta negativa a la pregunta de Kac en el caso de dominios con bordes suaves a trozos. En la actualidad aún no es sabido que dominios con bordes suaves pueden ser determinados a partir de su resonancia. Para resultados más recientes ver el trabajo de Osgood et al. [59].

- *Tomografía de impedancia eléctrica (TIE)*: es una técnica de diagnóstico por imágenes que consiste en inferir o recuperar la conductividad de una parte del cuerpo humano a través de mediciones de voltajes que se realizan en la superficie corporal. Esta técnica es usualmente utilizada para el monitoreo del funcionamiento de los pulmones, detección de cáncer de mamas y localización de focos epilépticos entre otras aplicaciones [42]. Las primeras experiencias médicas fueron detalladas en 1984 por David C. Barber y Brian H. Brown [9].

Desde el punto de vista matemático la TIE consiste en recuperar el mapa o distribución de conductividad de un dominio $\Omega \subset \mathbb{R}^n$, a partir de mediciones de voltajes que resultan de aplicar una o varias corrientes en la superficie del dominio. En la práctica se coloca un conjunto de electrodos sobre el cuerpo del paciente. Luego una o varias corrientes φ_i , $1 \leq i \leq I$, son aplicadas en algunos de los electrodos y los correspondientes voltajes f_i , $1 \leq i \leq I$, son medidos en los restantes puntos. Denotando por $\sigma(x)$ la conductividad que se desea hallar, el potencial u_i es la solución de la ecuación

$$\begin{cases} \operatorname{div}(\sigma \nabla u_i) = 0 & \text{en } \Omega, \\ \sigma \frac{\partial u}{\partial \eta} = \varphi_i & \text{sobre } \partial\Omega. \end{cases} \quad (3.2)$$

La formulación de este problema se atribuye al matemático Alberto Calderón [16] debido a su trabajo “On a inverse boundary problem” publicado en el año 1981 y hoy conocido en la literatura como el “Problema inverso de Calderón”. En la actualidad existe una extensa investigación desarrollada relacionada con la unicidad de la solución del problema y el desarrollo de algoritmos numéricos para su resolución [6–8, 77]. Más adelante daremos mayores detalles de este problema.

En todos los ejemplos presentados existe una diferencia fundamental entre el problema inverso y el problema directo correspondiente que tiene que ver con el concepto de problema bien o mal planteado. Hadamard [40] introdujo este concepto teniendo en cuenta que el modelo matemático considerado para describir un fenómeno físico debe poseer las propiedades de existencia, unicidad y estabilidad de la solución, o sea que el problema debe admitir solución, la solución de este debe ser única y además debe depender de los datos de manera continua. De no cumplirse alguna de estas propiedades estaríamos ante la presencia de un problema mal planteado. Mientras que los problemas directos resultan ser bien planteados, en la mayoría de los casos y aquí en todos los ejemplos presentados previamente, los problemas inversos son mal planteados en el sentido de Hadamard. Desde el punto de vista matemático podría decirse que las tres propiedades tienen diferentes grados de importancia. La primera propiedad, la cual trata sobre la existencia de la solución puede conseguirse extendiendo el espacio de soluciones, el concepto de distribución o función generalizada es un claro ejemplo de esto. Ahora, si el problema tiene más de una solución, esto puede deberse a falta de información en el modelo propuesto y tal vez suposiciones adicionales sobre la solución lleven a la unicidad del problema o simplemente podrá bastar con elegir la solución que resulte más adecuada de acuerdo al fenómeno en estudio; por ejemplo, se podría escoger la solución de menor norma. Por último, puede considerarse que el requisito de estabilidad es el más relevante desde el punto de vista de las aplicaciones. Si el problema carece de estabilidad entonces el cálculo de la solución verdadera es prácticamente imposible debido a que en las observaciones de un fenómeno real siempre estarán contaminados con ruido o al realizar los cálculos numéricos inevitablemente aparecerán errores. Luego, si la solución de un problema no depende de los datos de manera continua, la solución hallada puede distar bastante de la verdadera. Para conseguir la estabilidad de un problema mal planteado hay que acudir a métodos de regularización, ya sea a través de operadores, métodos iterativos, métodos continuos, mediante discretización o también en este caso considerando información adicional. En este trabajo no nos hemos centrado en explorar en detalle cada una de estas propiedades para los problemas inversos que aquí trataremos ya que el interés principal es desarrollar algoritmos numéricos para su resolución, sin embargo más adelante haremos comentarios sobre cada uno de estos conceptos.

El estudio de los problemas inversos ha sido una de las áreas dentro de la matemática aplicada que más ha crecido en los últimos años. Las aplicaciones en el campo de los problemas inversos

son muy variadas. Por ejemplo, en el area de ingeniería existen trabajos relacionados con el diseño de equipos térmicos y problemas de transferencia de calor [69, 71], dentro del área de diagnóstico por imágenes podemos citar trabajos relacionados con la tomografía computarizada y la reconstrucción de imágenes [82]. Por último en la biología, los problemas inversos han sido de utilidad para el estudio de dinámica de poblaciones [63].

3.2. Definición de los problemas inversos.

De acuerdo a [51, 52, 57, 67], la presencia de un tumor induce el incremento de la perfusión sanguínea local, la actividad metabólica y esto trae como consecuencia un incremento de la temperatura sobre la zona corporal que se encuentra sobre la región tumoral. Con este hecho como base, la idea es usar perfiles de temperaturas anormales sobre la superficie corporal de manera tal de predecir la localización y tamaño del tumor como así también estimar el valor de parámetros fisiológicos. Esto ha sido realizado considerando dos problemas inversos diferentes que serán explicados a continuación.

3.2.1. Problema Inverso 1: Estimación de la posición y tamaño del tumor.

Este primer problema inverso está asociado con la estimación de la localización y tamaño del tumor suponiendo que todos los otros parámetros son conocidos. Debido a que la motivación de este trabajo es la aplicación de esta metodología para tumores como melanomas nodulares o cáncer de mamas, suponemos entonces que el tumor tiene una forma esférica. Por lo tanto, para conocer su localización y tamaño es necesario conocer el centro del tumor y su radio, éstos serán entonces los parámetros geométricos a estimar. El problema que deseamos resolver es el siguiente:

- Conocido el valor de todos los parámetros fisiológicos, las temperaturas T_b y T_a y dado un perfil de temperatura u_{exp} medido sobre Γ_s , el problema consiste en hallar $\omega^* \subset \Omega$ tal que la solución u_{ω^*} al problema de borde (2.3) verifique $u_{\omega^*} = u_{exp}$ en Γ_s .

El ω^* óptimo se alcanza cuando la función objetivo:

$$\psi(\omega) = \Psi(\omega, u_\omega) = \frac{1}{2} \int_{\Gamma_s} (u_\omega - u_{exp})^2 d\Gamma \quad (3.3)$$

se anula, esto significa que ψ tiene un mínimo en ω^* .

La función costo $\psi(\omega)$ representa el error o diferencia entre un perfil de temperatura u_{exp} medido experimentalmente en la superficie corporal y la solución numérica u_ω , obtenida mediante diferencias finitas, del problema de transferencia de calor considerando el subdominio ω . Cabe observar que en el caso de dos dimensiones, hallar ω significa hallar una terna (x_c, y_c, R) que determina el centro y radio del tumor, mientras que en 3 dimensiones deben hallarse los parámetros (x_c, y_c, z_c, R) .

Para este problema inverso vamos a suponer la existencia de la solución. Es decir, que dado un perfil de temperatura medido sobre la superficie corporal asumimos que existirá al menos un vector de parámetros $p = (x_c, y_c, R)$ tal que la solución u del problema de borde (2.3), considerando el tumor centrado en el punto (x_c, y_c) y de radio R , es tal que $u = u_{exp}$ sobre Γ_s . Esta suposición es válida ya que aquí estamos utilizando datos artificiales o sintéticos.

Con respecto a la unicidad, sean ω_1 y ω_2 dos discos (bolas) en $\Omega \subset \mathbb{R}^2$ ($\Omega \subset \mathbb{R}^3$) y denotemos por u_j la solución al problema (2.3) considerando $\omega = \omega_j$. El estudio de la unicidad se traduce en responder la siguiente pregunta: Si $u_1 = u_2$ sobre Γ_s ,

¿podemos garantizar entonces que $\omega_1 = \omega_2$?

La respuesta a este interrogante, hasta donde nosotros sabemos, permanece abierta.

Notemos que el problema inverso que queremos resolver aquí tiene un enunciado muy similar al problema de la TIE que hemos presentado en la sección anterior. Este último problema consiste en determinar las propiedades eléctricas de un medio a partir de mediciones de corriente y voltaje sobre el borde del medio. Para llevarlo a un caso más cercano al tipo del problema inverso presentado en este trabajo, la TIE se puede formular de la siguiente manera: Dado un objeto cuya conductividad eléctrica es $\sigma \equiv 1$ y siendo $\omega \subset\subset \Omega$ una inclusión desconocida de conductividad $\sigma \equiv \mu$, tal que $0 < \mu \neq 1$, el problema inverso de conductividad, con sólo *una medición*, consiste en determinar ω a partir de una única medición del voltaje realizada sobre $\partial\Omega$ y el cual es consecuencia de haber aplicado un cierto flujo de corriente conocido también sobre $\partial\Omega$. En términos matemáticos, si u denota el potencial electrostático en Ω , uno desea recuperar ω en la ecuación

$$\left\{ \begin{array}{l} -\operatorname{div}((1 + (\mu - 1)\chi_\omega)\nabla u) = 0 \quad \text{en } \Omega, \\ \frac{\partial u}{\partial \eta} = g \quad \text{sobre } \partial\Omega, \\ \int_{\partial\Omega} u = 0 \end{array} \right. \quad (3.4)$$

a partir del conocimiento del dato $u = h$ sobre $\partial\Omega$. Con χ_ω denotamos la función característica. El problema de TIE ha sido estudiado desde hace varios años atrás, más precisamente a partir del trabajo de Alberto P. Calderón en el año 1980, [16]. A partir de ese año varios matemáticos reconocidos se han interesado en este problema y existe una extensa investigación desarrollada relacionada con la unicidad de la solución del problema; no obstante en la mayoría de los trabajos existentes se asume una cantidad infinita de mediciones sobre la frontera como así también se considera que las mediciones son realizadas en toda la frontera y no sólo en parte de ésta [49, 72]. A continuación listamos algunos de los resultados sobre unicidad del problema de conductividad en el caso de una única medición en la frontera.

En el año 1989 Friedman e Isakov, suponiendo que Ω es un dominio acotado en \mathbb{R}^n ($n = 2, 3$) y con frontera $C^{1,1}$ a trozos. Suponiendo además que la función $g \in L^\infty(\Omega)$ es tal que $g \neq 0$, de soporte compacto y contenido en un abierto S y tal que $\int_S u = 0$, probaron en [34] que:

Teorema(FI): Si ω_1, ω_2 son dos poliedros convexos tales que $\overline{\omega_j} \subset \Omega$ y además

$$\operatorname{diam} \omega_j < \operatorname{dist}(\omega_j, \partial\Omega) \quad (3.5)$$

Sea u_j la solución de (3.4) para $\omega = \omega_j$ y Γ_0 un abierto no vacío en $\partial\Omega$. Entonces, si $u_1 = u_2$ sobre Γ_0 debe pasar que $\omega_1 = \omega_2$.

Observación: Si Ω representa el semiplano el Teorema sigue siendo válido si la condición (3.5) es descartada. Más aún en el caso $n = 2$ el Teorema es válido si Ω es un disco.

Luego en el año 1996 Kang y Seo [44] probaron, considerando que Ω es un dominio acotado, simplemente conexo en \mathbb{R}^2 y con frontera Lipschitz, que la función $g \in L^2(\partial\Omega)$ es tal que $\int_{\partial\Omega} g = 0$ y $g \neq 0$ sobre $\partial\Omega$, el siguiente resultado.

Teorema(KS1): Sean ω_1, ω_2 discos en \mathbb{R}^2 tales que $\omega_j \subset\subset \Omega$ $j = 1, 2$. Denotemos por u_j las soluciones del problema (3.4) para $\omega = \omega_j$. Entonces, si $u_1 = u_2$ sobre $\partial\Omega$ debe ser $\omega_1 = \omega_2$.

Finalmente en 1999 Kang y Seo [45], asumiendo $\Omega \subset \mathbb{R}^3$ acotado y con frontera Lipschitz y en este caso los subdominios ω_j bolas en \mathbb{R}^3 , demostraron que

Teorema(KS2): Si u_j son soluciones de (3.4) para $\omega = \omega_j$ $j = 1, 2$, con g que satisface las mismas condiciones del teorema anterior y tales que $u_1 = u_2$ sobre $\partial\Omega$, entonces $\omega_1 = \omega_2$.

Cabe señalar que por más que el problema de TIE ha sido estudiado durante varios años el problema de unicidad con una única medición permanece aún abierto para el caso de subdominios suaves en dimensiones dos y tres. Más aún, la unicidad para el caso de ω elipse o elipsoide también es desconocida.

En cuanto a la estabilidad del problema podemos mencionar uno de los resultados para el caso de TIE y que se asemeja al del problema inverso de la localización del tumor. En [46] Kang et al. presentan resultados sobre la estabilidad y también sobre estimación del tamaño de la inclusión ω bajo la suposición de una única medición. En este trabajo presentaron cotas inferiores y superiores para el tamaño de la inclusión ω en términos de la conductividad μ y del dato en la frontera de Ω , esto para cualquier dimensión y sin conocimiento a priori respecto a la forma de ω , no obstante sí impusieron ciertas hipótesis sobre la función g . Con respecto a la estabilidad, probaron que si ω_1 y ω_2 son dos discos en \mathbb{R}^2 , si además el dato de Neumann g en (3.4) satisface las siguientes dos condiciones:

- (N1) Existe un número positivo M tal que $|g'(P)| > M$, si $|g(P)| < M$ para todo $P \in \partial\Omega$ (g' representa la derivada tangencial),
- (N2) $\{P \in \partial\Omega : g(P) \geq 0\}$ y $\{P \in \partial\Omega : g(P) \leq 0\}$ son subconjuntos conexos y no vacíos de $\partial\Omega$.

Por último denotando $h_1 = u_1$ restringida a $\partial\Omega$ y de manera similar $h_2 = u_2$ restringida a $\partial\Omega$, entonces vale el siguiente resultado de estabilidad logarítmica:

$$|\omega_1 \triangle \omega_2| \leq |\log \|h_1 - h_2\|_{L^2(\partial\Omega)}|^{-\frac{1}{\beta}} \quad (3.6)$$

donde $|\omega_1 \triangle \omega_2|$ es la medida de la diferencia simétrica de ω_1 y ω_2 definida por

$$|A \triangle B| = (A \cup B) \setminus (A \cap B),$$

es decir los elementos que pertenecen a la unión de A y B y que no están en su intersección. Por su parte β es un número determinado por el ángulo entre $\partial\omega_1$ y $\partial\omega_2$ en los puntos de intersección. Posteriormente los mismos autores en [30], asumiendo las mismas hipótesis para la función g , mejoraron el resultado de estabilidad logarítmica por el siguiente que asegura estabilidad de tipo Hölder:

$$|\omega_1 \triangle \omega_2| \leq C \|h_1 - h_2\|_{L^\infty(\partial\Omega)}^\kappa \quad (3.7)$$

donde C es una constante que depende de la distancia de los discos ω_j a la frontera $\partial\Omega$ y $0 < \kappa < 1$. Más resultados sobre estabilidad en el caso de una medición pueden hallarse en [4, 5]

3.2.2. Problema Inverso 2: Estimación de la fuente de calor metabólico.

En este problema suponemos que la posición y el tamaño del subdominio ω son conocidas, como así también el valor de todos los parámetros fisiológicos, salvo el valor correspondiente a la fuente de calor metabólico dentro de la región tumoral, o sea el valor de q_0 .

- Conocido el valor de todos los parámetros fisiológicos, las temperaturas T_b y T_a y dado un perfil de temperatura u_{exp} sobre Γ_s , el problema consiste en hallar q_0^* tal que la solución $u_{q_0^*}$ al problema de borde (2.3) verifique $u_{q_0^*} = u_{exp}$ en Γ_s .

El valor de q_0 óptimo se alcanza cuando la función objetivo:

$$\Psi(q_0) = \Psi(u_{q_0}) = \frac{1}{2} \int_{\Gamma_s} (u_{q_0} - u_{exp})^2 d\Gamma \quad (3.8)$$

se anula.

Similar al caso anterior, la función costo $\Psi(q_0)$ representa el error o diferencia entre un perfil de temperatura u_{exp} medido experimentalmente en la superficie corporal y el perfil correspondiente a la solución numérica u_{q_0} , del problema de transferencia de calor considerando el valor q_0 como la intensidad de la fuente de calor metabólico en la región tumoral.

Este problema es de particular interés ya que la estimación de los valores de los parámetros térmicos relacionados con el tumor pueden resultar de utilidad e importancia a la hora de estudiar la evolución del tumor luego de que éste haya recibido algún tipo de tratamiento, como lo es en el caso de BNCT [67]. La comparación de la intensidad de la fuente de calor metabólica dentro del tumor, antes y luego de haber recibido el tratamiento, puede ser utilizada como una medida para evaluar la efectividad del tratamiento recibido.

Para este problema inverso también vamos a suponer la existencia de una solución. Es decir que dado un perfil de temperatura u_{exp} medido sobre la superficie corporal asumimos que existirá al menos un valor q_0^* tal que la solución u del problema de borde (2.3) considerando la fuente de calor metabólico con intensidad igual a q_0^* dentro la región tumoral, es tal que $u = u_{exp}$ sobre Γ_s . Nuevamente esta suposición es válida ya que aquí estamos utilizando datos sintéticos.

Sobre el interrogante de la unicidad en la solución de este problema inverso, podemos probar que la respuesta es afirmativa. En efecto, supongamos que u^i con $i = 1, 2$ son las soluciones del problema (2.3) o equivalentemente del problema de transmisión

$$\left\{ \begin{array}{ll} -\sigma_1 \Delta u_1^i + k_1 u_1^i = q_1 + k_1 T_b, & \text{en } \Omega - \bar{\omega}, \\ -\sigma_0 \Delta u_0^i + k_0 u_0^i = q_0^i + k_0 T_b, & \text{en } \omega, \\ u_1^i = u_0^i, & \text{sobre } \partial\omega, \\ -\sigma_1 \frac{\partial u_1^i}{\partial \eta} = -\sigma_0 \frac{\partial u_0^i}{\partial \eta}, & \text{sobre } \partial\omega, \\ -\sigma_1 \frac{\partial u_1^i}{\partial \eta} = \alpha(u_1^i - T_a), & \text{sobre } \Gamma_s, \\ -\sigma_1 \frac{\partial u_1^i}{\partial \eta} = 0, & \text{sobre } \Gamma_l, \\ u_1^i = T_b, & \text{sobre } \Gamma_b. \end{array} \right. \quad (3.9)$$

O sea u^1 y u^2 son soluciones de problemas similares cuya única diferencia está en el valor de la fuente de calor metabólico dentro de la región tumoral que hemos denotado por q_0^i . Asumiendo que $u^1 = u^2$ sobre Γ_s queremos concluir que $q_0^1 = q_0^2$.

Consideremos la función $v = u_1^1 - u_1^2$ definida en el dominio $\Omega - \bar{\omega}$. Tenemos que esta función satisface la siguiente ecuación diferencial

$$-\sigma_1 \Delta v + k_1 v = 0, \quad \text{en } \Omega - \bar{\omega},$$

y además por hipótesis sabemos que

$$v = 0, \quad \text{sobre } \Gamma_s. \quad (3.10)$$

Usando esta última ecuación y apelando a las condiciones de borde que satisfacen u_1^1 y u_1^2 en la frontera Γ_s , tenemos que

$$-\sigma_1 \frac{\partial v}{\partial \eta} = -\sigma_1 \left(\frac{\partial u_1^1}{\partial \eta} - \frac{\partial u_1^2}{\partial \eta} \right) = \alpha(u_1^1 - u_1^2) = \alpha v = 0, \quad \text{sobre } \Gamma_s. \quad (3.11)$$

Luego, tenemos que v satisface una ecuación diferencial elíptica con coeficientes suaves (constantes) en el dominio $\Omega - \bar{\omega}$ y que además $v \equiv 0 \equiv \frac{\partial v}{\partial \eta}$ sobre Γ_s , luego por la propiedad de continuación de borde ¹ [74], podemos concluir que

$$v \equiv 0, \quad \text{en } \Omega - \bar{\omega}. \quad (3.12)$$

Por otra parte, consideremos ahora la función $\tilde{v} = u_0^1 - u_0^2$ definida en ω . Esta función satisface la ecuación diferencial

$$-\sigma_0 \Delta \tilde{v} + k_0 \tilde{v} = q_0^1 - q_0^2, \quad \text{en } \omega. \quad (3.13)$$

Además, usando el hecho de que $0 = v = u_1^1 - u_1^2$ en $\Omega - \bar{\omega}$ y apelando a las condiciones de borde que satisfacen u_0^1 y u_0^2 en la frontera $\partial\omega$ tenemos que

$$\begin{aligned} u_0^1 &= u_1^1 = u_1^2 = u_0^2, & \text{sobre } \partial\omega, \\ \frac{\partial u_0^1}{\partial \eta} &= \frac{\partial u_1^1}{\partial \eta} = \frac{\partial u_1^2}{\partial \eta} = \frac{\partial u_0^2}{\partial \eta}, & \text{sobre } \partial\omega, \end{aligned} \quad (3.14)$$

lo cual equivale a decir que $\tilde{v} = 0$ sobre la frontera $\partial\omega$. Por lo tanto \tilde{v} satisface el problema de borde

$$\begin{cases} -\sigma_0 \Delta \tilde{v} + k_0 \tilde{v} = q_0^1 - q_0^2, & \text{en } \omega, \\ \tilde{v} = 0, & \text{sobre } \partial\omega, \end{cases} \quad (3.15)$$

Ahora como q_0^1 y q_0^2 son constantes, podemos asumir que $q_0^1 > q_0^2$ y por el Lema de Hopf [29] tenemos que

$$\frac{\partial \tilde{v}}{\partial \eta} < 0, \quad \text{sobre } \partial\omega. \quad (3.16)$$

lo cual contradice (3.14). También llegaríamos a la misma contradicción si suponemos $q_0^1 < q_0^2$. Por lo tanto, es necesario que $q_0^1 = q_0^2$. Es decir que la intensidad interna de la fuente de calor metabólico tiene que ser la misma, resultado que implica la unicidad del problema inverso a tratar.

3.3. Resultados Numéricos.

En la sección previa presentamos y planteamos los dos problemas inversos que resultan de interés en este trabajo. Ambos problemas fueron planteados de manera tal que cada uno de ellos consiste en hallar el mínimo de una función costo Ψ . De este modo estos problemas inversos pueden ser vistos como problemas de optimización. El objetivo es hallar un conjunto de parámetros que sea óptimo en el sentido que la solución del problema de borde asociado a esa elección de parámetros sea la que minimice la función costo. En resumen, ambos problemas a resolver pueden escribirse como:

- Hallar $p \in \mathcal{A}$ tal que

$$p = \arg(\min_{p \in \mathcal{A}} \Psi(p, u_p)) \quad (3.17)$$

donde $\mathcal{A} = \{p \in \mathbb{R}^m : \underline{p} \leq p \leq \bar{p}, \quad L_p u = f_p\}$.

¹Para aplicar esta propiedad es necesario que el dominio $\Omega - \bar{\omega}$ sea un dominio Dini [74]

La primera restricción tiene que ver con las cotas inferiores y superiores que deben cumplir cada uno de los parámetros o variables, por ejemplo en el caso del problema inverso correspondiente a estimar la intensidad de la fuente de calor metabólico, debemos tener en cuenta que su valor está restringido a un cierto intervalo. La segunda restricción es simplemente que se debe cumplir cierta ecuación de estado, en nuestro caso el problema (2.3).

En este capítulo para hallar el mínimo de la función costo utilizaremos un algoritmo denominado Pattern Search [25]. Este algoritmo pertenece a la clase de algoritmos de optimización conocidos como algoritmos de búsqueda directa o también llamados de orden cero, ya que no hace uso de las derivadas para encontrar el mínimo deseado, lo cual lo convierte en un algoritmo de muy fácil implementación. Básicamente, el algoritmo consiste en lo siguiente: dado un punto inicial, y una vez definidas bajo cierto criterio las direcciones de exploración o patrón de búsqueda, estas se utilizan para definir un conjunto finito de puntos factibles en los cuales se realiza la evaluación de la función F a minimizar. A partir de estos valores se elige el punto en el que se alcanza el mínimo valor observado y así se define un nuevo punto “inicial” o iterando. Así por ejemplo en la k -ésima iteración tenemos el iterando $p_k \in \mathcal{A}$ y un paso $\delta_k > 0$. Con estos datos se definen de manera sucesiva los puntos $p_+ = p_k + \delta_k d_i$, $i \in \{1, 2, \dots, n\}$, donde d_i representan la direcciones de búsquedas, hasta encontrar un p_+ tal que $F(p_+) < F(p_k)$. Si no se llega a encontrar p_+ tal que $F(p_+) < F(p_k)$, se reduce el paso δ_k a su mitad y se continua con el proceso, si por el contrario se halló tal punto p_+ se definen $\delta_{k+1} = \delta_k$ y $p_{k+1} = p_+$. Esta iteración, descrita aquí en forma breve y simplificada, se repite de manera sucesiva hasta que δ_k sea suficientemente pequeño y por lo tanto no se produzcan cambios considerables en el valor que toma la función en el nuevo iterando. Otra opción es parar el proceso una vez que el valor de la función F sea menor al de cierto valor deseado de antemano. Más detalles sobre este algoritmo así como también resultados sobre su convergencia pueden encontrarse en [53, 76].

En resumen, los pasos requeridos para resolver los distintos problemas de optimización anteriormente presentados son los siguientes:

- Dado un vector inicial de parámetros p^i (para el caso de localización $p^i = (x^i, y^i, z^i, R^i)$, mientras que para la estimación de intensidad del calor metabólico $p^i = q_0^i$)

1. Definir los parámetros de diseño actuales como $p = p^i$.
2. Generar la grilla y las funciones σ , k y q correspondientes a la configuración determinada por p .
3. Resolver el problema de borde (2.3) para obtener el estado u .
4. Calcular el valor de la función costo ψ .
5. Si el criterio de parada es alcanzado, dar como salida los parámetros actuales como los de óptimo diseño y parar, sino continuar.
6. Actualizar p usando el algoritmo Pattern Search considerando el conjunto de restricciones.
7. Ir a 2.

3.3.1. Resultados: PI 1 Estimación de la posición y tamaño del tumor.

A continuación presentamos algunos de los resultados obtenidos. Estos resultados muestran que tanto para el caso en 2D como para el 3D es posible determinar los parámetros buscados

a partir de los datos de la temperatura sobre el borde superior del dominio. Más aún cuando los datos de entrada fueron contaminados con 10% de ruido aleatorio uniformemente distribuido, los resultados obtenidos fueron muy buenos. En el caso en el que a los datos de entrada se les sumó un 15% de ruido con distribución uniforme también resultaron ser buenos. En todos los ejemplos los valores correspondientes a los parámetros y las dimensiones del dominio fueron iguales a los valores expresados en la sección 2.5. Cabe señalar que varias corridas del algoritmo fueron realizadas utilizando diferentes condiciones iniciales aleatorias y en todos los casos los resultados obtenidos fueron similares.

En la Figura 3.1 se muestra un ejemplo del funcionamiento de la metodología propuesta para el caso de 2 dimensiones. El objetivo es localizar un tumor centrado en el punto $(0.029, 0.02)$ y de radio igual a 0.005 . En la Figura 3.1a, se puede ver en color verde el contorno del tumor que deseamos hallar, en color azul se encuentra graficado el contorno del dato inicial considerado. A su derecha, en la Figura 3.1, podemos ver la temperatura medida sobre el borde superior Γ_s correspondiente a este tumor. Este perfil de temperatura será utilizado como dato de entrada en el problema inverso de hallar el tumor. Abajo a la izquierda, en la Figura 3.1c mostramos el mismo perfil de temperatura pero al que se lo ha contaminado con 10% de ruido con distribución uniforme. Por último en la Figura 3.1d vemos nuevamente en color verde el contorno del tumor objetivo, en rojo el contorno del tumor hallado considerando los datos de entrada sin ruido y por último en negro está graficado el contorno del tumor hallado considerando los datos contaminados con 10% de ruido. Los valores correspondientes se encuentran en última fila del Cuadro 3.1. Vemos que utilizando la metodología propuesta es posible determinar la localización del centro y el valor del radio de la región tumoral con una muy buena precisión.

