

00365



**UNIVERSIDAD NACIONAL AUTONOMA  
DE MEXICO**

---

**FACULTAD DE CIENCIAS  
DIVISION DE ESTUDIOS DE POSGRADO**

**ESTIMACION DE PARAMETROS EN UN  
MODELO DE ACUIFEROS**

**T E S I S**  
QUE PARA OBTENER EL GRADO ACADEMICO DE  
**MAESTRIA EN CIENCIAS (MATEMATICAS)**  
**P R E S E N T A :**  
**ROSA MARGARITA ALVAREZ GONZALEZ**

DIRECTOR DE TESIS: DRA. SUSANA GOMEZ GOMEZ

MEXICO, D. F.

1997

**TESIS CON  
FALLA DE ORIGEN**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

**A José, hombre sensible y hermoso,  
con quien comparto la vida y a quien agradezco  
su dedicación y amor.**

# CONTENIDO

<b>1</b>	<b>Modelación Matemática</b>	<b>4</b>
1.1	Aguas subterráneas. Acuíferos. Clasificación. . . . .	4
1.2	Leyes que rigen el movimiento de las aguas subterráneas. . . . .	6
1.3	Conceptos geofísicos fundamentales. . . . .	8
<b>2</b>	<b>Problemas centrales.</b>	<b>13</b>
2.1	Problema directo . . . . .	13
2.1.1	Esquemas en diferencias . . . . .	14
2.1.2	Esquemas dependientes de un parámetro. . . . .	18
2.2	Problema inverso . . . . .	21
2.2.1	Método de Newton. . . . .	22
2.2.2	Métodos Quasi-Newton. . . . .	24
<b>3</b>	<b>Mal planteamiento del problema inverso y alternativas de solución.</b>	<b>32</b>
3.1	Mal planteamiento. . . . .	33
3.2	Regularización . . . . .	37
3.3	Optimización Multiescala. . . . .	40
3.3.1	Parametrización multiescala. . . . .	42
3.3.2	Proceso de optimización con base en multiescala. . . . .	46
<b>4</b>	<b>Experimentación y resultados computacionales.</b>	<b>50</b>
4.1	Caso sintético. . . . .	51
4.1.1	Datos sin error. . . . .	53
4.1.2	Presencia de error en los datos sintéticos . . . . .	53
4.2	Caso real: acuífero cubano de Ariguanabo. . . . .	57
4.2.1	Datos con error. . . . .	60

## **PROLOGO:**

El presente trabajo fue llevado a cabo como parte de las actividades desarrolladas en el marco del proyecto titulado " SOLUCION DEL PROBLEMA INVERSO PARA LA EXPLOTACION DE ACUIFEROS", el cual representa un punto más en la agenda de colaboración entre México y Cuba. En este contexto, constituye una primera aportación -exitosa en términos generales, según se verá- para la resolución de un problema considerado importante por las partes involucradas, tanto desde el punto de vista académico como del tecnológico.

Por supuesto, quedan cosas por hacer, pues los campos de investigación y desarrollo al respecto son inmensos. En este sentido, algunas propuestas y líneas de trabajo definidas ya en este primer producto, serán la base de mi doctorado.

Consciente de que una lista de agradecimientos nunca es completa, doy la mía:

A la UNAM, por el apoyo recibido, tanto en lo económico como en lo académico.

Al ICIMAF por haber avalado en todo momento mi iniciativa de realizar estudios de posgrado en este bello país.

A Susana Gómez, quien propuso y dirigió este trabajo.

A Angel Pérez, compañero de trabajo y colaborador importante en el soporte computacional de los resultados que aquí se reportan.

A Pablo Barrera, que mucho ha tenido que ver con mi formación en Optimización y Análisis Numérico.

A los demás miembros del jurado por sus opiniones, críticas y sugerencias.

A mis padres, por su constante apoyo.

A Lilliam Alvarez, por sus gestiones, comprensión y apoyo.

Al Dr. Rafael Pérez Pascual, por todas sus atenciones.

## RESUMEN:

La modelación de un acuífero de flujo horizontal, confinado, anisotrópico y no homogéneo, da lugar a una ecuación en derivadas parciales del tipo

$$\frac{\partial}{\partial x} \left( T(x, y) \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T(x, y) \frac{\partial h}{\partial y} \right) = S \frac{\partial h}{\partial t} + Q \quad (0.1)$$

con ciertas condiciones iniciales y de frontera.

Ahora bien, dada  $T(x, y)$  (transmisividad del acuífero) es posible resolver (0.1) y obtener la solución  $h$  (carga hidráulica) correspondiente. En términos operacionales esto puede escribirse como

$$L(T) = h \quad (0.2)$$

que se conoce como el problema directo.

En la práctica, sin embargo, es frecuente conocer ciertas observaciones  $\hat{h}$  de  $h$  y a partir de ellas tratar de obtener  $\hat{T}$  tal que

$$L(\hat{T}) = \hat{h} \quad (0.3)$$

Este es conocido como el problema inverso y ya sea debido a errores en  $\hat{h}$  o a inadecuaciones de (0.1) con el acuífero real en estudio, en general no tiene solución, por lo que es usual estimar  $\hat{T}$  como la  $T^*$  que resuelve

$$\min_T F(T) \quad F(T) = \|L(T) - \hat{h}\|^2 \quad (0.4)$$

La parte central de esta tesis consiste en resolver (0.4) de manera estable pues se trata de un problema de optimización no lineal originado en (0.3) que es típicamente mal-planteado a la *Hadamard*. Se discuten algunas soluciones clásicas y se presenta una alternativa, todavía en desarrollo, basada en multiescala que ha probado adaptarse con éxito a nuestro caso de estudio. Se dan también resultados computacionales para datos sintéticos y para datos correspondientes a una región del acuífero cubano de Ariguanabo.

## INTRODUCCION:

El agua subterránea constituye un factor muy importante de los recursos hidráulicos, abasteciendo la industria, la agricultura y a la población para uso doméstico. Por tal motivo es esencial hacer una correcta administración de los sistemas de aguas subterráneas, en particular de los **acuiferos**. Esto significa tomar decisiones en cuanto a la ubicación de pozos de bombeo y en cuanto a las cantidades o volúmenes de agua a extraer, evitando un posible agotamiento del acuífero. Se trata entonces de optimizar la extracción de manera que no se rebasen ciertos límites.

De esta manera, se hace necesario el uso de herramientas que permitan conocer la respuesta de un sistema en explotación, ante diferentes alternativas; esto es por ejemplo, pronosticar los posibles cambios en los parámetros de un acuífero como consecuencia de distintos niveles de extracción o recarga. En este contexto, resultan de gran ayuda **modelos matemáticos** que describen el comportamiento de un sistema frente a una alteración producida en él.

Uno de los problemas que se plantean los ingenieros hidráulicos es conocer el nivel de agua en varios puntos de un acuífero dado. Sucede que en general la extensión de un acuífero es grande y la construcción de pozos de medición y bombeo es cara, por lo que sólo se construye una cantidad pequeña de ellos, que no resulta suficiente para estimar el comportamiento de todo el sistema. Para esto se plantea el modelo que describe el flujo de agua en un acuífero y en el que intervienen, entre otros parámetros, la carga hidráulica o altura de la columna de agua, la transmisividad, el coeficiente de almacenamiento y otros, cuya discusión se hará en el capítulo 1. Cuando se conocen los valores de éstos parámetros o coeficientes del modelo, se puede calcular la solución o respuesta del mismo, que en este caso es la carga hidráulica. Calcular dicha solución se conoce también como resolver el **problema directo** y en el capítulo 2 se describe uno de los métodos clásicos que se usan para esto.

Sin embargo, dichos coeficientes son difíciles o caros de medir por lo que se trata de estimarlos, conociendo mediciones de la carga en ciertos puntos y el rango en el que varían los mismos.

El planteamiento anterior deriva en un problema de **estimación de parámetros o problema inverso**, con características muy particulares. Dado que los modelos no siempre reflejan exactamente el comportamiento

del fenómeno en estudio y que la respuesta conocida del modelo es el resultado de mediciones, o sea que tiene cierto error, estos problemas pueden no tener solución. Esto unido a otras características que se describen en el capítulo 3, lo hacen un problema **mal planteado**.

Como objetivo principal del trabajo, desarrollado en los capítulos 3 y 4, está el planteamiento, la implementación y experimentación numérica de una alternativa llamada **optimización multiescala** dirigida a la solución correcta de problemas mal planteados con características como la del que nos ocupa. Para llevar a cabo esto se estudiaron los métodos de diferencias finitas para calcular la solución del modelo, así como las técnicas de **regularización de Tjonov** y las de multiescala. Se diseñó un problema de laboratorio que permitiera comprobar el funcionamiento de dichas técnicas, antes de aplicarlas en un problema real. De igual manera se simuló el error que generalmente, según ya mencionamos, presentan los datos de estos problemas y se trata en el capítulo 4.

Así se presentan en este trabajo los primeros resultados, obtenidos hasta ahora usando dicha técnica en la estimación de parámetros en modelos de acuíferos, para **datos sintéticos y datos reales**.

# CAPITULO 1

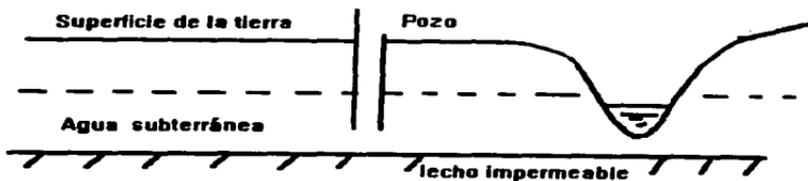
## Modelación Matemática

### 1.1 Aguas subterráneas. Acuíferos. Clasificación.

En todos los estudios hidrológicos, se considera como agua subterránea a aquella que se encuentra por debajo del nivel de saturación, en la llamada zona de saturación.

Como se muestra en la Figura 1.1, el agua se filtra por la superficie de la tierra. se mueve hacia abajo principalmente por efecto de la gravedad, a través de los intersticios del material rocoso o granular presente y se acumula por encima de una capa impermeable. Las características y formas que toman estas acumulaciones dependen directamente de las propiedades del material rocoso, por ejemplo sus intersticios, los cuales se agrupan en dos grandes clases: los originales -creados por procesos geológicos en el momento de formación de la roca- y los secundarios -en forma de fisuras desarrolladas después que la roca fue formada.

En el ámbito del agua subterránea los acuíferos son las formaciones geológicas más importantes y por lo tanto las más estudiadas. Se conoce como *acuífero* a una formación geológica con dos características fundamentales:



**Figura 1.1: Localización de un acuífero**

- contiene agua
- permite el flujo de ésta en cantidades significativas a través de él.

**Tipos de acuíferos de acuerdo con su sistema de presión:**

1. **Confinado:** se encuentra localizado entre dos capas de material impermeable, una por encima y otra por debajo. La recarga ocurre por flujo lateral.
2. **Artesiano:** es un acuífero confinado donde el nivel de presión alcanza o excede la superficie de la tierra, es decir, si se abre un pozo, el agua fluye libremente hasta la superficie sin necesidad de bombearla.
3. **Freático o no confinado:** es aquél donde la superficie freática, también llamada mesa de agua, constituye su frontera superior. En este caso la altura piezométrica o nivel de presión coincide con la superficie del agua en el subsuelo.
4. **Semiconfinado:** es aquél que presenta por encima o por debajo alguna zona semi-impermeable por donde puede ocurrir fuga o entrada de agua al mismo.

## 1.2 Leyes que rigen el movimiento de las aguas subterráneas.

Como punto de partida se tiene la conocida Ley de Darcy que describe el flujo de un líquido a través de un medio poroso en una columna vertical o filtro, indicando que el volumen de agua que fluye por unidad de tiempo es proporcional al área y a la diferencia de alturas (entrada-salida) e inversamente proporcional a la longitud recorrida. En general se escribe de la siguiente manera:

$$Q = \frac{K \cdot A (h_1 - h_2)}{L} \quad (1.1)$$

donde

- $Q$  = volumen de agua que fluye por unidad de tiempo ( $m^3/s$ )
- $\Delta h$  = diferencia entre los niveles de agua,  $\Delta h = h_1 - h_2$  ( $m$ )
- $L$  = longitud lineal de recorrido del flujo ( $m$ )
- $A$  = sección transversal de área ( $m^2$ )
- $K$  = conductividad hidráulica ( $m/s$ ) (véase más adelante)

Si se denomina  $\bar{q} = Q/A$  como el volumen de agua fluyendo por unidad de tiempo a través de una sección transversal de área normal a la dirección de flujo y a  $J = (h_1 - h_2)/L$  como gradiente hidráulico, podemos escribir (1.1) como

$$\bar{q} = KJ \quad (1.2)$$

y si en general el flujo se describe en tres dimensiones entonces  $\bar{q}$ , que indica la velocidad del flujo, será un vector tridimensional. al igual que el gradiente hidráulico, o sea

$$\bar{q} = \bar{q}(x, y, z) \quad (1.3)$$

En términos generales. el flujo en los acuíferos es predominantemente horizontal, de modo que al suponer que las superficies equipotenciales son verticales [3] entonces

$$\bar{q} = \bar{q}(x, y) \quad (1.4)$$

Además de la ley de Darcy, que expresa la cinemática del flujo, debe considerarse la ley de conservación de masa o de continuidad. Si

$$\bar{q} = \bar{q}_x e_1 + \bar{q}_y e_2 + \bar{q}_z e_3 \quad (1.5)$$

es el vector velocidad y  $\rho\bar{q}$  el flujo de masa por unidad de área, la variación espacial de esta propiedad es igual a la rapidez de variación de la masa de agua  $n\rho$  por unidad de volumen, es decir

$$\nabla \rho\bar{q} = \frac{\partial (n\rho)}{\partial t} \quad (1.6)$$

La variable  $n$  es la porosidad del medio pues, en efecto, no todo el volumen unitario contiene agua, solamente los poros o intersticios. Sustituyendo  $\bar{q}$  y aplicando el operador gradiente

$$\frac{\partial \rho\bar{q}_x}{\partial x} + \frac{\partial \rho\bar{q}_y}{\partial y} + \frac{\partial \rho\bar{q}_z}{\partial z} = \frac{\partial (n\rho)}{\partial t} \quad (1.7)$$

Si la densidad del flujo  $\rho$  se supone constante, entonces

$$\frac{\partial \bar{q}_x}{\partial x} + \frac{\partial \bar{q}_y}{\partial y} + \frac{\partial \bar{q}_z}{\partial z} = \frac{1}{\rho} \frac{\partial (n\rho)}{\partial t} \quad (1.8)$$

Al introducir los módulos de elasticidad del agua y del medio poroso y expresarlos en términos de la variación de la carga de hidráulica -  $h$  -, el segundo término de la ecuación (1.7) puede reescribirse como:

$$\frac{1}{\rho} \frac{\partial (n\rho)}{\partial t} = S_s \frac{\partial h}{\partial t} \quad (1.9)$$

donde  $S_s$  es el llamado coeficiente de almacenamiento específico.

Por otro lado, de acuerdo con la ley de Darcy, las componentes de velocidad del flujo se expresan como

$$\bar{q}_x = K_x J_x = K_x \frac{\partial h}{\partial x}, \quad \bar{q}_y = K_y J_y = K_y \frac{\partial h}{\partial y}, \quad \bar{q}_z = K_z J_z = K_z \frac{\partial h}{\partial z} \quad (1.10)$$

Tales expresiones se sustituyen a su vez en la ecuación (1.8) y con (1.9) obtener finalmente

$$\frac{\partial}{\partial x} \left( K_x \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( K_y \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left( K_z \frac{\partial h}{\partial z} \right) = S_s \frac{\partial h}{\partial t} \quad (1.11)$$

que se trata de una ecuación general, con la carga hidráulica o altura piezométrica  $h$  como única incógnita, la cual toma en cuenta simultáneamente las leyes de cantidad de movimiento (Ley de Darcy) y de conservación de masa (continuidad) del flujo de agua subterránea y es válida para cualquier condición y tipo de acuífero.

## 1.3 Conceptos geofísicos fundamentales.

Las propiedades generales de un acuífero para *transmitir, almacenar y retener* agua se definen mediante determinados parámetros como conductividad, transmisividad y otros.

- **Conductividad hidráulica:** es una magnitud escalar que se denota  $K$ , usualmente expresada en metros/día, que indica la facilidad con la cual un fluido es transportado a través de los intersticios de un medio poroso, resultando una propiedad combinada del medio y del fluido. Su expresión es

$$K = \frac{k\rho g}{\mu} \quad (1.12)$$

donde

$$\begin{aligned} g &= \text{aceleración de la gravedad } (m/s^2) \\ k &= \text{permeabilidad del medio poroso } (m^2) \\ \rho &= \text{densidad del fluido } (Kg/m^3) \\ \mu &= \text{viscosidad del fluido } (Kg/m.s) \end{aligned}$$

- **Transmisividad:** cuando el flujo de agua en un acuífero es esencialmente horizontal, la transmisividad, que se denota  $T$ , indica la facilidad del acuífero para dejar pasar el agua a través de su espesor. Es el producto de la conductividad hidráulica por el espesor  $b$  del acuífero y usualmente se expresa en ( $m^2/día$ )

$$T = Kb \quad (1.13)$$

Como ni  $b$  ni  $K$  son siempre constantes, entonces se considera  $T = T(x, y)$ .

