



Universidad Nacional de La Plata  
Facultad de Ciencias Astronómicas y Geofísicas

NUEVAS ESTRATEGIAS PARA LA INVERSIÓN  
SPARSE DE DATOS SÍSMICOS PRESTACK

Daniel O. Pérez

Director: Danilo R. Velis

Co-director: Mauricio D. Sacchi

Tesis presentada para obtener el grado académico de  
Doctor en Geofísica

LA PLATA, ARGENTINA  
- ABRIL DE 2015 -



NUEVAS ESTRATEGIAS PARA LA INVERSIÓN SPARSE DE DATOS  
SÍSMICOS PRESTACK

TESIS PRESENTADA EN CUMPLIMIENTO DE LOS REQUISITOS  
PARA OBTENER EL GRADO DE

DOCTOR EN GEOFÍSICA

en la

FACULTAD DE CIENCIAS ASTRONÓMICAS Y GEOFÍSICAS

UNIVERSIDAD NACIONAL DE LA PLATA

**Por:** Daniel Omar Pérez

**Director:** Dr. Danilo R. Velis

**Co-director:** Dr. Mauricio D. Sacchi

**Jurado:** Dra. Patricia Gauzellino

Dr. Juan Soldo

Dr. Deyan Draganov

Facultad de Ciencias Astronómicas y Geofísicas  
Universidad Nacional de La Plata  
Paseo del Bosque S/N  
(1900) La Plata  
Argentina

Abril 2015

© Daniel Omar Pérez, 2015.



# Agradecimientos

A mi familia y a mis amigos, por todo su apoyo.

A Danilo, por ser un excelente director con quien pude desarrollar mi doctorado con libertad, sin que faltase jamás su consejo y su vocación de guiar y enseñar.

A Mauricio, mi co-director, por sus valiosos aportes.

Al jurado, por dedicar su tiempo a la corrección de este trabajo. Su buena predisposición, comentarios y consejos hicieron de ésta una mejor tesis.

A mis compañeros y amigos del Departamento de Geofísica Aplicada.

A la Facultad de Ciencias Astronómicas y Geofísicas, la Universidad Nacional de La Plata y al CONICET.

A mi país, Argentina.



# Resumen

Uno de los objetivos centrales de la inversión de datos sísmicos *prestack* consiste en determinar contrastes entre las propiedades físicas de las rocas del subsuelo a partir de la información contenida en la variación en función del ángulo de incidencia de las amplitudes de las ondas sísmicas reflejadas en las interfaces geológicas. La inversión de datos sísmicos *prestack* es un problema mal planteado y mal condicionado, en el sentido de que pequeñas cantidades de ruido en el dato llevan a grandes inestabilidades en las soluciones estimadas. Además, debido a la naturaleza de los datos observados, que son ruidosos, incompletos y de banda limitada, coexiste el problema de la no-unicidad de las soluciones. Dichos problemas apremian la utilización de regularizaciones y restricciones con el fin de estabilizar el proceso de inversión y promover al mismo tiempo soluciones con alguna característica deseada. Las soluciones ralas, o *sparse*, son deseables debido a que permiten obtener reflectores bien definidos y de esa forma superar el problema de la baja resolución observada en las soluciones obtenidas por medio de métodos de inversión convencionales.

En este trabajo de tesis presentamos tres nuevas estrategias basadas en la utilización de diferentes regularizaciones que estabilizan el problema de inversión y promueven soluciones *sparse* a partir de datos sísmicos *prestack*. En la primera estrategia se procede a estimar soluciones sub-óptimas del problema de inversión regularizado mediante la norma  $l_0$  por medio de la utilización del algoritmo de optimización global *Very Fast Simulated Annealing* (VFSA). La segunda estrategia consta de dos etapas: primero se resuelve el problema de inversión regularizado mediante la norma  $l_1$  por medio de un eficiente algoritmo de optimización conocido como *Fast Iterative Shrinkage-Thresholding Algorithm* (FISTA) y luego se realiza un paso correctivo de las amplitudes estimadas utilizando mínimos cuadrados. Estas dos primeras estrategias permiten estimar con éxito soluciones *sparse* utilizando la aproximación de Shuey de dos términos, modelo que describe la variación con el ángulo de incidencia de los coeficientes de reflexión sísmica. La tercera estrategia utiliza como regularización la norma  $l_{2,1}$ , permitiendo incorporar información *a priori* por medio de matrices de covarianza o de escala. En este caso se estiman soluciones *sparse* de los parámetros de la aproximación de Aki & Richards de tres términos y, si la información *a priori* disponible es adecuada, es posible obtener también una estimación de tipo *blocky* de los parámetros elásticos del subsuelo.



# Índice general

Resumen	vii
<b>1. Introducción</b>	<b>1</b>
1.1. Antecedentes . . . . .	2
1.2. Motivación y alcances . . . . .	4
1.3. Objetivos . . . . .	4
1.4. Organización de la tesis . . . . .	5
<b>2. Problema inverso: regularización y optimización</b>	<b>11</b>
2.1. Introducción . . . . .	11
2.2. Formulación clásica del problema inverso . . . . .	13
2.3. Problemas de inversión lineales . . . . .	15
2.3.1. Problemas sobredeterminados . . . . .	16
2.3.2. Problemas subdeterminados . . . . .	18
2.3.3. Problemas mal planteados y mal condicionados . . . . .	20
2.4. Problemas de inversión no lineales . . . . .	22
2.5. Algoritmos de optimización . . . . .	23
2.5.1. Algoritmos de optimización local . . . . .	23
2.5.2. Algoritmos de optimización global . . . . .	25
2.6. Formulación Bayesiana del problema inverso . . . . .	26
2.6.1. Teorema de Bayes . . . . .	27
2.6.2. Ejemplos . . . . .	28
2.7. Sumario . . . . .	30
<b>3. Problema directo: de Zoeppritz a los datos sísmicos prestack</b>	<b>31</b>
3.1. Introducción . . . . .	31
3.2. Las ecuaciones de Zoeppritz . . . . .	32
3.2.1. Deducción de las ecuaciones de Zoeppritz . . . . .	33
3.3. Sobre las reflectividades . . . . .	38
3.4. Aproximaciones de las ecuaciones de Zoeppritz . . . . .	39
3.4.1. Aproximación de Aki y Richards . . . . .	40
3.4.2. Aproximación de Shuey . . . . .	41
3.4.3. Comparación entre la aproximación de Aki y Richards y la aproximación de Shuey . . . . .	42
3.4.4. Otras aproximaciones . . . . .	44
3.5. AVA y anisotropía . . . . .	45
3.6. Problema directo en AVA . . . . .	45
	<b>ix</b>

3.7. Sumario . . . . .	51
<b>4. Inversión de datos sísmicos <i>prestack</i> por mínimos cuadrados</b>	<b>53</b>
4.1. Introducción . . . . .	53
4.2. Datos sintéticos 1D . . . . .	54
4.3. Inversión por mínimos cuadrados . . . . .	54
4.4. Inversión por mínimos cuadrados amortiguados . . . . .	56
4.5. Conclusiones . . . . .	59
<b>5. Inversión <i>sparse</i> mediante <i>Simulated Annealing</i> y mínimos cuadrados</b>	<b>61</b>
5.1. Introducción . . . . .	61
5.2. Teoría . . . . .	63
5.3. Sobre <i>Very fast simulated annealing</i> . . . . .	64
5.4. Ejemplos . . . . .	65
5.4.1. Datos sintéticos 1D . . . . .	66
5.4.2. Datos sintéticos 2D . . . . .	72
5.4.3. Datos de campo . . . . .	75
5.5. Conclusiones . . . . .	77
<b>6. Inversión <i>sparse</i> mediante regularización con la norma <math>l_1</math></b>	<b>79</b>
6.1. Introducción . . . . .	79
6.2. Teoría . . . . .	81
6.2.1. Formulación bayesiana del problema . . . . .	83
6.3. Ejemplos . . . . .	86
6.3.1. Datos sintéticos 1D . . . . .	86
6.3.2. Datos sintéticos 2D . . . . .	96
6.3.3. Datos de campo . . . . .	99
6.4. Conclusiones . . . . .	102
<b>7. Inversión <i>sparse</i> mediante regularización con la norma <math>l_{2,1}</math></b>	<b>103</b>
7.1. Introducción . . . . .	103
7.2. Teoría . . . . .	104
7.3. Ejemplos . . . . .	109
7.3.1. Datos sintéticos 1D . . . . .	109
7.3.2. Datos sintéticos 2D . . . . .	115
7.3.3. Datos de campo . . . . .	120
7.4. Conclusiones . . . . .	123
<b>8. Conclusiones</b>	<b>125</b>
8.1. Conclusiones . . . . .	125
8.2. Discusión . . . . .	129
8.3. Contribución . . . . .	130
8.4. Trabajos a futuro . . . . .	131
8.5. Desarrollo . . . . .	132

<b>A. <i>Very fast simulated annealing</i></b>	<b>133</b>
A.1. Introducción . . . . .	133
A.2. <i>Simulated annealing</i> . . . . .	134
A.2.1. Función de generación de modelos . . . . .	135
A.2.2. Función de aceptación . . . . .	137
A.2.3. Esquema de enfriado . . . . .	137
A.3. <i>Very fast simulated annealing</i> . . . . .	138
A.3.1. Descripción del algoritmo . . . . .	138
A.3.2. Temperatura inicial . . . . .	139
<b>B. ISTA y FISTA</b>	<b>141</b>
B.1. Algoritmo ISTA . . . . .	141
B.1.1. Iteración de Landweber . . . . .	142
B.1.2. Función umbral . . . . .	143
B.1.3. <i>Iterative shrinkage-thresholding algorithm</i> . . . . .	144
B.2. Algoritmo FISTA . . . . .	145
<b>Bibliografía</b>	<b>147</b>



# Índice de figuras

2.1.	Ambigüedad del problema inverso . . . . .	12
2.2.	Clasificación de los problemas inversos. . . . .	13
2.3.	Problema directo, problema inverso y evaluación . . . . .	14
2.4.	Ajuste por mínimos cuadrados a un conjunto de observaciones. . . . .	17
2.5.	Función cuadrática de dos variables y curvas de nivel . . . . .	18
2.6.	Función de Rastrigin de dos variables y curvas de nivel. . . . .	22
3.1.	Ondas reflejadas y transmitidas en una discontinuidad. . . . .	35
3.2.	Coefficientes de reflexión $R_{PP}$ en función del ángulo de incidencia. . . . .	44
3.3.	Error debido a las aproximaciones a la ecuación de Zoeppritz. . . . .	44
3.4.	Parámetros de Aki y Richards y de Shuey del modelo de la Tabla 3.1. . . . .	49
3.5.	Dato sintético a partir del modelo de la Tabla 3.1. . . . .	50
4.1.	Modelo 1D de velocidades y densidad, y dato sintético con y sin ruido . . . . .	55
4.2.	Reflectividades de Aki y Richards y de Shuey del modelo 1D. . . . .	55
4.3.	Resultados de la inversión por mínimos cuadrados a partir del dato 1D. . . . .	56
4.4.	Número de condición vs parámetro de amortiguación $\mu$ . . . . .	57
4.5.	Estimación de los parámetros de amortiguación $\mu$ . . . . .	58
4.6.	Inversión por mínimos cuadrados amortiguados a partir del dato 1D . . . . .	59
4.7.	Datos estimados por los métodos de mínimos cuadrados . . . . .	59
5.1.	Dato sintético 1D, <i>intercept</i> $R_0$ y <i>gradient</i> $G$ . . . . .	66
5.2.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimados con VFSA a partir del modelo 1D . . . . .	68
5.3.	Comparación entre el dato observado y el estimado utilizando VFSA . . . . .	68
5.4.	<i>Misfit</i> vs número de reflectores $L_s$ . . . . .	69
5.5.	Efecto de la rotación de fase de la ondícula en el dato, para un reflector . . . . .	70
5.6.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimado utilizando una ondícula inexacta . . . . .	71
5.7.	Diferencia entre el dato observado (libre de ruido) y el estimado utilizando una ondícula inexacta . . . . .	72
5.8.	Modelo de velocidades y densidad 2D, <i>intercept</i> $R_0$ , <i>gradient</i> $G$ y <i>stack</i> . . . . .	73
5.9.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimado con VFSA a partir del dato sintético 2D, correspondientes desvíos estándar y <i>stacks</i> observado y estimado . . . . .	74
5.10.	Detalle del <i>intercept</i> y <i>stack</i> , verdaderos y estimados . . . . .	74
5.11.	Comparación entre el <i>gather</i> #1 observado y el estimado con VFSA. . . . .	75
5.12.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimados con VFSA a partir del dato de campo, correspondientes desvíos estándar y <i>stacks</i> observado y estimado . . . . .	76

## ÍNDICE DE FIGURAS

---

6.1.	Comparación entre la distribuciones gaussiana y laplaciana. . . . .	85
6.2.	Dato sintético 1D con y sin ruido, <i>intercept</i> $R_0$ y <i>gradient</i> $G$ . . . . .	86
6.3.	Estimación de los parámetros de compensación $\mu$ . . . . .	87
6.4.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimados con FISTA para diferentes niveles de ruido. . . . .	89
6.5.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimados con FISTA para diferentes valores de $\mu$ . . . . .	90
6.6.	Curva de Pareto y curvas de convergencia del FISTA. . . . .	91
6.7.	<i>Misfit</i> y valor de la norma $l_1$ versus el número de iteraciones del FISTA	91
6.8.	Curvas de Pareto para los datos con ruido. . . . .	92
6.9.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimados con FISTA+LS para diferentes niveles de ruido. . . . .	93
6.10.	Comparación entre las soluciones obtenidas utilizando mínimos cuadrados, VFSA, FISTA y FISTA+LS. . . . .	95
6.11.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ medios estimados a partir de 100 realizaciones del ruido . . . . .	96
6.12.	Modelo de velocidades y densidad 2D, <i>intercept</i> $R_0$ , <i>gradient</i> $G$ y <i>stack</i>	97
6.13.	Curva de Pareto para el primer <i>gather</i> del dato sintético 2D. . . . .	98
6.14.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimado con FISTA+LS a partir del dato sintético 2D y <i>stack</i> observado y estimado. . . . .	98
6.15.	Curva de Pareto para el primer <i>gather</i> del dato de campo . . . . .	99
6.16.	Comparación entre el <i>gather</i> #1 observado y el estimado con FISTA+LS.	100
6.17.	<i>Intercept</i> $R_0$ y <i>gradient</i> $G$ estimado a partir del dato de campo. . . . .	101
7.1.	Estructuras promovidas por las normas $l_2$ , $l_1$ , $l_{2,1}$ y $l_{1,2}$ . . . . .	107
7.2.	Modelo 1D de velocidades y densidades, y dato con y sin ruido. . . . .	110
7.3.	<i>Crossplots</i> entre las reflectividades. . . . .	111
7.4.	Valores estimados de $\alpha$ , $\beta$ y $\rho$ obtenidos por mínimos cuadrados y utilizando las normas $l_1$ y $l_{2,1}$ . . . . .	113
7.5.	Modelo 2D de velocidades, densidades y reflectividades. . . . .	115
7.6.	Valores estimados de $\alpha$ , $\beta$ y $\rho$ obtenidos por mínimos cuadrados y utilizando las normas $l_1$ y $l_{2,1}$ . . . . .	117
7.7.	Valores estimados de $R_\alpha$ , $R_\beta$ y $R_\rho$ obtenidos por mínimos cuadrados y utilizando las normas $l_1$ y $l_{2,1}$ . . . . .	118
7.8.	Valores estimados de $R_\alpha$ , $R_\beta$ y $R_\rho$ obtenidos las normas $l_1$ y $l_{2,1}$ con la matriz de escala. . . . .	120
7.9.	Comparación entre el <i>gather</i> #1 observado y el estimado utilizando la norma $l_{2,1}$ . . . . .	121
7.10.	$R_\alpha$ , $R_\beta$ y $R_\rho$ estimados a partir del dato de campo. . . . .	122
7.11.	<i>Stack</i> observado y estimado a partir del dato de campo. . . . .	123
A.1.	Diagrama de flujo del <i>Simulated annealing</i> . . . . .	136

# Índice de tablas

3.1. Modelos de velocidades $\alpha$ y $\beta$ , densidades $\rho$ y parámetros $\gamma$ . . . . .	42
3.2. Resumen de algunas aproximaciones a la ecuación de Zoeppritz para $R_{PP}$ . . . . .	46
3.3. Expresión de los parámetros $m_k$ y de las funciones $g_k(\theta_i)$ . . . . .	49



# Capítulo 1

## Introducción

El análisis de las variaciones de la amplitud de las ondas reflejadas en función del ángulo de incidencia, análisis llamado AVA (*Amplitude Variation with Angle*), es de gran importancia en el estudio del subsuelo para la detección de hidrocarburos y caracterización litológica (Chopra y Castagna, 2007). El análisis de AVA tiene su origen en la década de 1960, cuando se observaron en los datos sísmicos *poststack* anomalías de amplitud relacionadas con los depósitos de gas. La presencia de gas tiene como efecto una disminución considerable de la velocidad de propagación de las ondas compresionales  $P$ . Debido a esto también se observa un incremento en el valor del coeficiente de reflexión y se observa en el dato sísmico un reflector de amplitud anómala. Sin embargo estas anomalías de amplitud, llamadas *bright spots* (Sheriff, 2002), no sólo pueden ser consecuencia de la presencia de gas sino que también pueden deberse a cambios bruscos en la litología (Verm y Hilterman, 1995). Con el fin de separar el efecto sobre el dato de la presencia de fluidos del efecto debido a cambios bruscos en la litología se desarrollaron, en la década siguiente, técnicas y métodos basados en el análisis de la amplitud de los coeficientes de reflexión a partir de datos *poststack*. Ostrander (1984) finalmente introdujo una nueva clase de atributos sísmicos por medio del análisis de AVA. El análisis de AVA demostró ser capaz de proveer valiosa información para la caracterización del subsuelo y la prospección de hidrocarburos, ya que la variación de la amplitud no sólo depende del ángulo de incidencia sino también de las propiedades físicas del medio por el cual se propagan las ondas, y en particular del contenido de fluidos de las rocas. De esta variación se puede obtener estimaciones de parámetros elásticos tales como la velocidad de propagación de las ondas  $P$  y de las ondas  $S$ , la densidad  $\rho$  y el coeficiente de Poisson, entre otros.

Dada una onda plana que incide sobre una discontinuidad que separa dos medios adyacentes con propiedades diferentes, el AVA puede ser modelado por medio de las ecuaciones desarrolladas por Karl Bernhard Zoeppritz (1919). Debido a que no son lineales las ecuaciones de Zoeppritz resultan poco prácticas para ser aplicadas en problemas de inversión o de interpretación. Como consecuencia, desde la década de 1960 varios autores han desarrollado aproximaciones lineales de las ecuación de Zoeppritz como, por ejemplo, las aproximaciones de Bortfeld (1961), Aki y Richards (1980), Shuey (1985), Smith y Gidlow (1987), Fatti et al. (1994), Verm y Hilterman (1995). Los coeficientes de estas aproximaciones constituyen atributos AVA que sirven para caracterizar propiedades petrofísicas del subsuelo e indicar la presencia o no de hidro-

carburos (Ostrander, 1984; Castagna et al., 1998; Smith y Gidlow, 2000; Chopra y Castagna, 2007).

Dado un conjunto de datos *prestack* es posible desarrollar estrategias de inversión para estimar los valores de los coeficientes de las aproximaciones de las ecuaciones de Zoeppritz y, a través de estos, caracterizar el subsuelo. Sin embargo, al resolver el problema inverso se tiene que tener especial consideración en la existencia, unicidad y estabilidad de las soluciones estimadas (Aster et al., 2013). Al resolver un problema inverso la existencia de la solución no está garantizada, y en caso de existir podría no ser única. Además los problemas de inversión están en general mal condicionados; es decir, existen muchas soluciones que honran al dato con igual calidad, pero algunas de éstas pueden ser inestables y carecer de significado físico. Se pueden estimar soluciones más adecuadas regularizando el problema, es decir, promoviendo cierto tipo de soluciones mediante la minimización de una norma apropiada utilizada para incorporar información *a priori* útil al proceso de inversión (Tarantola, 2005; Ulrych y Sacchi, 2005). Por ejemplo, la simplicidad de las soluciones es una propiedad que puede ser obtenida por medio de diferentes tipos de regularizaciones. Desde el punto de vista del problema de inversión de datos sísmicos, estimar soluciones simples implica estimar soluciones ralas, o *sparse*, de los coeficientes de reflexión y/o soluciones *blocky* de las impedancias o de los parámetros elásticos. Una solución de tipo *sparse* es aquella que puede representarse por medio de un vector en el cual la mayoría de sus elementos son cero, ocupando los valores no nulos unos pocos elementos aislados. Por otro lado, una solución de tipo *blocky* es aquella que es constante por tramos. Las soluciones tipo *sparse* y/o *blocky* son deseables ya que a través de ellas es posible resolver mejor los reflectores cercanos, superando el problema de baja resolución de las soluciones obtenidas, por ejemplo, con el método de mínimos cuadrados (Debeye y van Riel, 1990). A pesar de que se sabe a través del estudio de pozos que la estructura real del subsuelo es más bien continua que *sparse* (Cooke y Schneider, 1983), es también verdad que en general la distribución de las amplitudes de los coeficientes de reflexión no es gaussiana (Walden y Hosken, 1986; Velis, 2003). Más aún, las secuencias de reflectividad derivadas de estudios de pozo indican que las principales unidades litológicas pueden asociarse con grandes coeficientes de reflexión que muestran un comportamiento esencialmente *sparse* (Walden y Hosken, 1986). Debido a esto es importante contar con algoritmos capaces de obtener soluciones *sparse* de los coeficientes de reflexión. Por ejemplo, las soluciones *sparse* han sido aplicadas con éxito en problemas de deconvolución de datos sísmicos por varios autores asumiendo que la reflectividad sigue una distribución de probabilidad laplaciana, lo que conduce a regularizar el problema mediante la norma  $l_1$  (Wiggins, 1978; Taylor et al., 1979; Oldenburg et al., 1983; Sacchi et al., 1996; Sacchi, 1997; Robinson y Ebrom, 1996, p.ej.).

### 1.1. Antecedentes

En el contexto de la inversión de datos sísmicos *prestack* se han desarrollado gran cantidad de métodos para obtener soluciones de diversos atributos con interesantes resultados. A continuación daremos una breve descripción de algunos trabajos representativos. La variedad de técnicas empleadas para la resolución del problema, así como la variedad de tipos de soluciones que se obtienen, hace difícil su clasificación. Si-

guiendo la temática principal de esta tesis separaremos aquellas técnicas que permiten obtener soluciones *sparse* y/o *blocky* de aquellas que no. A pesar de esta diferenciación hay características que son comunes a todos los métodos. Por ejemplo, es sabido que a partir de observaciones con ruido y con un rango limitado de ángulos de incidencia no es sencillo estimar simultáneamente los coeficientes de las aproximaciones de las ecuaciones de Zoeppritz sin tener alguna clase de información *a priori* acerca de las características esperadas de las soluciones y/o restricciones geológicas adicionales (Cambois, 1998a; Downton y Lines, 2001). Una forma de superar este problema es por medio de la utilización de una matriz de covarianza o matriz de escala que permita incorporar información de la correlación entre los parámetros a estimar. El desarrollo de los métodos de inversión, tanto los que promueven soluciones *sparse* como los que no, muchas veces están basados en la teoría de Bayes (1763). Esta teoría permite regularizar el proceso de inversión de manera sencilla, asumiendo que la solución sigue una función de densidad de probabilidad *a priori*.

Entre los métodos de inversión convencionales, podemos citar los trabajos de Simons y Backus (1996), quienes propusieron un método para estimar soluciones de las reflectividades de la aproximación de Aki y Richards asumiendo la validez de las ecuaciones de Gardner et al. (1974) y Castagna et al. (1985). En este método se incorpora información *a priori* a través de la utilización de matrices de covarianza. Por medio del enfoque bayesiano Buland y Omre (2003) propusieron utilizar la función de probabilidad *a priori* gaussiana para estimar soluciones de la velocidad de la onda  $P$ , velocidad de la onda  $S$  y de la densidad, método que también incorpora información *a priori* utilizando una matriz de covarianza. Una estrategia similar fue utilizada por Downton y Lines (2001, 2004). Hampson et al. (2005) y Hampson y Russell (2013) proponen una técnica para la estimación simultánea de las impedancias de las ondas  $P$  y  $S$ , y de la densidad basada en la aproximación de Aki y Richards. Dicha técnica asume que existe una relación lineal entre los logaritmos de las magnitudes a estimar, y los resultados se obtienen perturbando un modelo inicial de la impedancia de la onda  $P$ , por lo que es necesario contar con dicha información *a priori*. El método propuesto por Ma (2002) hace uso del VFSA, algoritmo de optimización global que también se utilizará en esta tesis, para obtener soluciones de las impedancias de las ondas  $P$  y  $S$ . Por otro lado, en el contexto de la inversión *sparse* y/o *blocky*, Alemie y Sacchi (2011) se basan en la teoría de Bayes y utilizan la distribución de probabilidad de Cauchy multivariable, como función *a priori*, para estimar reflectividades *sparse* asociadas a la aproximación de Aki y Richards. Con una estrategia similar, Downton y Lines (2003) obtienen soluciones *sparse* de las reflectividades asociadas a las impedancias de las ondas  $P$  y  $S$  y al *fluid stack*, asumiendo que dichas reflectividades siguen una distribución de probabilidad *a priori* laplaciana. En los trabajos de Jensås (2008); Eidsvik y Theune (2009) y Theune et al. (2010) se analiza el uso de las distribuciones de Gauss, Cauchy y Laplace como funciones de distribución *a priori* de las reflectividades de Aki y Richards para obtener soluciones *blocky* de las velocidades  $P$  y  $S$ , y de la densidad  $\rho$ . Con el mismo propósito Misra y Sacchi (2007, 2008) propusieron utilizar una combinación de VFSA y un filtro *edge-preserving-smoothing* (EPS). Mallick (1995) propone estimar soluciones *blocky* asumiendo que el subsuelo se compone de un número grande, pero finito, de capas y utilizan el algoritmo de optimización global *Genetic algorithm* (GA) para encontrar el mínimo de la función de costo del problema. Zhang et al. (2013) propusieron una

estrategia de inversión de tipo *basis pursuit* basada en la construcción de un diccionario de reflectividades que les permitió estimar soluciones *sparse* de las reflectividades asociadas a la onda  $P$  y a la onda  $S$ .

## 1.2. Motivación y alcances

Los métodos descritos son sólo una parte del basto, y en constante crecimiento, campo de estudio de la inversión de datos *prestack*. Aunque estos métodos comparten el objetivo común de estimar parámetros y atributos que sean útiles para la caracterización del subsuelo, son diferentes entre sí y, como toda técnica de inversión, presentan ventajas y desventajas. Algunos estiman soluciones que carecen de resolución, haciendo difícil su posterior interpretación. Otros requieren la utilización de matrices de covarianza para poder estimar correctamente las soluciones, sin embargo dichas matrices no son fáciles de estimar aún cuando la información *a priori* necesaria para ello esté disponible. Algunos de los métodos requieren de la inversión de matrices, hecho que puede ser muy costoso computacionalmente cuando se realiza la inversión de grandes volúmenes de datos. Por ello, y con el fin de superar algunos de los inconvenientes o limitaciones de los métodos mencionados, es necesario explorar el uso de técnicas alternativas para la resolución de este tipo de problema inverso, que sean tanto eficientes como eficaces desde el punto de vista computacional. Estas técnicas deberán permitir la estimación de parámetros tales como las reflectividades y/o las velocidades y densidades de las rocas valiéndose de estrategias de regularización que promuevan soluciones de tipo *sparse* (o *blocky* según corresponda) que honren al dato observado y que permitan, a su vez, incorporar la información *a priori* disponible de una manera simple y práctica.

## 1.3. Objetivos

El objetivo principal de esta tesis es el desarrollo de estrategias alternativas para la inversión de datos sísmicos *prestack*, que permitan superar los inconvenientes observados en los métodos descritos. En este sentido se espera obtener soluciones *sparse* para la estimación simultánea de diversos parámetros y atributos sísmicos y/o petrofísicos de interés como, por ejemplo, el *intercept* y *gradient* de la aproximación de Shuey de dos términos (Shuey, 1985) o las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  de la aproximación de Aki y Richards de tres términos (Aki y Richards, 1980). También se espera obtener, cuando la información disponible lo permita, soluciones de tipo *blocky* de las velocidades de propagación de las ondas  $P$  y  $S$ , y de la densidad. Con este fin, se utilizarán técnicas bayesianas para justificar y definir el uso de ciertas normas para una correcta regularización del problema inverso, y diferentes algoritmos de optimización eficientes basados en metaheurísticas de trayectoria y extensiones del método clásico del gradiente. Específicamente, se prevé:

- Analizar el uso de diferentes tipos de regularización del problema inverso, no sólo con el objetivo específico de obtener soluciones de tipo *sparse* y/o *blocky*,

sino también con el objetivo de incorporar convenientemente información adicional proveniente de otras fuentes (como ser correlaciones entre los parámetros obtenidas a partir de datos de pozo cercanos);

- Utilizar el enfoque bayesiano como una herramienta útil para derivar, cuando sea posible, las funciones de costo correspondientes a cada regularización planteada;
- Obtener soluciones *sparse* de los atributos *intercept* y *gradient* regularizando el problema de inversión mediante la norma  $l_0$ . Analizar el uso del algoritmo de optimización global VFSA como herramienta para minimizar la función de costo no lineal resultante;
- Obtener soluciones *sparse* de los atributos *intercept* y *gradient* basadas en la utilización la norma  $l_1$  como término de regularización. Analizar el uso del algoritmo proximal *Fast iterative shrinkage-thresholding algorithm* (FISTA) como herramienta para minimizar la función convexa resultante, y técnicas de *debiasing* para corregir las amplitudes de los atributos estimados;
- Estudiar la utilización de la norma mixta  $l_{2,1}$  como término de regularización con el fin de obtener soluciones *sparse* de las reflectividades de la aproximación de Aki y Richards de tres términos. En este contexto incorporar información *a priori* por medio de la utilización de matrices de covarianza o escala estimadas de datos de perfilaje de pozo, con el fin de restringir las amplitudes de las soluciones estimadas. Cuando el dato *a priori* lo permita, estimar soluciones *blocky* de los parámetros elásticos del modelo;
- Aplicar las técnicas desarrolladas a datos sintéticos realistas y a datos de campo con el fin de obtener estimaciones simultáneas de diferentes parámetros de interés para caracterizar el subsuelo con fines exploratorios. Realizar un análisis de las soluciones estimadas, cuantificando los errores de los atributos obtenidos.

## 1.4. Organización de la tesis

La tesis consta de ocho capítulos y dos apéndices. En los primeros tres capítulos, de los cuales esta introducción es el primero, planteamos las bases teóricas sobre las cuales desarrollamos los métodos de inversión presentados. En el capítulo 4 realizamos la inversión de datos sísmicos *prestack* por medio del método convencional de mínimos cuadrados. Los capítulos 5, 6 y 7 son el núcleo de este trabajo, allí presentamos tres nuevas estrategias para la inversión *sparse* de datos sísmicos *prestack*. En el capítulo 8 enunciamos las conclusiones en base a los resultados obtenidos en los capítulos anteriores y hacemos una breve discusión de los resultados obtenidos. En los apéndices A y B detallamos los métodos de optimización utilizados para resolver los problemas de inversión.

A continuación se hará una breve descripción de los próximos capítulos.

### Capítulo 2: Problema inverso: regularización y optimización

En el capítulo 2 hacemos un breve resumen de la teoría de inversión. Analizamos las formulaciones clásica y bayesiana del problema inverso. Dentro de la primera formulación se clasifican los problemas en lineales y no lineales, siendo los problemas lineales a la vez clasificados en sobredeterminados y subdeterminados. Asimismo se concluye que el método de mínimos cuadrados es adecuado para resolver problemas inversos lineales sobredeterminados, mientras que mínimos cuadrados amortiguados es adecuado para resolver problemas lineales subdeterminados. También estudiamos los problemas de inversión mal condicionados, ya que gran parte de los problemas de inversión en geofísica pertenecen a esta categoría. Para los problemas mal condicionados discutimos la regularización del mismo como un medio de poder estabilizar el proceso de inversión. También analizamos las complejidades que surgen de una relación no lineal entre el dato y los parámetros del modelo. En este mismo capítulo se puede encontrar un breve resumen sobre los métodos de optimización locales y globales, y hacemos una descripción de aquellos más conocidos y utilizados.

### Capítulo 3: Problema directo: de Zoeppritz a los datos sísmicos *prestack*

En el capítulo 3 resumimos la teoría de la variación de la amplitud de las ondas reflejadas en función del ángulo de incidencia (AVA). Se derivan las ecuaciones de Zoeppritz y hacemos una revisión de las aproximaciones lineales de las mismas, enfatizando las aproximaciones de Shuey de dos términos (Shuey, 1985) y de Aki y Richards de tres términos (Aki y Richards, 1980) ya que serán utilizadas en las estrategias de inversión desarrolladas en los capítulos 5, 6 y 7. También llevamos a cabo un análisis comparativo de estas dos aproximaciones. Sobre el final de este capítulo planteamos la resolución del problema directo que utilizaremos al resolver problemas de inversión. Dicha resolución está basada en el modelo convolucional, enumerándose las hipótesis en las cuales se basa y obteniéndose una expresión matricial para generar datos sintéticos a partir de las reflectividades y de la ondícula. A modo de ejemplo se generan datos sintéticos a partir de modelos que presentan bajo y alto contraste de sus parámetros elásticos, utilizando la ecuación de Zoeppritz, la aproximación de Aki y Richards y la aproximación de Shuey de dos términos, y luego se realiza una comparación cualitativa de los mismos.

### Capítulo 4: Inversión de datos sísmicos *prestack* por mínimos cuadrados

En el capítulo 4 analizamos el problema de inversión de datos sísmicos *prestack* sintéticos 1D utilizando los métodos de mínimos cuadrados y mínimos cuadrados amortiguados, contemplando la aproximación de Shuey de dos términos y la aproximación de Aki & Richard de tres términos. Para el caso de mínimos cuadrados amortiguados estimamos el parámetro de amortiguación mediante el criterio de discrepancia. También realizamos un breve análisis del número de condición. Para ambos métodos se ve que las soluciones estimadas, a pesar de ajustar al dato observado, no se corresponden con las soluciones esperadas. En el caso del método de mínimos cuadrados sin amortiguar esto ocurre porque el problema está mal condicionado y por lo tanto las soluciones estimadas son inestables, dando como resultado parámetros del modelo sin significado

físico y con amplitudes sobreestimadas. Cuando se realiza la inversión utilizando el método de mínimos cuadrados amortiguados el número de condición disminuye, sin embargo las soluciones estimadas siguen siendo poco aceptables. Esto se debe a que el método de mínimos cuadrados amortiguados contempla la regularización cuadrática de los parámetros del modelo y este tipo de regularizaciones promueve la suavidad en las soluciones, por lo que el carácter *sparse* de la solución esperada no será recuperado. Debido a esto planteamos la necesidad de regularizar el problema por medio de una norma que estabilice las soluciones y a la vez promueva soluciones de tipo *sparse*.

### Capítulo 5: Inversión *sparse* mediante *Simulated Annealing* y mínimos cuadrados

En el capítulo 5 presentamos la primera estrategia para obtener soluciones *sparse* a partir de datos sísmicos *prestack*. En la estrategia planteada se procede a estimar soluciones sub-óptimas del problema de inversión regularizado mediante la norma  $l_0$ . Dichas soluciones se estiman por medio de la utilización del algoritmo de optimización global VFSA (Ingber, 1989). Por medio de VFSA se establecen las posiciones en tiempo de un número fijo y pequeño de coeficientes de reflexión para luego estimar, en cada iteración, las amplitudes de los coeficientes de la aproximación de Shuey de dos términos por medio de la resolución de un problema de mínimos cuadrados de pequeña dimensión. Este mismo enfoque fue utilizado por Velis (2008) en la deconvolución de tipo *sparse-spike* de datos *poststack*. La técnica propuesta en este capítulo es una extensión para datos *prestack* del método de deconvolución mencionado. Previo a realizar el proceso de inversión establecemos, por medio del criterio de la curva-L, el número mínimo de reflectores necesarios para ajustar el dato observado con una dada tolerancia. Como parte del estudio del comportamiento del método propuesto, se realizó un análisis de las soluciones estimadas cuando la ondícula utilizada en la inversión es inexacta. Una ventaja del método de inversión propuesto es que, al ser VFSA un algoritmo estocástico, es posible estimar la incertidumbre de las soluciones. La estrategia propuesta demuestra ser robusta, permitiendo estimar soluciones *sparse* de la aproximación de Shuey de dos términos que honran al dato observado en presencia de altos niveles de ruido, tanto para datos sintéticos como para datos de campo. Sin embargo, el alto costo computacional inherente a la utilización de un algoritmo de optimización global hace que esta técnica sea poco práctica para procesar grandes volúmenes de datos.

### Capítulo 6: Inversión *sparse* mediante regularización con la norma $l_1$

En el capítulo 6 presentamos la segunda estrategia para obtener soluciones *sparse* a partir de datos sísmicos *prestack*. El método apunta a obtener simultáneamente la localización en tiempo de los reflectores más importante y los valores de los atributos AVA *intercept* y *gradient*. El método consta de dos etapas. La primera etapa es utilizada para estimar la ubicación en tiempo de los reflectores y consiste en resolver el problema de inversión utilizando la norma  $l_1$  como regularización, problema conocido como LASSO. Dicha regularización promueve soluciones *sparse* que pueden ser estimadas utilizando un eficiente algoritmo de optimización conocido como *Fast Iterative Shrinkage-Thresholding Algorithm* (FISTA) (Beck y Teboulle, 2009). En una segunda

etapa se procede a realizar un paso correctivo (*debiasing step*) por medio de un ajuste por mínimos cuadrados de las amplitudes. El paso correctivo es necesario ya que, debido a la naturaleza del dato y a la regularización utilizada, las soluciones obtenidas al resolver el LASSO presentan amplitudes severamente subestimadas. En la primera parte del capítulo explicamos el fundamento teórico de la estrategia propuesta y se hace un breve resumen de las características de FISTA. Luego procedemos a realizar la inversión de datos sintéticos sin realizar el paso correctivo, donde se observa que los resultados están lejos de ser aceptables debido a que el ruido presente en el dato influye considerablemente sobre la respuesta AVA. Luego probamos la estrategia de dos etapas viendo que las soluciones obtenidas, tanto sobre datos sintéticos como sobre datos de campo, son mucho más precisas que las obtenidas sin la etapa correctiva. Además se puede apreciar que las soluciones son comparables en calidad con aquellas estimadas con la estrategia presentada en el capítulo anterior, pero a una fracción del costo computacional. Como parte del análisis del método propuesto realizamos un análisis estadístico del mismo, utilizando datos con diferentes realizaciones de ruido. También realizamos una comparación entre las soluciones obtenidas con el método presentado en este capítulo y las obtenidas mediante mínimos cuadrados y VFSA. Debido a la calidad de los resultados obtenidos sobre el dato sintético realizamos una inversión sobre datos de campo, obteniendo soluciones *sparse* del *intercept* y el *gradient* que honran al dato observado.

### Capítulo 7: Inversión *sparse* mediante regularización con la norma $l_{2,1}$

La estrategia presentada en el capítulo 7 hace uso de la norma  $l_{2,1}$  como regularización con el propósito de obtener soluciones *sparse* utilizando la aproximación de Aki & Richards de tres términos. Esta regularización promueve soluciones de tipo *sparse* con estructura que pueden ser estimadas utilizando FISTA y permiten una mejor estabilización del proceso de inversión. A la vez es posible incorporar información *a priori* por medio de matrices de covarianza o de escala. La estrategia propuesta permite, si la información *a priori* disponible es adecuada, obtener también una estimación de tipo *blocky* de los parámetros elásticos del subsuelo. Además es versátil y fácil de adaptar a la información *a priori* disponible. Probamos el método propuesto en datos sintéticos 1D y 2D con altos niveles de ruido, y en datos de campo 2D. Sobre datos sintéticos se ve que, dada la información *a priori* adecuada, el método es capaz de obtener soluciones *blocky* de los parámetros elásticos con buena continuidad lateral y una correcta estimación de las magnitudes, a la vez que es posible obtener soluciones *sparse* de las reflectividades que honran a las verdaderas. También observamos que bajo condiciones en las cuales la información *a priori* incorporada no está completa o es incierta las soluciones estimadas utilizando la norma mixta  $l_{2,1}$  como término de regularización superaron en calidad a aquellas obtenidas utilizando la norma  $l_1$ . A partir de datos de campo, debido a la falta de información *a priori* adecuada, estimamos soluciones *sparse* de las reflectividades solamente. Las soluciones estimadas muestran, en general, una marcada característica *sparse* a la vez que presentan buena continuidad lateral y honran al dato observado.

### Capítulo 8: Conclusiones

En este capítulo se resumen las conclusiones obtenidas a partir de los capítulos 5, 6, 7. Además se describen los aportes, contribuciones y desarrollos resultantes de esta tesis y se enumeran las publicaciones realizadas. También se propone trabajos a futuro a partir de los resultados obtenidos de las estrategias presentadas.



# Capítulo 2

## Problema inverso: regularización y optimización

### 2.1. Introducción

La teoría de inversión da herramientas matemáticas para poder estimar los parámetros que caracterizan a un sistema físico a partir de datos observados e información *a priori*. Dicho conjunto de parámetros es llamado modelo. El dato observado son las observaciones realizadas experimentalmente previas a resolver el problema inverso, mientras que la información *a priori* es todo conocimiento independiente de las observaciones que se tiene sobre el modelo. Para resolver el problema de inversión se debe combinar correctamente las observaciones y la información *a priori*. Para resolver el problema inverso es necesario tener suficiente conocimiento de la física del problema como para poder generar datos sintéticos a partir del modelo. A este procedimiento se lo conoce como resolución del problema directo. La resolución de problemas inversos se ha extendido a casi todos los campos de la ciencia que hacen uso de la matemática como lenguaje, entre los que se encuentra la Geofísica.

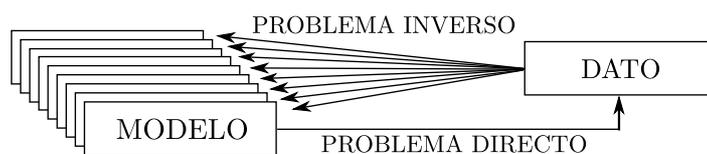
El estudio de un sistema físico puede dividirse en tres pasos (Tarantola, 2005):

1. Parametrización del problema: Es el procedimiento de establecer el mínimo número de parámetros con los que se puede caracterizar un sistema físico. En general la parametrización no es única, y la elección del conjunto de parámetros dependerá del problema a resolver y del tipo de solución buscada. El número de parámetros puede ser finito o infinito. Dos parametrizaciones son equivalentes si están relacionadas por una función biyectiva.
2. Problema directo: Es el proceso matemático de generar datos sintéticos a partir del conjunto de parámetros del modelo e información independiente apropiada. En general el dato sintético no será igual al dato observado debido a que tanto las leyes de la física utilizadas como el modelo son aproximaciones de la realidad. Además, las observaciones poseen ruido no contemplado por el modelo utilizado para resolver el problema directo. Resolver el problema directo puede representar un gran porcentaje del tiempo total del proceso de resolver un problema de inversión. Es importante analizar cuál problema directo permite generar datos sintéticos de la manera más sencilla y eficiente.

## 2. Problema inverso: regularización y optimización

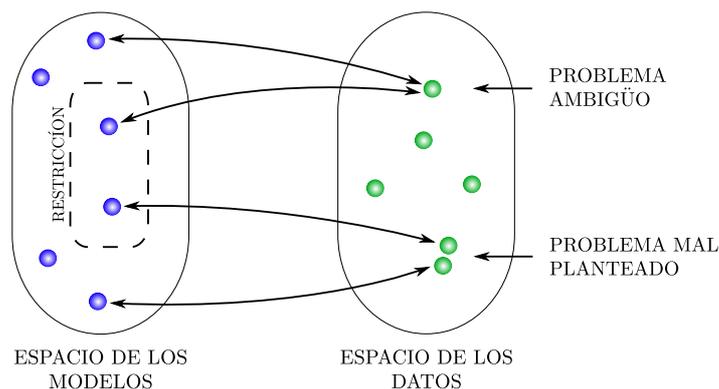
3. Problema inverso: Es el procedimiento matemático de estimar, conocida la física del problema, valores numéricos e información estadística de los parámetros del modelo a partir de datos observados y de información *a priori* independiente de los datos observados.

Desde el punto de vista matemático la resolución del problema inverso se plantea como un problema de optimización, tanto de minimización como de maximización. Con frecuencia la función a ser optimizada es llamada función objetivo, función de costo o función de ajuste. Muchas veces esta función, que nos permitirá evaluar el modelo, es alguna medida de distancia en el espacio de los datos observados. Por ejemplo, si el espacio es euclídeo una medida de distancia utilizada habitualmente es la suma de los cuadrados de la diferencia entre el dato observado y el dato modelado. La teoría del problema inverso permite estimar más que los parámetros del modelo. Puede ser utilizada para estimar la calidad del modelo estimado, para establecer que parámetros, o qué combinación de éstos, están mejor determinados, para determinar que tipo de dato acota mejor el espacio de posibles soluciones, para determinar la influencia del ruido en la estabilidad de las soluciones, o para determinar experimentalmente el tipo y calidad del dato observado necesario para poder determinar los parámetros de un dado modelo. A la hora de resolver el problema inverso se tiene que tener especial consideración en la existencia, unicidad y estabilidad de las soluciones estimadas (Aster et al., 2013). Al resolver un problema inverso la existencia de la solución no está garantizada, y en caso de existir ésta podría no ser única. Mientras el problema directo tiene una solución única, el problema inverso puede tener infinitas soluciones (Figura 2.1). Esto se debe a que tanto la parametrización utilizada como el problema directo con el cual se modelarán los datos sintéticos son aproximaciones de la realidad. Incluso pueden no haber sido correctamente elegidos, no siendo adecuados para representar el verdadero proceso físico que generó los datos observados. Por otro lado dichos datos observados presentan ruido, son finitos e incompletos, es decir que se dispone de una cantidad finita de observaciones imperfectas para reconstruir un modelo que en principio tiene infinitos grados de libertad (Snieder y Trampert, 1999).



**Figura 2.1.** Las soluciones del problema inverso no son únicas. Un conjunto de datos puede ser explicado por infinitos modelos.

Para superar la no unicidad inherente a un problema inverso es necesario hacer explícita todo tipo de información *a priori* que se tenga del modelo para limitar el espacio de posibles soluciones. Esta información *a priori* es incorporada a través de restricciones en el espacio del modelo (Tarantola, 2005). Algunos ejemplos de restricciones sobre el modelo podrían ser exigir que los parámetros estimados tomen sólo valores positivos, o que su suma esté acotada. La elección de una restricción en particular depende del problema a resolver y del tipo de solución buscada. Además de las complicaciones ya analizadas la resolución del problema inverso es un proceso inestable en el que pequeñas variaciones en las observaciones pueden llevar a grandes variaciones en los modelos



**Figura 2.2.** Ejemplo de problema de inversión ambiguo y problema mal planteado. La región delimitada por la línea punteada en el espacio de los modelos representa las soluciones obtenidas por medio de alguna restricción sobre los parámetros del modelo buscado, lo que contribuye a mitigar los problemas mencionados.

estimados. Los problemas de inversión que presentan inestabilidad son conocidos como problemas mal planteados (Hadamard, 1902). La Figura 2.2 ejemplifica el problemas de no unicidad (o ambigüedad) de las soluciones y el problema de inestabilidad que caracteriza a los problemas mal planteados. La mayoría de los problemas de inversión en geofísica entran en la categoría de problemas mal planteados (Menke, 1989). En muchos casos es posible estabilizar la inversión por medio de “regularizaciones”. Las regularizaciones son medidas, en el espacio del modelo, de alguna propiedad deseable sobre el modelo. La incorporación de regularizaciones permite obtener soluciones con significado físico, consistentes con el dato observado y con la información *a priori* disponible.

En general el modelo estimado al resolver el problema de inversión no es exactamente igual al modelo verdadero asociado al dato observado. Esto se debe a que en la mayoría de los problemas de inversión la cantidad de información contenida en el dato observado no es suficiente para determinar correctamente los parámetros del modelo. Como consecuencia la resolución de un problema inverso no sólo consiste en estimar un modelo, sino también realizar una evaluación cualitativa y/o cuantitativa que permita establecer la calidad de las soluciones base a los datos observados y la información *a priori* (Figura 2.3). Esta evaluación nos dará información sobre la relación que guarda el modelo estimado con el modelo verdadero. Es importante determinar qué propiedades del modelo verdadero, y con qué error, han sido recuperadas por el modelo estimado para poder realizar una correcta interpretación física del mismo.

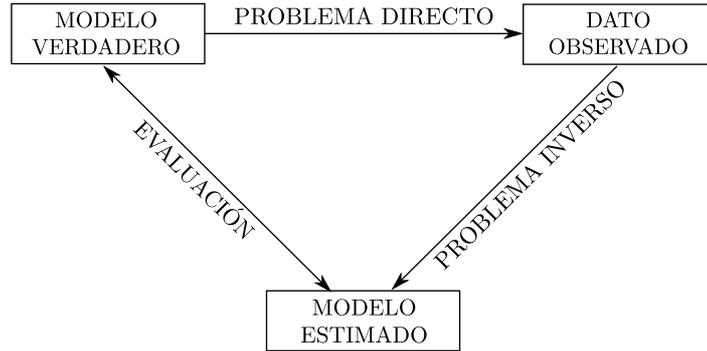
## 2.2. Formulación clásica del problema inverso

Asumiendo que la solución del problema directo es conocida, la relación entre el dato observado y el modelo verdadero se puede expresar como

$$d = g(m) + n, \tag{2.1}$$

## 2. Problema inverso: regularización y optimización

---



**Figura 2.3.** Resolver el problema inverso no solo consiste en estimar el modelo que ajusta al dato, sino también hacer una evaluación del mismo.

donde  $d$  representa al dato observado,  $m$  representa al modelo verdadero,  $g$  representa la relación entre  $d$  y  $m$ , y  $n$  representa el ruido observacional. En la práctica el ruido está siempre presente en las observaciones, en mayor o menor medida dependiendo del ambiente donde se realizaron las observaciones y la sensibilidad de los instrumentos utilizados. De esta forma, al dato observado  $d$  se lo puede expresar como la suma de observaciones sin ruido, dadas por  $g(m)$ , más el ruido  $n$ .

La relación  $g$  entre el modelo y el dato puede tomar muchas formas. Cuando  $d$  y  $m$  son funciones, entonces  $g$  será un operador. Si  $d$  y  $m$  son vectores,  $g$  será una función. El cálculo de  $g(m)$  podría implicar, por ejemplo, resolver un sistema de ecuaciones lineales, un sistema de ecuaciones diferenciales, realizar una integral o una derivación, realizar una multiplicación de matrices o aplicar un algoritmo complejo para el cual no necesariamente  $g$  tiene que tener una forma matemática explícita.

Usualmente sólo el problema directo puede ser resuelto de forma analítica o semi-analítica. Por el contrario, la solución del problema inverso se obtiene comparando el dato observado  $d$  con el dato modelado  $g(m)$ . El objetivo es estimar el modelo  $m$  para el cual  $d$  y  $g(m)$  son similares. Este grado de similitud será cuantificado por medio de alguna medida en el espacio de las observaciones. Esta medida puede ser representada matemáticamente como

$$J(d, m) = \|d - g(m)\|, \quad (2.2)$$

donde  $\|\cdot\|$  es alguna norma matemática que mide la discrepancia entre los datos observados y los datos modelados. Hay muchos tipos de normas, y métodos de inversión basados en diferentes normas que darán soluciones con diferentes características. Pedir que  $d$  y  $g(m)$  sean similares es equivalente a pedir que  $J(d, m)$  sea mínimo. Esto representa una restricción sobre el modelo  $m$ , que involucra tanto al problema directo  $g$  como al dato observado  $d$ .

Muchos otros tipos de restricciones pueden ser incorporadas a la inversión. Estas restricciones, como ya se dijo en la introducción y se ilustró en la Figura 2.2, permiten estabilizar el problema y acotar el espacio de posibles soluciones, disminuyendo la ambigüedad. Las restricciones pueden ser expresadas matemáticamente a través de alguna operación sobre  $m$ , en forma de igualdad:

$$C(m) = 0, \quad (2.3)$$

de desigualdad

$$C(m) \leq 0, \quad (2.4)$$

o de regularización

$$\min_m R(m). \quad (2.5)$$

Las regularizaciones, como fueron descritas anteriormente, son medidas en el espacio del modelo que permiten darle a la solución estimado alguna característica deseada, por ejemplo *smoothness*, *sparseness*, *blockyness*, alguna similitud con algún modelo en particular, etc. Al igual que el  $g$  dado en la ecuación (2.1),  $C$  y  $R$  pueden ser funciones u operadores que dependen de  $m$ , y su evaluación podría implicar realizar una integral, una derivada, un productos de matrices o, como se verá más adelante, el cálculo de cierta norma sobre  $m$ .

El modelo y el dato observado utilizados en la ecuación (2.1) son en general funciones continuas del tiempo y el espacio. Aquellos problemas de inversión en los que el dato y el modelo tienen esta característica son conocidos como problemas de inversión continuos. Sin embargo es común que el dato se encuentre discretizado, es decir, tiene un número fijo y limitado de valores tanto en tiempo como en el espacio. En principio esto no es un inconveniente ya que el teorema de Nyquist establece las condiciones para las cuales el dato puede ser discretizado sin que se pierda información de utilidad. En la práctica usualmente estas condiciones son fáciles de satisfacer para la discretización temporal de las observaciones pero no así para la discretización espacial, ya que muchas veces los lugares donde se desea realizar observaciones no son físicamente accesibles.

En este trabajo de tesis nos concentraremos en problemas de inversión donde tanto el dato como los parámetros del modelo son vectores de dimensión finita. Dichos problemas son llamados problemas de inversión discretos y la relación entre modelo y dato se expresa como

$$\mathbf{d} = g(\mathbf{m}) + \mathbf{n}, \quad (2.6)$$

donde  $\mathbf{d}$ ,  $\mathbf{m}$  y  $\mathbf{n}$  son vectores de dimensión finita representando, respectivamente, valores discretos del dato observado, de los parámetros del modelo y del ruido. Dependiendo de la relación  $g$  entre los parámetros del modelo y el dato, los problemas de inversión se pueden clasificar en lineales y no lineales. Dependiendo de las dimensiones de  $\mathbf{d}$  y  $\mathbf{m}$  los problemas de inversión se pueden clasificar en sobredeterminados o subdeterminados. Ciertos problemas, conocidos como problemas mal planteados, no pertenecen a ninguna de estas dos categorías y serán analizados aparte.

## 2.3. Problemas de inversión lineales

Cuando la relación entre el dato observado y los parámetros del modelo es lineal la ecuación (2.6) se puede expresar como:

$$\mathbf{d} = \mathbf{G}\mathbf{m} + \mathbf{n}, \quad (2.7)$$

donde  $\mathbf{d} = (d_1, d_2, \dots, d_N)^T$  es un vector columna de dimensión  $N$  representando al dato observado,  $\mathbf{m} = (m_1, m_2, \dots, m_M)^T$  es un vector columna de dimensión  $M$  representando

## 2. Problema inverso: regularización y optimización

---

los parámetros del modelo,  $\mathbf{n}$  es un vector de dimensión  $N$  representando al ruido y  $\mathbf{G}$  es una matriz de dimensión  $N \times M$  que relaciona linealmente a  $\mathbf{d}$  y  $\mathbf{m}$ . Dependiendo de las dimensiones de  $\mathbf{d}$  y  $\mathbf{m}$  los problemas de inversión lineales se clasificarán como problemas sobredeterminados si  $N > M$  o problemas subdeterminados si  $N < M$ . Si la solución al problema no existe, no es única o el proceso de inversión no es estable el problema se clasifica como mal planteado. No hay que perder de vista que esta clasificación no se limita a los problemas de inversión lineales, sino también se extiende a los problemas no lineales.

### 2.3.1. Problemas sobredeterminados

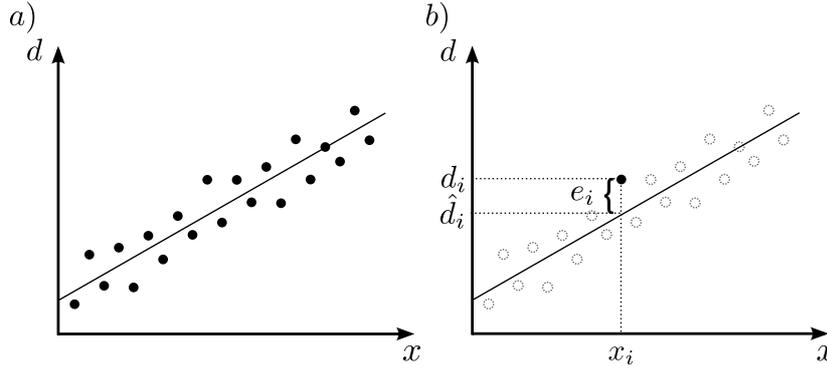
Un problema de inversión está sobredeterminado si las funciones lineales entre el modelo y el dato dadas por la ecuación (2.7) contienen más información de la requerida para resolver el problema exactamente. Se tiene dicha situación cuando la dimensión del vector  $\mathbf{d}$  es mayor que la del vector  $\mathbf{m}$ , es decir  $N > M$ . Un ejemplo sencillo de esta situación sería ajustar una recta utilizando más de dos puntos que no estén perfectamente alineados. Cuando el problema está sobredeterminado no existe una solución exacta de la ecuación (2.7), sin embargo se puede encontrar una solución aproximada que haga mínima alguna medida de la diferencia entre el dato observado y el modelado, es decir que minimice alguna medida del residuo  $\mathbf{e}$  dado por

$$\mathbf{e} = \mathbf{d} - \mathbf{G}\mathbf{m}. \quad (2.8)$$

Una medida muy utilizada para cuantificar esta diferencia es el cuadrado de la norma  $l_2$ , conocido como desajuste o *misfit*. El modelo que minimiza el *misfit* entre el dato observado y el dato modelado es llamado una solución de mínimos cuadrados. Dado el vector  $\mathbf{e}$  de dimensión  $N$ , se define su norma  $l_2$  como

$$\|\mathbf{e}\|_2 = \left( \sum_{i=1}^N |e_i|^2 \right)^{1/2}, \quad (2.9)$$

donde  $e_i = d_i - \hat{d}_i$ , siendo  $d_i$  los elementos de  $\mathbf{d}$  y  $\hat{d}_i$  los elementos del vector del dato modelado  $\mathbf{G}\mathbf{m}$ . La Figura 2.4 ejemplifica el caso en el que se ajusta una recta  $\hat{d} = ax + b$  a un conjunto de observaciones  $d_i$  a partir de un conjunto de abscisas  $x_i$ . En este caso  $M = 2$ , ya que  $\mathbf{m} = (a, b)$ , siendo  $a$  la pendiente y  $b$  la ordenada al origen de la recta con la que se quiere modelar los datos observados.



**Figura 2.4.** a) Ajuste por mínimos cuadrados de un conjunto de observaciones, b) error entre el ajuste y el  $i$ -ésimo dato observado.

Al buscar un solución por mínimos cuadrados el objetivo es encontrar el modelo que haga mínima la función de costo dada por

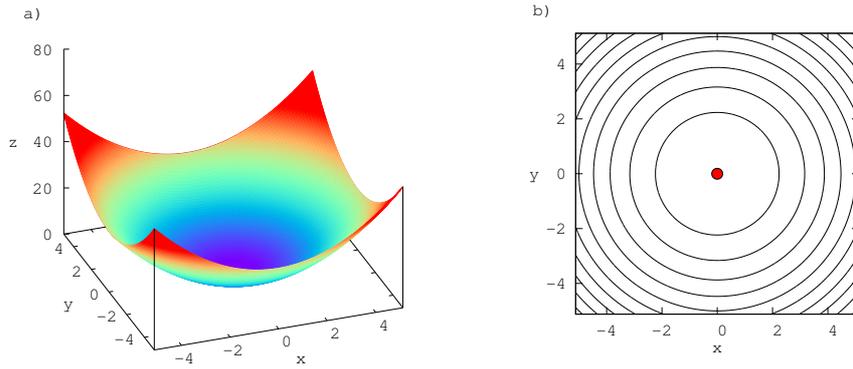
$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 = (\mathbf{d} - \mathbf{G}\mathbf{m})^T(\mathbf{d} - \mathbf{G}\mathbf{m}). \quad (2.10)$$

Notar que esta función de costo es un caso particular de la ecuación (2.2). Al desarrollar se tiene la suma de los elementos  $e_i$  al cuadrado, dando como resultado una función cuadrática cuyas variables son los  $M$  elementos de  $\mathbf{m}$ . La Figura 2.5 es un ejemplo de una función de costo cuadrática de dos variables que podría corresponder al ajuste de una recta ejemplificado en la Figura 2.4. El minimizador de una función cuadrática puede ser estimado analíticamente y es el valor  $\hat{\mathbf{m}}$  que hace cero su derivada respecto  $\mathbf{m}$ . De esta forma la solución por mínimos cuadrados es

$$\hat{\mathbf{m}} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} = \mathbf{B}\mathbf{d}. \quad (2.11)$$

Donde  $\mathbf{B}$  es conocida como la matriz pseudoinversa de  $\mathbf{G}$ . Si  $\mathbf{G}$  fuera una matriz cuadrada e invertible entonces  $\mathbf{B} = \mathbf{G}^{-1}$ . El vector  $\hat{\mathbf{m}}$  representa el modelo estimado asumiendo que  $(\mathbf{G}^T\mathbf{G})^{-1}$  pueda ser calculado.

Las soluciones obtenidas mediante mínimos cuadrados son de especial interés debido a que son fáciles de analizar y de entender desde un punto de vista geométrico, y muy convenientes desde el punto de vista numérico o computacional. Desde un punto de vista estadístico estas soluciones son aquellas más probables si el ruido es aleatorio y sigue una distribución gaussiana, como se verá más adelante en este mismo capítulo. En caso de que el ruido no pueda ser asumido como gaussiano el problema se vuelve más complejo desde el punto de vista computacional. Si bien el método de mínimos cuadrados es simple de aplicar, uno de los inconvenientes es su falta de robustez. El método es muy sensible a observaciones que presenten errores atípicamente grandes. Estas observaciones son llamadas *outliers*. La sola presencia de un *outlier* puede afectar significativamente la solución debido a que cada elemento de la suma implícita en la ecuación (2.10) está elevado al cuadrado. Algunas formas de superar este problema son eliminar los *outliers* utilizando algún criterio para discriminarlos, utilizar el método de mínimos cuadrados pesados (Menke, 1989) para disminuir su influencia, o considerar una distribución diferente a la gaussiana como se verá más adelante en esta tesis.



**Figura 2.5.** a) Función cuadrática de dos variables  $f(x, y) = x^2 + y^2$ , b) curvas de nivel de la función. El punto rojo indica la ubicación del mínimo.

### 2.3.2. Problemas subdeterminados

Un problema de inversión está subdeterminado si las relaciones lineales entre el modelo y el dato dadas por la ecuación (2.7) contienen menos información que la requerida para resolver el problema exactamente. Un ejemplo sencillo sería intentar ajustar una recta utilizando un solo punto como dato. Esta situación se da cuando la dimensión del vector  $\mathbf{d}$  es menor que la del vector  $\mathbf{m}$ , es decir  $N < M$ . En este tipo de problema existen infinitas soluciones que ajustan al dato y no se puede resolver utilizando mínimos cuadrados ya que  $(\mathbf{G}^T \mathbf{G})^{-1}$  no puede calcularse (Aster et al., 2013). Para resolver un problema de inversión subdeterminado se debe incorporar información *a priori* independiente de las observaciones con el fin de reducir el espacio de posibles soluciones, de tal forma que la solución estimada ajuste al dato y a la vez a la información *a priori*. Esta información *a priori* se incorpora al problema mediante la utilización de una regularización (ecuación (2.5)).

Un caso particular de problema inverso subdeterminado, llamado problema puramente subdeterminado, es aquel en el cual las infinitas soluciones ajustan el dato de forma exacta, es decir cuando el término del ruido  $\mathbf{n}$  en la ecuación (2.7) es cero. Para resolver este tipo de problema es común asumir *a priori* suavidad en la solución buscada. La idea detrás de este tipo de regularización es que de todas las soluciones posibles la más representativa es una solución suave. Este tipo de soluciones es conocida como de mínima longitud. En problemas de inversión en los cuales sólo se cuenta con observaciones y no se tiene ninguna clase de conocimiento *a priori* sobre el modelo, este tipo de regularización está justificada debido a que de antemano no se puede asumir que la solución tiene alguna clase de estructura compleja. Una regularización que promueve suavidad en las variaciones de los parámetros del modelo es la función cuadrática

$$R(\mathbf{m}) = \|\mathbf{m}\|_2^2 = \mathbf{m}^T \mathbf{m}. \quad (2.12)$$

Una explicación intuitiva para la elección de esta regularización es que  $R(\mathbf{m})$  se incrementa notablemente si  $\mathbf{m}$  tiene elementos grandes o *outliers*. Estos elementos grandes que aparecen en la solución son necesarios para ajustar al dato observado. Minimizar  $R(\mathbf{m})$  asegura que elementos grandes innecesarios no aparezcan en la solución regularizada. La utilización de esta función cuadrática como regularización nos conduce al

siguiente problema:

$$\min_{\mathbf{m}} \|\mathbf{m}\|_2^2 \quad \text{tal que} \quad \mathbf{G}\mathbf{m} - \mathbf{d} = \mathbf{0}. \quad (2.13)$$

Este problema puede ser resuelto por medio del método de los multiplicadores de Lagrange, dando como resultado:

$$\hat{\mathbf{m}} = \mathbf{G}^T(\mathbf{G}\mathbf{G}^T)^{-1}\mathbf{d}. \quad (2.14)$$

Si el problema no es puramente subdeterminado estimar un modelo que ajuste exactamente al dato observado no es deseable ya que también se estaría ajustando el ruido  $\mathbf{n}$ . En ese caso  $(\mathbf{G}\mathbf{G}^T)^{-1}$  puede ser inestable o no existir y  $R(\mathbf{m})$  tomar valores muy grandes. Para evitar ajustar ruido la solución estimada  $\hat{\mathbf{m}}$  debe ser tal que el *misfit* entre el dato observado  $\mathbf{d}$  y el calculado  $\hat{\mathbf{d}} = \mathbf{G}\hat{\mathbf{m}}$  sea menor o igual al valor esperado del residuo dado por la ecuación (2.8). Este valor, que llamaremos  $\epsilon$ , es en principio desconocido. Este planteo nos conduce al siguiente problema:

$$\min_{\mathbf{m}} \|\mathbf{m}\|_2^2 \quad \text{tal que} \quad \|\mathbf{G}\mathbf{m} - \mathbf{d}\|_2^2 \leq \epsilon. \quad (2.15)$$

Notar que un incremento en el valor de  $\epsilon$  significa que el espacio de los modelos factibles se agranda y que el mínimo valor de  $\|\mathbf{m}\|_2^2$  disminuye. Es decir, a medida que se permite un peor ajuste a los datos, un modelo con menor norma alcanzará para satisfacer este ajuste.