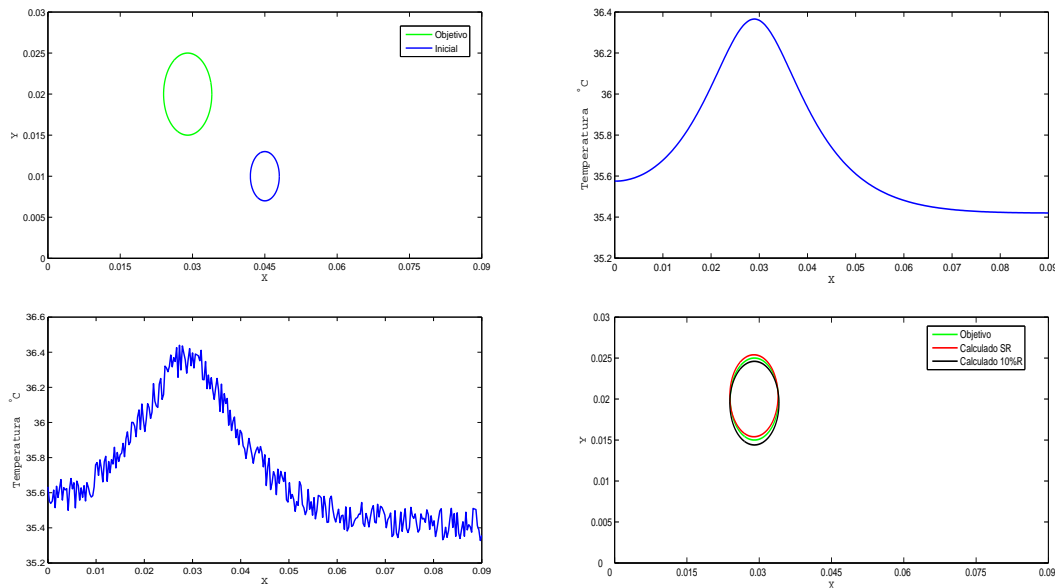


Figura 3.1: (a) Contorno dato inicial (azul) y contorno del subdominio objetivo (verde). (b) Temperatura sobre la frontera superficial Γ_s . Este perfil de temperatura es utilizado como datos de entrada. (c) Datos de entrada contaminados con 10% de ruido uniformemente distribuido. (d) Diferencia entre el subdominio objetivo (verde), el subdominio hallado utilizando datos de entrada sin ruido (rojo) y el subdominio hallado considerando datos de entrada contaminados con 10% de ruido (negro).

En el Cuadro 3.1 se listan los resultados del problema inverso relacionado con la localización del tumor en 2 dimensiones. Para el caso en que los datos de entrada no han sido contaminados con ruido todos los resultados tienen un error menor al 5%. Cuando al perfil de temperatura se le sumó 10% de ruido uniformemente distribuido, los valores estimados tienen un error menor al del 8%. Por último, cuando a los mismo datos de entradas se los contaminó con 15% los valores estimados presentaron un error menor al 12%.

Datos originales			Sin ruido			10% de ruido			15% de ruido		
x	y	R	x	y	R	x	y	R	x	y	R
0.040	0.020	0.003	0.0400	0.0209	0.0003	0.0403	0.0203	0.0029	0.0399	0.0215	0.0028
0.020	0.015	0.004	0.0206	0.0158	0.0039	0.0199	0.0162	0.0038	0.0195	0.0168	0.0045
0.029	0.020	0.005	0.0289	0.0204	0.0050	0.0290	0.0193	0.0051	0.0286	0.0204	0.0048

Cuadro 3.1: Estimación del centro y radio del tumor en 2D. Las diferentes columnas muestran los resultados obtenidos considerando los datos de entrada sin ruido y con 10% y 15% de ruido uniformemente distribuido respectivamente.

En el Cuadro 3.2 se pueden observar los resultados relacionados con el mismo problema inverso, pero en este caso en 3 dimensiones. Cabe aclarar que en este caso el algoritmo conlleva un mayor costo computacional, no obstante los resultados siguen siendo muy buenos. Para el primer ejemplo los errores en los valores estimados son menores al 5% incluso cuando al perfil de temperatura sobre la superficie se lo contaminó con 10% y 15% de ruido uniformemente distribuido. En el segundo ejemplo vemos que los errores en los valores estimados siguen siendo de este orden, pero cuando se contaminaron los datos de entradas con 15% los resultados también presentaron un error del 15%.

Datos originales		Sin ruido	10% ruido	15% ruido	Datos originales		Sin ruido	10% ruido	15% ruido
x	0.030	0.0301	0.0304	0.0298	x	0.020	0.0201	0.0198	0.0191
y	0.030	0.0300	0.0298	0.0303	y	0.015	0.0155	0.0156	0.0159
z	0.022	0.0221	0.0223	0.0231	z	0.019	0.0189	0.0179	0.0160
R	0.005	0.0050	0.0049	0.0048	R	0.006	0.0059	0.0063	0.0069

Cuadro 3.2: Estimación del centro y radio del tumor en 3D. Las diferentes columnas muestran los resultados obtenidos considerando los datos de entrada sin y con un ruido aleatorio del 10% y 15% respectivamente.

Observación: Cabe señalar que los muy buenos resultados obtenidos se deben en gran medida a la fuerte restricción impuesta sobre la clase de funciones para los coeficientes térmicos y para la fuente de calor metabólico. Desde el punto de vista matemático, suponer estas funciones como continuas a trozos implica una reducción notoria en la dimensión del conjunto de parámetros admisibles, que hemos denotado por \mathcal{A} y donde consideramos la minimización del funcional costo. De esta manera la minimización se realiza sobre un conjunto de dimensión finita. Esto puede ser interpretado como una regularización de problema mediante discretización [28].

3.3.2. Resultados: PI 2 Estimación de la intensidad de la fuente de calor metabólico.

Los resultados obtenidos para el problema de estimación de la intensidad de la fuente de calor metabólico en el caso de 2 dimensiones se muestran en el Cuadro 3.3 Nuevamente, se puede apreciar una gran similitud entre los datos originales y los estimados por el algoritmo. En todos los casos los valores estimados tienen un error menor al 3% comparados al del valor exacto.

Datos originales		Sin ruido	10 % de ruido	15 % de ruido
q_0	42000	42000	41881.17	42537.18
q_0	37000	37000	37177.50	37632.90
q_0	25000	25000	24891.71	25716.82

Cuadro 3.3: Estimación de la intensidad del calor metabólico dentro del tumor, caso 2D. En todos los ejemplos el tumor fue centrado en el punto (0.045, 0.02) y su radio igual a 0.005.

Finalmente, en el Cuadro 3.4 se presentan los resultados obtenidos para el problema de la estimación de la intensidad de la fuente metabólica dentro de la región tumoral, para el caso de 3 dimensiones. En los ejemplos que se muestran todos los resultados presentan un error menor al 5 % comparados al valor exacto.

Recordemos que este problema es de interés debido a que el conocimiento de los valores de los parámetros térmicos relacionados con el tumor pueden ser de mucha utilidad e importantes para poder estudiar la evolución del tumor luego de que éste haya recibido algún tipo de tratamiento, por ejemplo ver [67]. La comparación de la intensidad de la fuente de calor metabólica dentro de la región tumoral, antes y luego de haber recibido el tratamiento, podría ser utilizada también como una medida para evaluar la efectividad del tratamiento recibido.

Datos originales		Sin ruido	10 % de ruido	15 % de ruido
q_0	42000	42000	41690.09	40741.71
q_0	37000	37000	37957.31	36675.37
q_0	25000	25000	24609.03	26307.50

Cuadro 3.4: Estimación de la intensidad del calor metabólico dentro del tumor, caso 3D. En todos los ejemplos el tumor fue centrado en el punto (0.045, 0.045, 0.015) y su radio igual a 0.005.

3.4. Conclusión del capítulo

En este capítulo se presentó y desarrolló una metodología simple para la estimación de parámetros relacionados con la posición y tamaño del tumor, como así también parámetros fisiológicos asociados a la región tumoral. La idea fue utilizar mediciones de temperatura realizadas sobre la superficie corporal, las que en la práctica pueden ser obtenidas mediante cámaras termográficas y a partir de esta información detectar cualquier anomalía que pueda existir en el interior del cuerpo. Desde el punto de vista matemático los problemas a resolver pueden ser vistos como problemas inversos. Estos problemas fueron resueltos utilizando un método de diferencias finitas de segundo orden en conjunto con el algoritmo de minimización Pattern Search. Este es un algoritmo de búsqueda directa que no hace uso de las derivadas y por ende es de fácil implementación. Los resultados obtenidos demuestran la factibilidad de la metodología propuesta. Aún en el caso cuando 10% y 15% de ruido fue adicionado a los datos de entrada la metodología estima los diferentes parámetros con una muy buena precisión tanto en el caso 2D como en el caso 3D.

Los buenos resultados obtenidos se deben en mayor medida a la fuerte restricción impuesta sobre la clase de funciones para los coeficientes térmicos y para la fuente de calor metabólico.

De acuerdo a los resultados esta metodología puede ser considerada como una potencial ayuda para localizar regiones tumorales, en particular melanomas nodulares, como así también para estimar valores de parámetros fisiológicos relacionados con las regiones tumorales, valores que podrían ser de utilidad para estudiar la evolución de un tumor bajo tratamientos como por ejemplo BNCT [67].

Capítulo 4

Localización de tumores y estimación de parámetros: metodología 2

En este Capítulo presentaremos una segunda metodología para resolver los problemas inversos antes tratados. La metodología que proponemos hace uso de la información que provee la derivada del funcional a minimizar con respecto a las distintas variables a estimar. Para calcular tal derivada, acudiremos a las ideas del análisis de sensibilidad, ideas que introduciremos a continuación analizando un breve ejemplo.

4.1. Introducción al Análisis de Sensibilidad.

Consideremos una viga de espesor variable e , representada por el intervalo $[0, 1]$, la cual se encuentra con ambos extremos empotrados y a la que además se le aplica una fuerza uniforme f . La deformación transversal $u := u(e)$ de la viga es la solución del siguiente problema de borde:

$$\begin{cases} (\beta e^3 u''(x))'' = f, & \text{en } [0, 1], \\ u(0) = u(1) = 0, \\ (\beta e^3 u'')(0) = (\beta e^3 u'')(1) = 0, \end{cases} \quad (4.1)$$

donde β es una función positiva que depende de las propiedades del material y la cual asumimos conocida. Supongamos que disponemos de una cantidad fija de material y estamos interesados en el diseño de una viga tal que su rigidez sea máxima, teniendo en cuenta la fuerza f que se le aplicará y la cantidad de material que tenemos a nuestra disposición. Este problema puede traducirse en hallar una distribución para el espesor e , de manera tal que la viga tenga un diseño óptimo en el sentido que su rigidez es la máxima posible para esa cantidad de material. Para tal fin, es necesario analizar cómo varía la flexión o deformación de la viga de acuerdo a los cambios que realicemos en la distribución del espesor y cómo esto se ve reflejado en la rigidez de la viga. Para medir como los cambios en el espesor se reflejan en la rigidez de la viga, vamos a considerar el siguiente funcional:

$$J(u(e)) = \int_0^1 f u(e) dx, \quad (4.2)$$

donde $u(e)$ es la solución de (4.1). Este funcional, que representa energía externa de deformación, puede ser interpretado como una medida de la flexibilidad de la viga o también como una medida “inversa” de su rigidez. Mientras mayor es la rigidez de la viga menor es el valor que toma el funcional J . Por lo tanto, debemos hallar e^* tal que

$$J(u(e^*)) \leq J(u(e)), \quad \forall e \in \mathcal{A},$$

donde \mathcal{A} representa el conjunto de espesores admisibles. En este ejemplo vemos como un parámetro o más precisamente un coeficiente del modelo que describe el sistema que estamos analizando, resulta de fundamental importancia a la hora de obtener el diseño óptimo de este sistema. Tal como dijimos antes, un cambio en la distribución del espesor de la viga traerá como consecuencia un cambio en el comportamiento de la deformación de la viga y por ende en su rigidez.

El objetivo principal del análisis de sensibilidad es determinar de manera cuantitativa el cambio en el comportamiento de un sistema cuando, de alguna manera, un conjunto de sus variables de control son modificadas.

En el ejemplo anterior la variable de diseño o control era un coeficiente de la ecuación diferencial, otra de las posibilidades es considerar la geometría del dominio Ω , donde el sistema está definido, como la variable de diseño o control. Este caso es conocido como análisis de sensibilidad al cambio de forma. Aquí el problema típico es hallar la forma del dominio que es óptima en el sentido que minimiza un funcional costo apropiado. En muchos casos, el funcional considerado depende de la solución de un problema de borde definido justamente en dominio variable. Por ejemplo, en [32] se considera un problema inverso en el área de acústica y cuyo objetivo es hallar la forma de un objeto inmerso en un medio acústico. La forma óptima del objeto es la que minimiza un funcional que mide la diferencia entre la señal de scattering correspondiente a una solución de prueba y mediciones que contienen datos de la señal real medida en ciertos puntos. La señal de scattering satisface la ecuación de onda con condición de radiación de Sommerfeld en el infinito. Otro ejemplo que podemos mencionar está relacionado con el diseño aerodinámico. Un problema típico es el de encontrar la forma óptima de la superficie de elevación correspondiente al ala de un avión. El funcional costo en este caso depende de las variables que representan el flujo sobre la superficie del ala y estas variables son a su vez la solución de un problema de flujo que puede estar gobernado por la ecuación de Euler o Navier-Stokes. En ingeniería mecánica, la aparición y seguimiento del crecimiento de una fractura en un cuerpo elástico puede ser interpretado como un cambio en la forma del objeto y de esta manera es posible calcular la velocidad de liberación de energía. El funcional costo aquí es la energía potencial elástica del sistema y el problema variacional responde a una ecuación de equilibrio del cuerpo elástico [33, 66].

Lo que tienen en común todos los problemas anteriormente mencionados, es que pueden ser planteados como problemas de optimización. El objetivo en estos problemas es hallar un conjunto de parámetros, llamados variables de diseño o de control, las cuales dan el valor óptimo de una función costo apropiada que mide, de alguna manera, la eficiencia del sistema en estudio. A su vez, el sistema está gobernado por un problema de borde llamado problema directo o ecuación de estado y las variables de esta ecuación fundamental son llamadas variables de estado. Generalmente la función costo a minimizar depende de estas variables de estado. Por lo tanto, la ecuación de estado puede ser vista como una restricción a la que está sujeta la minimización del funcional costo.

Si el problema es diferenciable, algoritmos basados en el cálculo del gradiente son generalmente utilizados para resolver estos problemas de optimización. Además del cálculo del valor de la función costo y de las restricciones, estos métodos requieren el cálculo de las derivadas con respecto a las variables de diseño. En este Capítulo utilizaremos herramientas del análisis de sensibilidad para resolver los problemas inversos que hemos presentado en el Capítulo 3. Recordemos que en la sección 3.3 hemos reescrito ambos problemas inversos como problemas de optimización. Por lo tanto, la idea aquí es calcular la derivada de los funcionales costo utilizando las técnicas de análisis de sensibilidad. Por simplicidad comenzaremos con el problema que hemos denominado PI 2 y que consiste en hallar el valor correspondiente a la intensidad de la fuente de calor metabólico dentro del tumor. En este caso la variable de diseño es la fuente de actividad metabólica, de manera tal que la sensibilidad está asociada a la variación de un coeficiente o parámetro de la

ecuación de estado. Luego, trataremos el problema de la localización de la región tumoral. En este caso la variable de diseño es el dominio donde está definido el problema, por lo tanto introduciremos el concepto de derivada de forma con el objetivo de calcular la derivada del funcional costo con respecto al dominio.

4.2. Análisis de sensibilidad al cambio en los parámetros.

Imaginemos estar estudiando un fenómeno que es modelado a través de una ecuación diferencial y supongamos que el conjunto de variables de diseño o control está compuesto por parámetros o coeficientes que definen la ecuación diferencial. Denotemos este conjunto de parámetros de control por un vector $p = (p_1, \dots, p_d)$. De manera general, podemos plantear el problema como:

$$\begin{cases} \text{Hallar } u_p \text{ tal que} \\ a_p(u_p, v) = l_p(v), \quad \forall v \in \mathcal{W}(\Omega), \end{cases} \quad (4.3)$$

donde $\mathcal{W}(\Omega)$ es un espacio de Hilbert apropiado, $a_p(\cdot, \cdot)$ es una forma bilineal simétrica, coercitiva y continua, y $l_p(\cdot)$ una forma lineal continua. Con el subíndice p queremos enfatizar la dependencia de $a(\cdot, \cdot)$, $l(\cdot)$ y u en el vector de parámetros o variables de diseño p .

Supongamos además que para este sistema tenemos un funcional costo Ψ , por ahora genérico, que depende de los parámetros p de manera explícita pero también de manera implícita a través de la solución del problema variacional (4.3). Por lo tanto tenemos

$$\Psi(p) := \Psi(p, u_p). \quad (4.4)$$

El objetivo es encontrar un vector de parámetros o variables de diseño $p^* \in \mathcal{A}$, donde \mathcal{A} está definido como el conjunto de variables de diseño admisibles, tal que:

$$\Psi(p^*) \leq \Psi(p), \quad \forall p \in \mathcal{A}. \quad (4.5)$$

Observemos que resolver este problema conlleva a los siguientes: (i) existencia de un vector de variables de diseño óptimo, (ii) dar una caracterización de los posibles vectores de variables óptimos y (iii) construcción de un algoritmo que calcule el o los vectores con las variables de diseño óptimos. Estos dos últimos problemas están estrechamente relacionados con el cálculo de la derivada de la aplicación:

$$p \rightarrow \Psi(p) \in \mathbb{R}. \quad (4.6)$$

Consideremos entonces, una perturbación en el vector de parámetros del tipo

$$p_t = p + tV, \quad t \in \mathbb{R}^+, \quad (4.7)$$

donde $V = (\delta p_1, \delta p_2, \dots, \delta p_m)$ y $\mathbb{R}^+ = \{t \in \mathbb{R} : t \geq 0\}$.

Cabe señalar que para una perturbación suave más general, apelando al desarrollo de Taylor tenemos que

$$p(t) = p(0) + p'(0)t + O(t^2) = p + tV + O(t^2)V,$$

y para el cálculo de derivadas de primer orden los términos de $O(t^2)$ son despreciables, con lo cual basta considerar una perturbación del tipo (4.7).

Teniendo en cuenta esta perturbación, la nueva respuesta del sistema es la solución del siguiente problema:

$$\begin{cases} \text{Hallar } u_{p+tV} \text{ tal que} \\ a_{p+tV}(u_{p+tV}, v) = l_{p+tV}(v), \quad \forall v \in \mathcal{W}(\Omega) \end{cases} \quad (4.8)$$

y el nuevo valor del funcional costo está dado por

$$\psi(p+tV) = \Psi(p+tV, u_{p+tV}). \quad (4.9)$$

Tenemos entonces, que la sensibilidad del funcional ψ , causada por una perturbación en las variables de diseño en la dirección V , está dada por

$$\dot{\psi}(p; V) = \lim_{t \rightarrow 0} \frac{\Psi(p+tV) - \Psi(p)}{t}, \quad (4.10)$$

si este límite existe. En el caso que este límite exista para todo campo V , lo llamaremos derivada del funcional con respecto a las variables de diseño.

4.2.1. Una metodología para calcular la derivada con respecto a las variables de diseño.

En esta sección mostramos cómo calcular la derivada con respecto a los parámetros de diseño (4.10) utilizando el denominado método adjunto. Asumiremos que todas las funciones y funcionales tienen la regularidad suficiente de manera tal que el desarrollo que presentamos tenga sentido. Para un análisis riguroso de estos temas puede verse [70].

De acuerdo a lo visto en la sección previa la derivada de la función costo con respecto a las variables de diseño, en la dirección de perturbación V , está dada por

$$\dot{\psi}(p; V) = \left. \frac{d\Psi}{dt}(p+tV) \right|_{t=0}.$$

Con el objetivo de simplificar la escritura vamos a adoptar la siguiente notación:

$$a_t(u_t, v) = a_{p+tV}(u_{p+tV}, v), \quad (4.11)$$

$$l_t(v) = l_{p+tV}(v), \quad (4.12)$$

$$\Psi(t, u_t) = \Psi(p+tV, u_{p+tV}). \quad (4.13)$$

La dirección V en la que se calcula la derivada será clara a partir del contexto en el que estemos trabajando. La situación original se recupera cuando $t = 0$, obteniendo $u_0 = u_p$ y $\Psi(0, u_0) = \Psi(p, u_p)$.

Recordemos que en la sección anterior remarcamos que el funcional costo depende de p , tanto de manera explícita como de manera implícita a través de la solución del problema variacional (4.3), con lo cual tenemos que:

$$\dot{\psi}(p; V) = d\Psi(p, u_p; V) = \left. \frac{d\Psi}{dt}(t, u_t) \right|_{t=0}. \quad (4.14)$$

Ahora, si la función $\Psi : \mathbb{R} \times \mathcal{W}(\Omega) \rightarrow \mathbb{R}$ es derivable Fréchet, y por lo tanto Gâteaux, de acuerdo a los resultados listados en el apéndice, tenemos que

$$\left. \frac{d\Psi}{dt}(t, u_t) \right|_{t=0} = \left. \frac{d\Psi}{dt}(t, u_p) \right|_{t=0} + \left. \frac{d\Psi}{dt}(p, u_t) \right|_{t=0}, \quad (4.15)$$

donde los términos derechos de esta última ecuación están definidos por:

$$\left. \frac{d\Psi}{dt}(t, u) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{\Psi(t, u) - \Psi(0, u)}{t} =: \Psi'(p, u; V). \quad (4.16)$$

Notemos que hemos considerado $u \in \mathcal{W}(\Omega)$ fija, es decir, esta es la derivada parcial de Ψ con respecto a la primer variable p .

Por otra parte, aplicando la regla de la cadena, tenemos:

$$\frac{d\Psi}{dt}(p, u_t)|_{t=0} = \lim_{t \rightarrow 0} \frac{\Psi(p, u_t) - \Psi(p, u_0)}{t} =: \frac{\partial \Psi}{\partial u}(p, u_p; \dot{u}), \quad (4.17)$$

donde \dot{u} es la variación de la respuesta u , debida a la perturbación V y la cual viene dada por el límite

$$\lim_{t \rightarrow 0} \left\| \frac{u_t - u}{t} - \dot{u} \right\|_{\mathcal{W}(\Omega)} = 0. \quad (4.18)$$

Entonces, de acuerdo a las ecuaciones (4.15), (4.16) y (4.17) podemos reescribir la derivada de la función costo con respecto a las variables de diseño como

$$\psi(p; V) = \Psi'(p, u_p; V) + \frac{\partial \Psi}{\partial u}(p, u_p; \dot{u}). \quad (4.19)$$

Con lo cual, para obtener ψ necesitamos calcular \dot{u} , ya que es el único término desconocido en esta ecuación. Para esto derivamos la ecuación de estado con respecto a t y evaluamos en $t = 0$:

$$\frac{d}{dt} a_t(u_t, v; V)|_{t=0} = \frac{d}{dt} l_t(v; V)|_{t=0}.$$

De manera similar a lo realizado con la función costo, podemos definir las siguiente derivadas parciales de la forma bilineal y lineal:

$$\begin{aligned} a'(u, v; V) &:= \lim_{t \rightarrow 0} \frac{a_t(u, v) - a(u, v)}{t} = \frac{d}{dt} a_t(u, v; V)|_{t=0}, \\ \frac{\partial a}{\partial u}(u, v; \dot{u}) &:= \lim_{t \rightarrow 0} \frac{a(u_t, v) - a(u, v)}{t} = a(\dot{u}, v), \\ l'(v; V) &:= \lim_{t \rightarrow 0} \frac{l_t(v) - l(v)}{t} = \frac{d}{dt} l_t(v; V)|_{t=0}. \end{aligned}$$

Luego, asumiendo que estas derivadas son lineales y continuas en la dirección de la perturbación V , derivando la ecuación de estado obtenemos

$$\begin{aligned} \frac{d}{dt} a_t(u_t, v; V)|_{t=0} &= \frac{d}{dt} l_t(v; V)|_{t=0}, \\ a'(u_p, v; V) + a(\dot{u}, v) &= l'(v; V). \end{aligned}$$

Para evitar el cálculo de \dot{u} , definimos la siguiente ecuación adjunta [31]:

$$\begin{cases} \text{Hallar } \lambda \text{ tal que} \\ a(\lambda, \bar{\lambda}) = -\frac{\partial \Psi}{\partial u}(p, u_p; \bar{\lambda}), \quad \forall \bar{\lambda} \in \mathcal{W}(\Omega) \end{cases} \quad (4.20)$$

La solución de esta ecuación, denotada por λ , se denomina estado adjunto. Como la ecuación es válida para todo $\bar{\lambda} \in \mathcal{W}(\Omega)$, en particular es válida si consideramos $\bar{\lambda} = \dot{u}$ y usando la simetría de la forma bilineal y la ecuación (4.20), tenemos:

$$-\frac{\partial \Psi}{\partial u}(p, u_p; \dot{u}) = a(\dot{u}, \lambda) = l'(\lambda; V) - a'(u_p, \lambda; V). \quad (4.21)$$

Finalmente, reemplazando esta última expresión en (4.19) podemos expresar la derivada total de la función costo como:

$$\psi(p; V) = \Psi'(p, u_p; V) + a'(u_p, \lambda; V) - l'(\lambda; V), \quad (4.22)$$

donde λ es la solución del problema adjunto (4.20). Por lo tanto, para calcular la derivada de la función costo con respecto a los parámetros de diseño utilizando el método adjunto, no es necesario calcular \dot{u} , sino que simplemente debemos resolver el problema original (4.3) y el problema adjunto (4.20).

4.3. Análisis de sensibilidad al cambio de forma

En esta sección presentamos los conceptos fundamentales para calcular la derivada con respecto al cambio en la forma del dominio donde está definido el problema que se está estudiando. Desde el punto de vista matemático este caso es más complejo que el de la sensibilidad al cambio en los coeficientes que aparecen en la ecuación diferencial. La presentación será desarrollada siguiendo las ideas explicadas en la sección previa.

Supongamos que el fenómeno o sistema que estamos estudiando está gobernado por el siguiente problema:

$$\begin{cases} \text{Hallar } u_\Omega \in \mathcal{W}(\Omega) \text{ tal que} \\ a_\Omega(u_\Omega, v) = l_\Omega(v), \quad \forall v \in \mathcal{W}(\Omega), \end{cases} \quad (4.23)$$

donde $\mathcal{W}(\Omega)$ es un espacio de Hilbert apropiado, $a_\Omega(\cdot, \cdot)$ es una forma bilineal simétrica, coercitiva y continua, y $l_\Omega(\cdot)$ una forma lineal continua. Con el subíndice Ω queremos enfatizar la dependencia de $a(\cdot, \cdot)$, $l(\cdot)$ y u en el dominio Ω .