- **Coefficiente de almacenamiento:** indica la relación entre los cambios en la cantidad de agua almacenada en el acuífero y los correspondientes cambios en las elevaciones de la carga hidráulica, y se denota  $S$ . En acuíferos confinados está definido como el volumen de agua extraída o añadida al acuífero por unidad horizontal de área. Es adimensional y su expresión es

$$S = \frac{\Delta V_a}{A \Delta h} \quad (1.14)$$

donde

$$\begin{aligned}\Delta V_a &= \text{variación del volumen de agua} \\ A &= \text{unidad horizontal de área} \\ \Delta h &= \text{variación de la carga hidráulica}\end{aligned}$$

Ahora bien, igual que se clasifican los acuíferos como se vió anteriormente, también se dividen en grupos según las características del medio y del fluido, lo que produce ecuaciones del tipo anterior pero con determinadas particularidades.

### Clasificación del medio poroso:

- **Anisotrópico:** cuando determinada propiedad del medio varía con respecto a la dirección, esto es, de manera espacial.
- **Isotrópico:** cuando determinada propiedad es independiente de la dirección, por ejemplo cuando el flujo de agua ocurre de igual manera en cualquier dirección.
- **Homogéneo:** cuando determinada propiedad es constante para todo el acuífero.
- **No homogéneo:** si la propiedad cambia con la posición.

En particular interesan estas clasificaciones con respecto a la propiedad de permeabilidad o conductividad en acuíferos confinados.

Supóngase que un acuífero confinado tiene un espesor  $b$ , constante y además que el medio está expuesto a una variación de sus volúmenes de agua por efecto de recargas (flujo fuente) o extracciones (flujo sumidero). Entonces, la ecuación (1.11) nos lleva a la forma general para acuíferos confinados :

$$\frac{\partial}{\partial x} \left( T_x \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T_y \frac{\partial h}{\partial y} \right) + \frac{\partial}{\partial z} \left( T_z \frac{\partial h}{\partial z} \right) + R - P = S \frac{\partial h}{\partial t} \quad (1.15)$$

donde  $T_x = K_x b$ ,  $T_y = K_y b$  y  $T_z = K_z b$  son las transmisividades en las tres direcciones cartesianas ( $m^2/día$ ) ;  $R$  y  $P$ , los volúmenes de recarga y

extracción por unidad de área (*m/día*) ; y  $S = S, b$  el coeficiente de almacenamiento (adimensional).

Si se tratara de un medio homogéneo e isotrópico, se tendría

$$T (\nabla^2 h) + R - P = S \frac{\partial h}{\partial t} \quad (1.16)$$

Cuando el flujo es estacionario y no hay variaciones temporales de la carga hidráulica, el segundo término se hace cero. Además, cuando no hay cambios por recarga o extracción, el problema se reduce a la ecuación de Laplace y desaparecen todos los coeficientes

$$\nabla^2 h = 0 \quad (1.17)$$

Por otra parte, además de la ecuación que describe el flujo en un acuífero, las condiciones iniciales y de frontera juegan un papel fundamental en la modelación correcta de cualquier fenómeno físico. Si  $\Omega$  es la región que abarca el acuífero, o sea sobre la que se resuelve (1.15), y  $\partial\Omega$  su frontera, se verá entonces la forma general en que se presentan dichas condiciones en los problemas de flujo y almacenamiento de agua subterránea.

- Condición inicial:

$$h(x, y, z, 0) = g(x, y, z) \quad \text{conocida}, \quad x, y, z \in \Omega \quad (1.18)$$

- Condiciones de frontera. Pueden ser:

a) De Dirichlet

$$h = f_1(x, y, z, t) \quad \text{en } \partial\Omega \quad (1.19)$$

es decir, se conoce la carga hidráulica y su variación temporal en la frontera del acuífero.

b) De Neumann

$$\bar{q} = f_2(x, y, z, t) \quad \text{en } \partial\Omega$$

o sea se conoce la velocidad del flujo (gasto caudal o variación temporal del volumen por unidad de área) en la frontera.

c) De tercer género (combinación lineal de las dos anteriores).

Como se observa todas las ecuaciones planteadas están consideradas en tres dimensiones. Sin embargo, debido a la geometría de la mayor parte de los acuíferos sucede que el grosor o espesor del mismo es relativamente pequeño en relación con su extensión horizontal, por lo que es posible simplificar el modelo.

Esto significa que el flujo en el acuífero es, en cualquier punto, esencialmente horizontal, de donde surge el término *flujo tipo acuífero*, es decir, se desprecia la componente vertical. Dicha propiedad es estrictamente cierta para acuíferos confinados, isotrópicos y homogéneos de grosor constante, aunque se supone como válida cuando se modela el flujo en otro tipo de acuíferos, o con variaciones en  $(x, y)$  de sus coeficientes. Cuando la geometría de un acuífero permite suponer flujo horizontal, se simplifica el modelo matemático al tener en cuenta sólo dos variables espaciales; sucede también que el error introducido por aceptar esta condición es pequeño en la mayoría de los casos de interés práctico. Una de las regiones donde no se puede considerar este tipo de flujo es en la vecindad de los pozos de extracción, en cuyo caso el análisis se reduce normalmente a la variación del fenómeno en las direcciones  $(x, z)$ , suponiendo que hay isotropía en la dirección  $y$  (suposición frecuente en el estudio de la hidráulica de pozos). Sin embargo no es de nuestro interés profundizar en este aspecto. En este trabajo se trata con el modelo que resulta de suponer flujo horizontal en un acuífero confinado, no homogéneo y anisotrópico en presencia de condiciones iniciales y de frontera de tal modo que

$$\frac{\partial}{\partial x} \left( T(x, y) \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T(x, y) \frac{\partial h}{\partial y} \right) = S \frac{\partial h}{\partial t} + Q \quad (1.20)$$

donde  $Q = P - R$  representa el volumen neto de extracción menos recarga vertical; con este término se procura la conservación de masa, aunque su

efecto no es considerado en la dinámica del fluido. Las condiciones (1.18) y (1.19) se plantean de la siguiente forma:

$$\begin{array}{ll} \text{condición inicial} & h(x, y, 0) = h_0(x, y) \quad x, y \in \Omega \\ \text{condiciones de frontera} & h(x, y, t) = \text{cte} \quad \text{en } \partial\Omega \\ \text{condiciones de frontera} & \bar{q}(x, y, t) = \text{cte} \quad \text{en } \partial\Omega \end{array}$$

En el capítulo siguiente se tratará la solución del mismo (problema directo) y del problema inverso que surge al no conocer el coeficiente de transmisividad.

# CAPITULO 2

## Problemas centrales.

### 2.1 Problema directo

Como se ha visto, el modelo de agua subterránea (1.20) presentado en el capítulo anterior está dado por una ecuación diferencial en derivadas parciales de tipo parabólico, con condiciones iniciales y de frontera.

Suponiendo que se conocen los coeficientes  $T$ ,  $S$  y  $Q$ , el *problema directo* consiste en obtener la función  $h(x, t)$  que satisface la ecuación, misma que se sabe está bien definida por la presencia de las condiciones mencionadas. Esto, haciendo uso de la notación operacional puede escribirse como

$$L(T, S, Q) = h \quad (2.1)$$

Interesa aquí el cálculo de soluciones numéricas y a continuación se presentan los elementos fundamentales de uno de los métodos clásicos para obtenerlas, el de diferencias finitas. Este método consiste en lo siguiente: la región de variación continua de las variables  $x$ ,  $t$  se sustituye por un conjunto finito de puntos (nodos), llamado red, por lo que las funciones que intervienen en la ecuación se consideran de argumento discreto, definidas en los nodos de la red; en consecuencia, las derivadas se sustituyen por cocientes de diferencias y en lugar de la ecuación diferencial se trabaja con un sistema de ecuaciones

algebraicas. De la misma manera se procede con las condiciones iniciales y de frontera.

### 2.1.1 Esquemas en diferencias

Se analizará brevemente el caso más sencillo de ecuación parabólica que corresponde a tomar en el modelo (1.20)  $h_t = h(x, t)$ ,  $S = 1$ ,  $T = T(x) = 1$ , y  $Q = 0$ .

Si se usa la notación siguiente

$$h_t \equiv \frac{\partial h}{\partial t}, \quad h_{xx} \equiv \frac{\partial^2 h}{\partial x^2} \quad (2.2)$$

y

$$h^t(x) \equiv h(x, t) \quad (2.3)$$

se tendrá entonces

$$h_t = h_{xx} \quad (2.4)$$

$$\begin{aligned} h(x, 0) &= h^0(x) & 0 \leq x \leq 1 \\ h(0, t) &= 0 & 0 \leq t \leq t_f \\ h(1, t) &= 0 & 0 \leq t \leq t_f \end{aligned}$$

El dominio o región donde se quiere calcular la solución es  $\Omega = [0, 1] \times [0, t_f]$  y el primer paso es discretizar  $\Omega$  como se ve en la Figura 2.1.

Sea  $N_x$  el número de subintervalos en los que se desea dividir el intervalo  $[0, 1]$  y  $N_t$  el número de subintervalos en los que se desea dividir el intervalo  $[0, t_f]$ , entonces si

$$\begin{aligned} \Delta x &= \frac{1}{N_x} & \text{incremento en } x \\ \Delta t &= \frac{t_f}{N_t} & \text{incremento en } t \end{aligned}$$

los puntos de la red serán de la forma  $(x_i, t_j)$ , donde

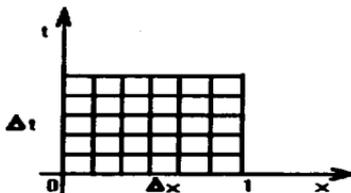


Figura 2.1: Región discretizada

$$\begin{aligned} x_i &= i\Delta x & i &= 0, 1, \dots, Nx \\ t_j &= j\Delta t & j &= 0, 1, \dots, Nt \end{aligned}$$

y al valor de la solución  $h$  del problema (2.4) en un punto  $(x_i, t_j)$  de la red se denotará como

$$h_i^j = h(x_i, t_j) \quad (2.5)$$

### Esquema explícito.

Al aproximar  $h_t$  usando diferencias del tipo

$$h_t(x_i, t_j) \approx \frac{h(x_i, t_{j+1}) - h(x_i, t_j)}{\Delta t} \quad i = 0, 1, \dots, Nx, \quad j = 0, 1, \dots, Nt \quad (2.6)$$

llamadas *diferencias hacia adelante* y aproximar  $h_{xx}$  usando diferencias del tipo

$$h_{xx}(x_i, t_j) \approx \frac{h(x_{i+1}, t_j) - 2h(x_i, t_j) + h(x_{i-1}, t_j))}{(\Delta x)^2} \quad (2.7)$$

llamadas *segundas diferencias centradas*, se obtiene, sustituyendo en (2.4), el siguiente esquema en diferencias, llamado *Explícito*, que constituye la expresión discretizada de dicho problema

$$h_i^{j+1} - h_i^j = \frac{\Delta t}{(\Delta x)^2} (h_{i+1}^j - 2h_i^j + h_{i-1}^j) \quad (2.8)$$

con

$$\begin{aligned} h_i^0 &= h^0(x_i) \quad i = 0, 1, \dots, Nx \\ h_0^j &= h_{Nx}^j = 0 \quad j = 0, 1, \dots, Nt \end{aligned}$$

Un aspecto muy importante a considerar en cuanto a la robustez de un método numérico es la propiedad de *estabilidad*. Veamos esto.

Sea  $h_i^n$  el vector solución en el paso de tiempo  $n$ , obtenido usando (2.8) con  $h_i^0$  como valor inicial y sea  $p_i^n$  el vector solución en el paso de tiempo  $n$ , obtenido usando (2.8) con  $p_i^0$  como valor inicial, con  $h_i^0 \neq p_i^0$ . Se dice que el esquema dado por (2.8) es **estable** en la norma  $\|\cdot\|$  si existe una constante  $M$  tal que

$$\|h_i^n - p_i^n\| \leq M \|h_i^0 - p_i^0\|$$

En [20] se muestran algunos resultados numéricos obtenidos al aplicar este esquema y se observa que dependen críticamente del valor de

$$\nu = \frac{\Delta t}{(\Delta x)^2} \quad (2.9)$$

es decir, de la relación entre el tamaño del paso por el tiempo y el tamaño del paso en el espacio. Haciendo uso de las series de Fourier para expresar la solución exacta de la ecuación diferencial y de la solución de la ecuación en diferencias, se analiza el error de truncamiento de dicho esquema y se llega a que (2.8) es estable si

$$\nu \leq \frac{1}{2}$$

e inestable si ocurre lo contrario. En la Figura 2.2a se muestran los puntos de la red que intervienen en este esquema, lo cual nos permite observar que basta con sustituir el valor conocido de  $h$  en la capa de tiempo  $j$  para obtener el de la capa  $j + 1$ , pero la condición de estabilidad es bastante restrictiva ya que se necesitaría dar muchos pasos en el tiempo, esto es  $\Delta t$  muy pequeña, si se desea mejorar la aproximación de una solución disminuyendo  $\Delta x$ . Por este motivo se han desarrollado otros esquemas que no presenten esta dificultad y que describimos a continuación.

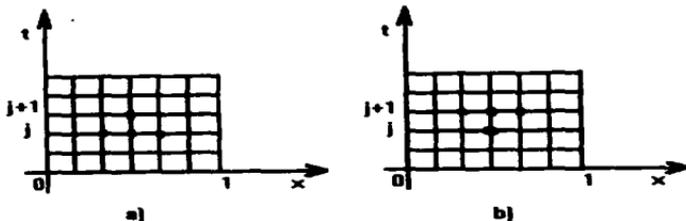


Figura 2.2: a) Esquema explícito b) Esquema implícito

### Esquema implícito.

Otra forma surge cuando se aproxima  $h_{xx}$  también usando segundas diferencias centradas correspondiente a la capa de tiempo  $j + 1$ , y aproximando  $h_t$  con diferencias del tipo

$$h_t(x_i, t_j) \approx \frac{h(x_i, t_j) - h(x_i, t_{j-1})}{\Delta t} \quad i = 0, 1, \dots, Nx, \quad j = 0, 1, \dots, Nt \quad (2.10)$$

llamadas *diferencias hacia atrás*. Se obtiene el siguiente esquema llamado *Implícito* el cual es incondicionalmente estable, esto es, no se tiene una restricción del tipo (2.9).

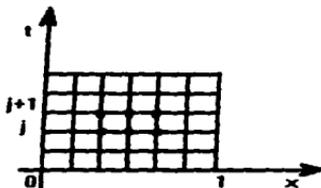
$$h_i^{j+1} = h_i^j + \frac{\Delta t}{(\Delta x)^2} (h_{i+1}^{j+1} - 2h_i^{j+1} + h_{i-1}^{j+1}) \quad (2.11)$$

$$\begin{aligned} h_i^0 &= h^0(x_i) & i &= 0, 1, \dots, Nx \\ h_0^j &= h_{Nx}^j = 0 & j &= 0, 1, \dots, Nt \end{aligned}$$

En la Figura 2.2b se muestran los puntos de la red que intervienen en este esquema.

Como se ve, aplicar el esquema explícito significa sustituir directamente en (2.8), mientras que el implícito no funciona así, pues para obtener la solución de la ecuación hay que resolver un sistema de ecuaciones lineal y tridiagonal. La matriz de este sistema es de diagonal predominante lo que asegura la estabilidad numérica del algoritmo de *Thomas* [20], utilizado generalmente para resolver dicho sistema.

La importancia del esquema implícito está dada por el hecho de que el paso por el tiempo puede ser mucho más grande que el paso por el espacio,



**Figura 2.3:** Esquema dependiente de un parámetro

sin que esto signifique inestabilidad en el proceso de cálculo de la solución de (2.3). Esto hace que dicho método sea muy usado a pesar de que en la medida en que la red se haga más fina, aumenta considerablemente el orden del sistema de ecuaciones ya mencionado.

### 2.1.2 Esquemas dependientes de un parámetro.

De lo anterior, resulta natural proponer una combinación de ambos esquemas de manera que no se presenten condiciones de estabilidad demasiado restrictivas y al mismo tiempo el proceso no se haga muy complicado. Una forma es usando seis puntos de la red para cada aproximación como se ve en la Figura 2.3, obteniendo la siguiente familia de esquemas dependientes del parámetro  $\theta$ , ( $0 \leq \theta \leq 1$ )

$$\frac{h_i^{j+1} - h_i^j}{\Delta t} = \frac{\theta (h_{i+1}^{j+1} - 2h_i^{j+1} + h_{i-1}^{j+1}) + (1 - \theta) (h_{i+1}^j - 2h_i^j + h_{i-1}^j)}{(\Delta x)^2} \quad (2.12)$$

la que incluye como casos particulares el explícito ( $\theta = 0$ ) y el implícito ( $\theta = 1$ ). Como se mencionó anteriormente, usando análisis de Fourier, se demuestran en [20] las siguientes condiciones necesarias y suficientes para garantizar la estabilidad de (2.12) con distintos valores de  $\theta$ :

- Si  $0 \leq \theta < \frac{1}{2}$ , se cumple que (2.12) es estable si y sólo si  $\nu \leq \frac{1}{2} (1 - 2\theta)^{-1}$ .
- Si  $\frac{1}{2} \leq \theta \leq 1$ , se cumple que (2.12) es estable para todo valor de  $\nu$ .

