

University of Alberta

Near real-time estimation of the seismic source parameters in a compressed domain

by

Ismael A. Vera Rodriguez

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Geophysics

Department of Physics

©Ismael A. Vera Rodriguez
Fall 2012
Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

To my mother

To my wife

Abstract

Seismic events can be characterized by its origin time, location and moment tensor. Fast estimations of these source parameters are important in areas of geophysics like earthquake seismology, and the monitoring of seismic activity produced by volcanoes, mining operations and hydraulic injections in geothermal and oil and gas reservoirs. Most available monitoring systems estimate the source parameters in a sequential procedure: first determining origin time and location (e.g., epicentre, hypocentre or centroid of the stress glut density), and then using this information to initialize the evaluation of the moment tensor. A more efficient estimation of the source parameters requires a concurrent evaluation of the three variables.

The main objective of the present thesis is to address the simultaneous estimation of origin time, location and moment tensor of seismic events. The proposed method displays the benefits of being: 1) automatic, 2) continuous and, depending on the scale of application, 3) of providing results in real-time or near real-time. The inversion algorithm is based on theoretical results from sparse representation theory and compressive sensing. The feasibility of implementation is determined through the analysis of synthetic and real data examples. The numerical experiments focus on the microseismic monitoring of hydraulic fractures in oil and gas wells, however, an example using real earthquake data is also presented for validation. The thesis is complemented with a resolvability analysis of the moment tensor. The analysis targets common monitoring geometries employed in hydraulic fracturing in oil wells. Additionally, it is presented an application of sparse representation theory for the denoising of one-component and three-component microseismicity records, and an algorithm for improved automatic time-picking using non-linear inversion constraints.

Acknowledgements

I am very grateful to Dr. Mauricio Sacchi for allowing me to join the Signal Analysis and Imaging Group, and for his guidance and support during my studies.

I would also like to acknowledge Dr. Jeffrey Gu for his support and help. All I learned from his Global Seismology class was particularly helpful for the development and put in practice of the ideas presented in this thesis.

I am grateful to Mr. David Bonar for sharing with me his MSc. work, which resulted in the development of the denoising method presented in Chapter 6 of this thesis.

Thanks to the Department of Physics, particularly to all those people who make easier the lives of grad students like me while we are here.

Finally, thanks to my colleagues and friends from the Signal Analysis and Imaging Group, the Department of Physics and the University of Alberta for the great time that I had during my studies. It has been a wonderful journey.

Contents

1	Introduction	1
1.1	Background	1
1.2	Contributions	5
1.3	Organization of the thesis	5
2	Theory	8
2.1	Quantitative description of the seismic source	8
2.1.1	The seismic moment tensor	10
2.1.2	The Green's function of an elemental force couple	13
2.2	Hydraulic fracturing	14
2.2.1	Microseismic monitoring	16
2.3	Compressive Sensing	19
3	Resolution analysis	25
3.1	Introduction	25
3.2	Parametrization	26
3.3	Resolution under single vertical monitoring well geometries	27
3.3.1	Resolution matrix	27
3.3.2	Maximum number of resolvable elements	29
3.3.3	Inversion stability and acquisition design	31
3.4	Minimum requirements for a full moment tensor inversion	33

3.4.1	Stability for non-vertical monitoring geometries	33
3.4.2	Sensitivity to noise and velocity perturbation	34
3.4.3	Full moment tensor inversion from a single deviated monitoring well	37
3.5	Summary	38
4	Simultaneous estimation of the source parameters	41
4.1	Introduction	41
4.2	Theoretical background	42
4.2.1	Block sparsity	42
4.2.2	Recovery of block K -sparse solutions	44
4.2.3	Displacement field due to a seismic source	45
4.3	Using sparse representation theory to invert for the source parameters	46
4.3.1	Algorithm to invert for the source parameters based on a block sparse representation	48
4.3.2	Proposed modification to the BOMP algorithm	51
4.3.3	Stopping criteria	53
4.4	Application to synthetic and real data examples	55
4.4.1	Hydraulic fracturing in an oil well	55
4.4.2	Application to the 18 June, 2002 Caborn Indiana earthquake	60
4.4.3	Comparison with other real-time inversion methodologies	67
4.5	Summary	69
5	Compressed domain inversion	71
5.1	Introduction	71
5.2	Compressed domain inversion	72
5.3	Numerical Modeling	74
5.3.1	Seismic monitoring with dense receiver coverage	74
5.3.2	Simultaneous events	75
5.3.3	Earthquake monitoring	78
5.4	Discussion	81
5.5	Summary	84

6	Microseismic data denoising	86
6.1	Introduction	86
6.2	One-component (1C) sparse time-frequency transform	87
6.3	Extension to the three-component (3C) case via group sparsity	89
6.4	Application to denoise microseismic data	92
6.5	Summary	97
7	Improvements in automatic time-picking	104
7.1	Introduction	104
7.2	Methodology	105
7.3	Examples	107
7.4	Summary	110
8	Conclusions	112
	Bibliography	116
	Appendices	
A	FISTA for group sparse solutions	128

List of Tables

4.1	Velocity model used in the Caborn Indiana earthquake inversion (Kim, 2003).	62
4.2	List of seismic stations used in the hypothetical monitoring system.	66
4.3	Decomposition in percentages of isotropic (ISO), double couple (DC) and compensated linear vector dipole (CLVD) of source mechanism solutions for the 18 June, 2002 Caborn earthquake. Negative percentages in ISO and CLVD components are associated to compressive faults, α is the angle formed between the fault plane and the slip vector (Vavrycuk, 2001).	67
5.1	Comparison of different solutions for the 18 June, Caborn earthquake. NNA stands for number of non-adaptive measurements per non-zero coefficient in the CS solution. DS stands for dictionary size after compression, where 100% corresponds to the uncompressed dictionary. Success rate refers to the percentage of times that the event was detected using the CS approach in 500 realizations of sensing matrix.	81
5.2	Comparison of the general steps involved in the estimation of the source parameters between the two-step Vs. the simultaneous approach. Advantages and disadvantages are considered in terms of the speed to generate results and their reliability.	85

List of Figures

2.1	Propagation of seismic energy in an elastic medium and its recording in surface stations. (Image created by Cecilia Camet)	9
2.2	The nine elemental force couples that form the seismic moment tensor. Modified after Aki and Richards (2009)	11
2.3	Point force approximation of a finite fault. a) Real fault with finite extension. b) Point force approximation. c) Principal axes for the point force approximation. Picture in a) modified from USGS website on San Andreas Fault. Original photograph taken by Robert E. Wallace.	11
2.4	Radiation pattern of P (light colors) and S (dark colors) waves for each of the elemental force couples in the seismic moment tensor.	14
2.5	Schematic deployment of a hydraulic fracturing job. (Image created by Cecilia Camet)	15
2.6	Summary of methodologies to retrieve information about properties of hydraulic fractures. Green: can determine. Yellow: may determine. Red: cannot determine. After Cipolla and Wright (2002).	17
2.7	Schematic representation of a microseismic monitoring job.	18
2.8	Relationship between fracture complexity index (FCI), fracture distribution (grey lines) and induced microseismic events (red dots). a) Scenario with low FCI. b) Scenario with high FCI.	19
2.9	Example of image compression using the JPEG method. When displayed on the screen, each of the four images occupies a memory of 10.2 MB with a size of 1030×1044 pixels. For storage purposes the JPEG method expresses each image as a linear combination of the elements of a dictionary of cosine functions (Discrete Cosine Transform). By thresholding the coefficients of smaller magnitude different levels of storage compression are achieved. When saved in hard disk, the JPEG file only requires the location and magnitude of the coefficients that expand the image. In this example, the displayed image has been recovered from a version in hard disk of (a) 1.3 MB, (b) 620 KB, (c) 350 KB and (d) 136 KB. Although there is not perceptual loss of information from one version to another, the fully sampled image has to be known first in order to compute its Discrete Cosine Transform. Once the desired coefficients are known, the original image is thrown away.	20

2.10	Compression of the set of six Green's functions for the three-components of eight seismic stations using the sensing matrix Φ . In this thesis, the seismic traces are first sampled/modeled at Nyquist rate and then compressed. Compressive sensing technology aims at the direct acquisition of the compressed measurements on the right. The red lines denote the start and end of the samples for each seismic station. This separation cannot be done in the compressed measurements.	23
2.11	Example of a matching pursuit methodology combined with group sparsity constraints in the setting of source monitoring. Every panel represents a metric related to the source strength of virtual sources located at each of the nodes of a grid. In the initial solution the virtual source with largest strength is identified and reclassified as an actual source, the seismic wavefield due to this source is forward modeled and extracted from the original observations, then the source strength metric is estimated again to obtain the next panel (It 1). Following this procedure iteratively three virtual sources are reclassified as actual sources after the third panel (It 2). The last two panels display only residuals (It 3 and It 4). After Vera Rodriguez et al. (2010a)	24
3.1	Source-receivers geometry for the microseismic monitoring experiment with one vertical monitoring well. The geographical and aligned reference systems have axes N-E-V and x' - y' -V, respectively. The axes x' and V are contained in the observational plane. The angle θ denotes the azimuth of the array of receivers (receiver azimuth) from North.	29
3.2	Resolution matrix for different combinations of source-receivers horizontal distance (d) and azimuth of the line of receivers (θ). Each diagonal box (delimited by white lines) represents the resolvability of the seismic moment tensor elements in the order $M_{11}, M_{12}, M_{13}, M_{22}, M_{23}$ and M_{33} (equivalently, m_1, m_2, m_3, m_4, m_5 and m_6). The scales showed on box R_{11} apply for any other box of the matrix. Colors close to one in a diagonal box mean that the corresponding element can be solved in the inversion. Colors different from zero in off-diagonal boxes (R_{ij}) denote the linear dependency between the corresponding elements of the seismic moment tensor (m_i and m_j).	30
3.3	Condition number (C_ϵ) of the matrix \mathbf{H} for different angular apertures between receivers with respect to the source location (α_1).	31
3.4	Normalized mean squared error ($NMSE$) function of the inversion for different combinations of aperture (α_1) and noise level.	33
3.5	Normalized mean squared error ($NMSE$) function of the inversion for different combinations of aperture (α_1) and noise level. The velocity model is biased by an increment of 5% in both P- and S-wave speeds.	34

3.6	Source-receiver geometry arrangements to compute the condition number for the full source mechanism inversion. Receivers R_1 , R_2 and R_3 are contained in the plane A, which lies perpendicular to the line that joins the source location with R_1 . The angle α_1 denotes the aperture between receivers from the source location and α_2 is the aperture between receivers R_2 and R_3 with respect to R_1 . a) source location and receiver R_1 are aligned with an axis of the reference system and the three receivers form a vertical plane. b) source location and receiver R_1 are not aligned with any of the axis of the reference system and the three receivers form a vertical plane. c) source location and receiver R_1 are not aligned with the reference system and the three receivers form a non-vertical plane. d) source location and receiver R_1 are not aligned with the reference system but receivers R_1 and R_2 are aligned in the vertical direction.	35
3.7	Condition number for different combinations of the angles α_1 and α_2 . The four graphs correspond to the source-receiver geometries presented in Figure 3.6. Condition numbers above 600 are set at 600 for better visualization. . . .	36
3.8	Normalized mean squared error (<i>NMSE</i>) function for all combinations of α_1 and α_2 with condition number < 100 in Figure 3.7-b. The percent value represents noise level.	37
3.9	Normalized mean squared error (<i>NMSE</i>) function for all combinations of α_1 and α_2 with condition number < 100 in Figure 3.7-b. The percent value represents noise level. In this case, the P- and S-wave speeds are biased by 5%. 38	38
3.10	Example of a receivers configuration that can solve the full seismic moment tensor inversion for the given source projection (white dot). Distances are measured over the plain that contains the receivers. The source location is at 400 m in the perpendicular direction to this view. The grey line represents a smooth trajectory joining the three receivers.	39
3.11	Receiver distribution (blue triangles) along an ideal deviated well (black line) that can solve a full moment tensor inversion. An isosurface value of 80 (condition number) is selected for this plot.	40
4.1	Mapping between the subsurface grid volume and the dictionary of Green's functions. Every node in the grid has spatial coordinates (x, y, z) and a linear index $l \in [1, N_t]$ where N_t is the total of nodes in the grid. Hence, the spatial coordinates of the l -th node are (x_l, y_l, z_l) . A set of 6 Green's functions corresponding to each of the independent elements of the seismic moment tensor are computed at each grid node. Every set of Green's functions form a block of the dictionary of Green's functions \mathbf{G} in equation (4.14). Correspondingly, every block in the solution vector \mathbf{m} is associated to a location in the subsurface grid volume. Hence, the blocks in \mathbf{m} with ℓ_2 -norm different from zero provide the location and moment tensor coefficients where sources in the grid have taken place. An extension in time of the dictionary provides also the origin times of the sources.	47

4.2	Real data example of the modification introduced to the BOMP algorithm. a) Function $f(k, l) = \ \mathbf{G}_k^T[l]\mathbf{r}_0\ _2^2$, the total number of nodes in the grid of this example is $N_t = 2205$, hence $l \in [1, 2205]$, the processing window is $N_w = 30$, hence $k \in [0, 29]$. b) Plot of the n_h largest values in a), in this example $n_h = 100$, therefore $p \in [1, 100]$. c) Normalized misfit for solutions computed at each (k_p, l_p) indexes that correspond to values of $f(k_p, l_p)$ displayed in b). The delay index k_p and position index l_p where the normalized misfit is minimum are considered the optimal origin time and location for a seismic source, and therefore saved in the index sets $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$. In the regular BOMP workflow the indexes (k, l) that correspond to the global maximum in a) are taken as the optimal origin time and location of the seismic source.	52
4.3	Idealized representation of the NM_i variable along time. The parameter α_d is the detection threshold. In this example, the maximum number of iterations is set at $N_s = 1$. Notice that in the limiting case $N_w = 1$, a seismic source would be declared at all time positions where $NM_1 < \alpha_d$ producing multiple spurious sources. On the other hand, if N_w is for example four times bigger than it is showed, N_s must have a minimum value of 2 in order to detect the two sources appearing as valleys in the NM_1 function.	54
4.4	Geometry of the microseismic monitoring experiment used to test the modified BOMP. The treatment well passes through the monitored volume, which is outlined by a grey box. The monitoring well is where the receiver locations are displayed.	56
4.5	Average source parameter inversion errors for different levels of SNR. Crosses corresponds to model 1, dashes to model 2, and circles to model 3. a) error in moment tensor estimation, the percentages relate to the number of realizations resolved by the algorithm in each SNR level (top values correspond to model 1, followed by models 2 and 3), b) error in origin time, c) error in location and d) average value of the index $p \in [1, n_h]$ where the solutions were found. Vertical bars are standard deviation.	59
4.6	Effects of inaccurate Green's functions in the inversion output. The inaccuracy in the Green's functions is due to source locations not coinciding with the grid nodes. a) error in moment tensor estimation, b) error in origin time, c) error in location and d) average value of the index $p \in [1, n_h]$ where the solutions were found. Vertical bars are standard deviation.	60
4.7	Effects of inaccurate Green's functions in the inversion output. After computing the observations, the velocity model is perturbed by 5% (circles) and 3% (dashes) to compute dictionaries of Green's functions. a) error in moment tensor estimation, b) error in origin time, c) error in location and d) average value of the index $p \in [1, n_h]$ where the solutions were found. Vertical bars are standard deviation.	61
4.8	Epicentral location of the 18 June 2002 Caborn, Indiana earthquake and distribution of seismic stations considered in this work (black triangles). The beach ball corresponds to the solution obtained by Kim (2003). The Wabash Valley Seismic Zone (WVZS) is delimited by the dotted line. The Wabash Valley Fault System (WVFS) is enclosed by the dashed line. The black square is the surface projection of the grid used in this study.	63

4.9	Source parameters of the 18 June 2002 Caborn, Indiana earthquake determined by Kim (2003). The source mechanism is represented by the beach-ball diagram (lower-hemisphere projection).	64
4.10	Source parameters of the 18 June 2002 Caborn, Indiana earthquake determined by the hypothetical monitoring system. The source mechanism is represented by the beach-ball diagram (lower-hemisphere projection). Solid line: observed records. Dotted line: synthetic records obtained from the forward modelling of the modified BOMP solution.	68
5.1	Diagram showing the encoding process of the seismic observations. The sensing matrix Φ acts as an encoder compressing the samples in the observations vector \mathbf{u} . Assuming that the original signal is sampled at a Nyquist rate, it turns out that the number of non-adaptive measurements that results from encoding \mathbf{u} is less than the number of samples required to sample \mathbf{u} through the Nyquist criterion. Compressive sensing (CS) provides conditions for the encoding and posterior decoding of \mathbf{f} to recover \mathbf{u} . Practical CS aims at the development of recording instruments that allow the acquisition of the encoded signal \mathbf{f} . In this thesis however, CS principles are applied to signals already recorded at a Nyquist rate, where further advantages arise from the processing of those signals in a compressed domain.	73
5.2	Distribution of recording stations (blue-red triangles) and grid of virtual sources (grey volume) used in the numerical example.	75
5.3	Estimation errors for 500 simulations of source parameter inversion using CS principles. (a) Normalized Misfit of the solution $NM = \ \mathbf{u} - \hat{\mathbf{u}}\ _2^2 / \ \mathbf{u}\ _2^2$. (b) Location error $= \ \mathbf{x} - \hat{\mathbf{x}}\ _2$, where \mathbf{x} is the true location vector of the source. (c) Origin time error $= OT - \hat{OT}$, where OT is the true origin time. (d) Normalized Mean Square Error of the estimated moment tensor $NMSE = \ \mathbf{m} - \hat{\mathbf{m}}\ _2^2 / \ \mathbf{m}\ _2^2$. Variables with a hat denote estimations from the inversion.	76
5.4	Resolvability of two seismic sources in the same vector of observations. Only observations for one receiver are presented. From top to bottom the traces correspond to the x (East), y (North) and vertical components of the receiver. NRE is the abbreviation for the number of resolved events by the inversion algorithm. a) and b) the algorithm cannot resolve the two sources. c) and d) the algorithm is able to resolve the two sources which is also reflected in a sharp decrease in the spectral norm operator $\rho(\cdot)$	78
5.5	Resolvability of two seismic sources in the same observations vector. The two seismic events have the same origin time and source mechanism. These plots show a snapshot of the objective function used in the modified BOMP at the exact origin time of the sources. Top: both sources have the same location. Middle: separation between locations is 4 nodes in the y direction (20 km). Bottom: separation between locations is 6 nodes in the y direction (30 km). (a) and (b) columns are respectively iterations one and two of the modified BOMP. Each iteration finds the source parameters of a seismic source. Variable ρ is the spectral norm measurement for each case. NRE stands for number of resolved events.	79

5.6	Results for 500 realizations of sensing matrix in 5 different scenarios of compression to detect and estimate source parameters for the 18 June, 2002 Caborn earthquake. NNA stands for number of non-adaptive measurements per non-zero coefficient in the solution. Vertical bars are standard deviations. a) Average source decompositions, y -axis is percentage for DC and CLVD components (Knopoff and Randall, 1970), and degrees for the angle between the slip vector and the fault plane (Vavrycuk, 2001). b) Average position of the value in the correlation function between compressed Green's functions and residuals in BOMP where the solution was found (values are sorted from maximum to minimum). c) Average processing time.	82
6.1	a) Synthetic trace composed of two arrivals with arbitrary phase and frequency content. b) Synthetic trace contaminated with random noise. c) Inversion results using the damped least-squares method (equation (6.4)). d) Data reconstruction from the time-frequency map in c). e) Inversion results using the ℓ_1 -norm regularization (equation 6.5). f) Data reconstruction from the time-frequency map in e).	90
6.2	Reference complex Ricker wavelet. Real part (dashed line), imaginary part (dotted line) and envelope (solid line).	91
6.3	Comparison of denoising results between the 1C and 3C versions of the sparse time-frequency transform. a) Synthetic 3C data. b) Synthetic data contaminated with random noise. c) Denoised data using the 1C sparse time-frequency transform. d) Denoised data using the 3C sparse time-frequency transform.	92
6.4	Velocity model used to compute the synthetic microseismic data. Triangles denote receiver depth locations. The star denotes the source depth.	93
6.5	Denoising results in a synthetic microseismic dataset. Columns a), b) and c) correspond to the x , y and z components, respectively. From top to bottom, the first row is the synthetic microseismic data. The second row is the synthetic data contaminated with random noise. The third row is the noisy data filtered with a low-pass frequency filter. The fourth row is the denoised data using the 3C sparse transform.	95
6.6	Comparison of hodogram analysis for receiver three of the synthetic microseismic dataset of Figure 6.5. Column a) synthetic data. Column b) synthetic data contaminated with random noise. Column c) noisy data filtered with a low-pass frequency filter. Column d) denoised data using the 3C sparse transform. Vertical lines in the first row delimit the hodogram window. Hodogram curves in column a) are repeated with dotted lines in columns b), c) and d) for comparison.	96
6.7	a) Denoised z component of the real microseismic dataset. b) Time-frequency map at central frequency 222 Hz for the data in a). Time-frequency map at central frequency 297 Hz for the data in a).	98
6.8	Denoising results in the x component of the real microseismic data. a) Raw microseismic data. b) Raw data filtered with a low-pass frequency filter. c) Denoised data using the 3C sparse transform.	99

6.9	Denoising results in the y component of the real microseismic data. a) Raw microseismic data. b) Raw data filtered with a low-pass frequency filter. c) Denoised data using the 3C sparse transform.	100
6.10	Denoising results in the z component of the real microseismic data. a) Raw microseismic data. b) Raw data filtered with a low-pass frequency filter. c) Denoised data using the 3C sparse transform.	101
6.11	Comparison of hodogram analysis with real microseismic data (receiver 7). Column a) raw data. Column b) data filtered with a low-pass frequency filter. Column c) denoised data using the 3C sparse transform. Vertical lines in the first row delimit the hodogram window. Hodogram curves in column a) are repeated with dotted lines in columns b) and c) for comparison.	102
7.1	Effect of noise on the build-up of the STA/LTA curve. The black traces are synthetics with decreasing SNR from bottom to top. Below each synthetic the corresponding STA/LTA curve is displayed in blue with the detection threshold in red. The black vertical line denotes the location of the onset of the wave arrival in the top synthetic.	105
7.2	From top to bottom: raw trace with seismic arrival; standard STA/LTA curve; derivative of the STA/LTA curve; minimization of the cost function $J = \ \mathbf{x}_{STA} - \mathbf{X}_{LTA}\mathbf{y}_{STA/LTA}\ _2^2 + \lambda\ \mathbf{D}\mathbf{y}_{STA/LTA}\ _2^2$ for a small value of λ ; same minimization of the previous cost function for a large value of λ	107
7.3	From top to bottom: raw trace with seismic arrival; standard STA/LTA curve; non-linear norm R applied to the STA/LTA ; non-linear norm R applied to the derivative of the STA/LTA ; minimization of the cost function in 7.4 with R defined as in 7.5.	108
7.4	From top to bottom: synthetic microseismic trace (s), synthetic with random noise (s + n), STA/LTA curve from noisy synthetic, blocky version of the STA/LTA curve from noisy synthetic, denoised synthetic using the 1C sparse time-frequency transform, STA/LTA curve for the denoised trace, blocky version of the STA/LTA curve for the denoised trace. Red vertical lines indicate the approximate position of P and S arrivals in the original synthetic trace.	109
7.5	From top to bottom: real microseismic trace (raw), STA/LTA curve from raw trace, blocky version of the STA/LTA curve from raw trace, denoised trace using the 1C sparse time-frequency transform, STA/LTA curve from the denoised trace, blocky version of the STA/LTA curve from the denoised trace. The red vertical line denotes the approximate position of a wave arrival.111	111

List of symbols

G_{ij}	point force Green's function
M_{jk}	seismic moment tensor coefficient
\dot{M}_{jk}	jk -th element of the moment rate tensor
R_{ijk}	radiation pattern coefficient
R_{ij}	element of the resolution matrix
C	condition number
C_ϵ	condition number considering only eigenvalues different from zero
N_w	number of samples in the processing window
N_x, N_y, N_z	grid dimensions
N_G	number of samples in a Green's function
N_{st}	number of receivers
N_s	number of expected sources in a vector of observations
N_t	total number of blocks in a dictionary
N_u	number of samples in a trace of observations
P	compressional wave
S	shear wave
V_p	speed of P waves
V_s	speed of S waves
\mathbb{G}	superdictionary of Green's functions
\mathbf{G}	dictionary of Green's functions
\mathbf{G}_j	dictionary \mathbf{G} delayed in time by j samples
\mathbf{L}	matrix of eigenvalues

\mathbf{L}^\dagger	pseudoinverse of \mathbf{L}
\mathbf{Q}	matrix of eigenvectors
\mathbf{R}	model resolution matrix
\mathbf{X}	matrix with radiation pattern coefficients
Γ_i, Λ_i	index sets after iteration i
Φ	sensing matrix
Ψ	dictionary of basis functions
$\Psi[l]$	column-block of atoms of the dictionary Ψ
$\hat{\Psi}$	perturbation to the dictionary
Θ_m	matrix of rotation for the moment tensor in vector form
\mathbf{d}, \mathbf{u}	vector of observations
\mathbf{n}	vector of noise
\mathbf{m}	seismic moment tensor in vector form
$\hat{\mathbf{m}}$	approximation of \mathbf{m}
$\tilde{\mathbf{m}}$	intermediate approximation of \mathbf{m}
\mathbf{m}_θ	seismic moment tensor rotated to a local reference system
\mathbf{r}	position vector
\mathbf{r}_i	vector of residuals after iteration i
$\tilde{\mathbf{s}}$	intermediate solution in the modified Block Orthogonal Matching Pursuit
\mathbf{x}	vector of representation coefficients
$\hat{\mathbf{x}}$	approximation of the representation coefficients
$\mathbf{x}[l]$	block of coefficients in vector \mathbf{x}
\mathbf{y}	time series in vector form
$\hat{\mathbf{y}}$	time series with inaccurate measurements
ξ	source location
Δt	sampling rate
α	angle between slip vector and fault's plane
α_1	angular aperture between two receivers with respect to a source location
α_2	angular aperture between two receivers with respect to a third one

α_d	detectability threshold
β	secondary detectability threshold
ψ_j	atom that belongs to the dictionary Ψ
λ_{max}	largest eigenvalue of a matrix
$\lambda_{min \neq 0}$	smallest eigenvalue of a matrix that is different from zero
μ_B	block-coherence of a dictionary
ν	sub-coherence of a dictionary
π	pi number
ρ	density
$\rho(\cdot)$	spectral norm operator
σ_n	standard deviation of noise
τ	duration of the source time function
θ	azimuth from North
a	distance between grid nodes
c	speed of seismic waves
d	length of a block in a dictionary
r	distance between a source and a receiver location
t	time variable
x, y, z	components of a cartesian reference system
$f(k, l)$	objective function in the Block Orthogonal Matching Pursuit
f_{max}	maximum frequency
m_i	element of the moment tensor in vector form
n_h	size of the search in the modified Block Orthogonal Matching Pursuit
$s(t)$	source time function
t_R	response time of the monitoring system
t_T	transmission time
t_p	processing time
t_{NG}	time duration of a Green's function
t_{Nu}	time duration of an observations trace

t_{N_w}	time duration of the processing window
u_i	displacement observed in the i -th component of a receiver
x_j	coefficient in vector \mathbf{x}
K	constant that depend on the medium properties
R_i	receiver tag
d	horizontal distance between source and receiver(s)

CHAPTER 1

Introduction

1.1 Background ¹

The recovery of origin time, location and moment tensor of seismic events is an inverse problem that has been widely investigated in the study of earthquakes. On a global scale, fast moment tensor inversions are routinely performed by institutes like the Global Centroid Moment Tensor (Global CMT) project (Dziewonski et al., 1981), the United States Geological Survey (USGS) (Sipkin, 1982), and the Earthquake Research Institute (ERI) (Kawakatsu, 1995). The CMT algorithm is mainly based on the pioneering work of Dziewonski et al. (1981), which expresses the source displacement field as a linear combination of *excitation kernels*. These kernels are subsequently modified through perturbations to an initial source model to minimize a misfit function. This approach was later implemented by Kawakatsu (1995) for the analysis of long-period body wave data archived at the ERI in Japan. Sipkin (1982) proposes two alternative methodologies, the first of which is based on the theory of optimal filter design. Under this approach, the seismic moment tensor is treated as a filter that produces the recorded displacement field through convolution with a set of Green's functions. Hence, using least-squares techniques, this method determines the optimum values for the source mechanism. The second methodology is a non-linear approach that seeks a set of filters to convolve with the observed seismograms and produce average values for each of the moment tensor elements. In all of the aforementioned methodologies, information about location and origin time is provided by an independent source and is required to initialize the inversion process.

At regional distances, fast solutions are also determined close to seismically active zones.

¹A version of this section has been published. Vera Rodriguez, I., Sacchi, M. and Gu, Y. 2012. *Geophysical Journal International* 188: 1188-1202.

For example, in Northern and Central California, the Berkeley Seismological Laboratory operates two different systems for automatic determination of source mechanisms. The first is a time-domain method based on the non-linear relationship between the double couple source model and the source displacement field (Dreger and Helmberger, 1993). The second is a frequency-domain approach designed to operate over surface wave data (Romanowicz, 1982; Romanowicz et al., 1993). These two methods have been automated (Pasyanos et al., 1996) and the time-domain approach has been integrated into the TDMT_INV distributable package (Dreger, 2003); the latter is currently used for moment tensor inversions by several institutions around the world (e.g., Dreger, 2003; Clinton et al., 2006; Scognamiglio et al., 2009). Bernardi et al. (2004) present a feasibility analysis of a system for automatic, near real-time source parameter recovery in the European-Mediterranean region. Similar to methods used for global monitoring, the aforementioned regional inversion routines are triggered only when pre-existing information about earthquake occurrence is available. For instance, a lag-time for the calculation of the information necessary to start the estimation of the source moment tensor is implicitly required in all of these monitoring systems. The feasibility of extracting continuous, automatic, real-time source parameters (origin time, location and moment tensor) is examined by Tajima et al. (2002). The system is based on the grid search algorithm (Kawakatsu, 1998) currently adopted for the monitoring of earthquakes in California (Guilhem and Dreger, 2010) and Japan (Tsuruoka et al., 2009). Also, the grid search method using 3D Green's functions has been tested to monitor seismic activity in Taiwan (Lee et al., 2010). The main difference of the grid search approach from other monitoring systems, aside from not requiring a priori information, is its capability to operate over a continuous flow of seismic data. On the other hand, a grid-based approach is limited by grid density and the maximum number of station recordings to be processed without compromising real-time response.

Besides earthquake monitoring, fast estimations of the source parameters can have important impact on applied geophysics, particularly the study of microseismic sources induced by hydraulic fracturing of oil and gas wells. Oil and gas wells are fractured to enhance production (Cipolla and Wright, 2002). This is achieved through the injection of fluids at high pressures, which percolate into rock formations and potentially trigger microseismic events that are recorded in nearby receiver locations (Maxwell, 2005; Maxwell et al., 2010; Duncan and Eisner, 2010). By analyzing the properties of the induced seismicity, interpreters generate estimates of fracture's geometry, complexity, permeability, and connectivity with pre-existing faults and/or fractures (e.g., House et al., 1996; Shapiro et al., 1999; Maxwell et al., 2002; Rothert and Shapiro, 2007; Fischer et al., 2008; Bayuk et al., 2009; Warpinski, 2009). A thorough understanding of the fractures induced through hydraulic injection is important for the planning and development of an oilfield (Cipolla and Wright, 2002). The source mechanism of microseismic events provides information to evaluate models of

fracture propagation (Warpinski, 1994; Warpinski et al., 2001), to determine the state of stresses in the neighbourhood of the injection well (House et al., 1996; Rutledge and Phillips, 2003) and, in general, as a source of information to analyze the effectiveness of the induced fractures. The importance in obtaining information about fracture propagation during the injection process has been largely recognized (Maxwell et al., 2002; LeCampion and Jeffrey, 2004; Le Calvez et al., 2006), however, the estimation of source mechanisms is still usually performed in a post-processing stage. Additionally, methodologies for the inversion of moment tensors with microseismic data do not produce simultaneous estimates of the three source parameters (e.g. Nolen-Hoeksema and Ruff, 2001; Rutledge and Phillips, 2003; Jechumtalova and Eisner, 2008; Sileny et al., 2009; Godano et al., 2010). Locations and origin times are estimated independently, which introduces a lag-time for the purpose of real-time implementation. Fast availability of source geometries is a valuable tool that can be incorporated into the decision-making process during a hydraulic injection. This thesis develops a framework for the development of new real-time or near real-time seismic source monitoring systems. This framework permits the automatic, continuous and simultaneous estimation of the source parameters using the theory of compressive sensing. Compressive sensing (CS) is a relatively new field of signal processing and applied mathematics that studies sampling and recovery conditions for signals susceptible to a sparse representation via a known basis or dictionary (Candes et al., 2006; Donoho, 2006). A major result of CS is the specification of protocols for the sampling of signals using a number of non-adaptive measurements that is below the number of samples required with the traditional Nyquist criterion. The non-adaptive measurements are linear combinations of the information contained in the signal weighted by coefficients prescribed in the form of a sensing matrix. In other words, using CS a signal can be acquired and transmitted in a compressed form, and ultimately uncompressed without perceptible loss of information. This represents an important improvement in efficiency from the traditional practice where a signal is acquired in full form, then compressed throwing away information, transmitted, and finally uncompressed. CS principles have found application in the fields of Magnetic Resonance Imaging (MRI) (Lustig et al., 2007), digital camera design (Takhar et al., 2006), acquisition of astronomical data (Bobin et al., 2008). In geophysical applications, CS has been implemented for earthquake location (Yao et al., 2011), simultaneous estimation of origin time, location and seismic moment tensor (Vera Rodriguez et al., 2010b), and acquisition of seismic data with simultaneous active sources (Herrmann, 2010).

Microseismicity induced during hydraulic injections is characterized for its small magnitude (Maxwell, 2005; Shemeta and Anderson, 2010). This is an important reason why microseismic records often display a low signal-to-noise ratio (SNR). Accuracy and reliability of the location and other event attributes derived from microseismic traces is influenced by this strong noise content (Maxwell, 2009). Thus, noise attenuation is a desirable step in

microseismic processing in order to improve the quality of subsequent processes. In general, noise can be considered as the part of the measured signal that is not of interest and it is usually divided into coherent and non-coherent. Two different approaches can be followed to improve the SNR in microseismic experiments, either improving the acquisition equipment and/or acquisition geometry or through the use of signal processing techniques. One of the most basic signal processing methods to attenuate noise is frequency filtering. Frequency filters are effective to attenuate frequencies outside user-defined cut-off values. However, signal and noise often share frequency bands which means that part of the signal is also filtered out with the noise, while the noise might not be fully attenuated. A more advanced denoising technique requires the use of matched filters (Eisner et al., 2008). A matched filter is designed by selecting a microseismic event with high SNR to be used as a "master" or reference event. Through the cross-correlation of the "master" event with other parts of the signal, events with lower SNR, similar source mechanism and nearby location are "matched" and found. Other examples of advanced denoising techniques are found in global seismology. For instance, in the analysis of earthquake precursors Sobolev and Lyubushin (2006) apply a wavelet transform and thresholding criteria in the time-frequency plane to facilitate the identification of microseismic events. Similarly, Baig et al. (2009) employ time-frequency analysis to denoise seismic noise cross-correlations. In reflection seismology, Bonar and Sacchi (2010) proposed a method for spectral decomposition to analyze seismic sections. The decomposition technique is based upon the transform of the time domain data into a sparse time-frequency map. When implemented for denoising purposes this method is equivalent to Basis Pursuit Denoising (BPDN) (Chen et al., 1998). This thesis presents an application of BPDN to microseismic data. Furthermore, the BPDN is extended to the three-component case through the use of group sparsity measurements.

Event location is an important step in the processing of microseismic data. For algorithms that determine location based on P- and S-wave arrival times, accurate time-picking is crucial. For small datasets accurate time-picks can be obtained by human interaction, however for large datasets, manual time-picks can be impractical, especially if a quick response is required. Different methodologies have been proposed to automate the process of time-picking with the short-term-average/long-term average (*STA/LTA*) filter (Allen, 1978) being the most common. The *STA/LTA* filter acts over a characteristic function (*CF*), which is usually the energy of the seismic trace. Allen (1978) also uses a *CF* that takes into consideration the time derivative of the seismic trace, giving it a weight that depends on the *SNR*. A disadvantage in the use of the *STA/LTA* filter resides in the rate of build-up of the curve when an arrival is detected. If the *SNR* is high, the *STA/LTA* curve will show a sharp increase at the location of the onset of the arrival, however as the *SNR* lowers, the *STA/LTA* curve displays a slower build-up that introduces error in the determination of the onset of the arrival. Improvements in automatic time-picking can be achieved through

the use of inverse theory. More specifically, this thesis presents the use of non-linear inversion techniques for the recovery of blocky STA/LTA curves that improve the performance of automatic time-pickers.

1.2 Contributions

The contributions from this thesis can be summarized as:

- Highlighting the importance of the reference system where the source mechanism solution is represented when the array of monitoring receivers forms a vertical line. If the moment tensor is not in a reference system aligned with the plane of observation from the source, any error in the value of the dipole perpendicular to this plane is propagated to two other elements of the tensor. An equation for the estimation of moment tensors expressed in the local aligned system is proposed.
- Exemplification of the use of the condition number as a tool for acquisition design in order to resolve 5 and 6 elements of the moment tensor in vertical and deviated wells, respectively.
- Introduction of sparse representation theory into the simultaneous estimation of origin time, location and moment tensor of seismic events. Benefits of the proposed method include: automatic and continuous processing of seismic observations.
- Introduction of compressive sensing into the simultaneous estimation of the source parameters. The use of compressive sensing extends the real-time applicability of the sparse representation method to scenarios where other currently available methods cannot perform in real-time (e.g., dense networks of monitoring stations).
- Proposing a denoising method for one-component and three-component microseismic traces based on a time-frequency representation constrained to display sparsity and group sparsity, respectively.
- Exemplifying the use of non-linear constraints for the determination of more accurate time-picks in automatic systems.

1.3 Organization of the thesis

Chapter 2 presents background theory related to the quantitative study of seismic sources. This chapter is important for the understanding of the thesis because it presents the theoretical assumptions used across the rest of the chapters. Additionally, this chapter provides

details on hydraulic fracturing technology and its characterization through the methodology known as microseismic monitoring. An important part of the conclusions in this thesis are obtained from numerical experiments in the setting of microseismic monitoring experiments. Finally, chapter 2 presents an overview on the theory of compressive sensing. This overview covers only the details of the theory that are relevant to the application proposed in this thesis. Further details on compressive sensing can be found in the cited references.

Chapter 3 investigates the resolvability of the complete seismic moment tensor in hydraulic fracture monitoring with a single array of receivers. In the case of a vertical array, the resolution matrix demonstrates that a correct representation of the 5 resolvable elements of the moment tensor is only possible in a local reference system. In these cases, a suitable choice of condition number can assist the acquisition design. For a non-vertical array, numerical modeling experiments suggest that the required distance and orientation of receivers for a full moment tensor inversion can be satisfied in a deviated well. In this case, information embedded in the condition number is valuable for determining the required distribution of receivers along the well.

In **Chapter 4** it is showed that through a proper parametrization, the source displacement field of a seismic event can be efficiently reconstructed by a redundant dictionary of Green's functions based on sparse representation theory. Then, by subjecting the pre-existing event records and pre-computed dictionary of Green's functions into a sparsity-promoting algorithm, it is possible to simultaneously evaluate the origin time, location coordinates and seismic moment tensor. In this chapter it is demonstrated the effectiveness and accuracy of the aforementioned approach via 1) detection of microseismic events produced during the hydraulic fracturing of oil and gas wells, and 2) inversion of a small-magnitude, regional earthquake (June 18, 2002 in Caborn, Indiana) data.

Chapter 5 introduces concepts from compressive sensing into the source parameter determination problem. Compressive sensing techniques enable determination of source parameters in a compressed space, where the dimensions of the variables involved in the inversion are significantly reduced. Results using a hypothetical monitoring network with a dense number of recording stations show that a compressed catalog of Green's functions with 0.004% of its original size recovers the exact source parameters in more than 50% of the tests. The gains in processing time in this case drop from an estimated 90 days to browse a solution in the uncompressed catalog to 41.57 s to obtain an estimation using the compressed catalog. For simultaneous events, the compressive sensing approach does not appear to influence the estimation results beyond the limitations presented by the uncompressed case. The main concern in the use of compressive sensing are detectability issues observed when the amount of compression is beyond the minimum number of required non-adaptive measurements. Tests using real data from the 18 June, 2002 Caborn Indiana earthquake show that the

presence of noise and inaccurate Green's functions require a larger number of non-adaptive measurements (i.e., less compression) to reproduce the solution obtained with the uncompressed catalog. In this case, numerical simulation enables the assessment of the amount of compression that provides a reasonable rate of detectability. Overall, the numerical experiments presented in this chapter demonstrate the effectiveness of the compressed domain inversion method in the real-time monitoring of seismic events.

In **Chapter 6** a non-coherent noise attenuation technique based on a constrained time-frequency transform is presented. When applied to one-component data, the transform corresponds to a sparse representation of the microseismic signal in terms of a dictionary of complex Ricker wavelets. A synthetic example illustrates the superior performance of the sparse constraint for denoising objectives when compared to the standard least squares regularization. As the arrival time and frequency content of any wavefront are equivalent in the three components of a single receiver, the extension of the sparse transform to three-component data is accomplished when the three components are considered to share the same sparsity pattern in the time-frequency plane. Application of the three-component sparse transform to synthetic and real microseismic datasets demonstrate the advantages of this technique when the denoised results are compared against the original and low-pass filtered version of the noisy data. Furthermore, a comparison of hodograms between original, low-pass and denoised traces shows that the denoising process preserves the phase and relative amplitude information present in the input data. The benefits of the three-component transform are highlighted particularly in cases where the wave arrivals are measured in the three components of a receiver but are only visible in two components due to the prevailing signal to noise ratio.

The *STA/LTA* filter is commonly used in automatic systems for the determination of arrival times. Time-picks are established at positions in time where the *STA/LTA* curve surpasses a user-defined threshold. The presence of noise produces a slow build-up of the *STA/LTA* curve that introduces errors in the automatically determined time-picks. **Chapter 7** exemplifies the use of a non-linear constraint for the estimation of blocky versions of the *STA/LTA* curve. The abrupt behaviour of the blocky curves improves the accuracy of the automatic time-picks. Synthetic and real microseismic data examples demonstrate the benefits of combining both the denoising method presented in Chapter 6 together with the estimation of blocky *STA/LTA* curves in the determination of accurate, automatic time-picks.

Finally, **Chapter 8** is dedicated to the Conclusions from this thesis.