Resolver este problema con restricciones es equivalente a minimizar la función de costo

$$\begin{aligned} J(\mathbf{m}) &= \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \mu\|\mathbf{m}\|_2^2 \\ &= (\mathbf{d} - \mathbf{G}\mathbf{m})^T(\mathbf{d} - \mathbf{G}\mathbf{m}) + \mu\mathbf{m}^T\mathbf{m}. \end{aligned} \quad (2.16)$$

El parámetro  $\mu$  es llamado parámetro de compensación o de *trade-off*. Este parámetro se utiliza para balancear el peso o impacto de ambos términos. Cuando es cero, o cercano a cero, no se estará incorporando información *a priori* a la inversión. Si su valor es demasiado grande se obtendrán soluciones que sólo honren a la información *a priori* pero no al dato observado. Es fundamental elegir un valor del parámetro de compensación  $\mu$  que permita un correcto balance entre los dos términos de la función de costo. Si se conoce el valor esperado del *misfit*, representado en la ecuación (2.15) por  $\epsilon$ , entonces el valor de  $\mu$  óptimo será aquel para el cual el error del ajuste entre la solución estimada y el dato observado tenga valor  $\epsilon$  y  $\|\mathbf{m}\|_2^2$  sea mínimo. A este criterio para determinar el  $\mu$  se lo conoce como principio de discrepancia. Sin embargo como generalmente no se conoce el valor de  $\epsilon$  no existe un método simple para poder estimar cuál es el  $\mu$  óptimo, y muchas veces éste se termina eligiendo por ensayo y error. Algunos métodos utilizados ampliamente para la estimación del parámetro  $\mu$  son el criterio de la curva-L o el *generalized cross-validation criterion* (Farquharson y Oldenburg, 2004; van den Berg y Friedlander, 2008; Hennenfent et al., 2008).

El valor estimado  $\hat{\mathbf{m}}$  que minimiza la función de costo dada por la ecuación (2.16) se obtiene, al igual que para el caso sobredeterminado, derivando respecto de  $\mathbf{m}$  e igualando a cero, obteniéndose como resultado:

$$\hat{\mathbf{m}} = (\mathbf{G}^T\mathbf{G} + \mu\mathbf{I})^{-1}\mathbf{G}^T\mathbf{d}, \quad (2.17)$$

## 2. Problema inverso: regularización y optimización

---

donde  $\mathbf{I}$  es la matriz identidad.

Nótese que cuando  $\mu \rightarrow 0$  la solución  $\hat{\mathbf{m}}$  estimada será similar a la solución por mínimos cuadrados (ecuación (2.11)). Si bien dicha solución minimizará el *misfit* entre el dato observado y el dato modelado, puede ser una solución inestable en el sentido de que alguno de sus valores pueden ser muy grandes, o no tener significado físico. Esto se debe a que algunos de los autovalores de la matriz  $\mathbf{G}^T \mathbf{G}$  podrían ser muy pequeños haciendo que su número de condición sea muy grande o incluso infinito. El número de condición es una medida de la sensibilidad de la solución de un sistema de ecuaciones lineales a los errores presentes en los datos. Se define el número de condición de una matriz cuadrada  $\mathbf{M}$  como (Groß, 2003)

$$\text{cond}(\mathbf{M}) = \|\mathbf{M}\| \|\mathbf{M}^{-1}\|, \quad (2.18)$$

donde  $\|\cdot\|$  es alguna norma matricial. En particular si la norma es  $\|\cdot\|_2$  y  $\mathbf{M}$  es una matriz normal el número de condición se puede expresar como

$$\text{cond}(\mathbf{M}) = \left| \frac{\lambda_{\max}(\mathbf{M})}{\lambda_{\min}(\mathbf{M})} \right|, \quad (2.19)$$

donde  $\lambda_{\max}(\cdot)$  y  $\lambda_{\min}(\cdot)$  representan el máximo y el mínimo autovalor (Hoffman y Kunze, 1973). Si la matriz  $\mathbf{M}$  es rectangular entonces el número de condición se calcula como

$$\text{cond}(\mathbf{M}) = \sqrt{\text{cond}(\mathbf{M}^T \mathbf{M})}. \quad (2.20)$$

Un problema lineal representado por una matriz  $\mathbf{M}$  tal que  $\text{cond}(\mathbf{M}) \simeq 1$  está bien condicionado. Contrariamente si  $\text{cond}(\mathbf{M}) \gg 1$  el problema está mal condicionado y la inversa de  $\mathbf{M}$  no existe o es inestable en el sentido de que alguno de sus elementos puede tomar valores muy grandes. En el problema planteado en la ecuación (2.17) la matriz cuadrada  $\mathbf{G}^T \mathbf{G}$  será la matriz  $\mathbf{M}$  y sumar un valor  $\mu$  a su diagonal logra que el valor de su número de condición disminuya. Esto sucede porque al sumar  $\mu$  a la diagonal de una matriz cuadrada  $\mathbf{M}$  su número de condición pasa a ser

$$\text{cond}(\mathbf{M}) = \left| \frac{\lambda_{\max}(\mathbf{M}) + \mu}{\lambda_{\min}(\mathbf{M}) + \mu} \right|, \quad (2.21)$$

y de esta forma se evita que tienda a infinito incluso si  $\lambda_{\min}(\mathbf{M}) = 0$ . Como consecuencia de esto la solución estimada utilizando la ecuación (2.17) es menos sensible a los errores presentes en el dato que la obtenida por medio de mínimos cuadrados. Esta suma es conocida como amortiguación o “pre-blanqueo”, y al parámetro de compensación  $\mu$  se lo llama parámetro de amortiguación. El modelo estimado con este método es conocido como solución por mínimos cuadrados amortiguados.

### 2.3.3. Problemas mal planteados y mal condicionados

Siguiendo la definición dada por Hadamard (1902) un problema bien planteado es aquel en el cual

1. La solución existe

2. La solución es única
3. La solución del problema depende continuamente de los datos, es decir, hay estabilidad respecto de los mismos.

Un problema mal planteado es aquel que no cumple alguna de las condiciones enumeradas. Este tipo de problemas no son sobredeterminados ni subdeterminados. Muchas veces el término “mal planteado” se utiliza como sinónimo de inestable (Kabanikhin, 2008). Para estos problemas no existe una solución del tipo deseada, no se tiene una única solución, o el proceso de encontrar dicha solución es inestable en el sentido que pequeñas variaciones en el dato observado llevan a grandes variaciones en las soluciones (Figura 2.2).

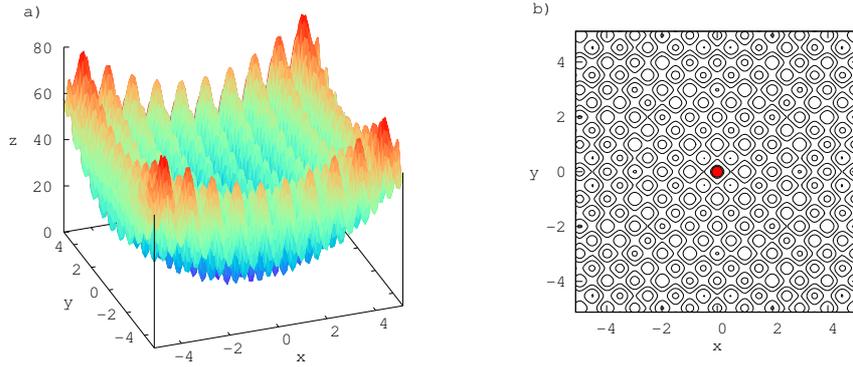
Es importante establecer la diferencia entre un problema mal planteado y un problema mal condicionado. En la sección anterior se analizó brevemente el concepto de número de condición de una matriz, que es una medida de la sensibilidad de las soluciones estimadas ante perturbaciones en los datos observados. Mientras que la definición de problema mal planteado hace referencia a cuestiones propias del problema que se quiere resolver, la definición de problema mal condicionado está relacionada con inestabilidades numéricas debidas a la forma en la cual se resuelve el problema. Cuando se trabaja con problemas lineales discretizados usualmente surge inestabilidad numérica debido a errores de redondeo y muestreo, que llevan a que el problema este mal condicionado. Un problema puede estar bien planteado, pero estar mal condicionado, y viceversa. En la secciones anteriores se analizaron los problemas puramente subdeterminado y subdeterminado, siendo ambos problemas mal planteados ya que en principio tienen infinitas soluciones. Si se resuelve el problema subdeterminado por medio de mínimos cuadrados (ecuación (2.11)) estará mal condicionado a la vez que mal planteado. Sin embargo, si se utiliza mínimos cuadrados amortiguados (ecuación (2.17)) el problema estará bien condicionado si  $\mu$  no es demasiado pequeño.

Como se comentó anteriormente, la mayoría de los problemas de inversión en geofísica entran en la categoría de mal planteados, pero también en la categoría de mal condicionados. La principal dificultad a la hora de resolver un problema mal planteado se debe a las inestabilidades en la solución. En general se cuenta con más observaciones que parámetros del modelo ( $N > M$ ), pero estas observaciones no son linealmente independientes e incompletas, por lo que hay redundancia de información para resolver el problema y éste pasa a ser un problema subdeterminado (Youzwishen, 2001).

Dado un problema mal planteado no existe un método determinado para resolverlo. En cada caso el principal inconveniente, que es la inestabilidad, tiene que ser resuelto de diferente manera. En esta tesis se procederá a la estabilización de los problemas mal planteados incorporando un término de regularización a la función de costo. Dicha función de costo tomará la forma

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{G}\mathbf{m}\| + \mu R(\mathbf{m}), \quad (2.22)$$

donde  $R(\mathbf{m})$  será un término de regularización que deberá ser elegido de acuerdo a las características del problema a resolver y a las características deseadas para las soluciones estimadas, y  $\|\cdot\|$  será una norma adecuada para medir la discrepancia entre el dato observado y el dato calculado. En la sección 2.3.2 se procedió a utilizar el cuadrado de



**Figura 2.6.** a) Función de Rastrigin de dos variables, b) curvas de nivel de la función de Rastrigin, el punto rojo indica la ubicación del mínimo absoluto.

la norma  $l_2$  tanto para medir la discrepancia entre el dato observado y el calculado como para regularizar el problema, es decir  $R(\mathbf{m}) = \|\mathbf{m}\|_2^2$ . Esta regularización dio como resultado una función de costo cuadrática para la cual el mínimo podía ser estimado de manera analítica y de forma fácil. Ésta es sólo una de las posibles regularizaciones utilizadas en los problemas de inversión mal condicionados. Otras regularizaciones llevarán a funciones de costo más complejas que requerirán de algoritmos iterativos para ser optimizadas. Cuando se trate de problemas de inversión lineales las regularizaciones que utilizaremos en esta tesis conducirán a funciones de costo convexas cuyo mínimo podrá ser estimado fácilmente mediante el uso de algoritmos de optimización local.

### 2.4. Problemas de inversión no lineales

Contrariamente a los problemas de inversión lineales, los problemas de inversión no lineales son aquellos en los cuales la relación entre el dato observado y los parámetros del modelo no es lineal. La relación entre  $\mathbf{m}$  y  $\mathbf{d}$  de este tipo de problemas, dada por la ecuación (2.6), ya no podrá expresarse como una matriz. Como ya se comentó, el cálculo de  $g(\mathbf{m})$  podría implicar el uso de un algoritmo complejo para el cual  $g(\mathbf{m})$  no necesariamente tiene que tener una expresión matemática explícita. Al igual que para los problemas lineales, los problemas no lineales se pueden clasificar en sobredeterminados, subdeterminados y mal planteados. Las relaciones no lineales entre el modelo y el dato incrementan la complejidad del problema inverso en comparación con los problemas lineales. La función de costo puede no ser cuadrática y ser multimodal, es decir tener varios mínimos. La Figura 2.6a muestra un ejemplo de función multimodal de dos variables, conocida como función de Rastrigin (Rastrigin, 1974). Esta función posee varios mínimos relativos y un mínimo absoluto como se muestra en la Figura 2.6b. Podría suceder que una función multimodal no tenga mínimo absoluto, sino que todos sean mínimos relativos. La minimización (o maximización) de las funciones multimodales es un proceso que requiere el uso de algoritmos de optimización global.

## 2.5. Algoritmos de optimización

En las secciones anteriores se clasificaron los problemas de inversión en lineales o no lineales dependiendo de la relación entre los datos y los parámetros del modelo. Para algunos casos particulares de problemas lineales se obtuvieron funciones de costo cuadráticas cuyo mínimo puede ser estimado analíticamente, como son los casos de las soluciones obtenidas por mínimos cuadrados (ecuación (2.11)), y mínimos cuadrados amortiguados (ecuación (2.17)). Sin embargo, muchas veces la relación entre el dato y el modelo no es lineal. Esto, sumado a las regularizaciones utilizadas para estabilizar el proceso de inversión, conduce a funciones de costo complejas cuyos mínimos no pueden ser estimados de forma analítica y deben ser estimados utilizando algoritmos de optimización iterativos. Dependiendo de las características de la función a minimizar se utilizan algoritmos de optimización local o algoritmos de optimización global.

### 2.5.1. Algoritmos de optimización local

Los algoritmos de optimización local son algoritmos iterativos que permiten encontrar un mínimo de una función a partir de un valor inicial. Si la función a minimizar es multimodal este tipo de algoritmos no garantiza la localización del mínimo absoluto. Generalmente son utilizados para buscar mínimos relativos o locales, de allí el nombre con el cual se los denomina. Desde el punto de vista del problema de inversión esto no es conveniente, ya que un modelo asociado a un mínimo local de la función de costo no será el modelo que mejor ajuste al dato observado y a las restricciones impuestas. El éxito de este tipo de algoritmos de converger al mínimo absoluto depende de la elección del valor inicial desde el cual se inicia el proceso iterativo y de la función a minimizar. Si el valor inicial está cerca del mínimo absoluto la probabilidad de converger a éste es grande, contrariamente el algoritmo puede converger a un mínimo local, o diverger. Existen una gran cantidad de métodos de optimización local y dependiendo de las características de la función a minimizar es conveniente optar por uno o por otro. Un grupo importante de métodos de optimización local son aquellos basados en el gradiente de la función. Este tipo de algoritmos requiere el cálculo de las derivadas primeras y muchas veces de las derivadas segundas de la función a minimizar. Dependiendo del método las derivadas se podrán obtener de manera analítica o numérica. Una ventaja de los métodos de optimización basados en el gradiente es que, cuando convergen, lo hacen de manera muy eficiente. Ejemplos de métodos de gradiente son el método de máxima pendiente, el método de gradientes conjugados y los métodos de Newton modificados. Una descripción detallada de cada método puede encontrarse en el libro de Scales (1985).

El método de máxima pendiente es el más simple de los métodos de gradiente. En cada iteración, y comenzando desde un punto inicial  $\mathbf{m}_0$ , el algoritmo busca un nuevo punto  $\hat{\mathbf{m}}$  en la dirección contraria a la de máxima pendiente de la función, moviéndose con un paso fijo. Este nuevo valor será el punto de origen de la siguiente iteración. De esta forma los valores estimados serán cada vez más cercanos a aquel que hace mínima la función. Las iteraciones necesarias para obtener un valor aceptable del mínimo dependen de muchos factores, como la complejidad de la función o el paso con el cual se va de un punto a otro. Es importante contar con un criterio de corte adecuado.

## 2. Problema inverso: regularización y optimización

---

Debido a que en cada iteración la dirección de búsqueda depende de las propiedades de la función en ese punto, muchas veces el algoritmo muestra un comportamiento errático, sin desplazarse en la dirección que más rápido lleva al mínimo. Además, si en el entorno del mínimo hacia el cual converge el algoritmo la función no presenta una concavidad marcada se necesitará una gran cantidad de iteraciones para obtener una solución aceptable.

El algoritmo de gradientes conjugados, que en muchos aspectos es similar al de máxima pendiente, es adecuado si se cuenta con las derivadas segundas de la función a minimizar. La principal diferencia con el método de máxima pendiente es que en lugar de buscar puntos en la dirección contraria a la de máxima pendiente, los puntos se buscan en la dirección perpendicular a la adoptada en la iteración anterior. De esta forma se evita el comportamiento errático del método de máxima pendiente y se tiene una convergencia más rápida al mínimo. Otro método que hace uso de las derivadas segundas de la función a minimizar es el método de Newton modificado. Si la función presenta un mínimo bien marcado, ésta se puede aproximar allí por una función cuadrática tal que su matriz Hessiana (matriz de las derivadas segundas) sea definida positiva. El método de Newton modificado hace uso de esta propiedad buscando el mínimo de dicha función cuadrática en lugar del mínimo de la función, de manera más fácil y eficiente. Este método está diseñado para converger rápidamente cuando se conocen analíticamente las derivadas segundas de la función de costo. Sin embargo, el método presenta problemas cuando la matriz Hessiana es indefinida o singular. Para evitar estas complicaciones se realizaron modificaciones al método de Newton en las cuales se utilizan aproximaciones definidas positivas de dicha matriz.

Una relación no lineal entre los datos observados y los parámetros del modelo lleva a funciones de costo complejas. Sin embargo, si la no linealidad de la función no es fuerte en cercanías del mínimo que se quiere estimar es posible utilizar, en el entorno del mismo, una aproximación lineal de la función. Es decir, el problema puede linealizarse en cercanías del mínimo. Esto permite desarrollar algoritmos basados en técnicas como las analizadas en la sección 2.3. Un algoritmo de optimización local que hace uso de esta ventaja es el método de mínimos cuadrados iterativo. En cada iteración el algoritmo aproxima la función de costo por medio de una función cuadrática tangente a algún valor inicial. Resolviendo un problema de mínimos cuadrados amortiguados (ecuación (2.17)) se obtiene una estimación de los parámetros del modelo  $\hat{\mathbf{m}}$ . Este valor estimado, que corresponde al mínimo de la función cuadrática, será utilizado en la siguiente iteración como punto de partida. El proceso iterativo termina cuando el *misfit* entre el dato observado y calculado es menor que un cierto valor preestablecido. Como en los demás algoritmos brevemente descritos, el éxito de convergencia al mínimo absoluto no está garantizado y dependerá de la elección del valor inicial desde el cual iniciar el proceso iterativo. Si la función cuadrática con la cual se inicia el proceso contiene al mínimo absoluto la probabilidad de converger a éste es grande, de otra forma el algoritmo puede converger a un mínimo local o no converger, mostrando un comportamiento oscilatorio.

### 2.5.2. Algoritmos de optimización global

Los algoritmos de optimización analizados en la sección anterior convergen al mínimo más cercano al punto desde cual se inicia el proceso iterativo, sea o no éste el mínimo absoluto. En la mayoría de los problemas de inversión no lineales la función de costo presenta una topología complicada y con varios mínimos. El mínimo absoluto sólo se podrá hallar si el proceso iterativo comienza desde un punto cercano al mínimo absoluto en cuestión. La correcta elección de dicho punto inicial sólo es posible si se cuenta con adecuada información *a priori* sobre la función en cercanías del mínimo absoluto, y esta información no siempre es fácil de obtener. La mayoría de los problemas de inversión en geofísica se encuentran dentro de la categoría de problemas no lineales en los cuales la función de costo tiene varios mínimos. Debido a esto es importante contar con algoritmos con los que sea posible estimar el mínimo absoluto de la función de costo sin que sea crítica la elección del punto desde el cual se comienza la iteración. Dichos algoritmos son conocidos como algoritmos de optimización global. En general, lo que permite a estos algoritmos converger al mínimo absoluto es su capacidad de evitar quedar atrapados en los mínimos locales, y continuar la búsqueda del mínimo absoluto en regiones diferentes del espacio de soluciones. Una desventaja de los algoritmos de optimización global es que estos son muy costosos desde el punto de vista computacional, ya que requieren por lo general una gran cantidad de iteraciones para converger al mínimo absoluto, lo que implica una gran cantidad de evaluaciones de la función de costo. Además el costo computacional se incrementa con el tamaño del espacio de búsqueda, por lo que al utilizar estos algoritmos es importante saber hacer uso de restricciones para reducir el espacio de las soluciones y evitar que el tiempo de búsqueda haga poco práctica su utilización. Algunas ventajas de estos métodos es que, a diferencia de los métodos de optimización local basados en el gradiente, no utilizan derivadas de la función de costo y pueden ser utilizados sobre funciones de costo discontinuas.

Existen una gran variedad de algoritmos de optimización global. Los métodos *Grid Search* y Monte Carlo son dos de las técnicas iterativas más sencillas e intuitivas. El método de *Grid Search* es una técnica de búsqueda exhaustiva en el cual se recorre el espacio de soluciones completo. Por otro lado, en el método Monte Carlo el espacio de soluciones se recorre aleatoriamente, deteniéndose la búsqueda tras un número determinado de iteraciones o cuando se satisface algún criterio de corte previamente establecido. Cuando dicho espacio es grande la búsqueda del mínimo absoluto de la función es muy demandante computacionalmente. A modo de ejemplo, si la parametrización del modelo con el cual se está trabajando tiene dimensión  $M = 10$ , y cada uno de dichos parámetros puede adoptar 10 posibles valores, para hallar el mínimo global utilizando *Grid Search* se requerirán  $10^{10}$  evaluaciones de la función de costo. Debido a esto fue necesario desarrollar otras técnicas en las cuales el espacio de soluciones se recorre de manera más inteligente, evitando calcular innecesariamente valores de la función donde es poco probable que se encuentre el mínimo absoluto. En esta tesis nos concentraremos en el estudio y uso de la familia de algoritmos iterativos conocidos como *Simulated Annealing* (recocido simulado), en particular en el *Very Fast Simulated Annealing* (Ingber, 1989; Velis, 1998). Estos algoritmos tienen su origen en el estudio y simulación del proceso de recocido de metales, una técnica metalúrgica que consiste en calentar y luego enfriar el material de manera controlada para variar sus propiedades

físicas. La evolución de las propiedades de un metal sometido al proceso de recocido fue estudiada por Metropolis et al. (1953) a través de métodos Monte Carlo. Kirkpatrick et al. (1983) generalizaron el concepto y lo aplicaron a problemas de optimización no lineal. Los parámetros desconocidos (el modelo) juegan el papel de las partículas en el metal, y el estado de energía del sistema está representado a través de una función de costo. En el caso del *Simulated Annealing* la convergencia al mínimo absoluto puede garantizarse estadísticamente dado un número suficiente de iteraciones (Velis, 1998). Sin embargo no siempre será práctico alcanzar dicho número de iteraciones por cuestiones de tiempo de cálculo.

Otros algoritmos iterativos de optimización global ampliamente utilizados es la familia de aquellos que se basan en procesos evolutivos (Fogel, 1994), entre los que se encuentran los algoritmos genéticos (*genetic algorithms*) (Davis, 1987) y las otras estrategias evolutivas y de programación evolucionaria (Fogel, 1993). Métodos menos utilizados son el *taboo search* (Cvijović y Klinowski, 1995), las redes neuronales (*neural networks*) (Ramanujan y Sadayappan, 1988), *Particle swarm optimization* (Shaw y Srivastava, 2007), entre otros. La bibliografía sobre métodos de optimización global es extensa y está en constante crecimiento.

Los algoritmos de optimización global, a pesar de ser diseñados para encontrar el mínimo absoluto de la función, pueden quedar atrapados en un mínimo local del cual no pueden escapar. Este inconveniente depende, entre otras cosas, de la complejidad que presente la función a minimizar y del tamaño del espacio del modelo. Para cada uno de los algoritmos nombrados se desarrollaron (y aún se desarrollan) técnicas y mejoras que muchas veces permiten superar este inconveniente sin incrementar demasiado el tiempo de búsqueda. A modo de ejemplo, la técnica de *re-annealing* (Ingber, 1989), utilizada en el método de *Simulated Annealing*, persigue este objetivo. Sin embargo con frecuencia ninguna de estas mejoras garantiza la convergencia al mínimo absoluto en tiempos computacionales adecuados, y la probabilidad de éxito varía dependiendo del problema en el cual se está aplicando el algoritmo de optimización.

### 2.6. Formulación Bayesiana del problema inverso

En las secciones previas se analizó el problema inverso desde el enfoque clásico. En esta sección nos concentraremos en estudiar, brevemente, el enfoque bayesiano. Una de las diferencias fundamentales entre el enfoque clásico y el enfoque bayesiano está en la naturaleza de la solución encontrada. Una característica inherente a los problemas de inversión es la aleatoriedad (Ulrych et al., 2001). La aleatoriedad se encuentra asociada a varias partes del proceso de inversión, por ejemplo el ruido siempre presente en los datos observados es muchas veces aleatorio. En el enfoque clásico se busca estimar un modelo  $\mathbf{m}$  desconocido, pero que se asume como determinado. En el enfoque bayesiano el modelo  $\mathbf{m}$  se lo considera una variable aleatoria y la solución encontrada es la función de distribución de probabilidad asociada al modelo  $\mathbf{m}$  (Aster et al., 2013).

El enfoque bayesiano se basa en asumir que se cuenta con información sobre el modelo a estimar antes de que se observe el dato (Scales y Tenorio, 2001). Esta forma de encarar el problema permite incorporar la información *a priori* de forma más elegante desde el punto de vista matemático. Esta información se expresa matemáticamente a través de una distribución de probabilidad llamada distribución *a priori*. La teoría

desarrollada por Bayes (1763) da el marco para poder combinar la distribución *a priori* con la información contenida en el dato observado. Como resultado de esta combinación se obtiene una distribución de probabilidad llamada distribución *a posteriori*. Esta distribución representa lo que sabemos del modelo, desde el punto de vista probabilístico, después de que realizamos las observaciones, considerando las restricciones impuestas por la distribución *a priori*.

Uno de los principales reparos que se hacen al enfoque bayesiano es el hecho de que a través de la distribución *a priori* se incorpora al modelo estimado características que no están fundamentadas por el dato observado. Sin embargo lo mismo sucede en el enfoque clásico al elegir alguna clase de regularización. Otra cuestión que surge es el hecho de que las probabilidades sólo deberían ser calculadas, no asignadas (Ulrych et al., 2001). En ese sentido la elección de la distribución *a priori* implica darle a la solución una característica que bien podría no ser del todo adecuada, de allí la importancia para el proceso de inversión en realizar un estudio previo que permita una correcta elección de dicha distribución.

### 2.6.1. Teorema de Bayes

La teoría de Bayes establece las bases para calcular probabilidades condicionales, que matemáticamente se pueden expresar como

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d})}. \quad (2.23)$$

La función de densidad de probabilidad *a posteriori*  $p(\mathbf{m}|\mathbf{d})$  es lo que que deseamos conocer. Es la distribución de probabilidad del modelo  $\mathbf{m}$  conocido el dato observado  $\mathbf{d}$ . Esta función nos permite saber cuál es la probabilidad de que un cierto modelo explique al dato observado. La función  $p(\mathbf{d}|\mathbf{m})$  es llamada función de verosimilitud (*likelihood function*). El significado de esta función varía dependiendo del dato observado  $\mathbf{d}$ . Antes de que  $\mathbf{d}$  haya sido observado la función de verosimilitud representa la densidad de probabilidad asociada a posibles realizaciones de  $\mathbf{d}$  a partir un modelo fijo  $\mathbf{m}$ . Una vez que el dato  $\mathbf{d}$  ha sido observado el significado cambia y la función representa la probabilidad de obtener dicho dato  $\mathbf{d}$  como función del modelo  $\mathbf{m}$ . La función  $p(\mathbf{m})$  es la distribución de probabilidad *a priori* del modelo. El denominador  $p(\mathbf{d})$  representa la probabilidad de que el dato haya sido observado y es una “constante” de normalización que garantiza que

$$\int p(\mathbf{m}|\mathbf{d})d\mathbf{m} = 1. \quad (2.24)$$

Esto último debe cumplirse ya que la función *a posteriori* es una función de densidad de probabilidad. Por ello, para asegurar que  $p(\mathbf{m}|\mathbf{d}) \in (0, 1)$  y que se cumpla la ecuación (2.24) se toma

$$p(\mathbf{d}) = \int p(\mathbf{d}|\mathbf{m})p(\mathbf{m})d\mathbf{m}, \quad (2.25)$$

por lo que el cálculo de  $p(\mathbf{d})$  puede ser obviado.

Si el valor numérico que tomará la distribución *a posteriori* no es de interés y sólo se requiere conocer su forma, se puede considerar que

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \quad (2.26)$$

Es importante destacar que la función *a posteriori*  $p(\mathbf{m}|\mathbf{d})$  no da como resultado un único modelo que pueda considerarse como la solución al problema de inversión, sino que da una distribución de posibles modelos. Si se está buscando un único modelo se debe elegir algún criterio para discriminarlo. Dependiendo del tipo de solución que se esté buscando, podría ser adecuado elegir aquel modelo que maximiza la función *a posteriori*. A ese modelo en particular se lo conoce como *maximum a posteriori model* (MAP).

### 2.6.2. Ejemplos

A modo de ejemplo estimaremos el MAP a partir de la relación lineal entre el dato observado y los parámetros del modelo dada por la ecuación (2.7), y considerandos que el ruido  $\mathbf{n} = \mathbf{d} - \mathbf{G}\mathbf{m}$  tiene distribución de probabilidad gaussiana. A partir de estas hipótesis la función de verosimilitud  $p(\mathbf{d}|\mathbf{m})$  será una función de densidad de probabilidad gaussiana multivariada y tendrá la forma (Sivia, 1996)

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{(2\pi)^N |\mathbf{C}_n|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{d} - \mathbf{G}\mathbf{m})^T \mathbf{C}_n^{-1} (\mathbf{d} - \mathbf{G}\mathbf{m}) \right], \quad (2.27)$$

donde  $N$  y  $\mathbf{C}_n$  son el número de muestras del vector  $\mathbf{d}$  y la matriz de covarianzas del vector  $\mathbf{n} = \mathbf{d} - \mathbf{G}\mathbf{m}$ , respectivamente. Tomando como hipótesis adicional que el ruido no está correlacionado entonces la matriz de covarianza del ruido  $\mathbf{n}$  se puede expresar como  $\mathbf{C}_n = \sigma_n^2 \mathbf{I}$ , donde  $\mathbf{I}$  es la matriz identidad y  $\sigma_n^2$  es la varianza.

Con respecto a la función de densidad  $p(\mathbf{m})$  vamos a considerar dos casos: que poseemos información *a priori* suficiente como para asumir con certeza que  $p(\mathbf{m})$  tiene una función de densidad uniforme o que tiene una función de densidad gaussiana.

En el primer caso se tendrá que

$$p(\mathbf{m}) = \mathbb{C}. \quad \forall \mathbf{m}, \quad (2.28)$$

En este caso todos los modelos  $\mathbf{m}$  son igualmente probables.

Remplazando las ecuaciones (2.27) y (2.28) en la ecuación (2.26), y tomando  $\mathbf{C}_n = \sigma_n^2 \mathbf{I}$  se obtiene la función de densidad de probabilidad *a posteriori*

$$p(\mathbf{m}|\mathbf{d}) \propto \frac{\mathbb{C}}{(2\pi)^N \sigma_n} \exp \left[ -\frac{1}{2\sigma_n^2} (\mathbf{d} - \mathbf{G}\mathbf{m})^T (\mathbf{d} - \mathbf{G}\mathbf{m}) \right]. \quad (2.29)$$

El modelo que maximiza la función  $p(\mathbf{m}|\mathbf{d})$  será aquel que minimiza el exponente. Es decir, el modelo  $\mathbf{m}$  más probable es aquel que minimiza la función de costo dada por

$$J(\mathbf{m}) = (\mathbf{d} - \mathbf{G}\mathbf{m})^T (\mathbf{d} - \mathbf{G}\mathbf{m}) = \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2,$$

que es igual a la ecuación (2.10).

Como ya se analizó en la sección 2.3.2, la estimación del modelo  $\mathbf{m}$  que minimiza esta función de costo es

$$\hat{\mathbf{m}} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d}. \quad (2.30)$$

De este resultado se deduce que la solución estimada utilizando el método de mínimos cuadrados tiene significado si se cumplen las hipótesis de que los errores en los datos que queremos ajustar linealmente siguen una distribución aproximadamente gaussiana y que todos los parámetros del modelo son igualmente probables (ecuación (2.28)). Esta solución es la más conservadora en el sentido de que no estamos haciendo ninguna suposición *a priori* especial acerca del modelo que queremos estimar. Todos los valores de  $\mathbf{m}$  son igualmente probables y la solución sólo está basada en el dato observado.

En el segundo caso, asumiendo que los parámetros  $\mathbf{m}$  siguen una distribución gaussiana y no están correlacionados, la función de densidad *a priori* tendrá la forma

$$p(\mathbf{m}) \propto \frac{1}{(2\pi)^M \sigma_m} \exp \left[ -\frac{1}{2\sigma_m^2} \mathbf{m}^T \mathbf{m} \right]. \quad (2.31)$$

donde  $M$  y  $\sigma_m^2$  son el número de elementos y la varianza del vector  $\mathbf{m}$ , respectivamente.

Siguiendo el mismo procedimiento llevado a cabo en el primer caso se obtiene que la función de densidad de probabilidad *a posteriori* tiene la siguiente forma

$$p(\mathbf{m}|\mathbf{d}) \propto \frac{1}{(2\pi)^{N+M} \sigma_n \sigma_m} \exp \left[ -\frac{1}{2\sigma_n^2} (\mathbf{d} - \mathbf{G}\mathbf{m})^T (\mathbf{d} - \mathbf{G}\mathbf{m}) - \frac{1}{2\sigma_m^2} \mathbf{m}^T \mathbf{m} \right]. \quad (2.32)$$

Por lo tanto, el modelo más probable es aquel que minimiza la función de costo dada por

$$\begin{aligned} J(\mathbf{m}) &= (\mathbf{d} - \mathbf{G}\mathbf{m})^T (\mathbf{d} - \mathbf{G}\mathbf{m}) + \frac{\sigma_n^2}{\sigma_m^2} \mathbf{m}^T \mathbf{m} \\ &= \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \frac{\sigma_n^2}{\sigma_m^2} \|\mathbf{m}\|_2^2, \end{aligned} \quad (2.33)$$

que es igual a la ecuación (2.16) con  $\mu = \sigma_n^2/\sigma_m^2$ . Como ya se analizó en la sección 2.3.2, la estimación del modelo  $\mathbf{m}$  que minimiza esta función de costo es

$$\hat{\mathbf{m}} = (\mathbf{G}^T \mathbf{G} + \mu \mathbf{I})^{-1} \mathbf{G}^T \mathbf{d}. \quad (2.34)$$

De este resultado se puede observar la relación entre el parámetro  $\mu$  y el dato. Si se cuenta con datos muy precisos y confiables su incertidumbre, representada por  $\sigma_n^2$ , es mucho menor que la incertidumbre del modelo  $\sigma_m^2$ , luego el segundo término de la ecuación (2.33) tendrá poco peso en la solución y ésta estará basada principalmente en el dato observado. En caso contrario, si las observaciones no son confiables, se le dará mayor peso al término de regularización y la solución estará fuertemente influenciada por este. Sin embargo la incerteza  $\sigma_m^2$  generalmente no es conocida por ello el parámetro  $\mu$  debe ser estimado por otros medios.

### 2.7. Sumario

En este capítulo se hizo una muy breve revisión de la teoría de inversión. La teoría de inversión da las herramientas para poder estimar modelos matemáticos a partir de observaciones y poder hacer un análisis de las soluciones estimadas. Una de las desventajas al resolver problemas de inversión es que éstos sufren de inestabilidad y ambigüedad. Para poder obtener resultados útiles y con significado físico dichas complicaciones deben ser sorteadas. En este capítulo se analizaron las formulaciones clásica y bayesiana del problema inverso. En la primera se clasificaron los problemas en lineales y no lineales, siendo a la vez clasificados en sobredeterminados y subdeterminados. Se vio que la solución por mínimos cuadrados es adecuada para resolver problemas inversos lineales sobredeterminados, mientras que la solución por mínimos cuadrados amortiguados es adecuada para resolver problemas lineales subdeterminados. Estos resultados fueron justificados posteriormente por medio de la formulación bayesiana, donde además se hizo evidente la cercana relación entre el método de mínimos cuadrados y la función de densidad de probabilidad gaussiana. Se hizo especial mención a problemas mal condicionados, ya que gran parte de los problemas de inversión en geofísica pertenecen a esta categoría. Se enunciaron las condiciones que definen un problema bien planteado, y a partir de esta definición se dieron las condiciones para determinar si un problema está bien o mal condicionado. Para los problemas mal condicionados se discutió su regularización como un medio de poder estabilizar el proceso de inversión. Este enfoque será utilizado en esta tesis para resolver diversos problemas de inversión. Se analizó la complejidad que surge de una relación no lineal entre el dato y los parámetros del modelo. A partir de esto se realizó un breve resumen de los métodos de optimización local y global, haciendo una descripción de aquellos más conocidos y utilizados. En esta tesis se estará trabajando con problemas de inversión lineales y no lineales, y las funciones de costo asociadas serán optimizadas utilizando métodos de optimización local y global dependiendo del caso. Es importante destacar que, más allá de la clasificación hecha en este capítulo, cada problema de inversión es único y se debe determinar la forma más efectiva de resolverlo.

# Capítulo 3

## Problema directo: de Zoeppritz a los datos sísmicos prestack

### 3.1. Introducción

En los métodos sísmicos de prospección la mayor parte de la información de interés es proporcionada por las ondas que se propagan por el subsuelo. Dichas ondas se separan en ondas compresionales, conocidas como ondas  $P$ , y en ondas de corte, conocidas como ondas  $S$ . Las ondas compresionales, como su nombre lo indica, se caracterizan por producir la compresión y dilatación del medio a medida que se propagan. El movimiento de las partículas que forman el medio es en la dirección de la propagación de la onda, por eso este tipo de onda también se conoce con el nombre de onda longitudinal. Por otro lado, las ondas de corte deben su nombre a que el movimiento de las partículas del medio es perpendicular a la dirección de propagación de las ondas. Una característica importante de este tipo de ondas es que no se propagan por los fluidos ya que éstos no soportan tensión de corte. Debido a la sensibilidad que muestran las ondas  $S$  frente a los fluidos presentes en las rocas es importante entender cual es su contribución en la información proporcionada por el método sísmico cuando se lo usa para caracterizar, por ejemplo, un reservorio. La velocidad de propagación de las ondas  $P$  varía dependiendo de una gran cantidad de parámetros físicos del medio por el cual se propaga, tal como la densidad, sus parámetros elásticos (modulo de Young, la constante de Lamé, etc), la cantidad de fluidos o la porosidad. Debido a esto, estudiar solamente las propiedades de la onda  $P$  suele no ser suficiente para poder caracterizar correctamente el medio, por lo que es necesario entender como las ondas  $P$  se relacionan con las ondas  $S$  para complementar la información provista.

Al utilizar el método sísmico de prospección se considera que la parte superior de la corteza terrestre está compuesta de capas rocosas con diferentes propiedades elásticas. A pesar de ser este un modelo muy sencillo del subsuelo es de gran utilidad al resolver problemas de inversión sísmica como los que se tratarán en los próximos capítulos. Cuando una onda sísmica se propaga hacia el interior de la Tierra su energía se divide a medida que incide en las discontinuidades que separan dichas capas. Si la onda que incide sobre una discontinuidad que separa dos medios es una onda compresional parte de su energía se convierte en una onda de corte, luego la onda compresional incidente y la onda de corte convertida son parcialmente reflejadas desde la discontinuidad y

parcialmente transmitidas a través de ésta.

En la década de 1960 se observaron, en los datos sísmicos *poststack*, incrementos anómalos de los valores de los coeficientes de reflexión relacionadas con depósitos de gas. Estas anomalías de amplitud, que son llamadas *bright spots* (Sheriff, 2002), podían también deberse a cambios bruscos en la litología de la zona explorada, por lo que en las siguiente década se desarrollaron, con variable éxito, técnicas y métodos basados en el análisis de datos *poststack* con el fin de resolver dicha ambigüedad. Finalmente Ostrander (1984) introdujo una nueva clase de atributos sísmicos por medio del análisis de AVA, ya que éste no sólo depende del ángulo de incidencia sino también de las propiedades físicas del medio por el cual se propagan las ondas, en particular del contenido de fluidos de las rocas. De esta forma el AVA mostró ser un método capaz de proveer valiosa información para la caracterización del subsuelo y la prospección de hidrocarburos. De esta variación se puede obtener estimaciones de parámetros físicos tales como la velocidad de propagación de las ondas  $P$ , la velocidad de propagación de las ondas  $S$ , la densidad, el coeficiente de Poisson, entre otros.

La variación de la amplitud de las ondas reflejadas y transmitidas en función del ángulo de incidencia ha sido estudiada desde finales del siglo XIX, siendo el trabajo de Knott (1899) uno de los primeros sobre el tema. A partir de este trabajo Zoeppritz (1919) desarrolló las ecuaciones que llevan su nombre y modelan dicha variación. En estas ecuaciones la relación entre los parámetros elásticos y el ángulo de incidencia es no lineal y complicada, por lo que a partir de la década de 1960 fueron desarrolladas varias aproximaciones lineales. Los coeficientes de estas aproximaciones constituyen atributos que pueden proveer importante información acerca del contenido de fluidos en las rocas, información que es clave para la caracterización de reservorios de hidrocarburos (Ostrander, 1984; Castagna et al., 1998; Smith y Gidlow, 2000; Chopra y Castagna, 2007). En este capítulo se deducirán las ecuaciones de Zoeppritz para una onda incidente compresional y también se hará un resumen de varias de las aproximaciones mas conocidas, estableciendo las hipótesis bajo las cuáles se derivaron. Se pondrá particular énfasis en detallar las aproximaciones de Aki y Richards (1980) y Shuey (1985), haciendo una comparación entre ambas, ya que serán ampliamente utilizadas en los próximos capítulos de esta tesis. De estas aproximaciones se obtienen expresiones que relacionan diversas magnitudes físicas que caracterizan al subsuelo con el ángulo de incidencia.

## 3.2. Las ecuaciones de Zoeppritz

Dada una onda plana que incide sobre una discontinuidad que separa dos medios con diferentes propiedades elásticas las ecuaciones de Zoeppritz (Zoeppritz, 1919; Yilmaz, 2001; Ikelle y Amundsen, 2005) permiten predecir las amplitudes de las ondas reflejadas y transmitidas (tanto onda  $P$  como  $S$ ) en función del ángulo con el cual la onda incidió. Las ecuaciones de Zoeppritz se derivan bajo la hipótesis de que los medios son semiespacios elásticos en condición de soldadura. Estos semiespacios están caracterizados por la densidad  $\rho$  y las velocidades  $\alpha$  y  $\beta$  de propagación de ondas compresionales y de corte, respectivamente. Para la derivación de las ecuaciones de Zoeppritz no es condición necesaria que los medios sean homogéneos e isótropos, sin embargo dichas características serán asumidas como verdaderas en esta tesis debido a que sí son necesarias para derivar las aproximaciones lineales que serán utilizadas en la

resolución de los problemas de inversión sísmica que serán estudiados en los siguientes capítulos. También se considera como hipótesis que el ángulo de incidencia es menor que el ángulo de incidencia crítica. Dicho ángulo crítico queda determinado por las velocidades de propagación de las ondas sísmicas en ambos medios y puede ser estimado a través de la Ley de Snell (Shearer, 1999).

Dados dos medios con las propiedades descritas y una onda plana, compresional o de corte, que se propaga por el primer medio y que incide en la discontinuidad que lo separa del segundo, las ecuaciones de Zoeppritz predicen que se tendrá una onda compresional y otra de corte reflejadas hacia el primer medio y una onda compresional y otra de corte transmitidas hacia el segundo (Figura 3.1). De la misma forma, cuando la onda incidente viaja inicialmente por el segundo medio e incide en la discontinuidad se tendrán otras dos ondas reflejadas y dos ondas transmitidas. De este análisis se observa que en general las ecuaciones de Zoeppritz permiten estimar, en función del ángulo de incidencia, la amplitud de 16 tipos diferentes de ondas dependiendo de la condición inicial de la onda incidente (Aki y Richards, 1980). En esta tesis nos limitaremos al caso en el que tanto la onda incidente como la reflejada son ondas  $P$ . Ondas transmitidas y ondas convertidas de  $P$  a  $S$  no serán estudio de este trabajo. La razón de esta limitación es que en exploración sísmica mayormente se generan y se registran sólo ondas compresionales. Sin embargo, con el desarrollo de nuevas tecnologías en el campo de la adquisición y el procesamiento de datos, la sísmica multicomponente está siendo utilizada cada vez más (Hampson et al., 2005; Hampson y Russell, 2013).

### 3.2.1. Deducción de las ecuaciones de Zoeppritz

En esta sección se deducirán las ecuaciones de Zoeppritz que permiten estimar los coeficientes de reflexión y transmisión para ondas planas en función del ángulo de incidencia y de seis parámetros físicos independientes, tres a cada lado de la discontinuidad que separa los dos medios. Las ecuaciones de Zoeppritz se obtienen imponiendo las siguientes condiciones de borde sobre los desplazamientos y los esfuerzos producidos por el campo de ondas en la discontinuidad que separa estos medios (Yilmaz, 2001; Ikelle y Amundsen, 2005):

1. Continuidad de la componente tangencial de los desplazamientos.
2. Continuidad de la componente normal de los desplazamientos.
3. Continuidad de la componente normal de los esfuerzos.
4. Continuidad de la componente tangencial de los esfuerzos.

A partir de estas condiciones de borde se obtendrán cuatro ecuaciones que relacionan las amplitudes de las ondas reflejadas y transmitidas, tanto compresionales como de corte, con el ángulo de incidencia de la onda compresional.

Por medio del teorema de Helmholtz (Lay y Wallace, 1995), el campo de desplazamiento  $\mathbf{u}$  producido por la propagación de las ondas a través del medio puede ser representado como

$$\mathbf{u} = \nabla\Phi + \nabla \times \Psi, \tag{3.1}$$

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack

donde  $\Phi$  es una función potencial escalar asociada a las ondas compresionales  $P$  y  $\Psi$  es una función potencial vectorial asociada a las ondas de corte  $S$ , tal que

$$\nabla \times \Phi = 0, \quad (3.2)$$

$$\nabla \cdot \Psi = 0. \quad (3.3)$$

Para un medio bidimensional de coordenadas cartesianas  $(x, z)$  dichos potenciales se relacionan con el campo de desplazamiento  $\mathbf{u} = u\hat{x} + w\hat{z}$ , donde  $\hat{x}$  y  $\hat{z}$  son los versores asociados a las coordenadas  $x$  y  $z$  respectivamente, a través de

$$u = \frac{\partial \Phi}{\partial x} + \frac{\partial \Psi}{\partial z}, \quad (3.4)$$

$$w = \frac{\partial \Phi}{\partial z} - \frac{\partial \Psi}{\partial x}, \quad (3.5)$$

donde  $\Psi$  es la componente de  $\Psi$  perpendicular al plano  $(x, z)$ . Dichos potenciales satisfacen las ecuaciones de onda

$$\frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Phi}{\partial z^2} = \frac{1}{\alpha^2} \frac{\partial^2 \Phi}{\partial t^2}, \quad (3.6)$$

y

$$\frac{\partial^2 \Psi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial z^2} = \frac{1}{\beta^2} \frac{\partial^2 \Psi}{\partial t^2}. \quad (3.7)$$

Dada una onda compresional  $P$  de amplitud  $A_0$  que incide con ángulo  $\theta_I$  en la discontinuidad que separa dos medios (Figura 3.1), el medio 1 con velocidades y densidad  $\alpha_1$ ,  $\beta_1$  y  $\rho_1$  y el medio 2 con velocidades y densidad  $\alpha_2$ ,  $\beta_2$  y  $\rho_2$ , los potenciales  $\Phi_1$ ,  $\Phi_2$ ,  $\Psi_1$  y  $\Psi_2$  de las ondas  $P$  y  $S$  reflejadas y transmitidas que satisfacen las ecuaciones (3.6) y (3.7) se expresan como

$$\Phi_1 = A_0 \exp \left( i \frac{\omega}{\alpha_1} \sin \theta_I x + i \frac{\omega}{\alpha_1} \cos \theta_I z - i \omega t \right) \quad (3.8)$$

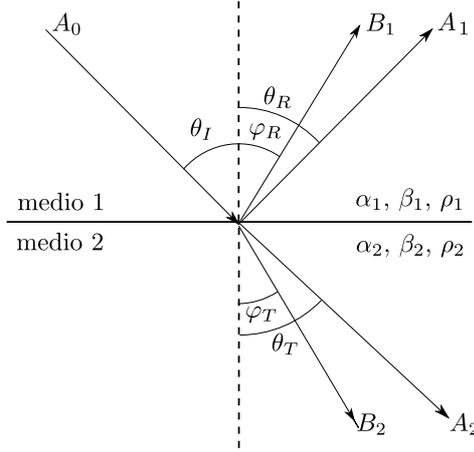
$$+ A_1 \exp \left( i \frac{\omega}{\alpha_1} \sin \theta_R x - i \frac{\omega}{\alpha_1} \cos \theta_R z - i \omega t \right), \quad (3.9)$$

$$\Psi_1 = B_1 \exp \left( i \frac{\omega}{\beta_1} \sin \varphi_R x - i \frac{\omega}{\beta_1} \cos \varphi_R z - i \omega t \right), \quad (3.10)$$

$$\Phi_2 = A_2 \exp \left( i \frac{\omega}{\alpha_2} \sin \theta_T x + i \frac{\omega}{\alpha_2} \cos \theta_T z - i \omega t \right), \quad (3.11)$$

$$\Psi_2 = B_2 \exp \left( i \frac{\omega}{\beta_2} \sin \varphi_T x + i \frac{\omega}{\beta_2} \cos \varphi_T z - i \omega t \right), \quad (3.12)$$

donde  $\omega$  es la frecuencia angular y  $A_1$ ,  $A_2$ ,  $B_1$  y  $B_2$  son, respectivamente, las amplitudes de las ondas  $P$  reflejada y transmitida y  $S$  reflejada y transmitida. La coordenada  $z$  se toma positiva hacia abajo, siendo  $z = 0$  en la discontinuidad. Los ángulos  $\theta_R$ ,  $\theta_T$ ,  $\varphi_R$ ,  $\varphi_T$  son respectivamente los ángulos de reflexión y transmisión de la onda  $P$  y los ángulos de reflexión y transmisión de la onda  $S$ .



**Figura 3.1.** Ondas compresionales y de corte, reflejadas y transmitidas, a partir de una onda compresional incidente en la discontinuidad que separa dos medios con propiedades elásticas diferentes.

El objetivo es calcular las amplitudes  $A_1$ ,  $A_2$ ,  $B_1$  y  $B_2$  de las ondas reflejadas y transmitidas. Para ello se hará uso de las expresiones de los potenciales dados en las ecuaciones (3.8) a (3.12), de las condiciones de borde antes enumeradas y de la ley de Snell

$$\frac{\sin \theta_I}{\alpha_1} = \frac{\sin \theta_R}{\alpha_1} = \frac{\sin \theta_T}{\alpha_2} = \frac{\sin \varphi_R}{\beta_1} = \frac{\sin \varphi_T}{\beta_2}. \quad (3.13)$$

**Primera ecuación de Zoeppritz:** Se observa la continuidad de la componente tangencial de desplazamientos a través de la discontinuidad que separa ambos medios.

Esta condición de borde se puede expresar como

$$u_1(z = 0) = u_2(z = 0). \quad (3.14)$$

Remplazando la ecuación (3.4) se obtiene

$$\left( \frac{\partial \Phi_1}{\partial x} + \frac{\partial \Psi_1}{\partial z} \right) \Big|_{z=0} = \left( \frac{\partial \Phi_2}{\partial x} + \frac{\partial \Psi_2}{\partial z} \right) \Big|_{z=0}, \quad (3.15)$$

luego, reemplazando con las funciones potenciales dadas en las ecuaciones (3.8) a (3.12) y utilizando la Ley de Snell se obtiene

$$\frac{\sin \theta_R}{\alpha_1} A_0 + \frac{\sin \theta_R}{\alpha_1} A_1 - \frac{\cos \varphi_R}{\beta_1} B_1 = \frac{\sin \theta_T}{\alpha_2} A_2 + \frac{\cos \varphi_T}{\beta_2} B_2, \quad (3.16)$$

que es la primera ecuación de Zoeppritz.

**Segunda ecuación de Zoeppritz:** Se observa la continuidad de la componente normal de los desplazamientos a través de la discontinuidad que separa ambos medios.

Esta condición de borde se puede expresar como

$$w_1(z = 0) = w_2(z = 0). \quad (3.17)$$

Remplazando la ecuación (3.5) se obtiene

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack

---

$$\left( \frac{\partial \Phi_1}{\partial z} - \frac{\partial \Psi_1}{\partial x} \right) \Big|_{z=0} = \left( \frac{\partial \Phi_2}{\partial z} - \frac{\partial \Psi_2}{\partial x} \right) \Big|_{z=0}, \quad (3.18)$$

luego, reemplazando con las funciones potenciales dadas en las ecuaciones (3.8) a (3.12) y utilizando la Ley de Snell se obtiene

$$\frac{\cos \theta_R}{\alpha_1} A_0 - \frac{\cos \theta_R}{\alpha_1} A_1 - \frac{\sin \varphi_R}{\beta_1} B_1 = \frac{\cos \theta_T}{\alpha_2} A_2 - \frac{\sin \varphi_T}{\beta_2} B_2, \quad (3.19)$$

que es la segunda ecuación de Zoeppritz.

Para poder aplicar las condiciones de borde restantes se deben expresar los esfuerzos normales y tangenciales en función de los potenciales  $\Phi$  y  $\Psi$ . Para ello vamos a utilizar la relación entre el esfuerzo y deformación dada por la Ley de Hook (Yilmaz, 2001)

$$\sigma_{ij} = \lambda \Delta \delta_{ij} + 2\mu e_{ij}, \quad i, j = x, y, z \quad (3.20)$$

donde  $\sigma_{i,j}$  es el esfuerzo,  $\delta_{i,j}$  es el delta de Kronecker,  $e_{i,j}$  es la deformación y  $\Delta$  representa el cambio relativo en volumen. Las constantes  $\lambda$  y  $\mu$  son los parámetros de Lamé, que se relacionan con las velocidades y densidades del medio a través de

$$\alpha = \sqrt{\frac{\lambda + 2\mu}{\rho}}, \quad \beta = \sqrt{\frac{\mu}{\rho}}. \quad (3.21)$$

Cuando se trabaja en volúmenes infinitesimales  $\Delta$  es igual a la suma de los tres componentes principales de la deformación:  $\Delta = e_{xx} + e_{yy} + e_{zz}$ . En este caso, al estar deduciendo las ecuaciones de Zoeppritz en el plano  $(x, z)$  se tiene que  $e_{yy} = 0$ . De esta forma la componente normal del esfuerzo puede expresarse, usando la ecuación (3.20) como

$$\sigma_{zz} = \lambda(e_{xx} + e_{zz}) + 2\mu e_{zz}, \quad (3.22)$$

donde

$$e_{xx} = \frac{\partial u}{\partial x}, \quad e_{zz} = \frac{\partial w}{\partial z}. \quad (3.23)$$

Utilizando las ecuaciones (3.4), (3.5), (3.21), (3.22) y (3.23) se puede expresar el esfuerzo normal como

$$\sigma_{zz} = \rho \frac{\partial^2 \Phi}{\partial t^2} - 2\rho\beta^2 \left( \frac{\partial^2 \Phi}{\partial x^2} + \frac{\partial^2 \Psi}{\partial z \partial x} \right). \quad (3.24)$$

Teniendo ya expresado el esfuerzo normal en función de las propiedades elásticas del medio y de los potenciales se puede hacer uso de la tercera condición de borde.

**Tercera ecuación de Zoeppritz:** Se observa la continuidad de la componente normal de los esfuerzos a través de la discontinuidad que separa ambos medios.

Esta condición de borde se puede expresar como

$$\sigma_{zz,1}(z = 0) = \sigma_{zz,2}(z = 0). \quad (3.25)$$

Reemplazando con la ecuación (3.24) se obtiene

$$\rho_1 \frac{\partial^2 \Phi_1}{\partial t^2} - 2\rho_1 \beta_1^2 \left( \frac{\partial^2 \Phi_1}{\partial x^2} + \frac{\partial^2 \Psi_1}{\partial z \partial x} \right) \Big|_{z=0} = \rho_2 \frac{\partial^2 \Phi_2}{\partial t^2} - 2\rho_2 \beta_2^2 \left( \frac{\partial^2 \Phi_2}{\partial x^2} + \frac{\partial^2 \Psi_2}{\partial z \partial x} \right) \Big|_{z=0}, \quad (3.26)$$

luego, reemplazando con las funciones potenciales dadas en las ecuaciones (3.8) a (3.12) y utilizando la Ley de Snell se obtiene

$$-\cos 2\varphi_R A_0 - \cos 2\varphi_R A_1 - \sin 2\varphi_R B_1 = -\frac{\rho_2}{\rho_1} \cos 2\varphi_T A_2 + \frac{\rho_2}{\rho_1} \sin 2\varphi_T B_2, \quad (3.27)$$

que es la tercera ecuación de Zoeppritz.

Siguiendo la ecuación (3.20), cuando se trabaja en el plano  $(x, z)$ , el esfuerzo tangencial puede expresarse como

$$\sigma_{xz} = 2\mu e_{xz}, \quad (3.28)$$

donde

$$e_{xz} = \frac{1}{2} \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right), \quad (3.29)$$

Utilizando las ecuaciones (3.4), (3.5), (3.21), (3.28) y (3.29) se puede expresar el esfuerzo tangencial como

$$\sigma_{xz} = \beta^2 \rho \left( 2 \frac{\partial^2 \Phi}{\partial z \partial x} + \frac{\partial^2 \Psi}{\partial z^2} - \frac{\partial^2 \Psi}{\partial x^2} \right) \quad (3.30)$$

Teniendo ya expresado el esfuerzo tangencial en función de las propiedades elásticas del medio y de los potenciales se puede hacer uso de la cuarta condición de borde.

**Cuarta ecuación de Zoeppritz:** Se observa la continuidad de la componente tangencial de los esfuerzos a través de la discontinuidad que separa ambos medios.

Esta condición de borde se puede expresar como

$$\sigma_{xz,1}(z=0) = \sigma_{xz,2}(z=0). \quad (3.31)$$

Reemplazando con la ecuación (3.30) se obtiene

$$\beta_1^2 \rho_1 \left( 2 \frac{\partial^2 \Phi_1}{\partial z \partial x} + \frac{\partial^2 \Psi_1}{\partial z^2} - \frac{\partial^2 \Psi_1}{\partial x^2} \right) \Big|_{z=0} = \beta_2^2 \rho_2 \left( 2 \frac{\partial^2 \Phi_2}{\partial z \partial x} + \frac{\partial^2 \Psi_2}{\partial z^2} - \frac{\partial^2 \Psi_2}{\partial x^2} \right) \Big|_{z=0}, \quad (3.32)$$

luego, reemplazando con las funciones potenciales dadas en las ecuaciones (3.8) a (3.12) y utilizando la Ley de Snell se obtiene

$$-\sin 2\theta_R A_0 + \sin 2\theta_R A_1 - \frac{\alpha_1^2}{\beta_1^2} \cos 2\varphi_R B_1 = -\frac{\rho_2 \beta_2^2 \alpha_1^2}{\rho_1 \beta_1^2 \alpha_2^2} \sin 2\theta_T A_2 - \frac{\rho_2 \alpha_1^2}{\rho_1 \beta_1^2} \cos 2\varphi_T B_2, \quad (3.33)$$

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack

que es la cuarta ecuación de Zoeppritz.

Las ecuaciones (3.16), (3.19), (3.27) y (3.33) constituyen un sistema de ecuaciones que puede ser resuelto para estimar las amplitudes  $A_1$ ,  $A_2$ ,  $B_1$  y  $B_2$  de los potenciales. Sin embargo es conveniente expresar estas ecuaciones en función de los coeficientes de reflexión y transmisión

$$R_{PP} = \frac{A_1}{A_0}, \quad R_{PS} = \frac{\alpha_1 B_1}{\beta_1 A_0}, \quad T_{PP} = \frac{\alpha_1 A_2}{\alpha_2 A_0}, \quad T_{PS} = \frac{\alpha_1 B_2}{\beta_2 A_0}. \quad (3.34)$$

Reescribiendo las ecuaciones de Zoeppritz en función de los coeficientes de reflexión y transmisión, y expresando el sistema en forma matricial se obtiene

$$\begin{pmatrix} \cos \theta_R & \sin \varphi_R & \cos \theta_T & -\sin \varphi_T \\ -\sin \theta_R & \cos \varphi_R & \sin \theta_T & \cos \varphi_T \\ -\cos 2\varphi_R & -\frac{\beta_1}{\alpha_1} \sin 2\varphi_R & \frac{\alpha_2 \rho_2}{\alpha_1 \rho_1} \cos 2\varphi_T & -\frac{\beta_2 \rho_2}{\alpha_1 \rho_1} \sin 2\varphi_T \\ \sin 2\theta_R & -\frac{\alpha_1}{\beta_1} \cos 2\varphi_R & \frac{\rho_2 \beta_2^2 \alpha_1}{\rho_1 \beta_1^2 \alpha_2} \sin 2\theta_T & \frac{\rho_2 \beta_2 \alpha_1}{\rho_1 \beta_1^2} \cos 2\varphi_T \end{pmatrix} \begin{pmatrix} R_{PP} \\ R_{PS} \\ T_{PP} \\ T_{PS} \end{pmatrix} = \begin{pmatrix} \cos \theta_R \\ \sin \theta_R \\ \cos 2\varphi_R \\ \sin 2\theta_R \end{pmatrix}. \quad (3.35)$$

### 3.3. Sobre las reflectividades

El objetivo de los métodos sísmicos de prospección es caracterizar el subsuelo. Éste consiste en una serie de capas de diferente estructura, litología, porosidad y contenido de fluidos. Desde el punto de vista geológico esta caracterización generalmente se realiza analizando datos de perfilaje de pozo a través de la medición de algún parámetro  $Z$ . Desde el punto de vista geofísico esta caracterización se realiza analizando los cambios de dicho parámetro en las discontinuidades que separan las capas que componen el subsuelo (Russell, 2014). Esta diferencia es llamada reflectividad  $R$ . La relación entre  $Z$  y  $R$  está dada por

$$R_Z = \frac{\Delta Z}{2Z}, \quad (3.36)$$

donde  $\Delta Z = Z_{i+1} - Z_i$  y  $Z = (Z_{i+1} + Z_i)/2$  y el subíndice  $i$  hace referencia a la  $i$ -ésima capa. Para cada discontinuidad, la reflectividad del parámetro  $T$  se obtiene dividiendo el cambio en dicho parámetro por el doble de su valor medio. Esta notación será utilizada a lo largo de esta tesis. El parámetro  $T$  puede tomar diferentes nombres y significados. Por ejemplo, la reflectividad asociada a la velocidad  $\alpha$ , será  $R_\alpha = \Delta\alpha/2\alpha$ .

Dado un modelo de velocidades y densidades, las ecuaciones de Zoeppritz permiten obtener coeficientes de reflexión y transmisión en función de los ángulos de incidencia y transmisión. Sin embargo, las expresiones de dichos coeficientes son complicadas y su dependencia con el ángulo de incidencia es poco intuitiva y no lineal. En efecto, resolviendo el sistema de ecuaciones (3.35), el coeficiente de reflexión exacto  $R_{PP}$  para la onda compresional incidente y reflejada, con el cual nos limitaremos a trabajar en esta tesis, puede expresarse como (Aki y Richards, 1980)

$$R_{PP} = \frac{\left(b \frac{\cos \theta_I}{\alpha_1} - c \frac{\cos \theta_T}{\alpha_2}\right) F - \left(a + d \frac{\cos \theta_I}{\alpha_1} \frac{\cos \varphi_T}{\beta_2}\right) H p^2}{D}, \quad (3.37)$$

donde

$$p = \frac{\sin \theta_I}{\alpha_1}, \quad (3.38)$$

$$a = \rho_2 (1 - 2\beta_2^2 p^2) - \rho_1 (1 - 2\beta_1^2 p^2), \quad (3.39)$$

$$b = \rho_2 (1 - 2\beta_2^2 p^2) + 2\rho_1 \beta_1^2 p^2, \quad (3.40)$$

$$c = \rho_1 (1 - 2\beta_1^2 p^2) + 2\rho_2 \beta_2^2 p^2, \quad (3.41)$$

$$d = 2 (\rho_2 \beta_2^2 - \rho_1 \beta_1^2), \quad (3.42)$$

$$F = b \frac{\cos \varphi_R}{\beta_1} + c \frac{\cos \varphi_T}{\beta_2}, \quad (3.43)$$

$$H = a - d \frac{\cos \theta_T}{\alpha_2} \frac{\cos \varphi_R}{\beta_1}, \quad (3.44)$$

$$D = \frac{\det(M)}{\alpha_1 \alpha_2 \beta_1 \beta_2}, \quad (3.45)$$

siendo  $\det(M)$  el determinante de la matriz del lado izquierdo de la ecuación (3.35).

### 3.4. Aproximaciones de las ecuaciones de Zoeppritz

Como puede observarse, la expresión del coeficiente de reflexión  $R_{PP}$  exacto es altamente no lineal y su relación con los parámetros del medio es compleja. Debido a esto las ecuaciones de Zoeppritz son poco prácticas para ser utilizadas en problemas de inversión. Para salvar estas dificultades varias aproximaciones lineales a las ecuaciones de Zoeppritz han sido desarrolladas por varios autores, tales como las aproximaciones de Bortfeld (1961), Aki y Richards (1980), Shuey (1985), Smith y Gidlow (1987), Fatti et al. (1994), Verm y Hilterman (1995), etc. En este trabajo nos limitaremos al estudio y uso de la aproximación de Aki y Richards de tres términos (Aki y Richards, 1980) y de la aproximación de Shuey de dos términos (Shuey, 1985), dos de la más utilizadas hoy en día.

La linealización de las ecuaciones de Zoeppritz parte de la hipótesis de que la variación relativa de los parámetros físicos es pequeña. Las expresiones resultantes depende del promedio entre el ángulo de incidencia  $\theta_I$  y el ángulo de transmisión  $\theta_T$  en cada discontinuidad donde se produce una reflexión. Por otro lado los datos sísmicos no se registran en función del ángulo de incidencia sino en función del *offset* (Sheriff, 2002). Antes de poder utilizar las aproximaciones en los problemas de inversión que serán estudiados, los datos tienen que ser convertidos de *offset-gathers* a *angle-gathers*. Esta conversión se realiza haciendo trazado de rayos a partir de un modelo de velocidades. Varios autores, como Dahl y Ursin (1991), Walden (1991) y Castle (1994), han propuesto diferentes técnicas para realizar dicha conversión.