Al analizar el error de truncamiento, usando desarrollo en serie de Taylor, se ve en [20] que para  $\theta = \frac{1}{2}$  se obtiene un mejor orden de aproximación de la solución que con cualquier otro valor. Esto dió lugar al conocido esquema de Crank-Nicolson (CN), aplicado por estos autores con resultados muy satisfactorios, y que resulta de usar (2.12) con ese valor de  $\theta$ .

Sucede entonces que cuando se intenta aplicar estos esquemas - por ejemplo el de (C-N)- a problemas en dos dimensiones espaciales como el que nos ocupa, éstos se hacen computacionalmente muy laboriosos y se desarrollan esquemas que aprovechan la eficiencia de los métodos implícitos en una dimensión. Uno de ellos es el:

### Método Implícito de Direcciones Alternadas.

La idea central de éste es considerar la red

$$W = \{(x_i, t_j) \text{ , } x_i = i\Delta x, i = 0, 1, \dots, Nx, y_j = j\Delta t, j = 0, 1, \dots, Ny\} \quad (2.13)$$

como un conjunto de nodos a lo largo de las filas  $j = 0, 1, \dots, Nt$  o como un conjunto de nodos a lo largo de las columnas  $i = 0, 1, \dots, Nx$ , entonces en cada fila (columna) se resuelve un sistema tridiagonal manteniendo fijo a  $j$  ( $i$ ). Esto es equivalente a aplicar el esquema implícito en una dimensión espacial y no en ambas al mismo tiempo.

Una variante de este método es la desarrollada por Peaceman y Rachford ([20]) motivada por el esquema de C-N y que denotando  $h_{i,j}^n = h(x_i, y_j, t_n)$ , se plantea del siguiente modo:

$$\begin{aligned} h_{i,j}^{n+1} - \frac{1}{2} \frac{\Delta t}{(\Delta y)^2} (h_{i,j+1}^{n+1} - 2h_{i,j}^{n+1} + h_{i,j-1}^{n+1}) - \frac{1}{2} \frac{\Delta t}{(\Delta x)^2} (h_{i+1,j}^{n+1} - 2h_{i,j}^{n+1} + h_{i-1,j}^{n+1}) + \\ \frac{(\Delta t)^2}{4(\Delta x)^2(\Delta y)^2} (h_{i,j+1}^{n+1} - 2h_{i,j}^{n+1} + h_{i,j-1}^{n+1}) (h_{i+1,j}^{n+1} - 2h_{i,j}^{n+1} + h_{i-1,j}^{n+1}) = \\ h_{i,j}^n + \frac{1}{2} \frac{\Delta t}{(\Delta y)^2} (h_{i,j+1}^n - 2h_{i,j}^n + h_{i,j-1}^n) + \frac{1}{2} \frac{\Delta t}{(\Delta x)^2} (h_{i+1,j}^n - 2h_{i,j}^n + h_{i-1,j}^n) + \end{aligned}$$

$$\frac{(\Delta t)^2}{4(\Delta x)^2(\Delta y)^2} (h_{i,j+1}^n - 2h_{i,j}^n + h_{i,j-1}^n) (h_{i+1,j}^n - 2h_{i,j}^n + h_{i-1,j}^n) \quad (2.14)$$

Usando una capa intermedia de tiempo se puede escribir el esquema anterior de forma más clara, teniendo que

$$h_{i,j}^{n+\frac{1}{2}} - \frac{1}{2} \frac{\Delta t}{(\Delta x)^2} (h_{i+1,j}^{n+\frac{1}{2}} - 2h_{i,j}^{n+\frac{1}{2}} + h_{i-1,j}^{n+\frac{1}{2}}) = h_{i,j}^n + \frac{1}{2} \frac{\Delta t}{(\Delta y)^2} (h_{i,j+1}^n - 2h_{i,j}^n + h_{i,j-1}^n) \quad (2.15)$$

y

$$h_{i,j}^{n+1} - \frac{1}{2} \frac{\Delta t}{(\Delta y)^2} (h_{i,j+1}^{n+1} - 2h_{i,j}^{n+1} + h_{i,j-1}^{n+1}) = h_{i,j}^{n+\frac{1}{2}} + \frac{1}{2} \frac{\Delta t}{(\Delta x)^2} (h_{i+1,j}^{n+\frac{1}{2}} - 2h_{i,j}^{n+\frac{1}{2}} + h_{i-1,j}^{n+\frac{1}{2}}) \quad (2.16)$$

Como en (2.15) los elementos del término de la derecha son conocidos del paso anterior se pueden calcular los  $h_{i,j}^{n+\frac{1}{2}}$  de la capa intermedia sin mayor problema, y luego se construye el término de la derecha de (2.16) y se obtiene  $h_{i,j}^{n+1}$  en la siguiente capa de tiempo. Es necesario señalar que cuando se trabaja con dos o más dimensiones surge otro problema importante, relacionado con la complejidad de la región en la que se resuelve la ecuación diferencial. El tratamiento de las condiciones de frontera en estos casos complica un poco más el proceso de cálculo, ver [20], pero no vamos a profundizar en este aspecto.

Siguiendo entonces el método implícito de direcciones alternadas, el proceso de solución de la ecuación es mucho más rápido que si se aplica el método de C-N tal cual. Esta variante propuesta por Peaceman y Rachford es la que está programada en el código que utilizamos para resolver el problema directo (2.1) y se dan otros detalles al respecto en el capítulo de implementación computacional.

## 2.2 Problema inverso

La transmisividad es un parámetro que físicamente se puede medir, pero el proceso para hacerlo implica la construcción de pozos específicamente para eso y la interrupción de operaciones de bombeo en todas las zonas aledañas al punto donde se va a medir. Esto lo hace muy costoso y por tanto casi imposible contar con esos valores. Por este motivo es usual que en el estudio del flujo y almacenamiento de agua en un acuífero no se conozcan los valores de los coeficientes  $T$ ,  $S$  y ocasionalmente  $Q$  sino observaciones -  $\hat{h}$  - de la carga hidráulica, que sí se puede medir en los pozos de bombeo sin gran complicación. Así lo que sucede es que se conoce la solución de (1.20) en algunos puntos de la región  $\Omega$ . Si se considera el caso  $S = 1$  y  $Q = 0$ , se tiene el siguiente problema:

Dada  $\hat{h}$ , determinar  $\hat{T}$ , tal que

$$L(\hat{T}) = \hat{h} \quad (2.17)$$

que se conoce como *problema inverso* y constituye el tema central de nuestro trabajo.

Como en general  $\hat{h}$  tiene algún error o bien el modelo (1.20) no corresponde exactamente con el fenómeno físico en estudio, resulta que (2.17) podrá no tener solución. Una alternativa usual en este caso es estimar  $\hat{T}$  como el valor de  $T$  que resuelve

$$\min_{\hat{T}} F(T), \quad F(T) = \|L(T) - \hat{h}\|^2 \quad (2.18)$$

Planteado así, la ecuación (2.17) se traduce en un problema de optimización con las siguientes características :

- Función objetivo de mínimos cuadrados no lineales.

- Restricciones de acotación sobre las variables, dadas por la física de cada aplicación.
- Crecimiento considerable de la dimensión en la medida en que se haga más fina la red  $W$ .

Existen varios algoritmos eficientemente implementados, para resolver diferentes problemas de optimización, de acuerdo con la función objetivo, sea o no lineal, presente o no restricciones, etcétera. En particular, para un problema como el que nos ocupa, son muy usados los métodos de optimización de tipo Quasi-Newton que combinan una estrategia globalmente convergente con la propiedad de convergencia local cuadrática del método de Newton [12].

Se han desarrollado también variantes de tales métodos, para hacerlos más eficientes y para ocupar menos memoria durante el procesamiento, sobre todo cuando la dimensión del problema es muy grande. Una de éstas es la basada en el método de Memoria Limitada [17], el cual considera además la presencia de cotas sobre las variables, que es nuestro caso; por ello en la experimentación computacional de este trabajo se ha usado un software desarrollado por Zhu y Nocedal de un método Quasi-Newton con memoria limitada para un problema de gran escala y con restricciones de acotación, del cual describimos a continuación sus características más importantes.

## 2.2.1 Método de Newton.

El método se desarrolla primeramente para resolver el problema siguiente:

Dada la función  $F : R^n \rightarrow R^n$  continuamente diferenciable, encontrar un punto  $x^* \in R^n$  tal que

$$F(x^*) = 0 \tag{2.19}$$

La idea es trabajar con una aproximación afín de  $F$  en cada iteración y generar la sucesión de puntos siguiente:

Dado  $x_0 \in R^n$ , para  $k = 0, 1, \dots$  hasta convergencia

**Paso 1 :** Resolver el sistema

$$J(x_k) s_k = -F(x_k) \quad (2.20)$$

donde  $J$  es el jacobiano de  $F$ .

**Paso 2 :**

$$x_{k+1} = x_k + s_k \quad (2.21)$$

En [12] se demuestra, con el siguiente resultado, que bajo ciertas condiciones la sucesión  $\{x_i\}$  converge a  $x^*$ , cuadráticamente, si  $x_0$  está en una vecindad de  $x^*$ .

**Teorema 2.2.1** Sea  $F : R^n \rightarrow R^n$  continuamente diferenciable en un conjunto abierto y convexo  $D \subset R^n$ . Sea  $N(x, r)$  una vecindad de radio  $r$  con centro en  $x$ . Supóngase que existen  $x^* \in R^n$  y  $r, \beta > 0$ , tales que  $N(x^*, r) \subset D$ ,  $F(x^*) = 0$ ,  $J(x^*)^{-1}$  existe con  $\|J(x^*)^{-1}\| \leq \beta$ , y que  $J$  cumple la condición de *Lipschitz* en  $N$  con constante  $\gamma$ . Entonces existe  $\varepsilon > 0$  tal que  $\forall x_0 \in N(x^*, \varepsilon)$  la sucesión  $x_1, x_2, \dots$  generada por (2.21) está bien definida, converge a  $x^*$  y cumple  $\|x_{k+1} - x^*\| \leq \beta\gamma \|x_k - x^*\|$ .

Ahora, cuando se tiene un problema de optimización del tipo

$$\min_{x \in R^n} f(x), \quad f : R^n \rightarrow R \quad (2.22)$$

donde  $f$  es dos veces continuamente diferenciable, se sabe que una condición necesaria para que  $x^*$  sea solución de (2.22) es que  $\nabla f(x^*) = 0$ . Dado que  $\nabla f(x) = 0$  resulta ser un sistema de ecuaciones es natural aplicar las ideas de Newton que dan lugar al esquema

Dado  $x_0 \in R^n$ , para  $k = 0, 1, \dots$  hasta convergencia

**Paso 1 :** Resolver el sistema

$$H(x_k) s_k = -\nabla f(x_k) \quad (2.23)$$

donde  $H(x_k) = \nabla^2 f(x_k)$ , es la Hessiana de  $f$ .

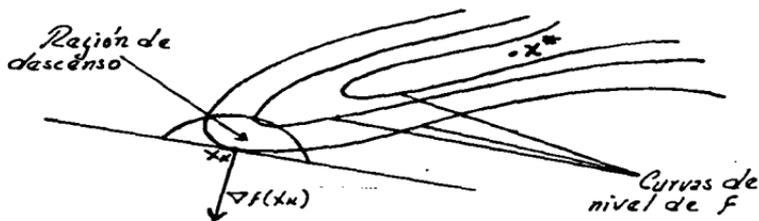


Figura 2.4: Región de descenso

**Paso 2 :**

$$x_{k+1} = x_k + s_k \quad (2.24)$$

De lo anterior, se tiene la misma propiedad de convergencia; sin embargo la implementación requiere tener las expresiones para  $\nabla f$  y  $H$  y además evaluar éstas para resolver (2.23) en cada iteración. Todo esto impone severas restricciones pues, en general :

1. Se depende de una buena aproximación inicial  $x_0$ ; esto es, no se trata de un método con convergencia global.
2. Puede no disponerse de  $\nabla f$  y  $H$  y aún en caso de tenerlos, su evaluación y la resolución del sistema (2.23) en cada paso puede resultar excesivamente costosa.

En la siguiente subsección se abordarán tales cuestiones.

### 2.2.2 Métodos Quasi-Newton.

Con respecto al problema de la convergencia, la idea básica es determinar en cada paso una  $s_k$  que sea de *descenso*. Geométricamente esto puede verse en la zona sombreada de la Figura 2.4.

De aquí, cualquier  $d_k$  que satisfaga

$$\nabla^T f(x_k) d_k < 0 \quad (2.25)$$

es una dirección que permitirá cumplir

$$f(x_k + \lambda d_k) < f(x_k) \quad (2.26)$$

para un cierto  $\lambda > 0$ .

Una vez determinado el valor de  $\lambda$ , digamos  $\lambda_k$ , un esquema general es hacer

$$x_{k+1} = x_k + \lambda_k d_k \quad (2.27)$$

Es importante notar que la dirección de Newton puede escribirse en función de la matriz inversa,  $H_k^{-1}$ , de la hessiana  $H_k$

$$s_k = -H_k^{-1} \nabla f(x_k), \quad H_k = H(x_k) \quad (2.28)$$

será de descenso si satisface

$$\nabla^T f(x_k) s_k = -\nabla^T f(x_k) H_k^{-1} \nabla f(x_k) < 0 \quad (2.29)$$

y esto implica el que  $H_k^{-1}$  o  $H_k$  sea definida positiva.

La idea general de los métodos Quasi-Newton se sustenta en buscar estrategias que no estén sujetas a la elección de un  $x_0$  necesariamente cercano al mínimo  $x^*$  de  $f$  y que al mismo tiempo busquen conservar la propiedad de convergencia de Newton en una vecindad de  $x^*$ .

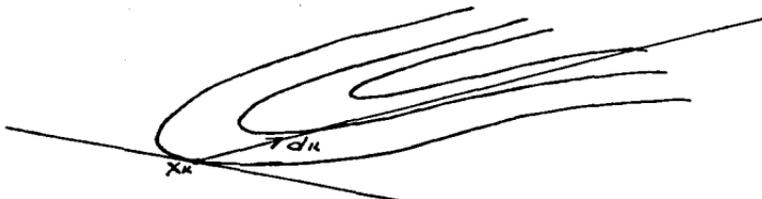
Continuando en este orden de ideas, una propuesta es:

Dada  $x_k$ , determinar  $d_k$  como

$$d_k = \begin{cases} s_k \text{ (dirección de Newton)} & \text{si } H_k \text{ es definida positiva} \\ -\widetilde{H}_k^{-1} \nabla f(x_k) & \text{en otro caso} \end{cases}$$

donde

$$\widetilde{H}_k = H_k + \mu_k I, \text{ y } \mu_k \text{ elegida de tal forma que } \widetilde{H}_k > 0$$



**Figura 2.5: Cálculo del nuevo punto**

A esta elección de  $d_k$  se le conoce como dirección Quasi-Newton.

Una vez fijada la dirección  $d_k$ , la cuestión consiste en determinar la correspondiente  $\lambda_k$  en (2.27) de tal forma que el nuevo punto resulte aceptable como nueva aproximación de  $x^*$ , como se puede apreciar en la Figura 2.5.

Resulta natural, en principio, pensar en la condición

$$f(x_{k+1}) < f(x_k) \quad (2.30)$$

sin embargo, pueden darse ejemplos [12] en los que se muestra que ésta no es suficiente. Un esquema general para nuestros propósitos es el siguiente :

Dadas  $x_k$  y  $d_k$  de descenso

- Calcular  $\lambda_k$  que satisfaga

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \alpha \lambda \nabla^T f(x_k) d_k, \alpha \in (0, 1) \quad (2.31)$$

y

$$\nabla^T f(x_k + \lambda_k d_k) d_k \geq \beta \nabla^T f(x_k) d_k, \beta \in (\alpha, 1) \quad (2.32)$$

- Hacer  $x_{k+1} = x_k + \lambda_k d_k$ .