CHAPTER 2

Theory

2.1 Quantitative description of the seismic source

Seismic energy is generated as a result of transient imbalances of stresses in an elastic medium (Lay and Wallace, 1995). Traveling in the form of elastic waves, seismic energy can be observed far from the location where it originated (Figure 2.1). The region covered by the seismic source presents finite physical proportions. However, the concept of equivalent point force allows a practical quantitative description at wavelengths larger than the source dimensions (Aki and Richards, 2009). An equivalent point source produces the same displacement field as the real source with the difference of being confined to a point in space. When approximating a finite source with a point equivalent the periods of the radiated waves must be comparable or larger to the transient imbalance or source duration (Lay and Wallace, 1995). Under this approximation, the location of the seismic event is assigned to the *centroid of the stress glut density* (Jost and Herrmann, 1989). In simple terms the stress glut is the difference between the actual stresses in the medium and the state of stresses described by Hooke's law. Most deviations from linear elasticity occur in the region of *instantaneous failure* (Dahlen and Tromp, 1998), which in consequence presents the largest values of stress glut density. In this thesis the point force approximation is employed to study seismic sources in the scenarios under consideration. In an isotropic homogeneous medium, the displacement due to an arbitrary point force is determined by solving the elastodynamic wave equation (Madariaga, 2007, equation 1)

$$\rho \frac{\partial^2}{\partial t^2} \mathbf{u}(\mathbf{r}, t) = (\lambda + \mu) \nabla(\nabla \cdot \mathbf{u}(\mathbf{r}, t)) + \mu \nabla^2 \mathbf{u}(\mathbf{r}, t) + \mathbf{f}(\mathbf{r}, t), \quad (2.1)$$

where ρ is density, and λ and μ are elastic constants. \mathbf{u} is the displacement field at position

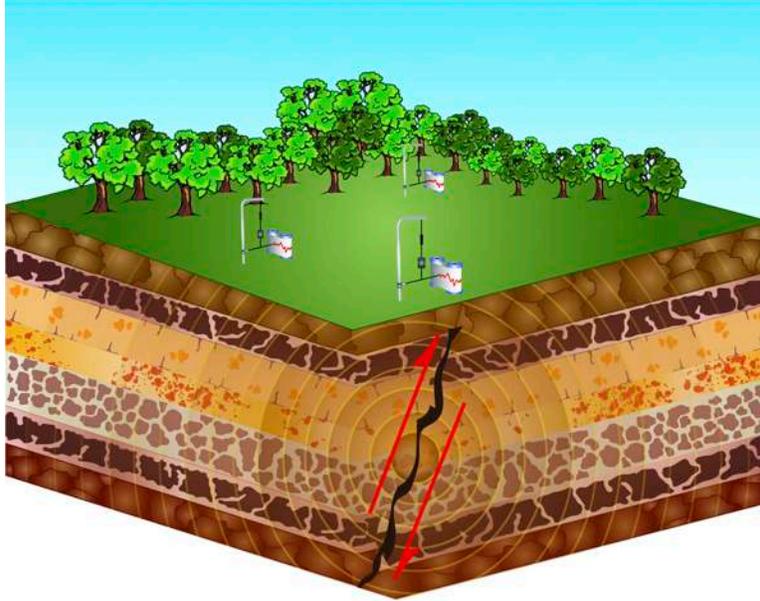


Figure 2.1: Propagation of seismic energy in an elastic medium and its recording in surface stations. (Image created by Cecilia Camet)

\mathbf{r} and time t , and $\mathbf{f}(\mathbf{r}, t)$ is a body force density, which is related to the source time function $s(t)$ by (Madariaga, 2007, equation 2)

$$\mathbf{f}(\mathbf{r}, t) = \mathbf{f}s(t)\delta(\mathbf{r} - \mathbf{r}_0). \quad (2.2)$$

In the last equation, \mathbf{f} is a unitary vector pointing in the direction where the source is being applied. The source time function represents the time history of the applied force (Lay and Wallace, 1995). The displacement field due to a point force at the origin of the reference system and applied in the coordinate direction specified by the index i is given by the Stokes solution (Lay and Wallace, 1995, equation 8.45)

$$\begin{aligned} u_i(\mathbf{r}, t) = & \frac{1}{4\pi\rho}(3\gamma_i\gamma_j - \delta_{ij})\frac{1}{r^3} \int_{r/V_p}^{r/V_s} \frac{r}{V_p} s(t - \frac{r}{V_p}) d\frac{r}{V_p} \\ & + \frac{1}{4\pi\rho V_p^2} \gamma_i\gamma_j \frac{1}{r} s(t - \frac{r}{V_p}) \\ & - \frac{1}{4\pi\rho V_s^2} (\gamma_i\gamma_j - \delta_{ij}) \frac{1}{r} s(t - \frac{r}{V_s}) \end{aligned}, \quad (2.3)$$

where $r = |\mathbf{r} - \boldsymbol{\xi}|$ is the distance between the source and the observation point, and $V_p = \sqrt{\frac{\lambda+2\mu}{\rho}}$ and $V_s = \sqrt{\frac{\mu}{\rho}}$ are the velocities of P and S waves, respectively. The variables γ_i are direction cosines and δ_{ij} is the Kronecker delta. Equation 2.3 can be analyzed in two parts. The first part corresponds to the term with the integral, which decays as $\frac{1}{r^2}$. Considering

that this term is only important at short distances from the source it is usually identified as the near-field displacement. The second part corresponds to the remaining two terms that decay as $\frac{1}{r}$, the contributions from these terms are more important as the observer is located farther from the source, therefore these terms are called the far-field displacement (Aki and Richards, 2009). In the applications considered in this thesis, the receiver positions are assumed to be located at a distance where only the far-field displacement plays an important role in the description of the wave propagation (e.g., Jost and Herrmann, 1989).

2.1.1 The seismic moment tensor

Equation 2.3 represents the displacement field due to a single force pointing in a coordinate direction. In reality, seismic sources are more complex than that. Advocating for the development of a simple, yet general theory, researchers observed that many seismic sources can be represented as a linear combination of a finite number of elemental force couples (Jost and Herrmann, 1989) (Figure 2.2). These elemental couples are arranged into a symmetric (Gilbert, 1970) second order tensor called the seismic moment tensor, this is

$$\mathbf{M} = \begin{pmatrix} M_{11} & M_{12} & M_{13} \\ M_{12} & M_{22} & M_{23} \\ M_{13} & M_{23} & M_{33} \end{pmatrix}. \quad (2.4)$$

The diagonal elements of the moment tensor correspond to dipoles, which account for the volumetric changes of the source. The off-diagonal elements are force couples. Seismic sources that result from internal deformation have zero net force and zero net moment (Madariaga, 2007). Dipoles and double-couples formed by symmetric off-diagonal components of the moment tensor not only comply with these conditions but also have been observed to correspond with radiation patterns from earthquakes (Aki and Richards, 2009). Using the moment tensor, a general seismic source with arbitrary orientation can be represented as a linear weighted-sum of equivalent point dipoles and double-couples (Gilbert, 1973) (Figure 2.3).

Moment tensor decompositions

Through eigenvalue decomposition, the seismic moment tensor can be rotated into a principal-axis system. The directions specified by the eigenvectors correspond to the maximum tensional deformation (T-axis), the maximum compressional deformation (P-axis), and the null axis (intermediate eigenvalue or B-axis) (Lay and Wallace, 1995) (see Figure 2.3). A diagonalized moment tensor can be further separated into an isotropic and a deviatoric part defined as (Jost and Herrmann, 1989, equation 23)

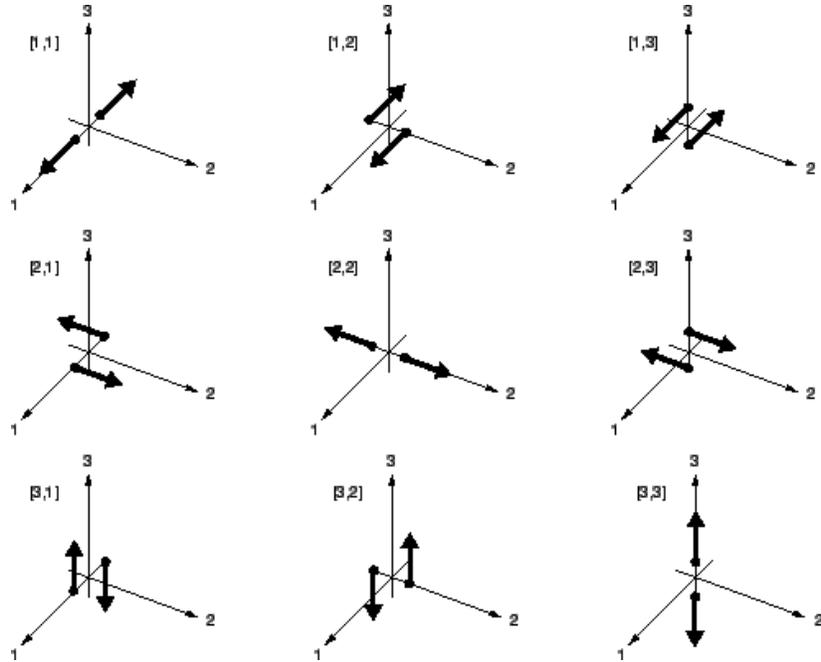


Figure 2.2: The nine elemental force couples that form the seismic moment tensor. Modified after Aki and Richards (2009)

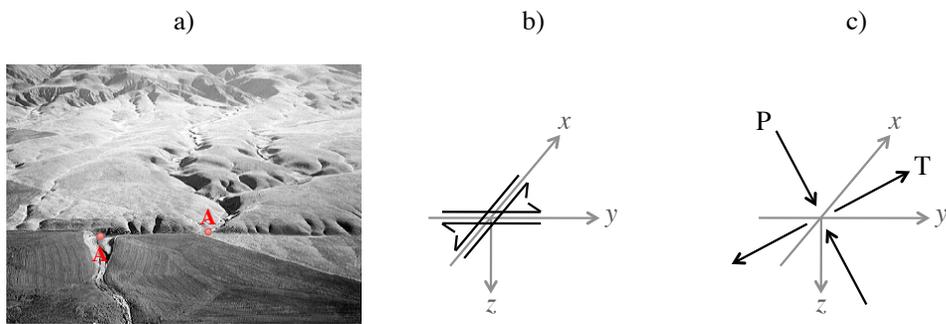


Figure 2.3: Point force approximation of a finite fault. a) Real fault with finite extension. b) Point force approximation. c) Principal axes for the point force approximation. Picture in a) modified from USGS website on San Andreas Fault. Original photograph taken by Robert E. Wallace.

$$\mathbf{M} = \begin{pmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} tr(\mathbf{M}) & 0 & 0 \\ 0 & tr(\mathbf{M}) & 0 \\ 0 & 0 & tr(\mathbf{M}) \end{pmatrix} + \begin{pmatrix} M_1^* & 0 & 0 \\ 0 & M_2^* & 0 \\ 0 & 0 & M_3^* \end{pmatrix} = \mathbf{M}_{ISO} + \mathbf{M}_{DEVI}, \quad (2.5)$$

where $M_i^* = M_i - \frac{1}{3}tr(\mathbf{M})$ are deviatoric eigenvalues. The sum of the deviatoric eigenvalues is zero; while the diagonal of \mathbf{M}_{ISO} adds to $tr(\mathbf{M})$, which stands for the trace of \mathbf{M} . The deviatoric moment tensor can also be decomposed into a variety of combinations of double-couples (DC) and compensated linear vector dipoles (CLVD). A pure DC expressed into its principal axes presents one eigenvalue equals to zero with a positive eigenvalue pointing in the T-axis and a negative eigenvalue pointing in the P-axis, both with the same magnitude. A CLVD presents two diagonal elements that are half the magnitude of the third element with opposite sign (Knopoff and Randall, 1970). A decomposition of interest in this thesis has the form (Jost and Herrmann, 1989, equation 36),

$$\mathbf{M} = \mathbf{M}_{ISO} + \mathbf{M}_{DC} + \mathbf{M}_{CLVD}, \quad (2.6)$$

where (Vavrycuk, 2001)

$$\mathbf{M}_{DC} = (1 - 2|\epsilon|) \begin{pmatrix} 0 & 0 & 0 \\ 0 & -M_3^* & 0 \\ 0 & 0 & M_3^* \end{pmatrix},$$

and

$$\mathbf{M}_{CLVD} = |\epsilon| \begin{pmatrix} -M_3^* & 0 & 0 \\ 0 & -M_3^* & 0 \\ 0 & 0 & 2M_3^* \end{pmatrix}.$$

The coefficient ϵ is a measure of the ratio of CLVD to DC components and is computed as (Julian et al., 1998, equation 18) $\epsilon = \frac{-M_1^*}{|M_3^*|}$, where it is considered that $|M_1^*| \leq |M_2^*| \leq |M_3^*|$. When $\epsilon = 0$ the moment tensor corresponds to a pure DC source. If $\epsilon < 0$ or $\epsilon > 0$ the moment tensor corresponds to a compressive or tensile source, respectively (Vavrycuk, 2001). For a faulting source that deviates from the pure DC behaviour, the angle that the slip vector makes with the fault's plane can be estimated with (Dufumier and Rivera, 1997)

$$\alpha = 90^\circ - \arcsin \left(3 \frac{M_{max}^* + M_{min}^*}{M_{max}^* - M_{min}^*} \right), \quad (2.7)$$

where M_{max}^* and M_{min}^* are the maximum and minimum deviatoric eigenvalues (not considering the absolute value), respectively. This thesis follows Vavrycuk (2001) for the decomposition of moment tensors into percentages of isotropic, CLVD and DC components. The formulae is

$$\begin{aligned} \%ISO &= \frac{1}{3} \frac{\text{tr}(\mathbf{M})}{|M_{|max|}|} 100\% \\ \%CLVD &= 2\epsilon(100\% - |\%ISO|), \\ \%DC &= \%100 - |\%CLVD| - |\%ISO| \end{aligned}$$

where $M_{|max|}$ is the eigenvalue of \mathbf{M} with largest absolute value. The values of $\%ISO$ and $\%CLVD$ are negative for compressive sources and positive for tensile sources.

Ambiguity in the orientation of the fault's plane for pure shearing sources

The moment tensor for a pure double-couple in an isotropic medium is given by (Aki and Richards, 2009, equation 3.22)

$$M_{jk} = \mu A(u_k \nu_j + u_j \nu_k), \quad (2.8)$$

where A is the fault's plane area, \mathbf{u} is the displacement vector on the fault's surface (slip vector), and $\boldsymbol{\nu}$ is the normal to the fault's plane. The contributions from \mathbf{u} and $\boldsymbol{\nu}$ in 2.8 are symmetric, in other words, they are indistinguishable from each other in \mathbf{M} . This results in an ambiguity where two fault geometries with coinciding slip and normal vectors generate the same observed displacements (Jost and Herrmann, 1989) (see Figure 2.3b).

2.1.2 The Green's function of an elemental force couple

Using the seismic moment tensor, the displacement field due to a point source in a general elastic medium is estimated through (Aki and Richards, 2009, equation 3.23)

$$u_i(\mathbf{r}, t) = M_{jk} \left[s(t) * \frac{\partial}{\partial \xi_k} G_{ij}(\mathbf{r}, t; \boldsymbol{\xi}, 0) \right]. \quad (2.9)$$

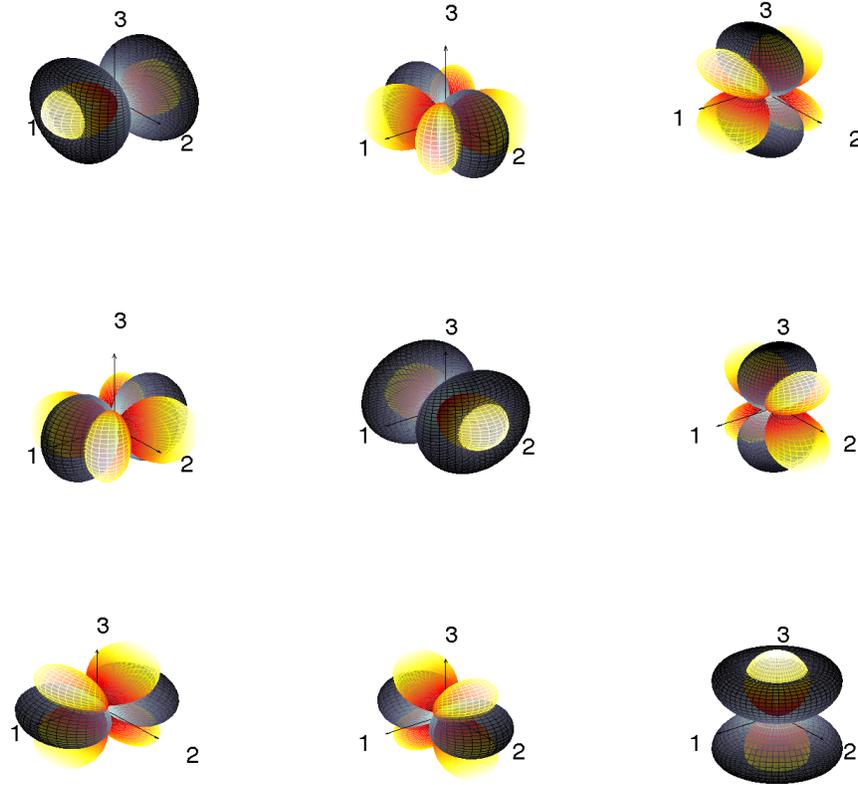


Figure 2.4: Radiation pattern of P (light colors) and S (dark colors) waves for each of the elemental force couples in the seismic moment tensor.

The variable $G_{ij}(\mathbf{r}, t; \boldsymbol{\xi}, 0)$ are Green's functions for each of the elemental force couples in the moment tensor. Green's functions are solutions to the elastodynamic wave equation for each elemental force couple considering an impulsive source. In the homogeneous, isotropic case, the Green's functions denote the radiation patterns of the moment tensor elements (equation 2.3, Figure 2.4). In a general elastic medium, Green's functions are estimated using wave modeling algorithms. Two different methodologies are employed in this thesis: one considers ray tracing times and amplitudes corrected by the source radiation pattern, and the second uses a reflectivity method (Randall, 1994) that provides full waveform solutions.

2.2 Hydraulic fracturing

Hydraulic fracturing is nowadays a widely used technology in the oilfield industry. Most authors refer the first commercial application in this field back to the late 1940s (Martinez et al., 1987). The fast growing implementation in the oilfield industry is attributed to

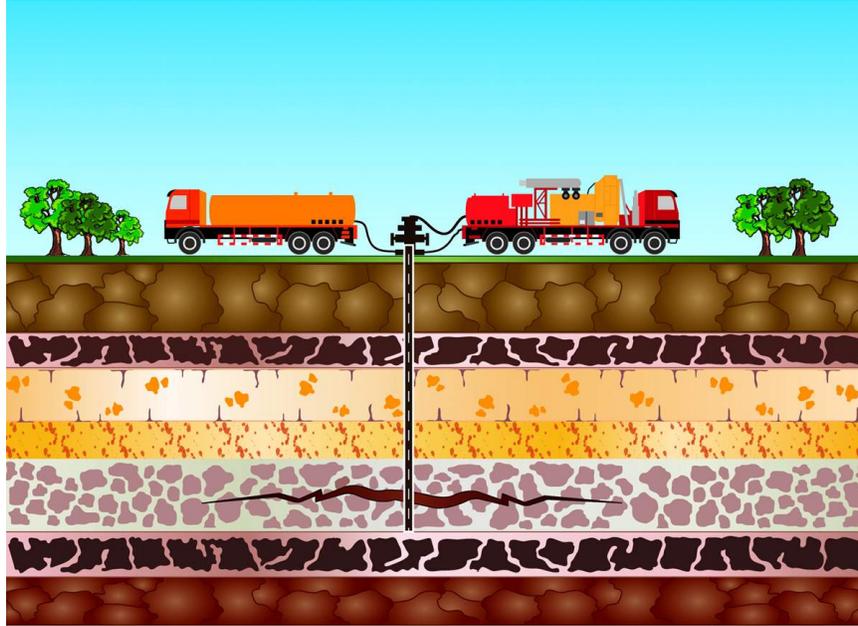


Figure 2.5: Schematic deployment of a hydraulic fracturing job. (Image created by Cecilia Camet)

the complexity of the producing formations currently under exploitation (e.g., tight gas reservoirs and shale gas, see Neale et al., 2009). The main purpose of hydraulic fracturing is to improve productivity by artificially increasing the permeability of a producing formation (Martinez et al., 1987) (Figure 2.5). The process generally starts with a mini-frac job, from which the engineering parameters (e.g., treating pressure) for the main fracture are determined. For both, the mini-frac and the main fracture, fluids at high pressure are injected into the borehole. When the pressure of the injected fluids is larger than the sum of the minimum stress in the formation plus the rock's tensile strength, the formation breakdown starts (Fjaer et al., 2008). After the fracture has been initiated, the fluids are kept flowing at high rates to prevent closure. The injected fluids also contain a solid phase or proppant. These solids are expected to pack inside the fracture to keep it open after the fluid pressure is dropped, thus creating a permanent high permeability zone in contact with the reservoir. The designing of a fracturing job requires the modeling of the entire process. A lot of effort has been put at this respect due to the increase in the use of non-vertical producing wells and the development of non-conventional fields, where the behaviour of the fracture's growth is more difficult to predict (Carter et al., 2000). The physics of the hydraulic fracturing process can be described making use of poroelastic theory. Two different sets of constitutive laws give form to poroelasticity: fluid flow (diffusion) and rock mechanics (Detournay and Cheng, 1993). An extensive knowledge of the properties of the

system (fluid and rock) is necessary in order to model the initiation and propagation of a hydraulic fracture. Realistic 3D fracture modeling must solve the complete set of constitutive poroelastic equations in a fully coupled manner (e.g., Yamamoto et al., 2000; Ji et al., 2009; Hsu et al., 2012).

Once a fracture has been designed and put into practice it is important to retrieve information about the effective fracture's geometry (volumetric distribution and azimuth) and its efficiency (productivity). This information is of fundamental interest, since it allows engineers to assess the initial fracture's model and if needed to introduce the necessary adjustments for future jobs, e.g., type of fluid, volumes of fluid and proppant, placement of producing and injecting wells, distance between wells, etc. Between the different techniques to measure the properties of an induced fracture can be mentioned: treatment pressure analysis, tiltmeter mapping and microseismic monitoring (LeCampion and Jeffrey, 2004). It is clear from Figure 2.6, that between all available alternatives for fracture mapping, microseismic monitoring is one of the methodologies that provide the most complete and reliable information.

2.2.1 Microseismic monitoring

The monitoring of acoustic emissions is a technique that has been used in geotechnical studies for almost 70 years (Hardy, 1981). Despite that, its implementation in the oil and gas industry as the microseismic monitoring technique has come until relatively recently (Maxwell et al., 2010). The first work where the potential of microseismic monitoring for the oil and gas industry was recognized is that of Albright and Pearson (1982). From this early study, important characteristic properties of hydraulically induced microseismic events were also identified, such as low local magnitudes in the order of -6 to -2 and high shear to compressional ratios in the amplitudes of the observed arrivals. Furthermore, the authors propose pore pressure increase and diffusion as the process through which preexisting planes of weakness displace in shearing mode giving origin to the microseismic events. The seismic monitoring of hydraulic fractures is a methodology that consists in the deployment of geophone arrays that "listen" to the acoustic emissions generated by the fracturing process (Figure 2.7). Usually, these geophone arrays are located downhole in nearby wells, nevertheless, recent years have witnessed the increase of microseismic monitoring from shallow and surface arrays (e.g., Kolinsky et al., 2009; Chambers et al., 2010; Duncan and Eisner, 2010). Traditionally, the most important information extracted from detected microseismic events is their location, from which the geometry of the induced fracture is interpreted (Bennett et al., 2005). The spatial distribution of the microseismic cloud has also been used to calculate a Fracture Complexity Index (FCI) (Cipolla et al., 2008). Larger FCI values correspond to fractures affecting larger rock volumes (Figure 2.8). From the magnitudes

CHAPTER 2. THEORY

Group	Diagnostic method	Main limitations	Ability to estimate							
			Length	Height	Asymmetry	Width	Azimuth	Dip	Volume	Conductivity
Direct, far-field	Surface-tilt fracture mapping	- Can't resolve individual fracture dimensions - Mapping resolution decreases with depth	Yellow	Yellow	Yellow	Red	Green	Green	Green	Red
	Downhole-tilt fracture mapping	- Resolution in fracture length and height decreases as offset well distance increases - Big picture fracture growth; not small details - No information about proppant distribution and effective or productive fracture geometry	Green	Green	Green	Yellow	Yellow	Red	Green	Red
	Microseismic fracture mapping	- Individual fracture planes can't be resolved - Adequate signal-to-noise ratio limits offset well distance - No information about proppant distribution and effective or productive fracture geometry	Green	Green	Green	Red	Green	Yellow	Red	Red
Direct, near-wellbore	Radioactive tracers	- Depth of investigation only about 1 to 2 ft from the fractured well - Provides only lower bound for fracture height if fracture and well path not aligned	Red	Yellow	Red	Yellow	Yellow	Red	Red	Red
	Temperature logging	- Thermal conductivity of different formations can vary, skewing temperature-log results - Post-treatment log requires multiple passes soon after the treatment - Provides only lower bound for fracture height if fracture and well path not aligned	Red	Yellow	Red	Red	Red	Red	Red	Red
	Production logging	- Provides only information about zones/perforations contributing to production in cased-hole applications	Red	Yellow	Red	Red	Red	Red	Red	Red
	Borehole image logging	- Only run in open hole - Provides fracture orientation and height for near-wellbore only	Red	Yellow	Red	Red	Yellow	Yellow	Red	Red
	Downhole video	- Run mostly in cased holes and only provides information about zones/perforations contributing to production (may have open hole applications)	Red	Yellow	Red	Red	Red	Red	Red	Red
	Caliper logging	- Only run in open hole - Formation wash-outs skew interpretation	Red	Red	Red	Red	Yellow	Red	Red	Red
Indirect	Net pressure fracture analysis	- Results depend on model assumptions and reservoir description - Requires calibration with direct observations	Yellow	Red	Red	Yellow	Red	Red	Red	Yellow
	Well testing	- Results dependent on model assumptions - Requires accurate permeability and reservoir pressure estimates	Yellow	Red	Red	Yellow	Red	Red	Red	Yellow
	Production analysis	- Results dependent on model assumptions - Results can also be time-dependent - Fracture conductivity difficult to determine - Requires accurate permeability and reservoir pressure estimates	Yellow	Red	Red	Yellow	Red	Red	Red	Yellow

Figure 2.6: Summary of methodologies to retrieve information about properties of hydraulic fractures. Green: can determine. Yellow: may determine. Red: cannot determine. After Cipolla and Wright (2002).

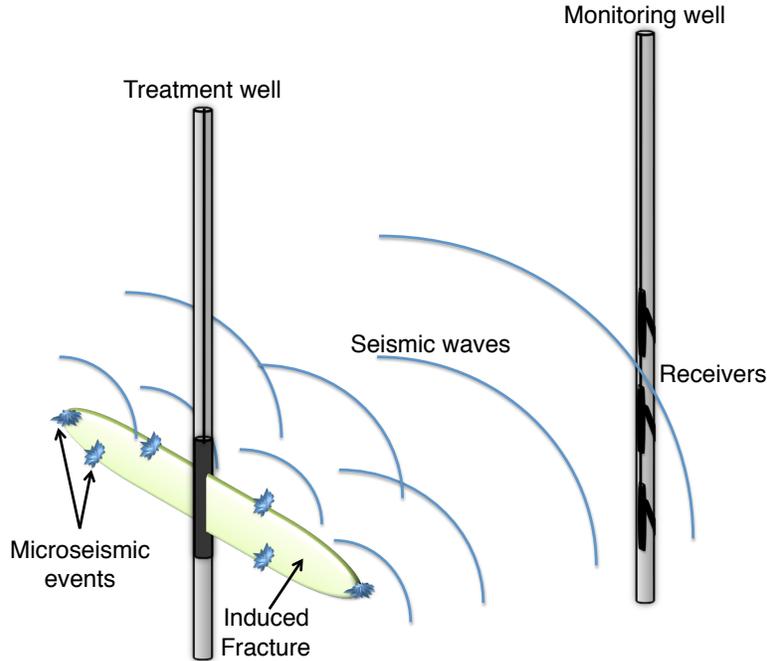


Figure 2.7: Schematic representation of a microseismic monitoring job.

of the microseismic events, graphs of cumulative seismic moment release are elaborated. Sudden changes in slope are interpreted in the context of the medium under investigation to determine possible hydraulic connection between different induced fractures, or between an induced fracture and a fault (Rutledge et al., 2004; Maxwell et al., 2008). Hydraulic connection between fractures is important to diagnose since competing fractures can potentially reduce production performance. Based on the assumption of pore-diffusion-driven microseismic events (Albright and Pearson, 1982), the rate of occurrence of microseismic locations in different directions is used to obtain an estimate of the permeability tensor of the medium (Shapiro et al., 1999). Anisotropy parameters have also been derived from microseismicity records (House et al., 1996; Verdon and Kendall, 2009). The source mechanism of microseismic events provides information to evaluate models of fracture propagation (Warpinski, 1994; Warpinski et al., 2001), to determine the state of stresses in the neighbourhood of the injection well (House et al., 1996; Rutledge and Phillips, 2003) and, in general, as a source of information to analyze the effectiveness of the induced fracture. The importance in obtaining information about fracture propagation during the injection process has been largely recognized (Maxwell et al., 2002; LeCampion and Jeffrey, 2004; Le Calvez et al., 2006), unfortunately, the estimation of source mechanisms is still usually performed in a post-processing stage.

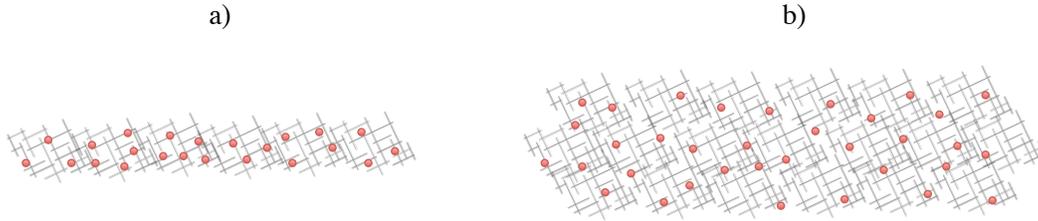


Figure 2.8: Relationship between fracture complexity index (FCI), fracture distribution (grey lines) and induced microseismic events (red dots). a) Scenario with low FCI. b) Scenario with high FCI.

2.3 Compressive Sensing

The current trend in science and technology poses a challenge in the resources required to manipulate ever increasing amounts of data. The most common strategy entails the acquisition of information about a signal at the Nyquist rate, then, the signal is expressed as a linear combination of a few elements of a base or dictionary. Since the dictionary is known, only the location and weighting coefficients of the elements that expand the signal are required for its manipulation, in other words, the signal is compressed. This strategy seems reasonable except for the fact that most of the information that was originally acquired is thrown away after compression (Figure 2.9). A more efficient procedure would require the direct identification of the weights and corresponding elements of the dictionary that expand the signal. An intermediate alternative is offered by the theory of Compressive Sensing (Candes et al., 2006; Donoho, 2006). In simple words, compressive sensing is a field of applied mathematics that studies sampling and recovery conditions for signals that admit a sparse representation in a known basis or dictionary. Two important elements in the last definition are sampling and recovery. The standard sampling protocol follows the Nyquist criterion, that states a minimum sampling rate of half the period of the highest frequency component in the signal. Compressive sensing establishes new protocols where the samples are linear combinations of the information contained in the signal. The weighting factors in these linear combinations are random numbers which do not depend on the signal itself. Hence, these new type of samples are usually referred to as non-adaptive measurements. Consider a signal given by the time series $\mathbf{y} = [y_1, y_2, \dots, y_N]$ sampled at the Nyquist rate. In order to detect the signal it is necessary a sampling operator, this is

$$\mathbf{f} = O(s(t)) = \mathbf{y}, \quad (2.10)$$

where $s(t)$ is the continuous signal and \mathbf{f} is the vector of measurements. The sampling operator $O(\cdot)$ takes a measurement of $s(t)$ at a rate in time defined by the Nyquist criterion.

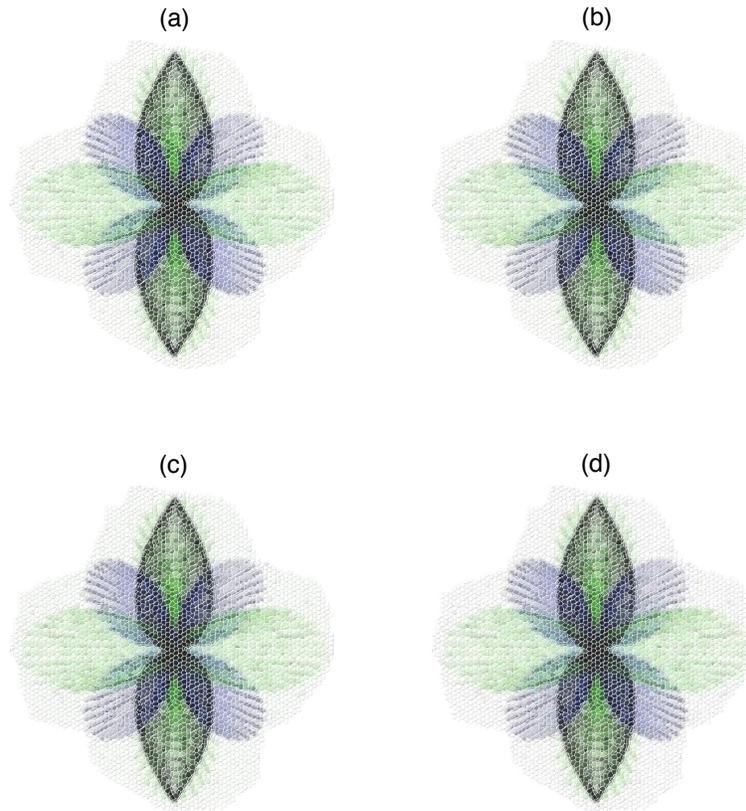


Figure 2.9: Example of image compression using the JPEG method. When displayed on the screen, each of the four images occupies a memory of 10.2 MB with a size of 1030×1044 pixels. For storage purposes the JPEG method expresses each image as a linear combination of the elements of a dictionary of cosine functions (Discrete Cosine Transform). By thresholding the coefficients of smaller magnitude different levels of storage compression are achieved. When saved in hard disk, the JPEG file only requires the location and magnitude of the coefficients that expand the image. In this example, the displayed image has been recovered from a version in hard disk of (a) 1.3 MB, (b) 620 KB, (c) 350 KB and (d) 136 KB. Although there is not perceptual loss of information from one version to another, the fully sampled image has to be known first in order to compute its Discrete Cosine Transform. Once the desired coefficients are known, the original image is thrown away.

Using compressive sensing, the measurements in vector \mathbf{f} correspond to linear combinations of the information acquired by the sampling operator. These linear combinations are obtained through the use of a sensing matrix (Φ). Sensing matrices present the characteristic of being rectangular, an important feature that introduces a dimensionality reduction in the size of the vector of measurements, in other words,

$$\mathbf{f}_{CS} = \Phi O(s(t)) = \Phi \mathbf{y}, \quad (2.11)$$

where $\Phi \in \mathbb{R}^{K \times N}$ and $K \ll N$. Equation 2.11 implies that the signal still needs to be sampled first at Nyquist rate before applying a compression through the application of Φ . In reality, technology employing compressive sensing principles incorporates the effects of Φ into the sampling operator $O(\cdot)$, so that the non-adaptive measurements are effectively all the information acquired by the instruments. In this thesis however, the application of compressive sensing follows the description in 2.11 (Figure 2.10). In order to ensure recoverability of the full signal \mathbf{y} , compressive sensing requires some conditions to be met. In the case of sensing matrices, the recoverability conditions require compliance with the Restricted Isometry Property (RIP) (Candes et al., 2006). Verifying that a matrix complies with the RIP is rather a non-trivial task, hence tools from probability theory have been resorted to determine probably good sensing matrices. Between the most recurred sensing matrix ensembles it can be counted random Gaussian and Bernoulli matrices. In the applications described in this thesis, random Gaussian ensembles are employed. In the case of the signal of interest, the recoverability conditions require the signal \mathbf{y} to be compressible or to present a sparse representation under a known dictionary. Assume that \mathbf{y} can be expanded from the linear combination of a subset of the columns of a dictionary Ψ , this is

$$\mathbf{y} = \Psi \mathbf{x}, \quad (2.12)$$

where the support of \mathbf{x} provides the location of the elements of Ψ that participate in the representation of \mathbf{y} . The size of the support of \mathbf{x} can be estimated through the ℓ_0 -norm that counts the number of non-zero coefficients. If the number of non-zero coefficients in \mathbf{x} is much smaller than the total of its elements, then \mathbf{x} is a sparse representation of \mathbf{y} under Ψ . On the other hand, if the non-zero coefficients of \mathbf{x} decay quickly in magnitude when sorted, such that \mathbf{y} can be well approximated using a small number of the largest magnitude coefficients, then \mathbf{y} is said to be compressible under Ψ . Notice that the sensing matrix only affects the dictionary under which \mathbf{y} is compressible but not its sparse representation, this is

$$\mathbf{f}_{CS} = \Phi \mathbf{y} = \Phi \Psi \mathbf{x}. \quad (2.13)$$

Hence the sparse representation \mathbf{x} is invariant for both the compressed measurements \mathbf{f}_{CS}

and the full signal \mathbf{y} . Since both Φ and Ψ are known, it does not matter whether we know \mathbf{f}_{CS} or \mathbf{y} , the representation \mathbf{x} can be determined from both of them and vice versa.

The minimum number of non-adaptive measurements K required to characterize a signal depends on its sparsity under a known dictionary. The sparser the representation, the lesser the required number of non-adaptive measurements. For this reason, the advantages of compressive sensing over the traditional Nyquist protocol exist only for signals that are compressible under a known dictionary. In the traditional practice, the signals sampled at the Nyquist rate are directly known. In the case of compressive sensing, the non-adaptive measurements are a representation in a compressed domain. In order to obtain a complete representation of the signal it is necessary to resort to recovery algorithms. The objective of the recovery algorithm is to estimate the representation of the signal under the sparsifying dictionary, then, from this representation the signal can be expanded. Common algorithms for recovery are basis pursuit (Chen et al., 1998) and orthogonal matching pursuit (Pati et al., 1993). In this thesis two different recovery methods are employed, the first of them corresponds to a modification of the Orthogonal Matching Pursuit (OMP) adapted for group sparsity (Eldar and Bolcskei, 2009) in the setting of seismic source monitoring. The second methodology is the Fast Iterative Soft-Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009) implemented with a group soft-thresholding criterion (Fornasier and Rauhut, 2008). In general, the theoretical bounds describing the limits of applicability of compressive sensing are over-restricting. The reason being that these bounds describe worst case scenarios. The presence of regular patterns in the support of sparse representations have been used in the estimation of less limiting bounds that ensure the recovery of complete signals from compressive sensing measurements (Eldar and Bolcskei, 2009; Eldar et al., 2010). In every case, there is a grey area outside the theoretical limits where the principles of compressive sensing can still be successfully implemented. The applications presented in this thesis fall within the grey area, therefore numerical modeling is employed to determine the feasibility of application of compressive sensing.

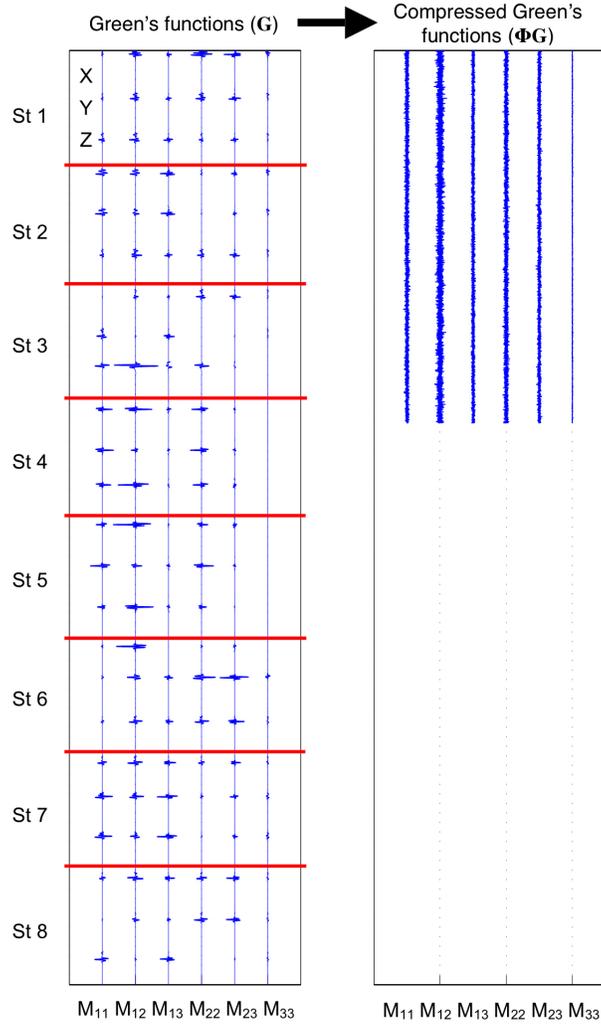


Figure 2.10: Compression of the set of six Green's functions for the three-components of eight seismic stations using the sensing matrix Φ . In this thesis, the seismic traces are first sampled modeled at Nyquist rate and then compressed. Compressive sensing technology aims at the direct acquisition of the compressed measurements on the right. The red lines denote the start and end of the samples for each seismic station. This separation cannot be done in the compressed measurements.

