Departamento de Explotación y Prospección de Minas

ANÁLISIS Y RESOLUCIÓN NUMÉRICA DE UN PROBLEMA INVERSO EN GEOFÍSICA MEDIOAMBIENTAL. APLICACIÓN AL CASO DE LOS SONDEOS ELÉCTRICOS VERTICALES

TESIS DOCTORAL

José Paulino Fernández Álvarez

DIRECTORES

Juan Luis Fernández Martínez César Omar Menéndez Pérez

Oviedo, Julio 2004.



1 Abstract

This thesis analyzes an inverse problem in resistive geophysics arising in the so called Vertical Electrical Sounding technique (V.E.S). The Schlumberger method has been chosen because of its traditional extended use.

The resistive geophysical method sees the terrain as a strictly dissipative electric circuit, characterized by a certain zonal resistivity distribution. In the most common V.E.S version, the goal of the direct problem is to calculate the data from the values of the resistivity and thicknesses of the layers of a multilayered earth. Data are called apparent resistivities because they are averages—to different depths-of the real values of the resistivity distribution. The mathematical formulation of the direct problem implies a convolution integral, in which the input function depends in a nonlinearly way of the parameters. To calculate its value, the so called Pekeris recurrence relations can be used and an approximate discrete filter function obtained to calculate the data.

There is a practical need of making comparisons between predicted data and observed values and also between different kind of plausible terrains – looking at the differences among their characteristic parameters. To this end it is very convenient for the model and data sets to be given a certain algebraic structure: The vector space is very useful and allows additional and easy construction of norms from which induced metric distances are derived. The influence of the use of different kind of norms, their interpretation in a statistical framework and the way they can be utilized to define errors and measure relevant properties of the sought model, are explained.

Some thoughts about the same concept of inverse problem are given. The starting point is the more classic idea of the inverse problem as an estimation one. Only one model is calculated from a given set of data, and as these latter are subject of measurements errors, the estimated model can be described as a random variable. It happens that, in general, this r.v do has a big dispersion. Statistical frameworks—frequentists and Bayesians—do have sense to describe different aspects of the solution behaviour.

In the frequentist approach concepts as bias, variances and confidence intervals are used. Instead, characteristic of the second approach are the a posteriori distributions where a priori information and data knowledge are combined. These distributions can be approximated well enough to support decision making.

There is a need to correctly introduce a priori information into the inverse problem to control the solution erratic behaviour. It will be explained how to translate this kind of information in mathematical terms. This has been achieved in two ways:

- a) Using the Affine structure and the reparameterization possibilities –particularly through the use of logarithms- in the space model.
 - b) Using norms (and seminorms).

This analysis is carried over using tools borrowed from numerical algebra as singular value decomposition and generalized singular value decomposition. The first one handles only information pertaining data and problem structure while the second one includes also, and simultaneously, a priori information for regularization.

Different kinds of linear systems likely to appear are examined and illustrated with examples.

There are two options for the linear problem unknown: The model itself or the variations

of it. This decision greatly influences the way regularization (stabilization) through a priori information is produced. After having showed and examined the regularization techniques, they are compared in different cases.

Local optimization methods are problematic in the sense they can converge to a local minimum and get trapped there. One alternative is the use of the global probabilistic methods. After introducing them, two of them are explained: simulated annealing (S.A) and genetic algorithms (G.A). The algorithms needed have been programmed for numerical and pedagogical analysis.

Global optimization methods have proved their ability to locate the global minimum and also the – yet most useful- property of being able to sample the so called equivalence regions. These are the zones in the space model consistent both with the measured data and with the other information, to a certain tolerance. The possibility of using G.A to sample this area according to the a posteriori probability distribution is explored. S.A is used with Gibbs-Botzmann expression for acceptance function, for this purpose, as theoretical contrast background.

Characterization of the equivalences is also attempted. Its origin is analyzed and traditional rules mentioned and translated in algebraic terms, allowing identification of the former with the linear approximation. Afterwards a comparison is made between linear equivalence zones and nonlinear ones. The work done with a priori information allow to find linear equivalent models consistent with certain desirable restrictions.

2 Resumen

En este trabajo se pretende analizar un problema inverso, en geofísica resistiva, bajo la modalidad denominada Sondeos Eléctricos Verticales (S.E.V). En particular se ha escogido el método Schlumberger por su extendido uso y larga tradición práctica.

El método resistivo concibe el terreno como un circuito eléctrico, estrictamente disipativo, caracterizado por la distribución de las resistividades de las distintas zonas. En la versión más comúnmente conocida de los S.E.V, las capas de distinta resistividad se disponen estratificadamente y, a partir de los valores de sus espesores y resistividades se pretende calcular el dato. En este método los datos se denominan resistividades aparentes debido a su carácter de promedios —a distintas profundidades— de los valores correctos de resistividades del terreno. En este trabajo el problema directo se formula estableciendo (Koefoed) la relación entre los parámetros del modelo y las resistividades aparentes mediante una integral de convolución en la cual la función de entrada depende de modo no lineal de los parámetros. El valor de la misma puede ser calculado usando unas relaciones de recurrencia (Pekeris). El cálculo efectivo de la resistividad aparentes se realiza utilizando un filtro aproximado discreto que procede de la discretización de la citada integral.

La necesidad práctica de efectuar comparaciones entre los datos predichos y los observados y de comparar distintos tipos posibles de terrenos —contrastando sus parámetros identificativos—aconsejan, y así se explica, dotar de la estructura de espacio vectorial tanto al conjunto de datos como al de parámetros y proceder en ellos a la construcción de normas de las cuales se deriven la posibilidad de medir distancias. En este trabajo se examina la utilización de diferentes tipos de normas, su interpretación en caso de adoptar un enfoque estadístico y su uso para decidir la tipología de los errores (relativos, absolutos) y para cuantificar algunas características relevantes del modelo.

Se aportan reflexiones sobre el propio concepto de problema inverso. Se parte la idea más clásica que concibe la solución en términos de la estimación de un único modelo sobre el que debe basarse la toma de decisiones. Como para cada conjunto de medidas se obtiene un modelo solución y dado que los datos están afectados de error aleatorio, las estimaciones pueden ser consideradas como una variable aleatoria que, además, se caracteriza por poseer gran dispersión. Es por tanto pertinente recurrir a los marcos estadísticos –tanto frecuentistas como bayesianos- para describir algunos aspectos del comportamiento de la solución.

A la luz del primero de ellos, tendría sentido valorar sesgos, dispersiones e intervalos de confianza para cuantificar el grado de desviación de la estimación respecto de la solución no perturbada. En el segundo, las distribuciones a posteriori son el concepto matemático donde se combina la información a priori y las incertidumbres de los datos. Se plantea que el conocimiento de éstas es posible en el grado suficiente como para permitir el apoyo de la toma de decisiones.

El control de las inestabilidades precisa poder vehicular la información a priori para restringir las variaciones en la solución. En este trabajo se explica la traducción de dicha información en términos matemáticos. Esto se ha hecho de dos formas diferentes:

- a) utilizando la estructura de espacio afín y las posibilidades de reparametrización —en particular la logarítmica- del espacio de modelos.
 - b) Utilizando las normas (y seminormas).

En este trabajo se examina la estructura de los métodos locales de optimización, los algoti-

mos más comúnmente utilizados y se explica—en ese contexto- en qué consiste la linealización. Se observará cómo la matriz jacobiana de la aplicación diferencial de la aproximación lineal de cada etapa iterativa juega un papel esencial en el análisis de las inestabilidades citadas, lo que justificará la gran atención dedicada al análisis del problema lineal.

Dicho análisis se llevará a cabo utilizando herramientas del álgebra numérica como la descomposición en valores singulares y la descomposición generalizada en valores singulares. La primera sólo maneja información referente a los datos y a la estructura del problema. La segunda, además, permite manejar también la información a priori.

Se examinan las distintas tipologías de sistema lineal que pueden presentarse, ilustrándolas con ejemplos prácticos.

Hay dos posibilidades de escoger la incógnita del problema: O bien el propio modelo o bien el incremento del mismo. Esta elección tiene grandes repercusiones en la forma en la que la información a priori produce estabilización. Tras haber presentado y analizado los métodos de regularización, se realiza una comparativa para diferentes casos.

Los métodos locales de optimización tienen el problema de la posible convergencia a mínimos locales. La alternativa es utilizar los métodos globales de optimización. Tras realizar una breve introducción a los mismos, se analizan en detalle dos de ellos: simulated annealing (S.A) y algoritmos genéticos (A.G). En este trabajo se han puesto a punto los programas precisos para su análisis numérico.

Además de mostrar su capacidad de encontrar los mínimos locales, han demostrado la –aun más util- propiedad de muestrear las denominadas regiones de equivalencia. Estas son, por así decirlo, el conjunto de todas las soluciones compatibles con los datos obtenidos y otras informaciones, dentro de cierta tolerancia especificada. En este trabajo se explora la posibilidad de utilizar los algoritmos genéticos para proceder a un muestreo de esta zona de acuerdo con la distribución de probabilidades a posteriori. Para ello se ha utilizado el método de simulated annealing con la función de aceptación de Gibbs-Boltzmann como fondo teórico de contraste.

La caracterización del fenómeno de la equivalencia se aborda también en este trabajo. Se analiza su origen y se explican las reglas tradicionalmente enunciadas en términos algebraicos que permiten la identificación de aquellas con los resultados de la aproximación lineal. Posteriormente se realiza una comparación entre dichas zonas via aproximación lineal o via solución no lineal. El trabajo realizado con la información a priori permite encontrar modelos linealmente equivalentes que cumplan ciertas restricciones deseables.

3 Agradecimientos

Dar las gracias no es tan fácil como aparenta ser. Al reflexionar sobre cómo escribir estas líneas, no obstante mantener firmes y nítidas algunas ideas, he sido obligado a recorrer mental y sentimentalmente todo el abanico de posibilidades -hasta las imposibles-; desde el tierno e injusto agradecer todo a todos hasta el provocador, algo tentador, pero no menos injusto, no agradecer nada a nadie. Ambas modalidades tienen la ventaja -no desdeñable allá por el fin de una tesis- de escribirse rápido.

Son las circunstancias de la vida -mucho más que la hipotética entidad de las aportaciones científicas aquí ofrecidas- las que, en mi caso, otorgan extraordinaria importancia a este trabajo. Desde mis primeros esbozos de tesis doctoral, allá por el año 1995 he recorrido un sinuoso (a veces tortuoso y en ocasiones algo tétrico) camino vital aferrado a esto que, permítaseme, me gustaría calificar de pedacito de ciencia útil. Han sido tantos los recovecos y tan raros, a veces, que mi cantautor preferido ha pasado a ser Javier Krahe. A veces, quisiera describirme como si esta tesis fuese un gran globo aerostático viajando ora dirigido, ora al capricho del viento, conmigo dentro de la barquilla. Otras partes del periplo quedan, sin duda, mejor descritas con la imagen nada bucólica y sí, quizás, algo vulgar, de un perro que, aferrado desesperadamente a un hueso que necesita para comer, es arrastrado sin miramientos calle abajo con el consiguiente perjuicio físico y psicológico.

Desde el singular y privilegiado punto de vista del protagonista -o sea, el mío- he constatado la enorme influencia que el contexto, inmediato y no tan inmediato, tiene en la producción científica. En mi caso y, por lo que conozco también en otros, la decisión de comenzar este tipo de trabajos, el momento y el tiempo de llevarlos a cabo, la elección de la temática, el enfoque de los contenidos e incluso las originalidades, las aplicaciones y la utilidad de las mismas y, no menos importante, el mismo hecho de poder darlos por finalizados (que no acabarlos), dependen, hasta extremos insospechados de la historia previa, de las posibilidades materiales, del entorno afectivo, de la constitución física y psicológica, etc...

Para mí es, por tanto, más evidente que nunca que -generalizando la modesta tarea aquí expuesta- cualquier contribución científica es un producto social en todos sus aspectos y a todas las escalas. Hasta ahora, que yo sepa, este hecho de relevancia imponderable sólo se reconoce explícitamente en el, un tanto sentimental, capítulo de agradecimientos. Por eso me resulta difícil dar las gracias a alguien por algo. Creo que antes de comenzar la tesis, me hubiese sido más fácil.

También he podido apreciar lo acertado de la expresión: "el que mucho abarca poco aprieta" que había oído casi desde siempre. He adquirido cierta experiencia en el ejercicio de la misma a lo largo de los últimos años y he decidido correr el riesgo de aplicarla también en los agradecimientos mencionando sólo a dos personas. Extender más la lista, sería extenderla del todo y mi gratitud se diluiría tanto que alguien -yo mismo- podría confundirla con la irritante indiferencia que rezuma cualquier protocolo social.

Para consolarme por -y justificar- tanta omisión he urdido excusas variopintas. Como ninguna es tan ingeniosa que consiga disimularse, prefiero no mencionarlas. A las personas que sí cito aquí, las veo en mi cabeza formando parte de la tesis como yo mismo, habiendo permitido la generación social de este trabajo de la misma forma que yo me he encargado de la parte científica. Si no las menciono como autoras, junto a mí, es sólo por impedimento administrativo. Espero que haya momentos en los que, de palabra o hecho, exprese la medida

de mi gratitud a los ausentes de este capítulo.

Lo que haya de bueno en esta tesis se lo quiero dedicar a mi madre, Loli.

Yo siempre he rechazado los actos solemnes porque suelen ser rescoldos de un orden injusto, por dispendiosos, por ser superfluos a mi entender o, a veces, porque sí. Sin embargo, quisiera invitar a hacer una excepción en este caso, para conseguir que toda la solemnidad, reconocimiento social y prestigio que albergue el acto de convertirse en doctor, sitúe la contribución de esta mujer admirable a la posición que se merece y que nunca le ha sido reconocida.

A mi madre, naturalmente, la quiero. Este sentimiento que no desaparecerá nunca, impregna estas líneas. Pero la razón profunda de mi gratitud hacia ella en este momento es mucho más profunda: yo admiro a mi madre. Veo en ella a la hija de un momento histórico en el que el hecho biológico de nacer mujer añadía mucha más miseria al ya perjudicial hecho de pertenecer a una familia humilde. Los designios de esta sociedad clasista, patriarcal y nacionalcatólica la privaron arbitrariamente de beber de las fuentes de la libertad humana. Nunca tuvo, porque no, acceso a la cultura. También se le negó, de formas más o menos violentas y más o menos sutiles, la posibilidad de realización económica y personal a través del trabajo como servicio a uno mismo y a la comunidad. El pesado yugo religioso esculpía -con la violencia necesaria- en los cerebros de hombres y mujeres la imagen de la perfecta casada como única posición legítima y moralmente aceptable de cualquier mujer que no fuese monja ni ramera¹.

Yo no consigo imaginar la fuerza de voluntad, la intuición y la inteligencia que ha tenido que poner en juego esta mujer, representante para mí de las mujeres luchadoras que han sido, no ya para sobrevivir, sino para conservar actualmente la sensibilidad que le conozco, la apertura de mente y de ideas, la falta de prejuicios y la energía para tratar de imbuir en mí estos atributos tan contrarios a lo que ella vivió y sufrió.

Loli. La mujer. Mi madre. Me ha dotado de alas libres de prejuicios y lastres absurdos impuestos socialmente. También le debo en gran parte el aliento preciso y la capacidad de ilusionarme para echar a volar. Que en lugar de un vuelo majestuoso sólo se observen saltos desgarbados, es ya cosa mía.

Mamá, infatigable luchadora. Gracias por esta tesis. Te admiro y te quiero.

La segunda persona que quiero mencionar es también una mujer: María.

Estoy profundamente enamorado de ella. Aparte de innumerables motivos -racionales e irracionales- para ello, a María le debo, sin duda, haber finalizado este trabajo gustándome el resultado y no habiendo dejado de gustarme yo mismo. Su amor por el conocimiento riguroso y público, al servicio de un bien social real y sin tapujos y su intensa antipatía por el falso conocimiento, críptico, prepotente y al servicio de intereses criticables ha sido un regalo que, gracias a ella, me ha acompañado sin descanso.

Simultáneamente ha sabido ver -y hacerme ver a mí- qué teclas tocar para convertir este trabajo en un pedacito ilusionante de vida común. Corre el rumor de que ella misma quiere,

¹En la España de mediados de el siglo XX, la otra posible razón social de la mujer, ser bruja, había quedado obsoleta. Aprovecho para recomendar dos lecturas edificantes: "Casadas, monjas, brujas y rameras" de Manuel Fernández Alvarez (historiador de la R.A.H) y "Si Aristóteles levantara la cabeza", de María Angeles Durán (socióloga).

en su momento, hacer su tesis: ¡Ánimo, mi amor!. Yo, personalmente, quisiera decir a quien competa que las tesis de los enamorados debieran seguir un cauce administrativo diferente.

María, te amo. Gracias por este regalo.

Índice

1.	MO	TIVACIÓN Y OBJETIVOS
	1.1.	Motivación y marco general
	1.2.	Objetivos y aportaciones
2.	\mathbf{EL}	PROBLEMA DIRECTO EN S.E.V.
	2.1.	Referencia histórica sobre S.E.V
	2.2.	Breves fundamentos del método resistivo
	2.3.	Dispositivos de campo y procedimiento operativo
	2.4.	Naturaleza de la observación en el método de S.E.V
	2.5.	
		2.5.1. Planteamiento y solución analítica del modelo matemático multicapa con inyección puntual
		2.5.2. Las relaciones de Pekeris
		2.5.2. Cálculo de la resistividad aparente para dispositivos simétricos
		2.5.4. Cálculo de la resistividad aparente para dispositivos simétricos utilizando
		la teoría de filtros
		la teoria de lititos
3.	\mathbf{CO}	NSIDERACIONES GENERALES SOBRE PROBLEMAS INVERSOS
	3.1.	Concepto de problema inverso
	ъ.	DDODLEMA INMEDICA LIMBAL
Ł.		PROBLEMA INVERSO LINEAL
	4.1.	Motivación para el estudio de los problemas lineales
		4.1.1. Un ejemplo de obtención de soluciones equivalentes
	4.2.	
		4.2.1. Parametrizaciones
		4.2.2. El espacio de los datos
		4.2.3. El espacio de modelos y su parametrización
	4.3.	La matriz jacobiana
		4.3.1. Obtención de J_{ρ}
		4.3.2. Interpretación física de J_T
		4.3.3. La estructura de la matriz de sensibilidad
	4.4.	Distancia y norma. Definiciones y repercusión de su elección
	4.5.	Influencia del tipo de norma en el espacio de los datos
		4.5.1. Construcción de un modelo matemático de las medidas. Las distribuciones
		de probabilidad
		4.5.2. La familia de las Gaussianas generalizadas
		4.5.3. Estimación de los parámetros de la distribución de datos. La máxima
		verosimilitud
		4.5.4. Casos particulares relevantes: distribución normal y exponencial
		4.5.5. El caso de varios puntos de medida
	4.6.	
	1.0.	4.6.1. Una formulación matemática para la información a priori
	17	Clasificación motivada de los sistemas lineales
	せ. 1 .	

		4.7.1. Sistema determinado	81
		4.7.2. Sistema subdeterminado	87
		4.7.3. Sistema sobredimensionado o sobredeterminado	94
	4.8.	La descomposición en valores singulares	97
	4.9.	La estructura geométrica de las inestabilidades	101
		. Clasificación de los problemas discretos lineales	
		. Concepto de regularización	
	4.12.	. La información a priori: otro modo de expresión	105
		. La descomposición generalizada en valores singulares	
		. Regularización de Thikonov	
		. Análisis lineal de los efectos de la regularización de Tikhonov	
5.	EL I	PROBLEMA INVERSO NO LINEAL	125
•		Formulaciones del problema no lineal	
		Construcción de la solución	
	5.3.	Métodos locales de optimización	
	0.0.	5.3.1. El metodo de Newton	
		5.3.2. El método de Gauss-Newton	
		5.3.3. El método de gradiente o del descenso más rápido (steepest-descent)	
		5.3.4. El método de Levenberg-Marquardt	
		5.3.5. Linealización del funcional F	
		5.3.6. Métodos de cuasi-Newton	
	5.4.	Incrementos .vs. parámetros (creeping .vs. jumping)	
	0.1.	5.4.1. Métodos locales con regularización "interna"	
		5.4.2. Regularización en la formulación global	
	5.5.	Métodos de optimización global	
6.		▼	151
	6.1.	Introducción	
		6.1.1. Las reglas de equivalencia de Maillet. Límites de validez	
		6.1.2. Las curvas de Dar Zarrouk	
	6.2.	El análisis lineal de equivalencias	
		6.2.1. Caso I: modelos linealmente equivalentes sin restricciones	
		6.2.2. Caso II: cálculo de modelos linealmente equivalentes con restricciones	
		6.2.3. Interés de la matriz jacobiana en el caso de parametrización logarítmica .	168
7.	MÉ	TODOS DE OPTIMIZACIÓN GLOBAL	177
	7.1.	Introducción	177
	7.2.		177
		7.2.1. Descripción básica del funcionamiento de un A.G	177
		7.2.2. Funcionamiento detallado de los algoritmos genéticos y bases matemáticas	180
	7.3.	El método de recocido simulado (Simulated Annealing)	190
		7.3.1. Problemas combinatorios y algoritmos numéricos	190
		7.3.2. El algoritmo de recocido simulado	192
			102

	7.4.	Aplicación de los métodos globales a un problema medioambiental de toma de	
			197
		v 1	198
		7.4.2. Modelos sintéticos	199
		7.4.3. Descripción de las funciones test y resultados numéricos	200
	7.5.	Caso de estudio: un problema de intrusión salina	204
		7.5.1. El problema inverso	204
		7.5.2. Modelización de la intrusión salina de la zona de Águilas	204
8.	ESC	QUEMA GENERAL DE PROGRAMCIÓN DE PROBLEMAS INVER-	-
			209
	8.1.	Breve descripción de MATLAB	209
		8.1.1. Cajas de herramientas en MATLAB	209
		8.1.2. MATLAB y el tratamiento de matrices	209
		8.1.3. Un tipo de objeto MATLAB: las estructuras	210
	8.2.	Estructuras relevantes para conceptualizar un problema inverso	211
		8.2.1. Concepto y clasificación de los datos en un problema inverso	212
		8.2.2. Estructura de almacenamiento de datos en el caso de S.E.V	213
		8.2.3. Organización conceptual de la estructura de parámetros de modelo	214
		8.2.4. Las opciones operativas	216
		8.2.5. La estructura de opciones en S.E.V	217
9.	CO	NCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN	223
10	BIE	BLIOGRAFIA	229
		Artículos consultados	
		Libros consultados	
	10.2	2.51.50 00.12.00.00.00.00.00.00.00.00.00.00.00.00.00	
Ír	adic	e de figuras	
	1.	Configuración de electrodos lineal simétrica	15
	2.	Curvas tipo para suelos de tres capas	16
	3.	Ejemplo sintético: a) Tres puntos de medida (sin error) sobre curva de resistivi-	
		dades aparentes. b) Predicciones de tres modelos obtenidos por análisis lineal,	
		junto a las medidas sin error.	40
	4.	Distintas regiones de equivalencia lineal. Relación de la estimación del error lineal	
		con el no lineal	41
	5.	Medidas de resistividad aparente en tres puntos de la superficie y muestras corres-	
		pondientes de la transformada de resistividad necesarias para la convolución	43
	6.	Familia de Gaussianas generalizadas	57
	7.	Forma de las diferentes Gaussianas generalizadas que asignan una probabilidad	
		del 5% al suceso $x = 30$	58
	8.	Expresión de la información apriori en coordenadas naturales	69
	9.	Expresión de la información a priori en coordenadas logarítmicas	70
	θ.	Expression de la información a priori en coordenadas logarismicas	

11.	Relación entre intervalo de búsqueda y restricciones de proporcionalidad	74
12.	Parametrización logarítmica para diferentes valores de α	75
13.	Diferencia entre un caso de optimización sin restricciones y otro con restricciones	
	de intervalo y proporcionalidad	78
14.	Datos con ECM del 0.1%	82
15.	Nube de modelos equivalentes con ECM=0.1 %	82
16.	Diferentes histogramas de la nube de modelos equivalentes con ECM= 0.1%	84
17.	Datos con ECM del 1%	84
18.	Nube de modelos equivalentes con ECM del 1 %	85
19.	Diferentes histogramas de la nube de modelos equivalentes con ECM=1 %	85
20.	Datos reducidos con error ECM=0,1%	86
21.	Nube de modelos equivalentes con ECM=0,1 %	87
22.	Caso en que las medidas están tomadas en dos puntos cercanos	89
23.	Soluciones del sistema de norma mínima	89
24.	Modelos equivalentes en el caso sobredeterminado con error del 5 %	95
25.	Soluciones equivalentes con ecm=5 % en el caso sobredeterminado con ecuaciones	
	"casi" redundantes	96
26.	Interpretación geométrica de la solución de mínimos cuadrados	97
27.	Datos medidos sobre curva de resistividades aparentes	
28.	Curva de valores singulares	
29.	Curva de valores singulares y filtros para la regularización de Tikhonov	
30.	Dispersión en las soluciones para distintos valores de λ	
31.	Curva L. El modelo de referencia minimiza profundidades y resistividades	
32.	Curva L. La información a priori es un modelo determinado de referencia	
33.	Soluciones: regularización de Tikhonov para distintos valores de λ	
34.	Valores singulares generalizados y filtros de Tikhonov correspondientes	
35.	Funciones de base para d.g.v.s	
36.	Dispersión del contraste de resistividades en función de λ	
37.	Curva L. El modelo de referencia pretende atenuar el contraste de resistividades.	122
38.	Problema inverso de desintegración radiactiva: comparación entre la solución	
00.	obtenida por linealización y la del problema no lineal original hallada mediante	
	un método cuasi-Newton	127
39.	Problema subdeterminado: diferencias entre creeping y jumping al solucionar me-	
	diante la pseudoinversa	141
40.	Problema sobredeterminado. Diferencia entre creeping y jumping al solucionar	
	utilizando truncamiento	142
41.	Evolución del proceso de optimización con algoritmo local empleando truncamien-	
	to de valores singulares en cada subproblema lineal	143
42.	Evolución de algoritmo local, tipo "jumping" con regularización de Tikhonov	146
43.	Evolución de algoritmo local, tipo "jumping" con regularización de Tikhonov y	
	nuevo modelo de referencia	146
44.	Regularización global	147
45.	Cálculo de los factores de reducción al plano inferior de la transformada de resis-	
	tividad	152

46.	Zonas de equivalencia del 5% y 10% de error en resistividades aparentes: a) suelo	
		155
47.	Curva de Dar-Zarrouk para los modelos de cuatro capas: $\mathbf{m}_1 = (10, 1, 100, 3, 10, 1, 100, 10, 10, 10, 10, 10, 10, 10$))
	$\mathbf{y} \ \mathbf{m}_2 = (10, 0.97, 80, 5, 10, 0.9, 100) \dots$	157
48.	Medidas generadas sintéticamente	169
49.	Curva de valores singulares. Las incógnitas son las resistividades y los espesores .	170
50.	Curva de valores singulares. Las incógnitas son sólo las resistividades	171
51.	Vectores singulares del espacio de datos: incógnitas resistividades y espesores	171
52.	Vectores singulares del espacio de datos: las incógnitas son las resistividades	172
53.	Vectores singulares para los modelos: las incógnitas son resistividades y espesores	172
54.	Vectores singulares de los modelos: las incógnitas son las resistividades	173
55.	Función sombrero de bruja. Funciones de distribución teórica y empírica utilizan-	
	do Simulated Annealing y varias ejecuciones de A.G. Para la función Gaussiana se	
	obtienen resultados similares siendo el grafo de la función de distribución teórica	
	una línea recta	202
56.	Sombrero de bruja. Histograma marginal para el parámetro \boldsymbol{x} en la región del	
	75% de verosimilitud con A.G. (los resultados obtenidos para el parámetro y son	
	idénticos)	203
57.	Sombrero de bruja $(a=1,\ b=2,\ C=2)$. Evolución del sesgo y de la dispersión	
	(ECM) de muestras de AG para diferentes umbrales de verosimilitud (probabili-	
	dad de mutación del $80\%)$ y 10 poblaciones iniciales generadas aleatoriamente	203
58.	Intrusión salina. Isolíneas de la función de desajuste (resistividad frente a espesor).	
	Parámetros del modelo para errores menores del 5 $\%$	205
59.	Intrusión. Función de distribución de los errores.	205
60.	Distribución marginal a posteriori de resistividades (región con errores inferiores	
	al 5%)	206

Capítulo I: INTRODUCCIÓN

1. MOTIVACIÓN Y OBJETIVOS

1.1. Motivación y marco general

Para comprender plenamente el sentido de este trabajo es importante reseñar que se ha llevado a cabo en el contexto de un departamento en una escuela de ingeniería. De la experiencia obtenida en la actividad docente y de la participación directa en la práctica geofísica, así como mediante numerosas conversaciones mantenidas con profesionales del sector, parece desprenderse que el corpus de conocimientos realmente puesto en juego por una cantidad representativa profesionales de la geofísica en España se nutre, casi en exclusiva, de dos tipos de fuentes:

- de los textos clásicos proceden las bases teóricas que sustentan la práctica de campo y
- de la experiencia personal *in situ*, conformada a lo largo de los años. Este conocimiento adquirido se traduce, básicamente, en un conjunto de reglas empíricas que cada cual sigue protocolariamente.

Los primeros suelen abordar en profundidad la física del problema y la resolución matemática del problema directo en casos que admiten solución analítica, es decir, la obtención de predicciones supuestos conocidos los parámetros del subsuelo. Por contra, la resolución del problema inverso -cuyas incógnitas son precisamente los parámetros del terreno- suele presentarse casi a modo de ilustración pedagógica: muy resumida en términos relativos, tratando sólamente casos muy simplificados que admitan solución analítica y describiendo métodos gráficos -cuando existen- de solución (Telford, 1990, Orellana, 1982).

Los casos reales son muy variados, complejos, necesitados, para ser resueltos, de la integración de informaciones "blandas", atinentes a características geológicas observadas o esperadas y de ese elemento subjetivo o "artístico" cuya dosificación correcta es atributo definitorio del buen profesional. Parece que existe un trecho difícil de salvar entre los materiales de consulta disponibles y la tipología y rango de dificultad de las dudas a solventar.

Los artículos de investigación son, tradicionalmente poco consultados. A esto contribuye sin duda el hecho de estar escritos en inglés y su carácter, a veces, más críptico de lo que estrictamente debiera derivarse del grado de abstracción de los contenidos tratados. Además, el desorden e inseguridad intrínsecos a los frentes de exploración científica en plena ebullición no parecen encajar bien con un tipo de innovación, más pausada y conservadora, propio del profesional.

Entre el profundo sustrato rocoso de los saberes clásicos y la excesiva frondosidad de los bosques superficiales de la investigación reciente se sitúan, metafóricamente hablando, sedimentos relativamente actuales con suficiente grado de consolidación como para ser utilizados con extraordinario provecho. En esta región tan útil del saber geofísico se hallan los avances habidos en las últimas dos décadas en las técnicas analíticas y numéricas de resolución de problemas inversos. Dado que el uso de éstos es ineludible en los paquetes informáticos usados en la interpretación geofísica, es preciso investigar sobre cómo trasponer aquellos conocimientos, relativamente abstractos, a un nivel asequible a las dudas que emanan de la práctica.

En la actualidad han aparecido textos, que parecen compartir esta motivación, que hacen más hincapié en el problema inverso en geofísica. Algunos son, sin embargo, de amplia cobertura y tratan cada técnica considerando con fines pedagógicos e ilustrativos un caso práctico de

referencia, mostrando, más que explicando, cuáles son las expresiones matemáticas a emplear en cada uno de ellos. Si bien estas obras reconocen el uso casi imprescindible de programas informáticos, suelen sugerir el conocimiento necesario en uno o dos capítulos finales (p. ej., Sharma, 1997). Existen también monografías con el mismo espíritu (Orellana, 1982, p.ej).

Otras obras (Parker, 1994) se preocupan de la traducción en términos prácticos de las distintas formulaciones matemáticas concretas de los problemas inversos. No obstante, reconociendo su pertinencia, precisión, rigor e interés práctico, emplea niveles de abstracción y conceptos con nivel alto de complejidad, lo cual produce que a pesar de estar escrito por un matemático con gran perspectiva y gusto excepcional, pueda quedar relativamente lejos de las posibilidades de la formación de ingeniería que se imparten en España. Se trata de una obra paradigmática del campo de la matemática aplicada.

Otras obras como (Meju, 1994) sí hacen un esfuerzo por abordar de manera asequible técnicas numéricas, pero sigue siendo de amplio propósito general -o sea, abarca muchas técnicas- y está en inglés.

Parece pues pertinente intentar la realización de un trabajo, en castellano, **que investigue** cómo plantear el corpus de conocimientos ya disponible sobre resolución del problemas inversos de forma que resulten razonablemente operativos a la luz de la formación impartida en las ingenierías españolas. Básicamente esto exige lo siguiente:

- Formulación en términos comprensibles del concepto de problema inverso. La solución de un problema inverso es una idea flexible que puede adaptarse a distintas necesidades prácticas.
- Utilización como herramientas propias del álgebra numérica y del cálculo de varias variables. Nótese que quedaría excluído, en principio, el análisis funcional.
- Tratamiento limitado a problemas discretos, es decir, parametrizables en términos de un número finito de incógnitas y datos.
- En lugar de ir desde la descripción general del marco hasta el problema concreto, se propone que todos los razonamientos teóricos tengan una fuerte raíz en ejemplos de una técnica concreta.
- Papel central de la programación en su vertiente pedagógica. De ahí la utilización de Matlab como plataforma de construcción, organización de rutinas e ilustraciones de alta calidad.

Por otro lado, desde el punto de vista del autor de este trabajo, la geofísica es todavía una técnica infrautilizada. En los casos prácticos reales se suele pasar sin solución de continuidad de la cartografía geológica a la perforación de sondeos de investigación, prescindiendo de un uso serio de la geofísica de superficie como etapa intermedia, de coste también intermedio, que multiplicaría el rendimiento de los hipotéticos sondeos que, posteriormente y en su caso, tuviesen que realizarse. La causa de esta omisión podría deberse, quizás, al uso generalizado de programas comerciales en la modalidad de "caja negra", uso que obliga a adoptar un estado de resignación ante la solución finalmente ofrecida por el programa. Esta resignación se torna en abierta desconfianza, explícitamente expresada, cuando se observa que, casi con los mismos datos de campo, pueden obtenerse soluciones con grandes diferencias entre sí. La inestabilidad, junto

al desconocimiento de sus causas y estructura, hace aparecer la geofísica como algo errático, poco fiable.

La percepción de que para extraer algo de información útil con las limitaciones citadas sería necesario acceder a un nivel matemático inalcanzable, desanima un tanto a los profesionales.

Hace falta, por tanto, **examinar y presentar la metodología que permite**, con rigor, **incorporar informaciones complementarias** a las propias medidas del aparato utilizado, sean datos geológicos, sean intuiciones empíricas, para entender en qué medida afectan a la solución final y evitar tener la sensación de que ésta es un resultado "ad hoc". Es preciso hacer hincapié en el uso de la denominada información a priori.

Por otro lado, debe **investigarse dónde está el origen de las inestabilidades** observadas y del **problema de la no unicidad** de la solución, hacer un seguimiento riguroso pero intuitivo de la propagación de las mismas a través del método de solución escogido y traducir en términos prácticos cada etapa.

El punto de vista adoptado en este trabajo es que la inseguridad provocada por la presencia de soluciones equivalentes se torna en franca utilidad cuando se comprende la estructura del conjunto de dichas soluciones. Por tanto, es preciso estudiar si es posible y de qué modos prácticos se puede describir dicho conjunto para que sea posible cuantificar riesgos y tomar decisiones.

Algunas motivaciones adicionales serán expuestas de modo más concreto bajo el formato de objetivos, pero se quisiera reseñar que el espíritu que anima al autor y directores de esta tesis es colocar la primera piedra de un estudio metodológico de propósito general que pueda servir, no sólo para resolver problemas aplicados (aunque también) sino como laboratorio de investigación y docencia sobre problemas inversos para personas que posean formación ingenieril.

1.2. Objetivos y aportaciones

Tal como sugieren las motivaciones precedentes, los propósitos de este trabajo acomodan una doble vertiente: la pedagógica, con su posible originalidad y la propiamente innovadora de contenidos. Como objetivos de nivel más general podrían destacarse los siguientes:

- 1. Investigar, elaborar y exponer la estructura anatómica del problema inverso geofísico, de forma que pueda utilizarse para el estudio ordenado y sistemático de su comportamiento en casuísticas diferentes.
- 2. Construir una plataforma de exploración de problemas inversos geofísicos. Este objetivo se pretende realizar a dos niveles,
 - a) Investigar y proponer razonadamente y con propósito general el mejor modo de organizar los algoritmos necesarios.
 - b) Programar el conjunto adecuado de rutinas.

A un nivel de detalle inferior, son también pretensiones de este trabajo las siguientes:

1. Definir y proponer vías de expresión matemática de la información a priori.

Mención especial merecen los casos en que, o bien se exige que la solución pertenezca a un subconjunto determinado del espacio de modelos, o bien se especifican condiciones

que deba cumplir alguna característica interesante de la solución (por ejemplo su grado de heterogeneidad, etc...). En el primer caso, se aporta una formulación, con su correspondiente interpretación geométrica, donde se emplea el concepto de variedad lineal en un espacio afín y se utilizan las posibilidades de reparametrización de los espacios de datos y parámetros.

- 2. Aclarar la génesis del fenómeno de las equivalencias y analizar, en el caso de los S.E.V, la amplitud y la estrutura de las regiones equivalentes, observando las diferencias entre las formulaciones lineales y no lineales. Analizar igualmente el uso de la información a priori para reducir el número de soluciones posibles. En este trabajo se reducirá este problema a un problema geométrico de intersecciones en un espacio afín.
- 3. Presentar de forma ordenada y clara las metodologías de solución del problema inverso no lineal en S.E.V, a saber:
 - a) Métodos locales.

El análisis de las etapas lineales permite identificar el origen de las inestabilidades -el denominado mal planteamiento del problema inverso- y proponer métodos paliativos de las mismas. Estos se conocen con el nombre de **métodos de regularización**. Se analizan también sus efectos prácticos.

Se aporta asimismo una clasificación, dentro del enfoque del álgebra numérica, de las soluciones de sistemas lineales de ecuaciones, que generaliza la puramente analítica.

Se realiza e ilustra una comparativa razonada entre dos grandes métodos generales de solución de problemas inversos lineales: algoritmos "creeping" (o formulación en incrementos) y "jumping" (donde la incógnita es el propio modelo). Al empleo no bien valorado de la formulación en incrementos se atribuye, recientemente (Parker, 1994), comportamientos no deseados de la secuencia de soluciones.

b) Métodos globales de optimización.

Se presenta la teoría de estos métodos como alternativa a los anteriores para solventar el problema de la convergencia a óptimos locales.

Una aportación de este trabajo consiste en la utilización de estos métodos globales para explorar de forma eficiente el conjunto de soluciones equivalentes. Si bien Simulated annealing ya ha sido utilizado con tal fin (Bhattacharya et al, 2003 y Sen et al, 2000), aquí se utilizan por primera vez, que el autor tenga constancia en el dominio de la geofísica, los algoritmos genéticos para conseguir un muestreo adecuado del espacio de soluciones a fin de extraer consecuencias de índole probabilística.

4. **Programar en Matlab los algoritmos** correspondientes a todos los apartados anteriores:

En referencia a los métodos locales, mención especial merecen técnicas algebraicas novedosas como las descomposiciones en valores singulares (y su correlato generalizado), y la programación de diferentes métodos de regularización (o atenuación de la inestabilidad).

Un esfuerzo particular se ha realizado en relación con los métodos globales de optimización, estando actualmente operativos dos conjuntos de rutinas:

a) El método de simulated annealing: esta técnica se ha programado de modo que, aprovechando las capacidades matriciales de Matlab, permite manejar simultáneamente una familia de modelos. Se han programado distintas opciones de función de aceptación, de temperaturas iniciales y patrones de enfriamiento. Asimismo se ha hecho con los criterios de convergencia. Así dispuestas, las rutinas pueden ser usadas con gran versatilidad y cómodamente para llevar a cabo experimentación numérica o simulaciones.

b) Los **algoritmos genéticos**: programados con la misma filosofía y finalidad que el precedente bajo dos modalidades diferentes: codificación binaria y aritmética (real).

Cabe señalar que la organización diseñada respondiendo al objetivo general (2.a) permite incorporar con coste mínimo los parámetros y rutinas correspondientes a nuevos métodos de solución o modificaciones de parámetros de los ya existentes.

Asimismo, se destaca que el organigrama ideado es válido, con cambios sólo de la ley física, parámetros y, quizás, de alguna metodología de solución (por ejemplo de las técnicas algorítmicas) para cualquier otro método geofísico distinto de S.E.V.

5. Finalmente se aporta la aplicación del método ideado de muestreo de la distribución a posteriori, mediante algoritmos genéticos, a un caso de intrusión salina.

Capítulo 2: El problema directo en S.I	Ξ.V.

2. EL PROBLEMA DIRECTO EN S.E.V.

2.1. Referencia histórica sobre S.E.V

El método de sóndeos eléctricos verticales (S.E.V) lo aplicó Conrad Schlumberger por primera vez en 1912. El propósito de dicha técnica es investigar el cambio de la resistividad de las formaciones del subsuelo con la profundidad. Inicialmente, las incógnitas del problema (resistividades y espesores de las formaciones) eran halladas mediante la superposición gráfica de las curvas de campo sobre las curvas de referencia adecuadas existentes en un album de curvas teóricas (Compagnie Génèrale de Géophysique,1963; Orellana y Mooney,1966 y otros) o mediante el método del punto auxiliar (Bhattacharya y Patra, 1968; Orellana y Mooney, 1966 y otros).

Con la llegada de la potencia de cálculo digital y las innovaciones aparecidas en los métodos numéricos, se simplificó notablemente el proceso de cálculo del dato, denominado resistividad aparente, sobre un suelo multicapa. Desde el punto de vista numérico, quizás el trabajo más relevante haya sido llevado a cabo por Gosh (Gosh,1970, 1971a,b). Como se verá posteriormente, la expresión que permite calcular la resistividad aparente en distintos puntos de la superficie sobre un suelo multicapa supone el cálculo de una integral impropia que involucra funciones de Bessel. El mérito de Gosh estriba en haber convertido dicha integral en una integral de convolución dando entrada, por tanto, a la potencia de cálculo de la teoría lineal de filtros.

La interpretación automática iterativa de los datos de resistividad puede llevarse a cabo en el dominio de la resistividad aparente -datos medidos- (Inman, 1975; Inman et al, 1973; Johansen, 1977, etc...) o en el dominio de la función núcleo integral (Bichara y Lackshmanan, 1976; Marsden, 1973 y otros).

El problema inverso en el dominio del núcleo fue resuelto por vez primera por Schlichter (1933) utilizando las expresiones de Langer (1933) para el potencial en la superficie del terreno. Schlichter calculó la función núcleo primeramente a partir de la resistividad aparente y después, la distribución de conductividades a partir de la función núcleo. Gosh (1971,a,b) obtuvo un conjunto de coeficientes para transformar los datos de campo, resistividades aparentes, en valores de la función núcleo y viceversa mediante aplicación de la teoría de filtros.

2.2. Breves fundamentos del método resistivo

Las técnicas de prospección geofísica son metodologías de trabajo no destructivas como resultado de las cuales se pretende obtener una suerte de imagen de la distribución de cierta propiedad física de interés en el subsuelo. Se precisan varias etapas para llevar a cabo una campaña geofísica (Plata, 2000), a saber:

- planteamiento o definición del problema geológico;
- planificación de la campaña geofísica que implica la elaboración de un modelo conceptual a comprobar y la elección de un método conveniente al objetivo perseguido;
- adquisición de medidas;
- procesado de los datos para ser puestos en condiciones de interpretación (a veces esta fase es bastante compleja);

 y, finalmente, la interpretación, es decir, traducción de los resultados en términos adecuados para la toma de decisiones (caracterización de litologías, presencia o ausencia de cierto tipo de fluidos, quimismo, contaminaciones, etc).

Cada tipo de método geofísico explora distintas propiedades físicas del terreno. Los métodos eléctricos permiten investigar las propiedades del subsuelo que afectan a la propagación de corrientes eléctricas y al valor de los campos que éstas generan. Existen numerosos métodos electricos con gran cantidad de aplicaciones prácticas. Este trabajo, centra toda la atención en un método eléctrico resistivo de corriente contínua¹. Idealmente, en este caso, la única modificación que sufre la energía eléctrica cuando se introduce en el terreno se debe a procesos de disipación². Por tanto, se considera el subsuelo como un circuito en el que existe cierta distribución espacial de resistencias eléctricas. La ley constitutiva que se supone aplicable a escala local para la deducción de las ecuaciones que rigen la distribución de corrientes es la ley de Ohm.

Parámetros importantes que es preciso definir para uso posterior son:

- 1. **Densidad de corriente**: es la intensidad (carga eléctrica por unidad de tiempo) de corriente que pasa a través de la unidad de área considerada perpendicularmente a la dirección del flujo. Se simboliza por $\mathbf{j} \left[A \cdot m^{-2} \right]$.
- 2. Resistividad: si se considera la expresión de la ley de Ohm para el caso de un conductor cilíndrico de longitud L y sección transversal S, la resistencia R que el mismo ofrece al paso de la corriente viene dada por la expresión:

$$R = \rho \frac{L}{S}.\tag{1}$$

Por tanto, si L=S=1, se tiene que $R=\rho$ $(\Omega\cdot m)$. El interés de trabajar con ρ en lugar de con R es que, como puede verse en la expresión (1) la resistencia depende de las dimensiones del conductor y, por tanto, del camino recorrido por las cargas eléctricas, mientras que la resistividad ρ sólo depende de la naturaleza del medio conductor. Por tanto, es ρ y no R quien puede correlacionarse (en su caso) con propiedades como litologías o características químicas de cierto fluido presente en el terreno (por ejemplo, una intrusión salina). A veces se utiliza la magnitud inversa, la conductividad, cuya unidad es Siemens/m.

El resultado importante de la aplicación del método resistivo es la obtención de la distribución de la resistividad en el subsuelo. Sin embargo, no es ésta la finalidad de la aplicación del método. El objetivo último, conseguido en la fase de interpretación, es establecer un vínculo entre la citada distribución de resistividades y una distribución de tipologías de terreno, para proceder a la toma de decisiones. Por lo tanto, la imagen de resistividades resultante deberá ser transformada en una imagen geológica o relacionarla con la presencia de cierto tipo de fluidos. En principio, parece necesario que distintos ambientes reales (geológicos, hidrogeológicos, etc) se caractericen por valores suficientemente diferentes de resistividad como para que su diferenciación sea posible. En la tabla 1 se ofrecen algunos datos orientativos al respecto:

¹El uso de corriente alterna sufre lo que se denomina el 'efecto piel'. Básicamente consiste en que la densidad de corriente alterna decrece exponencialmente con la profundidad en condiciones donde la contínua se matendría uniforme. Sólo en frecuencias muy bajas es este efecto despreciable.

²No se consideran relevantes, por tanto, los fenómenos inductivos o capacitivos.

MATERIAL	RESISTIVIDAD	CONDUCTIVIDAD
	$(\Omega \cdot \mathrm{m})$	(Siemens/m)
Rocas ígneas y metamórficas		
Granito	$5 \cdot 10^3 - 10^6$	$10^{-6} - 2 \cdot 10^{-4}$
Basalto	$10^3 - 10^6$	$10^{-6} - 2 \cdot 10^{-4}$
Esquistos	$6 \cdot 10^2 - 4 \cdot 10^7$	$2.5 \cdot 10^{-8} - 1.7 \cdot 10^{-3}$
Mármol	$10^2 - 2.5 \cdot 10^8$	$4 \cdot 10^{-9} - 10^{-2}$
Cuarcitas	$10^2 - 2.5 \cdot 10^8$	$5 \cdot 10^{-9} - 10^{-2}$
Rocas sedimentarias		
Areniscas	$8 - 4 \cdot 10^3$	$2.5 \cdot 10^{-4} - 0.125$
Pizarras	$20 - 2 \cdot 10^3$	$5 \cdot 10^{-4} - 0.05$
Calizas	$50 - 4 \cdot 10^2$	$2.5 \cdot 10^{-3} - 0.02$
Suelos y agua		
Arcillas	1 - 100	0.01 - 1
Aluvial	10 - 800	$1,25 \cdot 10^{-3} - 0,1$
Agua subterránea (dulce)	10 - 100	0.01 - 0.1
Agua de mar	0,2	5
Compuestos químicos		
Hierro	$9,074 \cdot 10^{-8}$	$1,102 \cdot 10^7$
Cloruro potásico 0.01M	0,708	1,413
Cloruro sódico 0.01M	0,843	1,185
Ácido acético 0.01M	6,13	0,163
Xyleno	$6,998 \cdot 10^{16}$	$1,429 \cdot 10^{-17}$

Tabla 1: resistividades características de distintos materiales (Loke, 1999).

Inspeccionando dicha tabla, se observa que las rocas ígneas y metamórficas tienen, en general, los valores de resistividad más altos, dependiendo los mismos altamente del grado de fracturación y del porcentaje de fracturas rellenas de agua. Las rocas sedimentarias, que generalmente tienen más porosidad y un contenido de agua superior, normalmente tienen valores de resistividad más bajos. Aún inferiores son los valores asignados a suelos húmedos y agua. Los rangos de resistividad entre los diferentes tipos de suelo se solapan, debido a que la resistividad de un tipo de material depende también de otros factores como la fracturación, el grado de saturación en agua y la concentración en sales disueltas. La resistividad del agua subterránea, por ejemplo, varía mucho en función del contenido en sales. Lo interesante es que las diferencias de resistividad que se manifiestan entre una arcilla y una arena embebida en agua, suelen posibilitar la detección por métodos eléctricos de un acuífero arenoso entre hastiales arcillosos. También la técnica resistiva es un buen método para delinear la interfase agua dulce-agua salada en acuíferos costeros.

Cabe reseñar que los valores de resistividad se extienden en un rango de valores mucho más amplio que el de cualquier otra magnitud física que intervenga en algún otro método geofísico. Por ejemplo, los valores de la densidad en un sondeo gravimétrico no suelen variar en más de un factor dos y las velocidades sísmicas suelen cambiar en un rango de 10. Esto hace de los métodos eléctricos (y electromagnéticos) técnicas muy versátiles.

La existencia de los mencionados solapes de rangos y la notable amplitud de los mismos aleja el proceso de interpretación de la pretendida objetividad y automatismo con que a veces se lo

trata. Más que los valores en sí, son muy informativos los contrastes de valores, las tendencias abarcando cierto rango espacial y los patrones consistentes que puedan observarse.

2.3. Dispositivos de campo y procedimiento operativo

El objetivo de los sondeos eléctricos verticales (S.E.V) es obtener la evolución de la resistividad con la profundidad. Es decir, no se busca -como ocurre con los perfilajes laterales o calicatas eléctricas- la caracterización de las diferencias que existen lateralmente entre los materiales del subsuelo sino, de manera similar a como ocurre en los sondeos mecánicos, describir las variaciones de resistividad a medida que se exploran terrenos más profundos.

La corriente se introduce en el terreno mediante dos electrodos (denominados por ello de corriente). Existen también dos electrodos de potencial, pudiendo adoptar los cuatro una gran variedad de posiciones relativas en superficie. Cada una de estas configuraciones electródicas se denomina dispositivo. Los diversos dispositivos reciben nombre distintos: Wenner, dipolo, Schlumberger, etc...

El empleo de uno u otro dispositivo depende del tipo de problema y de las circunstancias del entorno. Como cambia su posición relativa, varían las líneas de flujo de la corriente en el subsuelo y cambia la densidad de corriente en cada punto así como el grado de cobertura y la profundidad de investigación. La sensibilidad al ruido es diferente también.

En el método descrito en este trabajo, denominado $m\acute{e}todo$ Schlumberger, los cuatro electrodos se posicionan en superficie sobre una línea recta: los dos de corriente en el exterior (en el interior, por tanto, los de potencial) simétricamente respecto a un punto central. La corriente se inyecta y extrae por los electrodos de corriente y se mide la diferencia de potencial entre los dos electrodos centrales que están, en términos relativos, muy próximos. Se obtiene información sobre capas cada vez más profundas incrementando la distancia entre los electrodos de corriente (C_1C_2) mientras que los electrodos de potencial se mantienen en el mismo lugar. Cuando se hayan alejado tanto los electrodos de corriente que la ratio de la distancia entre electrodos de potencial (P_1P_2) respecto a los de corriente (C_1C_2) sea demasiado baja, se procede a distanciar los electrodos de potencial. De no hacerlo así, los voltajes medidos serán demasiado bajos y el error relativo cometido en su medida se haría intolerable³.

2.4. Naturaleza de la observación en el método de S.E.V

El procedimiento operativo especifica, pues, que introducida una corriente de intensidad I por uno de los electrodos de corriente, debe medirse la diferencia de potencial ΔV entre los electrodos centrales de potencial. El cociente $\frac{\Delta V}{I}$ es una resistencia eléctrica y la experiencia realizada sobre el terreno es, en cierto modo, idéntica a una que se hubiese llevado a cabo en un cable homogéneo (por ejemplo de cobre) introduciendo I por uno de sus extremos y midiendo la diferencia de potencial ΔV entre dos puntos cualesquiera del mismo.

Con este símil es fácil darse cuenta de que, si la corriente introducida se mantiene constante, el valor de ΔV y, por tanto, del cociente $\frac{\Delta V}{I}$ cambiaría si se alteran los puntos de inyección

³Al comienzo de la serie de medidas la ratio $r = \frac{MN}{AB}$ puede ser del orden de $\frac{1}{3}$, debiéndose modificar la distancia MN cuando r sea del orden de $\frac{1}{20}$ a $\frac{1}{50}$.

de corriente o de medida del potencial incluso sobre un terreno homogéneo. Esto ocurre porque cambia la geometría del circuito recorrido por las corrientes y, por tanto, su energía disipada por cada culombio de carga y por cada segundo. En el método Schlumberger esto ocurre cada vez que se aumenta la distancia C_1C_2 para conseguir mayor penetración en el terreno. Por tanto, en el valor de $\frac{\Delta V}{I}$ obtenido con cada separación C_1C_2 no sólo influyen las características conductivas del terreno recorrido por la corriente, sino otros factores geométricos derivados del cambio del punto de inyección.

Para que las sucesivas medidas de un sondeo resulten comparables es preciso corregir el valor del cociente $\frac{\Delta V}{I}$ mediante un factor geométrico⁴. En el caso del método Schlumberger (figura 1) el valor de dicho factor es:

$$k = \frac{\left(s^2 - b^2\right)}{4bs} \underset{b \to 0}{\to} \frac{s}{4b}.$$
 (2)

donde $s = \frac{C_1 C_2}{2}$ y $b = \frac{P_1 P_2}{2}$.

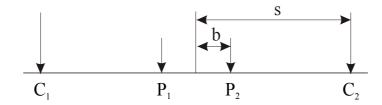


Figura 1: Configuración de electrodos lineal simétrica

Cuando se efectúan las medidas sobre un terreno homogéneo se puede calcular la resistividad real (salvo errores de medida) del suelo, directamente a partir de cada una de las medidas realizadas en cada posición de los electrodos de corriente mediante⁵:

$$\rho = k \frac{\Delta V}{I} \tag{3}$$

Sin embargo, se pretende que la técnica sea útil precisamente en terrenos no homogéneos donde las variaciones de resistividad importantes se asume que se producen sólo en función de la profundidad y no lateralmente. En este caso se sigue haciendo exactamente la misma operación indicada en (3) pero, conceptualmente, el resultado es una resistividad aparente $\rho_{ap}(s)$. Para cada espaciado entre electrodos de corriente se obtendrá, pues, un valor de

$$\rho_{ap}(s) = k(s) \frac{\Delta V(s)}{I} \tag{4}$$

Como se emplea una expresión válida sólo para terrenos homogéneos, el resultado, $\rho_{ap}(s)$, es la resistividad que tendría un subsuelo homogéneo (naturalmente no existente) que, respecto a esa posición de medida (con idéntica separación de electrodos), produjese la misma diferencia de potencial que ha sido observada al inyectar idéntica intensidad de corriente.

⁴Naturalmente éste cambia con el tipo de dispositivo.

⁵En un suelo homogéneo si la medida no tuviese error sería suficiente con hacer una sóla.

Se comprende que ρ_{ap} varíe con $s=\frac{C_1C_2}{2}$ ya que, al aumentar s, nuevas capas de terreno que antes no eran recorridas por la corriente introducida, ahora sí resultan afectadas. El valor de ρ_{ap} es, por tanto, una especie de media ponderada⁶ de las resistividades de las capas afectadas por la medida. Los datos de campo que resultan se grafican en la curva de ρ_{ap} frente a valores de $s=\frac{C_1C_2}{2}$. Suele representarse en ejes logarítmicos. Hay que resaltar que, aunque no se representen resistividades verdaderas frente a profundidades reales la forma de la curva y sus valores resultan afectados precisamente por éstos valores, permitiendo distinguir distintas tipologías. Para ilustrar la apariencia de las curvas de resistividades aparentes, se aporta un ejemplo -véase figura 109-.

Estos cuatro ejemplos aportados, generan una situación de relevancia práctica conocida por equivalencia (en el caso de las curvas de tipos H y K) y como supresión en los tipos A y Q.

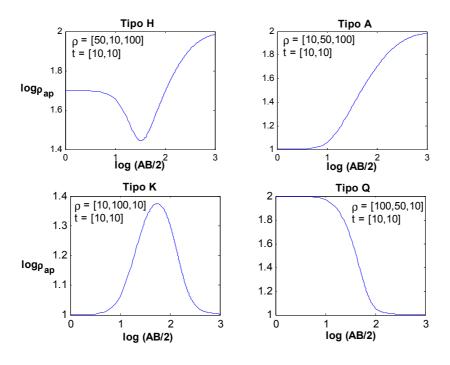


Figura 2: Curvas tipo para suelos de tres capas

2.5. El modelo matemático en S.E.V

Para poder realizar un tratamiento cuantitativo se sustituye el terreno existente por una abstracción conceptual capaz de ser expresada en términos de ecuaciones matemáticas. En el caso de los S.E.V las **hipótesis** adoptadas son las siguientes:

• el subsuelo consiste en un número finito de capas separadas por una frontera plana y

⁶Las potencias de las capas influyen en la ponderación.

horizontal. Todas las capas poseen un espesor finito, salvo la última, que se considera de potencia infinita;

- la extensión lateral de las capas es infinita;
- todas las capas son homogeneas e isótropas desde el punto de vista de su resistividad;
- todas las inyecciones de corriente son puntuales. Los únicos puntos de entrada de corriente al terreno son los electrodos habilitados para tal fin, cuyas dimensiones son despreciables a la escala de la experiencia;
- la corriente es continua.

2.5.1. Planteamiento y solución analítica del modelo matemático multicapa con inyección puntual

Inicialmente se plantea el caso en que sólo existe, en superficie, una única fuente puntual de inyección de corriente. Considérese un punto -distinto al de inyección- en el interior de cualquiera de las capas homogéneas que componen el medio estratificado. La ecuación que rige el potencial eléctrico, denotado por V, es la ecuación de Laplace (conservación de la carga). Debido a que se ha supuesto un único punto de inyección de corriente y a las hipótesis de isotropía y homogeneidad en cada capa, el potencial tiene una simetría axial respecto al eje vertical OZ que pasa por el punto de inyección, de modo que el uso de coordenadas cilíndricas resulta recomendado. La ecuación de Laplace en dicho sistema de coordenadas se escribe:

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial z^2} + \frac{1}{r^2} \frac{\partial^2 V}{\partial \theta^2} = 0,$$

que, debido a la simetría del potencial respecto al eje OZ, se reduce a

$$\frac{\partial^2 V}{\partial r^2} + \frac{1}{r} \frac{\partial V}{\partial r} + \frac{\partial^2 V}{\partial z^2} = 0. \tag{5}$$

La solución de la ecuación (5) puede obtenerse mediante el método de separación de variables:

$$V(r,z) = U(r)W(z). (6)$$

Sustituyendo (6) en (5) y dividiendo por V se obtiene:

$$\frac{1}{U(r)}\frac{d^{2}U}{dr^{2}} + \frac{1}{U(r)r}\frac{dU}{dr} + \frac{1}{W(z)}\frac{d^{2}W}{dz^{2}} = 0.$$

Dada la forma de la ecuación, ésta se satisface si y sólo si

$$\frac{1}{U\left(r\right)}\frac{d^{2}U}{dr^{2}} + \frac{1}{U\left(r\right)r}\frac{dU}{dr} = -\lambda^{2},\tag{7}$$

у

$$\frac{1}{W}\frac{d^2W}{dz^2} = \lambda^2,\tag{8}$$

donde λ es una constante real arbitraria. Dos soluciones linealmente independientes de la ecuación (8) son:

$$W_1(z) = e^{-\lambda z}, \ W_2(z) = e^{+\lambda z}.$$

La ecuación diferencial (7) es una ecuación de Bessel de orden cero y, por tanto, una solución particular es la función de Bessel de orden cero J_0 :

$$U_1 = J_0(\lambda r),$$

con lo que las soluciones particulares de la ecuación (5) son:

$$V_1 = e^{-\lambda z} J_0(\lambda r), \ V_2 = e^{+\lambda z} J_0(\lambda r)$$

Como cualquier combinación lineal de ellas resuelve también la ecuación diferencial, la solución general de (5) se puede expresar como:

$$V(r,z) = \int_0^\infty \left[\Phi(\lambda)e^{-\lambda z} + \Psi(\lambda)e^{+\lambda z} \right] J_0(\lambda r) d\lambda, \tag{9}$$

donde $\Phi(\lambda)$ y $\Psi(\lambda)$ son funciones arbitrarias de λ , cuya expresión particular depende de las condiciones de contorno. La intensidad I, introducida en el término fuente puntual puede ser explícitamente tenida en cuenta. En un medio homogéneo el potencial por ella generado es:

$$V = \frac{\rho_1 I}{2\pi\sqrt{r^2 + z^2}}. (10)$$

donde ρ_1 es la resistividad del suelo homogéneo (coincidente con la de la capa superficial del medio estratificado). Teniendo en cuenta la relación (integral de Lipschitz):

$$\int_0^\infty e^{-\lambda z} J_0(\lambda r) d\lambda = \frac{1}{\sqrt{r^2 + z^2}},$$

conocida en la teoría de funciones de Bessel, el potencial (10) se expresa como:

$$V = \frac{\rho_1 I}{2\pi} \int_0^\infty e^{-\lambda z} J_0(\lambda r) d\lambda. \tag{11}$$

Sumando y restando la expresión (11) a la solución general (9), aparece la solución general del problema no homogéneo, expresada como suma de un potencial primario, expresión (11) y un potencial perturbado

$$V = \frac{\rho_1 I}{2\pi} \int_0^\infty \left[\Theta(\lambda) \ e^{-\lambda z} + X(\lambda) \ e^{+\lambda z} \right] J_0(\lambda r) d\lambda,$$

en la forma

$$V = \frac{\rho_1 I}{2\pi} \int_0^\infty \left[e^{-\lambda z} + \Theta(\lambda) \ e^{-\lambda z} + X(\lambda) \ e^{+\lambda z} \right] J_0(\lambda r) d\lambda, \tag{12}$$

donde $\Theta(\lambda)$ y $X(\lambda)$ son funciones arbitrarias de λ .

Esta solución es válida para cualquier punto situado en el interior de cualquiera de las capas homogéneas del suelo estratificado. Es preciso ahora imponer las siguientes condiciones de comportamiento en los contactos entre capas (de transmisión) y de contorno:

1. Continuidad del potencial eléctrico en el contacto entre capas:

$$V_i(h_{i-1}) = V_{i-1}(h_{i-1}) \tag{13}$$

2. Continuidad del flujo eléctrico en los contactos intercapa. Siendo j_z la componente vertical de la densidad de corriente:

$$\mathbf{j_z} = -\sigma \frac{\partial V}{\partial \mathbf{z}} \tag{14}$$

$$\mathbf{j}_{\mathbf{z},i-1}(h_{i-1}) = -\frac{1}{\rho_{i-1}} \left(\frac{\partial V}{\partial z} \right)_{z=h_{i-1}} = \mathbf{j}_{\mathbf{z},i}(h_{i-1}) = \frac{1}{\rho_i} \left(\frac{\partial V}{\partial z} \right)_{z=h_{i-1}}$$
(15)

3. Comportamiento en la superficie del terreno: la componente vertical de la densidad de corriente debe anularse excepto en un entorno de la fuente de corriente, debido a que el aire se consiera como totalmente aislante.

$$\left(\frac{\partial V}{\partial z}\right)_{z=0} = 0. \tag{16}$$

4. Comportamiento en el origen: cerca de la fuente de corriente el potencial debe tener el siguiente comportamiento asintótico:

$$\lim_{r,z\to 0} V(r,z) = \frac{1}{\sqrt{r^2+z^2}}.$$

5. Comportamiento en el infinito: a profundidad infinita el potencial debe aproximarse a cero:

$$\lim_{z \to \infty} V(z) = 0$$

Aplicando la condición (1) en la frontera de separación de las capas $i \in i+1$, situada a la profundidad h_i , se obtiene:

$$\Theta_i(\lambda)e^{-\lambda h_i} + X_i(\lambda)e^{+\lambda h_i} = \Theta_{i+1}(\lambda)e^{-\lambda h_i} + X_{i+1}(\lambda)e^{+\lambda h_i}$$
(17)

La continuidad en los flujos (2) implica que:

$$\frac{1}{\rho_i} \int_0^\infty \left[\left\{ 1 + \Theta_i(\lambda) \right\} e^{-\lambda h_i} - X_i(\lambda) e^{+\lambda h_i} \right] J_0(\lambda r) \lambda d\lambda =
= \frac{1}{\rho_{i+1}} \int_0^\infty \left[\left\{ 1 + \Theta_{i+1}(\lambda) \right\} e^{-\lambda h_i} - X_{i+1}(\lambda) e^{+\lambda h_i} \right] J_0(\lambda r) \lambda d\lambda.$$

La satisfacción de esta ecuación se satisfaga para todo valor de r implica la igualdad de integrandos:

$$\frac{1}{\rho_i} \left[\{ 1 + \Theta_i(\lambda) \} e^{-\lambda h_i} - X_i(\lambda) e^{+\lambda h_i} \right] = \frac{1}{\rho_{i+1}} \left[\{ 1 + \Theta_{i+1}(\lambda) \} e^{-\lambda h_i} - X_{i+1}(\lambda) e^{+\lambda h_i} \right]. \tag{18}$$

La condición (3) restringe la expresión de $\left(\frac{\partial V}{\partial z}\right)_{z=0}$ y, por tanto:

$$\int_0^\infty \left[1 + \Theta_1(\lambda) - X_1(\lambda)\right] J_0(\lambda r) \lambda d\lambda = 0. \tag{19}$$

como

$$\int_0^\infty J_0(\lambda r)\lambda d\lambda = 0,\tag{20}$$

basta requerir que los dos últimos términos del integrando, correspondientes al campo perturbado, cumplan:

$$\Theta_1(\lambda) = X_1(\lambda). \tag{21}$$

La condición (4) se satisface automáticamente en la expresión del potencial primario. La condición (5) requiere que en la capa más profunda, indicada con el subíndice n, la función X valga cero, suprimiendo así la influencia del factor $e^{+\lambda z}$. Por tanto:

$$X_n(\lambda) = 0. (22)$$

El conjunto de ecuaciones (17)-(22) permite plantear un sistema de 2n ecuaciones con 2n funciones incógnita, $\Theta_i(\lambda)$ y $X_i(\lambda)$. Dado que el interés práctico se centra en la capa superficial, fundamentalmente hay que calcular $\Theta_1(\lambda) = X_1(\lambda)$. Si se introduce la notación:

$$u_{i} = e^{-\lambda h_{i}},$$

$$v_{i} = \frac{1}{u_{i}} = e^{+\lambda h_{i}},$$

$$p_{i} = \frac{\rho_{i}}{\rho_{i+1}}.$$
(23)

El citado sistema de ecuaciones es:

Su solución para Θ_1 se obtiene aplicando la regla de Cramer. De acuerdo con ella, Θ_1 es el cociente de dos determinantes:

$$\Theta_1 = \frac{N}{D},\tag{25}$$

donde

$$D = \begin{vmatrix} (u_1 + v_1) & -u_1 & -v_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ (v_1 - u_1) & +p_1u_1 & -p_1v_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & u_2 & +v_2 & -u_2 & -v_2 & \dots & 0 & 0 & 0 \\ 0 & -u_2 & +v_2 & +p_2u_2 & -p_2v_2 & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & u_{n-1} & +v_{n-1} & -u_{n-1} \\ 0 & 0 & 0 & 0 & 0 & \dots & -u_{n-1} & +v_{n-1} & +p_{n-1}u_{n-1} \end{vmatrix}$$
(26)

у

$$N = \begin{vmatrix} 0 & -u_1 & -v_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ (1-p_1)u_1 & +p_1u_1 & -p_1v_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & u_2 & +v_2 & -u_2 & -v_2 & \dots & 0 & 0 & 0 \\ (1-p_2)u_2 & -u_2 & +v_2 & +p_2u_2 & -p_2v_2 & \dots & 0 & 0 & 0 \\ \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & u_{n-1} & +v_{n-1} & -u_{n-1} \\ (1-p_{n-1})u_{n-1} & 0 & 0 & 0 & 0 & \dots & -u_{n-1} & +v_{n-1} & +p_{n-1}u_{n-1} \end{vmatrix}$$

Teniendo en cuenta las ecuaciones (??) y (21) el potencial en la superficie se escribe:

$$V(r,0) = \frac{\rho_1 I}{2\pi} \int_0^\infty \left[1 + 2\Theta_1(\lambda)\right] J_0(\lambda r) d\lambda. \tag{27}$$

La función $\Theta_1(\lambda)$ se suele denominar "núcleo integral de Stefanescu" y depende implícitamente de las resistividades, ρ_i , y de las profundidades de los contactos entre capas, h_i . Si se introduce la función $K_1(\lambda)$:

$$K_1(\lambda) = 1 + 2\Theta_1(\lambda),\tag{28}$$

La expresión (27) se transforma en:

$$V = \frac{\rho_1 I}{2\pi} \int_0^\infty K_1(\lambda) J_0(\lambda r). \tag{29}$$

 $K(\lambda)$ recibe el nombre de **núcleo integral de Slichter** (Slichter,1933). A partir de la ecuaciones (28), (25) la función $K(\lambda)$ puede ser calculada como:

$$K_1(\lambda) = \frac{(D+2N)}{D} \tag{30}$$

Las ecuaciones (26) y (30) combinadas con las relaciones (23), definen la relación entre el núcleo de Slichter y los parámetros de subsuelo $(h_i, \rho_i)_{i=1,\dots,n}$. Pekeris (1940) determina la misma relación a través de unas **relaciones de recurrencia** que llevan su nombre.

2.5.2. Las relaciones de Pekeris

Si se suma $e^{-\lambda h_i}$ a ambos miembros en la condición (17), y se dividen por la expresión (18), se llega a:

$$\rho_i \frac{(1 + \Theta_i(\lambda)) + X_i(\lambda)e^{2\lambda h_i}}{(1 + \Theta_i(\lambda)) - X_i(\lambda)e^{2\lambda h_i}} = \rho_{i+1} \frac{(1 + \Theta_{i+1}(\lambda)) + X_{i+1}(\lambda)e^{2\lambda h_i}}{(1 + \Theta_{i+1}(\lambda)) - X_{i+1}(\lambda)e^{2\lambda h_i}}$$
(31)

Se define ahora K_i como

$$K_i(\lambda) = \frac{(1 + \Theta_i(\lambda)) + X_i(\lambda)e^{2\lambda h_{i-1}}}{(1 + \Theta_i(\lambda)) - X_i(\lambda)e^{2\lambda h_{i-1}}}.$$
(32)

En la primera capa, $(h_{i-1} = 0 \text{ y } \Theta_i(\lambda) = X_i(\lambda)), K_i \text{ toma la forma}$

$$K_1(\lambda) = 1 + 2\Theta_1(\lambda),\tag{33}$$

y recupera, por tanto, el núcleo de Schlichter como caso particular. Manipulando algebraicamente los dos miembros de (31) con ayuda de (32) y definiendo:

$$t_i = h_i - h_{i-1},$$

como la potencia de la capa i-ésima y

$$p_i = \frac{\rho_i}{\rho_{i+1}},\tag{34}$$

como el contraste de resistividades entre capas sucesivas, se llega a (Koefoed, 1970):

$$K_{i+1}(\lambda) = p_i \frac{K_i - \tanh(\lambda t_i)}{1 - K_i \tanh(\lambda t_i)}.$$
(35)

o, despejando K_i ,

$$K_i(\lambda) = \frac{\left[K_{i+1} + p_i \tanh(\lambda t_i)\right]}{\left[p_i + K_{i+1} \tanh(\lambda t_i)\right]}.$$
(36)

Teniendo en cuenta la ecuación (32), y que $X_n(\lambda) = 0$ se obtiene que:

$$K_n\left(\lambda\right) = 1. \tag{37}$$

Comenzando con $K_n(\lambda)$ y supuestos conocidos los parámetros del terreno la aplicación de la relación (35), permite obtener $K_1(\lambda)$, relacionado con el potencial de superficie por la ecuación (29). La aplicación de la relación de recurrencia (35) "hacia arriba" puede deternerse en cualquier valor de subídice i. K_i representa, entonces, elnúcleo integral a emplear si las medidas de campo se llevasen a cabo, virtualmente, sobre la capa de índice i. La aplicación repetida de la ecuación (35) equivale a añadir una nueva capa por encima de la secuencia de capas ya tratada más el desplazamiento de la configuración de electrodos al techo de la nueva capa añadida.

Como la relación de Pekeris es reversible, la expresión (36) describe la supresión virtual de la capa actual y el desplazamiento de los electrododos al techo de la capa subyacente (proceso llamado "reducción a un plano frontera inferior").

Las relaciones de Pekeris pueden ser escritas en términos de una función denominada Transformada de resistividad (Koefoed, 1970), denotada por $T_i(\lambda)$ y definida como:

$$T_i(\lambda) = \rho_i K_i(\lambda). \tag{38}$$

Convirtiéndose las relaciones en:

$$T_i(\lambda) = \frac{\left[T_{i+1} + \rho_i \tanh(\lambda t_i)\right]}{\left[1 + \frac{T_{i+1} \tanh(\lambda t_i)}{\rho_i}\right]},$$
(39)

у

$$T_{i+1}(\lambda) = \frac{\left[T_i - \rho_i \tanh(\lambda t_i)\right]}{\left[1 - \frac{T_i \tanh(\lambda t_i)}{\rho_i}\right]}.$$
(40)

 $T_i(\lambda)$ tiene la misma dimensión de una resistividad y es una función de los parámetros de la capa y de λ , la cual tiene la dimensión recíproca de longitud $[L^{-1}]$. De las ecuaciones anteriores se deduce el comportamiento asintótico de dichas funciones:

- Cuando (1/λ) tiende a infinito, la tangente hiperbólica presente en la ecuaciones (39) y (40) tiende a cero, y por tanto T_i → T_{i+1}. Esto es cierto en todos lo pasos de la relación de recurrencia, de modo que todos los T_i (incluido T₁), se aproximan asintóticamente a T_n, la transformada de resistividad del sustrato. Y ésta, de acuerdo con la ecuaciones (37) y (38), vale ρ_n..
- Cuando $(1/\lambda)$ tiende a cero, la tangente hiperbólica tiende a uno y de la ecuación (39) se deduce que $T_i \to \rho_i$. Consecuentemente si $(1/\lambda)$ es pequeño, $T_1 \to \rho_1$.

2.5.3. Cálculo de la resistividad aparente para dispositivos simétricos

La relación (29) proporciona el potencial en un punto de la superficie de un suelo estratificado con inyección puntual de corriente. En el caso relativo a los S.E.V. hay que tener en cuenta la presencia de dos fuentes de corriente (inyección-extracción). La linealidad de la ecuación de Laplace permite calcular la diferencia de potencial medida entre electrodos situados en la superficie, δV , utilizando el principio de superposición:

$$\delta V(s) = 2\left[V(s-b) - V(s+b)\right] \tag{41}$$

donde V viene dado por la expresión (29). Operando sobre esta expresión se tiene

$$\rho_{ap}(s) = 2\rho_1 s \frac{s^2 - b^2}{4bs} \int_0^\infty K_1(\lambda) \left[J_0(\lambda s - \lambda b) - J_0(\lambda s + \lambda b) \right] d\lambda,$$

$$\rho_{ap}(s) = 2s \frac{s^2 - b^2}{4bs} \int_0^\infty T_1(\lambda) \left[J_0(\lambda s - \lambda b) - J_0(\lambda s + \lambda b) \right] d\lambda,$$
(42)

donde

$$T_1(\lambda) = \rho_1 K_1(\lambda),$$

es la transformada de resistividad en la primera capa. Introduciendo el factor de excentricidad, $c = \frac{b}{s}$, esta ecuación puede escribirse como

$$\rho_{ap}(s) = 2s \frac{1 - c^2}{4c} \int_0^\infty T_1(\lambda) \left[J_0(\lambda s (1 - c)) - J_0(\lambda s (1 + c)) \right] d\lambda. \tag{43}$$

Dicha expresión es válida para los dos principales tipos de dispositivos utilizados en la práctica que son el Wenner y el Schlumberger. En el caso del dispositivo Wenner, la resistividad aparente viene dada por la fórmula (43), con el valor concreto de la excentricidad ($c = \frac{1}{3}$). En el caso del dispositivo Schlumberger es posible simplificar dicha expresión, teniendo en cuenta particularidades de dicho dispositivo, como se ve a continuación.