Las curvas de AVA se obtienen midiendo la amplitud de cada traza sísmica de un *angle-gather* para un evento determinado. A partir del patrón de variación de la amplitud pueden llegarse a inferir propiedades de la zona de interés. La característica de la curva va a depender de la combinación de las propiedades físicas de las rocas y del contenido de fluido. La factibilidad de poder obtener información útil de este análisis

está determinada por la relación señal-ruido que presenta el dato y el rango de ángulos de incidencia presentes en el *angle-gather* (apertura).

#### 3.4.1. Aproximación de Aki y Richards

Sean las diferencias entre los parámetros elásticos de dos semiespacios  $\Delta\alpha = \alpha_2 - \alpha_1$ ,  $\Delta\beta = \beta_2 - \beta_1$ ,  $\Delta\rho = \rho_2 - \rho_1$  y sean los valores medios  $\alpha = \frac{1}{2}(\alpha_2 + \alpha_1)$ ,  $\beta = \frac{1}{2}(\beta_2 + \beta_1)$ ,  $\rho = \frac{1}{2}(\rho_2 + \rho_1)$ , los cambios relativos en los parámetros físicos están relacionados con las reflectividades correspondientes a través de

$$R_\alpha = \frac{\alpha_2 - \alpha_1}{\alpha_2 + \alpha_1} = \frac{1}{2} \frac{\Delta\alpha}{\alpha}, \quad (3.46)$$

$$R_\beta = \frac{\beta_2 - \beta_1}{\beta_2 + \beta_1} = \frac{1}{2} \frac{\Delta\beta}{\beta}, \quad (3.47)$$

$$R_\rho = \frac{\rho_2 - \rho_1}{\rho_2 + \rho_1} = \frac{1}{2} \frac{\Delta\rho}{\rho}. \quad (3.48)$$

Nótese que estas reflectividades tienen la misma forma matemática que la reflectividad debida al contraste de impedancia de las ondas  $P$  y  $S$  pero no el mismo significado físico. Si las propiedades elásticas de los medios son tales que  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  son mucho menores que la unidad entonces la relación del coeficientes de reflexión  $R_{PP}$  con el ángulo de incidencia puede ser aproximada por una ecuación lineal. A partir de esto último y asumiendo que los ángulos  $\theta_R$ ,  $\theta_T$ ,  $\varphi_R$  y  $\varphi_T$  son reales y menores a  $90^\circ$ , Aki y Richards (1980) se basaron en los trabajos de Bortfeld (1961) y de Richards y Frasier (1976) para desarrollar la aproximación de tres términos que lleva sus nombres:

$$R_{PP}(\theta) = (1 + \tan^2 \theta) R_\alpha - 8\gamma^2 \sin^2 \theta R_\beta + (1 - 4\gamma^2 \sin^2 \theta) R_\rho, \quad (3.49)$$

donde

$$\gamma = \frac{\beta}{\alpha}, \quad (3.50)$$

y

$$\theta = \frac{1}{2}(\theta_I + \theta_T) = \frac{1}{2} \left[ \theta_I + \arcsin \left( \frac{\alpha_2}{\alpha_1} \sin \theta_I \right) \right]. \quad (3.51)$$

Esta ecuación es la aproximación de Aki y Richards de tres términos para una onda  $P$  incidente y reflejada. Al igual que sucede con las ecuaciones de Zoeppritz, existen expresiones similares para todo tipo de onda plana incidente y reflejada. Esta aproximación nos da de manera clara la contribución independiente de  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  en la reflectividad  $R_{PP}$  en función del ángulo de incidencia.

Al linealizar el problema se asume que  $\gamma$  y  $\theta$  son conocidos *a priori*, pero no siempre esta información es conocida con precisión. La influencia de  $\gamma$  en el valor del coeficiente de reflexión depende del contraste de los parámetros elásticos del medio (Alemie, 2010). Cuanto mayor es dicho contraste el efecto de la incertidumbre en el  $\gamma$  utilizado sobre el coeficiente de reflexión es mayor. Como se verá en la sección 3.4.3, lo mismo sucede con la incertidumbre en el ángulo de incidencia utilizado.

### 3.4.2. Aproximación de Shuey

Shuey (1985) propuso modificar la ecuación de Aki y Richards reemplazando la velocidad  $\beta$  y su variación  $\Delta\beta$  a través de la incorporación del coeficiente de Poisson  $\sigma$ . La motivación de dicho remplazo se puede encontrar en los trabajos de Koefoed (1955, 1962), quien demostró que la dependencia del coeficiente de reflexión con el ángulo de incidencia está fuertemente influenciada por el valor de  $\sigma$ . El coeficiente de Poisson es la relación entre la deformación transversal y longitudinal del medio debida al esfuerzo producido, por ejemplo, por el paso de una onda compresional. Por otro lado está vinculado directamente con la elasticidad de las rocas y se relaciona con los parámetros elásticos del medio a través de

$$\sigma = \frac{\alpha^2 - 2\beta^2}{2(\alpha^2 - \beta^2)}. \quad (3.52)$$

De esta ecuación también se hace evidente que el coeficiente de Poisson se relaciona con el tipo de fluido contenido en las rocas, debido a que rocas saturadas con fluidos compresibles, como son los hidrocarburos poseen valores de  $\alpha$  menores, y valores de  $\beta$  levemente mayores, que aquellas que están saturadas con fluidos incompresibles, como es el agua (Simm y Bacon, 2014). En particular para un fluido se tiene que  $\beta = 0$ , por lo que  $\sigma = 0.5$ . Además de ser un indicador de la presencia de fluidos, el coeficiente de Poisson  $\sigma$  también es un indicador de litología.

Reemplazando  $\beta$  y  $\Delta\beta$  en la ecuación (3.49) por  $\sigma$  y  $\Delta\sigma$  obtenidos a través de la ecuación (3.52) se obtiene la ecuación de Shuey de tres términos

$$R_{PP}(\theta) = R_0 + \left[ R_0 H_0 + \frac{\Delta\sigma}{(1 - \sigma)^2} \right] \sin^2 \theta + R_\alpha (\tan^2 \theta - \sin^2 \theta), \quad (3.53)$$

donde

$$R_0 \approx \frac{1}{2} \left( \frac{\Delta\alpha}{\alpha} + \frac{\Delta\rho}{\rho} \right) = R_\alpha + R_\rho, \quad (3.54)$$

$$H_0 = H - 2(1 + H) \frac{1 - 2\sigma}{1 - \sigma}, \quad (3.55)$$

$$H = \frac{\Delta\alpha/\alpha}{\Delta\alpha/\alpha + \Delta\rho/\rho}, \quad (3.56)$$

$$\Delta\sigma = (\sigma_2 - \sigma_1), \quad (3.57)$$

$$\sigma = \frac{1}{2}(\sigma_2 + \sigma_1), \quad (3.58)$$

$$\theta = \frac{1}{2}(\theta_I + \theta_T) = \frac{1}{2} \left[ \theta_I + \arcsin \left( \frac{\alpha_2}{\alpha_1} \sin \theta_I \right) \right]. \quad (3.59)$$

Al derivarse de la ecuación de Aki y Richards, la ecuación de Shuey también se basa en la hipótesis de que el cambio relativo de las propiedades elásticas es pequeño, pero a diferencia de aquella la ecuación de Shuey está organizada de forma que permite analizar que combinación de parámetros elásticos contribuyen a  $R_{PP}$  según el valor del ángulo de incidencia  $\theta$ . El tercer término de la ecuación (3.53), que sólo se relaciona con cambios en  $\alpha$ , se comporta como  $\theta^4$  y por lo tanto es despreciable cuando  $\theta < 30^\circ$ ,

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack

sin embargo para ángulos mayores comienza a ser el término dominante. Si el tercer término es despreciado obtendremos la clásica aproximación de Shuey de dos términos

$$R_{PP}(\theta) = R_0 + G \sin^2 \theta, \quad (3.60)$$

donde

$$G = R_0 H_0 + \frac{\Delta\sigma}{(1 - \sigma)^2} \quad (3.61)$$

es un atributo llamado *gradient*. En la práctica tanto el *gradient* como  $R_0$ , llamado *intercept*, pueden ser estimados de los *angle-gathers* ajustando por mínimos cuadrados una recta a las amplitudes de un evento determinado en función de  $\sin^2 \theta$ . Este es esencialmente la base del AVA convencional. A partir de este procedimiento se obtienen dos secciones sísmicas, una relacionada con el *intercept* conteniendo información sobre la incidencia normal de las ondas y otra relacionada con el *gradient* conteniendo información sobre los cambios litológicos en el subsuelo. Sólo esta última provee información acerca del contraste del coeficiente de Poisson  $\Delta\sigma$  y por lo tanto información acerca de variaciones en el contenido de fluidos de las rocas.

#### 3.4.3. Comparación entre la aproximación de Aki y Richards y la aproximación de Shuey

En esta sección se comparan los valores del coeficiente de reflexión  $R_{PP}$  obtenidos mediante la ecuación de Zoeppritz, la aproximación de Aki y Richards de tres términos y la aproximación de Shuey de dos términos para los modelos detallados en la Tabla 3.1 (Nowak et al., 2008; Alemie, 2010). El segundo modelo representa el contacto entre arcilla y arena con gas (Rubino y Velis, 2009). Obviaremos la comparación entre la aproximación de Aki y Richards y de Shuey de tres términos ya que se tratan de la misma ecuación expresada en función de diferentes parámetros físicos, y por lo tanto son numéricamente idénticas. La principal diferencia entre los modelos es el contraste entre las velocidades y las densidades, siendo el contraste en el modelo 1 menor que en modelo 2.

Modelo	Capa	$\alpha$ (m/s)	$\beta$ (m/s)	$\rho$ (g/cm <sup>3</sup> )	$\gamma$	$\theta_C$
1	1	3000	1800	2.20	0.61	70°
	2	3200	2000	2.25		
2	1	3094	1515	2.40	0.56	50°
	2	4050	2526	2.21		

**Tabla 3.1.** Modelos de velocidades  $\alpha$  y  $\beta$ , densidades  $\rho$ , parámetros  $\gamma$  y ángulo de incidencia crítico  $\theta_C$  utilizados para crear las curvas de la Figura 3.2

Para estudiar el efecto de las incertezas en el ángulo de incidencia, los cálculos de los coeficientes de reflexión mediante las aproximaciones se realizaron utilizando los ángulos de incidencia  $\theta$  calculados con la ecuación (3.51) y los ángulos de incidencia  $\theta_I$  utilizados en la ecuación de Zoeppritz. Dichos ángulos se tomaron siempre menores al

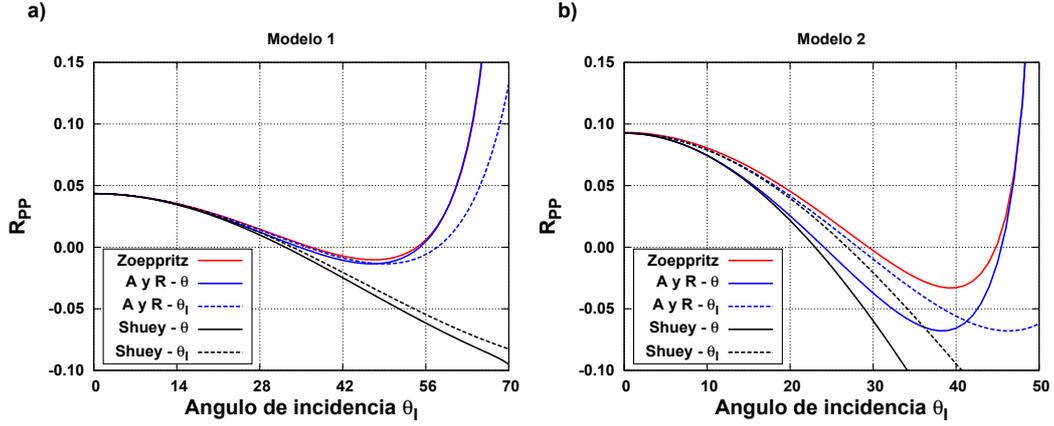
ángulo crítico  $\theta_C$ , valor que se especifica en la Tabla 3.1. Para el caso de la aproximación de Aki y Richards, el valor de  $\gamma$  utilizado fue el valor exacto dado en la Tabla 3.1.

En la Figura 3.2 se pueden ver las curvas creadas a partir de ambos modelos. Las curvas rojas se obtuvieron mediante la ecuación de Zoeppritz, las curvas azules continuas mediante la ecuación de Aki y Richards utilizando el ángulo  $\theta$  y las curvas discontinuas mediante el ángulo  $\theta_I$ . Similar criterio siguen las curvas color negro obtenidas por medio de la aproximación de Shuey. Como se puede observar en la Figura 3.2a, para el modelo de menor contraste y para ángulos de incidencia menores a  $30^\circ$ , todas las aproximaciones se ajustan de manera casi perfecta a la curva que representa el coeficiente de reflexión exacto. Las diferencias debidas a pequeñas incertezas en el ángulo de incidencia no son significativas. Por encima de los  $30^\circ$  la ecuación de Shuey de dos términos, para ambos ángulos, se aparta completamente del coeficiente de reflexión exacto. Los resultados de desprestigiar el tercer término son notables. En el caso de la ecuación de Aki y Richards se puede observar claramente el efecto sobre los resultados de las incertezas en el ángulo de incidencia. Cuando se utiliza el ángulo  $\theta$  el coeficiente  $R_{PP}$  aproximado se ajusta casi perfectamente al exacto para todo el rango de ángulos. Sin embargo cuando se utiliza el ángulo  $\theta_I$  comienzan a haber diferencias significativas entre el valor aproximado y el exacto para valores mayores a  $50^\circ$  del ángulo de incidencia.

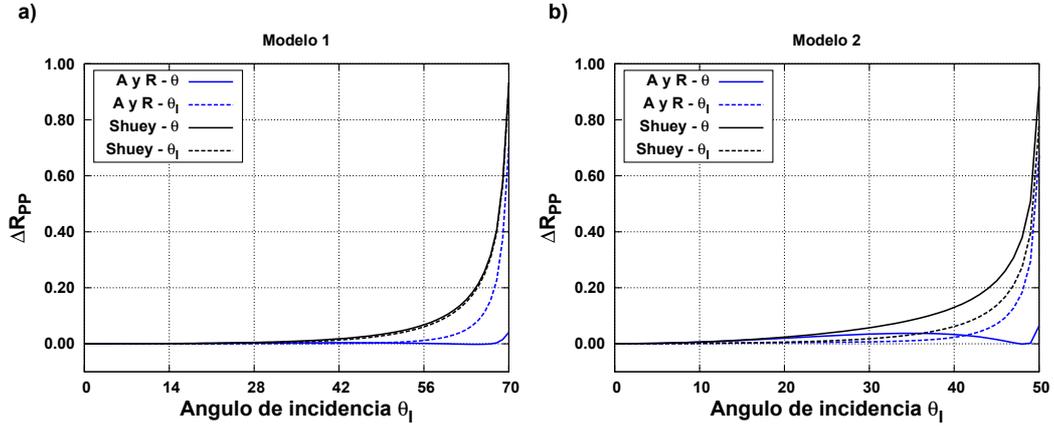
La Figura 3.2b muestra los resultados obtenidos utilizando el modelo de mayor contraste. Como se puede apreciar los coeficientes de reflexión aproximados muestran un notable apartamiento respecto al valor exacto. Las curvas obtenidas mediante las aproximaciones utilizando el ángulo  $\theta_I$  ajustan correctamente al valor exacto para ángulos menores a  $25^\circ$ . Para valores mayores ambas curvas se apartan del valor exacto. Para la aproximación de Shuey este comportamiento es similar al observado en el primer modelo, pero difiere notablemente para la aproximación de Aki y Richards. Como era de esperarse las curvas calculadas utilizando el ángulo  $\theta$  ajustan mejor al valor exacto que aquellas calculadas con el ángulo  $\theta_I$ , sin embargo estas soluciones muestran una desmejora que aquellas obtenidas con el primer modelo. Para el caso de la aproximación de Shuey se ve que el valor del ángulo de incidencia a partir del cual el coeficiente aproximado comienza a apartarse del exacto es menor que en el modelo de bajo contraste. La curva obtenida utilizando la aproximación de Aki y Richards ajusta mejor a los valores exactos que las otras curvas ya analizadas para este modelo, sin embargo para un rango de ángulos considerable se observan diferencias con respecto a los valores exactos que no se observan en el análisis sobre el primer modelo.

La Figura 3.3 muestra la diferencia entre los valores del coeficiente de reflexión calculados para ambos modelos de la Tabla 3.1 utilizando las aproximaciones de Aki y Richards y Shuey para los ángulos  $\theta$  y  $\theta_I$  (curvas negras y azules de la Figura 3.2) y el valor calculado utilizando la ecuación de Zoeppritz (curva roja en la Figura 3.2). Como era esperable, el error resultante de aproximar el coeficiente de reflexión por medio de la ecuación de Aki y Richards de tres términos es, en general, menor que el observado al utilizar la aproximación de Shuey de dos términos para todo el rango de ángulos utilizado.

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack



**Figura 3.2.** Coeficientes de reflexión  $R_{PP}$  en función del ángulo de incidencia  $\theta$  (curvas continuas) y  $\theta_I$  (curvas discontinuas), utilizando la ecuación de Zoeppritz (curvas rojas), la aproximación de Aki y Richards (curvas azules) y la aproximación de Shuey de dos términos (curvas negras) para a) el modelo 1 y b) el modelo 2 cuyos parámetros físicos están detallados en la Tabla 3.1.



**Figura 3.3.** Diferencia entre el coeficiente de reflexión calculado utilizando las aproximaciones de Aki y Richards y Shuey, para los ángulos  $\theta$  y  $\theta_I$  y el calculado utilizando la ecuación de Zoeppritz para los modelos de la Tabla 3.1. a) Modelo 1 y b) modelo 2.

#### 3.4.4. Otras aproximaciones

La Tabla 3.2 resume algunas características de las aproximaciones a las ecuaciones de Zoeppritz ya nombradas. A excepción de la aproximación de Bortfeld (1961), todas se obtienen partiendo de la aproximación de Aki y Richards (1980). La aproximación de Bortfeld fue el primer intento de obtener una relación que simplificara el análisis de AVA. Sin embargo, en esta aproximación, la relación con los parámetros físicos y el ángulo de incidencia sigue siendo compleja. Las ecuaciones de Smith y Gidlow (1987), Fatti et al. (1994) y Verm y Hilterman (1995) dan una relación simple entre el coeficiente de reflexión y el ángulo de incidencia. En estas aproximaciones la variación relativa entre los parámetros de los medios está expresada por medio de reflectividades. Al igual que en la ecuación de Aki y Richards, la ecuación de Smith y Gidlow depende de las reflectividades  $R_\alpha$  y  $R_\beta$  asociadas a las velocidades  $\alpha$  y  $\beta$ . Para derivar esta

aproximación los autores asumieron que se satisface la ecuación derivada por Gardner et al. (1974) que relaciona la velocidad  $\alpha$  con la densidad  $\rho$ . La ecuación desarrollada por Fatti y sus colaboradores está expresada en función de las reflectividades  $R_{IP}$  y  $R_{IS}$  asociadas al contraste de impedancia acústica de la onda  $P$  y de la onda  $S$ . La impedancia acústica se define como el producto de la densidad por la velocidad de propagación (Sheriff, 2002). La aproximación de Verm y Hilterman depende de  $R_0 = R_\alpha + R_\rho$  (ecuación (3.59)) y de  $R_\sigma$  que, si bien no sigue la forma de la ecuación (3.36), es llamada reflectividad de Poisson.

### 3.5. AVA y anisotropía

El desarrollo de las ecuaciones de Zoeppriz y sus correspondientes aproximaciones se hizo bajo las hipótesis de que los medios son isótropos y homogéneos, situación que con frecuencia no ocurre en la realidad. La variación de la velocidad de las ondas sísmicas en función de la dirección en la cual se propagan, fenómeno llamado anisotropía, depende, entre otras cosas, del estado de fracturación del medio. El conocimiento de la fracturación del subsuelo tiene especial importancia en el estudio de yacimientos no convencionales, como pueden ser reservas de *tight-oil* y *tight-gas*. La producción de hidrocarburos en un reservorio de baja porosidad, como por ejemplo aquellos formados por rocas carbonáticas, está relacionada muchas veces con la fracturación natural propia de las rocas que lo forman, ya que estas fracturas permiten el desplazamiento de los fluidos. Desde un punto de vista económico, la caracterización de estas fracturas es crucial para una producción eficiente y rentable de hidrocarburos. En los últimos años el estudio de la variación de la amplitud de la ondas reflejadas, en función del ángulo de incidencia y del azimut, o AVAz, se ha convertido en una técnica robusta para la caracterización de fracturas (Rüger y Gray, 2014). Así como el AVA es sensible a la presencia de fluidos en las rocas, el AVAz es sensible a los parámetros que caracterizan la anisotropía del medio, por ejemplo los parámetros de Thomsen (1986). El estudio de AVAz se basa en ecuaciones similares a las ecuaciones de AVA vistas en este capítulo, es decir, ecuaciones que permiten calcular los coeficientes de reflexión en función del ángulo de incidencia y del azimut (Rüger, 2002).

En esta tesis sólo se considerarán medio isótropos y homogéneos, por lo que los efectos de AVAz no serán tenidos en cuenta. No obstante, las diversas estrategias de inversión que se desarrollarán en los siguientes capítulos son lo suficientemente generales y flexibles como para poder incorporar el AVAz en la formulación y permitir su aplicación a casos donde los efectos de la anisotropía son significativos.

### 3.6. Problema directo en AVA

Como ya se observó en la sección 2.1, para resolver un problema inverso es necesario tener suficiente conocimiento de la física del problema como para poder generar datos sintéticos a partir de un modelo. Es decir, dada una parametrización, es necesario saber resolver el problema directo. En esta sección se describirá brevemente el modelo convolucional (Robinson, 1954, 1967) y se planteará el problema directo que será utilizado durante la resolución del problema de inversión de datos sísmicos *prestack* a tratar en

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack

Bortfeld (1961)	$R_{PP}(\theta_I) = \frac{1}{2} \ln \frac{\alpha_2 \rho_2 \cos \theta_I}{\alpha_1 \rho_1 \cos \theta_T} + \frac{\sin^2 \theta_I}{\alpha_1^2} (\beta_1^2 - \beta_2^2) \left( 2 + \frac{\ln \frac{\rho_2}{\rho_1}}{\ln \frac{\alpha_2}{\alpha_1} - \ln \frac{\alpha_2 \beta_1}{\alpha_1 \beta_2}} \right)$	
	Resuelve para:	Características y limitaciones:
		Fue la primera aproximación a las ecuaciones de Zoeppritz. Simplifica la forma en la que el coeficiente de reflexión se relaciona con el ángulo de incidencia. Sin embargo no es lineal con respecto a los parámetros elásticos haciendo difícil su aplicación en problemas de inversión.
Aki y Richards (1980)	$R_{PP}(\theta) = (1 + \tan^2 \theta) R_\alpha - 8\gamma^2 \sin^2 \theta R_\beta + (1 - 4\gamma^2 \sin^2 \theta) R_\rho$	
	Resuelve para:	Características y limitaciones:
	$R_\alpha, R_\beta, R_\rho$	En situaciones donde los parámetros elásticos presentan poco contraste, esta ecuación ajusta casi exactamente al coeficiente de reflexión obtenido mediante la ecuación de Zoeppritz hasta ángulos cercanos al ángulo crítico. Se asume conocido el valor de $\gamma$ .
Shuey (1985)	$R_{PP}(\theta) = R_0 + G \sin^2 \theta + R_\alpha (\tan^2 \theta - \sin^2 \theta)$	
	Resuelve para:	Características y limitaciones:
	$R_0, G$ y $R_\alpha$	Parte de la ecuación de Aki y Richards. El tercer término se desprecia para ángulos menores a $30^\circ$ . Variaciones en $G$ pueden estar relacionadas con cambios en la cantidad de fluidos, pero también se pueden deber a cambios en la litología.
Smith y Gidlow (1987)	$R_{PP}(\theta) = \left( \frac{3}{4} + \tan^2 \theta - \gamma^2 \sin^2 \theta \right) R_\alpha - 8\gamma^2 \sin^2 \theta R_\beta$	
	Resuelve para:	Características y limitaciones:
	$R_\alpha, R_\beta$	Se asume que la densidad cumple con la ecuación de Gardner. El ángulo de incidencia debe ser menor que el ángulo crítico $\theta_C$ . Se asume conocido el valor de $\gamma$
Fatti et al. (1994)	$R_{PP}(\theta) = \frac{1}{2} R_{IP} (1 + \tan^2 \theta) - 4\gamma^2 R_{IS} \sin^2 \theta$	
	Resuelve para:	Características y limitaciones:
	$R_{IP}, R_{IS}$	Funciona para ángulos menores a $50^\circ$ . Se obtiene asumiendo cambios pequeños de la densidad $\rho$ . Se asume conocido el valor de $\gamma$
Verm y Hilterman (1995)	$R_{PP}(\theta) = R_P \cos^2 \theta + R_\sigma \sin^2 \theta$	
	Resuelve para:	Características y limitaciones:
	$R_P, R_\sigma = \frac{\Delta\sigma}{(1-\sigma)^2}$	Está basada en la ecuación de Shuey. Funciona para ángulos menores a $30^\circ$ . Asume que $\gamma = 0.5$ . Cambios en la reflectividad de Poisson $R_\sigma$ pueden indicar cambios en el contenido de fluidos de las rocas.

**Tabla 3.2.** Resumen de algunas aproximaciones a la ecuación de Zoeppritz para  $R_{PP}$ .

los próximos capítulos.

Las ondas compresionales generadas en la superficie por alguna fuente impulsiva, como pueden ser explosivos, un cañón de aire o la caída de un peso, serán representadas matemáticamente a través de una ondícula de banda limitada y duración finita. A medida que las ondas sísmicas se propagan por el subsuelo e interactúan con las discontinuidades que separan las diferentes capas que lo conforman, parte de su energía se transmite y parte se refleja. La señal registrada en la superficie, correspondiente a la porción de energía reflejada, constituye el dato observado o sismograma (traza sísmica). El porcentaje de energía que se transmite y se refleja depende del valor del coeficiente de reflexión que caracteriza a cada discontinuidad, por lo que será natural caracterizar el subsuelo a través de una serie de coeficientes de reflexión de diferente amplitud.

El modelo convolucional de un sismograma se basa en las siguientes hipótesis:

1. El subsuelo está compuesto por capas horizontales y paralelas de propiedades elásticas constantes.
2. Las ondas sísmicas son ondas compresionales planas que inciden de forma normal sobre las discontinuidades que separan las diferentes capas que forman el subsuelo. Debido a esto no se generan ondas de corte reflejadas ni transmitidas.
3. La ondícula es estacionaria.

En la realidad estas hipótesis no se cumplen estrictamente. La primera hipótesis es incorrecta en lugares donde la geología presenta complejidad estructural o en lugares donde se producen variaciones laterales de las propiedades físicas del medio. La segunda hipótesis establece que las trazas del dato sísmico deben ser de *offset* cero, lo que es experimentalmente imposible, sin embargo esto se cumple aproximadamente si los reflectores son profundos ya que el ángulo con el cual incidirá la onda en la discontinuidad será pequeño. Al trabajar con *angle-gathers*, donde a veces el ángulo de incidencia puede llegar a ser mayor a  $40^\circ$ , esta hipótesis se aleja aún más de la realidad. Al incidir la onda oblicuamente sobre la discontinuidad se producen cambios de fase en la ondícula que no serán observados en el dato generado mediante el modelo convolucional. Por otro lado, a medida que las ondas se propagan por el subsuelo se producen fenómenos de atenuación y dispersión que se manifiestan en cambios en la forma y amplitud de la ondícula. Es decir, la ondícula no es estacionaria. Todo esto hace que, en principio, haya diferencias significativas entre el dato sintético generado utilizando el modelo convolucional y el dato observado. Un análisis completo de un sismograma debería contemplar los efectos de atenuación y dispersión, cambios de forma de la ondícula, y pérdidas de energía debida a la transmisión a través de las discontinuidades y reflexiones múltiples que no están siendo consideradas. A pesar de todas estas dificultades, a la hora de resolver problemas de inversión es posible procesar al dato observado para que todos estos fenómenos que nos apartan de las hipótesis antes enumeradas sean corregidos y el modelo convolucional puede ser aplicado de forma satisfactoria.

Utilizando el modelo convolucional es posible representar la  $i$ -ésima traza sísmica de un *angle-gather*, correspondiente al ángulo de incidencia  $\theta_i$ , como

$$\mathbf{d}(\theta_i) = \mathbf{w} * \mathbf{r}(\theta_i), \quad i = 1, \dots, N_\theta \quad (3.62)$$

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack

donde  $N_\theta$  es la cantidad de trazas en el *angle-gather*,  $\mathbf{r}$  es un vector de dimensión  $L_r$  representando los coeficientes de reflexión,  $\mathbf{w}$  es un vector de dimensión  $L_w$  representando la ondícula,  $\mathbf{d}$  es un vector de dimensión  $L_d = L_w + L_r - 1$  representando la traza sísmica, y “\*” es el operador de convolución.

Considerando la validez de las aproximaciones a las ecuaciones de Zoeppritz analizadas en las secciones previas, el coeficientes de reflexión para un dado reflector en función de un ángulo de incidencia  $\theta_i$  puede ser expresado, de forma general como

$$r(\theta_i) = \sum_{k=1}^n g_k(\theta_i) m_k, \quad (3.63)$$

donde  $m_k$  son parámetros que dependen de las propiedades físicas de las rocas a cada lado de la discontinuidad (velocidades y densidades),  $n$  es el orden de la aproximación utilizada (usualmente  $n = 2$  o  $3$ ), y  $g_k(\theta_i)$  son funciones que dependen del ángulo de incidencia, que debe ser menor al ángulo crítico (Ikelle y Amundsen, 2005). La Tabla 3.3 resume la forma que tomarán  $m_k$  y  $g_k(\theta_i)$  para la aproximación de Aki y Richards de tres términos y la aproximación de Shuey de dos términos. En la práctica el vector de los coeficientes de reflexión de la ecuación (3.62) será  $\mathbf{r}(\theta_i) = (r_1(\theta_i), \dots, r_{L_r}(\theta_i))$ , donde se tendrá una ecuación (3.63) asociada a cada uno de los  $r_l(\theta_i)$  con  $l = 1, \dots, L_r$ .

Combinando las ecuaciones (3.62) y (3.63) se tiene que la traza correspondiente al ángulo de incidencia  $\theta_i$  se puede expresar como

$$\mathbf{d}(\theta_i) = \mathbf{A}(\theta_i) \mathbf{m}, \quad i = 1, \dots, N_\theta \quad (3.64)$$

donde  $\mathbf{m} = (m_{11}, \dots, m_{L_r 1}, \dots, m_{1n}, \dots, m_{L_r n})^T$  y  $\mathbf{A}(\theta_i)$  es una matriz aumentada de dimensión  $L_d \times L_r n$  que se puede expresar como

$$\mathbf{A}(\theta_i) = [\mathbf{A}_1(\theta_i) | \dots | \mathbf{A}_n(\theta_i)]. \quad (3.65)$$

Aquí,  $\mathbf{A}_k(\theta_i)$  con  $k = 1, \dots, n$  son sub-matrices de dimensión  $L_d \times L_r$ , cuyos elementos están dados por

$$\{\mathbf{A}_k(\theta_i)\}_{hl} = g_k(\theta_i) w_{h-l+1}, \quad (3.66)$$

para  $h = 1, \dots, L_d$  y  $l = 1, \dots, L_r$ , y donde los  $w$  son los elementos de la ondícula  $\mathbf{w}$ . Entonces, los  $N_\theta$  sistemas de ecuaciones dadas por la ecuación (3.64) pueden ser re-organizados en un único sistema

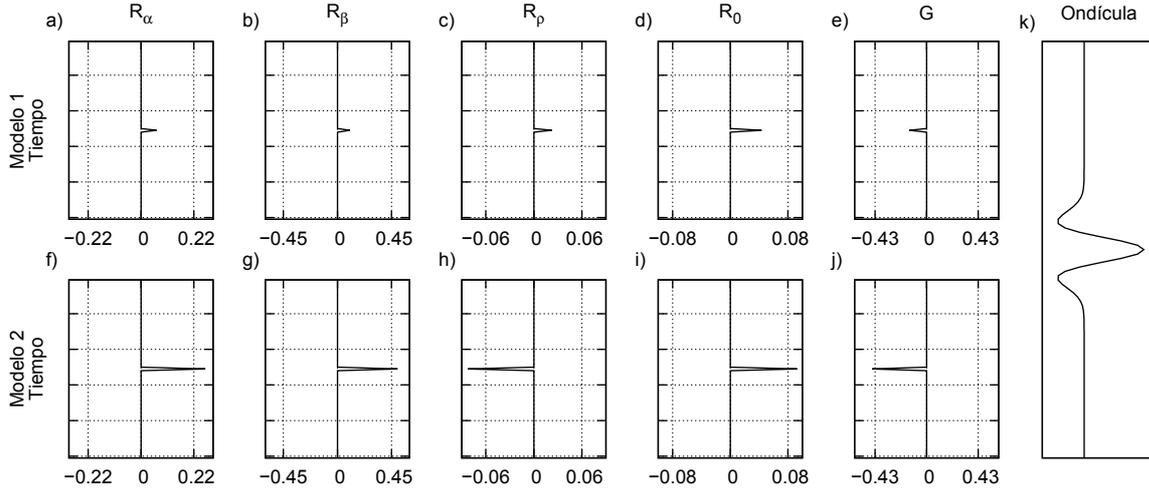
$$\mathbf{d} = \mathbf{A} \mathbf{m}, \quad (3.67)$$

donde  $\mathbf{A}$  es una matriz columna por bloques, donde cada bloque está dado por  $\mathbf{A}(\theta_i)$ , y  $\mathbf{d}$  es un vector columna por bloques, donde cada bloque está dado por  $\mathbf{d}(\theta_i)$ , para  $i = 1, \dots, N_\theta$ . Como se puede deducir de las ecuaciones (3.65) y (3.66), la matriz  $\mathbf{A}$  tomará diferente forma y dimensión dependiendo de la aproximación a las ecuaciones de Zoeppritz que se utilice. Asimismo la dimensión del vector  $\mathbf{m}$  dependerá del orden  $n$  de la aproximación elegida.

La ecuación (3.67) permite generar *angle-gathers* sintéticos libres de ruido, conocidos los parámetros del modelo de subsuelo utilizado, representados por el vector  $\mathbf{m}$ , y conocidos la ondícula  $\mathbf{w}$  y los valores de los ángulos  $\theta_i$ , representados por la matriz  $\mathbf{A}$ .

	$n$	$m_1$	$m_2$	$m_3$	$g_1(\theta_i)$	$g_2(\theta_i)$	$g_3(\theta_i)$
Aki y Richards	3	$R_\alpha$	$R_\beta$	$R_\rho$	$(1 + \tan^2 \theta_i)$	$-8\gamma^2 \sin^2 \theta_i$	$(1 - 4\gamma^2 \sin^2 \theta_i)$
Shuey	2	$R_0$	$G$	-	1	$\sin^2 \theta_i$	-

**Tabla 3.3.** Expresión de los parámetros  $m_k$  y de las funciones  $g_k(\theta_i)$  para la aproximación de Aki y Richards y la aproximación de Shuey de dos términos.



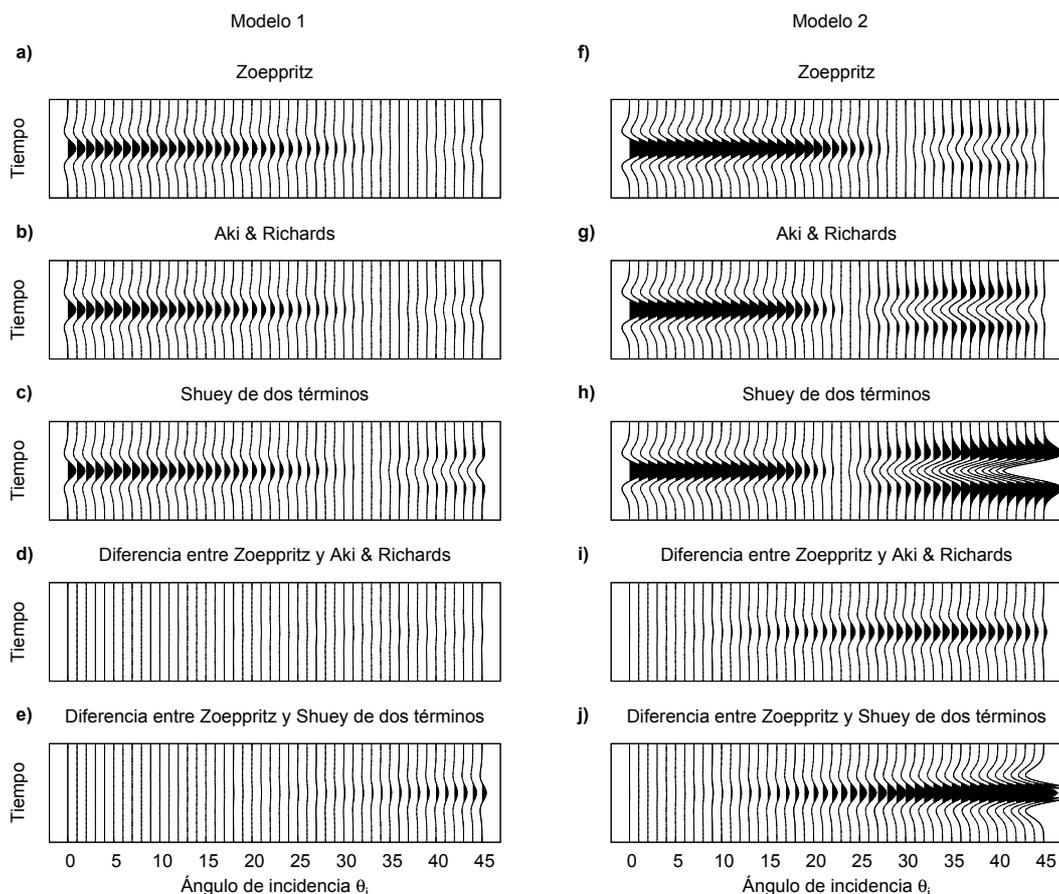
**Figura 3.4.** Parámetros de las aproximaciones de Aki y Richards y Shuey de dos términos para los modelos detallados en la Tabla 3.1.  $R_\alpha$ ,  $R_\beta$ ,  $R_\rho$ ,  $R_0$  y  $G$  para a)-e) el modelo de menor contraste (modelo 1) y f)-j) el modelo de mayor contraste (modelo 2), k) ondícula de Ricker con la cual se generó el dato de la Figura 3.5

Esta ecuación resuelve el problema directo que será la base de todos los problemas de inversión que serán resueltos en los siguientes capítulos.

A modo de ejemplo, y como contribución a la comparación llevada a cabo en la sección 3.4.3, se generó dato sintético a partir de los modelos dados en la Tabla 3.1. Al ser modelos compuestos por dos espacios semi-infinitos tendrán únicamente un reflector cada uno. En la Figura 3.4 se puede observar las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  de la aproximación de Aki y Richards, y el *intcept*  $R_0$  y el *gradient*  $G$  de la aproximación de Shuey de dos términos calculados a partir de dichos modelos. Estos valores se encuentran contenidos en el vector  $\mathbf{m}$ . Como se puede observar, las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  presentan grandes diferencias de magnitud entre si. En particular la pequeña magnitud de  $R_\rho$ , que se debe a la poca variación que presenta la densidad, es un indicativo de que dicha reflectividad tiene poca influencia en el dato observado. La ondícula utilizada en el modelo convolucional es una ondícula de Ricker (Ricker, 1940) cuya forma puede observarse en la Figura 3.4k. Se eligió un rango de ángulos desde  $\theta_1 = 0^\circ$  a  $\theta_{45} = 45^\circ$ , con paso de  $1^\circ$ . Esta información y la ondícula se encuentran contenidas en la matriz  $\mathbf{A}$ .

A partir de la ecuación (3.67) se crearon *angle-gathers* de 45 trazas cada uno para ambos modelos y ambas aproximaciones. Por otro lado, y con fines comparativos,

### 3. Problema directo: de Zoeppritz a los datos sísmicos prestack



**Figura 3.5.** Datos sintéticos generados a partir de las reflectividades y la ondícula de la Figura 3.5, utilizando a) y f) la ecuación de Zoeppritz, b) y g) la aproximación de Aki y Richards y c) y h) la aproximación de Shuey de dos términos. d) y i) Comparación entre Zoeppritz y Aki y Richards, e) y j) comparación entre Zoeppritz y Shuey de dos términos.

se crearon *angle-gathers* utilizando los coeficientes de reflexión calculados utilizando la ecuación de Zoeppritz. La Figura 3.5 muestra los resultados correspondientes. Un análisis general muestra, como es esperable, que para el modelo de menor contraste la diferencia entre el dato generado a partir de ambas aproximaciones y el dato generado a partir de las ecuaciones de Zoeppritz es menor que al utilizar el modelo de mayor contraste. En la Figura 3.5d se puede ver que para el modelo de menor contraste el dato creado utilizando la ecuación de Aki y Richards (Figura 3.5b) ajusta casi perfectamente al generado utilizando la ecuación de Zoeppritz (Figura 3.5a). Por otro lado, la Figura 3.5e muestra que el dato generado a partir de la ecuación de Shuey de dos términos (Figura 3.5c) presenta pequeñas diferencias con respecto al generado utilizando la ecuación de Zoeppritz a partir de los  $30^\circ$ . Estas diferencias son esperables ya que el tercer término de la aproximación de Shuey, que comienza a tener influencia en el dato a partir de los  $30^\circ$ , fue despreciado. Las diferencias entre el dato creado utilizando las aproximaciones y el creado utilizando la ecuación de Zoeppritz comienzan a ser considerables para el modelo de mayor contraste (Figuras 3.5i,j). Allí se pueden

observar diferencias significativas no sólo en los datos generados con la ecuación de Shuey sino también con la ecuación de Aki y Richards, especialmente para los ángulos más grandes.

### 3.7. Sumario

En este capítulo se realizó una revisión de la teoría de AVA. Se derivaron las ecuaciones de Zoeppritz que modelan la variación del coeficiente de reflexión en función del ángulo de incidencia para el caso particular en el cual que la onda incidente es una onda compresional. Se observó que cuando la onda reflejada también es una onda compresional la relación entre el coeficiente de reflexión y el ángulo de incidencia es no lineal, haciendo su uso poco práctico en problemas de inversión. Se realizó también una revisión de algunas aproximaciones lineales de la ecuación de Zoeppritz, planteando las hipótesis bajo las cuales se derivan. Se hizo énfasis en las aproximaciones de Aki y Richards de tres términos y la de Shuey de dos términos, ya que serán ampliamente utilizadas en esta tesis, realizando un análisis comparativo entre ellas. El análisis de los resultados mostró que la calidad del ajuste de las aproximaciones, respecto del valor exacto, se ve influenciada por los parámetros físicos del modelo de subsuelo utilizado. En particular el contraste entre los parámetros físicos influye notablemente en los resultados. Esto era esperable, ya que una de las hipótesis en las cuales se basan las aproximaciones es que los parámetros físicos deben presentar variaciones relativas pequeñas. Los resultados obtenidos sobre el modelo de menor contraste no sólo mostraron un mejor ajuste al valor exacto del coeficiente de reflexión, sino también una mayor tolerancia a incertezas en el ángulo de incidencia. También se observó que la calidad de los resultados desmejora a medida que el ángulo de incidencia aumenta. Además de las aproximaciones de Aki y Richards y Shuey, existen otras aproximaciones de gran utilidad, y a pesar de que no serán usadas en esta tesis se realizó resumen de las mismas, destacando sus características más importantes. Sobre el final de este capítulo se planteó la resolución del problema directo que será utilizado en los próximos capítulos al resolver problemas de inversión. Dicha resolución está basada en el modelo convolucional, del cual se enumeraron las hipótesis en las cuales se basa y se obtuvo una expresión matricial para generar datos sintéticos a partir de una serie de reflectividades y una ondícula. A modo de ejemplo se generaron datos sintéticos a partir de un modelo de bajo contraste y un modelo de alto contraste utilizando la ecuación de Zoeppritz, la aproximación de Aki y Richards de tres términos y la aproximación de Shuey de dos términos, y luego se procedió a realizar una comparación cualitativa de los mismos. De este análisis se desprende que la inversión de datos *prestack* deberá tener en cuenta el rango de ángulos a utilizar, así como la aproximación seleccionada, en función de las características del modelo geológico que da lugar a los datos empleados.



# Capítulo 4

## Inversión de datos sísmicos *prestack* por mínimos cuadrados

### 4.1. Introducción

En este capítulo procederemos a realizar inversión de datos sísmicos *prestack* por medio de los métodos de mínimos cuadrados y mínimos cuadrados amortiguados analizados en la sección 2.3. Allí se vio que estos métodos permiten estimar los parámetros del modelo por medio de las ecuaciones (2.11) y (2.17). Al realizar la inversión se asumirá que se cumplen las condiciones necesarias para poder utilizar el modelo convolucional y por lo tanto el dato sintético libre de ruido será generado a partir de la ecuación (3.67) desarrollada en la sección 3.6. Sin embargo, como ya fue mencionado en los capítulos previos, en la práctica siempre hay ruido presente en las observaciones, por lo que el dato observado  $\mathbf{d}$  tendrá la forma

$$\mathbf{d} = \mathbf{A}\mathbf{m} + \mathbf{n}, \quad (4.1)$$

donde la matriz  $\mathbf{A}$ , de dimensión  $N_\theta L_d \times nL_r$ , contiene los términos que dependen del ángulo de incidencia de las aproximaciones de orden  $n$  de las ecuaciones de Zoeppritz y los elementos de la ondícula  $\mathbf{w}$ , el vector  $\mathbf{m}$  de longitud  $nL_r$  contiene los parámetros del modelo a estimar, y el vector  $\mathbf{n}$  de longitud  $N_\theta L_d$  representa al ruido.

Ya fue demostrado en la sección 2.6 que el método de inversión por mínimos cuadrados se basa en la hipótesis de que el ruido presente en el dato sigue una distribución gaussiana. Se asumirá por simplicidad que dicho ruido, o al menos un estimado de las propiedades estadísticas del mismo, son conocidos. Si los errores entre el dato observado y el dato estimado no están correlacionados y siguen una distribución gaussiana con media cero y desviación estándar  $\sigma$ , entonces la variable

$$\chi^2 = \sum_{i=1}^{N_\theta} \sum_{h=1}^{L_d} \frac{(d_{ih} - [\mathbf{A}\mathbf{m}]_{ih})^2}{\sigma_{ih}^2} \quad (4.2)$$

sigue una distribución chi-cuadrado con  $N_\theta L_r$  grados de libertad. En este caso, como consideramos que el único ruido presente corresponde al dato observado se tiene que  $\sigma_{ih}^2 = \sigma_n^2$  para todo  $i, h$ , siendo  $\sigma_n^2$  el desvío estándar del ruido. El valor esperado de  $\chi^2$  está dado por (Sacchi et al., 1998)

$$E[\chi^2] = N_\theta L_r \pm \sqrt{2N_\theta L_r}. \quad (4.3)$$

Por lo tanto, el valor esperado  $E_n$  del *misfit* está dado por

$$E_n = E \left[ \sum_{i=1}^{N_\theta} \sum_{h=1}^{L_d} (d_{ih} - [\mathbf{A}\mathbf{m}]_{ih})^2 \right] = \sigma_n^2 E[\chi^2] = \sigma_n^2 (N_\theta L_r \pm \sqrt{2N_\theta L_r}). \quad (4.4)$$

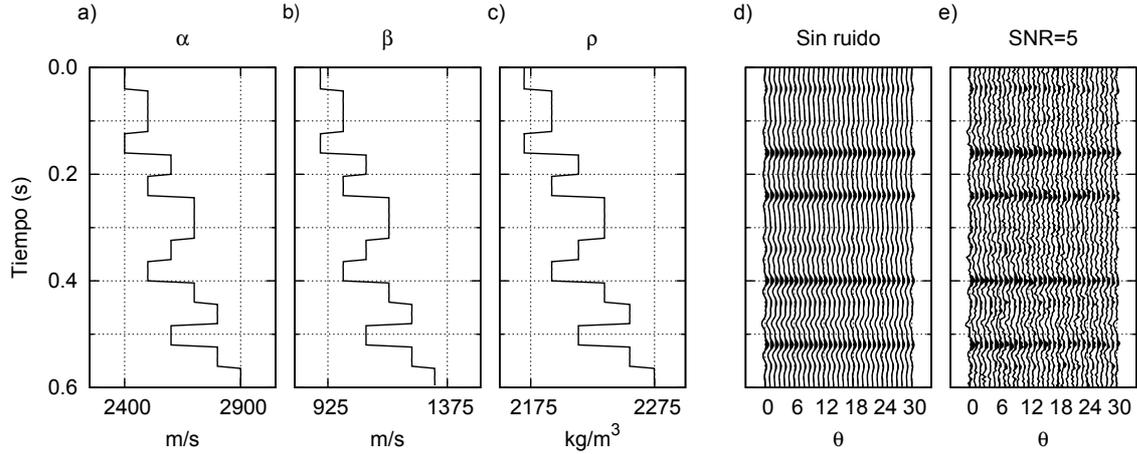
La solución  $\mathbf{m}$  estimada debe ser tal que el *misfit* correspondiente sea próximo al valor  $E_n$ . Esto nos dará un ajuste razonable al dato observado en función de la información que disponemos del nivel de ruido. Por otro lado, conocer el valor  $E_n$  será de utilidad al momento de estimar el parámetro  $\mu$  de la ecuación (2.17) utilizando, por ejemplo, el criterio de discrepancia.

## 4.2. Datos sintéticos 1D

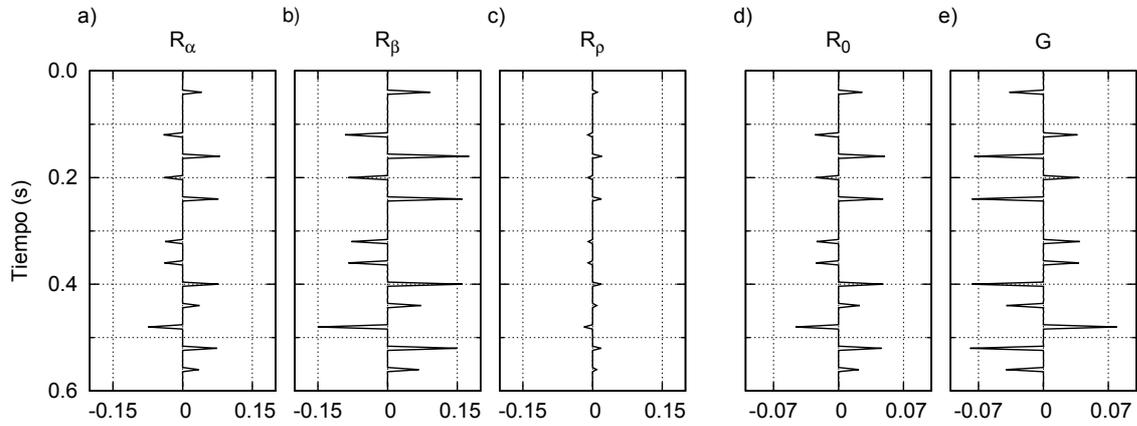
Para realizar las inversiones se generó un *angle-gather* sintético corregido por NMO a partir de los modelos de velocidades y densidad representados en las Figuras 4.1a-c. Este es un modelo estratificado, o tipo *blocky*, del subsuelo, en el que se usaron valores típicos de  $\alpha$ ,  $\beta$  y  $\rho$ , que se relacionan por medio de las ecuaciones desarrolladas por Gardner et al. (1974) y Castagna et al. (1985). El *gather* contiene 12 reflectores y consta de 31 trazas con ángulos de incidencia  $\theta_i \in (0^\circ, 30^\circ)$ . La respuesta AVA se calculó utilizando las ecuaciones de Zoeppritz para cada interfaz. Cada traza se obtuvo luego de convolucionar la serie de coeficientes de reflexión resultante con una ondícula de Ricker de frecuencia central  $f_0 = 20$  Hz, obteniéndose el *angle-gather* de la Figura 4.1d. Al dato se le agregó ruido gaussiano con  $\sigma_n = \text{máx}(\mathbf{d})/5$ , es decir una relación señal-ruido  $\text{SNR} = 5$  (Figura 4.1e). No obstante, la SNR para algunos reflectores individuales es mucho menor. El intervalo de muestreo utilizado fue de 4 ms. En las Figuras 4.2a-c se observan las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  de la aproximación de Aki y Richards correspondientes a dicho modelo de velocidades y densidades. Asimismo en las Figuras 4.2d y 4.2e se observan el *intercept*  $R_0$  y el *gradient*  $G$  de la aproximación de Shuey de dos términos. Como se puede observar, estas reflectividades presentan una simplicidad o *sparseness* muy marcada.

## 4.3. Inversión por mínimos cuadrados

A partir de las ecuaciones (3.65) y (3.66) se armó la matriz  $\mathbf{A}$  de la ecuación (4.1), tanto para la aproximación de Aki y Richards como para la aproximación de Shuey de dos términos, utilizando las expresiones de las funciones  $g_k(\theta_i)$  correspondientes a cada caso (ver Tabla 3.3). Por simplicidad la ondícula  $\mathbf{w}$  utilizada en la inversión fue la misma con la que se generaron los datos sintéticos. Para el caso en el que se usa la aproximación de Aki y Richards se tomó un valor  $\gamma = 0.5$ . Una vez armadas las matrices se procedió a realizar la inversión por mínimos cuadrados y por mínimos cuadrados amortiguados.



**Figura 4.1.** Modelo 1D de velocidades y densidad. Dato sintético de ese modelo, sin ruido y con ruido.

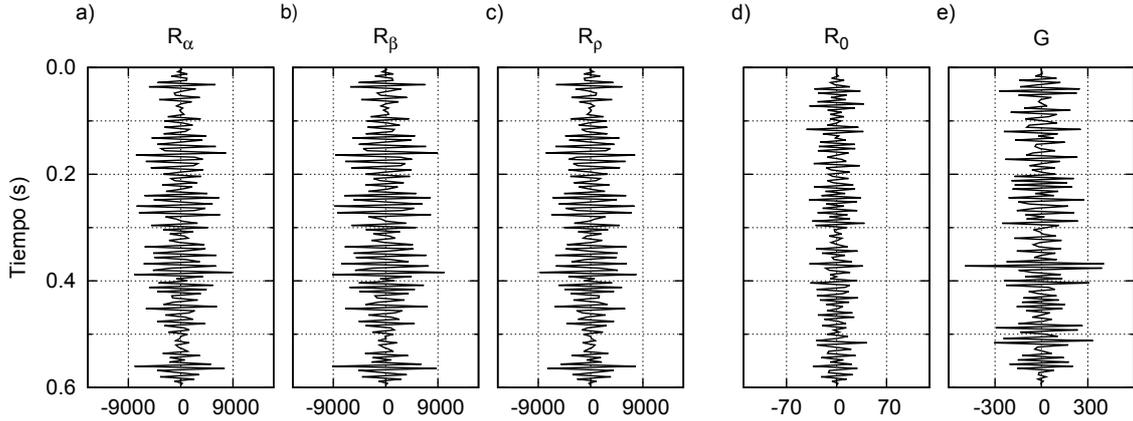


**Figura 4.2.** Reflectividades de Aki y Richards y Shuey correspondientes al modelo de la Figura 4.1.

Las Figuras 4.3a-e muestran los resultados de realizar la inversión por mínimos cuadrados, es decir

$$\hat{\mathbf{m}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{d}. \quad (4.5)$$

Como se puede observar las soluciones son altamente inestables y no guardan similitud alguna con los valores de las Figuras 4.2, tanto cuando se usa la aproximación de Aki y Richards como cuando se usa la aproximación de Shuey. Las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  estimadas (Figuras 4.3a-c) muestran valores exageradamente sobreestimados y lo mismo ocurre, aunque en una escala menor, con el *intercept*  $R_0$  y el *gradient*  $G$  (Figuras 4.3d,e). Este comportamiento de las soluciones estimadas se debe al mal condicionamiento del problema. Como ya fue analizado en la sección 2.3.2 el número de condición es una medida de la sensibilidad de la solución del sistema de ecuaciones lineales a los errores presentes en los datos observados. Si el número de condición es mucho mayor a la unidad entonces el problema está mal codicionado. Para este ejemplo se tiene que,



**Figura 4.3.** Resultados de la inversión por mínimos cuadrados a partir del dato de la Figura 4.1.

cuando  $\mathbf{A}$  representa la matriz del modelo utilizando la aproximación de Aki y Richards, entonces  $\text{cond}(\mathbf{A}^T \mathbf{A}) \propto 10^{15}$ . Asimismo cuando  $\mathbf{A}$  representa la matriz del modelo utilizando la aproximación de Shuey de dos términos, entonces  $\text{cond}(\mathbf{A}^T \mathbf{A}) \propto 10^{12}$ . Estos valores variarán dependiendo del rango de ángulos de incidencia, de la ondícula y del valor de  $\gamma$  elegidos, sin embargo su efecto sobre el número de condición no será tal que el problema planteado esté bien condicionado.

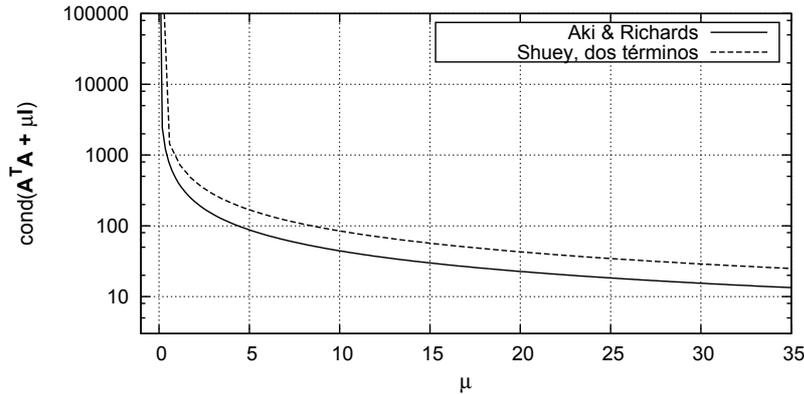
## 4.4. Inversión por mínimos cuadrados amortiguados

El problema de inversión por mínimos cuadrados de datos sísmicos *prestack* está mal condicionado y como consecuencia las soluciones  $\hat{\mathbf{m}}$  son muy inestables. Debido a esto es necesario incorporar alguna clase de regularización para obtener soluciones más estables. Como ya fue analizado en la sección 2.3.2, incluir una regularización cuadrática sobre los parámetros del modelo nos conduce al método de mínimos cuadrados amortiguados, donde los parámetros se obtienen mediante

$$\hat{\mathbf{m}} = (\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})^{-1} \mathbf{A}^T \mathbf{d}. \quad (4.6)$$

En la ecuación anterior,  $\mu$  es el parámetro de amortiguación que tendrá influencia directa en la estabilidad del sistema. La Figura 4.4 muestra el comportamiento del número de condición de la matriz  $(\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})$  en función de  $\mu$ . Allí se puede observar cómo el número de condición aumenta a medida que el  $\mu$  disminuye, tomando valores muy grandes cuando  $\mu \rightarrow 0$ . Contrariamente, utilizar valores grandes de  $\mu$  lleva a una disminución del número de condición. Sin embargo, la elección de dicho parámetro no sólo debe contemplar que el número de condición sea cercano a la unidad sino también que las soluciones obtenidas honren al dato observado.

El parámetro  $\mu$  debe ser tal que balancee el peso o impacto del *misfit* y la regularización cuadrática, y por lo tanto se requiere de una estrategia apropiada para elegirlo. Como ya fue comentado hay varios métodos reportados en la literatura que pueden

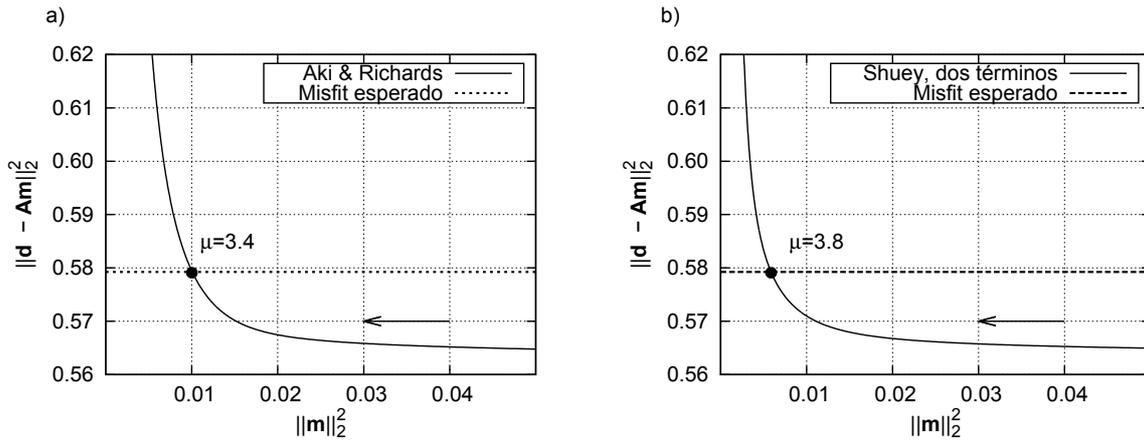


**Figura 4.4.** Número de condición de  $(\mathbf{A}^T \mathbf{A} + \mu \mathbf{I})$  para la aproximación de Shuey y Aki y Richards, en función del parámetro de amortiguación  $\mu$ .

ser utilizados para estimar un valor de  $\mu$ , como por ejemplo el criterio de la curva-L, el *generalized cross-validation criterion* o el principio de discrepancia (Farquharson y Oldenburg, 2004). Para estos ejemplos numéricos estimaremos  $\mu$  usando el principio de discrepancia y la curva-L. Esta estrategia es útil cuando se tiene, como en este caso, una estimación del nivel de ruido en el dato y por lo tanto un estimado del valor esperado del *misfit* por medio de la ecuación (4.4). El principio de discrepancia consiste en realizar la inversión utilizando diferentes valores de  $\mu$ , a partir de los resultados construir la curva-L, y luego elegir el valor óptimo como el mayor valor de  $\mu$  tal que el valor del *misfit* no supere a su valor esperado  $E_n$ . En la Figura 4.5 se pueden ver las curvas del *misfit* versus la regularización cuadrática obtenidas al realizar la inversión sobre los datos sintéticos para diferentes valores de  $\mu$ . Estas curvas se realizaron considerando la aproximación de Aki y Richards (Figura 4.5a) y la aproximación de Shuey de dos términos (Figura 4.5b). En ambas figuras la flecha indica la dirección en la cual aumentan los valores de  $\mu$ , mientras que la línea punteada indica el valor esperado del *misfit*, que en este ejemplo es  $E_n = 0.558$ . El punto ubicado en la intersección de la curva con el  $E_n$  se corresponde con el  $\mu$  óptimo según el criterio de discrepancia, que para el caso de la inversión usando la aproximación de Aki y Richards es  $\mu = 3.4$ , mientras que para la inversión utilizando la aproximación de Shuey es  $\mu = 3.8$ . Como se puede deducir a partir de la Figura 4.4, estos dos valores de  $\mu$  darán números de condición relativamente altos indicando que, aunque el problema ha sido regularizado, las soluciones todavía pueden llegar a ser inestables aunque en menor medida que en el caso de  $\mu = 0$  (Figura 4.3). A pesar de esto, dichos valores del parámetro de amortiguación son aquéllos que nos darán las soluciones que mejor ajustan al dato observado de acuerdo con la información disponible. Efectivamente, de las Figuras 4.5a,b se observa que para valores de  $\mu$  menores el *misfit* está por debajo del valor esperado  $E_n$ , lo que significa que las soluciones estimadas ajustarían ruido. Contrariamente, para valores mayores de  $\mu$  el *misfit* está por encima del valor esperado  $E_n$ , lo que significa que las soluciones estimadas no ajustarían adecuadamente al dato observado.