Al igual que en la sección previa asumimos que para este sistema tenemos un funcional costo Ψ que depende de Ω de manera explícita y también de manera implícita a través de la solución del problema (4.23). Por lo tanto tenemos

$$\psi(\Omega) = \Psi(\Omega, u_\Omega). \quad (4.24)$$

El objetivo es encontrar un dominio Ω^* perteneciente a una familia \mathcal{A} de dominios admisibles tal que:

$$\psi(\Omega^*) \leq \psi(\Omega), \quad \forall \Omega \in \mathcal{A}. \quad (4.25)$$

Nuevamente, notemos que hallar un dominio óptimo implica los siguientes problemas: (i) existencia de un dominio óptimo, (ii) dar una caracterización de los dominios óptimos y (iii) construcción de un algoritmo que calcule el o los dominios óptimos. Estos dos últimos problemas están asociados con el cálculo de la derivada de la aplicación:

$$\Omega \rightarrow \psi(\Omega) \in \mathbb{R}. \quad (4.26)$$

Un inconveniente para definir tal derivada es que el conjunto \mathcal{A} carece de una estructura de espacio vectorial. Una manera de resolver esto es utilizar el método propuesto originalmente en [81] y también desarrollado en [41], conocido en la literatura como “speed method”. La idea central del método consiste en: dado un campo vectorial $V : \mathbb{R}^n \rightarrow \mathbb{R}^n$, definir un mapa desde el dominio de referencia Ω a un nuevo dominio perturbado de la siguiente manera

$$x_t = p_t(x) = x + tV(x), \quad t \in \mathbb{R}^+. \quad (4.27)$$

donde $\mathbb{R}^+ = \{t \geq 0\}$. En palabras, este método simula una transformación gradual en la forma del dominio de referencia, de acuerdo a una dirección conocida V , a un dominio perturbado definido por:

$$\begin{aligned} \Omega_t &= \{x_t \in \mathbb{R}^n : \exists x \in \Omega, x_t = p_t(x)\}, \\ \partial\Omega_t &= \{x_t \in \mathbb{R}^n : \exists x \in \partial\Omega, x_t = p_t(x)\}. \end{aligned}$$

En esta nueva configuración perturbada Ω_t debemos considerar una nueva ecuación de estado:

$$\begin{cases} \text{Hallar } u_{\Omega_t} \in \mathcal{W}(\Omega_t) \text{ tal que} \\ a_{\Omega_t}(u_{\Omega_t}, v) = l_{\Omega_t}(v), \quad \forall v \in \mathcal{W}(\Omega_t), \end{cases}$$

Aquí u_{Ω_t} es la respuesta del sistema en la nueva configuración perturbada Ω_t y el nuevo valor de la función costo está dado por

$$\psi(\Omega_t) = \Psi(\Omega_t, u_{\Omega_t}).$$

De esta forma, tenemos que la sensibilidad de ψ causada por una perturbación del sistema en la dirección del campo V es

$$\dot{\psi}(\Omega; V) = \lim_{t \rightarrow 0} \frac{\psi(\Omega_t) - \psi(\Omega)}{t} \quad (4.28)$$

$$= \lim_{t \rightarrow 0} \frac{\Psi(\Omega_t, u_{\Omega_t}) - \Psi(\Omega, u)}{t}, \quad u_{\Omega_t} \in \mathcal{W}(\Omega_t), u \in \mathcal{W}(\Omega). \quad (4.29)$$

Hay que resaltar que en la expresión de arriba los elementos u y u_{Ω_t} pertenecen a espacios topológicos distintos. Una manera útil de calcular (4.28) se obtiene utilizando el mapa definido en la ecuación (4.27) de manera tal de transformar todas las expresiones definidas en la configuración Ω_t a expresiones definidas en la configuración de referencia Ω . Por ejemplo, la solución a la ecuación de estado será ahora

$$u_t(x) := u_{\Omega_t}(p_t(x)) = u_{\Omega_t}(x_t), \quad (4.30)$$

función definida en el dominio de referencia Ω , es decir $u_t \in \mathcal{W}(\Omega)$. Por su parte, la función costo está ahora definida por

$$\Psi(\Omega, t, u_t) := \Psi(\Omega_t, u_{\Omega_t}) = \Psi_{\Omega_t} \circ p_t(\Omega). \quad (4.31)$$

La situación original se recupera cuando $t = 0$, obteniendo $u_0 = u_{\Omega} =: u$ y $\Psi(\Omega, 0, u_0) = \Psi(\Omega, u)$. Realizando el mismo cambio de variables la ecuación de estado puede ser reescrita como:

$$\begin{cases} \text{Hallar } u_t \in \mathcal{W}(\Omega) \text{ tal que} \\ a_t(u_t, v) = l_t(v), \quad \forall v \in \mathcal{W}(\Omega), \end{cases} \quad (4.32)$$

donde $a_t : \mathcal{W}(\Omega) \times \mathcal{W}(\Omega) \rightarrow \mathbb{R}$ y $l_t : \mathcal{W}(\Omega) \rightarrow \mathbb{R}$. Finalmente la sensibilidad de ψ causada por una perturbación del sistema en la dirección de V está dada por

$$\dot{\psi}(\Omega; V) = \lim_{t \rightarrow 0} \frac{\Psi(\Omega, t, u_t) - \Psi(\Omega, 0, u)}{t}, \quad u_t, u \in \mathcal{W}(\Omega). \quad (4.33)$$

si este límite existe y lo llamaremos derivada de forma.

4.3.1. Una metodología para calcular la derivada de forma.

Como vimos en la sección previa, el método propuesto en [81] conocido como “speed method” puede ser interpretado de la siguiente manera. Dada una dirección de perturbación V , la variable de diseño Ω está controlada de manera única por el parámetro $t \in \mathbb{R}^+$. De esta manera la derivada de forma de la función costo, en la dirección del campo vectorial V , es simplemente

$$\dot{\psi}(\Omega; V) = \frac{d}{dt} \Psi(t, u_t) |_{t=0}. \quad (4.34)$$

Notemos que no hemos escrito Ω en el término derecho de ecuación (4.34) ya que ahora todas las expresiones están definidas en el dominio de referencia Ω .

4.3.1.A. Operaciones básicas de derivación.

Para escribir en forma explícita la expresión para la derivada de forma (4.34) es necesario utilizar ciertas operaciones básicas de derivación, las cuales serán listadas a continuación para facilitar la lectura del trabajo.

Recordemos que el mapa que transforma el dominio de referencia Ω en un dominio perturbado Ω_t , está dado por la ecuación

$$x_t = p_t(x) = x + tV(x),$$

donde $V(x)$ es un campo vectorial suficientemente regular definido en un abierto que contiene al conjunto Ω . Este campo vectorial describe la velocidad de cada punto en el instante t . La derivada de este mapa con respecto a un punto material x es:

$$Dp_t(x) = (Id + tDV)(x) =: F_t(x).$$

Utilizando esta última ecuación y apelando a resultados conocidos, obtenemos las siguiente identidades:

- $F_t^{-1} = Id - tDV + O(t^2)$ y por lo tanto $F_{t=0}^{-1} = F_{t=0}^{-1} = Id$,
- $\frac{d}{dt}(F_t) = DV$,
- $\frac{d}{dt}(F_t^{-1})|_{t=0} = -DV$,
- $\det(F_t) = \det(I + tDV) = 1 + t \operatorname{tr}(DV) + O(t^2) = 1 + t \operatorname{div}(V) + O(t^2)$,
- $\frac{d}{dt} \det(F_t)|_{t=0} = \operatorname{div}(V)$,
- $\det(F_t^{-1}) = \det(Id - tDV + O(t^2)) = 1 - t \operatorname{div}(V) + O(t^2)$,
- $\frac{d}{dt}(\det(F_t^{-1}))|_{t=0} = -\operatorname{div}(V)$,
- $dx_t = \det(F_t)dx$,
- $d\Gamma_t = \|F_t^{-T}\eta\| \det(F_t)d\Gamma$,
- $\frac{d}{dt}(\|F_t^{-T}\eta\|)|_{t=0} = -\eta \cdot (DV)^T \eta$,
- $\frac{d}{dt}(\|F_t^{-T}\eta\| \det(F_t))|_{t=0} = -\eta \cdot (DV)^T \eta + \operatorname{div}(V) =: \operatorname{div}_\Gamma(V)$.

Teniendo en cuenta los resultados listados arriba, a continuación mostramos cómo derivar funciones que pueden estar definidas por integrales en el dominio Ω o integrales sobre su frontera o parte de ésta, la que denotamos por Γ .

Caso 1: Sea ϕ una función definida por

$$\phi = \int_{\Omega} \varphi(x) dx$$

donde φ es una función suave definida en Ω . Utilizando el mapa (4.27), tenemos la familia de funcionales que dependen del parámetro t dada por

$$\Phi^t = \int_{\Omega_t} \varphi_{\Omega_t}(x_t) dx_t.$$

Denotando por $\varphi_t(x) = (\varphi_{\Omega_t} \circ p_t)(x) = \varphi_{\Omega_t}(x + tV(x))$, función definida en el dominio de referencia Ω , tenemos la nueva familia

$$\Phi_t = \int_{\Omega} \varphi_t(x) \det(F_t) dx.$$

Luego, la derivada de Φ_t con respecto a t , en $t = 0$ y en la dirección V , está dada por:

$$\begin{aligned} \dot{\Phi} &= \frac{d}{dt} \left(\int_{\Omega} \varphi_t(x) \det(F_t) dx \right) \Big|_{t=0} \\ &= \int_{\Omega} \frac{d}{dt} (\varphi_t(x) \det(F_t)) \Big|_{t=0} dx \\ &= \int_{\Omega} (\dot{\varphi}_t(x) + \varphi(x) \operatorname{div}(V)) dx. \end{aligned}$$

Caso 2: Sea ϕ una función definida por

$$\phi = \int_{\Gamma} \zeta(x) d\Gamma$$

donde ζ es una función suave definida en Γ . Nuevamente, mediante la aplicación del mapa (4.27), consideramos una familia de funciones que depende de t definidas por

$$\Phi^t = \int_{\Gamma_t} \zeta^t(x_t) d\Gamma_t.$$

Transformando estas funciones al dominio de referencia Ω , tenemos que la derivada de Φ_t con respecto a t , en $t = 0$ y en la dirección V , está dada por:

$$\begin{aligned} \dot{\Phi} &= \frac{d}{dt} \left(\int_{\Gamma} \zeta_t(x) \det(F_t) \|F_t^{-T} \eta\| d\Gamma \right) \Big|_{t=0} \\ &= \int_{\Gamma} \frac{d}{dt} (\zeta_t(x) \det(F_t) \|F_t^{-T} \eta\|) \Big|_{t=0} d\Gamma \\ &= \int_{\Gamma} (\dot{\zeta}_t(x) + \zeta(x) \operatorname{div}_{\Gamma}(V)) d\Gamma. \end{aligned}$$

4.3.1.B. Cálculo de la derivada de forma.

Consideremos un funcional costo que dependa explícita e implícitamente del dominio Ω , el caso más general es del tipo

$$\Psi(\Omega) = \Psi(\Omega, u_{\Omega}) = \int_{\Omega} \mathcal{G}(u_{\Omega}) dx + \int_{\Gamma} \mathcal{J}(u_{\Omega}) d\Gamma. \quad (4.35)$$

Utilizando los resultados presentados anteriormente podemos calcular la derivada de forma (4.34) para este funcional costo como mostramos a continuación:

$$\begin{aligned} \dot{\Psi}(\Omega; V) &= \frac{d}{dt} \Psi(t, u_t) \Big|_{t=0} \\ &= \int_{\Omega} \frac{d}{dt} (\mathcal{G}_t(u_t) \det(F_t)) dx + \int_{\Gamma} \frac{d}{dt} (\mathcal{J}_t(u_t) \det(F_t) \|F_t^{-T} \eta\|) \Big|_{t=0} d\Gamma \\ &= \int_{\Omega} (\dot{\mathcal{G}}(u) + \mathcal{G}(u) \operatorname{div}(V)) dx + \int_{\Gamma} (\dot{\mathcal{J}}(u) + \mathcal{J}(u) \operatorname{div}_{\Gamma}(V)) d\Gamma. \end{aligned} \quad (4.36)$$

Observemos que las funciones $\mathcal{G}_t(u_t) = \mathcal{G}(t, u_t)$ y $\mathcal{J}_t(u_t) = \mathcal{J}(t, u_t)$ dependen explícita e implícitamente de t . Por lo tanto, de manera similar a lo realizado en la sección (3.2.1) podemos escribir

$$\dot{\mathcal{G}}(u) = \mathcal{G}'(u) + \frac{\partial \mathcal{G}}{\partial u}(u; \dot{u}), \quad (4.37)$$

$$\dot{\mathcal{J}}(u) = \mathcal{J}'(u) + \frac{\partial \mathcal{J}}{\partial u}(u; \dot{u}), \quad (4.38)$$

donde \dot{u} es la variación de la respuesta u debida a la perturbación V dada por el límite

$$\lim_{t \rightarrow 0} \left\| \frac{u_t(x) - u(x)}{t} - \dot{u}(x; V) \right\|_{\mathcal{W}(\Omega)} = 0, \quad (4.39)$$

y la que usualmente se denomina derivada material.

Observemos que a partir de las ecuaciones (4.36), (4.37) y (4.38), al igual que para la derivada con respecto a los parámetros de diseño, obtenemos que la derivada con respecto a la forma puede escribirse como

$$\psi(\Omega; V) = \Psi'(0, u; V) + \frac{\partial \Psi}{\partial u}(0, u; \dot{u}). \quad (4.40)$$

Nuevamente, para obtener ψ es necesario conocer \dot{u} . Al igual que para el caso en que las variables de diseño eran los coeficientes de la ecuación, derivamos la ecuación de estado con respecto a t , en $t = 0$:

$$\frac{d}{dt} a_t(u_t, v; V)|_{t=0} = \frac{d}{dt} l_t(v; V)|_{t=0}.$$

Asumiendo que estas derivadas son lineales y continuas en la dirección de la perturbación V tenemos que

$$\frac{d}{dt} a_t(u_t, v; V)|_{t=0} = \frac{d}{dt} l_t(v; V)|_{t=0}, \quad (4.41)$$

$$a'(u, v; V) + a(\dot{u}, v) = l'(v; V). \quad (4.42)$$

nuevamente, para evitar el cálculo explícito de \dot{u} , definimos la siguiente ecuación adjunta:

$$\begin{cases} \text{Hallar } \lambda \text{ tal que} \\ a(\lambda, \bar{\lambda}) = -\frac{\partial \Psi}{\partial u}(0, u; \bar{\lambda}), \quad \forall \bar{\lambda} \in \mathcal{W}(\Omega) \end{cases} . \quad (4.43)$$

Como la ecuación es válida para todo $\bar{\lambda} \in \mathcal{W}(\Omega)$, en particular es válida si consideramos $\bar{\lambda} = \dot{u}$ y usando la simetría de la forma bilineal y la ecuación (4.42), obtenemos

$$-\frac{\partial \Psi}{\partial u}(0, u; \dot{u}) = a(\dot{u}, \lambda) = l'(\lambda; V) - a'(u, \lambda; V). \quad (4.44)$$

Finalmente, reemplazando esta última expresión en (4.40) podemos expresar la derivada de la función costo como:

$$\psi(\Omega; V) = \Psi'(u; V) + a'(u, \lambda; V) - l'(\lambda; V), \quad (4.45)$$

donde u es solución de (4.23) y λ es solución del problema adjunto (4.43).

Por lo tanto, para calcular la derivada de la función costo con respecto a cambios en la forma del dominio, al utilizar el método adjunto no es necesario calcular \dot{u} , sino que simplemente debemos resolver la ecuación de estado para u , la correspondiente ecuación adjunta y luego evaluar las expresiones Ψ' , a' y l' .

Observación: utilizar el método adjunto para evitar el cálculo explícito de \dot{u} es similar a emplear el método de Lagrange. En efecto, consideremos la función costo (4.24) a la que le sumamos la ecuación de estado (4.23) como una restricción, quedando así definido el Lagrangiano:

$$\mathcal{L}(\Omega, w, \tau) = \Psi(\Omega, \tau) + a(w, \tau) - l(\tau), \quad (4.46)$$

Utilizando el mapa (4.27), tenemos la familia que depende del parámetro t

$$\mathcal{L}^t(\Omega_t, w_{\Omega_t}, \tau_{\Omega_t}) = \Psi(\Omega_t, w_{\Omega_t}) + a_{\Omega_t}(w_{\Omega_t}, \tau_{\Omega_t}) - l_{\Omega_t}(\tau_{\Omega_t}).$$

Ahora, realizando el cambio de variables de manera tal que todas las expresiones queden definidas en Ω , tenemos:

$$\mathcal{L}_t(w_t, \tau_t) = \Psi(t, w_t) + a_t(w_t, \tau_t) - l_t(\tau_t). \quad (4.47)$$

Observemos que para $w_t = u_t$ solución de la ecuación de estado, resulta

$$\mathcal{L}_t(u_t, \tau_t) = \Psi(t, u_t). \quad (4.48)$$

Fijada una dirección de perturbación V y asumiendo suficiente regularidad en cada uno de los términos tenemos que la derivada de \mathcal{L}_t respecto a t , en $t = 0$, está dada por:

$$\dot{\mathcal{L}}(w, \tau; V) = \mathcal{L}'(w, \tau; V) + \frac{\partial \mathcal{L}}{\partial w}(w, \tau; \dot{w}) + \frac{\partial \mathcal{L}}{\partial \tau}(w, \tau; \dot{\tau}). \quad (4.49)$$

Donde, de acuerdo a la definición del Lagrangiano (4.46), tenemos

$$\frac{\partial \mathcal{L}}{\partial \tau}(w, \tau; \dot{\tau}) = a(w, \dot{\tau}) - l(\dot{\tau}), \quad (4.50)$$

$$\frac{\partial \mathcal{L}}{\partial w}(w, \tau; \dot{w}) = \frac{\partial \Psi}{\partial w}(0, w; \dot{w}) + a(\dot{w}, \tau). \quad (4.51)$$

Luego, si a estas dos últimas ecuaciones las evaluamos en $w = u$ solución del problema (4.23) y en $\tau = \lambda$ solución del problema adjunto (4.69) obtenemos

$$\frac{\partial \mathcal{L}}{\partial \tau}(u, \lambda; \dot{\lambda}) = a(u, \dot{\lambda}) - l(\dot{\lambda}) = 0, \quad (4.52)$$

$$\frac{\partial \mathcal{L}}{\partial w}(u, \lambda; \dot{u}) = \frac{\partial \Psi}{\partial w}(0, u; \dot{u}) + a(\dot{u}, \lambda) = 0. \quad (4.53)$$

Por lo tanto, de (4.49), (4.52), (4.53) y (4.22) concluimos que

$$\begin{aligned} \dot{\mathcal{L}}(\Omega, u, \lambda; V) &= \mathcal{L}'(u, \lambda; V) \\ &= \Psi'(0, u; V) + a'(u, \lambda; V) - l'(\lambda; V) \\ &= \dot{\Psi}(\Omega; V). \end{aligned}$$

4.4. Resolución del problema: PI 1

4.4.1. Cálculo de la derivada de forma

A continuación utilizaremos los resultados obtenidos en la sección previa para calcular la derivada de forma para el caso de nuestro problema particular que hemos denominado PI 1 y que consiste en estimar la localización y tamaño de la región tumoral.

En términos matemáticos, estamos interesados en hallar un subdominio $\omega^* \subset \Omega$, el cual representa la región tumoral dentro del cuerpo y sobre el cual asumimos que se encuentra alejado de la frontera, esto es $dist(\Omega, \omega^*) > \varepsilon > 0$. De esta última suposición vamos entonces a considerar que $\partial\Omega$ se mantiene fija durante todo el proceso, esto es $\partial\Omega_t = \partial\Omega, \forall t \in \mathbb{R}^+$ y sólo movimientos dentro del dominio están permitidos. Así, la diferencia entre Ω y Ω_t es simplemente la localización y forma de los respectivos subdominios ω y ω_t . Uno puede imaginar que el dominio está hecho de un material maleable el cual se encuentra fijo en sus extremos pero puede ser deformado en su interior. Con este propósito consideraremos V tal que su soporte esté incluido en Ω . Más adelante se dará una definición exacta del campo vectorial V .

En nuestro problema la función costo a minimizar es:

$$\psi(\omega) := \Psi(u_\omega) = \frac{1}{2} \int_{\Gamma_s} (u_\omega - u_{exp})^2 d\Gamma, \quad (4.54)$$

y representa el error o diferencia entre un perfil de temperatura u_{exp} medido experimentalmente sobre la superficie corporal y la solución u_ω del problema de transferencia de calor 2.3 considerando el subdominio ω . Cabe señalar que hemos denotado u_ω en lugar de u_Ω ya que para este problema Ω está fijo, y por lo tanto u depende precisamente del subdominio ω . También, notemos que en este caso la función costo Ψ no depende de manera explícita en el dominio Ω , ya que su frontera está fija durante todo el proceso y por lo tanto el borde superior Γ_s es mantenido fijo durante todo el proceso. De esta manera, la ecuación (4.40) se reduce a

$$\psi(\omega; V) = a'(u, \lambda; V) - l'(\lambda; V), \quad (4.55)$$

entonces sólo debemos calcular las expresiones $a'(\cdot, \cdot)$ y $l'(\cdot)$.

Para transformar las ecuaciones al dominio de referencia Ω , hacemos uso del mapa

$$x_t = p_t(x) = x + t V(x),$$

su jacobiano

$$Dp_t(x) = Id + tDV(x) =: F_t(x)$$

y de los resultados presentados en la sección 4.3.1.

De acuerdo a las ecuaciones (2.17) y (2.18) tenemos que la forma bilineal y la forma lineal quedan definidas por:

$$\begin{aligned} a_t(u, v; V) &= \int_{\Omega} [\sigma_t(\nabla u F_t^{-1}) \cdot (\nabla v F_t^{-1}) + k_t uv] \det F_t dx + \alpha \int_{\Gamma_s} uv \det F_t \|F_t^{-T}\| d\Gamma, \\ l_t(v; V) &= \int_{\Omega} q_t v \det F_t dx + \alpha(T_a - T_b) \int_{\Gamma_s} v \det F_t \|F_t^{-T}\| d\Gamma. \end{aligned}$$

Luego, teniendo en cuenta las definiciones (4.20) y (4.20) las derivadas de estas expresiones tiene la siguiente forma:

$$\begin{aligned} a'(u, v; V) &= \int_{\Omega} \frac{d}{dt} \{ [\sigma_t(\nabla u F_t^{-1}) \cdot (\nabla v F_t^{-1}) + k_t uv] \det F_t \}_{|_{t=0}} dx \\ &\quad + \alpha \int_{\Gamma_s} uv \frac{d}{dt} (\det F_t \|F_t^{-T}\|)_{|_{t=0}} d\Gamma, \end{aligned} \quad (4.56)$$

$$l'(v; V) = \int_{\Omega} \frac{d}{dt} (q_t v \det F_t)_{|_{t=0}} dx + \alpha(T_a - T_b) \int_{\Gamma_s} v \frac{d}{dt} (\det F_t \|F_t^{-T}\|)_{|_{t=0}} d\Gamma. \quad (4.57)$$

Ahora, usando el hecho que $\text{supp}(V) \subset \Omega$, tenemos que $\frac{d}{dt}(F_t) = DV = 0$ y $\text{div}(V) = 0$ sobre $\partial\Omega$, con lo cual

$$\int_{\Gamma_s} w \frac{d}{dt} (\det F_t \|F_t^{-T}\|)_{|_{t=0}} d\Gamma = 0, \quad \forall w(x) \in V(\Omega),$$

y entonces las expresiones (4.56) y (4.57) se transforman en:

$$a'(u, v; V) = \int_{\Omega} \frac{d}{dt} \{ [\sigma_t (\nabla u F_t^{-1}) \cdot (\nabla v F_t^{-1}) + k_t uv] \det F_t \}_{|_{t=0}} dx, \quad (4.58)$$

$$l'(v; V) = \int_{\Omega} \frac{d}{dt} (q_t v \det F_t)_{|_{t=0}} dx. \quad (4.59)$$

Asumiendo que el mapa $p_t : \Omega \rightarrow \Omega_t$ es además de regular, inyectivo y tal que

$$p_t^{-1}(\omega_t) = \omega_0 = \omega,$$

tenemos por ejemplo:

$$\sigma_t(x) = (\sigma_{\Omega_t} \circ p_t)(x) = \begin{cases} \sigma_0 & \text{si } x \in \omega; \\ \sigma_1 & \text{si } x \in \Omega - \omega \end{cases} \quad \forall t \in \mathbb{R}^+ \geq 0,$$

donde ω_0 es el subdominio original incluido en Ω . Por lo tanto, la conductividad σ_t es contante con respecto a la variable t y así

$$\frac{d}{dt} \sigma_t = 0.$$

El mismo resultado es válido para los coeficientes k_t and q_t . De esta manera las expresiones (4.58) y (4.59) pueden ser reescritas como

$$a'(u, v; V) = \int_{\Omega} \sigma \frac{d}{dt} [(\nabla u F_t^{-1}) \cdot (\nabla v F_t^{-1}) \det F_t]_{|_{t=0}} dx + \int_{\Omega} k uv \frac{d}{dt} (\det F_t)_{|_{t=0}} dx, \quad (4.60)$$

$$l'(v; V) = \int_{\Omega} q v \frac{d}{dt} (\det F_t)_{|_{t=0}} dx. \quad (4.61)$$

Finalmente, de las identidades listadas en la sección 4.3.1.A , obtenemos:

$$a'(u, v; V) = \int_{\Omega} \sigma [-(\nabla u DV) \cdot \nabla v - \nabla u \cdot (\nabla v DV) + \nabla u \cdot \nabla v \text{div}(V)] dx \quad (4.62)$$

$$+ \int_{\Omega} k uv \text{div}(V) dx,$$

$$l'(v; V) = \int_{\Omega} q v \text{div}(V) dx. \quad (4.63)$$

Por lo tanto, de (4.55), (4.62) y (4.63), la derivada de forma de la función costo (4.54) está dada por:

$$\begin{aligned} \psi(\omega; V) &= \int_{\Omega} \sigma [-(\nabla u DV) \cdot \nabla \lambda - \nabla u \cdot (\nabla \lambda DV) + \nabla u \cdot \nabla \lambda \text{div}(V)] dx \\ &+ \int_{\Omega} k u \lambda \text{div}(V) dx - \int_{\Omega} q \lambda \text{div}(V) dx, \end{aligned} \quad (4.64)$$

donde u es la solución de (2.3) en el dominio de referencia y λ es la solución del correspondiente problema adjunto (4.43).

4.4.2. Resolución numérica.

En esta parte discutimos la aproximación numérica utilizada para calcular la derivada de forma utilizando el método adjunto. Describiremos cada uno de los ingredientes necesarios para calcular ésta derivada: la resolución numérica del estado primitivo u y del estado adjunto λ , la parametrización del subdominio ω y la definición del campo vectorial V .

Tanto para el caso de resolver el problema (2.3) y así obtener el estado u como el caso de resolver su correspondiente problema adjunto y obtener λ , hemos considerado un esquema de diferencias finitas de segundo orden siguiendo las ideas presentadas en el Capítulo 2.

Nuevamente, motivados por el hecho de localizar melanomas nodulares, asumiremos que el tumor tiene una forma esférica. Por lo tanto para predecir la posición y tamaño es necesario determinar el centro y radio de la circunferencia. De esta manera, las variables de diseño para este problema son la coordenada horizontal x_c y la coordenada vertical y_c que determinan el centro del tumor como así también el radio R del mismo. Desde el punto de vista matemático esta suposición sobre la forma del tumor implica una reducción notable en la dimensión del conjunto de dominios admisibles, \mathcal{A} , donde minimizaremos el funcional ψ . Bajo esta hipótesis este conjunto queda definido como:

$$\mathcal{A} = \{\omega = B(x_c, y_c; R) : x_l \leq x_c \leq x_u, y_l \leq y_c \leq y_u, 0 \leq R \leq R_u\}$$

donde $B(x_c, y_c; R)$ denota la circunferencia centrada en el punto (x_c, y_c) y de radio igual a R , los valores x_l, x_u son cotas inferiores y superiores para la variable x y los mismos para y_l, y_u y R_u

Para cada variable de diseño hemos definido un campo vectorial particular V en Ω . Sin embargo, todos los campos comparten el hecho de que han sido definidos por partes, tomando cierto valor dentro del subdominio ω y por extensión fuera de este ω . Hemos considerado una extensión en forma de anillo y también que el campo vectorial se anula fuera de éste (ver Figura 4.1). De esta manera, tenemos que $\text{supp}(V) \subset \Omega$ como supusimos en la sección previa. Para la variable de diseño x_c consideraremos el siguiente campo

$$V(x, y) = \begin{cases} V_{int}(x, y) & \text{si } (x, y) \in \bar{\omega}; \\ V_{ext}(x, y) & \text{si } (x, y) \in \Omega - \omega, \end{cases}$$

donde $V_{int}(x, y) = (1, 0)$ para todos los puntos que están dentro de la circunferencia y V_{ext} es una extensión regular. En este trabajo hemos escogido construir la extensión de cada campo vectorial siguiendo las ideas que se presentan en [55] y que consideran la siguiente extensión:

$$V_{ext}(x, y) = \frac{V_{int}(x, y) \left(\frac{w_1^p}{\text{dist}((x, y), R)^\beta} \right)}{\frac{w_1^p}{\text{dist}((x, y), R)^\beta} + \frac{w_2^p}{\text{dist}((x, y), R_{ext})^\beta}}, \quad (4.65)$$

donde:

- x_c, y_c denotan las coordenadas del centro de la región tumoral ω , mientras que R denota su radio,
- R_{ext} es un número real a ser elegido y tal que $0 < R < R_{ext}$,
- $\text{dist}((x, y), R)$ denota la distancia del punto $(x, y) \in \Omega - \bar{\omega}$ al borde de la circunferencia de radio R ,

- $dist((x,y), R_{ext})$ denota la distancia del punto $(x,y) \in \Omega - \bar{\omega}$ al borde de la circunferencia de radio R_{ext} ,
- w_1, w_2 y β son parámetros a ser elegidos.