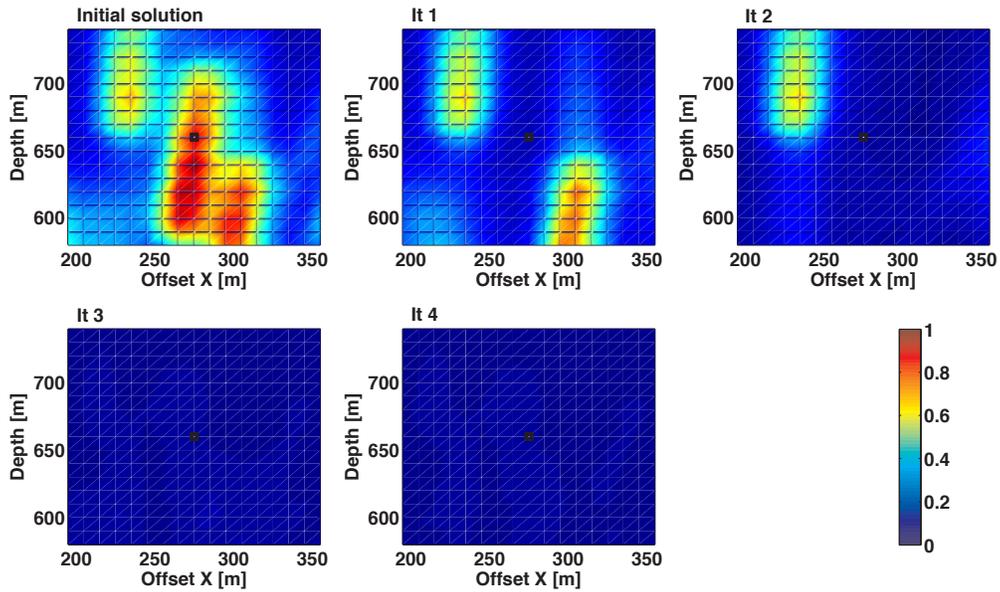


Figure 2.11: Example of a matching pursuit methodology combined with group sparsity constraints in the setting of source monitoring. Every panel represents a metric related to the source strength of virtual sources located at each of the nodes of a grid. In the initial solution the virtual source with largest strength is identified and reclassified as an actual source, the seismic wavefield due to this source is forward modeled and extracted from the original observations, then the source strength metric is estimated again to obtain the next panel (It 1). Following this procedure iteratively three virtual sources are reclassified as actual sources after the third panel (It 2). The last two panels display only residuals (It 3 and It 4). After Vera Rodriguez et al. (2010a)

CHAPTER 3

Resolution analysis of the seismic moment tensor from a borehole receiver array ¹

3.1 Introduction

Restrictions in acquisition geometry represent a considerable challenge for daily applications of moment tensor inversions in microseismic monitoring. While microseismic monitoring from multiple wells and/or surface receivers has been adopted (Bleakly et al., 2007; Kolinsky et al., 2009; Chambers et al., 2010; Duncan and Eisner, 2010), the single-well approach remains the most common practice. Like most inverse problems, an essential metric in moment tensor determination is resolvability. For a given acquisition geometry and seismic noise level, resolvability analyses assess if unknown parameters can be accurately retrieved from an inversion (Menke, 1989). Assessments of resolvability are routinely provided in, for example, studies of free oscillations (Wiggins, 1972), body and surface wave tomography (Trampert, 1998; Rao et al., 2006; Xia et al., 2008), and Amplitude-Versus-Offset (AVO) (Jing and Rape, 2004). Dufumier and Rivera (1997) estimated the resolvability of the isotropic component of the seismic moment tensor for surface stations. Trifu and Shumila (2002) investigated the accuracy of inverted source mechanisms, which is crucial for assessing risks associated with mining activities, based on statistical reliability. Resolvability of the full moment tensor for data acquired during microseismic monitoring surveys has been discussed in both single (Nolen-Hoeksema and Ruff, 2001; Vavrycuk, 2007) and multiple (Vavrycuk, 2007) vertical wells. For the case of a single vertical well, the array of receivers is assumed to reside within a plane containing the location of a given source (which I refer

¹A version of this chapter has been published. Vera Rodriguez, I., Gu, Y. and Sacchi, M., 2011. Bulletin of the Seismological Society of America 101: 2634-2642.

to as the observational plane). In isotropic and certain anisotropic (e.g., Vertical Transverse Isotropic) media, only 5 of the 6 elements of the moment tensor can be independently determined (Vavrycuk, 2007). The same restriction applies for a well that deviates from the vertical (deviated well) in the direction of wave propagation. In short, Nolen-Hoeksema and Ruff (2001) and Vavrycuk (2007) provide a blueprint for the inversion of the moment tensor in microseismic monitoring applications.

This study examines the resolvability of full moment tensor inversions under single-well monitoring geometries. In vertical geometries the experiments demonstrate that solutions of the moment tensor are only correctly represented in a reference system with two of its axes constrained to the observational plane (for short, the aligned system). Rotation of the tensor elements to a different reference system requires correct values for the dipole perpendicular to the observational plane; errors will propagate to two other moment tensor elements otherwise. Through a careful analysis on non-planar source-receiver geometries and condition numbers, I aim to provide quantitative criteria for resolving 5 and 6 moment tensor elements in vertical and deviated wells, respectively.

3.2 Parametrization

For an elastic homogeneous medium, the far field displacement at a distant station can be expressed as a function of the seismic moment tensor (Shearer, 1999, equations 9.13 and 9.17),

$$u_{i(P|S)}(\mathbf{r}, t) = \left[\frac{1}{4\pi\rho c^3} \right] \left[\frac{1}{r} \right] R_{ijk(P|S)} \dot{M}_{jk} \left(t - \frac{r}{c} \right), \quad (3.1)$$

where u_i is the displacement recorded on the i -th component of a receiver at a specific time t and position \mathbf{r} , and $(P|S)$ refers to compressional or shear waves. In this formulation ρ is the density of the medium, c is the speed of P or S waves, and r is the distance between the source and receiver. R_{ijk} represents the radiation pattern defined by the jk -th element of the moment rate tensor \dot{M}_{jk} along the i -th component of a receiver. It is also adopted the reference system of Aki and Richards (2009), where the x , y and z components point North, East, and downward, respectively. Integrating both sides of equation (3.1) over the source duration (τ), we have

$$\int_{\tau} u_{i(P|S)} dt = \left[\frac{1}{4\pi\rho c^3} \right] \left[\frac{1}{r} \right] R_{ijk(P|S)} \int_{\tau} \dot{M}_{jk} \left(t - \frac{r}{c} \right) dt, \quad (3.2)$$

where

$$\int_{\tau} \dot{M}_{jk} \left(t - \frac{r}{c} \right) dt = M_{jk}. \quad (3.3)$$

In equation (3.3), M_{jk} is the jk -th element of the seismic moment tensor of the source. Equation (3.2) can be written in matrix notation as

$$\mathbf{d}_{(P|S)} = \mathbf{K}_{(P|S)} \mathbf{X}_{(P|S)} \mathbf{m} = \mathbf{A}_{(P|S)} \mathbf{m}, \quad (3.4)$$

where $\mathbf{d}_{(P|S)}$ are the integrals of the displacement at each receiver component arranged in column vector form, $\mathbf{K}_{(P|S)}$ are constants that depend on the medium properties and $\mathbf{X}_{(P|S)}$ are matrices containing the radiation pattern. In this expression the six independent elements of the moment tensor have been arranged in the column vector $\mathbf{m} = [M_{11}, M_{12}, M_{13}, M_{22}, M_{23}, M_{33}]^T$. For P- and S- waves recorded by multiple receivers, one can rewrite equation (3.4) to a more compact form,

$$\mathbf{d} = \mathbf{A} \mathbf{m}. \quad (3.5)$$

The objective is to retrieve the seismic moment tensor \mathbf{m} from observations \mathbf{d} . The least squares solution to (3.5) requires the inverse of the matrix $\mathbf{H} = \mathbf{A}^T \mathbf{A}$. The following sections provide detailed analyses on the stability conditions of the above inverse problem.

3.3 Resolution under single vertical monitoring well geometries

3.3.1 Resolution matrix

I analyze the resolvability of the inverse problem presented in the previous section by using a model resolution matrix (Van Rijssen and Herman, 1991; Dufumier and Rivera, 1997; Jing and Rape, 2004; Vera Rodriguez and Sacchi, 2009) for single vertical monitoring well geometries (Figure 3.1). The symmetric matrix \mathbf{H} admits the eigendecomposition

$$\mathbf{H} = \mathbf{Q} \mathbf{L} \mathbf{Q}^T,$$

where \mathbf{L} is a diagonal matrix of the eigenvalues of \mathbf{H} and \mathbf{Q} contains the corresponding eigenvectors. The positive semi-definite nature of \mathbf{H} ensures that its eigenvalues are real and non-negative (Anton and Busby, 2003). The pseudoinverse of \mathbf{H} can be written as

$$\mathbf{H}^\dagger = \mathbf{Q} \mathbf{L}^\dagger \mathbf{Q}^T,$$

where \mathbf{L}^\dagger is the pseudoinverse of \mathbf{L} with the positive eigenvalues of \mathbf{H} replaced by their reciprocals (Horn and Johnson, 1985). The resolution of the system in equation (3.5) is estimated from the pseudoinverse of matrix \mathbf{H} ,

$$\hat{\mathbf{m}} = (\mathbf{H})^\dagger \mathbf{A}^T \mathbf{d}.$$

For noise free data, we substitute (3.5) in the last expression and obtain

$$\hat{\mathbf{m}} = \mathbf{H}^\dagger \mathbf{H} \mathbf{m}.$$

The product $\mathbf{H}^\dagger \mathbf{H}$ is the model resolution matrix (Aster et al., 2005), that is,

$$\mathbf{R} = \mathbf{H}^\dagger \mathbf{H}.$$

Matrix \mathbf{R} becomes the identity matrix when $\mathbf{H}^\dagger = \mathbf{H}^{-1}$, when all the model components are linearly independent and accurately determined (i.e., $\hat{\mathbf{m}} = \mathbf{m}$). If $\mathbf{H}^\dagger \neq \mathbf{H}^{-1}$, then \mathbf{R} provides important information on the resolvability of the model elements. For example, the existence of non-zero off-diagonal components ($R_{ij} \neq 0$) implies a linear correlation between model parameters m_i and m_j .

Numerical test

The numerical simulations begin by adopting a homogeneous medium with velocities $V_p = 2500 \frac{\text{m}}{\text{s}}$, $V_s = 1400 \frac{\text{m}}{\text{s}}$ and density $\rho = 2.5 \frac{\text{g}}{\text{cm}^3}$. Eight receivers are placed in a vertical array with separations of 30 m. I vary the horizontal distance between the array of receivers and the source location (d) in steps of 10 m between 100 m and 700 m, and examine array azimuths (see Figure 3.1) from $-\pi$ to π in steps of $\frac{\pi}{24}$ radians. The geographical reference system is used in this part of the analysis, though the main outcomes are equally applicable to other fixed reference systems. The resolvability of the model parameters is insensitive to the distance between the line of receivers and source location (Figure 3.2). However, different model parameters become linearly correlated when the azimuth of the line of receivers is altered. Five out of six model parameters can be correctly estimated when the observational plane is aligned with two axes of the geographical reference system, though the dipole perpendicular to the observational plane is not fully resolvable (see also Nolen-Hoeksema and Ruff, 2001; Vavrycuk, 2007). Representation of model parameters M_{13} , M_{23} and M_{33} are proper for all θ and d , but the resolution matrix suggests undesired correlations between M_{11} , M_{12} and M_{22} when the observational plane is misaligned with the geographical reference system. These results, which are obtained in a homogeneous medium, are equally valid for layered and anisotropic media, provided that the travel paths of the seismic energy recorded at the

receivers are contained in the observational plane (Vavrycuk, 2007).

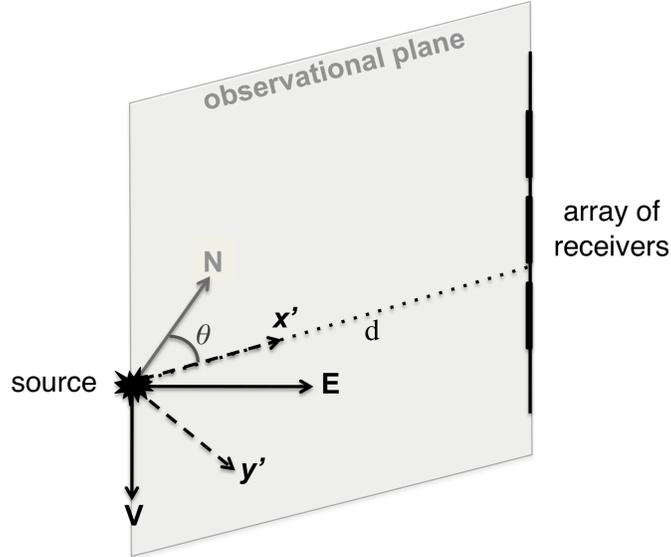


Figure 3.1: Source-receivers geometry for the microseismic monitoring experiment with one vertical monitoring well. The geographical and aligned reference systems have axes N-E-V and x' - y' -V, respectively. The axes x' and V are contained in the observational plane. The angle θ denotes the azimuth of the array of receivers (receiver azimuth) from North.

3.3.2 Maximum number of resolvable elements

It has been discussed that, without additional assumptions in the expected solutions, only 5 out of 6 independent elements of the seismic moment tensor are retrievable in single vertical monitoring well geometries. The existence of a null eigenvalue in the sensitivity matrix (\mathbf{H}) causes instability in the inversion process. The most common approach to minimize this instability is to seek deviatoric solutions (Dufumier and Rivera, 1997). This strategy does not guarantee the correctness of the solutions, however (Vavrycuk, 2007). It was first proposed that induced events during hydraulic fracturing were mainly deviatoric or double couples (Phillips et al., 1998; Rutledge and Phillips, 2003), though more recent studies have further suggested the presence of tensile events (Jechumtalova and Eisner, 2008; Sileny et al., 2009). Constrained inversions do not assist us in identifying tensile events. Furthermore, tensile events can be mistakenly fitted by deviatoric sources. The plausibility of a constrained inversion in microseismic monitoring is an important subject that requires an in-depth analysis beyond the scope of this study.

The resolution matrix shows that seismic moment tensor solutions in single vertical moni-

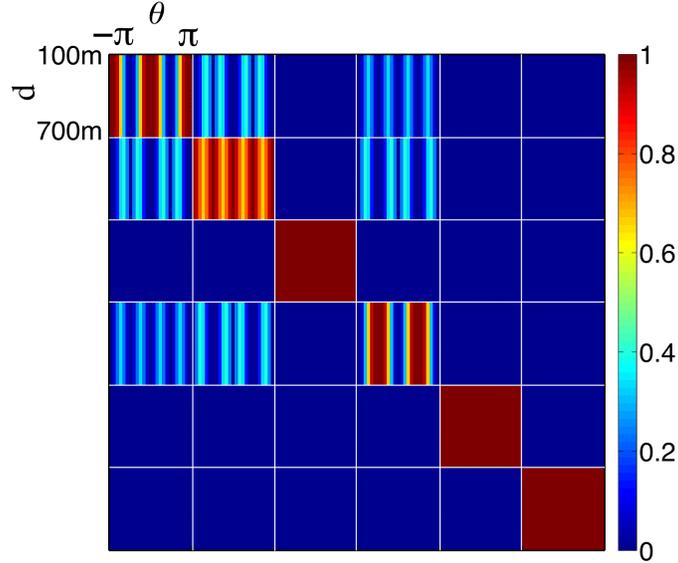


Figure 3.2: Resolution matrix for different combinations of source-receivers horizontal distance (d) and azimuth of the line of receivers (θ). Each diagonal box (delimited by white lines) represents the resolvability of the seismic moment tensor elements in the order M_{11} , M_{12} , M_{13} , M_{22} , M_{23} and M_{33} (equivalently, m_1 , m_2 , m_3 , m_4 , m_5 and m_6). The scales showed on box R_{11} apply for any other box of the matrix. Colors close to one in a diagonal box mean that the corresponding element can be solved in the inversion. Colors different from zero in off-diagonal boxes (R_{ij}) denote the linear dependency between the corresponding elements of the seismic moment tensor (m_i and m_j).

toring well geometries are only correctly represented in the aligned reference system (unless the value of the dipole perpendicular to the observational plane is known). In this study I propose to use

$$\mathbf{d} = \mathbf{A}\mathbf{\Theta}_m^{-1}\mathbf{m}_\theta, \quad (3.6)$$

where

$$\mathbf{m}_\theta = \mathbf{\Theta}_m \mathbf{m}$$

and

$$\mathbf{\Theta}_m = \begin{bmatrix} r_1^2 & -2r_1r_2 & 0 & r_2^2 & 0 & 0 \\ r_1r_2 & r_1^2 - r_2^2 & 0 & -r_1r_2 & 0 & 0 \\ 0 & 0 & r_1 & 0 & -r_2 & 0 \\ r_2^2 & 2r_1r_2 & 0 & r_1^2 & 0 & 0 \\ 0 & 0 & r_2 & 0 & r_1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

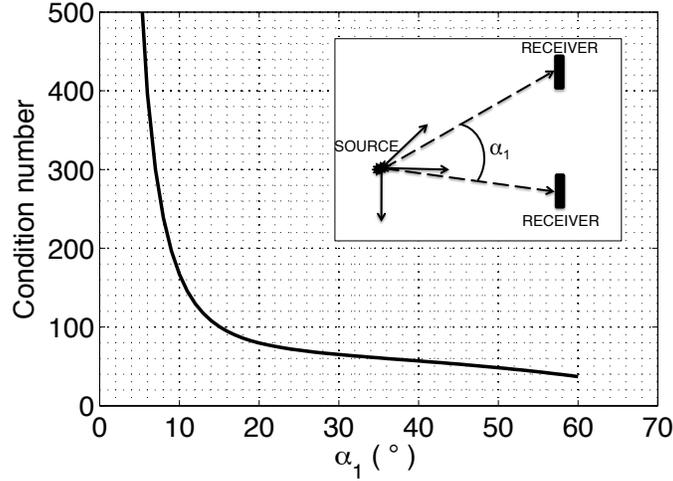


Figure 3.3: Condition number (C_ϵ) of the matrix \mathbf{H} for different angular apertures between receivers with respect to the source location (α_1).

The vector \mathbf{m}_θ denotes the elements of the seismic moment tensor expressed in the aligned reference system and Θ_m is a matrix that rotates the seismic moment tensor arranged in column vector form. For the reference system used in this study

$$r_1 = \cos \theta \text{ and } r_2 = -\sin \theta \text{ for a clockwise rotation,}$$

$$r_1 = \cos \theta \text{ and } r_2 = \sin \theta \text{ for a counterclockwise rotation.}$$

Solving the inversion using equation (3.6) eliminates the need to rotate the observations each time when a new source with a different location is introduced (Jechumtalova and Eisner, 2008). The change to the aligned system is accounted for by the matrix Θ_m^{-1} .

3.3.3 Inversion stability and acquisition design

The resolution matrix provides information about the resolvability of the model parameters in terms of the medium properties and acquisition geometry. However, it does not explicitly consider the noise content in the observations and uncertainties associated with other inversion-related parameters (e.g., inaccuracies in the medium velocity model). In this section, I evaluate the dependence of the inversion on angular aperture α_1 (Figure 3.3) and noise level for single vertical monitoring well applications.

The parameter of choice for stability evaluation is the condition number (C) of matrix \mathbf{H} (Dufumier and Rivera, 1997). Instead of using the traditional definition, I compute the

condition number of \mathbf{H} as

$$C_\epsilon = \lambda_{max}/\lambda_{min \neq 0}, \quad (3.7)$$

where λ_{max} is the largest eigenvalue of \mathbf{H} and $\lambda_{min \neq 0}$ is the smallest eigenvalue of \mathbf{H} that is different from zero. Under this condition, only eigenvalues related to the 5 solvable elements of the seismic moment tensor are considered. To compute the condition numbers the problem is formulated using equation (3.6) with the seismic moment tensor expressed in the aligned reference system. I choose a vertical array of 2 receivers and vary α_1 from 2° to 60° (see Figure 3.3). The azimuth (θ) and horizontal distance (d) of the line of receivers from the source are randomly selected. The results show that, as α_1 decreases, the condition number increases rapidly and thereby lowers the stability of the inversion (see Figure 3.3).

To illustrate the process of acquisition design, consider known arbitrary sources with their moment tensors referenced in the aligned system. I perform inversions for 5 different combinations of random noise level and angle α_1 . The problem is solved using the pseudo inverse. For a given combination, the inversion is performed 100 times and a new time series containing Gaussian noise is added to the observations in each iteration. The random noise level is defined by normalizing its maximum amplitude to the maximum amplitude of the signal from a single receiver, and then multiplying the outcome by the desired percentage. For each combination of α_1 and noise level, I compute a normalized mean squared error (*NMSE*) via (Figure 3.4)

$$NMSE = \frac{1}{100} \sum_{i=1}^{100} \frac{|\mathbf{m}_i - \hat{\mathbf{m}}_i|^2}{|\mathbf{m}_i|^2},$$

where $\hat{\mathbf{m}}_i$ is the result of the inversion and \mathbf{m}_i is the true moment tensor in the i -th realization of a noise level and α_1 . As expected, the *NMSE* of the inverted solution is zero for noise free data, but increases with decreasing α_1 when noise is added (see Figure 3.4). By setting cut-off values of 100 for C_ϵ and 0.025 for the *NMSE*, an aperture $\alpha_1 \sim 15^\circ$ is likely too small to lead to stable inversion results when the noise level exceeds 15%. This suggests that, for a line of receivers located at 100 m from the source location, the only receivers contributing to the solution of the inverse problem are those separated by more than ~ 26 m; the required separation is ~ 132 m for a line located at 500 m. A 5% perturbation in P- and S-wave speeds increases the *NMSE*, then the maximum allowable noise level for a reliable solution decreases to $\sim 10\%$ for α_1 close to 15° (Figure 3.5).

It should be noted that this acquisition design procedure requires a case-by-case analysis of the condition number C_ϵ and the *NMSE* function. The magnitudes of the condition number and the sensitivity to noise depend on the parametrization of the inverse problem. For example, Nolen-Hoeksema and Ruff (2001) adopted the same parametrization used in this study and obtained reliable solutions for condition numbers on the order of 100 – 500.

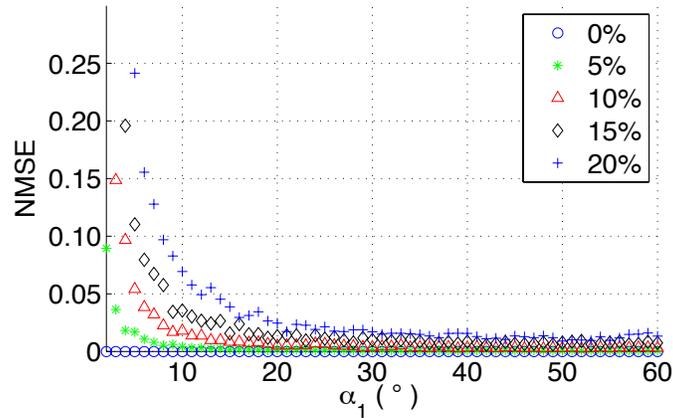


Figure 3.4: Normalized mean squared error ($NMSE$) function of the inversion for different combinations of aperture (α_1) and noise level.

On the other hand, Dufumier and Rivera (1997) introduced a different parametrization and report a much smaller cut-off value (~ 5) required for the condition number. The cut-off values for C_ϵ and the $NMSE$ depend on the level of desired accuracy of the solutions.

3.4 Minimum requirements for a full moment tensor inversion

3.4.1 Stability for non-vertical monitoring geometries

The resolution matrix becomes the identity when a receiver is placed outside a defined observational plane. This full resolvability is only true for noise-free data. For instance, the condition number of the full inversion (6 independent elements of the moment tensor) enables us to determine the minimum relative orientations and distances between receivers. For this part of the analysis, consider one receiver (R_1) at a fixed distance of 400 m from the source and two receivers (R_2 and R_3) embedded in a plane perpendicular to the straight line between R_1 and the source (Figure 3.6). For simplicity, the angle between receivers R_1 and R_2 (similarly for R_1 and R_3) will be referred to as the aperture from the source (α_1), and the angle between receivers R_2 and R_3 with vertex in R_1 will be referred to as the aperture between receivers (α_2). Assuming the same homogeneous medium, it is impractical to explore all possible combinations of receivers outside a single observational plane. However, the magnitude of the distances and orientations required by a full moment tensor solution can be assessed through numerical experiments.

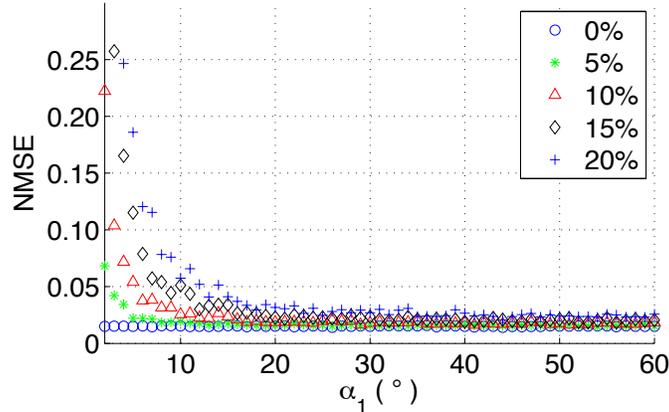


Figure 3.5: Normalized mean squared error ($NMSE$) function of the inversion for different combinations of aperture (α_1) and noise level. The velocity model is biased by an increment of 5% in both P- and S-wave speeds.

I calculate the condition number of the inversion for multiple combinations of angles α_1 and α_2 for four different acquisition geometries (Figures 3.6 and 3.7). The most stable solution space depends on factors such as the alignment of line R_1 - R_2 with the vertical axis of the reference system. A single high resolvability zone is obtained when these two receivers are not aligned (Figure 3.7-a, 3.7-b and 3.7-c); and two distinct stability zones are obtained otherwise (Figure 3.7-d). The azimuth (θ) of the fixed receiver (R_1) also appears to impact the location of maximum resolvability, especially with respect to angle α_2 (Figures 3.7-a and 3.7-b). In general, angle α_1 attains its optimal value from 40° to 55° . However, while a minimum value for the angle α_1 is required for stability conditions, large angles tend to introduce additional instability after the minimum condition number is reached. Furthermore, the optimal range for α_2 varies significantly for different geometries, despite a consistent increase in condition number near angles 0° and 180° (see Resolution under single vertical monitoring well geometries section) in all cases.

3.4.2 Sensitivity to noise and velocity perturbation

For the same scenario shown in Figure 3.7-b, I perform inversions under different combinations of random noise level and angles α_1 and α_2 . For each combination I obtain solutions for 100 realizations of random noise - moment tensor and compute their $NMSE$. As expected, higher condition numbers are associated with higher values of $NMSE$ and, for a fixed condition number the scatter in $NMSE$ increases with the noise level (Figure 3.8). These two observations are more evident when condition numbers exceed 100 (not presented here). Furthermore, background velocities have considerable influences on the inversion out-

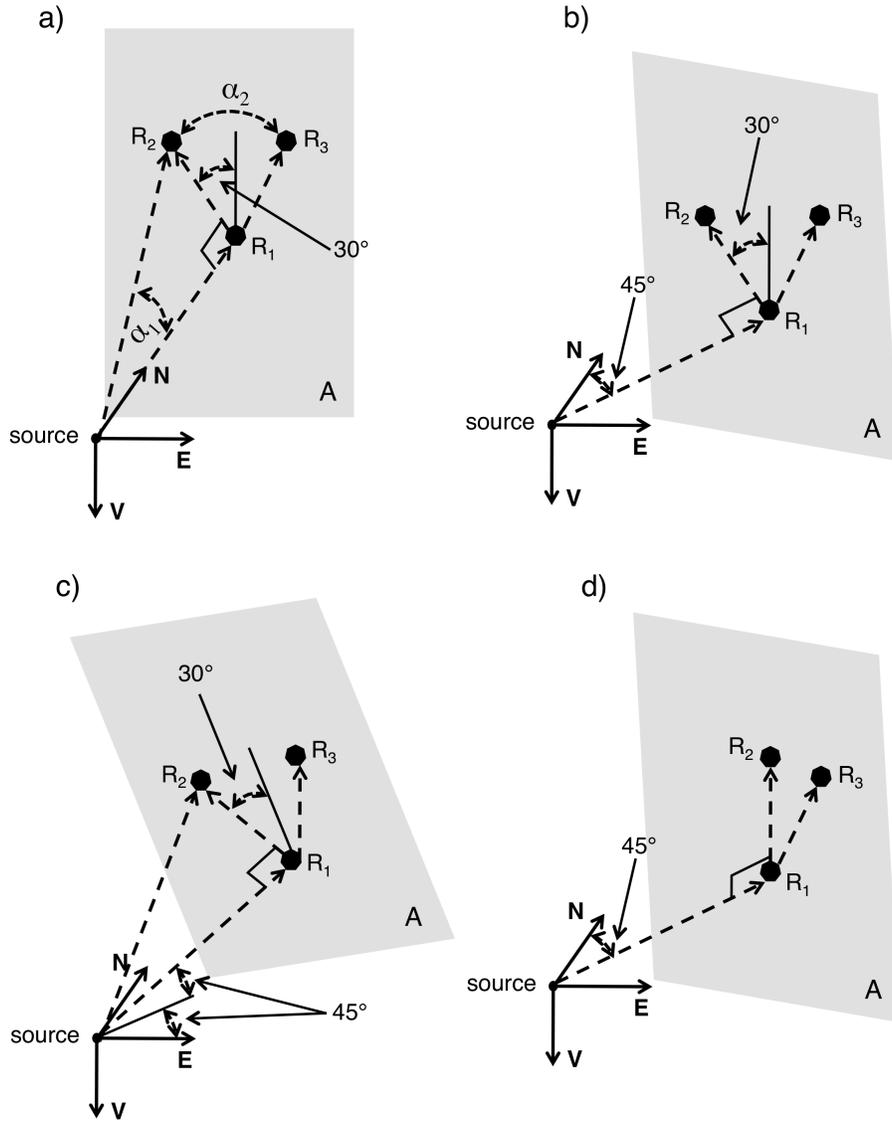


Figure 3.6: Source-receiver geometry arrangements to compute the condition number for the full source mechanism inversion. Receivers R_1 , R_2 and R_3 are contained in the plane A , which lies perpendicular to the line that joins the source location with R_1 . The angle α_1 denotes the aperture between receivers from the source location and α_2 is the aperture between receivers R_2 and R_3 with respect to R_1 . a) source location and receiver R_1 are aligned with an axis of the reference system and the three receivers form a vertical plane. b) source location and receiver R_1 are not aligned with any of the axis of the reference system and the three receivers form a vertical plane. c) source location and receiver R_1 are not aligned with the reference system and the three receivers form a non-vertical plane. d) source location and receiver R_1 are not aligned with the reference system but receivers R_1 and R_2 are aligned in the vertical direction.

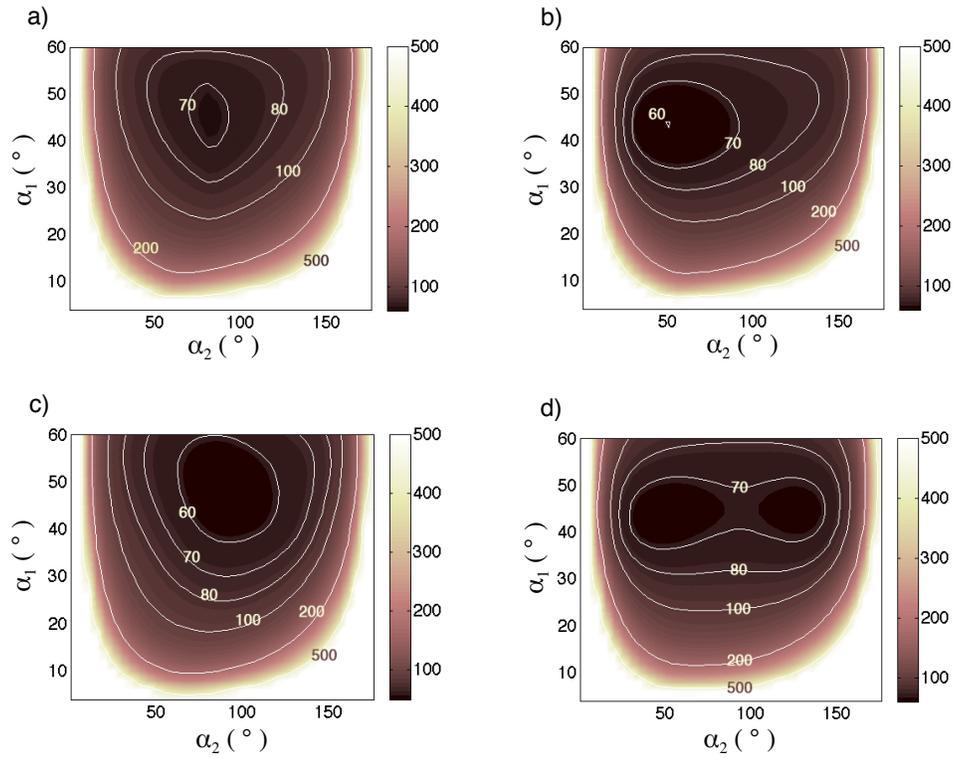


Figure 3.7: Condition number for different combinations of the angles α_1 and α_2 . The four graphs correspond to the source-receiver geometries presented in Figure 3.6. Condition numbers above 600 are set at 600 for better visualization.

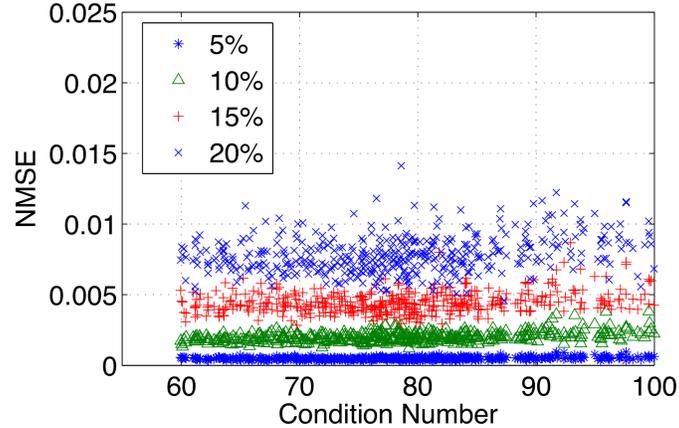


Figure 3.8: Normalized mean squared error ($NMSE$) function for all combinations of α_1 and α_2 with condition number < 100 in Figure 3.7-b. The percent value represents noise level.

come as, for instance, a 5% perturbation in P and S wave speeds increases the scatter and magnitudes of the $NMSE$ by a factor of 10 (Figure 3.9).

3.4.3 Full moment tensor inversion from a single deviated monitoring well

For microseismic monitoring applications, it is critical to identify the combinations of angles α_1 and α_2 that are 1) practical in borehole experiments, and 2) able to resolve the full moment tensor in an unconstrained inversion. For one of the examples presented in Figure 3.6, I have shown that various source-receivers geometries can lead to reliable solutions, with condition numbers < 80 for a cutoff $NMSE$ value of 0.025 and a noise level $< 15\%$ (see Figures 3.7-b and 3.9). One combination that also meets the two aforementioned objectives is $\alpha_1 = 45^\circ$ and $\alpha_2 = 130^\circ$ (see Figure 3.7-b). For a fixed receiver located 400 m from the source location, the angle α_1 translates to a distance of 400 m between receivers (Figure 3.10). Under this configuration, these three receivers form a smooth trajectory that does not exceed a dogleg severity (DLS) of $\sim 6^\circ/30$ m (Heisig et al., 2004). Current directional drilling technologies allow the perforation of wells with this degree of curvature (Bryan et al., 2009), and hence the full moment tensor can potentially be extracted from a carefully selected deviated monitoring well.

Monitoring from a deviated well requires information on the volume over which the inversion is stable. Consider the well trajectory from Figure 3.10 and an acquisition design with 9 receivers, the 9 receivers are distributed in 3 groups of 3 with the central receiver from each

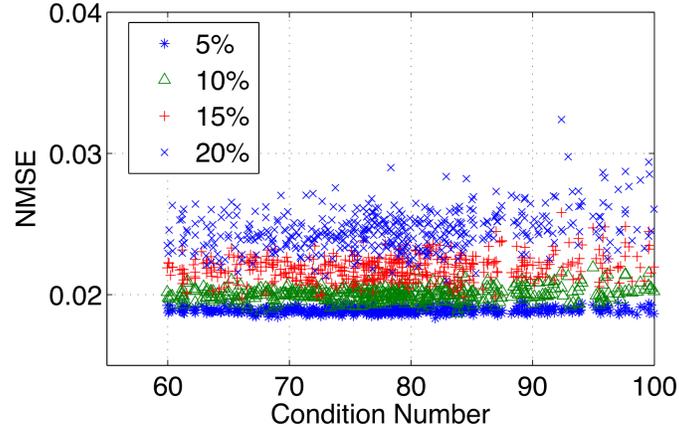


Figure 3.9: Normalized mean squared error ($NMSE$) function for all combinations of α_1 and α_2 with condition number < 100 in Figure 3.7-b. The percent value represents noise level. In this case, the P- and S-wave speeds are biased by 5%.

group at the locations showed in Figure 3.10. The remaining two receivers in each group are separated by 5 m from the center. For this monitoring geometry, an isosurface for a condition number of 80 (Figure 3.11) exhibits perfect symmetry between the two sides of the well. A relatively complex stability volume is obtained where the approximate spatial limits of x , y , z are (100 m, 500 m), (200 m, -500 m) and (1500 m, 1900 m), respectively. Depending on the location of the induced fracture(s), this monitoring geometry could facilitate moment tensor inversions of at least a significant fraction of induced microseismic events.

3.5 Summary

This chapter investigates two important problems associated with microseismic monitoring. First it is reviewed the case when the travel-path trajectories are constrained to a plane, where only 5 of the 6 elements of the seismic moment tensor are recoverable. The resolution matrix provides a broader overview of the moment tensor resolvability displaying linear correlation between the elements of the moment tensor. According to the simulations, the error due to an incorrect constraint on the dipole perpendicular to the observational plane can only be propagated to two further elements of the moment tensor. On the other hand, all the eigenvalues associated with the sensitivity matrix differ from zero when one or more receivers fall outside the observational plane. This indicates that, in theory, one can resolve all 6 elements of the seismic moment tensor. However, due to the presence of noise and small eigenvalues, the resulting solutions can be unstable. In this case the analysis based on the condition number of the sensitivity matrix can be highly advantageous. The condition

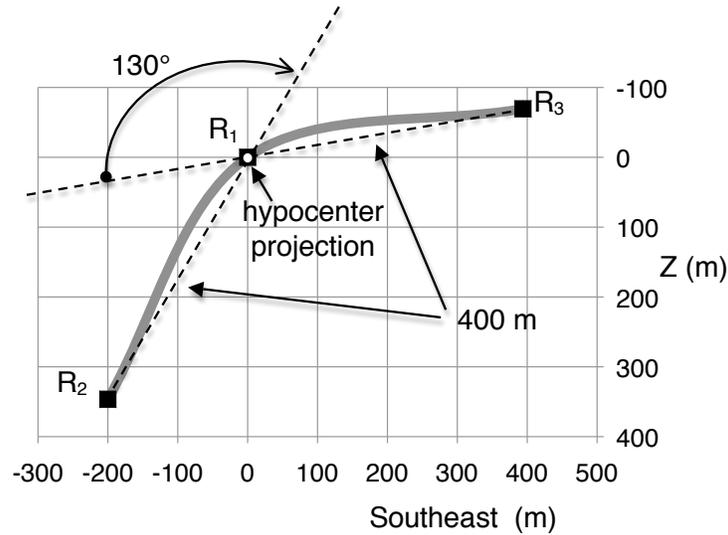


Figure 3.10: Example of a receivers configuration that can solve the full seismic moment tensor inversion for the given source projection (white dot). Distances are measured over the plain that contains the receivers. The source location is at 400 m in the perpendicular direction to this view. The grey line represents a smooth trajectory joining the three receivers.

number analysis enables us to access levels of resolvability for a given acquisition layout. The inversion of the full moment tensor with single borehole data is feasible in situations when the array of receivers is deployed in a deviated well. For instance, the simulations in homogeneous media showed that deviated wells with curvature (dogleg severity) on the order of $6^\circ/30$ m can be used to retrieve the full moment tensor. For case studies involving abrupt changes in velocity and/or anisotropy, the condition number analysis can also be an important indicator of the resolvability of the seismic moment tensor prior to monitoring tests.

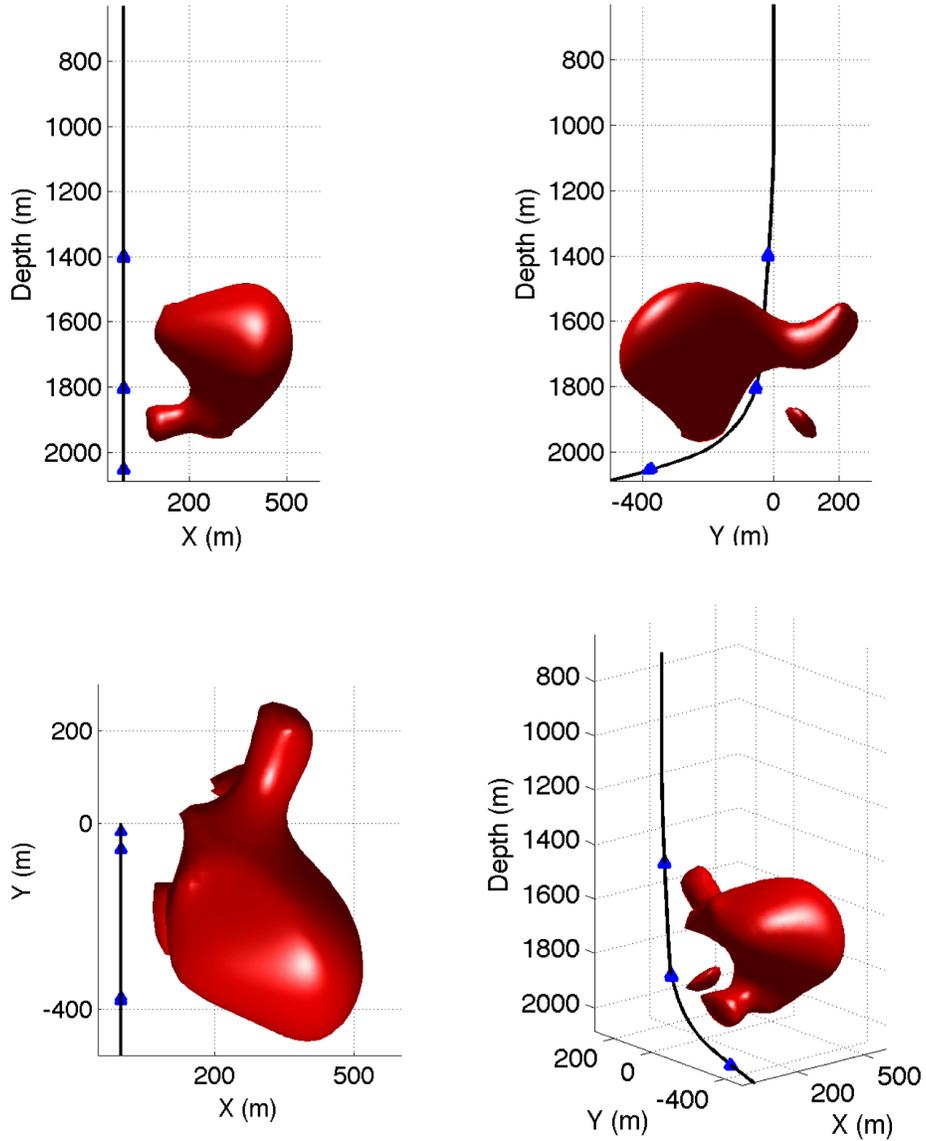


Figure 3.11: Receiver distribution (blue triangles) along an ideal deviated well (black line) that can solve a full moment tensor inversion. An isosurface value of 80 (condition number) is selected for this plot.