Aproximaciones en el caso del dispositivo Schlumberger

En el caso del dispostivo Schlumberger se considera la distancia entre electrodos de medida, 2b, despreciable frente a la que existe entre electrodos de corriente. Se tiene, por tanto:

$$\frac{\left(s^2 - b^2\right)}{4bs} \underset{b \to 0}{\to} \frac{s}{4b}.\tag{44}$$

$$\left(\frac{\partial V}{\partial r}\right)_{r=s} \simeq \frac{V\left(s+b\right) - V\left(s-b\right)}{2b} = -\frac{\delta V}{4b},\tag{45}$$

con lo cual

$$\rho_{ap_Schl}(s) = k \frac{\delta V}{I} = -\frac{2\pi s^2}{I} \left(\frac{\partial V}{\partial r}\right)_{r=s},\tag{46}$$

donde V es el potencial deducido en el caso de un único punto de inyección (29). Utilizando en las ecuaciones (46) y (29) la igualdad

$$\frac{dJ_0(x)}{dx} = -J_1(x), \qquad (47)$$

se obtiene

$$\rho_{ap_Schl}(s) = \rho_1 s^2 \int_0^\infty \lambda K(\lambda) J_1(\lambda s) d\lambda. \tag{48}$$

Problemas de convergencia de dicha integral impropia aconsejan expresarla en función del núcleo de Stefanescu, haciendo uso de la integral de Lipschitz (11), de modo que,

$$V(r) = \frac{\rho_1 I}{2\pi r} + \frac{\rho_1 I}{2\pi} \int_0^\infty \Theta_1(\lambda) J_0(\lambda r) d\lambda.$$
 (49)

Derivando la relación anterior y teniendo en cuenta (46), se obtiene:

$$\rho_{ap_Schl}(s) = \rho_1 + 2\rho_1 s^2 \int_0^\infty \Theta_1(\lambda) J_1(\lambda s) d\lambda.$$
 (50)

Finalmente la resistividad aparente Schlumberger puede expresarse en términos de la transformada de restistividad utilizando las relaciones (28) y (38):

$$\rho_{ap_Schl}(s) = \rho_1 + s^2 \int_0^\infty (T(\lambda) - \rho_1) J_1(\lambda s) \lambda d\lambda.$$
 (51)

2.5.4. Cálculo de la resistividad aparente para dispositivos simétricos utilizando la teoría de filtros

Las relaciones entre las funciones de resitividad aparente y la función transformada de resistividad para configuraciones de electrodos lineales y simétricas vienen dadas por la fórmulas (43) y (51), que es una simplificación de la anterior. El uso de la escala logarítmica da a las curvas una apariencia más regular, disminuyendo su contenido en frecuencias en todo el rango de trabajo. Esta transformación es además importante para, posteriormente, utilizar las técnicas de tratamiento digital de la señal.

Como λ tiene dimensión física de longitud, se definen las variables x e y como sigue:

$$x = \ln s,$$

$$y = \ln \frac{1}{\lambda} = -\ln \lambda,$$
(52)

donde s representa el semiespaciado, con lo cual

$$\lambda s = e^{x-y}. (53)$$

Con estos cambios la resistividad aparente (43) se expresa:

$$\rho_{ap}(x) = \frac{1 - c^2}{2c} \int_{-\infty}^{+\infty} T(y) \left(J_0 \left((1 - c) e^{x - y} \right) - J_0 \left((1 + c) e^{x - y} \right) \right) e^{x - y} dy.$$
 (54)

En el caso del dispositivo Schlumberger la expresión (51) se transforma en

$$\rho_{ap_Schl}(x) = \rho_1 + \int_{-\infty}^{+\infty} [T(y) - \rho_1] J_1(e^{x-y}) e^{2(x-y)} dy$$
 (55)

Si, en la expresión (54) se introduce la variable $\eta = x - y$ se tiene:

$$\rho_{ap}(x) = \frac{1 - c^2}{2c} \int_{-\infty}^{+\infty} T(x - \eta) \left(J_0 \left[(1 - c) e^{\eta} \right] - J_0 \left[(1 + c) e^{\eta} \right] \right) e^{\eta} dy = \frac{1 - c^2}{2c} T * \Psi(x), \quad (56)$$

donde la función Ψ se expresa en términos de la función de Bessel de orden cero, J_0 . Esta relación indica que la resistividad aparente en un punto de medida x, puede ser considerada como la salida de un filtro cuya entrada es la transformada de resistividad T y cuya respuesta impulsional viene dada por la función Ψ . La forma de esta función depende del problema en cuestión. La relación (56) será utilizada para resolver el **Problema Directo de los S.E.V:** "conocido un suelo, dado por sus resistividades y espesores, determinar la resistividad aparente en un punto de la superficie.

Para resolver el problema directo es necesario:

- calcular la respuesta impusional del sistema L.I.T, $\Psi(x)$, es decir, los coeficientes del filtro. Para ello como veremos es necesario utilizar funciones de prueba (pares I/O) para entrenar el filtro.
- Determinar (por ejemplo mediante las relaciones de Pekeris) la entrada del sistema T, o transformada de resistividad, la cual depende del tipo de suelo bajo consideración. En la práctica la transformada de resistividad se calcula para una muestra de abscisas λ , que permita calcular la resistividad aparente con el grado de precisión deseado.

Cabe reseñar que el mismo modo de proceder puede ser aplicado a la expresión (55) que modela el problema directo en el caso de un dispositivo Schlumberger.

En este trabajo se han reproducido los resultados hasta aquí reseñados (Gosh, 1971a, b y Koefoed, 1979) en lo referente al diseño y cálculo de filtros a partir de funciones test. Para ello se han programado las rutinas necesarias, que hacen uso de técnicas del tratamiento de la señal (transformadas de Fourier, transformadas Z, etc...). Siendo éste un tema interesante como objeto de investigación, se ha observado que la longitud del filtro no afecta sensiblemente a la solución del problema directo.

Capítulo 3: Problemas inversos. Consideraciones generales.

3. CONSIDERACIONES GENERALES SOBRE PROBLEMAS INVERSOS

3.1. Concepto de problema inverso

Bajo la denominación de problemas inversos se ubica "la teoría matemática que describe **cómo** puede derivarse información sobre un sistema físico parametrizado a partir de datos observables, relaciones teóricas entre los parámetros del modelo y los datos e información a priori" (Moseegard y Tarantola, 1995)⁷

Esta definición no es, naturalmente, la única que aparece en los artículos referentes a este tema, pero ha sido elegida frente a otras porque se caracteriza por una gran flexibilidad en el sentido de que no sólo no restringe el procedimiento empleado para extraer información sobre el sistema a partir de los datos medidos, sino que tampoco fija las características que debe tener la solución de un problema inverso⁸.

Es, precisamente, el objetivo de este capítulo decidir cómo debiera formularse un problema inverso (en particular el atinente a los S.E.V) y qué propiedades convendría que poseyese su solución. Por supuesto, para fijar líneas normativas respecto a esta cuestión hace falta asumir, a priori, un concepto rector o criterio de valoración. En el caso de este trabajo, dicho criterio es la utilidad práctica⁹. Por tanto, lo que en los párrafos siguientes se sugiere es un modo de concretar la definición citada al principio del capítulo, de forma que los resultados puedan ser empleados para tomar decisiones de relevancia práctica.

Cualquier problema inverso contiene los siguientes "elementos constituyentes" o básicos:

1. Un conjunto de modelos

Es decir, una clase M de elementos matemáticos -conjunto de funciones, de vectores, etc...-que, se asume, describen el sistema desconocido o, al menos, los aspectos del mismo que se busca conocer. En el caso de los S.E.V se utilizan vectores $\mathbf{m} \in \mathbb{R}^{2n-1}$, siendo n el número de capas del $modelo\ conceptual$.

Es importante comprender que una vez escogido M queda excluida la posibilidad de obtener cualquier solución que no esté en M. Por tanto dicha elección condiciona de modo claro la solución. M depende del modelo conceptual físico adoptado, con sus correspondientes hipótesis simplificadoras, así como de otras restricciones adicionales que pueden corresponder a experiencia práctica relativa al tema.

Los elementos de M están caracterizados por ciertos parámetros. Por ejemplo, en el caso de S.E.V aquí tratado son los valores de la resistividad y de espesor de las capas geoeléctricas. Pero, como se verá, también está justificado -e incluso es más ventajoso- el uso de los logaritmos de estas cantidades. Igualmente podría utilizarse la conductividad (en lugar de su inverso, la resistividad).

⁷El texto literal es: "Inverse problem theory is the mathematical theory describing how information about a parameterized physical system can be derived from observational data, theoretical relationships between model parameters and data, and prior information".

⁸En este artículo se aborda la resolución de los problemas inversos en un marco bayesiano.

⁹Otros criterios, más matemáticos, se refieren a la unicidad de la solución y a la robustez de la misma (estabilidad).

En S.E.V se utilizan parámetros cuyos correlatos físicos son el dónde (espesores) y el cuánto (resistividades y/o conductividades). No obstante, podría optarse por parametrizar otras propiedades interesantes del terreno como, por ejemplo, las distintas escalas de heterogeneidad¹⁰.

2. La física o ley de comportamiento del problema

Ésta se plasma en una expresión teórica que permita calcular las predicciones de observables asociadas a cada uno de los modelos presentes en la clase M elegida. El proceso de cálculo de dichas predicciones, para cada modelo, utilizando esta expresión teórica se denomina resolver el problema directo.

En el caso de S.E.V existe un funcional

$$\mathbf{F}: \mathbb{M} \subset \mathbb{R}^{2n-1} \to \mathbb{R}^s$$
,

 ${f F}$ depende del modelo de forma no lineal. Esta circunstancia da origen a bastantes complicaciones. Sin embargo, conocido un modelo ${f m}$, el cálculo de ${f F}$ (${f m}$) se efectúa mediante la aplicación explícita de las reglas recurrentes de Pekeris. O sea, el problema directo posee solución analítica. En otros casos éste es mucho más complejo ya que puede involucrar la solución de sistemas de ecuaciones diferenciales no lineales.

Implícito en la expresión de **F** (o de cualquier otra formulación matemática del problema directo) se halla el modelo conceptual escogido con todas las hipótesis simplificadoras. La aproximación con que las mismas se den en cada caso práctico real es el factor que regula la desviación (espuria y persistente) de las predicciones respecto de las mediciones en campo.

3. Un conjunto de datos \mathbb{D}

Al que pertenecen las medidas realizadas, así como las predicciones obtenidas al resolver el problema directo.

4. Estructura de los conjuntos de modelos \mathbb{M} y de datos \mathbb{D}

No es suficiente considerar la clase de los modelos como un "todo uno". Si, supóngase, el modelo "real" es $\mathbf{m}_v \in \mathbb{M}$ (en el mejor de los casos) y $\mathbf{m}_{est} \in \mathbb{M}$ es otro modelo postulado como posible solución, es natural preguntarse si hay mucha diferencia entre el estimado y el real o, más aún, si están "lejos" o "cerca". Otro tipo de averiguaciones podrían querer llevarse a cabo respecto a ciertas características de la solución obtenida: por ejemplo, caracterizar el grado de "rugosidad" (heterogeneidad) que tenga.

Idénticas preguntas podrían hacerse respecto a los datos medidos \mathbf{d}_{obs} y a los calculados $\mathbf{d}_{calc} = \mathbf{F}(\mathbf{m})$.

Es preciso dotar a los conjuntos \mathbb{M} y \mathbb{D} de relaciones y leyes de composición interna y externa con ciertas propiedades. Un ejemplo ilustrativo es definir la diferencia entre dos modelos mediante el operador "sustracción". La construcción de normas y distancias, de productos escalares, etc..., convierte a \mathbb{M} y \mathbb{D} en espacios métricos. Si a las leyes definidas en \mathbb{M} y \mathbb{D} se les atribuye significado físico, es posible realizar juicios aplicados a la práctica sobre las distintas partes de \mathbb{M} y \mathbb{D} .

¹⁰Esto es lo que se consigue cuando se trabaja en el dominio de la frecuencia.

5. Un planteamiento del problema inverso

El planteamiento básico del problema inverso se resume en la siguiente ecuación:

$$\mathbf{F}(\mathbf{m}) = \mathbf{d},\tag{57}$$

Es decir, obtenido $\mathbf{d} \in D$ mediante medidas de campo y conocido \mathbf{F} , determinar las $\mathbf{m} \in M$ que satisfacen la ecuación (57). La formulación del problema inverso (57), presenta varias dificultades, que serán ilustradas en capítulos posteriores. Aquí sólo se comentan brevemente.

En primer lugar, como **F** representa procesos físicos *simplificados* (y, por tanto, incorrectos en cierta medida), **m** se busca entre cierto conjunto *restringido*, y los datos observados están afectados por errores de medida, se puede concluir que, conceptualmente al menos, carece de sentido.plantear (57) como una igualdad. Por otro lado, parece que aparte de la elección de M -que antecede al resto de operaciones- dicha ecuación no permite introducir otras informaciones disponibles acerca del modelo solución.

Una versión más adecuada sería,

$$\mathbf{F}(\mathbf{m}) \simeq \mathbf{d},\tag{58}$$

o sea, calcular los $\mathbf{m} \in \mathbb{M}$ tal que las predicciones $\mathbf{F}(\mathbf{m})$ estén "cerca" de \mathbf{d} . Más que un problema único, esta formulación representa una familia de posibles problemas. Cada uno de ellos atribuye un significado cuantitativo diferente al signo \simeq . Por ejemplo, un planteamiento adscrito a la tipología indicada en (58) es el que identifica la solución como aquella que minimiza la norma (de cierto tipo) de la diferencia entre $\mathbf{F}(\mathbf{m})$ y \mathbf{d} , es decir, puede existir diferencia entre $\mathbf{F}(\mathbf{m}_{sol})$ y \mathbf{d} , pero ésta debe ser mínima.

Como se observa, se sigue pretendiendo estimar un modelo solución -aunque ahora haya algo más de flexibilidad en la definición del planteamiento matemático del problema-. Varias formulaciones posibles se analizan en capítulos posteriores. Este es el enfoque tradicional o clásico, si se quiere, del problema inverso.

Muchos problemas inversos en geofísica y, en particular los S.E.V, gozan ¹¹ de la propiedad de que existen modelos **m** muy diferentes que satisfacen la ecuación (58) de manera muy razonable. Dichos modelos se denominan equivalentes dado que ajustan los datos medidos con el mismo grado o nivel de tolerancia (calidad). Como los datos observados están afectados de error, el modelo finalmente estimado con uno u otro conjunto de medidas varía. El inconveniente serio de este tipo de problemas es que la propagación de los errores desde las observaciones de partida hasta el modelo calculado genera una gran inestabilidad en la solución estimada. La naturaleza del planteamiento del problema inverso en S.E.V es tal que la incertidumbre (en general razonablemente pequeña en una campaña cuidada de adquisición de datos) se amplifica notablemente y produce gran dispersión de las soluciones obtenidas. En la práctica es bastante común observar que cuando se obtiene un modelo utilizando un grupo de medidas y se considera otro grupo de medidas diferente (de la misma calidad pero no exactamente iguales) se puede obtener una solución radicalmente distinta. Este tipo de situaciones son percibidas con un grado alto de desasosiego por parte de la comunidad ingenieril pues, aparentemente al menos, permite un grado exagerado de

¹¹o sufren

subjetividad en la elección de una solución que, de otra parte, aparece como el resultado impredecible de una "caja negra".

Adquiere, por tanto, relevancia práctica la cuestión que versa sobre los criterios utilizados para escoger entre las diferentes soluciones. Asimismo, una vez elegida una, debería estar acompañada por una medida de la incertidumbre de dicha estimación, algún tipo de valoración de su distancia al modelo real (desconocido). La teoría de la propagación de errores, el análisis de intervalos de confianza o los núcleos de resolución son técnicas a las cuales se recurre si el problema es lineal. Sin embargo, cuando el problema es no lineal, las respuestas a este tipo de cuestiones sólo pueden ser aproximadas y, cuando la no linealidad es alta, éstas pueden dar lugar a resultados erróneos.

Parte de las erráticas fluctuaciones de la solución pueden eliminarse incorporando información "a priori", esto es, complementaria a la contenida en los datos medidos. El proceso de estabilización de las soluciones mediante el uso de información a priori se denomina **regularización**. Sin embargo, el grado de regularización sigue siendo una variable a decidir en cada problema.

En un marco bayesiano, cada modelo solución viene caracterizado por su probabilidad a posteriori. Es ésta una medida de su compatibilidad con los datos observados y con la información a priori. En este enfoque, el grado de regularización -que regula la intensidad y el modo en que la información "ajena" a los datos influye en la solución- queda especificada en el paso inicial que consiste en asignar probabilidades a los diferentes modelos que se consideran factibles. Esta asignación de probabilidades "a priori" es, a la vez, el talón de Aquiles del método y la causa de que su uso se haya extendido en geofísica. Efectivamente, la definición de "a prioris" adecuados no es un asunto simple y elecciones aparentemente inofensivas pueden producir efectos sorprendentes (Scales and Tenorio, 2001). Por otro lado, con este sistema, la información a priori fluye por el problema con la misma naturalidad con la que lo hacen los datos medidos facilitando su tratamiento.

A lo largo de este trabajo se analizan estos enfoques que conducen a la obtención de una solución estimada. Se realiza también algún esfuerzo por introducirse en la "caja negra" y determinar la importancia y significado físico de todos los factores que influyenen el proceso. Sin embargo, el decurso de la presentación lleva a un nuevo tipo de planteamiento que aparece como mucho más natural, poderoso y adaptado a las necesidades prácticas ingenieriles que el clásico recién mencionado.

El interés real de la interpretación geofísica es lograr traducir datos e información a priori en predicados sobre aquellas características de la solución sobre las que se pueda fundamentar una toma de decisiones. Para determinar lo que debe constar en el informe de resultados del intérprete geofísico hay que preguntarse por el tipo de frases que espera encontrar la persona que debe tomar decisiones sobre, por ejemplo, realizar o no un sondeo, o las actuaciones a emprender ante un problema medioambiental. ¿Le es útil a esta persona un informe en el que se le faciliten todos los parámetros -en este caso resistividades y espesores- del modelo estimado?. Si se le proporciona como resultado único, el decisor entenderá que éste es el único posible. Tras lo dicho sobre las equivalencias e inestabilidades, esto es erróneo. Por otra parte, este tipo de actitud presupone que lo que importa para la toma de decisiones es el valor de los parámetros. No obstante, esto

podría no ser así. La mayoría de las veces el interés real se centra en descubrir la presencia de patrones interpretables en la estructura del subsuelo. Por ejemplo, la constatación de grandes contrastes localizados en varios sondeos indicativa -posiblemente- de la presencia de un substrato rocoso, o de una intrusión salina. O bien, la observación de que las resistividades del suelo aumentan de cierta forma con la profundidad (más que el propio valor de la resistividad). Es decir, los patrones y tendencias adquieren bastante importancia. En la medida en que un mismo patrón -por ejemplo, altas resistividades a partir de cierta profundidad- puede ser compartido por muchos modelos que, por lo demás, pueden ser muy distintos, recorta el impacto de las mencionadas inestabilidades en la determinación de los parámetros.

Ahora bien, si la toma de decisiones no se fundamenta estrictamente hablando en los valores de los parámetros sino de ciertas características que dependen de ellos, debe ponerse en cuestión la utilidad misma de estimar un modelo. Podría no ser necesario cumplir con el ritual de presentar un modelo estimado en el informe final (aunque ello pudiese parecer una dejación de funciones) porque la característica relevante fuese compartida por otros modelos que no estarían presentes en el informe y que ajustan los datos con igual calidad.

La toma de decisiones en condiciones de riesgo se fundamenta en la teoría de probabilidades. Lo interesante sería disponer de una metodología en la cual, los modelos que mejor ajustasen los datos y menos se desviasen de lo indicado por experiencias de otro tipo apareciesen como los más probables. En este sentido, el marco bayesiano es una herramienta de valor incalculable. La distribución de probabilidades a posteriori hace precisamente eso: atribuir a cada modelo una probabilidad en función de dichos criterios.

Si se conoce -o se pudiera estimar- la distribución de probabilidades "a posteriori", se podrían también identificar los conjuntos de modelos que con mayor probabilidad se presentarían. Incluso, se podrían valorar las probabilidades relativas de que las soluciones posibles tengan tal o cual característica estructural (lo cual es mucho más interesante). Por ejemplo, si se está interesado en cierta seminorma de la solución, cabría obtener las probabilidades con las que serían observados distintos valores de dicha característica. Como este valor refleja una propiedad de interés, a saber: la heterogeneidad del suelo, sería posible tomar decisiones que dependiesen del valor de dicha heterogeneidad -independientemente de cuáles fuesen los valores particulares de los parámetros de los modelos incluídos.

En este sentido, el conjunto de soluciones equivalentes deja de ser un problema que produzca desorientación, perplejidad y conclusiones equivocadas y pasa a ser la población sobre la que se hace la estadística necesaria para tomar decisiones y asumir riesgos. Si se asume este concepto de solución de un problema inverso, en las páginas de los informes de resultados debieran aparecer histogramas de valores representativos de propiedades interesantes del suelo, de distribuciones marginales de algunos parámetros de ciertas capas, matrices de correlación, etc... Este es el enfoque que se sigue en la caracterización del problema práctico presentado en este trabajo.

La obtención de una expresión explícita de la distribución a posteriori es inviable en problemas no lineales y de gran número de parámetros. Además, la no linealidad suele dotar a esta distribución de propiedades nada habituales en los análisis teóricos tradicionales, es decir, multimodalidad, valores infinitos o excepcionalmente altos de algunos momentos, etc (Moseegard y Tarantola, 1995). Por ello es importante contar con un método que muestree la región "interesante" del espacio de modelos con una tasa proporcional al valor de la probabilidad a posteriori en cada punto. De esta forma, un simple contaje de los modelos muestreados permitirá identificar las frecuencias relativas de aparición con las probabilidades a posteriori necesarias para tomar decisiones. Se ha demostrado que el método de muestreo que posee las propiedades correctas para hacer esto es el método de "Metropolis-Gibbs" (Moseegard y Tarantola, 1995). Este es un método que ha sido implementado en el programa para ejecución de S.A. En esta tesis también se valora la utilidad de los algoritmos genéticos para servir de método de muestreo. Con estas dos técnicas de optimización global, se pueden obtener las muestras que permitan inferir los valores de probabilidad requeridos.

Capítulo 4: El problema inverso lineal

4. EL PROBLEMA INVERSO LINEAL

4.1. Motivación para el estudio de los problemas lineales

Conviene precisar qué se entiende por problema inverso lineal. Se denominará problema inverso lineal aquel que está asociado a un problema directo lineal (Parker, 1994). Esta definición incluye el caso en que el problema directo lineal se obtiene por linealización de un problema directo original no lineal.

El objetivo de este trabajo es poder **caracterizar** modelos de terreno -que satisfagan las hipótesis asumidas en la técnica de S.E.V- y tal que la operación sobre ellos del funcional **no lineal** del problema directo

$$\mathbf{F}: \mathbb{M} \subset \mathbb{R}^{2k-1} \to \mathbb{R}^s$$
,

predigan unos datos que aproximen "lo suficiente" las observaciones medidas en el terreno.

Como se verá, la condición de cercanía entre las predicciones $\mathbf{F}(\mathbf{m})$ del funcional no lineal y los datos medidos \mathbf{d} que, de manera intuitiva y un tanto vaga, podría expresarse como

$$\mathbf{F}(\mathbf{m}) \simeq \mathbf{d}$$

tiene otras formulaciones precisas y operativas que emplean el concepto de **norma** para medir distancias entre $\mathbf{F}(\mathbf{m})$ y \mathbf{d} en \mathbb{R}^s . En este sentido, la función

$$E\left(\mathbf{m}\right) = \left\|\mathbf{d} - \mathbf{F}\left(\mathbf{m}\right)\right\|_{N_{\mathbb{D}}} : \mathbb{R}^{2k-1} \to \mathbb{R}^{+} \cup \left\{0\right\}, \tag{59}$$

donde $N_{\mathbb{D}}$ es cualquier tipo de norma definida sobre el conjunto de datos, tiene una importancia central para el problema inverso porque se toma como una expresión del desajuste entre las predicciones del modelo estimado y los datos medidos. En este sentido, tiene cierta lógica definir la solución del problema inverso como el modelo \mathbf{m} que minimiza $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{N_{\mathbb{D}}}$, es decir, que "mejor" predice los datos, o también, elegir los modelos \mathbf{m} para los que $E(\mathbf{m})$ sea menor que cierta tolerancia.

Se verá que, a veces, en torno a un problema existe conocimiento que, estrictamente hablando, es "externo" al proceso de toma de datos (geofísicos) pero que, no obstante, informa sobre el tipo de terreno -piénsese en datos de sondeos, información geológica, experiencias previas en terrenos similares, etc. No solo sería un derroche no utilizar esta información, sino que, como se justificará a lo largo de este capítulo es esencial emplearla. En este caso, además de la norma del desajuste convendrá que aparezcan otros términos adicionales -que también se expresan mediante normasque reflejen el grado de cumplimiento de dicha información denominada "a priori". No obstante, de cara a justificar el estudio en profundidad de problemas lineales en este trabajo, se reseña que, la norma del desajuste, indicada mediante la expresión (59), aparece siempre, sóla o en compañía de otros términos portadores de información adicional, en cualquier formulación de un problema inverso, ya que es la única vía de entrada de la información contenida en las medidas d tomadas en el campo.

Se enumeran aquí tres motivaciones para el estudio del problema lineal:

1. Se puede construir la solución de un problema inverso no lineal resolviendo una secuencia de *problemas lineales*.

¹²En rigor, aunque el problema directo sea lineal, el problema inverso asociado puede no serlo, dependiendo de la formulación escogida para éste último.

2. Se pueden obtener soluciones equivalentes para el problema no lineal a través del estudio de las equivalencias del problema linealizado.

3. La teoría lineal puede utilizarse para realizar un análisis aproximado de la estructura y comportamiento de la solución (análisis de sensibilidad -model appraisal-, Scales and Snieder, 2000).

A continuación se explicitan más detalladamente estas razones.

1. Contrucción de la solución

Una de las estrategias para resolver el problema no lineal,

$$\mathbf{m}_{sol} = \min_{\mathbf{m} \in \mathbb{M}} \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{N_{\mathbb{D}}}.$$
 (60)

consiste en sustituir este problema por una secuencia iterativa de problemas lineales cuyas soluciones, se espera, converjan a la solución del problema original (177). Supóngase, entonces, que en el paso k-ésimo de dicha secuencia se ha llegado a un modelo \mathbf{m}_k . A partir del mismo, se tratará de determinar cuál es el modelo siguiente \mathbf{m}_{k+1} el cual, en teoría, se considera más "cercano" a la solución \mathbf{m}_{sol} .

El concepto de linealización del funcional $\mathbf{F}(\mathbf{m})$ es la idea clave en este momento. Como se explicará en más detalle en el capítulo siguiente, dedicado al problema no lineal, si en el punto \mathbf{m}_k el funcional no lineal $\mathbf{F}(\cdot)$ es diferenciable con continuidad, existe un entorno $\delta(\mathbf{m}_k)$ de \mathbf{m}_k tal que, para todo modelo \mathbf{m} que pertenezca a dicho entorno se puede escribir¹³:

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_k) + \mathbf{dF}_{\mathbf{m}_k}(\Delta \mathbf{m}) + \mathbf{o}\left(\|\Delta \mathbf{m}\|_{N_{\mathbb{M}}}\right). \tag{61}$$

Aquí, $\mathbf{dF}_{\mathbf{m}_k}(\Delta \mathbf{m})$, es la aplicación diferencial en \mathbf{m}_k , cuya expresión en las bases canónicas de \mathbb{M} y \mathbb{D} es:

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_k) + \begin{bmatrix} \mathbf{\nabla}^t F_1(\mathbf{m}_k) \\ \vdots \\ \mathbf{\nabla}^t F_s(\mathbf{m}_k) \end{bmatrix} [\mathbf{m} - \mathbf{m}_k] + \mathbf{o}\left(\|\Delta \mathbf{m}_k\|_{N_{\mathbb{M}}}\right), \tag{62}$$

donde
$$\mathbf{JF}_k \equiv \begin{bmatrix} \mathbf{\nabla}^t F_1(\mathbf{m}_k) \\ \vdots \\ \mathbf{\nabla}^t F_s(\mathbf{m}_k) \end{bmatrix}$$
 recibe el nombre de **matriz jacobiana** de **F** en \mathbf{m}_k y juega un papel fundamental tanto para hallar la solución del problema como para analizar sus

un papel fundamental tanto para hallar la solución del problema como para analizar sus características.

Según lo indicado por la expresión (187) si \mathbf{m} está tan "cerca" de \mathbf{m}_k como para que $\mathbf{o}\left(\|\Delta\mathbf{m}_k\|_{N_{\mathbb{M}}}\right)$ se pueda despreciar, entonces, sustituyendo en el problema (177) $\mathbf{F}\left(\mathbf{m}\right)$ por su aproximación lineal (187) se obtiene \mathbf{m}_{k+1} como:

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{N_{\mathbb{D}}} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \|\mathbf{F}(\mathbf{m}_k) - \mathbf{d} + \mathbf{J}\mathbf{F}_k (\mathbf{m} - \mathbf{m}_k)\|_{N_{\mathbb{D}}}.$$
 (63)

¹³ Al igual que se define una norma en el espacio de datos, N_D , debe hacerse también en el espacio de parámetros, N_M . Ambas, no tienen porqué coincidir.

Cabe observar que este modelo pertenece a un entorno de \mathbf{m}_k y minimiza una "buena" aproximación *local* de la norma del desajuste.

Si se denota por $\mathbf{d}_k = \mathbf{F}(\mathbf{m}_k) - \mathbf{d}$ el error de predicción no lineal correspondiente al modelo \mathbf{m}_k y por $\hat{\mathbf{d}}_k = \mathbf{F}(\mathbf{m}_k) - \mathbf{d} - \mathbf{J}\mathbf{F}_k\mathbf{m}_k$, el cálculo del modelo \mathbf{m}_{k+1} a partir del \mathbf{m}_k se puede formular de dos formas distintas:

• formulación en incrementos (creeping)

$$\Delta \mathbf{m}_{k} = \min_{\Delta \mathbf{m}_{k} \in \boldsymbol{\delta}(\mathbf{0})} \|\mathbf{d}_{k} + \mathbf{J} \mathbf{F}_{k} \Delta \mathbf{m}\|_{N_{\mathbb{D}}}, \mathbf{m}_{k+1} = \mathbf{m}_{k} + \Delta \mathbf{m}_{k},$$

$$(64)$$

donde se ha tomado como incógnita $\Delta \mathbf{m} = \mathbf{m} - \mathbf{m}_k$.

• formulación en el modelo (jumping)

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \left\| \hat{\mathbf{d}}_k + \mathbf{J} \mathbf{F}_k \mathbf{m} \right\|_{N_{\mathbb{D}}}, \tag{65}$$

en el que la incógnita es m.

Visto que se procede a construir la solución del problema no lineal (177) mediante la solución iterativa del problemas linealizados como (64) ó (189) es conveniente analizar qué información contienen las soluciones de éstos últimos, ver cómo se transmite ésta a través del proceso numérico, y cuáles son los inconvenientes que presenta cada uno de estas formulaciones.

2. Estudio de equivalencias

Supóngase que, de algún modo, se ha llegado a la solución del problema no lineal (177). En este tipo de problemas, cuando existe solución, la no unicidad suele ser una característica esencial. Dicho brevemente, suele haber otros modelos equivalentes al hallado en el sentido de ajustar los datos "igual de bien" (o sea, con una cierta tolerancia). Sería lógico pensar que, al menos algunas de estas soluciones equivalentes, estén en el entorno de la solución hallada. Nuevamente, un desarrollo lineal de $\mathbf{F}(\mathbf{m})$, esta vez en el entorno de la solución \mathbf{m}_{sol} , nos permite considerar que:

$$\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{N_{\mathbb{D}}} \simeq \|\mathbf{F}(\mathbf{m}_{sol}) - \mathbf{d} + \mathbf{J}\mathbf{F}_{sol}(\mathbf{m} - \mathbf{m}_{sol})\|_{N_{\mathbb{D}}}.$$
 (66)

Esta expresión brinda un modo mucho más simplificado y rápido, pero muy aproximado, de analizar los modelos equivalentes al obtenido.

3. Valoración de la solución estimada

Dado que el análisis de equivalencias puede abordarse, con buena aproximación, usando la simplificación linealizada de (66) en lugar de manejar el funcional no lineal $\mathbf{F}(\mathbf{m})$, y que mediante la solución reiterada de las ecuaciones linealizadas del tipo (64) ó (189) se puede llegar a construir una solución del problema no lineal, sería interesante disponer de una teoría consolidada -inexistente para el problema no lineal- que permita analizar estos nuevos problemas. Esta teoría existe y está bien desarrollada para los problemas inversos lineales, como se detallará en capítulos posteriores.

4.1.1. Un ejemplo de obtención de soluciones equivalentes

A continuación se aporta un ejemplo de linealización en el caso de un suelo con dos capas geoeléctricas

$$\mathbf{m}_{sol} = (\rho_1, t_1, \rho_2) = (10, 5, 1)$$
.

En este caso se han tomado datos en tres puntos de la superficie con un semiespaciado de 1,5 y 10 metros respectivamente. Se ha obtenido el vector correspondiente de medidas (en resistividades aparentes): $\mathbf{d} = (d_1, d_2, d_3) = (9.98, 8.75, 5.22)$. En este ejemplo se supone que las medidas están libres de error -por esta razón aparecen representadas en la figura 3 mediante estrellas rojas exactamente superpuestas a la curva de resistividades aparentes predicha por el citado modelo.

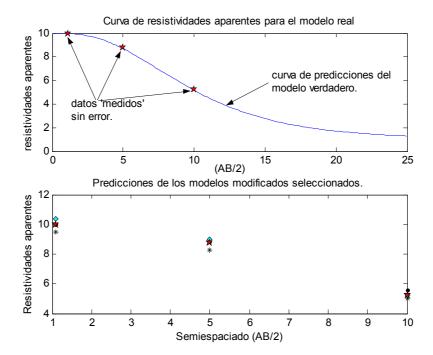


Figura 3: Ejemplo sintético: a) Tres puntos de medida (sin error) sobre curva de resistividades aparentes. b) Predicciones de tres modelos obtenidos por análisis lineal, junto a las medidas sin error.

Para medir distancias en \mathbb{D} , se ha empleado la norma euclídea (que es, con mucho, la más corriente). Como los datos están libres de error (para simplificar), el vector

$$\mathbf{d}_{sol} = \mathbf{F}\left(\mathbf{m}_{sol}\right) - \mathbf{d} = \mathbf{0}.$$

Por tanto la expresión (66) queda reducida a:

$$\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{2} \simeq \|\mathbf{J}\mathbf{F}_{sol}(\mathbf{m} - \mathbf{m}_{sol})\|_{2}.$$
 (67)

En la figura 4 se representan las curvas de nivel que delimitan los modelos para los que el $error\ lineal^{14}$ es del 5 %, 10 %, 25 % y 50 % respectivamente, cuando se permite que varíen las resistividades de la primera y de la segunda capa, pero se mantiene el espesor de la primera capa igual al de la solución.

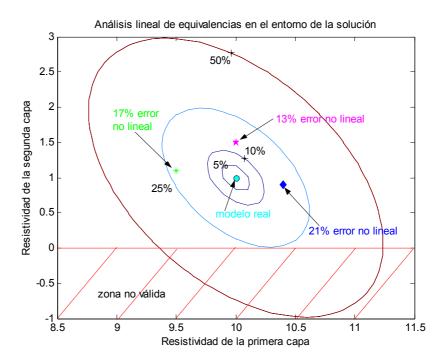


Figura 4: Distintas regiones de equivalencia lineal. Relación de la estimación del error lineal con el no lineal.

La forma y amplitud de dichas elipses está exclusivamente definida -tal como se vé en la expresión (67)- por la matriz jacobiana \mathbf{JF}_{sol} que, en este caso vale:

$$\mathbf{JF}_{sol} = \begin{bmatrix} 0.9977 & 0.0114 & 0.0036\\ 0.8454 & 0.6158 & 0.2387\\ 0.4330 & 1.2973 & 0.8270 \end{bmatrix}. \tag{68}$$

La estructura de la matriz jacobiana -también llamada matriz de sensibilidad- se discutirá a continuación. Los factores que influyen en la forma y tamaño que adquieren las zonas de equivalencia lineal -en este caso elipses por permitir sólo la variación de dos parámetros- se especificarán en detalle en el capítulo relativo al análisis de equivalencias.

En la zona de equivalencia lineal inferior al 25 %, se han tomado tres modelos que aparecen en la figura junto con sus errores *no lineales* respectivos. Las predicciones de dichos modelos en

$$\frac{\left\|\mathbf{J}\mathbf{F}_{sol}.\left(\mathbf{m}_{5\,\%}\mathbf{-m}_{sol}\right)\right\|_{2}}{\rho_{2}}=0.05.$$

¹⁴Medido como desviación cuadrática media relativa al valor del dato predicho más bajo. O sea, en la curva del 5% se cumple que:

los puntos de medida se representan en la parte inferior de la figura 3, donde se observa que casi se superponen a las predicciones del modelo verdadero (estrellas rojas). Se aprecia también la buena correspondencia entre el error lineal predicho para ellos y el correspondiente error no lineal. Éste puede ser, por tanto, un método sencillo de conseguir modelos equivalentes utilizando sólo información del problema linealizado. En la siguiente tabla se especifican las características de los modelos equivalentes:

modelos	1	2	3	m_{sol}
resistividades	(10, 1.5)	(9.5, 1.1)	(10.4, 0.9)	(10, 1)
potencias	5	5	5	5
error no lineal	21%	17%	13%	0

Características de los modelos mostrados en la figura 4.

4.2. Los espacios de datos y de modelos

4.2.1. Parametrizaciones

Se denomina parametrización, en el contexto de este trabajo, a la elección (adecuada) de las variables en las que se formula el problema.

Como se verá en el capítulo destinado al problema no lineal, el análisis de la existencia de solución es una tarea complicada. En la práctica se opta por un método constructivo de dicha solución (Parker, 1994), es decir, se emplean procedimientos numéricos que construyen posibles soluciones y se comprueba si satisfacen los requisitos adecuados. A veces, un cambio del espacio de modelos y/o de datos, puede permitir reformularlo como un problema lineal. Los problemas que permiten esta posibilidad podrían denominarse intrínsecamente lineales. Siempre es necesario prestar mucha atención a los cambios realizados en la parametrización debido a que:

- 1. El error de predicción o desajuste, considerado como variable aleatoria, se caracteriza por una distribución de probabilidades determinada. Al cambiar la parametrización se modifica el tipo de distribución de probabilidades y cambian por tanto las estimaciones obtenidas a partir de las mismas (Menke, 1989). Por ejemplo, la media del desajuste con una nueva parametrización podría no corresponderse con la media del desajuste en la parametrización original.
- 2. Cambia la interpretación física de los criterios (normas) estipulados para medir dicho error.
- 3. Se alteran todas las características del proceso numérico de solución que dependan de la elección del tipo de datos y de modelos. Por ejemplo, la matriz de sensibilidad y, por lo tanto, su número de condición, etc...

4.2.2. El espacio de los datos

Resistividades aparentes y transformadas de resistividad En un problema inverso, los datos son esenciales. En este caso, contienen la información sobre el modelo de capas del terreno recabada a través de un proceso de medida. Para trabajar con el problema inverso en S.E.V se puede optar, en principio, entre resistividades aparentes y

transformadas de resistividad. Asimismo, para cada tipología mencionada, pueden emplearse los propios datos o bien, sus logaritmos en la base adecuada.

La resistividad aparente -abreviadamente $\rho_{ap}(\mathbf{x})$ - se obtiene a partir de las diferencias de potencial medidas, de la intensidad de corriente introducida y de un factor geométrico. Puede considerarse, pues, como el **dato de campo**. Tal como se ha visto en el capítulo dedicado al tratamiento de la señal, los valores de las resistividades aparentes se pueden obtener por convolución de un filtro h característico del tipo de dispositivo (en este caso Schlumberger) con una función denominada transformada de resistividad, representada por $T(\lambda)$. En la práctica, la convolución se calcula empleando un conjunto discreto de valores del filtro y otro conjunto de valores procedente del muestreo de $T(\lambda)$. Los puntos de muestreo se eligen de modo que el error de discretización sea asumible. Por tanto:

$$\boldsymbol{\rho}_{ap}(x_i, \mathbf{m}) = (h * T)(x_i, \mathbf{m}) = \sum_{n=n1}^{n=n2} T(\lambda_n, \mathbf{m}) \cdot h(x_i - \lambda_n).$$
(69)

La figura 5 representa las medidas de resistividad aparente en tres puntos de la superficie y los valores de la transformada de resistividad asociados a cada una de ellas, por la relación expresada en (69).

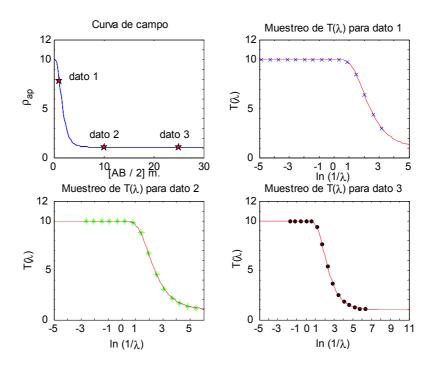


Figura 5: Medidas de resistividad aparente en tres puntos de la superficie y muestras correspondientes de la transformada de resistividad necesarias para la convolución

En realidad, los datos que tienen interés práctico son las resistividades aparentes, obtenidas a partir de las medidas de campo. Para poder operar con transformadas de resistividad, éstas debieran obtenerse por convolución del conjunto de medidas de resistividad aparente con un

filtro adecuado (inverso de h). Como se observa en la figura 5, cada uno de los puntos de medida de resistividades aparentes requiere, para ser calculado con cierta precisión por convolución con el filtro h, de un número elevado de valores muestreados sobre la función de transformadas de resistividad $T(\lambda)$. La situación sería la misma para el filtro inverso. Para calcular cada uno de los valores de transformada de resistividad, con alguna exactitud, debería utilizarse un amplio conjunto de medidas de resistividad aparente obtenidas en el campo¹⁵. En la práctica esta situación es -generalmente- inviable.

No obstante, las transformadas de resistividad juegan un papel relevante en el cálculo de las resistividades aparentes y en el análisis del problema y se han incluido en este capítulo como tipo de dato para que el tratamiento sea completo.

Volviendo a la expresión (69), de ella se deduce que el vector de resistividades aparentes puede expresarse en forma matricial como:

$$\rho_{ap} = H \cdot \mathbf{T},$$

donde H es la matriz del filtro y \mathbf{T} es el vector que contiene las muestras de $T(\lambda)$ para el conjunto de medidas. A continuación se detalla cada uno de estos términos para el ejemplo ilustrado en la figura 5. En él se han tomado tres medidas de resistividad aparente a 1.1, 10 y 25 metros respectivamente, obteniéndose el vector de medidas $\boldsymbol{\rho}_{ap} = (7.8, 1.04, 1.02)^{16}$. El terreno subyacente queda representado por el vector $\mathbf{m}_{sol} = (10, 5, 1)$. La matriz H tiene en este caso dimensiones (3, 45), ya que son tres los puntos de medida y 45 las abscisas de cálculo del filtro h.

El empleo de uno u otro tipo de dato $-\rho_{ap}$ ó T- cambia los valores de la matriz de sensibilidad \mathbf{JF}_{sol} mediante la relación (regla de la cadena):

$$\frac{\partial \rho_{ap}(x_i, \mathbf{m})}{\partial m_j} = \sum_{n=n_1}^{n=n_2} h(x_i - \lambda_n) \cdot \frac{\partial T(\lambda_n, \mathbf{m})}{\partial m_j}.$$
 (70)

Esta expresión se detallará en la siguiente sección al hablar de la estructura y propiedades de la matriz de sensibilidad. No obstante, obsérvese que la inmensa mayoría de los elementos del filtro poseen valores muy pequeños, hecho relevante para algunas propiedades de la matriz Jacobiana \mathbf{JF}_{sol} .

La influencia de la parametrización

En la literatura sobre problemas inversos suele suponerse que, el error en cada uno de los datos- entendido como la diferencia (en cada punto de medida) entre el dato medido y el dato predicho por el modelo verdadero- se distribuye según una ley *normal*.

Si los datos siguen una ley gaussiana debería asumirse que pueden tomar valores positivos y negativos. No obstante, tanto las resistividades aparentes medidas como las transformadas de resistividad son, siempre, cantidades positivas¹⁷.¿Cómo pueden entonces los datos, siendo

¹⁵ Además, sería preferible que las medidas tomadas estuvieran equiespaciadas en una escala logarítmica.

 $^{^{16}}$ Las casi igualdad de los valores de las dos últimas medidas se debe a que se ha alcanzado el comportamiento asintótico de la curva, reflejando la influencia dominante de la segunda (y última) capa ($\rho_2 = 1$).

¹⁷Si bien la lectura derivada de un circuito eléctrico puede subestimar la resistencia medida nunca proporcionará una lectura negativa. En el peor de los casos será nula.

positivos, seguir una distribución gaussiana?. Si las medidas siguiesen estrictamente una distribución normal, tras un número suficientemente grande de experimentos debería aparecer un cierto porcentaje de datos negativos. Como esto es imposible, no cabe, estrictamente hablando que ésta sea la distribución de los datos (ni, por tanto, la de los errores).

Sin embargo, para los errores relativos que se consideran asumibles en la práctica -del orden del 5%- (LaBrecque, 1993), la probabilidad de obtener datos negativos suponiendo que éstos siguen una distribución normal (ver apéndice-capítulo lineal) es 0.8%. Visto este pequeño valor, en la práctica tiene sentido, si el análisis estadístico de los datos no descarta esa hipótesis, suponer que han sido generados mediante un proceso gaussiano. Para errores operativos mayores, el porcentaje de datos negativos aumenta y, por tanto, seguir considerándolos gaussianos implicaría una sobreestimación del valor de su media y una subestimación del valor de su desviación típica. Esto es así porque el proceso gaussiano debe ser tal que justifique la ausencia, tras un número determinado de medidas, de valores negativos, lo que conduce a que la media adquiera un valor mayor del que realmente posee. Por la misma razón, la desviación típica se reduce, representando el hecho de que las medidas están más próximas unas a otras.

Para ejemplificar los efectos que tienen las reparametrizaciones, a continuación se expone el efecto producido por una transformación logarítmica de los datos -muy empleada en la práctica-sobre la estimación de la media poblacional de la distribución de datos.

Al ser los datos positivos está bien definida la transformación $\mathbf{d}_{\log} = \log(\mathbf{d})$. Se supone que los datos provienen de una distribución normal. Resolver el problema inverso mediante un ajuste por mínimos cuadrados sobre \mathbf{d} (datos sin transformar)

$$\mathbf{m}_{g} = \min_{\mathbf{m} \in \mathbb{M}} \left\| \mathbf{F} \left(\mathbf{m} \right) - \mathbf{d} \right\|_{2},$$

implica determinar un modelo, denotado por \mathbf{m}_g , que identifica $\mathbf{F}(\mathbf{m}_g)$ con la media muestral (aritmética) de los datos

$$F\left(\mathbf{m}_{g}, x_{i}\right) = \frac{1}{N\left(x_{i}\right)} \sum_{i=1}^{j=N} d_{j}\left(x_{i}\right), \tag{71}$$

ya que en caso de normalidad, la media muestral es el mejor estimador de la media poblacional.

Supóngase que tras la transformacion logarítmica, el método de mínimos cuadrados trata de encontrar un modelo, denotado por \mathbf{m}_{\log}

$$\mathbf{m}_{\log} = \min \|\log \mathbf{F}(\mathbf{m}) - \log \mathbf{d}\|_2$$
.

Esto supone identificar $\log \mathbf{F}(\mathbf{m}_{log},x_i)$ con la media aritmética de los logaritmos de los datos, es decir:

$$\log \mathbf{F}\left(\mathbf{m}_{\log}, x_i\right) = \frac{1}{N\left(x_i\right)} \sum_{j=1}^{j=N} \log d_j\left(x_i\right) = \log \sqrt[N]{d_1 d_2 \cdots d_N}. \tag{72}$$

Por lo tanto $\mathbf{F}(\mathbf{m}_{10g}, x_i) = \sqrt[N]{d_1 d_2 \cdots d_N}$ que es la media geométrica de los datos originales. Así pues \mathbf{m}_g y \mathbf{m}_{\log} son diferentes¹⁸.

Ambos modelos, \mathbf{m}_g y \mathbf{m}_{\log} , tienden a ser iguales si $N(x_i) \to \infty$ en cada punto de muestreo x_i .

La conveniencia de aplicar una transformación logarítmica a los datos

En la sección precedente se ha visto que el modelo estimado será diferente dependiendo de la parametrización adoptada. La valoración de la repercusión de la elección sigue dos líneas de razonamiento diferentes, mencionadas en los apartados 1 y 2 de la introducción a esta sección (véase 4.2.1).

1. El sentido físico de la definición del error

En relación a la interpretación "física", supóngase, para fijar ideas, que se trabaja con resistividades aparentes y sin transformación logarítmica. Se conocen los valores de las medidas, libres de error, $d_{1v}y \ d_{2v}^{19}$, en dos puntos de la superficie $x_1y \ x_2$. Por ejemplo $d_{1v} = 1$ y $d_{2v} = 100$ respectivamente. Imagínese que la realización de dos medidas con error, d_{1o} y d_{2o} , ha producido como resultados $d_{1o} = 5$ y $d_{2o} = 104$.

Si, provisionalmente, se adopta la norma euclídea para medir errores, el proceso algorítmico de solución del problema inverso intentará calcular el modelo que minimice la expresión:

$$D = (d_{1o} - d_1(\mathbf{m}))^2 + (d_{2o} - d_2(\mathbf{m}))^2,$$
(73)

con la esperanza de hallar un modelo \mathbf{m} cuyas predicciones $d_1(\mathbf{m})$ y $d_2(\mathbf{m})$ se parezcan "suficientemente" a los valores medidos 5 y 104, respectivamente.

En este caso, la medida en el primer punto es errónea en un 500% mientras que la del segundo punto lo es sólo en un 4%. Sin embargo,

$$(d_{1o} - d_{1v})^2 = (d_{2o} - d_{2v})^2 = 16.$$

Es decir: ¡los dos parecen tener el mismo error!. O sea, afirmar que el error absoluto es 4, no informa sobre la *calidad* de la solución.

La formulación del error en términos absolutos tiene, además, el inconveniente adicional de que su valor depende de las unidades. Por ejemplo, si en lugar de ohmios-metro, se expresan los resultados en miliohmios-metro, $[d_{1o} - d_{1v}]^2 = 16{,}10^6$ y el error pasaría de 4 a 4000. Naturalmente esto no es deseable.

La importancia que una equivocación supone, aparece cuando se compara la magnitud de dicho error con la magnitud de lo que se habría obtenido sin error, esto es, cuando se consideran errores relativos. En términos del error relativo, el funcional a minimizar en el ejemplo anterior sería:

$$D = \left(\frac{d_{1o} - d_1(\mathbf{m})}{d_{1v}}\right)^2 + \left(\frac{d_{2o} - d_2(\mathbf{m})}{d_{2v}}\right)^2.$$
 (74)

Si **m** fuese el modelo verdadero, el primer término es 16 y el segundo $16 \cdot 10^{-4}$.

Así pues, la formulación en términos de errores relativos tiene la virtud de identificar la importancia real de una discrepancia. Además es adimensional y, por tanto, invariante frente a un cambio de unidades.

 $^{^{19}}$ El subíndice v en d_{iv} indica que el i – $\acute{e}simo$ dato es verdadero, o sea, habría sido obtenido mediante un proceso $sin\ error$. El subíndice o alude a un dato realmente observado y, por tanto, sujeto a error.

Tanto con ésta como con la anterior expresión (73) del desajuste, ambos sumandos están igualmente ponderados en el cálculo de la solución del problema inverso. El método "detecta" que el error en el primer término es inasumible -lo cual está bien, porque es cierto-y tratará entonces de rebajar su valor a toda costa, aún a expensas de incrementar el valor del segundo. Esto ya no es deseable. Un método adecuado debería distinguir cuándo una alta discrepancia relativa se debe a una mala predicción de una buena medida -en cuyo caso hay que reducir esa discrepancia- y cuándo ese alto valor se debe a una buena predicción pero a una medida anómala. En este caso, no hay que intentar aproximar la predicción a dicha medida errónea. No obstante, esto ya no depende del uso de los errores relativos en la definición del funcional a minimizar. Se consigue introduciendo una ponderación adicional en cada uno de los términos del desajuste.

Como inconveniente de la formulación (74) cabe citar que no se dispone de los valores de los datos medidos sin error. Sin transformar el espacio de datos hay dos alternativas para solventar este problema. Una consiste en sustituir los valores verdaderos d_{1v} o d_{2v} por las predicciones del propio modelo. Esto es:

$$D_{rel} = \left(\frac{d_{1o} - d_1(\mathbf{m})}{d_1(\mathbf{m})}\right)^2 + \left(\frac{d_{2o} - d_2(\mathbf{m})}{d_2(\mathbf{m})}\right)^2.$$
 (75)

Aunque sería lo correcto, al proceder así se complican los cálculos de las matrices de sensibilidad, con el consiguiente coste en tiempo. Otro modo de salvar la situación es utilizar los valores observados como denominador:

$$D = \left(\frac{d_{1o} - d_1(\mathbf{m})}{d_{1o}}\right)^2 + \left(\frac{d_{2o} - d_2(\mathbf{m})}{d_{2o}}\right)^2$$

$$(76)$$

Si bien es simple la corrección a incluir en el cálculo de las matrices jacobianas, esta formulación tiene la desventaja de que, si se repite la toma de datos, cambia la importancia de los errores (porque cambian las medidas).

Todas estas consideraciones hacen ver la importancia de la presencia de los errores relativos en la formulación del desajuste y, también, las dificultades que se encuentran al realizar esta tarea sin transformar el espacio de datos.

Si se opta por cambiar el espacio de datos mediante una transformación logarítmica, en los puntos de medida x_1 y x_2 se obtendrán los datos medidos $d_{1o}^{\log} = \log(d_{1o})$ y $d_{2o}^{\log} = \log(d_{2o})$, respectivamente. Entonces:

$$D = \left(d_{1o}^{\log} - d_{1}^{\log}\left(\mathbf{m}\right)\right)^{2} + \left(d_{2o}^{\log} - d_{2}^{\log}\left(\mathbf{m}\right)\right)^{2} \stackrel{def}{=} \log^{2} \frac{d_{1o}}{d_{1}\left(\mathbf{m}\right)} + \log^{2} \frac{d_{2o}}{d_{2}\left(\mathbf{m}\right)} =$$

$$= \log^{2} \left(1 + \frac{d_{1o} - d_{1}\left(\mathbf{m}\right)}{d_{1}\left(\mathbf{m}\right)}\right) + \log^{2} \left(1 + \frac{d_{2o} - d_{2}\left(\mathbf{m}\right)}{d_{2}\left(\mathbf{m}\right)}\right),$$
(77)

y, para pequeños errores relativos, este funcional es aproximadamente igual que D_{rel} .

2. Las dificultades numéricas del problema, relevantes en particular cuando se aborda la solución del problema inverso mediante métodos de optimización local. Esto es así porque

el cambio de parametrización en el espacio de datos influye en la matriz jacobiana en cada iteración. Aunque la estructura detallada de dicha matriz se examinará posteriormente, es necesario precisar en este momento que, si hay s puntos de medida y n parámetros desconocidos, será una matriz $\mathbf{JF} \in \mathbf{M}(s,n)$. La fila i-ésima de \mathbf{JF} contiene el vector gradiente de la predicción del funcional en el punto de medida i-ésimo. Denotando $F_i \stackrel{def}{=} F(x_i, \mathbf{m})$, se tiene

$$\mathbf{JF}_{i} = \mathbf{\nabla}^{t} F\left(x_{i}, \mathbf{m}\right) \stackrel{def}{=} \left(\frac{\partial F_{i}}{\partial m_{1}}, \frac{\partial F_{i}}{\partial m_{2}}, \cdots, \frac{\partial F_{i}}{\partial m_{n}}\right).$$

En el caso en que el espacio de datos sea logarítmico, la fila i-ésima de la nueva matriz jacobiana será

$$\mathbf{JF}_{i}^{\log} = \left(\frac{\partial \log (F_{i})}{\partial m_{1}}, \frac{\partial \log (F_{i})}{\partial m_{2}}, \cdots, \frac{\partial \log (F_{i})}{\partial m_{n}}\right) = \frac{1}{F_{i}} \left(\frac{\partial F_{i}}{\partial m_{1}}, \frac{\partial F_{i}}{\partial m_{2}}, \cdots, \frac{\partial F_{i}}{\partial m_{n}}\right).$$

Por lo tanto, todos los términos de la fila i-ésima quedan divididos por el valor de la predicción funcional en el punto i-ésimo. Si, en este punto, los valores fuesen altos, se reduciría considerablemente el valor de los elementos de la matriz de sensibilidad, empeorando sus propiedades numéricas. No obstante, en el caso de valores bajos, el efecto sería el opuesto y, por lo tanto, beneficioso.

En el ejemplo de las figuras 3 y 5 donde el vector de datos es $\mathbf{d} = (9.98, 8.75, 5.22)$ y el modelo de terreno es $\mathbf{m}_{sol} = (10, 5, 1)$, se presenta la matriz jacobiana con transformación logarítmica y sin ella en el espacio de datos y sus números de condición:

٠٠.	datos naturales	datos logarítmicos	
JF	$ \begin{bmatrix} 0.998 & 0.011 & 0.0036 \\ 0.8454 & 0.6158 & 0.2387 \\ 0.433 & 1.2973 & 0.8270 \end{bmatrix} $	$ \begin{bmatrix} 0.0434 & 0.0005 & 0.0002 \\ 0.0422 & 0.0308 & 0.0119 \\ 0.0365 & 0.1092 & 0.0696 \end{bmatrix} $	
$Cond\left(\mathbf{JF} ight)$	19.05	28.71	

Se aprecia como en este caso el número de condición de la matriz de sensibilidad es mayor en el caso logarítmico. Un número de condición grande perjudica la estabilidad numérica de la solución de un problema del tipo:

$$\min_{\mathbf{m} \in \boldsymbol{\delta}(\mathbf{m}_k)} \| \mathbf{d}_J + J_k \mathbf{m} \|_{N_{\mathbb{D}}},$$

o

$$\min_{\Delta \mathbf{m}_k \in \boldsymbol{\delta}(\mathbf{0})} \| \mathbf{d}_0 + J_k \Delta \mathbf{m}_k \|_{N_{\mathbb{D}}},$$

donde

$$J_k \equiv \mathbf{J}\mathbf{F}_k, \ \mathbf{d}_J = \mathbf{d}_0 - J_k \mathbf{m}_k.$$

4.2.3. El espacio de modelos y su parametrización

Además de las cuestiones anteriormente mencionadas para valorar la influencia de una reparametrización en el espacio de datos, al hablar del espacio de modelos es necesario hacer precisiones sobre:

- 1. La elección de la magnitud física: conductividad .vs. resistividad.
- 2. El orden en que se disponen los parámetros característicos de cada capa.
- 3. Razones para el empleo de una parametrización logarítmica.

En el planteamiento de este trabajo se ha escogido la resistividad porque para resolver el problema directo se emplea la teoría de filtros, que hace uso de esta parametrización.

En cuanto al **orden** en que se disponen los parámetros de cada capa en el vector que caracteriza al modelo, se han valorado dos posibilidades:

- en la primera, la resistividad y la potencia de cada capa ocupan lugares consecutivos. Así, $\mathbf{m} = (\rho_1, t_1, \rho_2, t_2, \cdots \rho_n)$.
- En la segunda primero se colocan las resistividades de las capas y después los espesores de todas ellas. De esta otra forma, $\mathbf{m} = (\rho_1, \rho_2, \cdots, \rho_n, t_1, t_2, \cdots t_{n-1})$.

Se ha escogido la primera opción, básicamente por dos motivos. Como se verá después, el término \mathbf{JF}_{ij} de la matriz de sensibilidad puede interpretarse como una medida de la influencia que un cambio en el parámetro j-ésimo tiene en el valor del dato i-ésimo. Convendría pues, que se encuentre agrupada, en dicha matriz, la información que compete a los dos parámetros de la misma capa. En la segunda formulación, estos valores se encontrarían espacialmente separados. Por otro lado, la influencia de una capa sobre un dato no depende a veces del cambio de un parámetro (independientemente del otro), sino del cambio de cierta combinación de los dos parámetros de dicho estrato²⁰. Es mucho más fácil localizar este hecho en la matriz jacobiana si la información atinente a los dos parámetros de la misma capa está agrupada.

A continuación se exponen razones que justificarían el empleo de log (m) en lugar de m:

• si se utiliza \mathbf{m} , habría que excluir explícitamente del proceso de búsqueda de solución la posibilidad de hallar potencias o resistividades negativas. Es decir, hay que incorporar siempre la restricción²¹ $\mathbf{m} \geq \mathbf{0}$. Si se elimina la posibilidad de resistividades nulas o de espesores nulos (reemplazándolos por valores pequeños) la transformación logarítmica $\mathbf{m}_{\log} = \log(\mathbf{m})$, está bien definida y convierte el intervalo $(0, \infty)$ en $(-\infty, \infty)$, transformándose el problema inverso en uno sin restricciones.

²⁰En particular, las reglas de equivalencia de Maillet (1947) indican la relevancia de las combinaciones $\frac{t_i}{\rho_i}$ y $t_i\rho_i$ denominadas, respectivamente, conductancia longitudinal y resistencia transversal. Se analizarán posteriormente en el capítulo dedicado al análisis de las equivalencias.

²¹En ocasiones esta condición se impone estipulando un espacio de búsqueda cuyo extremo inferior sea estrictamente positivo.

■ Por otro lado, los valores de resistividad presentes simultáneamente en un terreno real pueden diferenciarse en varios órdenes de magnitud. Es un grave problema conseguir operar con cantidades como 10⁵ y 10⁻⁵ a la vez. El empleo de logaritmos convierte el problema en la operación con los exponentes, de diferencias manifiestamente menores (entre 5 y −5 en este caso).

- Algunas reglas empíricas de equivalencia están formuladas en términos de la diferencia de valor relativo entre los parámetros del modelo. Ya se ha visto al hablar de los datos, como los logaritmos pueden considerarse buenas aproximaciones de dichos errores.
- En su trabajo pionero sobre equivalencias en un suelo de tres capas, Maillet (1947) considera dos casos diferentes:
 - 1. capa intermedia de resistividad manifiestamente inferior a la resistividad de la capa subyacente. En este caso podrá ser sustituida por otra capa que mantenga constante el ratio $c_i = \frac{t_i}{\rho_i}$ (conductancia longitudinal) de la capa.
 - 2. Capa intermedia de resistividad notablemente superior a la resistividad de la capa subyacente. La sustitución por la capa equivalente exige que ésta mantenga constante el producto $r_i = \rho_i t_i$ (resistencia transversal).

La idea, que será desarrollada en más detalle en el apartado destinado a las equivalencias, es la siguiente: si se parte de un modelo \mathbf{m}_0 solución de un problema inverso y $\Delta \mathbf{m}_0$ es una modificación del mismo, en la que se supone -por simplicidad- que sólo han variado las componentes relativas a la capa *i*-ésima, se plantea la siguiente pregunta: ¿cómo se reconoce si se ha llegado a una situación equivalente?.

Bajo una parametrización natural, necesariamente hay que contar con los dos vectores \mathbf{m}_0 y $\Delta \mathbf{m}_0$. Sumando, se obtendría la situación final. La formulación en logaritmos, por otro lado, permite identificar directamente esta situación de equivalencia a partir únicamente del vector $\Delta \mathbf{m}_0$. En efecto, si $\mathbf{m}_0 = (\rho_1, t_1, \rho_2, t_2, \rho_3)$ es un suelo de tres capas, utilizando logaritmos se tiene que:

$$\mathbf{m}_{0 \log} = (\log \rho_1, \log t_1, \log \rho_2, \log t_2, \log \rho_3).$$

Por lo tanto

$$\Delta \mathbf{m}_{0 \log} = \left(\log \rho_1, \log t_1, \log \frac{\rho_2 + \Delta \rho_2}{\rho_2}, \log \frac{t_2 + \Delta t_2}{t_2}, \log \rho_3\right).$$

Se analiza ahora cómo se reflejan las condiciones de equivalencia mencionadas con anterioridad (véase casos 1 y 2) en la estructura de $\Delta m_{0 \log}$:esto va en equivalencias

1. Si $\rho_2 \ll \rho_3$ y

$$\frac{t_2}{\rho_2} \simeq \frac{t_2 + \Delta t_2}{\rho_2 + \Delta \rho_2},\tag{78}$$

se presenta una de las situaciones contempladas como equivalentes. La relación (264) implica

$$\log\left(\frac{\rho_2 + \Delta\rho_2}{t_2 + \Delta t_2} \cdot \frac{t_2}{\rho_2}\right) \simeq 0 \Rightarrow \log\frac{\rho_2 + \Delta\rho_2}{\rho_2} \simeq \log\frac{t_2 + \Delta t_2}{t_2},\tag{79}$$

es decir, la cuasi-igualdad de dos términos consecutivos (de la misma capa) en $\Delta m_{0 \log}$.

2. Si
$$\rho_2 \gg \rho_3$$
 y
$$(t_2 + \Delta t_2) (\rho_2 + \Delta \rho_2) \simeq \rho_2 t_2,$$
(80)

entonces se está ante el segundo de los casos de equivalencia. La relación (266) implica

$$\log \frac{(t_2 + \Delta t_2) \cdot (\rho_2 + \Delta \rho_2)}{\rho_2 \cdot t_2} \simeq 0 \Rightarrow \log \frac{\rho_2 + \Delta \rho_2}{\rho_2} \simeq -\log \frac{t_2 + \Delta t_2}{t_2}, \tag{81}$$

es decir, los dos términos consecutivos (de la misma capa) en $\Delta m_{0 \log}$ poseen casi el mismo valor absoluto **pero con signo contrario.**

Por lo tanto tras calcular $\Delta m_{0 \log}$ hay que observar si se cumple alguna de las condiciones (265 ó 267).

Naturalmente la reparametrización puede resultar desventajosa en algunos casos:

 la transformación logarítmica en el espacio de parámetros, al igual que en el espacio de datos, cambia también los valores de la matriz de sensibilidad. La columna j-ésima de la nueva matriz jacobiana es:

$$(\mathbf{JF}_{\log})_j^t = \left(\frac{\partial F_i}{\partial \log m_j}, \frac{\partial F_2}{\partial \log m_j}, \cdots, \frac{\partial F_m}{\partial \log m_j}\right) = m_j \cdot \left(\frac{\partial F_i}{\partial m_j}, \frac{\partial F_2}{\partial m_j}, \cdots, \frac{\partial F_3}{\partial m_j}\right).$$

2. Desde un punto de vista probabilístico, la reparametrización cambia la distribución de probabilidades de los modelos alrededor de la solución²².

4.3. La matriz jacobiana

La matriz jacobiana establece la relación lineal entre el espacio de datos \mathbb{D} y el de modelos \mathbb{M} . También denominada matriz de sensibilidad, proviene -recuérdense las expresiónes (186) y (187), de la linealización del funcional $\mathbf{F}(\mathbf{m})$ en el entorno de un punto \mathbf{m}_k , según la expresión:

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_k) + \begin{bmatrix} \mathbf{\nabla}^t F_1(\mathbf{m}_k) \\ \vdots \\ \mathbf{\nabla}^t F_s(\mathbf{m}_k) \end{bmatrix} (\mathbf{m} - \mathbf{m}_k) + \mathbf{o}(\|\Delta \mathbf{m}_k\|).$$
(82)

Los datos obtenidos en la práctica son las resistividades aparentes, por lo tanto se identifica $\mathbf{F}(\mathbf{m})$ con $\boldsymbol{\rho}_{ap}(\mathbf{m})$. Si se conviene en denotar a la medida *i*-ésima como $\boldsymbol{\rho}_{ap}^i$ y se toman los valores de resistividades y espesores como parámetros del modelo, la fila *i*-ésima de la matriz de sensibilidad será el vector

$$\mathbf{JF}_{i}^{t} = \left(\frac{\partial \boldsymbol{\rho}_{ap}^{i}}{\partial \rho_{1}}(\mathbf{m}_{k}), \frac{\partial \boldsymbol{\rho}_{ap}^{i}}{\partial t_{1}}(\mathbf{m}_{k}), \cdots, \frac{\partial \boldsymbol{\rho}_{ap}^{i}}{\partial \rho_{n-1}}(\mathbf{m}_{k}), \frac{\partial \boldsymbol{\rho}_{ap}^{i}}{\partial t_{n-1}}(\mathbf{m}_{k}), \frac{\partial \boldsymbol{\rho}_{ap}^{i}}{\partial \rho_{n}}(\mathbf{m}_{k})\right). \tag{83}$$

Obsérvese que

$$\mathbf{JF}_{i}^{2k} = \frac{\partial \boldsymbol{\rho}_{ap}^{i}}{\partial t_{k}} \left(\mathbf{m}_{k} \right), \ k = 1, \dots, n-1,$$

²²En la modelización es frecuente otorgar, a priori y en función de la experiencia, a unos modelos más "credibilidad" que a otros. Las distribuciones de probabilidad son la herramienta matemática que permite cuantificar este hecho.

es decir, representa -en aproximación de primer orden- la influencia que un "pequeño" cambio en el espesor de la capa k-ésima tiene en el valor del dato $\rho_{ap}^i \stackrel{def}{=} \rho_{ap}(x_i)$. Por su parte, las columnas de lugar impar, cuyos términos son de la forma \mathbf{JF}_i^{2k+1} para $k=1,\cdots,n-1$, contienen información análoga respecto al efecto de una "pequeña" modificación de la resistividad de la capa k-ésima sobre la medida ρ_{ap}^i . Por esta razón, a la matriz $\mathbf{JF}(\cdot)$ se la denomina matriz de sensibilidad: sus términos miden cuán sensible es un dato a una pequeña modificación de cada uno de los parámetros. Para calcular de forma efectiva la matriz jacobiana se utiliza la expresión ya mencionada:

$$\rho_{ap}(x_i, \mathbf{m}) = \sum_{n=n1}^{n=n2} T(\lambda_n) \cdot h(x_i - \lambda_n), \qquad (84)$$

que indica que la resistividad aparente ρ_{ap}^i se puede obtener como una combinación lineal de muestras de $T(\lambda_n)$ cuyos coeficientes provienen del muestreo de la función filtro h. La expresión (84) permite obtener la siguiente relación entre las matrices jacobianas de resistividades aparentes \mathbf{J}_{ρ} y de transformadas de resistividad \mathbf{J}_T

$$\mathbf{J}_{\rho(m,n)} = H \cdot \mathbf{J}_T. \tag{85}$$

 $H \in \mathcal{M}(s,k)$ es una matriz que contiene los valores del filtro empleado en la convolución (siendo k es el número de puntos empleados para calcularla) y \mathbf{J}_T posee expresión analítica recurrente basada en las relaciones de Pekeris. Como la mayor parte de los elementos de H tienen un valor muy pequeño (Ghosh, 1970), el efecto que tiene multiplicar la matriz \mathbf{J}_T por la matriz H suele ser empeorar el condicionamiento de la misma. Así pues, numéricamente, la matriz \mathbf{J}_T está mejor condicionada.

4.3.1. Obtención de J_{ρ}

La matriz de sensibilidad \mathbf{J}_{ρ} puede calcularse por distintas vías:

- numéricamente mediante algún esquema de diferencias finitas²³.
- A partir de la expresión (85) cuyo cálculo se detalla a continuación.

Las derivadas de la transformada de resistividad respecto a los parámetros del modelo son (Johansen, 1975 y Koefoed, 1977):

$$\frac{\partial T_m}{\partial t_m} = \frac{\rho_m - \frac{T_{m+1}^2}{\rho_m}}{\left(1 + T_{m+1} \frac{\tanh(\lambda t_m)}{\rho_m}\right)^2} \cdot \frac{\lambda}{\cosh(\lambda t_m)},\tag{86}$$

$$\frac{\partial T_m}{\partial \rho_m} = \frac{\tanh(\lambda t_m) \left(1 + \frac{T_{m+1}^2}{\rho_m^2} + 2 \frac{T_{m+1}}{\rho_m} \tanh(\lambda t_m) \right)}{\left(1 + T_{m+1} \frac{\tanh(\lambda t_m)}{\rho_m} \right)^2}.$$
 (87)

²³Este es el método que utilizan por defecto los programas de optimización de Matlab.

Para calcular las derivadas de la transformada de resistividad de la capa m-ésima respecto a los parámetros de la k-ésima capa hay que tener en cuenta que:

$$\frac{\partial T_m}{\partial t_k} = \begin{cases}
0, & \text{si } k < m, \\
\frac{\partial T_m}{\partial T_{m+1}} \frac{\partial T_{m+1}}{\partial T_{m+2}} \cdots \frac{\partial T_k}{\partial t_k}, & \text{si } k \ge m,
\end{cases}$$

$$\frac{\partial T_m}{\partial \rho_k} = \begin{cases}
0, & \text{si } k < m, \\
\frac{\partial T_m}{\partial T_{m+1}} \frac{\partial T_{m+1}}{\partial T_{m+2}} \cdots \frac{\partial T_k}{\partial \rho_k}, & \text{si } k \ge m.
\end{cases}$$
(88)

$$\frac{\partial T_m}{\partial \rho_k} = \begin{cases}
0, & \text{si } k < m, \\
\frac{\partial T_m}{\partial T_{m+1}} \frac{\partial T_{m+1}}{\partial T_{m+2}} \cdots \frac{\partial T_k}{\partial \rho_k}, & \text{si } k \ge m.
\end{cases}$$
(89)

Obsérvese que la transformada de resistividad de una capa no depende de los parámetros asociados a las capas superiores. Para evaluar dichas expresiones hace falta determinar primero las derivadas del tipo $\frac{\partial T_m}{\partial T_{m+1}}$, es decir, de la transformada de resistividad de una capa respecto a la transformada de resistividad de la capa inferior. Dicha derivada puede calcularse directamente a partir de las relaciones de recurrencia de Pekeris:

$$\frac{\partial T_m}{\partial T_{m+1}} = \frac{1 - \tanh^2(\lambda t_m)}{\left(1 + T_{m+1} \frac{\tanh(\lambda t_m)}{\rho_m}\right)^2}.$$
(90)

4.3.2. Interpretación física de J_T

Cada una de las filas de la matriz \mathbf{J}_T contiene el resultado de derivar la transformada de resistividad respecto a todos los parámetros del suelo dispuestos de acuerdo con lo indicado en el párrafo anterior, en un sólo punto (un único valor de la variable λ). Recuérdese que cuando $\frac{1}{\lambda}$ aumenta, el espaciado interelectródico también aumenta y viceversa. Así pues, valores de λ pequeños se corresponden con valores grandes del semiespaciado interelectródico

Los puntos en los que se ha calculado la transformada de resistividad tienen por abscisa $y = \ln(\frac{1}{\lambda})$. La primera fila de \mathbf{J}_T corresponde a la menor abscisa y la última fila a la mayor de ellas. Por tanto, "al aumentar el número de fila se observa lo que ocurre para mayores distancias entre los electrodos de corriente, o sea, a profundidades de investigación mayores".

Considerada individualmente la fila i-ésima, su columna j-ésima representa la sensibilidad de la transformada de resistividad en el punto y_i a los cambios en el parámetro m_i . Más concretamente, si j = 2k - 1 el parámetro m_j es la resistividad de la capa k-ésima. Cuando j = 2k, el parámetro m_i es el espesor de la misma capa.

Debido al orden escogido para el vector de parámetros, al aumentar el índice j y progresar hacia la derecha a lo largo de la fila escogida, aparecerán las sensibiliades de la tranformada de resistividad respecto a los parámetros de capas cada vez más profundas. Así pues, "al aumentar el número de columna se observa la dependencia respecto de capas más profundas".

La estructura de la matriz de sensibilidad

El número de filas de la matriz es igual al número de puntos en los que se quiere calcular la transformada de resistividad de la primera capa, lo cual depende del número de ciclos logarítmicos sobre los que se extiende y del número de intervalos en los que se divide cada ciclo

logarítmico. El número de columnas es igual al doble del número de capas (incluida la última) menos uno.