La función de los parámetros w_1, w_2 y β , es de controlar la regularidad del campo vectorial y determinar que partes del dominio contribuyen significativamente en la expresión de V , para más detalles ver [33].

De manera similar, para la variable y_c consideramos el siguiente campo vectorial:

$$V(x,y) = \begin{cases} V_{int}(x,y) & \text{si } (x,y) \in \bar{\omega}; \\ V_{ext}(x,y) & \text{si } (x,y) \in \Omega - \omega, \end{cases}$$

donde $V_{int}(x,y) = (0, 1)$ y V_{ext} está definido por (4.65). Finalmente para la última variable de diseño, el radio R , el campo vectorial en el interior de la circunferencia es $V_{int}(x,y) = (\frac{x-x_c}{R}, \frac{y-y_c}{R})$ mientras que V_{ext} también está definido de acuerdo a la expresión (4.65).

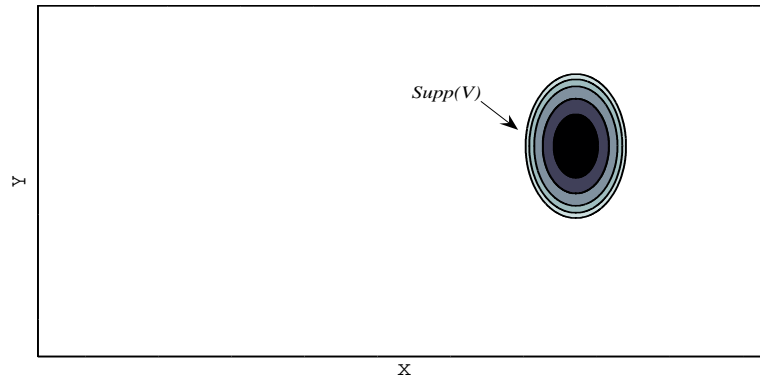


Figura 4.1: El campo vectorial V_{int} definido en la región tumoral ω (región en negro), es extendido a un campo vectorial V definido en todo el dominio Ω , pero tal que $supp(V) \subset \Omega$.

Luego de todas las consideraciones realizadas, el problema de hallar la localización y tamaño del tumor puede ser planteado como el siguiente problema de optimización:

$$\begin{aligned} & \text{Minimizar } \Psi(x_c, y_c, R) \\ & \text{sujeto a} \\ & (x_c, y_c, R) \in \mathcal{C} \end{aligned}$$

donde \mathcal{C} es un conjunto cerrado y convexo en \mathbb{R}^3 definido por

$$\mathcal{C} = \{(x, y, z) \in \mathbb{R}^3 : x_l \leq x \leq x_u, y_l \leq y \leq y_u, z_l \leq z \leq z_u\}.$$

Recordemos que la restricción impuesta por la ecuación de estado ha sido eliminada utilizando el problema adjunto.

Para resolver el problema de optimización hemos implementado un método de gradiente para problemas en conjuntos convexos siguiendo las ideas descritas en Birgin *et al.* [13]. A grandes

rasgos, este algoritmo combina el método del gradiente proyectado [12] y agrega dos nuevas características. La primera, es que extiende las estrategias típicas de globalización asociadas a estos métodos al caso de línea de búsqueda no monótona siguiendo el esquema desarrollado por Grippo *et al.* [38] para el método de Newton. En segundo lugar, también utiliza la elección del paso introducida en 1988 por Barzilai y Borwein [10] y conocida como método espectral. Esta elección de paso requiere de un pequeño trabajo computacional pero ayuda a acelerar notablemente la convergencia de los métodos de gradiente.

En resumen, los pasos requeridos para resolver el problema de optimización de forma son los siguientes:

- Dado un conjunto inicial de parámetros $p_0 = (x_0, y_0, R_0)$:
 1. Definir los parámetros de diseño actuales como $p = p_0$.
 2. Generar la grilla y las funciones σ , k y q correspondientes a la configuración determinada por p .
 3. Resolver el problema de borde para conocer el estado u y resolver el problema adjunto para conocer el estado adjunto λ .
 4. Calcular el valor de la función costo ψ .
 5. Si el criterio de parada es alcanzado, dar como salida los parámetros actuales como los de óptimo diseño y parar, sino continuar.
 6. Calcular las derivadas $\psi(\omega; V_k)$, para $k = 1, 2, 3$.
 7. Calcular el paso ε , considerando el conjunto de restricciones, y actualizar p .
 8. Ir a 2.

4.4.3. Resultados.

En esta parte del trabajo presentamos algunas simulaciones para mostrar el funcionamiento del algoritmo a la hora de estimar la posición y radio del tumor. En todos los casos los valores de los parámetros fisiológicos fueron iguales a los considerados en la sección 2.5. Las dimensiones del dominio Ω en todos los ejemplos que se presentan son $0,09 \times 0,03[m]$. El subconjunto cerrado y convexo C considerado fue el poliedro $[0,02, 0,08] \times [0,01, 0,025] \times [0, 0,007][m]$.

En la Figura 4.2 mostramos el resultado de la metodología propuesta para el caso en que el centro del tumor fue situado en el punto $(0,06, 0,02)$ y su radio se consideró igual a $0,006[m]$. En la Figura 4.2a está graficada en color verde la frontera del subdominio objetivo ω , mientras que en azul está graficada la frontera correspondiente a un subdominio considerado como condición inicial. Este subdominio está centrado en el punto $(0,45, 0,01)$ y su radio es igual a $0,003$. A su derecha, en la Figura 4.2b se muestra el perfil de temperatura medido sobre la frontera superior del dominio y al que se le adicionó 10% de ruido uniformemente distribuido. Este perfil de temperatura se utilizó como dato de entrada. Finalmente en 4.2c podemos ver la diferencia entre el subdominio original y el subdominio que fue hallado por el algoritmo. Es claro que el subdominio hallado se encuentra muy próximo al correcto, más allá de haber contaminado el dato de entrada con 10% de ruido aleatorio.

Cabe aclarar que en este ejemplo, como en los siguientes, el criterio de parada que adoptamos fue una tolerancia para el valor de la función objetivo dada por $\psi < 10^{-7}$ o un número máximo

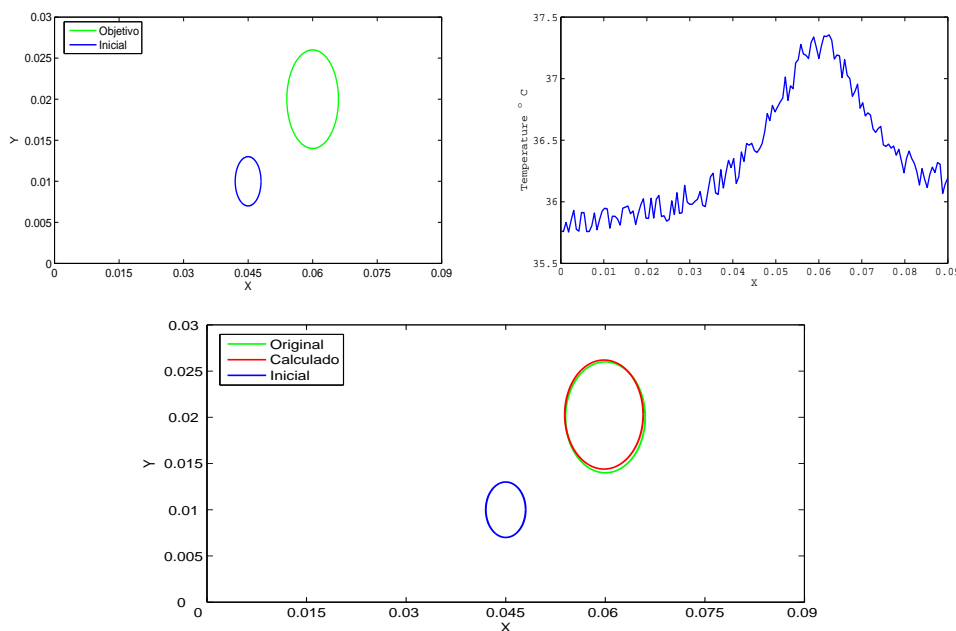


Figura 4.2: (a) Subdominio inicial (azul) y subdominio objetivo (verde). (b) Datos de entrada contaminados con 10% de ruido aleatorio. (c) Diferencia entre el subdominio objetivo (verde) y el subdominio hallado mediante la metodología 2 (rojo).

de 500 iteraciones. Cuando los datos de entrada no fueron contaminados con ruido aleatorio el algoritmo converge para casi todos los casos en a lo sumo 150 iteraciones.

En la Figura 4.3 se presenta otro ejemplo. En este caso, el tumor fue centrado en el punto $(0.029, 0.02)$ y su radio igual a $0.005[m]$. Nuevamente la condición inicial, graficada en azul en la Figura 4.3a, fue centrada en el punto $(0.45, 0.01)$ y su radio igual a $0.003[m]$. Cabe resaltar que hemos corrido el algoritmo considerando diferentes condiciones iniciales y en todos los casos los resultados obtenidos fueron similares. En este ejemplo también consideramos que los datos de entrada están contaminados con 10% de ruido aleatorio. En la Figura 4.3b podemos ver, graficado en rojo, la frontera del tumor hallado usando la metodología propuesta en este capítulo y en color negro el borde correspondiente al tumor hallado utilizando la metodología propuesta en el Capítulo 3 utilizando el algoritmo Pattern Search. De acuerdo a este ejemplo y a los que se muestran en el Cuadro 4.4.3, queda claro que utilizar la derivada de forma nos permite obtener los mismos o aún mejores resultados que los obtenidos utilizando la metodología propuesta en el Capítulo anterior, la cual hace uso del algoritmo Pattern Search. Además el hecho de calcular la derivada de la función costo y el uso de esta información implica un menor tiempo computacional a la hora de resolver estos problemas. En la Figura 4.3c las diferencias entre las dos metodologías se muestran en mayor detalle.

Por último, el Cuadro 4.4.3 muestra los resultados numéricos correspondientes a los ejemplos anteriores y a casos similares. En total 4 ejemplos son presentados. Para cada uno de estos ejemplos mostramos los resultados obtenidos en los casos en que los datos de entrada fueron considerados sin ruido aleatorio y para el caso en el que se les sumó 10% y 15% de ruido uniformemente distribuido. De estos ejemplos podemos ver que los errores en los valores estimados, incluso para los casos en que los datos de entrada fueron contaminados con ruido aleatorio, son siempre menores al 5%. Por lo tanto, los resultados obtenidos para la estimación de la posición y tamaño del tumor,

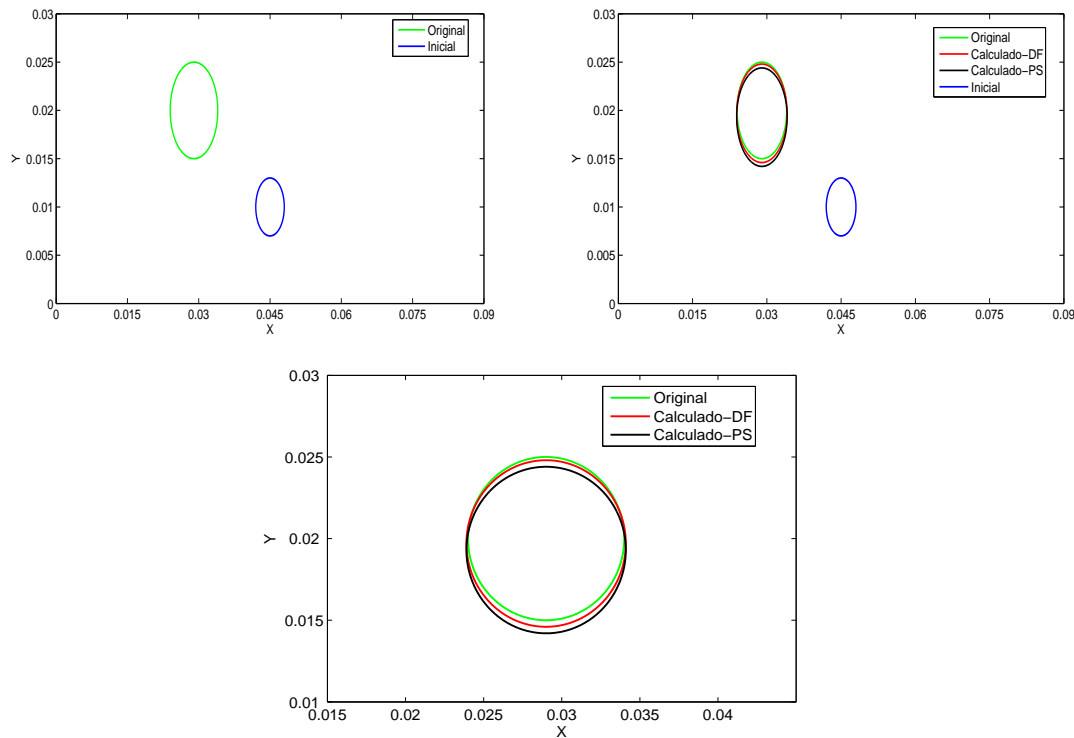


Figura 4.3: (a) Subdominio inicial (azul) subdominio objetivo (verde). (b) Diferencia entre el dominio objetivo y el subdominio ω hallado utilizando Pattern Search (negro) y mediante el cálculo de la derivada de forma (rojo). (c) Zoom del resultado graficado en (b).

utilizando esta metodología, mejoran los resultados obtenidos habiendo utilizado la metodología propuesta en el Capítulo anterior.

Datos originales			Sin ruido			10% de ruido			15% de ruido		
x	y	R	x	y	R	x	y	R	x	y	R
0.04	0.022	0.004	0.0400	0.0218	0.0040	0.0401	0.0219	0.0039	0.0000	0.0000	0.0000
0.029	0.02	0.005	0.0290	0.0199	0.0050	0.0290	0.0197	0.0051	0.0289	0.0203	0.0049
0.06	0.02	0.006	0.0599	0.0199	0.0060	0.0598	0.0203	0.0059	0.0600	0.0203	0.0059
0.07	0.016	0.007	0.0699	0.0158	0.0071	0.0701	0.0162	0.0069	0.0698	0.0159	0.0069

Cuadro 4.1: Estimación del centro y radio del tumor en 2D. Las diferentes columnas muestran los resultados obtenidos considerando los datos sin y con 10% y 15% de ruido aleatorio respectivamente.

4.5. Resolución del problema: PI 2

4.5.1. Cálculo de la derivada con respecto a la intensidad de calor metabólico q_0

En este problema asumimos conocida la posición y tamaño del subdominio ω como así también el valor de todos los parámetros fisiológicos, salvo el valor correspondiente a la fuente de calor metabólico dentro de la región tumoral, o sea el valor de q_0 . Recordemos entonces el problema inverso a resolver:

- Conocido el valor de todos los parámetros fisiológicos, las temperaturas T_b y T_a , y de un perfil de temperatura u_{exp} sobre Γ_s , el problema consiste en hallar q_0^* tal que la solución $u_{q_0^*}$ al problema de borde (2.3) verifique $u_{q_0^*} = u_{exp}$ en Γ_s .

El valor de q_0 óptimo se alcanza cuando la función costo:

$$\psi(q_0) = \Psi(u_{q_0}) = \frac{1}{2} \int_{\Gamma_s} (u_{q_0} - u_{exp})^2 d\Gamma, \quad (4.66)$$

se anula.

Similar al caso anterior, la función costo $\psi(q_0)$ representa el error o diferencia entre un perfil de temperatura u_{exp} medido experimentalmente en la superficie corporal y el perfil correspondiente a la solución numérica u_{q_0} del problema de transferencia de calor considerando el valor q_0 como la intensidad de la fuente de calor metabólico en la región tumoral. Es claro entonces que la variable de diseño en este problema es el coeficiente q_0 . Para calcular la derivada del funcional (4.66) utilizaremos los resultados presentados en la Sección 4.2. De acuerdo a la expresión (4.22) la derivada de la función costo con respecto a los parámetros está dada por

$$\dot{\Psi}(q_0; V) = \Psi'(u_0; V) + a'(u_0, \lambda; V) - l'(\lambda; V). \quad (4.67)$$

Notemos que en nuestro caso la función costo (4.66) no depende explícitamente del parámetro q_0 por lo tanto sólo debemos calcular

$$\dot{\Psi}(q_0; V) = a'(u_0, \lambda; V) - l'(\lambda; V),$$

donde la forma bilineal y lineal respectivamente, son las siguientes

$$\begin{aligned} a_{q_0}(u, v) &= \int_{\Omega} \sigma(x) \nabla u \cdot \nabla v \, dx + \int_{\Omega} k(x) uv \, dx + \alpha \int_{\Gamma_s} uv \, d\Gamma, \\ l_{q_0}(v) &= \int_{\Omega} q(x)v \, dx + \alpha(T_a - T_b) \int_{\Gamma_s} v \, d\Gamma. \end{aligned}$$

Ahora debido a que la forma bilineal no depende explícitamente del parámetro q_0 tenemos que

$$a'(u, \lambda; V) = 0,$$

y sólo debemos calcular

$$l'(v; V) = \lim_{t \rightarrow 0} \frac{l_{q_0+tV}(v) - l_{q_0}(v)}{t}$$

Luego si escogemos $V = 1$, tenemos que

$$\begin{aligned} l'(v; 1) &= \lim_{t \rightarrow 0} \frac{l_t(v) - l(v)}{t} \\ &= \lim_{t \rightarrow 0} \frac{\int_{\Omega} (q_t(x) - q(x))v \, dx}{t} \\ &= \lim_{t \rightarrow 0} \frac{\int_{\omega} (q_0 + tV - q_0)v \, dx}{t} \\ &= \lim_{t \rightarrow 0} \frac{\int_{\omega} tVv \, dx}{t} \\ &= \int_{\omega} Vv \, dx \\ &= \int_{\omega} v \, dx. \end{aligned}$$

Con lo cual tenemos que la derivada del funcional costo (4.66) con respecto al parámetro q_0 , en la dirección $V = 1$, es

$$\dot{\Psi}(q_0; V) = - \int_{\omega} \lambda \, dx, \quad (4.68)$$

donde λ es la solución de la siguiente ecuación adjunta

$$\begin{cases} \text{Hallar } \lambda \text{ tal que} \\ a(\lambda, v) = - \frac{\partial \Psi}{\partial u}(u_0; v), \quad \forall v \in \mathcal{V}(\Omega). \end{cases} \quad (4.69)$$

En este caso particular tenemos que

$$\begin{aligned} - \frac{\partial \Psi}{\partial u}(u_0; v) &= - \int_{\Gamma_s} (u_{q_0} - u_{exp}) v \, d\Gamma, \\ a(\lambda, v) &= \int_{\Omega} \sigma(x) \nabla \lambda \cdot \nabla v \, dx + \int_{\Omega} k(x) \lambda v \, dx + \alpha \int_{\Gamma_s} \lambda v \, d\Gamma, \end{aligned}$$

Por lo tanto, λ es solución de (4.69) si satisface la ecuación

$$\int_{\Omega} (\sigma(x) \nabla \lambda \cdot \nabla v + k(x) \lambda v) \, dx + \int_{\Gamma_s} (\alpha \lambda + (u_{q_0} - u_{exp})) v \, d\Gamma = 0, \quad \forall v \in \mathcal{V}(\Omega).$$

La formulación clásica de este problema es la siguiente:

$$\begin{cases} - \operatorname{div}(\sigma(x) \nabla \lambda(x)) + k(x) \lambda(x) = 0, & \text{en } \Omega, \\ -\sigma(x) \frac{\partial \lambda(x)}{\partial \eta} = \alpha \lambda + (u - u_{exp}), & \text{sobre } \Gamma_s, \\ -\sigma(x) \frac{\partial \lambda(x)}{\partial \eta} = 0, & \text{sobre } \Gamma_l, \\ \lambda(x) = 0, & \text{sobre } \Gamma_b. \end{cases} \quad (4.70)$$

4.5.2. Resolución Numérica

Luego de todas las consideraciones realizadas, estimar la intensidad de la fuente de calor metabólico dentro de la región tumoral puede ser planteado como el siguiente problema de optimización:

$$\begin{aligned} &\text{Minimizar} && \Psi(q_0) \\ &\text{sujeto a} && \\ & && q_0 \in \mathcal{C} \end{aligned}$$

donde \mathcal{C} es un intervalo cerrado en \mathbb{R} definido por

$$\mathcal{C} = \{q_0 \in \mathbb{R} : q_{min} \leq q_0 \leq q_{max}\}.$$

Nuevamente, notemos que la restricción impuesta por la ecuación de estado ha sido eliminada utilizando el problema adjunto.

Para resolver el problema de optimización, al igual que para el problema PI 1, hemos implementado el método de gradiente para problemas en conjuntos convexos siguiendo las ideas descritas en Birgin *et al.* [13].

En resumen, los pasos requeridos para resolver el problema de estimación del valor de la fuente de calor metabólico dentro de la región tumoral son los siguientes:

- Dado un valor inicial q_0 :

1. Definir el parámetro de diseño actual como $p = q_0$.
2. Generar la grilla y las funciones σ , k y q correspondientes a la configuración determinada por p .
3. Resolver el problema de borde para conocer el estado u y resolver el problema adjunto para conocer el estado adjunto λ .
4. Calcular el valor de la función costo ψ .
5. Si el criterio de parada es alcanzado, dar como salida el valor actual del parámetro como el de óptimo diseño y parar, sino continuar.
6. Calcular la derivada con respecto al parámetro p , en la dirección $V = 1$, $\Psi(q_0; 1)$.
7. Calcular el paso ϵ , considerando el conjunto de restricciones, y actualizar p .
8. Ir a 2.

4.5.3. Resultados

Los resultados obtenidos para el problema de estimación de la intensidad de la fuente de calor metabólico dentro de la región tumoral en el caso de 2 dimensiones se muestran en el Cuadro 4.2. Nuevamente, se puede apreciar una gran similitud entre los datos originales y los estimados por el algoritmo. El error de los valores estimados en todos los ejemplos es siempre menor al 2%.

Este problema es de nuestro particular interés ya que la estimación de los valores de los parámetros térmicos relacionados con el tumor, pueden ser de mucha utilidad e importantes para poder estudiar la evolución del tumor luego de que éste haya recibido algún tipo de tratamiento o terapia, como por ejemplo BNCT.

	Datos originales	Sin ruido	10% de ruido	15% de ruido
q_0	42000	41999.5	41931.73	41886.53
q_0	37000	36999.95	37230.41	37032.37
q_0	25000	25000.04	24959.47	25068.69

Cuadro 4.2: Estimación de la intensidad del calor metabólico dentro del tumor, caso 2D. En todos los ejemplos el tumor fue centrado en el punto (0.045, 0.02) y su radio igual a 0.005.

4.6. Conclusión del capítulo

En este Capítulo hemos presentado otra alternativa para resolver los problemas inversos que presentamos en el Capítulo anterior. La metodología propuesta hace uso de la información que provee la derivada de la función a minimizar con respecto a las distintas variables a estimar, motivo por el cual presentamos los conceptos básicos del análisis de sensibilidad: con respecto al cambio de los parámetros y con respecto al cambio de forma del dominio.

Para calcular las derivadas de las funciones costo hemos acudido en ambos casos al método adjunto. Este método implica la resolución de una problema de borde adicional, de manera tal que evitamos el cálculo explícito de \dot{u} . Vimos que este hecho es equivalente a emplear el método de Lagrange.

Aunque tradicionalmente el análisis de sensibilidad al cambio de forma ha estado muy ligado a la mecánica de las estructuras, de acuerdo a lo presentado en este Capítulo, estas herramientas pueden ser también aplicadas en problemas relacionados a la Medicina obteniendo excelente resultados.

El cálculo de la derivada del funcional costo y su posterior uso, permitió obtener una metodología más eficiente y la cual arrojó, de acuerdo a los ejemplos mostrados, mejores resultados que la propuesta en el capítulo anterior. Además, cabe destacar que la utilización de esta metodología reduce enormemente el tiempo computacional y en cuanto al costo computacional sólo es necesario resolver un problema de borde similar a la ecuación de estado original, por lo que hemos utilizado el mismo esquema de diferencias finitas.

4.7. Apéndice: Derivadas de Fréchet y Gâteaux.

Consideremos una aplicación $F : E \rightarrow W$, donde E y W son espacios de Banach. Una forma de estudiar el comportamiento de F en un entorno de un punto $x \in E$ es restringir su dominio a una dimensión, esto es analizar cómo se comporta F sobre cada recta que pasa por x . Con esta idea introduciremos la noción de derivada direccional que, como veremos luego, no es suficiente para realizar un estudio completo sobre el comportamiento local de F pero es de gran utilidad para establecer resultados parciales.

Definición: Sean E y W espacios de Banach y $F : E \rightarrow W$ una aplicación. Decimos que F tiene derivada de Gâteaux en el punto $x \in E$, en la dirección de $v \in E$, si el límite

$$\lim_{t \rightarrow 0} \frac{F(u + tv) - F(u)}{t}, \quad (4.71)$$

existe. En este caso, denotamos a este límite por $dF(u; v)$ y tenemos que

$$dF(x; v) = \left. \frac{d}{dt} F(u + tv) \right|_{t=0}. \quad (4.72)$$

Además si el límite existe para todo $v \in E$ decimos que F es Gâteaux diferenciable en $x \in E$.

Observación: Para cada punto $x \in E$, la derivada de Gâteaux define una función

$$dF(x; \cdot) : E \rightarrow W. \quad (4.73)$$

Esta función tiene la propiedad de ser homogénea, o sea para todo escalar α vale

$$dF(x; \alpha v) = \alpha dF(x; v). \quad (4.74)$$

Sin embargo, esta función no es necesariamente aditiva, con lo cual la derivada de Gâteaux puede resultar que no sea lineal.

A continuación daremos una definición de diferenciabilidad más fuerte que la anterior y para ello centraremos nuestra atención en las aplicaciones lineales y continuas entre espacios normados. Estas aplicaciones, además de respetar la estructura de espacio vectorial, poseen propiedades (consecuencia de su continuidad) que facilitan su estudio. De este modo, si podemos aproximar localmente a F en un punto x (tanto como se quiera) a una aplicación lineal y continua, es de esperar que en un entorno de x , F tenga un comportamiento análogo al de la aplicación en cuestión.

Definición: Sean E y W espacios de Banach. Una función $F : E \rightarrow W$ se dice Fréchet diferenciable en $x \in E$ si existe un operador lineal y acotado $DF_x : E \rightarrow W$ tal que

$$\lim_{v \rightarrow 0} \frac{\|F(x + v) - F(x) - DF_x(v)\|_W}{\|v\|_E} = 0. \quad (4.75)$$

Llamamos a la aplicación lineal y acotada $DF_x : E \rightarrow W$ la derivada de Fréchet de F en el punto $x \in E$.

Relación entre la derivada de Gâteaux y derivada de Fréchet: Sean $F : E \rightarrow W$ una función Fréchet diferenciable en $x \in E$ y v un elemento fijo de E . Sustituyendo v por tv en (4.75) tenemos

$$\begin{aligned} 0 &= \lim_{t \rightarrow 0} \frac{\|F(x+tv) - F(x) - DF_x(tv)\|_W}{\|tv\|_E} \\ &= \lim_{t \rightarrow 0} \frac{\|F(x+tv) - F(x) - tDF_x(v)\|_W}{t\|v\|_E} \\ &= \lim_{t \rightarrow 0} \left\| \frac{F(x+tv) - F(x)}{t} - DF_x(v) \right\|_W, \end{aligned} \quad (4.76)$$

por lo tanto F resulta Gâteaux diferenciable en ese punto y además tenemos que $DF_x(v) = dF(x; v)$. Sin embargo, no es válido que toda función Gâteaux diferenciable resulte ser Fréchet diferenciable. Ahora si F es Gâteaux diferenciable en un abierto $X \subset E$, podemos asegurar que F es también Fréchet diferenciable, si la derivada de Gâteaux es lineal y acotada en cada punto de X y además la derivada de Gâteaux resulta ser un mapa continuo de $X \rightarrow L(E, W)$.