CHAPTER 4

Simultaneous recovery of origin time, location and seismic moment tensor using sparse representation theory ¹

4.1 Introduction

Fast and accurate recovery of seismic source parameters (location, origin time and seismic moment tensor) is a problem of interest in different areas of geophysics. For earthquake monitoring, fast source parameters are routinely estimated for large events by institutions like the Global Centroid Moment Tensor project (Dziewonski et al., 1981), the United States Geological Survey (Sipkin, 1982), and the Earthquake Research Institute of Japan (Kawakatsu, 1995). On a regional scale, fast inversions are also performed close to seismically active areas (e.g., Dreger and Helmberger, 1993; Pasyanos et al., 1996; Dreger, 2003; Clinton et al., 2006; Bernardi et al., 2004; Scognamiglio et al., 2009). In mining applications, quick estimations of the source parameters are important for hazard mitigation during mining operations (Trifu and Shumila, 2002; Gibowicz, 2009). In the monitoring of hydraulic injections in the oil industry, the source parameters have been used to evaluate potential adjustments to the injection process (Maxwell et al., 2002; LeCampion and Jeffrey, 2004). Fast and accurate source parameter estimations are also important for monitoring fluid injections in geothermal areas (Julian et al., 2010), and volcano activity (McNutt, 1996; Foulger et al., 2004). Most inversion algorithms employed in the aforementioned applications do not provide simultaneous estimations of all three source parameters. Information about origin time and source location is determined by independent systems and/or algorithms, with the source moment tensor being estimated in a secondary step. Therefore, an inherent lag time is required, thus preventing efficient, simultaneous estimates of the three source parameters.

¹A version of this chapter has been published. Vera Rodriguez, I., Sacchi, M. and Gu, Y. 2012. *Geophysical Journal International* 188: 1188-1202.

In this chapter I present a method for the simultaneous recovery of the three source parameters (origin time, location and seismic moment tensor). This "waveform fitting" method is a modification of the Block Orthogonal Matching Pursuit (BOMP) from sparse representation theory (Eldar et al., 2010). The modified BOMP proposed in this chapter permits the recovery of optimal solutions in highly coherent dictionaries, which is the most common scenario for seismic source monitoring applications. Advantages of the proposed method with respect to other existing approaches include: 1) no *a priori* information about event occurrence is required, 2) potential for continuous, automatic, real-time performance, 3) simultaneous sources can be inverted from the same set of observations, and 4) the method's suitability to perform in a particular scale of application is controlled by the availability of appropriate Green's functions to "match" the observations.

The chapter is organized as follows. Section 4.2 reviews important definitions from sparse representation theory and the forward parametrization of the source inversion. Section 4.3 presents my two main contributions: the link between the source parameter inversion and sparse representation theory, and a methodology for the recovery of the source parameters based on BOMP. Section 4.4 presents applications of the proposed methodology to both synthetic and real cases. The synthetic case corresponds to the setting of a microseismic monitoring experiment. The real dataset is an earthquake scenario using records from the 18 June 2002 Caborn Indiana earthquake. Finally, section 4.5 is devoted to the summary of this study.

4.2 Theoretical background

4.2.1 Block sparsity

The discussion starts by introducing some definitions from sparse representation theory that apply to general signals. The description that follows is rather general, hence the reader is referred to the articles cited within the text for further details. Later in section 4.3, the connection between this general theory and the problem of interest is presented. Suppose a discrete signal $\mathbf{y} \in \mathbb{R}^L$ is a linear combination of a subset of the energy-normalized functions $\psi_j \in \mathbb{R}^L$, $j = 1, 2, \dots, N$, this is,

$$\mathbf{y} = \Psi \mathbf{x}, \quad (4.1)$$

where $\Psi = [\psi_1, \psi_2, \dots, \psi_N] \in \mathbb{R}^{L \times N}$, and $\mathbf{x} \in \mathbb{R}^N$ is a vector containing the coefficients representing the signal. By this definition, a coefficient x_j with zero value corresponds to a function ψ_j that does not participate in the representation of \mathbf{y} . Vector \mathbf{x} is said to be S -sparse (Baraniuk, 2007; Candes and Wakin, 2008) if $\|\mathbf{x}\|_0 \leq S \ll N$, where the symbol $\|\cdot\|_0$ refers to the ℓ_0 -norm of the variable between vertical bars. The ℓ_0 -norm is

defined as the count of the non-zero elements in the variable under consideration (e.g., if $\mathbf{x} = [1, 0, 3, 0, 0, -5, 0, 4]$, then $\|\mathbf{x}\|_0 = 4$) (Mallat, 2008). In the signal processing community, matrix Ψ and its columns are generally referred to as dictionary and atoms, respectively (Mallat and Zhang, 1993). The dictionary is overcomplete or redundant if $L < N$, which also corresponds to an underdetermined system of equations (Mallat, 2008). A vector \mathbf{x} exhibits block sparsity if its non-zero coefficients occur in clusters (Eldar and Bolcskei, 2009). Under this consideration, the vector \mathbf{x} can be expressed as a concatenation of blocks,

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}[1] & \mathbf{x}[2] & \cdots & \mathbf{x}[N_t] \end{bmatrix}^T, \quad (4.2)$$

where $\mathbf{x}[l] = [x_{d(l-1)+1}, x_{d(l-1)+2}, \dots, x_{d(l-1)+d}]$ is the l -th block and N_t is the total of blocks in \mathbf{x} . The length of each block is d , such that $N = N_t d$. Correspondingly, the dictionary Ψ is the concatenation of the column-blocks

$$\Psi = \begin{bmatrix} \Psi[1] & \Psi[2] & \cdots & \Psi[N_t] \end{bmatrix}, \quad (4.3)$$

where $\Psi[l] = [\psi_{d(l-1)+1}, \psi_{d(l-1)+2}, \dots, \psi_{d(l-1)+d}]$ is the l -th column-block. The vector \mathbf{x} is said to be block K -sparse if $\|\mathbf{x}\|_{2,0} \leq K \ll N_t$, where $\|\cdot\|_{2,0}$ denotes the mixed $\ell_{2,0}$ -norm (Kowalski and Torr sani, 2009; Eldar et al., 2010), that counts the number of blocks with non-zero ℓ_2 -norm. In sparse representation theory, an important property that characterizes a dictionary is the measurement of how alike are its atoms. A common metric to determine the similarity between the atoms of a dictionary is called mutual coherence (e.g., Tropp, 2004). In the case of block sparsity, two different measurements of coherence can be defined, one is the coherence within a block or sub-coherence and the second is the coherence between different blocks. The sub-coherence ν is computed as (Eldar et al., 2010, equation 8)

$$\nu = \max_l \left(\max_{i,j \neq i} |\psi_i^T \psi_j| \right), \quad \psi_i, \psi_j \in \Psi[l]. \quad (4.4)$$

On the other hand, the block-coherence is defined as (Eldar et al., 2010, equation 6)

$$\mu_B = \max_{l,r \neq l} \left(\frac{1}{d} \rho(\mathbf{M}[l,r]) \right), \quad (4.5)$$

where $\rho(\cdot)$ is the spectral norm (Strang, 2006) of the matrix $\mathbf{M}[l,r]$ given by

$$\mathbf{M}[l,r] = \Psi^T[l] \Psi[r]. \quad (4.6)$$

4.2.2 Recovery of block K -sparse solutions

The recovery of a block K -sparse solution \mathbf{x} from a set of observations \mathbf{y} is accomplished through modifications to Basis Pursuit (BP) (Chen et al., 1998) and Orthogonal Matching Pursuit (OMP) (Pati et al., 1993) algorithms. I further propose a source parameter inversion method based on Block Orthogonal Matching Pursuit (BOMP) (Eldar et al., 2010). BOMP is a greedy algorithm that works iteratively selecting one by one the blocks of the dictionary that participate in the representation of the signal. Using least squares criteria, BOMP selects after each iteration the block of the dictionary that provides the best improvement in quality for the signal representation. The details of BOMP are as follows,

Algorithm 1 Block Orthogonal Matching Pursuit (BOMP)

1. Initialize variables

residual $\mathbf{r}_0 = \mathbf{y}$
 index set $\Lambda_0 = \emptyset$
 counter $l = 1$

2. Identify the block that best matches the current residual

$n_l = \arg \max_n \|\Psi^T[n]\mathbf{r}_{l-1}\|_2^2$
 $\Lambda_l = \Lambda_{l-1} \cup n_l$

3. Estimate the best approximation coefficients with the blocks chosen so far

$\{\mathbf{x}_l[n]\}_{n \in \Lambda_l} = \arg \min_{\{\mathbf{a}[n]\}_{n \in \Lambda_l}} \|\mathbf{y} - \sum_{n \in \Lambda_l} \Psi[n]\mathbf{a}[n]\|_2^2$

4. Update the residual and iterate

$\mathbf{r}_l = \mathbf{y} - \sum_{n \in \Lambda_l} \Psi[n]\mathbf{x}_l[n]$
 $l = l + 1$

repeat 2 - 4 until the stopping criteria are met.

5. Output

solution vector $\mathbf{x}[n] = \mathbf{x}_l[n]$ for $n \in \Lambda_l$ and $\mathbf{x}[n] = 0$ otherwise.

Using BOMP and considering a model as in equation (4.1), a block K -sparse representation \mathbf{x} is uniquely recovered if (Eldar et al., 2010, Theorem 3),

$$Kd < \frac{1}{2} \left(\frac{1}{\mu_B} + d - (d-1) \frac{\nu}{\mu_B} \right). \quad (4.7)$$

In these cases, the algorithm stops when the updated residual is zero. In practical situations, however, we have models of the following form

$$\hat{\mathbf{y}} = (\Psi + \hat{\Psi})\mathbf{x} + \mathbf{n}, \quad (4.8)$$

where $\hat{\Psi}$ is a perturbation to the dictionary, \mathbf{n} is additive noise, and $\hat{\mathbf{y}}$ is a set of inaccurate observations. Under the more realistic model (4.8), one way to proceed is to seek an optimal block K -sparse approximation, i.e., the solution to

$$\min_{\|\mathbf{x}\|_{2,0}=K} \min_{\mathbf{x}} \|\hat{\mathbf{y}} - (\Psi + \hat{\Psi})\mathbf{x}\|_2^2 \quad (4.9)$$

In these cases, BOMP must be provided with a stopping criteria after K iterations. In practical scenarios, we always look for having dictionaries where the perturbation $\hat{\Psi}$ is small. If $\hat{\Psi}$ is negligible, Donoho et al. (2006) show that the error in the estimation of \mathbf{y} depends directly on the level of noise \mathbf{n} , and furthermore, that the optimal approximation $\hat{\mathbf{x}}$ shares the same support as the sparse representation \mathbf{x} of the noiseless signal. If $\hat{\Psi}$ is not negligible, Herman and Strohmer (2010) demonstrate both theoretically and numerically that the error in the estimation of \mathbf{y} increases linearly with the amount of perturbation introduced to the dictionary but anti-correlate with the sparsity of \mathbf{x} . Up to this point, I have described results from sparse representation theory that apply to general signals. After introducing the source parameter inversion problem in the following subsection, the connection between sparse representations and the seismic source monitoring problem will become clear.

4.2.3 Displacement field due to a seismic source

The far displacement field due to a point source is estimated using the expression (Aki and Richards, 2009, equation 3.23)

$$u_i(\mathbf{r}, t) = M_{jk}(t) * \frac{\partial}{\partial \xi_k} G_{ij}(\mathbf{r}, t; \boldsymbol{\xi}, 0), \quad (4.10)$$

where \mathbf{r} is the position of the recording station and $\boldsymbol{\xi}$ is the source location. The term $M_{jk}(t)$ is the moment tensor of the seismic source and G_{ij} are the point force Green's functions containing information about the wave propagation between \mathbf{r} and $\boldsymbol{\xi}$. The subscript i is

related to the ground motion component and the symbol $*$ denotes convolution. Expression (4.10) holds for any type of media with the appropriate use of Green's functions. Assuming that the source geometry can be separated from its time variation (Madariaga, 2007), we can write

$$u_i(\mathbf{r}, t) = M_{jk} g_{ij,k}(\mathbf{r}, t; \boldsymbol{\xi}, 0). \quad (4.11)$$

In equation (4.11), the seismic moment tensor has been split into a time invariant part M_{jk} and a source time function $s(t)$, where $g_{ij,k}(\mathbf{r}, t; \boldsymbol{\xi}, 0) = s(t) * \frac{\partial}{\partial \xi_k} G_{ij}(\mathbf{r}, t; \boldsymbol{\xi}, 0)$. Arranging terms in matrix form, equation (4.11) can also be expressed as

$$\mathbf{u} = \mathbf{G} \mathbf{m}, \quad (4.12)$$

where \mathbf{m} contains the 6 independent elements of the time invariant part of the moment tensor (from now on, the moment tensor) arranged in column vector form and the columns of matrix \mathbf{G} are the functions $g_{ij,k}$ (for short, the Green's functions). The formulation in (4.12) is developed for observations due to a single source recorded at multiple stations.

4.3 Using sparse representation theory to invert for the source parameters

The displacement field due to multiple sources can be expressed as

$$\mathbf{u} = \sum_s \mathbf{G}_s \mathbf{m}_s, \quad (4.13)$$

where the subscript s refers to a particular source. One way to implement equation (4.13) is by setting a grid over the space where seismic sources are expected to occur. Then, every node in the grid can be considered a potential source location or *virtual source* (Kawakatsu, 1998). Following this approach, equation (4.13) can be simplified to

$$\mathbf{u} = \mathbf{G} \mathbf{m}. \quad (4.14)$$

In this case, the matrix \mathbf{G} contains the six Green's functions for all node-station combinations, and vector \mathbf{m} includes the moment tensors for all the virtual sources in the grid (Fig. 4.1), i.e.

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}[1] & \mathbf{G}[2] & \cdots & \mathbf{G}[N_t] \end{bmatrix},$$

and

$$\mathbf{m} = \begin{bmatrix} \mathbf{m}[1] & \mathbf{m}[2] & \cdots & \mathbf{m}[N_t] \end{bmatrix}^T,$$

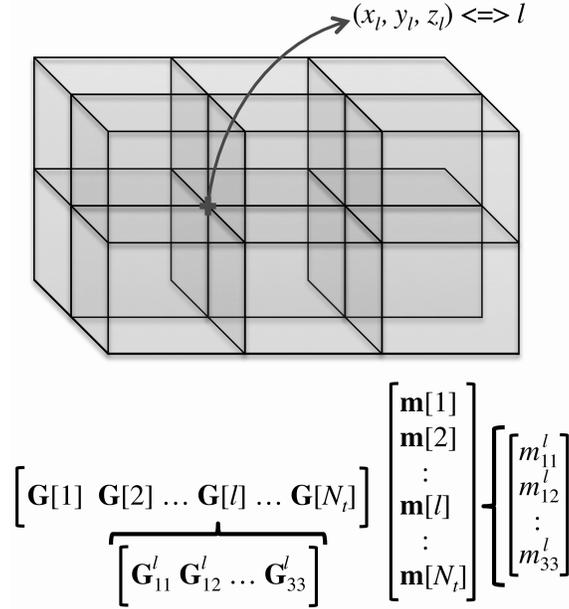


Figure 4.1: Mapping between the subsurface grid volume and the dictionary of Green’s functions. Every node in the grid has spatial coordinates (x, y, z) and a linear index $l \in [1, N_t]$ where N_t is the total of nodes in the grid. Hence, the spatial coordinates of the l -th node are (x_l, y_l, z_l) . A set of 6 Green’s functions corresponding to each of the independent elements of the seismic moment tensor are computed at each grid node. Every set of Green’s functions form a block of the dictionary of Green’s functions \mathbf{G} in equation (4.14). Correspondingly, every block in the solution vector \mathbf{m} is associated to a location in the subsurface grid volume. Hence, the blocks in \mathbf{m} with ℓ_2 -norm different from zero provide the location and moment tensor coefficients where sources in the grid have taken place. An extension in time of the dictionary provides also the origin times of the sources.

where the block $\mathbf{m}[l] = [m_{11}^l, m_{12}^l, m_{13}^l, m_{22}^l, m_{23}^l, m_{33}^l]$ is the seismic moment tensor of the l -th node (virtual source), $\mathbf{G}[l]$ the column-block of the corresponding Green's functions, and N_t is the total number of grid nodes. For a grid of $N_t = N_x \times N_y \times N_z$ nodes, the size of \mathbf{G} is $(L \times N)$, where $N = 6 \times N_t$ and $L = \text{number of receivers} \times \text{number of receiver's components} \times \text{number of samples in 1 component trace}$. For a given time period, the blocks with non-zero ℓ_2 -norm in the vector \mathbf{m} are those corresponding to the node locations where seismic sources have actually taken place. If, for example, in a set of observations \mathbf{u} we only have the displacement field due to one seismic source, then vector \mathbf{m} is block 1-sparse ($\|\mathbf{m}\|_{2,0} = 1 \ll N_t$). Therefore, the connection between the block sparse representation problem and (4.14) is straightforward. The matrix $\mathbf{G} \in \mathbb{R}^{L \times N}$ is a dictionary of Green's functions, while vector $\mathbf{m} \in \mathbb{R}^N$ is a block sparse representation of the source displacement field under that dictionary. By identifying the blocks $\mathbf{m}[l]$, with $d = 6$, that have ℓ_2 -norm different from zero, we are in fact determining the source mechanisms (magnitudes of the elements in each block $\mathbf{m}[l]$), the location coordinates (corresponding to the position of the Green's functions that expand the blocks $\mathbf{m}[l]$) and the origin times (given by the times where the blocks $\mathbf{m}[l]$ are found) of the seismic events represented in \mathbf{u} . An algorithm to accomplish such purpose is proposed in the following section.

4.3.1 Algorithm to invert for the source parameters based on a block sparse representation

The proposed methodology is a BOMP algorithm applied over an augmented dictionary of Green's functions. Hence, using the augmented dictionary, a set of Green's functions and a vector of observations are compared across a 3D volume and time to estimate optimal parameters for origin time and location of seismic events. The augmented dictionary of Green's functions is defined as

$$\mathbb{G} = \begin{bmatrix} \mathbf{G}_0 & \mathbf{G}_1 & \mathbf{G}_2 & \dots & \mathbf{G}_{N_w-1} \end{bmatrix}, \quad (4.15)$$

where \mathbf{G}_k is the dictionary of Green's functions for the whole grid delayed in time by k samples. Notice that \mathbf{G}_0 is basically the same as \mathbf{G} from equation (4.14) (for the rest of this chapter, the variable \mathbf{G} will refer to that of equation (4.14) unless otherwise stated). To simplify the presentation the augmented dictionary \mathbb{G} will be called superdictionary and the delayed dictionaries \mathbf{G}_k superblocks. Hence a superblock \mathbf{G}_k of \mathbb{G} is a k samples delayed version of the dictionary \mathbf{G} whose column blocks are the sets of 6 Green's functions $\mathbf{G}[l]$ for each node l in the grid. The superdictionary \mathbb{G} can be either seen as formed by the superblocks \mathbf{G}_k specified by a single index k , or by the blocks $\mathbf{G}_k[l]$ specified by the two indexes (k, l) . Considering its physical meaning, sometimes I will refer to the block index l

as a position index and to the index k as a delay index (which is related to the origin time). The maximum delay $N_w - 1$ is determined by the processing window selected by the user as I will show in the following paragraphs.

Consider a network of N_{st} three component seismic stations transmitting a continuous flow of seismic recordings to a computing centre. At the computing centre the recordings are cut in windows of duration t_{N_u} . If we consider a sampling rate of Δt this results in sets of $3 \times N_{st}$ traces of $N_u = t_{N_u}/\Delta t$ samples. The subsurface volume of interest is divided using a grid as established before. Then, Green's functions at each grid node are computed and stored to form a dictionary \mathbf{G} . The time duration of a Green's function is t_{N_G} , where t_{N_G} is determined by the longest arrival time for the wave phases of interest between all possible combinations of grid nodes and seismic stations. The number of samples in a Green's function is $N_G = t_{N_G}/\Delta t$. The row size of the dictionary of Green's functions \mathbf{G} is $L = 3 \times N_{st} \times N_G$. The length of a Green's function is related to the length of the observations by $N_u = N_G + N_w$. In order to compute the superdictionary \mathbb{G} , the dictionary \mathbf{G} is delayed and padded with zeros to create N_w superblocks \mathbf{G}_k with row length $3 \times N_{st} \times N_u$. An efficient implementation of this methodology does not require to create the delayed dictionaries \mathbf{G}_k . Instead, the input observation traces are successively cut to the length of the original Green's functions N_G to simulate the required time delays. In fact, the concept of superdictionary and superblocks is only used here as a means to simplify the presentation. The steps followed by BOMP for the recovery of the source parameters can be summarized as

Step 1. Initialization of variables

The observations are arranged in a column vector \mathbf{u} and a residual $\mathbf{r}_0 = \mathbf{u}$ is initialized. Also, the index sets $\mathbf{\Lambda}_0$ and $\mathbf{\Gamma}_0$ are initialized as empty, and the iteration counter i is set to 1.

Step 2. Identification of the best origin time and source location

The function $f(k, l) = \|\mathbf{G}_k^T[l]\mathbf{r}_{i-1}\|_2^2$ is computed for all the column blocks $l \in [1, N_t]$ and delays $k \in [0, N_w - 1]$. For the next step, we identify the delay index k (origin time), and position index l that correspond to the global maximum in the function $f(k, l)$. In other words, we find the indexes

$$n_i = \arg \max_k \left(f(k, l) \right), k \in [0, N_w - 1], l \in [1, N_t],$$

$$m_i = \arg \max_l \left(f(k, l) \right), k \in [0, N_w - 1], l \in [1, N_t].$$

The identified origin time and source location indexes are then saved with those found in previous iterations,

$$\mathbf{\Gamma}_i = \mathbf{\Gamma}_{i-1} \cup n_i,$$

$$\mathbf{\Lambda}_i = \mathbf{\Lambda}_{i-1} \cup m_i.$$

After step 2 of the i -th iteration, each of the sets $\mathbf{\Gamma}_i$ and $\mathbf{\Lambda}_i$ contain a total of i indexes. Step 2 is the most time consuming part of the method.

Step 3. Updating of the source mechanisms

The source mechanisms for all identified sources are updated. For such objective, a matrix \mathbf{A}_i is formed with the Green's function blocks specified by the index sets $\mathbf{\Gamma}_i$ and $\mathbf{\Lambda}_i$. These Green's functions correspond only to the locations and origin times of the sources identified in the previous step, hence, the matrix \mathbf{A}_i has a much smaller number of columns than matrix \mathbb{G} . The matrix \mathbf{A}_i is defined as (for $i > 2$)

$$\mathbf{A}_i = \left[\mathbf{G}_{\mathbf{\Gamma}_i[1]}[\mathbf{\Lambda}_i[1]] \quad \mathbf{G}_{\mathbf{\Gamma}_i[2]}[\mathbf{\Lambda}_i[2]] \quad \dots \quad \mathbf{G}_{\mathbf{\Gamma}_i[i]}[\mathbf{\Lambda}_i[i]] \right].$$

For $i = 1$ or 2 , matrix \mathbf{A}_i is formed by only the first one or two blocks, respectively. Then, the moment tensors from all sources so far identified are simultaneously updated,

$$\tilde{\mathbf{m}}_i = (\mathbf{A}_i^T \mathbf{A}_i)^{-1} \mathbf{A}_i^T \mathbf{u}.$$

In this step a constraint can also be imposed to the inversion. For example, the deviatoric constraint is linear and can be easily incorporated (Aki and Richards, 2009).

Step 4. Update the residual and iterate

The contributions to the displacement field corresponding to all the sources found up to the current iteration are subtracted from the original observations to generate a new residual vector,

$$\mathbf{r}_i = \mathbf{u} - \mathbf{A}_i \tilde{\mathbf{m}}_i.$$

Finally, the counter is updated to $i = i + 1$. Steps 2 through 4 are repeated until a stopping criteria is met. Every iteration of the code finds one seismic event.

Step 5. Output

The output from the inversion code are the origin times given by the index set $\mathbf{\Gamma}_i$, the locations given by the index set $\mathbf{\Lambda}_i$ and the updated moment tensors $\tilde{\mathbf{m}}_i$.

4.3.2 Proposed modification to the BOMP algorithm

Consider the ideal case with accurate Green's functions and no additive noise in the observations. If comparing the blocks $\mathbf{G}_k[l]$, the superdictionary \mathbb{G} meets inequality (4.7), then BOMP will always find the optimal set of source parameters. In other words, the origin times, locations and source mechanisms that correspond to the sparsest representation of the displacement field in terms of \mathbb{G} . Origin time and location will also correspond to the correct source parameters for the actual seismic events, however, the correctness of the source mechanisms will depend on the conditioning and constraint applied to the inversion in step 3. In practical terms, inequality (4.7) states that the columns of the superdictionary \mathbb{G} have to be sufficiently different to allow the algorithm to identify the correct set of Green's functions at each iteration. The sub-coherence and coherence are the way to quantify how alike are the columns within a block, and between different blocks of \mathbb{G} , respectively. It is immediately obvious that the coherence between blocks corresponding to the same node n_l at different time delays (e.g., $\mathbf{G}_{k_1}[n_l]$ and $\mathbf{G}_{k_2}[n_l]$, where $k_1 \neq k_2$) will in general have very high coherence. In this case, a possible alternative to reduce the coherence would be to increase the amount of time delay $k \cdot \Delta t$ between superblocks, however, this still does not guarantee compliance of inequality (4.7), while the resolution in origin time estimation would be negatively affected. The coherence between blocks of the same superblock n_k (e.g., between $\mathbf{G}_{n_k}[l_1]$ and $\mathbf{G}_{n_k}[l_2]$, where $l_1 \neq l_2$) is influenced by the frequency content and distance between the grid nodes. For the same distance between nodes, the lower the frequency content, we expect the coherence between blocks to increase due to the overlap of wave arrivals. Equivalently, fixing the frequency content, the coherence between blocks should increase if the distance between nodes is reduced. In a practical scenario, it is desirable to have combinations of node distance-frequency content where the wave arrivals between adjacent nodes overlap; otherwise, a considerable amount of error in the estimated moment tensors is introduced for seismic sources whose real location is not coincident with one of the grid nodes. The sub-coherence within a block is always high, since energy arrivals are located at similar times in the 6 Green's functions forming the block. It has to be pointed out, however, that while meeting inequality (4.7) is a guarantee of success, not meeting it is not a guarantee of failure. In other words, inequality (4.7) is a guarantee of success in the worst case scenario (Eldar et al., 2010), which means that even if \mathbb{G} does not comply with it, there will be multiple cases where BOMP will still be able to pick optimal

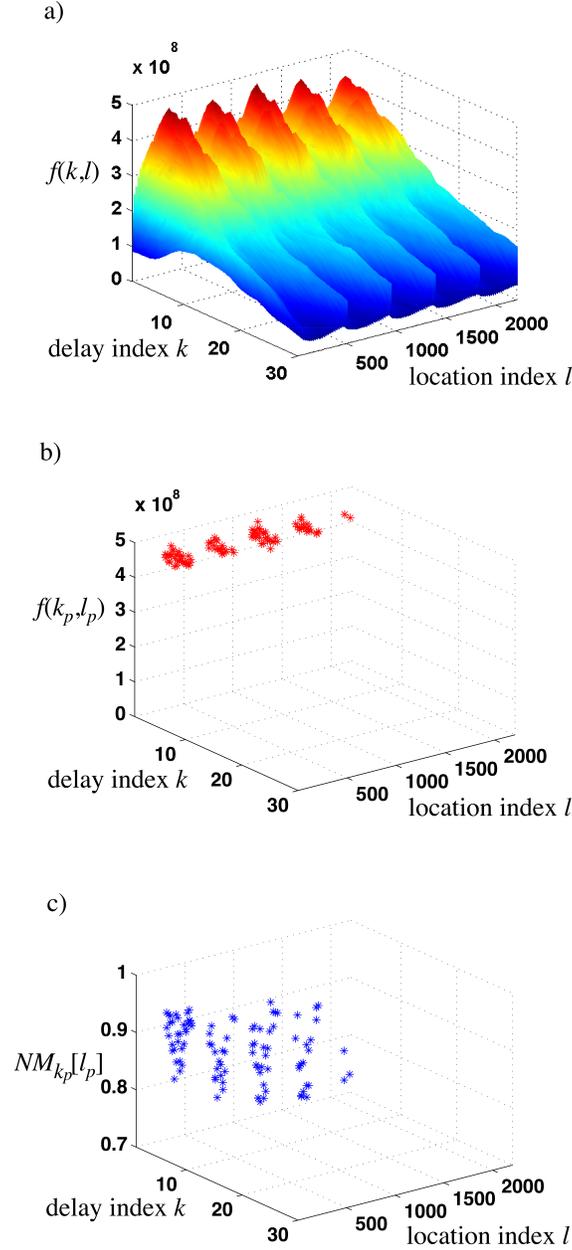


Figure 4.2: Real data example of the modification introduced to the BOMP algorithm. a) Function $f(k, l) = \|\mathbf{G}_k^T[l]\mathbf{r}_0\|_2^2$, the total number of nodes in the grid of this example is $N_t = 2205$, hence $l \in [1, 2205]$, the processing window is $N_w = 30$, hence $k \in [0, 29]$. b) Plot of the n_h largest values in a), in this example $n_h = 100$, therefore $p \in [1, 100]$. c) Normalized misfit for solutions computed at each (k_p, l_p) indexes that correspond to values of $f(k_p, l_p)$ displayed in b). The delay index k_p and position index l_p where the normalized misfit is minimum are considered the optimal origin time and location for a seismic source, and therefore saved in the index sets $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$. In the regular BOMP workflow the indexes (k, l) that correspond to the global maximum in a) are taken as the optimal origin time and location of the seismic source.

solutions. In these scenarios, the key is to create a \mathbb{G} where the chances of success are much higher than the chances of failure. Additionally, BOMP can also be modified to increase its chances of finding optimal solutions. Here, I introduce a modification that splits step 2 in section 4.3.1 into different sub-steps, these are

Step 2.1

Compute the function $f(k, l) = \|\mathbf{G}_k^T[l]\mathbf{r}_{i-1}\|_2^2$ as presented before (Fig. 4.2a). Take the indexes (k, l) of the n_h largest values in $f(k, l)$ and save them into a new variable $\mathbf{h}[p] = [k_p, l_p], p \in [1, n_h]$ (Fig. 4.2b). Then, moment tensor solutions at each index pair $\mathbf{h}[p]$ are calculated with

$$\tilde{\mathbf{s}}_{k_p}[l_p] = (\mathbf{G}_{k_p}^T[l_p]\mathbf{G}_{k_p}[l_p])^{-1}\mathbf{G}_{k_p}^T[l_p]\mathbf{r}_{i-1}, p \in [1, n_h].$$

If a constrained inversion is to be used, then it should also be considered at this sub-step.

Step 2.2

For every solution $\tilde{\mathbf{s}}_{k_p}[l_p]$ a normalized misfit measure is estimated, this is

$$NM_{k_p}[l_p] = \frac{\|\mathbf{G}_{k_p}[l_p]\tilde{\mathbf{s}}_{k_p}[l_p] - \mathbf{r}_{i-1}\|_2^2}{\|\mathbf{r}_{i-1}\|_2^2}, p \in [1, n_h].$$

Finally, the indexes (k_p, l_p) corresponding to the solution $\tilde{\mathbf{s}}_{k_p}[l_p]$ that minimizes the $NM_{k_p}[l_p]$ function are the ones saved in the index sets $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ (Fig. 4.2c). In general, $n_h \ll N_t$ to reduce the impact of this modification over the processing time. The search of minimum misfit solutions over the $NM_{k_p}[l_p]$ function allows BOMP the determination of optimal solutions in highly coherent dictionaries, making it suitable for source parameter determination.

4.3.3 Stopping criteria

In practical scenarios, both BOMP and the modified BOMP have to be provided with a stopping criteria after N_s iterations. Since the algorithms find the source parameters of a seismic event at every iteration, it is natural to set N_s equal to the number of maximum expected sources in each input observations vector. For example, in an earthquake scenario a straightforward value is $N_s = 1$. However, the algorithms will always output a solution after N_s iterations even if no seismic events are recorded in the observations. In order to discriminate between valid and spurious solutions, a detectability threshold α_d is implemented (Fig. 4.3). In each iteration i , a NM_i is computed using the updated residual \mathbf{r}_i , in other

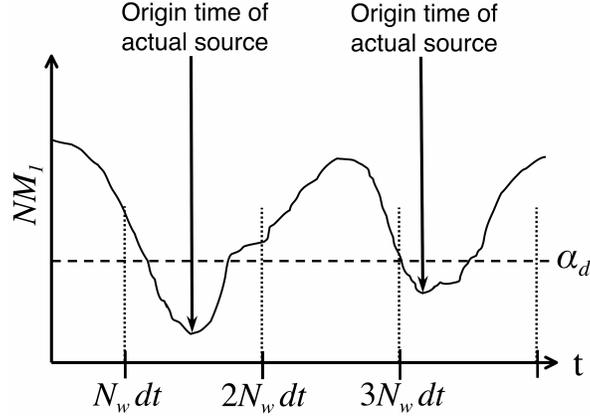


Figure 4.3: Idealized representation of the NM_i variable along time. The parameter α_d is the detection threshold. In this example, the maximum number of iterations is set at $N_s = 1$. Notice that in the limiting case $N_w = 1$, a seismic source would be declared at all time positions where $NM_1 < \alpha_d$ producing multiple spurious sources. On the other hand, if N_w is for example four times bigger than it is showed, N_s must have a minimum value of 2 in order to detect the two sources appearing as valleys in the NM_1 function.

words

$$NM_i = \frac{\|\mathbf{r}_i\|_2^2}{\|\mathbf{u}\|_2^2}.$$

Then, if $NM_i > \alpha_d$ the algorithm stops iterating and re-starts computations with a new observations vector. The seismic traces that form the new observations vector contain the last N_G entries from the previous input traces plus N_w newly recorded samples. In this way, the algorithm searches for seismic sources at all possible alignment positions allowed by the spatial and temporal sampling rates of the system (see example in Fig. 4.3). For multiple sources where $N_s > 2$, the fitting of the solution with the observations is shared between the contributions from each source. In these cases, a detectability criteria requires a more elaborated approach. One possibility is to set a second parameter β defining a minimum improvement in fitting required to declare a new source for all $i > 1$. The parameter α_d would still be the minimum fitting required to accept the whole solution. Multiple source scenarios are not analyzed in detail in this chapter. A synthetic example of multiple sources solved with the same fundamental ideas presented here can be consulted in Vera Rodriguez et al. (2010a). In the following section, the modified BOMP is tested in a synthetic scenario pertaining to the microseismic monitoring of a hydraulic fracture in an oil well, and with a real dataset from the 18 June, 2002 Caborn, Indiana earthquake. In both cases, the computational work was performed using a desktop computer with processor speed of 2.93 GHz and 8 GB in direct access memory. The code has not yet been parallelized, hence, all

the computations were performed using a single core of the computer.

4.4 Application to synthetic and real data examples

4.4.1 Hydraulic fracturing in an oil well

Oil and gas wells are fractured through the injection of fluids at high pressures in order to enhance production (Cipolla and Wright, 2002). Monitoring and processing of the microseismic activity generated during the injection process is a tool to study the properties of the induced fractures (Maxwell, 2005; Warpinski, 2009). For such objective, arrays of recording stations are positioned either on the surface or inside nearby wells (Maxwell et al., 2010; Duncan and Eisner, 2010). In this synthetic example, I consider the case of recording stations positioned downhole. For the remainder of this thesis, the well subjected to hydraulic fracturing will be referred to as the treatment well, while the well where the recording stations are positioned will be called the monitoring well. Both treatment and monitoring wells in this example are vertical (Fig. 4.4). The injection point in the treatment well is located at a depth of 2000 m, which also corresponds to the center of the grid where the algorithm will detect seismic events. Discretization of the subsurface using a grid introduces error in the estimated seismic locations. In a regular grid this error is at most $a\sqrt{3}/2$, where a is the distance between nodes, constant in all coordinate directions. This maximum discretization error corresponds to a source with actual location at the center of a grid cell, and whose estimated location can only be attributed to one of the nearby nodes. Although there is no restriction on the location of the nodes for the purpose of running the algorithm, the use of a regular grid simplifies the calculations. Three different combinations of node distance - maximum frequency content are used in these experiments: $a = 10$ m, $f_{max} = 100$ Hz (model 1), $a = 10$ m, $f_{max} = 400$ Hz (model 2), and $a = 40$ m, $f_{max} = 400$ Hz (model 3). The maximum errors due to spatial discretization in these three models are considered reasonable given the magnitude of location errors usually expected in a microseismic monitoring application (Eisner et al., 2009, 2010; Kidney et al., 2010). Furthermore, the maximum frequency content allowed in each model is also consistent with what could be expected for receivers located at the horizontal distances defined in these experiments (e.g., Shemeta and Anderson, 2010). Additional simplifications in the design of the experiment are the use of a homogeneous isotropic medium and a relatively small number of grid nodes ($N_x = N_y = N_z = 9$). The total number of nodes in the grid is $N_t = N_x \times N_y \times N_z = 729$ which, depending on the distance between nodes, corresponds to an also different monitored volume. The monitoring well is located at a horizontal distance of 400 m in the N35°E direction. An array of 8 receivers with equal separations of 30 m between them, spans a longitude

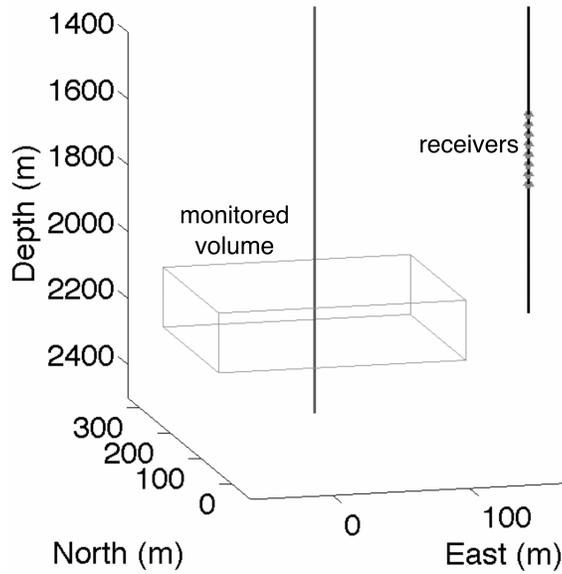


Figure 4.4: Geometry of the microseismic monitoring experiment used to test the modified BOMP. The treatment well passes through the monitored volume, which is outlined by a grey box. The monitoring well is where the receiver locations are displayed.

of 210 m from 1700 m to 1910 m of depth in the monitoring well (Fig. 4.4). All these parameters reflect realistic geometrical settings for actual monitoring applications (see e.g., Eisner et al., 2010). Green’s functions and synthetic sources are computed using ray tracing theory and corrected for the source radiation pattern (Aki and Richards, 2009). All traces are compared as velocity seismograms, rotated to a Vertical-Transverse-Radial system and low-pass filtered to the maximum frequency content referred in each model. Moment tensor inversions using a vertical array of receivers cannot resolve the dipole perpendicular to the plane formed by the source location and the receivers (Nolen-Hoeksema and Ruff, 2001; Vavrycuk, 2007; Vera Rodriguez et al., 2011), therefore, a deviatoric constraint is applied to all cases in these synthetic experiments.

Influence of noise content on the solutions

I gauge the effect of noise through the SNR, which is defined as $\text{SNR} = \max(|\mathbf{u}|)/\sigma_n$, where σ_n is the standard deviation of \mathbf{n} , a zero mean gaussian noise vector, and $\max(|\mathbf{u}|)$ is the maximum absolute value of the amplitudes in the current observation vector. At each node of the grid, 5 random sources are modelled for a total of $5 \times N_t = 3,645$ realizations in

each of the three previously described models. Each set of realizations is then subjected to the required values of SNR = 0.1, 0.5 and 1. The average location error (*ALE*) for a given model-SNR combination is computed by

$$ALE = \sum_{n=1}^{N_t} \left(\frac{\sum_{k=1}^{k_s} \|\boldsymbol{\xi}_n - \boldsymbol{\xi}_k\|_2^2}{k_s} \right) / N_t, \quad (4.16)$$

where k_s is the number of sources that the algorithm was able to resolve at node n (i.e., $k_s \in [1, 5]$). $\boldsymbol{\xi}_n$ is the actual source (node) location and $\boldsymbol{\xi}_k$ is the source location estimation for the k -th source. Similarly, the error in the moment tensors is estimated by an average normalized mean squared error (*ANMSE*),

$$ANMSE = \sum_{n=1}^{N_t} \left[\left(\frac{\sum_{k=1}^{k_s} \|\mathbf{m}_n^k - \hat{\mathbf{m}}_n^k\|_2^2}{\sum_{k=1}^{k_s} \|\mathbf{m}_n^k\|_2^2} \right) / k_s \right] / N_t, \quad (4.17)$$

where $\hat{\mathbf{m}}_n^k$ is the inversion output for the k -th source at n and \mathbf{m}_n^k its actual value. In this calculation I only consider the 5 elements of the moment tensor that are constrained by the monitoring geometry (Jechumtalova and Eisner, 2008). The average origin time error (*AOTE*) is computed from

$$AOTE = \sum_{n=1}^{N_t} \left(\frac{\sum_{k=1}^{k_s} |OT_n - OT_k|}{k_s} \right) / N_t, \quad (4.18)$$

where OT_n and OT_k represent the ground-truth and estimated values for the k th source at node n , respectively. As expected, increasing the SNR would improve the accuracy of the results. For moderate SNR levels (e.g., SNR = 1 and 0.5, Fig. 4.5), the three models display low errors in all estimated source parameters. The higher errors and standard deviations for location and origin time estimates are observed in models 1 and 2 with 10 m separation between nodes (Figs 4.5b and c). Between these two models, model 1 with 100 Hz frequency content present slightly higher magnitude of errors and standard deviations. These results can be analyzed in the context of the coherence between Green's functions. The lower the frequency content and smaller the distance between nodes, the Green's functions between different nodes in the grid are more alike, hence the modified BOMP has more trouble identifying the correct solution. In the case of the source parameters, the behaviour reverses (Fig. 4.5a), now it is model 3 which presents the higher magnitude of errors and standard deviations. This effect is in response to the higher separation between nodes (40 m) and frequency content (400 Hz) with respect to the other models. The greater overlaps of Green's functions in models 1 and 2 produces better estimates of moment tensors for slightly mislocated sources as compared to model 3 where the overlap is less. Additionally, this lack of overlapping increases the value of the normalized misfit in mislocated events,

generating more missed sources in model 3 (see percentages in Fig. 4.5a). Hence, there is a trade-off in the amount of overlapping between the Green's functions, which on one hand helps to obtain more robust moment tensor estimates, but on the other, makes it harder for the modified BOMP to identify the optimal locations and origin times. In the presence of very low SNR (e.g., $\text{SNR} = 0.1$, Fig. 4.5), the algorithm has difficulties detecting the synthetic events (see percentages in Fig. 4.5a). The only case where a considerable amount of events is resolved is model 1. However, this effect is only a result of the greater overlaps between the Green's functions in this model. In other words, since the Green's functions from multiple nearby nodes are highly coherent (alike), the solutions can have large errors in location and origin time and still produce a normalized misfit that will pass the detectability threshold. The less the overlap between Green's functions, the less the number of sources detected by the algorithm (with n_h constant). However, the quality of the source parameters of the sources that are detected is improved. For sources that pass the detectability criteria, the number of the index $p \in [1, n_h]$ where the solutions are found remains stable in all the experiments (Fig. 4.5d). In general, a possibility to increase the number of detected sources is to increase the value of n_h ; in this case, a trade-off exists with respect to the associated increase in processing time.