El vector de parámetros es un vector columna que comienza por los parámetros de la primera capa y finaliza por los de la última. Dentro de una capa, primero aparece en el vector de parámetros la resistividad y después el espesor. Esta ordenación condiciona la estructura de las columnas en la matriz jacobiana. Éstas se organizan por parejas. Las dos primeras columnas, representan las derivadas respecto a los parámetros de la primera capa; las dos siguientes las derivadas respecto a los parámetros de la segunda capa y así consecutivamente hasta llegar a la última columna. Ésta, contiene las derivadas respecto a la resistividad de la última capa. No hay ninguna columna adicional reservada para la derivada respecto al espesor de dicha capa, sencillamente porque este espesor, en el método de S.E.V, se considera infinito.

4.4. Distancia y norma. Definiciones y repercusión de su elección

En el presente trabajo hay dos conjuntos esenciales: el espacio de los datos $\mathbb{D} \subset \mathbb{R}^s$ y el espacio de los modelos $\mathbb{M} \subset \mathbb{R}^{2k-1}$. Es básico disponer, en el espacio de datos, de una forma precisa de "medir" la diferencia entre los datos observados en el campo y las predicciones del método de cálculo. Se verá, además, que pueden existir distintos modelos de terreno -elementos de \mathbb{M} - cuyas predicciones "disten" de los datos observados aproximadamente lo mismo. Aunque los datos admitan ambos modelos como igual de buenos o equivalentes, puede que se disponga de información externa -relativa al valor de algún parámetro o de alguna característica estructural-que, tomada como información de referencia, aconseje descartar los modelos que se "alejen demasiado" del comportamiento de referencia.

El concepto matemático que permite "medir" la cercanía (o lejanía) entre dos elementos de un conjunto -por ejemplo el de datos, \mathbb{D} , o el de modelos \mathbb{M} - se denomina, acertadamente, distancia.

Una distancia definida en un conjunto genérico C, es una aplicación $d: C \to \mathbb{R}^+ \cup \{0\}$ que satisface los siguiente requisitos:

- 1. $d(\mathbf{x}, \mathbf{y}) \ge 0$ para todo $\mathbf{x}, \mathbf{y} \in C$ y $d(\mathbf{x}, \mathbf{y}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{y}$.
- 2. $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$.
- 3. $d(\mathbf{x}, \mathbf{z}) \leq d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z})$.

En un conjunto, por ejemplo en el espacio de datos \mathbb{D} , sin dotarlo de la noción de métrica o distancia, cabe hablar de la diferencia entre las predicciones correspondientes a un modelo y los datos observados. Sin embargo no puede darse el paso siguiente que decide si las predicciones están "lejos" o "cerca" de las medidas. Para eso hace falta dotar al conjunto \mathbb{D} de una métrica o distancia d. Con ella, cada vez que se hace referencia a \mathbb{D} se considera realmente el par (\mathbb{D}, d) , y se tiene por tanto una medida de la distancia entre predicciones y observaciones.

La idea de métrica o distancia es muy general. Una métrica predica algo de cada par de elementos de un conjunto. ¡Pero no dice nada de un elemento considerado de forma aislada!. En un espacio métrico no existe, por ejemplo, nada parecido al concepto de "longitud" de un elemento. El instrumento matemático que permite incorporar esta idea se denomina norma. Cuando en un conjunto se define una de estas normas, de cada elemento se puede decir algo

que, intuitivamente, correspondería a afirmar que es "grande" o "pequeño". En términos más estrictos, una norma definida en un conjunto C es una aplicación

$$N: C \subset \mathbb{R}^n \to \mathbb{R}^+ \cup \{0\}$$
,

que cumple las siguientes propiedades:

- 1. $N(\mathbf{x}) \ge 0$ para todo $\mathbf{x} \in C$ y $N(\mathbf{x}) = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.
- 2. $N(\lambda \mathbf{x}) = |\lambda| \cdot N(\mathbf{x})$.
- 3. $N(\mathbf{x} + \mathbf{y}) \le N(\mathbf{x}) + N(\mathbf{y})$.

Lo interesante es que si un conjunto es normado (es decir, se ha definido en él una norma) siempre puede construirse una definición de distancia entre dos elementos del conjunto, derivada de la norma elegida. Basta definir la distancia entre dos elementos \mathbf{x} e \mathbf{y} de C como $d(\mathbf{x}, \mathbf{y}) = N(\mathbf{x} - \mathbf{y})$. Por lo tanto, con el sólo concepto de norma se dispone simultáneamente de una forma de medir "distancias" y "tamaños". En el caso de los S.E.V tanto el espacio de datos como el de modelos son espacios vectoriales de dimensión finita. En este tipo de espacios, hay varias posibilidades de definición de normas. Una familia muy importante, que se utilizará en exclusiva, es la simbolizada mediante $N \equiv \|\cdot\|_{l_p}$ con $p = 1, 2, 3, \cdots$. Su expresión es:

$$\|\mathbf{x}\|_{l_p} = \left(\sum_{i=1}^{i=n} |x_i|^p\right)^{1/p}.$$
 (91)

Según la norma l_p , la distancia entre dos elementos \mathbf{x} e \mathbf{y} se calcularía como:

$$d(\mathbf{x}, \mathbf{y}) = N(\mathbf{x} - \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_{l_p} = \left(\sum_{i=1}^{i=n} |x_i - y_i|^p\right)^{1/p}.$$
 (92)

Volviendo al ejemplo anterior en que los datos observados son $\mathbf{d}_o = (1, 100)$. y los predichos son. $\mathbf{d}_p = (5, 104)$, para cada valor de p se obtiene una medida de la distancia entre predicciones y observaciones. La diversidad de valores para la distancia es, simplemente, una manifestación del hecho, a veces mal comprendido, de que los números no sustituyen a la persona que modeliza en la toma de decisiones. En los epígrafes siguientes se intentarán suministrar los criterios de elección a través de la exposición razonada del significado de cada tipo de norma.

4.5. Influencia del tipo de norma en el espacio de los datos

4.5.1. Construcción de un modelo matemático de las medidas. Las distribuciones de probabilidad

Para poder dar un sentido a las operaciones que se realizan sobre los datos medidos es preciso disponer de una representación matemática adecuada de los mismos, esto es, un *modelo matemático de los datos*. Este modelo suele suponer la consideración del proceso de medida en cada punto como un proceso aleatorio que genera una variable aleatoria-la medida en dicho

punto. El proceso de medir en un punto de la superficie del terreno quedaría representado matemáticamente, pues, mediante una distribución de probabilidades.

Todas las metodologías de resolución del problema inverso consideran que esta distribución de probabilidades tiene, al menos, los dos momentos de primer y segundo orden, es decir: poseen **media** y **varianza**. La *media* se considera como la medida que se obtendría en ese punto si no hubiese ningún tipo de perturbación aleatoria en la medida. Por lo tanto, la *media* sería la predicción sobre el suelo real.

La dispersión inevitable de las medidas sujetas a perturbación, suele medirse mediante la varianza. Se supone que esta dispersión es íntegramente debida a las características fluctuantes e imprevisibles del proceso de medida en sí mismo y, por tanto, que no contiene información sobre la estructura del suelo.

Si se considera a la media de la distribución como lo que, se podría decir, se mediría sobre el suelo "verdadero", tiene sentido denominar error de una medida a la diferencia entre el valor de la misma (considerada como una realización de una variable aleatoria) y el valor de la media de la población. Por tanto, el error es también una variable aleatoria del mismo tipo que la variable aleatoria dato, con media nula y la misma varianza.

4.5.2. La familia de las Gaussianas generalizadas

Se ha mencionado ya anteriormente que es usual suponer que la variable aleatoria que modeliza el dato en un punto²⁴ -y la de los errores- sigue una distribución normal²⁵. Sin embargo, no es éste el único modelo de aleatoriedad posible. En principio, cabe suponer distintas funciones de distribución. En particular, existe una familia extraordinariamente interesante de cuyo análisis se extraerán las consecuencias del uso de diferentes normas para la medición de distancias en el espacio de datos. Se trata de la familia de las distribuciones gaussianas generalizadas. Sus miembros se caracterizan por su centro x_0 , un coeficiente de dispersión σ_p y un parámetro numérico, p. Su expresión es:

$$f(x) = \frac{p^{1-\frac{1}{p}}}{2\sigma_p \Gamma\left(\frac{1}{p}\right)} \exp\left(\frac{-1}{p} \frac{|x - x_0|^p}{(\sigma_p)^p}\right),\tag{93}$$

donde $\Gamma(\cdot)$ es la función gamma y el parámetro de dispersión σ_p tiene por expresión:

$$\left(\sigma_{p}\right)^{p} = \int_{-\infty}^{+\infty} \left|x - x_{0}\right|^{p} \rho\left(x\right) dx. \tag{94}$$

Esta familia contiene algunos miembros ilustres, conocidos con nombres más familiares debido a su importancia y frecuente uso. Por ejemplo, el valor p=1 caracteriza a la distribución laplaciana o doble-exponencial. El miembro con p=2 es, sencillamente, la distribución gaussiana. Para $p=\infty$ se recupera la distribución uniforme en el intervalo $[x_0-\sigma_\infty,\ x_0+\sigma_\infty]$. En la figura 6 se representan los miembros de esta familia con centro $x_0=0$ y $\sigma_p=1$, correspondientes a p=1,2,4 y 20 respectivamente.

²⁴En adelante, salvo cuando por el contexto no esté claro y por simplicidad, la palabra dato hará referencia tanto a los propios datos como a la variable aleatoria que los modeliza.

²⁵Con las debidas limitaciones mencionadas en dicho epígrafe y relacionadas con la magnitud de la varianza. La normalidad de la distribución viene sugerida por la ley de los grandes números en caso de que las variables aleatorias que modelizan las fluctuaciones se combinen de forma aditiva y sean suficientes en número.

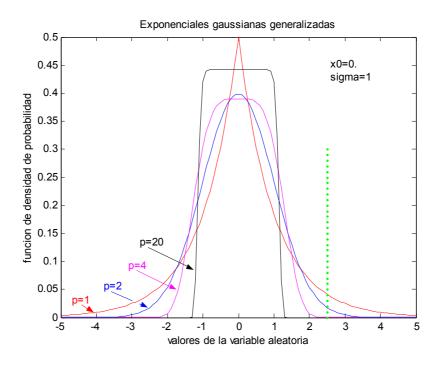


Figura 6: Familia de Gaussianas generalizadas.

En la figura se observa que los miembros que tienen una p menor, tienen colas mayores. Es decir, asignan mayor probabilidad a sucesos más distantes del centro. Si se acuerda en calificar a una realización de la variable aleatoria dato como anómala si le corresponde una probabilidad muy baja, a la vista de la figura 6 parece apropiado afirmar que según qué miembro de la familia se considere, el dato será o nó anómalo. A la misma distancia del centro común, el dato x=2.5no será anómalo para la distribución exponencial (p=1) y sí lo será para las distribuciones con p=1,2 y, por supuesto, 20. La misma idea, contemplada desde un ángulo diferente, consiste en tomar una realización posible de la medida, p.ej. x = 30 y estipular que debe suceder con cierta probabilidad -en el ejemplo ilustrado en la figura 7 ésta es del 5%- atribuirle a priori una de las distribuciones de la familia y preguntarse dónde tiene ésta su centro. El resultado puede visualizarse en la figura 7. Se han vuelto a considerar las distribuciones con p=1,2,4 y 20. Al parámetro de dispersión σ_p se le ha atribuido el valor arbitrario de 1. Se ha resaltado el punto de coordenadas (30,0.05) común a todas las gráficas. Claramente se aprecia cómo se ha producido un desplazamiento relativo entre ellas. La distribución exponencial (p=1) es la que mantiene su centro más alejado del punto x = 30. La normal (p = 2) debe acercarse un poco más al mismo y la misma tendencia, aumentada, persiste en el caso p=4. Para p=20, que se pretende considerar una buena aproximación de la norma uniforme, la gráfica ha precisado desplazarse hasta situar su extremo superior sobre dicho punto.

La relevante conclusión que se extrae es que, para poder atribuirse la realización de un dato con cierta probabilidad (no considerarlo anómalo), el centro de una distribución con valor de p mayor debe acercarse más a dicho dato que una distribución con un valor de p más pequeño.

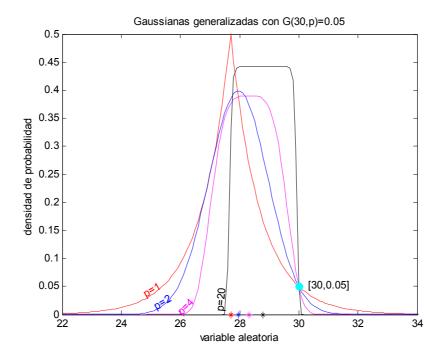


Figura 7: Forma de las diferentes Gaussianas generalizadas que asignan una probabilidad del 5% al suceso x=30.

4.5.3. Estimación de los parámetros de la distribución de datos. La máxima verosimilitud

Al introducir las distribuciones como modelos matemáticos de los datos medidos, se mencionaba el hecho de que se considera la *media* de la población como la predicción sobre el suelo real (incógnita en un problema inverso). Ya que esta *media poblacional* informa sobre la estructura del terreno, conviene calcularla.

Sin embargo, en cada punto de medida sólo se dispone-como máximo y en casos excepcionalesde un número finito de datos. Por tanto, la *media poblacional* no puede conocerse con certeza y es preciso conformarse con *estimarla* a partir de los datos muestrales.

Como es imposible que unos cuantos datos (medidas realizadas en un mismo punto de la superficie) contengan toda la información sobre la población, hay que suministrar un criterio de estimación que aporte la información faltante. Es ampliamente utilizado el criterio denominado de máxima verosimilitud²⁶. Para utilizar este criterio, previamente se asume a priori un tipo de distribución, o sea, se fija el parámetro p de la familia de gaussianas generalizadas. Es a la hora de calcular los parámetros restantes de esta distribución -media y varianza- cuando se aplica el criterio de máxima verosimilitud. Este criterio dice, básicamente, que cuando ocurre algo, ocurre lo más probable. Por lo tanto, se escogerán la media y la varianza que identifiquen a

²⁶La verosimilitud es un concepto que aparece en el contexto de la estimación bayesiana. Aunque en este trabajo no se adopta el marco bayesiano, se irán introduciendo los conceptos correspondientes donde sea pertinente.

una distribución concreta para la que los datos ocurridos sean los de máxima probabilidad (la moda de la distribución).

La necesidad de estimar y el uso del criterio de máxima verosimilitud proporcionan el enlace entre la distribución de los datos y la definición de la norma.

Supóngase que, mediante la repetición de N medidas sobre el mismo punto de la superficie, x_i , se ha obtenido el vector de medidas $\mathbf{d}_i = (d_1(x_i), d_2(x_i), \cdots d_N(x_i))$. Se hipotetiza que su modelo matemático es una distribución aleatoria de la familia de las gaussianas generalizadas. Las distintas repeticiones se supondrán -por simplicidad- independientes. Se pretende utilizar el criterio de máxima verosimilitud para estimar la media de la población μ_p . En principio, se supondrá conocida la dispersión σ_p .

Utilizar máxima verosimilitud (m.v) significa admitir que la probabilidad de ocurrencia de \mathbf{d}_i es máxima. Por tanto el problema a resolver es:

$$\max_{\hat{\mu}_p} \ f_p\left(\mathbf{d}_i, \hat{\mu}_p, \sigma_p\right) \overset{indep.}{=} \max_{\hat{\mu}_p} \prod_{j=1}^{j=N} f_p\left(d_j, \hat{\mu}_p, \sigma_p\right),$$

$$\begin{split} \max_{\hat{\mu}_p} \prod_{j=1}^{j=N} f_p\left(d_j, \hat{\mu}_p, \sigma_p\right) &= \frac{p^{1-\frac{1}{p}}}{2\sigma_p \Gamma\left(\frac{1}{p}\right)} \max_{\hat{\mu}_p} \prod_{j=1}^{j=N} e^{-\frac{1}{p} \frac{\left|d_j - \hat{\mu}_p\right|^p}{\left(\sigma_p\right)^p}} = \\ &= \frac{p^{1-\frac{1}{p}}}{2\sigma_p \Gamma\left(\frac{1}{p}\right)} \max_{\hat{\mu}_p} \left[e^{-\frac{1}{p} \sum_{j=1}^{j=N} \frac{\left|d_j - \hat{\mu}_p\right|^p}{\left(\sigma_p\right)^p}} \right] = \frac{p^{1-\frac{1}{p}}}{2\sigma_p \Gamma\left(\frac{1}{p}\right)} e^{-\frac{1}{p}} \max_{\hat{\mu}_p} \left[e^{-\frac{j=N}{p} \frac{\left|d_j - \hat{\mu}_p\right|^p}{\left(\sigma_p\right)^p}} \right]. \end{split}$$

Pero la solución de:

$$\max_{\hat{\mu}_p} \ e^{-\sum\limits_{j=1}^{j=N} \left| d_j - \hat{\mu}_p \right|^p},$$

es equivalente a:

$$\min_{\hat{\mu}_p} \left[\sum_{j=1}^{j=N} \left| d_j - \hat{\mu}_p \right|^p \right]. \tag{95}$$

Si se recuerda la expresión (91) para la norma, $\|\cdot\|_{l_p}$, el estimador $\hat{\mu}_p$ de la media poblacional buscado es -según la recién citada exprexión (95), el que minimiza $\|\mathbf{d}-\hat{\mu}_p\|_{l_p}^p$ y, por tanto, $\|\mathbf{d}-\hat{\mu}_p\|_{l_p}$. La interpretación que se expone es la siguiente: dado el vector de medidas \mathbf{d}_i , modelizado mediante la distribución f_p , la estimación por m.v. de la media poblacional μ_p es el valor $\hat{\mu}_p$ que "menos dista de las medidas tomadas" según la distancia inducida por la norma $\|\cdot\|_{l_p}$ (correspondiente al mismo valor p)²⁷.

²⁷Este es el punto de confluencia de las visiones determinista y probabilista de un problema de optimización (estimación).

4.5.4. Casos particulares relevantes: distribución normal y exponencial

En el caso de la **distribución gaussiana** (p=2) la estimación de máxima verosimilitud de la media poblacional, $\hat{\mu}_2$, es solución del problema de minimización:

$$\hat{\mu}_2 = \min_{\hat{\mu}_2} \sum_{j=1}^{j=N} (d_j - \hat{\mu}_2)^2.$$
 (96)

Esto ocurre para

$$\hat{\mu}_2 = \frac{1}{N} \sum_{i=1}^{j=N} d_j = \bar{d},\tag{97}$$

es decir, que la estimación de m.v. de la media poblacional es la media muestral.

Cuando, en la posición x_i , sólo hay un dato medido, la media coincide con él. Por tanto, él mismo es la estimación de m.v. de la media poblacional. Cuando se han tomado dos datos, la media muestral filtra las fluctuaciones de ambos y da una mejor estimación de la media poblacional que en el caso anterior. Supóngase que se dispone de un tercer punto de medida x_3 , cuyo dato muy distante de los anteriores (p.ej., un error del operador) y se acepta dicho dato como una realización legítima de la distribución gaussiana que modeliza el proceso de medida. Es decir, si se admite dicho dato como una medida más sin nada cualitativamente diferente, el valor de d_3 participa de pleno derecho en el cálculo de la nueva media muestral. Esto produce una desviación considerable -y espuria o ficticia- de la misma respecto a los valores anteriores.

Como $\hat{\mu}_2 = \bar{d}$ y la media muestral \bar{d} es tan extraordinariamente sensible a los datos anómalos, siempre que éstos estén presentes y no se reconozcan como tales, $\hat{\mu}_2$ será una mala estimación de la media poblacional μ_2 . Este comportamiento puede explicarse intuitivamente recurriendo nuevamente a las figuras 6 y 7. En ellas se aprecia cómo las colas de la distribución gaussiana son relativamente cortas. Por ello, para maximizar la probabilidad de ocurrencia de los tres datos tomados en x_3 , que es el producto

$$f_2(\mathbf{d}(x_3), \hat{\mu}_2, 0 \cdot 05) = f_2(d_1) \cdot f_2(d_2) \cdot f_2(d_3),$$

la probabilidad del nuevo dato $f_2(d_3)$ no puede tomar un valor demasiado bajo, lo cual obliga a la curva de la distribución gaussiana a acercarse bastante al nuevo punto, tal como se ilustraba en la figura 7, produciendo una desviación notable de su centro respecto al caso anterior.

La equivalencia de los problemas máx $f_2(\mathbf{d}_i, \hat{\mu}_2, \sigma_2)$ y mín $\sum_{\hat{\mu}_2}^{j=N} (d_j - \hat{\mu}_2)^2$ será importante en sucesivo. En particular afirmar que $\hat{\mu}_i$ so calcula minimizar en el caso anterior.

lo sucesivo. En particular afirmar que $\hat{\mu}_2$ se calcula minimizando la norma euclídea es equiavlente a decir que $\hat{\mu}_2$ estima la media de la población gaussiana que atribuye la máxima probabilidad justamente a los datos ocurridos. Por tanto, la minimización de la norma euclídea es una operación cuyo resultado puede ser bastante erróneo en caso de haber datos efectivamente anómalos, o sea, no procedentes de una distribución gaussiana.

En cuanto a la **distribución exponencial** (p = 1) el estimador $\hat{\mu}_1$ de la media poblacional μ_1 es según la expresión (95) particularizada para p = 1, la siguiente:

$$\hat{\mu}_1 = \min_{\hat{\mu}_1} \sum_{j=1}^{j=N} |d_j - \hat{\mu}_1|, \qquad (98)$$

Si se resuelve este problema, se obtiene como respuesta que $\hat{\mu}_1$ es la mediana muestral. En el caso de una única medida tomada en x_1 , la mediana muestral coincide con ella. Se vuelve a tener la distribución centrada en el dato $d_1(x_1)$. En el lugar de abscisa x_2 se han tomado dos medidas representables mediante el vector $\mathbf{d}(x_2) = (d_1(x_2), d_2(x_2))$. Ahora no hay punto central. ¿A qué corresponde, pués, la mediana muestral?. Se puede ver la respuesta intuitivamente si se piensa que se trata de minimizar la suma

$$E_1 = |d_1(x_2) - \hat{\mu}_1| + |d_2(x_2) - \hat{\mu}_1|.$$

Para fijar ideas considérese que $d_1(x_2) < d_2(x_2)$. La longitud del intervalo $[d_1(x_2), d_2(x_2)]$ es, por tanto, $l = d_2(x_2) - d_1(x_2)$.

Si, tentativamente, se colocase $\hat{\mu}_1$ a la izda (resp.dcha) de $d_1(x_2)$ (resp. de $d_2(x_2)$) el valor de E_1 sería $E_1 = l + 2(d_1 - \hat{\mu}_1)$, (resp. $E_1 = l + 2(\hat{\mu}_1 - d_2)$). Esta no puede ser la solución puesto que se puede disminuir aún el valor de E_1 haciéndolo coincidir con cualquiera de los dos extremos del intervalo, en cuyo caso $E_1 = l$.

Pero, ¿qué ocurriría si $\hat{\mu}_1$ perteneciese al intervalo $[d_1(x_2), d_2(x_2)]$. En ese caso

$$E_1 = (\hat{\mu}_1 - d_1) + (d_2 - \hat{\mu}_1) = d_2 - d_1 = l.$$

Es decir que cualquier $\hat{\mu}_1 \in [d_1(x_2), d_2(x_2)]$ minimiza la norma referida en (98) y es estimador de máxima verosimilitud de la media μ_1 de la población exponencial. Por lo tanto, la solución no es única²⁸. En caso de que el número de puntos sea par, el intervalo de soluciones es el comprendido entre los dos puntos centrales.

En el tercer lugar de toma de toma de datos, x_3 , el comportamiento del estimador $\hat{\mu}_1$ en presencia del dato extraño d_3 (x_3), es identificarse con la mediana, esto es, con el dato central, jindependientemente de lo desviado que esté el nuevo dato!.

Así pues, cuando la distribución de la que proceden los datos es exponencial, el estimador $\hat{\mu}_1$ de la media de la población se obtiene minimizando la norma (98), lo que da como resultado la mediana muestral. Este estimador es extraordinariamente estable o robusto frente a la presencia de valores anómalos, en el sentido de que, incluso cuando un pequeño número de éstos está en la muestra, estima bien la media μ_1 . Este comportamiento puede justificarse apelando -como se hizo a propósito del comentario de la figura 7- a la menor necesidad que tiene este tipo de distribución de acercarse a un dato para asumirlo con cierta probabilidad.

Sin embargo lo realmente interesante es que el estimador $\hat{\mu}_1$ también es un buen estimador de la media de la población incluso cuando ésta no es exponencial, por ejemplo en el caso de una distribución normal (Parker, 1994). Además, a diferencia del estimador m.v $\hat{\mu}_2$, media muestral, es mucho menos sensible a medidas anómalas.

4.5.5. El caso de varios puntos de medida

Cuando se lleva a la práctica el trabajo de campo de un S.E.V, la aplastante mayoría de las veces sólo se toma un dato en cada abscisa de medida (con cada apertura de electrodos). Si los datos $d(x_i)$ con $i = 1, 2, \dots n$ se considerasen realizaciones independientes de variables aleatorias

La norma $\|\cdot\|_{l_1}$ es la única, de entre todas las normas $\|\cdot\|_{l_p}$ que presenta esta no unicidad. Como se observa aquí, puede hacerse coincidir con uno de los vértices -reduciendo a cero el error de uno de los puntos de medida. Esta es una propiedad general comentada en Menke (1989).

con el mismo tipo de función de distribución -por ejemplo exponencial o gaussiana- pero con (posiblemente) distintos valores de μ y σ , el estimador de m.v de la media de la población coincidiría, en cada abscisa de medida, con el dato tomado.

Pero los datos no dependen solamente del proceso de medida, que es el que introduce aleatoriedad, sino también de la estructura del suelo. En particular la media de la distribución contiene la información sobre el terreno. Dicho de otro modo, el vector de medias, además de maximizar la probabilidad de que ocurran precisamente esos datos debe representar una predicción viable sobre alguno de los suelos admitidos como posibles.

Si el vector de datos es $\mathbf{d} = (d(x_1), d(x_2), \dots, d(x_s))$, un vector de medias $\boldsymbol{\mu} = \mathbf{d}$ maximizaría siempre la probabilidad de dichas medidas. Pero podría no existir ningún suelo *admitido* capaz de generar una predicción $\mathbf{F}(\mathbf{m}, \mathbf{x}) = \mu$. Por lo tanto, estas medias quedarían invalidadas como estimador compatible con un suelo razonable.

Como ejemplo ilustrativo considérese un suelo con una sola capa de resistividad desconocida. Supóngase que sobre dicho suelo, tras tomar medidas en dos puntos distintos, se obtiene el vector de datos $\mathbf{d} = (10 \cdot 2, 9 \cdot 8)$, resultante de la estructura del suelo y de la acción de un ruido que se asume gaussiano e independiente para cada medida. Consideradas como una pareja de datos procedentes de dos distribuciones normales, el vector $\hat{\boldsymbol{\mu}}_2$, estimador de m.v. de medias es: $\hat{\boldsymbol{\mu}}_2 = (10 \cdot 2, 9 \cdot 8)$. Por tratarse de un terreno homogéneo, $\mathbf{F}(\mathbf{m}, \mathbf{x}) = \rho$ en todo punto de la superficie. Como, además, también debe ser $\mathbf{F}(\mathbf{m}, \mathbf{x}) = \hat{\boldsymbol{\mu}}_2(\mathbf{x})$, resulta que el vector $\hat{\boldsymbol{\mu}}_2(\mathbf{x})$ tiene que ser constante con \mathbf{x} . Por lo tanto $\hat{\boldsymbol{\mu}}_2 = (10 \cdot 2, 9 \cdot 8)$ es incompatible con la clase supuesta de modelos homogéneos. No obstante, el estimador $\hat{\boldsymbol{\mu}}_2$ compatible con el tipo de suelo estipulado se puede calcular sistemáticamente si en la expresión:

$$\hat{\mu}_2 = \min \sum_{i=1}^{i=s} \left(\frac{d_i - \hat{\mu}}{\sigma_{i2}} \right)^2, \tag{99}$$

se sustituye $\hat{\mu}$ por su relación con un modelo **m** perteneciente al conjunto de modelos admisibles, es decir: $\hat{\mu} = \mathbf{F}(\mathbf{m}, \mathbf{x})$. Por lo tanto se trata de resolver el problema²⁹:

$$\hat{\mu}_2 = \min \sum_{i=1}^{i=s} \left(\frac{d_i - \mathbf{F}(\mathbf{m}, \mathbf{x})}{\sigma_{i2}} \right)^2.$$
(100)

En el caso considerado, al ser el suelo homogéneo, $\mathbf{F}(\mathbf{m}, \mathbf{x}) = \rho$. Al derivar respecto a ρ , se encuentra que:

$$\hat{\mu}_2 = \rho = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} d_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} d_2. \tag{101}$$

Esta expresión es una media aritmética ponderada de los datos d_1 y d_2 . Si se toman los valores σ_1^2 y σ_2^2 en la forma usual para las varianzas, como una medida de la dispersión de las variables aleatorias de las que proceden d_1 y d_2 alrededor de la media respectiva, la expresión (101) se interpreta de la siguiente manera: el factor $\frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$ tiende hacia la unidad cuando σ_2^2 crece o bien cuando σ_1^2 decrece. En el primer caso la dispersión de la variable aleatoria que genera la

²⁹En ambas expresiones, a diferencia de en (95) o de (96) se incluye la desviación típica porque ésta puede depender del lugar de medida. En las referencias citadas se consideraban sólo los datos tomados sobre el mismo punto de medida.

segunda medida es tan grande (al menos mucho más que la dispersión de la primera medida) que se asume que es bastante probable que haya ocurrido un dato lejano al valor medio. Por lo tanto no es razonable aproximar la estimación de la media $\hat{\mu}_2$ al segundo de los datos d_2 , lo cual se consigue porque su peso toma un valor muy bajo. Simultáneamente, es mucho más razonable aproximar la estimación de la media a d_1 -que procede de una distribución con dispersión mucho más pequeña- lo cual se logra si se observa como su coeficiente de ponderación tiende a uno en esas circunstancias. Este método se conoce también con el nombre de mínimos cuadrados sopesados.

4.6. Normas en el espacio de modelos e información a priori

En el ámbito de la resolución de problemas inversos es frecuente utilizar el concepto de información a priori. En el contexto de los S.E.V, y en particular en este trabajo, se considerará a priori toda información que, introducida en el proceso de resolución de un problema, influya en su solución final y aporte información adicional no contenida en los datos procedentes de las medidas de campo (comunmente denominados "duros").

Esta definición es bastante amplia. Conviene concretar algunas de las posibilidades que incluye y que, a veces, no son consideradas como a priori o, simplemente, se ignoran, y que sin embargo son consideradas en este trabajo.

1. Información de tipo 1

Es información sobre la que se admite "cierto grado de certeza", relativa a los parámetros geoeléctricos del terreno que se pretenden calcular. A esta clase pertenece por ejemplo, el admitir que una determinada resistividad (o espesor) de alguna(as) capas tomen valores fijados o que pueden variar en un rango de determinada amplitud. Esta información procede, normalmente, de experiencias previas sobre terrenos similares, de datos que, con otras técnicas se han obtenido sobre el mismo terreno (por ejemplo sondeos mecánicos) o, simplemente puede pretender plasmar la "intuición experta" de trabajadores experimentados en áreas relacionadas. Normalmente esta acepción es la más comunmente comprendida como información a priori.

2. Información numérica o de tipo 2

Es la información introducida (a veces de modo implícito) cuando se opta por una u otra formulación matemática y numérica del problema.

El camino utilizado para resolver el problema a partir de los datos y la información "tipo1" no es único. De hecho, no hay problema hasta que no se escoja una formulación de entre las distintas existentes. Esto se analizará posteriormente pero cabe ilustrarlo con lo expuesto hasta este momento. En el ejemplo que se visualiza en la figura ??, la información tipo1 atinente al terreno sería su clasificación como homogéneo³⁰ $\mathbf{m} = \rho$. Otra información, también de tipo1, relativa a las medidas consiste en su adscripción -antes de la confirmación mediante análisis estadístico de los datos³¹- a cierto tipo de distribución y, también, hipote-

³⁰Previamente, la persona que modeliza, ha considerado oportuno adoptar un modelo estratificado de capas geoeléctricas y admitir como plausibles todas las hipótesis que se precisan para la derivación del funcional **F**.

³¹Debido a la escasez de los mismos en las campañas normales, rara vez permite la estadística aceptar o descartar las hipótesis sobre la distribución que modela las medidas, *incluso a posteriori*.

tizar el valor de los parámetros de dispersión σ_1 y σ_2 , es decir, los pesos de ponderación de cada dato.

Si el problema se formula como:

$$\rho \equiv \mathbf{F}(\mathbf{m}, \mathbf{x}) = \mathbf{d} \equiv (10.2, 9.8) \tag{102}$$

no tiene solución.

Sí, considerando (y fijando) la información tipo 1 mencionada se decide plantear el problema como uno de minimización, resulta:

$$\rho_{\min} = \min_{\rho \in \mathbb{R}} \sum_{i=1}^{i=s} \left(\frac{d_i - \mathbf{F}(\mathbf{m}, \mathbf{x})}{\sigma_{i2}} \right)^2.$$
 (103)

Si, debido a las incertidumbres en σ_1 y σ_2 , calcular el mínimo en (103) fuese demasiado aventurado -si σ_1 y σ_2 no son los correctos se obtendría una $\rho_{\text{mín}}$ equivocada y se descartarían otras ρ que quizás fuesen más adecuadas- la prudencia aconsejaría, quizás, sustituir el problema (103) por el siguiente:

hallar
$$\rho: \sum_{i=1}^{i=s} \left(\frac{d_i - \rho}{\sigma_{i2}}\right)^2 \le T^2.$$
 (104)

La ecuación (104) tiene una interpretación geométrica. Si por simplicidad se toman sólo dos puntos de medida, para facilitar la comprensión, la expresión

$$\sum_{i=1}^{i=2} \left(\frac{d_i - \rho_i}{\sigma_{i2}} \right)^2 = T^2$$

es la ecuación de una elipse de semiejes $\sigma_1 T$ y $\sigma_2 T$ respectivamente, en el espacio $\{\rho_1, \rho_2\}$. Como el suelo es homogéneo y la resistividad es un parámetro positivo, puede interpretarse la solución de (104) como las intersecciones de dicha elipse con la bisectriz del primer cuadrante (en caso de igualdad) y como el intervalo comprendido entre los susodichos puntos en caso de desigualdad. Se observa en la figura ?? que este problema puede no tener solución, tener una o tener infinitas, dependiendo de cuáles sean los datos d, las desviaciones típicas σ_1 y σ_2 , y la tolerancia especificada T^2 .

Por lo tanto, la elección de la formulación del problema es, en sí misma una información "a priori" de bastante relevancia en la solución final. Esta tipología se designará, en este trabajo, como de $tipo\ 2$.

Otras formas de información "a priori" de tipo 2 incluyen, por ejemplo:

a) La elección de la norma. La influencia de este apartado puede apreciarse revisando lo mencionado en el apartado de "normas y espacio de datos" en cuanto al modo de considerar los datos anómalos, a la generación de no unicidad en las estimaciones (caso de la norma l_1) y a la robustez (disminución de la inestabilidad) en las estimaciones.

b) El problema de interés en S.E.V es no lineal. Su solución se aborda normalmente a través de un proceso iterativo. El camino recorrido por los resultados de cada iteración hasta el fin del proceso-que, se espera, produzca una solución-depende, por ejemplo, del tipo de algoritmo empleado, de la semilla inicial, de los criterios de convergencia estipulados, de las medidas algorítmicas tomadas para evitar inestabilidades excesivas (regularización), etc. Toda esta información debe ser suministrada por la persona que modela antes de haber interpretado los datos, siendo por tanto información a priori.

3. Información tentativa o de tipo 3

El objetivo último de la realización de inversión, en particular en S.E.V, suele ser la toma de decisiones. Cada decisión conlleva un cierto riesgo de equivocación, con el consiguiente coste. Por lo tanto, puede tener sentido pedir a un programa que ofrezca como solución, por ejemplo, un tipo de terreno extraordinariamente improbable desde el punto de vista de la experiencia anterior. No se trata de información basada en la experiencia o en datos medidos. Podría denominarse "información tentativa". Lo esencial es que, desde el punto de vista del problema, es a priori en el sentido de la definición. Se la denomina aquí, de tipo 3.

Estas tres tipologías de información "a priori" aparecerán frecuentemente en las argumentaciones posteriores.

4.6.1. Una formulación matemática para la información a priori

La información a priori de tipos 1 y 3, pueden incorporarse en la formulación del problema de la siguientes dos formas:

- 1. Estipulando que $\mathbf{m} \in \mathbf{S}$, donde \mathbf{S} es un determinado subconjunto del espacio de modelos.
- 2. Especificando información relativa al tamaño de ciertas características de los modelos aceptables. Por ejemplo, limitar los modelos válidos a aquellos cuya longitud sea menor que cierto umbral. Con este fin, según lo dicho al definir los conceptos de distancia y norma, se emplearán distintos ejemplos de normas sobre el espacio de modelos.

A continuación se eexplica más detalladamente la primera de las opciones. El análisis del segundo punto se postpone hasta que se hayan descrito las herramientas adecuadas (descomposición en valores singulares, etc).

Se recuerda que el subsuelo, en la aplicación de la técnica de los S.E.V, queda dividido en un número finito de capas geoeléctricas horizontales, isótropas y homogéneas. Cada una de las k capas, excepto la última, queda determinada por sus parámetros de resistividad eléctrica y espesor. La última capa, al suponerse infinito su espesor, queda determinada completamente por su resistividad eléctrica. Así pues, el subsuelo de un experimento de modelación en S.E.V queda perfectamente especificado mediante n=2k-1 parámetros, donde k es el número de capas geoeléctricas distintas.

Con vistas a la caracterización de la información a priori y, más tarde, al estudio de los modelos linealmente equivalentes a uno dado, es conveniente considerar esta n-upla de números reales desde una perspectiva doble. Por un lado, como elemento del espacio vectorial \mathbb{R}^n y, por

otra parte, como un punto del espacio afín \mathbb{M} asociado a dicho espacio vectorial. Así pues, todo modelo de subsuelo quedará representado como un punto del espacio afín \mathbb{M} asociado al espacio vectorial \mathbb{R}^n . Esta caracterización permitirá utilizar conceptos geométricos como las variedades lineales y las hipercuádricas.

A veces, el conocimiento que ya se tiene del subsuelo a partir de experiencias anteriores u otras mediciones -información tipo 1, pero también ocurre con la de tipo 3-, sugiere que ciertos parámetros satisfacen algunas restricciones en sus valores. Para establecerlas, en este estudio se han considerado relevantes para la práctica dos tipos de criterios:

- Fijar los valores de parámetros de ciertas capas geoeléctricas: este caso se presenta, por ejemplo, cuando dichos parámetros han sido determinados mediante experiencias de laboratorio a partir de testigos de sondeos en la zona, o por trabajos anteriores en afloramientos de las capas geoeléctricas aludidas.
 - En el proceso numérico pertinente, se permitirá que varíen nuevamente todos los parámetros excepto los de las capas citadas en este apartado, que permanecen inalteradas.
- Fijar las ratios entre dos parámetros distintos: caso que se da, por ejemplo, cuando a partir de estudios de geología de campo puede estimarse de manera aproximada la razón entre las potencias de dos capas de comportamiento geoeléctrico diferenciado. También podrían querer buscarse los modelos que mantengan entre dos de sus capas un determinado contraste de resistividades.

Caracterización matemática de las restricciones Se considera el subsuelo modelado por un punto \widehat{m} de un espacio afín M. Este punto forma parte de una variedad lineal de ecuación:

$$\widehat{m} = \widehat{m}_f + \mathbb{W},\tag{105}$$

siendo \widehat{m}_f un punto (fijo) del espacio de modelos M y W es un subespacio vectorial de \mathbb{R}^n (espacio de parámetros) cuya dimensión es p:

$$p = 2k - 1 - n_f - n_p = n - n_f - n_p, (106)$$

donde n_f es el número de parámetros fijados a priori y n_p es el número de restricciones de proporcionalidad. Esta formulación permite expresar *simultáneamente* los dos criterios anteriores de forma adecuada, es decir, reducir el número de grados de libertad del problema inverso *fijando* a la vez valores de parámetros del modelo y las razones entre los mismos.

Para precisar cómo se construyen tanto el punto fijo \widehat{m}_f como los vectores del subespacio \mathbb{W} es necesario especificar el tipo de parametrización por la que se ha optado, tanto en el espacio de modelos como en el espacio de datos: bien los valores de la resistividad y del espesor de las capas geoeléctricas, o bien sus logaritmos. A pesar de que en este trabajo se privilegia más la parametrización logarítmica, también se analiza cuál sería la estructura de \widehat{m}_f y de \mathbb{W} en el caso de que los parámetros fuesen los propios valores de resistividad y de espesor.

1. Caso I: coordenadas naturales

En este caso el punto fijo \widehat{m}_f contiene toda la información referente a los parámetros fijados, y sólo contiene esa información.

Si $\widehat{m}_f \in \mathbb{M}$ se expresase en la base canónica de \mathbb{R}^n sus coordenadas serían:

$$\mathbf{m}_f = (m_f^1, m_f^2, m_f^3, \cdots, m_f^n),$$
 (107)

donde:

- ullet m_f^i es nula si esta coordenada corresponde a un parámetro libre, y
- m_f^i es un valor predeterminado, si se trata de una coordenada que representa a un parámetro fijo.

Se busca ahora expresar un parámetro como combinación lineal de los parámetros restantes. Esta amplia formulación matemática permitirá acomodar fácilmente, como caso particular, la situación de proporcionalidad entre dos parámetros del mismo tipo (resistividades o espesores), que parece la única circunstancia físicamente plausible. Esta información queda incorporada en la estructura del subespacio vectorial W. Para tal fin, es necesario imponer condiciones a los vectores del espacio de parámetros, mediante la introducción de una aplicación lineal:

$$\mathcal{C}: \mathbb{R}^n \longrightarrow \mathbb{R}^n$$
,

cuya imagen es el subespacio $\mathbb{W} \subseteq \mathbb{R}^n$:

$$\mathbb{W} \equiv \operatorname{Im} \mathcal{C}$$
.

En el caso en que no existiesen restricciones de ningún tipo, $\mathbf{m}_f = \mathbf{0}$ y la aplicación \mathcal{C} sería la identidad, es decir, el problema conservaría todos sus grados de libertad ($\mathbb{W} \equiv \mathbb{R}^n$). En la práctica la aplicación \mathcal{C} se representa mediante su matriz $C \in \mathcal{M}(n,n)$, referida a la base canónica del espacio de parámetros. Así pues:

$$\mathbf{m}_w = C\mathbf{m}_l,\tag{108}$$

donde:

- $\mathbf{m}_w \in \mathbb{W}$ representa a un terreno que ha incorporado ya las relaciones de proporcionalidad, y
- $\mathbf{m}_l \in \mathbb{R}^n$ representa a un terreno sin ningún tipo de restricciones.

Por lo tanto $m_w^i = \mathbf{C}^i \mathbf{m}_l$, siendo \mathbf{C}^i la i-ésima fila de la matriz C. Esto permite expresar el parámetro m_w^i como combinación lineal de las coordenadas del vector \mathbf{m}_l . La estructura de la matriz C es como sigue:

- $\mathbf{C}^i = \mathbf{0}$, si el *i*-ésimo parámetro está fijado. Toda la información en este caso está en \mathbf{m}_f y \mathbf{m}_w es irrelevante.
- $\mathbf{C}^i = \mathbf{e}_i^t$ (vector *i*-ésimo de la base canónica de de \mathbb{R}^n), si el parámetro representado por la coordenada *i*-ésima puede variar libremente,

■ $\mathbf{C}^i = \begin{pmatrix} c_1^i, & c_2^i, & c_3^i, & \cdots, c_n^i \end{pmatrix}$ donde $c_j^i = \lambda_j^i \ (j = 1, ..., n)$, si $m_w^i = \lambda_j^i m_w^j$. Es decir, c_j^i representa el coeficiente de proporcionalidad entre los parámetros *i*-ésimo $\begin{pmatrix} m_w^i \end{pmatrix}$ y *j*-ésimo $\begin{pmatrix} m_w^j \end{pmatrix}$. Si no hubiese relación entre ambos parámetros, resultaría ser $c_j^i = 0$.

A modo de ejemplo, considérese un subsuelo de dos capas, representado por el vector:

$$\mathbf{m}_f = \left(\begin{array}{cc} \rho_1, & t_1, & \rho_2, \end{array} \right). \tag{109}$$

Se desearía, por ejemplo, comenzar una búsqueda de modelos equivalentes que dejasen inalterado el valor del espesor de la primera capa $-t_1$ - en 10 metros y que especificasen que la resistividad de la segunda capa y la de la primera no varíen de forma independiente, sino siguiendo una razón conocida, $\rho_2 = 100\rho_1$, o sea $\lambda_1^3 = 100$, permitiendo pues, en este caso, que ρ_1 varíe libremente.

El punto representativo del subsuelo buscado, \widehat{m} , tiene como vector asociado en \mathbb{R}^3 a **m** que, siguiendo las reglas enunciadas anteriormente, se expresa:

$$\mathbf{m} = \mathbf{m}_f + \mathbf{m}_w = \mathbf{m}_f + C\mathbf{m}_l = \begin{pmatrix} 0 \\ 10 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 100 & 0 & 0 \end{pmatrix} \begin{pmatrix} \rho_1 \\ t_1 \\ \rho_2 \end{pmatrix}.$$
(110)

Cabe observar que la dimensión del subespacio \mathbb{W} calculada de acuerdo con la expresión (106) coincide con el rango de la matriz C, como hemos anteriormente explicado.

El conjunto de modelos admisibles queda representado en la figura 8 mediante la línea azul. El punto de color verde es el punto fijo. Los segmentos rojos permiten comprobar que la pendiente es 100.

2. Caso II: coordenadas logarítmicas

También en este caso puede expresarse el punto representativo del modelo de un terreno siguiendo la estructura sugerida en (105). Pero, en este caso, las coordenadas del vector asociado al punto fijo, \mathbf{m}_f , difieren de las del caso anterior. Asimismo, tampoco pueden, en este caso, expresarse relaciones generales de combinación lineal entre los parámetros (no logarítmicos); sólo cabe expresar relaciones de multiplicidad entre dos parámetros (no logarítmicos) del mismo tipo, debido a las propiedades de la función logarítmica³².

A continuación se analiza el modo de obtener las coordenadas del vector asociado al punto fijo \widehat{m}_f y la estructura de los vectores pertenecientes al subespacio vectorial \mathbb{W} .

lacktriangle Construcción del punto fijo \widehat{m}_f

El punto fijo contiene en este caso información tanto de los parámetros cuyo valor está predeterminado, como relativa a los parámetros cuya ratio se estipula. Seguidamente se analiza en detalle esta circunstancia.

Las coordenadas de \mathbf{m}_f en la base canónica de \mathbb{R}^n se obtienen por suma de dos vectores de \mathbb{R}^n :

$$\mathbf{m}_f = \mathbf{m}_{f1} + \mathbf{m}_{f2}.$$

El primero, \mathbf{m}_{f1} , contiene información de los parámetros fijados a priori; es decir:

 $^{^{32}\}log\frac{a}{b} = \log a - \log b.$

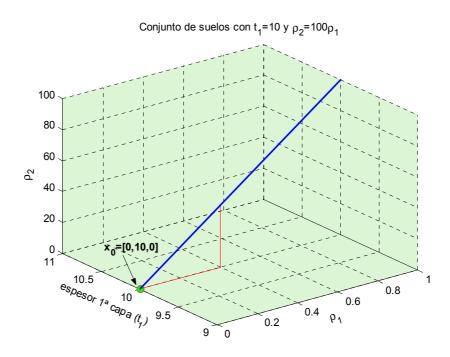


Figura 8: Expresión de la información apriori en coordenadas naturales.

- $m_{f1}^i=0$, si el parámetro i-ésimo es libre,
- $m_{f1}^i = \ln c$, si el parámetro *i*-ésimo se fija en el valor c.

El segundo vector, \mathbf{m}_{f2} , posee la siguiente estructura:

- $m_{f2}^i = 0$, si el *i*-ésimo parámetro se considera que varía libremente o, por contra, es un parámetro ya fijado en \mathbf{m}_{f1} .
- $m_{f2}^i = \ln \lambda_j^i$, si existe una relación de proporcionalidad entre los parámetros (no logarítmicos) m^i y m^j , tal que $m^i = \lambda_j^i m^j$, donde se asume que m^j varía libremente

\blacksquare Construcción del subespacio vectorial $\mathbb W$

La manera de operar es análoga al primer caso. No obstante, aquí la aplicación lineal C sólo informa de las parejas de parámetros que están relacionadas, pero no contiene ninguna información cuantitativa de dichas proporcionalidades, ya que esta información está presente en el vector \mathbf{m}_{f2} .

La estructura de la matriz C es como sigue:

- $\mathbf{C}^i = \mathbf{0}$, si el *i*-ésimo parámetro está fijado.
- $\mathbf{C}^i = \mathbf{e}^t_i$ si el parámetro representado por la coordenada *i*-ésima puede variar libremente. En este caso $m^i_{f2}=0$.

• $\mathbf{C}^i = \begin{pmatrix} c_1^i, & c_2^i, & c_3^i, & \cdots, c_n^i \end{pmatrix}$ donde $c_j^i = 1 \ (j = 1, \cdots, n)$ si $m_w^i = \lambda_j^i m_w^j$. Es decir, c_j^i vale la unidad si los parámetros *i*-ésimo $\begin{pmatrix} m_w^i \end{pmatrix}$ y *j*-ésimo $\begin{pmatrix} m_w^j \end{pmatrix}$ están relacionados. En caso contrario el valor de c_j^i es cero.

A modo de ejemplo se resuelve ahora el caso propuesto anteriormente:

$$t_1 = 10m,$$

$$\rho_3 = 100\rho_1 \Longleftrightarrow \log \rho_3 = \log 100 + \log \rho_1,$$
(111)

con lo cual:

$$\mathbf{m}_f = \mathbf{m}_{f1} + \mathbf{m}_{f2} = \begin{pmatrix} 0 \\ \log 10 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \log 100 \end{pmatrix}, \tag{112}$$

$$\mathbf{m}_{w} = C\mathbf{m}_{l} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \log \rho_{1} \\ \log t_{1} \\ \log \rho_{2} \end{pmatrix}. \tag{113}$$

La figura 9 representa gráficamente esta variedad lineal unidimensional en un espacio tridimensional donde los ejes son $\log \rho_1$, $\log t_1$ y $\log \rho_2$, respectivamente. Se puede ver cómo pasa por el punto $\mathbf{x}_0 = (0, \log 10, \log 100)$. Aunque no es fácil de apreciar en la vista ofrecida la recta tiene, en estos ejes, una pendiente unidad.

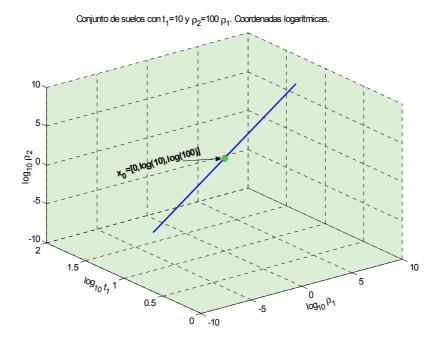
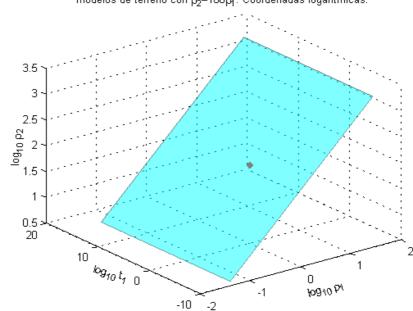


Figura 9: Expresión de la información a priori en coordenadas logarítmicas.

Se ha representado también, en la figura 10, el caso que considera los modelos para los cuales la única restricción es mantener la razón $\frac{\rho_2}{\rho_1}$ constante e igual a 100. Se dispone ahora de un grado más de libertad que en el caso anterior y, por lo tanto, los modelos admisibles se sitúan sobre un plano. El punto fijo resulta ser:



modelos de terreno con ρ₂=100ρ₁. Coordenadas logarítmicas.

Figura 10: Coordenadas logarítmicas. Restricción bidimensional.

$$\mathbf{m}_f = \mathbf{m}_{f1} + \mathbf{m}_{f2} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ \log 100 \end{pmatrix}.$$

y la dirección de la variedad lineal la indicada por:

$$\mathbf{m}_w = C\mathbf{m}_l = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \log \rho_1 \\ \log t_1 \\ \log \rho_2 \end{pmatrix},$$

donde se observa claramente que el número de columnas linealmente independientes es 2 y, por lo tanto, r(C) = 2. Se trata, pues, de un plano.

3. Caso III: Resistividades y/o espesores variando en un intervalo dado

Se puede utilizar la relación formal anterior, es decir:

$$\mathbf{m}^r = \mathbf{m}_f^r + \mathbf{m}_w^r = \mathbf{m}_f^r + C\mathbf{m}_l^r, \tag{114}$$

para obligar a las resistividades y/o espesores de algunas capas (o de todas) a permanecer dentro del rango comprendido entre los extremos de un intervalo del espacio de modelos. El superíndice r indica que el modelo no pertenecerá, en realidad, al espacio de modelos original \mathbb{M} sino a otro distinto \mathbb{M}^r obtenido mediante una reparametrización de \mathbb{M} .

La información a priori en este caso es del tipo:

$$\mathbf{m} \in [\mathbf{m}_{\min}, \mathbf{m}_{\max}], \mathbf{m} \in \mathbb{M},$$
 (115)

donde \mathbf{m}_{\min} representa el extremo inferior del intervalo *n*-dimensional y \mathbf{m}_{\max} representa el extremo superior del mismo. Las coordenadas de cualquier \mathbf{m} admisible cumplen:

$$m_i^{\min} \le m_i \le m_i^{\max}, i = 1, 2, \dots n.$$

Cuando el espacio de parámetros es cualquiera de los considerados en los dos casos anteriores (casos I y II) no parece posible representar información a priori del tipo indicado por (115) empleando la formulación sugerida en (114). La idea entonces consiste en obtener un nuevo espacio de parámetros \mathbb{M}^r distinto de \mathbb{M} , encontrando una función que transforme el intervalo $I_M = [\mathbf{m}_{\min}, \mathbf{m}_{\max}]$ en todo \mathbb{M}^r . De este modo, cualquier punto del nuevo espacio \mathbb{M}^r , sin restricciones, correspondería, de forma automática, a un terreno cuyo punto representativo en el espacio original \mathbb{M} posee resistividades y espesores que se mantienen dentro de los límites estipulados a priori por el intervalo I_M .

A modo de ejemplo, considérese un suelo de tres capas de resistividades $\rho_1 y \rho_2$, siendo t_1 el espesor de la primera capa. Debe ser posible considerar las siguientes restricciones:

$$\rho_1^{\min} \leq \rho_1 \leq \rho_1^{\max}, \ \rho_2^{\min} \leq \rho_2 \leq \rho_2^{\max}, t_1^{\min} \leq t_1 \leq t_1^{\max}. \tag{116}$$

Si el espacio de parámetros estuviese constituido por los valores de la resistividad y el espesor, el cumplimiento de las restricciones (116) impediría-por definición- el movimiento libre por todo el espacio. En un espacio de parámetros logarítmico, como el explicado en el caso II, el movimiento libre por el mismo sería imposible por idénticas razones. Basta ver que, las restricciones indicadas en (116) quedan transformadas, al tomar logaritmos, en:

$$\begin{array}{lcl} \log \rho_1^{\min} & \leq & \log \rho_1 \leq \log \rho_1^{\max}, \\ \log \rho_2^{\min} & \leq & \log \rho_2 \leq \log \rho_2^{\max}, \\ \log t_1^{\min} & \leq & \log t_1 \leq \log t_1^{\max}. \end{array}$$

En este trabajo se sugiere cómo definir la transformación del espacio de parámetros para llevar a cabo la reparametrización adecuada al objetivo mencionado.

La siguiente función logarítmica, ya utilizada por Mouiane y Pedersen (19??) en otro contexto, resulta adecuada para definir la reparametrización:

$$L: I_M \to \mathbb{R}^n: \mathbf{m}^L \equiv L(\mathbf{m}) = \ln \frac{\mathbf{m} - \mathbf{m}^{\min}}{\mathbf{m}^{\max} - \mathbf{m}} \Leftrightarrow m_i^L = \ln \frac{m_i - m_i^{\min}}{m_i^{\max} - m_i}, i = 1, 2, \dots,$$
(117)

donde:

$$I_M \subset \mathbb{R}^n = \left\{ \mathbf{m} : m_i^{\min} \le m_i \le m_i^{\max}, \ i = 1, 2, \dots n \right\}, \tag{118}$$

es la región paralelepipédica en el espacio de parámetros original. A continuación se caracteriza más detenidamente la acción de esta función.

Se parte del hecho de que tanto m_i^{\min} como m_i^{\max} son no negativos³³. Además, $m_i^{\min} > m_i^{\min}$. Por tanto, sólo mientras se verifique que $m_i^{\min} < m_i < m_i^{\min}$, el cociente $\frac{m_i - m_i^{\min}}{m_i^{\min} - m_i}$ será positivo y, por tanto, apto como argumento de la función logarítmica. En estas condiciones existirá m_i^L .

y, por tanto, apto como argumento de la función logarítmica. En estas condiciones existirá m_i^{ν} . Expuesto de otro modo, sólo tiene sentido esta transformación cuando los parámetros del suelo cumplen estrictamente las restricciones.

Por otro lado, si $m_i \longrightarrow m_i^{\min}$, entonces $m_i^L \longrightarrow -\infty$, mientras que, cuando $m_i \longrightarrow m_i^{\max}$, $m_i^L \longrightarrow +\infty$. Esto significa que, con m_i variando en el rango especificado por las restricciones, m_i^L recorre \mathbb{R} sin restricciones.

Es posible, además, introducir $m_i^{\min} = 0$, o sea, especificar explícitamente que la resistividad y el espesor han de ser estrictamente positivos. Sin embargo, no es posible eliminar la cota superior m_i^{\max} del intervalo. No obstante, en caso de no querer acotar superiormente los parámetros del terreno, bastaría introducir un número suficientemente elevado como valor de m_i^{\max} .

Por lo tanto, podría considerarse como nuevo espacio de parámetros la imagen de la aplicación L definida en (117). Los elementos de este espacio, denominados \mathbf{m}^L , pueden ser nuevamente transformados, de forma unívoca, a elementos del espacio natural mediante la aplicación inversa $L^{-1}: \mathbb{R}^n \to I_M$, donde:

$$\mathbf{m} = L^{-1} \left(\mathbf{m}^{L} \right) : m_{i} = \frac{m_{i}^{\text{máx}} \cdot e^{m_{i}^{L}} + m_{i}^{\text{mín}}}{1 + e^{m_{i}^{L}}}.$$
 (119)

La transformación L presenta dos inconvenientes:

- la mayor parte del espacio imagen de la transformación corresponde a valores del parametro muy próximos a los extremos del intervalo pero no a los propios extremos. Nunca podrá obtenerse, exactamente, un modelo en los extremos del intervalo. Cabe reseñar que es posible aproximarlos con precisión suficiente;
- la introducción de restricciones de proporcionalidad en el esquema propuesto por la expresión (105).

Supóngase, para fijar ideas, que se desea mantener el parámetro m_i proporcional a m_j con constante de proporcionalidad λ_i^j . Así pues: $m_i = \lambda_i^j \cdot m_j$. Si la proporcionalidad debe mantenerse, necesariamente que $m_i^{\min} = \lambda_i^j \cdot m_j^{\min}$ y también que $m_i^{\max} = \lambda_i^j \cdot m_j^{\max}$. O sea, no puede especificarse el rango de variación del parámetro dependiente m_i , sin tener en cuenta el del parametro independiente m_j .

En caso de encontrar un modo de solventar estas dificultades es preciso tener en cuenta que el requisito de que $\mathbf{m} \in I_M$ se puede querer establecer por razones independientes de las que motiven la restricción de proporcionalidad del tipo $m_i = \lambda_i^j \cdot m_j$. Cabe, por tanto, la posibilidad de incompatibilidades que hay que comprobar³⁴. Para ilustrar este punto considérese el siguiente

 $^{^{33}}$ físicamente no tendría sentido que lo fueran ya que representan a resistividades y espesores.

³⁴Es decir son dos informaciones a priori de procedencia independiente y, como reflejan circunstancias intuídas o, a lo sumo estimadas con cierto incertidumbre, pueden darse incompatibilidades.

ejemplo. En la figura 11, se trabaja sobre un terreno de dos capas en el cual el espesor se ha fijado en el valor $t_1 = 10$.

Las restricciones que se establecen sobre $\rho_1 y$ ρ_2 son las siguientes:

- $\bullet~1<\rho_1<10,~1<\rho_2<10.$ Región rectangular representada con borde negro en la figura.
- Se contemplan tres casos de proporcionalidad entre ρ_2 y ρ_1 :
 - $\rho_2 = 3 \rho_1$ (segmento azul).
 - $\rho_2 = \frac{1}{2} \rho_1$ (segmento verde oscuro).
 - $\rho_2 = 15 \ \rho_1$ (segmento magenta).

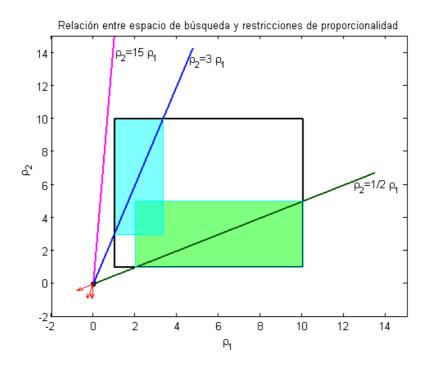


Figura 11: Relación entre intervalo de búsqueda y restricciones de proporcionalidad.

En el caso $\rho_2=3~\rho_1$, se observa cómo los puntos que forman la porción de segmento azul incluidos en el interior de la región rectangular negra, son modelos admisibles que cumplen los dos tipos de restricción al mismo tiempo. En realidad, para buscar estos modelos, en lugar de explorar el rectángulo negro grande, bastaría explorar el rectángulo reducido -incluido en el primero- representado con fondo azul claro. Ningún modelo fuera de este "rectángulo reducido" podría, nunca, ser candidato a cumplir la restricción de proporcionalidad. Se ha diseñado un programa, denominado "trans_intervalo.m" que localiza la región reducida compatible, en caso de que exista.

Otra nueva región "compatible" reducida, en este caso de fondo verde, se obtiene para la restricción $\rho_2 = \frac{1}{2} \rho_1$. También en este caso hay modelos que satisfacen simultáneamente los

dos modos de restricción. El caso $\rho_2 = 15\rho_1$, representado por los modelos de la recta color magenta, muestra claramente que ningún modelo puede satisfacer simultáneamente ambos tipos de restricción. No hay intersección entre la recta y el rectángulo negro.

Supóngase que se intenta emplear la parametrización indicada en (117). La relación, en el espacio reparametrizado, entre las imágenes $L(m_i)$ y $L(m_j)$ de los parámetros m_i y m_j que, en el espacio original M, guardan la proporción,

$$m_i = \lambda_i^j m_j,$$

es:

$$m_{i}^{L} = L\left(m_{i}\right) \overset{def}{=} \log \frac{m_{i} - m_{i}^{\min}}{m_{i}^{\max} - m_{i}} = \log \frac{\lambda_{i}^{j} m_{j} - \lambda_{i}^{j} m_{j}^{\min}}{\lambda_{i}^{j} m_{j}^{\min} - \lambda_{i}^{j} m_{j}} = L\left(m_{j}\right) = m_{j}^{L}.$$

Como se ve, las relaciones de proporcionalidad en el espacio de parámetros "natural"se convierten en identidades en el nuevo espacio, perdiéndose todo rastro de la constante de proporcionalidad.

Este inconveniente puede solventarse redefiniendo la función L, de forma que numerador y denominador no operen del mismo modo sobre el factor λ_i^j . En este trabajo, se propone la función:

$$L_{\alpha}(m_i) = \log \frac{\left(m_i - m_i^{\min}\right)^{\alpha}}{m_i^{\max} - m_i}.$$
 (120)

donde $0 < \alpha < 1$.

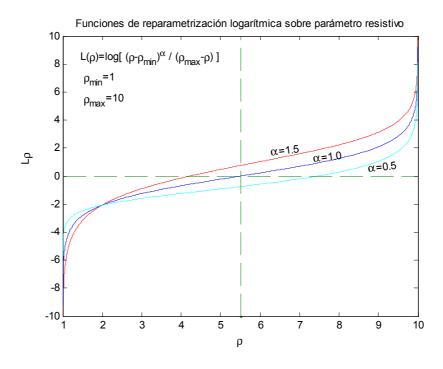


Figura 12: Parametrización logarítmica para diferentes valores de α .

Con la nueva función L_{α} no cambia el espacio imagen de la transformación³⁵, y tampoco lo hacen los comportamientos en los extremos. Como se verá a continuación, se estipulan sin inconvenientes las restricciones de igualdad y, en lo referente a las de proporcionalidad, no se pierde la constante λ_i^j puesto que, si $m_i = \lambda_i^j \cdot m_j$, se sigue que:

$$m_i^{L_{\alpha}} \stackrel{\text{def}}{=} L_{\alpha}(m_i) = \log \frac{\left(m_i - m_i^{\min}\right)^{\alpha}}{m_i^{\max} - m_i} = \log \frac{\left(\lambda_i^j m_j - \lambda_i^j m_j^{\min}\right)^{\alpha}}{\lambda_i^j m_j^{\max} - \lambda_i^j m_j} =$$

$$= (\alpha - 1) \log \lambda_i^j + \log \frac{\left(m_j - m_j^{\min}\right)^{\alpha}}{m_i^{\max} - m_j} = (\alpha - 1) \log \lambda_i^j + m_j^{L_{\alpha}}.$$

Lo interesante de esta expresión es que se trata de una relación afín, formulable mediante expresiones del tipo (114). Luego se verá en detalle esta posibilidad y se ilustrará mediante un ejemplo. El valor de α se puede elegir con diversas consideraciones. Una, importante y pragmática, puede ser facilitar el cálculo de la transformación inversa L_{α}^{-1} . Como $\alpha > 0$, la función $L_{\alpha}(m_i)$ es estrictamente creciente en el intervalo $[m_i^{\min}, m_i^{\max}]$, así que la ecuación

$$m_i^L = \log \frac{\left(m_i - m_i^{\min}\right)^{\alpha}}{m_i^{\max} - m_i},$$

donde m_i^L es conocido y m_i la incógnita, tendrá solución única en $[m_i^{\min}, m_i^{\max}]$. Para hallarla de la forma menos costosa, es conveniente tomar $\alpha = 0,5$. En este caso si se denomina $\beta = e^{m_i^L}$ resulta

$$\beta^2 = \frac{m_i - m_i^{\min}}{\left(m_i^{\max} - m_i\right)^2},$$

y m_i es solución de la siguiente ecuación de segundo grado:

$$\beta^2 m_i^2 - \left(1 + 2\beta^2 m_i^{\text{máx}}\right) m_i + \beta^2 \left(m_i^{\text{máx}}\right)^2 + m_i^{\text{mín}} = 0, \tag{121}$$

que tiene siempre dos soluciones, pero hay una sola en el intervalo $\left[m_i^{\min}, m_i^{\max}\right]$ -que es la correcta-. Por tanto, basta resolver la ecuación y tomar $la\ menor^{36}$ de sus dos raíces.

La información a priori que se puede estipular mediante esta reparametrización es la siguiente:

ullet Construcción del punto fijo 37 $\widehat{m}_f^{L_{lpha}}$

El punto fijo contiene información, al igual que en el caso de la reparametrización logarítmica simple, tanto de los parámetros cuyo valor está predeterminado como de aquellos

$$m_i = m_i^{\text{máx}} \pm \frac{\sqrt[2]{\Delta}}{\kappa^2},$$

donde se ha llamado Δ al discriminante de la ecuación. Evidentemente, la única solución que cumple $m_i < m_i^{\text{máx}}$ es la menor.

 $^{^{35}}$ Respecto al caso $\alpha = 1$.

 $^{^{36}}$ Que la raíz mayor queda descartada se comprueba fácilmente observando la expresión explícita de las soluciones:

 $^{^{37}}$ En lo sucesivo, para abreviar, se denotará $\mathbf{m}^{L_{\alpha}}$ como \mathbf{m}^{α} .

que guardan relación de proporcionalidad. Las coordenadas de \mathbf{m}_f^{α} en la base canónica de \mathbb{R}^n se siguen calculando sumando dos vectores de \mathbb{R}^n :

$$\mathbf{m}_f^{\alpha} = \mathbf{m}_{f1}^{\alpha} + \mathbf{m}_{f2}^{\alpha}.$$

El vector \mathbf{m}_{f1}^{α} , sigue almacenando la información de los parámetros fijados a priori, o sea:

- $m_{f1}^{\alpha,i} = 0$, si el parámetro *i*-ésimo es libre,
- $m_{f1}^{\alpha,i} = \log \frac{\left(c_i m_i^{\min}\right)^{\alpha}}{m_i^{\max} c_i}$, cuando el parámetro *i*-ésimo se fija en el valor c_i .

El vector \mathbf{m}_{f2}^{α} , tiene la siguiente estructura:

- $m_{f2}^{\alpha,i}=0$, si el *i*-ésimo parámetro se considera que varía libremente o, por contra, es un parámetro fijado previamente en \mathbf{m}_{f1}^{α} .
- $m_{f2}^{\alpha,i} = (\alpha 1) \log \lambda_j^i$, en caso en que los parámetros (no logarítmicos) m^i y m^j cumplan la relación $m^i = \lambda_j^i m^j$, variando m^j libremente.

• Construcción del subespacio vectorial \mathbb{W}_{α}

La aplicación lineal \mathcal{C} es idéntica a la del caso logarítmico normal ya mencionado antes. No obstante, se detalla aquí nuevamente:

- $\mathbf{C}^i = \mathbf{0}$, si el *i*-ésimo parámetro está fijado.
- $\mathbf{C}^i = \mathbf{e}_i^t$ si el parámetro representado por la coordenada i-ésima puede variar libremente. En este caso $m_{t2}^{\alpha,i} = 0$.
- $\mathbf{C}^i = \begin{pmatrix} c_1^i, & c_2^i, & c_3^i, & \cdots, & c_n^i \end{pmatrix}$ donde $c_j^i = 1 \ (j = 1, \cdots, n)$, si $m_w^i = \lambda_j^i m_w^j$. Es decir, c_j^i vale la unidad si los parámetros i-ésimo $\begin{pmatrix} m_w^i \end{pmatrix}$ y j-ésimo $\begin{pmatrix} m_w^j \end{pmatrix}$ están relacionados. En caso contrario el valor de c_j^i es cero.

Como ejemplo ilustrativo, se plantea la formulación de la misma información a priori del caso logarítmico anterior, es decir:

$$t_1 = 10m,$$

$$\rho_3 = 100\rho_1 \iff \log \rho_3 = \log 100 + \log \rho_1,$$
(122)

Pero, añadiendo además las siguientes restricciones:

$$1 < \rho_1 < 2, \ 1 < \rho_2 < 300,$$

Si se toma $\alpha = 0$. 5 entonces

$$\mathbf{m}_{f}^{\alpha} = \mathbf{m}_{f1}^{\alpha} + \mathbf{m}_{f2}^{\alpha} = \begin{pmatrix} 0 \\ \log \frac{(10-0)^{0.5}}{100-10} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ (0,5-1)\log 100 \end{pmatrix}, \tag{123}$$

$$\mathbf{m}_{w}^{\alpha} = C_{\alpha} \mathbf{m}_{l}^{\alpha} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} \log \rho_{1}^{\alpha} \\ \log t_{1}^{\alpha} \\ \log \rho_{2}^{\alpha} \end{pmatrix}. \tag{124}$$

En resumen, con la expresión (117) o la modificada (120) es posible imponer, en un proceso de optimización, por ejemplo, restricciones del tipo (118) sólamente, o bien-en el segundo caso (120), complementadas con la especificación ciertos parámetros y de relaciones de proporcionalidad entre otros.

En la figura 13 se observa la posición del mínimo de la función que cumple las restricciones en relación a la del mínimo global. Se ve cómo en el caso de $\alpha=0.5$ el método es capaz de hallar el punto que, además de cumplir las restricciones de intervalo, pertenece a la restricción unidimensional.

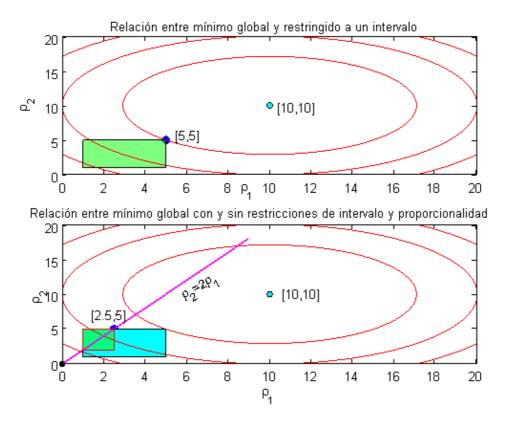


Figura 13: Diferencia entre un caso de optimización sin restricciones y otro con restricciones de intervalo y proporcionalidad.

La aparición de la nueva constante $\alpha=0$. 5 tampoco introduce demasiadas dificultades en el cálculo de la matriz jacobiana ya que según la regla de la cadena:

$$\frac{\partial T}{\partial m_i^L} = \frac{\partial T}{\partial m_i} \cdot \frac{\partial m_i}{\partial m_i^L}$$

 $\frac{\partial T}{\partial m_i}$ son derivadas ya calculadas para la matriz jacobiana relativa a la parametrización natural y

$$\frac{\partial m_i}{\partial m_i^L} = \frac{1}{\left(\frac{0.5}{m_i - m_i^{\min}} + \frac{1}{m_i^{\max} - m_i}\right)}.$$

4.7. Clasificación motivada de los sistemas lineales

En los epígrafes anteriores se ha justificado la necesidad de tratar con operadores lineales. Para resolver un problema inverso, aunque sea no lineal, es frecuente (aunque no estrictamente necesario) tener que resolver iterativamente varios problemas inversos lineales³⁸. En principio, cada uno de ellos podría ser planteado de la forma:

$$\mathbf{J}_F \cdot \mathbf{x} = \mathbf{b},\tag{125}$$

donde \mathbf{x} representa bien a \mathbf{m} o bien $\Delta \mathbf{m}$ y $\mathbf{J}_F \in \mathcal{M}(s, n)$.

Los sistemas lineales pueden ser clasificados en función del cardinal del conjunto de soluciones que admiten, en:

- Compatible y determinado: admite una y sólo una solución.
- Compatible e indeterminado: admite soluciones, pero en número infinito.
- Incompatible: No admite solución.

Hay, tradicionalmente, cierta tendencia a establecer relación unívoca entre cada uno de estos grupos y entre el número de ecuaciones -s- y el número de incógnitas -n-. En términos generales no hay relación unívoca entre la compatibilidad o incompatibilidad de un sistema y el número de ecuaciones e incógnitas. No obstante, se señala que:

• Si s = n, en caso de que ninguna ecuación pueda expresarse como combinación lineal de las restantes y que el cumplimiento de ninguna de ellas sea *incompatible* con el cumplimiento de alguna de las demás, el sistema es, también, *compatible y determinado*.

Si s = n, no obstante, la satisfacción de alguna de las s ecuaciones puede impedir el cumplimiento de alguna otra. En este caso, el sistema es *incompatible* y *no hay solución* del problema (125).

Cabe también considerar la situación donde no habiendo incompatibilidad, alguna de las ecuaciones es redundante. Por ejemplo, una de ellas es proporcional a otra. Las dos contendrían la misma información y bastaría considerar sólo una de ellas en cuyo caso, realmente, habría menos ecuaciones que incógnitas. El sistema (125) podría satisfacerse de forma exacta por infinidad de soluciones. Sería, por tanto, compatible e indeterminado.

Puede también darse el caso mixto.

³⁸Obtenidos por linealización del problema no lineal. Alternativas a esta estrategia la ofrecen los métodos de optimización global como simulated annealing o los algoritmos genéticos.

• Si s < n, es decir, hay menos ecuaciones que incógnitas y no hay incompatibilidades ni ecuaciones redundantes en el sentido aclarado en el párrafo anterior, hay infinitas soluciones y el sistema aludido es compatible e indeterminado.

No obstante, alguna de las ecuaciones puede ser incompatible con las demás. Entonces no habría solución posible y se tendría el caso *incompatible* nuevamente.

No hay forma de que un sistema así sea compatible y determinado.

Si s > n, el número de ecuaciones es mayor que el de incógnitas. Si alguna de las ecuaciones es incompatible con alguna de las restantes s - 1, el sistema es incompatible. Esta es, en verdad, la situacion más problable en la realidad. No obstante, podría ocurrir que todas las s-n ecuaciones en exceso fuesen redundantes, es decir, no aportasen información adicional. Por tanto podrían suprimirse y el sistema sería equivalente a uno de igual número de ecuaciones que de incógnitas y estaría sujeto a la discusión precedente.

Cabe notar la posibilidad de que el número de ecuaciones redundantes sea mayor que s-n, en cuyo caso habría infinitas soluciones -siempre en ausencia de incompatibilidad-.