A continuación listamos dos propiedades que serán muy útiles a lo largo del Capítulo 4.

Regla de la cadena: Sean E, W y U espacios de Banach y aplicaciones $F : E \rightarrow W$ y $G : W \rightarrow U$. Si F es Fréchet diferenciable en $x \in E$ y G es Fréchet diferenciable en $F(x) \in W$, entonces la aplicación $G \circ F : E \rightarrow U$ es Fréchet diferenciable en x y además vale

$$D(F \circ G)(x) = DG(F(x)) \circ DF(x). \quad (4.77)$$

Derivadas parciales: Sean E, U y W espacios de Banach y $F : O \rightarrow W$ una función, donde O es un conjunto abierto y no vacío de $E \times W$. Definimos la derivada parcial con respecto a la primer variable en el punto $(x_0, y_0) \in O$ como la derivada en x_0 de la aplicación $x \rightarrow F(x, y_0)$ definida en un entorno X de x_0 tal que $X \times Y$ resulta un entorno de (x_0, y_0) contenido en O . La derivada parcial es denotada por $DF_{(x_0, y_0)}^1$. La derivada parcial con respecto a la segunda variable y , la cual denotamos por $DF_{(x_0, y_0)}^2$, se define de manera análoga. Considerando estas definiciones tenemos los siguientes resultados:

(i) Si F es diferenciable Fréchet en el punto (x_0, y_0) entonces las derivadas parciales de F en (x_0, y_0) existen y además

$$DF_{(x_0, y_0)}(v, u) = DF_{(x_0, y_0)}^1 u + DF_{(x_0, y_0)}^2 v. \quad (4.78)$$

(ii) Si $(x, y) \rightarrow DF_{(x, y)}^1$ y $(x, y) \rightarrow DF_{(x, y)}^2$ son continuas en O entonces F es Fréchet diferenciable y vale (4.78).

Parte II

Clasificación y agrupamiento de datos

Capítulo 5

Clasificación y agrupamiento de datos mediante estimación de densidades

5.1. Motivación del problema

Tal como mencionamos en la introducción de este trabajo, el problema de hallar patrones dentro de un conjunto de datos ha sido de gran importancia a lo largo de la historia. En el ámbito de la Medicina, más precisamente en el área de la genética, dimos como ejemplo la importancia que existe en entender la relación que hay entre la variación en las secuencias de ADN de un ser humano y la susceptibilidad que éste puede tener a cierto tipo de enfermedad. Es decir, cómo los cambios en la secuencia de ADN de un individuo influyen en el riesgo de desarrollar enfermedades como por ejemplo el cáncer. Esto es de suma importancia para ayudar a mejorar el diagnóstico, prevención y tratamiento de las enfermedades.

En algunos problemas de clasificación de datos sólo un pequeño número de variables bastan para realizar la clasificación. Por ejemplo, para el diagnóstico clínico de ciertas enfermedades infecciosas, a un paciente que padece la infección se lo puede detectar por la mera presencia de un elevado número de glóbulos blancos en su sangre, mientras que pacientes que padecen una afección aguda en el hígado pueden ser identificados midiendo los niveles de amoníaco en la sangre. Si disponemos de un conjunto de datos de entrenamiento suficientemente grande, estas variables que usamos para realizar la clasificación pueden ser calibradas de manera tal que podamos clasificar en forma sencilla una nueva muestra cuando ésta se presente.

La manera más básica de realizar esto, sería elaborar una tabla con los rangos de valores que pueden tomar las variables o combinaciones de ellas en cada una de la clases y a partir de estos valores o de una función paramétrica de éstos (con parámetros ajustados considerando los datos de la población de entrenamiento), decidir el resultado cuando disponemos de una nueva observación. Otra manera más elaborada de realizar la clasificación es estimando la función densidad de probabilidad en el espacio de características asociada a cada clase. Luego, utilizando estas funciones podemos inferir la probabilidad de que la nueva muestra pertenezca a cada una de estas clases [14].

Sin embargo, la combinación de que sólo sean necesarios un número pequeño de variables para realizar la clasificación y disponer de una población de aprendizaje suficientemente grande, no es lo más usual. Esto es lo que sucede cuando se analizan fenómenos un tanto excepcionales como por ejemplo una enfermedad atípica o cuando realizar mediciones puede resultar muy difícil y costoso o cuando las causas fundamentales que determinan cierto fenómeno (y que posibilitan realizar la clasificación a partir de un número pequeño variables) no son bien entendidas, situación

que suele darse al estudiar sistemas complejos como el cuerpo humano, el clima de nuestro planeta o los mercados financieros. De todas maneras, la tecnología actual nos ofrece una posibilidad de compensar esta falta de entendimiento del problema ya que hoy es posible tener información de no sólo una decena de variables sino de una decena de miles de variables relacionadas con el problema y por ende con la clasificación buscada. Por ejemplo, en el área de la Medicina es posible tener el nivel de expresión asociado a miles de genes de un paciente, también hoy podemos obtener datos de la temperatura de la superficie del mar alrededor de todo el mundo, etc. Este hecho nos hace pensar que es entonces posible equilibrar, con la cantidad de datos a nuestra disposición, el desconocimiento que tenemos sobre las causas del fenómeno y que determinan la clasificación.

Sin embargo existe un problema. Consideremos, por ejemplo el procedimiento más simple de elaborar una tabla. Dado que el número de entradas necesarias para elaborar la tabla crece exponencialmente con la dimensión del espacio de las variables o características, el tamaño de la población de aprendizaje que necesitamos para confeccionar la tabla se convierte de pronto en algo impensado (necesitaríamos millones de los pacientes, de mediciones de temperatura mensuales promedio, etc). En la construcción de una fórmula paramétrica, esto se traduce en el problema de sobreajuste: siempre es posible encontrar un conjunto de parámetros que proporcionen un ajuste perfecto para los datos de la población de entrenamiento, sin embargo los resultados para las nuevas muestras pueden carecer de sentido (Pensemos en el caso sencillo de un ajuste polinomial. Dados n puntos podemos hallar un polinomio de grado menor o igual a n que pase exactamente por estos puntos, sin embargo si el orden del polinomio es grande éste oscilará mucho y la estimación en puntos cercanos a los del conjunto de aprendizaje distará seguramente bastante de la correcta).

De la misma manera, para el caso de estimación de densidades de probabilidades, uno necesita un número m de observaciones de entrenamiento que crece exponencialmente con la dimensión n del espacio donde estas se encuentran, esta es la llamada “maldición de la dimensión” [11]. A modo de ejemplo consideremos la situación cuando $m < n$. Debido a que todos los puntos de entrenamiento yacen en un hiperplano de dimensión m , éstos no nos proporcionan ningún tipo de información sobre cómo estimar las densidades de probabilidades fuera de este hiperplano. Luego, cuando dispongamos de una nueva muestra, esta muestra casi con probabilidad uno no estará contenida en el hiperplano. ¿Cómo podemos entonces decir algo sensato acerca de su probabilidad? El punto crucial es que al estar trabajando en un espacio de grandes dimensiones, seguramente la nuevas observaciones estarán “lejos” de las observaciones pertenecientes al conjunto de entrenamiento y por lo tanto el valor de las densidades estimadas no serán en esos puntos del todo correctas.

Este problema puede parecer en un principio muy difícil de superar, sin embargo veremos que existe la posibilidad de solucionarlo. Dado que no entendemos bien las causas fundamentales del fenómeno en estudio y por ende decidir cuales son las variables básicas para realizar la clasificación, hemos decidido recolectar más información (más variables). ¿Puede esta información adicional perjudicarnos? Imaginemos que un médico intenta diagnosticar la enfermedad de un paciente y que sólo a partir de los síntomas no logra identificar cuál es el tipo de afección que tiene este paciente. Luego, si el médico recibe análisis clínicos que le aportan más información sobre el estado del paciente, no parece lógico que los descarte por el sólo hecho de que sea más información que tiene que manejar y esto lo pueda confundir. Es evidente que si hay un problema, éste problema no es la disponibilidad de más información adicional, sino en nuestra manera de manejar esta información. En este capítulo proponemos sobrepasar este inconveniente a través de una metodología general para la clasificación y agrupamiento de datos.

El principio básico de esta metodología es sencillo. Dijimos que nos preocupa que los puntos de prueba pueden estar muy lejos de las observaciones pertenecientes a la clase de entrenamiento y por lo tanto en estos puntos las densidades están vagamente estimadas. Sin embargo, debemos

tener en cuenta que en realidad sí disponemos de información localizada en los puntos u observaciones de prueba, la información la aportan justamente los mismos puntos. No sabemos a cuál clase pertenece cada observación de prueba (este es justamente el problema de clasificación que nos interesa resolver), pero al menos sabemos que cada observación pertenece a una de las clases. La idea es utilizar esta información de manera de generar un procedimiento de clasificación que resulte robusto. Este enfoque de utilizar datos no etiquetados para el procedimiento de clasificación es clave en aprendizaje semisupervisado [18]. La metodología que aquí propondremos implica considerar las poblaciones de entrenamiento y de prueba de manera casi indistinta y este hecho permite poder unificar los procedimientos para clasificación y agrupamiento de datos, este último caso se da cuando directamente no disponemos de una clase de entrenamiento.

La metodología está basada en el espíritu del algoritmo Expectation-Maximization (EM) [23] y uno de sus pilares es la estimación de densidades, para lo cual usamos las ideas desarrolladas en [73]. Este procedimiento consiste en transformar, por medio de “flujos”, las muestras correspondientes a las p clases distintas a muestras correspondientes a una distribución Gaussiana isotrópica. En este contexto, las observaciones actúan como marcadores que guían los flujos marcando el descenso de la divergencia de Kullback-Leibler [50]. Todas las observaciones guían los p flujos, pero a medida que una observación es claramente asignada a una clase, esta observación comienza a cumplir un papel más activo en el flujo correspondiente a esa clase, y por lo tanto actúa de manera casi pasiva en el resto.

Muchas de las ideas aquí expuestas tienen puntos en común con temas en la literatura. Nuestra contribución no sólo proporciona novedosos ingredientes, sino también un nuevo enfoque de unificar los problemas de clasificación y agrupamiento de datos y una metodología para realizar esto. Un papel central es desempeñado por el algoritmo EM, pero aquí en el contexto de clasificación y agrupamiento. El procedimiento para la estimación de las densidades a partir de flujos de normalización fue desarrollado originalmente en [73], pero comparte algunas ideas con las desarrolladas en [20]. La selección de variables es un campo de gran interés y estudio (ver [39] para una revisión general). Basándonos en medir la capacidad que posee cada variable o un grupo de variables para particionar la muestras en clases distintas, proponemos un procedimiento destinado a la selección de un conjunto de variables. Estas ideas tienen relación con algunas provenientes de la teoría de la información [26]. Nuestra principal innovación es el uso de flujos suave y graduales, como una técnica de agrupamiento.

La estructura de este Capítulo 5 es la siguiente. Luego de esta introducción y motivación del problema, en la sección 2 presentamos la definición precisa de los problemas de clasificación y agrupamiento, describimos nuestro enfoque de unificar estos problemas y presentamos la ideas para realizar clasificación y agrupamiento en el espíritu del algoritmo EM. En la sección 3, describimos brevemente los conceptos para estimar densidades de probabilidades desarrollados en [73] y presentamos más detalladamente el algoritmo propuesto para realizar clasificación y agrupamiento. Luego, en la sección 4, proponemos el procedimiento para la elección de un conjunto de variables, el cual se desprende de las ideas presentadas en las secciones anteriores. En la sección 5, con el objetivo de ilustrar el desempeño de la metodología propuesta, presentamos los resultados obtenidos al aplicarla a dos conjuntos de datos disponibles en la literatura [36, 48], ambos relacionados con el diagnóstico de tumores a partir de niveles de expresión genética. Finalmente, en la sección 6, cerramos este capítulo con algunos comentarios y conclusiones.

5.2. Clasificación y Agrupamiento: una visión unificada

Los problemas de agrupamiento y clasificación de datos pueden definirse de la siguiente forma:

- Agrupamiento de datos: dada una matriz $Z \in \mathbb{R}^{n \times m}$, correspondiente a m observaciones z^j de n variables z_i , el objetivo es particionar las observaciones en p subconjuntos o grupos disjuntos (clusters), cada uno con “rasgos característicos”.
- Clasificación: Dada una matriz X , que contiene m observaciones x^j (de entrenamiento o aprendizaje), las cuales están particionadas en clases distintas $C_k, 1 \leq k \leq p$, y dada una nueva observación y^j (de prueba), el objetivo es asignar esta nueva observación a la clase con la cual posee más “características en común”. En algunos casos, además es posible disponer de antemano de algún tipo de información sobre la nueva observación que haga suponer que esta esté más próxima a una clase particular. Usualmente esta información a priori se denota por π_k^j y representa la probabilidad, antes de observar cualquier tipo de característica, de que la j -ésima muestra pertenezca a la k -ésima clase.

Este último problema, el de clasificación, puede ser generalizado y suavizado o debilitado, disminuyendo las pretensiones tanto en la respuesta buscada como así también en la información inicial requerida. Esto es, en lugar de asignar la nueva observación a una única clase, se podría simplemente querer conocer la probabilidad p_k^j que esta observación y^j pertenezca a la clase C_k y de igual forma para el conjunto de entrenamiento una opción sería la de sólo disponer de las probabilidades p_k^j que x^j esté en C_k . De manera similar, para el problema de agrupamiento, se pueden disminuir las pretensiones y dar como respuesta una partición en la cual cada observación z^j tenga una probabilidad p_k^j de pertenecer al cluster C_k .

De esta manera, es claro que ambos problemas pueden ahora ser planteados bajo una visión unificada: dada X , matriz que contiene m observaciones x^j de n variables x_i ; y para un subconjunto J_{train} de estas observaciones las probabilidades p_k^j que la observación x^j esté en la clase C_k , el objetivo es conocer las correspondientes probabilidades p_k^j para el resto de las observaciones x^j , con $j \in J_{test}$. De acuerdo a esta visión unificada la única diferencia entre los dos problemas, es que para agrupamiento, J_{train} es vacío.

Cabe mencionar que las versiones “originales” de cada problema se recuperan si las probabilidades p_k^j son cero o uno para clasificación y para el caso de clustering eligiendo como regla asignar la observación x^j a la clase C_k con máxima p_k^j .

5.2.1. Clasificación y Agrupamiento mediante estimación de densidades

En la sección anterior hablamos de clasificar y agrupar datos de acuerdo a “rasgos comunes”. La pregunta que surge entonces es: ¿cómo caracterizar los “rasgos comunes” que definen cada clase C_k ?

La manera más natural y general, es hacerlo a través de una función de densidad de probabilidad $\rho_k(x)$, que especifique cuán probable es encontrar un conjunto de observaciones en la clase C_k . Así, dadas cada una de las densidades $\rho_k(x)$, la probabilidad de la clase C_k dado que se observó x^j , puede ser calculada usando la formula de Bayes:

$$p_k^j = \frac{\pi_k^j \rho_k(x^j)}{\sum_q \pi_q^j \rho_q(x^j)} \quad (5.1)$$

Este valor nos dice la probabilidad de que la clase C_k contenga a la observación x^j .

Supongamos tener un algoritmo para estimar densidades de probabilidad de manera tal que dado un conjunto de m observaciones y^j de n variables, produce una estimación de la densidad $\rho(y)$. La manera clásica en que la estimación de densidades de probabilidad es utilizada en problemas

de clasificación de datos, consiste en estimar las densidades $\rho_k(x)$ a partir de los datos de entrenamiento y luego se aplica (5.1) a cada observación x_j (de prueba), para inferir la probabilidad que esta pertenezca a cada clase C_k .

El procedimiento antes mencionado no tiene en cuenta toda la información que tenemos a nuestra disposición. Esto genera un problema cuando y como es usualmente el caso, se tiene una población de entrenamiento relativamente pequeña, comparada a la cantidad de variables que tiene cada observación. (Por ejemplo, en el diagnóstico médico basado en el análisis de niveles de expresión genética mediante el uso de microarray, uno puede contar con el nivel de expresión correspondiente a miles de genes para cada uno de los individuos de una población, sin embargo puede pasar que esta población la compongan a lo sumo cientos de pacientes). Así, en los procedimientos clásicos, como por ejemplo regresión lineal, se puede caer en el problema de sobreajuste de datos: con tanta cantidad de variables a disposición uno puede realizar un ajuste perfecto en los datos de entrenamiento aunque seguramente obtendrá muy malos resultados en los datos de prueba. En el contexto de estimación de densidades este problema se conoce como “under-sampling”, es decir uno necesita estimar una densidad de probabilidad en un espacio de dimensión muy grande a partir de escasas o muy pocas observaciones. Al estar trabajando en un espacio de grandes dimensiones, seguramente las nuevas observaciones estarán “lejos” de las observaciones pertenecientes al conjunto de entrenamiento y por lo tanto les será asignada una probabilidad demasiado pequeña, aún en la clase a la cual realmente pertenece.

El interrogante que surge entonces es: ¿cómo solucionar este problema?. Dijimos que el problema surge debido a que las densidades de probabilidad estimadas en algunas o en todas las clases pueden estar escasamente resueltas en los puntos de prueba y esto es debido a la falta de puntos de entrenamiento cercanos a los puntos de prueba. Lo que debemos tener en cuenta es que en realidad sí disponemos de información localizada en los puntos u observaciones de prueba, la información la aportan justamente los mismos puntos. No sabemos a cuál clase pertenece cada observación de prueba (este es justamente el problema de clasificación que nos interesa resolver), pero al menos sabemos que cada observación pertenece a una de las clases. Por lo tanto, al menos unas de las ρ_k debe ser no nula en cada punto de prueba.

Consideremos un algoritmo que estime las densidades basado en la maximización de la función verosimilitud de los datos [14], esto es :

$$\rho = \arg \left(\max_{\rho} (L[\rho]) \right), \quad (5.2)$$

donde L es el logaritmo de la función verosimilitud,

$$L[\rho] = \sum_{j=1}^m \log(\rho(y^j)), \quad (5.3)$$

y la maximización es considerada sobre un conjunto permitido de distribuciones $\rho(y)$. El procedimiento clásico consiste en realizar esta maximización sobre cada una de las clases, considerando las observaciones de entrenamiento, para luego inferir las correspondientes probabilidades para la población de prueba mediante la formula de Bayes (5.1). Para cada clase, deseamos maximizar

$$L_k = \sum_{\substack{j \in C_k \\ \text{aprendizaje}}} \log(\rho_k(x^j)). \quad (5.4)$$

Notemos que la función verosimilitud definida de esta manera no tiene en cuenta las observaciones de prueba y en particular el hecho de que estas observaciones deben pertenecer a una de estas

clases. Como la densidad de probabilidad para una observación en el conjunto de prueba está dada por:

$$\rho(x^j) = \sum_k \pi_k^j \rho_k(x^j), \quad (5.5)$$

donde π_k^j es la probabilidad a priori de que la j -ésima observación pertenezca a la k -ésima clase, la función log-verosimilitud completa que involucra a todas las observaciones, entrenamiento y prueba, está dada por:

$$L = \sum_{\text{aprendizaje}} \sum_k p_k^j \log(\rho_k(x^j)) + \sum_{\text{prueba}} \log \left(\sum_k \pi_k^j \rho_k(x^j) \right), \quad (5.6)$$

donde p_k^j toma el valor igual a uno si la j -ésima observación pertenece a la clase k y cero en caso contrario.

Definida la función log-verosimilitud completa, consideremos la derivada del segundo sumando con respecto a $\rho_k^j = \rho_k(x^j)$:

$$\frac{\partial}{\partial \rho_k^j} \log \left(\sum_k \pi_k^j \rho_k^j \right) = \frac{\pi_k^j}{\sum_k \pi_k^j \rho_k^j} = \frac{p_k^j}{\rho_k^j}, \quad (5.7)$$

donde p_k^j es la probabilidad a posteriori calculada por (5.1). Observemos que esta derivada es igual a la derivada parcial correspondiente a función log-verosimilitud ponderada:

$$\sum_k p_k^j \log(\rho_k^j) \quad (5.8)$$

si las probabilidades a posteriori p_k^j son mantenidas fijas. Por lo tanto, la función log-verosimilitud completa (5.6) tiene las mismas derivadas parciales, con respecto a las densidades, que la suma:

$$L = \sum_k L_k, \quad \text{donde} \quad L_k = \sum_j p_k^j \log(\rho_k(x^j)), \quad (5.9)$$

donde la única diferencia entre las poblaciones de entrenamiento y de prueba es que las probabilidades a priori π_k^j correspondientes a las observaciones de entrenamiento están más sesgadas hacia una clase, posiblemente siendo deltas de Kronecker, como en la derivación que hicimos de esta fórmula.

Inicialmente las probabilidades p_k^j pueden ser consideradas igual a las probabilidades a priori π_k^j . Luego, maximizando las funciones L_k , manteniendo fija p_k^j , obtenemos como resultado un nuevo conjunto de densidades $\rho_k^0(x)$. Entonces, en el espíritu del algoritmo EM, podemos iterar este procedimiento con las probabilidades p_k^j ahora actualizadas por las a posteriores dadas por la fórmula (5.1), y luego actualizar $\rho_k^t(x)$ en $\rho_k^{t+1}(x)$, hasta que el procedimiento converja. Cabe observar que este procedimiento no sufre ninguna alteración si para la población de entrenamiento consideramos que las p_k^j no son deltas de Kronecker sino una probabilidad discreta más general. Finalmente, también hay que resaltar que este procedimiento es posible de implementarlo en el caso en que la población de entrenamiento sea inexistente, es decir cuando J_{train} es vacío, en este caso el algoritmo proporciona una metodología para realizar agrupamiento de datos.

Para aclarar y fijar ideas, a continuación enunciamos el algoritmo propuesto de manera más formal.

- Supongamos disponer de un procedimiento de para estimar densidades de manera paramétrica. Luego, dadas:
 - X matriz que contiene m observaciones x^j de n variables x_i .
 - un número p de clases C_k .
 - probabilidad a priori π_k^j que la observación j pertenezca a la clase C_k .
 - una familia paramétrica de distribuciones de probabilidades $\rho(x; \alpha)$.

podemos implementar un procedimiento iterativo que calcula en cada paso $t \geq 0$, una probabilidad estimada $P_k^j[t]$ que indica la probabilidad de que la observación j pertenezca a la clase C_k , y para los pasos $t > 0$ estima un conjunto de densidades de probabilidad $\rho_k^t(x)$ asociadas a cada clase C_k , de la siguiente manera:

- Para $t = 0$, definir $P_k^j[0] = \pi_k^j$.
- Para $t > 0$,
 1. calcular $\rho_k^t(x)$ maximizando sobre los parámetros α el valor esperado de la función log-verosimilitud,

$$L_k = \sum_j P_k^j[t-1] \log(\rho_k^t(x^j)), \quad \rho_k^t(x) \in \rho(x; \alpha), \quad (5.10)$$

para clase C_k ;

2. actualizar las probabilidades P_k^j usando la formula de Bayes

$$P_k^j[t] = \frac{\pi_k^j \rho_k^t(x^j)}{\sum_q \pi_q^j \rho_q^t(x^j)}, \quad (5.11)$$

hasta que cierto criterio de convergencia se alcanzado.

Observación: en el caso del problema de agrupamiento, o sea cuando las probabilidades a priori π_k^j no dependen de j , es necesario romper la simetría entre las clases para comenzar el algoritmo. Esto puede realizarse simplemente perturbando las probabilidades iniciales $P_k^j[0]$ de manera tal que sean diferentes. En la sección 5.5.3 se analiza esta situación.

5.3. Estimación de densidades

En el procedimiento mencionado en la sección anterior se busca hallar, para cada clase k , un ajuste apropiado dentro de una familia paramétrica de distribuciones $\rho_k(x; \alpha)$. Aquí proponemos utilizar una alternativa distinta, en la cual cada una de estas distribuciones está caracterizada por un mapa y_k y una distribución objetivo, con función densidad de probabilidad μ , y que es común para todas. De esta manera las densidades correspondientes a cada clase quedan definidas por:

$$\rho_k(x) = J_k(x) \mu(y_k(x)), \quad (5.12)$$

donde $J_k(x)$ es el jacobiano¹ del mapa $x \rightarrow y_k$. Ahora, si los mapas $y_k(x)$ son definidos en términos de un conjunto de parámetros α , la ecuación (5.13) parecería simplemente una manera de reescribir la familia de distribuciones paramétricas. Sin embargo, veremos que esta forma de escribirlo no sólo resulta natural, generando en términos geométricos, una visión “dual” del problema de clasificación y agrupamiento, sino que además proporciona un algoritmo efectivo a la hora de implementar su solución.

La dualidad resulta de considerar la ecuación (5.13) desde dos puntos de vistas distintos: dada una muestra x , buscamos hallar la densidad $\rho(x)$ que mejor la “representa” o en cambio podemos buscar el mapa $y(x)$ que mejor transforma a ésta en una muestra correspondiente a una densidad conocida $\mu(y)$. Esta última visión es la que desarrollaron Tabak y Vanden-Eijden en [73]. Ellos proponen realizar la estimación de densidades de probabilidad transformando las observaciones x a un nuevo conjunto de variables y las cuales tienen una función densidad de probabilidad $\mu(y)$ conocida. De esta manera la función densidad de probabilidad correspondiente a la muestra x está dada por la ecuación (5.13). En el contexto de clasificación y agrupamiento que aquí tratamos, cada mapa $y_k(x)$ adquiere un mayor grado de importancia ya que “absorbe” o “rechaza” cada observación, es decir la aleja del agrupamiento o la acerca a éste dependiendo de la clase a la que esta pertenezca. Si x pertenece a la clase C_{k_l} , entonces los mapas y_{k_s} , con $s \neq l$ la alejarán del agrupamiento geométrico correspondiente a la clase C_{k_s} , mientras que el mapa y_{k_l} la incluirá en este agrupamiento.

5.3.1. Un flujo normalizador

A continuación presentamos brevemente algunas de las principales ideas para estimar una densidad de probabilidad a partir de una muestra observada. Esta metodología, que se encuentra desarrollada de manera detallada en [73], será utilizada en conjunto con el algoritmo propuesto en la sección anterior para resolver los problemas de clasificación y agrupamiento de datos.

Para simplificar notación y tratar de ser más claros descartaremos el subíndice k . Es decir, supongamos estar simplemente trabajando con datos pertenecientes a una única clase y de la que nos interesa estimar su función densidad de probabilidad². Tal como mencionamos anteriormente la idea es realizar la estimación de la función densidad de la muestra x , transformando estas variables en un nuevo conjunto de variables y , cuya función densidad de probabilidad μ es conocida. De esta manera la densidad de probabilidad de la variables originales, $\rho(x)$, queda definida por la ecuación

$$\rho(x) = J_y(x)\mu(y(x)), \quad (5.13)$$

donde $J_y(x)$ es el jacobiano del mapa $y(x)$.

Consideremos que el mapa está construido a partir de una composición infinita de transformaciones infinitesimales. O sea, introducimos un flujo $z(x;t)$, tal que

$$z(x;0) = x, \quad z(x;\infty) = y(x).$$

Otra manera de verlo, sería pensar que el mapa $y(x)$ es el punto terminal del flujo z . Para cada tiempo t , tenemos una estimación actual de la densidad de probabilidad dada por la ecuación

$$\rho^t(x) = J_{z_t}(x)\mu(z_t(x)), \quad (5.14)$$

donde hemos utilizado la notación más compacta $z_t(x) = z(x;t)$.