Una vez estimados los valores óptimos de  $\mu$  por medio del principio de discrepan-

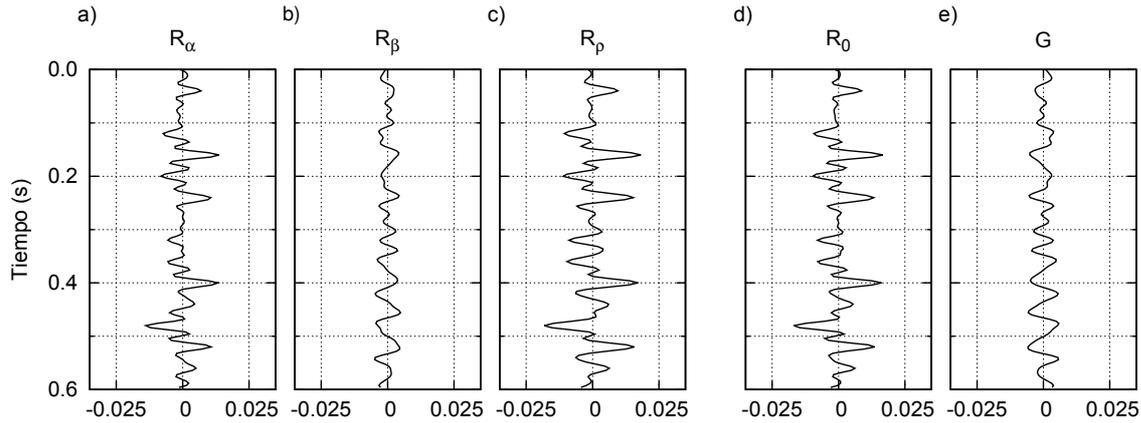
#### 4. Inversión de datos sísmicos *prestack* por mínimos cuadrados



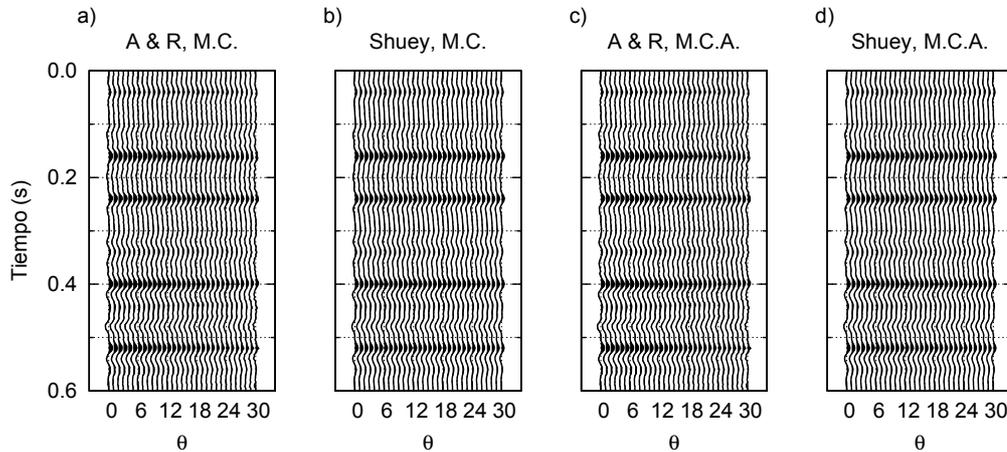
**Figura 4.5.** Parámetros de amortiguación estimados por el principio de discrepancia, para el dato de la Figura 4.1d.

cia se procedió a realizar la inversión por mínimos cuadrados amortiguados (ecuación (4.6)). Las Figuras 4.6a-c muestran las soluciones estimadas utilizando la aproximación de Aki y Richards y las Figuras 4.6d,e las soluciones utilizando la aproximación de Shuey de dos términos. Como se puede observar, a pesar de la regularización cuadrática las soluciones están lejos de ser óptimas. Aunque éstas guardan una mayor similitud con los valores de la Figura 4.2a-e que aquéllas obtenidas mediante el método de mínimos cuadrados sin amortiguar, son demasiado suaves (carecen de resolución) y las amplitudes resultan muy subestimadas. Estas son características propias de la regularización cuadrática que dispersa la energía de los reflectores, promoviendo soluciones suaves. Además de estar subestimados, los atributos estimados a partir de cada aproximación muestran un comportamiento diferente entre sí. Esto se debe a que cada parámetro del modelo influye de manera diferente en el dato observado, y al utilizar un único  $\mu$  durante la inversión algunos valores como  $R_\beta$  y  $G$  se ven más influenciados por la amortiguación que otros, mostrando un comportamiento más suave que el resto. Este inconveniente puede ser superado utilizando diferentes valores de  $\mu$  para cada parámetro, pero para eso se requiere contar con información *a priori* sobre el efecto de cada parámetro en el dato, información que no siempre es fácil de obtener y cuantificar. En el capítulo 7, en el contexto de la inversión *sparse* utilizando normas mixtas, se utilizarán matrices de escala estimadas a partir de datos de pozo para mitigar este inconveniente.

Las Figuras 4.7a-d muestran el dato reconstruido a partir de las soluciones estimadas con ambos métodos. Como se puede apreciar, a pesar de que las soluciones estimadas difieren considerablemente de los valores verdaderos, en todos los casos el dato estimado es casi idéntico al dato sintético sin ruido de la Figura 4.1d. Esto es esperable ya que los métodos utilizados en la inversión parten del hecho de que el *misfit* entre el dato observado y el dato modelado libre de ruido debe ser mínimo en un caso, e igual o parecido al valor esperado del *misfit*  $E_n$  en el otro. Esta figura también es un ejemplo conciso del problema de la ambigüedad en los problemas de inversión, es decir, diferentes parámetros del modelo pueden llevar al mismo dato observado.



**Figura 4.6.** Inversión por mínimos cuadrados amortiguados utilizando los  $\mu$  estimados por medio del criterio de discrepancia (Figura 4.5).



**Figura 4.7.** Datos reconstruidos a partir de las soluciones estimadas, para Aki y Richards (A&R) y Shuey de dos términos, por medio de a) y b) mínimos cuadrados (M.C.) y c) y d) mínimos cuadrados amortiguados (M.C.A.).

## 4.5. Conclusiones

A continuación se enumeran las principales conclusiones obtenidas utilizando las estrategias propuestas sobre dato sintético 1D con ruido:

- Las soluciones estimadas utilizando el método de mínimos cuadrados con y sin amortiguación se apartaron considerablemente de las soluciones esperadas.
- El problema está mal condicionado, presentando un número de condición muy grande. Por lo tanto las soluciones estimadas utilizando el método de mínimos cuadrados sin amortiguar son inestables, dando como resultado parámetros del modelo sin significado físico y con amplitudes muy sobreestimadas.
- Cuando se realiza la inversión utilizando el método de mínimos cuadrados amortiguados el número de condición disminuye. Sin embargo, las soluciones estima-

#### 4. Inversión de datos sísmicos *prestack* por mínimos cuadrados

---

das siguen siendo poco aceptables, ya que la regularización cuadrática dispersa la energía de los reflectores en lugar de concentrarla como se tendría en una solución *sparse*. Como consecuencia las amplitudes de las soluciones están subestimadas significativamente.

- Debido a que los métodos de mínimos cuadrados con y sin amortiguación no son capaces de estimar soluciones aceptables para los modelos de subsuelo planteados es necesario contar con métodos de inversión que permitan obtener soluciones de tipo *sparse* a partir de datos sísmicos *prestack*.

# Capítulo 5

## Inversión *sparse* mediante *Simulated Annealing* y mínimos cuadrados

### 5.1. Introducción

En el capítulo anterior se procedió a realizar la inversión por medio del método de mínimos cuadrados amortiguados, método que incorpora una regularización cuadrática. Sin embargo, a pesar de la regularización, las soluciones estimadas difirieron considerablemente de las esperadas. Esto se debió, en parte, a que se utilizó una norma que promueve suavidad con el fin de obtener soluciones que, por la naturaleza del dato usado, deberían ser *sparse*.

En este capítulo presentamos un método para obtener atributos AVA de tipo *sparse* a partir de datos sísmicos *prestack*. El método propuesto apunta a obtener el menor número de reflectores que, convolucionados con la ondícula, ajustan al dato. Este mismo enfoque fue utilizado por Velis (2008) en la deconvolución de tipo *sparse-spike* de datos *poststack*. La técnica propuesta en este capítulo es una extensión para datos *prestack* del método de deconvolución mencionado. Una de las principales diferencias entre el método propuesto en este capítulo y otros métodos que promueven *sparseness* es que estos últimos estiman una solución de tipo *sparse* considerando que el número de reflectores es igual al número de muestras en la ventana de tiempo utilizada para la inversión. Para ello, la solución se obtiene minimizando una norma que mide cuán *sparse* es la solución. Como alternativa, en el enfoque desarrollado en este capítulo las soluciones *sparse* se obtienen fijando *a priori* el número de reflectores. Con este criterio, intentamos obtener un conjunto de soluciones compuestas por el menor número posible de reflectores, lo que equivale a obtener soluciones sub-óptimas de las que se obtendrían al minimizar la norma  $l_0$  de la solución. Minimizar la norma  $l_0$  es un problema de optimización no convexo muy difícil de resolver que entra en la categoría de problemas NP-hard (*non-deterministic polynomial-time hard*) (Kreimer, 2013). Es decir, es un problema que a medida que el número de parámetros crece, el costo computacional para resolverlos no crece polinomialmente, sino exponencialmente o factorialmente. Estos problemas muchas veces pueden aproximarse reemplazando la norma  $l_0$  por la norma  $l_1$ , lo que lleva a un problema convexo como se verá en el próximo capítulo. Debido a

la alta no linealidad del problema inverso, que incluye la ubicación en tiempo de un número dado de reflectores, se busca la solución usando *Very fast simulated annealing* (VFSA) (Ingber, 1989). Una ventaja de utilizar VFSA es que la incertidumbre de las soluciones puede ser estimada estocásticamente, aprovechando el gran número de soluciones que se ponen a prueba durante el proceso de inversión. El método propuesto fue probado utilizando datos sintéticos prestack 1D y 2D corregidos por *normal-moveout* (NMO), y considerando la aproximación de Shuey de dos términos. Sin embargo la incorporación de otras aproximaciones es inmediata. Mostramos cómo los atributos AVA de tipo *sparse* pueden obtenerse con gran precisión a partir de datos con altos niveles de ruido, incluso cuando el número de reflectores es desconocido. Asimismo, se aplicó el método a datos de campo, obteniéndose soluciones de alta resolución que ajustan al dato observado de manera precisa y que exhiben una buena continuidad lateral. El análisis sugiere que el método es robusto ante la presencia de ruido e incertidumbre en el número de reflectores. También se observa la capacidad del mismo para resolver reflectores cercanos. Asimismo, los resultados son consistentes cuando se invierte un conjunto grande de registros sísmicos.

Al realizar la inversión considerando la aproximación de Shuey de dos términos, se estimarán soluciones *sparse* de los atributos *intercept* y *gradient*. Es bien conocida la dificultad de los métodos de inversión para estimar apropiadamente el *gradient* sin tener conocimiento *a priori* de su respuesta. Aunque tanto el *intercept* como el *gradient* son sensibles al ruido presente en el dato observado, el *gradient* muestra mayor sensibilidad (Whitcombe et al., 2004; Herrmann y Cambois, 2001). Las varianzas de estos atributos dependen parcialmente de la geometría con la cual se adquirió el dato (Downton et al., 2000). El desvío estándar del *gradient* es mayor que el del *intercept* y aumenta con el tiempo de viaje de la onda sísmica, haciendo que su estimación utilizando métodos convencionales sea inexacta y su interpretación cuestionable (Cambois, 1998b). Es más, la incorporación del tercer término de la aproximación de Shuey puede llevar a inexactitudes e inestabilidades aún mayores, debido a que el dato generalmente no contiene información suficiente para determinarlo correctamente, excepto para datos con un rango de ángulos de incidencia muy grande y una relación señal-ruido alta (Cambois, 1998a). Una forma de superar este problema es por medio de la inclusión de una matriz de covarianza o una matriz de escala que provea información de la correlación entre los coeficientes de la aproximación. Esto ha sido estudiado tanto en el contexto de la estadística gaussiana (Buland y Omre, 2003; Downton y Lines, 2001, 2004) como por medio del enfoque bayesiano utilizando funciones de distribución *a priori* derivadas de la distribución multivariable de Cauchy (Alemie y Sacchi, 2011). Sin embargo, estos métodos asumen que uno posee conocimiento previo de la correlación o la matriz de escala que relacione el *intercept* y el *gradient*. En el método propuesto en este capítulo obviamos conocer la relación entre los coeficientes. La exigencia de estimar soluciones *sparse* es utilizada para estimar la posición de los reflectores resolviendo un problema subdeterminado donde es difícil conseguir soluciones estables y fidedignas del *gradient*. Luego, estas posiciones de los reflectores son utilizadas para estimar simultáneamente el *intercept* y el *gradient* resolviendo por mínimos cuadrados un sistema más estable y sobredeterminado. Un ventaja clave del método propuesto es que estabiliza la inversión del *gradient* sin necesidad de incluir una matriz de covarianza o de escala. Esto es importante debido a que en general la matriz de covarianza sólo es conocida, aproxima-

damente, cuando se tiene información de pozo o un buen estimado del comportamiento AVA en el área de estudio (Downton, 2005).

## 5.2. Teoría

Una solución de tipo *sparse* es aquella que nos permite representar la traza sísmica como la convolución de una serie poco densa de reflectores con una ondícula. Dicha serie poco densa será un vector  $\mathbf{r}_s$  de dimensión  $L_r$  en el cual sólo habrá  $L_s$  elementos no nulos, con  $L_s \ll L_r$ , ubicados en las muestras  $\boldsymbol{\tau} = (\tau_1, \dots, \tau_{L_s})$ . Partiendo de la serie de coeficientes de reflexión  $\mathbf{r}(\theta_i)$  vista en la sección 3.6, una serie *sparse*  $\mathbf{r}_s(\theta_i)$  se puede expresar como

$$\mathbf{r}_s(\theta_i) = \mathbf{S}\mathbf{r}(\theta_i), \quad (5.1)$$

donde  $\mathbf{S}$  es una matriz diagonal de dimensión  $L_r \times L_r$  en la cual los elementos de la diagonal están dados por

$$\{\text{diag}(\mathbf{S})\}_l = \delta_{l-\tau_j}, \quad j = 1, \dots, L_s \quad (5.2)$$

para  $l = 1, \dots, L_r$ , siendo  $\delta_t$  la función impulso unidad.

Combinando la ecuación (5.1) con las ecuaciones (3.62) y (3.63) la traza correspondiente al ángulo de incidencia  $\theta_i$  se puede expresar como

$$\mathbf{d}(\theta_i) = \mathbf{A}_s(\theta_i)\mathbf{m}_s, \quad i = 1, \dots, N_\theta \quad (5.3)$$

donde  $\mathbf{m}_s = (m_{11}, \dots, m_{L_s 1}, \dots, m_{1n}, \dots, m_{L_s n})^T$  y  $\mathbf{A}_s(\theta_i)$  es una matriz aumentada de dimensión  $L_d \times L_s n$  tal que

$$\mathbf{A}_s(\theta_i) = [\mathbf{A}_{s1}(\theta_i) | \dots | \mathbf{A}_{sn}(\theta_i)]. \quad (5.4)$$

Aquí, las matrices  $\mathbf{A}_{sk}(\theta_i)$  con  $k = 1, \dots, n$  son sub-matrices de dimensión  $L_d \times L_s$ , formadas por las columnas  $l = \tau_1, \dots, \tau_{L_s}$  de las matrices  $\mathbf{A}_k(\theta_i)$  dadas en la ecuación (3.65). A partir de esto, los elementos de las matrices  $\mathbf{A}_{sk}(\theta_i)$  están dados por

$$\{\mathbf{A}_{sk}(\theta_i)\}_{hj} = g_k(\theta_i)w_{h-\tau_j+1}, \quad (5.5)$$

para  $h = 1, \dots, L_d$  y  $j = 1, \dots, L_s$ , y donde  $w$  son los elementos de la ondícula  $\mathbf{w}$ . De esta última ecuación se ve que la matriz que resuelve el problema directo depende ahora de la ondícula  $\mathbf{w}$ , de las funciones  $g_k(\theta_i)$ , pero también de  $L_s$  y de los valores que tomen los  $\tau_j$ .

Al igual que con la ecuación (3.64), los  $N_\theta$  sistemas de ecuaciones dados por la ecuación (5.3) pueden ser re-organizados en un único sistema

$$\mathbf{d} = \mathbf{A}_s\mathbf{m}_s, \quad (5.6)$$

donde  $\mathbf{A}_s$  es una matriz columna por bloques, donde cada bloque está dado por  $\mathbf{A}_s(\theta_i)$ , y  $\mathbf{d}$  es un vector columna por bloques, donde cada bloque está dado por  $\mathbf{d}(\theta_i)$ , para  $i = 1, \dots, N_\theta$ .

El vector de los parámetros del modelo óptimo puede estimarse minimizando la discrepancia entre el dato observado y el dato modelado, para lo cual se define la función de costo

$$J(\mathbf{m}_s, \boldsymbol{\tau}) = \|\mathbf{d} - \mathbf{A}_s \mathbf{m}_s\|_2^2. \quad (5.7)$$

A diferencia del problema analizado en la sección 4.3, donde se utilizó mínimos cuadrados para resolver un problema de inversión lineal, la función de costo dada por la ecuación (5.7) es altamente no lineal y multimodal debido a que la ecuación (5.6) representa un sistema lineal en términos de las incógnitas  $\mathbf{m}_s$ , pero no lineal en términos de los  $\tau_j$ . Como consecuencia, para poder minimizar la función de costo es necesaria la utilización de algún método de optimización global como los brevemente descritos en la sección 2.5.2.

En la práctica minimizaremos  $J(\mathbf{m}_s, \boldsymbol{\tau})$  respecto a  $\boldsymbol{\tau}$  utilizando (VFSA). En cada iteración del VFSA, dados los valores estimados del vector  $\boldsymbol{\tau}$ , el vector óptimo  $\hat{\mathbf{m}}$  es estimado resolviendo el problema de mínimos cuadrados dado por la ecuación (5.7), lo que conduce a

$$\hat{\mathbf{m}}_s = (\mathbf{A}_s^T \mathbf{A}_s)^{-1} \mathbf{A}_s^T \mathbf{d}. \quad (5.8)$$

El proceso es iterativo. A partir de valores  $\tau_j$  dados por el VFSA se construye la matriz  $\mathbf{A}_s$  y se calcula la solución óptima  $\hat{\mathbf{m}}_s$  utilizando la ecuación (5.8). Por medio de la ecuación (5.7) se calcula el valor de la función de costo. Dependiendo del valor que toma la función de costo, VFSA decide si aceptar, o no, la solución y propone nuevos valores de  $\tau_j$ . Con estos nuevos valores se vuelve a estimar una nueva solución óptima  $\hat{\mathbf{m}}_s$  y se vuelve a calcular el valor de la función de costo. El proceso iterativo termina cuando se alcanza un valor predeterminado de la función de costo, un número máximo de iteraciones o algún otro criterio de corte.

Una ventaja del método de inversión propuesto es que es posible estimar la incertidumbre de las soluciones. En la práctica esto se consigue estimando muchas soluciones diferentes, utilizando para cada una diferentes semillas para la generación de los números pseudoaleatorios que son utilizados por VFSA durante el proceso de optimización. Este proceso permite explorar estocásticamente el espacio de los parámetros del modelo, y de esta forma estimar no sólo una solución sino una familia de soluciones que honran al dato observado y a las restricciones impuestas. Usar el valor medio de las múltiples soluciones como resultado final, en lugar de una solución individual, permitirá disminuir el efecto de *spikes* espurios que frecuentemente aparecen en cada solución individual. Estos *spikes* espurios se deben al ruido presente en el dato observado, a una incorrecta elección del valor de  $L_s$  o a que no siempre el algoritmo converge al mínimo absoluto de la función de costo en el máximo número de iteraciones prefijado.

### 5.3. Sobre *Very fast simulated annealing*

A continuación se dará una breve descripción sobre el *Simulated annealing* (SA) y el VFSA. Una explicación más detallada sobre el funcionamiento de los mismos puede ser estudiada en el Apéndice A.

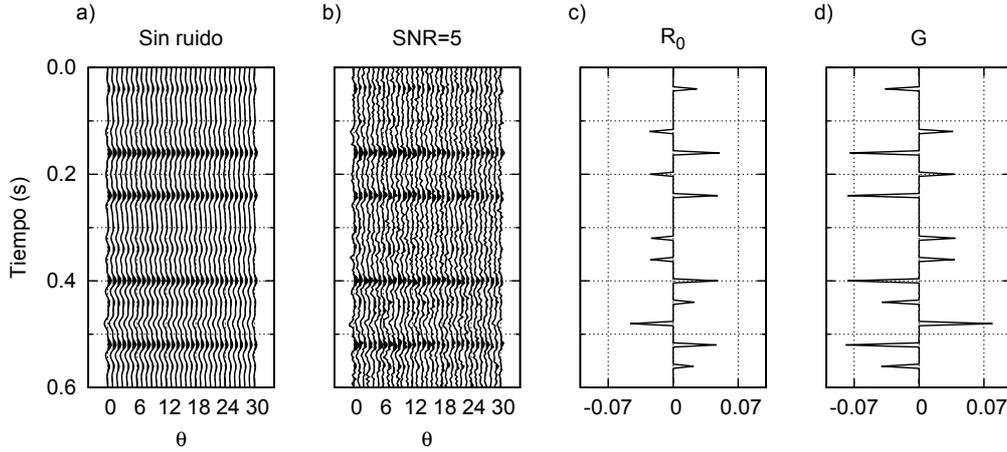
SA es un poderoso algoritmo de optimización no lineal cuyo objetivo es encontrar el mínimo global de una función independientemente del modelo inicial y que no requiere el cálculo de derivadas ni gradientes. El nombre proviene del proceso de recocido de metales, una técnica metalúrgica que consiste en calentar y luego enfriar el material de manera controlada para variar sus propiedades físicas. La evolución de las propiedades de un metal sometido al proceso de recocido fue estudiada por Metropolis et al. (1953) a través de métodos Monte Carlo. Kirkpatrick et al. (1983) generalizaron el concepto y lo aplicaron a problemas de optimización no lineal. Las incógnitas (el modelo) juegan el papel de las partículas en el metal y el estado de energía del sistema está representado a través de una función de costo. El proceso de simulación es iterativo. En cada paso la configuración del modelo (posición de las partículas) se selecciona a través de una distribución de probabilidad que depende de un parámetro de control llamado temperatura. La temperatura decrece gradualmente durante el proceso de optimización siguiendo un esquema de enfriado preestablecido. Cuando la temperatura es alta, el espacio de donde se escogen los parámetros del modelo es explorado de forma aproximadamente uniforme. A bajas temperaturas, se escoge preferentemente un modelo cuyo valor de energía (función de costo) es más bajo. Sin embargo, si el sistema es enfriado muy bruscamente, las partículas del material (incógnitas del modelo) no alcanzarán la configuración de mínima energía (mínimo global de la función), por lo que el esquema de enfriamiento es clave para un correcto funcionamiento del proceso. El modelo es aceptado o rechazado en cada iteración según el criterio de Metropolis. Este criterio establece que si el valor de la función de costo disminuye, el nuevo modelo es aceptado sin condiciones. En cambio, si el valor de la función de costo aumenta, el nuevo modelo es aceptado con una probabilidad no nula que depende de la energía y de la temperatura. A mayor temperatura esta probabilidad de aceptación es mayor. A menor temperatura, por otro lado, la probabilidad de aceptación es menor, pero siempre mayor que cero. Esta estrategia permite aceptar soluciones que producen un incremento en la función costo evitando así quedar atrapado en mínimos locales. Se consigue la convergencia cuando, a bajas temperaturas, no se observa un decrecimiento significativo en el valor de la función de costo.

VFSA es una modificación del SA clásico propuesta por Ingber (1989). Las principales diferencias residen en el esquema de enfriado y en la distribución de probabilidad utilizada para la elección de los modelos. VFSA utiliza una distribución del tipo Cauchy de cola larga que permite que el espacio de los modelos sea explorado más eficientemente que utilizando una distribución gaussiana o uniforme como hacen otros algoritmos de SA convencionales. Al mismo tiempo, eso permite utilizar un esquema de enfriado más veloz, lo que a la vez, permite acelerar la convergencia sin limitar la habilidad del método para escapar a los mínimos locales.

## 5.4. Ejemplos

En esta sección se probará el método propuesto sobre datos sintéticos 1D y 2D y sobre datos de campo. En todos los casos las inversiones serán realizadas considerando la aproximación de Shuey de dos términos:

$$R_{PP}(\theta_i) = R_0 + G \sin^2 \theta_i. \quad (5.9)$$



**Figura 5.1.** a) Dato sintético 1D sin ruido, b) dato con ruido (SNR=5), c) *intercept*  $R_0$  correspondiente, d) *gradient*  $G$  correspondiente.

Es decir, se estimarán soluciones *sparse-spike* del *intercept*  $R_0$  y del *gradient*  $G$ . Las funciones  $g_k(\theta_i)$  de la matriz  $\mathbf{A}_s$  en este caso serán  $g_1(\theta_i) = 1$  y  $g_2(\theta_i) = \sin^2 \theta_i$  (ver Tabla 3.3).

#### 5.4.1. Datos sintéticos 1D

En este ejemplo numérico analizamos el comportamiento del método propuesto utilizando datos sísmicos *prestack* 1D sintéticos. El dato es el mismo que utilizamos para probar los métodos de mínimos cuadrados y mínimos cuadrados amortiguados del capítulo 4. El *gather* contiene 12 reflectores y consta de 31 trazas con ángulos de incidencia  $\theta_i \in (0^\circ, 30^\circ)$ . La respuesta AVA se calculó utilizando las ecuaciones de Zoeppritz, se utilizó una ondícula de Ricker de frecuencia central  $f_0 = 20$  Hz, se agregó ruido gaussiano con una relación señal-ruido SNR = 5, y el intervalo de muestreo utilizado fue de 4 ms. La Figura 5.1 muestran los gathers sin ruido, con ruido, y el *intercept*  $R_0$  y *gradient*  $G$  correspondientes.

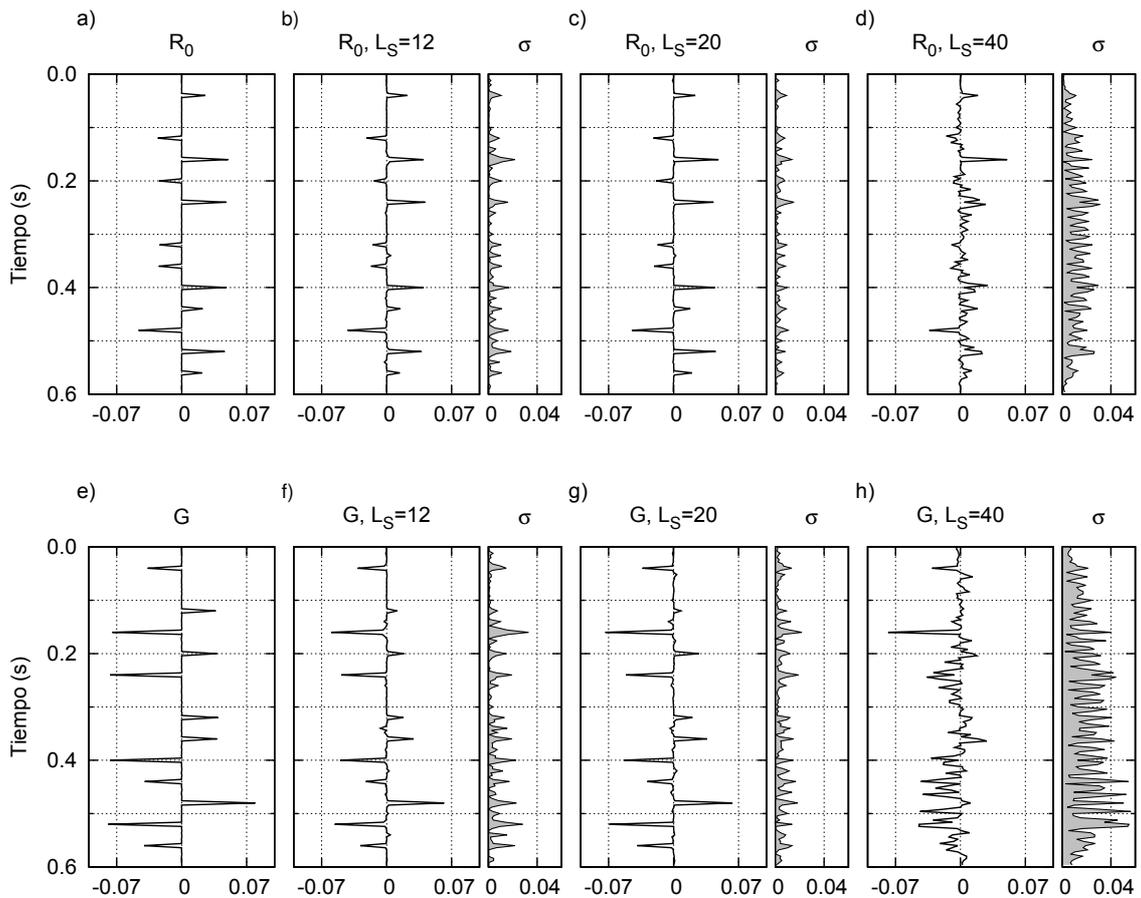
Para estudiar el comportamiento del método de inversión ante la incertidumbre en el número  $L_s$  de reflectores a buscar, se realizaron pruebas con  $L_s = 12$  (el número correcto),  $L_s = 20$  y  $L_s = 40$ . En cada prueba, por propósitos estadísticos, se procedió a realizar 100 inversiones donde en cada inversión se utilizó una semilla diferente para el generador de números pseudoaleatorios utilizados por el VFSA. Por simplicidad la ondícula utilizada para construir la matriz  $\mathbf{A}_s$  fue la misma que se utilizó para crear los datos. Como criterio de corte del VFSA se estableció que la inversión termine cuando el valor de la función de costo dado por la ecuación (5.7) fuera menor al valor esperado del misfit  $E_n$ , o cuando se alcanzaran las 10000 iteraciones. Como solución final se tomó el promedio de las 100 soluciones. Nótese que si bien el dato se creó utilizando las ecuaciones de Zoeppritz, la inversión se realizó utilizando la aproximación Shuey de dos términos.

La Figura 5.2 muestra los valores del *intercept*  $R_0$  y *gradient*  $G$  obtenidos a partir del modelo de velocidad y los estimados, así como la desviación estándar, con  $L_s = 12$ ,  $L_s = 20$  y  $L_s = 40$ . Se puede observar que para  $L_s = 12$  y  $L_s = 20$  el método brin-

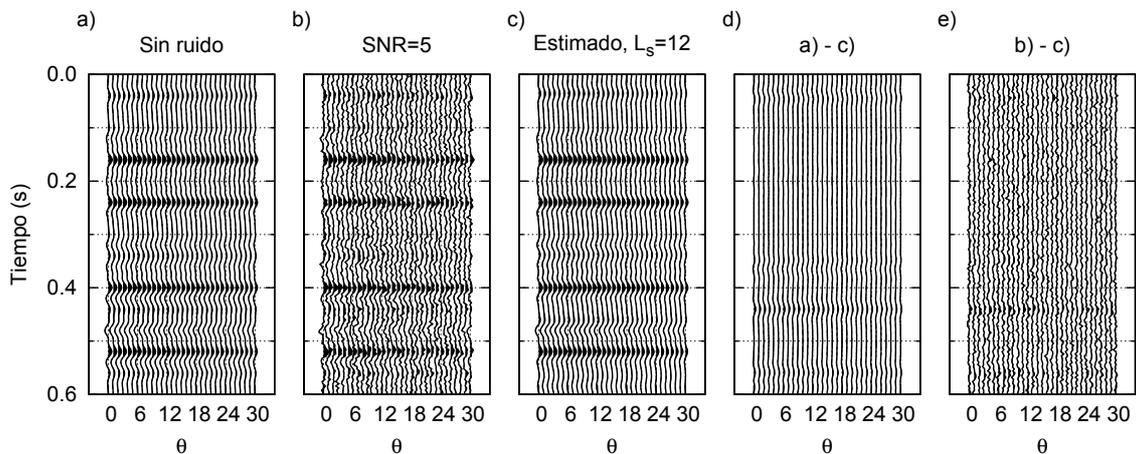
da muy buenos resultados incluso en presencia de un nivel de ruido significativo. Los parámetros de la aproximación resultan estimados con precisión, en tanto que los relativamente bajos valores de las desviaciones estándar resaltan la consistencia de las soluciones. Para el caso de  $L_s = 40$  la solución comienza a ser inestable. Este comportamiento es esperable, ya que en el caso extremo en el cual  $L_s \rightarrow L_r$  el método propuesto será equivalente a la inversión por mínimos cuadrados ya analizada en el capítulo 4. Es importante destacar que son esperables pequeñas diferencias entre los parámetros observados y estimados, no sólo debido a la presencia de ruido, sino también porque los datos fueron creados modelando la respuesta AVA por medio de las ecuaciones de Zoeppritz y la inversión fue realizada utilizando la aproximación de Shuey de dos términos. Además, al ser la solución un valor medio de 100 inversiones diferentes, aquellos reflectores que no hayan sido ubicados en la misma muestra las 100 veces comenzarán a mostrar amplitudes subestimadas. Nótese los impulsos espurios de pequeña amplitud en el caso de  $L_s = 20$  para el *gradient* (Figura 5.2g). Estos reflectores espurios son consecuencia de realizar la inversión buscando un número de reflectores mayor al que realmente posee el dato. En estos casos el algoritmo procede a ubicar estos reflectores adicionales en aquellos tiempos donde produzcan una disminución del valor de la función de costo. En el caso de trabajar sobre datos sin ruido, dichos reflectores espurios tendrán amplitud prácticamente nula y su ubicación en tiempo será consecuencia del comportamiento aleatorio del VFSA. En el caso de trabajar sobre datos que poseen ruido aleatorio, estos reflectores poseen una amplitud pequeña, y su ubicación en tiempo dependerá en parte de las características del ruido, siendo más probable que sean ubicados en aquellos lugares donde la amplitud del ruido sea mayor.

En la Figura 5.3a-e se puede ver la comparación entre el dato observado y el dato estimado con  $L_s = 12$ . El panel a) de la figura muestra el dato observado libre de ruido, mientras que en el panel b) se puede observar el dato con ruido sobre el cual se realizó la inversión. El panel c) muestra el dato estimado para  $L_s = 12$ , construido a partir de las soluciones medias del *intercept* y el *gradient* (Figuras 5.2b,f). Los paneles d) y e) muestran la diferencia entre el dato estimado y el dato observado con y sin ruido. Como se puede apreciar, el dato estimado honra el dato observado con mucha precisión. El residuo observado en los paneles d) y e) es casi nulo en el caso del primero y sólo ruido aleatorio en el caso del segundo, a excepción de alguna señal débil que puede ser observada a partir de los 0.4 s.

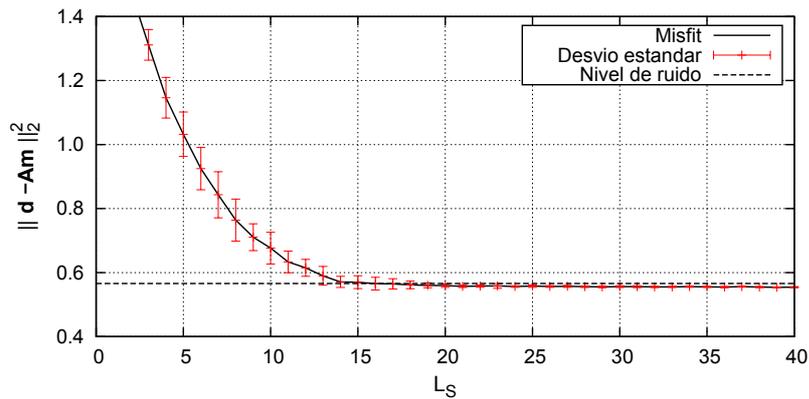
La Figura 5.4 muestra el comportamiento de la función de costo con respecto al número de reflectores  $L_s$  buscados durante la inversión. La curva muestra el valor medio del *misfit* calculado a partir de 50 realizaciones de la inversión para valores de  $L_s = 1, \dots, 40$ . También se puede apreciar el desvío estándar calculado para cada valor de  $L_s$  elegido. Como es esperable, el valor medio del *misfit* decrece a medida que el valor de  $L_s$  se incrementa. Sin embargo, la curva muestra una tendencia aproximadamente asintótica al nivel del ruido cuando el número de reflectores buscados es más grande que el número verdadero, que en este caso es igual a 12. Esto sucede debido a que para valores grandes todos los reflectores ya han sido encontrados e incrementar dicha cantidad sólo contribuye a ajustar ruido no correlacionado, lo que no produce una disminución significativa del valor de la función de costo. También se puede observar que el desvío estándar de las 50 soluciones aumenta para valores de  $L_s$  menores que el número correcto de reflectores. Esto se debe a que cuando el número  $L_s$  es menor



**Figura 5.2.** *Intercept*  $R_0$  y *gradient*  $G$  estimado con VFSA a) y e)  $R_0$  y  $G$  verdadero,  $R_0$  y  $G$  estimado medio y su correspondiente devío estándar usando b) y f)  $L_s = 12$ , c) y g)  $L_s = 20$  y d) y h)  $L_s = 40$ .



**Figura 5.3.** a) Dato libre de ruido, b) dato con ruido (SNR=5), c) dato estimado con  $L_s = 12$ , diferencia entre el dato estimado y d) el dato libre de ruido, y e) el dato con ruido.



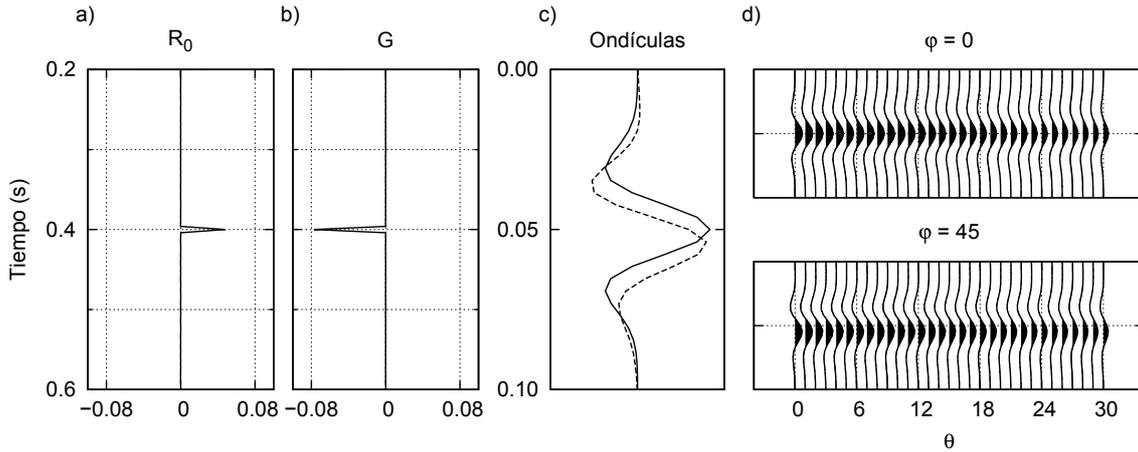
**Figura 5.4.** *Misfit* vs número de reflectores  $L_s$  (línea sólida) y nivel de ruido (línea punteada) estimados a partir del dato de la Figura 5.1. Las barras indican el desvío estándar del misfit para las 50 realizaciones.

que el número verdadero es poco probable que el VFSA converja siempre a la misma solución, dado un número fijo de iteraciones, en cada una de las 50 realizaciones. Como consecuencia, en cada realización se tiene un valor del *misfit* estimado diferente, haciendo que aumente el desvío estándar correspondiente. Contrariamente, cuando  $L_s$  es mayor que el número verdadero las probabilidades de que el VFSA converja a la misma solución en cada realización aumentan, disminuyendo el desvío estándar. Esta curva puede ser utilizada como guía para seleccionar el número apropiado de reflectores a buscar eligiendo, por ejemplo, el valor de  $L_s$  a partir de observar la zona en la cual la curva tiene máxima curvatura o comienza a aplanarse. En este sentido, cuando se utiliza un número  $L_s$  mayor al correcto los resultados muestran que los reflectores “extra” son ubicados a diferentes tiempos según la inversión, pero en todos los casos sus amplitudes son cercanas a cero y no son visibles en el dato reconstruido. Por ello la selección de este parámetro no es crítica dentro de cierto rango, más allá del incremento en el costo computacional que implica tomar valores grandes de  $L_s$ .

### Incertidumbre en la ondícula

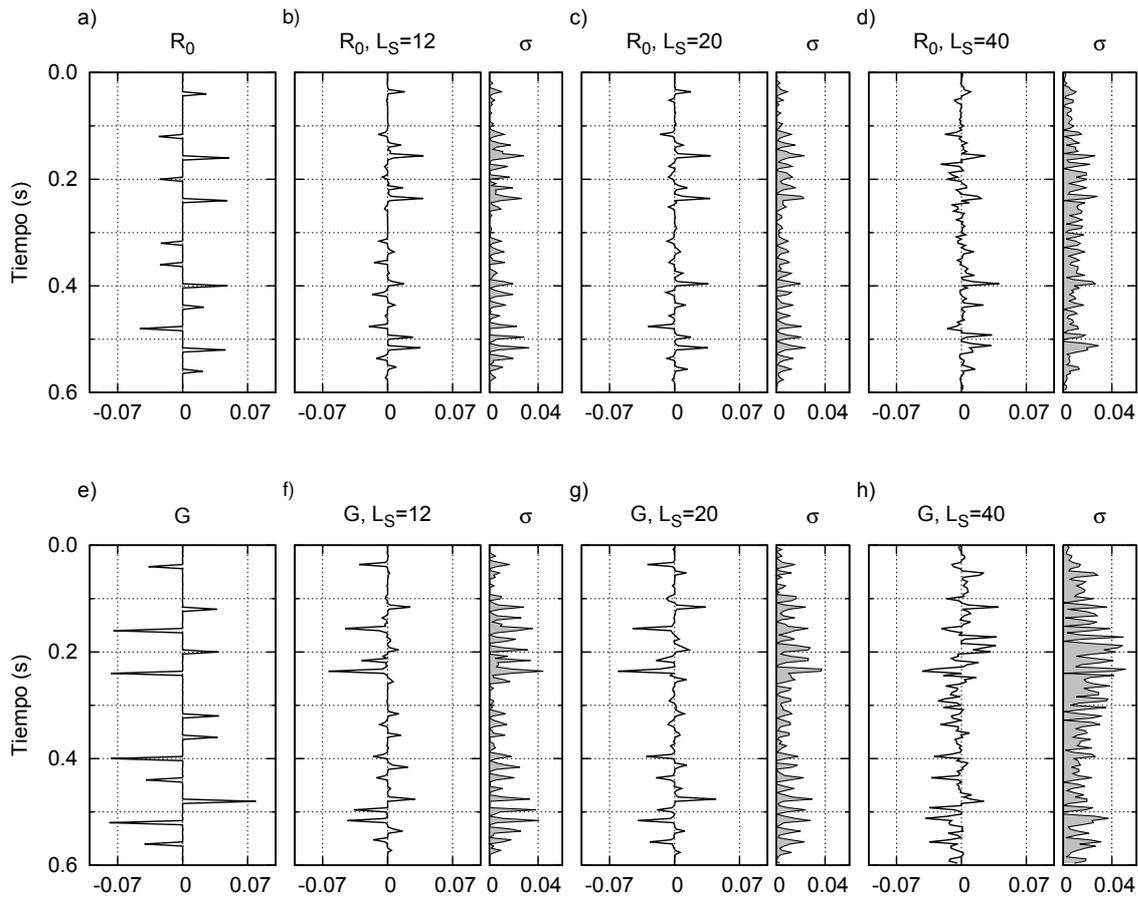
A pesar de que existen varios métodos de estimación de ondícula que proveen resultados razonables, incluso para ondículas que no son de fase mínima (Ulrych et al., 1995; Velis y Ulrych, 1996; Porsani y Ursin, 2000; Routh et al., 2003, p.ej.), generalmente la calidad de la ondícula estimada depende del dato. Ignorar las incertidumbres en la ondícula utilizada en la inversión llevará a soluciones poco confiables. En este sentido, el conocimiento de la fase de la ondícula es un primer paso crítico para llevar adelante el proceso de inversión que pretendemos realizar (Levy y Oldenburg, 1982).

En la Figura 5.5d se puede observar el aspecto del reflector ubicado a los 0.4 s del dato sintético (Figuras 5.5a,b) cuando es creado utilizando ondículas de Ricker con  $f_0 = 20$  Hz sin rotación de fase (línea sólida en la Figura 5.5c), y con una rotación de fase  $\varphi = 45^\circ$  (línea discontinua en la Figura 5.5c). Como se puede apreciar la diferencia entre las ondículas no es considerable y lo mismo ocurre para el dato generado con



**Figura 5.5.** a) y b) *intercept* y *gradient* correspondientes al reflector ubicado a los 0.4s del modelo de la Figura 5.1, c) ondículas de Ricker de  $f_0 = 20$  Hz de fase cero (línea sólida) y con rotación de fase  $\varphi = 45^\circ$  (línea punteada), d) dato generado a partir del *intercept* y *gradient* utilizando dichas ondículas.

éstas. Sin embargo el efecto sobre los parámetros estimados si se utiliza una ondícula incorrecta al realizar la inversión puede ser muy significativa. A modo de ejemplo la Figura (5.6) muestra los resultados del promedio de 100 inversiones llevadas a cabo sobre el dato con  $SRN=5$ , tomando  $L_s = 12$ ,  $L_s = 20$  y  $L_s = 40$  y utilizando la ondícula de Ricker con rotación de fase  $\varphi = 45^\circ$  cuando el dato fue generado con la ondícula de fase cero. Como se puede apreciar, para  $L_s = 12$  y  $L_s = 20$  los resultados están lejos de ser satisfactorios como sí lo fueron utilizando la ondícula correcta (Figura 5.2). Se observa una considerable subestimación de las amplitudes y un aumento de la cantidad de *spikes* espurios. Asimismo se observa un aumento en la incertidumbre de las soluciones estimadas. Esto se debe a que cuando la ondícula es incorrecta y el número  $L_s$  es mayor al número verdadero VFSA utilizará más de un *spike* por reflector para ajustar correctamente el dato observado y de esa forma minimizar el *misfit*. Por otro lado, si se utiliza un valor de  $L_s$  igual al número exacto de reflectores, utilizando una ondícula incorrecta nunca se podrá encontrar una solución tal que el *misfit* calculado sea menor o igual al valor esperado  $E_n$ . Para el caso en el cual  $L_s = 40$ , al igual que al utilizar la ondícula correcta, las soluciones se muestran muy inestables. Es importante destacar que incluso al usar una ondícula incorrecta el dato puede ser bien ajustado si se utiliza un valor de  $L_s$  suficientemente grande, hecho que ratifica la ambigüedad del problema inverso. La Figura 5.7 muestra la comparación entre los datos reconstruidos para los parámetros estimados con  $L_s = 12$  y  $L_s = 20$  y el dato libre de ruido. Como se puede apreciar en las Figuras 5.7b,c, y como era esperable al utilizar una ondícula incorrecta, si  $L_s = 12$  el método fue incapaz de ajustar correctamente el dato observado. Los reflectores ubicados a los 0.24 s y a los 0.4 s no fueron recuperados correctamente, es decir no se llegó a encontrar una solución tal que el *misfit* sea cercano al valor esperado  $E_n$ . En cambio, como se puede apreciar en las Figuras 5.7d,e, cuando se utiliza  $L_s = 20$  el dato estimado comienza a parecerse más al dato libre de ruido a pesar de que los parámetros estimados están lejos de ser óptimos. Esto demuestra que el uso de un número de reflectores  $L_s$  mayor al correcto incrementa la ambigüedad del problema.



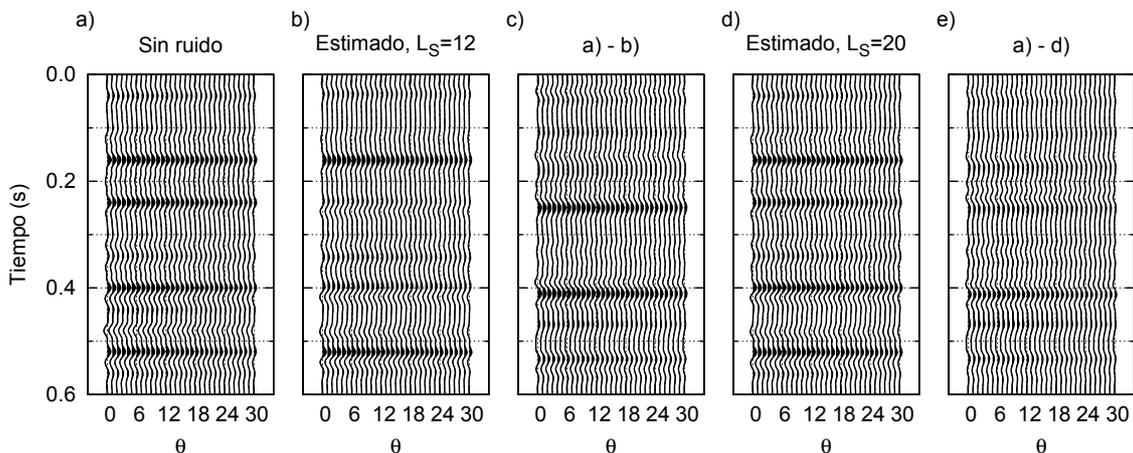
**Figura 5.6.** Intercept  $R_0$  y gradient  $G$  estimado con VFSA utilizando la ondícula con  $\varphi = 45^\circ$  a) y e)  $R_0$  y  $G$  verdadero,  $R_0$  y  $G$  estimado medio y su correspondiente devío estándar usando b) y f)  $L_s = 12$ , c) y g)  $L_s = 20$  y d) y h)  $L_s = 40$ .

Existen muchas soluciones *sparse* que pueden ajustar al dato observado incluso con la ondícula incorrecta.

Como estrategia para aliviar los problemas inherentes a utilizar una ondícula inexacta se puede incorporar, como una incógnita extra durante el proceso de inversión, una rotación de fase  $\varphi$  a la misma. Al realizar esta incorporación, la función de costo a minimizar dada por la ecuación (5.7) será ahora

$$J(\mathbf{m}_s, \tau, \varphi) = \|\mathbf{d} - \mathbf{A}_s \mathbf{m}_s\|_2^2. \quad (5.10)$$

Sin embargo es importante tener en cuenta que incorporar la rotación  $\varphi$  al proceso de inversión incrementa aún más la ambigüedad del problema a resolver. En particular si el número de reflectores  $L_s$  a buscar es mayor que el correcto, existen muchas soluciones *sparse* que pueden ajustar al dato observado con igual exactitud para diferentes valores de la rotación de fase  $\varphi$ . Debido a esto la estrategia propuesta sólo deberá ser aplicada con fines correctivos, en caso de ya contar con una estimación relativamente buena de la ondícula y del número de reflectores  $L_s$ .



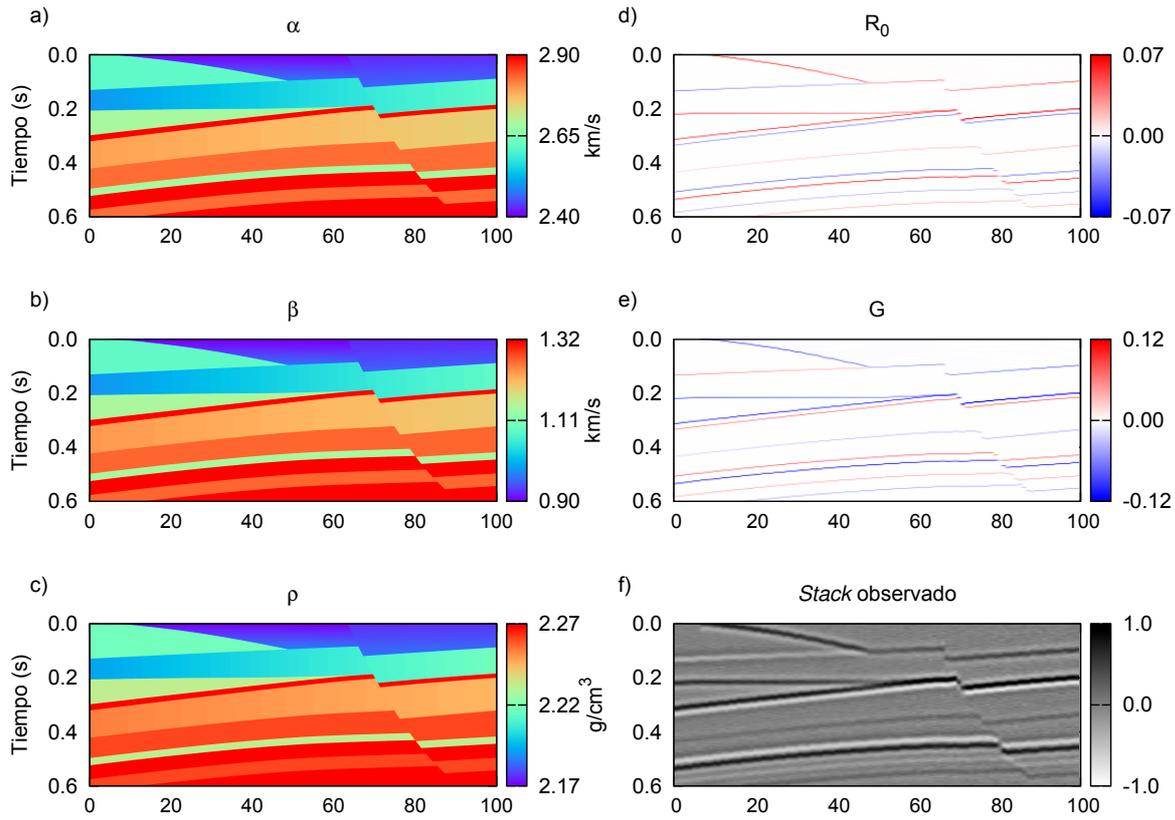
**Figura 5.7.** a) Dato observado libre de ruido, b) dato estimado utilizando la ondícula con  $\varphi = 45^\circ$  ( $L_s = 12$ ), c) diferencia entre el dato estimado con  $L_s = 12$  y el dato observado, d) dato estimado con  $L_s = 20$ , e) diferencia entre el dato estimado con  $L_s = 20$  y el dato observado.

#### 5.4.2. Datos sintéticos 2D

En este ejemplo se analiza la capacidad del método para preservar la continuidad lateral de los eventos horizontales en un modelo bidimensional. Para esto fue construido un modelo 2D de velocidad y densidad (Figura 5.8a-c). A partir de este modelo se crearon 100 *angle-gathers* corregidos por NMO, con un intervalo de muestreo de 4 ms. Al dato se le agregó ruido gaussiano con  $\text{SNR} = 5$ . Para generar el dato se utilizó el modelo convolucional y una ondícula de Ricker con  $f_0 = 20$  Hz. Como en el ejemplo 1D, la respuesta AVA se modeló utilizando las ecuaciones de Zoeppritz. Las Figuras 5.8d,e muestran el *intercept* y *gradient* correspondientes al modelo. La Figura 5.8f muestra la sección sísmica (*stack*) del dato observado. Para cada gather, con fines puramente estadísticos, se realizaron 50 inversiones en cada caso con  $L_s = 15$ . Como criterio de corte del VFSA se escogió que el proceso iterativo termine cuando el *misfit* de la solución estimada fuera menor al nivel de ruido o cuando se alcanzaran las 10000 iteraciones.

Las Figuras 5.9a,b muestran el *intercept*  $R_0$  y el *gradient*  $G$  estimados luego de promediar las 50 inversiones. Se aprecia que el método ofrece buenos resultados, conservando la continuidad lateral e invirtiendo satisfactoriamente los parámetros de la aproximación de Shuey. En las Figuras 5.9c,d se puede observar los desvíos estándar de las 50 realizaciones. Como se puede apreciar, los valores, que están dentro de lo aceptable considerando el nivel de ruido en el dato, son mayores en las inmediaciones de los reflectores para el *intercept*. Sin embargo se observa que la incertidumbre de los resultados para el *gradient* no se limita sólo a las cercanías de los reflectores, sino que está distribuida en toda la sección. Esto se debe a que, como ya fue establecido al principio de esta misma sección, el ruido tiene mayor influencia en el *gradient* que en el *intercept*. En las Figuras 5.9e,f se ven el *stack* observado y el estimado, la gran similitud de ambas muestra que el método fue capaz de estimar soluciones *sparse* que honran al dato observado.

En la Figura 5.10 se observa con mayor detalle el *intercept* y el *stack* verdadero, y

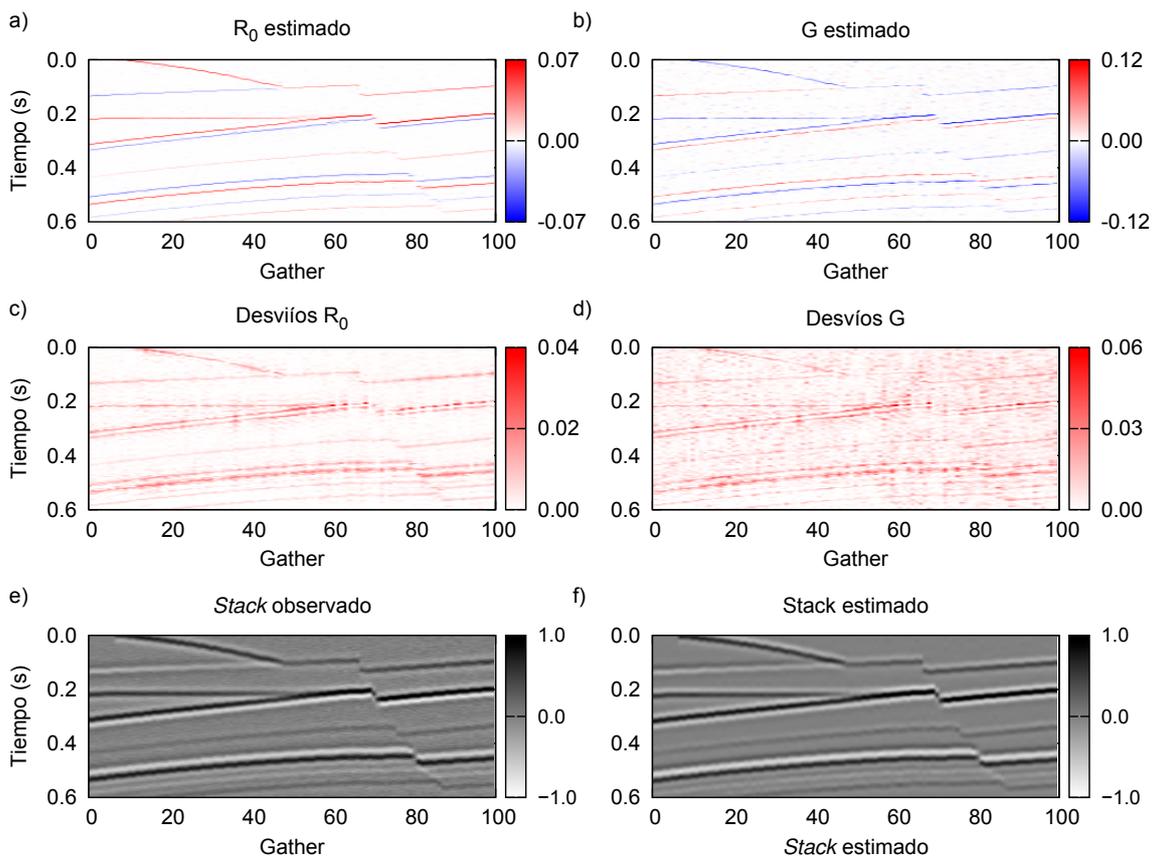


**Figura 5.8.** a), b) y c) modelo de velocidades y densidad 2D. d) y e) *intercept*  $R_0$  y *gradient*  $G$  correspondientes al modelo. f) *Stack-section* de los datos generados a partir del modelo.

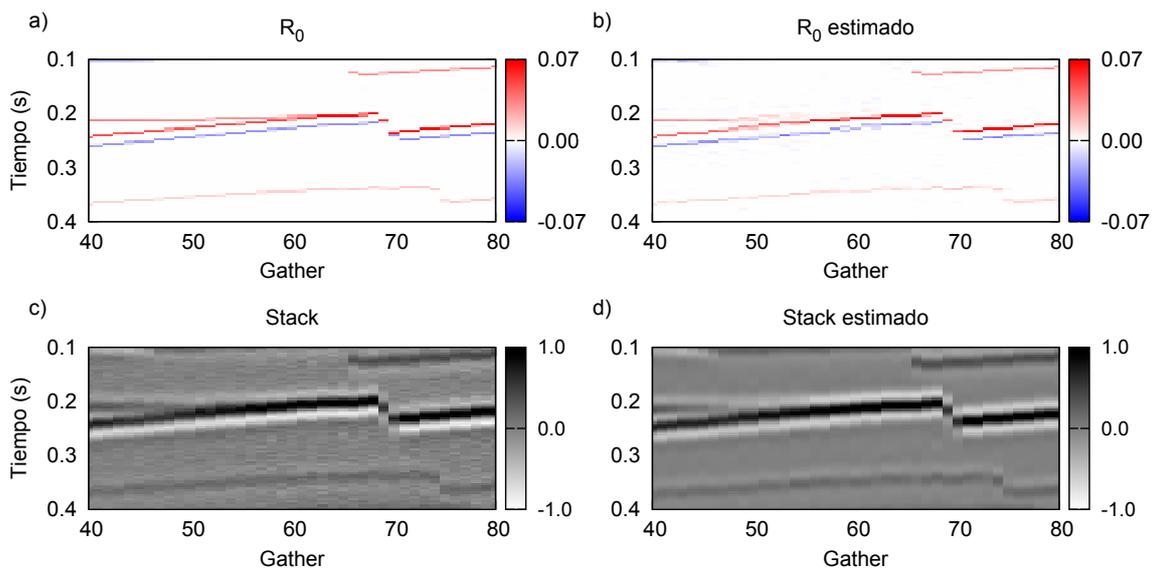
el *intercept* y el *stack* estimado seleccionando una ventana desde 0.1 s. a 0.4 s. y desde el gather #40 al #80. Allí queda en evidencia la capacidad del método para resolver reflectores cercanos. La diferencia, en tiempo, entre los reflectores que forman el par ubicado en el primer gather a los 0.32 s. se encuentra en el orden del espesor de *tuning*, que para el caso de los datos utilizados es aproximadamente 20 ms. El espesor de *tuning* es el mínimo espesor que debe poseer una capa para que los reflectores en el tope y la base se resuelvan como dos eventos separados. Para un espesor menor al del *tuning* las reflexiones comienzan a verse como un sólo evento tal como sucede con el *pinch out* (Sheriff, 2002) a los 0.2 s entre el gather #50 y #60. Para el caso de una ondícula de Ricker de frecuencia central  $f_0$  el espesor de *tuning* se puede expresar en unidades de tiempo como (Chung y Lawton, 1995)

$$T = \frac{\sqrt{6}}{2\pi f_0}. \quad (5.11)$$

En este caso, con  $f_0 = 20$  Hz se obtiene un valor para el espesor de *tuning* de aproximadamente 20 ms, lo que equivale a 5 muestras.



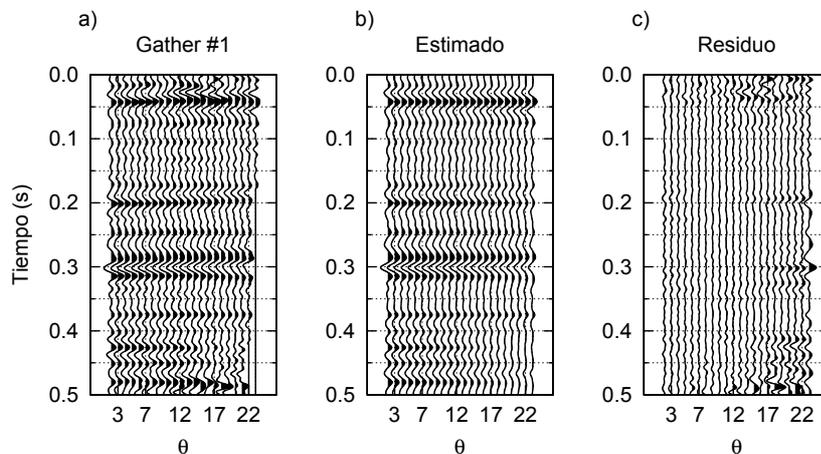
**Figura 5.9.** a) *Intercept*  $R_0$  estimado, b) *gradient*  $G$  estimado, c) desvío estándar del *intercept*, d) desvío estándar del *gradient*, e) *stack* observado, f) *stack* estimado.



**Figura 5.10.** a) Detalle del *intercept* verdadero del dato sintético 2D, b) detalle del *intercept* estimado, c) *stack* verdadero del dato sintético 2D, d) *stack* estimado.

### 5.4.3. Datos de campo

En esta sección probaremos el método en datos de campo. El dato de campo complica la aplicación del método propuesto debido no sólo a que el nivel de ruido no se conoce, sino también porque no se conoce la ondícula. Se asumirá que los datos han sido apropiadamente acondicionados para preservar las amplitudes. La ondícula fue estimada estadísticamente a partir de la raíz cuadrada de la transformada de Fourier de la autocorrelación de las trazas, asumiendo fase cero (Robinson y Treitel, 2002). Los datos de campo consisten en 160 *angle-gathers* ordenados por *Common Depth Point* (CDP) y corregidos por NMO, con un intervalo de muestreo de 2 ms. El rango de ángulos considerado para cada *angle-gather* es  $\theta_i \in (2^\circ, 24^\circ)$ . Se eligió una ventana temporal de 0.5 s. para realizar la inversión. Al igual que con los datos sintéticos, para la inversión se utilizó la aproximación de Shuey de dos términos. Asimismo se seleccionó  $L_s = 25$ . Para cada gather se obtuvo un arreglo que contiene las estimaciones de los coeficientes de la aproximación de Shuey. Una vez invertidos, se reconstruyó el gather y se calculó el *stack* correspondiente. Como en el ejemplo con datos sintéticos, y por razones estadísticas, se realizaron 50 inversiones y se tomó la media como solución final. A modo de ejemplo, la Figura 5.11 muestra el primer *gather* del dato, la correspondiente reconstrucción a partir del *intercept* y *gradient* medios estimados y la diferencia entre ambos. Como puede observarse la solución estimada honra al dato observado muy bien, sólo mostrando diferencias en lugares donde el dato observado es muy ruidoso, como al principio y al final del *gather*, y para los ángulos más grandes, donde la aproximación de Shuey de dos términos comienza a diferenciarse de las ecuaciones de Zoeppritz.

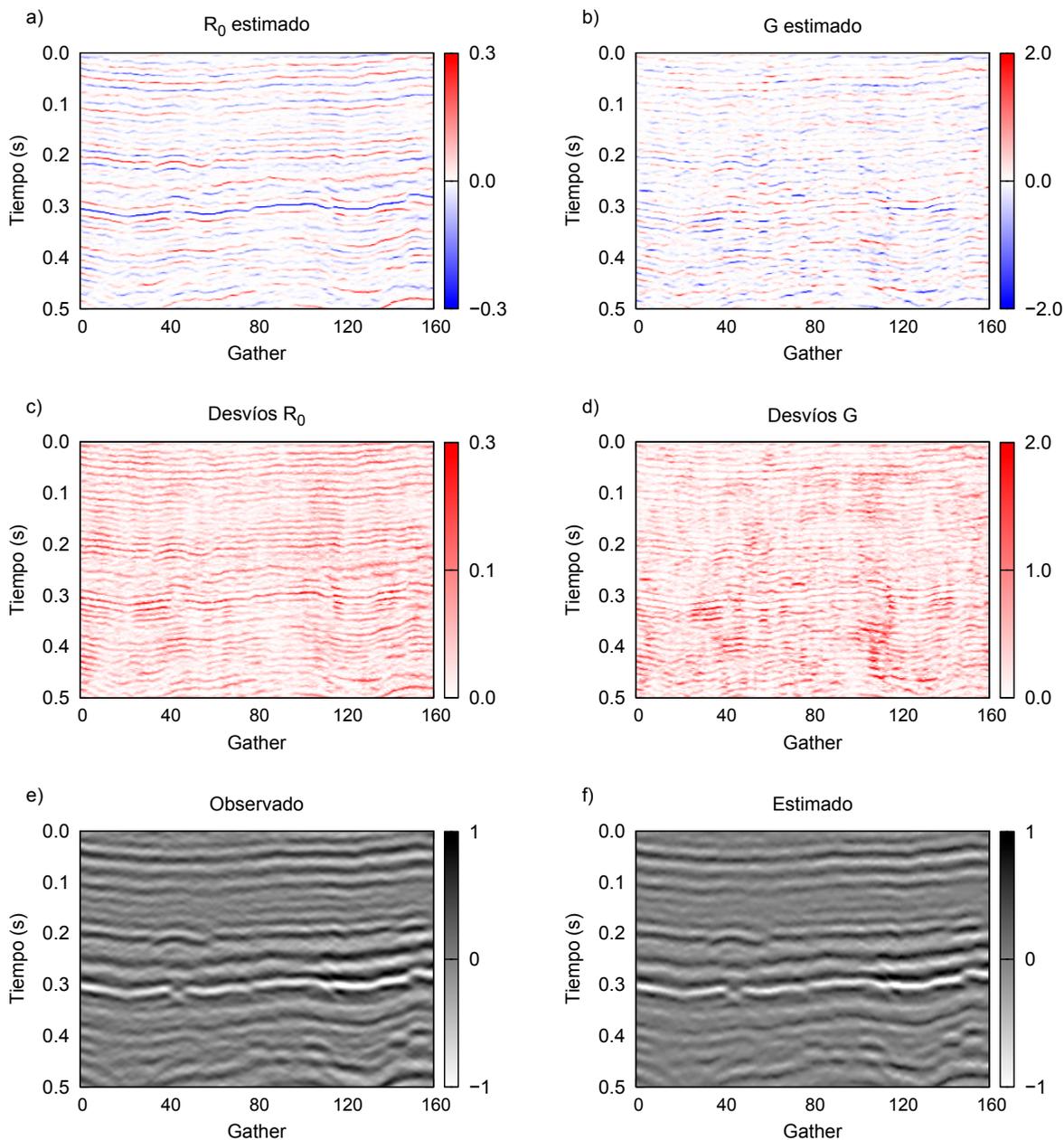


**Figura 5.11.** a) *Gather* observado #1 del dato de campo, b) estimado utilizando VFSA, c) residuo entre el *gather* observado y el estimado.