Influence of inaccurate Green's functions on the solutions

The effect of inaccurate Green's functions is examined using two examples. First, I compute Green's functions and a set of 400 observations vectors, both using the same velocity model. By design, the source location of these observations does not coincide with any of the grid nodes. In the second scenario, I introduce perturbations to the Green's function computations and adopt a set of 400 sources that overlap with the grid node locations. For actual source locations that do not coincide with the grid nodes, the errors in the estimated source parameters increase as the overlap between Green's functions of different nodes is reduced (Figs 4.6a through c). The effect is similar for inaccurate Green's functions that result from an inexact velocity model (Figs 4.7a through c). As one would expect, the more inexact the velocity model, the higher the errors in the estimated source parameters. Again, the number of the index $p \in [1, n_h]$ where the solutions are found remains stable in all cases (Figs 4.6d and 4.7d). These results suggest that a highly coherent dictionary provides estimations that are more robust to errors introduced by inaccuracies of the dictionary itself. On the other hand, highly coherent dictionaries present a difficulty for the modified BOMP to find optimal solutions. Increasing the variable n_h would help the algorithm, however, this would be at the expense of increasing computing time. A practical way to design a dictionary of Green's functions for a real application is through numerical experiments like the ones just presented. Following the results from the experiments, model 2 with distance between

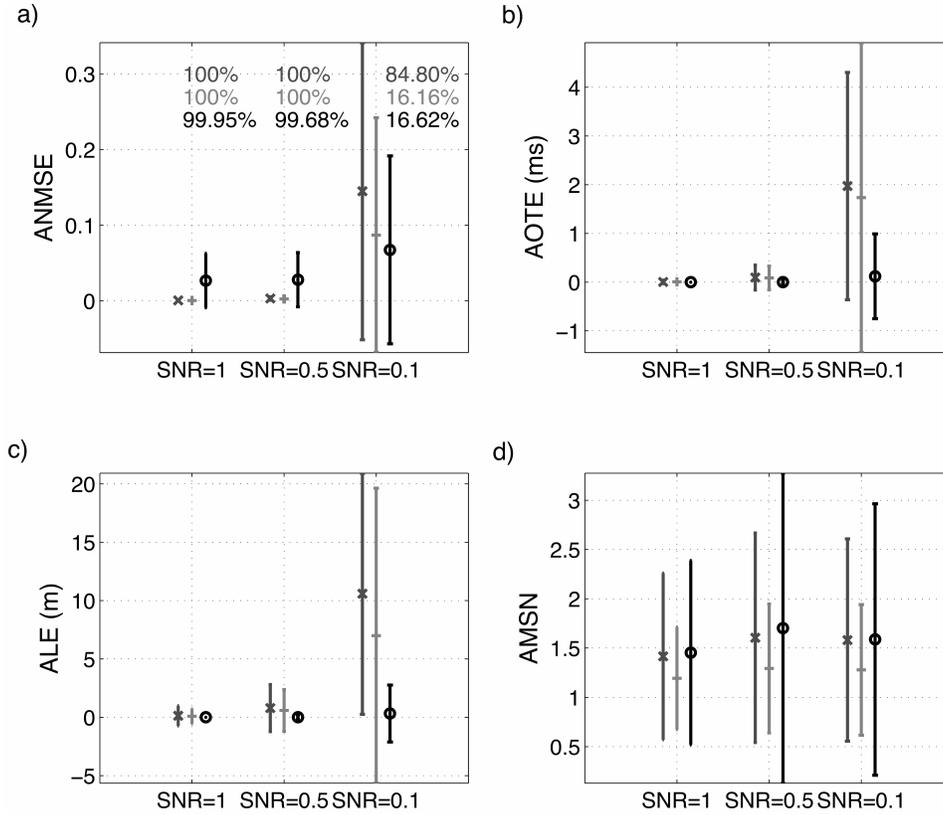


Figure 4.5: Average source parameter inversion errors for different levels of SNR. Crosses corresponds to model 1, dashes to model 2, and circles to model 3. a) error in moment tensor estimation, the percentages relate to the number of realizations resolved by the algorithm in each SNR level (top values correspond to model 1, followed by models 2 and 3), b) error in origin time, c) error in location and d) average value of the index $p \in [1, n_h]$ where the solutions were found. Vertical bars are standard deviation.

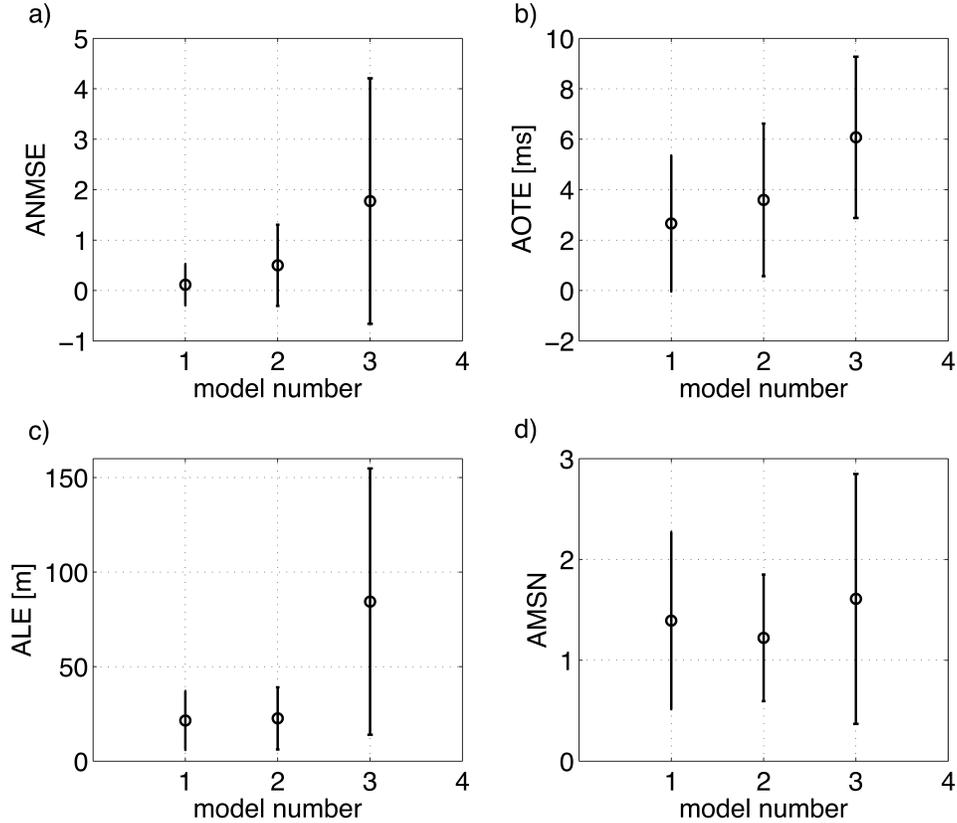


Figure 4.6: Effects of inaccurate Green's functions in the inversion output. The inaccuracy in the Green's functions is due to source locations not coinciding with the grid nodes. a) error in moment tensor estimation, b) error in origin time, c) error in location and d) average value of the index $p \in [1, n_h]$ where the solutions were found. Vertical bars are standard deviation.

nodes of 10 m and maximum frequency content of 400 Hz seems to provide a good trade-off between detectability and errors expected in the solutions. Furthermore, these properties can also be improved by setting the variable n_h at a level that is reasonable in terms of the required speed at which solutions have to be output.

4.4.2 Application to the 18 June, 2002 Caborn Indiana earthquake

The 18 June, 2002 Caborn Indiana earthquake is a moderate-sized earthquake with a well studied source location and seismic moment tensor. The event occurred in the Wabash Valley Seismic Zone (WVSZ), located in the central-eastern U.S. shared by the states of Indiana, Illinois and Kentucky (Langer and Bollinger, 1991) (Fig. 4.8). Considering its

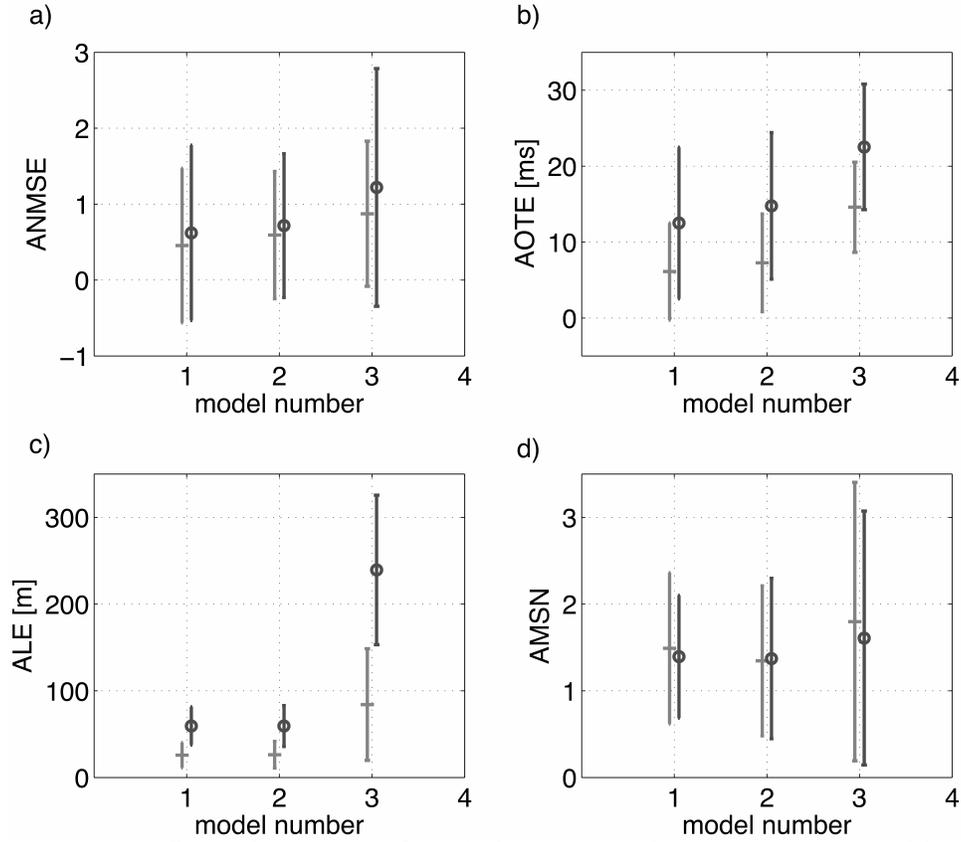


Figure 4.7: Effects of inaccurate Green's functions in the inversion output. After computing the observations, the velocity model is perturbed by 5% (circles) and 3% (dashes) to compute dictionaries of Green's functions. a) error in moment tensor estimation, b) error in origin time, c) error in location and d) average value of the index $p \in [1, n_h]$ where the solutions were found. Vertical bars are standard deviation.

Table 4.1: Velocity model used in the Caborn Indiana earthquake inversion (Kim, 2003).

Depth km	Vp $\frac{km}{s}$	Vs $\frac{km}{s}$
7	5.9	3.4
28	6.5	3.7
	8.0	4.5

seismic activity, the Wabash Valley is a prominent region of Illinois, since it is the only part that does not behave as an intra-plate region (Kontar et al., 2010). A focus of seismic activity in the WVSZ is the Wabash Valley Fault System (WVFS), which is a system of subsurface normal faults that run parallel to the Wabash River Valley. The faults are high angle, trending north-northeast and have been mapped down to more than 7 km penetrating the Precambrian basement (Bear et al., 1997). The WVFS covers an approximate area of 90 km by 50 km (see Fig. 4.8), the faults have been interpreted to split upward outlining horsts and grabens (Nelson, 1991). Kim (2003) infers that the WVFS could have initiated during the Iapetean, while the region was rifting, however Kontar et al. (2010) point out that timing and origin of the deep faults remains unknown. Based on an analysis that includes the 18 June 2002 earthquake, Kim (2003) suggests the reactivation of a possible Precambrian rift system by the contemporary regional stresses. A description of the source parameters of the 18 June 2002 event determined by Kim (2003) is presented in Fig. 4.9. These results were obtained from the use of a 1D velocity model (Table 4.1) and the recordings from 15 seismic stations with epicentral distances in the range 129-585 km. The epicentral location, source mechanism and depth of the event were determined separately. The source mechanism and depth were estimated using a waveform fitting method that seeks the best double couple solution within a grid spanning the (strike, dip, rake) space, and the best depth within a range of values (Zhao and Helmberger, 1994). The depth of the event was further corroborated by analyzing teleseismic P waves (~ 19 km). The dominant phases in the filtered regional seismograms are fundamental Rayleigh and Love waves, although body waves (P) from the eight closest stations were also considered. The epicentral location was ultimately refined using P- and S-wave time picks from 20 local and regional seismic stations and the depth from the source mechanism inversion. In order to test the inversion algorithm, I formulate a hypothetical monitoring system around the WVFS to detect and invert for the source parameters of the 18 June, 2002 event.

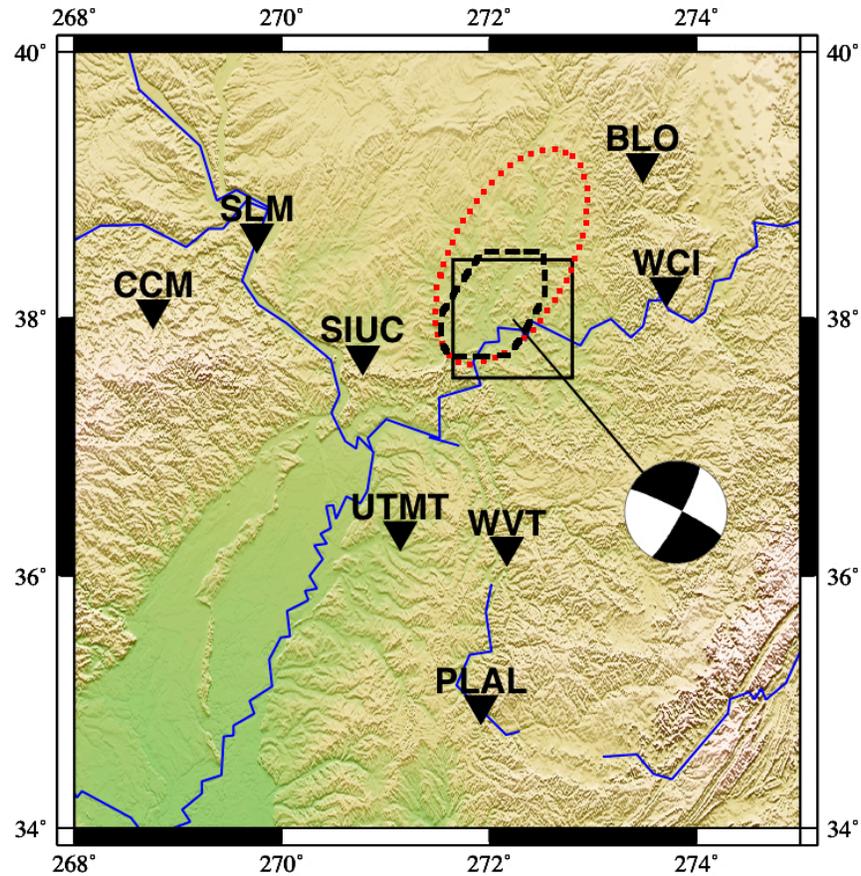


Figure 4.8: Epicentral location of the 18 June 2002 Caborn, Indiana earthquake and distribution of seismic stations considered in this work (black triangles). The beach ball corresponds to the solution obtained by Kim (2003). The Wabash Valley Seismic Zone (WVZS) is delimited by the dotted line. The Wabash Valley Fault System (WVFS) is enclosed by the dashed line. The black square is the surface projection of the grid used in this study.

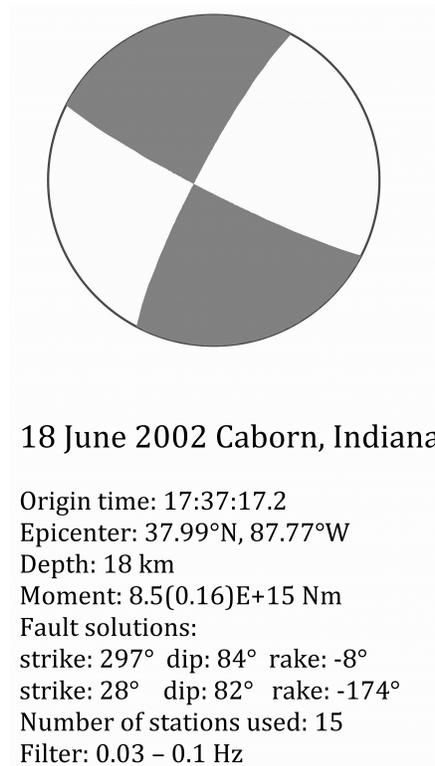


Figure 4.9: Source parameters of the 18 June 2002 Caborn, Indiana earthquake determined by Kim (2003). The source mechanism is represented by the beach-ball diagram (lower-hemisphere projection).

Description of the monitoring system

The hypothetical system consists of 8 broadband seismic stations in the approximate distance range of 100-300 km from WVFS (Fig. 4.8, Table 4.2). The monitored volume is divided using a grid of $21 \times 21 \times 5 = 2,205$ nodes or virtual sources. The distance between nodes is 5 km both in the horizontal x (North), y (East) and depth directions. The surface projection of the grid covers an area of 100 km \times 100 km outlining practically all WVFS. The node depths range from 5.5 to 25.5 km, which also cover the maximum proposed depth of the seismogenic region (~ 20 km). Following Kim (2003), the frequency range considered is 0.01-0.1 Hz. Similar combinations of distance between nodes and frequency content are used in the GridMT system (Tsuruoka et al., 2009), and for the study of earthquakes offshore Northern California (Guilhem and Dreger, 2010). In these two latter works, the detection threshold is set using a variance reduction (VR) measurement defined as

$$VR = 100(1 - NM). \quad (4.19)$$

In the GridMT system the detection threshold is $VR \geq 65.0$, solutions present variance reductions of similar magnitude in the offshore California system. Considering the relationship between the VR and the NM , the detection threshold in the hypothetical system is set at $\alpha_d \leq 0.35$. For each set virtual source-recording stations, six Green's functions corresponding to each of the elemental force couples in the seismic moment tensor are computed and arranged in a dictionary. Both the Green's functions and observations are frequency filtered, rotated to a Radial, Transverse, Vertical system and processed to be compared as velocity records. Furthermore, amplitudes are normalized, hence no scalar moment estimate is provided for this example. The Green's functions are computed using the velocity model displayed in Table 4.1 using the reflectivity method from Randall (1994). The duration of the Green's functions is 150 seconds, which considering a sampling rate of $\Delta t = 1$ s is equivalent to 150 samples. The total response time (t_R) of the system is

$$t_R = t_{N_G} + t_{N_w} + t_T + t_p, \quad (4.20)$$

where $t_{N_G} = 150$ s is the duration of the Green's functions, t_{N_w} is the length of the processing window, t_T is the transmission time from the recording stations to the computing centre, and t_p is the processing time of the algorithm. The duration of the observation vectors that input the inversion is $t_{N_u} = t_{N_G} + t_{N_w}$. A new vector of observations enter the inversion every $t_{N_w} + \Delta t$ seconds. The objective is that $t_p \leq t_{N_w} + \Delta t$, in other words, the system must provide an answer before new information is available to be inverted. The inversion is set to look for a maximum of 1 source at each input observation vector (i.e., $N_s = 1$). The search for optimal solutions in the modified BOMP is set to $n_h = 200$.

Table 4.2: List of seismic stations used in the hypothetical monitoring system.

Name	Lat (°)	Lon (°)	Elevation km
SIUC	37.71	-89.22	120
WCI	38.23	-86.29	210
BLO	39.17	-86.52	246
UTMT	36.35	-88.86	110
WVT	36.13	-87.83	153
SLM	38.64	-90.24	186
CCM	38.06	-91.24	222
PLAL	34.98	-88.08	165

Inversion of the 18 June, 2002 earthquake records

To simulate the detection and inversion of the 18 June, 2002 earthquake, the seismic recordings of the event are cut from 60 seconds before the origin time determined by Kim (2003) and up to 240 seconds after. The processing window is then set as $N_w = 30$ samples. Given the length of the Green's functions and the sampling rate, this means that a set of 5 observation vectors \mathbf{u} will enter the inversion process. The required processing time to guarantee a real-time response is $t_p \leq 31$ s. The average processing time observed during the tests was $\bar{t}_p \sim 4$ s, corroborating the feasibility of real-time monitoring using the hypothetical system, and that a bigger number of grid nodes and/or recording stations can be used. A disadvantage in the use of adjacent processing windows is the possibility to detect the same seismic event more than once when its origin time is close to the limit of the processing window. Alternatives to overcome this limitation are the use of overlapping windows and/or the comparison of results between different adjacent windows (e.g., Tsuruoka et al., 2009). In order to analyze with more detail the estimated source mechanism the inversion is run twice, once applying the deviatoric constraint and the other allowing a full moment tensor recovery. The percentages of isotropic (ISO), double couple (DC) and compensated linear vector dipole (CLVD) for each solution are presented in Table 4.3 (Knopoff and Randall, 1970; Vavrycuk, 2001). In the unconstrained solution, the percentage of isotropic component is negligible, on the other hand, in both results the source mechanism is predominantly double couple. The amount of CLVD percentage in both solutions is associated to the angle that the slip vector deviates from the fault plane (α). The small magnitude of α supports agreement between the results of this study and the assumption of pure double couple mechanism used by Kim (2003). Furthermore, the existence of an $\alpha \neq 0$ might be in part due to the presence of noisy components in the input records. In its current version, the algorithm does not have a mechanism to filter out of the inversion traces with low signal to noise

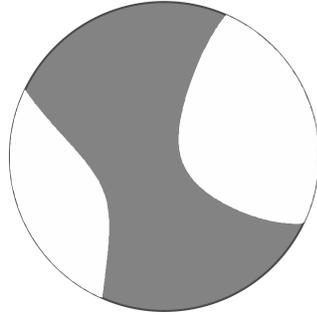
Table 4.3: Decomposition in percentages of isotropic (ISO), double couple (DC) and compensated linear vector dipole (CLVD) of source mechanism solutions for the 18 June, 2002 Caborn earthquake. Negative percentages in ISO and CLVD components are associated to compressive faults, α is the angle formed between the fault plane and the slip vector (Vavrycuk, 2001).

Type of constraint	%ISO	%DC	%CLVD	α [°]
No constraint	-4	67	-29	-14
Deviatoric	0	70	-30	-14

ratio, which can introduce biases into the inversion results. The origin time and location in both of the solutions are identical. Although the grid was designed on purpose not to coincide with the location determined by Kim (2003), the algorithm determined the event's position at the node that was closest to Kim's solution. Particularly, the results show the depth of the event at 20.5 km, which agrees better with the event's depth determined from teleseismic waves (~ 19 km). A summary of the solution using the deviatoric constraint is presented in Fig. 4.10.

4.4.3 Comparison with other real-time inversion methodologies

The Introduction of this thesis presents a rough overview of source parameter inversion methods in earthquake seismology and microseismic monitoring in oil and gas wells. The list of available methodologies to invert for the source parameters is even longer if we consider other areas of applied geophysics, e.g., mining applications (Gibowicz, 2009). Although the methodology presented in the current chapter has not been tested as extensively as the methodologies mentioned in the Introduction, it presents capabilities that are appealing for the development of new and more powerful source parameter inversion techniques using the theory of sparse representations. As it was mentioned before, most source parameter inversion techniques require the calculation of origin time and location (either epicentre, hypocentre or centroid) before starting the estimation of the source mechanism, since this is usually performed by independent algorithms, a waiting time is implicitly required for the estimation of the moment tensor. The BOMP method proposed here overcomes this limitation, making it suitable for the continuous automatic real-time recovery of the three source parameters simultaneously. Another methodology that presents these capabilities is the grid search approach proposed by Kawakatsu (1998). Similar to BOMP, this grid search approach deals with a continuous flow of seismic recordings, does not require a priori information about event occurrence, and the output from the system are the three source parameters (origin time, location and moment tensor). Additionally, in both cases the monitored volume is divided using a grid. It is also this parametrization of the subsurface



18 June 2002 Caborn, Indiana

Origin time: 17:37:16.0
 Epicenter: 37.988°N, 87.770°W
 Depth: 20.5 km
 Number of stations used: 8
 Filter: 0.01 - 0.1 Hz

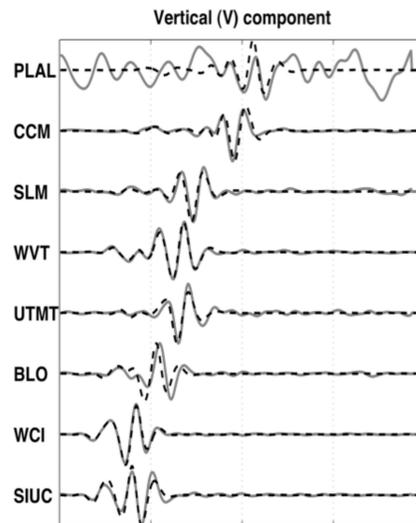
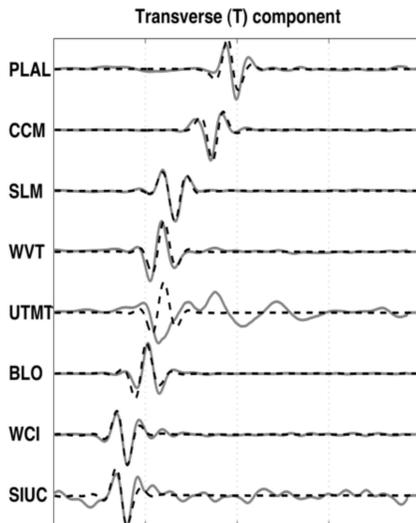
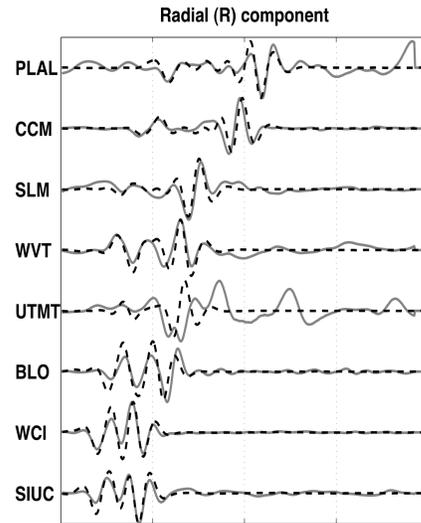


Figure 4.10: Source parameters of the 18 June 2002 Caborn, Indiana earthquake determined by the hypothetical monitoring system. The source mechanism is represented by the beach-ball diagram (lower-hemisphere projection). Solid line: observed records. Dotted line: synthetic records obtained from the forward modelling of the modified BOMP solution.

into a grid that imposes the main limitations for both systems, which are related to the maximum number of grid nodes (virtual sources) and recording stations that can be inverted in real-time. Kawakatsu’s method is currently implemented in the Earthquake Research Institute (ERI) of Japan under the name GridMT (Tsuruoka et al., 2009). GridMT divides the subsurface using a grid and then solves the system in (4.12) at each grid node. The least squares solution to (4.12) is $\mathbf{m}_n = (\mathbf{G}_n^T \mathbf{G}_n)^{-1} \mathbf{G}_n^T \mathbf{u}$, where n is a node index. In order to reduce computational effort during the inversion, matrix $\mathbf{A}_n = (\mathbf{G}_n^T \mathbf{G}_n)^{-1}$ is computed for all the grid nodes and saved in memory. Hence, at each time step, matrix \mathbf{A}_n is multiplied times the product $\mathbf{b}_n = \mathbf{G}_n^T \mathbf{u}$ between the observations \mathbf{u} and the Green’s functions from each node \mathbf{G}_n . Then, GridMT performs a grid search for the node with the solution \mathbf{m}_n that gives the highest VR . From this description, two differences between BOMP and GridMT are immediately obvious. First, since matrix \mathbf{A}_n is already computed in GridMT, it is not possible to alternate between deviatoric and full moment tensor solutions, unless two different versions of \mathbf{A}_n are pre-computed and saved. In the BOMP method, the constraint for deviatoric solutions is easily switched on and off at step 3 (and 2.1 in the modified BOMP). Furthermore, in single source scenarios like earthquake monitoring, a non-linear pure double couple constraint could also be incorporated in step 3 by replacing the least squares source mechanism determination with an algorithm that performs a search over a (dip, rake, strike) solution space (e.g., Zhu and Helmberger, 1996). The second difference is related to the comparison between Green’s functions and observations. In both methodologies the product \mathbf{b}_n is computed at each node (consider a single time step in both cases). This is the part of the process that takes the highest computational cost. In the BOMP method an extra step determines the ℓ_2 -norm of each \mathbf{b}_n . GridMT then calculates the product $\mathbf{A}_n \mathbf{b}_n$ for all nodes, while the modified BOMP method computes this product for only the n_h nodes that present the highest $\|\mathbf{b}_n\|_2^2$ values, where n_h is much smaller than the total number of nodes in the grid. This provides a subtle advantage in terms of computational speed. Additionally, an extra capability of BOMP-based methods is the possibility to identify more than one source contained in the same vector of observations \mathbf{u} .

4.5 Summary

Through a convenient parametrization of the seismic source forward problem, I showed that the source mechanism is a sparse representation of the source displacement field under a dictionary of Green’s functions. Hence, sparse representation theory provides alternative strategies for simultaneously inverting for the origin time, location and moment tensor of seismic events. The modified BOMP is suitable to enforce block sparsity in the solutions and to identify optimum source parameters in the presence of highly coherent dictionaries. The BOMP method used in the present chapter follows essentially the same workflow of

the regular BOMP algorithm. The main difference comes from the physical meaning of the superdictionary used in the source parameter inversion, where solutions are found after matching the input observations both in 3D space and time. Although, without giving consideration to its physical meaning, the superdictionary \mathbb{G} could be treated as a regular dictionary, the high coherence between its columns makes difficult the identification of optimal solutions using the regular BOMP method. Hence, a modification based on a search over the largest values of the ℓ_2 -norm of the correlation between the blocks of the dictionary and the current residuals makes the algorithm more suitable to work in highly coherent dictionaries such as those that arise in seismic monitoring applications. Although the BOMP method has not been tested as extensively as other methodologies currently in use in different areas of geophysics, it presents capabilities that are of great interest for the development of improved source parameter inversion systems. One characteristic that differentiates BOMP from most other source inversion approaches is that it is specifically designed to be implemented as a continuous monitoring system. Also, it can potentially solve for multiple sources in the same set of observations that enter the inversion.

CHAPTER 5

Compressed domain inversion of the seismic source parameters¹

5.1 Introduction

Most inversion algorithms employed in seismic source monitoring applications do not provide simultaneous estimations of the three source parameters. Information about origin time and source location is determined by independent systems and/or algorithms, with the source moment tensor being estimated in a secondary step. Therefore, an inherent lag-time is required preventing efficient, simultaneous estimates of the three source parameters. Algorithms that do not require a lag-time are grid search (Kawakatsu, 1998; Tsuruoka et al., 2009) and matching pursuit (Vera Rodriguez et al., 2012) approaches. Despite their advantages, however, these two methods are also constrained by the maximum numbers of grid nodes and recording stations that can be considered without compromising their real-time response.

Compressive sensing (CS) is a relatively new field of signal processing and applied mathematics that investigates sampling and recovery conditions for signals susceptible to a sparse representation via a known basis or dictionary (Candes et al., 2006; Donoho, 2006). A major result of CS is the specification of protocols for sampling of signals using a number of non-adaptive measurements that is below the number of samples required with the traditional Nyquist criterion. The non-adaptive measurements are linear combinations of the information contained in the signal weighted by coefficients prescribed in the form of a sensing matrix. In other words, using CS a signal can be acquired and transmitted in

¹A version of this chapter has been submitted for publication. Vera Rodriguez, I., Sacchi, M. and Gu, Y. 2012. Geophysical Journal International.

a compressed form, and ultimately uncompressed without perceptive loss of information. This represents an important improvement in efficiency from the traditional practice where a signal is acquired in full form, then compressed throwing away information, transmitted, and finally uncompressed. CS principles have found application in the fields of Magnetic Resonance Imaging (MRI) (Lustig et al., 2007), digital camera design (Takhar et al., 2006), acquisition of astronomical data (Bobin et al., 2008). In geophysical applications, CS has been implemented for earthquake location (Yao et al., 2011), simultaneous estimation of origin time, location and seismic moment tensor (Vera Rodriguez et al., 2010b), and acquisition of seismic data with simultaneous active sources (Herrmann, 2010).

In this chapter, I introduce the principles of CS into the problem of seismic source parameter estimation. I demonstrate through numerical experiments that the introduction of CS extends the real-time applicability of grid search and matching-pursuit-only methods. Improvements in processing time due to compression allow the monitoring of larger subsurface volumes using significantly denser receiver arrays without compromising real-time response.

5.2 Compressed domain inversion

As mentioned above, CS states conditions that permit the signal \mathbf{u} to be recovered from a number of non-adaptive measurements below the minimum number of samples required by the Nyquist criterion. The non-adaptive measurements are random combinations of the samples of the original signal obtained through a sensing matrix Φ (Figure 5.1), this is

$$\Phi \mathbf{u} = \Phi \mathbb{G} \mathbf{D} \mathbf{D}^{-1} \mathbf{m} = \Phi \mathbf{A} \mathbf{b}, \quad (5.1)$$

where $\mathbf{A} = \mathbb{G} \mathbf{D}$ and $\mathbf{b} = \mathbf{D}^{-1} \mathbf{m}$. The variable \mathbf{D} is a diagonal matrix with scalar factors that normalize by energy the columns of \mathbb{G} . The energy in \mathbf{g}_i , the i -th column of matrix \mathbb{G} , is given by

$$E_i = \mathbf{g}_i^T \mathbf{g}_i.$$

Hence the element d_{ii} of \mathbf{D} is given by the square root of the reciprocal of E_i . The sensing matrix Φ contains energy-normalized rows and complies with the Restricted Isometry Property (RIP) (Candes et al., 2006). While proving that a matrix complies with the RIP is rather a non-trivial task, it has been showed that matrices with random entries meet with high probability the RIP (Baraniuk et al., 2008). Between the most recurred sensing matrix ensembles it can be counted Gaussian and Bernoulli matrices. In the applications described in this chapter, Gaussian ensembles are employed. A Gaussian sensing matrix is constructed by drawing random entries from a normal probability density function with zero mean and variance $1/K$ (Baraniuk, 2007). The scalar K is the number of measurements left

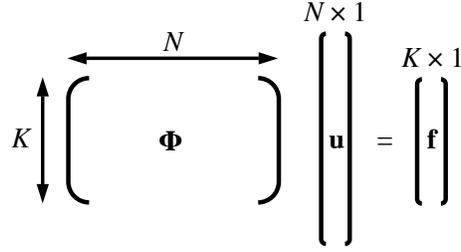


Figure 5.1: Diagram showing the encoding process of the seismic observations. The sensing matrix Φ acts as an encoder compressing the samples in the observations vector \mathbf{u} . Assuming that the original signal is sampled at a Nyquist rate, it turns out that the number of non-adaptive measurements that results from encoding \mathbf{u} is less than the number of samples required to sample \mathbf{u} through the Nyquist criterion. Compressive sensing (CS) provides conditions for the encoding and posterior decoding of \mathbf{f} to recover \mathbf{u} . Practical CS aims at the development of recording instruments that allow the acquisition of the encoded signal \mathbf{f} . In this thesis however, CS principles are applied to signals already recorded at a Nyquist rate, where further advantages arise from the processing of those signals in a compressed domain.

after applying Φ and for instance is related to the desired amount of compression. Clearly, the random variables that form a sensing matrix are non-adaptive, in other words, the same sensing matrix can be used for completely different applications. For this reason the compressed samples are called non-adaptive measurements. The practical effect of the sensing matrix in the system of equations is a dimensionality reduction in the row direction. Consider the dimensions of all the variables in equation (5.1): $\mathbf{u} \in \mathbb{R}^{M \times 1}$, where M = number of recording stations \times number of components in each recording station \times number of time samples in each component, $\mathbf{G} \in \mathbb{R}^{M \times N}$, where $N = 6$ (Green's functions) $\times N_t$ (number of nodes in the grid) $\times N_w$ (length of the processing window), and $\mathbf{m} \in \mathbb{R}^{N \times 1}$. Then, by definition, $\Phi \in \mathbb{R}^{K \times M}$, where $K < M$. Hence, by multiplying times Φ the resulting system of equations is "compressed" from an original domain with dimensions $M \times N$ to a "compressed domain" with dimensions $K \times N$. The minimum number of non-adaptive measurements K required to recover the original signal \mathbf{u} is related to the sparsity of its representation under \mathbf{A} but not to the original dimensions of \mathbf{A} . In the setting of the source monitoring problem this is a powerful result that indicates the amount of permissible compression is independent of the number of recording stations used for monitoring. As a practical rule of thumb in the noiseless case, more than four non-adaptive measurements per non-zero coefficient in \mathbf{m} (or \mathbf{b}) is the minimum required number of samples to recover \mathbf{u} from its compressed version $\Phi \mathbf{u}$ (Candes and Wakin, 2008). CS also imposes the following conditions on the signal \mathbf{u} and the matrix \mathbf{A} : 1) \mathbf{m} (or \mathbf{b}) is a sparse representation of \mathbf{u} or has non-zero coefficients that decay quickly in absolute magnitude (i.e., \mathbf{u} is compressible under \mathbf{A}); and 2) $\Phi \mathbf{A}$ also complies with the RIP. The first condition can be easily met since the observations vector

\mathbf{u} possesses a sparse representation under \mathbf{A} for most monitoring scenarios of interest (see previous chapter). However, the condition related to the product $\Phi\mathbf{A}$ is not met. The high coherence between the blocks $\mathbf{A}_j[i]$ hinder the recovery of sparse solutions (Vera Rodriguez et al., 2012). Given that a practical implementation of the source parameter inversion is far from the settings where the theoretical results of CS can be used to analyze the recovery of optimal solutions, I resort to numerical modelling to assess the applicability of CS in the source parameter inversion problem.

5.3 Numerical Modeling

5.3.1 Seismic monitoring with dense receiver coverage

The first numerical example highlights the advantages of the CS approach in a dense network of recording stations. The monitoring experiment consists of a network of 441 three-component receivers deployed on surface to monitor a grid of $45 \times 45 \times 11 = 22,275$ virtual sources (Figure 5.2). The separation between virtual sources is 5 km in the three coordinate directions. Green's functions in real applications should be ideally full waveform traces. In this example, however, the analysis is simplified by using only direct P- and S-wave arrivals in a homogeneous medium computed with ray tracing theory. Considering the longest arrival time of S-waves from the ray tracing results and a sampling rate $\Delta t = 0.5$ s, the length of a Green's function for a single component of a receiver is set at 603 samples (301.5 s). To study the impact of the compression on the detectability and accuracy of the source parameters, I select 500 sets of Green's functions with random locations inside the grid and generate synthetic wavefields for each location from a set of six random numbers as the independent elements of the moment tensor. Finally, the synthetic displacements are padded with zeros on both ends to produce observations with duration of 331.5 s with a common origin time at 15 s (this implies a processing window of 30 s or $N_w = 60$). The sensing matrix Φ contains independent, identically distributed (iid) random variables drawn from a Gaussian probability distribution, and five non-adaptive measurements are considered per non-zero coefficient in the expected solution. In other words, assuming a single source will be represented in each observations vector \mathbf{u} , only $K = 30$ samples (6 moment tensor coefficients in one source \times 5 non-adaptive measurements per coefficient) are used to invert for source parameters using the CS approach. If explicitly expressed, the matrix \mathbf{A} with the library of Green's functions should occupy a total memory size of ~ 52 Terabytes. On the other hand, the compressed version of the library of Green's functions ($\Phi\mathbf{A}$) requires a memory size of ~ 2 Gigabytes, and the inversion is solved using a modified version of Block Orthogonal Matching Pursuit (BOMP) (Vera Rodriguez et al., 2012) that only requires the compressed matrix $\Phi\mathbf{A}_0$ of ~ 32 Megabytes (\mathbf{A}_0 is a version of the dictionary \mathbf{G}_0

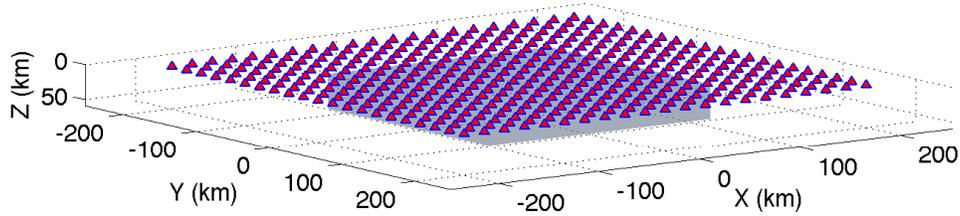


Figure 5.2: Distribution of recording stations (blue-red triangles) and grid of virtual sources (grey volume) used in the numerical example.

with columns that are energy-normalized) as input. Hence, with CS the inversion process becomes easily tractable using a desktop computer, e.g. a 4 cores desktop computer with 8 GB in RAM and 2.9 GHz processor speed in this study. In this numerical example I aim to evaluate the efficacy of the inversion method in the source parameter estimation when exact solutions exist, assuming accurate Green's functions and no noise.

The errors in the estimated source parameters for the numerical experiment are presented in Figure 5.3. For 500 trials, 51.6% of the sources were detected with no error in the three source parameters (black circles), while 86.4% had acceptable normalized misfits of 0.35 or below (black + blue circles). Finally, 5% of the results showed normalized misfits between 0.35 and 0.8 (red circles) that, in a practical setting, correspond to undetected sources. In this experiment 8.6% of the sources had normalized misfits > 0.8 (not presented in Figure 5.3 due to their large errors). The average processing time to obtain a solution was 41.57 s. In comparison, the estimated time required for the modified BOMP to browse a solution in the full 52 Terabytes library of Green's functions is about 90 days. For the processing window of 30 s, this result can be considered as near real-time for this hypothetical monitoring setting. Further improvements to the speed of the inversion algorithm could be achieved by an efficient implementation using parallel processing.