¹Para simplificar notación entendemos por Jacobiano, el módulo del determinante de la diferencial de la transformación, esto es $J_k(x) = |\det(Dy_k(x))|$

²Como consideramos una única clase las probabilidades p_k^j son iguales a 1 para todo j

Haciendo uso del método de máxima verosimilitud [14] se puede hallar los parámetros más adecuados para caracterizar esta densidad estimada. De acuerdo a este método mientras más grande sea el valor de la función log-verosimilitud (5.15), mejor será la estimación de $\rho(x)$.

$$L = \sum_j \log(\rho^t(x^j)). \quad (5.15)$$

Ahora, como la densidad queda definida por el flujo $z(x;t)$, podemos entonces pasar de un esquema de tiempos discretos a tiempos continuos t y hacer evolucionar este flujo de manera tal que su campo de velocidad $u = \frac{\partial}{\partial t} z(x;t)$, siga la dirección de ascenso de la función log-verosimilitud L , esto es:

$$u \propto \frac{\delta L}{\delta z_t}.$$

En el caso de estar ante una cantidad infinita de observaciones, la función log-verosimilitud (5.15) adopta la forma

$$L = \int \log(\rho^t(x;t)) \rho(x) dx, \quad (5.16)$$

donde $\rho(x)$ es la función de densidad de probabilidad que deseamos hallar. Utilizando (5.14), esta última ecuación puede escribirse como

$$L(z_t) = \int (\log(J_{z_t}(x)) + \log(\mu(z_t(x)))) \rho(x) dx, \quad (5.17)$$

Luego, para calcular la primera variación de L con respecto al flujo z_t observemos que (5.17) tiene la forma

$$L(z_t) = \int G(z_t, z'_t, x) dx, \quad (5.18)$$

por lo tanto la primera variación de L con respecto al flujo z_t puede ser calculada de manera directa, obteniendo

$$\frac{\delta L}{\delta z_t} = J_{z_t}(x) \left(\frac{\nabla_z \mu(z)}{\mu(z)} \rho_t(z) - \nabla_z \rho_t(z) \right), \quad (5.19)$$

donde $z = z_t(x)$ y

$$\rho_t(z) = \frac{\rho(x)}{J_{z_t}(x)}, \quad (5.20)$$

es la función densidad de probabilidad de la variable $z = z_t(x)$ dado que x está distribuida con función densidad de probabilidad $\rho(x)$ (no confundir esta función con $\rho^t(x)$ (5.14)).

Entonces, con el objetivo de incrementar el valor de la función log-verosimilitud, hacemos evolucionar el flujo $z(x;t)$ de manera tal que

$$\frac{\partial}{\partial t} z(x;t) = u(z(x;t)), \quad (5.21)$$

donde

$$u(z) = \left(\frac{\nabla_z \mu(z)}{\mu(z)} \rho_t(z) - \nabla_z \rho_t(z) \right). \quad (5.22)$$

Notemos que la velocidad $u(z)$ es simplemente el gradiente de la función log-verosimilitud dividido por el jacobiano (positivo) $J_{z_t}(x)$. Esta elección garantiza que el flujo siga una dirección de ascenso de la función log-verosimilitud y de esta manera nos aseguramos que el valor de esta función se incrementa.

Observación: en la expresión (5.19) es el Jacobiano $J_{z_t}(x)$ el único término que mantiene un registro de la muestra x donde el flujo dio inicio. Es en este punto donde la dualidad mencionada anteriormente cobra relevancia: si en lugar de estar realizando un proceso de estimación de densidad, i.e. hallar $\rho(x)$, simplemente pensamos que estamos realizando sobre las variables iniciales x un proceso de normalización, a partir de un mapa $z(x;t)$, que converge a las variables $y(x)$ con densidad μ , no hay necesidad entonces que en cada paso de tiempo recordemos con que x iniciamos el proceso. Simplemente los valores actuales de $z(x;t)$ es toda la información necesaria para continuar transformando estas variables en una muestra de μ . Así, adoptando este punto de vista podemos eliminar el Jacobiano $J_{z_t}(x)$ convirtiendo la dinámica en un proceso sin memoria y la ecuación (5.22) corresponde a la evolución en la que al tiempo t uno utiliza la muestra $z = z(x;t)$ en lugar de las original x .

Observación: de acuerdo a lo desarrollado en [73] la elección más natural de la distribución objetivo μ es considerar una gaussiana isotrópica, es decir

$$\mu(y) = \frac{1}{(2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}|y|^2}. \quad (5.23)$$

Para el caso en que contamos con un muestra finita de m observaciones x^j , la metodología presentada anteriormente se extiende de manera natural al escenario discreto. El único requerimiento es definir, para cada tiempo, una clase de transformaciones (de dimensión finita) de manera tal que sea posible calcular el gradiente de la función log-verosimilitud L con respecto a los correspondientes parámetros, lo cual representa la versión discreta de la variación de L respecto a $z(x;t)$ en el caso continuo.

En [73] los autores destacan que la familia transformaciones paramétricas $F(x;\alpha)$ debe satisfacer los siguientes requerimientos: depender de un número pequeño de parámetros α ; incluir la transformación identidad $F(x;0) = x$ y deben ser transformaciones básicas en el sentido que cualquier transformación general pueda ser generada a partir de la composición de varias F 's. Además muestran que es suficiente restringirse a considerar mapas unidimensionales. Cada fila de la matriz $X = x_i^j$ es una muestra de la densidad marginal con respecto a todos las demás variables. Luego con el objetivo de obtener densidades marginales en otras direcciones, es suficiente rotar la matriz X multiplicando por una matriz ortogonal U . Así en cada iteración, dada la matriz $Z = z_i^j$ correspondiente a la posición actual de las partículas, la rotamos multiplicando la por una matriz ortogonal elegida de manera aleatoria, obteniendo

$$Z \rightarrow UZ. \quad (5.24)$$

Este paso no conviene considerarlo como un movimiento de las partículas, si no que resulta más natural verlo como un cambio de coordenadas. Notemos que las transformaciones ortogonales tiene Jacobianos igual a 1, por lo tanto no afecta la densidad estimada ρ_t .

Luego de la rotación, se le aplica a cada variable z_i una transformación $z_i \rightarrow F^i(z_i)$ que la torna cada vez más Gaussiana. Tal como dijimos antes, la transformación F^i es escogida de una familia paramétrica $F(x;\alpha)$ teniendo en cuenta el ascenso de la log-verosimilitud, es decir

$$\alpha = \gamma \nabla_{\alpha} L_i |_{\alpha=0}, \quad (5.25)$$

donde

$$\gamma = \frac{\epsilon}{\sqrt{\epsilon^2 + \|\nabla_{\alpha} L_i\|^2}}, \quad (5.26)$$

$$L_i(\alpha) = \sum_{j=1}^m P^j \left[\log \left| F_z \left(z_i^j; \alpha \right) \right| + \log \mu \left(F \left(z_i^j; \alpha \right) \right) \right],$$

y $\mu(z)$ es la distribución normal unidimensional. La familia de transformaciones que utilizaremos aquí, y la cual fue introducida en [73], es la siguiente

$$F(x; \alpha) = (1 - \sigma)x + \varphi_0 + \gamma\sqrt{\eta^2 + [(1 - \sigma)x - x_0]^2}, \quad (5.27)$$

donde $\alpha = (\gamma, \sigma, \varphi_0)$. El parámetro σ está relacionado con el estiramiento, φ_0 con el desplazamiento y el parámetro γ con el cambio de pendiente en el punto x_0 , donde se produce un cambio de $DF/dx \approx 1 - \sigma - \gamma$ a $DF/dx \approx 1 - \sigma + \gamma$. El valor x_0 , donde ocurre el cambio de pendiente, es elegido de manera aleatoria de una distribución normal y por último el parámetro η , que está definido en términos de x_0 , cumple la función de suavizar las pendientes del mapa a la izquierda y derecha del punto x_0 .

Notemos que el flujo $z(x; t)$ y su Jacobiano necesitan ser calculados sólo en las observaciones x^j . Si estuviéramos ante el caso de contar con más puntos \tilde{x} en los cuales deseamos conocer el valor de ρ , estos puntos pueden ser transportados de manera “pasiva” por el algoritmo, sin afectar el valor de la log-verosimilitud y así calcular al finalizar el algoritmo $\rho(\tilde{x})$. En el contexto clásico de clasificación, los x^j representan las observaciones en cada una de las clases mientras que los \tilde{x} forman parte de la población de prueba. En la visión unificada de clasificación y agrupamiento que aquí proponemos, todas las observaciones constituyen puntos “activos” y la contribución de cada observación a la log-verosimilitud está ponderada por la probabilidad P^j , la cual se irá actualizando de acuerdo al algoritmo propuesto en la sección 5.2.

5.3.2. Algoritmo

El algoritmo que utilizaremos para resolver los problemas de clasificación y agrupamiento de datos es una mezcla de las ideas presentadas anteriormente para estimar densidades y las ideas presentadas en la sección 2 y el algoritmo propuesto

A continuación presentamos un esquema de la metodología propuesta. Empezando con una matriz X correspondiente a m observaciones x^j de n variables x_i y las probabilidades a priori π_k^j que la observación j pertenezca a la clase C_k , $k \in 1, \dots, p$. Consideramos p flujos $z_k(x; t)$ que evolucionan de acuerdo al siguiente esquema:

- **Precondicionamiento:** En este primer paso, para cada una de las clases C_k , consideramos todas las observaciones x^j las cuales están asociadas a esta clase de acuerdo a su probabilidad a priori π_k^j y calculamos la media ponderada

$$\bar{x}_k = \frac{\sum_j \pi_k^j x^j}{\sum_j \pi_k^j},$$

y la desviación estándar ponderada

$$\sigma_k = \sqrt{\frac{\sum_j \pi_k^j \|x^j - \bar{x}_k\|^2}{n \sum_j \pi_k^j}}.$$

Así, tenemos que para cada flujo las partículas se encuentran inicialmente centradas y normalizadas, es decir:

$$z_k^j = z_k(x^j; 0^+) = \frac{x^j - \bar{x}_k}{\sigma_k}, \quad (5.28)$$

y los correspondientes Jacobianos están dados por $J_k^j = \sigma_k^{-n}$.

Calcular las probabilidades a posteriori,

$$P_k^j = \frac{\pi_k^j \rho_k^j}{\sum_q \pi_q^j \rho_q^j}, \quad \text{donde } \rho_k^j = J_k^j \mu(z_k^j) \quad \text{y } q = 1, \dots, p. \quad (5.29)$$

■ **Flujo:** Para todo paso $t > 0$ (tiempo discreto), hacer lo siguiente:

1. Realizar, en cada clase k , un paso de normalización.

Para cada variable i

a) Aplicar una transformación ortogonal

$$z_k^j \rightarrow U_k z_k^j.$$

b) Aplicar una transformación

$$\begin{aligned} z_k^j(i) &\rightarrow F_k^i(z_k^j(i)), \\ J_k^j &\rightarrow \frac{d}{dz} F_k^i(z_k^j(i)) J_k^j. \end{aligned}$$

La transformación F_k^i es escogida considerando la dirección de ascenso de la función log-verosimilitud, esto es:

$$\alpha = \gamma \nabla_{\alpha} L_i |_{\alpha=0}, \quad (5.30)$$

donde

$$\gamma = \frac{\varepsilon}{\sqrt{\varepsilon^2 + \|\nabla_{\alpha} L_i\|^2}}, \quad (5.31)$$

$$L_i(\alpha) = \sum_{j=1}^m P_k^j \left[\log \left| F_z \left(z_k^j(i); \alpha \right) \right| + \log \mu \left(F \left(z_k^j(i); \alpha \right) \right) \right],$$

y $\mu(z)$ es la distribución normal unidimensional.

2. Actualizar las probabilidades a posteriori de acuerdo a la formula,

$$P_k^j = \frac{\pi_k^j \rho_k^j}{\sum_{q=1}^p \pi_q^j \rho_q^j}, \quad (5.32)$$

donde

$$\rho_k^j = J_k^j \mu(z_k^j). \quad (5.33)$$

Observación: La elección del valor ε en (5.31) es de considerable interés en la mayoría de los algoritmos de descenso. En todos los ejemplos que aquí presentaremos hemos considerado el valor constante $\varepsilon = 0,02$.

5.4. Elección de variables y evaluación del agrupamiento

En la sección previa presentamos una metodología que cuando es aplicada al problema de agrupamiento de datos nos provee de asignaciones p_k^j , las cuales representan la probabilidad que la observación j pertenezca a la clase k . En esta sección nos interesa analizar un problema que podríamos llamarlo recíproco; queremos evaluar que tan bien un conjunto de variables de las observaciones x confirman un agrupamiento dado de antemano. Es decir, supongamos que la población está distribuida en p clases distintas de acuerdo a las asignaciones (conocidas) q_k^j , buscamos conocer que tan bueno es un subconjunto de variables x_i para particionar la población en las distintas clases dadas por las probabilidades q_k^j . (Notemos que q_k^j no tiene porque ser igual a 1 o 0).

Existen varias situaciones en las que la evaluación del agrupamiento es de interés. La motivación en nuestro caso surge del problema de selección de variables: cuando el número de variables excede ampliamente al número de observaciones (esto es $m \ll n$), es necesario utilizar sólo un subconjunto de estas variables. Es natural entonces quedarnos con las que “mejor” agrupan los datos. Por ejemplo, para resolver el problema clásico de clasificación uno busca las variables que optimizan la partición de acuerdo a la clasificación dada para los datos perteneciente a la población de entrenamiento y para los cuales sabemos exactamente a que clase pertenece cada uno.

La pregunta que surge entonces es: ¿cómo evaluar que tan bien las variables x_i confirman un agrupamiento dado q_k^j ? Una manera de responder a este interrogante es considerando el valor M que nos proporciona la siguiente ecuación:

$$M = \sum_{k,j} q_k^j \log(p_k^j). \quad (5.34)$$

Esta ecuación puede interpretarse como menos la entropía-cruzada (el signo menos surge debido a que estamos buscando discriminar las observaciones en clases, i.e. orden, no desorden). Las probabilidades p_k^j provienen de las variables observadas x_i asumiendo el agrupamiento q_k^j dado y son calculadas a partir de la formula de Bayes (5.1), donde las densidades $\rho_k(x)$ maximizan las funciones verosimilitud

$$L_k = \sum_j q_k^j \log(\rho_k(x_i^j)). \quad (5.35)$$

(esta maximización puede realizarse considerando una familia paramétrica de distribuciones o por ejemplo mediante la metodología de los flujos propuesta anteriormente)

En resumen, debemos estimar las funciones densidades de probabilidad para cada clase, donde las clases están determinadas por las atribuciones q_k^j y las densidades son estimadas a partir del subconjunto de variables x_i consideradas. Luego, utilizamos estas densidades para calcular las asignaciones a posteriori p_k^j y finalmente, con estas probabilidades a priori, calculamos la entropía M . El valor que nos proporciona esta entropía (negativa) es la medida de que tan bien las variables x_i confirman el agrupamiento determinado por las atribuciones q_k^j . Para entender esto notemos que la medida (5.34) concuerda, salvo signo y sumada una constante, con la divergencia de Kullback–Leibler (KL) [50] de q y p y la cual está definida por

$$D_{KL}(q, p) = \sum_{k,j} q_k^j \log\left(\frac{q_k^j}{p_k^j}\right). \quad (5.36)$$

Esta función alcanza su máximo valor posible justamente cuando $p_k^j = q_k^j$, i.e. cuando las variables x_i recuperan o confirman las atribuciones dadas de antemano. En este caso tenemos que

$D_{KL}(p, p) = 0$ y M se convierte en el negativo de la entropía

$$M = \sum_{k,j} p_k^j \log(p_k^j). \quad (5.37)$$

Por lo tanto, mientras menor sea el valor de M , lo cual se da cuando mas próximos sean los valores de p_k^j a los de q_k^j , más útiles serán las variables x_i para utilizarlas en el proceso de discriminar las observaciones en las clases preestablecidas de acuerdo a q_k^j .

Para clarificar los conceptos anteriores, consideremos a continuación ejemplos de selección de un subconjunto de variables $i \in I$ de un conjunto mayor I^m .

- *Clasificación (versión 1)*: en la versión “clásica” del problema de clasificación, tenemos una población de entrenamiento de la que sabemos de antemano a que clase pertenece cada una de las observaciones que la componen, o sea tenemos $q_k^j = \delta_{k_j}^k$ (k_j^k es la clase a la que fue asignada la observación j). Seleccionar las variables x_i que maximizan M , corresponde en este caso a seleccionar las que hacen que $M = 0$ y esto sucede en el caso en que $p_k^j = \delta_{k_j}^k$, o sea cuando las variables x_i recuperan la partición en clases original.
- *Agrupamiento*: Para este caso uno no conoce las asignaciones q_k^j de antemano. Por lo tanto podemos suponer que $q_k^j = p_k^j$, i.e., realizar un procedimiento de agrupamiento a partir de las variables x_i y evaluar su performance de acuerdo al valor que nos de la entropía negativa. Ahora, como $M \leq 0$, su máximo valor se dará cuando las probabilidades p_k^j sean ceros o unos, lo cual corresponde a un agrupamiento de los datos con total certeza.
- *Clasificación (versión 2)*: En el caso en que la población de entrenamiento es pequeña y la de prueba resulta grande, uno puede utilizar la propuesta realizada para el caso de agrupamiento en lugar de la sugerida para clasificación. De hecho, siguiendo el espíritu de la visión unificadora de de los problemas de clasificación y agrupamiento de datos que aquí proponemos, es esto lo que debemos hacer. En efecto, si sólo utilizamos las observaciones pertenecientes a la población de entrenamiento perdemos toda la información disponible que tenemos en los puntos de prueba. Por lo tanto, lo que debemos hacer es utilizar ambas poblaciones, considerando los valores de q_k^j (conocidos) para las observaciones de prueba y las probabilidades a posteriori $q_k^j = p_k^j$ para los datos de prueba. Es decir realizar un mezcla de las dos propuestas anteriores.

Cabe señalar que la selección de un subconjunto $I \subset I^m$ de variables requiere realizar una búsqueda combinatoria que puede ser extremadamente costosa en el caso en que los cardinales de I e I^m sean grandes. Más aún, cada estimación de las densidades de probabilidad puede requerir también un trabajo significativo desde el punto de vista computacional. Para reducir estos costos podemos considerar las siguientes estrategias:

- No considerar todos los subconjuntos I posibles, si no sólo algunos casos elegidos bajo cierto criterio. El caso mas simple, seria evaluar el desempeño de cada variable por separado, ordenarlas de acuerdo a el valor de M que estas arrojen y luego seleccionar las primeras de la lista.
- No realizar una estimación para las ρ_k muy compleja sino considerar alguna sencilla, como por ejemplo una estimación paramétrica a una Gaussiana isotrópica.

Es claro que estrategias más sofisticadas pueden ser tenidas en cuenta. Todo depende de la importancia que tiene la reducción de variables y de los recursos disponibles. Si por ejemplo, reducir

el número de variables sólo se debe para que el problema resulte más manejable, entonces la estrategia más simple puede ser utilizado. Si en cambio el objetivo es identificar variables claves, tales como conjuntos de genes relacionados con una enfermedad particular, es indicado considerar una búsqueda más exhaustiva. En los ejemplos que presentaremos en la próxima sección correspondientes a clasificación de tumores a partir de datos de microarray, hemos consideramos la propuesta más simple de evaluar cada variable de manera individual y para la estimación de las densidades consideramos la metodología, obteniendo muy buenos resultados.

5.5. Ejemplos clínicos: clasificación de tumores

Con el objetivo de ilustrar el desempeño de la metodología propuesta, presentamos los resultados que hemos obtenidos al aplicarla a dos conjuntos de datos, ambos relacionados con el diagnóstico de tumores a partir de niveles de expresión genética y disponibles en la literatura. El primer conjunto de datos [48] está compuesto por 83 pacientes, cada uno de los cuales padece uno de las siguientes clases de tumor de células pequeñas, redondas y azules en la infancia³: neuroblastoma (NB), rhabdomyosarcoma (RMS), linfoma maligno no hodgkiniano (NHL) o sarcoma de Ewing extra óseo (EWS). Para cada uno de los 83 pacientes se dispone del nivel de expresión genética correspondiente a 2308 genes. El segundo conjunto de datos [36] contiene el nivel de expresión genética de 6817 genes correspondiente a células de la médula ósea. Estos datos corresponden a 72 pacientes los cuales padecen alguna de las siguientes clases de leucemia aguda: Leucemia linfoblástica aguda⁴ (ALL) o Leucemia mieloide aguda⁵ (AML).

Estos dos conjuntos de datos son cualitativamente diferentes (esto quedará reflejado en los gráficos que se presentarán a continuación). Consideramos que estas diferencias pueden deberse principalmente al hecho de que en realidad los 2308 genes en [48] son un subconjunto pertenecientes a un conjunto total de 6567, el cual ha sido filtrado de manera tal que los genes con los que trabajaremos cuenta al menos con un cierto grado de variación en su nivel de expresión. Por el contrario, los 6817 genes en [36] no han sido filtrados, de manera tal que para varios de los genes se tiene un nivel de expresión casi uniforme en muchas de las muestras. Más allá de estas diferencias veremos que la metodología propuesta funciona bien se haya o no filtrado los datos previamente.

5.5.1. Diagnóstico de una muestra por vez

Para comenzar nos enfocaremos en el problema de clasificación. Como primer ejemplo, mostraremos el caso en que hemos escogido un única observación, la cual queremos clasificar y utilizamos las restantes como población de entrenamiento. Este proceso lo hemos realizado con cada una de las observaciones y requiere de los siguientes pasos:

³Los tumores de células pequeñas redondas y azules en la infancia constituyen frecuentemente un desafío en el diagnóstico para el patólogo debido a que pueden presentarse como proliferaciones de células indiferenciadas en la microscopía de luz y muestran un alto grado de superposición histológica en su apariencia, careciendo de rasgos específicos para su diferenciación. Su subclasificación requiere muchas veces un diagnóstico multimodal que incluye el uso de técnicas como microscopía electrónica, inmunohistoquímica y técnicas moleculares y genéticas.

⁴Comprende un grupo de neoplasias malignas que afectan los precursores de los linfocitos en la médula ósea. Constituye una expansión clonal en una etapa de la hematopoyesis linfoide, expresada por una detención en la diferenciación celular, con proliferación y crecimiento no controlados de células leucémicas, que se originan en la médula ósea y luego se diseminan a sangre periférica, bazo, ganglios y al resto de los tejidos.

⁵Tipo de cáncer producido en las células de la línea mieloide de los leucocitos, caracterizado por la rápida proliferación de células anormales que se acumulan en la médula ósea e interfieren en la producción de glóbulos rojos normales.

- Asignar a la observación j una probabilidad a priori π_k^j de pertenecer a la k -ésima clase. Suponemos una probabilidad uniforme $\pi_k^j = \frac{1}{4}$ para el caso de los cuatro tumores de células redondas y pequeñas y una probabilidad a priori $\pi_k^j = \frac{1}{2}$ para el caso de los dos tipos de leucemia aguda.
- Selección de un conjunto de genes. Ordenamos los genes de acuerdo al valor provisto por la medida (5.37), donde las P_k^j son las probabilidades a posteriori calculadas por el algoritmo de agrupamiento usando un único gen por vez, las Q_k^j son unos o ceros para las observaciones de la clase de entrenamiento y $Q_k^j = P_k^j$ para el dato de prueba. Luego elegimos los primeros n genes de la lista.
- Clasificación de la observaciones de prueba (una única en este caso). Usando los genes seleccionados en el paso anterior, calculamos las probabilidades a posteriori P_k^j para la observación de prueba y asignamos esta observación a la clase cuya probabilidad a posteriori es mayor.

Cuando el número n de genes seleccionados es suficientemente grande, basta considerar sólo el paso de acondicionamiento del algoritmo para obtener resultados casi perfectos. Por ejemplo, con $n \geq 20$ los 83 casos correspondientes a tumores de células redondas y pequeñas son clasificados correctamente. En el caso de tumores de Leucemia 70 de los 72 pacientes son diagnosticados correctamente considerando $n \geq 10$. La parte no lineal del algoritmo nos permite reducir aún más el número de genes necesarios para obtener un diagnóstico correcto.

A modo de ejemplo, mostramos los resultados utilizando sólo dos genes para diagnosticar un paciente particular que padece linfoma maligno no hodgkiniano (NHL). La Figura 5.1 muestra los datos originales en el “espacio de genes”. Los dos genes corresponden a los seleccionados por el algoritmo de acuerdo a lo explicado anteriormente. Las diferentes clases de tumores se muestran en diferentes colores mientras que el paciente a ser diagnosticado en color verde para destacar que en un principio no conocemos a que clase éste pertenece (aunque es sabido que pertenece a la clase de color cyan). Vale la pena destacar que en este gráfico puede observarse la capacidad que tienen los dos genes seleccionados por el algoritmo para particionar los datos en las cuatro clases de tumores generando una partición bien clara.

La Figuras 5.2 y 5.3 describen, de manera gráfica, la evolución del proceso de diagnóstico correspondiente a este paciente particular con el paso de las iteraciones. En la Figura 5.2, podemos ver las variables transformadas $z = \phi_t(x)$, para las cuatro clases de tumores, incluyendo además en cada una, el paciente aún no diagnosticado. La fila superior contiene los mismo datos que la Figura 5.1, pero separados por clases. La segunda fila de gráficos muestra los resultados del paso de acondicionamiento para cada una de las clases. Estos gráficos son simplemente una versión desplazada y reescalada de lo que se muestra en la fila superior. A este nivel, con cero iteraciones, el algoritmo asigna incorrectamente al paciente a la clase EWS, este hecho se debe a que el punto verde se encuentra más cerca del centro de esta clase que del centro de la clase NHL. Sin embargo, vemos que aún ninguno de los agrupamientos corresponden a los de una Gaussiana isotrópica. A medida que las iteraciones progresan y como podemos ver en las últimas dos filas de gráficos, los agrupamientos se tornan más Gaussianos, mientras que el punto verde queda claramente incluido en la nube de observaciones de la clase NHL y lejos de la nube de puntos correspondientes a la clase EWS. Esta evolución, pero en términos de diagnóstico de acuerdo al valor de las probabilidades, puede observarse en la Figura 5.3. Aquí se muestra la evolución de las asignaciones $P[t]_k$, y su convergencia a las probabilidades p_k^j de que el tumor pertenezca a cada una de las cuatro clases. Inicialmente, todas las $P[t]_k$'s son iguales a 0.25 (el valor a priori) y luego, a pesar de que en un

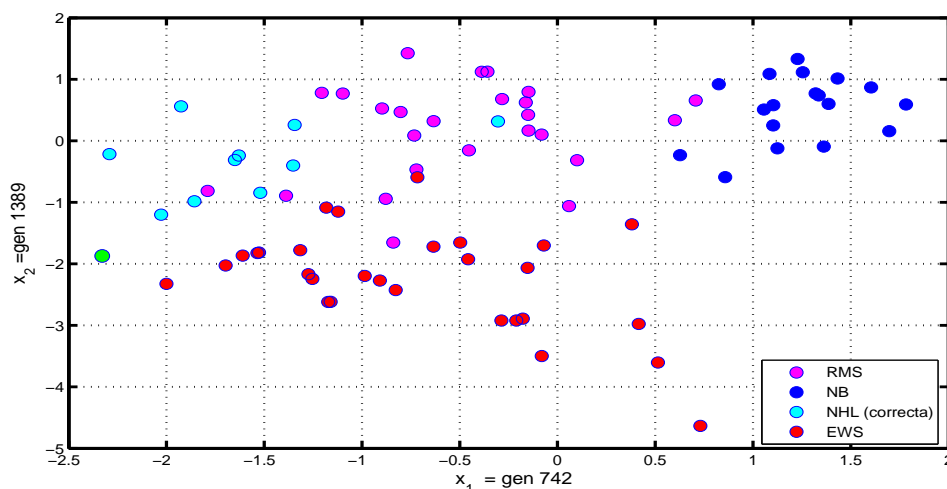


Figura 5.1: Datos sin procesar graficados considerando los dos genes seleccionados por el algoritmo de acuerdo a su capacidad para particionar la población. Para hacer el procedimiento de selección más expeditivo, se consideró cada gen de manera individual y la estimación de densidades se redujo sólo al paso de preconditionamiento, el cual consiste simplemente de un desplazamiento y “reescalado” de los datos. En el gráfico la población de entrenamiento está coloreada de acuerdo a su clase. En verde se muestra la observación que fue elegida como prueba. Esta elección fue realizada debido a que por su ubicación esta observación resulta una de las más difíciles de clasificar.

principio la probabilidad de pertenecer a la clase incorrecta (EWS) crece, ésta y las demás son superadas ampliamente por la probabilidad que da un diagnóstico correcto, NHL.