Trás estos razonamientos se observa que la clasificación atendiendo al número de soluciones es un tanto rígida en el sentido de que o bien no hay ninguna solución, o hay una única o hay infinitas. Pero no es posible, en principio, que haya un número finito (mayor que uno) de soluciones. Si bien cuando se habla de la resolución numérica del sistema (125) la situación, en términos matemáticos, no cambia, hay que incluir un matiz muy importante. El miembro derecho -el vector de datos b- procede, en general, de mediciones físicas sujetas a error. Por tanto, hay incertidumbre en el valor de sus componentes. Para un valor determinado de dicha incertidumbre³⁹ habría infinitos vectores de datos posibles y, por tanto, las soluciones física y numéricamente buscadas provendrían de la solución, exacta, de infinitos problemas del tipo (125). Es esperable que ellas mismas sean, por tanto, infinitas⁴⁰.

Un aspecto relevante es que, como se supone que la variación del vector de datos se debe al azar, todos los datos (vectores) obtenidos son representantes legítimos de la medida. Por ello, las distintas soluciones exactas obtenidas son, todas, soluciones igualmente legítimas. Si, en lugar de hacer hincapié en las diferencias entre ellas, se piensa que responden a la misma situación física subyacente, cabría pensar en el conjunto de soluciones como la solución del problema. En este sentido, la solución ya no tendría carácter puntual -un punto en el espacio de modelos- sino que quedaría caracterizada por una determinada región de dicho espacio.

$$\begin{array}{rcl} x + y & = & 2 \\ x + y & = & 2 \end{array}$$

tiene infinitas soluciones. Sin embargo, si se perturba ligeramente el vector de datos, el sistema resultante:

$$x + y = 2,$$

$$x + y = 2,001,$$

es incompatible, o sea, no aporta ninguna nueva solución.

³⁹ Por ejemplo la varianza de su modelo estadístico.

⁴⁰ Nótese que no es obligado que, por el hecho de resolver infinitos sistemas de ecuaciones, se obtengan infinitas soluciones distintas. Por ejemplo, el sistema:

4.7.1. Sistema determinado

Se puede ejemplificar esta situación utilizando el caso, ya presentado anteriormente de un terreno con dos capas de resistividades respectivas (10,1) y cuya primera capa tiene un espesor de 5. Las medidas de resistividad aparente se adquieren en tres puntos de la superficie cuyos semiespaciados interelectródicos $(\frac{C_1C_2}{2})$ quedan indicados en el vector $\mathbf{s} = (1.1, 10, 10.5)$. Las resistividades aparentes medidas son $\boldsymbol{\rho}_m = (9.98, 5.1708, 4.7967)$. Estos datos han sido obtenidos simulando un proceso de medida con errores gaussianos de media cero y error cuadrático medio⁴¹-relativo-del uno por mil. La matriz J_F en el modelo citado es:

$$\mathbf{J}_F = \begin{bmatrix} 0.9977 & 0.0114 & 0.0036 \\ 0.4330 & 1.2973 & 0.8270 \\ 0.3967 & 1.2942 & 0.8736 \end{bmatrix} . \tag{126}$$

Se ha simulado la repetición del experimento de medida 1500 veces. Los experimentos simulados están afectados por errores que, se asume, siguen una distribución gaussiana de media 0 y desviación típica $\sigma = 0.0019^{42}$. En cada ocasión se ha resuelto el problema:

$$\mathbf{F}\left(\mathbf{m}\right) = \mathbf{b},\tag{127}$$

mediante linealización de \mathbf{F} , operador no lineal del problema directo, en un entorno de $\mathbf{m}_0 = (10, 5, 1)$. Esto conduce a la formulación:

$$\mathbf{J}_F \mathbf{x} = \mathbf{b},\tag{128}$$

donde se ha escogido como incógnita el modelo, $\mathbf{x} = \mathbf{m}$. El vector de datos \mathbf{b} tiene por expresión, por tanto:

$$\mathbf{b} = \boldsymbol{\rho}_m - \mathbf{F}(\mathbf{m}_0) + \mathbf{J}_F \cdot \mathbf{m}_0. \tag{129}$$

La figura 14 representa los datos medidos -procedentes de la primera de las experiencias afectadas por error- representados sobre la curva de resistividades aparentes simulada sin error.

Posteriormente, la resolución del sistema (128) mediante eliminación gaussiana, permite la obtención de 1500 modelos que conforman una muestra del conjunto solución. La figura 15 ofrece distintas vistas de la nube de modelos resultante. Un punto rojo en el centro de la nube, representa la solución exacta del sistema no perturbado. El resto de los puntos quedan representados en color azul o verde.

El uso de estos dos colores lo motiva la parametrización elegida. No se emplea la parametrización logarítmica en el espacio de modelos. Se trabaja sobre los propios parámetros-resistividad y espesor-. Ahora bien, para mostrar íntegramente la forma de la nube de puntos resultante no se han incluido las restricciones de positividad que requiere el vector x. Por ello, como soluciones se obtienen tanto modelos aceptables del terreno -todos los parámetros positivos- coloreados de azul como modelos inaceptables de terreno -con algún parámetro negativo- coloreados de verde. Los cortes realizados paralelamente a los planos coordenados sobre la nube de puntos, facilitan la apreciación de la forma de la nube. Se observa que es un elipsoide con una gran excentricidad

⁴¹Véase apéndice (3)

⁴²Correspondiente a un error cuadrático medio de 0.001.

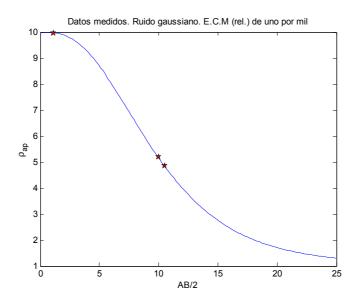


Figura 14: Datos con ECM del $0.1\,\%$

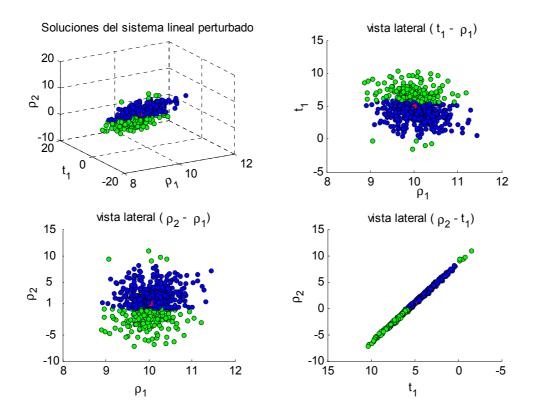


Figura 15: Nube de modelos equivalentes con ECM=0.1%.

-uno de los ejes es mucho más grande que los otros dos-. La presencia de errores en los datos parece provocar una dispersión mucho mayor en unas direcciones que en otras. Se destaca que pequeños errores en los datos provocan grandes variaciones respecto a la solución obtenida sin error en los datos. Se concluye, por lo tanto, que:

- existe gran inestabilidad en la determinación de las soluciones exactas del sistema;
- la inestabilidad no es de la misma magnitud en distintas direcciones del espacio de modelos;
- esta situación es intrínseca al proceso de inversión cuando se emplean medidas reales -con error-y métodos numéricos.

Esta inestabilidad se produce por el hecho de tratar de resolver el sistema con datos erróneos. De hecho, se han pretendido reproducir exactamente estos datos erróneos. La exigencia de perseguir los datos con exactitud -apartándose de aquellos libres de error- ha supuesto grandes desviaciones en las soluciones correspondientes. Parece intuitivo que si no se intentase resolver exactamente un sistema con segundo miembro erróneo, las inestabilidades en las soluciones podrían ser menores. Aparte, no tiene sentido reproducir exactamente un dato del que se ya sabe que es falso. Por todo ello, la formulación

$$\mathbf{J}_F \mathbf{x} = \mathbf{b},\tag{130}$$

debe quedar sustituido por otra más coherente

$$\mathbf{J}_F \mathbf{x} \simeq \mathbf{b},\tag{131}$$

donde se aclarará más adelante cómo interpretar estrictamente el símbolo \simeq .

En la figura 16 puede verse otra razón en apoyo de la formulación (131) y en detrimento de la (130). En la parte superior se representa el número de modelos en función de su *error no lineal*. Se puede observar que, aunque todos ellos resuelven exactamente el sistema lineal, ninguno resuelve exactamente el problema no lineal, lo cual después de todo, es lo que se pretende.

En la parte inferior de la citada figura se adjuntan dos histogramas que representan a los modelos (sobre 1500) agrupados en función del valor de la norma o de la seminorma. La norma es la euclídea del vector de parámetros solución y la seminorma es la norma del vector $\mathbf{s} = \mathbf{L}\mathbf{x}$. Como matriz \mathbf{L} se ha tomado, en este caso la siguiente:

$$L = [1, 0, -1],$$

que valora la diferencia entre las resistividades de la primera y de la segunda capa.

Estos histogramas podrían ser utilizados para investigar la estructura interna del conjunto de soluciones, ofreciendo la posibilidad de, por ejemplo, cuantificar la probabilidad de obtener una solución cuya norma esté entre 5 y 13.

Se ha procedido a repetir la experiencia con un ECM de un 1 %. En este caso se observa que la dispersión de las soluciones ha aumentado considerablemente, dado que la nube se extiende de modo que muchas de las soluciones son no válidas. En la figuras 17, 18 y 19 se ilustran estos hechos.

Si, en lugar de considerar tres puntos de medida, se considerasen sólo los dos últimos emplazamientos de los electrodos, lo cual sería equivalente a conservar solamente las dos últimas

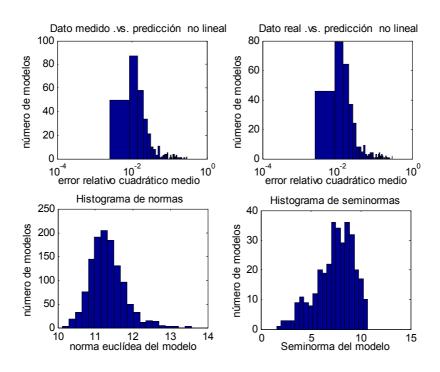


Figura 16: Diferentes histogramas de la nube de modelos equivalentes con ECM=0.1%

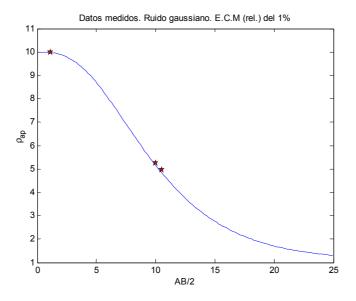


Figura 17: Datos con ECM del $1\,\%$

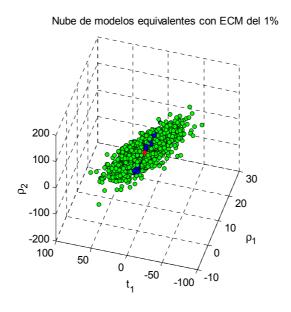


Figura 18: Nube de modelos equivalentes con ECM del $1\,\%$

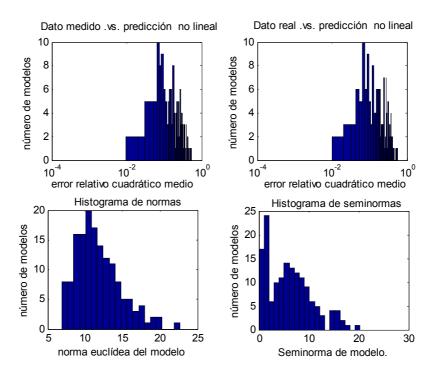


Figura 19: Diferentes histogramas de la nube de modelos equivalentes con ECM=1%

filas de \mathbf{J}_F y se optase por mantener invariable el valor del espesor de la primera capa, $t_1 = 5$, la nueva matriz \mathbf{J}_F' es:

$$\mathbf{J}_F' = \left[\begin{array}{cc} 0.4330 & 0.8270 \\ 0.3967 & 0.8736 \end{array} \right],$$

y las componentes del nuevo vector de datos se calculan según la expresión:

$$b'_{i} = b_{i} - \mathbf{J}_{F}(i, 2) t_{1}, \ i = 1, 2,$$
 (132)

con lo cual

$$\mathbf{b'} = \left[\begin{array}{c} 5 \cdot 1668 \\ 4 \cdot 8471 \end{array} \right].$$

El nuevo sistema restringido⁴³ tendría la siguiente expresión:

$$\mathbf{J}_F'\mathbf{m}=\mathbf{b}',$$

que, explícitamente, se escribe:

$$\begin{bmatrix} 0.4330 & 0.8270 \\ 0.3967 & 0.8736 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 5.1668 \\ 4.8471 \end{bmatrix}.$$
 (133)

El número de condición de \mathbf{J}_F' es 35.68, es decir, bastante alto, lo cual explica la inestabilidad del sistema. En la figura 20 se representan los dos primeros puntos medidos y su situación. En la figura 21, se representa la nube de soluciones.

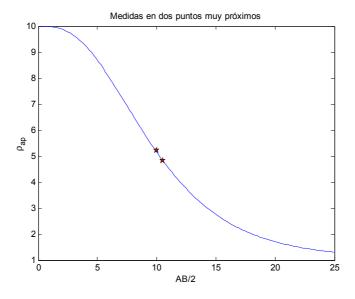


Figura 20: Datos reducidos con error ECM=0,1%.

 $^{^{43}}$ Se denomina restringido puesto que se mantiene fijo el valor del espesor de la primera capa. Las variables son, por lo tanto ρ_1 y ρ_2 .

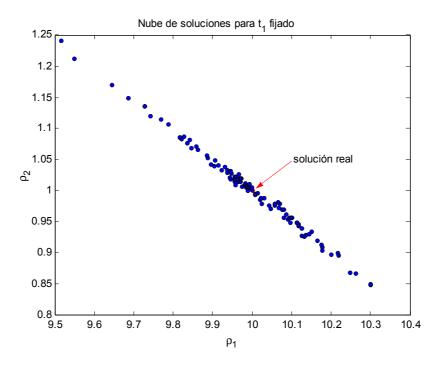


Figura 21: Nube de modelos equivalentes con ECM=0,1%.

Sobre 100 realizaciones diferentes se aprecia la gran dispersión existente respecto a la solución del problema no perturbado -indicada con punto rojo-. Asimismo aparece claramente ilustrado el hecho de que la dispersión no es la misma en todas direcciones, sino claramente mayor en la dirección, aproximada, de la diagonal del segundo y cuarto cuadrante. Este hecho guarda relación con el denominado rango numérico de la matriz (véase el apéndice relativo al capítulo lineal, donde se analiza en el contexto geométrico de intersección de dos rectas) y se entiende mejor utilizando la descomposición en valores singulares que se presentará en breve.

4.7.2. Sistema subdeterminado

Se analizará ahora el comportamiento de las soluciones en el caso en que los puntos de medida sean sólo dos, situados a corta distancia uno de otro. El terreno sobre el que se lleva a cabo la experiencia tiene, como antes, dos capas caracterizadas por idéntico vector de parámetros, a saber: $\mathbf{m}_0 = (1.1, 10, 10.5)$. Si las medidas no hubiesen sido afectadas por error alguno, las resistividades aparentes quedarían reflejadas en el vector:

$$\mathbf{b}_v = (5.22, 4.84)$$
.

Sin embargo, para simular lo que ocurriría en una situación real, donde existirían diferencias aleatorias entre las distintas repeticiones de la misma medida, se ha supuesto que el vector medida está afectado por un ruido aleatorio gaussiano de media nula y de desviación típica $\sigma = 0.0125$, correspondiente a un error cuadrático medio en el vector de medidas del 1 %. En

total, el número de experiencias simuladas ha sido de 200⁴⁴.

Observando atentamente la matriz jacobiana⁴⁵

$$\mathbf{J}_F = \begin{bmatrix} 0.4330 & 1.2973 & 0.8270 \\ 0.3967 & 1.2942 & 0.8736 \end{bmatrix}, \tag{134}$$

del problema representado mediante el sistema:

$$\mathbf{J}_F \mathbf{m} = \mathbf{b},\tag{135}$$

puede apreciarse que:

- los coeficientes de las dos filas son distintos y no proporcionales -es decir, una fila no es proporcional a la otra. Esto significa que las dos ecuaciones en (135), son diferentes. Por tanto, el rango de J_F es dos;
- el número de incógnitas es tres. Uno de los parámetros, pues, puede variar libremente sin influir en el resultado;
- a pesar de que los coeficientes sean distintos, puede decirse que casi son iguales. Entendida cada una de las ecuaciones en (135) como la ecuación de un plano en el el espacio tridimensional, ambos son casi paralelos. Como, además, el vector de datos $\bf b$ contiene dos medidas muy próximas en superficie, habrá, en general, poca diferencia entre los valores de sus dos coordenadas. Por eso, los planos serán casi coincidentes. Si lo fuesen completamente, su intersección sería un plano. Como no lo son completamente su intersección será, estrictamente hablando, siempre una recta -con un grado de libertad, coherente con el hecho de que un parámetro varíe libremente-. Sin embargo, dado que los planos están muy próximos, una pequeña incertidumbre en su situación relativa puede generar gran confusión en la posición real de la recta de intersección. La situación de ésta podrá barrer zonas relativamente grandes de un plano, aproximándose a lo que ocurre en el caso de planos coincidentes, situación que ocurre, recuérdese, cuando las dos filas de $\bf J_F$ -ampliadas con $\bf b$ son idénticas, siendo el rango de $\bf J_F$ igual a uno.

Es decir que, "numéricamente hablando" la situación es, en cierto modo, intermedia entre el rango dos y el rango uno.

La figura 22, sitúa los dos puntos de medida. Obsérvese la cercanía de los mismos lo cual motiva el comentario del epígrafe precedente.

La figura 23 representa las soluciones del sistema (135) cuando el segundo miembro ha sido perturbado de la forma comentada.

La recta azul indica la dirección del núcleo de J_F . Si un punto fuese capaz de solventar (135) exactamente -por ejemplo el suelo real \mathbf{m}_0 representado mediante un punto rojo- cualquier otro

⁴⁴ El hecho de que, en diferentes ejemplos, se simulen distinto número de repeticiones (en un caso 1500, en éste 200, etc) se debe a que se ha debido condicionar la cantidad de modelos representados a la consecución de la nitidez adecuada en la representación de la figura correspondiente. Es decir, tomando como ejemplo esta figura, la representación de 1500 modelos 'velaba' considerablemente la parte central de la nube de soluciones.

⁴⁵ préstese atención al hecho de que, dado que los puntos de medida actuales son los dos últimos puntos de la terna del ejercicio anterior, las filas de la nueva matriz en [134] son idénticas a las dos últimas filas de la matriz jacobiana en [126]

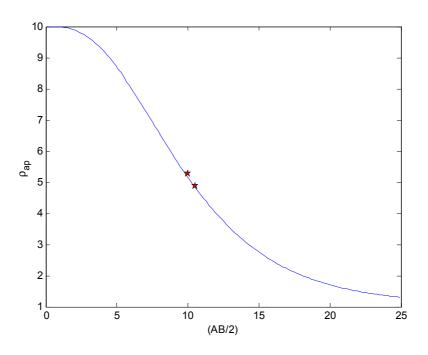


Figura 22: Caso en que las medidas están tomadas en dos puntos cercanos

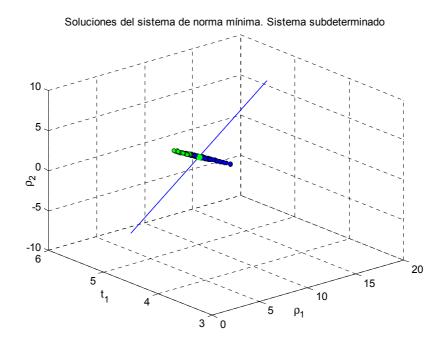


Figura 23: Soluciones del sistema de norma mínima

punto obtenido a partir de éste desplazándose en la dirección indicada por la recta azul, sería una solución capaz también de predecir **exactamente** los datos medidos. Sobre esta ilustración, parece importante comentar que:

■ El hecho de que el sistema sea compatible y subdeterminado se concreta visualmente en la existencia del conjunto de modelos de terreno representados por la dirección de la recta azul. Nótese que dicha recta no es paralela a ninguno de los ejes coordenados. Por lo tanto, el paso de un punto de ella -por ejemplo \mathbf{m}_0 - a otro cualquiera implica la modificación simultánea de los valores de \mathbf{todos} los parámetros. Es decir, introduciendo ciertos cambios $a\ la\ vez$ en los valores de las resistividades y espesores de todas las capas se llega a una predicción idéntica.

El uso matizado de la expresión "ciertos cambios" se debe a que, como se ve en la figura, para obtener a partir de un modelo solución otra solución equivalente la modificación debe tener lugar en la dirección de la recta y no de forma arbitraria. Por tanto, los cambios en los parámetros deben ser proporcionales a los cosenos directores de la recta en cuestión. O sea:

$$\Delta \rho_1 = \lambda \cos \alpha_1 = 0.6799\lambda,
\Delta t_1 = \lambda \cos \alpha_2 = -0.5419\lambda,
\Delta \rho_2 = \lambda \cos \alpha_3 = 0.4940\lambda.$$
(136)

Podría resumirse lo expuesto indicando que lo que influye en la predicción de los datos en un sistema subdeterminado no es la modificación libre de uno de los parámetros, sino la modificación de cierta combinación lineal de los incrementos⁴⁶ de los parámetros. En este ejemplo se ha supuesto que se dispone de una herramienta que permita, a partir de la formulación del problema (135) determinar qué combinaciones de parámetros (en este caso lineales) afectan y cuáles son irrelevantes en el cálculo del error de predicción de los datos.

• Una vez obtenida una solución -que en este caso reduce el desajuste a cero porque se trata de un sistema compatible- se ha visto que pueden modificarse sus parámetros de la forma indicada en (136) y obtener un terreno equivalente. Esta posibilidad tiene un aspecto negativo y también una lectura positiva. Para ilustrar el lado negativo imagínese obtenida -por algún procedimiento- la solución verdadera \mathbf{m}_0 . Se insiste en que se conoce que es verdadera, por ejemplo porque se tiene acceso a un corte geológico en un talud y se ha visto que así es en la realidad. Si la persona que esté resolviendo el sistema desconociese este hecho, podría estar tentada de escoger, en lugar de la solución \mathbf{m}_0 cualquiera otra de sus equivalentes hallada cambiando los parámetros de \mathbf{m}_0 de la forma indicada en

$$\mathbf{v} = (0.6799, -0.5419, 0.4940)$$

Por lo tanto, puede escribirse que, si \mathbf{m}_1 es una solución exacta del sistema:

$$\mathbf{m} = \mathbf{m}_1 + \lambda \mathbf{v}$$

también lo es.

 $^{^{46}}$ El núcleo de \mathbf{J}_F es, en este caso, el vector:

(136). No obstante correcta desde el punto de vista matemático, esta nueva solución ha introducido **características ficticias** que falsean la geología, en este ejemplo. La nueva solución incorpora lo que se denominan *artefactos matemáticos* sin correlato geológico real.

Dado que en un problema real no se conoce el terreno verdadero \mathbf{m}_0 -pues es precisamente el conocimiento que se persigue- no es fácil reconocer cuándo una solución matemáticamente correcta contiene artefactos. Esta dificultad se presenta, siempre, cuando al menos una parte del problema está subdeterminada en el sentido de que infinitas soluciones de "igual calidad" son posibles. Se ve, pues, la importancia que cobra disponer de una herramienta que permita discernir qué parte de un problema, si es que hay alguna, está subdeterminada. La descomposición en valores singulares responde a esta necesidad.

El aspecto positivo de esta situación es el siguiente: imagínese que el terreno real sobre el que se realiza la experiencia \mathbf{no} es \mathbf{m}_0 aunque este hecho se desconoce. Piénsese que se ha resuelto el problema (135) y que se ha obtenido como solución, **precisamente** \mathbf{m}_0 . En la misma zona se han llevado a cabo algunos sondeos casi a nivel de superficie y, tras ensayos de laboratorio sobre varias muestras, se ha concluido que la resistividad superficial no es inferior a 11 $\Omega \cdot m$. Justamente el hecho de que el problema esté subdeterminado permite calcular, a partir de la solución matemática obtenida \mathbf{m}_0 -inconsistente con la experiencia citada- una nueva solución compatible con los resultados de laboratorio. La nueva solución \mathbf{m}_{sol} cumple:

$$\mathbf{m}_{sol} = \mathbf{m}_0 + \lambda \mathbf{v} \Rightarrow \begin{bmatrix} 2 \\ t_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 10 + 0.6799\lambda \\ 5 - 0.5419\lambda \\ 1 + 0.4940\lambda \end{bmatrix}. \tag{137}$$

De la primera ecuación en (137) se obtiene $\lambda = 1.4708$. Con este valor se calculan t_1 y ρ_2 . Por tanto,

$$\mathbf{m}_{sol} = (20, 4.2, 1.73)$$
.

En consecuencia, en un problema subdeterminado pueden utilizarse los grados de libertad existentes para satisfacer la información a priori. Sobre la marcha se menciona que lo interesante es emplear esta información para conseguir reducir la incertidumbre en la parte subdeterminada del problema y que esta información "no afecte" a la otra parte del problema, que es la encargada de reproducir -en este caso exactamente- el valor de las medidas.

El efecto de la información a priori modificando la parte del problema perfectamente determinada por los datos -puede servir de ejemplo el caso compatible, determinado y con medidas libres de error, ilustrado antes- produciría una desviación o sesgo en la solución que se traduciría en una diferente capacidad para predecir las medidas. En el ejemplo compatible y determinado del primer caso, cuando la solución se determina sólo a partir de los datos, el error de predicción es nulo (es decir, el sistema se resuelve exactamente). Si se modificase alguno de los parámetros de esta solución en el sentido de satisfacer mejor alguna información a priori, la predicción del sistema sería distinta y, por tanto, siendo las medidas las mismas, el desacuerdo con ellas sería mayor que antes -donde era nulo-.

■ Una vez identificada la dirección v en la que, como se ha visto, pueden aparecer artefactos, la figura 23 muestra lo que ocurre cuando se pretende resolver el sistema (135) con la

condición de que -si los datos estuviesen libres de error- no apareciesen artefactos. O sea, que si las medidas no están afectadas por perturbaciones aleatorias se recupere el suelo verdadero \mathbf{m}_0 .

Lo sorprendente es lo que ocurre cuando el sistema se resuelve con medidas afectadas por errores aleatorios. Se observa que una pequeña perturbación en los datos -el error cuadrático medio de la perturbación aleatoria es del 1%- produce una gran modificación en las soluciones halladas. En la figura 23, el parámetro ρ_2 puede variar en más de un 400% y el parámetro ρ_1 hasta un 40%. Si "el dato" es interpretado como la realización de una variable aleatoria (debido a la incertidumbre que pesa sobre su valor), "la solución" ya no es puntual, sino una región (una nube de puntos).

La "nube" está en un plano perpendicular a la recta azul. Las soluciones halladas no tienen componentes en la dirección de dicha recta. La "incertidumbre" introducida por el ruido aleatorio en el vector de datos ha sido transmitida a la solución lo suficiente como para que la "nube" solución tenga un tamaño considerable. Sin embargo, el tamaño de la "nube" -que en realidad es un elipsoide- es mucho mayor en una dirección que en las otras dos. La disparidad es tal que la "nube" casi se percibe visualmente como un segmento perpendicular a la recta azul. Esto tiene su explicación.

Como ya se había sugerido, el extraordinario parecido de las dos filas de la matriz jacobiana, proveniente de la cercanía de los puntos de medida, es un hecho relevante. Mientras no sean iguales (rango 2), la intersección de los dos planos que las representan geométricamente es una recta que refleja la existencia de un grado de libertad. Perpendicularmente a dicha recta -la recta azul- el tamaño de la región solución es nulo.

En el caso en que las dos ecuaciones sean coincidentes (rango 1), las soluciones se sitúan en un plano. Éste queda definido por la propia recta azul más otra recta, perpendicular a ella, que contiene las componentes de las soluciones en la dirección perpendicular a la anterior.

Al no ser iguales, pero extraordinariamente parecidas, la incertidumbre en el dato produce un efecto extraño. La dirección azul, que no depende del dato sigue siendo la misma. Pero en la parte de la solución completamente determinada por el dato -y perpendicular a la recta azul- ésta se difumina convirtiéndose en una especie de segmento. Se trata de una situación intermedia entre las dos anteriores: la "rectilínea" (rango 2) y la "plana" (rango 1) en la cual el segmento aludido se "extendería" hasta alcanzar la condición de recta.

En esta circunstancia, las soluciones estarían comprendidas en una banda plana de tamaño infinito en la dirección indicada por la recta azul y limitada en anchura en la dirección del "segmento".

El hecho de que una pequeña perturbación en los datos pueda generar semejante desviación de la solución del problema (??) respecto a la solución verdadera es algo no deseable. Se ha visto que la información a priori puede ser utilizada para fijar las partes de la solución **no afectadas** por los datos -esto es, para decidir su situación sobre la recta azul-. Ahora bien: ¿podría utilizarse, también, la información a priori para reducir la gran variabilidad inducida en las soluciones por nimios cambios en las medidas?. Se verá que esto sí es posible. El precio que hay que pagar es que, al ser la solución que incorpora la información a priori distinta de la dictada exclusivamente por los datos, cambia la discrepancia con éstos

-aumentando-. ¿Es mejor una solución alejada de los conocimientos geológicos previos y que cumpla estrictamente con los datos medidos o una más aproximada a dicha información pero menos acorde con los datos medidos?. Esta pregunta es una de las razones por las que puede hablarse del "arte de modelizar".

■ Finalmente en el caso de un sistema puramente subdeterminado⁴⁷ cabe utilizar un tipo de información a priori numérica, como es la búsqueda de la solución de norma euclídea mínima (o solución natural):

$$\|\mathbf{m}_{NM}\|_2 = \min_{\mathbf{J}_F \mathbf{m} - \mathbf{b} = \mathbf{0}} \|\mathbf{b}\|_2. \tag{138}$$

Este es un problema de optimización con restricciones y puede ser resuelto mediante la técnica de multiplicadores de Lagrange. Una manera más directa de abordar el problema es mediante técnicas del álgebra lineal. Hemos visto que las soluciones del sistema subdeterminado pertenecen a la variedad lineal

$$\mathbf{m} = \mathbf{m}^* + \ker \mathbf{J}_F.$$

donde \mathbf{m}^* es cualquier solución particular del sistema lineal compatible subdimensionado. La solución de norma mínima geométricamente corresponde al punto de dicha variedad lineal que además se encuentra en el subespacio ortogonal a su subespacio director $[\ker \mathbf{J}_F = \operatorname{Im} \mathbf{J}_F^t]$:

$$\mathbf{m}_{NM} \in (\ker \mathbf{J}_F)^{\perp} \Leftrightarrow (\ker \mathbf{J}_F)^{\perp} = \operatorname{Im} \mathbf{J}_F^t,$$

por lo tanto:

$$\mathbf{m}_{NM} = \mathbf{J}_F^t \mathbf{y},\tag{139}$$

donde $\mathbf{J}_F^t \in \mathbb{M}(n,s)$ e $\mathbf{y} \in \mathbb{R}^s$. Como además se tiene:

$$\mathbf{J}_F \mathbf{m}_{NM} = \mathbf{b}$$
,

se debe verificar

$$\mathbf{J}_F \mathbf{J}_F^t \mathbf{y} = \mathbf{b},\tag{140}$$

sistema que recibe el apelativo de **ecuaciones normales del problema de norma mínima**. Si el problema es **puramente subdimensionado** entonces se verifica

$$rg\mathbf{J}_F\mathbf{J}_F^t = rg\mathbf{J}_F = s,$$

con lo cual la matriz $\mathbf{J}_F \mathbf{J}_F^t$ es regular y se tiene:

$$\mathbf{y} = (\mathbf{J}_F \mathbf{J}_F^t)^{-1} \mathbf{b}$$

$$\mathbf{m}_{NM} = \mathbf{J}_F^t (\mathbf{J}_F \mathbf{J}_F^t)^{-1} \mathbf{b},$$
(141)

es decir, existe una única solución de norma mínima.

 $[\]overline{^{47}\text{El}}$ rango de la matriz del sistema coincide con el número de datos (ecuaciones).

4.7.3. Sistema sobredimensionado o sobredeterminado

Una nueva situación se produce cuando el número de ecuaciones que condicionan los parámetros del terreno -coindicente con el número de medidas realizadas- es mayor que el número de ecuaciones⁴⁸. Para ejemplificar este caso, se presenta un experimento simulado en el cual se han llevado a cabo medidas en tres puntos diferentes de la superficie, bien distanciados entre sí, con semiespaciados electródicos dados por el vector:

$$\mathbf{s} = (1, 8, 30)$$

sobre el mismo terreno de dos capas de los ejemplos anteriores, $\mathbf{m}_0 = (10, 5, 1)$. A fin de reducir a dos el número de parámetros desconocidos, se supondrá conocido y fijo el espesor de la primera capa geoeléctrica $t_1 = 5$. Si las medidas realizadas estuviesen libres de error, las resistividades aparentes serían:

$$\rho_{ap} = (9.989, 6.633, 1.154).$$

Con el fin de simular lo más aproximadamente posible la realidad, han sido obtenidos 300 vectores de medidas sintéticas afectadas por un ruido gaussiano de media nula y desviación típica $\sigma = 0.094$, que permite obtener datos con un error cuadrático medio (relativo) del 5%. Uno cualquiera de estos tripletes es, por ejemplo:

$$\rho_m^{5\%} = (10.15, 6.97, 1.097),$$

y queda representado en la figura (lfig17datos1) junto con la curva de resistividades aparentes sin error. Asumiendo, como se ha dicho que $t_1 = 5$ siempre, el problema que se plantea es:

$$\mathbf{J}_F'\mathbf{m} = \mathbf{b} \tag{142}$$

donde

$$\mathbf{b} = \boldsymbol{\rho}_{m}^{5\%} - \mathbf{F}(\mathbf{m}_{0}) + \mathbf{J}_{F} \cdot \mathbf{m}_{0} - t_{1} \cdot \mathbf{J}_{F}(:, 2), \qquad (143)$$

siendo, $\mathbf{J}_F(:,2)$ la segunda columna de la matriz jacobiana \mathbf{J}_F original. El sistema (142) a resolver sería, pues:

$$\begin{bmatrix} 0.9983 & 0.0027 \\ 0.5952 & 0.6062 \\ 0.0049 & 1.1033 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \begin{bmatrix} 10.1455 \\ 6.8952 \\ 1.0953 \end{bmatrix}.$$
 (144)

La figura 24 presenta las tres rectas que traducen geométricamente cada una de las ecuaciones de dicho sistema en el plano (ρ_1, ρ_2) . Inmediatamente salta a la vista que las tres rectas no se cortan en un punto. Por lo tanto, **no existe solución del problema** (142). Es otra forma de decir que existen incompatibilidades entre las tres ecuaciones.

Sin embargo, recordando que el segundo miembro de (144) es una medida errónea y lo comentado en el caso del sistema determinado, lo que en realidad no existe es un suelo que prediga exactamente datos erróneos. Es decir, son los errores de los datos los que introducen incompatibilidad en las ecuaciones⁴⁹. No parece, nuevamente, razonable exigir el ajuste perfecto y cobra nuevamente sentido reformular el problema (142) como:

$$\mathbf{J}_F \mathbf{m} \simeq \mathbf{b}.$$
 (145)

⁴⁸Naturalmente se supone que todas las ecuaciones aportan información no contenida ya en las restantes, o sea, que no hay ninguna ecuación redundante.

⁴⁹En realidad, si los datos fuesen exactos -esto es libres de error- las tres rectas se cortarían en el mismo punto.

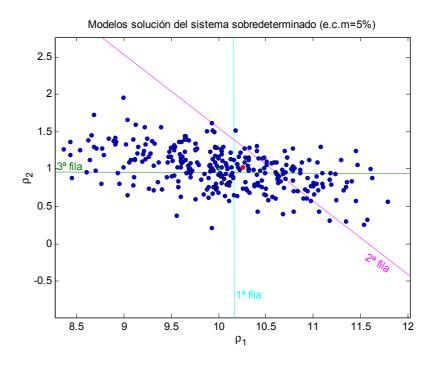


Figura 24: Modelos equivalentes en el caso sobredeterminado con error del 5 %.

iQué sentido debe darse al signo \simeq ?. Una vez obtenidas las medidas, cabe plantearse qué terreno permite unas predicciones más próximas a las mismas. La idea de proximidad exige la utilización del concepto de distancia y, por lo tanto, el uso de las normas. Así pues, el problema (145) podría ser sustituido por

$$\mathbf{m}_{sol} = \min_{\mathbf{m} \in \mathbb{R}^2} \left\| \widetilde{\mathbf{J}}_F \mathbf{m} - \mathbf{b} \right\|_p^p, \ p = 1, 2, 3, \dots \infty.$$
 (146)

Los criterios para elegir la norma p se han comentado en el apartado correspondiente. Se harán los razonamientos en el caso p=2. En esta situación, \mathbf{m}_{sol} cumple las denominadas ecuaciones normales del problema de mínimos cuadrados:

$$\left(\widetilde{\mathbf{J}}_F^t \widetilde{\mathbf{J}}_F\right) \mathbf{m}_{sol} = \widetilde{\mathbf{J}}_F^t \mathbf{b}. \tag{147}$$

En el caso anteriormente analizado, la solución del sistema se ha representado en la figura 24 como un punto rojo que, como se observa, no pertenece a ninguna de las tres rectas. Esto significa que dicho punto no cumple ninguna de las tres ecuaciones exactamente. Sin embargo, de alguna manera, es el que mejor predice las tres medidas.

¿Cuál es la solución del sistema cuando se perturba el dato b?. Cada uno de los puntos azules representa la solución de las ecuaciones normales en el caso de un sistema perturbado con el ruido gaussiano antes mencionado⁵⁰. Se vuelve a encontrar un fenómeno conocido: las perturbaciones en los datos producen dispersión de los puntos solución. En este caso en particular, como el

⁵⁰Se ha resuelto el sistema [147] mediante el método de eliminación gaussiana. Posteriormente se ha resuelto el problema de minimización utilizando una descomposición QR, teóricamente más estable. No hay diferencias en

número de condición de $\mathbf{J}_F^t \mathbf{J}_F$ es 1,716 es relativamente próximo a la unidad, la nube de puntos no es muy grande. No obstante, las desviaciones de algunos parámetros respecto a la solución puede ser del 30 %.

Es decir, también en el caso de minimización de un problema sobredeterminado la solución, siendo única, es inestable. ¿Podría reducirse esta inestabilidad?. A esta pregunta se responderá afirmativamente, utilizando la información a priori en un proceso denominado regularización. No obstante, la nueva solución obtenida producirá un desajuste con los datos medidos que ya no será mínimo.

La figura ?? representa el mismo caso cuando las dos últimas medidas se toman en puntos mucho más próximos. Puede observarse como las rectas representativas de dichas ecuaciones se "parecen" mucho más entre sí que en el caso anterior. Esto motiva que la nube de soluciones cambie de orientación. Como consecuencia del parecido entre las dos últimas filas, el número de condición de $\widetilde{\mathbf{J}}_F$ ha aumentado a 1,71 mientras que el de $\widetilde{\mathbf{J}}_F^t\widetilde{\mathbf{J}}_F$ vale ahora 2.95. La nube de puntos es ligeramente mayor en este caso.

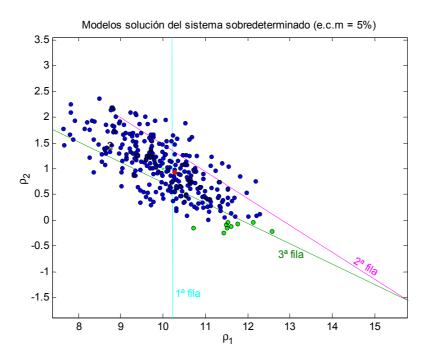


Figura 25: Soluciones equivalentes con ecm=5% en el caso sobredeterminado con ecuaciones "casi" redundantes.

El planteamiento del sistema (145) como problema de minimización (146) realizado en el caso de un sistema sobredeterminado puede emplearse para uno determinado. En este caso, la solución de mínimos cuadrados podrá reducir la discrepancia con los datos a cero. Las tres rectas

los resultados en este caso. Podría deberse a que la diferencia entre el número de condición de la matriz jacobiana (que es de 1,31) y el de la matriz $\mathbf{J}_F^t \mathbf{J}_F$ -de 1,71-, es muy reducida y los sistemas bastante estables -por ser su número de condición próximo a uno.

se cortan en un sólo punto que es la solución tanto del sistema de ecuaciones como del problema de minimización.

En el caso subdeterminado el problema tiene dos partes. Una, no condicionada por las medidas -son las componentes de la solución en el núcleo de \mathbf{J}_F - que es a la que se debe el nombre de subdeterminado. Esta parte de la solución puede variar libremente. La otra parte del problema transmite a la solución la información de los datos. Dada una solución -en el sentido de mínimos cuadrados- de esta segunda parte, cualquier otro modelo obtenido sumándole la componente del núcleo produce, exactamente el mismo desajuste.

Aquí, debe introducirse la información a priori para elegir una de ellas. Un tipo de demanda, muy frecuente, consiste en solicitar la solución que tenga la norma mínima. Esto, físicamente, equivale a interesarse por el terreno que tiene las resistividades más bajas en todas las capas y que, además, tiene las capas más delgadas y el sustrato más cerca de la superficie. Para que la norma sea mínima, bastaría que las componentes de la solución en el núcleo tomasen el valor nulo.

Interpretación geométrica del sistema de mínimos cuadrados

La figura 26 muestra la interpretación geométrica del sistema de mínimos cuadrados de un sistema lineal $\mathbf{J_F} \cdot \mathbf{x} = \mathbf{d}$. En realidad, el método consiste en determinar la proyección (ortogonal) \mathbf{p} del vector de datos \mathbf{d} sobre el subespacio $\mathrm{Im} \mathbf{JF}$ y en resolver el sistema lineal $\mathbf{J_F} \cdot \mathbf{x}_{ls} = \mathbf{p}$. Por lo tanto por génesis, las ecuaciones normales del sistema de mínimos cuadrados siempre son un sistema compatible, dicho de otro modo, dado que $\mathbf{J_F^t} \mathbf{b} \in \mathrm{Im} \mathbf{J_F^t} \mathbf{J_F}$ sea cual sea el vector de datos \mathbf{d} , es decir en general dicho sistema admite infinitas soluciones. Si además la matriz del sistema, $\mathbf{J_F^t}$, es completa en rango, entonces la solución de mínimos cuadrados es única y el sistema original recibe el apelativo de $\mathbf{puramente}$ sobredeterminado.

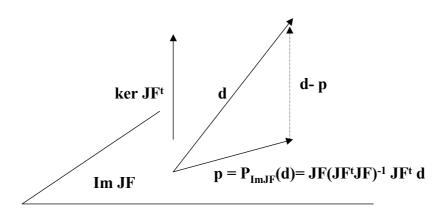


Figura 26: Interpretación geométrica de la solución de mínimos cuadrados

4.8. La descomposición en valores singulares

En los problemas anteriores se ha tenido la oportunidad de visualizar cómo pequeñas variaciones en los datos de un problema lineal -planteado como sistema o, de forma más amplia,

como problema de minimización- suponen la introducción de cambios relativamente grandes en la solución del sistema. El conjunto de soluciones sigue, a pesar de su aleatoriedad, un patrón característico. En las figuras anteriores se ha comprobado que la región en la que se hallan estos modelos parece un elipsoide cuyos ejes no son todos iguales. La orientación y tamaño parecen, además, tener que ver con lo cerca que estén las filas del jacobiano de ser linealmente dependientes. Geométricamente, la asimetría del elipsoide, así como el tamaño de su eje mayor, crece cuando las rectas que representan algunas de las filas del jacobiano tienden a parecerse. Resultaría interesante disponer de una herramienta que permitiese explicar, de manera sistemática, por qué esta región tiene forma de elipsoide, con qué está relacionada la orientación de los ejes del mismo y qué afecta a su tamaño. Sería recomendable que tal herramienta sugiriese claves de cómo reducir la variabilidad observada y cuáles son los cambios que se producen cuando esto ocurre.

Por otro lado, en un problema real, es posible que haya una parte del modelo que esté subdeterminada, es decir, que no se vea afectada por las medidas tomadas. Convendría poder identificarla de manera sistemática. Se ha visto cómo puede atribuirse valor a esta parte indeterminada utilizando la información a priori. Y, aun más, ayudaría, dentro de la parte del modelo que sí esté determinada -o sobredeterminada- por los datos, poder identificar componentes diferentes del modelo en función del grado en que están determinadas (o influyen en) las medidas.

El uso de la información a priori se ha venido sugiriendo que pudiese, también, ser utilizado para reducir la inestabilidad indeseable que surge cuando las medidas tienen error (o sea, siempre en la práctica). Esta herramienta buscada permitiría analizar cómo esto es posible y qué es lo que cambia en cada caso.

El objeto matemático que permite llevar a cabo estas tareas se denomina **descomposición** en valores singulares de la matriz **JF**. A continuación se presentan la filosofía y los detalles del mismo:

Filosofía de la descomposición

Considérese el problema lineal presentado en la forma -más general- de problema de optimización:

$$\min_{\mathbf{x}} \|\mathbf{J}\mathbf{F}\mathbf{x} - \mathbf{d}\|_{2}, \tag{148}$$

donde **x** representa bien **m** o bien $\Delta \mathbf{m}$. Supóngase que, para fijar ideas, $\mathbf{JF} \in M(s,n)$.

JF es la matriz de una aplicación lineal $\mathcal{J}: \mathbb{R}^n \to \mathbb{R}^m$, del espacio de modelos \mathbb{R}^n en el espacio de datos \mathbb{R}^s , referida a la base canónica de ambos espacios. La descomposición en valores singulares de J_F proporciona, en primer lugar, dos bases *ortonormadas* de vectores: una, $\{\mathbf{v}_i\}, i=1,2,\cdots n$, en el espacio de modelos y otra $\{\mathbf{u}_i\}, i=1,2,\cdots s$ en el de datos con una serie de características de gran utilidad para los propósitos enunciados en el párrafo anterior.

Los vectores \mathbf{u}_i hasta el *n*-ésimo están relacionados con su correspondiente \mathbf{v}_i , mediante un escalar σ_i . La terna $(\mathbf{u}_i, \sigma_i, \mathbf{v}_i)$ se caracteriza por las relaciones:

$$\mathbf{JFv}_i = \sigma_i \mathbf{u}_i, \tag{149}$$

у

$$\mathbf{JF}\mathbf{u}_i = \sigma_i \mathbf{v}_i.$$

Esto significa que la matriz **JF** se vuelve diagonal⁵¹ en estas dos bases. De la imagen por **JF** de la componente de un modelo (o incremento de modelo) cualquiera \mathbf{x} , en la dirección indicada por \mathbf{v}_i , puede decirse -analizando la expresión (149)- que:

- ullet está en el subespacio del espacio de datos generado por la ${f u}_i$ correspondiente.
- Es la propia componente del modelo sobre \mathbf{v}_i amplificada por el factor σ_i , véase la expresión (151).

A la luz de la relación(149), la descomposición en valores singulares podría decirse que desagrega o compartimenta los espacios de modelos y de datos en subespacios mutuamente perpendiculares, numerados mediante un subíndice i. La información relativa al modelo contenida en el subespacio i-ésimo (generado por \mathbf{v}_i), influye y es influida única y exclusivamente por el contenido informativo que las medidas poseen en el subespacio i-ésimo (generado por \mathbf{u}_i) del espacio de datos.

De la misma forma, supóngase que un vector de medidas \mathbf{d} pueda ser escrito como suma de la medida libre de perturbaciones \mathbf{d}_v y de un vector de errores \mathbf{e} . Si dicho error afectase sólo (o, sobre todo) a las componentes que el vector de datos tiene en los subespacios generados por algunas \mathbf{u}_i , dicho error afectaría sólo (resp. fundamentalmente) a las componentes que la solución posee sobre las \mathbf{v}_i correspondientes del espacio de modelos. Es decir, podría conocerse en qué direcciones -caracterizadas por los vectores \mathbf{v}_i - del espacio de modelos, la transmisión del error ocurrido en las componentes \mathbf{u}_i de los datos produce más difuminación o menos.

Los escalares σ_i , $i = 1, 2, \dots, n$, denominados valores singulares, son no negativos -pueden ser cero- y se presentan siempre en orden decreciente con el valor de i.

Los vectores \mathbf{u}_i y \mathbf{v}_i se denominan vectores singulares a la izquierda y a la derecha, respectivamente.

Dado que el sistema $\{\mathbf{v}_i\}$, $i=1,2,\cdots n$ es una base del espacio de modelos, cualquier modelo (o incremento) \mathbf{x} puede expresarse en dicha base como:

$$\mathbf{x} = \sum_{i=1}^{i=n} (\mathbf{x} \cdot \mathbf{v}_i) \, \mathbf{v}_i. \tag{150}$$

Utilizando la propiedad (149) la imagen por \mathbf{JF} de este modelo \mathbf{x} es:

$$\mathbf{JFx} = \sum_{i=1}^{i=n} (\mathbf{x} \cdot \mathbf{v}_i) \, \mathbf{JFv}_i = \sum_{i=1}^{i=n} \sigma_i (\mathbf{x} \cdot \mathbf{v}_i) \, \mathbf{u}_i.$$
 (151)

Según distintos autores (Hansen, 1998, Hanke, 2002), p.ej.- los vectores \mathbf{u}_i y \mathbf{v}_i tienen, en general, la propiedad de hacerse más oscilantes a medida que aumenta el índice i. Por lo tanto, expresar un modelo cualquiera \mathbf{x} o un dato \mathbf{d} , en sus bases respectivas, $\{\mathbf{v}_i\}$ y $\{\mathbf{u}_i\}$, pudiera ser considerado como una descomposición en componentes de distintas frecuencias. O sea, las componentes de \mathbf{x} (resp. \mathbf{d}) en la dirección de un \mathbf{v}_i (resp. \mathbf{u}_i) de índice i bajo describirían las partes estructurales del

 $^{^{51}}$ **JF** es, en general, una matriz rectangular de dimensiones (s, n). No puede, por tanto, estrictamente hablando, diagonalizarse. No obstante se utiliza esta palabra para aludir al hecho de que se consigue expresar esta matriz en función de otra, $\Sigma_{(s,n)}$, cuyos elementos de subíndice $i \neq j$ son siempre nulos.

modelo (resp. del vector de datos), es decir, serían las componentes de baja frecuencia. Asimismo, las componentes de \mathbf{x} (resp. \mathbf{d}) en la dirección de un \mathbf{v}_i (resp. \mathbf{u}_i) de índice i alto describirían los detalles del modelo (resp. del vector de medidas). Éstas últimas serían las componentes de alta frecuencia.

En los problemas relacionados con el método de S.E.V los valores singulares se hacen pequeños rápidamente y de manera contínua con i. Esto significa, según la expresión (151) y lo recientemente dicho que las componentes de alta frecuencia -las componentes más de detalledel modelo son amortiguadas o suavizadas, ya que su imagen se obtiene multiplicándolas por valores σ_i pequeños. Visto a la inversa, se obtiene la confirmación de un fenómeno conocido a partir de las gráficas de los ejemplos anteriores, a saber: Las componentes de alta frecuencia de las medidas se transmiten a las partes correspondientes de la solución amplificándose. Es decir, los detalles introducidos en los datos por el azar en las medidas se corresponden amplias variaciones en las partes correspondientes de la solución.

En un análisis posterior se confirmará el papel amortiguador o amplificador de las σ_i . No obstante, para la formulación tradicional en S.E.V se refutará el creciente comportamiento oscilante atribuido a las \mathbf{u}_i y \mathbf{v}_i a medida que aumenta i.

La expresión matricial de la descomposición en valores singulares (d.v.s) de la matriz JF es:

$$J = U\Sigma V^{t} = \begin{bmatrix} \mathbf{u}_{1}, \mathbf{u}_{2}, \cdots \mathbf{u}_{m} \end{bmatrix} \begin{bmatrix} \Sigma_{(r,r)} & 0_{(r,n-r)} \\ 0_{(m-r,r)} & 0_{(m-r,n-r)} \end{bmatrix} \begin{bmatrix} \mathbf{v}_{1}^{t} \\ \mathbf{v}_{2}^{t} \\ \vdots \\ \mathbf{v}_{n}^{t} \end{bmatrix},$$
(152)

donde la matriz $\Sigma_{(r,r)}$ es una matriz diagonal formada por los valores singulares no nulos en orden decreciente. Es decir:

$$\Sigma_{(r,r)} = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_r \end{bmatrix}, \ \sigma_1 \ge \sigma_2 \ge \dots \ge \sigma_r.$$
 (153)

Los términos restantes de la diagonal de la matriz Σ son nulos.

En (152) quedan relacionados los subespacios $N(\mathbf{JF}) = \{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \cdots \mathbf{v}_n\}$, núcleo de \mathbf{JF} , y $\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \cdots \mathbf{u}_m\}$. Como la descomposición en valores singulares dada por la expresión (152), relaciona univocamente y de forma exclusiva cada subespacio unidimensional generado por $\{\mathbf{v}_i\}$, $i = 1, 2, \cdots n$, con otro subespacio unidimensional del espacio de datos $\{\mathbf{u}_i\}$, $i = 1, 2, \cdots s$, a través de un factor de proporcionalidad σ_i , resulta que pueden suprimirse del problema las partes de los espacios de datos y de parámetros vinculadas a través de valores singulares nulos, sin que nada se vea afectado.

Supóngase que el problema a resolver sea sobredeterminado $(s \ge n)$. Si, en la matriz **JF**, el número de columnas (o filas) linealmente dependientes es menor que n, el rango de **JF**, $r(\mathbf{JF})$, cumple estrictamente que:

$$r(J) = r \le n$$
.

En este caso se pueden suprimir los subespacios generados por:

1. a) $\{\mathbf{u}_{n+1}, \mathbf{u}_{r+2}, \cdots \mathbf{u}_m\}$, porque ningún modelo tiene nada que ver con estos datos al no estar en la imagen de **JF**.

2. $\{\mathbf{u}_{r+1}, \mathbf{u}_{r+2}, \cdots \mathbf{u}_n\}$, porque, al corresponder a valores singulares nulos, no pertenecen tampoco a la imagen de **JF**.

3. $\{\mathbf{v}_{r+1}, \mathbf{v}_{r+2}, \cdots \mathbf{v}_n\}$, forman el núcleo del operador **JF** y, por tanto, no influyen en los datos

Así pues, cuando el rango del operador \mathbf{JF} es r y la matriz es deficiente en rango, basta sólo considerar los valores singulares no nulos y sus correspondientes subespacios para representar a la matriz \mathbf{JF} . Por tanto,

$$\mathbf{JF} = \mathbf{JF}_r = \sum_{i=1}^{i=r(\mathbf{JF})} \sigma_1 \mathbf{u}_i \mathbf{v}_i^t.$$

Si se considera, por tanto, la solución al problema sobredeterminado (o determinado),

$$\min_{\mathbf{x}} \|\mathbf{J}\mathbf{F}\mathbf{x} - \mathbf{d}\|_2$$
.

La descomposición en valores singulares simplifica notablemente los cálculos algebraicos en las ecuaciones normales correspondientes, obteniéndose al resolverlas que:

$$\mathbf{x}_{mc} = \mathbf{V} \left[\left(\mathbf{\Sigma}^t \mathbf{\Sigma} \right)^{-1} \mathbf{\Sigma}^t \right] \mathbf{U}^t \mathbf{d}$$
 (154)

La matriz $\mathbf{A}^{\dagger} \equiv \mathbf{V} \left[\left(\mathbf{\Sigma}^t \mathbf{\Sigma} \right)^{-1} \mathbf{\Sigma}^t \right] \mathbf{U}^t$ se denomina **matriz pseudoinversa**. Por tanto,

$$\mathbf{x}_{mc} = \mathbf{A}^{\dagger} \mathbf{d} \tag{155}$$

En el caso subdeterminado o mixto, la solución de *norma mínima* del problema (148) conduce a la misma fórmula que, desarrollada, es:

$$\mathbf{x}_{mc} = \sum_{i=1}^{i=r} rac{\mathbf{u}_i^t \mathbf{d}}{\sigma_i} \mathbf{v}_i.$$

4.9. La estructura geométrica de las inestabilidades

Las expresiones obtenidas sirven para aclarar la estructura que siguen las inestabilidades del problema de mínimos cuadrados y, por tanto, el patrón geométrico de la nube de soluciones.

Considérese primero el hecho de que, dado que la solución \mathbf{x}_{mc} se relaciona linealmente con el vector de datos \mathbf{d} mediante la expresión (155) y éste está afectado de error estadístico, la propia \mathbf{x}_{mc} puede considerarse como una variable aleatoria. Al vector de datos hipotéticamente libre de error, \mathbf{d}_v , le corresponde la solución "verdadera":

$$\mathbf{x}_v = \mathbf{A}^{\dagger} \mathbf{d}_v. \tag{156}$$

Una medida de cómo el error del dato contamina la solución viene dada por el error cuadrático medio definido como:

$$ECM = E\left\{ \left\| \mathbf{x}_{mc} - \mathbf{x}_v \right\|_{\mathbf{C}_x^{-1}}^2 \right\}$$

Como matriz de ponderación se utiliza la inversa de la covarianza de la variable aleatoria \mathbf{x}_{mc} , ya que la magnitud de la dispersión de la solución en distintas direcciones del espacio de modelos será, en general, diferente y por tanto los errores deben ser valorados de acuerdo a las distintas capacidades de variación.

El error cuadrático medio se puede reescribir como:

$$ECM = E\left\{\left\|\bar{\mathbf{x}} - \mathbf{x}_v\right\|_{\mathbf{C}_x^{-1}}^2\right\} + E\left\{\left\|\mathbf{x}_{mc} - \bar{\mathbf{x}}\right\|_{\mathbf{C}_x^{-1}}^2\right\}$$

El primer término es se conoce como sesgo de la estimación. En el caso en que los errores de los datos tengan un promedio nulo, debido a (155) y a (156) el sesgo también lo es.

El segundo término es una varianza ponderada para cuyo cálculo se precisa conocer la matriz \mathbf{C}_x de covarianza de la solución en el espacio de modelos. En el caso general de problema mixto de rango r < n, supóngase -por simplicidad- que los errores de los datos están incorrelados y tienen la misma varianza σ_d^2 . Entonces:

$$cov\left(\mathbf{x}_{mc}\right) = E\left\{\left(\mathbf{x}_{mc} - \mathbf{x}_{v}\right)\left(\mathbf{x}_{mc} - \mathbf{x}_{v}\right)^{t}\right\} = \mathbf{A}^{\dagger}cov\left(\mathbf{d}\right)\left(\mathbf{A}^{\dagger}\right)^{t},$$

utilizando la d.v.s la expresión se transforma en:

$$cov\left(\mathbf{x}_{mc}\right) = \mathbf{V}_r\left(\sigma_d \mathbf{\Sigma}_r^{-2}\right) \mathbf{V}_r^t.$$

Se puede representar más intuitivamente en forma matricial como:

$$cov\left(\mathbf{x}_{mc}
ight) = \left(\mathbf{v}_{1}, \mathbf{v}_{2}, \cdots, \mathbf{v}_{r}
ight) \left[egin{array}{ccc} rac{\sigma_{d}}{\sigma_{1}^{2}} & & & \\ & rac{\sigma_{d}}{\sigma_{2}^{2}} & & & \\ & & \ddots & & \\ & & & rac{\sigma_{d}}{\sigma_{r}^{2}} \end{array}
ight] \left(egin{array}{c} \mathbf{v}_{1}^{t} \\ \mathbf{v}_{2}^{t} \\ \vdots \\ \mathbf{v}_{r}^{t} \end{array}
ight).$$

Puede verse cómo la propagación del error del dato es mucho mayor en la dirección de las \mathbf{v}_i asociadas a los valores singulares σ_i más pequeños. El efecto amplificador de éstos pequeños valores es muy grande pues no sólo aparecen en el denominador de los términos de la matriz de covarianza, sino que lo hacen al cuadrado.

Siempre que en un problema existan valores singulares σ_i pequeños, habrá grandes variaciones en la solución y las direcciones de máxima inestabilidad -a lo largo de las cuales la nube de soluciones es más alargada- tienen por vectores directores asociados las v_i correspondientes a dichos valores.

En el decurso del capítulo, las distorsiones geométricas de las nubes solución han aparecido como consecuencia de la elección de dos (o más) puntos de medida muy próximos. Esto aseguraba la obtención de dos filas de la matriz jacobiana muy parecidas y, por tanto, la aparición de valores singulares pequeños indicativos de la cuasi-dependencia lineal del sistema. Sin embargo, en el problema de S.E.V, la aparición de valores singulares pequeños es una característica intrínseca del mismo. Para ilustrar este punto se ha tomado como ejemplo un suelo de 5 capas de resistividades representadas por el vector (10, 20, 1, 50, 2) y espesores dados por (5, 5, 5, 5). Se ha cuidado de que las abscisas superficiales de medida estuviesen regularmente repartidas, para tratar de evitar la posibilidad de recoger datos cuasi-redundantes. La curva de resistividades aparentes con las medidas puede verse en la figura (27) y la curva de los valores singulares del problema en la figura (28). Se observa que es una curva rápidamente decreciente.

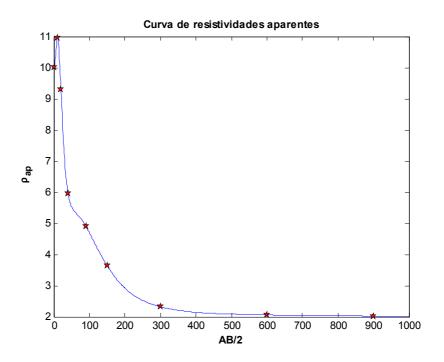


Figura 27: Datos medidos sobre curva de resistividades aparentes.

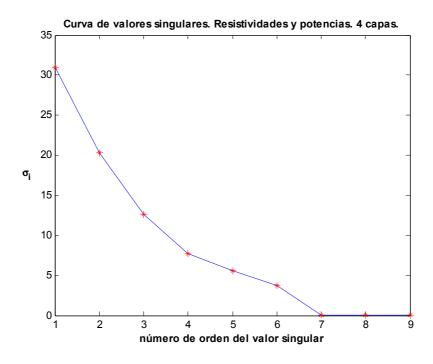


Figura 28: Curva de valores singulares

4.10. Clasificación de los problemas discretos lineales

A veces, la elección de una parametrización desacertada en el espacio de modelos puede conducir a un mal condicionamiento de la matriz del sistema lineal a resolver. Pero en ocasiones, como es el caso en los S.E.V, la matriz del sistema está correctamente mal condicionada⁵², siendo esta propiedad parte de la formulación del problema.

Dentro de los problemas inversos discretos mal condicionados, la literatura ha venido distinguiendo entre dos tipologías diferentes, según su modo de evolucionar hacia el caso contínuo (Hansen, 1998, Hanke, 2002):

- Sistemas denominados deficientes (numéricamente) en rango: caracterizados por la presencia de un grupo de valores singulares pequeños claramente distanciados de otro grupo de valores singulares mayores. Los métodos empleados para remediar las inestabilidades básicamente actúan sobre el grupo de valores singulares pequeños.
- Sistemas denominados tradicionalmente mal planteados: caracterizados porque los valores singulares evolucionan, sin solución de continuidad, desde los valores grandes a los valores pequeños. No es posible definir dos grupos.

Si bien tiene sentido mantener esta distinción, puesto que (en general) requieren tratamientos diferentes, en este trabajo se ha considerado más consistente y clarificador renunciar al uso del término **mal planteado** para los sistemas discretos que muestran un continuo de valores singulares. La razón es doble:

- Por un lado, preservar idéntico el significado del término "mal planteamiento" en el caso continuo y en el caso discreto.
- Incluir el hecho de que, por discretización de problemas continuos "mal planteados" se puede llegar a sistemas discretos deficientes (numéricamente) en rango.

Hadamard, a principios de siglo (Hadamard, 1923), caracterizaba a un problema contínuo mal planteado como aquel que cumplía alguna de estas tres propiedades: carecer de solución; poseer infinitas soluciones o contar con la presencia de discontinuidades en la evolución de la solución del problema en función de los datos del mismo. Esta última propiedad se refiere al hecho de que variaciones arbitrariamente pequeñas en los datos del problema puedan producir cambios arbitrariamente grandes en la solución del mismo.

En un problema discretizado queda descartada tercera posibilidad. Por tanto, sensu estricto, debiera llamarse problema mal planteado sólo a aquellos problemas discretos que no tengan solución o tengan infinitas.

Respecto a aquellos casos que tienen solución única, cabe la posibilidad de que su matriz de coeficientes esté bien condicionada o no lo esté. Y es en este segundo caso donde se recupera la distinción (tradicional) entre sistemas con dos grupos bien diferenciados de valores singulares y sistemas en los que esto no ocurre. A estos últimos se alude mediante la expresión "sistemas de valores singulares no agrupados". La clasificación puede ilustrarse así:

⁵²Se recuerda que el número de condición de una matriz se puede calcular como el cociente entre el valor singular mayor y el más pequeño.

$$Problemas \ discretos \ lineales \left\{ \begin{array}{l} \ Mal \ planteados \left\{ \begin{array}{l} Sin \ solución \\ Infinitas \ soluciones \\ \end{array} \right. \\ Bien \ planteados \left\{ \begin{array}{l} Bien \ condicionados \\ Mal \ condicionados \left\{ \begin{array}{l} v.s. \ agrupados \\ v.s. \ no \ agrupados \end{array} \right. \end{array} \right. \\ \end{array} \right.$$

El problema de S.E.V discreto pertenece al grupo denominado "v.s. no agrupados" aunque pueda asimilarse -sobre todo en discretizaciones de bajo número de parámetros- al caso "v.s. agrupados".

4.11. Concepto de regularización

Entre las patologías hasta aquí descritas se encuentra la no existencia de solución, la presencia de infinitas soluciones y la inestabilidad.

Bajo la denominación de **regularización** se agrupan una serie de procedimientos para incorporar información adicional, modificando el problema, de forma que exista *una* solución del nuevo problema transformado la cual debe ser, además, más estable. La regularización pues, o bien elimina o bien amortigua la influencia que en la solución tiene la parte de la misma "contaminada" por los errores de los datos o del procedimiento numérico. El modo de *filtrar* estas partes indeseadas de la solución depende del método de regularización elegido.

Son de gran utilidad, por tanto, herramientas numéricas que permiten identificar y segregar dichas componentes. Por ejemplo, la ya comentada descomposición en valores singulares (d.v.s) o su extensión, denominada descomposición generalizada en valores singulares (d.g.v.s) que se describirá posteriormente.

4.12. La información a priori: otro modo de expresión

Se ha visto que es posible especificar características de la solución exigiendo su pertenencia a ciertos conjuntos geométricos. Sin embargo, a veces se dispone de estimaciones razonables sobre las magnitudes de ciertas propiedades de la solución. Por ejemplo puede hablarse de soluciones más o menos "rugosas" o "heterogéneas"; de perfiles geoeléctricos que globalmente sean más o menos resistivos o de restricciones en la profundidad del sustrato rocoso -y, por tanto, en la suma de las potencias de las capas superiores-.

Las normas (o seminormas) adquieren nuevamente protagonismo para expresar matemáticamente estas propiedades físicas del terreno solución sobre las que se desea tener cierto grado de control. Una vez escogida la propiedad de interés (rugosidad, resistividad global, etc...) las normas permiten medir el "tamaño" de la solución de acuerdo con aquella. Se hablará pues, abreviadamente, de tamaño de la solución, sobreentendiéndose que se refiere, en realidad, al tamaño de ciertas características suyas. Una formulación suficientemente general para los propósitos de este trabajo es la siguiente:

$$\Omega\left(\mathbf{m}\right) = \left\|\mathbf{L}\left(\mathbf{m} - \mathbf{m}^*\right)\right\|_p,\tag{157}$$

donde \mathbf{m}^* representa alguna estimación disponible a priori, p es el indicativo del tipo de norma (en este trabajo se considera fundamentalmente que p=2 y se reflexionará, más adelante, sobre

el caso p=1) y **L** (π,n) , con $\pi \leq n$, es una matriz de rango π . Se aportan algunos ejemplos ilustrativos:

1. Se pretende tomar control sobre las variaciones de resistividad entre capas consecutivas tomando como base lo que le ocurre en este sentido al modelo de referencia \mathbf{m}^* . Si la parametrización escogida es $\mathbf{m} = (\rho_1, t_1, \rho_2, t_2, \cdots, \rho_n)$,

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 & 0 & -1 \end{bmatrix}.$$

La matriz **L** tiene de dimensiones $\left(\frac{n+1}{2}, n\right)$.

2. Se desea acotar la desviación relativa del valor de los parámetros de **m** respecto de los de **m***, penalizando más rigurosamente la discrepancia relativa en unos parámetros que en otros. Al tratarse de diferencias relativas, parece adecuado utilizar la parametrización logarítmica, de modo que $\mathbf{m} = (\log \rho_1, \log t_1, \log \rho_2, \log t_2, \cdots, \log \rho_n)$ y

$$\mathbf{L} = \begin{bmatrix} \alpha_1 & 0 & 0 & 0 \\ 0 & \alpha_2 & 1 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \alpha_n \end{bmatrix},$$

Hay, naturalmente, más posibilidades. Si \mathbf{m}^* es una buena estimación y \mathbf{L} permite una descripción adecuada de una propiedad relevante, cuanto menor sea $\Omega(\mathbf{m})$ más compatible será la solució hallada con la información a priori.

4.13. La descomposición generalizada en valores singulares

El objeto de la regularización es -como se ha mencionado- reducir el abanico de posibles soluciones del problema inverso (a una en particular) o estabilizar el problema frente a pequeñas modificaciones en los datos. Una de las formas de hacerlo, conocida como **regularización de Tikhonov**, utiliza la información a priori disponible -expresada del modo indicado en el apartado anterior- proponiendo la optimización del siguiente problema modificado:

$$\min_{\mathbf{x}} \|\mathbf{J}\mathbf{F}\mathbf{x} - \mathbf{d}\|_{2}^{2} + \lambda^{2} \|\mathbf{L}(\mathbf{x} - \mathbf{x}_{*})\|_{2}^{2}$$
(158)

 $\mathbf{L}:\mathbb{R}^n \to \mathbb{R}^\pi$ es el operador de regularización. El valor de λ establece el grado de penalización del valor $\Omega\left(\mathbf{m}\right)$ esto es, cuánto se penaliza la discrepancia respecto a la información de referencia. Si $\lambda \to 0$ se tiende a recuperar el problema inicial en el que la información a priori no juega ningún papel. Si $\lambda \to \infty$ se desprecia la información de los datos geofísicos medidos y se busca una solución con el criterio exclusivo de maximizar la compatibilidad con la información a priori. El valor adecuado de λ^{53} combinará del mejor modo ambos tipos de información. A

⁵³ Posteriormente se ofrecerán algunos criterios para su cálculo.

comprender el funcionamiento detallado de este método de regularización ayuda extraordinariamente una herramienta denominada descomposición generalizada en valores singulares, abreviadamente (d.g.v.s).

El funcional de regularización $\Omega(\mathbf{m})$ es una medida del grado de incumplimiento de los requisitos a priori. Un modelo $\mathbf{m} = \mathbf{m}^* + \mathbf{m}_0$ con $\mathbf{m}_0 \in \ker(\mathbf{L})$ anula $\Omega(\mathbf{m})$. Estos modelos, por tanto, serían ya plenamente consistentes con la información aportada y nada cambiaría para ellos. Sin embargo, con mayor generalidad, la diferencia $\mathbf{m} - \mathbf{m}^*$ se puede expresar como:

$$\mathbf{m} - \mathbf{m}^* = \Delta \mathbf{m}_0 + \Delta \mathbf{m}_r$$

donde $\Delta \mathbf{m}_0 \in \ker(\mathbf{L})$ pero $\Delta \mathbf{m}_r \notin \ker(\mathbf{L})$. Es ésta última componente de cada modelo la que, al ser penalizada en el funcional (158) hace que se encuentre una solución distinta que tienda a reducirla. Será denominada componente regularizada.

El operador $\mathbf{JF}: \mathbb{R}^n \to \mathbb{R}^s$ actúa tanto sobre las componentes regularizadas como sobre las no regularizadas. La inestabilidad del problema inverso puede analizarse a través de la $\mathbf{d.v.s}$ de la matriz JF. Se sabe que proviene del hecho de que aparecen valores singulares σ_i muy pequeños y, por tanto, las componentes inestables son las contenidas en los subespacios generados por las \mathbf{v}_i asociadas. Sin embargo, nada se conoce sobre el efecto que sobre dichos vectores tiene la regularización introducida en el problema a través del operador \mathbf{L} . La $\mathbf{d.v.s}$ de la matriz JF por sí sola, como es lógico, no puede clasificar los modelos desde el punto de vista de la acción de \mathbf{L} .

Por ejemplo, es posible que en el problema sin regularizar el vector \mathbf{m} tenga una componente asociada fundamentalmente a valores singulares σ_i pequeños. Por lo tanto, sería muy inestable frente a perturbaciones en los datos. Sin embargo si, tras incluir la regularización \mathbf{L} , resultase muy penalizada porque $\Omega(\mathbf{m})$ sea alto, se produciría un amortiguamiento de las inestabilidades a las que es aparentemente propenso.

Por otro lado, las componentes de un modelo \mathbf{m} más influyentes sobre los datos, asociadas a los valores σ_i más altos, podrían resultar ser las más incompatibles con la información a priori disponible. En ambos casos parece que la base propuesta \mathbf{V} para "desagregar" el comportamiento de un modelo \mathbf{m} bajo la acción del operador JF puede no ser la adecuada para considerar su comportamiento frente a la operación de \mathbf{L} .

Se precisa, pues, una herramienta que permita determinar simultáneamente y de forma sencilla dos cosas:

- 1. Cómo afecta un modelo dado a las predicciones del funcional (y viceversa). Este objetivo es también logrado por la $\mathbf{d.v.s}$ de JF.
- 2. Cómo se ve afectado ese mismo modelo por el proceso de regularización a través del operador L. Sería deseable que aparecieran, también aquí, coeficientes cuyo valor representase el grado de influencia que el operador tiene sobre cada vector. El efecto que un cambio en los datos produciría en la solución dependería de la combinación de los dos tipos de coeficientes. Convendría, por tanto, que dicha combinación relevante tuviese una forma fácilmente matematizable. Como se verá, se trata de su cociente.

Dadas dos matrices JF(s,n) y $\mathbf{L}(\pi,n)$, conjuntamente consideradas en el par (\mathbf{JF},L) , la **descomposición generalizada en valores singulares** de las dos matrices es una operación que genera los siguientes resultados:

lacktriangle Para la matriz JF

$$\mathbf{JF} = \bar{\mathbf{U}}_{(s,n)} \begin{bmatrix} \mathbf{\Sigma}_{(\pi,\pi)} & \mathbf{0}_{(\pi,n-\pi)} \\ \mathbf{0}_{(n-\pi,\pi)} & \mathbf{I}_{(n-\pi,n-\pi)} \end{bmatrix} \mathbf{X}_{(n,n)}^{-1} \Longrightarrow \mathbf{JFX} = \bar{\mathbf{U}} \begin{bmatrix} \mathbf{\Sigma}_{(\pi,\pi)} & \mathbf{0}_{(\pi,n-\pi)} \\ \mathbf{0}_{(n-\pi,\pi)} & \mathbf{I}_{(n-\pi,n-\pi)} \end{bmatrix}. \tag{159}$$

■ Para la matriz **L**

$$\mathbf{L} = \overline{\mathbf{V}}_{(\pi,\pi)} \begin{bmatrix} \mathbf{\Pi}_{(\pi,\pi)} & \mathbf{0}_{(\pi,n-\pi)} \end{bmatrix} \mathbf{X}_{(n,n)}^{-1} \Longrightarrow \mathbf{L} \mathbf{X} = \overline{\mathbf{V}} \begin{bmatrix} \mathbf{\Pi}_{(\pi,\pi)} & \mathbf{0}_{(\pi,n-\pi)} \end{bmatrix}. \tag{160}$$

La matriz \mathbf{X} es regular (no necesariamente bien condicionada) y de dimensiones (n, n). Sus columnas (o sus filas) son, por tanto, una base del espacio de modelos \mathcal{M} . Los términos a la derecha del signo de implicación en las expresiones (159 y 160) muestran un modo alternativo de expresar la acción de los operadores JF y \mathbf{L} sobre el sistema de vectores columna de \mathbf{X} -que es una base de \mathcal{M} -. Por tanto esta descomposición permite calcular el efecto de JF y \mathbf{L} sobre un vector arbitrario de \mathcal{M} , simultáneamente. Éste es uno de los objetivos deseados.

La descomposición generalizada produce dos matrices $\overline{\mathbf{U}}(s,n)$ y $\overline{\mathbf{V}}(p,p)$. Los vectores columna de $\overline{\mathbf{U}}$ forman una base ortonormada del subespacio imagen de JF. Así pues, el efecto que JF ejerce sobre cualquier modelo se puede expresar como una simple combinación lineal de las columnas de $\overline{\mathbf{U}}$. Por su parte, las columnas de $\overline{\mathbf{V}}$ forman una base ortonormada del subespacio imagen de \mathbf{L} . Es decir, el resultado de aplicar la regularización a cualquier vector se expresa como una combinación lineal de las columnas de $\overline{\mathbf{V}}$.