Las imágenes de alta resolución del *intercept* y del *gradient* pueden observarse en las Figuras 5.12a,b respectivamente, en tanto que las Figuras 5.12c,d muestran las correspondientes desviaciones estándar. Se puede observar que todos los reflectores más significativos fueron resueltos satisfactoriamente. Nótese además que se obtiene una continuidad lateral razonable a pesar de que no fue impuesta ninguna condición en este sentido durante la inversión, lo que demuestra la consistencia del método de inversión. Las Figuras 5.12e-f muestran, respectivamente, el *stack* del dato observado

## 5. Inversión *sparse* mediante *Simulated Annealing* y mínimos cuadrados

y el *stack* estimado luego de la inversión. Se observa que los datos estimados son muy similares a los datos observados. Los reflectores más importantes han sido hallados consiguiéndose un buen ajuste.



**Figura 5.12.** a) y b) *Intercept*  $R_0$  y *gradient*  $G$  estimados (promedio de 50 inversiones), c) y d) correspondientes desvíos estándar, e) y f) *stack* observado y estimado.

Como puede apreciarse, hay una considerable diferencia en las magnitudes del *intercept* y el *gradient* estimado. De acuerdo con Cambois (1998b,a), la razón de este comportamiento es el limitado rango de ángulos, de  $2^\circ$  a  $24^\circ$  en este caso, y la presencia de ruido. Por lo tanto no es inusual observar diferencias de estos órdenes de magnitud cuando se utilizan datos de campo. Otros autores atribuyen estas diferencias

a procesos físicos como la absorción, pérdidas por transmisión, anisotropía (Smith y Gidlow, 1987), o a un procesamiento inadecuado del dato antes de realizar la inversión, como puede ser una mala corrección por NMO o inexactitudes en la ondícula estimada (Castagna y Smith, 1994; Jin et al., 2000).

## 5.5. Conclusiones

A continuación se enumeran las principales conclusiones obtenidas utilizando la estrategia propuesta sobre dato sintético y dato de campo:

- Sobre datos sintéticos 1D el método demostró ser robusto en presencia de ruido aleatorio. Se obtuvieron muy buenos resultados en todas las magnitudes invertidas, estimándose con precisión la ubicación en tiempo de los reflectores y la respuesta AVA de los mismos.
- Se demostró que, dentro de un cierto rango, el número de reflectores  $L_s$  utilizado en la inversión no es crítico.
- Los pequeños valores de las desviaciones estándar de las magnitudes invertidas, resaltan la consistencia de las soluciones.
- Se observó que la utilización de una ondícula inexacta durante la inversión lleva a soluciones que honran al dato observado, pero que se alejan de las soluciones esperadas.
- Las pruebas sobre datos sintéticos 2D demostraron la capacidad del método para resolver reflectores cercanos y para obtener soluciones con buena continuidad lateral a pesar de que no se impuso ninguna clase de condición en este sentido durante la inversión.
- Sobre datos de campo el método mostró buen comportamiento a pesar de la incertidumbre inherente asociada a este tipo de datos. Los reflectores más importantes fueron encontrados y se obtuvieron imágenes de alta resolución del *intercept* y del *gradient* con buena continuidad lateral que honran al dato observado aceptablemente.
- En general las soluciones *sparse* obtenidas con el método propuesto son muy favorables en relación a aquellas obtenidas, por ejemplo, con los métodos de mínimos cuadrados.
- Aunque el número de reflectores involucrado en la inversión es limitado y los sistemas de ecuaciones a resolver son pequeños, el método es costoso computacionalmente debido a que se deben resolver dichos sistemas en cada iteración del VFSA.



# Capítulo 6

## Inversión *sparse* mediante regularización con la norma $l_1$

### 6.1. Introducción

En el capítulo anterior se procedió a regularizar el problema de inversión de datos *prestack* considerando *a priori* que las reflectividades son de tipo *sparse*. Para encontrar soluciones de tipo *sparse* se asumió como conocido el número de reflectores del subsuelo y se utilizó VFSA para obtener la ubicación de dichos reflectores, estimando las amplitudes de los coeficientes de la aproximación de Shuey de dos términos por medio de mínimos cuadrados. Sin embargo, debido al uso de un algoritmo de optimización global las soluciones de tipo *sparse* estimadas con esta estrategia se obtienen a expensas de un alto costo computacional, hecho que limita el uso del método en grandes volúmenes de datos.

Con el fin de superar dichos inconvenientes y poder obtener soluciones *sparse* de manera más eficiente, la estrategia desarrollada en este capítulo hace uso de un algoritmo simple y computacionalmente económico: el *Fast iterative shrinkage-thresholding algorithm* (FISTA). FISTA es un poderoso algoritmo iterativo utilizado para minimizar, junto con el *misfit*, la norma  $l_1$  de la solución estimada (Beck y Teboulle, 2009). Dicho problema de inversión es también conocido *least absolute shrinkage and selection operator* (LASSO) (Tibshirani, 1996). FISTA está basado en el *Iterative shrinkage-thresholding algorithm* (ISTA) (Daubechies et al., 2004), una extensión del algoritmo del gradiente clásico, utilizado para resolver problemas de inversión lineales de gran escala. Este algoritmo no requiere la inversión de matrices, y en cada iteración sólo se realizan multiplicaciones entre matrices y vectores. ISTA es un algoritmo con velocidad de convergencia lineal (Bredies y Lorenz, 2008), pero FISTA demostró ser varios órdenes de magnitud más veloz. El *misfit* y norma  $l_1$  serán combinadas en la función de costo a través de un parámetro de compensación o *trade-off* que deberá ser estimado por medio de algún criterio apropiado. Como algoritmo de optimización, FISTA provee soluciones de tipo *sparse* de una manera simple y efectiva. Sin embargo, como se verá en los ejemplos numéricos, en el caso de inversión de datos *prestack* contaminados con ruido las soluciones pueden ser correctas desde un punto de vista matemático pero carecer de significado físico. Este inconveniente surge por varias razones. Una de éstas es la propia naturaleza de los métodos de optimización basados en funciones del tipo

umbral (*soft-thresholding functions*), como lo son el ISTA y el FISTA. En este tipo de algoritmos es común observar una tendencia a subestimar las soluciones de mayor amplitud (Selesnick y Bayram, 2013; Bayram et al., 2014; Kowalski, 2014). Otra de las razones del problema se debe a la naturaleza de la norma  $l_1$  y a que las magnitudes implicadas en la inversión (p.ej. el *intercept* y *gradient* si se usa la aproximación de Shuey de dos términos) pueden mostrar diferentes escalas y, por lo tanto, no estar correctamente balanceados sus impactos en la norma  $l_1$ . Este problema ya fue analizado en la sección 4.4, donde se vio el mismo efecto sobre las amplitudes estimadas al minimizar la norma  $l_2$ . Para superar este inconveniente, proponemos utilizar la solución estimada por FISTA como información *a priori* para realizar, en un segundo paso, una inversión por mínimos cuadrados. Este segundo paso con fines correctivos, también llamado *de-biasing step* (Candes y Tao, 2007; Figueiredo et al., 2007), conduce a un método, al que llamaremos FISTA+LS, que promueve soluciones *sparse* robustas desde un punto de vista físico, y sin un incremento significativo del costo computacional.

El método propuesto en este capítulo conceptualmente tiene muchos puntos en común con el propuesto en el capítulo anterior. Ambos se basan en encontrar la ubicación en tiempo de un número pequeño de reflectores y a partir de esta información estimar las amplitudes de los coeficientes de la aproximación elegida. En el método del capítulo 5 dicha información era estimada por medio de VFSA, mientras que en este se obtiene, de una manera más eficiente, minimizando una función de costo utilizando FISTA. Mientras que al utilizar VFSA el número de reflectores no nulos era fijado previamente, al utilizar FISTA éste quedará determinado por el parámetro de compensación a estimar. Asimismo, en ambos métodos la estabilidad de la solución depende del número de reflectores no nulos, hecho que se vio al tomar valores grandes de  $L_s$  en la sección 5.4.1. Una de las principales diferencias es la eficiencia de los métodos. El VFSA es un algoritmo versátil que puede ser utilizado para minimizar prácticamente cualquier función, pero requiere gran cantidad de iteraciones. En el caso particular del capítulo 5, en cada iteración se resuelve por mínimos cuadrados un sistema de ecuaciones cuyo tamaño depende del valor de  $L_s$  elegido. En cambio el método propuesto en este capítulo, además de utilizar un algoritmo sumamente simple para minimizar la función de costo, sólo realiza un único ajuste por mínimos cuadrados al final del proceso iterativo. Otra diferencia es que mientras VFSA es un algoritmo estocástico y permite hacer un análisis de la incertidumbre de la solución estimada, el FISTA no lo es.

Este capítulo está organizado de la siguiente manera. Primero se explica el método propuesto, estableciendo las hipótesis necesarias y definiendo las ecuaciones relevantes. También se hace una descripción paso a paso de FISTA. Luego se procede a realizar pruebas del método en datos sísmicos *prestack* corregidos por *normal-moveout* (NMO), asumiendo que el AVA se puede modelar por medio de la aproximación de Shuey de dos términos. En el primer ejemplo numérico mostramos que, cuando el dato es ruidoso, FISTA no es capaz de proveer soluciones satisfactorias para el *intercept* y el *gradient*. Entonces, en un segundo ejemplo numérico, mostramos que las soluciones obtenidas usando FISTA+LS superan a las soluciones obtenidas por FISTA solo, siendo capaz de obtener estimaciones precisas tanto del *intercept* como del *gradient*. Finalmente probamos el método en datos de campo 2D, donde se puede ver que es posible obtener imágenes de alta resolución tanto del *intercept* como del *gradient* que honran al dato.

## 6.2. Teoría

Como en el caso de la deconvolución de datos sísmicos (Oldenburg et al., 1983), se pueden obtener soluciones *sparse* por medio de la minimización de la función de costo que utiliza el cuadrado de la norma  $l_2$  como discrepancia entre el dato observado y modelado, y la norma  $l_1$  como regularización. Entonces el problema se reduce a encontrar los parámetros del modelo que hagan mínima la función de costo

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + \mu\|\mathbf{m}\|_1, \quad (6.1)$$

donde  $\mathbf{m} = (m_{11}, \dots, m_{L_r1}, \dots, m_{1n}, \dots, m_{L_rn})^T$ , siendo  $\|\mathbf{m}\|_1 = (|m_{11}| + \dots + |m_{L_rn}|)$ , y  $\mathbf{A}$  es una matriz columna por bloques de dimensión  $L_d N_\theta \times L_r n$ , donde cada bloque está dado por  $\mathbf{A}(\theta_i)$ , y  $\mathbf{d}$  es un vector columna por bloques, donde cada bloque está dado por  $\mathbf{d}(\theta_i)$ , para  $i = 1, \dots, N_\theta$ . Una descripción detallada de la construcción de este sistema se puede encontrar en la sección 3.6. Como en los casos ya analizados, el primer término de la ecuación (6.1) representa el error entre el dato observado y el dato modelado. El segundo término es utilizado para penalizar las soluciones que no son *sparse*. El parámetro de compensación  $\mu$  es utilizado para balancear el peso, o impacto, de los dos términos. Debido a las propiedades de la norma  $l_1$  y la norma  $l_2$ , la función de costo dada por la ecuación (6.1) es una función convexa, lo que implica que sólo tiene un mínimo. La ecuación (6.1) será minimizada utilizando FISTA (Beck y Teboulle, 2009).

Es importante notar que, cuando el dato es ruidoso, la minimización de la ecuación (6.1) puede llevar a soluciones que honren al dato observado, pero muestren amplitudes mal estimadas. Esto se debe mayormente a que el parámetro de compensación  $\mu$  afecta de diferente manera a los elementos de  $\mathbf{m}$  dependiendo de cómo estos, según su escala, influyen en el dato observado. Este inconveniente será ilustrado en la próxima sección donde se verá que, cuando  $\mu$  es “grande” las soluciones honran el dato observado, pero sus amplitudes pueden resultar considerablemente subestimadas. Por otro lado, para valores de  $\mu$  “pequeños” las soluciones también honran al dato observado pero las amplitudes estimadas resultan muy sobreestimadas. Debido a esto proponemos llevar adelante la inversión por medio de un proceso de dos etapas. En la primer etapa utilizaremos FISTA para minimizar la función de costo  $J(\mathbf{m})$  usando un valor apropiado de  $\mu$  tal que la solución sea lo suficientemente *sparse*. La solución estimada en la primera etapa será usada entonces como información *a priori* que servirá de “soporte” para la segunda etapa del proceso. Dicha información *a priori* serán los tiempos asociados a los reflectores no nulos de la solución estimada por FISTA. La segunda etapa consiste en ajustar las amplitudes de los parámetros estimados por medio de mínimos cuadrados, es decir utilizando sólo el término del *misfit*. Es importante destacar que la estabilidad de la solución está garantizada siempre que la solución estimada en la primer etapa sea lo suficientemente *sparse*, de lo contrario la solución tendrá las características de las soluciones por mínimos cuadrados convencionales analizados en el capítulo 4 o las del método de inversión vía VFSA con  $L_s$  muy grande.

### Primera etapa: resolver el LASSO

En esta primera etapa del método propuesto se procede a minimizar la función de costo dada en la ecuación (6.1). Existen en la literatura varios algoritmos para estimar el mínimo de dicha función, en particular en esta tesis utilizaremos FISTA.

FISTA es un algoritmo iterativo, perteneciente a la familia de los algoritmos proximales (Bauschke et al., 2011), utilizado para resolver problemas de optimización convexos no diferenciables. Paso a paso, el algoritmo puede resumirse de la siguiente manera

1. Establecer  $\eta \geq \kappa_{max}$ , donde  $\kappa_{max}$  es el máximo autovalor de  $\mathbf{A}^T \mathbf{A}$ .
2. Establecer  $\mathbf{z}_0 = \mathbf{m}_0$  y  $t_0 = 1$ , donde  $\mathbf{m}_0$  es una solución inicial.
3. Para cada iteración  $k = 0, 1, 2, 3, \dots$ :

a)

$$\mathbf{m}_{k+1} = T_{\mu/2\eta} \left\{ \mathbf{z}_k + \frac{1}{\eta} \mathbf{A}^T (\mathbf{d} - \mathbf{A} \mathbf{z}_k) \right\}, \quad (6.2)$$

donde  $T_v\{\cdot\}$  es una función umbral suave (*soft-thresholding function*), que se aplica a cada elemento de su argumento vectorial y está definida como

$$T_v\{g\} = \begin{cases} g(1 - v/|g|), & |g| \geq v \\ 0, & |g| < v \end{cases} \quad (6.3)$$

b)

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}. \quad (6.4)$$

c)

$$\mathbf{z}_{k+1} = \mathbf{m}_{k+1} + \frac{t_k - 1}{t_{k+1}} (\mathbf{m}_{k+1} - \mathbf{m}_k). \quad (6.5)$$

d)

$$k = k + 1, \quad (6.6)$$

e) Verificar convergencia o criterio de corte.

Como puede verse FISTA es un algoritmo de optimización que no requiere inversión de matrices y que es sumamente sencillo de aplicar. El primer paso del proceso es necesario para evitar que el argumento de la función umbral sea negativo. Para encontrar el máximo autovalor puede utilizarse, por ejemplo, el método de las potencias de Rayleigh (Larson y Edwards, 1999, p.ej.), que es simple y eficiente. Una descripción detallada del algoritmo y sus propiedades puede encontrarse en el Apéndice B.

## Segunda etapa: ajuste de amplitudes por mínimos cuadrados

Como se verá en los ejemplos numéricos, minimizar la ecuación (6.1) llevará a una solución poco satisfactoria del problema de inversión que pretendemos resolver. Si bien los tiempos de los coeficientes de reflexión serán estimados con éxito y de forma eficiente, en algunos casos las amplitudes de los mismos estarán severamente subestimadas. Debido a esto, una vez que una solución del problema ha sido estimada utilizando FISTA, se realiza un paso correctivo o *debiasing step* (Candes y Tao, 2007; Figueiredo et al., 2007) con el fin de ajustar dichas amplitudes. De similar manera al método propuesto en el capítulo anterior, donde se utilizaba VFSA para estimar los tiempos de los coeficientes de reflexión y luego se estimaban sus amplitudes por mínimos cuadrados, aquí utilizaremos los tiempo de los coeficientes de reflexión estimados utilizando FISTA como información *a priori* para resolver un problema de mínimos cuadrados con el cual se estimarán las amplitudes de los mismos.

A partir de la solución  $\hat{\mathbf{m}}$  estimada por medio de la minimización de la ecuación (6.1) se obtiene el número  $L_s$  de reflectores distintos de cero ( $L_s \ll L_r$ ) y sus tiempos  $\tau_j$ . Luego se construye un sistema de ecuaciones lineales que relacionan el dato observado  $\mathbf{d}$  con los  $L_s n$  parámetros del modelo representado mediante

$$\mathbf{d} = \mathbf{A}_s \mathbf{m}_s, \quad (6.7)$$

donde  $\mathbf{m}_s = (m_{11}, \dots, m_{L_s 1}, \dots, m_{1n}, \dots, m_{L_s n})^T$  y  $\mathbf{A}_s$  es una matriz columna por bloques de dimensión  $L_d N_\theta \times L_s n$ , donde cada bloque está dado por  $\mathbf{A}_s(\theta_i)$ , y  $\mathbf{d}$  es un vector columna por bloques, donde cada bloque está dado por  $\mathbf{d}(\theta_i)$ , para  $i = 1, \dots, N_\theta$ . Una descripción detallada de la construcción de este sistema fue dada en la sección 5.2.

La solución  $\mathbf{m}_s$  de este sistema de ecuaciones lineales puede ser estimada por mínimos cuadrados, dando como resultado

$$\mathbf{m}_s = (\mathbf{A}_s^T \mathbf{A}_s)^{-1} \mathbf{A}_s^T \mathbf{d}. \quad (6.8)$$

En resumen, el método propuesto para realizar la inversión *sparse* a partir de dato sísmico *prestack* es como sigue:

1. Se estima una solución *sparse* de la ecuación (3.67) a través de la minimización de la ecuación (6.1) utilizando FISTA.
2. Se obtiene el número  $L_s$  y los tiempos  $\tau_j$ ,  $j = 1, \dots, L_s$  de los reflectores no nulos a partir de los elementos distintos de cero de  $\mathbf{m}$ , y se construye el sistema dado por la ecuación (6.7).
3. Se estima la solución  $\mathbf{m}_s$  por medio de mínimos cuadrados (ecuación (6.8)).

### 6.2.1. Formulación bayesiana del problema

En esta sección analizaremos el problema planteado por la ecuación (6.1) desde el punto de vista bayesiano. En la sección 2.6 se estableció que el enfoque bayesiano se basa en asumir que se cuenta con información sobre el modelo a estimar antes de que se observe el dato (Scales y Tenorio, 2001), y que esto permite incorporar la información *a priori* de forma más elegante desde el punto de vista matemático. Dicha

información *a priori* estará representada por la distribución de probabilidad *a priori*  $p(\mathbf{m})$ . En este caso, la función *a priori* deberá ser tal que promueva *sparseness* en las soluciones estimadas. La probabilidad de obtener el dato  $\mathbf{d}$  a partir del modelo  $\mathbf{m}$  queda representada por la función de verosimilitud  $p(\mathbf{d}|\mathbf{m})$ . Estas dos funciones se combinan en la función de probabilidad *a posteriori*  $p(\mathbf{m}|\mathbf{d})$  a través del teorema de Bayes, representado por la ecuación (2.23), resultando

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \quad (6.9)$$

El modelo  $\mathbf{m}$  más probable (MAP), dada la información *a priori* disponible, es aquel que hace máxima la función *a posteriori*.

Una de las formas de obtener soluciones *sparse* es asumiendo que a cada muestra  $m_q$  del vector  $\mathbf{m}$ , con  $q = 1, \dots, L_r n$ , se le puede asignar independientemente una distribución laplaciana de media cero (Figueiredo, 2003; Park y Casella, 2008; Babacan et al., 2010):

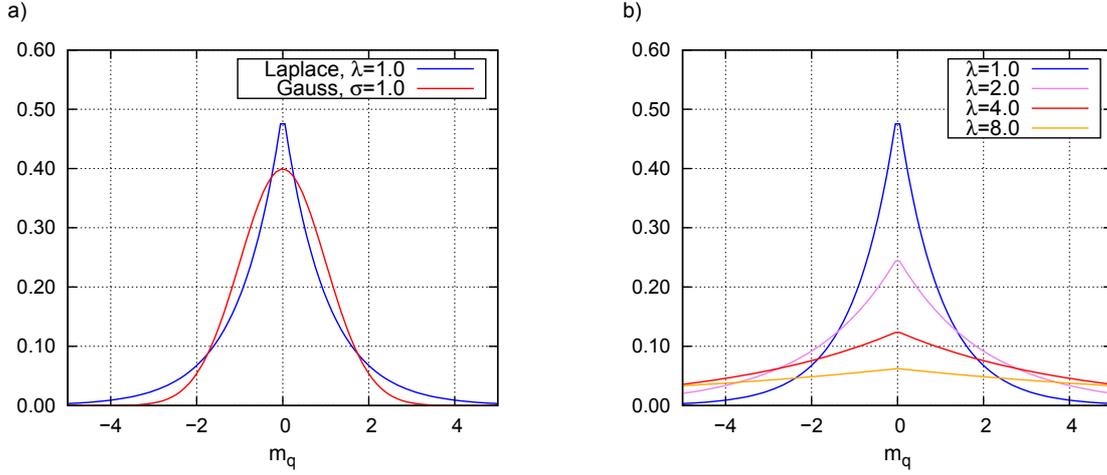
$$p(m_q) = \frac{1}{2\lambda} \exp\left(-\frac{|m_q|}{\lambda}\right), \quad (6.10)$$

donde  $\lambda$  es un parámetro de escala conocido como “diversidad”. En la Figura 6.1a se puede observar una comparación entre la distribución laplaciana de una variable y media cero (ecuación 6.10) y la distribución gaussiana de una variable, también de media cero. La distribución gaussiana fue utilizada como distribución *a priori* al analizar el enfoque bayesiano del método de mínimos cuadrados amortiguados (ecuación (2.31)). Como se puede apreciar la distribución laplaciana es, respecto a la gaussiana, más estrecha cerca del valor medio y presenta colas más largas. Estas características de la distribución laplaciana son las que la hacen una buena opción para estimar soluciones *sparse* ya que los valores más probable son aquellos cercanos a la media cero, pero sin que la probabilidad de obtener valores grandes no nulos sea demasiado pequeña como sí ocurre con la distribución gaussiana. Esta característica se corresponde con el comportamiento esperado de las reflectividades a estimar ya que, como ya se ha comentado, la distribución de las amplitudes de los coeficientes de reflexión no es gaussiana (Walden y Hosken, 1986; Velis, 2003), y estudios de pozo indican que las principales unidades litológicas pueden ser asociadas con grandes coeficientes de reflexión que muestran un comportamiento *sparse* (Walden y Hosken, 1986). La Figura 6.1b muestra el comportamiento de la distribución laplaciana en función del parámetro de escala  $\lambda$ . A medida que este parámetro aumenta la distribución se aplanan y tanto valores chicos como grandes comienzan a ser igualmente probables, es decir, disminuye la *sparseness*.

A partir de la distribución laplaciana para una variable dada por la ecuación (6.10) y asumiendo que cada  $m_q$  es independiente, la distribución de probabilidad conjunta de todos los  $m_q$  será el producto de las distribuciones individuales, por lo tanto la distribución de probabilidad *a priori* que representara al vector  $\mathbf{m}$  será

$$p(\mathbf{m}) = \prod_{q=1}^{L_r n} \frac{1}{2\lambda} \exp\left(-\frac{|m_q|}{\lambda}\right) = \left(\frac{1}{2\lambda}\right)^{L_r n} \exp\left(-\frac{1}{\lambda} \sum_{q=1}^{L_r n} |m_q|\right), \quad (6.11)$$

Al igual que en los ejemplos estudiados en la sección 2.6 se asumirá que el ruido presente en el dato observado tiene una distribución de probabilidad gaussiana, que



**Figura 6.1.** a) Comparación entre la distribución de probabilidad gaussiana y laplaciana, b) variación de la distribución laplaciana en función del parámetro de escala  $\lambda$ .

dicho ruido no está correlacionado y tiene varianza  $\sigma_n^2$ , por lo tanto la función de verosimilitud  $p(\mathbf{d}|\mathbf{m})$  será una función de densidad de probabilidad gaussiana multivariada de la forma

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{(2\pi)^{L_d} \sigma_n} \exp \left[ -\frac{1}{2\sigma_n^2} (\mathbf{d} - \mathbf{A}\mathbf{m})^T (\mathbf{d} - \mathbf{A}\mathbf{m}) \right]. \quad (6.12)$$

Reemplazando las ecuaciones (6.11) y (6.12) en la ecuación (6.9) se obtiene la función de densidad de probabilidad *a posteriori*

$$p(\mathbf{m}|\mathbf{d}) \propto \frac{1}{\lambda^{L_r n} 2^{(L_d + L_r n)} \pi^{L_d} \sigma_n} \exp \left[ -\frac{1}{2\sigma_n^2} (\mathbf{d} - \mathbf{A}\mathbf{m})^T (\mathbf{d} - \mathbf{A}\mathbf{m}) - \frac{1}{2\lambda} \sum_q^{L_r n} |m_q| \right]. \quad (6.13)$$

El modelo más probable, que maximiza la función  $p(\mathbf{m}|\mathbf{d})$ , será aquel que minimiza el exponente. Es decir, el modelo  $\mathbf{m}$  más probable es aquel que minimiza la función de costo dada por

$$J(\mathbf{m}) = (\mathbf{d} - \mathbf{A}\mathbf{m})^T (\mathbf{d} - \mathbf{A}\mathbf{m}) + \frac{\sigma_n^2}{\lambda} \sum_q^{L_r n} |m_q| = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + \frac{\sigma_n^2}{\lambda} \|\mathbf{m}\|_1, \quad (6.14)$$

que es igual a la ecuación (6.1) con  $\mu = \sigma_n^2/\lambda$ . Dado un valor de  $\sigma_n^2$  se deduce que las soluciones estimadas a partir de esta función de costo serán más *sparse* a medida que  $\mu$  se incremente, ya que esto implica que  $\lambda$  disminuye. Al igual que para el caso de mínimos cuadrados amortiguados el parámetro  $\mu$  tendrá que ser estimado por medio de algún criterio apropiado.

### 6.3. Ejemplos

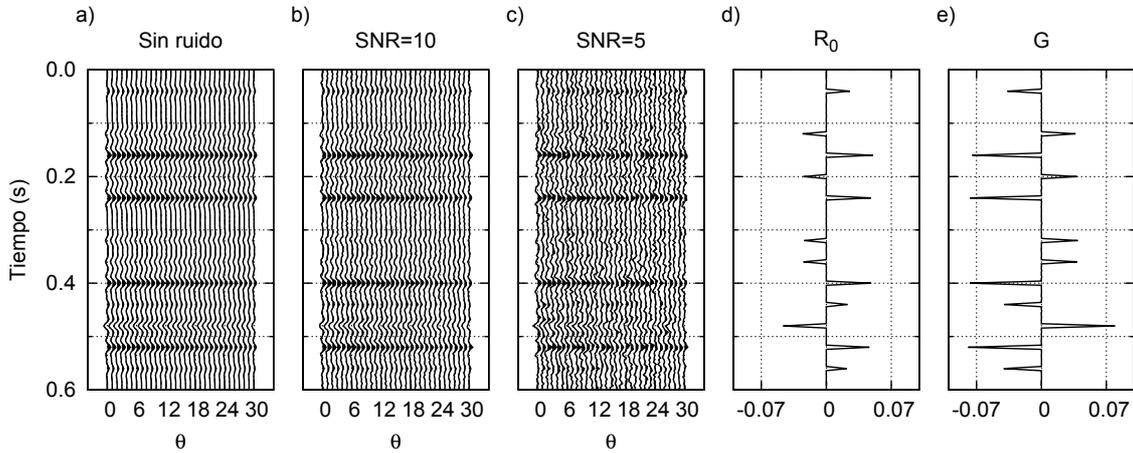
En esta sección probaremos el método propuesto utilizando datos sintéticos 1D y 2D y utilizando datos de campo. En todos los casos, al igual que en el capítulo anterior, las inversiones serán realizadas considerando la aproximación de Shuey de dos términos

$$R_{PP}(\theta_i) = R_0 + G \sin^2 \theta_i. \quad (6.15)$$

Es decir, se estimarán soluciones *sparse* del *intercept*  $R_0$  y del *gradient*  $G$ .

#### 6.3.1. Datos sintéticos 1D

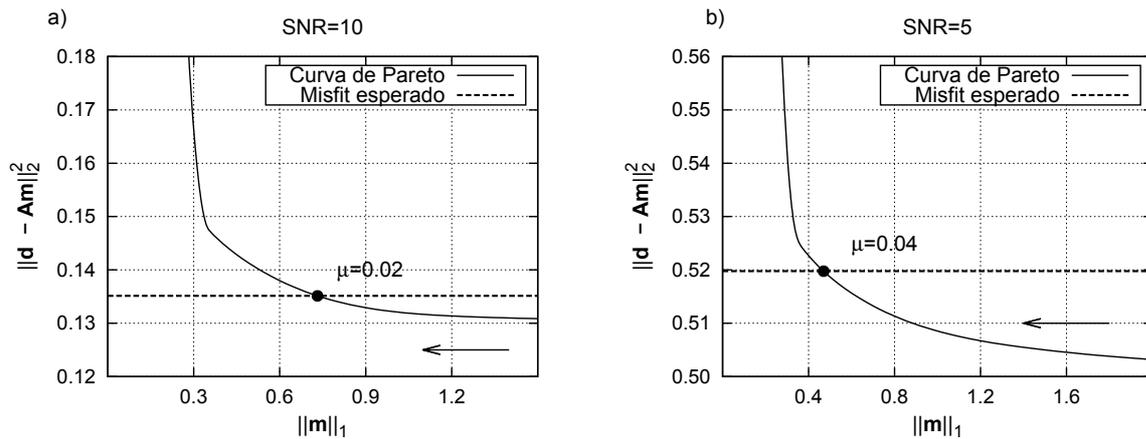
En este ejemplo numérico analizamos el comportamiento del método propuesto utilizando datos sísmicos *prestack* 1D sintético. Los datos están basados en el mismo modelo de velocidades y densidad utilizado para probar los métodos de mínimos cuadrados, y mínimos cuadrados amortiguados del capítulo 4, y el algoritmo basado en VFSA del capítulo 5 (Figura 4.1a-c). Cada *gather* contiene 12 reflectores y consta de 31 trazas con ángulos de incidencia  $\theta_i \in (0^\circ, 30^\circ)$ . La respuesta AVA se calculó utilizando las ecuaciones de Zoeppritz, se utilizó una ondícula de Ricker de frecuencia central  $f_0 = 30$  Hz, y el intervalo de muestreo utilizado fue de 4 ms. Para probar el algoritmo propuesto frente a diferentes niveles de ruido, se generaron dos *gathers*, uno con ruido gaussiano con relación señal-ruido SNR = 10 y otro con relación señal-ruido SNR = 5. La Figura 6.2 muestra el *gather* sin ruido, con SNR = 10, con SNR = 5, y el *intercept*  $R_0$  y *gradient*  $G$  correspondientes calculados a partir de los modelos de velocidades y densidad.



**Figura 6.2.** a) Dato sintético 1D libre de ruido, b) con SNR = 10, c) con SNR = 5, d) *intercept*  $R_0$  y e) *gradient*  $G$  correspondientes.

#### Inversión vía minimización de la norma $l_1$

Para analizar el comportamiento de las soluciones obtenidas mediante la minimización de la ecuación (6.1), en este primer ejemplo realizamos la inversión del dato con



**Figura 6.3.** Parámetros de compensación estimados por el principio de discrepancia para los datos con ruido de la Figura 6.2.

ruido de la Figura 6.2 omitiendo el segundo paso del método propuesto, es decir resolviendo sólo el LASSO. Para poder realizar la inversión, al igual que sucedió al aplicar el método de mínimos cuadrados amortiguados, primero necesitamos estimar un valor apropiado del parámetro de compensación  $\mu$ . La elección de este parámetro depende del nivel de ruido del dato observado. Claramente, si  $\mu$  es muy grande se estimará una solución muy *sparse*, pero puede que el dato no sea correctamente ajustado. Contrariamente, si  $\mu$  es muy pequeño puede haber un sobre-ajuste y la solución comenzará a ser inestable. Como ya fue comentado, hay varias estrategias apropiadas para seleccionar dicho parámetro (Farquharson y Oldenburg, 2004). En este caso, al igual que en el capítulo 4, utilizaremos el principio de discrepancia. Para ello se realiza la inversión utilizando diferentes valores de  $\mu$  y se construye la curva del *misfit* en función de la regularización. En este caso será la curva del *misfit* en función de la norma  $l_1$  de la solución estimada. A esta curva se la conoce como curva de Pareto (van den Berg y Friedlander, 2008). Luego se elige el  $\mu$  óptimo como aquél que minimiza la norma  $l_1$  mientras el *misfit* permanece menor o igual a su valor esperado  $E_n$  (ecuación (4.4)).

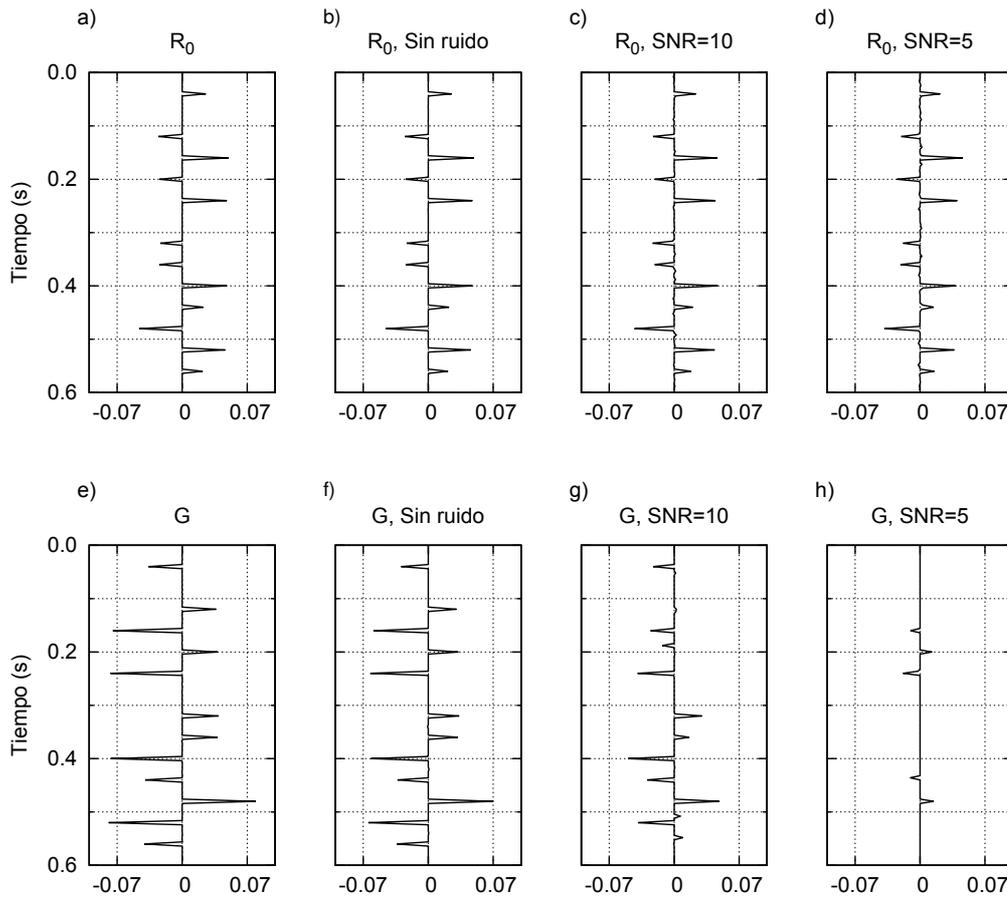
Además de proveer una herramienta para estimar el parámetro de compensación, las curvas de Pareto proveen un criterio objetivo para estudiar cómo diferentes algoritmos de optimización convergen hacia la solución estimada (Hennenfent et al., 2008). La Figura 6.3 muestra las curvas de Pareto correspondientes al dato observado con SNR = 10 y SNR = 5 (líneas sólidas). La línea horizontal discontinua indica el valor esperado del *misfit* y las flechas indican la dirección de incremento de  $\mu$ . Como es esperable, cuanto más grande el  $\mu$  más pequeño el valor de la norma  $l_1$  de la solución estimada y más grande el *misfit*, y viceversa. A través del principio de discrepancia se estimaron valores óptimos  $\mu = 0.02$  para el dato con SNR = 10, y  $\mu = 0.04$  para el dato con SNR = 5. Estos valores corresponden al cruce de la curva de Pareto con el valor esperado del *misfit*, y están indicados con los círculos negros en las figuras.

La Figura 6.4 muestra el *intercept*  $R_0$  y el *gradient*  $G$  del modelo, y los valores estimados correspondientes al dato sin ruido, con SNR = 5 y con SNR = 10, utilizando los valores de  $\mu$  estimados por medio del principio de discrepancia. Se puede observar que los resultados estimados del *intercept* son precisos incluso en presencia de una

significante cantidad de ruido. La localización en tiempo de todos los coeficientes de reflexión y sus amplitudes fueron estimadas correctamente. Al igual que en los ejemplos analizados en el capítulo 5, no hay que perder de vista que el dato fue creado utilizando las ecuaciones de Zoeppritz y la inversión se llevó a cabo a través de la aproximación de Shuey de dos términos, por lo que pequeñas diferencias entre los valores de las amplitudes estimadas y las del modelo son esperables. Aunque las soluciones muestran algunos coeficientes de reflexión espurios, especialmente para la relación señal-ruido más baja, sus amplitudes son pequeñas y pueden ser fácilmente eliminados por medio de algún filtro o función umbral adecuada, y de esa forma conseguir un ajuste casi perfecto con el valor del *intercept* esperado. Por otro lado, las soluciones estimadas para el *gradient*, cuando el dato es ruidoso, se apartan considerablemente de los valores esperados. La localización en tiempo es correcta, pero sus amplitudes no. A medida que el nivel de ruido aumenta, las amplitudes de los coeficientes de reflexión muestran una severa subestimación. Esto es esperable debido a que el ruido oculta más el efecto del AVA que la energía del reflector, y por lo tanto el término del *misfit* en la ecuación (6.1) es menos sensible a las variaciones en el *gradient* que la norma  $l_1$ . Por lo tanto, con el fin de reducir el término de regularización lo más posible FISTA tiende a reducir la norma  $l_1$  del *gradient* más que la norma  $l_1$  del *intercept*, sin que esto afecte al *misfit* significativamente. En casos de datos con gran cantidad de ruido donde para algunos reflectores el AVA es casi indistinguible, el algoritmo reduce las amplitudes del valor estimado del *gradient* a cero como puede apreciarse en las Figuras 6.4g,h.

Continuando con el análisis del comportamiento de las soluciones estimadas minimizando la ecuación (6.1), la Figura 6.5 muestra el resultado de realizar la inversión utilizando  $\mu = 0.1$ ,  $\mu = 0.04$ ,  $\mu = 0.01$  y  $\mu = 0.005$  en el caso del dato con  $\text{SNR} = 5$  (Figura 6.2c). Cuando  $\mu$  es más pequeño que el valor estimado utilizando el principio de discrepancia ( $\mu = 0.04$ ), las amplitudes del *gradient* estimado se incrementan, pero las soluciones comienzan a ser inestables. La cantidad y amplitudes de los reflectores espurios, tanto en el *intercept* como en el *gradient*, comienzan a ser más grandes que lo aceptable como puede verse en las Figuras 6.5a,b,e y f. Este fenómeno es similar al observado en el método propuesto en el capítulo 5 cuando se eligieron arbitrariamente valores de  $L_s$  demasiado grandes (ver Figura 5.2). Por otro lado, si  $\mu$  es más grande que el valor óptimo, los valores del *gradient* están aún más subestimados, siendo incluso cero como puede verse en la Figura 6.5h. Este comportamiento no se observa en el *intercept* estimado, debido a que su efecto en la norma  $l_1$  es diferente al del *gradient* como fue explicado en el párrafo anterior. En este caso, existe un amplio rango de valores de  $\mu$  que llevan a estimaciones del *intercept* aceptables. A partir de este análisis se concluye que no se puede estimar un único valor de  $\mu$  que lleve a soluciones aceptables del *intercept* y el *gradient* simultáneamente, excepto para valores de la relación señal-ruido altos, como muestran las Figuras 6.4b,f.

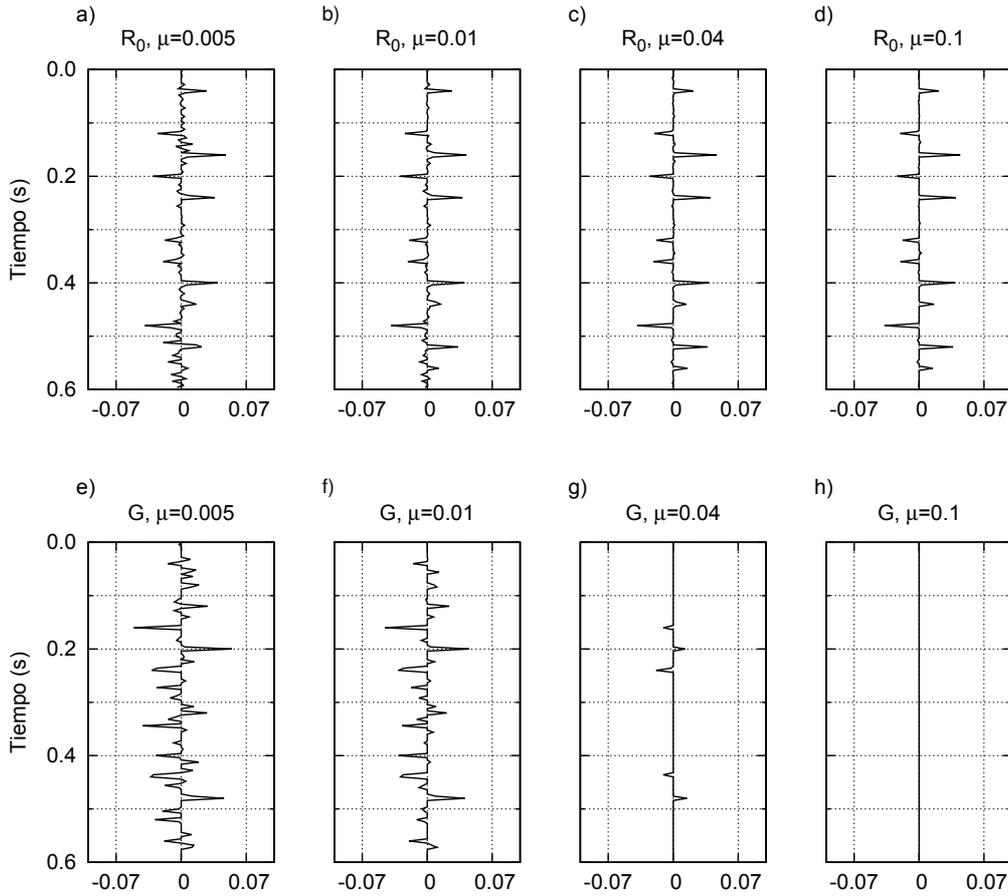
La Figura 6.6 nos permite hacer un análisis cualitativo del comportamiento de FISTA a medida que converge a la solución estimada, para los diferentes valores de  $\mu$  utilizados. La línea negra muestra la curva de Pareto correspondiente al dato con  $\text{SNR} = 5$ . Esta curva es la misma que la curva de Pareto de la Figura 6.3. Las líneas de color indican los valores de *misfit* y norma  $l_1$  calculados por FISTA a medida que transcurre el proceso iterativo, que finaliza cuando se alcanzan los valores marcados por los círculos negros. El punto verde marca el valor del *misfit* esperado y la norma



**Figura 6.4.** a) y e) *Intercept*  $R_0$  y *gradient*  $G$  del modelo, b) y f) estimados a partir del dato sin ruido, c) y g) a partir del dato con  $\text{SNR} = 10$ , d) y h) a partir del dato con  $\text{SNR} = 5$ .

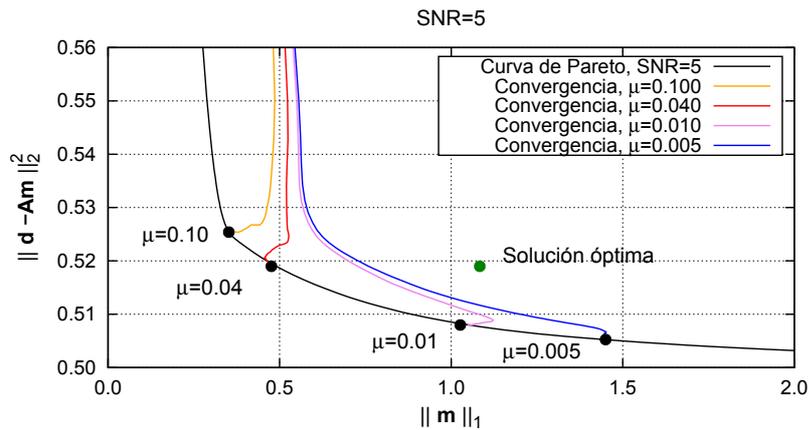
$l_1$  correspondientes al modelo verdadero, es decir, es la “mejor” solución que podría estimarse. Como ya fue visto en los resultados obtenidos, con ninguno de los valores de  $\mu$  se esta cerca de la solución óptima. Sólo  $\mu = 0.04$  tiene un valor del *misfit* cercano al esperado, debido a que fue elegido por medio del principio de discrepancia. Todos los demás valores de  $\mu$  dan soluciones que sobre-ajustan o sub-ajustan al dato observado. Es importante destacar que incluso en la mejor de las situaciones el valor óptimo del *misfit* es difícil de alcanzar, no sólo debido al ruido, sino también porque el modelo AVA con el que se generó el dato difiere del utilizado en la inversión. Asimismo sólo con la solución estimada utilizando  $\mu = 0.01$  se tiene un valor de la norma  $l_1$  cercano al verdadero, pero la solución es inestable, especialmente el *gradient* (Figura 6.5b,f).

El camino seguido por FISTA para llegar a la solución estimada es motivo de análisis. Como puede observarse en las primeras instancias del proceso iterativo el algoritmo tiende a minimizar el *misfit*, incluso a costas de sobreestimar el valor de la norma  $l_1$ . Una vez que el valor del *misfit* se estabiliza el algoritmo comienza a minimizar la norma  $l_1$  hasta converger a su valor final. La Figura 6.7 muestra el valor del *misfit* y la norma  $l_1$  en función del número de iteración para  $\mu = 0.04$ . Allí se puede observar más claramente el comportamiento descrito. El *misfit* converge rápidamente hacia el valor

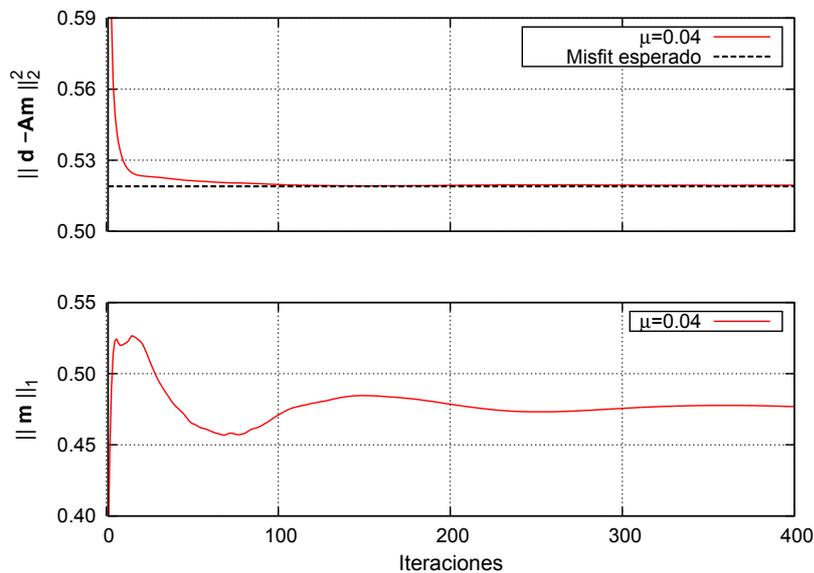


**Figura 6.5.** a) y e) *Intercept*  $R_0$  y *gradient*  $G$  estimados con  $\mu = 0.005$ , b) y f) estimados con  $\mu = 0.01$ , c) y g) estimados con  $\mu = 0.04$ , d) y h) estimados con  $\mu = 0.1$ .

esperado, mientras que el comportamiento de la norma  $l_1$  es oscilatorio, convergiendo más lentamente a su valor final. Esto ocurre porque en los casos analizados los valores de  $\mu$  son pequeños, es decir que la regularización tiene poco peso en la función de costo, entonces FISTA actúa primero minimizando el *misfit* y cuando éste es suficientemente pequeño el término de regularización comienza a tener un mayor efecto (Hennenfent et al., 2008). Este comportamiento es de especial interés a la hora de seleccionar el criterio de corte con el cual concluir el proceso iterativo, ya que alcanzar un valor determinado del *misfit* no necesariamente garantiza una solución *sparse*.



**Figura 6.6.** Curva de Pareto para el dato con  $\text{SNR} = 5$  (línea negra) y curvas de convergencia del FISTA para diferentes valores de  $\mu$  (líneas de colores).



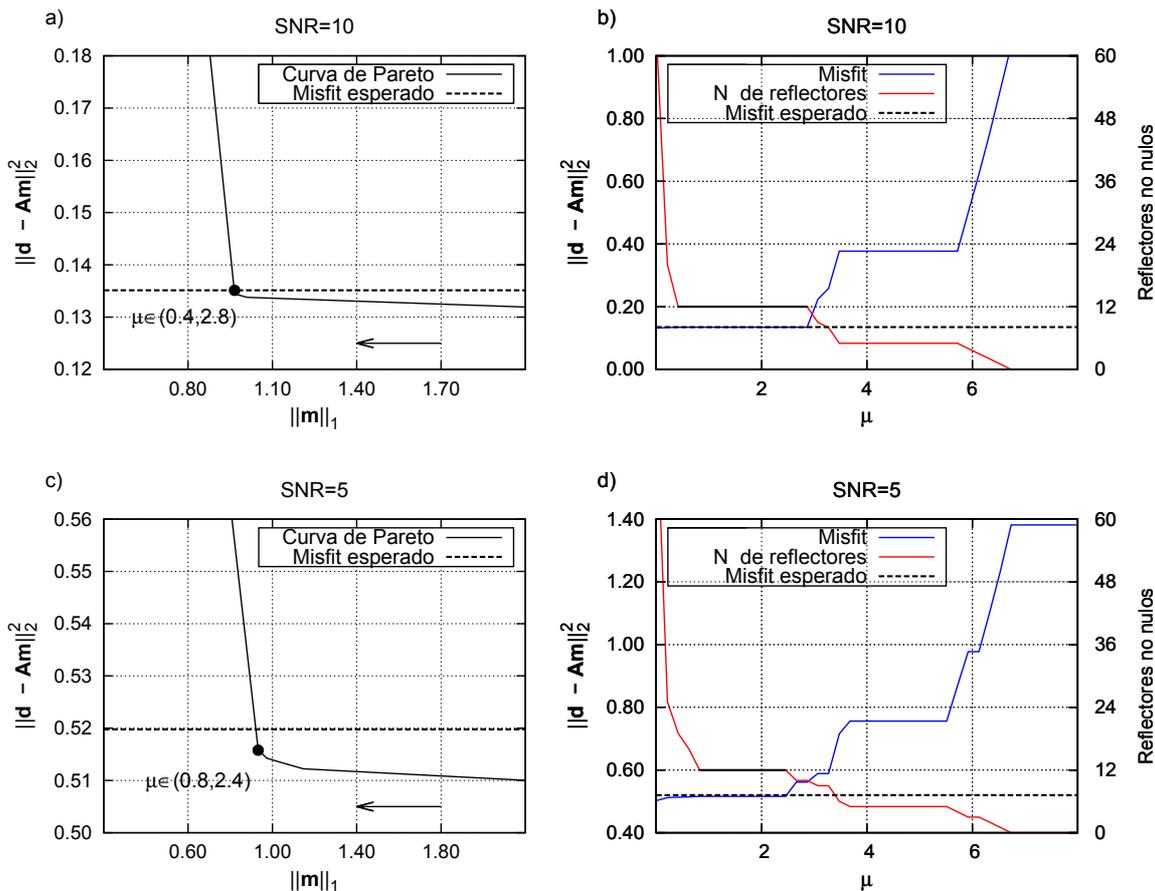
**Figura 6.7.** *Misfit* y valor de la norma  $l_1$  versus el número de iteraciones del FISTA para  $\mu = 0.04$ .

### Inversión vía minimización de la norma $l_1$ y mínimos cuadrados

Como se vio anteriormente, resolver el LASSO lleva a soluciones que no son satisfactorias. Las soluciones del *gradient* no son aceptables, estando las amplitudes severamente subestimadas. Sin embargo, el *intercept* puede ser estimado con un alto grado de precisión. El método propuesto en este capítulo saca ventaja de este resultado, utilizando los valores estimados del *intercept* como información *a priori* para mejorar la solución estimada del *gradient* a través de una inversión por mínimos cuadrados. Se procedió a probar el método en los datos sintéticos de la Figura 6.2. Las Figuras 6.8a,b muestran las curvas de Pareto obtenidas a través de FISTA+LS, para el dato con  $\text{SNR} = 10$  y  $\text{SNR} = 5$ . Una comparación con las curvas de Pareto obtenidas

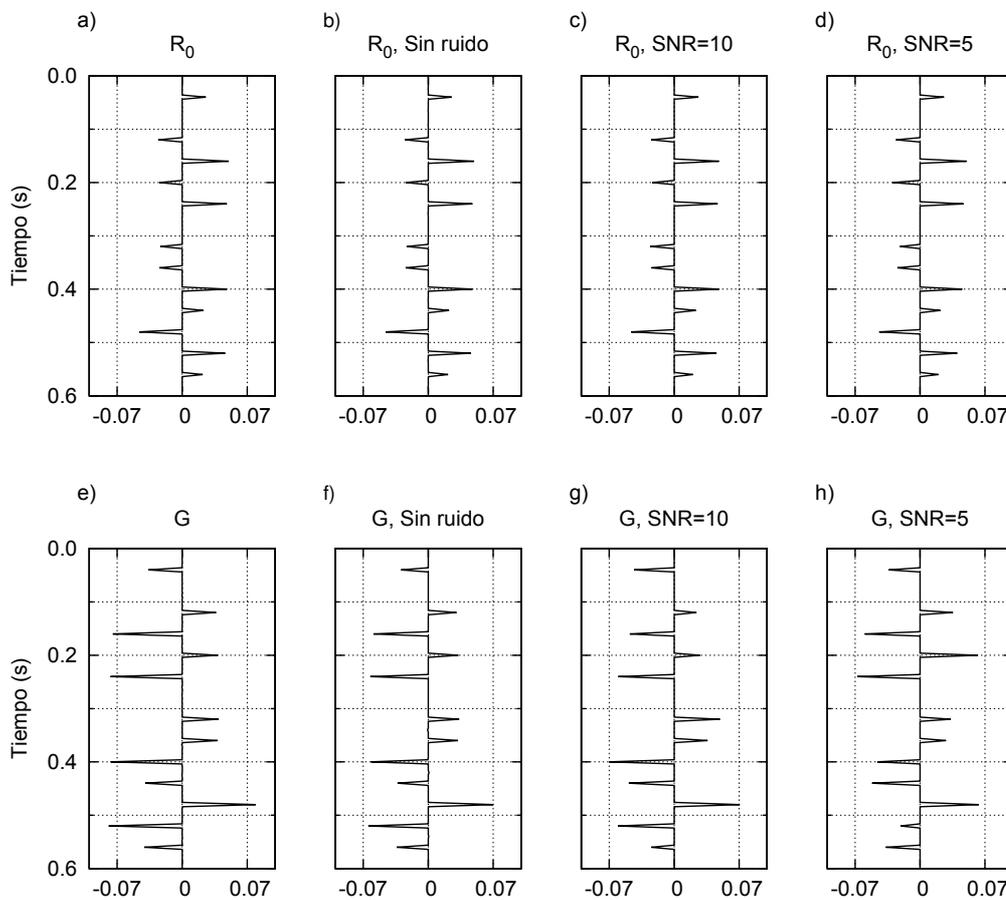
## 6. Inversión *sparse* mediante regularización con la norma $l_1$

utilizando únicamente FISTA muestra que éstas exhiben un quiebre cuando se aproximan al *misfit* esperado. Por lo tanto, cuando la estimación del nivel de ruido no está disponible, sería más simple estimar el valor de  $\mu$  utilizando el método de dos etapas que utilizando solamente FISTA. Por otro lado, las curvas rojas en las Figuras 6.8c,d muestran en número de reflectores no nulos, en función de  $\mu$ , estimados a partir de la inversión. Es importante marcar que para un dado *gather* existe un rango de valores de  $\mu$  para los cuales el número de reflectores no nulos estimado luego de la inversión es igual al número de reflectores verdadero. Este rango está representado por el segmento de color negro en la curva roja. Estos segmentos se mapean en los círculos negros de las Figuras 6.8a,b, que representan, en la práctica, un rango de valores de  $\mu$  en lugar de un único valor. En el caso de  $\text{SNR} = 10$  este rango es  $\mu \in (0.4, 2.8)$ , mientras que si  $\text{SNR} = 5$  este rango es  $\mu \in (0.8, 2.4)$ . Por lo tanto, a diferencia de cuando sólo se minimiza la ecuación (6.1), donde de acuerdo con el principio de discrepancia existe un único  $\mu$  óptimo, cuando se utiliza FISTA+LS existen varios valores de  $\mu$  que minimizan la norma  $l_1$  de la solución mientras el *misfit* toma valores menores o iguales a su valor esperado, un hecho que se muestra a través de las curvas azules en las Figuras 6.8c,d. Esto nos dice que las soluciones del método de dos etapas dependen, esencialmente, del número de reflectores provisto por FISTA.



**Figura 6.8.** a) y c) Curvas de Pareto para los datos con  $\text{SNR} = 10$  y  $\text{SNR} = 5$ , b) y d) *misfit* y número de reflectores no nulos en función del parámetros de compensación  $\mu$  para los datos con  $\text{SNR} = 10$  y  $\text{SNR} = 5$ .

En resumen, cuando se realiza la inversión resolviendo el LASSO sin el paso correctivo, como  $\mu$  balancea el impacto de los dos términos en la función de costo (ecuación 6.1) es esperable que las amplitudes estimadas de los reflectores disminuyan a medida que  $\mu$  se incrementa, incluso cuando el número de reflectores no nulos permanece constante (como se ha visto en la Figura 6.4). Pero, por otro lado, cuando se utiliza FISTA+LS, tanto el *misfit* como las amplitudes de los reflectores estimados son ajustados al mismo valor luego del paso correctivo independientemente del valor de  $\mu$  elegido dentro del rango en el cual el número de reflectores no nulos permanece constante (segmentos color negro en las Figuras 6.8c,d). A partir de estos resultados, y para este ejemplo en particular, elegimos como valor del parámetro de compensación  $\mu = 2$  cuando  $\text{SNR} = 10$  y  $\text{SNR} = 5$ . Sin embargo, cualquier otro valor dentro de los rangos correspondientes llevaría exactamente el mismo resultado.



**Figura 6.9.** a) y e) *Intercept*  $R_0$  y *gradient*  $G$  verdaderos, b) y f) estimados a partir del dato sin ruido, c) y g) a partir del dato con  $\text{SNR} = 10$ , d) y h) a partir del dato con  $\text{SNR} = 5$ .

La Figura 6.9 muestra el *intercept* y el *gradient* del modelo, y los valores estimados a partir del dato libre de ruido, con  $\text{SNR} = 10$  y con  $\text{SNR} = 5$ . Se puede observar que el método propuesto da, en todos los casos, soluciones más precisas tanto para el *intercept* como para el *gradient* en comparación al caso donde no se aplicó el paso correctivo (Figura 6.4). Todos los reflectores y sus amplitudes fueron adecuadamente

estimadas, incluso cuando el nivel de ruido en el dato es elevado.

### Comparación de resultados de los diversos métodos

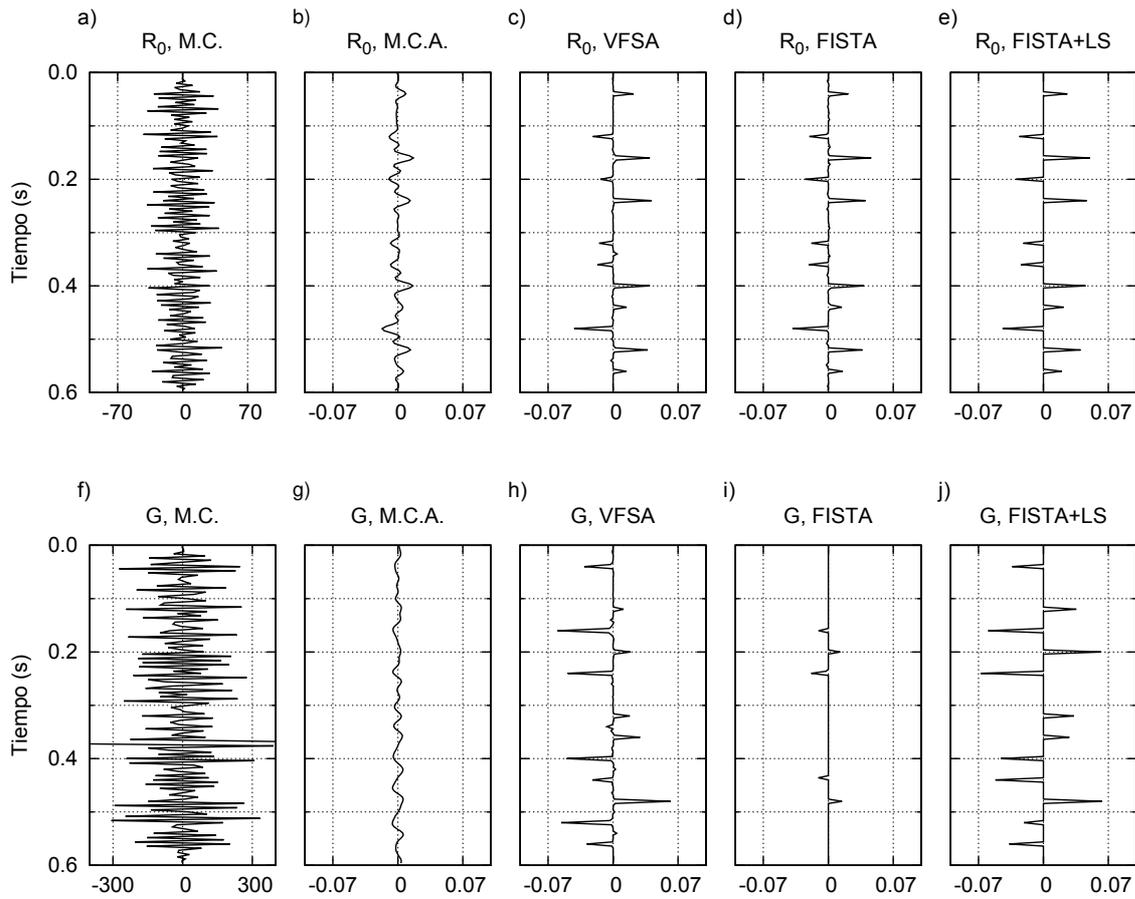
Finalmente, la Figura 6.10 muestra, a modo de comparación, las soluciones obtenidas a partir del dato 1D con  $\text{SNR} = 5$ , utilizando los métodos propuestos en los capítulos 4 y 5 y los propuestos en éste, es decir:

1. Solución por mínimos cuadrados sin amortiguar: capítulo 4, Figuras 4.3d,e.
2. Solución por mínimos cuadrados amortiguados: capítulo 4, Figuras 4.6d,e.
3. Solución estimada utilizando VFSA con  $L_S = 12$ : capítulo 5, Figuras 5.2b,f.
4. Solución estimada utilizando FISTA con  $\mu = 0.004$ : capítulo 6, Figuras 6.4d,h.
5. Solución estimada utilizando FISTA+LS con  $\mu = 2$ : capítulo 6, Figuras 6.9d,h.