Un ejemplo correspondiente a los datos relacionados con las dos clases de leucemia es presentado en las Figuras 5.4, 5.5, y 5.6. El diagnóstico que mostramos se realizó considerando 13 genes; sin embargo los gráficos muestran los dos primeros genes seleccionados para tiempo $t = 0$ y las correspondientes dos componentes de $z = \phi_t(x)$ para tiempos posteriores. Como se aprecia en la Figura 5.6, vemos que el paciente es diagnosticado correctamente de AML desde el tiempo cero, sin embargo con una probabilidad apenas mayor a 0.5. A medida que las iteraciones avanzan, esta probabilidad alcanza el valor uno. La Figura 5.5 muestra esta evolución en el espacio de las dos primeras variables. Nuevamente, a medida que las nubes de puntos se convierten cada vez más parecidas a una muestra de una distribución Gaussiana, el punto verde es claramente rechazado de la clase ALL a la cual no pertenece y es absorbido por su verdadera clase, AML.

5.5.2. Diagnóstico de múltiples muestras

A partir de los excelentes resultados obtenidos para el diagnóstico de una muestra por vez, nos preguntamos ahora si podemos también obtener tan buenos resultados pero disponiendo de menos información. Para ser más precisos la pregunta sería: ¿podemos obtener resultados similares invirtiendo las fracciones de población de muestra y prueba?, es decir, usando sólo un puñado de observaciones de entrenamiento para diagnosticar el resto de las observaciones.

Para responder este interrogante, reducimos la población de entrenamiento para la clasificación de los cuatro tipos de tumores de células redondas y pequeñas a tan sólo 5 observaciones por cada tipo de tumor y las usamos para clasificar los 63 pacientes restantes. Los resultados obteni-

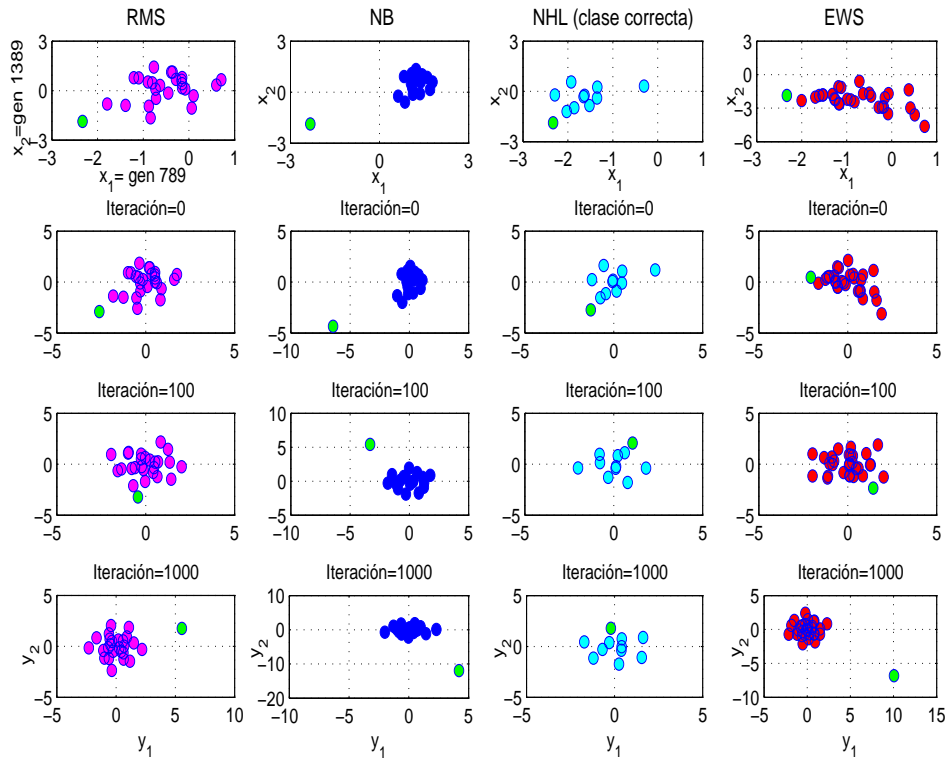


Figura 5.2: Las observaciones son transformadas de acuerdo a los flujos z_k , la observación de prueba (coloreada en verde) es “suavemente” asignada a las clases de acuerdo a su probabilidad a posteriori de pertenecer a cada una de éstas. Los gráficos superiores, al igual que en la Figura 5.1, muestran los datos originales en el plano definido por los dos genes seleccionados por el algoritmo pero aquí separados por clases. La segunda fila de gráficos es una versión centrada y reescalada de la de arriba. La tercer y cuarta muestran el progreso del algoritmo considerando ya la parte no lineal de este. A medida que cada clase se acerca a una muestra de una distribución Gaussiana, la observación de prueba es claramente absorbida o rechazada.

dos fueron también muy buenos: por ejemplo, utilizando 60 genes clasificamos correctamente el 95% de las observaciones. Resultados similares obtuvimos para el caso de leucemia: considerando una población de entrenamiento compuesta por sólo dos pacientes para cada clases, al utilizar 110 genes, obtuvimos el 90% de los diagnósticos de correctos. En todos los casos, la parte no lineal del algoritmo fue esencial para lograr esto, el paso de acondicionamiento por si solo arroja resultados más pobres.

Cabe señalar que, cuando la clase de entrenamiento es pequeña, la parte más débil de este procedimiento no es el proceso de clasificación sino la selección de los genes que se utilizan para esto. El problema reside en la dificultad para determinar cuáles son las variables que mejor particionan los datos en las distintas clases, siendo que para cada clase se tienen pocos datos. Es decir, las clases están poco representadas o vagamente determinadas. Esto además está agravado por el hecho de que al proceso de selección de genes lo hemos simplificado casi al máximo: evaluación de una variable a la vez y sólo considerando el paso de acondicionamiento.

Para corroborar lo expresado en el párrafo anterior, consideramos una población de entrenamiento mayor a la hora de seleccionar los genes y luego, para el proceso de clasificación, volvimos

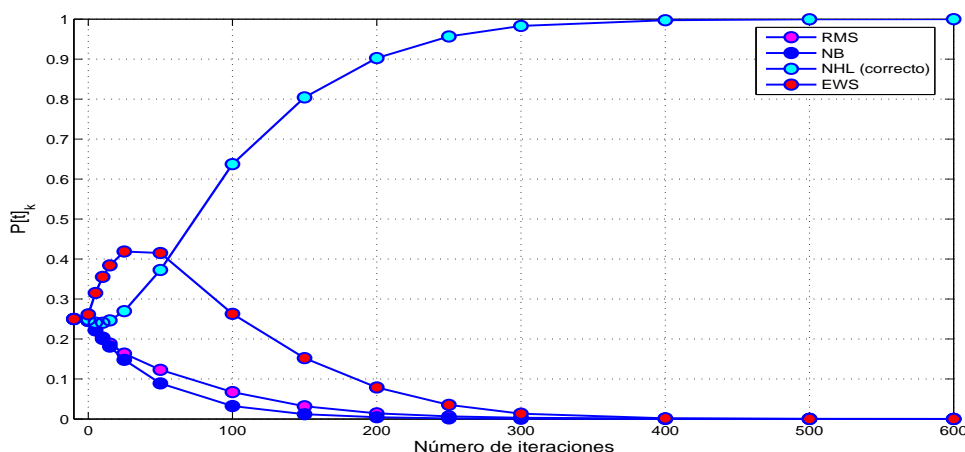


Figura 5.3: Evolución de las asignaciones $P[t]_k$ las cuales representan la probabilidad que la observación pertenezca a cada una de las clases C_k .

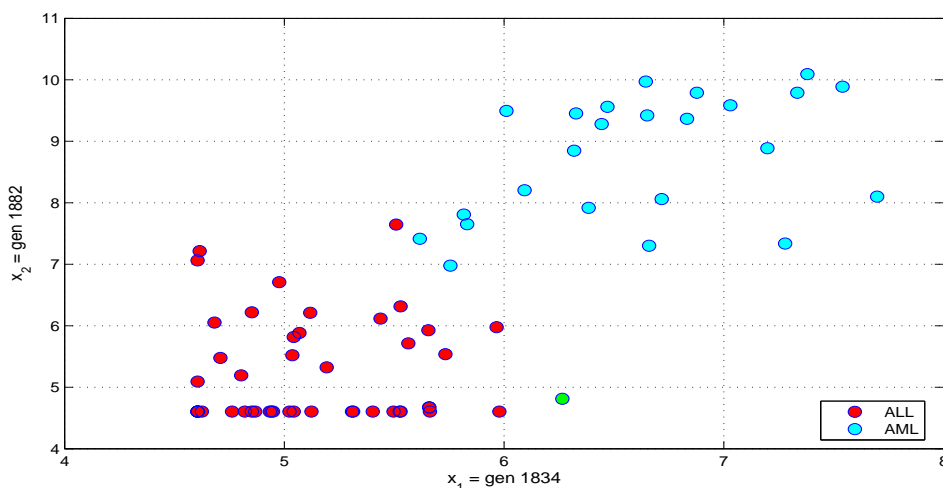


Figura 5.4: Ídem Figura 5.1, pero en este caso para la clasificación de tipos de leucemia aguda. Aunque aquí las observaciones son graficadas en el plano de los dos primeros genes elegidos por el algoritmo de selección, de acuerdo a su capacidad para agrupar, el número de genes tenido en cuenta para este ejemplo fue 13. Nótese, en este conjunto de datos sin filtrar, la presencia de varias observaciones para las cuales la expresión de los genes no muestra grandes variaciones y como si su expresión fuese la correspondiente a la de su valor normal. Esto hecho hace que se requiera de más de dos genes para realizar clasificación y agrupamiento. En particular, parecería imposible clasificar la observación graficada en verde, a partir de sólo estos dos genes seleccionados.

a utilizar sólo dos muestras por clase. Los resultados se acercaron nuevamente a casi el 100% de los diagnósticos correctos.

Este tema de la selección de genes será discutido más a fondo en el inciso a continuación y dentro del contexto del problema de agrupamiento de datos, donde la situación es aún más extrema ya que directamente no contamos con una población de entrenamiento.

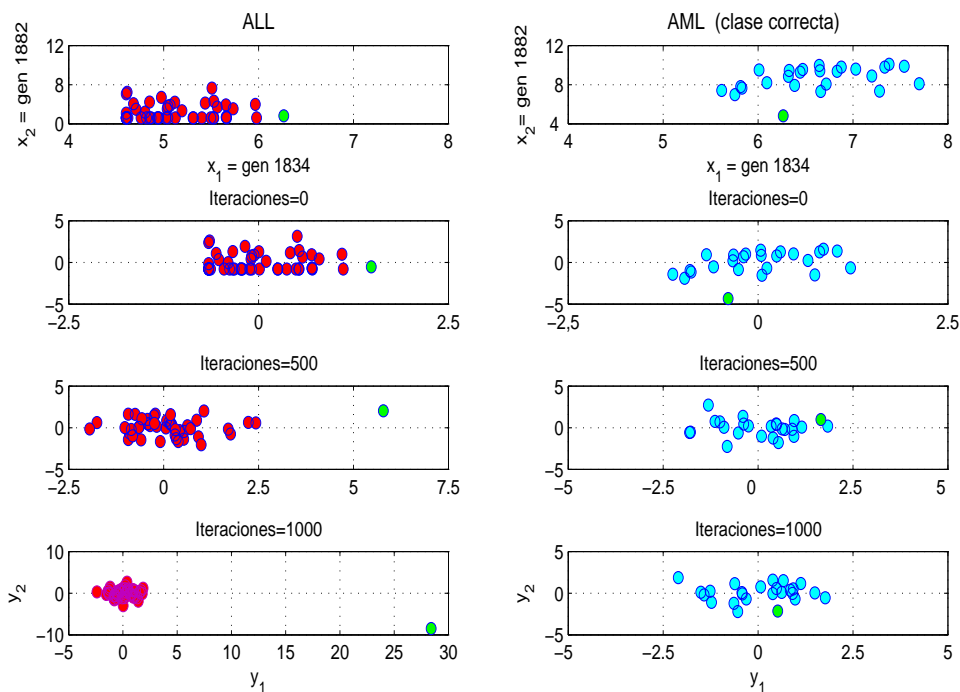


Figura 5.5: Ídem Figura 5.2, pero para los casos de leucemia. Se muestra la evolución en el plano de las dos primeras variables, las cuales corresponden a los dos genes seleccionados en el caso de $t = 0$. Sin embargo, el algoritmo realiza todo el trabajo en un espacio 13-dimensional y es a partir de utilizar estos 13 genes que el procedimiento puede identificar la clase correcta y tal como se ve en la última fila “rechaza” la observación de la nube de puntos correspondientes a la clase a la cual esta observación no pertenece.

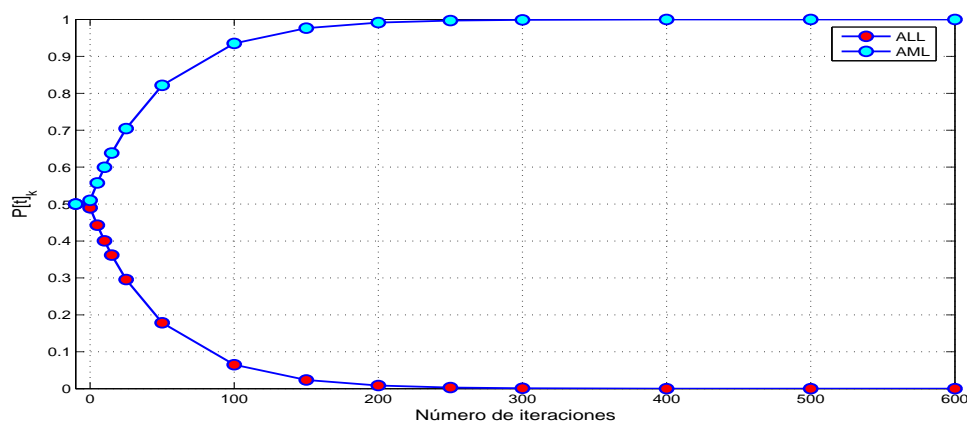


Figura 5.6: Ídem que la Figura 5.3, pero para la evolución de las asignaciones correspondientes a la observación de prueba perteneciente a la clase AML.

5.5.3. Agrupamiento

Llevando la idea del diagnóstico de múltiples muestras al límite, nos deshacemos de la población de entrenamiento y estamos entonces ante el problema de agrupamiento. La idea aquí es

verificar si el agrupamiento o partición de datos que propone el algoritmo coincide con la clasificación de tumores de acuerdo a sus tipos. En el lenguaje de [36], este proceso consiste en el “descubrimiento de clases”: patrones en los niveles de expresión genética que sugieran la existencia de diversos tipos de tumores.

Los pasos para realizar agrupamiento son los mismos que explicamos anteriormente, excepto que en este caso es necesario romper la simetría inicial entre todas las clases. Para esto, en vez de considerar inicialmente Q_k^j igual a las probabilidades a priori π_k^j , consideramos en cambio

$$Q_k^j = \pi_k^j + r_k^j, \quad (5.38)$$

donde r_k^j son número aleatorios, muy pequeños, y que sumados sobre k resultan igual a cero.

Tal como comentamos antes, el problema de la selección de genes aquí se agudiza debido a que la clase de entrenamiento desaparece por completo. No obstante, esto podría esperarse ya que se nos pide averiguar, a partir de un conjunto de miles de genes, cuáles son los que mejor particionan una población de un centenar de pacientes. Más aún, al no contar con una población de entrenamiento que brinde una referencia de los distintos tipos de cáncer, el procedimiento podría realizar un agrupamiento desde un ángulo totalmente diferente, ya sea por la edad del paciente, género, etnia, tipo de sangre, o enfermedad del corazón por ejemplo. Incluso variaciones aleatorias en los niveles de expresión genética, no atribuibles a ninguna causa biológica específica, podrían dar lugar a una partición que resulte robusta o coherente cuando la relación de variables vs pacientes es muy grande.

A partir de lo mencionado, decidimos en nuestros experimentos, realizar el proceso de selección de genes utilizando todas las observaciones como clase de entrenamiento. Una vez elegidos los genes, nos olvidamos de los diagnósticos (conocidos) para cada una de las observaciones y aplicamos el algoritmo para realizar el descubrimiento de clases. Los resultados fueron excelentes, acertando 100% de los diagnósticos considerando una cantidad suficiente de genes: para los tumores de células redondas y pequeñas 70 genes fueron suficientes, mientras que para el caso de los dos tipos de leucemia sólo 10 bastaron.

A continuación ilustramos estos experimentos, pero utilizando una cantidad menor de genes. La Figura 5.7 muestra los resultados utilizando 40 genes para realizar el agrupamiento de los diferentes tipos de tumores de células redondas y pequeñas: una única observación es asignada al grupo incorrecto. En la Figura 5.8 mostramos el resultado usando 6 genes para agrupar las observaciones de las dos tipos de leucemia. El “descubrimiento de clases” es igual al verdadero para el 93% de las observaciones.

A modo de ejemplo, en las Figuras 5.9 y 5.10 mostramos los resultados obtenidos para el problema de agrupamiento pero para el caso en el que al realizar la selección de genes no fue informado la clase a la cual pertenecía cada observación. Nuevamente 40 genes fueron utilizados para el agrupamiento de los tumores de células redondas y 6 genes para el agrupamiento de las muestras de leucemia. Podemos ver, que por más que estos genes fueron los que arrojaron mayores valores de la mediada M , la partición no fue realizada teniendo en cuenta los distintos tipos de tumores, sino simplemente considerando agruparla en cuatro y dos grupos respectivamente. O sea, fueron evaluados por su capacidad de agrupar la muestra en cuatro y dos grupos, posiblemente sin relación con los tipos de tumores. Los resultados obtenidos difieren notablemente de los obtenidos anteriormente. En la fila superior se muestran la verdadera partición o agrupamiento en clases donde cada una está coloreada en un tono distinto. En el gráfico de la derecha se puede apreciar lo poco natural que se ve la verdadera partición en clases vistas en el plano de los dos primeros genes seleccionados. En el gráfico de la izquierda se grafica lo mismo, pero en el plano de los dos genes seleccionados cuando se consideran todas las observaciones como población de entrenamiento

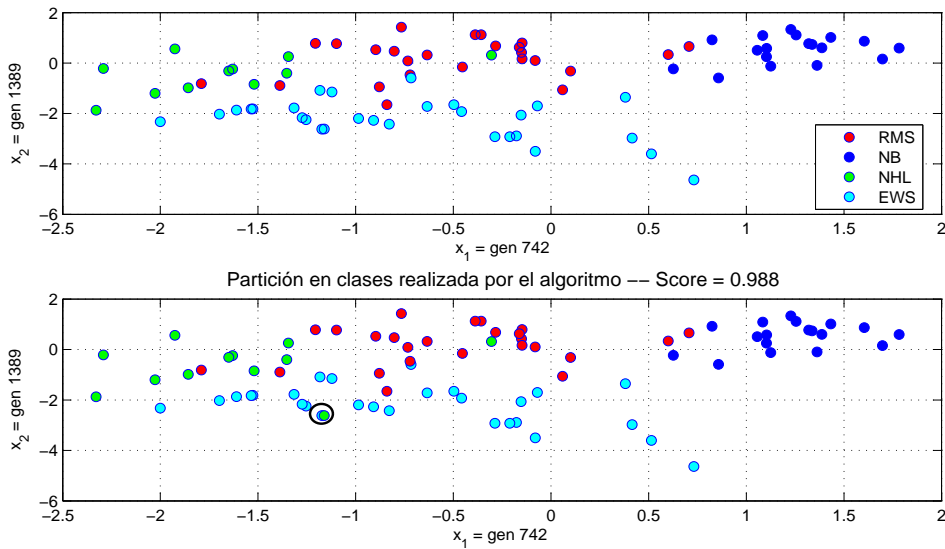


Figura 5.7: Agrupamiento de las muestras correspondientes a tumores de células redondas y pequeñas, considerando 40 genes pero graficado utilizando sólo los dos primeros. El panel superior muestra el agrupamiento correcto mientras que en el de abajo se ven las clases descubiertas por el algoritmo. Sólo una observación, resaltada con un círculo, fue asignada a una clase incorrecta.

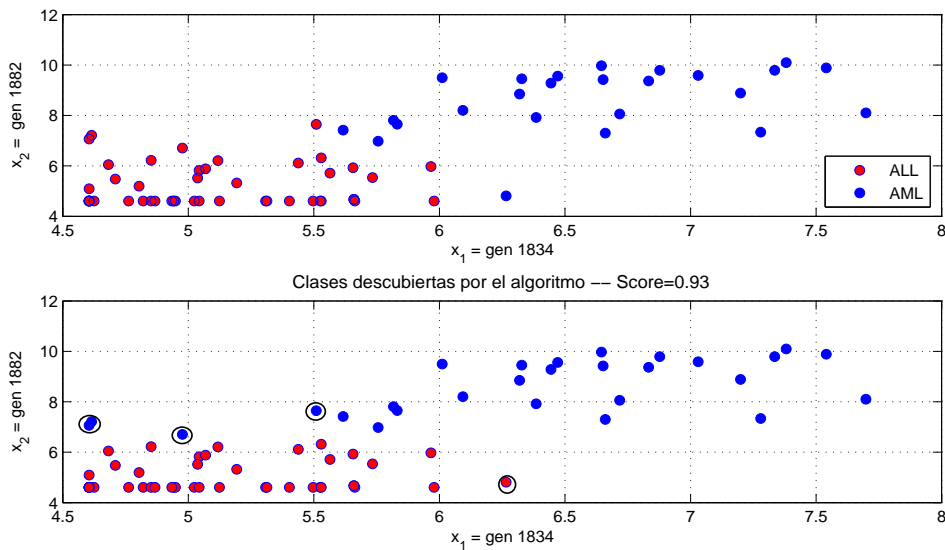


Figura 5.8: Ídem Figura 5.7, pero para los dos tipos de leucemia. El agrupamiento se realizó usando 6 genes. Encerrado en círculos se muestran los puntos que fueron erróneamente clasificados.

para realizar esta selección. La partición verdadera se ve reflejada de manera más natural. En fila inferior se muestran, los distintos agrupamientos obtenidos de acuerdo al conjunto de genes que fueron tenidos en cuenta. Para el caso en que los genes fueron elegidos sin información, podríamos decir que sólo una clase descubierta concuerda con la verdadera. A priori no está claro si podríamos llamar a este agrupamiento el “incorrecto” pues, debido a que los distintos tipos de tumores no fueron provistos, el algoritmo puede haber agrupado a esto tumores no de acuerdo al

tipo de tumor, sino por ejemplo de acuerdo al estado de desarrollo o a cualquier otra característica no observada a simple vista.

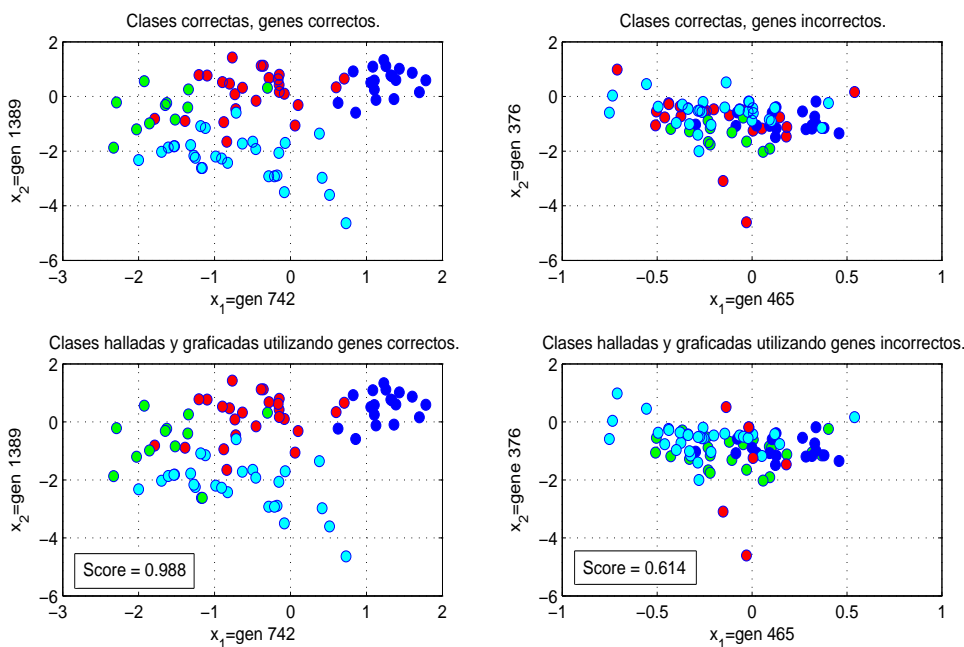


Figura 5.9: Agrupamiento de las muestras correspondientes a los distintos tipos de tumores de células redondas, utilizando 40 genes y graficado en el plano generado por los primeros dos. En el panel superior se muestra la partición de clases correcta, mientras que en el panel inferior se ve la partición realizada por el algoritmo. En los gráficos de la izquierda, se muestran los resultados para el caso en que los 40 genes fueron seleccionados por su capacidad de particionar las muestras de acuerdo a las distintas clases de tumores. En los de la derecha, se muestran los resultados cuando los genes fueron seleccionados de acuerdo a particionar las muestras sin usar observaciones etiquetadas con algún tipo de información.

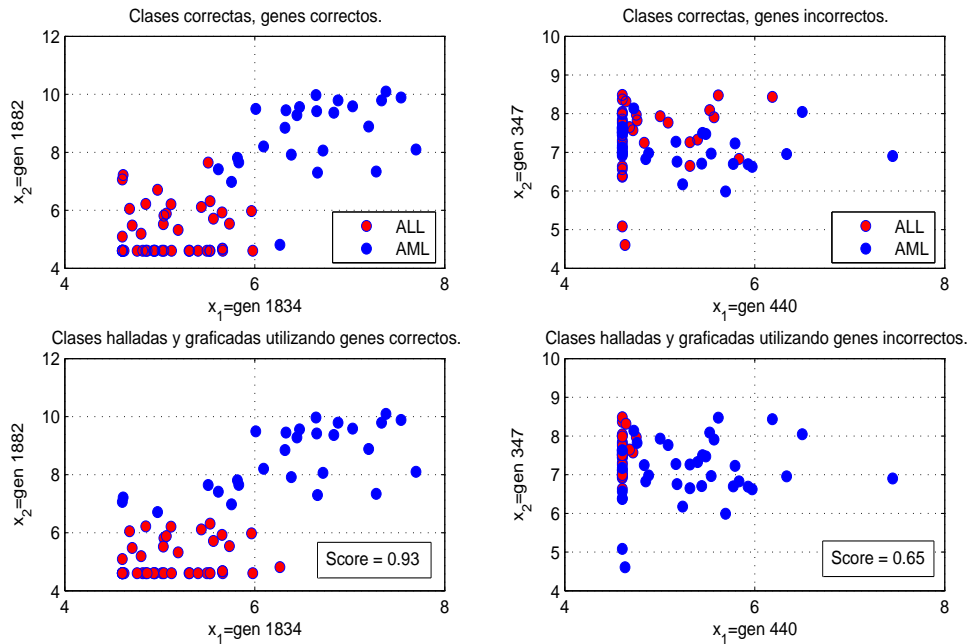


Figura 5.10: Ídem Figura 5.9, pero para el caso de leucemia y usando 6 genes.

5.6. Conclusiones del capítulo

En este capítulo hemos presentado una metodología destinada a resolver problemas de clasificación y agrupamiento de datos. La metodología propuesta implica considerar las poblaciones de entrenamiento y de prueba de manera casi indistinta. Este hecho resulta no sólo de utilidad desde el punto de vista de poder unificar los procedimientos para clasificación y agrupamiento de datos, sino que además resulta de mucha utilidad para atenuar el problema de la dimensión. Cuando estamos trabajando con una muestra pequeña de datos de dimensión muy grande, de acuerdo a las ideas aquí propuestas, podemos considerar de manera indistinta tanto a los datos de entrenamiento como a los de prueba y de esta forma contar con una única población de mayor volumen. La metodología está basada en el espíritu del algoritmo E-M y uno de sus pilares es la estimación de densidades. Este procedimiento consiste en la transformación de las muestras a través de “flujos”, a muestras correspondientes a una distribución Gaussiana isotrópica. En este contexto, las observaciones actúan como marcadores que guían el flujo. A medida que una observación es claramente asignada a una clase, esta observación comienza a cumplir un papel más activo en el flujo correspondiente a esa clase y a actuar de manera casi pasiva en el resto.

Cabe destacar, que de la metodología propuesta se desprende un procedimiento destinado a la selección de un conjunto de variables basado en la capacidad que estas poseen para particionar la muestras en clases distintas.