Las condiciones dadas en (2.31) y (2.32) están basadas en Armijo y Goldstein y tomadas conjuntamente establecen un equilibrio entre la magnitud del descenso en  $f$  de  $x_{k+1}$  a  $x_k$  y  $\|x_{k+1} - x_k\|$ . En [12], Wolfe prueba que, dada

$x_k \in R^n$  y  $d_k$  de descenso, es posible determinar  $\lambda_k$  de tal forma que la iteración  $x_{k+1} = x_k + \lambda_k d_k$  satisfaga las propiedades (2.31) y (2.32). Además se prueba el siguiente teorema:

**Teorema 2.2.2:** Sea  $f : R^n \rightarrow R$  continuamente diferenciable en  $R^n$ , y supóngase que existe  $\gamma \geq 0$  tal que

$$\|\nabla f(y) - \nabla f(x)\|_2 \leq \gamma \|y - x\|_2, \quad \text{para } x, y \in R^n$$

Entonces dado cualquier  $x_0 \in R^n$ ,  $f$  no está acotada inferiormente o existe una sucesión  $\{x_k\}$  que satisfice (2.31) y (2.32). También se cumple una de las siguientes propiedades

$$\nabla^t f(x_k) d_k < 0$$

o

$$\nabla f(x_k) = 0 \text{ y } d_k = 0 \text{ para cada } k \geq 0, \text{ donde } d_k = x_{k+1} - x_k$$

Además, para toda sucesión como esa, se cumple que:

- $\nabla f(x_k) = 0$ , para algún  $k \geq 0$ , o
- $\lim_{k \rightarrow \infty} f(x_k) = \infty$ , o
- $\lim_{k \rightarrow \infty} \frac{\nabla f(x_k)^t d_k}{\|d_k\|_2} = 0$ .

Dennis y Moré, con el siguiente teorema, prueban además un resultado en el que se garantiza convergencia superlineal.

**Teorema 2.2.3:** Sea  $f : R^n \rightarrow R$  dos veces continuamente en un conjunto convexo abierto  $D$ , y supóngase que  $\nabla^2 f$  cumple la condición de Lipschitz en  $D$  con parámetro  $\gamma$ . Considere una sucesión  $\{x_k\}$  generada por (2.27), donde  $\nabla^T f(x_k) d_k < 0 \forall k$  y el  $\lambda_k$  seleccionado satisfice (2.31) y (2.32). Si  $\{x_k\}$  converge a un punto  $x^* \in D$  para el cual  $\nabla^2 f(x^*)$  es definida positiva, y si

$$\lim_{k \rightarrow \infty} \frac{\|\nabla f(x_k) + \nabla^2 f(x_k) d_k\|_2}{\|d_k\|_2} = 0$$

entonces existe un índice  $k_0 \geq 0$  tal que  $\forall k \geq k_0$ , es permisible tomar  $\lambda_k = 1$ . Además  $\nabla f(x^*) = 0$ , y si  $\lambda_k = 1 \forall k \geq k_0$ , entonces  $\{x_k\}$  converge a  $x^*$  *q-superlinealmente*.

En resumen, si se tiene  $f$  acotada inferiormente, entonces la sucesión  $\{x_k\}$  generada con las condiciones

1. ángulo  $\langle d_k, \nabla f(x_k) \rangle > 90^\circ$
2.  $f(x_k + \lambda_k d_k) \leq f(x_k) + \alpha \lambda \nabla^T f(x_k) d_k$ ,  $\alpha \in (0, 1)$
3.  $\nabla^T f(x_k + \lambda_k d_k) d_k \geq \beta \nabla^T f(x_k) d_k$ ,  $\beta \in (\alpha, 1)$

satisfará

$$\lim_{k \rightarrow \infty} \nabla f(x_k) = 0$$

Más aun, si el algoritmo en cuestión intenta siempre como primera opción tomar la dirección de Newton o la de Quasi-Newton para cada iteración, entonces la sucesión generada converge *q-cuadráticamente* o *q-superlinealmente* a  $x^*$  si  $x_k$  está suficientemente cerca de  $x^*$  y se satisfacen las condiciones de convergencia local de Newton o Quasi-Newton respectivamente.

Respecto al punto 2, en buen número de casos se da la necesidad de encontrar buenas aproximaciones de  $H$  dado el caso de que no se tenga su expresión analítica o que su evaluación resulte muy cara.

Primeramente debemos analizar las ideas surgidas al estudiar el caso (2.19). Una de ellas se basa en aproximar  $J(x_{k+1})$  en (2.20) usando solamente información de la iteración anterior, siguiendo las ideas del método de la secante en una dimensión ( $n = 1$ ). Esto derivó en el método de Broyden, descrito en [12], que propone tomar

$$J_{k+1} = J_k + \frac{(y_k - J_k s_k) s_k^T}{s_k^T s_k} \quad (2.33)$$

donde  $J_k = J(x_k)$ ,  $y_k = F(x_{k+1}) - F(x_k)$ ,  $s_k = x_{k+1} - x_k$ . Al incorporarlo a (2.20) y (2.21) se tiene

Dados  $x_0 \in R^n$ ,  $J_0 \in R^{n \times n}$

Para  $k = 0, 1, \dots$  hasta convergencia

Resolver  $J(x_k) s_k = -F(x_k)$

$$\begin{aligned}
 x_{k+1} &= x_k + s_k \\
 y_k &= F'(x_{k+1}) - F'(x_k) \\
 J_{k+1} &= J_k + \frac{(y_k - J_k s_k) s_k^T}{s_k^T s_k}
 \end{aligned}$$

Se puede demostrar que esta variante del método de Newton tiene convergencia q-superlineal [12].

Al extender esto al problema de optimización (2.22) se presentan dos inconvenientes si se trata de usar (2.33) para aproximar  $H_{k+1}$  como

$$H_{k+1} = H_k + \frac{(y_k - H_k s_k) s_k^T}{s_k^T s_k} \quad (2.34)$$

pues puede suceder que  $H_{k+1}$  no sea ni simétrica ni definida positiva, aunque  $H_k$  lo sea. En este caso  $y_k = g_{k+1} - g_k$ , donde  $g_k = \nabla f(x_k)$ .

La simetría es una característica propia de toda matriz Hessiana y ya vimos la importancia que tiene que sea definida positiva, por lo que se han hecho variaciones sobre (2.34) para lograr que si  $H_k$  es simétrica y definida positiva,  $H_{k+1}$  también lo sea, asegurando así que si se parte de una  $H_0$  inicial con esas propiedades, éstas se mantengan en cada iteración.

Con ese objetivo trabajaron Broyden, Fletcher, Goldfarb y Shanno y descubrieron, independientemente cada uno, la conocida fórmula para aproximar  $H_{k+1}$

$$H_{k+1} = H_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} \quad (2.35)$$

conocida como la actualización BFGS, vista también en [12].

Si nos fijamos en (2.28), vemos que se puede usar también la correspondiente actualización de la inversa de  $H_{k+1}$ , la que se puede obtener a partir de (2.35) como se ve en [12]. Denotemos como  $\bar{H}_{k+1}$  a dicha aproximación

de la inversa de la hessiana en la iteración  $k + 1$ . Nótese que no es necesario formar esta matriz explícitamente, sino hacer el producto  $\bar{H}_{k+1}g_k$  mediante una sucesión de productos interiores entre  $g_k$  y los vectores  $s_i, y_i$  para  $i = 0, 1, \dots, k$ . Veamos esto con más detalle.

Como aparece en [21]  $\bar{H}_{k+1}$  se puede escribir en forma producto de la siguiente manera:

$$\bar{H}_{k+1} = V_k^t \bar{H}_k V_k + \rho_k s_k s_k^t \quad (2.36)$$

donde

$$\rho_k = \frac{1}{y_k^t s_k}, \quad V_k = I - \rho_k y_k s_k^t \quad (2.37)$$

Sin embargo, esto genera un problema serio cuando la dimensión del problema es muy grande ya que en la  $k$ -ésima iteración se necesita tener almacenados  $2k$  vectores de  $R^n$  para aplicar (2.36). Con el fin entonces de trabajar estos métodos buscando ahorro de memoria se han desarrollado los llamados de memoria limitada. Estos son del tipo Quasi-Newton con la variante de que se fija de entrada una cota para la cantidad de pares de vectores  $s_i, y_i$  que se van a considerar en las actualizaciones.

Sea  $m \ll n$  el número máximo de estos pares que se van a almacenar. Entonces, mientras  $k + 1 \leq m$  no hay ningún cambio en (2.36), lo cual se ve mejor si desarrollamos la forma producto

$$\begin{aligned} \bar{H}_{k+1} &= V_k^t V_{k-1}^t \cdots V_0^t \bar{H}_0 V_0 \cdots V_{k-1} V_k \\ &\quad + V_k^t \cdots V_1^t \rho_0 s_0 s_0^t V_1 \cdots V_k \\ &\quad \vdots \\ &\quad + V_k^t V_{k-1}^t \rho_{k-2} s_{k-2} s_{k-2}^t V_{k-1} V_k \\ &\quad + \rho_k s_k s_k^t. \end{aligned}$$

Cuando  $k + 1 > m$  se desecha la información de  $s_0$  y  $y_0$  y se guarda en su lugar la de  $s_{k+1}$  y  $y_{k+1}$ , y así sucesivamente se va eliminando la información más vieja y se dejan los  $m$  pares de vectores previos a la iteración actual. La forma producto queda como sigue

$$\begin{aligned} \bar{H}_{k+1} &= V_k^t V_{k-1}^t \cdots V_{k-m+1}^t \bar{H}_0 V_{k-m+1} \cdots V_{k-1} V_k \\ &\quad + V_k^t \cdots V_{k-m+2}^t \rho_{k-m+1} s_{k-m+1} s_{k-m+1}^t V_{k-m+1} V_{k-m+2} \cdots V_k \\ &\quad \vdots \\ &\quad + V_k^t \rho_{k-1} s_{k-1} s_{k-1}^t V_k \\ &\quad + \rho_k s_k s_k^t. \end{aligned}$$

Como se ve es muy importante la selección de la aproximación inicial  $\tilde{H}_0$ . En muchos casos se usa la matriz identidad y en otros ciertos escalamientos de la matriz identidad, pero no entraremos a especificar más en esta dirección.

El problema de optimización que se requiere resolver en este trabajo puede llegar a ser de gran dimensión pero además, como ya hemos mencionado, presenta restricciones sobre las variables (parámetros). Incorporar en un código de optimización la posibilidad de considerar restricciones en los problemas es algo bastante complicado y muy delicado por lo que decidimos utilizar la rutina LBFGSB [28] para nuestro caso. Este programa es uno de los más robustos y eficientes con que se cuenta en esta área de trabajo y está diseñado precisamente para resolver problemas de optimización no lineal con restricciones de acotación, como las del tipo que aquí se presentan. Las ideas de memoria limitada, antes descritas y que se usan en dicha rutina, se pueden encontrar en [17].

Cuando se supone que las observaciones  $\hat{h}$  no tienen errores, este algoritmo resuelve el problema (2.18) correctamente. Algunos ejemplos numéricos de esto se muestran en el capítulo 4 para un ejemplo de laboratorio. Pero cuando hay perturbaciones  $\delta$  en  $\hat{h}$ , como es usual que suceda por ser mediciones reales de campo, entonces la solución que nos devuelve LBFGSB para el problema

$$\min_T F(T) = \|L(T) - \hat{h}\|^2 \quad (2.38)$$

puede estar muy lejos de ser la correcta, es decir, no resulta suficiente aplicar este algoritmo solamente para resolverlo. Esto último está relacionado con las características de *mal planteamiento* del problema (2.17). En el próximo capítulo haremos un estudio de su comportamiento.

## CAPITULO 3

# Mal planteamiento del problema inverso y alternativas de solución.

Hemos ya establecido que para nuestro problema inverso: calcular  $\hat{T}$ , tal que

$$L(\hat{T}) = \hat{h} \quad (3.1)$$

donde  $L(\hat{T})$  significa -como ya se ha dicho- resolver el problema directo con coeficiente  $\hat{T}$ , una vía de solución es calcular la  $T$  que resuelve

$$\min_T \|L(T) - \hat{h}\|^2 \quad (3.2)$$

Para varios autores esto es conocido como el problema de *identificación de parámetros* por el criterio de error de salida (output error criterion).

Ahora bien, en la práctica puede tenerse que

1. No exista  $T^*$  solución de (3.2).
2. Multiplicidad de la solución, en caso de que exista.
3. Aun suponiendo existencia y unicidad, es posible que  $T^*$  no dependa continuamente de  $\hat{h}$ .

Esto nos dice que en general no es suficiente resolver (3.2) usando algún método confiable como el descrito en el capítulo anterior, pues a pesar de ello la posible  $T^*$  calculada puede estar lejos de la solución real que estamos buscando.

A la cuestión de dar una respuesta acerca de la bondad de una solución de (3.2) calculada numéricamente se le conoce como el problema de *identificabilidad de parámetros* a cuyo estudio dedicaremos lo que resta del capítulo.

### 3.1 Mal planteamiento.

Comenzamos introduciendo algunos conceptos y resultados clásicos que se pueden ver en [14], [19],[23]. Sean  $P$  y  $H$  espacios de Hilbert y

$$L : P \rightarrow H \quad (3.3)$$

un operador lineal continuo.

**Definición 3.1.** Dado  $h \in H$ , el problema de determinar  $p \in P$  tal que

$$L(p) = h \quad (3.4)$$

se dice *bien planteado*, si tiene solución única y ésta depende continuamente de los datos. En otro caso se dirá que es *mal planteado*.

Por datos consideraremos sólo a  $h$ , aunque en un sentido más general puede considerarse también a  $L$ .

**Definición 3.2.** Se dirá que  $p^+$  es una *solución generalizada Moore-Penrose* (M-P) de (3.4), si:

- $\|h - L(p^+)\|_H \leq \|h - L(p)\|_H, \quad \forall p \in P$

y

- Si  $\|h - L(p_0)\|_H \leq \|h - L(p)\|_H, \quad \forall p \in P$  entonces  $\|p_0\|_P \geq \|p^+\|_P$ .

En general,  $\|\cdot\|_V$  denota la norma en el espacio  $V$ .

**Teorema 3.1.** Sea  $L : P \rightarrow H$  lineal y continuo y sea  $\varphi_h$  la proyección ortogonal de  $h$  sobre la cerradura de la imagen de  $L$ ,  $\overline{Im(L)}$ . Entonces, (3.4) tiene solución generalizada única  $p^+$  si y sólo si  $\varphi_h \in Im(L)$ .

Este resultado se establece directamente al aplicar el teorema de la proyección ortogonal a  $\overline{Im(L)}$  y  $N(L)$  y su importancia en nuestro contexto es que establece condiciones necesarias y suficientes para la existencia y unicidad de la solución de nuestro problema inverso (3.1) en sentido generalizado.

Sea ahora

$$D(L^+) = \{h \in H : L(p) = h \text{ tiene solución generalizada M-P}\}$$

Al mapeo  $L^+ : D(L^+) \subset H \rightarrow P$ ,  $h \rightarrow p^+$  se le llama inverso generalizado de M-P de  $L$ .

**Teorema 3.2.** Si  $L : P \rightarrow H$  es lineal y continuo, entonces  $L^+$  es lineal y cerrado.

Demostración: la linealidad se prueba directamente. Para el caso de la cerradura, sea  $(h, p)$  un punto de acumulación de

$$G_r(L^+) = \{(h, L^+(h)) : h \in D(L^+)\}$$

la gráfica de  $L^+$ . Debemos probar que  $(h, p) \in G_r(L^+)$ ; para ello, sea

$$\{(h_n, L^+(h_n))\} \text{ tal que } (h_n, L^+(h_n)) \rightarrow (h, p)$$

Como  $L$  es continuo se tiene que

$$\{(h_n - L(L^+(h_n)))\} \subset Im(L)^\perp$$

converge a  $(h - L(p)) \in Im(L)^\perp$ ; luego,  $L^+(h - L(p)) = 0$ , o bien,  $L^+h = p$ ; por tanto

$$(h, p) \in G_r(L^+)$$

□

Con respecto a la continuidad de  $L^+$  se tiene

**Teorema 3.3.** Sea  $L : P \rightarrow H$  lineal y continuo. Si  $\overline{Im(L)} = Im(L)$ , entonces  $L^+ : H \rightarrow P$  es lineal y continuo.

Otro resultado determinante en este sentido es el siguiente:

**Teorema 3.4.** Si  $L : P \rightarrow H$  es lineal y compacto, entonces, para que  $D(L^+) = H$  y  $L^+ : H \rightarrow P$  sea continuo es necesario y suficiente que  $Im(L) \subset H$  sea de dimensión finita.

**Demostración:** llamemos  $\varphi : H \rightarrow \overline{Im(L)}$  a la proyección ortogonal de  $H$  sobre  $\overline{Im(L)}$ . Ahora, si  $L^+$  es continuo, entonces  $LL^+ = \varphi|_{Im(L)}$  es compacto, y  $\varphi|_{Im(L)}$  es compacto si y sólo si  $Im(L) \subset H$  es de dimensión finita. □

Podemos ahora en este contexto, reformular los conceptos de bien y mal planteamiento como sigue

**Definición 3.3.** Sea  $L : P \rightarrow H$  lineal. Dado  $h \in H$ , se dirá que el problema  $L(p) = h$  es bien planteado (a la Hadamard), si y sólo si

$$L^+ : D(L^+) \subset H \rightarrow P$$

es continuo. En otro caso, será mal planteado.

Tenemos pues que si  $L : P \rightarrow H$  es lineal, compacto y  $Im(L)$  es de dimensión infinita, entonces estamos ante un problema mal planteado. Cabe notar que si  $L^{-1}$  (el inverso de  $L$ ) existe, entonces  $L^+ = L^{-1}$  sobre  $D(L^{-1})$ . Pasaremos ahora a ver algunos ejemplos.