La matriz $\Sigma_{(\pi,\pi)}$ es una matriz diagonal que contiene π escalares $\overline{\sigma}_i, i=1,2,\cdots,\pi$ que cumplen:

$$0 \le \overline{\sigma}_1 \le \overline{\sigma}_2 \le \dots \le \overline{\sigma}_p \le 1.$$

Sobre estos valores cabe aclarar que:

- lacktriangle en general, no guardan relación con los valores singulares de la descomposición en valores singulares de la matriz JF.
- a diferencia dichos valores singulares de **JF**, los valores $\overline{\sigma}_i$ se presentan en orden creciente.
- Cuando JF opera sobre los vectores de la base X se distinguen dos casos:
 - si $1 \le i \le p$, $JF\mathbf{x}_i = \overline{\sigma}_i \overline{\mathbf{u}}_i$. O sea, los valores $\overline{\sigma}_i$ son los factores de amplificación de las coordenadas de un modelo cualquiera expresado en la base \mathbf{X} , al operar sobre él JF y expresar el resultado en la base ortonormada $\overline{\mathbf{U}}$. Según esto, las componentes de \mathbf{m} en las \mathbf{x}_i correspondientes a un $\overline{\sigma}_i$ de pequeño valor apenas influyen en las predicciones y, por tanto, estas componentes son las más susceptibles a las inestabilidades del problema inverso
 - si el índice de columna es $p+1 \le i \le n$ el vector y los datos por él predichos mantienen las mismas coordenadas en ambas bases $(\mathbf{X}, \overline{\mathbf{U}})$. Además, estas \mathbf{x}_i forman una base del núcleo del operador \mathbf{L} generando todos los modelos que no "sienten" la regularización. Así pues, las componentes de los modelos que no se ven afectados por la regularización conservan, al transformarse por acción de JF las mismas coordenadas en la base $\overline{\mathbf{U}}$ que las que tenían en la base \mathbf{X} .

• La matriz $\Pi_{(\pi,\pi)}$ también es una matriz diagonal. Contiene los π escalares μ_i , $i=1,2,\cdots,p$, para los que se cumple que:

$$1 \ge \mu_1 \ge \mu_2 \ge \dots \ge \mu_\pi \ge 0.$$

Como se ve, se presentan en sucesión decreciente:

- si $1 \le i \le p$, $L\mathbf{x}_i = \mu_i \overline{\mathbf{v}}_i$. Es decir, la regularización multiplica las coordenadas de un modelo en la base \mathbf{X} por los factores μ_i . Un valor de μ_i alto, significa una componente más penalizable en el funcional, o sea, indica que las componentes del modelo se alejan del comportamiento deseado.
- si $p+1 \le i \le n$, se trata de componentes que están en el núcleo de L.
- Se cumple además la relación:

$$\overline{\sigma}_i^2 + \mu_i^2 = 1,$$

y, por tanto,

$$\mu_i = \sqrt[2]{1 - \overline{\sigma}_i^2} \,.$$

Esta expresión indica que a las componentes más influyentes en JF y, por tanto, menos inestables en el problema inverso ($\overline{\sigma}_i$ grandes) les corresponden valores μ_i bajos. Esto muestra que tienden a ser también consistentes con la información a priori. Viceversa, las componentes más inestables del problema inverso ($\overline{\sigma}_i$ pequeñas) son, por tanto, como debe ser, las más afectadas por la regularización (μ_i altos).

Puede demostrarse que la solución al problema (158) es:

$$\mathbf{x}_{reg} = \sum_{i=1}^{i=p} f_i \frac{\left(\mathbf{u}_i^t \cdot \mathbf{d}\right)}{\overline{\sigma}_i} \mathbf{x}_i + \sum_{i=p+1}^{i=n} \left(\mathbf{u}_i^t \cdot \mathbf{d}\right) \mathbf{x}_i$$
 (161)

donde

 $f_i = \frac{\gamma_i^2}{\lambda^2 + \gamma_i^2},$

У

$$\gamma_i = \frac{\overline{\sigma}_i}{\mu_i}, \ i = 1, 2, \cdots, p,$$

son los denominados valores singulares generalizados. Se presentan en orden creciente con i y reflejan la influencia que en el subespacio generado por cada \mathbf{x}_i tienen simultáneamente el operador JF, a través de las $\overline{\sigma}_i$, y el operador \mathbf{L} , mediante las μ_i . Subespacios asociados a los primeros valores generalizados, γ_i pequeños, son susceptibles de sufrir inestabilidades y, por tanto, preferentemente afectados por la regularización. Por contra, los subespacios asociados a valores singulares de índice i mayor, γ_i grandes, son más estables y la regularización actúa menos sobre ellos.

En forma matricial (161) se puede expresar como:

$$\mathbf{x}_{reg} = \mathbf{X} \mathbf{\Theta} \left(egin{array}{cc} \mathbf{\Sigma}^\dagger & \mathbf{0} \ \mathbf{0} & \mathbf{I}_{(n-\pi,n-\pi)} \end{array}
ight) ar{\mathbf{U}}^t \mathbf{d},$$

donde la matriz Θ es una matriz diagonal,

$$\mathbf{\Theta} = \begin{pmatrix} \frac{\gamma_1^2}{\lambda^2 + \gamma_1^2} & 0 & 0 & 0\\ 0 & \frac{\gamma_2^2}{\lambda^2 + \gamma_2^2} & 0 & 0\\ 0 & 0 & \ddots & 0\\ 0 & 0 & 0 & \frac{\gamma_3^2}{\lambda^2 + \gamma_3^2} \end{pmatrix},$$

que filtra o amortigua los términos $\frac{(\mathbf{u}_i^t \cdot \mathbf{d})}{\overline{\sigma}_i}$. Cuanto mayor sea el valor del parámetro λ^2 menores son los términos de Θ y más se amortiguan las componentes correspondientes de la solución, evitando oscilaciones excesivamente grandes. La matriz de filtros es diferente en función del método de regularización escogido. Se examinarán al estudiar cada uno de ellos.

Finalmente es preciso comentar que en este trabajo sólo se hace alusión a los denominados **métodos de regularización directos**, en los que es preciso disponer de las matrices JF y/o JF^t explícitamente. Frente a éstos existen también métodos de regularización **iterativos**, los cuales sólo calculan los **productos** $JF\mathbf{x}$ y $JF^t\mathbf{y}$ sin calcular JF y/o JF^t y generan una secuencia iterativa de soluciones que convergen a la deseada. Estos últimos son métodos adecuados cuando se trata de grandes matrices, con estructura que merece la pena preservar, que obligan a ahorrar tiempo.

Dentro de los métodos directos, se realizará un análisis detallado del método de regularización de Tikhonov.

4.14. Regularización de Thikonov

El denominado método de regularización de Tikhonov ha sido desarrollado independientemente por Philips (1962) y Tikhonov (1963). En lugar de resolver el problema

$$\min_{\mathbf{x}} \|\mathbf{J}\mathbf{F}\mathbf{x} - \mathbf{d}\|_{2}^{2},\tag{162}$$

donde \mathbf{x} es, o bien \mathbf{m} o bien $\Delta \mathbf{m}$ se resuelve otro problema obtenido al añadir a (162) un término de penalización $\Omega(\mathbf{x}) = \|\mathbf{L}(\mathbf{x} - \mathbf{x}^*)\|_2^2$ que contiene la información a priori. El nuevo problema es:

$$\min_{\mathbf{x}} \|\mathbf{J}\mathbf{F}\mathbf{x} - \mathbf{d}\|_{2}^{2} + \lambda^{2} \|\mathbf{L}(\mathbf{x} - \mathbf{x}^{*})\|_{2}^{2}.$$
 (163)

Las ecuaciones normales asociadas son:

$$\left(\mathbf{J}\mathbf{F}^{t}\mathbf{J}\mathbf{F} + \lambda^{2}\mathbf{L}^{t}\mathbf{L}\right)\mathbf{x} = \mathbf{J}\mathbf{F}^{t}\mathbf{d} + \lambda^{2}\mathbf{L}^{t}\mathbf{L}\mathbf{x}^{*},$$

ecuaciones que muestran que la formulación (163) puede ser expresada también como:

$$\min_{\mathbf{x}} \left\| \begin{pmatrix} \mathbf{JF} \\ \lambda \mathbf{L} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{d} \\ \lambda \mathbf{L} \mathbf{x}^* \end{pmatrix} \right\|_{2}^{2}, \tag{164}$$

en la que se observa más fácilmente que si la intersección entre los nucleos de \mathbf{JF} y \mathbf{L} se reduce al vector nulo, existe una única solución de (164) que viene dada por:

$$\mathbf{x}_{mc,\lambda} = \left(\mathbf{J}\mathbf{F}^t\mathbf{J}\mathbf{F} + \lambda^2\mathbf{L}^t\mathbf{L}\right)^{-1} \left(\mathbf{J}\mathbf{F}^t\mathbf{d} + \lambda^2\mathbf{L}^t\mathbf{L}\mathbf{x}^*\right). \tag{165}$$

Si se denomina

$$\mathbf{J}\mathbf{F}_{\lambda}^{\#} \equiv \left(\mathbf{J}\mathbf{F}^{t}\mathbf{J}\mathbf{F} + \lambda^{2}\mathbf{L}^{t}\mathbf{L}\right)^{-1}\mathbf{J}\mathbf{F}^{t},$$

у

$$\mathbf{L}_{\lambda}^{\#} \equiv \left(\mathbf{J}\mathbf{F}^{t}\mathbf{J}\mathbf{F} + \lambda^{2}\mathbf{L}^{t}\mathbf{L}\right)^{-1}\mathbf{L}^{t},$$

la solución (165) se puede expresar de forma compacta como:

$$\mathbf{x}_{mc,\lambda} = \mathbf{J}\mathbf{F}_{\lambda}^{\#}\mathbf{d} + \lambda^{2}\mathbf{L}_{\lambda}^{\#}\mathbf{L}\mathbf{x}^{*}.$$

Si se opera sobre esta expresión utilizando la descomposición (resp. generalizada) en valores singulares de **JF** (resp. de **JF** y **L**) se pueden detallar más las componentes de $\mathbf{x}_{mc,\lambda}$. En particular, si se utiliza sólo la base **V** procedente de la **d.v.s** de **JF** y $\mathbf{L} = \mathbf{I}_n$,

$$\mathbf{x}_{mc,\lambda} = \sum_{i=1}^{i=n} \left(f_i \frac{\left(\mathbf{u}_i^t \cdot \mathbf{d} \right)}{\sigma_i} + (1 - f_i) \mathbf{v}_i^t \mathbf{x}^* \right) \mathbf{v}_i, \tag{166}$$

siendo

$$f_i = \frac{\sigma_i^2}{\lambda^2 + \sigma_i^2}.$$

Mientras que si $\mathbf{L} \neq \mathbf{I}_n$ es preferible utilizar la **d.g.v.s** del par (**JF**,**L**) para obtener,

$$\mathbf{x}_{mc,\lambda} = \sum_{i=1}^{i=p} \left(f_i \frac{\left(\mathbf{u}_i^t \cdot \mathbf{d} \right)}{\bar{\sigma}_i} + (1 - f_i) \, \boldsymbol{\phi}_i^t \mathbf{x}^* \right) \mathbf{x}_i + \sum_{i=p+1}^{i=n} \left(\mathbf{u}_i^t \cdot \mathbf{d} \right) \mathbf{x}_i, \tag{167}$$

con

$$f_i = \frac{\gamma_i^2}{\lambda^2 + \gamma_i^2}.$$

La presencia de los factores de filtro f_i es una característica relevante de estas expresiones. Observándo (166) y (167) puede verse que los errores en \mathbf{d} influyen en la solución $\mathbf{x}_{mc,\lambda}$ a través del término $\frac{\mathbf{u}_i^t \cdot \mathbf{d}}{\sigma_i}$ (resp. de $\frac{\mathbf{u}_i^t \cdot \mathbf{d}}{\bar{\sigma}_i}$) amplificando las incorrecciones asociadas a las componentes de la misma vinculadas a los valores singulares σ_i (resp. $\bar{\sigma}_i$) menores. Por tanto, es deseable que las f_i amortigüen precisamente estas variaciones disminuyendo su valor cuando i aumenta (resp. i disminuye).

4.15. Análisis lineal de los efectos de la regularización de Tikhonov

Tal como se sugiere en la introducción de este capítulo, el problema inverso lineal es una etapa intermedia de la solución del problema inverso no lineal en S.E.V, si se emplean algoritmos locales. Así que es importante estudiar el efecto de la regularización en cada una de esas etapas para entender cómo se transmite la inestabilidad de una a otra, hasta la solución final. Se comienza por la generación de datos sintéticos sobre un terreno interesante, abordando la solución y análisis mediante la **descomposición en valores singulares**. Posteriormente se propondrá otro caso donde se aprecian sustancialmente las ventajas del empleo de la **descomposición generalizada en valores singulares**.

En el primer caso (**d.v.s**), se comienza ilustrando la evolución de los filtros con λ e i. Considérese el suelo de 5 capas caracterizado por las resistividades (10, 1, 50, 10, 1) y espesores (5, 5, 5, 5). En la figura 29 puede observarse, arriba, su curva de valores singulares y en la parte inferior la evolución de los filtros de tikhonov en función del orden del vector singular.

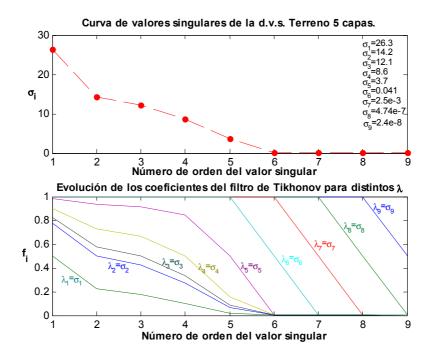


Figura 29: Curva de valores singulares y filtros para la regularización de Tikhonov.

Cada curva corresponde a uno de los valores de λ indicados. Se aprecian dos comportamientos diferentes:

1. Si $\lambda \leq \sigma_9$, el filtrado sólo afecta a la componente de la solución representada por el vector \mathbf{v}_9 que, tal como se vé en la tabla $\mathbf{\Phi}$, recoge casi exclusivamente información de ρ_3 y ρ_4 . La acción del filtro **no afecta** a la información situada en \mathbf{v}_i con i < 9, ni siquiera al anterior, i = 8. El mismo análisis y conclusión se mantiene para el resto de los valores de λ mientras éstos cumplan que $\lambda \leq \sigma_6$. La razón de este comportamiento estriba en que, en este tramo, $\sigma_i \ll \sigma_{i-1}$, por tanto, cuando λ es del orden de σ_i , el valor de $f_i = \frac{\sigma_i^2}{\lambda^2 + \sigma_i^2} \simeq 0,5$ pero el valor de $f_{i-1} = \frac{\sigma_{i-1}^2}{\lambda^2 + \sigma_{i-1}^2} \simeq \frac{\sigma_{i-1}^2}{\sigma_i^2 + \sigma_{i-1}^2} \simeq 1$.

Este filtrado es, por tanto, efectivo a la hora de disminuir las inestabilidades y no modifica la calidad de las predicciones del problema directo, ya que **sólo** afecta a la información asociada a los valores singulares pequeños.

2. Cuando $\lambda \geq \sigma_5$, el filtrado de una componente, por ejemplo \mathbf{v}_5 , afecta a todas las anteriores. Esto es así porque los σ_i con $i \leq 5$ tienen todos el mismo orden de magnitud. Por tanto,

cuando λ es capaz de reducir el valor del filtro f_i también lo hace (aunque en menor medida) con los anteriores. Este es un hecho importante porque significa que, a partir de cierto valor de λ se modifica información asociada a valores singulares grandes y, por tanto, de gran influencia en el ajuste de los datos. Necesariamente el desajuste tiene que aumentar.

A la luz de la expresión (166) y de la ilustración realizada sobre los filtros cabe valorar la influencia de la información a priori. Cuando $\lambda \leq \sigma_6$ la importancia de la misma viene dada por el factor $(1 - f_i) \simeq 1$. Esto significa que la regularización de tikkhonov casi "elimina" estas componentes de la solución "propuestas" por el dato erróneo **sustituyéndolas** por las correspondientes componentes del propio modelo de referencia. En el resto de los casos la información a priori entra de manera más distribuida.

La siguiente tabla contiene, en cada una de sus columnas, información relativa a cada uno de los vectores singulares \mathbf{v}_i del espacio de modelos. El número en cada casilla refleja el porcentaje aproximado de dicha información que influye en el parámetro de la misma fila. Por ejemplo, la cuadrícula situada en columna correspondiente a \mathbf{v}_9 y en la fila de ρ_4 contiene un 80,6, lo cual significa que aproximadamente el 80 % de la información que posee \mathbf{v}_9 se concentra en la resistividad de la cuarta capa en exclusiva. Se utilizará esta tabla ($\mathbf{\Phi}$) en lo sucesivo.

Φ	\mathbf{v}_1	\mathbf{v}_2	\mathbf{v}_3	\mathbf{v}_4	\mathbf{v}_5	\mathbf{v}_6	\mathbf{v}_7	\mathbf{v}_8	\mathbf{v}_9
$oldsymbol{ ho}_1$	64	15	2	3	0.2	0	0	0	0
\mathbf{t}_1	22	35	10	26	3,8	2.5	0.4	0	0
$oldsymbol{ ho}_2$	6.6	17.6	1.6	29.2	7.6	39.4	7.8	1	0
\mathbf{t}_2	5.1	14.8	1.7	30	7.7	40	8.5	1	1
$oldsymbol{ ho}_3$	0.0	7	29	0	12.6	5	18.3	35	13.6
\mathbf{t}_3	0	7	29	0	12.7	2	9	40	3
$oldsymbol{ ho}_4$	0	1.4	6	0	2.6	0	3	4	80.6
\mathbf{t}_4	0	1.3	5	0	2.5	10	52	19.6	1
$oldsymbol{ ho}_5$	0	1	14.8	11	50.3	0	0	0	0

La estructura resistiva del terreno presenta particularidades muy relevantes en la práctica. La segunda capa es mucho menos resistiva que la tercera y posee una potencia comparable a la del estrato suprayacente. En estas condiciones la reglas de equivalencia de Maillet afirman que existen múltiples combinaciones parámetros, ρ_2 y t_2 , del segundo estrato generadoras de predicciones prácticamente idénticas. La única condición es que el parámetro denominado conductancia longitudinal, definido como:

$$C = \frac{\rho_2}{t_2},$$

permanezca aproximadamente constante. Como en el presente análisis se ha utilizado parametrización logarítmica del espacio de modelos, la ecuación anterior se escribiría como:

$$\log \rho_2 = \log t_2 - \log C,\tag{168}$$

que, representada en un sistema ($\log t_2$, $\log \rho_2$), es una recta de pendiente unidad.

El problema inverso se ha resuelto generando, sintéticamente, datos que no difieren (tal como se supone que ocurre en las buenas campañas de campo) de las medidas libres de error en más

de un 5 %. A fin de analizar qué efectos tiene esta técnica de regularización sobre el fenómeno de equivalencia detectado en la segunda capa, han sido calculados y representados los parámetros $(\log t_2, \log \rho_2)$ de los modelos solución -correspondientes a dichos datos- obtenidos al resolver con diferentes valores del parámetro de regularización λ . Particular hincapié se ha hecho en el estudio del intervalo $\sigma_1 \leq \lambda \leq \sigma_9$. Como información a priori se ha considerado el modelo $\mathbf{m}^* = (10, 5, 10^5, 10^2, 50, 5, 10, 5, 1)$.

Las conclusiones obtenidas, su justificación y la ilustración -cuando es posible- de esta última, se exponen a continuación:

- 1. En la figura 30 se representan las proyecciones, sobre el plano $(\log t_2, \log \rho_2)$, de las nubes de modelos solución del problema planteado. Se concluye lo siguiente:
 - a) caso $\sigma_9 \leq \lambda \leq \sigma_6$,
 - 1) Se aprecia claramente la dispersión en los valores ρ_2 y t_2 , confirmándose empíricamente la equivalencia predicha.
 - 2) Las nubes solucion se disponen de manera bastante ajustada sobre una recta de pendiente +1, tal como indica la relación (168), confirmando que, al menos en el problema lineal las reglas de Maillet se cumplen. No obstante, las nubes tienen un tamaño que, aunque grande, es limitado.
 - 3) A medida que λ aumenta desde σ_9 hasta σ_6 la dispersión disminuye drásticamente (hasta 4 órdenes de magnitud). Esta evolución encuentra justificación teniendo en cuenta las relaciones entre magnitudes de valores singulares contiguos y los porcentajes indicados en la tabla Φ .

Cuando se aumenta λ , por ejemplo, del valor σ_9 a σ_8 , el posible efecto de \mathbf{v}_9 sobre la segunda capa desaparece⁵⁴ y se reduce a la mitad la influencia de \mathbf{v}_8 que, de por sí, tiene sólo del orden de un 2% de su contenido dedicado los parámetros de la segunda capa. El orden de magnitud aproximado de la dispersión es, entonces:

$$d_{(\rho_2,t_2)} \simeq \frac{\mathbf{u}_8^t \mathbf{e}}{\sigma_8} |\mathbf{v}_8|_{(\rho_2,t_2)} \simeq \frac{0.05}{4.7 \cdot 10^{-7}} 2 \cdot 10^{-2} \simeq 2 \cdot 10^3,$$

donde **e** es el vector de error, $|\mathbf{v}_8|_{(\rho_2,t_2)}$ es una medida del contenido de \mathbf{v}_i relativo a la segunda capa y $d_{(\rho_2,t_2)}$ la dispersión los parámetros de la misma (de sus logaritmos). Igualmente puede razonarse con los restantes valores de λ del intervalo aquí considerado.

- b) caso $\lambda \geq \sigma_5$,
 - 1) Ha desaparecido prácticamente la dispersión. La mayor dimensión de la nube observada es del orden de 0,05, lo cual supone una cota superior de los errores relativos del orden del 5 %.
 - 2) El hecho de que la dispersión remanente tenga una dirección perpendicular a la anterior, se debe que todas las componentes de la solución en la dirección precedente han sido filtradas.

 $^{^{54}}$ En este caso, de la tabla Φ se puede deducir que dicha contribución ya era nula.

3) Al aumentar λ se aprecia un desplazamiento de la nube hacia arriba -en dirección al valor log ρ₂ = 5- y hacia la izquierda -hacia log t₂ = 2-. Es la influencia de la información a priori introducida con el vector m*. Incluso la ligera asimetría observada en los valores de log ρ₂ hacia valores positivos, en el caso λ = σ₆ pudiera atribuirse a esta misma causa, indicando que la información a priori comienza a actuar con intensidad a partir de este valor de λ.

Por lo tanto, para valores de λ inferiores a $\sigma_6=0.041$, la solución está dominada por el error cometido en los datos de campo, existiendo dispersión apreciable. Para valores de λ superiores a $\sigma_5=3.7$, los errores de campo apenas influyen y, en cambio, domina la información a priori suministrada. Valores de λ tal que $0.041<\lambda<3.7$, caracterizan un rango en el que ambos -datos e información a priori- influyen de forma similar. La dispersión es razonablemente baja, aportando fiabilidad a los resultados, y además tiende a ser consistente con los conocimientos previos. Es, pues, una zona interesante desde el punto de vista práctico.

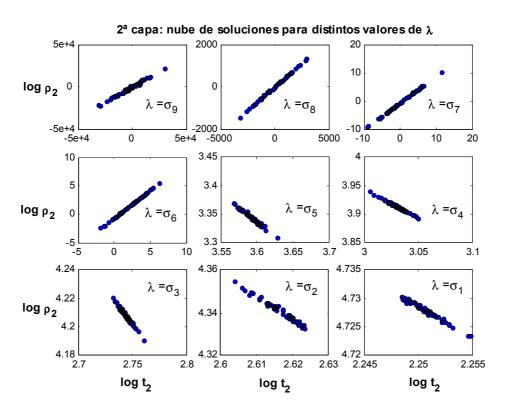


Figura 30: Dispersión en las soluciones para distintos valores de λ

2. **La curva L**: se han graficado de forma conjunta dos magnitudes de interés práctico: el residuo en la predicción de los datos medidos, $\|\mathbf{JFm} - \mathbf{d}\|_2$, y la norma de la solución, $\|\mathbf{m}\|_2$, en cuyo valor puede verse el grado de compatibilidad con la información a priori

suministrada. Dos casos han sido tomados en cuenta:

a) modelo de referencia representando estratos de potencias y resistividades unitarias,

$$\mathbf{m}^* = \log(1, 1, 1, 1, 1, 1, 1, 1, 1)$$

b) modelo de referencia

$$\mathbf{m}^* = (10, 5, 10^5, 10^2, 50, 5, 10, 5, 1),$$

considerado anteriormente.

Las gráficas se ilustran en las figuras 31 y 32, respectivamente. Se pueden extraer las sigu-

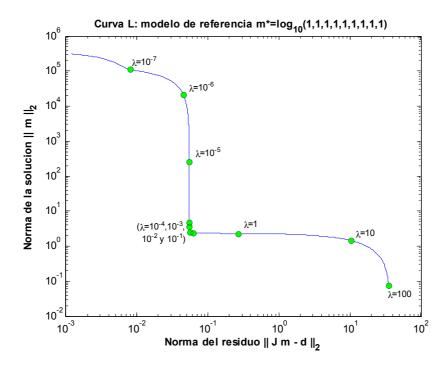


Figura 31: Curva L. El modelo de referencia minimiza profundidades y resistividades ientes conclusiones:

a) Para valores de $\lambda < \sigma_6$, la figura anterior mostraba como segura la obtención de un buen número de soluciones irreales. La curva graficada, en los dos casos considerados, recoge el hecho de que dichas soluciones producen un ajuste razonable, como corresponde al caso de equivalencia.

La curva se dispone, para estos valores de λ , o bien en un tramo horizontal de norma extraordinariamente alta, o bien en un tramo casi vertical en el que pequeñas variaciones de λ producen grandes cambios en la compatibilidad con la información a priori.

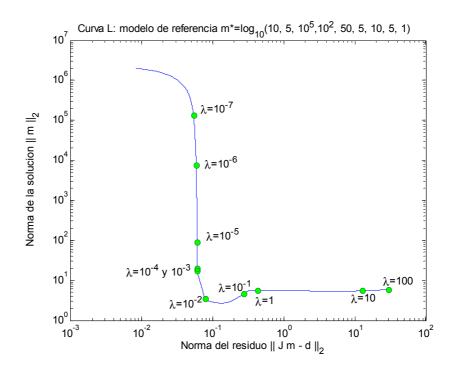


Figura 32: Curva L. La información a priori es un modelo determinado de referencia.

b) El tránsito desde $\lambda = \sigma_6$ hasta $\lambda = \sigma_5$, estabiliza los valores de la norma de la solución en valores cercanos a la del modelo de referencia -lo cual certifica la eliminación o amortiguación por filtrado de modelos irreales-. En la curva se observa cómo este sesgo progresivo de la solución hacia el modelo de referencia va acompañado de un desplazamiento hacia la derecha -valores crecientes del desajuste- del punto representativo. Un desplazamiento excesivo, daría predicciones con error intolerable. El filtrado completo de \mathbf{v}_6 -que concentra prácticamente toda su información en la segunda capa- ocasionado con $\lambda = \sigma_5$, eliminaría todas las equivalencias y, entre ellas, algunas soluciones como $\rho_2 = 200$ y $t_2 = 200$, que pudiesen ser factibles. En ausencia de otros criterios, es el peso dado a la información a priori el que decide, condicionando el grado de amortiguamiento.

Las curvas representadas tienen, representadas en escala logarítmica, un acodamiento característico que les proporciona una clara forma de L. Nuevamente se llega a la conclusión de la existencia de una zona óptima en los alrededores de dicho acodamiento.

- c) Existe diferencia entre las curvas L obtenidas para los dos modelos de referencia. En el segundo caso el acodamiento presenta un mínimo. Esto indica simplemente que, en el proceso, pueden obtenerse modelos cuya norma sea menor que la del modelo de referencia. En el primer caso es imposible puesto que la norma de este es nula.
- 3. Análisis del efecto sobre *una* solución particular. Puede verse en la figura 33 donde la solución correspondiente al dato sin error se ha graficado en primer lugar. Se aprecia que:

a) para $\lambda < \sigma_6$, p.ej. $\lambda \simeq \sigma_7$, (recuadro 1.3), se aprecian grandes inestabilidades. No suelen generarse soluciones aprovechables.

- b) para $\lambda \simeq \sigma_6$, (recuadro 2.1 y 1.2), aparecen comportamientos y valores razonables propios de la solución verdadera.
- c) para $\lambda \geq \sigma_5$, (resto de recuadros), se observa un claro desplazamiento hacia arriba del trazo horizontal correspondiente a la segunda capa, signo de la influencia de la información a priori-poseedora de una resistividad $\rho_2 = 10^5$ -.

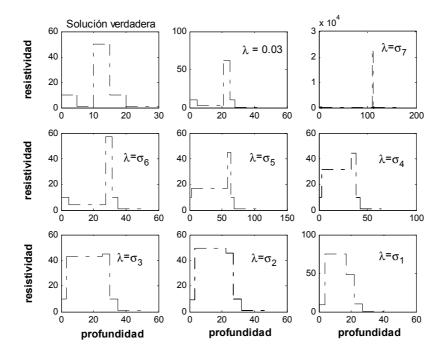


Figura 33: Soluciones: regularización de Tikhonov para distintos valores de λ

Se plantea ahora un problema ligeramente diferente al expuesto a fin de presentar las prestaciones de la descomposición generalizada en valores singulares respecto a la simple.

Se quiere resolver un problema inverso en S.E.V, buscando obtener un suelo que tenga las *mínimas variaciones posibles de resistividad*. Siguiendo las directrices indicadas en este capítulo para expresar la información a priori se tiene que:

$$\Omega(\mathbf{m}) = \|\mathbf{L}(\mathbf{m} - \mathbf{m}^*)\|_2$$

donde se emplea parametrización logarítmica y, por tanto,

$$\mathbf{m} = (\log \rho_1, \log t_1, \log \rho_2, \log t_2, \log \rho_3, \log t_3, \log \rho_4, \log t_4, \log \rho_5),$$

У

con $\mathbf{m}^* = 0$, expresando el hecho de que se desearían *anular* las diferencias entre las resistividades de las capas contiguas.

La **d.v.s** utilizada en el caso anterior propone, como base del espacio de modelos, unos vectores \mathbf{v}_i que distribuyen la información de los posibles terrenos solución de una forma insatisfactoria para los requerimientos del problema actual. Por ejemplo, no resulta obvio cómo afectaría a la magnitud log $\frac{\rho_2}{\rho_3}$ (y restantes como ella) y, simultáneamente al desajuste de las predicciones, un amortiguamiento de, digamos, el vector \mathbf{v}_8 . La **d.g.v.s** proporciona una nueva base cuyos vectores, \mathbf{x}_i , contienen estos dos tipos de información de manera jerarquizada.

De la **d.g.v.s** se han obtenido los valores singulares generalizados, γ_i , y se ha graficado la evolución de los filtros de Tikhonov -véase figura 34. Además, en la figura 35 se han representado las componentes de los vectores, \mathbf{x}_i , de la nueva base.

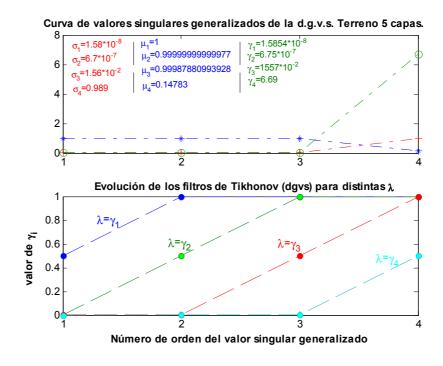


Figura 34: Valores singulares generalizados y filtros de Tikhonov correspondientes.

Se pueden extraer las siguientes conclusiones:

1. Dado que el rango de ${\bf L}$ es 4, sólamente hay cuatro valores singulares generalizados, γ_i , que responden a la fórmula

$$\gamma_i = \frac{\overline{\sigma}_i}{\mu_i}, i = 1, 2, \cdots, 4.$$

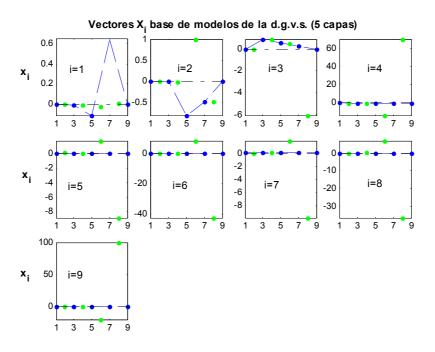


Figura 35: Funciones de base para d.g.v.s

Los valores de los tres parámetros aludidos en esta expresión aparecen graficados en la figura $(filtro5c_gen.fig)$ -parte superior-. El parámetro γ_i crece con el índice i.

- 2. El menor de todos, $\gamma_1 = 1,58 \cdot 10^{-8}$, posee un valor extraordinariamente bajo indicando que las componentes de la solución en la dirección de \mathbf{x}_1 son, tanto poco relevantes para la predicción de datos como muy poco consistentes con los requisitos exigidos por la información a priori. El valor de $\overline{\sigma}_1 = 1,58 \cdot 10^{-8}$ confirma lo primero. Lo segundo puede constatarse examinando en la figura ($vect_Vgen5c.fig$) -primer recuadro- las componentes impares de \mathbf{x}_1 , relacionadas con la resistividad, representadas por puntos unidos por una línea. Se aprecian -en las oscilaciones de la línea- las relativamente altas diferencias entre los valores de las resistividades contiguas. El valor $\mu_1 \simeq 1$ ratifica esta conclusión.
- 3. Los restantes valores γ_i , al aumentar de valor con i, se vinculan con componentes de la solución, \mathbf{x}_i , i=2,3,4, más adecuadas. En la $(vect_Vgen5c.fig)$ se aprecia cómo las oscilaciones aludidas disminuyen hasta desaparecer en i=4. Asimismo los valores $\overline{\sigma}_i$ crecen, indicando importancia creciente en la predicción de datos.
- 4. Si i > 4, todas las \mathbf{x}_i asociadas tienen capas de idéntica resistividad. Obsérvese que no ocurre lo mismo con las potencias de las capas. Esto es así porque sólo se dispone de información a priori relativa a las resistividades, quedando las potencias fuera de los efectos del proceso regularizador. Estos vecto res constituyen una base del núcleo de \mathbf{L} y la teoría afirma que son bastante estables.

5. El comportamiento de los filtros es prácticamente idéntico al del caso anterior. Al aumentar el valor de λ se amortiguan los efectos de las componentes más inestables y penalizables. Cuando $\lambda = \gamma_i$, el efecto de la componente en \mathbf{x}_i se reduce a la mitad, desaparece la influencia de \mathbf{x}_{i-1} y precedentes y no se altera la influencia de las posteriores.

Para ver, de manera más clara, cuál es el efecto del aumento de λ sobre las magnitudes de interés, se ha elaborado la figura

36. En ella se muestran las dispersiones que poseen, al variar λ , las magnitudes $\log \frac{\rho_i}{\rho_{i+1}}$ -de contraste

(relativo) de resistividad entre capas sucesivas- pertenecientes a las soluciones del problema de S.E.V mencionado, con datos

erróneos en no más de un 5%.

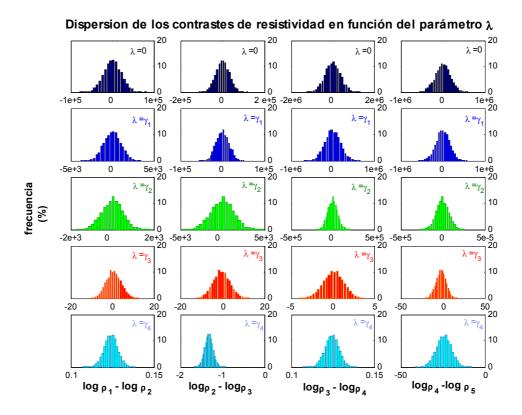


Figura 36: Dispersión del contraste de resistividades en función de λ

Se concluye lo siguiente:

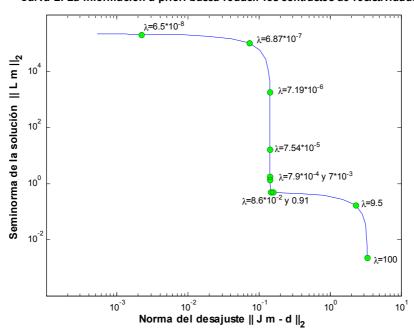
- 1. Si $\lambda \simeq 0$, la dispersión es del orden de 10^5 o superior. Las soluciones son potencialmente desaprovechables.
- 2. Si λ crece hasta el valor $\lambda \simeq \gamma_2$, la dispersión disminuye, pero continúa siendo demasiado elevada.

3. Cuando $\lambda \simeq \gamma_3$, se observa una disminución brusca de la dispersión del contraste de resistividades. La justificación estriba en que este valor elimina la contribución de las componentes, nada adecuadas, anteriores.

4. Para λ ≃ γ₄, el control de la solución lo tiene la información a priori como muestran los sesgos observados en los tres últimos histogramas y la pequeña dispersión del primero. En los tres últimos, referentes a capas de más profundidad y menos controladas por las medidas de superficie, el efecto de la información a priori ha tenido que manifestarse primero para estabilizar la situación. En el primero, de menor inestabilidad inherente por estar más cerca de la superficie las capas implicadas, esta influencia es menor.

A la luz de lo expuesto, un proceso que genere soluciones aprovechables debe funcionar con $\gamma_3 \leq \lambda \leq \gamma_4$.

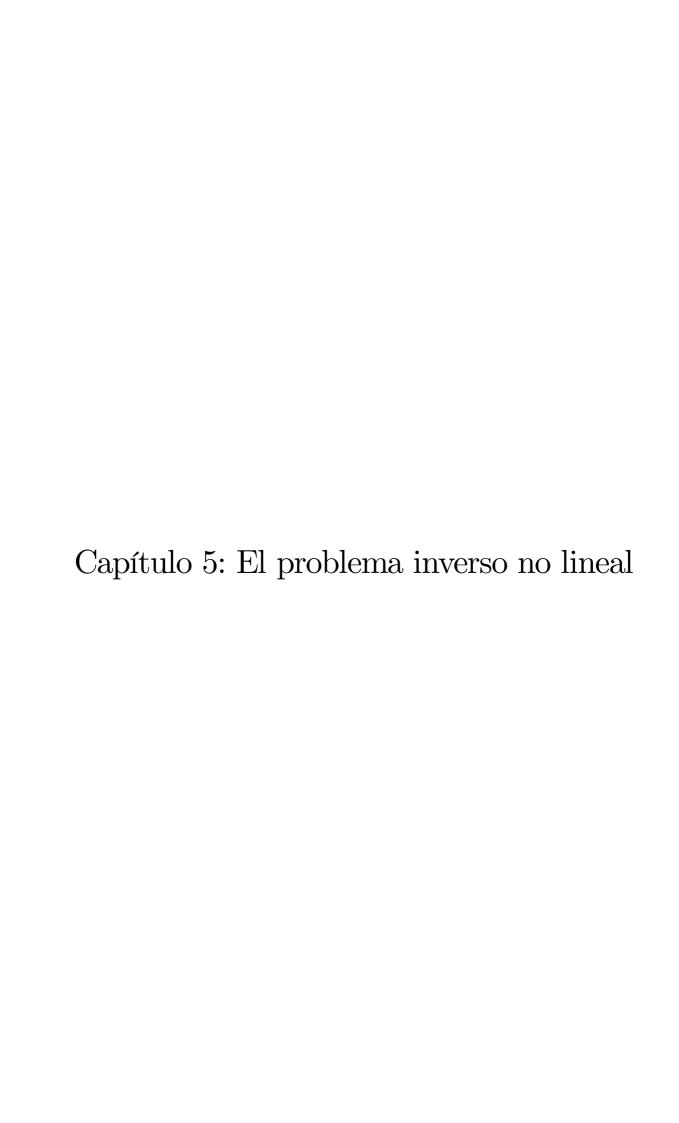
Se ha representado, adicionalmente, la curva que relaciona la norma de la magnitud de interés $\|\mathbf{Lm}\|_2$ con el desajuste en las predicciones $\|\mathbf{JFm} - \mathbf{d}\|_2$ -figura 37-. Se observa la clásica forma de L.



Curva L: La información a priori busca reducir los contrastes de resistividades

Figura 37: Curva L. El modelo de referencia pretende atenuar el contraste de resistividades.

El acodamiento de la curva coincide con la región de interés mencionada. Un valor de λ inferior, produce rápidamente valores altísimos de $\|\mathbf{Lm}\|_2$ -significando esto gran discrepancia con la información a priori- . Asimismo, un valor de λ superior produciría desajustes intolerables con los datos.



5. EL PROBLEMA INVERSO NO LINEAL

5.1. Formulaciones del problema no lineal

En el primer apartado del capítulo lineal, al exponer la motivación para estudiar los problemas lineales ha sido preciso plantear lo que se entiende por problema inverso no lineal. Se mostraba allí que es necesario disponer de un conjunto M, de modelos admisibles descritos mediante una determinada parametrización. La ley física que relaciona cada uno de estos modelos con las predicciones observables del problema directo viene representada matemáticamente por el funcional:

$$\mathbf{F}: \mathbb{M} \subset \mathbb{R}^n \to \mathbb{D} \subset \mathbb{R}^s$$
.

 ${f F}$ es, en este caso, no lineal 55 . ${\Bbb D}$ es el espacio, descrito mediante cierta parametrización, de las medidas observables.

El problema inverso, formulado resumidamente, pretendería **estimar y/o caracterizar** los modelos, que en este caso son modelos de terreno, para los cuales:

$$\mathbf{F}(\mathbf{m}) \simeq \mathbf{d}.\tag{169}$$

Naturalmente, en la búsqueda de tales modelos de terreno, se debe aprovechar al máximo la información que, sobre los mismos, se halle presente en las medidas de campo \mathbf{d} . Una forma operativa de conseguir este objetivo es interpretar el símbolo \simeq , en la ecuación aproximada anterior (169), de forma que la misma pueda reformularse en términos de un problema de optimización. Para tal fin, cabe servirse del concepto matemático de norma (véase el capítulo lineal). Con este concepto, la ecuación (169) se puede reescribir:

hallar
$$\mathbf{m} \in \mathbb{M}$$
: $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{p} < tol,$ (170)

El tipo de norma -determinado al elegir el valor del escalar p- y el umbral de tolerancia -toldependen del modelo matemático asumido para describir los errores en las medidas. Cuando la
estructura estadística de éstos se conoce con precisión cabe sustituir el problema (170) por el
problema de minimización:

hallar
$$\mathbf{m} = \min_{\mathbf{m} \in \mathbb{M}} \|\mathbf{W} (\mathbf{d} - \mathbf{F} (\mathbf{m}))\|_{p},$$
 (171)

donde W representa una matriz de ponderación de los errores. Tal como se ilustra en el capítulo dedicado al problema inverso lineal, el problema (171) podría no tener solución o tener infinitas. Asimismo, dado que en la práctica la muestra de datos suele ser única, las incertidumbres sobre la estructura estadística del error pueden ser demasiado altas como para pretender describirla con exactitud. Esto cuestiona la posibilidad de acertar con la ponderación adecuada en la formulación (171).

Por otro lado, incluso cuando la solución sea única, se ha visto que las perturbaciones aleatorias en los datos pueden transmitirse a la solución confiriéndole gran inestabilidad.

Por todo ello, si bien parece claro que la información aportada por las medidas geofísicas directas de campo debe estar presente y que formulaciones del tipo (170) y (171) permiten

 $^{^{55}}$ **F** es un funcional cuyo valor depende de la posición de medida en superficie (x = AB/2) y del modelo de terreno **m**. La no linealidad se refiere a la dependencia de **F** respecto a **m**.

aprovecharla al máximo, las dificultades anteriormente mencionadas precisan para ser solventadas o, al menos, amortiguadas, de la utilización de información complementaria: la denominada información a priori.

Como se detallará en las páginas sucesivas, cabe incorporar esta información de dos formas:

- Empleando la formulación (171) como planteamiento principal del problema. Como en ella no hay ningún elemento que incorpore **explícitamente** información adicional que no sean los datos **d**, la información a priori se deberá incluir en alguna **parte del algoritmo de optimización empleado**. Por ejemplo, si la resolución del problema de optimización se aborda mediante un método local, se tratará de llegar a una solución final resolviendo una secuencia de problemas aproximados más simples (lineales). La información a priori podría introducirse de forma adecuada al solventar cada uno de dichos pasos internos de manera que, propagándose a través de las sucesivas iteraciones, se consiga que esté presente y afecte a la solución final.
- Cambiando la formulación general del problema (171) por otra que incorpore explícitamente la información a priori. Por ejemplo, en lugar de (171), cabría tratar de resolver el siguiente problema de optimización:

hallar
$$\mathbf{m} = \min_{\mathbf{m} \in \mathbb{M}} \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{p_1} + \lambda^2 \|\mathbf{L}(\mathbf{m} - \mathbf{m}^*)\|_{p_2},$$
 (172)

donde ${\bf L}$ podría, por ejemplo, ser una aproximación discreta de un operador de derivación y λ es un escalar. O bien,

$$\mathbf{m} = \min_{\mathbf{m} \in \mathbb{M}} \|\mathbf{L} (\mathbf{m} - \mathbf{m}^*)\|_{p_2} \text{ tal que } \|\mathbf{d} - \mathbf{F} (\mathbf{m})\|_{p_1} \le tol,$$
 (173)

formulación en la cual el énfasis se pone en lograr el mayor grado posible de adecuación a la información a priori disponible, manteniendo al mismo tiempo cierto nivel prefijado de sintonía con las medidas de campo.

En este capítulo se comienza por tratar el problema (171) viendo sus particularidades. Posteriormente se trabajará sobre las formulaciones (172) y (173), analizando las similitudes y diferencias.

Los valores de p_1 y p_2 que determinan el tipo de norma en el espacio de modelos y de datos se tomarán, en principio, igual a 2. Cuando la norma sea no euclídea se indicará expresamente.

5.2. Construcción de la solución

En el párrafo anterior se ha hecho hincapié en la no linealidad del funcional **F**. En el problema de los sondeos eléctricos verticales no es posible soslayar esta circunstancia. Sin embargo existen algunos problemas⁵⁶ que admiten, mediante una reparametrización del espacio de modelos, del de datos o de ambos, una formulación lineal.

En la figura 38 se ilustra un caso sencillo correspondiente a dicha tipología.

⁵⁶Dichos problemas se denominan intrínsecamente lineales.

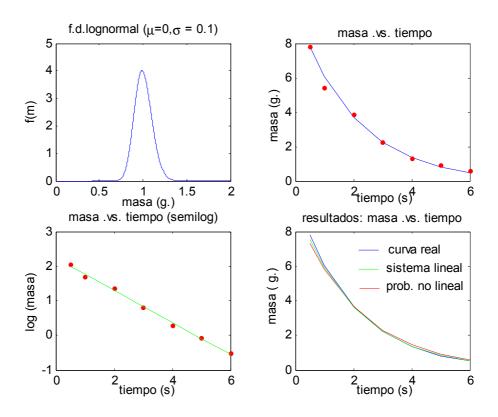


Figura 38: Problema inverso de desintegración radiactiva: comparación entre la solución obtenida por linealización y la del problema no lineal original hallada mediante un método cuasi-Newton.

El fenómeno físico considerado es la desintegración con el tiempo de determinada masa de material radiactivo. La ley que modeliza dicho fenómeno puede representarse en forma de ecuación diferencial ordinaria de primer grado como:

$$\frac{dm}{dt} = -km, k > 0,$$

$$m(t = 0) = m_0.$$
(174)

La solución de la ecuación (174) permite obtener explícitamente la masa restante en función del tiempo transcurrido desde el inicio de la experiencia:

$$m\left(t\right) = m_0 e^{-kt}. (175)$$

En el caso de estudio, la masa de partida, m_0 , es de 10 g. mientras que la constante de desintegración k vale 0.5. Supóngase que esto se ignora y que, a partir de cierto número de observaciones -en este caso 7- de la masa restante tras haber dejado transcurrir tiempos diferentes, se pretende conocer tanto la masa de partida m_0 como la constante k. Las medidas, lógicamente, están afectadas por errores aleatorios. En este ejemplo se supone que estos errores afectan multiplicativamente a dichas mediciones y que su distribución es lognormal con parámetros $\mu = 0$ y $\sigma = 0.1$. En la parte superior izquierda de la figura se representa la función de densidad representativa de la variable aleatoria observada. Las medidas se muestran en el gráfico superior derecho en realción a la curva teórica.

Si se realiza una transformación logarítmica del espacio de datos (en lugar de considerar las masas se consideran los logaritmos de las mismas), la expresión (175) se transforma en:

$$\log m = \log m_0 - kt. \tag{176}$$

Asimismo, el espacio de modelos se reparametriza de forma que cualquier vector (m_0, t) se transforme en $(\log m_0, t)$.

Con estos cambios, la ley (176) es lineal en los nuevos parámetros. Esto puede observarse al representar gráficamente (véase parte inferior izquierda) el nuevo dato $\log m$ frente al tiempo t. La resolución del sistema lineal correspondiente⁵⁷, permite obtener como parámetros:

$$\log m_0 = 2.222, \ k = 0.4631.$$

La calidad del ajuste puede verse observando la recta de color verde. Los parámetros resultantes son:

$$m_{0(l)} = 9.227,$$

 $k_{(l)} = 0.4631.$

Se ha resuelto también el sistema no lineal mediante un método de cuasi-Newton⁵⁸. La solución obtenida sin reparametrizar ha sido,

$$m_{0(nl)} = 9.6374$$

 $k_{(nl)} = 0.4869$

Los resultados son ligeramente diferentes. La repercusión que tiene esta discrepancia puede verse observando la gráfica inferior derecha. Las tres curvas (la real, la procedente de la inversión lineal y la resultante de la inversión no lineal) apenas se diferencian, pero no son idénticas.

Una cuestión notable que justifica el deseo de convertir un problema no lineal en lineal (o, al menos, en un problema aproximado lineal) es la referente a la existencia de solución. En el problema lineal siempre se puede encontrar una solución -en sentido matemático al menos, aun cuando ésta pueda carecer de sentido físico-. Además en el ámbito de los problemas lineales existe una teoría de análisis bien asentada. Sin embargo, cuando **F** es no lineal, la cuestión de la existencia de solución queda teóricamente, en muchas ocasiones, sin resolver. Para muchos problemas no lineales de interés práctico sólo se puede afirmar que hay una solución aceptable cuando, construyéndola, se ha encontrado (Parker, 1994).

Aparte de las cuestiones de existencia y unicidad, es interesante (y a menudo lo más importante) decidir cómo de buena es la solución estimada. Tal capacidad de análisis que sí está desarrollada para los problemas lineales no se ha alcanzado todavía para los no lineales. Cuestiones como, por ejemplo, cuál es el efecto *total* sobre la solución debido a la perturbación de un parámetro del problema o la acción del azar sobre los datos, no pueden resolverse aún de forma satisfactoria en el ámbito no lineal.

⁵⁷ Mediante la pseudoinversa de Moore-Penrose.

⁵⁸ Davidon-Fletcher-Powell.

Una de las modalidades de construcción de solución del problema no lineal consiste en sustituirlo por una secuencia de problemas **lineales aproximados**. La resolución del **problema inverso lineal** en cada paso, es el punto de partida del paso no lineal siguiente. Se pretende que dicha secuencia de puntos converja hacia una solución aceptable del problema original.

Naturalmente, la primera consecuencia destacada que produce la sustitución del problema inicial por una sucesión de *nuevos* problemas lineales, es que **siempre** se podrá construir una solución del problema no lineal. La cuestión clave es que el proceso iterativo lineal puede conducir a resultados finales que **no** sean solución del problema no lineal.

Algunos de estos métodos locales, que se comentarán a continuación, son muy aptos a la hora de encontrar mínimos del problema no lineal. Sin embargo, una vez hallado un mínimo local, estos procedimientos son incapaces de progresar hacia el descubrimiento de algún otro mínimo y, por tanto, pueden permanecer apartados del mínimo global -en caso de que éste exista-. Si se desea encontrar el mínimo global aplicando alguna de estas técnicas locales, deberá establecerse previamente la garantía teórica de que es el único mínimo que existe.

En otro orden de exigencias, la propuesta de estimar un sólo modelo como representación del único terreno capaz de reproducir los datos con aproximación asumible es incompatible con las observaciones realizadas en el capítulo relativo al problema inverso lineal, a propósito del efecto que pequeñísimos errores en las medidas tienen sobre los modelos estimados. Se puede concluir de aquel apartado que modelos **muy distintos** son capaces de producir predicciones indistinguibles dentro del rango de precisión de los aparatos de medida. Por otro lado, el fenónemo de las equivalencias en S.E.V está ampliamente documentado desde Maillet (1947). Por tanto, más que un único modelo del terreno, la realidad impone la existencia de un conjunto de modelos de terreno compatibles con los datos. Si, complementariamente, se hace uso de la información a priori, algunos de los modelos de este conjunto -indistinguibles desde el punto de vista de las medidas de campo- deberán ser descartados si se desvían excesivamente de la evidencia geológica disponible.

En todo caso, casi siempre habrá no uno, sino un conjunto de terrenos solución admisibles a la luz de **todos** los datos. Los métodos de optimización local sólo dan una única solución y, en este sentido, son bastante limitados. Los denominados métodos globales tienen la capacidad de examinar, potencialmente, todo el espacio de modelos. Una vez han localizado un mínimo local, son capaces de salir de él y seguir explorando en busca de otros mínimos locales o del mínimo global, caso de que exista e interese encontrarlo. Permiten, de este modo, obtener una cartografía de mínimos locales, o sea, el equivalente numérico de un mapa de los terrenos no descartables.

Algunos de estos métodos globales gozan de un paralelismo intrínseco en el sentido de que no trabajan con un único modelo -produciendo cambios en él hasta que resulte aceptable- sino, simultáneamente, con una gran cantidad de ellos. Por tanto estas técnicas pueden obtener -a la vez- varias propuestas asumibles de terrenos. Ejemplo de este último caso, tratado en detalle en esta tesis, son los denominados algoritmos genéticos.

No es el único método global al que hace referencia este trabajo. También la técnica denominada "simulated annealing" o "recocido simulado" es una técnica numérica con capacidad de cartografiar mínimos locales y de localizar el mínimo global. Si bien opera con los modelos de uno en uno, se ha programado aquí de manera que trabaje de forma simultánea -aunque independiente- con una familia de modelos.

5.3. Métodos locales de optimización

Una de las estrategias para resolver el problema no lineal,

$$\mathbf{m}_{sol} = \min_{\mathbf{m} \in \mathbb{M}} \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{2}, \tag{177}$$

consiste en sustituir este problema por una secuencia iterativa de minimización de funcionales cuadráticos cuyas soluciones, se espera, converjan a la solución del problema original (177). Supóngase, entonces, que en el paso k-ésimo de dicha secuencia se ha llegado a un modelo \mathbf{m}_k . A partir del mismo, se tratará de determinar cuál es el modelo siguiente \mathbf{m}_{k+1} el cual, en teoría, estará más cerca de la solución \mathbf{m}_{sol} .

El nuevo modelo \mathbf{m}_{k+1} , se puede escribir como:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k,$$

donde \mathbf{p}_k es un vector que indica la dirección en la que se produce la modificación de los parámetros del modelo y α_k es un escalar que condiciona la magnitud de dicha modificación. La elección del valor de α_k va, teóricamente, orientada a producir la mayor disminución en el valor del funcional a minimizar a lo largo de la dirección \mathbf{p}_k .

En lo sucesivo se denotará por e al residuo en las predicciones, es decir:

$$e = d - F(m)$$

La función a minimizar, denominada función objetivo, puede escribirse como⁵⁹:

$$\Phi\left(\mathbf{m}\right) = \|\mathbf{d} - \mathbf{F}\left(\mathbf{m}\right)\|_{2}^{2} = \mathbf{e}^{t}\mathbf{e}.$$
(178)

Normalmente el proceso de resolución del problema (178) se desarrolla en tres pasos:

- 1. Elección del funcional cuadrático $\hat{\Phi}_k(\mathbf{m})$ que aproximará localmente, en un entorno de \mathbf{m}_k , al funcional $\Phi(\mathbf{m})$. Los distintos métodos locales tratados en este trabajo difieren en este punto.
- 2. Cálculo de la dirección \mathbf{p}_k , sobre la cual se producirán los cambios en \mathbf{m}_k . Si $\hat{\Phi}_k(\mathbf{m})$ es una buena aproximación local de $\Phi(\mathbf{m})$ alrededor del punto \mathbf{m}_k , se asume que una buena forma de hacer descender el valor de $\Phi(\mathbf{m})$ es trasladarse al modelo \mathbf{m}_c donde la aproximación cuadrática toma su valor mínimo. Esto permite calcular \mathbf{p}_k como:

$$\mathbf{p}_k = \mathbf{m}_c - \mathbf{m}_k$$

3. Determinación de la magnitud del desplazamiento, α_k , a lo largo de \mathbf{p}_k . Si bien en teoría α_k debería escogerse de forma que en el punto $\mathbf{m}_k + \alpha_k \mathbf{p}_k$, la función objetivo $\Phi\left(\mathbf{m}\right)$ restringida sobre dicha dirección, alcanzase su mínimo, imponer esta condición es, en cierto sentido, un derroche de esfuerzo. Normalmente basta con garantizar que, a lo largo de dicha línea, se produce un descenso suficiente de $\Phi\left(\mathbf{m}_k + \alpha_k \mathbf{p}_k\right)$.

A continuación se exponen las características de los métodos de optimización local empleados en la tesis.

 $[\]overline{^{59}}$ El modelo **m** que minimiza $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_2$, dado que ésta es una cantidad siempre positiva, es el mismo que el que resulta de minimizar $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_2^2$. No obstante, como es natural, los valores de las funciones en el mínimo difieren.

5.3.1. El metodo de Newton

Sea \mathbf{m}_k la posición de inicio para el k-ésimo paso del proceso iterativo. En un entorno $\boldsymbol{\delta}(\mathbf{m}_k)$ de \mathbf{m}_k se realiza un desarrollo de Taylor de orden 2 del funcional $\Phi(\mathbf{m})$ a minimizar, verificándose:

$$\Phi\left(\mathbf{m}\right) = \Phi\left(\mathbf{m}_{k}\right) + \langle \nabla\Phi\left(\mathbf{m}_{k}\right), \mathbf{m} - \mathbf{m}_{k} \rangle + \frac{1}{2} \langle \mathbf{m} - \mathbf{m}_{k}, H\Phi\left(\mathbf{m}_{k}\right) (\mathbf{m} - \mathbf{m}_{k}) \rangle + o\left(\|\Delta\mathbf{m}\|_{2}^{2}\right).$$
(179) para todo $\mathbf{m} \in \boldsymbol{\delta}\left(\mathbf{m}_{k}\right).$

Si $\delta(\mathbf{m}_k)$ es lo suficientemente pequeño como para que $o\left(\|\Delta\mathbf{m}\|_2^2\right)$ pueda, efectivamente, despreciarse, y se define

$$\hat{\Phi}_{k}\left(\mathbf{m}\right) = \Phi\left(\mathbf{m}_{k}\right) + \nabla_{k}^{t}\Phi\left(\mathbf{m} - \mathbf{m}_{k}\right) + \frac{1}{2}\left(\mathbf{m} - \mathbf{m}_{k}\right)^{t}H\Phi_{k}\left(\mathbf{m} - \mathbf{m}_{k}\right)$$

en todo $\mathbf{m} \in \boldsymbol{\delta}(\mathbf{m}_k)$, es lícito asumir la aproximación

$$\Phi\left(\mathbf{m}\right) \simeq \hat{\Phi}_{k}\left(\mathbf{m}\right)$$
.

Si $\mathbf{m} = \mathbf{m}_c$ es el mínimo de $\hat{\Phi}_k(\mathbf{m})$, en él se cumple, necesariamente, la condición:

$$\nabla \hat{\Phi}_k \left(\mathbf{m}_c \right) = \mathbf{0},$$

es decir,

$$H\Phi_k\left(\mathbf{m}_c - \mathbf{m}_k\right) = -\nabla_k^t \Phi. \tag{180}$$

La expresión (180) muestra las denominadas ecuaciones normales del método de Newton, que propone que el nuevo modelo \mathbf{m}_{k+1} se busque en la dirección $\mathbf{p}_k = -H\Phi_k^{-1} \cdot \nabla_k^t \Phi$, dada por la solución de las citadas ecuaciones.

Para el caso particular -que es el considerado en este trabajo- en que $\Phi(\mathbf{m})$ sea una suma de cuadrados (véase la expresión

(178)) pueden obtenerse expresiones explícitas para el vector gradiente:

$$\nabla_k \Phi = 2 \mathbf{J}_{\mathbf{e}k}^t \cdot \mathbf{e},$$

y para la matriz Hessiana,

$$H\Phi_{k} = 2\mathbf{J}_{\mathbf{e}k}^{t}\mathbf{J}_{\mathbf{e}k} + 2\sum_{i=1}^{i=n} e_{i}He_{i}\left(\mathbf{m}_{k}\right),$$

donde J_{ek} es la matriz jacobiana de los residuos, que tiene por expresión:

$$\mathbf{J}_{\mathbf{e}k} \equiv \left[egin{array}{c} \mathbf{\nabla}^t e_1 \left(\mathbf{m}_k
ight) \\ dots \\ \mathbf{\nabla}^t e_s \left(\mathbf{m}_k
ight) \end{array}
ight],$$

y $He_i(\mathbf{m}_k)$ son las matrices hessianas en el punto \mathbf{m}_k de cada una de las componentes del vector de residuos⁶⁰. De este modo, el desarrollo de Taylor (179) se escribe como:

$$\hat{\Phi}_k(\mathbf{m}) = \Phi(\mathbf{m}_k) + 2\mathbf{J}_{\mathbf{e}k}^t \mathbf{e} \left(\mathbf{m} - \mathbf{m}_k\right) + \frac{1}{2} \left(\mathbf{m} - \mathbf{m}_k\right)^t \left(2\mathbf{J}_{\mathbf{e}k}^t \mathbf{J}_{\mathbf{e}k} + 2\sum_{i=1}^{i=n} e_i H_k e_i\right) \left(\mathbf{m} - \mathbf{m}_k\right). \tag{181}$$

⁶⁰ Por simplicidad denotaremos como $H_k e_i$ a $H e_i$ (\mathbf{m}_k).

donde, se recuerda que $\mathbf{e} = \mathbf{d} - \mathbf{F}(\mathbf{m})$ es el vector de residuos no lineales.

Esto hace que las ecuaciones normales (180), empleadas para el cálculo de la dirección de búsqueda \mathbf{p}_k adquieran, en este método, la forma

$$\left(\mathbf{J}_{\mathbf{e}k}^{t}\mathbf{J}_{\mathbf{e}k} + \sum_{i=1}^{i=n} e_{i}H_{k}e_{i}\right)\mathbf{p}_{k} = -\mathbf{J}_{\mathbf{e}k}^{t}\mathbf{e}$$
(182)

El método de Newton tiene, en ciertas condiciones, unas ventajas grandes sobre otras técnicas. Por ejemplo, si la matriz Hessiana $H_k\left(\Phi^2\right)$ fuese estrictamente definida positiva, entonces

$$\langle \mathbf{p}_k, \mathbf{\nabla}_k \Phi \rangle = -\mathbf{p}_k^t H \Phi_k^{-1} \mathbf{p}_k < 0,$$

lo que significa que en la dirección \mathbf{p}_k es posible el descenso. En estas condiciones, el ratio de convergencia es 2. Por construcción, se ve que el método encuentra el mínimo de una función cuadrática en un solo paso. Aunque, en general, necesita iterar varias veces para hallar el mínimo de una función no cuadrática, el método mejora su rendimiento en las proximidades de la solución dado que "suficientemente cerca" del mínimo cualquier función se "parece" a una cuadrática.

El método, tal como ha sido expuesto, presenta también serios inconvenientes que limitan su uso en esta forma:

- es necesario disponer de la matriz Hessiana. En general, el volumen de cálculo que esto implica aconseja decidirse por métodos que no precisen de dicha matriz o que la aproximen de alguna forma más económica;
- aún disponiendo de $H\Phi_k$, hay pocas funciones para las cuales pueda garantizarse que $H\Phi_k$ es siempre definida positiva;
- el paso \mathbf{p}_k para alcanzar el mínimo de $\hat{\Phi}_k$ (\mathbf{m}) puede estar tan alejado de \mathbf{m}_k que, a pesar de que $\hat{\Phi}_k$ (\mathbf{m}) haya descendido, la función objetivo no lineal Φ (\mathbf{m}) puede haber aumentado. Incluso cabe la posibilidad de que Φ (\mathbf{m}) no esté definida en $\mathbf{m}_k + \mathbf{p}_k$. En estos casos, lo que suele hacerse es utilizar alguna estrategia para reducir el paso localizando un punto intermedio $\mathbf{m}_{k+1} = \mathbf{m}_k + \alpha_k \mathbf{p}_k$ en el que, aunque $\hat{\Phi}_k$ (\mathbf{m}) ya no tome su valor mínimo, se cumpla que

$$\Phi\left(\mathbf{m}_{k}+\alpha_{k}\mathbf{p}_{k}\right)<\Phi\left(\mathbf{m}_{k}\right);$$

- El método da problemas cuando $H\Phi_k$ es singular. En este caso, el sistema (180) tiene infinitas soluciones o ninguna. Incluso cuando la matriz está cerca de la singularidad, de forma que $\hat{\Phi}_k(\mathbf{m})$ presente "largos valles" en alguna dirección, será un método con comportamiento muy veleidoso frente a pequeñas perturbaciones.
- Cuando \mathbf{m}_k es un punto crítico, es decir, $\nabla \hat{\Phi}_k = \mathbf{0}$, pero correspondiente a un punto de silla, la solución es $\mathbf{p}_k = \mathbf{0}$ que, naturalmente, no informa nada respecto a cómo seguir iterando.

5.3.2. El método de Gauss-Newton

Las propiedades de rápida convergencia que muestra el método de Newton, cuando funciona, motiva que se haya conservado como sustrato básico al que se han añadido modificaciones para eludir sus inconvenientes. Una primera modificación se emplea para conseguir aproximar la matriz Hessiana por otra que sea definida (o semidefinida) positiva siempre. La aproximación aludida es

$$\left(\mathbf{J}_{\mathbf{e}k}^{t}\mathbf{J}_{\mathbf{e}k} + \sum_{i=1}^{i=n} e_{i}H_{k}e_{i}\right) \simeq \mathbf{J}_{\mathbf{e}k}^{t}\mathbf{J}_{\mathbf{e}k}$$

con lo cual, la matriz Hessiana $H\Phi_k$ es

$$H\Phi_k = \mathbf{J}_{\mathbf{e}k}^t \mathbf{J}_{\mathbf{e}k}.$$

Uno de los méritos de esta expresión es que ofrece una aproximación de la matriz Hessiana que puede calcularse *conociendo sólo las derivadas primeras*. En este caso, el método de Newton recibe el nombre de **Gauss-Newton** y las ecuaciones normales (182) se escriben en este caso:

$$\left(\mathbf{J}_{\mathbf{e}k}^{t}\mathbf{J}_{\mathbf{e}k}\right)\mathbf{p}_{k} = -\mathbf{J}_{\mathbf{e}k}^{t}\mathbf{e}.\tag{183}$$

Se ha asumido, para "convertir" el método de **Newton** en el de **Gauss-Newton**, que la parte de la matriz Hessiana representada por

$$\Theta\left[\mathbf{e}\right] = \sum_{i=1}^{i=n} e_i H_k e_i,$$

puede suponerse nula. Esto será así cuando, por ejemplo, los residuos e_i sean "bastante" reducidos -esta hipótesis se denomina de los pequeños residuos- cosa que suele suceder cerca del mínimo si el desajuste no modelizable no es excesivamente grande. En tal circunstancia, el comportamiento del método de Gauss-Newton se parecerá al del método de Newton cerca de la solución. También cabe la posibilidad de que las distintas e_i , a pesar de ser relativamente grandes, se compensen en la expresión de Θ [e].

El método Gauss-Newton presenta también problemas de inestabilidad cuando $\mathbf{J}_{\mathbf{e}k}$ es casi singular pues $\mathbf{J}_{\mathbf{e}k}^t \mathbf{J}_{\mathbf{e}k}$ estará aún peor condicionada⁶¹. Si $\mathbf{J}_{\mathbf{e}k}$ es deficiente en rango, $\mathbf{J}_{\mathbf{e}k}^t \mathbf{J}_{\mathbf{e}k}$ es singular yy el sistema vuelve a ser, o incompatible o a poseedor de infinitas soluciones.

5.3.3. El método de gradiente o del descenso más rápido (steepest-descent)

Como su propio nombre indica, asume que la dirección propuesta es la que permite el más rápido descenso de la función objetivo. Esta dirección es, como se sabe, la indicada por el vector gradiente y su sentido opuesto. Así pues,

$$\mathbf{p}_{k} = -\frac{1}{\mu_{k}} \nabla \Phi \left(\mathbf{m}_{k} \right), \tag{184}$$

donde μ_k es un escalar positivo.

 $^{^{61}}$ Su número de condición es el cuadrado del de la matriz $\mathbf{J_{e}}_{k}$.

En los términos propuestos en el epígrafe anterior, ¿qué función cuadrática es la que, en este caso, se procura minimizar?. Obsérvese que la expresión (184) es la solución del sistema de ecuaciones normales

$$(\mu_k \mathbf{I}) \mathbf{p}_k = -\mathbf{J}_{\mathbf{e}k}^t \mathbf{e},$$

que indica, por analogía con (180) y (182), que la matriz Hessiana en este caso tiene por expresión

$$H\Phi_k = 2\mu_k \mathbf{I},$$

es decir, es proporcional a la matriz unidad.

Al igual que en el caso de los métodos de Newton y Gauss-Newton, el paso \mathbf{p}_k que resuelve las ecuaciones normales aquí pudiese ser demasiado grande y no suponer un descenso del funcional no lineal. Entonces, reduciendo suficientemente el valor de μ_k se garantiza que

$$\Phi\left(\mathbf{m}_{k}+\mathbf{p}_{k}\right)<\Phi\left(\mathbf{m}_{k}\right).$$

Si se emplea búsqueda lineal exacta, el método converge siempre a un punto estacionario. Si dicho punto es un mínimo estricto, la convergencia es lineal y el factor es:

$$k = \left(\frac{\sigma_{\text{máx}} - \sigma_{\text{mín}}}{\sigma_{\text{máx}} + \sigma_{\text{mín}}}\right)^2,$$

donde $\sigma_{\text{máx}}$ y $\sigma_{\text{mín}}$ representan respectivamente los valores singulares máximo y mínimo de la matriz Hessiana (Scales, 1987). Sólo cuando k=1, la convergencia es superlineal. Cuando la "forma" del valle en el que está el mínimo es problemática, la velocidad de convergencia se deteriora enormemente.

5.3.4. El método de Levenberg-Marquardt

Entre el método de Gauss-Newton, rápido pero que falla cuando la matriz $H\Phi_k$ es singular o casi y el método de descenso más rápido, para el que nunca la matriz $H\Phi_k$ es singular (ni casi singular)⁶², pero que es lentísimo justamente en las circunstancias que trata de superar ($H\Phi_k$ casi singular), se sitúa el método de Levenberg-Marquardt (abreviadamente L-M) que combina las ventajas de los dos anteriores.

Con la técnica (L-M) la matriz Hessiana $H\Phi_k$ se aproxima mediante la expresión:

$$H\Phi_k = \mathbf{J}_{\mathbf{e},k}^t \mathbf{J}_{\mathbf{e},k} + \mu_k \mathbf{I}.$$

Las ecuaciones normales son en este caso:

$$\left(\mathbf{J}_{\mathbf{e},k}^{t}\mathbf{J}_{\mathbf{e},k}+\mu_{k}\mathbf{I}\right)\mathbf{p}_{k}=-\mathbf{J}_{\mathbf{e},k}^{t}\mathbf{e}.\tag{185}$$

Cuando $\mu_k \to 0$, el método tiende a generar el paso de G-N con todos sus problemas de inestabilidad en caso de que $\mathbf{J}_{\mathbf{e},k}^t \mathbf{J}_{\mathbf{e},k}$ sea singular (o casi singular). A medida que μ_k crece, todos los valores singulares de $\left(\mathbf{J}_{\mathbf{e},k}^t \mathbf{J}_{\mathbf{e},k} + \mu_k \mathbf{I}\right)$ crecen en la misma medida convirtiéndose los valores nulos (caso de $\mathbf{J}_{\mathbf{e},k}^t \mathbf{J}_{\mathbf{e},k}$ singular) en positivos y aumentando los valores singulares pequeños (causa

 $^{^{62}}$ A menos que μ_k sea muy próximo a cero.

de inestabilidad en problemas con alto número de condición). Todo esto tiende a estabilizar las soluciones de las ecuaciones (185). Cuando $\mu_k \to \infty$, el método se parece al de descenso más rápido y el tamaño de paso tiende a hacerse nulo⁶³.

Como se desprende de estos comentarios, se trata de tomar el menor valor de μ_k capaz de conseguir un avance estable. Aparentemente, para esto han de resolverse las ecuaciones normales para toda una familia de parámetros μ_k distintos y, en función de los resultados, determinar -a posteriori- qué μ_k es el óptimo. Naturalmente esto acaba con cualquier ventaja que pudiera ofrecer el método. No obstante, se han propuesto distintas metodologías para escoger el parámetro μ_k en cada iteración, al comienzo de la misma, sin tener que resolver cada vez las ecuaciones normales.

Como se ha visto en el capítulo 4 a propósito de la regularización de Tikhonov, las ecuaciones normales (185) del método L-M son las que se obtienen al resolver el siguiente problema de mínimos cuadrados:

$$\Delta \mathbf{m}_{LM} = \min_{\Delta \mathbf{m}_k \in \delta(\mathbf{0})} \left\| \mathbf{J} \mathbf{F}_k \Delta \mathbf{m} - \mathbf{d}_k \right\|_2^2 + \mu_k^2 \left\| \Delta \mathbf{m} \right\|_2^2.$$

Esto explica cómo un aumento de μ_k reduce el tamaño del paso y aumenta la estabilidad de la solución del sistema (185). Al igual que en los casos anteriores, es posible que deba realizarse una búsqueda unidimensional para encontrar un descenso razonable del funcional no lineal en la dirección \mathbf{p}_k propuesta.

5.3.5. Linealización del funcional F

Tradicionalmente se ha llegado a plantear el funcional cuadrático local efectuando una linealización sobre el funcional \mathbf{F} . Básicamente, la mencionada linealización sustituye, en las cercanías de un modelo \mathbf{m}_k , el funcional \mathbf{F} por una aproximación lineal suya.

Si en el punto \mathbf{m}_k el funcional no lineal $\mathbf{F}(\cdot)$ es diferenciable con continuidad, existe un entorno de \mathbf{m}_k tal que, para todo modelo \mathbf{m} que pertenezca a dicho entorno se puede escribir:

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_k) + \mathbf{dF}_{\mathbf{m}_k}(\Delta \mathbf{m}) + \mathbf{o}(\|\Delta \mathbf{m}\|), \qquad (186)$$

donde $\mathbf{dF}_{\mathbf{m}_k}(\Delta \mathbf{m})$ es la aplicación diferencial en \mathbf{m}_k . La igualdad anterior tiene por expresión en las bases canónicas de \mathbb{M} y \mathbb{D} :

$$\mathbf{F}(\mathbf{m}) = \mathbf{F}(\mathbf{m}_k) + \mathbf{J}\mathbf{F}_k(\mathbf{m} - \mathbf{m}_k) + \mathbf{o}(\|\Delta \mathbf{m}_k\|), \qquad (187)$$

donde

$$\mathbf{JF}_{k} \equiv \left[\begin{array}{c} \mathbf{\nabla}^{t} F_{1}\left(\mathbf{m}_{k}\right) \\ \vdots \\ \mathbf{\nabla}^{t} F_{s}\left(\mathbf{m}_{k}\right) \end{array} \right],$$

recibe el nombre de matriz jacobiana de \mathbf{F} en \mathbf{m}_k y, como se sabe tras el análisis efectuado en el capítulo lineal, juega un papel fundamental tanto para hallar la solución del problema como para analizar sus características.

⁶³ Puede verse dividiendo ambos miembros de (185) por μ_k y tomando límites cuando $\mu_k \to \infty$.