Con el objetivo de mostrar el desempeño de la metodología propuesta, ésta fue utilizada en dos ejemplos de aplicaciones médicas relacionados con la clasificación de tumores a partir de datos de niveles de expresión genética. De acuerdo a los resultados obtenidos en estos ejemplos podemos decir que esta metodología puede ser considerada como una herramienta de potencial ayuda a la hora de realizar diagnóstico, prevención y tratamiento de las enfermedades a partir del conocimiento de los niveles de expresión genéticas del paciente.

Capítulo 6

Discusión y Conclusiones generales

Con el objeto de mostrar aplicaciones de modelos y herramientas matemáticas para el estudio, análisis y comprensión de diferentes procesos y fenómenos que surgen en otras ciencias, en este trabajo hemos considerado el área de la biomedicina. El trabajo fue dividido en dos partes. En la primera parte nos hemos abocado a la detección de tumores y estimación de parámetros asociados a regiones tumorales y en la segunda parte a la clasificación de tumores a partir de datos de niveles de expresión genética.

En la primera situación el objetivo principal fue el desarrollo de una metodología para estimar la localización, tamaño y parámetros térmicos asociados a un tumor, inmerso en el tejido corporal, usando como información perfiles de temperaturas medidas sobre la superficie corporal en la zona próxima al tumor. Desde el punto de vista matemático, estimar la localización y tamaño o parámetros térmicos usando como información perfiles de temperaturas medidas sobre la superficie corporal, fue considerado como un problema inverso: dada información sobre el borde, la idea es determinar características internas, ya sea dónde debe estar localizado el tumor y cuál es su tamaño o en caso de conocer la localización del tumor la idea es determinar qué valor tienen algunos de los parámetros fisiológicos asociados a esta región.

Para modelar la transferencia de calor en tejidos vivos hemos considerado la ecuación de Pennes estacionaria con condiciones de bordes mixtas. Para este problema de borde probamos existencia y unicidad de la solución. Luego, propusimos un esquema de diferencias finitas de segundo orden para resolver en forma numérica este problema directo. Por su parte, para resolver los problemas inversos, estos fueron reformulados como problemas de optimización y para cada uno se definió una función costo apropiada. En ambos casos la función costo representa el error o diferencia entre un perfil de temperatura u_{exp} medido experimentalmente en la superficie corporal y la solución numérica u_p , obtenida mediante diferencias finitas, del problema de transferencia de calor considerando un valor de parámetro p . La idea entonces fue hallar el parámetro p^* que minimice la función costo. Para resolver los problemas de optimización propusimos dos metodologías distintas.

En la primera metodología presentada, Capítulo 3, hicimos uso del algoritmo Pattern Search para minimizar la función costo. Este algoritmo pertenece a la clase de algoritmos de optimización conocidos como algoritmos de búsqueda directa o también llamados de orden cero, ya que no hace uso de las derivadas para encontrar el mínimo deseado, lo cual lo convierte en un algoritmo de muy fácil implementación. La utilización de algoritmos libres del cálculo de derivada ha sido también propuesta en [61], utilizando algoritmos del tipo genético y en [56] utilizando redes neuronales. La metodología que aquí propusimos y la cual está basada en el trabajo [1], es de más fácil implementación que las anteriores y sin embargo se obtuvieron iguales o mejores resultados. Estos resultados muestra la conveniencia de utilizar la metodología presentada. Aún en los casos

en que 10 % y 15 % de ruido aleatorio fue adicionado a los datos de entrada, la metodología estima los valores de los diferentes parámetros con una muy buena precisión tanto en el caso 2D como en el caso 3D.

En el Capítulo 4, desarrollamos la segunda metodología propuesta. Esta hace uso de la información que provee la derivada de la función a minimizar con respecto a las distintas variables a estimar. Para calcular tal derivada, acudimos al análisis de sensibilidad. Primero fue tratado el caso de la sensibilidad al cambio de los parámetros con el objetivo de resolver el problema de la estimación del valor de la fuente de calor metabólico dentro de la región tumoral. Luego, para el problema particular de localización y estimación del tamaño del tumor, en el que la variable de control es el dominio donde el problema está definido, presentamos algunos de los conceptos relacionados con el análisis al cambio de forma y hallamos la fórmula para la derivada de la función con respecto a la variación del dominio. Aunque tradicionalmente el análisis de sensibilidad al cambio de forma ha estado muy ligado a la mecánica de las estructuras, de acuerdo a lo presentado en este trabajo estas herramientas pueden ser también aplicadas en problemas relacionados con la Medicina obteniendo excelente resultados. El cálculo de la derivada de la función costo no insume un gran tiempo computacional, ya que al plantear una ecuación adjunta es necesario sólo resolver otro problema de borde muy parecido al original, por lo que hemos utilizado el mismo esquema de diferencias finitas. El cálculo de la derivada del funcional costo y su posterior uso permitió obtener una metodología más eficiente, la cual arrojó, de acuerdo a los ejemplos mostrados, mejores resultados que la propuesta en el capítulo anterior.

Cabe señalar que los muy buenos resultados obtenidos se deben en gran medida a la fuerte restricción impuesta sobre la clase de funciones para los coeficientes térmicos y para la fuente de calor metabólico. Desde el punto de vista matemático, la suposición de considerar estas funciones como continuas a trozos implica una reducción notoria en la dimensión del conjunto de parámetros admisibles donde consideramos la minimización del funcional costo. De esta manera la minimización se realizó sobre un conjunto de dimensión finita.

En conclusión, de acuerdo a los resultados obtenidos ambas metodologías y en particular la segunda propuesta, pueden ser consideradas de potencial ayuda para localizar regiones tumorales, principalmente melanomas nodulares, como así también para estimar parámetros relacionados con estos que resultan de utilidad para estudiar su evolución bajo terapias como por ejemplo el BNCT.

En la segunda parte del trabajo, Capítulo 5, el objetivo principal fue desarrollar un algoritmo capaz de extraer, de una gran base de datos, información que reside de manera implícita en estos datos. Dicha información es previamente desconocida y puede resultar útil para describir o tomar decisiones sobre el proceso o fenómeno que está bajo análisis. Las ideas aquí desarrolladas corresponden al trabajo [2].

La metodología propuesta implica considerar las poblaciones de entrenamiento y de prueba de manera casi indistinta. Este hecho resulta no sólo de utilidad desde el punto de vista de poder unificar los procedimientos para clasificación y agrupamiento de datos, sino que además resulta de mucha utilidad para atenuar el problema de la dimensión. Cuando se está trabajando con una muestra pequeña de datos de dimensión muy grande, de acuerdo a las ideas que aquí propusimos, es posible considerar de manera indistinta tanto a los datos de entrenamiento como a los de prueba, y de esta forma contar con una única población de mayor volumen.

La metodología está basada en el espíritu del algoritmo E-M y unos de sus pilares es la estimación de densidades. Este procedimiento consiste en la transformación de las muestras por medio de “flujos”, a muestras correspondientes a una distribución Gaussiana isotrópica. En este contexto, las observaciones actúan como marcadores que guían el flujo. A medida que una observación es claramente asignada a una clase, esta observación comienza a cumplir un papel más activo en el

flujo correspondiente a esa clase, y por lo tanto actúa de manera casi pasiva en el resto.

Cabe destacar, que de la metodología propuesta se desprende un procedimiento destinado a la selección de un conjunto de variables basado en la capacidad que estas poseen para particionar la muestras en clases distintas.

Con el objetivo de demostrar el desempeño de la metodología propuesta, la utilizamos en dos ejemplos de aplicaciones médicas relacionados con la clasificación de tumores a partir de datos de niveles de expresión genética. De acuerdo a los resultados obtenidos en estos ejemplos podemos decir que esta metodología puede ser considerada como una herramienta de potencial ayuda a la hora de realizar diagnóstico, prevención y tratamiento de las enfermedades a partir del conocimiento de los niveles de expresión genéticas del paciente.

Cabe señalar que aunque los problemas tratados en las distintas partes de este trabajo y las herramientas matemáticas necesarias para resolver cada uno de estos puedan parecer en un principio muy disímiles, queremos hacer notar que existe un grado de relación entre estos y que es posible aplicar el tipo de herramientas e ideas utilizadas en la Parte 2 para resolver los problemas inversos planteados en la Parte 1.

En las aplicaciones relacionadas con problemas inversos usualmente el interés reside en hallar un conjunto de parámetros correspondientes a un modelo que describe un fenómeno en estudio, a partir de una observación parcial de la solución del modelo (por ejemplo mediciones realizadas en el borde o sólo componentes de la solución). Es decir, tenemos una ecuación de la forma

$$y = Tx \quad (6.1)$$

donde T es un operador que relaciona los parámetros del modelo x con las observaciones y . Tal como comentamos en el Capítulo 3 este tipo de problemas son generalmente mal condicionados ya sea porque no existe la solución, existe pero no es única o posiblemente debido a que esta no depende continuamente de los datos y . Relacionado con este último concepto, debemos considerar que en las aplicaciones las mediciones que uno realiza siempre están contaminadas con ruido aleatorio y por lo tanto el problema que queremos resolver es del tipo

$$y = Tx + r \quad (6.2)$$

donde r representa ruido aleatorio con función densidad de probabilidad η .

Una manera de regularizar los problemas es introduciendo toda la información que conozcamos de antemano acerca del problema y en particular sobre los parámetros del modelo. Una de las formas posibles de realizar esto es utilizando un enfoque probabilístico. El uso de información a priori es la característica básica de los denominados métodos Bayesianos. Supongamos que los parámetros x tienen una cierta distribución de probabilidades dada por la densidad $\mu(x)$. Entonces, en lugar de buscar un único valor estimado para el conjunto de parámetros x , lo que podemos hacer es buscar un conjunto de estimaciones con diferentes probabilidades. Siguiendo esta idea podemos también suponer que las mediciones son el valor de una variable aleatoria con distribución ρ . La pregunta entonces es la siguiente: ¿qué podemos decir de los parámetros x cuando la observación y es conocida?. En otras palabras, el interés es conocer la densidad condicional del conjunto de parámetros dada la observación y .

Dijimos que el ruido tiene una distribución con función densidad η conocida, tenemos entonces que la densidad de las observaciones y dados los parámetros x está dada por

$$\rho(y/x) = \eta(y - Tx).$$

Luego utilizando la formula de Bayes tenemos que:

$$\mu(x/y) = \frac{\rho(y/x)\mu(x)}{\sum_x \rho(y/x)\mu(x)} \quad (6.3)$$

donde $\mu(x/y)$ es la función densidad de probabilidad condicional “a posteriori”, $\mu(x)$ es la función densidad de probabilidad a priori y $\rho(y/x)$ es la función de verosimilitud. La función $\mu(x/y)$ nos provee de toda la información sobre el conjunto de parámetros x , cuando conocemos su distribución a priori y hemos observado la medición y . Por ejemplo, maximizando la función densidad a posteriori podemos determinar el valor más probable de x dada la observación y . Este enfoque es empleado en [78] para la estimación del coeficiente de absorción correspondiente al tejido cerebral, coeficiente que resulta de importancia para detectar la presencia de un tumor. La mediciones son imágenes de tomografía óptica para las cuales suponen una distribución Gaussiana con media cero y matriz de covarianza conocida. Por su parte la distribución a priori considerada para el coeficiente de absorción es también un distribución Gaussiana con media cero pero con matriz de covarianza a determinar. Con estas suposiciones la densidad a posteriori también tendrá una distribución Gaussiana con media y covarianza definida en términos de las otras dos. Otro ejemplo de este enfoque se puede hallar en [22], quienes utilizan esta metodología para estimar parámetros asociados a tumores de mamas. Más detalles sobre la teoría de problemas inversos desde un enfoque probabilístico son presentados en [75].

De acuerdo a lo anteriormente expresado, abordar el estudio de los problemas inversos desde una perspectiva Bayesiana puede resultar de mucha utilidad en el caso en que los datos o mediciones presenten un comportamiento estocástico, así como también en el caso en que se conozca a priori información de relevancia acerca de la solución buscada. Este enfoque será considerado como una posible futura línea de trabajo con el objetivo de profundizar el estudio de los problemas inversos aquí tratados.

Bibliografía

- [1] AGNELLI J.P., BARREA A. AND TURNER C.V., *Tumor location and parameter estimation by thermography*, Mathematical and Computer Modelling, 53 (2011), pp. 1527-1534.
- [2] AGNELLI J.P, CADEIREAS M., TABAK E., TURNER C.V. AND VANDEN-EIJNDEN E., *Clustering and classification through normalizing flows in feature space*, SIAM: Multiscale Modeling and Simulation, 8 (2010), pp. 1784-1802.
- [3] AGNELLI J.P., PADRA C. AND TURNER C.V., *Shape optimization for tumor tumor location* Enviado: Computers and Mathematics with Applications, 2010.
- [4] ALESSANDRINI G., *Generic uniqueness and size estimates in the inverse conductivity problem with one measurement*, Le Matematiche, Vol LIV (1999), pp. 5-14.
- [5] ALESSANDRINI G. ROSSET E. AND SEO J.K., *Optimal size estimates for the inverse conductivity problem with one measurement*, Proceedings of the American Mathematical Society, Vol 128 (1999), No. 1, pp. 53-64.
- [6] AMMARI H., BONNETIER E., CAPDEBOSQ Y., TANTER M. AND FINK M., *Electrical impedance tomography by elastic deformation*, SIAM: Journal on Applied Mathematics, Vol. 68 (2008), No. 6, pp. 1557-1573.
- [7] AMMARI H. AND KANG H., *Reconstruction of Small Inhomogeneities from Boundary Measurements*, Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2004.
- [8] ASSENHEIMER M., LAVER-MOSKOVITZ O., MALONEK D., MANOR D., NAHALIEL U., NITZAN U., AND SAAD A., *The Tscan technology: electrical impedance as a diagnostic tool for breast cancer detection*, Physiological Measurement, 22 (2001), pp. 1-8.
- [9] BARBER D.C. AND BROWN B.H., *Applied Potential Tomography*, Journal of Physics E:Scientific Instruments, 17 (1984), pp. 723-733.
- [10] BARZILAI J. AND BORWEIN J.M, *Two point step size gradient methods*, IMA Journal of Numerical Analysis, Vol. 8 (1988), Number 1, pp. 141-148.
- [11] BELLMAN CR., *Adaptative control processe: A guide tour*, Princeton University Press, Princeton, 1961.
- [12] BERTSEKAS D.P., *On the Goldstein-Levitin-Polyak gradient projected method*, IEEE Transactions on Automatic Control, 21 (1976), pp. 174-184.
- [13] BIRGIN E.G, MARTINEZ J.M AND RAYDAN M., *Nonmonotone spectral projected gradient methods on convex sets*, SIAM Journal on Optimization, Vol. 10 (2000), Number 4, pp. 1196-1211.

- [14] BISHOP C.M., *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
- [15] BOWMAN H.F., *Estimation of tissue blood flow*, in: A. Shitzer, R.C. Eberhart (Eds.), *Heat Transfer in Medicine and Biology*, vol I, Plenum Press, NY, 1985.
- [16] CALDERÓN A.P., *On a inverse boundary problem*, Seminar on Numerical Analysis and its Application to Continuum Physics (Rio de Janeiro, 1980), Brazilian Mathematical Society, pp 65-73.
- [17] CÉA J., Problems of shape optimal design, in: E.J. Haug, J. Céa (Eds.), *Optimization of Distributed Parameters Structures*, Sijthoff and Noordhoff, Alphen a/d Rijn, 1981
- [18] CHAPPELLE O., SCHÖLKOPF B. AND ZIEN A., *Semi-supervised Learning*, MIT Press, Cambridge, MA, 2006.
- [19] CHATO J.C., *Measurement of thermal properties of biological materials*, in: A. Shitzer, R.C. Eberhart (Eds.), *Heat Transfer in Medicine and Biology*, vol I, Plenum Press, NY, 1985, pp.167-173.
- [20] CHEN S.S. AND GOPINATH R.A., *Gaussianization*, in *Advances in Neural Information Processing Systems 13*, T. K. Leen, T. G. Dietterich, and V. Tresp, eds., MIT Press, Cambridge, MA, 2001, pp. 423–429.
- [21] CHEN M.M., PEDERSEN C.O. AND CHATO J.C., *On the feasibility of obtaining three dimensional information from thermographic measurements*, *ASME Journal of Biomechanical Engineering*, 99 (1977), pp.58-64.
- [22] CHOUDHARY H. AND NEHORAI A., *Tumor detection using Bayesian conjugate prior in diffuse optical tomography*, *Proc. IEEE Engineering in Medicine and Biology Society* (2006), New York, USA, pp. 2655-2658.
- [23] DEMPSTER A., LAIRD N. AND RUBIN D., *Likelihood from incomplete data via the EM algorithm*, *J. Roy. Statist. Soc. Ser. B*, 39 (1977), pp. 1–38.
- [24] DENG Z. AND LIU J., *Mathematical modelling of temperature over skin surface and its implementation in thermal disease diagnostics*, *Computers in Biology and Medicine*, 34 (2004), pp.495-521.
- [25] DENNIS J. AND TORCZON V., *Derivative free pattern search method for multidisciplinary design problems*, *Proceeding of AIAA*, vol 26(20202), pp. 523-534.
- [26] DHILLON I.S., MALLELA S. AND KUMAR R., *A divisive information-theoretic feature clustering algorithm for text classification*, *J. Mach. Learn. Res.*, 3 (2003), pp. 1265–1287.
- [27] DIAKIDES N.A. AND BRONZINO J.D., *Medical infrared imaging*, CRC Press, Boca Raton, Florida, USA, 2007.
- [28] ENGL H. W., HANKE M. AND NEUBAUER A., *Regularization of inverse problems*, Kluwer Academic Publisher, Dordchet, 1996.
- [29] EVANS L.C., *Partial Differential Equations*, American Mathematical Society, 1998.
- [30] FABES E., KANG H. AND SEO J.K., *Inverse conductivity problem with one measurement: errors and estimate approximate identification for perturbed disks* *SIAM J. Math. Anal.*, vol 30 (1999), No. 4, pp. 699-720.

- [31] FANCELLO E., *Análise de sensibilidade, geração adaptativa de malhas e o método dos elementos finitos na otimização de forma em problemas de contato e mecânica de fratura*, Tese do grau de Doutor em Ciências em Engenharia Mecânica, Universidade Federal do Rio de Janeiro, Brasil, 1993.
- [32] FEIJOO G. , *On shape sensitivity analysis*, Dissertation for the degree of Doctor of Philosophy, Stanford University, 2003.
- [33] FEIJÓO R.A., PADRA C, SALIBA R., TAROCO E. AND VÉNERE M.J., *Shape sensitivity analysis for energy release rate evaluation and its application to the study of three-dimensional cracked bodies*, Computer Methods in Applied Mechanics and Engineering, Vol. 188 (2000), Issue 4, pp. 649-664
- [34] FRIEDMAN A. AND ISAKOV V., *On the uniqueness in the inverse conductivity problem with one measurement*, Indiana Univ. Math. J., 38 (1989), pp. 553-580.
- [35] FRIEDMAN J.H., W. STUETZLE W. AND SCHROEDER A., *Projection pursuit density estimation*, J. Amer. Statist. Assoc., 79 (1984), pp. 599-608.
- [36] T. R. GOLUB T.R., SLONIM D.K., TAMAYO P., HUARD C., GAASENBEEK M., MESIROV J.P., COLLER H., LOH M.L., DOWNING J.R., CALIGIURI M.A., BLOOMFIELD C.D. AND LANDER E.S., *Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring*, Science, 286 (1999), pp. 531-537.
- [37] GORDON C., WEBB D.L. AND WOLPERT S., *One Cannot Hear the Shape of a Drum*, Bulletin of the American Mathematical Society 27 (1992), pp. 134-138.
- [38] GRIPPO L., LAMPARIELLO F. AND LUCIDI S., *A nonmonotone line search technique for Newton's method*, SIAM Journal Numerical Analysis, Vol. 23 (1986), Number 4, pp. 707-716.
- [39] GUYON I. AND ELISSEEFF A., *An introduction to variable and feature selection*, J. Mach. Learn. Res., 3 (2003), pp. 1157-1182.
- [40] HADAMARD J., *Lectures on Cauchy problem in Linear Differential Equations*, Yale University Press, New Haven, 1923.
- [41] HAUG E.J., CHOI K.K. AND KOMKOV V. *Design Sensitivity Analysis of Structural Systems*, Academic Press, New York, 1986.
- [42] HOLDER D., *Electrical Impedance*, Bristol Institute of Physics Publishing, 2005.
- [43] KAC M., *Can one hear the shape of a drum?*, American Mathematical Monthly 73 (4, part 2, 1996), pp. 1-23.
- [44] KANG H. AND SEO J.K., *Layer potential technique for the inverse conductivity problem*, Inverse Problems, 12 (1996), pp. 267-278.
- [45] KANG H. AND SEO J.K., *Inverse conductivity problem with one measurement: uniqueness of balls in \mathbb{R}^3* , SIAM J. Appl. Math, vol 59 (1999), No. 5, pp. 1533-1539.
- [46] KANG H., SEO J.K AND SHEEN D., *The inverse conductivity problem with one measurement: stability and estimation of size*, SIAM J. Math. Anal., vol 28 (1997), No. 6, pp. 1389-1405.

- [47] KELLER J.B., *Inverse Problems*, The American Mathematical Monthly, Vol. 83 (1976), No. 2, pp. 107-118
- [48] J. KHAN, J. S. WEI, M. RINGNÉR, L. H. SAAL, M. LADANYI, F. WESTERMANN, F. BERTHOLOD, M. SCHWAB, C. R. ANTONESCU, C. PETERSON, AND P. S. MELTZER, *Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks*, Nature Medicine, 7 (2001), pp. 673–679.
- [49] KHON R. AND VOGELIUS M., *Determining conductivity by boundary measurements*, Communication in Pure and Applied Mathematics, 37 (1984), pp 289-298.
- [50] KULLBACK S. AND LEIBLER R.A., *On information and sufficiency*, Ann. Math. Statistics, 22 (1951), pp. 79–86.
- [51] LAWSON R.N., *Implications of surface temperatures in the diagnosis of breast cancer*, Canadian Medical Association Journal, vol. 75 (1956), pp.309-310.
- [52] LAWSON R.N. AND CHUGTAI M.S., *Breast cancer and body temperatures*, Canadian Medical Association Journal, vol. 88 (1963), pp.68-70.
- [53] LEWIS R.M. AND TORCZON V., *Pattern Search Algorithms for Bound Constrained Minimization*, SIAM Journal on Optimization, Vol. 9 (1999), Number 4, pp. 1082-1099.
- [54] J. LIU AND L.X. XU, *Boundary information based diagnostics on the thermal states of biological bodies*, Int. J. of Heat and Mass Transfer, 43 (2000), pp. 2827-2839.
- [55] MARROCCO A. *Simulations numériques dans la fabrication des circuits a semiconducteurs*, Rapport de Recherche INRIA, Rocquencourt 305 (1984).
- [56] MITAL M. AND SCOTT E.P., *Thermal Detection of Embedded Tumors using Infrared Imaging*, ASME Journal of Biomechanical Engineering, vol. 129 (2007), pp.33-39.
- [57] MIYAKAWA M. AND BOLOMEY J.C.(Eds), *Non-Invasive Thermometry of the Human Body*, CRC Press, Boca Raton, 1996.
- [58] NATTERER F., *The Mathematics of Computerized Tomography*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2001.
- [59] OSGOOD B., PHILLIPS R. AND SARNAK P., *Compact isospectral sets of surfaces*, Journal of Functional Analysis, Vol. 80 (1988), pp. 212-234.
- [60] ROMERO-MENDÉZ R., JIMÉNEZ-LOZANO J.N. AND GONZÁLEZ F.J., *Analytical solution of the Pennes equation for burn-depth determination from infrared thermographs*, Mathematical Medicine and Biology, (2009) In press.
- [61] PARUCH M. AND MAJCHRZAK E., *Identification of tumor region parameters using evolutionary algorithm and multiple reciprocity boundary element method*, Engineering Applications of Artificial Intelligence, 20 (2007), pp.647-655.
- [62] PENNES H., *Analysis of tissue and arterial blood temperature in the resting human forearm*, J. Appl. Physiol. 1 (1948), pp.93-122.
- [63] PERTHAME B. AND ZUBELLI J.P., *On the inverse problem for a size structured population model*, Inverse Problems , 23 (2007), pp. 1037-1052.

- [64] RADON J., *Über die Bestimmung von Funktionen durch ihre Integrawerte längs bestimmter Mannigfaltigkeiten*, Ber. Verb. Sächs. Akad. Wiss. Leipzig, Math.-Nat. Kl., 69 (1917), pp. 262-277.
- [65] OSWALD-TRANTA B., SORGER M. AND O'LEARY P., *Thermographic crack detection and failure classification*, J. Electron. Imaging, 19 (2010).
- [66] SALIBA R., PADRA C., VÉNERE M.J., TAROCO E. AND FEIJÓO R.A., *Adaptivity in linear elastic fracture mechanics based on shape sensitivity analysis*, Computer Methods in Applied Mechanics and Engineering, Vol. 194 (2005), Issues 34-35, pp. 3582-3606
- [67] SANTA CRUZ G.A., BERTOTTI J., MARÍN J., GONZÁLEZ S.J., GOSSIO S., ALVAREZ D., ROTH B.M.C., MENÉNDEZ P., PEREIRA M.D., ALBERO M., CUBAU L., ORELLANO P. AND LIBERMAN S.J., *Dynamic infrared imaging of cutaneous melanoma and normal skin in patients treated with BNCT*, Vol. 67 (2009), s54-s58.
- [68] SHARAFUTDINOV V.A., *Integral geometry of tensor fields*, Inverse and Ill-posed Problems Series. VSP, Utrecht, 1994. Appl. Radiat. Isotopes, vol. 67 (2009), pp s54-s58.
- [69] SILVA NETO A.J. AND OZISIK M.N., *Simultaneous estimation of location and timewise-varying strength of a plane heat source*, Num. Heat Transfer, Part A: Applications, 25 (1993), pp. 467-477.
- [70] SOKOLOWSKI J. AND ZOLESIO J.P., *Introduction to Shape Optimization: Shape sensitivity Analysis*, Springer-Verlag, Berlin, 1991.
- [71] SU J. AND SILVA NETO A.J., *Two-dimensional inverse heat conduction problem of source strength estimation in cylindrical rods*, Applied Mathematical Modelling, 25 (2001), pp. 861-872.
- [72] SYLVESTER J. AND UHLMANN G., *A global uniqueness theorem for an inverse boundary value problem* Annals of Mathematics, 125 (1987), pp 153-169.
- [73] TABAK E. AND VANDEN-EIJNDEN E., *Density estimation by dual ascent of the log-likelihood*, Commun. Math. Sci., 8 (2010), pp. 217-233.
- [74] TAO X. AND ZHANG S., *Unique continuation at the boundary for elliptic operators in Dini domains*, Southeast Asian Bulletin Mathematics, Vol. 26 (2002), pp. 523-534.
- [75] TARANTOLA A., *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.
- [76] TORCZON V., *On the convergence of Pattern Search Algorithms*, SIAM Journal on Optimization, Vol. 7 (1997), Number 1, pp. 1-25.
- [77] UHLMANN G., *Electrical impedance tomography and Calderón's problem*, Inverse Problems, Vol. 25 (2009), Number 12.
- [78] UMADEVI V., SURESH S. AND RAGHAVAN V. *Improved infrared thermography based imaged construction for biomedical applications using Markov chain Monte Carlo Method*, Proc. IEEE Engineering in Medicine and Biology Society (2009), Minneapolis, USA, pp. 5360-5363.
- [79] VAPNIK V.N., *Statistical Learning Theory*, Wiley, New York, 1998.

- [80] VIGNÉRAS M.F., *Variétés riemanniennes isospectrales et non isométriques*, Annals of Mathematics, 112 (1980), pp. 21-32.
- [81] ZOLESIO J.P. *The material derivative (or Speed) method for shape optimization*, Optimization for distributed parameters structures (Eds. Haug E.J. and Cea J.), Sijthoff & Noordhoff, Alphen aan den Rijn, Netherlands, 1981.
- [82] ZUBELLI J.P., MARABINI R., SORZANO C.O.S AND HERMAN G.T., *Three- dimensional reconstruction by Chahine's method from electron microscopic projections corrupted by instrumenstal aberrations*, Inverse Problems, 19 (2003), pp. 933-949.