5.3.2 Simultaneous events

An important benefit of matching pursuit methods when compared to other automatic source parameter inversion algorithms is the possibility to estimate multiple sources embedded in the same observations vector. The second numerical example is aimed to detect the effect of the inversion method over the estimation of simultaneous events. This objective is addressed by using the same hypothetical monitoring setting of the previous section and the modified BOMP. Two different settings are studied, the first corresponds to two seismic sources with the same source mechanism and location but different origin times. The second scenario consists of the two sources with the same source mechanism and origin time but different locations. For the first case, assume the seismic sources take place at location n

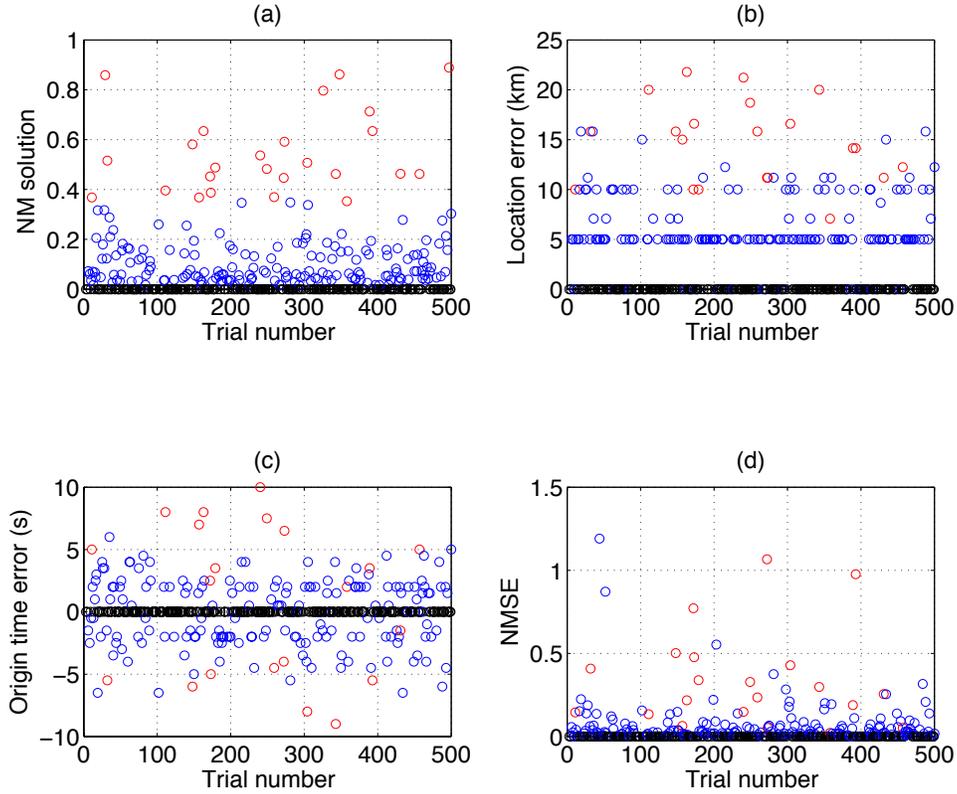


Figure 5.3: Estimation errors for 500 simulations of source parameter inversion using CS principles. (a) Normalized Misfit of the solution $NM = \|\mathbf{u} - \hat{\mathbf{u}}\|_2^2 / \|\mathbf{u}\|_2^2$. (b) Location error $= \|\mathbf{x} - \hat{\mathbf{x}}\|_2$, where \mathbf{x} is the true location vector of the source. (c) Origin time error $= OT - \hat{OT}$, where OT is the true origin time. (d) Normalized Mean Square Error of the estimated moment tensor $NMSE = \|\mathbf{m} - \hat{\mathbf{m}}\|_2^2 / \|\mathbf{m}\|_2^2$. Variables with a hat denote estimations from the inversion.

with origin times t_1 and t_2 . Setting the acquisition time of the first sample in the observations vector \mathbf{u} to zero, the origin times for the two seismic events in terms of the delay index j are $s_1 = t_1/\Delta t$ and $s_2 = t_2/\Delta t$, respectively. Random numbers are used to simulate a source mechanism for both sources. Using the CS approach the sources are simultaneously inverted by applying various differences in origin time between them. When the origin time difference is zero there is technically only one event and the algorithm detects it estimating a moment tensor that is double the size of each individual source (Figure 5.4a). As the difference in origin time between the two sources increases, the algorithm's output begins to be affected by the presence of the two displacement fields. For small time differences the algorithm is unable to resolve the two events. In some cases the events are not even detected, while biased solutions of the estimated source parameters are given by the algorithm in others. After a certain difference in origin time, the algorithm starts to resolve the presence of the two events, however we still observe cases where the estimated solutions are biased for one or both events. For the second scenario (Figure 5.5), the two seismic sources are inverted while varying the distance between their locations in the horizontal y direction. In this case both events have the same origin time and moment tensor. When the locations of the two sources are close, the group sparsity function calculated by the modified BOMP across space and time presents only a single maximum (Figure 5.5a-top and middle plots). After the first iteration this maximum is extracted and no further sources can be identified (see Figure 5.5b-top and middle plots). When the locations of the two sources are far enough the algorithm is able to determine the presence of the two seismic events (Figure 5.5a-bottom plot). The first event is extracted after the first iteration and the second source is identified as a maximum in the second iteration (Figure 5.5b-bottom plot). One way to identify the resolvability limit where the algorithm begins to resolve the two events is by analyzing the measure

$$\rho\left(\left(\mathbf{G}_{s_1}[n]\right)^T \mathbf{G}_{s_2}[n]\right),$$

where $\rho(\cdot)$ is the spectral norm operator (Strang, 2006). As the spectral norm measurement reduces the algorithm is able to resolve the presence of the two events. In the case of the events with different origin time, the limit of resolvability is particularly evident as a sharp drop in the spectral norm measurement (see Figure 5.4); this might also partly result from the simplicity of the Green's functions used in this example. In both multiple source experiments we identify detectability issues that seemed unrelated to the resolvability of the algorithm. The detectability of the algorithm is mainly influenced by the coherence of the dictionary of Green's functions. The high coherence nature of the dictionary is expressed in the presence of multiple local maxima in the objective function (correlation between compressed Green's functions and residuals) calculated by the modified BOMP (see Figure 5.5). The modified BOMP partially improves detectability by re-visiting a number of maxima in the objective function. The objective of re-visiting only a subset of

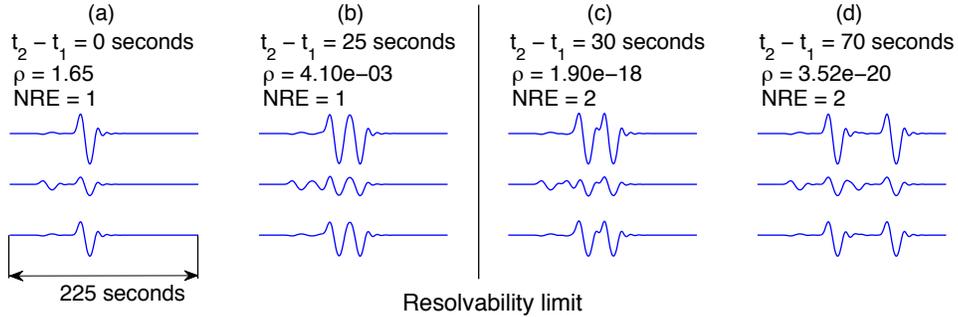


Figure 5.4: Resolvability of two seismic sources in the same vector of observations. Only observations for one receiver are presented. From top to bottom the traces correspond to the x (East), y (North) and vertical components of the receiver. NRE is the abbreviation for the number of resolved events by the inversion algorithm. a) and b) the algorithm cannot resolve the two sources. c) and d) the algorithm is able to resolve the two sources which is also reflected in a sharp decrease in the spectral norm operator $\rho(\cdot)$.

maxima values is to reduce computational time, however this is at the cost of potentially missing observed seismic events. In the limit when the modified BOMP re-visits the whole objective function this becomes equivalent to a standard grid search method. Resolvability and detectability considerations in these numerical examples arise from the combined use of sparse representation theory with CS. The following example using real earthquake data isolates the effect of the compression process on the detectability of seismic events.

5.3.3 Earthquake monitoring

The final example explores the influence of different compression rates and realizations of sensing matrices into the inversion results. This example utilizes the data from the 18 June, 2002, Caborn Indiana earthquake. The source parameters of this earthquake have been previously estimated in an earlier analysis using a combination of inversion methods (see Figure 4.9, Table 5.1), and also with a hypothetical real-time monitoring system based on a sparse representation (Chapter 3, Table 5.1). The objective of computing the source parameters of the Caborn earthquake in this chapter is not to obtain new or refined estimates of already accepted values, but to test if the CS approach can reproduce previous estimations. In the best case scenario the CS approach should reproduce the results obtained with the uncompressed dictionary (see Figure 4.10) with the advantage of improved processing speed.

For comparison purposes I test the CS approach using the same hypothetical monitoring system described in Chapter 3 (Vera Rodriguez et al., 2012). The system continuously inverts the recordings of eight broadband seismic stations located within 300 km from WVFS

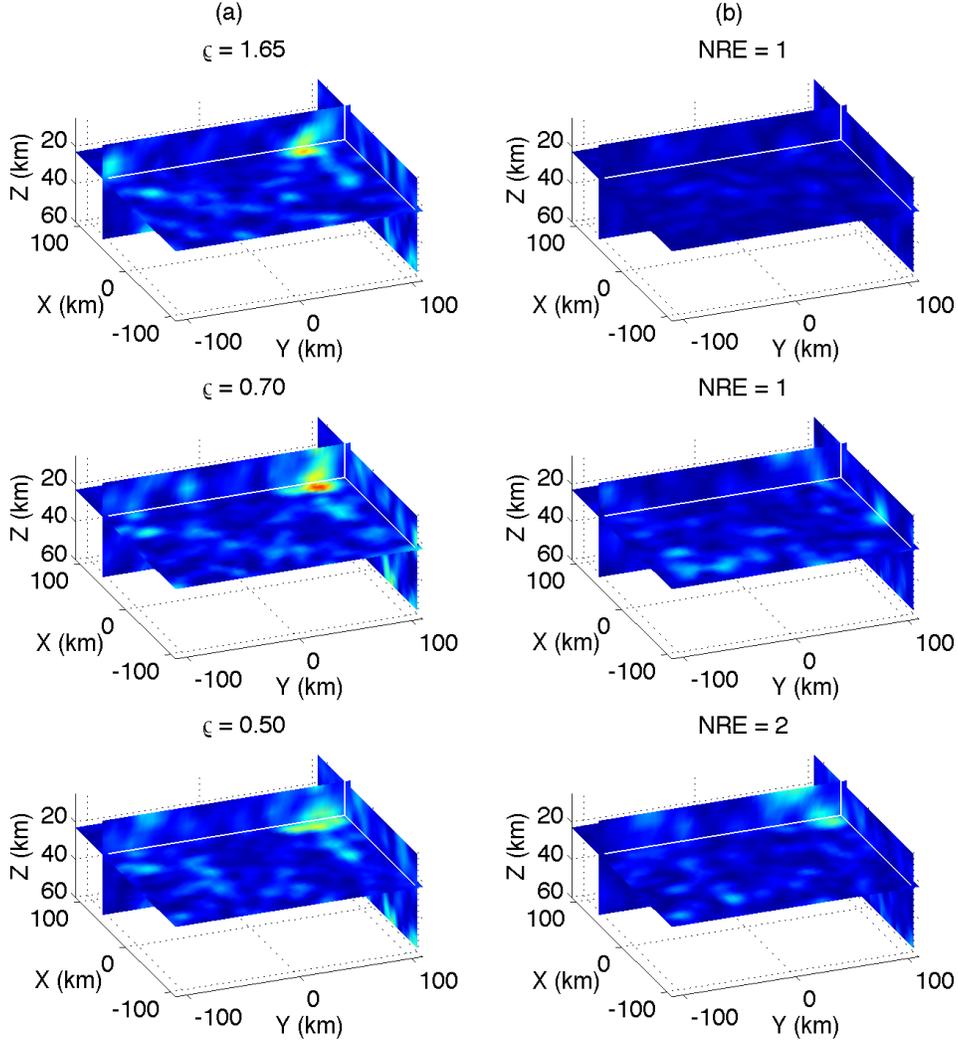


Figure 5.5: Resolvability of two seismic sources in the same observations vector. The two seismic events have the same origin time and source mechanism. These plots show a snapshot of the objective function used in the modified BOMP at the exact origin time of the sources. Top: both sources have the same location. Middle: separation between locations is 4 nodes in the y direction (20 km). Bottom: separation between locations is 6 nodes in the y direction (30 km). (a) and (b) columns are respectively iterations one and two of the modified BOMP. Each iteration finds the source parameters of a seismic source. Variable ρ is the spectral norm measurement for each case. NRE stands for number of resolved events.

(see Figure 4.8). The monitored volume comprises 2,205 virtual sources arranged in a regular grid of $21 \times 21 \times 5$ nodes with separations of 5 km in the three coordinate directions. Both Green's functions and observations are rotated to a radial, transverse, vertical system, and the velocity seismograms are band-limited to the frequency range 0.01-0.1 Hz. Green's functions are calculated with the reflectivity method of Randall (1994) using a three-layer 1D velocity model (see Table 4.1). The duration of a Green's function is 150 s, that corresponds also to 150 samples given the sampling rate of 1 s in the seismic traces. The processing window is $N_w = 30$ samples, hence the total length of a vector of observations \mathbf{u} is 180 samples. To simulate the continuous monitoring of seismic activity a time window of [-60, 240] s is selected relative to the origin time (Kim, 2003). The inversion is solved using the modified BOMP methodology with the same parameters employed in Chapter 4 that include 1) a deviatoric constraint, and 2) re-visiting the 200 highest local maxima of the objective function in the modified BOMP to look for optimal solutions. The monitoring simulation is performed six times with different degrees of compression corresponding to 5, 10, 20, 30, 40 and 120 non-adaptive samples per non-zero coefficient (NNA) in the solution vector \mathbf{m} . In each simulation 500 realizations of sensing matrix are examined, where each matrix is generated by drawing i.i.d. random variables from a Gaussian probability distribution. The CS approach recovers successfully the source parameters of the Caborn earthquake within a negligible variation margin (Table 5.1). The parameter with a slightly larger variation is the depth, though the CS results still fall between the depth of 18 km determined by Kim (2003) and 20.5 km by Vera Rodriguez et al. (2012). The main concern in the application of CS is detectability. The success rate for low NNA is around 4.6%. This small value contrasts with the 86.4% observed in the first synthetic experiment using the same amount of compression. The difference in success rate suggests an important impact in detectability due to the use of inaccurate Green's functions and the presence of noise in the real data. Reducing the amount of compression offsets the negative effect in detectability. For instance, a reasonable success rate of 96.8% is observed for $NNA = 120$, which corresponds to a compression of 16.7% relative to the uncompressed dictionary size of 100%. More remarkably, at this compression level the CS approach reproduced the source parameters for all the realizations of sensing matrix where the earthquake was detected. In other words, no effect of the compression process was visible from the results. As the NNA increases the CS results tend to converge towards the source parameters obtained without CS (Figure 5.6a). The faster the CS results converge the more advantageous the CS approach becomes since smaller dictionaries are needed to perform the inversion. For the source parameters, the convergence also implies that the results are unaffected by the choice of sensing matrix used for compression. In the case of the modified BOMP performance, the increase in the NNA also reduces the size of the search for optimal solutions (Figure 5.6b), which compensates the increase in processing time due to a lower compression rate. The gains in processing time with CS in this example

Table 5.1: Comparison of different solutions for the 18 June, Caborn earthquake. NNA stands for number of non-adaptive measurements per non-zero coefficient in the CS solution. DS stands for dictionary size after compression, where 100% corresponds to the uncompressed dictionary. Success rate refers to the percentage of times that the event was detected using the CS approach in 500 realizations of sensing matrix.

Solution	Success rate (%)	Origin time 17hr 37min + (s)	Location	Source Mechanism	
Kim (2003)	N/A	17.2	37.99°N 87.77°W Depth (18 ± 2) km		
Vera Rodriguez et al. (2012)	N/A	16.0	37.988°N 87.770°W Depth 20.5 km		
CS	NNA = 5 DS = 0.7%	4.6	16.0 ± 0.2	(37.988 ± 0.000)°N (87.770 ± 0.000)°W Depth (18.98 ± 2.35) km	
	NNA = 20 DS = 2.8%	31.4	16.0 ± 0.0	(37.988 ± 0.000)°N (87.770 ± 0.000)°W Depth (20.21 ± 1.17) km	
	NNA = 40 DS = 5.6%	60.0	16.0 ± 0.0	(37.988 ± 0.000)°N (87.770 ± 0.000)°W Depth (20.47 ± 0.41) km	
	NNA = 120 DS = 16.7%	96.8	16.0 ± 0.0	(37.988 ± 0.000)°N (87.770 ± 0.000)°W Depth (20.50 ± 0.00) km	

are rather moderate (Figure 5.6c), because the absolute size of the compressed dictionaries do not represent a substantial difference for the computer capabilities (the uncompressed dictionary size is ~ 350 MB). The CS approach is feasible in the hypothetical Caborn monitoring system and enables a large number of monitoring stations to be inverted in real-time. The benefits for real-time monitoring provided by CS would be more apparent as the number of recording stations in the monitoring system increases.

5.4 Discussion

The implementation of grid search approaches in different areas of science and engineering has become possible with the development of more powerful and less expensive computing systems. In the case of seismic source monitoring, the grid search approach still requires a

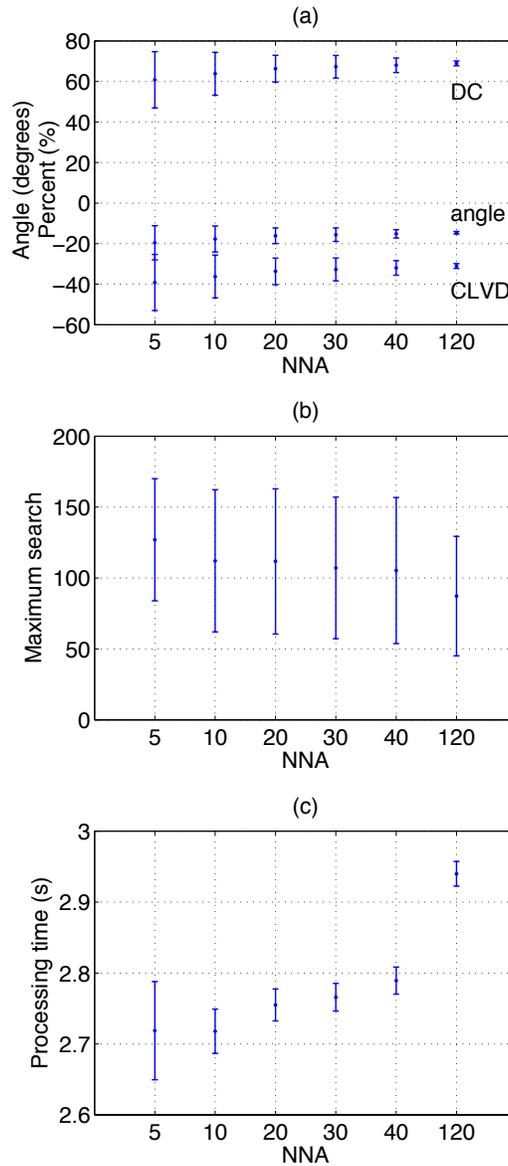


Figure 5.6: Results for 500 realizations of sensing matrix in 5 different scenarios of compression to detect and estimate source parameters for the 18 June, 2002 Caborn earthquake. NNA stands for number of non-adaptive measurements per non-zero coefficient in the solution. Vertical bars are standard deviations. a) Average source decompositions, y -axis is percentage for DC and CLVD components (Knopoff and Randall, 1970), and degrees for the angle between the slip vector and the fault plane (Vavrycuk, 2001). b) Average position of the value in the correlation function between compressed Green's functions and residuals in BOMP where the solution was found (values are sorted from maximum to minimum). c) Average processing time.

considerable amount of computing resources to handle large quantities of seismic records in a dense monitoring grid. As a reference, Tajima et al. (2002) evaluate the feasibility of the grid search approach using real data from the Berkeley Digital Seismic Network. Considering the computational resources available for their analysis, the authors estimated a total of 12 min for the processing of the records from three broadband seismic stations in a grid with 160 virtual sources using a processing window of 20 seconds. They also acknowledged that the task of reading Green's functions from hard disk occupies a significant amount of the processing time, suggesting parallel processing in a cluster of workstations as an alternative to achieve the real-time response of the system. Tsuruoka et al. (2009) present the results of implementation of the grid search monitoring approach in Japan. In this case, real-time response was achieved in a monitoring grid consisting of 6,875 virtual sources using three seismic stations. The authors further suggest that improvements in computer performance would allow the monitoring of more than 7,000 virtual sources using also a larger number of seismic stations. The sparse representation inversion presented in Chapter 4 performs with an ample margin the real-time inversion of the Caborn, Indiana earthquake using 2,205 virtual sources with the records from eight broadband seismic stations. Although less virtual sources are employed in this latter study the use of more seismic stations puts this result in a comparable setting to those of the previous grid search experiments. In contrast to these results, the compressed domain inversion presented in this chapter can perform in real-time using the same computational resources for a considerably larger amount of seismic records. Also, the reduction in variable size permits the use of denser monitoring grids as illustrated in the first synthetic example of this chapter, where a grid of 22,275 virtual sources is inverted for the case of 441 receivers. Simultaneous source parameter estimation approaches are an attempt to improve the response time of monitoring systems based on the two-step estimation procedure. For example, Tajima et al. (2002) presents the grid search approach as an alternative to the two-step system deployed in the Berkeley Seismological Laboratory, which relied from information about origin time and source location produced by the United States Geological Survey. Transmission of information about origin time and location from one place to another directly impacts the response time to provide estimations that include the moment tensor of the seismic event. Table 5.2 presents a comparison between the general steps involved in the source parameter determination followed by the two-step and the simultaneous approaches. In the best case scenario, the response time when the two-step approach is performed automatically by the same computer can be similar or better to the simultaneous approaches. However, even in this case, the source parameters obtained by the simultaneous algorithms are the best overall, while the two-step solutions present unrelated estimation errors. A larger amount of recordings can potentially be used in real-time if the moment tensor inversion method employed in the two-step procedure does not require full waveform information. Nevertheless, the maximum number of recordings will always be

limited at some point by the computational resources. The compressed domain inversion is, on the other hand, not directly limited by computational resources, as the maximum amount of compression does not depend on the number of input records but on the number of expected seismic sources in the solution. A major advantage in the two-step procedure is its flexibility, for instance, more accurate solutions can be obtained by manually manipulating and validating the results at each step. In the case of the simultaneous approaches, the use of fixed grids impacts the accuracy of the source locations by design, while there is no room for manual tuning of the intermediate inversion stages. Flexibility and accuracy is somehow sacrificed in favour of faster estimations in the simultaneous approaches. If the purpose of the system is real-time response, the trade-off between accuracy and speed is unavoidable, nevertheless simultaneous approaches can be specifically developed to handle efficiently this juncture.

5.5 Summary

Real-time simultaneous estimation of the source parameters has applications in different areas of geophysics. This chapter presents a new method for the automatic recovery of source parameters based on ideas of compressive sensing. The main advantage of this method is that a large number of recordings can be inverted without compromising real-time response. Furthermore, if the compression rate is correctly selected, the results obtained by the compressed domain methodology would be identical to those without using compression, with the added benefit of a shorter processing time. The application of compressive sensing does require a dictionary of Green's functions that embodies the properties of a low-coherency frame. Moreover, the lack of practical dictionaries that meet this assumption necessitates the use of numerical modelling as a vital tool to determine the feasibility of applying the compressive sensing approach in a given scenario. A complete implementation of compressive sensing to the source monitoring problem requires the setting of new acquisition protocols that allow the recording of the compressed measurements. Even in the absence of such protocols, a proper use of compressive sensing can be an effective tool for real-time monitoring of seismic events.

CHAPTER 5. COMPRESSED DOMAIN INVERSION

Table 5.2: Comparison of the general steps involved in the estimation of the source parameters between the two-step Vs. the simultaneous approach. Advantages and disadvantages are considered in terms of the speed to generate results and their reliability.

STEP	Two-step approach		Simultaneous approach	
	Advantages	Disadvantages	Advantages	Disadvantages
Feeding of data	Assuming continuous feeding of records from the seismic stations to the computing center in both cases			
Origin time and location estimation	Faster when only first breaks are required			Full waveform information needs to be available to start computations
Transmission	Depends on the location of the system that performs the source mechanism inversion. Best case: the source mechanism inversion is performed by the same computer. Worst case: origin time and location are received from a system in a different geographical location.		No further transmission of data required	
Start moment tensor estimation		Only after information about location and origin time is received	Continuous monitoring for the three source parameters in all available data	
Data preparation	Depends if the procedure is automatic, semi-automatic or manual, and the type of moment tensor inversion. Best case: automatic system with moment tensor inversion that does not require full waveform information Worst case: manual system with a moment tensor inversion that requires full waveform information		No further data preparation required	
Moment tensor inversion with full waveform method	More information per record to constraint the inversion	More sensitive to inaccurate velocity model Limited number of records that can be inverted in real-time	More information per record to constraint the inversion Compressed domain inversion extends the real-time limits	More sensitive to inaccurate velocity model Detectability considerations in the compressed domain inversion
Moment tensor inversion with method that does not require full waveform information	Faster results Larger monitoring volumes can be handled in real-time	Less information per record to constraint inversion (less robust than full waveform)		
Refinement of source parameters		Refinements are possible but require to re-initialize the two-step procedure	The first output is the best approximation achievable by the system. No refinement is required	
Accuracy of results	More accurate if there is manual validation and refinement but more time consuming	Source parameters are only optimal for the individual algorithms that obtained them, but not as a group. Other potential issues depend on the specific inversion methodologies.	Subjected to spatial and temporal sampling and accuracy of Green's functions. Detectability considerations in compressed inversion. Advantage: the three source parameters are optimal as a group. Disadvantage: noisy records cannot be discarded.	

CHAPTER 6

Microseismic data denoising using a 3C group sparsity constrained time-frequency transform ¹

6.1 Introduction

Microseismicity induced during hydraulic injections is characterized for its small magnitude (Maxwell, 2005; Shemeta and Anderson, 2010). This is an important reason why microseismic records often display a low signal-to-noise ratio (SNR). Accuracy and reliability of the location and other event attributes derived from microseismic traces is influenced by this strong noise content (Maxwell, 2009). Thus, noise attenuation is a desirable step in microseismic processing in order to improve the quality of subsequent processes.

In general, noise can be considered as the part of the measured signal that is not of interest and it is usually divided into coherent and non-coherent. Two different approaches can be followed to improve the SNR in microseismic experiments, either improving the acquisition equipment and/or acquisition geometry or through the use of signal processing techniques. One of the most basic signal processing methods to attenuate noise is frequency filtering. Frequency filters are effective to attenuate frequencies outside user-defined cut-off values. However, signal and noise often share frequency bands which means that part of the signal is also filtered out with the noise, while the noise might not be fully attenuated. A more advanced denoising technique requires the use of matched filters (Eisner et al., 2008). A matched filter is designed by selecting a microseismic event with high SNR to be used as a "master" or reference event. Through the cross-correlation of the "master" event with other parts of the signal, events with lower SNR, similar source mechanism and nearby location are

¹A version of this chapter has been published. Vera Rodriguez, I., Bonar, D. and Sacchi, M., 2012. *Geophysics* 77: V21-V29.

”matched” and found. Other examples of advanced denoising techniques are found in global seismology. For instance, in the analysis of earthquake precursors Sobolev and Lyubushin (2006) apply a wavelet transform and thresholding criteria in the time-frequency plane to facilitate the identification of microseismic events. Similarly, Baig et al. (2009) employ time-frequency analysis to denoise seismic noise cross-correlations. In reflection seismology, Bonar and Sacchi (2010) proposed a method for spectral decomposition to analyze seismic sections. The decomposition technique is based upon the transform of the time domain data into a sparse time-frequency map. When implemented for denoising purposes this method is equivalent to Basis Pursuit Denoising (BPDN) (Chen et al., 1998).

In this chapter, I show a procedure for attenuating non-coherent or ambient noise. Following the exposition by Bonar and Sacchi (2010), I seek a sparse representation of the microseismic signal in terms of a dictionary of complex Ricker wavelets. By imposing sparsity and misfit constraints to the decomposition process, I find a denoised version of the original signal. An extension of the time-frequency transform to 3C data is accomplished through the concept of group sparsity (Yuan and Lin, 2006; Eldar and Bolcskei, 2009). Group sparsity is a condition that arises naturally in problems such as linear regression (Yuan and Lin, 2006), brain activity imaging (Cotter et al., 2005) or the determination of the direction of arrival in antennas (Hyder and Mahata, 2010). In the field of microseismicity analysis, group sparsity has been recently used in the joint determination of origin time, location and moment tensor of microseismic events (Vera Rodriguez et al., 2012). Considering that the three components of a single receiver share the same sparsity pattern in the time-frequency plane, a group sparsity thresholding criteria (Fornasier and Rauhut, 2008) is implemented with a Fast Iterative Soft Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009) in order to simultaneously recover denoised versions of the three components. The resulting methodology has been tested with synthetic and real microseismic datasets. In all cases, the improvement in SNR with respect to the input noisy traces is clearly visible in the reconstructions obtained through the 1C and 3C sparse time-frequency transforms.

6.2 One-component (1C) sparse time-frequency transform

The decomposition of a discrete seismic trace $s[n]$ in terms of dilations and translations of a reference complex wavelet $w[n]$ is given by (Bonar and Sacchi, 2010)

$$s[n] = \sum_{i,k} a_k[i] w_k[n-i], \quad (6.1)$$

where the indexes i and k are related, respectively, to the translations and dilations of $w[n]$, and the complex coefficients $a_k[i]$ control the amplitude and phase of the wavelets. More specifically, the index k reflects a parameter/property of the wavelet that controls its dilation or frequency content. In a Ricker wavelet, for example, k represents the central frequency of the wavelet. In matrix-vector notation, equation (6.1) can be written as the convolution of a set of complex coefficient vectors $\{\mathbf{a}_k\}$ with a set of dilated versions of the reference wavelet $\{\mathbf{w}_k\}$,

$$\mathbf{s} = \sum_k \mathbf{W}_k \mathbf{a}_k, \quad (6.2)$$

where \mathbf{s} is the vector representation of the trace $s[n]$ and \mathbf{W}_k is the convolution matrix of the wavelet with dilation index k . In order to generalize this model, the convolution matrices of a set of N dilated versions of the reference wavelet can be incorporated to create a redundant dictionary \mathbf{R} (Mallat, 2008),

$$\mathbf{s} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 & \mathbf{W}_3 & \dots & \mathbf{W}_N \end{bmatrix} \begin{bmatrix} \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \vdots \\ \mathbf{a}_N \end{bmatrix} = \mathbf{R}\mathbf{a}. \quad (6.3)$$

Equation (6.3) is an under-determined system that accepts an infinite number of solutions. In other words, there are multiple possible representations of \mathbf{s} in terms of linear combinations of the columns of \mathbf{R} . A common approach to solve under-determined systems is through the damped least-squares (DLS) method. Mathematically, this entails solving the following problem

$$\hat{\mathbf{a}}_{\ell_2} = \arg \min_{\mathbf{a}} \left[\|\mathbf{s} - \mathbf{R}\mathbf{a}\|_2^2 + \lambda_{\ell_2} \|\mathbf{a}\|_2^2 \right]. \quad (6.4)$$

The cost function inside the square brackets consists of two terms. The first term is the misfit function, a measure of fidelity of fit between the observations and the synthesized data, and the second term is the ℓ_2 regularization term. Minimization of this cost function yields the DLS solution $\hat{\mathbf{a}}_{\ell_2}$. In this case, the solution vector of coefficients can be easily found utilizing the method of conjugate gradients (CG) (Hestenes and Stiefel, 1952). The trade-off parameter λ_{ℓ_2} is used to balance the relative strengths of the misfit and regularization terms. An alternative solution to the problem entails using an ℓ_1 regularization term. In other words, imposing a measure to promote sparsity on the coefficients needed to represent the seismogram. This can be accomplished by solving

$$\hat{\mathbf{a}}_{\ell_1} = \arg \min_{\mathbf{a}} \left[\|\mathbf{s} - \mathbf{R}\mathbf{a}\|_2^2 + \lambda_{\ell_1} \|\mathbf{a}\|_1 \right]. \quad (6.5)$$

Minimization of the cost function inside square brackets in (6.5) yields the sparse solution $\hat{\mathbf{a}}_{\ell_1}$. The trade-off parameter λ_{ℓ_1} balances the fidelity of fit versus the ℓ_1 regularization term. Sparse solutions from ℓ_1 regularization are estimated using the Fast Iterative Soft Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009) (Appendix A). The denoised version of the input data is obtained via

$$\hat{\mathbf{s}} = \mathbf{R}\hat{\mathbf{a}}_{\ell_p}, \quad (6.6)$$

where p can be either 1 or 2 depending on which regularization approach we are using. In Figure 6.1 I provide a synthetic example where I examine the two denoising schemes. In this case, a complex Ricker wavelet is used as the reference wavelet to create the dictionary \mathbf{R} (Figure 6.2), where the dilation index k is related to the central frequency that describes each wavelet in \mathbf{R} . Complex wavelets can be easily computed using the Hilbert Transform. Notice that, although the construction of the dictionary requires a fixed reference wavelet, in this case a complex Ricker wavelet, the dictionary is not adaptive and can be used to represent a wide range of source signatures in the input data through translations and changes in phase and frequency content of the reference wavelet. In order to produce a fair assessment of the denoising results, I have determined λ_{ℓ_1} and λ_{ℓ_2} using the χ^2 criterion for the goodness of fit (see e.g., Aster et al., 2005). Representations using DLS tend to be "spread" in the solution (Figure 6.1-c), whereas, sparse representations are well localized and produce cleaner signal estimations using a smaller number of non-zero coefficients (Figure 6.1-e). For the purpose of denoising, it is clear that sparse representations produce improved results when compared to DLS representations. Examples of application of the 1C sparse time-frequency transform to synthetic and real microseismic data are also presented in Vera Rodriguez et al. (2011).

6.3 Extension to the three-component (3C) case via group sparsity

The extension of the sparse time-frequency transform to 3C data is accomplished through the concept of group sparsity (Yuan and Lin, 2006; Eldar and Bolcskei, 2009). Group sparsity arises naturally in a wide variety of problems, for example, in microseismic monitoring a method based in group sparsity has been proposed for the simultaneous recovery of origin time, location and seismic moment tensor of microseismic events (Vera Rodriguez et al., 2012). In the present application, the concept of group sparsity is incorporated by considering the three components of a recording station as a multiple measurement vector (MMV) (e.g., Cotter et al., 2005; Hyder and Mahata, 2010). In other words, it is assumed that the three components of a single receiver are measuring the same information (seismic arrivals),

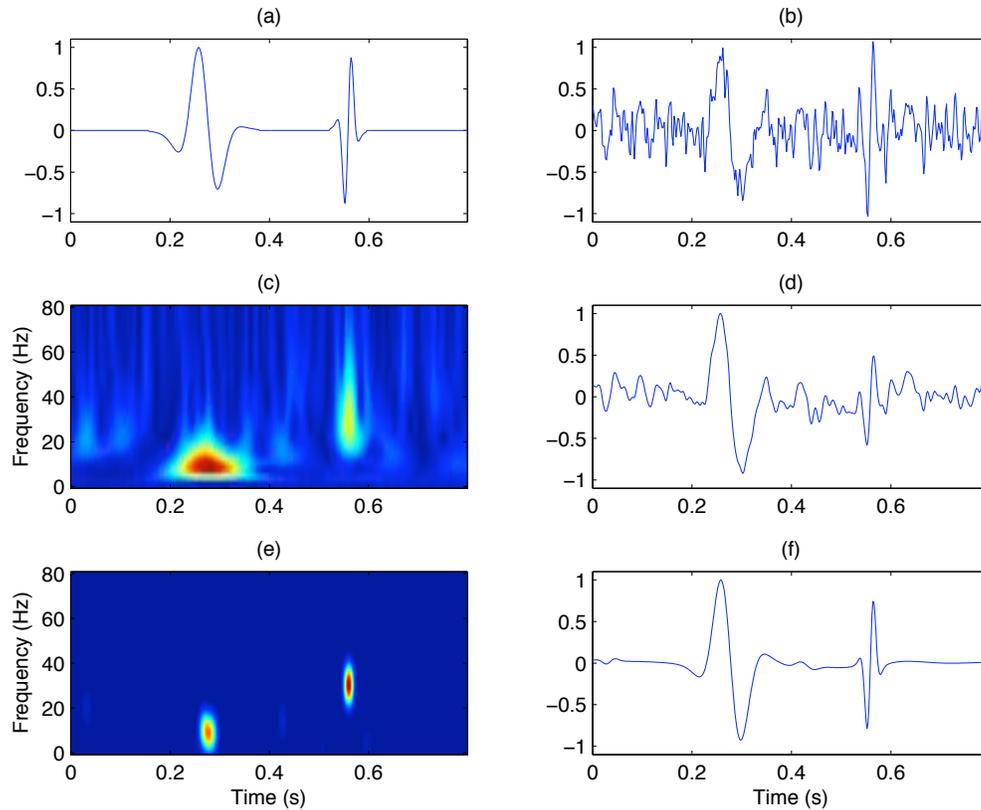


Figure 6.1: a) Synthetic trace composed of two arrivals with arbitrary phase and frequency content. b) Synthetic trace contaminated with random noise. c) Inversion results using the damped least-squares method (equation (6.4)). d) Data reconstruction from the time-frequency map in c). e) Inversion results using the ℓ_1 -norm regularization (equation 6.5). f) Data reconstruction from the time-frequency map in e).

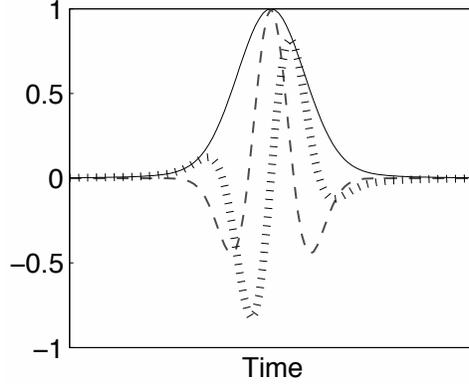


Figure 6.2: Reference complex Ricker wavelet. Real part (dashed line), imaginary part (dotted line) and envelope (solid line).

even though, the amplitudes of the measurements in each individual component might be different due to the arrival polarizations. The MMV assembled from the input data is

$$\mathbf{S} = \begin{bmatrix} \mathbf{x} & \mathbf{y} & \mathbf{z} \end{bmatrix}, \quad (6.7)$$

where \mathbf{x} , \mathbf{y} , and \mathbf{z} are the recordings from the three components of one receiver. Hence, the 3C version of equation (6.5) is

$$\hat{\mathbf{A}}_{\ell_{2,1}} = \arg \min_{\mathbf{A}} \left[\|\mathbf{S} - \mathbf{R}\mathbf{A}\|_{2,2}^2 + \lambda_{\ell_{2,1}} \|\mathbf{A}\|_{2,1} \right], \quad (6.8)$$

where the columns of $\hat{\mathbf{A}}_{\ell_{2,1}}$ are the sparse representations in the time-frequency plane of the input three components. The symbol $\|\cdot\|_{p,q}$ denotes the mixed $\ell_{p,q}$ -norm defined as (Kowalski and Torr sani, 2009)

$$\|\mathbf{X}\|_{p,q} = \left(\sum_n \left(\sum_k |x_{n,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}}, \quad (6.9)$$

where $x_{n,k}$ is the n -th row, k -th column element of the matrix \mathbf{X} . In order to apply FISTA to MMVs a multidimensional shrinkage-thresholding operator (Fornasier and Rauhut, 2008) is introduced into the inversion routine (see Appendix A). The denoised version of the 3C data is finally recovered via

$$\hat{\mathbf{S}} = \mathbf{R}\hat{\mathbf{A}}_{\ell_{2,1}}. \quad (6.10)$$

Both 1C and 3C versions of the sparse transform display comparable denoising results when applied to synthetic data (Figure 6.3). The real advantage of the 3C approach is evident in cases where the wave arrival, although present in the three components, is only above

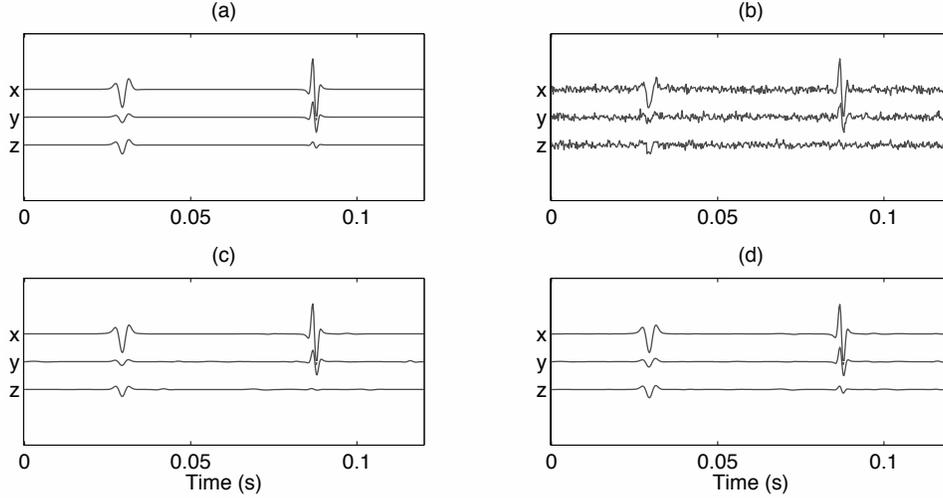


Figure 6.3: Comparison of denoising results between the 1C and 3C versions of the sparse time-frequency transform. a) Synthetic 3C data. b) Synthetic data contaminated with random noise. c) Denoised data using the 1C sparse time-frequency transform. d) Denoised data using the 3C sparse time-frequency transform.

the prevailing noise content in two of them. Using information of a single trace at a time does not allow the 1C sparse transform to recover the arrival where it is below the noise level (Figure 6.3-c). The 3C sparse transform, on the other hand, uses information from the components where the arrival is above the noise in order to identify and reconstruct the arrival in the component where the noise is dominant (Figure 6.3-d).

6.4 Application to denoise microseismic data

The 3C sparse transform is first tested with a synthetic microseismic dataset. The synthetic data consists of direct P and S arrivals computed in a 1D velocity model (Figure 6.4). The amplitude variations due to the source radiation pattern are included in the modelling, corresponding to a pure double couple source mechanism. Wave arrivals are computed in a vertical array of eight 3C receivers located at a horizontal distance of ~ 400 m from the source location. The resulting synthetic data is dominated by the amplitudes of the S arrivals (Figure 6.5, first row). The SNR is defined by

$$\text{SNR} = \frac{\mu_s}{\sigma_n}, \quad (6.11)$$

where σ_n is the standard deviation of the noise and μ_s the expected signal amplitude. For simplicity, the expected signal value is considered as the maximum absolute value of the data

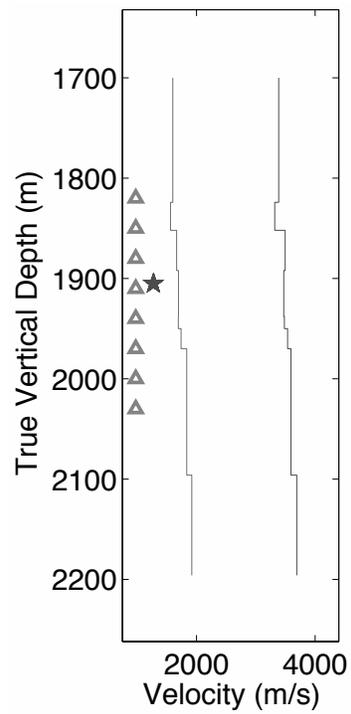


Figure 6.4: Velocity model used to compute the synthetic microseismic data. Triangles denote receiver depth locations. The star denotes the source depth.

samples in the synthetic traces. Hence, a $\text{SNR} = 5$ is set in the data (Figure 6.5, second row). Applying a low-pass filter to the noisy data provides some visible improvement in the SNR, however, the signal of interest is also negatively affected in its frequency content (Figure 6.5, third row). The denoised version of the data using the 3C sparse transform on the other hand, preserves the frequency content of the wave arrivals while removing also the random noise (Figure 6.5, fourth row). The advantage of denoising using the 3C sparse transform is particularly evident in the z component of this example, where the weak S arrivals in the first 6 receivers are clearly below the noise level in the noisy data, but are partially recovered by the transform (Figure 6.5, column c). Another important property of the transform is its ability to preserve the amplitude information of the wave arrivals. Amplitude information is important for analyses of, for example, polarization angles and source mechanisms. Polarization angles are usually estimated through hodogram analysis (Figure 6.6). Random noise introduces uncertainty in the estimation of polarization angles by causing the hodogram curves to deviate from the actual wave arrival angles (Figure 6.6 column b). The effect of reducing the frequency content in the input data is to smooth the hodogram curves, which can allow a more robust estimation of polarization angles (Figure 6.6 column c). Similar effects are observed when the traces are denoised using the 3C sparse transform (Figure 6.6 column d), with the advantage that the amplitude and frequency content of the arrivals of interest are preserved. Notice also that low-pass frequency filters introduce apparent delays in the wave arrivals, which are not present in the sparse transform.