Según lo indicado por la expresión (187) si \mathbf{m}_{k+1} está tan cerca de \mathbf{m}_k como para que $\mathbf{o}(\|\Delta\mathbf{m}_k\|)$ se pueda despreciar, se puede sustituir $\mathbf{F}(\mathbf{m})$ en el problema (177) por su aproximación lineal (187) obteniéndose \mathbf{m}_{k+1} como:

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_2 = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \|\mathbf{d} - \mathbf{F}(\mathbf{m}_k) - \mathbf{J}\mathbf{F}_k (\mathbf{m} - \mathbf{m}_k)\|_2.$$
(188)

Si se denota por $\mathbf{d}_k = \mathbf{d} - \mathbf{F}(\mathbf{m}_k)$ el error de predicción no lineal correspondiente al modelo \mathbf{m}_k y por

$$\hat{\mathbf{d}}_k = \mathbf{d} - \mathbf{F}(\mathbf{m}_k) + \mathbf{J}\mathbf{F}_k \cdot \mathbf{m}_k$$

se puede formular el cálculo del modelo \mathbf{m}_{k+1} a partir del \mathbf{m}_k de dos formas distintas:

$$\Delta \mathbf{m}_{k} = \min_{\Delta \mathbf{m}_{k} \in \boldsymbol{\delta}(\mathbf{0})} \left\| \mathbf{J} \mathbf{F}_{k} \Delta \mathbf{m} - \mathbf{d}_{k} \right\|_{2},$$
$$\mathbf{m}_{k+1} = \mathbf{m}_{k} + \Delta \mathbf{m}_{k},$$

donde se ha tomado como incógnita $\Delta \mathbf{m} = \mathbf{m} - \mathbf{m}_k$. En este caso, el funcional cuadrático de aproximación local $\hat{\Phi}_k$ tiene por expresión:

$$\hat{\Phi}_k(\Delta \mathbf{m}) = \|\mathbf{J}\mathbf{F}_k\Delta \mathbf{m} - \mathbf{d}_k\|_2^2$$

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \left\| \mathbf{J} \mathbf{F}_k \mathbf{m} - \hat{\mathbf{d}}_k \right\|_p, \tag{189}$$

en el que la incógnita es **m**. El funcional cuadrático $\hat{\Phi}_k$ es ahora,

$$\hat{\Phi}_k\left(\mathbf{m}
ight) = \left\|\mathbf{J}\mathbf{F}_k\mathbf{m} - \hat{\mathbf{d}}_k
ight\|_2^2.$$

En ambos casos cabe observar que el modelo resultante minimiza una buena aproximación local de la norma del desajuste. Sin embargo, si no se toman más precauciones, no se puede afirmar que el citado modelo pertenezca a un entorno local de \mathbf{m}_k . De hecho puede estár tan alejado, que el parecido (asumido cuando se dice que la aproximación linealizada es buena) entre $\hat{\Phi}_k(\cdot)$ y $\Phi(\cdot)$ sea escaso o inexistente. Es posible para ejemplificar la situación, en que $\hat{\Phi}_k(\mathbf{m}_{k+1})$ esté definido y, sin embargo, $\Phi(\mathbf{m}_{k+1})$ no lo ésté porque \mathbf{m}_{k+1} caiga fuera del conjunto de modelos admisibles \mathbb{M} . Cuando esto no ocurre es factible que, aunque \mathbf{m}_{k+1} haya conseguido reducir $\hat{\Phi}_k(\cdot)$ a su valor mínimo, no haya disminuido el valor de $\Phi(\cdot)$, que es lo que importa. La regularidad local asumida para el funciónal $\Phi(\cdot)$ en el entorno de \mathbf{m}_k permite, no obstante, reconducir estas circunstancias adversas garantizando la existencia de un punto

$$\mathbf{m}_{\alpha_k} = \mathbf{m}_k + \alpha_k \left(\mathbf{m}_{k+1} - \mathbf{m}_k \right), \ \alpha_k \in \mathbb{R}$$

para el cual se cumple que:

$$\Phi\left(\mathbf{m}_{\alpha_{k}}\right) < \Phi\left(\mathbf{m}_{k}\right)$$
.

5.3.6. Métodos de cuasi-Newton

Se comentan aquí brevemente por mor de completitud. Se trata de una clase de algoritmos basados en el método de Newton, pero que evitan el cálculo de las derivadas segundas de la matriz hessiana. En cada iteración, se utiliza una aproximación de esta matriz para cuyo cálculo sólo se necesitan los gradientes, de forma que se garantice que siempre sea definida positiva. Quizás el más conocido es el denominado BFGS (Broyden-Fletcher-Goldfarb-Shanno) aunque también es frecuente utilizar otros como DFP (Davidon-Fletcher-Powell).

5.4. Incrementos .vs. parámetros (creeping .vs. jumping)

La regularización es, como se ha visto, un proceso necesario para conseguir tener cierto grado de control sobre las características de la solución del problema inverso, así como para amortiguar la dispersión introducida debido a los errores que el azar produce en las medidas y a los que se generan, de forma inevitable, en el propio proceso numérico. El efecto que la regularización produce sobre la solución final, depende de cómo se introduzca en el problema. Se contemplan aquí dos posibilidades:

1. La formulación del problema inverso no lineal indicada en (177) no incluye aparentemente el uso de información a priori. Sin embargo, si se aborda su resolución mediante un algoritmo de optimización local (p.ej, cualquiera de los anteriormente mencionados), la citada información a priori puede introducirse en cada uno de los pasos iterativos lineales sirviendo, además, para regularizar la solución obtenida en cada etapa.

Cómo afecta la regularización al modelo actualizado \mathbf{m}_{k+1} obtenido a partir del \mathbf{m}_k , en el paso iterativo k-ésimo, depende:

- del tipo de método de regularización empleado (pseudoinversa, Tikhonov, truncamiento, principio de discrepancia, etc...),
- de la elección de la incógnita del problema.

Si lo que se busca en el paso k-ésimo es la modificación del modelo, o sea, $\Delta \mathbf{m}_k$, el problema es, como se sabe:

$$\Delta \mathbf{m}_{k} = \min_{\Delta \mathbf{m}_{k} \in \delta(\mathbf{0})} \| \mathbf{J} \mathbf{F}_{k} \Delta \mathbf{m} - \mathbf{d}_{k} \|_{2}, \mathbf{m}_{k+1} = \mathbf{m}_{k} + \Delta \mathbf{m}_{k},$$
(190)

donde $\mathbf{d}_k = \mathbf{d} - \mathbf{F}(\mathbf{m}_k)$. Esta modalidad se denomina "creeping" o formulación en incrementos o perturbaciones. Si la incógnita es el *propio modelo* actualizado, \mathbf{m}_{k+1} , la formulación es:

$$\mathbf{m}_{k+1} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \left\| \mathbf{J} \mathbf{F}_k \mathbf{m} - \hat{\mathbf{d}}_k \right\|_2, \tag{191}$$

donde $\hat{\mathbf{d}}_k = \mathbf{d} - \mathbf{F}(\mathbf{m}_k) + \mathbf{J}\mathbf{F}_k\mathbf{m}_k$. El nombre que recibe esta alternativa es "jumping" o formulación en parámetros. Los ejemplos y explicaciones aportarán justificación a tales nombres.

Uno y otro ofrecen soluciones diferentes, como se ilustrará a continuación, cuando se utiliza el mismo método de regularización.

2. La información a priori puede introducirse explícitamente en la formulación del problema no lineal, tal y como se ha sugerido en el apartado 1 de este capítulo. El efecto regularizador es, entonces, global ya que la información sobre el modelo está en el propio funcional no lineal a minimizar y éste está definido sobre toda parte del conjunto M. En este caso, la información afecta al propio modelo (sin perjuicio de que pueda optarse por un método local que, en cada etapa iterativa, tenga como incógnita el incremento del modelo).

Se desarrollarán en detalle las ecuaciones que resultan del proceso de linealización -es decir, se presentará la formulación del problema resultante en cada paso lineal- para cada caso de métodos iterativos locales.

Adicionalmente, la idea de incluir la información a priori -y, por tanto, sus efectos estabilizadoresen la propia función objetivo, de alguna manera la "define en" y la "transporta por" todas las partes del espacio de modelos M. Esto quiere decir que la misma está preparada para ser utilizada por métodos de optimización global (algoritmos genéticos o simulated annealing, en este caso). Estos métodos no están condicionados por problemas de continuidad y/o derivabilidad de la función objetivo. Esta circunstancia abre la puerta al empleo, de manera natural, de otras normas diferentes en la cuantificación tanto de la parte de la función objetivo que mide la discrepancia con los datos como de la que mide el grado de cumplimiento de la información a priori. Especial interés reviste el empleo de la norma p=1.

Por otro lado y ya en particular para los algoritmos genéticos⁶⁴, dado que no funcionan con un único modelo sino con una familia de ellos (generalmente muy numerosa), la propiedades de la regularización afectarían simultáneamente a todos los miembros de la familia.

5.4.1. Métodos locales con regularización "interna"

Considérese el caso descrito en el primero de los apartados anteriores (se utiliza un método iterativo local). Se analizará lo que ocurre en el paso k-ésimo cuando, para hallar \mathbf{m}_{k+1} a partir de \mathbf{m}_k se elige entre la formulación (190) o la (191).

1. Caso de sistema lineal determinado o sobredeterminado

En caso de que la formulación sea en incrementos -expresión (190)-, las ecuaciones normales del problema son:

$$\left(\mathbf{J}\mathbf{F}_{k}^{t}\mathbf{J}\mathbf{F}_{k}\right)\Delta\mathbf{m} = \mathbf{J}\mathbf{F}_{k}^{t}\mathbf{d}_{k},\tag{192}$$

mientras que, si la incógnita es \mathbf{m}_{k+1} , las ecuaciones normales son

$$\left(\mathbf{J}\mathbf{F}_{k}^{t}\mathbf{J}\mathbf{F}_{k}\right)\mathbf{m}_{k+1} = \mathbf{J}\mathbf{F}_{k}^{t}\hat{\mathbf{d}}_{k}.$$
(193)

Como $\hat{\mathbf{d}}_k = \mathbf{d}_k + \mathbf{J}\mathbf{F}_k \cdot \mathbf{m}_k$, sustituyendo en (193) y reordenando, se obtiene:

$$\left(\mathbf{J}\mathbf{F}_{k}^{\mathbf{t}}\mathbf{J}\mathbf{F}_{k}\right)\mathbf{m}_{k+1}-\left(\mathbf{J}\mathbf{F}_{k}^{\mathbf{t}}\mathbf{J}\mathbf{F}_{k}\right)\mathbf{m}_{k}=\mathbf{J}\mathbf{F}_{k}^{\mathbf{t}}\mathbf{d}_{k},$$

expresión que coincide con la (192) -caso de creeping-. Ambas formulaciones son, por lo tanto equivalentes aunque la solución del sistema (véase cap.4) puede ser inestable. Para solventar

 $^{^{64}}$ El método de simulated annealing, en la medida en que sea paralelizable, podría conseguir, en cierta manera, resultados comparables.

esta inestabilidad debería aplicarse algún tipo de regularización y cabe preguntarse si, al hacer esto, aparecen diferencias. Por el momento se aplaza brevemente la cuestión.

2. Caso subdeterminado o mixto

En este caso, al menos una parte del problema es subdeterminada. Esto significa que los datos geofísicos "duros" no son suficientes para determinar todas las componentes del modelo. Debe regularizarse el problema lineal tanto para fijar la parte no determinada, como para amortiguar la variabilidad de la solución.

Supóngase que se pretende resolver los sistemas (190) y (191) mediante el uso de la pseudoinversa. Este método anula las componentes de la solución en el núcleo de \mathbf{JF}_k . El matiz es que en el primer caso la solución es $\Delta \mathbf{m}$ mientras que en el segundo es, directamente, \mathbf{m}_{k+1} . O sea, si la incógnita es el propio modelo \mathbf{m}_{k+1} , es éste el que no tiene componentes en el núcleo de \mathbf{JF}_k . En cambio, en la formulación en incrementos, el hecho de que $\Delta \mathbf{m}$ no tenga componentes en dicho núcleo, no garantiza que a \mathbf{m}_{k+1} le ocurra lo mismo. Obsérvese que,

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta \mathbf{m},$$

de modo que, si \mathbf{m}_k -el modelo de partida- tiene componentes del núcleo, éstas se conservan en \mathbf{m}_{k+1} . ¿Qué traducción tiene esto en términos más tangibles?. Como se ha expuesto en el capítulo lineal, anular las componentes de un modelo -o de un incremento- en el núcleo de un operador es una forma de hacer mínima su norma euclídea. De hecho, resolver cualquiera de los dos problemas anteriores mediante la técnica de la pseudoinversa es equivalente a resolver,

$$\begin{aligned}
& \min \|\Delta \mathbf{m}\|_{2} \text{ sujeto a} \\
& \min_{\Delta \mathbf{m} \in \boldsymbol{\delta}(\mathbf{0})} \|\mathbf{J} \mathbf{F}_{k} \Delta \mathbf{m} - \mathbf{d}_{k}\|_{2}, \\
& \mathbf{m}_{k+1} = \mathbf{m}_{k} + \Delta \mathbf{m},
\end{aligned} \tag{194}$$

en el caso de la formulación "creeping" (190) o,

$$\min_{\mathbf{m} \in \boldsymbol{\delta}(\mathbf{m}_k)} \| \mathbf{J} \mathbf{F}_k \mathbf{m} - \hat{\mathbf{d}}_k \|_2, \tag{195}$$

si se ha optado por el método "jumping" (191).

En el caso del algoritmo jumping, si se busca un modelo de subsuelo de norma mínima, se busca un terreno que, de entre todos los posibles, sea el más conductivo (resistividades más bajas) y el que tenga el sustrato más superficial. Sin embargo, en el caso del algoritmo creeping, una modificación de modelos ($\Delta \mathbf{m}$) de norma mínima implica que se desea cambiar muy poco las resistividades y espesores del terreno ya existente. Pero -traduciendo lo expresado matemáticamente antes- dado que éste puede tener grandes resistividades y espesores, una pequeña modificación del mismo no lo convierte, lógicamente, en un subsuelo dráticamente diferente con bajas resistividades y espesores. Parece, por tanto, que tratar de minimizar la discrepancia con la información a priori en la formulación en incrementos -en este caso intentar rebajar la norma de $\Delta \mathbf{m}$ -, no garantiza que los modelos obtenidos sean cada vez más compatibles con dicha información.

• Formulación matemática

La solución de mínimos cuadrados (a través de la pseudoinversa) del problema (194) - formulado en términos de "creeping"- es

$$\mathbf{m}_{k+1}^{creep} = \mathbf{m}_k + \mathbf{J} \mathbf{F}_k^{\dagger} \left[\mathbf{d} - \mathbf{F} \left(\mathbf{m}_k \right) \right], \tag{196}$$

mientras que la solución del problema (195) es

$$\mathbf{m}_{k+1}^{jump} = \mathbf{J}\mathbf{F}_{k}^{\dagger} \left(\mathbf{d} - \mathbf{F}\left(\mathbf{m}_{k}\right) + \mathbf{J}\mathbf{F}_{k}\mathbf{m}_{k}\right). \tag{197}$$

 \mathbf{m}_k puede descomponerse en suma de dos vectores perpendiculares (véase descomposición en valores singulares),

$$\mathbf{m}_k = \mathbf{m}_k^{mc} + \mathbf{m}_k^0 = \left(\mathbf{J}\mathbf{F}_k^\dagger\mathbf{J}\mathbf{F}_k
ight)\mathbf{m}_k + \mathbf{m}_k^0,$$

donde \mathbf{m}_k^{mc} es la componente que \mathbf{m}_k tiene en la dirección perpendicular al núcleo de \mathbf{JF}_k y \mathbf{m}_k^0 es representa proyección de \mathbf{m}_k sobre dicho núcleo. Sustituyendo en (196) y operando,

$$\mathbf{m}_{k+1}^{creep} = \mathbf{J}\mathbf{F}_{k}^{\dagger}\left(\mathbf{d} - \mathbf{F}\left(\mathbf{m}_{k}\right) + \mathbf{J}\mathbf{F}_{k}\mathbf{m}_{k}\right) + \mathbf{m}_{k}^{0}.$$

pero, teniendo en cuenta (197)

$$\mathbf{m}_{k+1}^{creep} = \mathbf{m}_{k+1}^{jump} + \mathbf{m}_{k}^{0},$$

es decir, la solución "creeping" es la solución "jumping" a la que se añade las componentes de \mathbf{m}_k que están sobre el núcleo de \mathbf{JF}_k .

Ilustración

El resultado obtenido puede visualizarse en la figura 39 donde se representan las soluciones "creeping" y "jumping" en un caso sencillo. La función a minimizar en este caso es un cilindro parabólico cuyo eje es la recta x=0. Cuando el problema se plantea en incrementos, desde cualquier punto de inicio, el desplazamiento de menor longitud (norma mínima) que lleva a la función a alcanzar el mínimo es el segmento que une, en perpendicular, el punto inicial con la recta x=0. Los puntos azules son los puntos iniciales en tres casos diferentes. Los puntos rojos, a los que se llega mediante una flecha azul, los puntos solución en los tres casos. Como se ve, una de las consecuencias del "creeping" es que la solución depende enormemente del punto de partida. Si la incógnita es el propio modelo, la solución de norma mínima es, siempre, el punto (0,0) independientemente de cual haya sido el punto inicial. Ésta es una buena característica del método. Las flechas magenta unen los puntos iniciales con el punto solución.

3. Caso sobredeterminado con truncamiento

Supóngase ahora que el operador \mathbf{JF}_k no sea deficiente en rango. Aún así, una descomposición en valores singulares de $\mathbf{JF}_k^t\mathbf{JF}_k$ puede mostrar que algunos de los valores singulares no nulos son tan pequeños que generan problemas de inestabilidad en la solución. Se pretende atenuar dicha variabilidad regularizando el problema mediante truncamiento de la parte del problema del mismo correspondiente a los valores singulares conflictivos. Supóngase que, para ello, deben retenerse sólo los primeros r < n de ellos.

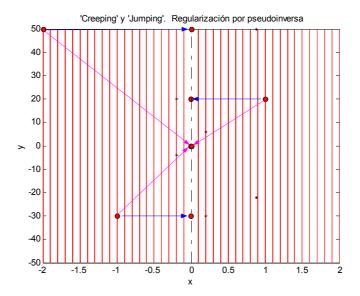


Figura 39: Problema subdeterminado: diferencias entre creeping y jumping al solucionar mediante la pseudoinversa.

• Formulación matemática

En la formulación "creeping"

$$\Delta^{c} \mathbf{m}_{k} = \sum_{i=1}^{i=r} \frac{\mathbf{u}_{i}^{t} \left(\mathbf{d} - \mathbf{F} \left(\mathbf{m}_{k} \right) \right)}{\sigma_{i}} \mathbf{v}_{i},$$

$$\mathbf{m}_{k+1}^{creep} = \mathbf{m}_{k} + \Delta^{c} \mathbf{m}_{k},$$

mientras que, al plantear "jumping"

$$\mathbf{m}_{k+1}^{jump} = \sum_{i=1}^{i=r} rac{\mathbf{u}_{i}^{t} \left(\mathbf{d} - \mathbf{F}\left(\mathbf{m}_{k}
ight) + \mathbf{J} \mathbf{F}_{k} \mathbf{m}_{k}
ight)}{\sigma_{i}} \mathbf{v}_{i}.$$

Esta expresión puede reescribirse como

$$\mathbf{m}_{k+1}^{jump} = \sum_{i=1}^{i=r} \frac{\mathbf{u}_{i}^{t} \left(\mathbf{d} - \mathbf{F}\left(\mathbf{m}_{k}\right)\right)}{\sigma_{i}} \mathbf{v}_{i} + \sum_{i=1}^{i=r} \frac{\mathbf{u}_{i}^{t} \left(\sum\limits_{j=1}^{j=r} \sigma_{j} m_{j} \mathbf{u}_{j} + \sum\limits_{j=n-r-1}^{j=n} \sigma_{j} m_{j} \mathbf{u}_{j}\right)}{\sigma_{i}} \mathbf{v}_{i}.$$

Teniendo en cuenta la ortogonalidad de los vectores \mathbf{u}_i resulta

$$\mathbf{m}_{k+1}^{jump} = \sum_{i=1}^{i=r} \frac{\mathbf{u}_{i}^{t} \left(\mathbf{d} - \mathbf{F} \left(\mathbf{m}_{k}\right)\right)}{\sigma_{i}} \mathbf{v}_{i} + \sum_{j=1}^{j=r} m_{j} \mathbf{v}_{j},$$

de donde, sustituyendo

$$\mathbf{m}_{k+1}^{jump} = \Delta^c \mathbf{m}_k + \sum_{j=1}^{j=r} m_j \mathbf{v}_j = \mathbf{m}_{k+1}^{creep} - \mathbf{m}_k + \sum_{j=1}^{j=r} m_j \mathbf{v}_j,$$

por tanto, finalmente

$$\mathbf{m}_{k+1}^{jump} = \mathbf{m}_{k+1}^{creep} - \sum_{j=s+1}^{j=n} m_j \mathbf{v}_j.$$

Es decir, la diferencia entre las dos soluciones vuelve a ser la proyección del modelo inicial \mathbf{m}_k sobre el núcleo numérico de \mathbf{JF}_k -es decir, el subespacio al que pertenece la información que no se considera-.

Ilustración

Supóngase que se quiere encontrar el mínimo de la función:

$$z = \|\mathbf{L}(\mathbf{x} - \mathbf{x}^*)\|_2,$$

donde

$$\mathbf{L} = \left[\begin{array}{cc} 10 & 0 \\ 0 & 0.1 \end{array} \right],$$

y $\mathbf{x}^* = (0,0)$. Las curvas de nivel de la misma se representan en rojo en la figura 40.

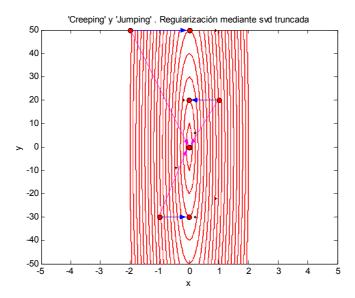


Figura 40: Problema sobredeterminado. Diferencia entre creeping y jumping al solucionar utilizando truncamiento

Se observa un mínimo con forma de valle muy alargado en la dirección del eje de las y. Para los mismos puntos iniciales que en el caso anterior, el comportamiento tanto del "creeping" como del "jumping" es similar al ilustrado en aquel caso. Cuando se utiliza "creeping" -flechas azules- la solución está a lo largo del valle de mínimos justo allí donde "caería" el punto inicial. La dependencia del lugar de comienzo es obvia a partir de la figura. La metodología "jumping" -flechas magenta- es, al contrario, básicamente robusta en el sentido de que no depende del punto inicial. La información a priori (que aquí transmite la consigna de que no puede haber componentes verticales) actúa, en este caso sobre todo el modelo.

Considérese un ejemplo algo menos académico. Sobre un terreno modelizado mediante el vector

$$\mathbf{m} = (100, 5, 1, 1, 100),$$

se lleva a cabo una campaña de 5 medidas con semiespaciados interelectródicos indicados por el vector

$$AB/2 = (1, 10, 50, 90, 200)$$
.

El vector de medidas está afectado por un error cuadrático medio del $5\,\%$ en términos relativos al valor sin error. Los datos medidos son

$$\rho_m = (88.693, 53.2199, 36.065, 44.735, 70.148)$$
.

En la figura 41 se representa la gráfica de la norma euclídea del error relativo para distintos modelos obtenidos al variar el espesor y la resistividad de la segunda capa. Como se observa existe un valle claramente delineado en el que los valores de la función son menores que en el resto de los modelos.

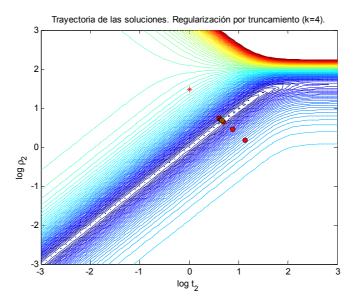


Figura 41: Evolución del proceso de optimización con algoritmo local empleando truncamiento de valores singulares en cada subproblema lineal.

El punto inicial del proceso iterativo aparece indicado mediante un asterisco rojo. El método empleado para regularizar ha sido truncar la información del problema relativa al último valor singular y resolver en incrementos. Es decir, se retienen sólo los cuatro mayores valores singulares de los cinco existentes. Se observa como la trayectoria de las soluciones parciales hasta la definitiva (bordeada con un círculo verde) es casi perpendicular a la dirección del valle. Esto demuestra que los ejemplos académicos anteriores son de gran utilidad para la comprensión de la problemática real en la aplicación de S.E.V.

4. Caso general. Regularización de Thikonov

El funcional lineal a minimizar en cada una de las iteraciones, puede ser:

• en el caso "creeping",

$$\Delta \mathbf{m} = \min_{\Delta \mathbf{m} \in \boldsymbol{\delta}(\mathbf{0})} \| \mathbf{J} \mathbf{F}_{k} \Delta \mathbf{m} - \mathbf{d}_{k} \|_{2}^{2} + \lambda^{2} \| \mathbf{L} (\Delta \mathbf{m} - \Delta \mathbf{m}^{*}) \|_{2}^{2},$$

$$\mathbf{m}_{k+1}^{creep} = \mathbf{m}_{k} + \Delta \mathbf{m},$$
(198)

• en el caso "jumping",

$$\mathbf{m}_{k+1}^{jump} = \min_{\mathbf{m} \in \delta(\mathbf{m}_k)} \left\| \mathbf{J} \mathbf{F}_k \mathbf{m} - \hat{\mathbf{d}}_k \right\|_2^2 + \lambda^2 \left\| \mathbf{L} \left(\mathbf{m} - \mathbf{m}^* \right) \right\|_2^2.$$
 (199)

La regularización es, como se sabe, un proceso que trata de amortiguar el efecto que los errores aleatorios tienen sobre las componentes de la solución expresadas en cierta base. El efecto corrector de la regularización tiende a estar concentrado, cuando se aplica correctamente, sobre todo en las componentes más inestables. Lo ideal sería que no afectase nada (o lo menos posible) a las componentes estables de la solución real.

Si la incógnita es $\Delta \mathbf{m}$ resultarán sobre todo modificadas las coordenadas de $\Delta \mathbf{m}$ en los vectores problemáticos de la base escogida⁶⁵. Estas modificaciones tenderán a hacer menos importantes estas componentes produciendo una mayor proximidad de $\Delta \mathbf{m}$ respecto de la información a priori. Considérese la posibilidad de que el modelo de partida \mathbf{m}_k tenga, él mismo, una componente muy importante sobre el subespacio generado por los vectores "problemáticos" de la base. Como dicho modelo \mathbf{m}_k no se somete a ningún proceso de regularización, esta componente problemática tiende a mantenerse lejos de la información a priori, ajena al hecho de que $\Delta \mathbf{m}$ ha "limado" las diferencias.

En la formulación "jumping" todas las partes inestables del modelo son modificadas de acuerdo con las pretensiones del proceso de regularización. Por tanto, ninguna parte del modelo inicial permanece inmune a dichos cambios. Si la regularización tiende a producir acercamiento a alguna característica del modelo de referencia \mathbf{m}^* , toda componente del modelo inicial \mathbf{m}_k debe cambiar de forma que dicha aproximación pueda producirse.

• Formulación matemática

Las ecuaciones normales correspondientes son en el caso de "creeping" (198),

$$\left(\mathbf{J}\mathbf{F}_{k}^{t}\mathbf{J}\mathbf{F}_{k}+\lambda\mathbf{L}^{t}\mathbf{L}\right)\Delta\mathbf{m}=\mathbf{J}\mathbf{F}_{k}^{t}\mathbf{d}_{k}+\mathbf{L}^{t}\mathbf{L}\Delta\mathbf{m}^{*},$$

⁶⁵ Normalmente los asociados a valores singulares pequeños

y, en términos de problema de minimización:

$$\min_{\Delta \mathbf{m}} \left\| \mathbf{\tilde{J}} \left(\mathbf{\Delta m} \right) - \mathbf{\tilde{d}} \right\|_{2},$$

donde

$$\begin{split} \tilde{\mathbf{J}} &= \left[\begin{array}{c} \mathbf{J}\mathbf{F}_k \\ \sqrt{\lambda}\mathbf{L} \end{array} \right], \\ \tilde{\mathbf{d}} &= \left[\begin{array}{c} \mathbf{d} - \mathbf{F}\left(\mathbf{m}_k\right) \\ \sqrt{\lambda}\mathbf{L}\boldsymbol{\Delta}\mathbf{m}^* \end{array} \right]. \end{split}$$

En el caso de "jumping" (199),

$$\left(\mathbf{J}\mathbf{F}_{k}^{t}\mathbf{J}\mathbf{F}_{k}+\lambda\mathbf{L}^{t}\mathbf{L}\right)\mathbf{m}=\mathbf{J}\mathbf{F}_{k}^{t}\mathbf{\hat{d}}_{k}+\mathbf{L}^{t}\mathbf{L}\mathbf{m}^{*},$$

que puede formularse como un problema de mínimos cuadrados como

$$\min_{\mathbf{m}} \left\| \mathbf{\tilde{J}m} - \mathbf{\tilde{d}} \right\|_2$$

donde

$$\begin{split} \mathbf{\tilde{J}} &= \begin{bmatrix} \mathbf{J}\mathbf{F}_k \\ \sqrt{\lambda}L \end{bmatrix}, \\ \mathbf{\tilde{d}} &= \begin{bmatrix} \mathbf{d} - \mathbf{F} \left(\mathbf{m}_k \right) + \mathbf{J}\mathbf{F}_k \mathbf{m}_k \\ \sqrt{\lambda} \mathbf{L} \mathbf{m}^* \end{bmatrix}. \end{split}$$

Si los núcleos de \mathbf{JF}_k y \mathbf{L} tienen intersección vacía, dicho sistema tiene una solución única. Comparando ambas formulaciones se tiene que:

$$\mathbf{m}_{k+1}^{creep} - \mathbf{m}_{k+1}^{jump} = \mathbf{ ilde{J}}^{\dagger} \left[egin{array}{c} \mathbf{0} \\ \sqrt{\lambda} \mathbf{L} \mathbf{m}_k \end{array}
ight] + \mathbf{ ilde{J}}^{\dagger} \left[egin{array}{c} \mathbf{0} \\ \sqrt{\lambda} \mathbf{L} \left(\mathbf{m}_{cr}^* - \mathbf{m}_j^* - \mathbf{m}_k
ight) \end{array}
ight].$$

Como ejemplo, cuando $\mathbf{L}\Delta\mathbf{m}^*=\mathbf{0}$, con lo cual $\mathbf{L}\mathbf{m}_{cr}^*=\mathbf{L}\mathbf{m}_k$, y $\mathbf{L}\mathbf{m}_j^*=\mathbf{0}$, la diferencia es

$$\mathbf{m}_{k+1}^{creep} - \mathbf{m}_{k+1}^{jump} = \mathbf{ ilde{J}}^{\dagger} \left[egin{array}{c} \mathbf{0} \ \sqrt{\lambda} L \mathbf{m}_k \end{array}
ight],$$

con lo cual se vuelve a observar la dependencia (generalizada) respecto al modelo inicial.

Ilustración

Sobre la misma función de error previamente ilustrada, se ha querido mostrar cómo la información a priori puede "dirigir" los destinos del proceso de inversión en el caso de regularización por el método de Tikhonov. A continuación se aportan dos ejemplos con distintos vector de información a priori. En el primer caso, figura 42, el vector de referencia es $\mathbf{m}^* = (100, 5, 50, 10, 100)$, y el punto inicial $\mathbf{m}_0 = (100, 5, 50, 1, 100)$.

Se han probado seis valores distintos -en orden creciente- del parámetro λ , desde 0 hasta 10. La figura muestra el acercamiento progresivo de la solución al vector \mathbf{m}^* incluso aunque deba "escalar" la ladera del valle. La figura 43 muestra lo mismo pero con distinto vector de referencia. Ahora $\mathbf{m}^* = (100, 5, 0.01, 0.01, 100)$, y se mantiene el mismo punto inicial. Los puntos más alejados corresponden a $\lambda = 0$ y 10^{-6} respectivamente. A partir de 10^{-3} las soluciones aparecen claramente dominadas por el punto de referencia.

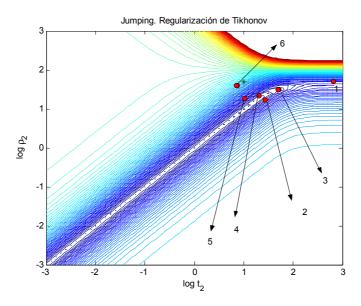


Figura 42: Evolución de algoritmo local, tipo "jumping" con regularización de Tikhonov.

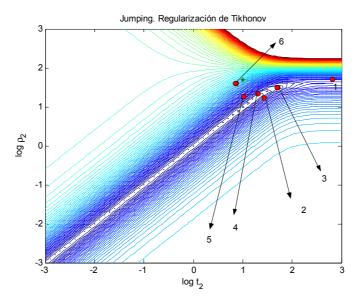


Figura 43: Evolución de algoritmo local, tipo "jumping" con regularización de Tikhonov y nuevo modelo de referencia.

5.4.2. Regularización en la formulación global

El planteamiento del problema de minimización en este caso es

$$\mathbf{m}_{sol} = \min_{\mathbf{m} \in M} \left\| \mathbf{d} - \mathbf{F} \left(\mathbf{m} \right) \right\|_{2}^{2} + \lambda^{2} \left\| \mathbf{L} \left(\mathbf{m} - \mathbf{m}^{*} \right) \right\|_{2}^{2},$$

que, tras linealización del funcional ${\bf F}$ en el entorno de ${\bf m}_k$ se convierte en

$$\min_{\mathbf{m} \in \boldsymbol{\delta}\left(\mathbf{m}_{k}\right)} \left\| \mathbf{d} - \mathbf{F}\left(\mathbf{m}_{k}\right) - \mathbf{J}\mathbf{F}_{k}\left(\mathbf{m} - \mathbf{m}_{k}\right) \right\|_{2}^{2} + \lambda^{2} \left\| \mathbf{L}\left(\mathbf{m} - \mathbf{m}^{*}\right) \right\|_{2}^{2}.$$

Esta expresión es idéntica a (199) y, por tanto, son idénticas sus soluciones. Merece la pena mencionar que, una vez expresado el problema de esta forma, puede resolverse respecto a $\Delta \mathbf{m}$. En este caso hay que solucionar

$$\min_{\Delta \mathbf{m}} \left\| \mathbf{\tilde{J}} \Delta \mathbf{m} - \mathbf{\tilde{d}} \right\|_2$$

donde

$$\mathbf{\tilde{J}} \! = \! \left[\begin{array}{c} \mathbf{J} \mathbf{F}_k \\ \sqrt{\lambda} L \end{array} \right], \; \mathbf{\tilde{d}} \! = \! \left[\begin{array}{c} \mathbf{d} - \mathbf{F} \left(\mathbf{m}_k \right) \\ \sqrt{\lambda} L \mathbf{m}^* \end{array} \right]$$

Ilustración

Siempre sobre la misma superficie de error que en casos precedentes, se ilustra ahora un caso de regularización global con modelo de referencia $\mathbf{m}^* = (100, 5, 0.01, 0.01, 100)$, y con punto inicial del proceso $\mathbf{m}_0 = (100, 5, 50, 1, 100)$. Estos resultados pueden observarse en la figura 44.

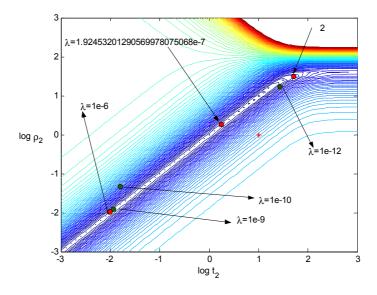


Figura 44: Regularización global

Para valores de λ del orden de $2 \cdot 10^{-7}$ o inferiores, se observa una gran inestabilidad en las soluciones que pueden obtenerse casi en cualquier lugar del valle. Para valores de lambda ligeramente superiores algunas de las soluciones se concentran en la parte central. Para mayores valores, invariablemente aparecen en las cercanías del modelo de referencia.

5.5. Métodos de optimización global

Los métodos de optimización global permiten la exploración potencial del espacio de búsqueda. En esta tesis se han utilizado los algoritmos genéticos con codificación binaria y aritmética así como simulated annealing. Dada la importancia que poseen en la metodología aquí desarrollada serán tratados de forma individual en un capítulo separado.

Capítulo 6: Análisis de modelos equivalentes en SEV

6. ANÁLISIS DE MODELOS EQUIVALENTES EN S.E.V.

6.1. Introducción

La aplicación de un método resistivo, tal como el de S.E.V, presupone la asimilación del terreno a un conjunto de resistencias eléctricas dispuestas de cierta forma.

De manera simplificada, supónganse dos porciones de conductores diferentes. La primera porción tiene de resistividad ρ_1 , longitud l_1 y el área de su sección transversal S_1 . Del mismo modo los parámetros de la segunda porción de conductor son ρ_2 , l_2 y S_2 .

Considérese un cable de longitud l construido por conexión en serie de las dos citadas porciones, suponiendo además que $S_1 = S_2 = S$. Las longitudes de ambas porciones cumplen que: $l = l_1 + l_2$. Si dicho cable es recorrido longitudinalmente por una corriente de intensidad I y se mide la diferencia de potencial ΔV entre los extremos del mismo, puede calcularse la magnitud de la resistividad ρ_s que debería tener un conductor homogéneo de longitud l y sección S tal que, al ser atravesado por la misma intensidad de corriente I produjese la misma caída de voltaje ΔV . Su expresión es:

$$\rho_s = \rho_1 \frac{l_1}{l} + \rho_2 \frac{l_2}{l}.$$

En esta expresión se resalta el hecho de que, si l se supone constante, ρ_s depende de los productos $\rho_1 l_1$ y $\rho_2 l_2$ y no de los valores de ρ_1 y l_1 (resp. ρ_2 y l_2) independientemente. O sea, desde el punto de vista de la resistividad aparente, una disminución de la resistividad, por ejemplo ρ_1 , puede verse compensada por un aumento de la longitud del tramo correspondiente l_1 . Si se plantea un problema en que, a partir de ciertas medidas de potencial, las incógnitas sean las resistividades y las longitudes de los tramos de cada conductor, se constata la presencia de cables $equivalentes^{66}$.

Igualmente cabe considerar la conexión en paralelo de los dos conductores mencionados, suponiendo ahora que tienen igual longitud. O sea, $l_1 = l_2 = l$. En este caso, la resistividad del conductor equivalente es:

$$\rho_p = \frac{S}{\frac{S_1}{\rho_1} + \frac{S_2}{\rho_2}}.$$

También en este caso son las ratios $\frac{S_1}{\rho_1}$ y $\frac{S_2}{\rho_2}$ las verdaderamente relevantes en el problema, permitiendo que la misma resistividad aparente ρ_p pueda estar representada por cables de distintos grosores y resistividades.

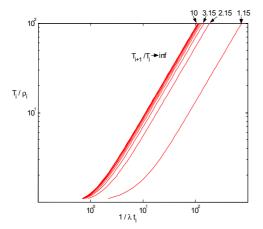
En una distribución más general de las resistividades, como la que se asume en el método de S.E.V, donde la dirección de la corriente no es ni estrictamente perpendicular a las capas (resistividades en serie) ni tampoco paralela a las mismas (conexión en paralelo) se tendrá una situación intermedia a las dos presentadas. No obstante, como se verá, existen ciertos tipos de estratificaciones que se aproximan lo suficiente a uno de los dos modelos anteriores como para que su análisis sea aproximadamente correcto.

En el caso de S.E.V se entiende por terrenos equivalentes aquellos caracterizados por una distribución (estratificada) de resistividades y espesores diferentes, pero que producen aproximadamente las mismas medidas de resistividad aparente. En realidad, tal como demostró Koefoed (1979), las razones para la equivalencia están implícitas en las relaciones de Pekeris. A

⁶⁶Desde el punto de vista de la caída de potencial.

partir de ellas, puede construirse un gráfico que represente el cociente $\frac{T_i}{\rho_i}$ frente a $\frac{1}{\lambda t_i}$ utilizando el valor $\frac{T_{i+1}}{T_i}$ como parámetro. Supóngase que se conoce T_1 para distintos valores de $\frac{1}{\lambda}$ (curva de transformadas de resistividad de la primera capa). Esta curva puede estimarse, por ejemplo, a partir de los datos de resistividades aparentes tomados en el campo. A partir de la misma, existen métodos que permiten hallar los parámetros de la primera capa, ρ_1 y t_1 . Como la gráfica propuesta se ha dibujado en coordenadas bilogarítmicas, la curva de campo (o la de transformadas de resistividad deducida de ella, $T_1 = T_1\left(\frac{1}{\lambda}\right)$), tras ser desplazada para situar en el origen del gráfico el punto (ρ_1,t_1) , queda situada sobre una línea caracterizada por un valor determinado del parámetro $\frac{T_{i+1}}{T_i}$. Este valor es el del factor por el que se debe de multiplicar la curva $T_1 = T_1\left(\frac{1}{\lambda}\right)$ para obtener la curva $T_2 = T_2\left(\frac{1}{\lambda}\right)$. A partir de ésta se procede a la estimación de los parámetros de la segunda capa ρ_2 y t_2 . El procedimiento continúa permitiendo el cálculo de todos parámetros geoeléctricos del terreno.

En las figura 45 se observa que las líneas se aglomeran hasta hacerse indistinguibles para valores de $\frac{T_{i+1}}{T_i}$ grandes (resp. $\frac{T_{i+1}}{T_i}$ tendiendo a cero).



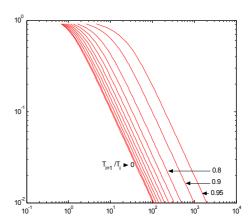


Figura 45: Cálculo de los factores de reducción al plano inferior de la transformada de resistividad

Si, en este proceso, se comete un pequeño error en la estimación del valor de T_i de forma que exista incertidumbre en el valor de $\frac{T_i}{\rho_i}$, en la zona de gran aglomeración de líneas, pueden producirse grandes inestabilidades en la determinación del valor del parámetro $\frac{T_{i+1}}{T_i}$ y, por lo tanto, la curva $T_{i+1} = T_{i+1} \left(\frac{1}{\lambda}\right)$ así obtenida ser errónea.

Sobre todo en la zona de gran concentración de líneas, cabe observar que las mismas tienden a disponerse con una pendiente +1 (resp. -1) en la figura 45. Supóngase que se ha cometido algún error en la determinación de los parámetros (ρ_i, t_i) y se desea saber cómo se propaga el

mismo a través del cálculo del factor $\frac{T_{i+1}}{T_i}$. De ambos gráficos se deduce que, sobre todo en zonas donde T_i sea relativamente grande (resp. pequeño) respecto a ρ_i hay errores en los parámetros que dejan inalterado el valor de $\frac{T_{i+1}}{T_i}$ y, por tanto, no influyen en los valores del resto de los parámetros de las capas subyacentes.

Si $\hat{\rho}_i$ y \hat{t}_i son valores estimados erróneamente de ρ_i y t_i que cumplen las relaciones

$$\hat{\rho}_i = c\rho_i, \ \hat{t}_i = ct_i, \ c \in \mathbb{R},$$

de modo que el cociente $\frac{\hat{\rho}_i}{\hat{t}_i} = k$, constante, entonces,

 $\log \frac{T_i}{\hat{\rho}_i} = \log \frac{T_{i+1}}{\rho_i} + \log \frac{1}{c},$

у

$$\log \frac{1}{\lambda \hat{t}_i} = \log \frac{1}{\lambda t_i} + \log \frac{1}{c}.$$

Esto significa que el punto correspondiente a los valores $(\hat{\rho}_i, \hat{t}_i)$ se sitúa sobre una recta que pasa por el punto correspondiente a (ρ_i, t_i) con una inclinación de 45°. Por tanto, según la figura 45, el valor de $\frac{T_{i+1}}{T_i}$ no cambia. Del mismo modo si el producto $\hat{\rho}_i \hat{t}_i = k$, entonces puede escribirse que

 $\hat{\rho}_i = c\rho_i, \ \hat{t}_i = \frac{t_i}{c}, \ c \in \mathbb{R},$

por lo tanto,

 $\log \frac{T_i}{\hat{\rho}_i} = \log \frac{T_i}{\rho_i} + \log \frac{1}{c},$

у

$$\log \frac{1}{\lambda \hat{t}_i} = \log \frac{1}{\lambda t_i} - \log \frac{1}{c},$$

lo que indica que el punto que corresponde a $(\hat{\rho}_i, \hat{t}_i)$ se sitúa sobre una recta que pasa por el relativo a (ρ_i, t_i) cuya pendiente es -1. No cambia tampoco en este caso el valor de $\frac{T_{i+1}}{T_i}$.

El primer gráfico de la figura 45 ilustra el caso donde $\frac{T_{i+1}}{T_i} > 1$ y, por tanto, ρ_{i+1} es mayor que ρ_i . La equivalencia, que se manifiesta como ausencia de efecto sobre el valor de $\frac{T_{i+1}}{T_i}$ en la reducción de horizonte, es tanto más evidente cuanto menor es el valor de ρ_i , mayor el de ρ_{i+1} -y, por tanto, el de T_{i+1} - y menor el espesor t_i . Un terreno de tres capas con esta descripción tendría una capa central delgada respecto al terreno suprayacente y una resistividad pequeña respecto a la de la capa inferior. Puede contemplarse el caso de que la resistividad de la capa de encima sea mayor -en cuyo caso la curva se denomina tipo cuenco ("bowl")- o bien que sea más pequeña, dando lugar a la tipología denominada ascendente.

De la misma forma, de la observación del segundo de los gráficos, se deducen los mismos efectos para el caso en que, sobre todo con t_i pequeño, ρ_{i+1} sea menor que ρ_i y, por tanto $\frac{T_{i+1}}{T_i} < 1$. Un terreno de tres capas con estas características puede presentar una pseudosección 67

⁶⁷Curva de resistividades aparentes

de tipo "campana" o bien "ascendente".

6.1.1. Las reglas de equivalencia de Maillet. Límites de validez

Varias veces a lo largo de los capítulos precedentes han sido citadas las denominadas reglas de Maillet. Maillet (1947) enuncia las características que deben cumplir los terrenos denominados equivalentes. Las situaciones susceptibles de producir equivalencia pueden resumirse en dos tipologías:

■ Perfil en "campana" y perfil "desdencente"

La capa (supóngase que es la i-ésima) cuyos parámetros (ρ_i, t_i) pueden presentarse en distintas combinaciones equivalentes tiene las siguientes características:

- 1. su potencia t_i es "pequeña" respecto a la profundidad a la que se halla,
- 2. su resistividad ρ_i es "grande" en relación a la de la capa subyacente ρ_{i+1} . El perfil en "campana" se produce cuando, además, ρ_i también es mayor que ρ_{i-1} , mientras que el "desdencente" aparece cuando ρ_i es menor que ρ_{i-1} .

El comportamiento de las líneas de corriente cuando se encuentran con la capa i-ésima, de elevada ρ_i como se ha dicho, es acortar al máximo su recorrido en el interior de la misma -minimizando la disipación de energía-, acercando su trayectoria a la vertical. La aproximación de recorrido estrictamente transversal adquiere más validez cuanto menos potente -menor t_i - sea la capa.

El paso de la corriente por la capa i-ésima podría pues describirse muy aproximadamente (véase la sección precedente) como la circulación por un conductor cilíndrico homogéneo, dispuesto verticalmente, de resistividad ρ_i y longitud t_i . Si se razona sobre la unidad de área horizontal, la resistencia es $R = \rho_i t_i$. Ésta, se denomina **resistencia transversal** y es el parámetro que realmente influye en estas condiciones.

• Perfil en "cuenco" y perfil "ascendente"

En este caso, los parámetros (ρ_i, t_i) cumplen que:

- 1. el espesor t_i es "pequeño" en relación al espesor del terreno suprayacente,
- 2. su resistividad ρ_i es "pequeña" comparada con la de la capa inferior ρ_{i+1}. El perfil denominado "en cuenco" se produce cuando, además, ρ_i es también "pequeña" en relación a ρ_{i-1}. En caso de que ρ_i sea mayor que ρ_{i-1} se obtiene el perfil "ascendente". Como ρ_i < ρ_{i+1}, la disipación de energía que sufre la corriente al circular por la capa i-ésima es menor que la que tendría lugar al hacerlo por la capa subyacente. Así pues, el recorrido por el interior de la capa i-ésima tiende a maximizarse y, por tanto, a aproximarse a la horizontalidad. Argumentando sobre la anchura unidad, la situación es similar a la producida al recorrer la corriente un conductor homogéneo, dispuesto horizontalmente, de resistividad ρ_i y sección t_i. Tal como se ha visto en la sección introductoria, el parámetro influyente entonces es C = ρ_i / t_i, que se denomina conductancia longitudinal. Por tanto, y dentro de los límites enunciados, cualquier

combinación de (ρ_i, t_i) con igual conductancia longitudinal dará lugar a una capa equivalente.

Las reglas de equivalencia enunciadas afirman que las combinaciones de parámetros

$$C = \frac{\rho_i}{t_i}, \ R = \rho_i t_i,$$

son las realmente importantes. En papel bilogarítmico ($\log \rho_i$, $\log t_i$) las parametrizaciones equivalentes de la capa i–ésima adquieren la forma de líneas rectas de pendiente +1 en los casos "campana" y "descendente", y de pendiente -1 en el caso de perfiles en "cuenco" y "ascendente".

Koefoed (1976a) mostró que las reglas de equivalencia de Maillet deben ser aplicadas con cautela porque tienen un rango de validez limitado. Sus resultados fueron escritos en términos de transformadas de resistividad. En la figura 46 se ilustra el comportamiento para resistividades aparentes.

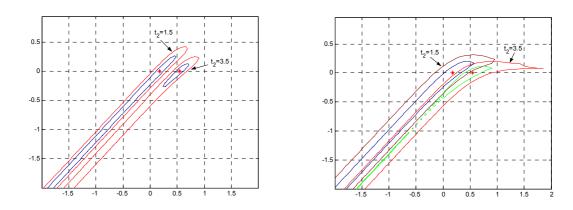


Figura 46: Zonas de equivalencia del 5 % y 10 % de error en resistividades aparentes: a) suelo de cuatro capas $\mathbf{m} = (10, 1, 1, t_2, 100, 1, 1)$. b) idem con $\mathbf{m} = (10, 1, 1, t_2, 10, 1, 1)$.

El terreno empleado en la simulación ha sido $\mathbf{m} = (10, 1, 1, t_2, 100, 1, 1)$ con valores de $t_2 = 1$ y 3.5. Se han obtenido así las regiones de equivalencia del 5% y del 10% de error relativo. Posteriormente se ha reducido el contraste de resistividades cambiando el valor de ρ_3 a 10. Las regiones ilustradas en el segundo gráfico muestran cierta distorsión en relación a las precedentes (mayor contraste de resistividad).

Las figuras anteriores muestran que los comportamientos en resistividades aparentes de las zonas de equivalencia son básicamente los mismos que los enunciados por Koefod para transformadas de resistividad, a saber:

• la región de equivalencia no es estrictamente una línea, sino una superficie. Su forma recuerda claramente, no obstante, en ciertas zonas, a una relación lineal con pendiente unidad;

- a veces, la pendiente con la que puede caracterizarse la zona de equivalencia no es la unidad. Esto es especialmente más acusado cuanto menor es el contraste de resistividades;
- las zonas de equivalencia son cerradas, apreciándose sobre todo este efecto en su parte superior. Esto confirma el carácter limitado de las reglas de Maillet;
- no puede olvidarse la influencia que tienen también los parámetros restantes en la consecución de modelos equivalentes.

6.1.2. Las curvas de Dar Zarrouk

Las curvas de Dar Zarrouk están íntimamente relacionadas con las reglas de equivalencia de Maillet y han sido descritas en el mismo artículo (Maillet, 1947). Se trata de otra forma de representar un suelo estratificado. A partir de los parámetros de un suelo, puede construirse de forma unívoca su curva de Dar Zarrouk y viceversa.

La propia curva de Dar Zarrouk tiene dos representaciones posibes. Ambas necesitan la definición previa de dos nuevas variables:

■ resistencia transversal equivalente (R): es una función del perfil de resistividades del suelo y de la profundidad. Su expresión es:

$$R(z) = \int_{0}^{z} \rho(z) dz,$$

• resistencia longitudinal equivalente (C), definida como:

$$C\left(z\right) = \int_{0}^{z} \frac{1}{\rho\left(z\right)} dz.$$

Para cada tipo de suelo se puede, en principio, eliminar la variable z de la definición obteniéndose una relación entre R y C (más concretamente $\ln R.vs. \ln C$). La expresión

$$r_s = \sqrt{\frac{R}{C}},$$

tiene dimensiones de resistividad y Maillet la denomina "pseudo-resistividad". Asimismo la magnitud

$$z_s = \sqrt{RC}$$
,

tiene dimensiones de profundidad y recibe el nombre de "pseudo-profundidad".

La representación alternativa, equivalente a la anterior, de la curva de Dar Zarrouk consiste en presentar $r_s.vs.z_s$.

Según Maillet, distribuciones de capas equivalentes producen cambios despreciables en la curva de Dar Zarrouk. No obstante, dado que las reglas de equivalencia tienen un rango limitado de validez pueden construirse curvas de Dar Zarrouk apreciablemente distintas que correspondan

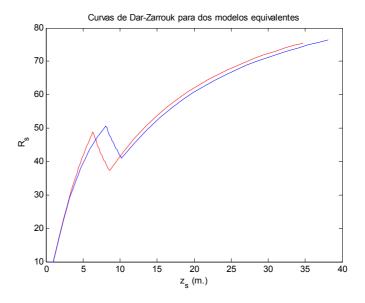


Figura 47: Curva de Dar-Zarrouk para los modelos de cuatro capas: $\mathbf{m}_1 = (10, 1, 100, 3, 10, 1, 100)$ y $\mathbf{m}_2 = (10, 0.97, 80, 5, 10, 0.9, 100)$.

a terrenos equivalentes. Este caso se ilustra en la figura 47 que representa las curvas de Dar Zarrouk para los modelos

$$\mathbf{m}_1 = (10, 1, 100, 3, 10, 1, 100) \text{ y } \mathbf{m}_2 = (10, 0.97, 80, 5, 10, 0.9, 100).$$

Aunque en este trabajo no se ha hecho mayor uso de dichas curvas, se incluye este apartado debido a su importancia histórica en el estudio de este tema.

6.2. El análisis lineal de equivalencias

Supóngase que se ha calculado, por algún método de los citados en el capítulo no lineal, alguna solución \mathbf{m}_0 del problema no lineal. Se trata ahora, de encontrar otros modelos que ajusten los datos dentro de cierta tolerancia especificada, sin tener que resolver para ellos el problema original.

La idea es sustituir en un entorno de la solución \mathbf{m}_0 la función objetivo no lineal , por ejemplo,

$$E\left(\mathbf{m}\right) = \left\|\mathbf{d} - \mathbf{F}(\mathbf{m})\right\|_{N_D}^2,\tag{200}$$

por una aproximación cuadrática de la misma. La denominación de análisis lineal proviene del hecho (detallado en el capítulo no lineal) de que dicha aproximación cuadrática puede obtenerse a partir de (200) linealizando \mathbf{F} . Al proceder así, cobra especial relevancia la matriz jacobiana $\mathbf{JF}(\mathbf{m}_0) \equiv \mathbf{JF}_{\mathbf{m}_0}$ para caracterizar el tamaño y la orientación de la zona de equivalencia lineal correspondiente a cierta tolerancia prefijada. En este capítulo se estudia la relación que existe entre la zona de equivalencia lineal y la no lineal para distintas familias de curvas tipo.

Asimismo, se desarrolla la posibilidad de obtener modelos linealmente equivalentes a uno dado, \mathbf{m}_0 , y que también cumplan ciertas restricciones que traducen información a priori. Con

este fin, se utilizará el formalismo desarrollado para este propósito (expresión matemática de la información a priori -véase el capítulo lineal-).

Finalmente, recurriendo a la descomposición en valores singulares de la matriz $\mathbf{JF}_{\mathbf{m}_0}$ se pone de manifiesto que en el caso de los S.E.V la información vinculada a valores singulares pequeños no describe -contra lo que comúnmente se afirma- características de alta frecuencia del modelo solución, sino que incluye propiedades claramente estructurales. Esta situación se presenta debido a que el valor del funcional del problema directo \mathbf{F} , depende de combinaciones de parámetros (fundamentalmente las mencionadas R y C) que pueden permanecer casi inalteradas a pesar de grandes variaciones relativas en los parámetros del modelo.

6.2.1. Caso I: modelos linealmente equivalentes sin restricciones

En esta sección se parte de un modelo geoeléctrico conocido del subsuelo. Si bien una posibilidad de excepcional interés es que éste se haya obtenido como resultado de la resolución de un problema de optimización, se aborda aquí una formulación más general en la que el modelo inicial, que se denominará \mathbf{m}_0 , puede ser cualquier punto⁶⁸. Naturalmente son del máximo interés aquellos que verifican la siguiente desigualdad:

$$\|\mathbf{d} - \mathbf{F}(\mathbf{m}_0)\|_{N_D}^2 \le \varepsilon^2,\tag{201}$$

donde \mathbf{F} es el operador (no lineal) que modeliza el problema directo, N_D es la norma en el espacio de datos y que corresponde a una determinada ponderación de los errores dada por la matriz \mathbf{W} y ε^2 es una tolerancia de ajuste dada.

El problema que se plantea es encontrar otros modelos \mathbf{m} para los cuales también se verifique la desigualdad (201). Una posibilidad es realizar una exploración directa, calculando $\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{N_D}^2$ sobre una familia elegida de modelos. La dificultad de esta metodología estriba en que no hay, a priori, criterios para escoger dicha familia y tratar de realizar una búsqueda sistemática puede ser totalmente ineficiente desde el punto de vista computacional. A modo de ejemplo, si se considera un modelo de suelo de cuatro capas es preciso determinar siete parámetros. Eligiendo diez valores para cada parámetro, el número de modelos posibles es de 10^7 que es, a su vez, el número de evaluaciones necesarias del funcional no lineal presente en (201), lo cual es prohibitivo en términos de tiempo de cálculo.

Otras técnicas, como los métodos probabilísticos de optimización global (G.A y simulated annealing), mencionados en el capítulo anterior, pueden ser utilizadas para resolver este problema. Sin embargo, para problemas con gran número de parámetros o casos en que la solución del problema directo sea mucho más costosa, lo conveniente sería obtener una solución y, a continuación, utilizar un método analítico que genere modelos equivalentes sin necesidad de volver a resolver el problema directo. Incluso podrían especificarse algunas características a priori de estos modelos equivalentes.

En este trabajo, se propone buscar esos modelos equivalentes dentro de la región del espacio de modelos donde la aproximación lineal es válida, lo cual reduce drásticamente el número de

 $^{^{68}}$ **m**₀ puede haber sido la solución de otro problema de optimización *distinto*, por ejemplo, uno en el cual el funcional haya sido regularizado. Este enfoque permite encontrar modelos equivalentes **m**₀ en el sentido de que verifiquen el desajuste dentro de cierta tolerancia, pero no estén obligados a mantener las restricciones estructurales que la regularización impone a **m**₀.

éstos que es necesario analizar y, por tanto, el tiempo de cálculo. Cabe recordar que para los modelos de dicha región es válida la siguiente aproximación:

$$E\left(\mathbf{m}\right) = \left\|\mathbf{d} - \mathbf{F}(\mathbf{m})\right\|_{N_D}^2 \approx \left\|\mathbf{d} - \mathbf{F}(\mathbf{m}_0) - \mathbf{J}\mathbf{F}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0)\right\|_{N_D}^2 = E_c\left(\mathbf{m}; \mathbf{m}_0\right). \tag{202}$$

Esta aproximación se obtiene, como se ve, por linealización de la función \mathbf{F} en el entorno del modelo inicial \mathbf{m}_0 . Los modelos linealmente equivalentes buscados, \mathbf{m} , serán aquellos que satisfagan la desigualdad:

$$E_c(\mathbf{m}; \mathbf{m}_0) = \|\mathbf{d} - \mathbf{F}(\mathbf{m}_0) - \mathbf{J}\mathbf{F}_{\mathbf{m}_0}(\mathbf{m} - \mathbf{m}_0)\|_{N_D}^2 \le \varepsilon^2.$$
(203)

Denotando por:

$$\mathbf{W} \cdot (\mathbf{d} - \mathbf{F}(\mathbf{m}_0)) \equiv \mathbf{e}_w,
\mathbf{W} \cdot \mathbf{J} \mathbf{F}_{\mathbf{m}_0} \equiv \mathbf{J}_{0w},
\mathbf{m} - \mathbf{m}_0 \equiv \Delta \mathbf{m},$$
(204)

y siendo $\mathbf{JF}_{\mathbf{m}_0} \in \mathcal{M}(s,n)$ la matriz de la aplicación diferencial de \mathbf{F} en \mathbf{m}_0 , referida a las bases canónicas:

$$\mathcal{J}_0: \mathbb{R}^n \longrightarrow \mathbb{R}^s,$$

la inecuación (203) se expresa abreviadamente:

$$E_c(\mathbf{m}; \mathbf{m}_0) = \|\mathbf{e}_w - \mathbf{J}_{0w} \Delta \mathbf{m}\|_2^2 \le \varepsilon^2, \tag{205}$$

donde la incógnita es $\Delta \mathbf{m}$. Desarrollando esta desigualdad se obtiene:

$$\Delta \mathbf{m}^{t} (\mathbf{J}_{0w}^{t} \mathbf{J}_{0w}) \Delta \mathbf{m} - 2 \mathbf{e}_{w}^{t} \mathbf{J}_{0w} \Delta \mathbf{m} + \mathbf{e}_{w}^{t} \mathbf{e}_{w} \le \varepsilon^{2}. \tag{206}$$

Este problema se puede reformular como: encontrar los $\Delta \mathbf{m}$ que satisfagan la igualdad

$$\Delta \mathbf{m}^{t} (\mathbf{J}_{0w}^{t} \mathbf{J}_{0w}) \Delta \mathbf{m} - 2 \mathbf{e}_{w}^{t} \mathbf{J}_{0w} \Delta \mathbf{m} + \mathbf{e}_{w}^{t} \mathbf{e}_{w} - T^{2} = 0, \tag{207}$$

para algún escalar T^2 menor que ε^2 .

Si existe alguna solución $\Delta \mathbf{m}$ de la ecuación (207), para calcular el modelo equivalente \mathbf{m} se emplea la relación:

$$\mathbf{m} - \mathbf{m}_0 \equiv \Delta \mathbf{m}$$
.

A continuación se seleccionan aquellos que satisfacen además la condición de equivalencia no lineal, es decir, los que verifican:

$$\|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_{N_D}^2 \le \varepsilon^2. \tag{208}$$

La igualdad (207) puede interpretarse como la ecuación asociada a una hipercuádrica en un sistema de referencia cartesiano n-dimensional $\{O, \mathbf{B}_c\}$. Así pues, la generación de modelos geoeléctricos de suelos linealmente equivalentes a \mathbf{m}_0 , es un proceso de determinación de puntos en esa hipercuádrica. Tanto desde el punto de vista teórico como para su cálculo, es conveniente obtener la ecuación de la misma tomando su centro como origen y referida a sus ejes principales⁶⁹. Para la determinación de los ejes principales se empleará la descomposición en valores singulares del operador lineal \mathcal{J}_{0w} , es decir, determinar las matrices \mathbf{U} , $\mathbf{\Sigma}$ y \mathbf{V} tales que:

$$\mathbf{J}_{0w} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^t, \tag{209}$$

 $^{^{69} \}mbox{Considerando}$ que no se trata de una hipercuádrica degenerada.

donde los vectores columna de V constituyen una base ortonormada $\{B_V\}$ del espacio de modelos, el sistema de vectores columna de U es una base ortonormada $\{B_U\}$ del espacio de datos, Σ es la matriz cuyos elementos no nulos son los valores singulares del operador \mathcal{J}_{0w} y \mathbf{J}_{0w} la matriz de este operador referida a las bases canónicas.

Se trata ahora de obtener la ecuación de la hipercuádrica (207) en el sistema de referencia $\{O, \mathbf{B}_V\}$. Se denominará $\Delta \mathbf{m_v}$ al vector $\Delta \mathbf{m}$ expresado en la base $\{\mathbf{B}_V\}$, de modo que:

$$\Delta \mathbf{m} = \mathbf{V} \Delta \mathbf{m}_{\mathbf{v}}.\tag{210}$$

Si se sustituyen las igualdades (210) y (209) en (207), la ecuación de la hipercuádrica en $\{O, \mathbf{B}_V\}$ es:

$$\Delta \mathbf{m}_{\mathbf{v}}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}} - 2 \mathbf{e}_{w}^{t} \left(\mathbf{U} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}} + \mathbf{e}_{w}^{t} \mathbf{e}_{w} - T^{2} = 0. \tag{211}$$

Esta ecuación puede simplificarse si se expresa \mathbf{e}_w como:

$$\mathbf{e}_w = \mathbf{U}\mathbf{e}_u,\tag{212}$$

donde \mathbf{e}_u puede interpretarse como las coordenadas del vector $\mathbf{e}_w \in \mathbb{R}^s$ en la base $\{\mathbf{B}_U\} \subset \mathbb{R}^s$. Por tanto se puede escribir que

$$\mathbf{e}_u = \mathbf{U}^{-1} \mathbf{e}_w = \mathbf{U}^t \mathbf{e}_w = \mathbf{U}^t \mathbf{W} (\mathbf{d} - \mathbf{F}(\mathbf{m}_0)). \tag{213}$$

Sustituyendo la expresión (213), en la igualdad (207), ésta se transforma en:

$$\Delta \mathbf{m}_{\mathbf{v}}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}} - 2 \mathbf{e}_{u}^{t} \mathbf{\Sigma} \Delta \mathbf{m}_{\mathbf{v}} + \mathbf{e}_{w}^{t} \mathbf{e}_{w} - T^{2} = 0.$$
 (214)

La ecuación (214), es la de una hipercuádrica cuyos ejes principales tienen la dirección indicada por los vectores columna de \mathbf{V} . Aun queda por trasladar el sistema de referencia al centro de la hipercuádrica, O_c . Las coordenadas, $\Delta \mathbf{m}_{\mathbf{v}c}$, del centro de la hipercuádrica en el sistema $\{O, \mathbf{B}_V\}$, verifican la ecuación :

$$\left(\mathbf{\Sigma}^{t}\mathbf{\Sigma}\right)\Delta\mathbf{m}_{\mathbf{v}c} - \mathbf{\Sigma}^{t}\mathbf{e}_{u} = \mathbf{0}.$$
 (215)

El sistema anterior puede resolverse de forma explícita utilizando la descomposición por cajas de la matriz Σ según el siguiente procedimiento:

$$\Sigma_{(s,n)} = \begin{pmatrix} \Sigma_{(r,r)} & \mathbf{0}_{(r,n-r)} \\ \mathbf{0}_{(s-r,r)} & \mathbf{0}_{(s-r,n-r)} \end{pmatrix} \Rightarrow \Sigma^{t} \Sigma = \begin{pmatrix} \Sigma_{(r,r)}^{2} & \mathbf{0}_{(r,n-r)} \\ \mathbf{0}_{(n-r,r)} & \mathbf{0}_{(n-r,n-r)} \end{pmatrix}, \tag{216}$$

(217)

$$\Delta \mathbf{m}_{\mathbf{v}c} = \begin{pmatrix} \Delta \mathbf{m}_{\mathbf{v}c}^r \\ \Delta \mathbf{m}_{\mathbf{v}c}^0 \end{pmatrix}, \tag{218}$$

(219)

$$\mathbf{e_u} = \begin{pmatrix} \mathbf{e_u^r} \\ \mathbf{e_u^0} \end{pmatrix}, \tag{220}$$

donde $r = \text{rango}(\mathbf{J}_{0w})$, que coincide con el número de valores singulares no nulos de dicha matriz. $\Delta \mathbf{m}_{\mathbf{v}c}^r$ contiene las primeras r coordenadas del vector $\Delta \mathbf{m}_{\mathbf{v}c}$ y $\Delta \mathbf{m}_{\mathbf{v}c}^0$ las n - r restantes, que pertenecen al núcleo del operador \mathcal{J}_{0w} . $\mathbf{e}_{\mathbf{u}}^r$ contiene las primeras r coordenadas del vector $\mathbf{e}_{\mathbf{u}}$

-modelizables linealmente- y $\mathbf{e}_{\mathbf{u}}^{0}$ las restantes s-r, que no pueden ser tenidas en cuenta por medio de la aproximación lineal⁷⁰. Por lo tanto el sistema (215) puede expresarse como:

$$\begin{pmatrix} \mathbf{\Sigma}_{(r,r)}^{2} & \mathbf{0}_{(r,n-r)} \\ \mathbf{0}_{(n-r,r)} & \mathbf{0}_{(n-r,n-r)} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{m}_{\mathbf{v}c}^{r} \\ \Delta \mathbf{m}_{\mathbf{v}c}^{0} \end{pmatrix} = \begin{pmatrix} \mathbf{\Sigma}_{(r,r)} & \mathbf{0}_{(r,s-r)} \\ \mathbf{0}_{(n-r,r)} & \mathbf{0}_{(n-r,s-r)} \end{pmatrix} \begin{pmatrix} \mathbf{e}_{\mathbf{u}}^{r} \\ \mathbf{e}_{\mathbf{u}}^{0} \end{pmatrix}, \tag{221}$$

que puede descomponerse en dos subsistemas,

$$\Sigma_r^2 \Delta \mathbf{m}_{\mathbf{v}c}^r = \Sigma_r \mathbf{e}_{\mathbf{u}}^r, \tag{222}$$

donde Σ_r es la caja cuadrada $\Sigma_{(r,r)}$ y,

$$\mathbf{0}_{(n-r,n-r)} \Delta \mathbf{m}_{\mathbf{v}c}^{0} = \mathbf{0}_{(n-r,s-r)} \mathbf{e}_{\mathbf{u}}^{0}. \tag{223}$$

La solución del sistema (222) es:

$$\Delta \mathbf{m}_{\mathbf{v}_c}^r = \mathbf{\Sigma}_r^{-2} \mathbf{\Sigma}_r \mathbf{e}_{\mathbf{u}}^r = \Sigma_r^{-1} \mathbf{e}_{\mathbf{u}}^r, \tag{224}$$

donde se ha tenido en cuenta que la matriz Σ_r es regular y diagonal.

Cuando se verifica que n > r, el sistema (221) tiene infinitas soluciones ya que el subsistema (223)se verica trivialmente. Por tanto, no existe centro. No obstante, en el caso de los sondeos eléctricos verticales los problemas con relevancia práctica son siempre sobredeterminados⁷¹, puesto que el número de datos no redundantes tomados en el campo es mayor o igual que el número de parámetros a determinar. Por tanto, en este caso, r = n, así que la solución (224), puede escribirse como:

$$\Delta \mathbf{m}_{\mathbf{v}c}^n = \Sigma_n^{-1} \mathbf{e}_{\mathbf{u}}^n. \tag{225}$$

Dado que $\mathbf{e}_{\mathbf{u}}^{0}$ corresponde a la parte del residuo no modelizable linealmente, éste no puede influir en la determinación de ninguna característica de la hipercuádrica. La ecuación (223) indica que tampoco influye en la posición del centro.

Si se denota con $\Delta \mathbf{m}_{\mathbf{v}f}$ a las coordenadas de un punto del espacio de modelos referido al sistema $\{O_c, \mathbf{B}_V\}$, se cumple que:

$$\Delta \mathbf{m_v} = \Delta \mathbf{m_{vc}} + \Delta \mathbf{m_{vf}}. \tag{226}$$

Sustituyendo esta igualdad en la ecuación (214) y teniendo en cuenta la igualdad (222), se obtiene:

$$\Delta \mathbf{m}_{\mathbf{v}f}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}f} + \left[\mathbf{e}_{w}^{t} \mathbf{e}_{w} - T^{2} + \Delta \mathbf{m}_{\mathbf{v}c}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}c} - 2 \mathbf{e}_{\mathbf{u}}^{t} \mathbf{\Sigma} \Delta \mathbf{m}_{\mathbf{v}c} \right] = 0.$$
 (227)

Introduciendo la notación:

$$I = \Delta \mathbf{m}_{\mathbf{v}c}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}c} - 2 \mathbf{e}_{\mathbf{u}}^{t} \mathbf{\Sigma} \Delta \mathbf{m}_{\mathbf{v}c}, \tag{228}$$

y teniendo en cuenta que

$$\left(\mathbf{\Sigma}^{t}\mathbf{\Sigma}\right)\Delta\mathbf{m}_{\mathbf{v}c} = \mathbf{\Sigma}^{t}\mathbf{e}_{\mathbf{u}},\tag{229}$$

⁷⁰Véase, a este respecto, la explicación acerca de los distintos subespacios sobre los que informa la descomposición en valores singulares.

 $^{^{71}}$ En la práctica habitual el número de datos, s, supera al número de parámetros del modelo, n, aunque en teoría se puede plantear también $s \leq n$.

se deduce que,

$$I = -\mathbf{e}_{\mathbf{u}}^{t} \mathbf{\Sigma} \Delta \mathbf{m}_{\mathbf{v}c}. \tag{230}$$

Además dado que,

$$\mathbf{e}_{\mathbf{u}}^{t} = \begin{pmatrix} \mathbf{e}_{\mathbf{u}}^{r} & \mathbf{e}_{\mathbf{u}}^{0} \end{pmatrix}^{t} \begin{pmatrix} \mathbf{\Sigma}_{r} & \mathbf{0}_{(r,n-r)} \\ \mathbf{0}_{(s-r,r)} & \mathbf{0}_{(s-r,n-r)} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\mathbf{u}}^{r} \end{pmatrix}^{t} \begin{pmatrix} \mathbf{\Sigma}_{r} & \mathbf{0}_{(r,n-r)} \end{pmatrix}, \tag{231}$$

la igualdad (230) puede reescribirse como:

$$I = -\left(\mathbf{e}_{\mathbf{u}}^{r}\right)^{t} \left(\begin{array}{c} \mathbf{\Sigma}_{r} & \mathbf{0}_{(r,n-r)} \end{array}\right) \left(\begin{array}{c} \Delta \mathbf{m}_{\mathbf{v}c}^{r} \\ \Delta \mathbf{m}_{\mathbf{v}c}^{0} \end{array}\right) = -\left(\mathbf{e}_{\mathbf{u}}^{r}\right)^{t} \mathbf{\Sigma}_{r} \Delta \mathbf{m}_{\mathbf{v}c}^{r}. \tag{232}$$

Retomando ahora la igualdad $\Delta \mathbf{m}_{\mathbf{v}c}^r = \mathbf{\Sigma}_r^{-1} \mathbf{e}_{\mathbf{u}}^r$, se llega a:

$$I = -(\mathbf{e}_{\mathbf{u}}^{r})^{t} \Sigma_{r} \Sigma_{r}^{-1} \mathbf{e}_{\mathbf{u}}^{r} = -(\mathbf{e}_{\mathbf{u}}^{r})^{t} \mathbf{e}_{\mathbf{u}}^{r} = -\|\mathbf{e}_{\mathbf{u}}^{r}\|_{2}^{2}.$$
(233)

Es decir, I representa la parte del error no lineal cometido en el punto \mathbf{m}_0 que puede eliminarse mediante modelización lineal, o sea, es posible encontrar un modelo \mathbf{m}_c -el centro de la aproximación cuadrática con la que se está trabajando aquí- que tenga el menor error lineal posible. Este error mínimo, se diferencia del existente en $\mathbf{m} = \mathbf{m}_0$ en el valor hallado para I.

Sustituyendo el valor de I en (227) se obtiene:

$$\Delta \mathbf{m}_{\mathbf{v}f}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}f} + \|\mathbf{e}_{w}\|_{2}^{2} - T^{2} - \|\mathbf{e}_{\mathbf{u}}^{r}\|_{2}^{2} = 0.$$
 (234)

Si se descompone $\mathbf{e}_{\mathbf{u}}$ de la siguiente manera,

$$\mathbf{e}_{\mathbf{u}} = \begin{pmatrix} \mathbf{e}_{\mathbf{u}}^{r} \\ \mathbf{e}_{\mathbf{u}}^{0} \end{pmatrix} = \begin{pmatrix} \mathbf{e}_{\mathbf{u}}^{r} \\ \mathbf{0} \end{pmatrix} + \begin{pmatrix} \mathbf{0} \\ \mathbf{e}_{\mathbf{u}}^{0} \end{pmatrix}, \tag{235}$$

se tiene la siguiente relación entre normas:

$$\|\mathbf{e_u}\|_2^2 = \|\mathbf{e_u}^r\|_2^2 + \|\mathbf{e_u}^0\|_2^2.$$
 (236)

Como, además, la norma euclídea es invariante frente a un cambio ortonormado de base:

$$\|\mathbf{e_w}\|_2^2 = \|\mathbf{e_u}\|_2^2 \tag{237}$$

con lo que la expresión (234) resulta:

$$\Delta \mathbf{m}_{\mathbf{v}f}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}f} + \left\| \mathbf{e}_{\mathbf{u}}^{0} \right\|_{2}^{2} = T^{2}, \tag{238}$$

ο,

$$\left\|\mathbf{\Delta}\mathbf{m}_{\mathbf{v}f}\right\|_{\mathbf{\Sigma}^{t}\mathbf{\Sigma}}^{2} + \left\|\mathbf{e}_{\mathbf{u}}^{0}\right\|_{2}^{2} = T^{2},\tag{239}$$

donde $\|\mathbf{\Delta m_{vf}}\|_{\mathbf{\Sigma}^t \mathbf{\Sigma}}^2$ indica la norma del vector $\mathbf{\Delta m_{vf}}$ derivada del producto escalar de matriz $\mathbf{\Sigma}^t \mathbf{\Sigma}$.