Es importante remarcar que en todos los casos las soluciones estimadas ajustan correctamente al dato observado a pesar de las considerables diferencias que presentan entre sí. Las Figuras 6.10a,f muestran el *intercept* y el *gradient* obtenidos utilizando el método de mínimos cuadrados, donde se observa que la solución es inestable y las amplitudes están muy sobreestimadas. Las Figuras 6.10b,g muestran las soluciones obtenidas por el método de mínimos cuadrados amortiguado. Se puede apreciar una menor inestabilidad en las soluciones estimadas, sin embargo las amplitudes resultan severamente subestimadas, y la baja resolución de éstas complica su interpretación. Las Figuras 6.10c,h muestran las soluciones estimadas utilizando VFSA con  $L_s = 12$ . A través de este método se pudo estimar soluciones *sparse* que ajustan al dato observado. Sin embargo hay que remarcar que el método es costoso desde el punto de vista computacional, y que la solución mostrada es el valor medio de 100 realizaciones del método, razón por la cual las amplitudes de algunos reflectores fueron levemente subestimadas. En las Figuras 6.10d,i se puede observar el resultado obtenido minimizando la función de costo dada por la ecuación (6.1) por medio del algoritmo FISTA. Como ya se estableció, el ruido presente en el dato enmascara el AVA y, como consecuencia, al minimizar la norma  $l_1$  de las soluciones se obtienen solución del *gradient* severamente subestimadas. Finalmente, las Figuras 6.10e,j muestran el resultado obtenido mediante el método de dos etapas propuesto. Como puede apreciarse las soluciones obtenidas superan al resto, todos los reflectores y sus amplitudes fueron estimados correctamente, de una manera sencilla y económica computacionalmente, a pesar del ruido presente en el dato observado.

### Análisis estadístico

FISTA, a diferencia de VFSA, no es un algoritmo estocástico. Es decir, si se aplica varias veces el algoritmo sobre un mismo dato, con  $\mu$  fijo, obtendremos soluciones idénticas. A pesar de esto se puede obtener información estadística útil al aplicar el algoritmo una única vez sobre una familia de datos, donde cada miembro de esta familia es creado a partir de contaminar un único dato inicial con diferentes realizaciones de ruido gaussiano. Dicho ruido tiene que ser tal que la relación señal-ruido SNR sea

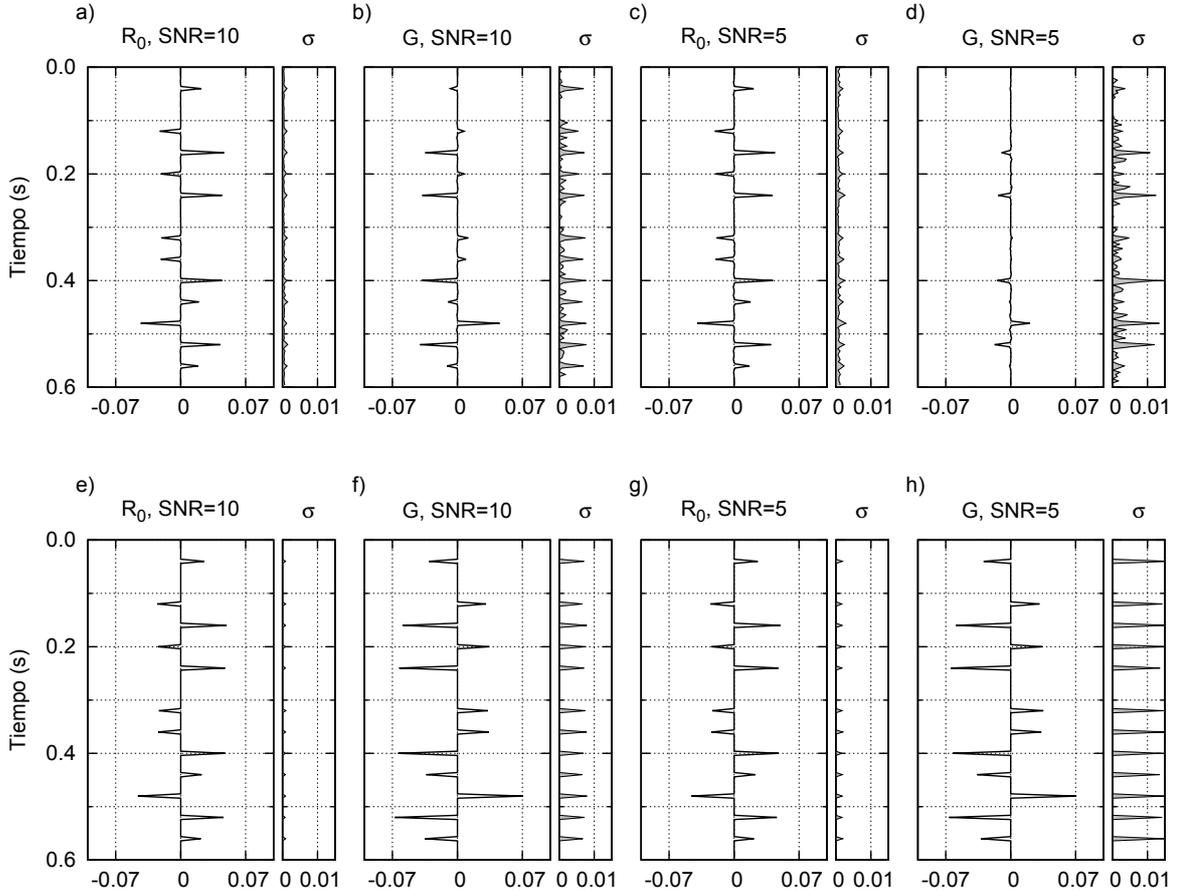


**Figura 6.10.** Comparación entre los métodos vistos. *Intercept*  $R_0$  y *gradient*  $G$  estimados utilizando a) y f) mínimos cuadrados (M.C.), b) y g) mínimos cuadrados amortiguados (M.C.A.), c) y h) usando VFSA ( $L_s = 12$ ), d) y i) usando FISTA ( $\mu = 0.004$ ), e) y j) usando FISTA+LS ( $\mu = 2$ ).

igual para cada realización. Es decir, en lugar de obtener una familia de soluciones aplicando múltiples veces un algoritmo, sobre un único dato, obtendremos una familia de soluciones aplicando el algoritmo una única vez sobre múltiples datos similares. Para este ejemplo, se creó una familia de 100 *gathers* con  $\text{SNR} = 10$  y 100 *gathers* con  $\text{SNR} = 5$  a partir del modelo de la Figura 6.2. Debido a que el  $\text{SNR}$  será igual para cada uno de los *gathers* de cada familia de datos, en la inversión se utilizarán los valores de  $\mu$  ya estimados mediante el principio de discrepancia.

La Figura 6.11 muestra los valores medios y los desvíos estándar obtenidos a partir de las soluciones obtenidas sobre las 100 realizaciones del dato con  $\text{SNR} = 10$  y con  $\text{SNR} = 5$ , resolviendo sólo el LASSO (Figuras 6.11a-d) y aplicando FISTA+LS (Figuras 6.11e-f). Para las soluciones estimadas del *intercept* se puede observar que en todos los casos los valores medios son similares a los valores del modelo. Por otro lado, los desvíos estándar son pequeños en todos los casos analizados, indicando que este atributo fue estimado correctamente con ambos métodos para todos los datos. En el caso del *gradient*, como era esperable, los valores medios estimados utilizando sólo FISTA están lejos de ser óptimos, no así los valores medios estimados con el método de dos etapas.

Los desvíos estándar, en todos los casos, son mayores que aquellos obtenidos para el *intercept*, demostrando que el *gradient* es más sensible al ruido en el dato.

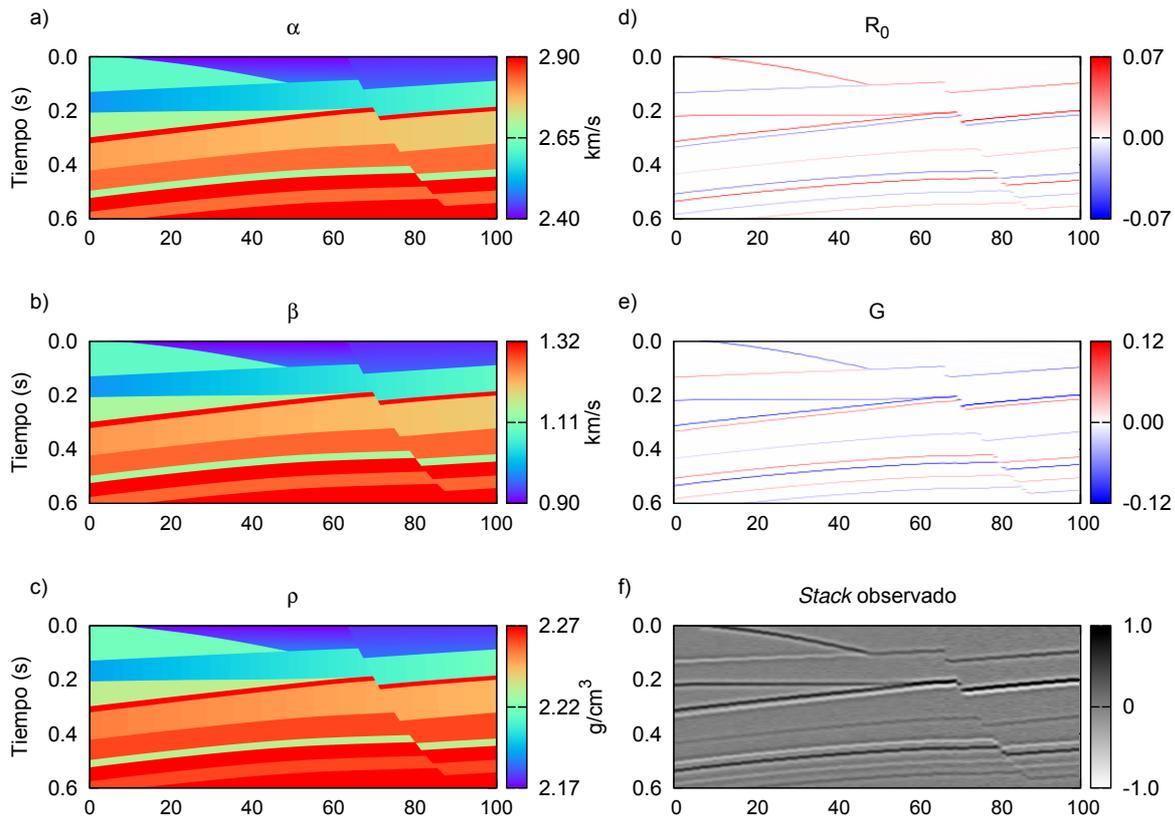


**Figura 6.11.** *Intercept*  $R_0$  y *gradient*  $G$  medios estimados a partir de 100 realizaciones del ruido, sobre el dato con SNR = 10 y SNR = 5, utilizando a)-d) FISTA y e)-h) FISTA+LS.

### 6.3.2. Datos sintéticos 2D

En esta sección se procederá a probar el método propuesto en datos sintéticos 2D. A partir de los modelos de velocidades y densidad de las Figuras 6.12a-c se crearon 100 *angle-gathers* corregidos por NMO, con un intervalo de muestreo de 4 ms. Al dato se le agregó ruido gaussiano con SNR = 5. Para generar el dato se utilizó el modelo convolucional y una ondícula de Ricker con  $f_0 = 30$  Hz de fase cero. Como en el ejemplo 1D, la respuesta AVA se modeló utilizando las ecuaciones de Zoeppritz. Las Figuras 6.12d,e muestran el *intercept* y *gradient* correspondientes al modelo. La Figura 6.12f muestra la sección sísmica (*stack*) del dato observado.

A partir de la ecuación (6.14) se dedujo que  $\mu = \sigma_n^2/\lambda$ . Esta relación permite estimar un valor de  $\mu$  diferente para cada *gather* del dato 2D a partir de nivel de ruido presente en cada uno de ellos y de un valor de  $\lambda$  fijo. Este valor fijo de  $\lambda$  puede ser estimado a partir de un único *gather* por medio de la curva de Pareto y el principio

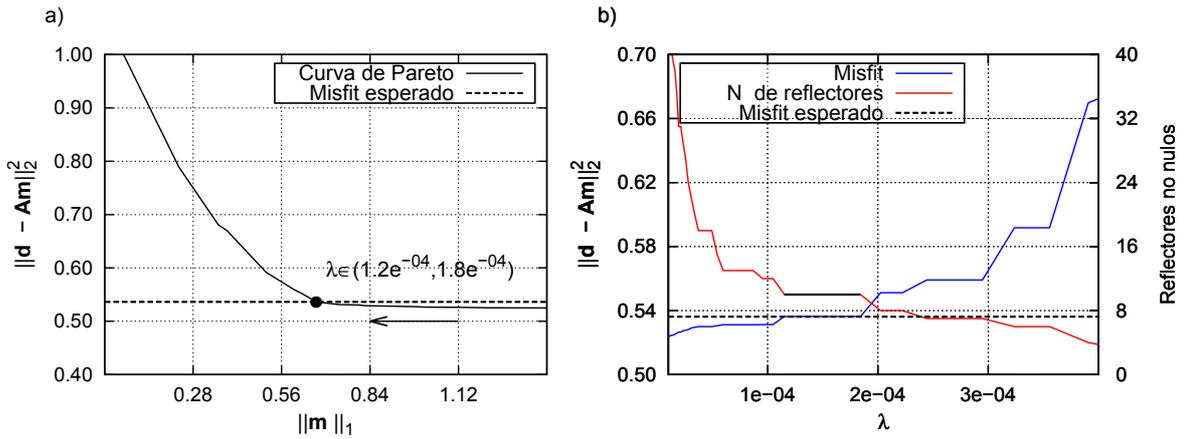


**Figura 6.12.** a), b) y c) modelo de velocidades y densidad 2D. d) y e) *intercept*  $R_0$  y *gradient*  $G$  correspondientes al modelo. f) *Stack* de los datos generados a partir del modelo.

de discrepancia. Esta estrategia permitirá un mejor balance *gather* a *gather* entre los términos de la función de costo dada por la ecuación (6.1), y de esta forma obtener una solución estimada con mejor continuidad lateral. En el método de inversión propuesto en el capítulo anterior, esta estrategia sería similar a tomar un valor diferente de  $L_s$  para cada *gather*, pero al no tener una relación explícita entre  $L_s$  y  $\sigma_n^2$  habría que construir una curva similar a la de la Figura 5.4 para cada *gather*, proceso que no es práctico por su alto costo computacional.

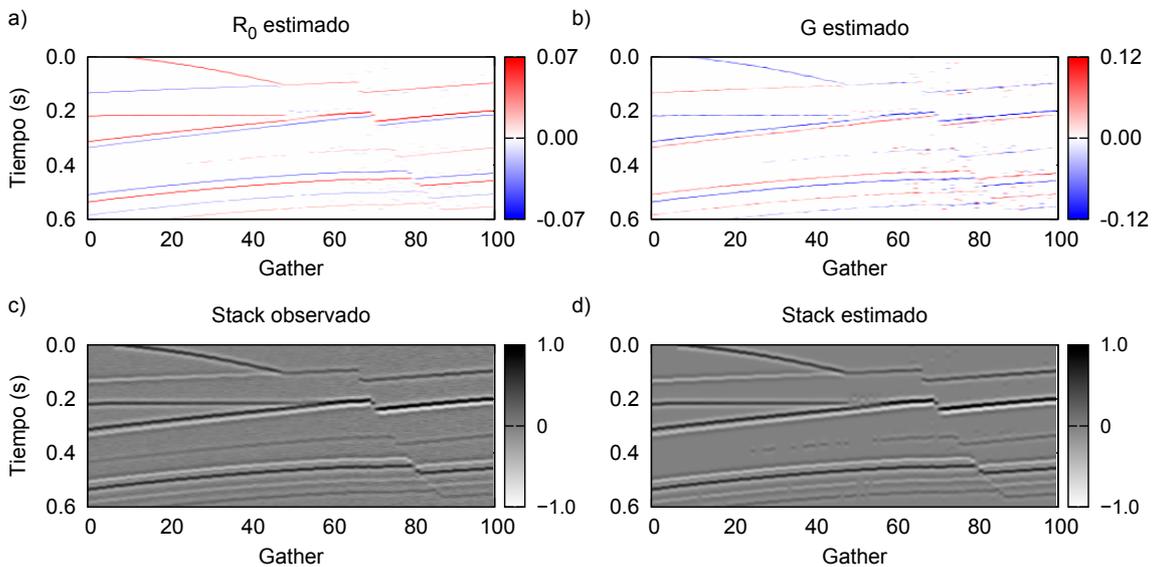
Para estimar el  $\lambda$  con el cual realizar la inversión se construyó la curva de Pareto a partir del primer *gather* del dato 2D (Figura 6.13a). El círculo negro representa los valores de  $\lambda$  óptimos de acuerdo al criterio de discrepancia. La Figura 6.13b muestra el número de reflectores no nulos y el *misfit* en función de  $\lambda$ . Como se puede ver, para los valores de  $\lambda$  representados por el círculo negro en la Figura 6.13a, el *misfit* de las soluciones estimadas está por debajo del nivel de ruido. Sin embargo, el número de reflectores no nulos, representado por el segmento negro en la curva roja, no corresponde al número verdadero, que es  $L_s = 8$ , sino que es mayor. Es decir que para este dato en particular se requieren más reflectores que los verdaderos para alcanzar valores del *misfit* cercanos al valor esperado. Como ya fue analizado, si el número de reflectores no nulos es demasiado grande la inversión comenzará a ser inestable, pero en este caso este valor es cercano al número verdadero. Utilizar un  $\lambda$  que asegure el número correcto

## 6. Inversión *sparse* mediante regularización con la norma $l_1$



**Figura 6.13.** a) Curva de Pareto para el primer *gather* del dato sintético 2D con  $\text{SNR} = 5$ , b) *misfit* y número de reflectores no nulos en función de los parámetros de compensación  $\lambda$  para el dato  $\text{SNR} = 5$ .

de reflectores no nulos nos dará soluciones que no ajustan al dato correctamente. A partir del análisis de la curva de Pareto obtenida a partir del primer *gather* se eligió un valor  $\lambda = 1.5 \times 10^{-4}$ . Luego se obtuvo el  $\mu$  de cada *gather* a partir del valor de  $\sigma_n^2$  correspondiente.



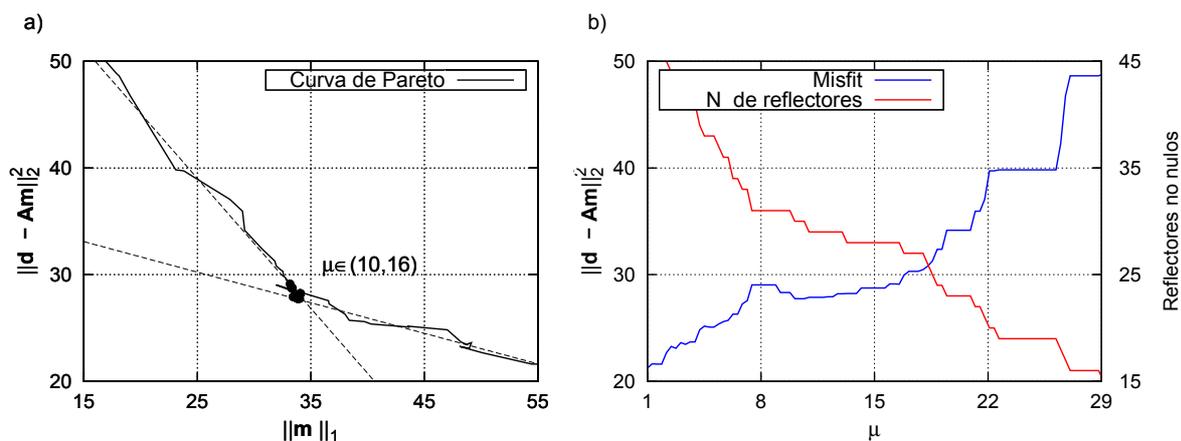
**Figura 6.14.** a) y b) *intercept*  $R_0$  y *gradient*  $G$  estimado a partir del dato sintético 2D, c) y d) *stack* observado y estimado.

En las Figuras 6.14a,b se pueden ver las soluciones *sparse* obtenidas para el *intercept* y el *gradient*. Se aprecia que el método ofrece buenos resultados, conservando la continuidad lateral e invirtiendo satisfactoriamente los parámetros de la aproximación de Shuey de dos términos. Queda además en evidencia la capacidad del método de resolver reflectores cercanos. Como es esperable, los resultados del *gradient* son ligeramente más inestables que los del *intercept*, debido a la falta de robustez del primero

frente al ruido presente en el dato. Los resultados son comparables con aquellos obtenidos mediante VFSA, pero a una fracción considerable del costo computacional. Las Figuras 6.14c,d muestran los *stack* del dato observado y del dato reconstruido a partir de la solución *sparse*. Como puede verse las soluciones honran al dato observado con precisión.

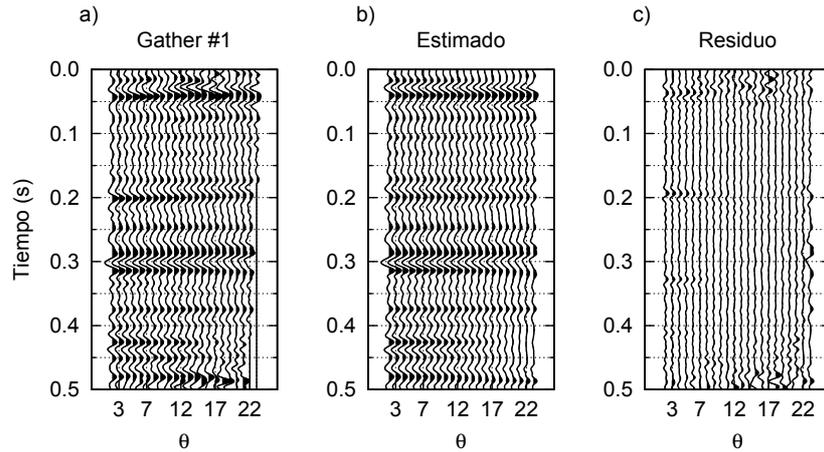
### 6.3.3. Datos de campo

Los datos de campo dificultan la inversión no sólo porque el nivel de ruido no está disponible, sino también porque la ondícula no es conocida. Se asumirá que el dato ha sido apropiadamente acondicionado para preservar las amplitudes y la ondícula es de éste es de fase cero. Los datos de campo consisten en 400 *angle-gathers* corregidos por NMO, con un intervalo de muestreo de 2 ms, una ventana temporal de 0.5 s y un rango de ángulos  $\theta_i = (2^\circ - 24^\circ)$ . En lugar de procesar cada *gather* individualmente, con el fin de mejorar la continuidad lateral de los resultados, la inversión fue realizada en “*super-gathers*”, donde cada *super-gather* es el promedio pesado de cinco *gathers* adyacentes. Una vez realizada la inversión, el arreglo conteniendo el *intercept* y el *gradient* estimados es asignado al *gather* central. La ondícula utilizada durante la inversión fue estimada estadísticamente, para cada *super-gather*, a partir de la raíz cuadrada de la transformada de Fourier de la autocorrelación de las trazas (Robinson y Treitel, 2002).



**Figura 6.15.** a) Curva de Pareto para el primer *gather* del dato de campo, b) *misfit* y número de reflectores no nulos en función del parámetro de compensación  $\mu$ .

En la Figura 6.15a se puede observar la curva de Pareto obtenida a partir del primer *gather* del dato. Si bien, debido a la complejidad del datos de campo, la curva no muestra un quiebre tan marcado como el observado en las pruebas sobre datos sintéticos (Figuras 6.8a,c y 6.13a), se puede observar claramente que ésta está compuesta por dos segmentos aproximadamente rectilíneos. Las líneas discontinuas muestran rectas ajustadas por mínimos cuadrados a ambos segmentos. Los círculos negros se corresponden a valores de *misfit* y norma  $l_1$  de las soluciones obtenidas para un rango de  $\mu \in (10, 16)$ , que se corresponden con los valores donde se produce el quiebre de la curva. La Figura 6.15b muestra el valor del *misfit* y el número de reflectores no nulos



**Figura 6.16.** a) *Gather* observado #1 del dato de campo, b) estimado utilizando FISTA+LS, c) residuo entre el *gather* observado y el estimado.

estimados en función de  $\mu$ . Se puede apreciar que para el rango  $\mu \in (10, 16)$  el número de reflectores no nulos varía entre 28 y 29. También se puede observar que el *misfit* muestra en general una clara tendencia a crecer con  $\mu$ , pero este comportamiento ya no es monótono como si lo era en los ejemplos sobre datos sintéticos. Para los valores de  $\mu \in (10, 16)$  el *misfit* se conserva aproximadamente estable tomando valores entre 28 y 30. Las diferencias de comportamiento en las curvas estimadas respecto de aquellas obtenidas utilizando el dato sintético se deben a la complejidad inherente a los datos de campo. La ondícula es inexacta, la corrección por NMO no es perfecta y el ruido presente no es exactamente gaussiano, entre muchas otras causas.

A partir de la información obtenida de la Figura 6.15 se eligió un valor  $\mu = 15$  para realizar la inversión de todos los *gathers*. Las Figuras 6.16a-c muestran el *gather* a partir del cual se estimaron las curvas de la Figura 6.15, la correspondiente reconstrucción a partir del *intercept* y *gradient* estimado utilizando  $\mu = 15$  y la diferencia entre ambos. Como se puede observar las soluciones estimadas honran al dato observado, mostrando mayores diferencias en aquellos tiempos donde el dato es muy ruidoso, como al principio y al final del *gather*. En las Figuras 6.17a,b se pueden ver las imágenes correspondientes al *intercept* y el *gradient* estimadas a partir de todos los *gathers*. Como se puede apreciar el método propuesto logró obtener soluciones de tipo *sparse* para ambos atributos. Dichas soluciones presentan buena continuidad lateral y honran al dato observado, como puede verse de la comparación entre el *stack* observado, presentado en la Figura 6.17b, y el *stack* estimado a partir de las soluciones obtenidas, presentado en la Figura 6.17c. Al igual que en la inversión realizada utilizando VFSA (Figura 5.12), se observa una considerable diferencia de magnitud entre el *intercept* y el *gradient* estimado debido a fenómenos ajenos al método utilizado.

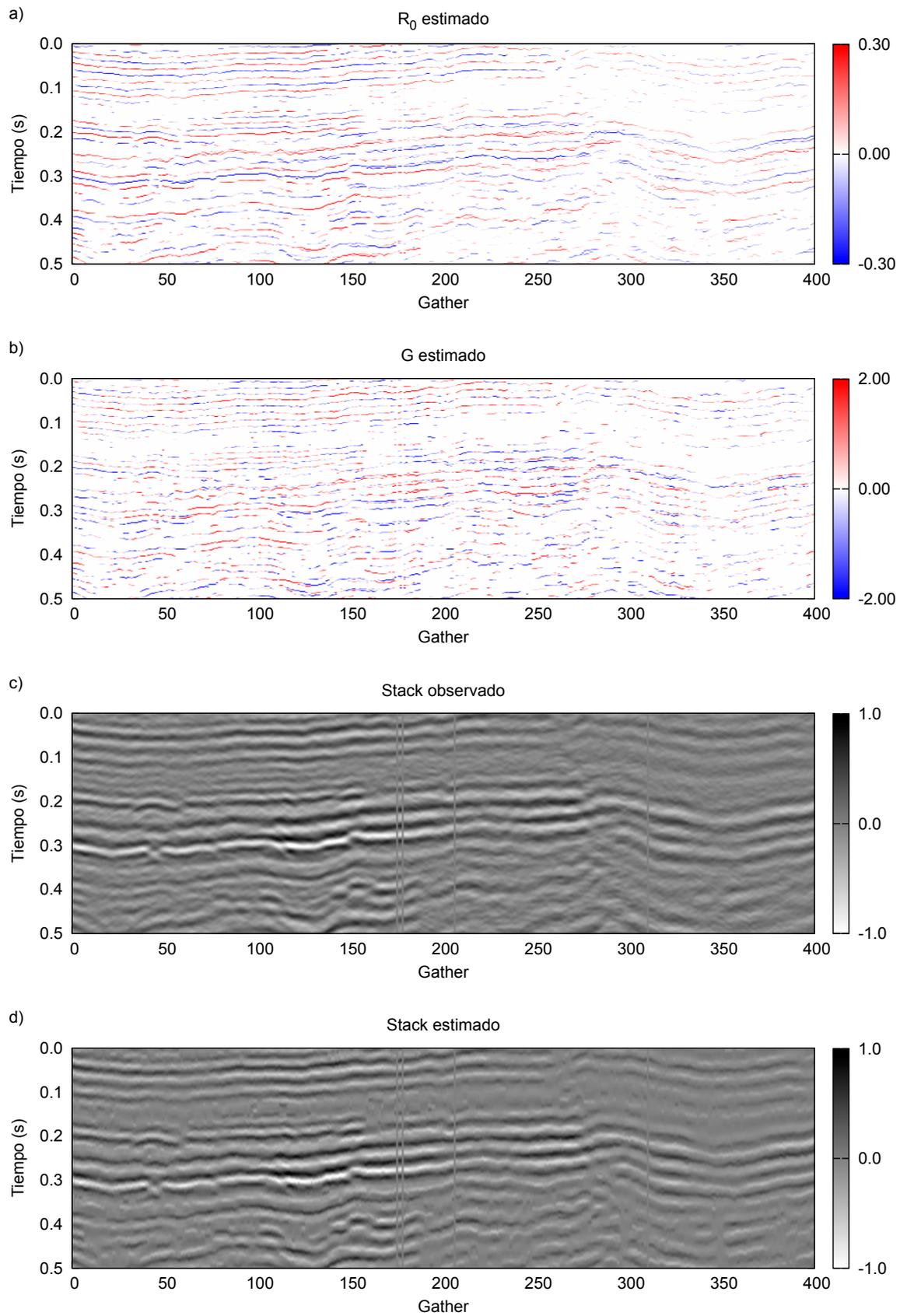


Figura 6.17. a) y b) *Intercept*  $R_0$  y *gradient*  $G$  estimado a partir del dato de campo, c) y d) *stack* observado y estimado.

### 6.4. Conclusiones

A continuación se enumeran las principales conclusiones obtenidas utilizando la estrategia propuesta sobre datos sintéticos y datos de campo:

- Resolver el problema regularizado utilizando la norma  $l_1$  (LASSO) no permite obtener una adecuada solución del atributo *gradient*, ya que éste tiene una fuerte dependencia con el nivel de ruido presente en el dato y el valor del parámetro de compensación  $\mu$  elegido para llevar a cabo la inversión.
- El uso de las soluciones obtenidas al resolver el LASSO como información *a priori* para realizar una segunda etapa correctiva mediante mínimos cuadrados, demostró ser una estrategia apropiada para obtener atributos *sparse* que honran al dato observado.
- Sobre datos sintéticos 1D con niveles de ruido significativo la estrategia propuesta demostró ser robusta y estable, siendo capaz de estimar soluciones *sparse* muy precisas.
- La curva de Pareto obtenida mediante la estrategia de dos etapas muestra un quiebre asociado al valor óptimo de  $\mu$ , según el principio de discrepancia, que facilita su estimación.
- Los pequeños valores de las desviaciones estándar de las magnitudes invertidas resaltan la consistencia de las soluciones obtenidas con la estrategia de dos etapas.
- Las pruebas realizadas sobre datos sintéticos 2D muestran que el método es capaz de obtener soluciones *sparse* que honran al dato y presentan buena continuidad lateral.
- Pruebas en datos de campo mostraron que la técnica propuesta es capaz de ofrecer imágenes *sparse* del *intercept* y del *gradient* que honran el dato observado con precisión.
- Debido a que se utiliza el algoritmo FISTA, que sólo realiza multiplicación de matrices y vectores, la técnica propuesta es considerablemente más económica computacionalmente que la propuesta en el capítulo anterior, sin que esto signifique un detrimento significativo en la calidad de las soluciones estimadas.

# Capítulo 7

## Inversión *sparse* mediante regularización con la norma $l_{2,1}$

### 7.1. Introducción

En los capítulos previos se resolvió el problema utilizando la aproximación de Shuey de dos términos (Shuey, 1985), ubicando en tiempo la posición de los coeficientes de reflexión y luego ajustando por mínimos cuadrados la amplitud de los coeficientes de dicha aproximación. Sin embargo, los problemas de ambigüedad, inestabilidad y subestimación de las amplitudes estimadas analizados y parcialmente resueltos en los capítulos anteriores, se vuelven más críticos al utilizar una aproximación de tres términos debido a que las incertidumbres asociadas a los parámetros del modelo son mucho más grandes (Downton y Lines, 2001, e.g.).

En este capítulo utilizaremos la aproximación de Aki y Richards de tres términos (Aki y Richards, 1980) para obtener soluciones *blocky* de los parámetros físicos a partir de datos sísmicos *prestack*. La estimación de soluciones *blocky* de  $\alpha$ ,  $\beta$  y  $\rho$  es un problema equivalente a la estimación de soluciones *sparse* de las correspondientes reflectividades dadas por la ecuación de Aki y Richards. El método se basa en la utilización de restricciones específicas para imponer *blockiness* y *sparseness* donde la información *a priori* indique que los parámetros físicos deban ser constantes por tramos y las reflectividades *sparse*.

De manera similar a las estrategias propuestas en los capítulos anteriores, donde se regularizó el problema de inversión mediante la norma  $l_0$  y  $l_1$ , en este capítulo utilizamos como regularización la norma mixta  $l_{2,1}$ . A diferencia de la norma  $l_1$  que sólo promueve soluciones *sparse*, las normas mixtas promueven soluciones *sparse* con estructura (propiedad llamada *group-sparsity*) (Kowalski y Torrèsani, 2009; Gramfort et al., 2012; Kowalski et al., 2013). En el contexto del presente trabajo, una de las ventajas de utilizar la norma  $l_{2,1}$  es que permite estimar soluciones *sparse* de las reflectividades en el tiempo (es decir soluciones *blocky* de los parámetros físicos) y, a la vez, permite incorporar fácilmente restricciones a las amplitudes de los parámetros estimados por medio de la utilización de una matriz de covarianza o una matriz de escala, si dicha información está disponible. Esto ayuda a mitigar la subestimación de las amplitudes asociadas a las soluciones *sparse* como, por ejemplo, sucedió en el capítulo anterior al realizar la inversión sin el *debiasing step*. En el caso de inversión *prestack* de tres

términos, la norma  $l_{2,1}$  promueve *group-sparsity* de los parámetros de la aproximación utilizada, asegurando que las reflectividades resultantes estén correlacionadas incluso cuando no se utiliza una matriz de covarianza. Este es un aspecto importante ya que aunque se cuente con información de pozo suficiente, hecho que no siempre sucede, las matrices de covarianza no son fáciles de construir (Downton, 2005; Theune et al., 2010). Por otro lado, la estrategia propuesta incluye un término adicional en la función de costo conteniendo información de la tendencia de baja frecuencia de los parámetros físicos. Esta información es necesaria para una correcta estimación de los mismos. El método propuesto es económico en términos de costo computacional debido a que la norma mixta  $l_{2,1}$  lleva a una función de costo convexa que puede ser minimizada utilizando algoritmos proximales, como el algoritmo FISTA que ya hemos utilizado en el capítulo anterior.

Este capítulo está organizado de la siguiente manera. Primero se explica en detalle el método propuesto, estableciendo las hipótesis necesarias y definiendo todas las ecuaciones relevantes. Luego se probará el método en datos sintéticos *prestack* 1D y 2D, asumiendo la validez de la aproximación de Aki y Richards de tres términos. Se hará una comparación entre los resultados obtenidos con el método propuesto y los resultados obtenidos utilizando, como regularización, las norma  $l_1$  y la norma  $l_2$ . Además se analizará el impacto de utilizar una matriz de covarianza como información *a priori*. Se verá que utilizar la norma  $l_{2,1}$  permite mejorar las soluciones cuando la matriz de covarianza no está disponible o sólo está disponible parcialmente. Finalmente se probará el método sobre datos de campo.

## 7.2. Teoría

Como fue analizado en detalle en el capítulo 3, la aproximación de Aki y Richards (1980) de tres términos permite expresar al coeficiente de reflexión de una onda plana  $P$  que incide en la discontinuidad que separa dos medios, en función del  $i$ -ésimo ángulo de incidencia  $\theta_i$ , como

$$R_{PP}(\theta_i) = \frac{1}{2} (1 + \tan^2 \theta_i) R_\alpha - (4\gamma^2 \sin^2 \theta_i) R_\beta + \frac{1}{2} (1 - \gamma^2 \tan^2 \theta_i) R_\rho, \quad (7.1)$$

donde,  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  son las reflectividades asociadas a  $\alpha$ ,  $\beta$  y  $\rho$  respectivamente (ecuaciones (3.46), (3.47) y (3.48)), y  $\gamma$  es la relación entre la velocidad de corte  $\beta$  y la velocidad compresional  $\alpha$  (ecuación (3.50)).

Cuando el contraste entre los parámetros físicos a cada lado de las discontinuidades que separan los medios son pequeños comparados con sus valores absolutos se pueden establecer relaciones entre dichos parámetros y sus correspondientes reflectividades dadas por

$$\sum_{k=1}^l R_{\alpha_k} \approx e_{\alpha_l}, \quad \sum_{k=1}^l R_{\beta_k} \approx e_{\beta_l}, \quad \sum_{k=1}^l R_{\rho_k} \approx e_{\rho_l}, \quad l = 1, \dots, L_r, \quad (7.2)$$

donde  $e_{\alpha_l} = \ln(\alpha_l/\alpha_1)$ ,  $e_{\beta_l} = \ln(\beta_l/\beta_1)$ , y  $e_{\rho_l} = \ln(\rho_l/\rho_1)$ , y  $\alpha_1$ ,  $\beta_1$  y  $\rho_1$  son los correspondientes valores de los parámetros físicos para  $l = 1$ . A modo de ejemplo,

$$R_{\alpha_k} = \frac{\Delta\alpha_k}{\alpha_k} = \Delta \log(\alpha_k), \quad (7.3)$$

para algún  $k$ , entonces

$$\sum_{k=1}^l R_{\alpha_k} \simeq \log\left(\frac{\alpha_l}{\alpha_1}\right) = e_{\alpha l}, \quad (7.4)$$

Como se verá más adelante  $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_{L_r})^T$ , siendo  $\mathbf{e}_l = (e_{\alpha l}, e_{\beta l}, e_{\rho l})$ , puede ser utilizado como una restricción al modelo que permitirá incorporar la tendencia de baja frecuencia de los parámetros físicos, información que no está presente en el dato sísmico. En la práctica esta información puede ser obtenida aplicando un filtro pasabajos a valores de los parámetros físicos obtenidos de un pozo, si esta información está disponible.

Como en los casos estudiados en los capítulos anteriores, basaremos nuestra estrategia de inversión en el modelo convolucional, por lo tanto asumiremos que para cada ángulo de incidencia  $\theta_i$ , con  $i = 1, \dots, N_\theta$ , el dato observado  $\mathbf{d}(\theta_i)$  tiene la forma

$$\mathbf{d}(\theta_i) = \mathbf{A}(\theta_i)\mathbf{m}, \quad (7.5)$$

donde  $\mathbf{m} = (R_{\alpha_1}, R_{\beta_1}, R_{\rho_1}, \dots, R_{\alpha_{L_r}}, R_{\beta_{L_r}}, R_{\rho_{L_r}})^T$ , la matriz  $\mathbf{A}$  tiene dimensión  $L_d \times nL_r$ , con  $n = 3$ , y contiene los términos que dependen del ángulo  $\theta_i$  de la ecuación (7.1) y los elementos de la ondícula  $\mathbf{w}$ .

Los  $N_\theta$  sistemas de ecuaciones dados por la ecuación (7.5) forman un único sistema dado por

$$\mathbf{d} = \mathbf{A}\mathbf{m}, \quad (7.6)$$

donde  $\mathbf{A}$  es una matriz columna por bloques, donde cada bloque está dado por  $\mathbf{A}(\theta_i)$ , y  $\mathbf{d}$  es un vector columna por bloques, donde cada bloque está dado por  $\mathbf{d}(\theta_i)$ , para  $i = 1, \dots, N_\theta$ .

Para obtener la función de costo a minimizar nos valdremos del teorema de Bayes. Es decir, buscaremos la función de probabilidad *a priori*  $p(\mathbf{m})$  y la función de verosimilitud  $p(\mathbf{d}|\mathbf{m})$  que nos conducirá a la función de probabilidad *a posteriori*

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \quad (7.7)$$

En el contexto del método que desarrollaremos la función de distribución *a priori* representará la información conocida de los parámetros físicos. Esta información tiene que ser independiente del dato sísmico y debe ser obtenida desde fuentes independientes, como por ejemplo un pozo.

Siguiendo los trabajos de Jensås (2008) y Theune et al. (2010), la distribución *a priori* puede ser expresada como el producto de una distribución “lateral”  $p_{lat}(\mathbf{m})$  y una distribución “vertical”  $p_{vert}(\mathbf{m})$ :

$$p(\mathbf{m}) = p_{lat}(\mathbf{m})p_{vert}(\mathbf{m}). \quad (7.8)$$

Asumiendo que  $\mathbf{e}$  tienen un comportamiento gaussiano (Buland y Omre, 2003), y por medio de las relaciones dadas por la ecuación (7.2), se tiene

$$p_{lat}(\mathbf{m}) = \frac{1}{(2\pi)^{L_r} |\boldsymbol{\Sigma}_l|^{1/2}} \exp \left[ -\frac{1}{2} \sum_{l=1}^{L_r} (\mathbf{e}_l - \mathbf{I}_l \mathbf{m})^T \boldsymbol{\Sigma}_l^{-1} (\mathbf{e}_l - \mathbf{I}_l \mathbf{m}) \right], \quad (7.9)$$

donde  $\mathbf{I}_l \mathbf{m}$  es la forma matricial del lado izquierdo de las ecuaciones (7.2) y

$$\boldsymbol{\Sigma}_l = \begin{pmatrix} \sigma_{e_{\alpha l}}^2 & 0 & 0 \\ 0 & \sigma_{e_{\beta l}}^2 & 0 \\ 0 & 0 & \sigma_{e_{\rho l}}^2 \end{pmatrix}. \quad (7.10)$$

Es importante notar que  $p_{lat}(\mathbf{x})$  incorpora la información de la tendencia de baja frecuencia y las correspondiente varianzas o “escalas” de los diferente parámetros físicos del modelo. Al igual que la tendencia de baja frecuencia, la información de la escala podría ser estimada a partir de información obtenida de un pozo.

Por otro lado, necesitamos desarrollar una expresión para la función de distribución *a priori* vertical  $p_{vert}(\mathbf{m})$ . Para ello se tiene en cuenta el siguiente análisis. Dado que el problema de estimar soluciones *blocky* para  $\alpha$ ,  $\beta$  y  $\rho$  es equivalente al problema de estimar soluciones *sparse* de las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ , se pretende encontrar un vector  $\mathbf{m}$  tal que la función *a posteriori*  $p(\mathbf{d}|\mathbf{m})$ , dada por la ecuación (2.26), sea maximizada. Que el problema a resolver esté mal planteado y mal condicionado nos indica la necesidad de favorecer soluciones estables y con significado físico a través de una regularización apropiada. Regularizar el problema mediante la norma  $l_1$ , como fue hecho en el capítulo anterior, surge como la opción más lógica, pero como se verá en los ejemplos numéricos no permite obtener soluciones correctas. Debido a esto proponemos la utilización de una norma mixta  $l_{p,q}$  (Kowalski y Torrèsani, 2009). Como se verá en las próximas secciones, este tipo de regularización nos permitirá obtener soluciones *sparse* con estructura, una cualidad que describe correctamente las soluciones esperadas.

### Normas mixtas

Sea  $\mathbf{v}$  un vector genérico cuyos elementos están organizados en  $G$  grupos, cada uno de estos formado por  $M$  miembros (Figura 7.1a). Dado un vector de pesos  $\boldsymbol{\omega}$  tal que  $\omega_{g,m} > 0$  para todo  $g, m$ , entonces la norma mixta pesada  $l_{\boldsymbol{\omega};p,q}$  del vector  $\mathbf{v}$  se puede expresar como

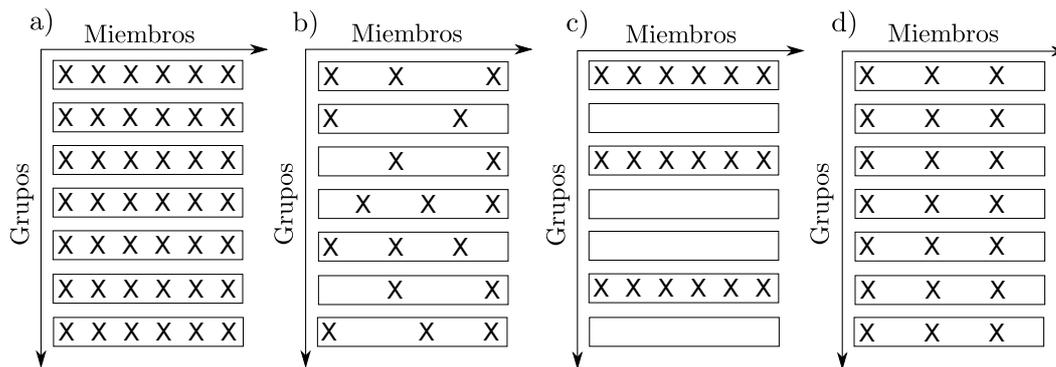
$$l_{\boldsymbol{\omega};p,q} = \|\mathbf{v}\|_{\boldsymbol{\omega};p,q} = \left( \sum_{g=1}^G \left( \sum_{m=1}^M \omega_{g,m} |v_{g,m}|^p \right)^{q/p} \right)^{1/q}. \quad (7.11)$$

Es de remarcar que si  $p, q \geq 1$ , entonces se puede demostrar que  $\|\mathbf{v}\|_{\boldsymbol{\omega};p,q}$  es convexo (Kowalski, 2009), una característica muy importante desde el punto de vista computacional para asegurar la optimalidad global. Es más, cuando  $p = q$ ,

$$\|\mathbf{v}\|_{\boldsymbol{\omega};p,q} = \|\mathbf{v}\|_{\boldsymbol{\omega};p} = \left( \sum_g \sum_m \omega_{g,m} |v_{g,m}|^p \right)^{1/p} = \left( \sum_{g,m} \omega_{g,m} |v_{g,m}|^p \right)^{1/p}, \quad (7.12)$$

es decir, las normas mixtas son una generalización de las normas  $l_p$  convencionales.

La principal motivación para usar normas mixtas como término de regularización es que, a diferencia de las normas  $l_p$ , las normas mixtas  $l_{p,q}$  permiten incorporar cierta



**Figura 7.1.** Un ejemplo de las estructuras promovidas por las normas mixtas: a) elementos de un dado vector  $\mathbf{v}$  organizados en  $G$  grupos de  $M$  miembros (solución por mínimos cuadrados), b) soluciones *sparse* promovidas por la norma  $l_1$ , c) soluciones *sparse* entre grupos promovidas por la norma  $l_{2,1}$ , d) soluciones *sparse* dentro de cada grupo promovida por la norma  $l_{1,2}$ .

dependencia o estructura entre los parámetros involucrados. Cuando se minimiza la norma  $l_p$  con  $p < 2$  se promueven soluciones *sparse*, pero contrariamente, este tipo de soluciones son penalizadas si  $p > 2$  (Debye y van Riel, 1990). En particular, cuando  $p = q = 2$  y todos los pesos son iguales a la unidad, la norma mixta es igual a la norma  $l_2$  usada en la regularización de Tikhonov convencional, que lleva a la solución de mínimos cuadrados amortiguados. Esta solución está representada en la Figura 7.1a, donde todos los elementos de  $\mathbf{v}$  son uniformes. Cuando  $p = q = 1$  estamos en presencia de la norma  $l_1$  pesada (Khajehnejad et al., 2011), que cuando es utilizada como regularización favorece las soluciones *sparse*, como se representa en la Figura 7.1b. Notar que estas dos soluciones no presentan ninguna clase de estructura en particular, excepto aquella que podría llegar a aparecer por forzar a la solución a honrar al dato observado, hecho que no está representado en la figura.

Utilizando las normas mixtas  $l_{p,q}$  es posible obtener soluciones que presenten formas estructuradas de *sparsity*, controladas por la elección de  $p$  y  $q$ . En el desarrollo del método propuesto en este capítulo vamos a utilizar la norma mixta  $l_{2,1}$ , también conocida como *group-sparsity*, *group-lasso*, o *multiple measurement vector* o *joint sparsity* (Yuan et al., 2006). A fines prácticos, la norma mixta  $l_{2,1}$  es una composición de la norma  $l_1$  y la norma  $l_2$ , como puede verse en la ecuación (7.11). El concepto clave es que la regularización mediante la norma  $l_1$  se realiza entre grupos, mientras que la regularización mediante la norma  $l_2$  se realiza dentro de cada grupo. Como consecuencia, la minimización de la norma mixta  $l_{2,1}$  promueve soluciones *sparse* entre los grupos, pero promueve uniformidad dentro de cada grupo (Figure 7.1c). Por completitud, si consideramos la minimización de la norma mixta  $l_{1,2}$ , se promoverían soluciones *sparse* dentro de cada grupo y uniformidad entre los grupos (Figure 7.1d).

Siguiendo la ecuación (7.11), la norma mixta  $l_{2,1}$  pesada de un dado vector  $\mathbf{v}$  puede expresarse como

$$\|\mathbf{v}\|_{2,1} = \sum_{g=1}^G \left( \sum_{m=1}^M \omega_{g,m} v_{g,m}^2 \right)^{\frac{1}{2}} = \sum_{g=1}^G \|\boldsymbol{\omega}_g^{1/2} \cdot \mathbf{v}_g\|_2, \quad (7.13)$$

## 7. Inversión *sparse* mediante regularización con la norma $l_{2,1}$

donde  $\mathbf{v}_g = (v_{g,1}, \dots, v_{g,M})$  y  $\boldsymbol{\omega}_g^{1/2} = (\sqrt{\omega_{g,1}}, \dots, \sqrt{\omega_{g,M}})$ . En el contexto de la inversión de datos sísmicos *prestack* utilizando la ecuación de Aki y Richards de tres términos, cada grupo (muestra  $l$  del vector  $\mathbf{r}$  de los coeficiente de reflexión) estará compuesto por tres miembros:  $R_{\alpha l}$ ,  $R_{\beta l}$  y  $R_{\rho l}$ . Por lo tanto, la regularización a través de la norma  $l_{2,1}$  propuesta promoverá soluciones *sparse* en el tiempo y reflectividades restringidas en el sentido de que  $R_{\alpha l}$ ,  $R_{\beta l}$  y  $R_{\rho l}$  tenderán a ser nulos o distintos de cero, de forma simultánea. Pretender soluciones de este tipo tiene sentido desde un punto de vista geológico, ya que siempre que exista un contraste de  $\alpha$  en general se espera que exista contrastes de  $\beta$  y  $\rho$ , y de la misma forma, si no hay cambios en las velocidades es esperable que no haya cambios en la densidad. Además, los pesos  $w_{g,m}$  pueden ser utilizados para incorporar información de pozo para restringir aún más el espacio de soluciones.

Podemos incorporar la norma mixta  $l_{2,1}$  al enfoque bayesiano propuesto a través de la siguiente distribución (Babacan et al., 2014):

$$p_{vert}(\mathbf{m}) \propto \prod_{l=1}^{L_r} \exp \left\{ -\frac{1}{2\lambda} \sqrt{\mathbf{m}_l^T \boldsymbol{\Omega}_l^{-1} \mathbf{m}_l} \right\}, \quad (7.14)$$

donde  $\mathbf{m}_l = (R_{\alpha l}, R_{\beta l}, R_{\rho l})$ ,  $\lambda$  es un parámetro de compensación, y  $\boldsymbol{\Omega}_l$  es la matriz de covarianzas de las reflectividades

$$\boldsymbol{\Omega}_l = \begin{pmatrix} \sigma_{R_{\alpha l}}^2 & \nu_{R_{\alpha l} R_{\beta l}} & \nu_{R_{\alpha l} R_{\rho l}} \\ \nu_{R_{\beta l} R_{\alpha l}} & \sigma_{R_{\beta l}}^2 & \nu_{R_{\beta l} R_{\rho l}} \\ \nu_{R_{\rho l} R_{\alpha l}} & \nu_{R_{\rho l} R_{\beta l}} & \sigma_{R_{\rho l}}^2 \end{pmatrix}, \quad l = 1, \dots, L_r. \quad (7.15)$$

Como es usual,  $\lambda$  representa un parámetro utilizado para balancear el impacto de la regularización, y debe ser estimado antes de la inversión. La estimación de las matrices  $\boldsymbol{\Omega}_l$  también debe ser realizada antes de la inversión. Claramente, estas matrices proveen fuertes restricciones a la estructura de la solución estimada. Sin embargo, si la matriz de covarianza no está disponible o la incertidumbre asociada a ésta es grande, el uso de una matriz de escala simple (por ejemplo, asumiendo que las covarianzas son nulas), junto con la minimización de la norma  $l_{2,1}$  llevará a soluciones aceptables que presentan *group-sparsity*, como se verá en los ejemplos numéricos.

Al igual que en los ejemplos estudiados en la sección 2.6 se asumirá que el ruido presente en el dato observado tiene una distribución de probabilidad gaussiana, que dicho ruido no está correlacionado y tiene varianza  $\sigma_n^2$ , por lo tanto la función de verosimilitud  $p(\mathbf{d}|\mathbf{m})$  será una función de densidad de probabilidad gaussiana de la forma (Sivia, 1996)

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{(2\pi)^{L_d} \sigma_n} \exp \left[ -\frac{1}{2\sigma_n^2} (\mathbf{d} - \mathbf{A}\mathbf{m})^T (\mathbf{d} - \mathbf{A}\mathbf{m}) \right]. \quad (7.16)$$

Ya estando definidas la función de verosimilitud  $p(\mathbf{d}|\mathbf{m})$  y las funciones de probabilidad *a priori*  $p_{lat}(\mathbf{m})$  y  $p_{vert}(\mathbf{m})$ , ecuaciones (7.16), (7.9) y (7.14) respectivamente, el modelo  $\mathbf{m}$  que maximiza la función *a posteriori* dada por la ecuación (2.26) puede ser obtenido minimizando la función de costo dada por

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + \sigma_n^2 \sum_{l=1}^{L_r} \|\Sigma_l^{-1/2}(\mathbf{e}_l - \mathbf{I}_l \mathbf{m})\|_2^2 + \mu \|\Omega_l^{-1/2} \mathbf{m}_l\|_{2,1}, \quad (7.17)$$

donde  $\mu = \sigma_n^2/\lambda$  es un nuevo, pero único, parámetro de compensación que deberá ser estimado por medio de algún criterio apropiado.

Tomando  $\mathbf{y}_l = \Omega_l^{-1/2} \mathbf{m}_l$ , esta función de costo se puede expresar como

$$J(\mathbf{y}) = \|\mathbf{f} - \mathbf{C}\mathbf{y}\|_2^2 + \mu \|\mathbf{y}\|_{2,1}, \quad (7.18)$$

donde  $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_{L_r})^T$ , la matriz  $\mathbf{C}$  está formada a partir de las matrices  $\Omega_l^{-1/2}$ ,  $\mathbf{I}_l$ ,  $\Sigma_l^{-1/2}$  y  $\mathbf{A}$ , y el vector  $\mathbf{f}$  está formado a partir de los vectores  $\mathbf{d}$  y  $\mathbf{e}_l$ .

La función de costo dada por la ecuación (7.18) puede ser eficientemente minimizada utilizando FISTA (Beck y Teboulle, 2009). Si bien FISTA fue, en principio, desarrollado para minimizar problemas regularizados con la norma  $l_1$ , redefiniendo la función umbral dada en la ecuación (6.3) como

$$T_\beta\{g\} = \begin{cases} g(1 - \beta/\|g\|_2), & \|g\|_2 \geq \beta \\ 0, & \|g\|_2 < \beta \end{cases} \quad (7.19)$$

es aplicable a problemas regularizados con la norma mixta  $l_{2,1}$  (Fornasier y Rauhut, 2008; Kowalski et al., 2013).

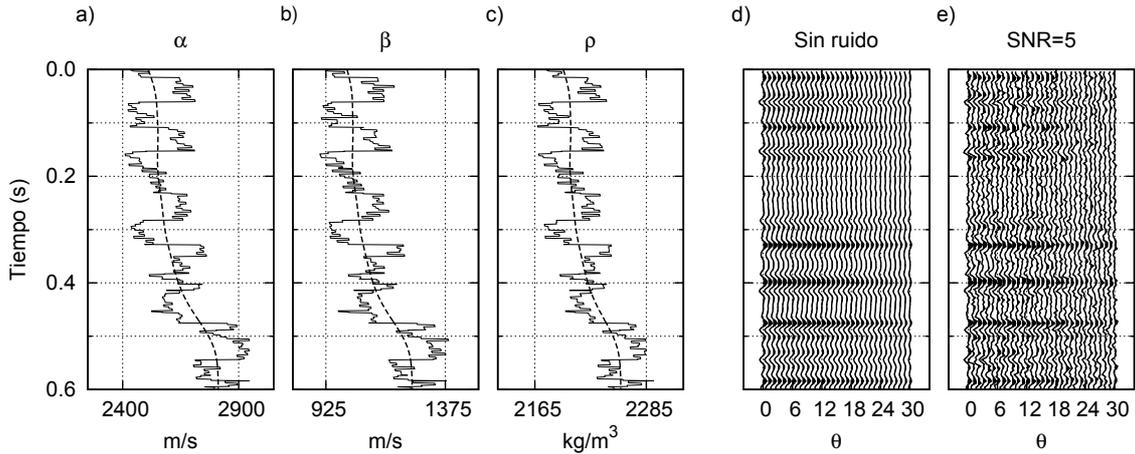
Una vez obtenido el minimizador  $\hat{\mathbf{y}}$  de la función de costo dada por la ecuación (7.18) la solución *blocky* de los parámetros  $\alpha$ ,  $\beta$  y  $\rho$  puede ser obtenida mediante la ecuación (7.2), previo cambio de variable de  $\hat{\mathbf{y}}$  a  $\hat{\mathbf{m}}$  por medio de las matrices  $\Omega_l^{-1/2}$ .

## 7.3. Ejemplos

En esta sección se probará el método propuesto sobre datos sintéticos 1D y 2D y sobre datos de campo. A diferencia de los capítulos anteriores donde las inversiones fueron realizadas considerando la aproximación de Shuey de dos términos, en estos ejemplos nos concentraremos en la aproximación de Aki y Richards de tres términos. Cuando la información *a priori* disponible lo permita se obtendrán soluciones *blocky* de los parámetros  $\alpha$ ,  $\beta$  y  $\rho$ , en caso contrario se estimarán soluciones *sparse* de las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ .

### 7.3.1. Datos sintéticos 1D

En este ejemplo se procederá a probar en datos sintéticos 1D el algoritmo propuesto con el fin de obtener soluciones *blocky* de  $\alpha$ ,  $\beta$  y  $\rho$ . Es decir, se procederá a resolver el problema de inversión minimizando la ecuación (7.17). El dato consiste en un *angle-gather* generado a partir del modelo de velocidades y densidad de las Figuras 7.2a-c. Este modelo fue creado en base al modelo Marmousi2 (Martin et al., 2006). Las líneas discontinuas muestran las tendencias de baja frecuencia obtenidas al aplicar un filtro pasa-bajos a los valores de  $\alpha$ ,  $\beta$  y  $\rho$ , indicados con las líneas sólidas. El *angle-gather* se compone de 31 trazas, abarcando un rango de ángulos  $\theta_i \in (0^\circ, 30^\circ)$ . El intervalo de



**Figura 7.2.** a), b), c) Las líneas continuas indican el modelo, mientras que las líneas discontinuas indican las tendencias de baja frecuencia. d) Dato sintético libre de ruido, e) dato sintético con  $SNR = 5$ .

muestreo es de 4 ms. El dato fue creado basado en el modelo convolucional, utilizando una ondícula de Ricker de  $f_0 = 30$  Hz. La respuesta AVA fue modelada utilizando las ecuaciones de Zoeppritz. Al dato se le agregó ruido gaussiano no correlacionado con  $SNR = 5$ . Las Figuras 7.2d,e muestran respectivamente el dato libre de ruido y el dato con ruido. Como puede apreciarse, el nivel del ruido agregado es lo suficientemente elevado como para perturbar considerablemente la respuesta AVA de algunos reflectores.

### Estimación de las matrices de covarianza

Para aplicar el método propuesto primero es necesario estimar las matrices  $\Sigma_l$ ,  $\Omega_l$  y el parámetro de compensación  $\mu$ . En este ejemplo ambas matrices serán consideradas invariantes en la ventana de tiempo en la cual se realizará la inversión, por lo tanto  $\Sigma_l = \Sigma$  y  $\Omega_l = \Omega$ . Una adecuada estimación de estas matrices es crucial para balancear los primeros dos términos de la ecuación (7.17) y de esta forma obtener soluciones que honren la información *a priori*. Si las varianzas en  $\Sigma$  son demasiado pequeñas, el segundo término de la función de costo estará demasiado pesado y las soluciones estimadas estarán demasiado cercanas a la tendencia de baja frecuencia, representada por e. Como consecuencia se perderá la característica *blocky* de las soluciones estimadas y el dato observado no se honrará correctamente. Contrariamente, si las varianzas son demasiado grandes el segundo término tendrá poco peso, las soluciones estimadas no seguirán la tendencia de baja frecuencia y el dato estará sobre-ajustado. En este ejemplo, estimaremos la matriz  $\Sigma$  directamente de la diferencia entre los logaritmos del modelo y la correspondiente tendencia de baja frecuencia. Con respecto a la matriz  $\Omega$ , una incorrecta estimación de ésta puede llevar a estimaciones de los parámetros físicos que no estarán correlacionados con los parámetros físicos del subsuelo, incluso cuando el modelo estimado honre correctamente al dato observado. Para estimar dicha matriz seguiremos el método propuesto por Downton (2005). Para una ventana de tiempo relativamente pequeña,  $\Omega$  puede expresarse como

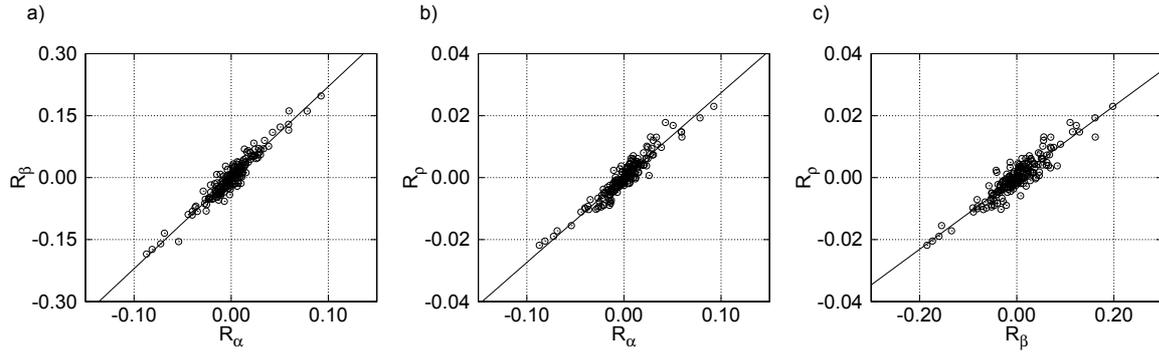


Figura 7.3. Crossplots entre a)  $R_\alpha$  y  $R_\beta$ , b)  $R_\alpha$  y  $R_\rho$  y c)  $R_\beta$  y  $R_\rho$

$$\mathbf{\Omega} = \sigma_{R_\alpha}^2 \begin{pmatrix} 1 & \frac{\tau_{R_\alpha R_\beta}^2}{a} & b \\ \frac{\tau_{R_\alpha R_\beta}^2}{a} & \left(\frac{\tau_{R_\alpha R_\beta}}{a}\right)^2 & c \left(\frac{\tau_{R_\alpha R_\beta}}{a}\right)^2 \\ b & c \left(\frac{\tau_{R_\alpha R_\beta}}{a}\right)^2 & \left(\frac{b}{\tau_{R_\alpha R_\rho}}\right)^2 \end{pmatrix}, \quad (7.20)$$

donde  $a$ ,  $b$  y  $c$  son estimados por medio de un ajuste de mínimos cuadrados a partir de los correspondientes *crossplots* de las reflectividades, y  $\tau_{R_\alpha R_\beta}$  y  $\tau_{R_\alpha R_\rho}$  son los correspondientes coeficientes de correlación. Las Figuras 7.3a-c muestran los *crossplots* entre  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  para el modelo de la Figura 7.2, y los correspondientes ajustes por mínimos cuadrados. La matriz  $\mathbf{\Omega}^{-1/2}$  se obtiene mediante la factorización de Cholesky.

### Elección del parámetro de compensación

Como ya se vio en los capítulos anteriores, la elección del parámetro de compensación  $\mu$  depende del nivel de ruido presente en el dato. Si  $\mu$  es demasiado grande, las soluciones estimadas serán demasiado *blocky* y quizás no honren al dato observado. Contrariamente, si  $\mu$  es demasiado pequeño el dato puede ser sobre-ajustado y las soluciones pueden ser inestables y no presentar *blockyness*. En los capítulos previos se utilizó el método de discrepancia y el análisis de la curva de Pareto para poder estimar un valor de  $\mu$  óptimo, según la información disponible. En este capítulo utilizaremos nuevamente el principio de discrepancia, basado en el test de  $\chi^2$  (Sacchi, 1997). Como se vio en el capítulo 4, se define la variable  $\chi^2$  como

$$\chi^2 = \sum_{i=1}^{N_\theta} \sum_{h=1}^{L_r} \frac{(d_{ih} - [\mathbf{Am}]_{ih})^2}{\sigma_n^2}, \quad (7.21)$$

donde  $\sigma_n^2$  es el desvío estándar del dato observado y  $(d_{ih} - [\mathbf{Am}]_{ih})^2$  es el *misfit* entre el dato observado y el dato modelado para un dado valor del parámetro de compensación  $\mu$ . Dicha variable sigue la distribución chi-cuadrado con  $N_\theta L_r$  grados de libertad. El valor esperado de  $\chi^2$  es

$$E[\chi^2] = N_\theta L_r \pm \sqrt{2N_\theta L_r}, \quad (7.22)$$

En este ejemplo en particular el dato sísmico consiste en 31 trazas, cada una formada por 150 muestras, por lo tanto el valor óptimo de  $\mu$  será aquel tal que  $\|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 \simeq \sigma_n^2(4650 \pm 100)$ .