Figures 6.7 through 6.11 present the results of applying the 3C sparse transform to a real microseismic dataset. The time-frequency maps obtained from the sparse transform provide an alternative domain to analyze the microseismic data (Figure 6.7). Furthermore, the time-frequency analysis of the microseismic traces can reveal variations in the dominant frequency of the wave arrivals in different receivers. In the example shown in Figure 6.7, the arrivals in the deeper 3 receivers display a lower dominant frequency than in the rest of the array of receivers. If time picks were to be obtained from these time-frequency maps, different frequency slides can be analyzed to determine where the onset of the arrivals is more evident. In this example, the maximum dominant frequency observed in all receivers for the arrivals around 60 ms is used to low-pass filter the raw microseismic data (Figure 6.8b, 6.9b, 6.10b and Figure 6.11 column b). This frequency filter is not optimum however for the receivers where the dominant frequency is lower. Additionally, the component of the random noise that is below the cut-off frequency of the filter is preserved in the data. As a consequence, the low-pass filtered traces display higher noise content before the first arrivals when compared with the traces that were denoised with the sparse transform. The presence of cleaner wave arrivals in the denoised traces is of advantage for processes such as automatic time-picking, selection of hodogram windows and amplitude stacking. After the first arrival, the sparse transform retains the oscillations of the signal that represent potential information. For

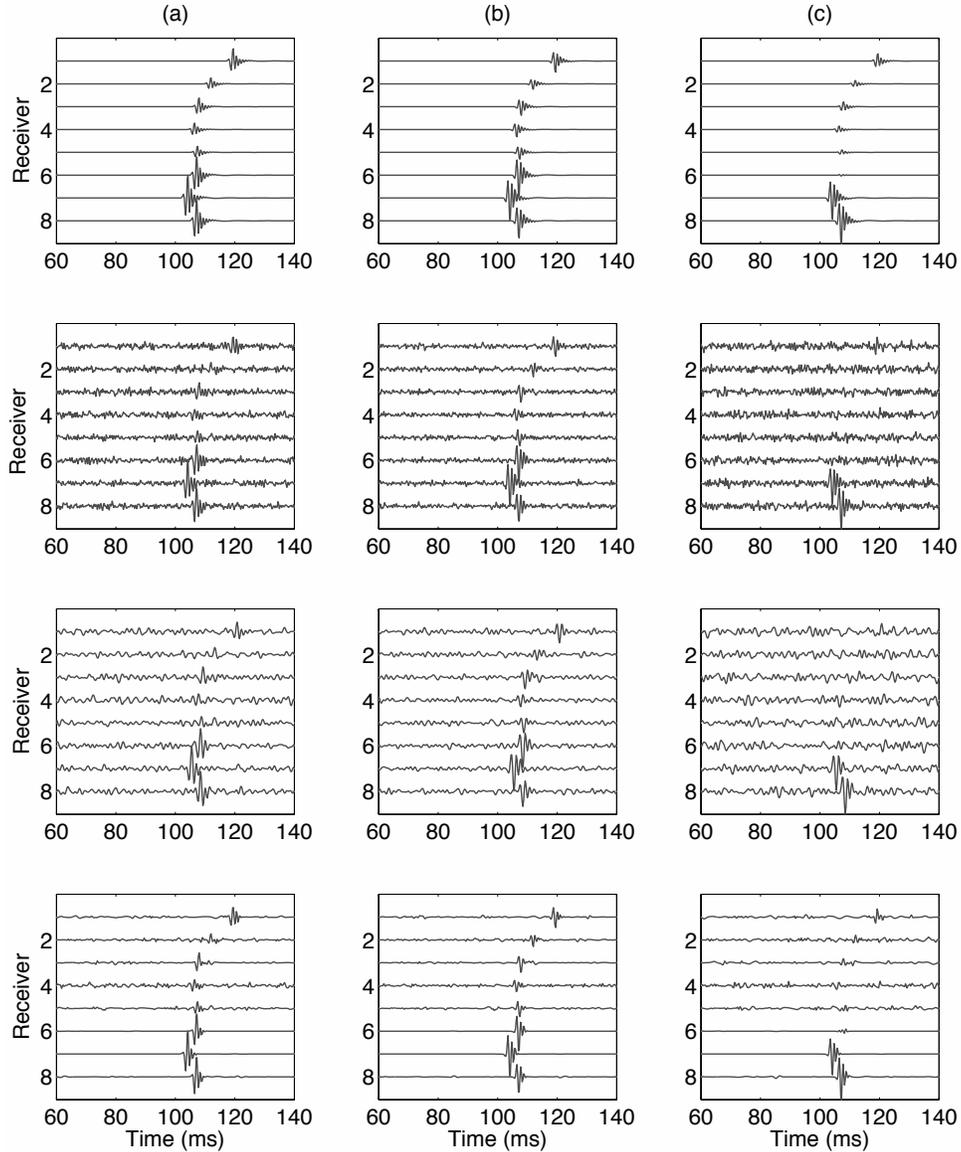


Figure 6.5: Denoising results in a synthetic microseismic dataset. Columns a), b) and c) correspond to the x , y and z components, respectively. From top to bottom, the first row is the synthetic microseismic data. The second row is the synthetic data contaminated with random noise. The third row is the noisy data filtered with a low-pass frequency filter. The fourth row is the denoised data using the 3C sparse transform.

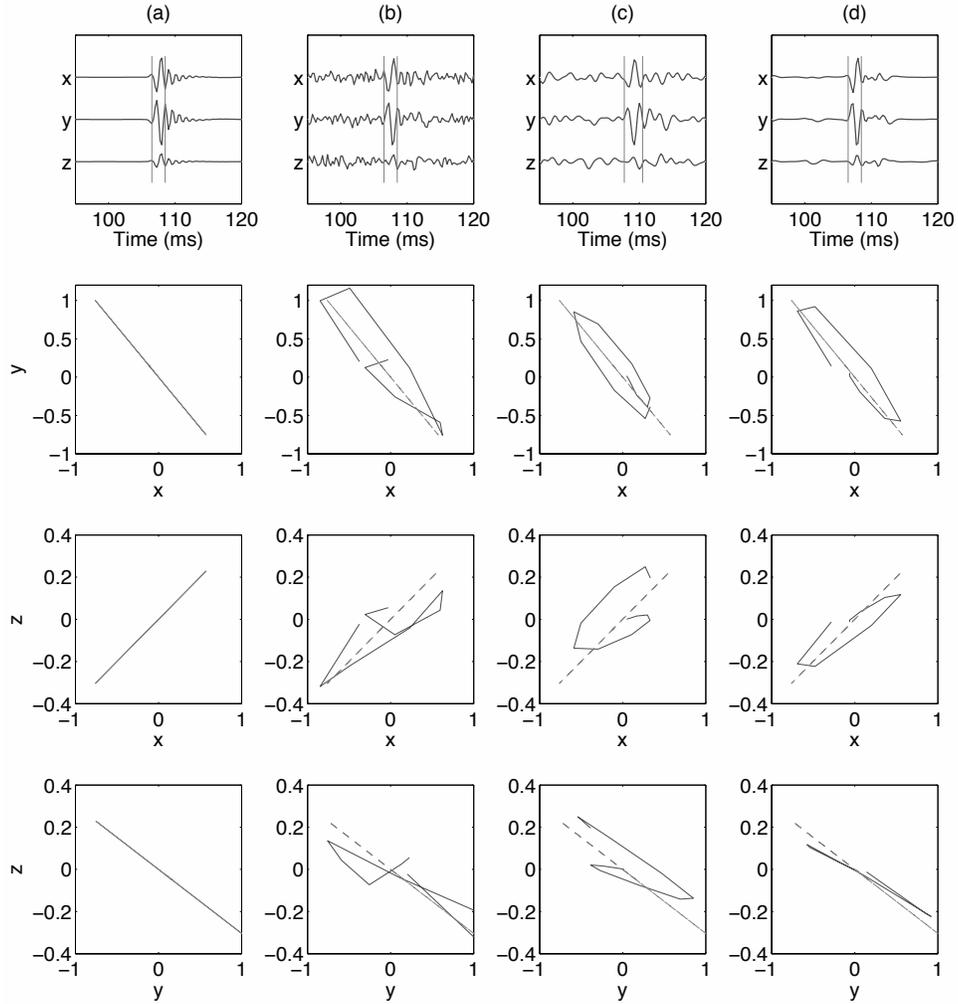


Figure 6.6: Comparison of hodogram analysis for receiver three of the synthetic microseismic dataset of Figure 6.5. Column a) synthetic data. Column b) synthetic data contaminated with random noise. Column c) noisy data filtered with a low-pass frequency filter. Column d) denoised data using the 3C sparse transform. Vertical lines in the first row delimit the hodogram window. Hodogram curves in column a) are repeated with dotted lines in columns b), c) and d) for comparison.

example, consider the arrival around ~ 110 ms in the x component of receiver 12 (Figure 6.8). While the arrival is noticeable in both the denoised and low-pass filtered versions of the data, the denoised trace is visibly cleaner before the onset. Similarly, the wave arrivals between 100 ms and 150 ms of the y component (Figure 6.9) display cleaner onsets in the denoised results. The amount of noise rejected by the transform is related to the sparsity imposed into the time-frequency representation of the data, which is controlled by the trade-off parameter $\lambda_{\ell_2,1}$. By normalizing the maximum absolute amplitude of the input data to 1, the estimated optimum value for $\lambda_{\ell_2,1}$ oscillated between ~ 0.9 and ~ 1.3 in both synthetic and real data examples analyzed in this study. The best $\lambda_{\ell_2,1}$ value for a particular dataset can be found by constructing a Pareto curve (e.g., Berg and Friedlander, 2008) or by simple trial and error. The preservation of the amplitude and phase properties of the arrivals in the input data is corroborated by hodogram analysis. Hodogram curves for receiver 7 of the real dataset (Figure 6.11) show smoother behaviour in the denoised traces than in the original and low-pass filtered versions of the data. The low-pass filtered version of the data is noisier before the onset of the arrivals and displays a lower frequency content that is translated into a small delay in the selected time window for hodogram analysis (Figure 6.11, first row). Cleaner onsets in the wave arrivals facilitates the selection of the hodogram windows improving the results of automatic algorithms that perform this task.

6.5 Summary

Processing of microseismicity records is a challenge due to, among other problems, the low signal-to-noise ratio displayed by the signals of interest. Application of techniques and processing tools developed in other knowledge areas can be of great advantage to deal with microseismicity records. The method presented in this chapter is a time-frequency transform constrained to impose sparsity into the time-frequency representation of the signal. When compared against the equivalent ℓ_2 -norm transform, the sparsity constrained transform using the ℓ_1 -norm displays improved results for denoising purposes. This transform can be implemented component by component (1C algorithm) or by simultaneously utilizing the three components of a receiver (3C algorithm). In particular, the 3C version of the sparse transform is advantageous in cases where the wave arrival exists in the three components of a receiver but the amplitude of the arrival is below the noise level in one component. In such cases, the 3C algorithm uses information from the components where the arrival amplitudes are above the noise level to recover the amplitude information in the other component. The results observed in synthetic and real data show the efficacy of the method to remove ambient noise without significantly affecting the frequency content and/or amplitude of the signal of interest. Hodogram analyses of the denoised results further verifies the preservation of amplitude and phase information after the denoising process. Additionally, denoised results

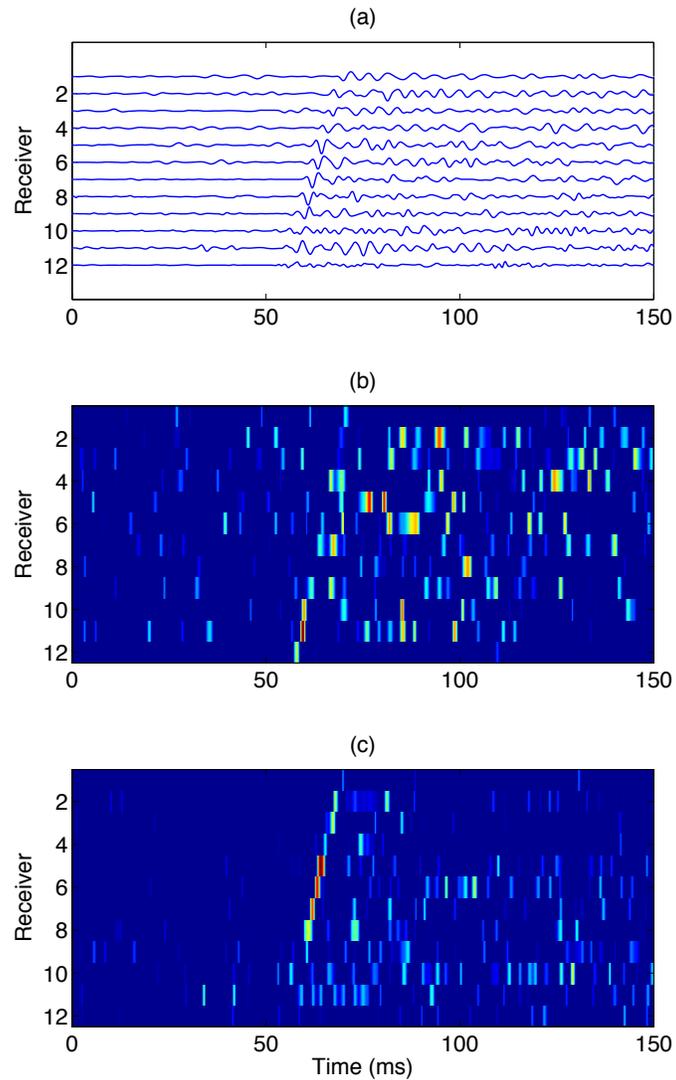


Figure 6.7: a) Denoised z component of the real microseismic dataset. b) Time-frequency map at central frequency 222 Hz for the data in a). Time-frequency map at central frequency 297 Hz for the data in a).

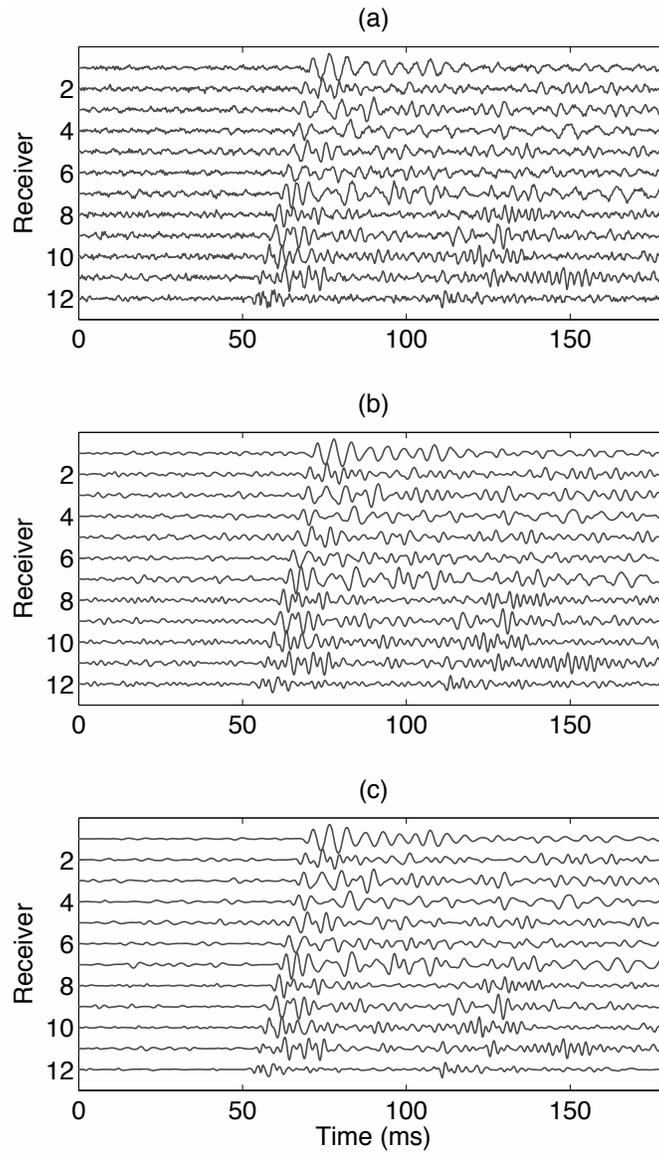


Figure 6.8: Denoising results in the x component of the real microseismic data. a) Raw microseismic data. b) Raw data filtered with a low-pass frequency filter. c) Denoised data using the 3C sparse transform.

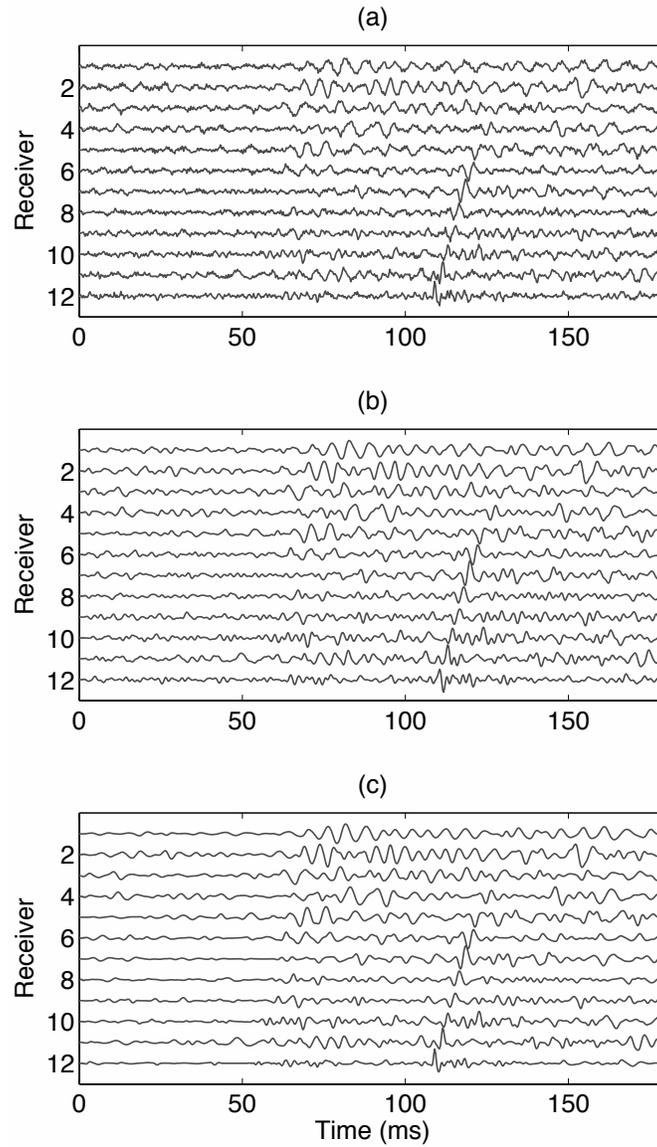


Figure 6.9: Denoising results in the y component of the real microseismic data. a) Raw microseismic data. b) Raw data filtered with a low-pass frequency filter. c) Denoised data using the 3C sparse transform.

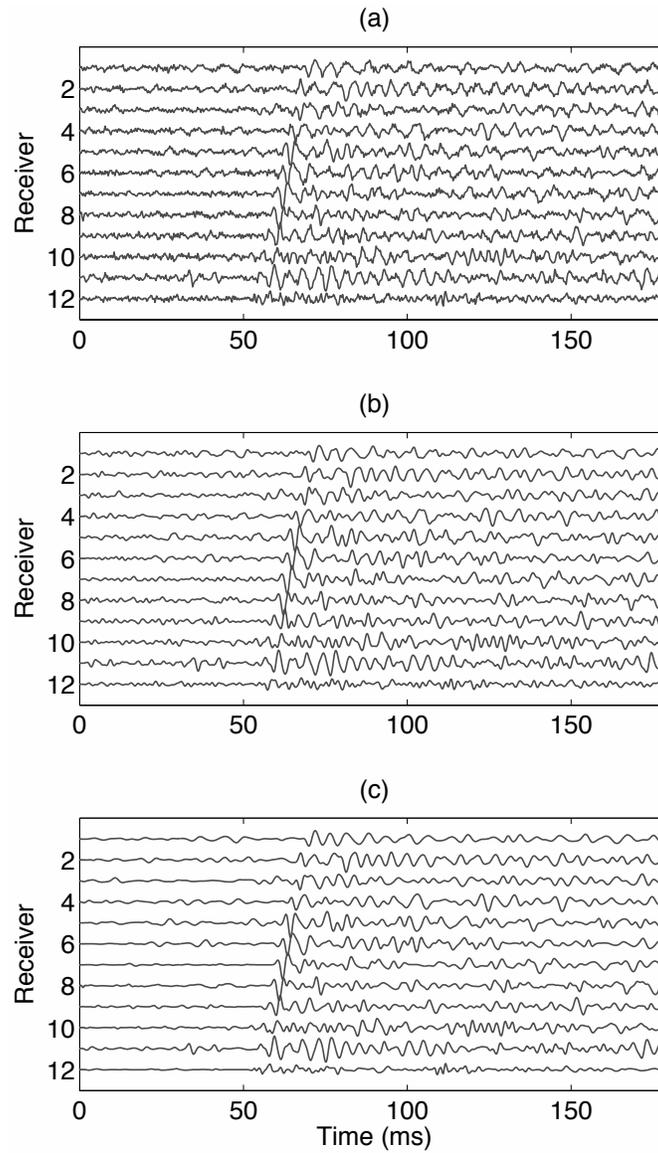


Figure 6.10: Denoising results in the z component of the real microseismic data. a) Raw microseismic data. b) Raw data filtered with a low-pass frequency filter. c) Denoised data using the 3C sparse transform.

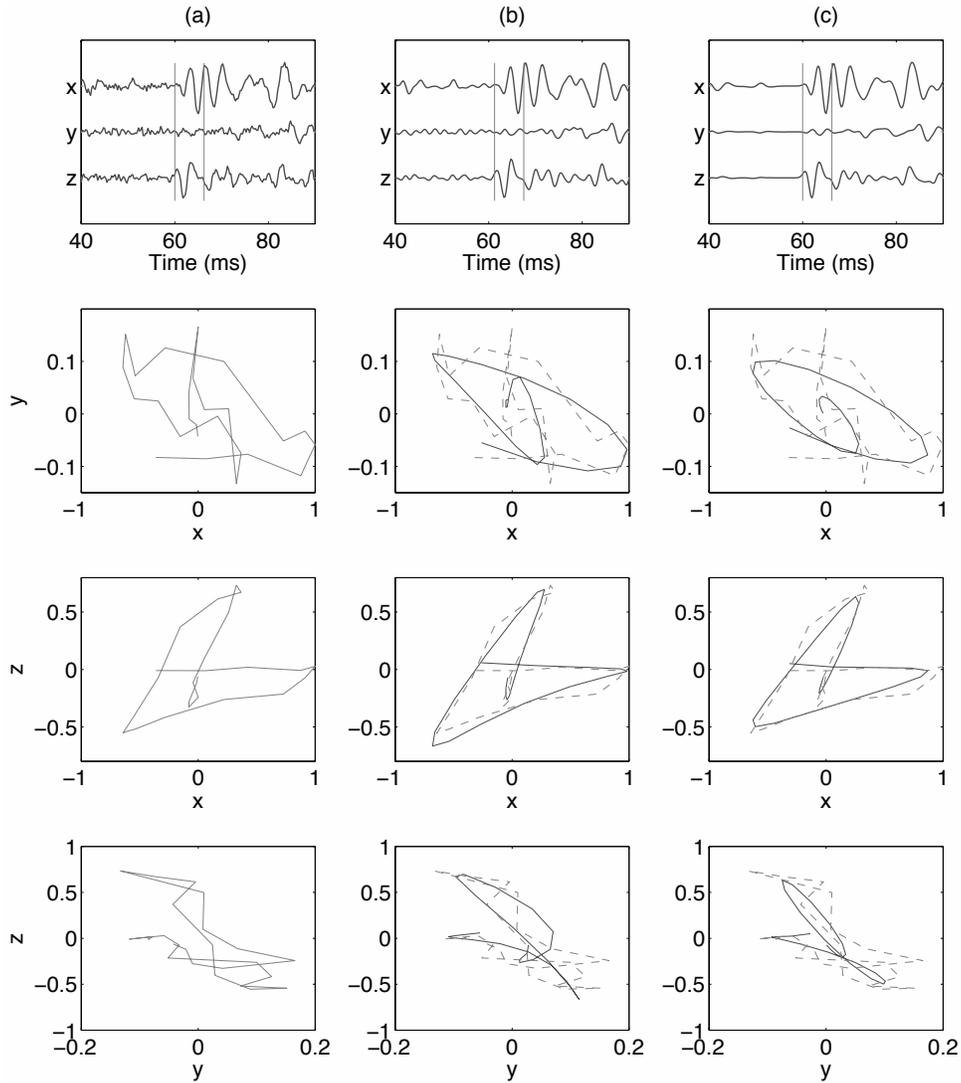


Figure 6.11: Comparison of hodogram analysis with real microseismic data (receiver 7). Column a) raw data. Column b) data filtered with a low-pass frequency filter. Column c) denoised data using the 3C sparse transform. Vertical lines in the first row delimit the hodogram window. Hodogram curves in column a) are repeated with dotted lines in columns b) and c) for comparison.

displayed cleaner onsets useful for manual or automatic selection of the hodogram windows and higher frequency content than low-pass filtered versions of the original data. This extra information allowed the hodogram curves of the denoised data to display a smooth behaviour keeping the main trends of particle movement. Through the use of the denoising method in a pre-processing stage, improved results can be expected in the estimation of polarization angles, automatic time-picking and amplitude stacking of microseismic traces.

CHAPTER 7

Improvements in automatic time-picking using non-linear inversion constraints ¹

7.1 Introduction

Event location is an important step in the processing of microseismic data. For algorithms that determine location based on P- and S-wave arrival times, accurate time-picking is crucial. For small datasets accurate time-picks can be obtained by human interaction, however for large datasets, manual time-picks can be impractical, especially if a quick response is required. Different methodologies have been proposed to automate the process of time-picking with the short-term-average/long-term average (*STA/LTA*) filter (Allen, 1978) being the most common. The *STA/LTA* filter acts over a characteristic function (*CF*), which is usually the energy of the seismic trace. Allen (1978) also uses a *CF* that takes into consideration the time derivative of the seismic trace, giving it a weight that depends on the signal-to-noise ratio (*SNR*). A disadvantage in the use of the *STA/LTA* filter resides in the rate of build-up of the curve when an arrival is detected. If the *SNR* is high, the *STA/LTA* curve will show a sharp increase at the location of the onset of the arrival, however as the *SNR* lowers, the *STA/LTA* curve displays a slower build-up that introduces error in the determination of the onset of the arrival (Figure 7.1).

The relatively recent growth of interest in microseismicity analysis in the petroleum industry opens an area of opportunity to apply the vast amount of signal processing knowledge previously developed in signal analysis and imaging processing. In this chapter, a step into this area of opportunity is given by implementing an automatic time-picking technique using

¹A version of this chapter has been published. Vera Rodriguez, I., Bonar, D. and Sacchi, M., 2011. CSEG Recorder 36: 26-28.

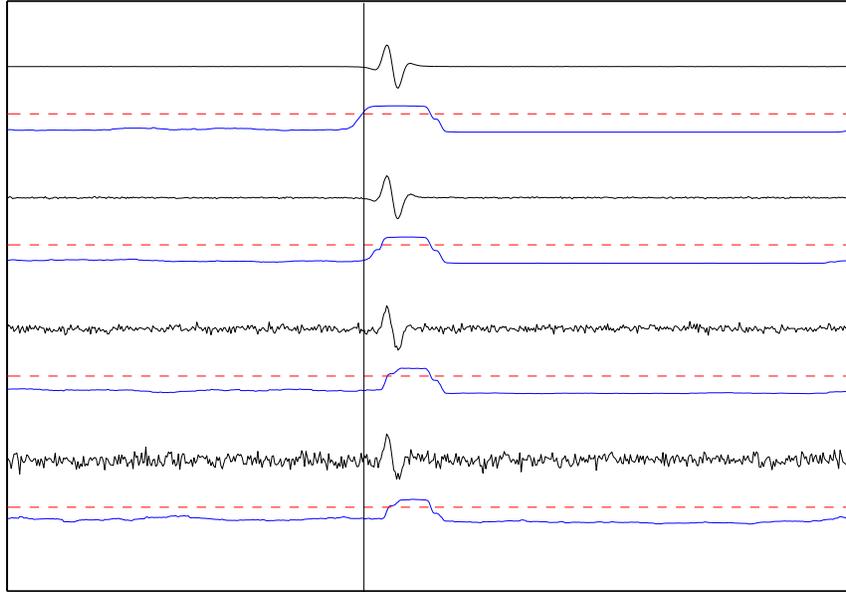


Figure 7.1: Effect of noise on the build-up of the *STA/LTA* curve. The black traces are synthetics with decreasing *SNR* from bottom to top. Below each synthetic the corresponding *STA/LTA* curve is displayed in blue with the detection threshold in red. The black vertical line denotes the location of the onset of the wave arrival in the top synthetic.

inverse regularization schemes taken from other areas of geophysics and engineering. The algorithm is based upon the recovery of the standard *STA/LTA* curve using inversion with a constraint that results in a non-linear system of equations. The selection of the constraint is directed towards the recovery of blocky versions of the *STA/LTA* curve (Sabbione and Velis, 2010). Such a constraint has also been used to invert for the earth’s relative impedance profile (Walker and Ulrych, 1983) and medical imaging applications (Charbonnier et al., 1997). Examples using synthetic and real microseismic data demonstrate the advantages of the blocky *STA/LTA* curve in the recovery of more accurate automatic time-picks in the presence of low *SNR*.

7.2 Methodology

The *STA/LTA* curve is estimated as

$$STA(i) = \frac{1}{n_s} \sum_{j=1}^i s_j^2, \quad (7.1)$$

$$LTA(i) = \frac{1}{nl} \sum_{j=1-nl}^i s_j^2, \quad (7.2)$$

$$STA/LTA(i) = \frac{STA(i)}{LTA(i)}, \quad (7.3)$$

where s_i is the i -th sample of the observed seismic trace, and ns and nl are the number of samples in the short and long averaging windows, respectively. Time picks are determined through amplitude thresholding of the STA/LTA curve. Inaccurate time-picks are commonly the result of the slow build-up of the STA/LTA curve from the time of the wave arrival until the time where the curve surpasses the user-defined threshold. An alternative method to improve the accuracy of the time-picks is through the recovery of blocky STA/LTA curves using inversion. This is obtained by minimization of the cost function,

$$J = \|\mathbf{x}_{STA} - \mathbf{X}_{LTA}\mathbf{y}_{STA/LTA}\|_2^2 + \lambda R(\mathbf{D}\mathbf{y}_{STA/LTA}), \quad (7.4)$$

where \mathbf{X}_{LTA} is a diagonal matrix with non-zero elements given by equation 7.2, and vectors \mathbf{x}_{STA} and $\mathbf{y}_{STA/LTA}$ are the time series given by equations 7.1 and 7.3, respectively. The scalar λ is a trade-off parameter, \mathbf{D} is the first derivative matrix, and $R(\mathbf{x})$ is a non-linear norm. The derivative operator by itself enhances the presence of sudden changes or boundaries in the solution and in principle it could be another alternative to onset detection. However, the derivative tends to be too sensitive to variations in the STA/LTA curve that are not of interest (e.g., see double picks at the onset in Figure 7.2). The minimization of the ℓ_2 -norm of the derivative performs a smoothing effect that seems helpful to facilitate the onset detection, although some error can be introduced due to an apparent delay in the STA/LTA build-up (see Figure 7.2). Since we are interested in blocky rather than smooth solutions, the non-linear operator R is introduced. One possibility for $R(\mathbf{x})$ is

$$R(\mathbf{x}) = \frac{\left(\frac{x_i}{\delta}\right)^2}{1 + \left(\frac{x_i}{\delta}\right)^2}, \quad (7.5)$$

where δ is a stabilization factor to avoid dividing by zero. The effect of the non-linear operator is to "square" the shape of the STA/LTA curve, while applying it to the derivative of the STA/LTA produces a cleaner non-negative curve. Nevertheless the sensitivity of the derivative to non-meaningful variations remains (Figure 7.3). In principle, the results showed in Figures 7.2 and 7.3 suggest that a good selection of thresholds and detection parameters would allow the use of either the minimization of the ℓ_2 -norm of the derivative, the non-linear norm applied over the standard STA/LTA curve or the minimization of 7.4 as improved alternatives to the standard STA/LTA filter. Nevertheless, this chapter focus in the benefits obtained through the inversion of the non-linear norm. The solution to the

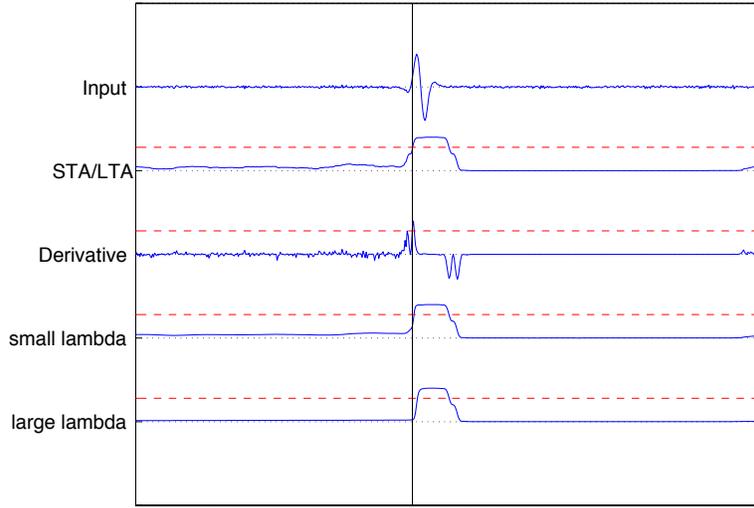


Figure 7.2: From top to bottom: raw trace with seismic arrival; standard STA/LTA curve; derivative of the STA/LTA curve; minimization of the cost function $J = \|\mathbf{x}_{STA} - \mathbf{X}_{LTA}\mathbf{y}_{STA/LTA}\|_2^2 + \lambda\|\mathbf{D}\mathbf{y}_{STA/LTA}\|_2^2$ for a small value of λ ; same minimization of the previous cost function for a large value of λ .

minimization of 7.4 is given by the iterative system

$$\tilde{\mathbf{y}}_{STA/LTA}^k = [\mathbf{X}_{LTA}^T \mathbf{X}_{LTA} + \lambda \mathbf{Q}^{k-1}]^{-1} \mathbf{X}_{LTA}^T \mathbf{x}_{STA}, \quad (7.6)$$

where \mathbf{Q} is a diagonal matrix with elements

$$Q_i = \frac{1}{[1 + (\frac{y_{STA/LTA(i)}}{\delta})^2]^2}. \quad (7.7)$$

In this way, the matrix \mathbf{Q}^{k-1} is constructed from the elements of the solution vector $\tilde{\mathbf{y}}_{STA/LTA}^{k-1}$. The iterative solver in 7.6 corresponds to the Iterative Re-weighted Least Squares (IRLS) method.

7.3 Examples

The application of the previously described method is first demonstrated on a synthetic microseismic trace computed in a horizontally layered medium. Random noise is added to the synthetic trace using the relationship $SNR = \mu_s/\sigma_n$, where μ_s is the expected signal value and σ_n the standard deviation of the noise. The SNR is set to a value of $SNR = 3$. Figure 7.4 presents the results of applying the standard and blocky STA/LTA

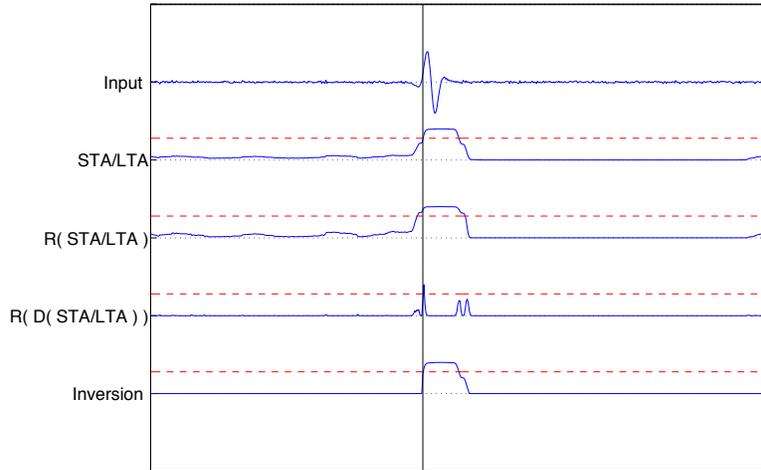


Figure 7.3: From top to bottom: raw trace with seismic arrival; standard STA/LTA curve; non-linear norm R applied to the STA/LTA ; non-linear norm R applied to the derivative of the STA/LTA ; minimization of the cost function in 7.4 with R defined as in 7.5.

filters to the previously described synthetic trace. The most straightforward method to set time picks in the STA/LTA curve is through amplitude thresholding. A visible difference in the maximum amplitude of the "boxes" that denote the arrivals of P and S waves in the STA/LTA curves is observed when the filters are applied over the noisy data. This means that if the thresholding level is set too high, the P arrivals could be missed. On the other hand, a thresholding level that is too low would produce multiple time picks that do not correspond to actual wave arrivals. Although the problem of false picks is partially alleviated using the blocky STA/LTA curve, improved results are obtained after combining the 1C denoising method from last chapter and the blocky STA/LTA curve. In the denoised version of the data, the difference in the maximum amplitude between the P and S "boxes" is reduced. Furthermore, the blocky STA/LTA curve is cleaner than the standard STA/LTA curve before the first arrival.

The second data example corresponds to an actual microseismic record (Figure 7.5). In this trace a wave arrival is visible around 0.56 s with ambient noise affecting the entire record. The trace is first denoised using the 1C sparse time-frequency transform, after which displays a visible improvement in SNR while preserving the pulse shape and time location of the arrival of interest. This suggests that the sparse transform has not substantially affected either the phase or frequency content of the arrival. The standard STA/LTA curves in both the raw and denoised traces display a slow build-up at the position of the arrival, which as stated before, introduces error in the automatic time-picking process whose amount depends on the threshold value selected by the user. The blocky behaviour on the inverted STA/LTA

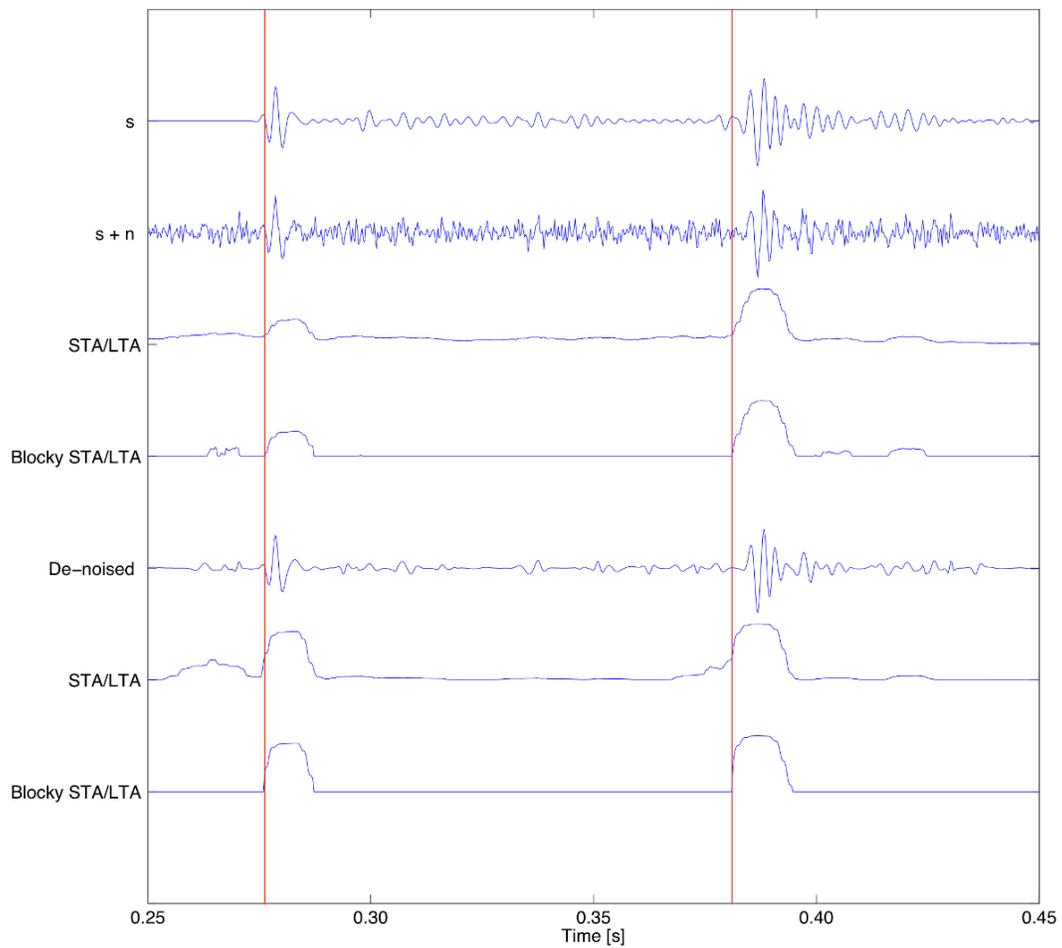


Figure 7.4: From top to bottom: synthetic microseismic trace (s), synthetic with random noise ($s + n$), STA/LTA curve from noisy synthetic, blocky version of the STA/LTA curve from noisy synthetic, denoised synthetic using the 1C sparse time-frequency transform, STA/LTA curve for the denoised trace, blocky version of the STA/LTA curve for the denoised trace. Red vertical lines indicate the approximate position of P and S arrivals in the original synthetic trace.

curves helps to reduce this error.