**Ejemplo 1.** Considérese la ecuación

$$\int_a^b K(x, t) f(t) dt = g(x) \quad , c \leq x \leq d \quad (3.5)$$

con  $f, g \in U = L_2[a, b]$ ,  $K \in L_2([a, b] \times [c, d])$ .

(3.5) se conoce como ecuación de Fredholm de primera clase, y si hacemos  $V = L_2[c, d]$ , podemos reescribirla en la forma operacional

$$K(f) = g, \text{ con } K : U \rightarrow V \quad (3.6)$$

donde

$$K(f)x = \int_a^b K(x, t) f(t) dt \quad (3.7)$$

$K$  resulta ser un caso típico de operador lineal compacto, por lo que (3.6) es un problema mal planteado ya que  $K^{-1}$  no es continuo. Esto se muestra directamente en [16].

**Ejemplo 2.** Dada una función  $f(x)$ , calcular su  $k$ -ésima derivada  $u(x) = f^{(k)}(x)$ . Este es un problema común en análisis numérico y es usual que sólo se cuente con una tabla de valores  $T_j$  de  $f$ .

Sin pérdida de generalidad puede suponerse que

$$f^{(i)}(0) = 0, \quad i = 1, \dots, k-1$$

Del desarrollo de Taylor para  $f(x)$  con residuo en forma integral se tiene

$$\begin{aligned} f(x) &= f(0) + f^{(1)}(0)x + \dots + \frac{1}{(k-1)!} f^{(k-1)}(0)x^{k-1} + \int_0^x \frac{(x-\alpha)^{k-1}}{(k-1)!} u(\alpha) d\alpha \\ &= \int_0^a g(x, \alpha) u(\alpha) d\alpha, \quad 0 \leq x \leq a \end{aligned} \quad (3.8)$$

donde

$$g(x, \alpha) = \begin{cases} 0 & \text{si, } \alpha > x \\ \frac{(x-\alpha)^{k-1}}{(k-1)!} & \text{si, } \alpha \leq x \end{cases}$$

Esto es,  $u(x)$  resulta ser la solución de

$$\int_0^a g(x, \alpha) u(\alpha) d\alpha = f(x), \quad 0 \leq x \leq a$$

que es justamente una ecuación del tipo Fredholm, vista anteriormente.

Este ejemplo resulta particularmente importante en nuestro contexto pues en los problemas de optimización aparece de manera natural el cálculo de derivadas, como una etapa fundamental en el proceso de solución.

Con respecto a nuestro problema, si consideramos el caso de estado estacionario y una sola variable espacial, la ecuación correspondiente puede escribirse como

$$-(p(x)h_x)_x = g(x), \quad x \in [0, 1] \quad (3.9)$$

con condiciones de frontera, por ejemplo de Neumann, dadas.

Si  $h_x(x) \neq 0 \forall x$ , entonces  $p(x)$  está dada por

$$p(x) = \frac{1}{h_x(x)} \left[ p(0)h_x(0) - \int_0^x g(s) ds \right] \quad (3.10)$$

y resulta ser única.

Como se observa, el cálculo de  $p(x)$  depende de la derivada de  $h(x)$  que como vimos anteriormente, es un problema mal planteado. Además, si para alguna región se tiene  $h_x(x) \approx 0$ , resulta que pequeños errores en  $g$  pueden verse amplificados por el factor  $\frac{1}{h_x(x)}$ . En [23] se prueba el mal planteamiento de (3.9). Para el caso de dos variables espaciales y estado no estacionario no conocemos una demostración de este tipo; aunque es usual considerarlo como mal planteado, y en los experimentos numéricos que presentaremos más adelante se puede apreciar ese comportamiento.

Los resultados documentados en esta sección, en relación con la existencia y continuidad del inverso del operador lineal  $L$ , se pueden extender según Engl[15] al caso en que dicho operador sea no lineal y establecer condiciones equivalentes a las aquí vistas.

## 3.2 Regularización

Se trata de una de las alternativas más usuales y mejor estructuradas para tratar los problemas mal planteados. Su creación y desarrollo están estrechamente ligados al nombre de Tikhonov y en términos generales puede definirse como la aproximación de un problema mal planteado por una familia

de problemas vecinos bien planteados, dependiente de un parámetro. En nuestro caso, esto equivale a sustituir

$$\min_p \|L(p) - \hat{h}\|_H^2 \quad (3.11)$$

por, digamos

$$\min_{p \in SCP} R_\alpha(p) \quad (3.12)$$

donde

$$R_\alpha(p) = \|L(p) - \hat{h}\|_H^2 + \alpha^2 \|p - p_0\|_P^2 \quad (3.13)$$

con:

- $\alpha > 0$  un parámetro, llamado de regularización, estrechamente ligado al ruido presente en  $\hat{h}$ .
- $S$ , conjunto de parámetros permisibles, denso en  $P$ .
- $p_0 \in P$  alguna estimación a priori de  $p^*$ , solución de (3.11). Es posible tomar  $p_0 = 0$ .

$R_\alpha(p)$  se conoce como funcional regularizante y es posible [9] probar que existe  $p_\alpha$  llamada solución regularizada, la que bajo ciertas hipótesis resulta ser continua con respecto a  $\hat{h}$  y además, satisface

$$p_\alpha \rightarrow p^+, \quad \alpha = \alpha(\delta) \rightarrow 0$$

siendo  $\delta$  una medida del ruido en  $\hat{h}$ . En el siguiente capítulo se muestran resultados obtenidos con esta propuesta.

Un esquema general es, tomar

$$R_\alpha(p) = \|L(p) - \hat{h}\|_H^2 + \alpha S(p) \quad (3.14)$$

con  $\alpha > 0$  y  $S$  una cierta funcional no negativa, llamada estabilizante, cuyo objetivo es justamente controlar la sensibilidad de la solución regularizada  $p_\alpha$  con respecto a las perturbaciones en  $\hat{h}$ . Algunas elecciones frecuentes de  $S$  son:

1.  $S(p) = \|p\|^2$
2.  $S(p) = \|\Phi p\|^2$

donde  $\Phi$  es un operador diferencial, por ejemplo  $\Phi p = p'$  y las normas se toman en los espacios correspondientes; con frecuencia  $L_2$ .

En [1] se dan resultados sobre:

- Existencia de solución  $p_\alpha$  de

$$\min_p R_\alpha(p)$$

para el caso  $S(p) = \|p\|^2$ .

- Continuidad de  $p_\alpha$  con respecto a  $\hat{h}$ .
- Convergencia de  $p_\alpha$  a  $p^+$  cuando  $\alpha \rightarrow 0$ .

En [15] se muestra, además, que si el ruido en  $\hat{h}$  es menor o igual que  $\delta$ , es posible tomar el parámetro de regularización  $\alpha = \alpha(\delta)$  a-priori como

$$\alpha(\delta) = \delta \frac{2}{1+2\nu}$$

con  $\nu \leq 1$  para obtener soluciones regularizadas que satisfagan

$$\|p^+ - p_\alpha\| = O\left(\delta \frac{2\nu}{1+2\nu}\right)$$

En la práctica, las elecciones más usuales son  $\nu = \frac{1}{2}$  y  $\nu = 1$ .

Cabe decir, sin embargo, que la cuestión del cálculo del parámetro  $\alpha$  de regularización, cuando se desconoce  $\delta$ , es una de las grandes dificultades que deben resolverse al aplicar esta metodología en la resolución de un problema mal planteado. Aunque existen algunos algoritmos eurísticos para el cálculo automático de  $\alpha$ , su uso no es siempre exitoso para cualquier tipo de problema.

Pasaremos para concluir nuestro estudio, a la presentación de una técnica, todavía en desarrollo, que podría llegar a ser una herramienta importante en el tratamiento de problemas mal planteados, ya sea por sí misma o ligada con regularización.

### 3.3 Optimización Multiescala.

Volvamos a nuestro problema en estudio, que consiste en resolver

$$\min_T F(T), \quad F(T) = \|L(T) - \hat{h}\|^2 \quad (3.15)$$

donde  $\hat{h}$  es una matriz de observaciones de la función  $h(x, y; t)$  en ciertos puntos, digamos,  $(x_i, y_j) \in [0, 1] \times [0, 1]$ ,  $i = 0, \dots, Nx$ ,  $j = 0, \dots, Ny$ . Esto es, se tiene

$$\hat{h}_{ij} = h(x_i, y_j; t) + \delta_{ij} \quad (3.16)$$

para un cierto tiempo  $t$  y donde  $\delta_{ij}$  representa el error de medición en  $(x_i, y_j)$ . Si hacemos

$$\begin{aligned} x_i &= i\Delta x, \quad i = 0, \dots, Nx; \quad \Delta x = \frac{1}{Nx} \\ y_j &= j\Delta y, \quad j = 0, \dots, Ny; \quad \Delta y = \frac{1}{Ny} \end{aligned}$$

(3.15) se transforma en

$$\min_{T_\Delta} F_\Delta(T_\Delta), \quad F_\Delta(T_\Delta) = \|L(T_\Delta) - \hat{h}_\Delta\|^2 \quad (3.17)$$

donde  $\Delta$  es el nivel de discretización correspondiente a  $\Delta x$  y  $\Delta y$ .

Estamos ahora en un problema de optimización en dimensión finita, y la cuestión es determinar  $T_\Delta^+$ , con

$$T_\Delta^+ \equiv T_{ij}^+ = T^+(x_i, y_j) \quad (3.18)$$

solución generalizada discreta de (3.15).

Consideraremos

$$T(x, y) \in L^2(\mathfrak{R}), \quad \mathfrak{R} = [a, b] \times [c, d] \quad (3.19)$$

donde  $\mathfrak{R}$  es la región que se ve en la Figura 3.1, y lo que nos interesa es conocer  $T(x, y)$  en los puntos  $(x_i, y_j)$  de la malla dada en la Figura 3.2.

Esto puede interpretarse como, determinar  $T_\Delta$  en un subespacio de dimensión finita  $W \subset L^2(\mathfrak{R})$  definido por la discretización de  $\mathfrak{R}$ .

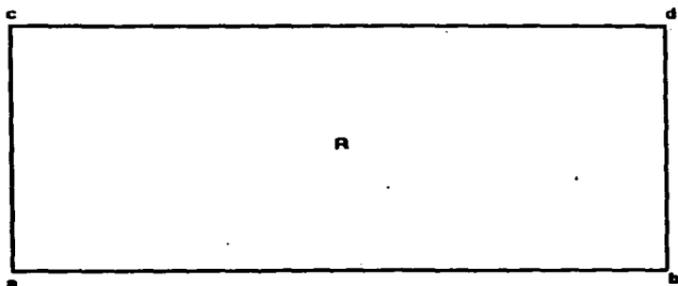


Figura 3.1: Región rectangular

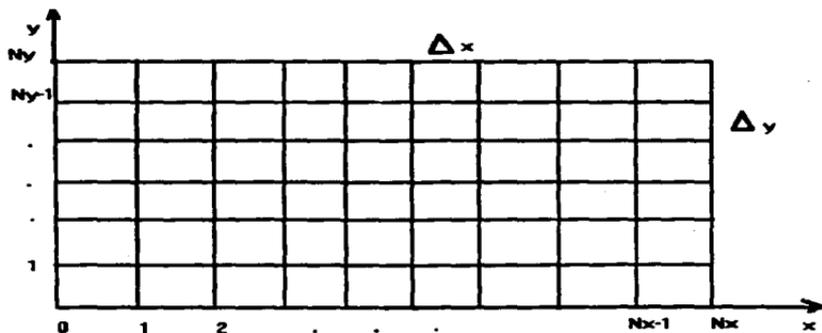


Figura 3.2: Subespacio  $W$

Se tiene que

$$\dim W = (Nx + 1) * (Ny + 1)$$

Hagamos  $\dim W = N_w$  y sea

$$\{w_{ij}\} \quad i = 0, 1, \dots, Nx; j = 0, 1, \dots, Ny \quad (3.20)$$

una base de  $W$ ; de esto resulta

$$T_{\Delta}(x, y) = \sum \alpha_{ij} w_{ij}, \quad (x, y) \in \mathfrak{R} \quad (3.21)$$

y entonces resolver (3.17) equivale a determinar la matriz de coeficientes

$$A = (\alpha_{ij}) \quad (3.22)$$

tal que (3.17) alcanza el mínimo con respecto a la base (3.20). Es importante notar que la elección de ésta es un punto fino y delicado pues para bases diferentes pueden obtenerse resultados dramáticamente distintos [10].

Por supuesto, con los elementos dados podría procederse a la estimación de  $T_{\Delta}^*$  directamente; sin embargo, lo que se propone es una estrategia distinta, llamada *optimización multiescala*, cuya adopción produce generalmente resultados superiores a los obtenidos al resolver el problema de estimación de parámetros directamente en el subespacio  $W$ , según mostraremos para nuestro problema en el siguiente capítulo. Esta técnica ya ha sido usada para resolver problemas inversos [7],[11], aunque esta es la primera vez que se usa para mostrar que se pueden resolver dichos problemas en el caso de errores en los datos.

### 3.3.1 Parametrización multiescala.

En términos generales, la idea consiste en construir una serie de subespacios  $W_0, W_1, \dots, W_{NE}$  de  $W$ , tales que

$$W_{k-1} \subset W_k \quad \text{y} \quad W_{NE} = W \quad (3.23)$$

Una propuesta es partir de un  $W_0$  dado por la discretización dada en la Figura 3.3, y proceder por bipartición de cada uno de los rectángulos  $\mathfrak{R}_{ij}$ , para obtener el correspondiente  $W_1$ , y así sucesivamente. En general, si se tiene  $W_{k-1}$  que corresponde a la discretización dada en la Figura 3.4, el paso a

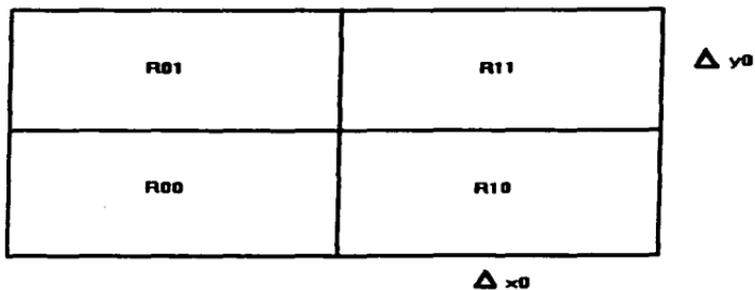


Figura 3.3: Espacio  $W_0$

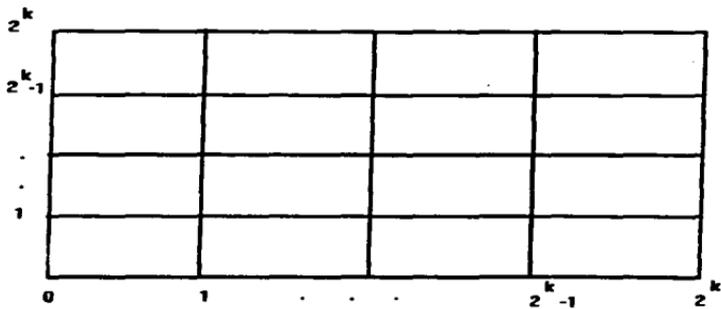


Figura 3.4: Subespacio  $W_{k-1}$

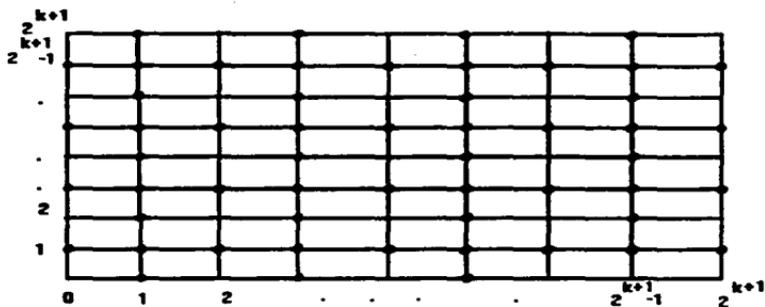


Figura 3.5: Puntos que se agregan para pasar a  $W_k$

$W_k$  se hace agregando a la malla para  $W_{k-1}$  los puntos marcados en negritas, y que se muestran en la Figura 3.5, en la discretización correspondiente a  $W_k$ .

Si llamamos  $V_k$  al subespacio generado a partir de estos nuevos puntos agregados a  $W_{k-1}$  para llegar a  $W_k$ , tendremos que

$$W_k = W_{k-1} \oplus V_k \quad , \quad W_{k-1} \subset W_k \quad (3.24)$$

Se tiene que

$$\dim W_k = (2^{k+1} + 1) * (2^{k+1} + 1)$$

y como base para este subespacio se propone elegir funciones

$$w_{ij}, i = 1, \dots, 2^{k+1} + 1 ; j = 1, \dots, 2^{k+1} + 1 \quad (3.25)$$

continuas y bilineales por pedazos sobre cada uno de los rectángulos  $\mathcal{R}_{ij}^k$ .