Siguiendo el trabajo de Gramfort et al. (2013), cuando se utiliza la norma  $l_{2,1}$  como término de regularización existe

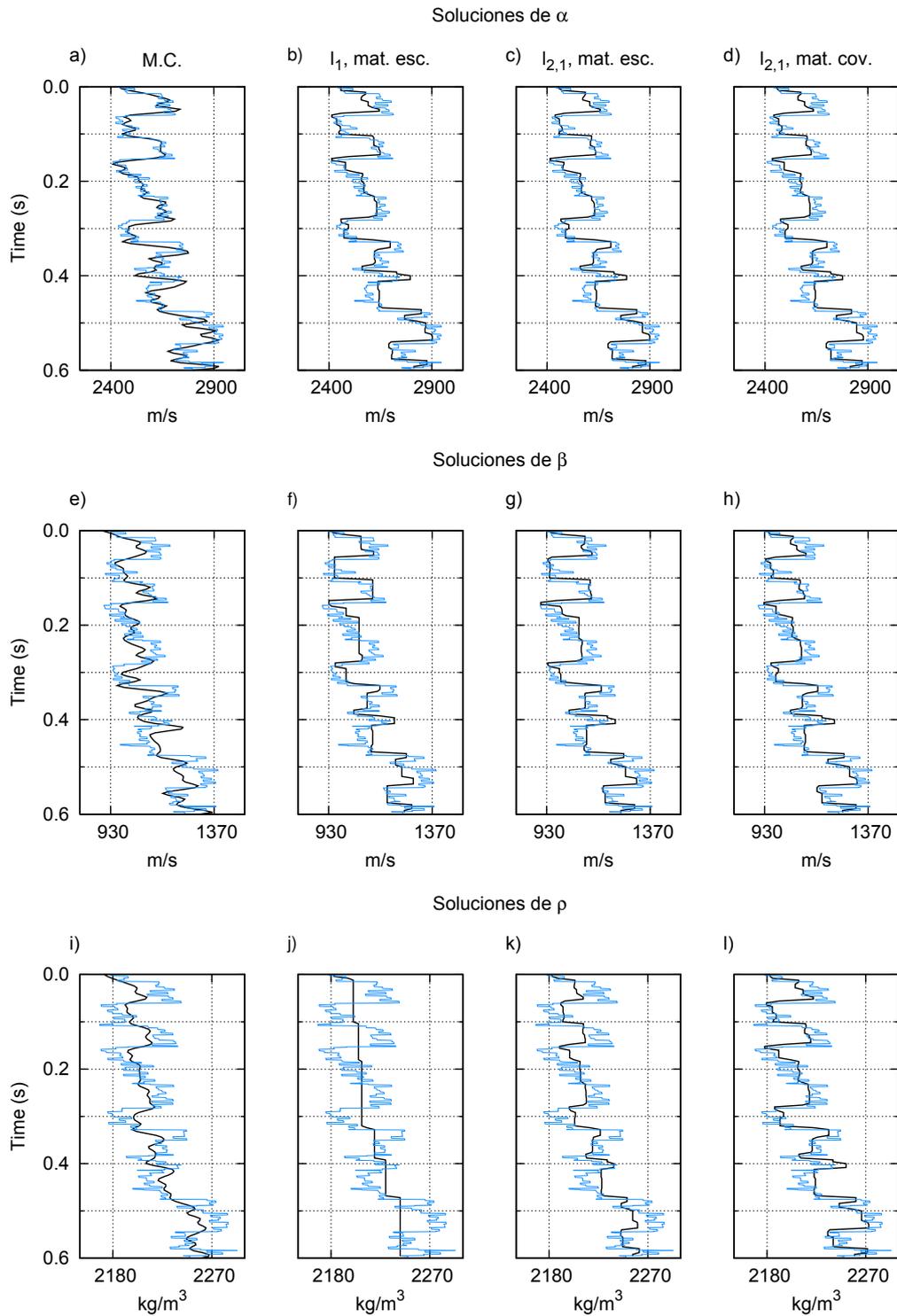
$$\mu_{max} = \max_i \|(\mathbf{C}^T \mathbf{f})[i, \cdot]\|_2, \quad (7.23)$$

tal que si  $\mu \geq \mu_{max}$ , entonces la solución estimada es igual al vector nulo. Por lo tanto, en lugar de buscar el parámetro de compensación óptimo  $\mu$ , estimaremos  $\tilde{\mu} = \mu/\mu_{max}$ , donde  $\tilde{\mu} \in (0, 1)$ . Esta reparametrización no sólo hace al parámetro de compensación menos sensible al dato, sino que también simplifica su estimación. Como el valor de  $\chi^2$  se incrementa con el valor de  $\tilde{\mu}$ , y el rango de posibles valores de éste está acotado, podemos utilizar un algoritmo de búsqueda de raíces clásico, como puede ser el método de la secante, para estimar el valor óptimo de  $\tilde{\mu}$  dado un valor aproximado de  $E[\chi^2]$ .

El método utilizado para estimar el parámetro de compensación es, en varios aspectos, similar al utilizado en el capítulo anterior al realizar la inversión sin el *debiasing step*, es decir, haciendo un análisis de las curvas de Pareto. Ambos métodos están basados en el principio de discrepancia y requieren de un estimado del nivel de ruido en el dato observado. Es importante notar que el valor estimado del parámetro de compensación puede no ser único, y existir un rango de valores de  $\mu$  que honren al dato observado con igual exactitud. Muchas veces la elección de un determinado valor de este parámetro estará basada en el estudio de las soluciones estimadas y en el juicio personal, especialmente cuando el nivel de ruido en el dato no es conocido (Farquharson y Oldenburg, 2004).

## Resultados

La Figura 7.4 muestra los resultados de realizar la inversión sobre el dato de la Figura 7.2 utilizando la inversión por mínimos cuadrados (es decir  $\mu = 0$ ), utilizando la regularización mediante la norma  $l_1$  pesada ( $p = q = 1$ ) y la regularización mediante la norma  $l_{2,1}$  pesada ( $p = 2, q = 1$ ). Es importante notar que en todos los casos el segundo término de la ecuación (7.17) fue incorporado en la inversión. Este término no sólo permite recuperar la tendencia de baja frecuencia sino que también al realizar la inversión por mínimos cuadrados actúa como una regularización de Tikhonov que balancea las amplitudes de las soluciones estimadas. Como se vio en la sección 4.2, Figura 4.6, sin dicha matriz de escala las soluciones por mínimos cuadrados no sólo carecen de resolución sino que también presentan amplitudes inexactas. Por simplicidad, la inversión utilizando la norma  $l_1$  se llevó a cabo considerando que las covarianzas entre las tres reflectividades son cero, es decir, utilizando una matriz de escala diagonal para restringir las amplitudes relativas de los parámetros estimados.



**Figura 7.4.** Valores estimados de  $\alpha$ ,  $\beta$  y  $\rho$  utilizando a), e) y i) mínimos cuadrados, b), f) y j) norma  $l_1$  con la matriz de escala, c), g) y k) norma  $l_{2,1}$  con la matriz de escala y d), h) y l) con la matriz de covarianzas.

Al utilizar la norma  $l_{2,1}$ , con propósitos comparativos, se utilizó la matriz de covarianza completa y la matriz de escala diagonal como dos casos separados. Las líneas color azul en la Figura 7.4 muestran los valores verdaderos de  $\alpha$ ,  $\beta$  y  $\rho$ , mientras que las

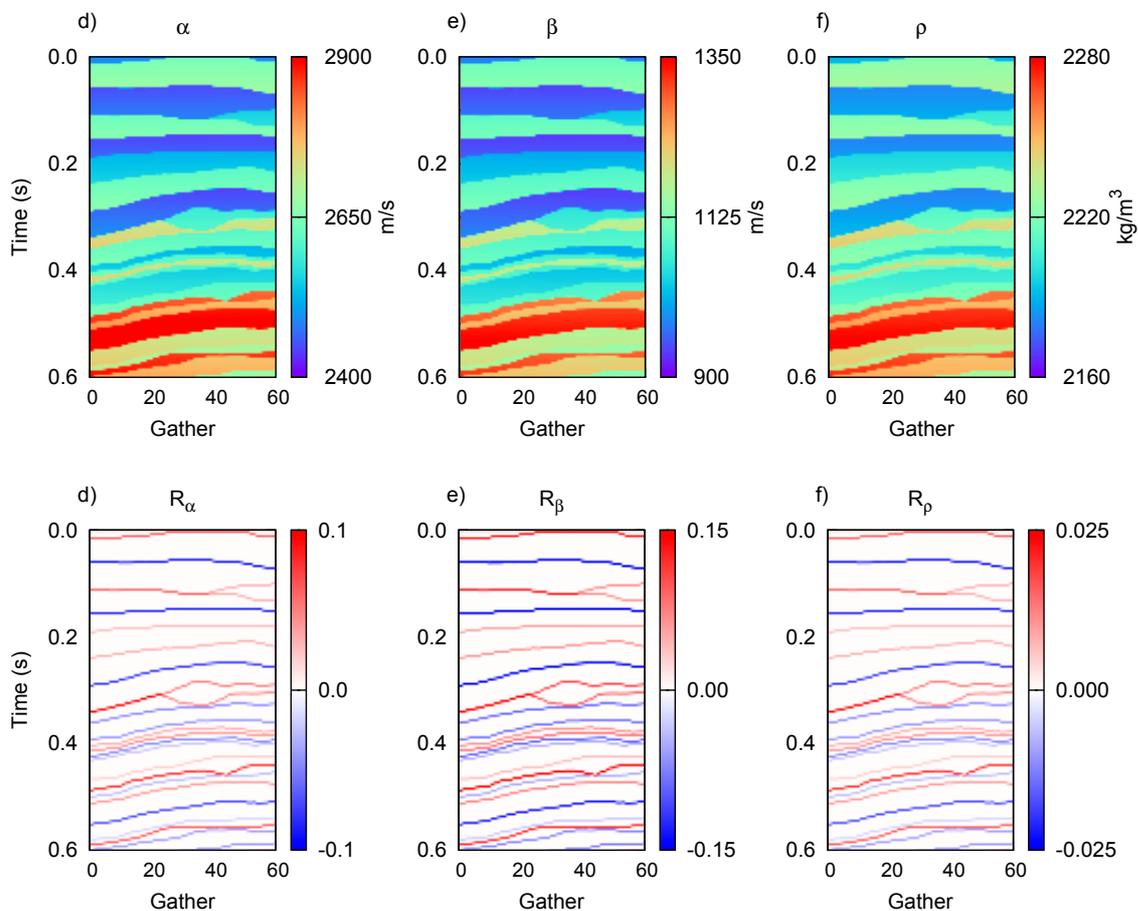
líneas color negro muestran los valores estimados. Se puede observar que en general el valor de  $\alpha$  fue estimado correctamente en todos los casos, aunque, como es esperable, la inversión por mínimos cuadrados presenta mucho menos definición que las otras, haciendo difícil la resolución de estructuras delgadas, como aquella localizada a los 0.5 s. También se pueden observar inestabilidades de alta frecuencia en las amplitudes estimadas. Estos inconvenientes no están presentes en las soluciones regularizadas. En dichos casos se puede observar que las amplitudes de  $\alpha$  están levemente subestimadas alrededor de los 0.4 s, probablemente porque la matriz de escala utilizada fue considerada invariante en toda la ventana temporal. Sin embargo, las soluciones regularizadas muestran una estructura que se asemeja al modelo actual con más precisión que la solución por mínimos cuadrados.

Las amplitudes estimadas de  $\beta$ , por otro lado, comienzan a mostrar algunas imprecisiones, excepto en el caso en el cual se regularizó el problema utilizando la norma  $l_{2,1}$  y la matriz de covarianza completa. La falta de *blockyness* y las inestabilidades de la solución por mínimos cuadrados hace imposible resolver muchas de las capas que forman el subsuelo. La solución obtenida mediante la norma  $l_1$  muestra una leve subestimación de las amplitudes y algunas discontinuidades, como aquella ubicada a los 0.2 s, no fueron resueltas con claridad. La solución obtenida utilizando la norma  $l_{2,1}$  y la matriz de escala diagonal muestra una mejor estimación de las amplitudes, aunque se siguen observando leves subestimaciones de las mismas, especialmente a partir de los 0.4 s. Cuando se realiza la inversión utilizando la norma  $l_{2,1}$  y la matriz de covarianza completa, el modelo estimado honra al verdadero con más precisión.

En el caso de la densidad  $\rho$ , las diferencias entre los modelos estimados son más significativas. La solución obtenida con mínimos cuadrados muestra amplitudes subestimadas, una considerable falta de *blockyness* e inestabilidades, y prácticamente ninguna de las discontinuidades puede ser resuelta adecuadamente. En el caso de utilizar la norma  $l_1$  como regularización, sólo la tendencia de baja frecuencia fue recuperada, y las amplitudes estimadas están severamente subestimadas haciendo que la mayoría de las discontinuidades sean imposibles de resolver. Las diferencias entre las soluciones estimadas utilizando la norma  $l_1$  y la norma  $l_{2,1}$  son ahora considerables. La solución obtenida utilizando la norma  $l_{2,1}$  considerando que no se tiene información de las covarianzas es cualitativamente similar a aquellas obtenidas para  $\alpha$  y  $\beta$ . Aún cuando se observan subestimaciones en las amplitudes, la característica *blocky* del modelo fue recuperada en gran parte. Al igual que en el caso de  $\alpha$  y  $\beta$ , la solución obtenida utilizando la norma  $l_{2,1}$  y la matriz de covarianza completa supera en calidad las soluciones obtenidas en los otros casos. El modelo estimado honra al verdadero con precisión.

### 7.3.2. Datos sintéticos 2D

La Figura 7.5 muestra un modelo 2D de velocidades  $\alpha$  y  $\beta$  y densidad  $\rho$ , que es una extensión del modelo 1D utilizado en sección anterior, también obtenido a partir de una sección del modelo Marmousi2 (Martin et al., 2006). Asimismo se pueden ver las correspondientes reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  del modelo 2D. A partir de este modelo se crearon 60 *angle-gathers*, cada uno abarcando un rango de ángulos  $\theta_i \in (0^\circ, 30^\circ)$ , con un intervalo de muestreo de 4 ms. El dato fue creado utilizando una ondicula de Ricker con  $f_0 = 30$  Hz. La respuesta AVA fue modelada utilizando las ecuaciones de Zoeppritz. Al dato se le agregó ruido gaussiano no correlacionado con  $\text{SNR} = 5$ . El *angle-gather* analizado en la sección anterior corresponde al primer *angle-gather* del modelo 2D analizado aquí.



**Figura 7.5.** a), b), c) Modelo 2D de velocidades y densidad, d), e) y f) reflectividades correspondientes.

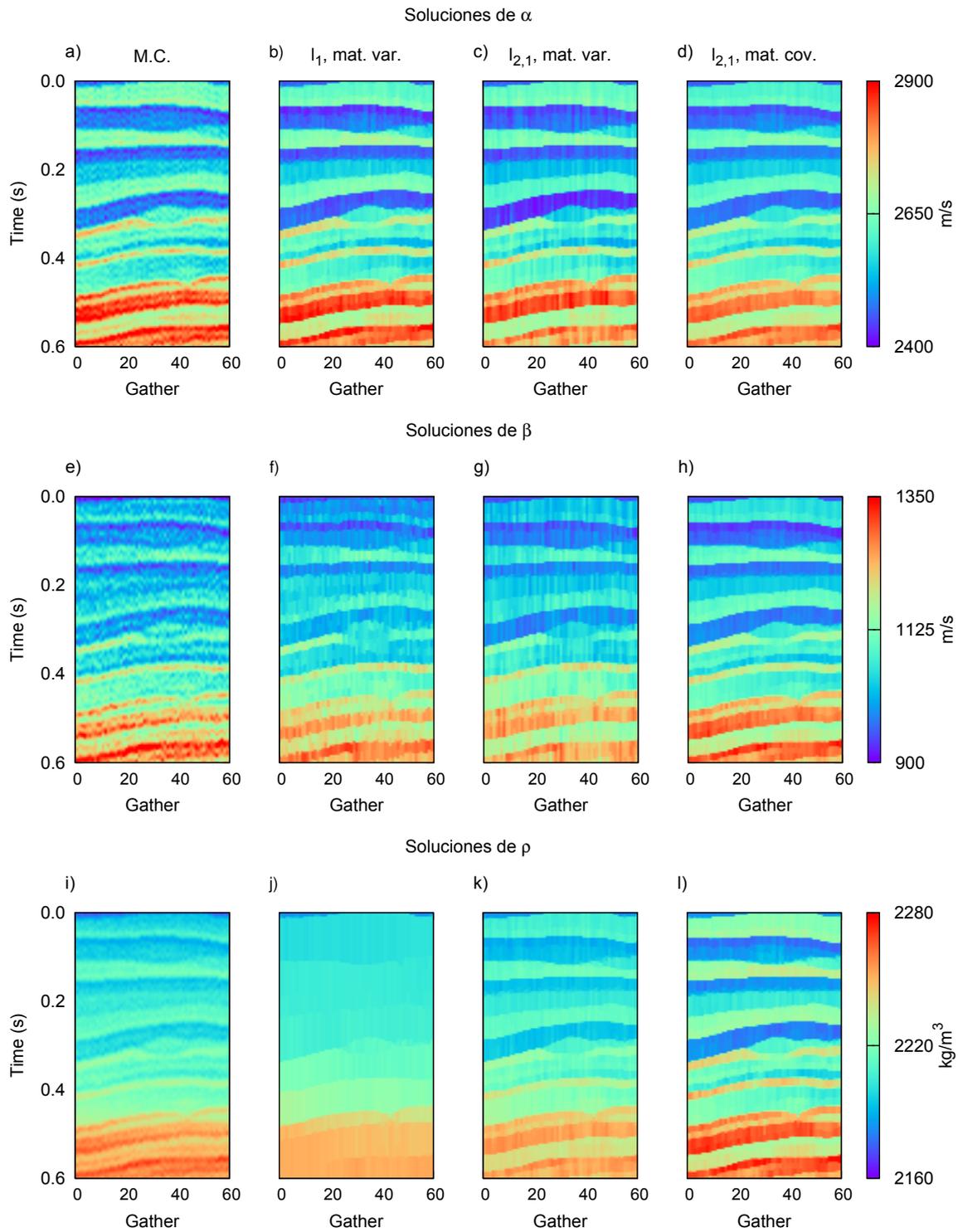
Como primer paso procederemos a estimar soluciones *blocky* de los parámetros físicos, es decir que para cada *angle-gather* encontraremos la solución que hace mínima la ecuación (7.17). Al igual que en la sección anterior, con propósitos comparativos, se procedió a realizar la inversión por medio del método de mínimos cuadrados ( $\mu = 0$ ), utilizando la regularización  $l_1$  pesada ( $p = q = 1$ ) y utilizando la regularización  $l_{2,1}$  pesada ( $p = 2, q = 1$ ). Para todos los *gathers* se utilizaron las mismas matrices de

escala  $\Sigma$  y  $\Omega$ , tendencia de baja frecuencia y parámetro de compensación  $\mu$  estimados en la sección anterior a partir del primer *gather*. La Figura 7.6 muestra los valores de  $\alpha$ ,  $\beta$  y  $\rho$  estimados para todas las inversiones realizadas. Analizando la Figura 7.6 en comparación a la Figura 7.5 se observa, como era esperable, que las soluciones regularizadas con la norma  $l_{2,1}$  son las que mejor honran al modelo. Las soluciones obtenidas por el método de mínimos cuadrados carecen de resolución y en particular la solución estimada de la densidad  $\rho$  está muy subestimada. Regularizar el problema mediante la norma  $l_1$  permitió obtener estimaciones de  $\alpha$  y  $\beta$  con buena resolución y correctas amplitudes pero, al igual que la solución por mínimos cuadrados, las amplitudes estimadas de  $\rho$  resultan muy subestimadas. Las soluciones estimadas mediante la norma  $l_{2,1}$  muestran una considerable mejoría respecto de las otras. Si bien la solución estimada mediante la norma  $l_{2,1}$  utilizando la matriz de escala muestra subestimación de los valores estimados de  $\rho$ , son más cercanas a los valores deseados. Las principales estructuras geológicas, así como su variación lateral y vertical, fueron recuperadas correctamente. Sin embargo se hacen más notables algunos problemas de continuidad lateral, especialmente en las soluciones de  $\beta$ . Estos problemas de continuidad lateral surgen de realizar la inversión *gather a gather* utilizando las mismas matrices de escala y un único valor de  $\mu$  en todos los *gather*s. En el caso de utilizar la norma  $l_{2,1}$  junto con la matriz de covarianza las soluciones son mucho más precisas que aquellas soluciones estimadas con las otras estrategias. Las amplitudes de los tres parámetros fueron estimados correctamente y las soluciones muestran una muy buena continuidad lateral. Estructuras como la lente ubicada a los  $t \simeq 0.3$  s, entre los *gather*s #20 y #40 están claramente delineadas, haciendo sencilla su interpretación.

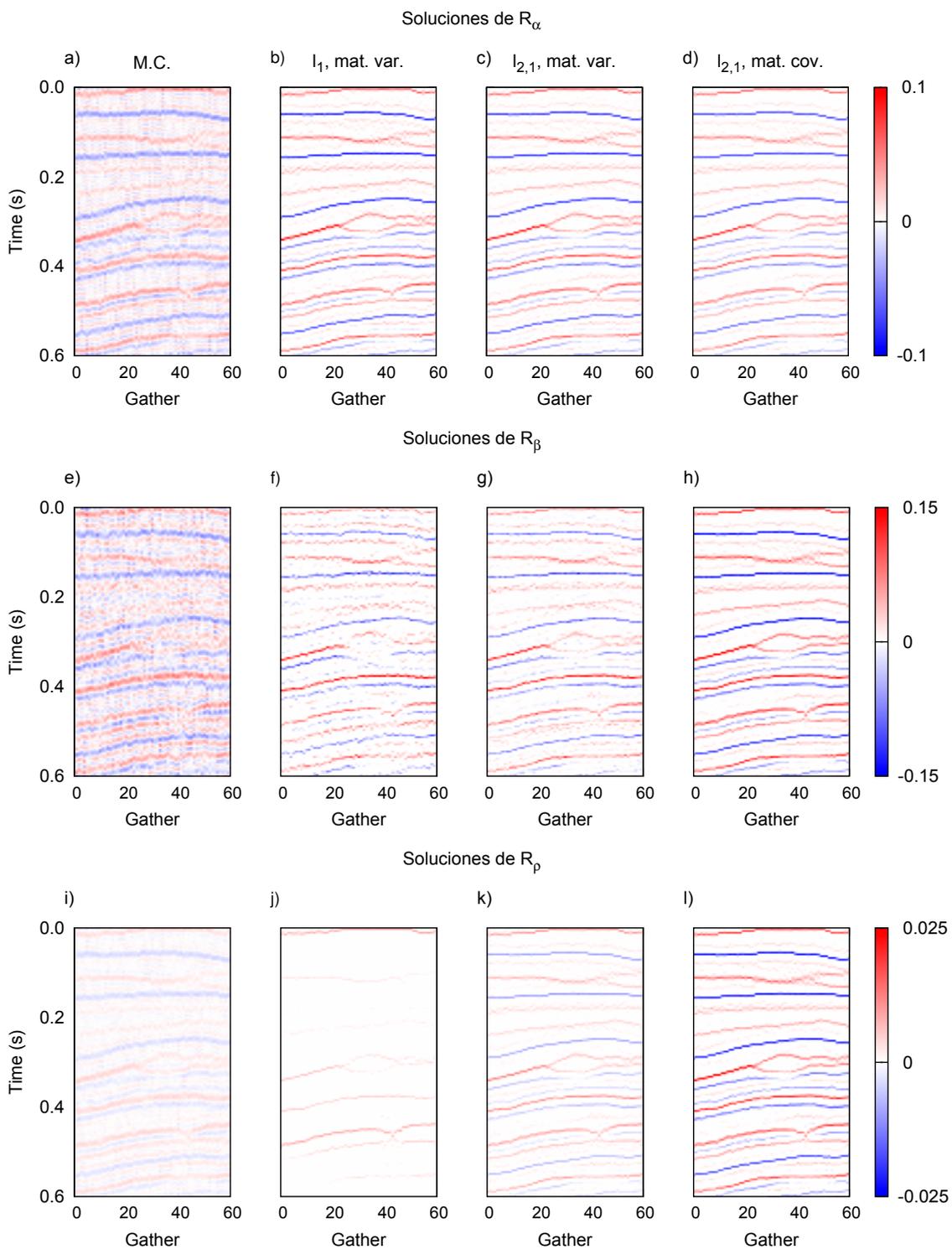
La Figura 7.7 muestra las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  estimadas. Como puede apreciarse, las principales características observadas en las soluciones de  $\alpha$ ,  $\beta$  y  $\rho$  pueden observarse también aquí. Es de destacar que en el caso de la inversión por mínimos cuadrados, es decir con  $\mu = 0$ , las reflectividades estimadas, aunque levemente ruidosas, muestran un comportamiento mucho más estable que el visto en el capítulo 4 al utilizar mínimos cuadrados amortiguados. Esto se debe a que el segundo término de la función de costo dada por la ecuación (7.17) actúa como una regularización de Tikhonov, estabilizando la solución. Además, la presencia de la matriz  $\Sigma_l^{-1/2}$  dentro de dicho término tiene el mismo efecto que utilizar un factor de amortiguación diferente para cada parámetro físico. Esto permite que las amplitudes de cada uno de éstos estén mejor restringidas que cuando se utiliza un único factor de amortiguación. Sin embargo las soluciones carecen de resolución y la correspondiente a  $R_\rho$  muestra subestimación a pesar de la matriz  $\Sigma_l^{-1/2}$ .

Cuando se utiliza la norma  $l_1$  como regularización se puede observar que las soluciones estimadas de  $R_\alpha$  y  $R_\beta$  son correctas, aunque en esta última se puede ver que no todos los reflectores han sido resueltos correctamente y hay una leve falta de continuidad lateral. La solución estimada de  $R_\rho$  está lejos de ser aceptable, sólo aparecen unos pocos reflectores débiles. Estos reflectores son estimados con el único fin de recuperar la tendencia de baja frecuencia presente en el segundo término de la ecuación (7.17). De no estar dicho término la inversión utilizando la norma  $l_1$  sería incapaz de recuperar información de la reflectividad  $R_\rho$ .

Las reflectividades estimadas por medio de la norma  $l_{2,1}$  son en todos los casos correctas, mostrándose más cercanas a los valores verdaderos que aquellas estimadas



**Figura 7.6.** Valores estimados de  $\alpha$ ,  $\beta$  y  $\rho$  utilizando a),e) y i) mínimos cuadrados, b),f) y j) norma  $l_1$  con la matriz de escala, c), g) y k) norma  $l_{2,1}$  con la matriz de escala y d),h) y l) con la matriz de covarianzas.



**Figura 7.7.** Valores estimados de  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  utilizando a),e) y i) mínimos cuadrados, b),f) y j) norma  $l_1$  con la matriz de escala, c), g) y k) norma  $l_{2,1}$  con la matriz de escala y d),h) y l) con la matriz de covarianzas.

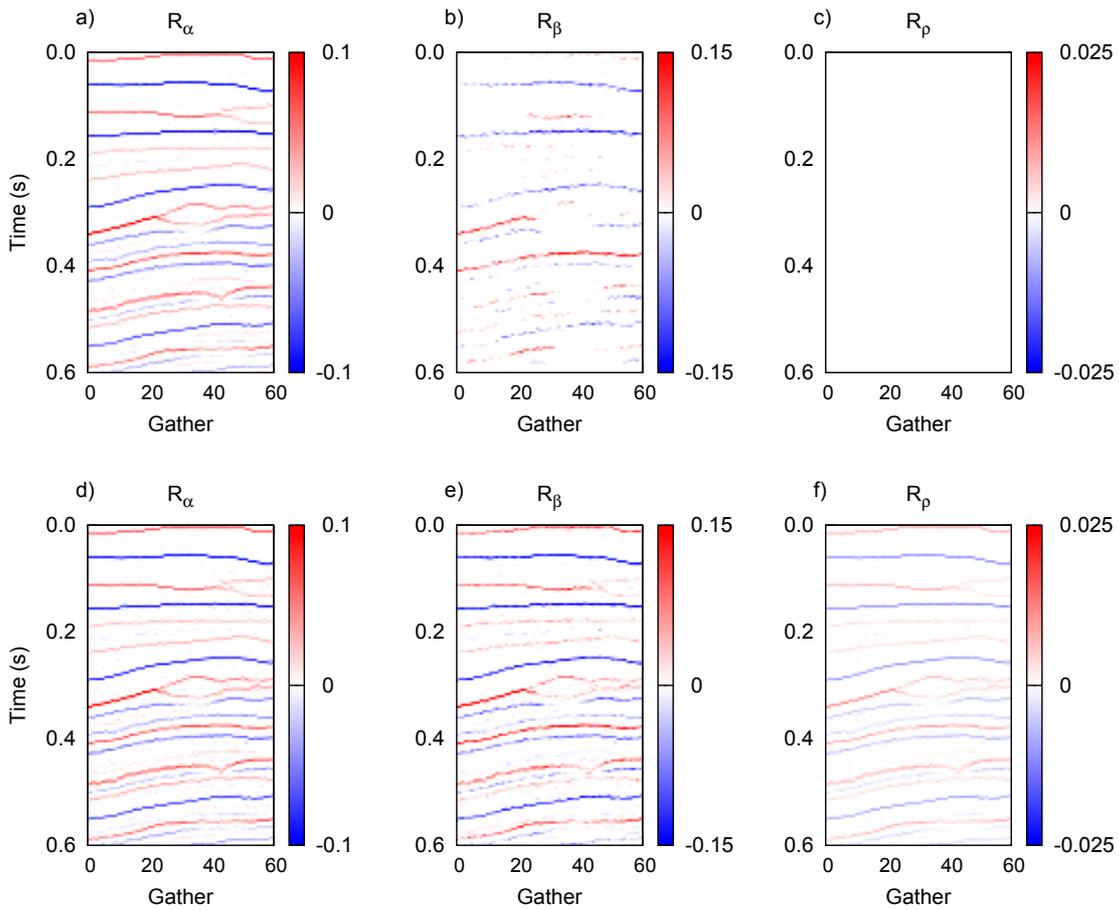
por los otros métodos. Cuando no se utiliza la matriz de covarianza completa se puede observar una leve subestimación de las amplitudes estimadas, especialmente para  $R_\rho$ , sin embargo la solución es más clara y fácil de interpretar que la estimada utilizando la norma  $l_1$ . Cuando se utiliza la matriz de covarianza completa las soluciones estimadas honran correctamente a las reflectividades verdaderas. Sin embargo, al trabajar con datos de campo no siempre se tiene información de pozo disponible para poder estimar dicha matriz.

### Soluciones $l_1$ vs $l_{2,1}$ sin información de pozo

La Figura 7.8 muestra los resultados de realizar la inversión utilizando como regularización la norma  $l_1$  (Figuras 7.8a-c) y la norma mixta  $l_{2,1}$  (Figuras 7.8e-f) suponiendo que no se tiene información de pozo. Es decir se descartó el segundo término de la ecuación (7.17), que contiene la información de la tendencia de baja frecuencia e información de las escalas de los parámetros físicos. Asimismo se consideró a la matrices  $\Omega_l$  como matrices diagonales, es decir que no se incorporó a la inversión información de la correlación entre las reflectividades, pero sí de las varianzas. Ésta es la menor cantidad de información necesaria para restringir las amplitudes de las soluciones estimadas a partir de la aproximación de Aki y Richards de tres términos. Como se puede observar, bajo estas condiciones, regularizar el problema mediante la norma  $l_1$  nos lleva a soluciones poco satisfactorias. Mientras que la reflectividad  $R_\alpha$  fue correctamente estimada, la reflectividad  $R_\beta$  muestra subestimación de las amplitudes y falta de continuidad lateral. Muchos de los reflectores no fueron resueltos y la mayoría de las estructuras geológicas no están definidas. Aún con la presencia de la matriz de escala, cuyo fin es balancear las amplitudes de las soluciones estimadas, la inversión regularizada con la norma  $l_1$  no permite estimar soluciones *sparse* de la reflectividad  $R_\rho$ . Este comportamiento es similar al ya visto en el capítulo 5, donde las amplitudes de las soluciones estimadas del *gradient*  $G$  estaban severamente subestimadas.

Bajo las mismas condiciones, la regularización con la norma mixta  $l_{2,1}$  ofrece mejores resultados. Como puede apreciarse, las soluciones *sparse* de  $R_\beta$  y  $R_\rho$  se asemejan más a los valores verdaderos. La solución de  $R_\beta$  muestra mejor continuidad lateral que aquella estimada con la norma  $l_1$ . Todos los reflectores y sus amplitudes han sido recuperados satisfactoriamente, y estructuras como la lente ubicada a los 0.3 s comienzan a ser visible. La solución de  $R_\rho$  comienza a ser visible también, y aunque se observa cierta subestimación en las amplitudes se pueden distinguir claramente los principales reflectores y estos muestran buena continuidad lateral.

Como puede apreciarse, a pesar de que la inversión se llevó a cabo sobre dato con niveles de ruido elevado e incorporando la mínima cantidad de información *a priori*, los resultados obtenidos con la norma  $l_{2,1}$  superaron en calidad a aquellos estimados utilizando la norma  $l_1$ , brindando interesantes soluciones *sparse* que permiten una caracterización clara del subsuelo. Esto se debe a que la norma  $l_{2,1}$  restringe más el espacio de las soluciones, forzando una mejor distribución de la energía de los reflectores, de manera tal que dicha energía está mejor balanceada entre  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ . Dicho comportamiento se debe a que la norma mixta  $l_{2,1}$ , que es una combinación de la norma  $l_2$  y la norma  $l_1$ , por lo que permite obtener soluciones *sparse* sin la tendencia propia de la norma  $l_1$  de subestimar agresivamente las amplitudes.

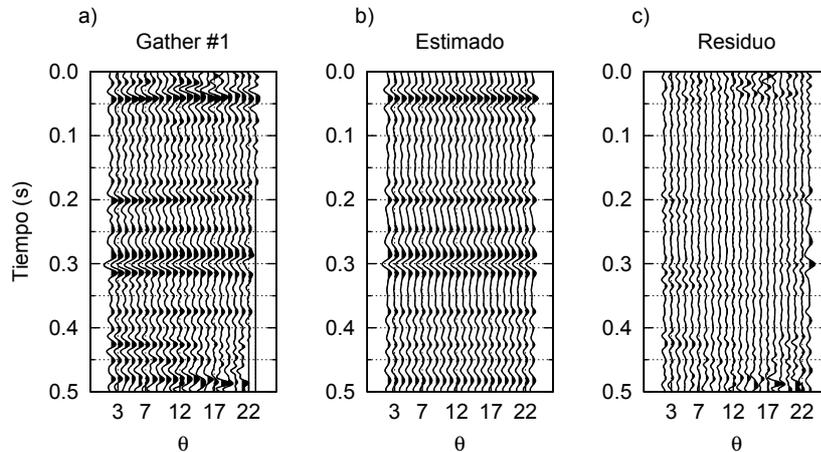


**Figura 7.8.** Valores estimados de  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  utilizando a), b) y c) la norma  $l_1$  con la matriz de escala, d), e) y f) norma  $l_{2,1}$  con la matriz de escala.

### 7.3.3. Datos de campo

Debido a falta de información *a priori* adecuada la inversión sólo se llevará a cabo con el fin de estimar soluciones *sparse* de las reflectividades, es decir que se obviaré el segundo término de la ecuación (7.17). Los datos de campo complican la inversión no sólo porque el nivel de ruido no está disponible, sino también porque la ondícula no es conocida. En el caso del método propuesto en este capítulo, la inversión se complica aún más debido a la inestabilidad inherente de las aproximaciones de tres términos y a las incertidumbres a la hora de estimar las matrices de escala  $\Omega_l$ . Al igual que en los ejemplos sobre datos sintéticos ya analizados, se asumirá que la matriz es invariante en la ventana de tiempo en la cual se realizará la inversión, por lo tanto  $\Omega_l = \Omega$ . En este ejemplo, al no contar con datos de pozo de la zona donde fue adquirido el dato sísmico, dicha matriz de pesos fue estimada asumiendo que las reflectividades se derivan de variaciones de velocidades  $\alpha$ ,  $\beta$  y de la densidad  $\rho$  que se relacionan entre si por medio de las ecuaciones de Gardner et al. (1974) y Castagna et al. (1985), y que toman valores similares a los del modelo sintético 2D. Se asumirá además que el dato ha sido apropiadamente acondicionado para preservar las amplitudes, y que la ondícula fue estimada estadísticamente a partir de la raíz cuadrada de la transformada

de Fourier de la autocorrelación de las trazas, asumiendo fase cero (Robinson y Treitel, 2002). El dato de campo utilizado, que es el mismo utilizado en el capítulo 6, consiste en 400 *angle-gathers* corregidos por NMO, con un intervalo de muestreo de 2 ms, una ventana temporal de 0.5 s y un rango de ángulos  $\theta_i \in (2^\circ, 24^\circ)$ .



**Figura 7.9.** a) *Gather* observado #1 del dato de campo, b) estimado utilizando la norma  $l_{2,1}$ , c) residuo entre el *gather* observado y el estimado.

La Figura 7.9 muestra el primer *gather* del dato, la correspondiente reconstrucción a partir del  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  estimados y la diferencia entre ambos. Como puede observarse las soluciones estimadas honran al dato observado correctamente, sólo mostrando diferencias en lugares donde el dato observado es muy ruidoso, como al principio y al final del *gather*. En la Figura 7.10 se pueden ver las imágenes correspondientes a  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  estimadas a partir de todos los *gathers*. El método propuesto logró obtener soluciones de tipo *sparse* para las tres reflectividades. Dichas soluciones presentan, en general, buena continuidad lateral y honran al dato observado, como puede verse de la comparación entre el *stack* del dato observado presentado en la Figura 7.11a, y el *stack* estimado a partir de las soluciones obtenidas, presentado en la Figura 7.11b. Debido al limitado rango de ángulos utilizado y a la incertidumbre en la estimación de la matriz  $\Omega$ , la solución estimada para  $R_\beta$  presenta inestabilidades y menor continuidad lateral que las soluciones de  $R_\alpha$  y  $R_\rho$ .

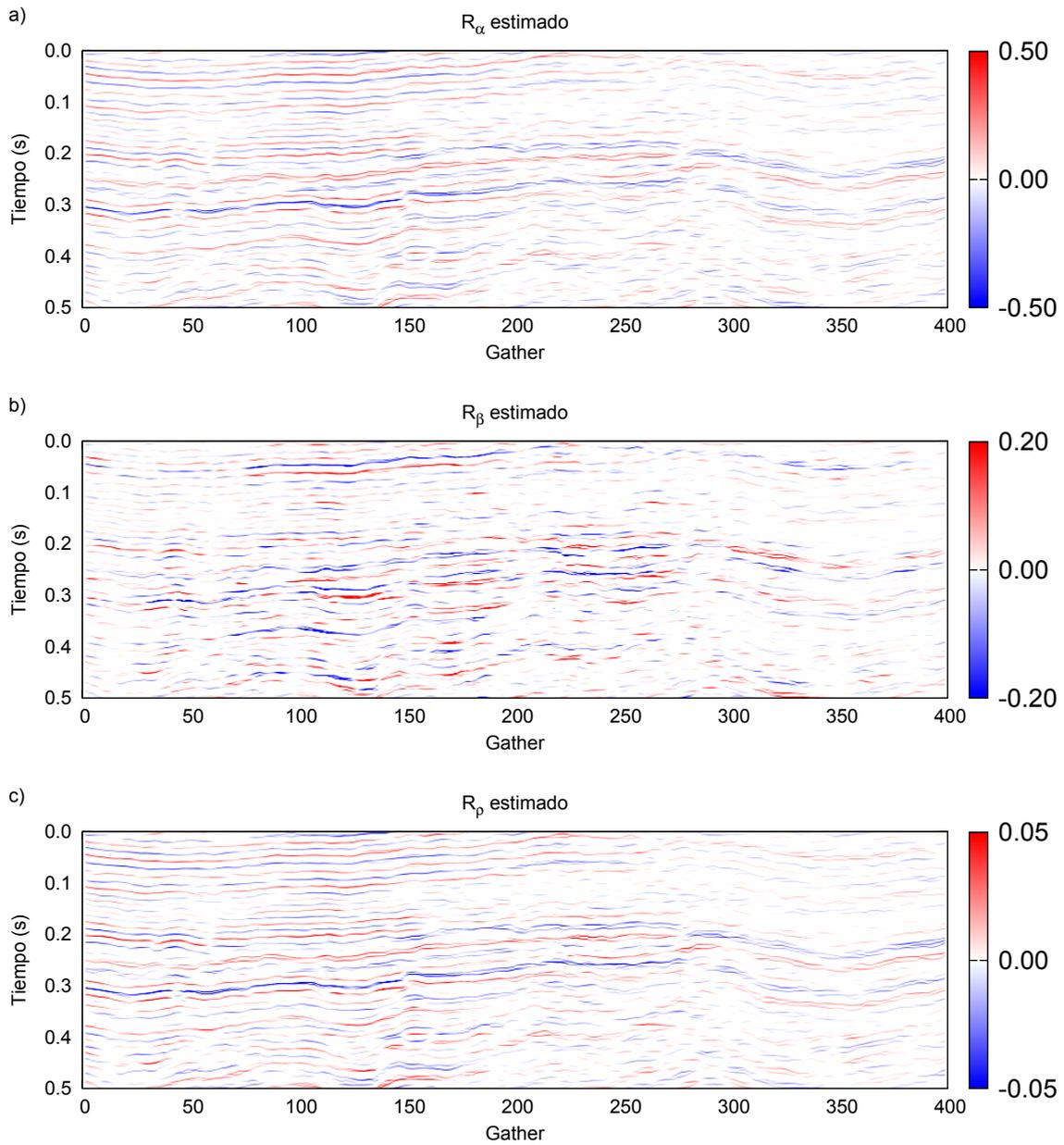


Figura 7.10. a), b) y c)  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$  estimados a partir del dato de campo.

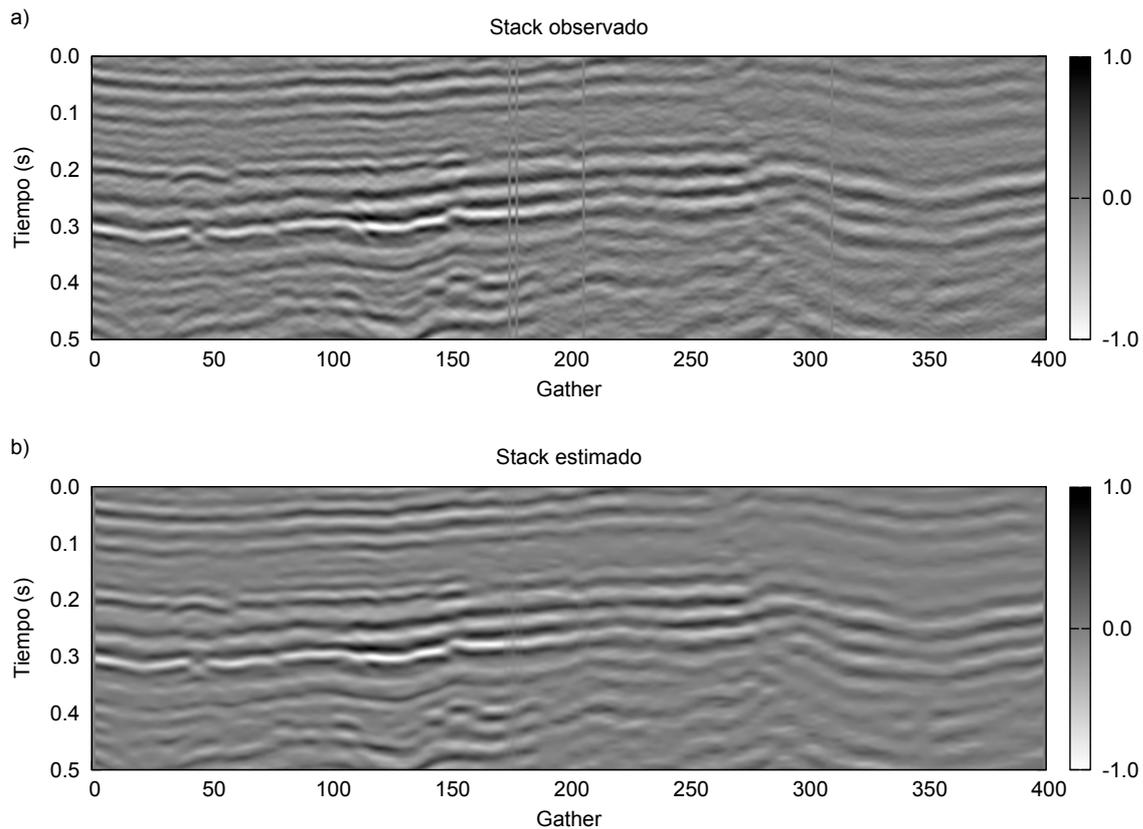


Figura 7.11. a) y b) *stack* observado y estimado a partir del dato de campo.

## 7.4. Conclusiones

A continuación se enumeran las principales conclusiones obtenidas utilizando la estrategia propuesta sobre datos sintéticos y datos de campo:

- La estrategia propuesta es versátil, permitiendo estimar soluciones *sparse* de las reflectividades de la aproximación de Aki y Richards de tres términos, o soluciones *blocky* de los parámetros físicos involucrados (velocidades y densidades), dependiendo de la información *a priori* disponible.
- Pruebas sobre datos sintéticos 1D muestran que la estrategia propuesta es capaz de obtener soluciones *blocky* de los parámetros físicos que honran al dato observado, muy favorables en relación a las estimadas utilizando la norma  $l_1$ .
- En pruebas sobre datos sintéticos 2D se observa que la estrategia es capaz de obtener soluciones *blocky* de los parámetros físicos y soluciones *sparse* de las reflectividades que honran al dato y presentan buena continuidad lateral.
- Se observó que bajo condiciones en las cuales la información *a priori* utilizada no está completa o es incierta, las soluciones estimadas utilizando la norma mixta  $l_{2,1}$  superaron en calidad a aquellas obtenidas utilizando la norma  $l_1$ .

- Las soluciones estimadas a partir de datos de campo muestran una marcada *sparse*ness a la vez que presentan buena continuidad lateral y honran al dato observado. Como consecuencia de la incertidumbre que se tiene de la matriz de escala utilizada y debido al limitado rango de ángulos utilizados durante la inversión se observaron algunas inestabilidades en la solución  $R_\beta$ .
- Al igual que en la estrategia desarrollada en el capítulo anterior, como FISTA sólo realiza multiplicaciones entre matrices y vectores, el método propuesto es económico en términos de costo computacional y de fácil aplicación.

# Capítulo 8

## Conclusiones

El tema central de esta tesis fue la resolución de un problema fundamental para la detección de hidrocarburos y la caracterización litológica: la inversión de datos sísmicos *prestack*. Debido a que este es un problema mal planteado y mal condicionado se requiere la utilización de regularizaciones y restricciones con el fin de estabilizar el proceso de inversión, mitigar el problema de la no-unicidad y, a la vez, darle a la solución estimada alguna característica deseada de acuerdo a la información *a priori* disponible. Más precisamente, en esta tesis se desarrollaron estrategias para estimar soluciones de tipo *sparse* de las reflectividades asociadas a las aproximaciones de Shuey de dos términos, de Aki y Richards de tres términos y, cuando la información disponible lo permitió, soluciones tipo *blocky* de las velocidades y densidades. Con este fin se utilizaron como término de regularización diferentes normas matemáticas que promueven soluciones *sparse*. Esto condujo a funciones de costo que debieron ser minimizadas utilizando algoritmos de optimización global y local dependiendo de las características que presentaron dichas funciones.

### 8.1. Conclusiones

A continuación se enunciarán las principales conclusiones obtenidas de las pruebas realizadas sobre datos sintéticos y datos de campo para cada una de las estrategias propuestas, y que constituyen el aporte central de esta tesis.

#### Capítulo 5: Inversión *sparse-spike* mediante *Simulated Annealing* y mínimos cuadrados

En el capítulo 5 se presentó la primera estrategia para obtener soluciones *sparse* a partir de datos sísmicos *prestack*. Dicha estrategia fue desarrollada para obtener soluciones *sparse* de los atributos *intercept* y *gradient* de la aproximación de Shuey de dos términos. La estrategia planteada se basa en estimar soluciones sub-óptimas del problema de inversión regularizado mediante la norma  $l_0$ . Esta regularización conduce a un problema de optimización NP-hard. Este tipo de problemas son los más costosos de resolver desde un punto de vista computacional, y muchas veces son reemplazados por problemas regularizados utilizando la norma  $l_1$ . La estrategia propuesta consiste en determinar las posiciones en tiempo de un número fijo y pequeño de coeficientes de re-

flexión por medio de VFSA. En cada iteración del VFSA la amplitud de los coeficientes de reflexión se determina resolviendo un problema de mínimos cuadrados de pequeña dimensión. La utilización de un número limitado de coeficientes de reflexión estabilizó el problema de inversión, permitiendo obtener soluciones *sparse* con sentido físico que honran al dato observado incluso en presencia de niveles de ruido significativos.

Una desventaja de la estrategia propuesta es que, previo a utilizar VFSA, se debe establecer el mínimo número de coeficientes de reflexión  $L_s$  necesarios para ajustar el dato observado con una dada tolerancia. Obtener esta información puede ser costoso computacionalmente, ya que se debe resolver el problema varias veces utilizando diferentes valores del  $L_s$  y construir una curva-L que permita discriminar el  $L_s$  óptimo de acuerdo a algún criterio adecuado. En este caso se utilizó el criterio de la discrepancia, que es de utilidad cuando se tiene un estimado del nivel de ruido en el dato. A partir de este análisis se concluyó que si el número  $L_s$  de coeficientes de reflexión utilizados no es mucho mayor que el verdadero, la estabilidad de las soluciones no se ven afectadas significativamente. Sin embargo utilizar un número menor lleva a soluciones que no honran el dato observado.

Una ventaja del método de inversión propuesto es que, al ser VFSA un algoritmo estocástico, fue posible estimar la incertidumbre de las soluciones aprovechando el gran número de soluciones que se pusieron a prueba durante el proceso de inversión. Sobre dato sintético 1D la estrategia propuesta demostró ser robusta en presencia de ruido aleatorio. Se obtuvieron muy buenos resultados en todas las magnitudes invertidas, estimándose con precisión la ubicación en tiempo de los coeficientes de reflexión y la respuesta AVA de los mismos. Los pequeños valores de las desviaciones estándar de las magnitudes invertidas resaltan la consistencia de las soluciones. Las pruebas sobre datos sintéticos 2D demostraron la capacidad del método para resolver reflectores cercanos y para obtener soluciones con buena continuidad lateral a pesar de que no se impuso ninguna clase de condición en este sentido durante la inversión. Se observó que, cuando la ondícula utilizada es exacta, la estrategia propuesta fue capaz de resolver reflectores cuya separación se encuentra cercana al espesor de *tuning*. Sobre datos de campo el método mostró buen comportamiento a pesar de los problemas que implica trabajar con este tipo de datos. Los reflectores más importantes fueron recuperados y se obtuvieron imágenes de alta resolución del *intercept* y del *gradient* con buena continuidad lateral que honran al dato observado con una precisión aceptable.

Como parte del estudio del comportamiento del método propuesto, se realizó un análisis de las soluciones estimadas cuando la ondícula utilizada en la inversión es inexacta. De este análisis se vio que una incorrecta estimación de la ondícula conducen a soluciones *sparse* que, si bien ajustan al dato observado, se alejan de las soluciones esperadas. Se propone entonces, como parte del proceso de inversión realizar un pequeño ajuste de la ondícula utilizada. Este ajuste debe ser realizado sólo si se cuenta con un valor aproximado del número de coeficientes de reflexión  $L_s$ , de otra forma el proceso de inversión podría no converger a una solución aceptable.

En general la estrategia propuesta demostró ser robusta, permitiendo estimar soluciones *sparse* de la aproximación de Shuey de dos términos que honran al dato observado en presencia de altos niveles de ruido, tanto para datos sintético como para datos de campo. Dichas soluciones son muy favorables en relación a aquellas obtenidas con el método de mínimos cuadrados analizado en el capítulo 4, donde la baja resolución de

las soluciones estimadas dificultan su interpretación. Sin embargo, el alto costo computacional como consecuencia de utilizar un algoritmo de optimización global hace que la estrategia propuesta sea poco práctica para ser utilizada sobre grandes volúmenes de datos.

## Capítulo 6: Inversión *sparse* mediante regularización con la norma $l_1$

En el capítulo 6 presentamos la segunda estrategia para obtener soluciones *sparse* a partir de datos sísmicos *prestack*. La estrategia se basa en regularizar el problema de inversión mediante el uso de la norma  $l_1$ , problema conocido como LASSO, y en ajustar, en un segundo paso correctivo, las amplitudes estimadas mediante mínimos cuadrados. La norma  $l_1$  y el *misfit* se relacionan a través del parámetro de compensación  $\mu$ . Para resolver el LASSO utilizamos el algoritmo FISTA. Este algoritmo sólo realiza multiplicaciones de matrices y vectores y no requiere invertir matrices, por lo que es económico en términos de costo computacional. Al igual que la estrategia propuesta en el capítulo 5, el método apunta a obtener simultáneamente la localización en tiempo de los reflectores más importante y los valores de los atributos *intercept* y *gradient* de la aproximación de Shuey de dos términos, pero con un costo computacional significativamente menor.

Como se vio en los ejemplos numéricos, resolver el LASSO no permite obtener una adecuada solución del *gradient*, aunque si se obtienen soluciones del *intercept* aceptables. Esto sucede debido a que el *gradient* tiene una fuerte dependencia con el nivel de ruido presente en la dato y el valor del parámetro  $\mu$  utilizado para llevar a cabo la inversión. Este comportamiento fue estudiado utilizando dato sintético 1D, donde se vio que la calidad de la solución estimada del *gradient* decrece a medida que el nivel de ruido en el dato aumenta y el valor de  $\mu$  disminuye. El uso de las soluciones obtenidas al resolver el LASSO como información *a priori* para realizar un segundo paso correctivo mediante mínimos cuadrados, demostró ser una opción apropiada para obtener atributos *sparse* que honran al dato observado.

Una de las ventajas de la estrategia de dos etapas presentada, a la que llamamos FISTA+LS, es que permite obtener soluciones *sparse* del *gradient* sin el uso de información *a priori*, o el uso de una matriz de covarianza o escala que aporte información sobre las amplitudes de los atributos. Además, como la función de costo a minimizar al resolver el LASSO es una función convexa no es necesario empezar el proceso de inversión desde un modelo inicial cercano a la solución óptima. Por otro lado, al utilizar FISTA+LS la elección del parámetro de compensación  $\mu$  se vio facilitada debido a que existe un amplio rango de valores que conducen a una solución similar. En las pruebas numéricas dicho parámetro  $\mu$  fue estimado utilizando el principio de discrepancia, por medio de la construcción de una curva de Pareto. Se vio que la curva de Pareto correspondiente a FISTA+LS presenta un quiebre en los valores de  $\mu$  óptimos, a diferencia de la curva obtenida al resolver el LASSO, que muestra un comportamiento suave. Esta característica permite una estimación más sencilla de  $\mu$  cuando el nivel de ruido en el dato no es conocido como sucede, por ejemplo, al trabajar con datos de campo.

En pruebas sobre datos sintéticos 1D con niveles de ruido significativo el método propuesto probó ser robusto y estable, ofreciendo soluciones *sparse* muy precisas, comparables a las obtenidas con VFSA pero a un costo computacional menor. A pesar de que FISTA no es un método de optimización estocástico, se realizó un análisis esta-

dístico, utilizando datos con diferentes realizaciones de ruido aleatorio, con el fin de estudiar la estabilidad de las soluciones estimadas por FISTA+LS frente al ruido en el dato. Los pequeños valores de las desviaciones estándar de las magnitudes invertidas resaltan la consistencia de las soluciones. Las pruebas con datos sintéticos 2D demostraron que el método es capaz de obtener soluciones *sparse* que honran al dato y presentan buena continuidad lateral. Sin embargo, algunos reflectores cercanos, como los que se observan en el *pinch-out* ubicado a los 0.2 s entre los *gathers* #40 y #60, no fueron resueltos correctamente. Pruebas en datos de campo mostraron que la técnica propuesta es capaz de ofrecer imágenes del *intercept* y el *gradient* que honran el dato observado con precisión. Debido al bajo costo computacional del algoritmo utilizado es posible trabajar con volúmenes grandes de datos.

### Capítulo 7: Inversión *sparse* mediante regularización con la norma $l_{2,1}$

En el capítulo 7 presentamos una estrategia de inversión para obtener soluciones *sparse* utilizando la aproximación de Aki y Richards de tres términos. El método propuesto apunta a obtener soluciones *sparse* de  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ , y si la información *a priori* disponible es suficiente, estimar soluciones *blocky* de los parámetros elásticos  $\alpha$ ,  $\beta$  y  $\rho$ . Como término de regularización utilizamos la norma mixta  $l_{2,1}$ . Esta norma, que es una extensión de la norma  $l_1$ , permite estimar soluciones *sparse* con estructura. La razón por la cual se optó por incorporar dicha norma como término de regularización es que permite obtener soluciones *sparse/blocky* y, al mismo tiempo, restringir los valores de las amplitudes de los parámetros estimados utilizando información *a priori*. A diferencia de lo analizado en el capítulo 6, donde para estimar soluciones *sparse* del *intercept* y el *gradient* no se necesitó de información *a priori*, al utilizar la aproximación de Aki y Richards de tres términos esta información es necesaria si se pretende obtener una correcta estimación de las amplitudes de  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ . Dicha información se incorpora al proceso de inversión por medio de una matriz de covarianza o una matriz de escala. Estas matrices deben ser estimadas previo a la inversión utilizando, por ejemplo, información de pozo. Una correcta estimación de dichas matrices es crucial para obtener soluciones que honren al dato observado y, al mismo tiempo, a la información *a priori*.

Al igual que en el capítulo 6 la función de costo obtenida es convexa, porque la norma  $l_{2,1}$  también lo es, por lo que no es necesario empezar el proceso de inversión desde un modelo inicial cercano a la solución óptima para estimar el minimizador de la función. Además la función de costo puede ser minimizada utilizando FISTA con pequeños ajustes en la función umbral. La estrategia de inversión propuesta es versátil, permitiendo sumar a la función de costo un término adicional conteniendo información sobre la tendencia de baja frecuencia y recuperarla al estimar soluciones *blocky* de las velocidades y densidades.

El método propuesto fue probado en datos sintéticos 1D y 2D con significativos niveles de ruido, y en datos de campo 2D. Sobre dato sintético 1D se vio que, dada la información *a priori* adecuada, el método es capaz de obtener soluciones *blocky* de los parámetros físicos que presentan una correcta estimación de las magnitudes. Los resultados obtenidos superaron a aquellos estimados utilizando mínimos cuadrados y utilizando la norma  $l_1$  como regularización. En este caso, la inversión llevada a cabo utilizando la norma  $l_1$  no consistió en utilizar un método de dos etapas como en la estrategia planteada en el capítulo 6, sino que se optó por estabilizar las amplitudes

utilizando una matriz de escala.

En las soluciones estimadas utilizando la norma  $l_{2,1}$  se observa que cuando se dispone de suficiente información *a priori* como para construir la matriz de covarianza completa los resultados obtenidos superan en calidad a aquellos estimados utilizando sólo una matriz de escala. Sin embargo, estos últimos igual son aceptables, mostrando un marcado comportamiento *blocky* y una correcta estimación de las amplitudes. Sobre datos sintéticos 2D se mostró que la estrategia es capaz de conseguir soluciones *blocky* con buena continuidad lateral y correcta estimación de las amplitudes, a la vez que fue posible obtener soluciones *sparse* de las reflectividades que honran a las verdaderas. Del análisis llevado a cabo utilizando la norma  $l_1$  y la norma  $l_{2,1}$ , usando en ambos casos sólo la matriz de escala, se vio que las soluciones estimadas utilizando la norma mixta  $l_{2,1}$  superaron en calidad a aquellas obtenidas utilizando la norma  $l_1$ . Esto se debe a que la norma  $l_{2,1}$  restringe mejor el espacio de posibles soluciones y distribuye de manera más equilibrada la energía entre las reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ .

Las pruebas sobre datos de campo se realizaron sólo con el propósito de estimar reflectividades  $R_\alpha$ ,  $R_\beta$  y  $R_\rho$ , debido a la falta de información *a priori* necesaria para estimar soluciones *blocky* de los parámetros elásticos. La inversión, realizada sobre 400 *angle-gather*, muestra que la estrategia propuesta es capaz de estimar soluciones *sparse* de las tres reflectividades, que honran al dato observado correctamente y presentan una buena continuidad lateral. Al igual que en la estrategia desarrollada en el capítulo anterior, el método propuesto es económico en términos de costo computacional y de fácil aplicación debido al algoritmo de optimización utilizado.

## 8.2. Discusión

Aunque utilizamos dos técnicas diferentes, los resultados obtenidos en el capítulo 5 son comparables con los obtenidos en el capítulo 6, ya que ambas técnicas fueron desarrolladas para estimar soluciones *sparse* utilizando la aproximación de Shuey de dos términos. Una comparación cualitativa de las Figuras 6.17 y 5.12 muestra que ambos métodos fueron capaces de estimar soluciones *sparse* que honran a las observaciones cuando se realiza la inversión sobre datos de campo. Sin embargo, se puede ver que las soluciones estimadas utilizando VFSA presentan una mejor continuidad lateral que aquellas estimadas utilizando FISTA+LS, especialmente del *gradient*. Esto se debe a que, para evitar las inestabilidades que podrían llegar a surgir en el *debiasing step*, la segunda etapa del método se realiza sobre soluciones estimadas a partir de resolver el LASSO utilizando un valor grande del parámetro de compensación  $\mu$ , es decir, soluciones muy *sparse*. Por otro lado, comparando las Figuras 5.9 y 6.14, se ve que las soluciones obtenidas con VFSA en algunos casos resuelven mejor los reflectores cercanos. Esto puede observarse en el *pinch-out* ubicado a los 0.2 s entre los *gathers* #40 y #60. Allí VFSA fue capaz de resolver correctamente dicha estructura, mientras que la solución obtenida con FISTA+LS no. La razón de este comportamiento, dado un valor de  $\mu$ , es que las soluciones estimadas con la estrategia de dos etapas en general son menos *sparse* que las estimadas utilizando VFSA. Esto se debe a que el número  $L_S$  de coeficientes de reflexión no nulos que utiliza VFSA se fija previamente, mientras que el número de reflectores no nulos que estimará el FISTA al resolver el LASSO depende de  $\mu$ , que es sensible al ruido en el dato. Por lo tanto, cuando se resuelve el LASSO en

lugares donde dos reflectores se aproximan, como en el *pinch-out* mencionado, en vez de ajustar el dato utilizando varios coeficientes de reflexión, lo que haría inestable la solución por mínimos cuadrados del *debiasing step*, FISTA+LS ubicará un solo reflector o incluso ninguno, dependiendo de qué tanto afecte esto al *misfit* y la norma  $l_1$ .

A pesar de estas aparentes desventajas, las soluciones estimadas por el FISTA+LS se obtuvieron en una fracción del costo computacional que la requerida por VFSA. Mientras que el método basado en VFSA resuelve un sistema de ecuaciones en cada iteración, FISTA+LS sólo lo hace una única vez al final del proceso iterativo. Por otro lado es más sencillo y económico estimar el valor del parámetro de compensación  $\mu$  que del número mínimo de coeficientes de reflexión  $L_s$ . Tampoco hay que perder de vista que las soluciones obtenidas con VFSA son el resultado de varias realizaciones del método, mientras que las soluciones obtenidas con FISTA+LS se obtuvieron realizando la inversión una única vez.

Si bien las soluciones obtenidas en el capítulo 7 no son directamente comparables con las obtenidas en los capítulos 5 y 6 pues se utilizaron aproximaciones distintas a las ecuaciones de Zoeppritz, es importante notar algunas diferencias entre los resultados. Una comparación cualitativa de las Figuras 6.17 y 7.10 muestra una mejor continuidad lateral en las soluciones estimadas sobre dato de campo cuando se utiliza la estrategia propuesta en el capítulo 7. Al igual que sucedió al comparar las soluciones estimadas con FISTA+LS y las estimadas con VFSA, esto se debe a que el *debiasing step* utilizado en el capítulo 6 se realiza sobre soluciones *sparse* estimadas con un valor grande de  $\mu$  con el fin de evitar las inestabilidades que podrían llegar a surgir en la resolución del problema de mínimos cuadrados. Como ya fue visto en los ejemplos sobre datos sintéticos de este capítulo, la norma  $l_{2,1}$  subestima menos las amplitudes estimadas que la norma  $l_1$ . Como consecuencia las primeras muestran una mejor continuidad lateral. En este sentido, es importante destacar que el hecho de utilizar la norma  $l_{2,1}$  pesada suplanta la necesidad de realizar el *debiasing step*. Dada una matriz de escala que relacione el *intercept* y el *gradient* las soluciones obtenidas en el capítulo 6 podrían ser estimadas utilizando la estrategia propuesta en el capítulo 7, quizás incluso con mejor continuidad lateral. Sin embargo, como ya fue explicado, la determinación de dichas matrices no es fácil, incluso cuando la información de pozo está disponible, por lo que la técnica propuesta en el capítulo 6 sigue siendo una buena opción a pesar de los mínimos problemas de continuidad lateral que presentan las soluciones.

### 8.3. Contribución

Si bien las tres estrategias presentadas en los capítulos 5, 6 y 7 comparten el mismo objetivo, cada una posee características diferentes. Al momento de realizar una inversión *sparse* la elección de una u otra de las estrategias presentadas dependerá, al final, del dato disponible sobre el cual se realizará la inversión y de las características de las soluciones que se pretende estimar.

En la primer estrategia, presentada en el capítulo 5, se procedió a estimar soluciones sub-óptimas del problema de inversión regularizado mediante la norma  $l_0$  y resuelto mediante VFSA, lo que permitió estimar soluciones *sparse* del *intercept* y del *gradient*. La estrategia presentada en dicho capítulo ha sido publicada en la revista de circulación nacional *Geoacta* (Pérez y Velis, 2012) y presentada como resúmenes expandidos en la

“XXV Reunión Científica de la Asociación Argentina de Geofísicos y Geodestas, Córdoba, Argentina” (Pérez y Velis, 2010), en la “81<sup>st</sup> Society of Exploration Geophysicists International Exhibition and Annual Meeting, San Antonio, Texas, USA” (Pérez y Velis, 2011a) y como trabajo completo en el “VIII Congreso de Exploración y Desarrollo de Hidrocarburos, Mar del Plata, Argentina” (Pérez y Velis, 2011b).