Interpretación geométrica

La relación (239) puede interpretarse en términos de distribución del error lineal:

- lacktriangle la tolerancia T^2 admitida para el desajuste total del problema lineal consta de dos contribuciones:
 - $\|\mathbf{e}_{\mathbf{u}}^{0}\|_{2}^{2}$ representa la parte del desajuste no modelizable mediante la teoría lineal, es decir, $\mathbf{e}_{\mathbf{u}}^{0} \notin \operatorname{Im}(\mathcal{J}_{0w})$. Si \mathbf{m}_{0} es el resultado de la resolución de un problema de optimización no lineal por algún método local, $\|\mathbf{e}_{\mathbf{u}}^{0}\|_{2}^{2}$ es, no sólo el valor del funcional cuadrático de error, sino también, se espera que además

$$\left\|\mathbf{e}_{\mathbf{w}}\right\|_{2}^{2}=\left\|\mathbf{d}-\mathbf{F}(\mathbf{m}_{min})\right\|_{\mathbf{W}}^{2}=\left\|\mathbf{e}_{\mathbf{u}}^{0}\right\|_{2}^{2}.$$

- Si \mathbf{m}_0 no proviene de la solución del problema de minimización asociado, $\|\mathbf{e}_{\mathbf{u}}^0\|_2^2$ es el valor del *mínimo error lineal* cometido, el cual se alcanza en el centro \mathbf{m}_c del hiperparaboloide obtenido al linealizar. Como debe verificarse la igualdad indicada en (236), cabe concluir que el mínimo error *lineal* siempre es menor que el error no lineal cometido en el punto \mathbf{m}_0 .
- $\|\Delta \mathbf{m}_{\mathbf{v}f}\|_{\mathbf{\Sigma}^t \mathbf{\Sigma}}^2$ que es la parte adicional del error lineal correspondiente al cambio de \mathbf{m}_0 a \mathbf{m} .
- ullet Para que existan modelos linealmente equivalentes a ${f m}_0$ es necesario especificar

$$T^2 \ge \left\| \mathbf{e}_{\mathbf{u}}^0 \right\|_2^2$$

Expresión del modelo linealmente equivalente

Una vez que se han calculado las coordenadas de múltiples vectores $\Delta \mathbf{m}_{\mathbf{v}f}$ que verifican la restricción (239), puede calcularse el vector $\Delta \mathbf{m}$ de la siguiente forma:

$$\Delta \mathbf{m} = \mathbf{V} \Delta \mathbf{m}_{\mathbf{v}} = V \left(\Delta \mathbf{m}_{\mathbf{v}c} + \Delta \mathbf{m}_{\mathbf{v}f} \right) = \mathbf{V} \left(\left(\mathbf{\Sigma}^{t} \right)^{-1} \mathbf{e}_{\mathbf{u}} + \Delta \mathbf{m}_{\mathbf{v}f} \right) =$$

$$= V \left(\left(\mathbf{\Sigma}^{t} \right)^{-1} \mathbf{U}^{t} \mathbf{W} \left(\mathbf{d} - \mathbf{F}(\mathbf{m}_{0}) \right) + \Delta \mathbf{m}_{\mathbf{v}f} \right), \tag{240}$$

o sea

$$\mathbf{m} = \mathbf{m}_0 + V\left(\left(\mathbf{\Sigma}^t\right)^{-1} \mathbf{U}^t \mathbf{W} \left(\mathbf{d} - \mathbf{F}(\mathbf{m}_0)\right) + \mathbf{\Delta} \mathbf{m}_{\mathbf{v}f}\right). \tag{241}$$

Cuando $\mathbf{m}_0 = \mathbf{m}_{\min}$, el residuo definitivo $(\mathbf{d} - F(\mathbf{m}_{\min}))$ no puede ser modelado por la aproximación lineal -si no podría conseguirse mayor reducción del error, incompatible con el hecho de que sea la solución final- lo que significa que el vector $(\mathbf{d} - F(\mathbf{m}_0))$ pertenece al complemento ortogonal del subespacio de vectores columna de la matriz \mathbf{U} . Por ello se verifica que:

$$\mathbf{U}^t \mathbf{W} (\mathbf{d} - \mathbf{F}(\mathbf{m}_0)) = \mathbf{0}, \tag{242}$$

y por tanto se cumple que:

$$\Delta \mathbf{m} = V \Delta \mathbf{m}_{\mathbf{v}f},\tag{243}$$

coherente con el hecho de hacer $\Delta \mathbf{m}_{\mathbf{v}c} = \mathbf{0}$ en la ecuación [240].

Procedimiento de cálculo

Se ha implementado el siguiente algoritmo para el cálculo de modelos linealmente equivalentes:

• Cálculo del umbral mínimo de tolerancia: a partir de \mathbf{m}_0 y de la descomposición en valores singulares de \mathbf{J}_{0w} , se determina

$$\left\|\mathbf{e}_{\mathbf{u}}^{0}\right\|_{2}^{2}=E_{c}\left(\mathbf{m}_{\min};\mathbf{m}_{0}\right).$$

Especificación del error lineal asumido por los modelos equivalentes buscados

$$T^2 > E_c\left(\mathbf{m}_{\min}; \mathbf{m}_0\right)$$
.

• Sorteo aleatorio de puntos que verifiquen la ecuación de la hipercuádrica:

$$\Delta \mathbf{m}_{\mathbf{v}f}^{t} \left(\mathbf{\Sigma}^{t} \mathbf{\Sigma} \right) \Delta \mathbf{m}_{\mathbf{v}f} = T^{2} - E_{c} \left(\mathbf{m}_{\min}; \mathbf{m}_{0} \right) = k^{2}.$$
 (244)

Se verifica que:

$$\sum_{i=1}^{n} \left(\frac{\sigma_i}{k} \Delta m_{\mathbf{v}f}^i \right)^2 = 1, \tag{245}$$

donde σ_i es el valor singular i-ésimo de \mathbf{J}_{0w} . Mediante el sorteo se obtiene un vector unitario de cosenos directores que cumplen la igualdad

$$\sum_{i=1}^{n} \cos^2 \theta_i = 1, \tag{246}$$

con lo cual basta identificar:

$$\cos \theta_i = \frac{\sigma_i}{k} \Delta m_{\mathbf{v}f}^i, \ i = 1, ..., n \tag{247}$$

y por tanto

$$\Delta m_{\mathbf{v}f}^i = \frac{k}{\sigma_i} \cos \theta_i, \ i = 1, ..., n.$$

• Cálculo de las coordenadas en la base original según la expresión (241).

6.2.2. Caso II: cálculo de modelos linealmente equivalentes con restricciones

En el apartado anterior se indicaba cómo, a partir de un modelo \mathbf{m}_0 (que podría ser, por ejemplo, una solución \mathbf{m}_{\min} de un problema de optimización), se puede especificar el conjunto de todos los modelos linealmente equivalentes de los cuales, posiblemente, algunos lo sean también en el sentido no lineal. Obtener sin restricciones uno de ellos, se reduce a escoger arbitrariamente un punto de la hipercuádrica de equivalencia.

A veces no se busca simplemente un modelo equivalente sino uno que posea, además, ciertas características introducidas como *información a priori*. Se admite aquí que dichas propiedades son expresables mediante alguna de las metodologías indicadas en el capítulo lineal. En este apartado se indica cuáles de los puntos de esta hipercuádrica satisfacen, si existen, las restricciones de este tipo.

El procedimiento comienza por la consideración en el punto \mathbf{m}_0 de una aproximación cuadrática $E_c(\mathbf{m}; \mathbf{m}_0)$, del funcional de error $E(\mathbf{m}) = \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_W^2$. $E_c(\mathbf{m}; \mathbf{m}_0)$ resulta de reemplazar $\mathbf{F}(\mathbf{m})$ por su aproximación lineal en el entorno del punto \mathbf{m}_0 . Así pues:

$$E_c(\mathbf{m}; \mathbf{m}_0) = E_c(\mathbf{m}_0) - 2\mathbf{e}_w^t \mathbf{J}_{0w} \Delta \mathbf{m} + \Delta \mathbf{m}^t (\mathbf{J}_{0w}^t \mathbf{J}_{0w}) \Delta \mathbf{m}, \tag{248}$$

donde $E_c(\mathbf{m}_0) = \mathbf{e}_w^t \mathbf{e}_w$ se denota el valor de $E_c(\mathbf{m}_0, \mathbf{m}_0)$.

Si el modelo **m** que finalmente se desea obtener, debe incorporar las restricciones adecuadas, su expresión puede escribirse como se ha visto en el capítulo lineal:

$$\mathbf{m} = \mathbf{m}_f + \mathbf{m}_w = \mathbf{m}_f + C\mathbf{m}_l,$$

donde \mathbf{m}_f es un vector y C una matriz determinados por el tipo de información a priori; \mathbf{m}_l es un vector que puede moverse libremente por todo el espacio de modelos. Así pues:

$$\Delta \mathbf{m} = \mathbf{m} - \mathbf{m}_0 = (\mathbf{m}_f - \mathbf{m}_0) + C\mathbf{m} = \Delta \mathbf{m}_f^* + C\mathbf{m}_l. \tag{249}$$

Si la igualdad (249) se introduce en la ecuación (248) y se desarrolla, se obtiene:

$$E_c(\mathbf{m}, \mathbf{m}_0) = E_c(\mathbf{m}_f, \mathbf{m}_0) - \nabla E_c(\mathbf{m}_f, \mathbf{m}_0) \cdot \mathbf{m}_w + \mathbf{m}_w^t (\mathbf{J}_{0w}^t \mathbf{J}_{0w}) \mathbf{m}_w,$$

donde:

$$E_c(\mathbf{m}_f, \mathbf{m}_0) = E_c(\mathbf{m}_0) - 2\mathbf{e}_w^t \mathbf{J}_{0w} \Delta \mathbf{m}_f^* + \Delta \mathbf{m}_f^{*t} (\mathbf{J}_{0w}^t \mathbf{J}_{0w}) \Delta \mathbf{m}_f^*,$$

$$\nabla E_c(\mathbf{m}_f, \mathbf{m}_0) \cdot \mathbf{m}_w = 2 \left(\mathbf{e}_w - \mathbf{J}_{0w} \Delta \mathbf{m}_f^* \right)^t \mathbf{J}_{0w} \mathbf{m}_w.$$

siendo $\mathbf{m}_w^t(\mathbf{J}_{0w}^t\mathbf{J}_{0w})\mathbf{m}_w$ el término de segundo grado.

El vector $\mathbf{m}_w \in W \subset \mathbb{R}^n$ puede ser expresado respecto a una base local del subespacio vectorial W. Se denominará $\{\mathbf{B}_P\} \subset W$ a dicha base. Se puede definir una aplicación lineal:

$$\mathcal{P}: W \subset \mathbb{R}^p \longrightarrow \mathbb{R}^n. \tag{251}$$

que a todo vector de W expresado en la base $\{\mathbf{B}_P\}$ le asocia el mismo vector referido a la base canónica de \mathbb{R}^n . Denominando $\mathbf{P} \in \mathcal{M}(n,p)$ a la matriz de dicha aplicación lineal \mathcal{P} , se tiene entonces:

$$\mathbf{m}_w = C\mathbf{m}_l = \mathbf{P}\boldsymbol{\beta},\tag{252}$$

donde $\beta \in \mathbb{R}^p$ contiene por tanto las coordenadas del vector \mathbf{m}_w expresado en la base $\{\mathbf{B}_P\}$. De modo que la expresión (??) en la nueva base es:

$$E_c(\mathbf{m}, \mathbf{m}_0) = E_c(\mathbf{m}_f, \mathbf{m}_0) - 2\left(\mathbf{e}_w - \mathbf{J}_{0w}\Delta\mathbf{m}_f^*\right)^t \mathbf{J}_{0w}\mathbf{P}\boldsymbol{\beta} + (\mathbf{P}\boldsymbol{\beta})^t (\mathbf{J}_{0w}^t \mathbf{J}_{0w})\mathbf{P}\boldsymbol{\beta}.$$
(253)

Dado que $\mathbf{J}_{0w} \in \mathcal{M}(m,n)$ es la matriz de la diferencial, referida a las bases canónicas:

$$\mathcal{J}_0: \mathbb{R}^n \longrightarrow \mathbb{R}^s$$
,

se define la aplicación

$$\mathcal{JP} \equiv \mathcal{J}_0 \circ \mathcal{P} : W \longrightarrow \mathbb{R}^s$$
,

cuya matriz es, $(\mathbf{J}_{0w}\mathbf{P}) \in \mathcal{M}(m,p)$, respecto a las bases $\{\mathbf{B}_P\} \subset W$ y la base canónica de \mathbb{R}^s . La descomposición en valores singulares de este operador es:

$$\mathbf{J}_{0w}\mathbf{P} = \overline{\mathbf{U}}\boldsymbol{\Lambda}\overline{\mathbf{V}}^t \tag{254}$$

donde $\overline{\mathbf{U}} \in \mathcal{M}(s,s)$, $\overline{\mathbf{V}} \in \mathcal{M}(p,p)$ y $\mathbf{\Lambda} \in \mathcal{M}(s,p)$. Además, el rango de la matriz $\mathbf{\Lambda}$ es p. Se hace notar que cada uno de los vectores columna de $\overline{\mathbf{V}}$, viene expresado por sus coordenadas en la base $\{\mathbf{B}_P\} \subset W$.

El vector $\boldsymbol{\beta}$ que estaba expresado en la base $\{\mathbf{B}_P\}\subset W$, se expresará ahora en una nueva base, la formada por los vectores columna de $\overline{\mathbf{V}}$. En esta base, las coordenadas de $\boldsymbol{\beta}$ se denotan $\boldsymbol{\beta}_v$, de modo que:

$$\boldsymbol{\beta} = \overline{\mathbf{V}}\boldsymbol{\beta}_{v}.\tag{255}$$

Sustituyendo (255)y (254) en la expresión (253) se obtiene:

$$E_c\left(\mathbf{m}, \mathbf{m}_0\right) = E_c\left(\mathbf{m}_f, \mathbf{m}_0\right) - 2\left(\mathbf{e}_w - \mathbf{J}_{0w} \Delta \mathbf{m}_f^*\right)^t \overline{\mathbf{U}} \mathbf{\Lambda} \boldsymbol{\beta}_v + \boldsymbol{\beta}_v^t \left(\mathbf{\Lambda}^t \mathbf{\Lambda}\right) \boldsymbol{\beta}_v.$$

Si se tiene en cuenta que $\left(\mathbf{e}_w - \mathbf{J}_{0w} \Delta \mathbf{m}_f^*\right)$ es un vector de \mathbb{R}^s , conviene hallar su expresión, $\boldsymbol{\mu}$, en la base ortonormada $\left\{\mathbf{B}_{\overline{U}}\right\}$. En este caso se tiene la siguiente relación:

$$\mathbf{e}_w - \mathbf{J}_{0w} \Delta \mathbf{m}_f^* = \overline{\mathbf{U}} \boldsymbol{\mu}, \tag{256}$$

con lo cual $E_c(\mathbf{m}, \mathbf{m}_0)$ toma la forma:

$$E_c(\mathbf{m}, \mathbf{m}_0) = E_c(\mathbf{m}_f, \mathbf{m}_0) - 2\mu^t \Lambda \beta_v + \beta_v^t (\Lambda^t \Lambda) \beta_v.$$
 (257)

El punto sobre la restricción en el que esta función alcanza el mínimo es solución del sistema de ecuaciones:

$$(\mathbf{\Lambda}^t \mathbf{\Lambda}) \, \boldsymbol{\beta}_{vc} = \mathbf{\Lambda}^t \boldsymbol{\mu} \tag{258}$$

donde el vector $\boldsymbol{\beta}_{vc} \in W$ representa las coordenadas de dicho punto en la base $\overline{\mathbf{V}}$. Por tanto, se cumple la relación:

$$\boldsymbol{\beta}_v = \boldsymbol{\beta}_{vc} + \overline{\boldsymbol{\beta}_v},\tag{259}$$

siendo $\overline{\beta_v}$ las coordenadas de un punto genérico sobre la restricción cuando el origen del sistema de referencia es el punto β_{vc} . Reemplazando en (257) resulta:

$$E_{c}\left(\mathbf{m},\mathbf{m}_{0}\right)=E_{c}\left(\mathbf{m}_{f},\mathbf{m}_{0}\right)-2\boldsymbol{\mu}^{t}\boldsymbol{\Lambda}\boldsymbol{\beta}_{vc}+\boldsymbol{\beta}_{vc}^{t}\left(\boldsymbol{\Lambda}^{t}\boldsymbol{\Lambda}\right)\boldsymbol{\beta}_{vc}-2\boldsymbol{\mu}^{t}\boldsymbol{\Lambda}\overline{\boldsymbol{\beta}_{v}}+2\boldsymbol{\beta}_{vc}^{t}\left(\boldsymbol{\Lambda}^{t}\boldsymbol{\Lambda}\right)\overline{\boldsymbol{\beta}_{v}}+\overline{\boldsymbol{\beta}_{v}}^{t}\left(\boldsymbol{\Lambda}^{t}\boldsymbol{\Lambda}\right)\overline{\boldsymbol{\beta}_{v}}.$$

El sentido geométrico de esta expresión puede aclararse explicando el significado de algunas agrupaciones de términos:

$$E_{c}\left(\mathbf{m}_{c}, \mathbf{m}_{0}\right) = E_{c}\left(\mathbf{m}_{f}, \mathbf{m}_{0}\right) - 2\boldsymbol{\mu}^{t}\boldsymbol{\Lambda}\boldsymbol{\beta}_{vc} + \boldsymbol{\beta}_{vc}^{t}\left(\boldsymbol{\Lambda}^{t}\boldsymbol{\Lambda}\right)\boldsymbol{\beta}_{vc},$$

$$\nabla E_{c}\left(\mathbf{m}_{c}, \mathbf{m}_{0}\right) \cdot \left(\mathbf{m} - \mathbf{m}_{c}\right) = -2\boldsymbol{\mu}^{t}\boldsymbol{\Lambda}\overline{\boldsymbol{\beta}_{v}} + 2\boldsymbol{\beta}_{vc}^{t}\left(\boldsymbol{\Lambda}^{t}\boldsymbol{\Lambda}\right)\overline{\boldsymbol{\beta}_{v}} = 0,$$
(260)

más el término de segundo grado resultante

$$\overline{m{eta}_{v}}^{t}\left(m{\Lambda}^{t}m{\Lambda}\right)\overline{m{eta}_{v}}.$$

Por tanto

$$E_c(\mathbf{m}, \mathbf{m}_0) = E_c(\mathbf{m}_c, \mathbf{m}_0) + \overline{\beta_v}^t(\mathbf{\Lambda}^t \mathbf{\Lambda}) \overline{\beta_v}.$$
 (261)

A la hora de calcular modelos \mathbf{m} equivalentes desde el punto de vista no lineal a \mathbf{m}_0 , se trata de imponer que:

$$|E\left(\mathbf{m}\right) - E\left(\mathbf{m}_0\right)| < T^2,$$

siendo $E(\mathbf{m}) = \|\mathbf{d} - \mathbf{F}(\mathbf{m})\|_W^2$ y T la tolerancia que sea oportuno considerar. La condición de equivalencia no lineal es, pues:

$$E(\mathbf{m}_0) - T^2 < E(\mathbf{m}) < E(\mathbf{m}_0) + T^2.$$
 (262)

Las tolerancias de interés son del orden de algunas unidades por ciento y los modelos \mathbf{m}_0 que usualmente se consideran interesantes son el resultado de un proceso de optimización previo que, supuestamente, ha llevado el valor de $E\left(\mathbf{m}_0\right)$ por debajo del valor de T^2 . Así pues, el extremo izquierdo de la inecuación (262) puede aproximarse por 0. A veces, el valor de $E\left(\mathbf{m}_0\right)$ puede resultar pequeño respecto a la tolerancia estipulada o, sencillamente, se quisieran obtener todos los modelos para los cuales $E\left(\mathbf{m}\right) < T^2$. En ambos casos se puede reescribir la desigualdad (262) como,

$$0 < E\left(\mathbf{m}\right) < T^2.$$

Si se reemplaza $E(\mathbf{m})$ por $E_c(\mathbf{m}, \mathbf{m}_0)$, los modelos que cumplan

$$0 < E_c\left(\mathbf{m}, \mathbf{m}_0\right) < T^2,$$

se denominan linealmente equivalentes a \mathbf{m}_0 . Con ayuda de la expresión (261) resulta que,

$$0 < \overline{\beta_v}^t (\Lambda^t \Lambda) \overline{\beta_v} < T^2 - E_c (\mathbf{m}_c, \mathbf{m}_0).$$

Para que existan modelos linealmente equivalentes debe, evidentemente, ocurrir que

$$T^2 \ge E_c\left(\mathbf{m}_c, \mathbf{m}_0\right)$$
.

puesto que $\Lambda^t \Lambda$ es siempre (semi)definida positiva.

Expresión de los modelos linealmente equivalentes

Una vez han sido obtenidas las coordenadas $\overline{\beta_v}$ correspondientes a los modelos linealmente equivalentes, es preciso deshacer las transformaciones realizadas a fin de disponer de las coordenadas de los vectores que representan a dichos modelos en la base canónica inicial, de forma que:

$$\mathbf{m} = \mathbf{m}_f + \mathbf{P}\overline{\mathbf{V}} \left(\boldsymbol{\beta}_{vc} + \overline{\boldsymbol{\beta}_v} \right), \tag{263}$$

donde $\boldsymbol{\beta}_{vc} = \Lambda_p^{-1} \boldsymbol{\mu}_p$.

6.2.3. Interés de la matriz jacobiana en el caso de parametrización logarítmica

El comportamiento del problema inverso lineal en S.E.V viene definido por la estructura de la matriz de sensibilidad o matriz jacobiana **JF**. Se pretende analizar aquí en qué partes de la misma se fundamentan las reglas de equivalencia de Maillet. Con este fin, el uso de la parametrización logarítmica en el espacio de modelos es especialmente adecuado.

Supóngase que se ha determinado, por resolución del problema inverso, un terreno de tres capas

$$\mathbf{m}_0 = [\rho_1, t_1, \rho_2, t_2, \rho_3].$$

Como la parametrización es logarítmica se tiene que:

$$\mathbf{m}_{0 \log} = [\log(\rho_1), \log(t_1), \log(\rho_2), \log(t_2), \log(\rho_3)].$$

El paso a un nuevo modelo, candidato a ser equivalente a \mathbf{m}_0 según cierta tolerancia T previamente especificada, queda definido por el vector $\Delta \mathbf{m}_{0 \log}$ cuya expresión es:

$$\boldsymbol{\Delta m_{0 \log}} = \left[\log\left(\rho_{1}\right), \log\left(t_{1}\right), \log\left(\frac{\rho_{2} + \Delta \rho_{2}}{\rho_{2}}\right), \log\left(\frac{t_{2} + \Delta t_{2}}{t_{2}}\right) \log, \log\left(\rho_{3}\right)\right].$$

Se analiza ahora cómo se reflejan las condiciones de equivalencia mencionadas por Maillet en la estructura de $\Delta m_{0 \log}$:

1. Si
$$\rho_2 \ll \rho_3$$
 y

$$\frac{t_2}{\rho_2} \simeq \frac{t_2 + \Delta t_2}{\rho_2 + \Delta \rho_2},\tag{264}$$

se presenta una de las situaciones contempladas como equivalentes. La relación (264) implica

$$\log\left(\frac{\rho_2 + \Delta\rho_2}{\rho_2}\right) \simeq \log\left(\frac{t_2 + \Delta t_2}{t_2}\right),\tag{265}$$

es decir, la cuasi-igualdad de dos términos consecutivos (de la misma capa) en $\Delta m_{0 \log}$.

2. Si
$$\rho_2 \gg \rho_3$$
 y

$$(t_2 + \Delta t_2) \left(\rho_2 + \Delta \rho_2\right) \simeq \rho_2 t_2,\tag{266}$$

entonces se está ante el otro de los casos de equivalencia. La relación (266) implica

$$\log\left(\frac{\rho_2 + \Delta\rho_2}{\rho_2}\right) \simeq -\log\left(\frac{t_2 + \Delta t_2}{t_2}\right),\tag{267}$$

es decir, los dos términos consecutivos (de la misma capa) en $\Delta m_{0 \log}$ poseen casi el mismo valor absoluto **pero con signo contrario.**

Por tanto si el modelo $\mathbf{m}_{\log} = \mathbf{m}_{0\log} + \Delta \mathbf{m}_{0\log}$, es equivalente a $\mathbf{m}_{0\log}$, inicialmente obtenido, los dos términos consecutivos correspondientes a la capa aludida de $\Delta \mathbf{m}_{0\log}$ deben tener el mismo valor absoluto. Utilizando los vectores \mathbf{v}_i procedentes de la $\mathbf{d.v.s}$ de \mathbf{JF} , el vector $\Delta \mathbf{m}_{0\log}$ se expresa:

$$\Delta \mathbf{m}_{0\log} = \sum_{i=1}^{i=n} \alpha_i \mathbf{v}_i,$$

para ciertos escalares α_i . Como la capa analizada es la segunda, los parámetros de $\Delta m_{0 \log}$ implicados son el tercero y el cuarto. Las condiciones especificadas se traducen en,

$$\sum_{i=1}^{i=n} \alpha_i \left(|\mathbf{v}_{i,3}| - |\mathbf{v}_{i,4}| \right) \simeq 0.$$
(268)

En conclusión, siempre que las componentes de los vectores \mathbf{v}_i correspondientes a los parámetros de la misma capa tengan valores absolutos muy similares, cualquier cambio en el modelo -incluso de gran magnitud- dejará prácticamente inalterado el valor del cociente o del producto de los dos parámetros (situación de equivalencia). Por otro lado, las desviaciones de la igualdad en (268), deshacen rápidamente la equivalencia en índices i bajos, de cuyas componentes dependen mucho las predicciones del modelo y, por tanto, la posibilidad de exceder la tolerancia especificada en la equivalencia. Así pues, los modelos equivalentes que aporten cambios apreciables en la solución obtenida están asociadas a valores de índices i altos y, por tanto, a valores singulares pequeños.

Se han estudiado, además, los vectores asociados a las bases de la **d.v.s** de **JF** para comprobar si, como se sostiene repetidamente en la literatura -véase (Hansen, 1998, y Hanke, 2002, p.ej.)- a medida que crecen los índices i, y disminuyen los valores singulares σ_i los vectores \mathbf{u}_i y \mathbf{v}_i representan componentes de más alta frecuencia de la solución. Para ello se ha investigado el problema asociado a varios suelos, uno de los cuales es el terreno de cuatro capas con resistividades (10, 20, 1, 50, 2) y espesores (5, 5, 5, 5). Los datos del problema, superpuestos a la curva de resistividades aparentes quedan representados en la figura 48. Los valores singulares se han calculado y representado en la figura 49, donde se observa su carácter rápidamente decreciente.

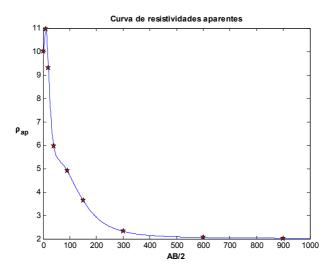


Figura 48: Medidas generadas sintéticamente.

Los resultados se han comparado con los correspondientes a un terreno idéntico, pero parametrizado de otra forma. Cada uno de los estratos del suelo inicial, excepto el último que tiene extensión ilimitada, se ha dividido en dos capas de resistividad idéntica a la inicial. El nuevo vector de resistividades es (10, 10, 20, 20, 1, 1, 50, 50, 2) y el de potencias (5, 5, 5, 5, 5, 5, 5, 5, 5). La

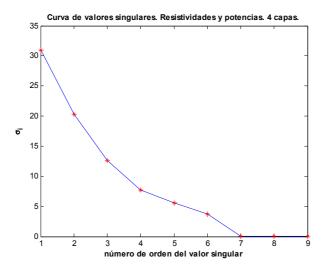


Figura 49: Curva de valores singulares. Las incógnitas son las resistividades y los espesores

principal diferencia es que los espesores se consideran ahora fijos y las incógnitas del problema son sólo las resistividades. El número de grados de libertad (9), es idéntico al del caso anterior. La curva de valores singulares del nuevo problema se representa en la figura 50. El patrón -y los valores- se mantienen.

A continuación se muestran las figuras -representativas del conjunto de la experimentaciónsobre las que pueden sacarse conclusiones. En las figuras 51 y 52 se muestran los vectores singulares \mathbf{u}_i en ambos casos. No parece haber diferencias apreciables.

En las figuras $\mathbf{54}$ y $\mathbf{53}$ aparecen representados los vectores singulares \mathbf{v}_i . Se puede observar que:

- 1. En la formulación tradicional -en resistividades y espesores- no se cumple que al aumentar el índice i, las \mathbf{v}_i asociadas representen componentes de la solución de mayor frecuencia.
- 2. En la formulación en resistividades -con espesores fijados-si ocurre así. Las \mathbf{v}_i asociadas a valores singulares menores aportan los detalles de frecuencia creciente.
- 3. Como justificación de la discrepancia observada se aduce que, tal como se ha visto anteriormente, la presencia de equivalencias se debe a la existencia de parámetros contiguos que, al variar, interactúan mutuamente compensando sus efectos. La correlación mutua aparece en la forma antes indicada en la matriz V y evita la introducción de nuevos grados de libertad. Esta causa desaparece al fijar los espesores.

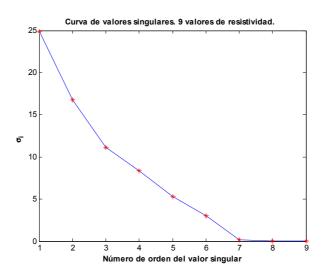


Figura 50: Curva de valores singulares. Las incógnitas son sólo las resistividades

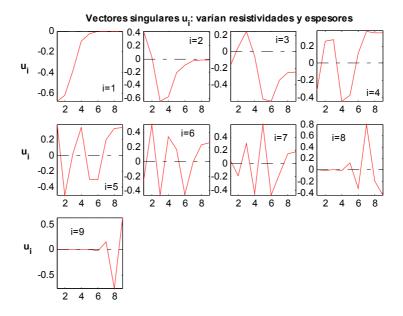


Figura 51: Vectores singulares del espacio de datos: incógnitas resistividades y espesores

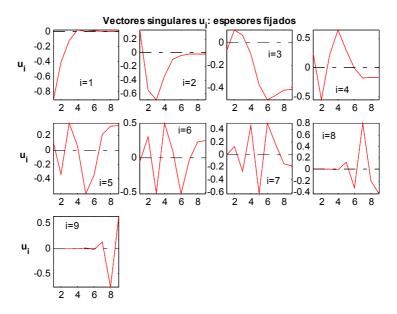


Figura 52: Vectores singulares del espacio de datos: las incógnitas son las resistividades

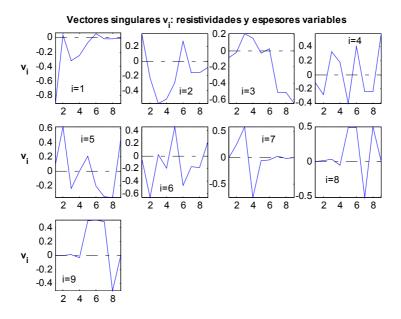


Figura 53: Vectores singulares para los modelos: las incógnitas son resistividades y espesores

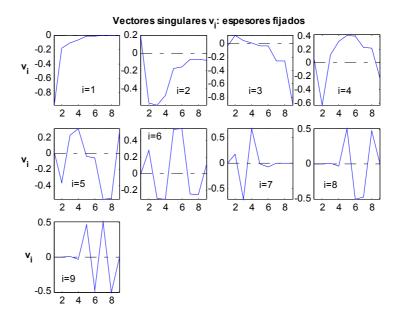
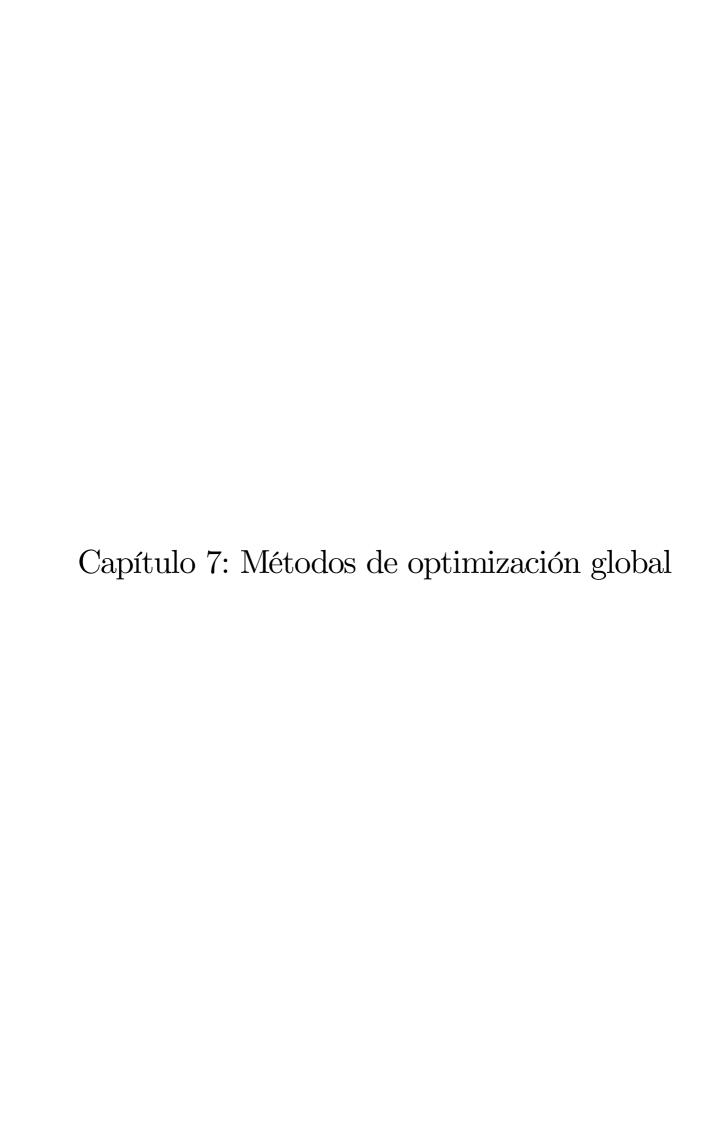


Figura 54: Vectores singulares de los modelos: las incógnitas son las resistividades



7. MÉTODOS DE OPTIMIZACIÓN GLOBAL

7.1. Introducción

Los métodos de optimización global son métodos numéricos probabilísticos que exploran potencialmente todo el espacio de modelos. Permiten eludir la limitación inherente a los métodos locales relativa al entrampamiento en óptimos locales. Su interés radica, como se verá, no sólo en la garantía de una mejor aproximación al óptimo global, sino en que permiten enfocar la existencia de las equivalencia de forma que se puedan extraer conclusiones útiles en términos de probabilidades para la toma de decisiones.

7.2. Algoritmos genéticos

Los algoritmos genéticos han sido una de las ramas de investigación más fructíferas en el campo de la optimización en los últimos años. Ha habido un enorme número de artículos y conferencias sobre el tema desde que Holland expusiese por primera vez la idea en los años 70. El éxito inicial que estas técnicas tuvieron al abordar problemas de difícil resolución por métodos clásicos, las popularizó rápidamente.

Los algoritmos genéticos son una técnica que pertenece a una rama de la investigación aplicada denominada genéricamente computación evolutiva. Este término se inspira en la base teórica de los procesos naturales asociados a la evolución biológica, en particular con la teoría darwinista de selección natural. Según esta teoría, las especies evolucionan hacia estructuras genéticas mejor adaptadas al entorno mediante la transformación del código genético indidual por mutación y recombinación sexual, siendo inviables los cambios introducidos que perjudiquen al individuo en el medioambiente en el que vive.

Los algoritmos genéticos, a pesar de su uso especialmente indicado para problemas discretos, se han utilizado extensivamente para funciones continuas. Son notablemente útiles para problemas discontinuos, no diferenciables o mal condicionados. A diferencia del algoritmo de simulated annealing, en los algoritmos genéticos no es un único modelo el que cambia en una dirección de cambio estipulada, sino que es toda una población de modelos la que se transforma, de una iteración a otra, evolucionando hacia el óptimo global. Los algoritmos genéticos pueden verse como una variante de los métodos de multicomienzo en la cual los individuos de la población de soluciones potenciales no siguen un proceso clásico de optimización local, sino leyes que pretenden simular los principios admitidos para la evolución natural.

7.2.1. Descripción básica del funcionamiento de un A.G

Se parte de una población inicial, generada normalmente de modo aleatorio, cada uno de cuyos individuos se representa mediante una cadena de elementos de un *alfabeto*, cadena denominada *cromosoma*. En términos de un problema de optimización, cada cromosoma representa un modelo candidato a resolver éste.

A cada una de las posiciones que existen en el cromosoma se la denomina locus genético y, a la información allí situada, gen. Puede haber distintas formas del mismo gen, los denominados alelos. En este contexto, los alelos son los distintos elementos del alfabeto escogido.

La elección del alfabeto -Λ- para codificar las variables del problema, así como la longitud del cromosoma son cuestiones delicadas que influyen en el éxito de la técnica. El alfabeto binario,

por ejemplo, emplea los símbolos $\Lambda_{binario} = \{0, 1\}$. En este alfabeto, cada modelo del espacio de parámetros quedaría representado (codificado) mediante una cadena -cromosoma- de ceros y unos:

modelo $1 \longmapsto$	1	0	1	 0
modelo $2 \longmapsto$	0	1	0	 1

La utilización de otros alfabetos también es posible, por ejemplo, puede utilizarse como alfabeto el conjunto de todos los números reales - $\Lambda_{aritmetico} \equiv \mathbb{R}$ - (codificación que denominaremos aritmética) o, incluso, variables no numéricas.

El esquema fundamental de funcionamiento de los algoritmos genéticos es el siguiente:

- Se parte de una población inicial, cuyo tamaño -número de cromosomas- es necesario especificar. A partir de ésta se van produciendo sucesivas poblaciones -generaciones- que tienen, normalmente, el mismo tamaño que la inicial. Este número tiene que ser lo suficientemente pequeño para que el proceso sea manejable y lo bastante grande como para que la población considerada tenga posibilidades de explorar todo el espacio de búsqueda (véase más adelante el epígrafe tamaño de la población).
- A cada individuo -cromosoma- de la población hay que atribuirle un valor que lo ordene respecto a los demás. Este valor se denomina aptitud o idoneidad. Para este fin se utiliza una función relacionada con la función objetivo del problema de optimización. Si el problema es de maximización, la propia función objetivo puede servir como criterio para asignar aptitudes: a mayor valor de la función objetivo, mayor aptitud. Sin embargo, si el problema es de minimización, la función de aptitud no puede ser la función objetivo, puesto que el algoritmo genético pretenderá siempre maximizar la función de aptitud (i.e maximizar la función objetivo), sino que deberá ser una función relacionada con ella según distintos procedimientos que se discutirán más adelante y que denominaremos de escalado.

La función escogida como criterio de aptitud toma sus valores, en muchos problemas de optimización, sobre un espacio de parámetros (subconjunto de \mathbb{R}^n) no codificado con el alfabeto Λ . Por ello, para que las aptitudes puedan ser evaluadas, es necesario descodificar los cromosomas con anterioridad a la evaluación de las mismas.

- De acuerdo con los valores de la aptitud, se procede a la selección de los cromosomas más aptos, que son los que intervendrán en el proceso de reproducción y, a partir de los cuales, se construirá la siguiente generación. Existen distintos procedimientos de selección, así como distintos criterios (criterios de reemplazo) para determinar qué individuos de la generación actual son sustituidos en la siguiente.
 - El proceso de transformación de una generación de cromosomas en la siguiente se lleva a cabo a través de la aplicación de distintos operadores genéticos. Intervienen, esencialmente, dos tipos de operadores: la recombinación o cruce y la mutación.
- A la nueva población obtenida se le aplica el mismo proceso previo hasta que se cumplan los criterios de parada especificados. Normalmente, éstos tienen que ver con que se haya alcanzado un número estipulado de generaciones; con que se hayan explorado un número determinado de individuos o con que haya sido alcanzado un grado dado de aptitud por un número predeterminado de individuos.

Una vez explicado de forma somera e intuitiva el proceso, es conveniente formalizar estas ideas de forma más rigurosa.

Definición formal de AG

Para que un algoritmo se pueda catalogar como algoritmo genético necesita los siguientes elementos:

- Una población de individuos cada uno de los cuales representa una posible solución al problema.
- Una función que hace las veces de entorno⁷², permitirá clasificar a los individuos como aptos o no aptos según una determinada escala de selección.
- Unos operadores genéticos que alteren la composición de los individuos durante la creación de la siguiente generación.
- Un conjunto de parámetros como tamaño de la población, probabilidades asociadas a los operadores, etc.

Todos estos elementos aparecen en la siguiente definición: un algoritmo genético es un algoritmo probabilístico especificado por la n-upla $(N, P, f, \Theta, \Omega, \Psi, \tau)$ donde:

- P es una población de N individuos o cromosomas, $P = \{S_1, S_2, ..., S_N\}$. Cada individuo S_i es una cadena de enteros⁷³, de longitud especificada, pertenecientes al alfabeto Λ .
- f es la función de aptitud que asigna un valor real a cada individuo.

$$f: S_i \to \mathbb{R}^+, \quad i = 1, ..., N$$

ullet Θ es un operador de selección que selecciona u individuos de P.

$$\Theta: P \to \{P_1, ..., P_n\}$$

lacktriangledown Ω es un conjunto de operadores genéticos capaces de producir v individuos nuevos a partir de los seleccionados por Θ

$$\Omega = \{\Omega_c \text{ (cruce)}, \ \Omega_m \text{ (mutación)}, ...\} : \{P_1, ..., P_u\} \rightarrow \{C_1, ..., C_v\}$$

• Ψ es un operador de borrado que elimina v individuos (tantos como han sido creados) de la población P.

au es un criterio de terminación que determina cuándo el proceso evolutivo debe finalizar.

Con estos elementos se define la población en un instante dado en función de la población de instante anterior como:

$$P(t+1) = P(t) - \Psi(P(t)) + \{C_1, ..., C_v\}$$

donde los signos -y + tienen el significado de eliminar y añadir respectivamente.

⁷²Es decir sustituye, en la metáfora del proceso de evolución biológica, al medio ambiente, el cual selecciona los individuos más adaptados al mismo.

 $^{^{73}{\}rm Muy}$ frecuentemente los elementos del alfabeto Λ son números enteros.

7.2.2. Funcionamiento detallado de los algoritmos genéticos y bases matemáticas

Para analizar el funcionamiento de los algoritmos genéticos, Holland (1992) introdujo el concepto de esquema. Un esquema es un conjunto de individuos-cromosomas-con similitudes en ciertas posiciones. Holland sugirió esta definición porque, en el proceso de evolución de un algoritmo genético, se observó que cromosomas con ciertos parecidos se conservaban en las sucesivas generaciones.

Para ejemplificar la noción de esquema, supongamos que la codificación se realiza con un alfabeto binario- $\Lambda_{binario} = \{0, 1\}$ -. Se introduce ahora, además, el símbolo * para denotar al conjunto de todos los elementos del alfabeto, es decir * $\equiv \{0, 1\}$. Se define un esquema como una cadena en el nuevo alfabeto ampliado, $\Lambda_{ampliado} = \{0, 1, *\}$.

Los esquemas no son todos iguales. Se diferencian mediante los conceptos de orden y de longitud de definición. El orden de un esquema H, representado como o(H) es el número de alelos que están fijados en dicho esquema. O sea, el número de elementos que \mathbf{no} son *. Por ejemplo o(*,*,1,*,*,*,0)=2. La longitud de definición de un esquema $\delta(H)$ es la distancia, medida en número de posiciones, entre el primero y el último de los alelos que SÍ han sido fijados.

El concepto de esquema, junto a los de orden y longitud de definición permiten examinar cómo afecta el proceso de reproducción y la acción de los operadores genéticos a los esquemas representados por los cromosomas presentes en la población. Aporta algo de luz al comportamiento de los algoritmos genéticos.

Algoritmos de Selección

La selección dirige el proceso de búsqueda en el espacio de posibles soluciones. El mecanismo de selección asigna a cada individuo una probabilidad de reproducirse, es decir, de permanecer en la siguiente generación, que está en relación directa con su aptitud. Los mecanismos de selección tienden a concentrar la exploración en determinadas regiones del espacio de búsqueda, cada vez más prometedoras desde el punto de vista de la aptitud de sus cromosomas miembros. El objetivo de los algoritmos de selección es conseguir un balance entre la búsqueda en la región más prometedora y el resto de regiones de forma que se evite la convergencia demasiado temprana a una buena zona pero que sólo contenga un óptimo local

Existen varios tipos de mecanismos de selección que asignan a cada individuo una determinada probabilidad de reproducción -es decir, de contribuir a la siguiente generación. Una vez determinadas las probabilidades individuales, el mecanismo selecciona un cierto número de progenitores mediante sorteo. La técnica para hacer el muestreo es del tipo "ruleta rusa": a cada individuo se le asigna un subintervalo dentro del intervalo [0,1] de amplitud igual a su probabilidad; a continuación se genera un número aleatorio en [0,1] y se escoge el individuo en cuyo intervalo haya caído el número aleatorio. Es decir, se define:

$$P_0 = 0,$$

$$P_i = \sum_{j=0}^{i} P_j \Longrightarrow P_N = 1.$$
(269)

Al individuo S_i se le asigna el intervalo $[P_{i-1}, P_i]$. Se genera un número aleatorio y se escoge el individuo correspondiente al intervalo al que pertenece dicho número. Este proceso se repite

tantas veces como sea necesario para obtener el número de individuos especificado en el algortimo de selección. Con este procedimiento los individuos seleccionados pueden repetirse.

Los mecanismos de selección más importantes difieren en la forma de realizar la asignación de probabilidades:

1. Selección proporcional

A cada individuo se le asigna una probabilidad de selección proporcional a su aptitud:

$$p(S_i, t) = \frac{f(S_i, t)}{\sum_{i=1}^{N} f(S_j, t)} \qquad i = 1, ..., N,$$

donde $f(S_i, t)$ es la aptitud del cromosoma S_i en la generación t.

Un gran inconveniente que presenta este mecanismo es su alta dependencia del valor concreto de la aptitud, dado que si, por ejemplo, existiese un individuo cuyo valor de aptitud es mucho más alto que el del resto, pero sin ser óptimo, este tipo de sorteo hace muy probable que este individuo sea seleccionado varias veces. El cromosona de dicho individuo dominará rápidamente la población, causando lo que se denomina convergencia prematura hacia un óptimo local. Para evitar este suceso, deben asignarse las probabilidades de selección de forma que no dependan tanto de las aptitudes atribuidas durante la evaluación. Para ello se utiliza por ejemplo la selección por posición.

2. Selección por posición

En este método los individuos son ordenados en orden ascendente de aptitud, esto es:

$$f(S_1, t) \le f(S_2, t) \le ... \le f(S_N, t)$$
.

La probabilidad de que un individuo sea seleccionado se atribuye en función de la posición que éste ocupa éste en dicho reordenamineto de aptitudes. La probabilidad de seleccionar el individuo i será:

$$p(S_i, t) = \frac{2i}{N(N+1)}$$
 $i = 1, ..., N.$ (270)

3. Selección por torneo

Es el más usado de todos los métodos de selección. La selección k-torneo selecciona k individuos de la población de forma aleatoria y luego selecciona el más apto de estos k individuos. Este proceso se repite hasta que se han seleccionado el número de individuos que constituye la talla de la población. Si suponemos que los individuos están ordenados de menor a mayor aptitud,

$$f(S_1,t) < f(S_2,t) < ... < f(S_N,t)$$

se puede demostrar que la probabilidad de que el individuo i-ésimo sea seleccionado es:

$$p(S_1, t) = \frac{i^k - (i-1)^k}{N^k} \qquad i = 1, ..., N.$$
(271)

Un tamaño típico para k es 2, se llama entonces selección por torneo binario. Si se incrementa el tamaño del torneo se aumenta la presión sobre la selección, ya que cuanto mayor sea k más difícil será para los individuos por debajo de la media competir con los que están por encima de la media. Obsérvese que en el caso extremo, k=N, siempre se selecciona el individuo más apto.

Análisis del efecto de la selección

Dada una población de N cromosomas, sea H un esquema representado por alguno (posiblemente más de uno) de los miembros de dicha población. Se puede calcular el número de probables representantes de H seleccionados para la siguiente generación: como el tamaño de la población es N, es necesario realizar N sorteos aleatorios, para seleccionar N cromosomas -algunos de los cuales pueden estar repetidos. La probabilidad de seleccionar un esquema H en cada una de esas experiencias aleatorias es la suma de las probabilidades de seleccionar cada uno de los cromosomas que lo representan en la población. De acuerdo con la regla proporcional, la probabilidad de seleccionar un esquema H en cada sorteo en la generación t es:

$$p(H,t) = \frac{\sum_{i \in H} f(S_i, t)}{\sum_{j=1}^{N} f(S_j, t)}.$$

Así pues, el número medio de cromosomas del esquema H en la siguiente generación se obtiene multiplicando el número de sorteos N por la probabilidad de que, en cada sorteo, se incorpore un cromosoma adicional del esquema H. Por tanto, el número medio de representantes de H en la generación siguiente es:

$$m\left(H,t
ight) = N \cdot p\left(H,t
ight) = N \cdot rac{\sum\limits_{i \in H} f\left(S_{i},t
ight)}{N}. \ \sum\limits_{j=1}^{N} f\left(S_{j},t
ight)$$

Si se denomina f(H,t) a la aptitud media en la generación t:

$$f(H,t) = \frac{\sum_{i \in H} f(S_i, t)}{m(H, t)},$$

se tiene que:

$$m\left(H,t+1
ight) = m\left(H,t
ight) \cdot N \cdot rac{f\left(H,t
ight)}{\displaystyle\sum_{j=1}^{N} f\left(S_{j}
ight)}.$$

Como la aptitud media de la población en el instante t es:

$$\overline{f}(t) = \frac{\sum_{j=1}^{N} f(S_j, t)}{N},$$

resulta que:

$$m(H, t + 1) = m(H, t) \cdot \frac{f(H, t)}{\overline{f}(t)}.$$

Esto signigifica que el número de representantes de un esquema presente en la población, con aptitud por debajo de la media tiende a disminuir, mientras que el número de representantes de un esquema con aptitud por encima de la media tiende a incrementarse. Así por ejemplo, si se asume que el esquema H posee una aptitud media superior en un ϵ % a la aptitud media de la población

$$f(H,t) = \overline{f}(t)(1+\epsilon)$$
,

el efecto a largo plazo del algoritmo de selección es el siguiente:

$$m(H,t) = m(H,0) \cdot (1+\epsilon)^t,$$

es decir, la presencia de dicho esquema crece exponencialmente en la generaciones sucesivas.

Operadores genéticos

Los operadores genéticos son los encargados de transformar los individuos de la población en nuevos individuos, ocasionando así la modificación necesaria para que se produzca su evolución. Los dos operadores genéticos más empleados y que aquí se comentarán son el operador de *cruce* y el operador de *mutación*.

Operador de cruce (crossover)

Es el más importante de los operadores de búsqueda dentro de los algorítmos genéticos. Consiste en combinar partes de los individuos padres para formar un nuevo individuo (o más de uno). Los cromosomas progenitores son escogidos aleatoriamente para proceder, luego, al intercambio de material genético. Mediante este operador resultan descendientes diferentes a sus progenitores, pero que mantienen partes de cromosomas comunes.

La forma tradicional de hacer el cruce de individuos es lo que se llama cruce puntual. Consiste en generar de forma aleatoria un punto de cruce-el mismo en los dos cromosomas progenitores- y después intercambiar los segmentos de los padres antes o después de ese punto generando así uno o dos individuos nuevos. Formalmente, si P_1 y P_2 son dos individuos codificados por dos cadenas de bits:

$$P_1 = [p_{11}, ..., p_{1n}],$$

 $P_2 = [p_{21}, ..., p_{2n}],$

respectivamente y se ha generado aleatoriamente un punto de cruce en la posición $1 \le k \le n$ entonces los dos individuos C_1 y C_2 generados a partir de P_1 y P_2 son:

$$C_1 = [p_{11}, ..., p_{1k} | p_{2(k+1)}, ..., p_{2n}],$$

$$C_2 = [p_{21}, ..., p_{2k} | p_{1(k+1)}, ..., p_{1n}].$$

Este operador de cruce es válido para todo tipo de codificación y, en particular, para la binaria y la real (aritmética). Los cromosomas engendrados deben pertenecer al espacio

de búsqueda. Si esto no fuese así, siempre podría conseguirse esto mediante un método de proyección. De la misma forma se puede definir un cruce *multi-punto* donde los individuos se fragmentan por varios puntos en más de dos segmentos, los cuales se intercambian entre los dos cromosomas progenitores para dar lugar a dos nuevos cromosomas.

Otra técnica que suele ser usada es el *cruce uniforme*. En un cruce uniforme cada gen en el hijo se crea copiando un gen de un padre elegido de forma aleatoria entre los dos.

Otro mecanismo de cruce es el denominado *cruce aritmético*. En él, dados dos cromosomas progenitores

$$\mathbf{p}_1 = [x_1, x_2, ..., x_k, ..., x_n],$$

у

$$\mathbf{p}_2 = [y_1, y_2, ..., y_k, ..., y_n],$$

se selecciona aleatoriamente un parámetro de interpolación $\alpha \in [0,1]$ y se generan los descendientes:

$$\mathbf{c_1} = \mathbf{p_1} * \alpha + \mathbf{p_2} * (1 - \alpha),$$

 $\mathbf{c_2} = \mathbf{p_1} * (1 - \alpha) + \mathbf{p_2} * \alpha.$

Este mecanismo es propio de la codificación aritmética, aunque puede ser adaptado a la codificación binaria mediante descodificación de los cromosomas progenitores y codificación de los descendientes resultantes del cruce. El mecanismo de cruce heurístico, es un caso particular del algoritmo de cruce aritmético en el cual sólo se genera un hijo dado que en el proceso de interpolación se da más peso al padre cuya aptitud es mayor.

 Existen en la bibliografía otros algoritmos de cruce (Michalewicz ,1994), como el cruce aritmético puntual que combina ambos mecanismos anteriormente comentados, es decir, se realiza un cruce aritmético en una única posición seleccionada al azar, quedando los demás genes inalterados.

Se resalta que una característica común a todas las técnicas de cruce es que todos los bits que son comunes a ambos padres se heredan en los descendientes.

• Análisis del efecto del cruce

El valor clave para evaluar el efecto del cruce en la conservación de un esquema H es la longitud de definición $\delta(H)$. En efecto, si se considerasen dos esquemas

$$H_1 = [*, *, 1, 0, *, *, *],$$

 $H_2 = [*, 1, *, *, *, 0, *],$

el primer esquema seguiría vigente tras una partición en cualquier posición salvo cuando la misma ocurra entre las posiciones tercera y cuarta. La probabilidad de que al escoger una posición aleatoria de cruce se obtenga precisamente esa es $\frac{1}{6}$. En el segundo caso el esquema se destruye, en principio, cuando la posición del corte esté entre la segunda y la cuarta. La probabilidad de destrucción de este esquema para un sólo cromosoma es $\frac{4}{6}$.

La probabilidad de que un corte en una posición aleatoria destruya el esquema H en un cromosoma, es:

$$p_d = \frac{\delta\left(H\right)}{n-1}.$$

Naturalmente el cruce con la correspondiente porción del esquema del otro cromosoma progenitor pudiera, en teoría, volver a restituir el esquema destruido. Si el cruce se produce con una probabilidad p_c . La probabilidad de que un descendiente **no** conserve el mismo esquema será:

$$p_{nc} = p_c \cdot \frac{\delta(H)}{n-1}.$$

Teniendo en cuenta que existe la posibilidad de que el efecto del cruce con el esquema del otro cromosoma progenitor sea restituir el esquema destruido del primer cromosoma, se tiene:

$$p_{nc} \leq p_c \frac{\delta\left(H\right)}{n-1},$$

siendo la probabilidad de que se mantenga en la siguiente generación:

$$p_c \ge 1 - p_c \frac{\delta(H)}{n - 1}.$$

De la fórmula anterior puede deducirse que los esquemas cortos *presentes en la población* en un instante determinado tienen más probabilidades de conservarse en la siguiente generación por efecto del cruce.

• Operador de mutación

Este operador cambia, en posiciones aleatoriamente escogidas del cromosoma, un alelo por otro. Su efecto es introducir información nueva de manera aleatoria. Considerado uno cualquiera de los alelos fijados, la probabilidad de que resulte mutado se suele mantener pequeña (entre 0.001 y 0.1). En general $p_m \ll 1$. Se suele aceptar que una cota inferior para la probabilidad de mutación debe ser $p_m = 1/o(H)$. Así es casi seguro que muta alguna de las posiciones fijas.

Este operador es también útil para reintroducir información perdida en la población debido a convergencia prematura.

Existen diferentes mecanismos de mutación para alfabetos reales, la mayoría ideados por Michalewicz (1994). Los más importantes son:

1. Mecanismo de mutación uniforme

Dado un individuo $\mathbf{p}_1 = [x_1, x_2, ..., x_k, ..., x_n]$, se selecciona al azar uno de sus genes, denotado por k, y se genera un descendiente

$$\mathbf{c}_1 = [x_1, x_2, ..., x'_k, ..., x_n]$$

donde $x'_k \in U(a_k, b_k)$, es decir es el resultado de un sorteo con ley de probabilidad uniforme en el intervalo (a_k, b_k) , es decir, intervalo de búsqueda para el gen en la posición k-ésima.

2. Mutación no uniforme

En este caso el valor del gen seleccionado se modifica según una ley de probabilidad no uniforme que se hace más apuntada según avanza el algoritmo (número de iteraciones). El espacio de búsqueda se explora uniformemente al inicio del algoritmo y la búsqueda se vuelve más localizada en torno a las cuencas de soluciones potenciales (cambios menos drásticos) según el algoritmo avanza. En este caso el valor del gen mutado es:

$$x'_{k} = \begin{cases} x_{k} + (b_{k} - x_{k}) \cdot f(g) & \text{si } r = U(0, 1) < 0.5, \\ x_{k} - (x_{k} + a_{k}) \cdot f(g) & \text{si } r = U(0, 1) \ge 0.5, \end{cases}$$

donde

$$f(g) = U(0,1) \cdot \left(1 - \frac{g}{g_{\text{máx}}}\right)^{b}$$

siendo g el número de la iteración en curso, $g_{\text{máx}}$ el número máximo de iteraciones del algoritmo y b un parámetro de forma de la distribución de probabilidad (Houck et al, 1996).

3. Mutación de borde

Es un caso particular de la mutación uniforme, donde

$$x'_{k} = \begin{cases} a_{k} & \text{si } r = U(0, 1) < 0.5, \\ b_{k} & \text{si } r = U(0, 1) \ge 0.5. \end{cases}$$

4. Mutación multivariante

Es una variante de los algoritmos de mutación uniforme y no uniforme, que actúa o bien sobre todos los genes del individuo seleccionado, o sobre una lista prefijada de genes. Una manera que hemos ideado para prefijar la lista de genes a mutar en el problema realtivo a los S.E.V es estudiar la lista histórica de mejores soluciones encontradas por el algoritmo y mutar preferentemente un cierto número de ellos, aquellos que poseen en la historia un mayor coeficiente de variación, dado que la alteración de estos parámetros genera un sin fin de soluciones equivalentes.

Análisis del efecto de la mutación

Si p_m es la probabilidad de que, considerado uno de los alelos fijos, éste mute; $(1 - p_m)$ es la probabilidad de que este alelo no mute. La probabilidad de que ninguno de los alelos fijados mute y, por tanto, el esquema no se destruya es:

$$p_{nm} = (1 - p_m)^{o(H)}$$

Como $p_m \ll 1$ se tiene que

$$(1 - p_m)^{o(H)} \approx 1 - p_m \cdot o(H).$$
 (272)

De la fórmula anterior se puede deducir que el efecto que tiene la mutación depende del orden de los esquemas a los que pertenezcan los cromosomas de una población en un momento determinado. Cuanto mayor sea el orden mayor es la probabilidad de que la mutación afecte a alguno de los alelos fijos y, por tanto menor la probabilidad de que se mantenga, dada por la expresión (272).

Teorema de los esquemas

La expresión matemática del teorema de los esquemas es:

$$m(H, t+1) \ge m(H, t) \cdot \frac{f(H, t)}{\overline{f}(t)} \cdot \left(1 - p_c \frac{\delta(H)}{n-1} - p_m \Phi(H)\right)$$
(273)

Su interpretación, en función de lo explicado por bloques hasta ahora, se puede enunciar como sigue: si en un momento determinado hay en la población esquemas cortos, de orden bajo y con idoneidad por encima de la media, éstos aumentarán su presencia en generaciones sucesivas.

Hay que destacar que el teorema de los esquemas informa sobre los esquemas *presentes* en la población. No dice nada sobre los esquemas futuros que puedan incorporarse a la misma, es decir, sobre las nuevas áreas de exploración que puedan aparecer.

Hipótesis del bloque básico

Los esquemas cortos y de orden bajo presentes en un momento dado en la población, se usarán como bloques básicos para construir individuos cada vez mejores. Esta idea es lo que Goldberg ha denominado "hipótesis del bloque básico". Goldberg afirma que los algoritmos genéticos inicialmente detectan, de entre la población inicial, esquemas de orden bajo y aptitud alta, por encima de la media, y converge hacia la parte del espacio de búsqueda que alberga estos esquemas. Posteriormente, a medida que las generaciones se suceden, el algoritmo genético construye a apartir de estos esquemas de orden bajo y utilizando el operador cruce, esquemas de ordenes más altos, de forma que el espacio de búsqueda prometedor se va reduciendo hasta que en algún momento finalice la búsqueda.

Otros parámetros influyentes en el desarrollo de un A.G.

Tamaño de la población

El requisito fundamental de la población de individuos considerada es que represente de forma adecuada a la población entera. Dos factores se tienen en cuenta cuando al determinar el tamaño de la población: diversidad y convergencia. Es claro que cuanto más pequeña sea la población inicial menos diversidad habrá y por lo tanto la población convergerá más rápidamente, siendo probable que lo haga a un extremo local. Al contrario, un tamaño de población grande tendrá una diversidad mucho mayor y por lo tanto más posibilidades de encontrar el óptimo global, siendo en este caso la convergencia mucho más lenta al tener que tratar con más datos.

Existe una teoría de tamaño óptimo de población inicial (Goldberg) que afirma que el tamaño óptimo de la población inicial crece exponencialmente con la longitud de los cromosomas. Para la inmensa mayoría de problemas prácticos esto significa una población muy grande. Sin embargo, a pesar de lo que esta teoría afirma, los resultados empíricos muestran que con tamaños de población en el rango $50 \le N \le 200$ se obtienen buenos resultados.

Población inicial

Normalmente la población inicial se genera de forma aleatoria. Así se asegura su distribución uiniforme en todo el espacio de búsqueda. Sin embargo si se posee algún dato sobre la solución, bien debido a restricciones del problema o bien debido a alguna otra técnica heurística, se puede forzar la población inicial a pertenecer a este subconjunto del espacio de soluciones. Es de esperar que al hacerlo se consiga la convergencia más rápidamente.

• Estrategia de reemplazamiento

Una vez que se han generado u individuos de la población y a partir de ellos, mediante los cruces y las mutaciones, se han generado ${\bf u}$ individuos hijos, es necesario introducir estos nuevos individuos en la población. Dado que el tamaño de población permanece constante hay que eliminar algunos individuos de la población para dejar sitio a los nuevos, es decir, debe producirse un reemplazo de individuos. Las dos técnicas más ampliamente usadas para realizarlo son:

• Reemplazo generacional y elitismo

Es la forma tradicional de proceder. Si N es el tamaño de la población, se generan siempre N descendientes nuevos y toda la población es sustituida en cada iteración. El problema de este método es que no garantiza que el mejor cromosoma de la población actual sobreviva en la próxima generación. Se podría, por tanto, perder información muy valiosa -en la forma de modelo cercano a la solución. Para solucionar este inconveniente se usa la estrategia elitista en la cual los mejores individuos de la población actual se mantienen en la siguiente generación.

• Reemplazo estacionario

En esta técnica sólo un pequeño número de individuos, normalmente uno o dos, son reemplazados en cada generación. Para ser sustituidos se suele escoger el individuo o individuos menos aptos o, a lo sumo, alguno cuya aptitud esté por debajo de la media.

Las ventajas del reemplazo estacionario es que la mejor solución encontrada permanece siempre en la población y que los nuevos individuos generados están inmediatamente disponibles para ser seleccionados. Estas son propiedades útiles teniendo en cuenta que lo que interesa es encontrar el mejor individuo lo antes posible. El esquema estacionario suele converger más rápido que el método de reemplazo generacional. Desde el punto de vista computacional el esquema estacionario es un poco más caro, ya que las probabilidades de selección deben ser recalculadas cada vez que se genera y se reintroduce un nuevo individuo. Manteniendo la población ordenada este coste se reduce algo ya que sólo hay que insertar el individuo en la posición adecuada. Al usar el esquema estacionario debe prestarse atención a la posibilidad de repetir individuos.

• Definición de la aptitud

Como el propósito de un problema de optimización es encontrar el punto óptimo de la función objetivo, intuitivamente podría identificarse la función de aptitud con la propia función objetivo. Si el óptimo buscado es un máximo, esta identificación puede hacerse. Si el problema, como en el caso de los sondeos eléctricos verticales, es localizar el mínimo global, hay dos procedimientos:

- Convertir el problema de minimización en uno de maximización y prodecer con las técnicas propias.
- Considerar que la función de aptitud es una función convenientemente escalada obtenida a partir de la función objetivo. Existen diferentes maneras de realizar ésta tarea:

 - escalado lineal $f(S_{j},t) = K f_{obj}(S_{j},t)$; escalado inverso $f(S_{j},t) = \frac{1}{f_{obj}(S_{j},t)}$;
 - escalado exponencial $f(S_i, t) = e^{-k \cdot f_{obj}(S_i, t)}$;
 - o escalado logarítmico $f(S_j, t) = -\log f_{obj}(S_j, t)$.

La función de aptitud permite deformar, de forma en cierto modo controlada, las características de la función objetivo. Esto permite solucionar dos problemas:

- el caso de individuos cuya aptitud esté muy próxima numéricamente;
- el caso de individuos con evaluación muy superior a la media, que dominan en el proceso de selección produciendo, posiblemente, convergencia prematura hacia un óptimo local.

Definición de parámetros ajustables

Estos métodos pretenden especificar cuál será, para cada ciclo del proceso, el valor óptimo de los parámetros ajustables (probabilidades de mutación, cruce, etc...). Por ejemplo, la probabilidad de cruce es razonable que sea alta al comienzo. En ese momento la población está bien repartida por el espacio de búsqueda y los cromosomas de la misma no tendrán -en principio- muchas similitudes. El cruce aumenta el parecido entre los cromosomas tras varias generaciones. Cuando los individuos presenten características similares, señal de que están aproximadamente en la misma zona del espacio de búsqueda, se corre el riesgo de estancarse en un óptimo local. Para introducir información procedente de nuevas zonas, cabe aumentar la probabilidad de mutación produciendo la consiguiente dispersión en los individuos. Se puede optar entre:

- Mantener dichos parámetros constantes a lo largo del desarrollo del algoritmo genéti-
- el uso de técnicas de interpolación lineal, en las que el parámetro ajustable varía linealmente entre dos valores, siendo actualizado su valor cada p generaciones.
- el uso de técnicas de interpolación adaptativa, en las que el valor de un parámetro en un ciclo dado se calcula teniendo en cuenta su valor en ciclos anteriores. Naturalmente esto requiere mantener almacenadas las poblaciones anteriores con el consiguiente consumo de memoria.

Criterios de parada

En los algoritmos genéticos el critero de parada se suele controlar especificando un número de iteraciones máximo, o en el caso de reemplazo estacionario de un número máximo de hijos generados. Otros criterios que se pueden usar son: mejora relativa en cada generación, nivel de diversidad en la población, Obtención de determinado número de individuos por encima de cierto nivel de aptitud, etc.

Algoritmo

A modo de ejemplo, el esquema básico de un algoritmo genético con reemplazo generacional y criterio de parada por iteraciones es el siguiente:

```
Inicializar P(0) := \{S_1, S_2, ..., S_N\};

Evaluar P(0) := \{f(S_1), ..., f(S_N)\},

Encontrar S^* \in P(0) t.q. f(S^*) \geq f(S), \forall S \in P(0);

Mientras (t < t_{\text{máx}}) hacer

Seleccionar: P_S(t) := \Phi(P(t));

Cruce: P_C(t) := \Omega_c(P_S(t), p_c);

Mutación: P(t+1) \longleftarrow \Omega_m(P_C(t), p_m);

Evaluar: f(P(t+1));

Encontrar óptimo: \widetilde{S} \in P(t+1) t.q. f(\widetilde{S}) \geq f(S), \forall S \in P(t+1)

Reemplazar el óptimo: si f(\widetilde{S}) > f(S^*) entonces S^* \longleftarrow \widetilde{S};

Incrementar generación t \longleftarrow t+1;

Fin del bucle

Devolver S^*, f(S^*).
```

Es usual dividir el proceso de convergencia del algoritmo genético en dos fases diferentes, que serán:

- Explotación o selección natural: es la etapa en la que el AG selecciona los individuos con aptitud más alta. Los mecanismos para la explotación son la selección y el reemplazo.
- Exploración: es el proceso de explorar nuevas áreas del espacio de búsqueda mediante la creación de nuevos individuos a partir de los ya existentes.

Según la hipótesis de Goldberg, en las primeras iteraciones del algoritmo predominará la exploración sobre la explotación y en las últimas fases predominará la explotación sobre la exploración.

7.3. El método de recocido simulado (Simulated Annealing)

7.3.1. Problemas combinatorios y algoritmos numéricos

En muchos problemas es necesario escoger el mejor conjunto de parámetros (configuración del sistema), para los caules una función objetivo determinada, alcanza el mínimo (o máximo) valor, entre un conjunto de configuraciones admisibles. Cuando las configuraciones a valorar son elementos de un conjunto finito o infinito numerable, los problemas se denominan combinatorios. Dos prototipos -diferentes- de problema combinatorio son el problema de asignación y el problema del viajante (Otten et al, 1989).

Algunos problemas tienen métodos de solución algorítmicos que son eficientes en el sentido de que el tiempo de cálculo puede ser acotado por el valor de un polinomio cuya variable es tamaño del problema⁷⁴. Un ejemplo es el problema de asignación. Sin embargo, hay otros problemas para los que parece que no existen algoritmos eficientes, por ejemplo, el problema del viajante. Para

⁷⁴Podría considerarse el tamaño del problema como el número de bits necesario para definirlo.

éstos últimos lo más aconsejable es proceder con métodos aleatorios de búsqueda de solución. Dichos métodos pueden ser clasificados en diferentes categorías:

1. Métodos de búsqueda aleatoria 'ciega'

En este tipo de métodos los candidatos a solución se proponen mediante un sorteo aleatorio, desechando aquellos que no pertenecen a la zona admisible, en el caso de que haya restricciones. Se calcula el valor de la función objetivo y, si el valor obtenido es inferior a los óptimos ya calculados anteriormente, la nueva solución es adoptada. Tras esto, se procede a realizar otro sorteo sin ningún tipo de memoria, en todo el espacio de búsqueda. Cabe introducir modificaciones que aceleren la obtención de valores bajos de la función objetivo. En este tipo tipo se encuentran los métodos de mejora iterativa.

2. Métodos de mejora iterativa

Si, una vez realizado el sorteo aleatorio sobre todo el espacio de búsqueda, el algoritmo está situado en una configuración determinada, se define ahora una vecindad de dicha configuración. Una vecindad es un conjunto de configuraciones -o modelos- que el algoritmo explorará o podrá explorar con cierta probabilidad, mediante movimientos realizados desde la configuración actual. Antes de proceder a un nuevo sorteo global, el algoritmo selecciona, de todos los componentes del entorno local asignado a la configuración actual el que tenga el menor valor de la función objetivo. Este valor es almacenado y, sólo entonces, se procede a un sorteo "ciego" sobre todo el espacio de búsqueda, para obtener una nueva configuración. Este método produce secuencias de óptimos menores que los hallados mediante la búsqueda aleatoria "ciega". La razón para ello es que la elección de los óptimos en la búsqueda aleatoria "ciega" tiene que ver con la distribución global de los valores de la función objetivo. Sin embargo, en el caso de la mejora iterativa, los óptimos obtenidos se distribuyen más de acuerdo a los mínimos locales de la función objetivo.

El único mecanismo que permite a la mejora iterativa escapar de un mínimo local es, por así decirlo, comenzar de cero sin ningún tipo de memoria. Otra solución más eficaz a este incoveniente se consigue con los métodos de escalado probabilístico.

3. Métodos de escalado probabilístico

En éstos, se consigue "escapar" de un mínimo local mediante el empleo de una función de aceptación. Ésta atribuye cierta probabilidad a la admisión, como nueva configuración, de uno de los modelos pertenecientes a la vecindad local de la configuración actual, aunque este movimiento suponga un aumento de la función objetivo. De este modo, en este proceso existe probabilidad de obtener una secuencia de configuraciones de valor creciente de la función objetivo. Esta probabilidad, regulable, permitiría salir de los mínimos locales. Estos algoritmos reciben el nombre genérico de algoritmos de escalado probabilístico. Cuando la función de aceptación permanece fija durante todo el proceso, el rendimiento no parece ser superior al de la técnica de mejora iterativa. Lo interesante es modificar la función de aceptación mediante el cambio de un parámetro de control a lo largo del algoritmo.

El método de recocido simulado (simulated annealing) pertenece a la clase de algoritmos de escalado probabilístico con un sólo parámetro que, por razones que se explicarán más adelante, se denomina temperatura. Este parámetro es positivo y, a lo largo del proceso, decrece.

7.3.2. El algoritmo de recocido simulado

Se trata de un método de optimización global cuyo nombre remite a un procedimiento de obtención de diferentes estructuras cristalinas estables. Si se piensa en un material en estado líquido, el enfriamiento instantáneo a temperaturas suficientemente bajas produciría su congelación y los átomos quedarían 'atrapados' en una posición muy similar a la que ocupaban previamente en el líquido. Este estado congelado tiene una energía superior a la del ordenamiento cristalino, pero no tiende a ella espontáneamente porque existe una barrera de energía intermedia. Para 'elevar' la energía por encima de la de esta barrera, se puede calentar el sólido 'congelado' hasta llevarlo nuevamente al estado líquido. Si, ahora, en lugar de enfriar bruscamente, se disminuye lentamente la temperatura, los átomos tendrán tiempo de 'explorar' numerosas posiciones antes de quedar atrapados en aquella que corresponda a la de menor energía de todas las visitadas. Cuanto más lento sea el enfriamiento, más probable es que la sustancia 'cristalice' en el estado de mínima energía de todos los posibles a cierta temperatura. Este proceso se denomina recocido y se utiliza para eliminar defectos en cristales.

Origen y aplicaciones del método

El recocido simulado se introdujo en el ámbito de los métodos numéricos en la década de los 50 con el trabajo de Metrópolis y sus colaboradores (1953). En este trabajo se pretendía simular el comportamiento de sistemas en mecánica estadística. A una temperatura por encima del cero absoluto, un sistema puede encontrarse con cierta probabilidad no nula en estados de distinta energía. La ley que vincula la energía que tiene un estado y la probabilidad de que un sistema se halle en dicho estado a una temperatura dada es:

$$p(E) \propto e^{-\frac{E}{\kappa T}},$$
 (274)

donde T es la temperatura absoluta, E es la energía y κ es la denominada constante de Boltzman. En su trabajo se simula el paso del sistema de un estado a otro, proponiendo aleatoriamente un nuevo estado, evaluando su energía y aceptándolo como nuevo estado en dos circunstancias:

- 1. Cuando la energía del nuevo estado sea menor que la del original.
- 2. Cuando la energía del nuevo estado sea mayor que la del original, la probabilidad de aceptarlo será evaluada mediante la expresión:

$$p\left(\Delta E\right) = e^{-\frac{\Delta E}{T}}. (275)$$

La primera aplicación, en el campo de la optimización numérica, de la técnica de simulated annealing se debe a Kirkpatrick (1983). Se trataba de minimizar la longitud de las conexiones entre los distintos componentes electrónicos de un circuito integrado.

7.3.3. Elementos característicos del método de simmulated annealing

Considerada como técnica de optimización, el recocido simulado precisa que existan lo siguientes elementos:

- Una regla para calcular nuevos modelos que integren lo que se ha denominado vecindad del modelo actual.
- 2. Un criterio de aceptación -o rechazo- de cada uno de los modelos de dicha vecindad, como nuevo estado del sistema.
- 3. Un criterio para modificar los parámetros del método de forma que el modelo obtenido tienda a estabilizarse alrededor del mínimo global.
- 4. Deben existir criterios de parada del algoritmo cuyo cumplimiento indique que se ha alcanzado un modelo de características adecuadas para considerarlo solución del problema.

A continuación se aportan detalles de cada uno de estos elementos:

1. Sorteo aleatorio de propuesta de modelos

Partiendo de un modelo (configuracion) inicial, debe un sorteo aleatorio que produzca nuevos modelos (configuraciones) al azar. Estos modelos serán posteriormente valorados y se decidirá la conveniencia o no de aceptarlos.