Por supuesto, como  $W_{k-1} \subset W_k$ , lo natural para obtener la base (3.25) es agregar a la de  $W_{k-1}$ , las  $w_{ij}$  correspondientes al subespacio  $V_k$ . Resulta fácil ver que  $V_k$  es un espacio suplementario a  $W_{k-1}$  en  $W_k$ , ya que:

- Por construcción, se tiene

$$V_k \subset W_k$$

- Dada  $T_{\Delta k}^W \in W_k$ , si consideramos que ésta coincide con  $T_{\Delta k-1}^W \in W_{k-1}$  en todos los puntos de la malla correspondiente a  $W_{k-1}$  y hacemos

$$T_{\Delta k}^V = T_{\Delta k}^W - T_{\Delta k-1}^W$$

Por construcción se tiene que  $T_{\Delta k}^V \in W_k$ , y si  $T_{\Delta k}^V$  se hace cero en todos los nodos correspondientes a  $W_k$  que no están en  $V_k$ , entonces  $T_{\Delta k}^V \in V_k$ , de donde

$$T_{\Delta k}^W = T_{\Delta k-1}^W + T_{\Delta k}^V \quad (3.26)$$

con

$$T_{\Delta k-1}^W \in W_{k-1} \quad \text{y} \quad T_{\Delta k}^V \in V_k$$

- Puede probarse sin grandes dificultades que la descomposición (3.26) es única, y con esto queda mostrado que  $V_k$  es suplementario de  $W_{k-1}$  en  $W_k$ .

El proceso descrito anteriormente se repite hasta arribar a un subespacio  $W_{NE}$ , tal que  $W_{NE} = W$ , esto es, se tiene  $W_0 \subset W_1 \subset \dots \subset W_{NE} = W$  y como  $W_k = W_{k-1} \oplus V_k$  resulta

$$\begin{aligned} W_1 &= W_0 \oplus V_1 \\ W_2 &= W_1 \oplus V_2 = W_0 \oplus V_1 \oplus V_2 \\ &\vdots \\ W_{NE} &= W_0 \oplus V_1 \oplus \dots \oplus V_{NE} = W \end{aligned}$$

Así,  $B = \{w_{ij}\}$  base de  $W$  se obtiene como la unión de las bases de  $W_0, V_1, \dots, V_{NE}$ ; esto es

$$B = \left\{ w_{ij}^k \mid k = 0, 1, \dots, NE \quad ; \quad i = 0, 1, \dots, 2^{k+1} - 1, \quad j = 0, 1, \dots, 2^{k+1} - 1 \right\}$$

Existen distintas opciones para la elección de  $B$  entre las llamadas bases

tipo multiescala. Por nuestra parte hemos elegido

$$w_{ij}(x, y) = \begin{cases} \left[ \frac{x}{h} - (i-1) \right] \left[ \frac{y}{h} - (j-1) \right] & \text{si } i-1 \leq \frac{x}{h} \leq i, \quad j-1 \leq \frac{y}{h} \leq j \\ \left[ \frac{x}{h} - (i-1) \right] \left[ (j+1) - \frac{y}{h} \right] & \text{si } i-1 \leq \frac{x}{h} \leq i, \quad j \leq \frac{y}{h} \leq j+1 \\ \left[ (i+1) - \frac{x}{h} \right] \left[ \frac{y}{h} - (j-1) \right] & \text{si } i \leq \frac{x}{h} \leq i+1, \quad j-1 \leq \frac{y}{h} \leq j \\ \left[ (i+1) - \frac{x}{h} \right] \left[ (j+1) - \frac{y}{h} \right] & \text{si } i \leq \frac{x}{h} \leq i+1, \quad j \leq \frac{y}{h} \leq j+1 \end{cases}$$

### 3.3.2 Proceso de optimización con base en multiescala.

Partiendo de lo anterior, la idea consiste en resolver la siguiente colección de problemas

$$\min_{T_{\Delta k}} F_{\Delta}(T_{\Delta k}), \quad T_{\Delta k} \in W_k, \quad 0 \leq k \leq NE \quad (3.27)$$

donde, como hemos visto, la dimensión de cada nuevo problema va creciendo, dado que  $W_{k-1} \subset W_k$ . En términos algorítmicos el procedimiento es el siguiente:

- Para  $k = 0$ , se calcula la solución  $T_{\Delta 0}^+$  de

$$\min_{T_{\Delta 0}} F_{\Delta}(T_{\Delta 0})$$

partiendo de una aproximación inicial

$$A_0 = (\alpha_{ij}^0), \quad i = 0, \dots, 2^k, \quad j = 0, \dots, 2^k$$

arbitraria, donde, recordemos,  $T_{\Delta 0}^0 = \sum \alpha_{ij} w_{ij}$ .

- Una vez calculada  $T_{\Delta k-1}^+$ , se resuelve el correspondiente

$$\min_{T_{\Delta k}} F_{\Delta}(T_{\Delta k}), \quad 1 \leq k \leq NE$$

usando como condición inicial los  $\alpha_{ij}^{k-1}$  obtenidos como resultado de la optimización  $k-1$ , más los correspondientes a  $V_k$  de la descomposición

$$W_k = W_{k-1} \oplus V_k$$

calculados por interpolación a partir de  $T_{\Delta k-1}^+$ , pues según vimos antes

$$T_{\Delta k}^W = T_{\Delta k-1}^W - T_{\Delta k}^V$$

con

$$T_{\Delta k-1}^W \in W_{k-1} \quad \text{y} \quad T_{\Delta k}^V \in V_k$$

Es importante destacar que, dentro del proceso de minimización de  $F_{\Delta}(T_{\Delta k})$  para cualquier  $k$ , cada vez que se requiere la evaluación de  $F_{\Delta}$  en una  $T_{\Delta k}$  dada hay que resolver la ecuación diferencial que ya vimos que forma parte del modelo.

En el próximo capítulo se muestran resultados obtenidos con este método, que son superiores a los obtenidos con optimización directa, lo que constituye una evidencia experimental acerca de aplicar multiescala en un problema inverso como el tratado aquí.

Existen también algunos estudios, Chavent y Liu[10] por ejemplo, en los que se analiza multiescala en los siguientes términos:

Dado el modelo elíptico unidimensional

$$-\frac{d}{dx} \left( T(x) \frac{dh}{dx} \right) = f(x), \quad x \in [0, 1]$$

con

$$h(0) = h(1) = 0$$

y como la solución depende de  $T$ , se define el operador

$$\begin{aligned} \Phi : L^2(0, 1) &\rightarrow L^2(0, 1) \\ \Phi(T) &= h \end{aligned}$$

o, para el caso discreto,

$$\Phi_{\Delta}(T_{\Delta}) = h_{\Delta}$$

donde, análogamente a lo visto antes,

$$T_{\Delta} = \sum \alpha_i e_i$$

siendo  $e_1, \dots, e_n$  una base del subespacio de  $L^2(0, 1)$  correspondiente al nivel de discretización  $\Delta$ .

A partir de esto, en [10] se definen los siguientes conceptos:

- Velocidad

$$v_i = \Phi'_\Delta(T_\Delta) e_i$$

que es justamente la derivada de  $\Phi_\Delta$  con respecto a los  $\alpha_i$ , que son ahora las variables independientes dada la base.

- Aceleración

$$a_i = \Phi''_\Delta(T_\Delta)(e_i, e_i)$$

y finalmente

- Radio de curvatura

$$\rho_i = \frac{\|v_i\|^2}{\|a_i\|^2} \left( 1 - \left\langle \frac{v_i}{\|v_i\|^2}, \frac{a_i}{\|a_i\|^2} \right\rangle \right)^{\frac{1}{2}}$$

Tomando luego  $T_\Delta = 1$  y  $f(x) = 2$  se muestra que para un nivel de discretización  $\Delta$  dado, la resolución del problema de optimización

$$\min_{\alpha_\Delta} \left\| \Phi_\Delta \left( \sum_{i=1}^n \alpha_i e_i \right) - h_\Delta \right\|^2, \quad \alpha_\Delta = [\alpha_1, \dots, \alpha_n]^t \quad (3.28)$$

reporta mejores resultados cuando se toma como  $\{e_i\}_{i=1}^n$  una base multiescala en vez de una base local.

Por su parte, Liu en [18] tomando  $T_\Delta = 1$  y  $f(x) = 1$  calcula expresiones para la velocidad y la curvatura de  $\Phi$  a partir de

$$\Phi(T(x) + te(x))$$

donde

$$e(x) = \begin{cases} 0 & x \in (0, t - \Delta) \\ -1 & x \in (t - \Delta, t) \\ 1 & x \in (t, t + \Delta) \\ 0 & x \in (t + \Delta, 1) \end{cases}$$

y llega a las siguientes expresiones

$$v^2 \approx \frac{2\Delta^2 \left(t - \frac{1}{2}\right)}{3}, \quad K^2 \approx 6\Delta^{-2}, \quad \text{cuando } \Delta \rightarrow 0$$

De esto resulta que a medida que la discretización (dada por  $\Delta$ ) se hace cada vez más fina:

- La velocidad decrece, y como ésta mide la variación de  $\Phi$  con respecto a  $T$ , resulta entonces que la función objetivo (3.28) es menos sensible a cambios en los coeficientes de la representación de  $T$  correspondiente a escalas finas.
- La curvatura crece (o el radio de curvatura  $\rho = \frac{1}{|K|}$  disminuye), lo que se traduce en mayor no linealidad de (3.28) a medida que se tengan escalas más finas.

En consecuencia, se sugiere no resolver (3.28) directamente sino con optimización multiescala, partiendo como se ha propuesto antes, de una escala cero en la que se espera que el problema sea lo menos no lineal posible, pero con alta sensibilidad.

Vista de esta manera, multiescala puede interpretarse como un tipo de regularización en donde la función regularizada tendría la expresión

$$J(\alpha_i; \beta) = \left\| \Phi_{\Delta} \left( \sum_{i=1}^n \alpha_i e_i \right) - h_{\Delta} \right\|^2 + \sum \beta_i \sum \alpha_i$$

siendo las  $\beta_i$  ciertos pesos. Así, si fuera posible conocer cuales son las  $\alpha_i$  más sensibles, se podría dar mayor peso a estos coeficientes en las primeras iteraciones de la optimización, lo que resultaría en una buena aproximación a esos  $\alpha_i$ . Posteriormente la optimización podría calcular los  $\alpha_i$  más no lineales. Desde luego no se conocen a priori cuales son los coeficientes sensibles y menos no lineales, por lo que regularizar de esta forma no es posible en la práctica, y entonces usar mallas multiescala se convierte en un método válido para resolver problemas mal planteados de este tipo.

Llevar a cabo un análisis del estilo propuesto por estos autores para nuestro problema (ecuación parabólica) en dos dimensiones es una tarea que queda por hacer.

## CAPITULO 4

# Experimentación y resultados computacionales.

Como se ha mencionado antes, para llevar a buen término el proceso de estimar  $T$ , se propone resolver para cada escala  $k$  el siguiente problema de optimización

$$\min_{T_{\Delta_k}} F_{\Delta}(T_{\Delta_k}), \quad F_{\Delta}(T_{\Delta_k}) = \|L(T_{\Delta_k}) - h\| \quad (4.1)$$

En este caso, evaluar la función objetivo del problema de mínimos cuadrados (4.1) implica resolver una ecuación en derivadas parciales, para obtener la carga en cada punto de la malla. Es decir, para estimar la solución del problema inverso, que es nuestro interés, es necesario resolver una gran cantidad de veces el modelo (problema directo). Esto hace que el proceso sea más costoso aun y se trata, por tanto, de usar programas robustos y rápidos en cada etapa.

Para resolver el problema directo

$$\frac{\partial}{\partial x} \left( T(x, y) \frac{\partial h}{\partial x} \right) + \frac{\partial}{\partial y} \left( T(x, y) \frac{\partial h}{\partial y} \right) = S \frac{\partial h}{\partial t} + Q \quad (4.2)$$

con condiciones iniciales y de frontera dadas, se usó una implementación de alta precisión del método implícito de direcciones alternadas, mientras

que para la optimización se utilizó el método de Quasi-Newton de memoria limitada implementado en el código L-BFGSB [28], que contempla la posibilidad de cotas en las variables. Este punto es importante ya que, si bien es cierto que estamos estimando  $T$  -ya que es muy caro obtener su valor por mediciones-, también es cierto que si se conoce el rango posible en el que ésta varía, por lo que se introducen restricciones de acotación en 4.1 según el criterio de los expertos. Esta rutina de optimización, como muchas de su estilo, requieren de la expresión de la función y del gradiente. En este caso, dada la complejidad de la función a minimizar, se usa una aproximación del gradiente por diferencias finitas.

Como herramienta de cómputo se usó una microcomputadora tipo PC compatible, y se hicieron cálculos a partir de datos sintéticos y reales.

## 4.1 Caso sintético.

Un estudio de este tipo se lleva a cabo en la medida que permite calibrar, en condiciones en las que el experimentador tiene control de las variables del proceso. Ahora bien, dado que se trata de calcular  $T(x, y)$  conociendo  $h$ , definimos que en el óptimo la transmisividad es una función dada por la siguiente expresión:

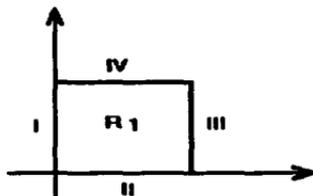
$$T^*(x, y) = x^2 + y^2 + xy + x + y + 1 \quad (4.3)$$

Esto es, partiendo de  $T^*$  y tomando  $S = 1$ ,  $Q = 0$  en (4.2), se resolvió

$$L(T^*) = h \quad (4.4)$$

para obtener  $h$ . El hecho de conocer la solución  $T^*$  nos permite comprobar si el ajuste es correcto o no.

Hecho esto, el problema inverso consistió en recuperar  $T^*$  a partir de esta  $h$ . Los resultados que presentaremos corresponden a la solución de dicho problema sobre el cuadrado unitario. Las condiciones de frontera consideradas fueron las siguientes, que corresponden a la Figura 4.1.



**Figura 4.1: Región R1**

escala	intervalos por x e y	nodos
0	2,2	3*3
1	4,4	5*5
2	8,8	9*9
3	16,16	17*17
4	32,32	33*33

**Tabla 4.1: No. de intervalos y de nodos en cada escala para R1**

$$T(0, y) \frac{\partial h}{\partial x} = 0 \quad \text{en I} \qquad T(x, 0) \frac{\partial h}{\partial y} = 0 \quad \text{en III}$$

$$h(1, y; t) = 1 \quad \text{en II} \qquad T(x, 1) \frac{\partial h}{\partial y} = 0 \quad \text{en IV}$$

y como condición inicial se tomó  $h(x, y; 0) = 1, \forall x, y \in [0, 1] \times [0, 1]$ .

Para resolver el problema directo fue necesario definir el nivel de discretización en el que se debe resolver la ecuación diferencial que ahí aparece o, lo que es lo mismo, definir en cual de las escalas dadas en la tabla 4.1 se calcularía la carga hidráulica  $-h-$  para una  $T$  dada. Esto se determinó aplicando el proceso de optimización multiescala usando diferentes tolerancias. En la Gráfica 4.1 se ve que con la tolerancia de  $1.d - 15$  el valor de la función objetivo se hace suficientemente pequeño, en la escala 3, con lo que se logra reproducir adecuadamente la transmisividad. Además si se continúa a la siguiente escala empeora el valor de la función. Por tal motivo el problema directo se resuelve siempre en la malla de  $17 \times 17$ .

En la Gráfica 4.2 se ilustra la necesidad de utilizar una alta precisión

	eval. de la función	error en T
sin multiescala	50	4.6
con multiescala	43	0.547

**Tabla 4.2: Datos sintéticos, sin ruido y con tol=1.d-8**

	eval. de la función	error en T
sin multiescala	670	4.3
con multiescala	75	0.202

**Tabla 4.3: Datos sintéticos, sin ruido y con tol=1.d-12**

(1.d-12) para llegar a reproducir  $T$ , con la precisión requerida, sin hacer multiescala.

#### 4.1.1 Datos sin error.

En las tablas 4.2 y 4.3 se puede apreciar el efecto de aplicar la optimización multiescala, descrita en el capítulo 3, aun cuando los datos no tienen ninguna perturbación o error. Aplicar esta técnica presenta dos ventajas, lo mismo en el caso en que se usa tolerancia de 1.d-8 como cuando se usa 1.d-12:

- Se reproducen las transmisividades, cosa que se aprecia (tablas 4.2, 4.3 y Gráfica 4.3) por la norma infinito del error entre la  $T$  calculada ( $T_c$ ) y la real ( $T_r$ ).

$$\text{error en } T = \|T_c - T_r\|_{\infty}$$

- El costo es menor que cuando no se usa multiescala, lo que se mide por el número de evaluaciones de la función.

En conclusión, para este caso, el uso de multiescala acelera el proceso.

#### 4.1.2 Presencia de error en los datos sintéticos

Como resulta que en los problemas reales hay presencia de errores en los datos, aun para el caso sintético tiene sentido trabajar con ciertas perturbaciones de la solución teórica. El estudio que se hace en estos casos es muy

importante ya que, como el problema que estamos tratando es mal planteado, un pequeño error en los datos provoca oscilaciones en la solución y resulta interesante observar que también en esta situación la optimización multiescala produce buenos resultados, mismos que sería razonable esperar aplicando una técnica de regularización.