La segunda estrategia, presentada en el capítulo 6, consistió en un método de dos etapas basado en minimizar la norma  $l_1$  por medio de FISTA, realizando luego un paso correctivo de las amplitudes estimadas utilizando mínimos cuadrados. Esta estrategia también permitió estimar soluciones *sparse* del *intercept* y del *gradient*. La estrategia presentada en dicho capítulo ha sido publicada en la revista de circulación internacional *Geophysics* (Pérez et al., 2013a) y presentada como resumen expandido en la “83<sup>rd</sup> Society of Exploration Geophysicists International Exhibition and Annual Meeting, Houston, Texas, USA” (Pérez et al., 2013b), y como trabajos completos en el “X Congreso Argentino de Mecánica Computacional, MECOM 2012, Salta, Argentina” (Pérez et al., 2012), y en el “XX Congreso sobre Métodos Numéricos y sus Aplicaciones, ENIEF 2013, Mendoza, Argentina.” (Pérez et al., 2013c),

La tercer estrategia, presentada en el capítulo 7 utilizó como regularización la norma  $l_{2,1}$ , lo que permitió incorporar información *a priori* por medio de matrices de covarianza o de escala. Esta estrategia permitió estimar soluciones *sparse* de los parámetros de la aproximación de Aki y Richards de tres términos y, dada información *a priori* adecuada, estimar soluciones de tipo *blocky* de los parámetros físicos del subsuelo. La estrategia presentada en dicho capítulo ha sido presentada como resumen expandido en la “84<sup>th</sup> Society of Exploration Geophysicists International Exhibition and Annual Meeting, Denver, Colorado, USA” (Pérez et al., 2014a), y en la “XXVII Reunión Científica de la Asociación Argentina de Geofísicos y Geodestas, San Juan, Argentina” (Pérez et al., 2014b).

## 8.4. Trabajos a futuro

Aunque en esta tesis se procedió a realizar inversión de datos sísmicos *prestack* utilizando las aproximaciones de Shuey de dos términos y la aproximación de Aki y Richards de tres términos, es importante destacar que las pruebas sobre datos sintéticos y datos de campo realizadas no deben ser consideradas limitantes de las capacidades de cada uno de los métodos presentados. Todas las técnicas analizadas permiten la utilización de otras aproximaciones a las ecuaciones de Zoeppritz. Por ejemplo, la aproximación de Fatti et al. (1994) de dos términos podría ser utilizada junto con el FISTA+LS para estimar soluciones *sparse* de las reflectividades  $R_{I_P}$  y  $R_{I_S}$ , o incluso soluciones *blocky* de las impedancias  $I_P$  e  $I_S$ . Asimismo, herramientas utilizadas en una estrategia pueden ser incorporadas a las otras. Por ejemplo, la estrategia de incorporar información *a priori* por medio de matrices de covarianza o de escala que fue utilizada en el capítulo 7, podría también ser utilizada para restringir las amplitudes estimadas al utilizar VFSA o FISTA+LS. También es importante destacar que, como se describió en el capítulo 2, hay gran cantidad de métodos de optimización, además de VFSA y FISTA, que pueden ser utilizados para minimizar las funciones de costo.

En esta tesis nos concentramos en utilizar información obtenida a través de las ondas  $P$  reflejadas, pero estas no son las únicas que pueden ser utilizadas para realizar

la inversión *prestack*. A modo de ejemplo se pueden citar los trabajos de Hampson et al. (2005) y Hampson y Russell (2013), quienes proponen incorporar información de las ondas convertidas *PS* para realizar estimación simultánea de la impedancia de las ondas *P* y *S*, y de la densidad. Las técnicas propuestas en esta tesis pueden ser modificadas con el fin de incorporar dicha información, hecho que podría contribuir a disminuir el espacio de posibles soluciones y, por lo tanto, la ambigüedad del problema inverso.

Si bien en los ejemplos numéricos realizados para probar las estrategias propuestas se asumió que los medios son homogéneos e isotrópicos, con frecuencia estas hipótesis no se satisfacen completamente cuando se trabaja con datos de campo. En este sentido, el estudio de las propiedades anisotrópicas del medio es de vital importancia para el estudio de yacimientos no convencionales. Particularmente, en lo que se refiere al tema que abarca esta tesis, sería importante extender las técnicas propuestas al estudio de la *Azimuthal amplitude-variation-with-offset* (AVAz) (Rüger y Gray, 2014). El análisis de AVAz consiste en estudiar las variaciones de las amplitudes de las ondas reflejadas no solamente en función del ángulo de incidencia, sino también en función del azimut (o dirección) de la cual provienen. Esto permitiría realizar estudios de anisotropía y de los sistemas de fracturas presentes en los reservorios de hidrocarburos.

### 8.5. Desarrollo

Para la realización de esta tesis fue necesario el desarrollo de herramientas computacionales, la mayoría programadas en FORTRAN 2003, en GNU Octave, y en MATLAB<sup>®</sup>, entre las que se encuentran:

- El algoritmo *Very fast simulated annealing* programado en FORTRAN 2003, basado en el código desarrollado por Velis (1998).
- El algoritmo *Fast iterative shrinkage-thresholding algorithm*, programado en FORTRAN 2003.
- Programas para generar datos sintéticos a partir de modelos de velocidades y densidades, contemplando respuesta AVA por medio de las ecuaciones de Zoeppritz, Shuey de dos términos y Aki y Richards de tres términos.
- Programas de soporte a las estrategias presentadas. Por ejemplo, programas para estimar el máximo autovalor de una matriz, utilizado por el FISTA, por medio del método de las potencias de Rayleigh, o programas para generar las matrices de covarianza utilizadas en el capítulo 7, entre otras.

Todas estas herramientas son parte de la batería de programas y códigos que implementan las diversas estrategias descritas a lo largo de la tesis. Además de las herramientas desarrolladas, se hizo uso de otras ya existentes de gran utilidad, como el *software open source* de procesamiento de datos sísmicos *Seismic Unix*, el *software open source* de interpretación sísmica *OpendTect* y algoritmos numéricos varios desarrollados en el libro *Numerical Recipes in Fortran 90: The Art of Parallel Scientific Computing* (Press et al., 1996).

# Apéndice A

## *Very fast simulated annealing*

### A.1. Introducción

*Simulated annealing* (SA) es un algoritmo iterativo estocástico para estimar soluciones casi-óptimas de problemas de optimización complejos, siendo capaz de estimar mínimos globales de funciones multimodales y/o discontinuas. El algoritmo se deriva de la mecánica estadística, y su nombre proviene del proceso de recocido de metales, una técnica metalúrgica que consiste en calentar y luego enfriar el material de manera controlada para variar sus propiedades físicas. En 1953, Metropolis et al. (1953) presentaron una técnica Monte Carlo para modelar la evolución de las propiedades de un sólido a una temperatura dada. Tres décadas después la técnica fue generalizada y aplicada a problemas de optimización no lineales (Kirkpatrick et al., 1983). En el caso de problemas no lineales, las incógnitas juegan el rol de las partículas en el sólido, y la función de costo representa la energía del sistema. En cada iteración del SA, el espacio del modelo es perturbado aleatoriamente y el nuevo modelo es aceptado (o rechazado) dependiendo si el valor de la función de costo disminuyó o se incrementó. El SA acepta ocasionalmente incrementos en el valor de la función de costo, esta característica le permite al algoritmo escapar de los mínimos locales. Inicialmente la temperatura (que es un parámetro de control) es lo suficientemente elevada como para que todos los modelos que producen un aumento de la función de costo sean aceptados. Por el contrario, a bajas temperaturas la probabilidad de aceptar un modelo que produce un aumento de la función de costo es pequeña. Las perturbaciones aleatorias de los modelos decrecen gradualmente con la temperatura siguiendo un esquema de enfriado previamente establecido, y el proceso iterativo concluye al alcanzar un número máximo de iteraciones o cuando no hay un cambio significativo del valor de la función de costo tras cierta cantidad de iteraciones. En ese momento, al concluir el proceso, se dice que el sistema está “cristalizado” en el sentido que se alcanza el mínimo estado de energía, lo que equivale al mínimo absoluto de la función de costo.

Un sistema formado por un número grande partículas interactuando entre sí a una temperatura dada, como es el caso de un fluido o un sólido fundido, puede ser descrito por la posición de dichas partículas. A dicha configuración o estado lo llamaremos  $x$ . Si  $T$  es la temperatura de equilibrio térmico del sistema, que posee energía  $E(x)$ , entonces la probabilidad  $\pi_T(x)$  de que el sistema esté en el estado  $x$  está dada por la distribución de Boltzmann

$$\pi_T(x) = \frac{1}{Z(T)} \exp \left\{ \frac{-E(x)}{\kappa T} \right\}, \quad (\text{A.1})$$

donde  $\kappa$  es la constante de Boltzmann, y

$$Z(T) = \sum_{\omega \in X} \exp \left[ \frac{-E(\omega)}{\kappa T} \right], \quad (\text{A.2})$$

es la función de partición, siendo  $X$  el conjunto de todos los posibles estados  $x$ .

El algoritmo de Metropolis (MA) (Metropolis et al., 1953) es un simple algoritmo estocástico para simular esta clase de procesos. En cada iteración se realiza, sobre el estado  $x$  con energía  $E(x)$ , una pequeña perturbación aleatoria  $\Delta x$  y se calcula el cambio en la energía  $\Delta E = E(x + \Delta x) - E(x)$ . La relación  $h$  entre las probabilidades de ambos estados está dada por

$$h = \frac{\pi_T(x + \Delta x)}{\pi_T(x)} = \exp \left[ \frac{-\Delta E}{\kappa T} \right]. \quad (\text{A.3})$$

Si la energía del nuevo estado es menor que la del estado previo, es decir si  $h > 1$ , entonces la nueva configuración es aceptada sin restricciones. Si  $h < 1$ , es decir, si la energía del nuevo estado es mayor que la del estado previo entonces la nueva configuración es aceptada con probabilidad  $h$ . Este criterio para aceptar o rechazar los nuevos estados es conocido como ‘‘Criterio de Metropolis’’. Puede demostrarse que cuando el número de iteraciones tiende a infinito, la probabilidad de que el sistema presente una dada configuración  $x$  es igual a  $\pi_T(x)$ , y por lo tanto la distribución de la configuración generada converge a la distribución de Boltzmann (Geman y Geman, 1984).

Es de interés estudiar el comportamiento del sistema en estados de baja energía con el fin de comprender la estructura cristalina del sólido fundido a medida que la temperatura disminuye gradualmente. Para ello la temperatura debe decrecer lo suficientemente despacio como para que predomine la probabilidad de obtener estados de baja energía de acuerdo con la distribución de Boltzmann. Estos estados se pueden obtener manteniendo el sistema en equilibrio térmico. Por el contrario, cuando la temperatura baja rápidamente, proceso conocido como templado, el sistema no puede alcanzar el equilibrio térmico, la probabilidad de alcanzar un estado de baja energía se reduce y la estructura resultante del material puede estar lejos de constituir un tejido cristalino.

## **A.2. Simulated annealing**

En 1983 Kirkpatrick et al. (1983) hizo una analogía entre el sistema físico descrito y el problema, en el campo de la ingeniería electrónica, de encontrar soluciones sub-óptimas del particionado, ensamblado y cableado de *very large scale integration* (VLSI). Básicamente SA consiste en aplicar el algoritmo de Metropolis en un proceso iterativo en el cual la temperatura decrece gradualmente. Para aplicar SA en la resolución de problemas de optimización se deben hacer las siguientes analogías:

- Identificar la configuración de los parámetros del modelo con la configuración de las partículas del sistema.
- Identificar la función objetivo con la energía del sistema.
- Identificar el problema de encontrar una solución sub-óptima con el proceso de estimar el menor estado de energía del sistema.

La constante de Boltzmann  $\kappa$  se toma igual a la unidad y la temperatura  $T$  se convierte en un parámetro de control. Además, se deben definir tres funciones de gran importancia:

1.  $g(\Delta x)$ : La función de densidad de probabilidad para generar nuevas configuraciones del modelo.
2.  $h(\Delta\Phi)$ : La función de densidad de probabilidad de aceptar la nueva configuración.
3.  $T_k$ : el esquema de enfriado, es decir la temperatura  $T$  en la iteración  $k$ .

Aquí  $x$  representa el espacio del modelo de  $M$  parámetros. Ambas funciones de densidad de probabilidad dependen también de la temperatura. Existen muchas variaciones de SA y la principal diferencias entre ellas reside en la elección de las tres funciones. La Figura A.1 muestra el diagrama de flujo de un algoritmo SA típico. En cada iteración un nuevo estado es generado por la función  $g(\Delta x)$ . Se lo evalúa según el criterio de Metropolis y, después de bajar la temperatura de acuerdo con el esquema de enfriado  $T_k$ , se comienza una nueva iteración. El proceso se detiene cuando se alcanza un número máximo de interacciones  $k_{max}$  o se satisface algún otro criterio de corte. Cuando el valor de la función de costo  $\Phi$  se incrementa, la probabilidad de aceptar dicho estado está dado por  $h$ . En la práctica, el nuevo estado es aceptado si  $h$  es mayor que un número aleatorio  $r$ , que es generado a partir de una función de densidad de probabilidad uniforme  $U[0, 1]$ .

### A.2.1. Función de generación de modelos

Inicialmente el algoritmo de Metropolis realizaba una búsqueda del espacio del modelo utilizando una función de probabilidad  $g$  uniforme. Tiempo después se demostró que se puede conseguir una convergencia más rápida a las soluciones utilizando una función de probabilidad gaussiana, que se concentra en los estados de menor energía estrechándose a medida que la temperatura disminuye (Szu y Hartley, 1987). En ese caso, la función de probabilidad gaussiana de  $M$  dimensiones puede ser expresada como

$$g(\Delta x) = (2\pi T)^{-M/2} \exp \left[ \frac{-\Delta x^2}{2T} \right] \quad (\text{A.4})$$

donde  $\Delta x = x - x_0$  es la perturbación aplicada a la configuración previa  $x_0$ . Aquí la temperatura  $T$  puede ser vista como una medida de la fluctuación de  $g$  alrededor de  $x_0$ . Cuando la temperatura es alta, el espacio del modelo es explorado de manera aproximadamente uniformemente. Contrariamente, cuando la temperatura disminuye las perturbaciones tienden a ser menores.

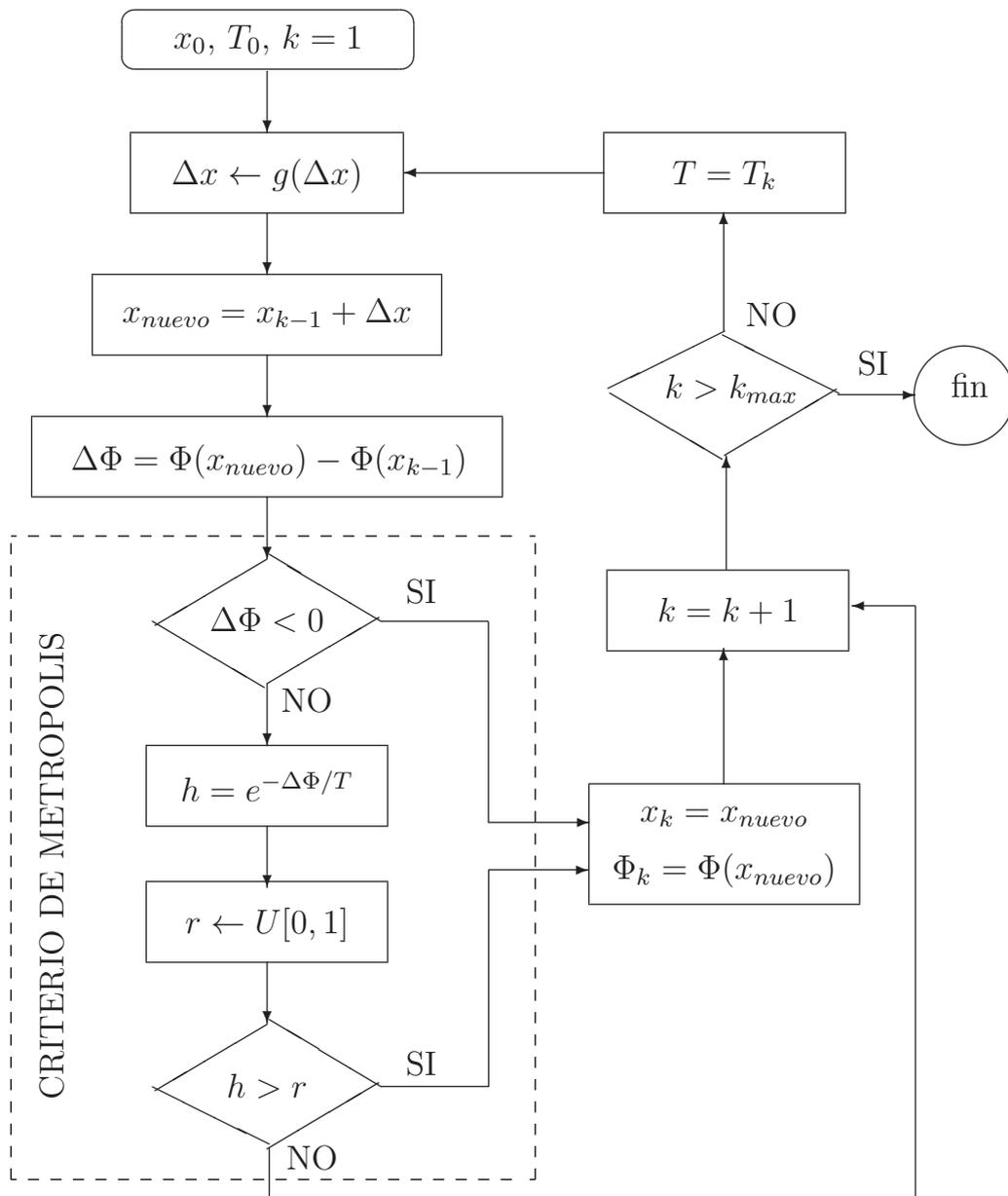


Figura A.1. Diagrama de flujo del *Simulated annealing* (Velis, 1998)

### A.2.2. Función de aceptación

Generalmente, el criterio de Metropolis es utilizado para aceptar o rechazar una nueva configuración:

$$h(\Delta\Phi) = \frac{1}{1 + \exp\left[\frac{\Delta\Phi}{T}\right]} \simeq \exp\left[\frac{-\Delta\Phi}{T}\right] \quad (\text{A.5})$$

Esta expresión se basa en la probabilidad de obtener un nuevo estado con energía  $\Phi_{k+1}$  relativa al estado previo con energía  $\Phi$ , donde  $\Delta\Phi = \Phi_{k+1} - \Phi$ . Esa es básicamente la distribución de Boltzmann. Siguiendo a Metropolis, una nueva configuración es aceptada sin condición si  $\Delta\Phi < 0$ , y aceptada con probabilidad  $h(\Delta\Phi)$  si  $\Delta\Phi \geq 0$ . Claramente, si la temperatura se acerca a cero la probabilidad de aceptación decrece exponencialmente y sólo los estados más bajos de energía son aceptados. Esto es fundamental en el funcionamiento de SA. Mientras otros algoritmos de optimización sólo buscan modelos que hacen disminuir la función de costo, el SA permite explorar modelos que incrementan el valor de la misma. Este comportamiento permite explorar de manera más amplia el espacio del modelo, evitando quedar atrapado en mínimos locales.

### A.2.3. Esquema de enfriado

El esquema utilizado para disminuir la temperatura del proceso juega un rol clave en el proceso de optimización llevado a cabo por el SA. Es deseable poder bajar la temperatura lo más rápido posible con el fin de alcanzar el estado de mínima energía en la menor cantidad de iteraciones posibles, sin embargo la teoría indica que para alcanzar el mínimo absoluto de la función de costo la velocidad de enfriamiento no puede ser más rápida que una cierta función prefijada. Volviendo a la analogía del proceso térmico descrito en la primer sección, si el material es enfriado con la lentitud suficiente, la estructura cristalina resultante presenta una configuración de mínima energía. Por el contrario, si la temperatura disminuye demasiado rápido, se empezaran a observar distorsiones en la estructura cristalina que representan estados de energía más altos.

Por lo tanto, para alcanzar el estado de mínima energía se debe conseguir el estado de equilibrio térmico antes de disminuir la temperatura. Originalmente, este estado era alcanzado manteniendo la temperatura a un valor constante (Metropolis et al., 1953). Fueron Geman y Geman (1984) quienes demostraron que el mínimo global  $\Phi$  puede ser alcanzado (estadísticamente hablando) si la temperatura no disminuye más rápido que

$$T_k = \frac{T_0}{\ln(1 + k)}, \quad (\text{A.6})$$

donde  $T_0$  es una constante y  $k$  es el número de iteración. Si la temperatura disminuye más rápido que lo que expresa dicha ecuación, se tendrá una cristalización prematura y el sistema puede quedar atrapado en un mínimo local.

### A.3. *Very fast simulated annealing*

Ingber (1989) propuso una técnica de SA que permite un esquema de enfriado más rápido, el *Very Fast Simulated annealing* (VFSA). Esta técnica, en algunos casos, es incluso superior a los métodos evolutivos tales como el *genetic algorithm* (Ingber y Rosen, 1992). VFSA utiliza una función de generación  $g$  que permite una mejor exploración del espacio de los modelos. El espacio de búsqueda está restringido y además toma en cuenta las diferencias en la dimensión de cada parámetro, permitiendo que  $g$  se expanda o se contraiga de forma adaptativa dependiendo de la sensibilidad de la función de costo respecto de cada parámetro.

#### A.3.1. Descripción del algoritmo

En el VFSA, los nuevos parámetro  $x_i^{(k)}$  en la dimensión  $i$  son generados por medio de la variable aleatoria  $y_i$

$$x_i^{(k+1)} = x_i^k + y_i(B_i - A_i), \quad y_i \in [-1, 1]. \quad (\text{A.7})$$

Los nuevos parámetros  $x_i^{(k+1)}$  están restringidos dentro del rango  $[A_i, B_i]$ , y la ecuación (A.7) se repite para todo  $i$  hasta que se obtienen  $M$  nuevos parámetros. La función de generación  $g$ , utilizada por el VFSA está definida como

$$g(y) = \prod_{i=1}^M \frac{1}{2(|y_i| + T_i) \ln(1 + 1/T_i)}, \quad (\text{A.8})$$

donde se ha omitido el subíndice  $k$  por simplicidad. La distribución de probabilidad conjunta puede ser obtenida como

$$G(y) = \int_{-1}^{y_1} \cdots \int_{-1}^{y_M} g(y) dy_1 \cdots g(y) dy_M = \prod_{i=1}^M G_i(y_i), \quad (\text{A.9})$$

donde

$$G_i(y_i) = \frac{1}{2} + \frac{\text{sgn}(y_i) \ln(1 + |y_i|/T_i)}{2 \ln(1 + 1/T_i)}. \quad (\text{A.10})$$

Como resultado, se pueden generar nuevos puntos que sigan esta distribución a partir de una distribución uniforme  $u_i \in [0, 1]$  considerando

$$y_i = G^{-1}(u_i), \quad (\text{A.11})$$

por lo tanto

$$y_i = \text{sgn} \left( u_i - \frac{1}{2} \right) T_i \left[ (1 + 1/T_i)^{(2u_i - 1)} - 1 \right]. \quad (\text{A.12})$$

Habiendo definido  $g(y)$ , el esquema de enfriado que garantiza estadísticamente la convergencia al mínimo global está dado por

$$T_i = T_{0i} \exp \left[ -c_i k^{(1/M)} \right]. \quad (\text{A.13})$$

Este esquema de enfriamiento es mucho más veloz que el propuesto en la ecuación (A.6). Las constantes  $c_i$  no afectan la convergencia del algoritmo y su finalidad depende de cada aplicación en particular. Por ejemplo, para obtener una temperatura  $T_f$  en  $k_f$  iteraciones se elige

$$c_i = k_f^{(-1/M)} \log \left( \frac{T_{0i}}{T_{fi}} \right) \quad (\text{A.14})$$

Además del VFSA, existen otras variaciones del SA dependiendo de la forma de la distribución  $g(\Delta x)$  elegida. Por ejemplo, si  $g(\Delta x)$  es una distribución de Cauchy se obtiene el método conocido como *Fast Annealing* (FA). Sin embargo, el método más eficiente es VFSA.

### A.3.2. Temperatura inicial

La elección de la temperatura inicial  $T_0$  con la cual empezar el proceso iterativo es a veces crítica en el SA. Si  $T_0$  es demasiado alta, al sistema le tomará muchas iteraciones para alcanzar un estado de baja energía. Contrariamente, si  $T_0$  es demasiado baja, el sistema cristalizará demasiado pronto. Esto no sucede con VFSA, donde  $T_0$  es relativamente arbitraria. Además como el esquema de enfriado es mucho más rápido, elegir un valor de  $T_0$  demasiado alto no es crítico. En la práctica la temperatura inicial deberá ser tal que el valor de la función de costo a dicha temperatura esté a un desvío estándar de su valor medio (White, 1984). Esto es así para asegurarse que al principio del proceso casi todas las configuraciones propuestas sean aceptadas (es decir, el sistema está completamente “fundido”). Con este propósito alcanza con elegir  $T_0$  igual al valor medio de la función de costo para un conjunto arbitrario de valores elegidos al azar del espacio del modelo. Es importante remarcar, para el caso del VFSA, la diferencia entre la temperatura de aceptación, asociada al criterio de Metropolis, y la temperatura de los parámetros, asociada a la función de generación  $g$ . Aunque el esquema de enfriamiento para ambas temperatura es generalmente el mismo, los valores iniciales de cada temperatura pueden ser diferentes. En el primer caso  $T_0$  se escribe de la forma ya descrita. En el segundo caso  $T_0$  generalmente es elegido igual a la unidad, para que de esta forma el espacio del modelo sea explorado independiente de la configuración inicial.



# Apéndice B

## ISTA y FISTA

### B.1. Algoritmo ISTA

El desarrollo de algoritmos eficientes para resolver el problema LASSO, es decir, encontrar el mínimos de

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + \mu\|\mathbf{m}\|_1, \quad (\text{B.1})$$

es un campo de investigación activo y en constante desarrollo. Un importante algoritmo es el *Iterative shrinkage-thresholding algorithm* (ISTA) (Daubechies et al., 2004). El algoritmo ISTA es una combinación de una iteración de Landweber (Landweber, 1951) y una función tipo umbral (*soft-thresholding function*). Debido a esto el ISTA también es llamado *Thresholded-Landweber Algorithm* (TL). La derivación que haremos de ISTA en este apéndice está basada en el método de mayorización-minimización (MM). Dicho método consiste en reemplazar un problema de difícil minimización por una secuencia de problemas más sencillos. Al hacer MM se genera una secuencia de vectores  $\mathbf{m}_k$ ,  $k = 0, 1, 2, \dots$  que converge a la solución deseada. Debido a que  $J(\mathbf{m})$  de la ecuación (B.1) no puede ser fácilmente minimizada, utilizar el algoritmo MM es enfoque útil para resolver el problema.

La idea de la mayorización-minimización se puede describir de la siguiente manera. Supongamos que tenemos un vector  $\mathbf{m}_k$  que será un estimado del vector que minimiza  $J(\mathbf{m})$ . Basándonos en  $\mathbf{m}_k$  nos interesa encontrar un nuevo vector  $\mathbf{m}_{k+1}$  que haga decrecer  $J(\mathbf{m})$ ; es decir, busquemos un  $\mathbf{m}_{k+1}$  tal que

$$J(\mathbf{m}_{k+1}) < J(\mathbf{m}_k). \quad (\text{B.2})$$

Para esto se debe encontrar una función  $G(\mathbf{m})$  que mayorice  $J(\mathbf{m})$ , es decir  $G(\mathbf{m}) \geq J(\mathbf{m})$  para todo  $\mathbf{m}$ , y luego minimizar dicha función para obtener  $\mathbf{m}_{k+1}$ . Además  $G(\mathbf{m})$  debe ser igual a  $J(\mathbf{m})$  para  $\mathbf{m}_k$ . Para que el método sea útil  $G(\mathbf{m})$  debe ser una función fácil de minimizar. Además, la función  $G(\mathbf{m})$  será diferente en cada iteración, por lo que la llamaremos  $G_k(\mathbf{m})$ .

El algoritmo de mayorización-minimización MM puede resumirse en estos pasos:

1. Establecer  $k = 0$ . Inicializar  $\mathbf{m}_0$ .
2. Elegir  $G_k(\mathbf{m})$  tal que

- a)  $G_k(\mathbf{m}) \geq J(\mathbf{m})$ .
  - b)  $G_l(\mathbf{m}_k) = J(\mathbf{m}_k)$ .
3. Tomar  $\mathbf{m}_{k+1}$  igual al minimizador de  $G_k(\mathbf{m})$ .
  4. Establecer  $k = k + 1$  e ir al paso 2.

A continuación se describirán los pasos necesarios para deducir el algoritmo ISTA. La deducción consta de dos etapas: una iteración de Landweber y la aplicación de una función umbral.

### B.1.1. Iteración de Landweber

Aunque estamos interesados en minimizar la función  $J(\mathbf{m})$  de la ecuación (B.1) primero nos ocuparemos del problema más sencillo de minimizar la función de costo dada por

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2. \quad (\text{B.3})$$

Como se vio en el capítulo 2, el vector  $\hat{\mathbf{m}}$  que minimiza la ecuación (B.3) es

$$\hat{\mathbf{m}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{d}. \quad (\text{B.4})$$

Sin embargo la matriz  $(\mathbf{A}^T \mathbf{A})$  podría no ser invertible o estar mal condicionada. Utilizando el método de mayorización-minimización se puede evitar tener que resolver un sistema de ecuaciones para encontrar el mínimo de la ecuación (B.3). En cada iteración  $k$  del MM debemos encontrar una función  $G_k(\mathbf{m})$  tal que mayorice a  $J(\mathbf{m})$  pero que ambas tomen el mismo valor en  $\mathbf{m}_k$ . Además  $G_k(\mathbf{m})$  debe ser fácilmente minimizable. Dicha función se puede construir sumándole una función no-negativa a  $J(\mathbf{m})$ . Además la función no-negativa debe ser igual a cero en  $\mathbf{m}_k$  para que  $G_k(\mathbf{m}_k) = J(\mathbf{m}_k)$ . Una función que satisface todas las características deseadas es

$$G_k(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + (\mathbf{m} - \mathbf{m}_k)^T (\eta \mathbf{I} - \mathbf{A}^T \mathbf{A}) (\mathbf{m} - \mathbf{m}_k), \quad (\text{B.5})$$

donde el parámetro  $\eta$  debe ser elegido igual o mayor al máximo autovalor de  $\mathbf{A}^T \mathbf{A}$  para asegurar que el término sumado sea no-negativo para todo  $\mathbf{m}$ .

Siguiendo el método MM se debe encontrar el minimizador de  $G_k(\mathbf{m})$  para obtener  $\mathbf{m}_{k+1}$ . Desarrollando la ecuación (B.5) se tiene

$$G_k(\mathbf{m}) = \mathbf{d}^T \mathbf{d} + \mathbf{m}_k^T (\eta \mathbf{I} - \mathbf{A}^T \mathbf{A}) \mathbf{m}_k - 2(\mathbf{d}^T \mathbf{A} + \mathbf{m}_k^T (\eta \mathbf{I} - \mathbf{A}^T \mathbf{A})) \mathbf{m} + \eta \mathbf{m}^T \mathbf{m}. \quad (\text{B.6})$$

Luego, derivando respecto de  $\mathbf{m}$  se tiene

$$\frac{\partial G_k(\mathbf{m})}{\partial \mathbf{m}} = -2\mathbf{A}^T \mathbf{d} - 2(\eta \mathbf{I} - \mathbf{A}^T \mathbf{A}) \mathbf{m}_k + 2\eta \mathbf{m}. \quad (\text{B.7})$$

Igualando a cero y despejando  $\mathbf{m}$ , se obtienen los valores de  $\mathbf{m}_k$  para las diferentes iteraciones  $k$

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \frac{1}{\eta} \mathbf{A}^T (\mathbf{d} - \mathbf{A} \mathbf{m}_k). \quad (\text{B.8})$$

A este proceso se lo conoce como iteración de Landweber. Debido a las hipótesis sobre las cuales se basa el método MM, está garantizado que el valor  $J(\mathbf{m})$  de la ecuación (B.3) decrece con cada iteración. Notar que la ecuación anterior solo requiere de multiplicación de matrices y vectores.

La función  $G_k(\mathbf{m})$  dada en la ecuación (B.6) se puede expresar de forma más útil como

$$G_k(\mathbf{m}) = \eta \left\| \mathbf{m}_k + \frac{1}{\eta} \mathbf{A}^T (\mathbf{d} - \mathbf{A} \mathbf{m}_k) - \mathbf{m} \right\|_2^2 + K, \quad (\text{B.9})$$

donde  $K$  no depende de  $\mathbf{m}$ . A partir de esta expresión se puede ver que  $G_k(\mathbf{m})$  es una función cuadrática.

### B.1.2. Función umbral

Al igual que en la sección anterior, como parte del desarrollo del ISTA, nos concentraremos primero en la minimización de una función más sencilla que la dada en la ecuación (B.1). En este caso buscaremos minimizar la función

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{m}\|_2^2 + \mu \|\mathbf{m}\|_1. \quad (\text{B.10})$$

Este es un caso especial de la ecuación (B.1), tomando  $\mathbf{A} = \mathbf{I}$ . Esta función es convexa pero no diferenciable debido a  $\|\mathbf{m}\|_1$ . Sin embargo su mínimo puede ser estimado de forma sencilla.

Desarrollando la ecuación (B.10) se tiene

$$J(\mathbf{m}) = (d_1 - m_1)^2 + \mu |m_1| + (d_2 - m_2)^2 + \mu |m_2| + \dots + (d_N - m_N)^2 + \mu |m_N|. \quad (\text{B.11})$$

Notar que las variables  $m_i$ , con  $i = 1, \dots, N$ , no están acopladas por lo que se puede encontrar el mínimo de la función  $J(\mathbf{m})$  minimizando individualmente cada término. Debido a esto analizaremos el problema de minimizar la función escalar

$$f(m) = (d - m)^2 + \mu |m|. \quad (\text{B.12})$$

Tomando la derivada e igualando a cero se tiene

$$d = m + \frac{\mu}{2} \text{sign}(m), \quad (\text{B.13})$$

donde  $\text{sign}(\cdot)$  es la función signo. A partir de este resultado se deduce que el mínimo de la función (B.12) puede obtenerse aplicando una función umbral (*soft-thresholding function*) definida, para una variable genérica  $g$ , como

$$T_v\{g\} = \begin{cases} g(1 - v/|g|) & |g| \geq v \\ 0 & |g| < v \end{cases} \quad (\text{B.14})$$

donde  $v$  es el valor del umbral (*threshold*). La ecuación (B.14) puede ser expresada de manera más simple como

$$T_v\{g\} = \text{sign}(g)\max(0, |g| - v). \quad (\text{B.15})$$

Siguiendo esta idea, el minimizador de la ecuación (B.12) se obtendrá aplicando la función  $T_v\{g\}$  a los valores  $d$  con un valor umbral  $v = \mu/2$ , es decir

$$m = T_{\mu/2}\{d\}. \quad (\text{B.16})$$

Considerando que la variables de la función  $J(\mathbf{m})$  de la ecuación (B.10) no están acopladas, y que la solución se obtiene minimizando individualmente cada  $m_i$ , entonces el minimizador de dicha función se obtiene aplicando la función umbral a cada elemento de  $\mathbf{m}$  con  $v = \mu/2$ :

$$\mathbf{m} = T_{\mu/2}\{\mathbf{d}\}. \quad (\text{B.17})$$

Notar que la minimización de esta función no requiere de un proceso iterativo.

### B.1.3. *Iterative shrinkage-thresholding algorithm*

Consideremos ahora la función de costo  $J(\mathbf{m})$  dada por la ecuación (B.1)

$$J(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + \mu\|\mathbf{m}\|_1. \quad (\text{B.18})$$

Esta función puede ser minimizada combinando las dos técnicas explicadas en las secciones anteriores.

Para aplicar el algoritmo MM vamos construir la función  $G_k(\mathbf{m})$  sumando a  $J(\mathbf{m})$  la misma función no-negativa que en la ecuación (B.5), obteniendo

$$G_k(\mathbf{m}) = \|\mathbf{d} - \mathbf{A}\mathbf{m}\|_2^2 + (\mathbf{m} - \mathbf{m}_k)^T(\eta\mathbf{I} - \mathbf{A}^T\mathbf{A})(\mathbf{m} - \mathbf{m}_k) + \mu\|\mathbf{m}\|_1, \quad (\text{B.19})$$

que por medio de la ecuación (B.9) se puede expresar como

$$G_k(\mathbf{m}) = \eta \left\| \mathbf{m}_k + \frac{1}{\eta}\mathbf{A}^T(\mathbf{d} - \mathbf{A}\mathbf{m}_k) - \mathbf{m} \right\|_2^2 + \mu\|\mathbf{m}\|_1 + K, \quad (\text{B.20})$$

donde  $K$  es una constante que no depende de  $\mathbf{m}$ . Minimizar  $G_k(\mathbf{m})$  es equivalente a minimizar  $(1/\eta)G_k(\mathbf{m})$ , por lo tanto el paso  $\mathbf{m}_{k+1}$  de la iteración de Landweber se obtiene minimizando

$$\frac{1}{\eta}G_k(\mathbf{m}) = \left\| \mathbf{m}_k + \frac{1}{\eta}\mathbf{A}^T(\mathbf{d} - \mathbf{A}\mathbf{m}_k) - \mathbf{m} \right\|_2^2 + \frac{\mu}{\eta}\|\mathbf{m}\|_1, \quad (\text{B.21})$$

donde por simplicidad despreciamos  $K$  ya que no depende de  $\mathbf{m}$ . Notar que esta ecuación tiene exactamente la misma forma que la ecuación (B.10), que puede ser minimizada haciendo uso de la función umbral dada por la ecuación (B.15). Por lo tanto el paso  $\mathbf{m}_{k+1}$  de la iteración de Landweber se obtiene de la siguiente manera

$$\mathbf{m}_{k+1} = T_{\mu/2\eta} \left\{ \mathbf{m}_k + \frac{1}{\eta} \mathbf{A}^T (\mathbf{d} - \mathbf{A} \mathbf{m}_k) \right\}, \quad (\text{B.22})$$

donde  $\eta$  es mayor o igual al máximo autovalor de  $\mathbf{A}^T \mathbf{A}$ . Este proceso iterativo, que converge al minimizador de la ecuación (B.18), es el *Iterative Shrinkage-Thresholding Algorithm* (ISTA).

Paso a paso, el algoritmo ISTA puede resumirse de la siguiente manera

1. Establecer  $\eta \geq \kappa_{max}$ , donde  $\kappa_{max}$  es el máximo autovalor de  $\mathbf{A}^T \mathbf{A}$ .
2. Establecer una solución inicial  $\mathbf{m}_0$ ,
3. Para cada iteración  $k = 0, 1, 2, \dots$ :

a)

$$\mathbf{m}_{k+1} = T_{\mu/2\eta} \left\{ \mathbf{m}_k + \frac{1}{\eta} \mathbf{A}^T (\mathbf{d} - \mathbf{A} \mathbf{m}_k) \right\}, \quad (\text{B.23})$$

donde  $T_v\{\cdot\}$  es una función umbral suave (*soft-thresholding function*), que se aplica a cada elemento de su argumento vectorial y está definida como

$$T_v\{g\} = \begin{cases} g(1 - v/|g|), & |g| \geq v \\ 0, & |g| < v \end{cases} \quad (\text{B.24})$$

b)

$$k = k + 1, \quad (\text{B.25})$$

c) Verificar convergencia o criterio de corte.

El primer paso del proceso es necesario para evitar que el argumento de la función umbral sea negativo. Para encontrar el máximo autovalor puede utilizarse, por ejemplo, el método de las potencias de Rayleigh (Larson y Edwards, 1999, p.ej.), que es simple y eficiente.

ISTA es un algoritmo de optimización que no requiere inversión de matrices y que es sumamente sencillo de aplicar. Sin embargo está demostrada su lenta convergencia a la solución (Bredies y Lorenz, 2008). Si se descarta el segundo término de la ecuación (B.18), ISTA se reduce al método clásico del gradiente, y en ese caso se puede demostrar que el método converge al valor mínimo de la función  $J(\mathbf{m})$  con una tasa de convergencia no menor a  $O(1/k)$ , o “sublineal”.

## B.2. Algoritmo FISTA

Como se estableció en la sección anterior ISTA converge al valor mínimo de la función  $J(\mathbf{m})$  con una tasa de convergencia no menor a  $O(1/k)$ , o “sublineal”. Cuando se descarta el segundo término de la ecuación (B.18) el problema se reduce a minimizar una función convexa suave y ISTA se reduce al método del gradiente. Se ha probado que existen métodos de gradiente con una tasa de razón de convergencia  $O(1/k^2)$  que, al igual que el método del gradiente clásico, no requieren más de una evaluación del

gradiente en cada iteración, tan solo el uso de un  $\mathbf{m}$  adicional elegido ingeniosamente y fácil de calcular (Nesterov, 1983). Basados en esta idea Beck y Teboulle (2009) desarrollaron el *Fast iterative shrinkage-thresholding algorithm* (FISTA).

Paso a paso, el algoritmo FISTA puede resumirse de la siguiente manera

1. Establecer  $\eta \geq \kappa_{max}$ , donde  $\kappa_{max}$  es el máximo autovalor de  $\mathbf{A}^T \mathbf{A}$ .
2. Establecer  $\mathbf{z}_0 = \mathbf{m}_0$  y  $t_0 = 1$ , donde  $\mathbf{m}_0$  es una solución inicial.
3. Para cada iteración  $k = 0, 1, 2, 3, \dots$ :

a)

$$\mathbf{m}_{k+1} = T_{\mu/2\eta} \left\{ \mathbf{z}_k + \frac{1}{\eta} \mathbf{A}^T (\mathbf{d} - \mathbf{A} \mathbf{z}_k) \right\}, \quad (\text{B.26})$$

donde  $T_v\{\cdot\}$  es una función umbral suave (*soft-thresholding function*), que se aplica a cada elemento de su argumento vectorial y está definida como

$$T_v\{g\} = \begin{cases} g(1 - v/|g|), & |g| \geq v \\ 0, & |g| < v \end{cases} \quad (\text{B.27})$$

b)

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}. \quad (\text{B.28})$$

c)

$$\mathbf{z}_{k+1} = \mathbf{m}_{k+1} + \frac{t_k - 1}{t_{k+1}} (\mathbf{m}_{k+1} - \mathbf{m}_k). \quad (\text{B.29})$$

d)

$$k = k + 1, \quad (\text{B.30})$$

e) Verificar convergencia o criterio de corte.

La principal diferencia entre ISTA y FISTA es que en la función umbral  $T_v\{\cdot\}$  no se utiliza el punto  $\mathbf{m}_k$ , sino una combinación lineal muy específica de los puntos  $\mathbf{m}_{k+1}$  y  $\mathbf{m}_k$ . Con esta mejora se consigue una tasa de convergencia mejor que la del ISTA con un incremento despreciable del costo computacional, ya que el mayor costo reside en calcular  $\mathbf{m}_{k+1}$  y no en calcular un nuevo punto a partir de  $\mathbf{m}_k$  y  $\mathbf{m}_{k+1}$ .

# Bibliografía

- Aki, K. y Richards, P., 1980. *Quantitative seismology: theory and methods*, W.H. Freeman and Co.
- Alemie, W. M., 2010. *Regularization of the AVO inverse problem by means of a multivariate Cauchy probability distribution*, Master's thesis, The University of Alberta.
- Alemie, W. M. y Sacchi, M. D., 2011. High-resolution three-term AVO inversion by means of a trivariate Cauchy probability distribution, *Geophysics*, **76**(3), R43–R55.
- Aster, R., Borchers, B., y Thurber, C., 2013. *Parameter Estimation and Inverse Problems*, Academic Press.
- Babacan, D., Molina, R., y Katsaggelos, A., 2010. Bayesian compressive sensing using laplace priors, *IEEE Transactions on Image Processing*, **19**(1), 53–63.
- Babacan, S. D., Nakajima, S., y Do, M. N., 2014. Bayesian group-sparse modeling and variational inference, *IEEE Transactions on Signal Processing*, **62**(11), 2906–2921.
- Bauschke, H., Burachik, R., Combettes, P., Elser, V., Luke, D., y Wolkowicz, H., 2011. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer New York.
- Bayes, T., 1763. An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S., *Philosophical Transactions of the Royal Society*, **53**, 370–418.
- Bayram, I., Chen, P., y Selesnick, I. W., 2014. Fused lasso with a non-convex sparsity inducing penalty, en *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2014, Florence, Italy, May 4-9, 2014*, pp. 4156–4160.
- Beck, A. y Teboulle, M., 2009. A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal of Imaging Sciences*, **2**, 183–202.
- Bortfeld, R., 1961. Approximations to the reflection and transmission coefficients of plane longitudinal and transverse waves, *Geophysical Prospecting*, **9**(4), 485–502.
- Bredies, K. y Lorenz, D. A., 2008. Linear convergence of iterative soft-thresholding, *Journal of Fourier Analysis and Applications*, **14**, 813–837.

## BIBLIOGRAFÍA

---

- Buland, A. y Omre, H., 2003. Bayesian linearized avo inversion, *Geophysics*, **68**(1), 185–198.
- Cambois, G., 1998a. *AVO attributes and noise: Pitfalls of crossplotting*, chap. 64, pp. 244–247, SEG.
- Cambois, G., 1998b. AVO attributes and noise: pitfalls of crossplotting, en *Expanded Abstracts*, vol. 17, pp. 244–247, SEG.
- Candes, E. y Tao, T., 2007. The dantzig selector: Statistical estimation when p is much larger than n, *Ann. Statist.*, **35**(6), 2313–2351.
- Castagna, J. P. y Smith, S. W., 1994. Comparison of AVO indicators: A modeling study, *Geophysics*, **59**(12), 1849–1855.
- Castagna, J. P., Batzle, M. L., y Eastwood, R. L., 1985. Relationships between compressional wave and shear wave velocities in clastic silicate rocks, *Geophysics*, **50**(4), 571–581.
- Castagna, J. P., Swan, H. W., y Foster, D. J., 1998. Framework for AVO gradient and intercept interpretation, *Geophysics*, **63**(3), 948–956.
- Castle, R., 1994. A theory of normal moveout, *Geophysics*, **59**(6), 983–999.
- Chopra, S. y Castagna, J., 2007. Introduction to this special section – AVO, *The Leading Edge*, **26**, 1506–1507.
- Chung, H. y Lawton, D., 1995. Frequency characteristics of seismic reflections from thin beds, *Canadian Journal of Exploration Geophysicists*, **31**, 32–37.
- Cooke, D. y Schneider, W., 1983. Generalized linear inversion of reflection seismic data, *Geophysics*, **48**(6), 665–676.
- Cvijović, D. y Klinowski, J., 1995. Taboo search: An approach to the multiple minima problem, *Science*, **267**, 664–666.
- Dahl, T. y Ursin, B., 1991. Parameter estimation in a one-dimensional anelastic medium, *Journal of Geophysical Research: Solid Earth*, **96**(B12), 20217–20233.
- Daubechies, I., Defrise, M., y Mol, C. D., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Communications on Pure and Applied Mathematics*, **57**, 1413–1457.
- ed. Davis, L., 1987. *Genetic algorithms and simulated annealing*, Morgan Kaufmann Publishers, Los Altos, CA.
- Debye, H. W. J. y van Riel, P., 1990.  $L_p$ -norm deconvolution, *Geophysical Prospecting*, **38**, 381–403.
- Downton, J., Lines, L., Goodway, B., Xu, Y., y Li, Y., 2000. Predicting the statistical properties of the fluid stack, en *Expanded Abstracts*, vol. 19, pp. 224–227, SEG.

- Downton, J. E., 2005. *Seismic parameter estimation from AVO inversion*, Ph.D. thesis, Department of geology and geophysics, University of Calgary, Canada.
- Downton, J. E. y Lines, L. R., 2001. Constrained three parameter avo inversion and uncertainty analysis, *SEG Technical Program Expanded Abstracts*, **20**(1), 251–254.
- Downton, J. E. y Lines, L. R., 2003. High-resolution AVO analysis before NMO, en *Expanded Abstracts*, vol. 22, pp. 219–222, SEG.
- Downton, J. E. y Lines, L. R., 2004. Three term avo waveform inversion, *SEG Technical Program Expanded Abstracts*, **23**(1), 215–218.
- Eidsvik, J. y Theune, U., 2009. Blocky inversion of timelapse seismic avo data, *SEG Technical Program Expanded Abstracts*, **28**(1), 3820–3824.
- Farquharson, C. G. y Oldenburg, D. W., 2004. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems, *Geophysical Journal International*, **156**, 411–425.
- Fatti, J. L., Smith, G. C., Vail, P. J., Strauss, P. J., y Levitt, P. R., 1994. Detection of gas in sandstone reservoirs using AVO analysis: A 3-D seismic case history using the Geostack technique, *Geophysics*, **59**(9), 1362–1376.
- Figueiredo, M., 2003. Adaptive sparseness for supervised learning, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **25**(9), 1150–1159.
- Figueiredo, M. A. T., Nowak, R. D., y Wright, S. J., 2007. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, *IEEE journal of selected topics in signal processing*, **1**, 586–597.
- Fogel, D., 1993. On the philosophical differences between genetic algorithms and evolutionary algorithms, en *Proceedings*, pp. 23–29, Evolutionary Programming Society, La Jolla, CA.
- Fogel, D., 1994. An introduction to evolutionary optimization, *IEEE Transactions on Neural Networks*, **5**, 3–14.
- Fornasier, M. y Rauhut, H., 2008. Recovery algorithms for vector-valued data with joint sparsity constraints, *SIAM Journal on Numerical Analysis*, **46**(2), 577–613.
- Gardner, G., Gardner, L., y Gregory, A., 1974. Formation velocity and density - the diagnostic basics for stratigraphic traps, *Geophysics*, **39**(6), 770–780.
- Geman, S. y Geman, D., 1984. Stochastic relaxation, Gibbs' distributions, and the Bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **PAMI-6**, 721–741.
- Gramfort, A., Kowalski, M., y Hämmäläinen, M., 2012. Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods, *Physics in Medicine and Biology*, **57**(7), 1937–1961.

## BIBLIOGRAFÍA

---

- Gramfort, A., Strohmeier, D., Haueisen, J., Härmäläinen, M., y Kowalski, M., 2013. Time-Frequency Mixed-Norm Estimates: Sparse M/EEG imaging with non-stationary source activations, *NeuroImage*, **70**, 410–422.
- Groß, J., 2003. *Linear regression*, Lecture notes in statistics, Springer, Berlin, Heidelberg, Paris.
- Hadamard, J., 1902. Sur les problèmes aux dérivées partielles et leur signification physique, *Princeton University Bulletin*, **13**, 49–52.
- Hampson, D. y Russell, B., 2013. Joint simultaneous inversion of pp and ps angle gathers, *CSEG Recorder*, **38**, 32–39.
- Hampson, D. P., Russell, B. H., y Bankhead, B., 2005. *Simultaneous inversion of pre-stack seismic data*, chap. 417, pp. 1633–1637, SEG.
- Hennenfent, G., van den Berg, E., Friedlander, M. P., y Hermann, F. J., 2008. New insights into one-norm solvers from the pareto curve, *Geophysics*, **73**(4), 23–26.
- Herrmann, P. y Cambois, G., 2001. Statistical properties of seismically derived AVO attributes, en *Expanded Abstracts*, vol. 20, pp. 231–234, SEG.
- Hoffman, K. y Kunze, R., 1973. *Álgebra lineal*, Prentice-Hall Internacional.
- Ikelle, L. T. y Amundsen, L., 2005. *Introduction to petroleum seismology*, Investigations in Geophysics, Society of Exploration Geophysicists.
- Ingber, L., 1989. Very fast simulated re-annealing, *Journal of Mathematical Computation and Modelling*, **12**, 967–973.
- Ingber, L. y Rosen, B., 1992. Genetic algorithms and very fast simulated reannealing: a comparison, *Journal of Mathematical Computation and Modelling*, **16**, 87–100.
- Jensås, I. Ø., 2008. *A blockyness Constraint for seismic AVA Inversion*, Master's thesis, Norwegian University of Science and Technology Department of Mathematical Sciences.
- Jin, S., Cambois, G., y Vuillermoz, C., 2000. Shear-wave velocity and density estimation from PS-wave AVO analysis: Application to an OBS dataset from the North Sea, *Geophysics*, **65**(5), 1446–1454.
- Kabanikhin, S. I., 2008. Definitions and examples of inverse and ill-posed problems, *Journal of Inverse and Ill-posed Problems*, **16**(4), 317–357.
- Khajehnejad, M. A., Xu, W., Avestimehr, A. S., y Hassibi, B., 2011. Analyzing weighted minimization for sparse recovery with nonuniform sparse models, *IEEE Transactions on Signal Processing*, **59**(5), 1985–2001.
- Kirkpatrick, S., Gelatt, C. D., y Vecchi, M. P., 1983. Optimization by simulated annealing, *Science*, **220**(4598), 671–680.

- Knott, C., 1899. Reflection and refraction of elastic waves with seismological applications, *Philosophical Magazine*, **48**, 64–97.
- Koefoed, O., 1955. On the effect of poisson's ratios of rock strata on the reflection coefficients of plane waves, *Geophysical Prospecting*, **3**(4), 381–387.
- Koefoed, O., 1962. Reflection and transmission coefficients for plane longitudinal incident waves, *Geophysical Prospecting*, **10**(3), 304–351.
- Kowalski, M., 2009. Sparse regression using mixed norms., *Appl. Comput. Harmon. Anal.*, **27**(3), 303–324.
- Kowalski, M., 2014. Thresholding rules and iterative shrinkage/thresholding algorithm: a convergence study, en *International Conference on Image Processing (ICIP) 2014*, La Défense, Paris, France.
- Kowalski, M. y Torrèsani, B., 2009. Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients, *Signal, Image and Video Processing*, **3**(3), 251–264.
- Kowalski, M., Siedenburg, K., y Dörfler, M., 2013. Social sparsity! Neighborhood systems enrich structured shrinkage operators, *IEEE Transactions on Signal Processing*, **61**(10), 2498–2511.
- Kreimer, N., 2013. *Multidimensional seismic data reconstruction using tensor analysis*, Ph.D. thesis, The University of Alberta.
- Landweber, L., 1951. *An Iteration Formula for Fredholm Integral Equations of the First Kind with Application to the Axially Symmetric Potential Flow about Elongated Bodies of Revolution*, University of Maryland, College Park.
- Larson, R. y Edwards, B. H., 1999. *Elementary Linear Algebra*, Houghton Mifflin Company, 4th edn.
- Lay, T. y Wallace, T., 1995. *Modern Global Seismology*, International Geophysics, Elsevier Science.
- Levy, S. y Oldenburg, D., 1982. The deconvolution of phase-shifted wavelets, *Geophysics*, **47**(9), 1285–1294.
- Ma, X., 2002. Simultaneous inversion of prestack seismic data for rock properties using simulated annealing, *Geophysics*, **67**(6), 1877–1885.
- Mallick, S., 1995. Model-based inversion of amplitude-variations-with-offset data using a genetic algorithm, *Geophysics*, **60**(4), 939–954.
- Martin, G. S., Wiley, R., y Marfurt, K. J., 2006. Marmousi2: an elastic upgrade for Marmousi, *The Leading Edge*, **25**, 156–166.
- Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, International Geophysics, Elsevier Science.

## BIBLIOGRAFÍA

---

- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., y Teller, E., 1953. Equation of state calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087–1092.
- Misra, S. y Sacchi, M. D., 2007. Nonlinear one-dimensional prestack seismic inversion with edge-reserving smoothing filter, en *Extended Abstracts*, Eur. Assn. Geosci. Eng., London, United Kingdom.
- Misra, S. y Sacchi, M. D., 2008. Global optimization with model-space preconditioning: Application to AVO inversion, *Geophysics*, **73**(5), R71–R82.
- Nesterov, Y., 1983. A method for solving a convex programming problem with convergence rate  $o(1/k^2)$ , *Soviet Mathematics Doklady*, **27**(2), 372–376.
- Nowak, E., Swan, H., y Lane, D., 2008. Quantitative thickness estimates from the spectral response of avo measurements, *Geophysics*, **73**(1), C1–C6.
- Oldenburg, D. W., Scheuer, T., y Levy, S., 1983. Recovery of the acoustic impedance from reflection seismograms, *Geophysics*, **48**(10), 1318–1337.
- Ostrander, W., 1984. Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence, *Geophysics*, **49**(10), 1637–1648.
- Park, T. y Casella, G., 2008. The bayesian lasso, *Journal of the American Statistical Association*, **103**(482), 681–686.
- Pérez, D. O. y Velis, D. R., 2010. Inversión sparse-spike de datos sísmicos prestack, pp. 164–169, Asociación Argentina de Geofísicos y Geodestas, Cordoba, Argentina.
- Pérez, D. O. y Velis, D. R., 2011a. Sparse-spike AVO/AVA attributes from prestack data, en *SEG Technical Program Expanded Abstracts 2011*, vol. 30, pp. 340–344, SEG.
- Pérez, D. O. y Velis, D. R., 2011b. Atributos AVO/AVA de alta resolución a partir de datos prestack, en *Integración: acercando la ondícula al trépano*, vol. 1, pp. 145–154, Instituto Argentino del Petróleo y del Gas.
- Pérez, D. O. y Velis, D. R., 2012. Atributos AVO/AVA de alta resolución utilizando Very Fast Simulated Annealing, *Geoacta*, **37**(1), 19–36.
- Pérez, D. O., Velis, D. R., y Sacchi, M. D., 2012. Inversion of prestack seismic data using fista, en *Mecánica Computacional Vol. XXXI*, pp. 3255–3263, Asociación Argentina de Mecánica Computacional.
- Pérez, D. O., Velis, D. R., y Sacchi, M. D., 2013a. High-resolution prestack seismic inversion using a hybrid FISTA least-squares strategy, *Geophysics*, **78**(5), R185–R195.
- Pérez, D. O., Velis, D. R., y Sacchi, M. D., 2013b. Estimating sparse-spike attributes from ava data using a fast iterative shrinkage-thresholding algorithm and least squares, en *SEG Technical Program Expanded Abstracts 2013*, pp. 3062–3067.

- Pérez, D. O., Velis, D. R., y Sacchi, M. D., 2013c. A hybrid strategy based on fast iterative shrinkage-thresholding algorithm and very fast simulated annealing: Application to the prestack seismic inverse problem, en *Mecánica Computacional Vol. XXXII*, pp. 1701–1713, Asociación Argentina de Mecánica Computacional.
- Pérez, D. O., Velis, D. R., y Sacchi, M. D., 2014a. Blocky inversion of prestack seismic data using mixed-norms, en *SEG Technical Program Expanded Abstracts 2014*, pp. 3106–3111.
- Pérez, D. O., Velis, D. R., y Sacchi, M. D., 2014b. The use of mixed norms to constrain the inversion of prestack seismic data, Asociación Argentina de Geofísicos y Geodestas, San Juan, Argentina.
- Porsani, M. J. y Ursin, B., 2000. Mixed-phase deconvolution and wavelet estimation, *The Leading Edge*, **19**(1), 76–79.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., y Flannery, B. P., 1996. *Numerical Recipes in Fortran 90 (2Nd Ed.): The Art of Parallel Scientific Computing*, Cambridge University Press, New York, NY, USA.
- Ramanujan, J. y Sadayappan, P., 1988. Optimization by neural networks, en *Proc.*, vol. II, pp. 325–332.
- Rastrigin, L. A., 1974. *Sistemy ekstremal'nogo upravleniya*, Nauka Moskva.
- Richards, P. y Frasier, C., 1976. Scattering of elastic waves from depth dependent inhomogeneities, *Geophysics*, **41**(3), 441–458.
- Ricker, N., 1940. The form and nature of seismic waves and the structure of seismograms, *Geophysics*, **5**, 348–366.
- Robinson, E., 1954. *Predictive decomposition of time series with application to seismic exploration*, Ph.D. thesis, Massachusetts Institute of Technology, Massachusetts, United States of America.
- Robinson, E., 1967. Predictive decomposition of time series with application to seismic exploration, *Geophysics*, **32**(3), 418–484.
- eds Robinson, E. A. y Ebro, D. A., 1996. *Deconvolution 2*, no. 17 en Geophysics Reprint Series, SEG.
- Robinson, E. A. y Treitel, S., 2002. *Geophysical Signal Analysis*, SEG.
- Routh, P. S., Anno, P. D., Baumel, R. T., y Chavarria, J. A., 2003. *Inversion for source wavelet and AVA parameters from prestack seismic data*, chap. 179, pp. 698–701, SEG.
- Rubino, J. y Velis, D., 2009. Thin-bed prestack spectral inversion, *Geophysics*, **74**(4), R49–R57.
- Rüger, A., 2002. *Reflection Coefficients and Azimuthal AVO Analysis in Anisotropic Media*, Geophysical monograph series, Society of Exploration Geophysicists.

## BIBLIOGRAFÍA

---

- Rüger, A. y Gray, D., 2014. *Wide-azimuth amplitude-variation-with-offset analysis of anisotropic fractured reservoirs*, chap. 1, pp. N1–1–N1–14, SEG.
- Russell, B., 2014. Prestack seismic amplitude analysis: An integrated overview, *Interpretation*, **2**(2), SC19–SC36.
- Sacchi, M. D., 1997. Reweighting strategies in seismic deconvolution, *Geophysical Journal International*, **129**(3), 651–656.
- Sacchi, M. D., Velis, D. R., y Ulrych, T. J., 1996. Wavelet via polycepstra, en *Expanded Abstracts*, pp. 1583–1586, SEG, Denver, USA.
- Sacchi, M. D., Ulrych, T. J., y Walker, C. J., 1998. Interpolation and extrapolation using a high-resolution discrete fourier transform, *IEEE Transaction on Signal Processing*, pp. 31–38.
- Scales, J. y Tenorio, L., 2001. Prior information and uncertainty in inverse problems, *Geophysics*, **66**(2), 389–397.
- Scales, L. E., 1985. *Introduction to Non-Linear Optimization*, Macmillan computer science series, Springer-Verlag.
- Selesnick, I. W. y Bayram, I., 2013. Sparse signal estimation by maximally sparse convex optimization, *CoRR*, **abs/1302.5729**.
- Shaw, R. y Srivastava, S., 2007. Particle swarm optimization: A new tool to invert geophysical data, *Geophysics*, **72**(2), F75–F83.
- Shearer, P., 1999. *Introduction to Seismology*, Cambridge University Press.
- Sheriff, R., 2002. *Encyclopedic Dictionary of Exploration Geophysics*, Geophysical Reference Series No. 13, SEG, 4th edn.
- Shuey, R., 1985. A simplification of the Zoeppritz equations, *Geophysics*, **50**(4), 609–614.
- Simm, R. y Bacon, M., 2014. *Seismic Amplitude: An Interpreter's Handbook*, Cambridge University Press.
- Simmons, J. L. y Backus, M. M., 1996. Waveform-based avo inversion and avo prediction-error, *Geophysics*, **61**(6), 1575–1588.
- Sivia, D. S., 1996. *Data Analysis: A Bayesian Tutorial*, Clarendon (Oxford Univ. Press), 2nd edn.
- Smith, G. C. y Gidlow, M., 1987. Weighted stacking for rock property estimation and detection of gas, *Geophysical Prospecting*, **35**(9), 993–1014.
- Smith, G. C. y Gidlow, M., 2000. A comparison of the fluid factor with  $\lambda$  and  $\mu$  in AVO analysis, en *Expanded Abstracts*, pp. 122–125, Soc. of Expl. Geophys.

- Snieder, R. y Trampert, J., 1999. Inverse problems in geophysics, en *Wavefield Inversion*, pp. 119–190, ed. Wirgin, A., Springer Verlag.
- Szu, H. y Hartley, R., 1987. Fast simulated annealing, *Phys. Lett. A.*, **122**, 157–162.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Available at the Web.
- Taylor, H. L., Banks, S. C., y McCoy, J. F., 1979. Deconvolution with the  $l_1$  norm, *Geophysics*, **44**(1), 39–52.
- Theune, U., Jensås, I. Ø., y Eidsvik, J., 2010. Analysis of prior models for a blocky inversion of seismic avo data, *Geophysics*, **75**(3), C25–C35.
- Thomsen, L., 1986. Weak elastic anisotropy, *Geophysics*, **51**(10), 1954–1966.
- Tibshirani, R., 1996. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, **58**, 267–288.
- Ulrych, T., Sacchi, M., y Woodbury, A., 2001. A bayes tour of inversion: A tutorial, *Geophysics*, **66**(1), 55–69.
- Ulrych, T. J. y Sacchi, M. D., 2005. *Information-based inversion and processing with applications*, Elsevier.
- Ulrych, T. J., Velis, D. R., y Sacchi, M., 1995. Wavelet estimation revisited, *The Leading Edge*, **10**, 1139–1143.
- van den Berg, E. y Friedlander, M., 2008. Probing the pareto frontier for basis pursuit solutions, *SIAM Journal of Scientific Computing*, **31**(2), 890–912.
- Velis, D. R., 1998. *Application of simulated annealing to some seismic problems*, Ph.D. thesis, University of British Columbia, Vancouver, Canada.
- Velis, D. R., 2003. Estimating the distribution of primary reflection coefficients, *Geophysics*, **68**(4), 1417–1422.
- Velis, D. R., 2008. Stochastic sparse-spike deconvolution, *Geophysics*, **73**, R1–R9.
- Velis, D. R. y Ulrych, T. J., 1996. Simulated annealing wavelet estimation via fourth-order cumulant matching, *Geophysics*, **61**, 1939–1948.
- Verm, R. y Hilterman, F., 1995. Lithology color coded seismic sections: The calibration of avo crossplotting to rock properties, *The Leading Edge*, **14**(8), 847–853.
- Walden, A. y Hosken, J., 1986. The nature of the non-Gaussianity of primary reflection coefficients and its significance for deconvolution, *Geophysical Prospecting*, **34**, 1038–1066.
- Walden, A. T., 1991. Making avo sections more robust, *Geophysical Prospecting*, **39**(7), 915–942.

## BIBLIOGRAFÍA

---

- Whitcombe, D. N., Dyce, M., McKenzie, C. J. S., y Hoerber, H., 2004. Stabilizing the AVO gradient, en *Expanded Abstracts*, vol. 23, pp. 232–235, SEG.
- White, S., 1984. Concepts of scale in simulated annealing, *Proceedings of IEEE*, pp. 646–651.
- Wiggins, R., 1978. Minimum entropy deconvolution, *Geoexploration*, **16**, 21–35.
- Yilmaz, O., 2001. *Seismic Data Analysis: processing, inversion, and interpretation of seismic data*, Investigations in Geophysics, Society of Exploration Geophysicists.
- Youzwishen, C. F., 2001. *Non-linear sparse and blocky constraints for seismic inverse problems*, Master's thesis, The University of Alberta.
- Yuan, M., Yuan, M., Lin, Y., y Lin, Y., 2006. Model selection and estimation in regression with grouped variables, *Journal of the Royal Statistical Society, Series B*, **68**, 49–67.
- Zhang, R., Sen, M., y Srinivasan, S., 2013. A prestack basis pursuit seismic inversion, *Geophysics*, **78**(1), R1–R11.
- Zoeppritz, K., 1919. Über Reflexion and Durchgang seismischer Wellen durch Unstetigkeits-flächen, *Gott. Nachr. Math. Phys*, **K1**, 66–84.