7.4 Summary

For location algorithms based on arrival times, time-picking is a crucial step in the determination of accurate positions. If the amount of data to be processed is large and the required response time is short, automatic time-picking algorithms are a compelling alternative for the processing of the seismic data. The most popular strategy for automatic time-picking entails amplitude thresholding of the short-term-averaging/long-term-averaging (*STA/LTA*) of the observations. However, the presence of noise produces a slow build-up of the *STA/LTA* curve that introduces errors in the determination of the onsets of wave arrivals. This chapter exemplifies several alternatives that improve the performance of the standard *STA/LTA* filter. These alternatives correspond to the minimization of the ℓ_2 -norm of the derivative of the *STA/LTA* curve, the application of a non-linear operator over the *STA/LTA* curve, and the minimization of the non-linear operator applied over the derivative of the *STA/LTA* curve. An adequate selection of thresholds and detection parameters show that any of these three alternatives can produce more accurate time-picks than the standard *STA/LTA* curve. In particular, synthetic and real microseismic data examples demonstrate the benefits brought by the last of the aforementioned strategies. A combination of this alternative with the denoising methodology from the previous chapter, presents additional advantages for the identification of accurate onsets.

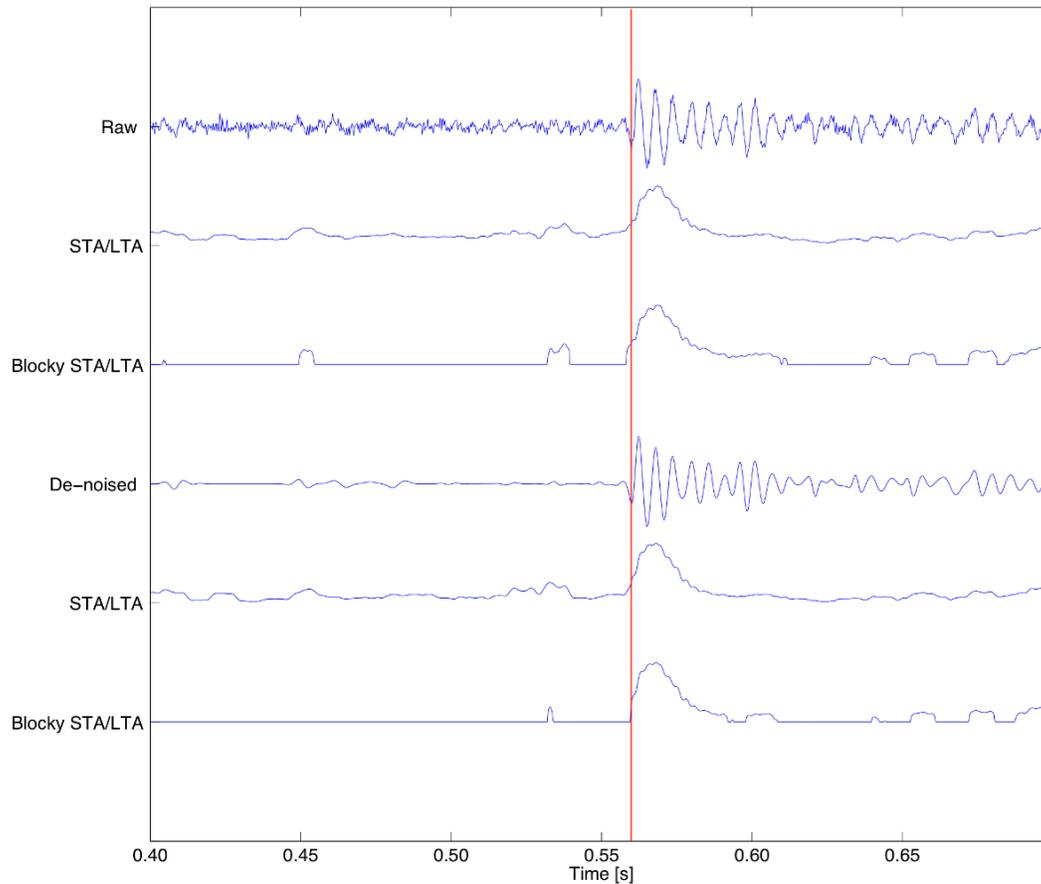


Figure 7.5: From top to bottom: real microseismic trace (raw), STA/LTA curve from raw trace, blocky version of the STA/LTA curve from raw trace, denoised trace using the 1C sparse time-frequency transform, STA/LTA curve from the denoised trace, blocky version of the STA/LTA curve from the denoised trace. The red vertical line denotes the approximate position of a wave arrival.

CHAPTER 8

Conclusions

The conclusions from this thesis can be summarized as follows:

Resolution analysis

This chapter investigates two important problems associated with microseismic monitoring. First it is reviewed the case when the travel-path trajectories are constrained to a plane, where only 5 of the 6 elements of the seismic moment tensor are recoverable. The resolution matrix provides a broader overview of the moment tensor resolvability displaying linear correlation between the elements of the moment tensor. According to the simulations, the error due to an incorrect constraint on the dipole perpendicular to the observational plane can only be propagated to two further elements of the moment tensor. On the other hand, all the eigenvalues associated with the sensitivity matrix differ from zero when one or more receivers fall outside the observational plane. This indicates that, in theory, one can resolve all 6 elements of the seismic moment tensor. However, due to the presence of noise and small eigenvalues, the resulting solutions can be unstable. In this case the analysis based on the condition number of the sensitivity matrix can be highly advantageous. The condition number analysis enables us to access levels of resolvability for a given acquisition layout. The inversion of the full moment tensor with single borehole data is feasible in situations when the array of receivers is deployed in a deviated well. For instance, the simulations in homogeneous media showed that deviated wells with curvature (dogleg severity) on the order of $6^\circ/30$ m can be used to retrieve the full moment tensor. For case studies involving abrupt changes in velocity and/or anisotropy, the condition number analysis can also be an important indicator of the resolvability of the seismic moment tensor prior to monitoring tests.

Simultaneous estimation of the source parameters

Through a convenient parametrization of the seismic source forward problem, I showed that the source mechanism is a sparse representation of the source displacement field under a dictionary of Green's functions. Hence, sparse representation theory provides alternative strategies for simultaneously inverting for the origin time, location and seismic moment tensor of seismic events. The modified BOMP is suitable to enforce block sparsity in the solutions and to identify optimum source parameters in the presence of highly coherent dictionaries. The BOMP method used in this thesis follows essentially the same workflow of the regular BOMP algorithm. The main difference comes from the physical meaning of the superdictionary used in the source parameter inversion, where solutions are found after matching the input observations both in 3D space and time. Although, without giving consideration to its physical meaning, the superdictionary \mathbb{G} could be treated as a regular dictionary, the high coherence between its columns makes difficult the identification of optimal solutions using the regular BOMP method. Hence, a modification based on a search over the largest values of the ℓ_2 -norm of the correlation between the blocks of the dictionary and the current residuals makes the algorithm more suitable to work in highly coherent dictionaries such as those that arise in seismic monitoring applications. Although the BOMP method has not been tested as extensively as other methodologies currently in use in different areas of geophysics, it presents capabilities that are of great interest for the development of improved source parameter inversion systems. One characteristic that differentiates BOMP from most other source inversion approaches is that it is specifically designed to be implemented as a continuous monitoring system. Also, it can potentially solve for multiple sources in the same set of observations that enter the inversion.

Compressed domain inversion

Real-time simultaneous estimation of the source parameters has applications in different areas of geophysics. This thesis introduces a new method for the automatic recovery of source parameters based on ideas of compressive sensing. The main advantage of this method is that a large number of recordings can be inverted without compromising real-time response. Furthermore, if the compression rate is correctly selected, the results obtained by the compressed domain methodology would be identical to those without using compression, with the added benefit of a shorter processing time. The application of compressive sensing does require a dictionary of Green's functions that embodies the properties of a low-coherency frame or an orthonormal basis. Moreover, the lack of practical dictionaries that meet this assumption necessitates the use of numerical modelling as a vital tool to determine the feasibility of applying the compressive sensing approach in a given scenario. The overall effect of deviating from the theoretical assumptions outlined in compressive sensing theory

(e.g., the use of an inaccurate dictionary that is not a frame or orthonormal basis) is that an increased number of non-adaptive measurements is required to recover optimal solutions. A complete implementation of compressive sensing requires the setting of new acquisition protocols that allow the recording of the compressed measurements. Even in the absence of such protocols, a proper use of compressive sensing can be an effective tool for real-time monitoring of seismic events.

Microseismic data denoising

Processing of microseismicity records is a challenge due to, among other problems, the low signal-to-noise ratio displayed by the signals of interest. Application of techniques and processing tools developed in other knowledge areas can be of great advantage to deal with microseismicity records. The method presented in this chapter is a time-frequency transform constrained to impose sparsity into the time-frequency representation of the signal. When compared against the equivalent ℓ_2 -norm transform, the sparsity constrained transform using the ℓ_1 -norm displays improved results for denoising purposes. This transform can be implemented component by component (1C algorithm) or by simultaneously utilizing the three components of a receiver (3C algorithm). In particular, the 3C version of the sparse transform is advantageous in cases where the wave arrival exists in the three components of a receiver but the amplitude of the arrival is below the noise level in one component. In such cases, the 3C algorithm uses information from the components where the arrival amplitudes are above the noise level to recover the amplitude information in the other component. The results observed in synthetic and real data show the efficacy of the method to remove ambient noise without significantly affecting the frequency content and/or amplitude of the signal of interest. Hodogram analyses of the denoised results further verifies the preservation of amplitude and phase information after the denoising process. Additionally, denoised results displayed cleaner onsets useful for manual or automatic selection of the hodogram windows and higher frequency content than low-pass filtered versions of the original data. This extra information allowed the hodogram curves of the denoised data to display a smooth behaviour keeping the main trends of particle movement. Through the use of the denoising method in a pre-processing stage, improved results can be expected in the estimation of polarization angles, automatic time-picking and amplitude stacking of microseismic traces.

Improvements in automatic time-picking

For location algorithms based on arrival times, time-picking is a crucial step in the determination of accurate positions. If the amount of data to be processed is large and the required response time is short, automatic time-picking algorithms are a compelling alternative for

the processing of the seismic data. The most popular strategy for automatic time-picking entails amplitude thresholding of the short-term-average/long-term-average (*STA/LTA*) of the observations. However, the presence of noise produces a slow build-up of the *STA/LTA* curve that introduces errors in the determination of the onsets of wave arrivals. This chapter exemplifies several alternatives that improve the performance of the standard *STA/LTA* filter. These alternatives correspond to the minimization of the ℓ_2 -norm of the derivative of the *STA/LTA* curve, the application of a non-linear operator over the *STA/LTA* curve, and the minimization of the non-linear operator applied over the derivative of the *STA/LTA* curve. An adequate selection of thresholds and detection parameters show that any of these three alternatives can produce more accurate time-picks than the standard *STA/LTA* curve. In particular, synthetic and real microseismic data examples demonstrate the benefits brought by the last of the aforementioned strategies. A combination of this alternative with the denoising methodology presented in chapter 6, presents additional advantages for the identification of accurate onsets.

Bibliography

- Aki, K. and P. Richards. *Quantitative seismology*. Sausalito, CA: University Science Books, 2009.
- Albright, J. and C. Pearson. “Acoustic Emissions as a Tool for Hydraulic Fracture Location: Experience at the Fenton Hill Hot Dry Rock Site.” *SPE Journal* 22 (1982): 523–530.
- Allen, R. “Automatic earthquake recognition and timing from single traces.” *Bulletin of the Seismological Society of America* 68 (1978): 1521–1532.
- Anton, H. and R. Busby. *Contemporary Linear Algebra*. John Wiley & Sons, 2003.
- Aster, R., B. Borchers, and C. Thurber. *Parameter estimation and inverse problems*. Elsevier Academic Press, 2005.
- Baig, A., M. Campillo, and F. Brenguier. “Denoising seismic noise cross correlations.” *Journal of Geophysical Research* 114 (2009): B08310.
- Baraniuk, R. “Compressive Sensing.” *IEEE Signal Processing Magazine* 24 (2007): 118–121.
- Baraniuk, R., M. Davenport, R. DeVore, and M. Wakin. “A Simple Proof of the Restricted Isometry Property for Random Matrices.” *Constructive Approximation* 28 (2008): 253–263.
- Bayuk, I., M. Ammerman, and E. Chesnokov. “Estimation of shale’s permeability from microseismicity.” *79th Annual International Meeting, SEG, Expanded Abstracts*. 2009, 1581–1585.
- Bear, G., J. Rupp, and A. Rudman. “Seismic interpretation of the deep structure of the Wabash Valley Fault System.” *Seismological Research Letters* 68 (1997): 624–640.
- Beck, Amir and Marc Teboulle. “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems.” *SIAM Journal on Imaging Sciences* 2 (2009): 183–202.

BIBLIOGRAPHY

- Bennett, L., J. Le Calvez, D. Sarver, K. Tanner, W. Birk, G. Waters, J. Drew, G. Michaud, P. Primiero, L. Eisner, R. Jones, D. Leslie, M. Williams, J. Govenlock, R. Klem, and K. Tezuka. “The Source for Hydraulic Fracture Characterization.” *Oilfield Review* (2005): 42–57.
- Berg, E. and M. Friedlander. “Probing the Pareto frontier for basis pursuit solutions.” *SIAM Journal on Scientific Computing* 31 (2008): 890–912.
- Bernardi, Fabrizio, Jochen Braunmiller, Urs Kradolfer, and Domenico Giardini. “Automatic regional moment tensor inversion in the European-Mediterranean region.” *Geophysical Journal International* 157 (2004): 703–716.
- Bleakly, D., J. Wolfe, I. Leslie, M. Prince, V. Shumila, and T. Urbancic. “Using multi-well microseismicity to identify fracture types associated with hydraulic fracture stimulations.” *CSEG Recorder* 32 (2007): 49–51.
- Bobin, J., J. Starck, and R. Ottensamer. “Compressed Sensing in Astronomy.” *IEEE Journal of Selected Topics in Signal Processing* 2 (2008): 718–726.
- Bonar, D. and M. Sacchi. “Complex spectral decomposition via inversion strategies.” *SEG expanded abstracts*. 2010, 1408–1412.
- Bryan, S., J. Cox, and D. Blackwell. “High dogleg rotary steerable system: a step change in drilling process.” *SPE Annual Technical Conference and Exhibition*. 2009.
- Candes, E., J. Romberg, and T. Tao. “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information.” *IEEE Transactions on Information Theory* 52 (2006): 489–509.
- Candes, E. and M. Wakin. “An introduction to compressive sampling.” *IEEE Signal Processing Magazine* 25 (2008): 21–30.
- Carter, B., J. Desroches, A. Ingraffea, and P. Wawrzynek. *Simulating fully 3D hydraulic fracturing*. Modeling in Geomechanics. Wiley. 2000.
- Chambers, K., J. Kendall, S. Brandsberg-Dahl, and J. Rueda. “Testing the ability of surface arrays to monitor microseismic activity.” *Geophysical Prospecting* 58 (2010): 821–830.
- Charbonnier, P., L. Blanc-Feraud, Aubert, G., and M. Barlaud. “Deterministic edge-preserving regularization in computed imaging.” *IEEE Transactions on Image Processing* 6 (1997): 298–311.
- Chen, S., D. Donoho, and M. Saunders. “Atomic decomposition by basis pursuit.” *SIAM Journal on Scientific Computing* 20 (1998): 33–61.

BIBLIOGRAPHY

- Cipolla, C., N. Warpinski, M. Mayerhofer, E. Lolon, and M. Vincent. “The Relationship Between Fracture Complexity, Reservoir Properties, and Fracture Treatment Design.” *SPE Annual Technical Conference and Exhibition*. Denver, CO: Society of Petroleum Engineers, 2008, SPE115769.
- Cipolla, C. and C. Wright. “Diagnostic techniques to understand hydraulic fracturing: what? why? and how?.” *SPE Production & Facilities* 17 (2002): 23–35.
- Clinton, J., E. Hauksson, and K. Solanki. “An evaluation of the SCSN moment tensor solutions: robustness of the M_W magnitude scale, style of faulting and automation of the method.” *Bulletin of the Seismological Society of America* 96 (2006): 1689–1705.
- Cotter, S., B. Rao, K. Engan, and K. Kreutz-Delgado. “Sparse solutions to linear inverse problems with multiple measurement vectors.” *IEEE Transactions on Signal Processing* 53 (2005): 2477–2488.
- Dahlen, F. and J. Tromp. *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press, 1998.
- Daubechies, I., M. Defrise, and C. De Mol. “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint.” *Communications on Pure and Applied Mathematics* 57 (2004): 1413–1457.
- Detournay, E. and A. Cheng *Fundamentals of Poroelasticity*. Volume Volume 2, Analysis and Design Method of Comprehensive Rock Engineering: Principles, Practice and Projects. Oxford, UK: Pergamon Press. 1993. 113–171.
- Donoho, D. “Compressed Sensing.” *IEEE Transactions on Information Theory* 52 (2006): 1289–1306.
- Donoho, D., M. Elad, and V. Temlyakov. “Stable recovery of sparse overcomplete representations in the presence of noise.” *IEEE Transactions on Information Theory* 52 (2006): 6–17.
- Dreger, D. and D. Helmberger. “Determination of source parameters at regional distances with single station or sparse network data.” *Journal of Geophysical Research* 98 (1993): 8107–8125.
- Dreger, Douglas S. “85.11 TDMT_INV: Time domain seismic moment tensor inversion.” *International Handbook of Earthquake and Engineering Seismology*. Ed. Paul C. Jennings William H.K. Lee, Hiroo Kanamori and Carl Kisslinger. Volume 81, Part B of International Geophysics. Academic Press, 2003. 1627 – 1627.

BIBLIOGRAPHY

- Dufumier, H. and L. Rivera. “On the resolution of the isotropic component in moment tensor inversion.” *Geophysical Journal International* 131 (1997): 595–606.
- Duncan, P. and L. Eisner. “Reservoir characterization using surface microseismic monitoring.” *Geophysics* 75 (2010): 75A139–75A146.
- Dziewonski, A., T. Chou, and J. Woodhouse. “Determination of earthquake source parameters from waveform data for studies of global and regional seismicity.” *Journal of Geophysical Research* 86 (1981): 2825–2852.
- Eisner, L., D. Abbott, W. Barker, J. Lakings, and M. Thornton. “Noise suppression for detection and location of microseismic events using a matched filter.” *SEG expanded abstracts*. 2008, 1431–1435.
- Eisner, L., P. Duncan, H. Werner, and W. Keller. “Uncertainties in passive seismic monitoring.” *The Leading Edge* 28 (2009): 648–655.
- Eisner, L., B. Hulse, P. Duncan, D. Jurick, H. Werner, and W. Keller. “Comparison of surface and borehole locations of induced seismicity.” *Geophysical Prospecting* 58 (2010): 809–820.
- Eldar, Yonina C. and Helmut Bolcskei. “Block-sparsity: Coherence and efficient recovery.” *Acoustics, Speech, and Signal Processing, IEEE International Conference on* 0 (2009): 2885–2888.
- Eldar, Yonina C., Patrick Kuppinger, and Helmut Bolcskei. “Block-Sparse Signals: Uncertainty Relations and Efficient Recovery.” *IEEE Transactions on Signal Processing* 58 (June 2010): 3042–3054.
- Fischer, T., S. Hainzl, L. Eisner, S. Shapiro, and J. LeCalvez. “Microseismic signatures of hydraulic fracture growth in sediment formations: Observations and modeling.” *Journal of Geophysical Research* 113 (2008): B02307.
- Fjaer, E., R. Holt, P. Horsrud, A. Raaen, and R. Risnes. “Chapter 11 Mechanics of hydraulic fracturing.” *Petroleum Related Rock Mechanics 2nd Edition*. Ed. E. Fjaer, R. Holt, P. Horsrud, A. Raaen, and R. Risnes. Volume 53 of Developments in Petroleum Science. Elsevier, 2008. 369 – 390.
- Fornasier, Massimo and Holger Rauhut. “Recovery Algorithms for Vector-Valued Data with Joint Sparsity Constraints.” *SIAM J. Numerical Analysis* 46 (2008): 577–613.
- Foulger, G., B. Julian, D. Hill, A. Pitt, P. Malin, and E. Shalev. “Non-double-couple microearthquakes at Long Valley caldera, California, provide evidence for hydraulic fracturing.” *Journal of Volcanology and Geothermal Research* 132 (2004): 45–71.

BIBLIOGRAPHY

- Gibowicz, Slawomir J. “Seismicity Induced by Mining: Recent Research.” *Advances in Geophysics*. Ed. Renata Dmowska. Volume 51 . Elsevier, 2009. 1 – 53.
- Gilbert, F. “Excitation of the Normal Modes of the Earth by Earthquake Sources.” *Geophysical Journal of the Royal Astronomical Society* 22 (1970): 223–226.
- Gilbert, F. “Derivation of Source Parameters from Low-Frequency Spectra.” *Philosophical Transactions of the Royal Society A* 274 (1973): 369–371.
- Godano, M., E. Gaucher, T. Bardainne, M. Regnier, A. Deschamps, and M. Valette. “Assessment of focal mechanisms of microseismic events computed from two three-component receivers: application to the Arkema-Vauvert field (France).” *Geophysical Prospecting* 58 (2010): 775–790.
- Guilhem, A. and D. Dreger. Towards a real-time earthquake source determination and tsunami early warning in Northern California. Technical report, Berkeley Seismological Laboratory, 2010.
- Hardy, H. *Applications of Acoustic Emission Techniques to Rock and Rock Structures: A State-of-the-Art Review*. Acoustic Emissions in Geotechnical Engineering Practice. Baltimore, MD: American Society for Testing and Materials. 1981. 4–92.
- Heisig, G., G. Cavallaro, P. Jogi, J. Hood, and I. Forstner. “Continuous borehole curvature estimates while drilling based on downhole bending moment measurements.” *SPE Annual Technical Conference and Exhibition*. 2004.
- Herman, M. and T. Strohmer. “General deviants: an analysis of perturbations in compressed sensing.” *IEEE Journal of Selected Topics in Signal Processing* 4 (2010): 342–349.
- Herrmann, F. “Randomized sampling and sparsity: Getting more information from fewer samples.” *Geophysics* 75 (2010): WB173–WB187.
- Hestenes, Magnus R. and Eduard Stiefel. “Methods of Conjugate Gradients for Solving Linear Systems.” *Journal of Research of National Bureau Standards* 49 (1952): 409–436.
- Horn, R. and Ch. Johnson. *Matrix analysis*. Cambridge University Press, 1985.
- House, L., R. Flores, and R. Whitters. “Microearthquakes induced by a hydraulic injection in sedimentary rock, East Texas.” *66th Annual International Meeting, SEG, Expanded Abstracts*. 1996, 110–113.
- Hsu, Y., X. Dang, W. Chilton, P. Chan, I. Stelin, D. Gusain, N. Northington, and H. Pater. “New Physics-Based 3D Hydraulic Fracture Model.” *Hydraulic Fracturing Technology Conference*. The Woodlands, TX: Society of Petroleum Engineers, 2012, SPE152525.

BIBLIOGRAPHY

- Hyder, M. and K. Mahata. “Direction-of-Arrival estimation using a mixed $\ell_{2,0}$ norm approximation.” *IEEE Transactions on Signal Processing* 58 (2010): 4646–4655.
- Jechumtalova, Z. and L. Eisner. “Seismic source mechanism inversion from a linear array of receivers reveals non-double-couple seismic events induced by hydraulic fracturing in sedimentary formation.” *Tectonophysics* 460 (2008): 124–133.
- Ji, L., A. Settari, and R. Sullivan. “A Novel Hydraulic Fracturing Model Fully Coupled with Geomechanics and Reservoir Simulator.” *SPE Journal* 14 (2009): 423–430.
- Jing, Ch. and T. Rape. “Resolvability analysis for rock property inversions of multicomponent seismic data.” *74th International Exposition and Annual Meeting, SEG, Expanded Abstracts*. 2004.
- Jost, M. and R. Herrmann. “A Student’s Guide to and Review of Moment Tensors.” *Seismological Research Letters* 60 (1989): 37–57.
- Julian, B., G. Foulger, F. Monastero, and S. Bjornstad. “Imaging hydraulic fractures in a geothermal reservoir.” *Geophysical Research Letters* 37 (2010): L07305, doi:10.1029/2009GL040933.
- Julian, B., A. Miller, and G. Foulger. “Non-double-couple earthquakes.” *Reviews of Geophysics* 36 (1998): 525–549.
- Kawakatsu, H. “Automated near-realtime CMT inversions.” *Geophysical Research Letters* 94 (1995): 2569–2572.
- Kawakatsu, H. “On the realtime monitoring of the long-period seismic wavefield.” *Bulletin of the Earthquake Research Institute* 73 (1998): 267–274.
- Kidney, R., U. Zimmer, and N. Boroumand. “Impact of distance-dependent location dispersion error on interpretation of microseismic event distributions.” *The Leading Edge* 29 (2010): 284–289.
- Kim, W. “The 18 June 2002 Caborn, Indiana, earthquake: reactivation of ancient rift in the Wabash Valley seismic zone?.” *Bulletin of the Seismological Society of America* 93 (2003): 2201–2211.
- Knopoff, L. and M. Randall. “The compensated linear-vector dipole: a possible mechanism for deep earthquakes.” *Journal of Geophysical Research* 75 (1970): 4,957–4,963.
- Kolinsky, P., L. Eisner, V. Grechka, D. Jurick, and P. Duncan. “Observation of shear-wave splitting from microseismicity induced by hydraulic fracturing - a non-VTI story.” *71st EAGE Conference & Exhibition*. 2009.

BIBLIOGRAPHY

- Kontar, Y., F. Boadu, P. Carpenter, R. Harrison, M. Long, R. Nelson, S. Olson, A. Raef, R. Singh, and M. Zhdanov. “Is the Wabash Valley Seismic Zone related with the ancient Reelfoot Rift?” *GeoPRISMS Implementation Workshop: Rift Initiation and Evolution*. 2010.
- Kowalski, Matthieu and Bruno Torr sani. “Sparsity and persistence: mixed norms provide simple signal models with dependent coefficients.” *Signal, Image and Video Processing* 3 (September 2009): 251–264.
- Langer, C. and G. Bollinger. “The southeastern Illinois earthquake of 10 June 1987: the later aftershocks.” *Bulletin of the Seismological Society of America* 81 (1991): 423–445.
- Lay, T. and T. Wallace. *Modern Global Seismology*. San Diego, CA: Academic Press, 1995.
- Le Calvez, J., K. Tanner, S. Glenn, P. Kaufman, D. Sarver, L. Bennett, R. Panse, and J. Palacio. “Using induced microseismicity to monitor hydraulic fracture treatment: a tool to improve completion techniques and reservoir management.” *SPE Eastern Regional Meeting*. 2006, SPE104570.
- LeCampion, B. and R. Jeffrey. “Real-time estimation of fracture volume and hydraulic fracture treatment efficiency.” *North America Rock Mechanics Symposium: Rock mechanics across borders and disciplines*. 2004, ARMA/NARMS04–519.
- Lee, S., B. Huang, W. Liang, and K. Chen. “Grid-based moment tensor inversion technique by using 3-D Green’s functions database: a demonstration of the 23 October Taipei earthquake.” *Terrestrial Atmospheric and Oceanic Sciences* 21 (2010): 503–514.
- Lustig, M., D. Donoho, and J. Pauly. “Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging.” *Magnetic Resonance in Medicine* 58 (2007): 1182–1195.
- Madariaga, R. “Seismic Source Theory.” *Treatise on Geophysics*. Ed. Gerald Schubert. Amsterdam: Elsevier, 2007. 59 – 82.
- Mallat, S. and Z. Zhang. “Matching pursuits with time-frequency dictionaries.” *IEEE Transactions on Signal Processing* 41 (1993): 3397–3415.
- Mallat, Stphane. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3rd edition. Academic Press, 2008.
- Martinez, S., R. Steanson, and A. Coulter. *Formation Fracturing*. Petroleum Engineering Handbook. Richardson, TX: Society of Petroleum Engineers. 1987. 1–12.
- Maxwell, S. “A brief guide to passive seismic monitoring.” *CSEG National Convention*. 2005, 177–178.

BIBLIOGRAPHY

- Maxwell, S. “Confidence and accuracy of microseismic images.” *CSPG CSEG CWLS Convention*. 2009, 480–483.
- Maxwell, S., J. Rutledge, R. Jones, and M. Fehler. “Petroleum reservoir characterization using downhole microseismic monitoring.” *Geophysics* 75 (2010): 75A129–75A137.
- Maxwell, S., J. Shemeta, E. Campbell, and D. Quirk. “Microseismic Deformation Rate Monitoring.” *SPE Annual Technical Conference and Exhibition*. Denver, CO: Society of Petroleum Engineers, 2008, SPE116596.
- Maxwell, S., T. Urbancic, N. Steinsberger, and R. Zinno. “Microseismic imaging of hydraulic fracture complexity in the Barnett Shale.” *SPE Annual Technical Conference and Exhibition*. 2002, SPE77440.
- McNutt, S. “Seismic Monitoring and Eruption Forecasting of Volcanoes: A Review of the State-of-the-Art and Case Histories.” *Monitoring and Mitigation of Volcano Hazards*. Ed. Scarpa and Tiling. Berlin: Springer-Verlag, 1996. 99 – 146.
- Menke, W. *Geophysical data analysis: discrete inverse theory*. Academic Press, 1989.
- Neale, C., M. Macaulay, M. Manning, and D. MacNeil. “The Value of Microseismic Monitoring.” *The Source* 6 (2009): 20–24.
- Nelson, John. “Structural Styles of the Illinois Basin.” *Interior Cratonic Basins*. Ed. D. F. Oltz M. W. Leighton, D. R. Kolata and J. J. Eidel. American Association of Petroleum Geologists, 1991. 209 – 243.
- Nolen-Hoeksema, R. and L. Ruff. “Moment tensor inversion of microseisms from the B-sand propped hydrofracture, M-site, Colorado.” *Tectonophysics* 336 (2001): 163–181.
- Pasyanos, M., D. Dreger, and B. Romanowicz. “Toward real-time estimation of regional moment tensors.” *Bulletin of the Seismological Society of America* 86 (1996): 1255–1269.
- Pati, Y., R. Rezaifar, and P. Krishnaprasad. “Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition.” *Proceedings of the 27th Annual Asilomar Conference in Signals, Systems and Computers*. 1993.
- Phillips, W., T. Fairbanks, J. Rutledge, and D. Anderson. “Induced microearthquake patterns and oil-producing fracture systems in the Austin chalk.” *Tectonophysics* 289 (1998): 153–169.
- Randall, G. “Efficient calculation of complete differential seismograms for laterally homogeneous earth models.” *Geophysical Journal International* 118 (1994): 245–254.

BIBLIOGRAPHY

- Rao, Y., Y. Wang, and J. Morgan. "Crosshole seismic waveform tomography - II. Resolution analysis." *Geophysical Journal International* 166 (2006): 1237–1248.
- Romanowicz, B. "Moment tensor inversion of long period Rayleigh waves: a new approach." *Journal of Geophysical Research* 87 (1982): 5395–5407.
- Romanowicz, B., D. Dreger, M. Pasyanos, and R. Uhrhammer. "Monitoring of strain release in central and northern California using broadband data." *Geophysical Research Letters* 20 (1993): 1643–1646.
- Rothert, E. and A. Shapiro. "Statistics of fracture strength and fluid-induced microseismicity." *Journal of Geophysical Research* 112 (2007): B04309.
- Rutledge, J. and W. Phillips. "Hydraulic simulation of natural fractures as revealed by induced microearthquakes, Carthage Cotton Valley gas field, east Texas." *Geophysics* 68 (2003): 441–452.
- Rutledge, J., W. Phillips, and M. Mayerhofer. "Faulting Induced by Forced Fluid Injection and Fluid Flow Forced by Faulting: An Interpretation of Hydraulic-Fracture Microseismicity, Carthage Cotton Valley Gas Field, Texas." *Bulletin of the Seismological Society of America* 94 (2004): 1817–1830.
- Sabbione, J. and D. Velis. "Automatic first-breaks picking: New strategies and algorithms." *Geophysics* 75 (2010): V67–V76.
- Scognamiglio, L., E. Tinti, and A. Michelini. "Real-time determination of seismic moment tensor for the Italian region." *Bulletin of the Seismological Society of America* 99 (2009): 2223–2242.
- Shapiro, S., P. Audigane, and J. Royer. "Large-scale in situ permeability tensor of rocks from induced microseismicity." *Geophysical Journal International* 137 (1999): 207–213.
- Shearer, P. *Introduction to Seismology*. Cambridge University Press, 1999.
- Shemeta, J. and P. Anderson. "It's a matter of size: Magnitude and moment estimates for microseismic data." *The Leading Edge* 29 (2010): 296–302.
- Sileny, J., D. Hill, L. Eisner, and F. Cornet. "Non-double-couple mechanisms of microearthquakes induced by hydraulic fracturing." *Journal of Geophysical Research* 114 (2009): B08307.
- Sipkin, Stuart A. "Estimation of earthquake source parameters by the inversion of waveform data: synthetic waveforms." *Physics of The Earth and Planetary Interiors* 30 (1982): 242 – 259.

BIBLIOGRAPHY

- Sobolev, G. and A. Lyubushin. “Microseismic impulses as earthquake precursors.” *Izvestiya, Physics of the Solid Earth* 42 (2006): 721–733.
- Strang, G. *Linear algebra and its applications*. Thomson Brooks/Cole, 2006.
- Tajima, F., Ch. Megnin, D. Dreger, and B. Romanowicz. “Feasibility of real-time broadband waveform inversion for simultaneous moment tensor and centroid location determination.” *Bulletin of the Seismological Society of America* 92 (2002): 739–750.
- Takhar, D., J. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. Kelly, and R. Baraniuk. “A New Compressive Imaging Camera Architecture using Optical-Domain Compression.” *Proceedings Computational Imaging IV*. San Jose, CA: SPIE, 2006, 43–52.
- Trampert, J. “Global seismic tomography: the inverse problem and beyond.” *Inverse Problems* 14 (1998): 371–385.
- Trifu, C. and V. Shumila. “Reliability of seismic moment tensor inversions for induced microseismicity at Kidd Mine, Ontario.” *Pure and Applied Geophysics* 159 (2002): 145–164.
- Tropp, J. “Greed is good: algorithmic results for sparse approximation.” *IEEE Transactions Information Theory* 50 (2004): 2231–2242.
- Tsuruoka, Hiroshi, Hitoshi Kawakatsu, and Taku Urabe. “GRiD MT (grid-based real-time determination of moment tensors) monitoring the long-period seismic wavefield.” *Physics of the Earth & Planetary Interiors* 175 (2009): 8 – 16.
- Van Rijssen, E. and G. Herman. “Resolution analysis of band-limited and offset-limited seismic data for plane-layered subsurface models.” *Geophysical Prospecting* 39 (1991): 61–76.
- Vavrycuk, V. “Inversion for parameters of tensile earthquakes.” *Journal of Geophysical Research* 106 (2001): 16,339–16,355.
- Vavrycuk, V. “On the retrieval of moment tensors from borehole data.” *Geophysical Prospecting* 55 (2007): 381–391.
- Vera Rodriguez, I., D. Bonar, and M. Sacchi. “Microseismic record de-noising using a sparse time-frequency transform.” *81st Annual International Meeting, SEG, Expanded Abstracts*. 2011, 1693–1698.
- Vera Rodriguez, I., Y. Gu, and M. Sacchi. “Resolution of Seismic-Moment Tensor Inversions from a Single Array of Receivers.” *Bulletin of the Seismological Society of America* 101 (2011): 2634–2642.

BIBLIOGRAPHY

- Vera Rodriguez, I. and M. Sacchi. “Resolvability analysis of single azimuth seismic moment tensor inversion.” *CSPG CSEG CWLS Convention*. 2009, 476–479.
- Vera Rodriguez, I., M. Sacchi, and Y. Gu. “Continuous hypocenter and source mechanism inversion via a Green’s function-based matching pursuit algorithm.” *The Leading Edge* 29 (2010): 334–337.
- Vera Rodriguez, I., M. Sacchi, and Y. Gu. “Towards a near real-time system for event hypocenter and source mechanism recovery via compressive sensing.” *80th Annual International Meeting, SEG, Expanded Abstracts*. 2010, 2140–2145.
- Vera Rodriguez, I., M. Sacchi, and Y. Gu. “Simultaneous recovery of origin time, hypocentre location and seismic moment tensor using sparse representation theory.” *Geophysical Journal International* 188 (2012): 1188–1202.
- Verdon, J. and J. Kendall. “Constraining Fracture Geometry from Shear-wave Splitting Measurements of Passive Seismic Data.” *71st EAGE Conference & Exhibition*. Amsterdam: European Association of Geoscientists & Engineers, 2009.
- Walker, C. and T. Ulrych. “Autoregressive recovery of the acoustic impedance.” *Geophysics* 45 (1983): 1338–1350.
- Warpinski, N. “Interpretation of hydraulic fracture mapping experiments.” *SPE Centennial Petroleum Engineering Symposium*. 1994, SPE27985.
- Warpinski, N.. “Integrating microseismic monitoring with well completions, reservoir behavior, and rock mechanics.” *SPE Tight Gas Completions Conference*. 2009, SPE125239.
- Warpinski, N., S. Wolhart, and C. Wright. “Analysis and prediction of microseismicity induced by hydraulic fracturing.” *SPE Annual Technical Conference and Exhibition*. 2001, SPE71649.
- Wiggins, R. “The general linear inverse problem: implications of surface waves and free oscillations on earth structure.” *Reviews of Geophysics and Space Physics* 10 (1972): 251–285.
- Xia, J., R. Miller, and Y. Xu. “Data-resolution matrix and model-resolution matrix for Raleigh-wave inversion using a damped least-squares method.” *Pure and Applied Geophysics* 165 (2008): 1227–1248.
- Yamamoto, K., M. Gutierrez, T. Shimamoto, and S. Maezumi. “Verification of a 3D Hydraulic Fracturing Model Against a Field Case.” *Improved Oil Recovery Symposium*. Tulsa, OK: Society of Petroleum Engineers, 2000, SPE59373.

BIBLIOGRAPHY

- Yao, H., P. Gerstoft, P. Shearer, and C. Mecklenbrauker. “Compressive sensing of the Tohoku-Oki Mw 9.0 earthquake: Frequency-dependent rupture modes.” *Geophysical Research Letters* 38 (2011): L20310.
- Yuan, Ming and Yi Lin. “Model selection and estimation in regression with grouped variables.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (2006): 49–67.
- Zhao, L. and D. Helmberger. “Source estimation from broadband regional seismograms.” *Bulletin of the Seismological Society of America* 84 (1994): 91–104.
- Zhu, Lupei and Donald V. Helmberger. “Advancement in source estimation techniques using broadband regional seismograms.” *Bulletin of the Seismological Society of America* 86 (1996): 1634–1641.

APPENDIX A

FISTA for group sparse solutions ¹

To incorporate the mixed $\ell_{2,1}$ -norm within the FISTA algorithm for the purposes of promoting group sparsity, a brief introduction to the Iterative Soft Thresholding Function (ISTA) is presented. The mathematical framework for this algorithm was first presented in Daubechies et al. (2004). Minimizing the cost function defined in Equation 6.5 through a gradient based approach, the ISTA algorithm iterates towards the solution, from an initial estimate \mathbf{a}_0 , by,

$$\mathbf{a}_{k+1} = \mathcal{T}_{\lambda_{\ell_1}/2\alpha} \left(\mathbf{a}_k - \frac{1}{\alpha} \mathbf{R}^T (\mathbf{s} - \mathbf{R}\mathbf{a}_k) \right). \quad (\text{A.1})$$

The constant α is chosen to be greater than or equal to the maximum eigenvalue of $\mathbf{R}^T \mathbf{R}$ to prevent the argument, i.e. gradient step, within the operator \mathcal{T}_τ from becoming a negative function. The operator \mathcal{T}_τ is known as the soft thresholding function and is defined by

$$\mathcal{T}_\tau(g) = (|g| - \tau)_+ \text{sign}(g), \quad (\text{A.2})$$

where g is a scalar value, τ is a value greater than zero, and $(|g| - \tau)_+$ is simply the maximum between $(|g| - \tau)$ and 0. To accelerate the convergence rate of ISTA, the FISTA algorithm, developed in Beck and Teboulle (2009), uses a very specific linear combination of the previous two points or iterations. Specifically, this combination at the k^{th} iteration is

$$\mathbf{y}_k = \mathbf{a}_{k-1} + \left(\frac{t_{k-2}}{t_k} \right) (\mathbf{a}_{k-1} - \mathbf{a}_{k-2}), \quad (\text{A.3})$$

¹A version of this appendix has been published. Vera Rodriguez, I., Bonar, D. and Sacchi, M., 2012. Geophysics 77: V21-V29.

where the variable t_k is defined by

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

For the first iteration of the FISTA algorithm \mathbf{y}_1 is chosen to be \mathbf{a}_0 and t_1 is chosen to be 1. The next iterate of the solution model, \mathbf{a}_{k+1} , is then calculated using \mathbf{y}_k ,

$$\mathbf{a}_{k+1} = \mathcal{T}_{\lambda_{\ell_1}/2\alpha} \left(\mathbf{y}_k - \frac{1}{\alpha} \mathbf{R}^T (\mathbf{s} - \mathbf{R}\mathbf{y}_k) \right). \quad (\text{A.5})$$

By changing the regularization term to the mixed $\ell_{2,1}$ -norm as in Equation 6.8, the derivation of the ISTA algorithm, and by extension the FISTA algorithm, is relatively unchanged (Fornasier and Rauhut, 2008). The only difference that arises within the ISTA or FISTA algorithm is the definition of the soft thresholding function. Rather than the soft thresholding operator \mathcal{T}_τ , a new multidimensional soft thresholding operator \mathcal{G}_j is required. This new multidimensional soft thresholding operator \mathcal{G}_τ is defined by

$$\mathcal{G}_\tau(\mathbf{g}) = (\|\mathbf{g}\|_2 - \tau)_+ \frac{\mathbf{g}}{\|\mathbf{g}\|_2} \quad (\text{A.6})$$

where \mathbf{g} is a vector group. For the purposes of the FISTA algorithm incorporating group sparsity, each predefined group is thresholded separately using the \mathcal{G}_τ soft thresholding operator. Therefore, the FISTA algorithm incorporating group sparsity becomes

$$\mathbf{A}_{k+1} = \sum_{i=1}^N \mathcal{G}_{\lambda_{\ell_{2,1}}/2\alpha} \left(\mathbf{Y}_{k,i} - \frac{1}{\alpha} \mathbf{R}^T (\mathbf{S}_i - \mathbf{R}\mathbf{Y}_{k,i}) \right) \quad (\text{A.7})$$

where the index i refers to each separate group of the total N groups, which in the present formulation represents the vectors containing the components \mathbf{x} , \mathbf{y} , and \mathbf{z} for each defined time-frequency location.