Para simular esto en el problema de laboratorio, se perturbaron los valores de la carga hidráulica obtenidos en (4.4). Teniendo en cuenta que los valores de  $h$ , en este caso, están entre 0 y 1 se dejaron sólo dos cifras significativas correctas. Esto representa un ruido del 1.1% con respecto al rango en que fluctúan estos valores.

En la tabla 4.4 se ve que si no se aplica multiescala, o sea si se resuelve el problema de mínimos cuadrados directamente en la escala 3, la norma infinito del error entre  $T$  calculada y real es grande. Esto significa que no se reproduce la transmisividad y la superficie correspondiente a esta  $T$  se muestra en la Gráfica 4.4. Por otro lado si se usa multiescala sí se reproducen los valores de  $T$  correctos, lo que se muestra en la Gráfica 4.5.

Con el objetivo de probar la optimización multiescala en un problema como el nuestro pero donde se asume que la función de transmisividad no sea convexa y además tenga varios mínimos locales, tomamos la  $T$  óptima como sigue:

$$T_1^*(x, y) = \left( \sum_{i=1}^5 i \cos((i+1)x + 1) \right) \left( \sum_{i=1}^5 i \cos((i+1)y + 1) \right) \quad (4.5)$$

Al igual que con el caso anterior se resolvió (4.4) para  $T^* = T_1^*$ . Se perturbó la  $h$  resultante de la misma manera que se hizo cuando trabajamos con la primera  $T^*$  y los resultados de resolver el problema inverso correspondiente se muestran en la Gráfica 4.3.

Como se puede observar éstos no son tan buenos como cuando se toma la transmisividad dada por (4.3). Sucede que en este caso la expresión (4.5) es mucho más no lineal que la primera y multiescala no es un método global, por lo que los resultados en este caso no son los definitivos. Según se sabe, los

	eval. de la función	error en T
sin multiescala	105	4.511
con multiescala	49	0.601

**Tabla 4.4: Datos sintéticos, con ruido y con  $\text{tol}=1.d-8$**

alfa	eval. de la función	error en T
1.d-3	5	4.75
1.d-5	15	4.75
1.d-7	48	4.75
1.d-9	137	4.52

**Tabla 4.5: Resultados de aplicar Tijonov según (4.5), con  $\text{tol}=1.d-8$**

valores de la transmisividad en la mayor parte de los acuíferos cambian suavemente, como la primera expresión de  $T(x, y)$  dada, o se mantiene constante por secciones como veremos en la segunda parte de este capítulo cuando trabajamos con los datos reales.

### Comparación entre multiescala y regularización de Tijonov.

Alternativamente, como en este caso conocemos el error en los datos, que es del orden de

$$\delta = \|h - h_0\| = 1.d - 3$$

podemos aplicar regularización de Tijonov. Esto significa que debemos ahora resolver el problema

$$\min_T J(T; \alpha), \quad J(T; \alpha) = [\|L(T) - h\|_H^2 + \alpha \|T - T_{opt}\|_2^2] \quad (4.6)$$

donde  $T_{opt}$  es la  $T$  dada por (4.3) y  $\alpha$  es el parámetro de regularización.

Como se vio en el capítulo 3, tomamos  $\alpha = (\delta)^{\frac{2}{3}}$ . Como es de esperar, los resultados son correctos y la superficie dada por la  $T$  solución de (4.1) se muestra en la Gráfica 4.5. En realidad, no tiene sentido comparar la solución que se obtiene por esta vía con la que se obtiene de multiescala pues en el primer caso se asume que se conoce la solución real  $T_{opt}$ , mientras que multiescala no usa esa información.

También se probó aplicar regularización de Tijonov, sin usar ninguna información de la solución real, resolviendo:

$$\min_T J(T; \alpha), \quad J(T; \alpha) = [\|L(T) - h\|_H^2 + \alpha \|T\|_2^2] \quad (4.7)$$

y los resultados están dados en la tabla 4.5. En la Gráfica 4.6 se aprecia que para ninguno de los valores de  $\alpha$ , con los que se probó, se reproduce la transmisividad correcta.

Se probó también con otra norma

$$\min_T J(T; \alpha), \quad J(T; \alpha) = \left[ \|L(T) - h\|_H^2 + \alpha \left( \sum_{i=1}^{np} \left( \frac{\partial T_i}{\partial x} \right)^2 + \left( \frac{\partial T_i}{\partial y} \right)^2 \right) \right] \quad (4.8)$$

donde  $np$  es la cantidad de puntos de la malla sobre la que se discretizó  $T$ , sugerida por Chardigny en [7].

Como ya se ha mencionado, determinar el valor del parámetro de regularización  $\alpha$  es un problema complicado y estrechamente ligado a las particularidades de cada problema. Para este caso en particular se hicieron algunos experimentos con distintos valores del parámetro en (4.8) y no se lograron buenos resultados. Esto se muestra en la Gráfica 4.7, y en la tabla 4.6 se puede apreciar el número de evaluaciones de la función y la norma infinito del error en la  $T$  calculada.

En resumen, de los experimentos realizados usando regularización de Tijonov se puede decir que sólo se obtuvieron buenos resultados en el caso en que se conoce una aproximación, suficientemente buena, de la solución. A pesar de que, al resolver (4.7) o (4.8) no se obtienen resultados satisfactorios, no se puede afirmar que tal regularización no funciona correctamente sino que para los valores del parámetro usados no se reproduce la transmisividad esperada. Esto deja, por lo tanto, un problema abierto. En cambio con multiescala se logra estimar correctamente el parámetro sin usar ninguna aproximación inicial del mismo.

Es importante recordar que, un problema adicional que lleva consigo la regularización de Tijonov es que depende de una correcta selección del

parámetro de regularización para la obtención de buenos resultados, mientras que la *optimización multiescala no depende de ningún parámetro*, lo que la hace más robusta.

## 4.2 Caso real: acuífero cubano de Ariguanabo.

Trabajaremos ahora con un caso real, el Acuífero de Ariguanabo, que se muestra en la Figura 4.2 y tiene las siguientes características:

- Es de tipo libre o freático.
- Abarca un área de  $259.7 \text{ Km}^2$ .

Para su estudio, se divide en 4 subcuencas: Laguna, Aeropuerto, Govea y Zona circundante y aunque no es confinado se puede modelar por (1.20), debido a que las variaciones de la carga hidráulica  $-h-$  en el tiempo, son muy pequeñas en comparación con su valor promedio, por lo que la transmisividad no depende de  $h$ . Así la no linealidad típica de las ecuaciones diferenciales parciales para acuíferos libres, no aparece.

La zona sombreada de la Figura 4.2 corresponde a la subcuenca Aeropuerto que es la que estudiamos aquí. Los límites superficiales norte, sur y oeste se modelan como contornos impermeables. Se consideró el límite este como una línea de flujo y el límite nordeste como una zona de drenaje subterráneo.

Esta subcuenca abarca  $78.2 \text{ Km}^2$ . El período seleccionado para usar las mediciones fue de 3 años, dividido en trimestres, es decir 12 intervalos de tiempo.

Después de varios ensayos de prueba y error, los ingenieros a cargo lograron determinar que la transmisividad  $T(x, y)$  se mantiene constante en tres áreas de esta subcuenca, tal como se ve en la Figura 4. Se trata entonces de reproducir esa solución estimada por prueba y error, automáticamente. Los resultados de la optimización directa se muestran en la Gráfica 4.6. Asimismo, en la Figura 4.3 se muestran los datos para  $T$  y  $h$ .

escala	intervalos por x e y	nodos
0	2,1	3*2
1	6,4	7*5
2	12,8	13*9
3	24,16	25*17

**Tabla 4.6: Escalas para la región irregular**

Para resolver el problema directo se hizo una adaptación del programa utilizado para el caso sintético, pues ahora la región no es cuadrada sino irregular. Además, el coeficiente de almacenamiento  $S$  tiene valor de 0.15 distinto al del caso sintético que se tomó como 1. También el valor de  $Q$ , que representa la extracción y carga del acuífero, tiene valores en cada punto de la malla y en cada paso de tiempo, que fueron aportados por los ingenieros, y no como en el caso sintético donde se consideró  $Q = 0$ .

La idea básica fue inscribir la región irregular en un rectángulo y sobre éste construir una malla regular uniforme. Esta malla se construyó de tal forma que coincidiera con la correspondiente a alguna escala de las generadas a partir de la escala cero del proceso de multiescala (ver tabla 4.6). De manera general las condiciones de frontera se pueden escribir de la forma

$$\alpha T(x, y) \frac{\partial h}{\partial n} = \beta h - \gamma \quad \text{en } \Omega$$

con  $\alpha, \beta$  y  $\gamma$  constantes y al menos en una parte de  $\partial\Omega$  se tiene  $\beta \neq 0$  para garantizar unicidad de la solución del problema directo. Los valores de  $\alpha, \beta$  y  $\gamma$  son calculados en cada punto de la frontera.

En la Gráfica 4.9 se muestran los valores de  $T(x, y)$  calibrados por prueba y error. Ahí mismo se da la gráfica de  $T(x, y)$  obtenida por optimización directa, partiendo de una transmisividad constante (como si no se tuviera ninguna información)  $T^0(x, y_j) = 1$  como aproximación inicial y usando una tolerancia de 1.d-12. En la tabla 4.7 se puede comparar el efecto de usar multiescala para los datos sin perturbaciones. En este caso es interesante notar que no se comporta como en el caso sintético, ya que multiescala re-

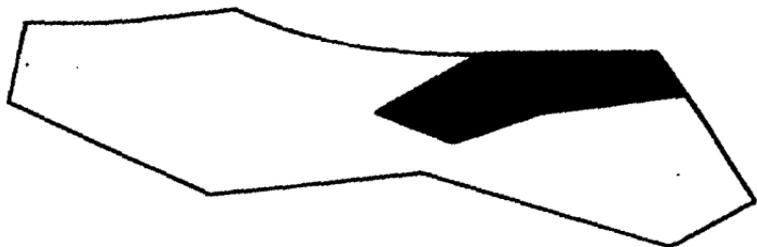


Figura 4.2: Acuífero de Ariguanabo

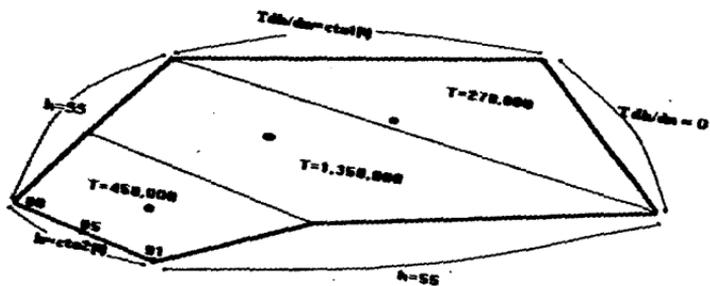


Figura 4.3: Subcuenca Aeropuerto

	eval. de la función	error en T
sin multiescala	78	0.04
con multiescala	220	0.025

**Tabla 4.7: Datos reales, sin ruido y con  $\text{tol}=1.d-8$**

sulta ser mucho más costosa. Finalmente, multiescala para el problema de acuíferos está diseñado para regularizar en el caso de error en los datos.

### 4.2.1 Datos con error.

En el caso de los datos reales debemos anotar lo siguiente. Dado que los ingenieros tienen mediciones (que ya contienen un error) de la carga hidráulica sólo en algunos puntos de la región, es necesario interpolar los que faltan y esto ya introduce un error en las observaciones  $-h-$  con las que se trabaja. No obstante se procedió al igual que para el caso sintético y se provocó una perturbación, aun menor, del 0.01% con respecto al rango de valores de  $h$  (considerados en metros). Este error implica dejar solamente dos decimales correctos, lo que ya significa un error muy pequeño.

En la Gráfica 4.10 se muestran los resultados obtenidos con optimización directa y aplicando optimización multiescala. Como se observa cuando no se aplica ésta, no se reproduce la  $T$ , pero no sólo eso. Anteriormente se habló de cotas o valores extremos entre los cuales se sabe está  $T$ , y si se observa en la misma gráfica, se ve que la solución del proceso sin multiescala oscila tanto que se va a las cotas. Si éstas no existieran, tomaría valores disparatados. Es notable, por tanto, la calidad de la solución cuando se aplica optimización multiescala.

Por otra parte, al igual que en el caso sintético, se probó resolver el problema usando regularización de Tijonov

$$\min_T J(T; \alpha), \quad J(T; \alpha) = \left[ \|L(T) - \hat{h}\|_H^2 + \alpha \left( \sum_{i=1}^{np} \left( \frac{\partial T_i}{\partial x} \right)^2 + \left( \frac{\partial T_i}{\partial y} \right)^2 \right) \right] \quad (4.9)$$

con un valor de  $\alpha = 1.d - 06$ , como el propuesto en [7], y se lograron buenos resultados. En la Gráfica 4.11 se muestra la  $T$  obtenida por esta vía y la co-

	eval. de la función	error en T
sin multiescala	94	0.36
con multiescala	77	0.08

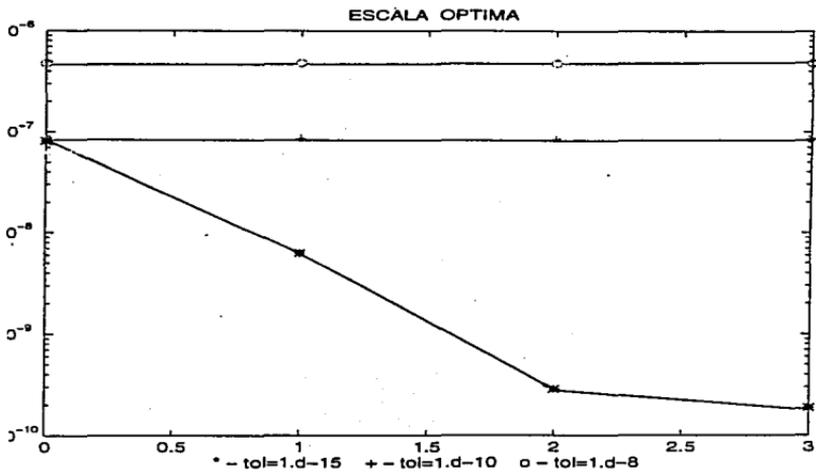
**Tabla 4.8: Datos reales, con ruido y con  $\text{tol}=1.d-8$**

	eval. de la función	error en T
con multiescala	77	0.08
Regularización	122	0.011

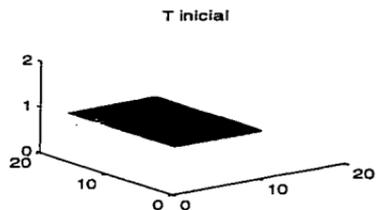
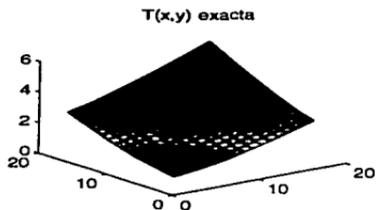
**Tabla 4.9: Tijonov (4.8) y multiescala. Datos reales y  $\text{tol}=1.d-8$**

respondiente al cálculo con multiescala, con una tolerancia de  $1.d-8$ . Como se observa, la regularización reproduce mejor la transmisividad conocida, aunque como se muestra en la tabla 4.9, a un costo mucho mayor.

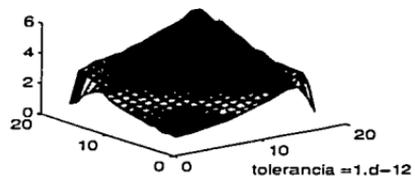
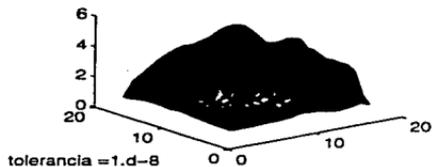
Según comentamos antes, estos experimentos han sido realizados para una parte del acuífero cubano de Ariguanabo y lo acertado de los resultados obtenidos sólo sienta pautas para extender todo lo expuesto en este trabajo al acuífero en toda su extensión. Esto es por tanto uno de los caminos en los que, de hecho, ya se ha comenzado a trabajar.