Inicialmente, conviene que las propuestas exploren zonas relativamente alejadas del modelo actual porque se sabe poco de si éste está cerca o lejos de la solución. Sin embargo, a medida que se producen progresos en términos de lograr buenas aproximaciones a la solución, sería bueno no perder demasiado tiempo analizando propuestas que, según se sabe a través de la información ya obtenida, no pertenecen -con bastante probabilidada la zona interesante.

El método admite soluciones en el interior de un dominio paralelepipédico del espacio de modelos -el denominado espacio de búsqueda- que se estipulado a priori. Puede especificarse una ley aleatoria de propuesta de modelos vecinos de forma que el tamaño de la vencindad aproxime el tamaño de todo el espacio de búsqueda en los inicios del algoritmo. A medida que progresa el cálculo, el tamaño de la vecindad debiera ir reduciéndose, ciñéndose sólo a la zona de interés que contiene a la solución.

En este trabajo se han incorporado dos métodos diferentes de sorteo de paso aleatorio:

Sorteo con ley Gaussiana

Este sorteo produce un vector $\Delta \mathbf{m}$ que sigue una distribución Gaussiana de media $\mathbf{0}$ y desviación típica dada, para cada una de las coordenadas de $\Delta \mathbf{m}$ mediante la siguiente expresión:

$$\sigma_i = \gamma \cdot (s_i - l_i) \cdot \frac{f(\mathbf{m}) - f_{\inf}}{f_{\sup} - f_{\inf}}.$$
(276)

 σ_i es la desviación típica de la ley gaussiana que sigue la componente *i*-ésima del vector aleatorio de paso. $\gamma \in (0,1)$, es una constante denominada parámetro de localización⁷⁵. s_i (resp. l_i) es la coordenada *i*-ésima del limite superior (resp. inferior) del espacio de búsqueda -que es común para todos los modelos-. $f(\mathbf{m})$ es el valor

 $^{^{75}}$ El valor del parámetro de localización $\gamma \in (0,1)$ aparece en Applied Numerical Methods for Engineers, pags. 285-286 tomando valores entre 0.1 y 0.2.

actual de la función objetivo. Finalmente, f_{sup} y f_{inf} son las cotas superior e inferior estipuladas para la citada función objetivo.

Del análisis de la expresión citada en (276) se deduce que, cuando los modelos encontrados producen un acercamiento de $f(\mathbf{m})$ hacia el óptimo (y, por tanto, reducen su diferencia respecto a $f_{\rm inf}$), σ_i disminuye en todas las componentes del vector. Asimismo, sobre cada eje-representado por valores de i diferentes- la dispersión de las propuestas será proporcional al diámetro del espacio de búsqueda global sobre dicho eje i-ésimo. Un parámetro γ más pequeño, produce vecindades más concentradas alrededor del modelo actual.

Sorteo con ley uniforme

El vector $\Delta \mathbf{m}$ resulta uniformemente distribuido en un intervalo comprendido entre los límites superior e inferior. La coordenada *i*-ésima de este intervalo se obtiene mediante:

$$d_i = \frac{T}{T_0} \left(s_i - l_i \right). \tag{277}$$

En este caso también, la búsqueda se hace más local a medida que progresa el algoritmo. La constante de proporcionalidad $\frac{T}{T_0}$, precisa para su interpretación del concepto algorítmico de temperatura. Cuando, seguidamente, se trate la cuestión de los criterios de aceptación, se entenderá cómo la temperatura es un parámetro del recocido simulado que desciende a medida que progresa el algoritmo. Si T_0 es el valor de la temperatura al inicio del proceso y T el valor en un momento posterior -el de realización del sorteo bajo consideración- el cociente $\frac{T}{T_0}$ disminuye al desarrollarse el algoritmo. Por tanto el intervalo de búsqueda va siendo cada vez menor.

Los valores de energía (funcion objetivo) que en estos nuevos modelos se alcancen dependen de varios factores:

- a. de la ley de probabilidades que rija el sorteo aleatorio, pues la probabilidad de que un determinado modelo sea propuesto es diferente si, por ejemplo, la ley del sorteo es gaussiana que si es uniforme en todo el espacio de búsqueda.
- b. del punto inicial, dao que la probabilidad de que un modelo nuevo sea alcanzado depende de cuál sea el punto de partida del posible movimiento, naturalmente suponiendo siempre utilizada la misma ley de azar.
- c. de la forma de la función objetivo, que determinará la distribución de los valores de energía entre los nuevos modelos obtenidos.

Así pues, un importante parámetro de control para optimizar el rendimiento de la técnica de S.A en cada tipo de problema particular es la elección de la ley de probabilidades que define los movimientos posibles.

2. Criterios de aceptación

Estas transiciones aleatorias de un modelo a otro pueden ser aceptadas o rechazadas. Como ejemplo de ley que se utiliza para decidir sobre la admisión de la nueva posición

cabe, precisamente, la de **Boltzman** enunciada en la expresión (275). Ésta es, además, la que dió origen al método:

 $p\left(\Delta E\right) = e^{-\frac{\Delta E}{T}}. (278)$

En esta ley, cuando $\Delta E < 0$, el movimiento se acepta siempre. Es decir, siempre que se proponga una mejora se acepta. Por otro lado, cuando $\Delta E > 0$, el movimiento podría aceptarse. De hecho, se admite la propuesta con probabilidad $p(\Delta E)$, indicada por la ley de Boltzman. T es un parámetro denominado temperatura. Se observa en (278) que, cuando T es alta, es también elevada la probabilidad de aceptar valores de $\Delta E > 0$ grandes. Es decir, es bastante posible que se pueda aceptar un cambio de un modelo a otro con una función objetivo mayor. Cuando, al contrario, se consideran valores de T bajos, decrece la probabilidad con la que se admiten modelos que supongan un aumento de la función objetivo. En general, con esta expresión, la probabilidad de que un modelo situado en un mínimo local de 'profundidad' aproximada T, salga de él es aproximadamente igual al 37%.

Al inicio, conviene no quedar entrampado en mínimos locales -pues se busca el mínimo global-. Por tanto, el valor de T debe ser alto. A medida que continúa el proceso, debe ir disminuyéndose la probabilidad de salir de los mínimos encontrados -sin terminar por anularla totalmente-. Esto se consigue disminuyendo T. El modo como T evoluciona, desde su valor inicial hasta su valor final recibe el nombre de patrón de enfriamiento y es un aspecto clave del éxito del algoritmo.

La elección de la ley de Boltzman como criterio probabilístico de aceptación de nuevos modelos, aporta al método las siguientes características deseables:

- Se aceptan todos los movimientos que supongan descenso de la función objetivo, esto es, inevitablemente, el camino aleatorio por distintos modelos progresa hacia el mínimo (o mínimos).
- No se rechazan todas las transiciones hacia nuevos modelos que supongan un aumento de la energía o función objetivo. La utilidad de esta propiedad se aprecia imaginando un modelo situado en un mínimo local. En cada nueva iteración se proponen para aceptación nuevos modelos generados aleatoriamente y que, posiblemente (esto depende de la ley aleatoria de generación de modelos) caigan más allá de la cuenca de atracción del mínimo local (aunque, posiblemente en un modelo de energía mayor). La existencia de cierta probabilidad de que estos modelos sean aceptados, garantiza la posibilidad de no entramparse definitivamente en un mínimo local.
- Es una ley decreciente. Esto indica que grandes incrementos de energía tienen muy poca probabilidad de ser aceptados (pero conservan alguna).
- Es acotada en el origen.
- Es realmente una ley de comportamiento simple, con pocos parámetros y de fácil implementación.

La ley de Boltzman no es la única propuesta válida para decidir la aceptación o el rechazo de un modelo propuesto. De hecho, el conjunto de programas que permite la ejecución del

algoritmo de simmulated annealing en este trabajo, considera además la ley de Cauchy. Su expresión es:

$$p_c = \frac{2T}{\pi \left(T^2 + \Delta f^2 \right)}. (279)$$

Esta función atribuye a los movimientos grandes probabilidades mayores que la ley de Boltzman.

3. El patrón de enfriamiento

Se considera en este apartado tanto la elección de la temperatura inicial como la evolución de la misma a lo largo del algoritmo.

• Elección de la temperatura inicial

Un criterio normalmente utilizado en la práctica, ([?]) consiste en calcular la temperatura inicial como aquella que permite aceptar (según la ley de Boltzman) un incremento del 50 % (u otro factor porcentual) del valor inicial de la función objetivo con una probabilidad p (por ejemplo del 70 %). Matemáticamente:

$$p_0 = e^{-\frac{\delta \cdot f_0}{T_0}} \Longrightarrow T_0 = -\frac{\delta \cdot f_0}{\log p_0},\tag{280}$$

El parámetro δ es el incremento porcentual en tanto por uno. Por ejemplo, en el caso de un 50% mencionado antes, $\delta = 0.5$. f_0 es el valor de la función objetivo en el modelo inicial y p_0 es la probabilidad con la que se estipula debe aceptarse el movimiento que genere el citado incremento.

Otras formas de cálculo de T_0 son posibles. Por ejemplo Kirkpatrick (1983) calcula la temperatura inicial como aquella para la que la probabilidad de aceptar el valor medio de todos los *ascensos* posibles es de un 80 %. Así pues:

$$T_0 = -\frac{\delta E}{\ln 0.8}.\tag{281}$$

4. La curva de enfriamiento

El patrón de enfriamiento refleja la evolución de la temperatura con el número de iteraciones. Dos tipos diferentes de consideraciones influyen en la forma de dicha curva:

- El criterio para decidir cuándo se modifica la temperatura.
 Básicamente, en una temperatura determinada se permanece un tiempo limitado:
 - bien hasta que transcurra un número máximo de iteraciones totales impuestas a priori,
 - bien hasta que se haya obtenido una cantidad -estipulada también a priori- de nuevos modelos aceptados,
 - o bien, finalmente, cuando un número dado de óptimos haya sido superado.
- El criterio que especifica cómo y cuánto se cambia.
 Resumiendo las distintas modalidades indicadas en la literatura existente, se mencionan cuatro métodos que están implementados:

a) Patrón aritmético:

$$T_{n+1} = T_0 - n\Delta T.$$

 T_0 es la temperatura inicial y ΔT es la variación de temperatura en cada paso. Si el número total de pasos es N, una opción (Dittmer et al, 1995) es hacer $\Delta T = \frac{T_0}{N}$.

b) Patrón geométrico:

$$T_n = f^n \cdot T_0, \ 0 \le f \le 1.$$

Normalmente la razón f tiene un valor del orden de 0.9 (Schilling et al, 19??)

c) Patrón logarítmico:

$$T_n = \frac{T_0}{1 + \log\left(1 + n\right)}.$$

d) Patrón inverso:

$$T_n = \frac{T_0}{(1+n)}.$$

5. Criterios de parada

Es preciso determinar cuándo se ha llegado a una solución razonable. Se contemplan las siguientes posibilidades:

- El valor de la función objetivo en el último modelo es inferior a la tolerancia especificada.
- Parar si el valor de la función objetivo no ha mejorado en un porcentaje determinado, tras un número de iteraciones estipulado
- Parar si el número de transiciones aceptadas es menor que un porcentaje dado, tras un número fijado de iteraciones.

7.4. Aplicación de los métodos globales a un problema medioambiental de toma de decisiones

En la práctica sería conveniente formular el problema inverso geofísico en un marco adecuado a la valoración de riesgos para la toma de decisiones y, por tanto, las probabilidades debieran estar involucradas. Se ha venido mencionando que, en los últimos tiempos el énfasis se ha desplazado desde el problema clásico de estimación de un modelo a la caracterización del conjunto de modelos equivalentes. Dada una tolerancia de medida, ligada a las incertidumbres en los datos de campo, las predicciones de los modelos equivalentes aproximan las observaciones mejor que la citada tolerancia.

El enfoque bayesiano para los problemas inversos -Tarantola and Valette (1982)- permite construir una distribución a posteriori sobre el espacio de modelos que incluya tanto la información que proviene de las medidas de campo como la información a priori y el modelo físico. Esta metodología asume conocido o estimable el modelo estadístico de los errores en los datos y la distribución correspondiente de los modelos en el espacio de soluciones que cuantificaría el grado de certeza que se atribuye al conocimiento geológico existente. Naturalmente, debido a

las limitaciones económicas y específicas propias del trabajo de campo, puede resultar escasa la disponibilidad de ambos tipos de información.

Normalmente las expresiones explícitas de la citada distribución a posteriori no están disponibles, especialmente en problemas no lineales. Asimismo, la distribución a posteriori puede ser difícil de describir -Mosegaard y Tarantola (1995)- en el sentido de que podría carecer de alguno de sus momentos estadísticos, no ser unimodal, etc. Por tanto, sería extraordinariamente valioso disponer de un algoritmo capaz de realizar un muestreo de la región interesante del espacio de modelos ponderando cada zona de acuerdo con la distribución a posteriori (incluso aunque esto se consiga asintóticamente). Mosegaard y Tarantola (1995) han demostrado que la metodología teóricamente correcta para conseguir visitar cada región del espacio de modelos un número de veces proporcional a su probabilidad a posteriori está basada en el algoritmo de Metropolis-Gibss, implementado como función de aceptación más común en el método de simulated annealing. Esta capacidad del citado algoritmo se mantiene independientemente de la forma que tenga la distribución a posteriori, no siendo necesario hacer hipótesis sobre la misma. El método de simulated annealing en versión reformada (VFSA)⁷⁶ ha sido utilizado por Bhattacharya et al (2003), para caracterizar los dos primeros momentos de la distribución.

En este trabajo se muestran las posibilidades que, al mismo respecto, ofrece un algoritmo genético estándar. Se ha realizado también un estudio numérico de la influencia y obtención de los valores adecuados de los distintos parámetros del algoritmo utilizando unas funciones test para las cuales es posible expresar los resultados en términos de expresiones analíticas simples. Sobre estas funciones se han llevado a cabo comparaciones de los resultados analíticos y S.A frente a los algoritmos genéticos. Una vez confirmadas las posibilidades del método, han pasado a aplicarse sobre S.E.V realmente obtenidos con motivo de un problema de intrusión salina en Murcia.

7.4.1. Características relevantes y parámetros empleados

Se ha mencionado ya la capacidad de los algoritmos genéticos para alcanzar un compromiso entre la exploración y la convergencia hacia las regiones prometedoras. En este caso el objetivo fundamental es la exploración o muestreo de la zona de interés a un ritmo proporcional a la distribución a posteriori. Incidentalmente, en el diseño de los espacios de búsqueda puede tenerse en cuenta la información geológica disponible.

Los parámetros utilizados del algoritmo genético son los siguientes:

- El alfabeto utilizado: para las soluciones potenciales. Los modelos binarios pertenecen a \mathbb{R}^k , where $k = \sum_{i=1}^{2n-1} l_i$, donde l_i es la longitud del alfabeto del parámetro i-ésimo.
- un mecanismo para crear las poblaciones iniciales. El modo más simple e intuitivo es un sorteo aleatorio uniforme en todo el espacio de parámetros (o en la zona de búsqueda). Es el que ha sido utilizado.
- una funcion de evaluación (función de aptitud). Para clasificar los modelos en términos de su verosimilitud. Los modelos más verosímiles se considerarían más aptos. La

⁷⁶Very Fast Simmulated Annealing

evaluación de la función de aptitud es inmediata en el caso de las funciones test utilizadas en esta investigación -ellas mismas son la función de aptitud- pero es más complicada en el ejemplo real de V.E.S donde hay que proceder resolviendo, primero, el problema directo para todos los representantes de una generación; en segundo lugar hay que calcular los residuos con respecto a los datos medidos (para lo cual, se ha utilizado aquí la norma L_2) y, finalmente, eligiendo un modo de re-escalar dichos residuos de modo que se asignen mayores niveles de aptitud a los individuos con residuos menores.

- un conjunto de operadores genéticos capaces de alterar la población de soluciones potenciales, privilegiando el intercambio de información entre ellas. Los operadores utilizados aquí han sido la selección, el cruce y la mutación.
 - Selección. Influye enormemente sobre la probabilidad de que un individuo aceptable mantenga o incremente su influencia genética a lo largo de la ejecución del algoritmo. Un aumento de la presión selectiva disminuye la diversidad en la población y puede causar una convergencia prematura. Por contra, una presión selectiva excesivamente débil puede volver ineficiente el algoritmo. Los distintos métodos difieren en el modo de asignar las probabilidades de selección. Para los resultados aquí presentados se han considerado: ruleta, torneo y clasificación geométrica normalizada.
 - Cruce. Se ha utilizado el cruce puntual y el doble cruce.
 - Mutacion. Permite escapar de las cuencas locales de atracción, mejorando la tarea de exploración y haciendo más lenta la convergencia. Se ha empleado la mutación uniforme y no uniforme.

7.4.2. Modelos sintéticos

A pesar de que en el problema inverso en S.E.V el espacio de modelos tiene 2k-1 parámetros donde k representa el número de capas, para comprobar la tesis de que los algoritmos genéticos son capaces de muestrear adecuadamente el espacio de búsqueda, se han empleado ejemplos sintéticos en los cuales el espacio de modelos es 2D facilitándose así el análisis y la interpretación geométrica.

Las funciones test analizadas juegan el papel de funciones de distribución a posteriori conocidas. Se investiga si los algoritmos genéticos son capaces de muestrearlas adecuadamente. La estrategia seguida para conseguirlo consiste en *introducirlas también como funciones de aptitud* en el propio algoritmo genético. Como se conoce la expresión analítica de las citadas funciones de prueba, este proceso es relativamente simple. En el caso de los S.E.V la función de densidad de probabilidad a posteriori (f.d.p. a posteriori, en lo sucesivo) que, en la metodología propuesta, también hace las veces de función de aptitud en el algoritmo genético utilizado, se relaciona con el funcional de desajuste⁷⁷.

$$P(\mathbf{m}) \propto e^{-r(\mathbf{m}, \mathbf{d})}$$

suele asumirse para la f.d.p a posteriori. (Si en el funcional no se incluye ningúna información a priori, que es el caso considerado).

⁷⁷Generalmente si $r(\mathbf{m}, \mathbf{d}) \in \mathbb{R}_*^+$ es el desajuste, una expresión de la forma:

Al ejecutar el algoritmo genético, se generará una población de soluciones que, se postula, es una muestra representativa adecuada del espacio de modelos. Los modelos obtenidos son entonces clasificados de acuerdo con sus funciones de aptitud (y, por lo tanto, de acuerdo con su verosimilitud). Además, a partir de dicha muestra, es posible obtener algunas estadísticas muestrales de utilidad.

Los resultados empíricos obtenidos (histogramas muestrales) serán comparados con las expresiones teóricas para las probabilidades de las regiones muestreadas. Para las funciones de prueba utilizadas se han obtenido fórmulas analíticas que permiten evaluar las integrales de estas p.d.f sobre dominios particulares del espacio de modelos. El cálculo de estas integrales es necesario para asignar probabilidades de ocurrencia de los modelos pertenecientes a las citadas zonas. Para las funciones escogidas, se muestra que la región del espacio de modelos cuya probabilidad es igual o mayor que un cierto umbral f_0 -que, recuérdese, se interpreta aquí también como aptitud- tienen formas elípticas que resultan muy similares a los valles estrechos y alargados que se observan en el problema inverso de S.E.V. Además, resulta que dicha región está determinada de forma unívoca por el valor f_0 . Ocurre, además, que el valor de la probabilidad de dicha región es una función simple de f_0 permitiendo, por tanto, la comparación de resultados numéricos y teóricos.

Supóngase que:

$$D_{f_0} \equiv \left\{ \mathbf{m} = (x, y) \in \mathbb{R}^2 : f(x, y) \ge f_0 \right\},$$
 (282)

es la región de modelos caracterizados porque todos ellos tienen una f.d.p a posteriori, f(x,y), mayor que un determinado valor de corte f_0 . Supóngase además que se denomina $E \subseteq D_{f_0}$ a cualquier subregión de D_{f_0} . Es válida entonces la siguiente relación:

$$P\left(\mathbf{m} \in E/D_{f_0}\right) = \frac{P\left(\mathbf{m} \in E\right)}{P\left(\mathbf{m} \in D_{f_0}\right)} \stackrel{def}{=} \frac{\iint_E f\left(x, y\right) dx dy}{\iint_{D_{f_0}} f\left(x, y\right) dx dy}.$$
 (283)

En el caso de que E esté asociado a un valor determinado de aptitud, f_E :

$$E \equiv D_{f_E} = \left\{ \mathbf{m} = (x, y) \in \mathbb{R}^2 : f(x, y) \ge f_E > f_0 \right\},\,$$

se tiene que:

$$P(\mathbf{m} \in D_{f_E}/D_{f_0}) = \frac{V(f_E)}{V(f_0)},$$
 (284)

siendo V(f) el volumen asociado a la porción de función de aptitud con valores mayores que f. En este trabajo se sugiere que es esta probabilidad teórica (284) la que hay que comparar con el porcentaje siguiente, obtenido por contaje en la población solución proporcionada por los A.G:

$$n_{E/0} = \frac{N_E \left(\mathbf{m} : f\left(\mathbf{m}\right) \ge f_E\right)}{N_0 \left(\mathbf{m} : f\left(\mathbf{m}\right) \ge f_0\right)}.$$
(285)

Si el método funciona, ambos porcentajes debieran ser muy aproximados.

7.4.3. Descripción de las funciones test y resultados numéricos

A continuación se muestran las expresiones que poseen los dos tipos de funciones test utilizadas en este trabajo, a saber:

■ La función **sombrero de bruja:** Se aporta aquí como ejemplo de función no diferenciable. Además de su propia expresión matemática f(x,y), se adjuntan los valores de la constante de normalización k precisa para poder interpretarla en términos de probabilidad.

Asimismo se ha calculado la expresión D_{f_E} de la región (elíptica) en la que todos los modelos tienen una probabilidad superior a f_E -valor umbral- junto con la probabilidad total asociada a dicha región (y, por tanto al valor umbral), $V(f_E)$.

$$f(x,y) = \begin{cases} k\left(C - \sqrt{\frac{x^2}{a^2} + \frac{y^2}{b^2}}\right), & \text{if } \frac{x^2}{a^2} + \frac{y^2}{b^2} \le C^2, \\ 0 & \text{if } \frac{x^2}{a^2} + \frac{y^2}{b^2} > C^2, \end{cases}$$
 $k, C \in \mathbb{R}^+,$ (286)

$$k = \frac{3}{\pi abC^3},\tag{287}$$

$$D_{f_E} = \left\{ (x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \le \left(C - \frac{f_E}{k} \right)^2 \right\}, \tag{288}$$

$$V(f_E) = \frac{1}{C^3} \left(C - \frac{f_E}{k} \right)^2 \left(C + 2\frac{f_E}{k} \right). \tag{289}$$

• funcion gaussiana: usualmente tomada en geofísica como modelo de la verosimilitud de los desajustes en los datos medidos. Se han calculado también las mismas expresiones relevantes.

$$f(x,y) = \frac{1}{2\pi ab} e^{-\frac{1}{2}\left(\frac{x^2}{a^2} + \frac{y^2}{b^2}\right)},$$
(290)

$$D_{f_E} = \left\{ (x, y) \in \mathbb{R}^2 : \frac{x^2}{a^2} + \frac{y^2}{b^2} \le -2\ln\left(2\pi a b f_E\right) \right\},\tag{291}$$

$$V(f_E) = 1 - 2\pi a b f_E.$$
 (292)

Las conclusiones que se derivan de los experimentos numéricos planificados son las siguientes:

- 1. Los algoritmos genéticos realizan, de modo consistente a lo largo de toda la experimentación, un buen muestreo de la región de modelos de verosimilitud relativamente alta (desajuste bajo), que es la zona sobre la cual se deben tomar las decisiones prácticas, (figura 55, conjunto de curvas a la derecha). La capacidad de muestreo del algoritmo se deteriora cuando se incluyen regiones donde hay modelos con baja verosimilitud, (figura 55, conjunto de curvas a la derecha). El método de simulated annealing, utilizando el algoritmo de Metropolis-Gibbs, muestrea todo el espacio de búsqueda con idéntica calidad, pudiendo observarse un ajuste casi perfecto con la curva teórica.
- 2. Este comportamiento se ha observado para probabilidades de mutación relativamente altas (siempre mayores que 0,1), independientemente de la semilla inicial. Para probabilidades de mutación bajo el umbral de 0,1 el resultado del muestreo depende extraordinariamente del punto de partida y el sesgo es fuerte.

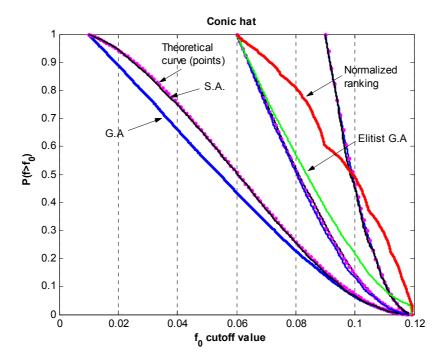


Figura 55: Función sombrero de bruja. Funciones de distribución teórica y empírica utilizando Simulated Annealing y varias ejecuciones de A.G. Para la función Gaussiana se obtienen resultados similares siendo el grafo de la función de distribución teórica una línea recta.

- 3. La influencia del cruce no parece ser determinante. Entre los métodos de selección ensayados, parece que el método de ruleta es el adecuado.
- 4. La presencia de elitismo tiende a sobremuestrear los modelos más cercanos al óptimo global, comportamiento consistente con la tendencia a una convergencia más rápida (figura 55, ver flechas)
- 5. Los histogramas de las distribuciones marginales de los parámetros del modelo, utilizados para el análisis que precede a la toma de decisión, han mostrado ser estables bajo las circunstancias mencionadas (figura 56). Los histogramas 2D varían moderadamente manteniendo, no obstante, similar apariencia general y el mismo sesgo y dispersión (figura 57).

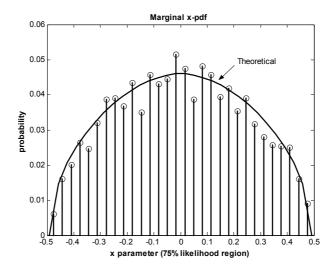


Figura 56: Sombrero de bruja. Histograma marginal para el parámetro x en la región del 75 % de verosimilitud con A.G. (los resultados obtenidos para el parámetro y son idénticos).

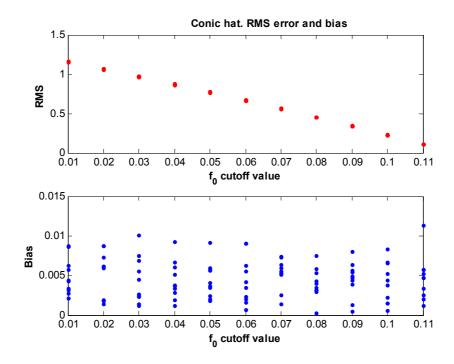


Figura 57: Sombrero de bruja $(a=1,\ b=2,\ C=2)$. Evolución del sesgo y de la dispersión (ECM) de muestras de AG para diferentes umbrales de verosimilitud (probabilidad de mutación del 80%) y 10 poblaciones iniciales generadas aleatoriamente.

7.5. Caso de estudio: un problema de intrusión salina

7.5.1. El problema inverso

El problema clásico de estimación consiste en hallar $\widetilde{\mathbf{m}} \in \mathbf{U} \subseteq \mathbb{R}^{2k-1}$ que minimiza el funcional:

$$\begin{split} &\Phi\left(\widetilde{\mathbf{m}}\right) = \min_{\mathbf{m} \in \mathbf{U}} \Phi\left(\mathbf{m}\right), \\ &\Phi\left(\mathbf{m}\right) = \left\|\rho_{ap}\left(\mathbf{m}, \mathbf{r}\right) - \rho_{ap}^{obs}\left(\mathbf{r}\right)\right\|_{W_{d}}^{2} + \epsilon^{2} \left\|\mathbf{m} - \mathbf{m}^{*}\right\|_{W_{m}}^{2}, \end{split}$$

donde **U** representa al espacio de búsqueda de modelos; W_d , W_m son normas adecuadas en espacio de datos y de modelos, respectivamente; $\rho_{ap}^{obs}(\mathbf{r})$ son las resistividades aparentes observadas; \mathbf{m}^* es un modelo de referencia (si se utiliza información a priori), y ϵ^2 es el parámetro de regularización, el cual puede verse como un botón para sintonizar la contribución de la información a priori en el modelo estimado.

Incluso con la presencia del término de regularización, el problema de S.E.V genera estimaciones inestables de los parámetros del modelo cuyas predicciones, sin embargo, ajustan los datos igual de bien (modelos equivalentes). En un problema de toma de decisiones la clave está en clasificar esos modelos equivalentes

$$\widetilde{\mathbf{m}} \in \mathbf{M} \subseteq \mathbf{U} : \Phi(\mathbf{m}) < tol,$$
 (293)

de acuerdo a su distribucióna posteriori.

Se ha utilizado simulated annealing y los algoritmos genéticos binarios para generar muestras adecuadas de \mathbf{M} a partir de las cuales puedan realizarse inferencias que fundamenten la toma de decisiones. Se han obtenido los histogramas para los parámetros de las capas y se ha caracterizado la interdependencia y balances entre los mismos de modo que pueden ser comparados estos resultados con los provenientes del análisis lineal.

7.5.2. Modelización de la intrusión salina de la zona de Águilas

Se presentan brevemente en esta sección los resultados obtenidos por aplicación de la metodología citada a un sondeo costero en Murcia (España) orientado a la localización de una intrusión salina.

Las conclusiones son las siguientes:

- 1. El método de simulated annealing y los A.G han identificado la misma zona de modelos equivalentes (figura 58). Además, las funciones de distribución acumuladas son casi coincidentes (figura 59) al igual que ocurre con las distribuciones marginales de probabilidad (figura 60). Las ejecuciones del programa han tenido una duración de alrededor 2 minutos en ambos métodos, S.A y A.G, con una población de 40.000 modelos, en un Pentium 4, 1.7 Ghz y 130 Mb RAM.
- 2. En el caso analizado los valores más probables de resistividad se central alrededor de los $10~\Omega.m$ confirmando la presencia de una intrusión salina.

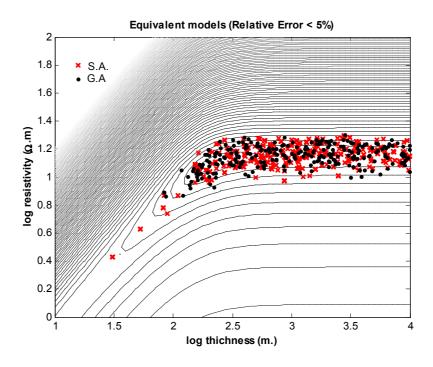


Figura 58: Intrusión salina. Isolíneas de la función de desajuste (resistividad frente a espesor). Parámetros del modelo para errores menores del $5\,\%$.

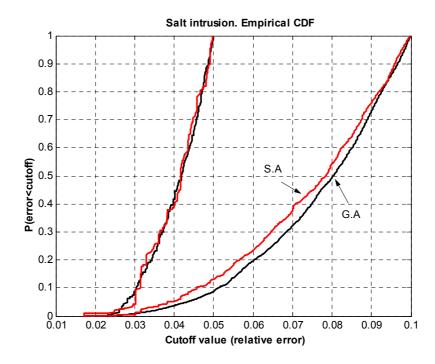


Figura 59: Intrusión. Función de distribución de los errores.

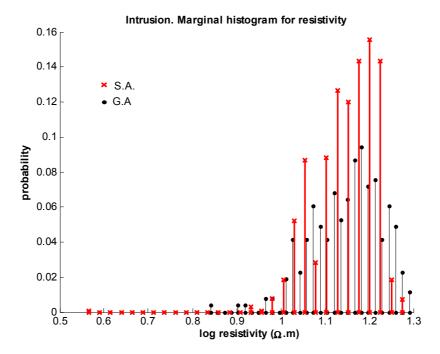


Figura 60: Distribución marginal a posteriori de resistividades (región con errores inferiores al 5%).

Capítulo 8: Esquema general de programación de problemas inversos en Matlab

8. ESQUEMA GENERAL DE PROGRAMCIÓN DE PROB-LEMAS INVERSOS EN MATLAB

8.1. Breve descripción de MATLAB

MATLAB $^{\textcircled{R}}$ es paquete informático que posee un lenguage de programación de alto nivel para realizar cálculos matemáticos. Permite el cálculo matemático en sí (simbólico y con aritméticas de diferente precisión), la programación y desarrollo de algoritmos-mediante la generación de ficheros de extensión .m y la visualización, a través de su potente capacidad de generación y manejo de gráficos. Además se presenta en un entorno relativamente intuitivo, muy interactivo y sencillo de utilizar y se sirve (en general) de la notación matemática estándar.

Con MATLAB el usuario puede generar a voluntad su propia biblioteca de funciones adaptadas que funcionarán junto con las ya existentes, completando y personalizando las denominadas toolboxes o cajas de herramientas. De éstas se hablará a continuación. Asimismo permite el desarrollo de interfaces gráficas que proporcionan al usuario un entorno más agradable y sencillo de uso del código programado.

8.1.1. Cajas de herramientas en MATLAB

MATLAB incluye una familia de funciones programadas para su uso en aplicaciones específicas. Estas familias se denominan toolboxes o cajas de herramientas. Las mismas pueden ser completadas por funciones programadas por el usuario. Algunos ejemplos de cajas de herramientas son la de procesado de la señal, sistemas de control, redes reuronales, lógica difusa, ondeletas, simulación, optimización y otras. De especial utilidad para la resolución del problema inverso en el caso de los Sondeos Eléctricos Verticales han sido las de optimización, procesado de la señal y estadística.

8.1.2. MATLAB y el tratamiento de matrices

Una de las ventajas básicas MATLAB es que posee dimensionamiento dinámico de sus variables, lo que significa no requiere dimensionamiento previo de sus arrays de elementos, lo cual lo vuelve extraordinariamente flexible, especialmente para los cálculos en formulaciones matriciales y vectoriales. MATLAB en realidad significa laboratorio matricial, e inicialmente fue concebido para obtener un acceso fácil al software desarrollado en los proyectos LINPACK y EISPACK. Hoy en día, MATLAB usa software desarrollado en los proyectos LAPACK y ARPACK que representan los conocimientos punteros en software para cálculo matricial.

En Matlab existe lo que se denomina sobredefinición de operadores⁷⁸. En consecuencia, MAT-LAB no maneja las matrices elemento a elemento, sino como un único objeto. Al mismo tiempo, permite acceder fácilmente a cada uno de sus elementos constituyentes y modificarlo e incluso variar sobre la marcha-añadiendo o eliminando elementos- las dimensiones del objeto (matriz) completo.

Además de en matrices, también pueden almacenarse datos numéricos y no numéricos de formas alternativas como los conjuntos de celdas y las estructuras.

⁷⁸También llamada polimorfismo de operadores. Esto significa que cada operador de Matlab, conoce el modo específico de operar sobre cada tipo de objeto.

8.1.3. Un tipo de objeto MATLAB: las estructuras

Las estructuras en MATLAB son conjuntos de datos multidimensionales a cuyos elementos se accede mediante indicadores textuales de campos. Por ejemplo, las órdenes:

Inversion.rutina='lsqnonlin'; Inversion.reiniciar='si';

al ser tecleadas en la ventana de comandos o ejecutadas mediante un fichero .m, crean un objeto de tipo estructura llamado *Inversion* que tiene dos *campos* distintos: *rutina*, cuyo contenido es el nombre de la rutina empleada en la inversión ('lsqnonlin') y *reiniciar*, que contiene una variable carácter que indica si, tras acabar el proceso actual, es necesario volver a comenzar.

En cada uno de los campos de esta estructura se pueden almacenar cualesquiera otros objetos que puedan ser manejados por MATLAB, con cualquier nivel de heterogeneidad. Es decir, dada una estructura, en uno de sus campos puede almacenarse una matriz; en otro una variable tipo carácter y en otro, por ejemplo, una nueva estructura en cuyos campos puede introducirse información adicional. Las siguientes facetas realtivas a las estructuras pueden facilitar muchísimo el almacenamiento y flujo organizado de información en la resolución de un problema inverso:

- 1. Capacidad para almacenar elementos de distinta clase,
- 2. Capacidad para representar organización jerarquizada (anidando estructuras),
- 3. Capacidad de dimensionamiento dinámico.

Ejemplo:

$$\begin{cases} & \text{tipo} & \text{variable carácter} \\ & \text{penalización} & \text{Matriz diagonal de penalización} \\ & \text{operador} & \text{Matriz operadora} \\ & \text{Vector_H} & \text{vector} \\ & \text{Modelo_priori} & \text{vector} \\ & \text{Modelo_priori} & \text{tipo1} \\ & \text{subobjeto2} & \text{tipo2} \\ & \dots & \dots \\ & \text{subobjeto n} & \text{tipo n} \end{cases}$$

En esta estructura, que es a su vez una subestructura de la estructura general de *opciones*, hay dos campos. El primero es una estructura también: "*opciones.inversion.fobj_regular*" que contiene varios campos. Se muestra, a la derecha del nombre de los mismos, los distintos tipos de objetos que pueden almacenar.

El segundo campo principal queda disponible para almacenamiento de cualquier otro tipo de variable que sea preciso. Pueden definirse campos adicionales sobre la marcha, sin haberlo previsto anticipadamente, como consecuencia del mencionado dimensionamiento dinámico.

Para acceder a los contenidos almacenados en los campos de una estructura se pueden utilizar idénticos procedimientos a los usados para referenciar una posición en una matriz. Se añaden los subíndices correspondientes al nombre del campo. Si el contenido del campo fuese algún

elemento estructurado, por ejemplo una matriz, se usarían nuevamente subíndices para referirse a la posición de cierto elemento dentro de esa matriz.

También existen funciones en Matlab que permiten obtener o modificar el contenido de cierto campo conociendo su nombre y su situación en la estructura⁷⁹.

En los programas para resolver el problema inverso, las estructuras pueden ser básicamente utilizadas para transportar, a lo largo del proceso, tres tipos de información, correspondientes a tres propósitos distintos:

- 1. Elementos decisores: Algunos campos contienen información que permite decidir entre dos o más alternativas. Por ejemplo el campo: opciones.inversión.regularización, cuyo contenido es la variable de tipo carácter 'si" o 'no" decide si en la función objetivo a minimizar debe considerarse un término de regularización o no.
- 2. Elementos indicadores: Los campos que contienen este tipo de elementos informan sobre el estado de algún otro campo de la estructura. Por ejemplo, en el caso de S.E.V, el campo suelo.indicador vale 1 si el campo suelo.formato_lineal está activo para la inversión y 0 si no lo está.
- 3. Elementos operativos: Estos campos almacenan la información precisa para llevar a cabo ciertas operaciones que ya se ha decidido previamente que son necesarias. Por ejemplo, en el caso anterior un ejemplo de elemento operativo sería el contenido del campo opciones.inversión.fobj_regular.vector_H que contiene un vector de referencia. Otro ejemplo de almacenamiento de elementos operativos lo aporta los campos: suelo.resistividades o datos.observados.valores que contienen, respectivamente, el vector de resistividades de un modelo y los valores de los datos observados en el campo.

La información adecuada es transportada en el proceso de modo parecido a como lo hacen los ocupantes de un autobús. Éstos se bajan del mismo cuando deben hacer algo (operar) y vuelven a subirse, más o menos modificados, cuando han finalizado. En la metáfora que compara al autobús con la estructura, cada pasajero tendría el asiento asignado de forma permanente. De este modo, una buena forma de saber lo que está pasando es mirar en las estructuras -en el interior del autobús- y comprobar lo que hay en los asientos.

Además, cuando estas estructuras son argumentos de entrada a funciones de Matlab, sólo se pasan los valores cuando, en el interior de la función, se modifica ese valor. Si la información almacenada en la estructura participa en otras operaciones, sin modificarse, o bien no se utiliza-Matlab pasa sólo la referencia al lugar de memoria donde se almacena la misma y no una copia. Esto aumenta mucho la eficiencia.

8.2. Estructuras relevantes para conceptualizar un problema inverso

El flujo de información que se produce en cualquier problema inverso queda ilustrado, de forma gráfica, en la siguiente figura:

(Importar figura)

⁷⁹En particular las funciones *qetfield* y *setfield*.

En la misma, aparecen como necesarios tres tipos de bloques de información que se corresponden con tres estructuras de MATLAB:

- 1. La estructura de datos,
- 2. la estructura de parámetros y
- 3. la estructura de opciones.

A continuación se explican la organización y contenidos de cada uno de estos bloques conceptuales.

8.2.1. Concepto y clasificación de los datos en un problema inverso

Resulta conveniente considerar dos tipos de cantidades como adecuadas para ser denominadas datos.

- En primer lugar datos como sinónimo de observaciones: son el punto de partida de un problema inverso. Estos datos observados convendrá clasificarlos en dos subclases distintas, empleando como criterio la génesis de los mismos. Esta división está sugerida por los objetivos que se espera cumpla el programa en su uso. Una propiedad que poseen los datos observados, es que tienen una tipología⁸⁰ completamente independiente de las tipologías de las variables con las que se opere para resolver el problema inverso.
- En segundo lugar para solucionar un problema inverso es necesario efectuar algún tipo de comparación-según un criterio conveniente-entre estos datos observados y las predicciones efectuadas por un modelo candidato a solución sugerido previamente. En el momento en el que vayan a ser comparados, los datos observados y las predicciones deberán estar en el mismo formato. Sin embargo, el formato de las predicciones puede estar dictado por consideraciones que afecten a la algoritmia de la solución, es decir por factores operacionales. La propuesta que se hace aquí es que estas predicciones sean consideradas como datos calculados, cuyo formato estará especificado por consideraciones operativas y será, por tanto, una propiedad operativa.

A la hora de comparar, el programa deberá tener en cuenta no sólo el par (dato calculado, datos observados) sino también sus tipologías respectivas a fin de reducirlas a la de uno de ellos. Hasta ese momento, datos calculados y datos observados serán dos entidades completamente independientes.

Si la realización de un programa que resuelva problemas inversos se entiende como un modo de solucionar problemas de la vida real, los datos son las cantidades que resultan de medir ciertas magnitudes físicas de interés. En el caso de los S.E.V, de realizar una campaña geofísica -con este tipo de dispositivo- en el campo. Combinando las diferencias de voltaje entre los electrodos de potencial, la medida de la corriente introducida y una constante geométrica dependiente del tipo de dispositivo se obtiene la magnitud resistividad aparente.

⁸⁰Se entiende por *tipología* de una variable almacenada en una estructura, el conjunto de propiedades que se predican de la misma, por ejemplo en S.E.V: unidades, si es logarítmica o no, si es resistividad aparente o transformada de resistividad,...

Complementariamente se puede entender el programa como herramienta de exploración de las propiedades que poseen los diferentes enfoques teóricos de resolución del problema inverso. Se trataría de construir un laboratorio en el que pudiesen generarse datos sintéticos sistemáticamente. Así pues, cabe hablar de dos tipos de datos observados, según la génesis que tengan:

- 1. datos de campo: obtenidos realmente a través de un proceso de medida.
- 2. Datos *sintéticos*: producidos en el ordenador, generalmente mediante una operación de resolución del problema directo seguida de una perturbación aleatoria controlada⁸¹.

Marcar explícitamente la diferencia entre ambos tipos de datos tiene sentido porque la primera fase en la resolución de un problema inverso es cargar los datos y la vía para realizar esta operación depende de si el dato es "de campo" o "sintético". A partir de ahí, son indiferenciables.

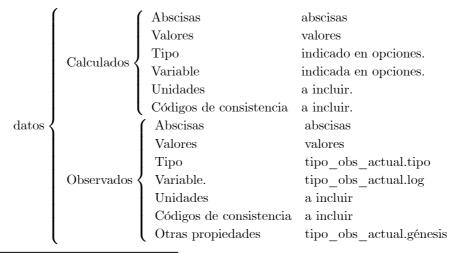
El diagrama de flujo muestra como la *génesis* de los datos es una propiedad de los mismos crítica al ejecutar el programa que los carga. En un caso el programa debe primero producirlos. En otro sólo debe ir a recogerlos.

(Introducir diagrama de flujo)

La génesis de los datos también establece diferencias en el momento en que deben ser utilizadas (y por tanto deben estar disponibles) las otras propiedades de los datos observados. Por ejemplo, en caso de que génesis sea sintéticos, ANTES de calcular los valores de los datos deben definirse las propiedades: abscisas, tipo de observación y formato. En cambio, si génesis es campo, las propiedades son "recuperadas" (desde el punto de vista del programa), simultáneamente con los datos, sus tipologías y sus formatos.

8.2.2. Estructura de almacenamiento de datos en el caso de S.E.V

El patrón organizativo general que sigue la estructura que en concreto se ha empleado es el siguiente.



⁸¹La generación de datos sintéticos es un procedimiento muy utilizado para comprobar la fiabilidad de los algoritmos.

Comentario

Los campos abscisas y valores, son autoexplicativos respecto a su contenido. El campo "Tipo", distingue distintos posibles cálculos que puedan considerarse como magnitud medida. Por ejemplo, en S.E.V podrían ser resistividades aparentes o transformadas de resistividad. El campo "variable" contiene información sobre si se ha producido un cambio de variable: por ejemplo, una transformación logarítmica. El campo "unidades" contiene las unidades en las que se expresan las magnitudes utilizadas. Los códigos de consistencia producen mensajes de error cuando se pretende introducir en algún campo de la estructura información no compatible con las propiedades especificadas.

8.2.3. Organización conceptual de la estructura de parámetros de modelo

En un problema inverso se buscan, precisamente, los parámetros desconocidos del modelo. Pero no es éste el único modelo que aparece. Para el cálculo del modelo incógnita es necesario, a lo largo del proceso y con distintos propósitos, especificar otros modelos. Por ejemplo, si se emplea un método local de optimización hay que aportar la denominada semilla o modelo inicial. A partir de éste y por medio de sucesivas iteraciones, se alcanza la solución final. La función del citado modelo inicial es realizar la primera simulación o predicción.

La marcha correcta del método de solución puede precisar la introducción de otros modelos diferentes con otros propósitos distintos al de simulación antes citado. Por ejemplo, a veces, con el objetivo de servir de información a priori (propósito), hay que especificar un modelo de referencia. En otras ocasiones, hay que limitar la región admisible (propósito) de búsqueda de la solución a la comprendida entre dos modelos-inferior y superior-. Hace falta reservar espacio para almacenar estas variables y estructurar su acceso.

Por lo tanto, el almacenamiento de los propios modelos utilizados así como de las propiedades atribuibles a los mismos, se puede jerarquizar en *propósitos*.

modelos
$$\begin{cases} \text{prop\'osito } 1^o \\ \text{prop\'osito } 2^o \\ \text{prop\'osito } 3^o \end{cases}$$

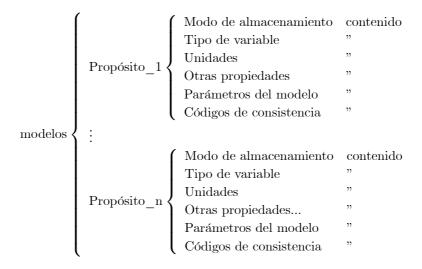
Para la realización de cada propósito, además de especificar los valores de los propios parámetros, hay que indicar las *propiedades* que poseen dichos valores y que permitirán su utilización posteriormente. En general, para cada propósito, deben ser enunciadas las siguientes propiedades:

- 1. Modo de almacenamiento: Los parámetros pueden ser almacenados siguiendo distintos patrones. Por ejemplo, en el caso de S.E.V la modalidad de patrón puede ser:
 - a) Modo matricial: Las resistividades de cada modelo son la fila de una matriz. Lo mismo ocurre con los espesores de cada modelo.
 - Modo vectorial: Resistividades y espesores son almacenados en un único vector (formato lineal)
- 2. Tipo de variable: Esta propiedad indica si los parámetros almacenados han sufrido algún tipo de transformación. Por ejemplo, si son logarítmicos o no lo son.

3. *Unidades*: Contiene información que indica el sistema de unidades y las unidades en las que están expresadas las magnitudes parametrizadas del modelo.

Finalmente, tras haber especificado la misión que cumple en el proceso y las propiedades del modelo aportado para cumplirla, deben introducirse los propios parámetros. Naturalmente, éstos deben cumplir las propiedades que se predican de ellos. Es decir, si previamente se ha dicho que el modo de almacenamiento es matricial no deberá introducirse en modo vectorial. Para evitar esto, debe incorporarse a la propia estructura código para comprobar la consistencia interna.

El patrón general organizativo para la estructura de modelos es:



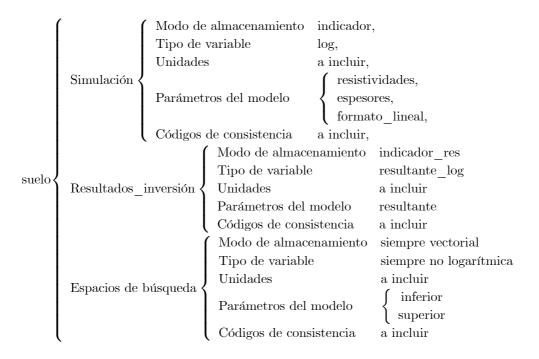
Otro modo de organizar esta información es el siguiente:

La estructura en árbol es ligeramente diferente. La primera tiene tantos campos primarios como funciones. Cada uno de estos campos contiene una subestructura con el conjunto de propiedades características de cada función. Una propiedad que sólo tenga sentido para una función, sólo aparece como campo en la subestructura de esa función. En las demás no existe.

En la segunda modalidad, la función es un campo más. De este modo, el conjunto de propiedades es común a todas las funciones. Si alguna de las propiedades no tiene sentido para alguna función, se puede atribuir el valor 'sin_sentido' o, simplemente dejarlo vacío de contenido. Los códigos de consistencia deben detectar esta circunstancia.

Las propiedades que se incluyen en esta estructura, son propiedades previas a la operación. Esto es, cuando se opera, puede ser necesaria una transformación de estas propiedades en otras más adecuadas al caso. Estas transformaciones auxiliares se realizarán mediante un proceso intermedio de formateo que, en el caso de S.E.V, se realizan mediante un programa denominado formatear.m

Según el patrón diseñado, la estructura suelo se ha diseñado de la siguiente forma:



8.2.4. Las opciones operativas

Los contenidos de esta estructura regulan todas las decisiones que deben tomarse a lo largo del proceso de solución y aportan los elementos operativos necesarios en cada parte del proceso.

La estructura está organizada en *bloques conceptuales* cada uno de los cuales es, también, una estructura. Cada bloque responde a una finalidad determinada. En el caso de los S.E.V la división en bloques es la siguiente:

```
opciones

Bloque básico.

Bloque criterio_error.

Bloque equivalencias.

Bloque inversión_general.

Bloque inversión_global.

Bloque inversión local.
```

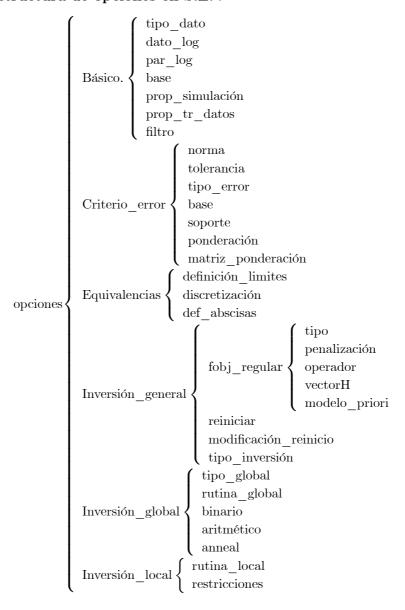
Debe comprenderse que es en esta estructura donde figuran la mayor parte de los elementos decisores. Es decir, información optativa de cuyo valor depende que se emprendan ciertas acciones. Por ejemplo, opciones.inversion.tipo_inversion es un campo cuyo contenido debe ser una variable de tipo carácter como 'global' o 'local'. Dependiendo del valor de este elemento decisor, el programa accederá a la los elementos operativos almacenados en el campo opciones.inversion.global para proceder a la inversión.

Una de las ventajas de contar con esta estructura, es que la misma es pasada como argumento de entrada de las distintas funciones que resuelven el problema y algunos de sus campos pueden ser modificados durante la ejecución. Esto permite extremar el rigor organizativo (ninguna información o variable está dispersa) y la eficacia de los códigos de control (que chequean la admisibilidad de cada variable que pretenda almacenarse en un campo dado).

Para que esta estructura funcione, sus contenidos deben poder sustituirse tanto individualmente como por bloques-subestructuras que agrupen a varios campos a la vez. Se ha programado una función para realizar esta tarea.

Asimismo, se precisa la acción de códigos de comprobación para admitir los contenidos como válidos o rechazarlos como inadecuados para un campo dado. En este sentido las órdenes 'switchcase' de MATLAB permiten hacer esto de manera estructurada y eficiente.

8.2.5. La estructura de opciones en S.E.V



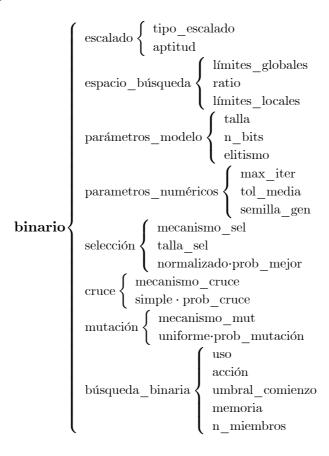
Comentarios

La estructura de opciones contiene los campos básicos de un problema inverso: informaciones relativas a los datos y/o del propósito de simulación (datos sintéticos o datos de

campo), informaciones relativas al criterio de error utilizado (tipo de norma y soporte de la misma, ponderación y posibles transformaciones logarítmicas); el campo relativo a la búsqueda de modelos equivalentes y finalmente campos relativos a la inversión: función objetivo, información a priori, algoritmos de inversión local y de inversión global.

La subestructuras de inversión globlal son:

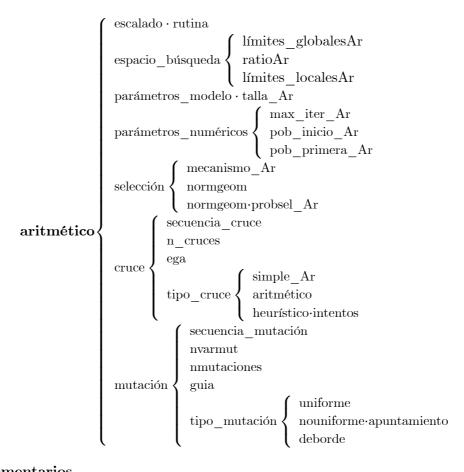
Algoritmos genéticos con codificación binaria



Comentarios

La estructura relativa a los algoritmos genéticos con codificación binaria contienen: el tipo de escalado para convertir el error de minimización en un escalar que haga las funciones de aptitud, el espacio de búsqueda de las soluciones potenciales (paralelepípedo n-dimensional), parámetros relativos a los modelos (número de bits de codificación, talla de las soluciones potenciales y presencia o ausencia de elitismo), parámetros numéricos relativos al máximo número de iteraciones, tipo de semilla (aleatoria o predefinida) y tolerancia especificada para la aceptación de un modelo elitista medio. Finalmente se especifican los campos relativos a las técnicas de selección, cruce, mutación y algoritmos específicos como el aumento binario, que es un método de creación de soluciones potenciales (en función de los bits más aptos) y que puede ser alternativo y/o complementario a los algoritmos de selección (Fan et al. 2000)

Algoritmos genéticos con codificación aritmética



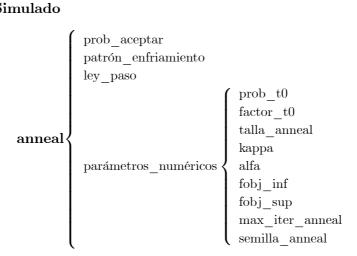
Comentarios

Los campos relativos a A.G con codificación aritmética poseen una estructura similar a la de los binarios. Una diferencia relevante es la posibilidad de éstos útimos de establecer secuencias de operación de mecanismos, combinando diferentes tipos de ellos -p.ej, varios tipos de cruce (resp. mutación, selección, etc) pueden usarse consecutivamente-.

Además, pueden definirse operadores genéticos exclusivos de la codificación aritmética (p.ej. el cruce aritmético). Por último el mecanismo denominado ega^{82} (Fan et Al, 2000) que optimiza el proceso de cruce.

 $^{^{82} \}mbox{Evolutionary Directional Operator}$

■ Recocido Simulado



Comentarios

Los campos de simmulated annealing aluden a las principales propiedades del algoritmo, detalladas en el capítulo correspondiente. Se han programado varias opciones para cada campo, especificadas en su momento. Cabe destacar el funcionamiento matricial del método no sólo en los modelos, sino también en los parámetros de cada opción. Por ejemplo, distintas temperaturas pueden evolucionar simultáneamente para cada modelo.

Capítulo 9: Conclusiones y futuras líneas de investigación

9. CONCLUSIONES Y FUTURAS LÍNEAS DE INVESTIGACIÓN

En la exposición de los contenidos de este trabajo, el acento se ha puesto en la obtención de una *metodología de análisis* de los problemas inversos. Algunas conclusiones relevantes obtenidas se enuncian a continuación:

- 1. Papel relevante en este trabajo posee la información a priori. Existen distintos modos de traducir dicha información en formulaciones operativas para un problema inverso. En particular, se ha propuesto una formulación de la misma en términos de pertenencia a cierto conjunto, utilizando la estructura de espacio afín. La idea, que se ha mostrado viable, es resolver un problema con restricciones utilizando un planteamiento sin restricciones. En este sentido, la posibilidad de redefinir la parametrización del espacio de parámetros haciendo uso de las parametrizaciones logarítmicas- facilita la tarea. Se ha propuesto un nuevo tipo de parametrización logarítmica que permite tratar de manera consistente y con generalidad restricciones de igualdad, proporcionalidad y pertenencia a un intervalo, permitiendo la solución como un problema de optimización sin restricciones.
- 2. La información a priori puede incluirse dentro de un término de penalización, que hace uso de operadores lineales y de los conceptos de normas (o seminormas). Este enfoque da mayor flexibilidad al anterior permitiendo introducir la idea de "grado de aproximación" al cumplimiento de las restricciones anteriormente mencionadas, de forma cuantificable.
- 3. En relación con el fenómeno de las equivalencias, lo tratado confirma la limitación de las denominadas reglas de equivalencia de Maillet. Se muestra que la organización de espesores y resistividades de las capas de terreno que sustentan dichas reglas hacen muy fiable la aproximación lineal del problema, siendo ésta la razón por la cual las regiones de equivalencia lineal siguen las citadas reglas. La comparación entre las zonas de equivalencia lineal y no lineal, para resistividades aparentes, muestra que las zonas lineales son más amplias confirmando que no se pueden aplicar indiscriminadamente las reglas de equivalencia de Maillet. Dentro de las posibles reparametrizaciones del espacio de modelos, se confirma que la parametrización logarítmica tiene cierto efecto linealizador de la región de equivalencia en S.E.V. Asimismo se ha empleado la (aquí denominada) parametrización log-α para posibilitar la obtención de modelos linealmente equivalentes con restricciones de desigualdad.
- 4. El problema tradicional de los algoritmos locales de optimización suele ser la convergencia a mínimos locales distintos del mínimo global. En el caso de S.E.V, el mínimo global parece ser único, pero está situado en zonas donde la función objetivo tiene forma de valle largo y estrecho. A lo largo del valle los gradientes son prácticamente nulos y los valores muy parecidos a los del mínimo global. En estas condiciones, la aproximación cuadrática de la función objetivo -obtenida por linealización del funcional del problema directocapta bastante bien la dirección y las dimensiones menores del valle aunque lo prolonga excesivamente en su eje mayor.
- 5. El análisis de la matriz jacobiana de la aproximación diferencial del funcional del problema directo se ha revelado clave para comprender el origen de las inestabilidades, avanzar métodos de solución y analizar el rendimiento de los mismos. La interpretación física de

- la estructura de dicha matriz se ha hecho comprensible gracias al empleo de herramientas numéricas como la descomposición (tb. generalizada) en valores singulares.
- 6. A la luz de la acción de estas herramientas sobre la matriz de sensibilidad se muestra que el problema discreto de S.E.V es un problema mal condicionado con una curva de valores singulares (tb. generalizados) que decrece (resp. crece) sin solución de continuidad. No obstante, en ocasiones, dado el reducido número de parámetros implicados puede tratarse como un problema de valores singulares agrupados.
- 7. Los métodos de regularización planteados son, se ha mostrado, al mismo tiempo que un modo de estabilizar el problema, una vía de entrada de la información a priori. En este sentido, los distintos métodos de regularización presentan ventajas e inconvenientes recomendándose, en general, la resolución con varios de ellos para considerarlos de forma comparada.
- 8. De las experiencias numéricas realizadas sobre el problema inverso lineal, se deduce que existe un intervalo relativamente pequeño del parámetro de regularización que proporciona soluciones razonables. En este intervalo, los efectos de los errores son adecuadamente compensados por la influencia de la información a priori.
 - La curva L se ha revelado como un instrumento muy útil para localizar este intervalo. No obstante, debe siempre tenerse en cuenta que el filtrado producido por el parámetro de regularización en la solución puede eliminar modelos útiles.
- 9. La elección de la incógnita del problema lineal tiene un gran efecto en la solución de cada iteración del método de optimización local. Si la incógnita es el incremento de la variable ("creeping"), ha quedado expuesto el proceso por el que distintas componentes no deseadas de las soluciones de la secuencia iterativa pueden ir acumulándose hasta la obtención de una solución final inservible, incluso bajo la influencia de distintos métodos de regularización. Asimismo, se ha analizado el efecto de estas técnicas cuando el problema lineal tiene por incógnita al propio modelo, observándose que la información a priori se incorpora así a las soluciones de la manera deseada.
- 10. Se ha comparado teóricamente el efecto de introducir la regularización en la propia formulación del problema inverso como un funcional de penalización, con la consideración de la regularización en cada paso lineal iterativo. Las simulaciones numéricas han mostrado resultados comparables entre la regularización global y la regularización interna con "jumping" (no así en el "creeping"). Si se escoge un método de optimización local, se recomienda el uso del propio modelo como incógnita.
- 11. Del trabajo con los métodos de optimización global se deduce que las dos técnicas programadas: algoritmos genéticos (A.G) y simulated annealing (S.A) son, ambos, capaces de encontrar el óptimo global en un tiempo razonable y, lo que es mucho más importante, de explorar la región de modelos aceptables dentro de cierta tolerancia de modo que puedan hacerse interpretaciones estadísticas de la nube de soluciones equivalentes. En particular, y dentro de un enfoque bayesiano, se han puesto a punto estos métodos para proceder a muestrear la distribución a posteriori orientada a la toma de decisiones.

- 12. Se ha probado, numéricamente, la capacidad de los algoritmos genéticos para reproducir las distribuciones en el muestreo del método de Metropolis-Gibbs, que es el teóricamente correcto para el muestreo de la distribución a posteriori. En particular, dos distribuciones muy apreciadas para la toma de decisiones como son las marginales y las regiones de cierta tolerancia de desajuste aparecen como casi coincidentes. La presencia de elitismo en la selección destruye esta capacidad.
- 13. La naturaleza global de los métodos probabilísticos citados permite la incorporación, de forma natural, de la información a priori dentro del propio funcional a minimizar.

En cuanto a los contenidos de futuros esfuerzos investigadores sugeridos por la realización de este trabajo, pudieran mencionarse los siguientes:

- 1. La inestabilidad del problema y el grado de regularización (o tratamiento) preciso para solventarla dependen de un buen conocimiento de la estructura estadística del error experimental. Parece, pues, que resultaría fructífero atacar este problema no sólo desde el punto de vista de descripción teórica del error, sino que, dado que resulta conocida la vinculación que existe entre error en el dato e incertidumbre en la solución, debiera tener por objetivo poder contribuir al diseño económico y práctico de una campaña de toma de datos suficiente para permitir caracterizar la incertidumbre en la solución obtenida.
- 2. De extraordinaria relevancia práctica se están revelando, especialmente en aplicaciones medioambientales someras, las técnicas tomográficas 2-D y 3-D. Con mínimas (o es posible que ninguna) modificaciones en la estructura de programación propuesta, se podría analizar dicho problema. Al menos dos nuevos niveles de conocimiento es preciso incorporar con este objetivo
 - a) el análisis del problema contínuo: y en particular el estudio de los efectos de la discretización, para el que podría utilizarse el mismo problema de S.E.V por ser 1-D.
 - b) Investigación de las técnicas iterativas de regularización: en general se precisará hacer hincapié en técnicas eficientes de resolución numérica, debido al aumento del número de variables del problema. Es interesante además, explorar las capacidades de Matlab en este sentido más relacionado con la eficiencia así como comprobar el rendimiento en alto número de dimensiones de los A.G y S.A.
 - c) Investigar en detalle la problemática relacionada con la elección de tipo y magnitud de las condiciones de contorno.
- 3. Analizar e investigar el problema inverso más en su vertiente de toma de decisiones, tratando de perfilar el marco teórico e informático que lo haría posible. En este sentido moverse en el marco bayesiano parece recomendable, pero también lo sería tratar, siempre que sea posible de encontrar referencias en el marco de la estadística frecuentista.
- 4. Estudiar la viabilidad de recoger "firmas geofísicas" de terrenos locales en diferentes situaciones. El conocimiento geológico e hidrogeológico regional ha mejorado sustancialmente en los últimos tiempos y hay, por tanto, a disposición cierta cantidad de información a priori que podría ser útil para encontrar soluciones consistentes.

- 5. Investigar la relación entre situaciones hidrogeológicas de "quimismo controlado" y sus correlatos geofísicos. Basándose en las experiencias que ya existen a escala de laboratorio, se podría tratar de abordar el tema a escala piloto en el campo.
- 6. Investigación pedagógica: Se trataría de realizar un esfuerzo serio en la elección, organización y presentación de los contenidos así como en la búsqueda de modos de trabajo conjunto para que estas técnicas puedan llegar a los profesionales del sector.
- 7. La organización de contenidos educativos en la web -de libre disposición- sería un tema, no menor, que seguramente impulsaría el uso de este tipo de conocimientos a nivel nacional.



10. BIBLIOGRAFIA

10.1. Artículos consultados

Referencias

- [1] Bhattacharya, B.B., Mrinal K.Sen., (2003), "Use of VFSA for resolution, sensitivity and uncertainty analysis in 1D DC resistivity and IP inversion", Geophysical Prospecting, 51, pp.393-408.
- [2] BERTERO, M., DE MOL, C., PIKE, E.R., (1985), "Linear inverse problems with discrete data I: General formulation and singular system analysis", *Inverse problems*, 1, pp.301-330.
- [3] Constable S.C, Parker R.L, Constable C.G., (1987), "Occam's Inversion: A practical algorithm for generating smooth models from electromagnetic sounding data", *Geophysics*, vol 52, n° 3, pp.289-300.
- [4] Curtis, A., Lomax, A., (2001), "Tutorial: Prior information, sampling distributions and the curse of dimensionality", *Geophysics*, vol 66, no 2, pp.372-378.
- [5] DITTMER, J.K., SZYMANSKY, J.E., (1995), "Stochastic inversion of magnetics and resistivity data using the simulated annealing algorithm", Geophysical Prospecting, 43, pp.397-416.
- [6] MUIUANE, E.A., PEDERSEN, L.B., (2001), "1D inversion of DC resistivity data using a quality-based truncated SVD", Geophysical prospecting, 49, pp.387-394.
- [7] FAN, H., WEI-ZHEN, J., ZONG-BEN, X., (2000), "An empirical comparison of three novel genetic algorithms", Engineering Computations, 17, 8, pp.981-1001.
- [8] Geman, S., Geman, D., (1984), IEEE Trans. Patt. Anan. Mach. Int. PAMI-6, 721
- [9] Goldberg D., (1991), "What every computer scientist should know about floating-point arithmetic", ACM. Computing Surveys, 23(1).
- [10] Gosh, D.P., (1971a), "The application of linear filter theory to the direct interpretation of geoelectrical resistivity sounding measurements", Geophysical prospecting, 19, pp.192-217.
- [11] Gosh, D.P., (1971b), "Inverse filter coefficients for the for the computation of apparent resistivity standard curves for a horizontally stratified earth", Geophysical prospecting, 19, pp.769-777.
- [12] HADAMARD, J.,(1923), "Lectures on Cauchy problem in linear partial differential equations", Yale University Press, New Haven, CT.
- [13] HOUCK, C.R., et Al (1995), "A genetic algorith for function optimization: a Matlab implementation", NCSU-IE TR.
- [14] INGBER, L., (1989), "Very fast simulated reannealing", Math. Comput. Modelling, vol 12, nº 8, pp.967-973.

[15] INGBER, L., ROSEN, B., (1992), "Genetic algorithms and very fast simulated reannealing: a comparison", *Math. Comput. Modelling*, vol 16, no 11, pp.87-100.

- [16] JOHANSEN, H.K., (1977), "A man/computer interpretation system for resistivity soundings over a horizontally stratified earth", Geophysical prospecting, 25, pp.667-691.
- [17] KAIPIO, J.P., KOLEHMAINEN, V., SOMERSALO, E., VAUKHONEN, M., (2000), "Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography", *Inverse problems*, 16, pp. 1487-1522.
- [18] KIRKPATRICK,S., GELATT, D.D., and VECCHI, M.P., (1983), "Optimization by simmulated annealing", Science, vol 220, pp.671-680.
- [19] LABRECQUE, D.J., MILETTO, M., DAILY, W., RAMIREZ, A., OWEN, E., (1996), "The effects of noise on Occam's inversion of resistivity tomography data". Geophysics, vol 61, no 2, pp.538-548.
- [20] METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A., TELLER, E., (1953), "Equations of state calculations by fast computing machines", *Journal of Chemical Physics*, vol.21, pp.1087-1092.
- [21] MOSEGAARD, K., TARANTOLA, A., (1995), "Monte Carlo sampling of solutions to inverse problems", Journal of geophysical research, vol 100, no B7, pp.12,431-12,447.
- [22] MRINAL, K.Sen, BHATTACHARYA, B.B., and STOFFA, P.L., (1993), "Nonlinear inversion of resistivity sounding data". *Geophysics*, vol 58, no 4, pp.496–507.
- [23] MRINAL K.Sen, Stoffa, P.L., (), "Nonlinear one-dimensional seismic waveform inversion using simulated annealing", *Geophysics*, vol 56, no 10, pp.1624-1638.
- [24] NORMARK, E., MOSEGAARD, K., (1993), "Residual statics estimation: scaling temperature schedules using simulated annealing". *Geophysical Prospecting*, 41, pp.565-578.
- [25] Parker, R.L., McNutt, M.K., (1980), "Statistics for the one-norm misfit measure". J.Geophys. Res, 85, pp.4429-4430-
- [26] Penna, T.J.P., (1995), "Fitting curves by simmulated annealing", Computers in physics, vol 9, no 3, pp.341-343.
- [27] PHILLIPS, D.L., (1962), "A technique for the numerical solution of certain integral equations of the first kind", J.Assoc. Comput. Math, 9, pp.84-97.
- [28] Porsani, J.M., Niwas, S., Ferreira, N.R., (2001), "Robust inversion of vertical electrical sounding data using a multiple reweighted least-squares method". *Geophysical Prospecting*, 49, pp.255-264.
- [29] Roy, G.I., (1999), "An efficient non-linear least squares 1D inversion scheme for resistivity and IP sounding data". Geophysical Prospecting, 47, pp.527-550.

[30] Scales, J.A., Docherty, P., Gersztenkorn, A., (1990), "Regularization of nonlinear inverse problems: imaging the near-surface weathering layer", *Inverse problems*, 6, pp.115-131.

- [31] Scales, J.A., Gersztenkorn, A., (1988), "Robust methods in inverse theory", *Inverse problems*, pp.1071-1091.
- [32] Scales, J.A., Snieder R., (1997), "To bayes or not to bayes?", Geophysics, vol 62, no 4, pp.1045-1046.
- [33] Scales, J.A., Snieder, R., (2000), "The anatomy of inverse problems", Geophysics, vol 65, no 6, pp.1708-1710.
- [34] Scales, J.A., Tenorio, L., (2001), "Prior information and uncertainty in inverse problems?", Geophysics, vol 66, no 2, pp.389-397.
- [35] Sharma, S.P., Kaikkonen, P., (1999), "Appraisal of equivalence and suppression problems in 1D EM and DC measurements using global optimization and joint inversion". *Geophysical prospecting*, 47, pp.219-249.
- [36] STARIOLO, D.A., TSALLIS, C., (1994), in Annual review of computational physics II, edited by D.Stauffer (World Scientific, Singapore, 1994).
- [37] Straub, A., (1995), "General formulation of the electric stratified problem with a boundary integral equation". *Geophysics*, vol 60, no 6, pp.1656-1670.
- [38] TARANTOLA, A., VALETTE, B., (1982), "Generalized nonlinear inverse problems solved using the linear least squares criterion", Rev. Geophys. Space Phys, 20, 2, pp.219-232.
- [39] TARANTOLA, A., VALETTE, B., (1982), "Inverse problems =quest for information", J. Geo-phys, 50, pp.159-170.
- [40] TIKHONOV, A.N., (1963), "Solutions of incorrectly formulated problems and the regularization method", Soviet Math. Dokl, 4, pp. 1035-1038.
- [41] Thompson, A.M., Kay, J.W., Titterington, D.M., (1989), "A cautionary note about crossvalidatory choice", J. Statist. Comput. Simul, vol 33, pp.199-216.

10.2. Libros consultados

Referencias

- [1] Brannan, D.A., Esplen, M.F., Gray, J.J., (1999), Geometry. Cambridge University Press, Open University.
- [2] Engl, H.W., Hanke, M., Neubauer, A., Regularization of inverse problems. Dordrecht/Boston/London, Kluwer academic publishers.
- [3] ETIENNE, R., (1999), Vitesses et techniques d'imagerie en sismique de réflexion: principes et méthodes. Paris, editions Technique et documentation.
- [4] Fuster, J.F., Olaizola, J., Olaizola, J., (1996), Curso de algebra lineal. Pamplona, ediciones Universidad de Navarra S.A.
- [5] GNEDENKO, B., (1996), Teoría de las probabilidades. Madrid-Moscú, Rubiños.
- [6] Goldberg, D.E., (1989), Genetic algorithm in search, optimization and machine learning. Add. Wesley.
- [7] Hansen, C., (1998), Rank deficient and discrete Ill-posed problems: numerical aspects of linear inversion. Philadelpia, Library of Congress.
- [8] IGLESIAS, T., (1998), Algoritmos genéticos generalizados: variaciones sobre un tema. servicio de publicaciones de la Universidad de La Coruña.
- [9] INGLE, V.K., PROAKIS, J.G., (2000), Digital Signal processing using Matlab. Canadá, Library of Congress.
- [10] KOEFOED, O., (1979), Geosounding principles, 1. Amsterdam, Elsevier Scientific publishing company.
- [11] LAKUBOVSKII, IU.V., LIAJOV, L.L., (1980), Exploración eléctrica, Moscú, Editorial Nedra.
- [12] Meju, M.A., (1994), Geophysical data analysis: understanding inverse problem theory and practice. USA, Society of exploration geophysics.
- [13] Menke, W., (1989), Geophysical Data analysis: discrete inverse theory (revised edition). California, Library of Congress.
- [14] MICHALEWICZ, A., (1992), Genetic algorithm + data structures = evolution programs. Springer Verlag.
- [15] MIETTINEN, K., MAKELA, M.M., NEITTAANMAKI, P., PERIAUX, J., (editores). Evolutionary algorithms in engineering and computer science. John Wiley & sons, ltd.1999, ISBN: 0 471 99902 4
- [16] Montgomery, D.C., Runger, G.C., (1996), Probabilidad y estadística aplicadas a la ingeniería. México, Mc Graw Hill.

[17] NOCEDAL, J., WRIGHT, S.J., (1999), Numerical optimization. USA, Springer series in operations research.

- [18] Orellana, E., (1982), Prospección geoeléctrica en corriente contínua. Madrid, ed. Paraninfo.
- [19] Otten, R.H.J.M and Van Ginneken, L.P.P.P., (1989), The annealing algorithm. USA, Kluwer Academic Publishers.
- [20] Patra, H.P., Nath, S.K., (1999), Schlumberger Geoelectric Sounding in Ground Water (principles, interpretation and application). A.A.Balkema.
- [21] Parker, R.L., (1994), Geophysical Inverse Theory. Princeton University Press.
- [22] ROBINSON, S.E., CORUH, C., (), Basic exploration geophysics. USA, John Wiley & Sons.
- [23] Scales, L.E., (1985), Introduction to non-linear optimization. Hong Kong, MacMillan Publishers ltd.
- [24] Scales, J.A., Smith, M.L., (1997), Introductory geophysical inverse theory (draft). Samizdat press.
- [25] Schilling, R. J., Harris, S.L., Applied numerical methods for engineers using MATLAB and C
- [26] Sharma, V.P., (1997), Environmental and engineering geophysics. U.K, Cambridge University Press.
- [27] SNIEDER, R., TRAMPERT, J., (2000), Inverse problems in geophysics. www-draft.
- [28] TARANTOLA, A., (1987), Inverse problem theory. Amsterdam, Elsevier Scientifif Publishing company.
- [29] Telford, W.M., Geldart, L.P., Sheriff, R.E., (1990), Applied geophysics. Cambridge University Press, second edition.
- [30] WEAVER, H.J., (1989), Theory of discrete and continuous Fourier analysis. USA, John Wiley & sons, ltd.