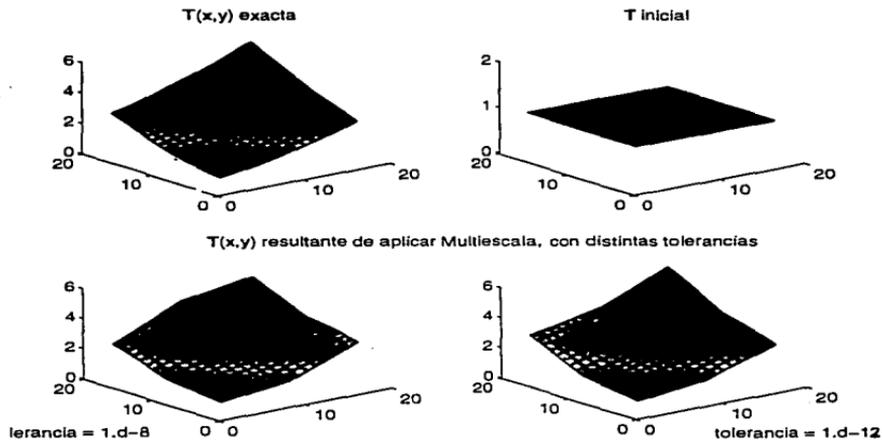
Gràfica 4.1



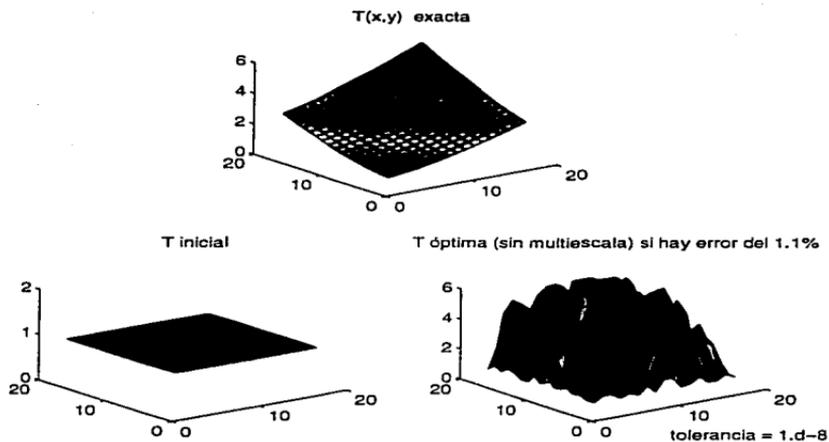
**T(x,y) resultante de optimizar en 17x17, con distintas tolerancias**



**Gráfica 4.2**

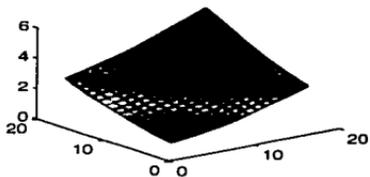


Gráfica 4.3

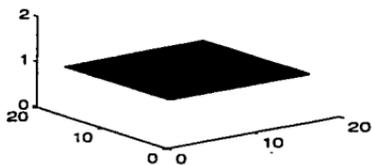


**Gráfica 4.4**

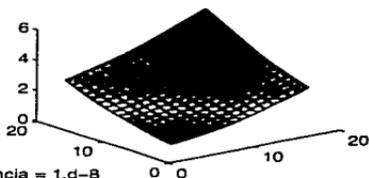
$T(x,y)$  exacta



T inicial

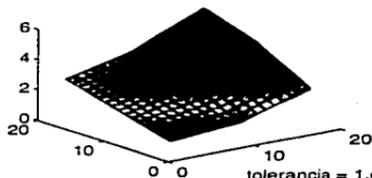


$T(x,y)$  óptima usando Tijonov



tolerancia = 1.d-8

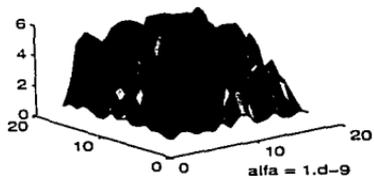
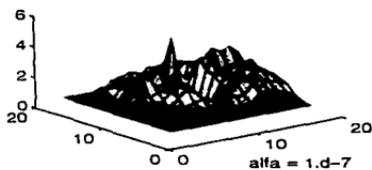
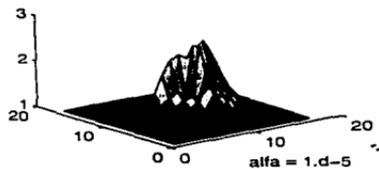
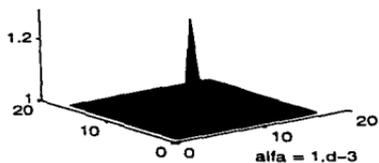
$T(x,y)$  óptima usando Multiescala



tolerancia = 1.d-8

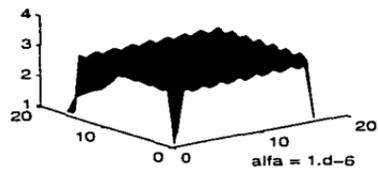
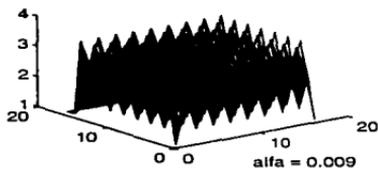
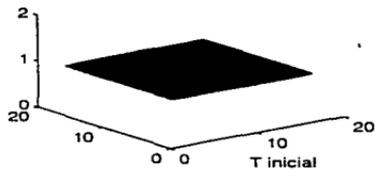
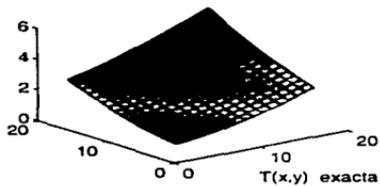
Gráfica 4.5

Resultados de minimizar  $\|L(T)-h\|^2 + \alpha\|T\|^2$ , con  $\text{tol} = 1.d-8$  y diferentes valores de  $\alpha$

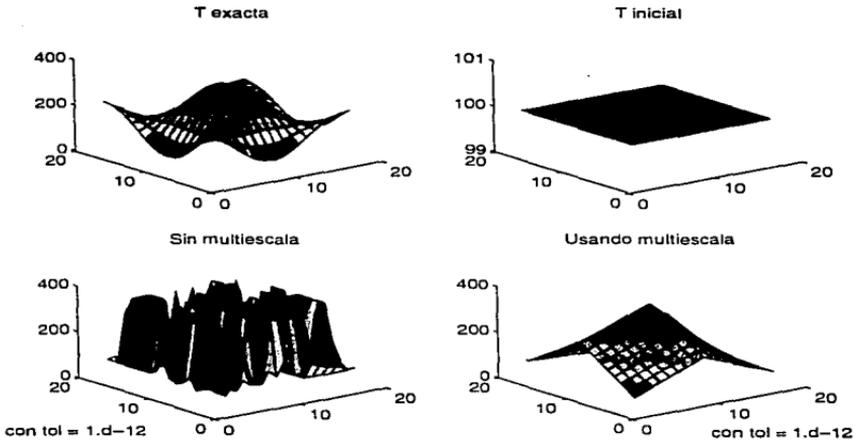


Gráfica 4.6

Resultados de  $\min \|L(T)-h\|^2 + \alpha((dT/dx)^2+(dT/dy)^2)$ , para dos valores distintos de  $\alpha$



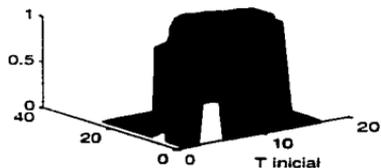
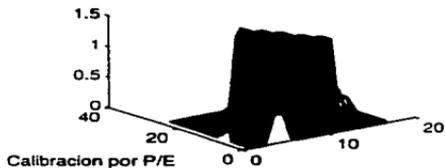
Gráfica 4.7



Gráfica 4.8

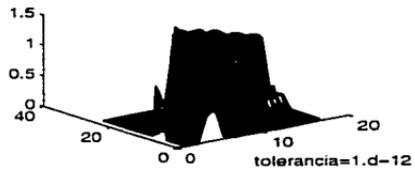
**ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA**

**ARIGUANABO (Subcuenca Aeropuerto) , Datos sin error**



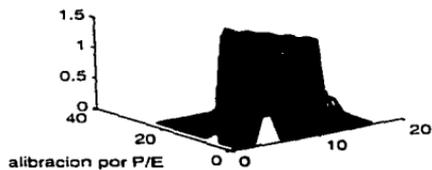
**Calibración Automática Sin Multiescala**

**Calibración Automática Con Multiescala**

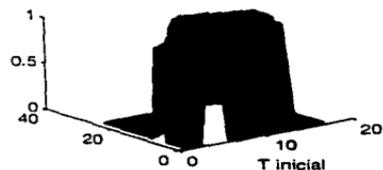


**Gráfica 4.9**

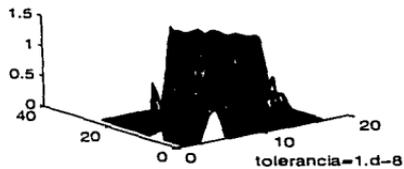
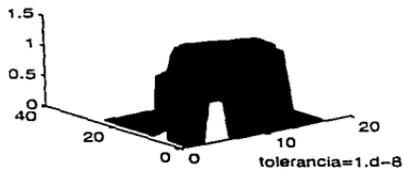
**ARIGUANABO (Subcuenca Aeropuerto) , Datos con error**



**Calibracion Automática Sin Multiescala**

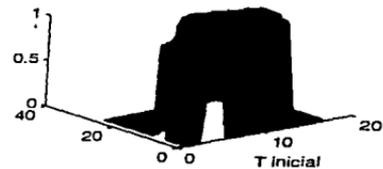
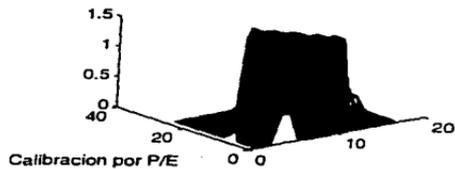


**Calibracion Automática Con Multiescala**



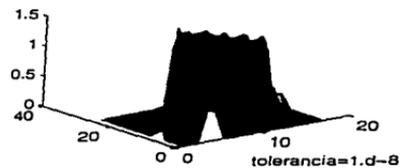
**Gráfica 4.10**

ARIGUANABO (Subcuenca Aeropuerto) , Datos con error



Calibracion Automática Con Multiescala

Calibracion usando Tijnov



Gráfica 4.11

## COMENTARIOS FINALES.

A partir de los resultados obtenidos hasta el momento, podemos hacer los siguientes comentarios:

- La optimización multiescala se presenta como una alternativa interesante y viable para tratar problemas inversos que cuando son mal planteados resultan difíciles de resolver directamente con los algoritmos usuales de optimización, especialmente cuando hay presencia de errores en los datos. Esto resulta claro en nuestro caso si se comparan los resultados obtenidos con multiescala y optimización directa, en términos de calidad y costo computacional.
- Multiescala funciona bien cuando el parámetro a estimar resulta ser constante por pedazos (acuifero real) o cuando su curvatura no muestra zonas de fuerte variación. De la opinión de expertos en hidráulica sabemos que ambos casos son usuales en la práctica.
- Aproximar el gradiente por diferencias finitas, como se ha hecho en este trabajo, hace mucho más caro todo el proceso de multiescala. Usando el paquete de diferenciación automática, ADIFOR2.0, se generó una rutina para evaluar el gradiente y en estos momentos contamos con una implementación del método de multiescala donde se tiene el gradiente exacto. Esta variante requiere de gran cantidad de memoria operativa por lo que no la hemos podido usar, a pesar de ser mucho más eficiente que la que usa gradiente por diferencias, debido a limitaciones en el equipo de cómputo. Para el trabajo futuro, que incluye el acuifero completo, será imprescindible adaptar el equipo de cómputo para poder usar esta implementación mencionada ya que disminuye notablemente el tiempo de proceso -aspecto de vital importancia cuando aumente la dimensión del problema-.

Con respecto a regularización, se trata de una teoría más hecha y estudiada, y en nuestro caso se ha mostrado exitosa una vez más, particularmente en el tratamiento del caso real. Sin embargo, la elección del parámetro de

regularización sigue siendo un problema no resuelto en general, lo que se ve claramente en el caso sintético, para el cual usando Tijonov no se obtienen buenos resultados. Debido a que multiescala *no depende de un parámetro*, esta alternativa para el problema que aquí estudiamos representa una opción más robusta.

En cuanto a cuestiones pendientes y posibles vías de trabajo y desarrollo futuro, podemos mencionar:

- Completar el estudio del acuífero de Ariguanabo.
- Estudiar el problema de estimación de parámetros distribuidos cuando el número de observaciones es pequeño.
- Llevar a cabo un estudio teórico con el objetivo de establecer resultados en los que se den condiciones sobre la aplicabilidad y convergencia del método de optimización multiescala; esto es, que permitan saber cuándo funciona y cuál es su eficiencia.
- Estudiar y establecer, dado el caso, posibles relaciones entre multiescala y regularización.

Hasta donde sabemos, este es el primer trabajo en el que se usa la optimización multiescala para resolver un problema inverso cuando hay errores en los datos.

## REFERENCIAS

- [1] Banks, H.T. y Kunish, K. (1989), *Estimation Techniques for Distributed Parameter System*, Birkhäuser, Boston.
- [2] Bear, J. (1972), *Dynamics of Fluids in Porous Media*, American Elsevier, New York.
- [3] Bear, J. y Verruijt, A. (1990), *Modeling Groundwater Flow and Pollution*, D. Reidel Publishing Company.
- [4] Byrd R. H, Lu P, Nocedal J. y Zhu C. (1995), A limited memory algorithm for bound constrained optimization, *SIAM J. Sci. Comput.* Vol. **17**, No. 5, 1190-1208.
- [5] Chang, S. y Yeh, W-G (1976), A proposed algorithm for the solution of the large-scale inverse problem in groundwater, *Water Resources Research*, Vol **4**, No. 3, 365-374.
- [6] Chardaire-Riviere C, Chavent G, Jaffre J. y Jun L. (1990), Multiscale representation for simultaneous estimation of relative permeabilities and capillary pressure, *Society of Petroleum Engineers*, SPE 20501.
- [7] Chardigny, E., Siegel, P., Mosé, R. y Ackerer, P.(1996), Parameter identification for complex groundwater systems, *XI International Conference on Computational Methods in Water Resources*, Cancún, México.
- [8] Chavent, G. y IRIA-Laboria (1979), Identification of distributed parameter system: about the output least square method, its implementation, and identifiability, *Proceedings of the 5<sup>th</sup> IFAC Symposium on " Identification and System Parameter Estimations"*, Pregamon Press.

- [9] Chavent, G. (1985), On parameter identifiability, *Proceedings of the 7<sup>th</sup> IFAC Symposium on Identification and System Parameter Estimations*, Pergamon Press.
- [10] Chavent, G y Liu, J. (1989), Multiscale parametrization for the estimation of a diffusion coefficient in elliptic and parabolic problems, *Preprints of IFAC Symposium on Control of Distributed Parameter System*, Perpignan, France.
- [11] Chavent, G, Clemente, F. y Gómez S.(1995), Migration-Based travel-time Inversion of 2D simple structures: the Synclay model, *Geophysics*.
- [12] Dennis, J. E. y Schnabel, R. B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.
- [13] Engl, H. W. (1987), Discrepancy principles for Tikhonov regularization of ill-posed problems leading to optimal convergence rates, *Journal of Optimization Theory and Applications*, **52**, No. 2, 209-215.
- [14] Engl H W, Kunish K y Neubauer A. (1989), Convergence rates for Tikhonov regularisation of non-linear ill-posed problems, *Inverse Problems* Vol **5**, 523-540.
- [15] Engl, H. W. (1993), *Inverse Problems*, Johannes-Kepler-Universität Linz, Austria.
- [16] Guerrero, J. (1993), Algoritmo de Gene H. Golub para el problema de cuadrados mínimos amortiguados: un estudio numérico y su aplicación en la resolución de ecuaciones integrales de Fredholm de primera clase, Tesis de Maestría en ciencias (Matemáticas), Facultad de Ciencias de la UNAM.
- [17] Liu, D. C. y Nocedal, J. (1989), On the limited memory BFGS method for large scale optimization, *Mathematical Programming*. **45**, 503-528.
- [18] Liu, J. (1993), A multiresolution method for distributed parameter estimation, *SIAM J. Sci. Comput.* Vol **14**, No 2, 389-405.

- [19] López E, J. (1988), Principio de Pseudo-optimalidad en la resolución de problemas mal-planteados por el método de regularización de Tikhonov, Tesis de Doctorado en Ciencias (Matemáticas), Facultad de Ciencias de la UNAM.
- [20] Morton, K. W. y Mayers, D. F (1994), *Numerical Solution of Partial Differential Equations*, Cambridge University Press.
- [21] Nocedal, J. (1980), Updating quasi-Newton matrices with limited storage, *Mathematics of Computation*, 35.
- [22] Seidman, T. I. y Vogel, C. R. (1989), Well posedness and convergence of some regularisation methods for non-linear ill-posed problems, *Inverse Problems* 5, 227-238.
- [23] Tikhonov, A. N. y Arsenin, V. Y. (1977), *Solutions of Ill-Posed Problems*, John Wiley & Sons.
- [24] Vainikko, G. ( ), Inverse problem of groundwater filtration: identifiability, discretization and regularization.
- [25] Wood, W. L. (1993), *Introduction to Numerical Methods for Water Resources*, Clarendon Press-Oxford.
- [26] Yeh, W-G. (1981), Aquifer parameter identification with optimum dimension in parameterization, *Water Resources Research*, Vol. 17, No. 3, 664-672.
- [27] Yeh, W-G. (1986), Review of parameter identification procedures in groundwater hidrology: the inverse problem, *Water Resources Research*, Vol. 22, No. 2, 95-108.
- [28] Zhu C, Byrd R H, Lu P. y Nocedal, J. (1994), L-BFGS-B Fortran sub-routines for large-scale bound constrained optimization, Northwestern University, Department of Electrical Engineering and Computer Science.